Playing hard exploration games by watching YouTube

Yusuf Aytar, Tobias Pfaff, David Budden,
Tom Le Paine, Ziyu Wang, Nando de Freitas

Learning by watching YouTube

People learn many tasks by watching online videos

Despite huge gaps in visual appearance,
sensing modalities,
body differences, etc..

knitting

construction

playing games

Learning by watching YouTube

People learn many tasks by watching online videos

Despite huge gaps in visual appearance,
sensing modalities,
body differences, etc..

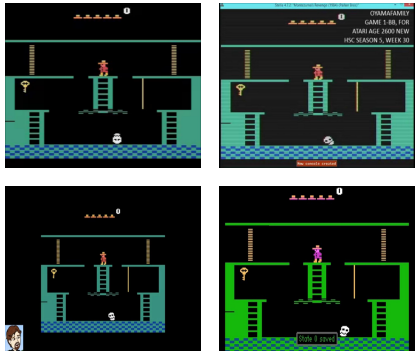
knitting

construction

playing games

Challenges

Domain Gap

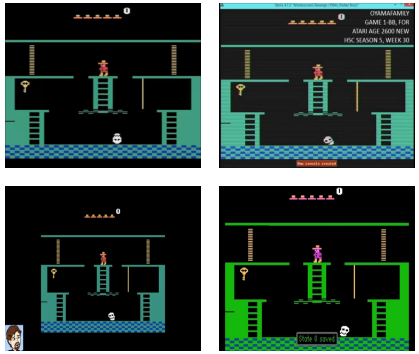


No Actions

No Rewards

Challenges

Domain Gap



Self-Supervised Domain Alignment

No Actions

Learn to Play with Imitation (RL)

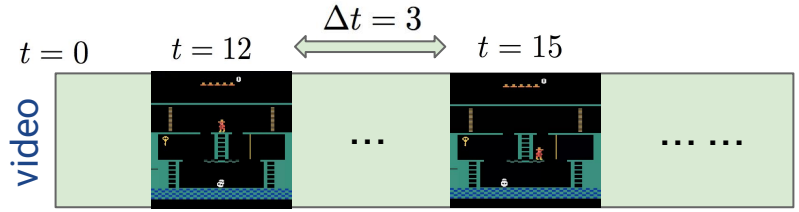
No Rewards

Rewards Learned from Expert Sequence

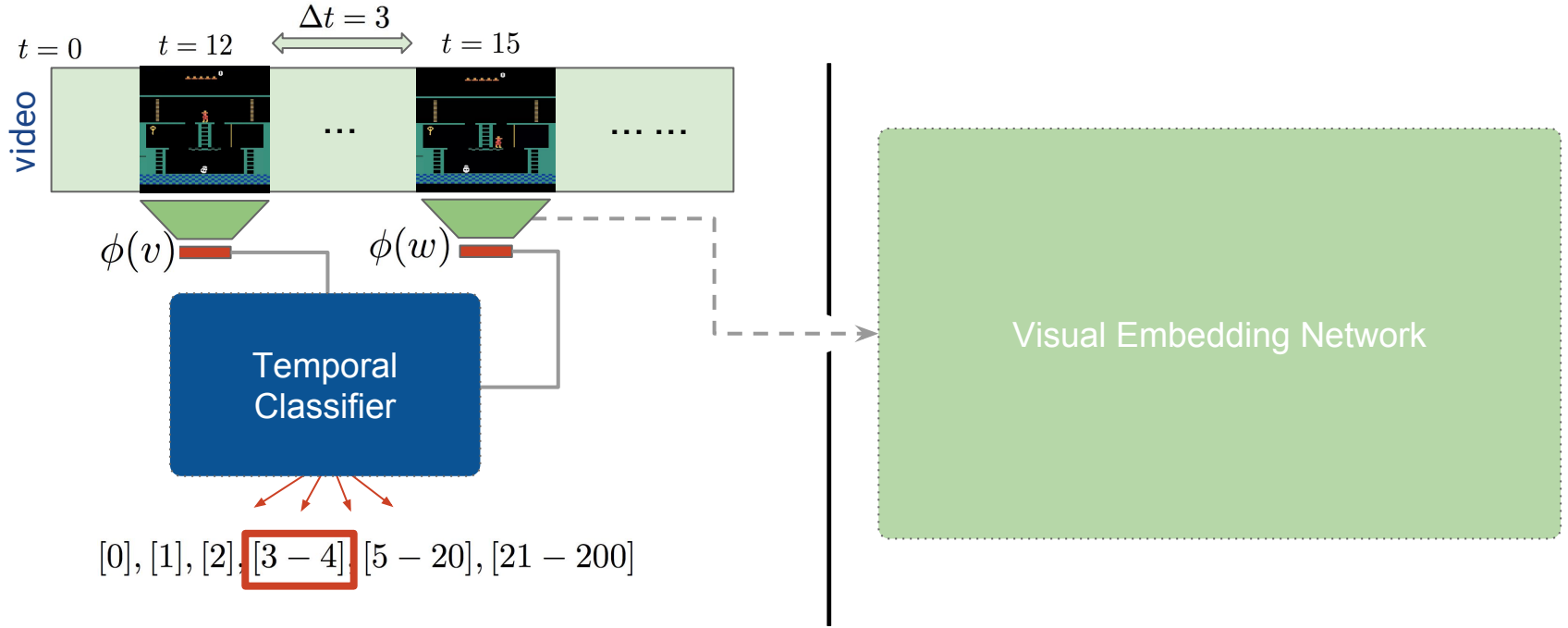
Temporal distance classification (TDC)

demonstration sequence

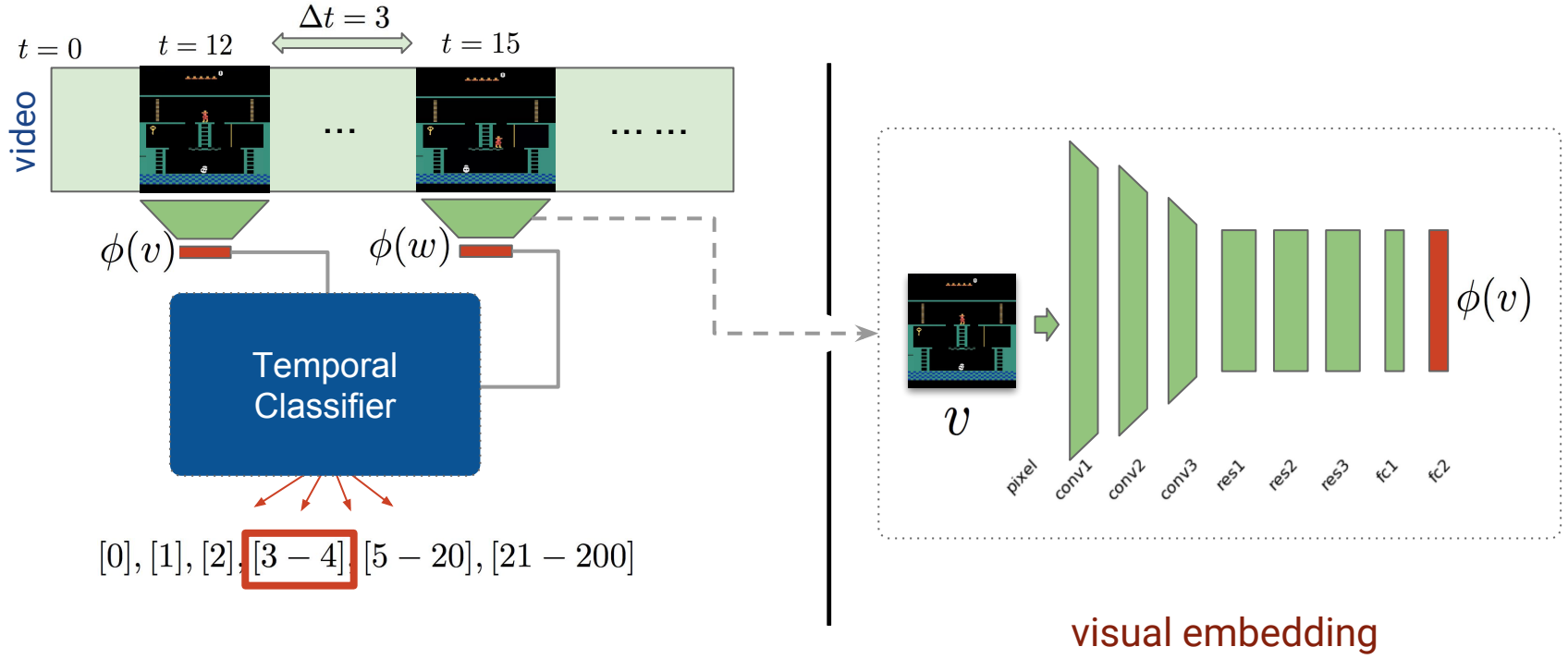
Temporal distance classification (TDC)



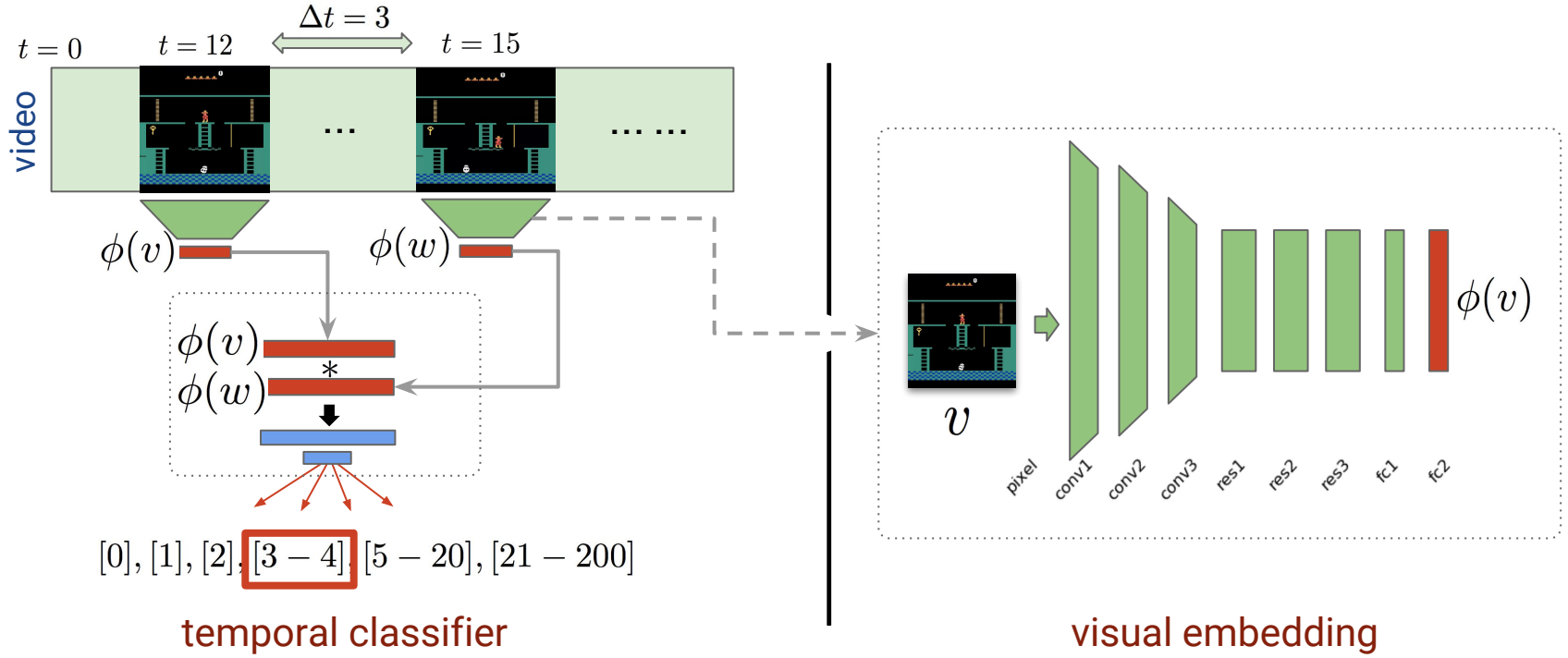
Temporal distance classification (TDC)



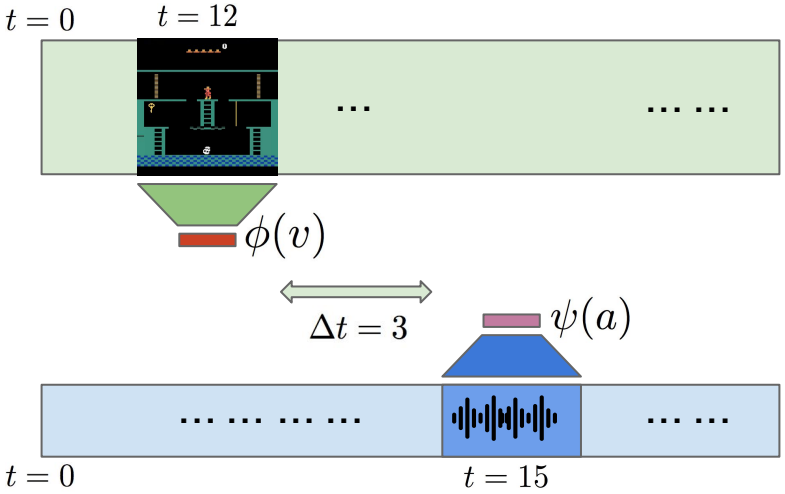
Temporal distance classification (TDC)



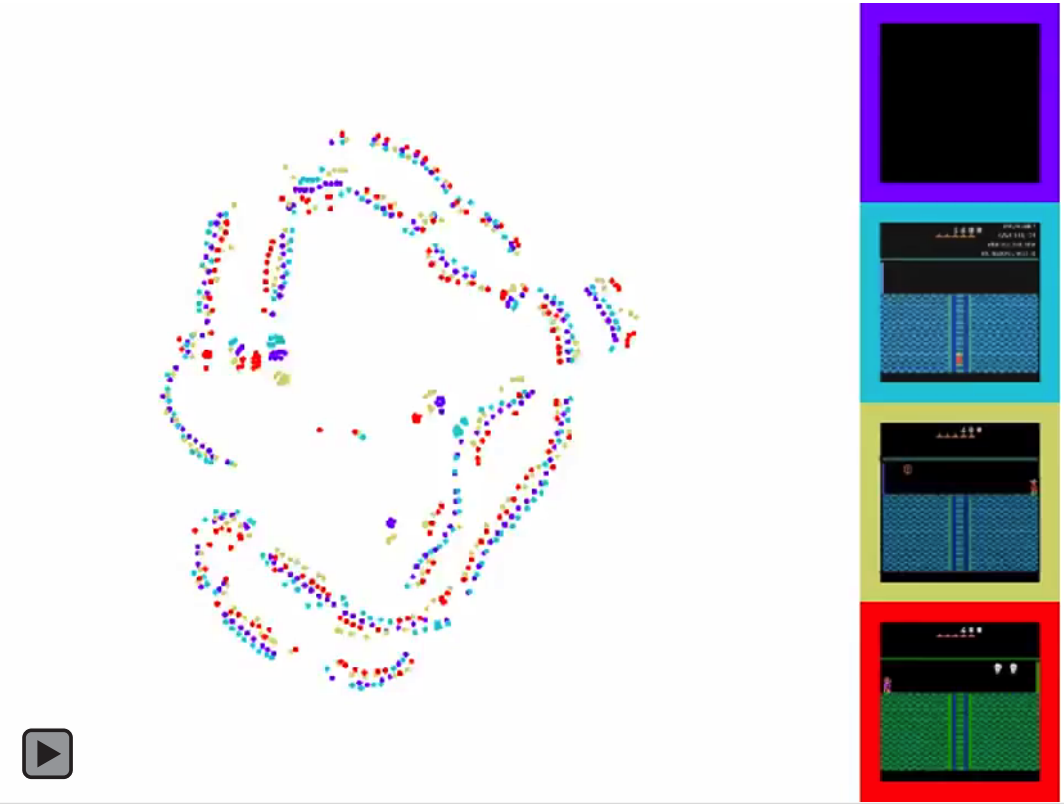
Temporal distance classification (TDC)



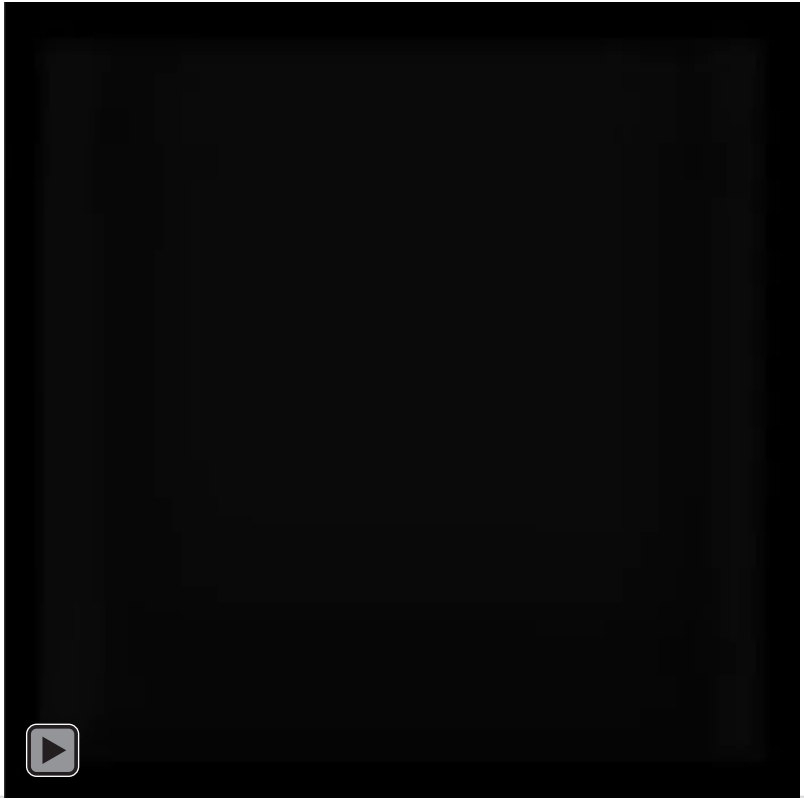
Cross-modal distance classification (CMC)



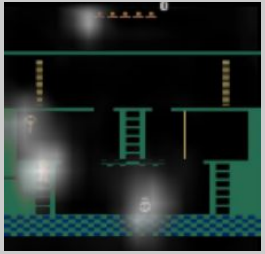
Model successfully aligns different videos



What does the embedding focus on?

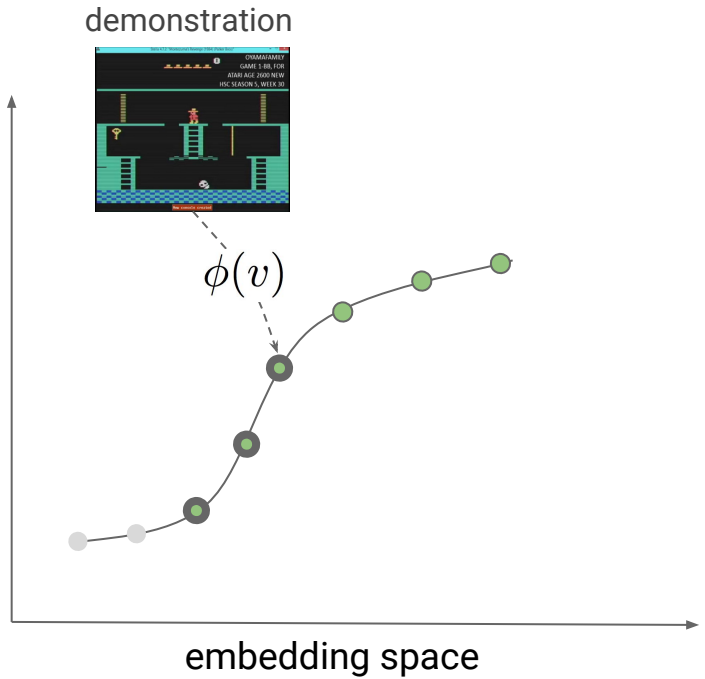


Visual only

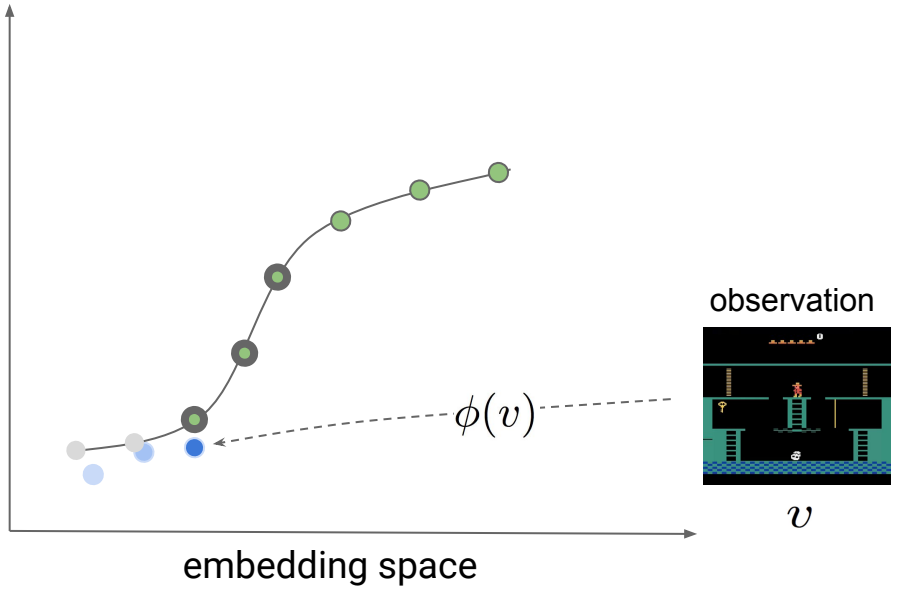


Cross-modal

Imitation through RL



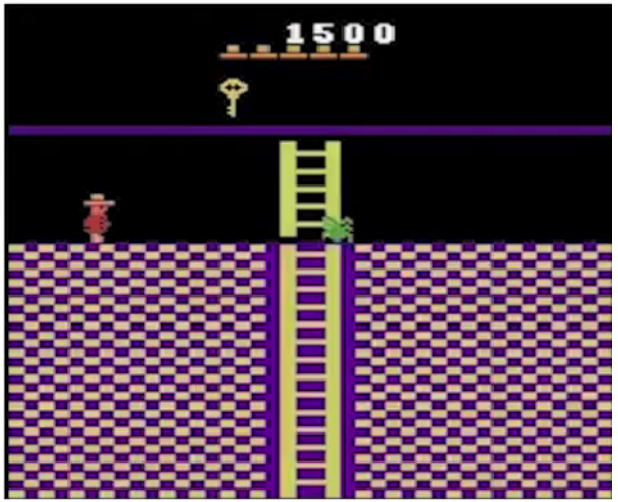
Imitation through RL



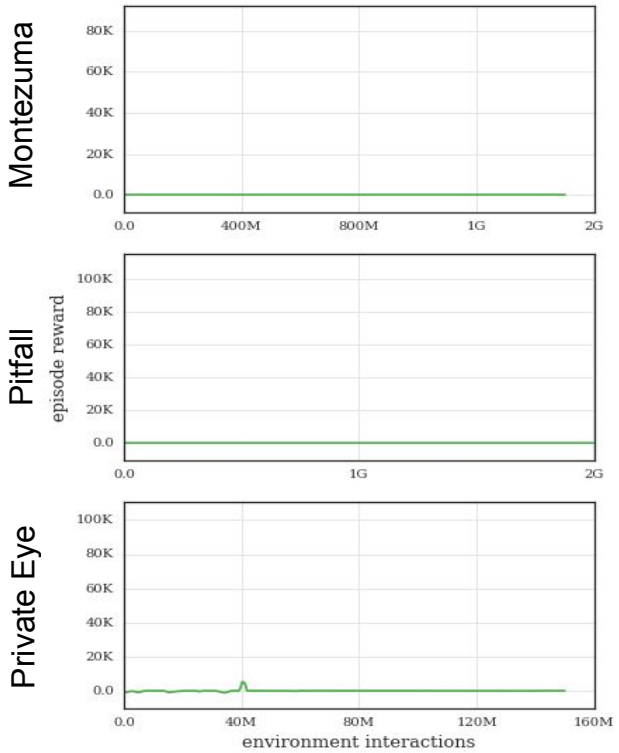
RL makes imitation more **robust**

demonstration

learnt agent



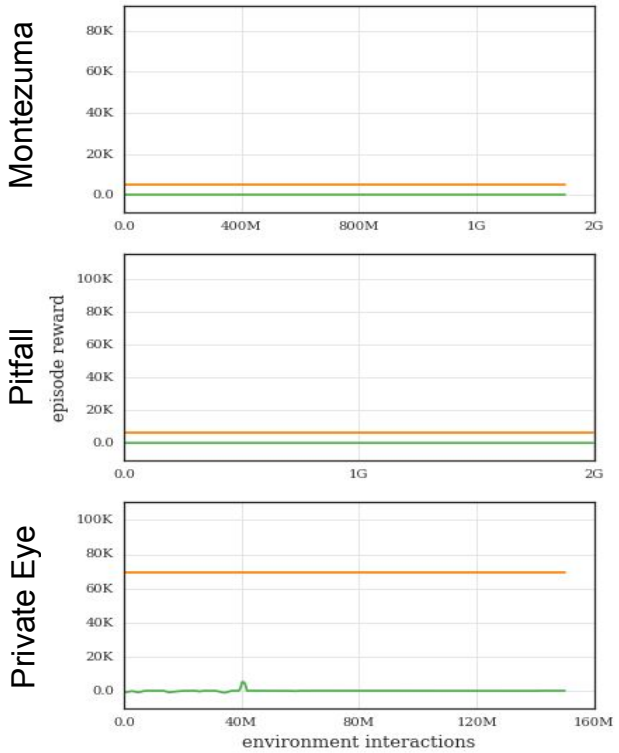
Results



	Montezuma	Pitfall!	Private Eye
Pure RL	**~ 2,500**	**~ 0**	**~ 50**
Avg. Human	4,743	6,464	69,571
DQfD (2018)	29,384	3,997	100,747
Ours	58,175	74,323	98,763

Averaged score of best policy

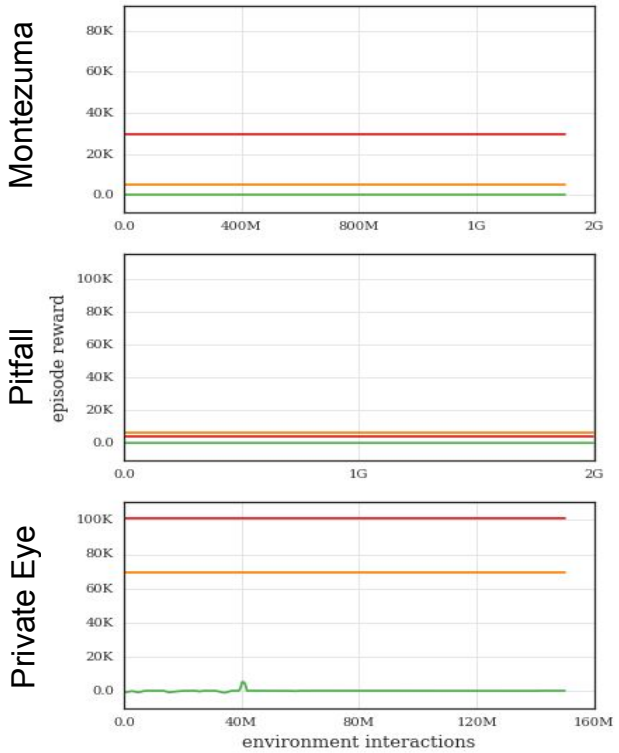
Results



	Montezuma	Pitfall!	Private Eye
Pure RL	~ 2,500	~ 0	~ 50
Avg. Human	**4,743**	**6,464**	**69,571**
DQfD (2018)	29,384	3,997	100,747
Ours	58,175	74,323	98,763

Averaged score of best policy

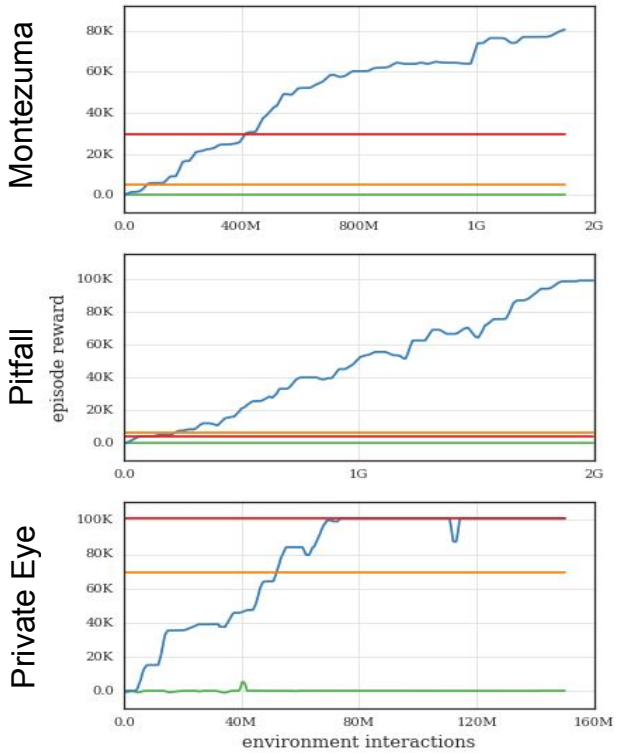
Results



	Montezuma	Pitfall!	Private Eye
Pure RL	~ 2,500	~ 0	~ 50
Avg. Human	4,743	**6,464**	69,571
DQfD (2018)	**29,384**	3,997	**100,747**
Ours	58,175	74,323	98,763

Averaged score of best policy

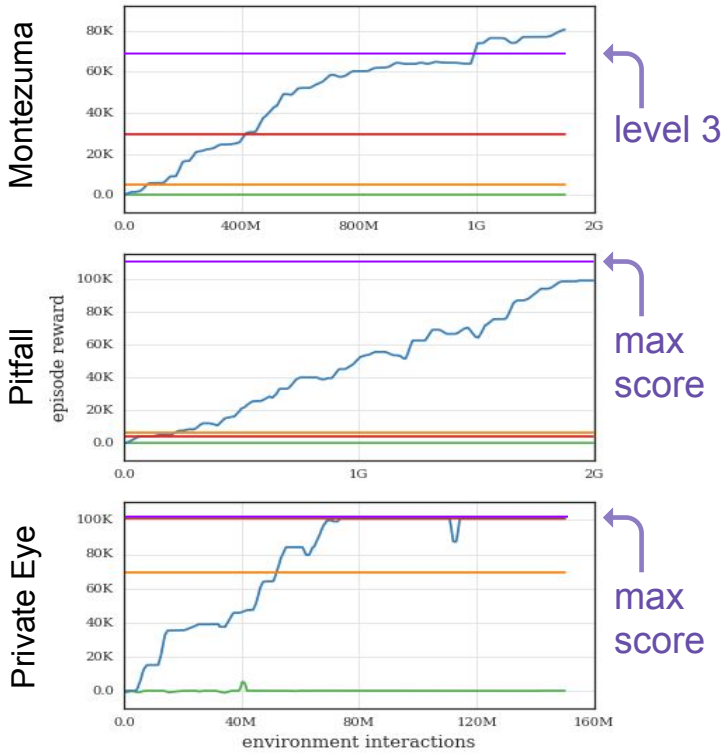
Results



	Montezuma	Pitfall!	Private Eye
Pure RL	~ 2,500	~ 0	~ 50
Avg. Human	4,743	6,464	69,571
DQfD (2018)	29,384	3,997	**100,747**
Ours	**58,175**	**74,323**	98,763

Averaged score of best policy

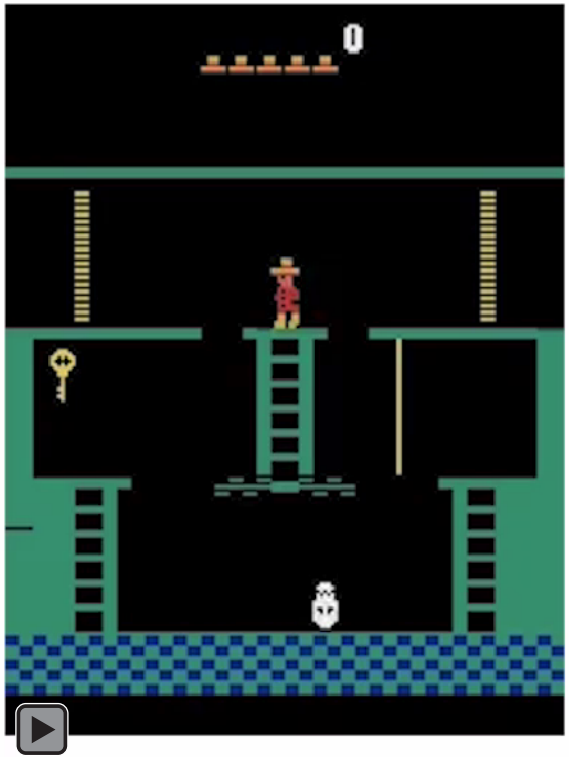
Results



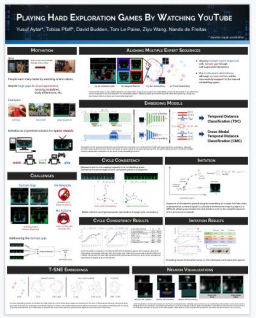
	Montezuma	Pitfall!	Private Eye
Pure RL	~ 2,500	~ 0	~ 50
Avg. Human	4,743	6,464	69,571
DQfD (2018)	29,384	3,997	**100,747**
Ours	**58,175**	**74,323**	98,763

Averaged score of best policy

Visit our poster !



Playing hard exploration games by watching Youtube



#142