Two Birds with One Stone: Series Saliency for Accurate and Interpretable Multivariate Time Series Forecasting

Qingyi Pan, Wenbo Hu, Ning Chen*

Dept. of Comp. Sci. & Tech., Institute for AI, BNRist Lab, THBI Lab, Tsinghua University

Tsinghua University

IJCAI 2021 MONTREAL

Introduction

❑ It is important yet challenging to perform **accurate** and **interpretable** time series forecasting. Traditional parametric model are easy-to-interpret, but their predictive capabilities are limited. Deep architectures can boost the forecasting accuracy, they sacrifice interpretability. It is relatively unexplored to develop both accurate and interpretable methods for **multivariate time series forecasting.**

❑ **Existing work:**

1. Interpretation methods for general neural networks [Ribeiro *et al.,* 2016; Shrikumar *et al.,* 2017; Lundberg & Lee 2017]:

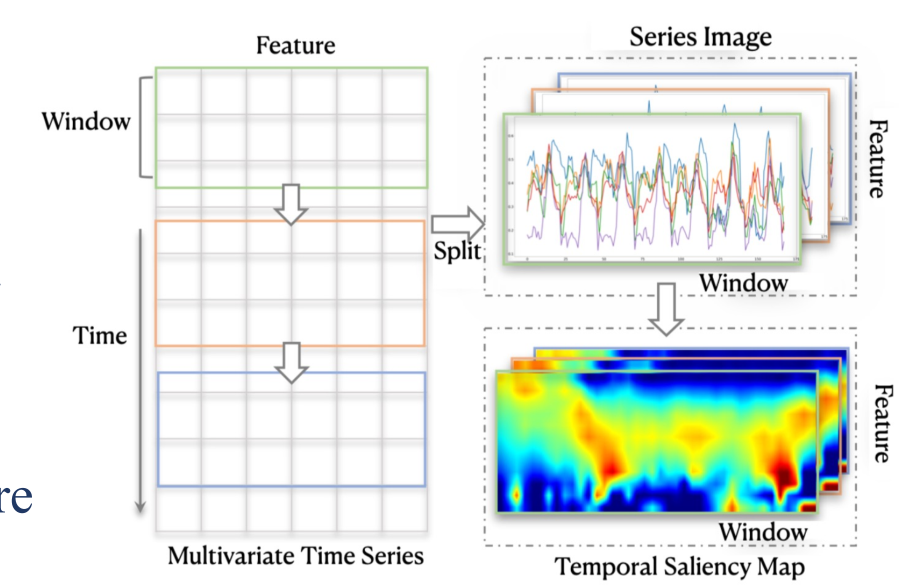
   Use gradient information to extract feature information for after the back- propagation

2. Transfer attention methods from the fields of language or vision [Bahadanau *et al.,* 2014, Shih *et al.,* 2019]:

   Attention values are calculated via the relative importance of the different time steps.

❑ **Key: Considering the time and feature dimensions in coherent manner**

- Represent the multivariate time series as a set of *window* × *feature* 2D series images.
- Each series image corresponds to a part of the multivariate time series within a given time window.
- Each row corresponds to one feature dimension.



- Follow the perturbation strategy in the smallest destroying region (SDR) principle [Dabkowski & Gal, 2017] for the reference series image.

$$\hat{x}_{t,i} = \begin{cases} x_{t,i} + \epsilon_{\sigma_1} & \text{noise} \\ g_{\sigma_2}(x_{t,i}) & \text{blur} \end{cases}$$

- Each series can be composed of three parts: trend, seasonality and residual [Hyandman & Athanasopoulos, 2018]
- Adding Gaussian blur extracts the trend information, and adding some noise enhances local information.

❑ **Motivation: Sensitivity over perturbation for DNNs in time domain**

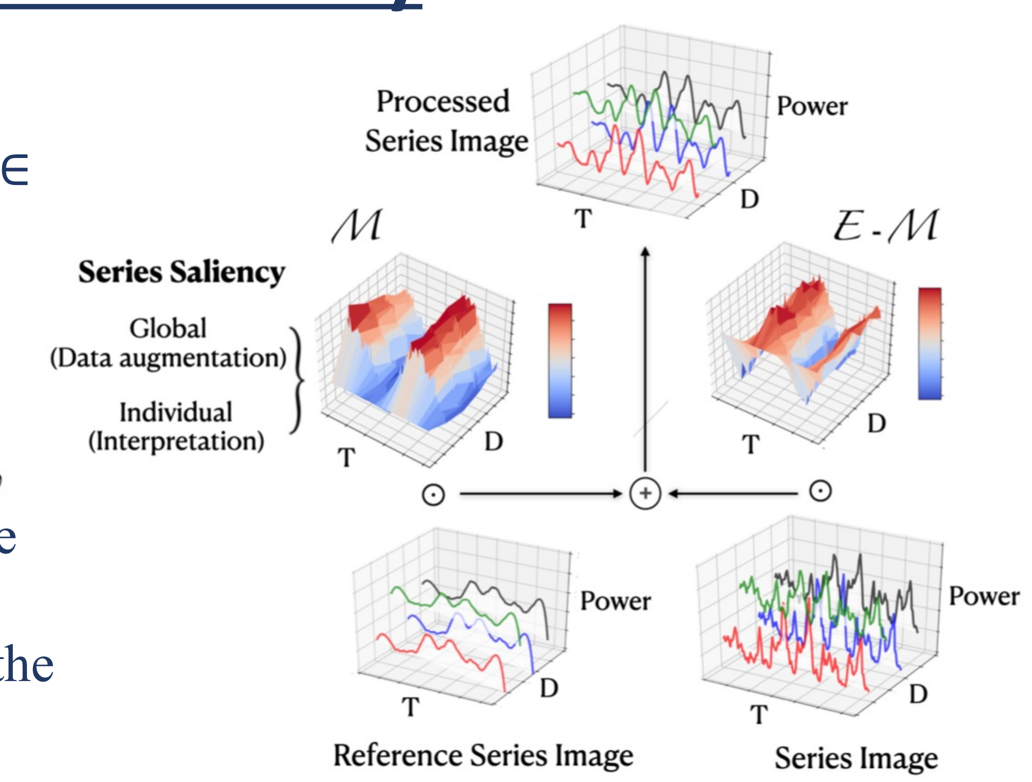
- When the amount of injected noise or blurring is small, the reference series image can be treated as data augmentation in time domain for deep models.
- If the perturbation is not set properly (e.g., too large), the blurring will introduce irregular roughness to cover the original series, making it difficult for DNNs to learn temporal patterns.

Series Saliency

■ Series saliency module introduces a learnable mask $M \in [0,1]^{D\times T}$ and selectively combines the reference series images and original one:

$$\tilde{X} = M \odot \hat{X} + (E - M) \odot X,$$

■ Series saliency can generate data that cover the unexplored input space while maintaining the important characteristics of the original series image.



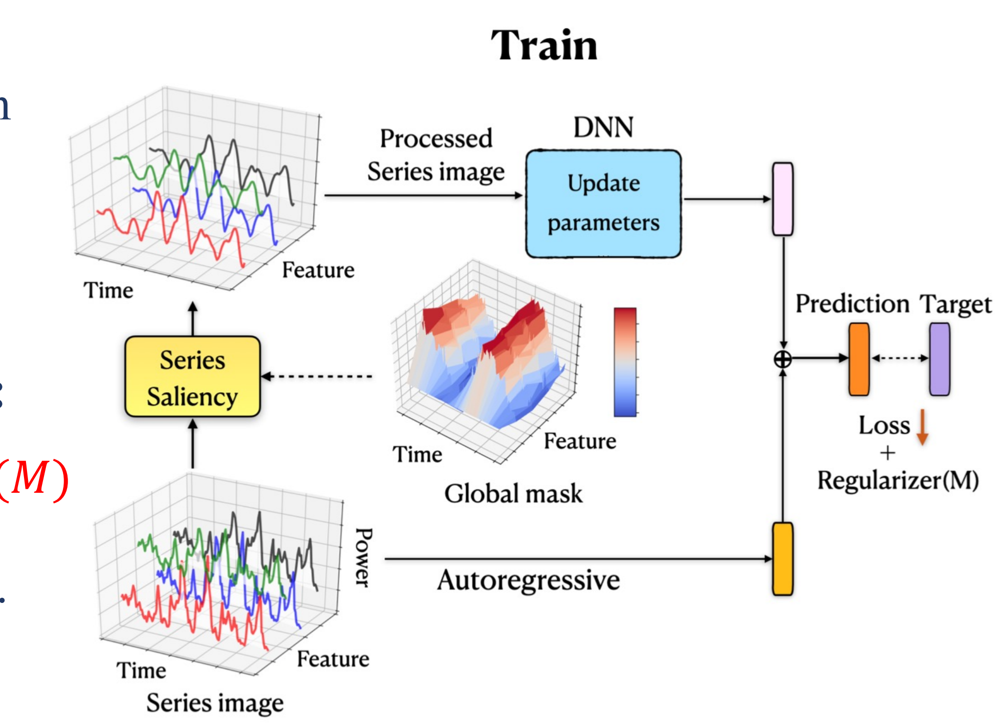
Mixup mechanism

Methodology

❑ **Dual-path architecture**

- Scale of input data often changes in a non-periodic manner

$$\hat{y} = \underbrace{y^{(o)}}_{\text{Linear}} + \underbrace{y^{(r)}}_{\text{Non-Linear}}$$



❑ **Training with Series Saliency:**

$$\min_{\phi,M} \ell_1(\phi, M) + \lambda_1 \ell_m(M) + \lambda_2 \ell_r(M)$$

where λ_1 and λ_2 are the coefficient.

❑ **Training loss ℓ_1:**

- $\ell(\phi, M) = \left\|f(\tilde{X}_i) - y_i\right\|^2 + \|f(X_i) - y_i\|^2$
- Generate the augmented series image and fully leverage the training set.

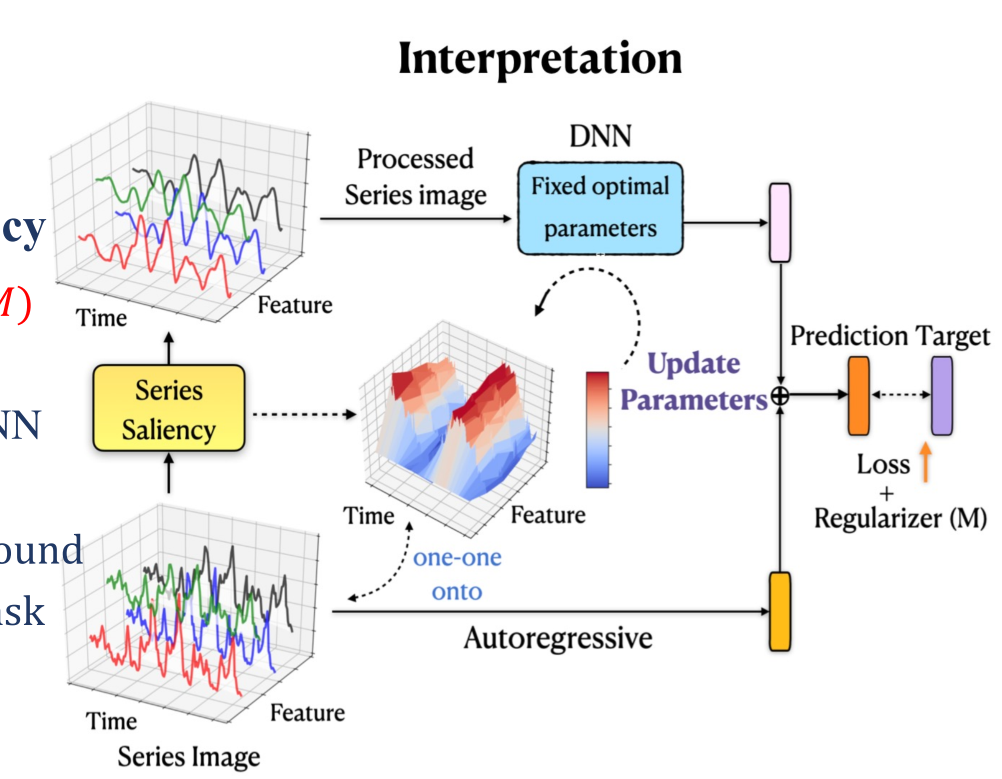
❑ **Regularization term:**

- $\ell_m(M) = \|M\|_2$
- $\ell_r(M) = \|MM^T - I\|_F$

❑ **Interpretation with Series Saliency**

$$\min_{M} -\|\tilde{y}^* - y^*\| + \lambda_1 \ell_m(M) + \lambda_2 \ell_r(M)$$

- Fixed optimal parameters ϕ^* of the DNN and AR models after training.
- The most salient feature region are found by identifying the representative mask
- For AR part, the weights are easy-to-interpret because of linearity.



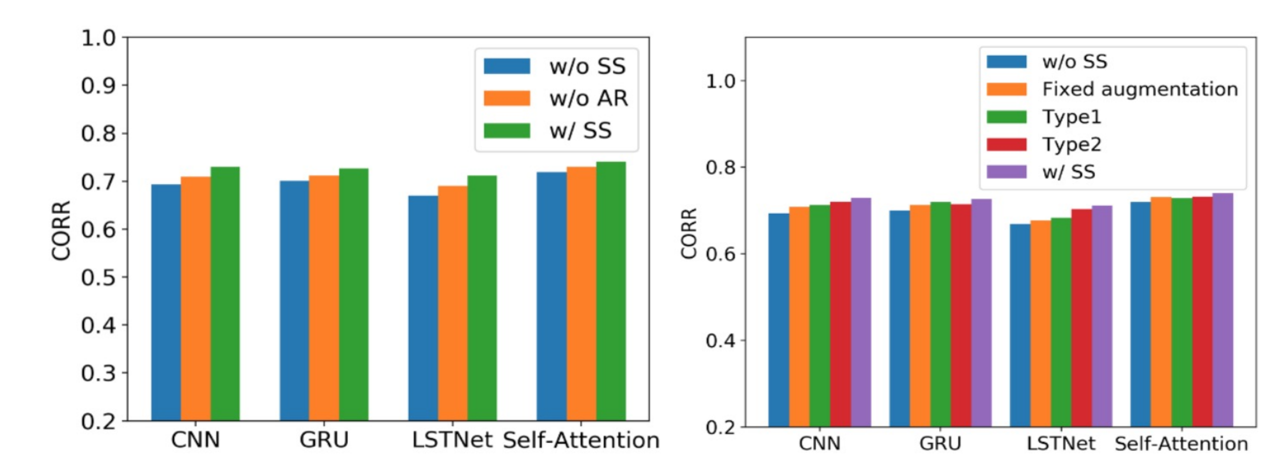
Experiments

❑ **Results on Forecasting**

➢ The three datasets are representative (from difficult to easy) and dimension from 13 to 321.

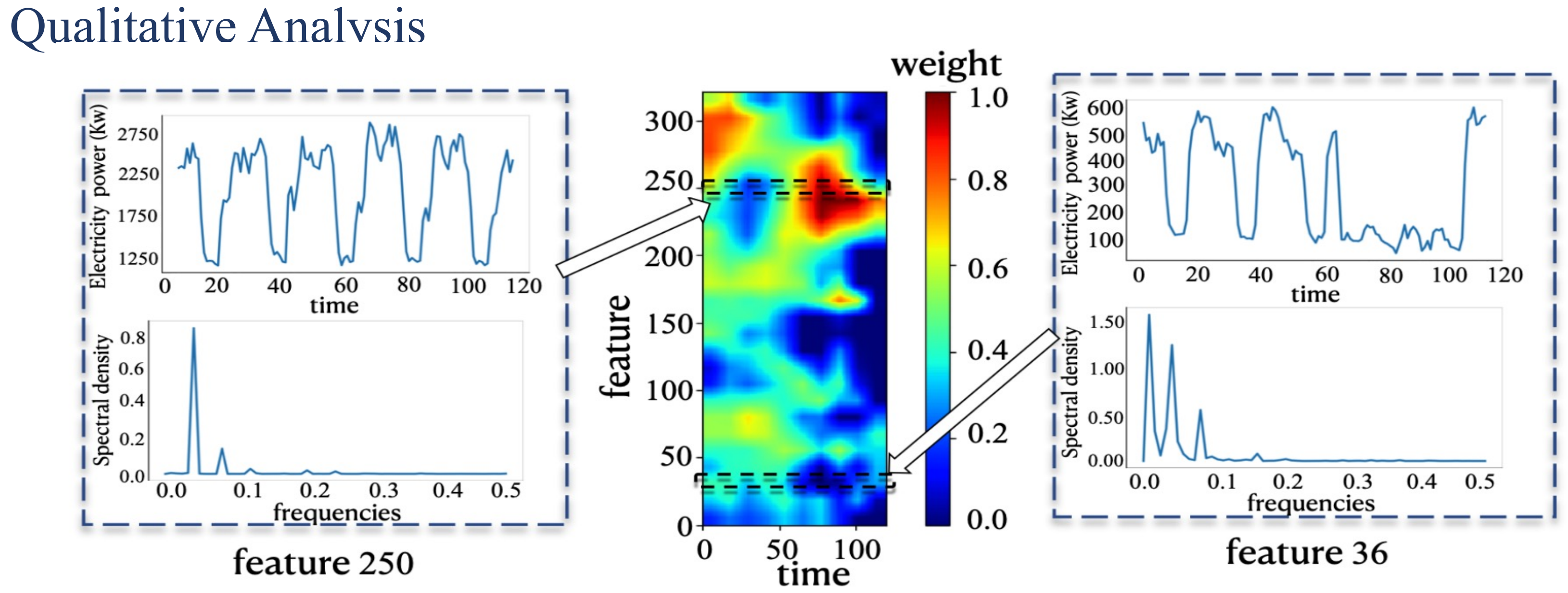
Methods	Air Quality $\tau=3$	Air Quality $\tau=6$	Air Quality $\tau=12$	Industry $\tau=3$	Industry $\tau=6$	Industry $\tau=12$	Electricity $\tau=3$	Electricity $\tau=6$	Electricity $\tau=12$
CNN	0.775 ± 0.003	0.701 ± 0.001	0.636 ± 0.001	0.919 ± 0.022	0.909 ± 0.019	0.841 ± 0.008	0.883 ± 0.004	0.871 ± 0.002	0.866 ± 0.004
GRU	0.804 ± 0.003	0.712 ± 0.002	0.639 ± 0.003	0.953 ± 0.003	0.936 ± 0.013	0.904 ± 0.011	0.878 ± 0.001	0.877 ± 0.003	0.867 ± 0.002
LSTNet	0.777 ± 0.001	0.708 ± 0.004	0.624 ± 0.004	0.949 ± 0.004	0.934 ± 0.003	0.876 ± 0.011	0.922 ± 0.004	0.913 ± 0.002	0.906 ± 0.002
SA	0.813 ± 0.002	0.722 ± 0.003	0.643 ± 0.003	0.961 ± 0.002	0.942 ± 0.005	0.905 ± 0.009	0.919 ± 0.007	0.907 ± 0.001	0.902 ± 0.003
CNN w/ SS	0.779 ± 0.005	0.723 ± 0.009	0.641 ± 0.007	0.941 ± 0.006	0.927 ± 0.004	0.881 ± 0.001	0.898 ± 0.004	0.893 ± 0.002	0.892 ± 0.007
GRU w/ SS	0.809 ± 0.003	0.716 ± 0.012	0.649 ± 0.003	0.955 ± 0.001	0.935 ± 0.002	0.912 ± 0.003	0.905 ± 0.004	0.889 ± 0.008	0.878 ± 0.003
LSTNet w/ SS	0.794 ± 0.008	0.724 ± 0.002	0.641 ± 0.003	0.959 ± 0.004	0.938 ± 0.001	0.901 ± 0.002	$\mathbf{0.928} \pm 0.003$	$\mathbf{0.918} \pm 0.003$	0.907 ± 0.001
SA w/ SS	$\mathbf{0.819} \pm 0.003$	$\mathbf{0.732} \pm 0.009$	$\mathbf{0.658} \pm 0.001$	$\mathbf{0.965} \pm 0.003$	$\mathbf{0.955} \pm 0.016$	$\mathbf{0.916} \pm 0.004$	0.923 ± 0.003	0.915 ± 0.001	$\mathbf{0.911} \pm 0.002$

➢ Ablation study



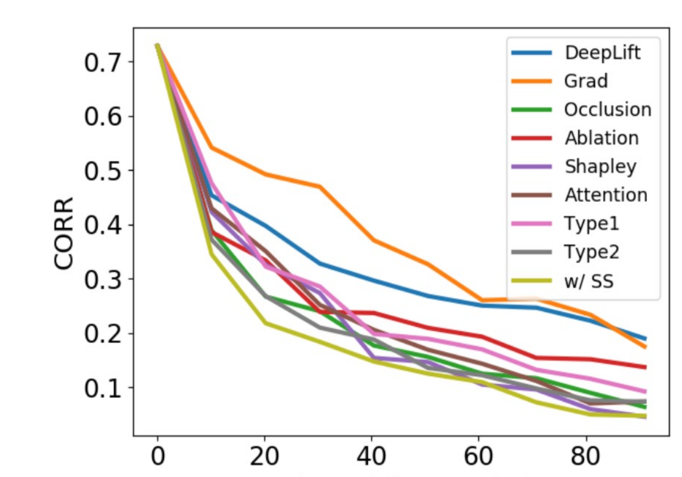
❑ **Results on Interpretation**

➢ Qualitative Analvsis



➢ Quantitative Comparison

Methods	Industry	Air Quality	Electricity
Grad	0.214 ± 0.007	0.297 ± 0.007	0.199 ± 0.008
DeepLift	0.211 ± 0.008	0.241 ± 0.006	0.174 ± 0.003
Ablation	0.204 ± 0.008	0.225 ± 0.006	0.213 ± 0.009
Occlusion	0.124 ± 0.004	0.221 ± 0.011	0.142 ± 0.005
Shapley	0.145 ± 0.006	0.211 ± 0.010	0.171 ± 0.005
Attention	0.141 ± 0.007	0.203 ± 0.007	0.139 ± 0.003
Type1	0.131 ± 0.005	0.205 ± 0.008	0.143 ± 0.004
Type2	0.122 ± 0.002	0.201 ± 0.005	0.135 ± 0.001
w/ SS	$\mathbf{0.117} \pm 0.002$	$\mathbf{0.192} \pm 0.003$	$\mathbf{0.131} \pm 0.002$



Conclusion

❑ **We propose Series Saliency to boost both accuracy and interpretability for multivariate time series forecasting.**

❑ **Series saliency module acts as an adaptive data augmentation method for training deep models while can be optimized for interpretable forecasting in both feature and time dimensions.**

www.PosterPresentations.com