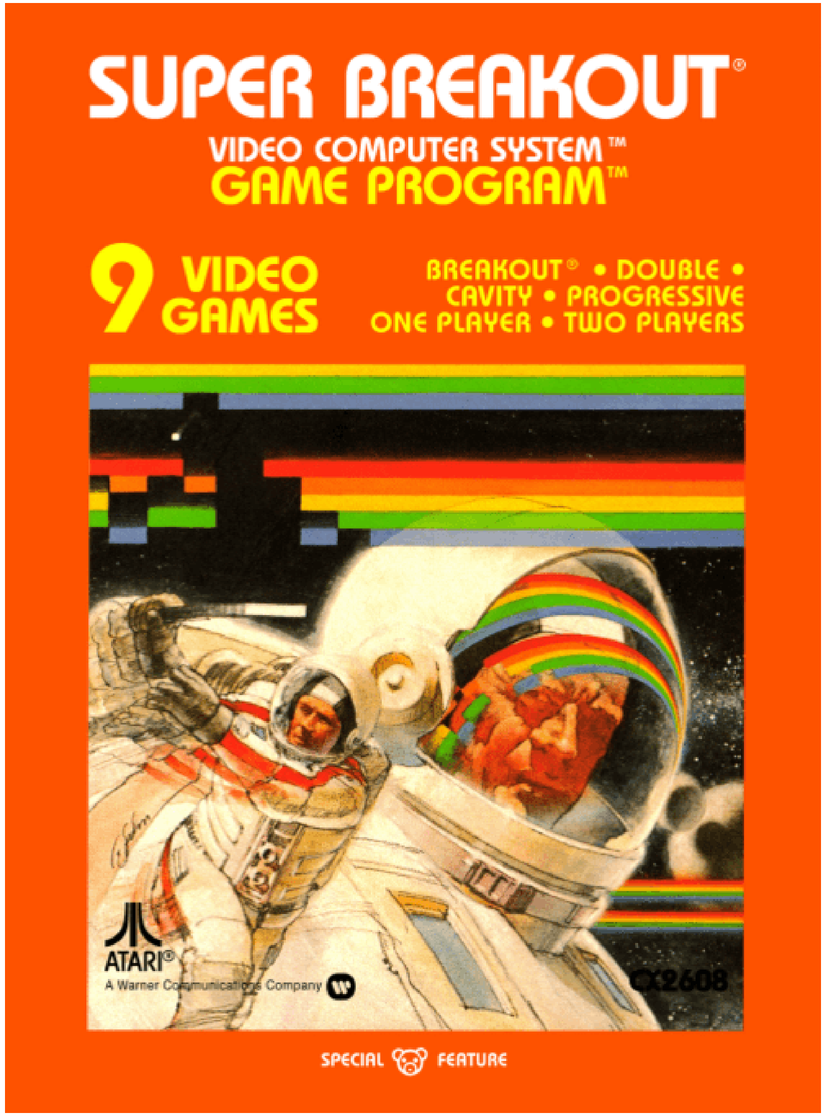
RL for POMDPs

4 Approaches

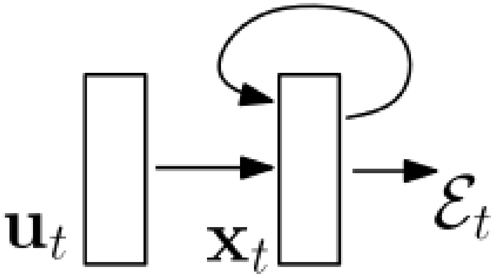
Approach 1: k-Markov

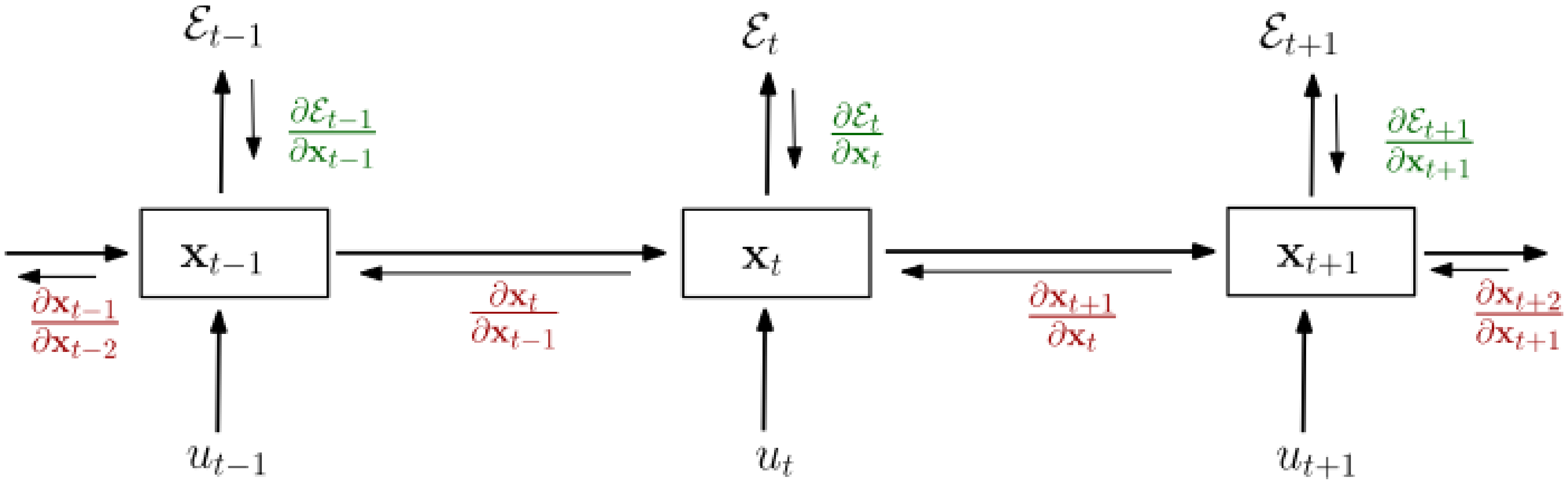


Approach 2: RNN (e.g. LSTM)

$$\mathbf{x}_t = F(\mathbf{x}_{t-1}, \mathbf{u}_t, \theta)$$

$$\mathbf{x}_t = \mathbf{W}_{rec}\sigma(\mathbf{x}_{t-1}) + \mathbf{W}_{in}\mathbf{u}_t + \mathbf{b}$$





Input: u_t

State/Output: x_t

Cost: $\mathcal{E}_t$

e.g. Alpha Star

Images: Pascanu 2013

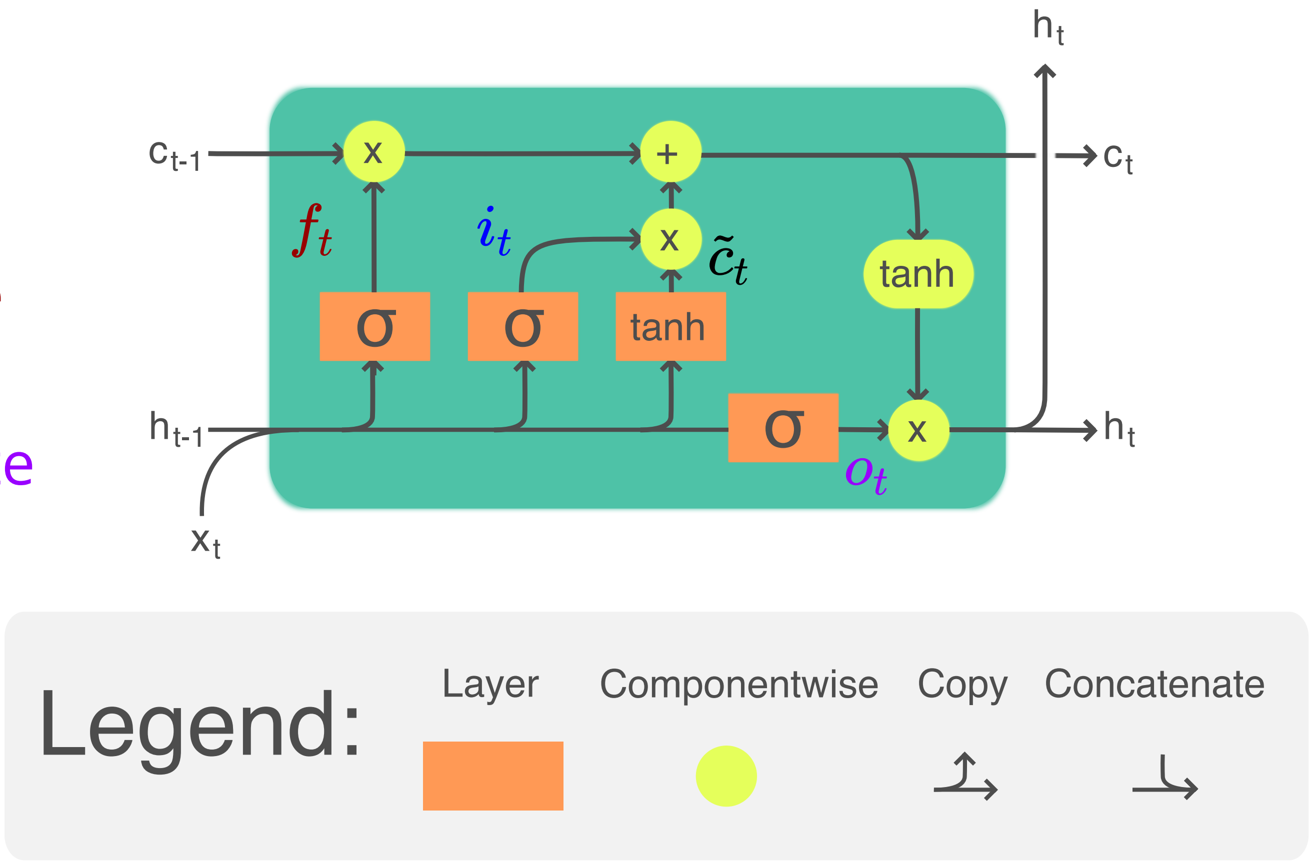
Approach 2: RNN (e.g. LSTM)

Input: x_t

Output: h_t

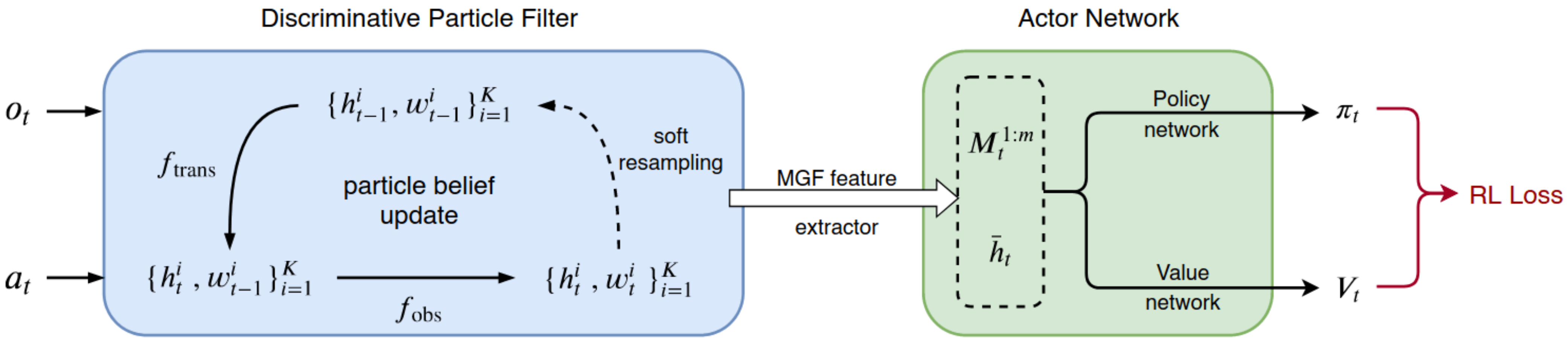
Cell state: c_t

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad \text{Forget gate}$$
$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad \text{Input Gate}$$
$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad \text{Output Gate}$$
$$\tilde{c}_t = \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$
$$h_t = o_t \odot \sigma_h(c_t)$$



Hochreiter, S. and Schmidhuber, J. (1997). *Long Short-Term Memory*. *1780*, 1735–1780.

Approach 3: Particle Filters and MGF Features



$$M_{\mathbf{X}}(\mathbf{v}) = \mathbb{E}\left[e^{\mathbf{v}^\top \mathbf{X}}\right]$$

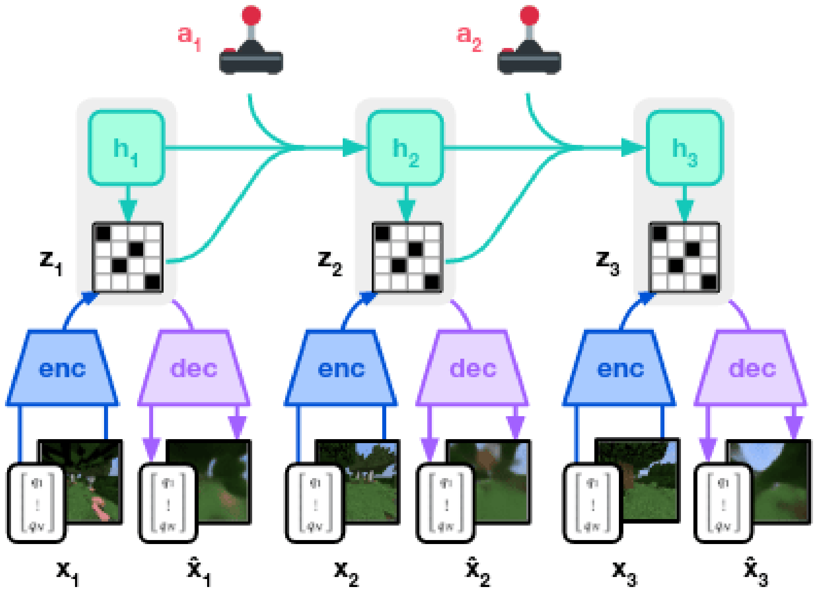
$$M_j = \frac{d^j M_{\mathbf{X}}}{d\mathbf{v}^j}\Big|_{\mathbf{v}=0}$$

Published as a conference paper at ICLR 2020

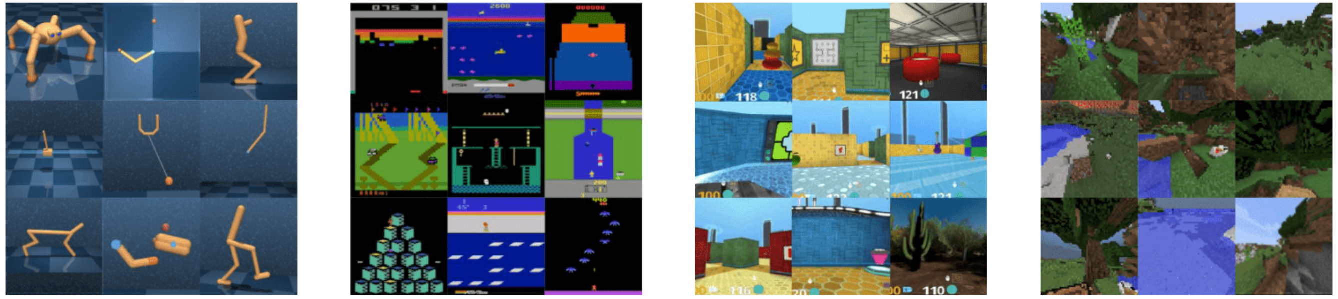
DISCRIMINATIVE PARTICLE FILTER REINFORCEMENT LEARNING FOR COMPLEX PARTIAL OBSERVATIONS

Xiao Ma[1], **Peter Karkus**[1], **David Hsu**[1] **and Wee Sun Lee**[1], **Nan Ye**[2]
[1]National Unviersity of Singapore, [2]The University of Queensland
{xiao-ma,karkus,dyhsu,leews}@comp.nus.edu.sg, nan.ye@uq.edu.au

Approach 4: Latent Representations (e.g. World Models)



(a) World Model Learning



(a) Control Suite **(b)** Atari **(c)** DMLab **(d)** Minecraft

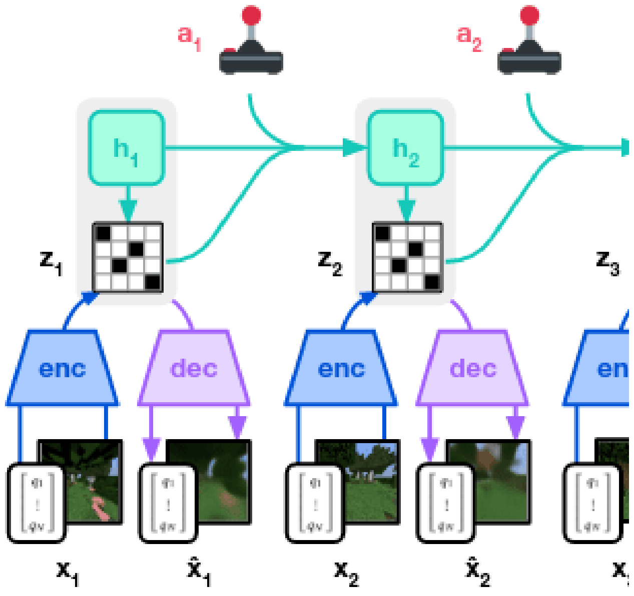
Mastering Diverse Domains through World Models

Danijar Hafner[1,2] Jurgis Pasukonis[1] Jimmy Ba[2] Timothy Lillicrap[1]

[1]DeepMind [2]University of Toronto

Approach
entations

(
)



(a) World Model Learning

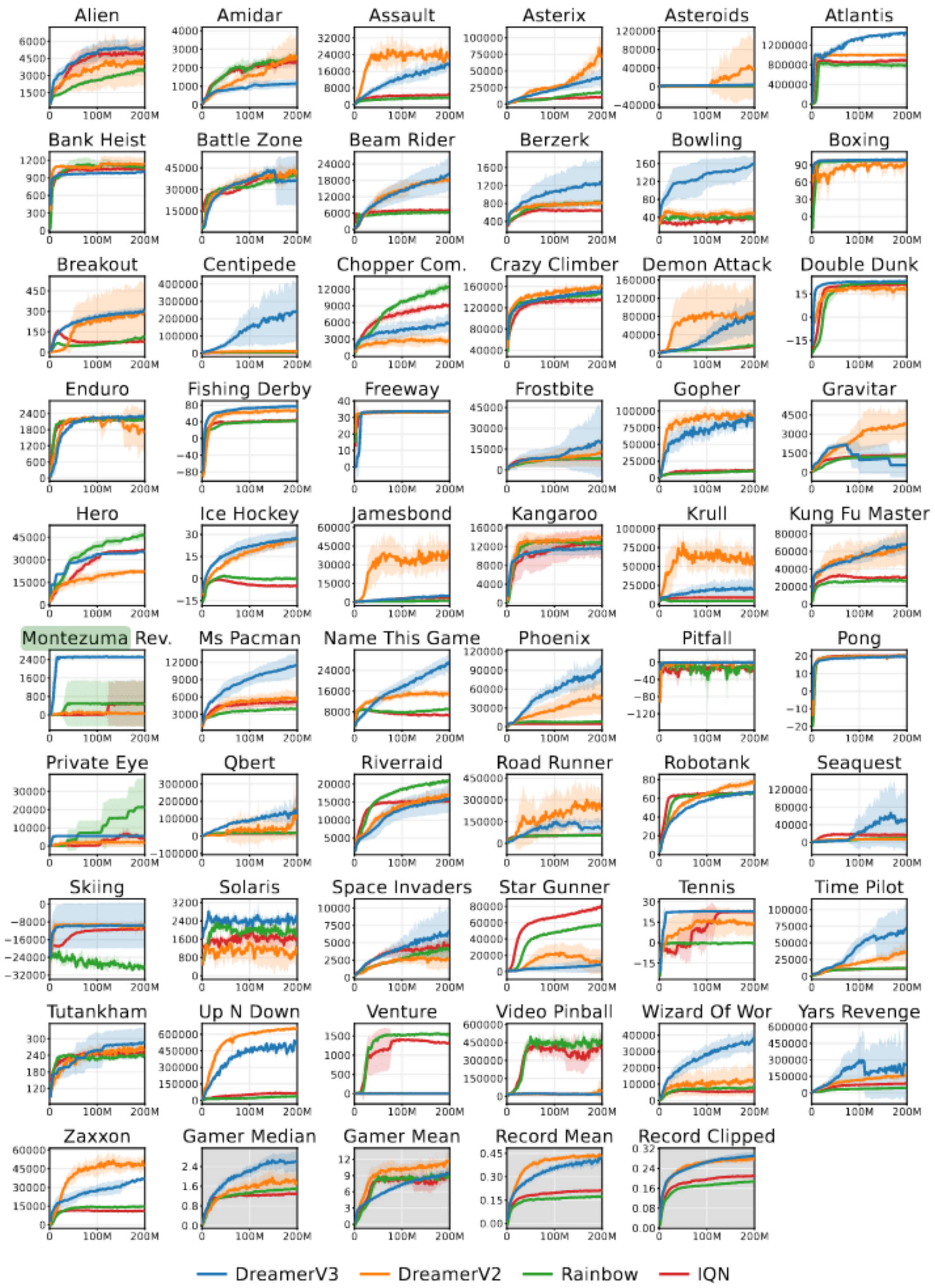
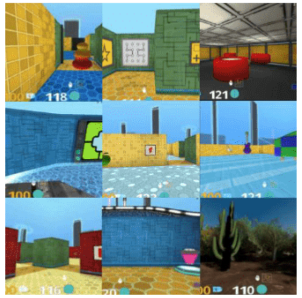
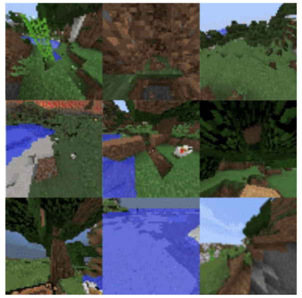


Figure U.1: Atari training curves with a budget of 200M frames.



(c) DMLab



(d) Minecraft

g Diverse Domains through World Models

Hafner,[1,2] Jurgis Pasukonis,[1] Jimmy Ba,[2] Timothy Lillicrap[1]

[1]DeepMind [2]University of Toronto