Generative Causal Representation Learning for Out-of-Distribution Motion Forecasting

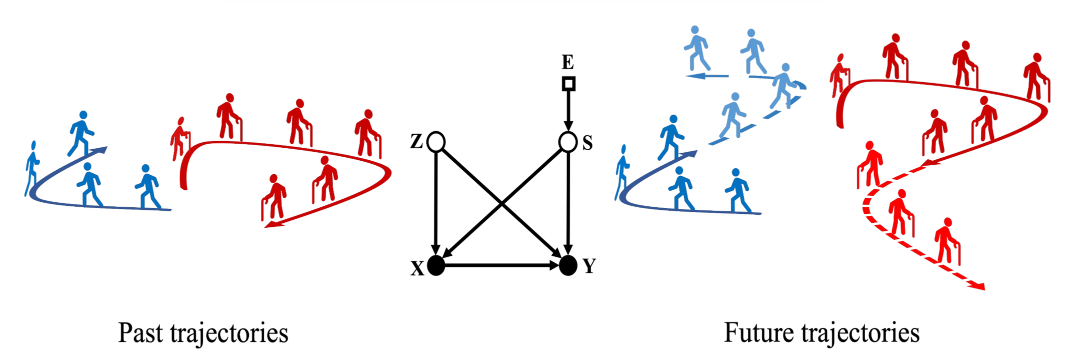
Shayan Shirahmad Gale Bagi[1], Zahra Gharaee[2], Oliver Schulte[3], Mark Crowley[1]

[1]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada
[2]Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada
[3]School of Computing Science, Simon Fraser University, Burnaby, Canada

ABSTRACT

Conventional supervised learning methods typically assume i.i.d samples and are found to be sensitive to out-of-distribution (OOD) data. We propose Generative Causal Representation Learning (**GCRL**) which leverages causality to facilitate knowledge transfer under distribution shifts. First, we propose a novel causal model that explains the generative factors in motion forecasting datasets using features that are common across all environments and with features that are specific to each environment. Selection variables are used to determine which parts of the model can be directly transferred to a new environment without fine-tuning. Second, we propose an end-to-end variational learning paradigm to learn the causal mechanisms that generate observations from features. Experimental results on synthetic and real-world motion forecasting datasets show the robustness and effectiveness of our proposed method for knowledge transfer under zero-shot and low-shot settings.

PROBLEM DEFINITION

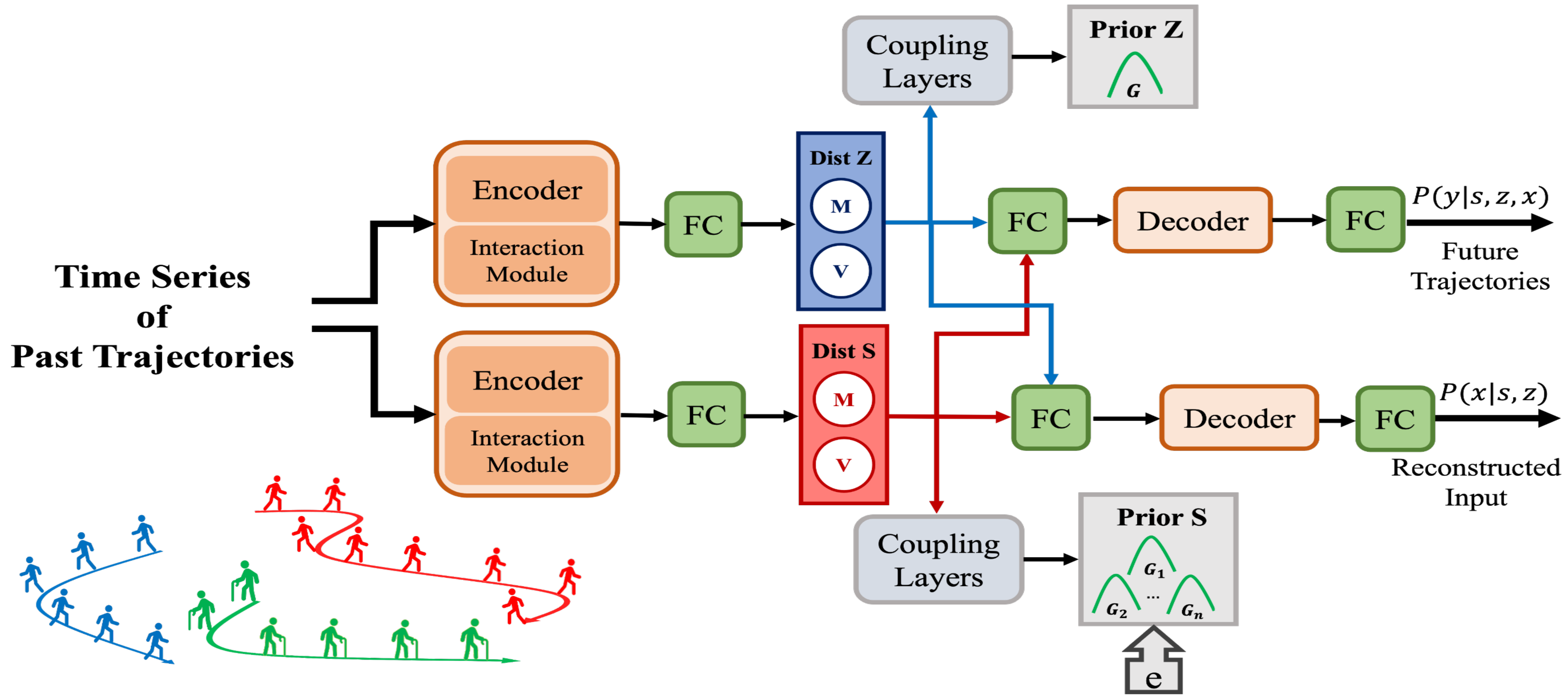


- **Given:**
  - 2D coordinates of agents in the observed T_{obs} time steps
  - Training data with multiple environments.
- **Objectives:**
  - Predicting the states of agents in the next T_{pred} time steps in a new environment.
- **Examples scenes from ETH-UCY dataset:**

Image taken from (Amirian et al., 2020)

PROPOSED METHOD

- **Endogenous variables:**
  - Past trajectories (X)
  - Future trajectories (Y)
  - Invariant features (Z)
  - Variant features (S)
- **Exogenous variables:**
  - Selection variable (E)
- **Examples:**
  - Invariant features are common across domains. These features can be associated with physical laws.
  - variant features vary across domains and can be associated with the motion styles.
  - The selection variable acts as an identifier of an environment



METHOD	ADE/FDE $\alpha = 8$	ADE/FDE $\alpha = 16$	ADE/FDE $\alpha = 32$	ADE/FDE $\alpha = 64$
BASELINE (HUANG ET AL., 2019)	**0.80/1.37**	2.15/3.80	2.64/4.44	2.68/4.48
COUNTERFACTUAL (CHEN ET AL., 2021)	0.80/1.59	1.62/2.68	2.32/3.90	2.68/4.52
INVARIANT $\lambda = 1.0$ (LIU ET AL., 2022)	0.94/1.65	1.04/**1.76**	1.52/2.55	1.96/3.35
INVARIANT $\lambda = 3.0$ (LIU ET AL., 2022)	0.91/1.67	0.99/1.87	1.18/2.20	1.27/2.33
INVARIANT $\lambda = 5.0$ (LIU ET AL., 2022)	0.98/1.79	1.00/1.83	1.06/1.90	1.56/2.58
GCRL (OURS)	0.97/1.8	**0.97**/1.8	**0.97/1.8**	**0.97/1.8**

- **The objective function in GCRL is given as:**

$$\max_{p,q} \mathrm{E}_{p^*(\mathrm{x,y})}[\log q(y|x) + \frac{1}{q(y|x)} \mathrm{E}_{q(s|x),q(z|x)}[p(y|x,s,z) \log \frac{p(x|s,z)p(s)p(z)}{q(s|x)q(z|x)}$$

- $p(s) = \sum_{e \in E} p(s|e)p(e)$ which means that S has a Gaussian mixture prior
- $q(y|x) = E_{q(s|x),q(z|x)}[p(y|s,z,x)] = p(y|do(x))$ which can be calculated by ancestral sampling and eliminates the confounding bias.
- Consequently, **GCRL** learns:
  1. To minimize the distance between ground-truth future trajectories and predicted future trajectories via maximizing $\log q(y|x)$
  2. To eliminate the confounding effect by estimating the causal effect of X on Y via $p(y|do(x))$
  3. To reconstruct past trajectories via maximizing $\log p(x|s,z)$
  4. Invariant representations via maximizing $\log \frac{p(z)}{q(z|x)}$
  5. Variant representations via maximizing $\log \frac{p(s)}{q(s|x)}$
- **GCRL** learns to predict the future trajectories with a generative approach, hence, it can tackle the multi-modality of trajectories.
- Z is invariant, we can directly transfer it to the new domain without any fine-tuning
- In *test-time adaptation*, we may need to fine-tune the components of the GMM and obtain a new prior for S depending on how related the test domains are to the training domains.

RESULTS

- Table: our method is robust against observation noise while performing comparably with other motion forecasting models.
- Our proposed method also learns to reconstruct inputs: eliminates the effect of noise by reconstructing uncorrupted inputs
- Figure: **GCRL** adapts to the new environment faster than IM (Liu et al., 2022)
- is more robust to OOD-Extra shifts.
- **GCRL** improves the ADE by 34.3% from IM on average.

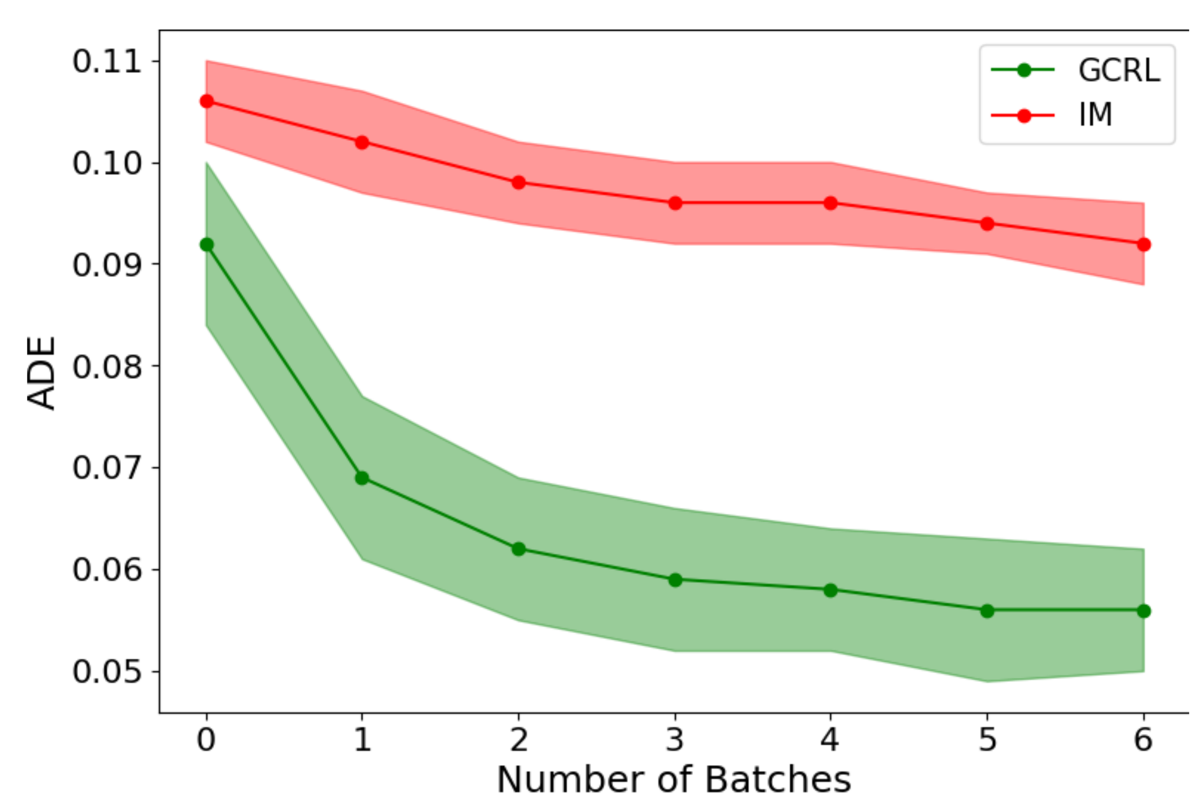
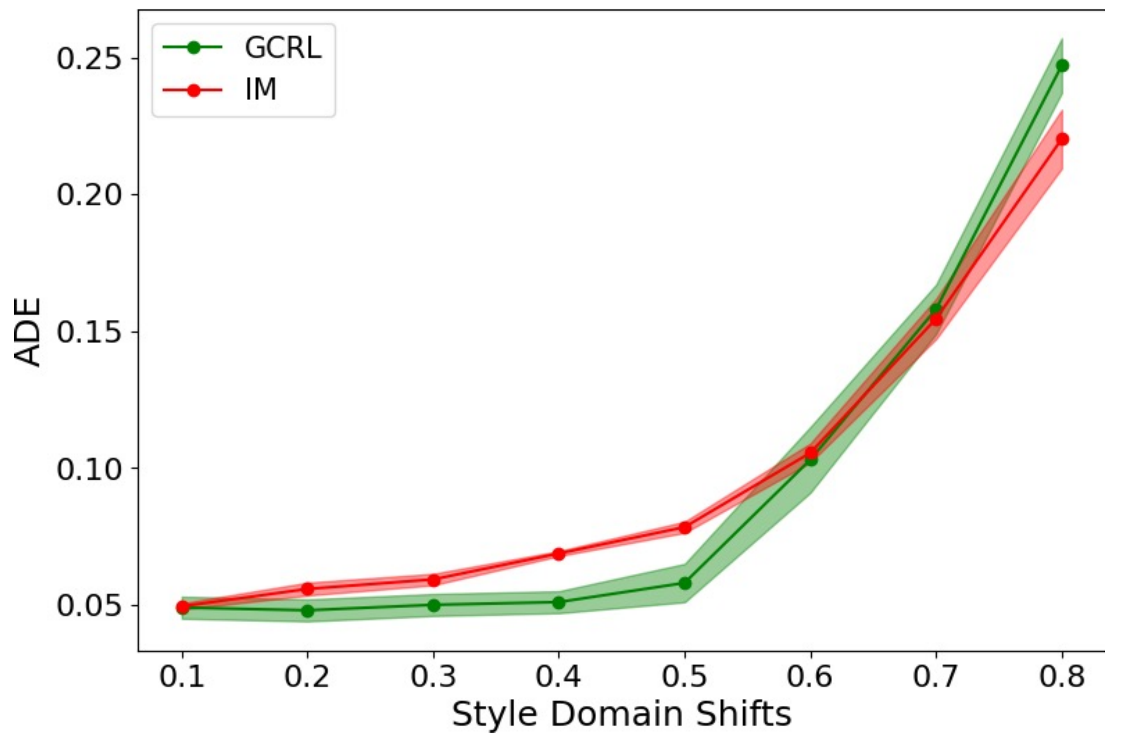


Figure below:

- Our method is more robust to domain shifts compared to the baseline model
- Achieves slightly better ADE (8.8% on average).
- For OOD-Inter cases ($Style\ Domain\ Shifts \leq 0.5$) where the test domain shift is within the range of training domain shifts ($Style\ Domain\ Shifts = \{0.1, 0.3, 0.5\}$) e.g., test domain shift=0.4, **GCRL** is reusable since ADE is insensitive to the domain shifts.
- On the other hand, for the test domain shifts out of the range of training domain shifts, the OOD-Extra cases ($Style\ Domain\ Shifts > 0.5$) , the model needs to be fine-tuned.



CONCLUSIONS

We propose a method that leverages causality to learn meaningful features that can increase the *robustness and transferability of deep learning models*. In presence of spurious correlation, we demonstrated the robustness of our method while other human trajectory prediction models performed poorly compared to our method. Furthermore, our augmented causal model was able to *enhance the transferability in a zero-shot and low-shot settings*. It can be concluded from our results that incorporating causality in deep learning is a promising research direction towards robustness and explainability of deep learning models.

REFERENCES

1. Scholkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalch- ¨ brenner, N., Goyal, A., and Bengio, Y. Toward causal representation learning. Proceedings of the IEEE, 109(5): 612–634, 2021. doi: 10.1109/JPROC.2021.3058954
2. Pearl, J. and Bareinboim, E. External Validity: From DoCalculus to Transportability Across Populations. Statistical Science, 29(4):579 – 595, 2014. doi: 10.1214/ 14-STS486. URL https://doi.org/10.1214/ 14-STS486.
3. Liu, Y., Cadei, R., Schweizer, J., Bahmani, S., and Alahi, A. Towards robust and adaptive motion forecasting: A causal representation perspective. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, pp. 17060– 17071. IEEE, 2022. doi: 10.1109/CVPR52688.2022. 01657.
4. Amirian, J., Zhang, B., Castro, F. V., Baldelomar, J. J., Hayet, J.-B., and Pettre, J. Opentraj: Assessing prediction complexity in human trajectories datasets, 2020.

SFU
SIMON FRASER UNIVERSITY

UNIVERSITY OF WATERLOO

WATERLOO.AI
WATERLOO ARTIFICIAL INTELLIGENCE INSTITUTE

CONTACT

If you have any questions/suggestions, please get in touch!

Shayan: sshirahm@uwaterloo.ca
Lab: https://markcrowley.ca/