Use Value Iteration Network to Play Games

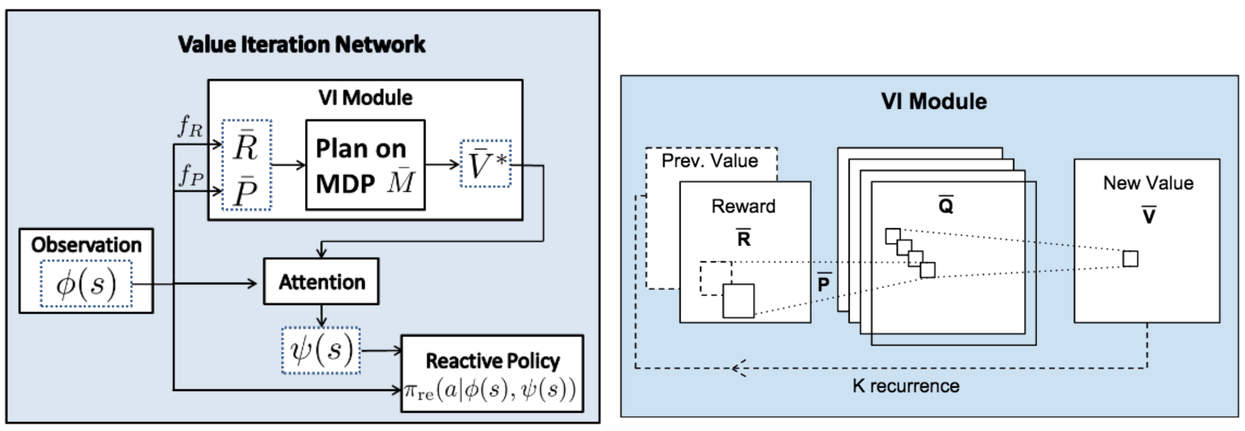
Machine Intelligence and Understanding Laboratory

Yi Ting Yeh and Yun-Nung (Vivian) Chen

Overview

- Goal: add the **Value Iteration Module (VI module)** to improve **Deep Q-Network (DQN)** to play games
- DQN learns to estimate the expected future reward of the current state
- The VI module provides additional information for better DQN estimation

*Value Iteration Network (*Tamar et al., *NIPS 2016)*



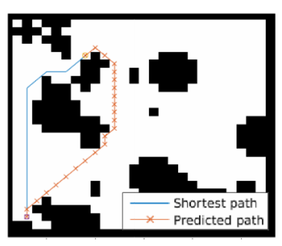
$$Q_n(s,a) = R(s,a) + \gamma \sum_{s'} P(s'|s,a) V_n(s'), \quad V_{n+1}(s) = \max_a Q_n(s,a)$$

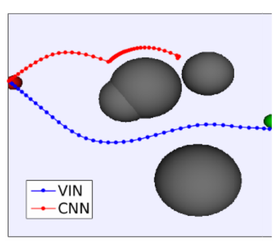
$$\bar{Q}_{\bar{a},i',j'} = \sum_{l,i,j} W^{\bar{a}}_{l,i,j} \bar{R}_{l,i'-i,j'-j}$$

$$\pi^*(s) = \arg\max_a Q_\infty(s,a)$$

a: action
s: state
R: reward

- The value iteration learns **how to plan**, viewed as the CNN.
- The attention module chooses the useful information for the current observation
- With planning, VIN can generalize well shown in experiments on the Grid world and continuous control



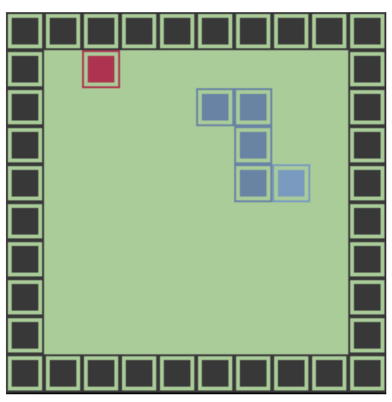


➢VIN learns the planning computation, not reactive policy.

Experimental Setup

Environment for RL: Snake

- The snake wants to eat fruits as many as possible and cannot run into itself
- Players can control the trail
- *The fruit and obstacle (snake) is changing over time*



➢This project investigates whether the VIN can help the snake move with learned **planning** strategy on the **moving-obstacle** environment

Attention Module

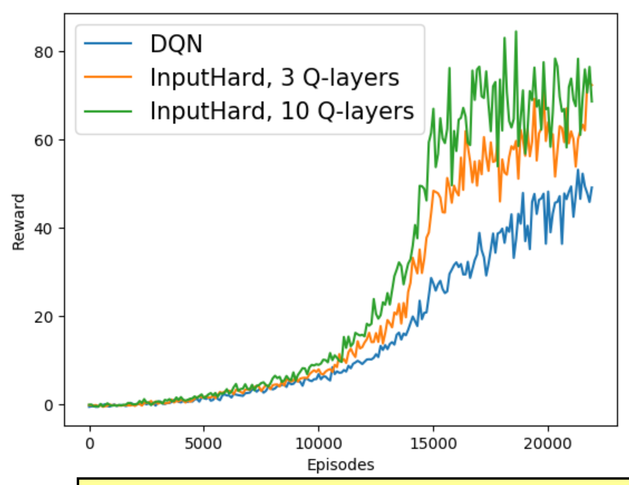
- Attention Mechanism
  - ✓ **InputConv**: convolution layer with a 3x3 filter on the input image
  - ✓ **InputVIConv**: convolution layer with a 3x3 filter on the stacked input image and the values from the VI module
  - ✓ **InputDense**: dense layer fed by the input image
  - ✓ **InputHard**: hard attention on the position of the snake head

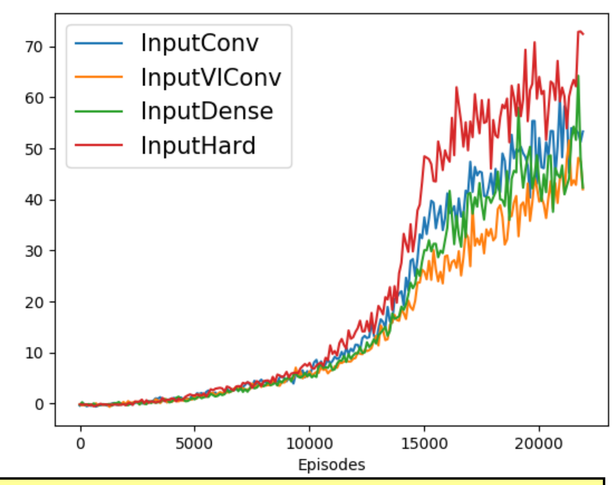
Experimental Results

- **Average reward after training 30000 episodes**

Model	Attention	Reward
DQN		15.8
Proposed: DQN w/ VIN	InputConv	15.8
	InputVIConv	14.4
	InputDense	15.3
	InputHard	**17.7**

- **Learning curves**

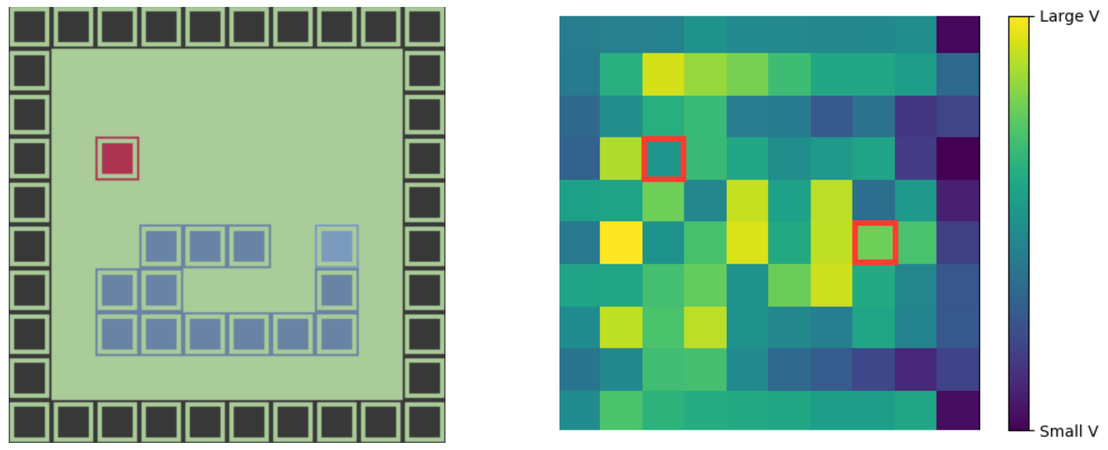




➢The proposed model (InputHard) outperforms the baseline DQN, and the larger Q-layer performs best

➢The hard attention on snake's head performs the best

- **Visualization of the VI module**



➢The VI module successfully learns to not only **eat fruits** but also **twist the trail**