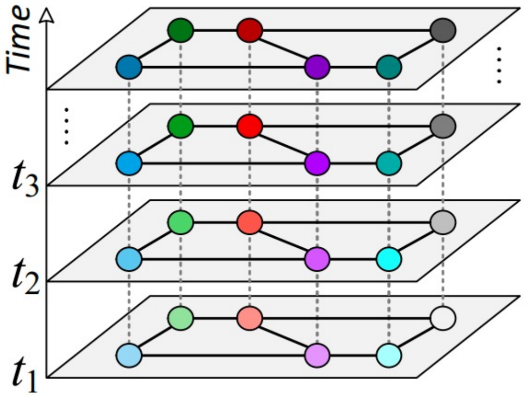
Deciphering Spatio-Temporal Graph Forecasting: A Causal Lens and Treatment

Yutong Xia, Yuxuan Liang, Haomin Wen, Xu Liu, Kun Wang, Zhengyang Zhou, Roger Zimmermann

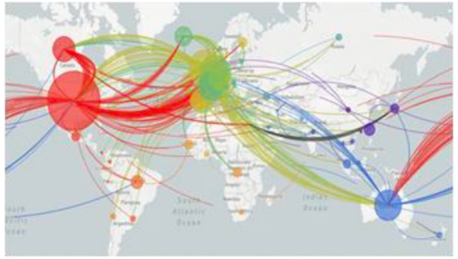
National University of Singapore

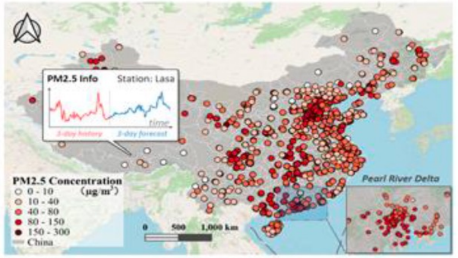
Background

- **Spatio-Temporal Graph (STG)** represents the spatial and temporal relationships between nodes or entities, which is widely used in various fields (e.g., transportation, environment and epidemiology)



Transportation


Epidemiology



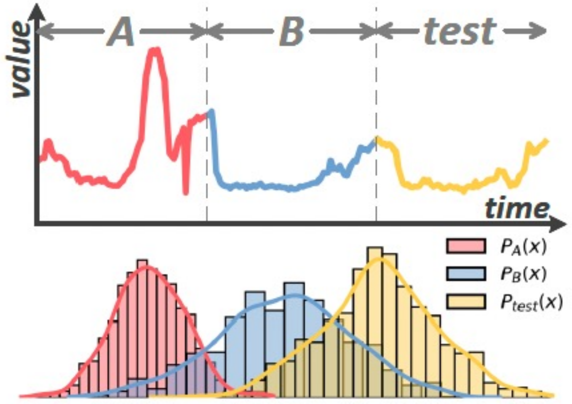
Environment

- **STG forecasting** has become crucial in the context of smart cities (e.g. informed decision-making, sustainable environments)

Challenges

- Temporal Distribution Shift

$$P_A(x) \neq P_B(x) \neq P_{test}(x)$$

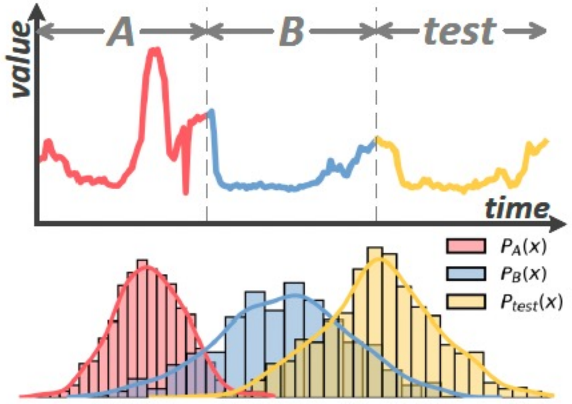


(a) Temporal Distribution Shift

Challenges

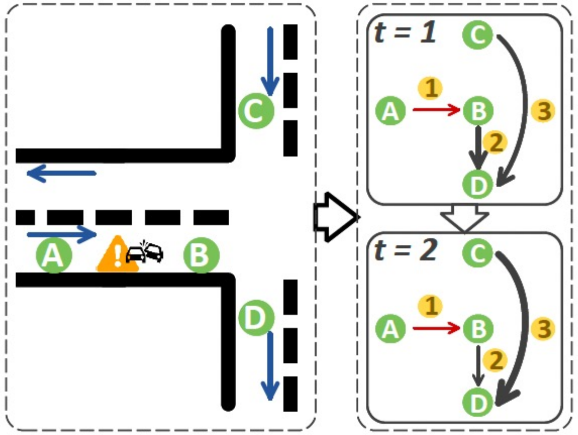
- Temporal Distribution Shift

$$P_A(x) \neq P_B(x) \neq P_{test}(x)$$



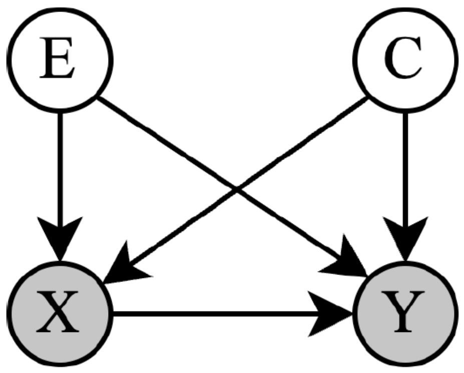
(a) Temporal Distribution Shift

- Dynamic Spatial Causation
  - Existing work:
    - Distance-based adjacency matrices
    - Attention mechanism
  - However, the ripple effect of causations



(b) Ripple Effect of Causations

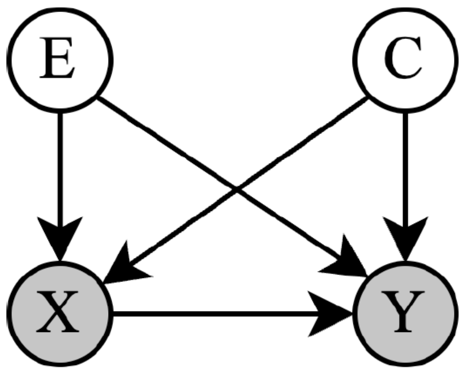
A Causal Lens



(a)

E: temporal environment
C: spatial context
X: historical signal
Y: future signal

A Causal Lens

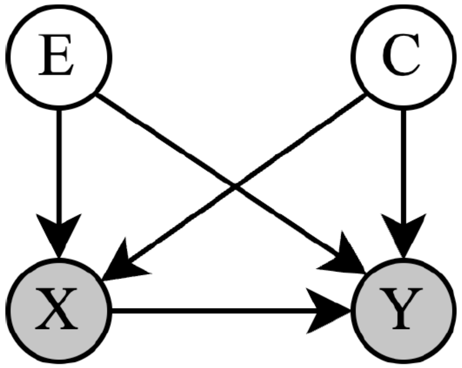


(a)

E: temporal environment
C: spatial context
X: historical signal
Y: future signal

- **X ← E → Y**
- **X ← C → Y**
- **X → Y**

A Causal Lens

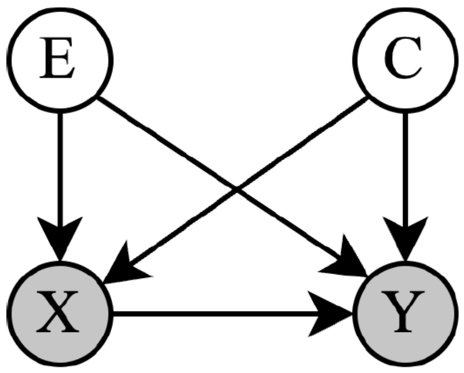


(a)

E: temporal environment
C: spatial context
X: historical signal
Y: future signal

- **X ← E → Y** The temporal OoD can arise due to changes in external variables over time. (e.g., weather can affect traffic flow observations)
- **X ← C → Y**
- **X → Y**

A Causal Lens

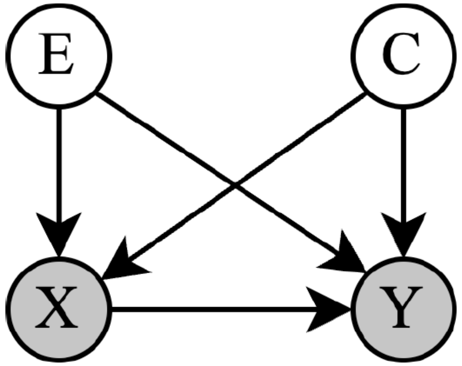


(a)

E: temporal environment
C: spatial context
X: historical signal
Y: future signal

- **X ← E → Y** The temporal OoD can arise due to changes in external variables over time. (e.g., weather can affect traffic flow observations)
- **X ← C → Y** X and Y are intrinsically affected by the surrounding spatial context, comprising both spurious and genuine causal components.
- **X → Y**

A Causal Lens

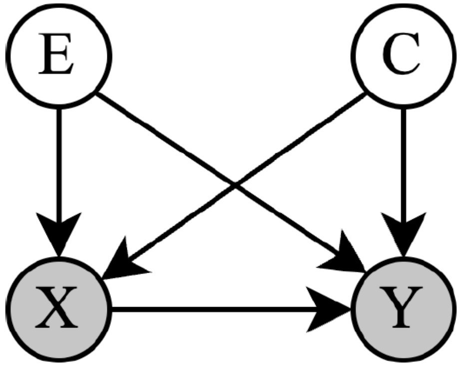


(a)

E: temporal environment
C: spatial context
X: historical signal
Y: future signal

- **X ← E → Y** The temporal OoD can arise due to changes in external variables over time. (e.g., weather can affect traffic flow observations)
- **X ← C → Y** X and Y are intrinsically affected by the surrounding spatial context, comprising both spurious and genuine causal components.
- **X → Y** Our primary goal.

A Causal Lens



(a)

E: temporal environment
C: spatial context
X: historical signal
Y: future signal

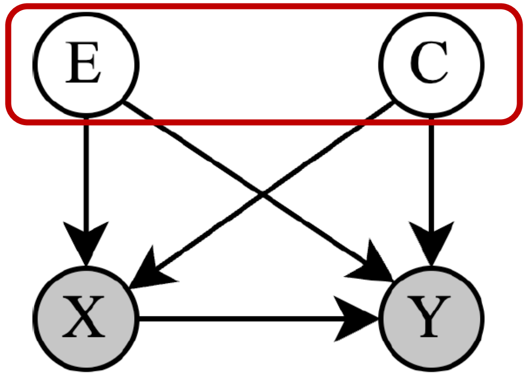
Backdoor paths

- **X ← E → Y** The temporal OoD can arise due to changes in external variables over time. (e.g., weather can affect traffic flow observations)
- **X ← C → Y** X and Y are intrinsically affected by the surrounding spatial context, comprising both spurious and genuine causal components.
- **X → Y** Our primary goal.

A Causal Lens

NUS National University of Singapore

Confounding factors



(a)

E: temporal environment
C: spatial context
X: historical signal
Y: future signal

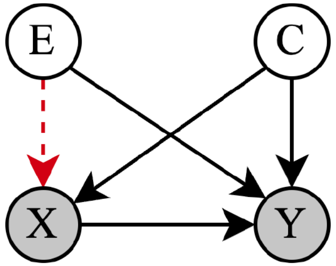
Backdoor paths

- **X ← E → Y** The temporal OoD can arise due to changes in external variables over time. (e.g., weather can affect traffic flow observations)
- **X ← C → Y** X and Y are intrinsically affected by the surrounding spatial context, comprising both spurious and genuine causal components.
- **X → Y** Our primary goal.

Causal Treatments

- Back-door adjustment for E

$$
\begin{aligned}
P(Y|do(X)) &= \sum\nolimits_e P(Y|do(X), E=e)P(E=e|do(X)) \\
&= \sum\nolimits_e P(Y|do(X), E=e)P(E=e) \\
&= \sum\nolimits_e P(Y|X, E=e)P(E=e)
\end{aligned}
$$

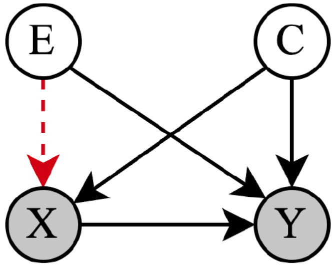


(b)

Causal Treatments

- Back-door adjustment for E

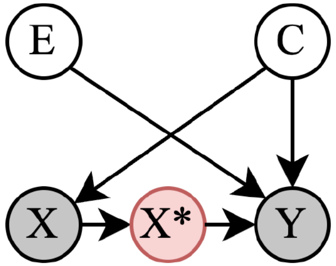
$$\begin{aligned} P(Y|do(X)) &= \sum\nolimits_e P(Y|do(X), E=e)P(E=e|do(X)) \\ &= \sum\nolimits_e P(Y|do(X), E=e)P(E=e) \\ &= \sum\nolimits_e P(Y|X, E=e)P(E=e) \end{aligned}$$



(b)

- Front-door adjustment for C

$$\begin{aligned} P(Y|do(X)) &= \sum\nolimits_{x^*} P(Y|do(X^*=x^*))P(X^*=x^*|do(X)) \\ &= \sum\nolimits_{x^*}\sum\nolimits_{x'} P(Y|X^*=x^*, X=x')P(X=x')P(X^*=x^*|do(X)) \\ &= \sum\nolimits_{x^*}\sum\nolimits_{x'} P(X^*=x^*|X)P(Y|X^*=x^*, X=x')P(X=x') \end{aligned}$$



(c)

Model Instantiations

- Causal Spatio-Temporal neural network (CaST)

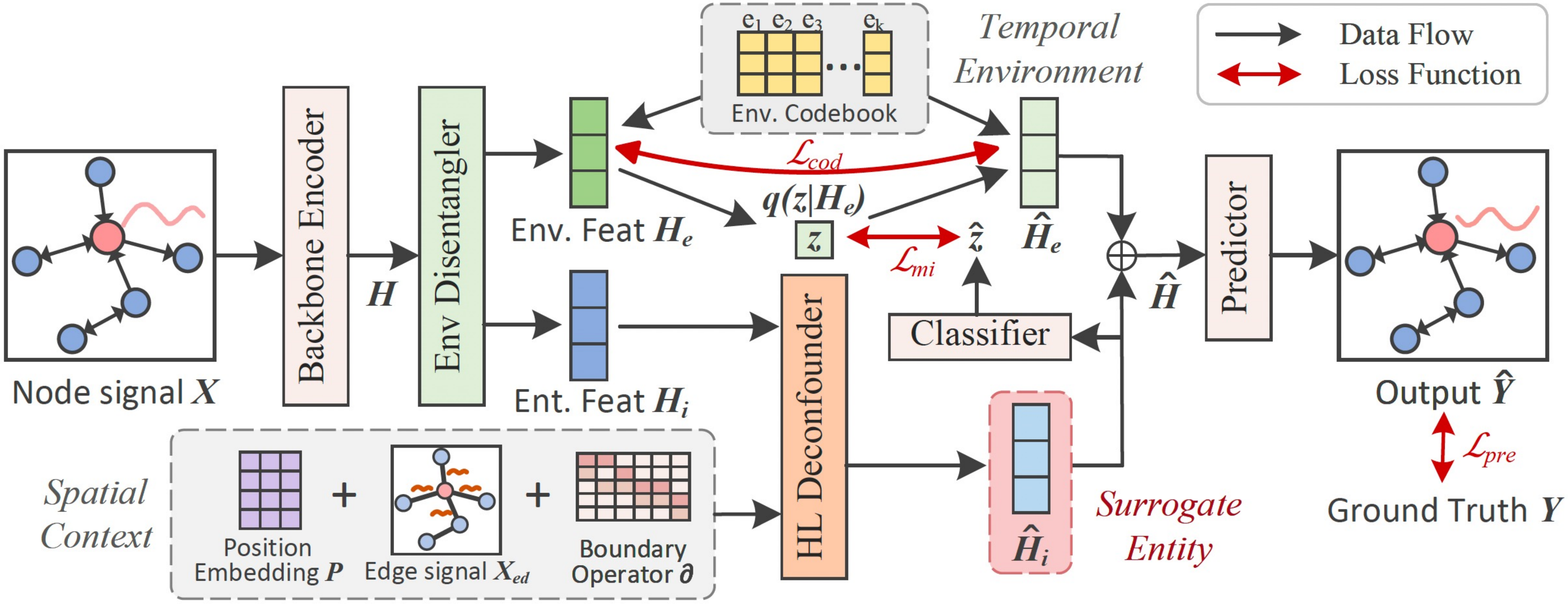


Figure 3: The pipeline of CaST. Env: Environment. Ent: Entity. Feat: Feature.

Model Instantiations

- Causal Spatio-Temporal neural network (CaST)

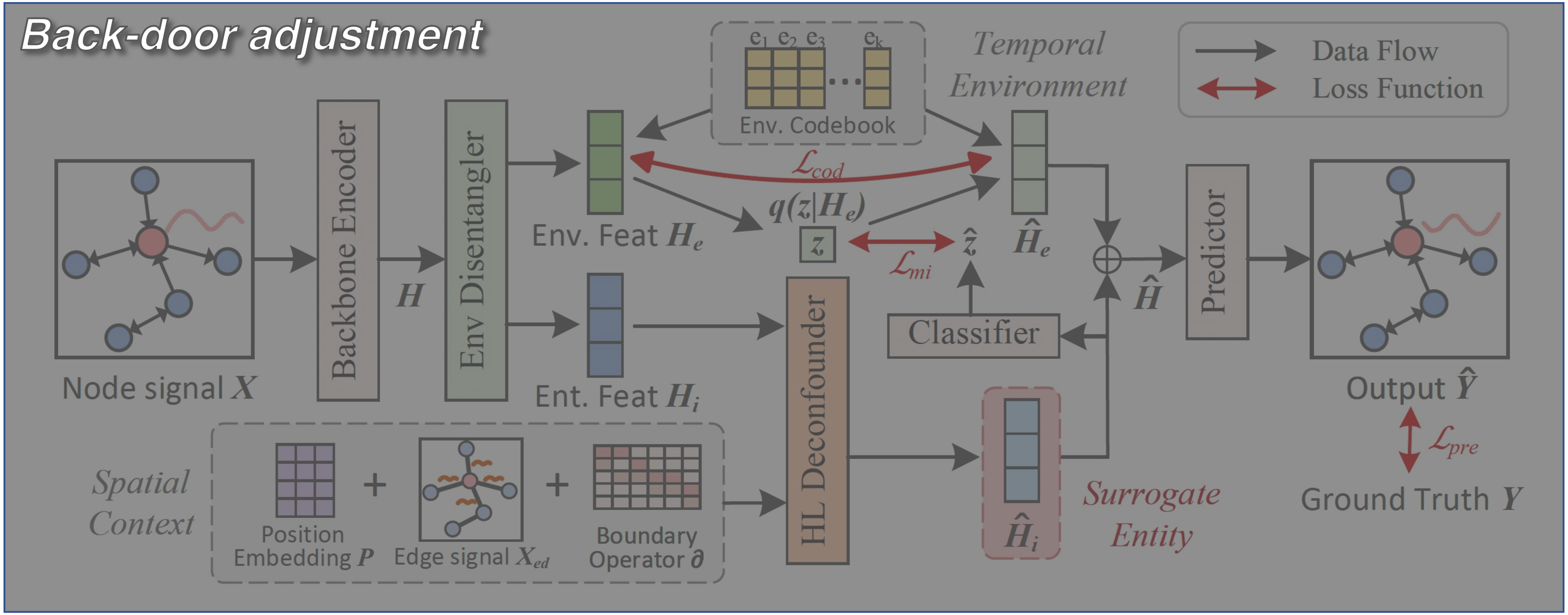


Figure 3: The pipeline of CaST. Env: Environment. Ent: Entity. Feat: Feature.

Model Instantiations

- Causal Spatio-Temporal neural network (CaST)

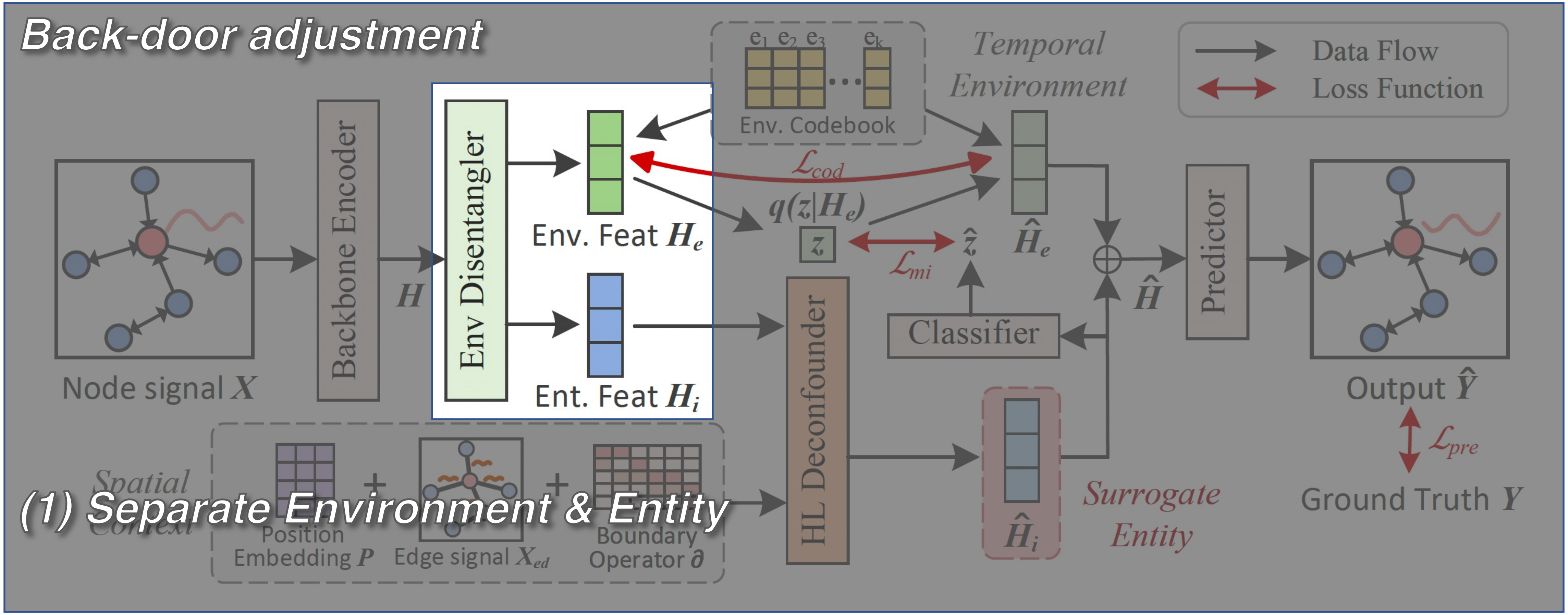


Figure 3: The pipeline of CaST. Env: Environment. Ent: Entity. Feat: Feature.

Model Instantiations

- Causal Spatio-Temporal neural network (CaST)

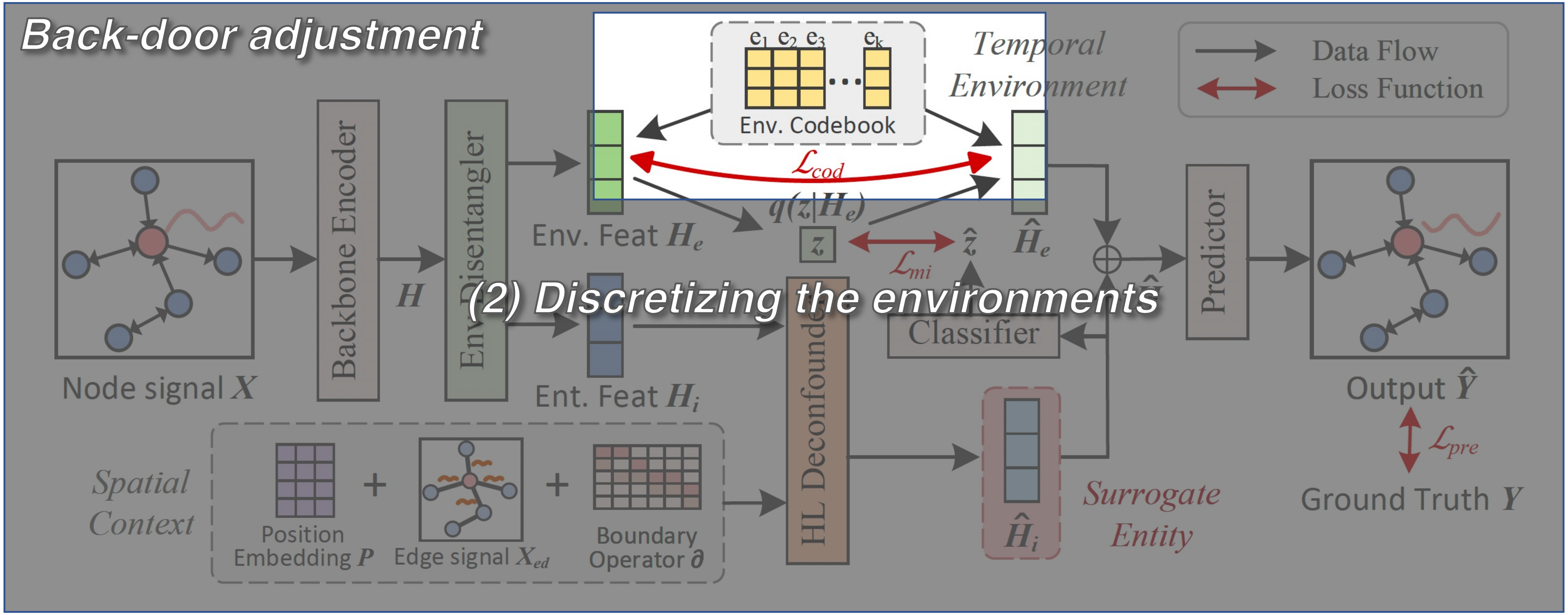


Figure 3: The pipeline of CaST. Env: Environment. Ent: Entity. Feat: Feature.

Model Instantiations

- Causal Spatio-Temporal neural network (CaST)

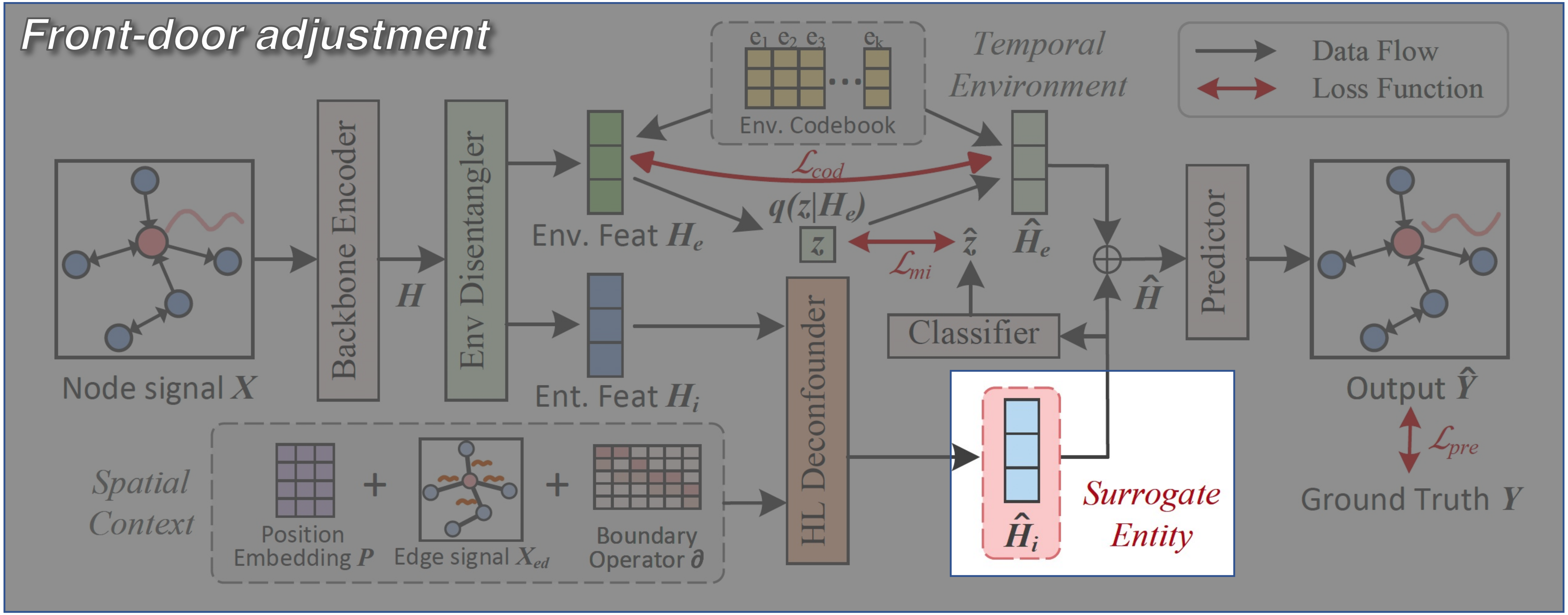


Figure 3: The pipeline of CaST. Env: Environment. Ent: Entity. Feat: Feature.

Model Instantiations

- Causal Spatio-Temporal neural network (CaST)

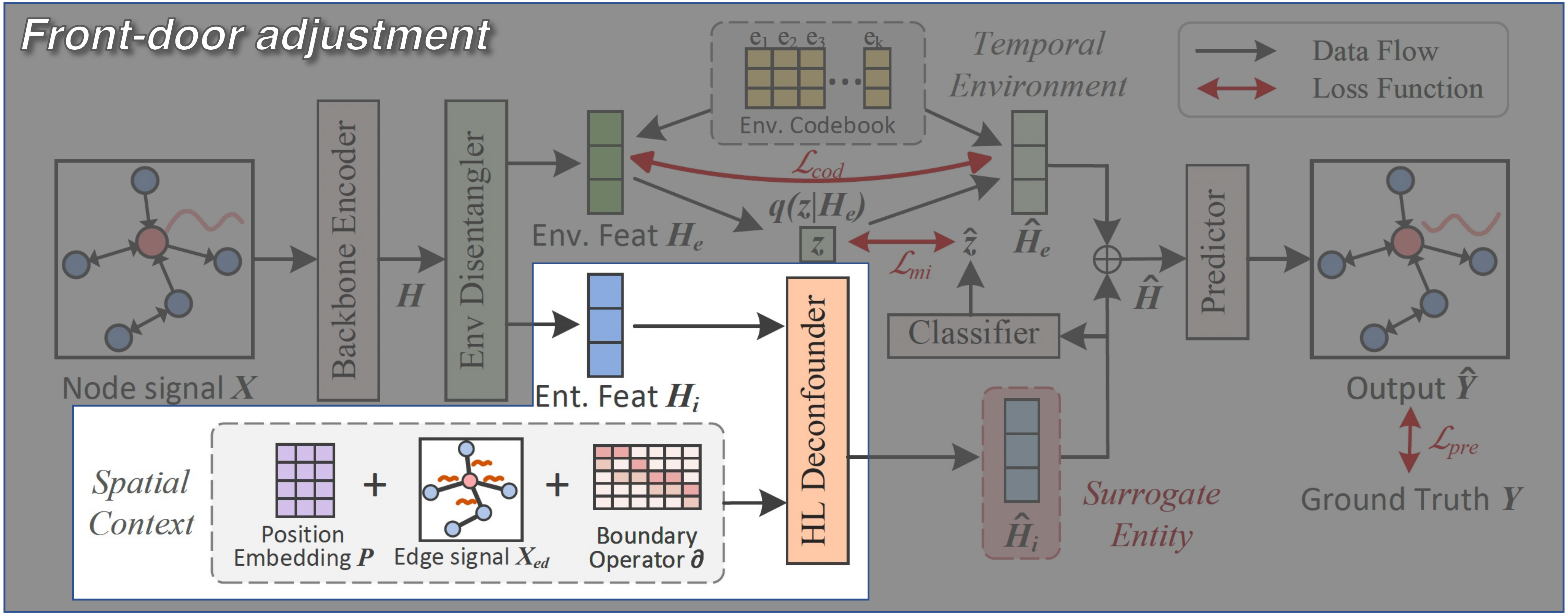


Figure 3: The pipeline of CaST. Env: Environment. Ent: Entity. Feat: Feature.

Model Instantiations

- Causal Spatio-Temporal neural network (CaST)

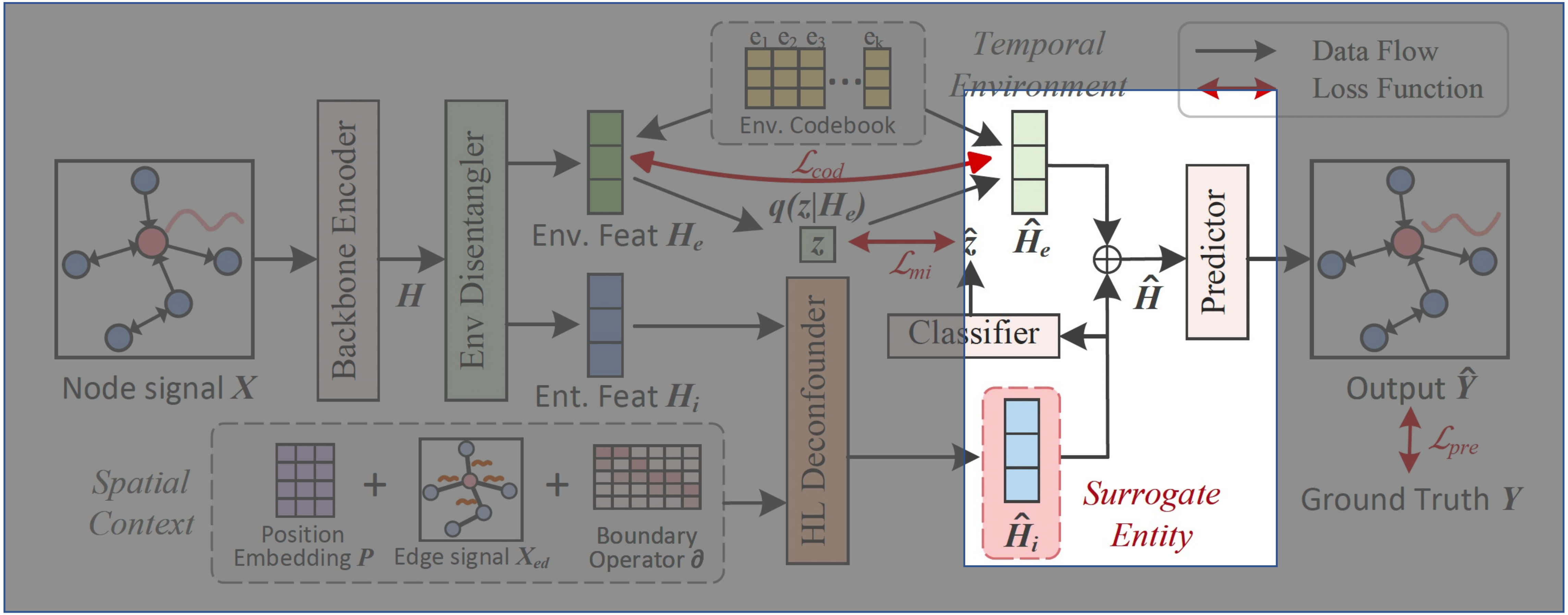


Figure 3: The pipeline of CaST. Env: Environment. Ent: Entity. Feat: Feature.

Model Instantiations

- Causal Spatio-Temporal neural network (CaST)

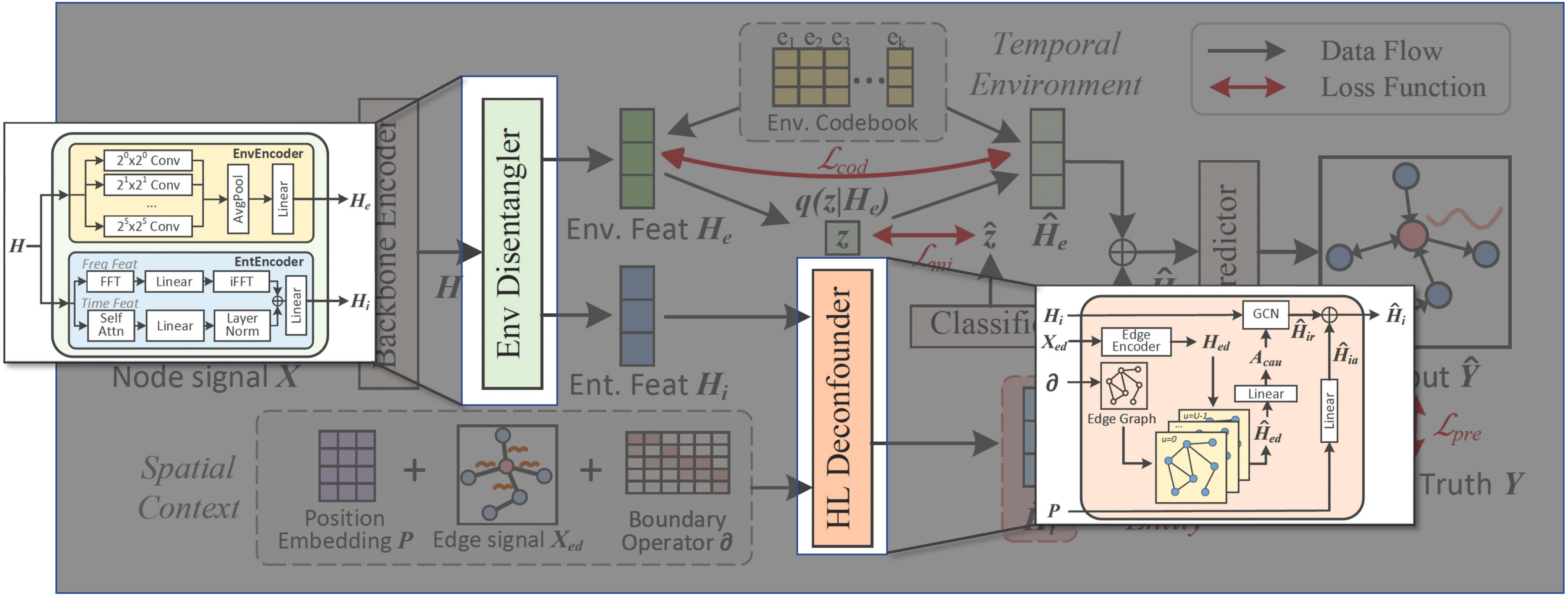


Figure 3: The pipeline of CaST. Env: Environment. Ent: Entity. Feat: Feature.

Experiments

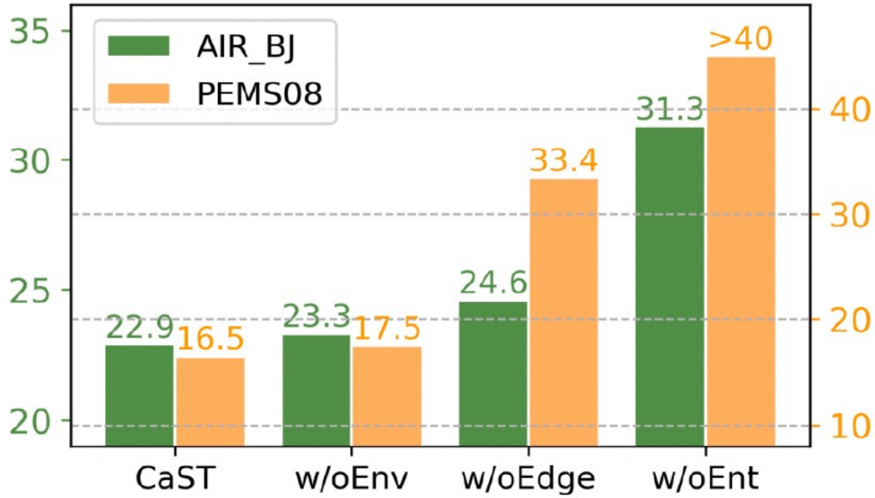
- Datasets: PEMS08, AIR-BJ, AIR-GZ
- Experiment settings: predict over the next 24 steps given the past 24 steps
- Evaluation metrics: MAE, RMSE

Table 1: 5-run error comparison. The bold/underlined font means the best/the second-best result.

Model	**PEMS08** (24→24)		**AIR-BJ** (24→24)		**AIR-GZ** (24→24)	
	MAE	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
HA(2017)	58.83	81.96	32.12	43.95	19.56	25.77
VAR(1991)	37.04	53.08	29.79	42.04	14.97	20.61
DCRNN(2017)	22.10 ± 0.45	33.96 ± 0.59	23.72 ± 0.36	35.84 ± 0.56	12.99 ± 0.26	18.27 ± 0.41
STGCN(2018)	18.60 ± 0.08	28.44 ± 0.15	23.71 ± 0.21	36.30 ± 0.58	12.69 ± 0.04	17.66 ± 0.09
ASTGCN(2019)	20.36 ± 0.48	30.87 ± 0.55	23.78 ± 0.22	35.91 ± 0.11	12.91 ± 0.15	18.02 ± 0.27
MTGNN(2020)	18.13 ± 0.10	28.85 ± 0.12	24.35 ± 0.74	38.97 ± 1.81	12.43 ± 0.11	17.99 ± 0.18
AGCRN(2020)	17.06 ± 0.14	26.80 ± 0.15	23.43 ± 0.29	35.66 ± 0.57	12.74 ± 0.01	17.49 ± 0.01
GMSDR(2022)	18.34 ± 0.68	28.36 ± 1.01	25.92 ± 0.52	39.60 ± 0.44	13.47 ± 0.31	19.04 ± 0.46
STGNCDE(2022)	17.55 ± 0.30	27.28 ± 0.36	24.35 ± 0.31	35.91 ± 0.48	13.70 ± 0.10	19.15 ± 0.07
CaST (ours)	**16.44** ± 0.10	**26.61** ± 0.15	**22.90** ± 0.09	**34.84** ± 0.11	**12.36** ± 0.01	**17.25** ± 0.05

Ablation Study & Interpretation Analysis

- Effects of Core Components
  - w/o Env: excludes environment features for prediction.
  - w/o Ent: omits entity features for prediction.
  - w/o Edge: not utilize the causal score to guide the spatial message passing



Ablation Study & Interpretation Analysis

NUS National University of Singapore

- Effects of Edge Convolution
    - CaST-ADP: using a self-adaptive adjacency matrix
    - CaST-GAT: using the graph attention mechanism for causal scoring

Table 2: Variant results on MAE over AIR-BJ. s: steps.

Variant	**Overall**	**1-8s**	**9-16s**	**17-24s**
CaST-ADP	24.28	16.42	26.06	30.36
CaST-GAT	23.77	14.76	25.75	30.80
CaST	**22.90**	**13.79**	**24.86**	**30.05**

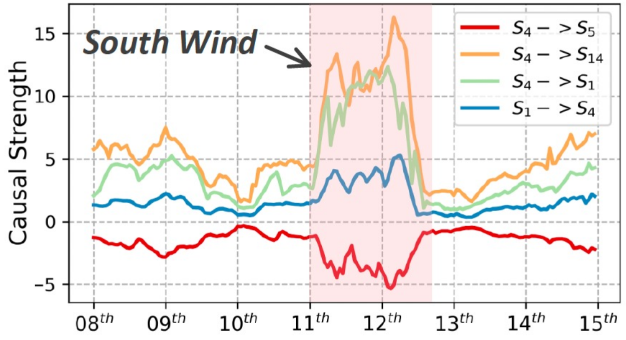
Ablation Study & Interpretation Analysis

- Effects of Edge Convolution
  - CaST-ADP: using a self-adaptive adjacency matrix
  - CaST-GAT: using the graph attention mechanism for causal scoring

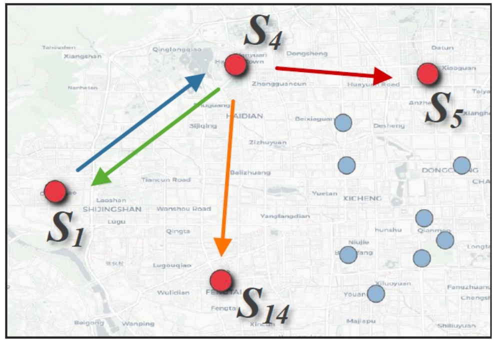
Table 2: Variant results on MAE over AIR-BJ. s: steps.

Variant	Overall	1-8s	9-16s	17-24s
CaST-ADP	24.28	16.42	26.06	30.36
CaST-GAT	23.77	14.76	25.75	30.80
CaST	**22.90**	**13.79**	**24.86**	**30.05**

- Visualization of Dynamic Spatial Causation



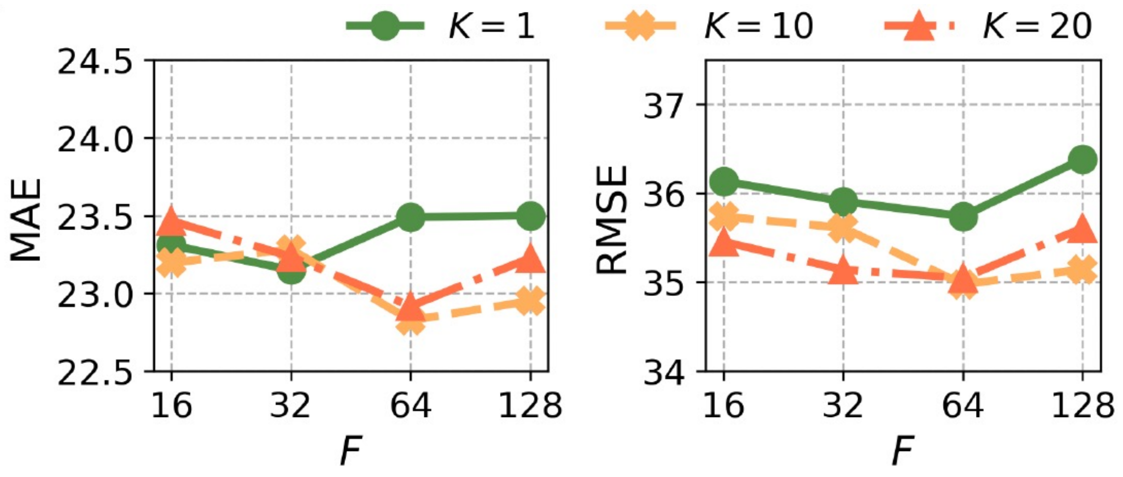
(b) Dynamic Causal Relations



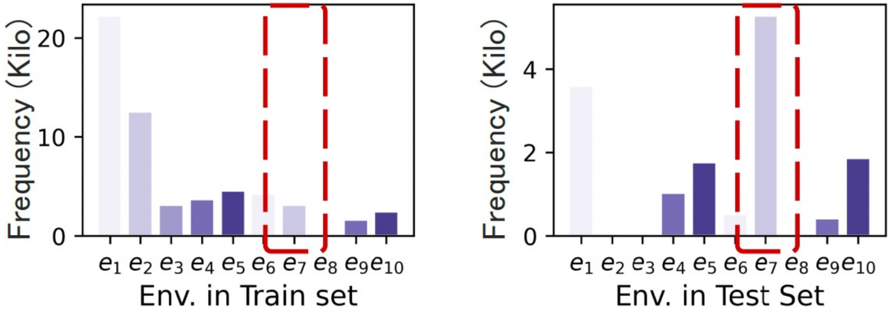
(c) Geographical Distribution

Ablation Study & Interpretation Analysis

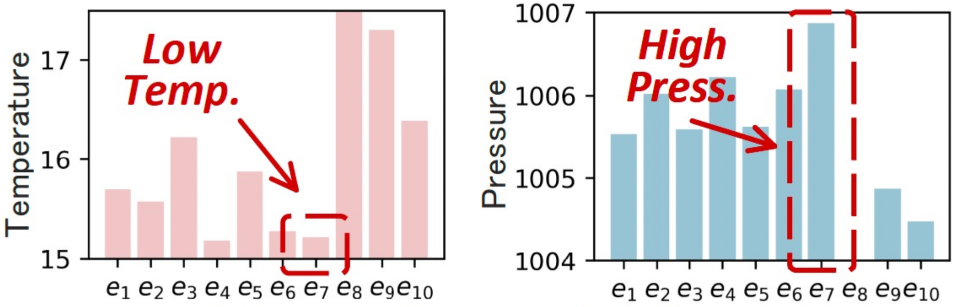
- Analysis on Environmental Codebook



- Interpretation of Temporal Environments



(b) Frequency of Environments



(c) External Factors of Environments

Conclusion

NUS National University of Singapore

- Took a causal look at the STG forecasting problem
- Utilized back-door and front-door adjustments for resolving challenges
- Introduced a novel Causal Spatio-Temporal neural network (CaST)
- Verified effectiveness, generalizability, and interpretability through extensive experiments on three datasets

Thanks!

Xia et al., Graph Forecasting: A Causal Lens and Treatment. NeurIPS, 2023.