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

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

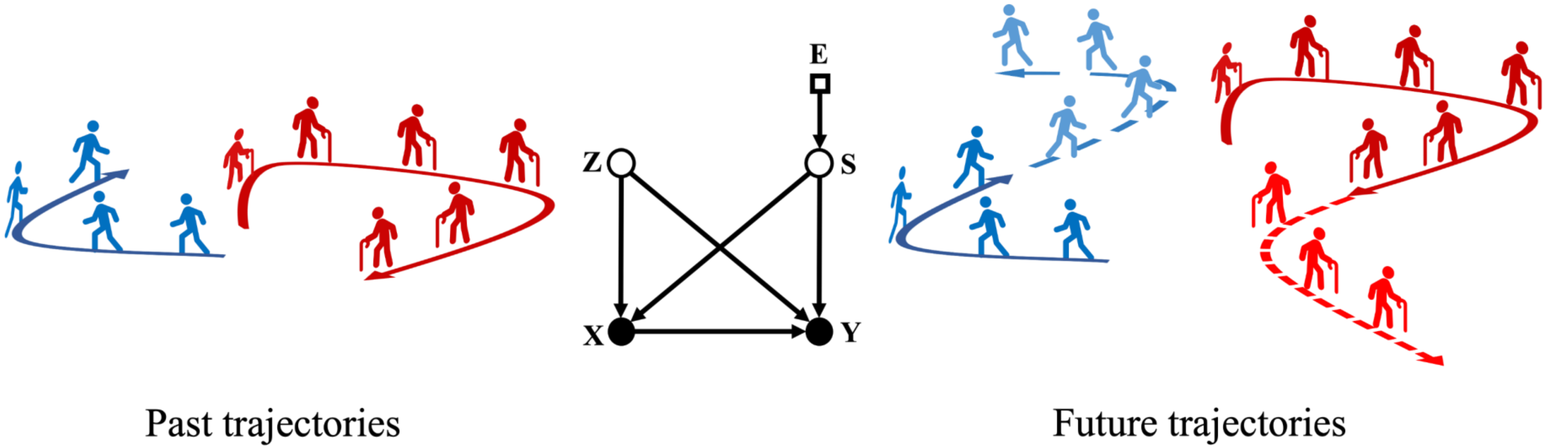
[1]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada
[2]Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada
[3]School of Computing Science, Simon Fraser University, Burnaby, Canada

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Robust Motion Forecasting

- Given past states of M agents in the scene, we predict the states of each agent in the next timesteps.
- Robustness comes into play when there are multiple domains/distributions in the data. These shifts are due to interventions in the generative/causal factors of the dataset.
- Statistical inference is only valid across environments with identical experimental conditions whereas learning about interventions trains robust models against environmental changes.

Proposed Causal Model



- There are four endogenous variables: S, Z, X, and Y which represent variant features such as motion styles, invariant features such as physical laws, past trajectories, and future trajectories, respectively.
- There is one exogenous variables E i.e., the selection variable.
- We assume the model is causally sufficient i.e., it can explain the data without adding further causal variables.
- Only X and Y i.e., past and future trajectories, are observed.

Loss Function

- The loss function of GCRL is given as:

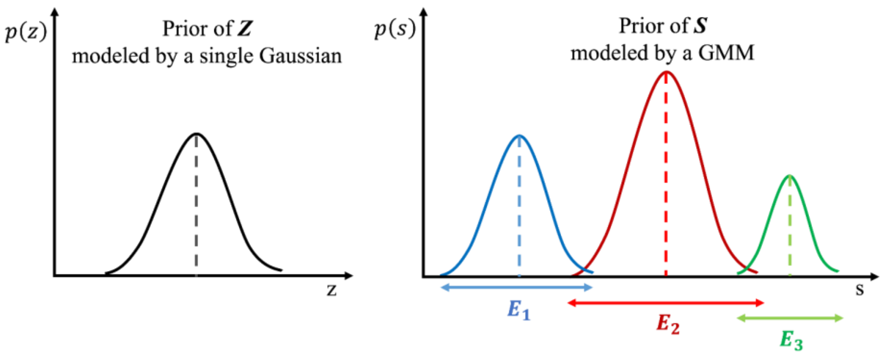
$$\max_{p,q} E_{p^*(x,y)}\left[\log q(y|x) + \frac{1}{q(y|x)} E_{q(s|x),q(z|x)}\left[p(y|x,s,z)\log\frac{p(x|s,z)p(s)p(z)}{q(s|x)q(z|x)}\right]\right]$$

- Through this loss function, GCRL learns to:

1. To minimize the distance between groundtruth 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 $\log \frac{p(s)}{q(s|x)}$
6. Since GCRL learns to predict the future trajectories with a generative approach, it can tackle the multi-modality of trajectories

Domain Adaptation

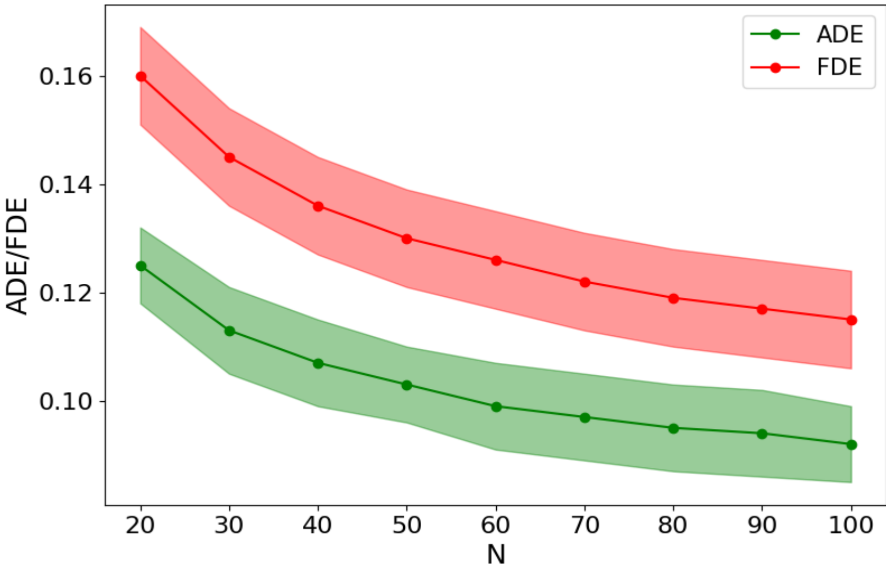
- All representations generated by q(z|x) will be in the same range, whereas the representations of q(s|x) will form clusters, each modeled by a component of the GMM.
- Since Z is invariant, we can directly transfer it to the new domain without any fine-tuning.
- However, S can be interpreted as a weighted sum of the representations learnt from different environments of the training domains, which may be used in the test domains as well.
- Depending on how related the test domains are to the training domains, we may need to fine-tune the components of the GMM and obtain a new prior for S.

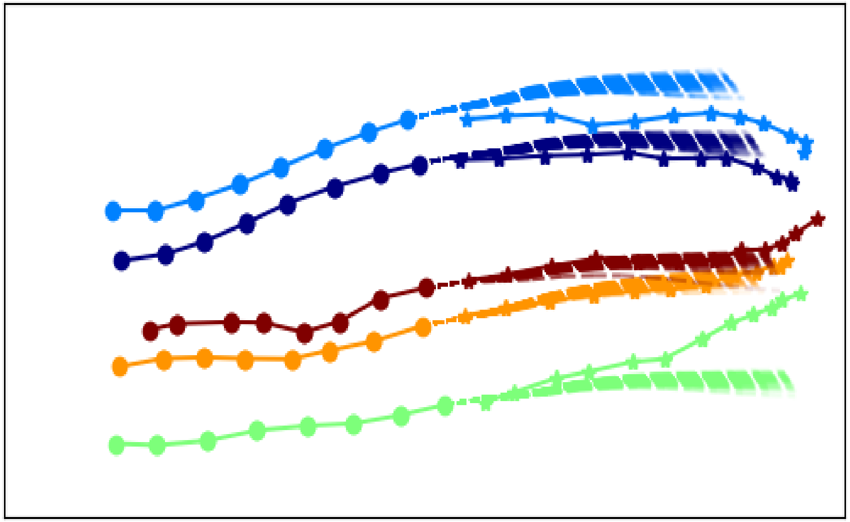


Robustness of GCRL

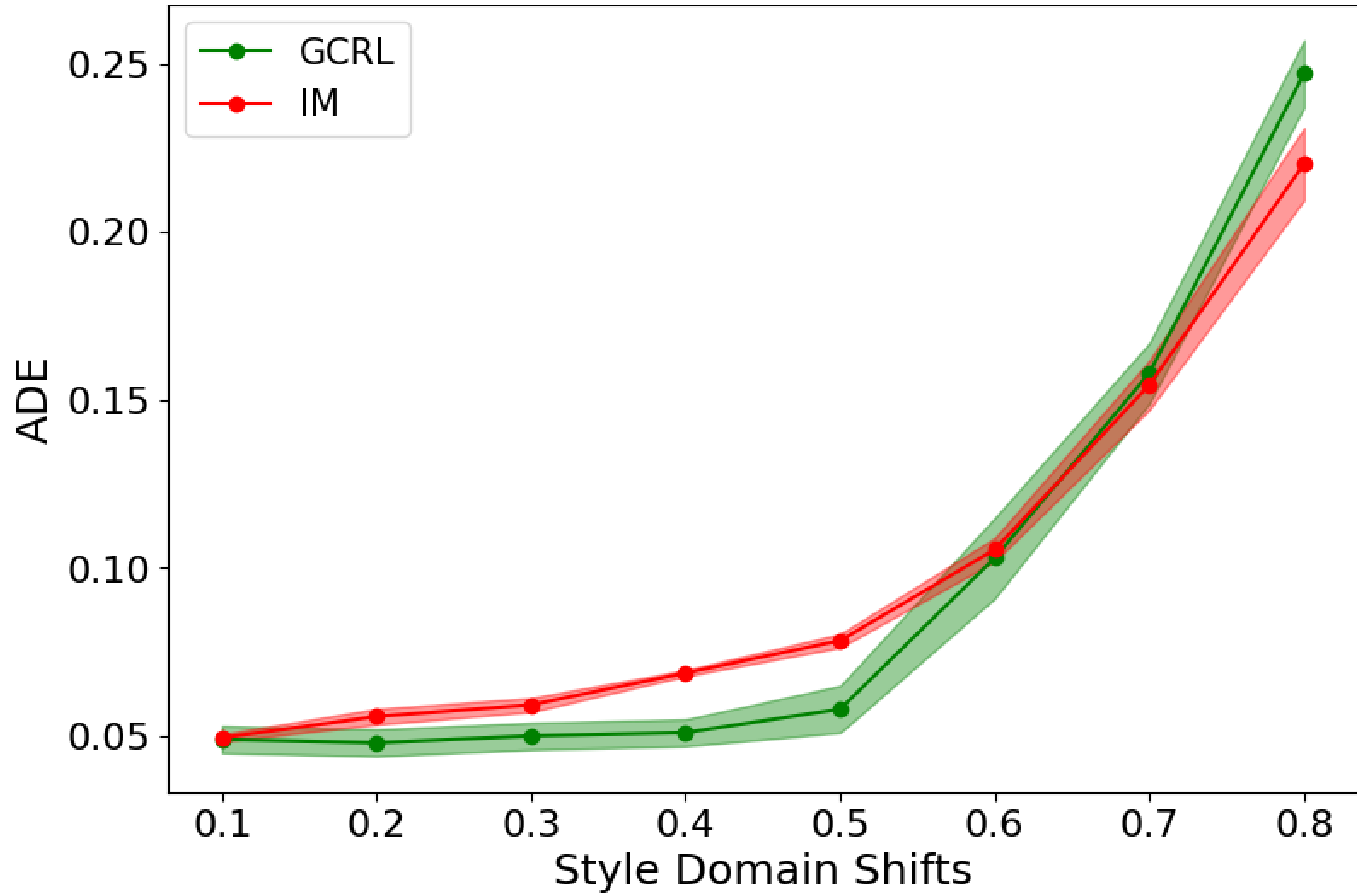
Method	ADE/FDE $\alpha = 8$	ADE/FDE $\alpha = 16$	ADE/FDE $\alpha = 32$	ADE/FDE $\alpha = 64$
Baseline (Huang et al., 2019)	**0.80/1.37**	2.15/3.80	2.64/4.44	2.68/4.48
Counterfactual (Chen et al., 2021)	0.80/1.59	1.62/2.68	2.32/3.90	2.68/4.52
Invariant $\lambda = 1.0$ (Liu et al., 2022)	0.94/1.65	1.04/1.76	1.52/2.55	1.96/3.35
Invariant $\lambda = 1.0$ (Liu et al., 2022)	0.91/1.67	0.99/1.87	1.18/2.20	1.27/2.33
Invariant $\lambda = 1.0$ (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**

Multimodal Trajectories in GCRL

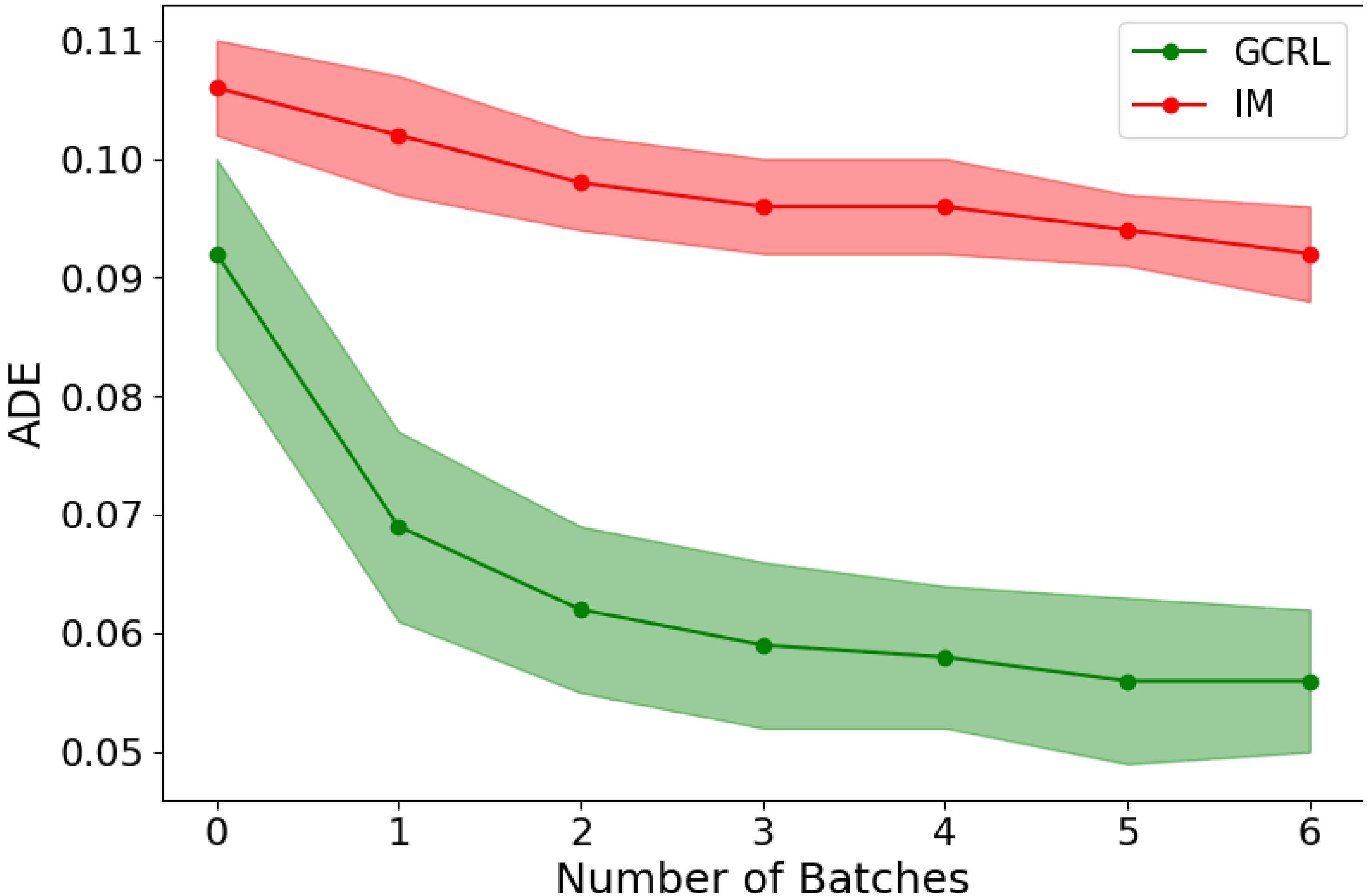




Domain Generalization



Domain Adaptation



Identifiability

Weak MCC of S	Weak MCC of Z	Strong MCC of S	Strong MCC of Z0.956
0.956	0.049	-0.16	-0.025

Method	ADE	FDE
Z Only	0.1054	0.1347
S Only	0.2188	0.2418

Contact

- You can always reach to me via my email sshirahm@uwaterloo.ca. ☻

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

References

1. Huang, Y., Bi, H., Li, Z., Mao, T., and Wang, Z. Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6271–6280, 2019. doi: 10.1109/ICCV.2019.00637.
2. Chen, G., Li, J., Lu, J., and Zhou, J. Human trajectory prediction via counterfactual analysis. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9804–9813, 2021. doi: 10.1109/ICCV48922.2021. 00968.
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.