Causal Discovery in Physical Systems from Videos

Yunzhu Li, Antonio Torralba, Animashree Anandkumar, Dieter Fox, and Animesh Garg

V-CDN: Visual Causal Discovery Network finds object variables, discovers the dependency structures, and models the causal mechanisms end-to-end from images in an unsupervised way.

Introduction

- Causal discovery is at the core of human cognition.
- The interactions in a scene causally affect the system's behavior.

We propose a V-CDN (Visual Causal Discovery Network) that

- extracts a structured keypoint-based representation from videos,
- discovers the causal relationships between different components,
- identifies the hidden confounding variables, and
- makes future predictions.

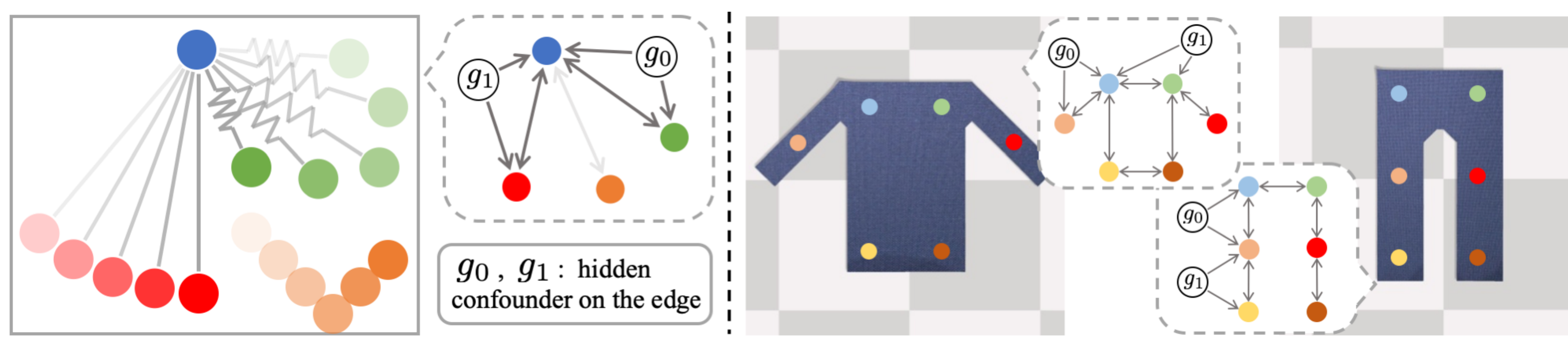

Figure 1: **Causal discovery in physical systems from videos.**

- Left: balls moving around. Hidden confounders on the physical interactions causally affect the system's behavior.
- Right: we can find a reduced-order representation from the images and infer the causal relationships to reflect the topology of the cloth.

V-CDN: Visual Causal Discovery Network

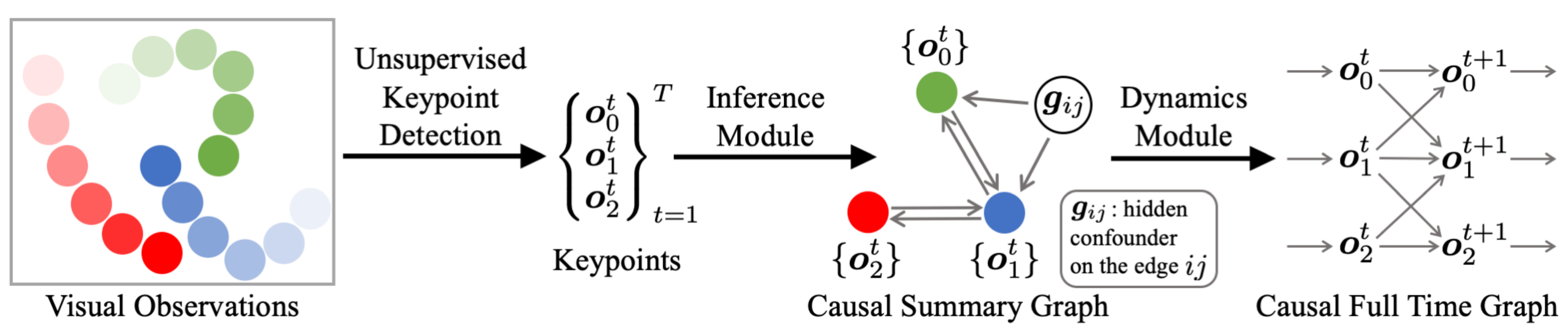

Figure 2: **Model overview**

Our model consists of three components:

- A perception module extracts unsupervised keypoints from images.
- An inference module observes the movements of the keypoints,
  - determines the existence of the causal relations and
  - the associated hidden confounders.
- A dynamics module predicts the future by conditioning on the current state and the inferred causal summary graph.

Unsupervised Keypoint Detection

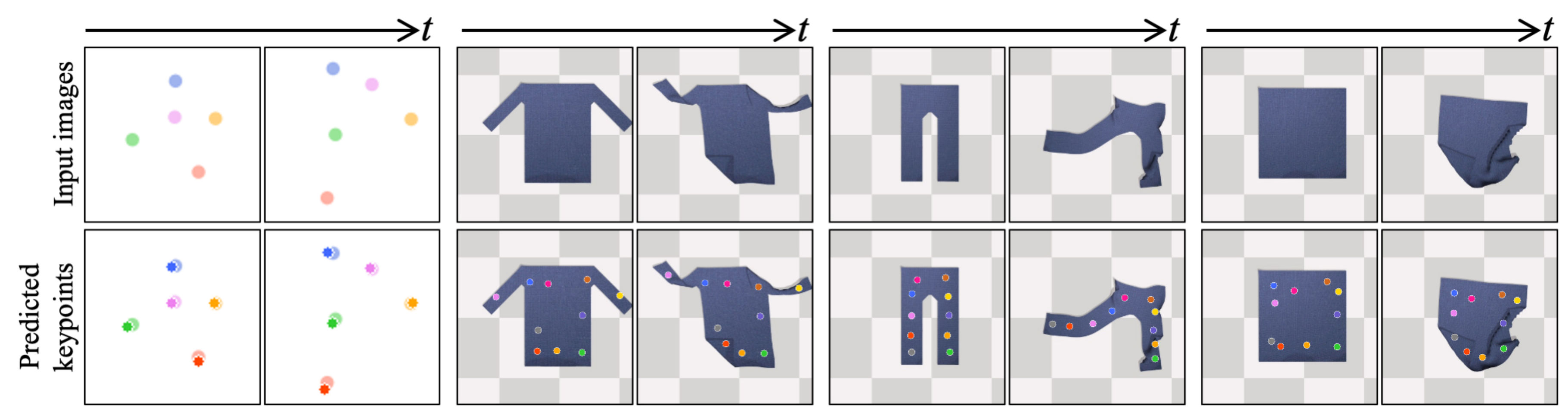

Figure 3: **Unsupervised keypoint detection.**

- The perception module assigns keypoints over the foreground and consistently tracks the objects across different frames.

Predict the *Causal Summary Graph* and the Future

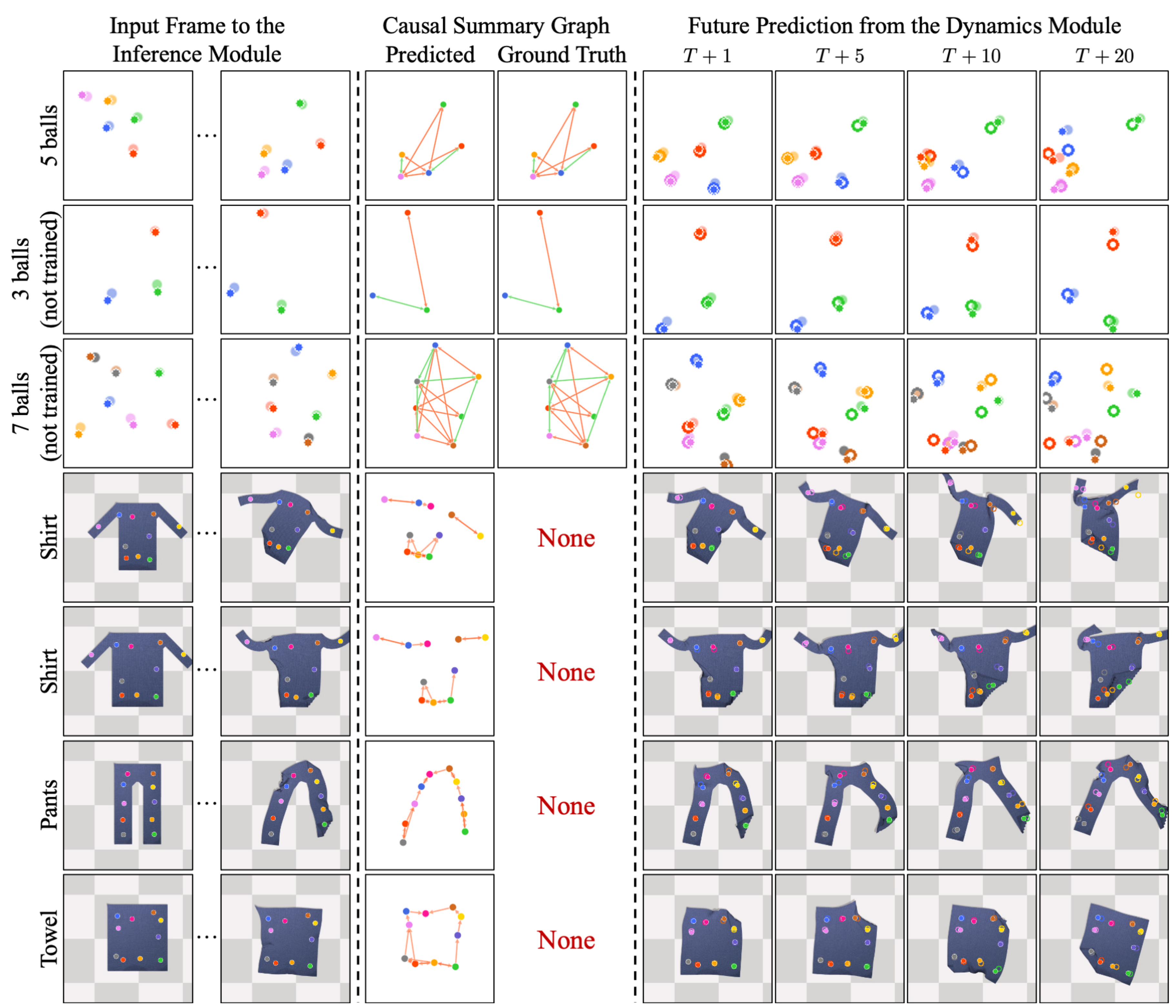

Figure 4: **Predict the *Causal Summary Graph* and the future.**

Our inference module

- recovers the causal graph in the Multi-Body environment, and
- captures the connectivity structures in the Cloth environment.

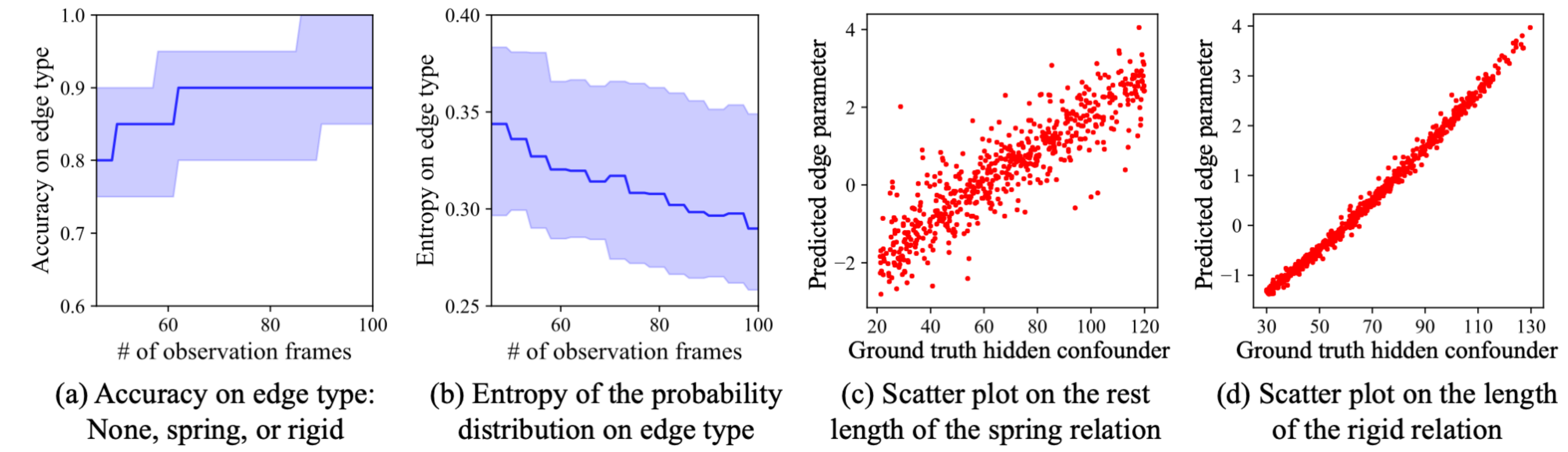

Figure 5: **Results on discovering the Causal Summary Graph.**

- More observation frames lead to higher edge classification accuracy (a) and lower uncertainty (b).
- The inferred continuous variables correlate with the ground truth hidden confounders (c & d).

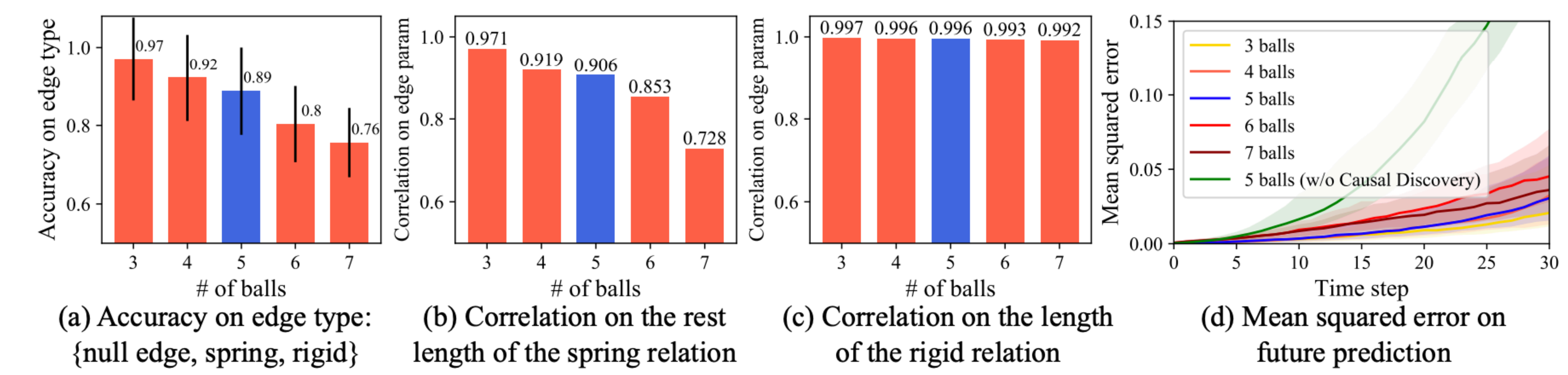

Figure 6: **Extrapolating to unseen graphs of different sizes.**

- Our inference and dynamics modules, trained only on 5 masses, generalize to different numbers of masses from training.

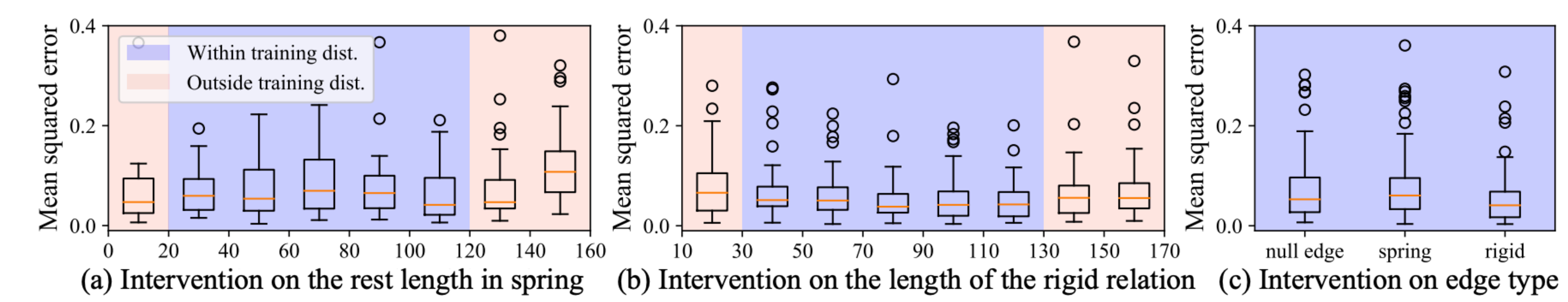

Figure 7: **Results on counterfactual prediction.**

- Counterfactual predictions via intervening in the identified graph.
- Allow extrapolation to param ranges outside the training distribution.

References

[1] Kulkarni et al., "Unsupervised Learning of Object Keypoints for Perception and Control", in **NeurIPS 2019**

[2] Kipf et al., "Neural Relational Inference for Interacting Systems", in **ICML 2018**

[3] Löwe et al., "Amortized Causal Discovery: Learning to Infer Causal Graphs from Time-Series Data", in **arXiv 2020**

Website (video & code)

https://bit.ly/2GFykji