Curb Your Attention: Causal Attention Gating for Robust Trajectory Prediction

UNIVERSITY OF ALBERTA

Ehsan Ahmadi, Ray Mercurius, Soheil Alizadeh, Kasra Rezaee, Amir Rasouli

Background & Ideas

Context

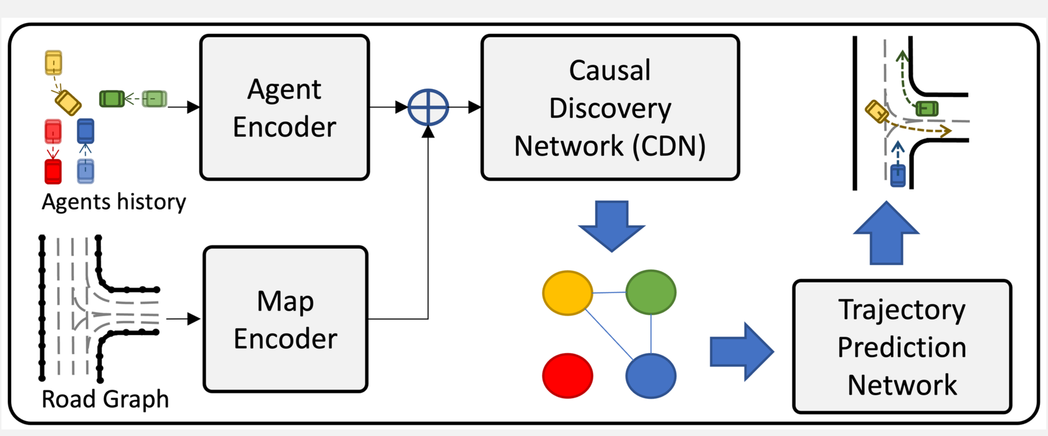
- Autonomous vehicles must robustly predict other agents' trajectories under **distribution shifts** and avoid relying on **spurious correlations**.
- State-of-the-art models suffer large performance drops if non-causal agents are removed.
- We have causal agent labels annotated by humans.

Key Challenge

- How to **identify** and **attend only** to truly causal agents?

Our Goal

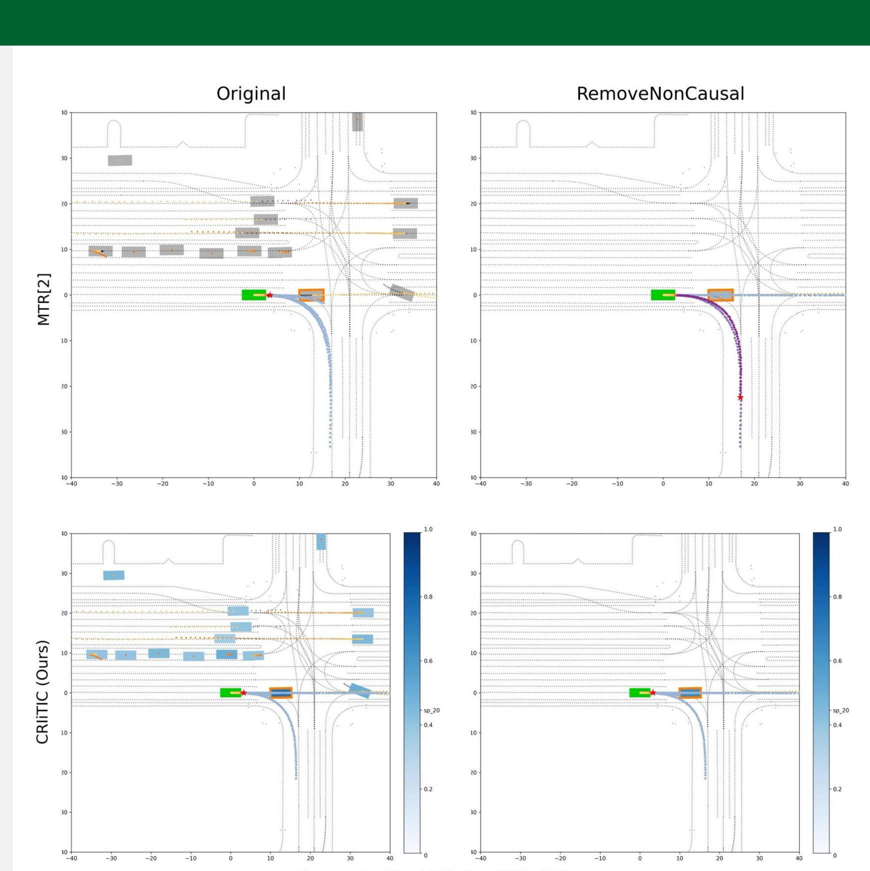
- Improve **robustness** and **generalization** by integrating explicit causal structure into prediction model's attention layers.



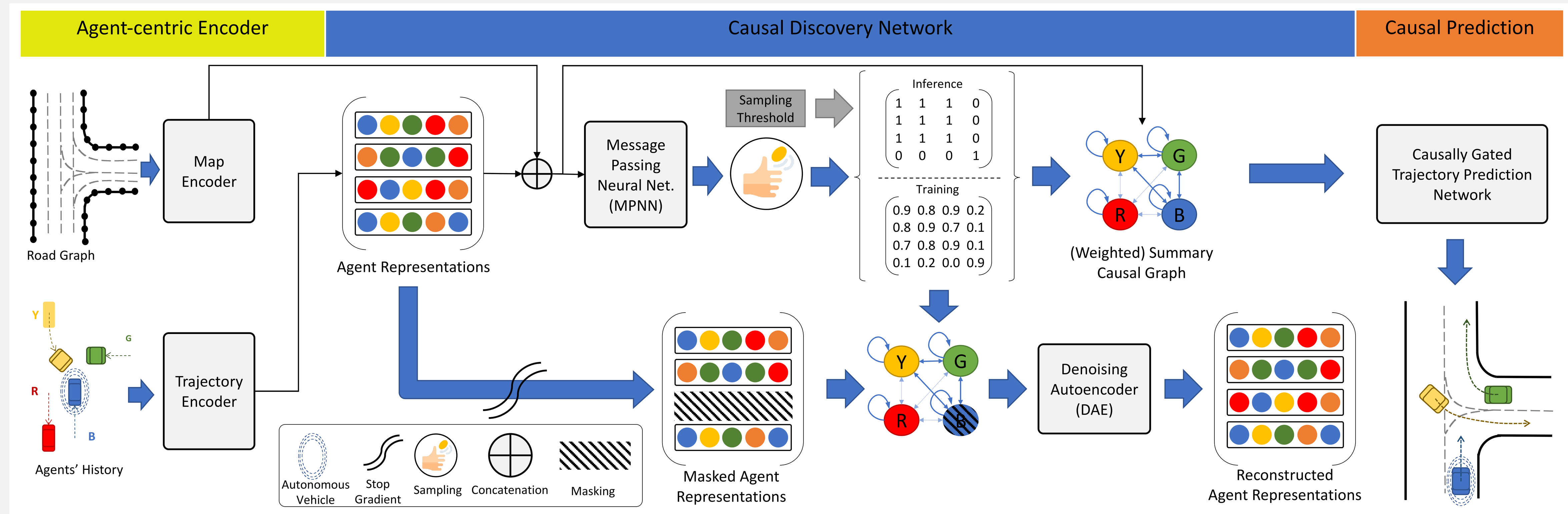
Contributions

1. **CRiTIC**: an agent-centric causal trajectory predictor.
2. **Causal Discovery Network** (CDN): learns a sparse inter-agent causal graph via an information-bottleneck regularization + self-supervised denoising autoencoder.
3. **Causal Attention Gating** (CAG): injects the learned graph into the Transformer's attention, suppressing non-causal inputs.
4. **Empirical Gains**: up to **54 %** robustness improvement under RemoveNonCausal perturbation and **29 %** cross-domain gains.

Causal Robustness Qualitative Samples



The AV is shown in green. The orange borderline indicates the causal agents. The ground truth, and predictions are shown in orange, and purple colors. Star shows the predicted most confident mode. The color saturation indicates the estimated probability of an agent being causal.



Causal Discovery Network receives the agent representations and generates a causality adjacency matrix. The matrix is used by a Transformer-based prediction backbone to shape the attention toward the causal agents.

Model Overview

1-Inputs:

- **Agent Encoder**: encodes each agent + map into disentangled embeddings.

2-Causal Discovery Network:

- **MPNN** produces soft adjacency A via a BinConcrete relaxation.
- **Sparsity** loss (KL to Bernoulli(p)) enforces an information bottleneck.
- **Auxiliary GCN** denoising autoencoder to further assist causal discovery network's training

3-Transformer-based Prediction Backbone

- **Causal Attention Gating (CAG)** to apply the discovered causal adjacency matrix in the attention layers of the backbone network

Causal Attention Gating and Information Bottleneck

We apply a learned soft adjacency **A** to the standard self-attention weights and drown non-causal links in noise.

Let $\Phi = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)$

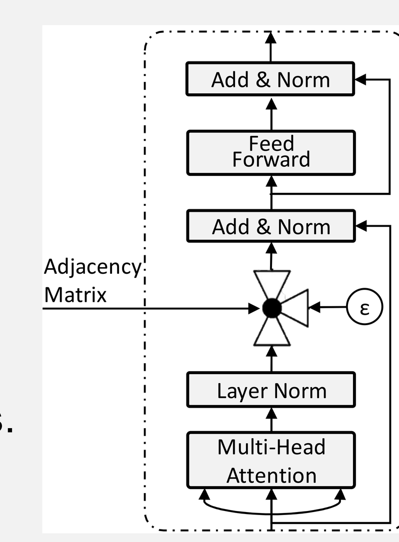
then

$$\mathrm{CausalAttn}(Q,K,V;A) = (\Phi \odot A)\,V + \alpha\,(\Phi \odot (1-A))\,\mathcal{N}(0,I)$$

where the second term is a noise term creating the attention bottleneck.
A is generate by the causal discovery network, α=0 at inference time.

A regularization loss term is used to encourage sparse causal adj. matrices.

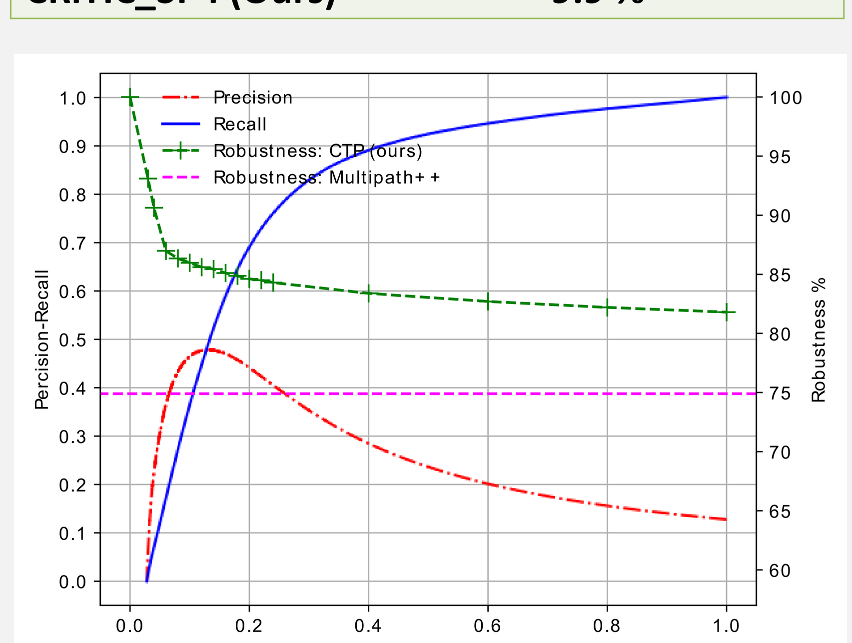
$$L_{\text{sparsity}} = \sum_{i \neq j} \mathrm{KL}(\mathrm{Bern}(e_{ij}) \,\|\, \mathrm{Bern}(p))$$



Robustness to Non-causal Agent Removal

- **Causal robustness**: The Prediction accuracy change caused by removing non-causal agents (ΔminADE / minADE (%))
- Here we show that our model at sparsity value of 4% outperforms the baselines significantly from the perspective of causal robustness.
- Sparsity is defined based on the number connections in the discovered causal relation graph from the perspective of autonomous vehicle.

Model	ΔminADE / minADE (%) ↓
MultiPath++	37.5 %
SceneTransformer	26.8 %
Wayformer	25.7 %
MTR	21.5 %
CRiTIC_SP4 (Ours)	**9.9 %**



- The relation between the sparsity of the identified causal graph, the precision and recall values for the causal agent identification, and the causal robustness measure.
- We argue that causal robustness should be reported along with the sparsity values.

Conclusion

Takeaways

- Explicit causal structure leads to dramatically improved robustness and cross-domain performance.
- Causal gating imposes an effective information bottleneck.

Next Steps

- Incorporate **agent–map** causal edges (e.g., traffic lights).
- Model unobserved **confounders**.

Contact

Corresponding author
Ehsan Ahmadi

Project Page

IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION

ATLANTA 2025

HUAWEI