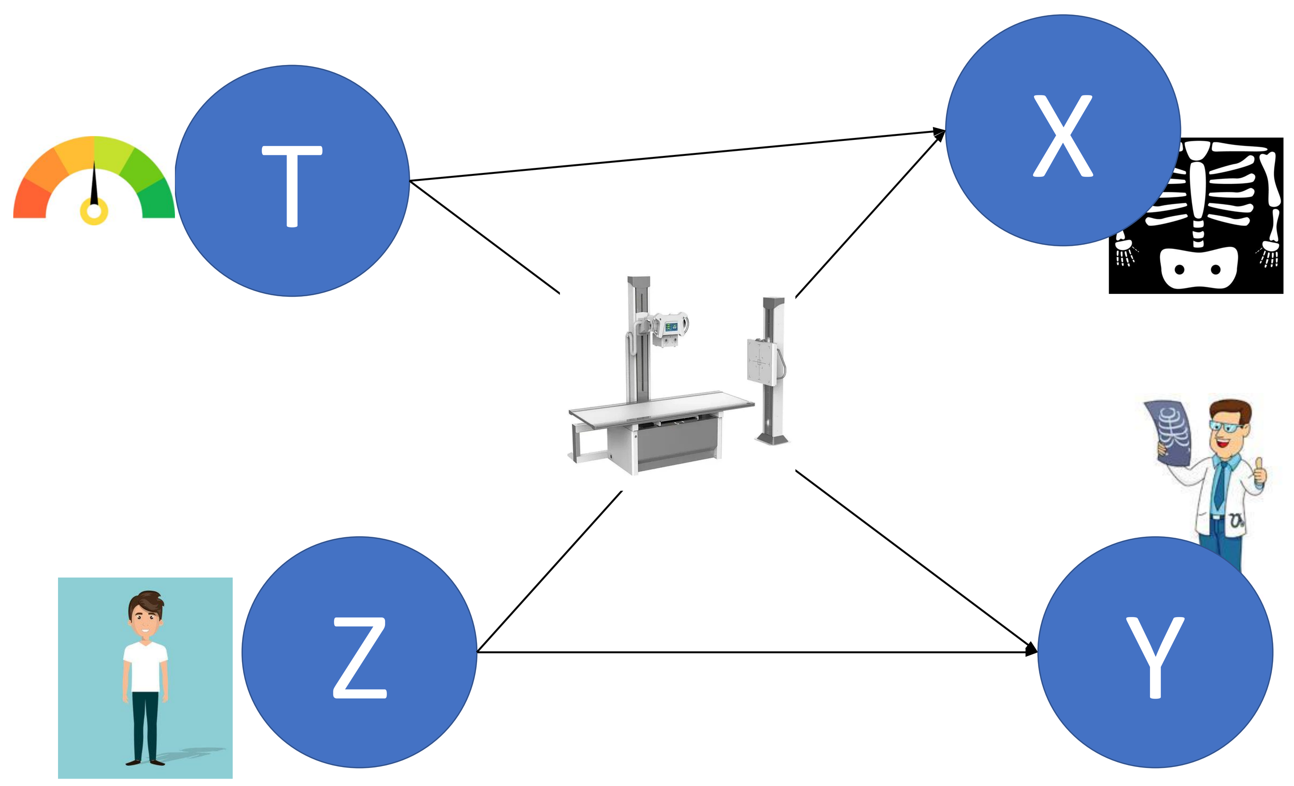
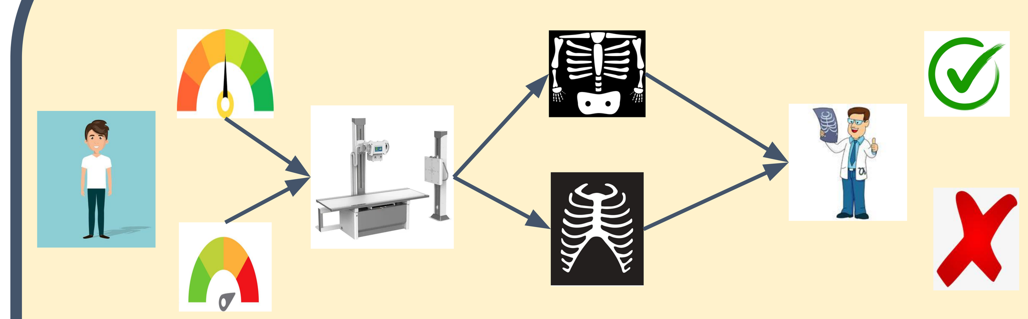
# Leveraging a Simulator for Learning Causal Representations for CATE

Lokesh N, Pranava S, Avishek Ghosh, Sunita Sarawagi




## Problem Statement



Given post-T input, determine if we need to change the T

**Treatment Effect Estimation:** Predict change in Y as T changes for a particular Z

## Observational Dataset

Logged Dataset comprising:

- Assigned Treatment T
- Post-Treatment Covariates X(T)
- Observed Outcome Y(T)

Obs Data

## Impossibility Result

*Lemma 1. The causal effect of T on Y is* ***not identifiable*** *from i.i.d. samples of the observational Dataset.*

**Insights:**

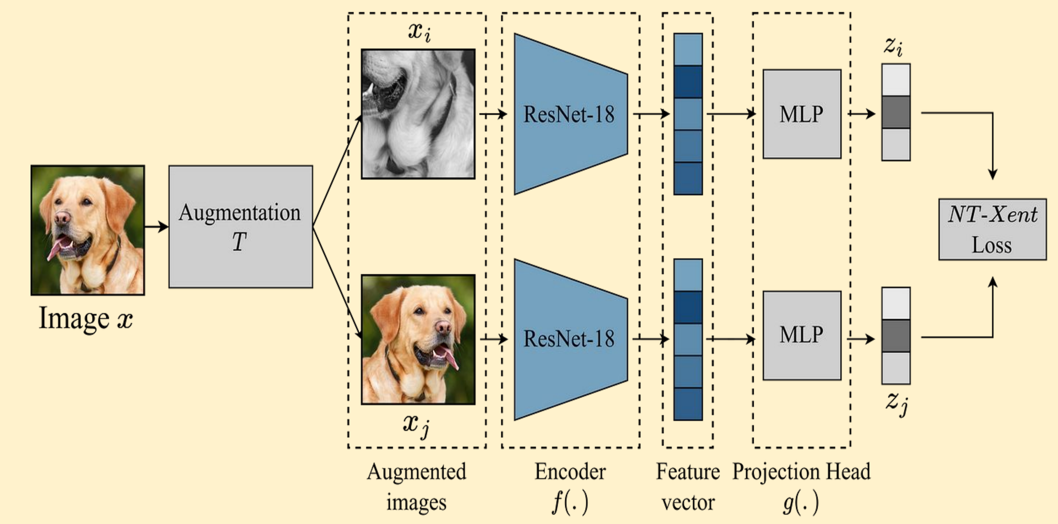
→ Certain Assumptions are inevitable.

→ Need to extract **causal** representation Z from X that affect Y

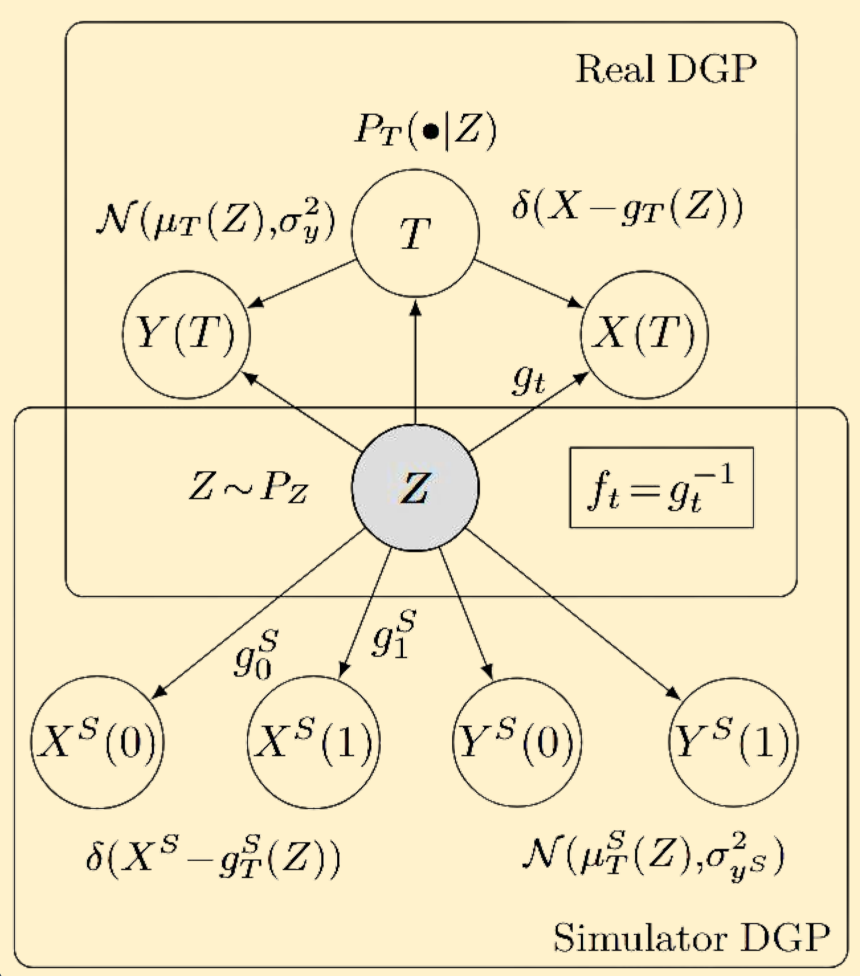
## Recovery of Causal Reps Z from X(T)

Possible using contrastive training over counterfactual pairs of the form X( ), X( ) in Obs Data

Julius von Kügelgen et. al.




## Real/Sim Data Generating Process




| | | |
|---|---|---|
| **SimOnly** | Train for CATE leveraging counterfactual samples from the Simulator | Needs Agreement in **μ** and ***τ*** |
| **RealOnly** | Train only on the real samples in the Observational Data | Good factual prediction **μ**(**x**(t), t). But fails for CF **μ**(**x**(t), 1-t). |
| **Realμ-Simf** | Learns Causal reps on Synthetic CF Samples and applies as-in to real samples. | Works well only when **g** = **gs** |

## Our Approach: SimPONet

### Error Bound on CATE from port-T covariates

**Lemma 4.** *Assume $\tau$ is $K_\tau$-Lipschitz, and $\widetilde{f}^S$ and $\widetilde{\tau}^S$ are estimates from the simulator DGP obtained from the optimization in Eq. 1, 2. Then, the CATE error on the estimates $\widehat{f}_t$ and $\widehat{\tau}$ admits the following bound:*

$$\mathcal{E}^t_{CATE}(\widehat{f}_t,\widehat{\tau})\leq[8\mathcal{E}^t_F+12d_h(\widehat{\tau},\widetilde{\tau}^S)+12K_\tau^2 d_{\boldsymbol{x}|t}(\widehat{f}_t,\widetilde{f}^S_t)]+[12d_z(\tau,\tau^S)+12K_\tau^2 d_{\boldsymbol{x}|t}(f_t,f^S_t)]$$

### SimPONet's Joint Training Objective

$$\min_{\{\widehat{\mu}_t,\widehat{f}_t\}} \underbrace{\sum_{D_{\text{trn}}}\Big(y_i-\widehat{\mu}_{t_i}(\widehat{f}_{t_i}(\boldsymbol{x}_i))\Big)^2}_{\text{Factual Loss on } D_{\text{trn}}}+\lambda_f\underbrace{\sum_{D_{\text{trn}}}\|\widetilde{f}^S_{t_i}(\boldsymbol{x}_i)-\widehat{f}_{t_i}(\boldsymbol{x}_i)\|_2^2}_{d(\widetilde{f}^S_t,\widehat{f}_t)\text{ regularizer}}+\lambda_\tau\underbrace{\sum_{D_{\text{syn}}}\sum_{t\in\{0,1\}}\Big(\tau^S_i-\widehat{\tau}(\widetilde{f}^S_t(\boldsymbol{x}^S_i(t)))\Big)^2}_{\tau^S\text{ regularizer on } D_{\text{syn}}} \quad (3)$$

where $\tau^S_i=y^S_i(1)-y^S_i(0)$ and $\lambda_\tau,\lambda_f>0$ are loss weights. $\widehat{\tau}(\bullet)=\widehat{\mu}_1(\bullet)-\widehat{\mu}_0(\bullet)$ denotes the estimated CATE.

## Comparison with SOTA

| Method | IHDP | ACIC-2 | ACIC-7 |
|---|---|---|---|
| RNet (Nie & Wager, 2021) | 1.54 (0.00) | 3.30 (0.00) | 5.91 (0.04) |
| XNet (Künzel et al., 2019) | 1.0 (0.00) | 0.43 (0.15) | 5.49 (0.17) |
| DRNet (Schwab et al., 2020) | 0.96 (0.00) | 0.24 (0.59) | 5.53 (0.15) |
| CFRNet (Shalit et al., 2017) | 0.96 (0.00) | 0.36 (0.26) | 5.55 (0.15) |
| FlexTENet (Curth & van der Schaar, 2021) | 0.96 (0.00) | 0.32 (0.32) | 5.46 (0.19) |
| DragonNet (Shi et al., 2019) | 0.96 (0.00) | 0.29 (0.41) | 5.57 (0.14) |
| IPW (Robins et al., 1994) | 0.96 (0.00) | 0.36 (0.24) | 5.56 (0.15) |
| $k$-NN (Stuart, 2010) | 0.96 (0.00) | 0.33 (0.33) | 5.48 (0.18) |
| PerfectMatch (Schwab et al., 2018) | 0.98 (0.00) | 0.56 (0.11) | 5.75 (0.08) |
| StableCFR (Wu et al., 2023) | 1.01 (0.00) | 1.09 (0.03) | 5.56 (0.15) |
| ESCFR (Wang et al., 2024) | 0.96 (0.00) | 0.27 (0.47) | 5.55 (0.15) |
| PairNet (Nagalapatti et al., 2024b) | 0.97 (0.00) | 0.12 (0.85) | 5.46 (0.23) |
| SimOnly | 0.94 (0.00) | **0.00 (0.98)** | 6.65 (0.00) |
| RealOnly | 0.83 (0.13) | 11.23 (0.01) | 14.81 (0.05) |
| Real$_\mu$Sim$_f$ | 0.96 (0.00) | 0.17 (0.76) | 5.57 (0.14) |
| *SimPONet* | **0.79 (0.00)** | 0.26 (0.00) | **5.04 (0.00)** |

## Across Real/Sim Gaps

| | | | Synthetic-Gaussian | | | | Real World-IHDP | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $d(f_0,f_1)$ | $d(f_t,f_t^S)$ | $d(\tau,\tau^S)$ | SimOnly | RealOnly | Real$_\mu$Sim$_f$ | *SimPONet* | SimOnly | RealOnly | Real$_\mu$Sim$_f$ | *SimPONet* |
| 0.00 | high | high | 2.82 (0.27) | **0.00 (1.00)** | 15.75 (0.01) | 2.58 (0.00) | 3.57 (0.11) | **0.00 (1.00)** | 48.76 (0.05) | 3.20 (0.00) |
| low | low | low | 0.63 (0.00) | 2.47 (0.02) | 1.19 (0.01) | **0.54 (0.00)** | 1.00 (0.44) | 3.43 (0.02) | 2.73 (0.00) | **0.97 (0.00)** |
| low | low | high | 1.57 (0.16) | 2.47 (0.08) | **1.19 (0.83)** | 1.39 (0.00) | 1.62 (0.26) | 3.43 (0.04) | 2.73 (0.02) | **1.49 (0.00)** |
| low | high | low | 2.14 (0.22) | 2.47 (0.00) | 15.75 (0.01) | **1.85 (0.00)** | 3.67 (0.31) | 3.43 (0.48) | 48.76 (0.05) | **3.37 (0.00)** |
| low | high | high | 2.82 (0.26) | **2.47 (0.56)** | 15.75 (0.01) | 2.57 (0.00) | 3.57 (0.11) | 3.43 (0.39) | 48.76 (0.05) | **3.19 (0.00)** |
| high | low | low | 0.63 (0.00) | 13.86 (0.02) | 1.19 (0.01) | **0.54 (0.00)** | 1.00 (0.47) | 47.78 (0.06) | 2.73 (0.00) | **0.98 (0.00)** |
| high | low | high | 1.57 (0.16) | 13.86 (0.03) | **1.19 (0.83)** | 1.39 (0.00) | 1.62 (0.27) | 47.78 (0.06) | 2.73 (0.02) | **1.50 (0.00)** |
| high | high | low | 2.14 (0.21) | 13.86 (0.03) | 15.75 (0.01) | **1.85 (0.00)** | 3.67 (0.31) | 47.78 (0.06) | 48.76 (0.05) | **3.38 (0.00)** |
| high | high | high | 2.82 (0.26) | 13.86 (0.04) | 15.75 (0.01) | **2.57 (0.00)** | 3.57 (0.11) | 47.78 (0.06) | 48.76 (0.05) | **3.19 (0.00)** |

Real and Simulator DGPs assume linear functions

$d(f_0,f_1)$ : Impact of T on covariates.

$d(f_t,f_t^S)$ : Gap between real and simulated Covariates

$d(\tau,\tau^S)$ : Gap between real and simulated Treat. Effects

## Conclusion

- We estimated CATE from post-Treatment covariates using a simulator.
- SimPONet uses the simulator only to the extent it enhances CATE beyond what training on real data alone can achieve.
- Across DGPs, and across CATE Baselines, SimPONet proved to be the best CATE estimator for post-T covariates

Contact: N Lokesh <nlokeshiisc@gmail.com>