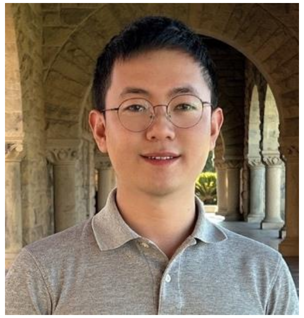
KDD2024 BARCELONA, SPAIN

Griffith UNIVERSITY Queensland, Australia

Time Series Foundation Models



Ming Jin

Assistant Professor
Griffith University

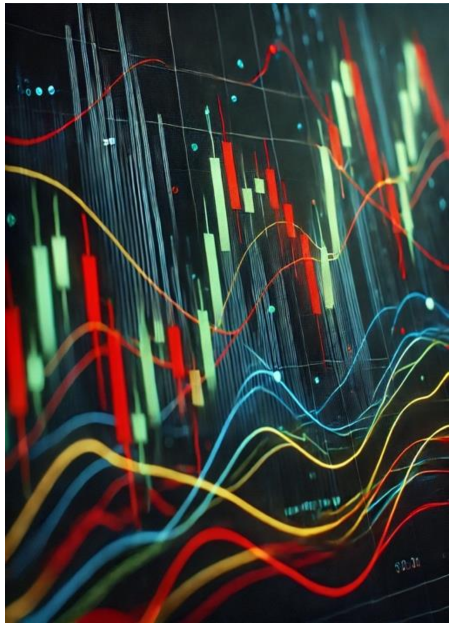
https://mingjin.dev/

(Generated by DALL·E)

Time Series Are Everywhere

- **What are time series?**

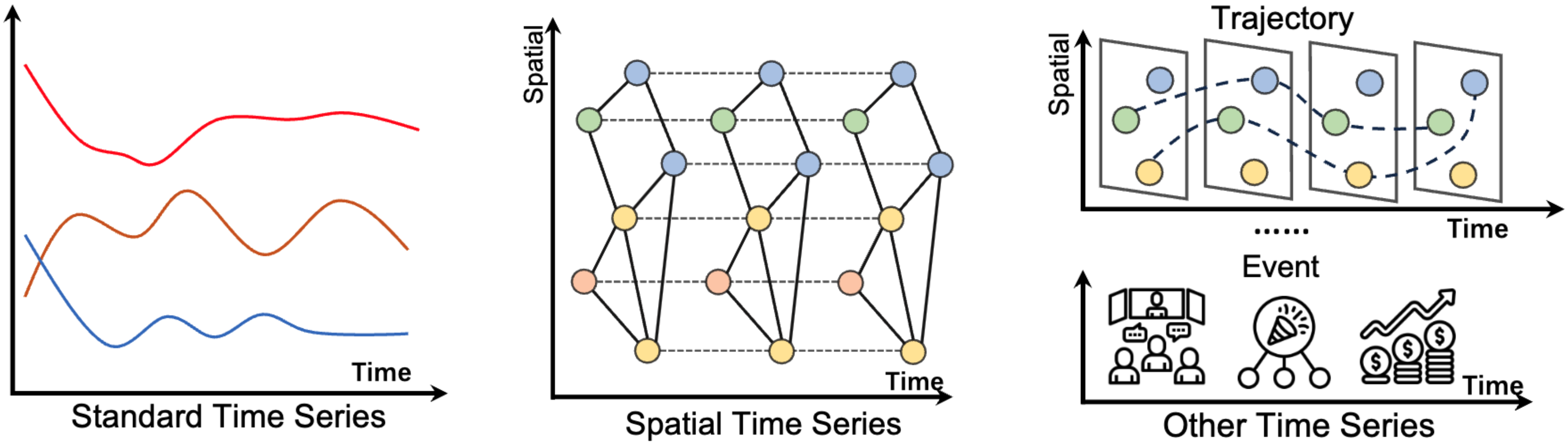
- **Time series are everywhere**



(Figures in this page are generated by DALL·E)

Time Series Data

- **What are time series?**



Liang, Y., Wen, H., Nie, Y., Jiang, Y., Jin, M., Song, D., ... & Wen, Q. (2024). Foundation models for time series analysis: A tutorial and survey. In KDD'24

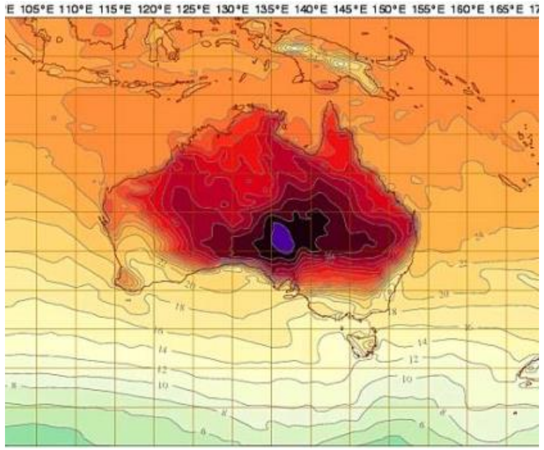
Time Series Analysis

- **Forecasting**

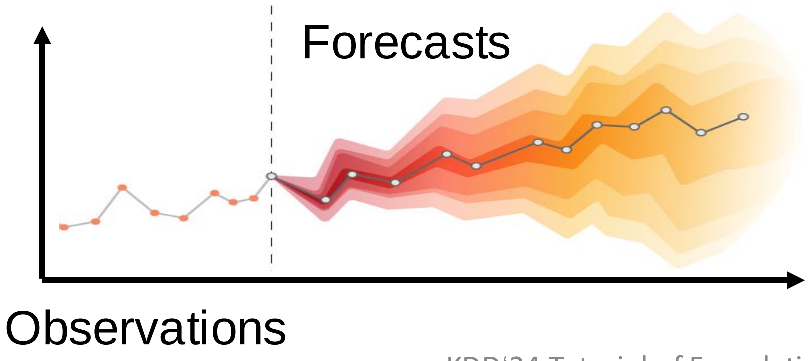
ETA

Disease propagation

Electricity demand



Global Weather



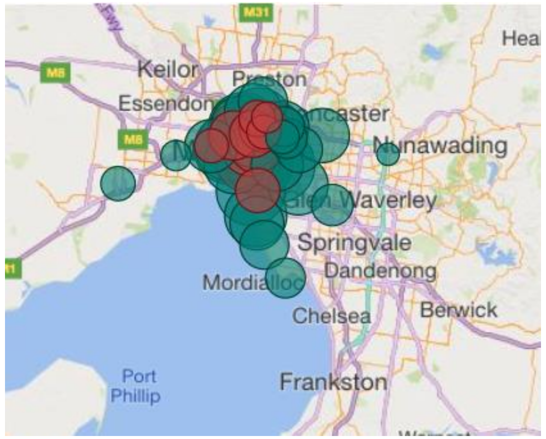
- Long-term planning
- Early warning
- Better management

(Figures in this page are generated by DALL·E)

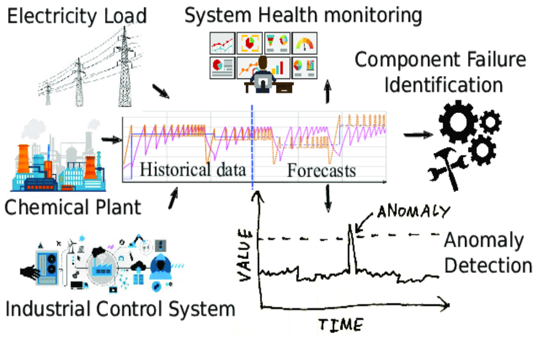
Time Series Analysis

- **Classification**

ECG diagnose

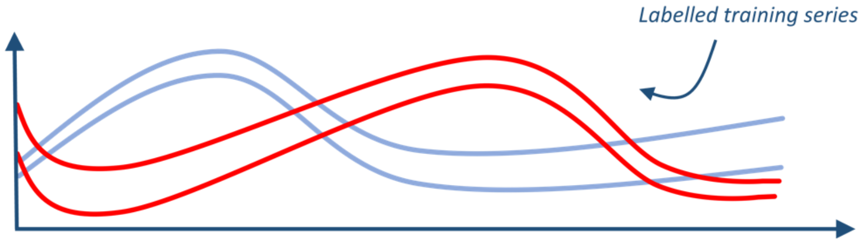


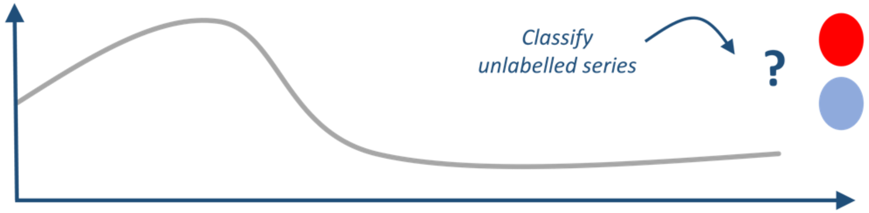
Traffic condition



Anomaly detection

AQI category



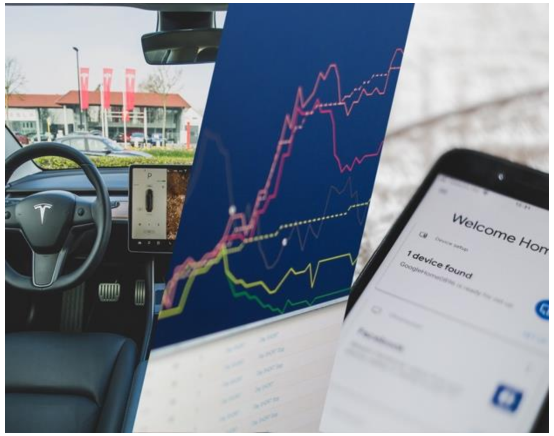


https://www.aeon-toolkit.org/en/stable/examples/classification/classification.html

Time Series Analysis

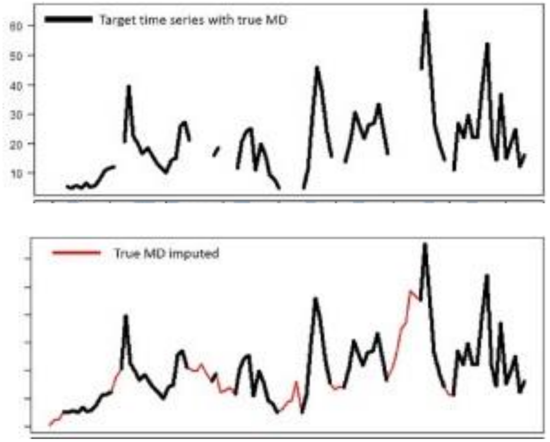
- **Generation**

Simulation

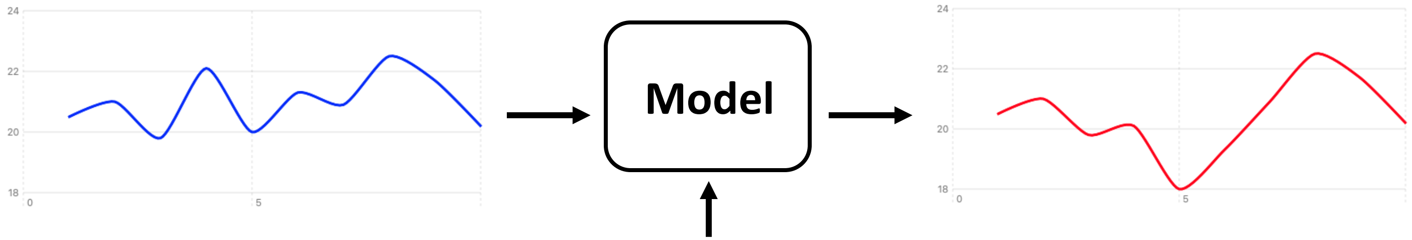


Anonymization

Data augmentation



Imputation

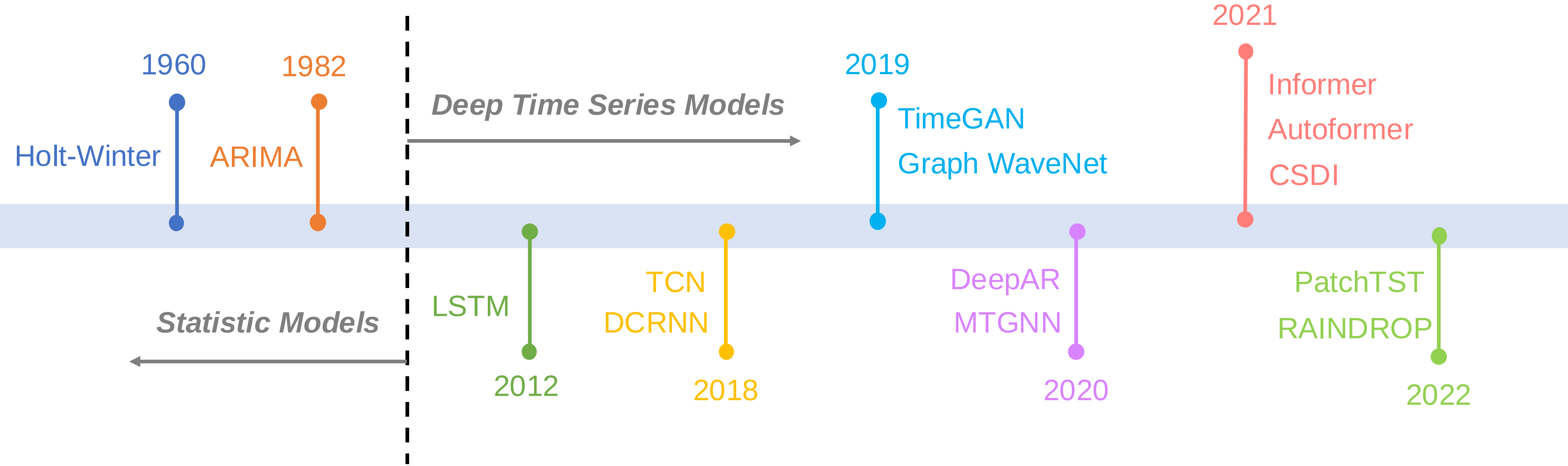


"A cold front moved through the area on Day 4, lasting until Day 6"

- Better planning and management
- Privacy preserving
- More data and applications

Timeline

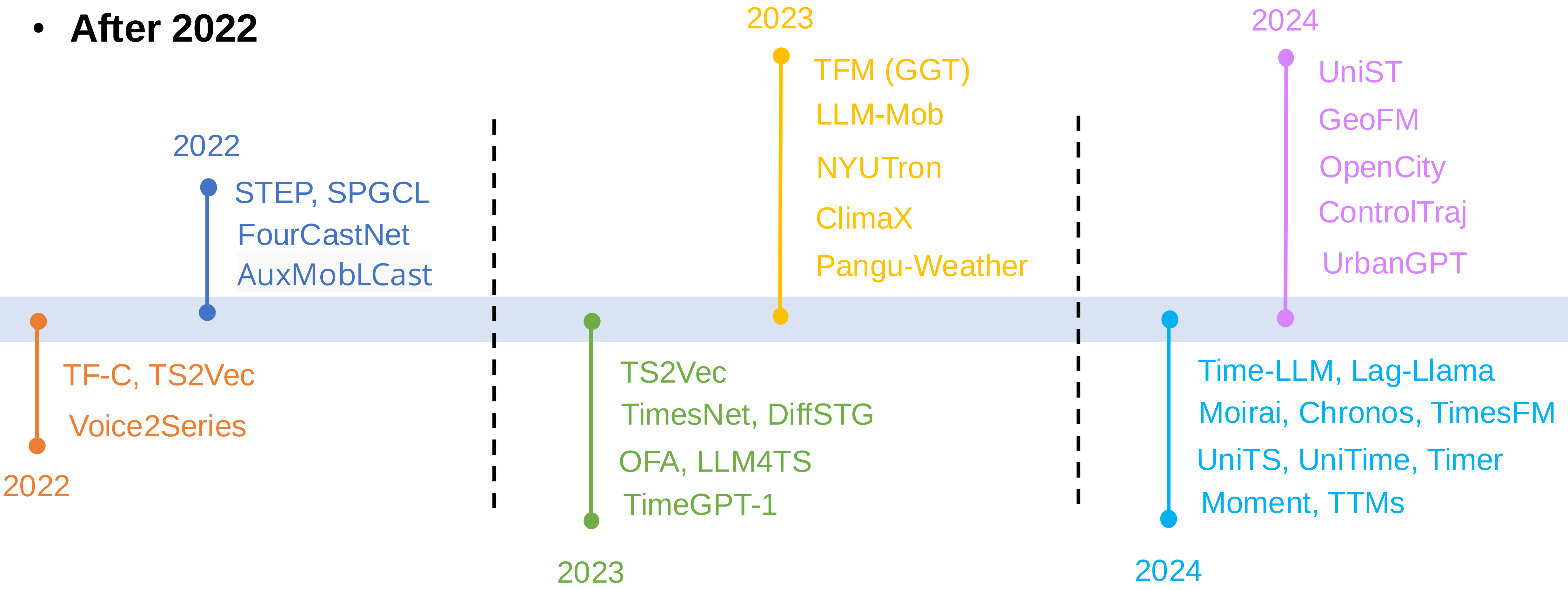
- **Before 2022**



https://ise.thss.tsinghua.edu.cn/~mlong/doc/foundation-models-for-time-series-analysis-gaitc23.pdf

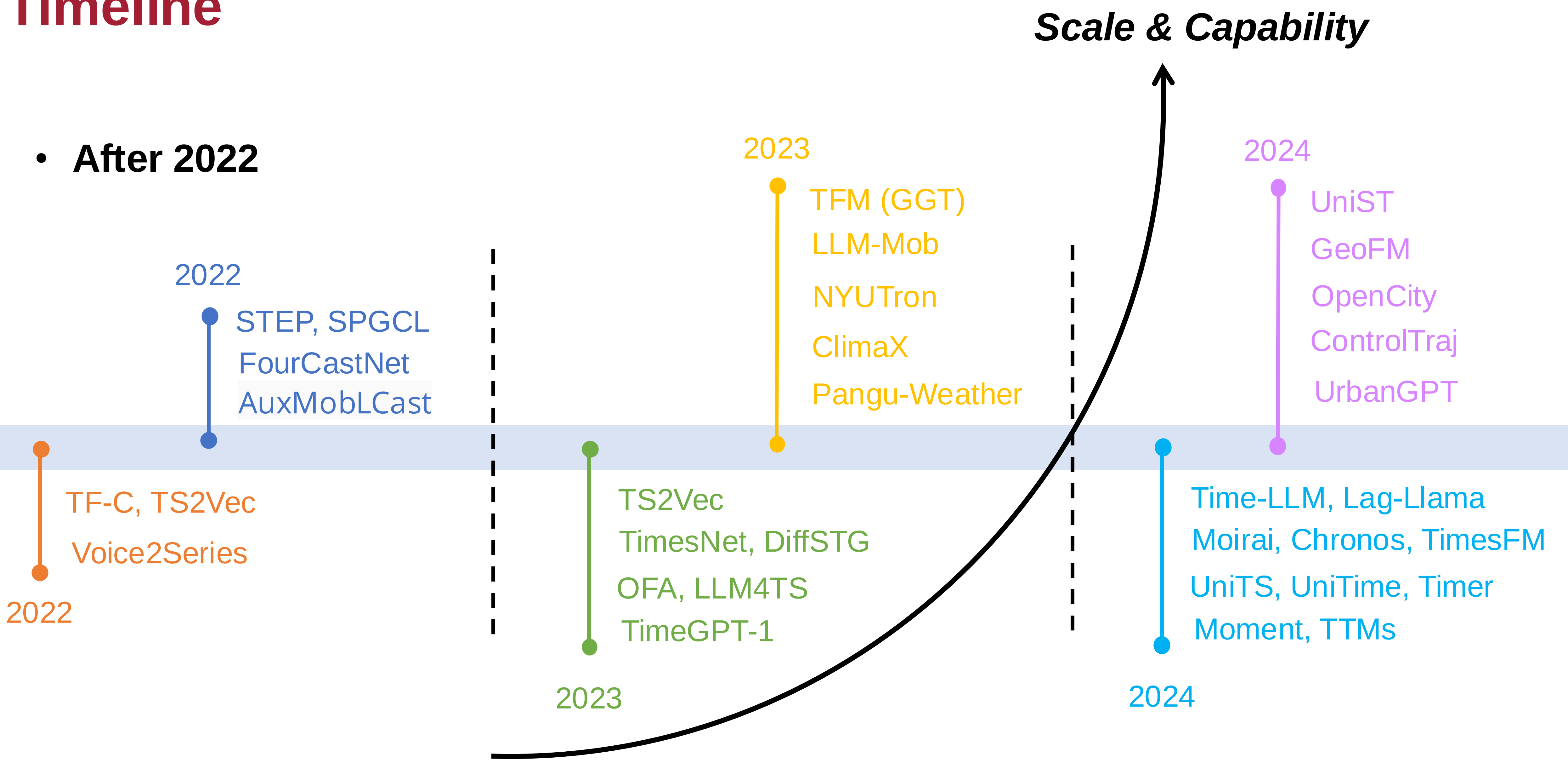
Timeline

- **After 2022**



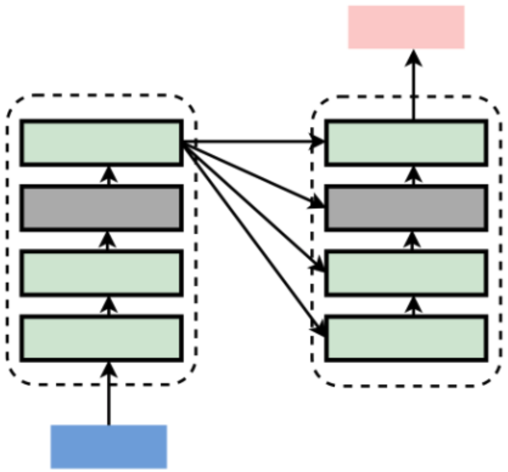
Timeline

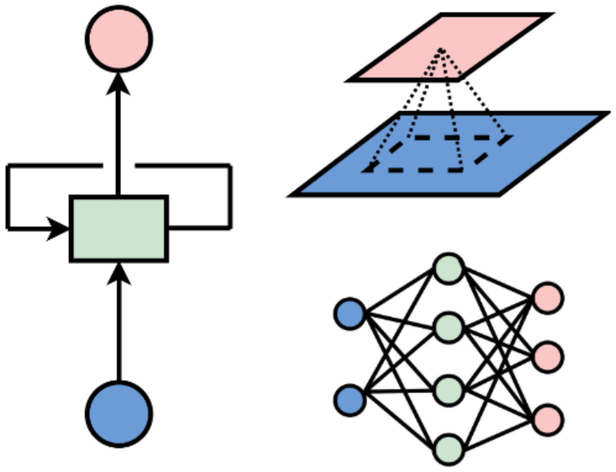
- **After 2022**

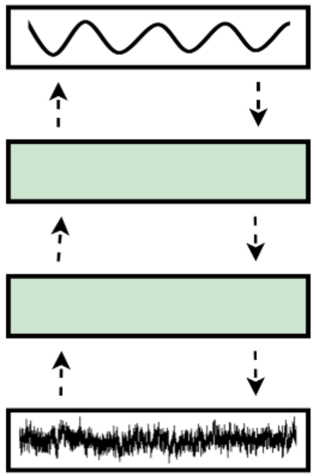


Deep Time Series Models

- **Architecture**







Transformer Models

- *Encoder-only*
- *Decoder-only*

Non-Transformer Models

- *RNNs* - *MLP* - *TCNs*
- …

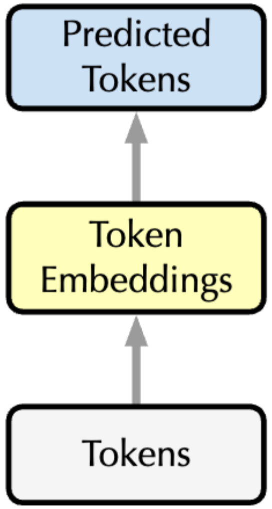
Diffusion Models

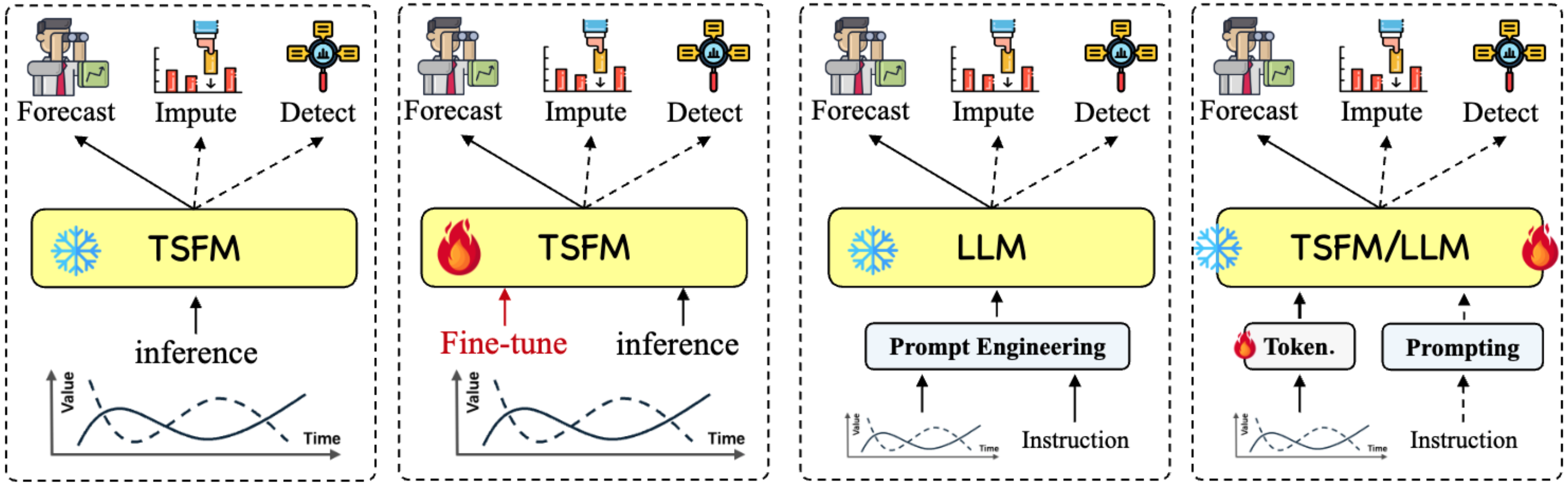
- *Unconditioned*
- *Conditioned*

Liang, Y., Wen, H., Nie, Y., Jiang, Y., Jin, M., Song, D., ... & Wen, Q. (2024). Foundation models for time series analysis: A tutorial and survey. In KDD'24

Deep Time Series Models

- **Pipeline**





(a) Direct usage (b) Tuning-based (c) Prompting-based (d) Tokenization-based

Task-specific training
or
Pre-training

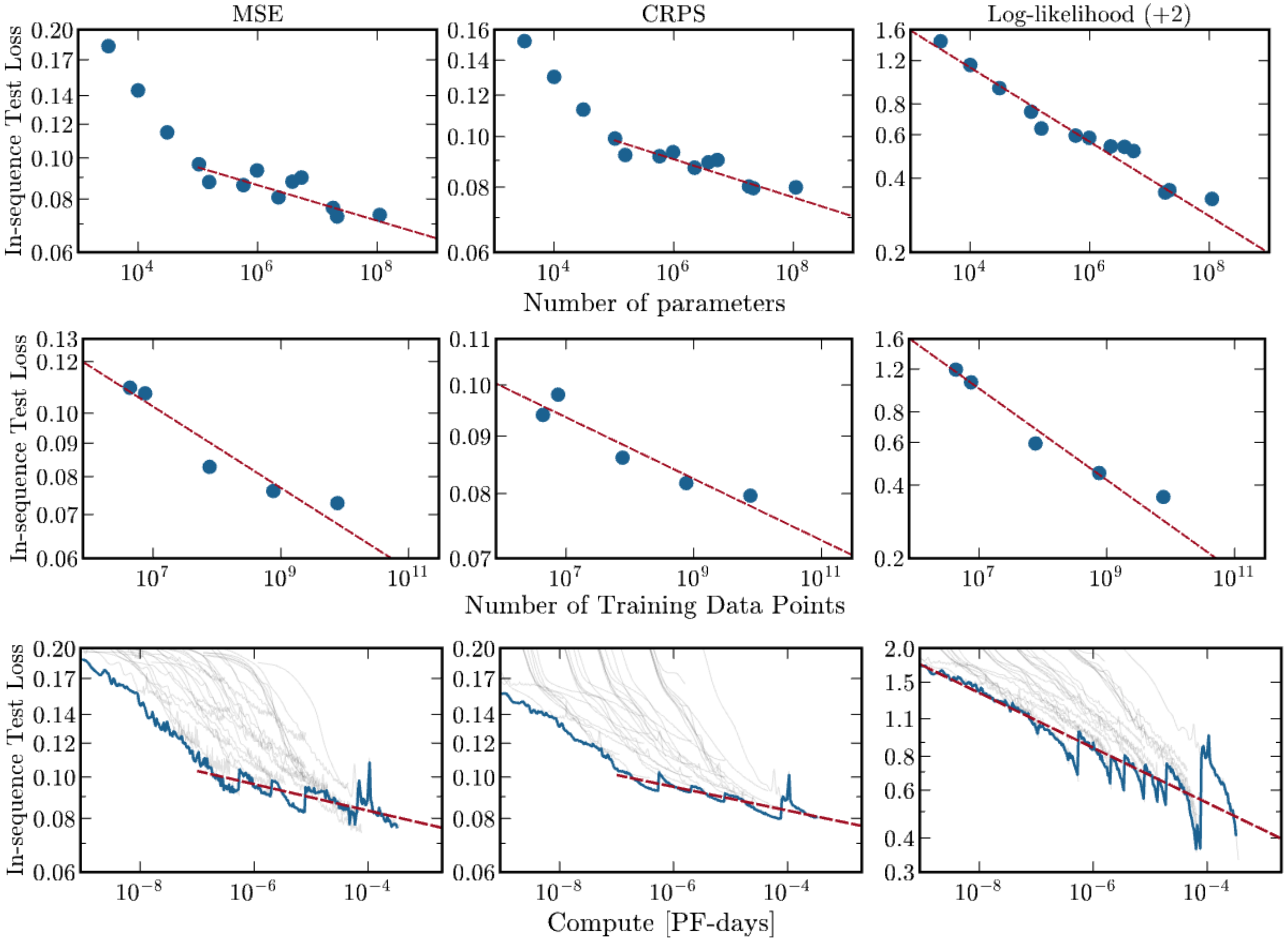
Adaptation

Liang, Y., Wen, H., Nie, Y., Jiang, Y., Jin, M., Song, D., ... & Wen, Q. (2024). Foundation models for time series analysis: A tutorial and survey. In KDD'24
Liu, Y., Zhang, H., Li, C., Huang, X., Wang, J., & Long, M. Timer: Generative Pre-trained Transformers Are Large Time Series Models. In ICML'24.

Scaling Laws

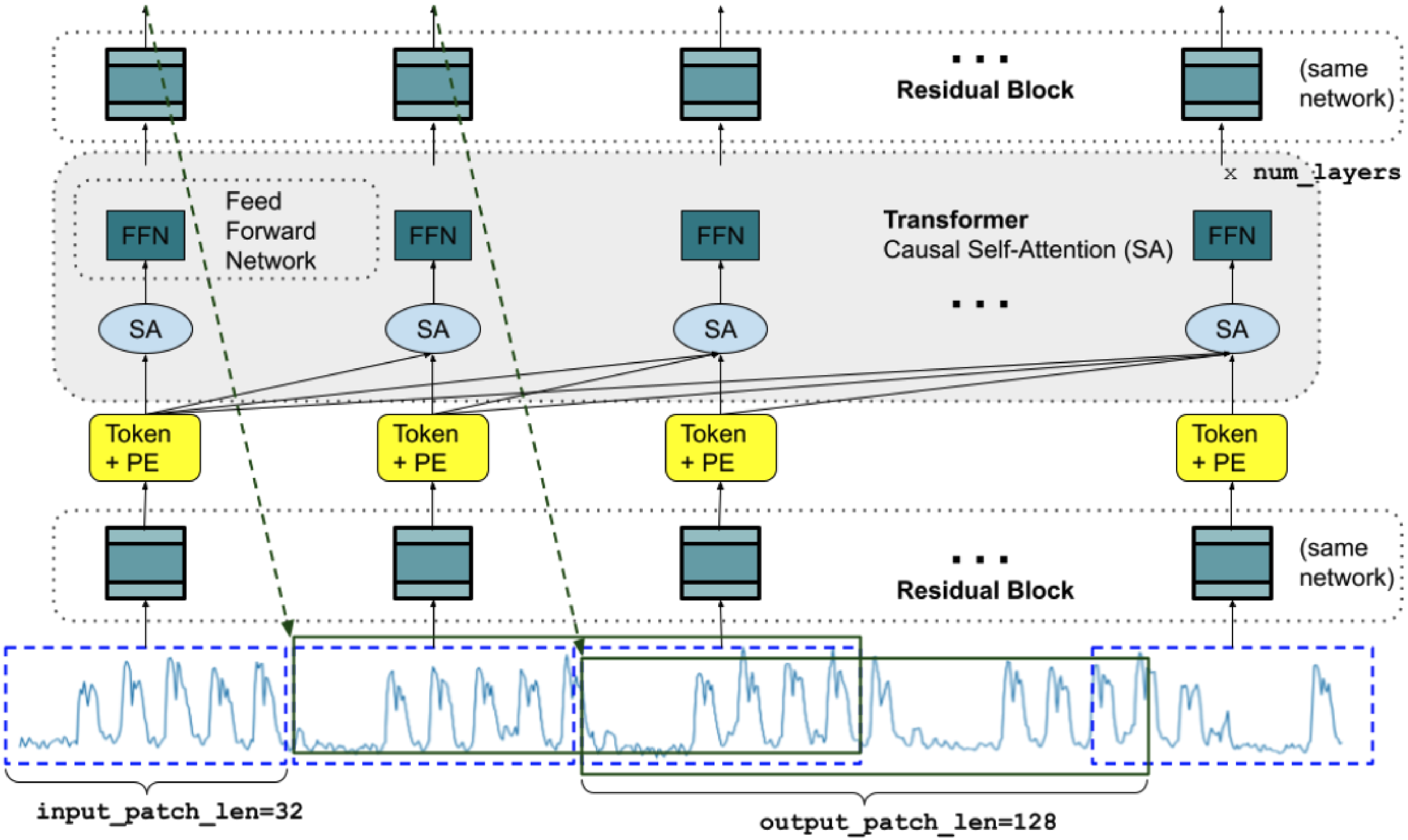
- **Three key aspects**
  - Model parameters (e.g., 10K to 100M)
  - Training tokens (e.g., 10M to 8B)
  - Computation (e.g., PF-day budget)

"Large time series models scales approximately as a power law with all three quantities" -- Edwards et al.

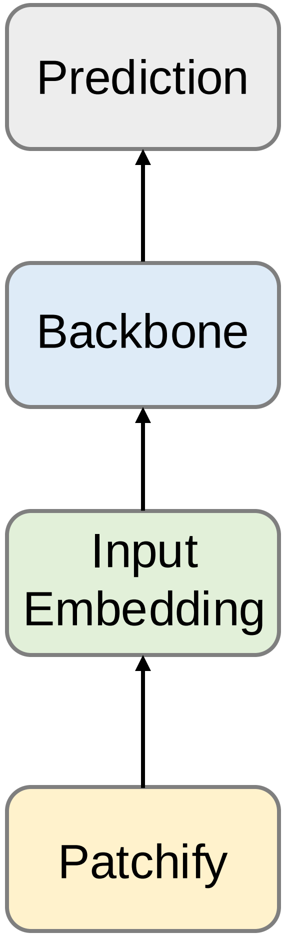


Edwards, T. D., Alvey, J., Alsing, J., Nguyen, N. H., & Wandelt, B. D. (2024). Scaling-laws for Large Time-series Models. arXiv preprint arXiv:2405.13867.

Transformer-based Models



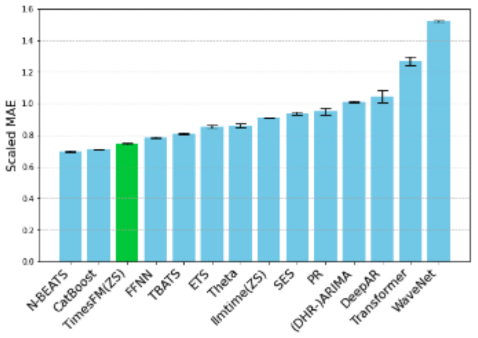
(Decoder-only)



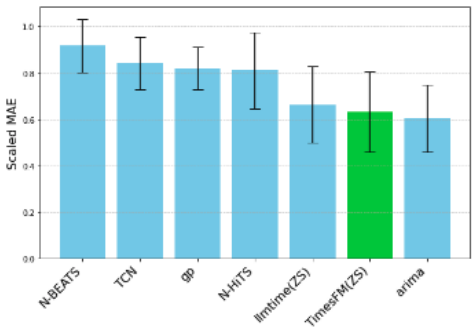
Das, A., Kong, W., Sen, R., & Zhou, Y. A decoder-only foundation model for time-series forecasting. In ICML'24.

Transformer-based Models

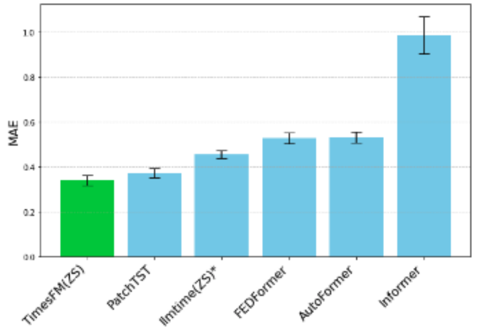
(a) Forecasting performance



(a) Monash Archive (Godahewa et al., 2021)

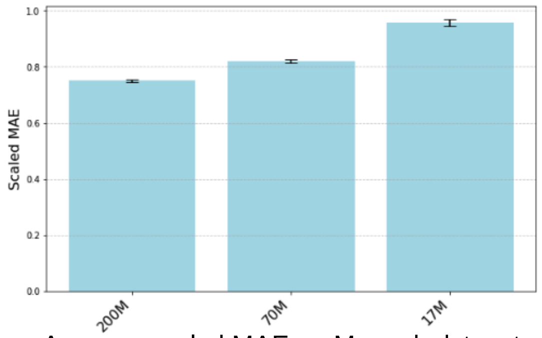


(b) Darts (Herzen et al., 2022)



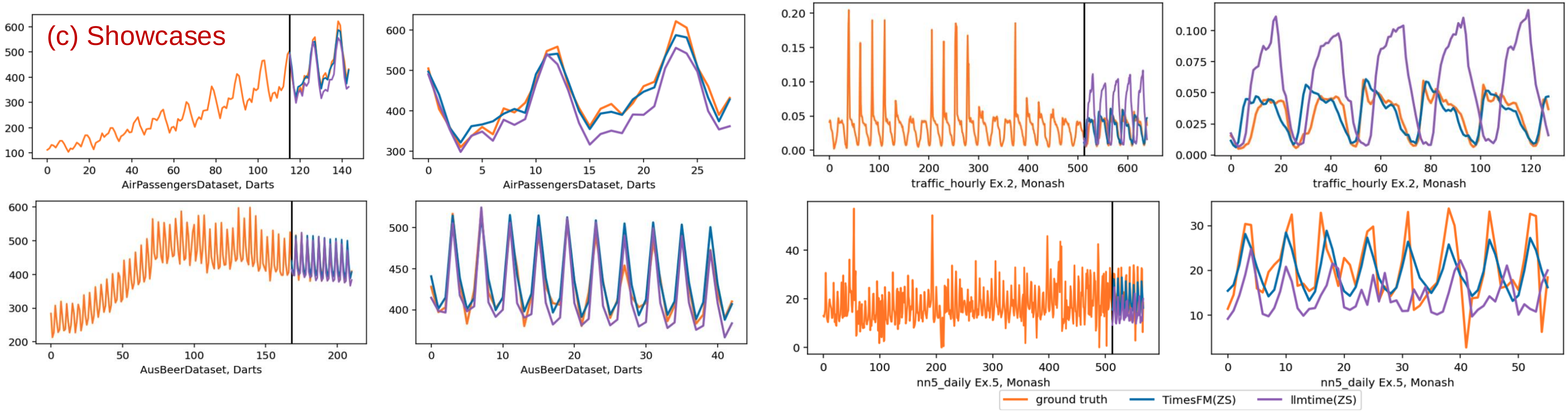
(c) ETT (Horizons 96 and 192) (Zhou et al., 2021)

(b) Scalability



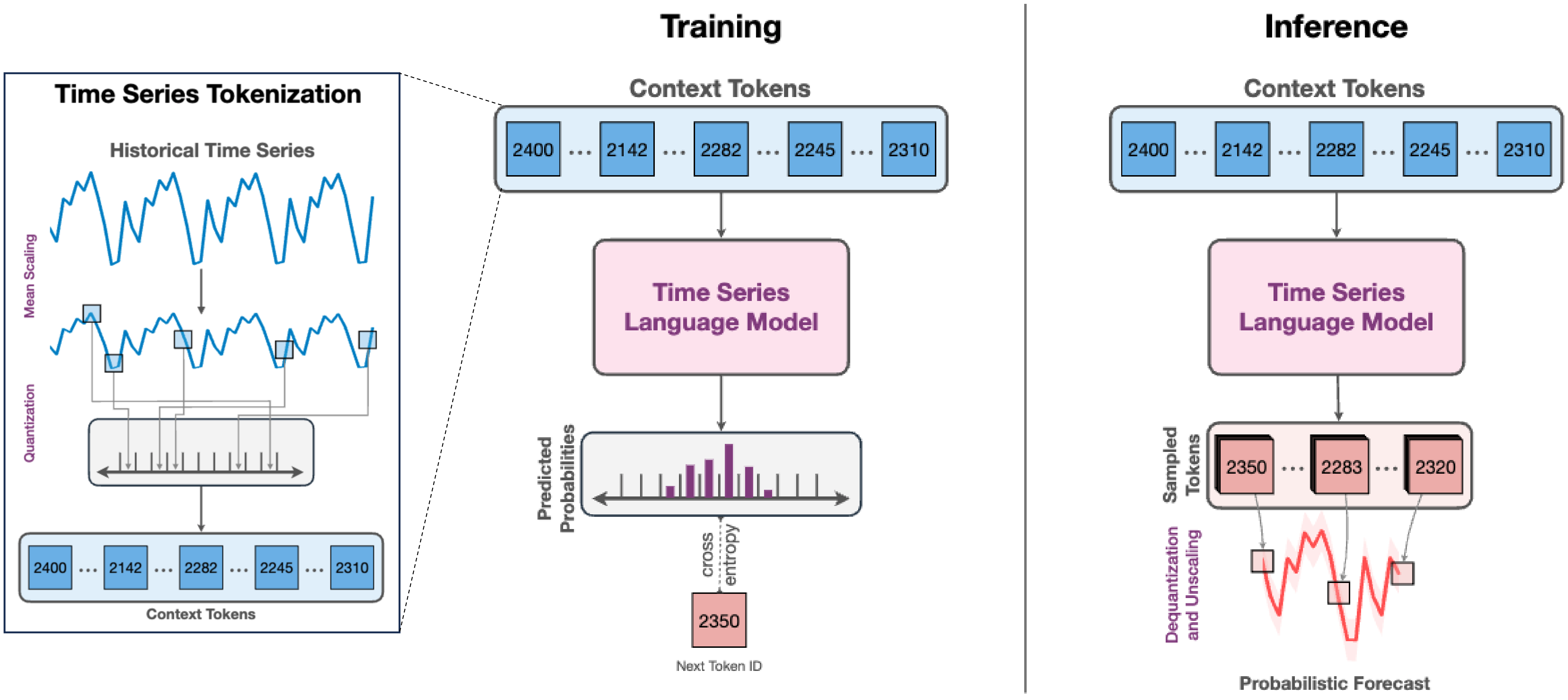
Average scaled MAE on Monash datasets for three different TimesFM model sizes

(c) Showcases



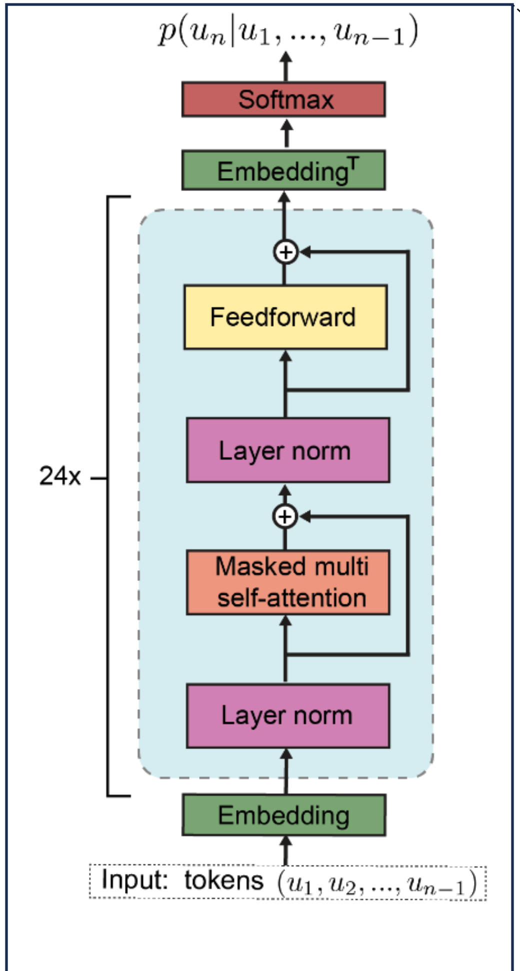
Das, A., Kong, W., Sen, R., & Zhou, Y. A decoder-only foundation model for time-series forecasting. In ICML'24.

Transformer-based Models

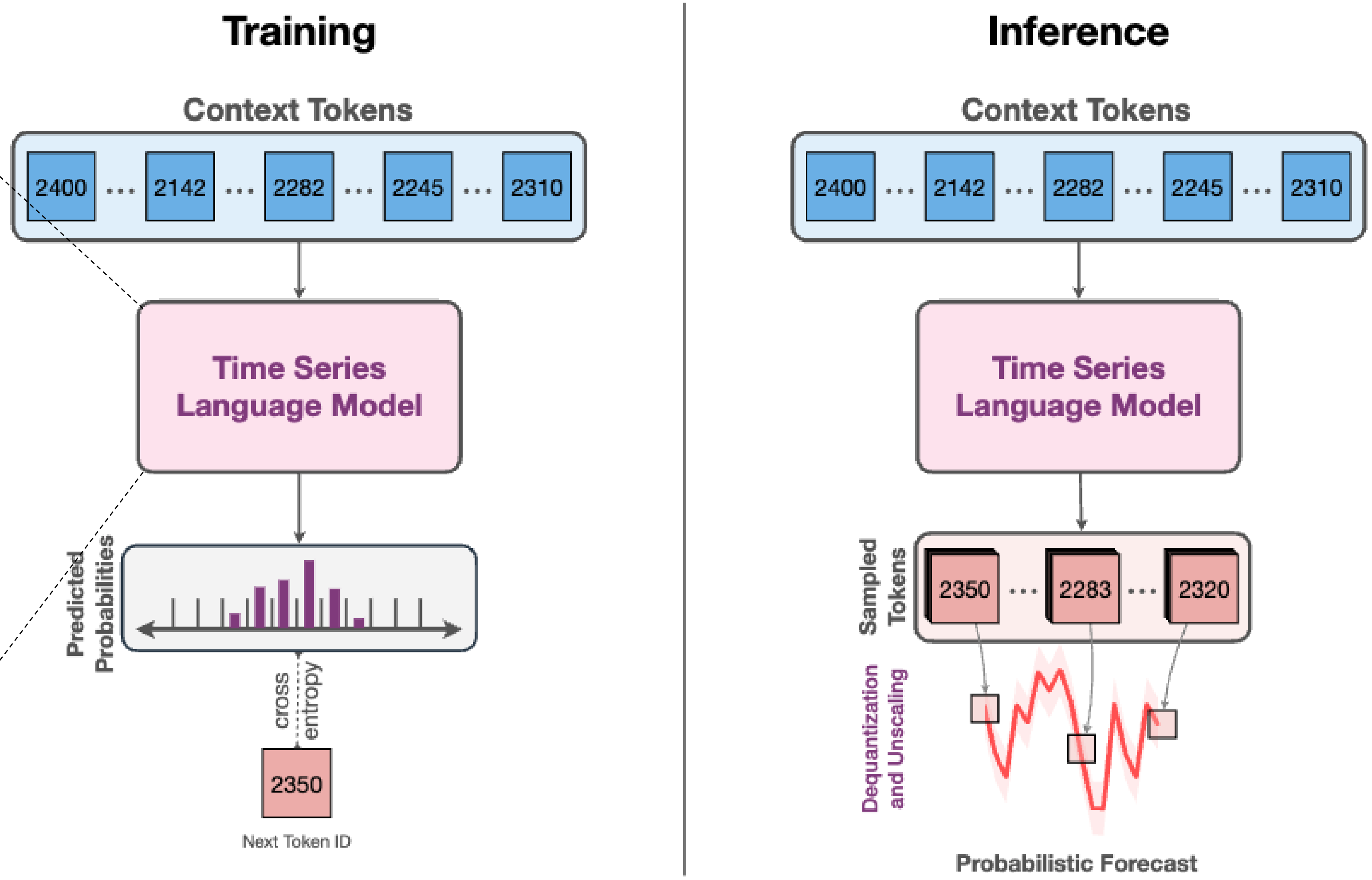


Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., ... & Wang, Y. (2024). Chronos: Learning the language of time series. arXiv preprint arXiv:2403.07815.
Heilbron, M., Ehinger, B., Hagoort, P., & De Lange, F. P. (2019). Tracking naturalistic linguistic predictions with deep neural language models. arXiv preprint arXiv:1909.04400.

Transformer-based Models



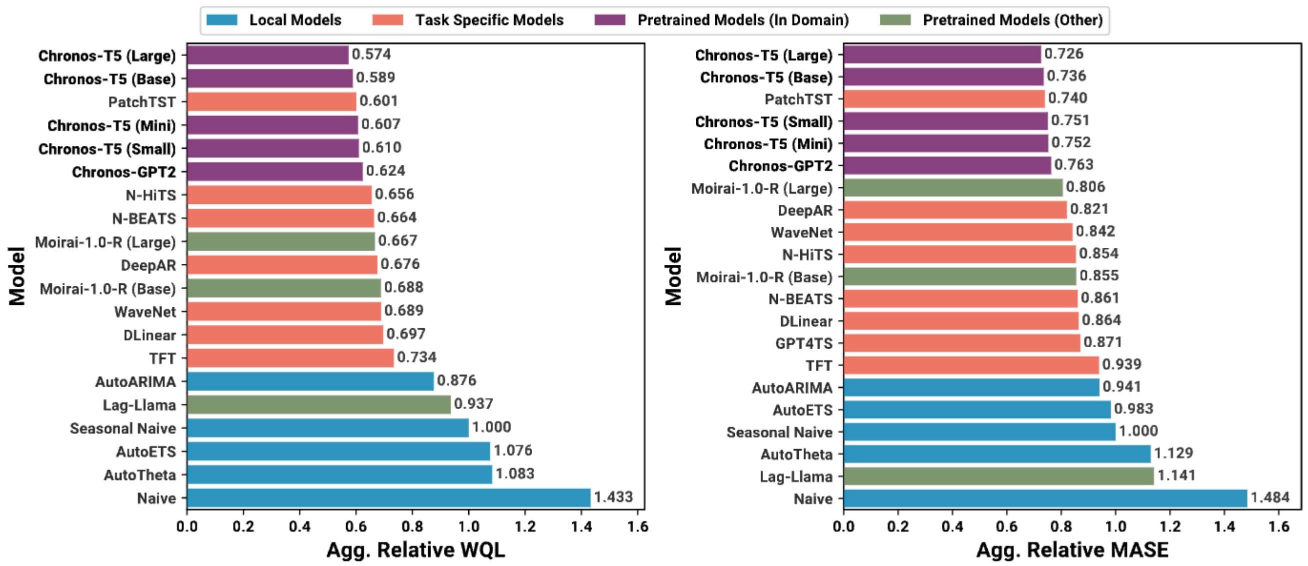
Example: GPT-2



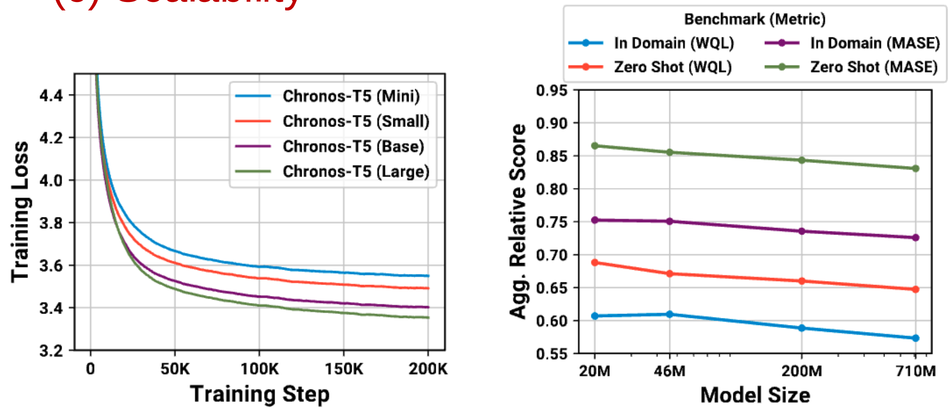
Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., … & Wang, Y. (2024). Chronos: Learning the language of time series. arXiv preprint arXiv:2403.07815.
Heilbron, M., Ehinger, B., Hagoort, P., & De Lange, F. P. (2019). Tracking naturalistic linguistic predictions with deep neural language models. arXiv preprint arXiv:1909.04400.

Transformer-based Models

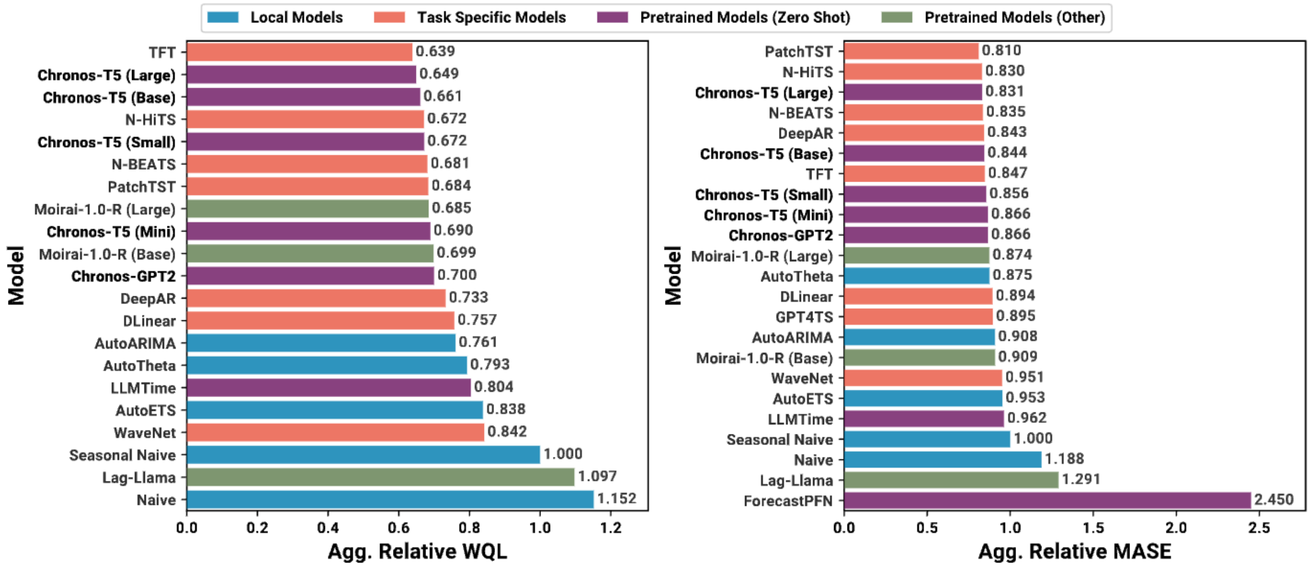
(a) In-domain avg. results



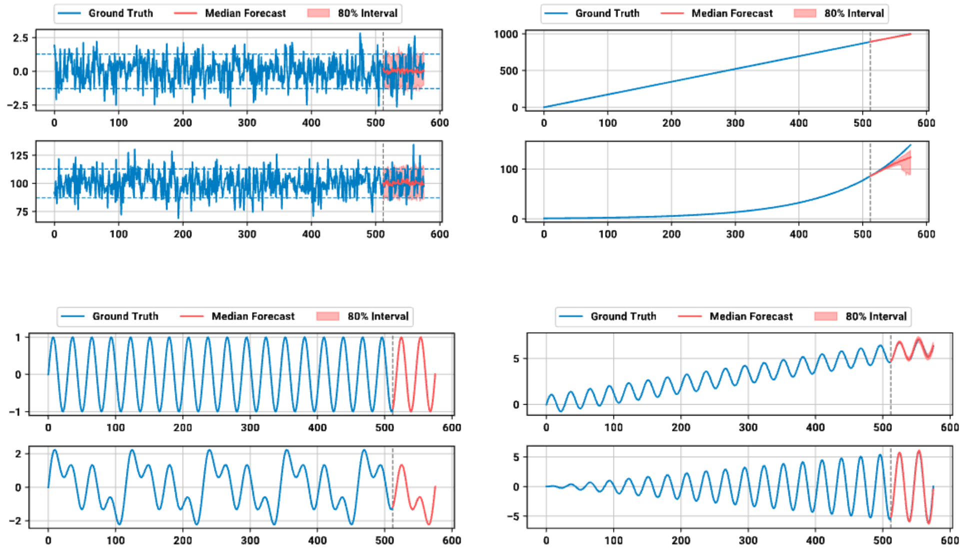
(c) Scalability



(b) Zero-shot avg. results

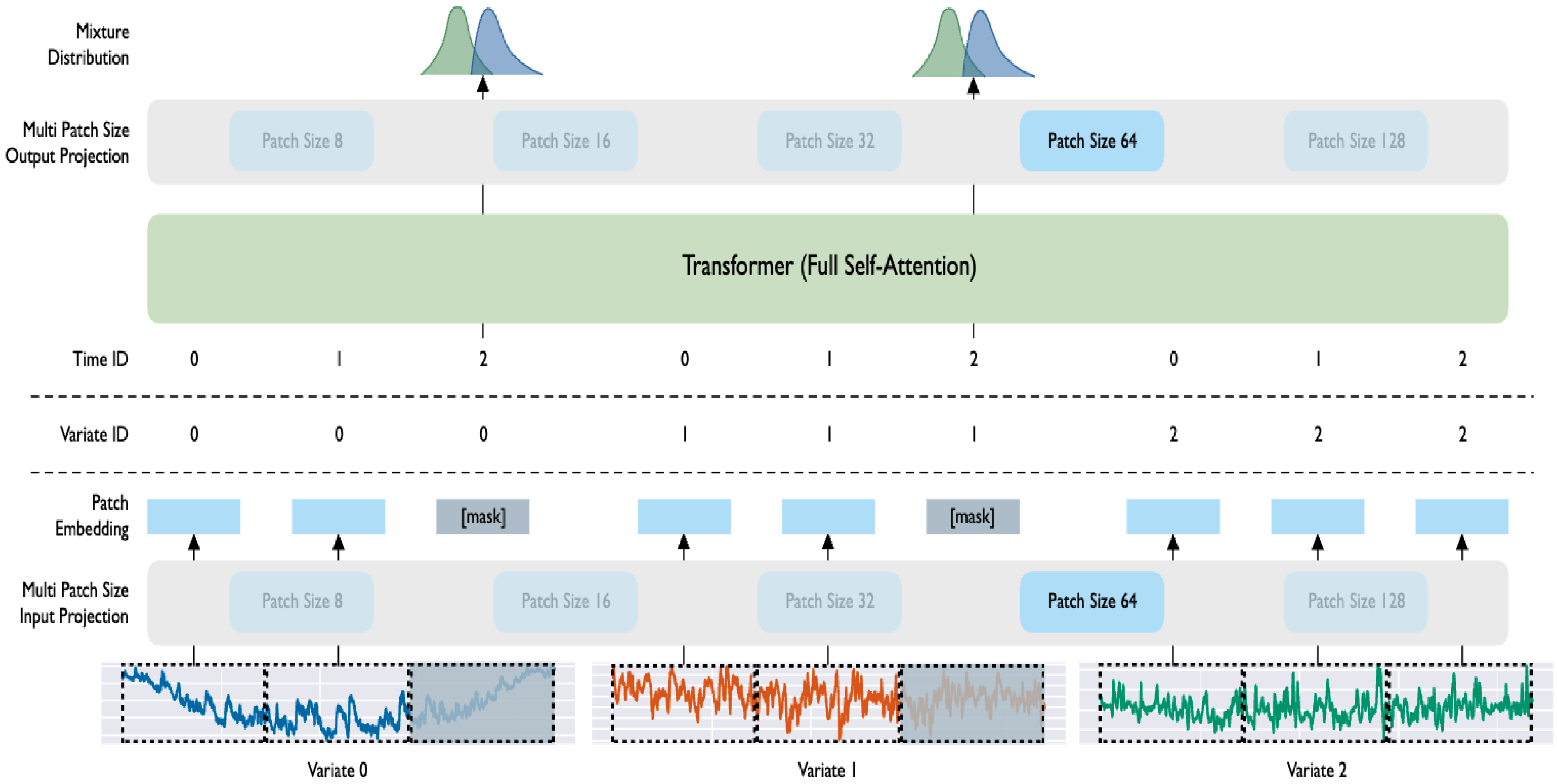


(d) Showcases

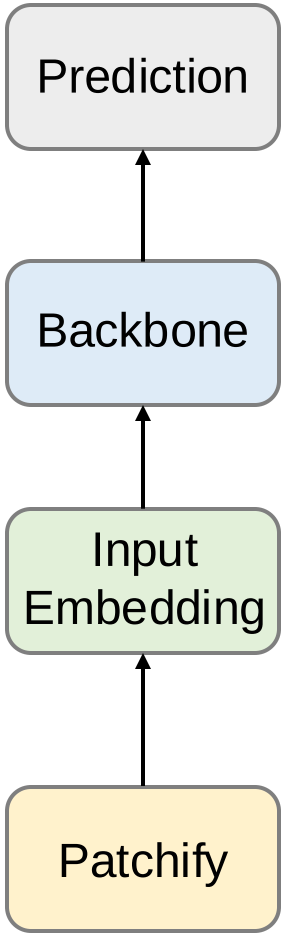


Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., ... & Wang, Y. (2024). Chronos: Learning the language of time series. arXiv preprint arXiv:2403.07815.

Transformer-based Models



(Encoder-only)

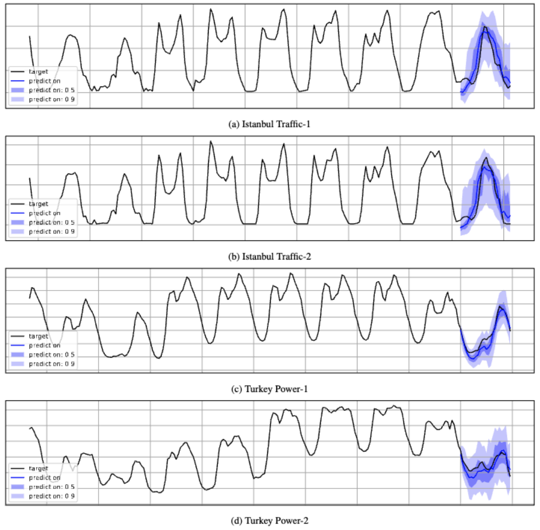


Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., & Sahoo, D. Unified Training of Universal Time Series Forecasting Transformers. In ICML'24.

Transformer-based Models

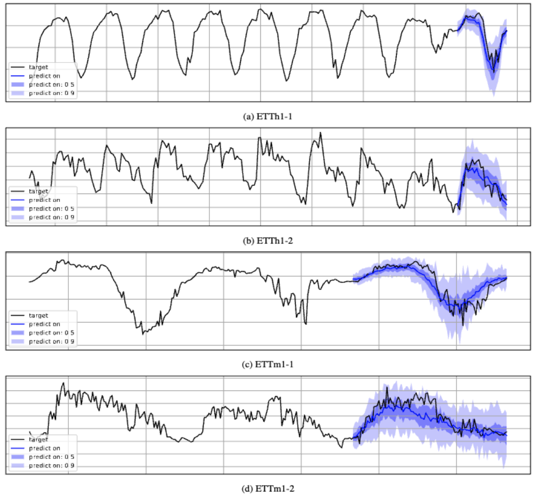
(a) Probabilistic forecasting

		Zero-shot			Full-shot				Baseline	
		MOIRAI$_{Small}$	MOIRAI$_{Base}$	MOIRAI$_{Large}$	PatchTST	TiDE	TFT	DeepAR	AutoARIMA	Seasonal Naive
Electricity	CRPS	0.072	0.055	0.050	0.052±0.00	**0.048±0.00**	0.050±0.00	0.065±0.01	0.327	0.070
	MSIS	7.999	6.172	5.875	5.744±0.12	**5.672±0.08**	6.278±0.24	6.893±0.82	29.412	35.251
Solar	CRPS	0.471	0.419	**0.406**	0.518±0.09	0.420±0.00	0.446±0.03	0.431±0.01	1.055	0.512
	MSIS	8.425	7.011	**6.250**	8.447±1.59	13.754±0.32	8.057±3.51	11.181±0.67	25.849	48.130
Walmart	CRPS	0.103	0.093	0.098	0.082±0.01	**0.077±0.00**	0.087±0.00	0.121±0.00	0.124	0.151
	MSIS	9.371	8.421	8.520	**6.005±0.21**	6.258±0.12	8.718±0.10	12.502±0.03	9.888	49.458
Weather	CRPS	0.049	**0.041**	0.051	0.059±0.01	0.054±0.00	0.043±0.00	0.132±0.11	0.252	0.068
	MSIS	5.236	5.136	**4.962**	7.759±0.49	8.095±1.74	7.791±0.44	21.651±17.34	19.805	31.293
Istanbul Traffic	CRPS	0.173	0.116	0.112	0.112±0.00	0.110±0.01	0.110±0.01	**0.108±0.00**	0.589	0.257
	MSIS	5.937	4.461	4.277	**3.813±0.09**	4.752±0.17	4.057±0.44	4.094±0.31	16.317	45.473
Turkey Power	CRPS	0.048	0.040	**0.036**	0.054±0.01	0.046±0.01	0.039±0.00	0.066±0.02	0.116	0.085
	MSIS	7.127	6.766	**6.341**	8.978±0.51	8.579±0.52	7.943±0.31	13.520±1.17	14.863	36.256


(a) Istanbul Traffic-1
(b) Istanbul Traffic-2
(c) Turkey Power-1
(d) Turkey Power-2

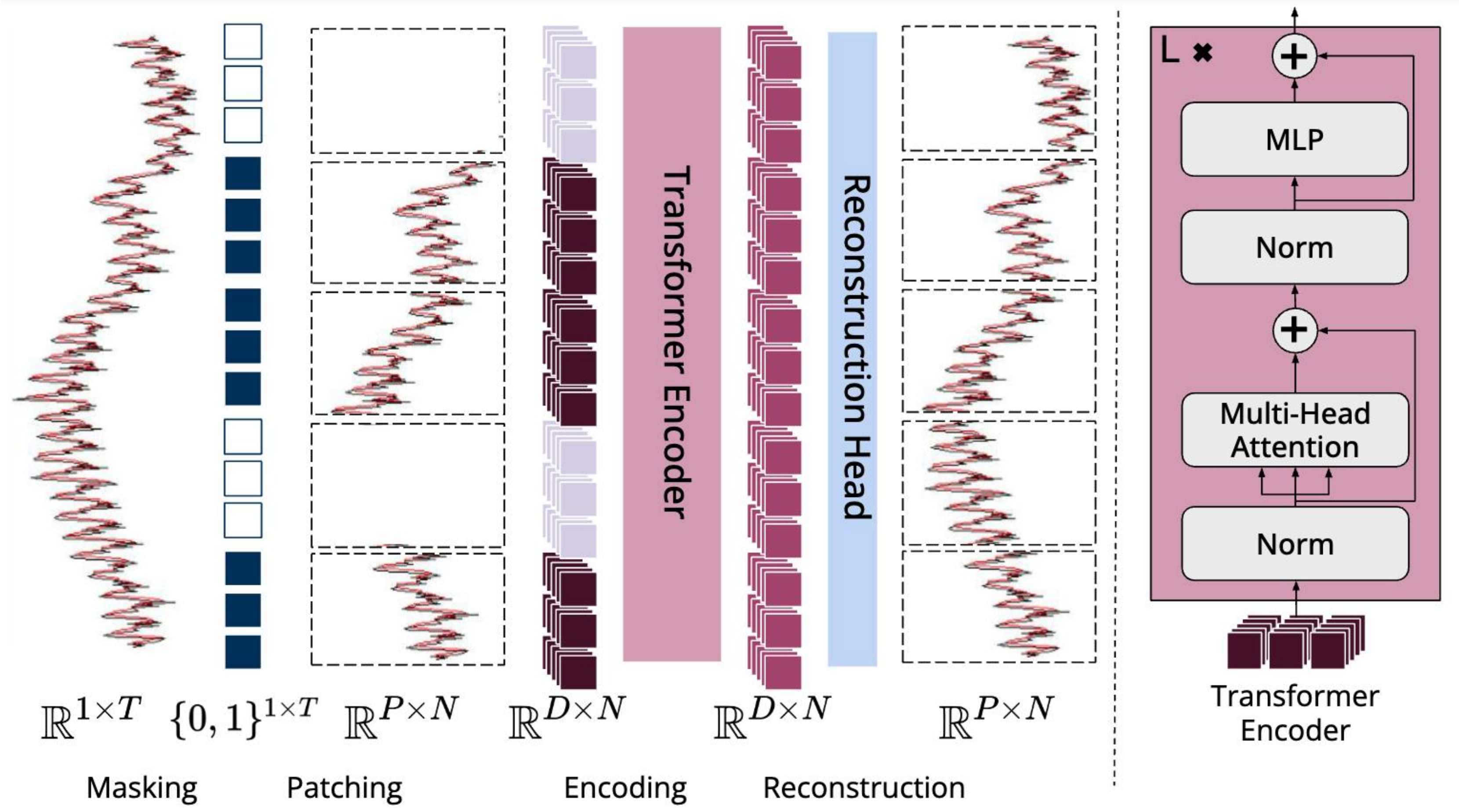
(b) Long sequence forecasting

		Zero-shot			Full-shot							
		MOIRAI$_{Small}$	MOIRAI$_{Base}$	MOIRAI$_{Large}$	iTransformer	TimesNet	PatchTST	Crossformer	TiDE	DLinear	SCINet	FEDformer
ETTh1	MSE	**0.400**	0.434	0.510	0.454	0.458	0.469	0.529	0.541	0.456	0.747	0.44
	MAE	**0.424**	0.438	0.469	0.448	0.450	0.455	0.522	0.507	0.452	0.647	0.46
ETTh2	MSE	**0.341**	0.345	0.354	0.383	0.414	0.387	0.942	0.611	0.559	0.954	0.437
	MAE	0.379	0.382	**0.376**	0.407	0.497	0.407	0.684	0.550	0.515	0.723	0.449
ETTm1	MSE	0.448	**0.381**	0.390	0.407	0.400	0.387	0.513	0.419	0.403	0.486	0.448
	MAE	0.409	**0.388**	0.389	0.410	0.406	0.400	0.495	0.419	0.407	0.481	0.452
ETTm2	MSE	0.300	**0.272**	0.276	0.288	0.291	0.281	0.757	0.358	0.35	0.571	0.305
	MAE	0.341	0.321	**0.320**	0.332	0.333	0.326	0.611	0.404	0.401	0.537	0.349
Electricity	MSE	0.233	0.188	0.188	**0.178**	0.193	0.216	0.244	0.252	0.212	0.268	0.214
	MAE	0.320	0.274	0.273	**0.270**	0.295	0.304	0.334	0.344	0.3	0.365	0.327
Weather	MSE	0.242	**0.238**	0.259	0.258	0.259	0.259	0.259	0.271	0.265	0.292	0.309
	MAE	0.267	**0.261**	0.275	0.278	0.287	0.281	0.315	0.320	0.317	0.363	0.36

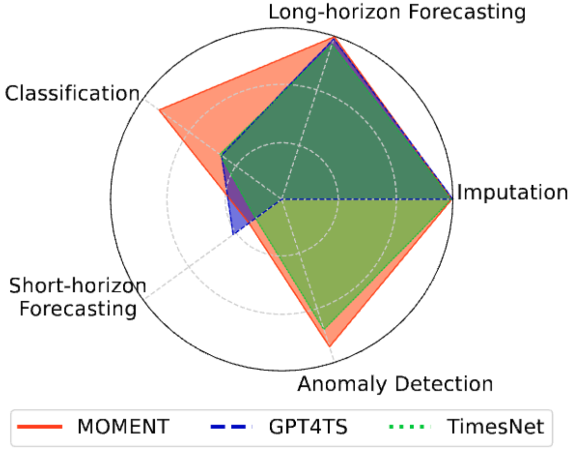

(a) ETTh1-1
(b) ETTh1-2
(c) ETTm1-1
(d) ETTm1-2

Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., & Sahoo, D. Unified Training of Universal Time Series Forecasting Transformers. In ICML'24.

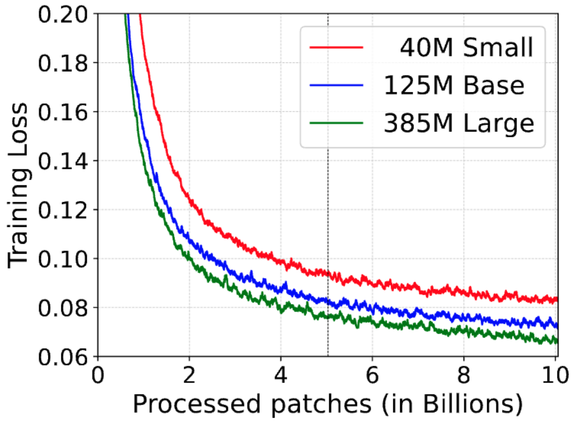
Transformer-based Models



(a) Performance



(b) Scalability

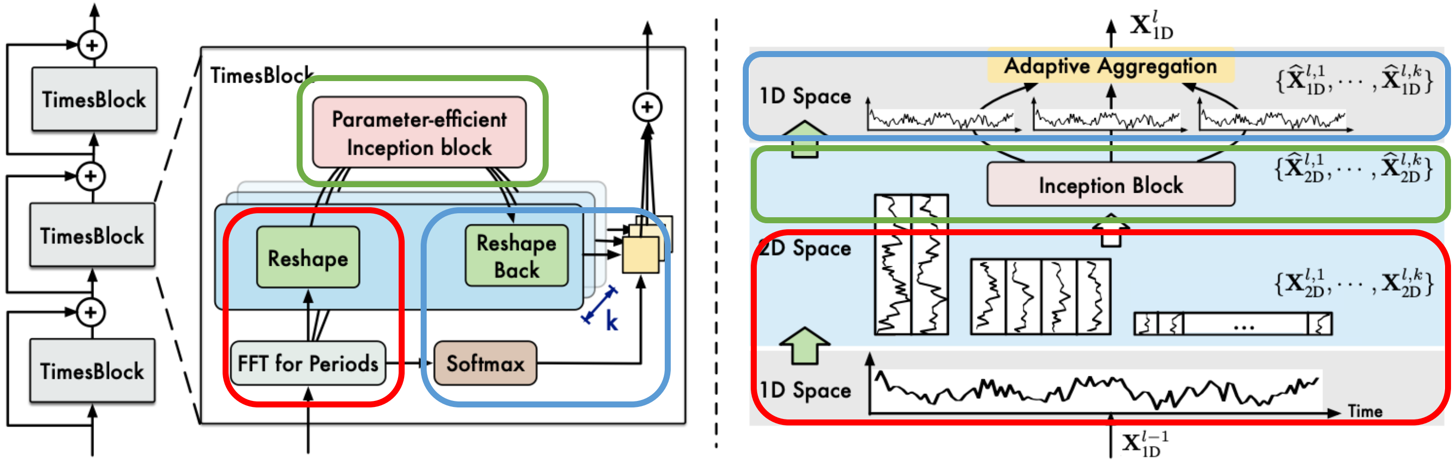


Goswami, M., Szafer, K., Choudhry, A., Cai, Y., Li, S., & Dubrawski, A. MOMENT: A Family of Open Time-series Foundation Models. In ICML'24.

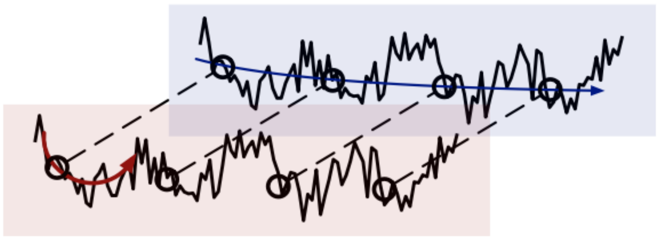
Non-Transformer Models

TimesNet is stacked by TimesBlocks in a residual way

TimesBlock learns representations in 2D space

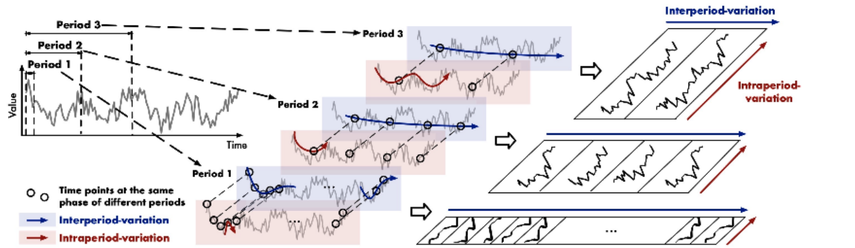


1 1D to 2D 2 2D representation learning 3 2D to 1D

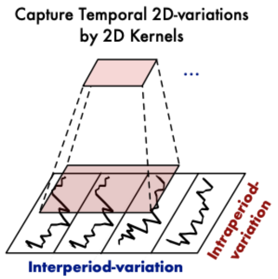


✓ **Intraperiod**: adjacent area, **short-term variations**

✓ **Interperiod**: same phase in adjacent periods, **long-term variations**



Unify intraperiod- and interperiod-variations in 2D space by reshape



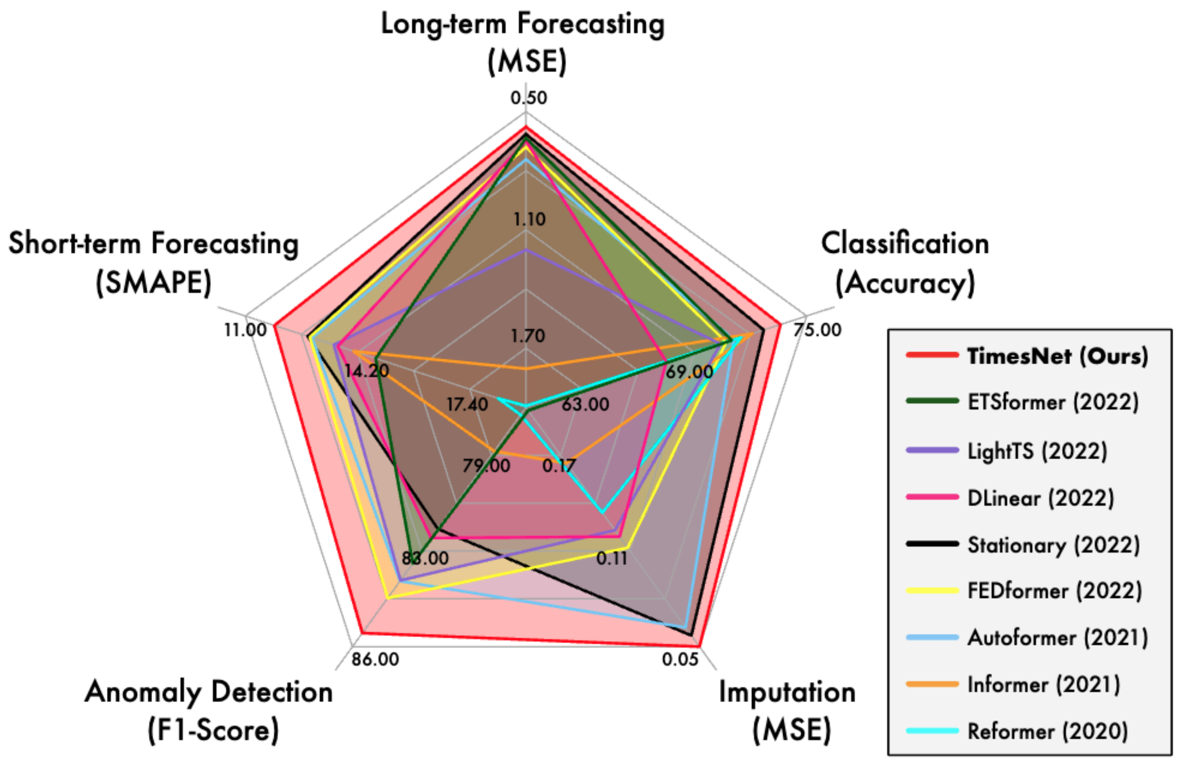
With temporal 2D-variations, we can

✓ Unify intraperiod- interperiod-variations

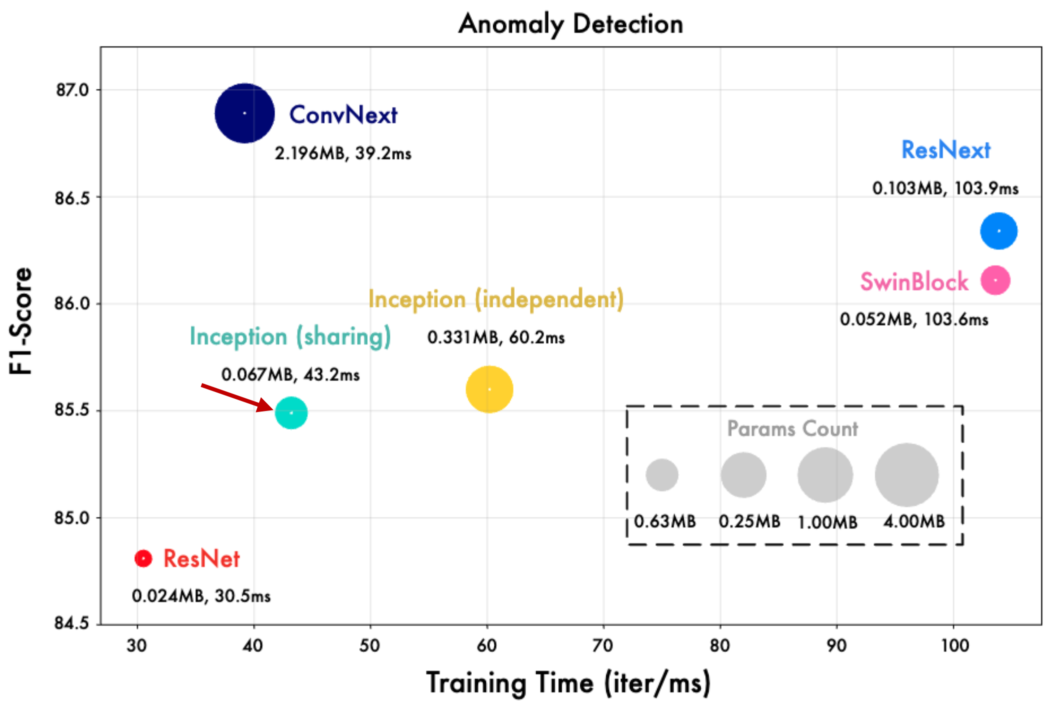
✓ Learn representations by 2D kernels

Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In ICLR'23.
https://ise.thss.tsinghua.edu.cn/~mlong/doc/foundation-models-for-time-series-analysis-gaitc23.pdf

Non-Transformer Models



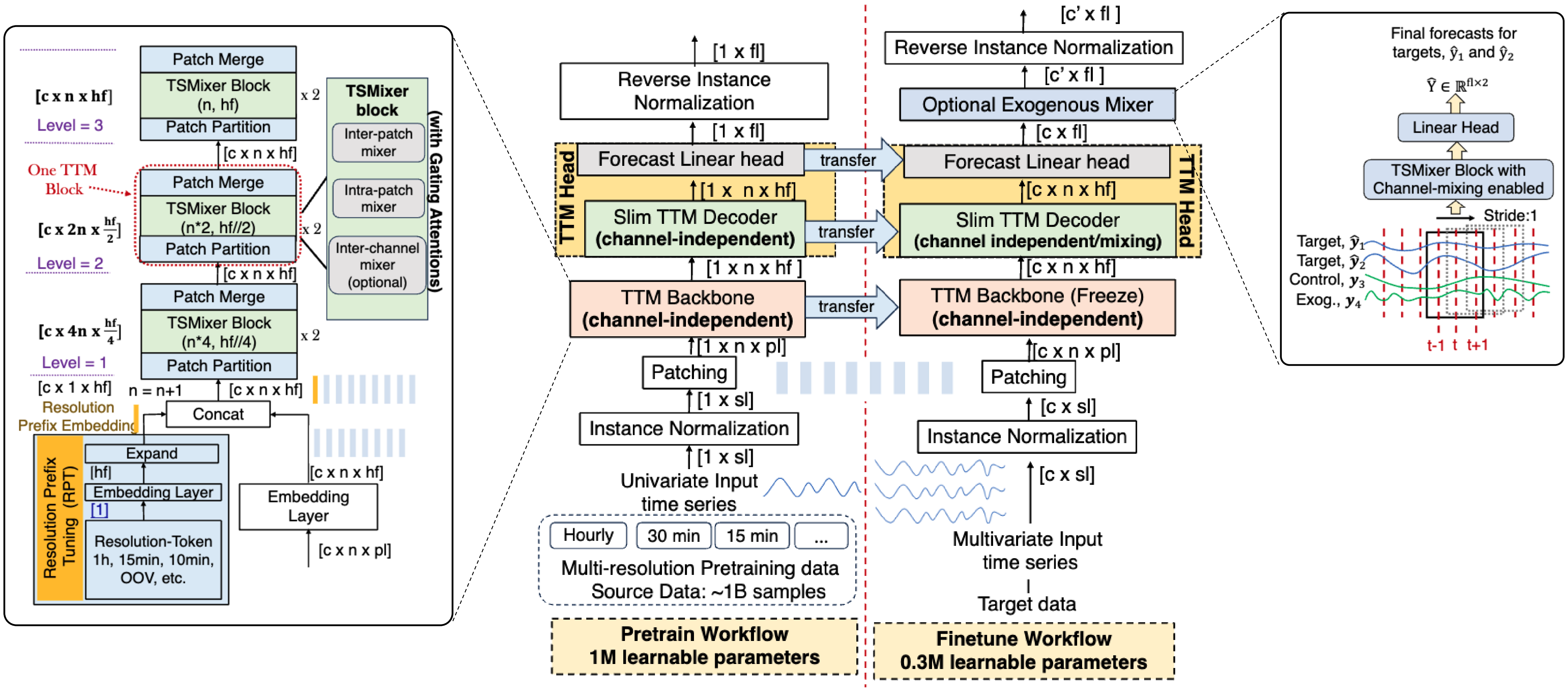
(a) Performance overview



(b) Model generality

Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In ICLR'23.
https://ise.thss.tsinghua.edu.cn/~mlong/doc/foundation-models-for-time-series-analysis-gaitc23.pdf

Non-Transformer-based Models



Ekambaram, V., Jati, A., Dayama, P., Mukherjee, S., Nguyen, N. H., Gifford, W. M., ... & Kalagnanam, J. (2024). Tiny Time Mixers (TTMs): Fast Pre-trained Models for Enhanced Zero/Few-Shot Forecasting of Multivariate Time Series. CoRR

Non-Transformer-based Models

Data	TTM_B	TTM_E	TTM_A	$Moirai_S$	$Moirai_B$	$Moirai_L$	TimesFM
ETTH1	**0.394**	0.404	0.4	0.4	0.434	0.51	0.479
ETTH2	0.345	0.335	**0.333**	0.341	0.346	0.354	0.403
ETTM1	0.386	0.38	**0.362**	0.448	0.382	0.39	0.429
ETTM2	0.281	0.271	**0.252**	0.3	0.272	0.276	0.334
Weather	0.237	0.238	**0.231**	0.242	0.238	0.26	-
Electricity	0.205	0.194	0.192	0.233	**0.188**	**0.188**	-
Size	**1M**	**4M**	**5M**	**14M**	**91M**	**311M**	**200M**
TTM_B *f-imp(%) s-imp(X)*				**6% ↑ 14X ↑**	**1% ↓ 91X ↑**	**4% ↑ 311X ↑**	**15% ↑ 200X ↑**
TTM_E *f-imp(%) s-imp(X)*				**7% ↑ 4X ↑**	**1% ↑ 23X ↑**	**6% ↑ 78X ↑**	**16% ↑ 50X ↑**
TTM_A *f-imp(%) s-imp(X)*				**10% ↑ 3X ↑**	**4% ↑ 18X ↑**	**9% ↑ 62X ↑**	**19% ↑ 40X ↑**

Zero-shot forecast-improvement (f-imp) and model size-improvement (s-imp) of TTM over Moirai and TimesFM.

Data	TTM_B	TTM_E	TTM_A	$Chronos_T$	$Chronos_S$	$Chronos_B$	$Chronos_L$	Lag-llama
ETTH1	**0.204**	0.227	0.214	0.311	0.302	0.252	0.266	0.334
ETTH2	**0.131**	0.151	0.162	0.177	0.16	0.164	0.155	0.168
ETTM1	0.206	0.239	**0.19**	0.839	0.486	0.49	0.538	0.842
ETTM2	0.124	0.128	**0.117**	0.206	0.174	0.19	0.187	0.308
Weather	0.039	0.032	0.043	0.043	0.046	**0.03**	0.033	0.126
Electricity	**0.335**	0.351	0.349	0.423	0.377	0.344	0.339	0.393
Traffic	0.246	**0.24**	0.244	0.291	0.3	0.28	0.269	0.243
Size	**1M**	**4M**	**5M**	**8M**	**46M**	**201M**	**709M**	**3M**
TTM_B *f-imp(%) s-imp(X)*				**32% ↑ 8X ↑**	**26% ↑ 46X ↑**	**17% ↑ 201X ↑**	**18% ↑ 709X ↑**	**40% ↑ 3X ↑**
TTM_E *f-imp(%) s-imp(X)*				**30% ↑ 2X ↑**	**24% ↑ 12X ↑**	**15% ↑ 50X ↑**	**16% ↑ 177X ↑**	**37% ↑ 1X ↓**
TTM_A *f-imp(%) s-imp(X)*				**28% ↑ 2X ↑**	**22% ↑ 9X ↑**	**12% ↑ 40X ↑**	**13% ↑ 142X ↑**	**37% ↑ 2X ↓**

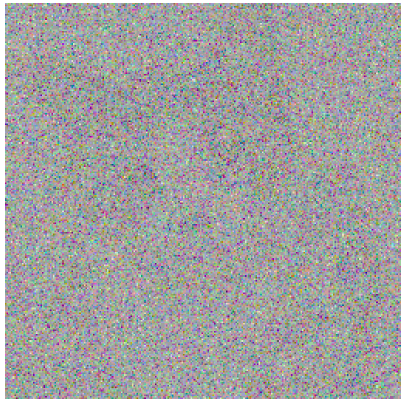
Zero-shot forecast-improvement (f-imp) and model size-improvement (s-imp) of TTM over Chronos and Lag-Llama.

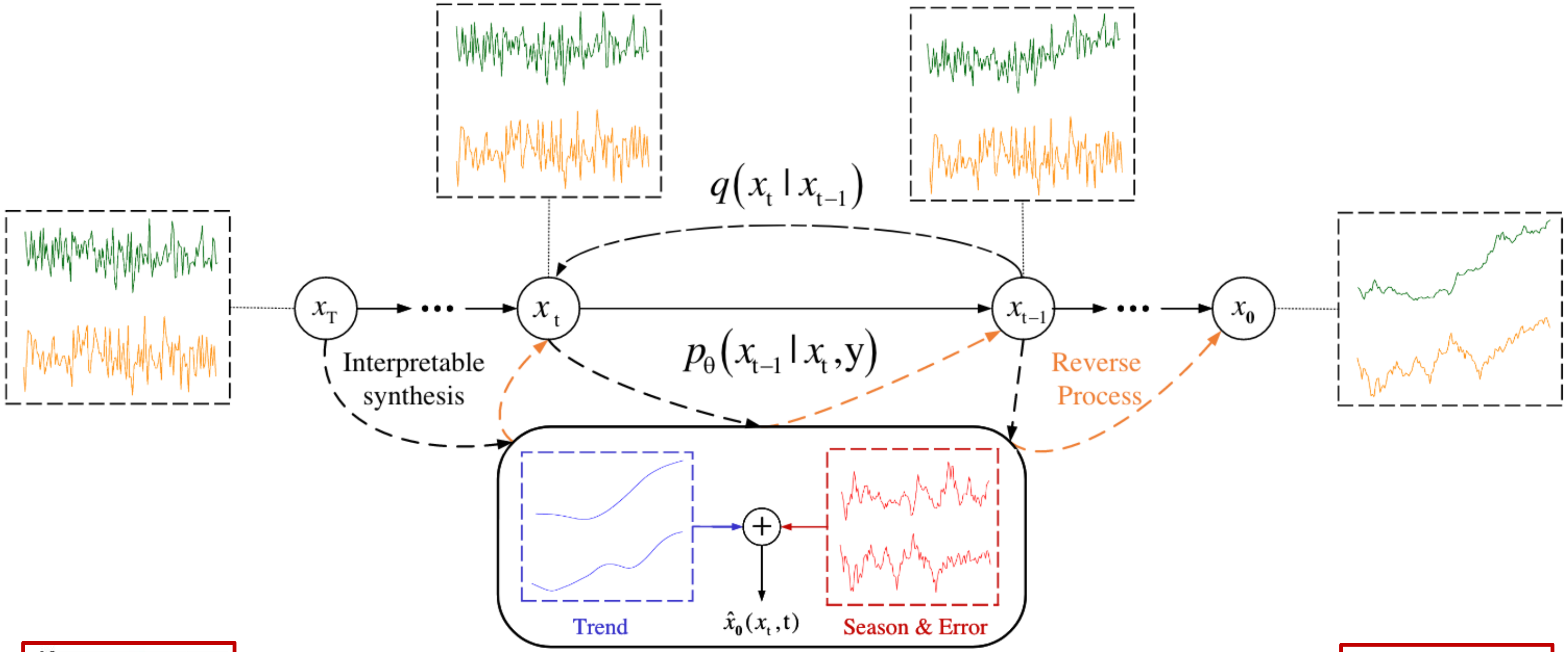
Model	GPU TIME (ms)	Params (M)	MEM (GB)	CPU TIME (s)
TTM_B	**4.7**	**0.8**	**0.06**	**0.01**
$Chronos_B$ (2024)	1395 (298X)	201 (251X)	16 (267X)	2340 (239KX)
$Chronos_L$ (2024)	1393 (298X)	709 (886X)	41 (683X)	2352 (240KX)
$Chronos_S$ (2024)	1386 (296X)	46 (58X)	6 (100X)	2349 (240KX)
$Chronos_T$ (2024)	1389 (297X)	8 (10X)	2 (33X)	2504 (256KX)
GPT4TS (NeurIPS '23)	13.9 (3X)	87 (109X)	1.34 (36X)	0.3 (26X)
Lag-Llama (2024)	1619 (346X)	2.4 (3X)	0.2 (3X)	37.5 (3830X)
$Moirai_S$ (ICML '24)	205 (44X)	14 (18X)	0.1 (2X)	1.4 (141X)
$Moirai_L$ (ICML '24)	693 (148X)	311 (389X)	2 (33X)	10.5 (1070X)
$Moirai_B$ (ICML '24)	335 (72X)	91 (114X)	1 (17X)	4.1 (421X)
Moment-L (ICML '24)	88 (19X)	348 (435X)	8 (133X)	1.4 (144X)
TimesFM (ICML '24)	24 (5X)	200 (250X)	2 (33X)	0.4 (46X)

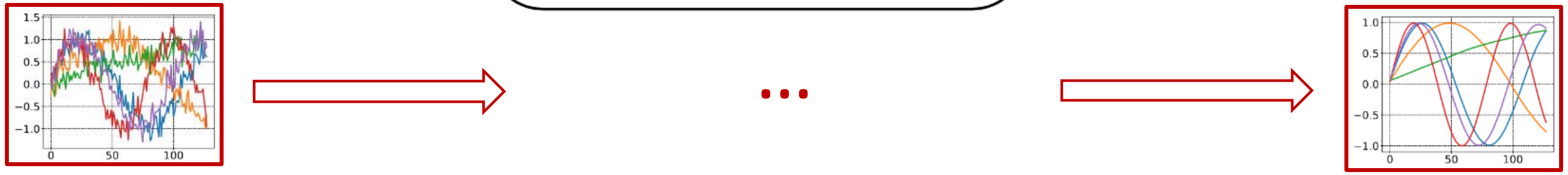
Computational improvement of TTM w.r.t. existing TS pre-trained models. Inference time per-batch in GPU and CPU, total parameters (Params), and maximum GPU memory usage (MEM) are reported.

Ekambaram, V., Jati, A., Dayama, P., Mukherjee, S., Nguyen, N. H., Gifford, W. M., ... & Kalagnanam, J. (2024). Tiny Time Mixers (TTMs): Fast Pre-trained Models for Enhanced Zero/Few-Shot Forecasting of Multivariate Time Series. CoRR

Diffusion Models





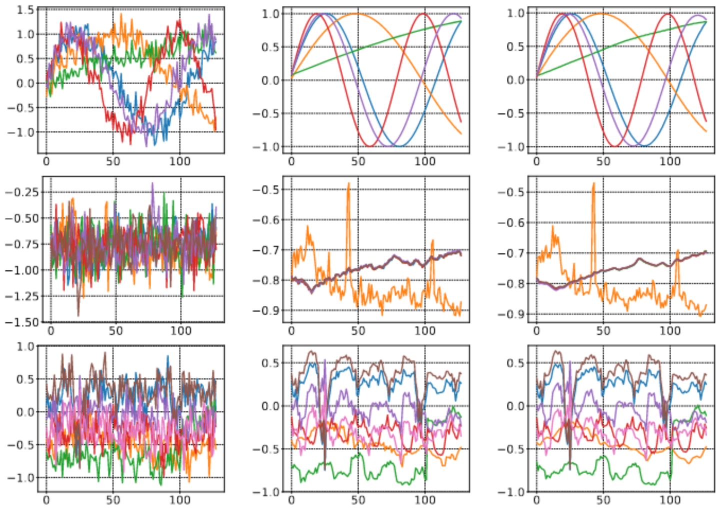


- Generating time series data using a diffusion model that maps Gaussian vectors to signals resembling those in a given dataset

Yuan, X., & Qiao, Y. Diffusion-TS: Interpretable Diffusion for General Time Series Generation. In ICLR'24

Diffusion Models

(a) Reconstruction



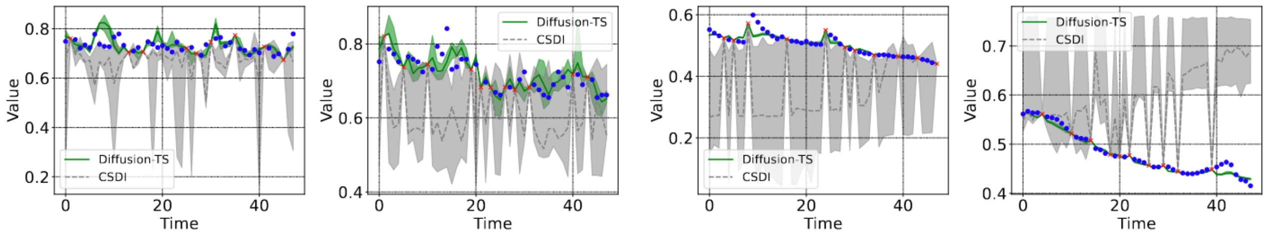
(a) Input (b) Origin (c) Output

(b) Unconditional gen.

Table 1: Results on Multiple Time-Series Datasets (Bold indicates best performance).

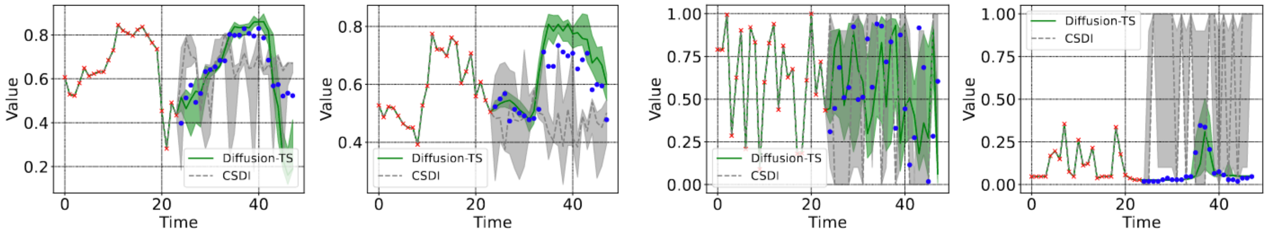
Metric	Methods	Sines	Stocks	ETTh	MuJoCo	Energy	fMRI
Context-FID Score (Lower the Better)	Diffusion-TS	**0.006±.000**	0.147±.025	**0.116±.010**	**0.013±.001**	**0.089±.024**	**0.105±.006**
	TimeGAN	0.101±.014	**0.103±.013**	0.300±.013	0.563±.052	0.767±.103	1.292±.218
	TimeVAE	0.307±.060	0.215±.035	0.805±.186	0.251±.015	1.631±.142	14.449±.969
	Diffwave	0.014±.002	0.232±.032	0.873±.061	0.393±.041	1.031±.131	0.244±.018
	DiffTime	0.006±.001	0.236±.074	0.299±.044	0.188±.028	0.279±.045	0.340±.015
	Cot-GAN	1.337±.068	0.408±.086	0.980±.071	1.094±.079	1.039±.028	7.813±.550
Correlational Score (Lower the Better)	Diffusion-TS	**0.015±.004**	**0.004±.001**	**0.049±.008**	**0.193±.027**	**0.856±.147**	**1.411±.042**
	TimeGAN	0.045±.010	0.063±.005	0.210±.006	0.886±.039	4.010±.104	23.502±.039
	TimeVAE	0.131±.010	0.095±.008	0.111±020	0.388±.041	1.688±.226	17.296±.526
	Diffwave	0.022±.005	0.030±.020	0.175±.006	0.579±.018	5.001±.154	3.927±.049
	DiffTime	0.017±.004	0.006±.002	0.067±.005	0.218±.031	1.158±.095	1.501±.048
	Cot-GAN	0.049±.010	0.087±.004	0.249±.009	1.042±.007	3.164±.061	26.824±.449
Discriminative Score (Lower the Better)	Diffusion-TS	**0.006±.007**	**0.067±.015**	**0.061±.009**	**0.008±.002**	**0.122±.003**	**0.167±.023**
	TimeGAN	0.011±.008	0.102±.021	0.114±.055	0.238±.068	0.236±.012	0.484±.042
	TimeVAE	0.041±.044	0.145±.120	0.209±.058	0.230±.102	0.499±.000	0.476±.044
	Diffwave	0.017±.008	0.232±.061	0.190±.008	0.203±.096	0.493±.004	0.402±.029
	DiffTime	0.013±.006	0.097±.016	0.100±.007	0.154±.045	0.445±.004	0.245±.051
	Cot-GAN	0.254±.137	0.230±.016	0.325±.099	0.426±.022	0.498±.002	0.492±.018
Predictive Score (Lower the Better)	Diffusion-TS	**0.093±.000**	**0.036±.000**	**0.119±.002**	**0.007±.000**	**0.250±.000**	**0.099±.000**
	TimeGAN	0.093±.019	0.038±.001	0.124±.001	0.025±.003	0.273±.004	0.126±.002
	TimeVAE	0.093±.000	0.039±.000	0.126±.004	0.012±.002	0.292±.000	0.113±.003
	Diffwave	0.093±.000	0.047±.000	0.130±.001	0.013±.000	0.251±.000	0.101±.000
	DiffTime	0.093±.000	0.038±.001	0.121±.004	0.010±.001	0.252±.000	0.100±.000
	Cot-GAN	0.100±.000	0.047±.001	0.129±.000	0.068±.009	0.259±.000	0.185±.003
	Original	0.094±.001	0.036±.001	0.121±.005	0.007±.001	0.250±.003	0.090±.001

(c) Conditional gen.



(a) ETTh (b) Energy

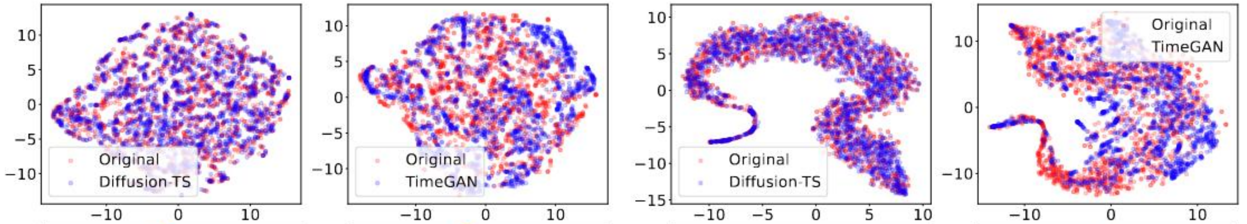
Imputation



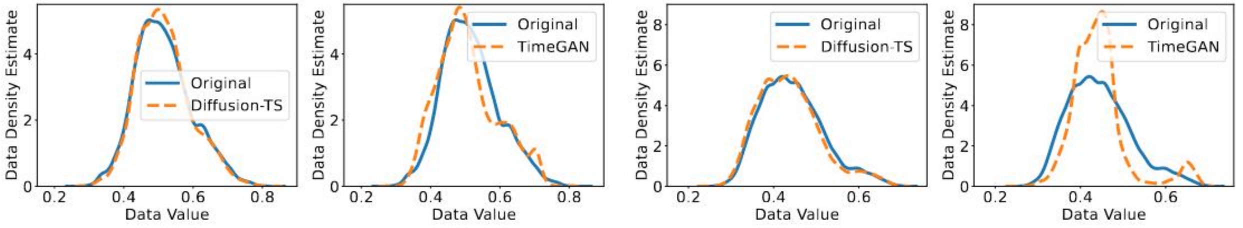
(a) ETTh (b) Energy

Forecasting

(d) Visualization



t-SNE

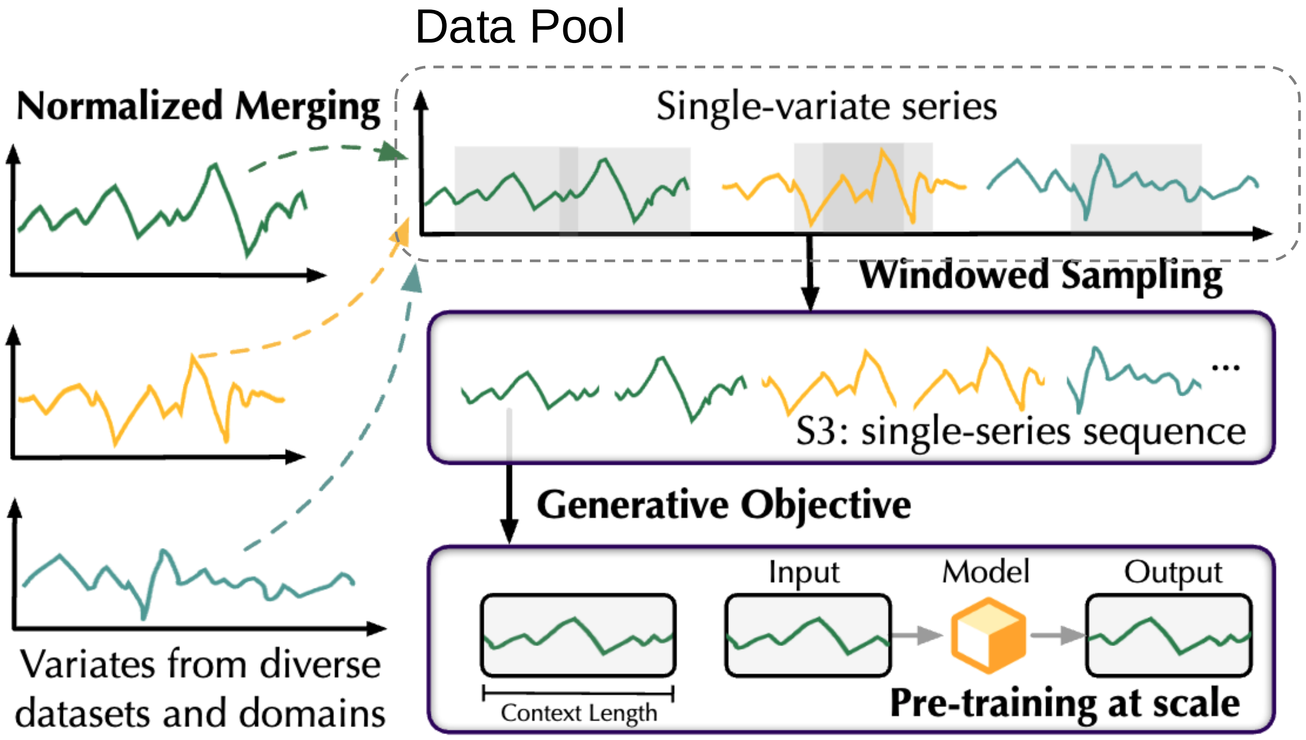


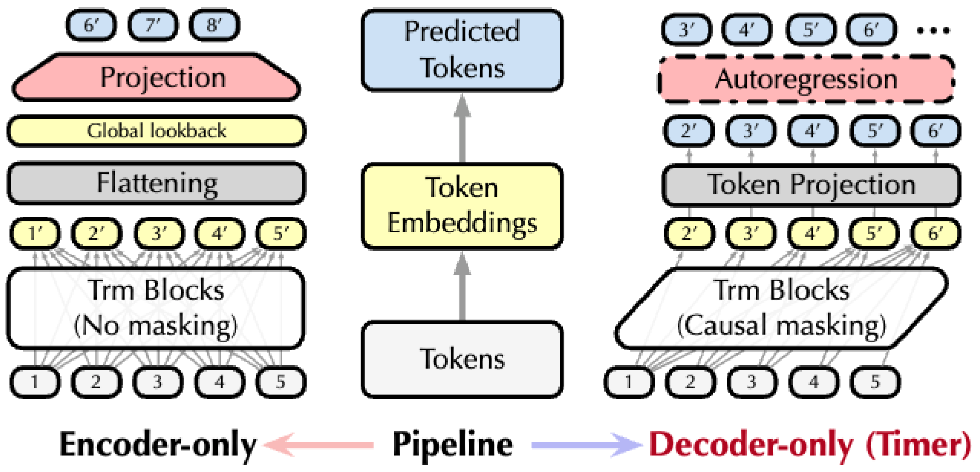
(a) ETTh (b) Energy

Data Dist.

Yuan, X., & Qiao, Y. Diffusion-TS: Interpretable Diffusion for General Time Series Generation. In ICLR'24

Pre-training Pipelines

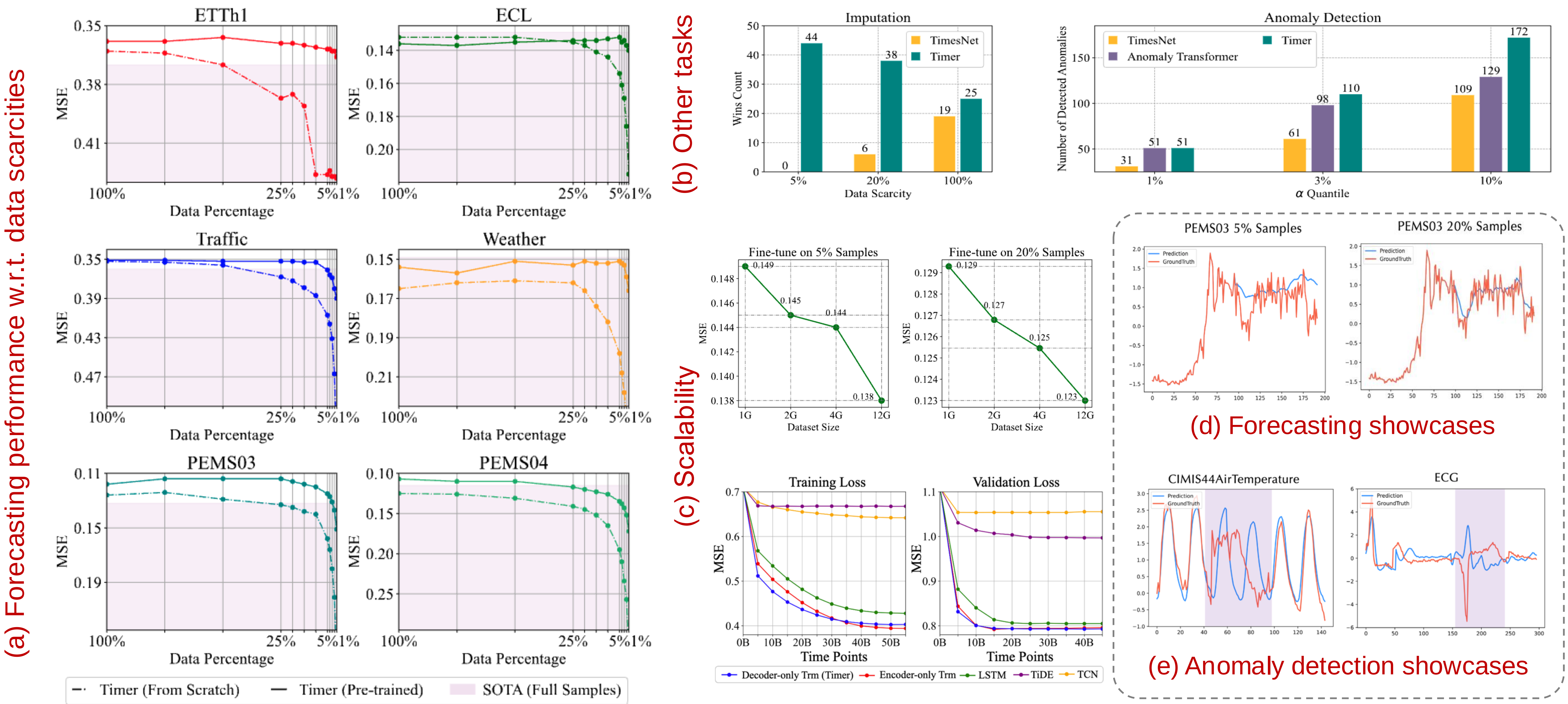




- The model observes sequences from different periods and different datasets
- Increasing the pre-training difficulty and directing more attention to the temporal variation
- S3 does not require time alignment, and single-series sequences are regarded as standard sentences of time series

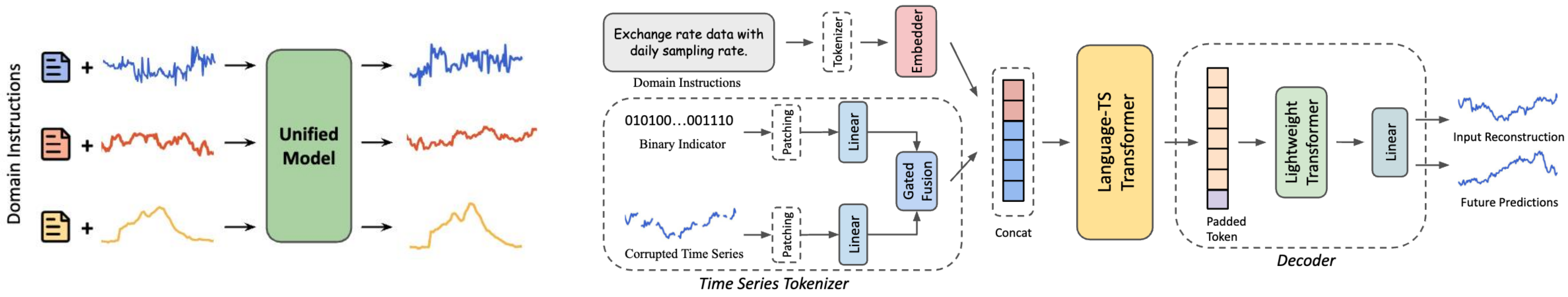
Liu, Y., Zhang, H., Li, C., Huang, X., Wang, J., & Long, M. Timer: Generative Pre-trained Transformers Are Large Time Series Models. In ICML'24

Pre-training Pipelines



Liu, Y., Zhang, H., Li, C., Huang, X., Wang, J., & Long, M. Timer: Generative Pre-trained Transformers Are Large Time Series Models. In ICML'24

Pre-training Pipelines



- Cross-domain learning + Domain Instructions
- Construct batches of data by randomly selecting instances from the data pool
- Data pool consists of training data across 8 different time series dataset

Liu, X., Hu, J., Li, Y., Diao, S., Liang, Y., Hooi, B., & Zimmermann, R. (2024, May). Unitime: A language-empowered unified model for cross-domain time series forecasting. In WWW'24

Pre-training Pipelines

Table 6: Details of the training, validation, and testing set partitions, as well as the configurations specific to different domains.

Dataset	#Training	#Validation	#Testing	Batch Size	Oversample Times	Stride	Domain Instructions
ETTm1	34,465	11,521	11,521	64	0	16	Electricity transformer A data with fifteen minutes sample rate.
ETTm2	34,465	11,521	11,521	64	0	16	Electricity transformer B data with fifteen minutes sample rate.
ETTh1	8,545	2,881	2,881	32	0	16	Electricity transformer A data with one hour sample rate.
ETTh2	8,545	2,881	2,881	32	0	16	Electricity transformer B data with one hour sample rate.
Electricity	18,317	2,633	5,261	24	0	16	Power consumption data with hourly sample rate.
Weather	36,792	5,271	10,540	64	0	16	Meteorological indicator data with ten minutes sample rate.
Exchange	5,120	665	1,422	24	0	16	Exchange rate data with one day sample rate.
Illness	617	74	170	16	12	4	Patient number data with one week sample rate.

Table 7: Variants of domain instructions.

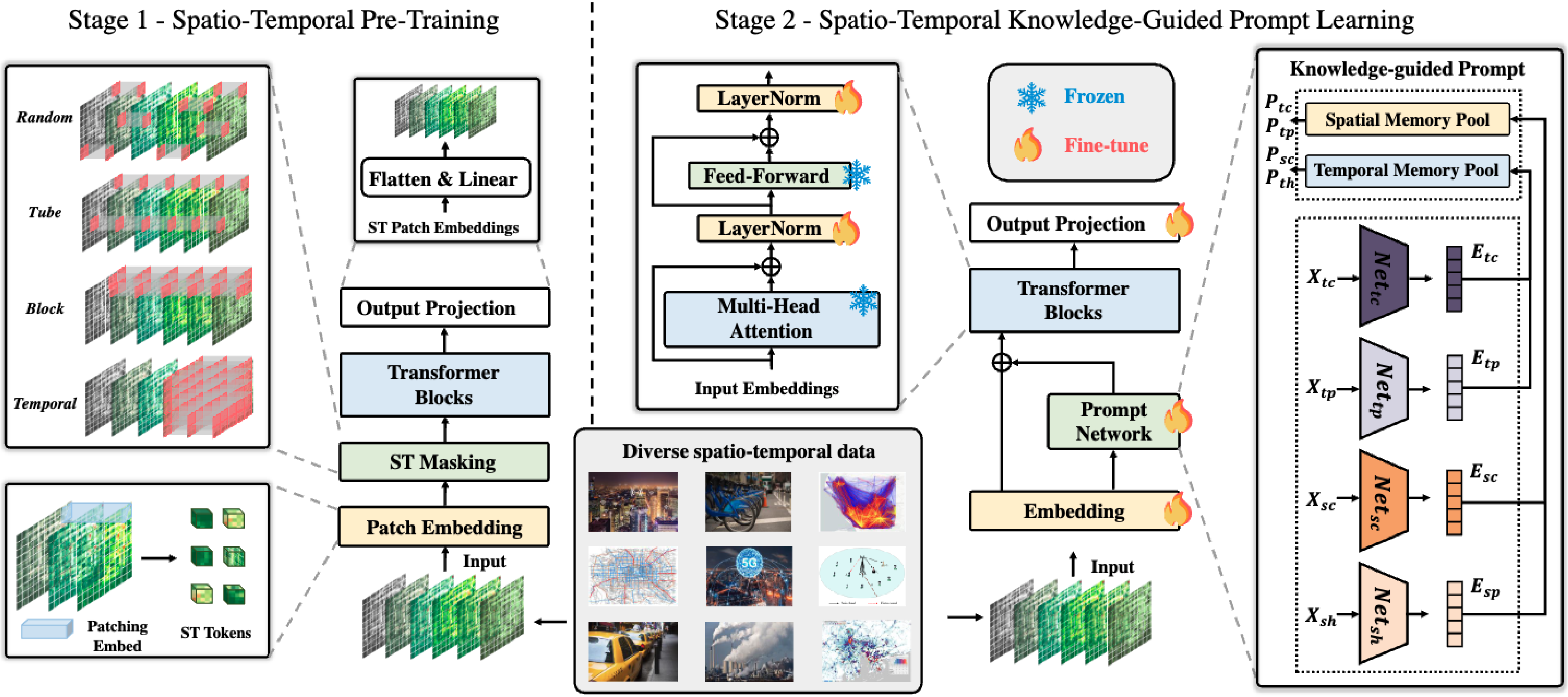
Variants	Prompts for ChatGPT	Example 1	Example 2
Original	–	meteorological indicator data with ten minute sample rate.	exchange rate data with one day sample rate.
Short	Rephrase the following text shorter: {instruction}.	ten-minute meteorological data.	daily exchange rate data.
Expand	Rephrase the following text longer: {instruction}.	the dataset for meteorological indicators presents detailed information, with data points collected at specific ten-minute intervals, facilitating a thorough analysis of meteorological conditions and trends over time.	the dataset for exchange rates provides comprehensive information, with data points recorded at consistent one-day intervals, enabling a detailed examination of currency fluctuations and trends over time.
Detail	Rephrase the following text: {instruction}, by adding the information: {information}.	the dataset includes meteorological indicators sampled every ten minutes, collected in the year 2020, and features information on 21 meteorological indicators, including temperature and humidity.	the dataset comprises exchange rate data sampled on a daily basis, documenting the daily exchange rates of eight distinct countries spanning the period from 1990 to 2016.

Table 2: Forecasting performance comparisons. The input sequence length is set to 36 for the Illness dataset and 96 for the others. The predictive lengths are set to $\{24, 36, 48, 60\}$ for Illness, and $\{96, 192, 336, 720\}$ for others. Avg is averaged over all predictive lengths. Note that we bold the best performance among models trained across datasets, which is on the left-hand side of the two vertical lines, and we bold and underline the best performance for the entire row.

		Models Trained Across Datasets						*Models Trained on Each Dataset*															
Method		UniTime		GPT4TS†		PatchTST†		GPT4TS*		PatchTST*		TimesNet		DLinear		NSformer		FEDformer		Autoformer		Informer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	96	**0.322**	**0.363**	0.509	0.463	0.927	0.604	0.335	0.369	0.344	0.373	0.338	0.375	0.345	0.372	0.386	0.398	0.379	0.419	0.505	0.475	0.672	0.571
	192	**0.366**	**0.387**	0.537	0.476	0.964	0.620	0.374	**0.385**	0.367	0.386	0.374	0.387	0.380	0.389	0.459	0.444	0.426	0.441	0.553	0.496	0.795	0.669
	336	**0.398**	**0.407**	0.564	0.488	1.041	0.656	0.407	**0.406**	**0.392**	0.407	0.410	0.411	0.413	0.413	0.495	0.464	0.445	0.459	0.621	0.537	1.212	0.871
	720	**0.454**	**0.440**	0.592	0.504	0.950	0.636	0.469	0.442	0.464	0.442	0.478	0.450	0.474	0.453	0.585	0.516	0.543	0.490	0.671	0.561	1.166	0.823
	Avg	**0.385**	**0.399**	0.551	0.483	0.971	0.629	0.396	0.401	0.392	0.402	0.400	0.406	0.403	0.407	0.481	0.456	0.448	0.452	0.588	0.517	0.961	0.734
ETTm2	96	**0.183**	**0.266**	0.229	0.304	0.240	0.318	0.190	0.275	**0.177**	**0.260**	0.187	0.267	0.193	0.292	0.192	0.274	0.203	0.287	0.255	0.339	0.365	0.453
	192	**0.251**	**0.310**	0.287	0.338	0.301	0.352	0.253	0.313	**0.246**	**0.305**	0.249	0.309	0.284	0.362	0.280	0.339	0.269	0.328	0.281	0.340	0.533	0.563
	336	**0.319**	**0.351**	0.337	0.367	0.367	0.391	0.321	0.360	**0.305**	**0.343**	0.321	0.351	0.369	0.427	0.334	0.361	0.325	0.366	0.339	0.372	1.363	0.887
	720	**0.420**	**0.410**	0.430	0.416	0.451	0.432	0.411	0.406	0.410	0.405	**0.408**	**0.403**	0.554	0.522	0.417	0.413	0.421	0.415	0.433	0.432	3.379	1.338
	Avg	**0.293**	**0.334**	0.321	0.356	0.340	0.373	0.294	0.339	**0.285**	**0.328**	0.291	0.333	0.350	0.401	0.306	0.347	0.305	0.349	0.327	0.371	1.410	0.810
ETTh1	96	**0.397**	0.418	0.449	0.424	0.409	**0.403**	0.398	0.424	0.404	0.413	0.384	0.402	0.386	**0.400**	0.513	0.491	**0.376**	0.419	0.449	0.459	0.865	0.713
	192	**0.434**	**0.439**	0.503	0.453	0.467	0.444	0.449	**0.427**	0.454	0.440	0.436	0.429	0.437	0.432	0.534	0.504	**0.420**	0.448	0.500	0.482	1.008	0.792
	336	**0.468**	**0.457**	0.540	0.477	0.509	0.472	0.492	0.466	0.497	0.462	0.491	0.469	0.481	0.459	0.588	0.535	**0.459**	0.465	0.521	0.496	1.107	0.809
	720	**0.469**	**0.477**	0.515	0.489	0.503	0.485	0.487	0.483	0.496	0.481	0.521	0.500	0.519	0.516	0.643	0.616	0.506	0.507	0.514	0.512	1.181	0.865
	Avg	**0.442**	**0.448**	0.502	0.461	0.472	0.451	0.457	0.450	0.463	0.449	0.458	0.450	0.456	0.452	0.570	0.537	**0.440**	0.460	0.496	0.487	1.040	0.795
ETTh2	96	**0.296**	**0.345**	0.303	0.349	0.314	0.361	0.312	0.360	0.312	0.358	0.340	0.374	0.333	0.387	0.476	0.458	0.358	0.397	0.346	0.388	3.755	1.525
	192	**0.374**	**0.394**	0.391	0.399	0.407	0.411	0.387	0.405	0.397	0.408	0.402	0.414	0.477	0.476	0.512	0.493	0.429	0.439	0.456	0.452	5.602	1.931
	336	**0.415**	**0.427**	0.422	0.428	0.437	0.443	0.424	0.437	0.435	0.440	0.452	0.452	0.594	0.541	0.552	0.551	0.496	0.487	0.482	0.486	4.721	1.835
	720	**0.425**	**0.444**	0.429	0.449	0.434	0.448	0.433	0.453	0.436	0.449	0.462	0.468	0.831	0.657	0.562	0.560	0.463	0.474	0.515	0.511	3.647	1.625
	Avg	**0.378**	**0.403**	0.386	0.406	0.398	0.416	0.389	0.414	0.395	0.414	0.414	0.427	0.559	0.515	0.526	0.516	0.437	0.449	0.450	0.459	4.431	1.729
Electricity	96	**0.196**	**0.287**	0.232	0.321	0.198	0.290	0.197	0.290	0.186	**0.269**	**0.168**	0.272	0.197	0.282	0.169	0.273	0.193	0.308	0.201	0.317	0.274	0.368
	192	**0.199**	**0.291**	0.234	0.325	0.202	0.293	0.201	0.292	0.190	**0.273**	0.184	0.289	0.196	0.285	**0.182**	0.286	0.201	0.315	0.222	0.334	0.296	0.386
	336	**0.214**	**0.305**	0.249	0.338	0.223	0.318	0.217	0.309	0.206	**0.290**	**0.198**	0.300	0.209	0.301	0.200	0.304	0.214	0.329	0.231	0.338	0.300	0.394
	720	**0.254**	**0.335**	0.289	0.366	0.259	0.341	0.253	0.339	0.247	0.322	**0.220**	**0.320**	0.245	0.333	0.222	0.321	0.246	0.355	0.254	0.361	0.373	0.439
	Avg	**0.216**	**0.305**	0.251	0.338	0.221	0.311	0.217	0.308	0.207	**0.289**	**0.192**	0.295	0.212	0.300	0.193	0.296	0.214	0.327	0.227	0.338	0.311	0.397
Weather	96	**0.171**	**0.214**	0.212	0.251	0.213	0.260	0.203	0.244	0.177	0.218	0.172	0.220	0.196	0.255	0.173	0.223	0.217	0.296	0.266	0.336	0.300	0.384
	192	**0.217**	**0.254**	0.261	0.288	0.269	0.300	0.247	0.277	0.222	0.259	0.219	0.261	0.237	0.296	0.245	0.285	0.276	0.336	0.307	0.367	0.598	0.544
	336	**0.274**	**0.293**	0.313	0.324	0.330	0.341	0.297	0.311	0.277	0.297	0.280	0.306	0.283	0.335	0.321	0.338	0.339	0.380	0.359	0.395	0.578	0.523
	720	**0.351**	**0.343**	0.386	0.372	0.404	0.389	0.368	0.356	0.352	0.347	0.365	0.359	**0.345**	0.381	0.414	0.410	0.403	0.428	0.419	0.428	1.059	0.741
	Avg	**0.253**	**0.276**	0.293	0.309	0.304	0.323	0.279	0.297	0.257	0.280	0.259	0.287	0.265	0.317	0.288	0.314	0.309	0.360	0.338	0.382	0.634	0.548
Exchange	96	**0.086**	**0.209**	0.142	0.261	0.137	0.260	0.091	0.212	0.109	0.236	0.107	0.234	0.088	0.218	0.111	0.237	0.148	0.278	0.197	0.323	0.847	0.752
	192	**0.174**	**0.299**	0.224	0.339	0.222	0.341	0.183	0.304	0.205	0.327	0.226	0.344	0.176	0.315	0.219	0.335	0.271	0.380	0.300	0.369	1.204	0.895
	336	**0.319**	**0.408**	0.377	0.448	0.372	0.447	0.328	0.417	0.356	0.436	0.367	0.448	**0.313**	0.427	0.421	0.476	0.460	0.500	0.509	0.524	1.672	1.036
	720	**0.875**	**0.701**	0.939	0.736	0.912	0.727	0.880	0.704	0.888	0.716	0.964	0.746	**0.839**	**0.695**	1.092	0.769	1.195	0.841	1.447	0.941	2.478	1.310
	Avg	**0.364**	**0.404**	0.421	0.446	0.411	0.444	0.371	0.409	0.390	0.429	0.416	0.443	**0.354**	0.414	0.461	0.454	0.519	0.500	0.613	0.539	1.550	0.998
Illness	24	**2.460**	**0.954**	3.322	1.278	4.289	1.485	2.732	1.100	2.335	0.989	2.317	**0.934**	2.398	1.040	**2.294**	0.945	3.228	1.260	3.483	1.287	5.764	1.677
	36	**1.998**	**0.912**	3.696	1.374	4.360	1.510	2.664	1.063	2.561	1.035	1.972	0.920	2.646	1.088	**1.825**	**0.848**	2.679	1.080	3.103	1.148	4.755	1.467
	48	**1.979**	**0.912**	3.765	1.402	4.209	1.481	2.617	1.041	2.465	1.022	2.238	0.940	2.614	1.086	2.010	**0.900**	2.622	1.078	2.669	1.085	4.763	1.469
	60	**2.109**	**0.938**	3.928	1.432	3.981	1.444	2.478	1.035	2.189	0.997	**2.027**	**0.928**	2.804	1.146	2.178	0.963	2.857	1.157	2.770	1.125	5.264	1.564
	Avg	**2.137**	**0.929**	3.678	1.372	4.210	1.480	2.623	1.060	2.388	1.011	2.139	0.931	2.616	1.090	**2.077**	**0.914**	2.847	1.144	3.006	1.161	5.137	1.544
1st Count		37		0		0		3		13		10		6		7		4		0		0	

Liu, X., Hu, J., Li, Y., Diao, S., Liang, Y., Hooi, B., & Zimmermann, R. (2024, May). Unitime: A language-empowered unified model for cross-domain time series forecasting. In WWW'24

Pre-training Pipelines



- Masking reconstruction
- Prompt learning enhances generalization ability

Yuan, Y., Ding, J., Feng, J., Jin, D., & Li, Y. (2024). UniST: A Prompt-Empowered Universal Model for Urban Spatio-Temporal Prediction. In KDD'24.

Pre-training Pipelines

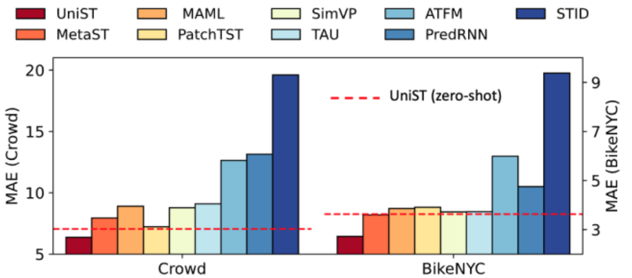
(a) Short-term prediction

Model	TaxiBJ		Crowd		Cellular		BikeNYC		TrafficJN		TDrive		TrafficSH	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
HA	53.77	29.82	17.80	6.79	72.94	27.57	11.41	3.43	1.38	0.690	150.2	74.5	1.24	0.771
ARIMA	56.70	39.53	21.87	10.23	81.31	40.22	12.37	3.86	1.20	0.651	211.3	108.5	1.17	0.769
STResNet	45.17	30.87	5.355	3.382	24.30	14.32	8.20	4.98	0.964	0.556	220.1	117.4	1.00	0.723
ACFM	37.77	21.59	4.17	2.34	22.79	12.00	3.93	1.67	0.920	0.559	98.1	51.9	0.833	0.566
STID	27.36	14.01	3.85	1.63	18.77	8.24	4.06	1.54	0.880	0.495	47.4	23.3	0.742	0.469
STNorm	29.37	15.71	4.44	2.09	19.77	8.19	4.45	1.66	0.961	0.532	54.3	47.9	0.871	0.579
STGSP	45.04	28.28	7.93	4.56	39.99	21.40	5.00	1.69	0.882	0.490	94.6	47.8	1.02	0.749
MC-STL	29.14	15.83	4.75	2.39	21.22	10.26	4.08	2.05	1.19	0.833	54.2	28.1	1.00	0.720
PromptST	27.44	14.54	3.52	1.54	15.74	7.20	4.36	1.57	0.953	0.490	47.5	22.8	0.811	0.523
MAU	38.14	20.13	4.94	2.35	39.09	18.73	5.22	2.06	1.28	0.697	48.8	22.1	1.37	0.991
PredRNN	27.50	14.29	5.13	2.36	24.15	10.44	5.00	1.74	0.852	0.463	54.9	25.2	0.748	0.469
MIM	28.62	14.77	5.66	2.27	21.38	9.37	4.40	1.62	1.17	0.650	51.4	22.7	0.760	0.505
SimVP	32.66	17.67	3.91	1.96	16.48	8.23	4.11	1.67	0.969	0.556	46.8	22.9	0.814	0.569
TAU	33.90	19.37	4.09	2.11	17.94	8.91	4.30	1.83	0.993	0.566	51.6	28.1	0.820	0.557
PatchTST	42.74	22.23	10.25	3.62	43.40	15.74	5.27	1.65	1.25	0.616	106.4	51.3	1.10	0.663
iTransformer	36.97	19.14	9.40	3.40	37.01	13.93	7.74	2.53	1.11	0.570	86.3	42.6	1.04	0.655
PatchTST(one-for-all)	43.66	23.16	13.51	5.00	56.80	20.56	9.97	3.05	1.30	0.645	127.0	59.26	1.13	0.679
UniST(one-for-all)	**26.84**	**13.95**	**3.00**	**1.38**	**14.29**	**6.50**	**3.50**	**1.27**	**0.843**	**0.430**	**44.97**	**19.67**	**0.665**	**0.405**

(b) Long-term prediction

Model	TaxiNYC		Crowd		BikeNYC	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
HA	61.03	21.33	19.57	8.49	11.00	3.66
ARIMA	68.0	28.66	21.34	8.93	11.59	3.98
STResNet	29.54	14.46	8.75	5.58	7.15	3.87
ACFM	32.91	13.72	6.16	3.35	4.56	1.86
STID	24.74	11.01	4.91	2.63	4.78	2.24
STNorm	31.81	11.99	9.62	4.30	6.45	2.18
STGSP	28.65	10.38	17.03	8.21	4.71	1.54
MC-STL	29.29	17.36	9.01	6.32	4.97	2.61
MAU	26.28	9.07	20.13	8.49	6.18	2.13
PredRNN	21.17	7.31	19.70	10.66	5.86	1.97
MIM	63.36	29.83	15.70	8.81	7.58	2.81
SimVP	20.18	9.78	5.50	3.13	4.10	1.71
TAU	24.97	10.93	5.31	2.81	3.89	1.73
PatchTST	30.64	17.49	5.25	2.83	5.27	1.65
iTransformer	33.81	11.48	6.94	2.63	6.00	2.02
PatchTST(one-for-all)	34.50	10.63	6.39	2.92	6.02	1.83
UniST (one-for-all)	**19.83**	**6.71**	**4.25**	**2.26**	**3.56**	**1.31**

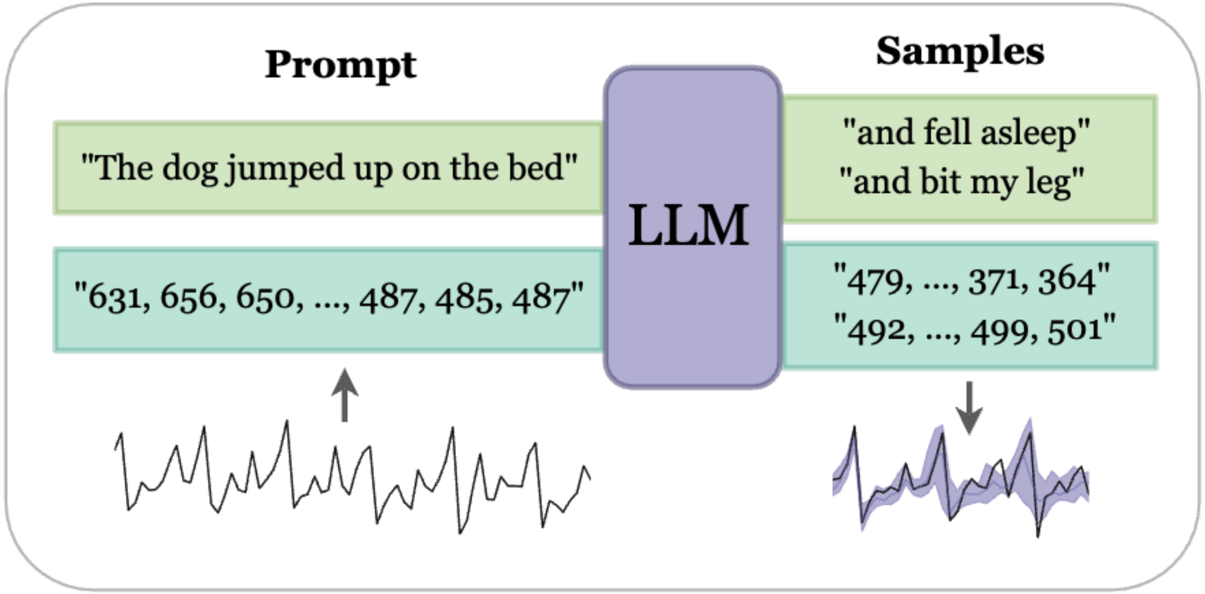
(c) Zero/few-shot performance

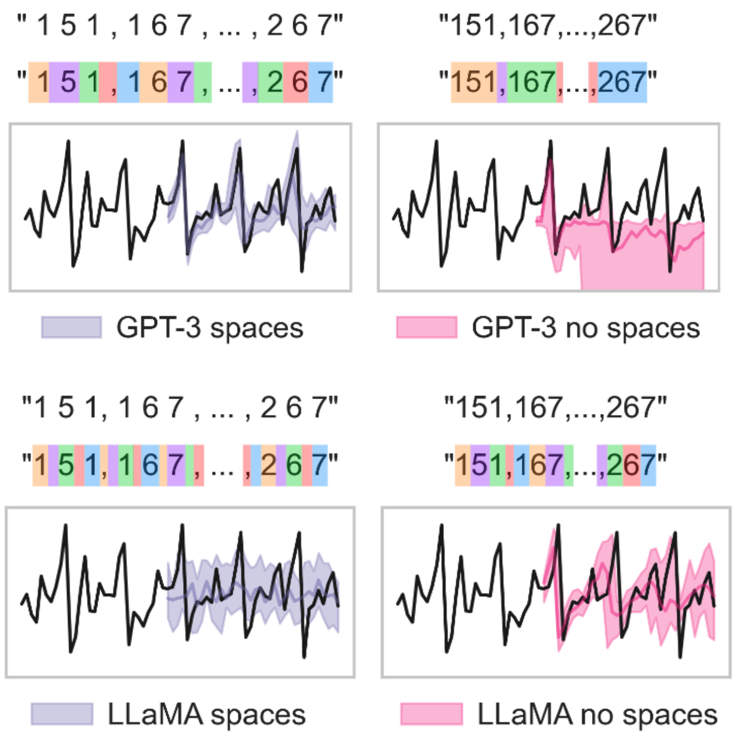


Performance on Few-shot (1%) and Zero-shot Scenarios

Yuan, Y., Ding, J., Feng, J., Jin, D., & Li, Y. (2024). UniST: A Prompt-Empowered Universal Model for Urban Spatio-Temporal Prediction. In KDD'24.

Adaptation

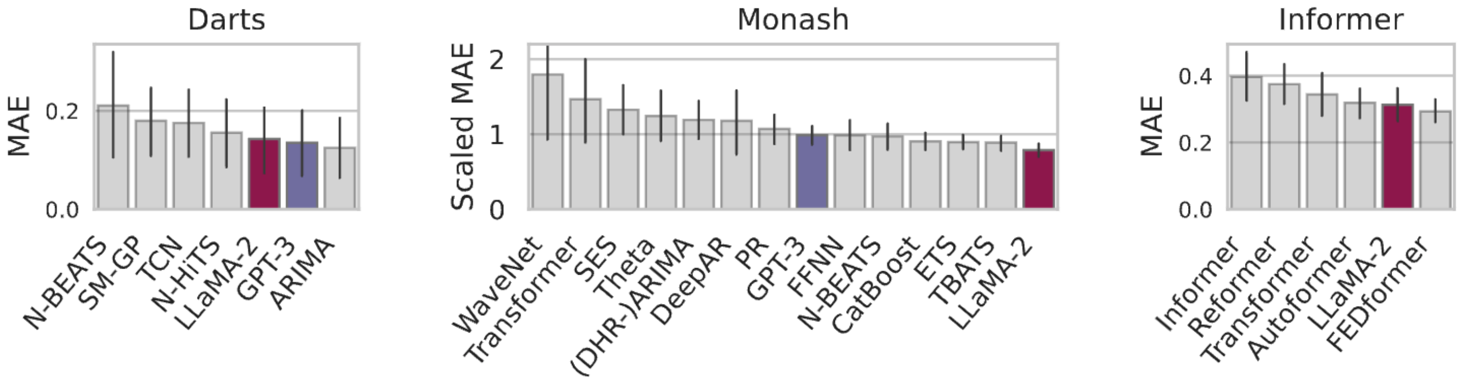




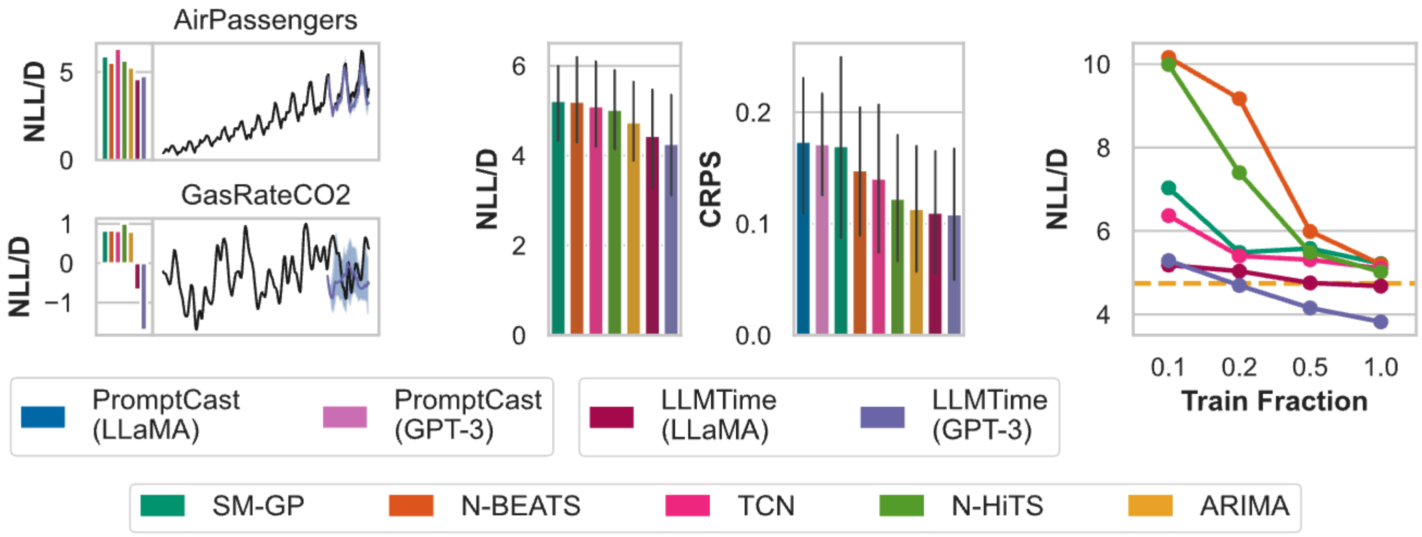
Gruver, N., Finzi, M., Qiu, S., & Wilson, A. G. (2024). Large language models are zero-shot time series forecasters. In NeurIPS'23

Adaptation

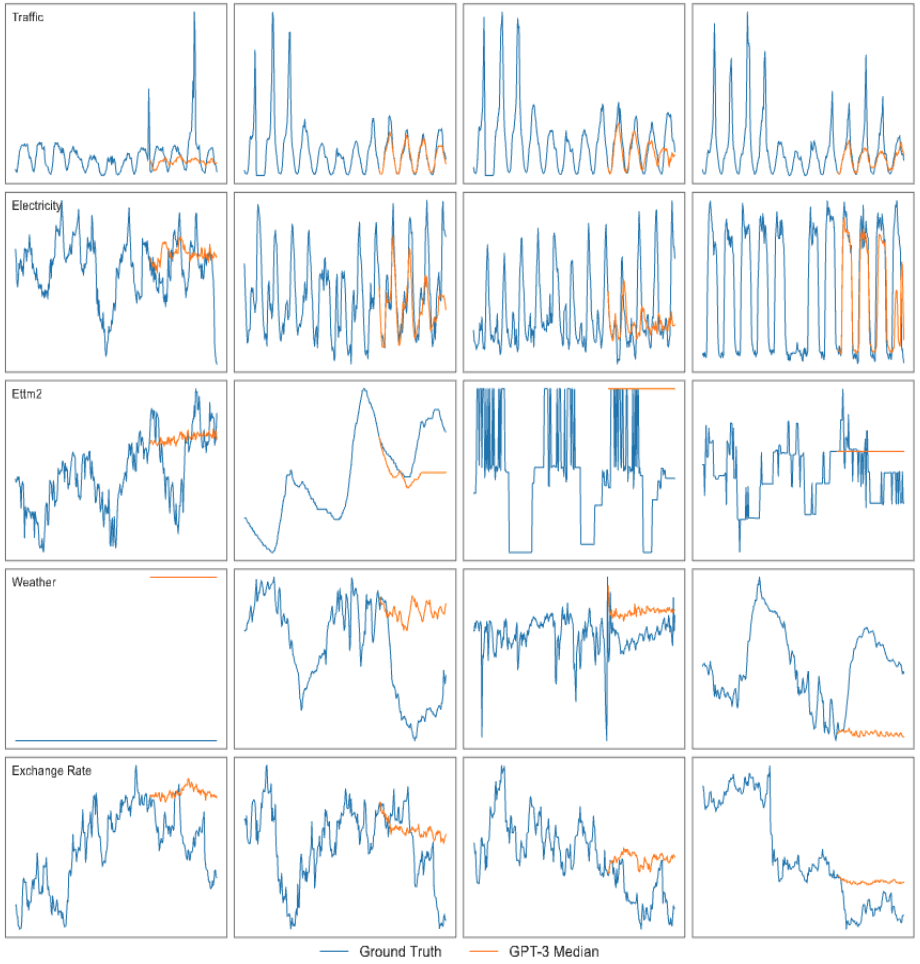
(a) Forecasting performance



(b) Left & Middle: prob. forecasting; Right: sample efficiency

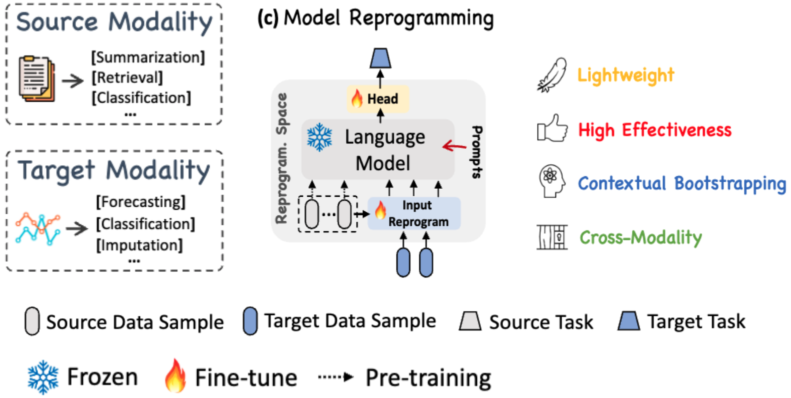


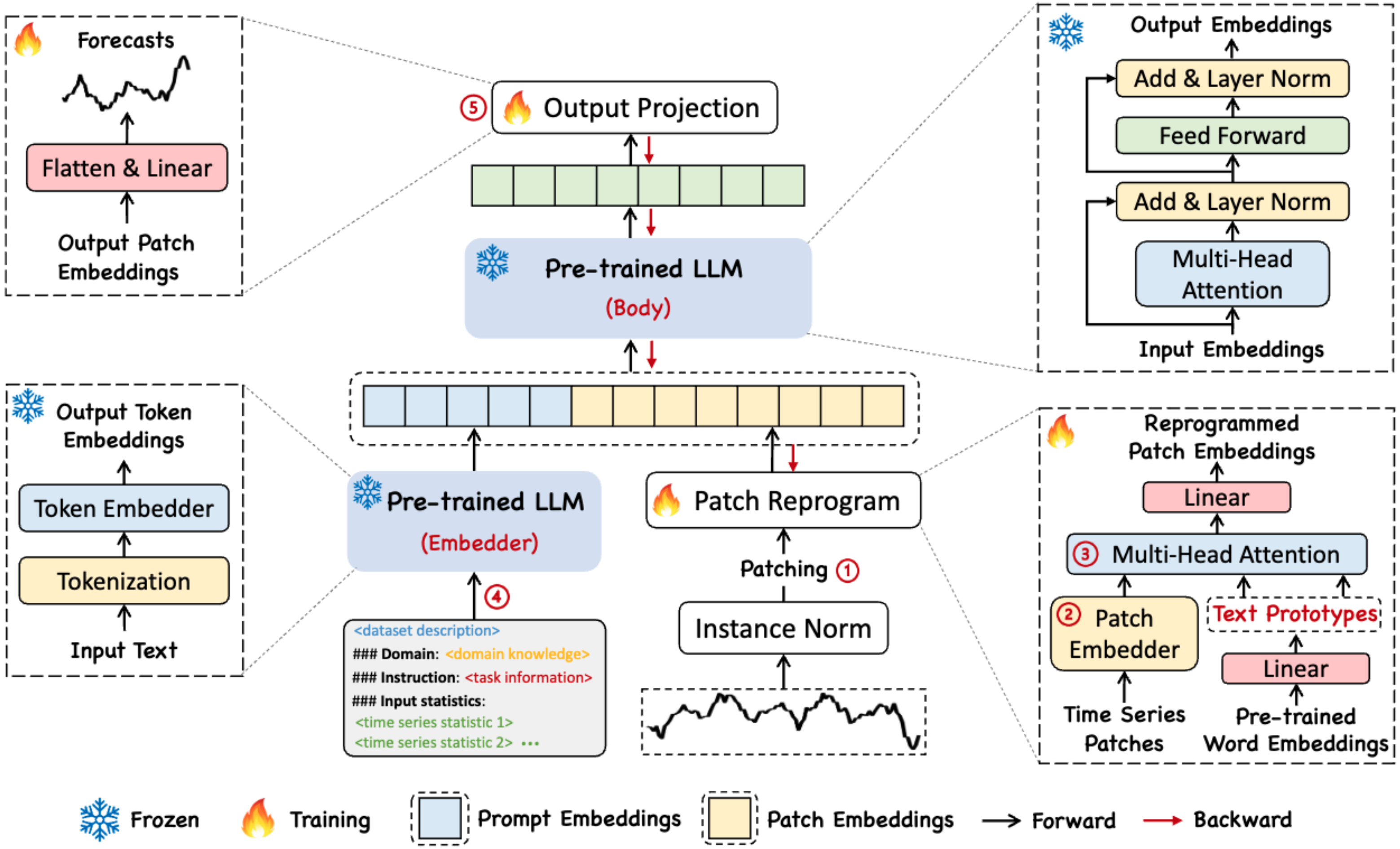
(c) Visualization



Gruver, N., Finzi, M., Qiu, S., & Wilson, A. G. (2024). Large language models are zero-shot time series forecasters. In NeurIPS'23

Adaptation





Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J., Shi, X., ... & Wen, Q. (2024, May). Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. In ICLR'24

Adaptation

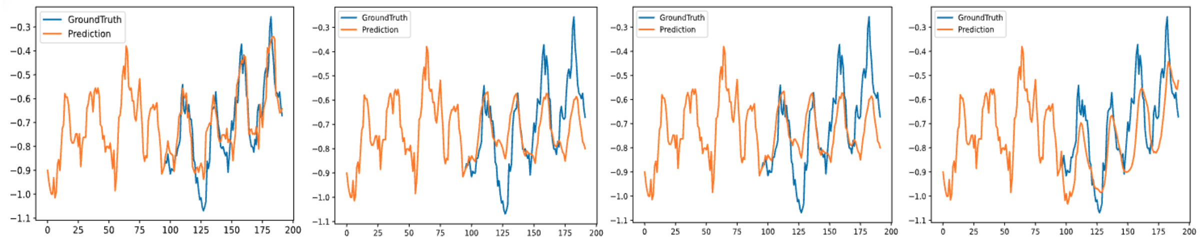
Table 3: Few-shot learning on 10% training data. We use the same protocol in Tab. 1. All results are averaged from four different forecasting horizons: $H \in \{96, 192, 336, 720\}$. Our full results are in Appendix E.

Methods	TIME-LLM (Ours)		GPT4TS (2023a)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)		FEDformer (2022)		Autoformer (2021)		Stationary (2022)		ETSformer (2022)		LightTS (2022a)		Informer (2021)		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	MSE	MAE
$ETTh1$	0.556	0.522	0.590	0.525	0.691	0.600	0.633	0.542	0.869	0.628	0.639	0.561	0.702	0.596	0.915	0.639	1.180	0.834	1.375	0.877	1.199	0.809	1.249	0.833
$ETTh2$	0.370	0.394	0.397	0.421	0.605	0.538	0.415	0.431	0.479	0.465	0.466	0.475	0.488	0.499	0.462	0.455	0.894	0.713	2.655	1.160	3.872	1.513	3.485	1.486
$ETTm1$	0.404	0.427	0.464	0.441	0.411	0.429	0.501	0.466	0.677	0.537	0.722	0.605	0.802	0.628	0.797	0.578	0.980	0.714	0.971	0.705	1.192	0.821	1.426	0.856
$ETTm2$	0.277	0.323	0.293	0.335	0.316	0.368	0.296	0.343	0.320	0.353	0.463	0.488	1.342	0.930	0.332	0.366	0.447	0.487	0.987	0.756	3.370	1.440	3.978	1.587
$Weather$	0.234	0.273	0.238	0.275	0.241	0.283	0.242	0.279	0.279	0.301	0.284	0.324	0.300	0.342	0.318	0.323	0.318	0.360	0.289	0.322	0.597	0.495	0.546	0.469
ECL	0.175	0.270	0.176	0.269	0.180	0.280	0.180	0.273	0.323	0.392	0.346	0.427	0.431	0.478	0.444	0.480	0.660	0.617	0.441	0.489	1.195	0.891	0.965	0.768
$Traffic$	0.429	0.306	0.440	0.310	0.447	0.313	0.430	0.305	0.951	0.535	0.663	0.425	0.749	0.446	1.453	0.815	1.914	0.936	1.248	0.684	1.534	0.811	1.551	0.821
1^{st}Count	7		1		0		1		0		0		0		0		0		0		0		0	

Table 5: Zero-shot learning results. **Red**: the best, Blue: the second best. Appendix E shows our detailed results.

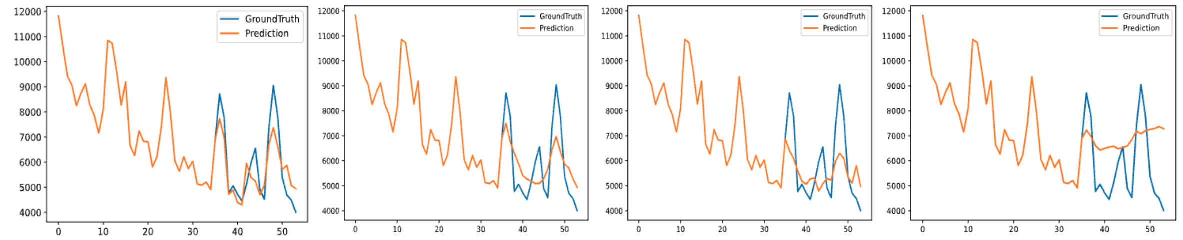
Methods	TIME-LLM (Ours)		GPT4TS (2023a)		LLMTime (2023)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh1 \rightarrow ETTh2$	0.353	0.387	0.406	0.422	0.992	0.708	0.493	0.488	0.380	0.405	0.421	0.431
$ETTh1 \rightarrow ETTm2$	0.273	0.340	0.325	0.363	1.867	0.869	0.415	0.452	0.314	0.360	0.327	0.361
$ETTh2 \rightarrow ETTh1$	0.479	0.474	0.757	0.578	1.961	0.981	0.703	0.574	0.565	0.513	0.865	0.621
$ETTh2 \rightarrow ETTm2$	0.272	0.341	0.335	0.370	1.867	0.869	0.328	0.386	0.325	0.365	0.342	0.376
$ETTm1 \rightarrow ETTh2$	0.381	0.412	0.433	0.439	0.992	0.708	0.464	0.475	0.439	0.438	0.457	0.454
$ETTm1 \rightarrow ETTm2$	0.268	0.320	0.313	0.348	1.867	0.869	0.335	0.389	0.296	0.334	0.322	0.354
$ETTm2 \rightarrow ETTh2$	0.354	0.400	0.435	0.443	0.992	0.708	0.455	0.471	0.409	0.425	0.435	0.443
$ETTm2 \rightarrow ETTm1$	0.414	0.438	0.769	0.567	1.933	0.984	0.649	0.537	0.568	0.492	0.769	0.567

(i) Long-term forecasting



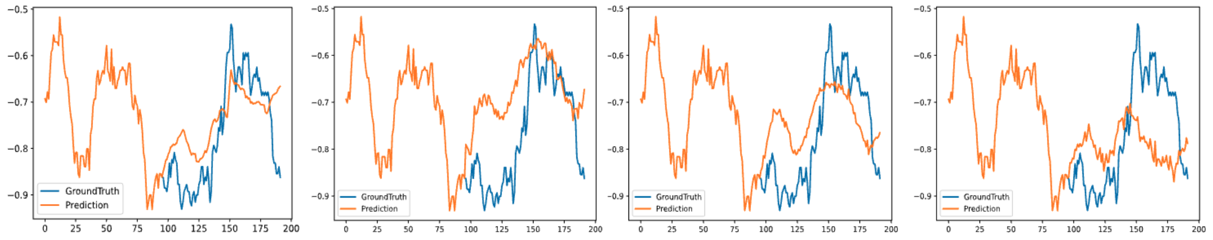
(a)Time-LLM (b)GPT4TS (c)PatchTST (d) Autoformer

(ii) Short-term forecasting



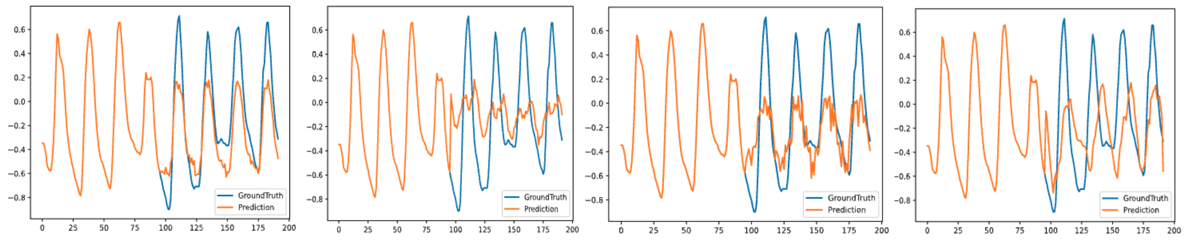
(a)Time-LLM (b)GPT4TS (c) TimesNet (d) Autoformer

(iii) Few-shot forecasting



(a)Time-LLM (b)GPT4TS (c)PatchTST (d) Autoformer

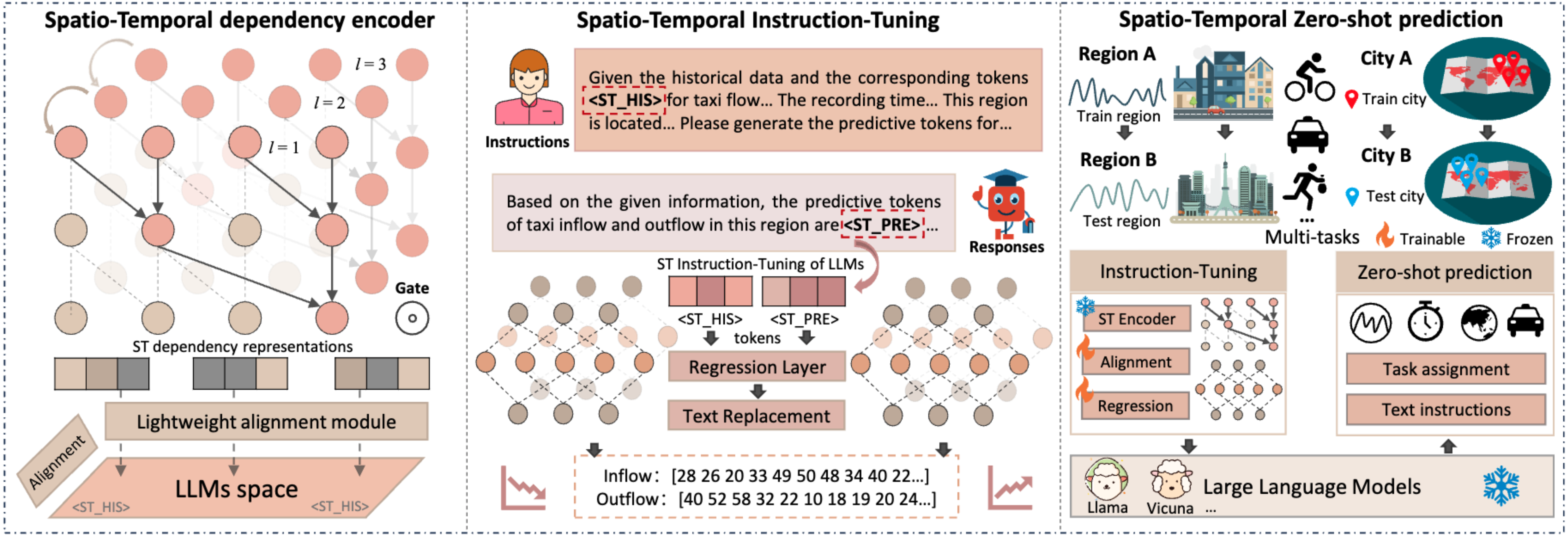
(iv) Zero-shot forecasting



(a)Time-LLM (b)GPT4TS (c)PatchTST (d) Autoformer

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J., Shi, X., ... & Wen, Q. (2024, May). Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. In ICLR'24

Adaptation



- Enabling LLMs to comprehend spatial-temporal dependencies in data for downstream urban tasks
- Spatio-temporal encoder + instruction-tuning = UrbanGPT

Li, Z., Xia, L., Tang, J., Xu, Y., Shi, L., Xia, L., ... & Huang, C. (2024). Urbangpt: Spatio-temporal large language models. In KDD'24

Adaptation

(a) Forecasting performance

Table 1: Our model's performance in zero-shot prediction is evaluated on three diverse datasets: NYC-taxi, NYC-bike, and NYC-crime, providing a comprehensive assessment of its predictive capabilities in unseen situations.

Model	Dataset	NYC-taxi				NYC-bike				NYC-crime			
	Type	Inflow		Outflow		Inflow		Outflow		Burglary		Robbery	
	Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	Macro-F1	Recall	Macro-F1	Recall
AGCRN		10.86	26.51	13.15	36.45	3.41	7.98	3.42	8.08	0.48	0.00	0.49	0.01
ASTGCN		9.75	24.12	12.42	33.28	5.58	11.58	5.78	12.29	0.49	0.01	0.55	0.09
GWN		10.73	26.50	9.67	26.74	3.32	8.17	3.07	7.52	0.48	0.00	0.52	0.04
MTGNN		10.16	25.84	12.59	35.38	3.18	7.62	3.20	7.65	0.64	0.27	0.65	0.30
STWA		11.28	28.97	13.54	38.61	4.59	10.94	4.35	10.67	0.48	0.00	0.51	0.03
STSGCN		18.97	41.38	20.07	45.79	6.85	14.98	6.54	14.77	0.48	0.00	0.48	0.00
STGCN		12.54	30.80	14.32	39.58	4.11	9.21	4.45	9.62	0.48	0.00	0.64	0.30
TGCN		10.04	25.10	10.98	30.03	2.88	6.55	2.91	6.42	0.56	0.10	0.58	0.13
DMVSTNET		11.00	28.29	10.59	29.20	3.80	9.87	3.65	9.21	0.48	0.01	0.59	0.15
ST-LSTM		16.97	34.43	18.93	44.10	7.78	15.41	6.92	17.12	0.48	0.00	0.49	0.03
UrbanGPT		**6.16**	**16.92**	**6.83**	**21.78**	**2.02**	**5.16**	**2.01**	**5.03**	**0.67**	**0.34**	**0.69**	**0.42**

(b) Showcase

Table 3: We examine the zero-shot predictions of different LLMs for bicycle flow in NYC with the provided instructions.

Instructions: Given the historical data for bike flow over 12 time steps in a specific region of New York City, the recorded bike inflows are [12 8 9 15 10 12 14 9 6 12 25 20], and the recorded bike outflows are [5 12 12 15 10 11 6 9 4 6 10 15]. The recording time of the historical data is 'January 14, 2020, 12:00, Tuesday to January 14, 2020, 17:30, Tuesday, with data points recorded at 30-minute intervals'. Here is the region information: This region is located within the Staten Island borough district and encompasses various POIs within a one-kilometer radius, covering Education Facility, Cultural Facility, Commercial, Transportation Facility categories. We now aim to predict the bike inflow and outflow for the next 12 time steps during the time period of 'January 14, 2020, 18:00, Tuesday to January 14, 2020, 23:30, Tuesday, with data points recorded at 30-minute intervals'. To improve prediction accuracy, a spatio-temporal model is utilized to encode the historical bike data as tokens <ST_Start><ST_HIS><ST_HIS><ST_End>, where the first and the second tokens correspond to the representations of bike inflow and outflow. Please conduct an analysis of the traffic patterns in this region, taking into account the provided time and regional information, and then generate the predictions (the predictive tokens for regression).

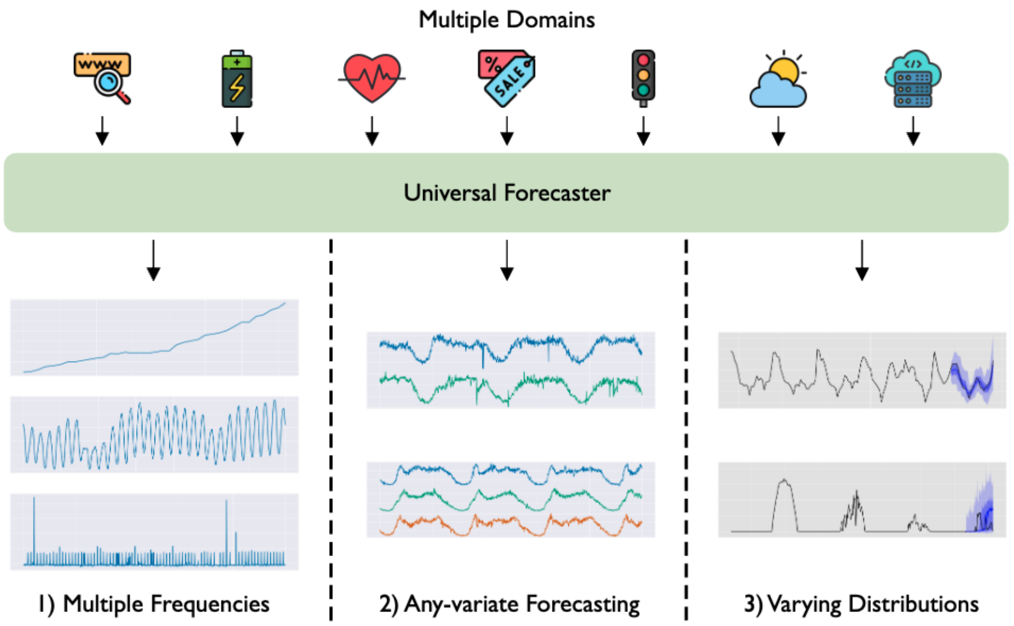
Ground true: Based on the given information, the predictions of bike inflow and outflow in this region are **[26 38 28 24 20 13 16 9 5 3 5 2]** and **[12 19 10 10 11 9 6 5 4 4 2 2]**.

UrbanGPT Responds: Based on the given information, the predictions (predicted tokens) of bike inflow and outflow in this region are <ST_PRE> **[20 21 22 21 17 12 9 7 6 5 5 5]** and <ST_PRE> **[11 10 8 7 6 6 5 4 4 3 3 2]**.

Li, Z., Xia, L., Tang, J., Xu, Y., Shi, L., Xia, L., ... & Huang, C. (2024). Urbangpt: Spatio-temporal large language models. In KDD'24

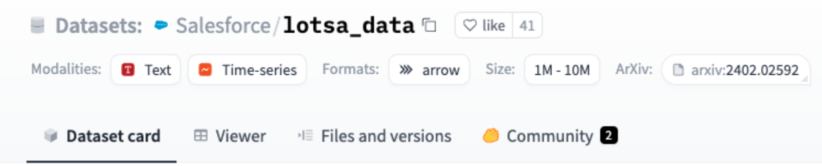
Single Modality

- **Moirai** **(ICML'24)**



	Any-variate (Zero-shot)	Probabilistic Forecasting	Flexible Distribution	Pre-training Data (Size)	Open-source
MOIRAI	✓	✓	✓	LOTSA (> 27B)	✓
TimeGPT-1	✓	✓	✗	Unknown (100B)	✗
ForecastPFN	✗	✗	-	Synthetic Data (60M)	✓
Lag-Llama	✗	✓	✗	Monash (< 1B)	✓
TimesFM	✗	✗	-	Wiki + Trends + Others (> 100B)	✓
TTM	✗	✗	-	Monash (< 1B)	✓
LLMTime	✗	✓	✓	Web-scale Text	✓

	Energy	Transport	Climate	CloudOps	Web	Sales	Nature	Econ/Fin	Healthcare
# Datasets	30	23	6	3	3	6	5	23	6
# Obs.	16,358,600,896	4,900,453,419	4,188,011,890	1,518,268,292	428,082,373	197,984,339	28,547,647	24,919,596	1,594,281
%	59.17%	17.73%	15.15%	5.49%	1.55%	0.72%	0.09%	0.10%	0.01%



Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., & Sahoo, D. Unified Training of Universal Time Series Forecasting Transformers. In ICML'24

Single Modality

- **Chronos** (arXiv'24)

Data Subset	# Datasets	# Series	Usage	Baselines
Pretraining-only	13	795,936	pretraining	–
Benchmark I	15	97,272	pretraining and in-domain evaluation	Naive, SeasonalNaive, AutoETS, AutoTheta, AutoARIMA, DeepAR, TFT, PatchTST, DLinear, WaveNet, N-BEATS, N-HiTS, GPT4TS, Lag-Llama, Moirai-1.0-R
Benchmark II	27	190,674	zero-shot evaluation	All the above, LLMTime and ForecastPFN

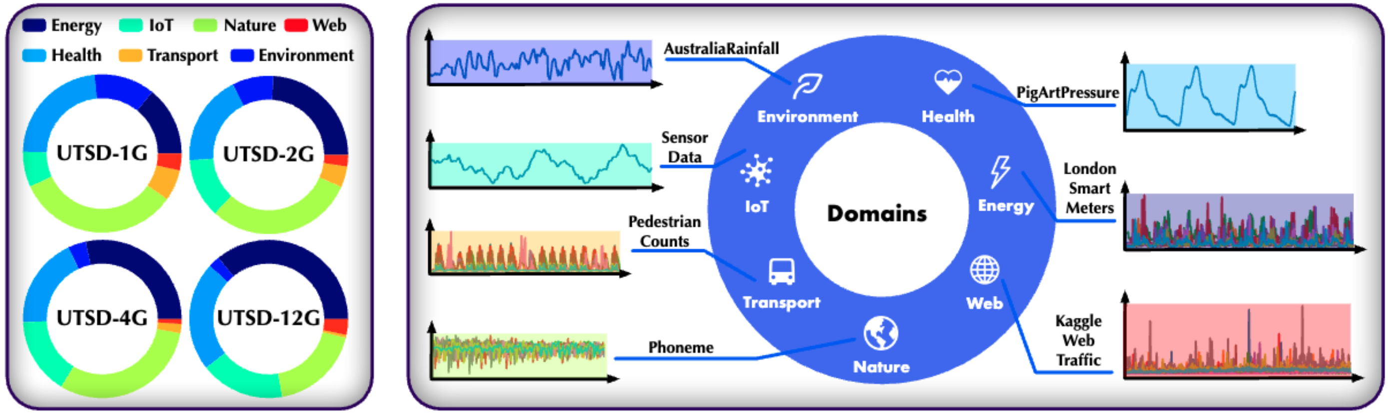
Dataset	Domain	Freq.	Num. Series	Series Length			Prediction
				min	avg	max	Length (H)
Pretraining-only							
Brazilian Cities Temperature	nature	M	12	492	757	1320	-
Mexico City Bikes	transport	1H	494	780	78313	104449	-
Solar (5 Min.)	energy	5min	5166	105120	105120	105120	-
Solar (Hourly)	energy	1H	5166	8760	8760	8760	-
Spanish Energy and Weather	energy	1H	66	35064	35064	35064	-
Taxi (Hourly)	transport	1H	2428	734	739	744	-
USHCN	nature	1D	6090	5906	38653	59283	-
Weatherbench (Daily)	nature	1D	225280	14609	14609	14610	-
Weatherbench (Hourly)	nature	1H	225280	350633	350639	350640	-
Weatherbench (Weekly)	nature	1W	225280	2087	2087	2087	-
Wiki Daily (100k)	web	1D	100000	2741	2741	2741	-
Wind Farms (Daily)	energy	1D	337	71	354	366	-
Wind Farms (Hourly)	energy	1H	337	1715	8514	8784	-

Dataset	Domain	Freq.	Num. Series	Series Length			Prediction
				min	avg	max	Length (H)
In-domain evaluation							
Electricity (15 Min.)	energy	15min	370	16032	113341	140256	24
Electricity (Hourly)	energy	1H	321	26304	26304	26304	24
Electricity (Weekly)	energy	1W	321	156	156	156	8
KDD Cup 2018	nature	1H	270	9504	10897	10920	48
London Smart Meters	energy	30min	5560	288	29951	39648	48
M4 (Daily)	various	1D	4227	107	2371	9933	14
M4 (Hourly)	various	1H	414	748	901	1008	48
M4 (Monthly)	various	1M	48000	60	234	2812	18
M4 (Weekly)	various	1W	359	93	1035	2610	13
Pedestrian Counts	transport	1H	66	576	47459	96424	48
Rideshare	transport	1H	2340	541	541	541	24
Taxi (30 Min.)	transport	30min	2428	1469	1478	1488	48
Temperature-Rain	nature	1D	32072	725	725	725	30
Uber TLC (Daily)	transport	1D	262	181	181	181	7
Uber TLC (Hourly)	transport	1H	262	4344	4344	4344	24
Zero-shot evaluation							
Australian Electricity	energy	30min	5	230736	231052	232272	48
CIF 2016	banking	1M	72	28	98	120	12
Car Parts	retail	1M	2674	51	51	51	12
Covid Deaths	healthcare	1D	266	212	212	212	30
Dominick	retail	1D	100014	201	296	399	8
ERCOT Load	energy	1H	8	154854	154854	154854	24
ETT (15 Min.)	energy	15min	14	69680	69680	69680	24
ETT (Hourly)	energy	1H	14	17420	17420	17420	24
Exchange Rate	finance	1B	8	7588	7588	7588	30
FRED-MD	economics	1M	107	728	728	728	12
Hospital	healthcare	1M	767	84	84	84	12
M1 (Monthly)	various	1M	617	48	90	150	18
M1 (Quarterly)	various	3M	203	18	48	114	8
M1 (Yearly)	various	1Y	181	15	24	58	6
M3 (Monthly)	various	1M	1428	66	117	144	18
M3 (Quarterly)	various	3M	756	24	48	72	8
M3 (Yearly)	various	1Y	645	20	28	47	6
M4 (Quarterly)	various	3M	24000	24	100	874	8
M4 (Yearly)	various	1Y	23000	19	37	841	6
M5	retail	1D	30490	124	1562	1969	28
NN5 (Daily)	finance	1D	111	791	791	791	56
NN5 (Weekly)	finance	1W	111	113	113	113	8
Tourism (Monthly)	various	1M	366	91	298	333	24
Tourism (Quarterly)	various	1Q	427	30	99	130	8
Tourism (Yearly)	various	1Y	518	11	24	47	4
Traffic	transport	1H	862	17544	17544	17544	24
Weather	nature	1D	3010	1332	14296	65981	30

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., … & Wang, Y. (2024). Chronos: Learning the language of time series. arXiv preprint arXiv:2403.07815.

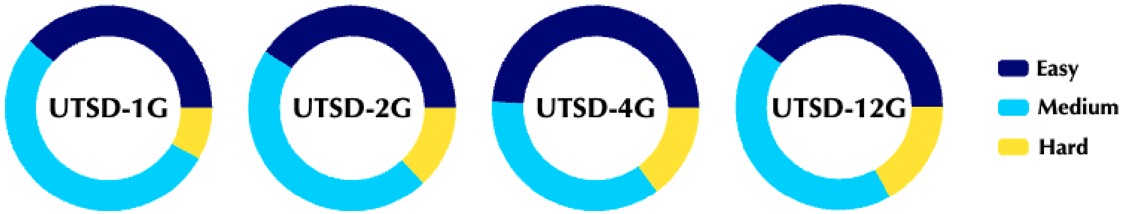
Single Modality

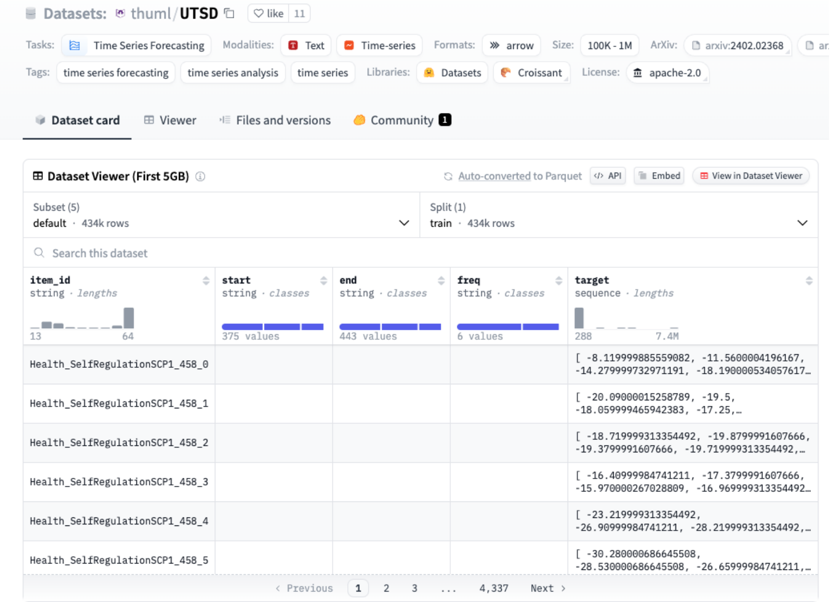
- **Timer** **(ICML'24)**



Unified Time Series Dataset

Diverse Shapes, Patterns and Domains





- Unified Time Series Dataset (UTSD) encompasses seven domains with up to 1B time points (UTSD-12G)
- Data complexity is measured by Augmented Dickey-Fuller (ADF) test (that reflects the degree of non-stationarity)

Liu, Y., Zhang, H., Li, C., Huang, X., Wang, J., & Long, M. Timer: Generative Pre-trained Transformers Are Large Time Series Models. In ICML'24

Single Modality

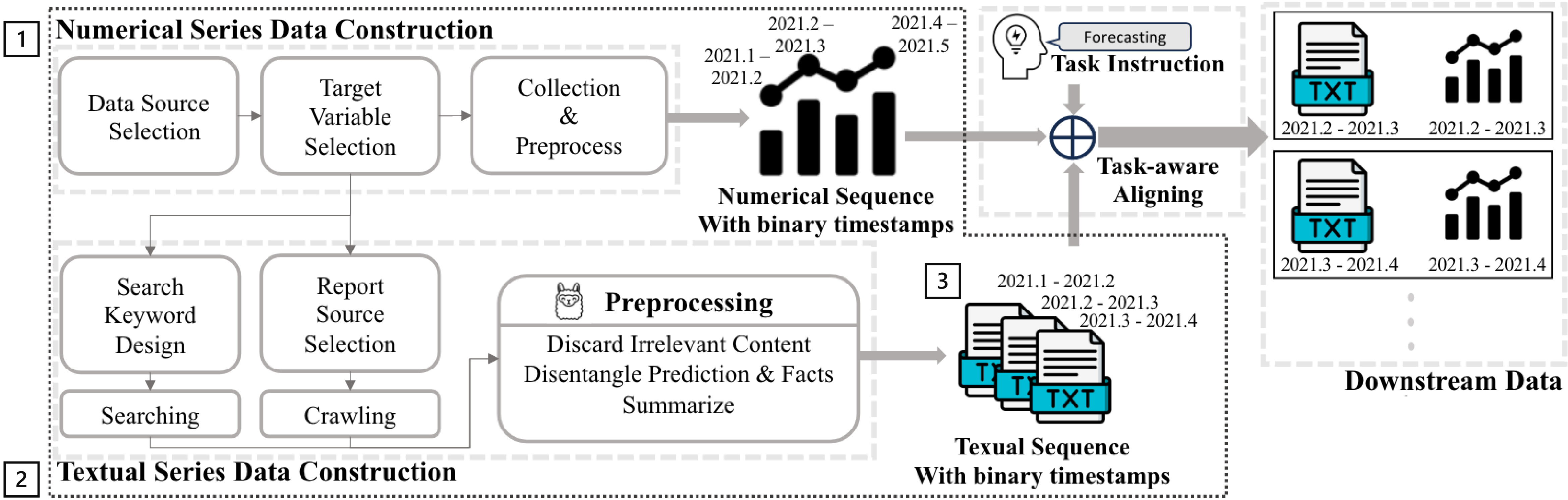
- **UniST** **(KDD'24)**

Dataset	Min value	Max value	Mean value	Standard deviation
TaxiBJ	0.0	1285	107	133
Cellular	0.0	2992532	75258	149505
TaxiNYC-1	0.0	1517	32	94
TaxiNYC-2	0.0	1283	37	102
BikeNYC-1	0.0	266	9.2	18.1
BikeNYC-2	0.0	299	4.4	14.6
TDrive	0.0	2681	123	229
Crowd	0.0	593118	21656	40825
TrafficCS	0.0	22.25	6.22	4.79
TrafficWH	0.0	22.35	6.22	4.68
TrafficCD	0.0	22.25	7.33	4.36
TrafficJN	0.0	25.04	5.72	4.71
TrafficNJ	0.0	24.82	5.38	4.73
TrafficSH	0.0	21.83	7.92	3.86
TrafficSZ	0.0	22.12	5.11	4.75
TrafficGZ	0.0	25.16	5.26	4.79
TrafficGY	0.0	28.89	5.95	7.03
TrafficTJ	0.0	25.24	6.32	5.05
TrafficHZ	0.0	29.50	3.81	4.38
TrafficZZ	0.0	23.26	6.67	4.32
TrafficBJ	0.0	22.82	6.30	4.22

Dataset	Domain	City	Temporal Duration	Temporal interval	Spatial partition
TaxiBJ	Taxi GPS	Beijing, China	20130601-20131030 20140301-20140630 20150301-20150630 20151101-20160410	Half an hour	32×32
Cellular	Cellular usage	Nanjing, China	20201111-20210531	Half an hour	16 * 20
TaxiNYC-1	Taxi OD	New York City, USA	20160101-20160229	Half an hour	16 * 12
TaxiNYC-2	Taxi OD	New York City, USA	20150101-20150301	Half an hour	20 * 10
BikeNYC-1	Bike usage	New York City, USA	20160801-20160929	One hour	16 * 8
BikeNYC-2	Bike usage	New York City, USA	20160701-20160829	Half an hour	10 * 20
TDrive	Taxi trajectory	New York City, USA	20150201-20160602	One hour	32×32
Crowd	Crowd flow	Nanjing, China	20201111-20210531	Half an hour	16 * 20
TrafficCS	Traffic speed	Changsha, China	20220305-20220405	Five minutes	28×28
TrafficWH	Traffic speed	Wuhan, China	20220305-20220405	Five minutes	30×28
TrafficCD	Traffic speed	Chengdu, China	20220305-20220405	Five minutes	28×26
TrafficJN	Traffic speed	Jinan, China	20220305-20220405	Five minutes	32×18
TrafficNJ	Traffic speed	Nanjing, China	20220305-20220405	Five minutes	32×24
TrafficSH	Traffic speed	Shanghai, China	20220127-20220227	Five minutes	28×32
TrafficSZ	Traffic speed	Shenzhen, China	20220305-20220405	Five minutes	24×18
TrafficGZ	Traffic speed	Guangzhou, China	20220305-20220405	Five minutes	32×26
TrafficGY	Traffic speed	Guiyang, China	20220305-20220405	Five minutes	26×28
TrafficTJ	Traffic speed	Tianjin, China	20220305-20220405	Five minutes	24×30
TrafficHZ	Traffic speed	Hangzhou, China	20220305-20220405	Five minutes	28×24
TrafficZZ	Traffic speed	Zhengzhou, China	20220305-20220405	Five minutes	26×26
TrafficBJ	Traffic speed	Beijing, China	20220305-20220405	Five minutes	30×32

Liu, Y., Zhang, H., Li, C., Huang, X., Wang, J., & Long, M. Timer: Generative Pre-trained Transformers Are Large Time Series Models. In ICML'24

Multimodality



- 1 Gather numerical data from reputable sources
- 2 Textual data is collected for fine-grained matching with the numerical data
- 3 Binary timestamps (start, end) are leveraged to mark the start and end dates as a universal temporal alignment method between numerical and textual data

Liu, H., Xu, S., Zhao, Z., Kong, L., Kamarthi, H., Sasanur, A. B., ... & Prakash, B. A. (2024). Time-MMD: A New Multi-Domain Multimodal Dataset for Time Series Analysis. arXiv preprint arXiv:2406.08627.

Multimodality

Numerical Data

Numerical data of each domain contains a csv file with has the following format:

```
start_date, end_date, OT, (other variable 1), (other variable 2), ...
```

Here, OT represents the default target variable for prediction in each dataset. Its specific meaning is as follows:

Table 1: Overview of numerical data in Time-MMD, covering key variables across nine domains with daily, weekly, or monthly frequencies, sourced from reputable government departments. Eight domains are updated to May 2024; the environment domain update is scheduled for June 2024.

Domain	Target	Frequency	Timestamps	Timespan
Agriculture	Retail Broiler Composite	Monthly	496	1983 - Present
Climate	Drought Level	Monthly	496	1983 - Present
Economy	International Trade Balance	Monthly	423	1989 - Present
Energy	Gasoline Prices	Weekly	1479	1996 - Present
Environment	Air Quality Index	Daily	11102	1982 - 2023
Health	Influenza Patients Proportion	Weekly	1389	1997 - Present
Security	Disaster and Emergency Grants	Monthly	297	1999 - Present
Social Good	Unemployment Rate	Monthly	900	1950 - Present
Traffic	Travel Volulmn	Monthly	531	1980 - Present

Textual Data

Textual data of each domain contains two csv file, one for report data and another for search data. All data are in a unified format:

```
start_date, end_date, fact, pred
```

Visualization of relevant report (a. left) and search (b. right) counts in Time-MMD over time is as follows:

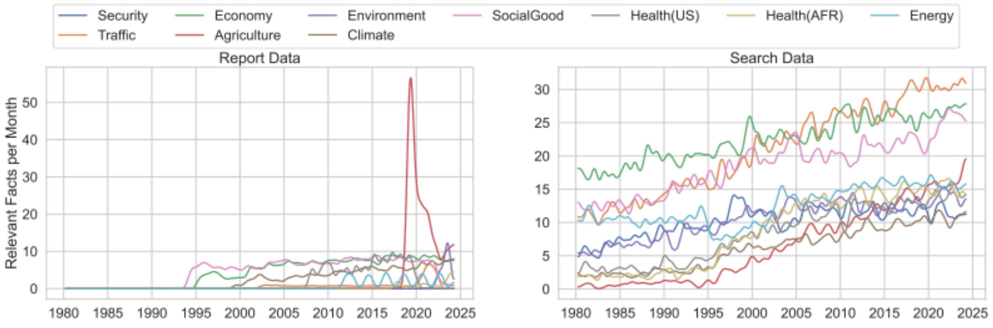
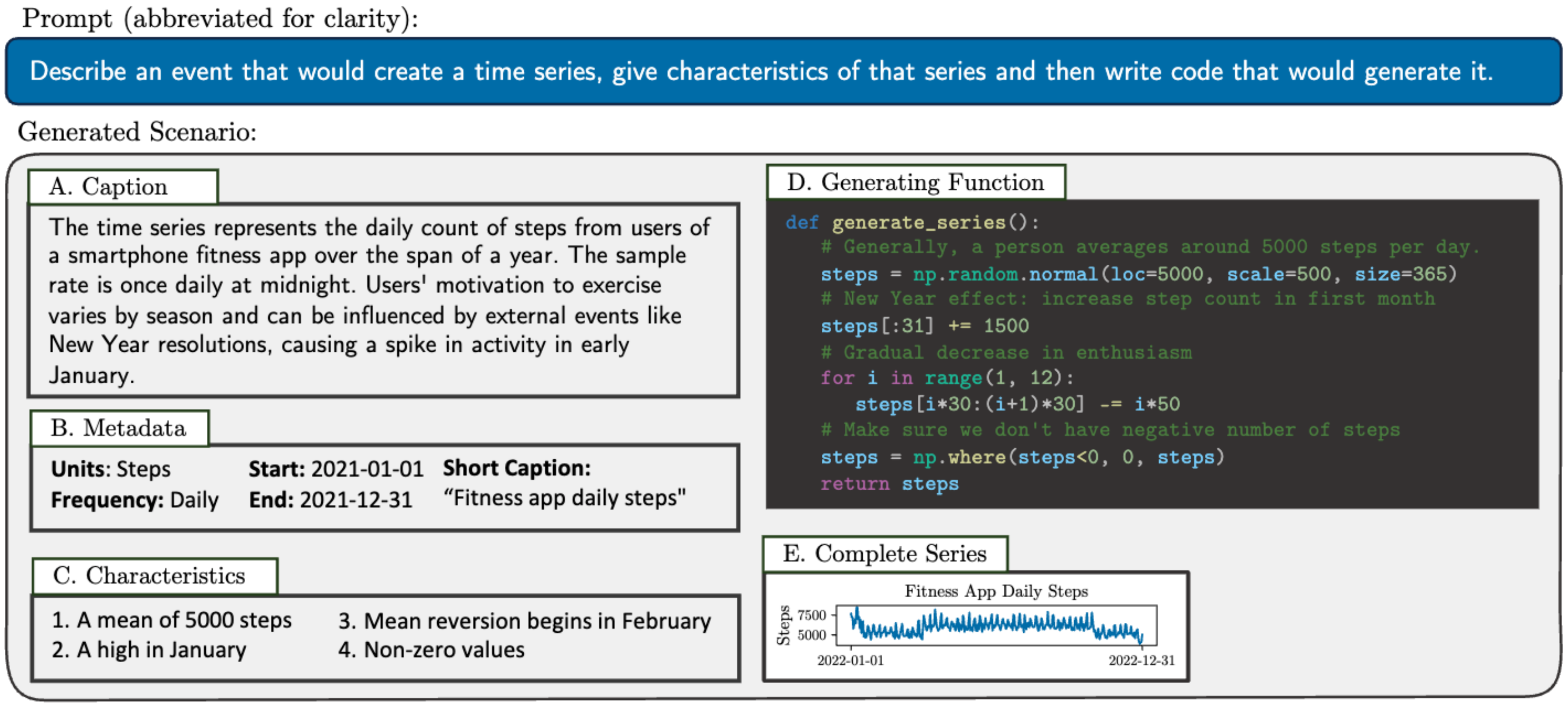


Figure 3: Visualization of relevant report (a, left) and search (b, right) counts in Time-MMD over time. Textual counts from both reports and searches increase over time. Domains with higher attention, such as the economy, have more relevant text data.

Liu, H., Xu, S., Zhao, Z., Kong, L., Kamarthi, H., Sasanur, A. B., ... & Prakash, B. A. (2024). Time-MMD: A New Multi-Domain Multimodal Dataset for Time Series Analysis. arXiv preprint arXiv:2406.08627.

Multimodality



1. Describe a scenario that might produce a time series. This scenario should include an external event and how it might influence the reading. Be sure to describe the sample rate of the time series and the duration over which it is sampled. The description should be less than 100 words in length. Delimit this description with the XML tag <description>.

   The time series must be less than 1000 observations in length, be a single variable, have no values greater than 1e6, and have no missing values.

   Also add a summary of the description, no more than 25 words in length with the tag <description_short>. Also add summary, no more than three words in length with the tag <description_tiny>. The scenario should be as different as possible from any of the following: [<previous_descriptions>]

2. You will generate a list of up to five characteristics of this specific time series, including patterns that you might expect to see in the series and how external events might cause distribution shifts in the data generating process Delimit these characteristics with the XML tag <characteristics>.

3. You will write a numpy function called 'generate_series' that takes no arguments and outputs a time series that matches the description. All parameters from the data generating process should be drawn from reasonable distributions. The function must return a single numpy array. Place this code inside a python markdown block and delimit your code with the XML tag < generator>. Do not call the function, simply define it. You should also make sure that the scale of time series is realistic. For example, a time series of a quantity like stock price should never be less than zero.

4. Return a json string, delimited by the tag <metadata> that contains the units of the time series and the timestamps corresponding to the first and last values. Remember that in JSON format datetimes must be passed as strings. Also include a string that relects the frequency of the time series.

Here is an example of a complete response:

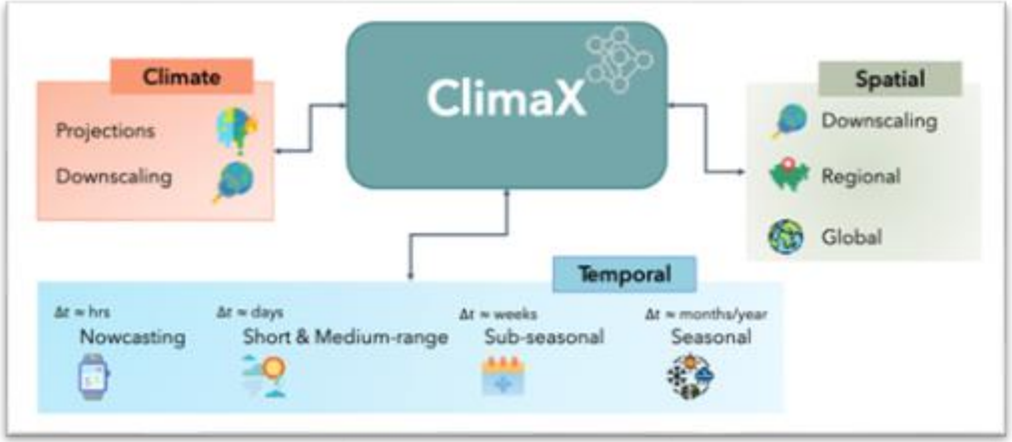
```
<description> *your description* </description>
<description_short> *your description* </description_short>
<description_tiny> *your description* </description_tiny>
<characteristics> *your characteristics* </characteristics>
<generator>
    ```python
    def generate_series():
        # your code here
        return x
    ```
</generator>
<metadata>
        {
        "start": x,
        "end": y,
        "units": z,
        "frequency" : freq
        }
</metadata>
```

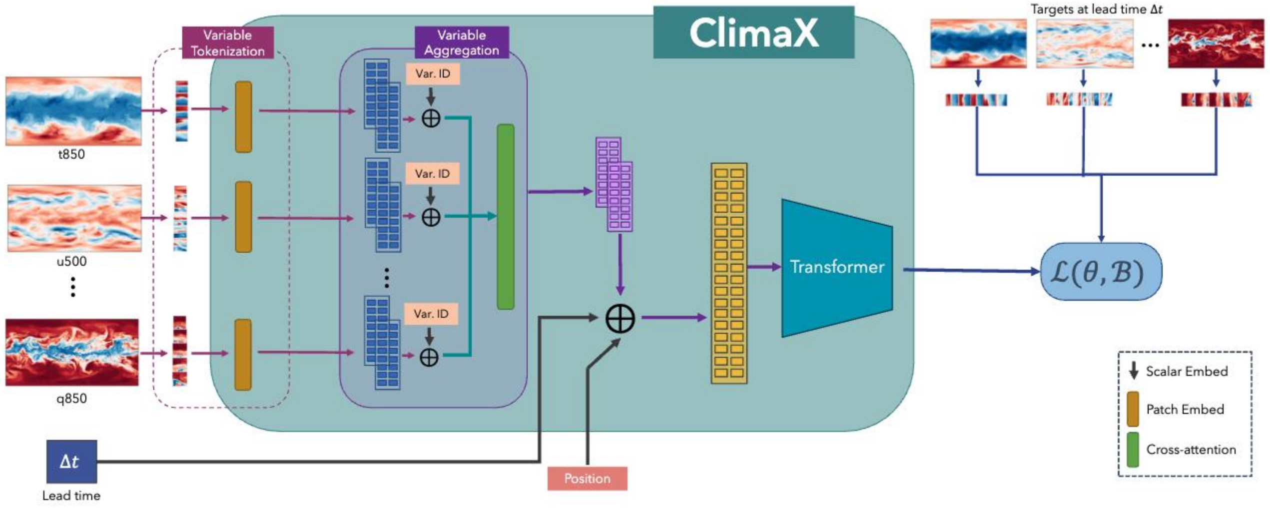
- We can ask the model to "imagine a scenario" that would produce a time series
- We then yield the following data for each scenario: caption, metadata, characteristics and a script for time series generation

Merrill, M. A., Tan, M., Gupta, V., Hartvigsen, T., & Althoff, T. (2024). Language Models Still Struggle to Zero-shot Reason about Time Series. arXiv preprint arXiv:2404.11757.

Application

- **Global Weather Forecasting**



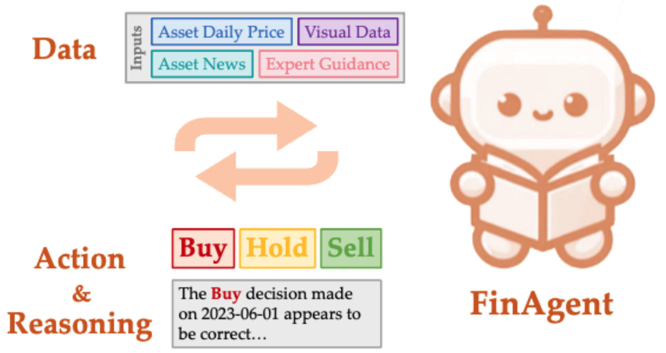


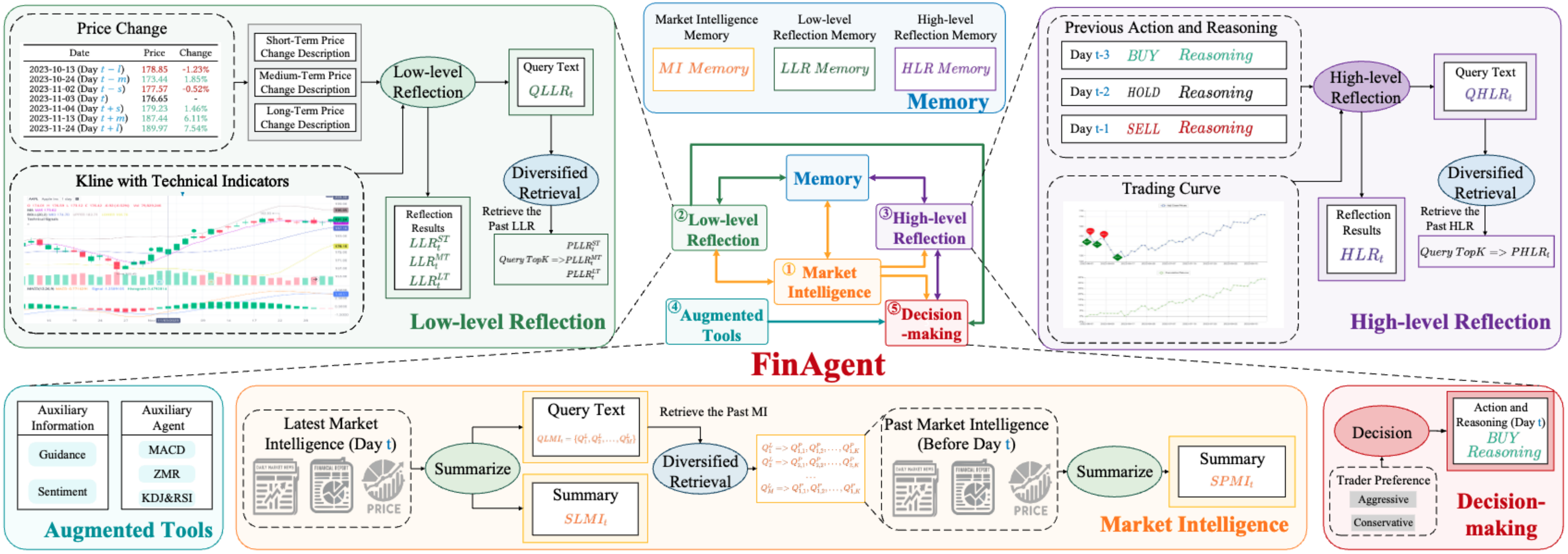
- ***Left***: ClimaX is built as a foundation model for diverse weather and climate modeling tasks
- ***Right***: Pretraining phase of ClimaX. Variables are encoded using variable-separate tokenization, and subsequently aggregated using variable aggregation. Together with position embedding and lead time embedding those are fed to the ViT backbone.

Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., & Grover, A. (2023). ClimaX: A foundation model for weather and climate. In ICML'23

Application

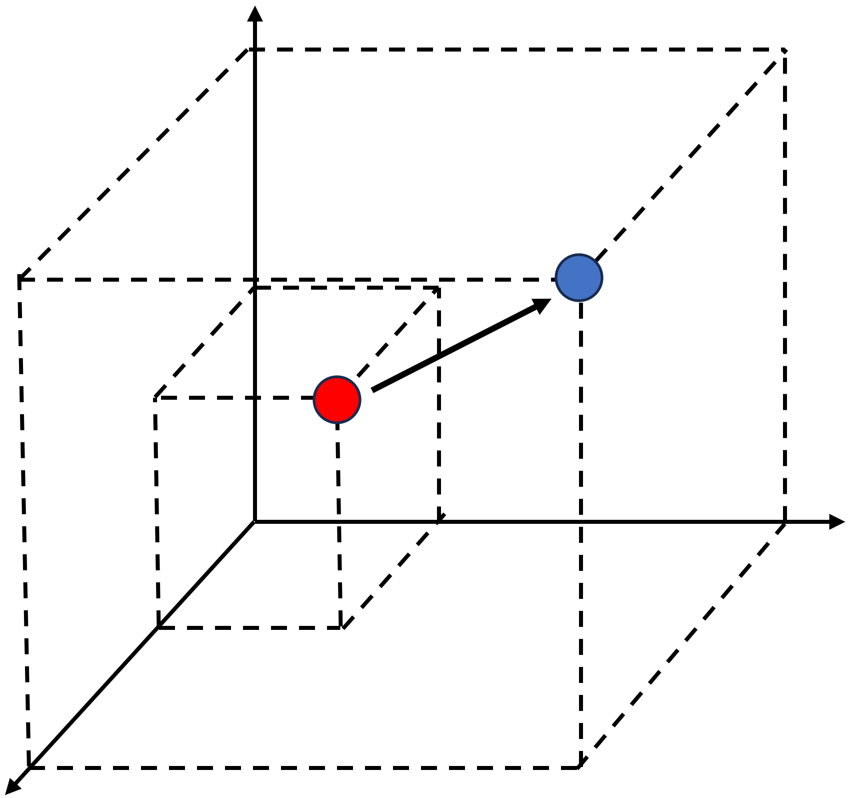
- **Financial Agent**





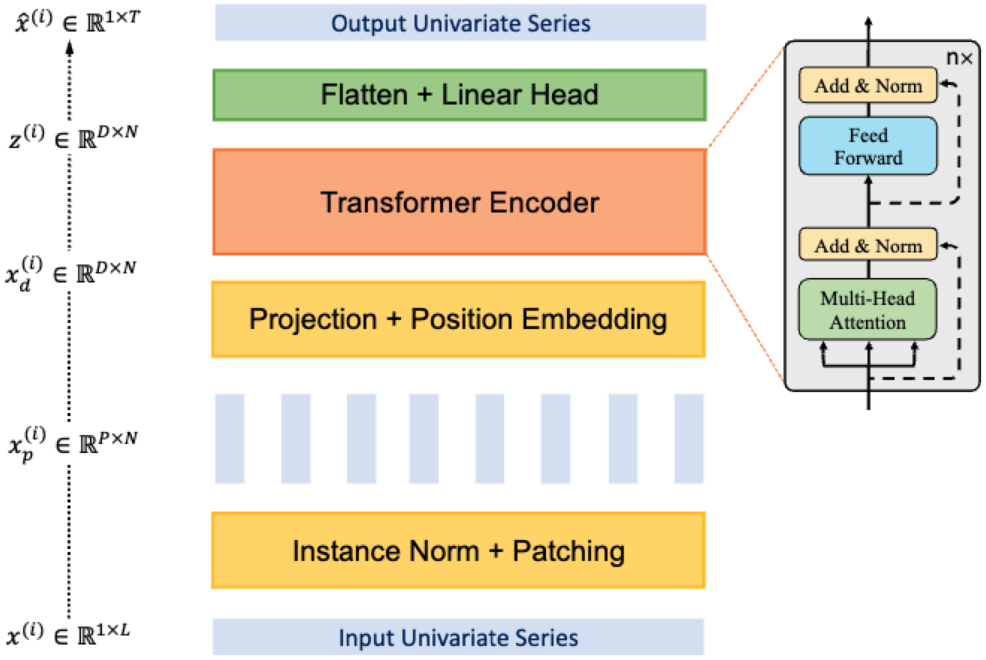
Zhang, W., Zhao, L., Xia, H., Sun, S., Sun, J., Qin, M., … & An, B. (2024). FinAgent: A Multimodal Foundation Agent for Financial Trading: Tool-Augmented, Diversified, and Generalist. In KDD'24

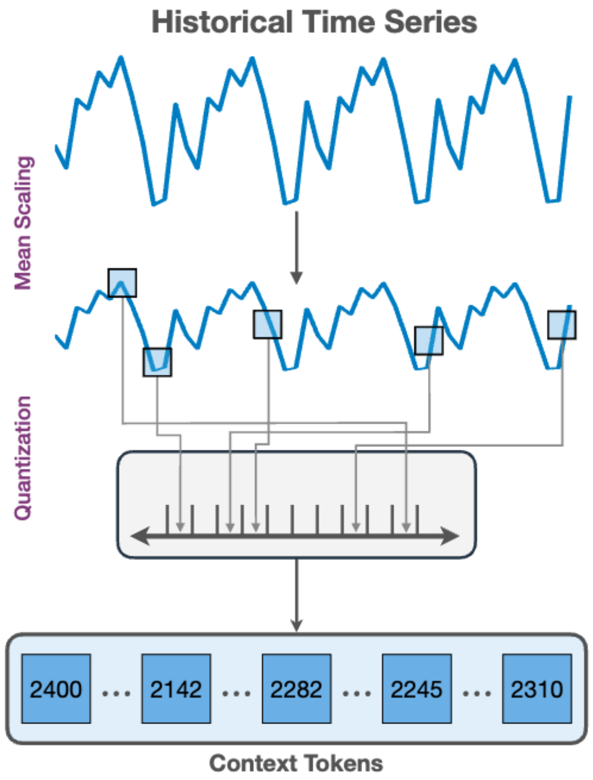
Where Are We



Tokenization

- **Time series tokenization is not easy**





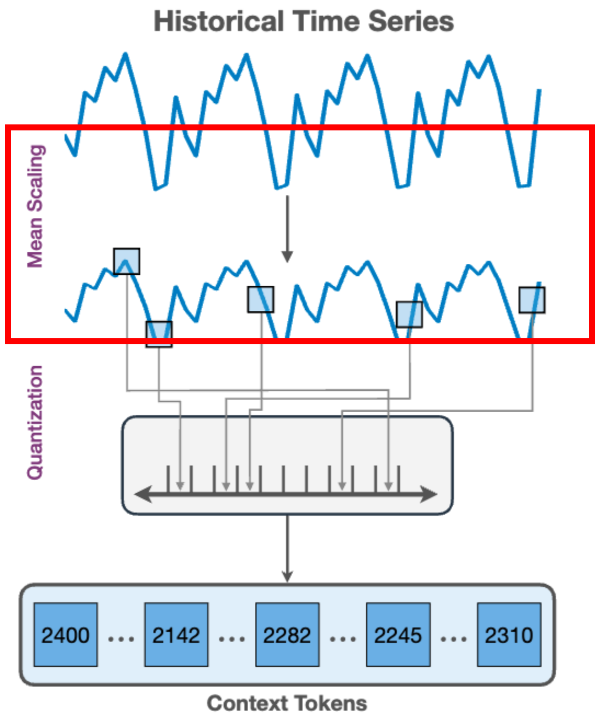
Patchify **Quantization**

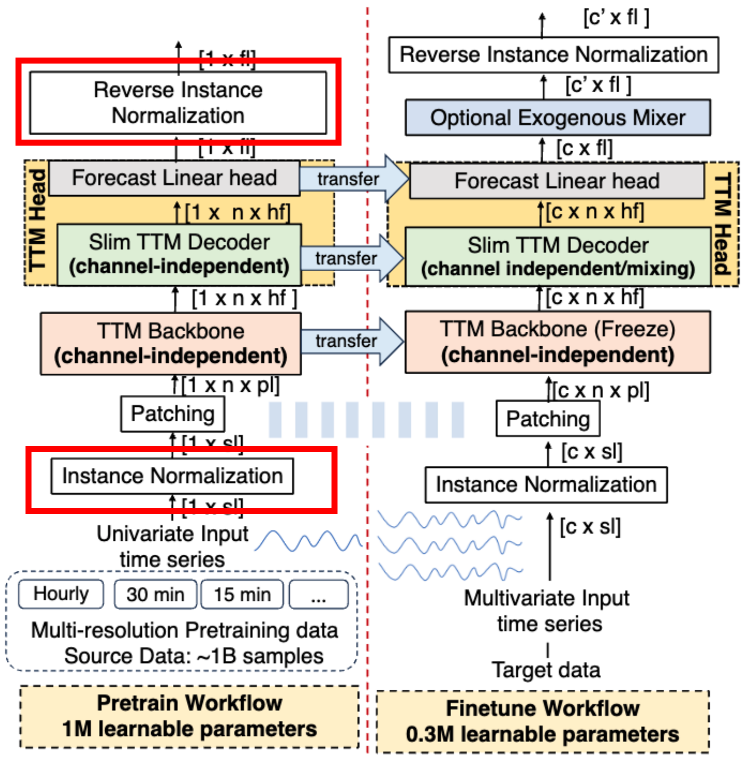
Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In ICLR'23

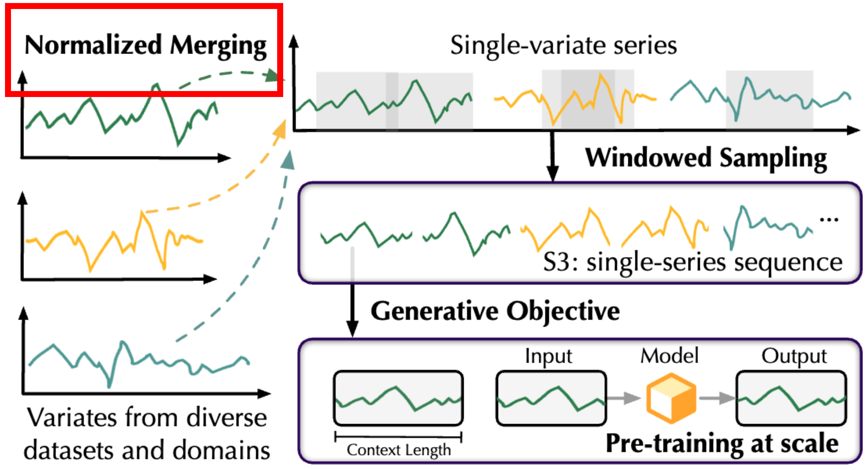
Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., ... & Wang, Y. (2024). Chronos: Learning the language of time series. arXiv preprint arXiv:2403.07815.

Normalization

- **Normalization is overlooked**







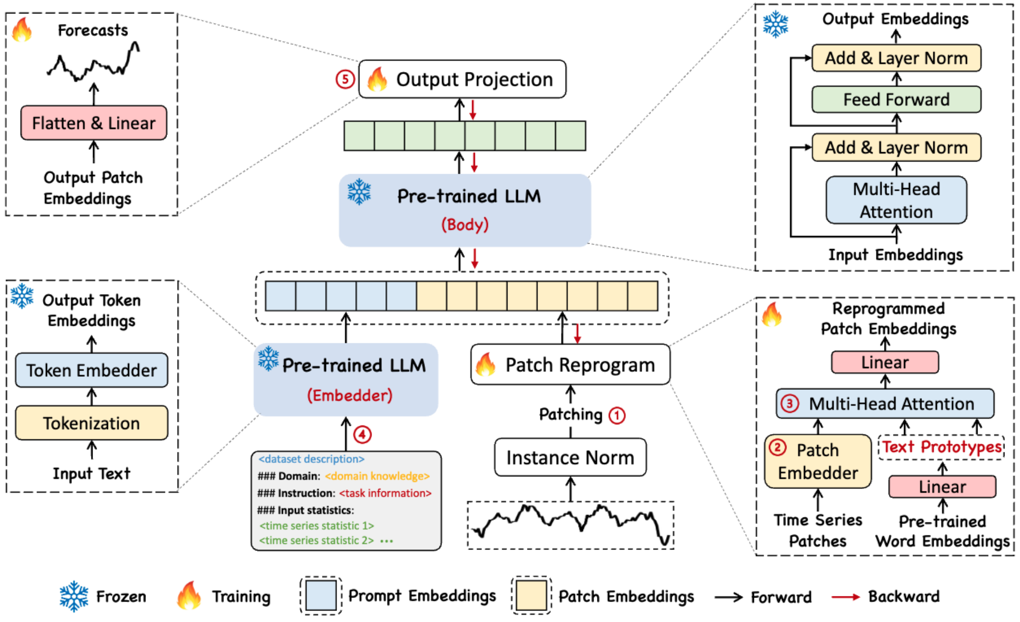
Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., ... & Wang, Y. (2024). Chronos: Learning the language of time series. arXiv preprint arXiv:2403.07815.

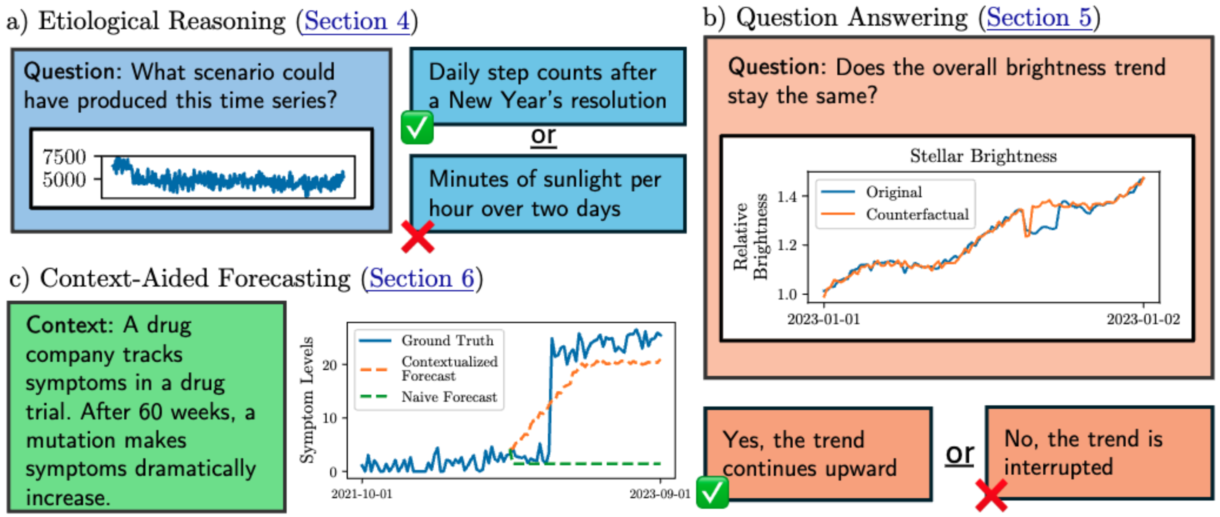
Ekambaram, V., Jati, A., Dayama, P., Mukherjee, S., Nguyen, N. H., Gifford, W. M., ... & Kalagnanam, J. (2024). Tiny Time Mixers (TTMs): Fast Pre-trained Models for Enhanced Zero/Few-Shot Forecasting of Multivariate Time Series. CoRR

Liu, Y., Zhang, H., Li, C., Huang, X., Wang, J., & Long, M. Timer: Generative Pre-trained Transformers Are Large Time Series Models. In ICML'24

Data Modality

- **Time series reasoning is promising**





Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J., Shi, X., ... & Wen, Q. (2024, May). Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. In ICLR'24

Merrill, M. A., Tan, M., Gupta, V., Hartvigsen, T., & Althoff, T. (2024). Language Models Still Struggle to Zero-shot Reason about Time Series. arXiv preprint arXiv:2404.11757.

Scaling Laws & Capabilities

- **Clear and robust scaling laws in language modeling**

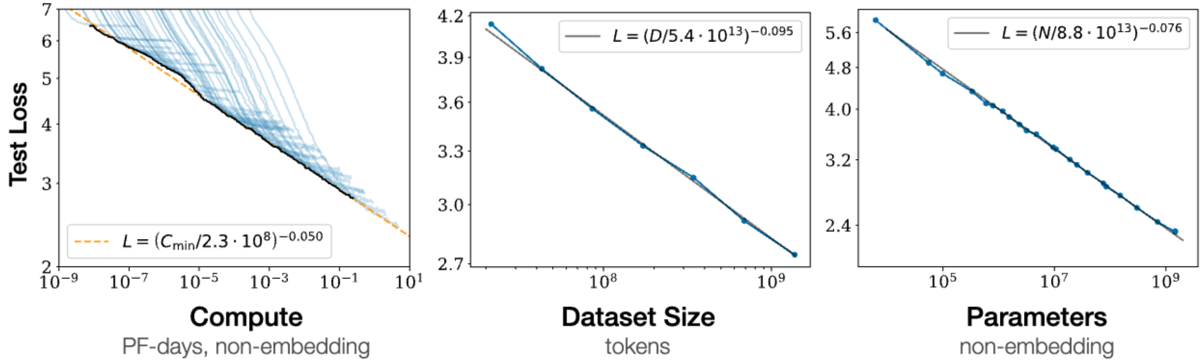
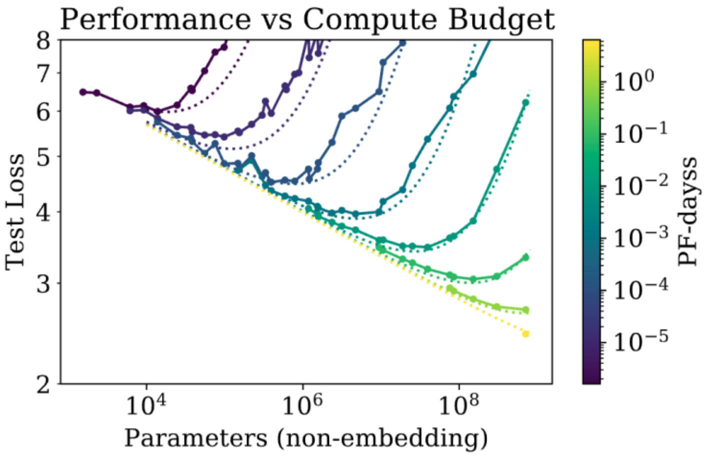
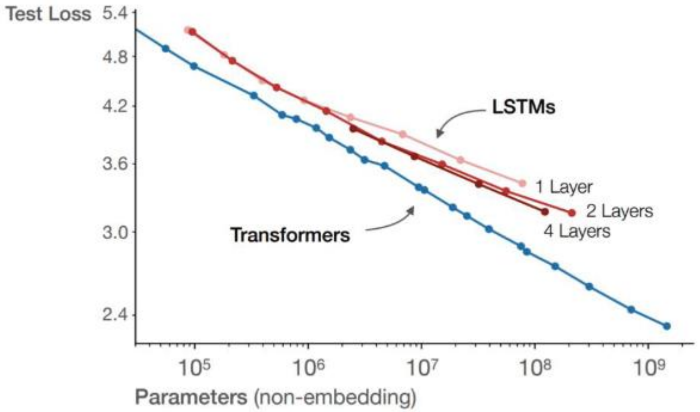


Figure 1 Language modeling performance improves smoothly as we increase the model size, datasetset size, and amount of compute[2] used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when not bottlenecked by the other two.

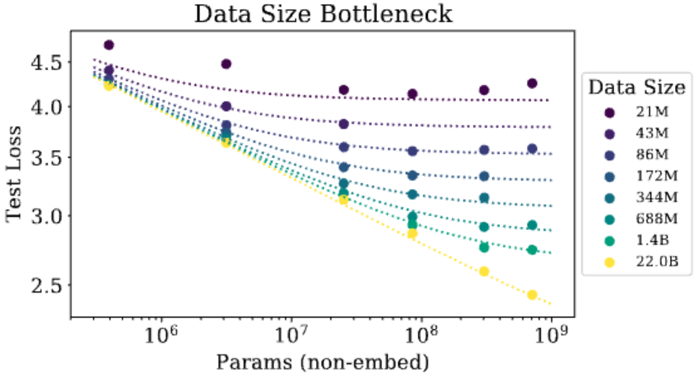
- Big model that is undertrained or small model that is well trained?

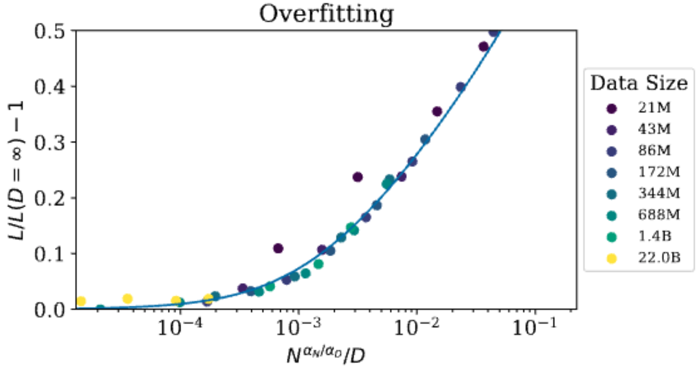


- Are Transformers better than LSTMs?



- Do we have enough data to feed our model?

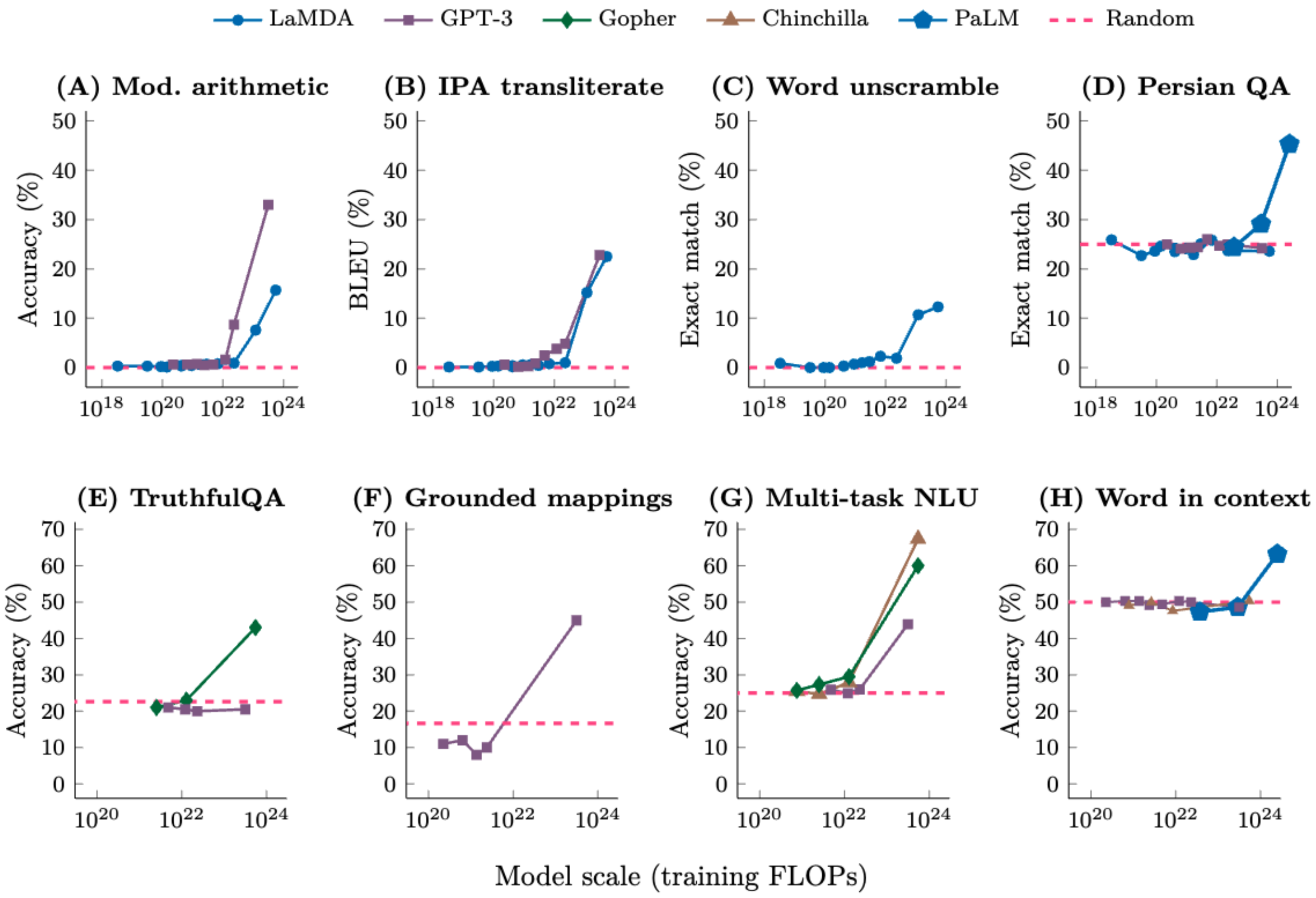


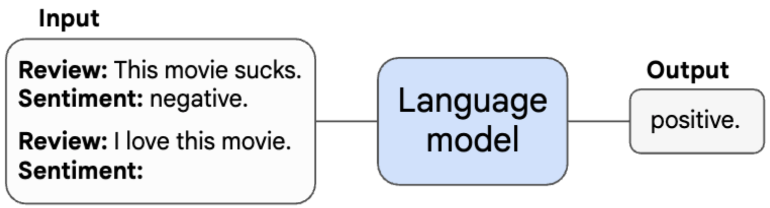


Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... & Amodei, D. (2020). Scaling laws for neural language models. arXiv preprint arXiv:2001.08361.
https://stanford-cs324.github.io/winter2022/assets/pdfs/Scaling%20laws%20pdf.pdf

Scaling Laws & Capabilities

- **Large models but why?**





Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022). Emergent abilities of large language models. arXiv preprint arXiv:2206.07682.

Scaling Laws & Capabilities

- **All we know so far…**

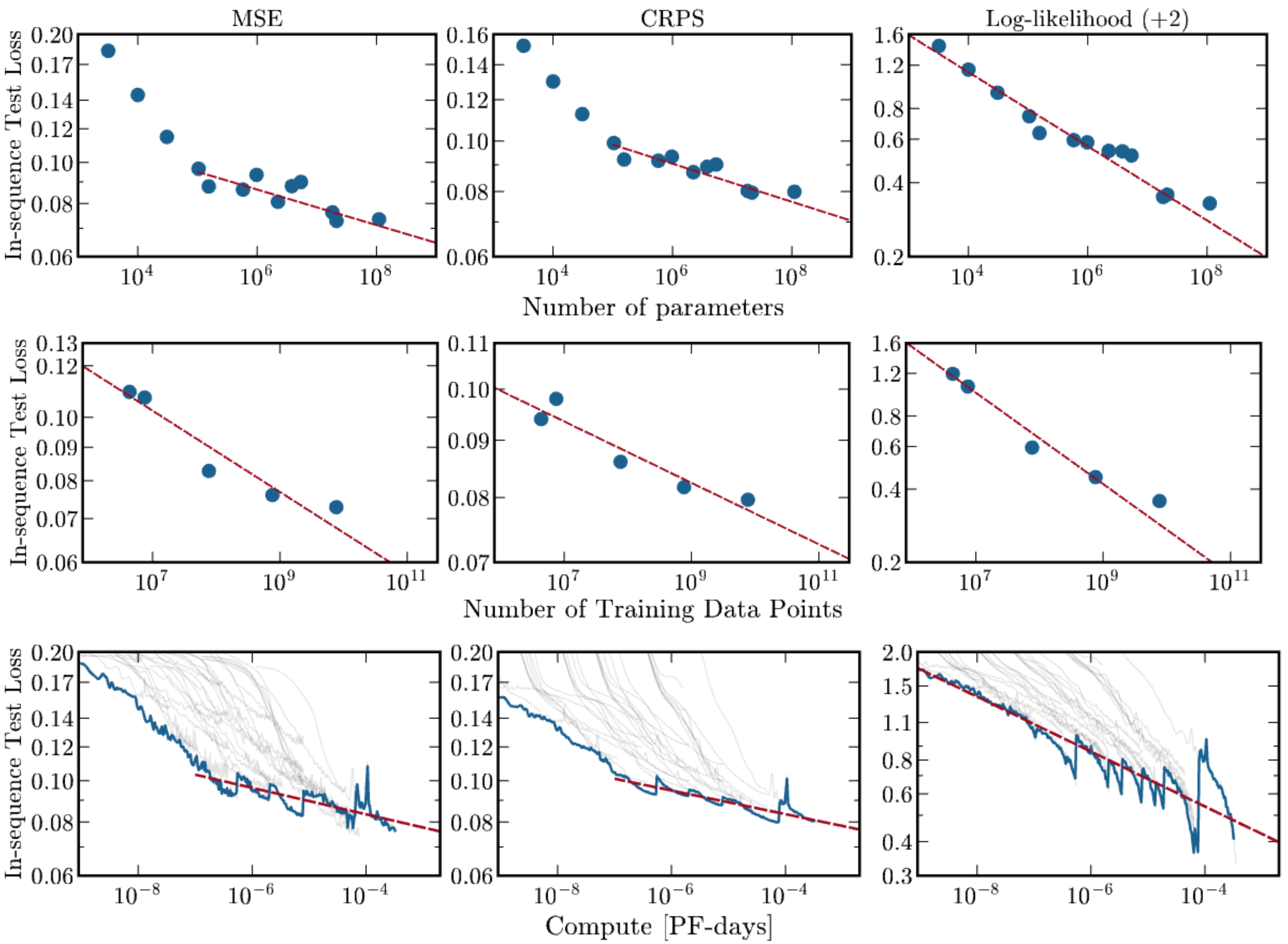
- Model parameters (e.g., 10K to 100M)

- Training tokens (e.g., 10M to 8B)

- Computation (e.g., PF-day budget)

> *"Large time series models scales approximately as a power law with all three quantities" -- Edwards et al.*

…



[2] Edwards, T. D., Alvey, J., Alsing, J., Nguyen, N. H., & Wandelt, B. D. (2024). Scaling-laws for Large Time-series Models. arXiv preprint arXiv:2405.13867.

Thank You

ORGANIZERS


Yuxuan Liang
Hong Kong University of Science and Technology (GZ)


Dongjin Song
University of Connecticut


Shirui Pan
Griffith University


Qingsong Wen
Squirrel AI, USA

CONTRIBUTORS


Haomin Wen
Carnegie Mellon University


Ming Jin
Griffith University


Yuqi Nie
Princeton University


Yushan Jiang
University of Connecticut