Cooperative Exploration for Multi-Agent Deep Reinforcement Learning

Iou-Jen Liu, Unnat Jain, Raymond A. Yeh, Alexander G. Schwing

ILLINOIS

ICML International Conference On Machine Learning

1. Introduction

Problem:

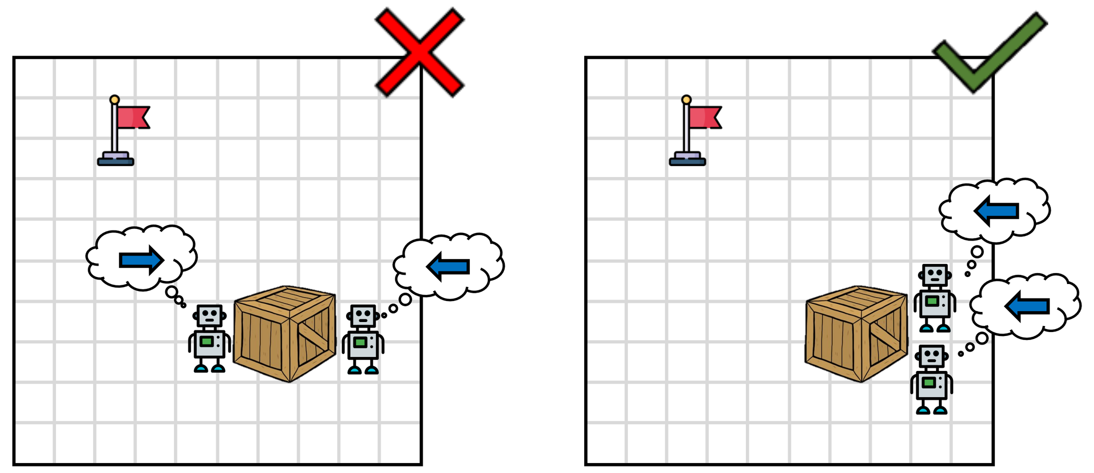
- Multi-agent deep reinforcement learning (MARL)
- Learning policies in sparse-reward settings is challenging
- Efficient exploration strategies are needed

Challenges for Multi-Agent Exploration:

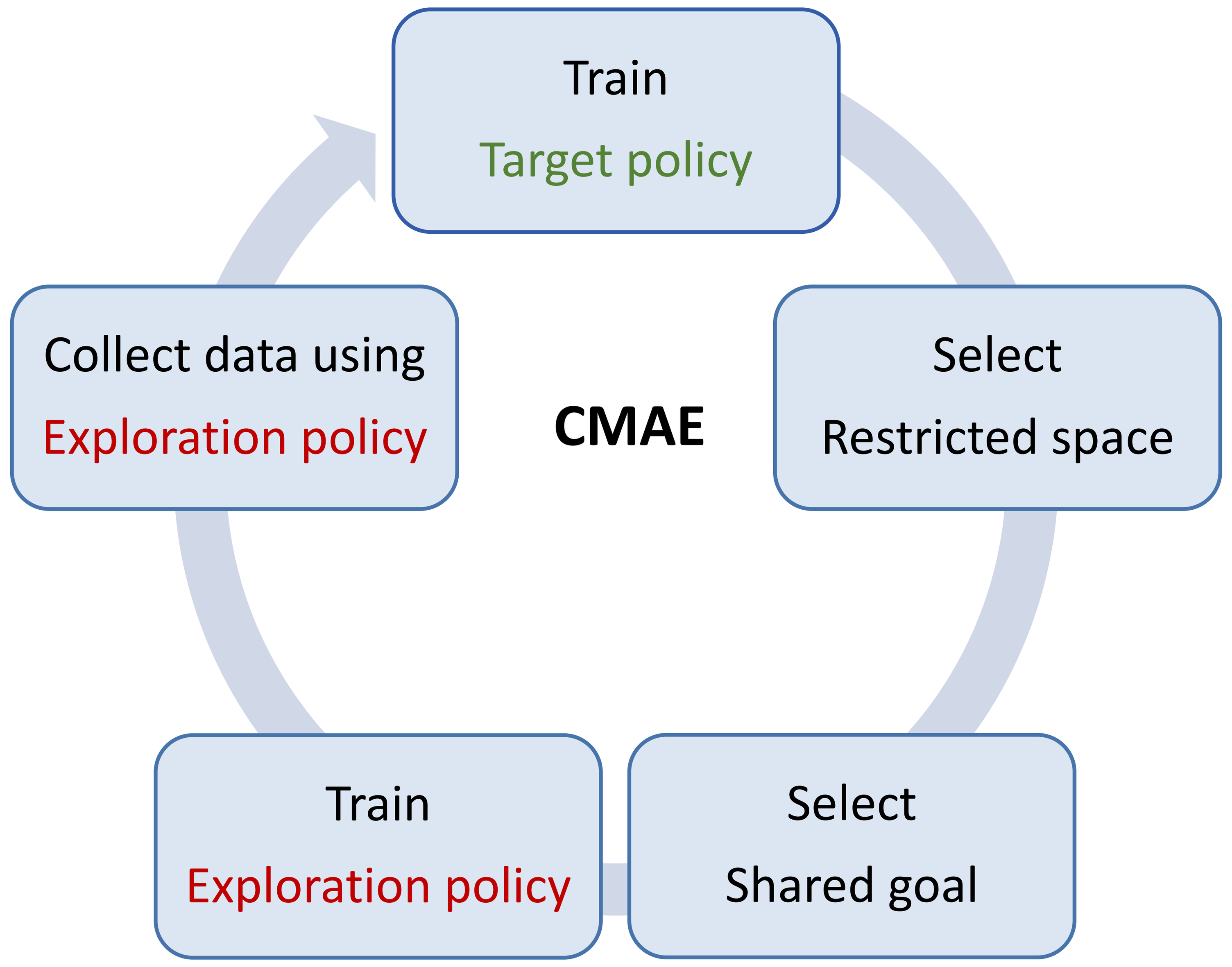
- **Identify states that are worth exploring:**
  - ➢ Number of states grow exponentially with number of agents
  - ➢ Infeasible to explore all states
  - ➢ Example: N-agent push-box task in $L \times L$ grid
    - – $(L^2)^{1+N}$ states to explore
- **Coordinate agents' exploration efforts:**
  - ➢ Uncoordinated exploration is inefficient
  - ➢ Example: two-agent push-box task in $L \times L$ grid
    - – Agents need to push the box toward same direction to move the box

Ours: Cooperative Multi-Agent Exploration (CMAE) :

- **Restricted space exploration**:
  - ➢ Identify under-explored low-dimensional restricted space
  - ➢ Avoid exploring the exponentially-growing full state space
- **Shared goal exploration** :
  - ➢ Agents share a common goal while exploring
  - ➢ Enable coordinated multi-agent exploration



2. Overview



Approach:

Decouple exploration and target policy

- **Train target policy:** Maximize the external environment reward using off-policy RL algorithms
- **Collect data using exploration policy:** Data contains under-explored states
- **Train exploration policy:** Positive reward when reaching a shared goal
- **Select shared goal:** Select a rarely visited state as shared goal
- **Select restricted space:** Select low-dimensional and under explored subspace as restricted space

Please see paper for details

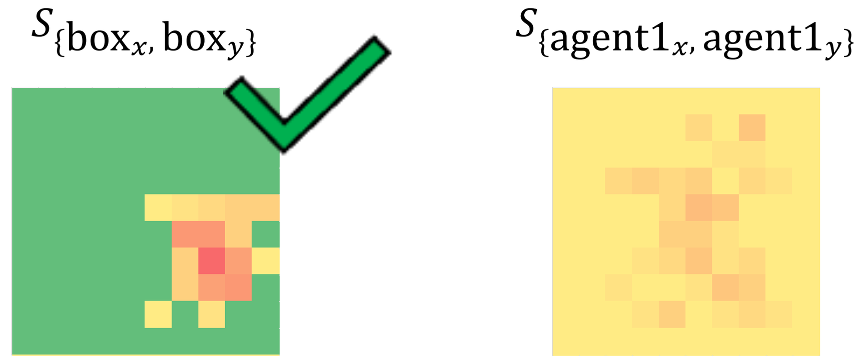
3. Restricted Space

Why Restricted Space?

- **Observation:** Reward function typically depends on a low-dimensional subspace of the state space
- **Example:** N-agent push-box task in $L \times L$ grid.
  - ➢ Size of state space: $(L^2)^{1+N}$
  - ➢ Reward function depends only on the box location L^2

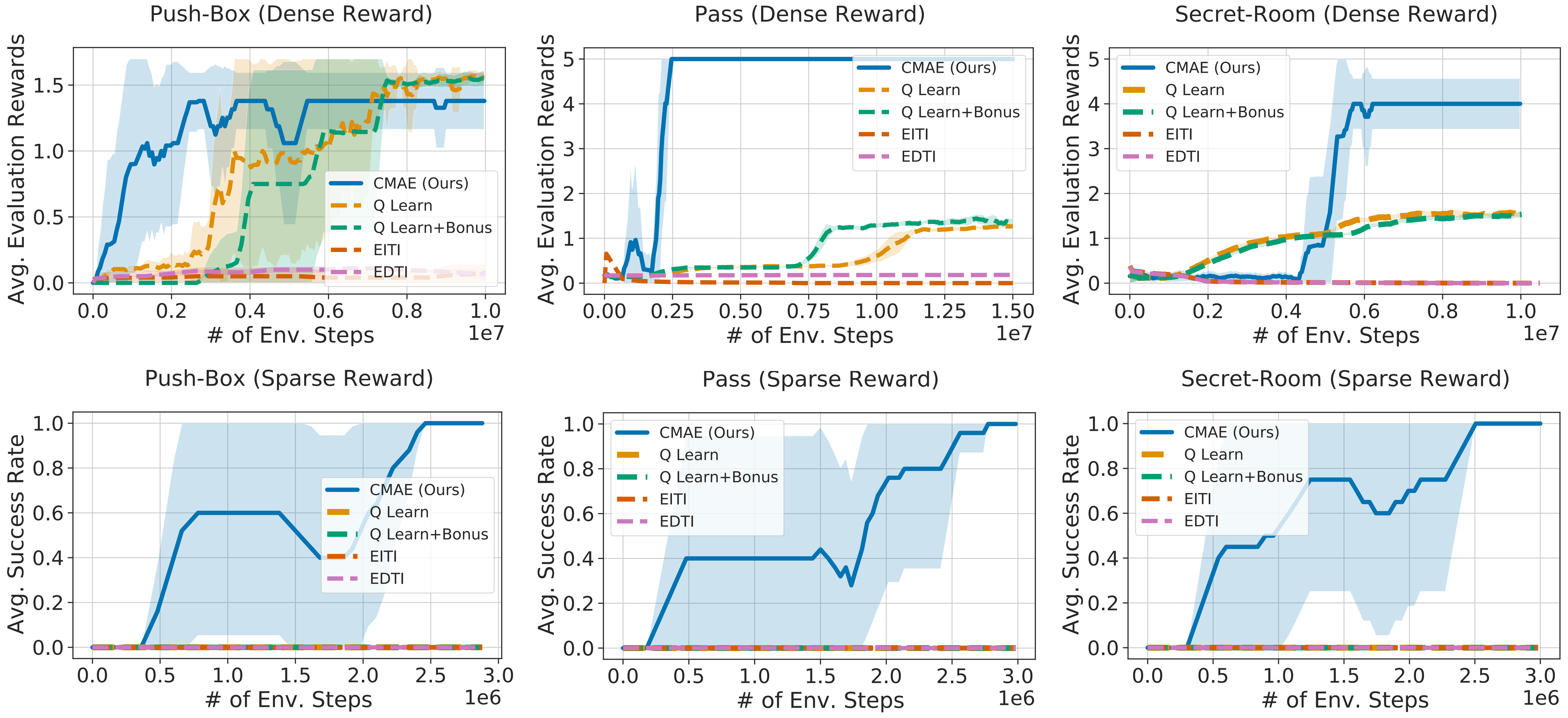
How to find under-explored restricted space?

- Each restricted space S_k has a counter
- Under-explored restricted space has smaller entropy

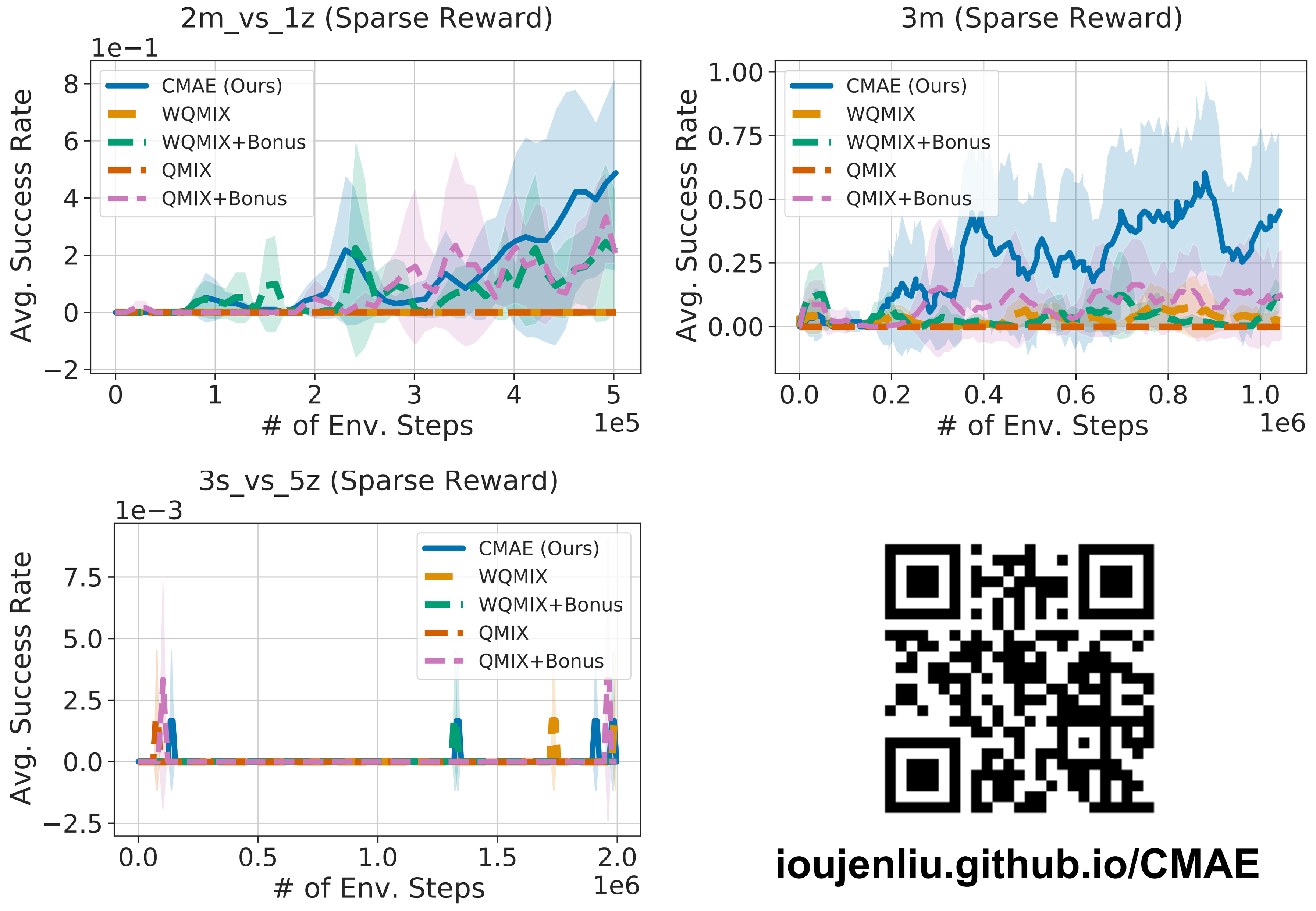


4. Experimental Results

Results on multi-agent grid world



Results on StarCraft II Multi-Agent Challenge (SMAC)



ioujenliu.github.io/CMAE