

## 1 NUMERICAL RESULTS FOR BARPLOTS

**Table 1: Performance evaluation of the grid-world task**

| Group          | Method        | Best Episode Return           | Comparison       |
|----------------|---------------|-------------------------------|------------------|
| Global agent   | DDPG          | 1.636 (0.956,2.317)           | -81.721%         |
|                | PPO           | 0.525 (-1.049,2.101)          | -94.134%         |
|                | SAC           | 0.586 (-1.315,2.486)          | -93.453%         |
| CTDE           | MADDPG        | 0.322 (-0.327,0.971)          | -96.402%         |
|                | MAPPO         | 8.950 (6.190,11.711)          | Best baseline    |
| Ablation study | MF            | 1.731 (-2.232,5.694)          | -80.659%         |
| <b>Ours</b>    | <b>GAT-MF</b> | <b>33.288 (18.657,49.919)</b> | <b>+271.933%</b> |

**Table 2: Performance evaluation of the real-world metropolitan task**

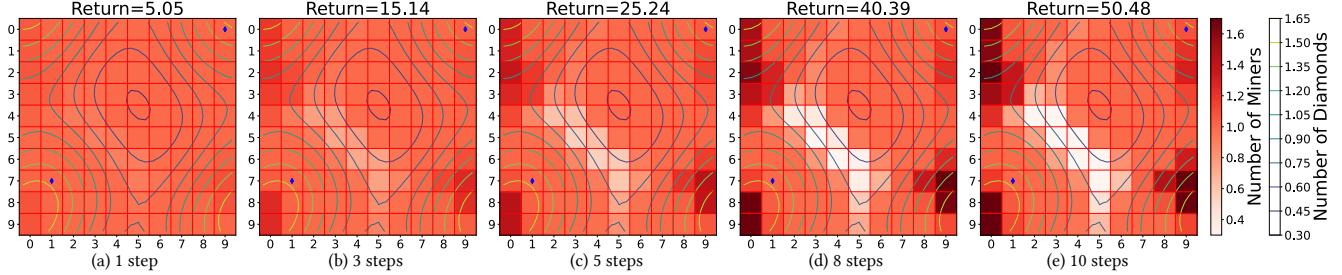
| Group          | Method        | Atlanta             |                 | Miami               |                 |
|----------------|---------------|---------------------|-----------------|---------------------|-----------------|
|                |               | Best Episode Return | Comparison      | Best Episode Return | Comparison      |
| Global agent   | DDPG          | -9127.401           | -1.261%         | -15518.533          | -8.012%         |
|                | PPO           | -9185.278           | -1.903%         | -15659.398          | -8.992%         |
|                | SAC           | -9293.688           | -3.106%         | -15681.702          | -9.148%         |
| CTDE           | MADDPG        | -9023.012           | -0.103%         | -14367.416          | Best baseline   |
|                | MAPPO         | -9013.738           | Best baseline   | -15409.643          | -7.254%         |
| Ablation study | MF            | -6572.228           | +27.087%        | -13439.023          | +6.462%         |
| <b>Ours</b>    | <b>GAT-MF</b> | <b>-5162.126</b>    | <b>+42.730%</b> | <b>-11080.564</b>   | <b>+22.877%</b> |

**Table 3: Computational efficiency in the real-world metropolitan task**

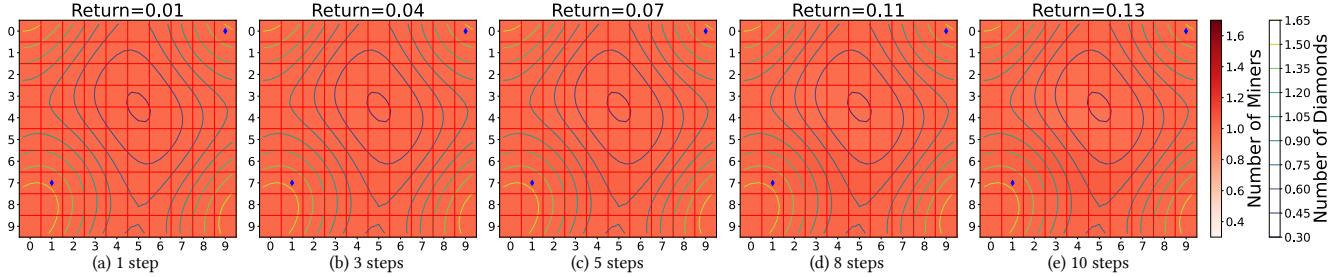
| Method      | Atlanta       |                 |                |                 | Miami         |                 |                |                 |
|-------------|---------------|-----------------|----------------|-----------------|---------------|-----------------|----------------|-----------------|
|             | Training Time | Comparison      | GPU Memory     | Comparison      | Training Time | Comparison      | GPU Memory     | Comparison      |
| MADDPG      | 94041s        | +28.457%        | 53553MB        | +94.423%        | 141727s       | +34.686%        | 68519MB        | +94.905%        |
| MAPPO       | 73208s        | Best baseline   | 27687MB        | Best baseline   | 105228s       | Best baseline   | 35155MB        | Best baseline   |
| <b>Ours</b> | <b>9980s</b>  | <b>-86.368%</b> | <b>22379MB</b> | <b>-19.171%</b> | <b>11316s</b> | <b>-89.246%</b> | <b>28289MB</b> | <b>-19.531%</b> |

## 2 LEARNED POLICY VISUALIZATIONS IN THE MANUAL GRID-WORLD TASK

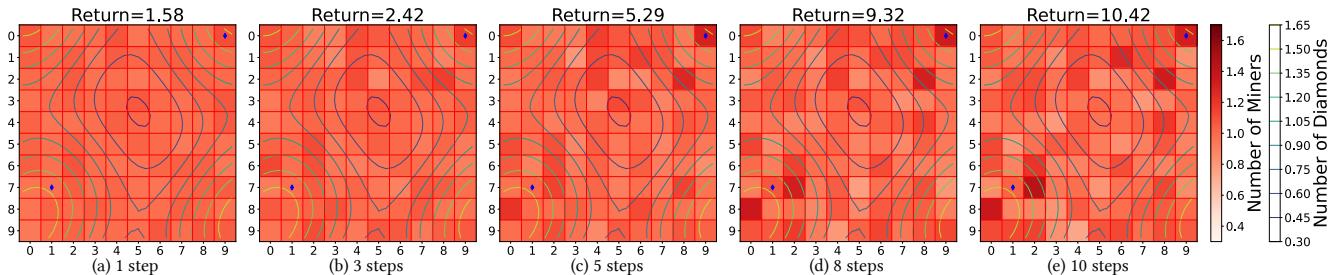
In this section, we provide more visualizations of the learned policy with different methods in the grid-world task. We show the distribution of miners after applying the policy from a trained GAT-MF model, a trained ablation model, a trained MAPPO model, a trained MADDPG model, a trained SAC model, a trained PPO model, and a trained DDPG model for 1, 3, 5, 8, and 10 steps respectively in Figure 1-7.



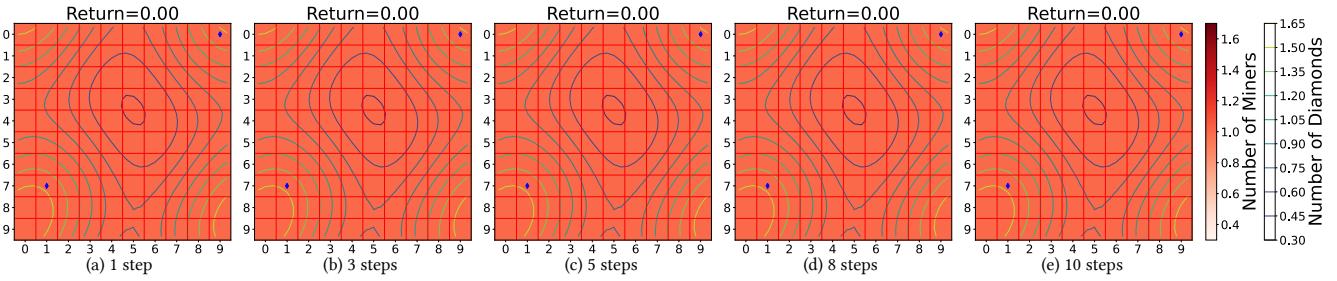
**Figure 1: More visualizations of the learned policies. (a)-(e) The distribution of miners after applying the policy from a trained GAT-MF model for 1, 3, 5, 8, and 10 steps.**



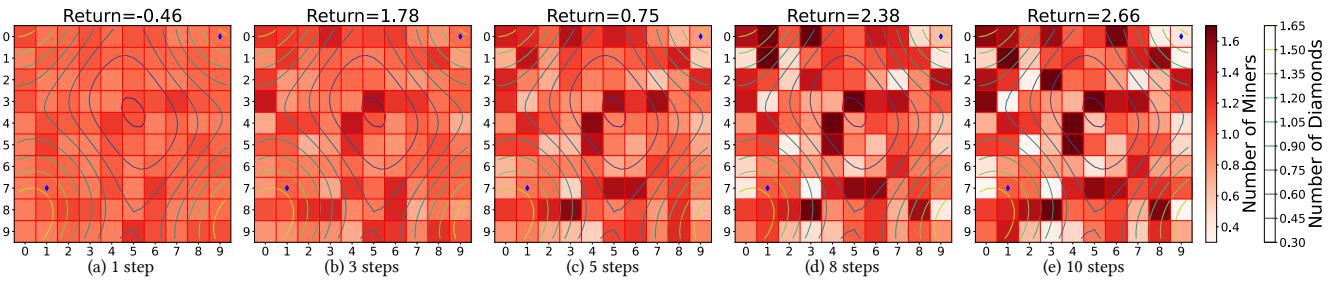
**Figure 2: More visualizations of the learned policies. (a)-(e) The distribution of miners after applying the policy from a trained ablation model for 1, 3, 5, 8, and 10 steps.**



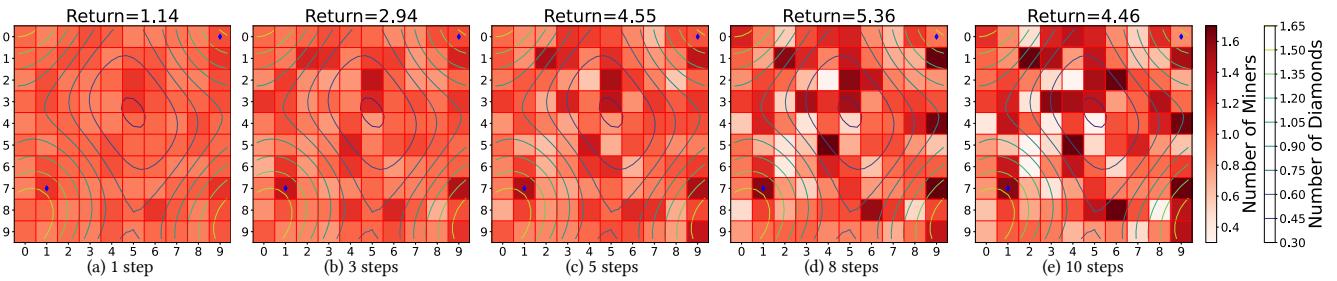
**Figure 3: More visualizations of the learned policies. (a)-(e) The distribution of miners after applying the policy from a trained MAPPO model for 1, 3, 5, 8, and 10 steps.**



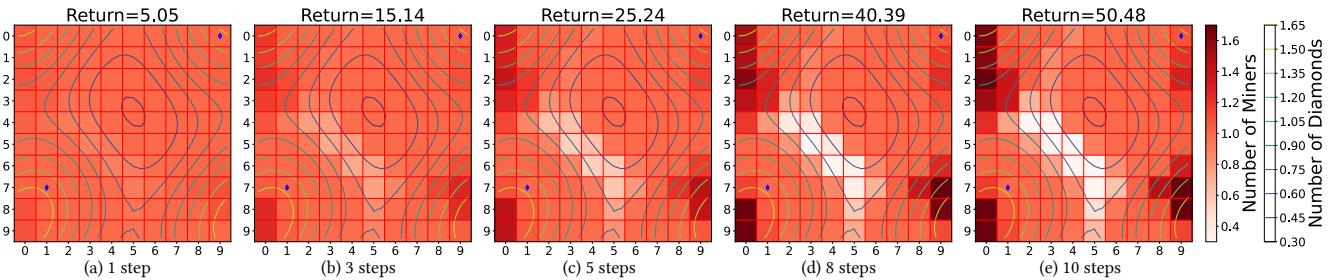
**Figure 4:** More visualizations of the learned policies. (a)-(e) The distribution of miners after applying the policy from a trained MADDPG model for 1, 3, 5, 8, and 10 steps.



**Figure 5:** More visualizations of the learned policies. (a)-(e) The distribution of miners after applying the policy from a trained SAC model for 1, 3, 5, 8, and 10 steps.



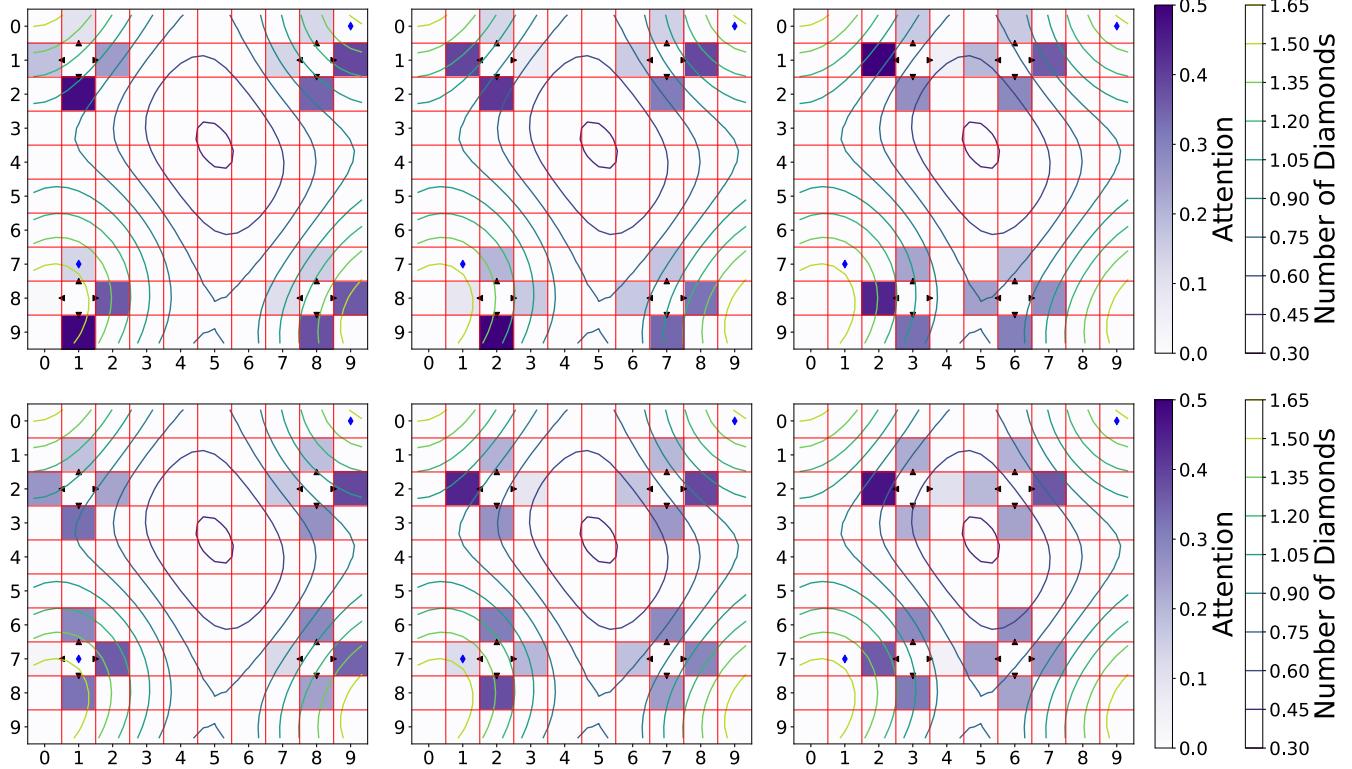
**Figure 6:** More visualizations of the learned policies. (a)-(e) The distribution of miners after applying the policy from a trained PPO model for 1, 3, 5, 8, and 10 steps.



**Figure 7:** More visualizations of the learned policies. (a)-(e) The distribution of miners after applying the policy from a trained DDPG model for 1, 3, 5, 8, and 10 steps.

### 3 LEARNED ATTENTION VISUALIZATIONS IN THE MANUAL GRID-WORLD TASK

In this section, we provide visualizations of the learned attention of some more actors at the first step from a trained GAT-MF model in the grid-world task in Figure 8.



**Figure 8: More visualizations of the learned policies. The learned attentions of some more actors at the first step from a trained GAT-MF model. Since one panel cannot hold the attentions of all the actors, we show the attentions of four actors in each panel.**