Towards Diverse and Coherent Augmentation for Time-Series Forecasting

Xiyuan Zhang[1], Ranak Roy Chowdhury[1], Jingbo Shang[1], Rajesh Gupta[1], Dezhi Hong[2]

[1]UC San Diego , [2] Amazon

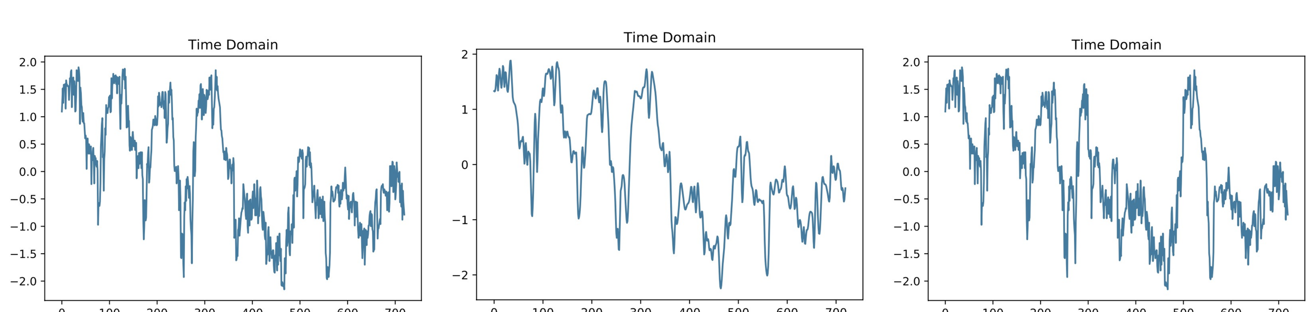
Background and Motivation

Data scarcity in time series is very common

- Malfunctioning sensors: battery depletion
- Privacy concerns: medical data
- Cold-start prediction: new station, new store

Data augmentation mitigates the data scarcity issue

❌ However, existing augmentation methods are designed for classification, where class labels can be preserved even if augmentation changes the temporal dynamics



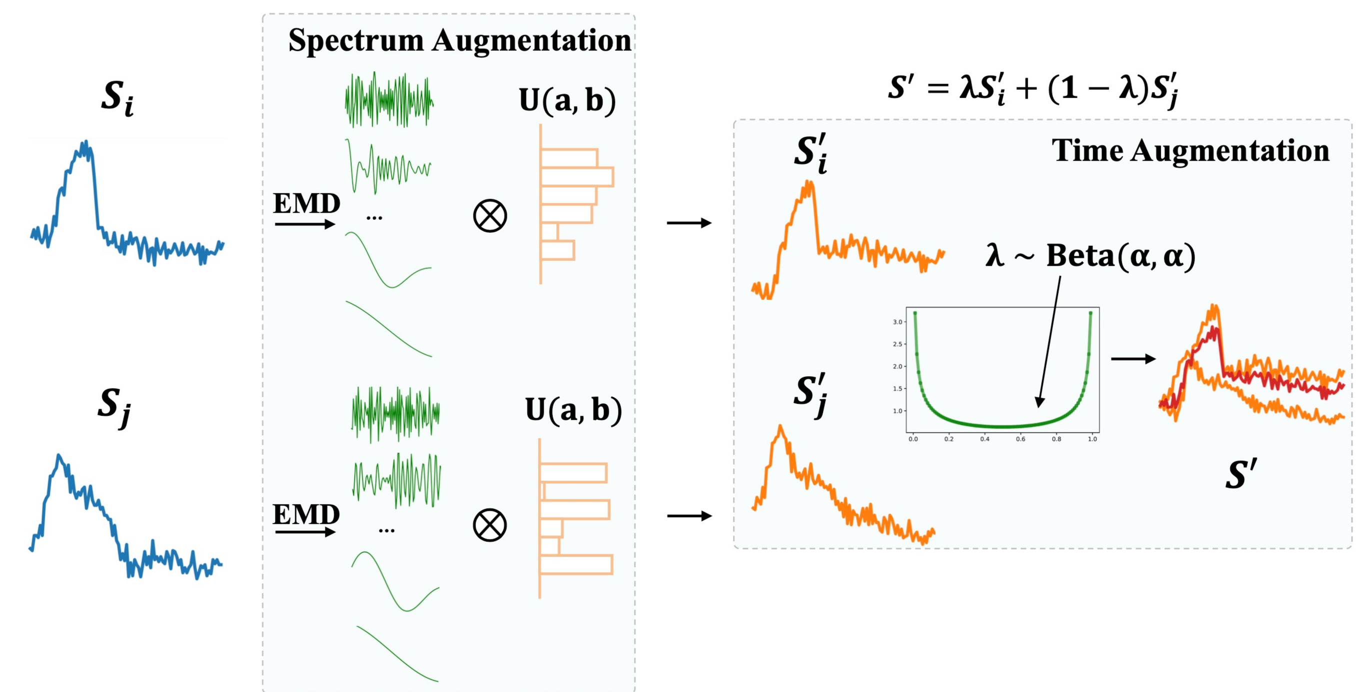
original filtering permutation

✅ Augmentation for forecasting requires ***diversity*** as well as ***coherence*** with the temporal dynamics

💡 Time-series data generated by real-life physical processes have characteristics in both ***time*** and ***frequency*** domains to capture temporal dynamics

Augmentation Method

Combine frequency-domain augmentation and time-domain augmentation



Frequency-domain augmentation: decompose time series with Empirical Mode Decomposition (EMD), and reassemble subcomponents with random weights sampled from $U(a, b)$

$$\mathbf{S} = \sum_{i=1}^{n} \mathrm{IMF}_i + \mathbf{R} \qquad \mathbf{S}' = \sum_{i=1}^{n} w_i \cdot \mathrm{IMF}_i$$

Diversity: random weights to combine components

Coherence: augmented samples have the same set of base components

Time-domain augmentation: mix up values at both history and future time steps with $\lambda \sim Beta(\alpha, \alpha)$

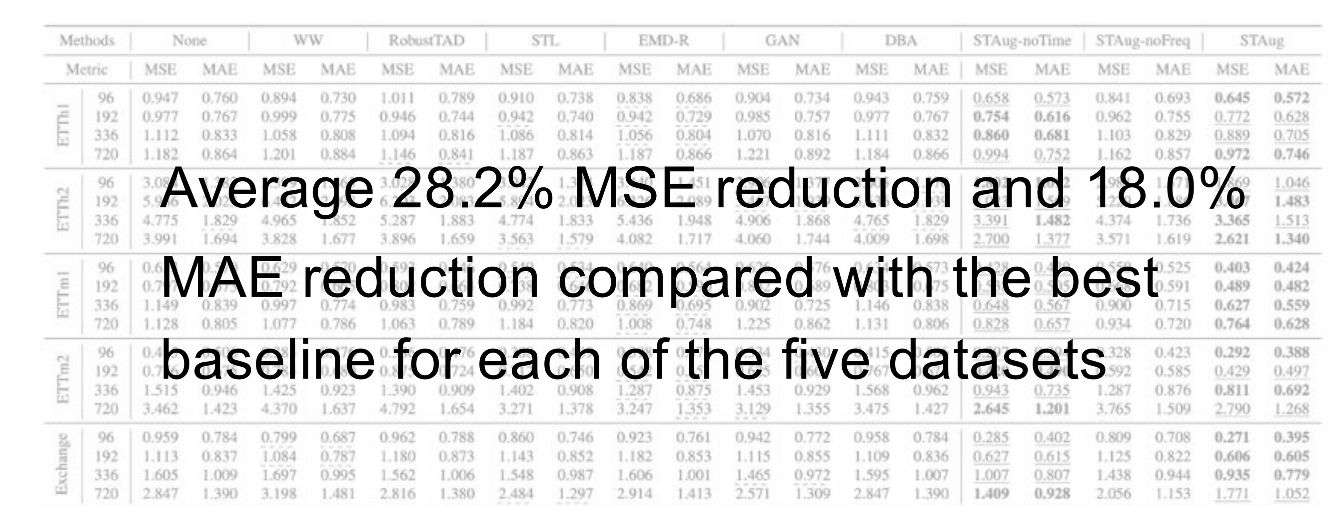
$$\mathbf{H}' = \lambda \mathbf{H_i} + (1-\lambda)\mathbf{H_j} \qquad \mathbf{F}' = \lambda \mathbf{F_i} + (1-\lambda)\mathbf{F_j}$$

Diversity: random weights to combine series pairs

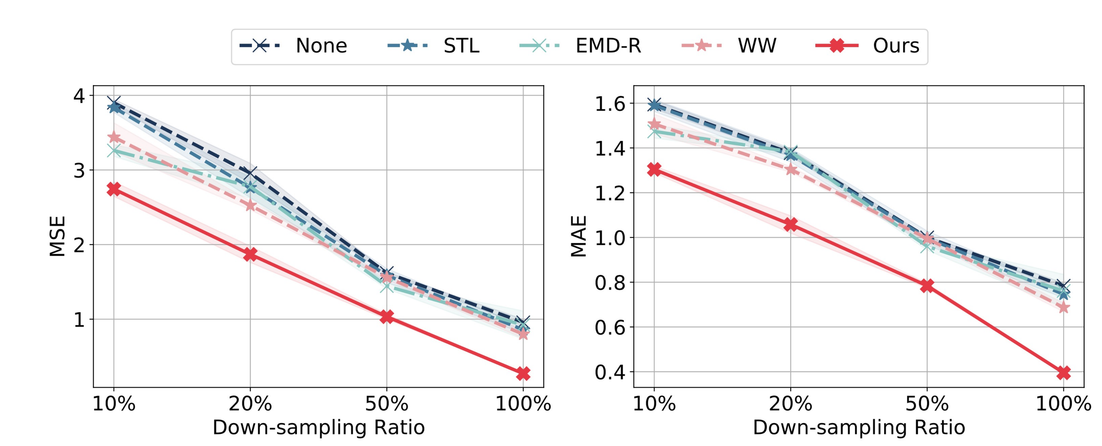
Coherence: only augment linearly in-between samples by the interpolation nature of “mix up”

Experiments

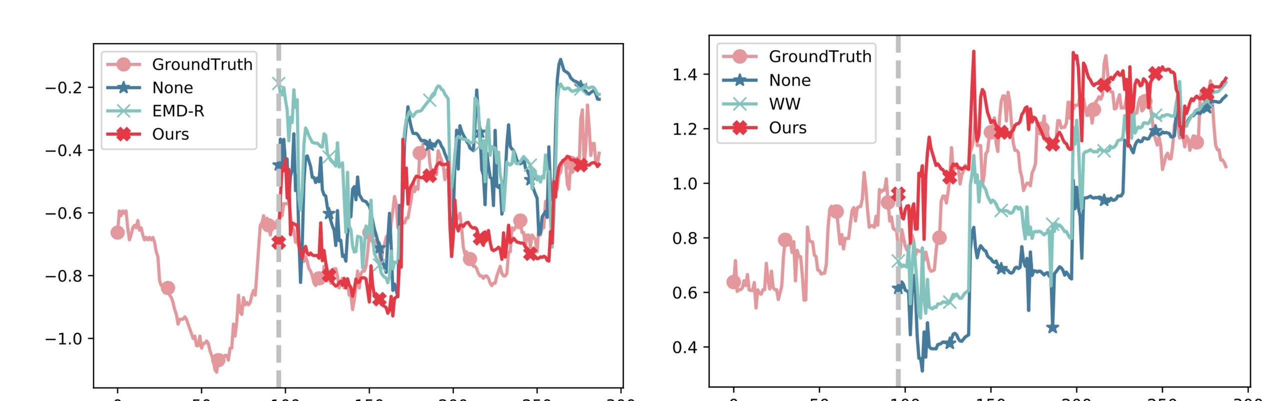
Forecasting backbone: Informer (AAAI’ 2021 best paper)



Our method is the most robust with respect to down-sampling training data



Case study: prediction better aligns with ground truth



ETTm1 dataset Exchange dataset