

Can a MISL Fly?

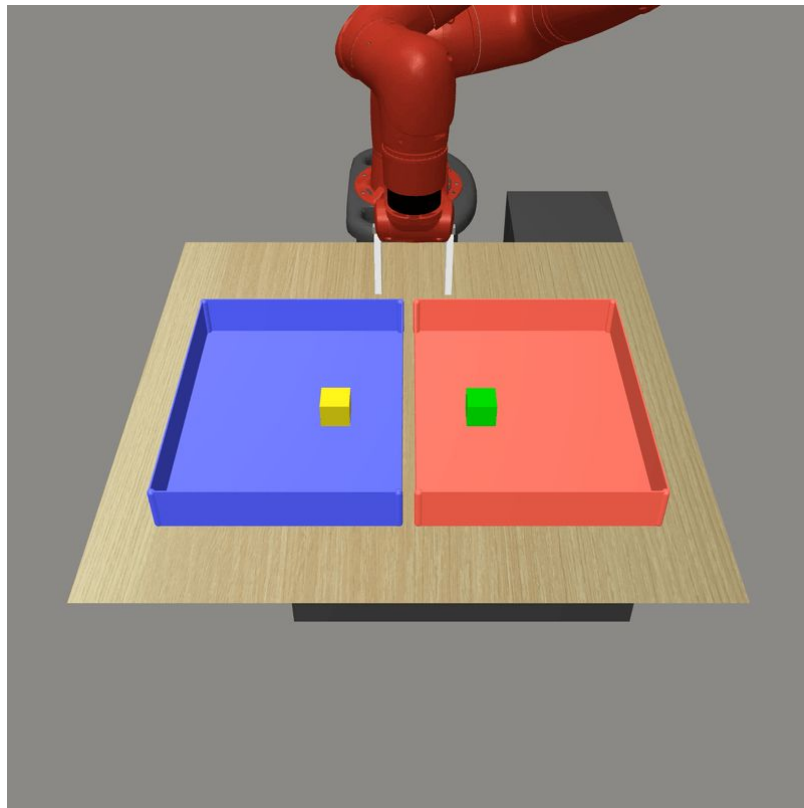
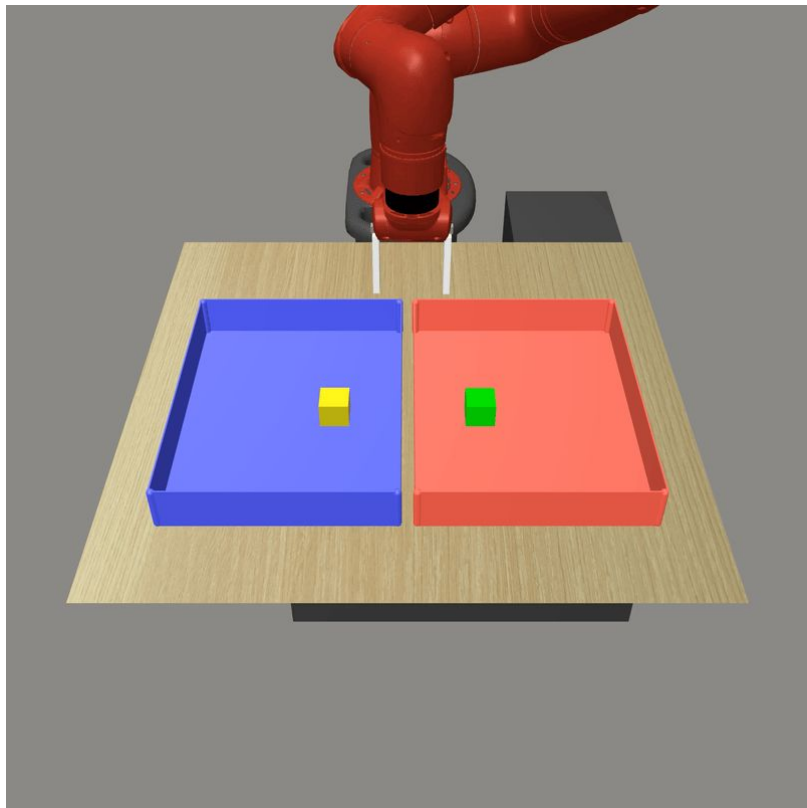
Analysis and Ingredients for Mutual Information Skill Learning

Chongyi Zheng*, **Jens Tuyls***, Joanne Peng,
Benjamin Eysenbach

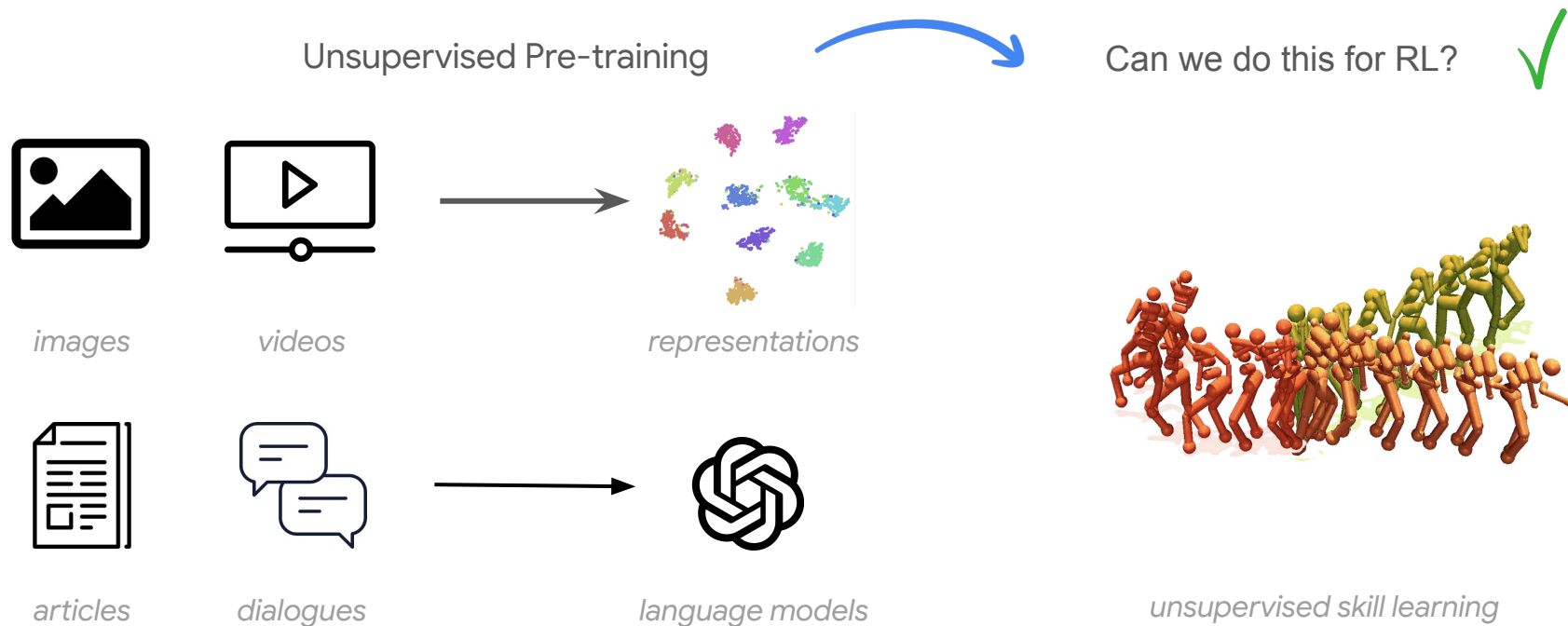
{chongyiz, jtuyls}@princeton.edu



Our work: unsupervised pre-training for RL (demo)



Unsupervised pre-training has proven successful in CV and NLP.



[1] Chuang et al. Debiased contrastive learning. 2020.

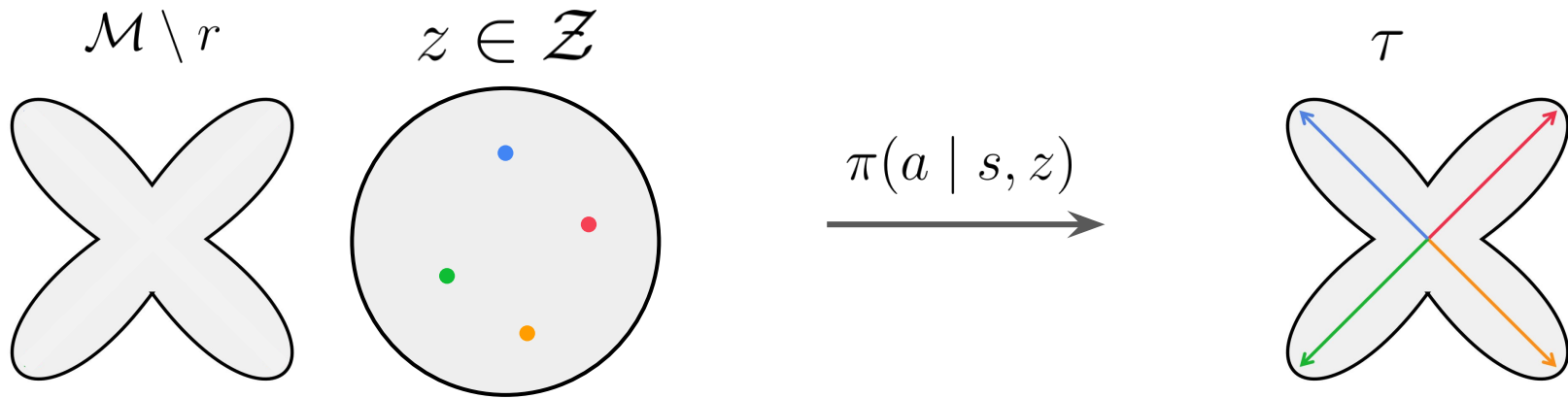
[2] He et al. Masked Autoencoders Are Scalable Vision Learners. 2021.

[3] Radford et al. Improving language understanding by generative pre-training. 2018.

[4] Sharma et al. Dynamics-Aware Unsupervised Discovery of Skills. 2020.

Mutual Information Skill Learning (MISL)

$$I^\pi(S, S'; Z) \stackrel{\text{diverse}}{=} H^\pi(S, S') \stackrel{\text{distinguishable}}{-} H^\pi(S, S' \mid Z)$$



$p(z) = \text{Gaussian or Uniform}$

- [1] Park et al. *METRA: Scalable Unsupervised RL with Metric-Aware Abstraction*. 2024.
- [2] Eysenbach et al. *Diversity is all you need: Learning skills without a reward function*. 2018.
- [3] Sharma et al. *Dynamics-Aware Unsupervised Discovery of Skills*. 2020.
- [4] Laskin et al. *CIC: Contrastive Intrinsic Control for Unsupervised Skill Discovery*. 2022.

Alternative ideas to unsupervised skill learning

MISL

$$\textcircled{\text{O}} \text{ METRA [7]} \quad \max_{\pi} I_{\mathcal{W}}^{\pi}(S; Z)$$

Wasserstein dependency measure

Can we build effective skill learning algorithms within the
MISL framework?

Learning state representations $\phi(s)$ to approximate the WDM

Updating the policy $\pi(a | s, z)$ using the successor features

$\textcircled{\text{O}}$ DIYAN [2]

$\textcircled{\text{O}}$ VISR [1]

$\textcircled{\text{O}}$ CIC [3]

$\textcircled{\text{O}}$ VIC [4]

$\textcircled{\text{O}}$ DADS [5]

$\textcircled{\text{O}}$ EDL [6]

[1] Hansen et al. Fast task inference with variational intrinsic successor features. 2020.

[2] Eysenbach et al. Diversity is all you need: Learning skills without a reward function. 2018.

[3] Laskin et al. CIC: Contrastive Intrinsic Control for Unsupervised Skill Discovery. 2022.

[4] Gregor et al. Variational intrinsic control. 2016.

[5] Sharma et al. Dynamics-Aware Unsupervised Discovery of Skills. 2020.

[6] Campos et al. Explore, discover and learn: Unsupervised discovery of state-covering skills. 2020.

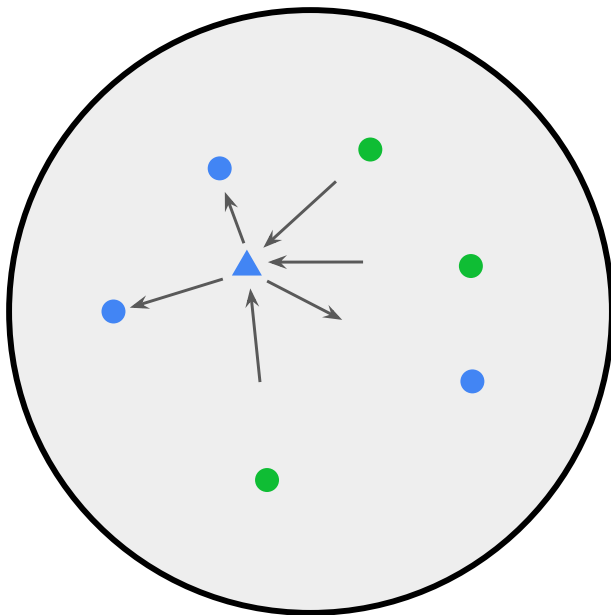
[7] Park et al. METRA: Scalable Unsupervised RL with Metric-Aware Abstraction. 2024.

Intuitions of METRA's representation objective

$$\min_{\lambda \geq 0} \max_{\phi} \underbrace{\mathbb{E}_{p^{\beta}(s, s', z)} [(\phi(s') - \phi(s))^{\top} z]}_{\text{Contrastive learning?}} - \lambda \underbrace{(\mathbb{E}_{p^{\beta}(s, s')} [\|\phi(s') - \phi(s)\|_2^2] - 1)}$$

$\phi(s') - \phi(s) :$ ▲

$z :$ ● ●



Contrastive learning?



Relating METRA's representation objective to contrastive learning

Ours

InfoNCE loss

$$I^\beta(S, S'; Z) \geq \underbrace{\mathbb{E}_{p^\beta(s, s', z)} [(\phi(s') - \phi(s))^\top z]}_{\text{blue line}} - \underbrace{\mathbb{E}_{p^\beta(s, s')} \left[\log \mathbb{E}_{p(z')} \left[e^{(\phi(s') - \phi(s))^\top z} \right] \right]}_{\text{green line}}$$

+: push together representations and skills from the same trajectory.

- : push away representations from other trajectories.

Second-order Taylor approximation



METRA

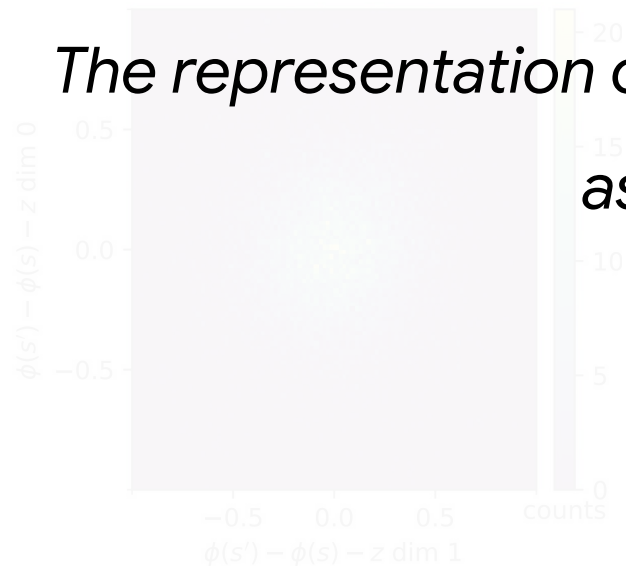
$$\underbrace{\mathbb{E}_{p^\beta(s, s', z)} [(\phi(s') - \phi(s))^\top z]}_{\text{blue line}} - \underbrace{\lambda \left(\mathbb{E}_{p^\beta(s, s')} [\|\phi(s') - \phi(s)\|_2^2] - 1 \right)}_{\text{green line}}$$

[1] Poole et al. *On variational bounds of mutual information*. 2019.

[2] Park et al. *METRA: Scalable Unsupervised RL with Metric-Aware Abstraction*. 2024.

METRA learns contrastive representations

Gaussianity ← contrastive representations [1] → Uniformity

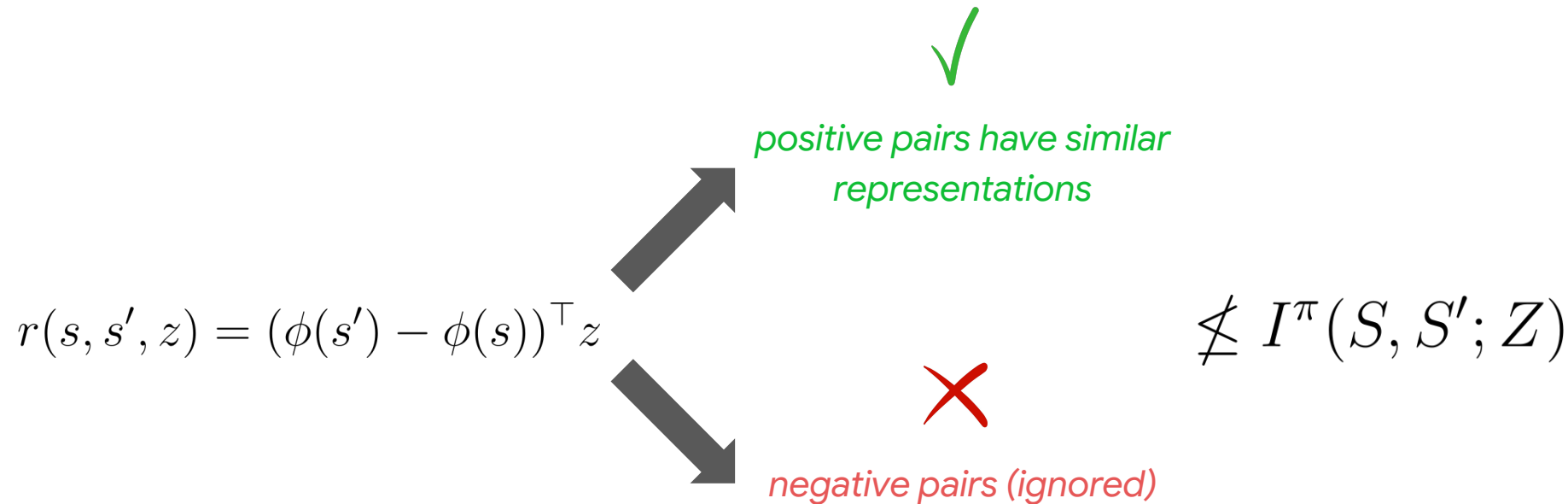


*The representation objective in METRA has the same effect as a **contrastive loss**.*



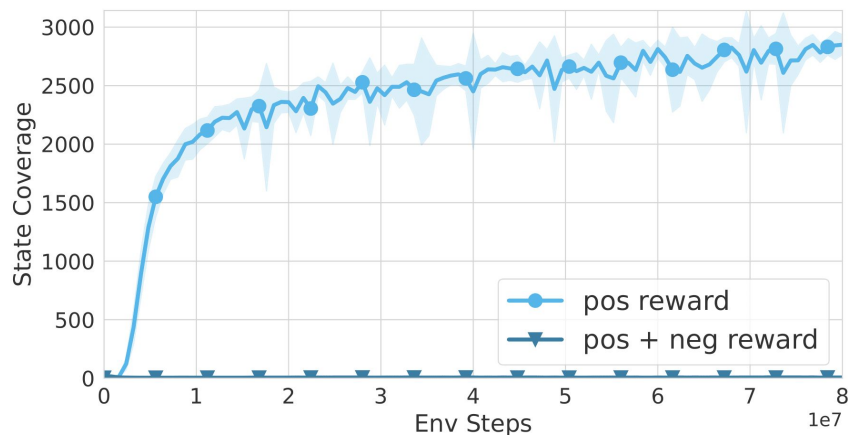
[1] Wang & Isola. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. 2020.

The intrinsic reward of METRA



Relating METRA's policy objective to an information bottleneck

Maximizing MI is not enough



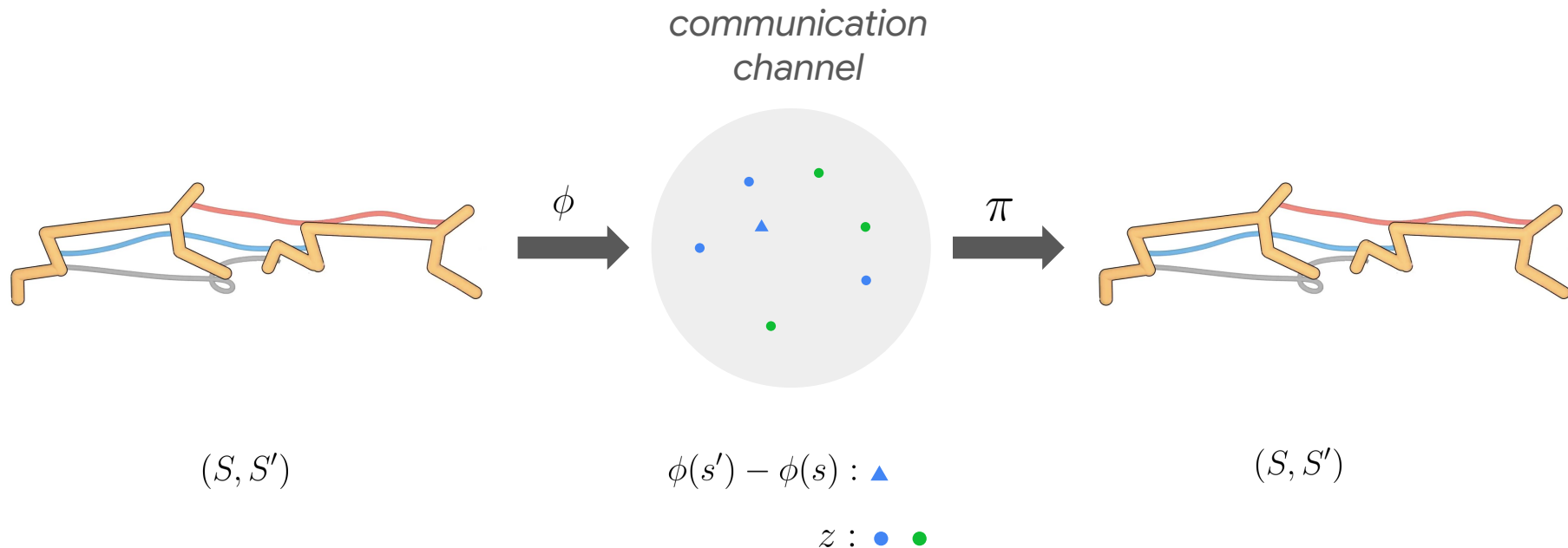
$$r(s, s', z) = (\phi(s') - \phi(s))^{\top} z \quad ?$$

*Lower bound on a
Information bottleneck*

[1] Tishby et al. *The information bottleneck method*. 2000.

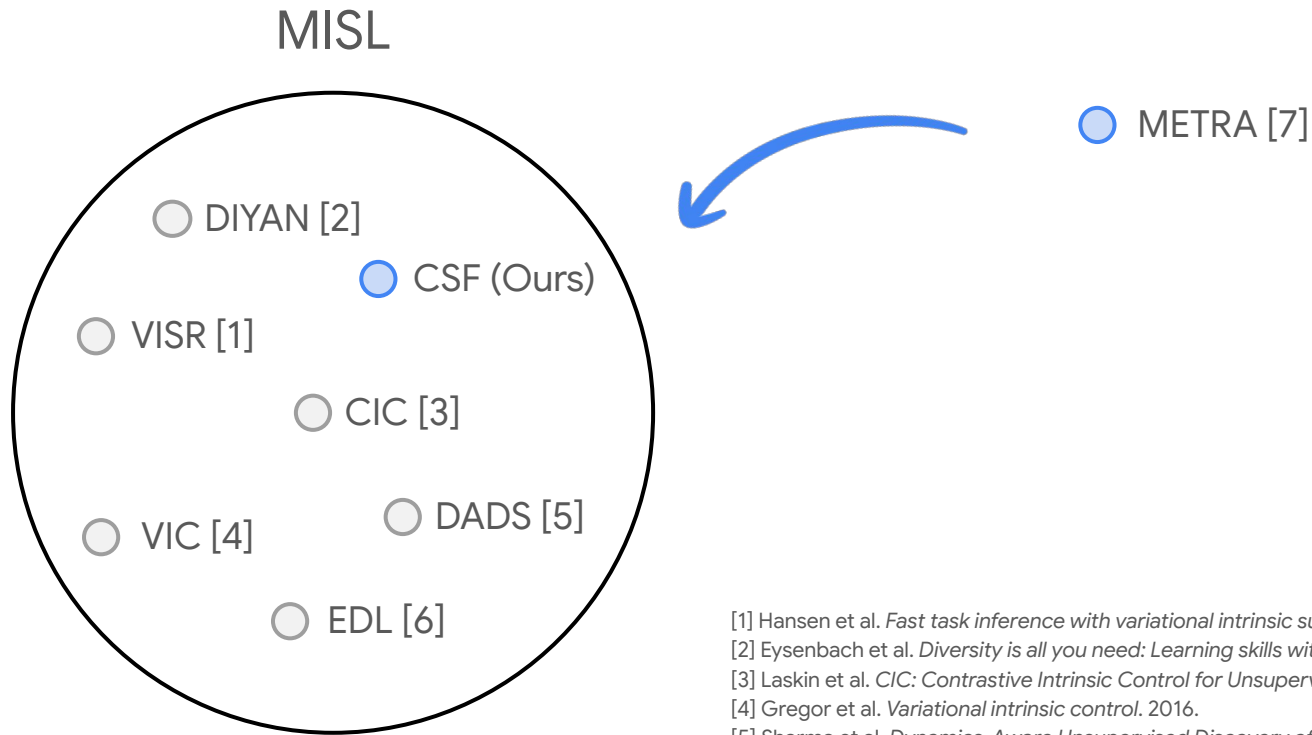
[2] Alemi et al. *Deep variational information bottleneck*. 2016.

Intuition for the information bottleneck



- [1] Janner & Du et al. *Planning with Diffusion for Flexible Behavior Synthesis*. 2022.
- [2] Tishby et al. *The information bottleneck method*. 2000.
- [3] Alemi et al. *Deep variational information bottleneck*. 2016.
- [4] Kingma. *Auto-Encoding Variational Bayes*. 2016.

Why do we need new interpretations?



[1] Hansen et al. *Fast task inference with variational intrinsic successor features*. 2020.

[2] Eysenbach et al. *Diversity is all you need: Learning skills without a reward function*. 2018.

[3] Laskin et al. *CIC: Contrastive Intrinsic Control for Unsupervised Skill Discovery*. 2022.

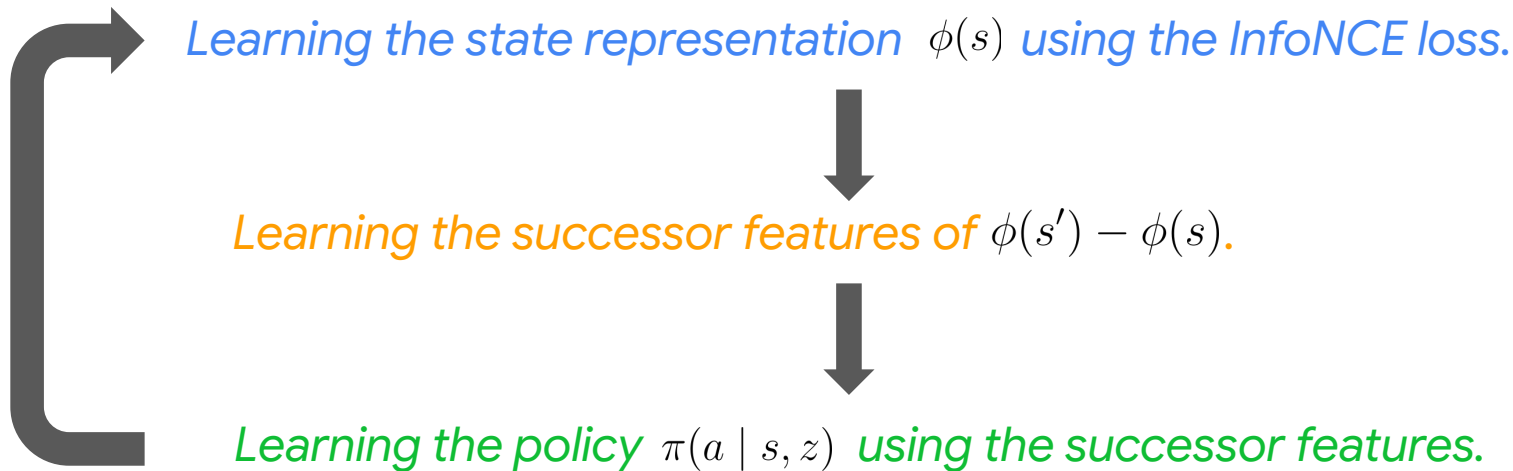
[4] Gregor et al. *Variational intrinsic control*. 2016.

[5] Sharma et al. *Dynamics-Aware Unsupervised Discovery of Skills*. 2020.

[6] Campos et al. *Explore, discover and learn: Unsupervised discovery of state-covering skills*. 2020.

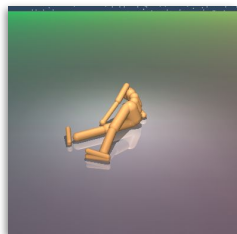
[7] Park et al. *METRA: Scalable Unsupervised RL with Metric-Aware Abstraction*. 2024.

Ideas of contrastive successor features (CSF)



Learning skills to explore the state space from pixels

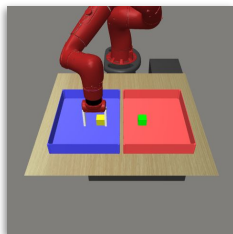
Humanoid



Kitchen

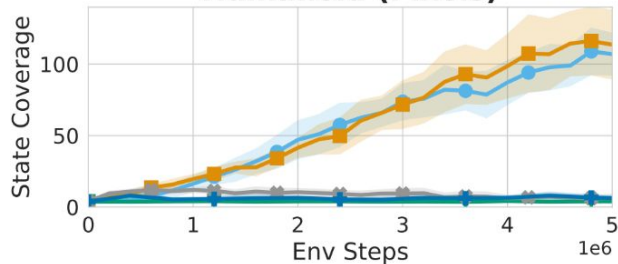


Robobin

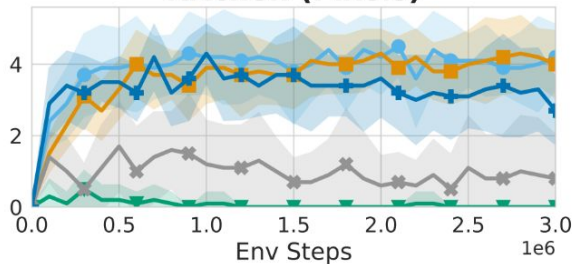


1.5x higher coverage!

Humanoid (Pixels)

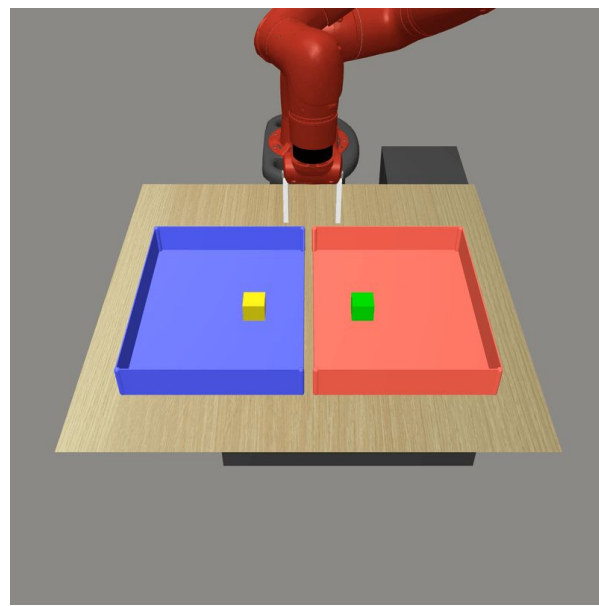
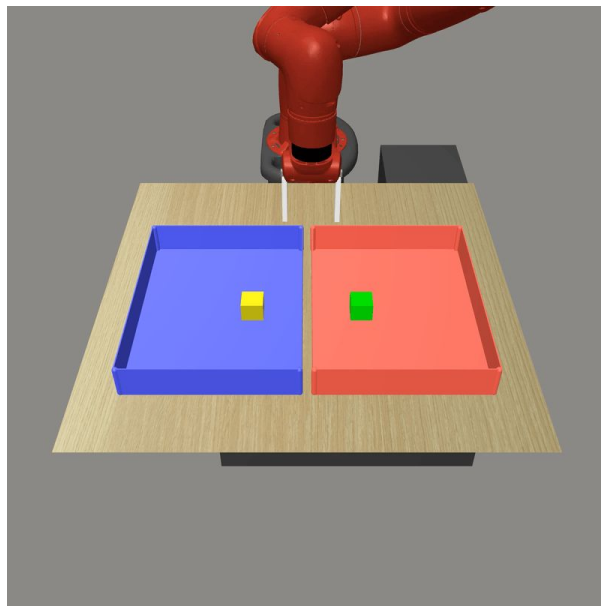
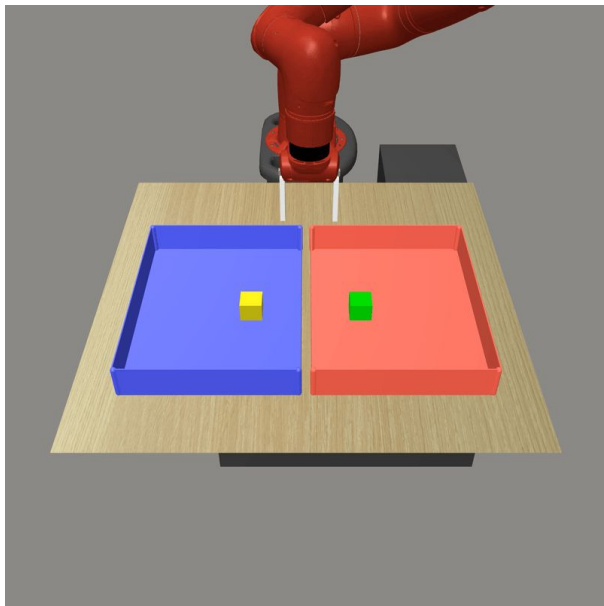


Kitchen (Pixels)



—●— CSF (ours) —■— METRA —▼— DIAYN —◆— DADS —×— CIC —+— VISR

CSF can learn manipulation skills without a reward function.

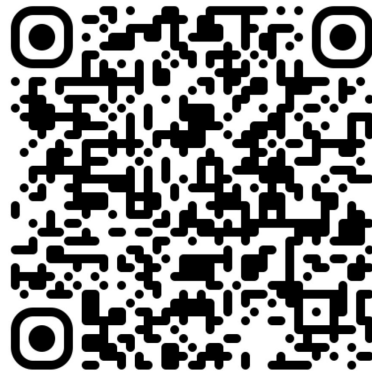


Summary and connections

tldr: explain and simplify METRA within MISL

- Representation learning - contrastive learning
- Policy learning - information bottleneck
- Simplified version: contrastive successor features
- Connections with many other areas:
 - ◆ Forward backward representations
 - ◆ Goal-conditioned RL
 - ◆ Zero-shot adaption

Video, code, and paper!



<https://princeton-rl.github.io/contrastive-successor-features/>