Given an agent...

- Observation state is a single monocular image
- Discrete action space: → ↘ ↖
Learn the environment...
And learn the second environment...

While preserving the knowledge about previous environment that has not been visited for longer period in time!
Dense rewards in range \([-1, 1]\).
Network Architecture

- Input image (1,50,70)
- Input action (1,3)
- Output is the reward prediction for a given state-action pair in range [-1, 1]

$Q(s,a|\theta)$
Experience Replay

• Playing Atari with Deep Reinforcement Learning (Mnih et al. in 2013)
• Learn from previous experiences
• **Replay Memory** - Maintain a list of previously gathered sample triplets (state, action, reward)
• During the training phase, randomly sample a mini-batch from the replay memory:
  – > Global solution for the given environment
• Possible Alternative: Multi-Agent strategies
Experience Replay Flowchart

\[
\text{mse}(mb) = \frac{1}{128} \sum_{(s,a,r) \in mb} (r - Q(s, a; \theta))^2
\]
Experience Replay, Sample Selectivity

• Which samples to replay?
  – Prioritized Experience Replay
    *(Schaul et al. in 2015)*

• Which samples to keep?
  – Standard approach can lead to agent
catastrophically forgetting the less frequently
appearing experiences
Replay Memory Management (RMM)

• Replay Memory is limited in size:
  – What can we do once it is full?

NoRMM

Similar Samples are marked red.

RMM

RMM preserves less frequently appearing samples in the replay memory!
Replay Memory Management (RMM)

- Measuring Sample Similarities:
  - Extract binary descriptors from images
  - When similar descriptors are found, mark the older one as a discard candidate
  - Compare only the samples containing same actions
- Update rewards for samples occasionally:
  \[ r_t = 0.7r_t + 0.3Q(s_t, a_t | \theta) \]
Lag due to very high CPU usage!
Every prediction takes about equal amount of time!
I16c Experiment
I16c Experiment

(a)  
(b)  
(c)  
(d)  
(e)  
(f)  
(g)  
(h)  

Action color legend | forward | forward/right | forward/left
Conclusion

• Advantages of RMM:
  – Enables preserving knowledge about less frequent experiences
  – Enables a higher variation of experiences if memory budget is low

• Possible Future Work:
  – Evaluation, learning multiple ATARI 2600 video games at the same time
  – Adaptation for Continuous Action Space
Thank you for your attention!
Questions?