Transforming CoPhy

Alexandre St-Aubin

The University of British Columbia

CPSC 532Y – Causal Machine Learning

Problem Definition

Background: CoPhy

- **Origin:** Baradel et al. (Facebook AI, SFU), ICLR 2020.
- **Objective:** Teach a model to reason about counterfactual scenarios in physics.
- **Contributions:** A testing benchmark (CoPhy) and a model (CoPhyNet).

CoPhy: Counterfactual Learning of Physical Dynamics

Fabien Baradel[1] **Natalia Neverova**[2] **Julien Mille**[3] **Greg Mori**[4] **Christian Wolf**[1,5]

[1]Université Lyon, INSA Lyon, CNRS, LIRIS, Villeurbanne, France
[2]Facebook AI Research, Paris, France
[3]Laboratoire dInformatique de lUniv. de Tours, INSA Centre Val de Loire, Blois, France
[4]Simon Fraser University and Borealis AI, Vancouver, Canada
[5]Inria, Chroma group, CITI Laboratory, Villeurbanne, France

Abstract

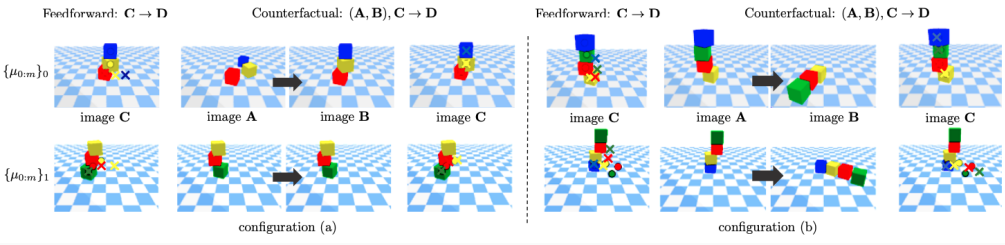
Understanding causes and effects in mechanical systems is an essential component of reasoning in the physical world. This work poses a new problem of counterfactual learning of object mechanics from visual input. We develop the CoPhy benchmark to assess the capacity of the state-of-the-art models for causal physical reasoning in a synthetic 3D environment and propose a model for learning the physical dynamics in a counterfactual setting. Having observed a mechanical experiment that involves, for example, a falling tower of blocks, a set of bouncing balls or colliding objects, we learn to predict how its outcome is affected by an arbitrary intervention on its initial conditions, such as displacing one of the objects in the scene. The alternative future is predicted given the altered past and a latent representation of the confounders learned by the model in an end-to-end fashion with no supervision of confounders. We compare against feedforward video prediction baselines and show how observing alternative experiences allows the network to capture latent physical properties of the environment, which results in significantly more accurate predictions at the level of super human performance.

Project Objectives

Core Goal: Modernize the CoPhyNet architecture by replacing the sequential RNN encoder with a **Transformer**.

1. **Accelerate Training Efficiency**
   Leverage the parallel nature transformers to process time-series data faster.
2. **Enhance Physical Reasoning**
   Improve the estimation of latent confounders (U).
3. **Benchmark Performance**
   Validate improvements in stability classification and trajectory MSE on the CoPhy benchmark.

Problem Formulation



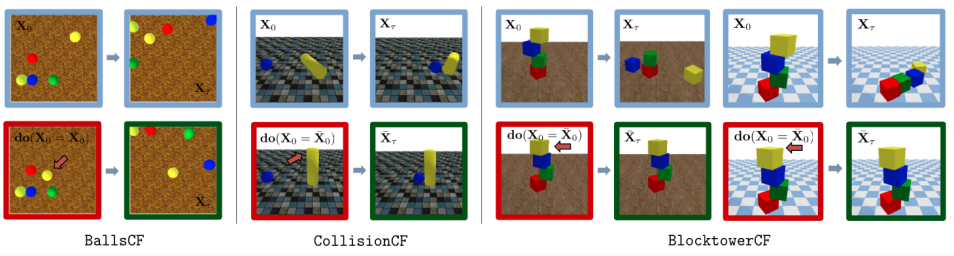
The Objective

Given a triplet $\{A, B, C\}$:

- $A := X_0$ (Initial State)
- $B := \{X_1, \ldots, X_\tau\}$ (Observed Outcome)
- $C := \overline{X}_0$ (Intervened Initial State)

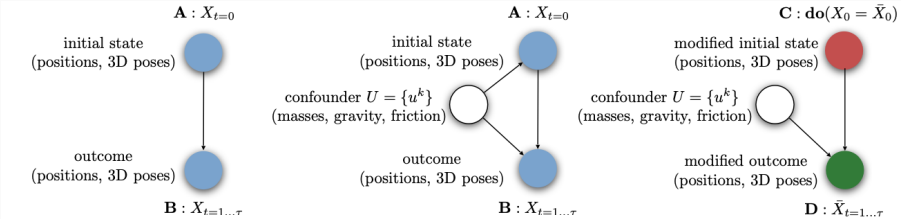
Goal: Predict the counterfactual outcome $D := \{\overline{X}_1, \ldots, \overline{X}_\tau\}$.

The Tasks



▸ Watch Demo Video (YouTube)

Structural Causal Model (SCM)



(a) feedforward future forecasting

(b) counterfactual future forecasting

The physical laws U determine valid initial states X_0 and outcomes.

The CoPhy Guarantee: Anti-Shortcut

The Constraint

"We enforce existence of at least two different confounder configurations resulting in significantly different trajectories."

Example: The Ambiguous Tower

- **Scenario A (High Friction):** The tower stands.
- **Scenario B (Low Friction):** The exact same tower collapses.

$$\textit{Visually identical } X_0 \rightarrow \textit{Divergent Futures.}$$

Implication: Simple extrapolation fails. Estimating U is necessary.

Applications

Why does this matter?

- **Robotics (CausalCF):** Use latent confounder representation to train RL agents.
- **Robustness:** Makes agents robust to unseen changes in the environment (e.g., slippery floors, heavier payloads).
- **Trustworthiness:** Agents make decisions based on physical understanding, not pixel statistics.

CoPhyNet

The Pearl Framework (2009)

The theoretical foundation for the 3-step process:

Theorem 7.1.7
Given model $\langle M, P(u)\rangle$, *the conditional probability* $P(B_A \mid e)$ *of a counterfactual sentence "If it were A then B," given evidence e, can be evaluated using the following three steps.*

1. ***Abduction*** – *Update* $P(\mathrm{u})$ *by the evidence e to obtain* $P(u \mid e)$.
2. ***Action*** – *Modify M by the action* $do(A)$, *where A is the antecedent of the counterfactual, to obtain the submodel* M_A.
3. ***Prediction*** – *Use the modified model* $\langle M_A, P(u \mid e)\rangle$ *to compute the probability of B, the consequence of the counterfactual.*

Pearl, J. (2009). Causality: Models, reasoning, and inference.

The 3-Step Process in CoPhyNet

1. **Abduction (Infer U):** Use observed data $\mathrm{A} = X_0, \mathrm{B} = X_{1:\tau}$ to compute the latent representation U.

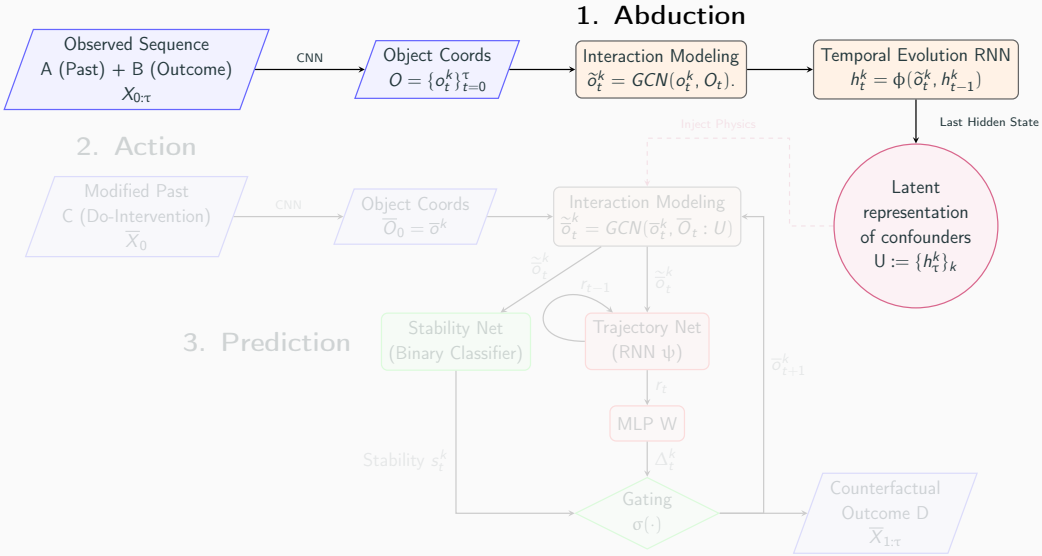
The 3-Step Process in CoPhyNet

1. **Abduction (Infer U):** Use observed data $\mathrm{A} = X_0, \mathrm{B} = X_{1:\tau}$ to compute the latent representation U.
2. **Action (Intervene):** Update the causal model. Keep identified confounders U, but apply $do(X_0 = \overline{X}_0)$.

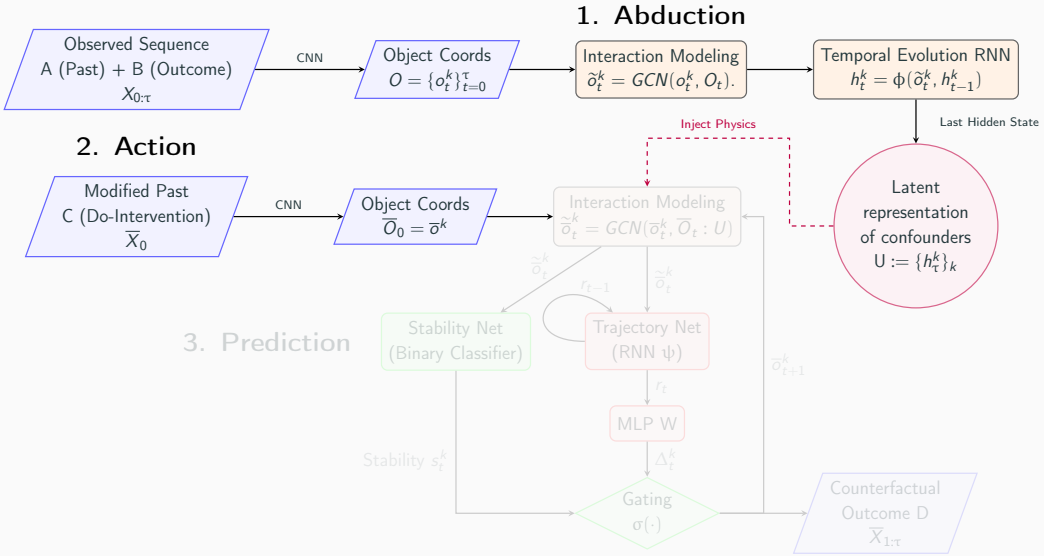
The 3-Step Process in CoPhyNet

1. **Abduction (Infer U):** Use observed data $\mathrm{A} = X_0, \mathrm{B} = X_{1:\tau}$ to compute the latent representation U.
2. **Action (Intervene):** Update the causal model. Keep identified confounders U, but apply $do(X_0 = \overline{X}_0)$.
3. **Prediction (Simulate):** Compute counterfactual outcome $\mathrm{D} = \overline{X}_{1:\tau}$ using the modified causal model.

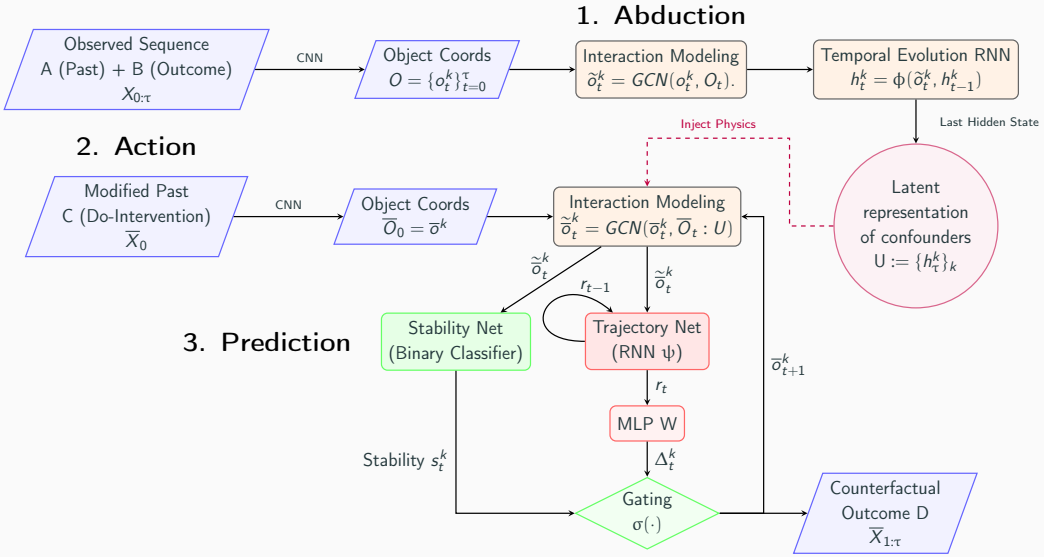
The 3-Step Process in Architecture



The 3-Step Process in Architecture



The 3-Step Process in Architecture



Loss Function: "Whether" vs "How"

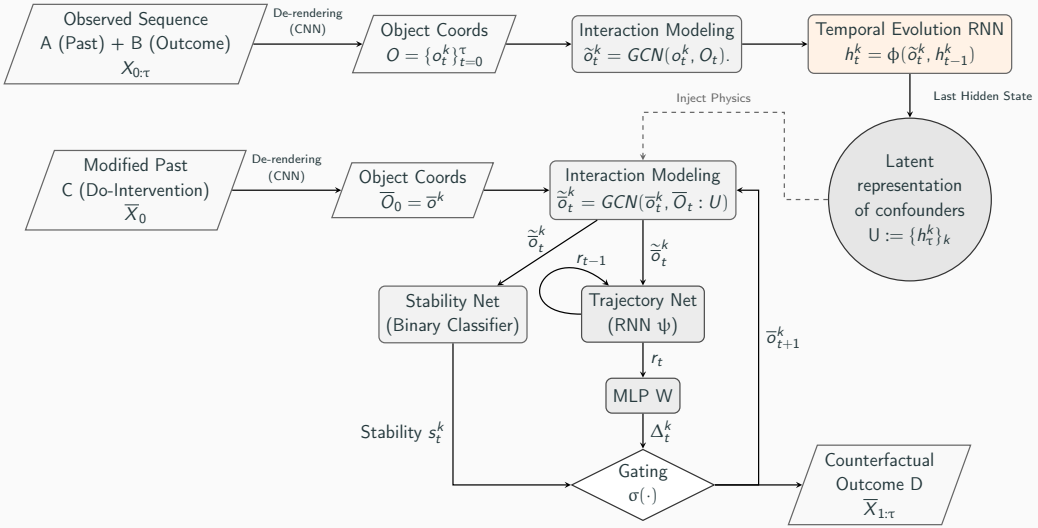
Separating **Stability** (Whether-causation) from **Trajectory** (How-causation):

$$\mathcal{L}_{e2e} = \gamma \cdot \sum_{k=1}^{K} \underbrace{\mathcal{L}_{ce}(\mathrm{s}^k, \mathrm{s}^{k*})}_{\text{Whether (Classification)}} + \alpha \cdot \sum_{t=0}^{\tau} \left[\sum_{k=1}^{K} \underbrace{\mathcal{L}_{mse}(\overline{o}_t^k, \overline{o}_t^{k*})}_{\text{How (Regression)}} \right]$$

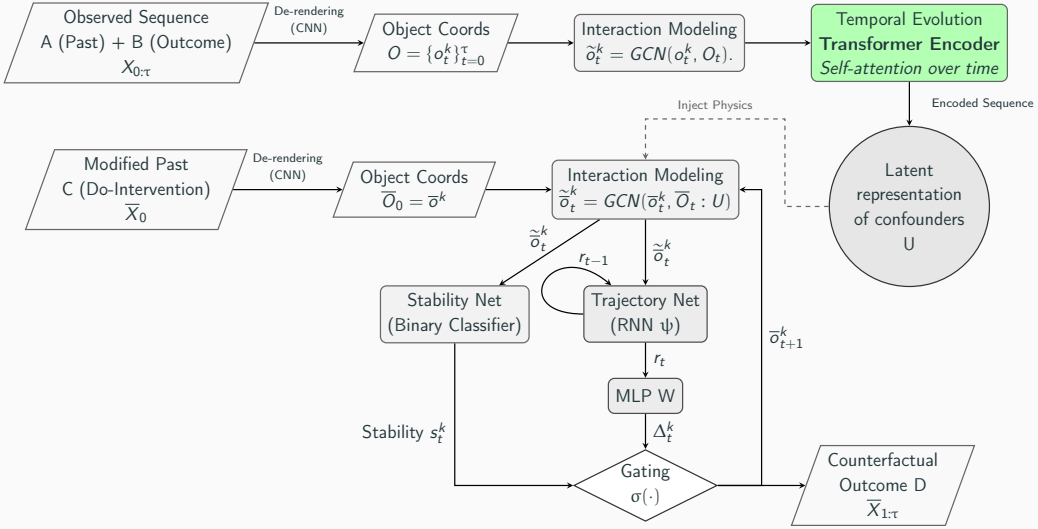
Inspired by: Gerstenberg et al. (2015). "How, whether, why: Causal judgments as counterfactual contrasts."

Transforming CoPhyNet

Original Architecture

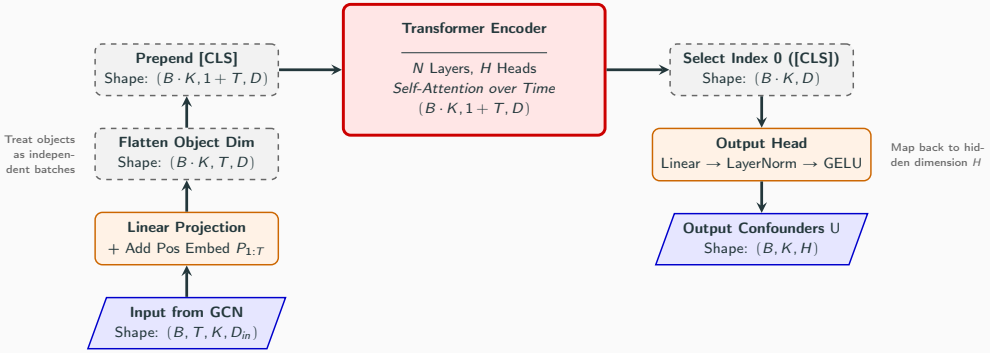


New Architecture



Proposal: Transformer Encoder

Replacing the RNN with a Transformer:



Hypothesis: RNN vs. Transformer

Feature	RNN (GRU)	Transformer
Time View	Sequential	**Simultaneous**
Info Flow	Bottleneck at h_t	Direct access to history
Event Detection	Hard (Vanishing Gradients)	**Easy** (Self-Attention)
Training	Slow (Sequential)	Fast (Parallelizable)

Experiments

Experimental Setup

Dataset & Task

- **Data:** CoPhy `BallsCF` (3 Objects, Normal Split)
- **Goal:** Predict counterfactual stability and 3D trajectory.

Comparison Framework

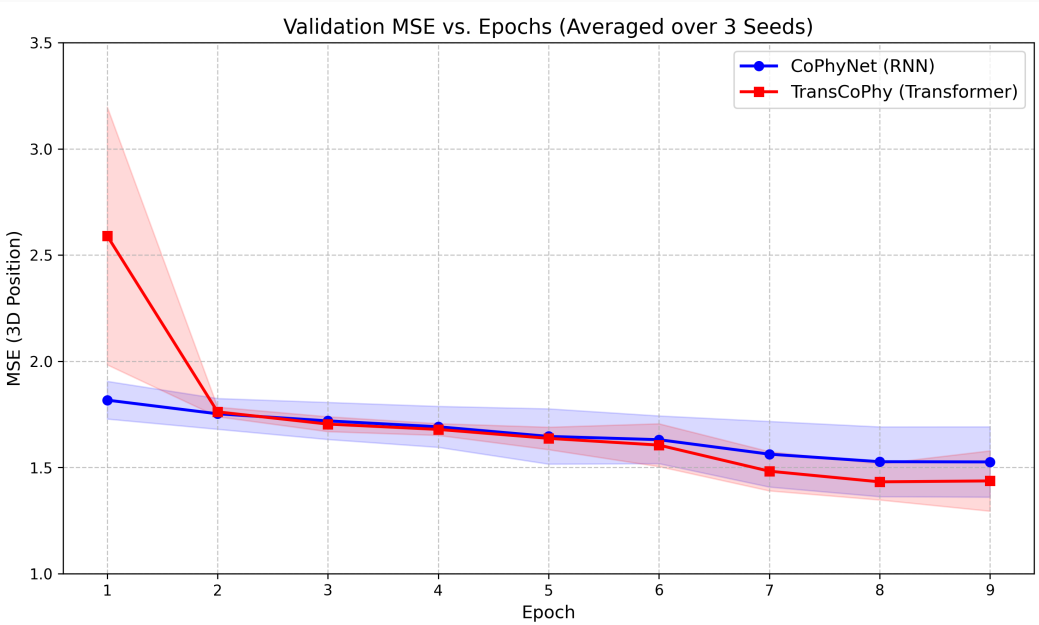
- **Models:** Baseline CoPhyNet vs. TransCoPhy.
- **Metrics:** Mean Squared Error (MSE) on 3D positions and computational cost (Training Time).

Both models use identical batch sizes and hardware. *Note: Learning rates and warmup schedules were tuned specifically for each architecture's stability requirements.*

Results

Model	CoPhyNet (RNN)	TransCoPhy (Ours)
epoch train time	$165s$	$135s$ (-18%)

Results


Validation MSE vs. Epochs (Averaged over 3 Seeds)
CoPhyNet (RNN)
TransCoPhy (Transformer)
MSE (3D Position)
Epoch
3.5
3.0
2.5
2.0
1.5
1.0
1
2
3
4
5
6
7
8
9

Conclusion

Future Work & Conclusion

- **Current Limitation:** Benchmark restricted to 6 seconds @ 5fps.
- **Scaling Up:** Newer physics datasets use 25fps+ and longer horizons.
- **The Transformer Advantage:** RNNs struggle with long sequences (vanishing gradients). Transformers should show drastic improvements on longer, higher-frequency simulations.

Thank You.