CERiL: Continuous Event-based Reinforcement Learning

CVSSP Centre for Vision, Speech and Signal Processing

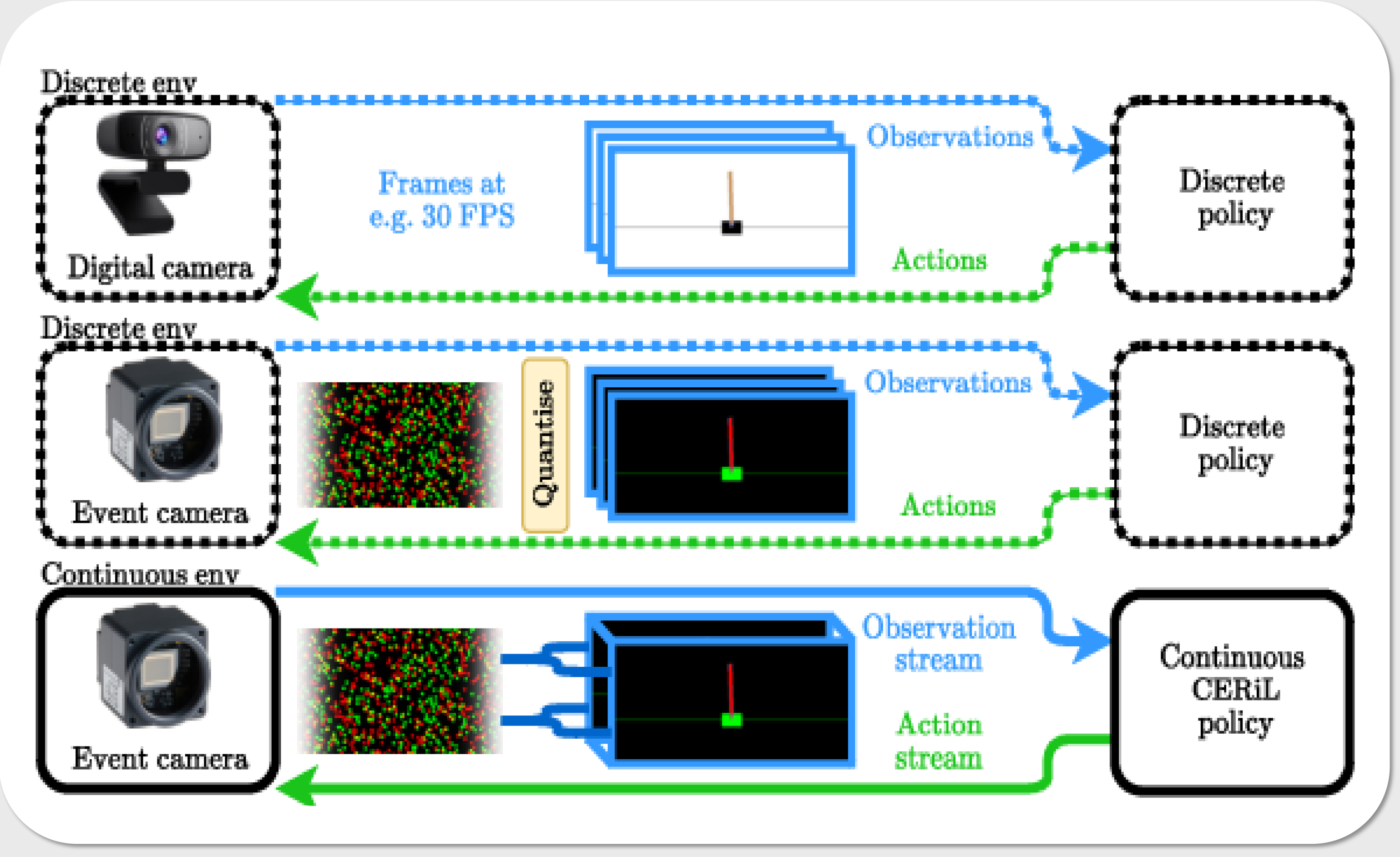
Celyn Walters (celynwalters@gmail.com) Simon Hadfield (S.Hadfield@surrey.ac.uk)

UNIVERSITY OF SURREY

1 - Abstract

- A new approach for **reinforcement learning** with **event cameras.**
- Operates **directly** on the event stream using **EDeNNs** (see our oral @BMVC)
  - No intermediate aggregation
  - **Continuous** input observation stream and output action stream
- **Wrapper code** for simple application to any "Gymnasium" style RL environment

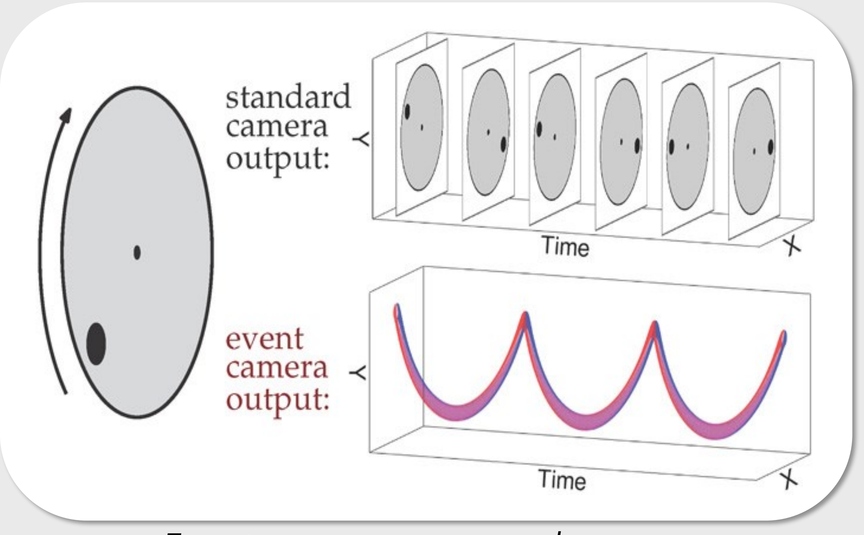
Code available



Traditional vision based RL, vs discrete event based RL, vs CERiL

2 - Background - Event Cameras

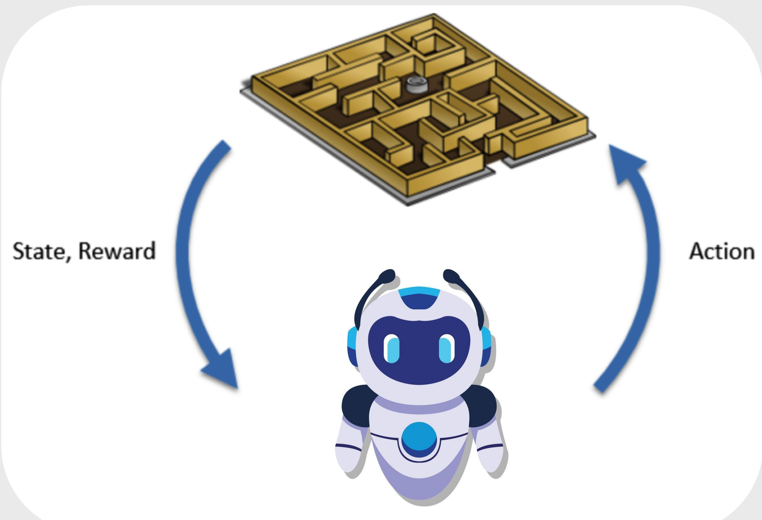
- Event cameras are **asynchronous** visual sensors
- Brightness changes cause **immediate** signals from the sensor, with no shutter based poling
- Numerous advantages: low-power, low bandwidth, high dynamic range, and low **world-to-sensor** latency
- Disadvantage: **No images**, unclear how to apply traditional computer-vision tools



Event camera vs normal camera

3 - Background - Reinforcement Learning

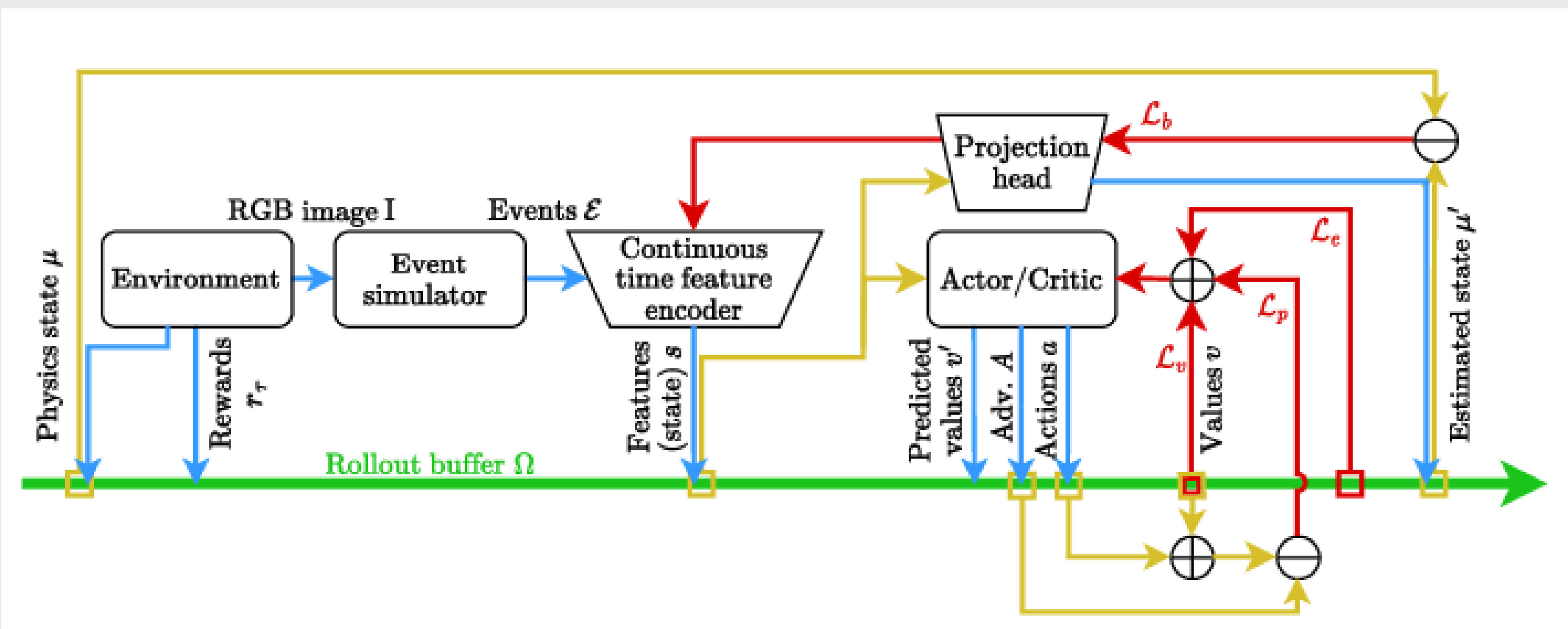
- Long horizon **strategic** machine learning (e.g. game playing)
  - **No ground truth** right or wrong answers
  - **Maximize reward** function across game
- Traditionally iterative process
  - Agent chooses actions based on current environmental state
  - Environment executes actions and returns new state



Reinforcement learning process

4 - CERiL overview

- **Modules** asynchronously **insert** items to the rollout buffer
- **Losses query** the buffer and are computed on any entries found
- **Environment** logs **rewards** and **states**
- **Encoder** logs **feature volumes** based on simulated events
- **Actor/Critic** logs **actions**, **values** and **advantage** functions
- **Projection head** logs estimated **state**



CERiL system flow. Blue arrows are insertions into the rollout buffer. Yellow arrows are extractions. Red are losses.

5 - CERiL Details

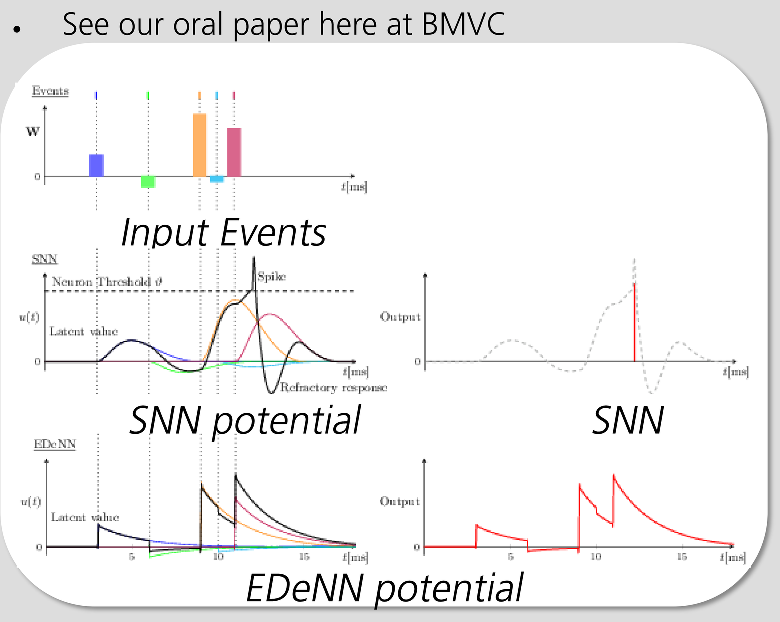
5.1 - Continuous rollout generation

- Generic **wrapper** for OpenAI **Gym** environments
  - Render environment every step
  - Event-Camera Simulator (**ESIM**) turns discrete renders into continuous-time event stream

ESIM converts environment renders to events

5.2 - Continuous feature encoding

- Event Decay Neural Network (**EDeNN**) on events
  - Specialised spatio-temporal convolution
  - CNN style **spatial convolution** kernel
  - SNN style **temporal decay** (learned per neuron)
  - Dense feature encoding from sparse events
- See our oral paper here at BMVC

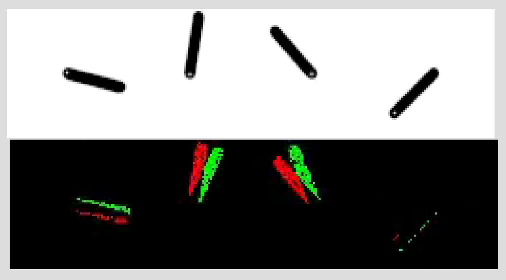


5.3 - Losses

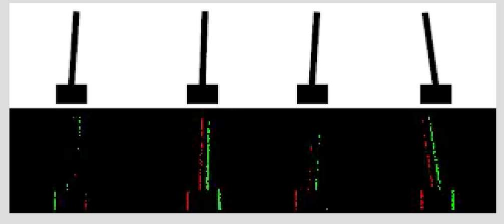
- **Projection head loss:** regularises vision system
  - Requires that states are recoverable from features
- Continuous variant of Proximal Policy Optimisation
  - **Policy loss**: integral of clipped advantage function
  - **Critic loss:** integral of critic/reward disagreement
- Evaluated at discrete times based on control loop speed.

6 - Evaluation

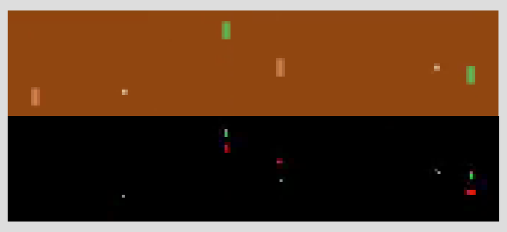
6.1 - Environments



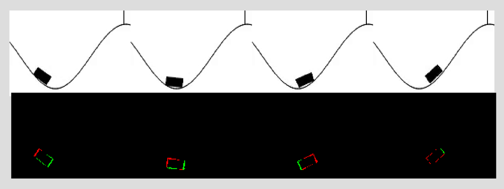
- **Pendulum:** Apply torque to swing up & balance pole
- **Dense Reward** = pole verticality angle - torque use



- **CartPole.** Use linear actuated cart to balance inverted pole
- **Keep-alive reward** = +1 per step pole is upright



- **Pong.** Move a paddle up and down to deflect bouncing balls past an autonomous opponent paddle
- **Sparse reward** = +1 if agent wins a round, -1 if it loses



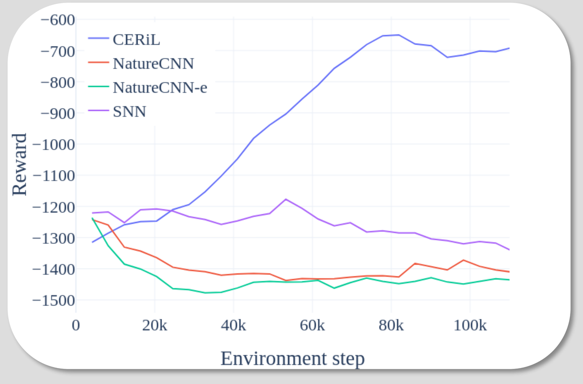
- **MountainCar.** Oscillate linear actuated cart to escape valley
- **Terminate ASAP reward** = -1 for every step in episode

6.2 - State-of-the-art comparison

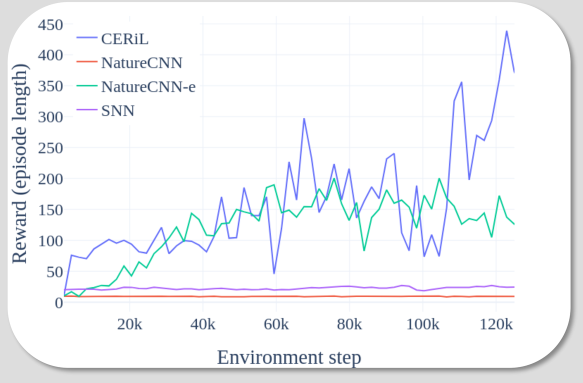
- Compared against SOTA visual RL algorithms (**CNNs** and **SNNs**) on **events** and **RGB**

Approach	Input data format	Average rewards Pendulum	CartPole	Pong	MountainCar
NatureCNN [9]	RGB	−1242.2	9.4	**17.9**	−200.0
NatureCNN-e [9]	2D event image	−1236.6	137.4	15.0	−200.0
SNN [6]	Event stream	−1177.1	87.3	−17.2	−200.0
CERiL (ours)	Event stream	**-638.7**	**438.8**	1.0	**-97.6**

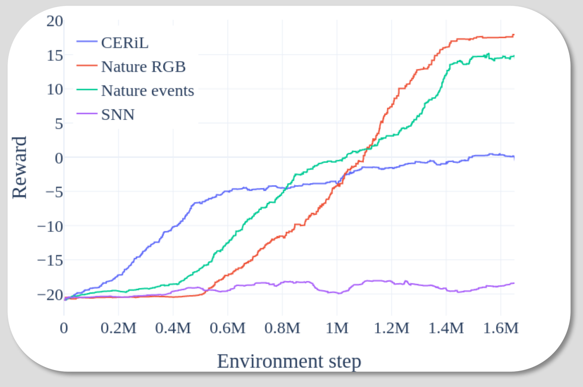
Average rewards of different RL techniques



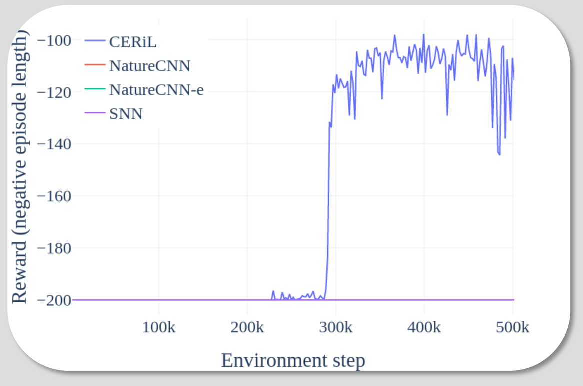
Pendulum training curves



CartPole training curves



Pong training curves



Mountaincar training curves

- CERiL performs very favourably compared to all other visual RL approaches
- Pong is challenging: long term planning required vs short term event aggregation
- **Only CERiL** can solve **MountainCar**

Funded by EPSRC grant "Reflexive robotics" (EP/S035761/1)

EPSRC Engineering and Physical Sciences Research Council