Relative Velocity-Based Reward Model for Socially-Aware Navigation with Deep Reinforcement Learning

Vinu Maddumage[1], Sarath Kodagoda[1], Marc G. Carmichael[1], Amal Gunatilake[1], Karthick Thiyagarajan[2] and Jodi Martin[3]

***Abstract*— Mobile robots are increasingly deployed in shared environments where they must learn to navigate alongside humans. Deep Reinforcement Learning (DRL) techniques have shown promise in developing navigation policies that account for interactions within crowds, fostering socially acceptable movement. However, these techniques often depend heavily on collision avoidance rewards to ensure safe navigation. In this study, we introduce a novel reward component based on relative velocity for collision avoidance, which integrates both the robot's and humans' kinematics within personal distance constraints. We conducted a thorough evaluation comparing this new reward model against a conventional one in simulated environments using advanced DRL methods. Our findings indicate that the proposed reward model improves the robots' ability to avoid collisions and navigate towards their goals while being socially acceptable.**

***Index Terms*— Human-Aware Motion Planning, Social HRI, Collision Avoidance, Motion and Path Planning.**

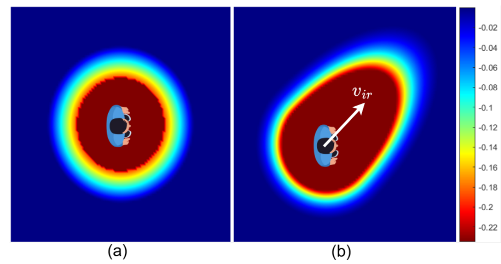

Fig. 1. Collision avoidance reward model component: (a) baseline distance-based model and (b) proposed relative velocity-based model.

I. INTRODUCTION

Advances in perception, navigation, and control technologies have enabled autonomous mobile robots to be successfully deployed across various industries, significantly enhancing task efficiency. This widespread adoption has extended beyond industrial applications, with mobile robots increasingly integrated into service-oriented environments as guide robots. In public spaces such as airports [1], museums [2], [3], and universities [4], [5], mobile robots are used to assist people navigating complex map layouts. Moreover, these guide robots have evolved into assistive devices designed to aid individuals with disabilities that impair their ability to navigate, thereby enhancing their independence [6], [7].

The increasing deployment of mobile robots in these diverse environments has led to more frequent and complex interactions between robots and humans. In these human-shared spaces, safety is highly prioritised regardless of the application [8]. Robots must not only reach their destinations efficiently but also navigate in a way that is socially acceptable and avoids causing discomfort or harm to humans. This necessity introduces unique challenges in robot navigation [9], requiring robots to interpret and respond appropriately to human behaviours and movements.

To address these challenges, various navigation strategies have been developed in the literature. Although reactive navigation methods quickly respond to avoid collisions, they oversimplify human interactions while treating humans as non-responsive obstacles [10], [11]. On the other hand, proactive approaches attempt to predict human behaviours through probability models and expected trajectories, offering more advanced navigation by anticipating future states [12], [13]. However, the inherent complexity and unpredictability of human behaviours [14] can cause robots to hesitate or freeze [15] when facing uncertainties.

Deep Reinforcement Learning (DRL) methods offer a promising approach for developing navigation policies that capture the cooperative nature of human navigation [16]–[18]. Even though social navigation is inherently a Partially Observable Markov Decision Process (POMDP), if we assume full observability of the state, we can apply DRL to train a robot agent to navigate toward a goal while avoiding other agents [19]. DRL allows robots to learn from their interactions with the environment, optimizing for collision avoidance and efficiency. Reward models have been designed to encourage socially acceptable behaviours beyond simple distance-based measures, incorporating elements that better capture social dynamics [20].

Despite significant advancements, the existing literature on DRL-based navigation lacks comprehensive ablation studies focused specifically on the collision avoidance components

This research was supported by the Australian Government through the Australian Research Council's Linkage Projects funding scheme (LP220100430) and the industry partner Guide Dogs NSW/ACT.

[1]Vinu Maddumage, Sarath Kodagoda, Marc G. Carmichael and Amal Gunatilake are with the Robotics Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Broadway, Ultimo NSW 2007, Australia. Email: VinuVihan.Maddumage@student.uts.edu.au; {Sarath.Kodagoda; Marc.Carmichael; Amal.Gunatilake}@uts.edu.au

[2]Karthick Thiyagarajan is with the Smart Sensing and Robotics Laboratory (SensR Lab), Centre for Advanced Manufacturing Technology (CfAMT), School of Engineering, Design and Built Environment (SoEDBE), Kingswood, NSW 2747, Australia. Email: K.Thiyagarajan@westernsydney.edu.au

[3]Jodi Martin is with Guide Dogs New South Wales / Australian Capital Territory, Sydney, New South Wales, Australia. Email: JMartin@guidedogs.com.au

of reward models. This leaves a critical gap in understanding how different aspects of these reward models influence overall performance, especially in dynamic, crowded environments where safe navigation is essential. Addressing this gap is vital for enhancing the safety and effectiveness of these navigation algorithms.

The contributions of the paper are: 1) We propose a novel relative velocity-based reward model (Fig. 1(b)) for addressing above mentioned gaps. Unlike the original distance-based models (Fig. 1(a)), this approach considers both the robot's and human agents' velocities in addition to personal distance constraints, providing a more accurate representation of interaction dynamics via the reward model. 2) We conduct an ablation study using state-of-the-art DRL methods to offer new insights into the effectiveness of the collision avoidance component in these methods. By employing metrics such as success rate and minimum time to collision, our study contributes to a deeper understanding of how different reward model components impact the safety of robot navigation in complex, dynamic environments.

II. Related Work

A. Deep Reinforcement Learning for Social Navigation

Since early work demonstrated the potential of DRL systems [21], these methods have been adapted to various applications, including socially aware robot navigation. The seminal work in [19] formulated collision avoidance in decentralized, non-communicating multi-agent systems as an Markov Decision Process (MDP) based on the observable states of other agents and the robot state. However, this approach was limited to two-agent systems and did not scale effectively in crowded environments. This scalability challenge associated with fixed-size state inputs was addressed by [22], which utilized Long Short-Term Memory (LSTM) networks to embed the crowd state, prioritizing humans based on their proximity to the robot. Although this approach improved scalability, it did not adequately capture the interactions between agents. To overcome this limitation, [16] introduced a network architecture with a social-attentive mechanism, which effectively captures both human-robot and human-human interactions, enabling the robot to learn cooperative behaviours in crowded environments. However, all these methods were not capable of distinguishing between humans and static obstacles. Recognizing this, [17] proposed a network architecture that incorporates an additional channel in the state representation specifically for capturing static obstacles. This enhancement extends the applicability of DRL-based navigation methods to more varied and realistic scenarios.

B. Reward Functions

Reward model design is crucial in DRL-based navigation, shaping the robot's behavior. For socially aware navigation, these models prioritize goal achievement, efficiency, social comfort, and collision avoidance. The robot faces penalties for failing to meet these objectives, except when it successfully reaches the goal. Reward models often include components based on path length [23], [24] and navigation time [24], [25] to improve navigation efficiency, encouraging the robot to find the shortest and fastest route to its goal. To enhance social compliance, various studies have introduced additional components designed to promote socially acceptable behaviours. These include inducing the robot to develop a side preference to mimic human navigation patterns [20], as well as penalizing sudden movements to create smoother interactions and minimize discomfort for nearby humans [17], [23], [25].

Although adherence to social norms can reduce the likelihood of collisions [26], collision avoidance remains the primary component that drives safe behaviour during the learning process. Most existing approaches employ simple distance-based reward models, which penalize the robot for getting too close to obstacles or humans, thus promoting safer navigation paths [16], [17], [19], [22]–[25], [27]–[29]. These position-based models effectively induce passive collision avoidance but often fail to consider the observable velocity of other agents, which is crucial for anticipating and responding to dynamic interactions in real-time.

Very few studies have utilized velocity in reward models. For instance, [30], [31] uses velocity to define an adaptive penalty region, considering only the velocity of the humans. Another study [32] introduces a penalty component based on a desired direction, evaluated using velocity obstacles. However, there are no comprehensive ablation studies specifically focused on these collision avoidance reward components, leaving a gap in understanding their impact on overall navigation performance across different DRL methods.

III. Methodology

A. Relative Velocity-based Reward Model

The proposed reward model is a handcrafted exponential function to assess a penalty from each pedestrian in the crowd based on their relative velocity with respect to the robot. These functions are subjected to distance-based constraints, which allow them to be modelled based on distance parameters related to personal space models of humans.

Considering the observable state $\mathbf{s}^o \in \mathbb{R}^5$ of any agent in the environment to be denoted as $\mathbf{s}^o = \begin{bmatrix}\mathbf{p}^\top & \mathbf{v}^\top & r\end{bmatrix}^\top$, where $\mathbf{p}$ and $\mathbf{v}$ are the position and velocity vectors of the agent and r is its radius (size) [19]. We define the state of the crowd with n pedestrians as, $\mathbf{S} = \bigcup_{i=1}^{n} \mathbf{s}_i^o$, where i denotes the i^{th} human in the crowd. Similarly, by using subscript r to denote the robot, we can evaluate the velocity $\mathbf{v}_{ir}$ of a human relative to the robot and represent it in polar notation:

$$\mathbf{v}_{ir} = \mathbf{v}_i - \mathbf{v}_r = \begin{pmatrix} v_{ir} \\ \theta_{ir} \end{pmatrix} \tag{1}$$

We initially define a frame of reference $\{i\}$ for each human based on $\mathbf{v}_{ir}$ with the x-axis of the frame oriented in the direction θ_{ir} (Fig. 2). The position of the robot $\mathbf{p}_r^{\{i\}}$ in the human reference frame $\{i\}$ can be evaluated as follows,

$$\mathbf{p}_r^{\{i\}} = \begin{bmatrix} x_r^i \\ y_r^i \end{bmatrix} = \begin{bmatrix} \cos\theta_{ir} & \sin\theta_{ir} \\ -\sin\theta_{ir} & \cos\theta_{ir} \end{bmatrix} (\mathbf{p}_r - \mathbf{p}_i) \tag{2}$$

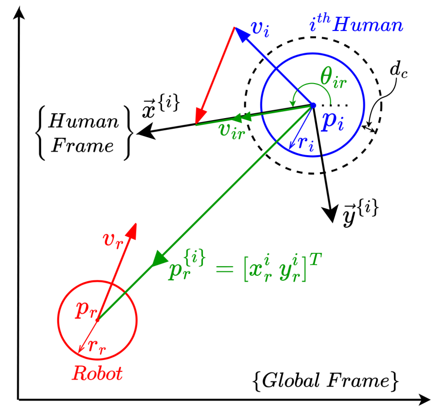

Fig. 2. Defining humans reference frame based on relative velocity.

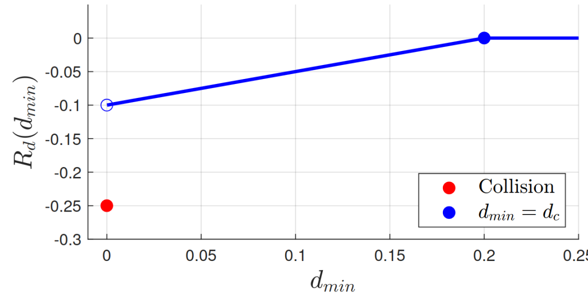

Fig. 3. Minimum distance-based reward.

Based on this, a piece-wise function (3) is defined to model the reward based on the magnitude and direction of the human's velocity relative to the robot.

$$R_i = \begin{cases} A_i \exp\left(-c_i \begin{bmatrix} (v_{ir}+1)^{-\alpha} \\ (v_{ir}+1)^{\beta} \end{bmatrix}^\top \begin{bmatrix} (x_r^i)^2 \\ (y_r^i)^2 \end{bmatrix}\right) & x_r^i \geq 0 \\ A_i \exp\left(-c_i \left(v_{ir}+1\right)^{\beta} \|\mathbf{p}_r^{\{i\}}\|^2\right) & x_r^i < 0 \end{cases} \tag{3}$$

such that when $\mathbf{v}_{ir} = \mathbf{0}$,

$$\|\mathbf{p}_r^{\{i\}}\|^2 = (r_r + r_i)^2 \implies R_i = R_{coll} \tag{4a}$$

$$\|\mathbf{p}_r^{\{i\}}\|^2 = (r_r + r_i + d_c)^2 \implies R_i = R_{min} \tag{4b}$$

where d_c is the comfort distance and $R_{coll} < R_{min} < 0$ are the collision penalty and the minimum penalty offset of the reward model. α and β are hyperparameters that factor the reward model's deformation based on the relative velocity's magnitude. The main focus of α is to tune the decay rate of the exponential reward function in the positive direction of the relative velocity (i.e. R_i when $x_r^i \geq 0$), elongating the reward function in that direction based on the magnitude of the relative velocity. Similarly, β deforms the reward function in the remaining axis directions. The continuity of the reward function is sustained by sharing the same deformation factor for the components in the orthogonal direction. Although the R_{min} has an impact on the decay rate of the function, its primary purpose is to define the boundary of the reward function (4b) while creating an offset between the collision penalty and the minimum reward value of $\hat{R}_i$. Applying distance constraints (4a) and (4b) at relative stationary condition enables the evaluation of the coefficients c_i and A_i as follows.

$$c_i = -\frac{\ln\left(R_{min}/R_{coll}\right)}{d_c\left(2\left(r_r + r_i\right) + d_c\right)} \tag{5}$$

$$A_i = \frac{R_{coll}}{\exp\left(-c_i\left(r_r + r_i\right)^2\right)} \tag{6}$$

The constrained reward function of each pedestrian is then bound between zero and the collision penalty as,

$$\hat{R}_i(\mathbf{s}_r^o, \mathbf{S}) = \min\{\max\{R_i(\mathbf{s}_r^o, \mathbf{S}), R_{coll}\} - R_{min}, 0\}. \tag{7}$$

$$R_v(\mathbf{s}_r, \mathbf{S}) = \begin{cases} R_{comp} & \mathbf{p}_r = \mathbf{p}_g \\ R_{coll} & d_{min} \leq 0 \\ \min_{i\in[n]}\{\hat{R}_i(\mathbf{s}_r^o, \mathbf{S})\} & otherwise \end{cases} \tag{8}$$

Based on this, Equation (8) is used to evaluate the reward value at each time step. A reward $R_{comp} > 0$ is assigned for reaching the robot's goal ($\mathbf{p}_g$). The R_{coll} penalty is given at collisions (i.e. when minimum distance between the robot and any pedestrian $d_{min} \leq 0$) and otherwise based on the state of the system, the minimum reward assigned, considering all pedestrians in the crowd.

B. Reward Model Parameters

The proposed reward model was evaluated considering the distance-based reward model (R_d) as a baseline [19]. Therefore, the same values are considered for parameters common in both reward models. To isolate and better understand the impact of the pedestrian collision component within the reward model, certain penalty components originally included in some of the DRL methods being evaluated have been excluded.

1) Distance-based Model: The distance-based reward model (Fig. 1(a)) used as a baseline for this study comprises a penalty component assigned based on the minimum distance between the robot and any human agent in the environment (d_{min}) (9), where we consider $R_{coll} = -0.25$ and $d_c = 0.2m$ (Fig. 3). In addition, a task completion reward of $R_{comp} = 1$ is assigned upon reaching the goal.

$$R_d(\mathbf{s}_r, \mathbf{S}) = \begin{cases} R_{comp} & \mathbf{p} = \mathbf{p}_g \\ R_{coll} & d_{min} \leq 0 \\ -0.1 + (d_{min}/2) & 0 < d_{min} \leq d_c \\ 0 & otherwise \end{cases} \tag{9}$$

2) Relative Velocity-based Reward Model: In addition to the common parameters from the distance-based model, the proposed reward model introduces three new parameters: α, β and R_{min}. α is considered based on a collision course scenario where two agents walk directly towards each other. Considering an average walking speed of $1.5m/s$, an average response time of $0.5s$ is deemed necessary, resulting in a minimum distance of $2.5m$ between the agents to prevent

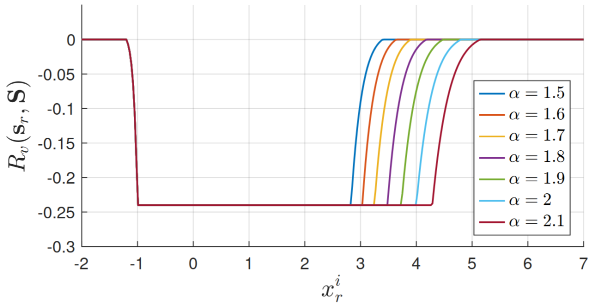

Fig. 4. Relative velocity-based reward along the direction of the relative velocity ($y_r^i = 0$ plane) for varying α ($r_r = 0.5$, $r_i = 0.5$, $\beta = 0$).

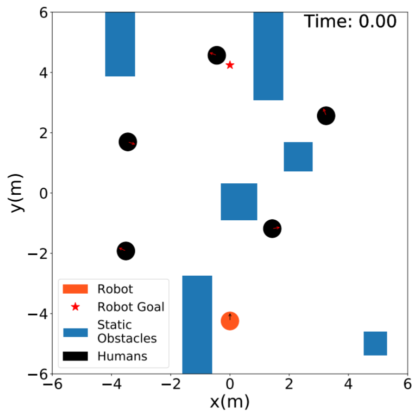

Fig. 5. CrowdNav simulator environment with static obstacles.

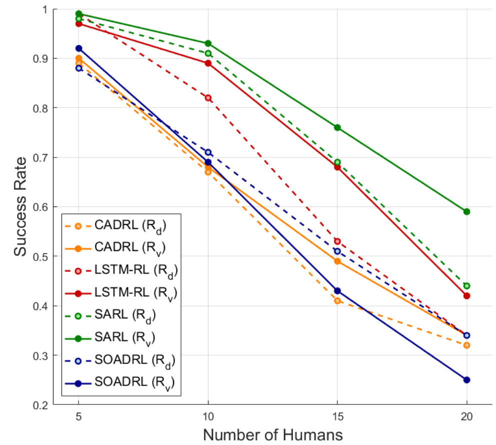

Fig. 6. Success rate (SR) of models trained independently using the R_d and R_v evaluated in varying complexities.

collision by stopping [33]. Fig. 4 shows the effect of α on the reward model in this scenario where the size of both agents, $r = 0.5m$. The required stopping distance corresponds to selecting $\alpha = 1.8$. We consider $\beta = 0.2$ and $R_{min} = -0.01$ in our experiments.

IV. Results and Discussion

To comprehensively analyze the effect of the proposed reward function, we implemented and integrated it with CADRL [19], LSTM-RL [22], SARL [16], and SOADRL [17] methods. These state-of-the-art DRL-based methods were independently trained using the distance-based reward model (R_d) and our proposed reward model (R_v) under the same initialization conditions.

A. Simulation Environment Configuration

The algorithms are trained and evaluated using the circular crossing scenario [19] in a 12x12m open space in the CrowdNav simulator (Fig 5). Human agents, robot and their goals are spawned randomly in an annular region with an inner radius of 2m and an outer radius of 5m. The generated human agents utilize the ORCA algorithm for multi-agent navigation. To simulate non-cooperative behaviour, we introduce randomness in the human agents' ability to perceive the robot [17]. The training environments feature 5 dynamic humans, and each episode considers a time threshold of 25s. The performances of the trained models are then evaluated in environments with different levels of complexity configured by increasing the crowd size of 5, 10, 15 and 20 humans. For each configuration, 500 random test cases initialized based on seed values are used to evaluate the performance.

B. Performance Metrics

We evaluate the performance of the algorithms using two quantitative metrics. We use the success rate (SR) as a success metric and the minimum time to collision (TTC) as a social metric [9].

1) Success Rate: SR indicates the proportion of test cases where the robot successfully reaches its goal location within the specified time threshold.

TABLE I

Success Rate (SR) of Models Trained Independently via the R_d and Proposed R_v Evaluated in Varying Crowd Density.

Method	Reward Model	Humans 5	10	15	20
CADRL	R_d	0.89	0.67	0.41	0.32
	R_v	**0.90**	**0.68**	**0.49**	**0.34**
LSTM-RL	R_d	**0.99**	0.82	0.53	0.34
	R_v	0.97	**0.89**	**0.68**	**0.42**
SARL	R_d	0.98	0.91	0.69	0.44
	R_v	**0.99**	**0.93**	**0.76**	**0.59**
SOADRL	R_d	0.88	**0.71**	**0.51**	**0.34**
	R_v	**0.92**	0.69	0.43	0.25
SOADRL (fine-tuned)	R_d	0.87	0.73	0.57	**0.40**
	R_v	**0.91**	**0.76**	**0.58**	0.37

Fig. 6 shows a steady decline in SR as environmental complexity (number of human agents) increase, highlighting the impact of crowd density on the robot's navigation. Comparing the effects of R_v and R_d, we observe that except for the SOADRL method, all other evaluated methods (CADRL, LSTM-RL, and SARL) generally demonstrate higher SR when trained with R_v (Fig. 7). Notably, in scenarios with 5 human agents, the LSTM-RL method trained with R_d shows a marginally higher SR compared to R_v, but this difference diminishes as the crowd size increases.

Interestingly, the SOADRL method did not follow this trend, showing a lower SR when trained with R_v in more complex environments than those used during its initial

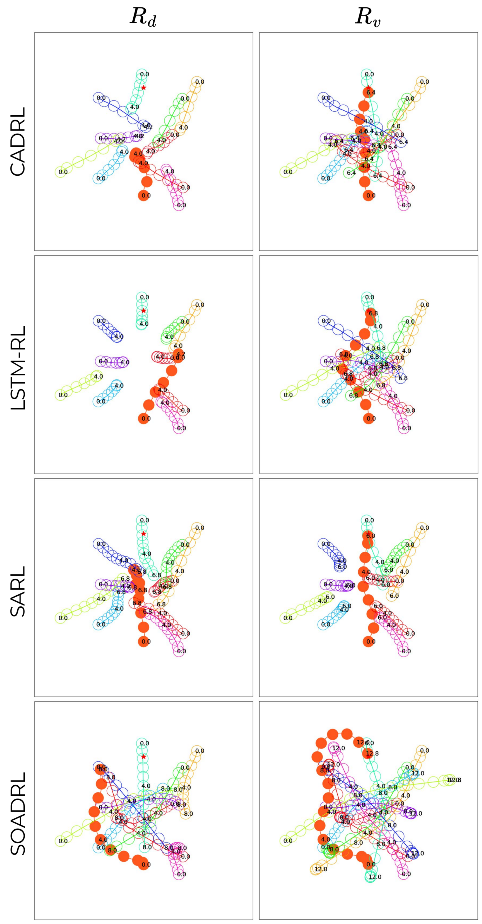

Fig. 7. Performance of Robot (solid red circle) navigating towards its goal (red star) using methods trained with R_d and R_v in an open space with ten human agents with common initialization conditions.

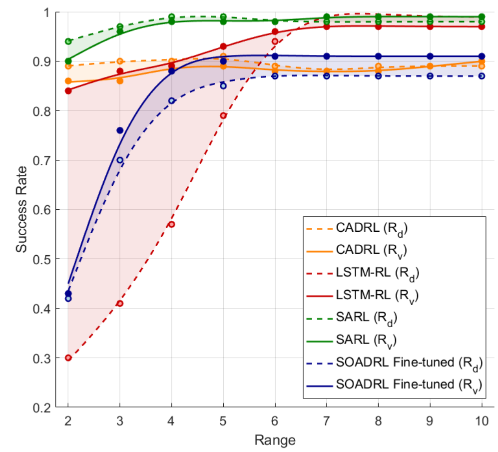

Fig. 8. Success rate (SR) of methods trained with R_d and R_v under range restriction in an open environment with five humans.

training. We hypothesized that this discrepancy is due to the model architecture of the SOADRL method, which includes an additional channel specifically aimed at capturing static objects in the environment. Since the initial training was conducted in open space scenarios for a fair comparison, the model parameters associated with this channel were not effectively trained. Consequently, the SOADRL method's ability to generalize to more complex environments may be impaired, resulting in lower SR. To verify this, we fine-tuned the SOADRL method in an environment with static obstacles (introduced 6 random static obstacles as in Fig. 5). The performance of the fine-tuned SOADRL model was reevaluated in the open space configuration (Table I), and a clear improvement was observed with R_v in moderately crowded scenarios, though this declined in highly dense scenarios. Fine-tuning the other methods in similar configurations by decomposing static obstacles as static humans drastically reduced the performance of these methods.

The improved performance of R_v in most cases highlights that incorporating relative velocity information into the reward model provides a better account for the dynamic interactions between the robot and human agents. This allows the algorithms to make more informed decisions, leading to higher SR, especially in moderately crowded scenarios.

Furthermore, we also analyzed how restricting the robot's visibility range impacted the success of each method. Since all methods, except SOADRL, were not designed to handle scenarios without humans, we restricted the range by distance but included the closest human as the crowd state when no humans were detected within the defined range. This ensured that the crowd state was never empty. The visibility range restriction logic was applied and evaluated in an environment with five humans across all methods, including SOADRL, to ensure a fair comparison.

Overall, except for the LSTM-RL method, a similar trend is observed between R_d and R_v for each method under range restriction (Fig. 8). The CADRL method was largely unaffected due to its inherent design as a two-agent system. The SARL method consistently maintained high performance across all ranges for both reward models, while the fine-tuned SOADRL method showed noticeable drops in SR at closer ranges for both reward models.

However, the LSTM-RL method with R_d had a significant performance drop at ranges below 6 meters, highlighting limitations in handling restricted ranges. Notably, the same method trained with the R_v performed much better under restricted visibility, suggesting that the R_v model improves the method's emphasis on nearby humans, compensating for the shortcomings observed with the R_d model.

2) Minimum Time to Collision: TTC serves as a metric for evaluating the social aspect of the robot's navigation [9], [34]. It measures the minimum time for potential collisions between the robot and any human agent, assuming all agents maintain their current linear motion at a given instance. To ensure a realistic assessment, our TTC evaluation accounts for the physical sizes of the robot and the human agents.

As shown in Table II, the average TTC values of most algorithms using R_v are higher compared to their R_d counterparts. This indicates that the robot maintains safer velocities and distances (SOADRL method in Fig. 7) from the human agents, thus enhancing social compliance and reducing the likelihood of collisions.

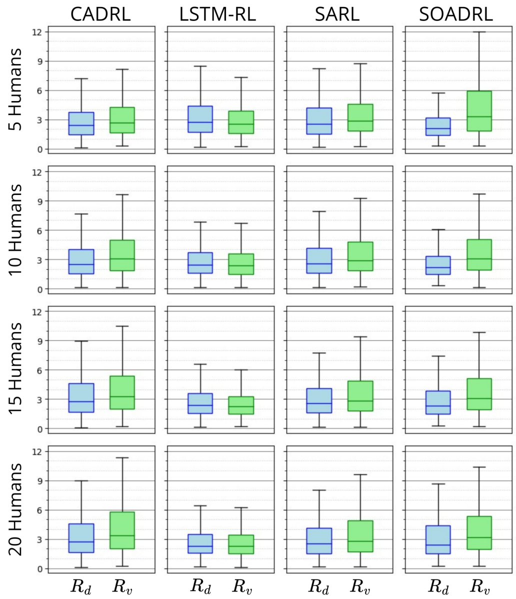

Fig. 9. Box plot representation of TCC of models trained using the R_d (blue) and R_v (green) evaluated in varying complexities.

TABLE II

AVERAGE TTC OF METHODS TRAINED INDEPENDENTLY USING R_d AND PROPOSED R_v EVALUATED IN VARYING COMPLEXITIES.

Method	Reward Model	Humans 5	10	15	20
CADRL	R_d	3.03	3.20	3.60	3.60
	R_v	3.40	3.87	4.14	**4.31**
LSTM-RL	R_d	**3.49**	3.06	3.02	2.84
	R_v	3.19	2.97	2.70	2.74
SARL	R_d	3.31	3.36	3.34	3.29
	R_v	3.61	3.83	**3.88**	3.80
SOADRL	R_d	2.67	2.87	3.39	3.64
	R_v	4.26	3.96	4.20	**4.32**

Fig. 9 illustrates a box plot representation of TTC values across the test cases, offering insight into the distribution of these TTC values. For the CADRL, SARL, and SOADRL methods, the TTC values tend to skew towards higher values when using R_v, indicating that these methods benefit significantly from the incorporation of relative velocity information in their reward functions. By integrating R_v, these methods can more effectively assess and respond to dynamic changes.

In contrast, the LSTM-RL method exhibits a slightly different trend. While LSTM-RL trained with R_v shows improvements in success rate, its TTC values are marginally lower compared to those trained with R_d. However, the overall difference in TTC between R_v and R_d for LSTM-RL remains small relative to the gains observed in the other methods. We believe that this difference in the LSTM-RL method's TTC performance stems from its focus on nearby human agents noted in the success rate analysis (Fig. 8). The LSTM processes human agents based on their distance from the robot, beginning with the furthest agent. This ordering increases the emphasis on agents in close proximity. The R_v reward model further amplifies this focus. As a result, while this method maintains high success rates with R_v, the robot may be making more aggressive maneuvers when near

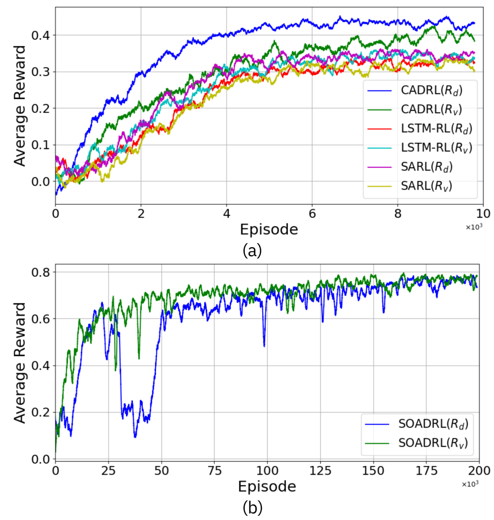

Fig. 10. Training (a) CADRL, LSTM-RL and SARL (b) SOADRL

human agents, which could explain the slightly lower TTC values observed with R_v.

Overall, the higher TTC values observed with Rv in the majority of cases reinforce the benefit of incorporating relative velocity into the reward model. This allows the robot to gain a better understanding of dynamic interactions with surrounding agents, leading to more socially compliant and safer navigation behaviours.

C. Training Convergence

Furthermore, we analysed the training convergence of the methods trained under each reward setting (Fig. 10). Notably, the SOADRL method required significantly more episodes to converge compared to the other methods. Additionally, a significant, sharp drop in average reward was observed during early episodes with R_d, whereas R_v demonstrated a much more stable increase, indicating smoother learning progression (Fig. 10(b)).

V. FUTURE WORK

For future work, we plan to evaluate performance using a broader set of metrics and analyze its effectiveness across different scenarios. Additionally, we are in the process of implementing our approach on the Spot robot [35] to assess the impact of real-world noise on perception.

VI. CONCLUSIONS

We proposed a novel relative velocity-based reward model, subject to personal distance constraints, for collision avoidance in DRL navigation systems, providing a better understanding of interaction dynamics. Our results demonstrated the reward model enhanced both the safety and social acceptability of robot behaviours. Furthermore, our work emphasizes the importance of carefully designing reward components and selecting suitable network architectures in DRL to enhance navigation performance in complex, dynamic settings. This advancement offers a meaningful step toward developing more reliable and socially aware navigation systems.

References

[1] R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila, M. Chetouani, D. Cremers, V. Evers, and M. Fiore, "Spencer: A socially aware service robot for passenger guidance and help in busy airports," in *Field and Service Robotics: Results of the 10th International Conference*. Springer, 2015, Conference Proceedings, pp. 607–622.

[2] W. Burgard, A. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "The interactive museum tour-guide robot," *Proceedings of AAAI-98*, pp. 11–18, 01 1998.

[3] S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "Minerva: a second-generation museum tour-guide robot," in *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, vol. 3, 1999, Conference Proceedings, pp. 1999–2005 vol.3.

[4] M. Liu, Z. Hou, Z. Sun, N. Yin, H. Yang, Y. Wang, Z. Chu, and H. Kong, "Campus guide: A lidar-based mobile robot," in *2019 European Conference on Mobile Robots (ECMR)*, 2019, pp. 1–6.

[5] B. Greenberg, D. Nakhimovich, R. Magnotti, H. Purohit, S. Shah, A. S. Kulkarni, U. Gonzalez-Bravo, and N. R. Carver, "Development of a socially cognizant robotic campus guide," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2024, Conference Proceedings, pp. 1229–1232.

[6] V. K. Narayanan, A. Spalanzani, and M. Babel, "A semi-autonomous framework for human-aware and user intention driven wheelchair mobility assistance," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, Conference Proceedings, pp. 4700–4707.

[7] K. Thiyagarajan, S. Kodagoda, M. Luu, T. Duggan-Harper, D. Ritchie, K. Prentice, and J. Martin, "Intelligent guide robots for people who are blind or have low vision: A review," *Vision Rehabilitation International*, vol. 13, no. 1, pp. 1–15, 2022. [Online]. Available: https://doi.org/10.2478/vri-2022-0003

[8] N. Tsoi, J. Romero, and M. Vázquez, "How do robot experts measure the success of social robot navigation?" in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 1063–1066. [Online]. Available: https://doi.org/10.1145/3610978.3640636

[9] A. Francis, C. Pérez-D'Arpino, C. Li, F. Xia, A. Alahi, R. Alami, A. Bera, A. Biswas, J. Biswas, R. Chandra, H.-T. L. Chiang, M. Everett, S. Ha, J. Hart, J. P. How, H. Karnan, T.-W. E. Lee, L. J. Manso, R. Mirksy, S. Pirk, P. Teja Singamaneni, P. Stone, A. V. Taylor, P. Trautman, N. Tsoi, M. Vázquez, X. Xiao, P. Xu, N. Yokoyama, A. Toshev, and R. Martín-Martín, "Principles and Guidelines for Evaluating Social Robot Navigation Algorithms," *arXiv preprint*, p. arXiv:2306.16740, Jun. 2023.

[10] F. Farina, D. Fontanelli, A. Garulli, A. Giannitrapani, and D. Prattichizzo, "Walking ahead: The headed social force model," *PLOS ONE*, vol. 12, p. e0169734, 01 2017.

[11] D. T. Nguyen, V. B. Hoang, T. N. Le, T. H. Nguyen, and X. T. Truong, "Social hybrid reciprocal velocity obstacle-based socially aware mobile robot navigation framework," in *Industrial Networks and Intelligent Systems*, N.-S. Vo, V.-P. Hoang, and Q.-T. Vien, Eds. Springer International Publishing, 2021, pp. 301–314.

[12] X.-T. Truong and T. D. Ngo, "Toward socially aware robot navigation in dynamic and crowded environments: A proactive social motion model," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 4, pp. 1743–1760, 2017.

[13] M. Kamezaki, Y. Tsuburaya, T. Kanada, M. Hirayama, and S. Sugano, "Reactive, proactive, and inducible proximal crowd robot navigation method based on inducible social force model," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3922–3929, 2022.

[14] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: a survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020. [Online]. Available: https://doi.org/10.1177/0278364920917446

[15] P. Trautman and A. Krause, "Unfreezing the robot: Navigation in dense, interacting crowds," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 797–803.

[16] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, "Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning," in *IEEE International Conference on Robotics and Automation*, 2019, Conference Proceedings, pp. 6015–6022.

[17] L. Liu, D. Dugas, G. Cesari, R. Siegwart, and R. Dube, "Robot navigation in crowded environments using deep reinforcement learning," in *IEEE International Conference on Intelligent Robots and Systems*, 2020, Conference Proceedings, pp. 5671–5677.

[18] Y. Chen, C. Liu, B. E. Shi, and M. Liu, "Robot navigation in crowds by graph convolutional networks with attention learned from human gaze," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2754–2761, 2020.

[19] Y. F. Chen, M. Liu, M. Everett, and J. P. How, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *IEEE International Conference on Robotics and Automation*, 2017, Conference Proceedings.

[20] Y. F. Chen, M. Everett, M. Liu, and J. P. How, "Socially aware motion planning with deep reinforcement learning," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2017-September, 2017, Conference Proceedings, pp. 1343–1350.

[21] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," 2013. [Online]. Available: https://arxiv.org/abs/1312.5602

[22] M. Everett, Y. F. Chen, and J. P. How, "Motion planning among dynamic, decision-making agents with deep reinforcement learning," in *IEEE International Conference on Intelligent Robots and Systems*, 2018, Conference Proceedings.

[23] P. Long, T. Fan, X. Liao, W. Liu, H. Zhang, and J. Pan, "Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 6252–6259.

[24] C. Pérez-D'Arpino, C. Liu, P. Goebel, R. Martín-Martín, and S. Savarese, "Robot navigation in constrained pedestrian environments using reinforcement learning," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 1140–1146.

[25] X. Huang, H. Deng, W. Zhang, R. Song, and Y. Li, "Towards multi-modal perception-based navigation: A deep reinforcement learning method," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4986–4993, 2021.

[26] C. Yao, C. Liu, M. Liu, and Q. Chen, "Navigation in multi-agent system with side preference path optimizer," *IEEE Access*, vol. 9, pp. 113 944–113 953, 2021.

[27] S. Pohland, A. Tan, P. Dutta, and C. Tomlin, "Stranger danger! identifying and avoiding unpredictable pedestrians in rl-based social robot navigation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, Conference Proceedings, pp. 15 217–15 224.

[28] M. Everett, Y. F. Chen, and J. P. How, "Collision avoidance in pedestrian-rich environments with deep reinforcement learning," *IEEE Access*, vol. 9, pp. 10 357–10 377, 2021.

[29] R. Guldenring, M. Görner, N. Hendrich, N. J. Jacobsen, and J. Zhang, "Learning local planners for human-aware navigation in indoor environments," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 6053–6060.

[30] E. E. Montero, H. Mutahira, N. Pico, and M. S. Muhammad, "Dynamic warning zone and a short-distance goal for autonomous robot navigation using deep reinforcement learning," *Complex and Intelligent Systems*, vol. 10, no. 1, pp. 1149–1166, 2024.

[31] S. S. Samsani and M. S. Muhammad, "Socially compliant robot navigation in crowded environment by human behavior resemblance using deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5223–5230, 2021.

[32] Z. Xie and P. Dames, "Drl-vo: Learning to navigate through crowded dynamic scenes using velocity obstacles," *IEEE Transactions on Robotics*, vol. 39, no. 4, pp. 2700–2719, 2023.

[33] M. Wood, T. Ayres, R. Kelkar, and R. Khatua, "Walking and jogging: An analysis of pedestrian stopping times and distances," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 54, no. 19, pp. 1435–1439, 2010.

[34] A. Biswas, A. Wang, G. Silvera, A. Steinfeld, and H. Admoni, "Socnavbench: A grounded simulation testing framework for evaluating social navigation," *ACM Transactions on Human-Robot Interaction*, vol. 11, no. 3, 2022.

[35] Boston Dynamics, "Spot – The Agile Mobile Robot," 2025, accessed: Feb. 24, 2025. [Online]. Available: https://bostondynamics.com/products/spot/