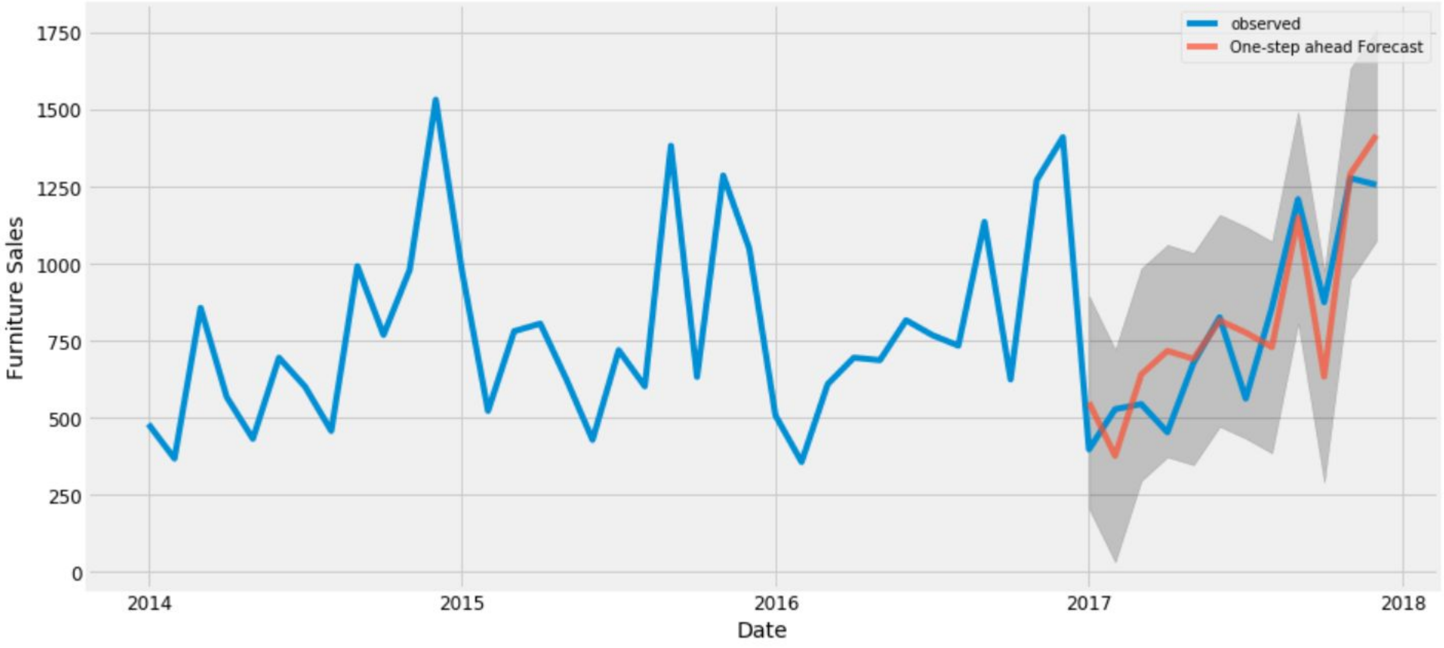
CoST: Contrastive Learning of Disentangled Seasonal-Trend Representations for Time Series Forecasting

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, Steven Hoi

Problem Setting

- **Problem Formulation**:
  - Given the look-back window $(X_1, \ldots, X_h) \in R^{h \times m}$ with length h, the goal is to forecast the values of the next k steps $(X_{h+1}, \ldots, X_{h+k}) \in R^{k \times m}$



Introduction

Deep Time Series Forecasting - State of the Art

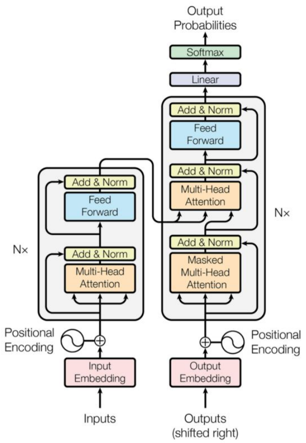
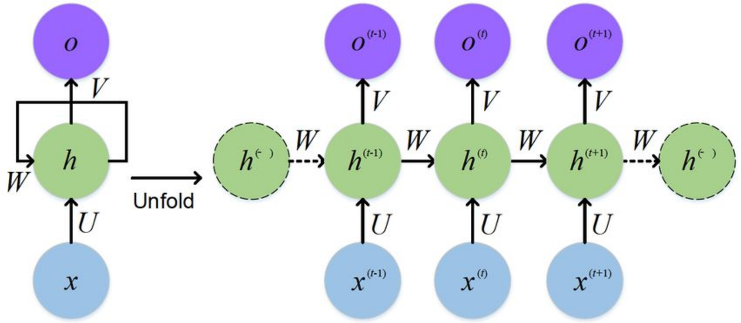
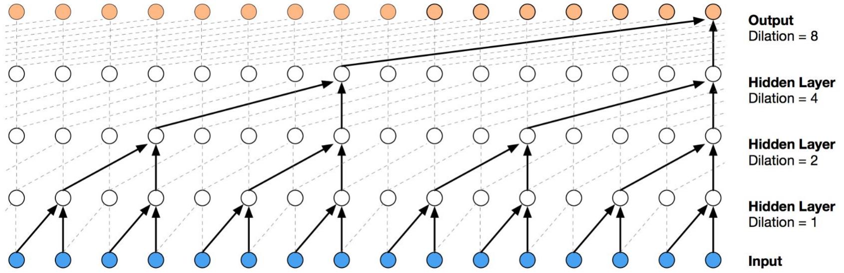


Figure 1: The Transformer - model architecture.





Transformer

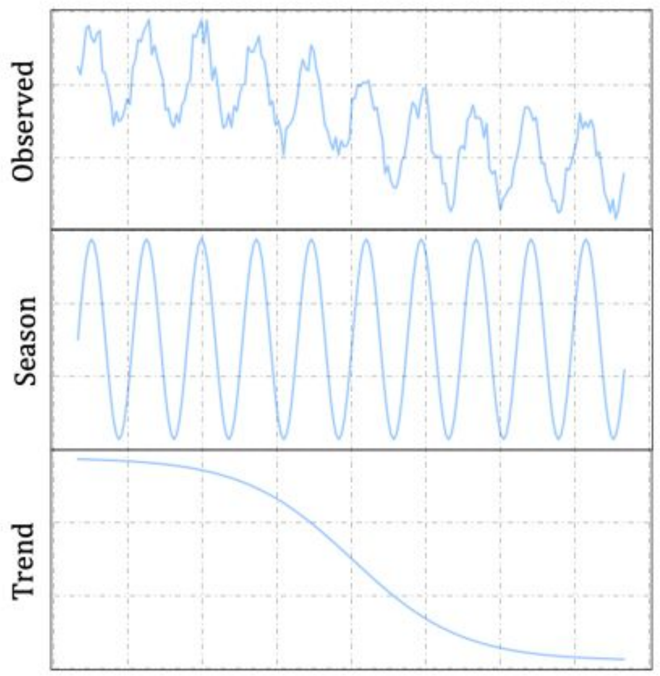
RNN

TCN

- Recent efforts focus on designing advanced deep architectures
- Stacking a series of non-linear layers + a regression layers
- Jointly learning these layers end-to-end

Joint Learning for Time Series Forecasting

- Problems associated with joint learning:
  - Can lead to **overfitting**
  - Captures **spurious correlations** of the unpredictable noise contained in the observed data
  - **Entangled** representations



What is the problem with **entangled representations?**

Distribution shift over seasonal module

- Time series - local independent modules of season and trend, distribution shifts occurs to local components.
- If we learn an **entangled** representation, it is challenging to handle distribution shift.

Goal

Can we **directly** learn **disentangled representations** for time series forecasting task?

Causal Interpretation for Disentangled Representation Learning for Time Series

- Introduce structural priors inspired by seasonal-trend decomposition
  - X is generated by error E and error-free latent variable X^*
  - X^* is generated by season S and trend T

- Merits:
  - Learn the **error-free latent variable X^*** to avoid capturing spurious correlation
  - Learn the **seasonal and trend two independent module,** enhance transferability in non-stationary environment

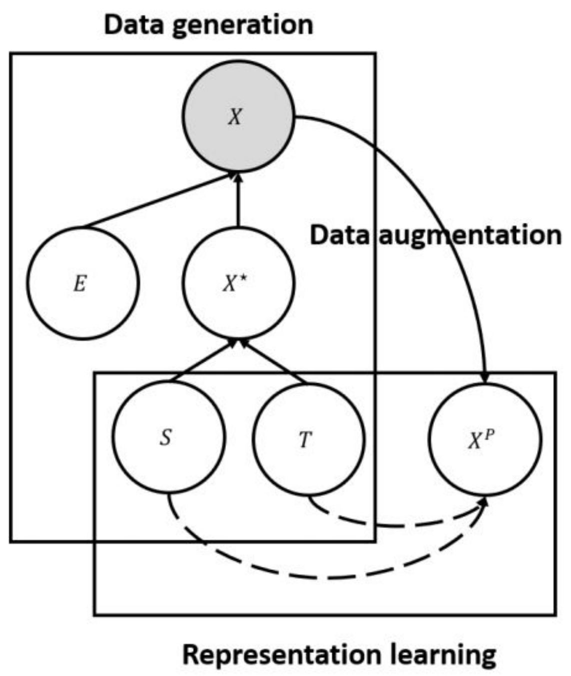


Figure 2: A causal graph demonstrating the generative process of time series data.

Disentangled Representation Learning for Time Series

- **Key idea**: Construct a proxy contrastive learning objective to learn X^*
  - Error E does not influence $P(X^*|T, S)$

    $$P^{do(E=e_i)}(X^\star|T,S) = P^{do(E=e_j)}(X^\star|T,S)$$

  - Use data augmentations as interventions on the error E and learn invariant representations of T and S via contrastive learning

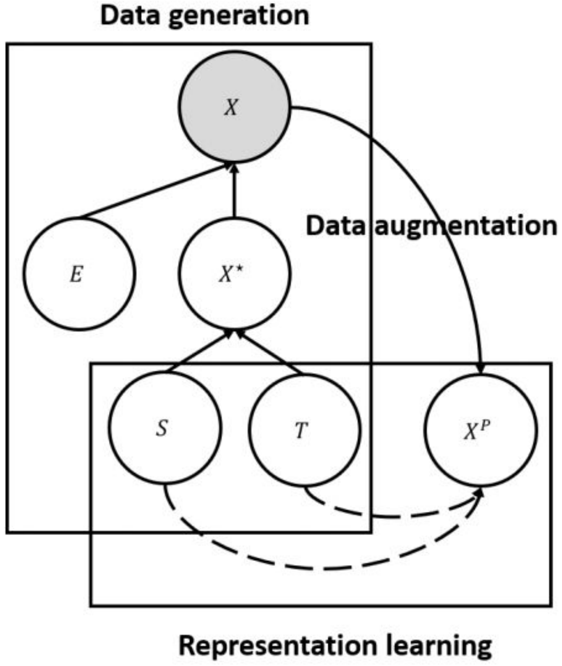


Figure 2: A causal graph demonstrating the generative process of time series data.

Seasonal-Trend Contrastive Learning Framework

- Goal: given look-back window h, learn the disentangled seasonal-trend representations for each time steps

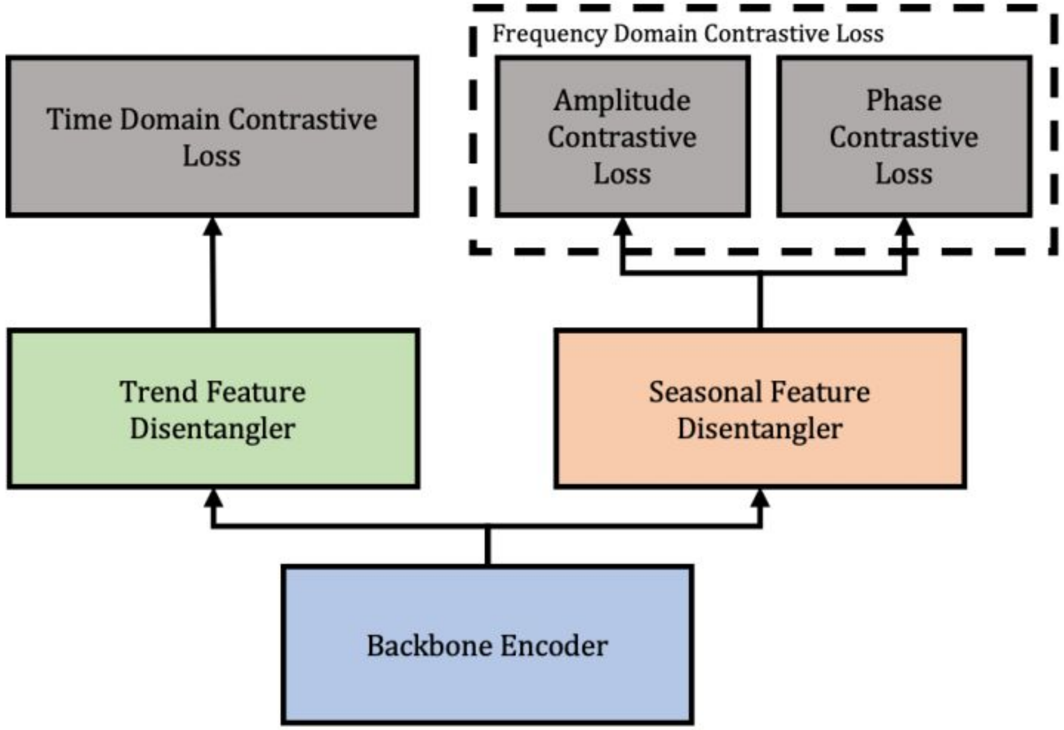
Representations:

$$\boldsymbol{V} = [\boldsymbol{V}^{(T)}; \boldsymbol{V}^{(S)}] \in \mathbb{R}^{h \times d}$$

$\boldsymbol{V}^{(T)} \in \mathbb{R}^{h \times d_T}$ and $\boldsymbol{V}^{(S)} \in \mathbb{R}^{h \times d_S}$, such that $d = d_T + d_S$.

Objective:

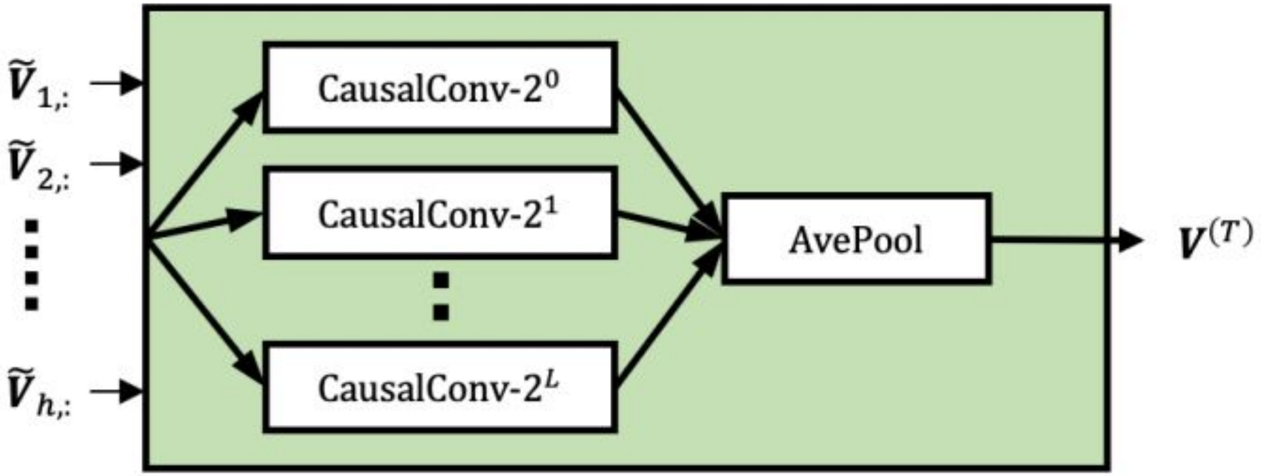
$$\mathcal{L} = \mathcal{L}_{\text{time}} + \frac{\alpha}{2}(\mathcal{L}_{\text{amp}} + \mathcal{L}_{\text{phase}}),$$



(a) Overall Framework

Trend Feature Disentangler

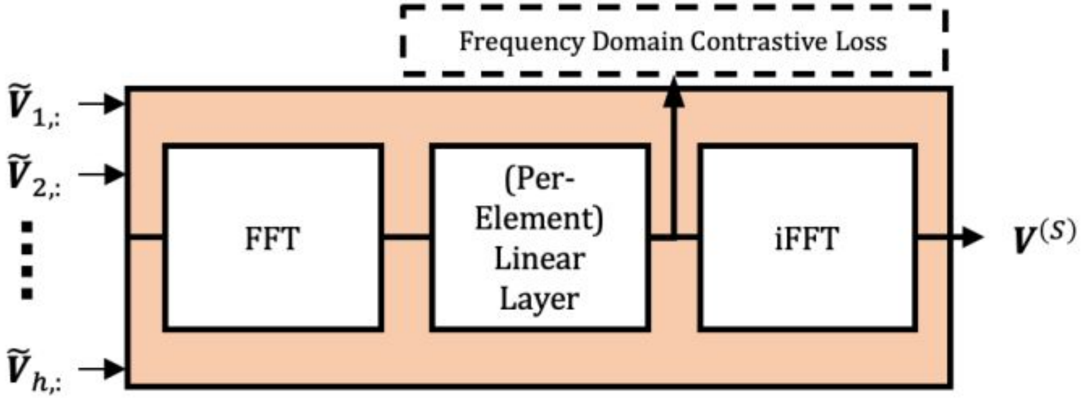
- Key Idea:
  - Extract the trend representations via **a mixture of autoregressive expert**
  - Learn it via **a time domain contrastive loss**



(b) Trend Feature Disentangler

Seasonal Feature Disentangler

- **Key Idea**:
  - Turn to the **frequency domain** to handle the learning of seasonal representations inspired by the spectral analysis
- **Implementation**
  - Incorporate a **learnable Fourier layer** to uncover intra-frequency interactions
  - Develop **a frequency domain contrastive loss** w.r.t. **amplitude** and **phase**



(c) Seasonal Feature Disentangler

Experiments

- Beat the best performing end-to-end forecasting approach by **39.3%** and **18.22%** (MSE) in the **multivariate** and **univariate** settings
- CoST also beat next best performing feature-based approach by **21.3%** and **4.71%** (MSE) in the **multivariate** and **univariate** settings respectively

Table 1: Multivariate forecasting results. Best results are highlighted in bold.

Methods		Representation Learning								End-to-end Forecasting					
		CoST		TS2Vec		TNC		MoCo		Informer		LogTrans		TCN	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	24	**0.386**	**0.429**	0.590	0.531	0.708	0.592	0.623	0.555	0.577	0.549	0.686	0.604	0.583	0.547
	48	**0.437**	**0.464**	0.624	0.555	0.749	0.619	0.669	0.586	0.685	0.625	0.766	0.757	0.670	0.606
	168	**0.643**	**0.582**	0.762	0.639	0.884	0.699	0.820	0.674	0.931	0.752	1.002	0.846	0.811	0.680
	336	**0.812**	**0.679**	0.931	0.728	1.020	0.768	0.981	0.755	1.128	0.873	1.362	0.952	1.132	0.815
	720	**0.970**	**0.771**	1.063	0.799	1.157	0.830	1.138	0.831	1.215	0.896	1.397	1.291	1.165	0.813
ETTh2	24	0.447	0.502	**0.423**	**0.489**	0.612	0.595	0.444	0.495	0.720	0.665	0.828	0.750	0.935	0.754
	48	0.699	0.637	**0.619**	**0.605**	0.840	0.716	0.613	0.595	1.457	1.001	1.806	1.034	1.300	0.911
	168	**1.549**	**0.982**	1.845	1.074	2.359	1.213	1.791	1.034	3.489	1.515	4.070	1.681	4.017	1.579
	336	**1.749**	**1.042**	2.194	1.197	2.782	1.349	2.241	1.186	2.723	1.340	3.875	1.763	3.460	1.456
	720	**1.971**	**1.092**	2.636	1.370	2.753	1.394	2.425	1.292	3.467	1.473	3.913	1.552	3.106	1.381
ETTm1	24	**0.246**	**0.329**	0.453	0.444	0.522	0.472	0.458	0.444	0.323	0.369	0.419	0.412	0.363	0.397
	48	**0.331**	**0.386**	0.592	0.521	0.695	0.567	0.594	0.528	0.494	0.503	0.507	0.583	0.542	0.508
	96	**0.378**	**0.419**	0.635	0.554	0.731	0.595	0.621	0.553	0.678	0.614	0.768	0.792	0.666	0.578
	288	**0.472**	**0.486**	0.693	0.597	0.818	0.649	0.700	0.606	1.056	0.786	1.462	1.320	0.991	0.735
	672	**0.620**	**0.574**	0.782	0.653	0.932	0.712	0.821	0.674	1.192	0.926	1.669	1.461	1.032	0.756
Electricity	24	**0.136**	**0.242**	0.287	0.375	0.354	0.423	0.288	0.374	0.312	0.387	0.297	0.374	0.235	0.346
	48	**0.153**	**0.258**	0.309	0.391	0.376	0.438	0.310	0.390	0.392	0.431	0.316	0.389	0.253	0.359
	168	**0.175**	**0.275**	0.335	0.410	0.402	0.456	0.337	0.410	0.515	0.509	0.426	0.466	0.278	0.372
	336	**0.196**	**0.296**	0.351	0.422	0.417	0.466	0.353	0.422	0.759	0.625	0.365	0.417	0.287	0.382
	720	**0.232**	**0.327**	0.378	0.440	0.442	0.483	0.380	0.441	0.969	0.788	0.344	0.403	0.287	0.381
Weather	24	**0.298**	**0.360**	0.307	0.363	0.320	0.373	0.311	0.365	0.335	0.381	0.435	0.477	0.321	0.367
	48	**0.359**	**0.411**	0.374	0.418	0.380	0.421	0.372	0.416	0.395	0.459	0.426	0.495	0.386	0.423
	168	**0.464**	**0.491**	0.491	0.506	0.479	0.495	0.482	0.499	0.608	0.567	0.727	0.671	0.491	0.501
	336	**0.497**	0.517	0.525	0.530	0.505	0.514	0.516	0.523	0.702	0.620	0.754	0.670	0.502	**0.507**
	720	0.533	0.542	0.556	0.552	0.519	0.525	0.540	0.540	0.831	0.731	0.885	0.773	**0.498**	**0.508**
Avg.		**0.590**	**0.524**	0.750	0.607	0.870	0.655	0.753	0.608	1.038	0.735	1.180	0.837	0.972	0.666

Ablation Study

Table 3: Ablation study of various components of CoST on ETT datasets). TFD: Trend Feature Disentangler, MARE: Mixture of Auto-regressive Experts (TFD without MARE refers to the TFD module with a single AR expert with kernel size $\lfloor h/2 \rfloor$), SFD: Seasonal Feature Disentangler, LFL: Learnable Fourier Layer, FDCL: Frequency Domain Contrastive Loss. † indicates a model trained end-to-end with supervised forecasting loss. ‡ indicates † with an additional contrastive loss.

	TFD	MARE	SFD	LFL	FDCL	Multivariate		Univariate	
						MSE	MAE	MSE	MAE
Trend	✓					0.882	0.674	0.115	0.243
	✓	✓				0.789	0.630	0.105	0.235
Seasonal			✓		✓	0.905	0.675	0.105	0.237
			✓	✓		0.895	0.721	0.103	0.239
			✓	✓	✓	0.862	0.668	0.129	0.255
CoST†			-			1.376	0.834	0.228	0.366
CoST‡			-			1.477	0.909	0.965	0.883
MoCo			-			0.996	0.721	0.112	0.248
SimCLR			-			1.021	0.730	0.113	0.248
CoST	✓	✓	✓	✓	✓	**0.781**	**0.625**	**0.102**	**0.233**

Ablation Study

Table 4: Ablation study of various backbone encoders on the ETT datasets.

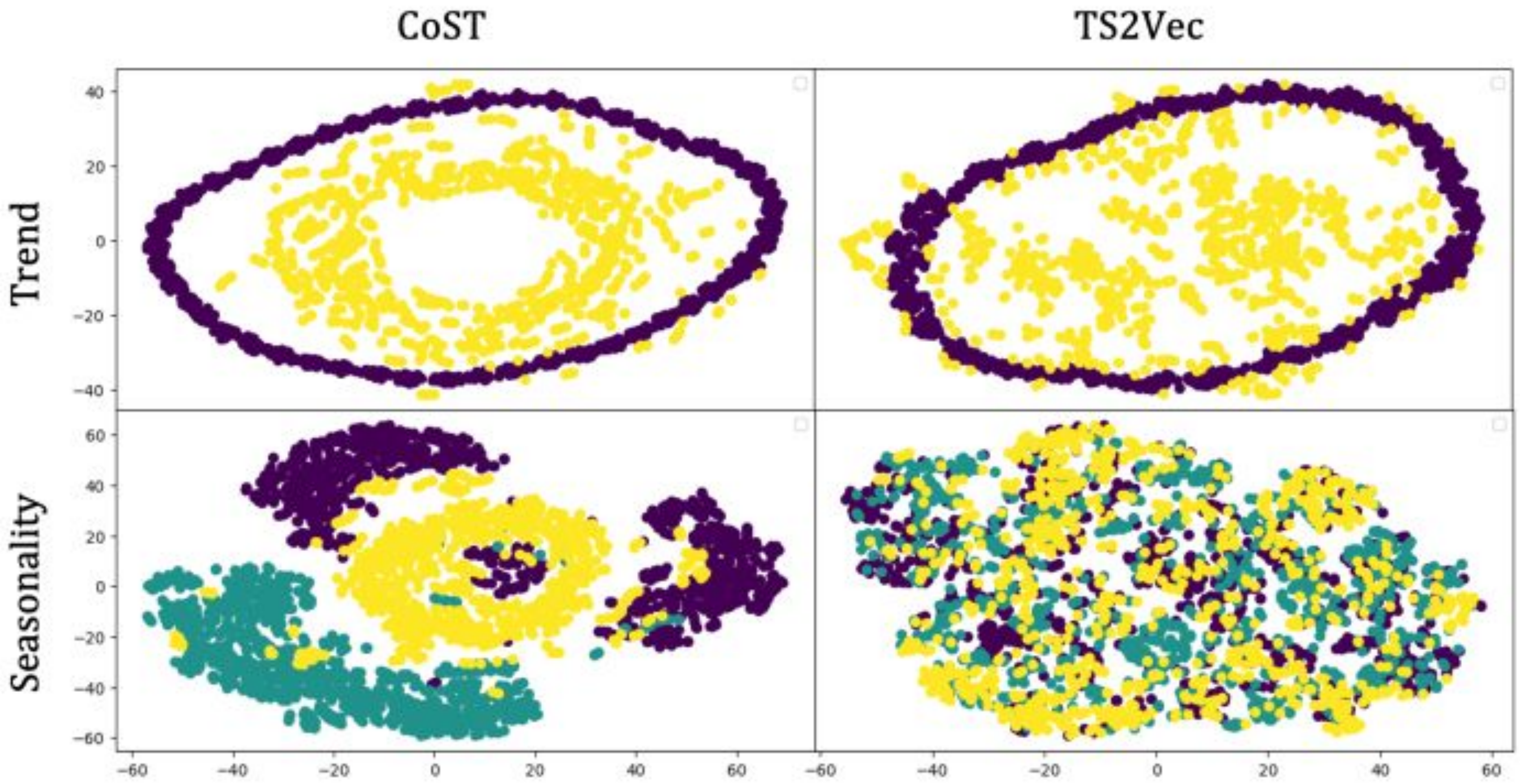
Backbones	TCN				LSTM				Transformer			
Methods	TS2Vec		CoST		TS2Vec		CoST		TS2Vec		CoST	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Multivariate	0.990	0.717	**0.781**	**0.625**	1.415	0.903	**0.928**	**0.706**	1.092	0.766	**0.863**	**0.674**
Univariate	0.116	0.253	**0.102**	**0.233**	0.544	0.596	**0.148**	**0.301**	0.172	0.328	**0.159**	**0.320**

Table 5: Ablation study of various regressors on the ETT datasets

Regressors	Ridge				Linear				Kernel Ridge			
Methods	TS2Vec		CoST		TS2Vec		CoST		TS2Vec		CoST	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Multivariate	0.990	0.717	**0.781**	**0.625**	1.821	0.944	**1.472**	**0.781**	1.045	0.738	**0.868**	**0.686**
Univariate	0.116	0.253	**0.102**	**0.233**	0.304	0.414	**0.182**	**0.310**	0.132	0.273	**0.109**	**0.243**

CoST is robust to various backbones and regressors!

T-SNE Visualizations



Thank You

Code is available at: https://github.com/salesforce/CoST