ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

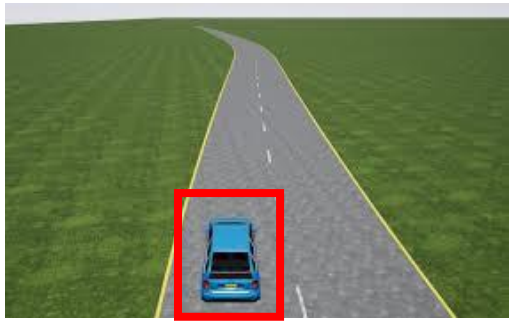
AVD-Net: Attention Value Decomposition Network For Deep Multi-Agent Reinforcement Learning

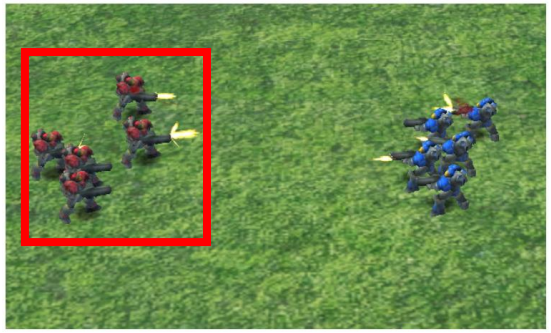
Yuanxin Zhang[1], Huimin Ma[2*], Yu Wang[1]

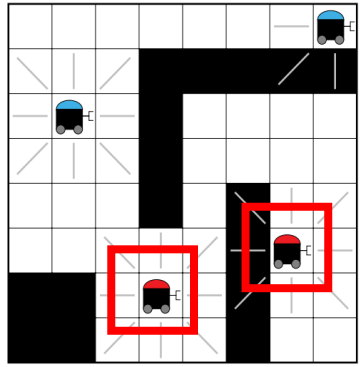
[1]Tsinghua University

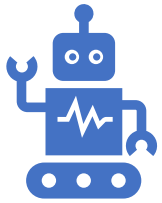
[2]University of Science and Technology Beijing

Introduction









Single Agent → **Multi-Agent**

Challenge in Multi-Agent RL

- The Curse of Dimensionality
- The Nonstationarity of Environment
- The Exploration-Exploitation Trade-off
- The Multi-Agent Goal Setting
- Agents' Coordination

Contribution

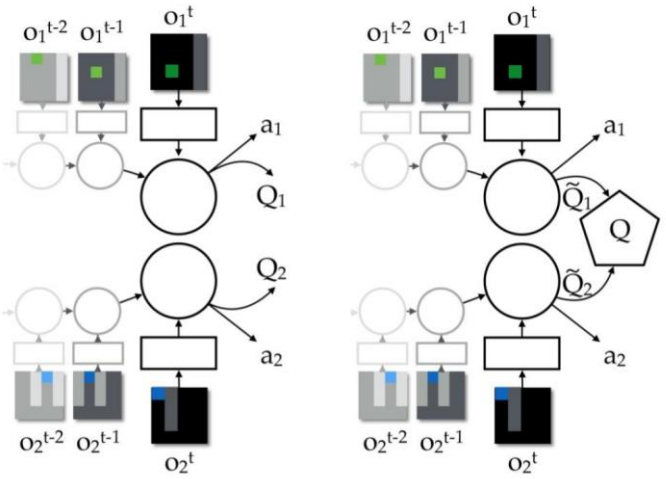
- We propose a value-based architecture which factorizes the joint value function with only the partial observations and actions of local agents.
- We adopt attention mechanism to learn the correlations between agents and compute the decomposition weight of each agent’s action-value function.
- Our proposal effectively exploits the information in multiagent system and achieves state-of-the-art performance in different cooperative MARL environments.

Related Work

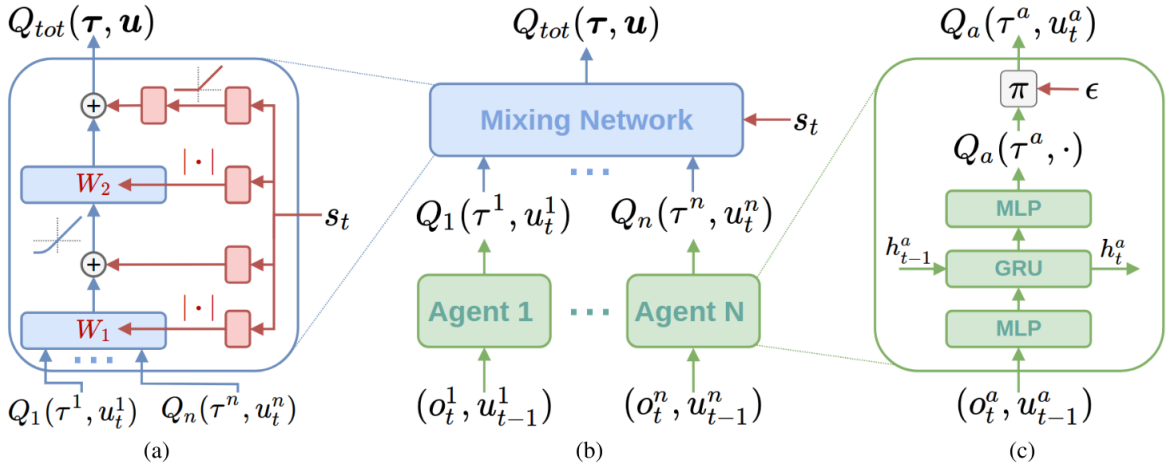
VDN

Partially observed $s_t \rightarrow h_t = a_1 o_1 r_1, \ldots, a_{t-1} o_{t-1} r_{t-1}$

$$Q((h^1, h^2, \ldots, h^t), a(a^1, a^2, \ldots, a^t) \approx \sum_{i=1}^{d} \tilde{Q}_i \left(h^i, a^i\right)$$



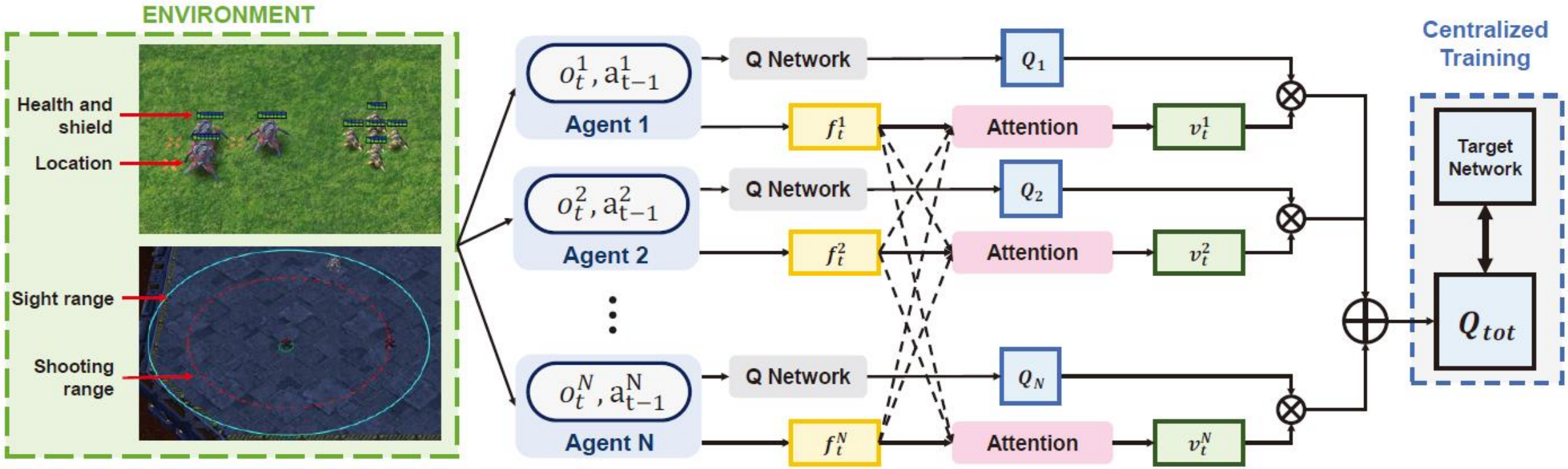
QMIX



[1] Sunehag, Peter, et al. "Value-Decomposition Networks For Cooperative Multi-Agent Learning Based On Team Reward." AAMAS. 2018.
[2] Rashid, Tabish, et al. "QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning." arXiv preprint arXiv:1803.11485 (2018).

Our Work

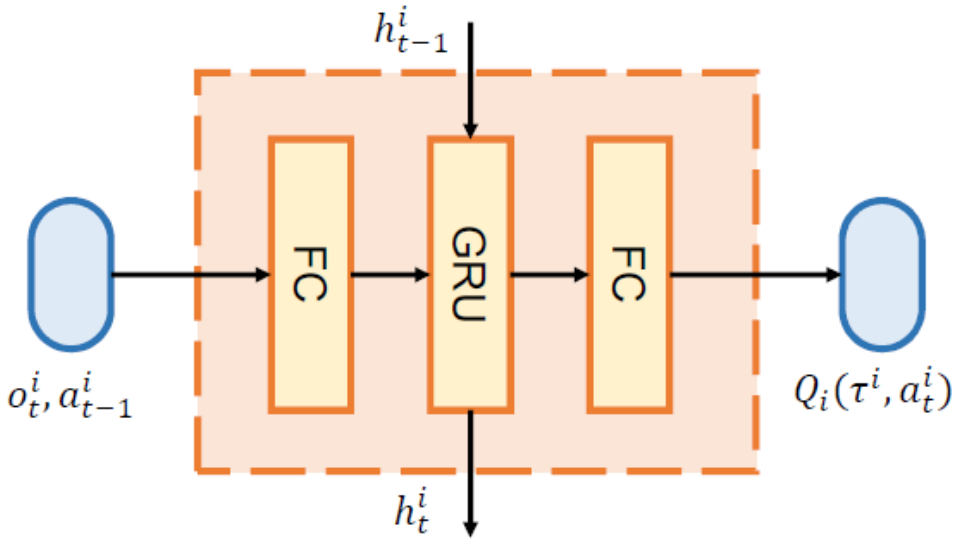
Overall structure of AVD-Net



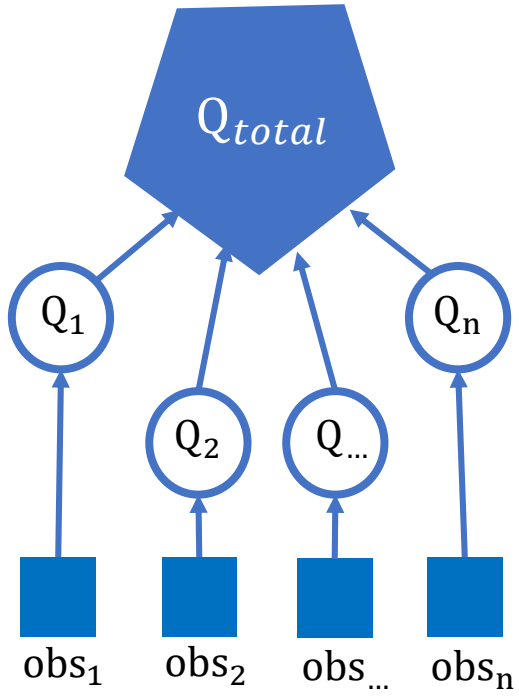
Our Work

Value Decomposition

- For each agent, there is an action-value network, which adopts the DRQNs structure and receives the current agent observation o_t^i and the last action a_{t-1}^i as input at each time step.



$$Q_{tot}(\boldsymbol{\tau}, \boldsymbol{a}) = \sum_{i=1}^{N} v_t^i Q_i(\tau_t^i, a_t^i), v_t^i = fc([fc(f_t^i), z_t^i])$$



Our Work

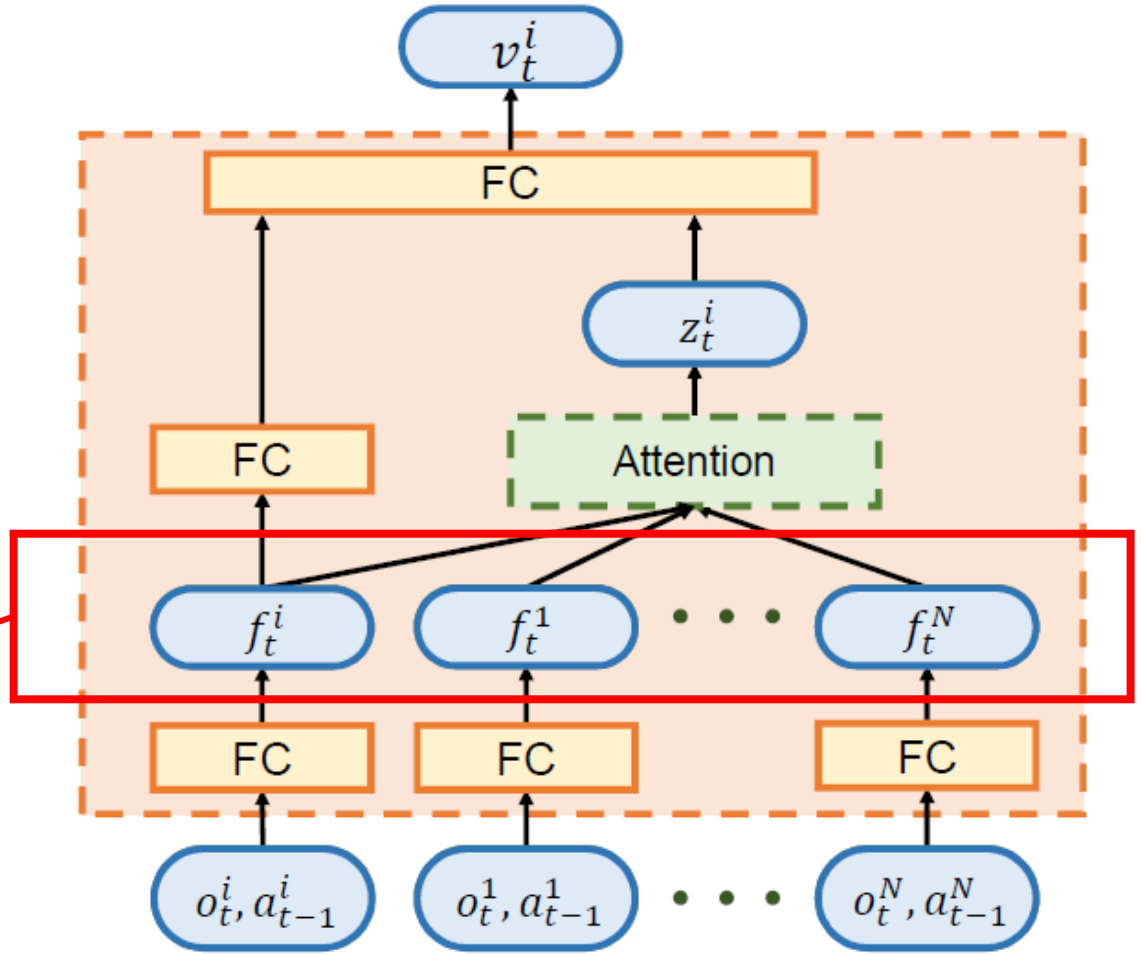
Attention mechanism

- First define an agent attention encoder which takes in the observation o_t^i and the last action a_{t-1}^i from agent i. The output f_t^i represents the local agent attention embedding.

$$Q_{tot}(\boldsymbol{\tau}, \boldsymbol{a}) = \sum_{i=1}^{N} v_t^i Q_i(\tau_t^i, a_t^i), v_t^i = fc([fc(f_t^i), z_t^i])$$

$$z_t^i = \sum\nolimits_{j \neq i} \alpha_{i,j} f_t^j (o_t^j, a_{t-1}^j)$$

$$\alpha_{i,j} = \frac{exp(\beta_{i,j})}{\sum_{j \neq i} exp(\beta_{i,j})}, \beta_{i,j} = f_t^i W_a f_t^{jT}$$



Our Work

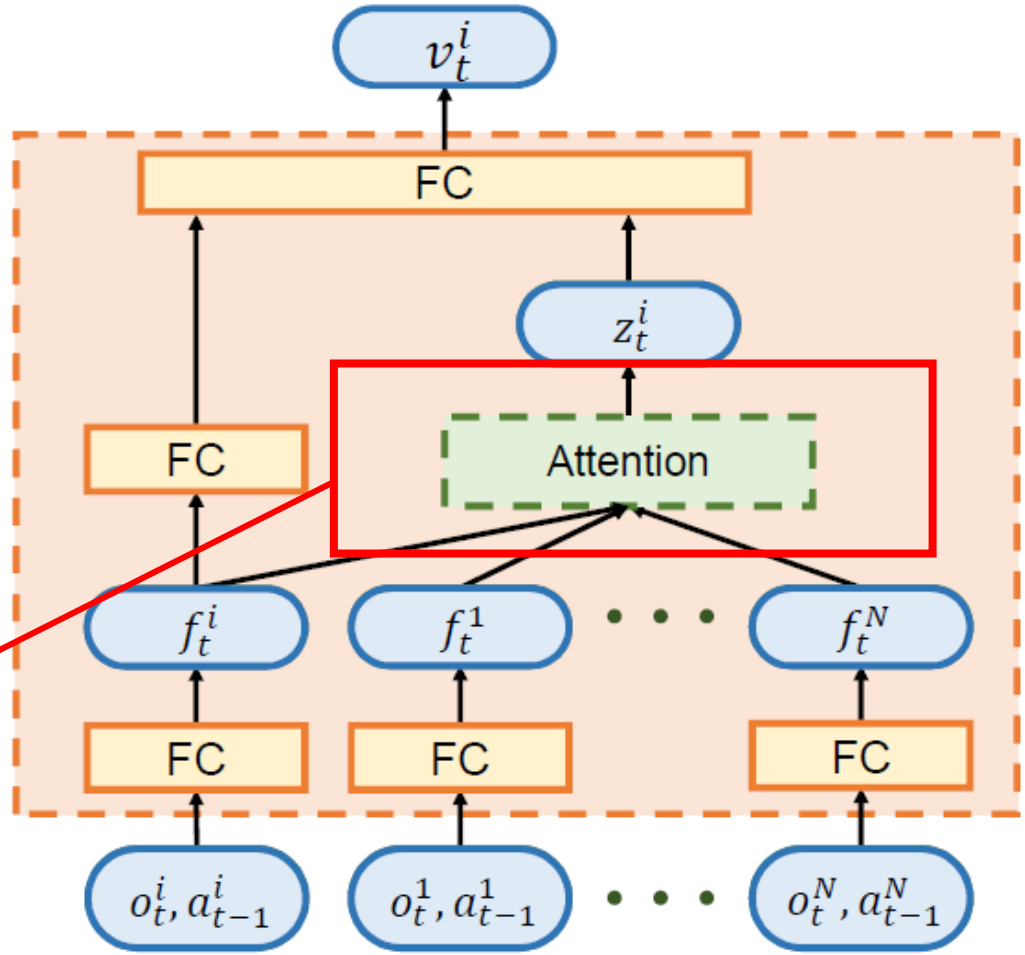
Attention mechanism

- Calculate the similarity between the attention embeddings of agent i and every other agent.
- Calculate the attention score $\alpha_{i,j}$

$$Q_{tot}(\boldsymbol{\tau}, \boldsymbol{a}) = \sum_{i=1}^{N} v_t^i Q_i(\tau_t^i, a_t^i), v_t^i = fc([fc(f_t^i), z_t^i])$$

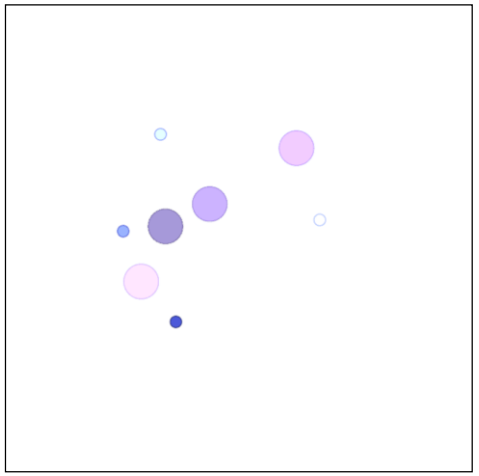
$$z_t^i = \sum\nolimits_{j \neq i} \alpha_{i,j} f_t^j(o_t^j, a_{t-1}^j)$$

$$\alpha_{i,j} = \frac{exp(\beta_{i,j})}{\sum_{j \neq i} exp(\beta_{i,j})}, \beta_{i,j} = f_t^i W_a f_t^{jT}$$

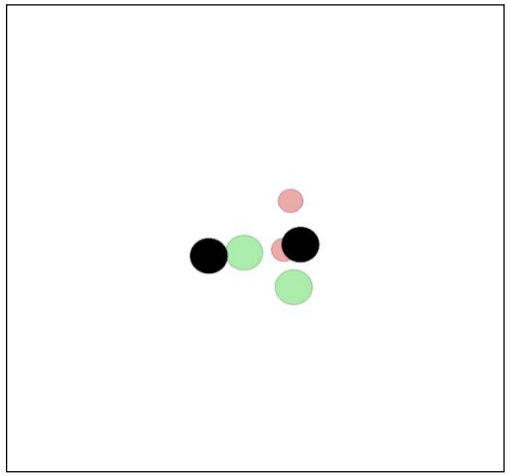


Experiment

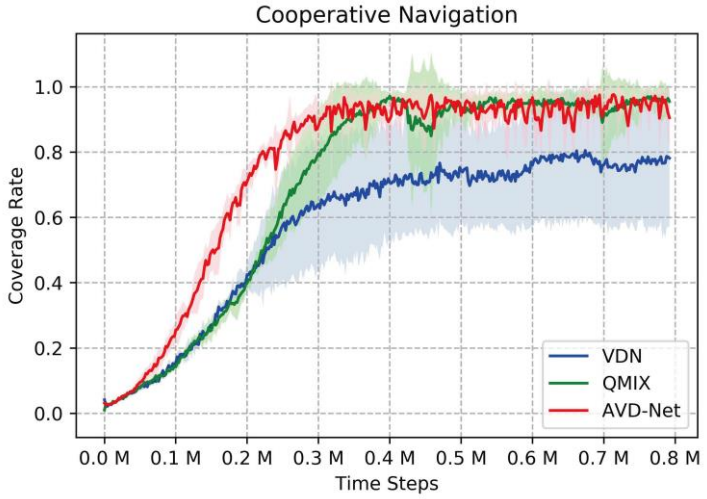
Multi-Agent Particle Environment

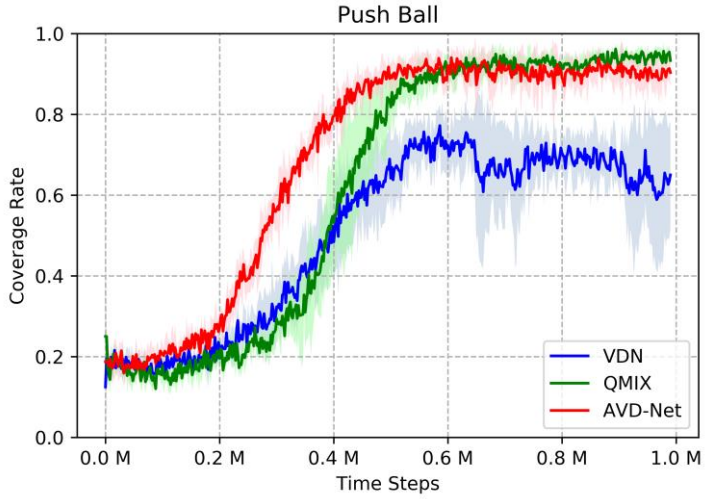


Cooperative Navigation
N agents need to cover
N landmarks



Push Ball
N agents need to push
K balls into K landmarks



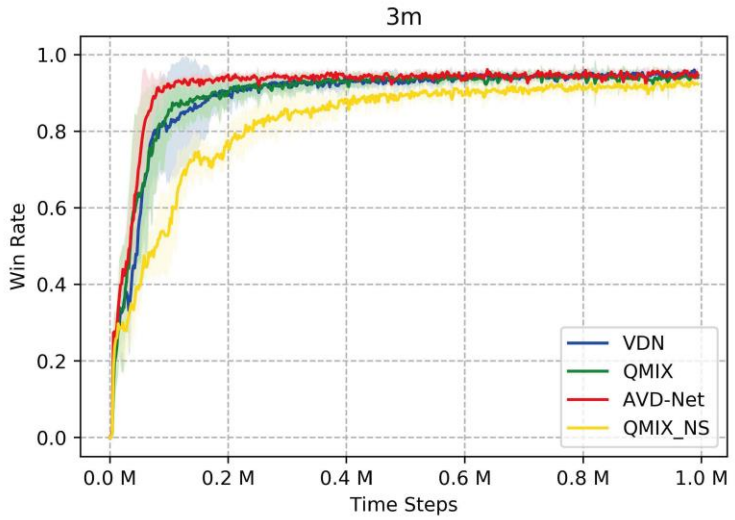


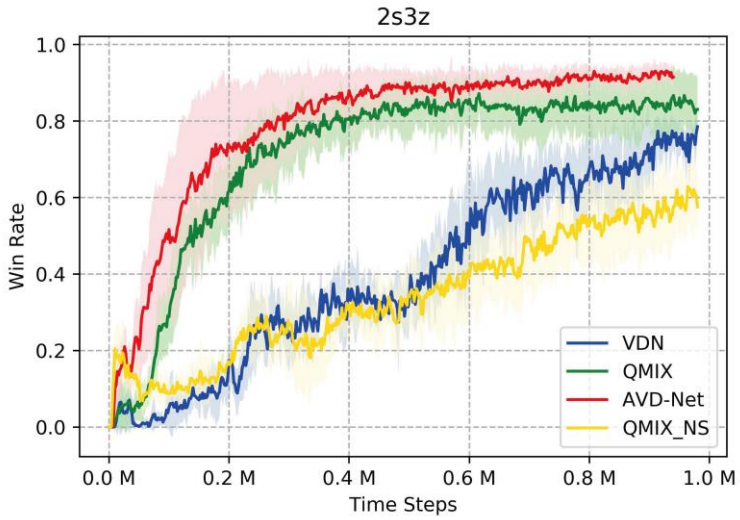
https://github.com/openai/multiagent-particle-envs

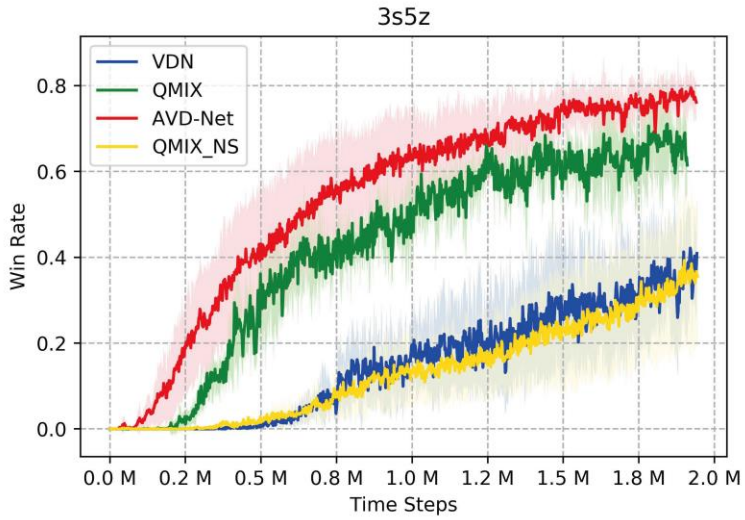
Experiment

StarCraft Multi-Agent Challenge

Each unit be controlled by an independent RL agent that conditions only on local observations restricted to a limited field of view centered on that unit.







https://github.com/oxwhirl/smac

Thanks for Listening!