MoFo: Empowering Long-term Time Series Forecasting with Periodic Pattern Modeling

Jiaming Ma (Presenter), Binwu Wang*, Qihe Huang, Guanjun Wang,

Pengkun Wang, Zhengyang Zhou, Yang Wang*

University of Science ang Technology of China (USTC)

Brief Introduction

➢ Challenges of Long-term Time Series Forecasting

❖ ***Continuous but Low-autocorrelated Time Steps of Input Patch.***

SOTA Transformer-based models adopt patching to aggregate adjacent steps for efficiency.

❖ ***Weak Inductive Bias for Periodicity.***

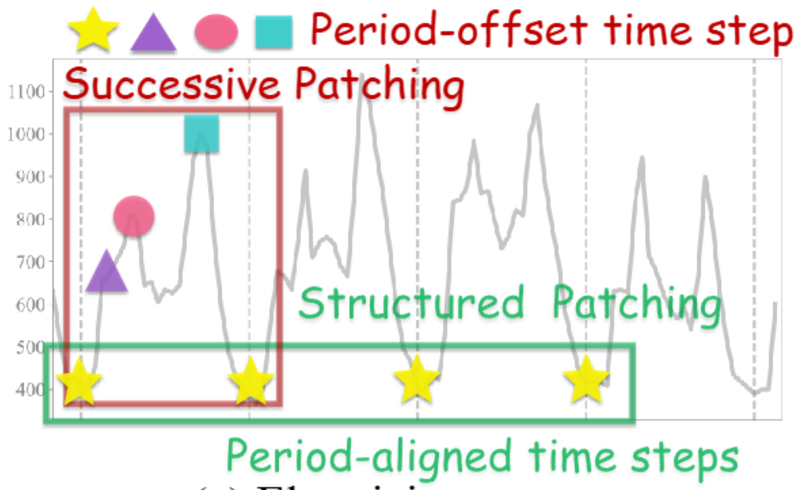
Transformer-based models lack an properly inductive bias for periodicity.

Brief Introduction

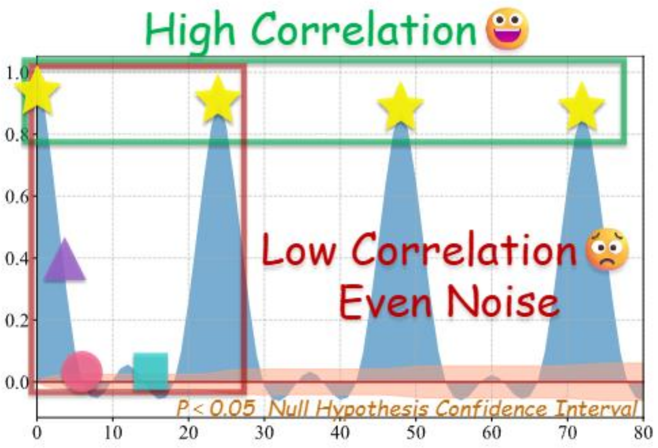
➢ Challenges of Long-term Time Series Forecasting

❖ *Continuous but Low-autocorrelated Time Steps of Input Patch.*

SOTA Transformer-based models adopt patching to aggregate adjacent steps for efficiency.



(a) Electricity Dataset



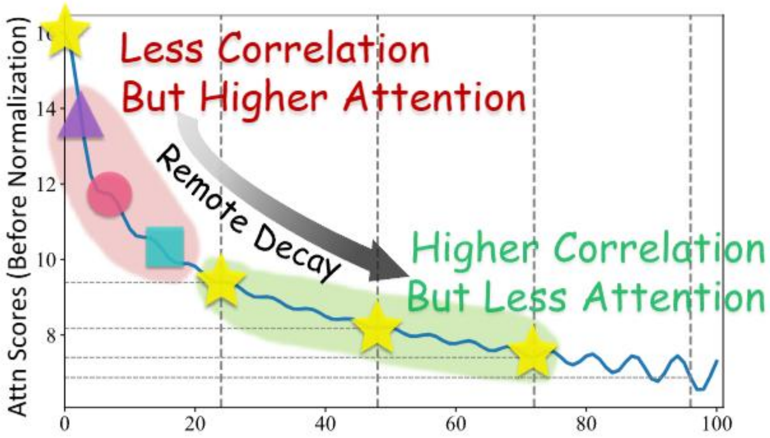
(b) Autocorrelation Function

Brief Introduction

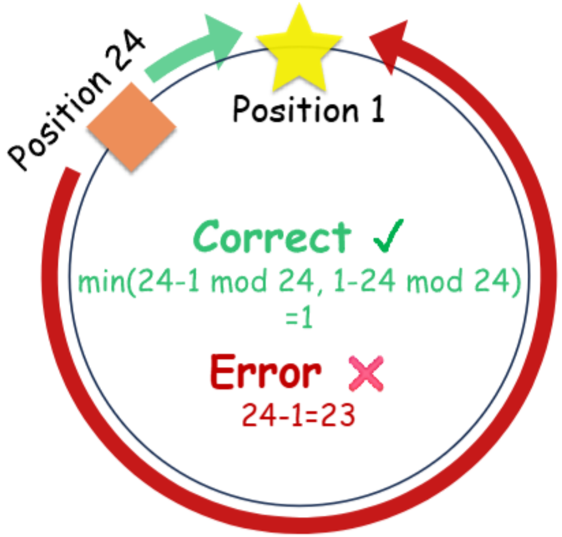
➢ Challenges of Long-term Time Series Forecasting

❖ *Weak Inductive Bias for Periodicity.*

Transformer-based models lack an properly inductive bias for periodicity.

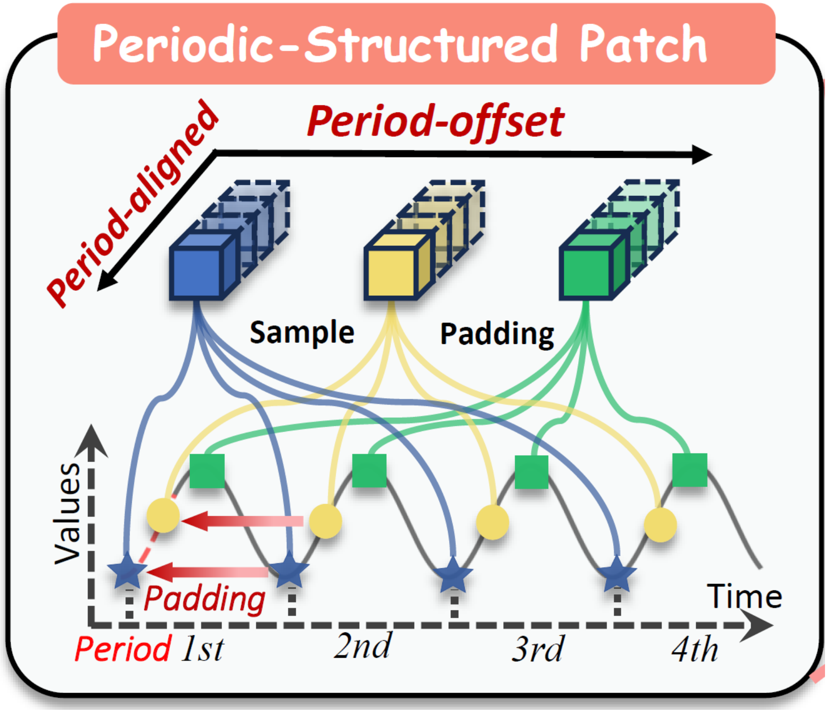


(c) Sinuous Position Encoding in LTSF



(d) Period Position Calculation

Methods: Period-structured Patch



Input Padding. Our proposed padding strategy fills incomplete periodic segments with data from adjacent periods. Specifically, we start from the current time step and move backward to delineate periods of length P, and if necessary, we prepend the input time series with the first $P-(T \bmod P)$ time steps of the first complete period, as shown in Figure 2. This ensures a seamless continuation of the sequence while retaining its underlying periodic structure. As a result, the input series is extended to $\mathbf{X}_{pad} \in \mathbb{R}^{T'}$ with $T'=P*\lceil T/P \rceil$, and the padded series can be formally expressed as:

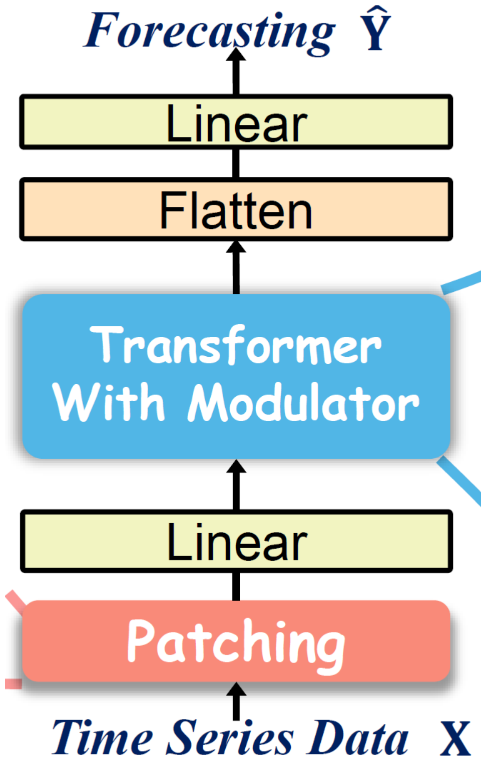
$$\mathbf{X}_{pad} = \begin{cases} \text{Concat}\left(\mathbf{X}_{(T \bmod P):P}, \mathbf{X}\right), & \text{if } T \bmod P > 0, \\ \mathbf{X}, & \text{if } T \bmod P = 0. \end{cases} \tag{1}$$

Sampling Patch and Unflatten. We sample time steps at periodic intervals (i.e., period-aligned time steps) from $\mathbf{X}_{pad}$ and group them into the same patch. For example, for i-th patch, it can be denoted as $\overline{\mathbf{X}}^i = [x_i, x_{i+P}, \cdots, x_{i+P*\lceil T/P \rceil}] \in \mathbb{R}^{\lceil T/P \rceil}$, where x_{i+P} means the data point at the time step $(i+P)$ in $\mathbf{X}_{pad}$. Then we unflatten $\overline{\mathbf{X}}$ to generate the patch-structure input $\mathbf{X}_{in}$,

$$\mathbf{X}_{in} = \text{Unflatten}\left(\overline{\mathbf{X}}\right) \in \mathbb{R}^{P \times \lceil T/P \rceil}. \tag{2}$$

where P is the number of patches (also equal to the period length).

Methods: Attention with Period Mask



(1) Attention with mask on period-offset dimension

$$\text{Attention}\,(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_h}} + \log \mathbf{M}\right)\mathbf{V},$$

$$\text{where} \quad \mathbf{Q} = \mathbf{Z}\mathbf{W}_Q^j \in \mathbb{R}^{P\times d_h}, \mathbf{K} = \mathbf{Z}\mathbf{W}_K^j \in \mathbb{R}^{P\times d_h}, \mathbf{V} = \mathbf{Z}\mathbf{W}_V^j \in \mathbb{R}^{P\times d_h},$$

(2) Period-distance mask (*non-differentiable*)

$$\mathbf{M}_{ij} = \begin{cases} 1, & \text{if } \gamma_{ij} \le \beta, \\ 0, & \text{if } \gamma_{ij} > \beta, \end{cases} \iff \log \mathbf{M}_{ij} = \begin{cases} 0, & \text{if } \gamma_{ij} \le \beta, \\ -\infty, & \text{if } \gamma_{ij} > \beta. \end{cases}$$

(3) Period distance calculation

$$\gamma_{ij} = \min\{(i-j) \bmod P, (j-i) \bmod P\} \in [0, \lfloor P/2 \rfloor],$$

Methods: Smooth Modulator of Mask

✓ **Note:** the Heaviside Step function $\mathcal{H}:\mathbb{R}\rightarrow\{0,1\}$: $\mathbf{M}_{ij}=\mathcal{H}\left(\beta-\gamma_{ij}\right)$,

Theorem 1. Regulated Relaxation Function

Define a continuous differentiable function $\mathcal{S}\left(\cdot;\alpha,\beta\right):\mathbb{R}^{+}\cup\{0\}\rightarrow[0,1]$ as follows,

$$\mathcal{S}\left(\gamma;\alpha,\beta\right)=\frac{1}{1+\exp\left(\alpha\left(\gamma-\beta\right)\right)}+\frac{\exp\left(-\gamma\right)}{1+\exp\left(\alpha\beta\right)}\in[0,1]\,. \tag{9}$$

where the regulated parameter $\alpha>0$ control the gradient of attenuation and $\beta>0$ is the distance penalty threshold. This function has following properties:

(1) $\mathcal{S}\left(\gamma;\alpha,\beta\right)$ is the smooth approximation of $\mathcal{H}\left(\beta-\gamma_{ij}\right)$ for arbitrary $\gamma\geq 0$ satisfies,

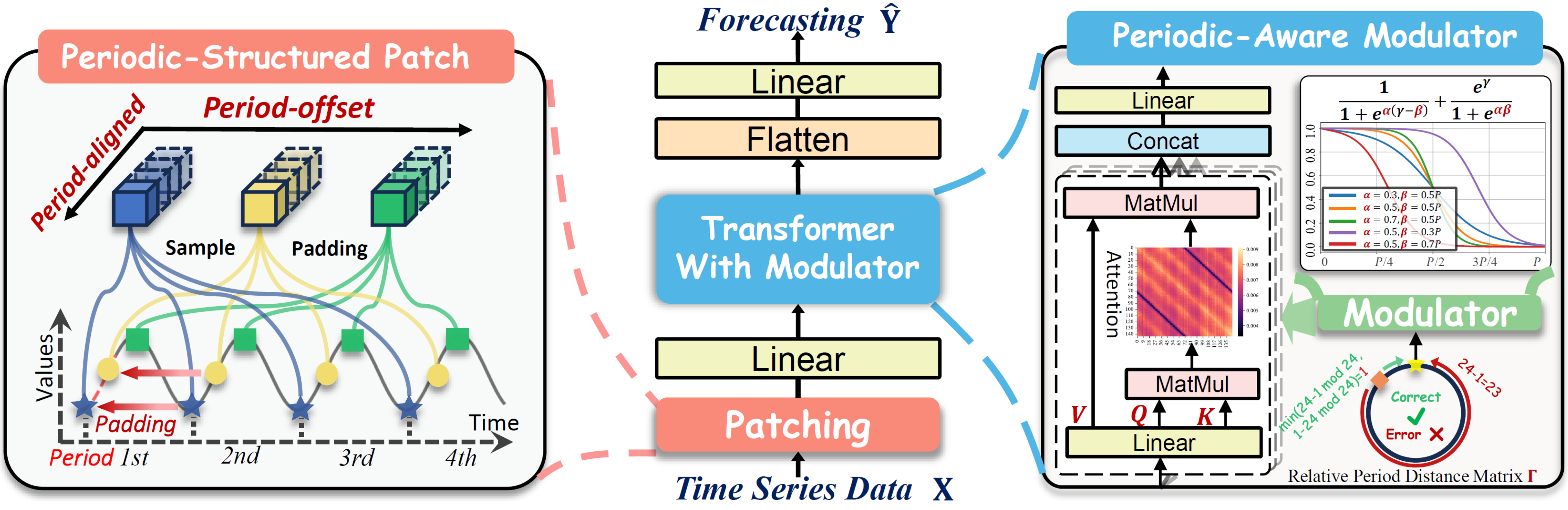
$$\mathcal{S}\left(0;\alpha,\beta\right)=1,\quad \mathcal{S}\left(+\infty;\alpha,\beta\right)=0,\quad \forall\alpha,\beta>0. \tag{10}$$

(2) The cumulative error upper bound of this smooth approximation satisfies,

$$\int_{0}^{+\infty}\left|\mathcal{H}\left(\beta-\gamma\right)-\mathcal{S}\left(\gamma;\alpha,\beta\right)\right|\mathrm{d}\gamma<\frac{2\log 2}{\alpha}+\frac{1}{1+\exp\alpha}\rightarrow 0^{+}\quad(\alpha\rightarrow+\infty). \tag{11}$$

**Proof is available in the paper.*

MOFO: Transformer with Modulator

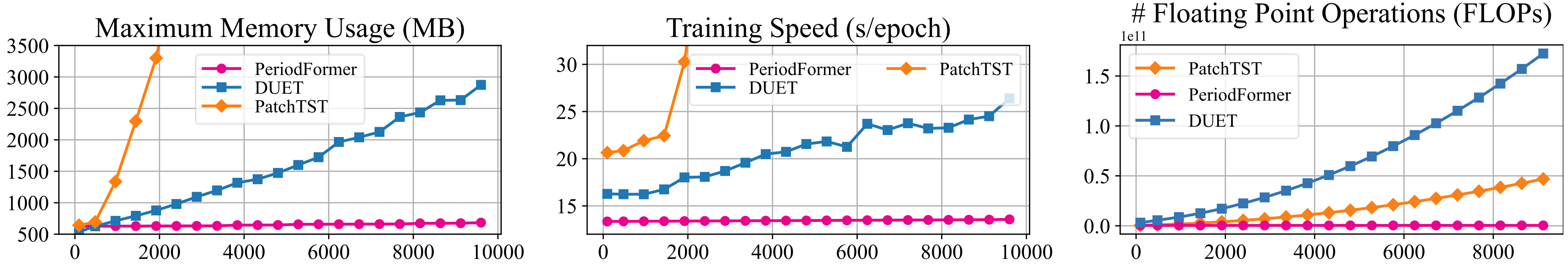


Experimental Results

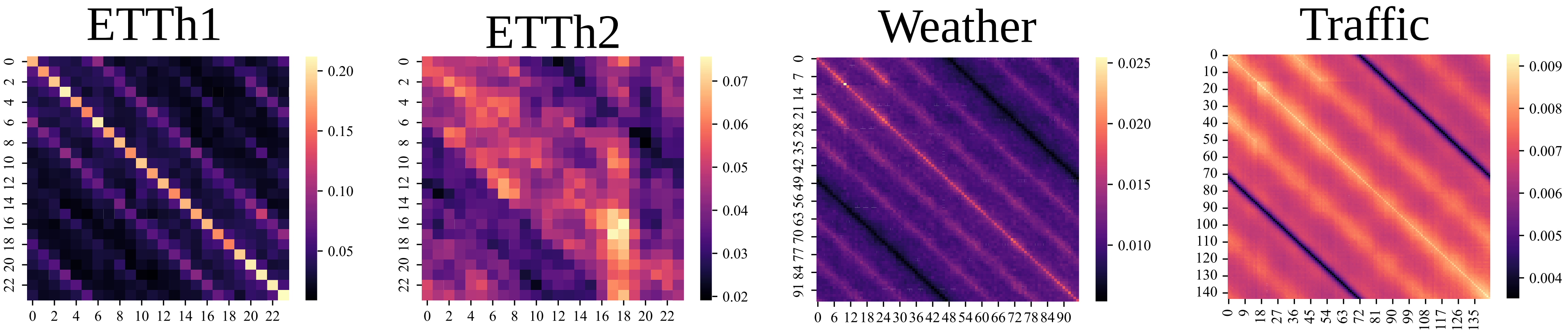
Method		MoFo (Ours)		DUET (2025)		PDF (2024)		iTransformer (2024)		Pathformer (2024)		CycleNet (2024)		TimeMixer (2024)		PatchTST (2023)		Crossformer (2023)		DLinear (2023)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.360	0.389	**0.352**	**0.384**	0.360	0.391	0.386	0.405	0.372	0.392	0.372	0.394	0.372	0.401	0.377	0.397	0.411	0.435	0.379	0.403
	192	0.397	0.413	0.398	**0.409**	**0.392**	0.414	0.424	0.440	0.408	0.415	0.404	0.417	0.413	0.430	0.409	0.425	0.409	0.438	0.408	0.419
	336	**0.407**	**0.424**	0.414	0.426	0.418	0.435	0.449	0.460	0.438	0.434	0.430	0.429	0.438	0.450	0.431	0.444	0.433	0.457	0.440	0.440
	720	0.447	**0.454**	**0.429**	0.455	0.456	0.462	0.495	0.487	0.450	0.463	0.448	0.464	0.486	0.484	0.457	0.477	0.501	0.514	0.471	0.493
ETTh2	96	0.273	**0.334**	**0.270**	0.336	0.276	0.341	0.297	0.348	0.279	0.336	0.277	0.341	0.281	0.351	0.274	0.337	0.728	0.603	0.300	0.364
	192	**0.327**	**0.373**	0.332	0.374	0.339	0.382	0.372	0.403	0.345	0.380	0.341	0.385	0.349	0.387	0.348	0.384	0.723	0.607	0.387	0.423
	336	0.361	0.405	**0.353**	**0.397**	0.374	0.406	0.388	0.417	0.378	0.408	0.370	0.411	0.366	0.413	0.377	0.416	0.740	0.628	0.490	0.487
	720	**0.379**	**0.425**	0.382	**0.425**	0.398	0.433	0.424	0.444	0.437	0.455	0.424	0.451	0.401	0.436	0.406	0.441	1.386	0.882	0.704	0.597
ETTm1	96	0.286	0.335	**0.279**	**0.333**	0.286	0.340	0.300	0.353	0.290	0.335	0.297	0.344	0.293	0.345	0.289	0.343	0.314	0.367	0.300	0.345
	192	**0.320**	0.363	**0.320**	**0.358**	0.321	0.364	0.341	0.380	0.337	0.363	0.332	0.365	0.335	0.372	0.329	0.368	0.374	0.410	0.336	0.366
	336	**0.347**	0.382	0.348	**0.377**	0.354	0.383	0.374	0.396	0.374	0.384	0.366	0.386	0.365	0.386	0.362	0.390	0.413	0.432	0.367	0.386
	720	**0.388**	0.411	0.405	**0.408**	0.408	0.415	0.429	0.430	0.428	0.416	0.417	0.414	0.426	0.412	0.413	0.423	0.753	0.613	0.419	0.416
ETTm2	96	**0.155**	**0.240**	0.161	0.248	0.163	0.251	0.175	0.266	0.164	0.250	0.157	0.247	0.165	0.256	0.165	0.255	0.296	0.391	0.164	0.255
	192	**0.211**	**0.283**	0.214	0.287	0.219	0.290	0.242	0.312	0.219	0.288	0.214	0.286	0.225	0.298	0.221	0.293	0.369	0.416	0.224	0.304
	336	**0.258**	**0.314**	0.267	0.320	0.269	0.330	0.282	0.337	0.267	0.319	0.269	0.322	0.277	0.332	0.276	0.327	0.588	0.600	0.277	0.337
	720	**0.342**	**0.368**	0.348	0.374	0.349	0.382	0.375	0.394	0.361	0.377	0.363	0.382	0.360	0.387	0.362	0.381	0.750	0.612	0.371	0.401
Weather	96	**0.141**	**0.186**	0.146	0.191	0.147	0.196	0.157	0.207	0.148	0.195	0.166	0.222	0.147	0.198	0.150	0.200	0.143	0.210	0.170	0.230
	192	**0.186**	**0.230**	0.188	0.231	0.193	0.240	0.200	0.248	0.191	0.235	0.213	0.259	0.192	0.243	0.191	0.239	0.195	0.261	0.216	0.273
	336	**0.233**	0.272	0.234	**0.268**	0.245	0.280	0.252	0.287	0.243	0.274	0.262	0.291	0.247	0.284	0.242	0.279	0.254	0.319	0.258	0.307
	720	0.312	0.331	**0.305**	**0.319**	0.323	0.334	0.320	0.336	0.318	0.326	0.329	0.338	0.318	0.330	0.312	0.330	0.335	0.385	0.323	0.362
Solar	96	**0.169**	0.214	**0.169**	**0.195**	0.181	0.247	0.190	0.244	0.218	0.235	0.201	0.252	0.179	0.232	0.170	0.234	0.183	0.208	0.199	0.265
	192	**0.177**	0.231	0.187	**0.207**	0.199	0.257	0.193	0.257	0.196	0.220	0.221	0.261	0.201	0.259	0.204	0.302	0.208	0.226	0.220	0.282
	336	**0.186**	0.238	0.199	**0.213**	0.208	0.269	0.203	0.266	0.195	0.220	0.233	0.269	0.190	0.256	0.212	0.293	0.212	0.239	0.234	0.295
	720	**0.193**	0.248	0.202	**0.216**	0.212	0.275	0.223	0.281	0.208	0.237	0.236	0.271	0.203	0.261	0.215	0.307	0.215	0.256	0.243	0.301
Electricity	96	**0.122**	**0.215**	0.128	0.219	0.128	0.222	0.134	0.230	0.135	0.222	0.126	0.221	0.153	0.256	0.143	0.247	0.134	0.231	0.140	0.237
	192	**0.140**	**0.234**	0.145	0.235	0.147	0.242	0.154	0.250	0.157	0.253	0.144	0.239	0.168	0.269	0.158	0.260	0.146	0.243	0.154	0.251
	336	**0.157**	**0.252**	0.163	0.255	0.165	0.260	0.169	0.265	0.170	0.267	0.161	0.253	0.189	0.291	0.168	0.267	0.165	0.264	0.169	0.268
	720	**0.191**	0.284	0.193	**0.281**	0.199	0.289	0.194	0.288	0.211	0.302	0.199	0.286	0.228	0.320	0.214	0.307	0.237	0.314	0.204	0.301
Traffic	96	0.362	0.247	**0.360**	**0.238**	0.368	0.252	0.363	0.265	0.384	0.250	0.389	0.276	0.369	0.257	0.370	0.262	0.526	0.288	0.395	0.275
	192	**0.379**	0.254	0.383	**0.249**	0.382	0.261	0.384	0.273	0.405	0.257	0.406	0.280	0.400	0.272	0.386	0.269	0.503	0.263	0.407	0.280
	336	**0.390**	**0.258**	0.395	0.259	0.393	0.268	0.396	0.277	0.424	0.265	0.425	0.291	0.407	0.272	0.396	0.275	0.505	0.276	0.417	0.286
	720	**0.424**	0.281	0.435	**0.278**	0.438	0.297	0.445	0.308	0.452	0.283	0.450	0.303	0.461	0.316	0.435	0.295	0.552	0.301	0.454	0.308

Experimental Results

(1) Efficiency on extremely long input series.



(2) Visualization of attention logits



中国科学技术大学 University of Science and Technology of China

数据智能实验室 Data Intelligence Lab

NEURAL INFORMATION PROCESSING SYSTEMS

◊ Connection & Cooperation

- ❖ Available Code: https://github.com/PoorOtterBob/MoFo.
- ❖ Contact Emails: JiamingMa@mail.ustc.edu.cn.
- ❖ Personal Websit: https://poorotterbob.github.io/.
- ❖ WeChat:

See you San Diego!!!