



# Teach AI to play Atari iceHockey game using gameplay videos



Presenter: Xu Yangyang [u6325688]  
Supervisor: Dr. Penny Kyburz  
COMP8755

# Content

- ▶ Problem statement & challenges
- ▶ Background
- ▶ Methods
- ▶ Q&A

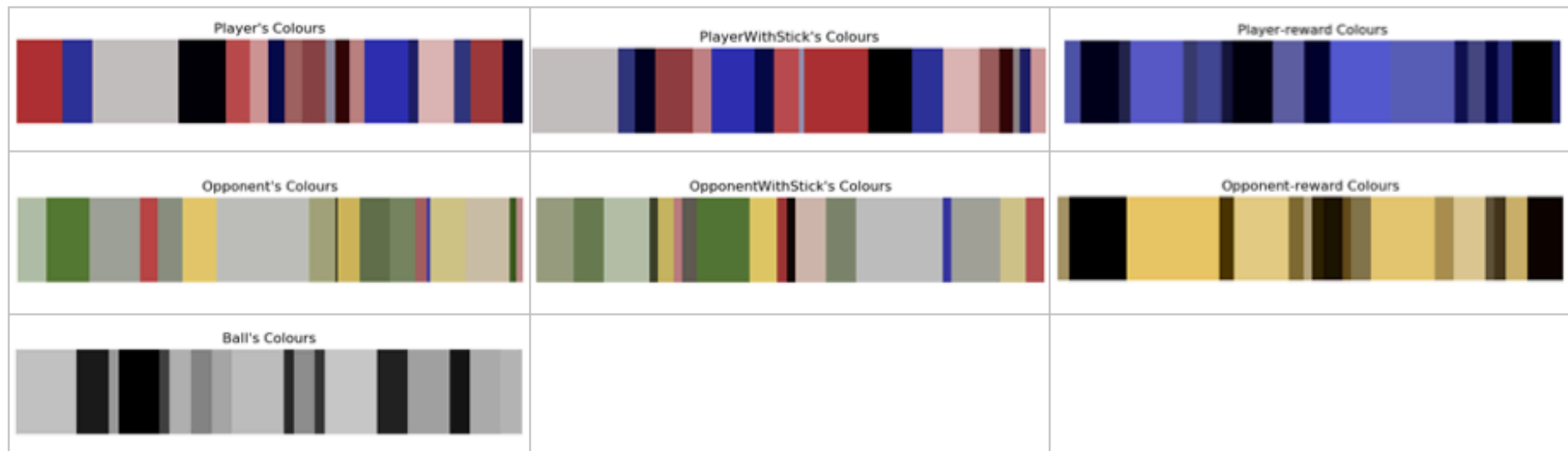
# Problem statement

Use historical gameplay videos to teach AI play game (choose action).

## Challenges

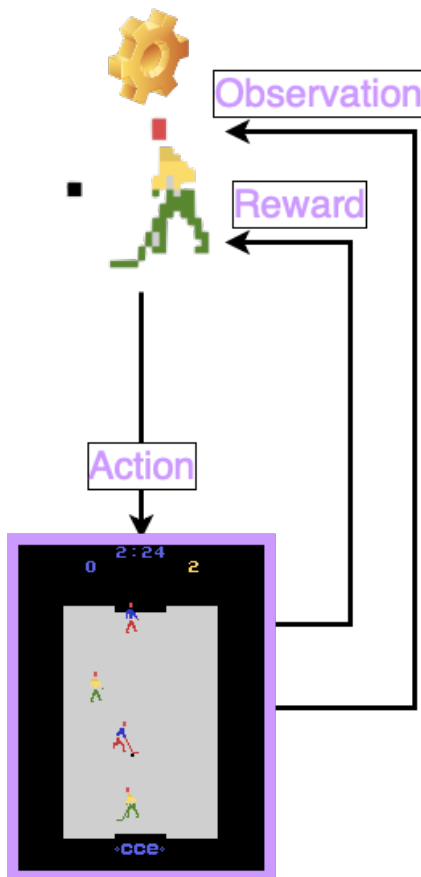
Videos have different resolutions and sizes.

Youtube gameplay videos do not have action records.



# Background - Reinforcement Learning

Markov Decision Process (MDP): (State, Actions, Transition Probabilities, Reward)



**Forward  
Reinforcement  
Learning**

**Inverse  
Reinforcement  
Learning**

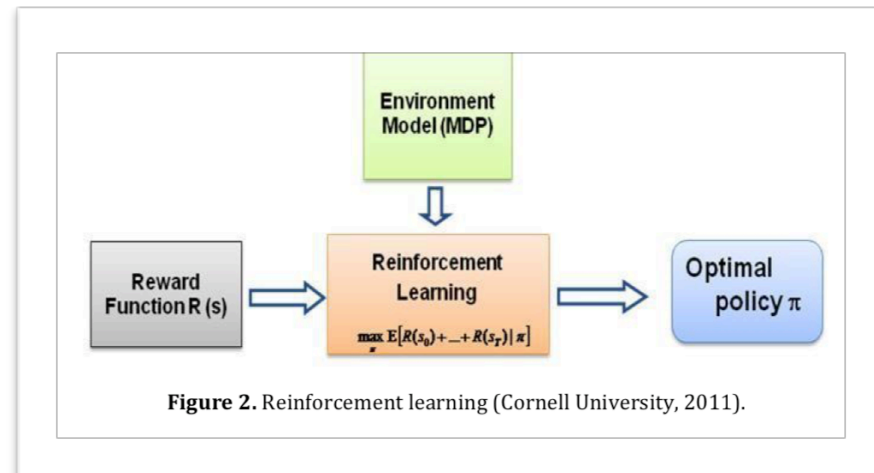


Figure 2. Reinforcement learning (Cornell University, 2011).

Pavoloiu, A., & Kose, U. (2017). Ethical artificial intelligence-an open question. *arXiv preprint arXiv:1706.03021*.

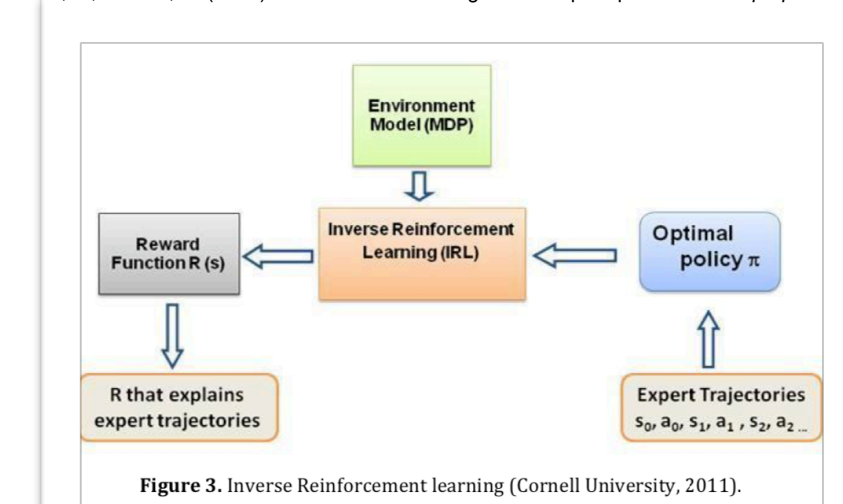
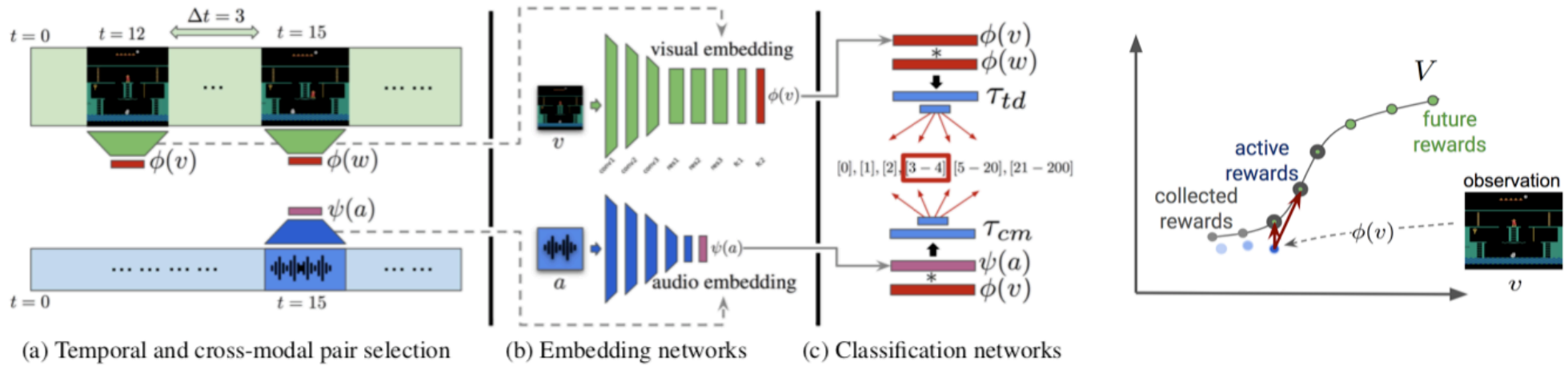


Figure 3. Inverse Reinforcement learning (Cornell University, 2011).

Pavoloiu, A., & Kose, U. (2017). Ethical artificial intelligence-an open question. *arXiv preprint arXiv:1706.03021*.

# Background - Related work <sup>[1]</sup>

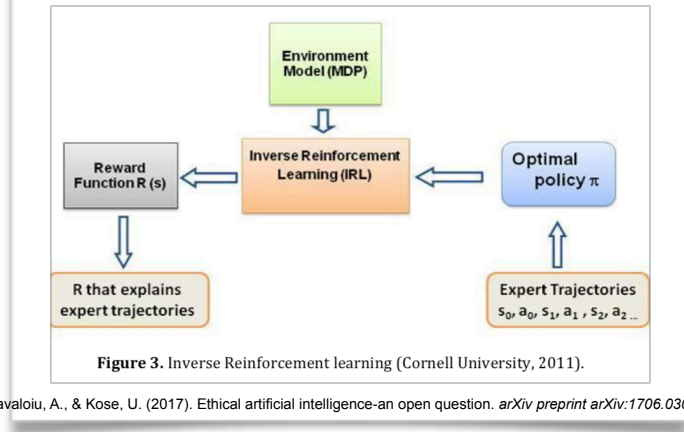


	MONTEZUMA'S REVENGE	PITFALL!	PRIVATE EYE
Rainbow [19]	384.0	0.0	4,234.0
ApeX [22]	2,500.0	-0.6	49.8
DQfD [20]	4,659.7	57.3	42,457.2
Average Human [43]	4,743.0	6,464.0	69,571.0
Ours ( $r_{\text{imitation}}$ only)	37,232.7	54,912.4	98,212.5
Ours ( $r_{\text{imitation}} + r_{\text{env}}$ )	<b>58,175.1</b>	<b>76,812.5</b>	<b>98,763.2</b>

Table 1: Comparison of our best policy (mean of 200 evaluation episodes) to previously published results across MONTEZUMA'S REVENGE, PITFALL! and PRIVATE EYE. Our agent is the first to exceed average human-level performance on all three games, even without environment rewards.

# Method - GAIL

GAIL: (Generative Adversarial Imitation Learning)<sup>[2]</sup>



# Method - State representation

## Current Object Detection Method:

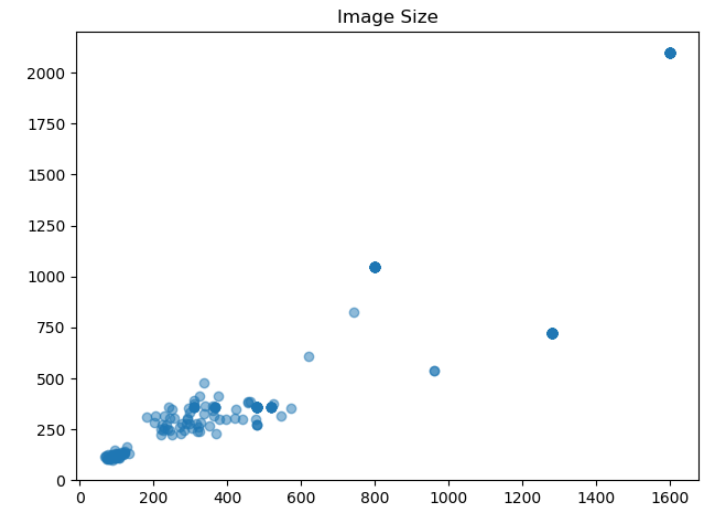
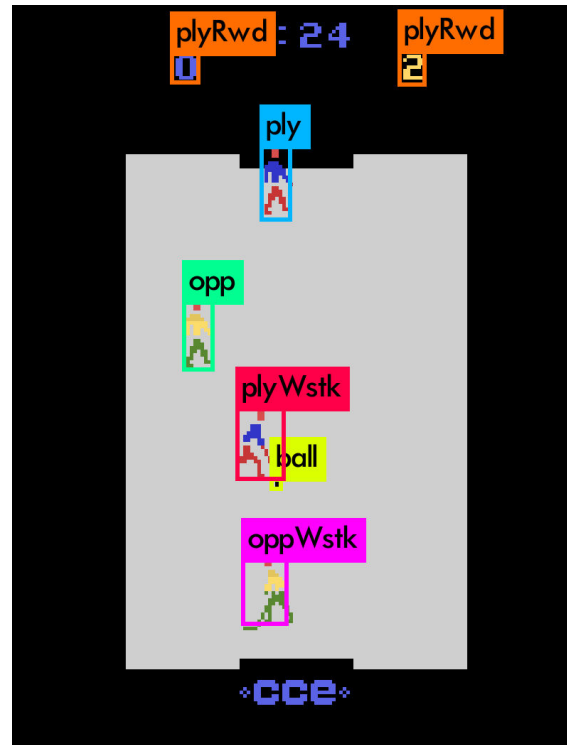
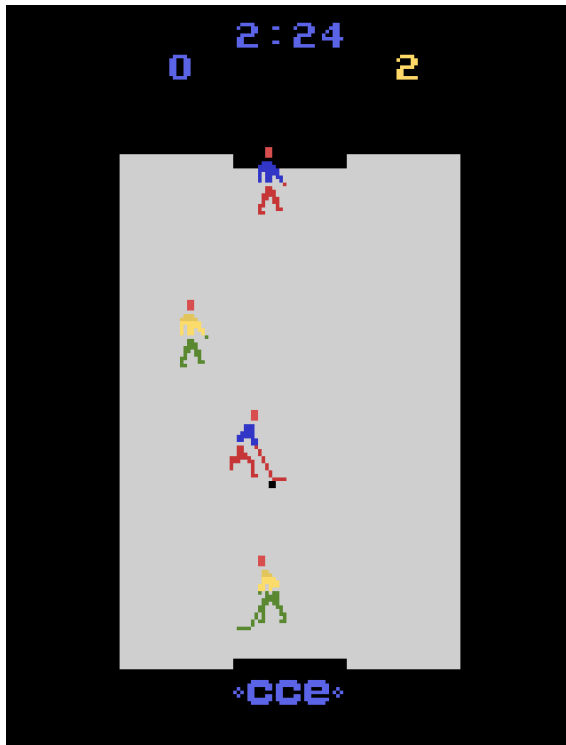
### One-Stage

YOLO(You Only Look Once)[3]

Total: 264 training images

Framework: Darknet[4]

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Objects Types	Ball	Opp	Player	OppWstk	PlyWstk	OppRwd	PlyRwd
Training Data	151(12.5%)	184(15.2%)	190(15.7%)	185(15.3%)	188(15.6%)	156(12.9%)	154(12.7%)



Player:  Player With Stick:  Player Reward:  Opponent:  Opponent With Stick:  Opponent Reward: 

# References

- [1] Aytar, Yusuf, et al. "Playing hard exploration games by watching youtube." *Advances in Neural Information Processing Systems*. 2018.
- [2] Ho, J., & Ermon, S. (2016). Generative adversarial imitation learning. In *Advances in neural information processing systems* (pp. 4565-4573).
- [3] PavaloIU, A., & Kose, U. (2017). Ethical artificial intelligence-an open question. *arXiv preprint arXiv:1706.03021*.
- [4] J. Redmon, "YOLO: Real-Time Object Detection", *Pjreddie.com*, 2019. [Online]. Available: <https://pjreddie.com/darknet/yolo/>. [Accessed: 08- Aug- 2019].