The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Reinforcement Learning with Fuzzy Human Attention-Guided Graph for Heterogeneous Multiagent Systems

Dingbang Liu[1,2], Fenghui Ren[1]*, Jun Yan[1], Guoxin Su[1], Shohei Kato[2], Wen Gu[2]
[1]School of Computing and Information Technology, University of Wollongong, Australia
[2]Department of Computer Science, Nagoya Institute of Technology, Japan
dl956@uowmail.edu.au, fren@uow.edu.au, jyan@uow.edu.au
guoxin@uow.edu.au, shohey@nitech.ac.jp, wgu@nitech.ac.jp

Abstract

Effective agent coordination is crucial in cooperative Multiagent Reinforcement Learning (MARL). While recent advances have significantly improved cooperation by modeling agent interactions through various graph structures, most existing approaches primarily focus on homogeneous agents. Despite the ubiquity of heterogeneous agents, constructing a comprehensive graph that captures their diverse attributes and relationships from scratch is notoriously labor-intensive for both humans and agents, which makes policy learning extremely challenging. To tackle this difficulty, we propose a novel method that utilizes a fuzzy human attention-guided graph to model inter-agent relationships. Instead of learning the graph entirely from scratch, we incorporate abstract human attention, with its uncertainty captured through fuzzy logic, to guide the graph development process. To further accommodate the varying attributes and objectives of heterogeneous agents while maintaining their learning capabilities, the attention-guided graph is fine-tuned through a hyper-network. Our proposed approach is end-to-end trainable and agnostic to specific MARL methods. Empirical evaluations conducted on challenging heterogeneous scenarios from the StarCraft Multiagent Challenge (SMAC) and SMACv2 validate the effectiveness of the proposed method.

1 Introduction

Multiagent reinforcement learning (MARL) has emerged as an extremely active research field and has achieved notable success in multiagent systems (MAS) (Zhang, Yang, and Başar 2021; Du and Ding 2021). Through trial-and-error, MARL agents leverage their inherent learning capabilities to develop sophisticated policies in increasingly complex scenarios. Despite the abundance of MARL methods, several key challenges remain unresolved. One of the foremost issues is scalability, where MARL methods suffer from the curse of dimensionality when dealing with increasing number of agents (Yu 2023). To address this, vector representations and relational abstractions via graphical models have been introduced, enabling the integration of MARL with graph neural networks (GNNs) as a new paradigm to mitigate scalability concerns (Wang et al. 2022; Hu et al. 2024). However, despite attracting extensive attention, most existing studies in this direction have focused on homogeneous agents, while constructing effective GNNs for heterogeneous environments has received relatively less discussion compared to the prevalence of heterogeneous agents (Wang et al. 2023; Peng et al. 2024). Indeed, agent heterogeneity, with the diversity of relationships and interactions, presents substantial challenges to effective policy learning.

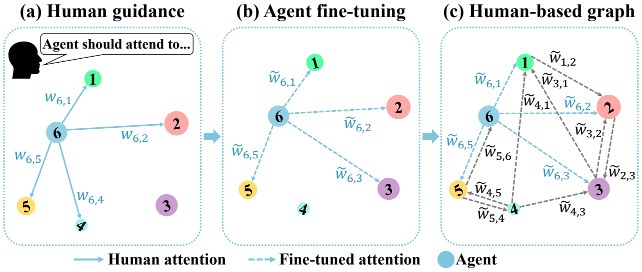


Figure 1: Human attention-guided graph. (a) Based on local observations, humans indicate their attention tendencies for each agent. (b) Each agent then fine-tunes the proposed attention using its own local observation and internal knowledge. (c) By aggregating the fine-tuned attentions from all agents, a human attention-guided graph is constructed.

Compared to homogeneous cooperation, heterogeneous agents maintain individual goals, possess personal knowledge, and exhibit varied abilities, significantly expanding the action-state space and requiring distinct coordination mechanisms (Gronauer and Diepold 2022). Given that agents typically make decisions based on their unique information from potentially hundreds of input features, effective inter-agent communication becomes increasingly critical (Nguyen, Nguyen, and Nahavandi 2020). Consequently, using a homogeneous fully connected graph may inadequately represent the complexity and diversity inherent in heterogeneous interactions (Duan, Lu, and Xuan 2024). Furthermore, due to inherent start-up and free-rider problems (Feng et al. 2023), developing a comprehensive coordination graph from scratch is intricate and challenging. These diverse agent attributes and complex interactions naturally impose greater demands on graph design. To alleviate these difficulties, knowledge transfer methods have attracted increased atten-

---

*Corresponding Author: Fenghui Ren.

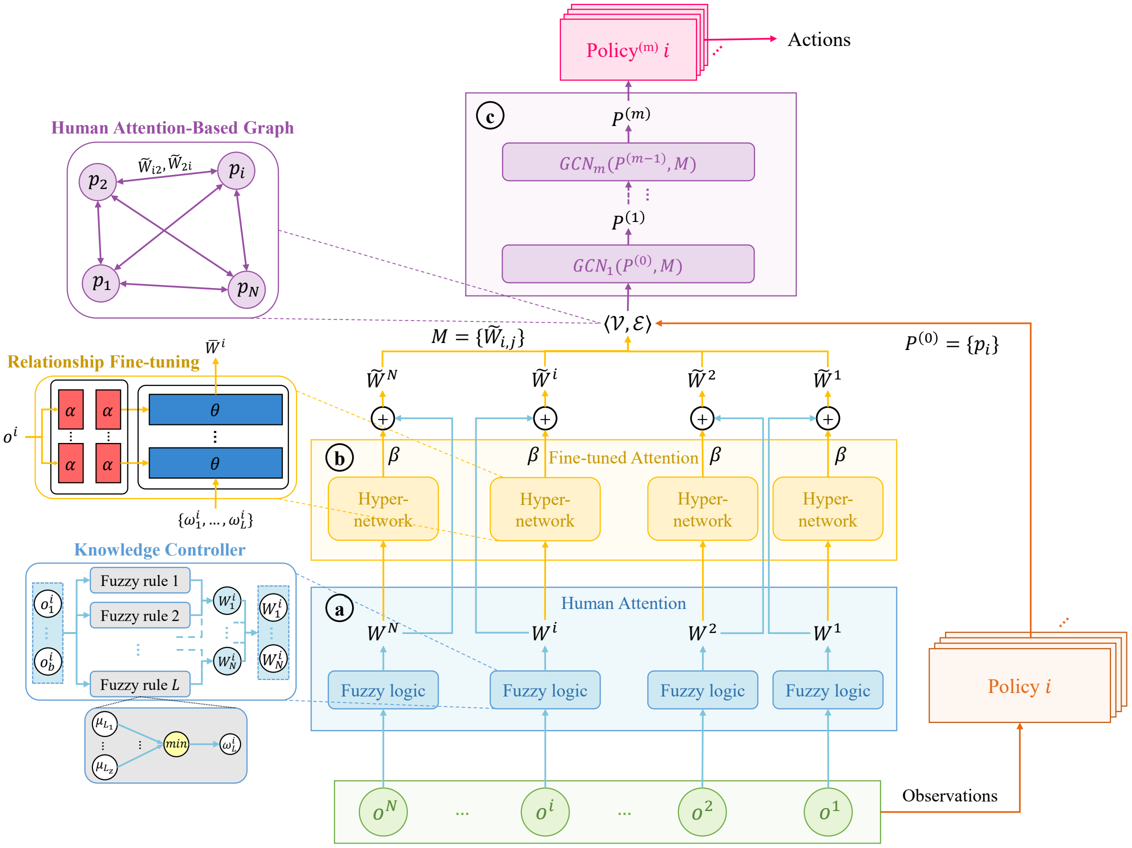


Figure 2: The overall framework of HAGG. (a) Based on local observations, human attention guidance is generated using built-in fuzzy logic rules; (b) The proposed guidance is fine-tuned via a hyper-network to maintain the learning capabilities of agents and adapt to agent heterogeneity; (c) Graph-based policies are derived through GCNs to enhance overall coordination.

tion. Particularly, transferring prior human knowledge has demonstrated significant promise in improving interpretability and performance, despite the challenge of obtaining high-quality expert demonstrations (Glanois et al. 2024). Yet, the effective integration of abstract human guidance into heterogeneous agent learning remains a topic worth discussing.

To address these challenges, we propose constructing a fuzzy human attention-guided graph for heterogeneous multiagent environments, combining the benefits of knowledge transfer with agent adaptive learning capabilities. Motivated by the commonality of human attention across different populations (Langley et al. 2023) and its adaptation into distinct mechanisms to support diverse capabilities (Xia et al. 2024), we leverage human attention as a general paradigm for knowledge transfer. As illustrated in Figure 1a, instead of training from scratch, abstract human attention guides initial graph construction, significantly reducing reliance on expert domain knowledge through the capability of fuzzy logic to handle uncertainty and ambiguity. Since humans and agents have different knowledge structures, agent learning flexibility is maintained via a hyper-network, which fine-tunes human attention to form connections tailored to diverse attributes and objectives of agents, as shown in Figure 1b. By aggregating these fine-tuned connections, the proposed fuzzy human attention-guided graph enhances cooperation among heterogeneous agents, as illustrated in Figure 1c. By integrating the fuzzy human attention-guided graph into the MARL process, the proposed framework, which is end-to-end and can be flexibly combined with a wide range of MARL algorithms, balances knowledge transfer and autonomous learning while avoiding both negative knowledge transfer and the difficulties of learning from scratch.

2 Preliminaries

We focus on cooperative multiagent tasks modeled as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP) (Oliehoek and Amato 2016), defined by the tuple $\langle S, U, P, r, Z, O, n, \gamma \rangle$, where $s \in S$ denotes the global environment state. At each time step, each agent $i \in n \equiv \{1, \dots, N\}$ observes the state partially by receiving an observation $o^i \in O^i$ based on the observation function $Z(s, i)$ and selects an action $u^i \in U^i$ according to its stochastic policy $\pi_i : O^i \times U^i \to [0, 1]$. The individual actions form a joint action $\vec{u} \in U$, which leads to the next state s' via the transition function $P(s'|s, \vec{u})$, and generates a reward $r(s, \vec{u})$ shared by all agents. Agents learn to collectively maximize the global return $\mathbb{E}_{\vec{u}_{t+1} \sim \vec{\pi}, s_{t+1} \sim P}[\sum_{a=0}^{Th-t} \gamma^a r_{t+a}(s_t, \vec{u}_t)|s_t, \vec{u}_t]$ where γ

is the discount factor and Th denotes the time horizon.

Learning the coordination among agents can be viewed as the development of a meaningful dynamic graph topology. This graph is represented as $\langle \mathcal{V}, \mathcal{E} \rangle$, where $\mathcal{V} = n$ is the set of nodes (agents), and $\mathcal{E}$ is the set of edges (relationships) between them. Since the proposed guidance is allowed to be abstract, fuzzy logic is employed for knowledge representation. The formal definition of fuzzy logic and an intuitive example of graph construction and human knowledge integration are provided in Appendix A for further clarification.

3 Methodology

In this section, we present our proposed method, Human Attention-Guided Graph (HAGG). As illustrated in Figure 2, fuzzy human guidance is integrated into the learning process of RL agents to enhance cooperation among heterogeneous agents, with the use of fuzzy logic, the hyper-network, and graph convolutional networks (GCNs).

3.1 Fuzzy Human Attention Integration

Human attention mechanisms exhibit both commonalities across ages and races (Langley et al. 2023), as well as distinctions that support diverse abilities and behaviors (Xia et al. 2024). Motivated by its unique properties, human attention is leveraged to facilitate cooperation among heterogeneous agents and avoid learning from scratch. Since even experts face challenges in providing comprehensive demonstrations in heterogeneous MAS, abstract human knowledge, which is represented through fuzzy logic, is applied to alleviate the burden on humans. By adapting to ambiguity, fuzzy logic aligns with human reasoning and provides both enhanced interpretability (Zhang et al. 2020) and generalization capabilities (Li, Shi, and Hwang 2022). Generally, a fuzzy logic rule for modeling human attention is formed as:

- *Rule* l: IF O_1^i is F_{l_1} AND O_2^i is F_{l_2} AND $\ldots$ AND O_z^i is F_{l_z}, THEN *Agent* is *Attended*

Here, O_z^i denotes the environment variable, and F_{l_z} represents the corresponding fuzzy set in rule l. For fairness, human attention guidance is proposed based on the local observation of agents. Following the principle of fuzzy logic, the strength of the conclusion for each rule is calculated as:

$$\omega_l^i = \left\{\omega_{l,1}^i, ..., \omega_{l,N}^i\right\} = min[\mu_{l_1}(o_1^i), ..., \mu_{l_z}(o_z^i)] \quad (1)$$

where $o_z^i = \{o_{z,1}^i, ..., o_{z,N}^i\}$ represents the observation values of O_z^i related to agents, and μ_{l_z} denotes the membership function for the fuzzy set F_{l_z}. The proposed human attention guidance, together with the corresponding fuzzy logic rules, is shared among all agents to maintain scalability. As illustrated in Figure 2a, for agent i, the local observation o^i is used to generate L human attention vectors $\{\omega_1^i, ..., \omega_L^i\}$ based on built-in fuzzy logic rules. Therefore, the human attention weights for agent i are represented as:

$$W^i = \left\{W_1^i, ..., W_N^i\right\} = mean[\omega_1^i, ..., \omega_L^i] \quad (2)$$

where the mean operator is applied to aggregate all guidance, and W_j^i denotes the human attention weight indicating the level of attention humans recommend agent i to pay to agent j. A higher value corresponds to a higher degree of attention assigned (an example is provided in Section 4.1).

Algorithm 1: MARL with human attention-guided graph

```
Input: MARL algorithm, human attention
Output: HAGG
Initialize the parameters of the MARL algorithm, the
  fine-tuning hyper-network, and GCN layers.
for episode = 1 to max-episode do
  for $t$ = 1 to max-episode-length do
    for agent $i$ = 1 to $N$ do
      Obtain agent policy $p_i$ from its local network.
      Derive human attention $W^i$ from the local ob-
        servation using built-in fuzzy logic.
      Generate fine-tuned $\widetilde{W^i}$ via the hyper-network.
    end for
    Aggregate the edges $\widetilde{W^i}$ and policies $p_i$ from all
      agents to construct the graph $\langle \mathcal{V}, \mathcal{E} \rangle$.
    Gain graph-based policies $P^{(m)}$ through the GCN.
    Sample action $u^i$ from $p_i^{(m)}$.
    Execute actions $(u^1, \ldots, u^N)$ in the environment
      to obtain reward $r$ and next observation $o'$.
    Update the networks following the MARL process.
  end for
end for
```

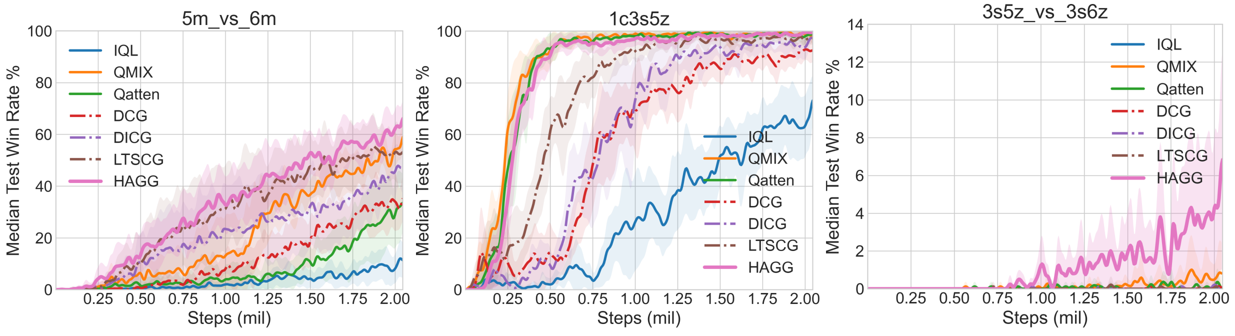
3.2 Fine-Tuning with Hyper-Network

Despite the advantages offered by knowledge transfer from human attention, agents and humans possess different perceptions and knowledge structures, making it inappropriate for agents to follow the guidance directly. As the proposed human attention is highly abstract, it is essential to maintain the learning ability of agents to avoid negative knowledge transfer and accommodate heterogeneity. To address the dynamic knowledge requirements of heterogeneous agents under varying states, the hyper-network is applied to fine-tune the proposed human guidance, as illustrated in Figure 2b. Unlike traditional neural networks, the hyper-network dynamically generates the weights of another network through a feed-forward architecture. This mechanism facilitates the development of context-dependent functions for effective knowledge utilization (Zhang et al. 2020), while also alleviating gradient conflicts and catastrophic forgetting during multiagent policy learning, which is critical for heterogeneous agent cooperation (Tessera et al. 2025). Formally, two networks are involved in the hyper-network structure for the fine-tuning process of each agent. According to Equation 4, the first network dynamically generates the parameters θ of the second network based on the local observation o^i of agent i. Without distorting the proposed human attention while also resolving potential knowledge conflicts, the agent fine-tuned attention weights $\overline{W}^i$ are formed through the generated parameters θ, taking the human attention vectors $\{\omega_1^i, ..., \omega_L^i\}$ as inputs, as expressed in Equation 3:

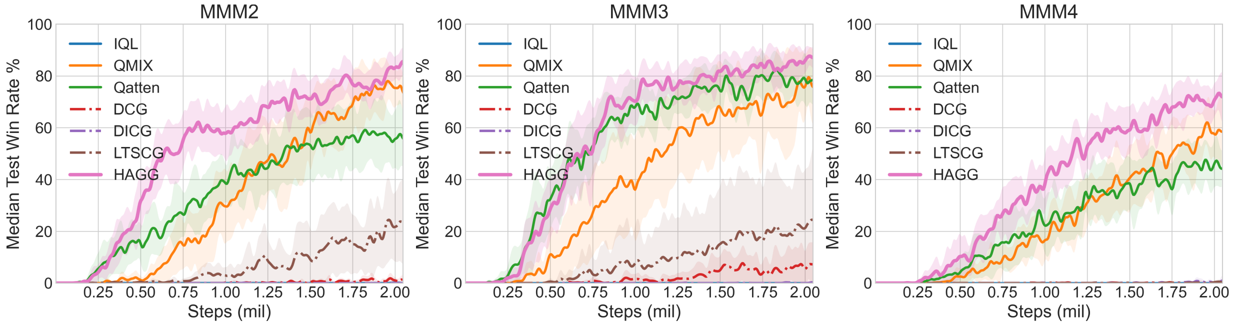
$$\overline{W}^i = \left\{\overline{W}_1^i, ..., \overline{W}_N^i\right\} = h_\theta(\{\omega_1^i, ..., \omega_L^i\}) \quad (3)$$

where:

$$\theta = g_\alpha(o^i) \quad (4)$$



(a) 5m vs 6m (b) 1c3s5z (c) 3s5z vs 3s6z



(d) MMM2 (e) MMM3 (f) MMM4

Figure 3: Experimental results in SMAC scenarios. Shaded regions show the standard deviation of the average over 5 trials.

Here, $g_\alpha(\cdot)$ represents the first network with trainable parameters α, which generates the parameters θ of the second network $h_\theta(\cdot)$. Similarly, this hyper-network structure is shared across agents for scalability. During the initial training stage, the proposed human attention guidance should be frequently utilized to maximize the benefits of knowledge transfer before further adaptation by the agents. To achieve this, a hyperparameter β is introduced:

$$\widetilde{W}^i = \beta \cdot \overline{W}^i + W^i \tag{5}$$

where $\widetilde{W}^i$ represents the fuzzy human attention-guided weights for agent i, which are used to construct the graph edge (relationship) set. The hyperparameter β is initially set to a small value and is then rapidly increased to 1 to preserve agent autonomy during the fine-tuning process.

3.3 Human Attention-Guided Graph for MARL

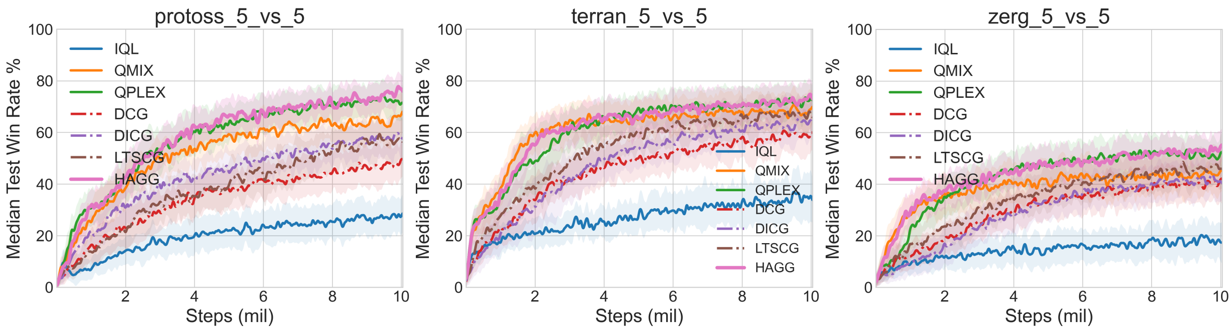
To enhance cooperation among heterogeneous agents, a fuzzy human attention-guided graph is integrated into the MARL process, as illustrated in Figure 2. Instead of learning from scratch or relying on comprehensive expert demonstrations, abstract human attention, represented through fuzzy logic, is applied to guide the development of the agent relationship graph. Specifically, at each time step, human attention weights W^i and vectors $\{\omega_1^i, ..., \omega_L^i\}$ are generated for each agent i. To avoid negative knowledge transfer and accommodate agent heterogeneity, a hyper-network is applied, allowing agents to autonomously utilize and fine-tune the proposed human attention guidance. Based on Equation 5, the fuzzy human attention-guided weights $\widetilde{W}^i$ are derived, benefiting from knowledge transfer while preserving agent learning ability. By aggregating the fuzzy human attention-guided weights from all agents, an edge set $M = \{\widetilde{W}^1, ..., \widetilde{W}^N\}$ is formed to represent the inter-agent relationships. In parallel, local neural networks of each agent generate the initial policies $P^{(0)} = \{p_1, ..., p_N\}$, which are treated as node features. Based on the edge set M and node set $P^{(0)}$, a fuzzy human attention-guided graph $\langle \mathcal{V}, \mathcal{E} \rangle$ is constructed for relationship representation. To enhance heterogeneous agent cooperation, a set of GCNs is deployed to effectively leverage the graph structure, shown in Figure 2c:

$$P^{(k+1)} = GCN_{(k+1)}(P^{(k)}, M), \quad k = 0, \dots, m-1 \tag{6}$$

where m is the number of GCN layers. Through Equation 6, the final graph-based policies $P^{(m)} = \{p_1^{(m)}, ..., p_N^{(m)}\}$ are obtained and used for action selection, driving the environment to transition to the next state. The proposed framework is end-to-end trainable and can be flexibly combined with various MARL algorithms. Its training process follows standard reinforcement learning procedures. Taking QMIX (Rashid et al. 2020) as an example, the selected action Q values from the graph-based policies will be fed into the mixing network of QMIX. The pseudo-code of the proposed method (HAGG) is elaborated in Algorithm 1.

4 Experiments

In our experiments, we aim to answer the following questions: (1) Can the proposed fuzzy human attention-guided graph enhance cooperation among heterogeneous agents



(a) protoss 5 vs 5

(b) terran 5 vs 5

(c) zeg 5 vs 5

Figure 4: Experimental results in SMACv2 scenarios. Shaded regions show the standard deviation of the average over 5 trials.

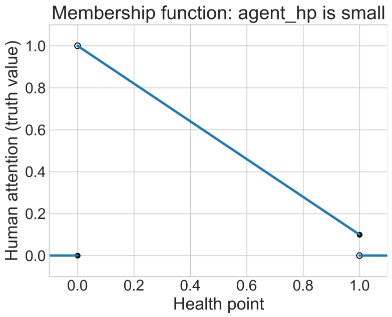


Figure 5: Membership function: '$agent\_hp$ is $small$'. The X-axis represents the observation value of the variable $agent\_hp$, and the Y-axis represents the truth value.

(Section 4.2)? (2) Does our method alleviate the reliance on high-quality expert demonstrations (Section 4.3)? (3) What are the benefits of introducing the hyper-network and avoiding learning from scratch (Section 4.4 and 4.6)? (4) Does the proposed framework exhibit acceptable scalability (Section 4.5)? (5) Is the proposed method end-to-end trainable and compatible with various MARL algorithms (Section 4.5)?

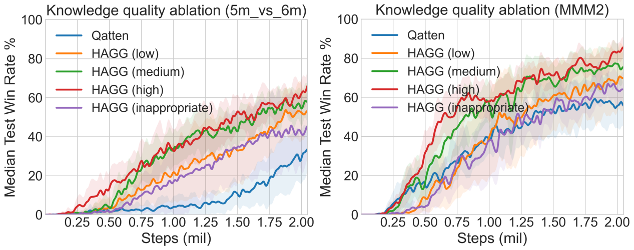
4.1 Experimental Setting

To answer the above questions, we conduct experiments on challenging scenarios from the StarCraft Multiagent Challenge (SMAC) (Samvelyan et al. 2019) and its updated version, SMACv2 (Ellis et al. 2023), both of which demand effective coordination among agents. To evaluate the effectiveness of our end-to-end framework, we consider a range of value-based, policy-based, and graph-based MARL baselines, including IQL (Tampuu et al. 2017), QMIX (Rashid et al. 2020), Qatten (Yang et al. 2020), QPLEX (Wang et al. 2021), IPPO (De Witt et al. 2020), DCG (Boehmer, Kurin, and Whiteson 2020), DICG (Li et al. 2021), and LTSCG (Duan, Lu, and Xuan 2024). In our experiments, without the need for careful crafting, abstract human attention guidance is applied to alleviate dependence on expert domain knowledge. The design of guidance follows the standard fuzzy logic process (Zhang et al. 2020) with basic domain knowledge involved. For further clarification, we illustrate a fuzzy human attention instruction used in our experiments *'Pay attention to agents with low health points'*, which can be represented by the fuzzy logic rule *'IF* $agent\_hp$ *is* $small$*, THEN* $agent$ *is* $attended$*'*. Here, $agent\_hp$ denotes the health points of agents, and the corresponding membership function for the fuzzy set $small$ is defined based on a simple linear function, as shown in Figure 5. As demonstrated, the proposed human attention guidance is both easy to construct and highly interpretable from a human perspective, where agents with lower health points are assigned higher human attention weights. The proposed framework is user-friendly, maintaining the knowledge design freedom for users. To further validate that our framework does not depend on high-quality expert input and requires reduced human effort, the same human attention guidance is consistently applied across all SMAC and SMACv2 scenarios. In this study, all experimental results are obtained over five separate trials with different random seeds, and both the average and standard deviation are reported as evaluation metrics. Additional experimental details, including computational consumption, scenario configurations, hyperparameter settings, tables with summarized results, and the applied human attention guidance, are provided in Appendix A.

4.2 SMAC and SMACv2

Designed for micromanagement control, SMAC has become a widely recognized benchmark for evaluating MARL algorithms due to its diverse scenarios with varying cooperation requirements. Achieving victory in SMAC demands effective collaboration among agents, necessitating both strong collective coordination and precise individual decision-making. In this study, the proposed HAGG method is combined with the Qatten algorithm and evaluated on challenging heterogeneous scenarios from SMAC 'Hard' and 'Super Hard' categories, with the game AI difficulty level set to 'Very Hard'. The learning curves of all compared methods across different tasks are shown in Figure 3. As illustrated, the proposed fuzzy human attention-guided graph significantly improves performance among heterogeneous agents by effectively transferring abstract human knowledge while preserving adaptation to dynamic states.

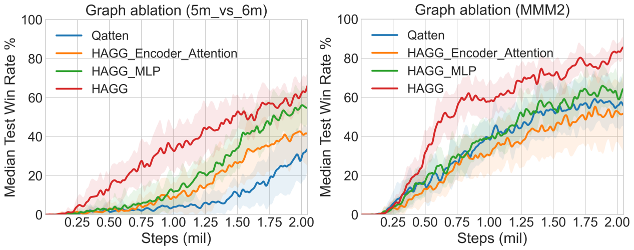
To further validate the advantages of the proposed fuzzy human attention-guided graph, we integrate our method



(a) 5m vs 6m

(b) MMM2

Figure 6: Ablation study on the effect of knowledge quality.



(a) 5m vs 6m

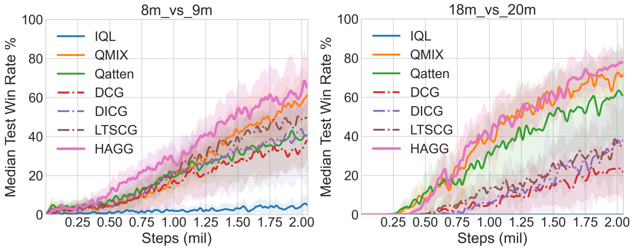
(b) MMM2

Figure 7: Ablation study on graph development.

with the QPLEX algorithm and evaluate its performance on scenarios from the more challenging SMACv2 benchmark, which introduces greater agent heterogeneity and complexity. Unlike SMAC, agents in SMACv2 are constructed using realistic parameters derived from StarCraft II, and both their initial positions and unit types vary randomly across episodes. These variations necessitate the development of more diverse and adaptive strategies. The learning curves obtained from SMACv2 scenarios are presented in Figure 4. Despite the increased challenges, our end-to-end method consistently enhances the performance of MARL algorithms across various tasks, demonstrating the effectiveness of knowledge transfer from fuzzy human attention.

4.3 Knowledge Quality Impact

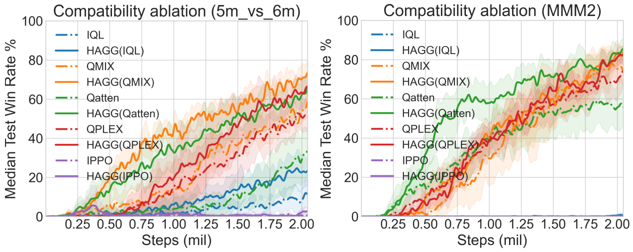
Instead of relying on comprehensive expert domain knowledge, highly abstract human attention is applied to guide the learning process of agents. For further validation, we perform ablation studies by incorporating human attention guidance of varying quality into the Qatten algorithm within our HAGG framework, where knowledge is inherited across ablation approaches, and knowledge covering a broader action-state space is regarded as being of higher quality. Additionally, to assess robustness against unsuitable guidance and knowledge adaptation, noise is integrated to represent inappropriate knowledge. The ablation results are shown in Figure 6. As presented, the proposed framework effectively adapts to abstract human attention guidance, yielding superior performance with higher-quality knowledge, while still maintaining performance improvements even when confronted with inappropriate guidance. Hence, our approach significantly reduces the required human effort and offers greater flexibility in knowledge design and application.



(a) 8m vs 9m

(b) 18m vs 20m

Figure 8: Evaluation of method scalability.



(a) 5m vs 6m

(b) MMM2

Figure 9: Compatibility with various MARL algorithms.

4.4 Hyper-Network and Knowledge Transfer

To validate the benefits of human attention and the hyper-network, additional ablation experiments are conducted using Qatten as the base algorithm (Figure 7). Specifically, at similar capacity, the hyper-network is replaced by multilayer perceptrons for fine-tuning, and attention-encoder layers (Li et al. 2021) are employed to construct the graph from scratch, respectively. As demonstrated, compared to learning from scratch, leveraging the commonalities and distinctions captured by human attention significantly enhances cooperation among agents, especially in heterogeneous settings where intricate relationships impose higher requirements for graph development. Furthermore, the proposed hyper-network facilitates knowledge adaptation across varying states, effectively accommodating agent heterogeneity, which poses challenges for concatenated neural networks.

4.5 Scalability and Compatibility

In this work, a sparse graph is constructed based on the local observations of agents to enhance cooperation among agents while maintaining computational efficiency and scalability. To demonstrate the scalability of the proposed framework, we integrate our method with Qatten and evaluate its performance across scenarios involving an increasing number of agents. The learning curves, depicted in Figure 3a and 8, show scenarios with agent counts ranging from 5 to 18. Despite the growing complexity associated with a greater number of agents, our fuzzy human attention-guided graph still consistently improves performance, whereas baseline algorithms find it increasingly challenging to learn from scratch.

To further demonstrate the compatibility of HAGG, the proposed graph is incorporated into various MARL algorithms. For fairness, each algorithm and its corresponding version integrated with the fuzzy human attention-guided graph share similar hyperparameters and neural network architectures. The experimental results are presented in Figure 9. As illustrated, our end-to-end framework readily combines with diverse value-based and policy-based MARL algorithms, consistently enhancing their performance across challenging homogeneous and heterogeneous scenarios.

4.6 Graph Visualization

To further illustrate the effectiveness of the fuzzy human attention-guided graph, the development of the graph at different training stages in the 'MMM2' scenario is visualized in Figure 10. At the initial stage, the human attention graph is frequently leveraged with minimal modifications, helping agents overcome early-stage training difficulties. As learning progresses, the proposed human attention graph is adaptively fine-tuned to enhance cooperation among heterogeneous agents by adjusting edge weights and dynamically forming or pruning connections. Compared with learning from scratch, the proposed framework more effectively accommodates agent heterogeneity, highlighting the importance of balancing knowledge transfer with learning ability.

5 Related Work

Graphs have been extensively applied to enhance agent cooperation, modeling inter-agent relationships and simplifying complex interactions through local feature aggregation and dimensionality reduction (Khan et al. 2019; Yang et al. 2022). To address the inherent complexities of graph design in MAS, various structures have been explored to facilitate communication among homogeneous agents, aiming for improved scalability and portability (Wang et al. 2022; Hu et al. 2024). Furthermore, graph-based structures show potential in promoting cooperation among heterogeneous agents, although learning effective graphs from scratch remains a non-trivial challenge (Du et al. 2023; Wang et al. 2023; Peng et al. 2024). To alleviate the difficulties of exploration, knowledge transfer techniques are widely employed to reuse previously acquired knowledge across tasks (Da Silva and Costa 2019; Liu et al. 2024). Specifically, fundamental pattern transfer and selective knowledge sharing have been examined to prevent negative knowledge transfer among heterogeneous agents (Shi et al. 2023; Zhong et al. 2024). As a prevalent knowledge source, human-derived knowledge transfer has attracted increasing attention due to its enhanced interpretability compared to 'black-box' agent-to-agent approaches (Glanois et al. 2024). Typically, human demonstrations (action-state pairs) can significantly facilitate RL agent training, despite the requirement for high-quality expert knowledge (Zhang et al. 2018; Yang, Ma, and Xia 2021). To mitigate human effort further, higher-level and more abstract forms of human guidance have been introduced. Benefiting from its ability to represent abstract information, fuzzy logic has been utilized to effectively capture human knowledge in single-agent or homogeneous scenarios (Zhang et al. 2020; Li, Shi, and Hwang 2022). Nonetheless, exploiting abstract human guidance within heterogeneous MAS settings remains an area warranting deeper investigation. Seraj et al. (Seraj et al. 2023) recently combined human guidance with RL to facilitate heterogeneous communication. However, humans only provided action guidance, and the communication policy was still learned from scratch, reportedly due to difficulties in obtaining appropriate demonstrations. In contrast to existing studies, our work leverages fuzzy human attention to guide the construction of the agent relationship graph, avoiding the need to learn from scratch. Concurrently, agents retain autonomy in determining the utilization and adaptation of the proposed graph, enabling more effective accommodation of the unique attributes present in heterogeneous agents.

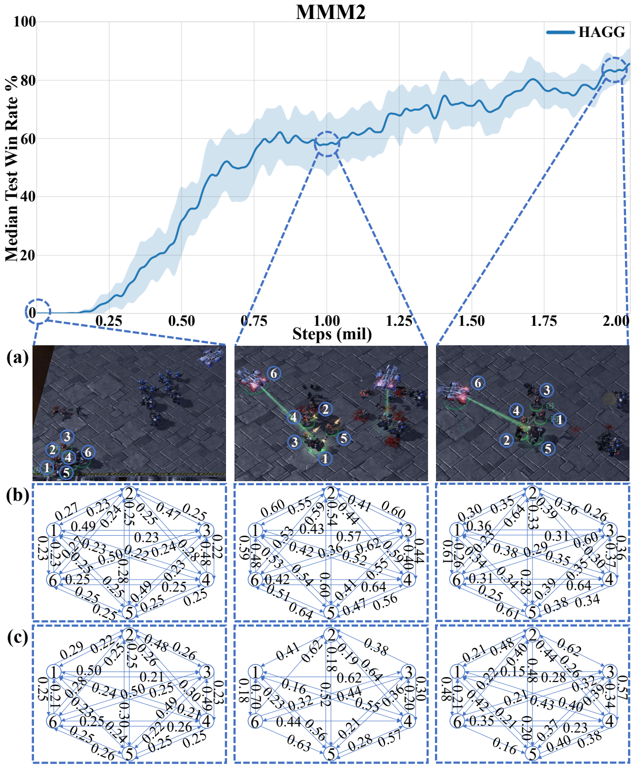


Figure 10: Visualization of graph learning and development on the 'MMM2' map. (a) Screenshots of a specific scenario state at different training stages; (b) The corresponding human attention graph; (c) The agent fine-tuned graph.

6 Conclusion

In this work, we focus on policy learning for heterogeneous agents characterized by distinct attributes within multiagent environments. To effectively represent relationships among agents, we introduce a novel fuzzy human attention-guided graph framework (HAGG), compatible with various MARL algorithms. Experimental results on the SMAC and SMACv2 benchmarks demonstrate that the HAGG framework, leveraging a dynamically adaptive graph, achieves enhanced performance compared to existing MARL methods. For future research, we intend to extend the proposed method to more realistic heterogeneous multiagent settings.

Acknowledgments

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology-Japan, Grant–in–Aid for Scientific Research under grant #JP24H00741, part by the commissioned research by National Institute of Information and Communications Technology (NICT), JAPAN, and part by the JSPS KAKENHI Grant Number #JP24K20900.

References

Boehmer, W.; Kurin, V.; and Whiteson, S. 2020. Deep Coordination Graphs. In *Proceedings of the 37th International Conference on Machine Learning*, 980–991. Virtual Event: PMLR.

Da Silva, F. L.; and Costa, A. H. R. 2019. A survey on transfer learning for multiagent reinforcement learning systems. *Journal of Artificial Intelligence Research*, 64: 645–703.

De Witt, C. S.; Gupta, T.; Makoviichuk, D.; Makoviychuk, V.; Torr, P. H.; Sun, M.; and Whiteson, S. 2020. Is independent learning all you need in the starcraft multi-agent challenge? arXiv:2011.09533.

Du, W.; and Ding, S. 2021. A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications. *Artificial Intelligence Review*, 54(5): 3215–3238.

Du, W.; Ding, S.; Zhang, C.; and Shi, Z. 2023. Multiagent Reinforcement Learning With Heterogeneous Graph Attention Network. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10): 6851–6860.

Duan, W.; Lu, J.; and Xuan, J. 2024. Inferring Latent Temporal Sparse Coordination Graph for Multiagent Reinforcement Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13.

Ellis, B.; Cook, J.; Moalla, S.; Samvelyan, M.; Sun, M.; Mahajan, A.; Foerster, J.; and Whiteson, S. 2023. Smacv2: An improved benchmark for cooperative multi-agent reinforcement learning. In *Advances in Neural Information Processing Systems*, 37567–37593. New Orleans, USA: Curran Associates, Inc.

Feng, Z.; Hu, G.; Dong, X.; and Lü, J. 2023. Adaptively Distributed Nash Equilibrium Seeking of Noncooperative Games for Uncertain Heterogeneous Linear Multi-Agent Systems. *IEEE Transactions on Network Science and Engineering*, 10: 3871–3882.

Glanois, C.; Weng, P.; Zimmer, M.; Li, D.; Yang, T.; Hao, J.; and Liu, W. 2024. A survey on interpretable reinforcement learning. *Machine Learning*, 1–44.

Gronauer, S.; and Diepold, K. 2022. Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*, 55(2): 895–943.

Hu, S.; Shen, L.; Zhang, Y.; and Tao, D. 2024. Learning Multi-Agent Communication from Graph Modeling Perspective. In *Proceedings of the 12th International Conference on Learning Representations*. Vienna Austria: OpenReview.net.

Khan, A.; Tolstaya, E. I.; Ribeiro, A.; and Kumar, V. 2019. Graph Policy Gradients for Large Scale Robot Control. In *Proceedings of the 3rd Annual Conference on Robot Learning*, 823–834. Osaka Japan: PMLR.

Langley, M. D.; Van Houghton, K.; McBeath, M. K.; and Lucca, K. 2023. Children and adults exhibit a common vertical attention bias for object tops and scene bottoms. *Developmental psychology*, 59(8): 1377.

Li, J.; Shi, H.; and Hwang, K.-S. 2022. Using fuzzy logic to learn abstract policies in large-scale multiagent reinforcement learning. *IEEE Transactions on Fuzzy Systems*, 30(12): 5211–5224.

Li, S.; Gupta, J. K.; Morales, P.; Allen, R. E.; and Kochenderfer, M. J. 2021. Deep Implicit Coordination Graphs for Multi-agent Reinforcement Learning. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems*, 764–772. Virtual Event: ACM.

Liu, D.; Ren, F.; Yan, J.; Su, G.; Gu, W.; and Kato, S. 2024. Scaling Up Multi-Agent Reinforcement Learning: An Extensive Survey on Scalability Issues. *IEEE Access*, 12: 94610–94631.

Nguyen, T. T.; Nguyen, N. D.; and Nahavandi, S. 2020. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE transactions on cybernetics*, 50(9): 3826–3839.

Oliehoek, F. A.; and Amato, C. 2016. *A concise introduction to decentralized POMDPs*, volume 1. Springer International Publishing.

Peng, K.; Ma, T.; Jia, L.; and Rong, H. 2024. Enhancing Collaboration in Heterogeneous Multiagent Systems Through Communication Complementary Graph. *IEEE Transactions on Cybernetics*, 54(11): 6881–6894.

Rashid, T.; Samvelyan, M.; de Witt, C. S.; Farquhar, G.; Foerster, J.; and Whiteson, S. 2020. Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. *Journal of Machine Learning Research*, 21(178): 1–51.

Samvelyan, M.; Rashid, T.; De Witt, C. S.; Farquhar, G.; Nardelli, N.; Rudner, T. G.; Hung, C.-M.; Torr, P. H.; Foerster, J.; and Whiteson, S. 2019. The StarCraft Multi-Agent Challenge. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 2186–2188. Montreal, Canada: IFAAMAS.

Seraj, E.; Xiong, J.; Schrum, M.; and Gombolay, M. 2023. Mixed-Initiative Multiagent Apprenticeship Learning for Human Training of Robot Teams. In *Advances in Neural Information Processing Systems*, 35426–35440. New Orleans, USA: Curran Associates, Inc.

Shi, H.; Li, J.; Mao, J.; and Hwang, K.-S. 2023. Lateral transfer learning for multiagent reinforcement learning. *IEEE Transactions on Cybernetics*, 53(3): 1699–1711.

Tampuu, A.; Matiisen, T.; Kodelja, D.; Kuzovkin, I.; Korjus, K.; Aru, J.; Aru, J.; and Vicente, R. 2017. Multiagent cooperation and competition with deep reinforcement learning. *PLOS ONE*, 12(4): 1–15.

Tessera, K.-a. A.; Rahman, A.; Storkey, A.; and Albrecht, S. V. 2025. HyperMARL: Adaptive Hypernetworks for Multi-Agent RL. arXiv:2412.04233.

Wang, J.; Ren, Z.; Liu, T.; Yu, Y.; and Zhang, C. 2021. Qplex: Duplex dueling multi-agent q-learning. In *Proceedings of the 9th International Conference on Learning Representations*. Virtual Event, Austria: OpenReview.net.

Wang, J.; Yuan, M.; Li, Y.; and Zhao, Z. 2023. Hierarchical Attention Master-Slave for heterogeneous multi-agent reinforcement learning. *Neural Networks*, 162: 359–368.

Wang, T.; Zeng, L.; Dong, W.; Yang, Q.; Yu, Y.; and Zhang, C. 2022. Context-Aware Sparse Deep Coordination Graphs. In *Proceedings of the 10th International Conference on Learning Representations*. Virtual Event: OpenReview.net.

Xia, R.; Chen, X.; Engel, T. A.; and Moore, T. 2024. Common and distinct neural mechanisms of attention. *Trends in Cognitive Sciences*, 28(6): 554–567.

Yang, B.; Ma, C.; and Xia, X. 2021. Drone formation control via belief-correlated imitation learning. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, 1407–1415. Virtual Event, United Kingdom: ACM.

Yang, Q.; Dong, W.; Ren, Z.; Wang, J.; Wang, T.; and Zhang, C. 2022. Self-Organized Polynomial-Time Coordination Graphs. In *Proceedings of the 39th International Conference on Machine Learning*, 24963–24979. Baltimore USA: PMLR.

Yang, Y.; Hao, J.; Liao, B.; Shao, K.; Chen, G.; Liu, W.; and Tang, H. 2020. Qatten: A general framework for cooperative multiagent reinforcement learning. arXiv:2002.03939.

Yu, C. 2023. Hierarchical Mean-Field Deep Reinforcement Learning for Large-Scale Multiagent Systems. In *Proceedings of the 37th AAAI conference on artificial intelligence*, 11744–11752. Washington, USA: AAAI Press.

Zhang, K.; Yang, Z.; and Başar, T. 2021. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control*, 325: 321–384.

Zhang, P.; Hao, J.; Wang, W.; Tang, H.; Ma, Y.; Duan, Y.; and Zheng, Y. 2020. KoGuN: Accelerating Deep Reinforcement Learning via Integrating Human Suboptimal Knowledge. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence*, 2291–2297. Virtual, Japan: IJCAI Organization.

Zhang, R.; Liu, Z.; Zhang, L.; Whritner, J. A.; Muller, K. S.; Hayhoe, M. M.; and Ballard, D. H. 2018. AGIL: Learning Attention from Human for Visuomotor Tasks. In *Proceedings of the 15th European Conference on Computer Vision*, 692–707. Munich, Germany: Springer.

Zhong, Y.; Kuba, J. G.; Feng, X.; Hu, S.; Ji, J.; and Yang, Y. 2024. Heterogeneous-Agent Reinforcement Learning. *Journal of Machine Learning Research*, 25(32): 1–67.