DEEP FEATURE EXTRACTION FOR SAMPLE-EFFICIENT REINFORCEMENT LEARNING

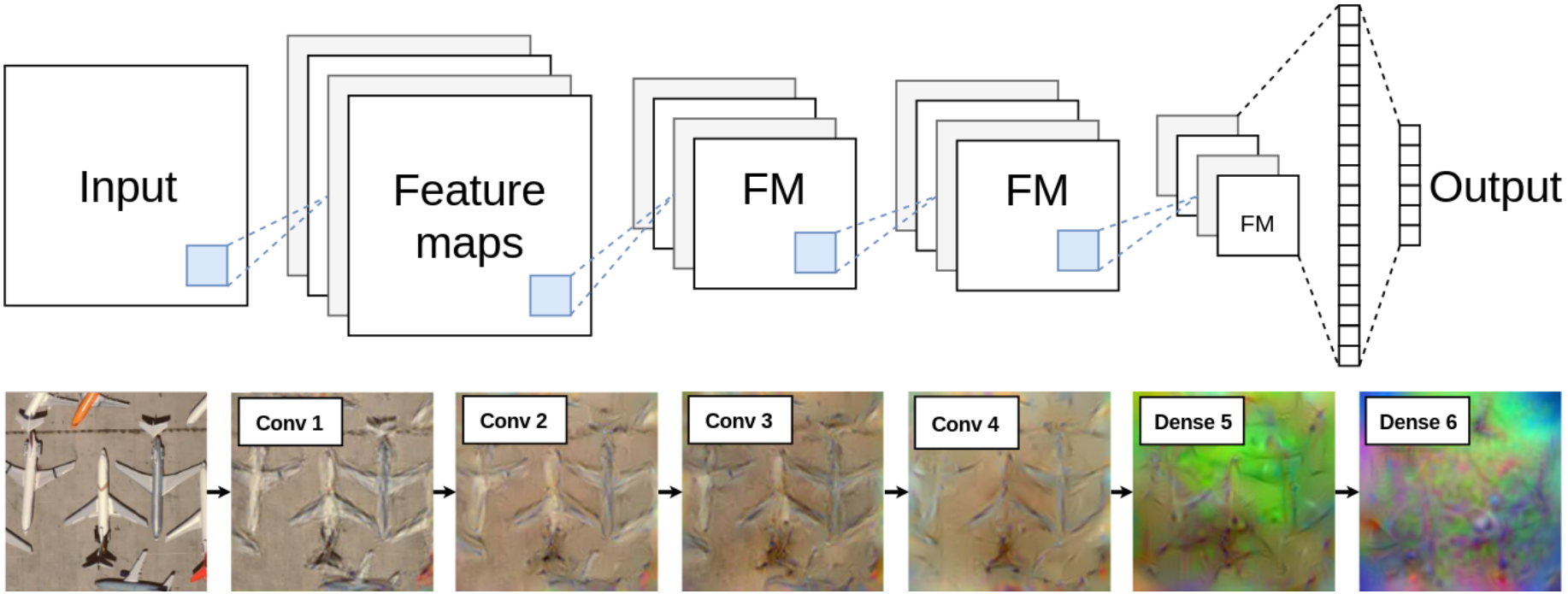
DANIELE GRATTAROLA Author
Prof. MARCELLO RESTELLI Supervisor
Dott. CARLO D'ERAMO Co-supervisor
Dott. MATTEO PIROTTA Co-supervisor

October 3, 2017

TABLE of CONTENTS

DEEP LEARNING

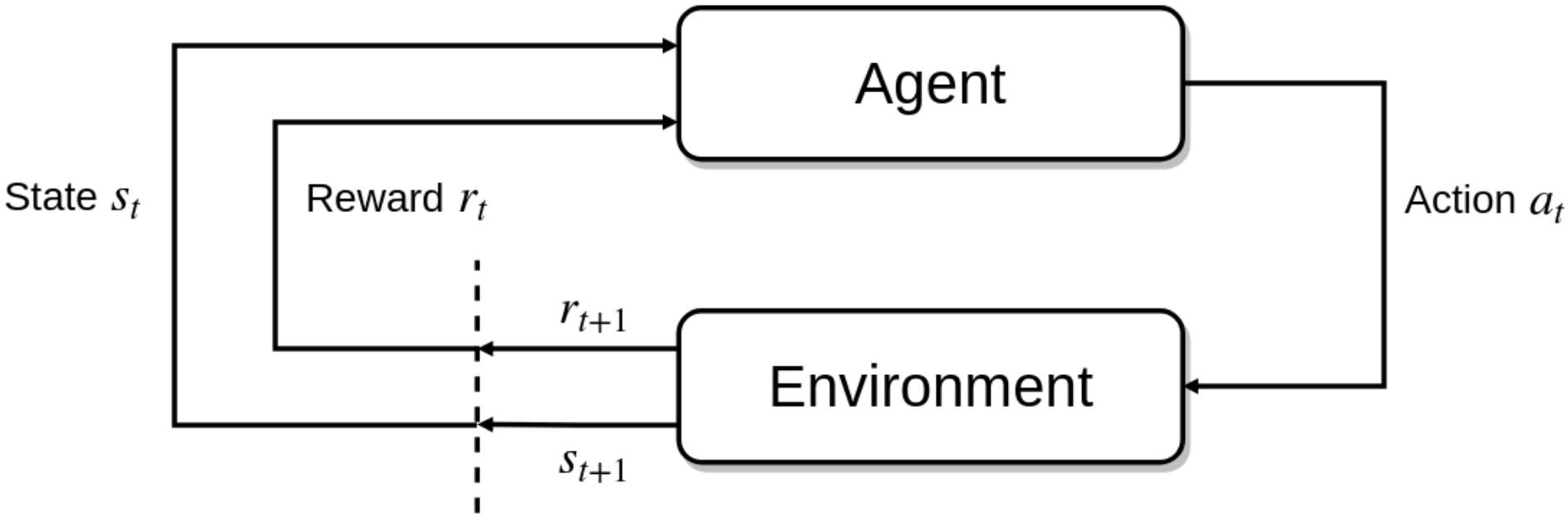
Hierarchical abstractions



Deep learning aims to **learn an abstraction of the input space**, by composing non-linear layers that sequentially transform the representation. (Bengio, LeCun, Hinton – 2015)

REINFORCEMENT LEARNING

Maximizing reward in an environment (Sutton, Barto – 1998)

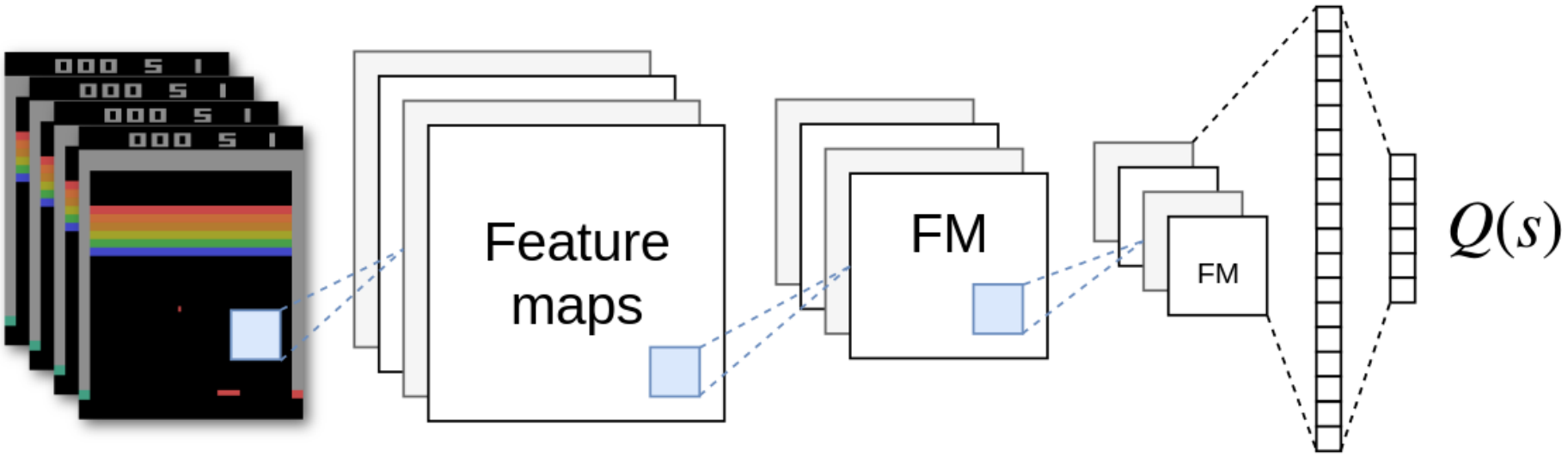


$$Q^{\pi}(s,a) = E_{\pi}[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a]$$

$$= E_{\pi}[r_{t+1} + \gamma Q^{\pi}(s_{t+1}, a_{t+1}) | s_t = s, a_t = a]$$

DEEP REINFORCEMENT LEARNING

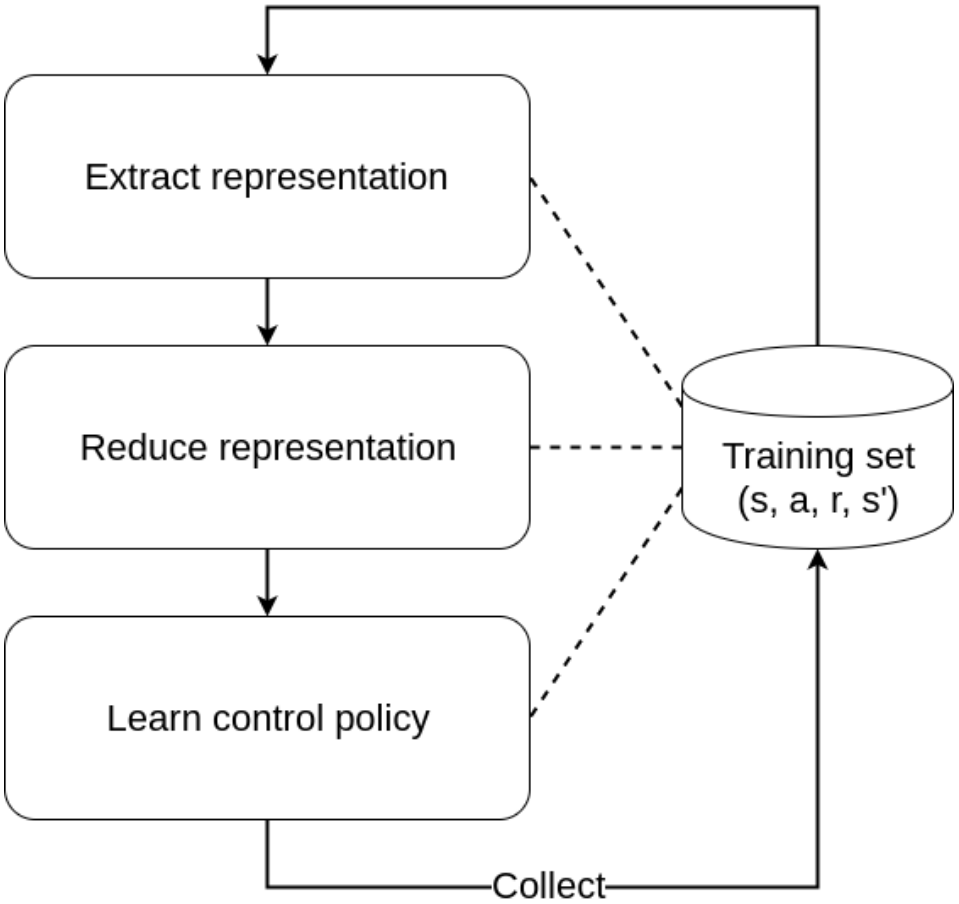
State-of-the-art control using deep representations (Mnih et al. – 2013)



$$Q(s) = [Q(s, a_0), \ldots, Q(s, a_k)]$$

OUR ALGORITHM

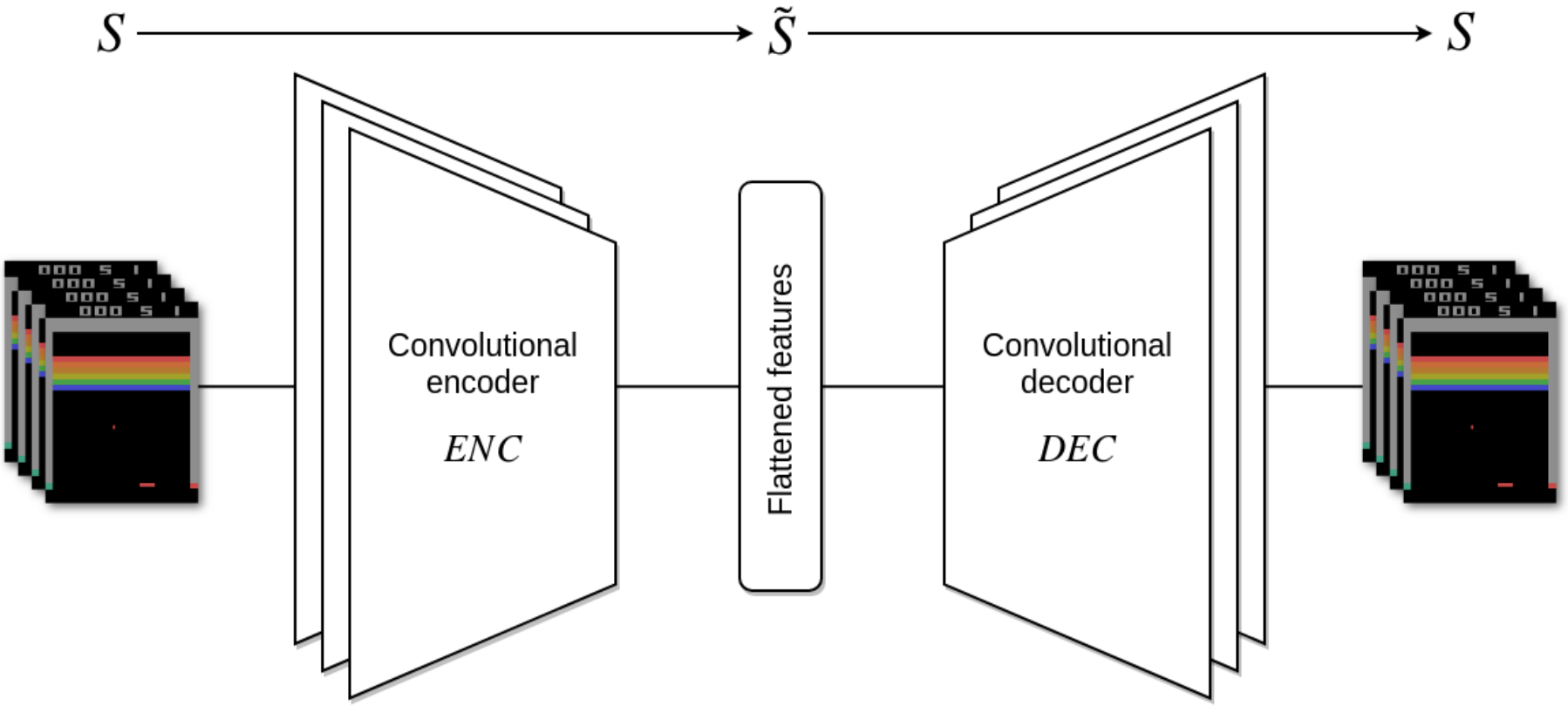
Deep features and batch reinforcement learning



- Goal: use as few samples as possible
- Learn optimal Q function
- Unsupervised feature extraction
- Control-oriented feature selection
- Semi-batch reinforcement learning
- Decreasing exploration rate

OUR ALGORITHM

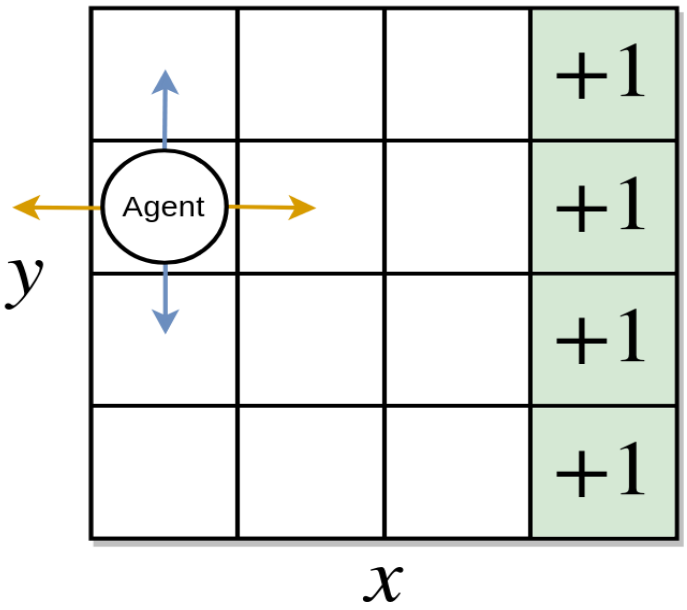
Deep convolutional autoencoder

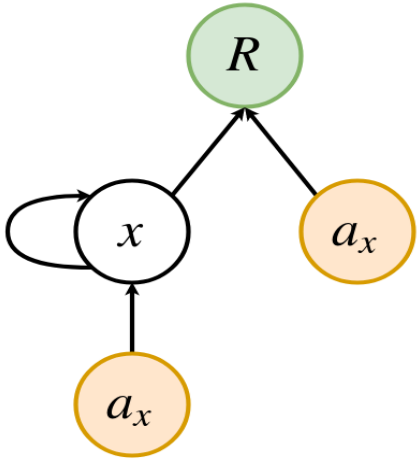


$$ENC : S \rightarrow \tilde{S}$$

OUR ALGORITHM

Recursive Feature Selection (Castelletti, Restelli et al. – 2011)



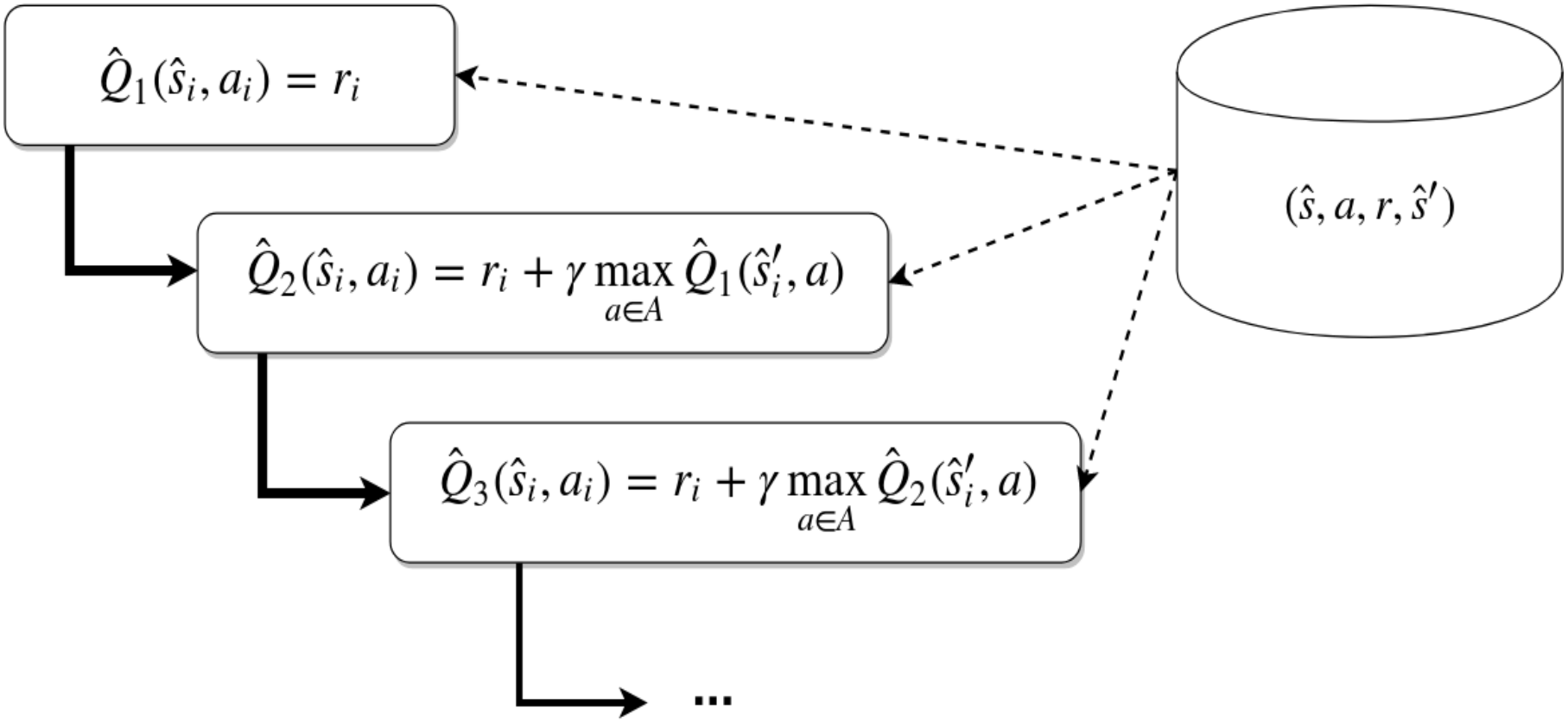


To explain changes in the reward, we only need to know the current value of X and the action that controls it. A similar consideration holds for explaining X.

$$RFS : \tilde{S} \rightarrow \hat{S}$$

OUR ALGORITHM

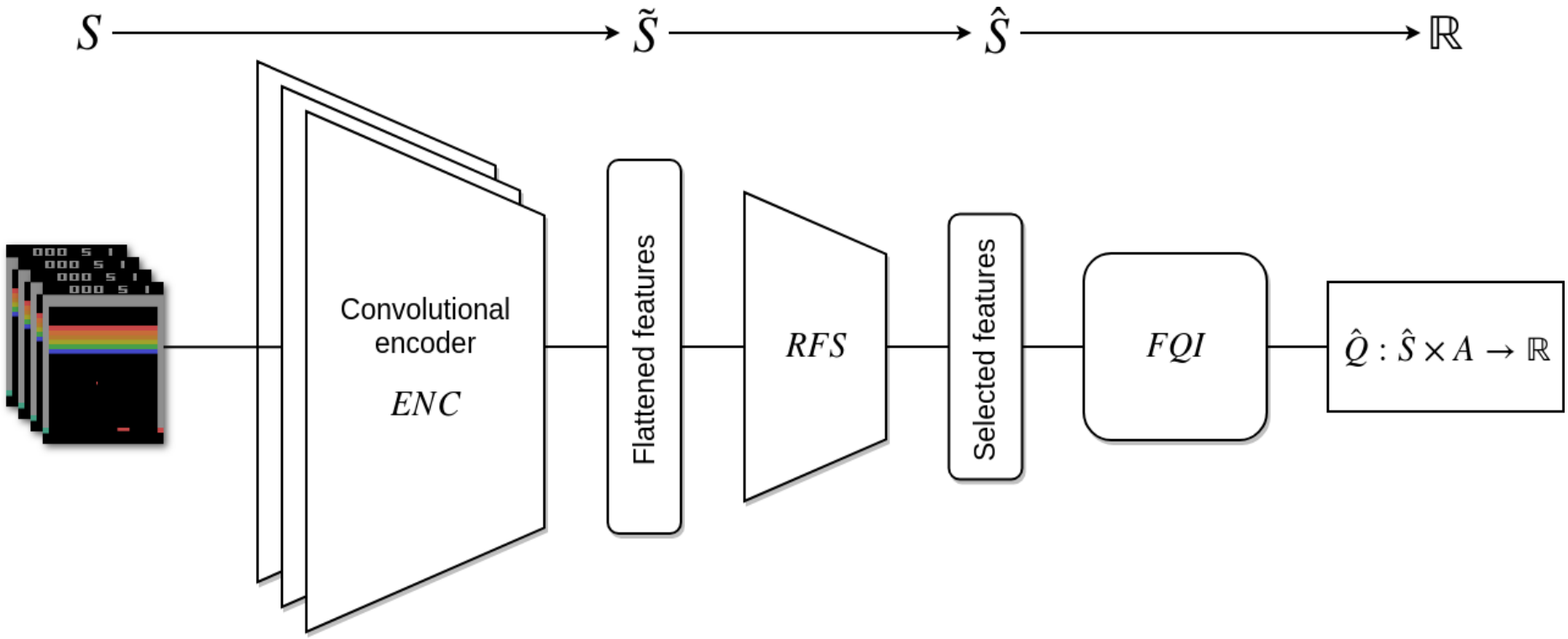
Fitted Q-Iteration (Ernst et al. – 2005)



$$\hat{Q} : \hat{S} \times A \rightarrow \mathbb{R}$$

OUR ALGORITHM

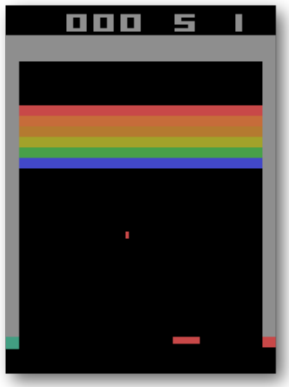
Final agent composition



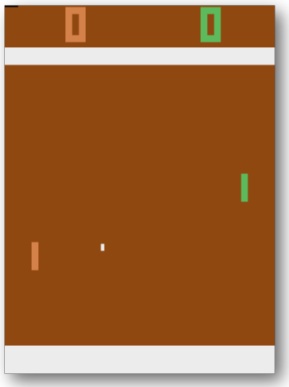
$$Q(s,a) = \hat{Q}(RFS(ENC(s)), a)$$

EXPERIMENTS

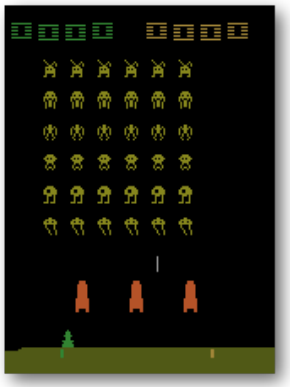
Testing our algorithm on Atari games



Breakout



Pong

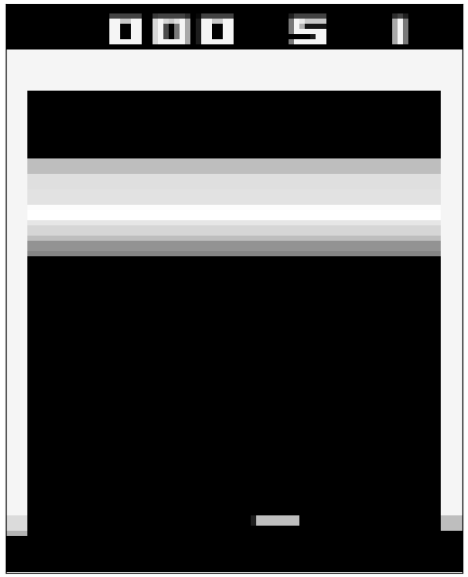


Space Invaders

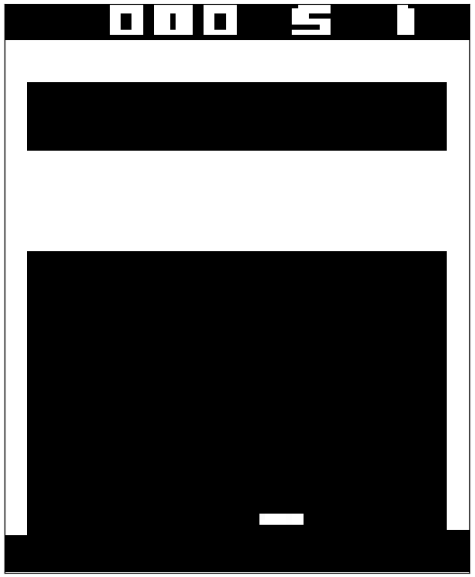
EXPERIMENTS

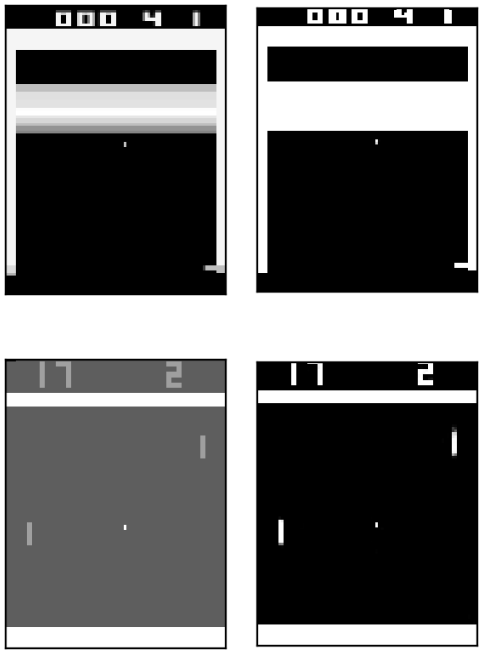
Reconstruction accuracy

Input

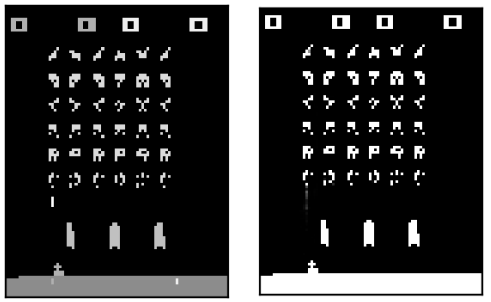


Output



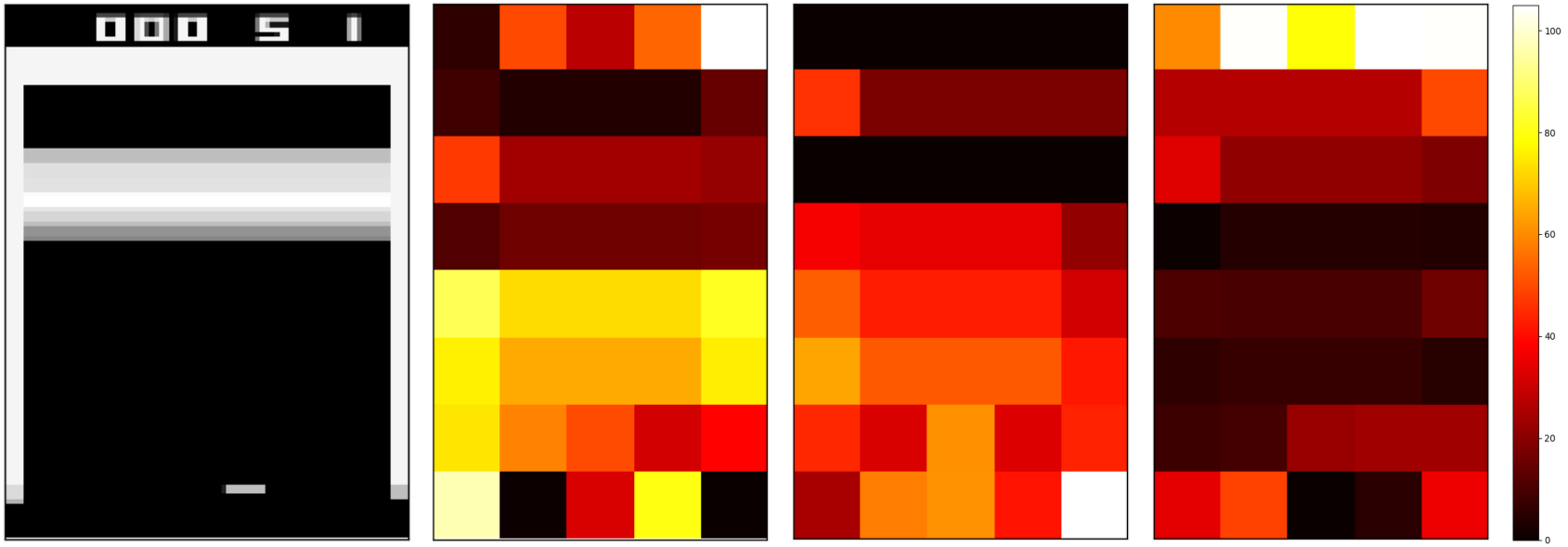


- Almost **100% reconstruction accuracy**
- Reconstruction only lacks sharpness, not elements



EXPERIMENTS

Feature analysis

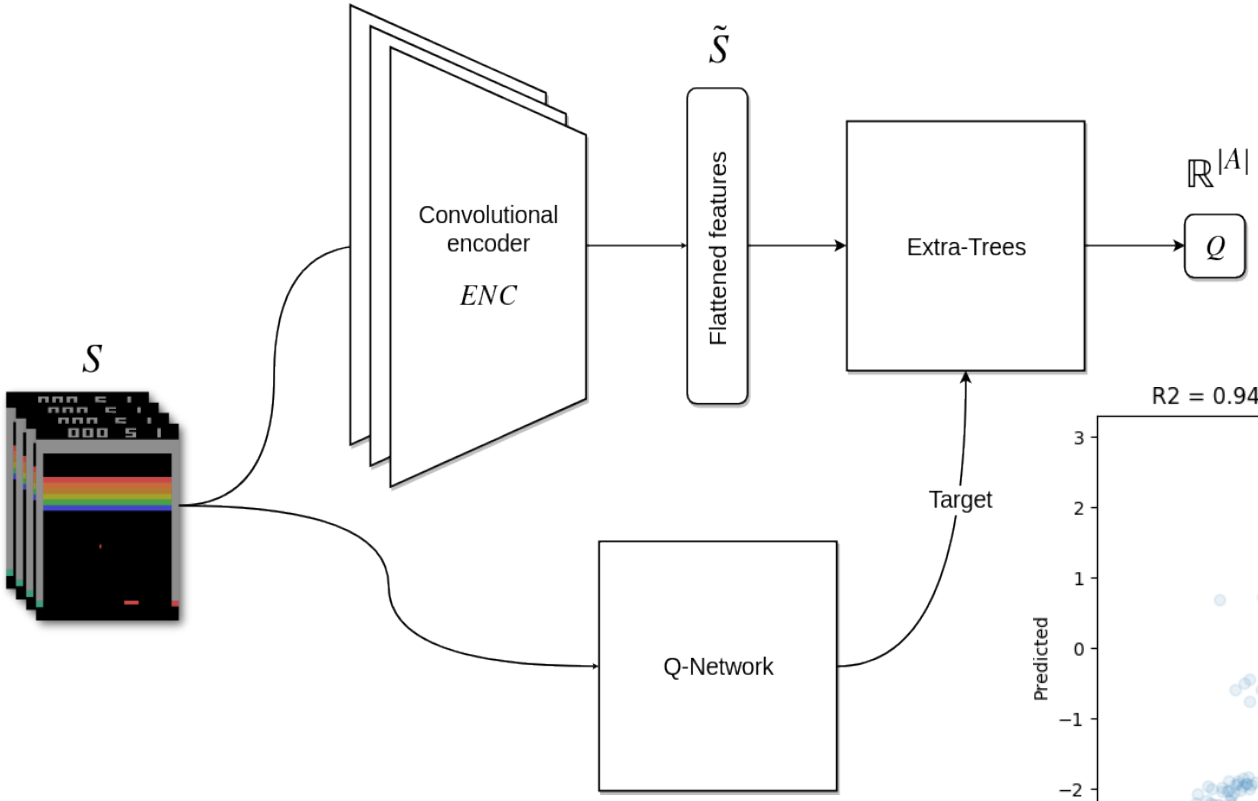


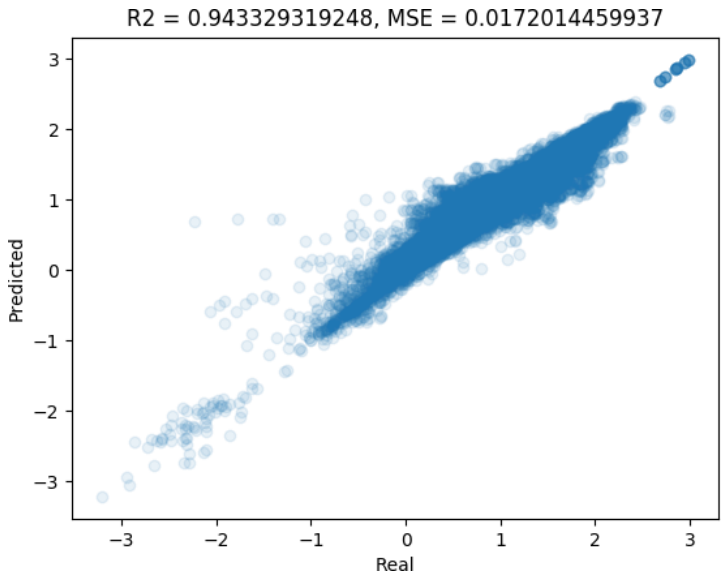
Feature space is **sufficiently abstract**

Feature behavior is **partially interpretable**

EXPERIMENTS

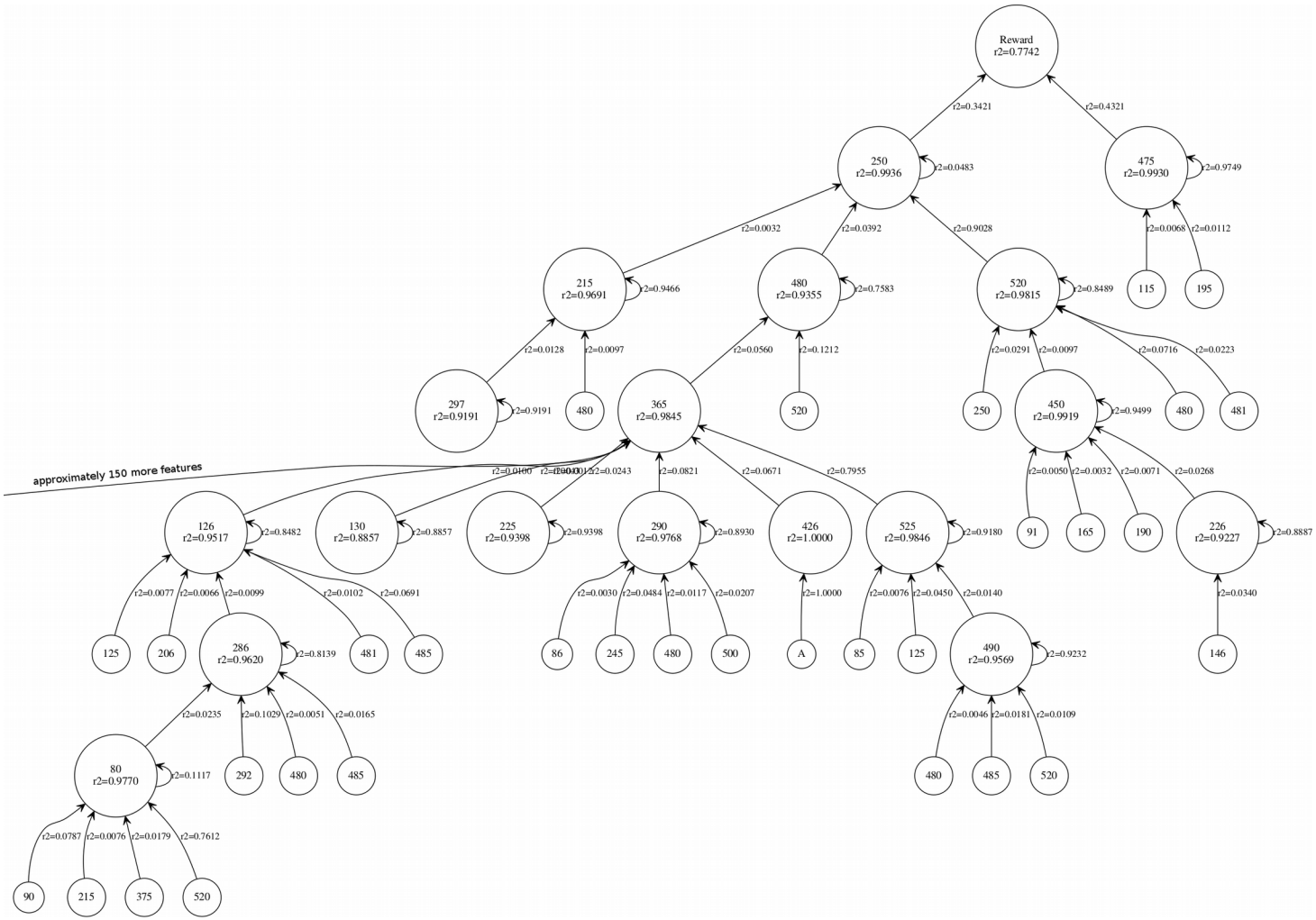
Suitability for control





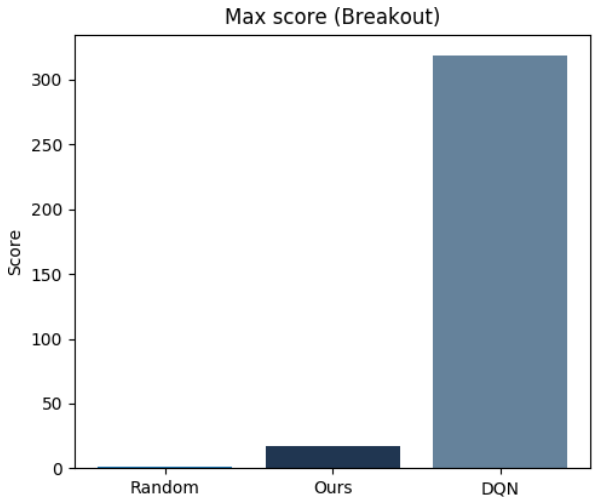
EXPERIMENTS

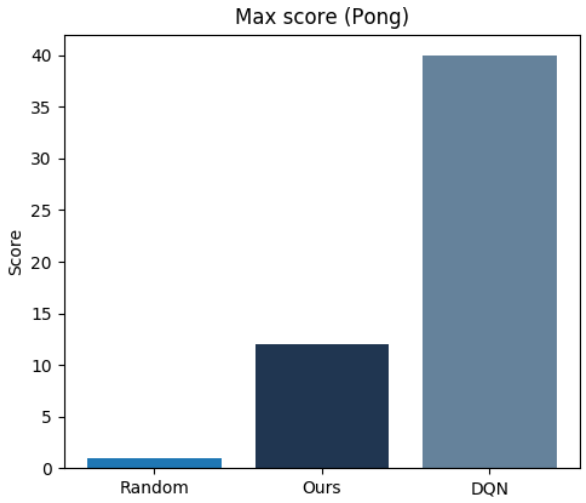
Recursive Feature Selection

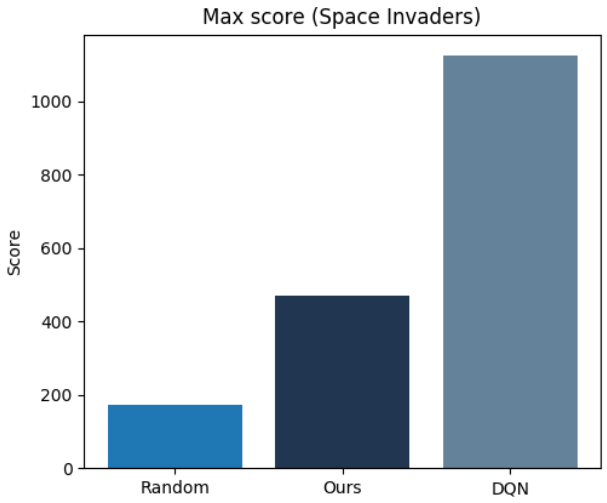


EXPERIMENTS

Main results: top score





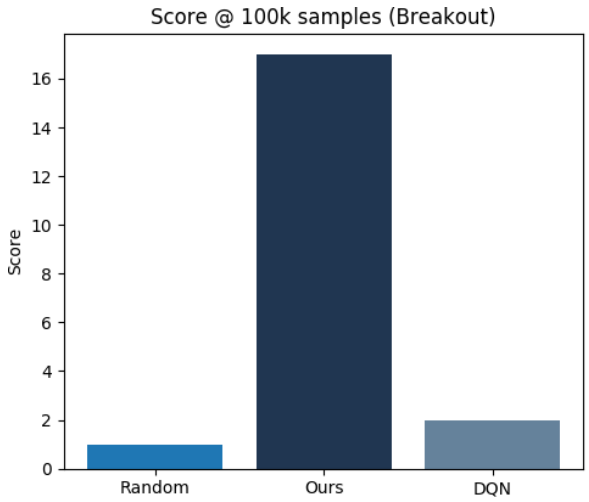


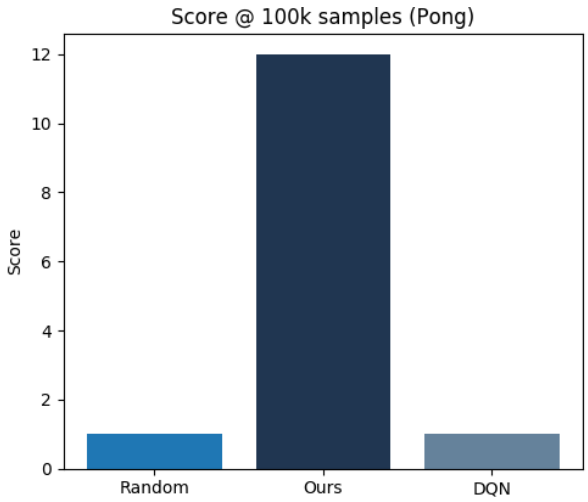
Always learning non-trivial policies that do **better than random**

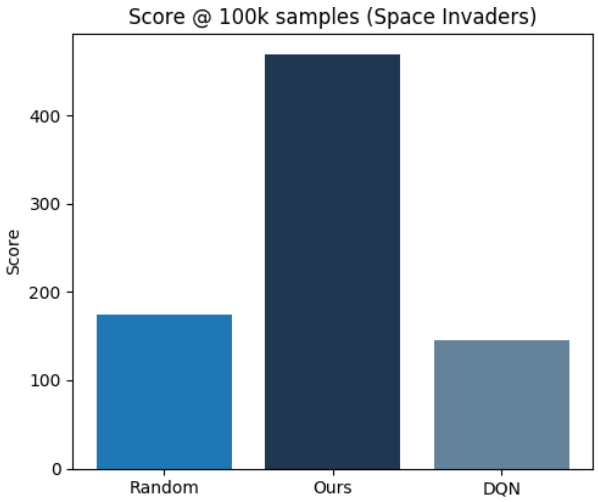
Getting to **25% of DQN's score**, on average (up to 40%)

EXPERIMENTS

Main results: small datasets





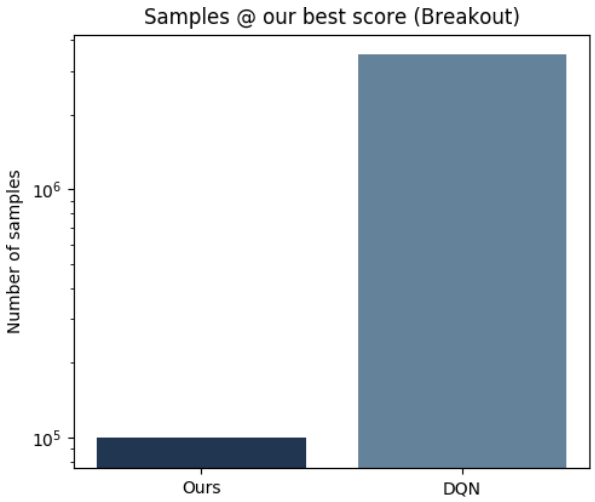


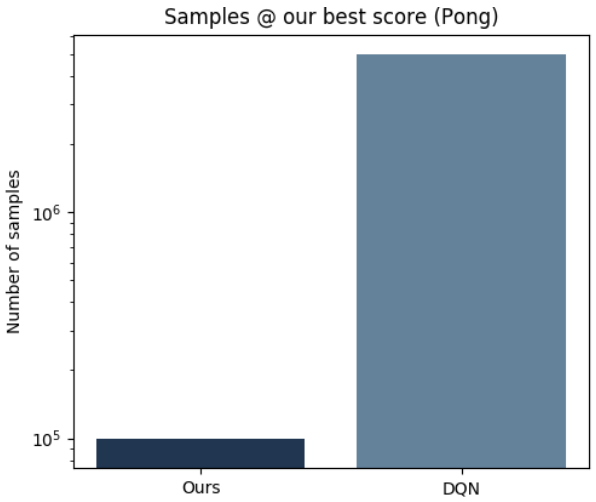
Limit training to **100k collected samples** (minimum amount that we use)

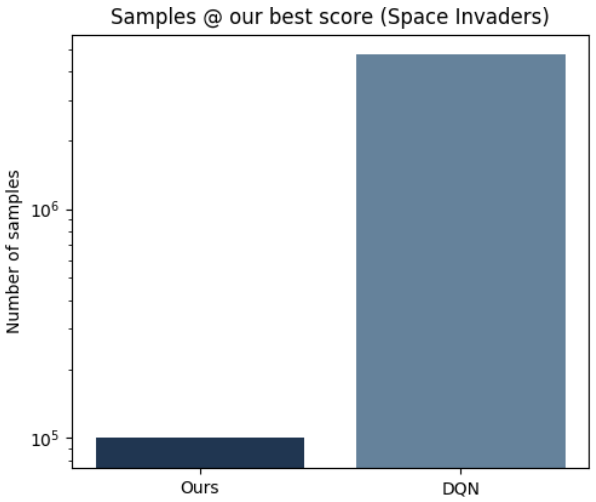
On average **8x better than DQN** (up to 12x)

EXPERIMENTS

Main results: samples





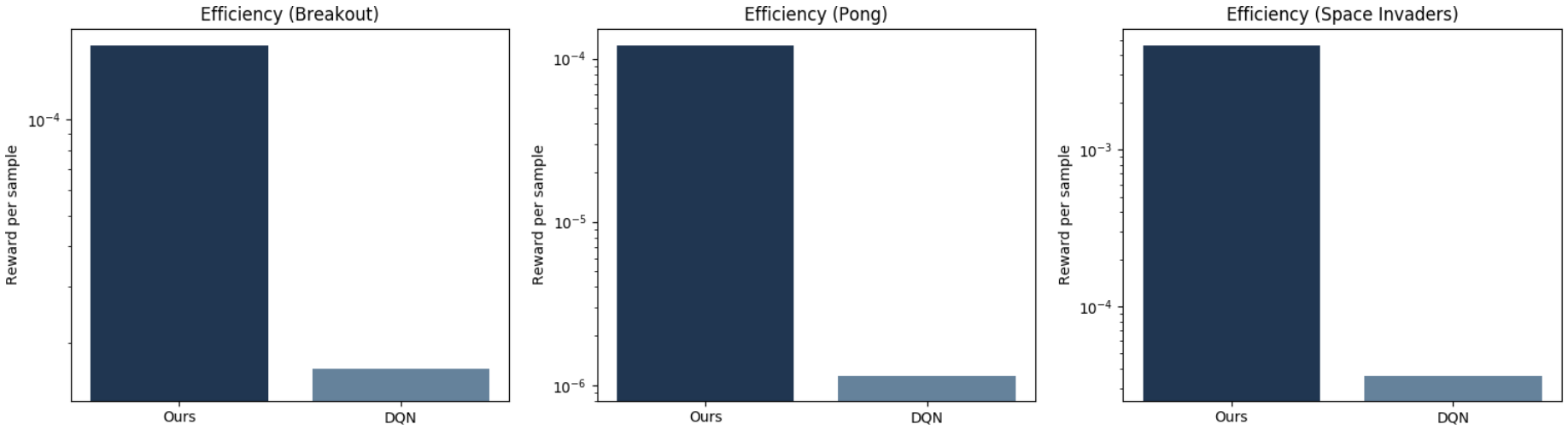


Number of samples required by both algorithms to reach our top score

Almost **50x less samples** required by our agent

EXPERIMENTS

Main results: efficiency



Efficiency in using samples (computed as: max_score / #_training_samples)

Goal achieved: up to **100x more efficient** than DQN

CONCLUSIONS

Recap and future work

100x more **efficient**

25% performance

0.3% training samples

Better on **small datasets**

Improve **exploration**

Improve feature **extraction**

Optimize RFS

Test other **models**

POLITECNICO
MILANO 1863

THANK YOU

DANIELE GRATTAROLA Author
Prof. MARCELLO RESTELLI Supervisor
Dott. CARLO D'ERAMO Co-supervisor
Dott. MATTEO PIROTTA Co-supervisor

October 3, 2017