Hybrid-TSR: A Distributed MAPF Fusion Framework with Global Map Prior Injection

Huan Peng[1], Lingzhi Li[2*], Nan Che[3], Zhijun Li[4], Rui Lin[2]
[1]School of Computer Science and Technology, Soochow University, Suzhou, China
[2]School of Future Science and Engineering, Soochow University, Suzhou, China
[3]School of Computer Science, Harbin University of Science and Technology, Harbin, China
[4]Faculty of Computing, Harbin Institute of Technology, Harbin, China
20235227024@stu.suda.edu.cn, lilingzhi@suda.edu.cn, chenan@hrbust.edu.cn, zjli@suda.edu.cn, 20245258012@stu.suda.edu.cn

***Abstract*—Centralized solvers for Multi-Agent Path Finding (MAPF) possess global conflict resolution capabilities but incur significant computational and communication overhead in large-scale scenarios. Conversely, distributed planning under local observation offers excellent scalability yet is prone to erroneous decisions and deadlocks in high-density environments due to insufficient information. To address this, we propose Map-Informed Partially Observable MAPF (MI-POMAPF) that incorporates global map prior knowledge and designs a hybrid decision framework, Hybrid-TSR, that integrates deep imitation learning with the distributed heuristic algorithm 1-step collisions (CS-PIBT). This framework trains a Transformer decision model using global prior knowledge and local observations during the offline phase, then embeds it into CS-PIBT during online execution to enhance local decision quality. Experimental results demonstrate that this method significantly improves execution success rates in high-density scenarios while maintaining stable planning overhead, validating its effectiveness and robustness in large-scale multi-robot systems.**

***Index Terms*—MAPF, PO-MAPF, CS-PIBT, Transformer, Imitation learning, Algorithm fusion.**

I. Introduction

As a core foundational problem in multi-robot systems, multi-robot path planning is typically formalized as the Multi-Agent Path Finding (MAPF) problem [1], [2], whose objective is to generate feasible paths for all agents while ensuring collision-free operation. Traditional centralized solvers rely on global maps and position sharing but are constrained by their NP-hard computational nature [3], making them struggle to meet real-time requirements in large-scale, high-density scenarios. Although numerous subsequent suboptimal solvers have sought to balance solution quality and speed, their scalability remains bottlenecked.

In contrast, distributed solvers enable agents to make autonomous decisions based solely on local observations and limited communication, significantly enhancing system scalability. These often incorporate imitation or reinforcement learning for model optimization [4]. Subsequently, researchers considered perception-constrained real-world scenarios, leading to the introduction of the Partially Observable Multi-Agent Path Finding (POMAPF) problem, which requires agents to plan online using only local information [5]. However, under partially observable environments, agents struggle to perceive global conflict structures. Purely learning-based methods often result in overly conservative behaviors, leading to reciprocal avoidance or even livelock [6], [7]. Notably, while real-time observations are local, global map priors are readily available in most practical scenarios. Integrating such priors with local observations enables agents to resolve conflicts more effectively by providing a global context for decision-making.

To address these challenges, this paper proposes Hybrid-TSR (Hybrid Transformer with State-cycle Resolution), a hybrid decision framework guided by a state-cycle detection mechanism (SCDM). First, the Transformer [8] decision model is trained using locally observed data enriched with global map information generated by LaCAM. During execution, the SCDM detects local cycles and dynamically coordinates the locally learned decision model with the 1-step collision-avoidance model (CS-PIBT) [9]. Experimental results demonstrate that this hybrid framework effectively reduces collision frequency and mitigates deadlock issues while maintaining low computational overhead. Furthermore, by incorporating global map information, expert algorithm data, and the SCDM algorithm-switching mechanism, the framework achieves global conflict resolution capabilities that approach those of centralized methods while preserving distributed scalability.

Specifically, the main contributions of this paper are as follows:

- We constructed a large-scale expert dataset with global map priors, enabling Transformers to learn global conflict resolution in partially observable environments.
- We developed Hybrid-TSR, an SCDM-based framework integrating Transformer decision-making with CS-PIBT to enhance deadlock handling through global-distributed strategy fusion.
- Experimental results demonstrate that our method reduces conflicts and deadlocks with low computational overhead, validating its efficacy in partially observable MAPF.

*Corresponding author

II. Related Works

A. Centralized Multi-Agent Path Finding

Centralized MAPF methods rely on a global planner to search in the joint state space. Multi-agent A* [10] explicitly models spatiotemporal conflicts but suffers from exponential complexity, limiting scalability [11]. Methods such as MA-RRT* [12] improve efficiency but remain constrained by agent scale. CBS [13] reduces complexity by decoupling path planning and conflict resolution, while IDCBS [14] further enhances scalability. Learning-based techniques, such as GNN-guided heuristics [15], and suboptimal solvers like ECBS [16] improve efficiency for real-time applications. Priority-based methods such as PIBT [17] and its extensions LaCAM further support large-scale scenarios by reducing coordination overhead. Overall, centralized methods provide high-quality solutions but require global information and incur significant communication and computational costs, limiting their applicability in large-scale systems.

B. Distributed Multi-Agent Path Finding

Distributed MAPF enhances scalability and robustness by leveraging local sensing and limited communication, bypassing centralized bottlenecks. Early methods focus on reinforcement learning–based policies, such as PRIMAL and its variants [18], [19]. To improve stability, communication mechanisms have been introduced, including GNN-based interaction modeling [20], attention-based coordination such as SACHA [6], and global communication strategies like SCRIMP [21]. Recent approaches integrate learning with search and imitation learning, including large-model policies such as MAPF-GPT [22].

Distributed methods still struggle with robustness in dense scenarios. To bridge the gap between global priors and adaptive local coordination, this paper introduces MI-POMAPF, leveraging global map priors to enhance conflict resolution in large-scale, partially observable environments.

III. Problem Definition

A. Map-Informed Partially Observable MAPF

Traditional MAPF relies on full observability, which incurs high costs and limits scalability. While PO-MAPF utilizes local sensing to scale, it neglects accessible global map priors. However, in most practical systems, the environment's static topology is known beforehand.

Motivated by this observation, we propose Map-Informed Partially Observable MAPF (MI-POMAPF). MI-POMAPF preserves the partial observability assumption while incorporating readily available global static map information to complement local observations and enables agent coordination through limited communication. This formulation achieves a better balance between modeling capability and system deployability, making it more suitable for real-world multi-robot systems.

Unlike the classical MAPF, which assumes all agents have access to the global state, agents in the PO-MAPF setting can only rely on local observations and constrained communication for online decision-making.

Specifically, the observation of agent a_i at time step t is defined as:

$$o_i^t = \left(o_{local}^t(i), o_{comm}^t(i)\right) \tag{1}$$

where $o_{local}^t(i)$ denotes local spatial observation information. The observation range is a matrix centered on the agent with radius r, resulting in a total observation size of $2r + 1 \times 2r + 1$. Local observation information typically includes the distribution of obstacles within the observation range, navigable areas, and the relative positions of neighboring agents. Simultaneously, agents can combine known static maps to obtain global static information without involving other agents. Meanwhile, $o_{comm}^t(i)$ denotes communication information from other agents within the local observation range. Under the aforementioned partially observable and communication-constrained conditions, each agent independently selects actions based on its observation o_i^t.

Like standard MAPF, MI-POMAPF seeks conflict-free distributed policies to optimize overall path costs. However, limited local observability restricts preemptive conflict avoidance, rendering agents vulnerable to deadlocks in dense or narrow environments and posing a critical challenge for distributed solvers.

IV. Local Decision Model for Imitation Learning Based on Map Priors

A. Local Observation Data Formats

1) Multi-source heterogeneous local observation metadata: As shown in Fig. 1, Multi-Source Partial Observation integrates a 3D observation matrix with serialized communication data. The 3D matrix encodes local obstacle distributions, relative agent positions, and global cost-to-go information. Simultaneously, neighbors' current/target positions and greedy actions are serialized into a text sequence. This process effectively merges spatial sensing with inter-agent communication for decentralized decision-making. The observation matrix is a three-channel 2D grid within the sensing range. The obstacle channel encodes static barriers (∞) and relative time-step deltas; the agent channel marks occupancy; and the target channel projects the destination into the local grid to provide directional guidance.

Communication information is represented as text sequences summarizing key states of neighboring agents, including positions, targets, recent actions, and greedy decisions, thereby compensating for limited local perception.

This multi-source representation combines the spatial modeling strengths of convolutional networks with high-level semantic reasoning enabled by text encoders, providing effective inputs for distributed decision models.

2) Semantic Serialization and Encoding of Partial Observations: To simplify architecture and unify inputs, we also propose a Text-Based Partial Observation representation (Fig. 1, left). It integrates map-based cost-to-go and inter-agent communication data into a single sequence, which is

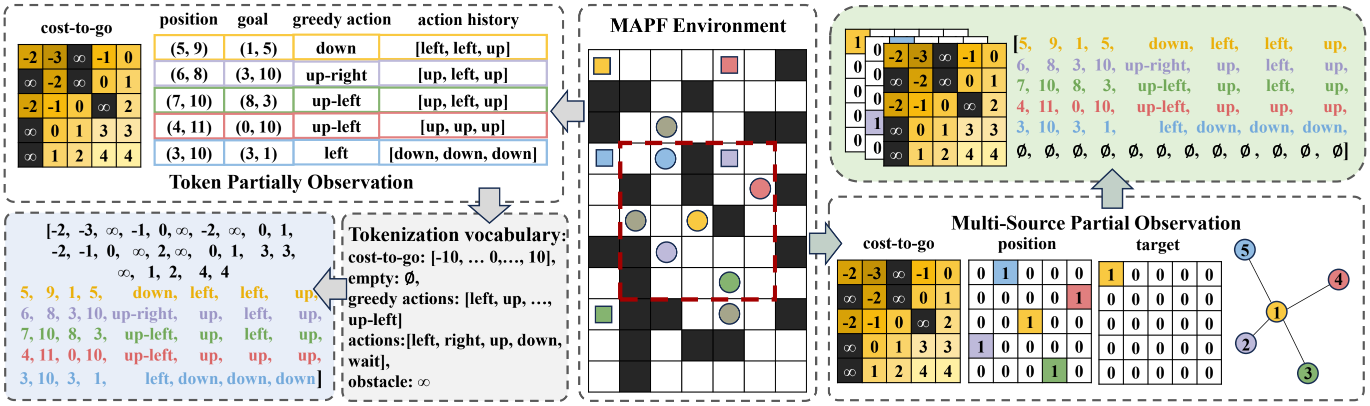

Fig. 1. Representations of partially observable multi-agent environments. The right shows multi-source local observations, while the left shows text-based local observations.

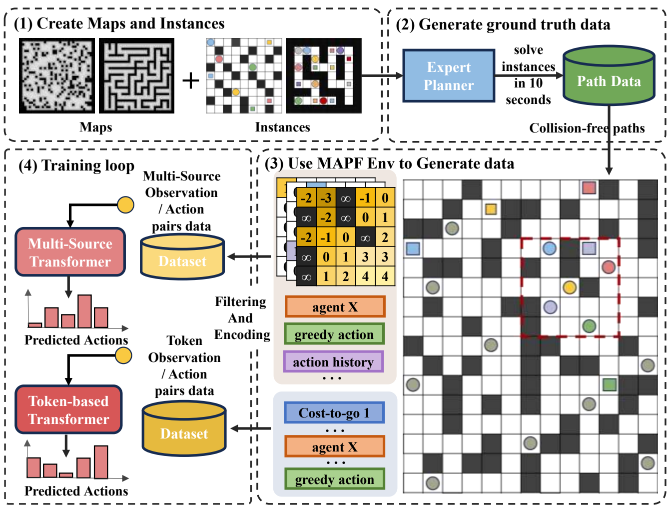

Fig. 2. Framework for constructing local observation data from expert planners. Demonstration paths generated across multiple maps are converted into local observations, filtered, encoded, and used for model training.

then processed by a custom encoder into a unified text-only format. Greedy environmental observations summarize local states comprising movement directions, shortest-path guidance, and conflict risks into structured textual features. These features are concatenated with neighbor communication data in a predefined order to form a unified input sequence for distributed decision-making.

By removing explicit spatial grid structures, this representation is well-suited for Transformer-based architectures. It also offers improved computational efficiency and scalability in large-scale multi-agent settings.

B. Dataset Construction

To construct a high-quality expert dataset, we employ LaCAM [23] as the primary expert algorithm for imitation learning. The LaCAM planner is used to solve a wide range of maps and instances from the MAPF Benchmark [2], with a runtime limit of 10 seconds per instance to ensure the acquisition of high-quality, collision-free path data.

Subsequently, global trajectories are decomposed into time-stepped localized observation-action pairs. Unlike conventional local-only approaches, these observations are enriched with global map priors such as single-agent Cost-to-Go maps and Greedy Action guidance to instill implicit global conflict awareness. The data then undergoes post-processing to remove exact duplicates and filter out redundant endpoint-waiting samples to optimize training performance. Finally, the refined data is structured into Multi-Source and Token-based formats to support distributed decision-making models. The specific dataset construction process is shown in Fig. 2.

Ultimately, 1.19M pairs of local observations and actions were collected. The entire dataset was divided into training and test sets at a 9:1 ratio for different decision models to perform imitation learning.

V. Method

To resolve deadlocks and oscillations in distributed MAPF, this paper introduces Hybrid-TSR (Hybrid Transformer with State-cycle Resolution). Its core mechanism, SCDM (the State-cycle Detection Mechanism), identifies real-time state cycles to trigger adaptive strategy switching, effectively breaking local loops and enhancing robustness.

A. Distributed Decision Model Training

This Transformer-based distributed architecture uses observation-specific processing layers. With a fixed input length of 256 and a 5D discrete action output, final actions are selected via softmax-parameterized multinomial sampling.

During training, the model is optimized using a cross-entropy loss to imitate expert behavior. Mini-batch SGD with the AdamW optimizer is applied, enabling the model to fit the actions generated by the centralized expert planner LaCAM.

While LaCAM has access to full global state and map information, the distributed model relies only on local observations of agent i. To address this information gap, the model

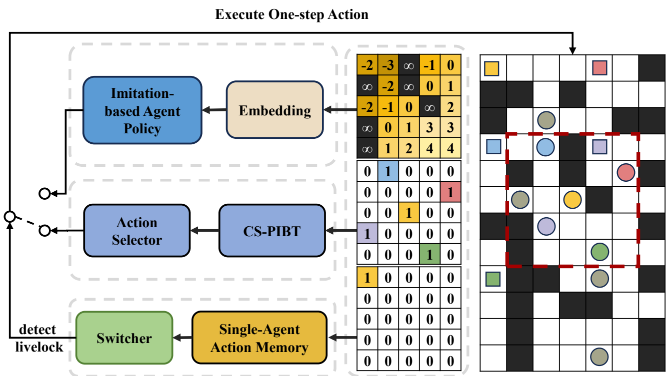

Fig. 3. Hybrid decision framework integrating policy networks with CS-PIBT. An embedding layer provides action priors, while SCDM detects cycles and triggers a switch between the imitation policy and CS-PIBT for conflict resolution.

incorporates global static map priors m, thereby enhancing decision quality under partial observability.

$$\mathcal{L}(\theta) = -\log p_\theta(a_i^{LaCAM}(s) \mid o_i, m_i) \tag{2}$$

Through iterative training, the model strategy progressively adapts to LaCAM and performs action sampling based on the predicted action distribution:

$$\hat{a}^i(o_i) \sim p_\theta(o_i, m_i) \tag{3}$$

where p_θ denotes the action distribution predicted by the distributed decision model, o_u represents the local observation of agent u, and m_u is the information computed by this agent based on the global static map.

Training utilizes Cosine Annealing to decay the learning rate from 6×10^{-4} to 6×10^{-5}, following 2,000 warm-up iterations. To ensure stability and generalization, gradient clipping is set to 1.0 and weight decay to 0.1, facilitating effective model optimization on the constructed dataset.

B. Hybrid Decision Framework: Hybrid-TSR

To address local deadlocks, the Hybrid-TSR framework incorporates a State Cycle Detection Mechanism (SCDM). This mechanism monitors the historical trajectories of individual agents to swiftly resolve local conflicts while maintaining overall system efficiency, as shown in Fig. 3.

Hybrid-TSR operates using a distributed policy model and CS-PIBT. At each time step t, agent i determines if its current state s_i^t has appeared in its historical state set $\mathcal{H}_i$, as tracked by SCDM. When repeated states surpass a set cycle threshold, the Switcher changes between the policy model and CS-PIBT. The chosen algorithm acts based on local observations. Before taking action, the SCDM memories of relevant agents are reset to avoid false cycle detections and ensure precise monitoring with the new strategy. With SCDM, Hybrid-TSR enables on-demand algorithm switching. This approach maintains planning efficiency and utilizes the strengths of both algorithms to mitigate deadlocks in distributed planning, improving overall system completeness and success in complex tasks.

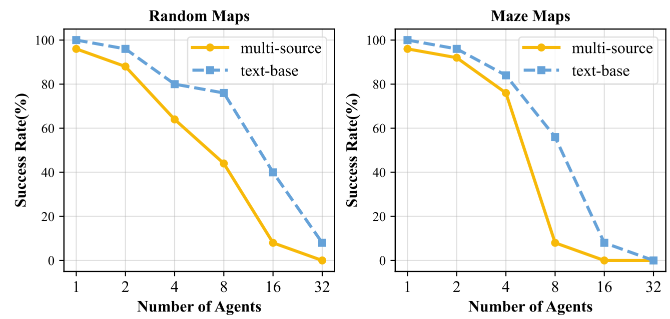

Fig. 4. Local Observation Data Format Experiment.

VI. Experiment

To evaluate Hybrid-TSR in MI-POMAPF, this paper analyzes observation formats, benchmarks against baselines, and conducts information ablation studies.

A. Local Observation Data Format Experiment

Local observation representation is crucial for distributed decision-making. To evaluate this, we compare the planning success rates of multi-source heterogeneous data and text sequences across Random and Maze maps.

The final experimental results are shown in Fig. 4. These experiments demonstrate that text-based serialization of local observations achieves superior performance across various topological environments. Further analysis reveals that text sequences explicitly encode complex agent interactions, shifting intricate relational reasoning from the model's learning phase to the representation stage, thereby significantly reducing Transformer modeling complexity. By contrast, multi-source heterogeneous formats suffer from complex feature alignment and fusion noise, ultimately leading to insufficient decision robustness.

B. Algorithm Comparison Experiment

To validate the effectiveness of the proposed state-cycle detection-based algorithm fusion method in complex environments, this paper conducts comparative experiments using several representative MAPF solvers: DCC [24], DHC [20], When to Switch [5], and SCRIMP [21]. The selection of these solvers was guided by their established performance as reported in previous studies, ensuring a comprehensive reflection of current mainstream algorithms.

The comparative experiments were conducted across four typical map types in the MAPF Benchmark: Empty, Maze, Room, and Random. For consistency, all map instances were sourced from the publicly available MAPF Benchmark [2]. Throughout these experiments, Success Rate and Sum of Costs (SoC) were primarily used as evaluation metrics to fairly compare performance.

Fig. 5 shows the final success rates. Hybrid-TSR consistently outperforms other algorithms, especially on high-density, complex maps. In Empty and Random maps, all methods perform well with few agents, but most degrade as

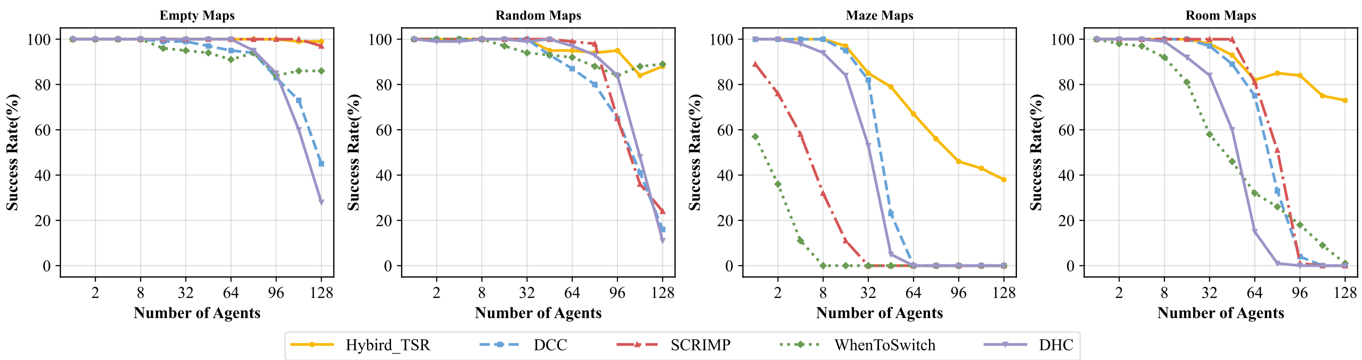

Fig. 5. Success rate of the evaluated MAPF solvers on different maps.

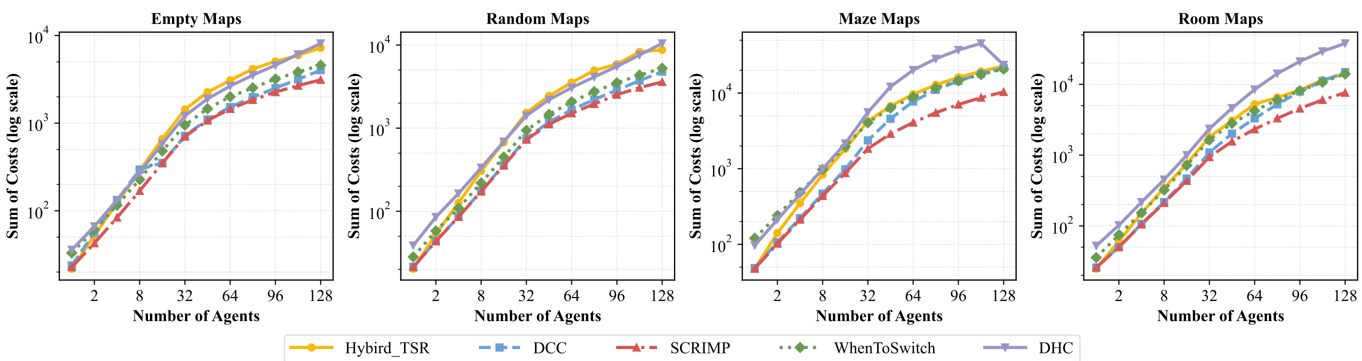

Fig. 6. Comparison of Solution Quality Among MAPF Solvers (Lower SoC values indicate better quality).

agent numbers rise, with only When to Switch and SCRIMP maintaining moderate success. In Maze and Room maps, Hybrid-TSR still leads, though with some decline.

Fig. 6 presents the total solution cost (SoC). Costs increase logarithmically with more agents. Hybrid-TSR achieves near-optimal efficiency, but its SoC is slightly higher at large agent counts due to the SCDM mechanism and long-range avoidance, which proactively resolve deadlocks, increasing per-step cost but ensuring solutions within a finite number of steps.

C. Ablation Study

To validate the algorithm fusion and its deadlock-resolution capabilities in narrow passages, we conduct ablation studies against CS-PIBT. Experiments utilize numerous 17×17 Maze and Random maps featuring bottleneck paths, generated alongside corresponding MAPF instances using POGEMA [25].

As shown in Table I, Hybrid-TSR significantly outperforms CS-PIBT on both 17×17 Maze and Random maps. Comparative analysis reveals that Hybrid-TSR achieves substantial improvements in perception and conflict-resolution capabilities within high-density, complex environments, particularly in deadlock resolution through the effective integration of global map prior information injection and algorithm fusion mechanisms.

TABLE I
SUCCESS RATE COMPARISON BETWEEN CS-PIBT AND HYBRID-TSR ON DIFFERENT MAPS.

Method	Maze Maps		Random Maps	
	8	16	8	16
CS-PIBT	83.5% ±6.5	51.5% ±14.5	92.0% ±1.0	62.0% ±1.0
Hybrid-TSR	**95.5%** ±0.5	**78.5%** ±3.5	**99.0%** ±1.0	**92.5%** ±0.5

TABLE II
SUCCESS RATES OF THE BASE MODEL AND ABLATED VARIANTS ACROSS MAPS.

Map	Agents	Base	noC2G	noLoc	noGoal	noGA
Empty-32	1	**100%**	84%	88%	96%	92%
	4	**98%**	36%	72%	92%	80%
	16	**75%**	8%	8%	72%	36%
Random-32	1	**100%**	52%	84%	84%	84%
	4	**100%**	8%	44%	48%	48%
	16	**82%**	0%	0%	12%	12%
Maze-32	1	**100%**	12%	20%	12%	16%
	4	**100%**	0%	0%	0%	0%
	16	**92%**	0%	0%	0%	0%
Room-32	1	**100%**	12%	24%	28%	32%
	4	**100%**	0%	4%	0%	4%
	16	**93%**	0%	0%	0%	0%

To analyze the impact of different observation dimensions in Hybrid-TSR, four ablation experiments were conducted on Empty, Random, Maze, and Room maps. The Base model uses the full Transformer decision model with text-serialized observations. Ablations sequentially removed key inputs: noC2G (global cost guidance), noLoc (agent location), noGoal (agent goal), and noGA (global greedy actions). Comparing success rates across agent scales highlights each feature's role in complex path planning.

Table II shows that the Base model maintains strong performance, exceeding 68% success even with 32 agents in complex maps. C2G is critical for movement decisions; its removal causes success to drop to 0 in Random maps. GA helps with a few agents but loses effectiveness as density rises. Loc and Goal are essential for avoiding physical conflicts; without them, planning in dense Maze maps nearly fails.

In short, integrating diverse observation features allows agents to make correct decisions and provides a strong foundation for SCDM-based deadlock resolution.

VII. Conclusion

This paper addresses the challenges of local information constraints and cooperative deadlocks in MI-POMAPF by proposing a Hybrid-TSR decision framework. This framework resolves the aforementioned issues from a distributed perspective by integrating an imitation learning decision model based on a global map prior with CS-PIBT. By incorporating global map prior knowledge, this paper employs the centralized expert algorithm LaCAM to construct a local observation dataset and trains the distributed decision model Transformer. Finally, SCDM effectively integrates this decision model with CS-PIBT to form Hybrid-TSR. Extensive experiments demonstrate Hybrid-TSR's capability to solve MI-POMAPF problems. Future research will focus on enhancing the framework's generalization ability in dynamic obstacle avoidance and heterogeneous communication environments, while deeply exploring the tightly coupled collaborative mechanism between learning models and symbolic rules.

Acknowledgment

This work is supported by Suzhou Key Core Technologies Breakthrough Project under Grant SYG2024144 and SNG2025010.

References

[1] J. Li, A. Tinka, S. Kiesel, J. W. Durham, T. S. Kumar, and S. Koenig, "Lifelong multi-agent path finding in large-scale warehouses," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 13, 2021, pp. 11 272–11 281.

[2] R. Stern, N. Sturtevant, A. Felner, S. Koenig, H. Ma, T. Walker, J. Li, D. Atzmon, L. Cohen, T. Kumar *et al.*, "Multi-agent pathfinding: Definitions, variants, and benchmarks," in *Proceedings of the International Symposium on Combinatorial Search*, vol. 10, no. 1, 2019, pp. 151–158.

[3] J. Yu and S. M. LaValle, "Optimal multirobot path planning on graphs: Complete algorithms and effective heuristics," *IEEE Transactions on Robotics*, vol. 32, no. 5, pp. 1163–1177, 2016.

[4] J.-M. Alkazzi and K. Okumura, "A comprehensive review on leveraging machine learning for multi-agent path finding," *IEEE Access*, vol. 12, pp. 57 390–57 409, 2024.

[5] A. Skrynnik, A. Andreychuk, K. Yakovlev, and A. I. Panov, "When to switch: planning and learning for partially observable multi-agent pathfinding," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[6] Q. Lin and H. Ma, "Sacha: Soft actor-critic with heuristic-based attention for partially observable multi-agent path finding," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 5100–5107, 2023.

[7] K. Okumura, F. Bonnet, Y. Tamura, and X. Défago, "Offline time-independent multiagent path planning," *IEEE Transactions on Robotics*, vol. 39, no. 4, pp. 2720–2737, 2023.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[9] R. Veerapaneni, Q. Wang, K. Ren, A. Jakobsson, J. Li, and M. Likhachev, "Improving learnt local mapf policies with heuristic search," in *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 34, 2024, pp. 597–606.

[10] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, 1968.

[11] T. Standley, "Finding optimal solutions to cooperative pathfinding problems," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 24, no. 1, 2010, pp. 173–178.

[12] M. Čáp, P. Novák, J. Vokřínek, and M. Pěchouček, "Multi-agent rrt*: Sampling-based cooperative pathfinding," *arXiv preprint arXiv:1302.2828*, 2013.

[13] G. Sharon, R. Stern, A. Felner, and N. R. Sturtevant, "Conflict-based search for optimal multi-agent pathfinding," *Artificial intelligence*, vol. 219, pp. 40–66, 2015.

[14] E. Boyarski, A. Felner, D. Harabor, P. J. Stuckey, L. Cohen, J. Li, and S. Koenig, "Iterative-deepening conflict-based search," in *International Joint Conference on Artificial Intelligence-Pacific Rim International Conference on Artificial Intelligence 2020*. Association for the Advancement of Artificial Intelligence (AAAI), 2020, pp. 4084–4090.

[15] C. Yu, Q. Li, S. Gao, and A. Prorok, "Accelerating multi-agent planning using graph transformers with bounded suboptimality," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 3432–3439.

[16] M. Barer, G. Sharon, R. Stern, and A. Felner, "Suboptimal variants of the conflict-based search algorithm for the multi-agent pathfinding problem," in *Proceedings of the international symposium on combinatorial Search*, vol. 5, no. 1, 2014, pp. 19–27.

[17] K. Okumura, M. Machida, X. Défago, and Y. Tamura, "Priority inheritance with backtracking for iterative multi-agent path finding," *Artificial Intelligence*, vol. 310, p. 103752, 2022.

[18] G. Sartoretti, J. Kerr, Y. Shi, G. Wagner, T. S. Kumar, S. Koenig, and H. Choset, "Primal: Pathfinding via reinforcement and imitation multi-agent learning," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2378–2385, 2019.

[19] M. Damani, Z. Luo, E. Wenzel, and G. Sartoretti, "Primal _2: Pathfinding via reinforcement and imitation multi-agent learning-lifelong," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2666–2673, 2021.

[20] Z. Ma, Y. Luo, and H. Ma, "Distributed heuristic multi-agent path finding with communication," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 8699–8705.

[21] Y. Wang, B. Xiang, S. Huang, and G. Sartoretti, "Scrimp: Scalable communication for reinforcement-and imitation-learning-based multi-agent pathfinding," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 9301–9308.

[22] A. Andreychuk, K. Yakovlev, A. Panov, and A. Skrynnik, "Mapf-gpt: Imitation learning for multi-agent pathfinding at scale," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 22, 2025, pp. 23 126–23 134.

[23] K. Okumura, "Lacam: Search-based algorithm for quick multi-agent pathfinding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 10, 2023, pp. 11 655–11 662.

[24] Z. Ma, Y. Luo, and J. Pan, "Learning selective communication for multi-agent path finding," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1455–1462, 2021.

[25] A. Skrynnik, A. Andreychuk, A. Borzilov, A. Chernyavskiy, K. Yakovlev, and A. Panov, "Pogema: A benchmark platform for cooperative multi-agent pathfinding," *arXiv preprint arXiv:2407.14931*, 2024.