Experiences running Deep Reinforcement Learning on the IN2P3 GPU cluster

Edward Beeching

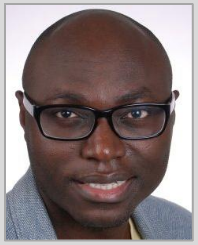
04/04/2019

Overview

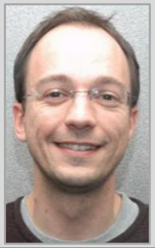
- Brief intro to RL
- Simulators
- Running Deep RL on the in2p3 cluster

Edward Beeching
PhD
INRIA Chroma

Christian Wolf
INRIA Chroma

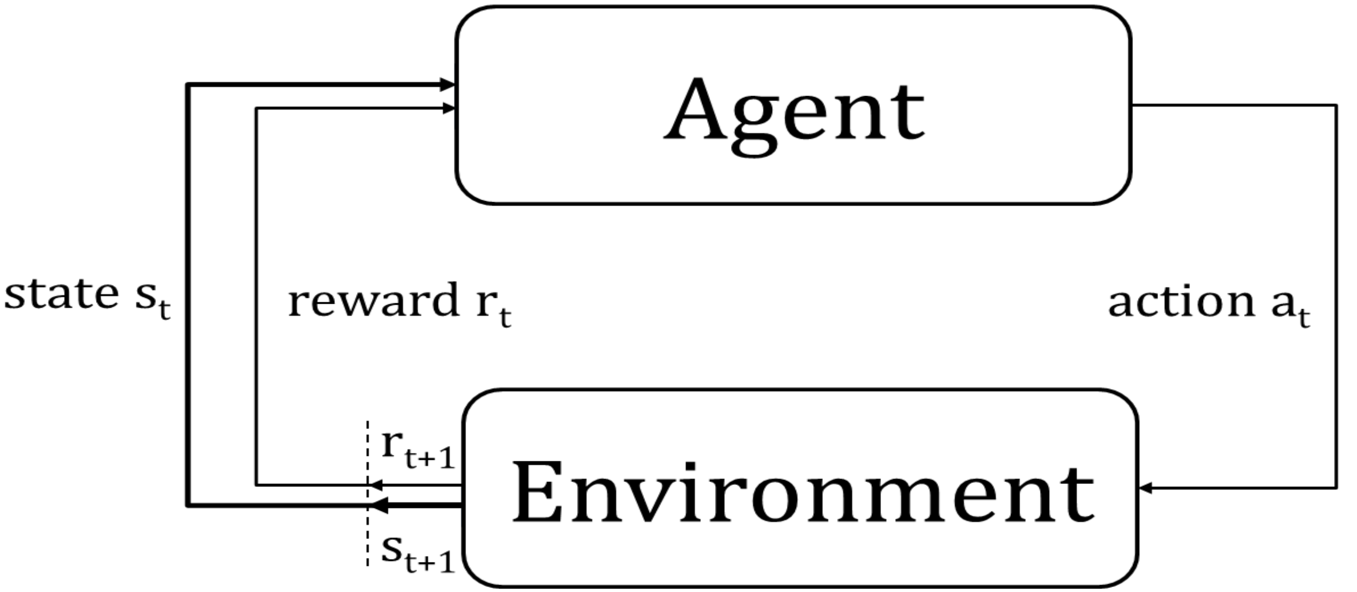


Jilles Dibangoye
INRIA Chroma



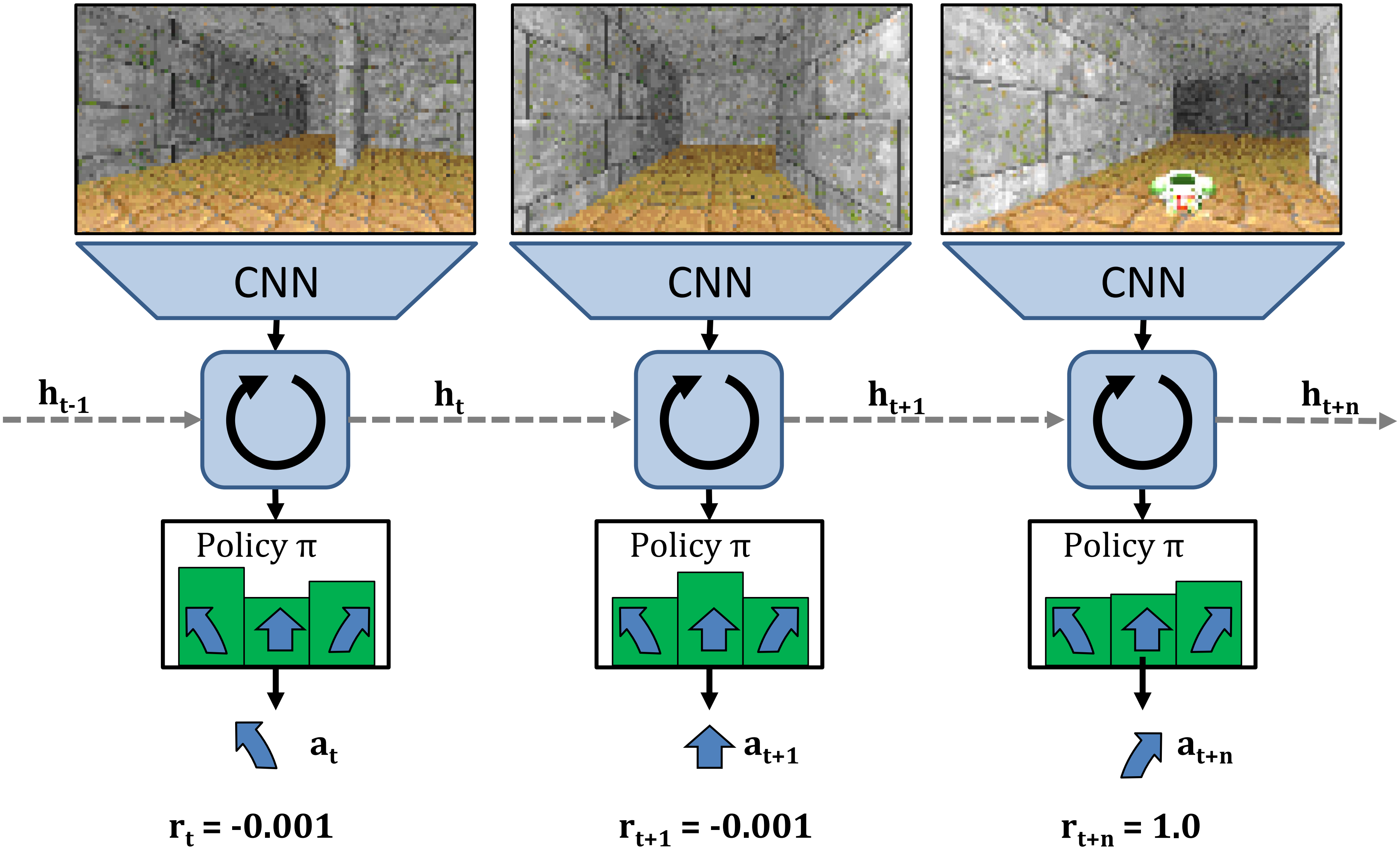
Olivier Simonin
INRIA Chroma

Introduction to RL



- Agent acts in a state based environment.
- **Observable**: Agent receives full state (Board Games, Grid Worlds, Atari Games, Cartpole)
- **Partially observable**: Agent receives a snapshot of the environment. (3D environments, Real time strategy games …)

Introduction to RL: Concrete example



RL Algorithms

Name	Refactored [1]	Recurrent	Box	Discrete	Multi Processing
A2C	Advantage Actor Critic				
ACER	✔	✔	× [5]	✔	✔
ACKTR	✔	✔	× [5]	✔	✔
DDPG	✔	×	✔	×	×
DQN	✔	×	×	✔	×
GAIL [2]	✔	✔	✔	✔	✔ [4]
PPO1	✔	×	✔	✔	✔ [4]
PPO2	✔	✔	✔	✔	✔
SAC	✔	×	✔	×	×
TRPO	✔	×	✔	✔	✔ [4]

https://stable-baselines.readthedocs.io

Simulators and training RL agents

Scenarios for Deep RL: Simulators

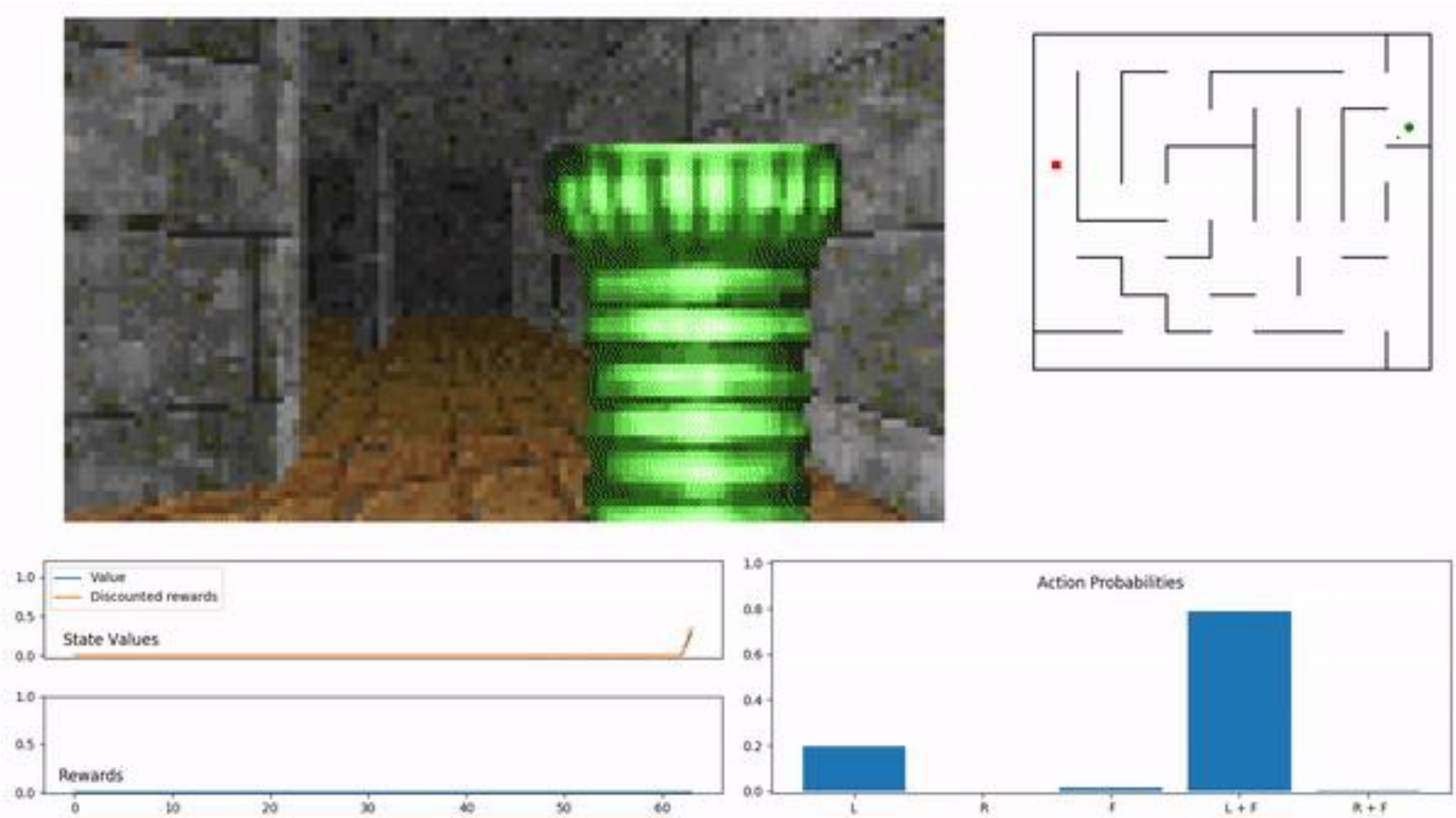
ViZDoom, Kempka et al. 2016

DeepMind lab Beattie et al. 2016

House3D, Wu et al. 2018

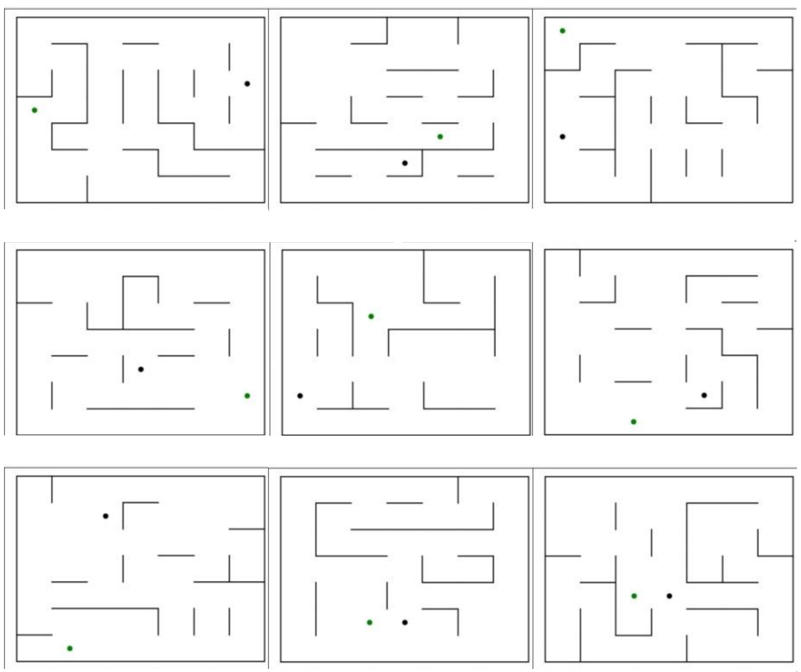
Habitat, Savva et al. 2018

Find and Return scenario

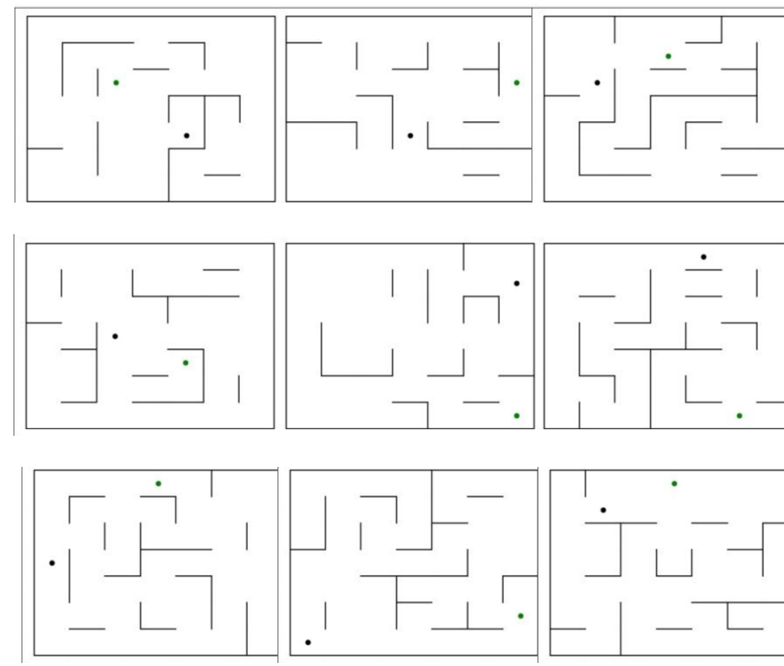


Generalization

N - Training configurations

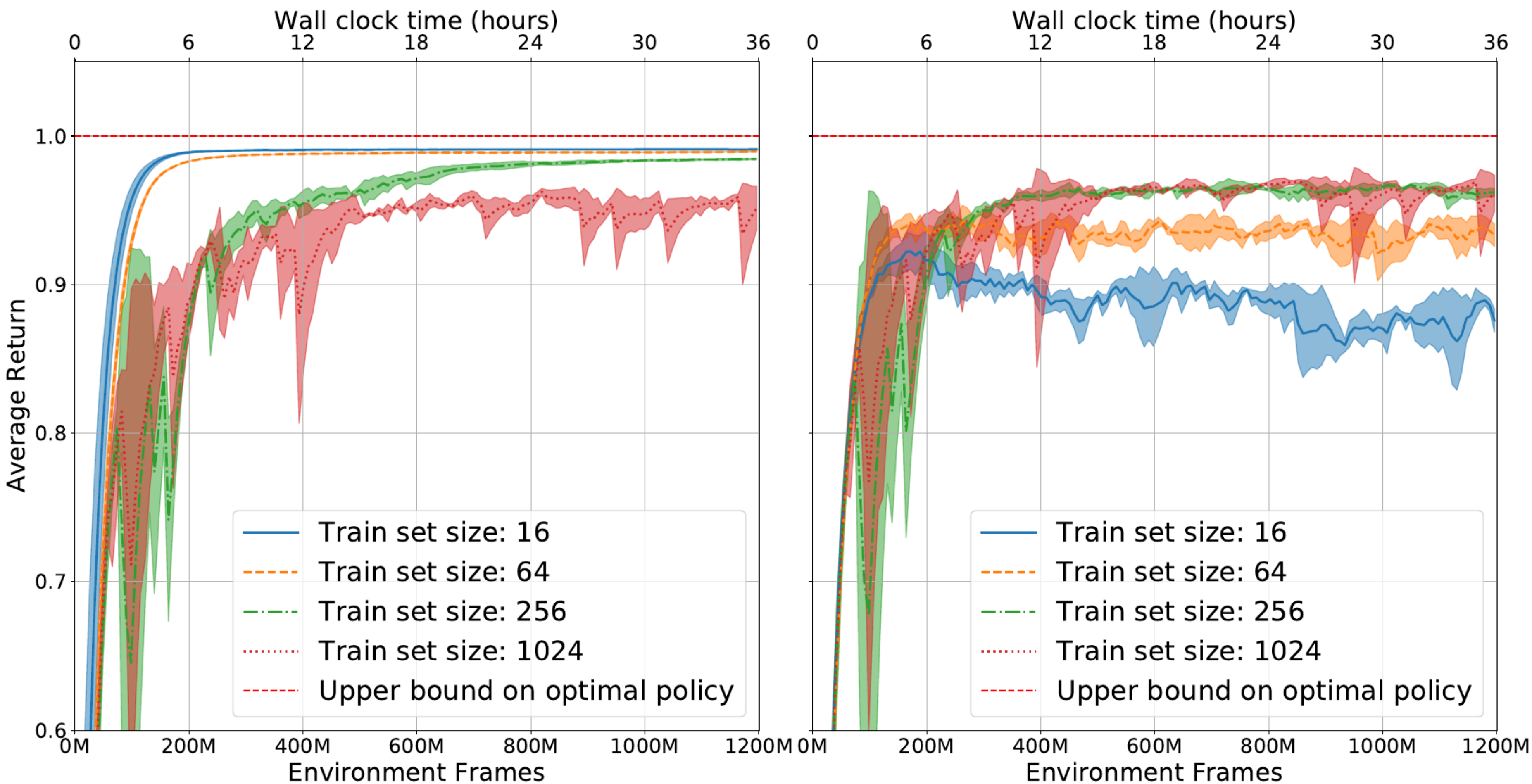


K - Testing configurations



How many training configurations are required to generalize to unseen test configurations?

Generalization results



Running Deep RL on in2p3

Running Deep RL on in2p3

- Singularity images
- PyTorch 1.0
- Tested simulators:
  - ViZDoom (Kempka et al. 2016)
  - House3D (Wu et al. 2018)
  - Habitat (Savva et al. 2019)
- Algorithms
  - Advantage Actor Critic (A2C) (Mnih et al. 2016)
  - Proximal Policy Optimization (PPO) (Schulman et al. 2017)
  - Deep Q-Learning (DQN) (Mnih et al. 2015)

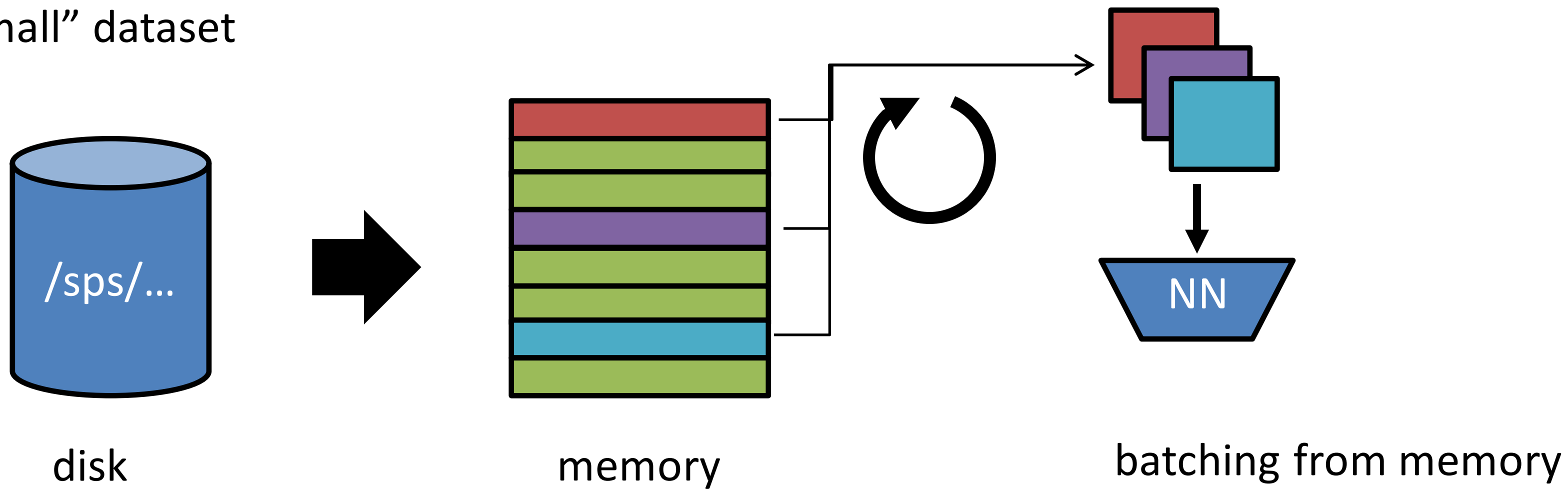
Recap

- Supervised learning
  - A set of pairs (x, y) input/output
  - Strong supervision in the form of labels

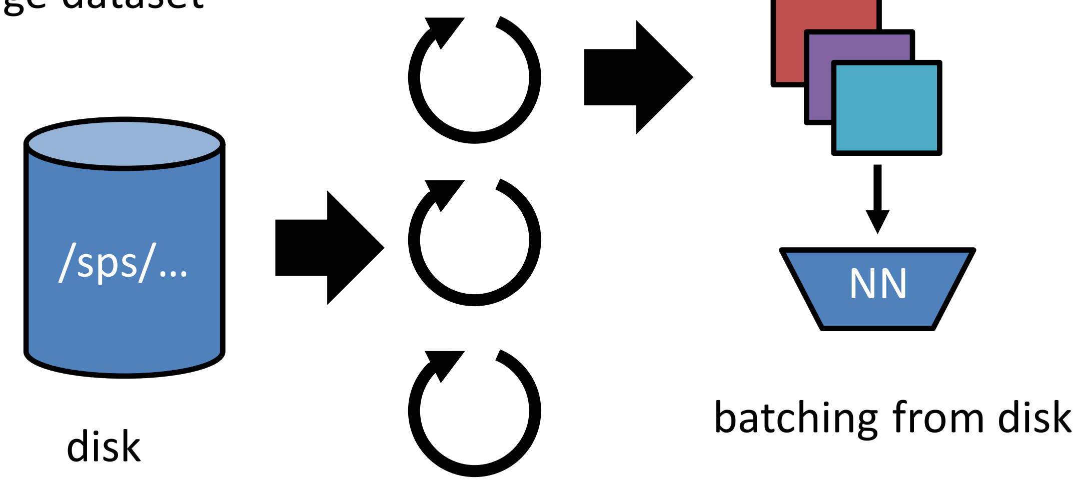
- Reinforcement learning
  - Learning to optimize cumulative reward through interaction in a (black box) environment.
  - Extremely noisy gradient signal

Supervised learning

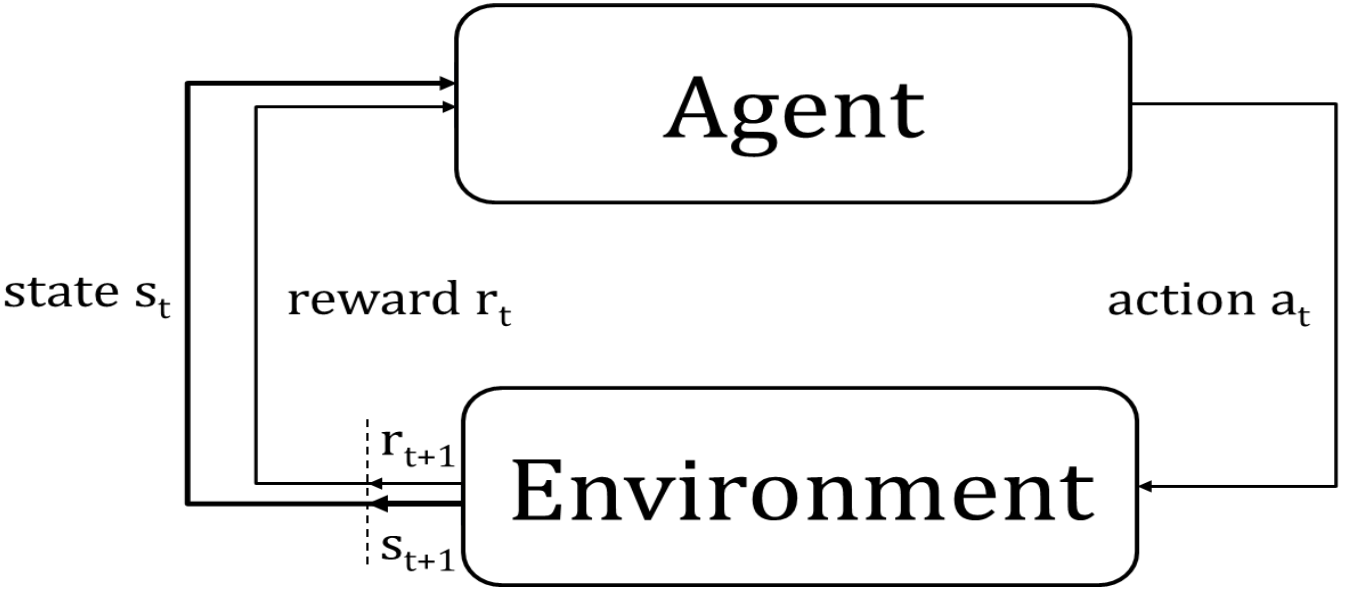
“small” dataset



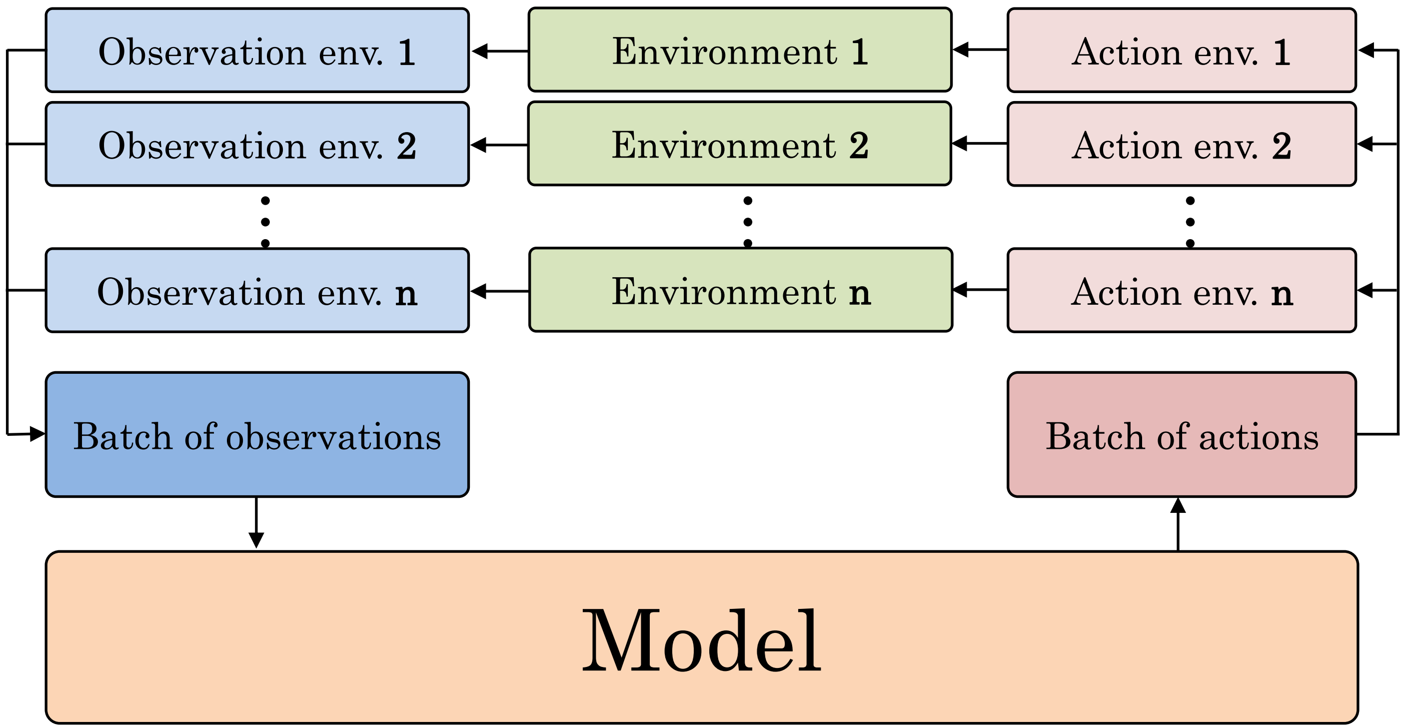
large dataset



Reinforcement Learning



Batching of simulations



Job submission

```bash
#!/bin/bash

echo "Running on $HOSTNAME"
echo "job id : $JOB_ID"
echo "starting"

if ! echo ${LD_LIBRARY_PATH} | /bin/grep -q /opt/cuda-9.2/lib64 ; then
    LD_LIBRARY_PATH=/opt/cuda-9.2/lib64:${LD_LIBRARY_PATH}
fi

TESTNAME="habi_ppo_026_$JOB_ID"

mkdir -p data/checkpoints/$TESTNAME
PARAMS="--use-gae --sim-gpu-id 0 --pth-gpu-id 0 --lr 2.5e-4 --clip-param 0.1 --value-loss-coef 0.5 --num-processes 32 --num-steps 128 --num-mini-batch 4 --num-updates 400000 --entropy-coef 0.01 --log-file logs/$TESTNAME-train.log --log-interval 1 --checkpoint-folder data/checkpoints/$TESTNAME --checkpoint-interval 1000 --task-config tasks/pointnav_gibson_64.yaml --num_gpu 4 --hidden-size 512"

BINDS='--bind /pbs/home/b/beeching/work/:/home/edward/work/ --bind /sps/liris/beeching/storage/:/home/edward/storage/'

IMG='/sps/liris/beeching/sing/sing_habi_no_docker/habi_no_docker.img'
HH='--home /pbs/home/b/beeching/work/habitat_comp'

export SINGULARITYENV_MAGNUM_LOG="quiet"
export SINGULARITYENV_GLOG_minloglevel=2

echo $PARAMS
singularity exec --nv --writable $HH $BINDS $IMG python baselines/train_ppo.py $PARAMS

echo "finishing"
```

Conclusions

- 917 GPU jobs run on the cluster since October.
- Singularity images provide an easy and flexible way to manage dependencies.
- Thanks to the in2p3 cluster team for their working maintaining and running the cluster.