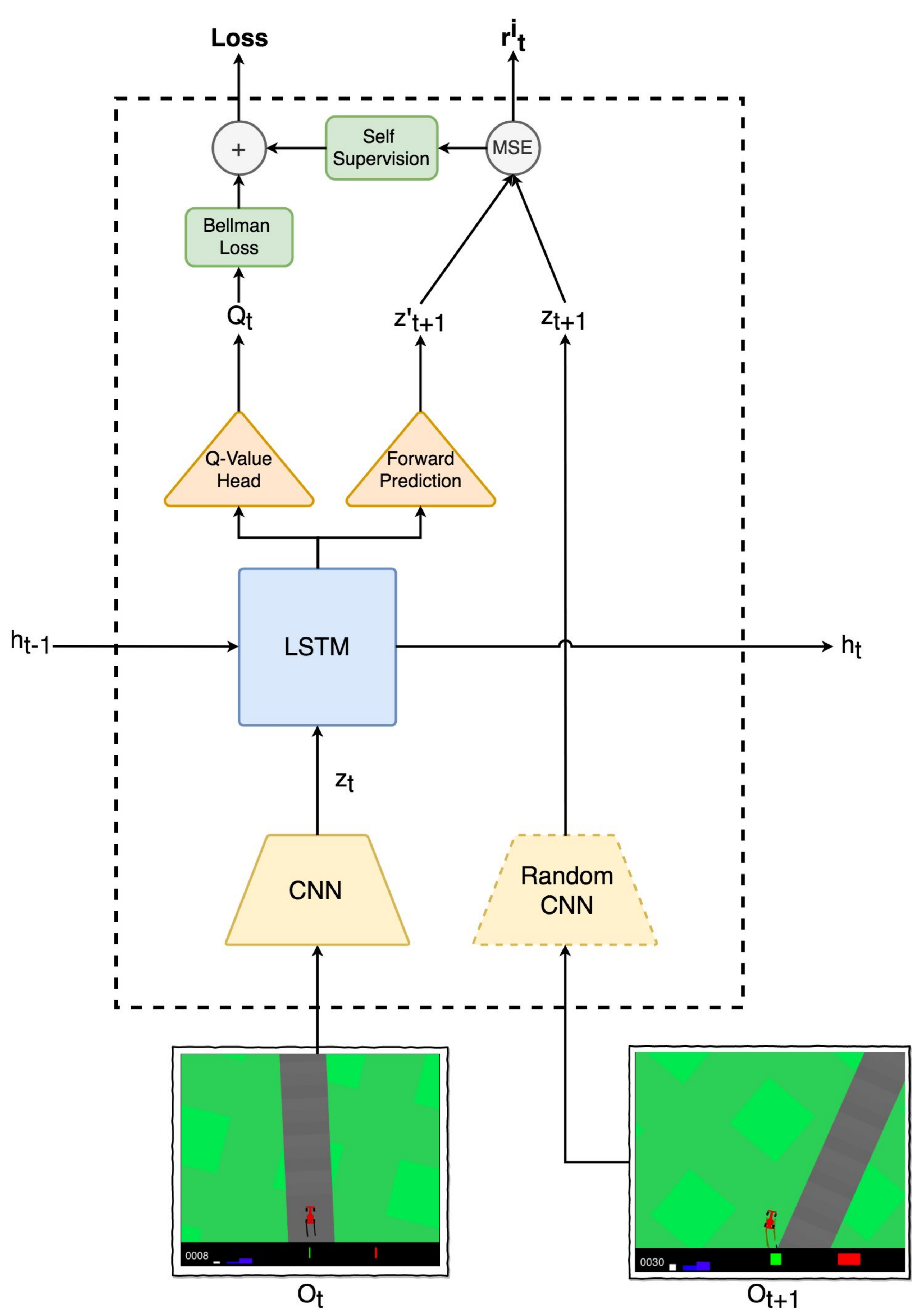
A Curious Model Is All You Need

Shubham Chandel
Tulsi Jain

Motivation

- Reinforcement Learning agents need to learn the **consequences of their actions**.
- Current RL approaches are highly sample **inefficient**. A model of the world is required, to efficiently use each sample.
- Each environment requires hand-engineered extrinsic reward, which is not scalable.
- Curiosity is a type of **intrinsic reward** function which uses prediction error as reward signal.
- A model based curious agent can learn to solve a task, **without** the need of **extrinsic rewards**.

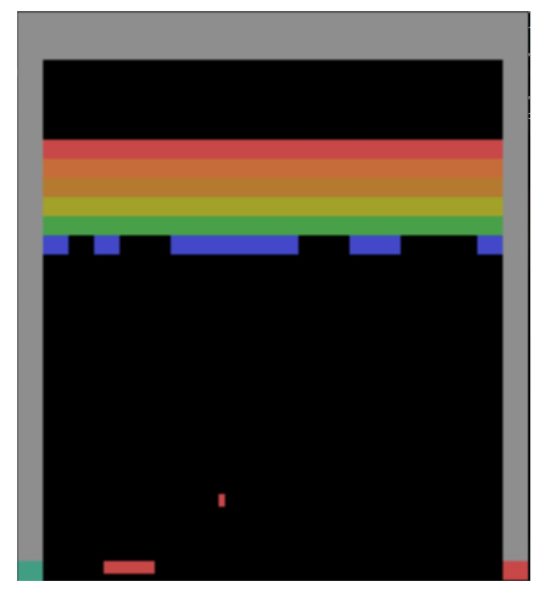
Approach

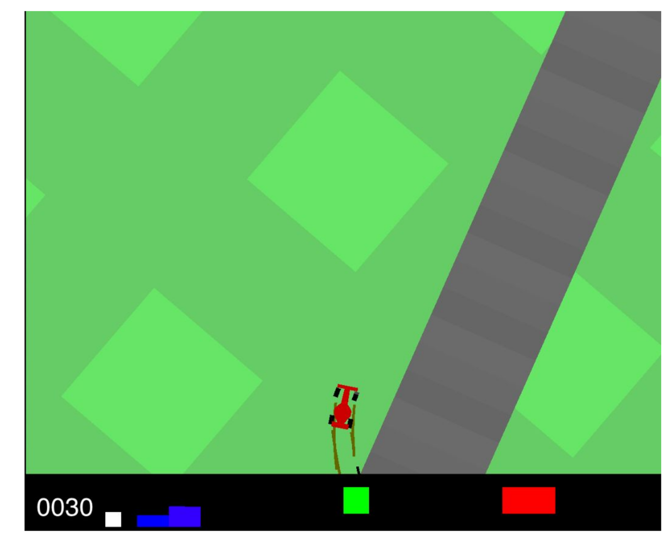


- We propose a generative LSTM, which serves two tasks:
  - Predict **Q-value** for each action, Q_{t+1}.
  - Predict the **future state representation**, z'_{t+1}.
- A **random CNN**, provides a **stable** future state representation, z_{t+1}.
- The error in prediction, $E(z_{t+1}, z'_{t+1})$, is the **intrinsic reward,** r^i_t, to agent, which we show is aligned to extrinsic reward by the environment.
- Further, the agent is trained to minimize the **sum of two losses**:
  - **Bellman Loss**, encountered from deviation from expected Q-value.
  - **Self-Supervision Loss**, as an auxiliary forward-prediction loss for extra learning signal.
- Deep Q Learning, with experience replay buffer and target network is used to minimize the bellman error.

Experiments and Results

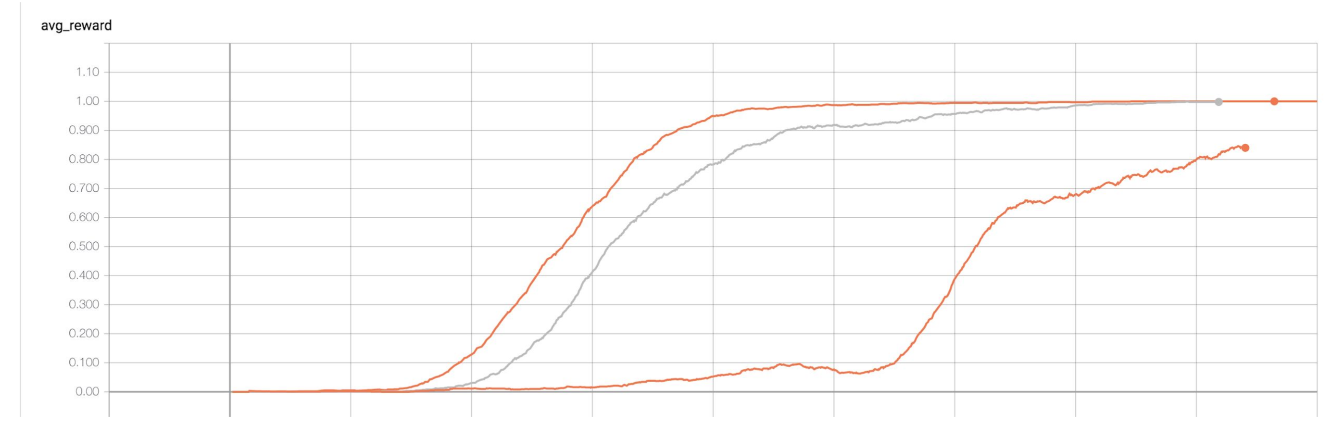
- We use two ALE environments, Breakout and CarRacing.





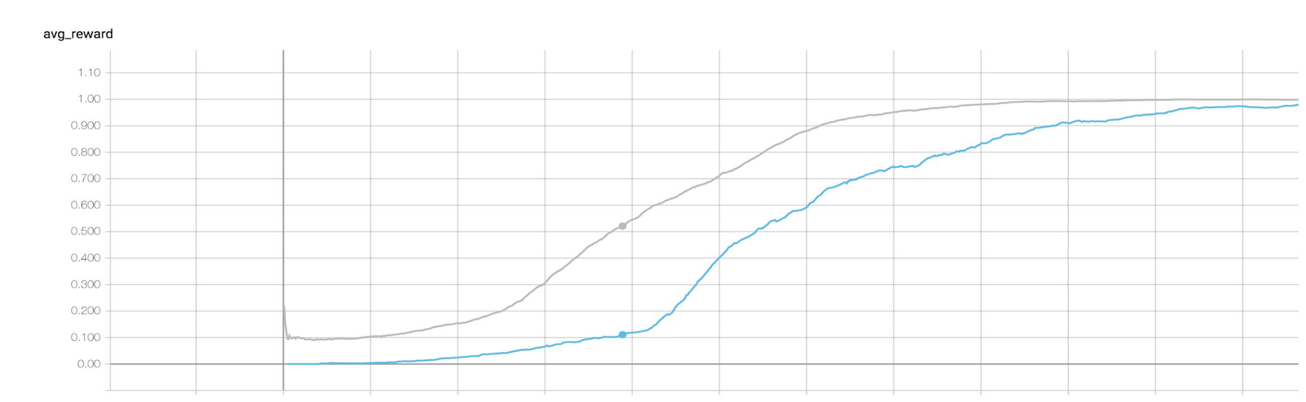
Ablation over Components

- Curious Model
- with random features
- w\t LSTM



Ablation over Auxiliary task

- Self Supervision
- w\t Self Supervision



Method	Avg. Score
DQN	343 ± 18
A3C (continuous)	591 ± 45
A3C (discrete)	652 ± 10
Intrinsic Curiosity Module (ICM)	813 ± 42
World Model	906 ± 21
Curious Model	712 ± 11

Table 1: `CarRacing-v0` scores achieved using various methods.

Related Work

- Deep Recurrent Q-Learning (Hausknecht et al., 2015)
- Curiosity-driven Exploration (Pathak et al., 2017)
- World Model (Ha et. al., 2018)
- Large-Scale Study of Curiosity (Burda et al., 2018)
- Random Network Distillation (Burda et al., 2018)
- Diversity is All You Need (Eysenbach al., 2018)

Conclusion

- Purely curiosity-driven learning, i.e. **without any extrinsic rewards**, is sufficient for certain environments.
- A **stable but random** feature-space performs competitive to a learned feature-space.
- Model based reinforcement learning, along with model free is a good tradeoff, to **maximize sample efficiency**.