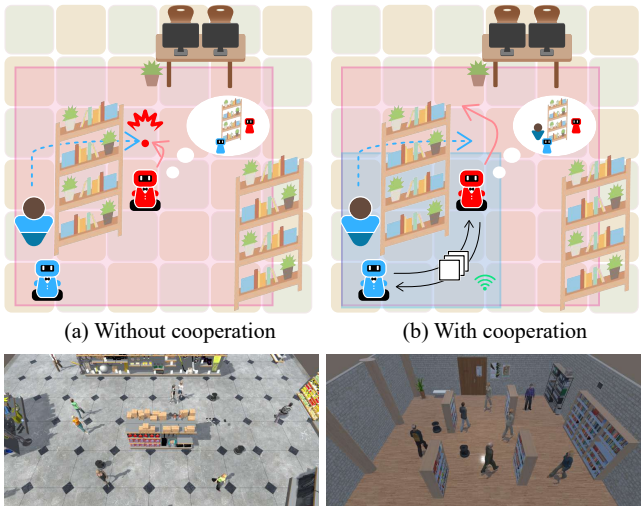
CoopUQ: Uncertainty-Aware Multi-Robot Social Navigation with Cooperative Occupancy Prediction

***Abstract*— While reinforcement learning (RL) has enabled social robots to navigate in complex and dynamic real-world environments without hand-crafted rules, in shared human environments with multiple robots, directly extending single-robot approaches often fails to ensure safety and efficiency in constrained spaces. Leveraging cooperation effectively for both prediction and navigation remains a major challenge, especially under occlusions and partial observations. We propose CoopUQ, an end-to-end multi-robot social navigation framework that integrates uncertainty-aware cooperative occupancy grid prediction with constrained multi-agent reinforcement learning (MARL) to handle occlusions and partial observations in environments populated by static obstacles and moving pedestrians. Each robot shares either raw sensor data or locally predicted occupancy grid maps, which significantly improves situational awareness and boosts the prediction accuracy by 29.8%. To further enhance robustness and generalization, CoopUQ estimates the uncertainty of each robot's prediction and performs adaptive fusion across agents, which enables agents to weigh contributions based on the prediction confidence. The resulting uncertainty map is used to augment RL observations and guide the constrained MARL policy to avoid navigating into predicted occupied or highly uncertain regions. This approach achieves state-of-the-art performance across various in-distribution and out-of-distribution scenarios in our high-fidelity Unity3D-based simulation environments. We further validate CoopUQ in real-world environments to demonstrate its effectiveness and robustness. Code and video demonstrations are provided in the supplementary materials.**

I. INTRODUCTION

Social navigation typically involves mobile robots operating in constrained environments with static (e.g., walls, bookshelves) and dynamic obstacles (e.g., pedestrians) [1], [2]. Prior work has largely focused on enabling a single robot to navigate safely and efficiently among humans by leveraging advances in human motion prediction and socially compliant planning [3]. However, in many real-world settings, such as hospitals, airports, or office buildings, multiple robots may operate simultaneously in the shared human environment. In such cases, effective coordination among robots is essential to ensure safety, efficiency, and human comfort.

Single-robot systems are often limited by occlusions and constrained local observations, making it difficult to consistently identify safe and efficient paths. For instance, a robot may get stuck in a confined area due to the static obstacles or fail to detect pedestrians hidden behind a wall when turning a corner (see Fig. 1a), which may lead to unsafe or inefficient navigation. In contrast, cooperative multi-robot systems can exchange information to enhance situational awareness. By sharing observations and predictions (e.g., a nearby robot detecting a pedestrian in an occluded area), robots can augment their perception and make more informed decisions (see Fig. 1b). While cooperative perception and prediction have been widely studied [4]–[6], cooperative decision making for multi-robot social navigation remains largely underexplored.



(a) Without cooperation (b) With cooperation

(c) Our Unity3D-based simulation platform

Fig. 1: (a) The single-agent approach relies only on the robot's own observations and misses an incoming pedestrian behind the obstacle. (b) The proposed approach fuses shared information, anticipates dynamic obstacle motions under uncertainty, and plans a safe trajectory. (c) Our high-fidelity 3D simulations with diverse layouts and human behaviors.

A recent study on cooperative planning [7] focuses on behavior modeling via trajectory prediction but cannot handle static obstacles. In contrast, some methods use occupancy grid maps (OGMs) to assign occupancy probabilities to grid cells, enabling real-time prediction and planning in constrained environments [8]. However, dynamic environments introduce uncertainty due to evolving pedestrian intentions, diverse motion patterns or speeds, and occluded high-risk regions. These factors can cause distribution shifts in observations, leading to degraded predictions or unsafe planning. Accurately predicting future dynamics and estimating uncertainty in real time remains challenging for complex navigation tasks. While a prior work [9] addresses uncertainty-aware single-robot navigation, it cannot be extended naturally to multi-robot settings. In particular, single-agent prediction may misguide the uncertainty module due to occlusions or noisy ego motion, whereas cooperative prediction with adaptive fusion improves reliability by leveraging diverse viewpoints and shared uncertainty. Moreover, prior work often overlooks how OGM predictions and uncertainty estimates can explicitly guide the learning process of navigation.

To address these gaps, we propose an end-to-end coopera-

tive social navigation framework that integrates uncertainty-aware cooperative OGM prediction into downstream decision making via constrained multi-agent reinforcement learning (CMARL) [10]. Specifically, we develop a communication-efficient, latency-robust cooperative OGM predictor that leverages raw sensor data and shared predictions from multiple robots to enhance situational awareness. Our method achieves state-of-the-art performance on our challenging dataset collected in Unity3D-based simulation environments. To further mitigate the negative effects of imperfect predictions in robot navigation, we introduce an adaptive fusion mechanism that allows each robot to incorporate lower-uncertainty predictions from neighbors for more robust forecasting. Finally, we integrate uncertainty-aware fusion into constrained MARL by formulating a cost function that penalizes navigation through regions predicted to be occupied or highly uncertain, enabling safer and more generalizable policies, especially in out-of-distribution (OOD) scenarios.

To the best of our knowledge, this is the first work in multi-robot social navigation in constrained environments that leverages adaptive cooperative OGM prediction and handles environmental uncertainties to enable safe and efficient navigation. The main contributions are as follows:

- We propose a novel end-to-end framework, CoopUQ, for multi-robot social navigation in dynamic constrained environments, which integrates uncertainty-aware cooperative occupancy prediction into decision making.
- We design cooperative OGM predictors with various fusion strategies to enhance each robot's ability to anticipate future environmental states through collaboration. By adaptively incorporating uncertainty estimates from neighboring robots and integrating them into a constrained policy learning, we further improve the navigation safety in three different OOD scenarios.
- We develop a high-fidelity Unity3D-based simulator for realistic scenarios with diverse obstacle layouts and human behaviors. Our method achieves state-of-the-art performance in both cooperative occupancy prediction and multi-agent navigation, and we further validate its safety and robustness in real-world robot experiments.

II. Related Work

Social Robot Navigation. Current mobile robots must navigate shared spaces with humans while adhering to social norms [11]–[13]. Social navigation is often modeled as a Partially Observable Markov Decision Process (POMDP), where agents act based on incomplete local observations [14]. A key challenge is effectively modeling these partial and uncertain observations in real-world environments. Many prior works focus on predicting trajectories of dynamic obstacles, modeling pedestrians as simple position-radius tuples. These rule-based or learning-based methods perform well in simulation [7], [15]–[17], but often assume full visibility in open spaces and overlook static obstacles, limiting their applicability in occluded or cluttered environments. In contrast, we adopt predicted OGMs with quantified uncertainties as a holistic spatiotemporal representation and aggregate complementary observations through multi-robot cooperation, enabling safer decisions in constrained environments.

Occupancy Grid Map (OGM) Prediction. Predicting the future environment remains challenging due to its inherent stochasticity and uncertainty. OGM is an effective spatial representation that unifies both static and dynamic obstacles. Since OGMs can be treated as images, multi-step prediction is often formulated as a video prediction problem [8], [18]. Recent studies use OGMs to model dynamic human crowds [9], [19]. However, most prior work focuses on single-robot OGM prediction. In contrast, we investigate uncertainty-aware cooperative prediction, where robots adaptively fuse predictions based on their uncertainty estimate to enhance robustness and reliability in OOD scenarios.

Cooperative Perception and Prediction. Cooperative perception and prediction are crucial for overcoming partial observability and perceptual limitations, particularly in environments with occlusions. By sharing sensor data, intermediate features, and predictions, multiple agents can construct a more comprehensive scene understanding, enhancing detection and motion prediction in constrained settings [5], [6]. In this work, we investigate end-to-end cooperative prediction and develop an uncertainty-aware planning framework that leverages the predictions and corresponding uncertainties.

III. Problem Formulation

We consider the problem of navigating multiple mobile robots in an environment with static obstacles (e.g., walls, bookshelves) and dynamic obstacles (e.g., pedestrians, other robots) with specific goals. Our objective is to learn an end-to-end cooperative policy that enables the robots to reach their goal positions safely and efficiently. Formally, we formulate multi-robot social navigation as a constrained Markov game [20], defined by the tuple $< \mathcal{I}, \mathcal{S}, \boldsymbol{\mathcal{O}}, \boldsymbol{\Omega}, \boldsymbol{\mathcal{A}}, \mathcal{T}, R, C, \rho_0, \gamma >$. The agent set $\mathcal{I} = \{1, 2, \ldots, N\}$ represents the social robots. N is the number of social robots. Social robots navigate in a shared space with different obstacles alongside M pedestrians. The state space $\mathcal{S}$ includes each state s_t at time t, capturing the position and speed of robots and dynamic obstacles, and static obstacles in the environment. The joint observation space is $\boldsymbol{\mathcal{O}} = \times_{i \in \mathcal{I}} \mathcal{O}^i$, where $\mathcal{O}^i$ is the local observation of agent i. The observation functions $\boldsymbol{\Omega} = \{\Omega^i\}_{i \in \mathcal{I}}$ map the global state to each agent's local observation $\Omega^i : \mathcal{S} \to \mathcal{O}^i$. The joint action space is $\boldsymbol{\mathcal{A}} = \times_{i \in \mathcal{I}} \mathcal{A}^i$, where each action $a_t^i \in \mathbb{R}^2$ represents the desired 2D velocity of agent i at time t. The transition function $\mathcal{T} : \mathcal{S} \times \boldsymbol{\mathcal{A}} \to \mathcal{S}$ governs the state evolution, with each robot tracking its desired speed via a low-level controller. The reward function $R : \mathcal{S} \times \boldsymbol{\mathcal{A}} \to \mathbb{R}$ and cost function $C : \mathcal{S} \times \boldsymbol{\mathcal{A}} \to \mathbb{R}$ aggregate individual agent rewards and costs, respectively. Finally, ρ_0 is the initial state distribution and $\gamma \in [0, 1)$ is the discount factor. We aim to optimize a shared policy π_θ that maximizes the expected cumulative reward while ensuring the cumulative cost remains within specified constraints. This can be represented as:

$$\begin{aligned} \max_{\pi_\theta} \quad & J^R(\pi_\theta) = \mathbb{E}_{s_0, \mathbf{a}_0, \mathbf{o}_0, \ldots} \left[\sum_{t=0}^{T} \gamma^t R(s_t, \mathbf{a}_t) \right] \\ \text{s.t.} \quad & J^C(\pi_\theta) = \mathbb{E}_{s_0, \mathbf{a}_0, \mathbf{o}_0, \ldots} \left[\sum_{t=0}^{T} \gamma^t C(s_t, \mathbf{a}_t) \right] \leq \epsilon, \end{aligned} \tag{1}$$

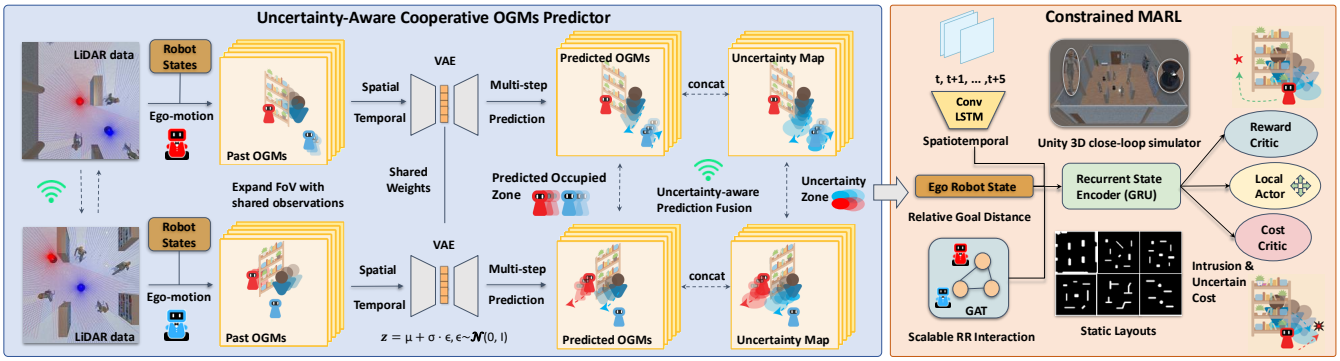


Fig. 2: The overall architecture of the proposed cooperative end-to-end system for multi-robot social navigation.

where $\mathbf{a}_t = (a_t^1, ..., a_t^N)$, $a_t^i \sim \pi_\theta(a_t^i|\mathbf{o}_t^i)$, and ϵ is a user-specified cost limit.

All robots are assumed to be equipped with a 2D LiDAR sensor. Let $\mathbf{y}_t^i$ denote the LiDAR measurements at time t for robot $i \in \mathcal{I}$. Binary OGMs can be obtained from sensor measurements. We assume that the robot can obtain accurate short-term estimates of its own relative pose and velocity through onboard odometry or other localization methods. The pose of robot i at time t is denoted as $\mathbf{p}_t^i = [p_t^{x_i}, p_t^{y_i}]$, its velocity is $\mathbf{v}_t^i = [v_t^{x_i}, v_t^{y_i}]$, and its goal position is represented by $\mathbf{g}_t^i = [g_t^{x_i}, g_t^{y_i}]$. To align with the real-world sensors and multi-agent communication ability, the observation space is comprised of three components $\mathbf{o}_t = [\mathbf{o}_t^{\text{ego}}, \mathbf{o}_t^{\text{map}}, \mathbf{o}_t^d]$. $\mathbf{o}_t^{\text{ego}} = [\mathbf{p}_t, \mathbf{v}_t, \mathbf{g}_t]$ is the robot's ego state information in the current frame. $\mathbf{o}_t^{\text{map}}$ stores the sequence of OGMs obtained from the h-step predicted by the cooperative prediction model, aligned and appended to the robot's local frame. It is stored as a sequence of 2D images with the shape $(h+1, m_x, m_y)$, where m_x and m_y are the height and width of the map, respectively, indicating the occupancy probability at each grid cell at each time step. $\mathbf{o}_t^d$ represents the current pose of other robots within the communication range, which provides spatial information crucial for collision avoidance.

IV. Method

A. Method Overview

Fig. 2 provides an overall diagram of the proposed CoopUQ framework. We first propose an uncertainty-aware cooperative OGMs predictor to generate and fuse multi-step OGM forecasts. By incorporating pixel-wise entropy as an uncertainty estimate, our method performs uncertainty-aware fusion, allowing robots to emphasize confident predictions. The resulting forecasts and uncertainty maps provide a holistic and risk-aware view of the future scene, which is used to augment the observations for the policy network. Further, we adopt a safety-related cost to guide safe and efficient multi-robot social navigation via constrained MARL.

B. Uncertainty-Aware Cooperative OGM Prediction

Due to the ray-tracing nature of LiDAR, the local OGMs obtained by each robot may contain occlusions, which can result in failures when the robot tries to predict dynamic agents behind obstacles. For instance, a pedestrian suddenly appearing from behind a wall can lead to a potential collision. Additionally, LiDAR only captures robot-facing surface, making it challenging to infer accurate object size and shape, which in turn limits the situational awareness and degrades planning efficiency. To overcome these challenges, we propose a cooperative predictor that generates multi-step OGMs forecasts by either fusing raw sensor inputs from all agents (i.e., early fusion) or aggregating individual predictions from each agent (i.e., late fusion). This design balances prediction accuracy with communication and computation overhead. Our cooperative OGM predictor leverages shared information to generate holistic and uncertainty-aware forecasts of the future environment. Specifically, the prediction model learns the distribution over h-step future OGMs, denoted by $\mathbf{o}_t^{\text{map}}$, conditioned on the past τ frames observed up to time t. The probabilistic prediction is formulated as:

$$\mathbf{o}_t^{\text{map}} \sim p_\delta(\mathbf{o}_t^{\text{map}} \mid \mathbf{y}_{t-\tau:t}^{1:N}, \mathbf{p}_{t-\tau:t}^{1:N}), \tag{2}$$

where $\mathbf{y}_{t-\tau:t}^{1:N}$ and $\mathbf{p}_{t-\tau:t}^{1:N}$ represent the historical sequences of LiDAR point clouds and agent poses for all N cooperative agents, respectively. The model is parameterized by δ. We employ the SOGMP++ backbone [9] for multi-agent cooperative prediction with various fusion strategies. SOGMP++ is a lightweight single-agent model that compensates for ego-motion, segments static and dynamic elements, encodes spatiotemporal features via a ConvLSTM, and uses a variational autoencoder (VAE) to predict future stochastic OGMs.

Early and Late Fusion. In the *early fusion* setting, each robot i shares its point clouds with neighbors within a communication radius at each timestep t. Each ego robot transforms the received data into its local frame, masks out points hitting its own body, and converts them into OGMs at a specified resolution and range. As ego motion is the main source of discrepancy between consecutive observations, aligning past OGMs to the current frame significantly improves prediction accuracy, enabling holistic predictions from fused observations. This process is formulated as:

$$\mathbf{o}_t^{\text{map}_i} \sim p_\delta(\mathbf{o}_t^{\text{map}_i} \mid \mathbf{y}_{t-\tau:t}^i, \{\hat{\mathbf{y}}_{t-\tau:t}^k\}_{k\neq i}, \mathbf{p}_{t-\tau:t}^i), \tag{3}$$

$$\hat{\mathbf{y}}_{t-\tau:t}^k = \Gamma(\mathbf{y}_{t-\tau:t}^k, \mathbf{p}_{t-\tau:t}^i, \mathbf{p}_{t-\tau:t}^k) \text{ for all } k \neq i, k \in \mathcal{I}, \tag{4}$$

where $\Gamma(\cdot)$ is a spatial transformation function that aligns point clouds from other agents k to the ego robot i's frame.

In the *late fusion* setting, each robot i independently predicts future OGMs at each time step t based on its own observations, then shares its prediction with neighboring robots within the communication range. The ego robot transforms the received OGMs into its local coordinate frame and fuses them with its own prediction by taking the element-

wise maximum probability across all maps. The process can be formulated as:

$$\bar{\mathbf{o}}_t^{\text{map}_i} = \beta(\tilde{\mathbf{o}}_t^{\text{map}_i}, \{\tilde{\mathbf{o}}_t^{\text{map}_k}\}_{k \neq i}), \tag{5}$$

$$\tilde{\mathbf{o}}_t^{\text{map}_i} \sim p_\delta(\tilde{\mathbf{o}}_t^{\text{map}_i} \mid \mathbf{y}_{t-\tau:t}^i, \mathbf{p}_{t-\tau:t}^i), \tag{6}$$

$$\tilde{\mathbf{o}}_t^{\text{map}_k} \sim p_\delta(\tilde{\mathbf{o}}_t^{\text{map}_k} \mid \mathbf{y}_{t-\tau:t}^k, \mathbf{p}_{t-\tau:t}^k), \text{for all} \quad k \neq i, k \in \mathcal{I}, \tag{7}$$

where $\tilde{\mathbf{o}}_t^{\text{map}_i}$ and $\tilde{\mathbf{o}}_t^{\text{map}_k}$ are the predicted OGMs from the ego robot i and surrounding robots k, respectively, and $\beta(\cdot)$ is a spatial fusion function.

Uncertainty Estimation and Fusion. We employ a VAE to model the distribution over future OGMs and generate probabilistic estimates of multi-step occupancy states. This framework captures the inherent uncertainty in forecasting environmental changes in complex, partially observable scenarios. To quantify this uncertainty over time, we compute Shannon entropy for each predicted OGM. For predicted map $\bar{\mathbf{o}}_t^{\text{map}_i}$ at time t from robot i, the entropy is defined as:

$$H(\bar{\mathbf{o}}_t^{\text{map}_i}) = -\sum_{c \in \bar{\mathbf{o}}_t^{\text{map}_i}} p(c) \log p(c) + (1 - p(c)) \log(1 - p(c)), \tag{8}$$

where c denotes each grid cell and $p(c)$ is the predicted occupancy probability. Each cell is assumed to be sampled independently. To estimate pixel-wise uncertainty, we sample M future OGMs from the VAE and compute the entropy at each grid cell based on the distribution of occupancy probabilities. The uncertainty map $\bar{\mathbf{e}}_t^{\text{map}_i}$ is computed as:

$$\bar{\mathbf{e}}_t^{\text{map}_i} = -\frac{1}{M} \sum_{j=1}^{M} p(c_j) \log p(c_j) + (1 - p(c_j)) \log(1 - p(c_j)), \tag{9}$$

where c_j represents the occupancy probability of a given cell in the j-th sample. The shape of $\bar{\mathbf{e}}_t^{\text{map}_i}$ matches that of the predicted OGMs $\bar{\mathbf{o}}_t^{\text{map}_i}$. We normalize the entropy values. The number of samples S allows a trade-off between computational cost and uncertainty estimation fidelity. To enhance both robustness and consistency in prediction fusion, we apply uncertainty-weighted fusion across agents. Let $\{\bar{\mathbf{o}}_t^{\text{map}_i}, \bar{\mathbf{e}}_t^{\text{map}_i}\}_{i=1}^N$ and $\{\mathbf{o}_t^{\text{map}_i}, \mathbf{e}_t^{\text{map}_i}\}_{i=1}^N$ denote the original and final predicted OGMs and their corresponding uncertainty maps from N cooperative agents, respectively. For the ego robot i, OGMs are computed via uncertainty-aware fusion, where each agent's prediction is weighted by the inverse of its uncertainty:

$$\mathbf{o}_t^{\text{map}_i} = \frac{\sum_{i=1}^N w_t^{\text{map}_i} \cdot \bar{\mathbf{o}}_t^{\text{map}_i}}{\sum_{i=1}^N w_t^{\text{map}_i}}, \quad w_t^{\text{map}_i} = \frac{1}{\bar{\mathbf{e}}_t^{\text{map}_i} + \epsilon}, \tag{10}$$

$$\mathbf{e}_t^{\text{map}_i} = \frac{1}{\sum_{i=1}^N w_t^{\text{map}_i}}, \tag{11}$$

where ϵ is a small constant to ensure numerical stability. This strategy gives higher weight to predictions with lower uncertainty, encouraging the fusion process to favor confident agents while still leveraging information from all sources. Such uncertainty-aware fusion enhances spatial coverage and robustness in scenarios where an individual robot makes an error prediction or a prediction with high uncertainty.

C. Policy Network

We concatenate the predicted OGMs and uncertainty maps before feeding them into a ConvLSTM module, which allows robots to account for both spatial forecasts and prediction uncertainty in their decision making. The ego robot's state is encoded using an MLP, while the states of surrounding robots are processed using a graph attention network (GAT), allowing for scalable and structured feature aggregation. The ConvLSTM output is flattened and concatenated with the encoded robot features, then passed through an additional MLP for feature fusion. A GRU unit models temporal dependencies, producing representations used for recurrent policy learning.

D. Constrained MARL with Prediction and Uncertainty Map

We go beyond traditional reward shaping by introducing explicit cost constraints and adopting constrained MARL, which enables safer, uncertainty-aware navigation in dynamic, constrained environments. Specifically, we design a safety radius around the robot, where a cost is incurred if any cell in the predicted OGM within this region exceeds an occupancy threshold. This cost is weighted by the cell's distance to the robot. Similarly, a risk cost is applied within a risk radius on the uncertainty map, incurred when any cell exceeds an uncertainty threshold. The total cost is computed as a weighted sum of the maximum safety and risk costs over the prediction horizon. We define the joint cost functions C as follows:

$$\begin{aligned} C_i(s_t, \mathbf{a}_t) = \Delta t \cdot & \max_{(x,y) \in \mathbf{o}_t^{\text{map}_i}} \lambda_{\text{occ}} \cdot \mathbb{I}\left(\mathbf{o}_t^{\text{map}_i}(x,y) > \phi_{\text{occ}}\right) \cdot d(x,y) \\ & + \max_{(x,y) \in \mathbf{e}_t^{\text{map}_i}} \lambda_{\text{unc}} \cdot \mathbb{I}\left(\mathbf{e}_t^{\text{map}_i}(x,y) > \phi_{\text{unc}}\right) \cdot d(x,y), \end{aligned} \tag{12}$$

where $\mathbb{I}(\cdot)$ returns 1 when the occupancy probability $\mathbf{o}_t^{\text{map}_i}(x,y)$ or the uncertainty $\mathbf{e}_t^{\text{map}_i}(x,y)$ exceeds their respective thresholds ϕ_{occ} and ϕ_{unc}. The distance $d(x,y)$ is measured from the grid center (i.e., ego robot position), and the discomfort factors λ_{occ} and λ_{unc} weight the respective costs. The final cost is scaled by the time step duration Δt. The joint reward function R is defined as the sum of individual rewards $R_i(s_t, \mathbf{a}_t)$ for each robot:

$$R_i(s_t, \mathbf{a}_t) = \begin{cases} 10 & \text{if } d_{\text{target}} < D_{\text{goal}}, \\ -10 & \text{if collision with obstacles}, \\ -15 & \text{if collision with other robots}, \\ -20 & \text{if collision with human}, \\ r_{\text{dis}} & \text{otherwise}, \end{cases} \tag{13}$$

where d_{target} is the Euclidean distance to the goal, and $D_{\text{goal}} = 0.5\,\text{m}$ is the goal radius. The dense progress reward r_{dis} encourages efficient navigation and is defined as:

$$r_{\text{dis}} = \Delta d \cdot f + r_{\text{time}}, \quad \Delta d = d_{\text{target}}^{t+1} - d_{\text{target}}^{t}, \tag{14}$$

with scaling factor $f = 5$ and time penalty $r_{\text{time}} = -0.1$.

The multi-agent variant of Proximal Policy Optimization (MAPPO) has proven effective for MARL [21]. To address constrained MARL, we extend PPO-Lagrangian to MAPPO-Lagrangian, which enables solving an unconstrained dual problem using the Lagrangian relaxation method. Specifically, instead of optimizing the constrained objective in Eq. 1, we optimize the following Lagrangian function:

$$\max_{\pi_\theta} \min_{\lambda} L(\pi_\theta, \lambda) = J^R(\pi_\theta) - \lambda \left(J^C(\pi_\theta) - \epsilon\right) \tag{15}$$

where $\lambda \in \mathbb{R}^+$ is the Lagrangian multiplier. We set up two

TABLE I: Comparison of OGM prediction under different cooperation settings.

Methods	Cooperation Strategy	Prediction Horizon (steps)	SSIM ↑		WMSE ↓		IoU ↑		Bandwidth (MB/s) ↓	FPS ↑
			no delay	200ms delay	no delay	200ms delay	no delay	200ms delay		
SOGMP++ [9]	No Cooperation	1	0.73 ± 0.00	-	0.17 ± 0.00	-	0.55 ± 0.01	-	-	79.6 ± 0.2
		5	0.68 ± 0.00	-	0.20 ± 0.00	-	0.52 ± 0.01	-		
Ours	Late Fusion	1	0.86 ± 0.01	0.85 ± 0.01	0.09 ± 0.00	0.09 ± 0.01	0.69 ± 0.01	0.67 ± 0.02	0.39	70.7 ± 0.3
		5	0.80 ± 0.01	0.79 ± 0.01	0.13 ± 0.01	0.13 ± 0.01	0.64 ± 0.01	0.62 ± 0.02		
	Early Fusion	1	**0.90 ± 0.00**	**0.89 ± 0.01**	**0.06 ± 0.00**	**0.06 ± 0.01**	**0.75 ± 0.01**	**0.73 ± 0.01**	0.02	73.2 ± 0.3
		5	**0.86 ± 0.00**	**0.85 ± 0.01**	**0.09 ± 0.00**	**0.09 ± 0.01**	**0.69 ± 0.01**	**0.68 ± 0.02**		

critics to estimate the values for reward and cost, respectively.

V. Experiments

A. Simulation Environment and Dataset

Most existing environments for social navigation are in 2D or overly simplified 3D simulators and only support a single robot [15]. To address these limitations, we develop a novel Unity3D-based simulation platform that supports multiple cooperative robots, human agents, and diverse static obstacles (see Fig. 1c). Static environments are generated by randomly placing convex polygons to represent non-navigable regions. Human agents navigate using Unity's built-in NavMesh system, which distinguishes between walkable and non-walkable areas based on the detected obstacles, enabling realistic simulation of crowd movement in complex layouts. Humans with randomized maximum speeds ranging from $0.8\,\mathrm{m/s}$ to $1.5\,\mathrm{m/s}$ and radii between $0.3\,\mathrm{m}$ and $0.4\,\mathrm{m}$. Additionally, each pedestrian randomly changes their goal every 3 to 5 seconds, further increasing the unpredictability and dynamics of the environment. Robots are modeled as invisible to humans, assuming full responsibility for collision avoidance. Each robot is equipped with a simulated 2D LiDAR of $7\,\mathrm{m}$ sensing range. The robot radius is set to $0.2\,\mathrm{m}$ and the maximum speed is set to $1.0\,\mathrm{m/s}$. During policy training, three robots spread across the arena populated with 5–10 pedestrians with randomized positions and goals to emulate crowds. The arena is randomly selected from three distinct layouts with varying static obstacles, spanning between $10\,\mathrm{m} \times 10\,\mathrm{m}$ and $15\,\mathrm{m} \times 15\,\mathrm{m}$. The time limit of each episode is $40\,\mathrm{s}$.

For the in-distribution test setting, robots navigate within the same three arenas as training, with 7 to 10 pedestrians. For the OOD evaluation, we test in three scenarios: a) unseen environments with different obstacle shapes and orientations, b) increased pedestrian density (12–15), and c) higher crowd dynamics with maximum speeds up to $2\,\mathrm{m/s}$. Moreover, to train and evaluate our cooperative OGM predictors, we construct an offline dataset. LiDAR and pose data are collected at $10\,\mathrm{Hz}$, with occupancy grids generated at a resolution of $0.1\,\mathrm{m}$ per cell, resulting in 64×64 OGMs. We set the historical window size τ to 5. To enhance data diversity, each robot collects sequences of LiDAR scans and poses during both policy sampling and training, conditioned on current and historical OGMs. The dataset comprises scenarios with three robots and 10–15 pedestrians across three distinct environments, totaling 361,152 training samples and 30,498 testing samples.

B. Evaluation Metrics and Baselines

Cooperative Prediction. We adopt standard metrics from the video prediction literature [18] to evaluate the quality of future OGM predictions. Specifically, we compute the weighted mean squared error (WMSE) and the structural similarity index measure (SSIM) over the next 5 predicted time steps. We also report the intersection over union (IoU) to assess spatial overlap accuracy. To evaluate the communication and computational efficiency of different cooperation strategies, we measure the bandwidth usage and the FPS. To demonstrate the effectiveness of our cooperative prediction approach, we compare against the SOGMP++ baseline [9].

Social Navigation. The navigation metrics [15] include: success rate (SR), collision rate (CR), timeout rate (TR), average navigation time (NT), and path length (PL). Additionally, to evaluate trajectory smoothness, we use average acceleration (AA), defined as the mean step-to-step change in velocity. We compare our method (CMARL with uncertainty fusion in the early fusion setting) against the NavMesh-based non-cooperative policy. We also assess the impact of cooperative prediction by evaluating policies trained with and without cooperative prediction. For the ablation studies, we analyze performance across the following settings:

- Perception Only: without a prediction model, use only the past five steps OGMs as observations.
- Prediction: with current and predicted OGMs as observations, but without the uncertainty map.
- CMARL with Uncertainty: MAPPO-Lag. policy constrained on predicted OGMs and uncertainty maps.
- CMARL with Uncertainty Fusion: MAPPO-Lag. policy constrained on uncertainty-aware prediction and uncertainty fusion.

C. Quantitative Results

Cooperative Prediction. Table I compares the OGM prediction performance under different cooperation strategies. We evaluate models trained with the same training seeds across three random test seeds, and report the mean and standard deviation of each metric. Without cooperation, SOGMP++ [9] yields the lowest performance across both short and long prediction horizons, highlighting the limitations of a single agent in reasoning about dynamic obstacles or occluded regions without contextual information from others.

In contrast, both early and late fusion strategies significantly improve prediction quality and remain robust under a $200\,\mathrm{ms}$ communication delay, exhibiting only a 1.5% average SSIM drop at the 5-step horizon. Among them, early fusion achieves the best overall performance, with an SSIM of

TABLE II: In-distribution navigation performance.

Methods	Coop.	SR↑	CR↓			TR↓	NT↓	PL↓	AA↓
			Obs	NPC	Robot				
NavMesh	No	0.39 ± 0.05	**0.00 ± 0.00**	0.41 ± 0.06	0.19 ± 0.04	**0.00 ± 0.00**	12.20 ± 0.53	13.89 ± 0.62	0.10 ± 0.05
Perception Only		0.44 ± 0.03	0.13 ± 0.02	0.30 ± 0.02	0.11 ± 0.02	0.02 ± 0.01	15.56 ± 0.46	17.89 ± 0.16	0.17 ± 0.04
Prediction		0.71 ± 0.01	0.12 ± 0.02	0.13 ± 0.01	0.02 ± 0.00	0.01 ± 0.00	14.36 ± 0.38	16.94 ± 0.26	0.14 ± 0.02
CMARL with Uncertainty		0.75 ± 0.02	0.10 ± 0.02	0.10 ± 0.03	0.01 ± 0.00	0.04 ± 0.02	10.82 ± 0.18	13.74 ± 0.16	0.10 ± 0.00
Perception Only	Late	0.47 ± 0.03	0.12 ± 0.02	0.30 ± 0.01	0.09 ± 0.03	0.01 ± 0.01	11.50 ± 0.56	14.70 ± 0.37	0.10 ± 0.03
Prediction		0.78 ± 0.03	0.08 ± 0.02	0.10 ± 0.02	0.01 ± 0.01	0.02 ± 0.02	9.89 ± 0.18	12.23 ± 0.33	0.09 ± 0.07
CMARL with Uncertainty		0.81 ± 0.03	0.07 ± 0.02	0.09 ± 0.02	0.01 ± 0.01	0.01 ± 0.02	10.93 ± 0.16	13.27 ± 0.36	0.09 ± 0.05
CMARL with Uncertainty Fusion		0.88 ± 0.02	0.04 ± 0.01	0.05 ± 0.01	**0.00 ± 0.00**	0.03 ± 0.02	10.45 ± 0.09	13.09 ± 0.39	0.08 ± 0.02
Perception Only	Early	0.50 ± 0.02	0.10 ± 0.02	0.32 ± 0.02	0.07 ± 0.02	0.01 ± 0.01	12.33 ± 0.56	15.22 ± 0.16	0.11 ± 0.08
Prediction		0.77 ± 0.02	0.11 ± 0.01	0.06 ± 0.02	0.01 ± 0.01	0.05 ± 0.02	9.94 ± 0.19	12.55 ± 0.33	0.10 ± 0.06
CMARL with Uncertainty		0.83 ± 0.02	0.06 ± 0.01	0.05 ± 0.01	0.00 ± 0.01	0.06 ± 0.02	11.55 ± 0.36	14.30 ± 0.47	0.15 ± 0.05
CMARL with Uncertainty Fusion (CoopUQ)		**0.91 ± 0.03**	0.04 ± 0.01	**0.02 ± 0.01**	**0.00 ± 0.00**	0.03 ± 0.01	**9.32 ± 0.15**	**11.72 ± 0.29**	**0.07 ± 0.05**

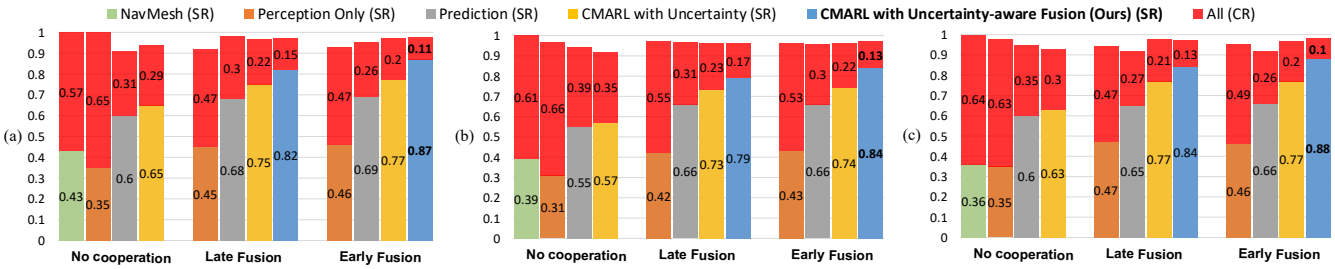


Fig. 3: Success / Collision Rate comparison of OOD tests in (a) unseen layouts; (b) denser crowds; (c) rushing human.

0.86 (+26.5%) and IoU of 0.69 (+32.7%) at 5 time steps. It also incurs minimal communication overhead, requiring only $0.02\,\mathrm{MB/s}$ bandwidth, and maintains efficient runtime with a modest 8% drop in FPS (from 79.6 to 73.2). While late fusion performs slightly below early fusion, likely due to its lack of context aggregation before prediction, which leads to ambiguous predictions of uncertain regions. However, it still substantially outperforms the non-cooperative baseline. These results demonstrate that cooperative prediction enhances spatial-temporal understanding and remains robust under realistic communication constraints.

Social Navigation. The in-distribution experimental results are summarized in Table II. We evaluate 1,000 test episodes across three random seeds and report the mean and standard deviation across three cooperative robots. Among all methods, the rule-based NavMesh planner yields the lowest success rate (SR) and performs poorly in dynamic environments, showing the highest overall collision rate (CR) with pedestrians. This is due to its inability to anticipate agents exhibiting highly dynamic or unpredictable behaviors. However, it maintains a zero collision rate with static obstacles and a minimal timeout rate (TR), owing to its complete knowledge of the static map. Among RL methods, policies that incorporate OGM prediction into their observations significantly improve SR and reduce CR with humans across all three cooperation settings. This highlights the value of OGM prediction for safe navigation in crowded environments. Furthermore, both early and late fusion strategies outperform the no-fusion setting. Notably, the late fusion policy increases SR from 0.71 to 0.78 (+9.9%) and reduces CR with static obstacles from 0.12 to 0.08, and with humans from 0.13 to 0.10. It also offers a balanced trade-off in TR, navigation time (NT), and path length (PL), which indicates improved efficiency and safer interaction with dynamic agents due to the enhancement of prediction.

In addition, CMARL with uncertainty estimation further boosts SR across all cooperation settings. This suggests that incorporating uncertainty estimates and safety constraints into the decision making process enhances the agent's ability to handle ambiguous or high-risk scenarios. SR increases from 0.71 to 0.75 (+5.6%) for the no-fusion baseline, 0.77 to 0.83 (+7.8%) in early fusion, and 0.78 to 0.81 (+3.8%) in late fusion. The larger performance gain observed in early fusion highlights its advantage in leveraging richer contextual cues and benefiting from higher prediction accuracy. Finally, in our constrained and dynamic testing environments where occlusions exist, CoopUQ achieves the best performance. It improves SR from 0.75 to 0.91 (+21.3%) and reduces CR with humans from 0.10 to 0.02. Unlike CMARL with uncertainty-aware fusion in the late fusion setting, where each robot still relies on its local prediction in others' occluded regions, CoopUQ enables robots not only to share sensor data before making predictions but also to collaboratively fuse uncertainty-aware predictions. This adaptive reasoning over a more holistic uncertainty allows agents to better incorporate the predictions from cooperative robots with lower uncertainty, and thus construct a more comprehensive situational awareness and more accurately estimate potential risks. As a result, it leads to more robust decision making.

The out-of-distribution (OOD) test results are presented in Fig. 3, where all learning-based methods exhibit some degree of performance degradation. However, CoopUQ shows the smallest decline. While the rule-based NavMesh approach maintains stable but consistently low performance, our method achieves the highest robustness across all three challenging scenarios. CoopUQ shows greater generalization ability to unseen layouts with the highest SR of 0.87 among others, which is attributed to enhanced situational awareness from cooperative prediction and constrained decision making. Notably, as the number of pedestrians increases, the SR drops most by 24% for the non-cooperative baseline,

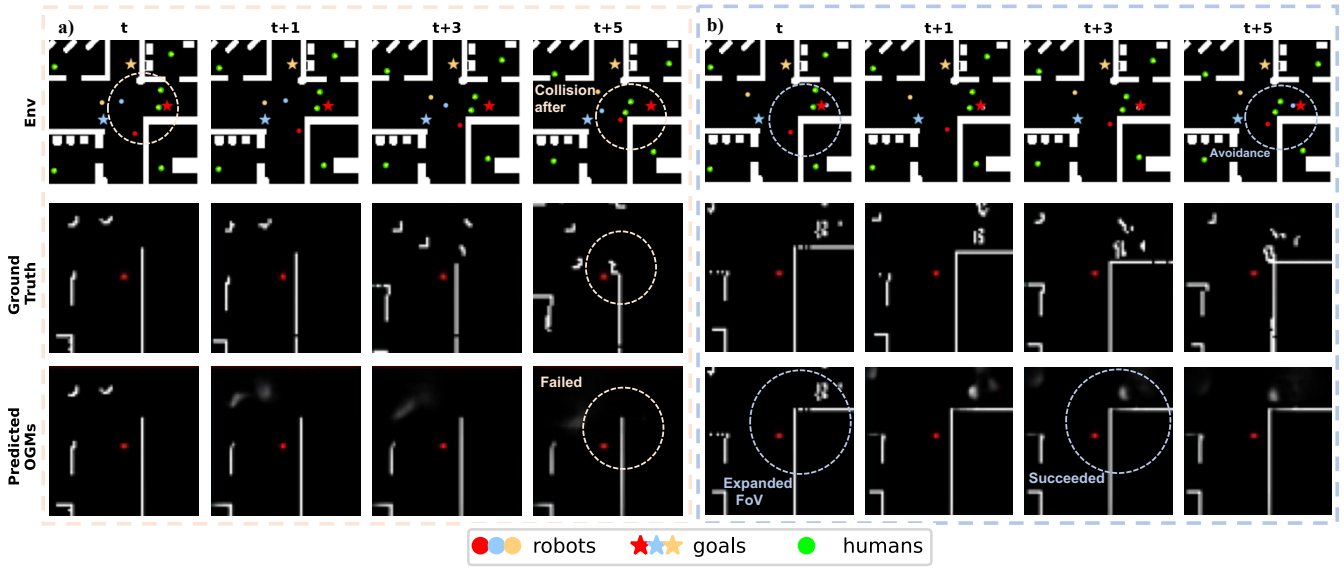


Fig. 4: (a) Prediction without cooperation: The red robot maintains its speed and direction and reacts only after the pedestrian fully emerges into its field-of-view (FoV), resulting in a sharp brake and an eventual collision. (b) Prediction with cooperation: After fusion with the surrounding blue and orange robots, the red ego robot detects potential risks earlier by expanding FoV, thus decelerating and changing direction in advance to avoid a collision.

whereas CoopUQ exhibits only a 7.7% reduction. Compared to CMARL without uncertainty-aware fusion, which leads to a CR increase to 0.22, CoopUQ maintains a clear advantage with a CR of 0.13, which demonstrates that the uncertainty-aware fusion further improves navigation safety in highly dynamic environments where prediction becomes more challenging. Furthermore, CoopUQ performs the best in the scenario involving rushing pedestrians, showing the smallest SR drop of only 3.3%, which implies its ability to effectively handle sudden and unpredictable human behaviors. This widening performance gap under OOD conditions highlights the effectiveness of our approach in improving the reliability of decision making in unfamiliar, dynamic environments.

D. Qualitative Results

Cooperative Prediction. We visualize the predicted OGMs for both the non-cooperative baseline and our cooperative method in Fig. 4. The first row depicts a scenario with diverse obstacles of varying geometries and orientations, where the red ego robot must navigate through crowds toward its goal position. The ground truth OGMs include realistic obstacle shapes, such as double-connected cylinders simulating human legs. Our cooperative method predicts a more complete and accurate scene geometry, significantly enhancing situational awareness. A key challenge in this scenario occurs when a pedestrian emerges from behind a wall corner and crosses the robot's path. In the non-cooperative baseline, the robot fails to anticipate the pedestrian's motion due to occlusion, reacting only after full visibility. This delayed response leads to abrupt braking and ultimately results in a collision. In contrast, CoopUQ leverages shared observations from nearby blue and orange robots, enabling the ego robot to perceive occluded regions and predict the human behind the wall. This improved awareness allows the robot to detect potential risks earlier, proactively adjust its speed and direction, and successfully avoid the pedestrian.

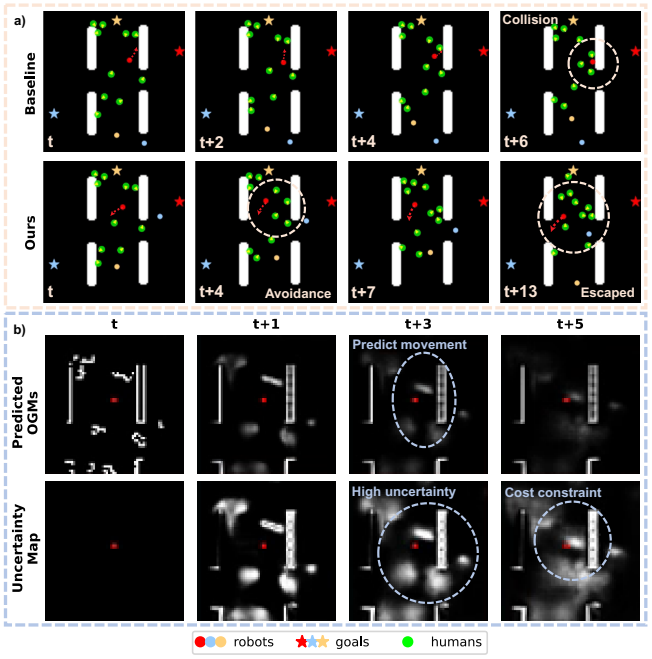


Fig. 5: (a) Comparison of key frames of navigation without and with cooperation in the same test scenario. (b) Predicted OGMs and uncertainty map from Ours at the first step t. Constrained by our policy, the red ego robot proactively avoids entering high-uncertainty areas, thereby preventing potential collisions in advance.

The full episode is available in supplementary videos.

Social Navigation. Fig. 5 illustrates a representative comparison between the baseline and CoopUQ. In this scenario, the red ego robot navigates toward its goal in a cluttered environment with static obstacles and randomly moving pedestrians. In the baseline policy (Fig. 5a), the robot lacks both cooperative prediction and uncertainty estimation. It

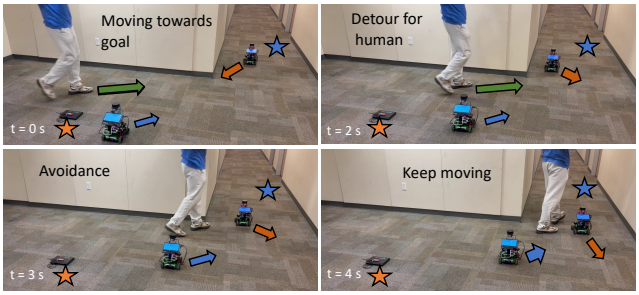


Fig. 6: The upper robot navigates toward its goal while detours early to avoid colliding with a pedestrian emerging from occlusion. Demo is available in supplementary videos.

reacts late, braking abruptly in front of obstacles without proactively adjusting its speed. When surrounded by pedestrians, it becomes trapped and unable to escape. Its limited FoV also leads to misjudgment of the bookshelf's size and shape, causing it to inefficiently follow along the edge of obstacles when the goal is behind it, which results in unsafe and inefficient behaviors. In contrast, CoopUQ (Fig. 5b) leverages shared perception and prediction from the blue robot, significantly expanding the ego robot's FoV. This enables a clearer understanding of the bookshelf's geometry and earlier deceleration. Moreover, the uncertainty maps further enlarge the predicted occupied regions around pedestrians, helping the robot anticipate risks and adjust its trajectory accordingly. When the path is fully blocked, the robot is able to re-plan and select a safer route. Overall, our approach leads to more robust, efficient, adaptive, and socially compliant navigation.

Real Robot Experiments. We deploy our method on two ROSMASTER X3 robots with Mecanum wheels. Each robot uses an Ultra-Wideband (UWB) module for localization, with tag measurements fused via a Kalman filter. Robots connect to a laptop with an NVIDIA RTX 3070 GPU through a router for real-time inference. The policy is trained in Unity simulation with holonomic dynamics and deployed to the robot without fine-tuning. We test in two indoor corridor environments with humans walking around. Fig. 6 shows a scenario where the robot initially navigates directly toward its goal. With the help of shared predictions, it detours early to avoid a potential collision. Once clear, it resumes smooth navigation. This demonstrates CoopUQ in handling real-world interactions via cooperative prediction and uncertainty-aware decision making.

VI. Conclusions

In this work, we introduce CoopUQ, a novel uncertainty-aware multi-robot social navigation framework that integrates cooperative occupancy prediction into multi-agent decision making. CoopUQ enables robots to share and integrate spatial forecasts in a way that prioritizes confident prediction based on uncertainty estimation. By incorporating a safety- and risk-constrained MARL policy, CoopUQ demonstrates strong generalization to unseen environments and robust adaptation to variations in crowd density and velocity in our high-fidelity Unity3D simulator. Real-world experiments also validate the practical effectiveness of our method. In future work, we will investigate multi-modal sensor fusion for heterogeneous robots and explore active communication strategies that dynamically adapt the information sharing based on task demands and bandwidth constraints.

References

[1] 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)*. IEEE, 2021, pp. 1140–1146.

[2] 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.

[3] J. Yao, X. Zhang, Y. Xia, Z. Wang, A. K. Roy-Chowdhury, and J. Li, "Sonic: Safe social navigation with adaptive conformal inference and constrained reinforcement learning," *arXiv preprint arXiv:2407.17460*, 2024.

[4] Z. Bai, G. Wu, M. J. Barth, Y. Liu, E. A. Sisbot, K. Oguchi, and Z. Huang, "A survey and framework of cooperative perception: From heterogeneous singleton to hierarchical cooperation," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[5] R. Xu, Z. Tu, H. Xiang, W. Shao, B. Zhou, and J. Ma, "Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers," in *Conference on Robot Learning*. PMLR, 2023.

[6] Z. Wang, Y. Wang, Z. Wu, H. Ma, Z. Li, H. Qiu, and J. Li, "Cmp: Cooperative motion prediction with multi-agent communication," *IEEE Robotics and Automation Letters*, 2025.

[7] W. Wang, L. Mao, R. Wang, and B.-C. Min, "Multi-robot cooperative socially-aware navigation using multi-agent reinforcement learning," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 12 353–12 360.

[8] N. Mohajerin and S. L. Waslander, "Multistep prediction of dynamic systems with recurrent neural networks," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3370–3383, 2019.

[9] Z. Xie and P. Dames, "Stochastic occupancy grid map prediction in dynamic scenes," in *Conference on Robot Learning*. PMLR, 2023.

[10] Z. Yang, H. Jin, R. Ding, H. You, G. Fan, X. Wang, and C. Zhou, "Decom: Decomposed policy for constrained cooperative multi-agent reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 9, 2023, pp. 10 861–10 870.

[11] H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda, and R. Nakatsu, "Robovie: A robot generates episode chains in our daily life," in *Proceedings of the International Symposium on Robotics*, 2001.

[12] C. S. Chen, C. J. Lin, and C. C. Lai, "Non-contact service robot development in fast-food restaurants," *IEEE Access*, vol. 10, pp. 31 466–31 479, 2022.

[13] 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.

[14] R. Mirsky, X. Xiao, J. Hart, and P. Stone, "Conflict avoidance in social navigation—a survey," *ACM Transactions on Human-Robot Interaction*, vol. 13, no. 1, pp. 1–36, 2024.

[15] 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. 12 015–12 021.

[16] J. Li, C. Hua, H. Ma, J. Park, V. Dax, and M. J. Kochenderfer, "Multi-agent dynamic relational reasoning for social robot navigation," *arXiv preprint arXiv:2401.12275*, 2024.

[17] 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.

[18] M. Toyungyernsub, M. Itkina, R. Senanayake, and M. J. Kochenderfer, "Double-prong convlstm for spatiotemporal occupancy prediction in dynamic environments," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 931–13 937.

[19] D. Dugas, J. Nieto, R. Siegwart, and J. J. Chung, "Navrep: Unsupervised representations for reinforcement learning of robot navigation in dynamic human environments," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 7829–7835.

[20] S. Gu, J. G. Kuba, Y. Chen, Y. Du, L. Yang, A. Knoll, and Y. Yang, "Safe multi-agent reinforcement learning for multi-robot control," *Artificial Intelligence*, vol. 319, p. 103905, 2023.

[21] C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu, "The surprising effectiveness of ppo in cooperative multi-agent games," *Advances in neural information processing systems*, vol. 35, pp. 24 611–24 624, 2022.