TimeMixer

Decomposable MultiScale Mixing for Time Series Forecasting

（TimeMixer 可分解多尺度混合用于时间序列预测）

论文分享

汇报时间：2024.07.03

汇报人：梁琨

目 录

CONTENT

壹 项目介绍 03

贰 项目流程 04

叁 项目进展 07

肆 项目后续安排 20

问题描述

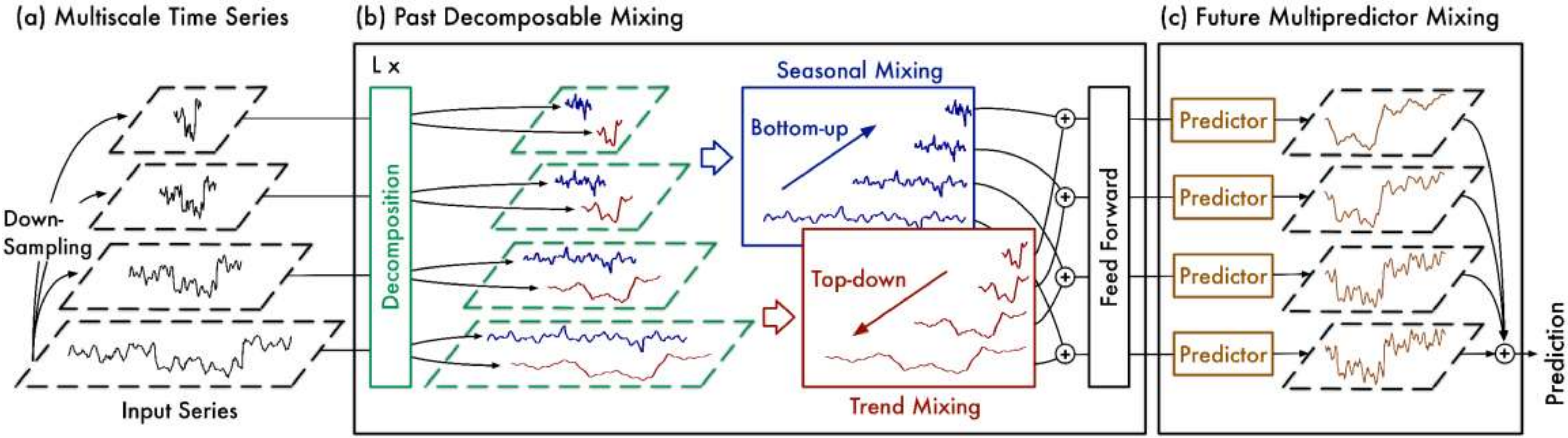
- 目前被广泛认可的处理时间序列的范式：
  - **序列分解**：例如Autoformer，Dlinear，Fedformer
  - **多尺度（周期）分析**：例如TimesNet，WITRAN等
- 不同的采样尺度下呈现出不同的时间变化
  - ✓ 对于时间序列来说：未来的变化是由多个尺度的变化共同决定的

创新点

- 完全基于MLP的模型：**简单而高效**
- 多尺度混合
  - **解纠缠变化**（历史信息提取阶段）：分解多个尺度的变化
  - **互补预测**（预测阶段）：多预测器混合多尺度信息
- 在长时和短时预测任务上都达到了SOTA，且**计算效率高**

模型结构

多尺度季节分量混合



多尺度分解

季节趋势分解

多尺度趋势分量混合

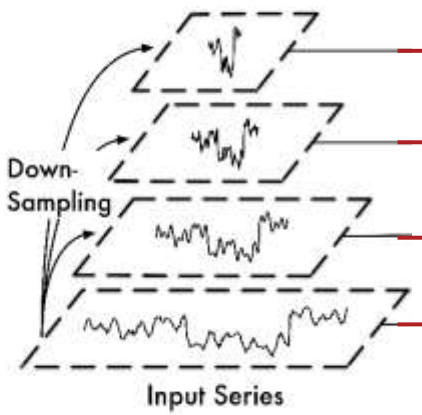
多预测器混合

算法实现：多尺度混合架构

- 用过去 **P** 步预测未来 **F** 步，**C** 表示变量数： $x \in \mathbb{R}^{P \times C}$
- 降采样：分解序列为 **M** 个尺度

$$\mathcal{X} = \{\mathbf{x}_0, \cdots, \mathbf{x}_M\}, \text{where}\sim \mathbf{x}_m \in \mathbb{R}^{\lfloor \frac{P}{2^m} \rfloor \times C}, m \in \{0, \cdots, M\}$$

(a) Multiscale Time Series



以$P = 96$为例（timestep=15min 即24小时）

$$\mathcal{X} = \{\mathbf{x}_0, \cdots, \mathbf{x}_M\}, \text{where}\sim \mathbf{x}_m \in \mathbb{R}^{\lfloor \frac{P}{2^m} \rfloor \times C}, M = 3, m \in \{0, 1, 2, 3\}$$

宏观

$\frac{96}{2^3} = 12$ ----------每8个点取1个平均值 —— x_3

$\frac{96}{2^2} = 24$ ----------每4个点取1个平均值 —— x_2

Avg Pooling

$\frac{96}{2^1} = 48$ ----------每2个点取1个平均值 —— x_1

$\frac{96}{2^0} = 96$ ----------原序列 —— x_0

最细finest

算法实现：多尺度混合架构

- 用过去 **P** 步预测未来 **F** 步，**C** 表示变量数： $x \in \mathbb{R}^{P\times C}$
- 降采样：分解序列为 **M** 个尺度

$$\mathcal{X} = \{\mathbf{x}_0, \cdots, \mathbf{x}_M\}, \text{where}\sim \mathbf{x}_m \in \mathbb{R}^{\left\lfloor \frac{P}{2^m} \right\rfloor \times C}, m \in \{0, \cdots, M\}$$

- 数据嵌入：token嵌入+PE嵌入+时序嵌入， $C \to d_{model}$

$$\mathcal{X}^0 = Embed(\mathcal{X})$$

- **PDM模块**： **L** 层堆叠，在每一层都混合多尺度信息

$$\mathcal{X}^l = \mathrm{PDM}(\mathcal{X}^{l-1}), l \in \{0, \cdots, L\}, \mathcal{X}^{l} = \{\mathrm{x}_0^l, \cdots, \mathrm{x}_M^l\}, \mathrm{x}_m^l \in \mathbb{R}^{\left\lfloor \frac{P}{2^m} \right\rfloor \times d_{\text{model}}}$$

- **FMM模块**：组装提取的多尺度历史信息，生成未来预测

$$\hat{\mathbf{x}} = \mathrm{FMM}(X^L), \quad \hat{\mathbf{x}} \in \mathbb{R}^{F\times C}$$

算法实现：PDM模块（Past Decomposable Mixing）

- **PDM模块**：$\mathcal{X}^l = \mathrm{PDM}(\mathcal{X}^{l-1})$
  - 每个尺度分别做趋势分解

$$\mathcal{X}_{\mathrm{t}} = \mathrm{AvgPool}(\mathrm{Padding}(\mathcal{X}))$$
$$\mathcal{X}_{\mathrm{s}} = \mathcal{X} - \mathcal{X}_{\mathrm{t}}$$

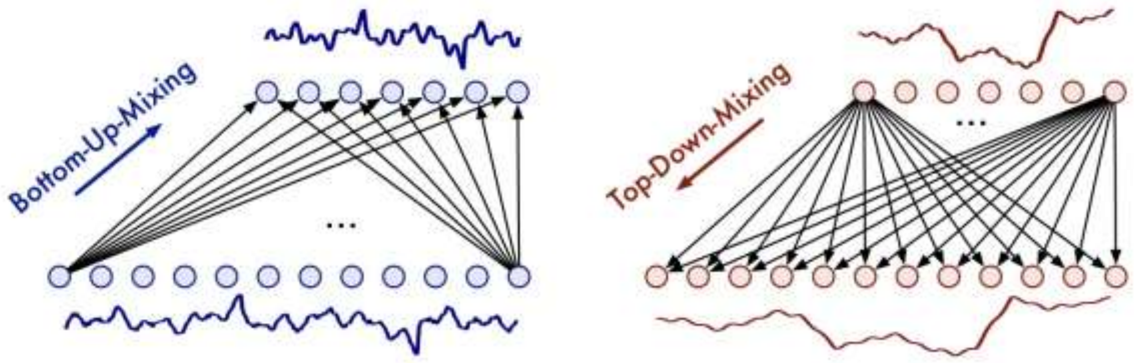
$$\mathbf{s}_m^l, \mathbf{t}_m^l = \mathrm{SeriesDecomp}(\mathbf{x}_m^l), m \in \{0, \cdots, M\},$$

$$\mathcal{S}^l = \{\mathbf{s}_0^l, \cdots, \mathbf{s}_M^l\}$$
$$\mathcal{T}^l = \{\mathbf{t}_0^l, \cdots, \mathbf{t}_M^l\}$$

$$\mathcal{X}^l = \mathcal{X}^{l-1} + \mathrm{FeedForward}\left(\text{S-Mix}\left(\{\mathbf{s}_m^l\}_{m=0}^M\right) + \text{T-Mix}\left(\{\mathbf{t}_m^l\}_{m=0}^M\right)\right)$$

for m: $1 \to M$ do: $\mathbf{s}_m^l = \mathbf{s}_m^l + \text{Bottom-Up-Mixing}(\mathbf{s}_{m-1}^l)$.

for m: $(M-1) \to 0$ do: $\mathbf{t}_m^l = \mathbf{t}_m^l + \text{Top-Down-Mixing}(\mathbf{t}_{m+1}^l)$



(a) Seasonal Mixing (b) Trend Mixing

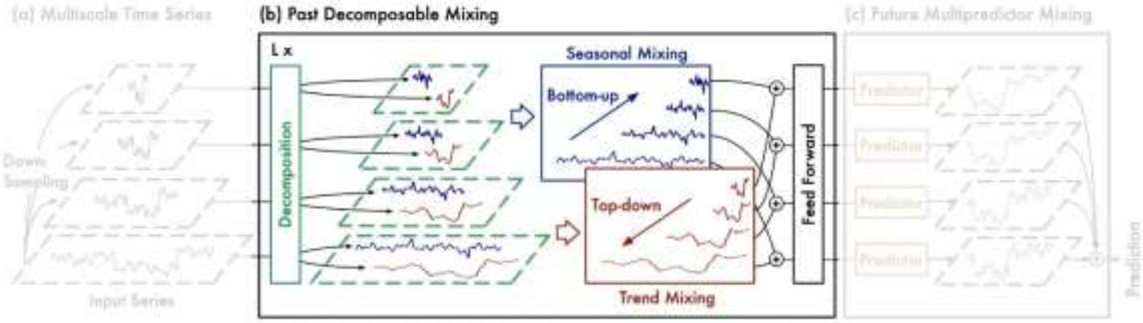
算法实现：PDM模块——Seasonal Mixing

➢ $\mathcal{S}^l = \{\mathbf{s}_0^l, \cdots, \mathbf{s}_M^l\}$ ： Bottom-Up-Mixing layer

- **自底而上合并信息**：残差的方式

for $m: 1 \rightarrow M$ do:

$$\mathrm{s}_m^l = \underline{\mathrm{s}_m^l} + \mathrm{Bottom} - \mathrm{Up} - \mathrm{Mixing}\underline{(\mathrm{s}_{m-1}^l)}$$



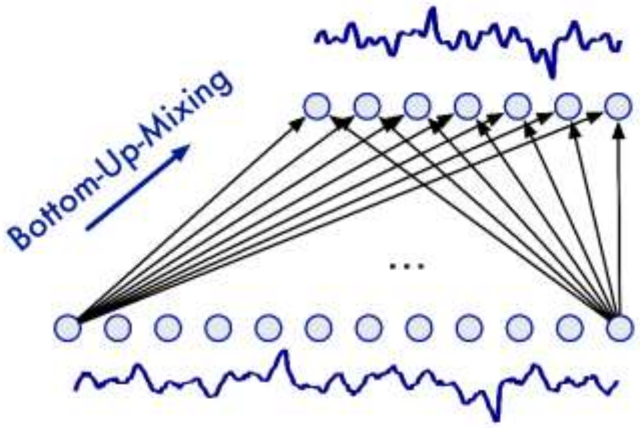
$$\mathbf{s}_m^l, \mathbf{t}_m^l = \mathrm{SeriesDecomp}(\mathbf{x}_m^l), m \in \{0, \cdots, M\},$$

$$\mathcal{X}^l = \mathcal{X}^{l-1} + \mathrm{FeedForward}\left(\text{S-Mix}\left(\{\mathbf{s}_m^l\}_{m=0}^M\right) + \text{T-Mix}\left(\{\mathbf{t}_m^l\}_{m=0}^M\right)\right)$$

$m = 1$ $\quad x_1^l \in \mathbb{R}^{48 \times d_{model}}$ $\quad x_0^l \in \mathbb{R}^{96 \times d_{model}}$

$\quad\quad\quad s_1^l \in \mathbb{R}^{48 \times d_{model}}$ $\quad s_0^l \in \mathbb{R}^{96 \times d_{model}}$

$s_{m=1}^l$: $\quad s_1^l = s_1^l + \underline{Mix}(s_0^l)$ $\left\{ \begin{array}{l} \text{nn.Linear(96, 48)} \\ \text{nn.GELU} \\ \text{nn.Linear(48, 48)} \end{array} \right.$

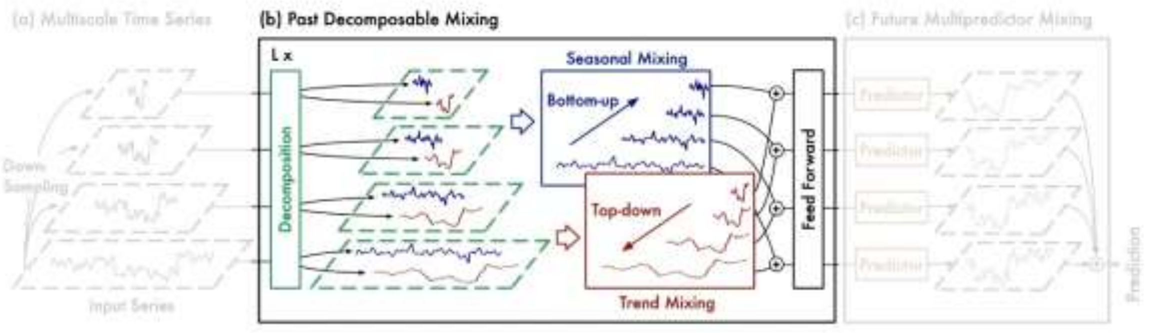


(a) Seasonal Mixing

算法实现：PDM模块——Trend Mixing

- $\mathcal{S}^l=\{\mathbf{s}_0^l,\cdots,\mathbf{s}_M^l\}$ ： Top-Down-Mixing layer

  - **自上而下合并信息**：残差的方式



$$\mathbf{s}_m^l,\mathbf{t}_m^l=\mathrm{SeriesDecomp}(\mathbf{x}_m^l),m\in\{0,\cdots,M\},$$

$$\mathcal{X}^l=\mathcal{X}^{l-1}+\mathrm{FeedForward}\left(\text{S-Mix}\left(\{\mathbf{s}_m^l\}_{m=0}^M\right)+\text{T-Mix}\left(\{\mathbf{t}_m^l\}_{m=0}^M\right)\right)$$

$$\text{for } \ m:(M-1)\rightarrow 0 \ \text{ do:}$$

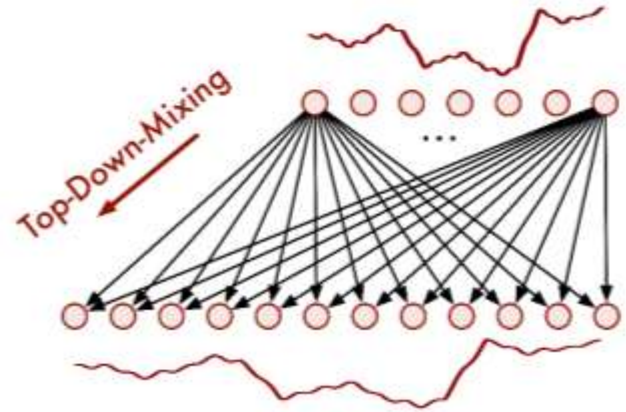
$$\mathrm{t}_m^l=\underline{\mathrm{t}_m^l}+\mathrm{Top}-\mathrm{Down}-\mathrm{Mixing}(\underline{\mathrm{t}_{m+1}^l})$$

$m=1$

$x_2^l\in\mathbb{R}^{24\times d_{model}}$	$x_1^l\in\mathbb{R}^{48\times d_{model}}$	$x_0^l\in\mathbb{R}^{96\times d_{model}}$
$\downarrow$	$\downarrow$	$\downarrow$
$t_2^l\in\mathbb{R}^{24\times d_{model}}$	$t_1^l\in\mathbb{R}^{48\times d_{model}}$	$t_0^l\in\mathbb{R}^{96\times d_{model}}$

$t_{m=1}^l$: $t_1^l=t_1^l+\underline{Mix}(t_2^l)$

nn.Linear(24, 48)
nn.GELU
nn.Linear(48, 48)

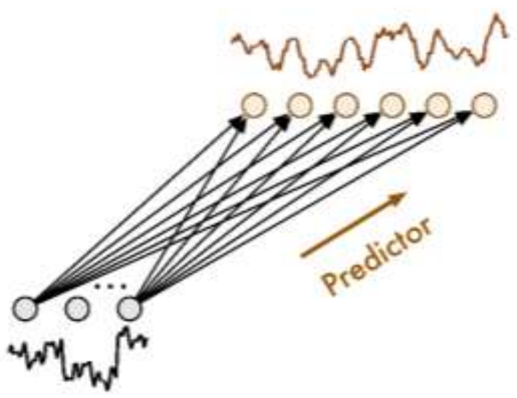


(b) Trend Mixing

算法实现：FMM模块（Future Multipredictor Mixing)

多预测器混合： $\hat{x}_m \in \mathbb{R}^{F\times C}$

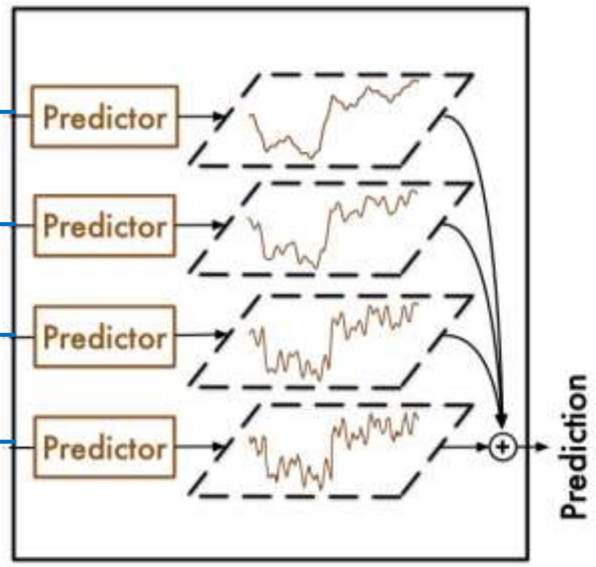
$$\hat{\mathbf{x}}_m = \mathrm{Predictor}_m(\mathbf{x}_m^L), m \in \{0, \cdots, M\}, \hat{\mathbf{x}} = \sum_{m=0}^{M} \hat{\mathbf{x}}_m$$



(c) Future Prediction

任务说明：输入序列长度 **P** = 96， 预测序列长度为 **F** = 192

$$\sum \begin{cases} m=3 & \text{nn.Linear(12, 192)} & d_{model} \to C \\ m=2 & \text{nn.Linear(24, 192)} & d_{model} \to C \\ m=1 & \text{nn.Linear(48, 192)} & d_{model} \to C \\ m=0 & \text{nn.Linear(96, 192)} & d_{model} \to C \end{cases}$$



实验数据集

表 6: 数据集详细描述。数据集大小分为（训练，验证，测试）三部分。

Tasks	Dataset	Dim	Series Length	Dataset Size	Frequency	Forecastability∗	Information
	ETTm1	7	{96, 192, 336, 720}	(34465, 11521, 11521)	15min	0.46	Temperature
	ETTm2	7	{96, 192, 336, 720}	(34465, 11521, 11521)	15min	0.55	Temperature
	ETTh1	7	{96, 192, 336, 720}	(8545, 2881, 2881)	15 min	0.38	Temperature
	ETTh2	7	{96, 192, 336, 720}	(8545, 2881, 2881)	15 min	0.45	Temperature
Long-term	Electricity	321	{96, 192, 336, 720}	(18317, 2633, 5261)	Hourly	0.77	Electricity
Forecasting	Traffic	862	{96, 192, 336, 720}	(12185, 1757, 3509)	Hourly	0.68	Transportation
	Weather	21	{96, 192, 336, 720}	(36792, 5271, 10540)	10 min	0.75	Weather
	Solar-Energy	137	{96, 192, 336, 720}	(36601, 5161, 10417)	10min	0.33	Electricity
	PEMS03	358	12	(15617,5135,5135)	5min	0.65	Transportation
	PEMS04	307	12	(10172,3375,3375)	5min	0.45	Transportation
	PEMS07	883	12	(16911,5622,5622)	5min	0.58	Transportation
Short-term	PEMS08	170	12	(10690,3548,265)	5min	0.52	Transportation
Forecasting	M4-Yearly	1	6	(23000, 0, 23000)	Yearly	0.43	Demographic
	M4-Quarterly	1	8	(24000, 0, 24000)	Quarterly	0.47	Finance
	M4-Monthly	1	18	(48000, 0, 48000)	Monthly	0.44	Industry
	M4-Weakly	1	13	(359, 0, 359)	Weakly	0.43	Macro
	M4-Daily	1	14	(4227, 0, 4227)	Daily	0.44	Micro
	M4-Hourly	1	48	(414, 0, 414)	Hourly	0.46	Other

∗ The forecastability is calculated by one minus the entropy of Fourier decomposition of time series (Goerg, 2013). A larger value indicates better predictability.

Forecastability可预测性:（2013 ForeCA算法）

计算时间序列的傅里叶分解的熵，再用1减去这个熵。值越大，预测能力越好。

实验：长时预测

表 2: 长期预测结果。所有结果均取自4种不同的预测长度，即{96, 192, 336, 720}。MSE或MAE越低，预测效果越好。我们在所有实验中将输入长度固定为96。完整结果见附录表13。

Models	**TimeMixer (Ours)**		PatchTST (2023)		TimesNet (2023a)		Crossformer (2023)		MICN (2023)		FiLM (2022a)		DLinear (2023)		FEDformer (2022b)		Stationary (2022b)		Autoformer (2021)		Informer (2021)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	**0.240**	**0.271**	0.265	0.285	0.251	0.294	0.264	0.320	0.268	0.321	0.271	0.291	0.265	0.315	0.309	0.360	0.288	0.314	0.338	0.382	0.634	0.548
Solar-Energy	**0.216**	**0.280**	0.287	0.333	0.403	0.374	0.406	0.442	0.283	0.358	0.380	0.371	0.330	0.401	0.328	0.383	0.350	0.390	0.586	0.557	0.331	0.381
Electricity	**0.182**	**0.272**	0.216	0.318	0.193	0.304	0.244	0.334	0.196	0.309	0.223	0.302	0.225	0.319	0.214	0.327	0.193	0.296	0.227	0.338	0.311	0.397
Traffic	**0.484**	**0.297**	0.529	0.341	0.620	0.336	0.667	0.426	0.593	0.356	0.637	0.384	0.625	0.383	0.610	0.376	0.624	0.340	0.628	0.379	0.764	0.416
ETTh1	**0.447**	**0.440**	0.516	0.484	0.495	0.450	0.529	0.522	0.475	0.480	0.516	0.483	0.461	0.457	0.498	0.484	0.570	0.537	0.496	0.487	1.040	0.795
ETTh2	**0.364**	**0.395**	0.391	0.411	0.414	0.427	0.942	0.684	0.574	0.531	0.402	0.420	0.563	0.519	0.437	0.449	0.526	0.516	0.450	0.459	4.431	1.729
ETTm1	**0.381**	**0.395**	0.406	0.407	0.400	0.406	0.513	0.495	0.423	0.422	0.411	0.402	0.404	0.408	0.448	0.452	0.481	0.456	0.588	0.517	0.961	0.734
ETTm2	**0.275**	**0.323**	0.290	0.334	0.291	0.333	0.757	0.610	0.353	0.402	0.287	0.329	0.354	0.402	0.305	0.349	0.306	0.347	0.327	0.371	1.410	0.810

实验：短时预测——PEMS多变量数据集

表 3: 在PEMS数据集中,多变量的短期预测结果。 所有输入长度均为96，预测长度为12。MAE、MAPE或RMSE数值越低，表示预测效果越好。

Models		TimeMixer (Ours)	SCINet (2022a)	Crossformer (2023)	PatchTST (2023)	TimesNet (2023a)	MICN (2023)	FiLM (2022a)	DLinear (2023)	FEDformer (2022b)	Stationary (2022b)	Autoformer (2021)	Informer (2021)
PEMS03	MAE	**14.63**	15.97	15.64	18.95	16.41	15.71	21.36	19.70	19.00	17.64	18.08	19.19
	MAPE	**14.54**	15.89	15.74	17.29	15.17	15.67	18.35	18.35	18.57	17.56	18.75	19.58
	RMSE	**23.28**	25.20	25.56	30.15	26.72	24.55	35.07	32.35	30.05	28.37	27.82	32.70
PEMS04	MAE	**19.21**	20.35	20.38	24.86	21.63	21.62	26.74	24.62	26.51	22.34	25.00	22.05
	MAPE	**12.53**	12.84	12.84	16.65	13.15	13.53	16.46	16.12	16.76	14.85	16.70	14.88
	RMSE	**30.92**	32.31	32.41	40.46	34.90	34.39	42.86	39.51	41.81	35.47	38.02	36.20
PEMS07	MAE	**20.57**	22.79	22.54	27.87	25.12	22.28	28.76	28.65	27.92	26.02	26.92	27.26
	MAPE	**8.62**	9.41	9.38	12.69	10.60	9.57	11.21	12.15	12.29	11.75	11.83	11.63
	RMSE	**33.59**	35.61	35.49	42.56	40.71	35.40	45.85	45.02	42.29	42.34	40.60	45.81
PEMS08	MAE	**15.22**	17.38	17.56	20.35	19.01	17.76	22.11	20.26	20.56	19.29	20.47	20.96
	MAPE	**9.67**	10.80	10.92	13.15	11.83	10.76	12.81	12.09	12.41	12.21	12.27	13.20
	RMSE	**24.26**	27.34	27.21	31.04	30.65	27.26	35.13	32.38	32.97	38.62	31.52	30.61

实验：短时预测——M4单变量数据集

表 4: 单变量M4数据集的短期预测结果。所有预测长度在[6, 48]范围内。SMAPE、MASE和OWA值越低，预测效果越好。*former名称后的*表示。*Stationary*表示非平稳Transformer。

Models		TimeMixer (Ours)	TimesNet (2023a)	N-HiTS (2023)	N-BEATS* (2019)	SCINet (2022a)	PatchTST (2023)	MICN (2023)	FiLM (2022a)	LightTS (2022)	DLinear (2023)	FED. (2022b)	Stationary (2022b)	Auto. (2021)	Pyra. (2021)	In. (2021)
Yearly	SMAPE	**13.206**	13.387	13.418	13.436	18.605	16.463	25.022	17.431	14.247	16.965	13.728	13.717	13.974	15.530	14.727
	MASE	**2.916**	2.996	3.045	3.043	4.471	3.967	7.162	4.043	3.109	4.283	3.048	3.078	3.134	3.711	3.418
	OWA	**0.776**	0.786	0.793	0.794	1.132	1.003	1.667	1.042	0.827	1.058	0.803	0.807	0.822	0.942	0.881
Quarterly	SMAPE	**9.996**	10.100	10.202	10.124	14.871	10.644	15.214	12.925	11.364	12.145	10.792	10.958	11.338	15.449	11.360
	MASE	**1.166**	1.182	1.194	1.169	2.054	1.278	1.963	1.664	1.328	1.520	1.283	1.325	1.365	2.350	1.401
	OWA	**0.825**	0.890	0.899	0.886	1.424	0.949	1.407	1.193	1.000	1.106	0.958	0.981	1.012	1.558	1.027
Monthly	SMAPE	**12.605**	12.670	12.791	12.677	14.925	13.399	16.943	15.407	14.014	13.514	14.260	13.917	13.958	17.642	14.062
	MASE	**0.919**	0.933	0.969	0.937	1.131	1.031	1.442	1.298	1.053	1.037	1.102	1.097	1.103	1.913	1.141
	OWA	**0.869**	0.878	0.899	0.880	1.027	0.949	1.265	1.144	0.981	0.956	1.012	0.998	1.002	1.511	1.024
Others	SMAPE	**4.564**	4.891	5.061	4.925	16.655	6.558	41.985	7.134	15.880	6.709	4.954	6.302	5.485	24.786	24.460
	MASE	**3.115**	3.302	3.216	3.391	15.034	4.511	62.734	5.09	11.434	4.953	3.264	4.064	3.865	18.581	20.960
	OWA	**0.982**	1.035	1.040	1.053	4.123	1.401	14.313	1.553	3.474	1.487	1.036	1.304	1.187	5.538	5.879
Weighted Average	SMAPE	**11.723**	11.829	11.927	11.851	15.542	13.152	19.638	14.863	13.525	13.639	12.840	12.780	12.909	16.987	14.086
	MASE	**1.559**	1.585	1.613	1.559	2.816	1.945	5.947	2.207	2.111	2.095	1.701	1.756	1.771	3.265	2.718
	OWA	**0.840**	0.851	0.861	0.855	1.309	0.998	2.279	1.125	1.051	1.051	0.918	0.930	0.939	1.480	1.230

∗ The original paper of N-BEATS (2019) adopts a special ensemble method to promote the performance. For fair comparisons, we remove the ensemble and only compare the pure forecasting models.

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

$$\mathrm{MASE} = \frac{1}{F}\sum_{i=1}^{F}\frac{|\mathbf{X}_i - \widehat{\mathbf{X}}_i|}{\frac{1}{F-s}\sum_{j=s+1}^{F}|\mathbf{X}_j - \mathbf{X}_{j-s}|}$$

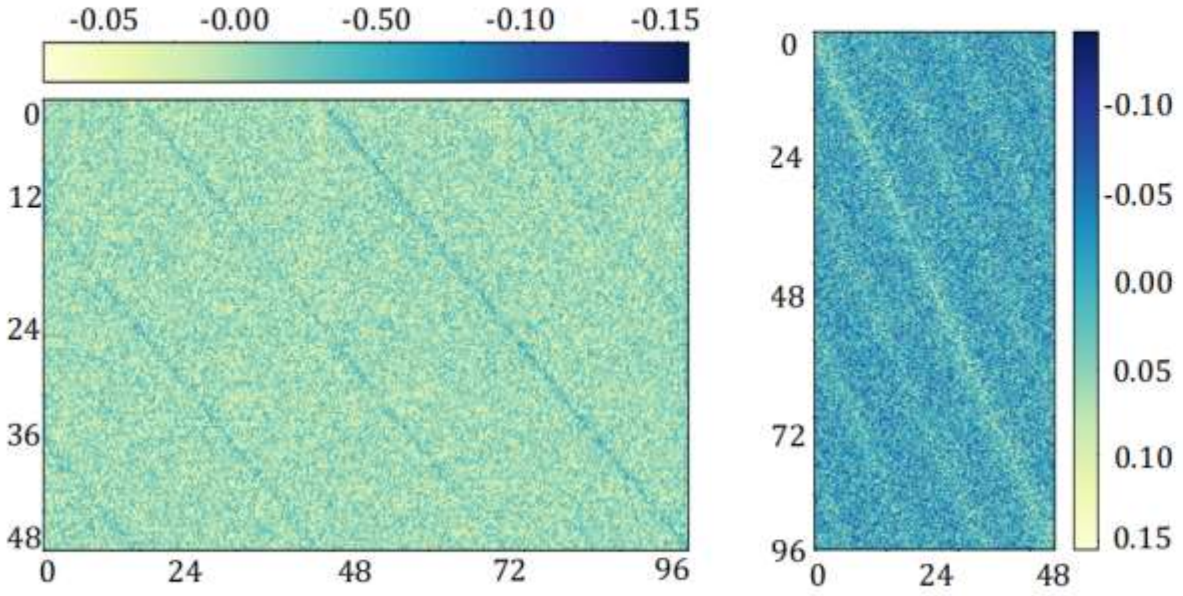
$$\mathrm{OWA} = \frac{1}{2}\left[\frac{\mathrm{SMAPE}}{\mathrm{SMAPE}_{\text{Naïve2}}} + \frac{\mathrm{MASE}}{\mathrm{MASE}_{\text{Naive2}}}\right]$$

实验：消融实验

表 5：对M4、PEMS04数据集和ETTm1的predict-336设置下的PDM（分解、季节混合、趋势混合）和FMM块进行消融研究。↗表示自底向上的混合，↙表示自顶向下的。勾选标记✓表示包含某组件，错误标记×表示不包含。①是TimeMixer的官方设计（见附录F获取完整消融结果）。

Case	Decompose	Past mixing		Future mixing	M4			PEMS04			ETTm1	
		Seasonal	Trend	Multipredictor	SMAPE	MASE	OWA	MAE	MAPE	RMSE	MSE	MAE
①	✓	↗	↙	✓	**11.723**	**1.559**	**0.840**	**19.21**	**12.53**	**30.92**	**0.390**	**0.404**
②	✓	↗	↙	×	12.503	1.634	0.925	21.67	13.45	34.89	0.402	0.415
③	✓	×	↙	✓	13.051	1.676	0.962	24.49	16.28	38.79	0.411	0.427
④	✓	↗	×	✓	12.911	1.655	0.941	22.91	15.02	37.04	0.405	0.414
⑤	✓	↙	↙	✓	12.008	1.628	0.871	20.78	13.02	32.47	0.392	0.413
⑥	✓	↗	↗	✓	11.978	1.626	0.859	21.09	13.78	33.11	0.396	0.415
⑦	✓	↙	↗	✓	13.012	1.657	0.954	22.27	15.14	34.67	0.412	0.429
⑧	×	↗		✓	11.975	1.617	0.851	21.51	13.47	34.81	0.395	0.408
⑨	×	↙		✓	11.973	1.622	0.850	21.79	14.03	35.23	0.393	0.406
⑩	×	×		✓	12.468	1.671	0.916	24.87	16.66	39.48	0.405	0.412

实验：季节性和趋势混合可视化



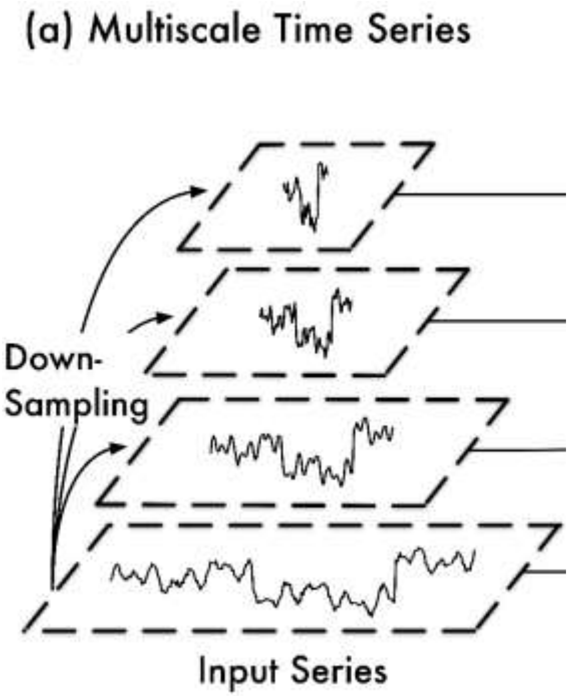
(a) Seasonal Mixing Weights
(bottom-up: from 96 to 48)

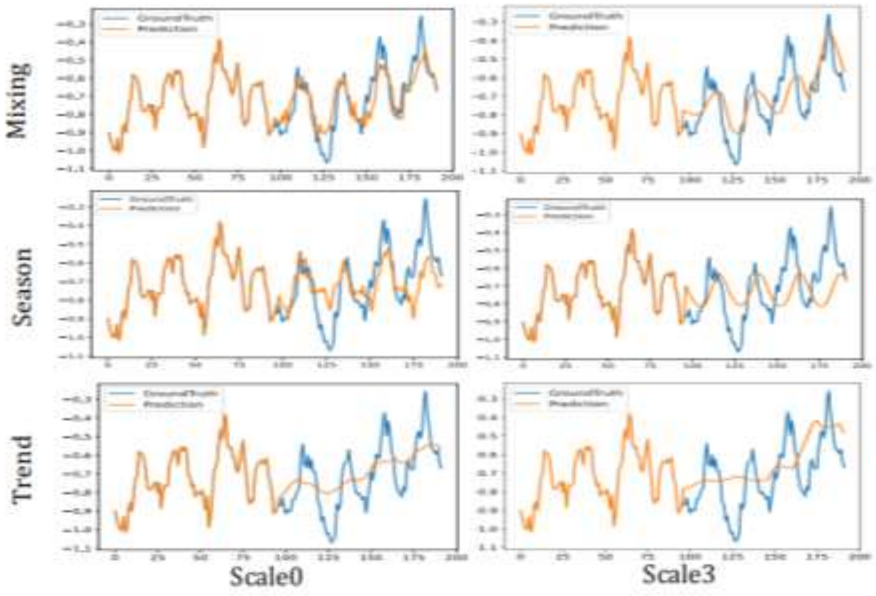
(b) Trend Mixing Weights
(top-down: from 48 to 96)

季节混合层呈现周期性变化
((a)中重复的蓝线)

趋势混合层以局部聚集为主
((b)中占主导地位的对角线黄线)

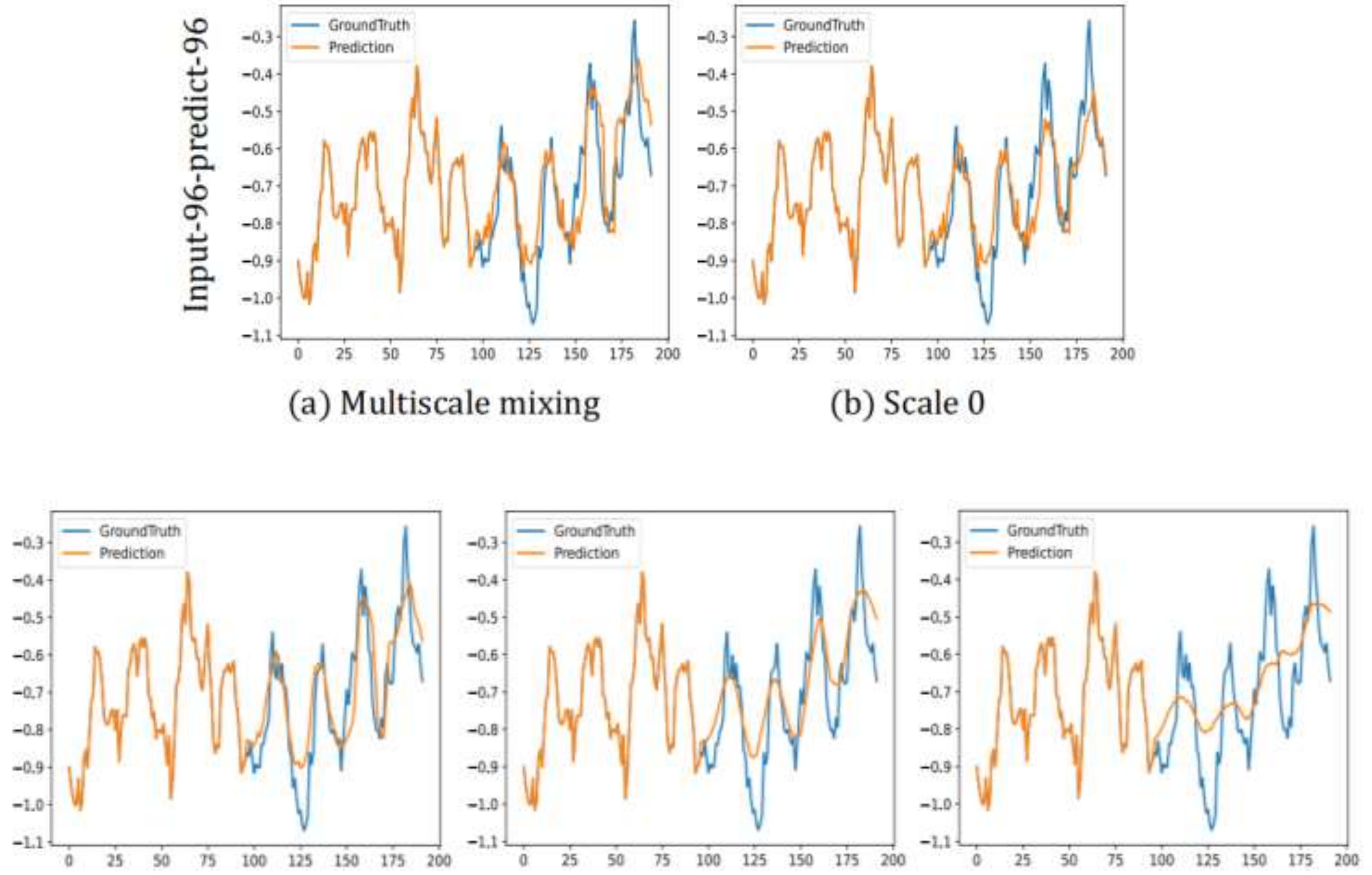
实验：季节性和趋势混合可视化





(c) Multiscale Season-trend Predictions
(input-96-predict-96)

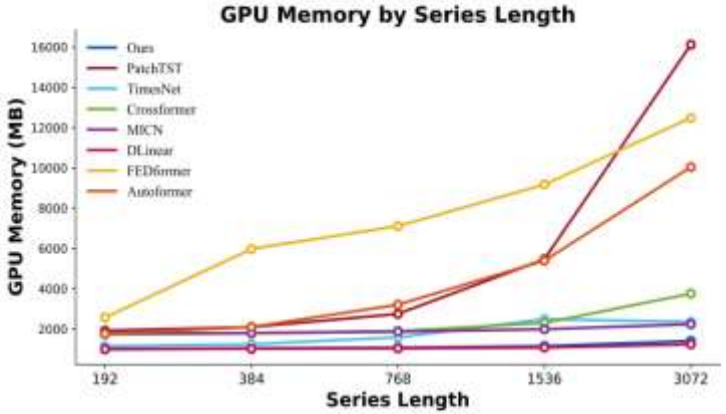
实验：多预测器可视化



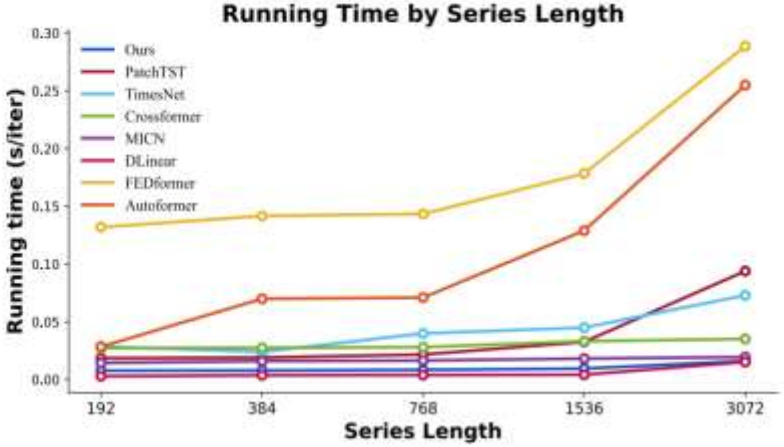
(a) Multiscale mixing (b) Scale 0

(c) Scale 1 (d) Scale 2 (e) Scale 3

实验：效率分析



(a) Memory Efficiency Analysis



(b) Running Time Efficiency Analysis

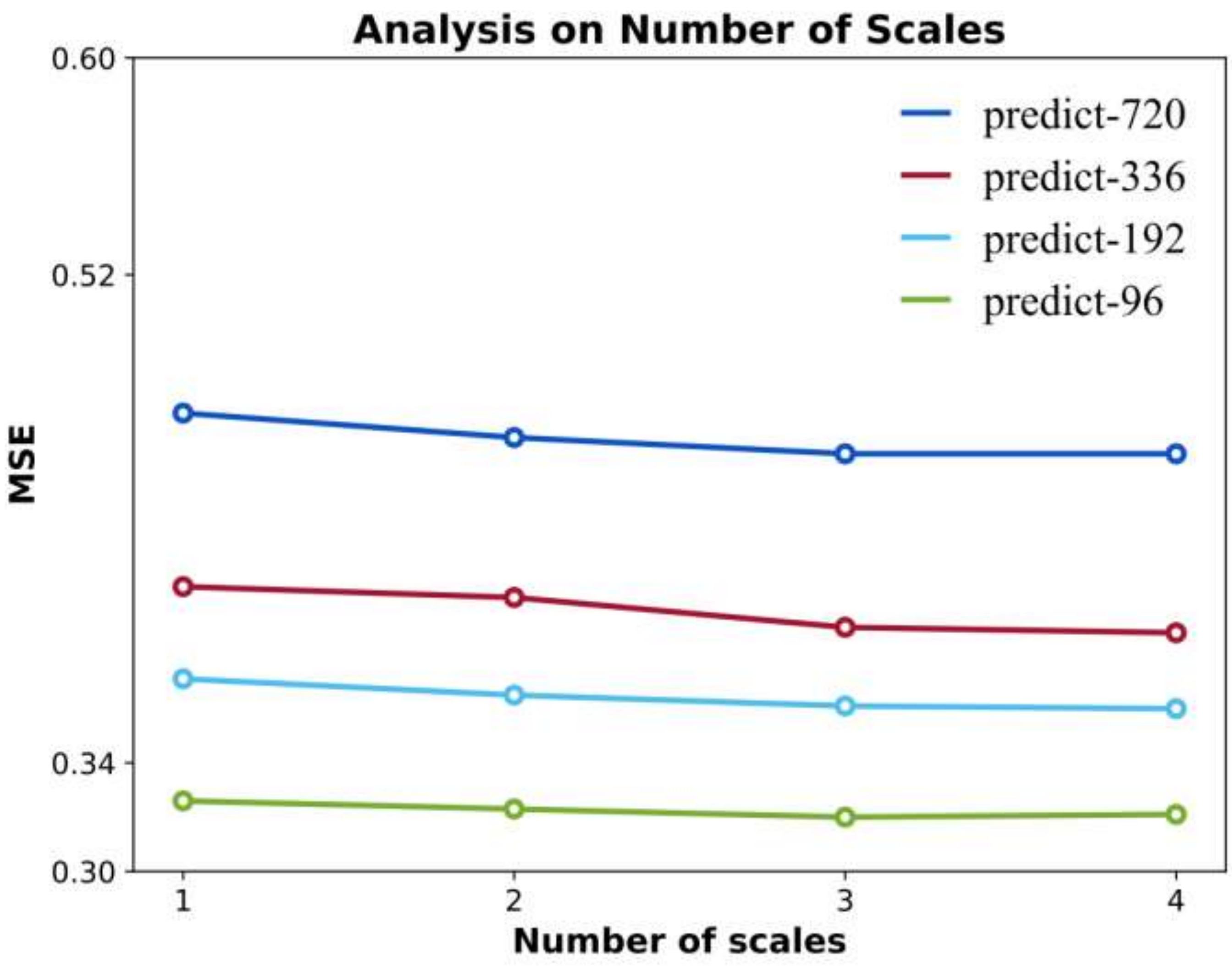
Table 8: The GPU memory (MiB) and speed (running time, s/iter) of each model.

Series Length	192		384		768		1536		3072	
Models	Mem	Speed	Mem	Speed	Mem	Speed	Mem	Speed	Mem	Speed
TimeMixer (**Ours**)	1003	0.007	1043	0.007	1075	0.008	1151	0.009	1411	0.016
PatchTST (2023)	1919	0.018	2097	0.019	2749	0.021	5465	0.032	16119	0.094
TimesNet (2023a)	1148	0.028	1245	0.024	1585	0.042	2491	0.045	2353	0.073
Crossformer (2023)	1737	0.027	1799	0.027	1895	0.028	2303	0.033	3759	0.035
MICN (2023)	1771	0.014	1801	0.016	1873	0.017	1991	0.018	2239	0.020
DLinear (2023)	1001	0.002	1021	0.003	1041	0.003	1081	0.0.004	1239	0.015
FEDFormer (2022b)	2567	0.132	5977	0.141	7111	0.143	9173	0.178	12485	0.288
Autoformer (2021)	1761	0.028	2101	0.070	3209	0.071	5395	0.129	10043	0.255

实验：尺度数分析

考虑性能增益

M设置为：长期预测为3，短期预测为1



实验：尺度数分析

1. 不同的趋势分解方法
   - DFT增强季节趋势分解

表 18：在M4、PEMS04和ETTm1的predict-336设置中的替代分解方法。

Decomposition methods	ETTm1		M4			PEMS04		
	MSE	MAE	SMAPE	MASE	OWA	MAE	MAPE	RMSE
DFT-based high- and low-frequency decomposition	0.392	0.404	12.054	1.632	0.862	19.83	12.74	31.48
DFT-based season-trend decomposition	0.383	0.399	11.673	1.536	0.824	18.91	12.27	29.47
Moving-average-based season-trend decomposition (TimeMixer)	0.390	0.404	11.723	1.559	0.840	19.21	12.53	30.92

2. 不同的降采样方法
   - 一维卷积

表 19：在M4、PEMS04和ETTm1的predict-336设置中,对替代降采样方法的探讨。

Downsampling methods	ETTm1		M4			PEMS04		
	MSE	MAE	SMAPE	MASE	OWA	MAE	MAPE	RMSE
Moving average	0.390	0.404	11.723	1.559	0.840	19.21	12.53	30.92
1D convolutions with stride as 2	0.387	0.401	11.682	1.542	0.831	19.04	12.17	29.88

3. 不同的集成策略
   - 差别不大

表 20：在M4、PEMS04和ETTm1的predict-336设置中的替代ensemble策略。

Ensemble strategies	ETTm1		M4			PEMS04		
	MSE	MAE	SMAPE	MASE	OWA	MAE	MAPE	RMSE
Sum ensemble	0.390	0.404	11.723	1.559	0.840	19.21	12.53	30.92
Average ensemble	0.391	0.407	11.742	1.573	0.851	19.17	12.45	30.88

THANK FOR ATTENTION

厚德　求真　励学　笃行