

HeR-DRL:Heterogeneous Relational Deep Reinforcement Learning for Decentralized Multi-Robot Crowd Navigation

Xinyu Zhou¹, Songhao Piao^{1*}, Wenzheng Chi², Liguo Chen², and Wei Li²

Abstract—Crowd navigation has received significant research attention in recent years, especially DRL-based methods. While single-robot crowd scenarios have dominated research, they offer limited applicability to real-world complexities. The heterogeneity of interaction among multiple agent categories, like in decentralized multi-robot pedestrian scenarios, are frequently disregarded. This “interaction blind spot” hinders generalizability and restricts progress towards robust navigation algorithms. In this paper, we propose a heterogeneous relational deep reinforcement learning(HeR-DRL), based on customised heterogeneous GNN, in order to improve navigation strategies in decentralized multi-robot crowd navigation. Firstly, we devised a method for constructing robot-crowd heterogeneous relation graph that effectively simulates the heterogeneous pair-wise interaction relationships. We proposed a new heterogeneous graph neural network for transferring and aggregating the heterogeneous state information. Finally, we incorporate the encoded information into deep reinforcement learning to explore the optimal policy. HeR-DRL are rigorously evaluated through comparing it to state-of-the-art algorithms in both single-robot and multi-robot circle crowssing scenario. The experimental results demonstrate that HeR-DRL surpasses the state-of-the-art approaches in overall performance, particularly excelling in safety and comfort metrics. This underscores the significance of interaction heterogeneity for crowd navigation. The source code will be publicly released in <https://github.com/Zhouxy-Debugging-Den/HeR-DRL>.

I. INTRODUCTION

Recently, to fulfill essential daily service needs for individuals, the frontiers of mobile robot system research has transitioned from unmanned factories to dynamic environments coexisting with humans, such as offices[1], hospitals[2], canteens[3] and other public places. Robot navigation is evolving, aiming for superior safety, efficiency, and user comfort. This comfort focus prioritizes minimizing disruptions and managing stress[4]. Nevertheless, navigating crowded, unpredictable environments remains challenging due to complex pedestrian movements and the difficulty of modeling comfort.

The research on navigation in pedestrian-rich scenarios is frequently called social navigation or crowd navigation. In this paper, we will refer to it as crowd navigation for convenience. Throughout the past 30 years, the solutions to

¹X. Zhou and S. Piao are with the Faculty of Computing, Harbin Institute of Technology, Harbin, 150001 China. E-mail: zhouxy@stu.hit.edu.cn, piaosh@hit.edu.cn.

²W. Chi, L. Chen and W. Li are with Soochow University, Soochow, 215031, China. E-mail: wzchi@suda.edu.cn, chenliguo@suda.edu.cn, li7713@163.com

*S. Piao is the Corresponding author.

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

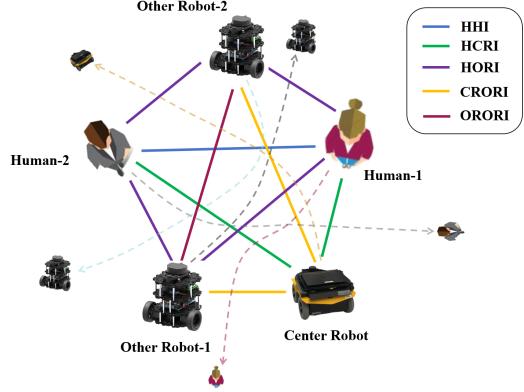


Fig. 1: Heterogeneous interaction relationships in decentralized multi-robot crowd navigation. The center robot is controlled by trained policy, while the other robots controlled by other unknown policies. Consequently, there are five heterogeneous pair-wise interactions: HHI refers to the interaction specifically occurring among human, HCRI refers to the interaction between human and the center robot, HORI refers to the interaction between human and other robots, CRORI refers to the interaction between the center robot and other robots, and ORORI refers to the interaction specifically occurring among other robot.

crowd navigation have been generally classified into three primary categories: reaction-based, trajectory-based, and DRL-based methods. Reaction-based methods automate immediate actions against obstacles based on predefined rules, i.e. RVO[5], ORCA[21], SFM[6], IGP[7]. However, its decision-making process primarily hinges on the current state, highly likely triggering the short-sighted behaviour. Moreover, there’s a risk of “reciprocal dance” on account of neglecting pedestrian reactions[8]. Trajectory-based methods devise the feasible planning accordingly after predicting the intended trajectories of other agents [9], [10], [11], [12], [13]. Unfortunately, the computational expense of online prediction and path search in a vast state space can be significant[14]. In addition, predicted trajectories enlarge the spatially infeasible region, prone to result in overly conservative robot movement[15] and the more serious “freezing robot problem”[7].

Deep Reinforcement Learning (DRL) offers an alternative method to implicitly merge interactive prediction and planning[16], [22], [17], [19], [12], [20]. This effectively shifts the burdensome online computation to an offline train-

ing procedure. Crowd navigation exhibit high dynamism and complexity, making the effective extraction of interaction information a challenging endeavor. Numerous explorations have structured deep learning approaches for encoding interaction information, through employing various models, such as LSTM (Long Short-Term Memory)[22], Self-Attention Mechanisms[17], GCN (Graph Convolutional Networks)[19], [12], Transformers[20], and more.

However, these studies often assume that the relations among agents are homogenous, disregarding the inherent heterogeneity originating from the discrepancy in the interaction relationships. This will lead to under-constructed associations between agents in crowd navigation, and social performance will inevitably be limited. The more complex the scenario, the more pronounced the impact of heterogeneity. In particular, a few recent studies[23], [24] have focused on multi-robot crowd navigation, mainly in view of the low efficiency of single-robot crowd navigation. Regrettably, these studies primarily tackle the issue of multi-agent cooperative control without concentrating on the heterogeneity of complex interactions. Therefore, this paper introduces the pioneering notions of heterogeneous interaction relationships in decentralized multi-robot crowd navigation, as depicted in Fig. 1. Building upon this viewpoint, a heterogeneous GNN-based approach called HeR-DRL is proposed. The primary contributions of this letter are outlined as follows.

- 1) We introduce a innovative heterogeneous relational deep reinforcement learning(HeR-DRL) to represent the interactive states. HeR-DRL attains safe and comfort performances in the multi-robot crowded scenarios.
- 2) We constructed a novel form of heterogeneous graph neural network to extract diverse pairwise spatial interactions. New state information for each category is configured to enhance the network in effectively encodding the heterogeneity of interactions.
- 3) A decentralised multi-robot 2D circle crossing simulation is innovatively designed. Experimental results in a variety of configurations demonstrate that our approach outperforms previous state-of-the-art methods in safety and comfort, and presents exceptionally superior generalization.

II. RELATED WORKS

In 2017, Yufan Chen et al.[16] introduced the CADRL method, pioneering a DRL-based algorithm to illustrate interactions between pairs of humans and robots. However, its use of a fixed-dimension input in a typical feed-forward neural network led to decreased scene adaptability due to the variable number of agents in the scenario. Subsequently, in 2018, Michael Everett et al.[22] proposed the GA3C-CADRL algorithm, addressing the CADRL's fixed dimensionality issue by employing LSTM to encode spatial relationships of HRI(Human-Robot Interaction). Nevertheless, this approach overlooked the presence of HHI(Human-Human Interaction) in the scenario. Consequently, Changan Chen et al.[17] revisited the pairwise interaction and proposed a new approach called SARL, utilizing self-attention to jointly model HRI and HHI.

Although this method only provided a locally coarse-grained characterization of HHI, it exhibited promising performance. Building on prior work, Weixian Shi et al.[18] propose the ESA by amalgamating LSTM and self-attention mechanisms, leveraging the spatial graph as a parallel branch of the modified attention graph. However, most of the aforementioned methods primarily concentrate on modeling HRI spatial interactions using various neural networks, with comparatively lesser emphasis on HHI spatial interactions.

As research progressed, successive studies began adopting a relational graph perspective to address this challenge. Changan Chen et al.[12] , for instance, introduces a relational graph to understand the relationship between HHI and HRI, employing a bi-layered GCN for computing interactions among agents, showcasing superior performance compared to SARL. Yuying Chen et al.[25] utilizes a bi-layered GCN to train an attention network, predicting human attention. Xueyou Zhang et al.[19] captures spatial relationships of HRI and HHI using GCNs. Zhiqian Zhou et al.[26] explores integrating social attention mechanisms to more effectively extract spatial relationships between HHI and HRI using GAT(Graph Attention Network). While the mentioned approach encodes HHI and HRI spatial interactions, it doesn't explicitly differentiate between the relationship characteristics inherent in HRI and HHI.

Subsequent research has emphasized that agent associations aren't limited to the spatial domain, changes in the temporal domain can also significantly impact interactions. Shuijing Liu et al.[27] proposed a new method called DSRNN, which utilises capturing spatio-temporal interactions between robots and multiple pedestrians. While it integrates spatio-temporal features of human-robot interaction to some degree, there's limited description of spatio-temporal features of HHI. The later proposed a new approach incorporates intent information and employs temporal edges to establish a comprehensive social environment graph, aiming to portray potential associations of multi-robot systems[28]. Despite integrating spatio-temporal features of HHI and HRI, compared to DSRNN, it doesn't adequately model potential spatio-temporal associations. Additionally, recent approaches employing transformers for feature representation, such as ST²[20], exhibit remarkable performance in encoding spatio-temporal features and associations of HHI and HRI. Nonetheless, these approaches haven't yet explored the heterogeneity present in the multi-agent crowd navigation. Kai Zhu et al.[29] introduced DH-SARL, an approach utilizing OBC(Orientated Bounding Capsule) and velocity-related collision risk function to capture the shape and velocity heterogeneity of agents. Aided by reward shaping, it optimizes the success rate and security of SARL. However, this method disregards interaction heterogeneity, where various types of interactions affect each other differently. Consequently, to fill the research gap, this paper concentrates on investigating the spatial heterogeneity within interaction relations, specifically in decentralized multi-robot crowd navigation.

III. METHOD

Crowd navigation can be conceptualized as a POMDP (Partially Observable Markov Decision Process). The concrete

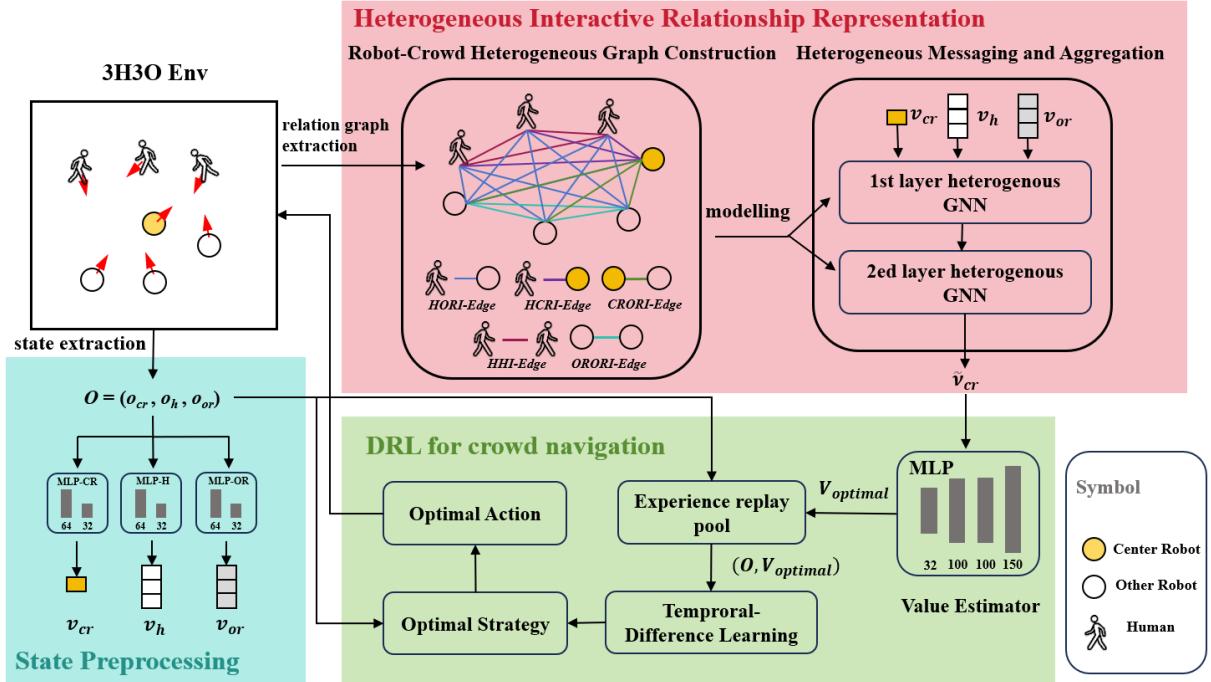


Fig. 2: The overall framework of our HeR-DRL. Here is an illustration of a multi-robot crowd scenario with a 3H3O configuration.

workflow of the method is illustrated in Fig. 2. Step one involves extracting agents state information from the scenario and embedding it into the same-dimensional vector. In the second step, a robot-crowd heterogeneous relation graph will be constructed to transfer and aggregate heterogeneous message using two-layer heterogenous graph neural network(heterogenous GNN). Ultimately, center robot node incorporated with interaction relations of the neighbour agents, are utilized as input for the value network to acquire optimal strategy.

A. State Information Preprocessing

In this section, the significant state information within a multi-robot crowd navigation are initially parameterised. These details are preprocessed as node information for the subsequent construction of the robot-crowd heterogeneous relation graph.

1) *State Parameterisation*: In previous generic state parameterization definitions[17], [20], [12], the state space comprises both observable and unobservable states of all agents. The observable state includes the velocity $\mathbf{v} = [v_x, v_y]$, position $\mathbf{p} = [p_x, p_y]$ and the radius r of the agent itself. Meanwhile, the unobservable state comprises the target position $\mathbf{p}_g = [g_x, g_y]$, preferred velocity v_{pref} and the direction angle ψ . Compared to single-robot crowd navigation, decentralised multi-robot crowd navigation introduces an additional category of agents, namely "other robots", as shown in Fig. 1. To make a distinction, the robot our policy control are designated as the "center robot". The center robot is fully observable, whereas the humans and the other robots are partially observable. We use o_{cr} , $o_h = [o_{h_1}, o_{h_2}, \dots, o_{h_n}]$, and $o_{or} = [o_{or_1}, o_{or_2}, \dots, o_{or_m}]$

to represent the state of the center robot ,the humans and other robots respectively.

$$\begin{aligned} o_{cr} &= [d_g, v_{pref}, \theta, r, v_x, v_y], \\ o_{h_i} &= [p_x^{h_i}, p_y^{h_i}, v_x^{h_i}, v_y^{h_i}, r^{h_i}, d_a^{h_i}, r^{h_i} + r, c^{h_i}], \\ o_{or_j} &= [p_x^{or_j}, p_y^{or_j}, v_x^{or_j}, v_y^{or_j}, r^{or_j}, d_a^{or_j}, r^{or_j} + r, c^{or_j}], \end{aligned} \quad (1)$$

where o_h^i is the observable state of the i -th human, o_{or}^j is the observable state of the j -th other robot, $d_g = \|\mathbf{p} - \mathbf{p}_g\|_2$ is the center robot's distance to the goal, $d_a^* = \|\mathbf{p} - \mathbf{p}_*\|_2$ is the robot's distance to the agent *. Notably, compared to previous studies[17], [20], this paper proposes a new state parameter, the agent category c^* which can be received by object detection[30] in the real world. 1 if human, 0 if other robot. This parameter is established with the attention to capture the category heterogeneity of all agents except center robot, thus boosting the extraction of information about the heterogeneous relations.

2) *State Embedding*: In this section, o_{cr} , o_h , o_{or} are embedded into a fixed-length embedding vector v_{cr} , v_h , v_{or} separately, employed as the node of the subsequent heterogenous relation graph.

$$\begin{aligned} v_{cr} &= f_{cr}(o_{cr}, W_{cr}^{embedding}), \\ v_h &= f_h(o_h, W_h^{embedding}), \\ v_{or} &= f_{or}(o_{or}, W_{or}^{embedding}), \end{aligned} \quad (2)$$

where f_{cr} , f_h and f_{or} are the distinct 2-layer MLP with ReLU activations, the weights are denoted by $W_{cr}^{embedding}$, $W_h^{embedding}$ and $W_{or}^{embedding}$ distinctly.

B. Heterogeneous Interactive Relationship Representation

This part focuses on capturing complex interactions among agents based on relational graph learning[12]. Firstly, We introduce a method for constructing the robot-crowd heterogeneous graph, facilitating the extraction of diverse categories of interactions. In the second step, we will innovatively employ a heterogeneous graph neural network approach to characterisation of robot-crowd heterogeneous graph for information transfer among different nodes.

1) *Robot-Crowd Heterogeneous Graph Construction:* In previous graph-based studies[12], [25], [19], graphs were typically constructed under the assumption of homogeneity. While these graphs encoded both HRI and HHI, they failed to differentiate between these two types of interactions. Consequently, Limited information encoding hampered the network's effectiveness, potentially hindering its performance in decentralized multi-robot crowd navigation. Real-world graphs often exhibit a diverse array of nodes and edges, a concept widely recognized as a heterogeneous information network(HIN)[35]. In this context, we construct the robot-crowd heterogeneous graph, aiming to comprehensively capture the intricacies inherent in multi-robot crowd navigation scenarios.

In this letter, the robot-crowd heterogeneous graph, denoted as the $\mathcal{G}_{general} = (\mathcal{A}, \mathcal{L})$, consists of an agents set \mathcal{A} and a link set \mathcal{L} . The robot-crowd heterogeneous graph is also associated with a node type mapping function $\varsigma : \mathcal{A} \rightarrow \mathcal{V}$ and a link type mapping function $\tau : \mathcal{L} \rightarrow \mathcal{E}$. where \mathcal{V} is the set of types of embedding features and \mathcal{E} is the set of types of edges. This mapping function ς is actually the state embedding part mentioned above, so the $\mathcal{V} = \{v_{cr}, v_h, v_{or}\}$. As illustrated in Fig. 1, this letter proposes five distinct types of pair-wise interactions. when constructing the robot-crowd heterogeneous graph, a total of five different edges exist, as shown in Fig. 2. so the $\mathcal{E} = \{\mathcal{E}_{HORI}, \mathcal{E}_{HCRI}, \mathcal{E}_{CRORI}, \mathcal{E}_{HHI}, \mathcal{E}_{ORORI}\}$.

Next, We break down the robot-crowd interaction network(composed of all pairwise interactions) into five smaller networks, each focusing on a specific type of interaction. These five homogeneous subgraphs actually are represented as the five relations in Fig. 3. For ease of expression, we define the subgraph \mathcal{G}_{CRHI} to be the relation between the center robot and the humans, the subgraph \mathcal{G}_{ORHI} to be the relation between the other robots and the humans, the subgraph \mathcal{G}_{CRORI} to be the relation between the center robot and the other robots, the subgraph \mathcal{G}_{HHI} to be the relation among the humans themselves, the subgraph \mathcal{G}_{ORORI} to be the relation among the other robots themselves. The specific formula is as follows.

$$\begin{aligned} \mathcal{G}_{CRHI} &= (\mathcal{V}_{(v_{cr}, v_h)}, \mathcal{E}_{HCRI}), \\ \mathcal{G}_{ORHI} &= (\mathcal{V}_{(v_{or}, v_h)}, \mathcal{E}_{HORI}), \\ \mathcal{G}_{ORCRI} &= (\mathcal{V}_{(v_{or}, v_{cr})}, \mathcal{E}_{CRORI}), \\ \mathcal{G}_{HHI} &= (\mathcal{V}_{(v_h, v_h)}, \mathcal{E}_{HHI}), \\ \mathcal{G}_{ORORI} &= (\mathcal{V}_{(v_{or}, v_{or})}, \mathcal{E}_{ORORI}), \end{aligned} \quad (3)$$

In the above equation, $\mathcal{V}_{(*,*)}$ refers to the node category of each pair of interaction relations.

$$\tilde{v}_{cr} = F_{Hetero}^2(F_{Hetero}^1((v_{cr}, v_h, v_{or}))), \quad (4)$$

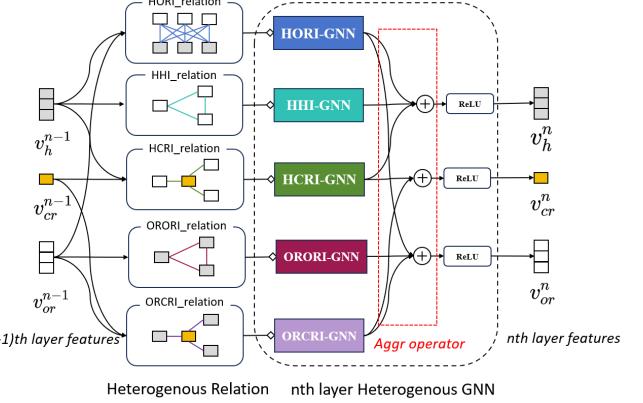


Fig. 3: Diagram of the nth-layer heterogeneous GNN with a 3H3O configuration.

where \tilde{v}_{cr} is the center robot node-level feature after the processing of heterogeneous messaging and aggregation.

2) *Heterogeneous Messaging and Aggregation:* In previous studies, many models like GCN[12], [25] and GAT[26] were commonly employed to handle data aggregation in relational graphs. However, these graph neural networks are limited in their ability to accommodate diverse node and edge types, making them suitable only for homogeneous graphs. In this section we propose a new heterogeneous GNN for processing the robot-crowd heterogeneous graph constructed in the previous section.

First, inspired by [31], we apply the following graphical neural network (GNN) operator to each of the five different types of homogeneous relation subgraphs. The rationale behind GNN lies in the concept that graphs can be viewed as frameworks for "information transfer" algorithms between nodes, aligning more with the core principles of this letter's heterogeneous interaction information setting, rather than GCN or GAT which mere aggregations of information from neighboring nodes. The following outlines the computation process of a layer of our GNN operators.

$$h^{(n)}(v) = h^{(n-1)}(v) \cdot W_1^{(n)} + \sum_{w \in N(v)} h^{(n-1)}(w) \cdot W_2^{(n)}, \quad (5)$$

where $h^{(n-1)}(v)$ is the v node-level feature of layer $n - 1$, $h^{(n-1)}(w)$ is the v 's neighbour node-level feature at layer $n - 1$. W_1^n , W_2^n are layer-specific trainable weight matrix when passed from layer $n - 1$ to layer n respectively.

We employ the mentioned GNN operators for each heterogeneous subgraph, forming ORCRI-GNN, HCRI-GNN, ORORI-GNN, HHI-GNN, ORHI-GNN, respectively. We can find that after a layer of GNN computation, the information of pairs-wise nodes transfers and aggregates with each other in the each homogeneous relation subgraph. Inspired by R-GCNs[32], we devised the effective heterogenous aggregation method for node information from different pairs-wise subgraphs after messaging, As shown in Fig. 3. The nth layer

heterogenous GNN is calculated as follows.

$$F_{Hetero}^n = \text{ReLU}(\text{Aggr}_{i \in \mathcal{E}}(\sum_{v_i \in \mathcal{G}_i} (h^{(n-1)}(v_i) \cdot W_{1i}^{(n)}) + \sum_{w_i \in N(v_i)} h^{(n-1)}(w_i) \cdot W_{2i}^{(n)})), \quad (6)$$

where F_{Hetero}^n refers to the operator of the nth layer of the heterogeneous GNN. v_i refers to the node in the subgraph named i.

It has been verified that GNN layers typically perform optimally with 2-3 layers. Therefore, this work use a two-layer heterogeneous GNN, as illustrated in Fig. 2.

C. Deep Reinforcement Learning for Crowd Navigation

In most of the previous studies[17], [20], the objective of DRL is to determine the optimal policy that maximizes the state-action value function $Q^*(s_t, a_t)$:

$$\pi^*(s_t) = \arg \max_{a_t} Q^*(s_t, a_t), \quad (7)$$

then $Q^*(s_t, a_t)$ is established by the Bellman optimal equation:

$$Q^*(s_t, a_t) = \sum_{s_{t+1}, r_t} P(s_{t+1}, r_t | s_t, a_t) \times [r_t + \gamma \max_{a_{t+1}} Q^*(s_{t+1}, a_{t+1})], \quad (8)$$

where $\gamma \in (0, 1)$ is a discount factor.

This letter introduces a discrete action space with feasible options for the agent. At each time step t, the agent chooses an action (denoted as a_t) consisting of a speed and a heading angle. We offer 5 distinct speeds, exponentially distributed between 0 and v_{pref} , and 16 evenly spaced heading angles across 0 to 2π , resulting in a total of 80 fine-grained discrete actions.

As the scenario continues to expand, the previous setting of mismatched rewards makes the training less convergent[12]. Hence, this letter adopts the shaped reward proposed in SG-DQN[26]. The reward R^t with time step t includes rewards for reaching goals, comfort rewards, penalties for collision, etc. Please refer to SG-DQN for details. This optimized reward function propels agents to peak performance, bypassing the need for imitation learning initialization. We adhere primarily to algorithmic flow of the ST²[20] for anticipating a fairer contrast in subsequent comparisons.

We use a 4-layer MLP with ReLU activations as a value estimator to approach $Q^*(s_t, a_t)$, As shown in Fig. 2.

IV. EXPERIMENTS

A. Simulation Setup

1) *Simulation Environment*: Fig. 4 displays our 2D simulation environment based on CrowdNav¹. Test scenarios are categorized into two groups: one is the single-robot circle crossing scenario, and the other is the multi-robot circle crossing scenario. The single-robot circle crossing scenario is set up with 5 humans, or 5H for short, while the multi-robot circle crossing scenario is set up with 5 humans and 2 other

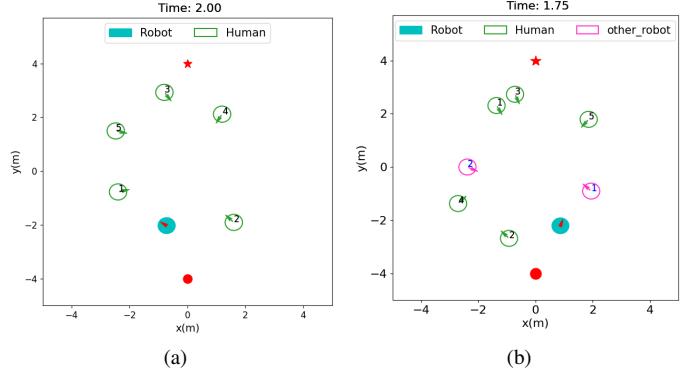


Fig. 4: **Demonstration of our simulation environment.** We mainly utilize two training and testing scenarios: (a) Single-robot circle crossing scenario; (b) Multi-robot circle crossing scenario.

robots, or 5H2O for short. Unlike center robot and other robots, humans have comfort requirements and invisible settings to avoid training overly aggressive strategies. We employ unicycle kinematics for all agents. The human and other robots are controlled by ORCA. We also assume that all agents can achieve the desired velocities immediately, and they will keep moving with these velocities for δt seconds.

2) *Baselines and Ablation Models*: We choose four existing state-of-the-art methods as baseline, including LSTM-RL[22], LM-SARL[17] and ST2[20]. To assess the contributions of the heterogeneous GNN and the newly proposed parameters of the category on the performance, thus we set up the ablation models separately. Where HoR-DRL-nocate denotes the policy using a homogenous GNN without state information of the category. HoR-DRL denotes the policy using a homogenous GNN with state information of the category. HeR-DRL-nocate denotes the policy using the heterogeneous GNN without state information of the category. It is noteworthy that this homogeneous GNN particularly takes the states of all agents as nodes in a comprehensive graph, feeding them into the 2-layer GNN in Equation 4. Since there is only one class of agent in addition to the center robot in the single-robot scenario, only the HoR-DRL ablation method is trained.

3) *Evaluation Metrics*: The performance metrics include *Success Rate*, *Collision Rate*, *Average Time*, *Discomfort Rate* and *Min Distance*. *Success Rate*(SR): the ratio of the center robot reaching its goal without a collision. *Collision Rate*(CR): the ratio of the center robot colliding with other agents. *Average Time*(AT): the average time taken by the center robot to reach its goal (in seconds). *Discomfort Rate*(DR): the ratio of the minimum separating distance between a robot and other agents is less than risk but has not yet collided. *Min Distance*(MD): the average robot minimum separating distance to the nearest obstacle agent in risk cases. Among them, SR and CR can reflect safety, AT reflects efficiency, and DR and MD reflect comfort.

4) *Training Process*: We implemented the policy in PyTorch[33] and trained it using the Adam[34] optimizer

¹<https://github.com/vita-epfl/CrowdNav.git>

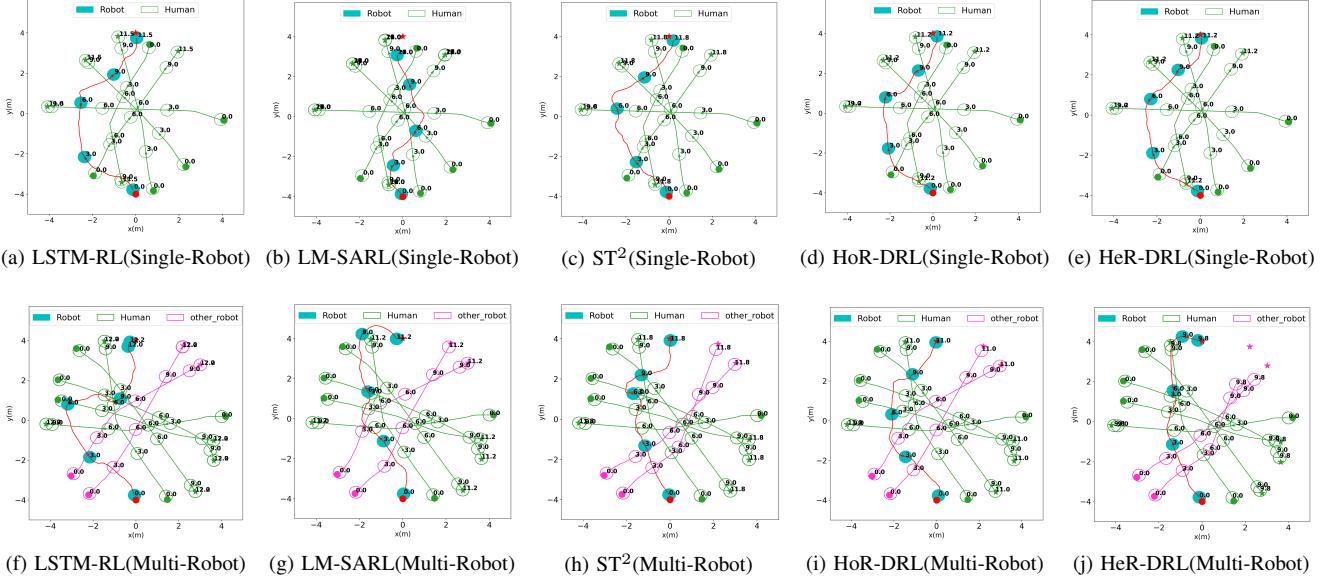


Fig. 5: The comparison of navigation trajectories in single-robot and multi-robot circle crossing.

with a learning rate of $1e^{-3}$. The batch size is set to 100 and discount factor $\gamma = 0.9$. Due to differences in scenario complexity, for optimal performance after convergence, each method can be trained on 10,000 episodes for single-robot circle crossing scenario and 15,000 episodes for multi-robot circle crossing scenario. The same parameter settings were used for training in both scenarios. Subsequent to each episode, the agent's position and goal point are randomly resampled, with the exception of the center robot. The exploration rate of the ϵ -greedy policy decays linearly from 0.5 to 0.1 in the first $4k$ episodes and keeps 0.1 for the other episodes. To ensure fairness, deep reinforcement learning training process applied to all compared methods are uniform.

B. Quantitative Evaluation

1) Performance Comparison with Baseline In Original Configuration: Table I presents the test results of the original configuration in the single-robot scenario and the multi-robot scenario, respectively. HeR-DRL convincingly outperformed all baselines in both scenarios, with respect to SR, DR and MD. meanwhile, the advantage of HeR-DRL is larger in multi-robot scenarios than in single-robot scenarios. The analysis suggests that the superior performance in multi-robot scenarios is likely attributed to the presence of five heterogeneous relations, compared to the two(HHI and HCRI) in single-robot scenarios. This increased diversity of interactions enhances the effectiveness of the method. Analyzing the performance of AT, it is evident that our method exhibits a relatively longer overall navigation time. This can be attributed to the deliberate choice of paths prioritizing comfort. Opting for paths that enhance comfort may involve navigating through long detours or adopting lower speeds, ultimately leading to less efficient navigation. However, the difference in navigation time is not significant relative to the other methods, with a difference of 0.248s from the minimum in the single-robot scenario

and 0.428s from the minimum in the multi-robot scenario. Meanwhile the frequency of discomfort improved significantly, with a minimum reduction of 34.2% for single-robot scenarios and 33.3% for multi-robot scenarios. We were very pleased to see the positive results of sacrificing a small amount of time for a substantial increase in comfort.

TABLE I: QUANTITATIVE RESULTS IN CIRCLE CROSSING ENVIRONMENT

Scenario	Method	SR(%)	CR(%)	AT(s)	DR(%)	MD(m)
Single-Robot	LSTM-RL	95.5	4.5	10.167	0.06	0.16
	LM-SARL	98.5	1.3	10.187	0.038	0.17
	ST ²	98.9	1.1	10.038	0.04	0.17
	HoR-DRL	97.7	2.3	10.328	0.022	0.17
	HeR-DRL(Ours)	99	1	10.415	0.025	0.16
Multi-Robot	LSTM-RL	89.2	10.7	11.776	0.03	0.13
	LM-SARL	94.1	4.7	10.698	0.039	0.15
	ST ²	95.5	4.5	10.949	0.035	0.14
	HoR-DRL-nocate	93.3	6.6	10.841	0.025	0.15
	HoR-DRL	95.7	3.6	10.779	0.021	0.15
	HeR-DRL-nocate(Ours)	95.8	4.1	10.964	0.034	0.15
	HeR-DRL(Ours)	96.3	3.5	11.126	0.02	0.16

Bold values highlight the top performers within each environment configuration.

2) Scenario Adaptation Comparison with Baseline: In order to test the algorithm's scenario adaptation, we tested the trained algorithm under different configurations. To ensure fair and meaningful comparisons, we limited the scenario configurations to a maximum of 10 agents except center robot for mitigating potential performance crashes in overly dense environments. The selected test configurations include 5H3O, 5H4O, 5H5O, 6H2O, 7H2O, and 8H2O. The experimental results are presented in Table II. Table II unveils a clear trend: the overall performance of all algorithms consistently declines as the number of agents in the configuration increases. This inverse relationship is visually evident, suggesting a direct impact of agent density on algorithm effectiveness. Nevertheless, in terms of comfort, the HeR-DRL has a substantial advantage. Obviously, although HeR-DRL performs

sub-optimally in terms of the SR and slightly inferior to ST², the CR, DR and MD metrics are overall superior to ST². Fortunately, the AT metrics of HeR-DRL lags behind relative to it by 0.77 s on average. This further reinforces the point mentioned in the previous section: the current methodology tends to trade a small amount of navigational efficiency for a significant improvement in comfort. Since only HRI was coded and no distinction was made between HCRI and HORI accordingly, It's apparent that LSTM-RL exhibits the worst overall performance. While LM-SARL demonstrates the best navigational efficiency, a substantial performance gap exists in other metrics when compared to ST² and HeR-DRL. This further underscores the importance of encoding complex heterogeneous relations.

TABLE II: QUANTITATIVE ADAPTATION ANALYSIS IN MULTIPLE CONFIGURATIONS OF MULTI-ROBOT CROWD NAVIGATION

Methods	Configure	SR(%)	CR(%)	AT(s)	DR(%)	MD(m)
LSTM-RL	5H3O	83.3	7.8	11.672	0.032	0.14
	5H4O	84.2	7.7	11.782	0.033	0.16
	5H5O	82.5	9.1	12.079	0.032	0.15
	6H2O	80.4	8.1	11.734	0.047	0.14
	7H2O	75.8	14.3	12.187	0.045	0.13
	8H2O	68.5	18.1	12.411	0.06	0.14
LM-SARL	5H3O	88.3	10.1	10.7	0.044	0.14
	5H4O	85.8	13.3	10.794	0.052	0.14
	5H5O	83.9	14.7	10.942	0.062	0.13
	6H2O	87.7	11.2	10.859	0.057	0.14
	7H2O	84.6	14.5	11.011	0.079	0.14
	8H2O	79.6	18.5	11.196	0.112	0.14
ST ²	5H3O	95.7	4	10.887	0.037	0.15
	5H4O	93.4	6.3	10.929	0.046	0.15
	5H5O	92.7	7.2	10.926	0.048	0.15
	6H2O	95.2	4.4	11.087	0.046	0.14
	7H2O	91.4	8.2	11.139	0.063	0.15
	8H2O	89	10.5	11.235	0.082	0.15
HeR-DRL	5H3O	93.4	4.3	11.283	0.021	0.17
	5H4O	91.5	4.8	11.653	0.025	0.17
	5H5O	85.6	6.3	12.492	0.02	0.17
	6H2O	92.3	4.9	11.434	0.035	0.17
	7H2O	87.5	7.9	11.756	0.045	0.16
	8H2O	80.6	11.3	12.216	0.059	0.15

Bold values highlight the top performers within each environment configuration.

C. Qualitative Evaluation

We further investigated the effectiveness and superiority of our method by performing trajectory analyses in single-robot scenarios and multi-robot scenarios respectively, as shown in Fig.5. Firstly, in the test case of the single robot scenario, LM-SARL crosses the pedestrians from the middle and freezes at the end, and LSTM-RL, ST², and HeR-DRL all choose to go around the middle to reach the end point. However, it can be clearly seen that HeR-DRL's trajectory is the one with the least abrupt change in the direction. Then in the multi-robot circle crowssing scenario. LSTM-RL and ST² have similar trajectories, favouring obstacle avoidance and bypassing from after the agent, while LM-SARL and HeR-DRL avoid obstacles and bypass from before the agent. After comparison, LSTM-RL deliberately avoids from the back of the trajectory after the 6th second, resulting in a long route

around the far side. ST² performs better than LSTM-RL in terms of smoothness, but its average distance from the other agents during the movement is significantly worse than that of HeR-DRL. Although the trajectories of LM-SARL and HeR-DRL are similar, the trajectory show that LM-SARL tends to go around at a low speed even though it can accelerate in front of it, while HeR-DRL accelerates when it has the conditions to accelerate, which gives it an advantage in terms of direction change and efficiency. Despite focusing only on spatial aspects, HeR-DRL achieves comfort and trajectory quality equal to or even better than ST2, which models both spatial and temporal interactions.

D. Ablation Analysis

Comparing HoR-DRL-nocate and HoR-DRL reveals that the addition of state information of the category not only enhances the success rate but also improves the comfort metrics. Similar conclusions can be drawn from the performance of the comparison between HER-DRL-nocate and HeR-DRL. This indicates that the state information of the category comprehensively improves the extraction of interactive relations in scenarios.

Comparison of the two ablation models, HoR-DRL-nocate and HeR-DRL-nocate, revealed that HeR-DRL-nocate sacrificed a certain level of efficiency and comfort to improve the success rate. The same conclusion was found when comparing the HoR-DRL and HeR-DRL in single-robot scenario. This seems to imply that our proposed heterogeneous GNN are more inclined to improve success rates, even at the expense of a small amount of comfort. Nevertheless, a comparison of HoR-DRL and HeR-DRL in a multi-robot scenario reveals that the addition of the heterogeneous GNN leads to an improvement in overall performance in addition to AT. The analysis reveals that incorporating state information with the category enhances the effectiveness of the heterogeneous GNN in managing diverse relations. This enables a notable increase in the success rate while maintaining, or possibly slightly improving, overall comfort. Trajectory analysis of HoR-DRL and HeR-DRL in both scenarios demonstrates that HeR-DRL has fewer changes in the direction of motion than HoR-DRL, and the trajectory is smoother, reflecting the excellence of heterogeneous GNN for modelling interactive relations.

V. CONCLUSION

In this letter, HeR-DRL is proposed to solve the navigation problem in multi-robot pedestrian scenarios. The proposed algorithm fully improves the representation of scenario interactive relations by constructing a robot-crowd heterogeneous relation graph with the help of the state information of the category. The experimental results show that the proposed method achieves the best performance in terms of success rate and comfort metrics, in contrast to the state-of-the-art baselines. However, there are some limitations to the methodology of this letter, which only explores spatial heterogeneity and does not address the effects of spatial-temporal heterogeneity. Thus, This will be explored in future research.

REFERENCES

- [1] L. Huber, J.-J. Slotine, and A. Billard, “Avoiding dense and dynamic obstacles in enclosed spaces: Application to moving in crowds,” *IEEE Transactions on Robotics*, vol. 38, no. 5, pp. 3113–3132, 2022.
- [2] K. Kodur and M. Kyararini, “Patient–robot co-navigation of crowded hospital environments,” *Applied Sciences*, vol. 13, no. 7, p. 4576, 2023.
- [3] T. Fan, X. Cheng, J. Pan, P. Long, W. Liu, R. Yang, and D. Manocha, “Getting robots unfrozen and unlost in dense pedestrian crowds,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1178–1185, 2019.
- [4] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, “Human-aware robot navigation: A survey,” *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.
- [5] J. Van den Berg, M. Lin, and D. Manocha, “Reciprocal velocity obstacles for real-time multi-agent navigation,” in *2008 IEEE international conference on robotics and automation*. Ieee, 2008, pp. 1928–1935.
- [6] D. Helbing and P. Molnar, “Social force model for pedestrian dynamics,” *Physical review E*, vol. 51, no. 5, p. 4282, 1995.
- [7] P. Trautman and A. Krause, “Unfreezing the robot: Navigation in dense, interacting crowds,” in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 797–803.
- [8] F. Feurtey, “Simulating the collision avoidance behavior of pedestrians,” *The University of Tokyo, School of Engineering, Department of Electronic Engineering*, 2000.
- [9] G. S. Aoude, B. D. Luders, J. M. Joseph, N. Roy, and J. P. How, “Probabilistically safe motion planning to avoid dynamic obstacles with uncertain motion patterns,” *Autonomous Robots*, vol. 35, pp. 51–76, 2013.
- [10] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard, “Socially compliant mobile robot navigation via inverse reinforcement learning,” *The International Journal of Robotics Research*, vol. 35, no. 11, pp. 1289–1307, 2016.
- [11] P. Trautman, J. Ma, R. M. Murray, and A. Krause, “Robot navigation in dense human crowds: the case for cooperation,” in *2013 IEEE international conference on robotics and automation*. IEEE, 2013, pp. 2153–2160.
- [12] C. Chen, S. Hu, P. Nikdel, G. Mori, and M. Savva, “Relational graph learning for crowd navigation,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10007–10013.
- [13] C. Cao, P. Trautman, and S. Iba, “Dynamic channel: A planning framework for crowd navigation,” in *2019 international conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 5551–5557.
- [14] K. Driggs-Campbell, V. Govindarajan, and R. Bajcsy, “Integrating intuitive driver models in autonomous planning for interactive maneuvers,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 12, pp. 3461–3472, 2017.
- [15] K. Driggs-Campbell, R. Dong, and R. Bajcsy, “Robust, informative human-in-the-loop predictions via empirical reachable sets,” *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 3, pp. 300–309, 2018.
- [16] Y. F. Chen, M. Liu, M. Everett, and J. P. How, “Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning,” in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 285–292.
- [17] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, “Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning,” in *2019 international conference on robotics and automation(ICRA)*. IEEE, 2019, pp. 6015–6022.
- [18] W. Shi, Y. Zhou, X. Zeng, S. Li, and M. Bennewitz, “Enhanced spatial attention graph for motion planning in crowded, partially observable environments,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4750–4756.
- [19] X. Zhang, W. Xi, X. Guo, Y. Fang, B. Wang, W. Liu, and J. Hao, “Relational navigation learning in continuous action space among crowds,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 3175–3181.
- [20] Y. Yang, J. Jiang, J. Zhang, J. Huang, and M. Gao, “St²: Spatial-temporal state transformer for crowd-aware autonomous navigation,” *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 912–919, 2023.
- [21] 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.
- [22] M. Everett, Y. F. Chen, and J. P. How, “Motion planning among dynamic decision-making agents with deep reinforcement learning,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3052–3059.
- [23] W. Wang, L. Mao, R. Wang, and B.-C. Min, “Multi-robot cooperative socially-aware navigation using multi-agent reinforcement learning,” *arXiv preprint arXiv:2309.15234*.
- [24] R. Chandra, V. Zinage, E. Bakolas, J. Biswas, and P. Stone, “Decentralized multi-robot social navigation in constrained environments via game-theoretic control barrier functions,” *arXiv preprint arXiv:2308.10966*, 2023.
- [25] 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.
- [26] Z. Zhou, P. Zhu, Z. Zeng, J. Xiao, H. Lu, and Z. Zhou, “Robot navigation in a crowd by integrating deep reinforcement learning and online planning,” *Applied Intelligence*, vol. 52, no. 13, pp. 15600–15616, 2022.
- [27] S. Liu, P. Chang, W. Liang, N. Chakraborty, and K. Driggs-Campbell, “Decentralized structural-rnn for robot crowd navigation with deep reinforcement learning,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 3517–3524.
- [28] S. Liu, P. Chang, Z. Huang, N. Chakraborty, K. Hong, W. Liang, D. L. McPherson, J. Geng, and K. Driggs-Campbell, “Intention aware robot crowd navigation with attention-based interaction graph,” in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12015–12021.
- [29] K. Zhu, B. Li, W. Zhe, and T. Zhang, “Collision avoidance among dense heterogeneous agents using deep reinforcement learning,” *IEEE Robotics and Automation Letters*, vol. 8, no. 1, pp. 57–64, 2022.
- [30] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [31] W. L. Hamilton, R. Ying, and J. Leskovec, “Representation learning on graphs: Methods and applications,” *arXiv preprint arXiv:1709.05584*, 2017.
- [32] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, “Modeling relational data with graph convolutional networks,” in *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*. Springer, 2018, pp. 593–607.
- [33] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in pytorch,” 2017.
- [34] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [35] C. Shi, Y. Li, J. Zhang, Y. Sun, and S. Y. Philip, “A survey of heterogeneous information network analysis,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 17–37, 2016.