TimeMixer++: A General Time Series Pattern Machine for Universal Predictive Analysis

Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju,

Zhixuan Chu, Ming Jin

Griffith University

The Hong Kong University of Science and Technology (Guangzhou)

Massachusetts Institute of Technology

Zhejiang University

April 25, 2025, Singapore

ICLR 2025

Time Series Data

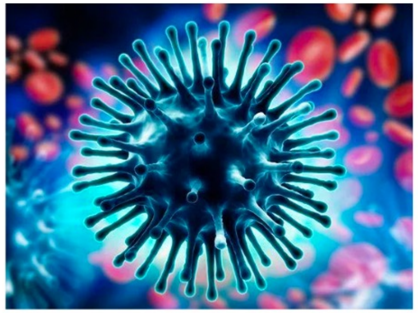
- **Time series are a crucial data modality**

Energy Consumption

Traffic Flow

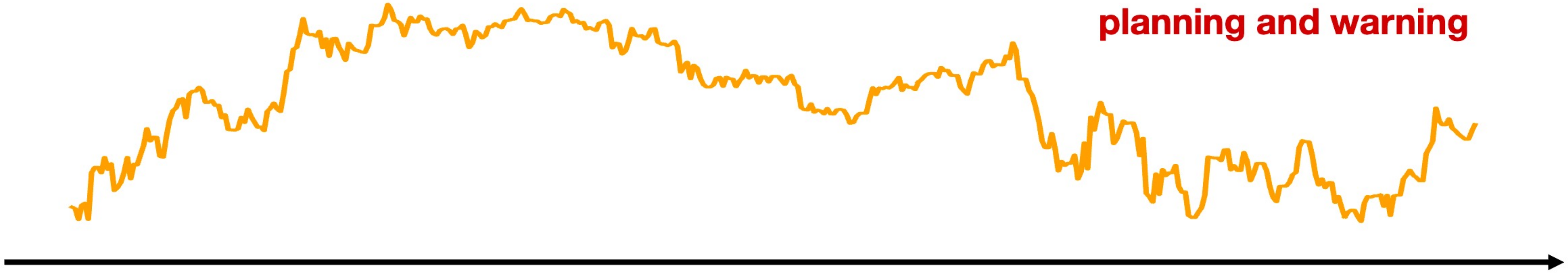
Economic Changes

Weather Variations



Disease Propagation

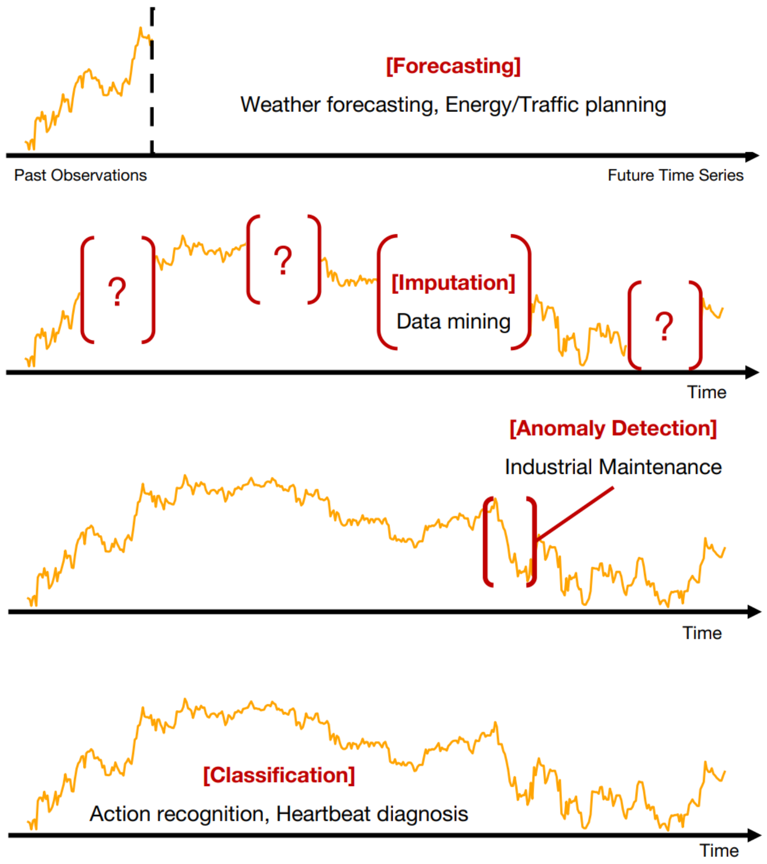
Predict the future for planning and warning

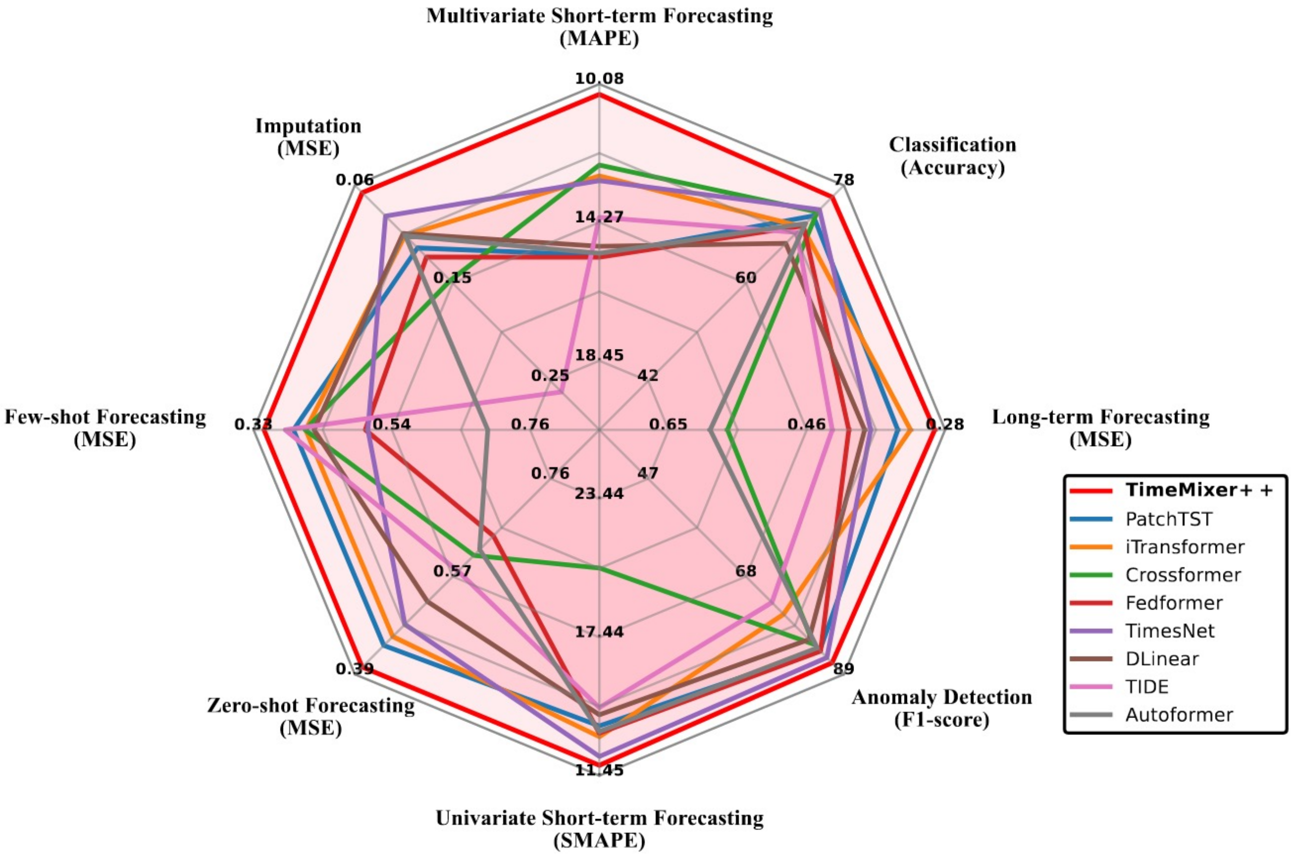


(Figures in this page are generated by GPT-4o)

Time Series Intelligence

- **Unlocking general-purposed time series analysis**

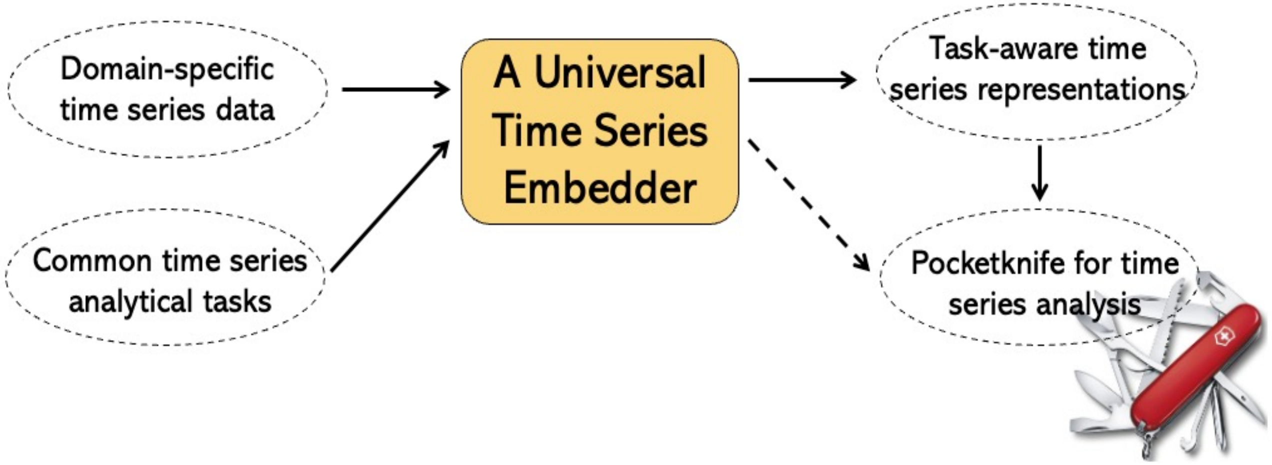




xx

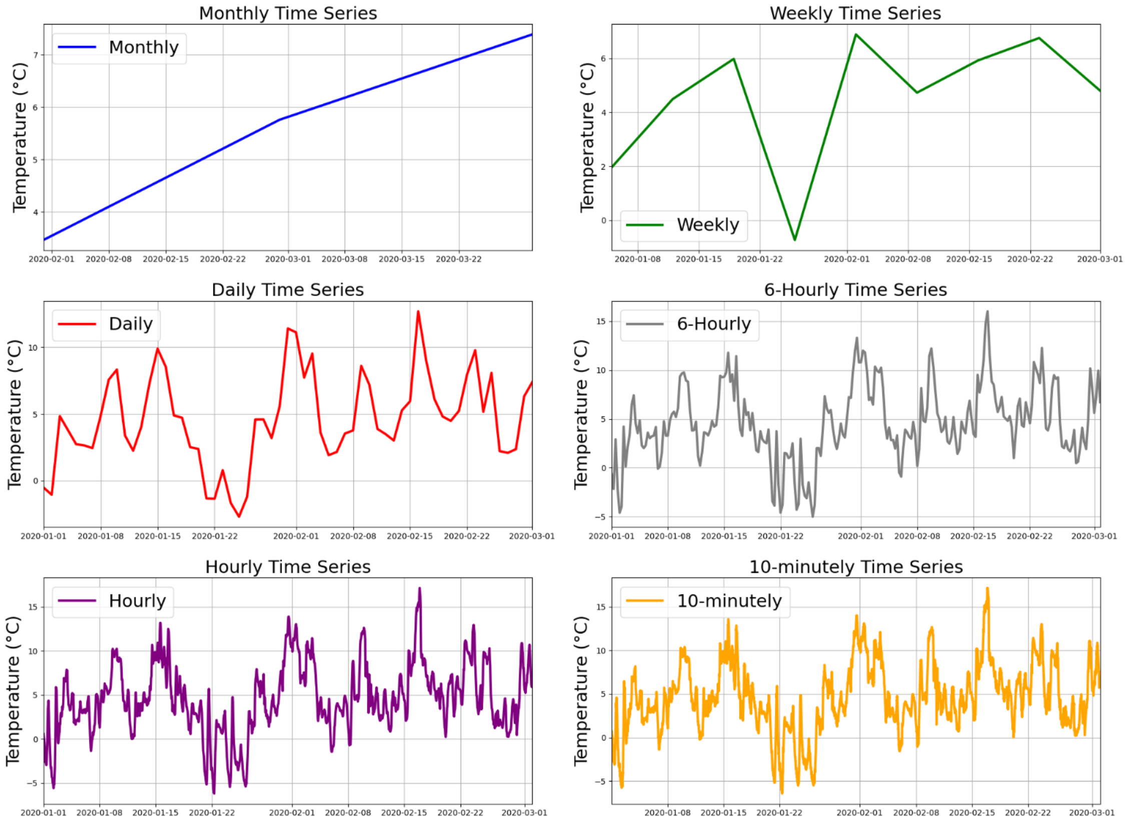
Time Series Intelligence

- **Introducing Time Series Pattern Machine (TSPM)**

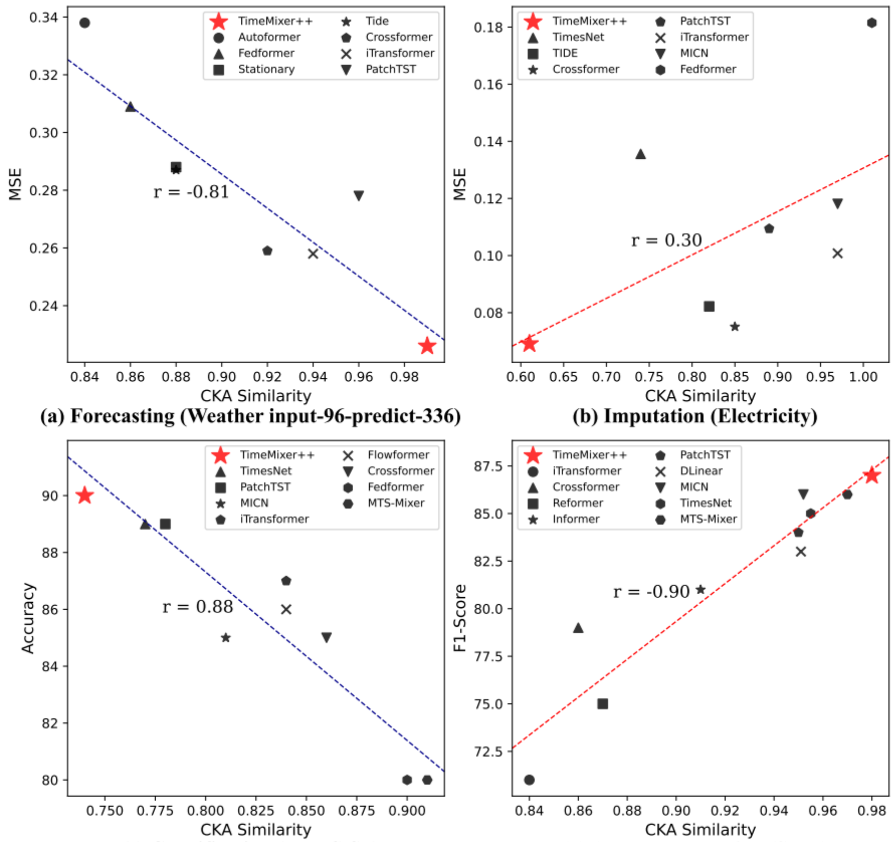


Time Series Intelligence

- **What capabilities must a model possess to function as a TSPM?**



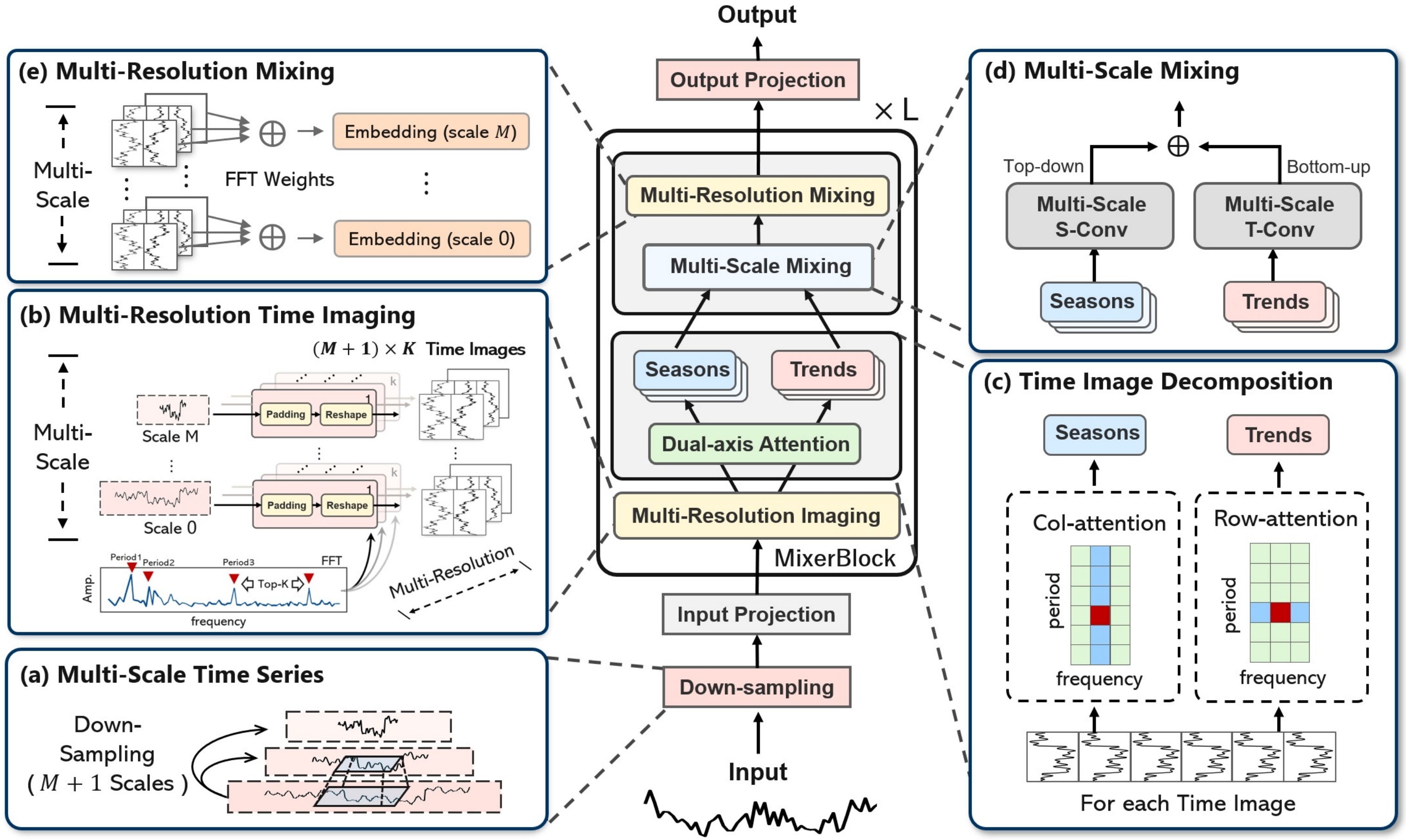
The multi-scale and multi-periodicity characteristics of the Weather Dataset.



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

Solution

- **Introducing TimeMixer++**
  - Transforming time series into a series of time series images via **time imaging**
  - Capturing both **multi-scale and multi-periodicity dynamics** in latent space



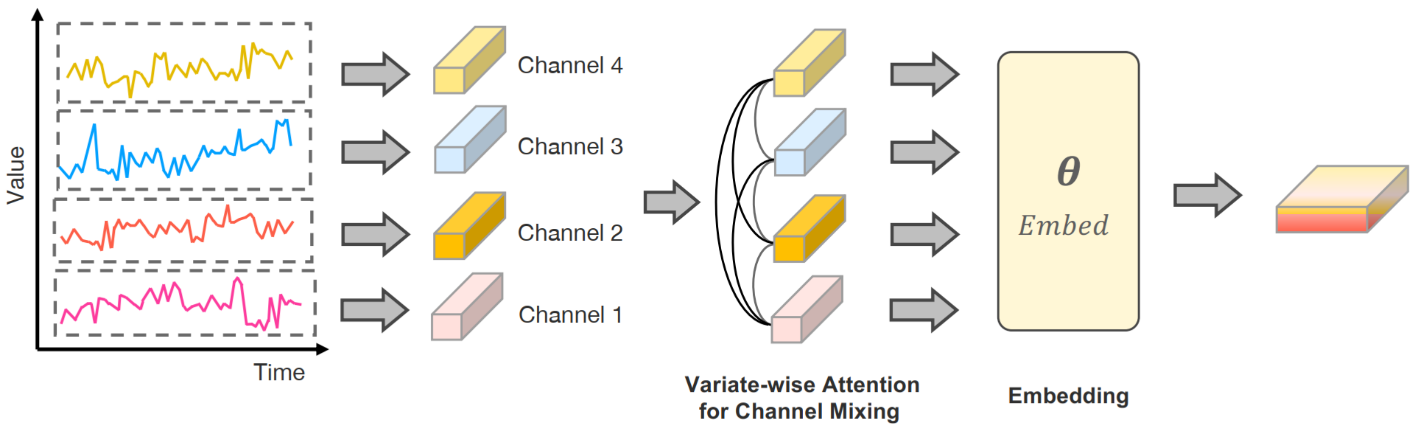
TimeMixer++

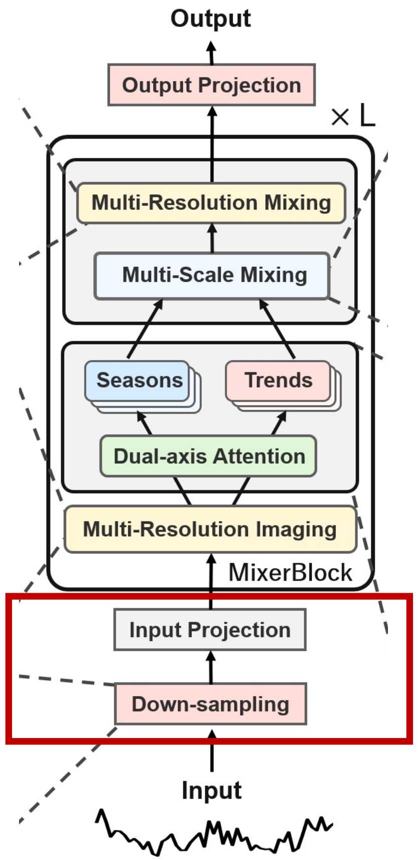
- **Multi-Scale Time Series and Input Projection**
  - We generate a multi-scale representation by using Conv1D for **down-sampling**
  - We adopt a **channel mixing** strategy using variate-wise attention at the coarsest scale
  - Each scale is embedded separately

$$\mathbf{x}_m = \mathrm{Conv}(\mathbf{x}_{m-1}, \mathrm{stride}=2), \quad m \in \{1, \cdots, M\} \tag{1}$$

$$\mathbf{x}_M = \text{Channel-Attn}(\mathbf{Q}_M, \mathbf{K}_M, \mathbf{V}_M) \tag{2}$$

$$\mathcal{X}^0 = \{\mathbf{x}_0^0, \cdots, \mathbf{x}_M^0\} = \mathrm{Embed}(\mathcal{X}_{init})$$

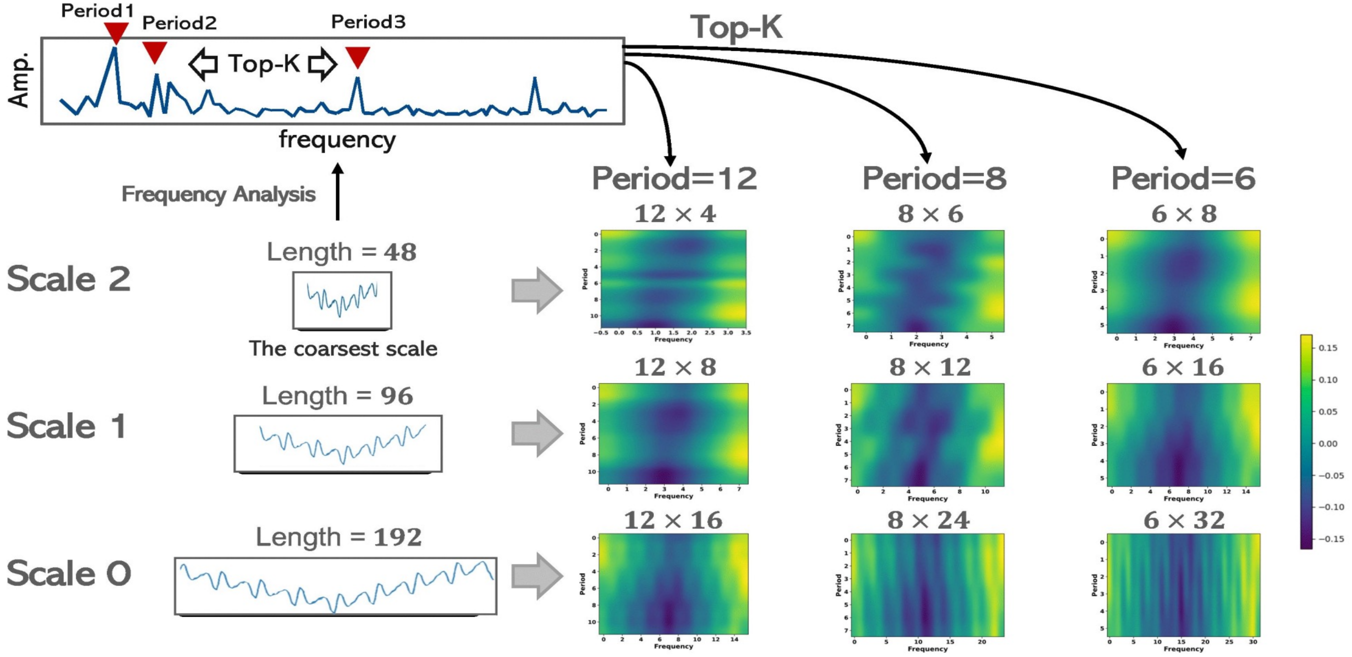


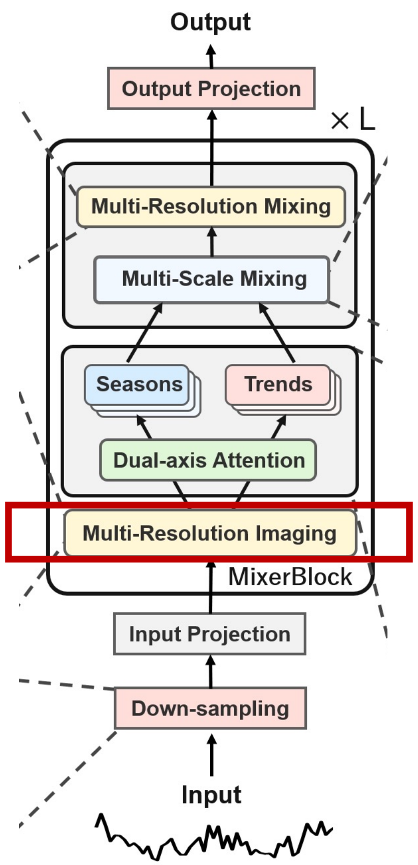


MixerBlock

- **Multi-Resolution Time Series Imaging**
  - We convert the input into **multi-resolution time images** with frequency analysis

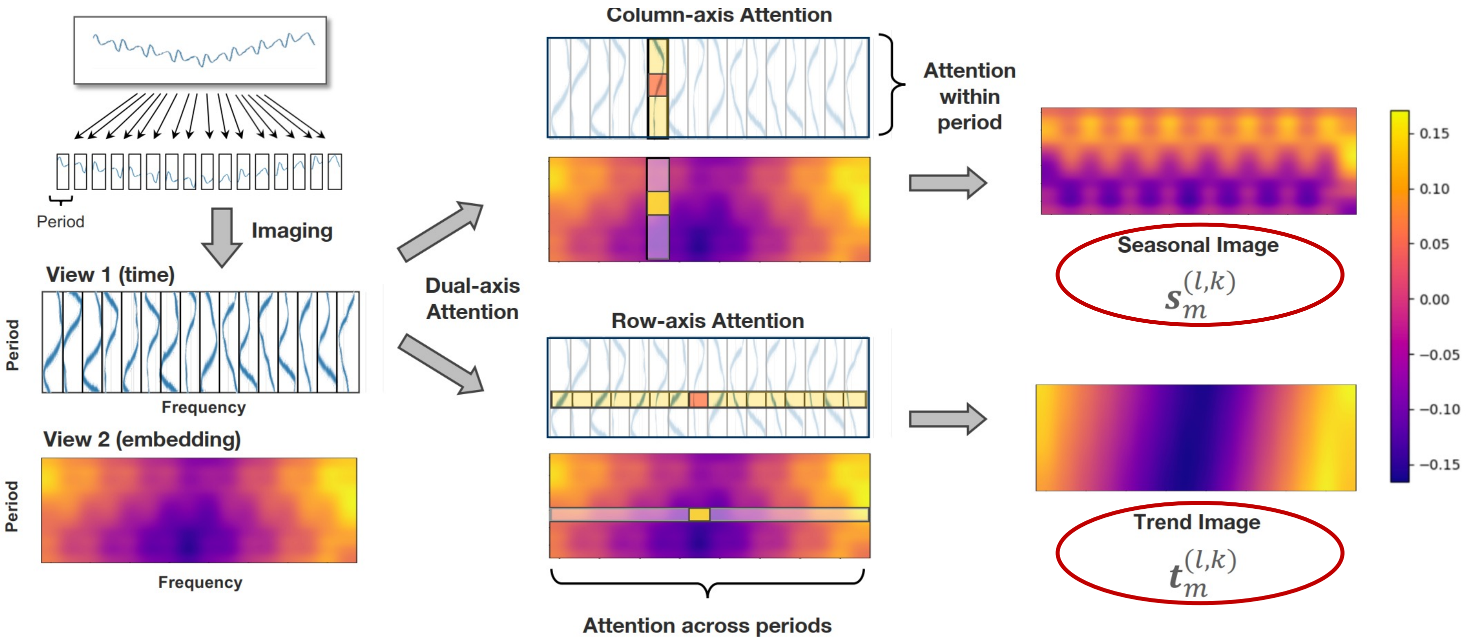
$$\mathrm{MRTI}(\mathcal{X}^l) = \{\mathcal{Z}_m^l\}_{m=0}^M = \left\{\mathbf{z}_m^{(l,k)} \mid m = 0, \ldots, M;\ k = 1, \ldots, K\right\}$$
$$= \left\{\underset{1D\to 2D}{\mathrm{Reshape}_{m,k}}(\mathrm{Padding}_{m,k}(\mathbf{x}_m^l)) \mid m = 0, \ldots, M;\ k = 1, \ldots, K\right\}, \quad (6)$$

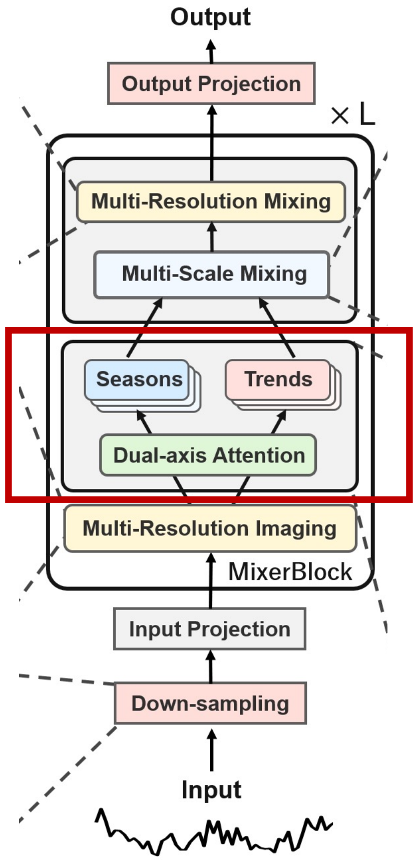




MixerBlock

- **Time Image Decomposition**
  - Columns represent time series segments within each period, while the rows track consistent time points across periods
  - Column-axis attention captures **seasonality within period**, and row-axis attention extracts **trend across periods**





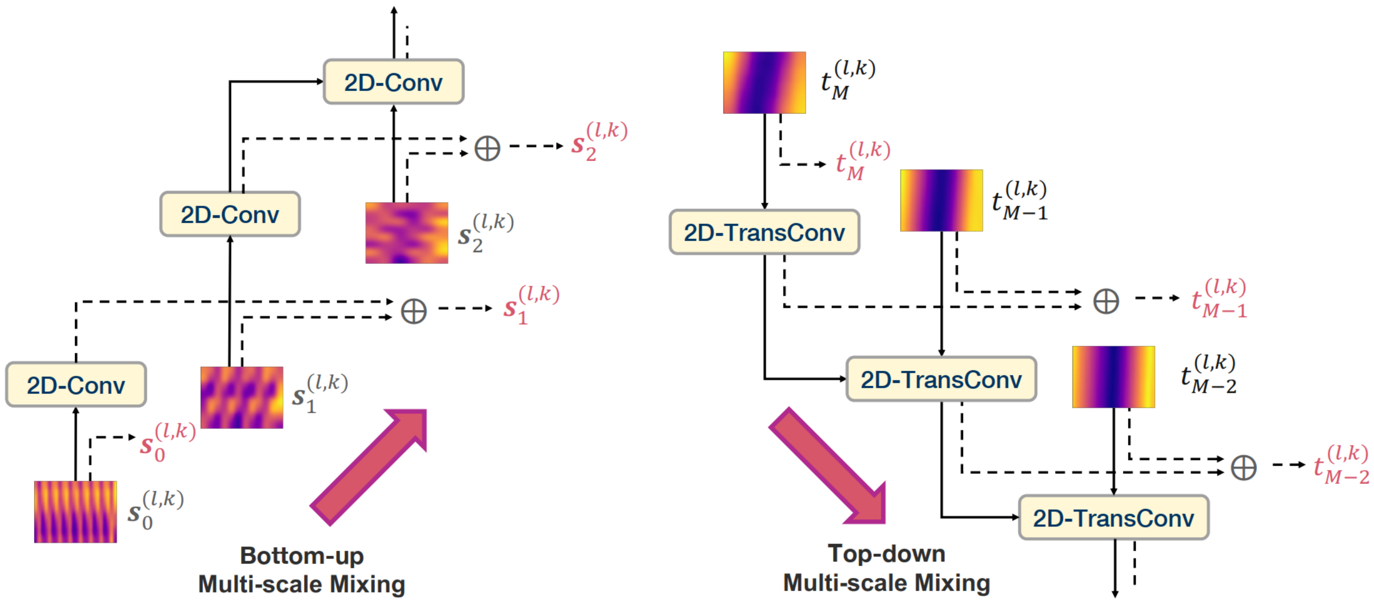
MixerBlock

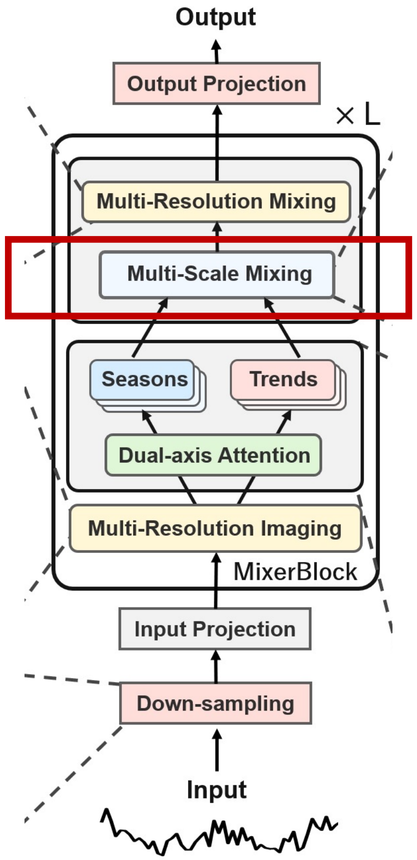
- **Multi-Scale Mixing**
  - Mixing the seasonal patterns from fine-scale to coarse-scale **(Bottom-up)**

$$\text{for } m: 1 \to M \text{ do:} \quad \mathbf{s}_m^{(l,k)} = \mathbf{s}_m^{(l,k)} + \text{2D-Conv}(\mathbf{s}_{m-1}^{(l,k)}), \tag{8}$$

  - Mixing the trend patterns from coarse-scale to fine-scale **(Top-down)**

$$\text{for } m: M-1 \to 0 \text{ do:} \quad \mathbf{t}_m^{(l,k)} = \mathbf{t}_m^{(l,k)} + \text{2D-TransConv}(\mathbf{t}_{m+1}^{(l,k)}) \tag{9}$$

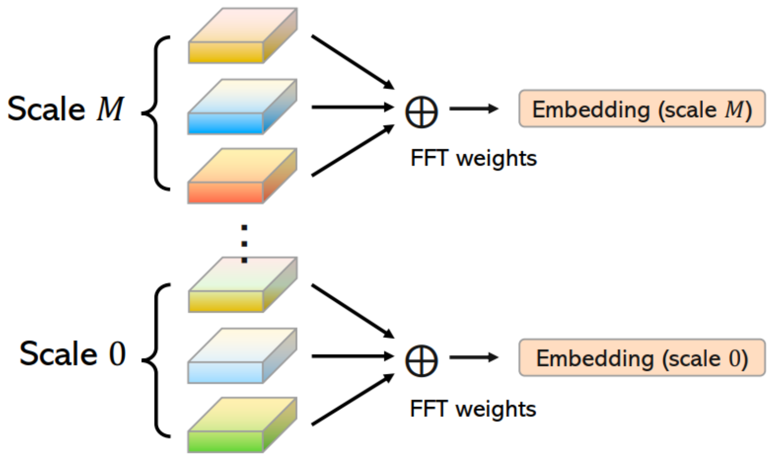


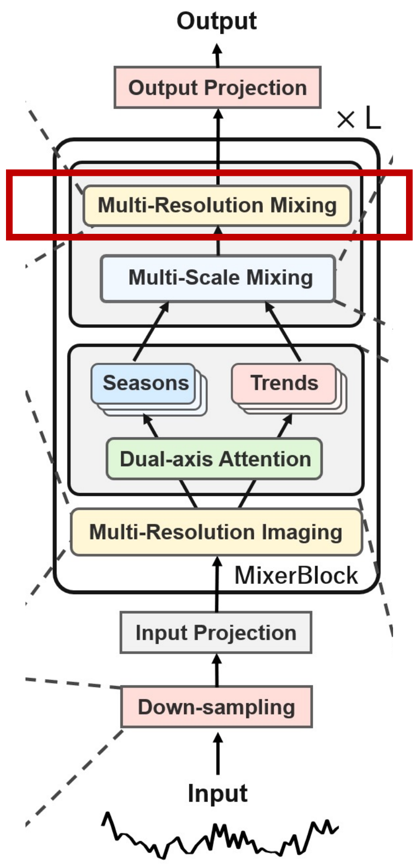


MixerBlock

- **Multi-Resolution Mixing**
  - Finally, at each scale, we mix K periods adaptively

$$\{\hat{\mathbf{A}}_{f_k}\}_{k=1}^{K} = \mathrm{Softmax}(\{\mathbf{A}_{f_k}\}_{k=1}^{K}), \quad \mathbf{x}_m^l = \sum_{k=1}^{K} \hat{\mathbf{A}}_{f_k} \circ \mathbf{z}_m^{(l,k)}, \quad m \in \{0, \cdots, M\} \qquad (11)$$





Main Results

- **Forecasting**

Long-term forecasting results. We average the results across 4 prediction lengths: {96, 192, 336, 720}.

Models	**TimeMixer++ (Ours)**		TimeMixer (2024)		iTransformer (2023)		PatchTST (2023)		Crossformer (2023)		TiDE (2023b)		TimesNet (2023)		DLinear (2023)		SCINet (2022a)		FEDformer (2022b)		Stationary (2022c)		Autoformer (2021)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	**0.165**	**0.253**	0.182	0.272	0.178	0.270	0.205	0.290	0.244	0.334	0.251	0.344	0.192	0.295	0.212	0.300	0.268	0.365	0.214	0.327	0.193	0.296	0.227	0.338
ETT (Avg)	**0.349**	0.399	0.367	0.388	0.383	**0.377**	0.381	0.397	0.685	0.578	0.482	0.470	0.391	0.404	0.442	0.444	0.689	0.597	0.408	0.428	0.471	0.464	0.465	0.459
Exchange	0.357	0.391	0.391	0.453	0.378	**0.360**	0.403	0.404	0.940	0.707	0.370	0.413	0.416	0.443	**0.354**	0.414	0.750	0.626	0.519	0.429	0.461	0.454	0.613	0.539
Traffic	**0.416**	**0.264**	0.484	0.297	0.428	0.282	0.481	0.304	0.550	0.304	0.760	0.473	0.620	0.336	0.625	0.383	0.804	0.509	0.610	0.376	0.624	0.340	0.628	0.379
Weather	**0.226**	**0.262**	0.240	0.271	0.258	0.278	0.259	0.281	0.259	0.315	0.271	0.320	0.259	0.287	0.265	0.317	0.292	0.363	0.309	0.360	0.288	0.314	0.338	0.382
Solar-Energy	**0.203**	**0.238**	0.216	0.280	0.233	0.262	0.270	0.307	0.641	0.639	0.347	0.417	0.301	0.319	0.330	0.401	0.282	0.375	0.291	0.381	0.261	0.381	0.885	0.711

Main Results

- **Forecasting**

Results of **multivariate short-term forecasting**, averaged across all PEMS datasets.

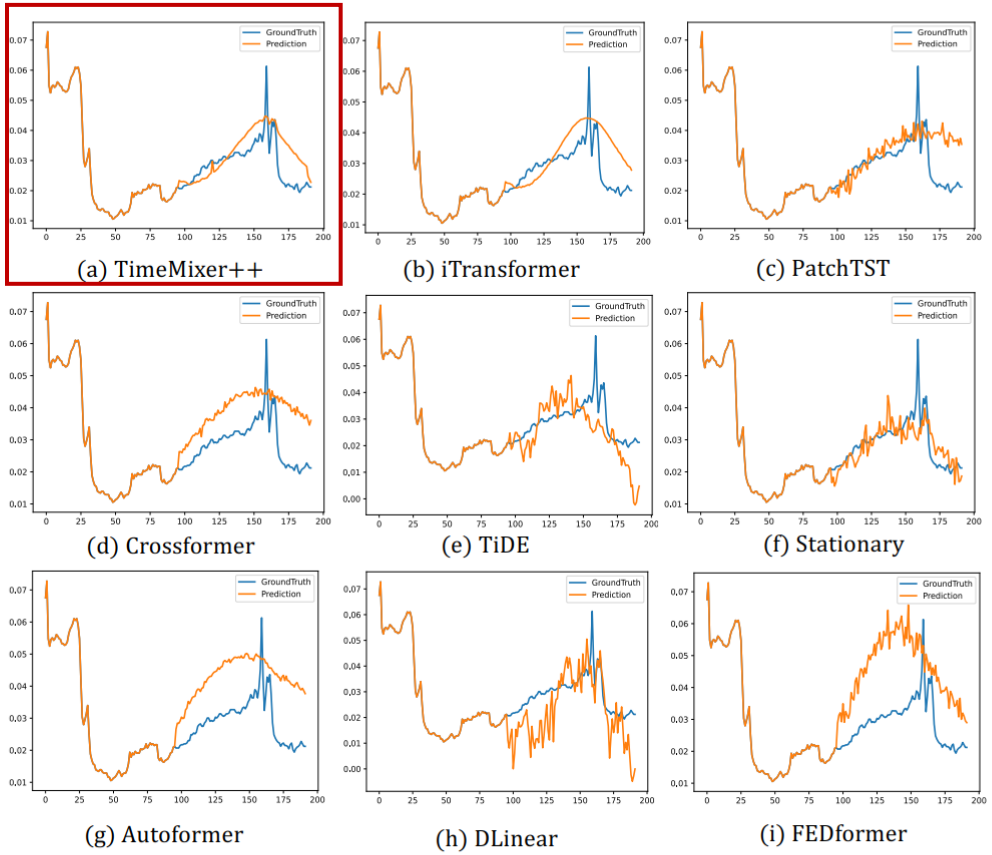
Models	**TimeMixer++ (Ours)**	TimeMixer (2024)	iTransformer (2023)	TiDE (2023b)	TimesNet (2023)	N-HiTS (2023)	N-BEATS (2019)	PatchTST (2023)	MICN (2023)	FiLM (2022a)	LightTS (2022a)	DLinear (2023)	FED. (2022b)	Stationary (2022b)	Auto. (2021)
SMAPE	**11.448**	11.723	12.684	13.950	11.829	11.927	11.851	13.152	19.638	14.863	13.525	13.639	12.840	12.780	12.909
MASE	**1.487**	1.559	1.764	1.940	1.585	1.613	1.559	1.945	5.947	2.207	2.111	2.095	1.701	1.756	1.771
OWA	**0.821**	0.840	0.929	1.020	0.851	0.861	0.855	0.998	2.279	1.125	1.051	1.051	0.918	0.930	0.939

Univariate short-term forecasting results, averaged across all M4 subsets.

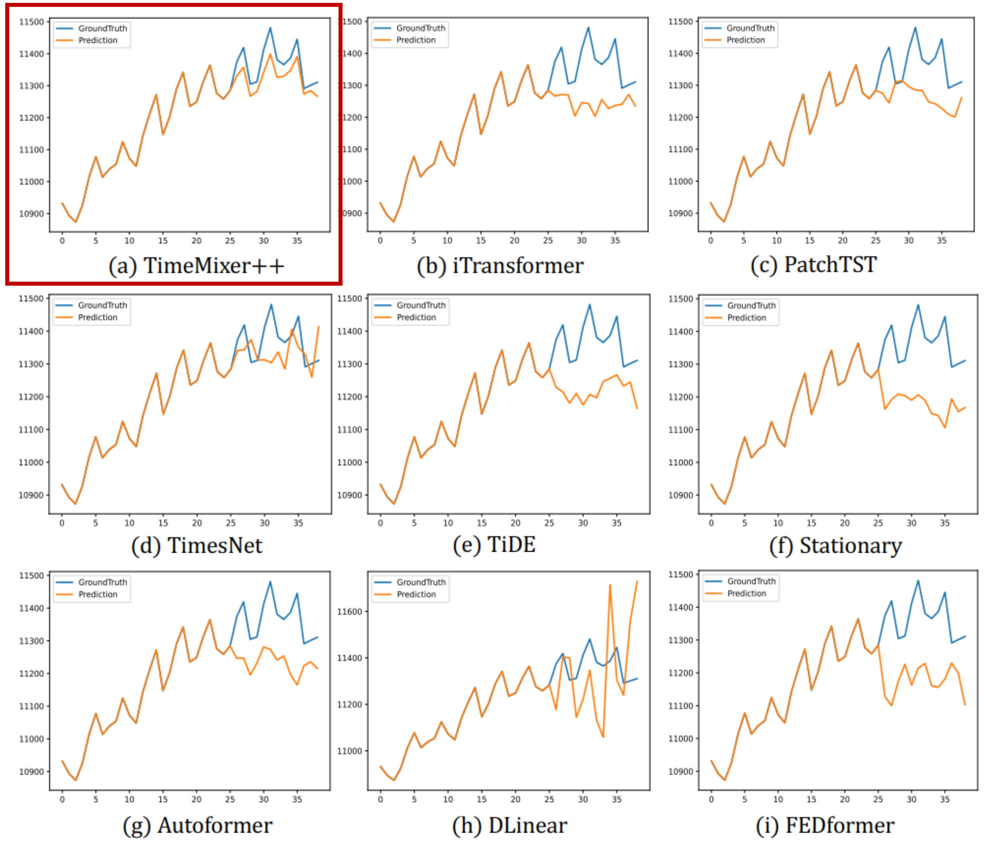
Models	**TimeMixer++ (Ours)**	TimeMixer (2024b)	iTransformer (2024)	TiDE (2023a)	SCINet (2022a)	Crossformer (2023)	PatchTST (2023)	TimesNet (2023)	MICN (2023a)	DLinear (2023)	FEDformer (2022b)	Stationary (2022c)	Autoformer (2021)
MAE	**15.91**	17.41	19.87	21.86	19.12	19.03	23.01	20.54	19.34	23.31	23.50	21.32	22.62
MAPE	**10.08**	10.59	12.55	13.80	12.24	12.22	14.95	12.69	12.38	14.68	15.01	14.09	14.89
RMSE	**27.06**	28.01	31.29	34.42	30.12	30.17	36.05	33.25	30.40	37.32	36.78	36.20	34.49

Main Results

- **Forecasting**



Input-96-predict-96 in the Weather dataset



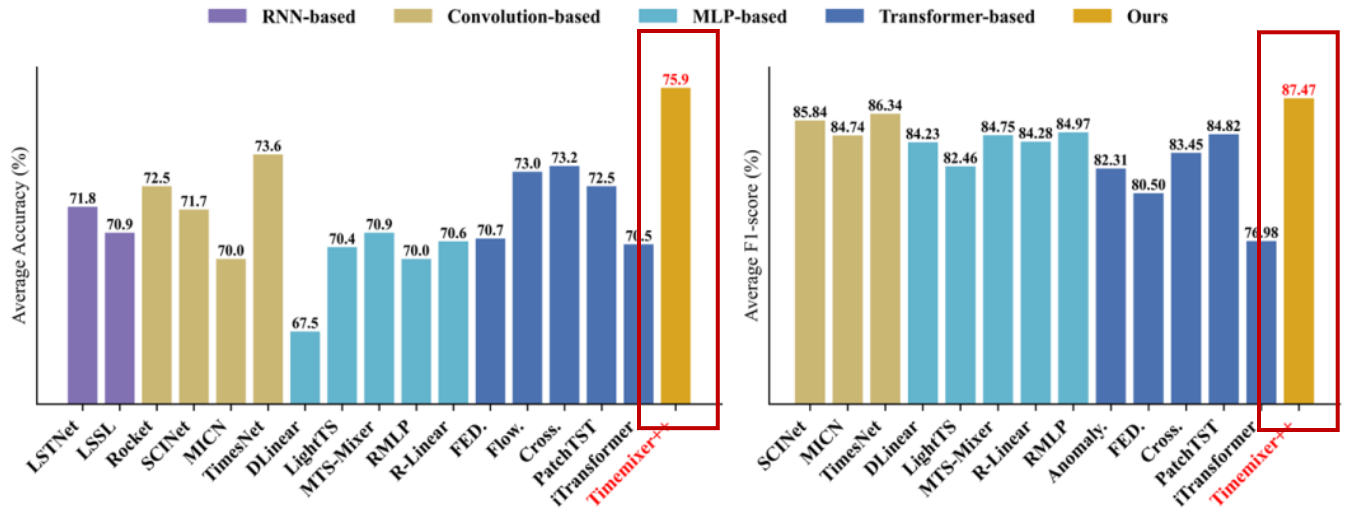
Input-36-predict-18 in the M4 dataset

Main Results

- **Imputation, Classification and Anomaly Detection**

Models	TimeMixer++ (Ours)		TimeMixer (2024b)		iTransformer (2024)		PatchTST (2023)		Crossformer (2023)		FEDformer (2022b)		TIDE (2023a)		DLinear (2023)		TimesNet (2023)		MICN (2023a)		Autoformer (2021)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETT(Avg)	**0.055**	**0.154**	0.097	0.220	0.096	0.205	0.120	0.225	0.150	0.258	0.124	0.230	0.314	0.366	0.115	0.229	0.079	0.182	0.119	0.234	0.104	0.215
ECL	**0.109**	**0.197**	0.142	0.261	0.140	0.223	0.129	0.198	0.125	0.204	0.181	0.314	0.182	0.202	0.080	0.200	0.135	0.255	0.138	0.246	0.141	0.234
Weather	**0.049**	**0.078**	0.091	0.114	0.095	0.102	0.082	0.149	0.150	0.111	0.064	0.139	0.063	0.131	0.071	0.107	0.061	0.098	0.075	0.126	0.066	0.107

Results of **classification and anomaly detection**. The results are averaged from several datasets. Higher accuracy and F1 score indicate better performance.

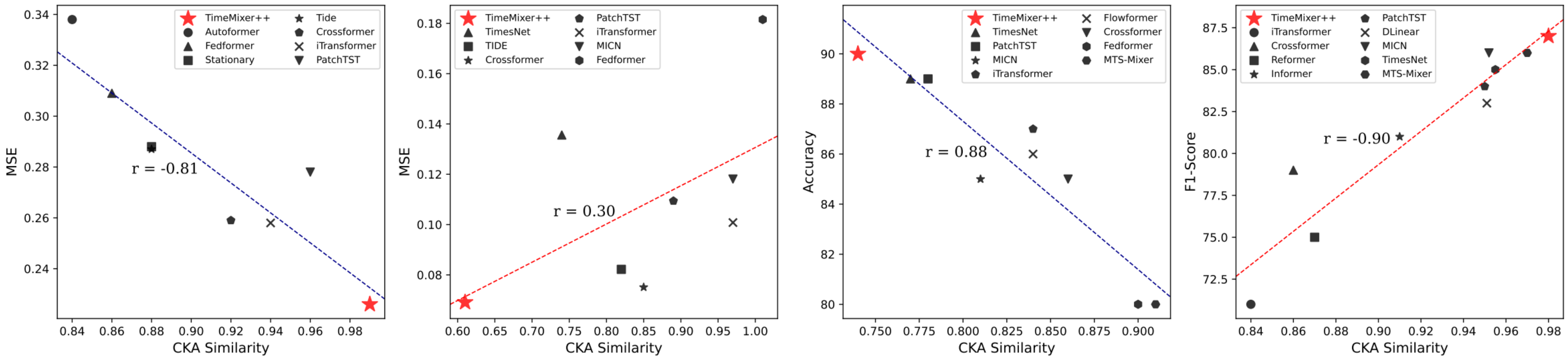


(a) Classification Results (b) Anomaly Detection Results

Model Analysis

- **Ablation Study & Representation Analysis**

	ETTh1	ETTh2	ETTm1	ETTm2	ECL	Traffic	Weather	Solar	Average	Promotion
TimeMixer++	**0.419**	**0.339**	**0.369**	**0.269**	**0.165**	**0.416**	**0.226**	**0.203**	**0.300**	-
w/o channel mixing	0.424	0.346	0.374	0.271	0.197	0.442	0.233	0.245	0.317	5.36%
w/o time image decomposition	0.441	0.358	0.409	0.291	0.198	0.445	0.251	0.241	0.329	8.81%
w/o multi-scale mixing	0.447	0.361	0.391	0.284	0.172	0.427	0.239	0.234	0.320	6.25%
w/o multi-resolution mixing	0.431	0.350	0.374	0.280	0.181	0.432	0.241	0.233	0.316	5.10%



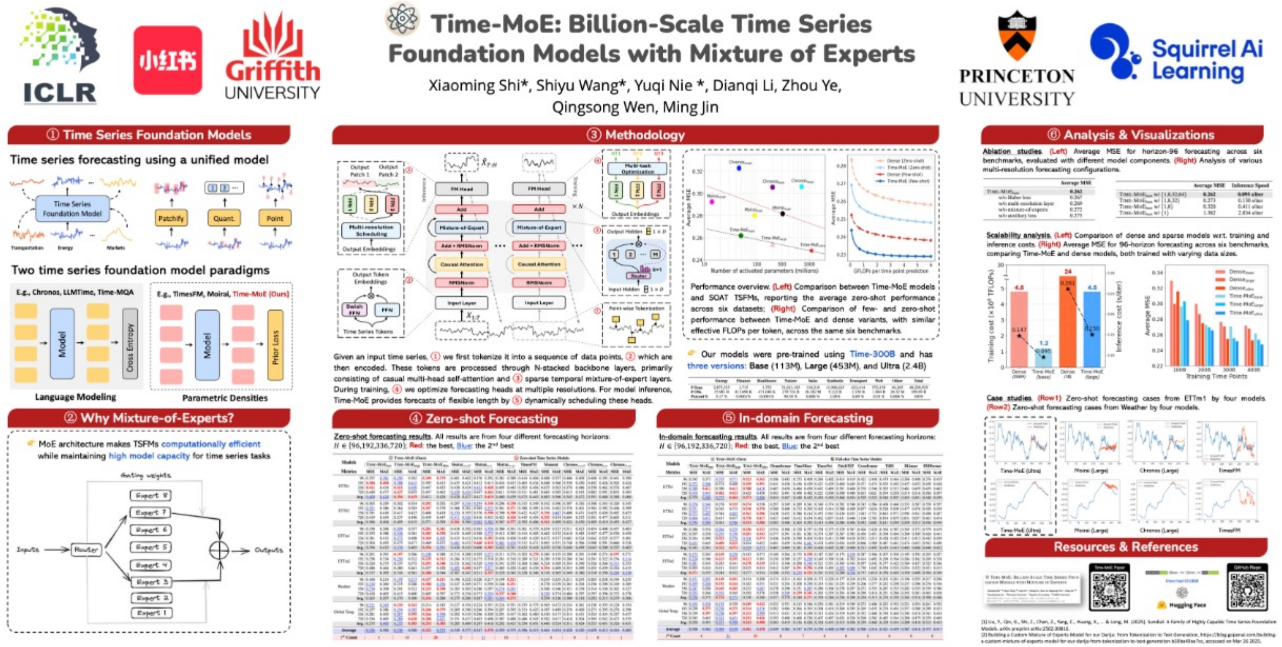
(a) Forecasting (Weather input-96-predict-336) **(b) Imputation (Electricity)** **(c) Classification (PEMS-SF)** **(d) Anomaly Detection (SMD)**

Representation analysis in four tasks. For each model, the centered kernel alignment (CKA) similarity is computed between representations from the first and the last layers. Lower CKA similarity indicates more diverse representations across layers.

Read More

- **Poster Session: TimeMixer++**
  - This afternoon (3 - 5:30 p.m. 25 April) Hall 3 + Hall 2B #345

- **Poster Session: Time-MoE**
  - Tomorrow (3 - 5:30 p.m. 26 April) Hall 3 + Hall 2B #318



ICLR International Conference On Learning Representations | Griffith UNIVERSITY Queensland, Australia | 香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY | MIT | ZHEJIANG UNIVERSITY 1897

TimeMixer++: A General Time Series Pattern Machine for Universal Predictive Analysis

Thank You

April 25, 2025, Singapore

ICLR 2025