CSCI1470

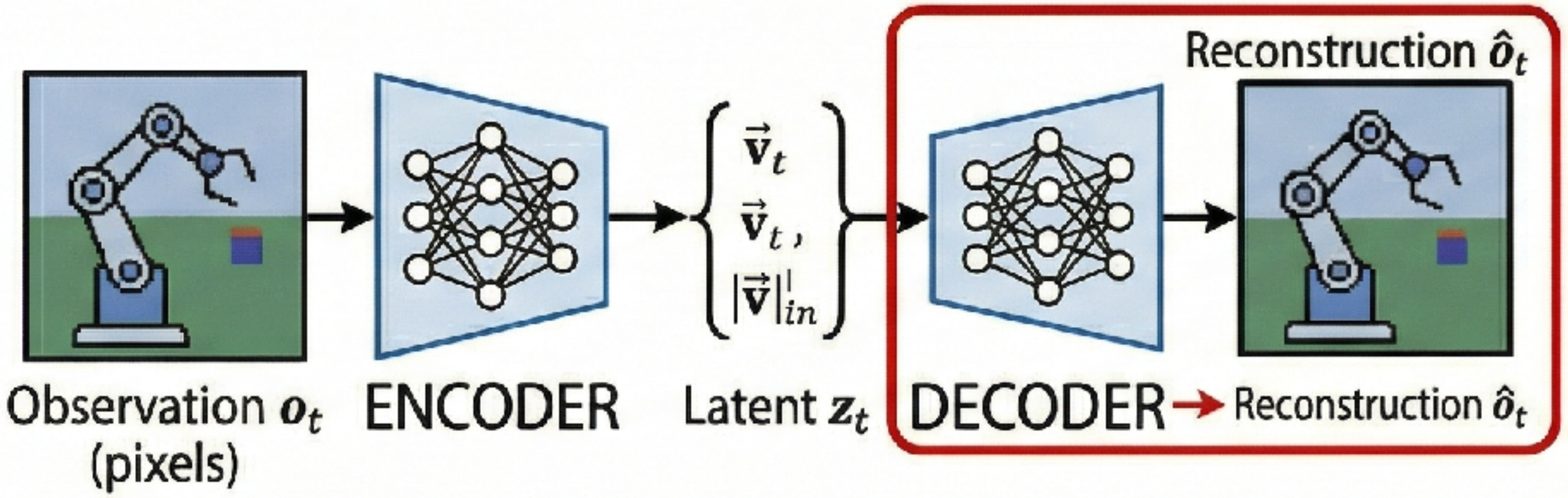
Deep Learning

Randall Balestriero

Recap

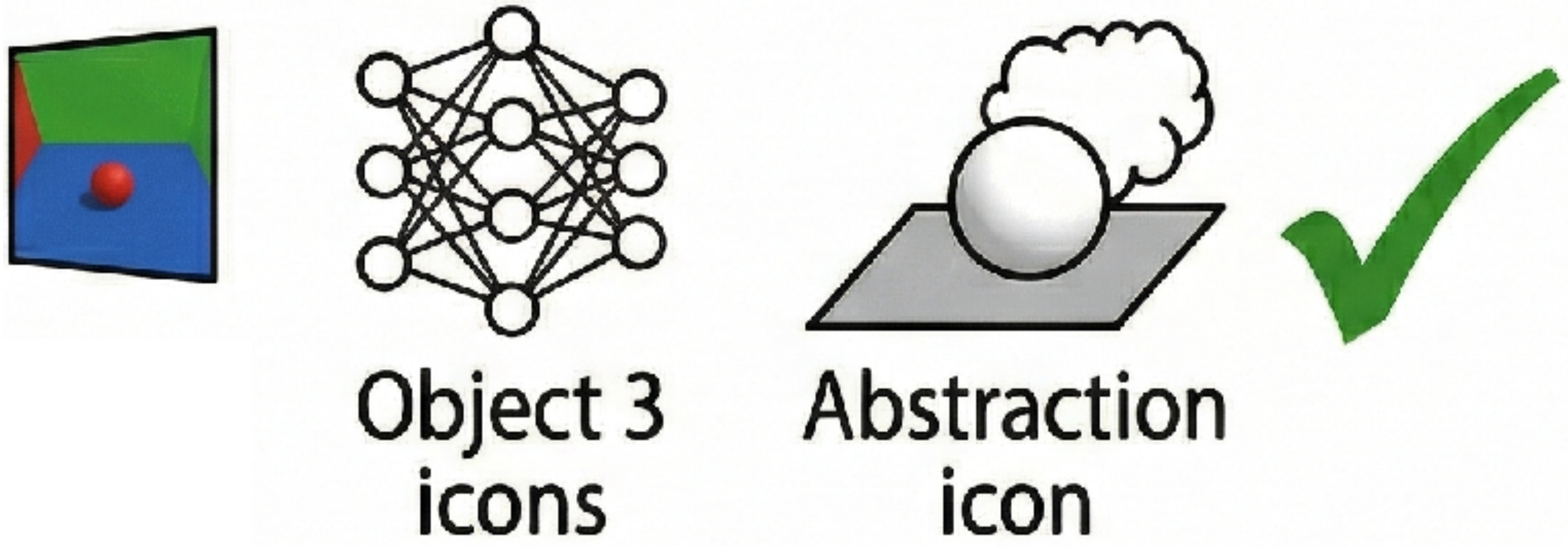
SUMMARY: THE PARADIGM SHIFT

PIXEL-SPACE WORLD MODEL



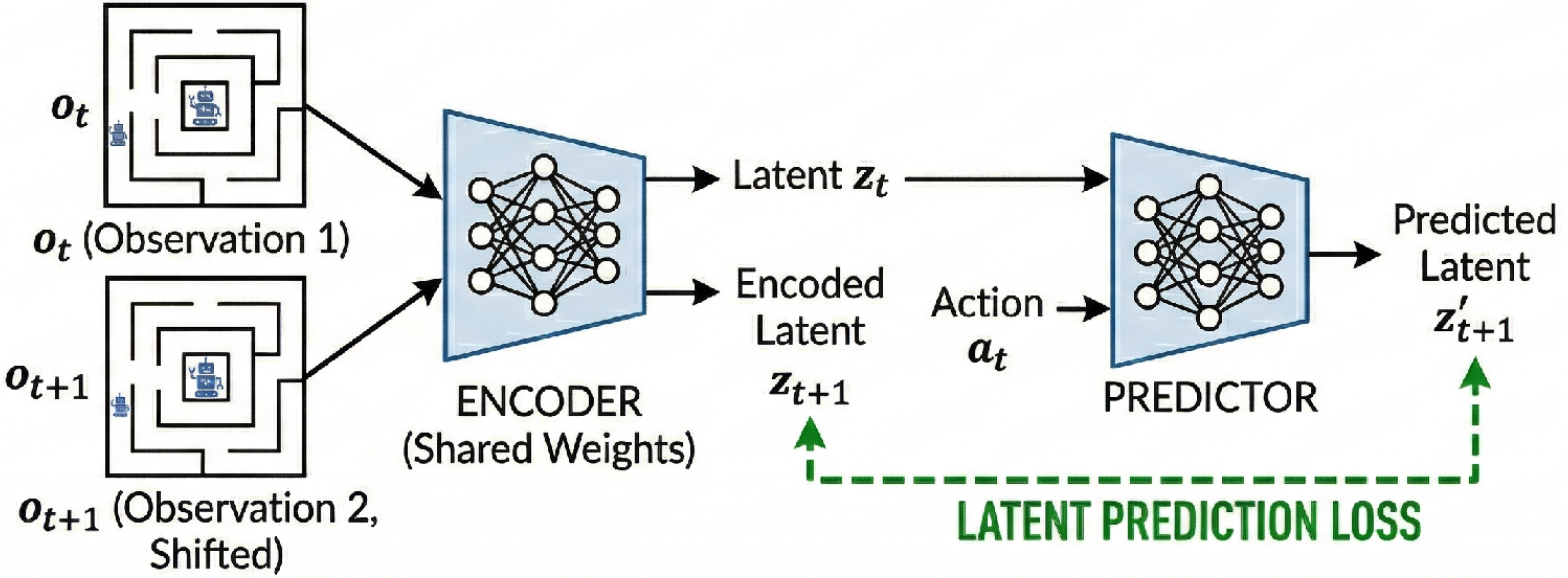
- High Reconstruct Cost
- Metric = MSE/Pixels
- Focus on *Appearance*
- Data Hungry

RECONSTRUCTION-FREE WORLD MODEL



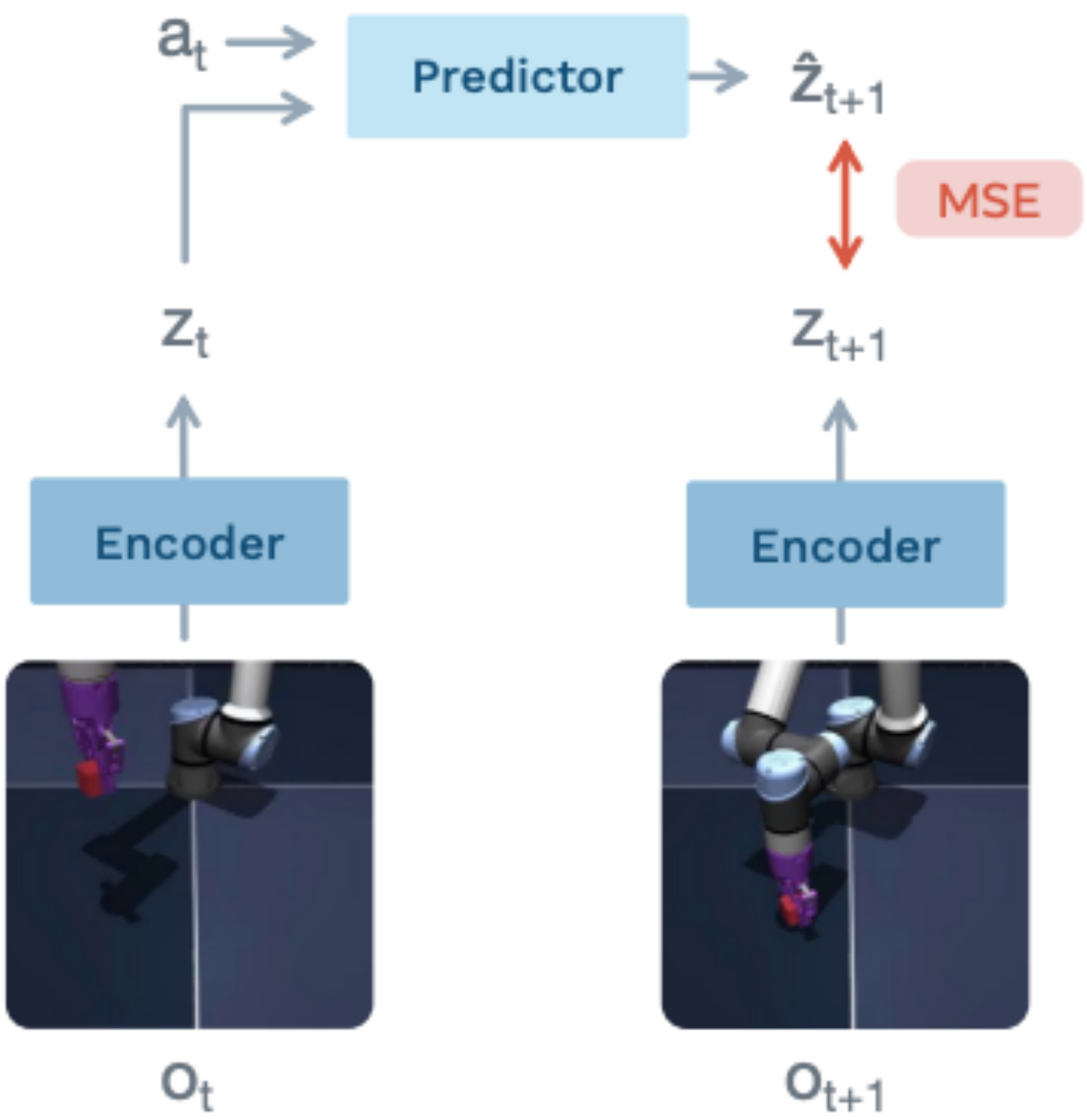
- ✓ Zero Reconstruct Cost
- Metric = Dynamics/Function
- Focus on *Structure*
- Data Efficient

JOINT-EMBEDDING PREDICTIVE ARCHITECTURE (JEPA)



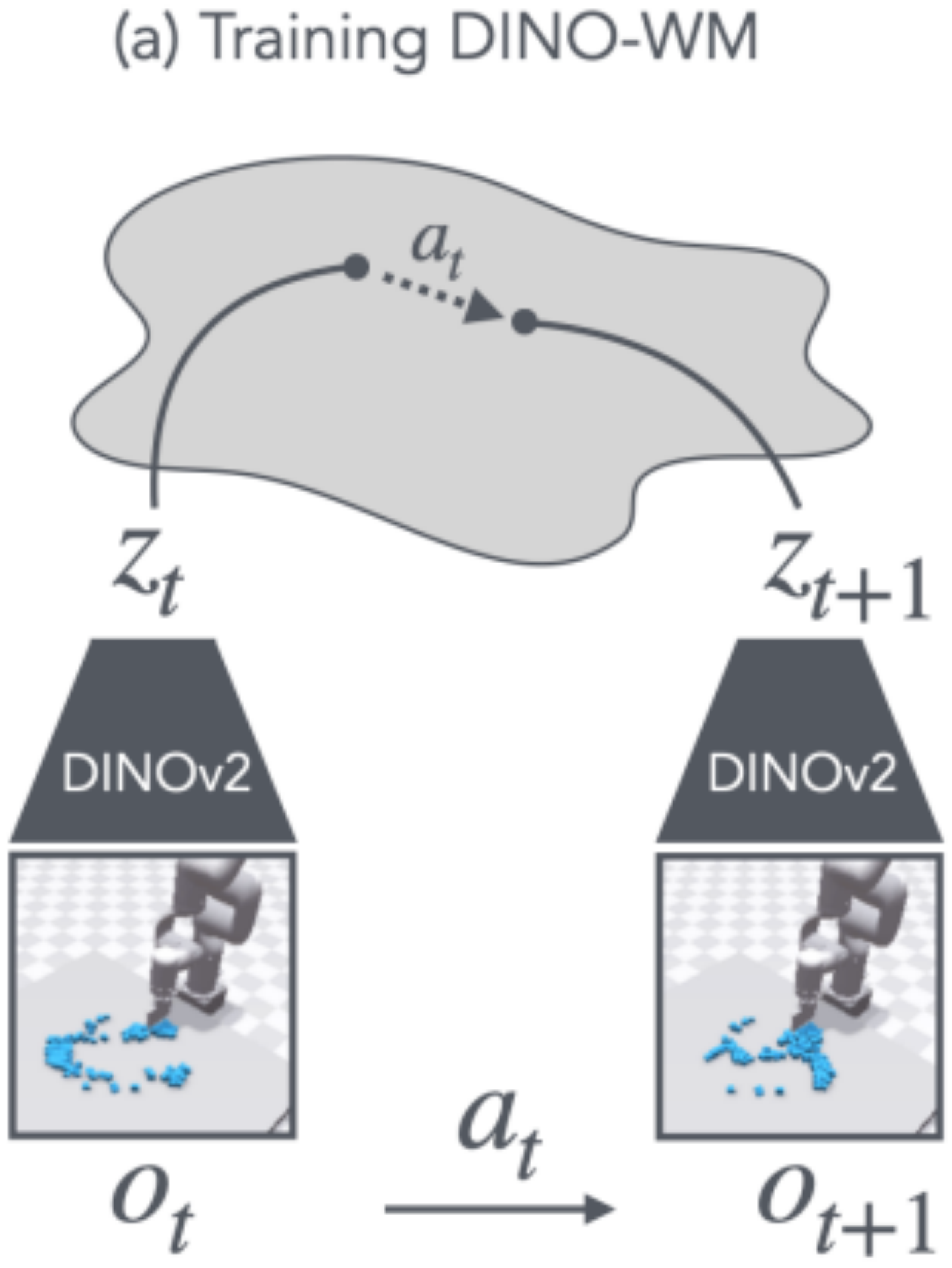
WILL THAT WORK OUT OF THE BOX?

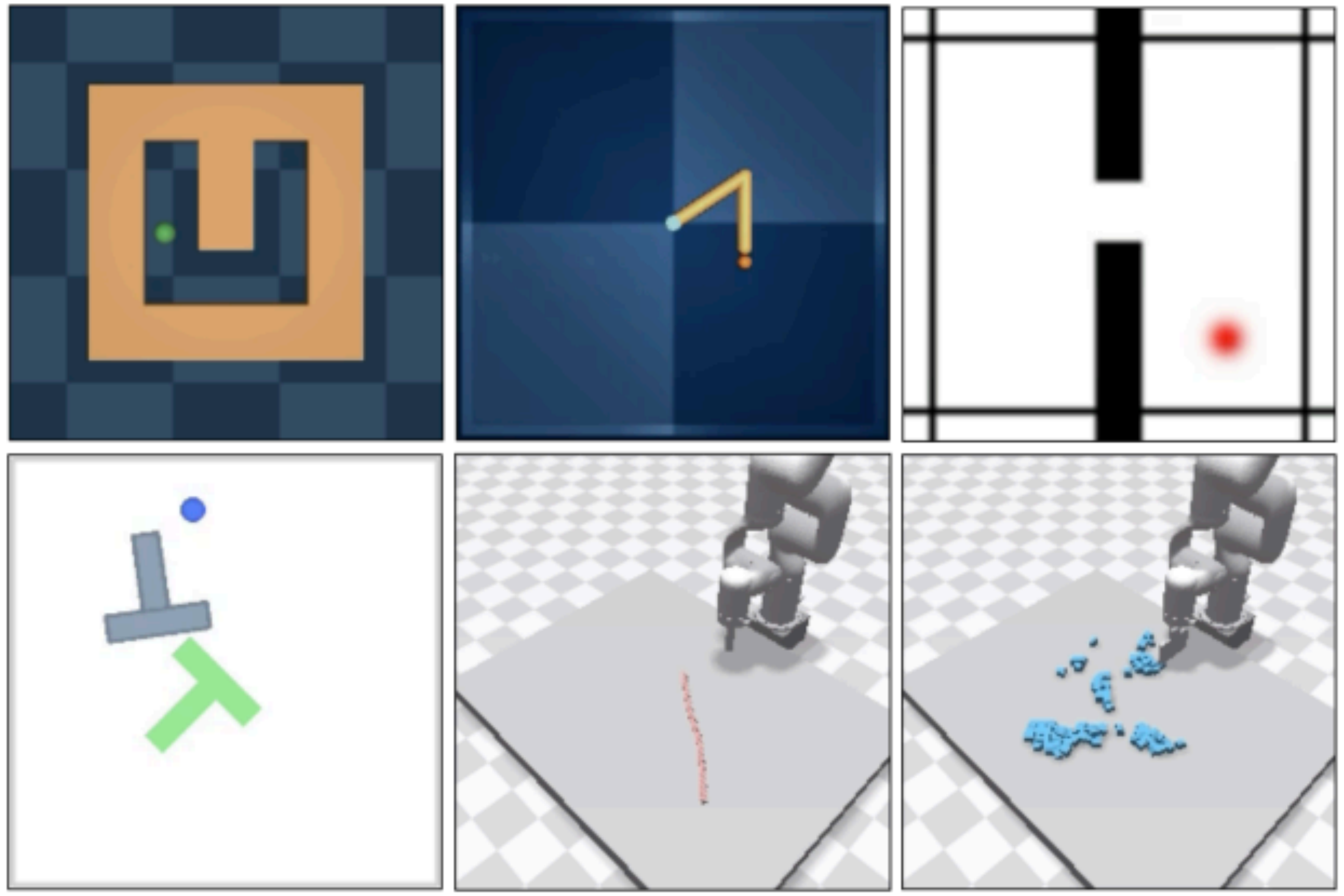
Consider: Ambiguous Futures? Degenerate Solutions? Metric in Latent Space?



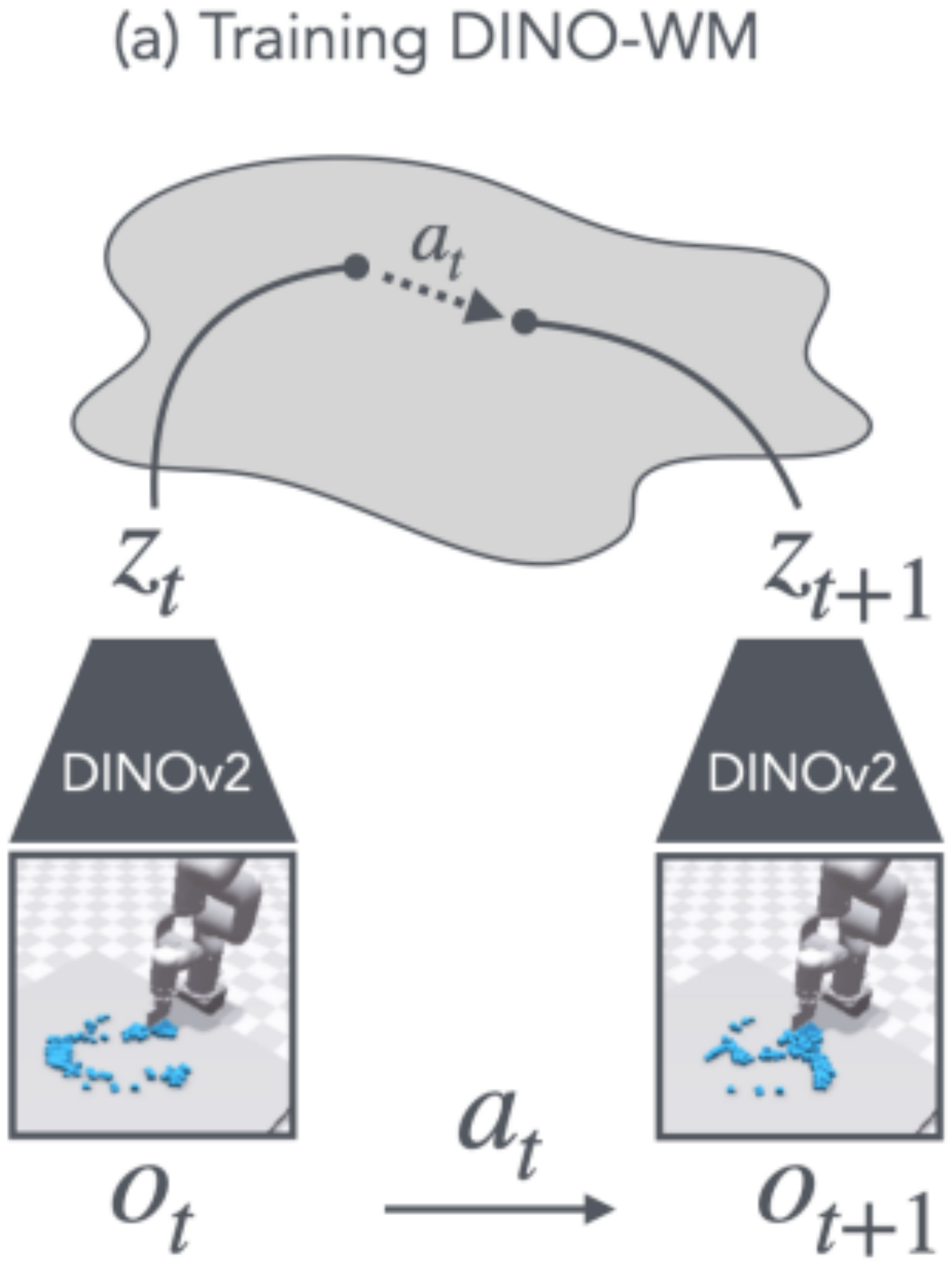
a_t
Predictor
$\hat{z}_{t+1}$
MSE
z_t
z_{t+1}
Encoder
Encoder
o_t
o_{t+1}

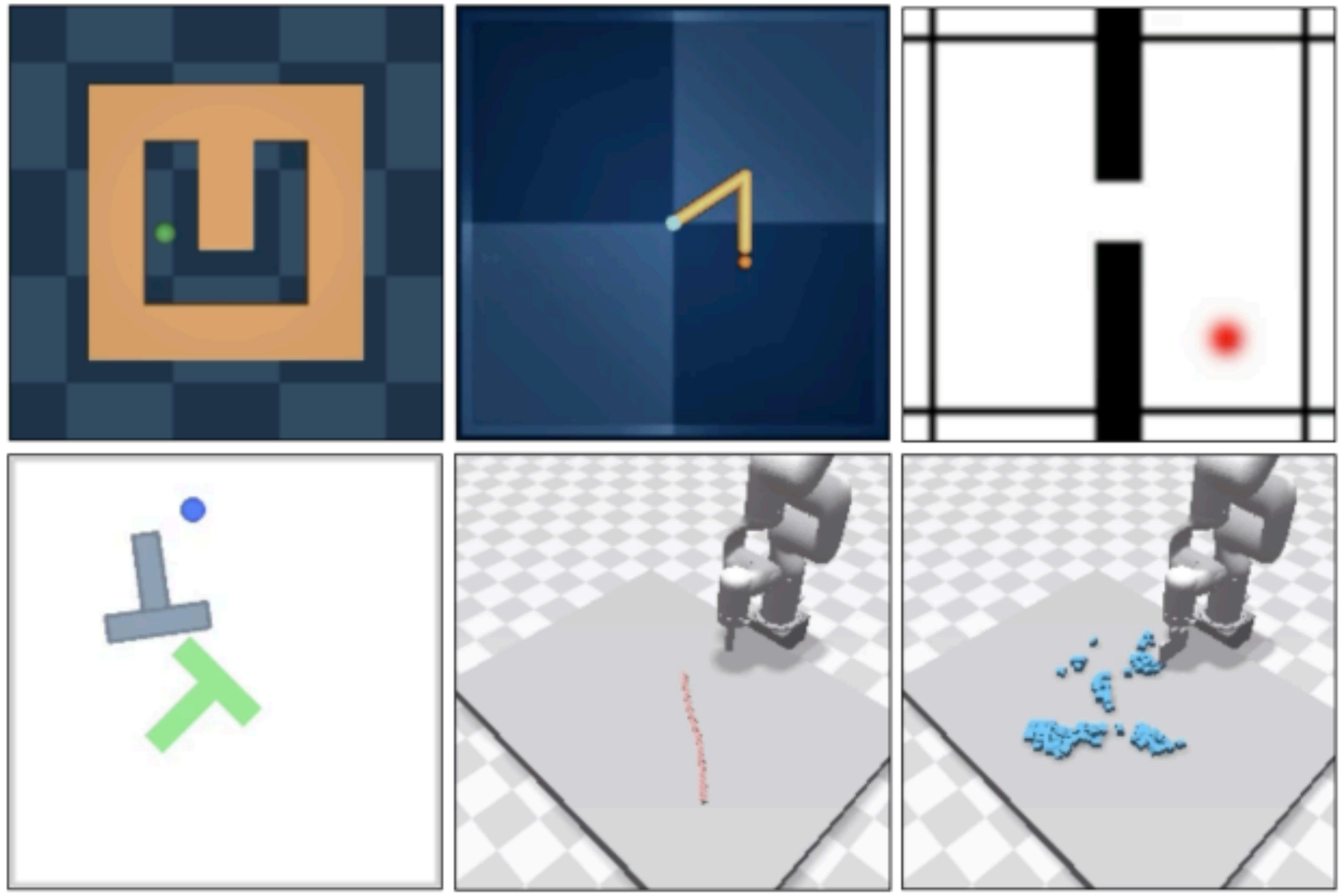
DINO-WM





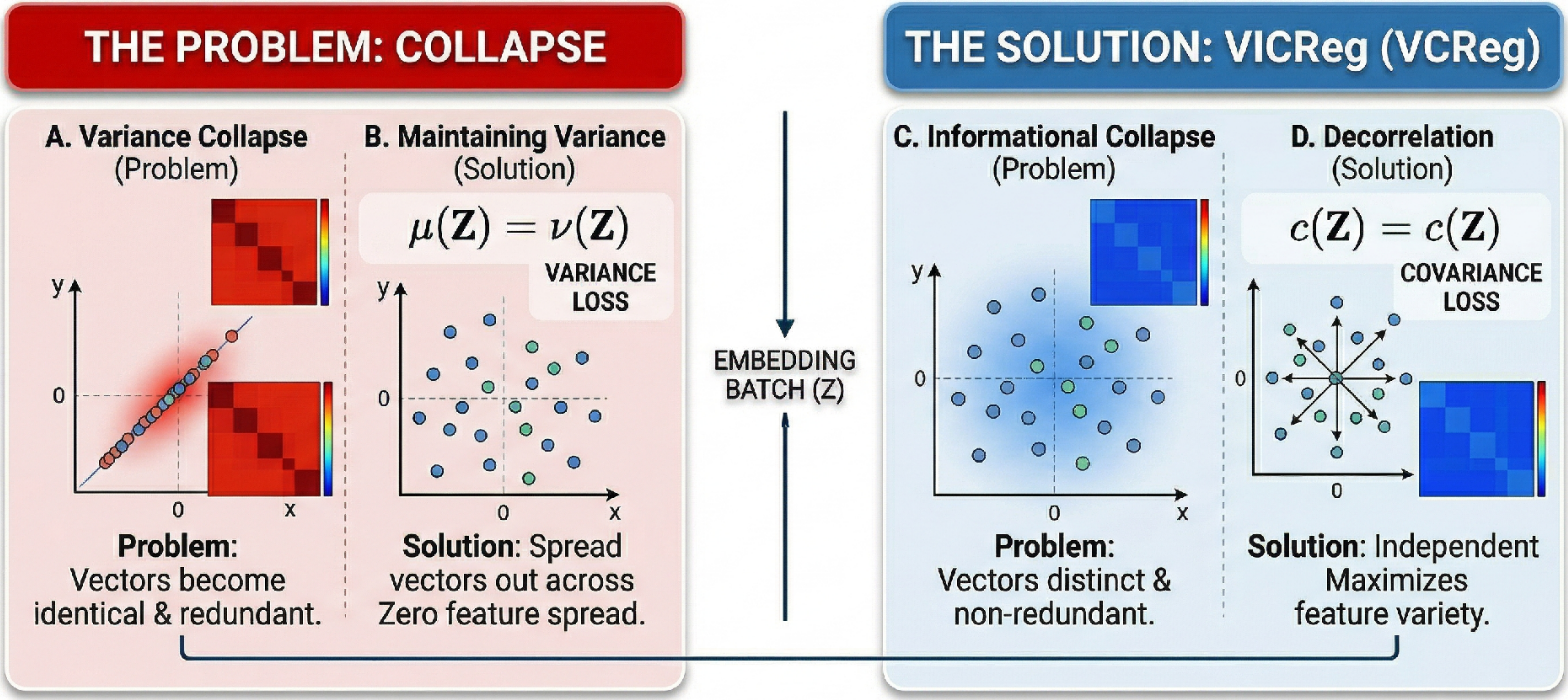
DINO-WM





Q: Any limitations?

VICReg: Direct Collapse Prevention (VCReg)

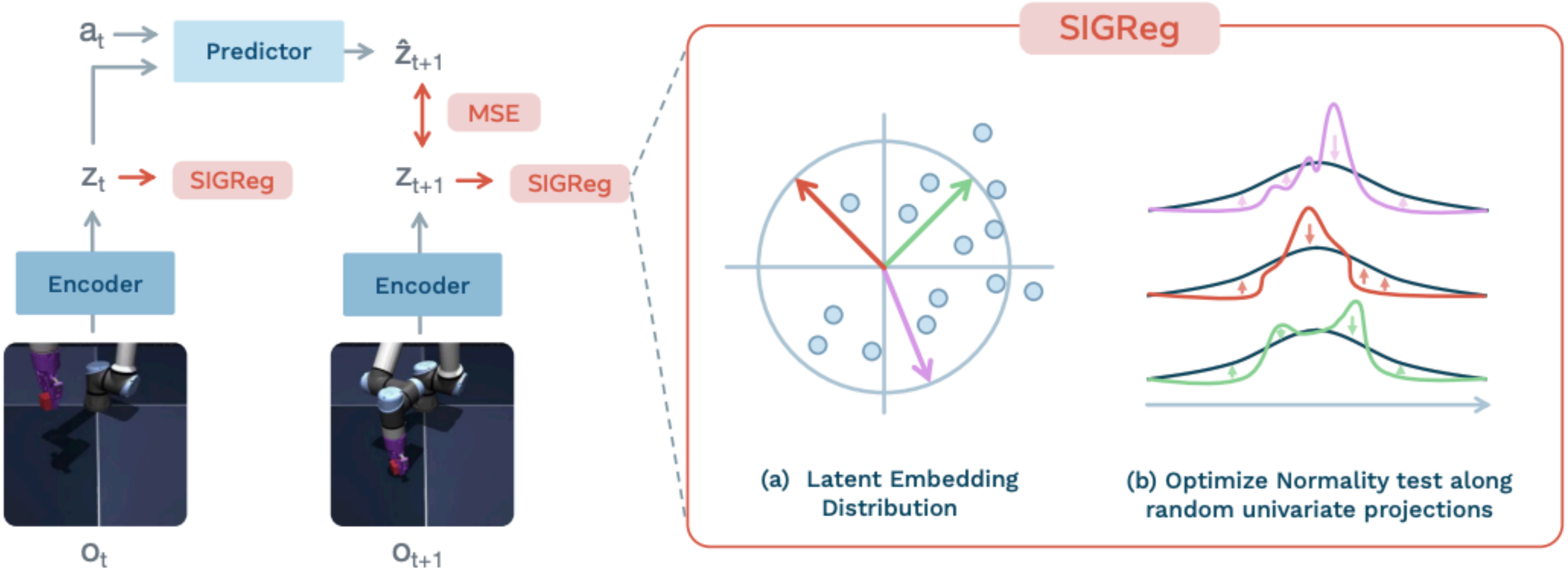


SUMMARY: VCReg terms ensure vectors are diverse (spread out) and unique (independent).

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{inv}}(\mathbf{Z}, \mathbf{Z}') + \underbrace{[\mathcal{L}_{\text{var}}(\mathbf{Z}) + \mathcal{L}_{\text{cov}}(\mathbf{Z})]}_{\text{VCReg}}$$


ANTI-COLLAPSE IN CONTRASTIVE LEARNING
THE COLLAPSE PROBLEM (Degenerate Solution)
Diverse Inputs
(e.g., varied categories)
Single, Indistinguishable Cluster (Collapse)
INFORMATION LOSS
y
0
0
x
Input variety is lost.
NEGATIVE PAIR REPULSION (Solution)
y
Easy Negatives
Hard Negative
s_{t+1}
s_{t+10}
s_t
0
Anchor
s_{t+2}
s_{t+3}
Wider Separation Achieved
Reduced Hardness
0
x
$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{negative}}(\mathbf{Z}_{\text{p}}, \mathbf{Z}_{\text{neg}})$

LeWorldModel



Sufficiency

Lemma 3: Hyperspherical Cramér-Wold

Let X, Y be $\mathbb{R}^d$-valued random vectors, then

$$\langle \boldsymbol{u}, X\rangle \overset{d}{=} \langle \boldsymbol{u}, Y\rangle, \forall \boldsymbol{u} \in \mathbb{S}^{d-1} \iff X \overset{d}{=} Y.$$

Convergence in distribution also holds. (Proof in Section B.8.)

Sufficiency

Lemma 3: Hyperspherical Cramér-Wold

Let X, Y be $\mathbb{R}^d$-valued random vectors, then

$$\langle \boldsymbol{u}, X\rangle \overset{d}{=} \langle \boldsymbol{u}, Y\rangle, \forall \boldsymbol{u} \in \mathbb{S}^{d-1} \iff X \overset{d}{=} Y.$$

Convergence in distribution also holds. (Proof in Section B.8.)

Q: What is the univariate distribution obtained by projecting a isotropic Gaussian into a unit-norm vector?

LeWorldModel

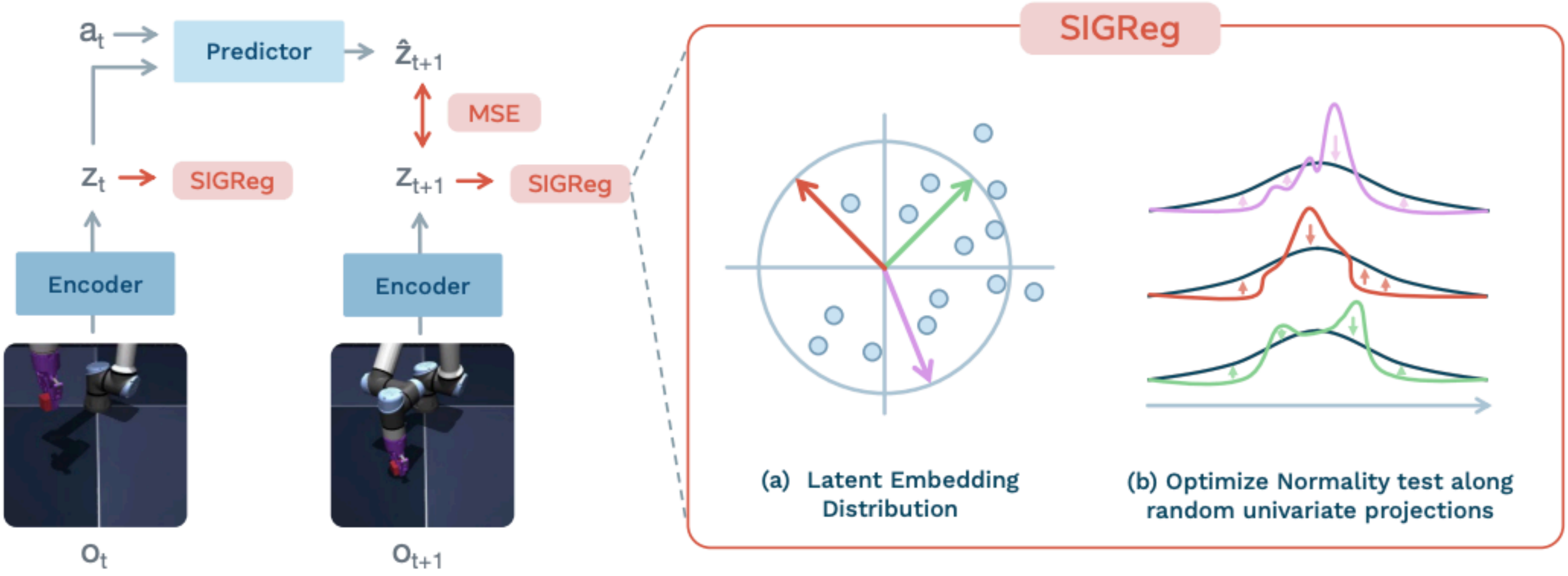
Definition 2: SIGReg (PyTorch code in algorithm 1)

SIGReg sketches a statistical test T towards isotropic Gaussian

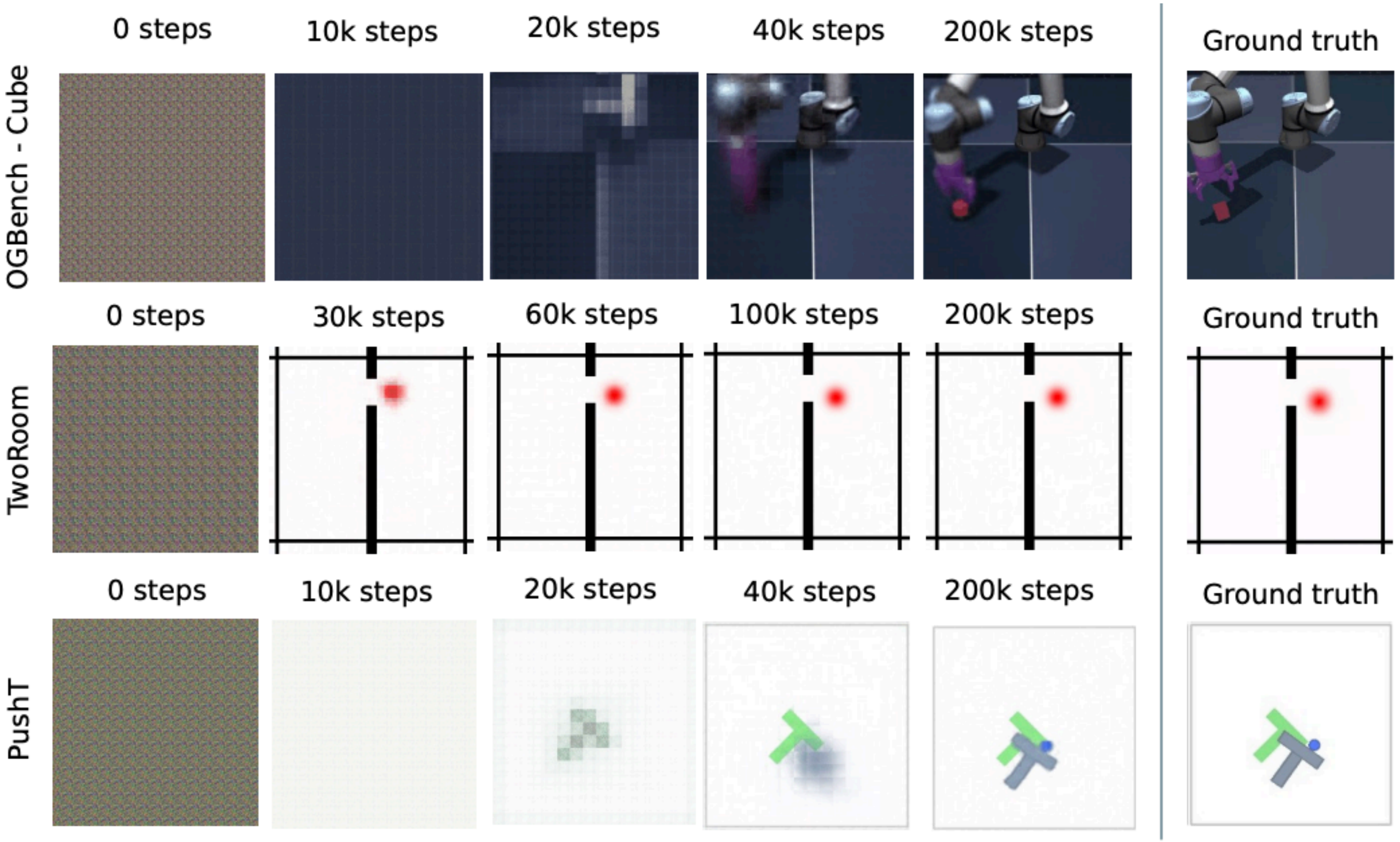
$$\text{SIGReg}_T(\mathbb{A}, \{f_{\boldsymbol{\theta}}(\boldsymbol{x}_n)\}_{n=1}^{N}) \triangleq \frac{1}{|\mathbb{A}|} \sum_{\boldsymbol{a} \in \mathbb{A}} T(\{\boldsymbol{a}^\top f_{\boldsymbol{\theta}}(\boldsymbol{x}_n)\}_{n=1}^{N}), \quad \text{(SIGReg)}$$

where we recommend the Epps-Pulley test (Section 4.2.3) for T.

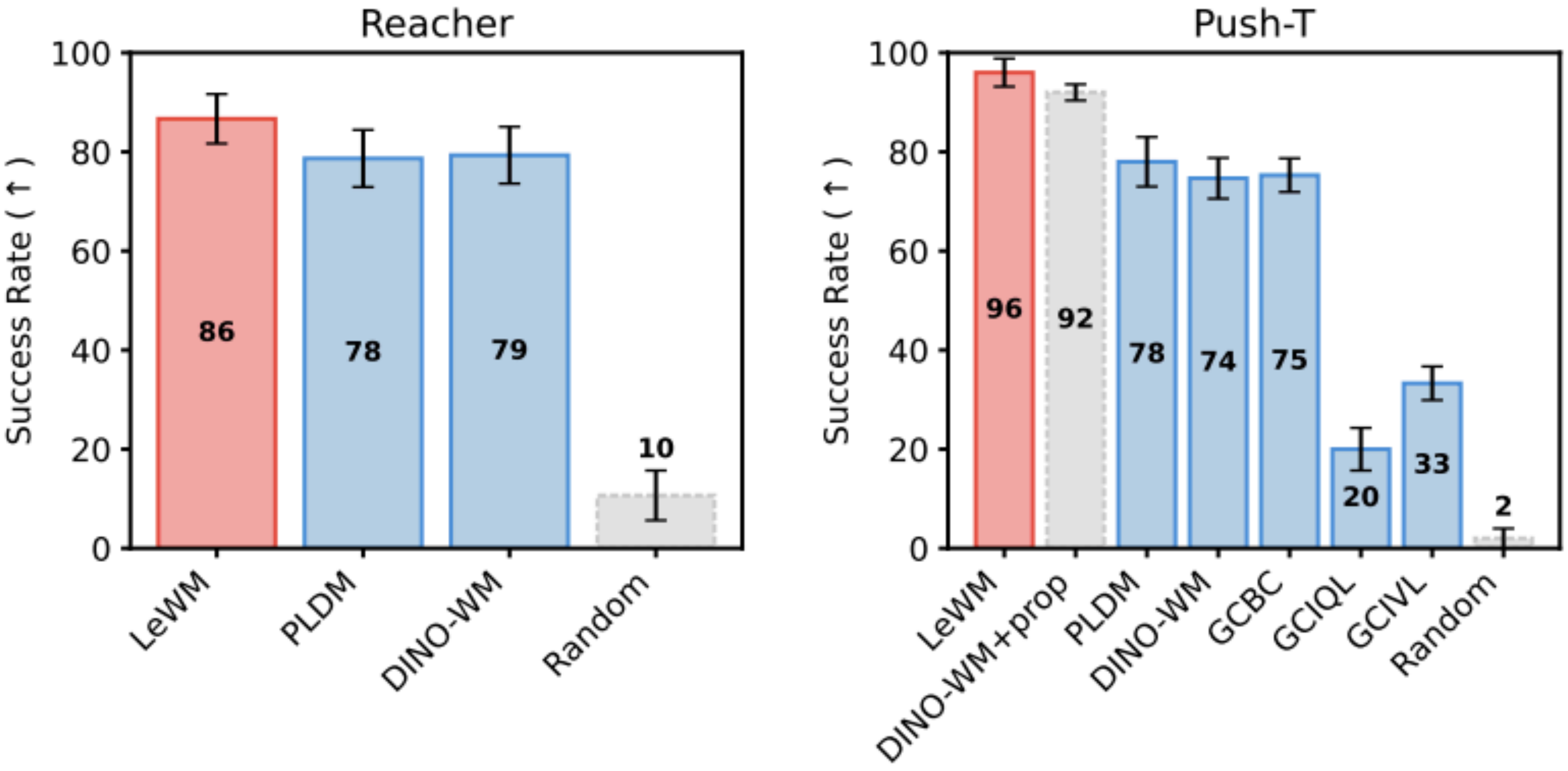
LeWorldModel



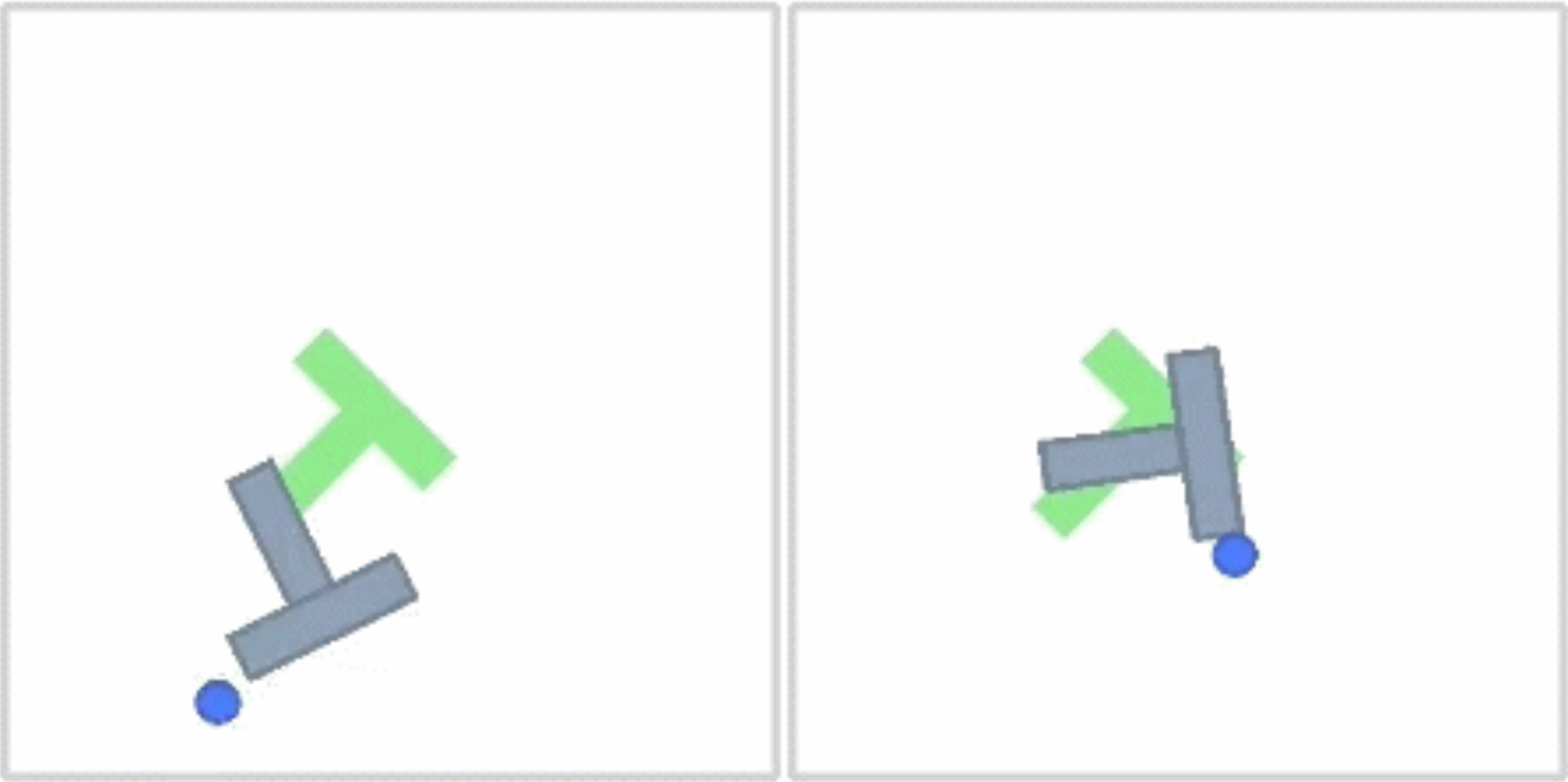
LeWorldModel

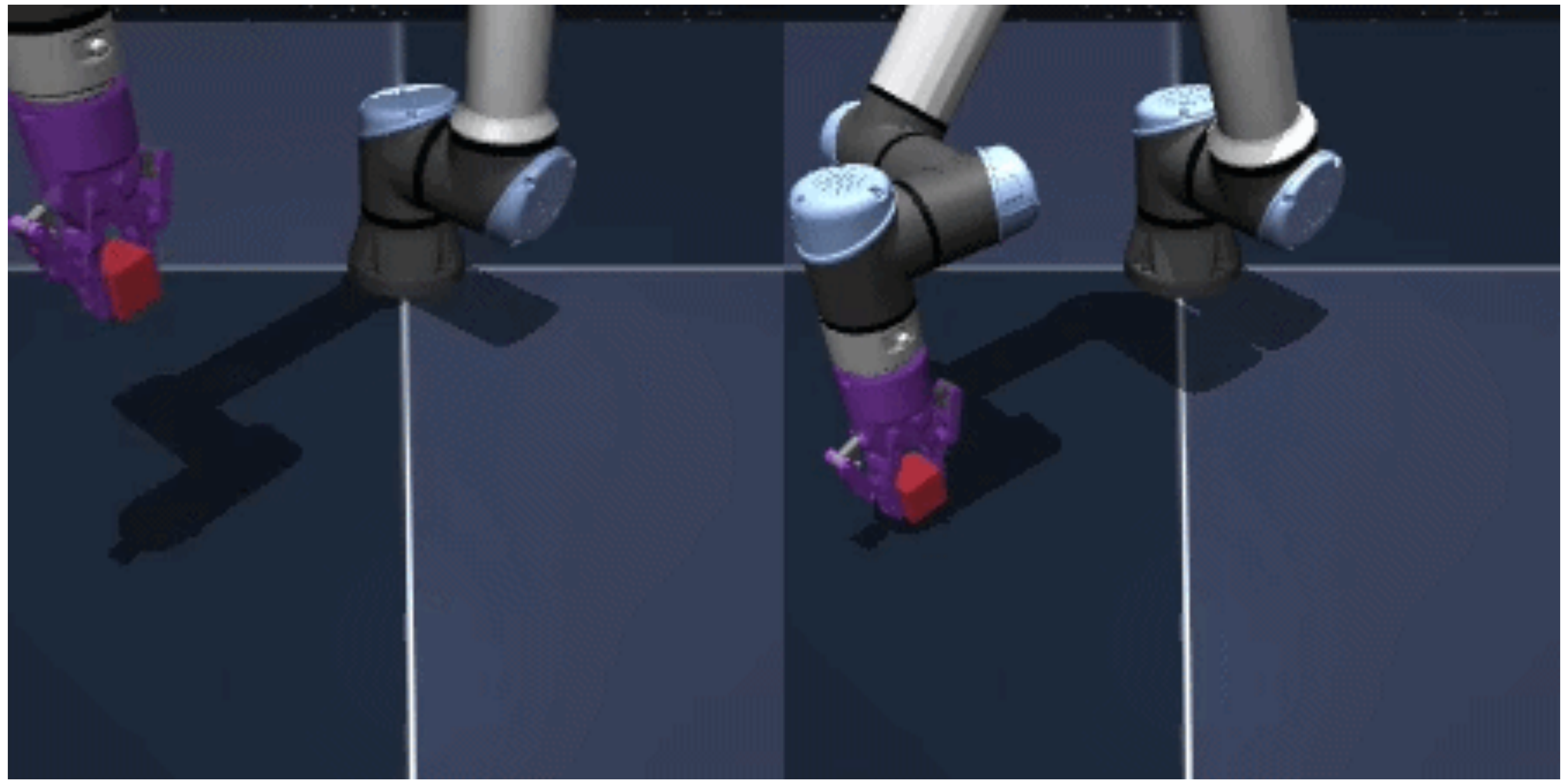


LeWorldModel

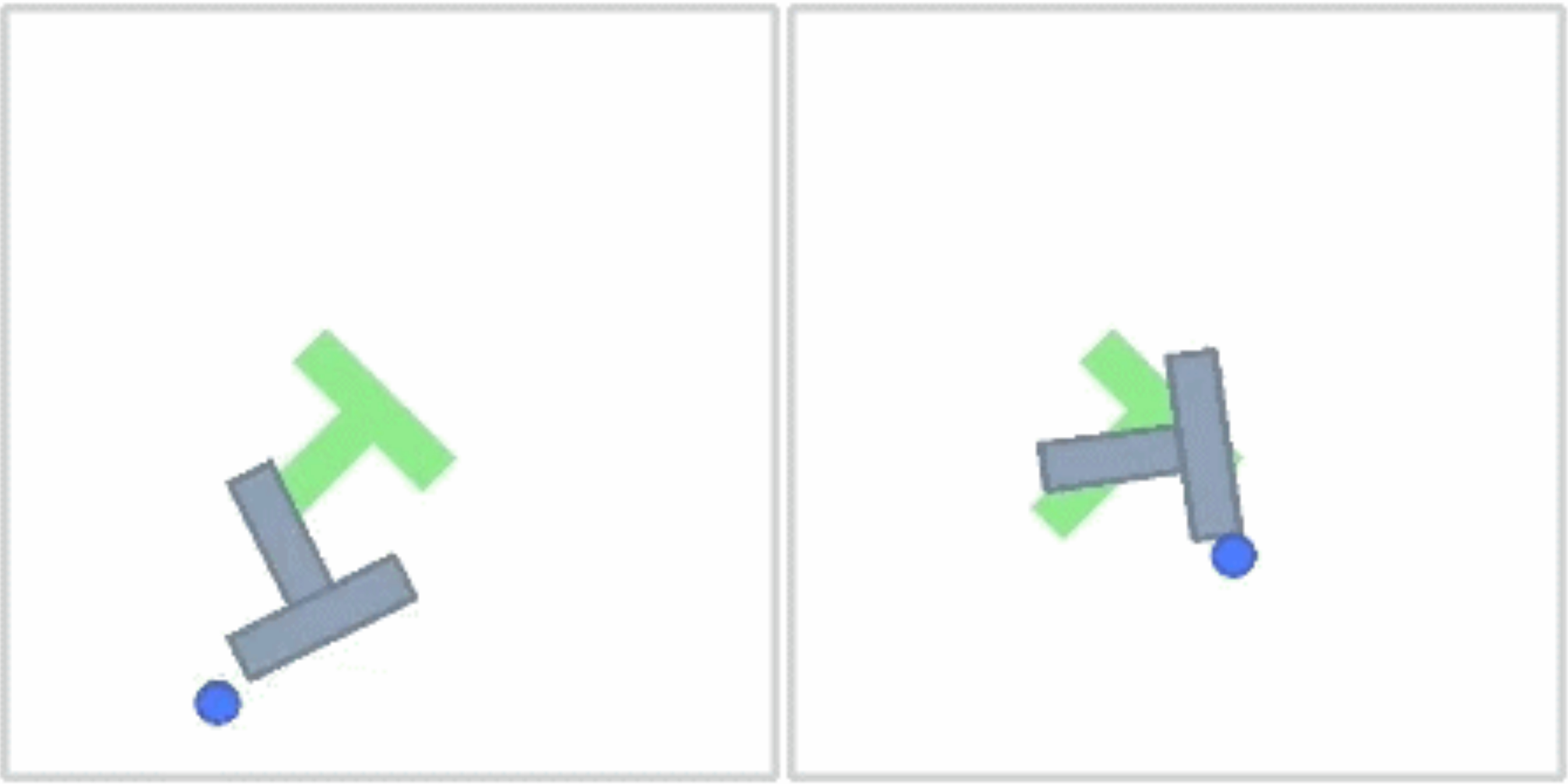


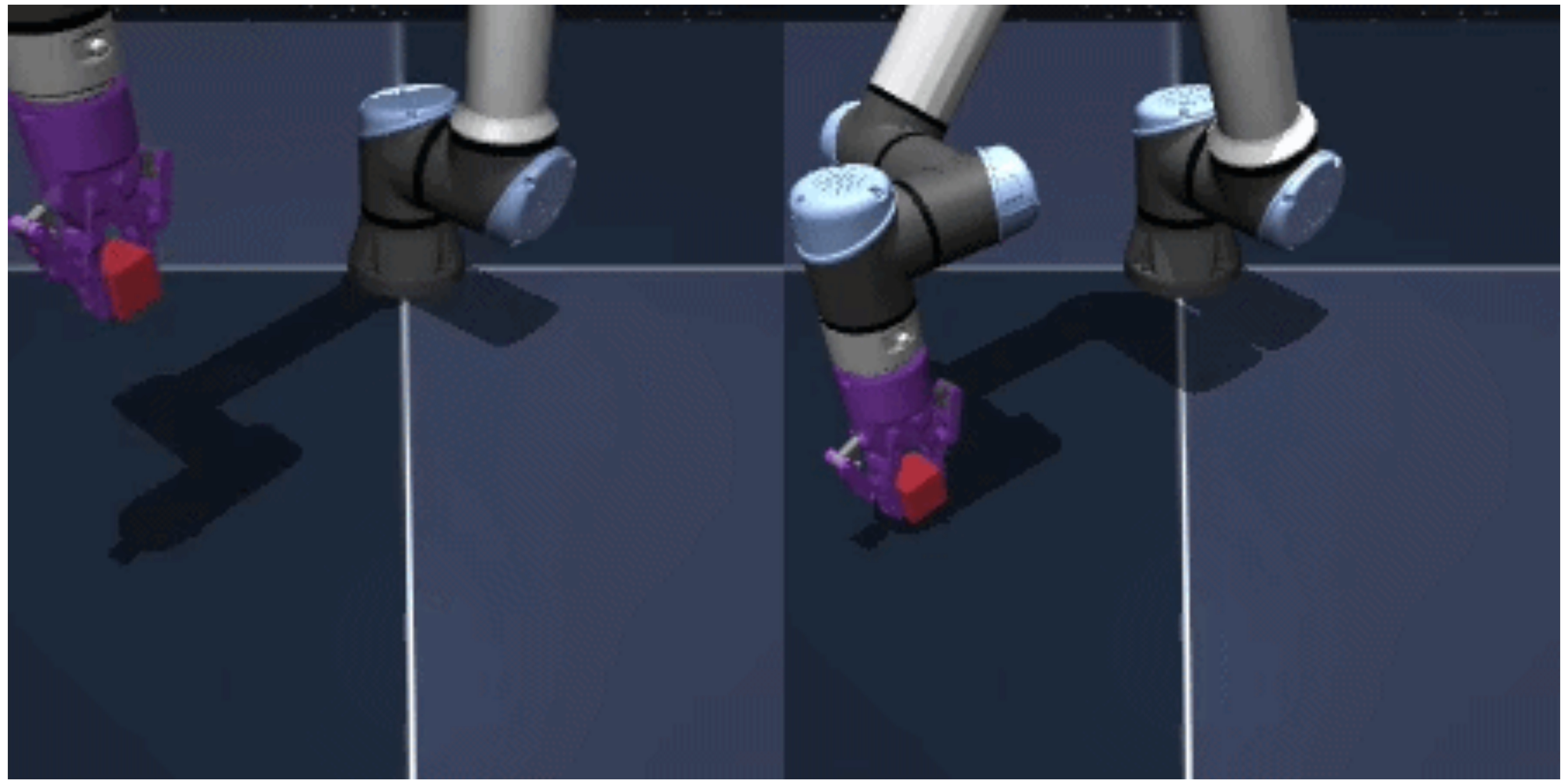
LeWorldModel





LeWorldModel





Thank you!
See you Monday!