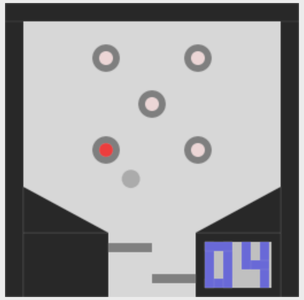






Learning Causal Variables from Temporal Observations

Phillip Lippe
PhD Student, University of Amsterdam

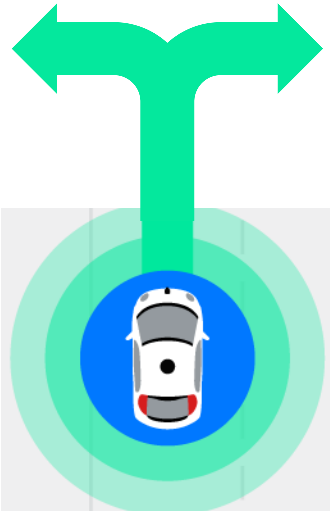
21. February 2023

Introduction
Representation Learning

Vision

Neural Network

?



Autonomous driving

Not interpretable
Unknown robustness
…

Figure credits:
[1] Waymo tech block, 2017
[2] Cordts et al., The Cityscapes dataset. CVPR 2016.

Introduction
Representation Learning

Vision

Neural Network

Structured Representation

AD: Human guidance what to model, causal factors

Interpretable

Generalizable

Robust

Reasoning-oriented

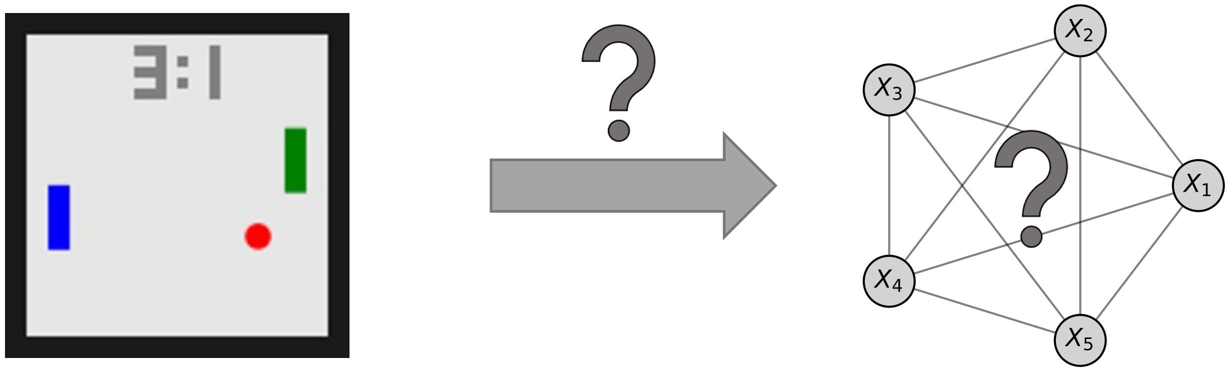
Structure learned?

Causal Representation Learning

Figure credits:
[1] Waymo tech block, 2017
[2] Cordts et al., The Cityscapes dataset. CVPR 2016.

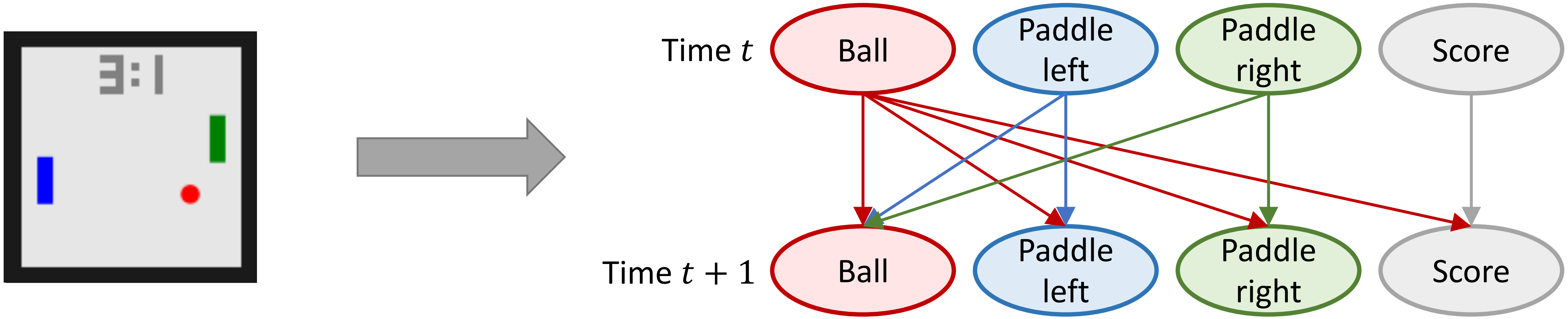
Causal Representation Learning

- Given high-dimensional observations of a (dynamical) system, what is its latent causal structure?
  - Causal variables
  - Their cause-effect relations

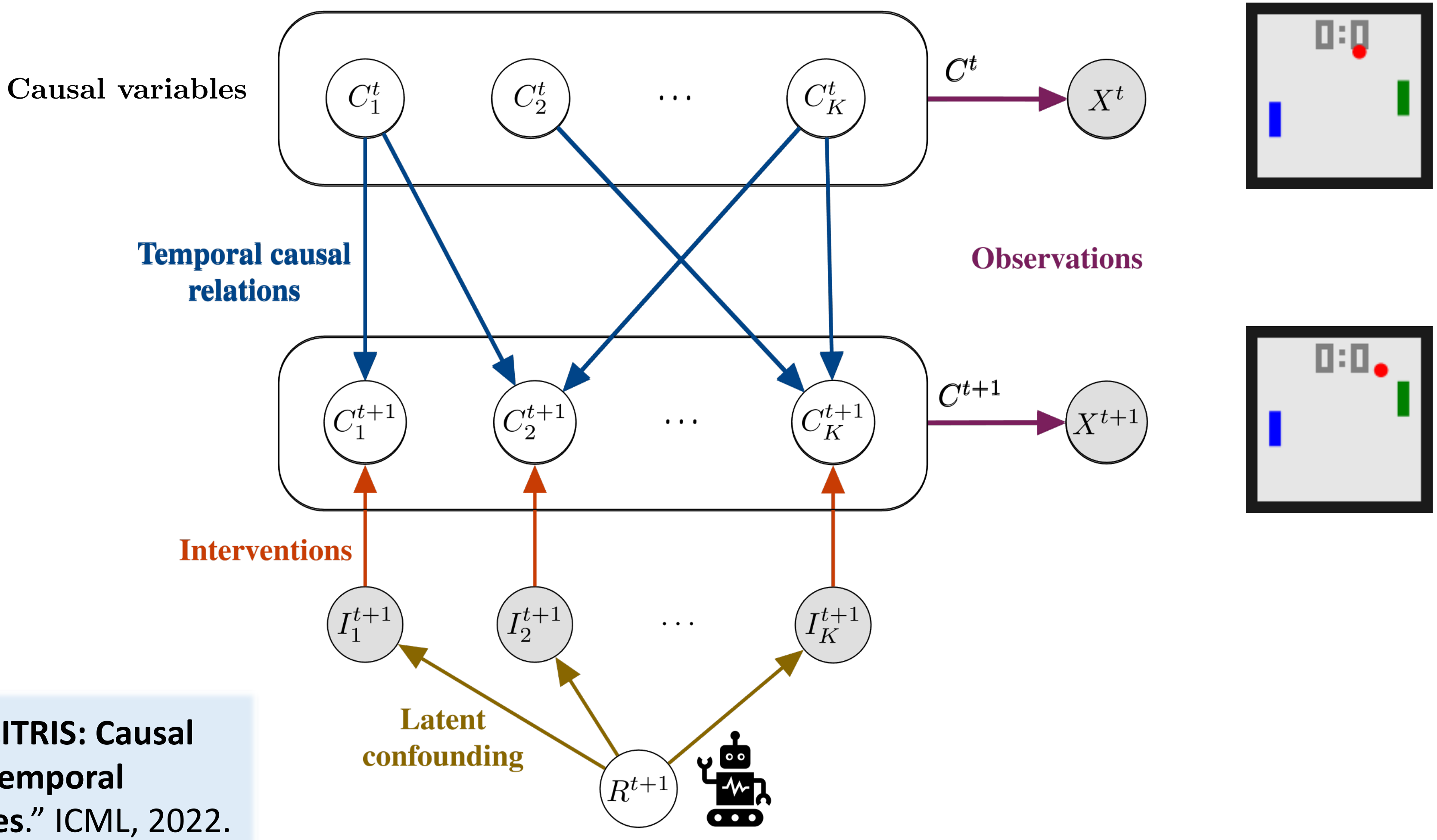


Causal Representation Learning

- Given high-dimensional observations of a (dynamical) system, what is its latent causal structure?
  - Causal variables
  - Their cause-effect relations

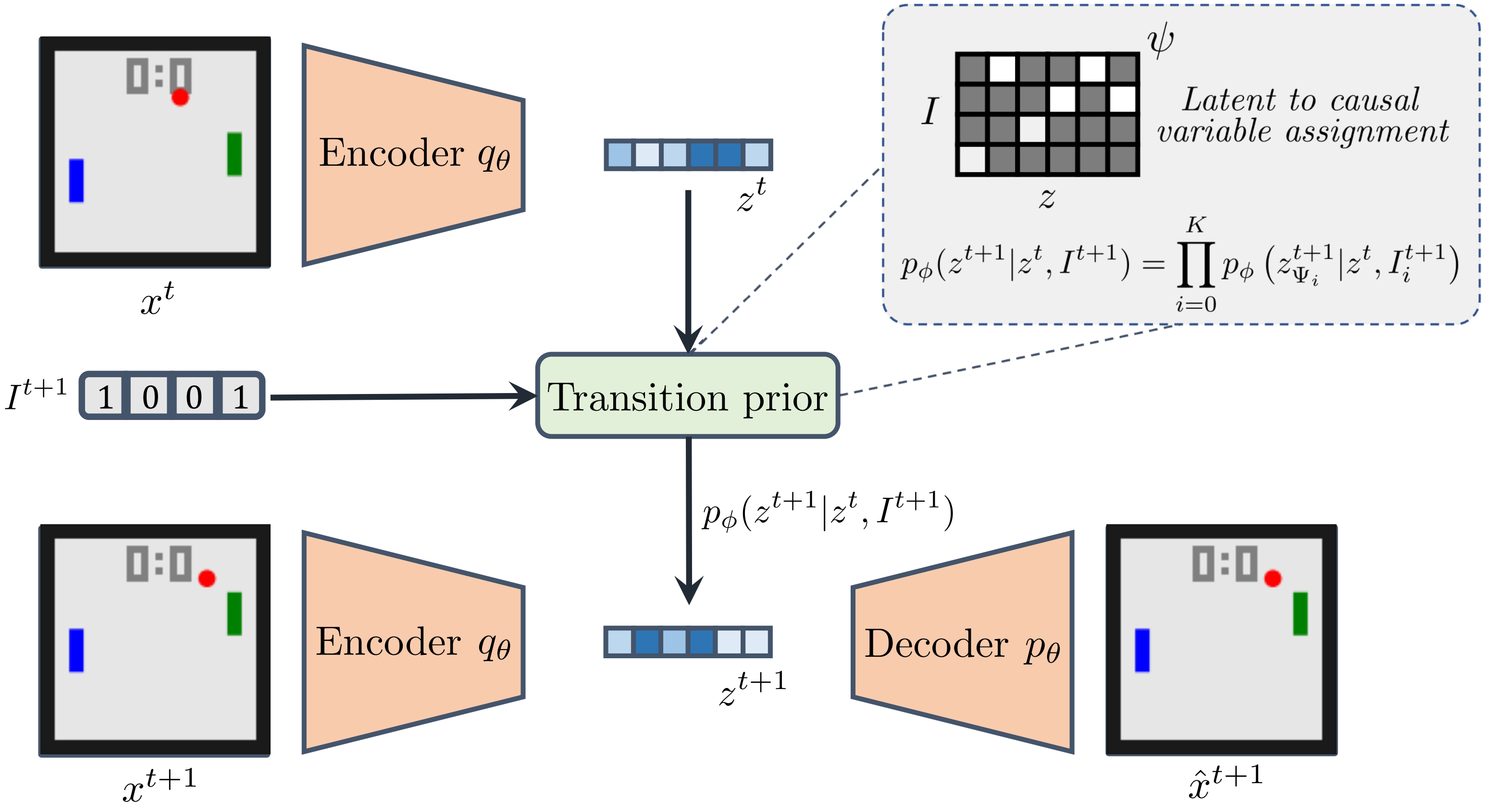


CITRIS: Causal Identifiability from Temporal Intervened Sequences
Setup

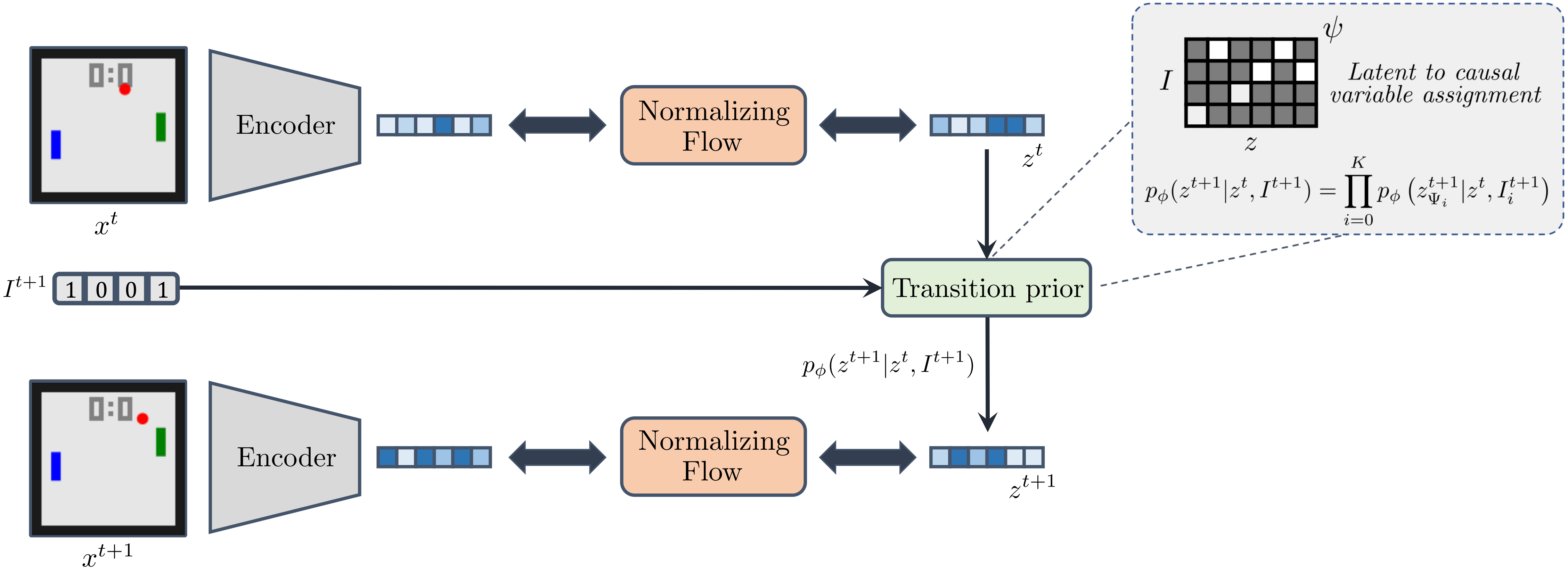


Lippe, Phillip et al. “**CITRIS: Causal Identifiability from Temporal Intervened Sequences.**” ICML, 2022.

CITRIS Architecture
CITRIS-VAE

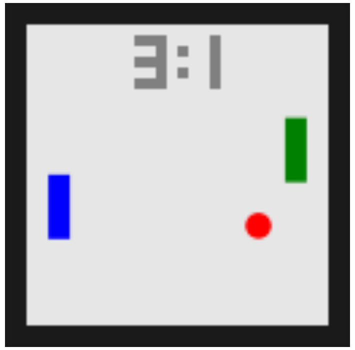


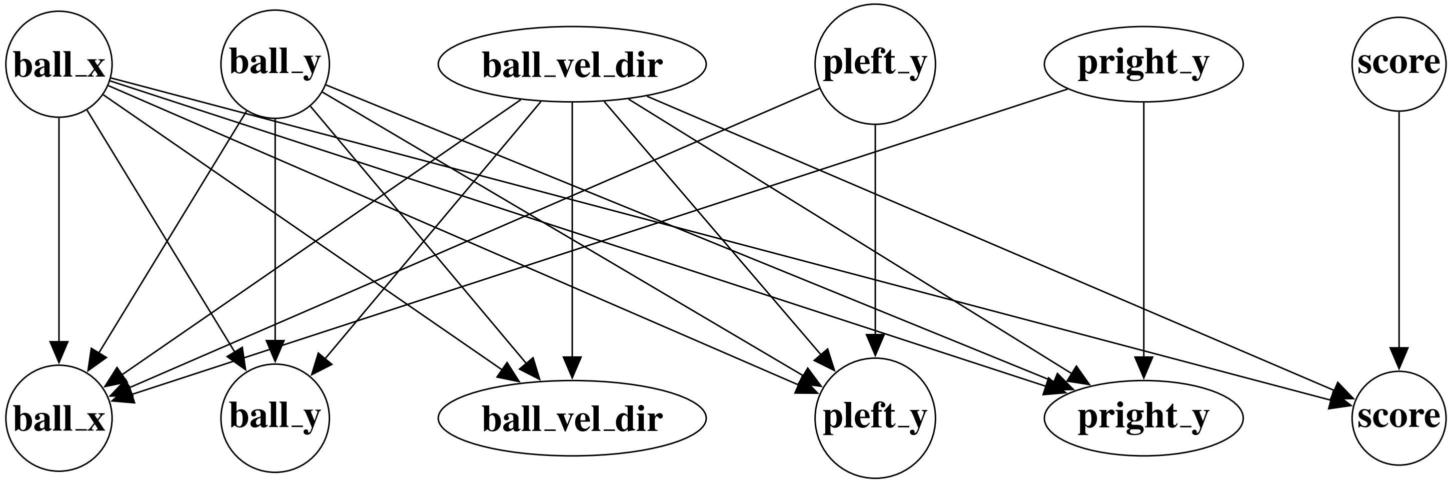
CITRIS Architecture
CITRIS-NF



CITRIS Experiments
Pong

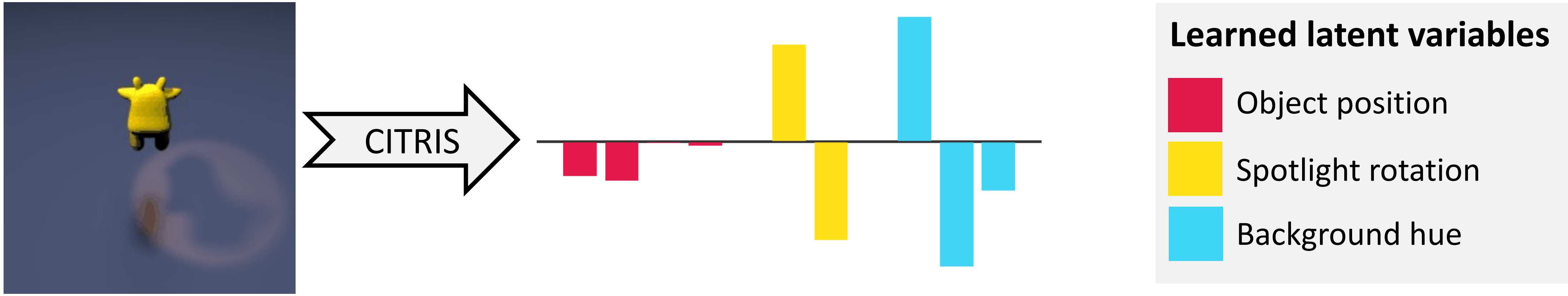
- CITRIS identifies the causal variables accurately
- Identified cause-effect relations closely follow ground truth





CITRIS Experiments

Visualizing the latent space



Novel combinations of causal factors

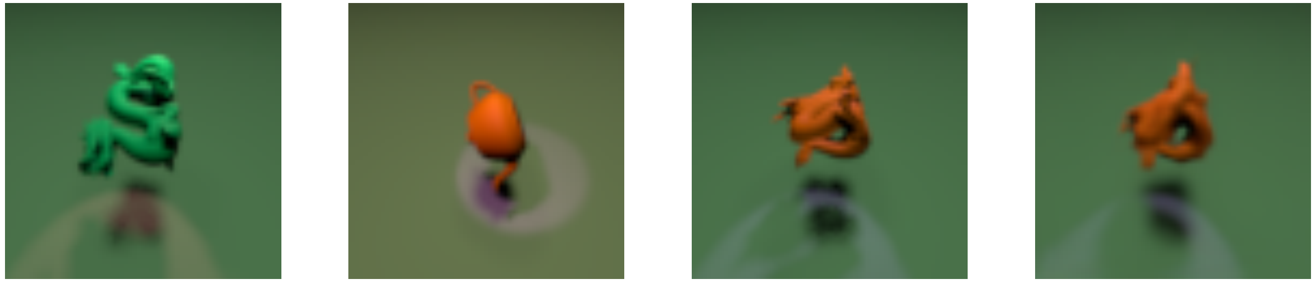
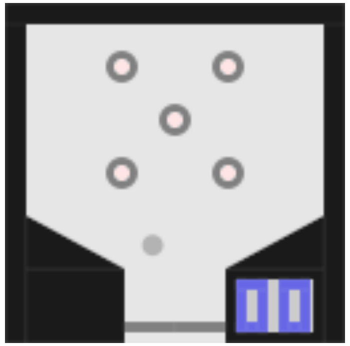
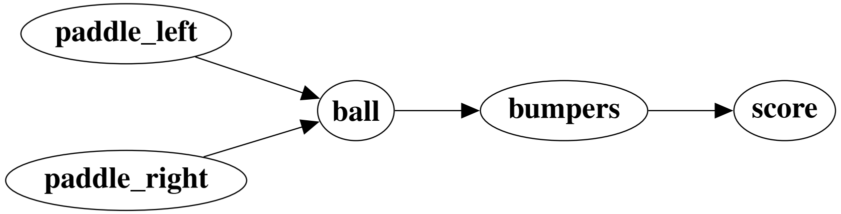
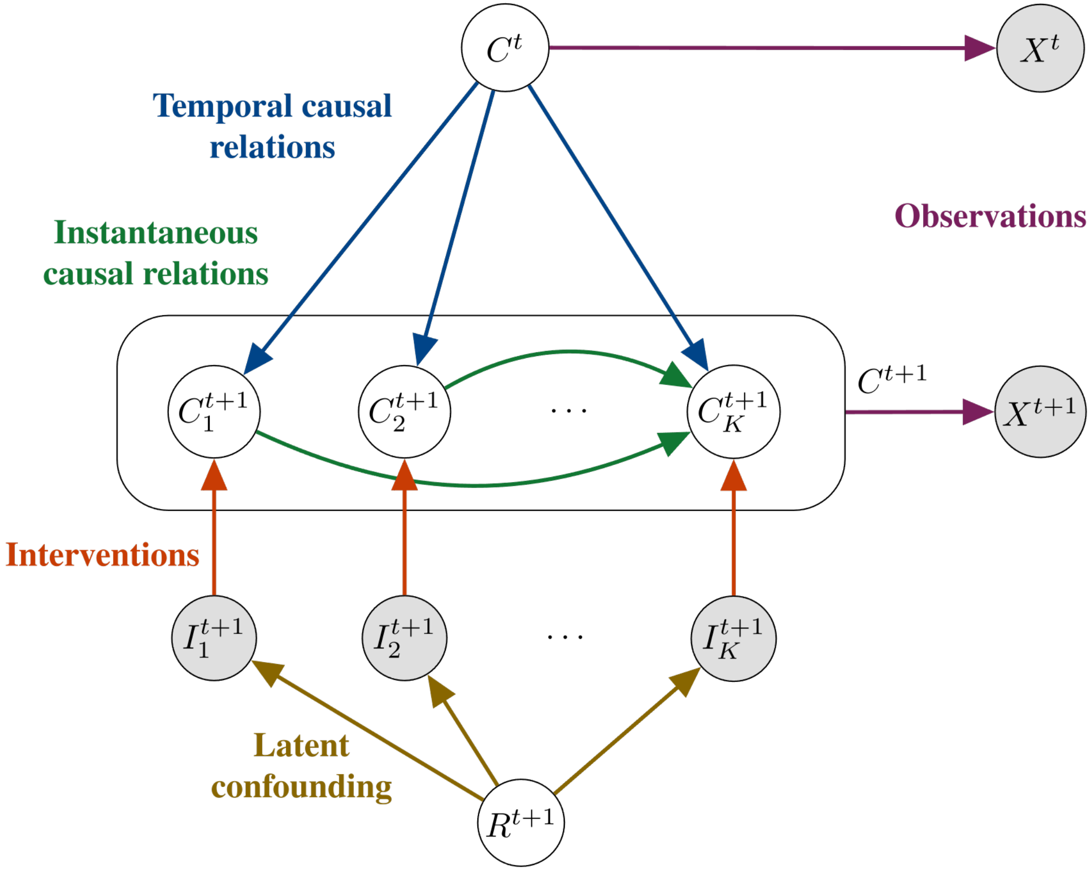


Image 1 Image 2 Ground Truth Prediction

Causal Representation Learning for Instantaneous Effects







Lippe, Phillip et al. "**Causal Representation Learning for Instantaneous and Temporal Effects**." ICLR, 2023.

Summary

- **Causal Representation Learning** aims to learn generalizable, robust representations of causal variables in an environment
- **CITRIS** identifies causal variables in variety of environments by information about interventions
- Allows for interpretable, controllable latent spaces
- Opportunity for learning representations in complex, interactive environments like Embodied AI

Figure credit: [1] Szot, Andrew, et al. "Habitat 2.0: Training home assistants to rearrange their habitat." NeurIPS 2021.

References

Phillip Lippe, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco Cohen, and Efstratios Gavves. "**CITRIS: Causal Identifiability from Temporal Intervened Sequences**." In International Conference on Machine Learning (ICML). PMLR, 2022.

Phillip Lippe, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco Cohen, and Efstratios Gavves. ” **Causal Representation Learning for Instantaneous and Temporal Effects.**" In International Conference on Learning Representations (ICLR), 2023.

Johann Brehmer, Pim de Haan, Phillip Lippe, Taco Cohen. “**Weakly supervised causal representation learning**.“ In Advanced in Neural Information Processing Systems (NeurIPS), 2022.

Slides and Papers