# CITRIS: Causal Identifiability from Temporal Intervened Sequences

Phillip Lippe[1] (p.lippe@uva.nl), Sara Magliacane[2,3], Sindy Löwe[4], Yuki M. Asano[1], Taco Cohen[5], Efstratios Gavves[1]

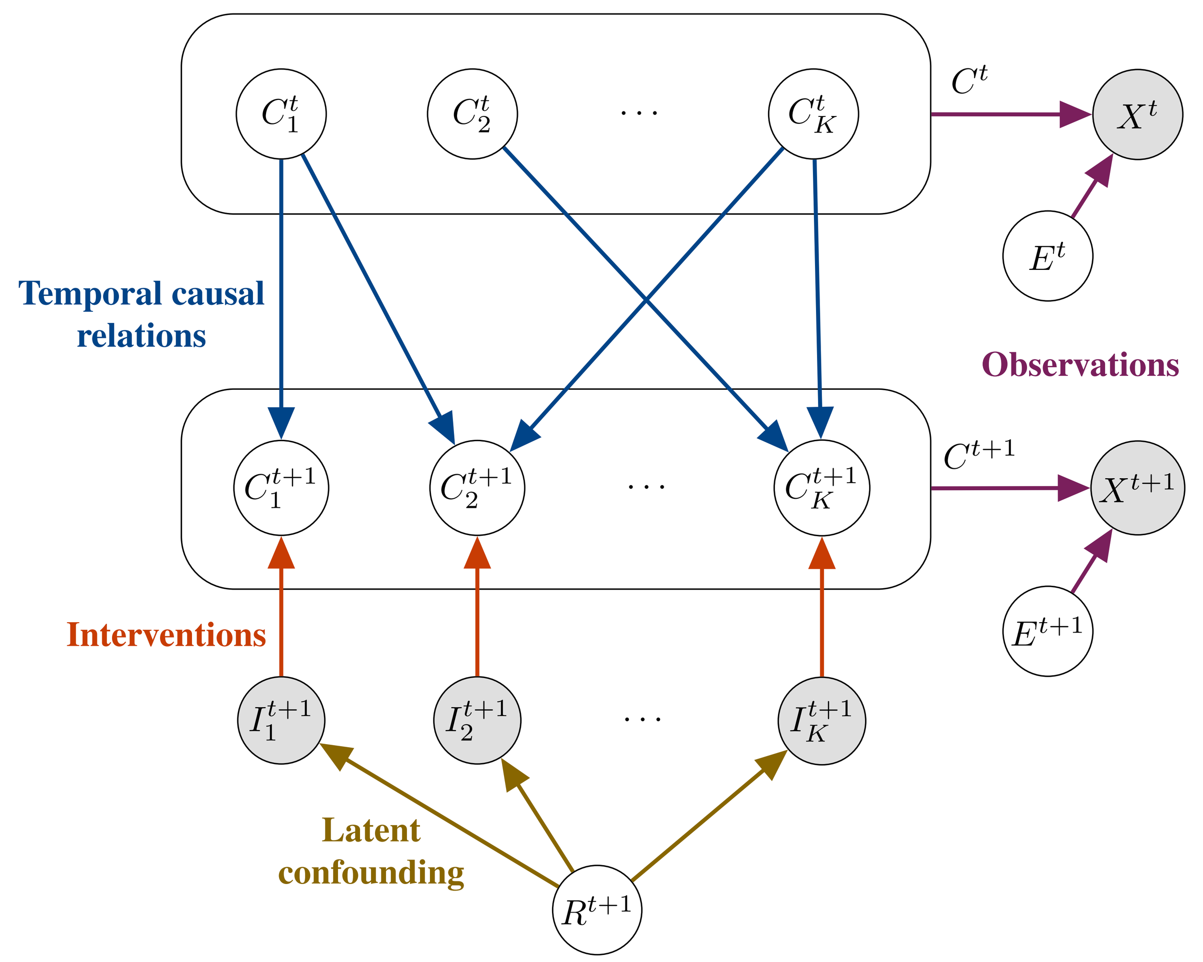
[1]University of Amsterdam, QUVA lab; [2]University of Amsterdam, INDE lab; [3]MIT-IBM Watson AI Lab; [4]University of Amsterdam, Bosch-Delta lab; [4]Qualcomm AI Research, The Netherlands

We identify **multidimensional causal variables** from **videos** with **interventions.**

**Paper and code**

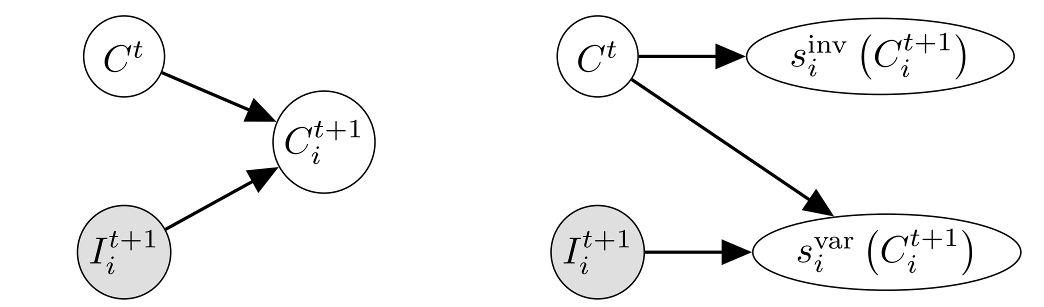
## Setup

Latent **Dynamical Bayesian Network** with **temporal causal relations** and **observed intervention targets.**



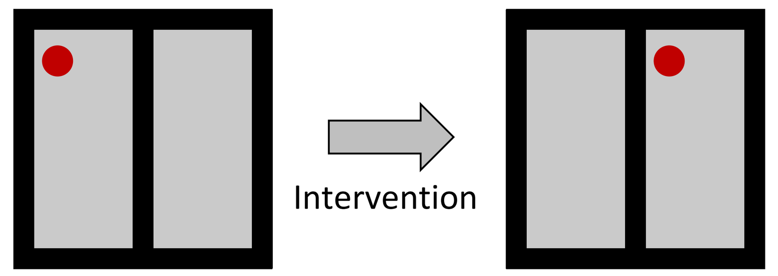

## What are Minimal Causal Variables?

Interventions may not affect all dimensions of a causal variable
⇒ **Minimal causal variables**: the part of a causal variable whose mechanism is strictly influenced by the provided interventions



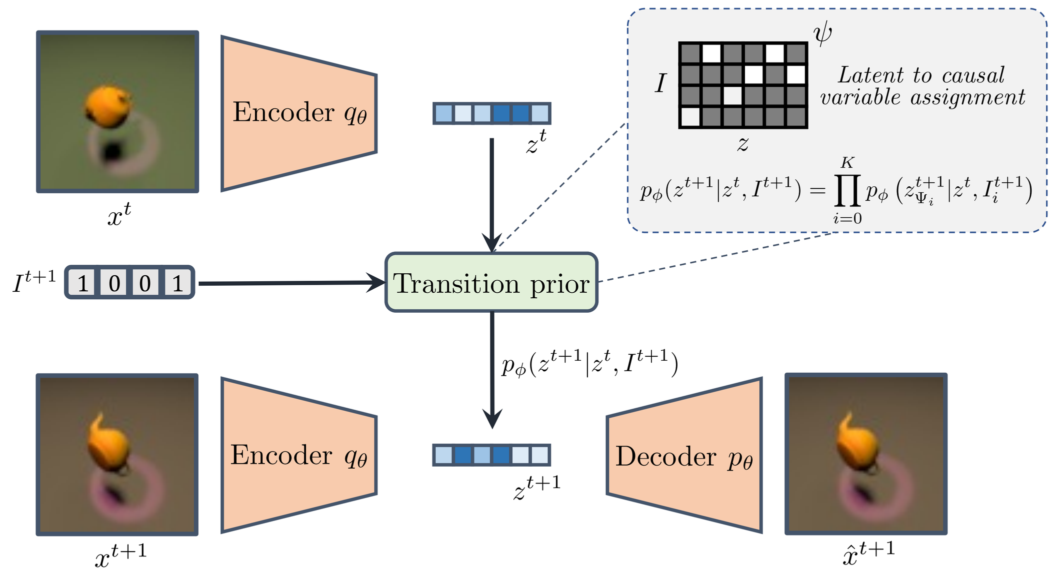

(a) Original causal graph of $C_i$ (b) Minimal causal split graph of $C_i$

Multidimensional factors are only identifiable up to the minimal causal variable. Example: an intervention swaps boxes of the ball. Is the x-position or only the box-ID the multidimensional causal variable?

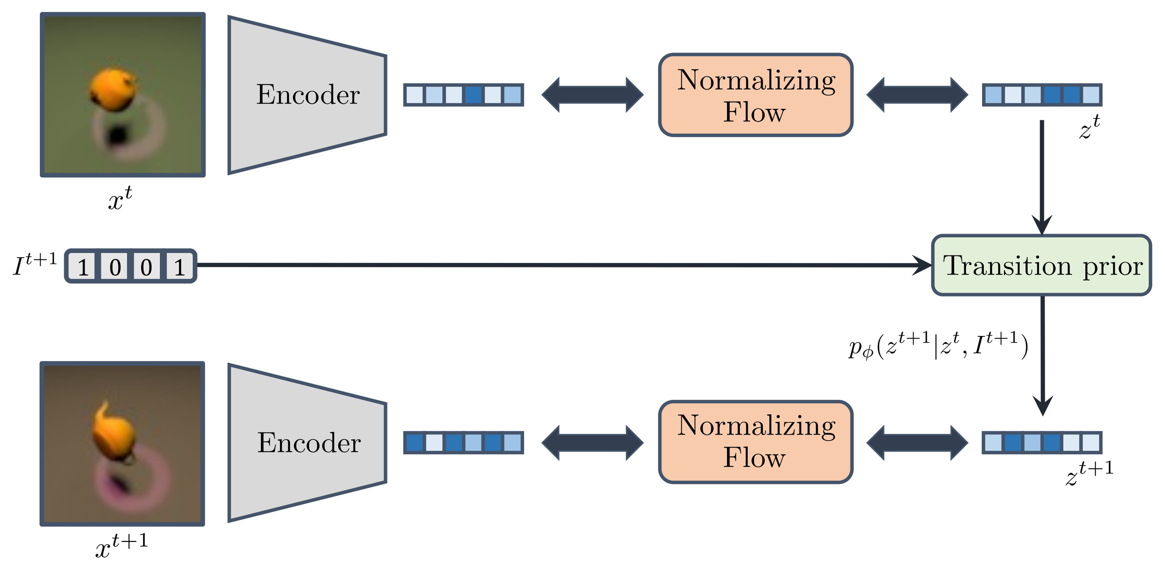



## Architectures

**CITRIS-VAE:**
Learn latent-to-causal assignment by conditioning only on one target

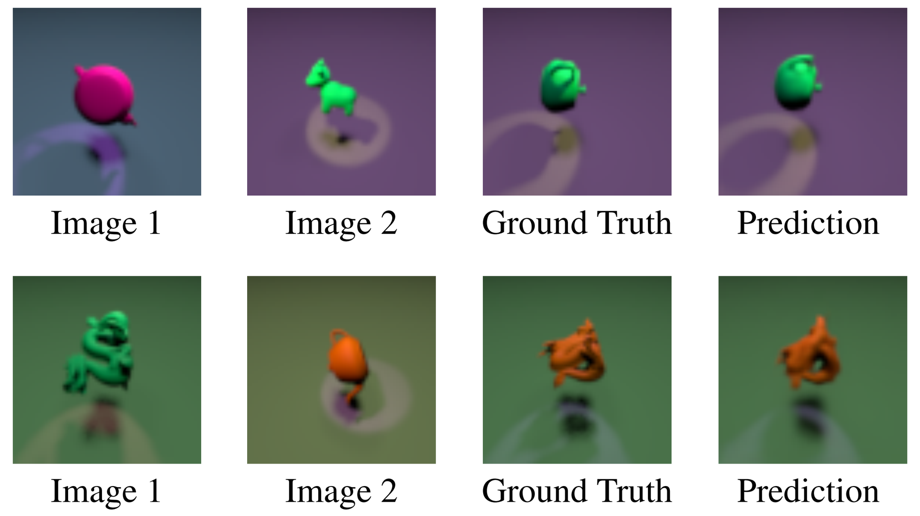



**CITRIS-NF:**
Combining Pretrained Autoencoders with Normalizing Flows

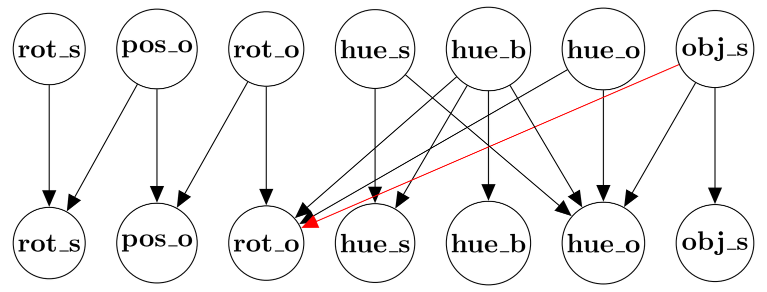



## Experiments

**Temporal Causal3DIdent**
3D rendered images with 7 causal factors

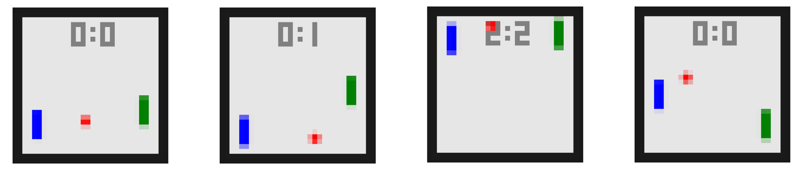



| | Triplet dist ↓ | Correlation metrics | | | |
|---|---|---|---|---|---|
| | Mean | $R^2$ diag ↑ | $R^2$ sep ↓ | Spearman diag ↑ | Spearman sep ↓ |
| **SlowVAE** | 0.40 | 0.61 | 0.23 | 0.59 | 0.27 |
| **iVAE*** | 0.27 | 0.80 | 0.29 | 0.77 | 0.28 |
| **CITRIS-VAE** | 0.21 | 0.89 | 0.10 | 0.88 | 0.12 |
| **CITRIS-NF** | **0.07** | **0.98** | **0.04** | **0.97** | **0.08** |
| - 5 shapes | 0.09 | 0.98 | 0.05 | 0.97 | 0.10 |
| - 2 unseen shapes | 0.23 | 0.94 | 0.15 | 0.93 | 0.19 |




**CITRIS disentangles all seven causal variables well, while finding the causal graph and generalizing beyond seen instances!**

**Interventional Pong**
Pong game dynamics with 5 intervened causal factors




| | Triplets ↓ | $R^2$ diag ↑ | $R^2$ sep ↓ | Spearman diag ↑ | Spearman sep ↓ |
|---|---|---|---|---|---|
| **SlowVAE** | 0.34 | 0.61 | 0.17 | 0.66 | 0.23 |
| **iVAE*** | 0.09 | 0.91 | 0.04 | 0.92 | 0.06 |
| **CITRIS-VAE** | 0.03 | 0.99 | **0.01** | 0.99 | **0.05** |
| **CITRIS-NF** | **0.02** | **1.00** | 0.04 | **1.00** | 0.10 |

**CITRIS identifies the five causal variables while disentangling them from the unintervened causal variables!**