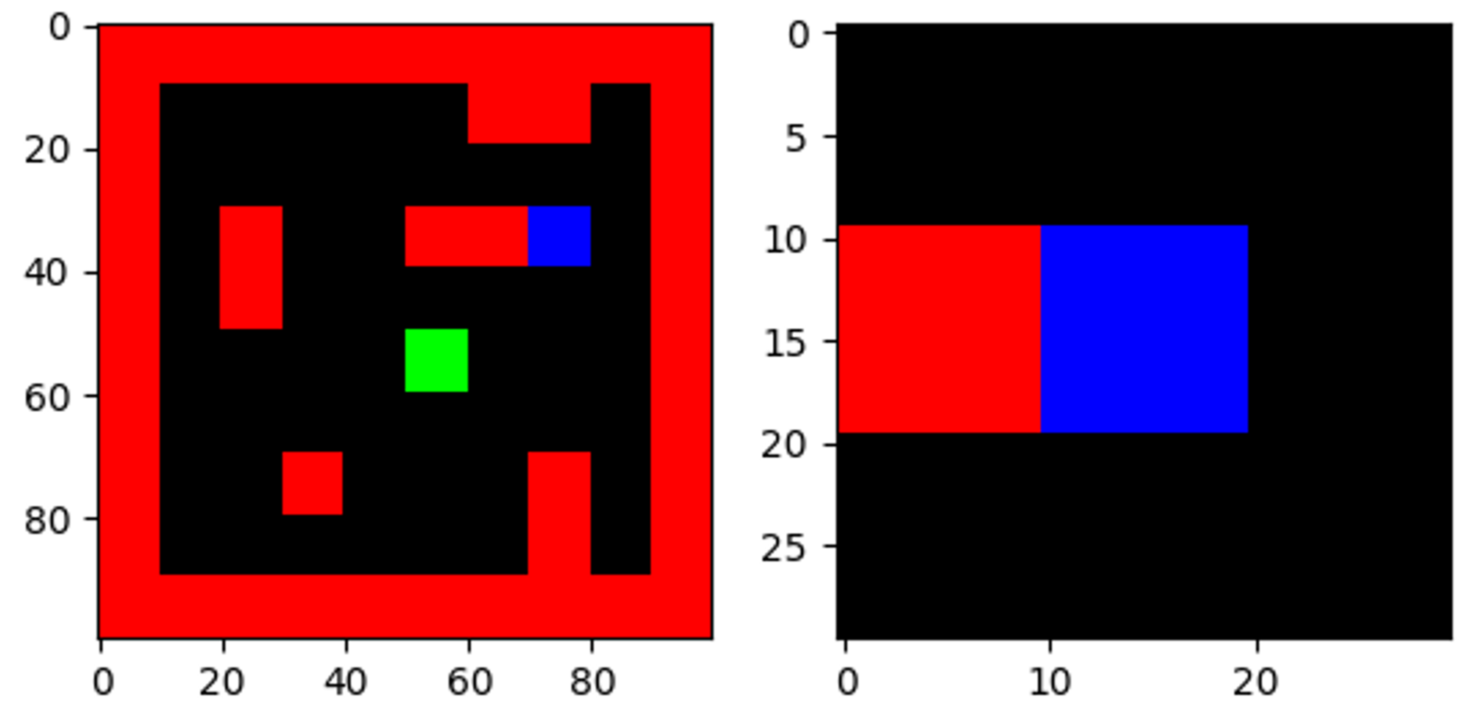
Deep Q-learning variants

Archit Basal, Ilia Dobrusin

Deep Q-learning (DQN)

- Use Deep Neural Network to approximate Q-function
- We implement and compare three **extensions to regular Deep Q-learning**
- **Visual Planning task**: find route to goal in a grid maze map with small local view
- Train for 300.000 steps with random sampling from previous experience



Target Network

- Simple extensions of regular DQN
- **Use second (target) network** to calculate target Q-values and next action (for loss calculation)
- **Keep target network static**, only update source network weights
- Synchronize source and target network periodically

Double DQN (DDQN)

- Regular DQN overestimates Q-values
- Similar to target network
- **Use source network for action prediction**
- Calculate **target Q-value with target network**

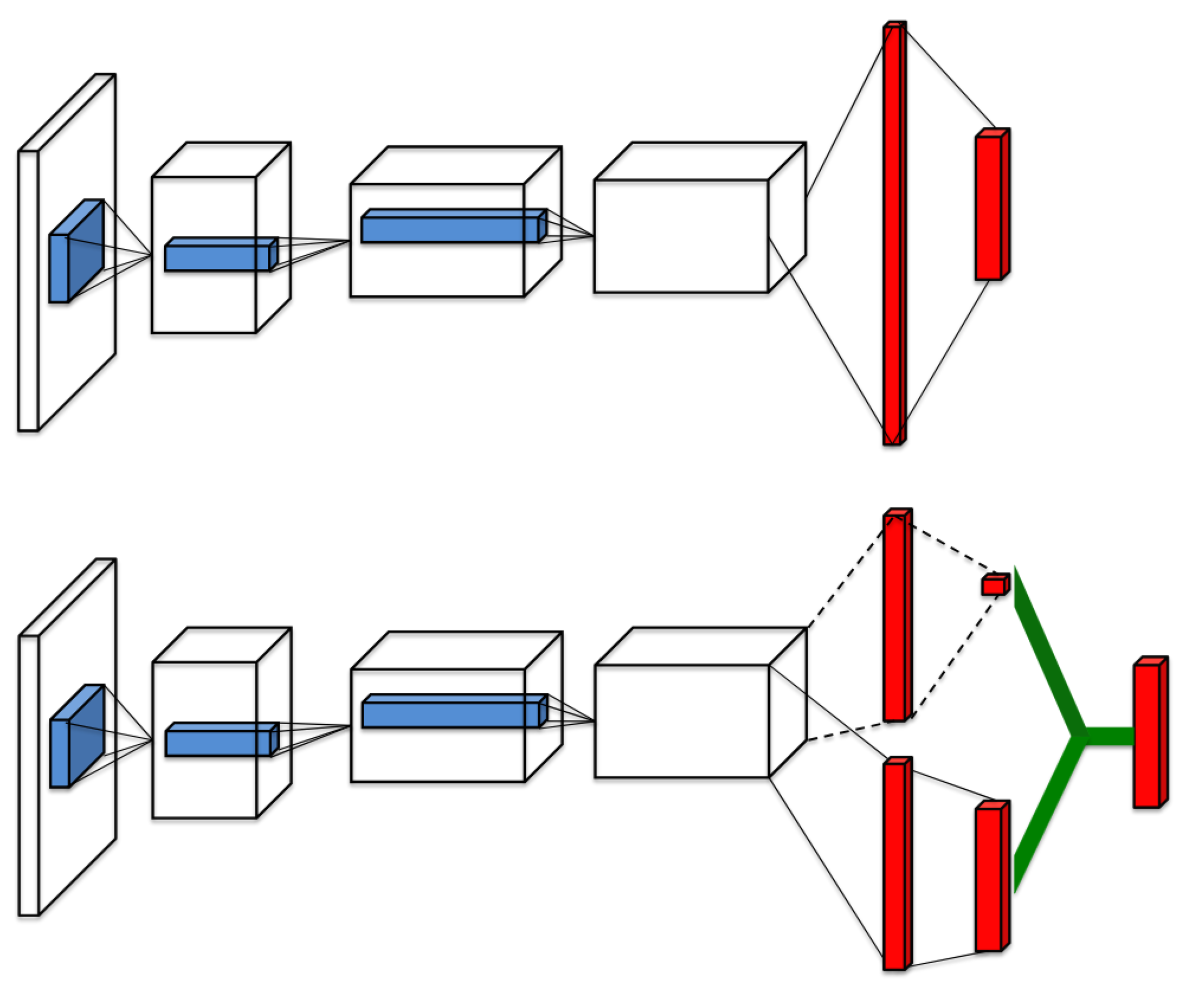
$$Q_{target} = r + \gamma Q(s', argmax(Q(s', a, \theta)), \theta')$$

Duelling Networks

- Q-values indicate how good each action is for given state
- Split Q-value calculation into two:
  - **Value function** V(s): indicates value of current state
  - **Advantage function** A(a): Comparrison of actions compared to each other

$$Q(s,a) = V(s) + A(s,a)$$

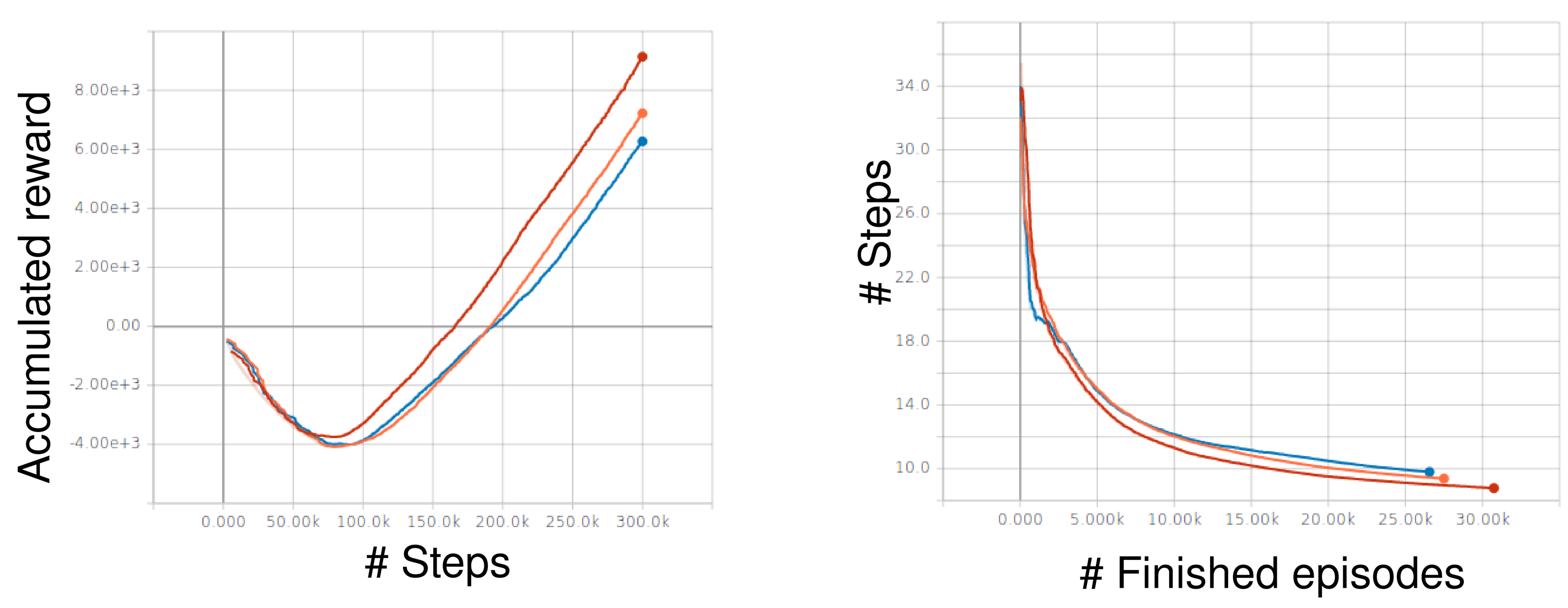
- **Duelling approach:** Let network compute seperate values for V and A by splitting network internally after convolution layers



[1]

Results

- **Training performance** Colors: DQN, Target Network, DDQN



- **Test performance**
  - Regular DQN, Target Network and DDQN reach the goal in 100 % of test episodes
  - Dueling DQN not shown (implementation issue)
- Regular DQN can overestimate Q-value, implement extensions stabilize results
- Target network keeps predicted Q-value fixed for a period of time – stabilize Q-value prediction
- Doube DQN decouples action selection and Q-valuation into seperate networks which increases performance further
- Dueling DQN performs best in theory

References:
[1] Ziyu Wang, Nando de Freitas, Marc Lanctot, *Dueling Network Architectures for Deep Reinforcement Learning,* in CoRR, 2015, arXiv:1511.06581v3