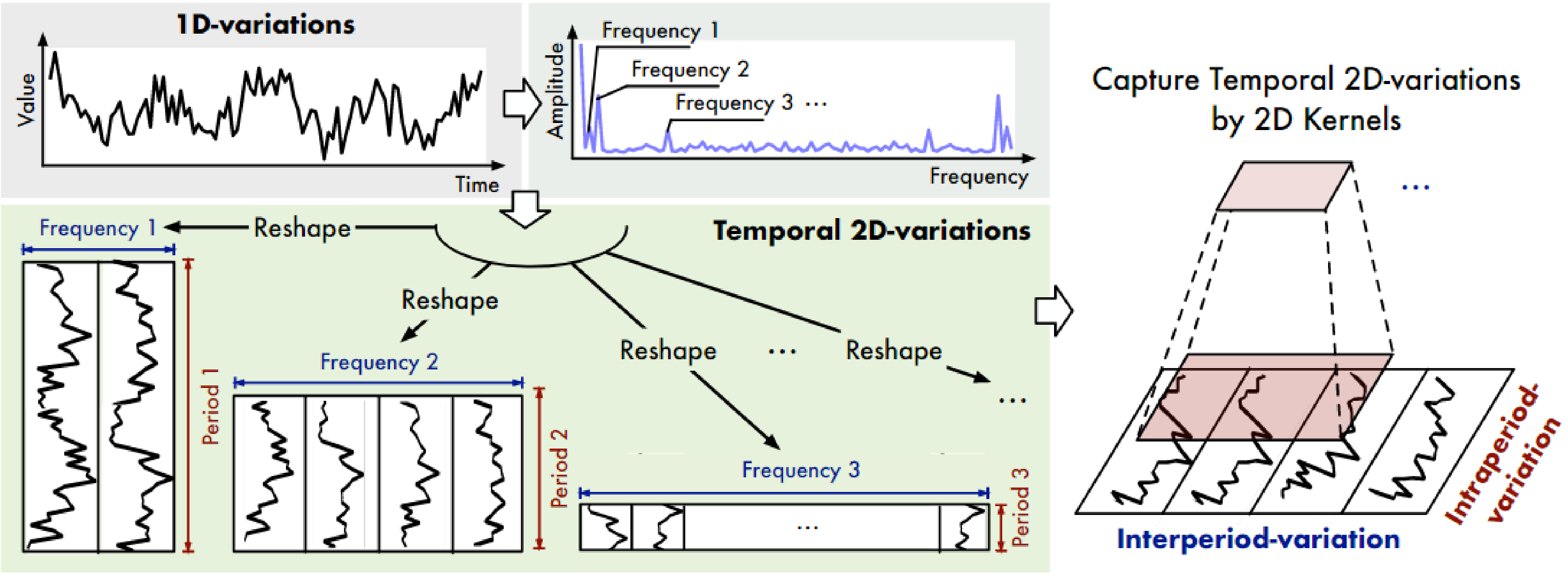
TimesNet

Temporal 2D-Variation Modeling for General Time Series Analysis



ICLR 2023

TimesNet

Temporal 2D-Variation Modeling for General Time Series Analysis

23.12.19

Presented by Yyyq

00

作者团队介绍：清华软院机器学习组 THUML

THUML ...

清华大学软件学院机器学习实验室，专注于迁移学习、深度学习、科学学习等基础理论方法及在人工智能和系统软件中的应用研究，负责人为王建民教授和龙明盛副教授，顾问为Michael I. Jordan院士。

- 龙明盛个人主页：
  http://ise.thss.tsinghua.edu.cn/~mlong/
- 小组的官方公众号：THUML-LAB
- 小组的github主页：https://github.com/thuml

Highlights

- Big nowcasting model for extreme precipitation (NowcastNet) was reported in News and Views and published in Nature 2023
- Unified forecasting model for worldwide stations (Corrformer) was published as the Cover Article in Nat. Mach. Intell. 2023
- Base forecasting model for time series (Autoformer) was ranked 14th of the most influential papers in NeurIPS 2021
- Conditional Domain Adversarial Network (CDAN) was ranked 6th of the most influential papers in NeurIPS 2018
- Joint Adaptation Network (JAN) was ranked 12th of the most influential papers in ICML 2017
- Deep Adaptation Network (DAN) was ranked 5th of the most influential papers in ICML 2015, Test of Time Award at FTL-IJCAI
- Joint Distribution Adaptation (JDA) was ranked 2nd of the most influential papers in ICCV 2013

00 作者团队介绍：清华软院机器学习组 THUML

时序库 **TSlib：https://github.com/thuml/Time-Series-Library**

Till October 2023, the top three models for five different tasks are:

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

01

问题描述

时间序列的多周期性：相互重叠、相互影响

交通流领域，比较常见的：Recently，Daily，Weekly

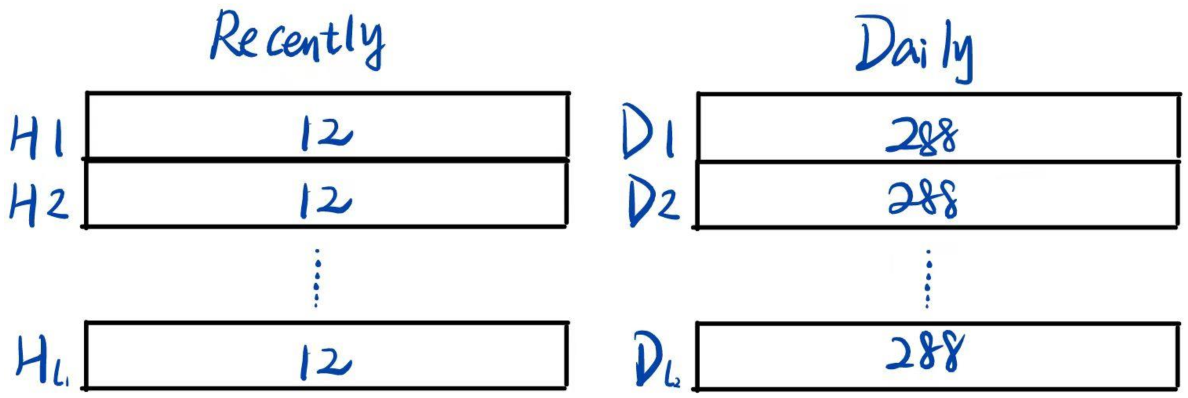
周期内(intraperiod)变化：相邻区域时间（短期时间模式）

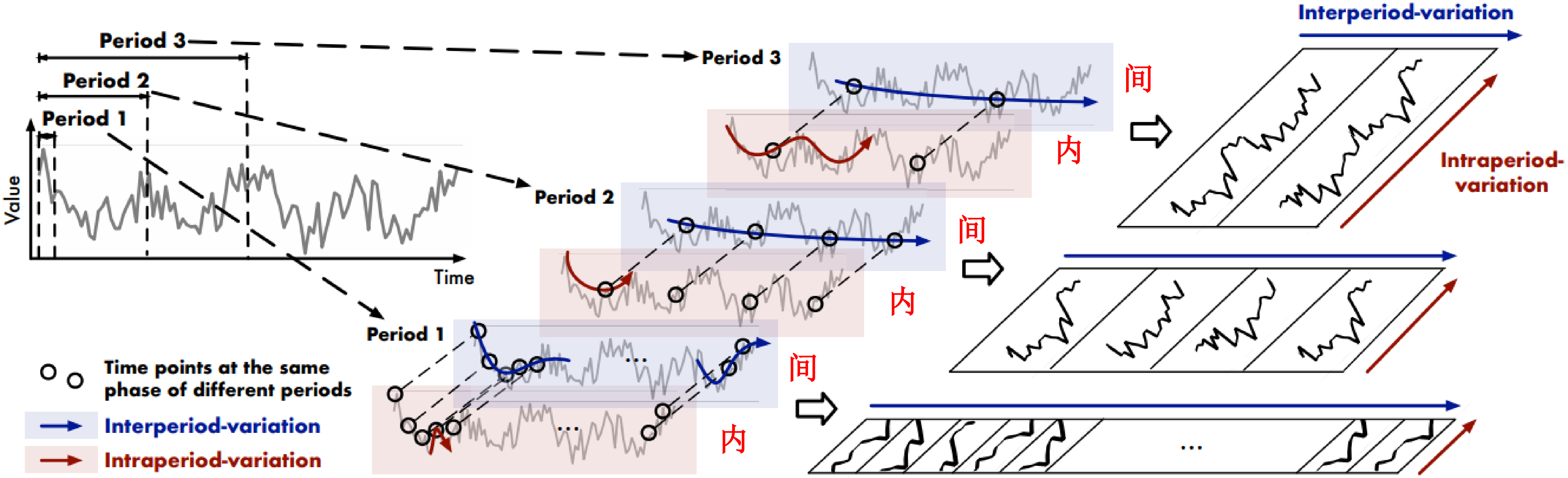
周期间(interperiod)变化：相邻时段的变化（连续不同时期的长期趋势）

不同的周期，会导致不同的周期内和周期间变化。

02 创新点

➢ **模块化**的时间变化建模方法：**一维→二维**，





02 创新点

- **模块化**的时间变化建模方法：一维→二维，同时表示周期内和周期间的变化
- **TimesNet with TimesBlock**
  - 自适应地发现多周期
  - 从二维张量中，捕获时间变化
- **任务通用型基础模型**

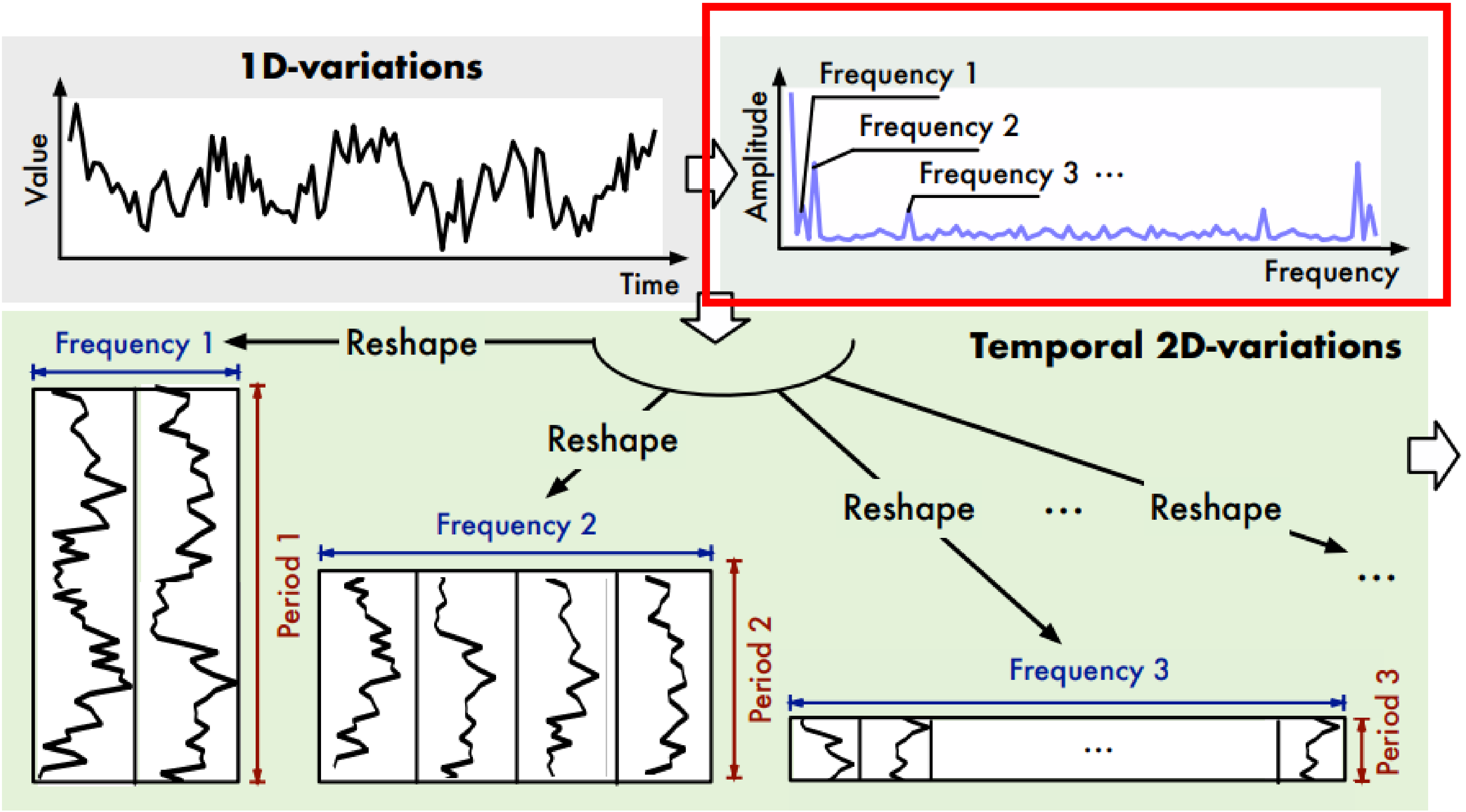
03 TimesNet模型介绍

➢ 自适应发现多周期

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

$$f_1, \cdots, f_k = \underset{f_* \in \{1, \cdots, [\frac{T}{2}]\}}{\arg\mathrm{Topk}}(\mathbf{A})$$

$$p_1, \cdots, p_k = \left\lceil \frac{T}{f_1} \right\rceil, \cdots, \left\lceil \frac{T}{f_k} \right\rceil,$$



03

TimesNet模型介绍

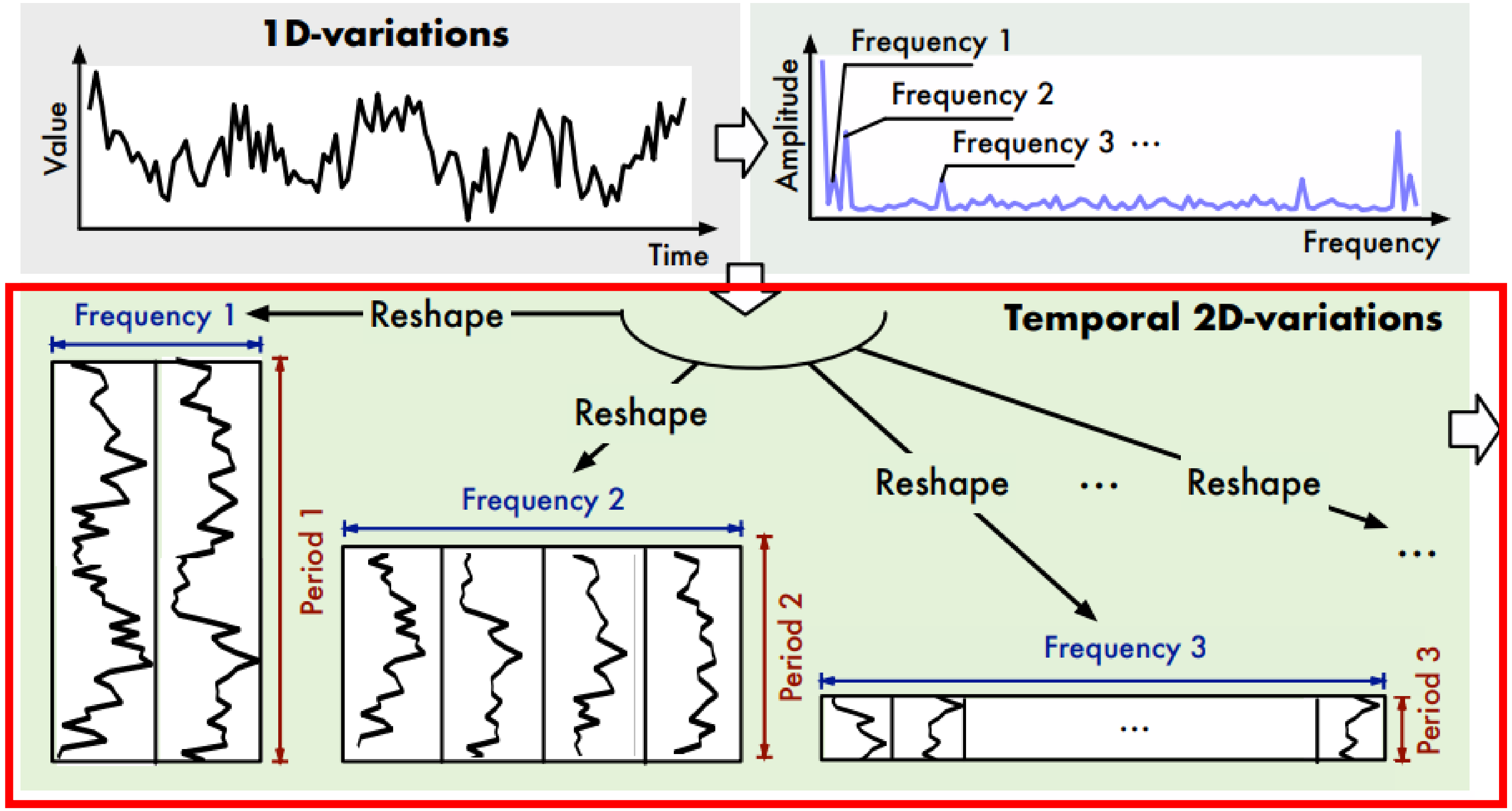
➢ 一维 → 二维

$$\mathbf{X}_{1D} \in \mathbb{R}^{T \times C}.$$

$$\mathbf{X}^{i}_{2D} = \mathrm{Reshape}_{p_i, f_i}\left(\mathrm{Padding}(\mathbf{X}_{1D})\right),$$

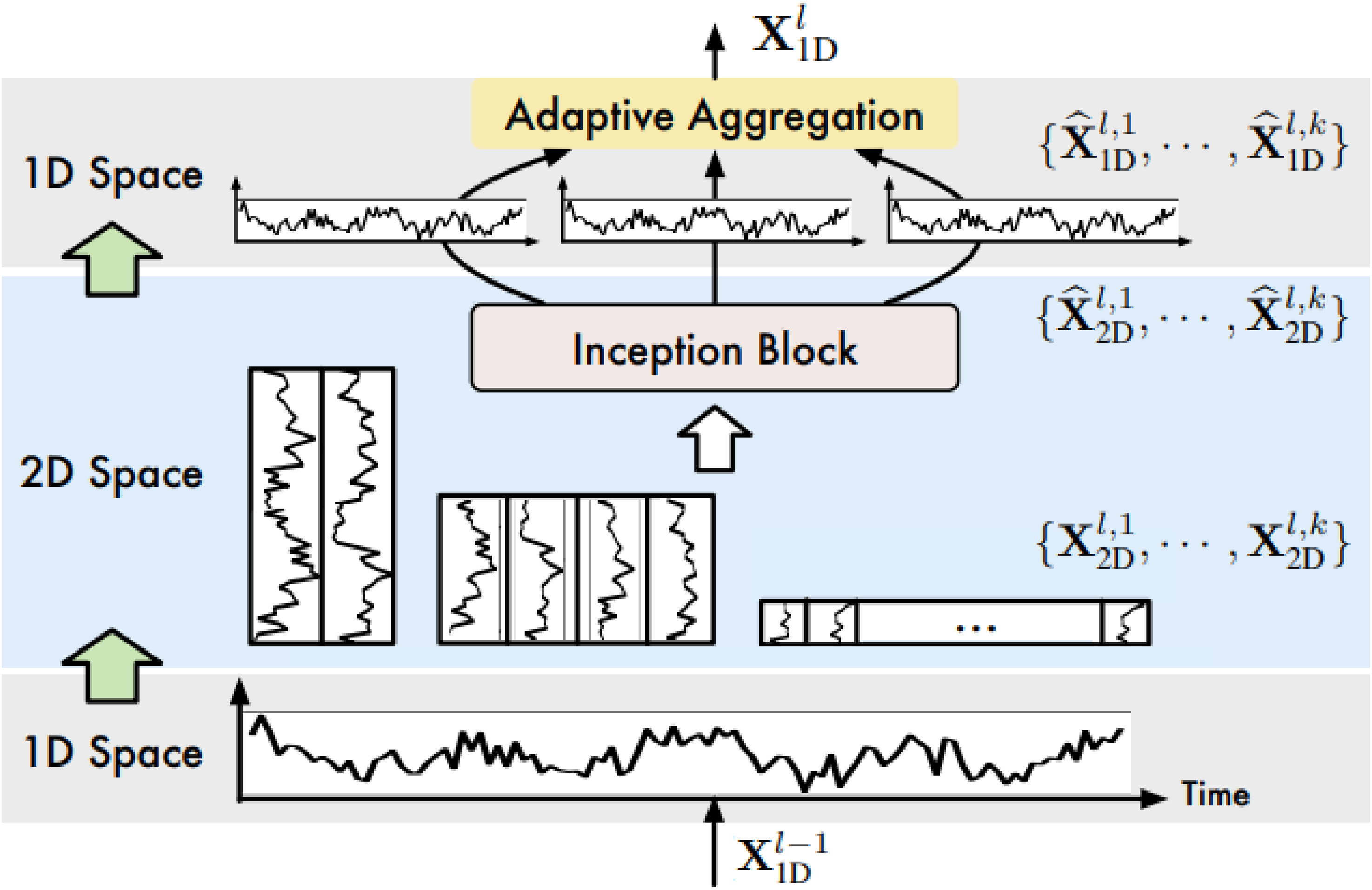
$$\mathbf{X}^{i}_{2D} \in \mathbb{R}^{p_i \times f_i \times C}$$

$$\{\mathbf{X}^{1}_{2D}, \cdots, \mathbf{X}^{k}_{2D}\}$$



03 TimesNet模型介绍

- **TimesBlock**



自适应融合

二维→一维

$$\widehat{\mathbf{X}}^{l,i}_{1D} = \mathrm{Trunc}\left(\mathrm{Reshape}_{1,(p_i\times f_i)}\left(\widehat{\mathbf{X}}^{l,i}_{2D}\right)\right),$$

提取二维时序变化表征

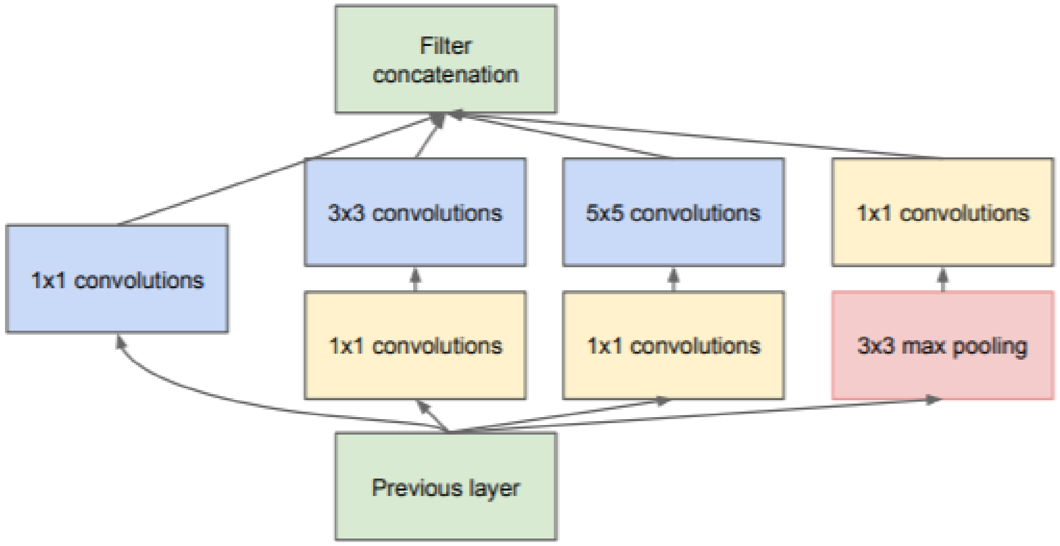
一维 → 二维

$$\mathbf{X}^{l,i}_{2D} = \mathrm{Reshape}_{p_i,f_i}\left(\mathrm{Padding}(\mathbf{X}^{l-1}_{1D})\right)$$

03 TimesNet模型介绍

➤ **TimesBlock**

$$\widehat{\mathbf{X}}_{2D}^{l,i} = \mathrm{Inception}\left(\mathbf{X}_{2D}^{l,i}\right).$$



(b) Inception module with dimension reductions

自适应融合

二维→一维

$$\widehat{\mathbf{X}}_{1D}^{l,i} = \mathrm{Trunc}\left(\mathrm{Reshape}_{1,(p_i \times f_i)}\left(\widehat{\mathbf{X}}_{2D}^{l,i}\right)\right),$$

提取二维时序变化表征

一维 → 二维

$$\mathbf{X}_{2D}^{l,i} = \mathrm{Reshape}_{p_i,f_i}\left(\mathrm{Padding}(\mathbf{X}_{1D}^{l-1})\right)$$

03 TimesNet模型介绍

➢ **TimesBlock**

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

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

自适应融合

⬆

二维→一维

⬆ $\widehat{\mathbf{X}}_{\mathrm{1D}}^{l,i} = \mathrm{Trunc}\left(\mathrm{Reshape}_{1,(p_i \times f_i)}\left(\widehat{\mathbf{X}}_{\mathrm{2D}}^{l,i}\right)\right),$

提取二维时序变化表征

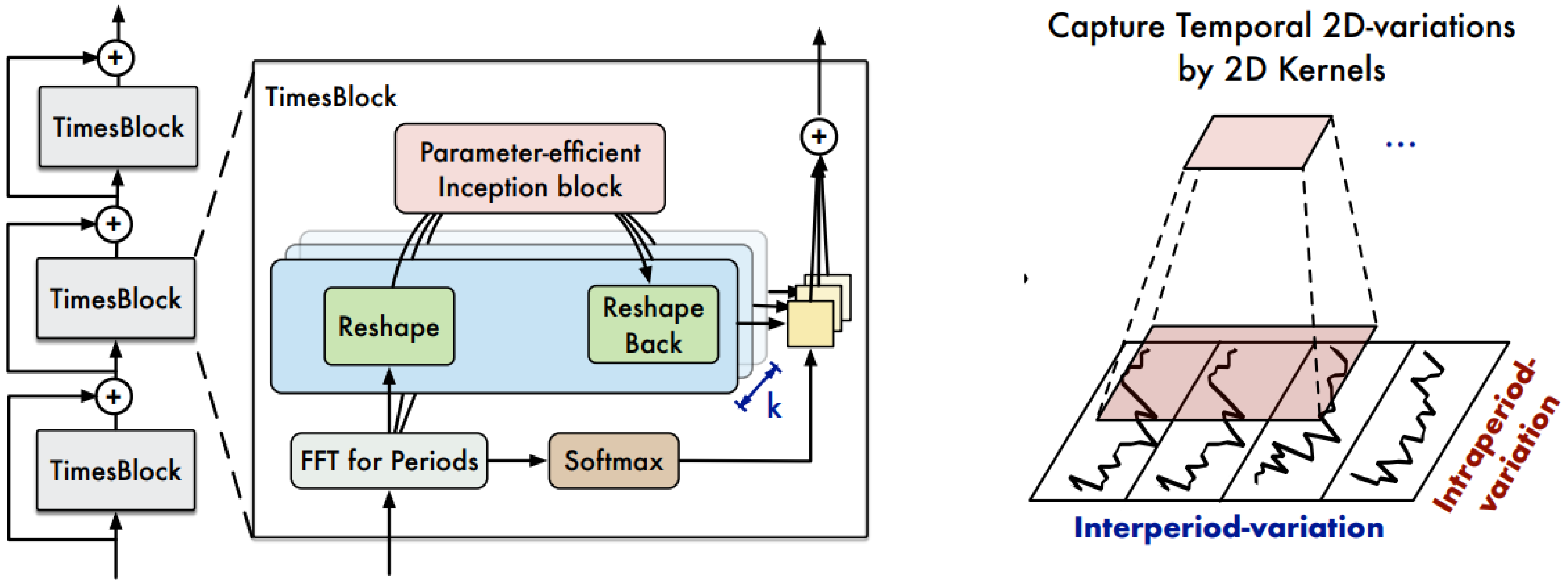
⬆

一维 → 二维

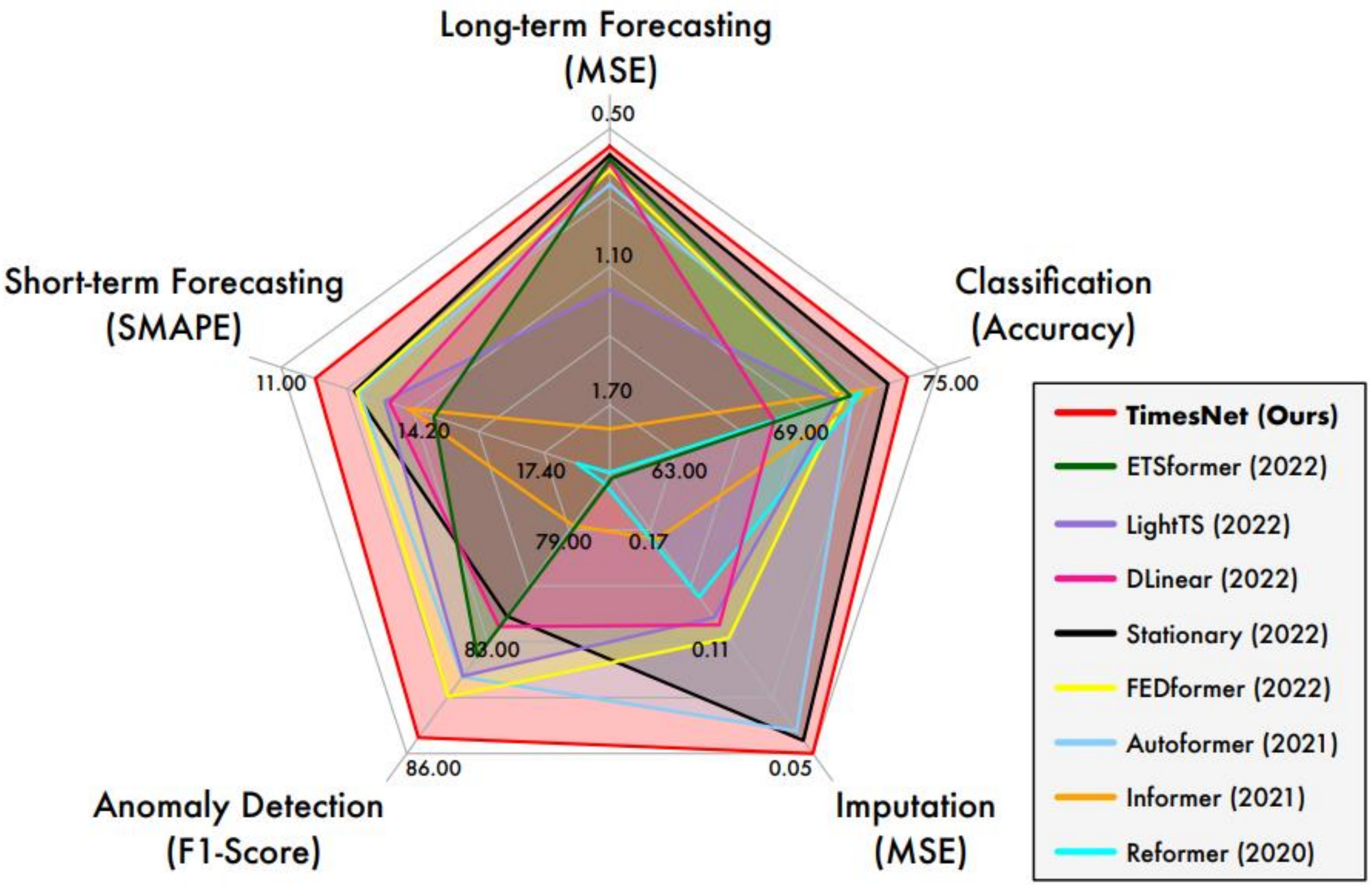
$\mathbf{X}_{\mathrm{2D}}^{l,i} = \mathrm{Reshape}_{p_i,f_i}\left(\mathrm{Padding}(\mathbf{X}_{\mathrm{1D}}^{l-1})\right)$

03 TimesNet模型介绍

➢ **TimesBlock堆叠：** $$\mathbf{X}_{\mathrm{1D}}^{l} = \mathrm{TimesBlock}\left(\mathbf{X}_{\mathrm{1D}}^{l-1}\right) + \mathbf{X}_{\mathrm{1D}}^{l-1}.$$



04 实验结果



Long-term Forecasting (MSE)
0.50
1.10
1.70
Classification (Accuracy)
75.00
69.00
63.00
Short-term Forecasting (SMAPE)
11.00
14.20
17.40
Anomaly Detection (F1-Score)
86.00
83.00
79.00
Imputation (MSE)
0.05
0.11
0.17
TimesNet (Ours)
ETSformer (2022)
LightTS (2022)
DLinear (2022)
Stationary (2022)
FEDformer (2022)
Autoformer (2021)
Informer (2021)
Reformer (2020)

04 实验结果：预测

关于长期预测和短期预测，作者在评论中如下回复：

“目前的长期预测的9个数据集的特点在于，他们一个数据集就是一个超级长的序列，每个batch是从整体序列中切分出来的。因此不管是哪个batch，他们的temporal pattern都是类似的（因为是同一个场景下的观测）”——滑动窗口取样本

“除了上面9个数据集之外，TimesNet还在M4上做了实验，M4的特点是每一个sample都是独立的，也就是意味着他们的temporal pattern相差很多” ——不需要滑动窗口

Long-term: ETT (4 subsets), Electricity, Traffic, Weather, Exchange, ILI

Short-term: M4 (6 subsets)

M4来自第四届预测竞赛，10万个时间序列(Foredeck)，这些数据来自多个、不同的和公开可访问的来源。

Foredeck强调商业预测应用，包括来自相关领域的系列，如工业、服务、旅游、进出口、人口统计、教育、劳动和工资、政府、家庭、债券、股票、保险、贷款、房地产、运输以及自然资源和环境.

04

实验结果：长期预测

Table 2: Long-term forecasting task. The past sequence length is set as 36 for ILI and 96 for the others. All the results are averaged from 4 different prediction lengths, that is $\{24, 36, 48, 60\}$ for ILI and $\{96, 192, 336, 720\}$ for the others. See Table 13 in Appendix for the full results.

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	**0.400**	**0.406**	0.429	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**	0.293	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**	0.542	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**	0.439	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**	0.208	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**	0.621	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**	0.271	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	0.410	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	2.497	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

04 实验结果：短期预测

Table 3: Short-term forecasting task on M4. The prediction lengths are in [6, 48] and results are weighted averaged from several datasets under different sample intervals. See Table 14 for full results.

Models	**TimesNet (Ours)**	N-HiTS (2022)	N-BEATS (2019)	ETSformer (2022)	LightTS (2022)	DLinear (2023)	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

$$\mathrm{SMAPE} = \frac{200}{H}\sum_{i=1}^{H}\frac{|\mathbf{X}_i - \widehat{\mathbf{X}}_i|}{|\mathbf{X}_i| + |\widehat{\mathbf{X}}_i|},$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$$

$$\mathrm{OWA} = \frac{1}{2}\left[\frac{\mathrm{SMAPE}}{\mathrm{SMAPE}_{\text{Naïve2}}} + \frac{\mathrm{MASE}}{\mathrm{MASE}_{\text{Naïve2}}}\right],$$

$$\mathrm{SMAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2}$$

04 实验结果：插补

Table 4: Imputation task. We randomly mask $\{12.5\%, 25\%, 37.5\%, 50\%\}$ time points in length-96 time series. The results are averaged from 4 different mask ratios. See Table 16 for full results.

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

04

实验结果：分类

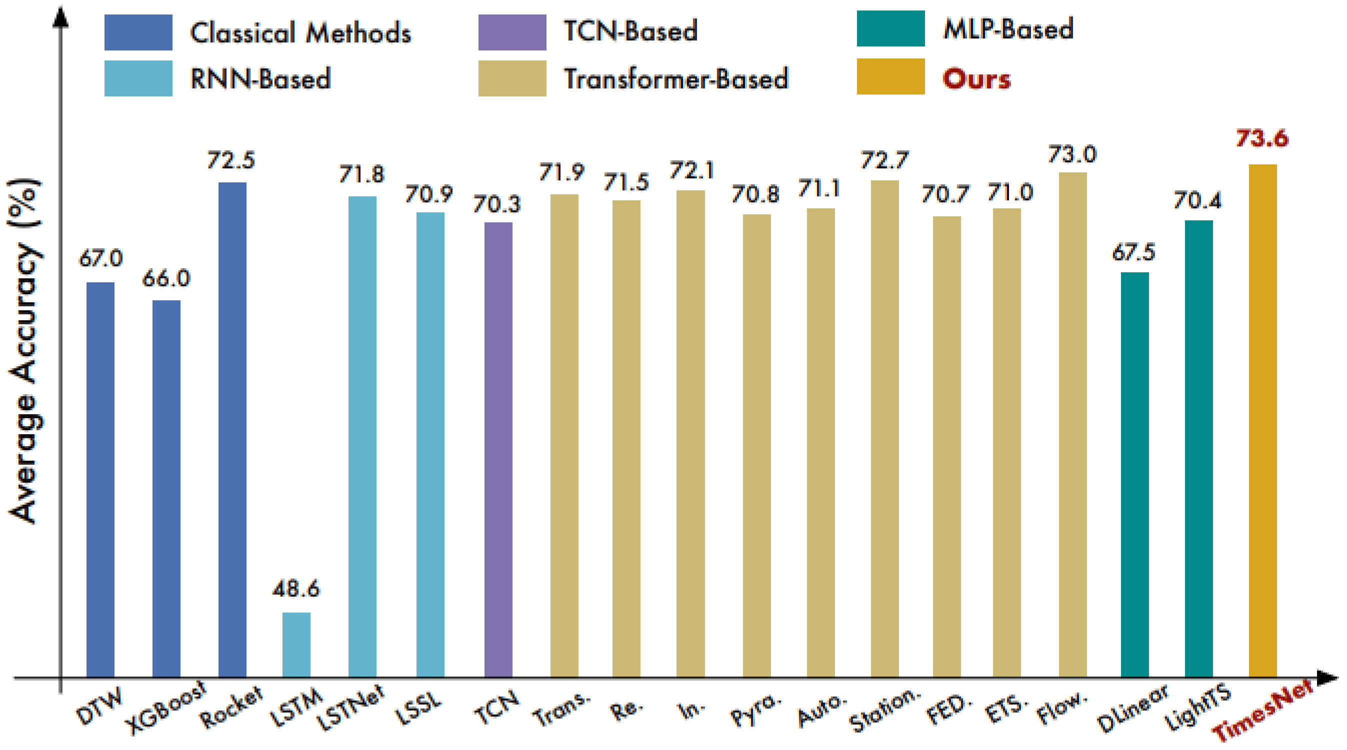


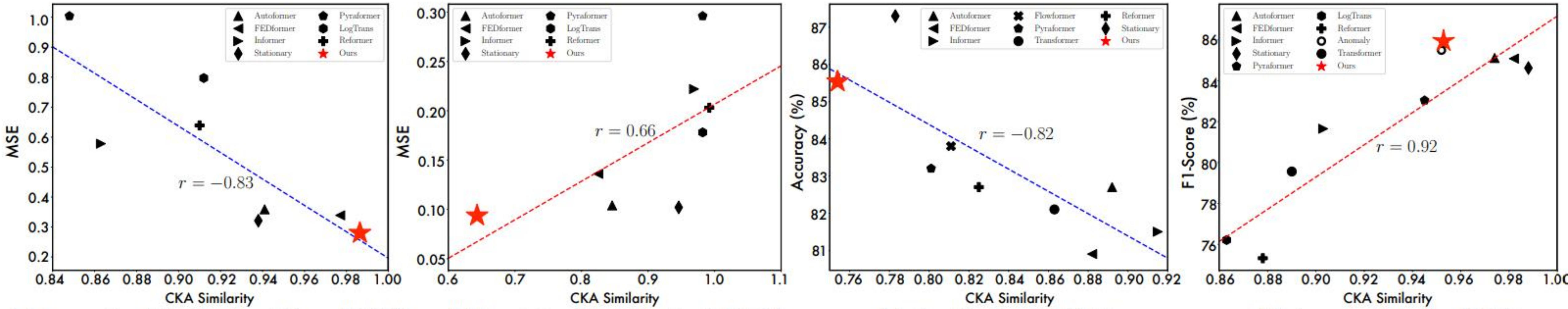
Figure 5: Model comparison in classification. "∗." in the Transformers indicates the name of ∗former. The results are averaged from 10 subsets of UEA. See Table 17 in Appendix for full results.

04 实验结果：异常检测

Table 5: Anomaly detection task. We calculate the F1-score (as %) for each dataset. *. means the *former. A higher value of F1-score indicates a better performance. See Table 15 for full results.

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

04 实验结果：表征分析



(a) Forecasting (Weather input-96-predict-336) (b) Imputation (Electricity Mask 37.5%) (c) Classification (PEMS-SF) (d) Anomaly Detection (SMD)

CKA相似度：值越低，代表模型底层-顶层之间的表征差异越大，即在不同层次上具有区分性。

- **在预测与异常检测任务中，**效果越好的模型往往底层-顶层的表征相似度越高，表明任务期待更加底层的表征（low-level representations）；
- **在分类与缺失值填补任务中，**效果越好的模型往往底层-顶层的表征相似度越低，表明该任务需要层次化表征（hierarchical representation）。

05

补充

1、“在Timesnet模块前面加一个线性层，将输入序列的形状从[B, seq_len, C]变成[B, seq_len + pred_len, C]。这样做是出于什么的考虑？”

作者回复：我们提出timesnet的目的是捕捉时序变化，预测任务的核心是捕捉过去和未来时序变化的关联，因此我们首先在底层进行对齐，然后随着深度模型层数的加深，通过timesnet不断refine预测结果。

2、有人提出：“classification和anomoly detection的基线模型不具有代表性，目前卷积类模型是优于transformer系模型的，基线中主要是former系的模型。”

3、*作者补充：“通用表征学习能力是foundation model的基础。TimesNet中折叠的目的是：使简单的2D卷积就可以**同时覆盖两种locality**，从而使的特征提取更加高效。这种设计和之前的直接作用在1D原始序列上的深度模型是有很大区别的。”*

来源：https://zhuanlan.zhihu.com/p/606575441

谢谢观看

MANY THANKS !

23.12.19