ADAPT DATA TO MODEL: ADAPTIVE TRANSFORMATION OPTIMIZATION FOR DOMAIN-SHARED TIME SERIES FOUNDATION MODELS

ICLR 2026 CONFERENCE PRESENTATION

Yunzhong Qiu*, Zhiyao Cen*, Zhongyi Pei†, Chen Wang, Jianmin Wang

School of Software, BNRist, Tsinghua University, China

Report Outline

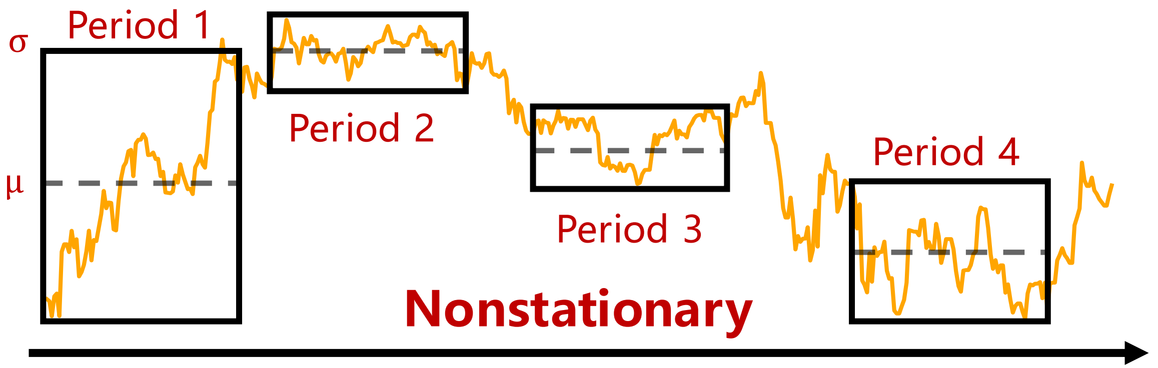
Background & Challenge

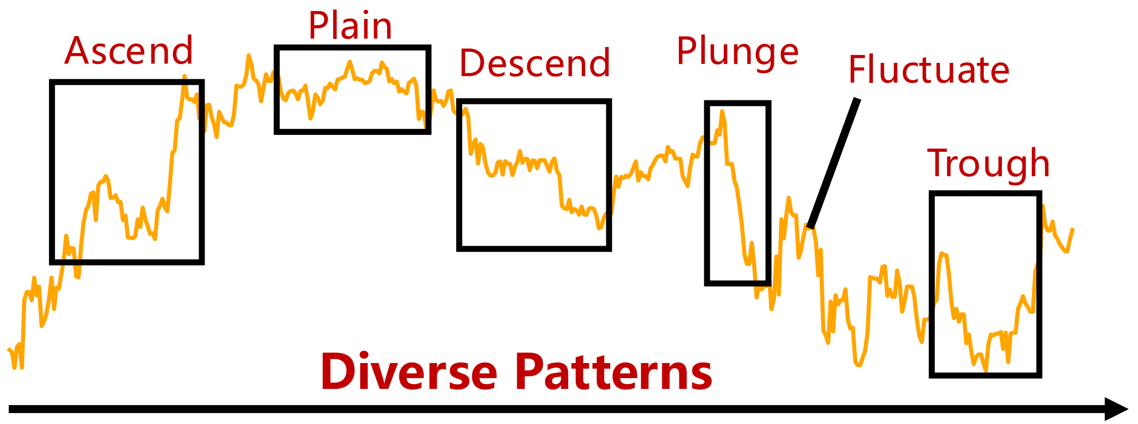
Research Background

Large Time Series Models (LTMs) are powerful pretrained **universal forecasters**. The **generalization ability** of LTMs exceeds traditional models.

Generalization Challenges

Real-world time series data are **diverse** and **non-stationary**. A single pre-trained LTM is difficult to adapt to all domains.
Traditional **fine-tuning** process can harm **generalization** and is often **computationally expensive**.





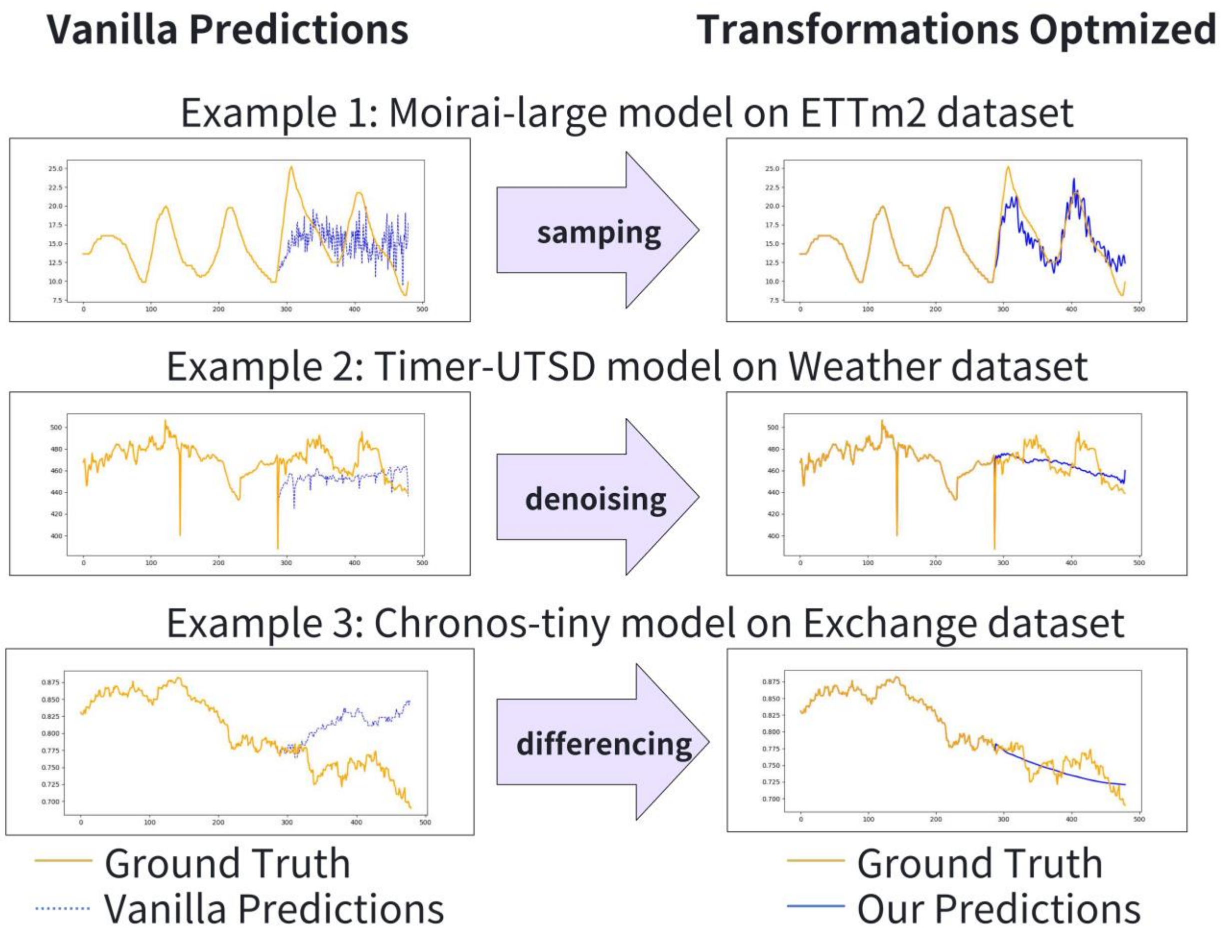
Core Insight & Motivation

Core Insights

By applying **data transformations** to targeted time series data, the prediction performance of the frozen LTM can be **significantly improved without modifying the model architecture**.

Realistic Requirements

Can we adapt **frozen LTM** to multiple downstream domains through optimizing **data transformation pipelines**?
Can we do this in an **efficient** way without harming **accuracy** and **generalization**?



FrozenForecasting: A New Paradigm

Paradigm Definition

Use a single **frozen pre-trained foundation model** and applies **lightweight, domain-specific adaptive transformations** only to the input data to achieve accurate forecasting across multiple downstream domains.

Input
Time Series
Historical Data

→

Stage1
Adaptive Transformation
Find the optimal transformation pipeline h* in the transformation space H to minimize the model's loss.

→

Stage2
Frozen Pre-trained LTM
Conduct zero-shot inference using optimized transformation pipeline

→

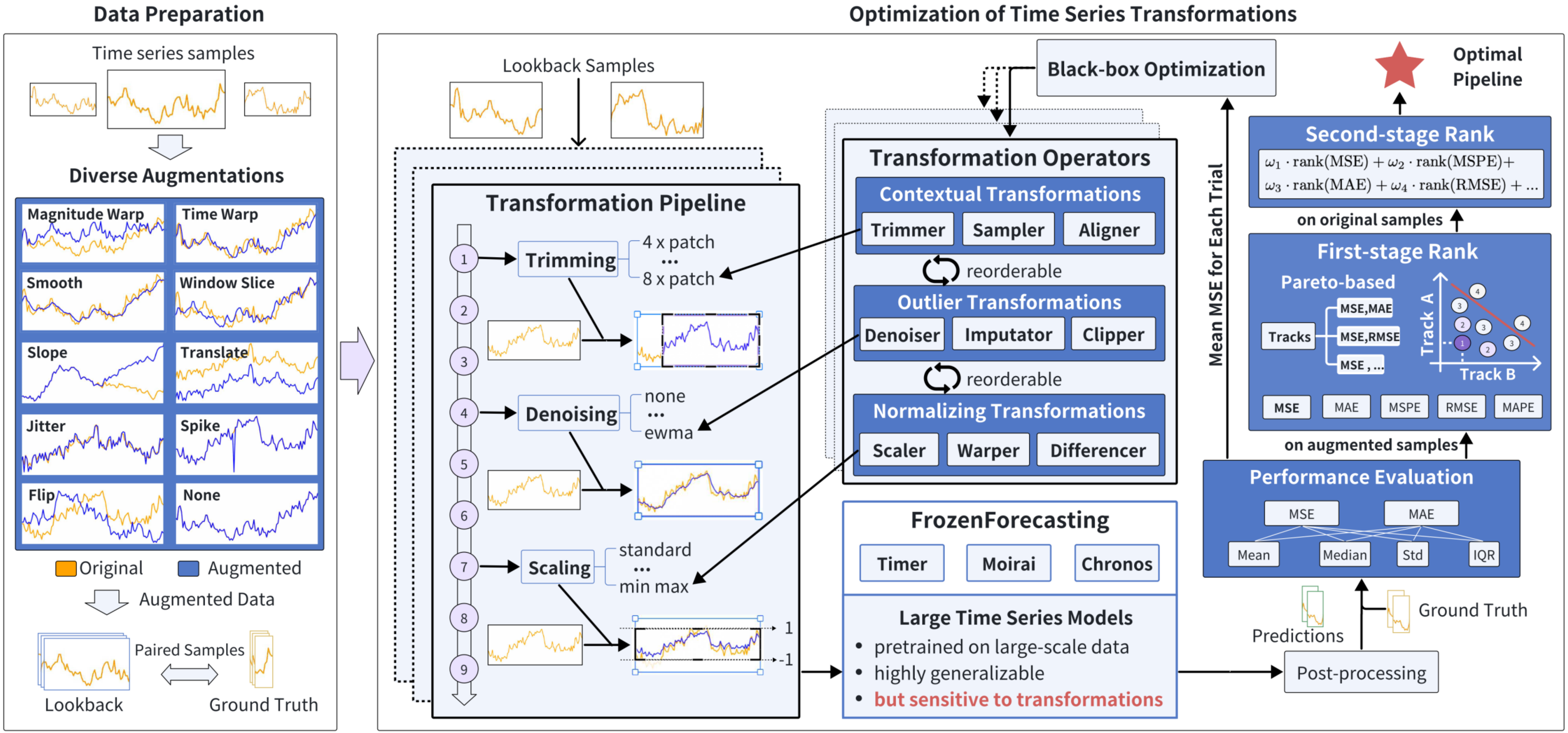
Output
Time Series
Predictions

Model Frozen
Preserves native generalization ability

Data-Centric
Extremely low computational cost

Model-Agnostic
Compatible with any SOTA LTM

TATO Framework Overview



Key Experiment Results: Performance Improvement

Verified on **8** mainstream time series datasets, **6** open-source baseline LTMs, and **4** forecasting horizons. Including **192** different forecasting scenarios.

Models		Timer-UTSD		Timer-LOTSA		Moirai-small		Moirai-base		Moirai-large		Chronos-tiny		Average	
Error Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	Vanilla	0.4096	0.4846	0.4372	0.5010	0.4628	0.5163	0.4901	0.5249	0.4381	0.4989	0.4371	0.5055	0.4458	0.5052
	with TATO	0.3901	0.4762	0.4379	0.5010	0.4460	0.5044	0.4503	0.5081	0.4141	0.4823	0.4460	0.5036	0.4307	0.4960
	%Promotion	4.8%	1.7%	-0.2%	0.0%	3.6%	2.3%	8.1%	3.2%	5.5%	3.3%	-2.0%	0.4%	3.4%	1.8%
ETTh2	Vanilla	0.3258	0.4413	0.3646	0.4777	0.5249	0.5660	0.4657	0.5287	0.4494	0.5119	0.4181	0.4892	0.4247	0.5025
	with TATO	0.3203	0.4375	0.3431	0.4535	0.4691	0.5344	0.4538	0.5211	0.3688	0.4692	0.4297	0.5001	0.3975	0.4860
	%Promotion	1.7%	0.9%	5.9%	5.1%	10.6%	5.6%	2.6%	1.4%	17.9%	8.3%	-2.8%	-2.2%	6.4%	3.3%
ETTm1	Vanilla	0.2963	0.3979	0.3687	0.4550	0.2554	0.3682	0.2648	0.3701	0.2380	0.3549	0.2815	0.3831	0.2841	0.3882
	with TATO	0.2890	0.3910	0.2189	0.3380	0.2259	0.3467	0.2286	0.3478	0.2253	0.3445	0.2457	0.3561	0.2389	0.3540
	%Promotion	2.5%	1.8%	40.6%	25.7%	11.6%	5.8%	13.7%	6.0%	5.3%	2.9%	12.7%	7.0%	15.9%	8.8%
ETTm2	Vanilla	0.4644	0.5224	0.3146	0.4166	0.2969	0.3996	0.3747	0.4391	0.3288	0.3984	0.4006	0.4307	0.3633	0.4345
	with TATO	0.4368	0.4909	0.2506	0.3520	0.2374	0.3405	0.2512	0.3477	0.2277	0.3257	0.2594	0.3447	0.2772	0.3669
	%Promotion	5.9%	6.0%	20.3%	15.5%	20.1%	14.8%	33.0%	20.8%	30.7%	18.3%	35.3%	20.0%	23.7%	15.5%
Elec.	Vanilla	0.1753	0.2904	0.1714	0.2895	0.6550	0.6305	0.5761	0.5749	0.6783	0.6352	0.2733	0.3632	0.4216	0.4639
	with TATO	0.1729	0.2859	0.1710	0.2885	0.5539	0.5566	0.4434	0.4949	0.5100	0.5350	0.2716	0.3629	0.3538	0.4206
	%Promotion	1.4%	1.6%	0.3%	0.3%	15.4%	11.7%	23.0%	13.9%	24.8%	15.8%	0.6%	0.1%	16.1%	9.3%
Exch.	Vanilla	0.5190	0.5004	0.8317	0.7013	0.2897	0.3855	0.2524	0.3615	0.2450	0.3570	0.4013	0.4404	0.4232	0.4577
	with TATO	0.3073	0.3905	0.2881	0.3727	0.2797	0.3801	0.2570	0.3586	0.2635	0.3673	0.2529	0.3541	0.2748	0.3706
	%Promotion	40.8%	22.0%	65.4%	46.9%	3.5%	1.4%	-1.8%	0.8%	-7.5%	-2.9%	37.0%	19.6%	35.1%	19.0%
Traffic	Vanilla	0.0605	0.1380	0.0616	0.1404	0.3646	0.4982	0.3364	0.4690	0.3049	0.4233	0.0762	0.1548	0.2007	0.3040
	with TATO	0.0589	0.1353	0.0602	0.1384	0.3326	0.4663	0.2743	0.4112	0.2519	0.3766	0.0769	0.1554	0.1758	0.2805
	%Promotion	2.5%	2.0%	2.3%	1.5%	8.8%	6.4%	18.5%	12.3%	17.4%	11.0%	-0.9%	-0.4%	12.4%	7.7%
Weather	Vanilla	0.4937	0.4967	0.5888	0.5533	0.4984	0.4785	0.4929	0.4606	0.5422	0.4958	0.7282	0.5692	0.5573	0.5090
	with TATO	0.6048	0.5290	0.5898	0.5343	0.4944	0.4814	0.4961	0.4687	0.5107	0.4818	0.5980	0.5245	0.5490	0.5033
	%Promotion	-22.5%	-6.5%	-0.2%	3.4%	0.8%	-0.6%	-0.7%	-1.8%	5.8%	2.8%	17.9%	7.9%	1.5%	1.1%
Average	Vanilla	0.3431	0.4090	0.3923	0.4419	0.4185	0.4804	0.4066	0.4661	0.4031	0.4594	0.3770	0.4170	0.3901	0.4456
	with TATO	0.3225	0.3920	0.2950	0.3723	0.3799	0.4513	0.3568	0.4323	0.3465	0.4228	0.3225	0.3877	0.3372	0.4097
	%Promotion	6.0%	4.1%	24.8%	15.7%	9.2%	6.0%	12.2%	7.3%	14.0%	8.0%	14.5%	7.0%	13.6%	8.1%

84.3%

TATO's performance is on par native frozen LTM in 84.3% of scenarios.

13.6%

MSE is reduced by 13.6% on average; MAE is reduced by 8.1% on average.

65.4%

TATO brings significant improvements in scenarios where the native LTM performs poorly.

Efficiency & Scalability Verification

Optimization Efficiency

The optimization process of TATO can be completed within **2 minutes** in most scenarios.

Inference Overhead

The additional overhead brought by transformations is less than **3ms** per sample during inference.

Strong Scalability

The optimization effect **steadily improves** with the increase of search iterations and sample size.

< 2 Min	< 3 ms	Stable Growth
Optimization Time	Per-sample Overhead	Scalability Trend

Ablation Study & Additional Verification

Operators Ablation

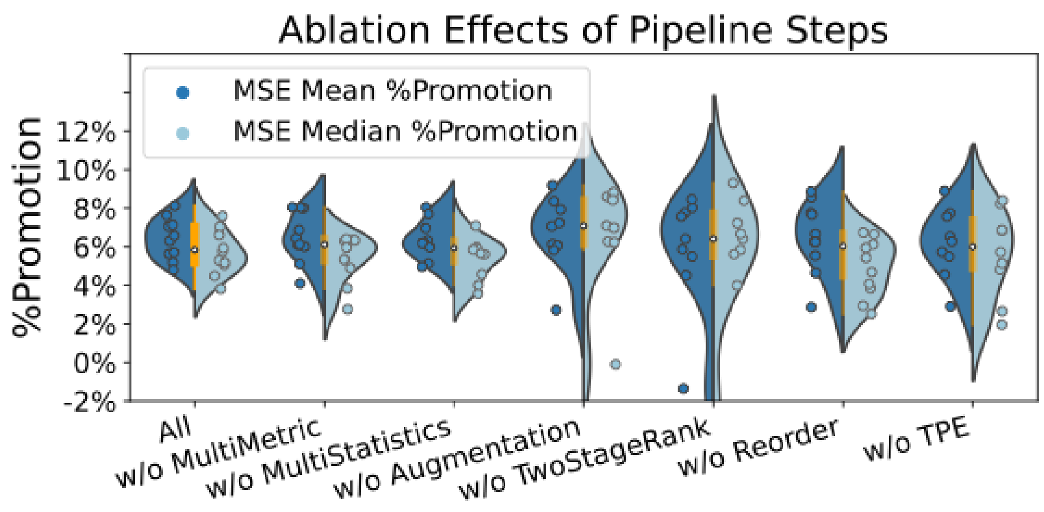
All operators contribute positively to the final result, with **Trimmer** and **Scaler** having the greatest impact.

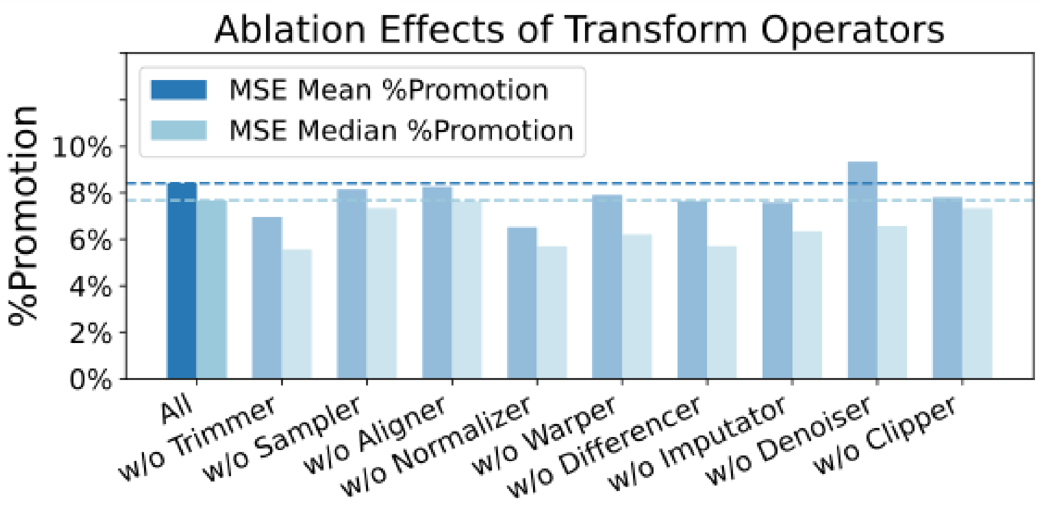
Modules Ablation

Denoising operators and the two-stage ranking mechanism significantly improve the **robustness** of the results.

Additional Verification

TATO can further reduce MSE by **7.3%** on LTMs after full-domain fine-tuning, complementing the fine-tuning scheme.





Conclusion & Future Work

Key Conclusions

- Proposed **FrozenForecasting paradigm** to enable a single frozen LTM to adapt across multiple domains.
- Designed the **TATO framework**, which significantly improves the prediction performance and robustness of frozen LTMs by automatically optimizing time series transformation pipelines.
- Extensive experiments verify the **general effectiveness and high efficiency** of TATO, which can be directly deployed in real-world scenarios.

Future Work

- Extend to **multivariate** time series forecasting scenarios.
- Expand transformation operators **for specific domains**.
- Combine with **test-time adaptation** schemes to further improve robustness under distribution shifts.

Acknowledgements

Acknowledgments to the **National Key R&D Program of China, BNRist project, and the National Engineering Research Center for Big Data Software** for their support.

Acknowledgments to **Mengren Zheng** for his contribution in supplementing experiments during the rebuttal phase.

Thank You for Your Attention

qiuyz24@mails.tsinghua.edu.cn, cenzy23@mails.tsinghua.edu.cn

https://github.com/thulab/TATO

清華大學
Tsinghua University