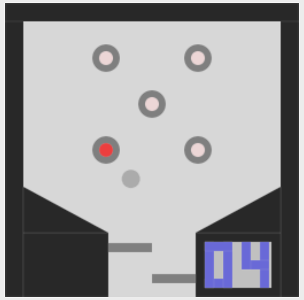


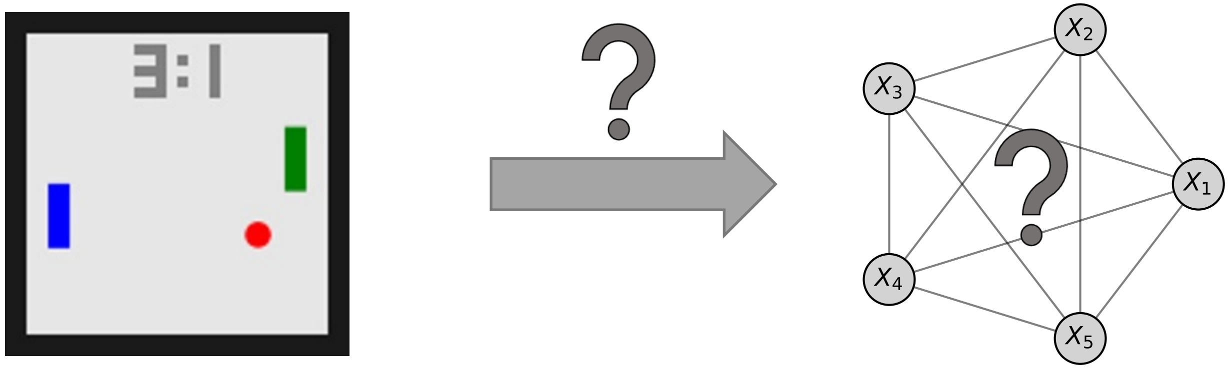
Learning Causal Variables from Temporal Sequences with Interventions

Phillip Lippe

05. August 2022

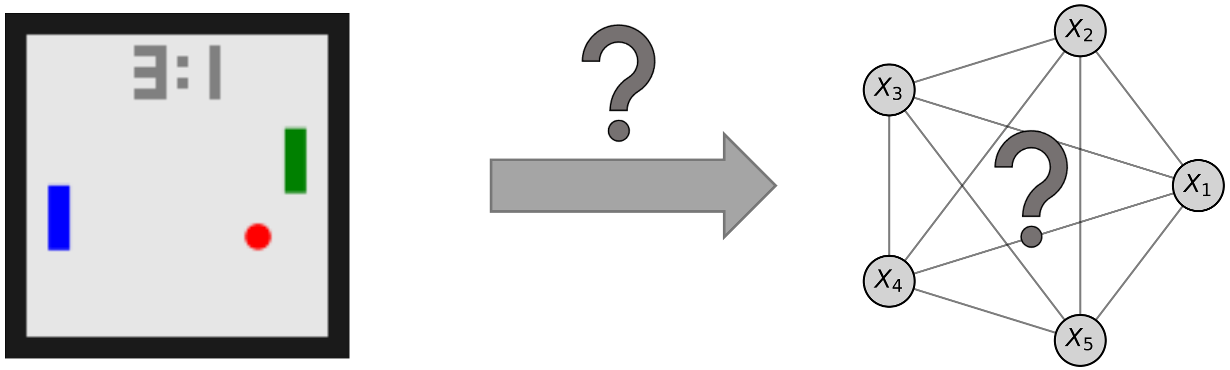
Causal Representation Learning

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



Causal Representation Learning
Why Temporal?

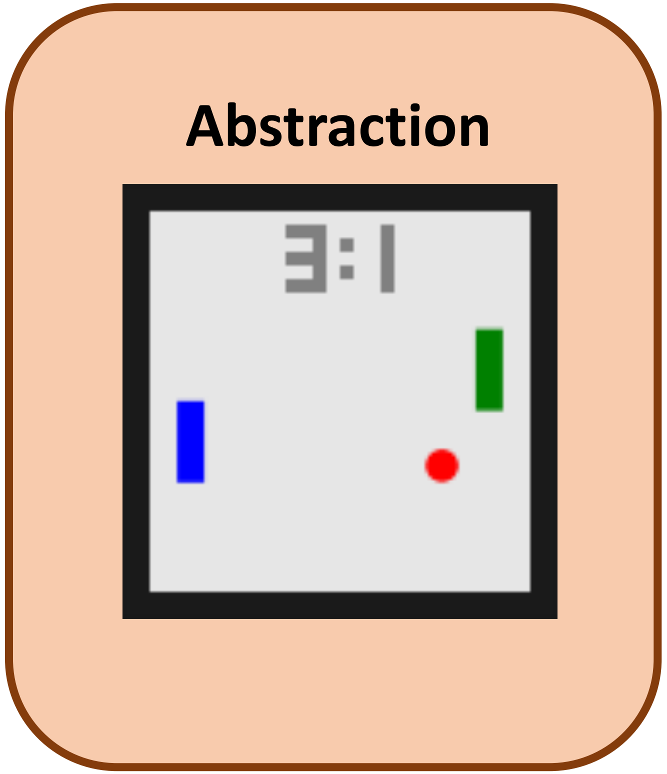
- Temporality gives strong bias
- Interact with an environment ⇒ see and reason about effect of intervention

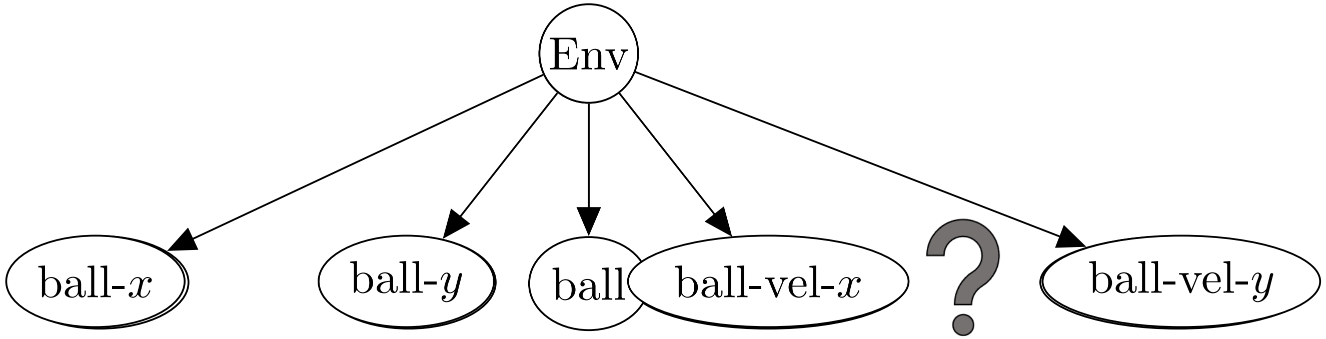


What is a Causal Variable?

Greff, Klaus, et al. "Kubric: A scalable dataset generator." *CVPR,* 2022.

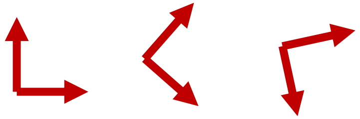
What is a Causal Variable?
Challenges





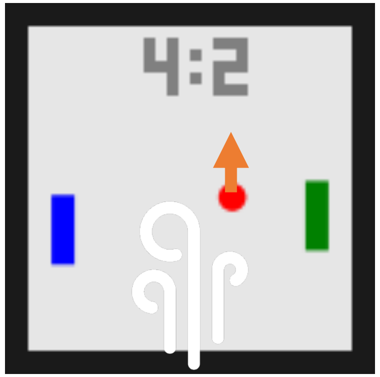
Abstraction allows for:

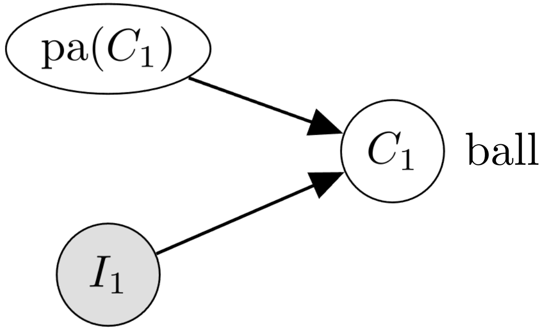
- Simpler graphs
- Fewer requirements to find it
- Scalability



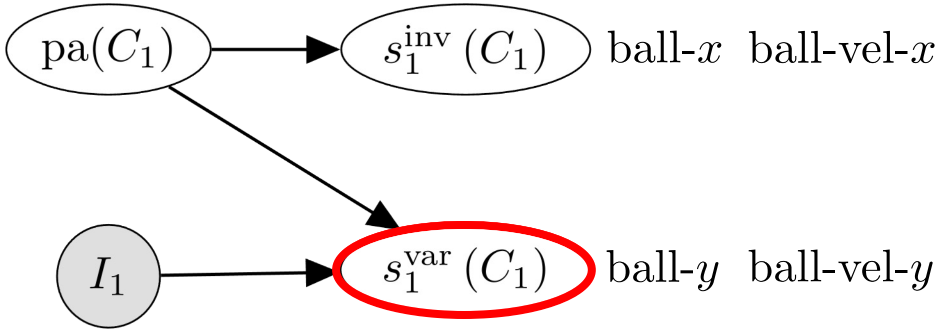
What is a Causal Variable?
Minimal Causal Variables

- 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





(a) Original causal graph of C_1



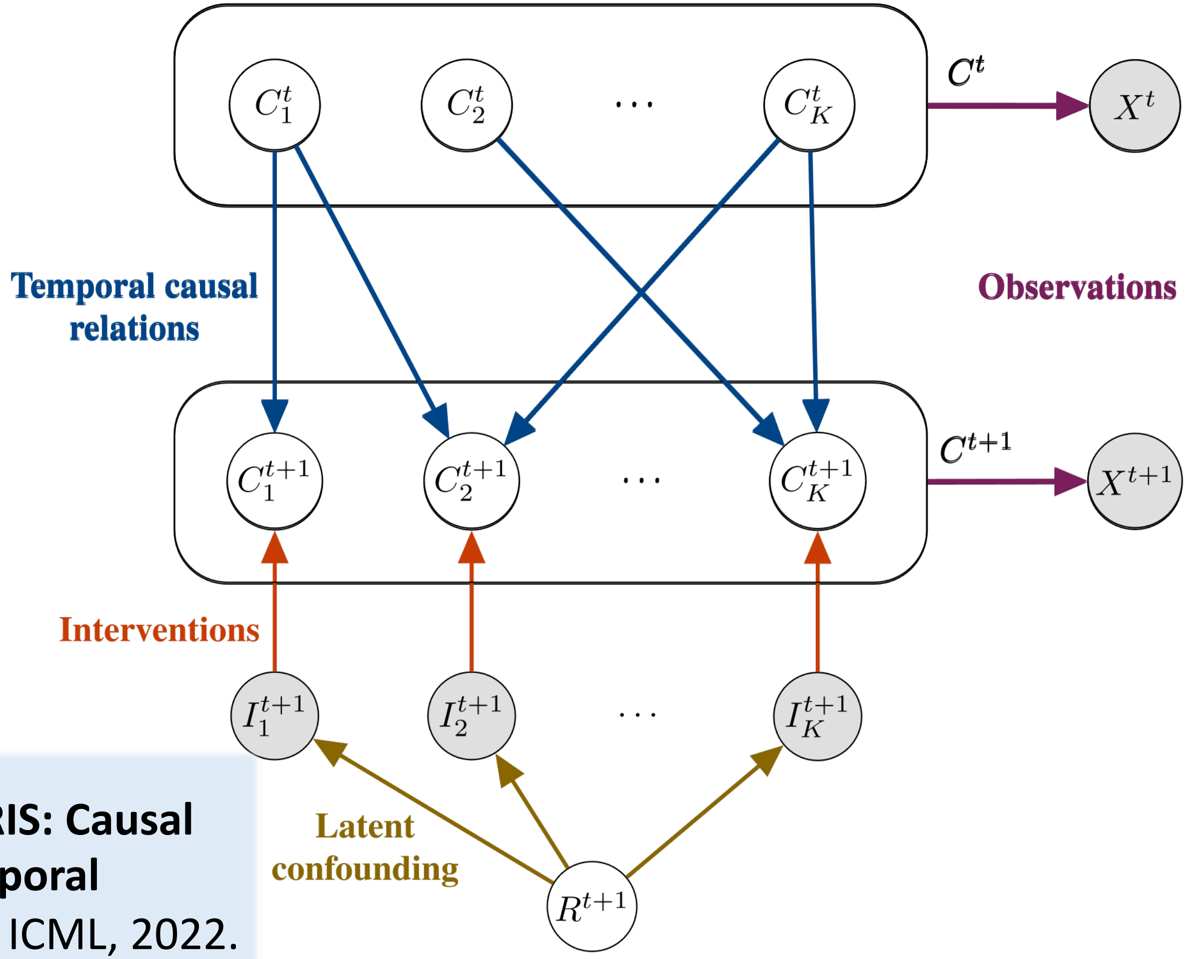
(b) Minimal causal split graph of C_1

Summary

- Abstraction simplifies graphs and identifiability
- Multidimensional causal variables needed for modeling different levels of abstractions
- Minimal causal variables: define causal variables by the interventions we have

- How can we identify minimal causal variables?

Causal Identifiability from Temporal Intervened Sequences
Setup

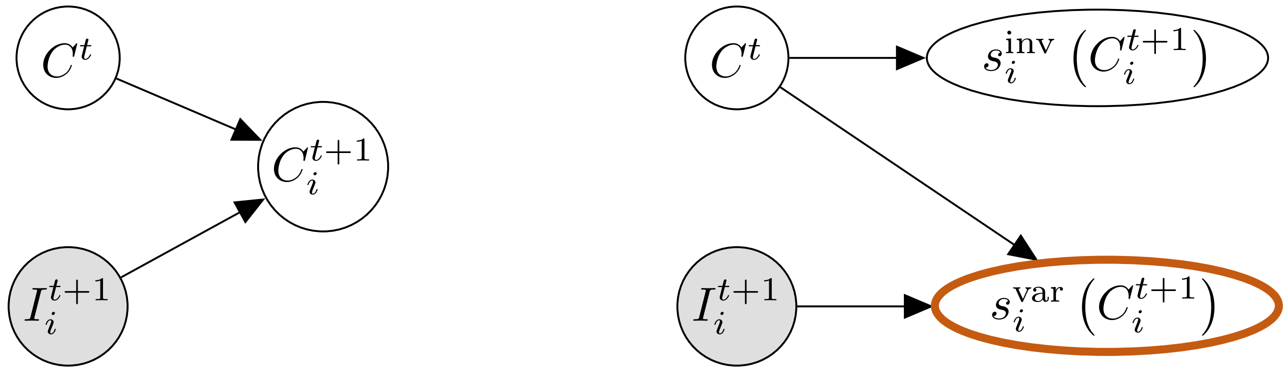


Lippe, Phillip et al. "**CITRIS: Causal Identifiability from Temporal Intervened Sequences**." ICML, 2022.

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:

$$C_i^{t+1} \not\perp\!\!\!\perp I_i^{t+1} | C^t, I_j^{t+1}$$

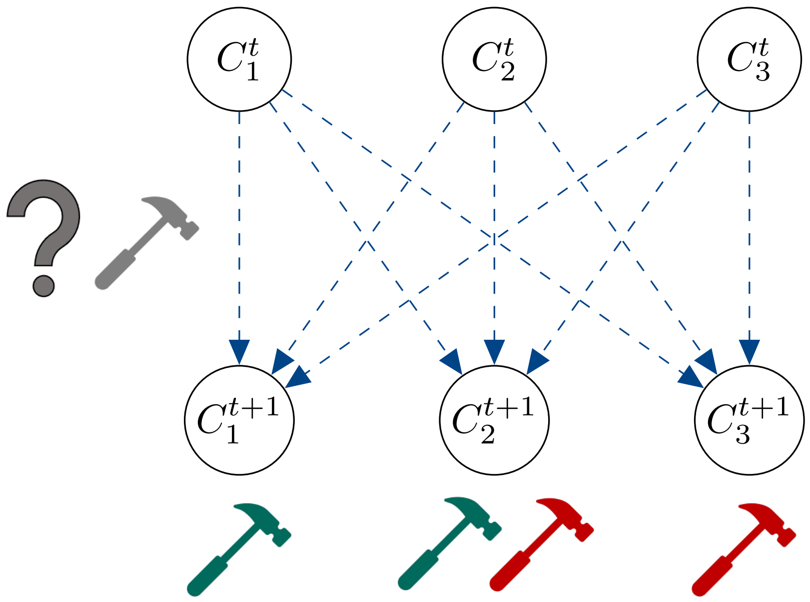


(a) Original causal graph of C_i (b) Minimal causal split graph of C_i

Causal Identifiability from Temporal Intervened Sequences

Intervention Experiments

- How many (soft) interventions are needed? $C_i^{t+1} \not\perp\!\!\!\perp I_i^{t+1} | C^t, I_j^{t+1}$
- Every variable needs to be unique in the sets of experiments it is in



Experiments

 $\{\,\}$

 $\{C_1, C_2\}$

 $\{C_2, C_3\}$

	I_1	I_2	I_3
	0	0	0
	1	1	0
	0	1	1

Causal Identifiability from Temporal Intervened Sequences

Intervention Experiments

- How many (soft) interventions are needed? $C_i^{t+1} \not\perp\!\!\!\perp I_i^{t+1} | C^t, I_j^{t+1}$
- Every variable needs to be unique in the sets of experiments it is in
- Turns out:

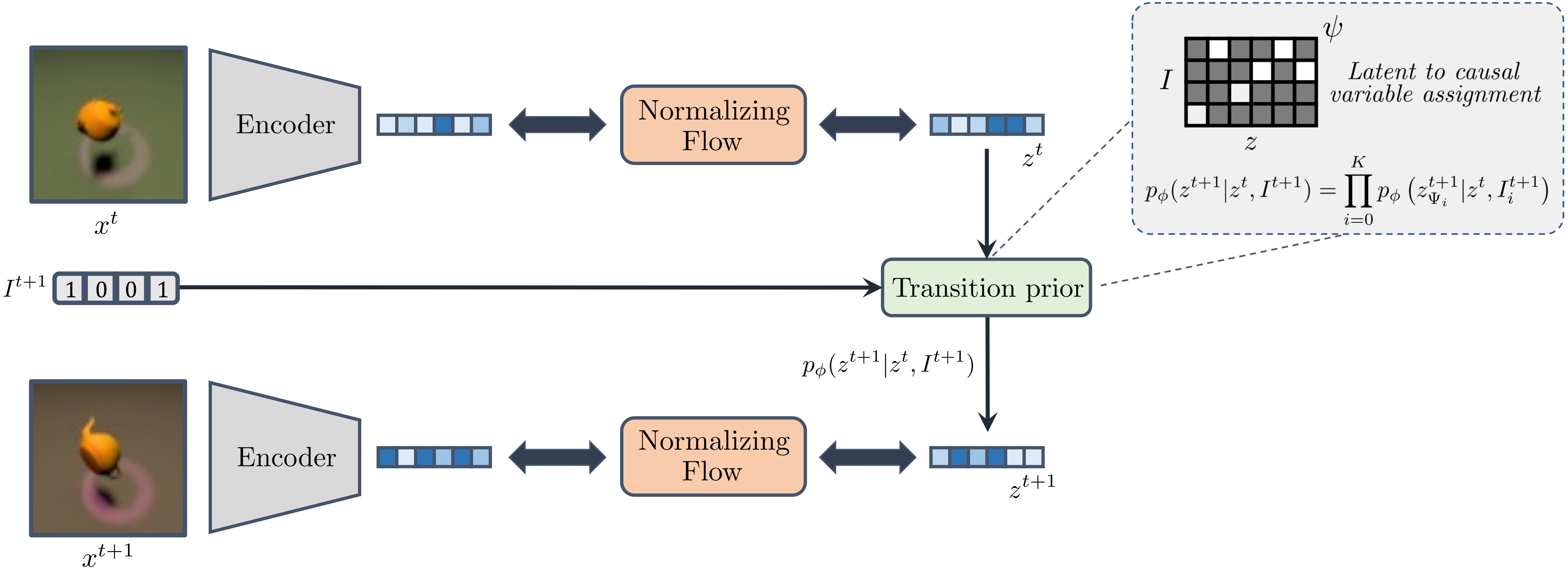
$\lfloor \log_2 K \rfloor + 2$ experiments identify the minimal causal variables

- Just one more than Intervention Design bound for causal discovery: $\lfloor \log_2 K \rfloor + 1$

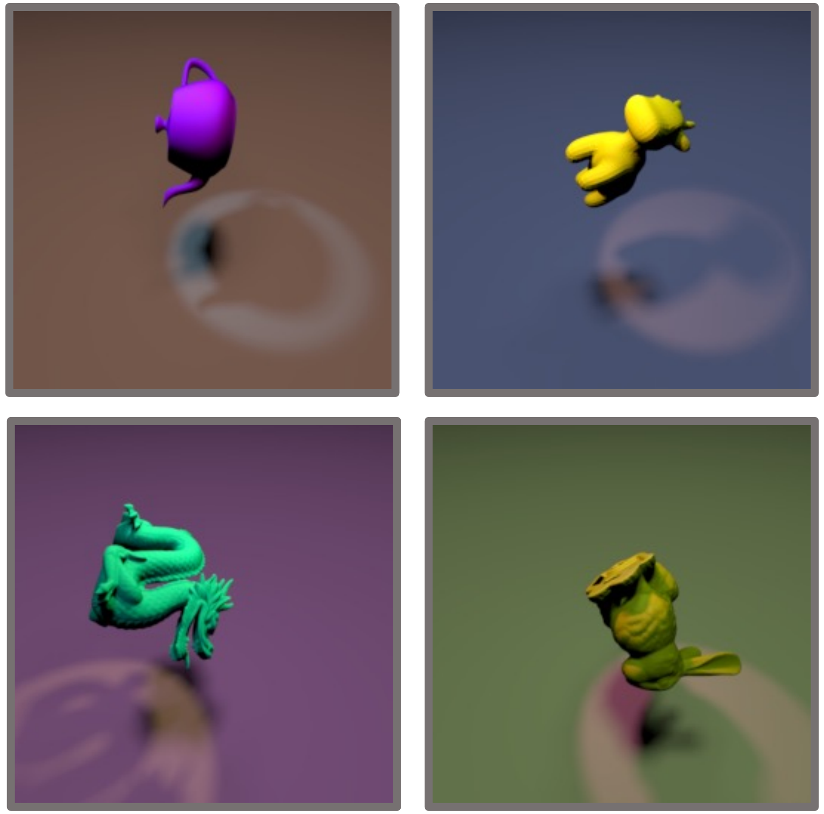
Lippe, Phillip et al. "**Intervention Design for Causal Representation Learning.**" CRL@UAI 2022.

Come to our poster later!

CITRIS Architecture
CITRIS-NF



CITRIS Experiments
Temporal Causal3DIdent



Causal Factors

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

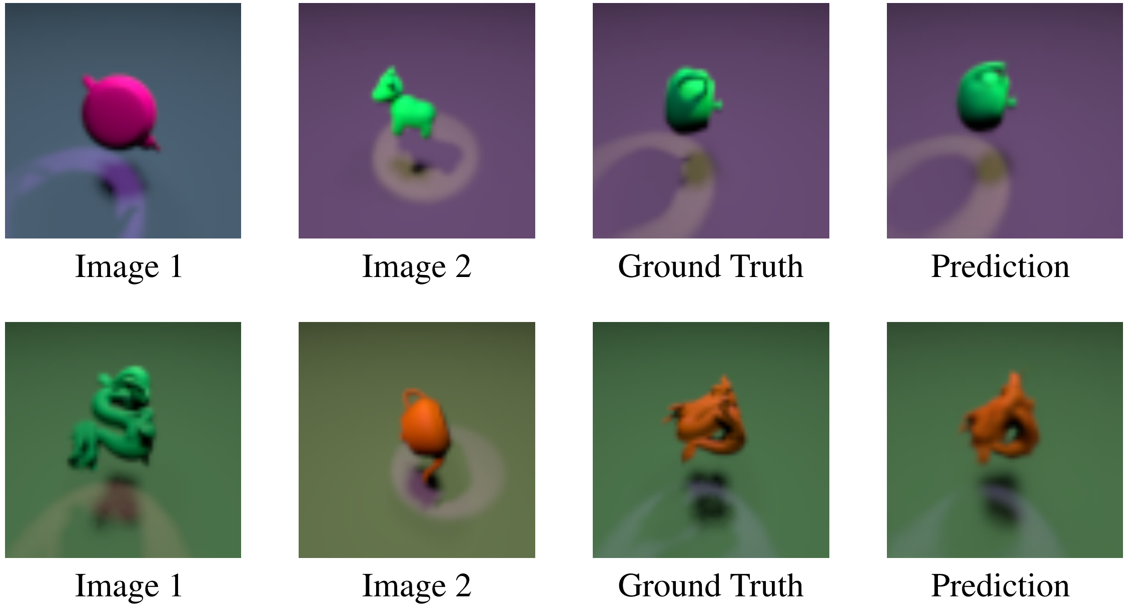
categorical
continuous
angle / circular

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

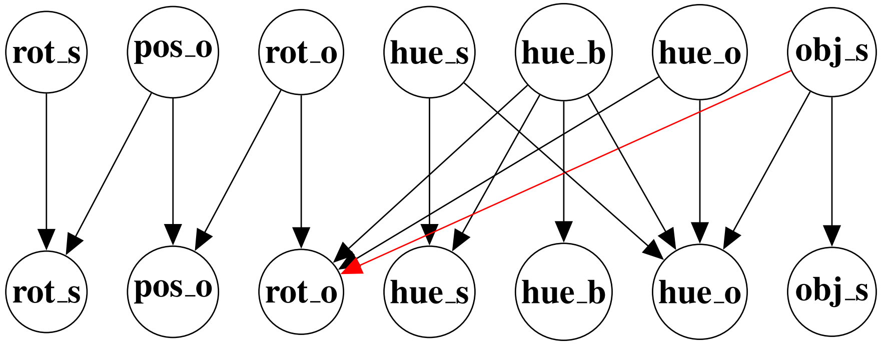
Von Kügelgen, Julius, et al. "Self-supervised learning with data augmentations provably isolates content from style." *NeurIPS,* 2021.

CITRIS Experiments
Temporal Causal3DIdent

Novel combinations of causal factors



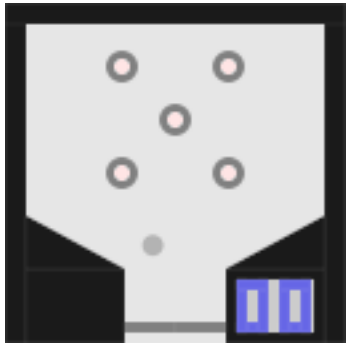
Learned Causal Graph

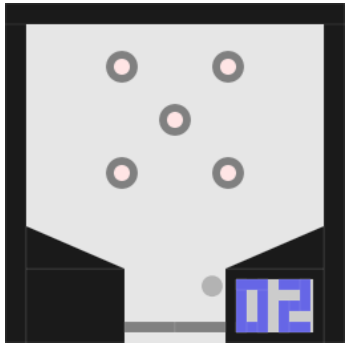


Instantaneous Effects in Temporal Sequences

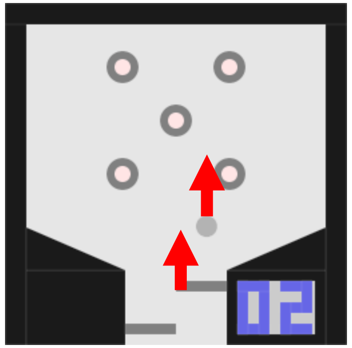
- Common assumption: time resolves causal effects
- But what about observations at low frame rates?

⇒ Instantaneous Effects!





time step t



time step $t+1$

Instantaneous Effects in Temporal Sequences
Challenges

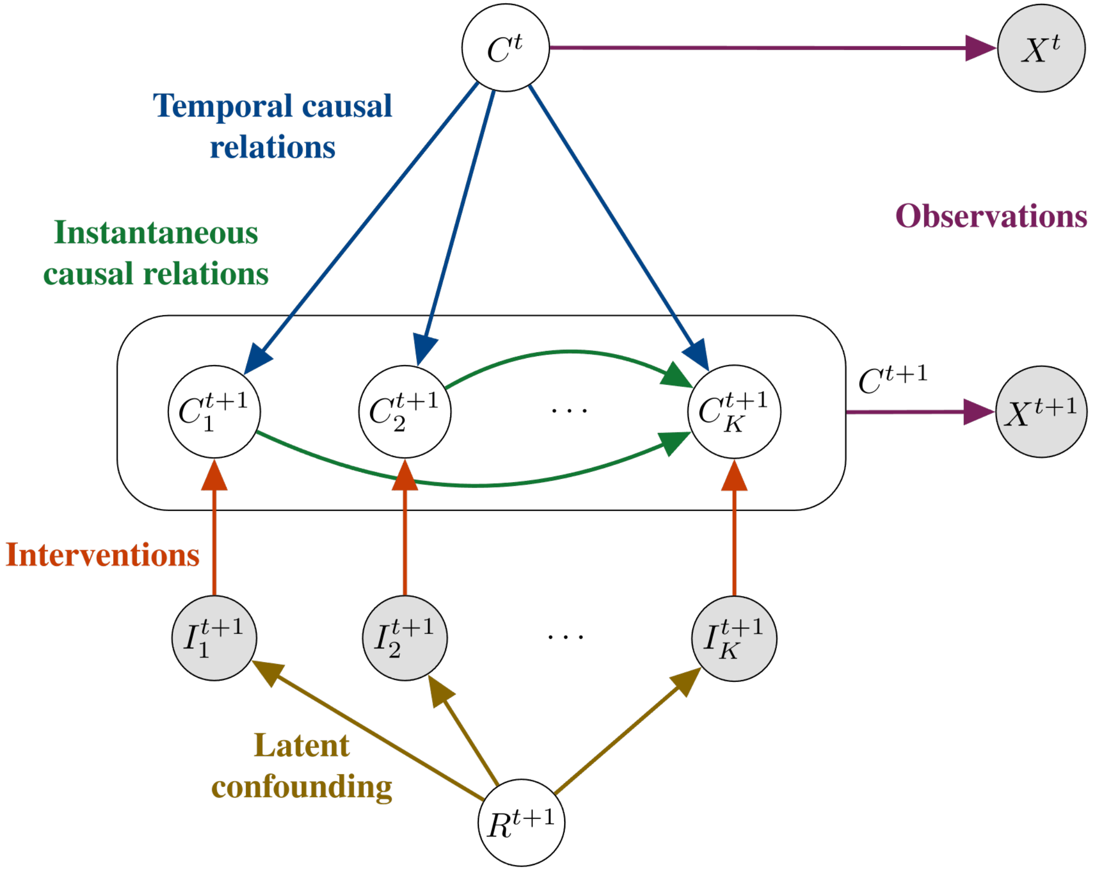
- Many more pitfalls, e.g.:

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

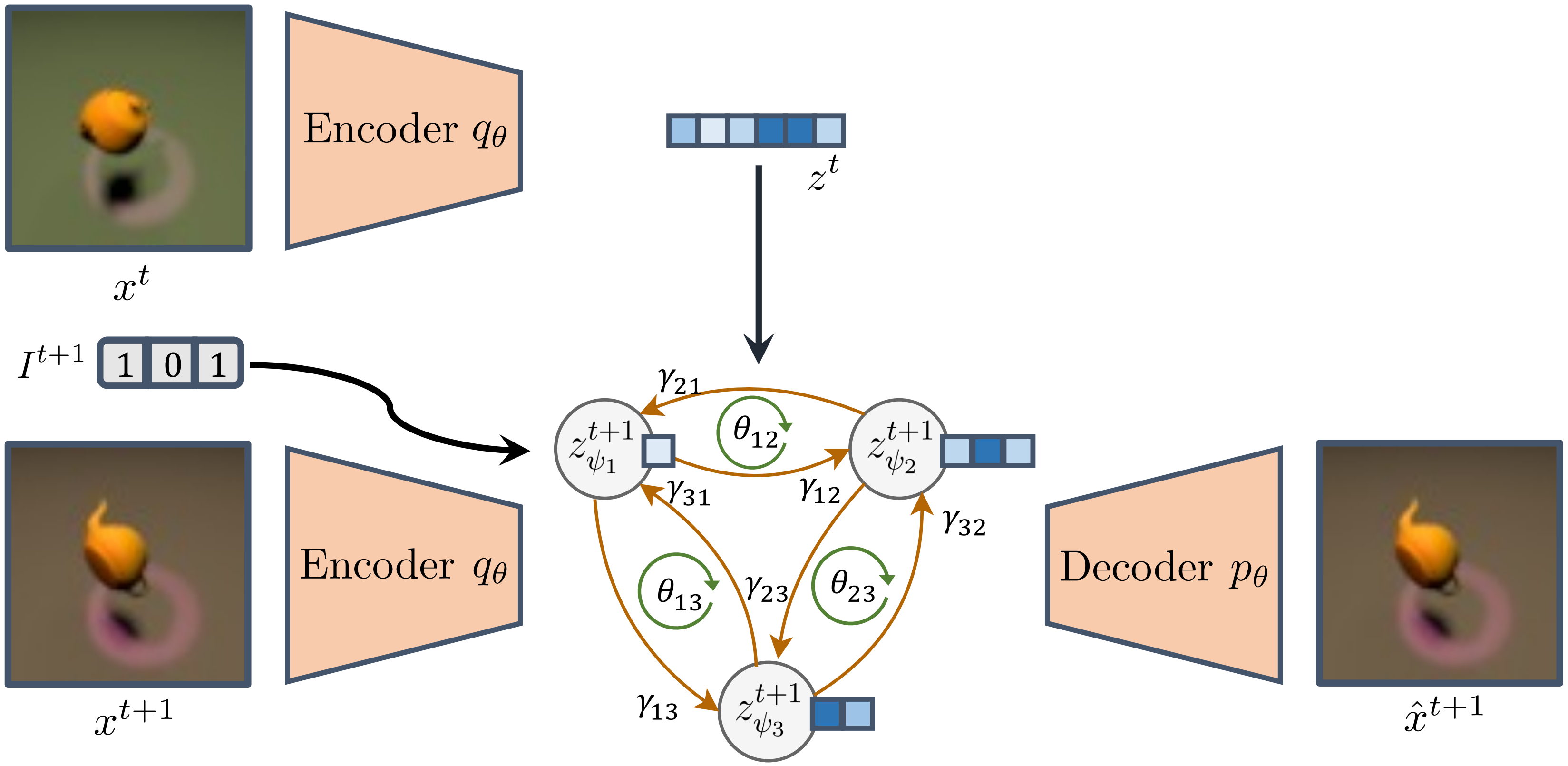
- Solution: *perfect* interventions!

  $\Rightarrow$ Minimal causal variables become identifiable

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

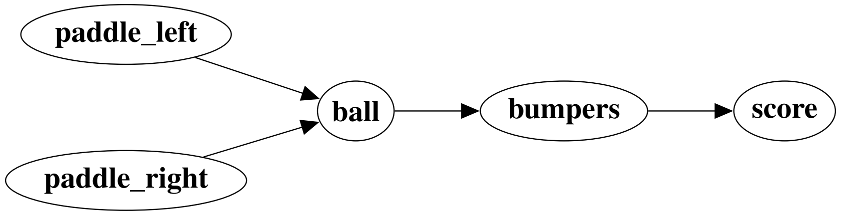


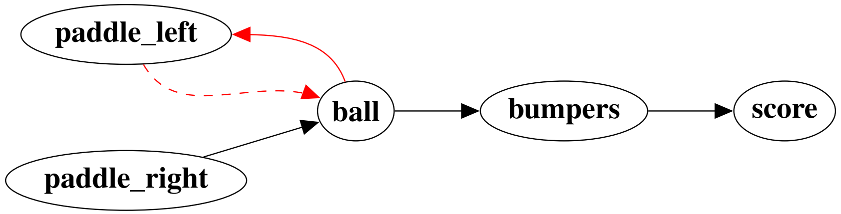
iCITRIS: CRL for Instantaneous Temporal Effects
Architecture

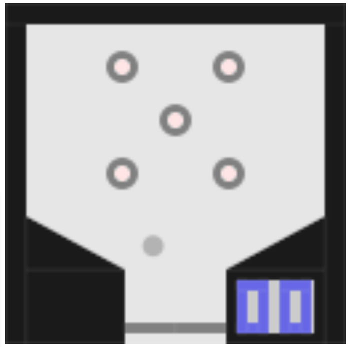


iCITRIS: CRL for Instantaneous Temporal Effects
Experiments

Learned Causal Graphs







Lippe, Phillip et al. "**iCITRIS: Causal Representation Learning for Instantaneous Temporal Effects**." CRL@UAI 2022.

Come to our poster later!

Summary

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

- **Intervention Design**: $\lfloor \log_2 K \rfloor + 2$ experiments identify the minimal causal variables, just one more than in causal discovery

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

Challenges in CRL

Open world

Low-level actions

Observability

Guarantees

Evaluation

Sample efficiency

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

Collaborators


Sara Magliacane


Sindy Löwe


Yuki Asano


Taco Cohen


Efstratios Gavves


Johann Brehmer


Pim de Haan

Thank You!

[1] 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.

[2] 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.

Poster today!

[3] 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.

Poster today!

[4] Brehmer, Johann, Pim de Haan, Phillip Lippe, Taco Cohen. "**Weakly supervised causal representation learning**." First Workshop on Causal Representation Learning (CRL), UAI 2022.

Poster today!