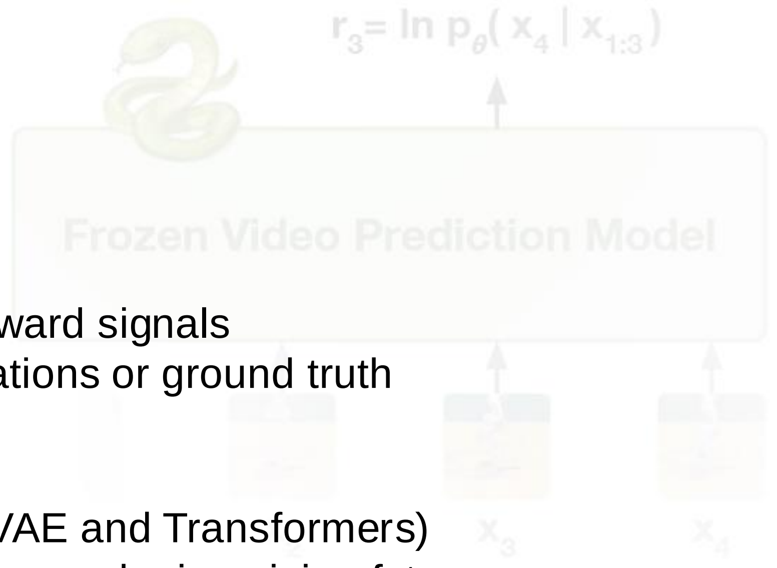
Video prediction models as rewards for reinforcement learning (Escontrela, et al. 2023)

Thorben Klamt - 20.06.2024

Advanced Topics in Reinforcement Learning, Theresa Eimer, Prof. Dr. rer. nat. Marius Lindauer,
Gottfried Wilhelm Leibniz University Hannover

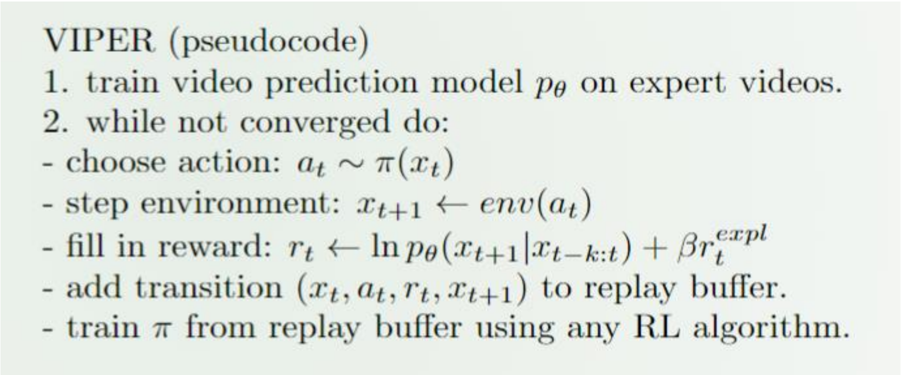
Video prediction models as rewards for reinforcement learning (Escontrela, et al. 2023)


$r_3 = \ln p_\theta(x_4 | x_{1:3})$
Frozen Video Prediction Model
x_3
x_4

VIPER

- Pretrained video prediction models to derive reward signals
- learns complex behaviors without action annotations or ground truth rewards

- VideoGPT (*2021,* Video Generation using VQ-VAE and Transformers)
- DreamerV3 (*2023,* learns a env-model and improves by imagining future scenarios)

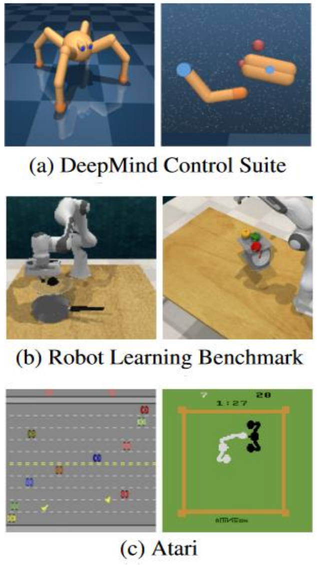


```
VIPER (pseudocode)
1. train video prediction model $p_\theta$ on expert videos.
2. while not converged do:
- choose action: $a_t \sim \pi(x_t)$
- step environment: $x_{t+1} \leftarrow env(a_t)$
- fill in reward: $r_t \leftarrow \ln p_\theta(x_{t+1}|x_{t-k:t}) + \beta r_t^{expl}$
- add transition $(x_t, a_t, r_t, x_{t+1})$ to replay buffer.
- train $\pi$ from replay buffer using any RL algorithm.
```

Escontrela, Alejandro, et al. "Video prediction models as rewards for reinforcement learning." *Advances in Neural Information Processing Systems* 36 (2024).

Video prediction models as rewards for reinforcement learning (Escontrela, et al. 2023)

- Open source https://escontrela.me/viper
  - Notebooks, checkpoints and example applications
- Straight forward algorithm
  - Custom or improved VideoGPT models
  - Widely available training data
  - Highly applicable to human-like models


(a) DeepMind Control Suite

(b) Robot Learning Benchmark

(c) Atari

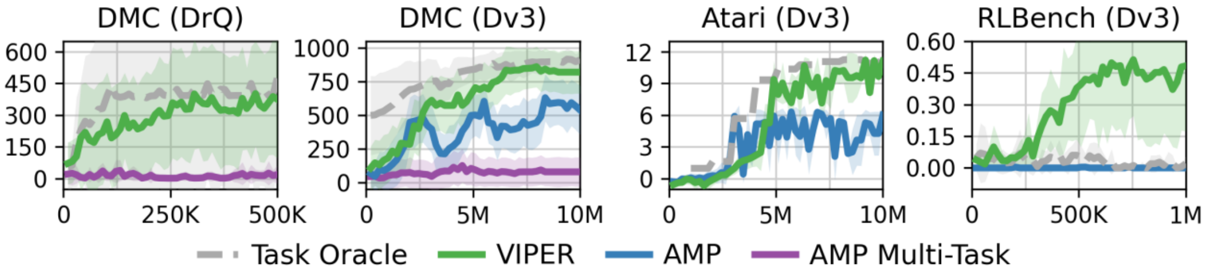


Fig.1 and Fig.2: Aggregated results across 15 DMC tasks (top a), 7 Atari games (top c), and 6 RLBench tasks (top b). DMC results are provided for DrQ and DreamerV3 (Dv3) agents. Atari and RLBenchmark results are reported for DreamerV3. Atari scores computed using Human-Normalized mean.

Escontrela, Alejandro, et al. "Video prediction models as rewards for reinforcement learning." *Advances in Neural Information Processing Systems* 36 (2024).