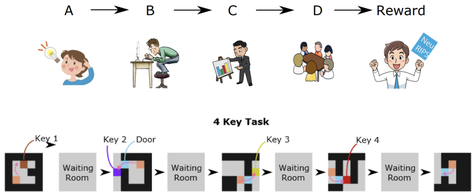
# Contrastive Retrospection: honing in on critical steps for rapid learning and generalization in RL

Chen Sun[1,2], Wannan Yang[3], Thomas Jiralerspong[1,2], Dane Malenfant[4], Benjamin Alsbury-Nealy[5], Yoshua Bengio[1,2,3], Blake Richards[1,2,3]
[1]Mila, [2]Université de Montréal, [3]New York University, [4]McGill University, [5]SilicoLabs Incorporated, [6]CIFAR

Bourses postdoctorales Banting Postdoctoral Fellowships

## INTRODUCTION




In real life, critical steps are few in number, and are separate from each other and from the reward by many other irrelevant events. But with such structured experience, learning by contemporary RL algorithms often collapse.

## METHODS

We designed a new contrastive loss for honing in on sparse critical steps that separate success from failure. This strategy simultaneously helps OOD generalization and rapid long term credit assignment in RL .

$$\mathcal{L}_{\text{ConSpec}} = \sum_{i=1}^{H}\left[\frac{1}{M_{\mathcal{S}}}\sum_{k\in\mathcal{S}}|1-\max_{t\in\{1\ldots T\}} s_{ikt}| + \frac{1}{M_{\mathcal{F}}}\sum_{k\in\mathcal{F}}|\max_{t\in\{1\ldots T\}} s_{ikt}|\right] + \alpha\cdot\frac{1}{H}\sum_{i\neq j}\sum_{k\in\mathcal{S}}|\cos(\boldsymbol{s_{ik}},\boldsymbol{s_{jk}})|$$

Successes · Failures · Diversity

$$s_{ikt} = \cos(h_i, g_\theta(z_{kt}))$$

## CONCLUSION

$$p(\text{success}|\{O_t\})$$

$$p(\{O_{t_1}..O_{t_k}\}|\text{success}=1) \overset{\text{ConSpec}}{=} \prod p(O_{t_i}|\text{success}=1)$$

**Premises:** Outcomes in the world are dependent on a sparse number of critical states.
**Strategy:** Use of episodic memory, prioritized sampling of success & inductive biases from hippocampal neuroscience.

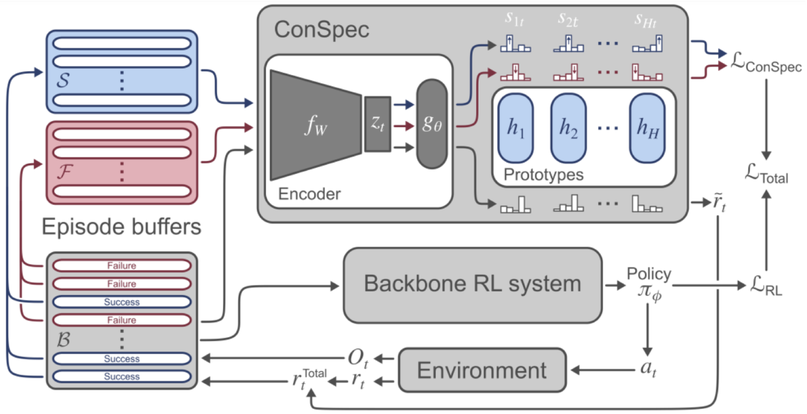
***It's easier to try and figure out a few critical steps than to try and model the whole world.***

## RELATED WORK

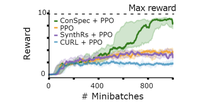
ConSpec centralizes several key intuitions shared with classical works:

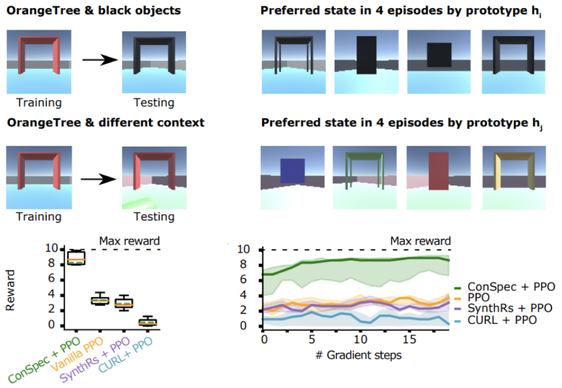
- **Long term credit assignment:** TVT, Synthetic Returns, RUDDER, RND
- **Temporal abstraction:** Hierarchical RL, options discovery, bottleneck states
- **Contrastive learning:** CPC, MOCO, SimCLR in computer vision, CURL, Contrastive RL
- **Neuroscience:** ESR cells, retrospective dopamine, prioritized replay

## Principles of ConSpec (Contrastive Retrospection)

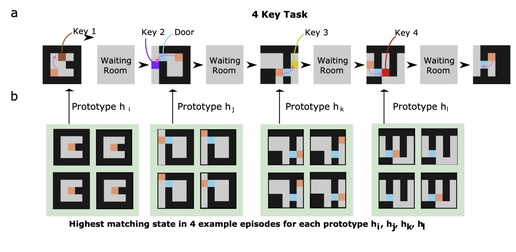


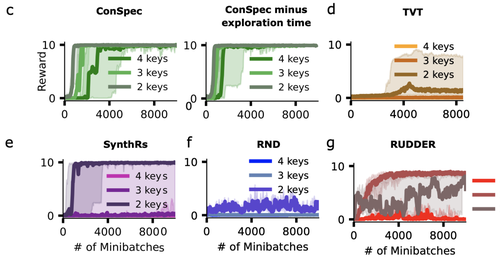

## 1.ConSpec's invariant representations help OoD generalization



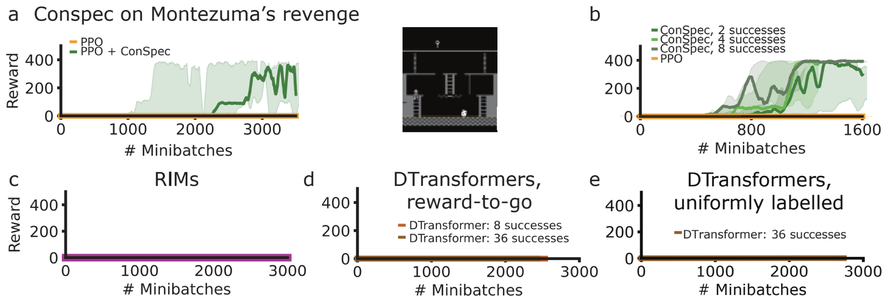




## 2.ConSpec enables rapid long-term credit assignment in grid-world tasks with multiple critical steps




## 3.ConSpec enables credit assignment in Montezuma's Revenge with only 2 spurious success trajectories, and no specialized exploration design




## 4.ConSpec helps credit assignment in continuous control tasks