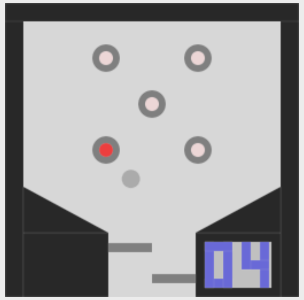


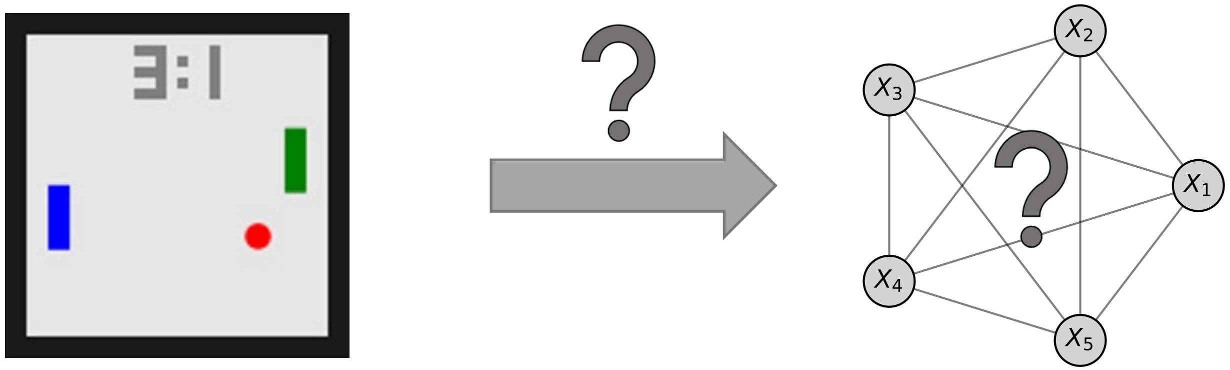

# Learning Causal Variables from Temporal Observations

Phillip Lippe

04. October 2022

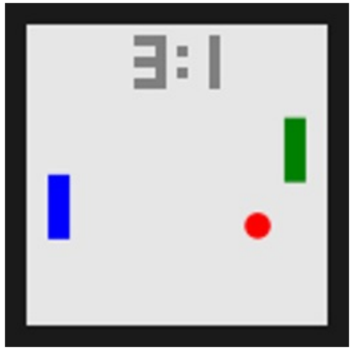
# Causal Representation Learning

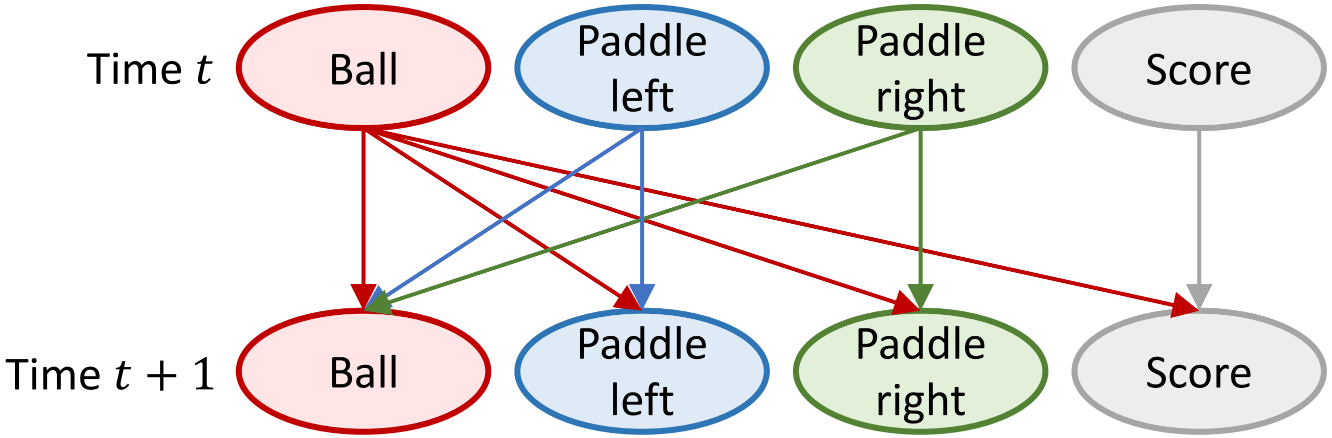
- Given high-dimensional observations of a (dynamical) system, what is its latent causal structure?
- Crucial for reasoning, planning, generalization, identifying cause-effect relations, etc.



# Causal Representation Learning

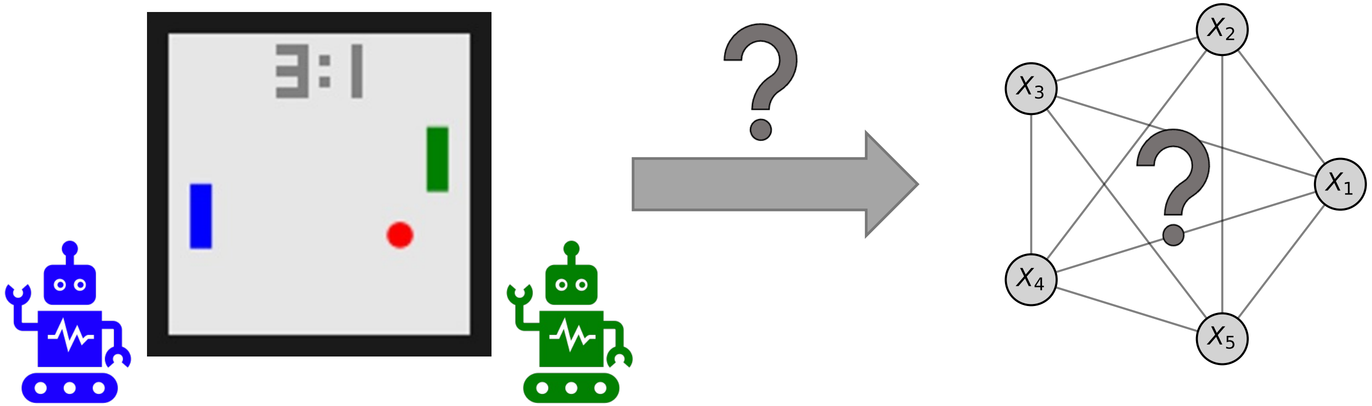
- Given high-dimensional observations of a (dynamical) system, what is its latent causal structure?
- Crucial for reasoning, planning, generalization, identifying cause-effect relations, etc.



# Causal Representation Learning
## Challenges

- High-dimensional input ↔ low-dimensional causal system
- Causal variables depend on each other
- Multiple (non-)causal representations can describe the same system
- Is a ‘causal’ representation unique?



# Causal Representation Learning
## Forms

### Counterfactual CRL

- Pairs of images where only a subset of variables change
- Requires a lot of control over system; not possible in real world (Pearl, 2009)

Examples: [Brehmer et al., 2022; Locatello et al., 2020; von Kügelgen et al., 2021; Ahuja et al., 2022]



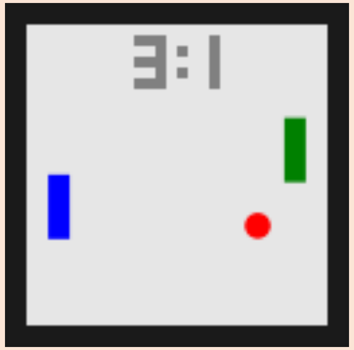




### Temporal CRL

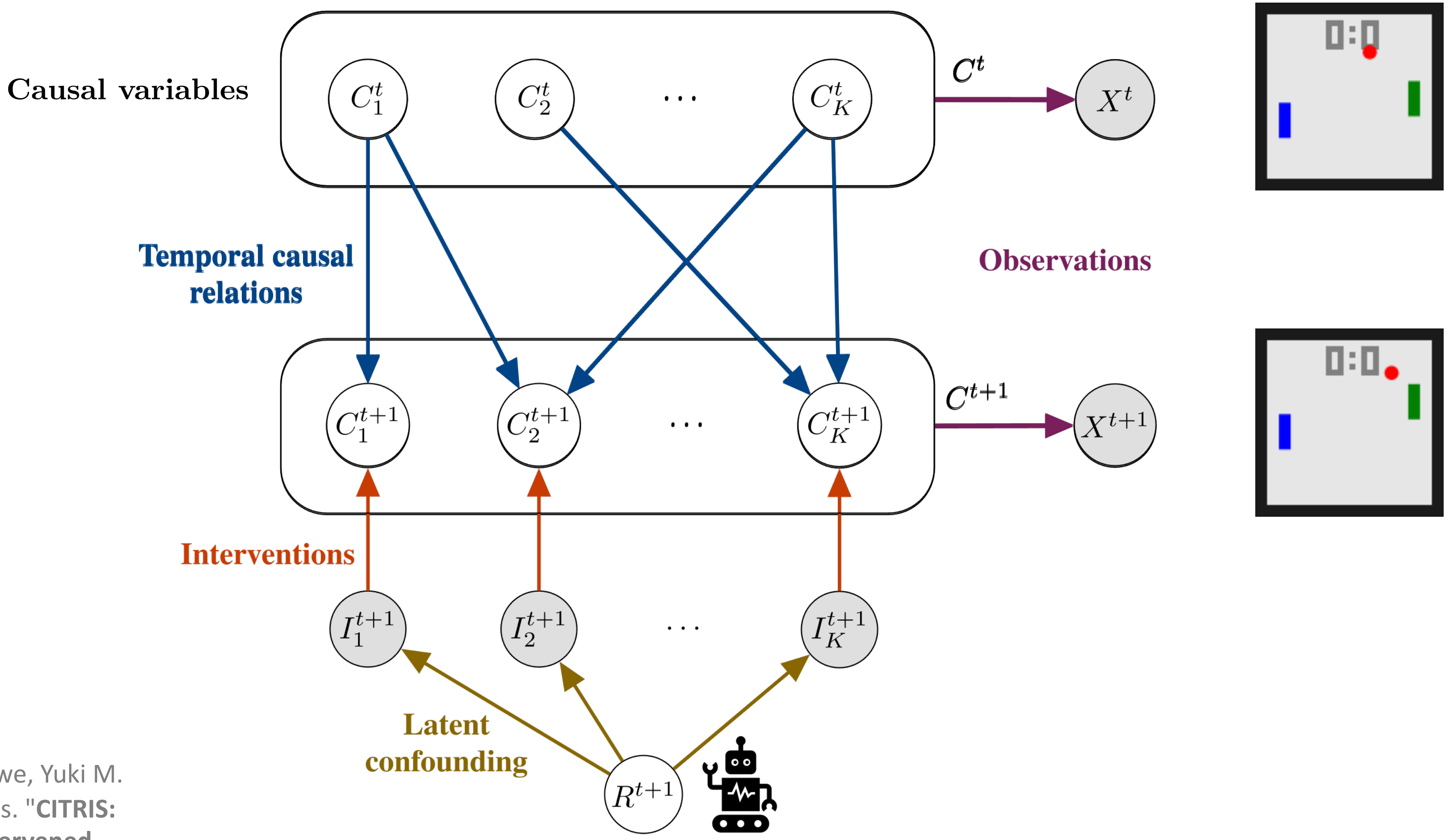
- Temporal sequences; all causal variables evolve over time
- Common RL environments
- Temporality gives strong bias

Examples: [Lippe et al., 2022ab; Lachapelle et al., 2022 ab; Yao et al., 2022ab; Khemakhem et al., 2020; Hyvärinen et al.; 2019]



# Causal Identifiability from Temporal Intervened Sequences

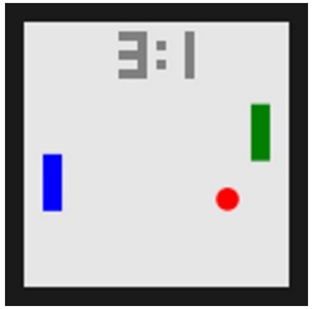
## Setup

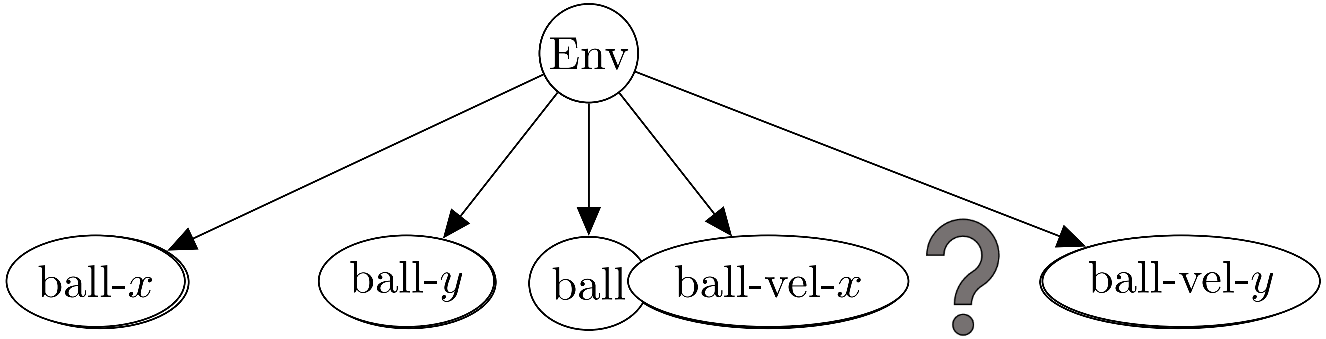



Lippe, Phillip, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco Cohen, and Efstratios Gavves. "**CITRIS: Causal Identifiability from Temporal Intervened Sequences**." In International Conference on Machine Learning, pp. 13557-13603. PMLR, 2022.

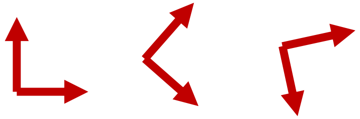
# Causal Identifiability from Temporal Intervened Sequences

## What is a Causal Variable?



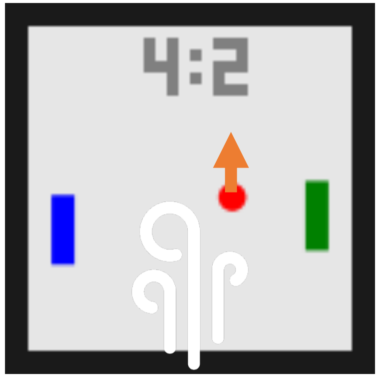

Abstraction allows for:

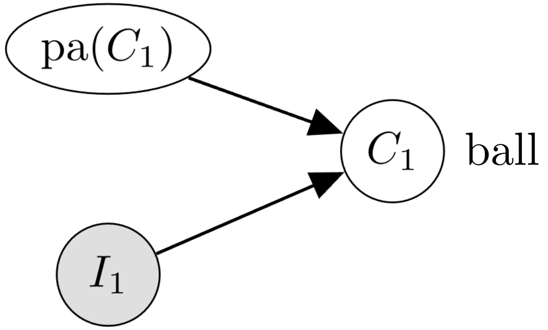


- Simpler graphs
- Fewer requirements to find it
- Scalability

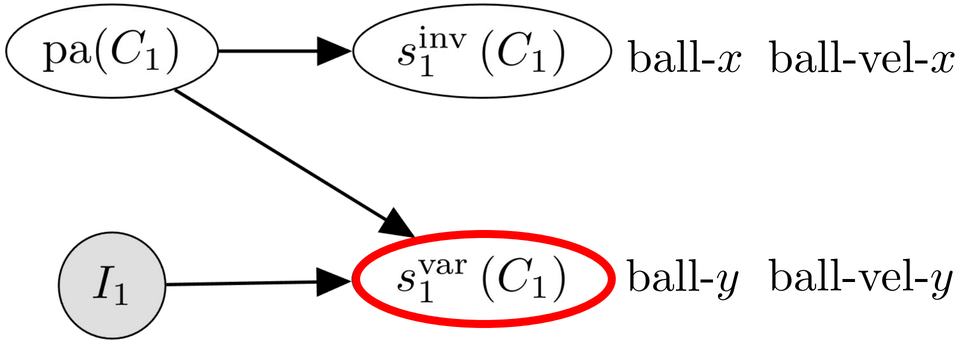
# Causal Identifiability from Temporal Intervened Sequences
## Minimal Causal Variables

- Abstraction ⇒ Multidimensional causal variables
- Identifying abstraction level ⇒ Interventions
- Augment causal graph with intervention targets
    - $I_1 = 1$ ⇒ Intervention on $C_1$
    - $I_1 = 0$ ⇒ Passively observing $C_1$
- Minimal causal variable $s_1^{\text{var}}(C_1)$: intervention-dependent part of a multidimensional causal variable
- Causal representation depends on the abilities of an agent/expert



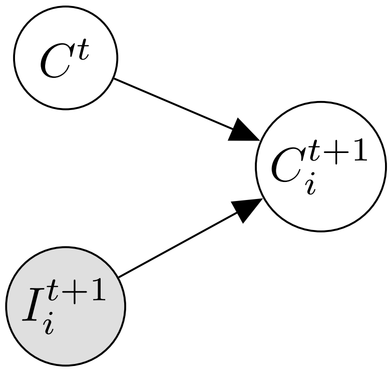

(a) Original causal graph of $C_1$



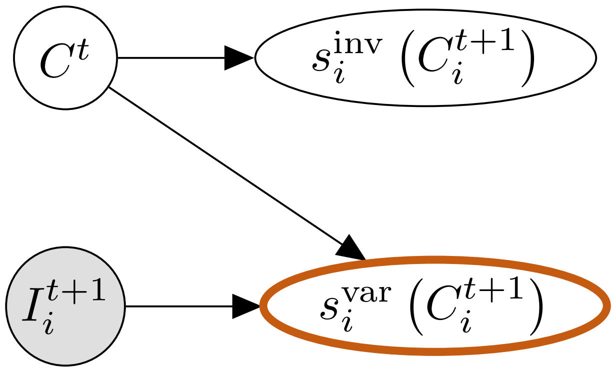

(b) Minimal causal split graph of $C_1$

# Causal Identifiability from Temporal Intervened Sequences
## Theoretical Results

- Main theoretical result: we can identify the ***minimal causal variables*** up to invertible, component-wise transformations if:
  - No intervention target $I_i^{t+1}$ is a deterministic function of any other
  - Following intervention design, $\lfloor \log_2 K \rfloor + 2$ experiments are sufficient for this [Lippe et al., 2022c]

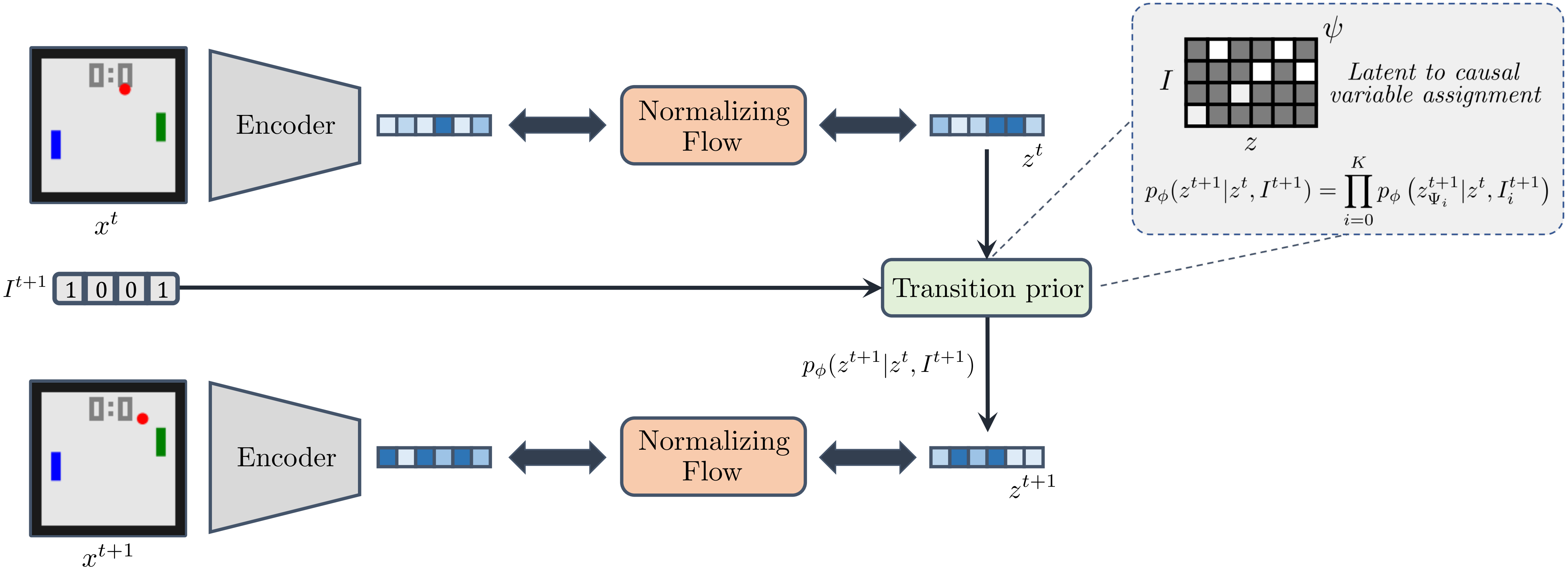



(a) Original causal graph of $C_i$



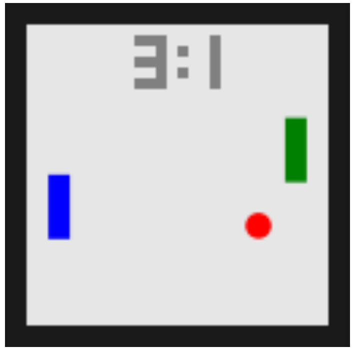

(b) Minimal causal split graph of $C_i$

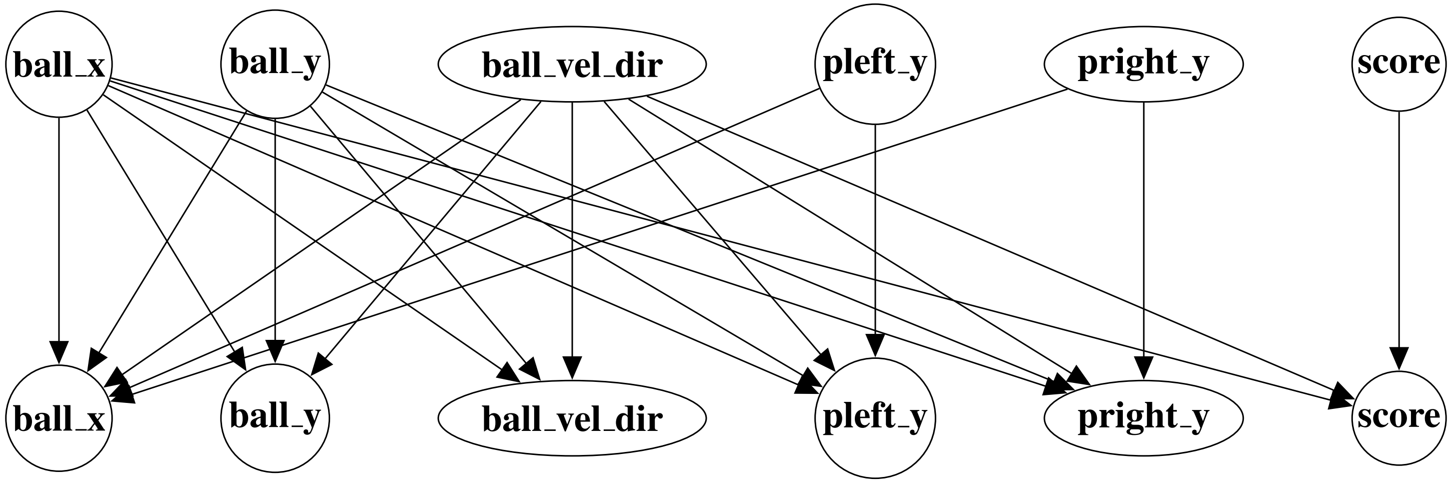
# CITRIS Architecture
## CITRIS-NF



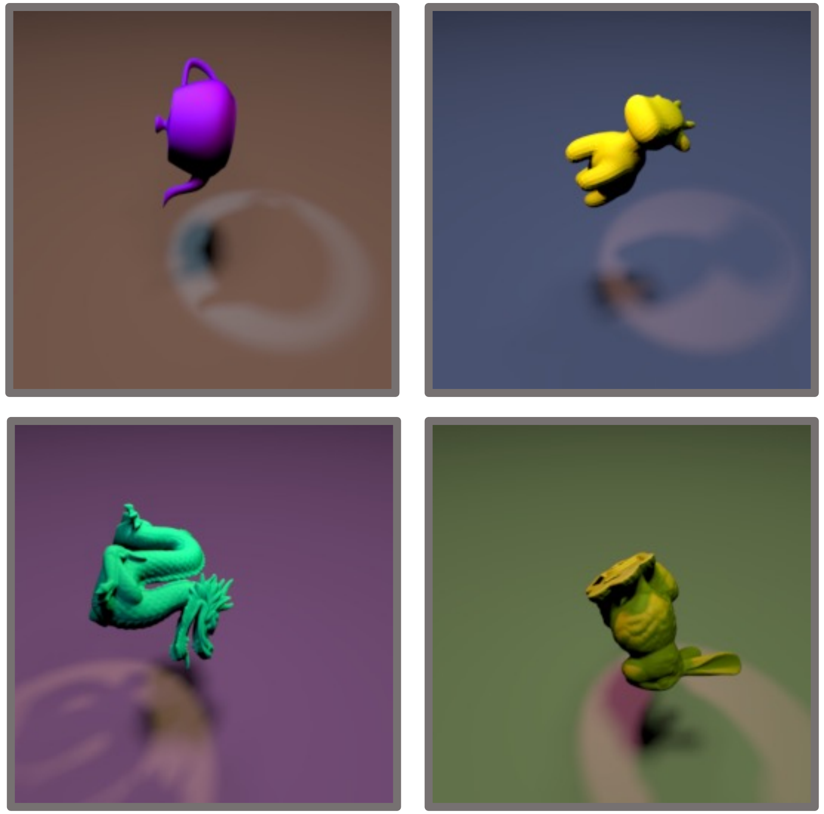
# CITRIS Experiments
## Pong

- CITRIS identifies the causal variables accurately
- Interventions on paddles changed their policy
- Assumption of Independent Causal Mechanisms



# CITRIS Experiments
## Temporal Causal3DIdent



**Causal Factors**

object-shape
object-position
object-hue
object-rotation
spotlight-hue
spotlight-rot
background-hue

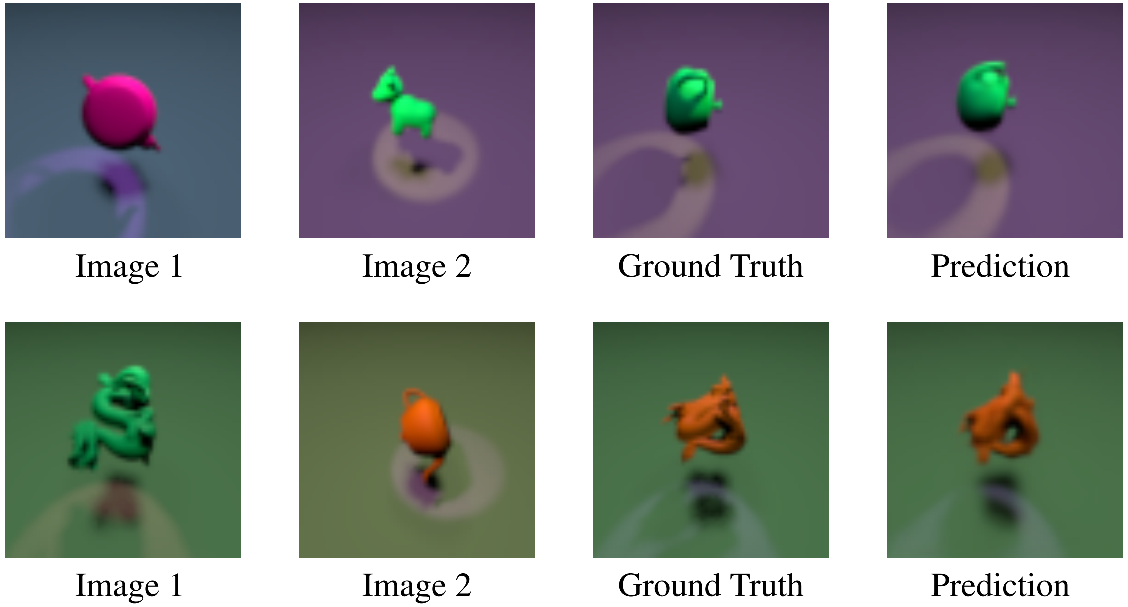
*categorical*
*continuous*
*angle / circular*

Zimmermann, Roland S., et al. "Contrastive learning inverts the data generating process." *ICML*, 2021.

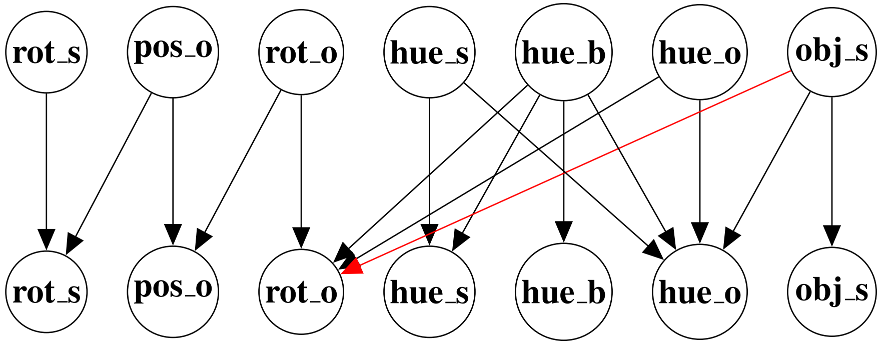
Von Kügelgen, Julius, et al. "Self-supervised learning with data augmentations provably isolates content from style." *NeurIPS,* 2021.

# CITRIS Experiments
## Temporal Causal3DIdent

**Novel combinations of causal factors**



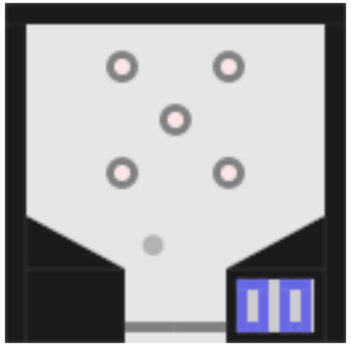

**Learned Causal Graph**

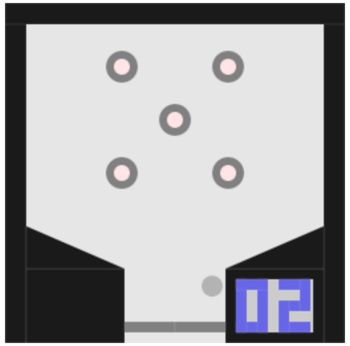


# Instantaneous Effects in Temporal Sequences

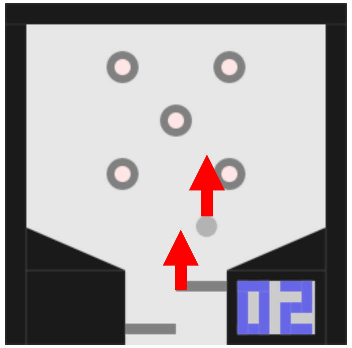
- Common assumption: time resolves causal effects
- But what about observations at low frame rates?

$\Rightarrow$ Instantaneous Effects!





time step $t$



time step $t + 1$

Lippe, Phillip, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco Cohen, and Efstratios Gavves. "**iCITRIS: Causal Representation Learning for Instantaneous Temporal Effects.**" First Workshop on Causal Representation Learning (CRL), UAI 2022.

# Instantaneous Effects in Temporal Sequences
## Challenges

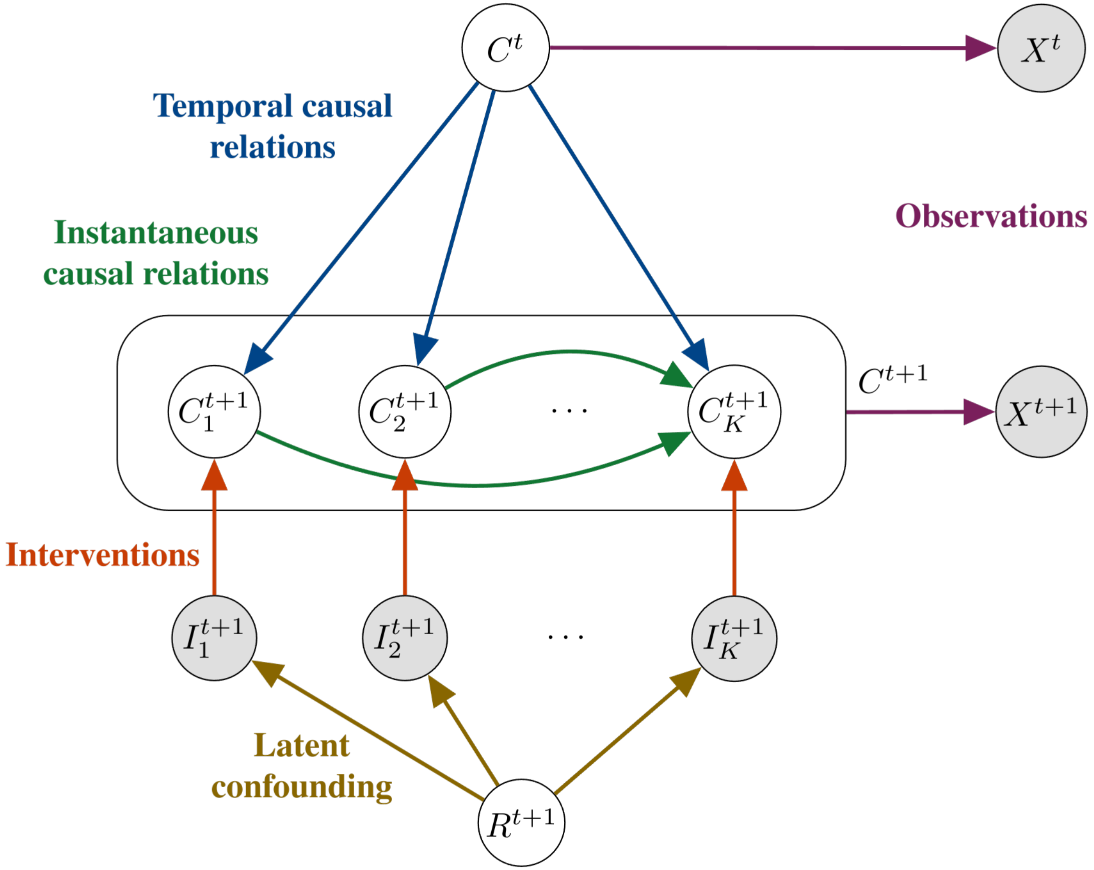
- Many more pitfalls, e.g.:

  $$p_1(C_1)p_2(C_2) \quad \text{vs} \quad p_1(C_1)\hat{p}_2(C_2 + C_1|C_1)$$

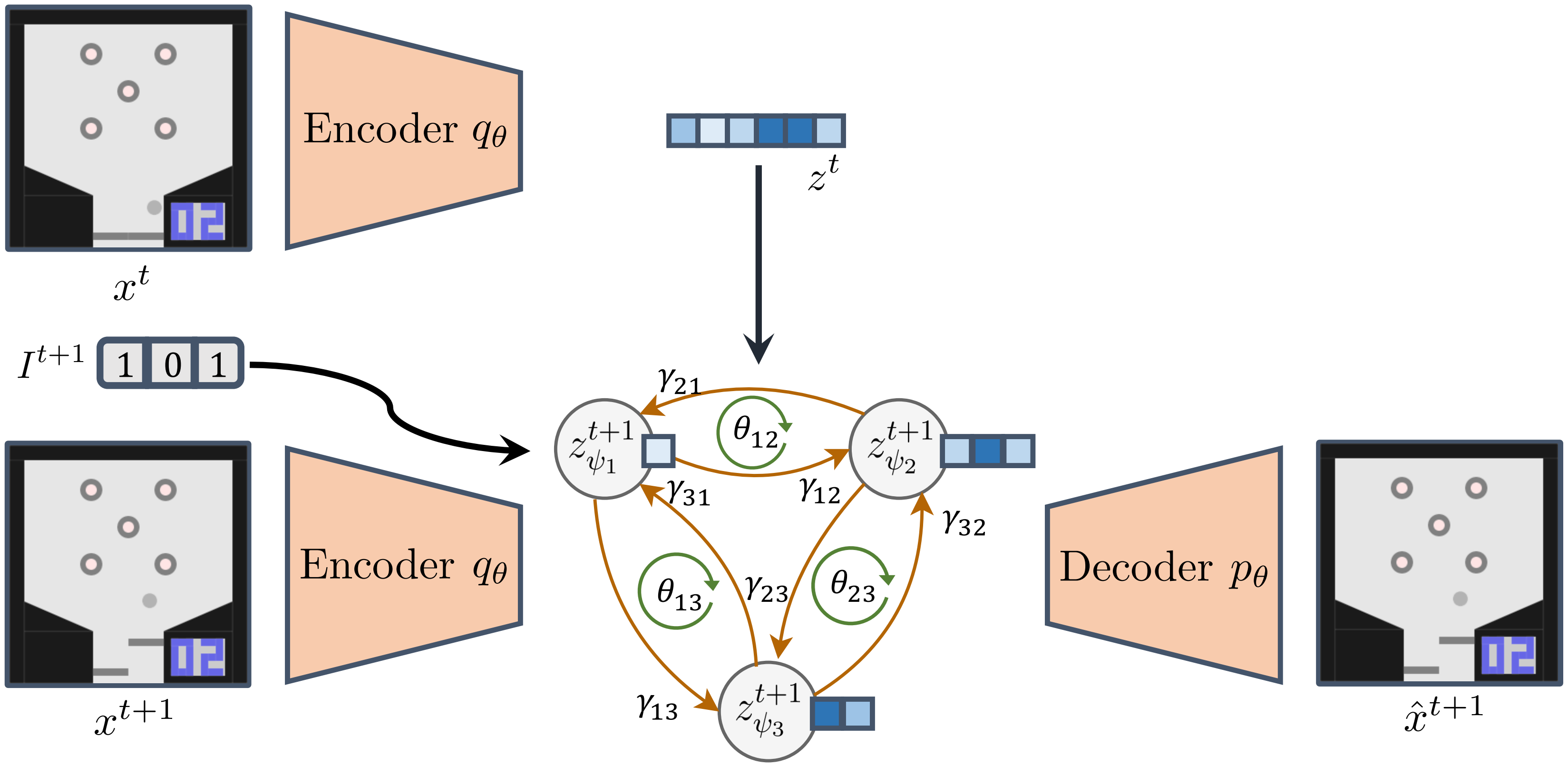
- Solution: *partially-perfect* interventions that remove instantaneous parents

  ⇒ Minimal causal variables become identifiable

- Chicken-and-egg situation:
  - Without graph, no causal variables
  - Without causal variables, no graph



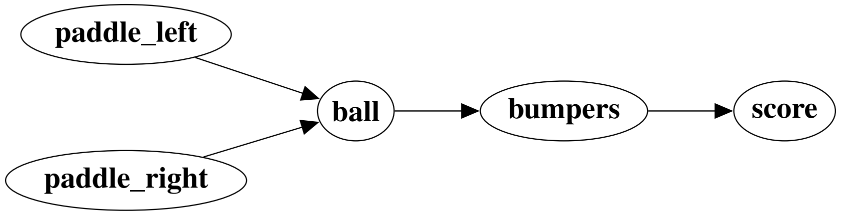
# iCITRIS: CRL for Instantaneous Temporal Effects
## Architecture

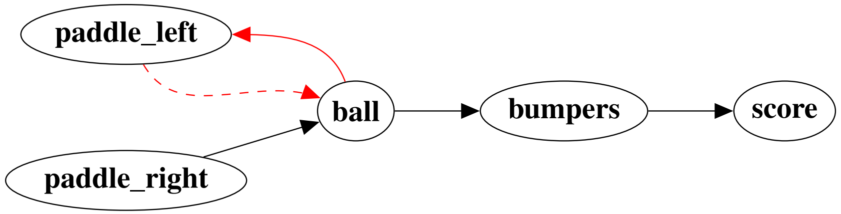


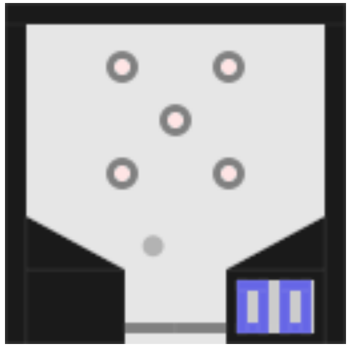
# iCITRIS: CRL for Instantaneous Temporal Effects

## Experiments

**Learned Causal Graphs**



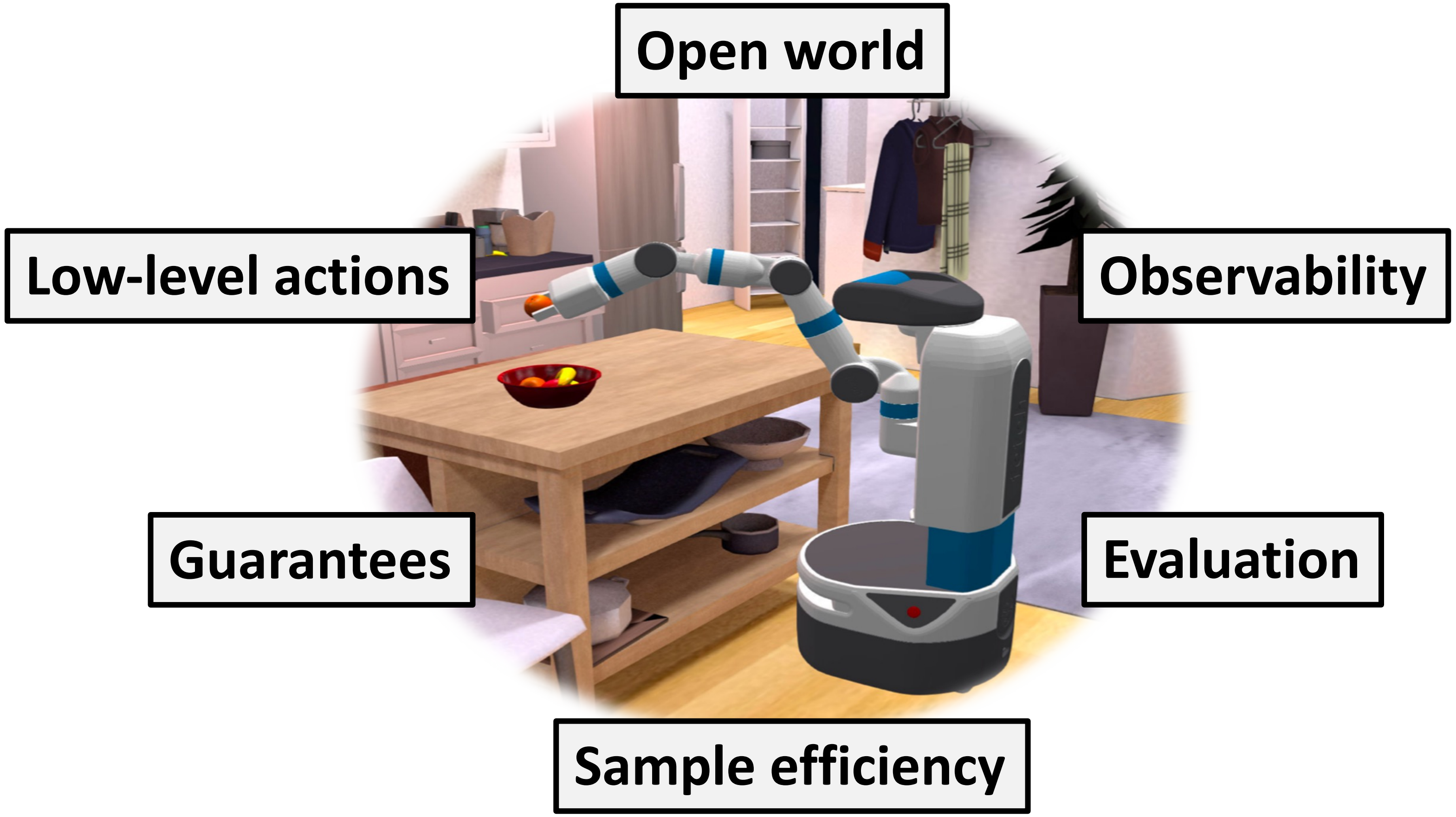



# Summary

- **CITRIS**: Identify multidimensional causal variables from temporal sequences with soft interventions and known intervention targets
- Identifies minimal causal variables, i.e., part of the variables that depends on interventions
- CITRIS-NF scales to visually complex scenes with pretrained autoencoder

- **iCITRIS**: Extension to instantaneous effects within a time step
- Need for partially-perfect interventions
- End-to-end learning with joint causal discovery and causal representation learning

# Challenges in CRL




Szot, Andrew, et al. "Habitat 2.0: Training home assistants to rearrange their habitat." NeurIPS 2021.

# References

[Lippe et al., 2022a] Lippe, Phillip, 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, pp. 13557-13603. PMLR, 2022.

[Lippe et al., 2022b] Lippe, Phillip, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco Cohen, and Efstratios Gavves. "**iCITRIS: Causal Representation Learning for Instantaneous Temporal Effects.**" First Workshop on Causal Representation Learning (CRL), UAI 2022.

[Lippe et al., 2022c] Lippe, Phillip, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco Cohen, and Efstratios Gavves. ”**Intervention Design for Causal Representation Learning.**" First Workshop on Causal Representation Learning (CRL), UAI 2022.

[Brehmer et al., 2022] Brehmer, Johann, Pim de Haan, Phillip Lippe, Taco Cohen. “**Weakly supervised causal representation learning**.“ Advances in Neural Information Processing Systems, NeurIPS 2022.

# References


Kartik Ahuja, Jason Hartford, and Yoshua Bengio. Weakly Supervised Representation Learning with Sparse Perturbations. In Advances in Neural Information Processing, NeurIPS 2022.

Aapo Hyvärinen, Hiroaki Sasaki, and Richard Turner. Nonlinear ICA Using Auxiliary Variables and Generalized Contrastive Learning. In Kamalika Chaudhuri and Masashi Sugiyama (eds.), Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, volume 89 of Proceedings of Machine Learning Research, pp. 859–868. PMLR, 2019

Ilyes Khemakhem, Diederik Kingma, Ricardo Monti, and Aapo Hyvarinen. Variational Autoencoders and Nonlinear ICA: A Unifying Framework. In Silvia Chiappa and Roberto Calandra (eds.), Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, volume 108 of Proceedings of Machine Learning Research, pp. 2207–2217. PMLR, 2020

Sebastien Lachapelle and Simon Lacoste-Julien. Partial Disentanglement via Mechanism Sparsity. arXiv preprint arXiv:2207.07732, 2022a

Sebastien Lachapelle, Pau Rodriguez, Rémi Le, Yash Sharma, Katie E Everett, Alexandre Lacoste, and Simon Lacoste-Julien. Disentanglement via Mechanism Sparsity Regularization: A New Principle for Nonlinear ICA. In First Conference on Causal Learning and Reasoning, 2022b

Francesco Locatello, Ben Poole, Gunnar Rätsch, Bernhard Schölkopf, Olivier Bachem, and Michael Tschannen. Weakly-Supervised Disentanglement Without Compromises. In Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event, volume 119 of Proceedings of Machine Learning Research, pp. 6348–6359. PMLR, 2020

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. Proceedings of the IEEE, 109(5):612–634, 2021

Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, Bernhard Schölkopf, Michel Besserve, and Francesco Locatello. Self-Supervised Learning with Data Augmentations Provably Isolates Content from Style. In Thirty-Fifth Conference on Neural Information Processing Systems, 2021

Weiran Yao, Guangyi Chen, and Kun Zhang. Learning Latent Causal Dynamics. arXiv preprint arXiv:2202.04828, 2022a

Weiran Yao, Yuewen Sun, Alex Ho, Changyin Sun, and Kun Zhang. Learning Temporally Causal Latent Processes from General Temporal Data. In International Conference on Learning Representations, 2022b