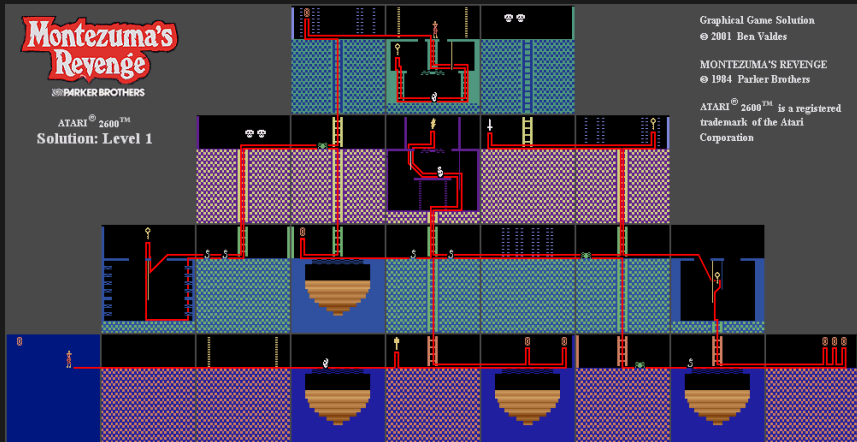


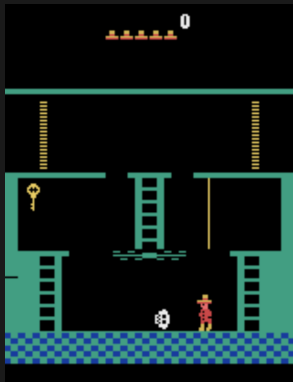
A dramatic painting depicting a Native American ceremony on a cliff. A central figure, a Native American man, stands on a high platform, facing away from the viewer with arms raised in a gesture of praise or prayer. He wears an elaborate feathered headdress with yellow, black, and red feathers, and a loincloth with a red and blue pattern. To his right, a large, ornate banner with a central emblem is visible. In the foreground, the back of a person's head with a feathered headdress is seen, looking towards the central figure. The background shows a vast, hazy landscape with a large, stepped pyramid or temple structure in the distance. The overall atmosphere is one of a significant, ancient event.

# challenging Montezuma's Revenge intrinsic motivation in RL Michal CHOVANEC

# Montezuma's Revenge



# Montezuma's Revenge



- **very sparse rewards** - hundreds of steps
- **huge state space**
- **hard exploration**
- **needs returns back**

# highlighted score

<https://paperswithcode.com/sota/atari-games-on-atari-2600-montezumas-revenge>

year	name	score
2015	Deep Reinforcement Learning with Double Q-learning	0
2021	MuZero	2500
2018	Count-Based Exploration with Neural Density Models <sup>1</sup>	3705
<b>2019</b>	<b>Exploration by Random Network Distillation <sup>2</sup></b>	<b>8152</b>
2021	GoExplore* <sup>3</sup>	43 000

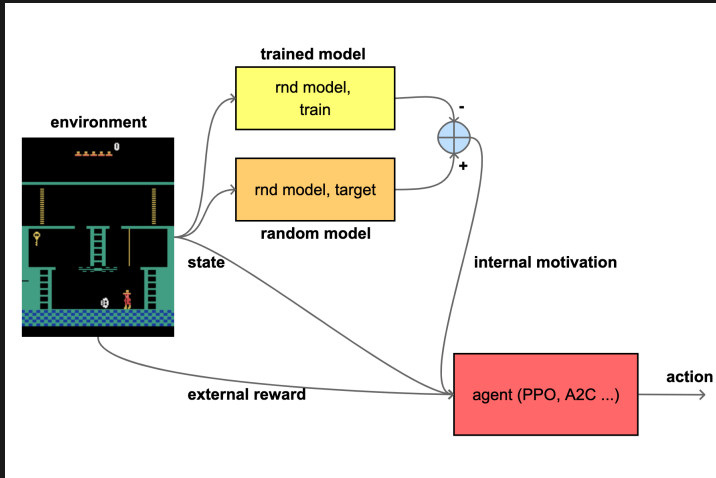
\* : requires environment state saving/loading

<sup>1</sup><https://arxiv.org/abs/1703.01310>

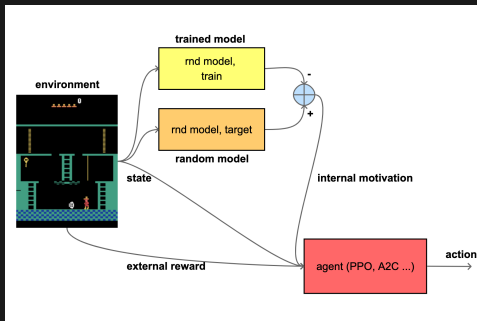
<sup>2</sup><https://arxiv.org/abs/1810.12894>

<sup>3</sup><https://arxiv.org/abs/2004.12919>

# random network distillation

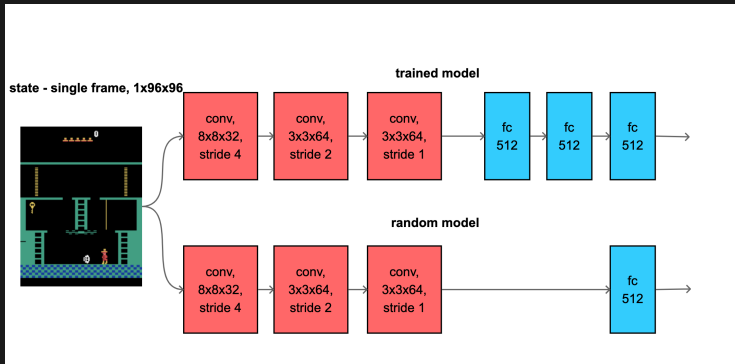


# random network distillation

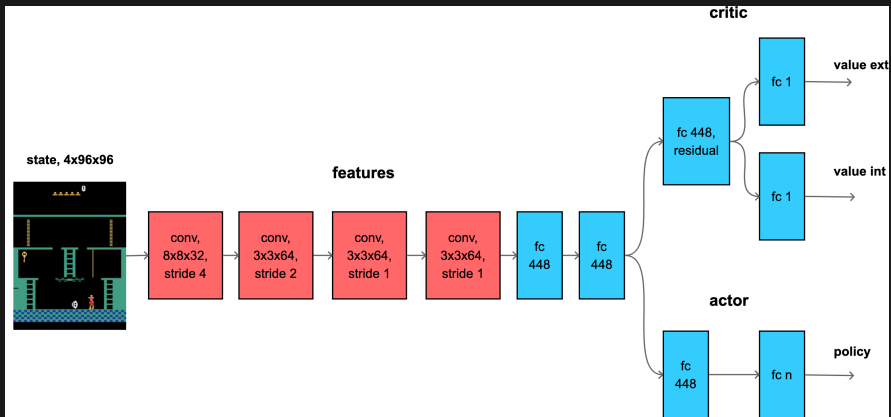


- neural network works as **novelty detector**
- model learns to imitate random (target) model
- **less visited states produce bigger motivation signal**
- orthogonal weights initialisation ( $g = 2^{0.5}$ ) for strong signal
- lot of fully connected layers **to avoid generalisation**

# random network distillation architecture

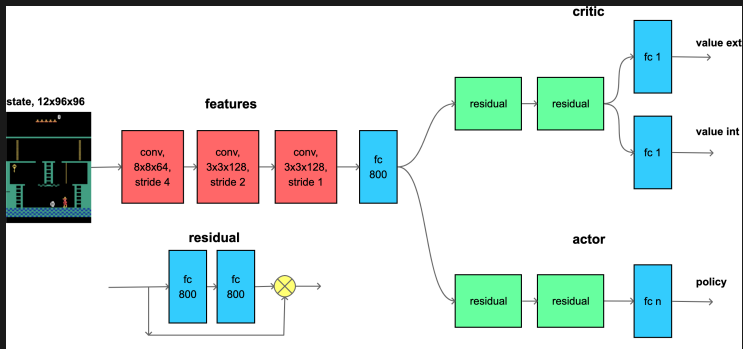


# ppo model architecture

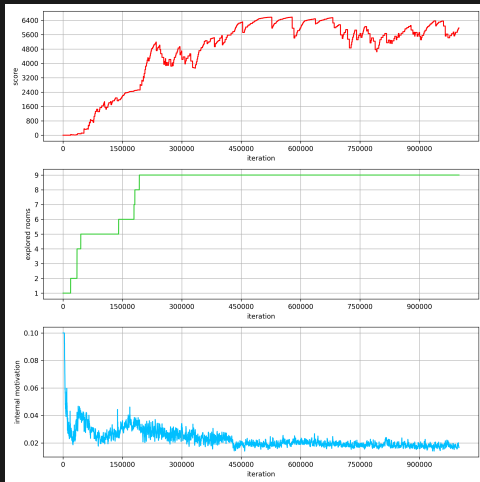




# ppo model architecture



# results



- 1M steps - 20% of original paper
- 128 parallel envs = total 128M steps
- score 6400
- 9 rooms explored

# Q&A



Michal CHOVANEC, PhD

random network distribution