Tencent AI Lab

Grid-Wise Control for Multi-Agent Reinforcement Learning in Video Game AI

Lei Han*[1], Peng Sun*[1], Yali Du*[2], Jiechao Xiong[1], Qing Wang[1], Xinghai Sun[1], Han Liu[3], Tong Zhang[4]

1 Tencent AI Lab, Shenzhen, China
2 University of Technology Sydney, Australia
3 Northwestern University, IL, USA
4 Hong Kong University of Science and Technology, Hong Kong, China
* Equal contribution

Email: leihan.cs@gmail.com

Introduction

Tencent AI Lab

❑Considered Problem

- Multi-agent reinforcement learning (MARL)
- Grid-world environment (video game)
- Challenge
  - ➢ flexibly control an **arbitrary number** of agents
  - ➢ while achieving **effective collaboration**

❑Existing MARL Approaches

- Decentralized learning
  - ➢ IQL, IAC (Tan, 1993; Foerster et al., 2017)
- Centralized learning
  - ➢ CommNet, BicNet (Sukhbaatar et al., 2016; Peng et al., 2017)
- Mixture
  - ➢ COMA, QMIX, Mean-Field (Foerster et al., 2017; Rashid et al., 2018; Yang et al., 2018)

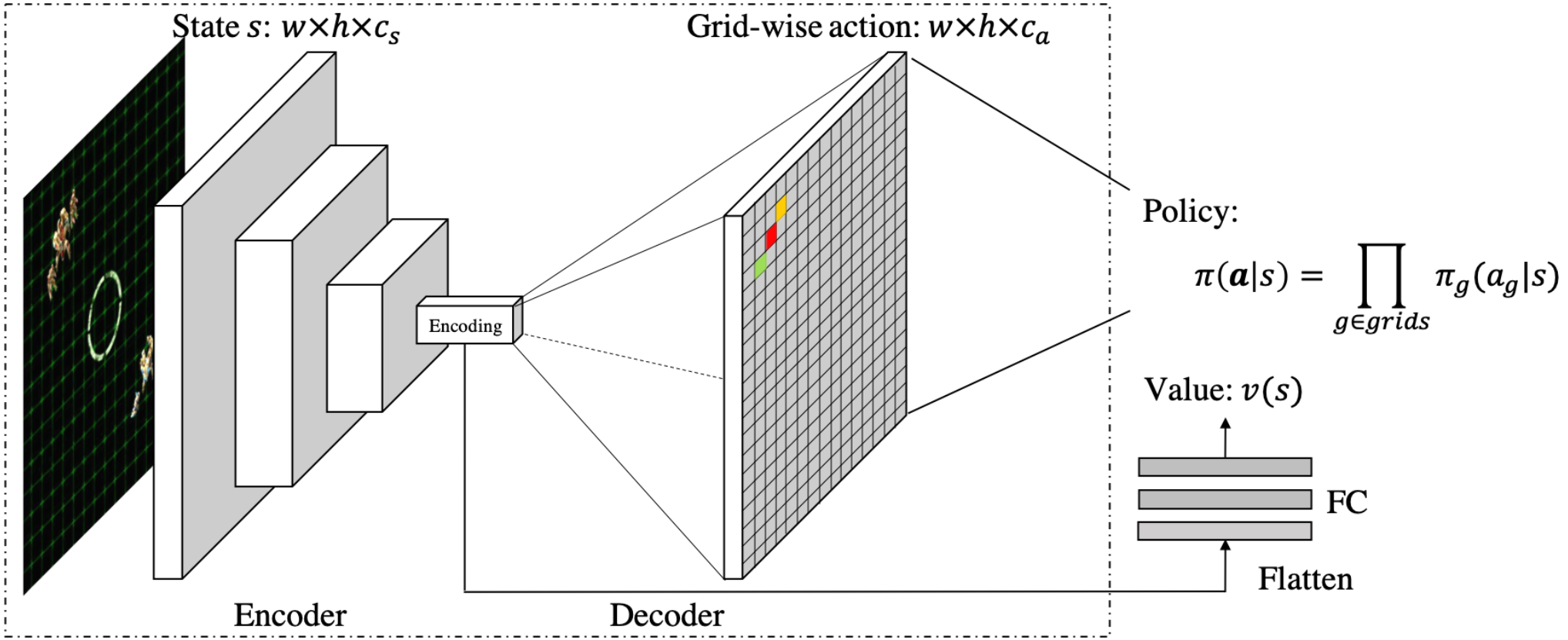
❖**Unable/instable to deal with variant agent number**

GridNet

Tencent AI Lab

❑Architecture

- **Encoder**
  - ➢ Inputs are represented as an **image-like** structure
  - ➢ Using conv/pooling layers to generate an **embedding**
- **Decoder**
  - ➢Up-sampling to construct an **action map**
  - ➢An agent will **take the action in the grid it occupies**



GridNet

Tencent AI Lab

❑Algorithms

- Can be integrated with many **general** RL algorithms
  - ➢ Q-learning
  - ➢ Actor-critic

❑Properties

- **Collaboration is natural**
  - ➢ Stacked convolutional and/or pooling layers provide a large receptive field
  - ➢ Each agent is aware of other agents in its neighborhood
- **Fast parallel exploration**
  - ➢ Convolutional parameters are shared by all the agents
  - ➢ Once an agent takes a beneficial action during its own exploration, the other agents will acquire the knowledge as well
- **Transferrable policy**
  - ➢ The trained policy is easy to be transferred to other settings with a various number of agents

Experiments on Battle Games in StarCraft II

Tencent AI Lab

❑Scenarios

- 5Immortals vs. 5Immortals (**5I**)
- 3Immortals+2Zealots vs. 3Immortals+2Zealots (**3I2Z**)
- mixed army battle (**MAB**) with a random number of various Zerg units
  - including Baneling, Zergling, Roach, Hydralisk and Mutalisk.

❑Training Strategies

- Against handcraft policies: **random (Rand)**, **attack-nearest (AN)**, **hit-and-run (HR)**
- Against self historic versions: **self-play (SP)**

❑Compared Methods

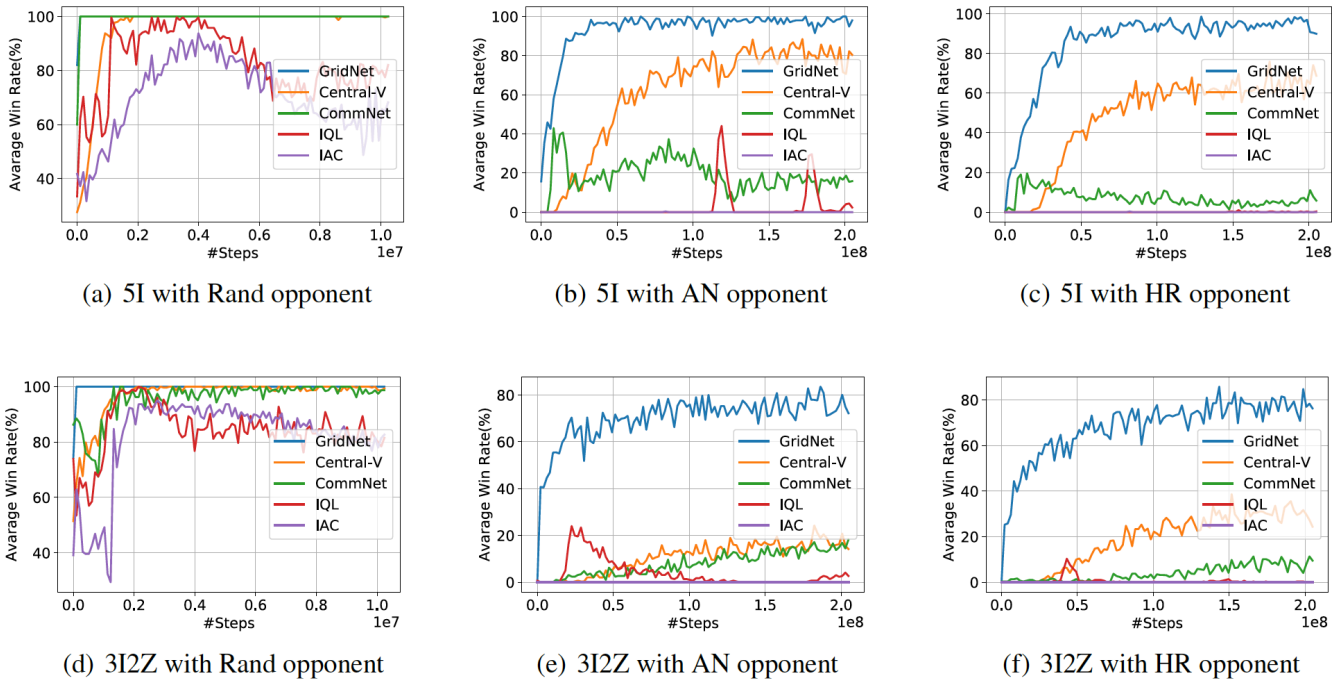
- **IQL**: independent Q-learning [Tan, 1993]
- **IAC**: independent actor-critic [Foerster et al., 2017]
- **Central-V**: centralized value with decentralized policy [Foerster et al., 2017]
- **CommNet**: communication net [Sukhbaatar et al., 2016]

❑Video link: https://youtu.be/LTcr01iTgZA

Experiments on Battle Games in StarCraft II

Tencent AI Lab

- **On 5I and 3I2Z**
  - Performance (against handcraft policies)



(a) 5I with Rand opponent (b) 5I with AN opponent (c) 5I with HR opponent

(d) 3I2Z with Rand opponent (e) 3I2Z with AN opponent (f) 3I2Z with HR opponent

Figure 3. Average test wining rate vs. training steps (i.e., number of passed data points) on 5I and 3I2Z when training against Rand, AN and HR. We omit the curves on SP, because using self-play, the winning rate will always be a value slightly above 50%.

  - Performance (against each other)

Table 1. Cross combat winning rate over the selected strongest policies of the 5 methods on 5I. The winning rates are for "row against column". Any pair of symmetric values sum to 1.

	IAC	IQL	Central-V	CommNet	GridNet
IAC	–	0.43	0.14	0.05	**0.00**
IQL	0.57	–	0.39	0.08	**0.06**
Central-V	0.86	0.61	–	0.52	**0.27**
CommNet	0.95	0.92	0.48	–	**0.01**
GridNet	**1.00**	**0.94**	**0.73**	**0.99**	–

Table 2. Cross combat winning rate over the selected strongest policies of the 5 methods on 3I2Z. The table format follows that used in Table 1.

	IAC	IQL	Central-V	CommNet	GridNet
IAC	–	0.60	0.24	0.20	**0.00**
IQL	0.40	–	0.17	0.24	**0.04**
Central-V	0.76	0.83	–	0.29	**0.12**
CommNet	0.80	0.76	0.71	–	**0.32**
GridNet	**1.00**	**0.96**	**0.88**	**0.68**	–

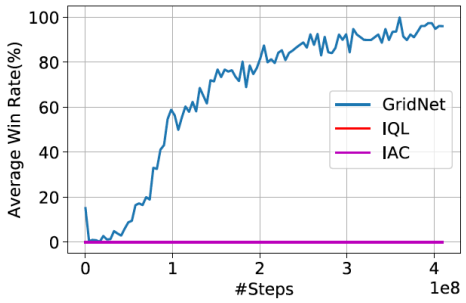
Experiments on Battle Games in StarCraft II

Tencent AI Lab

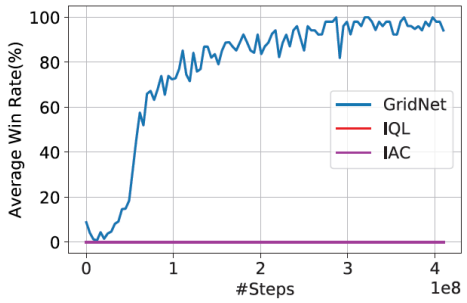
- **Transferability On 5I and 3I2Z**
  - **Directly apply** the trained policy **to maps with more agents**
  - **10I**, **20I**, **5I5Z**, **10I10Z**

Scenario	10I		20I		5I5Z		10I10Z	
Opponent	AN	HR	AN	HR	AN	HR	AN	HR
IAC	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00
IQL	0.03	0.00	0.01	0.00	0.02	0.00	0.01	0.00
Central-V	0.78	0.76	0.55	0.49	0.02	0.10	0.02	0.05
GridNet	**1.00**	**1.00**	**1.00**	**1.00**	**0.70**	**0.87**	**0.79**	**0.71**

- **Performance On MAB**
  - CommNet and Central-V cannot be applied

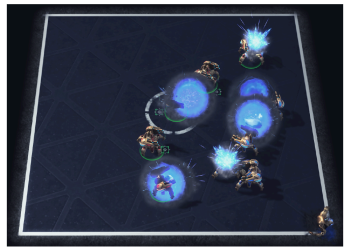


(a) MAB against AN

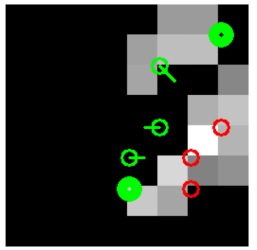


(b) MAB against HR

❑ Learned Tactics

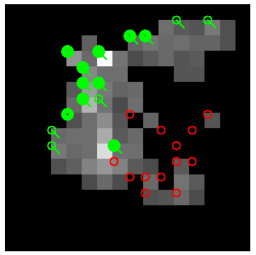


(a) Hit-and-run



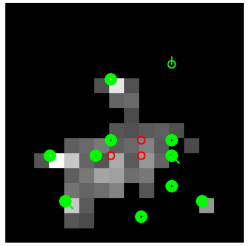
(b) Saliency map

(c) Scatter



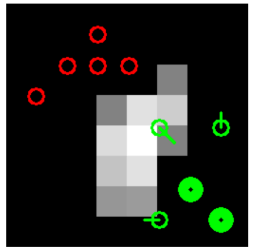
(d) Saliency map

(e) Surround



(f) Saliency map

(g) Retreat



(h) Saliency map

Tencent AI Lab

Thanks!

Poster at Pacific Ballroom #243

Jun 11th, 6:30 pm