Published as a conference paper at ICLR 2023

TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis

Haixu Wu[*], Tengge Hu[*], Yong Liu[*], Hang Zhou, Jianmin Wang, Mingsheng Long[✉]
School of Software, BNRist, Tsinghua University, Beijing 100084, China
{whx20,liuyong21,htg21,h-zhou18}@mails.tsinghua.edu.cn
{jimwang,mingsheng}@tsinghua.edu.cn


Haixu Wu


Tengge Hu


Yong Liu


Hang Zhou


Jianmin Wang


Mingsheng Long

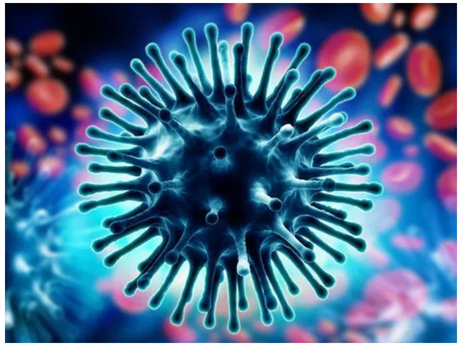
Time Series In Real World

Energy
Consumption

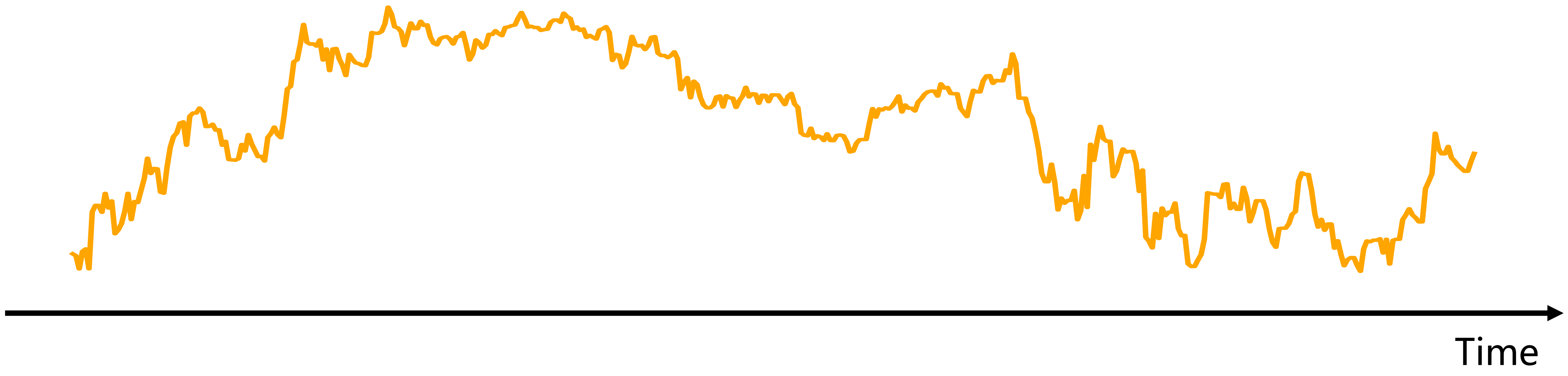
Traffic
Flow

Economic
Changes

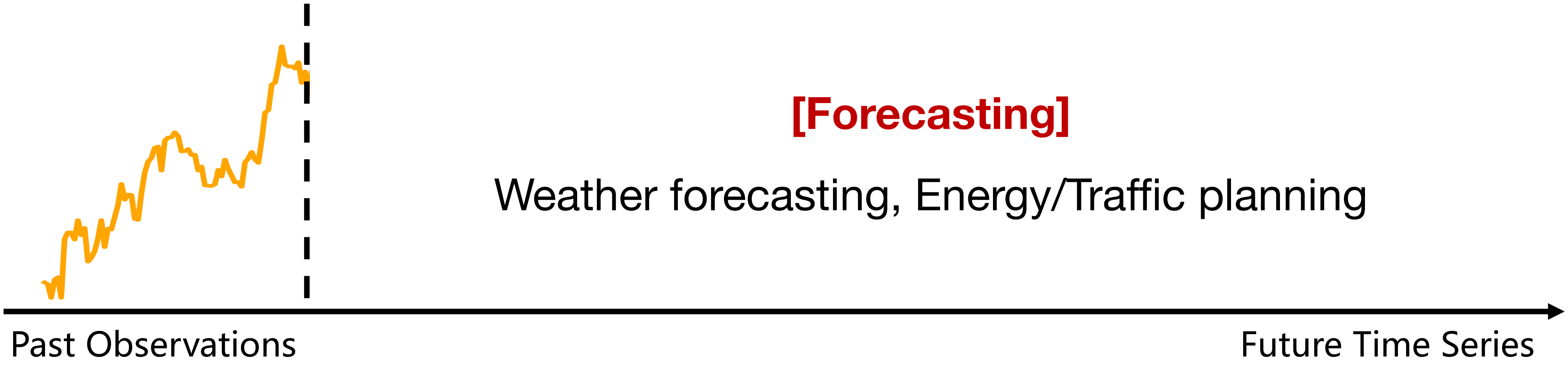
Weather
Variations



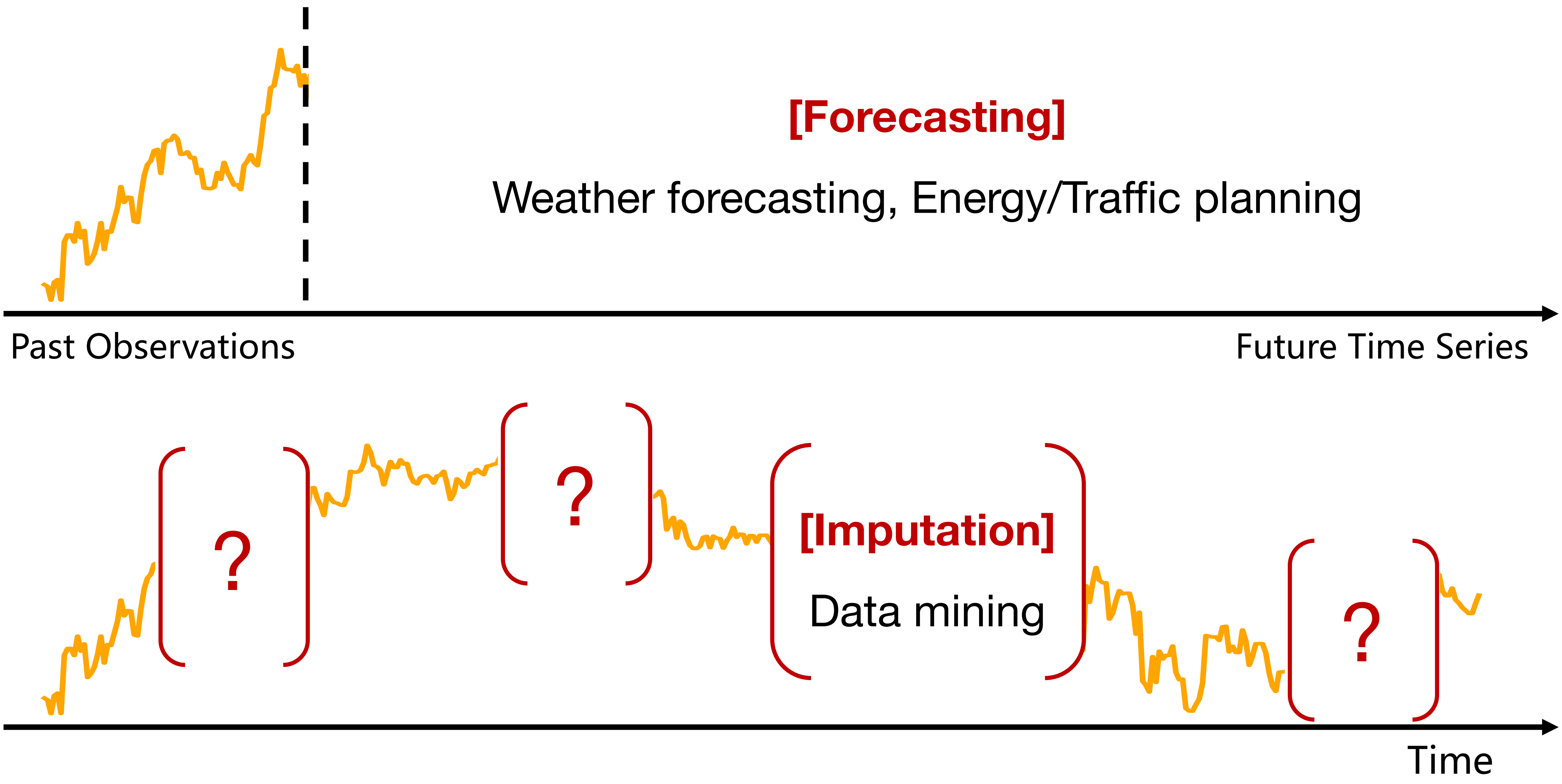
Disease
Propagation



Time Series Analysis

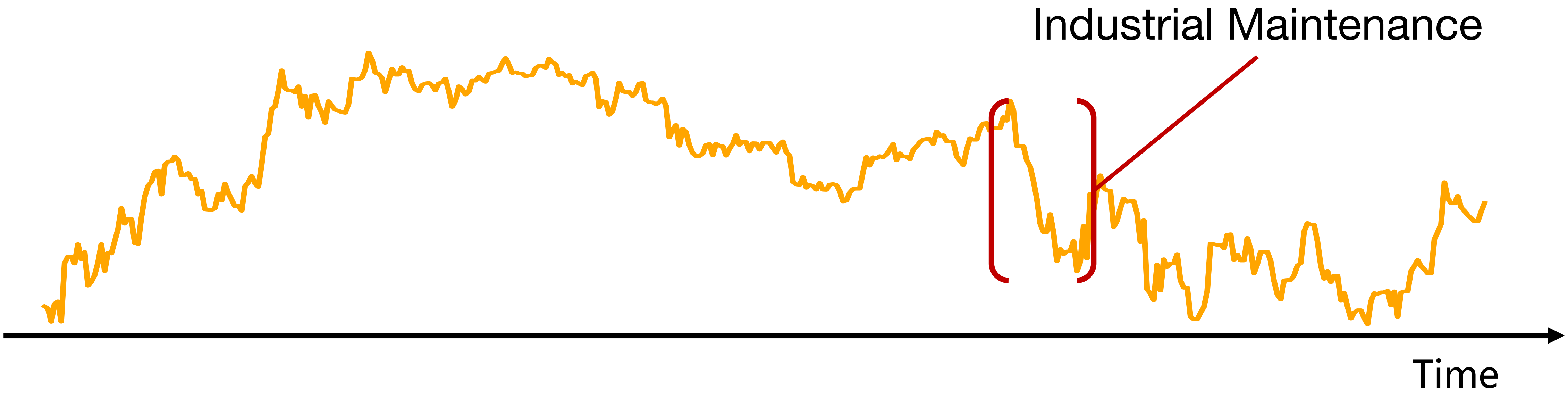


Time Series Analysis

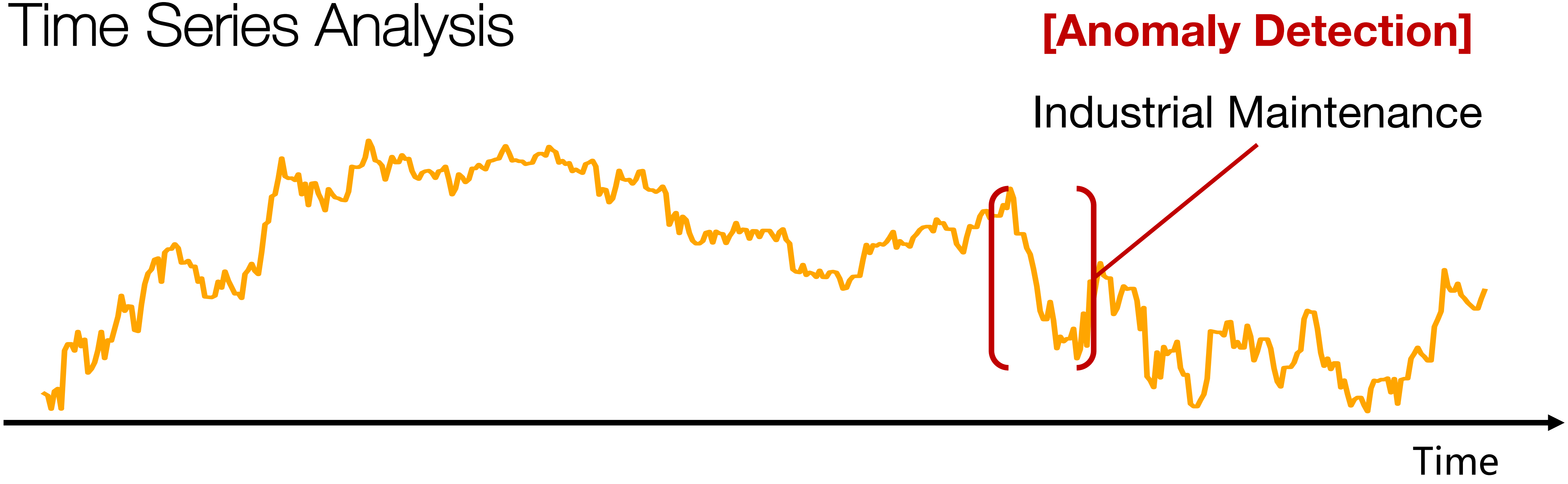


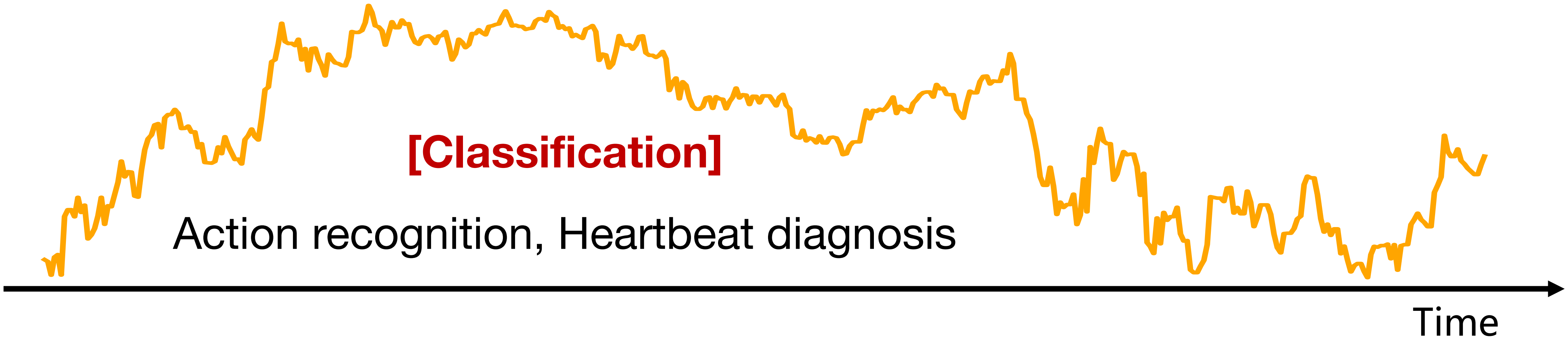
Time Series Analysis

[Anomaly Detection]

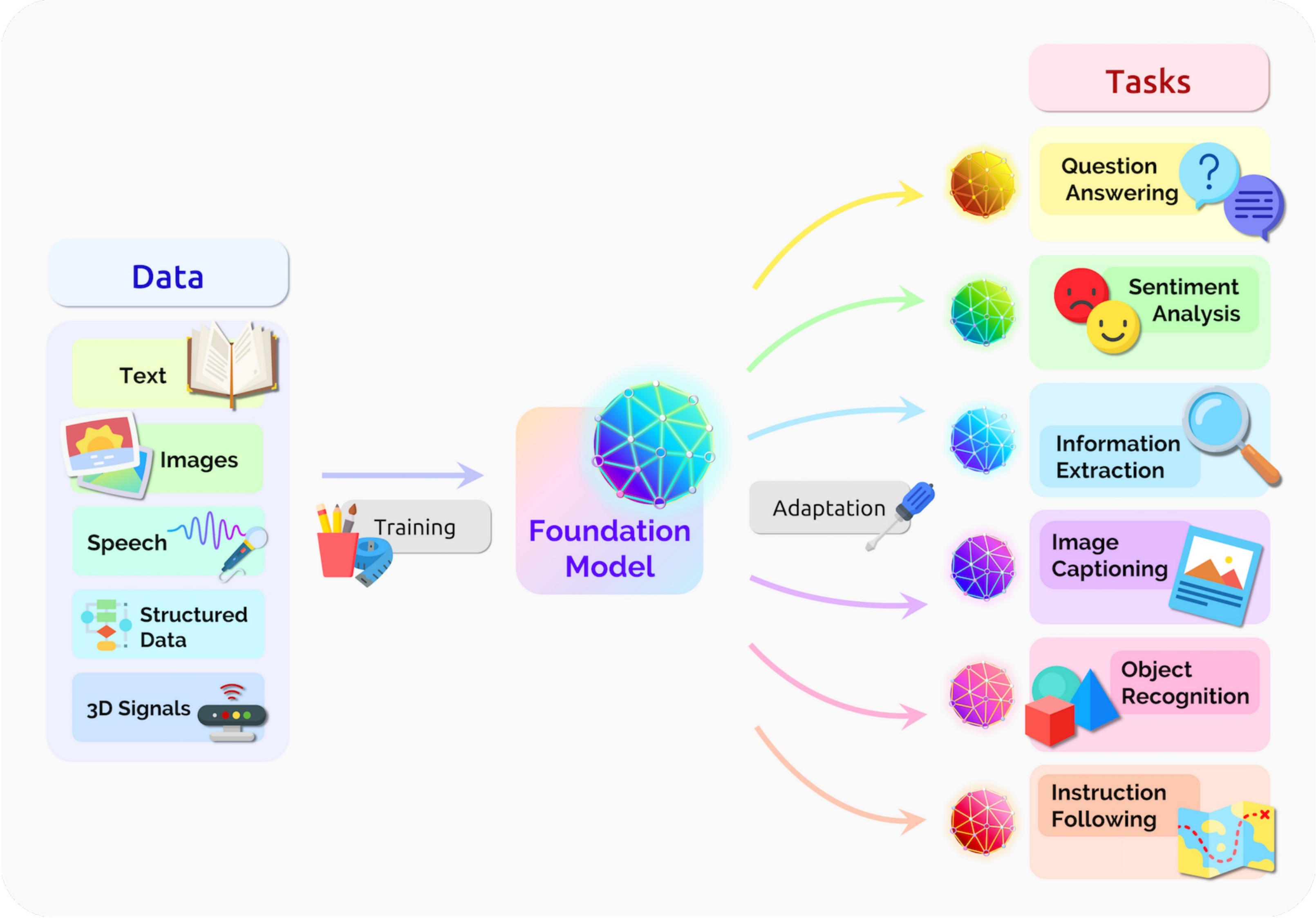


Time Series Analysis





In Pursing Foundation Models



[Data Universal]

Learn from various modalities

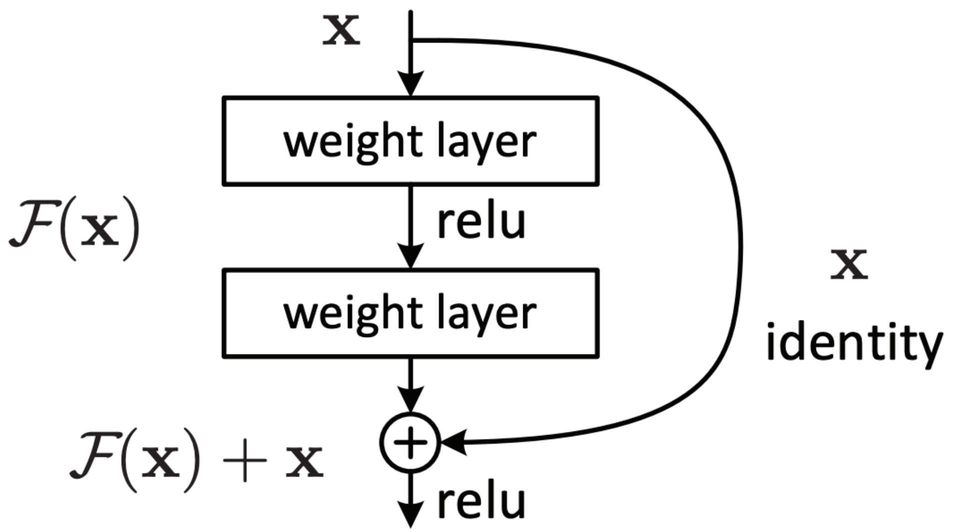
[Task Universal]

Adapt to a wide range of downstream tasks

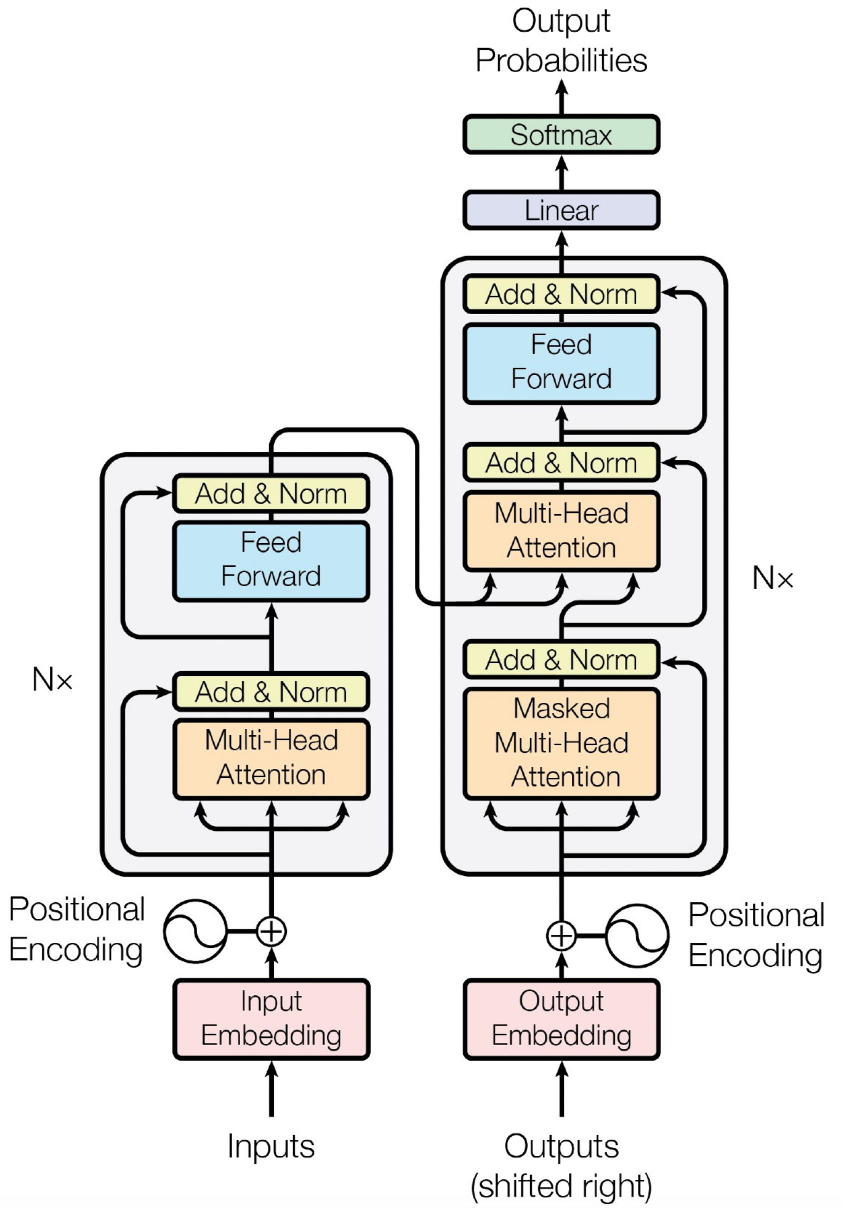
Bommasani et al. On the Opportunities and Risks of Foundation Models. *Arxiv 2021.*

Foundation Models in CV and NLP

Universal backbone with task-specific heads for different tasks.



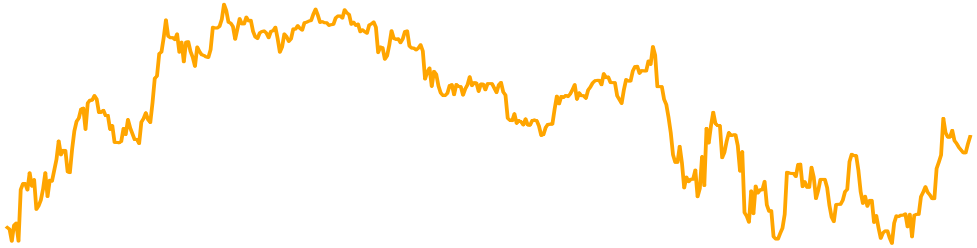
Classification, Object detection, Segmentation



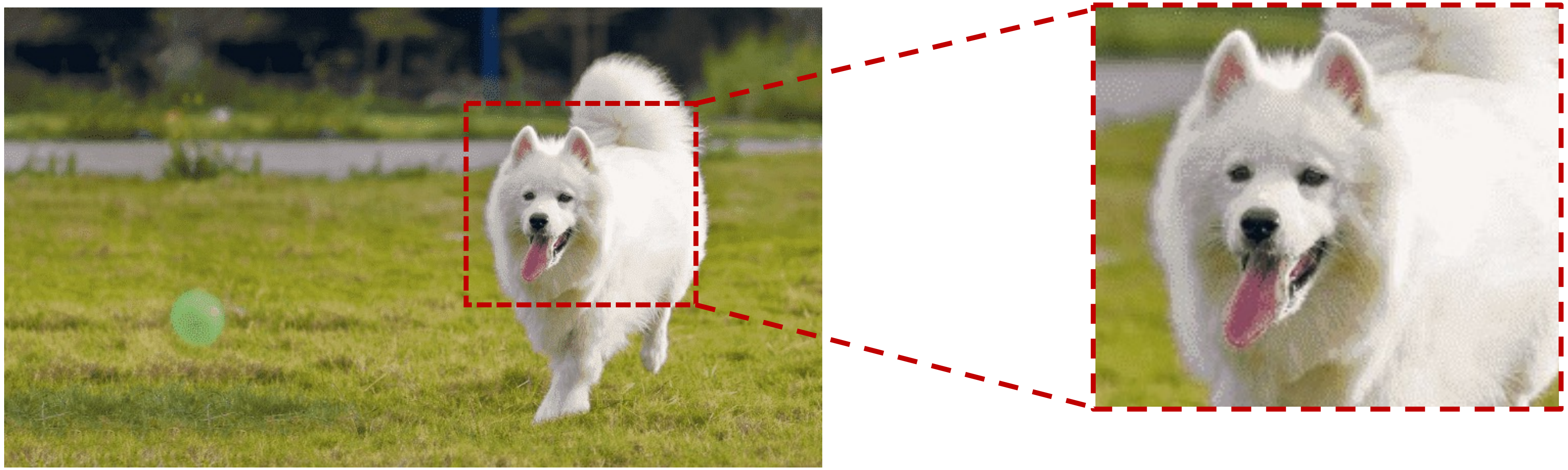
Classification, Generation

Differences among Image, Language, Time Series

TimesNet is for time series analysis.



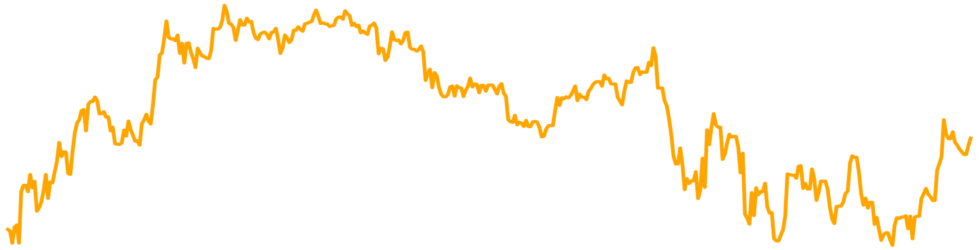
Differences among Image, Language, Time Series



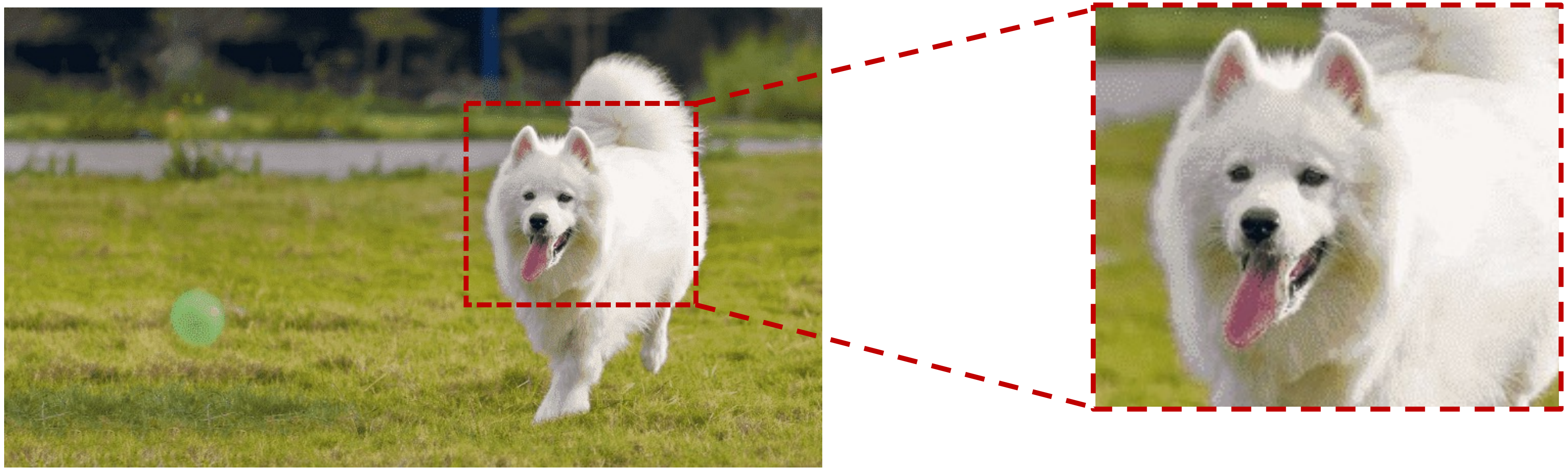
TimesNet is for time series analysis.

Analysis is the process of breaking a complex topic into smaller parts for a better understanding.

Wikipedia
The Free Encyclopedia



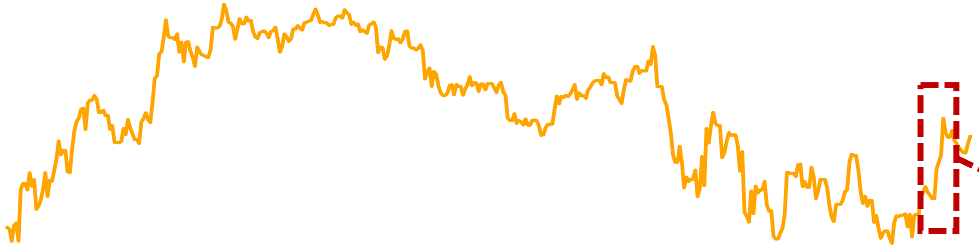
Differences among Image, Language, Time Series



TimesNet is for time series analysis.

Analysis is the process of breaking a complex topic into smaller parts for a better understanding.

Wikipedia
The Free Encyclopedia

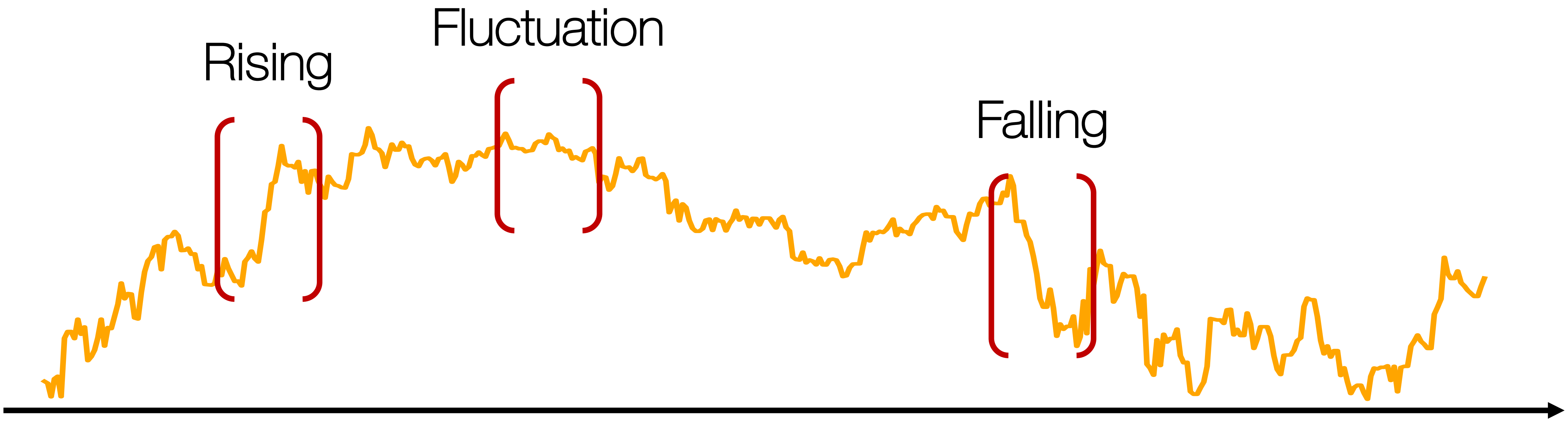


Each time point only saves some scalars.

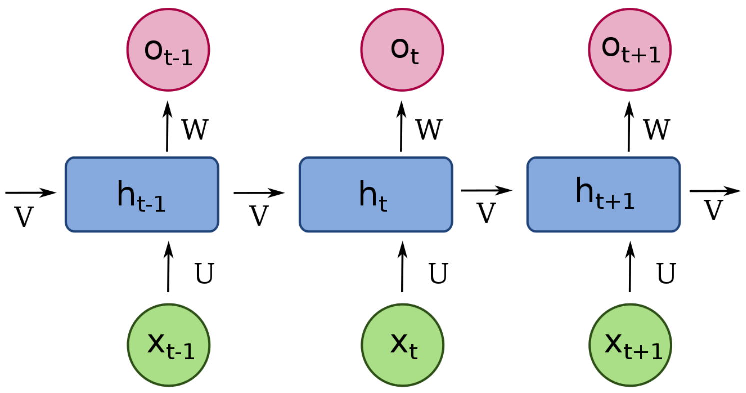
Temporal Variations of Time Series

More information of time series is in temporal variations,

such as continuity, periodicity, trend and etc.

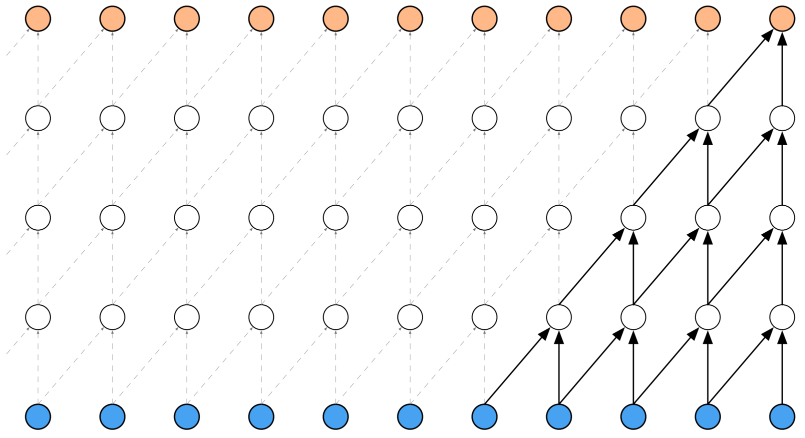


Temporal Variation Modeling (Previous work)



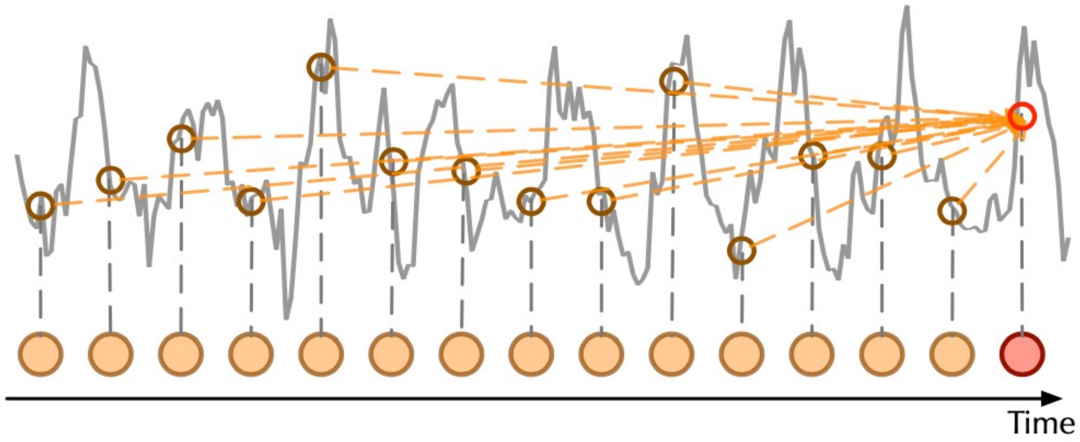
RNNs?

Markov ☹



TCNs?

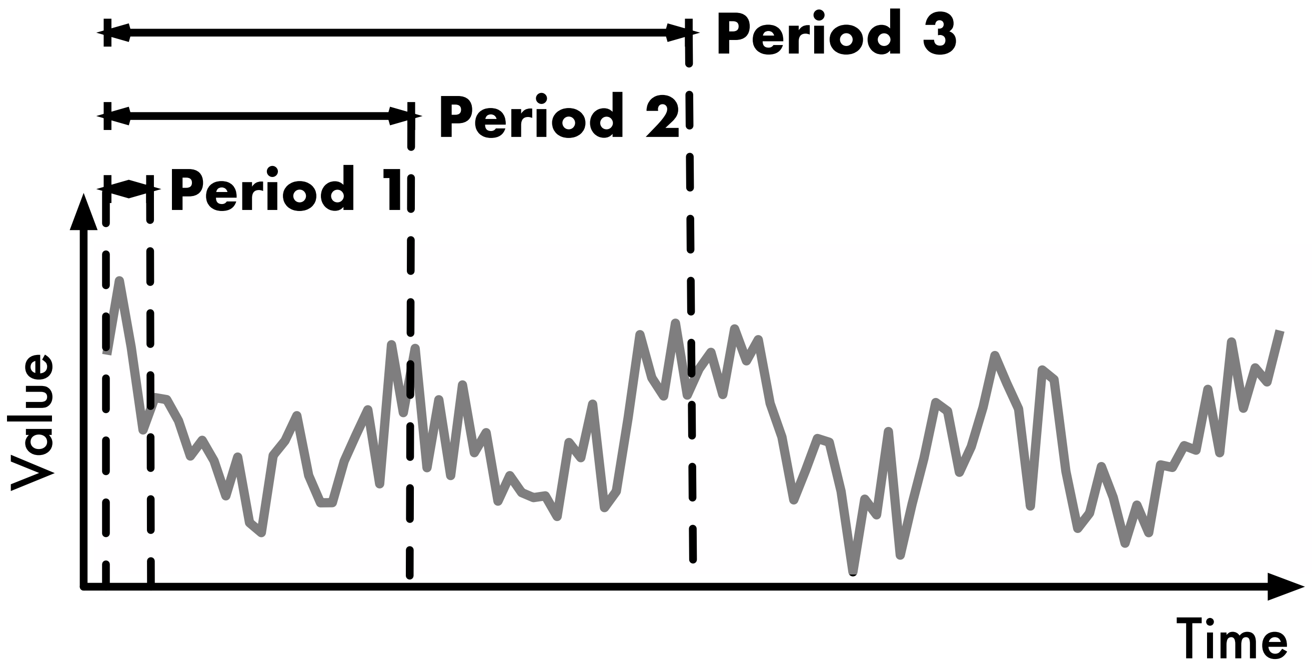
Locality ☹



Transformers?

Temporal dependencies can be obscured deeply in intricate temporal patterns ☹

Multi-periodicity View of Time Series



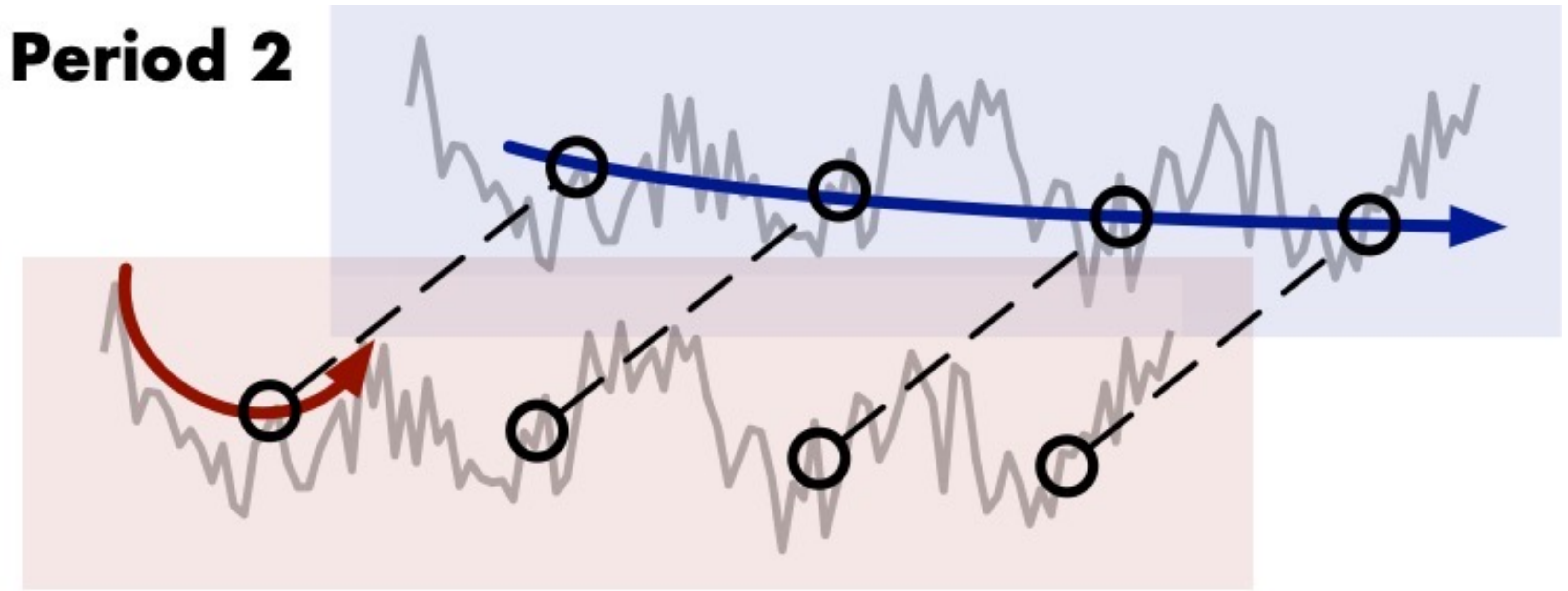
✓ Traffic: daily and weekly

✓ Weather: daily and yearly

Real-world time series usually present multi-periodicity.

Multiple periods overlap and interact with each other.

Intraperiod- and Interperiod-variations

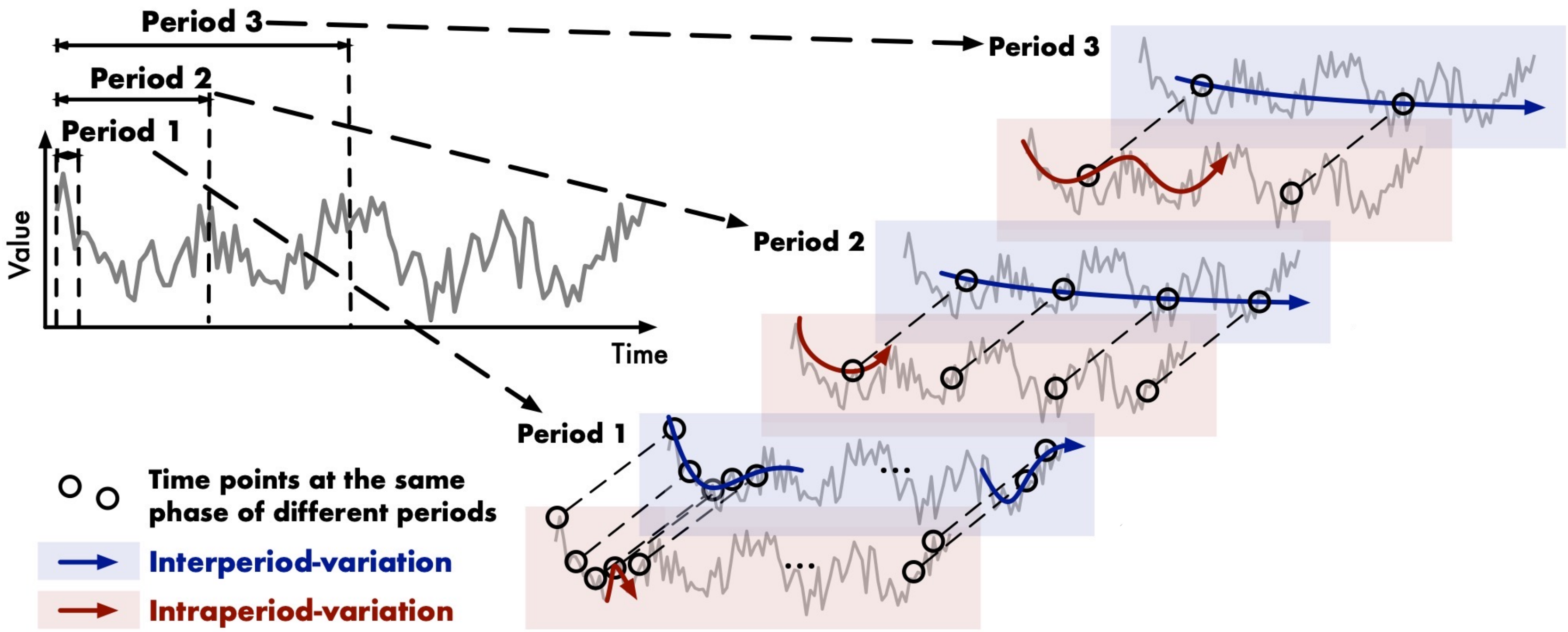


✓ **Intraperiod**: adjacent area, **short-term variations**

✓ **Interperiod**: same phase in adjacent periods, **long-term variations**

Non-periodic cases, the variations will be dominated by intraperiod-variations.

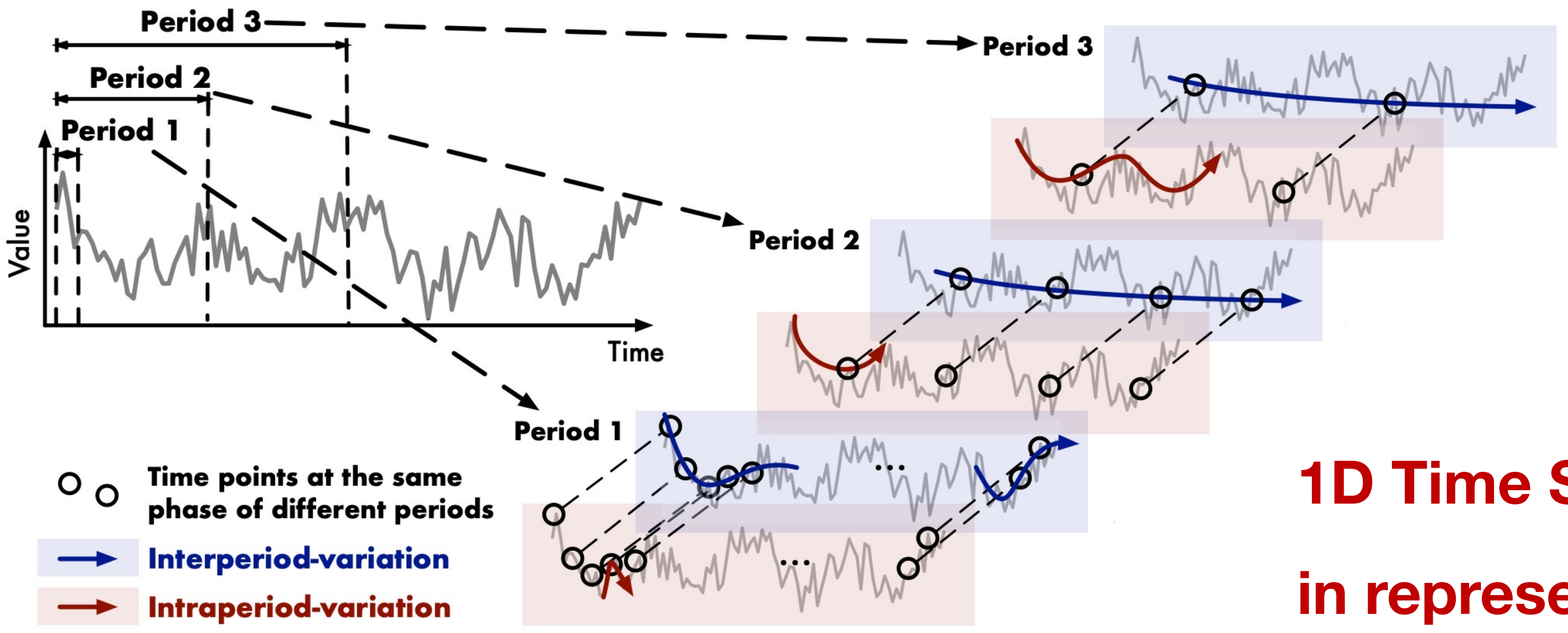
Overall design



① Multi-periodicity

A modular architecture to disentangle intricate temporal patterns

Overall design

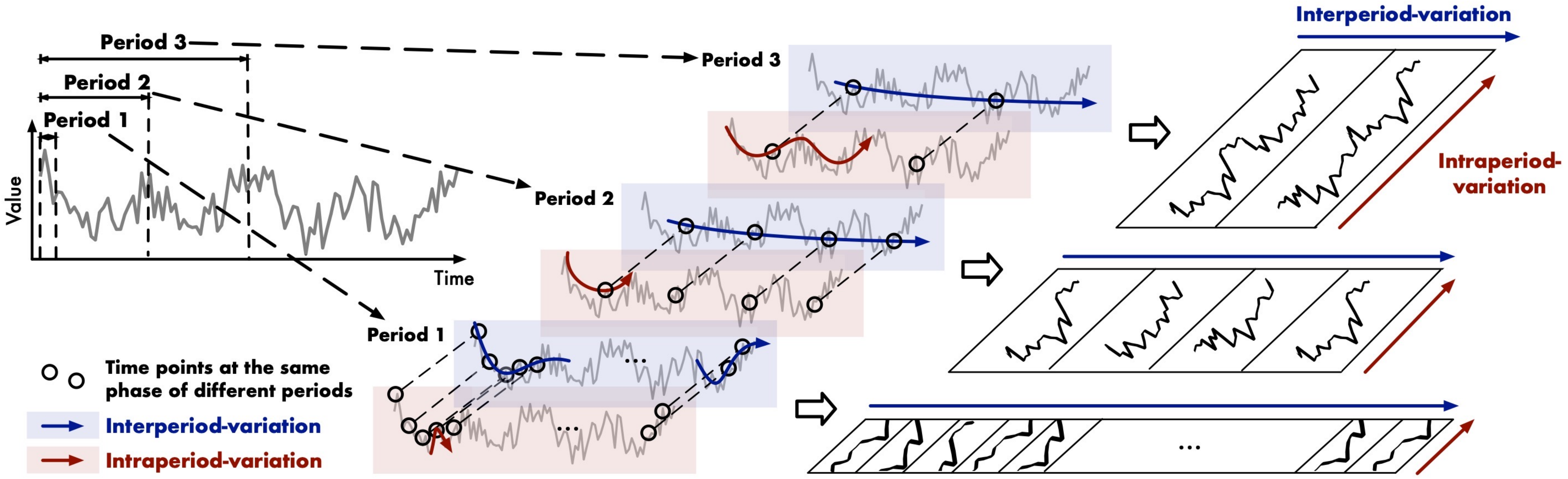


1D Time Series has limitations in representation capability.

① Multi-periodicity

A modular architecture to disentangle intricate temporal patterns

Overall design

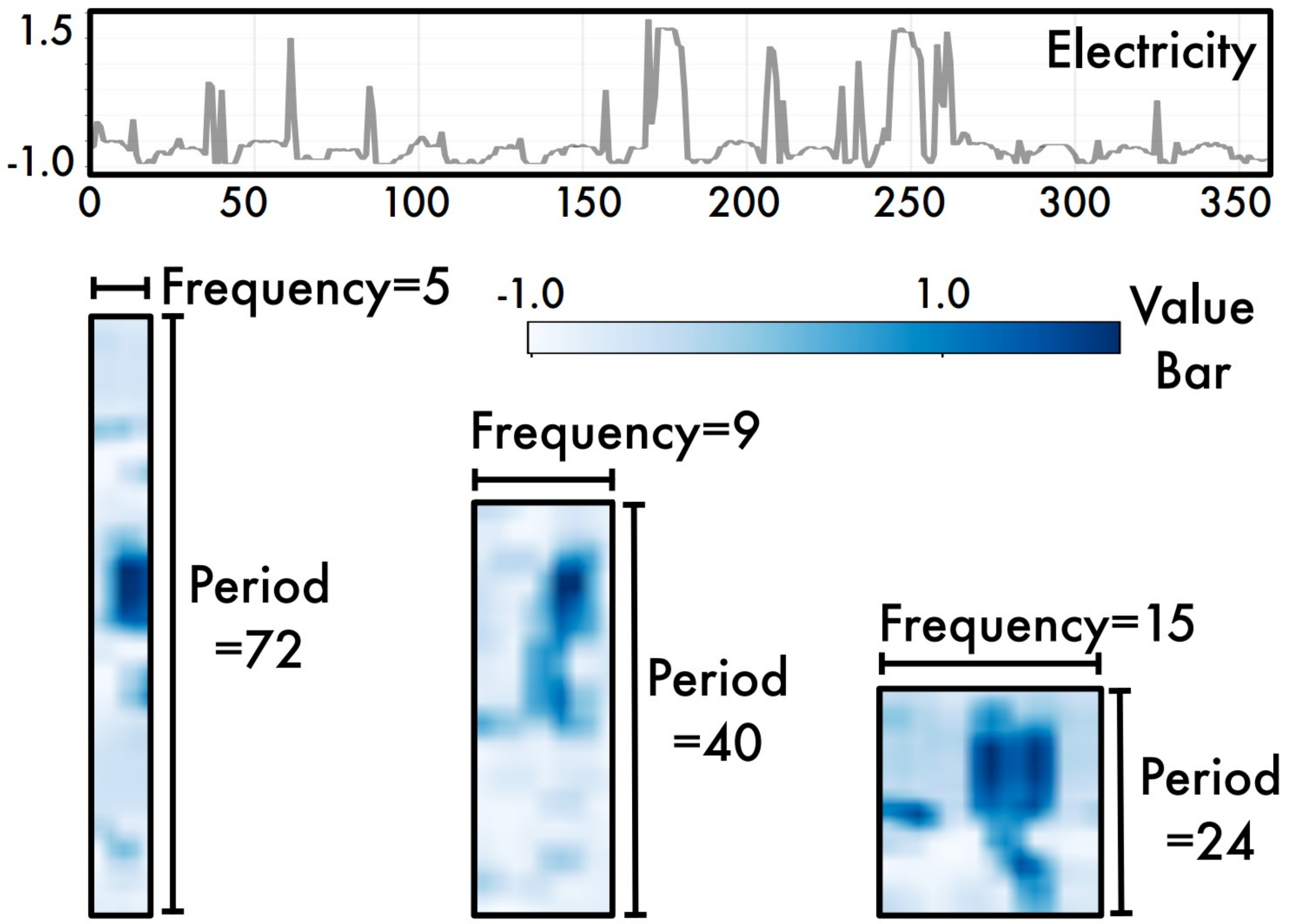


① Multi-periodicity ② Temporal 2D-variation

Unify intraperiod- and interperiod-variations in 2D space by reshape

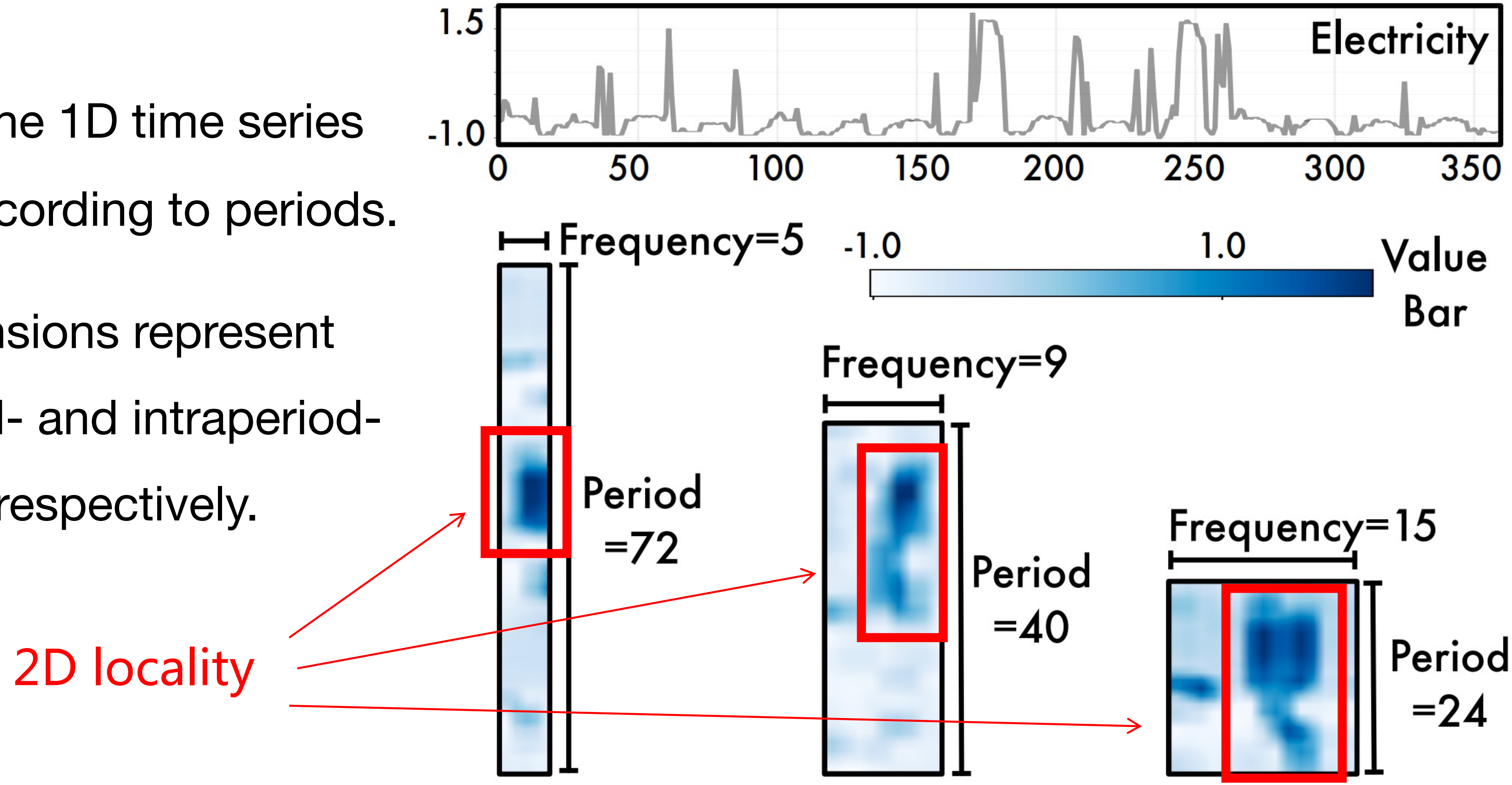
Temporal 2D-variation: A Case Study

- ✓ Reshape the 1D time series into 2D according to periods.
- ✓ Two dimensions represent interperiod- and intraperiod-variations respectively.

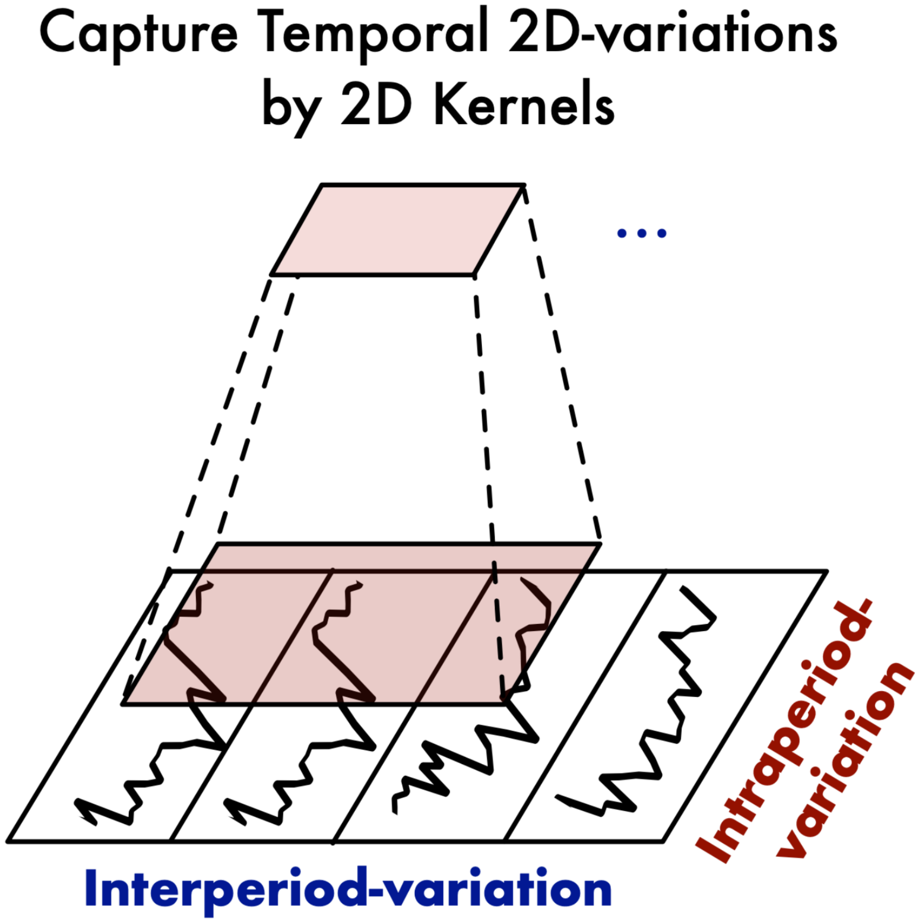


Temporal 2D-variation: A Case Study

- ✓ Reshape the 1D time series into 2D according to periods.
- ✓ Two dimensions represent interperiod- and intraperiod-variations respectively.



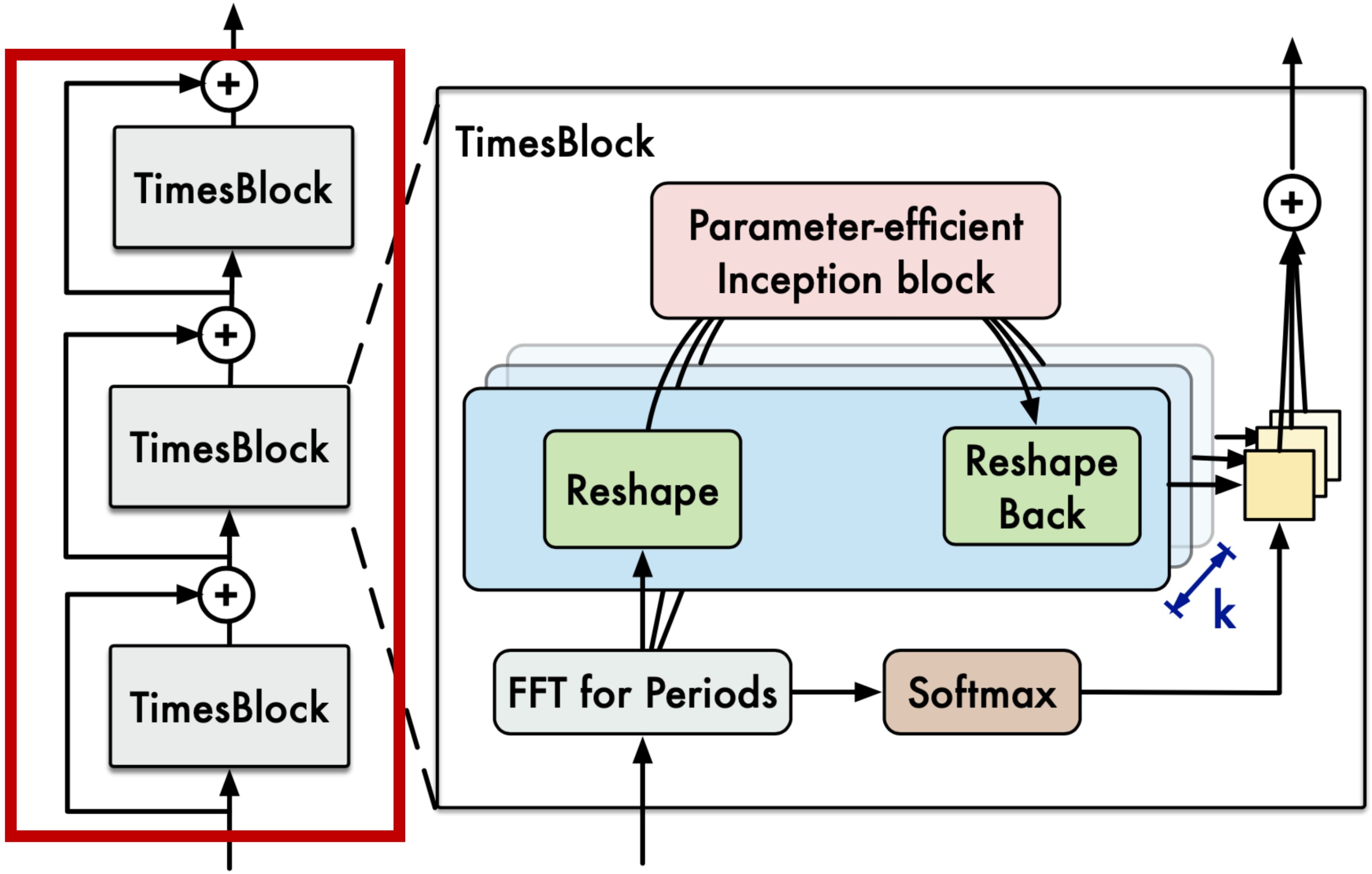
Temporal 2D-variation: A Case Study



With temporal 2D-variations, we can

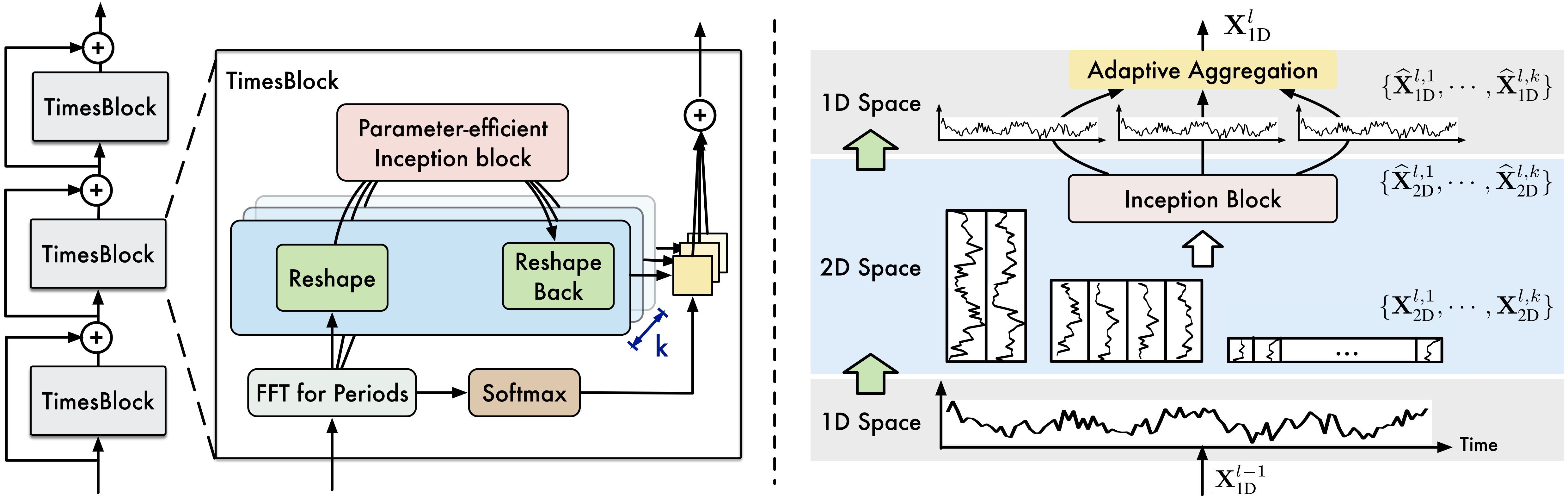
- ✓ Unify intraperiod- interperiod-variations
- ✓ Learn representations by 2D kernels

TimesNet



TimesNet consists of residual-connected **TimesBlocks**.

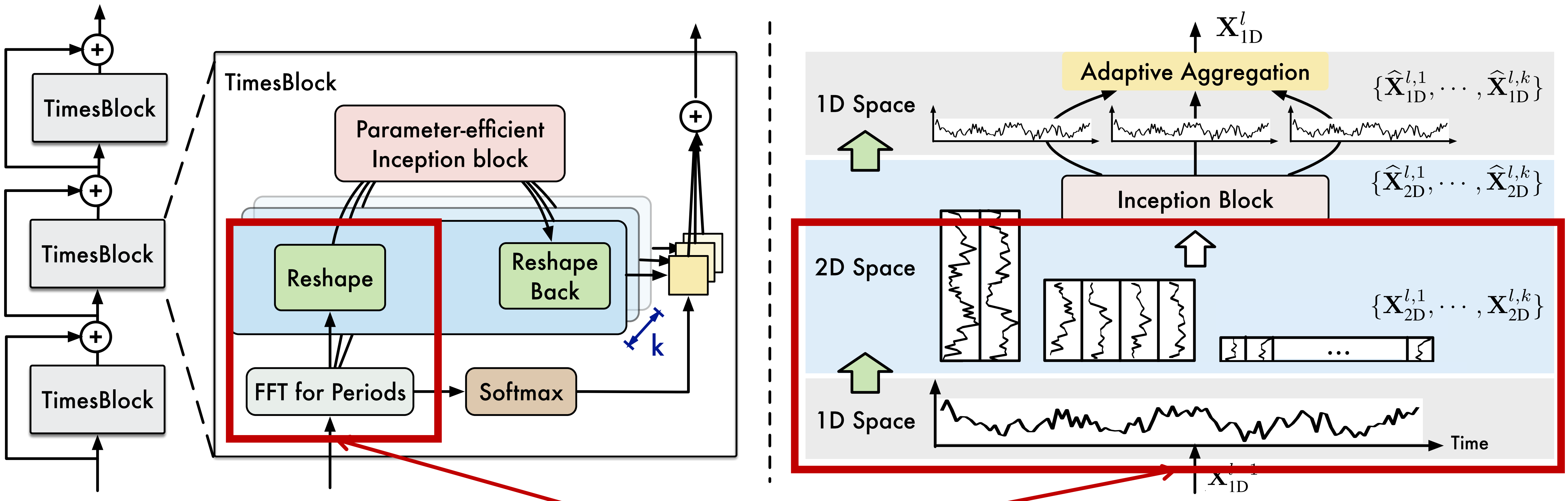
TimesBlock



TimesBlock learns representations in 2D space.

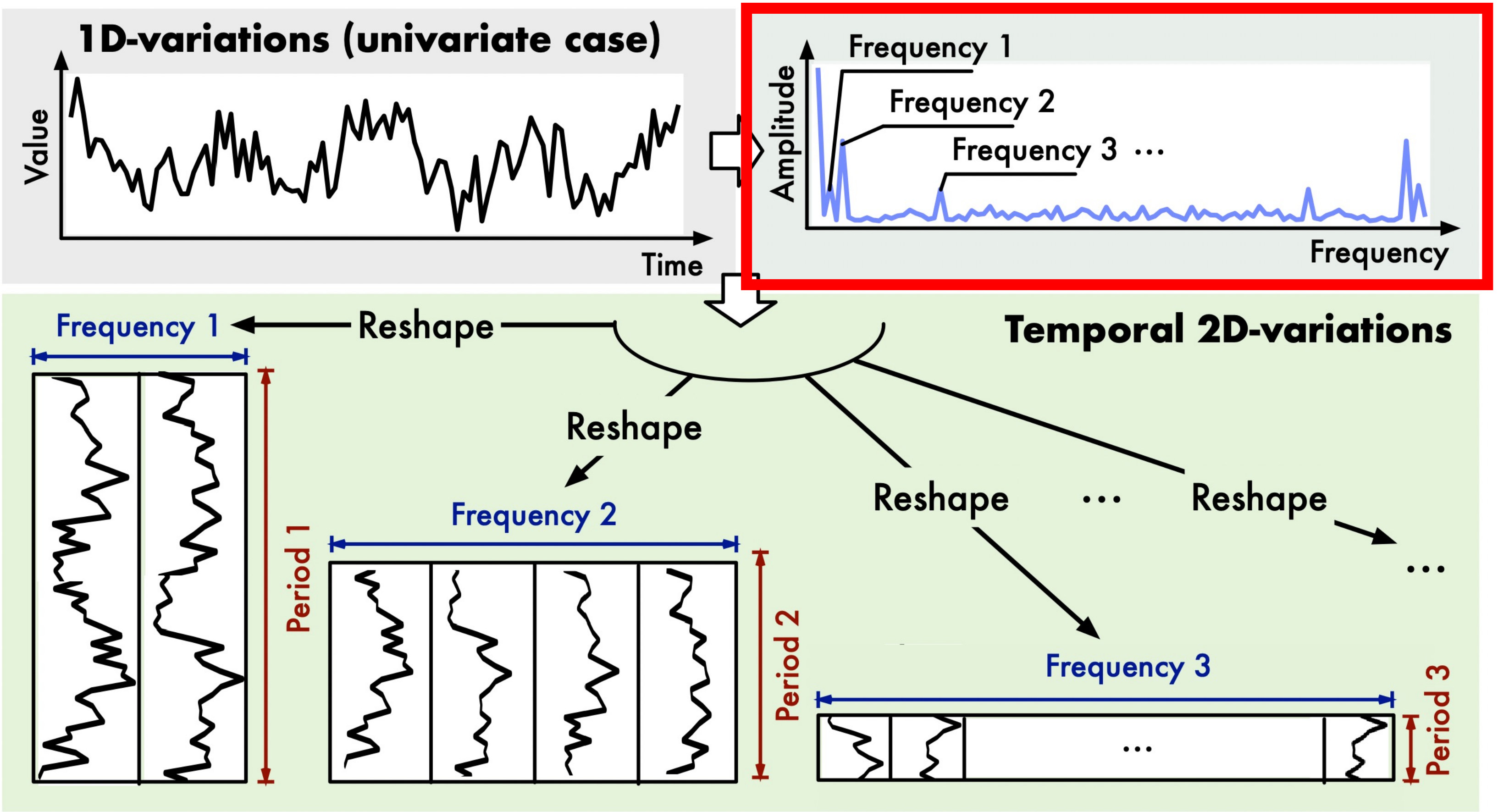
① 1D → 2D ② 2D representation learning ③ 2D → 1D

TimesBlock



① 1D→2D
Obtain temporal 2D-variations

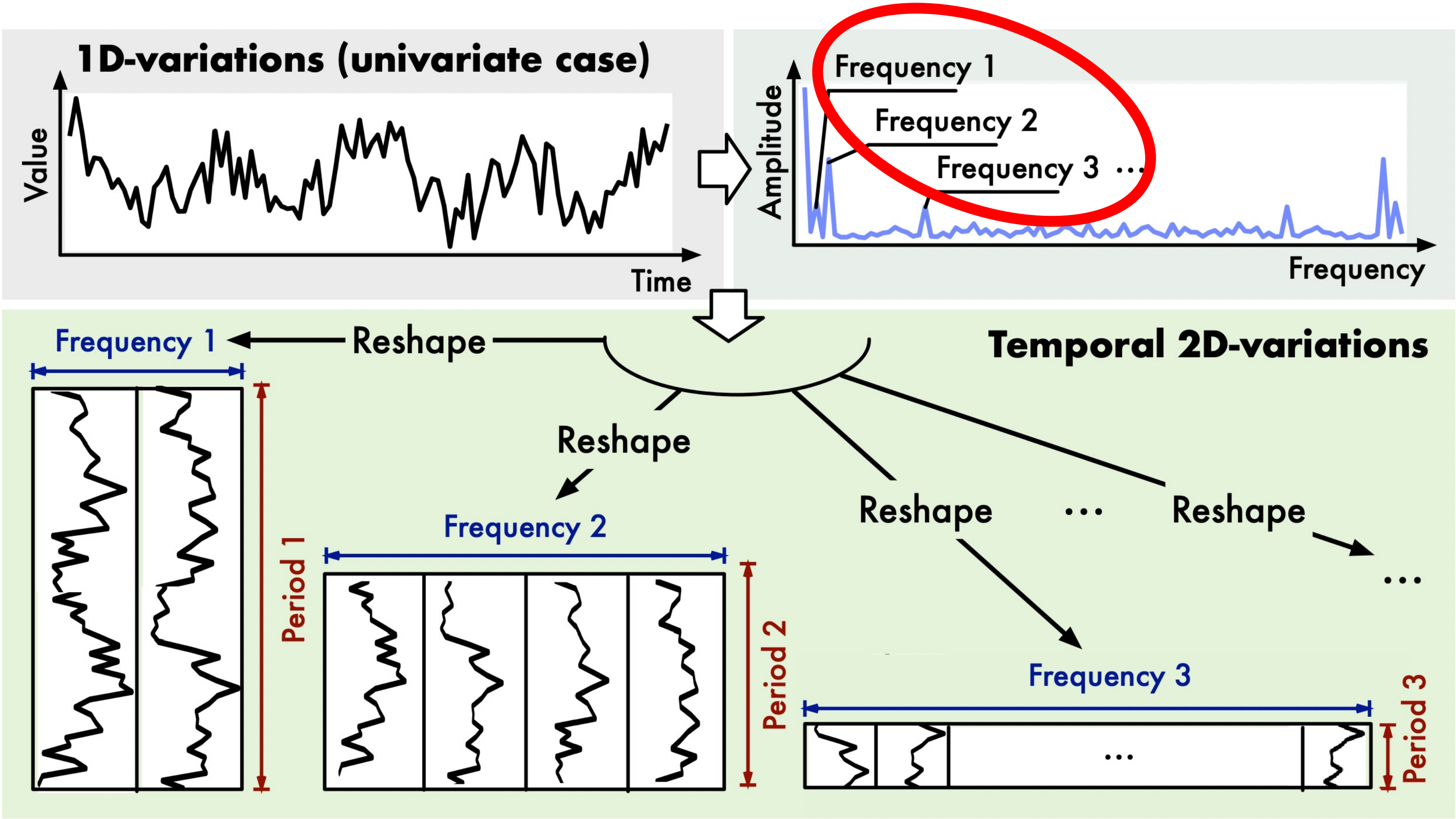
1D → 2D



1. Calculate the spectrum by Fast Fourier Transform

$$\mathbf{A} = \mathrm{Avg}\left(\mathrm{Amp}\left(\mathrm{FFT}(\mathbf{X}_{1\mathrm{D}})\right)\right),$$

1D → 2D

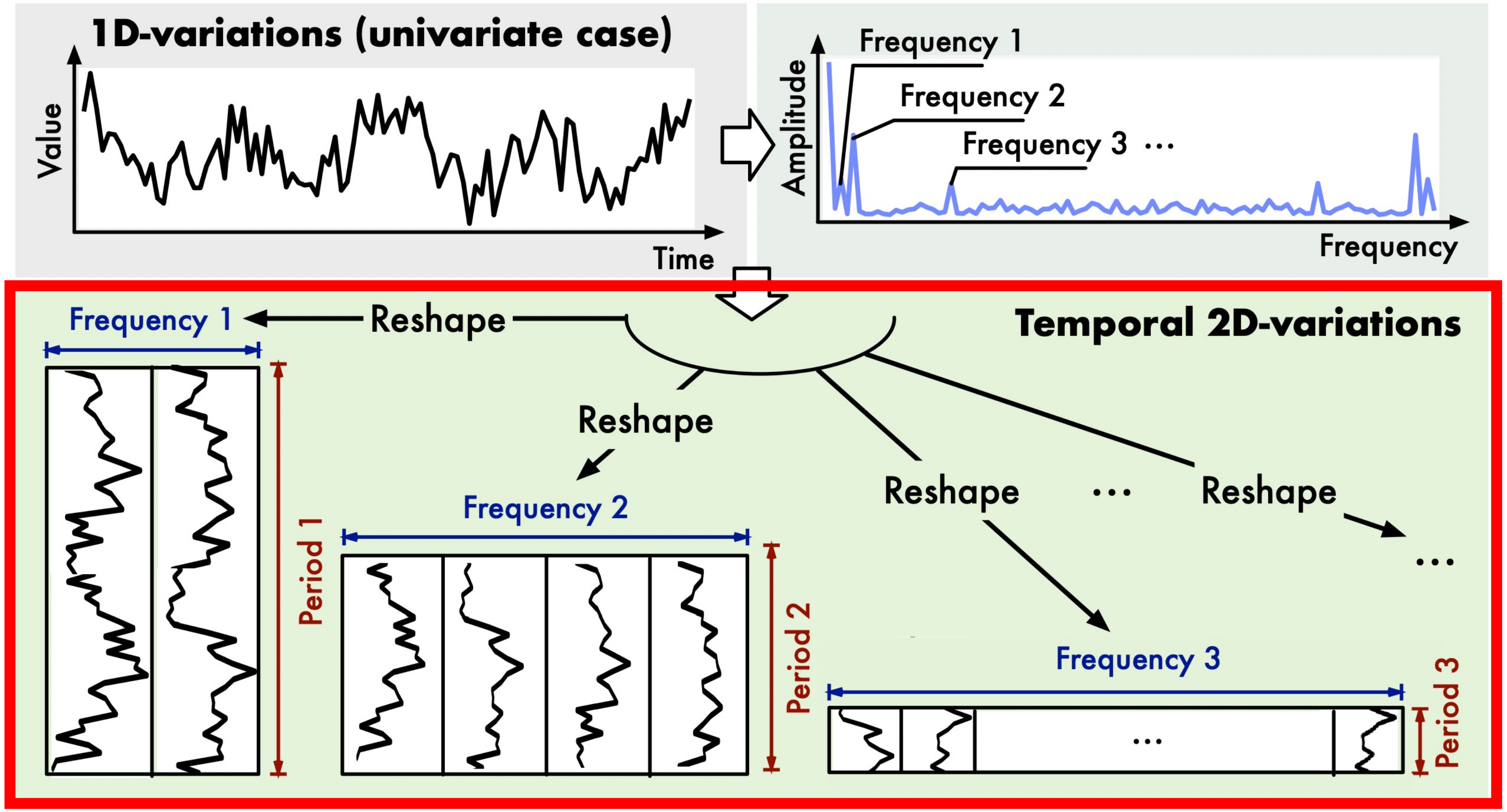


1. Calculate the spectrum by Fast Fourier Transform

2. Choose Topk Frequency

$$\mathbf{A} = \mathrm{Avg}\bigg(\mathrm{Amp}\left(\mathrm{FFT}(\mathbf{X}_{1\mathrm{D}})\right)\bigg),\ \{f_1, \cdots, f_k\} = \underset{f_* \in \{1, \cdots, [\frac{T}{2}]\}}{\arg\mathrm{Topk}} (\mathbf{A}),\ p_i = \left\lceil \frac{T}{f_i} \right\rceil, i \in \{1, \cdots, k\}.$$

1D → 2D

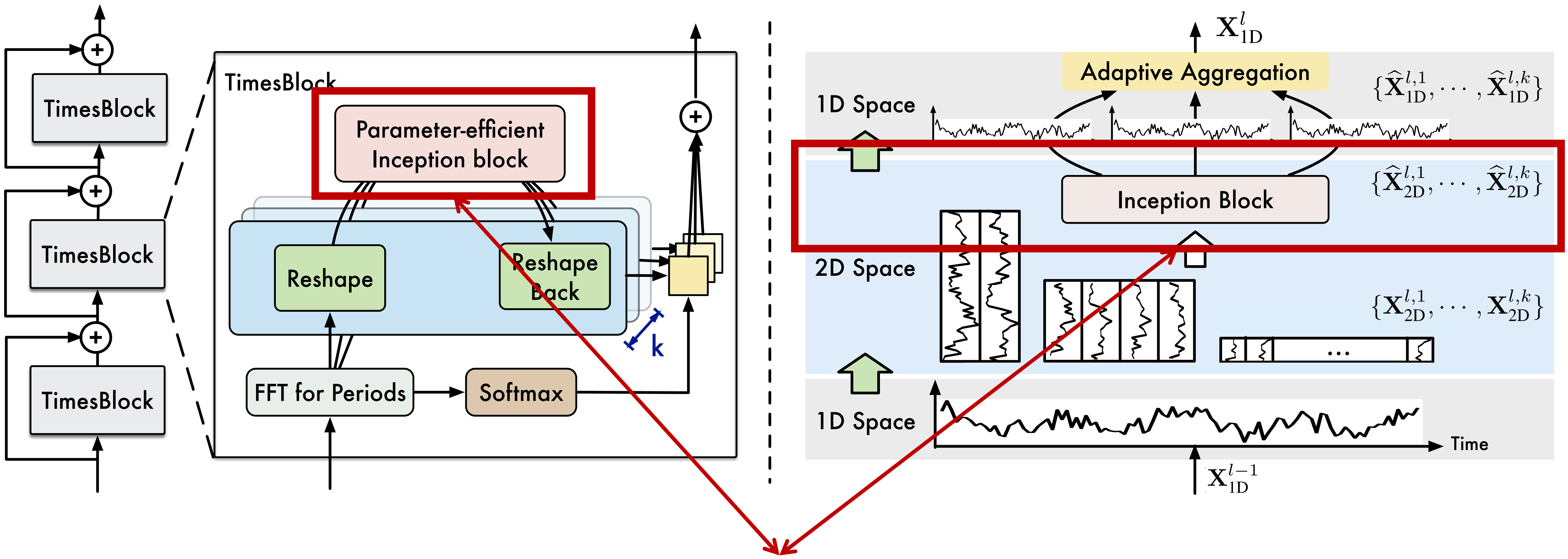


1. Calculate the spectrum by Fast Fourier Transform

2. Choose Topk Frequency

3. For each frequency, reshape 1D time series into 2D tensor

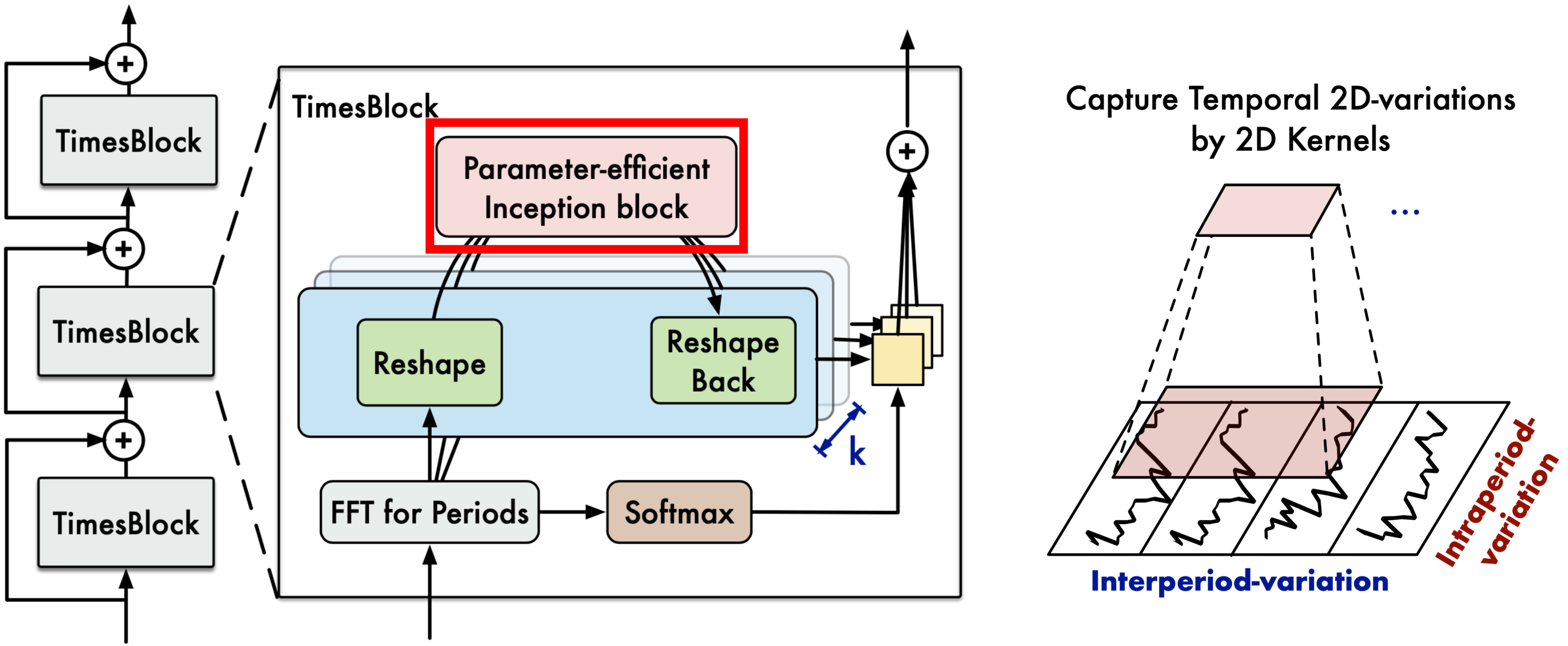
TimesBlock



② 2D representation learning

Extract temporal 2D-variations by 2D kernels

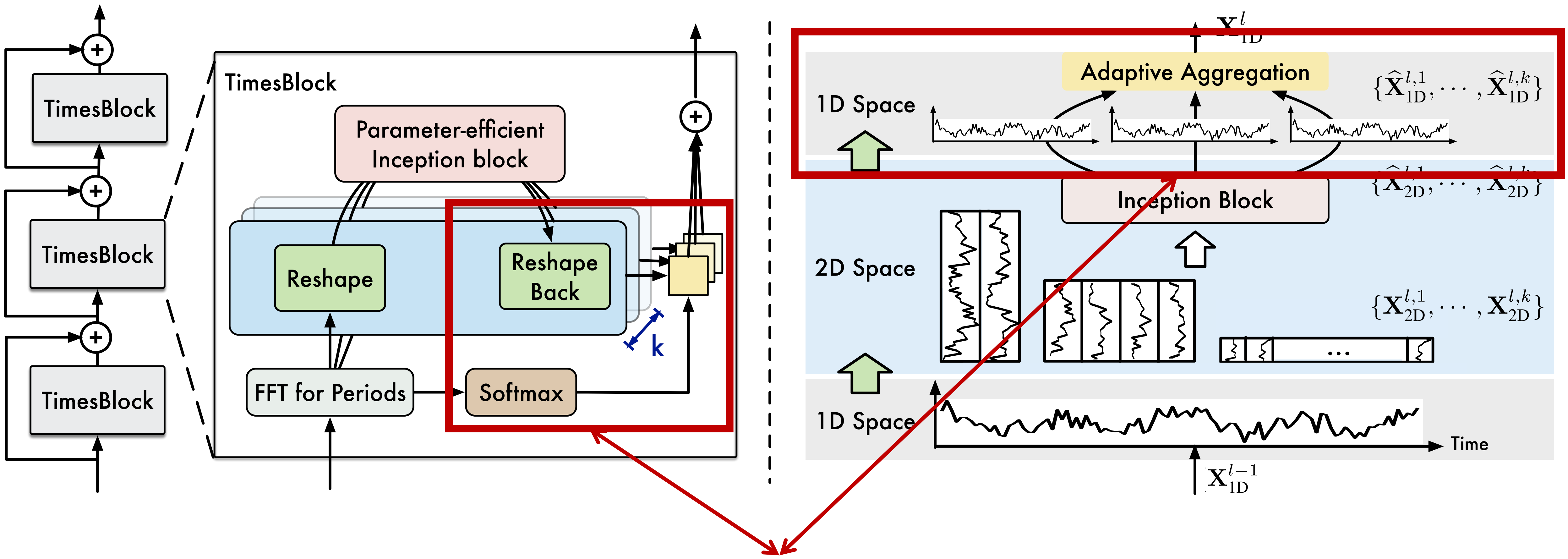
2D Representation Learning



✓ Inception block is shared in all selected periods for parameter efficiency.

✓ It can be replaced by any vision backbones, bridging time series and CV.

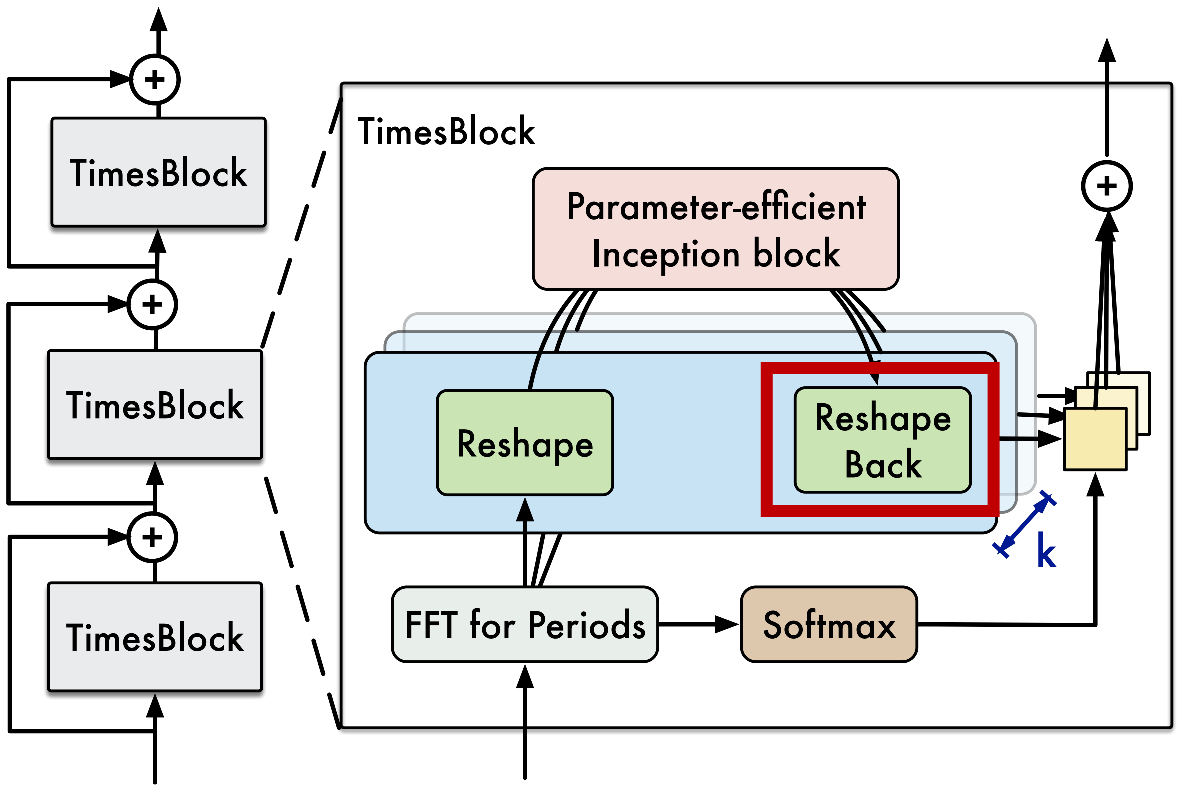
TimesBlock



③ 2D → 1D

Reshape to 1D space to aggregation

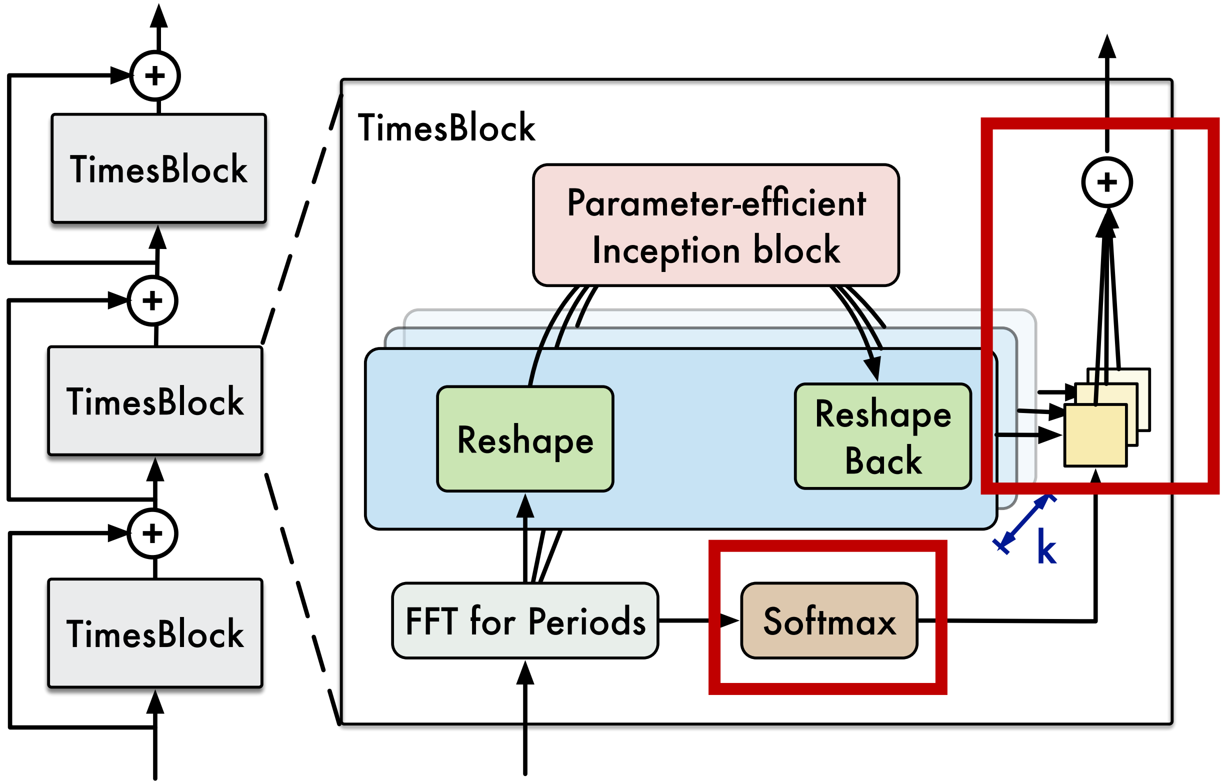
2D → 1D



✓ Reshape back to 1D space for k different representations:

$$\{\widehat{\mathbf{X}}^{l,1}_{\mathrm{1D}}, \cdots, \widehat{\mathbf{X}}^{l,k}_{\mathrm{1D}}\}$$

2D → 1D



✓ Reshape back to 1D space for k different representations.

$$\{\widehat{\mathbf{X}}^{l,1}_{\mathrm{1D}}, \cdots, \widehat{\mathbf{X}}^{l,k}_{\mathrm{1D}}\}$$

✓ Aggregate k different representations based on the periods' amplitudes $\mathbf{A}$:

$$\widehat{\mathbf{A}}^{l-1}_{f_1}, \cdots, \widehat{\mathbf{A}}^{l-1}_{f_k} = \mathrm{Softmax}\left(\mathbf{A}^{l-1}_{f_1}, \cdots, \mathbf{A}^{l-1}_{f_k}\right)$$

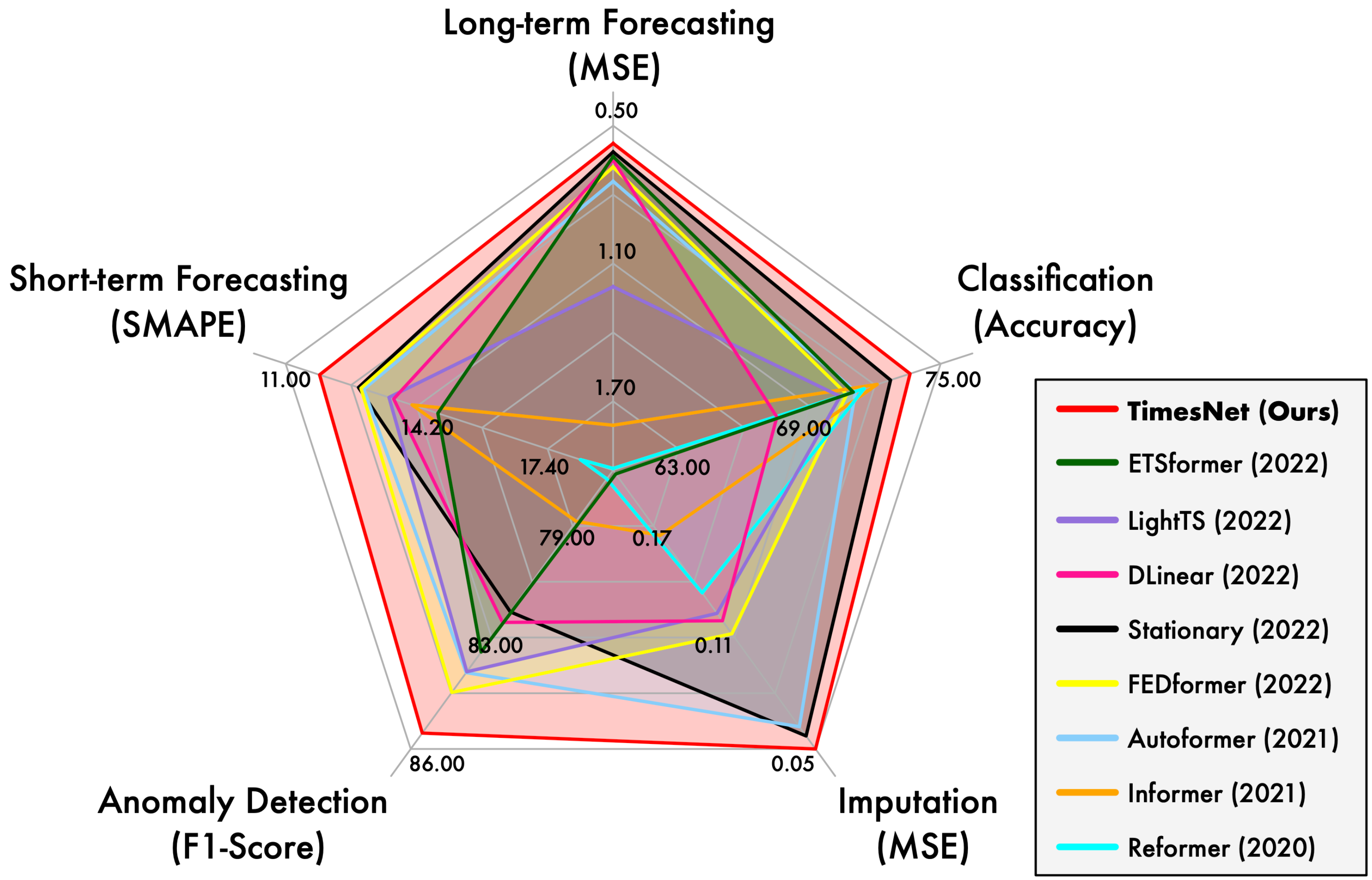
$$\mathbf{X}^{l}_{\mathrm{1D}} = \sum_{i=1}^{k} \widehat{\mathbf{A}}^{l-1}_{f_i} \times \widehat{\mathbf{X}}^{l,i}_{\mathrm{1D}}.$$

Experiment: Overall

Tasks	Benchmarks
Forecasting	**Long-term**: ETT (4 subsets), Electricity, Traffic, Weather, Exchange, ILI
	Short-term: M4 (6 subsets)
Imputation	ETT (4 subsets), Electricity, Weather
Classification	UEA (10 subsets)
Anomaly Detection	SMD, MSL, SMAP, SWaT, PSM

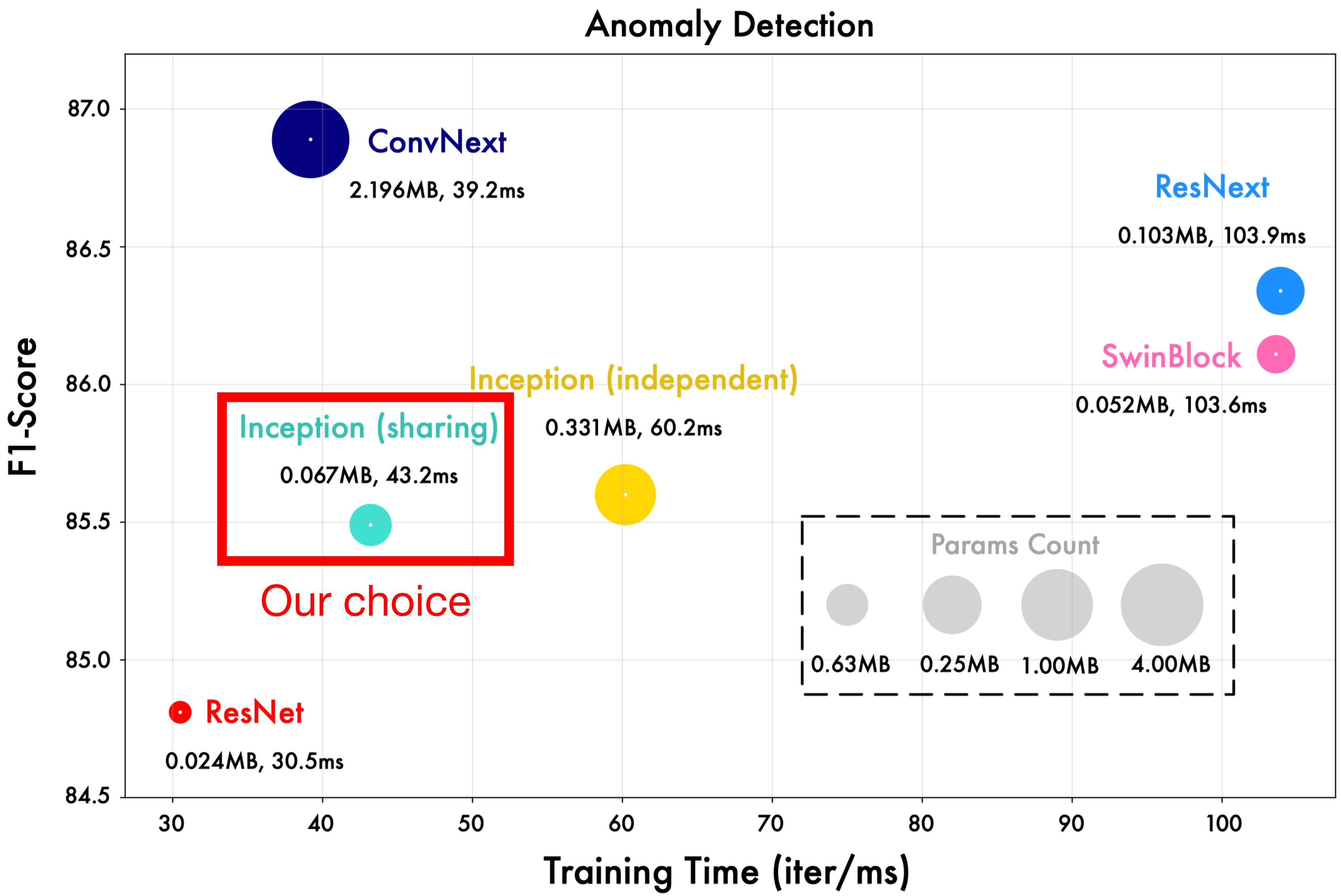
- ✓ Five mainstream time series analysis tasks.
- ✓ 36 datasets, 81 settings, 20+ baselines

Experiment: Overall



TimesNet achieves state-of-the-art in all five tasks!

Model Generality



Better vision backbones,

Better performance 🏆

Bridge Time Series and

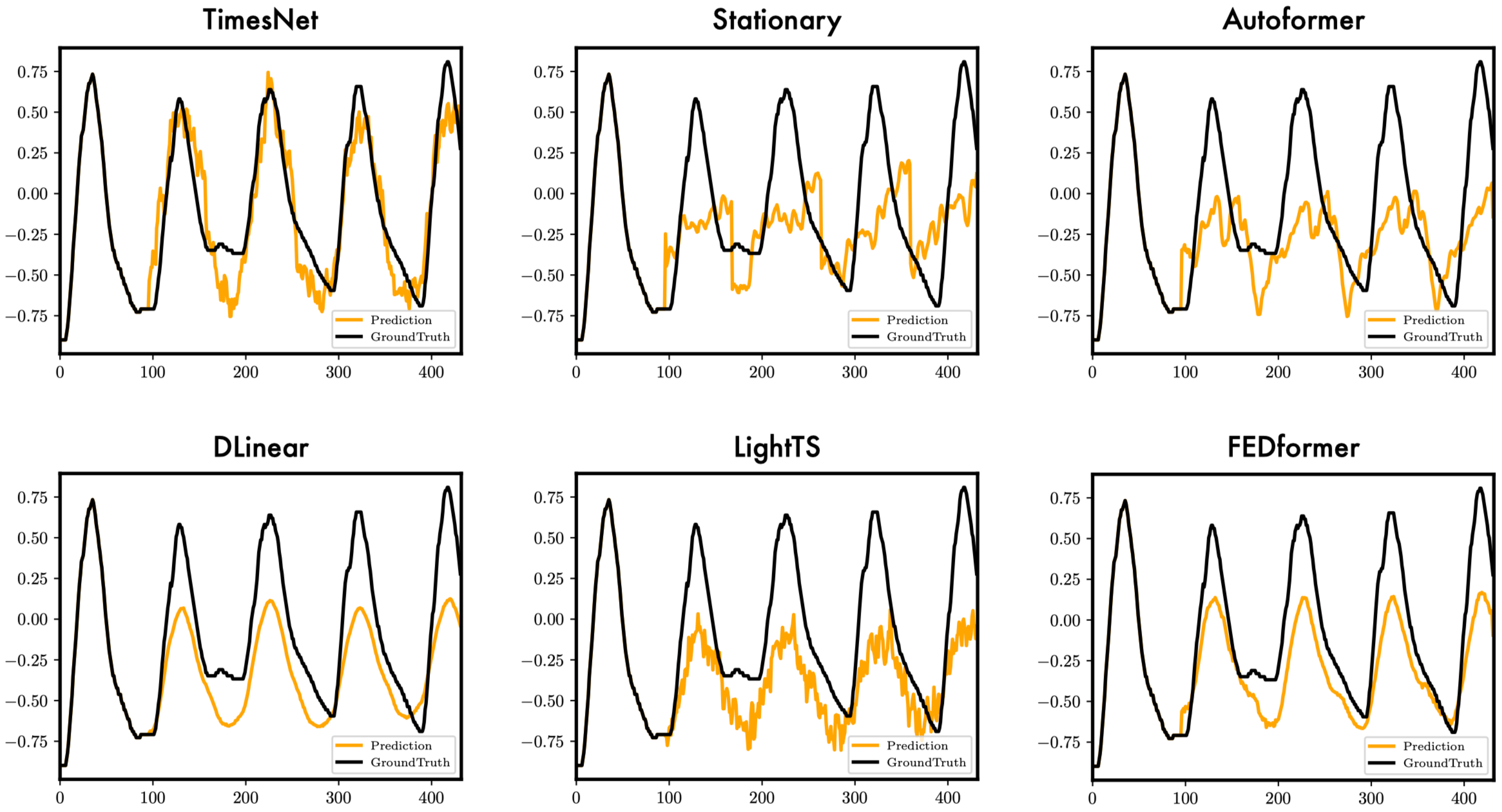
vision backbones 🏆

Experiment: long-term forecasting

Models	TimesNet (Ours)		ETSformer (2022)		LightTS (2022)		DLinear (2022)		FEDformer (2022)		Stationary (2022a)		Autoformer (2021)		Pyraformer (2021a)		Informer (2021)		LogTrans (2019)		Reformer (2020)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	**0.400**	**0.406**	.429	0.425	0.435	0.437	0.403	0.407	0.448	0.452	0.481	0.456	0.588	0.517	0.691	0.607	0.961	0.734	0.929	0.725	0.799	0.671
ETTm2	**0.291**	**0.333**	.293	0.342	0.409	0.436	0.350	0.401	0.305	0.349	0.306	0.347	0.327	0.371	1.498	0.869	1.410	0.810	1.535	0.900	1.479	0.915
ETTh1	0.458	**0.450**	.542	0.510	0.491	0.479	0.456	0.452	**0.440**	0.460	0.570	0.537	0.496	0.487	0.827	0.703	1.040	0.795	1.072	0.837	1.029	0.805
ETTh2	**0.414**	**0.427**	.439	0.452	0.602	0.543	0.559	0.515	0.437	0.449	0.526	0.516	0.450	0.459	0.826	0.703	4.431	1.729	2.686	1.494	6.736	2.191
Electricity	**0.192**	**0.295**	.208	0.323	0.229	0.329	0.212	0.300	0.214	0.327	0.193	0.296	0.227	0.338	0.379	0.445	0.311	0.397	0.272	0.370	0.338	0.422
Traffic	0.620	**0.336**	.621	0.396	0.622	0.392	0.625	0.383	**0.610**	0.376	0.624	0.340	0.628	0.379	0.878	0.469	0.764	0.416	0.705	0.395	0.741	0.422
Weather	**0.259**	**0.287**	.271	0.334	0.261	0.312	0.265	0.317	0.309	0.360	0.288	0.314	0.338	0.382	0.946	0.717	0.634	0.548	0.696	0.602	0.803	0.656
Exchange	0.416	0.443	.410	0.427	0.385	0.447	**0.354**	**0.414**	0.519	0.500	0.461	0.454	0.613	0.539	1.913	1.159	1.550	0.998	1.402	0.968	1.280	0.932
ILI	2.139	0.931	.497	1.004	7.382	2.003	2.616	1.090	2.847	1.144	**2.077**	**0.914**	3.006	1.161	7.635	2.050	5.137	1.544	4.839	1.485	4.724	1.445

TimesNet surpasses advanced Transformer-based and MLP-based models.

Experiment: long-term forecasting



Experiment: short-term forecasting

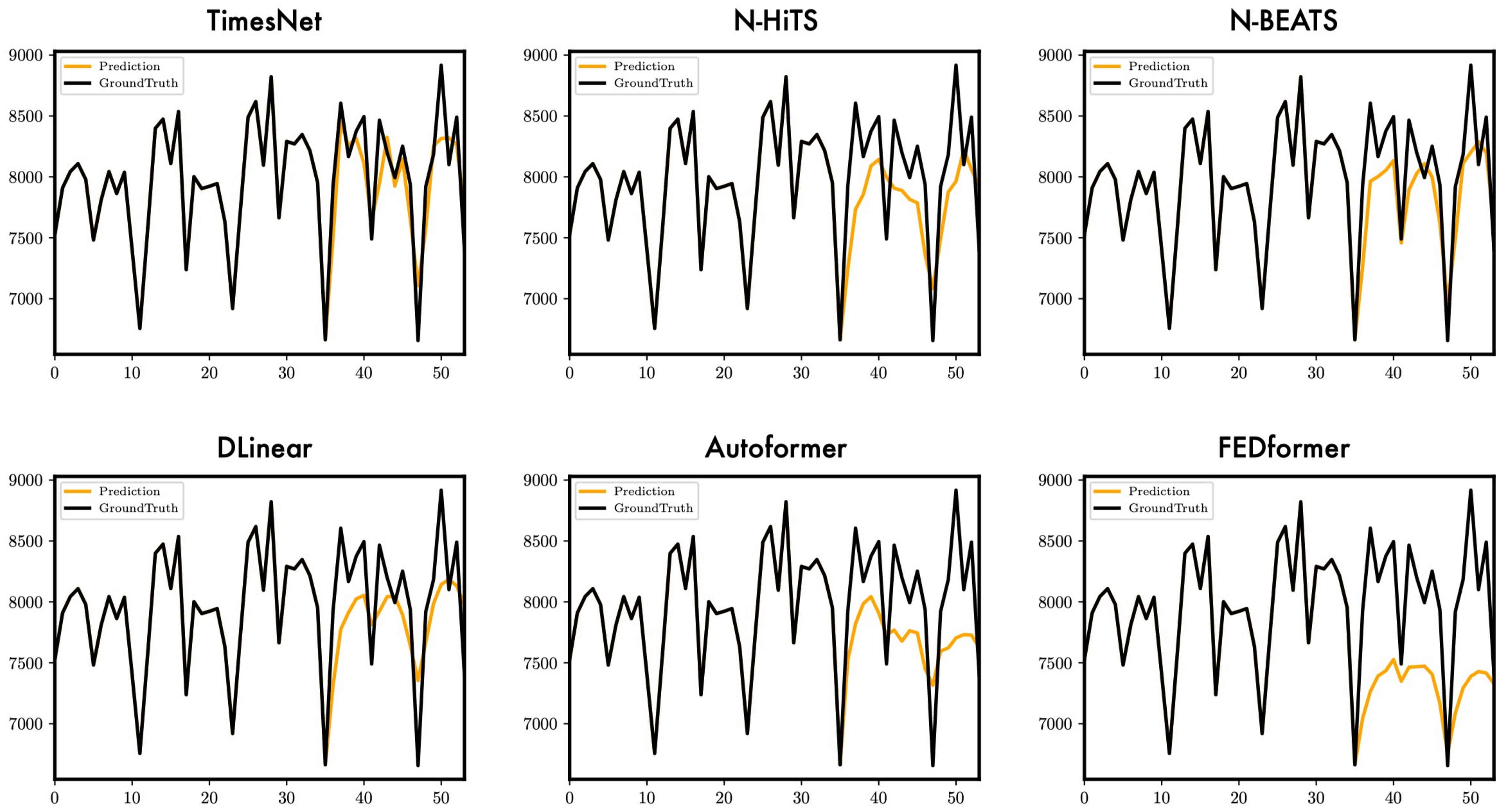
✓ More complex temporal patterns: M4 dataset is composed of yearly, monthly, weekly, daily, hourly and quarterly collected univariate marketing data.

✓ TimesNet surpasses N-HiTs and N-BEATS.

✓ Simple Linear methods degenerate a lot.

Models	**TimesNet (Ours)**	N-HiTS (2022)	N-BEATS (2019)	ETSformer (2022)	LightTS (2022)	DLinear (2022)	FEDformer (2022)	Stationary (2022a)	Autoformer (2021)	Pyraformer (2021a)	Informer (2021)	LogTrans (2019)	Reformer (2020)
SMAPE	**11.829**	11.927	11.851	14.718	13.525	13.639	12.840	12.780	12.909	16.987	14.086	16.018	18.200
MASE	**1.585**	1.613	1.599	2.408	2.111	2.095	1.701	1.756	1.771	3.265	2.718	3.010	4.223
OWA	**0.851**	0.861	0.855	1.172	1.051	1.051	0.918	0.930	0.939	1.480	1.230	1.378	1.775

Experiment: short-term forecasting

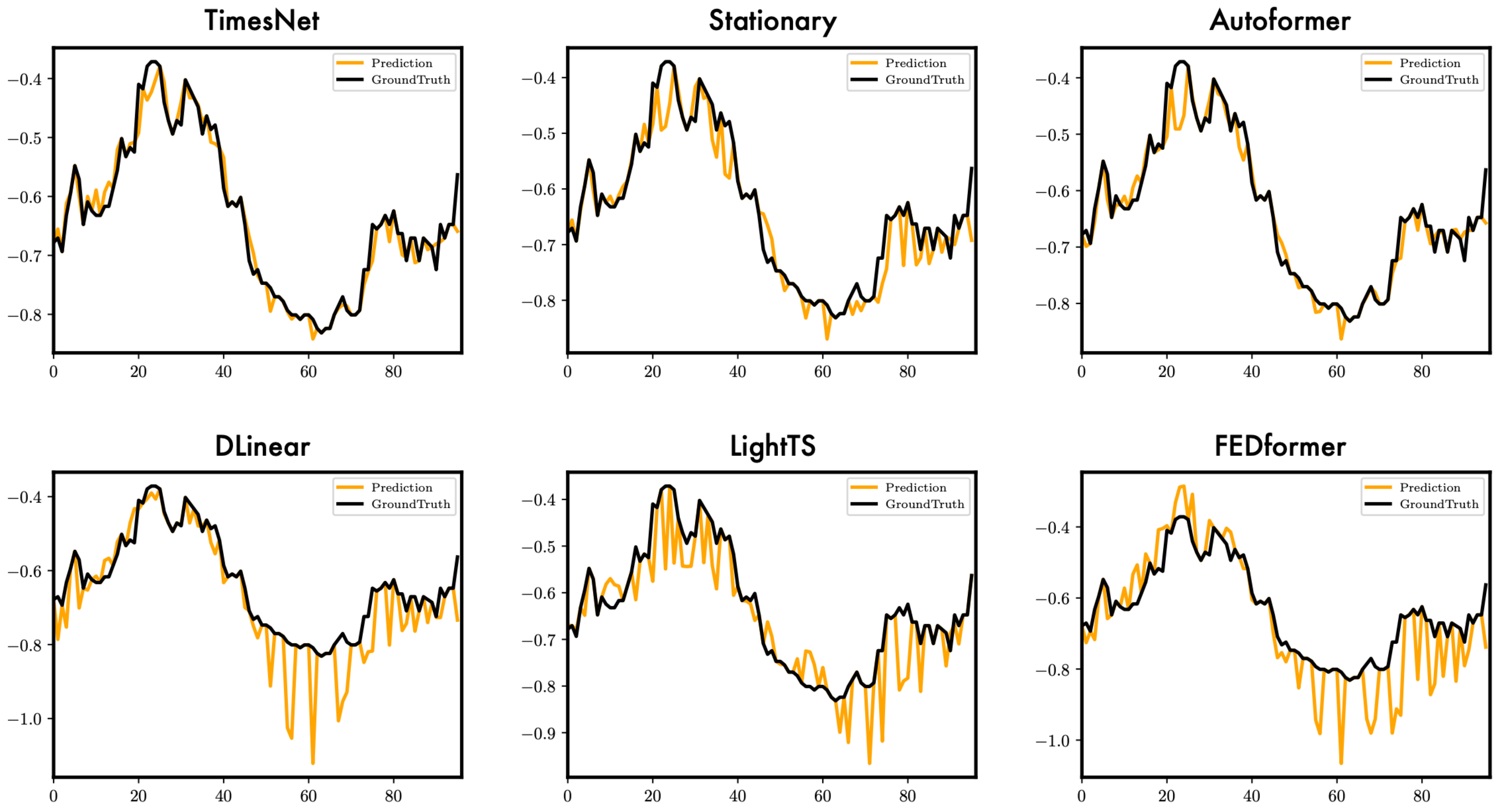


Experiment: imputation

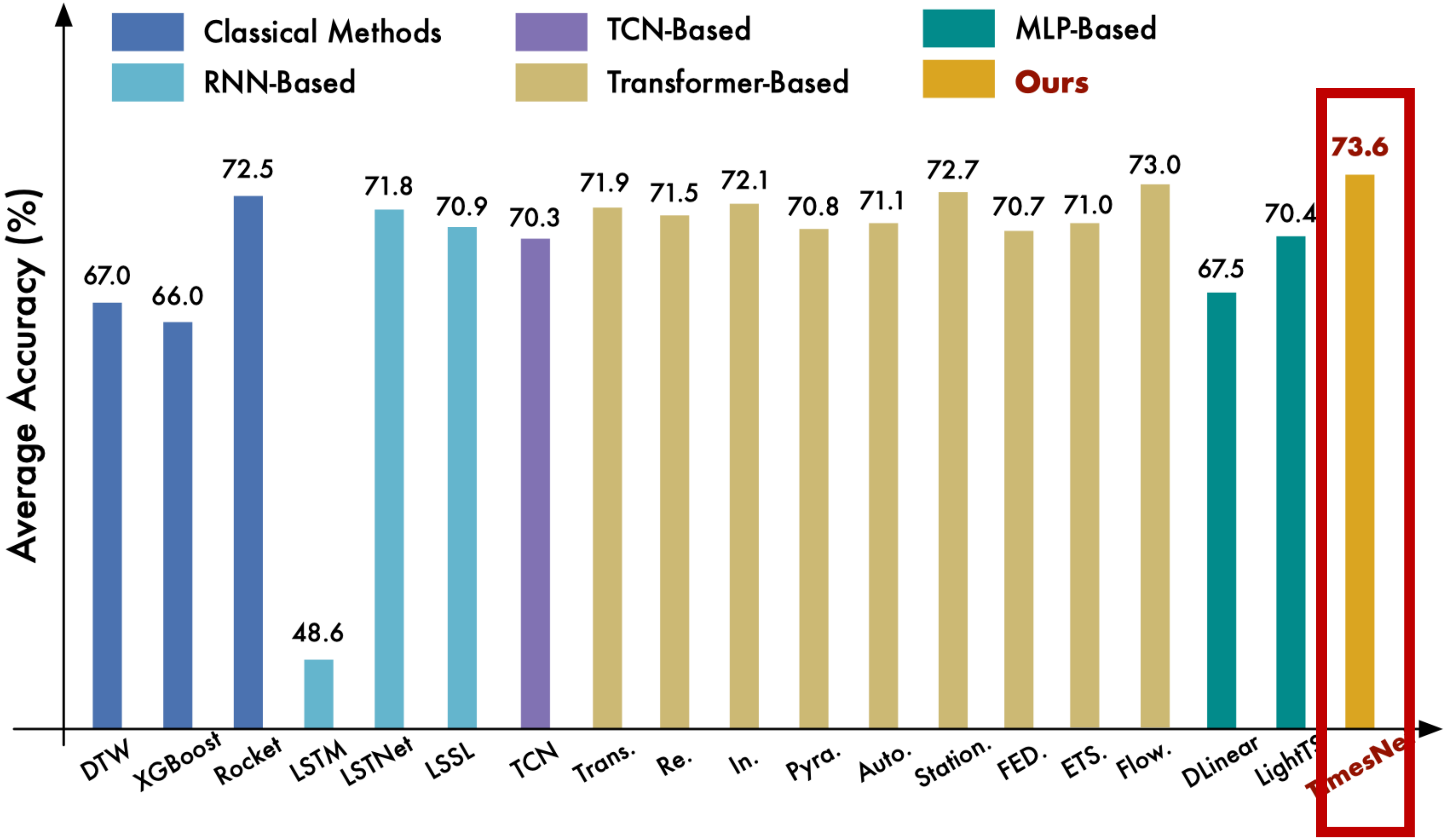
- ✓ Averaged from 4 different mask ratios: 12.5%, 25%, 37.5%, 50%
- ✓ Requires the model to handle irregular inputs.
- ✓ Non-stationary Transformer performs well but MLP-based models fail in this task.

Models	**TimesNet (Ours)**		ETSformer (2022)		LightTS (2022)		DLinear (2023)		FEDformer (2022)		Stationary (2022a)		Autoformer (2021)		Pyraformer (2021a)		Informer (2021)		LogTrans (2019)		Reformer (2020)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	**0.027**	**0.107**	0.120	0.253	0.104	0.218	0.093	0.206	0.062	0.177	0.036	0.126	0.051	0.150	0.717	0.570	0.071	0.188	0.050	0.154	0.055	0.166
ETTm2	**0.022**	**0.088**	0.208	0.327	0.046	0.151	0.096	0.208	0.101	0.215	0.026	0.099	0.029	0.105	0.465	0.508	0.156	0.292	0.119	0.246	0.157	0.280
ETTh1	**0.078**	**0.187**	0.202	0.329	0.284	0.373	0.201	0.306	0.117	0.246	0.094	0.201	0.103	0.214	0.842	0.682	0.161	0.279	0.219	0.332	0.122	0.245
ETTh2	**0.049**	**0.146**	0.367	0.436	0.119	0.250	0.142	0.259	0.163	0.279	0.053	0.152	0.055	0.156	1.079	0.792	0.337	0.452	0.186	0.318	0.234	0.352
Electricity	**0.092**	**0.210**	0.214	0.339	0.131	0.262	0.132	0.260	0.130	0.259	0.100	0.218	0.101	0.225	0.297	0.382	0.222	0.328	0.175	0.303	0.200	0.313
Weather	**0.030**	**0.054**	0.076	0.171	0.055	0.117	0.052	0.110	0.099	0.203	0.032	0.059	0.031	0.057	0.152	0.235	0.045	0.104	0.039	0.076	0.038	0.087

Experiment: imputation



Experiment: classification



✓ TimesNet still achieves the best performance.

✓ Transformer-based models generally outperform MLP-based models

Experiment: anomaly detection

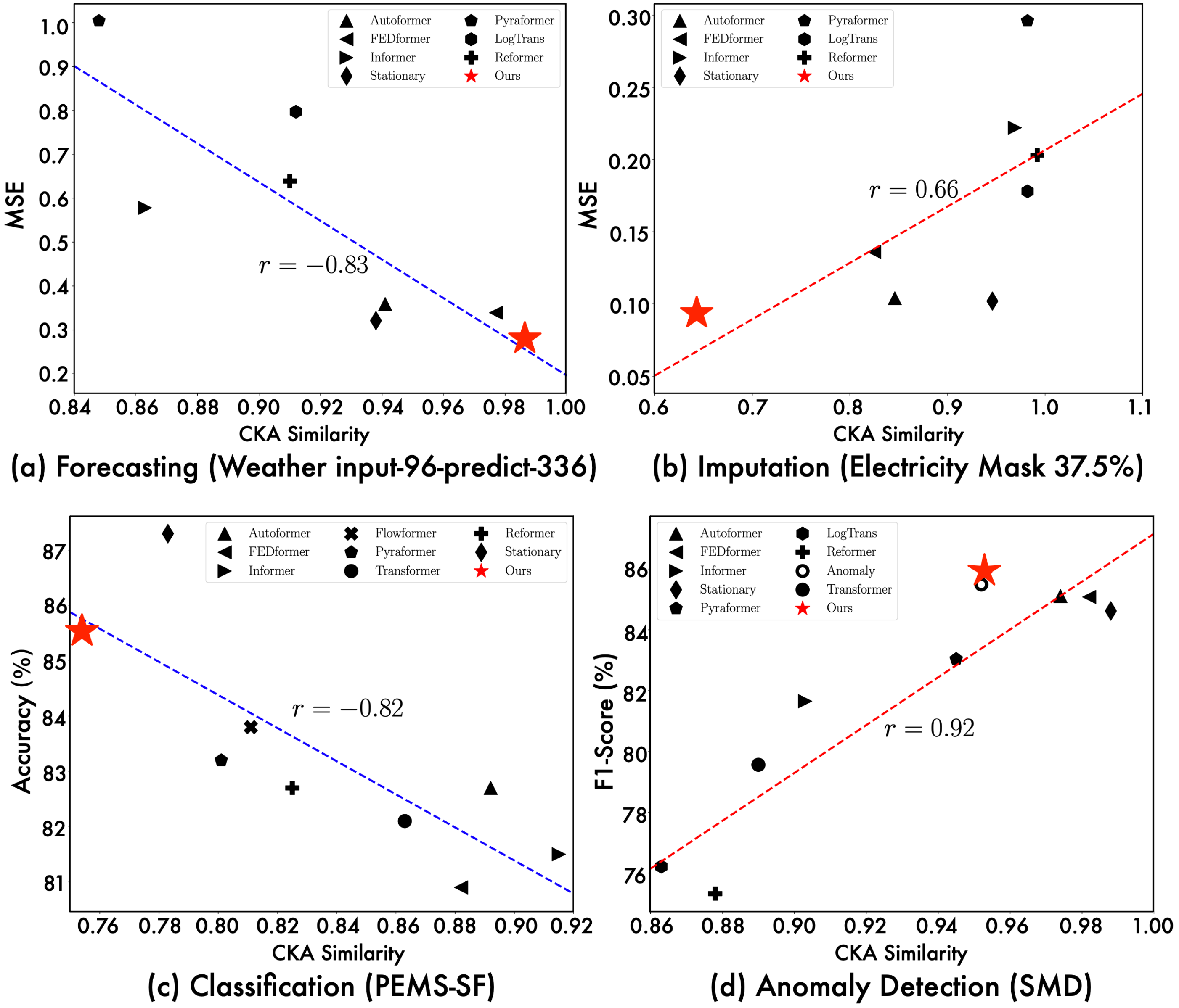
✓ Adopt the reconstruction error as the anomaly criterion.

✓ Better 2D backbones bring better performances.

✓ Transformer-based models performs well.

Models	**TimesNet (ResNeXt)**	**TimesNet (Inception)**	ETS. (2022)	FED. (2022)	LightTS (2022)	DLinear (2023)	Stationary (2022a)	Auto. (2021)	Pyra. (2021a)	Anomaly* (2021)	In. (2021)	Re. (2020)	LogTrans (2019)	Trans. (2017)
SMD	**85.81**	85.12	83.13	85.08	82.53	77.10	84.72	85.11	83.04	85.49	81.65	75.32	76.21	79.56
MSL	**85.15**	84.18	85.03	78.57	78.95	84.88	77.50	79.05	84.86	83.31	84.06	84.40	79.57	78.68
SMAP	**71.52**	70.85	69.50	70.76	69.21	69.26	71.09	71.12	71.09	71.18	69.92	70.40	69.97	69.70
SWaT	91.74	92.10	84.91	93.19	**93.33**	87.52	79.88	92.74	91.78	83.10	81.43	82.80	80.52	80.37
PSM	**97.47**	95.21	91.76	97.23	97.15	93.55	97.29	93.29	82.08	79.40	77.10	73.61	76.74	76.07
Avg F1	**86.34**	85.49	82.87	84.97	84.23	82.46	82.08	84.26	82.57	80.50	78.83	77.31	76.60	76.88

Representation Analysis

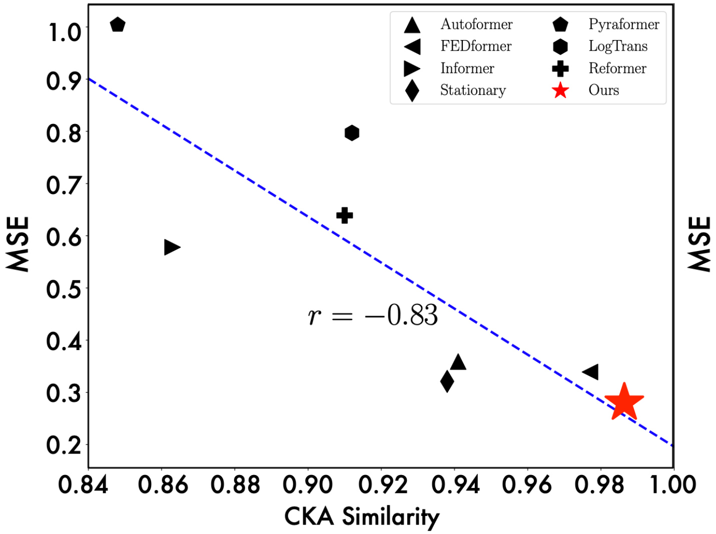


Relation between top-bottom layer CKA similarity and performance

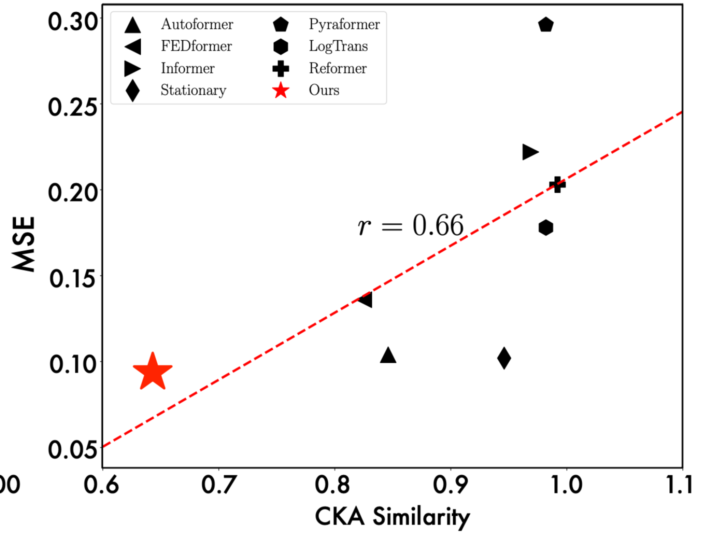
✓ **Why TimesNet achieves SOTA?**

Benefiting from temporal 2D-variations, it can learn proper representations for different tasks.

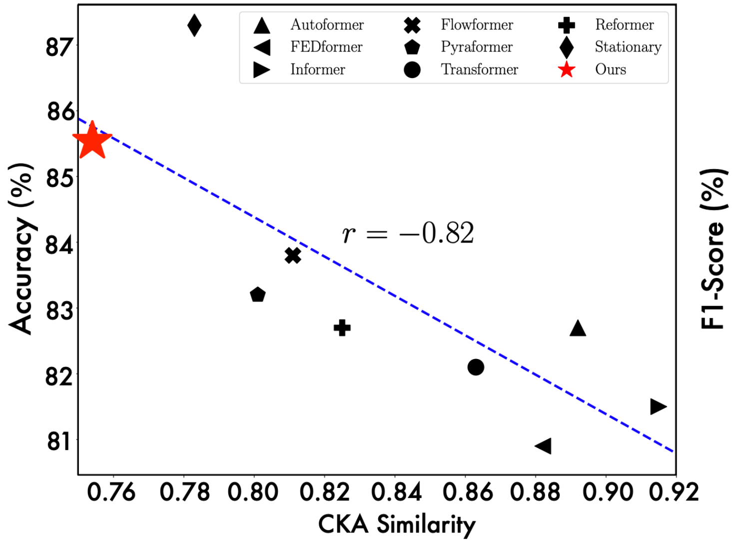
Representation Analysis



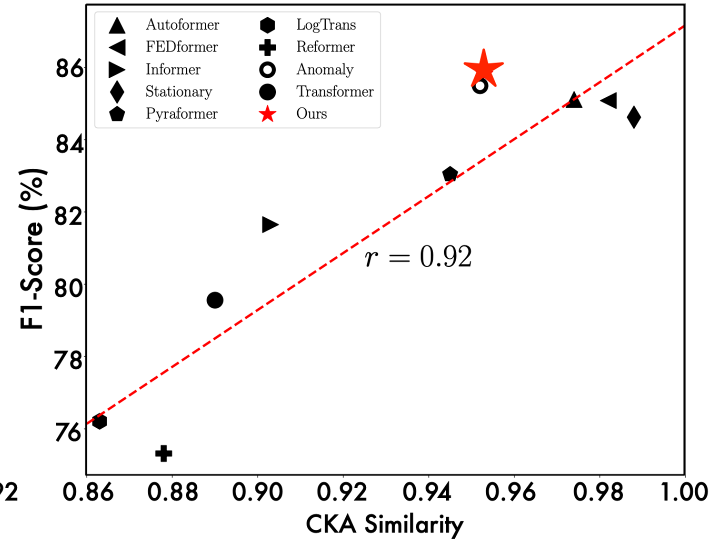
(a) Forecasting (Weather input-96-predict-336)



(b) Imputation (Electricity Mask 37.5%)



(c) Classification (PEMS-SF)



(d) Anomaly Detection (SMD)

Relation between top-bottom layer CKA similarity and performance

✓ **What is the design principle?**

- Classification & imputation need hierarchical representations.

- Anomaly detection & Forecasting expect low-level representations.

Model Performance Ranking

✓ After comparing more than 20+ baselines, we get:

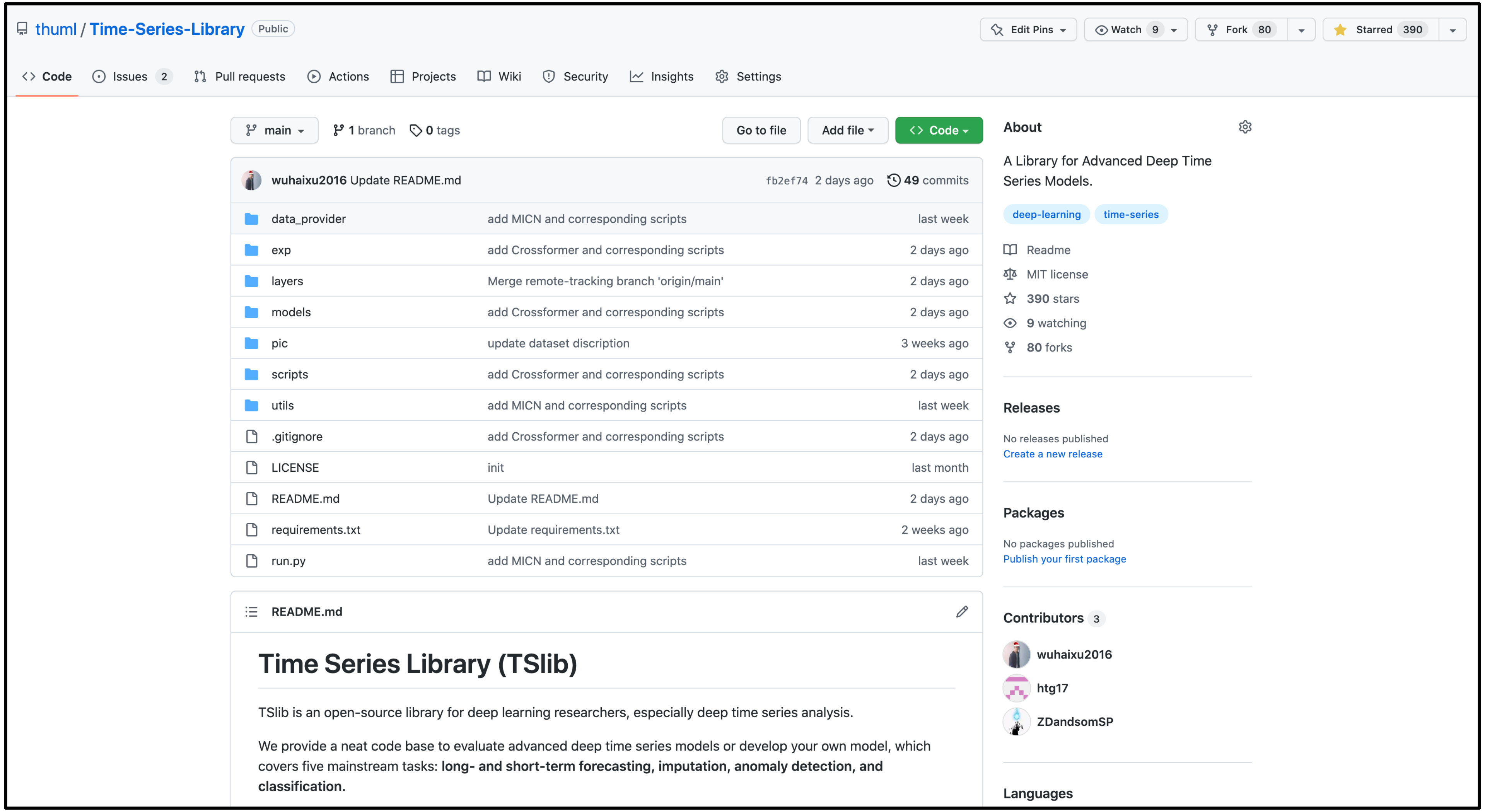
Model Ranking	Long-term Forecasting	Short-term Forecasting	Imputation	Anomaly Detection	Classification
🥇 1st	TimesNet	TimesNet	TimesNet	TimesNet	TimesNet
🥈 2nd	DLinear	Non-stationary Transformer	Non-stationary Transformer	Non-stationary Transformer	FEDformer
🥉 3rd	Non-stationary Transformer	FEDformer	Autoformer	Informer	Autoformer

Until 2023.02 (Keep updating)

Efficiency comparison for top 4 models

Models / Series Length		Parameter (MB)	GPU Memory (MiB)	Running Time (s / iter)	Ranking: Five tasks	Ranking: Avg Ranking
TimesNet (ours)	384	0.067	1245	0.024	(1, 1, 1, 1, 1)	1.0
	768	0.067	1585	0.040		
	1536	0.067	2491	0.045		
	3072	0.067	2353	0.073		
Non-stationary Transformer	384	1.884	2321	0.046	(3, 2, 2, 2, 8)	3.4
	768	1.910	4927	0.118		
	1536	1.961	/	/		
	3072	/	/	/		
Autoformer	384	1.848	2101	0.070	(7, 4, 3, 5, 3)	4.4
	768	1.848	3209	0.071		
	1536	1.848	5395	0.129		
	3072	1.848	10043	0.255		
FEDformer	384	2.901	5977	0.807	(4, 3, 6, 9, 2)	4.8
	768	2.901	7111	1.055		
	1536	2.901	9173	1.482		
	3072	2.901	/	/		

Open Source



Code is available at https://github.com/thuml/Time-Series-Library

Thank You!

whx20@mails.tsinghua.edu.cn