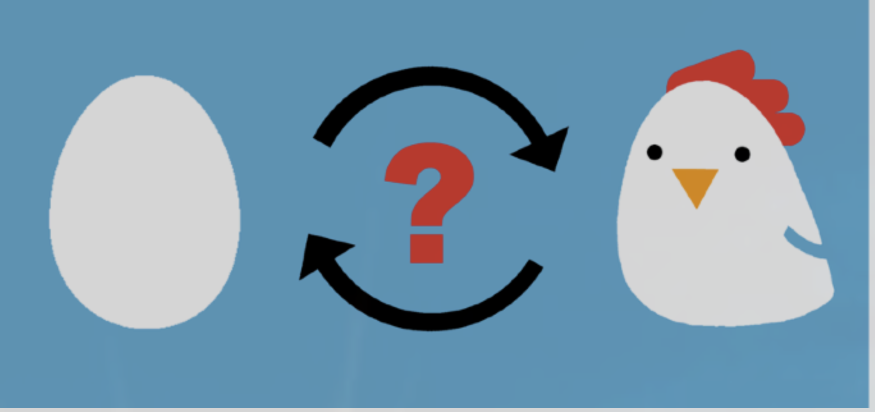


Causality and Machine Learning
(80-816/516)

Classes 20 (March 27, 2025)

Causal Representation Learning 3: Benefits from Temporal Constrains

Instructor:

Kun Zhang (kunz1@cmu.edu)
Zoom link: https://cmu.zoom.us/j/8214572323)
Office Hours: W 3:00–4:00PM (on Zoom or in person); other times by appointment

We Mainly Focused on the IID Case: Recent Advances in Causal Representation Learning

i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	*No*	*No*	*(Different types of) equivalence class*
		Yes	
	Yes	**No**	**Unique identifiability (under structural conditions)**
		Yes	
Non-I, but I.D.	No/Yes	No	(Extended) regression
		Yes	Latent temporal causal processes identifiable!
I., but non-I.D.	No	No	More informative than MEC (CD-NOD)
	Yes		May have unique identifiability
	No	Yes	Changing subspace identifiable
	Yes		Variables in changing relations identifiable

- PC, FCI, etc.
- LiNGAM
- Rank-based, GIN…
- TDRL
- TDRL with instantaneous relations…
- CD-NOD
- CRL from multiple distributions
- Causal GenAI

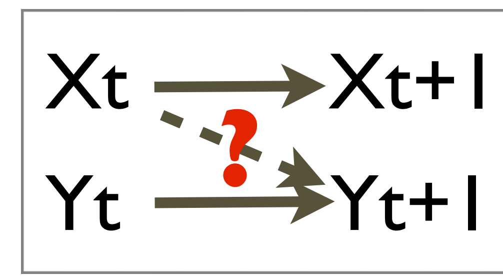
CRL with Temporal Constraints

Non-I, but I.D.	No/Yes	No	(Extended) regression
		Yes	Latent temporal causal processes identifiable!

- Discovering causal relations among the measured time series
    - Granger causality (but be aware of temporal resolution)
- Temporally disentangled representation learning
    - With invertible or non-invertible mixing functions
- With instantaneous relations

Granger Causality: Original Definition & Practical Constraints

- Two principles (Granger, '80)
  - Future cannot cause past
  - No redundant info: Cause contains unique information about effect
- X causes Y if $P(Y_{t+1} \in A \mid \Omega_t) \neq P(Y_{t+1} \in A \mid \Omega_t^{-X})$

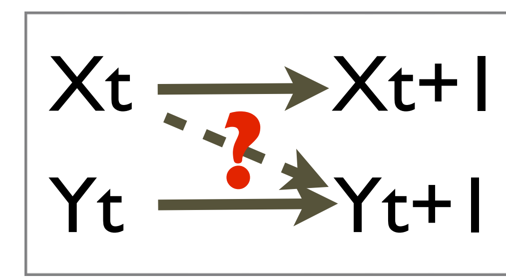


  - Completely nonparametric; $Y_{t+1} \not\perp X_t$ given all the remaining information until time t
- In practice: causality in mean; linear Granger causality

- C.W.J. Granger, Testing for causality: A personal viewpoint. Journal of Economic Dynamics & Control 2: 329–352, 1980

Conditional Independence-Based Method for Causal Discovery from Time Series

- Two principles (Granger, '80)
  - **Future cannot cause past**
  - No redundant info: Cause contains unique information about effect



- X causes Y if $P(Y_{t+1} \in A \mid \Omega_t) \neq P(Y_{t+1} \in A \mid \Omega_t^{-X})$
  - Completely nonparametric; $Y_{t+1} \not\perp\!\!\!\perp X_t$ given all the remaining information until time t
- In practice: causality in mean; linear Granger causality

- The PC algorithm still applies; additional temporal constraints!

Two Schemes of Temporal Aggregation

- **Subsampling (syst... sampling)**

X...

- T...

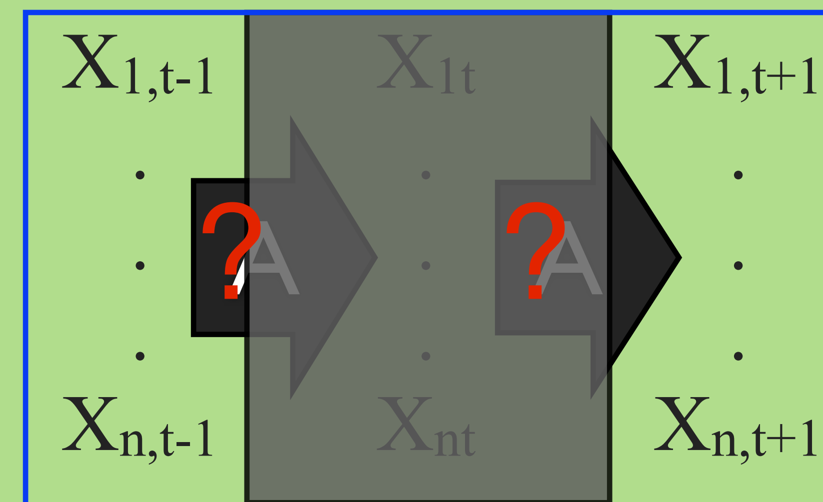
$\mathbf{x}_1, \ldots$

$\tilde{\mathbf{x}}_1 = \frac{1}{k} \ldots$

Can we recover the causal influence matrix A?

- Examples: temperature data, stock daily returns, GDP, fMRI...

Assume $\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{E}_t$



ausal info *tends to disappear* *as k⟶∞*

...usal info *tends to be ...antaneous* *as k⟶∞* :

$$\tilde{\mathbf{X}}_t \approx \mathbf{A}\tilde{\mathbf{X}}_t + \tilde{\mathbf{E}}_t$$

CRL with Temporal Constraints

Non-I, but I.D.	No/Yes	No	(Extended) regression
		Yes	Latent temporal causal processes identifiable!

- Discovering causal relations among the measured time series
  - Granger causality (but be aware of temporal resolution)
- Temporally disentangled representation learning
  - With invertible or non-invertible mixing functions
- With instantaneous relations

Learning Latent Causal Dynamics

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

"Time-delayed" influence (no instantaneous dependence) renders latent processes & their relations identifiable

Time-series Inputs $\{x_t\}_{t=0}^{T}$

Unsupervised Representation Learning

$$\mathbf{x}_t = \mathbf{g}(\mathbf{z}_t)$$

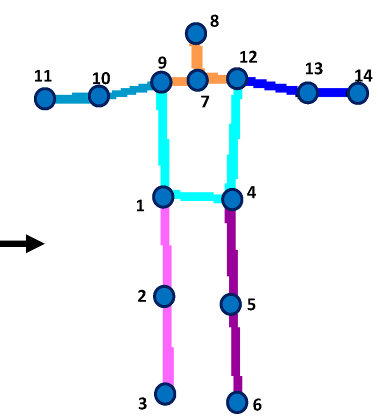
Latent processes

Temporal VAE with causal prior

Latent temporal causal processes z_{it} follow

- completely **nonparametric** model; or furthermore,
- **non-stationary** noise or causal influence, or
- **Parametric** constraints

Causal Skeleton Recovery



Recovered latent processes

- *Yao, Chen, Zhang, "Causal Disentanglement for Time Series," NeurIPS 2022*
- *Yao, Sun, Ho, Sun, Zhang, "Learning Temporally causal latent processes from general temporal data," ICLR 2022*
- *Chen et al., "CaRiNG: Learning Temporal Causal Representation under Non-Invertible Generation Process," ICML 2024*

Learning Latent Causal Dynamics

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes		
No		

“Time-delayed” influence (no
ers latent
tifiable

$$
\begin{cases}
z^{\text{fix}}_{s,t} &= f_s\left(\{z_{i,t-\tau} | z_{i,t-\tau} \in \mathbf{Pa}(z^{\text{fix}}_{s,t})\}, \epsilon_{s,t}\right), \\
z^{\text{chg}}_{c,t} &= f_c\left(\{z_{i,t-\tau} | z_{i,t-\tau} \in \mathbf{Pa}(z^{\text{chg}}_{c,t})\}, \boldsymbol{\theta}^{\text{dyn}}_r, \epsilon_{c,t}\right), \\
z^{\text{obs}}_{o,t} &= f_o\left(\boldsymbol{\theta}^{\text{obs}}_r, \epsilon_{o,t}\right), \\
\mathbf{x}_t &= \mathbf{g}(\mathbf{z}_t).
\end{cases}
$$

Time-series Inputs $\{x_t\}_{t=0}^{T}$

Unsupervised Representation Learning

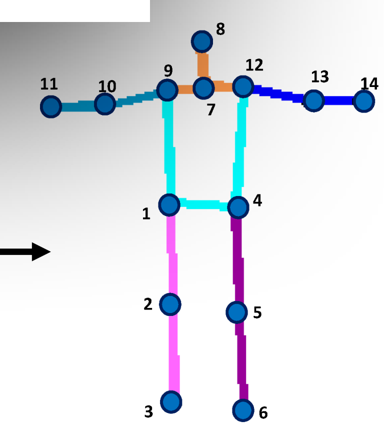
$\mathbf{x}_t = \mathbf{g}(\mathbf{z}_t)$

Latent processes

Latent temporal causal processes z_{it} follow

- completely **nonparametric** model; or furthermore,
- **non-stationary** noise or causal influence, or
- **Parametric** constraints

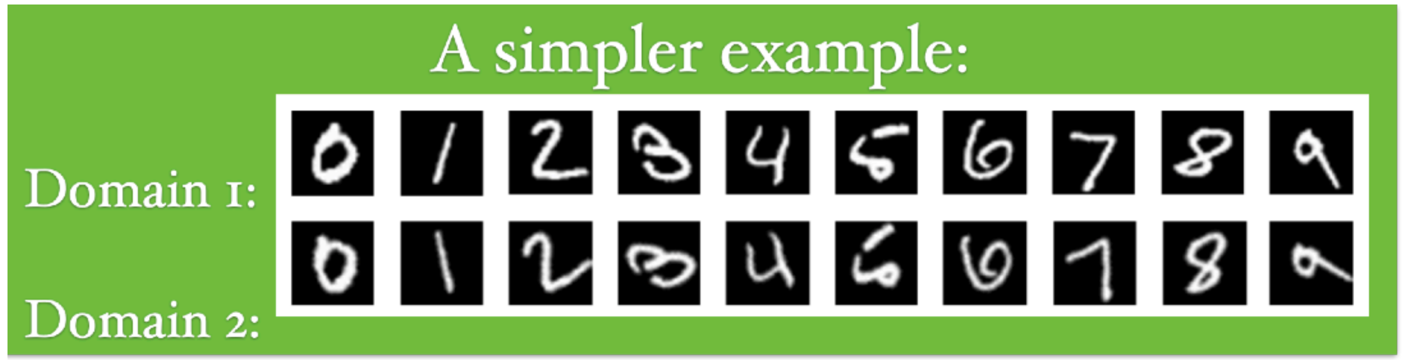
Causal Skeleton Recovery



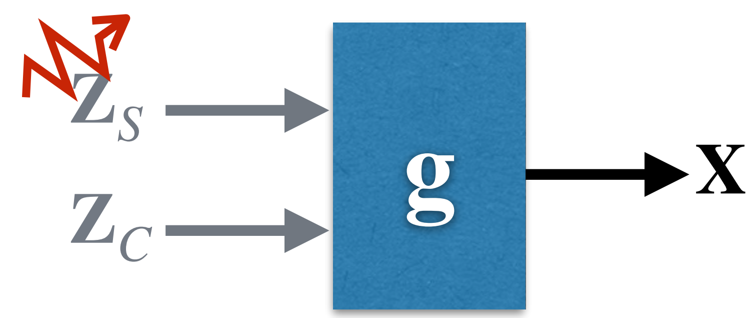
Recovered latent processes

- *Yao, Chen, Zhang, “Causal Disentanglement for Time Series,” NeurIPS 2022*
- *Yao, Sun, Ho, Sun, Zhang, “Learning Temporally causal latent processes from general temporal data,” ICLR 2022*
- *Chen et al., “CaRiNG: Learning Temporal Causal Representation under Non-Invertible Generation Process,” ICML 2024*

Remember the *Multi-Domain Case?*



i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes



We exploit the conditional independence among Z_{Si} given the surrogate (domain info)!

- Underlyi
- Changing components $\mathbf{Z}_S$ are identifiable, invariant part $\mathbf{Z}_C$ is identifiable up to its subspace
- Using $\mathbf{Z}_C$ and transformed changing part $\tilde{\mathbf{Z}}_S$ for transfer learning

Temporally Disentangled Representation Learning

Weiran Yao
CMU
weiran@cmu.edu

Guangyi Chen
CMU & MBZUAI
guangyichen1994@gmail.com

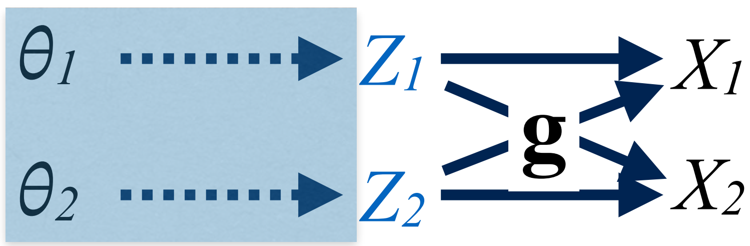
Kun Zhang
CMU & MBZUAI
kunz1@cmu.edu

Abstract

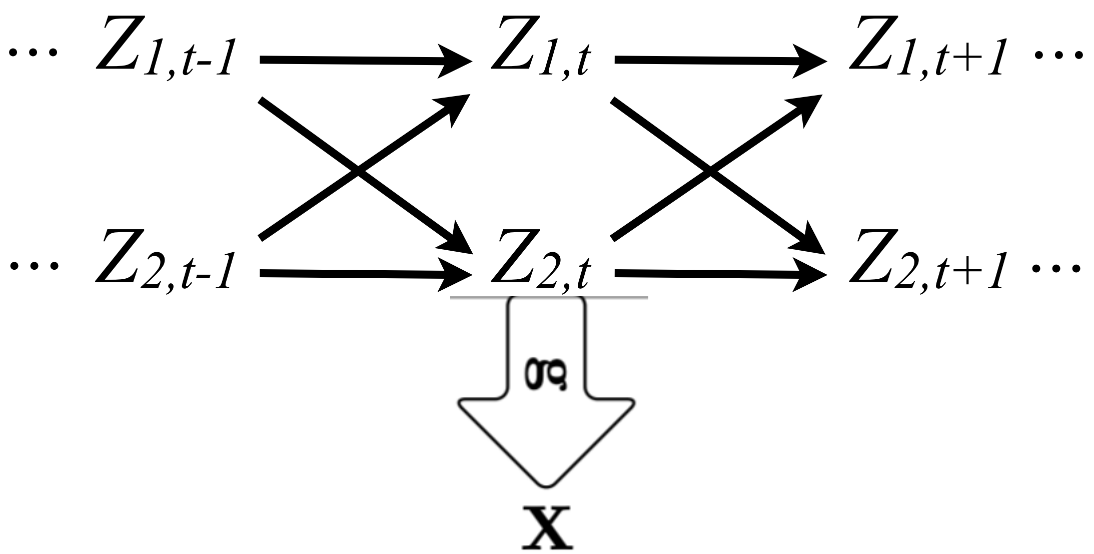
Recently in the field of unsupervised representation learning, strong identifiability results for disentanglement of causally-related latent variables have been established by exploiting certain side information, such as class labels, in addition to independence. However, most existing work is constrained by functional form assumptions such as independent sources or further with linear transitions, and distribution assumptions such as stationary, exponential family distribution. It is unknown whether the underlying latent variables and their causal relations are identifiable if they have arbitrary, nonparametric causal influences in between. In this work, we establish the identifiability theories of nonparametric latent causal processes from their nonlinear mixtures under fixed temporal causal influences and analyze how distribution changes can further benefit the disentanglement. We propose **`TDRL`**, a principled framework to recover time-delayed latent causal variables and identify their relations from measured sequential data under stationary environments and under different distribution shifts. Specifically, the framework can factorize unknown distribution shifts into transition distribution changes under fixed and time-varying latent causal relations, and under observation changes in observation. Through experiments, we show that time-delayed latent causal influences are reliably identified and that our approach considerably outperforms existing baselines that do not correctly exploit this modular representation of changes. Our code is available at: https://github.com/weirayao/tdrl.

Why? Let's Derive it…

- Multi-domain case:



- Temporal case:



Comparison of the Identifiability Result

- Multi-domain case:

Theorem 1 (Identifiablity under a Fixed Temporal Causal Model). *Suppose there exists invertible function $\hat{\mathbf{g}}$ that maps $\mathbf{x}_t$ to $\hat{\mathbf{z}}_t$, i.e.,*

$$\hat{\mathbf{z}}_t = \hat{\mathbf{g}}(\mathbf{x}_t) \tag{3}$$

such that the components of $\hat{\mathbf{z}}_t$ are mutually independent conditional on $\hat{\mathbf{z}}_{t-1}$. Let

$$\mathbf{v}_{k,t} \triangleq \left(\frac{\partial^2 \eta_{kt}}{\partial z_{k,t}\partial z_{1,t-1}}, \frac{\partial^2 \eta_{kt}}{\partial z_{k,t}\partial z_{2,t-1}}, ..., \frac{\partial^2 \eta_{kt}}{\partial z_{k,t}\partial z_{n,t-1}}\right)^{\mathsf{T}}, \ \mathring{\mathbf{v}}_{k,t} \triangleq \left(\frac{\partial^3 \eta_{kt}}{\partial z_{k,t}^2\partial z_{1,t-1}}, \frac{\partial^3 \eta_{kt}}{\partial z_{k,t}^2\partial z_{2,t-1}}, ..., \frac{\partial^3 \eta_{kt}}{\partial z_{k,t}^2\partial z_{n,t-1}}\right)^{\mathsf{T}}. \tag{4}$$

If for each value of $\mathbf{z}_t$, $\mathbf{v}_{1,t}, \mathring{\mathbf{v}}_{1,t}, \mathbf{v}_{2,t}, \mathring{\mathbf{v}}_{2,t}, ..., \mathbf{v}_{n,t}, \mathring{\mathbf{v}}_{n,t}$, as $2n$ vector functions in $z_{1,t-1}$, $z_{2,t-1}$, ..., $z_{n,t-1}$, are linearly independent, then $\mathbf{z}_t$ must be an invertible, component-wise transformation of a permuted version of $\hat{\mathbf{z}}_t$.

- Temporal case:

such that $\hat{z}_{s,j} = h_{s,i}(z_{s,i})$. For ease of exposition, we assume that the $\mathbf{z}_c$ and $\mathbf{z}_s$ correspond to components in $\mathbf{z}$ with indices $\{1, \ldots, n_c\}$ and $\{n_c + 1, \ldots, n\}$ respectively, that is, $\mathbf{z}_c = (z_i)_{i=1}^{n_c}$ and $\mathbf{z}_s = (z_i)_{i=n_c+1}^{n}$.

Theorem 4.1. *We follow the data generation process in Equation 1 and make the following assumptions:*

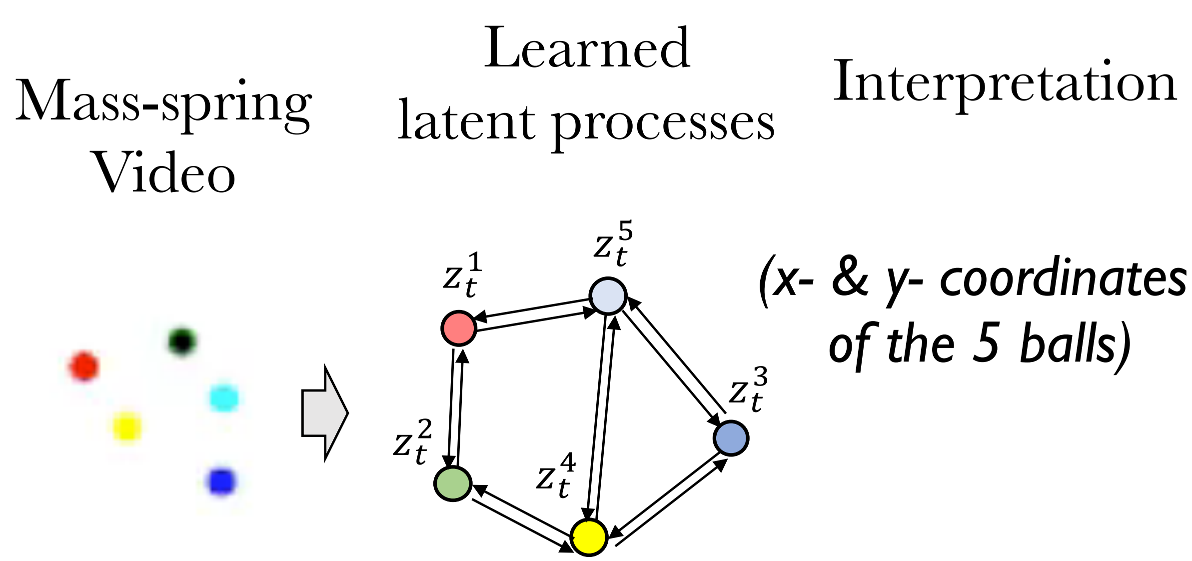
- *A1 (Smooth and Positive Density): The probability density function of latent variables is smooth and positive, i.e. $p_{\mathbf{z}|\mathbf{u}}$ is smooth and $p_{\mathbf{z}|\mathbf{u}} > 0$ over $\mathcal{Z}$ and $\mathcal{U}$.*
- *A2 (Conditional independence): Conditioned on $\mathbf{u}$, each z_i is independent of any other z_j for $i, j \in [n]$, $i \neq j$, i.e. $\log p_{\mathbf{z}|\mathbf{u}}(\mathbf{z}|\mathbf{u}) = \sum_i^n q_i(z_i, \mathbf{u})$ where q_i is the log density of the conditional distribution, i.e., $q_i := \log p_{z_i|\mathbf{u}}$.*
- *A3 (Linear independence): For any $\mathbf{z}_s \in \mathcal{Z}_s \subseteq \mathbb{R}^{n_s}$, there exist $2n_s + 1$ values of $\mathbf{u}$, i.e., $\mathbf{u}_j$ with $j = 0, 1, \ldots, 2n_s$, such that the $2n_s$ vectors $\mathbf{w}(\mathbf{z}_s, \mathbf{u}_j) - \mathbf{w}(\mathbf{z}_s, \mathbf{u}_0)$ with $j = 1, ..., 2n_s$, are linearly independent, where vector $\mathbf{w}(\mathbf{z}_s, \mathbf{u})$ is defined as follows:*

$$\mathbf{w}(\mathbf{z}_s, \mathbf{u}) = \Big(\frac{\partial q_{n_c+1}(z_{n_c+1}, \mathbf{u})}{\partial z_{n_c+1}}, \ldots, \frac{\partial q_n(z_n, \mathbf{u})}{\partial z_n}, \frac{\partial^2 q_{n_c+1}(z_{n_c+1}, \mathbf{u})}{\partial z_{n_c+1}^2}, \ldots, \frac{\partial^2 q_n(z_n, \mathbf{u})}{\partial z_n^2}\Big). \tag{3}$$

By learning $(\hat{g}, p_{\hat{\mathbf{z}}_c}, p_{\hat{\mathbf{z}}_s|\mathbf{u}})$ to achieve Equation 2, $\mathbf{z}_s$ is component-wise identifiable.

Results on Simple Video Data

- For easy interpretation, consider a simple video data set
  - Mass-spring system: a video dataset with ball movement and invisible springs



A Causal Perspective on Reinforcement Learning

- Potential issues in deep RL algorithms
  - Lack interpretability
  - Not generalize well
  - Data hungry
- Mitigate such issues through causal representations and graph structures

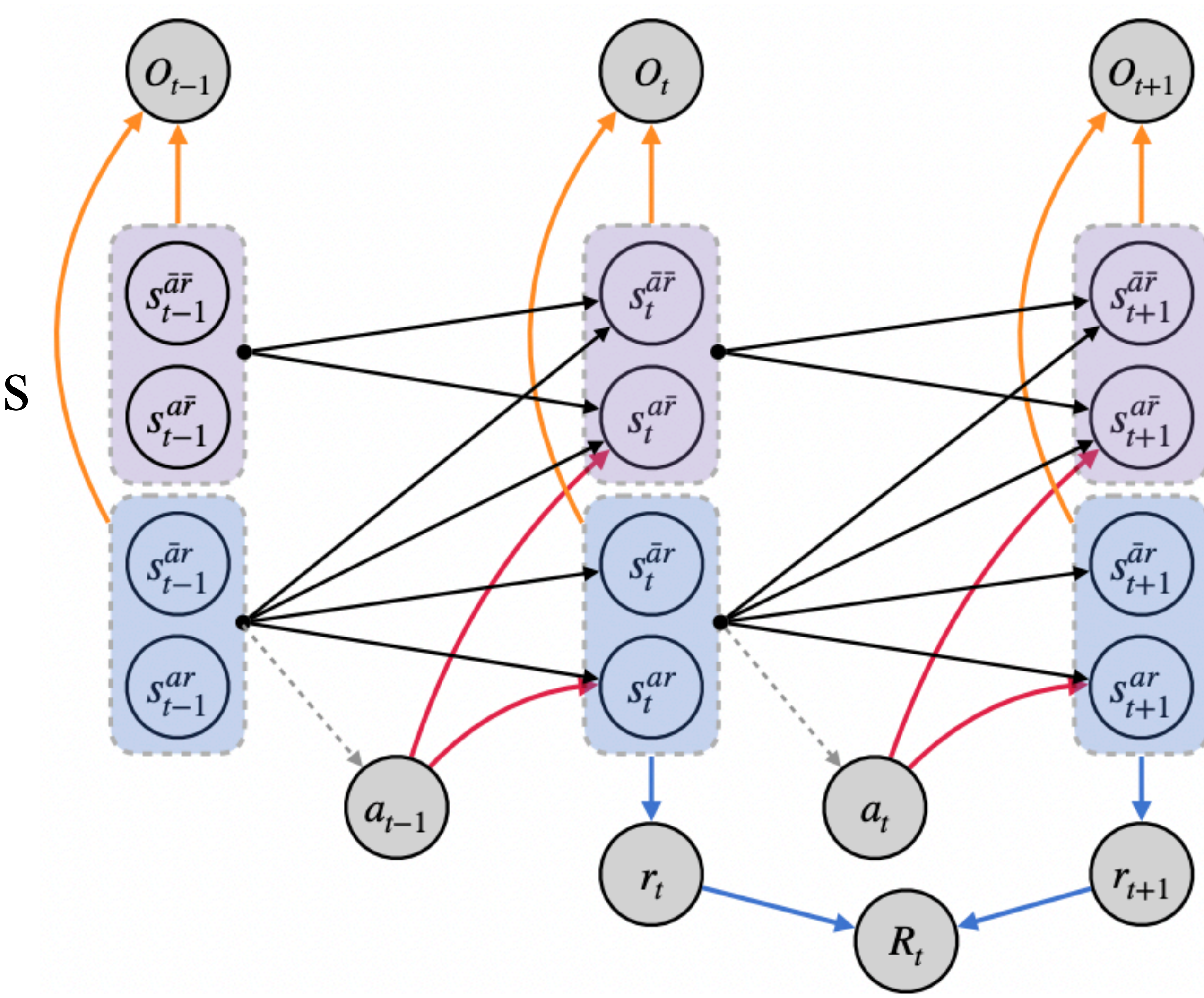
A Causal Perspective on Reinforcement Learning

- Potential issues in deep RL algorithms
  - Lack interpretability
  - Not generalize well
  - Data hungry
- Mitigate such issues through causal representations and graph structures

Four Categories of State Representations in RL

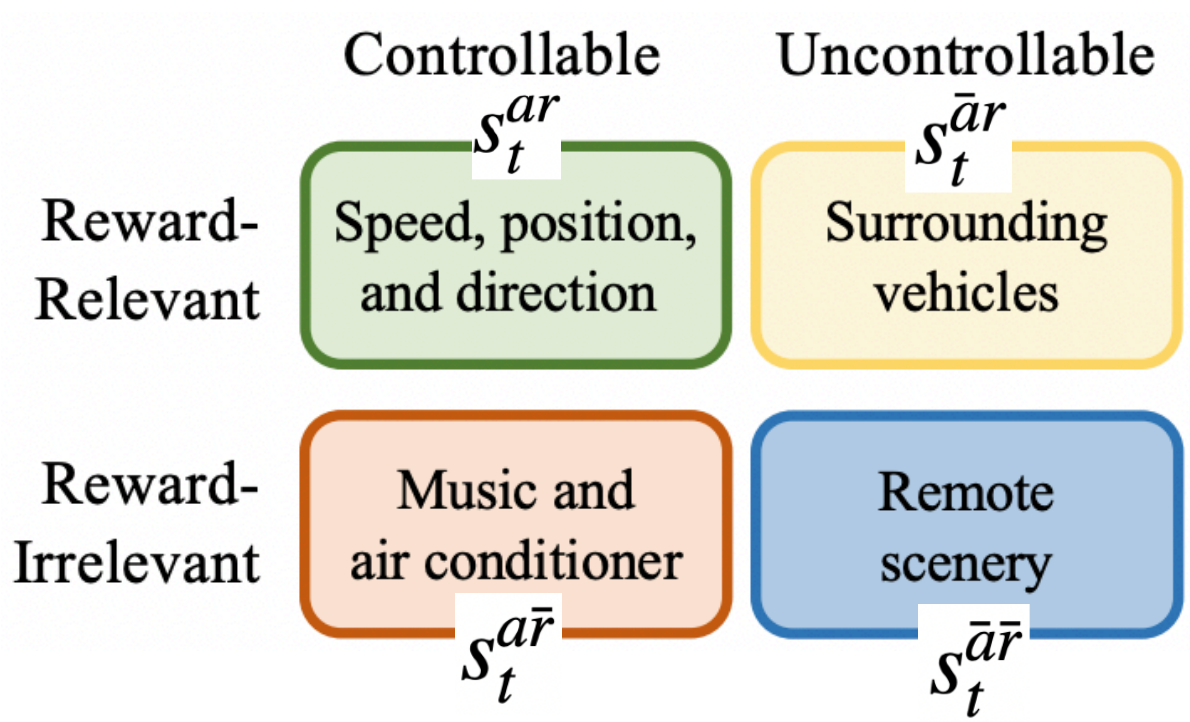
	Controllable	Uncontrollable
Reward-Relevant	Speed, position, and direction	Surrounding vehicles
Reward-Irrelevant	Music and air conditioner	Remote scenery

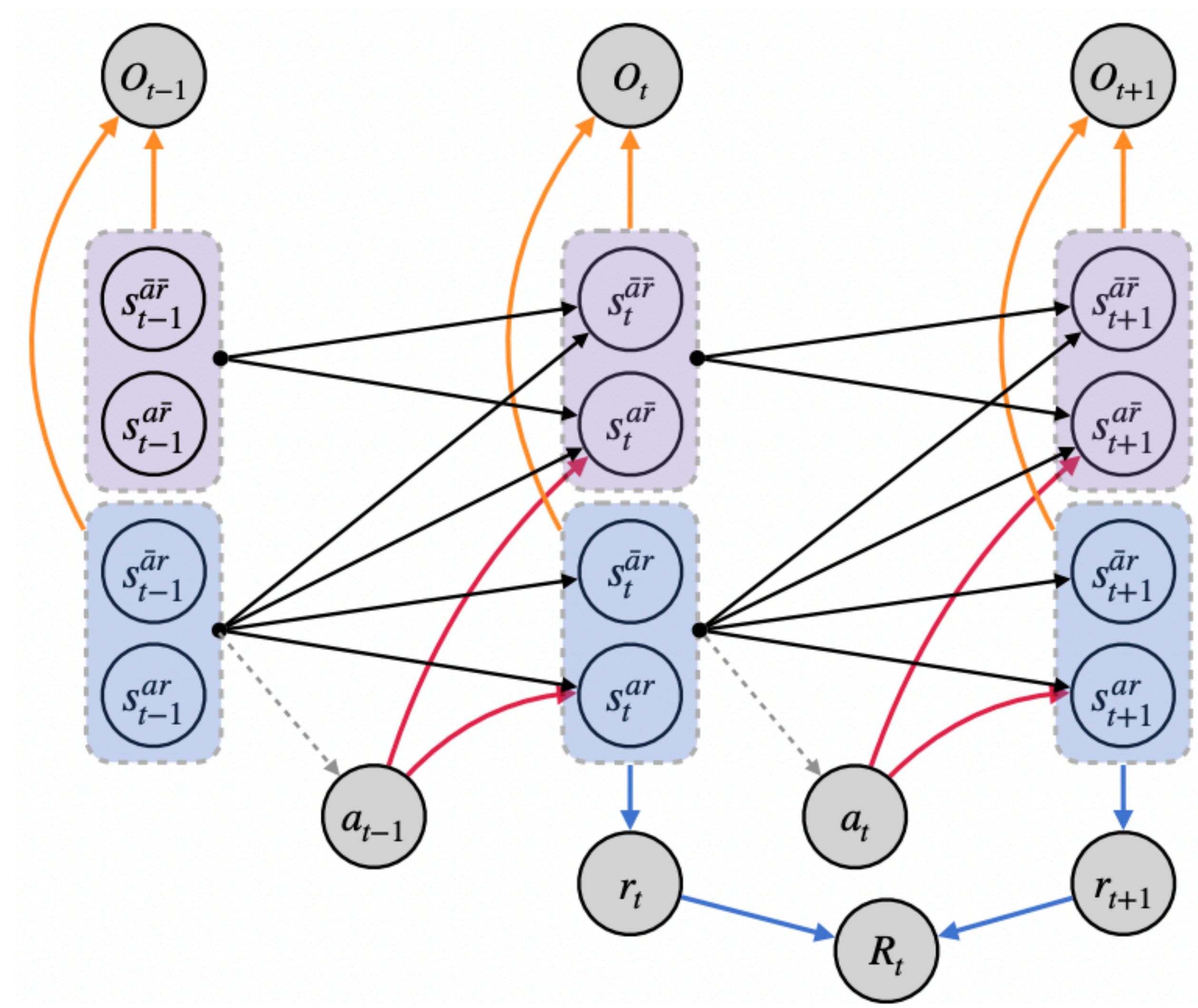
- s_t^{ar} : controllable and reward-relevant state variables
- $s_t^{\bar{a}r}$: reward-relevant state variables that are beyond our control
- $s_t^{a\bar{r}}$: controllable but reward-irrelevant factors
- $s_t^{\bar{a}\bar{r}}$: uncontrollable and reward-irrelevant latent variables



Liu*, Huang*, Zhu, Tian, Gong, Yu, Zhang. Learning world models with identifiable factorization. Arxiv, 2023.

Four Categories of State Representations in RL

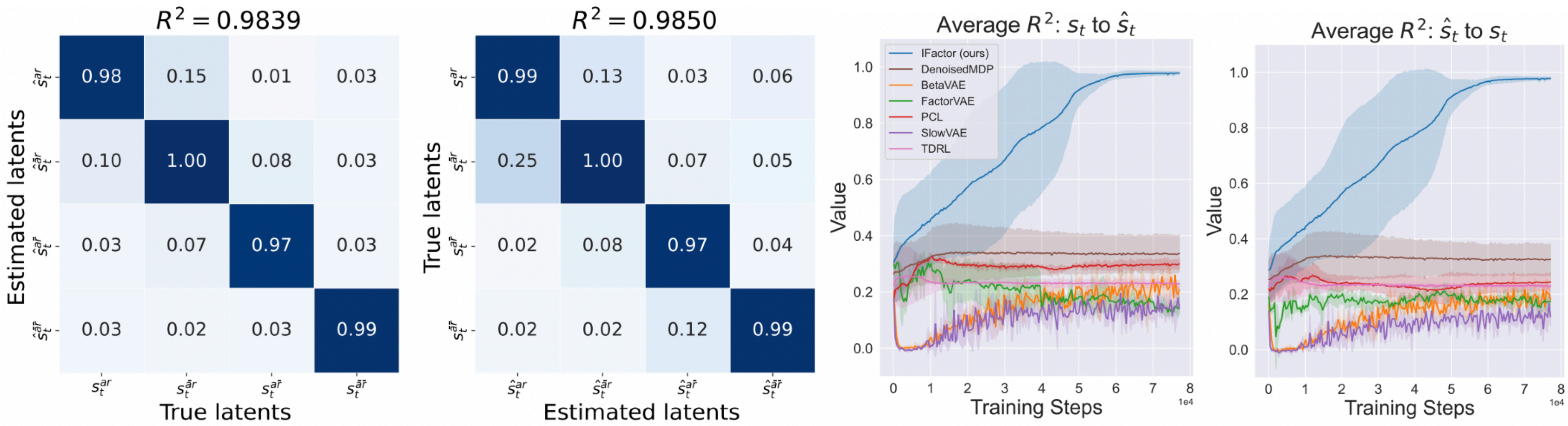




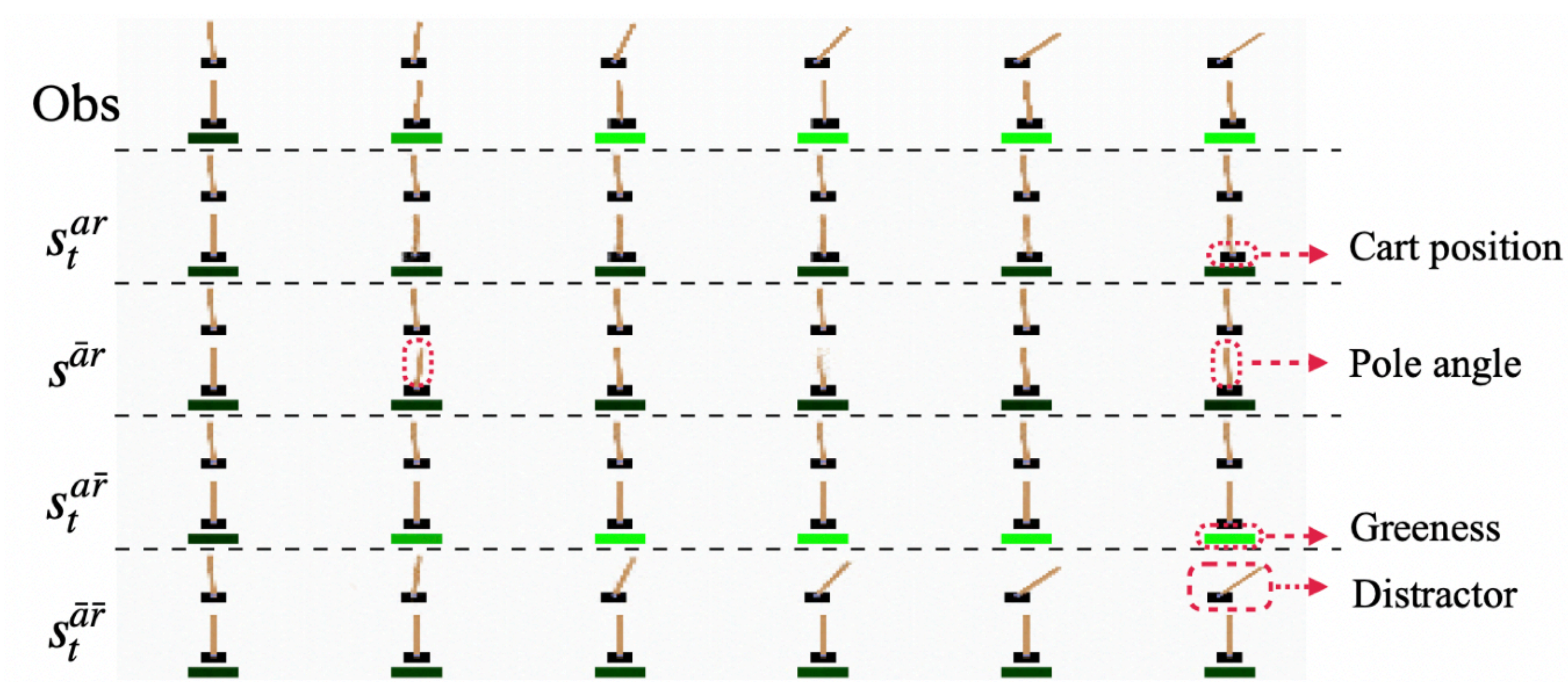
- Each category is identifiable!

- Liu*, Huang*, Zhu, Tian, Gong, Yu, Zhang. Learning world models with identifiable factorization. NeurIPS 2023

Experimental Results on Latent States Recovery



(a) (b) (c) (d)



Modified Cartpole with two distractors

CaRiNG: Learning Temporal Causal Representation under Non-Invertible Generation Process

Guangyi Chen [*1,2] Yifan Shen [*1] Zhenhao Chen [*1] Xiangchen Song [2]
Yuewen Sun [1] Weiran Yao [3] Xiao Liu [1] Kun Zhang [1,2]

We can go further: mixing function g can be nonparametric and noisy

Abstract

Identifying the underlying time-delayed latent causal proces
grasping temp
stream reasoni
robustly identi
rely on strict a
eration process from latent variables to observed data. However, these assumptions are often hard to satisfy in real-world applications containing information loss. For instance, the visual perception

1. Introduction

Sequential data, including video, stock, and climate observations, are integral to our daily lives. Gaining an under-
series data
t al., 2012;
racted con-
dentify the
underlying causal dynamics in the data we observe.

Towards this goal, we focus on Independent Component Analysis (ICA) (Hyvärinen & Oja, 2000), which is a classical method for decomposing the latent signals from mixed

CRL with Temporal Constraints

Non-I, but I.D.	No/Yes	No	(Extended) regression
		Yes	Latent temporal causal processes identifiable!

- Discovering causal relations among the measured time series
  - Granger causality (but be aware of temporal resolution)
- Temporally disentangled representation learning
  - With invertible or non-invertible mixing functions
- With instantaneous relations

ON THE IDENTIFICATION OF TEMPORALLY CAUSAL REPRESENTATION WITH INSTANTANEOUS DEPENDENCE

Zijian Li[†•*] **Yifan Shen**[•*] **Kaitao Zheng**[‡] **Ruichu Cai**[‡] **Xiangchen Song**[†] **Mingming Gong**[⋆]
Guangyi Chen[†•] **Kun Zhang**[†•]
[†]Carnegie Mellon University, Pittsburgh PA, USA
[•]Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
[‡]Guangdong University of Technology, Guangzhou, China
[⋆]The University of Melbourne

ABSTRACT

Temporally causal representation learning aims to identify the latent causal process from time series observations, but most methods require the assumption that the latent causal processes do not have instantaneous relations. Although some recent methods achieve identifiability in the instantaneous causality case, they require either interventions on the latent variables or grouping of the observations, which are in general difficult to obtain in real-world scenarios. To fill this gap, we propose an **ID**entification framework for instantane**O**us **L**atent dynamics (**IDOL**) by imposing a sparse influence constraint that the latent causal processes have sparse time-delayed and instantaneous relations. Specifically, we establish identifiability results of the latent causal process up to a Markov equivalence class based on sufficient variability and the sparse influence constraint by employing contextual information. We further explore under what conditions the identification can be extended to the causal graph. Based on these theoretical results, we incorporate a

Comparison

On the Identification of Temporally Causal Representation with Instantaneous Dependence

Zijian Li[†•*] **Yifan Shen**[•*] **Kaitao Zheng**[‡] **Ruichu Cai**[‡] **Xiangchen Song**[†] **Mingming Gong**[⋆]
Guangyi Chen[†•] **Kun Zhang**[†•]
[†]Carnegie Mellon University, Pittsburgh PA, USA
[•]Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
[‡]Guangdong University of Technology, Guangzhou, China
[⋆]The University of Melbourne

Table 1: The summary of related work of causal representation learning.

	No Intervention	No Grouping	Stationarity	Instantaneous Effect	Temporal Data
IDOL	✓	✓	✓	✓	✓
Yao et al. (2022)	✓	✓	✓	×	✓
Morioka & Hyvärinen (2023)	✓	×	✓	✓	✓
Lippe et al. (2023)	×	✓	✓	✓	✓
Zhang et al. (2024)	✓	✓	×	✓	×

time-delayed and instantaneous relations. Specifically, we establish identifiability results of the latent causal process up to a Markov equivalence class based on sufficient variability and the sparse influence constraint by employing contextual information. We further explore under what conditions the identification can be extended to the causal graph. Based on these theoretical results, we incorporate a

Empirical Results

On the Identification of Temporally Causal Representation with Instantaneous Dependence

Zijian Li[†•*] Yifan Shen[•*] Kaitao Zheng[‡] Ruichu Cai[‡] Xiangchen Song[†] Mingming Gong[⋆] Guangyi Chen[†•] Kun Zhang[†•]
[†]Carnegie Mellon University, Pittsburgh PA, USA
[•]Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE

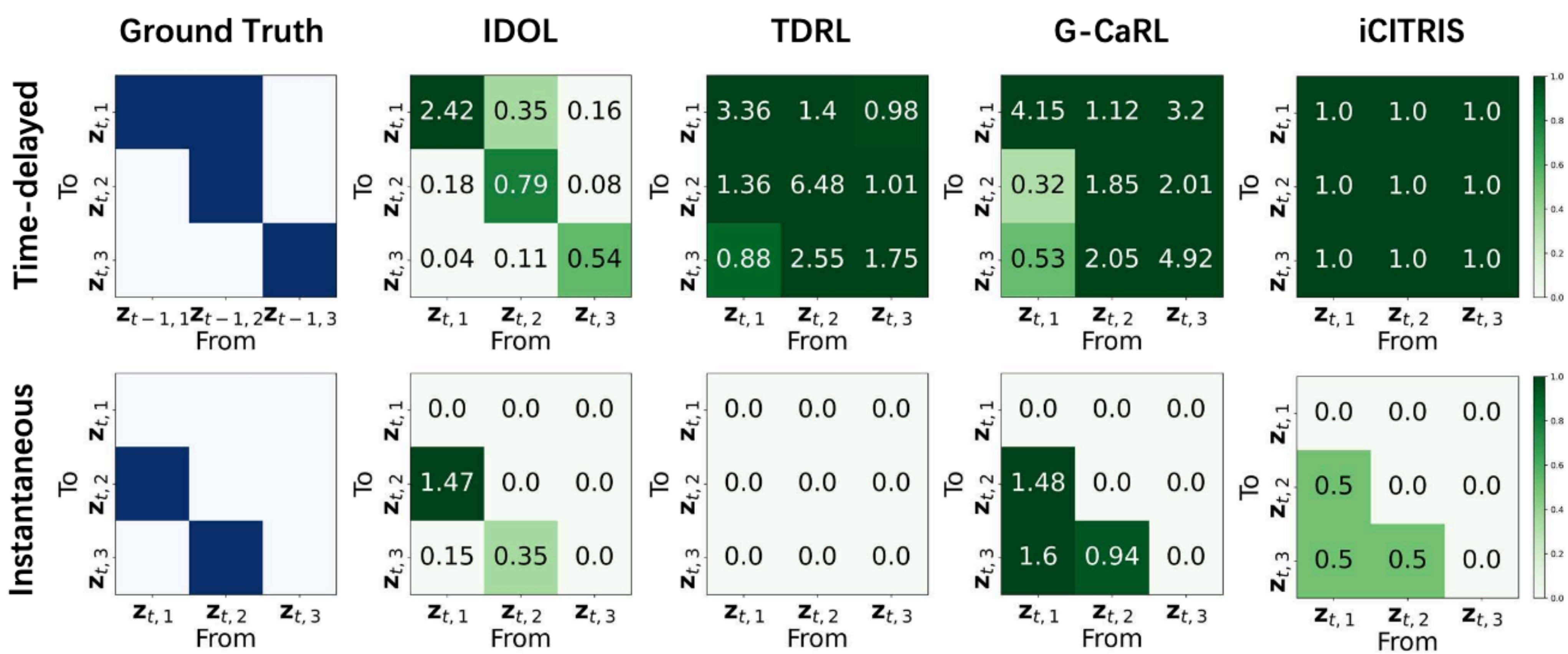
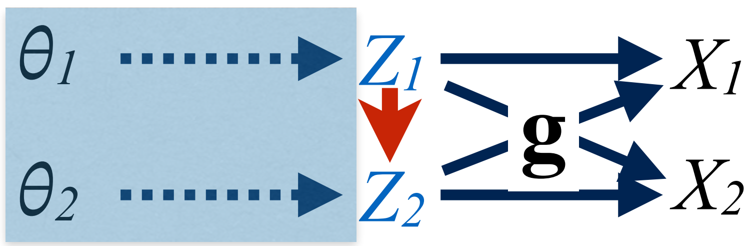


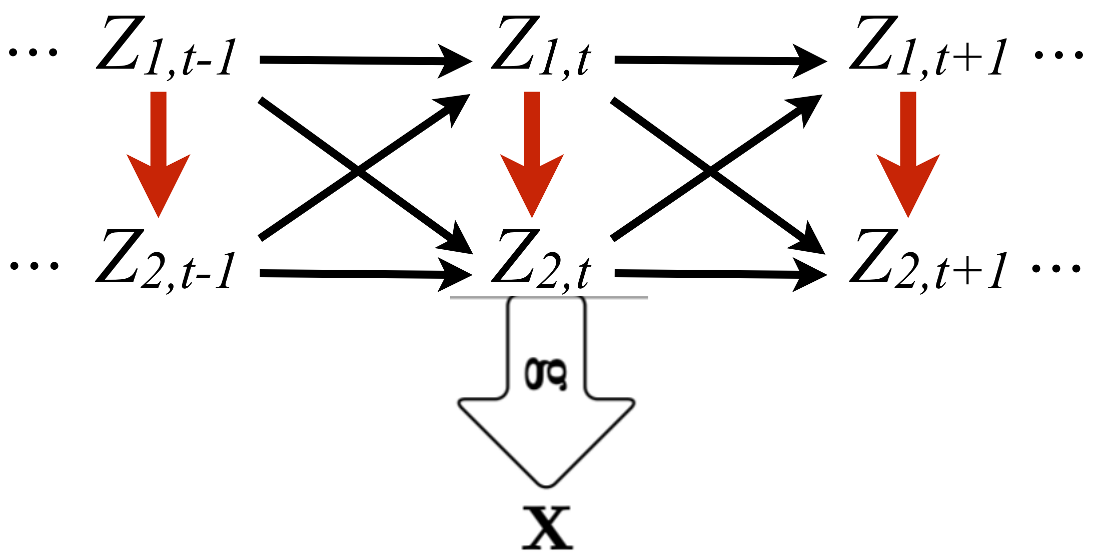
Figure 4: Visualization results of directed acyclic graphs of latent variables of different methods. The first and second rows denote time-delayed and instantaneous causal relationships of latent variables.

Sounds New. But You can See the Connection

- Multi-domain case:

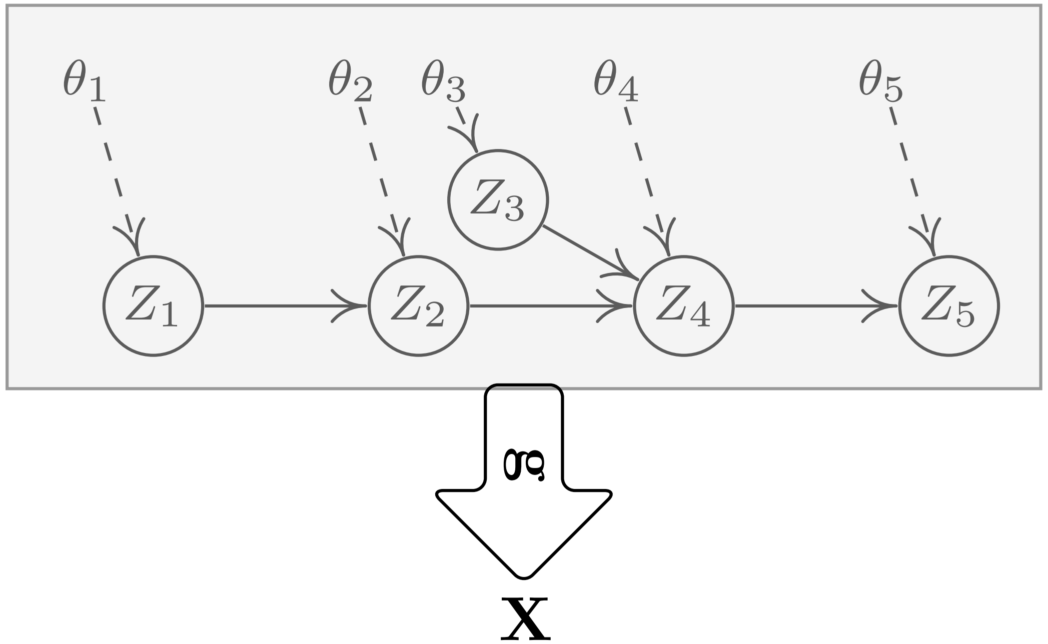


- Temporal case:



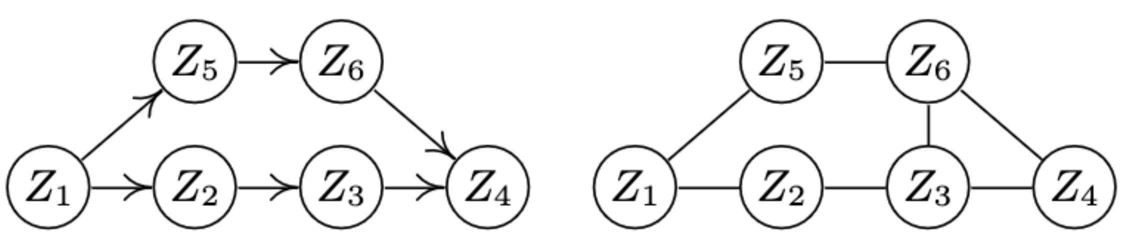
Remember? Causal Representation Learning from Multiple Distributions: A *General* Setting

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes



- Goal: Uncovering hidden variables Z_i with changing causal relations from $\mathbf{X}$ in nonparametric settings
- What is identifia
  - Markov netw
  - Each estimated variable $\tilde{Z}_i$ is a function of Z_i and it **intimate neighbors**
- In this example, each Z_i ($i \neq 4$) can be recovered up to component-wise transformation

We exploit the changes in causal mechanisms along with domain!



(a) $\mathcal{G}_Z$, the DAG over true latent variables Z_i.

(b) The corresponding Markov network $\mathcal{M}_Z$.

- *Zhang, Xie, Ng, Zheng, "Causal Representation Learning from Multiple Distributions: A General Setting," ICML 2024*

Summary: CRL from Temporal Data

- Discovering causal relations among the measured time series
- Temporally disentangled representation learning
- With instantaneous relations

- **Unification—connection with the multi-domain case!**