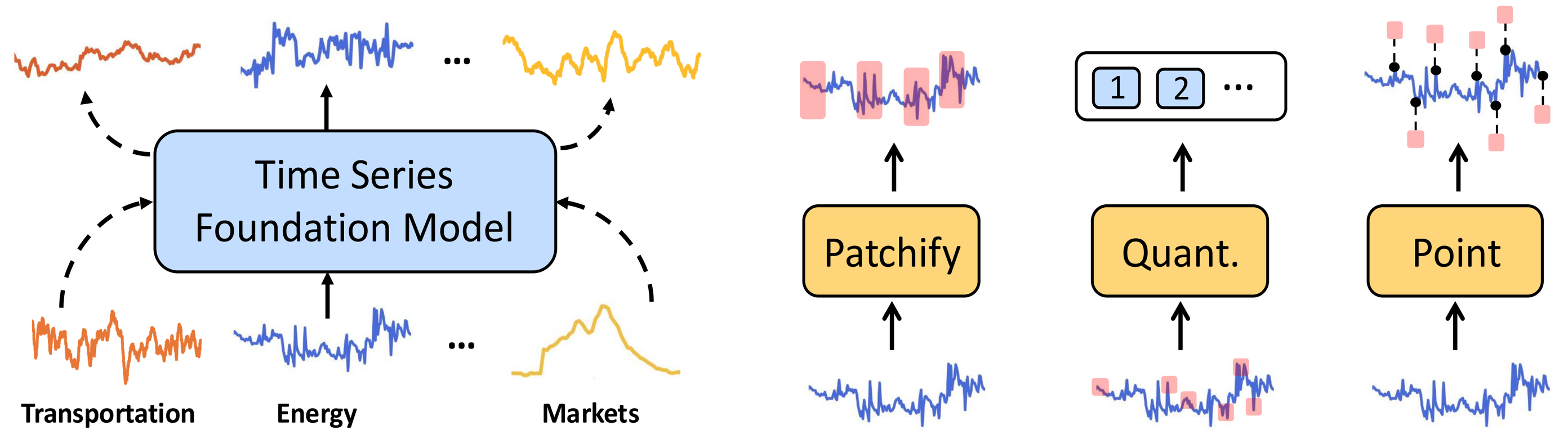
Time-MoE: Billion-Scale Time Series Foundation Models with Mixture of Experts

Xiaoming Shi*, Shiyu Wang*, Yuqi Nie *, Dianqi Li, Zhou Ye,
Qingsong Wen, Ming Jin

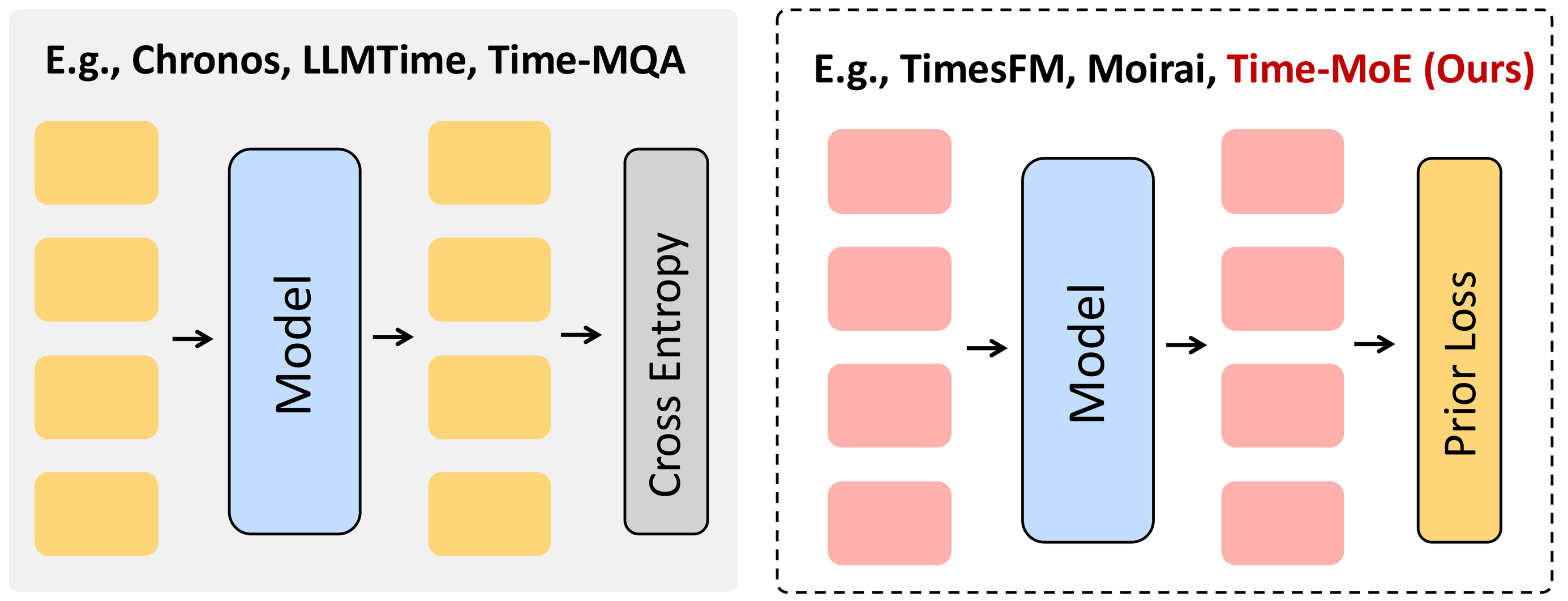
ICLR · 小红书 · Griffith University · Princeton University · Squirrel Ai Learning

① Time Series Foundation Models

Time series forecasting using a unified model



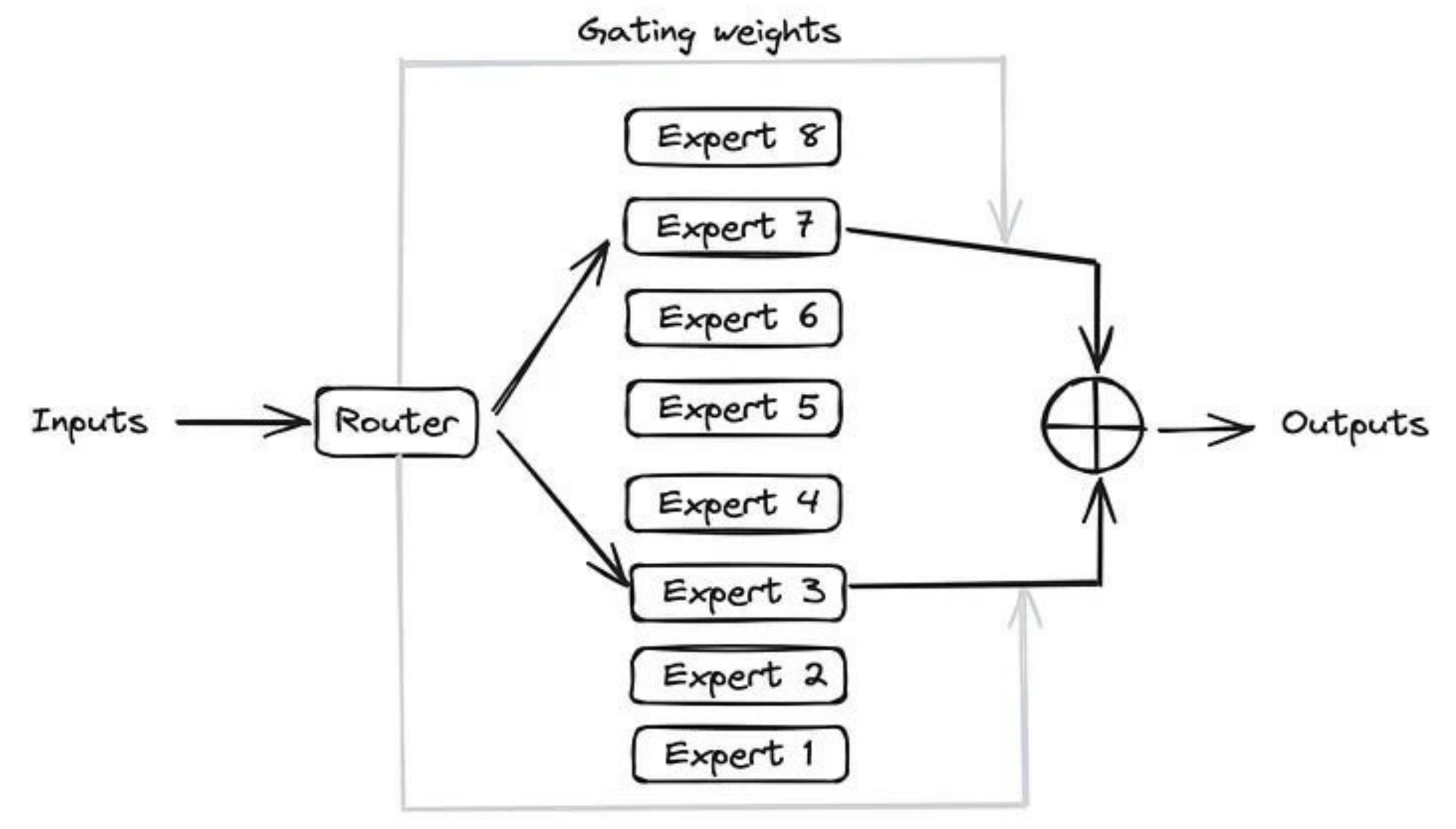
Two time series foundation model paradigms



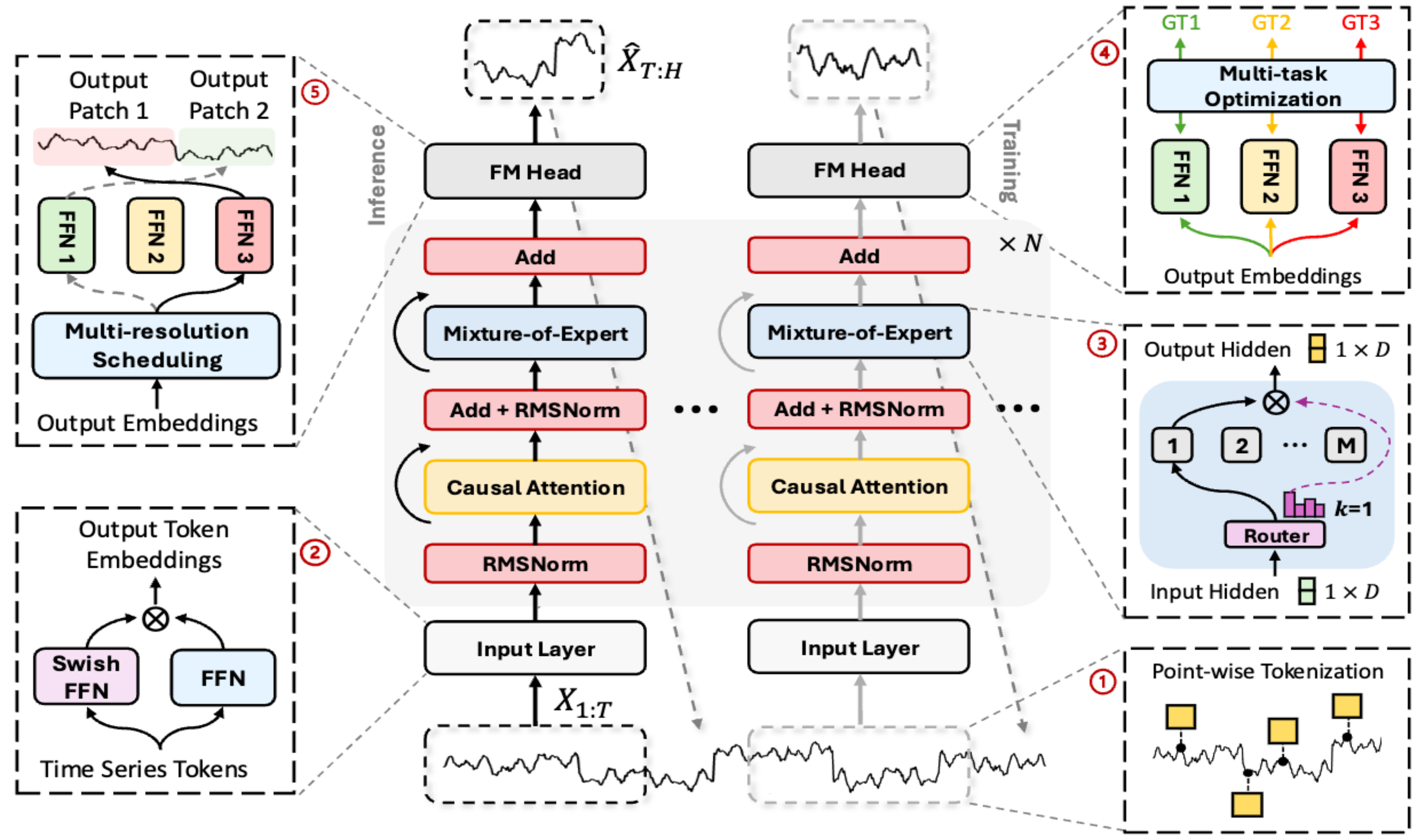
Language Modeling | Parametric Densities

② Why Mixture-of-Experts?

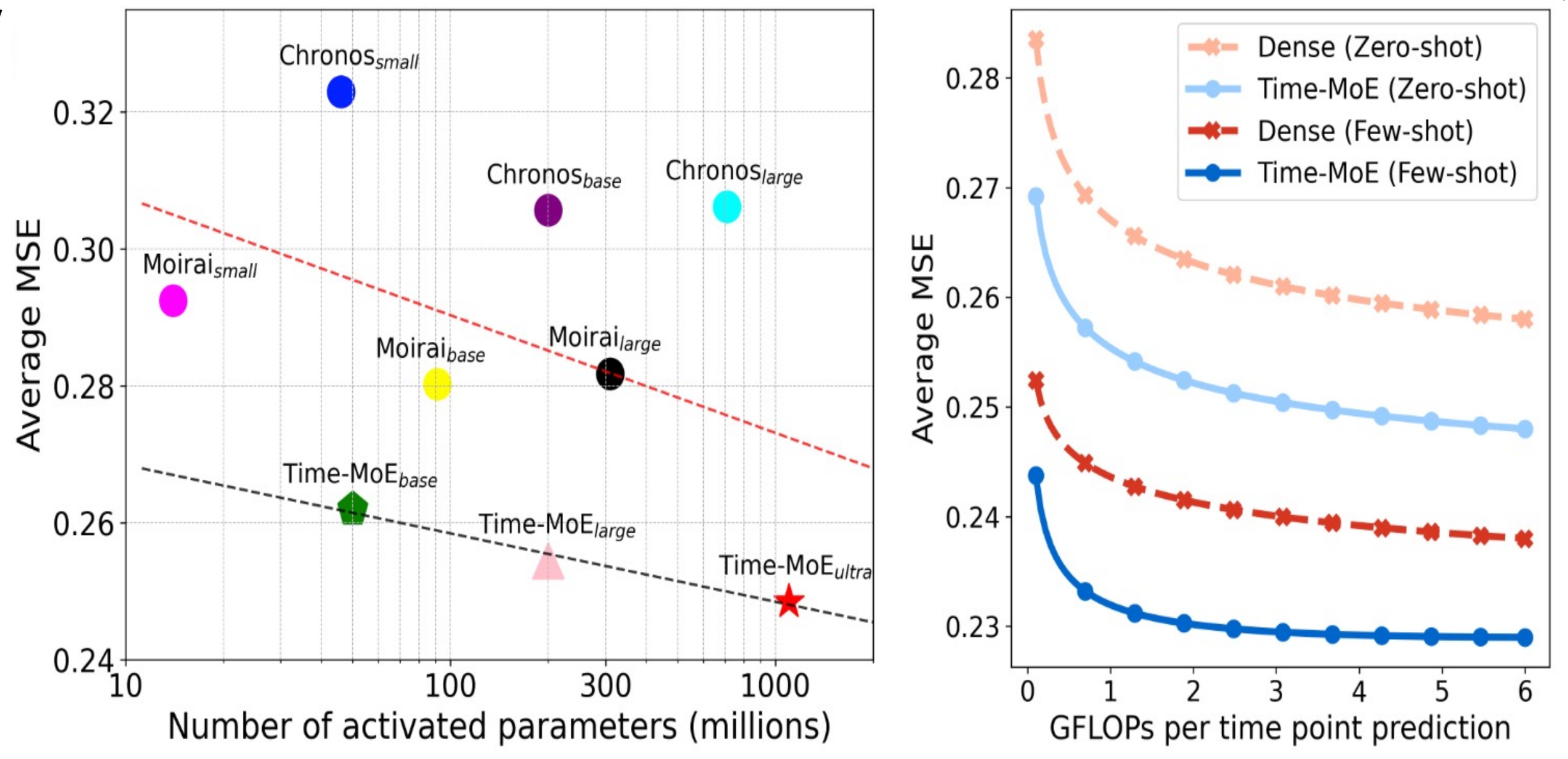
👉 MoE architecture makes TSFMs **computationally efficient** while maintaining **high model capacity** for time series tasks



③ Methodology



Given an input time series, ① we first tokenize it into a sequence of data points, ② which are then encoded. These tokens are processed through N-stacked backbone layers, primarily consisting of casual multi-head self-attention and ③ sparse temporal mixture-of-expert layers. During training, ④ we optimize forecasting heads at multiple resolutions. For model inference, Time-MoE provides forecasts of flexible length by ⑤ dynamically scheduling these heads.



Performance overview. **(Left)** Comparison between Time-MoE models and SOAT TSFMs, reporting the average zero-shot performance across six datasets; **(Right)** Comparison of few- and zero-shot performance between Time-MoE and dense variants, with similar effective FLOPs per token, across the same six benchmarks.

👉 Our models were pre-trained using **Time-300B** and has **three versions:** **Base** (113M), **Large** (453M), and **Ultra** (2.4B)

	Energy	Finance	Healthcare	Nature	Sales	Synthetic	Transport	Web	Other	Total
# Seqs.	2,875,335	1,715	1,752	31,621,183	110,210	11,968,625	622,414	972,158	40,265	48,220,929
# Obs.	15.981 B	413.696 K	471.040 K	279.724 B	26.382 M	9.222 B	2.130 B	1.804 B	20.32 M	309.09 B
Percent%	5.17 %	0.0001%	0.0001%	90.50 %	0.008 %	2.98%	0.69 %	0.58 %	0.006 %	100%

④ Zero-shot Forecasting

Zero-shot forecasting results. All results are from four different forecasting horizons: $H \in \{96,192,336,720\}$; Red: the best, Blue: the 2nd best

Models		Time-MoE (Ours): Time-MoE$_{base}$		Time-MoE$_{large}$		Time-MoE$_{ultra}$		Zero-shot Time Series Models: Moirai$_{small}$		Moirai$_{base}$		Moirai$_{large}$		TimesFM		Moment		Chronos$_{small}$		Chronos$_{base}$		Chronos$_{large}$	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.357	0.381	0.350	0.382	0.349	0.379	0.401	0.402	0.376	0.392	0.381	0.388	0.414	0.404	0.688	0.557	0.466	0.409	0.440	0.393	0.441	0.390
	192	0.384	0.404	0.388	0.412	0.395	0.413	0.435	0.421	0.412	0.413	0.434	0.415	0.465	0.434	0.688	0.560	0.530	0.450	0.492	0.426	0.502	0.424
	336	0.411	0.434	0.411	0.430	0.447	0.453	0.438	0.434	0.433	0.428	0.495	0.445	0.503	0.456	0.675	0.563	0.570	0.486	0.550	0.462	0.576	0.467
	720	0.449	0.477	0.427	0.455	0.457	0.462	0.439	0.454	0.447	0.444	0.611	0.510	0.511	0.481	0.683	0.585	0.615	0.543	0.882	0.591	0.835	0.583
	Avg.	0.400	0.424	0.394	0.419	0.412	0.426	0.428	0.427	0.417	0.419	0.480	0.439	0.473	0.443	0.683	0.566	0.545	0.472	0.591	0.468	0.588	0.466
ETTh2	96	0.305	0.359	0.302	0.354	0.292	0.352	0.297	0.336	0.294	0.330	0.296	0.330	0.315	0.349	0.342	0.396	0.307	0.356	0.308	0.343	0.320	0.345
	192	0.351	0.386	0.364	0.385	0.347	0.379	0.368	0.381	0.365	0.375	0.361	0.371	0.388	0.395	0.354	0.402	0.376	0.401	0.384	0.392	0.406	0.399
	336	0.391	0.418	0.417	0.425	0.406	0.419	0.370	0.393	0.376	0.390	0.390	0.390	0.422	0.427	0.356	0.407	0.408	0.431	0.429	0.430	0.492	0.453
	720	0.419	0.454	0.537	0.496	0.439	0.447	0.411	0.426	0.416	0.433	0.423	0.418	0.443	0.454	0.395	0.434	0.604	0.533	0.501	0.477	0.603	0.511
	Avg.	0.366	0.404	0.405	0.415	0.371	0.399	0.361	0.384	0.362	0.382	0.367	0.377	0.392	0.406	0.361	0.409	0.424	0.430	0.405	0.410	0.455	0.427
ETTm1	96	0.338	0.368	0.309	0.357	0.281	0.341	0.418	0.392	0.363	0.356	0.380	0.361	0.361	0.370	0.654	0.527	0.511	0.423	0.454	0.408	0.457	0.403
	192	0.353	0.388	0.346	0.381	0.305	0.358	0.431	0.405	0.388	0.375	0.412	0.383	0.414	0.405	0.662	0.532	0.618	0.485	0.567	0.477	0.530	0.450
	336	0.381	0.413	0.373	0.408	0.369	0.395	0.433	0.412	0.416	0.392	0.436	0.400	0.445	0.429	0.672	0.537	0.683	0.524	0.662	0.525	0.577	0.481
	720	0.504	0.493	0.475	0.477	0.469	0.472	0.462	0.432	0.460	0.418	0.462	0.420	0.512	0.471	0.692	0.551	0.748	0.566	0.900	0.591	0.660	0.526
	Avg.	0.394	0.415	0.376	0.405	0.356	0.391	0.436	0.410	0.406	0.385	0.422	0.391	0.433	0.418	0.670	0.536	0.640	0.499	0.645	0.500	0.555	0.465
ETTm2	96	0.201	0.291	0.197	0.286	0.198	0.288	0.214	0.288	0.205	0.273	0.211	0.274	0.202	0.270	0.260	0.335	0.209	0.291	0.199	0.274	0.197	0.271
	192	0.258	0.334	0.250	0.322	0.235	0.312	0.284	0.332	0.275	0.316	0.281	0.318	0.289	0.321	0.289	0.350	0.280	0.341	0.261	0.322	0.254	0.314
	336	0.324	0.373	0.337	0.375	0.293	0.348	0.331	0.362	0.329	0.350	0.341	0.355	0.360	0.366	0.324	0.369	0.354	0.390	0.326	0.366	0.313	0.353
	720	0.488	0.464	0.480	0.461	0.427	0.428	0.402	0.408	0.437	0.411	0.485	0.428	0.462	0.430	0.394	0.409	0.553	0.499	0.455	0.439	0.416	0.415
	Avg.	0.317	0.365	0.316	0.361	0.288	0.344	0.307	0.347	0.311	0.337	0.329	0.343	0.328	0.346	0.316	0.365	0.349	0.380	0.310	0.350	0.295	0.338
Weather	96	0.160	0.214	0.159	0.213	0.157	0.211	0.198	0.222	0.220	0.217	0.199	0.211	-	-	0.243	0.255	0.211	0.243	0.203	0.238	0.194	0.235
	192	0.210	0.260	0.215	0.266	0.208	0.256	0.247	0.265	0.271	0.259	0.246	0.251	-	-	0.278	0.329	0.263	0.294	0.256	0.290	0.249	0.285
	336	0.274	0.309	0.291	0.322	0.255	0.290	0.283	0.303	0.286	0.297	0.274	0.291	-	-	0.306	0.346	0.321	0.339	0.314	0.336	0.302	0.327
	720	0.418	0.405	0.415	0.400	0.405	0.397	0.373	0.354	0.373	0.354	0.337	0.340	-	-	0.350	0.374	0.404	0.397	0.397	0.396	0.372	0.378
	Avg.	0.265	0.297	0.270	0.300	0.256	0.288	0.275	0.286	0.287	0.281	0.264	0.273	-	-	0.294	0.326	0.300	0.318	0.292	0.315	0.279	0.306
Global Temp	96	0.211	0.343	0.210	0.342	0.214	0.345	0.227	0.354	0.224	0.351	0.224	0.351	0.255	0.375	0.363	0.472	0.234	0.361	0.230	0.355	0.228	0.354
	192	0.257	0.386	0.254	0.385	0.246	0.379	0.269	0.396	0.266	0.394	0.267	0.395	0.313	0.423	0.387	0.489	0.276	0.400	0.273	0.395	0.276	0.398
	336	0.281	0.405	0.267	0.395	0.266	0.398	0.292	0.419	0.296	0.420	0.291	0.417	0.362	0.460	0.430	0.517	0.314	0.431	0.324	0.434	0.327	0.437
	720	0.354	0.465	0.289	0.420	0.288	0.421	0.351	0.437	0.403	0.498	0.387	0.488	0.486	0.545	0.582	0.617	0.418	0.504	0.505	0.542	0.472	0.535
	Avg.	0.275	0.400	0.255	0.385	0.253	0.385	0.285	0.409	0.297	0.416	0.292	0.413	0.354	0.451	0.440	0.524	0.311	0.424	0.333	0.431	0.326	0.431
Average		0.336	0.384	0.336	0.380	0.322	0.372	0.349	0.377	0.347	0.370	0.359	0.373	0.396	0.413	0.461	0.454	0.428	0.420	0.429	0.412	0.416	0.405
1st Count		3		10		28		2		11		10		1		4		0		0		1	

⑤ In-domain Forecasting

In-domain forecasting results. All results are from four different forecasting horizons: $H \in \{96,192,336,720\}$; Red: the best, Blue: the 2nd best

Models		Time-MoE (Ours): Time-MoE$_{base}$		Time-MoE$_{large}$		Time-MoE$_{ultra}$		Full-shot Time Series Models: iTransformer		TimeMixer		TimesNet		PatchTST		Crossformer		TiDE		DLinear		FEDformer	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.345	0.373	0.335	0.371	0.323	0.365	0.386	0.405	0.375	0.400	0.384	0.402	0.414	0.419	0.423	0.448	0.479	0.464	0.386	0.400	0.376	0.419
	192	0.372	0.396	0.374	0.400	0.359	0.391	0.441	0.436	0.436	0.429	0.421	0.429	0.460	0.445	0.471	0.474	0.525	0.492	0.437	0.432	0.420	0.448
	336	0.389	0.412	0.390	0.412	0.388	0.418	0.487	0.458	0.484	0.458	0.491	0.469	0.501	0.466	0.570	0.546	0.565	0.515	0.481	0.459	0.459	0.465
	720	0.410	0.443	0.402	0.433	0.425	0.450	0.503	0.491	0.498	0.482	0.521	0.500	0.500	0.488	0.653	0.621	0.594	0.558	0.519	0.516	0.506	0.507
	Avg.	0.379	0.406	0.375	0.404	0.373	0.406	0.454	0.447	0.448	0.442	0.454	0.450	0.468	0.454	0.529	0.522	0.540	0.507	0.455	0.451	0.440	0.459
ETTh2	96	0.276	0.340	0.278	0.335	0.274	0.338	0.297	0.349	0.289	0.341	0.340	0.374	0.302	0.348	0.745	0.584	0.400	0.440	0.333	0.387	0.358	0.397
	192	0.331	0.371	0.345	0.373	0.330	0.370	0.380	0.400	0.372	0.392	0.402	0.414	0.388	0.400	0.877	0.656	0.528	0.509	0.477	0.476	0.429	0.439
	336	0.373	0.402	0.384	0.402	0.362	0.396	0.428	0.432	0.386	0.414	0.452	0.541	0.426	0.433	1.043	0.731	0.643	0.571	0.594	0.541	0.496	0.487
	720	0.404	0.431	0.437	0.437	0.370	0.417	0.427	0.445	0.412	0.434	0.462	0.657	0.431	0.446	1.104	0.763	0.874	0.679	0.831	0.657	0.463	0.474
	Avg.	0.346	0.386	0.361	0.386	0.334	0.380	0.383	0.406	0.364	0.395	0.414	0.496	0.386	0.406	0.942	0.683	0.611	0.549	0.558	0.515	0.436	0.449
ETTm1	96	0.286	0.334	0.264	0.325	0.256	0.323	0.334	0.368	0.320	0.357	0.338	0.375	0.329	0.367	0.404	0.426	0.364	0.387	0.345	0.372	0.379	0.419
	192	0.307	0.358	0.295	0.350	0.281	0.343	0.377	0.391	0.361	0.381	0.374	0.387	0.367	0.385	0.450	0.451	0.398	0.404	0.380	0.389	0.426	0.441
	336	0.354	0.390	0.323	0.376	0.326	0.374	0.426	0.420	0.390	0.404	0.410	0.411	0.399	0.410	0.532	0.515	0.428	0.425	0.413	0.413	0.445	0.459
	720	0.433	0.445	0.409	0.435	0.454	0.452	0.491	0.459	0.454	0.441	0.478	0.450	0.454	0.439	0.666	0.589	0.487	0.461	0.474	0.453	0.543	0.490
	Avg.	0.345	0.381	0.322	0.371	0.329	0.373	0.407	0.409	0.381	0.395	0.400	0.405	0.387	0.400	0.513	0.495	0.419	0.419	0.403	0.406	0.448	0.452
ETTm2	96	0.172	0.265	0.169	0.259	0.183	0.273	0.180	0.264	0.175	0.258	0.187	0.267	0.175	0.259	0.287	0.366	0.207	0.305	0.193	0.292	0.203	0.287
	192	0.228	0.306	0.223	0.295	0.223	0.301	0.250	0.309	0.237	0.299	0.249	0.309	0.241	0.302	0.414	0.492	0.290	0.364	0.284	0.362	0.269	0.328
	336	0.281	0.345	0.293	0.341	0.278	0.339	0.311	0.348	0.298	0.340	0.321	0.351	0.305	0.343	0.597	0.542	0.377	0.422	0.369	0.427	0.325	0.366
	720	0.403	0.424	0.451	0.433	0.425	0.424	0.412	0.407	0.391	0.396	0.408	0.403	0.402	0.400	1.730	1.042	0.558	0.524	0.554	0.522	0.421	0.415
	Avg.	0.271	0.335	0.284	0.332	0.277	0.334	0.288	0.332	0.275	0.323	0.291	0.332	0.280	0.326	0.757	0.610	0.358	0.403	0.350	0.400	0.304	0.349
Weather	96	0.151	0.203	0.149	0.201	0.154	0.208	0.174	0.214	0.163	0.209	0.172	0.220	0.177	0.218	0.158	0.230	0.202	0.261	0.196	0.255	0.217	0.296
	192	0.195	0.246	0.192	0.244	0.202	0.251	0.221	0.254	0.208	0.250	0.219	0.261	0.225	0.259	0.206	0.277	0.242	0.298	0.237	0.296	0.276	0.336
	336	0.247	0.288	0.245	0.285	0.252	0.287	0.278	0.296	0.251	0.287	0.280	0.306	0.278	0.297	0.272	0.335	0.287	0.335	0.283	0.335	0.339	0.380
	720	0.352	0.366	0.352	0.365	0.392	0.376	0.358	0.349	0.339	0.341	0.365	0.359	0.354	0.348	0.398	0.418	0.351	0.386	0.345	0.381	0.403	0.428
	Avg.	0.236	0.275	0.234	0.273	0.250	0.280	0.257	0.278	0.240	0.271	0.259	0.286	0.258	0.280	0.258	0.315	0.270	0.320	0.265	0.316	0.308	0.360
Global Temp	96	0.192	0.328	0.192	0.329	0.189	0.322	0.223	0.351	0.215	0.346	0.250	0.381	0.219	0.349	0.272	0.406	0.223	0.352	0.221	0.354	0.261	0.392
	192	0.238	0.375	0.236	0.375	0.234	0.376	0.282	0.404	0.266	0.393	0.298	0.418	0.269	0.395	0.305	0.435	0.278	0.401	0.257	0.388	0.299	0.423
	336	0.259	0.397	0.256	0.397	0.253	0.399	0.313	0.431	0.313	0.430	0.315	0.434	0.319	0.435	0.352	0.468	0.330	0.440	0.294	0.418	0.341	0.454
	720	0.345	0.465	0.322	0.451	0.292	0.426	0.393	0.488	0.468	0.536	0.407	0.497	0.452	0.526	0.508	0.562	0.485	0.544	0.380	0.479	0.359	0.469
	Avg.	0.258	0.391	0.251	0.388	0.242	0.380	0.303	0.419	0.316	0.426	0.318	0.433	0.315	0.426	0.359	0.468	0.329	0.434	0.288	0.410	0.315	0.435
Average		0.306	0.362	0.304	0.359	0.301	0.358	0.349	0.382	0.337	0.375	0.356	0.400	0.349	0.382	0.560	0.516	0.421	0.439	0.387	0.416	0.375	0.417
1st Count		4		21		33		0		7		0		0		0		0		0		0	

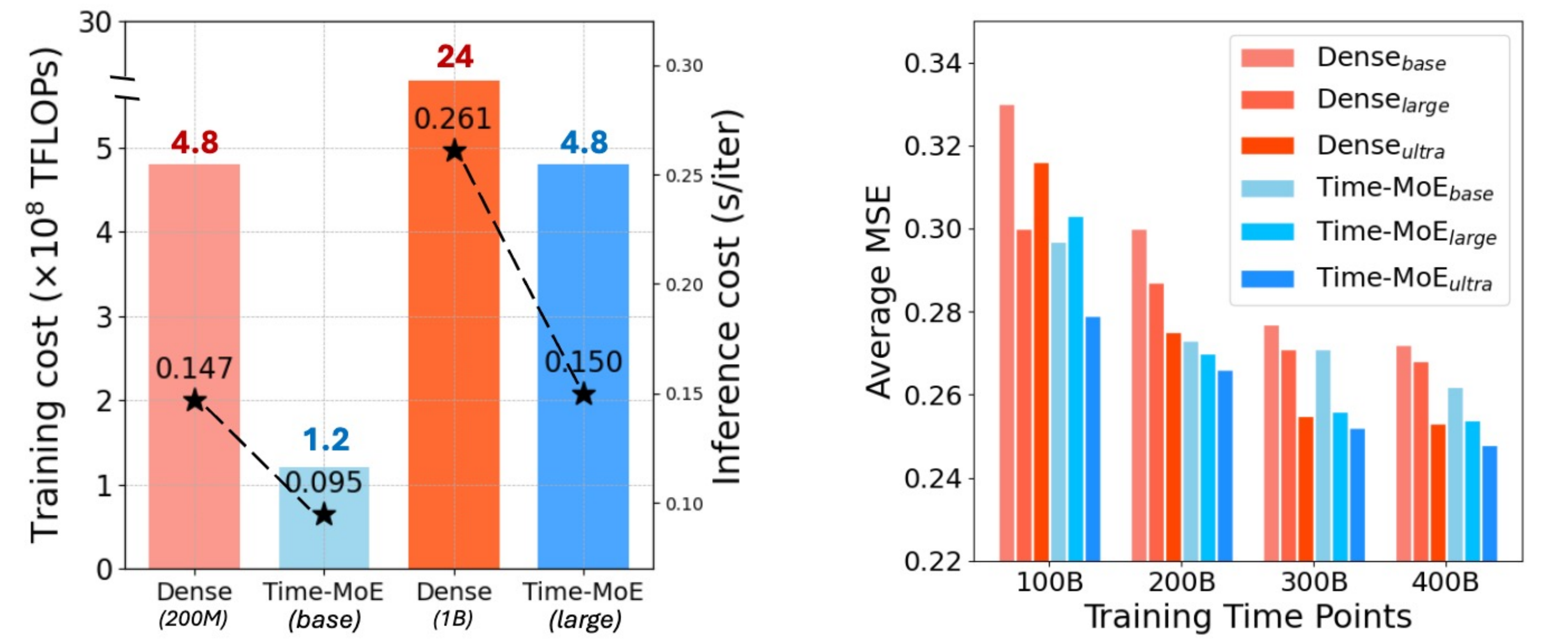
⑥ Analysis & Visualizations

Ablation studies. **(Left)** Average MSE for horizon-96 forecasting across six benchmarks, evaluated with different model components. **(Right)** Analysis of various multi-resolution forecasting configurations.

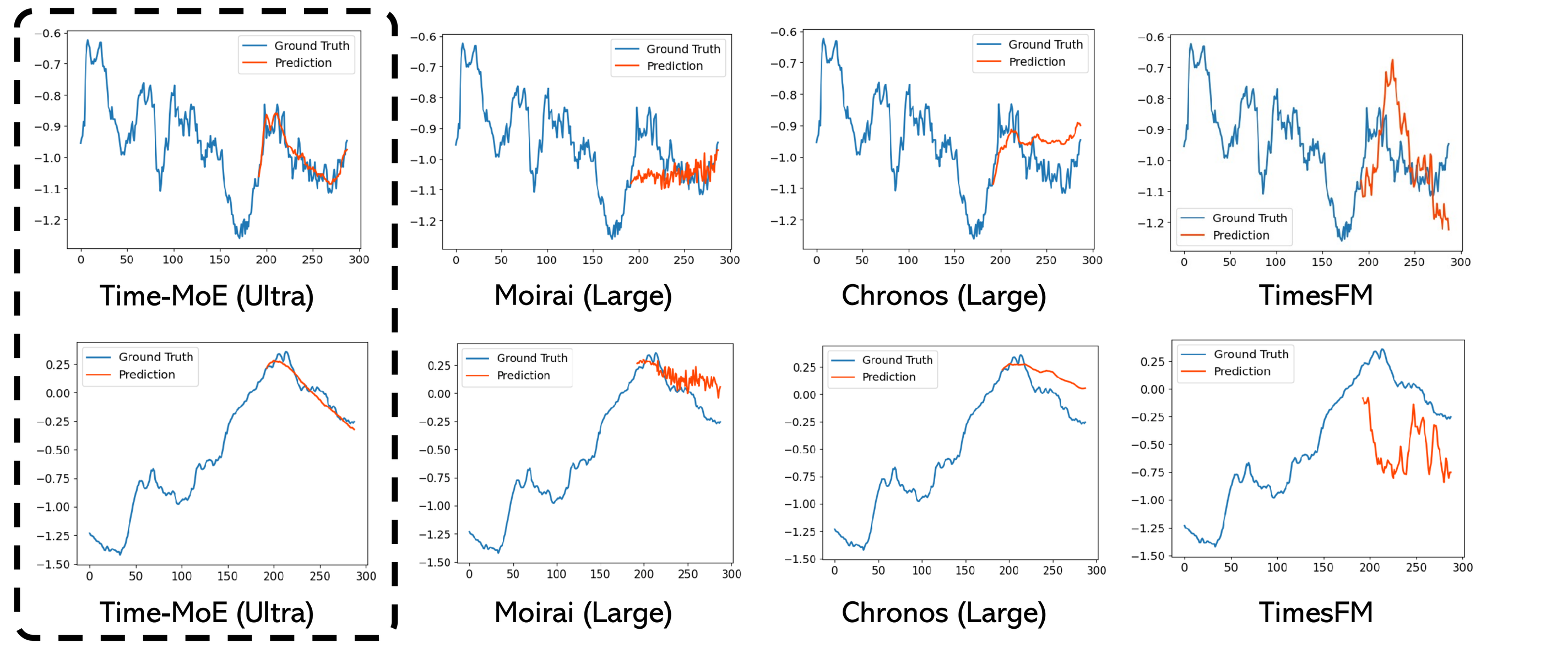
	Average MSE
Time-MoE$_{base}$	**0.262**
w/o Huber loss	0.267
w/o multi-resolution layer	0.269
w/o mixture-of-experts	0.272
w/o auxiliary loss	0.275

	Average MSE	Inference Speed
Time-MoE$_{base}$ w/ {1,8,32,64}	**0.262**	**0.095 s/iter**
Time-MoE$_{base}$ w/ {1,8,32}	0.273	0.130 s/iter
Time-MoE$_{base}$ w/ {1,8}	0.320	0.411 s/iter
Time-MoE$_{base}$ w/ {1}	1.382	2.834 s/iter

Scalability analysis. **(Left)** Comparison of dense and sparse models w.r.t. training and inference costs. **(Right)** Average MSE for 96-horizon forecasting across six benchmarks, comparing Time-MoE and dense models, both trained with varying data sizes.



Case studies. **(Row1)** Zero-shot forecasting cases from ETTm1 by four models. **(Row2)** Zero-shot forecasting cases from Weather by four models.



Resources & References

Time-MoE Paper

last commit march · Stars 513 · Forks 40 · PRs Welcome

[Paper Page] [中文解读]

Hugging Face

GitHub Repo

[1] Liu, Y., Qin, G., Shi, Z., Chen, Z., Yang, C., Huang, X., ... & Long, M. (2025). Sundial: A Family of Highly Capable Time Series Foundation Models. arXiv preprint arXiv:2502.00816.
[2] Building a Custom Mixture of Experts Model for our Darija: From Tokenization to Text Generation, https://blog.gopenai.com/building-a-custom-mixture-of-experts-model-for-our-darija-from-tokenization-to-text-generation-b30ba49ae7cc, accessed on Mar 26 2025.