Problem statement
• Causal representation learning considers the identifiability of causal factors from high-dimensional observations such as images for causal reasoning.
• We consider causal factors that are temporally related, and have possibly been intervened upon between time steps.
• The causal factors may be multidimensional, e.g. position in 3D or rotation in multiple axes. Up to what degree can we yet identify the factors?

Causal Identifiability from Temporal Intervened Sequences
• The general setup in terms of a causal graph is shown on the right.
• $C^t$ - a causal factor at time step $t$, $I^{t+1}_t$ - observed intervention target, $X^t$ - high-dimensional observation (e.g. image), $R^{t+1}_t$ - confounder among interventions.
• Key aspect: each causal variable is causally related to only one intervention target.
• Main theoretical result: we can identify the minimal causal variables, i.e. the information/mechanism of a causal variable which strictly depends on the interventions.

Causal Identiﬁability Assumptions:
• Causal structure assumptions: Causal graph is identical to the ground truth graph. Since we have a single object shape, we can only observe a high-dimensional observation.
• Evaluation metrics: (1) correlation between latents and causal factors; (2) novel combinations of causal factors by performing combination in latent space.
• CITRIS outperforms baselines and AE-NF combination works very well.
• Generalization experiments: train NF on 5-object shapes, check its disentanglement on 2 unseen shapes – still works reasonable well.

Experiments
• Experiments on videos with interventions.

Temporal Causal3DIdent
• 3D rendered images with 7 causal factors: 3D position, 2D angles, spotlight rotation, spotlight hue, background hue, object hue, object shape.
• Evaluation metrics: (1) correlation between latents and causal factors; (2) novel combinations of causal factors by performing combination in latent space.
• CITRIS outperforms baselines and AE-NF combination works very well.
• Generalization experiments: train NF on 5-object shapes, check its disentanglement on 2 unseen shapes – still works reasonable well.

Takeaways
• We can identify multidimensional causal variables from high-dimensional, temporal data up to their minimal causal variable.
• CITRIS disentangles factors by enforcing structure in the prior and learns alignment of latent to causal variables.
• CITRIS-NF generalizes to unseen instances of causal factors, holding promise for future work.

Check out our paper and code for details!