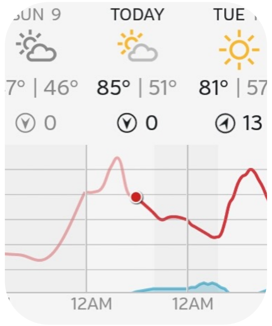
AutoDA-Timeseries: Automated Data Augmentation for Time Series

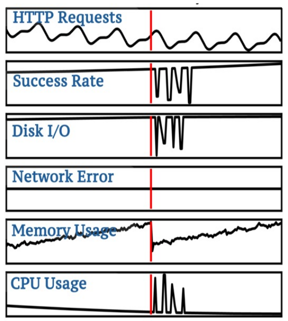
Zijun Dou[1], Zhenhe Yao, Zhe Xie, Xidao Wen, Tong Xiao, Dan Pei

1. Presenter. Email: douzj25@mails.tsinghua.edu.cn
Tsinghua NetMan Lab

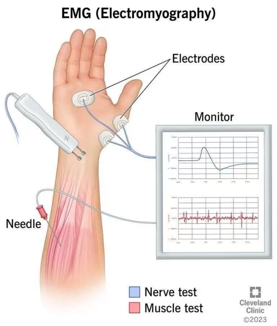
Time Series Data



Weather

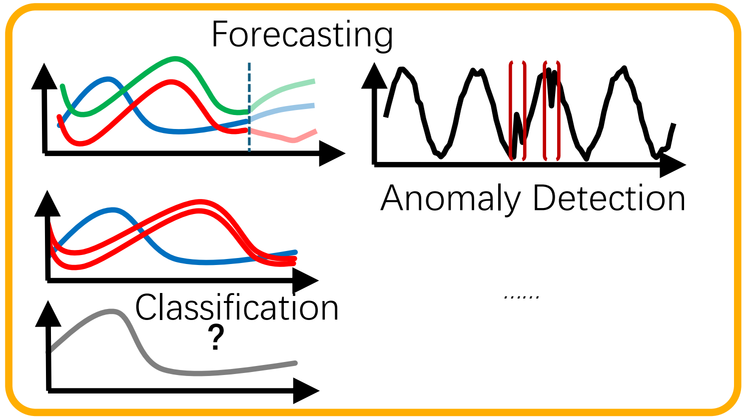


AIOps



Healthcare

Time series data has been widely applied to different domains



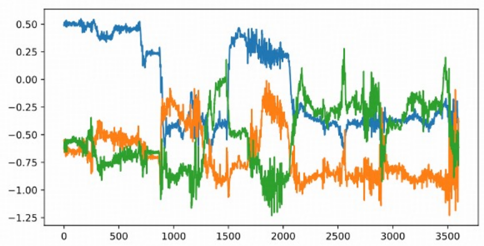
Diverse downstream tasks

High-quality time series data is scarce

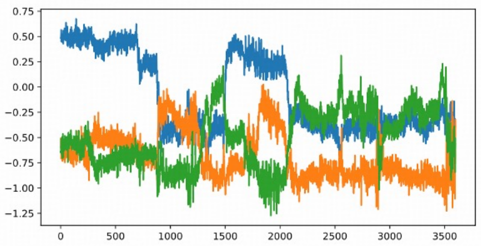
- High cost of data acquisition devices
- Privacy and security constraints
- Missing values are common

Time Series Data Augmentation

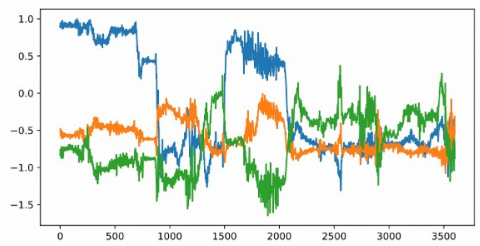
Time series data augmentation leverages transformations in the ***amplitude***, ***time***, ***frequency***, and ***time-frequency*** domains



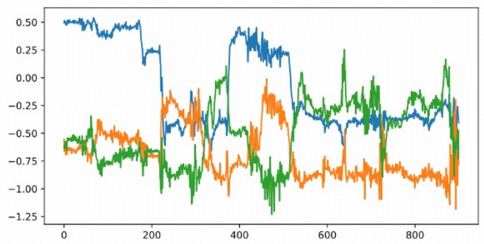
Original Time Series



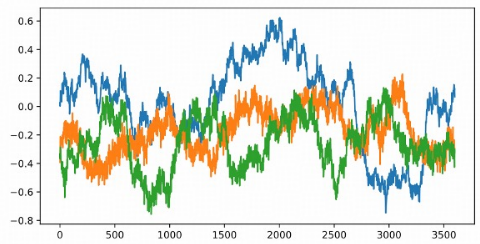
Jittering



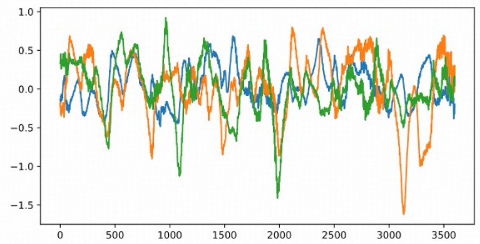
Scaling



Downsampling



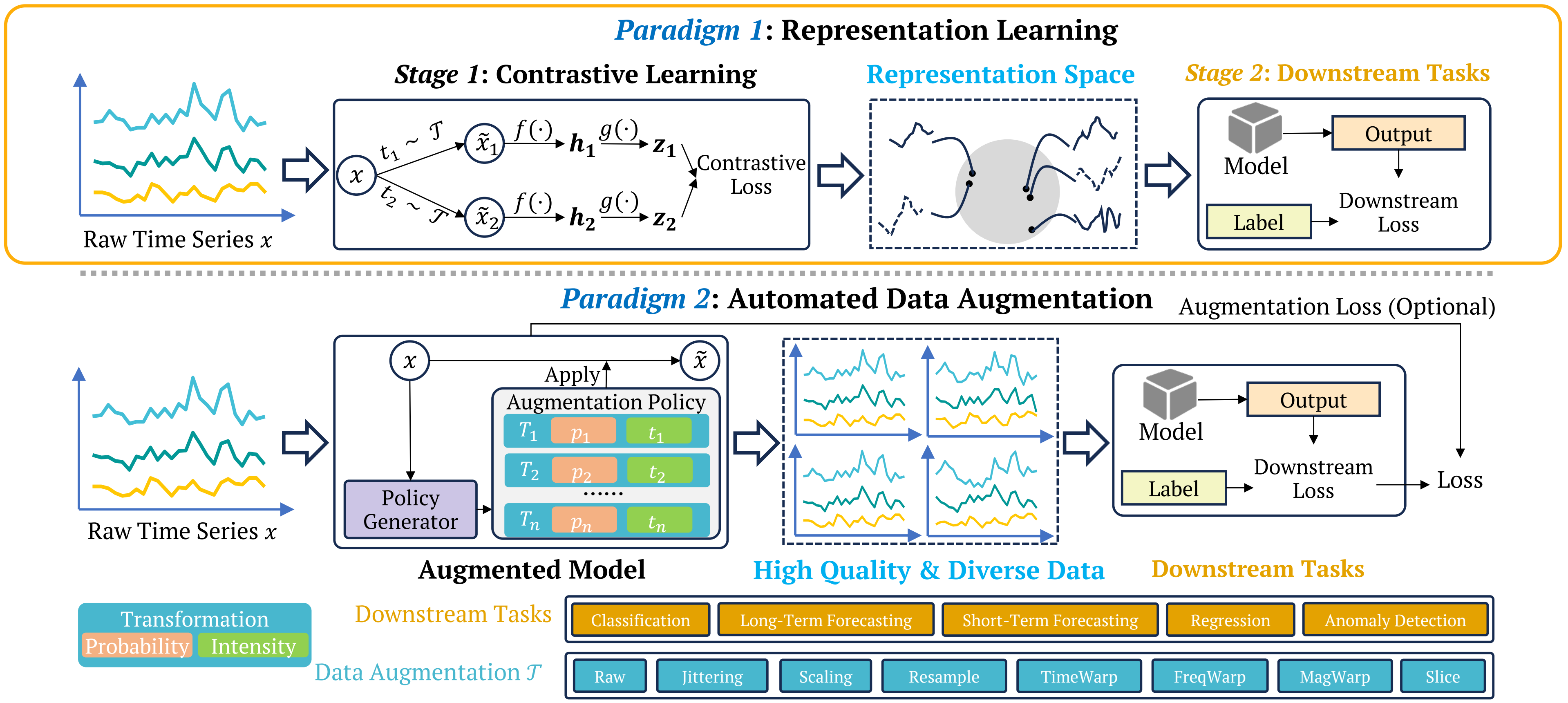
AAFT



STFT

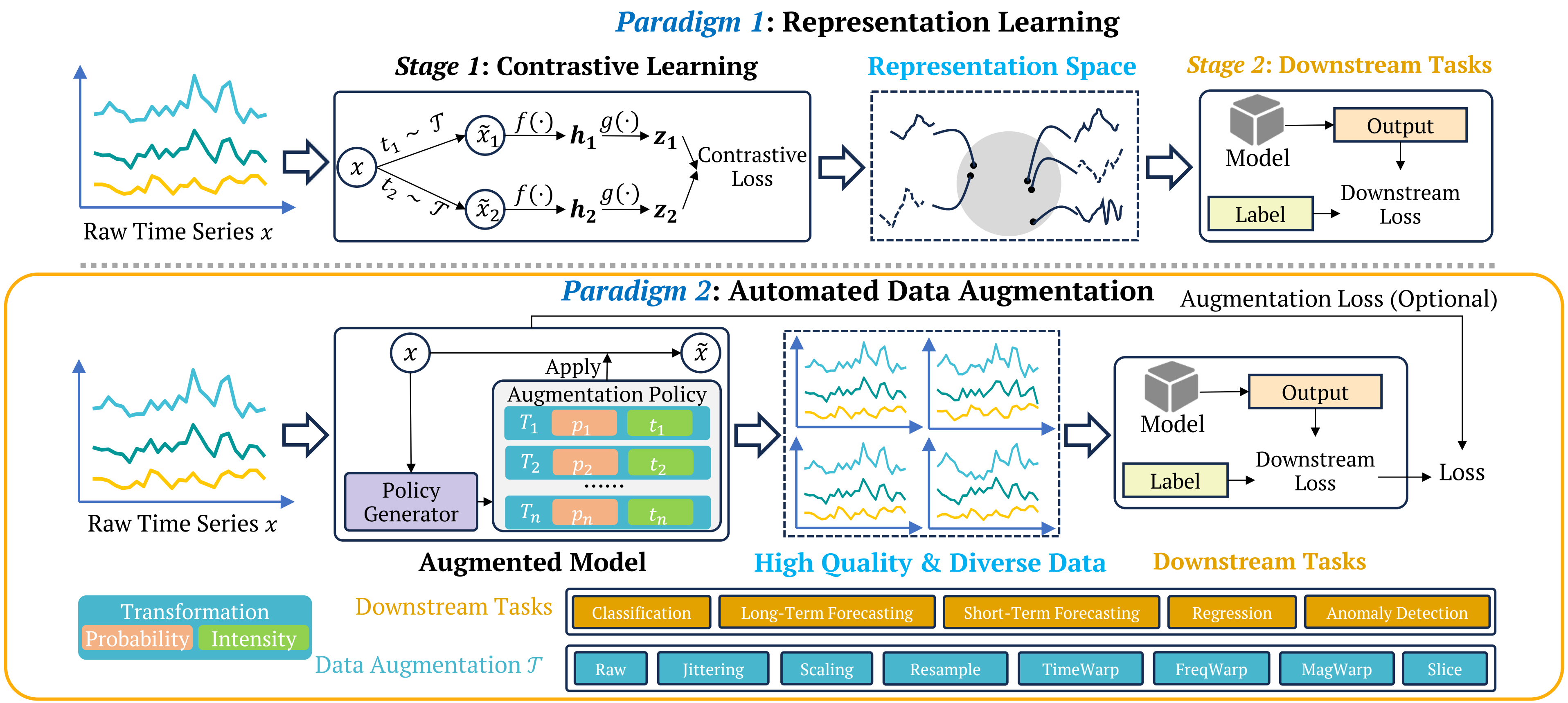
These transformations generate diverse training samples, enabling downstream models to achieve ***better generalization and improved performance***

Two Application Paradigms of Time Series Augmentation



- **Representation Learning**
  - *Stage 1*: augmentations are used to construct contrastive samples, enabling models to learn task-agnostic representations
  - *Stage 2*: train downstream model to adapt the learned encoder representations

Two Application Paradigms of Time Series Augmentation



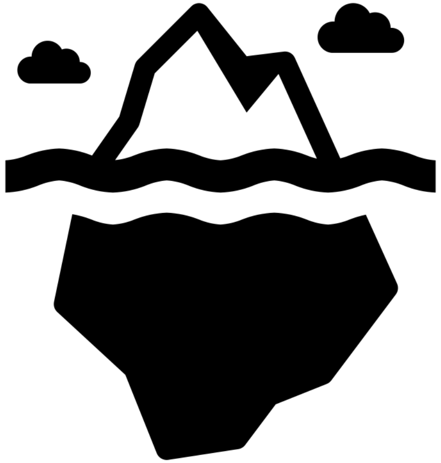
- **Automated Data Augmentation (AutoDA)**
  - Focuses on *automatically generating augmentation strategies* that *directly optimize downstream model performance* while reducing the reliance on manual design

Representation Learning vs. AutoDA

- **Key Limitation of Representation Learning: Two-Stage Misalignment**
  - The *adaptability* of downstream models to the learned representations
  - E.g., RNNs are inherently designed for sequence-to-sequence prediction, excelling at modeling long-term dependencies rather than capturing invariant representations

- **AutoDA** follows a one-stage scheme where augmentations are ***jointly optimized*** with the downstream model

Key Challenges



Limited Task Generalization

- Most existing AutoDA methods are validated on *a single task*
- This narrow evaluation *overlooks the poor generalization* of augmentation policies across different time series tasks

Neglect of Time Series Characteristics

- Existing methods *ignore time series-specific features* when generating augmentation policies
- Modality-agnostic transformations can *disrupt intrinsic properties*, resulting in degraded downstream performance

Lack of Adaptive Policy Learning

- Previous SOTA methods rely on *uniform sampling* to determine augmentation transformations
- This ignores *the heterogeneous impact* of different transformations, leading to inappropriate augmentation policies

Our Goals

Domain Data

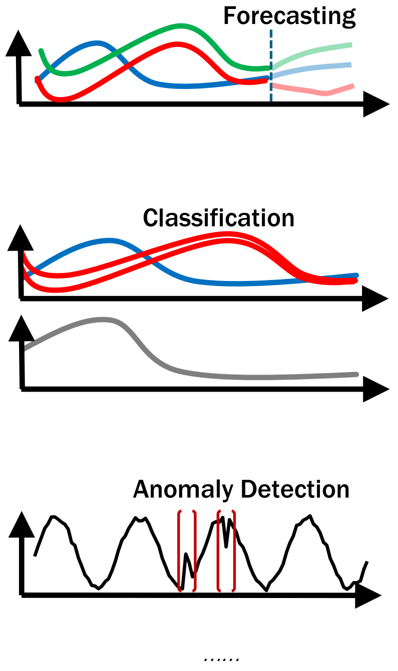
- Energy
- Healthcare
- AIOps
- Weather
- Economy
- ......

AutoDA-Timeseries

- Feature-Driven
- Adaptive
- Plug-and-Play
- Physically Meaningful
- Efficient
- Interpretable

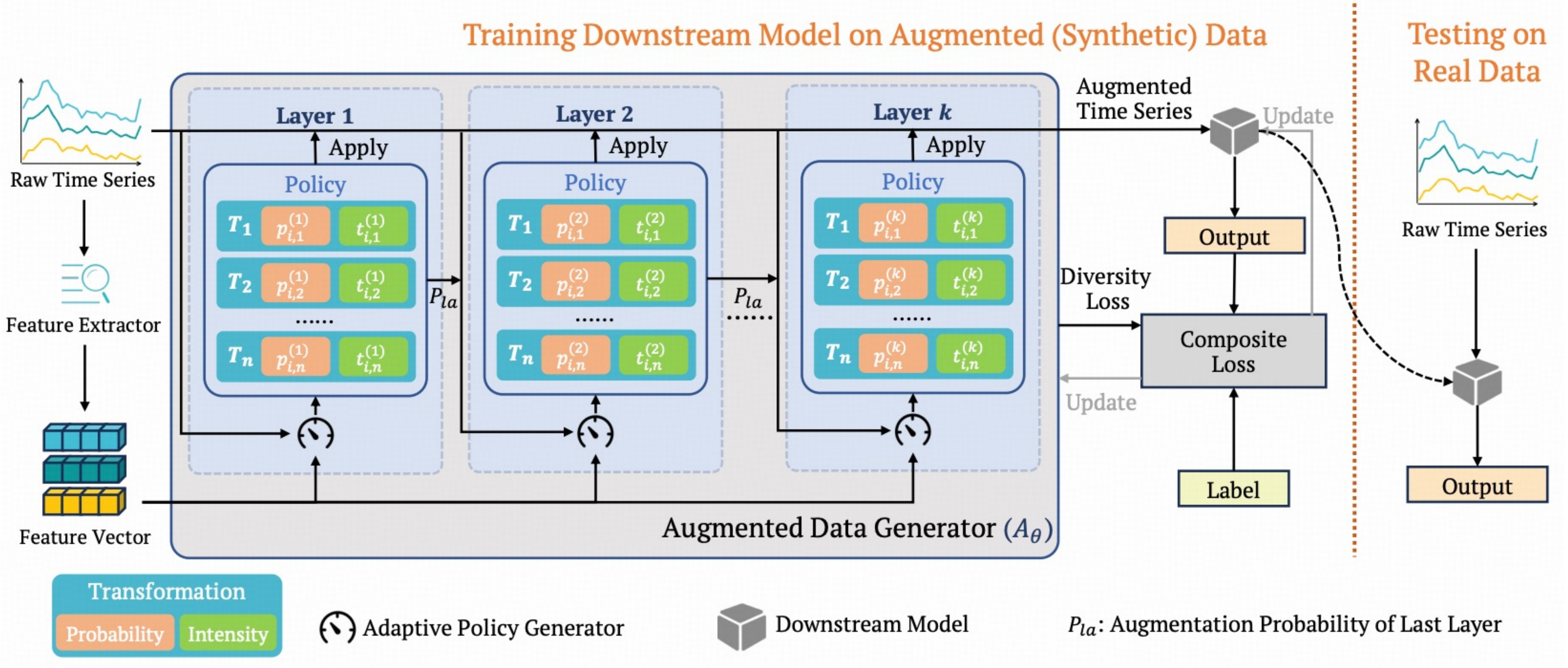
A general-purpose, multi-task automated data augmentation framework for time series

Downstream Tasks



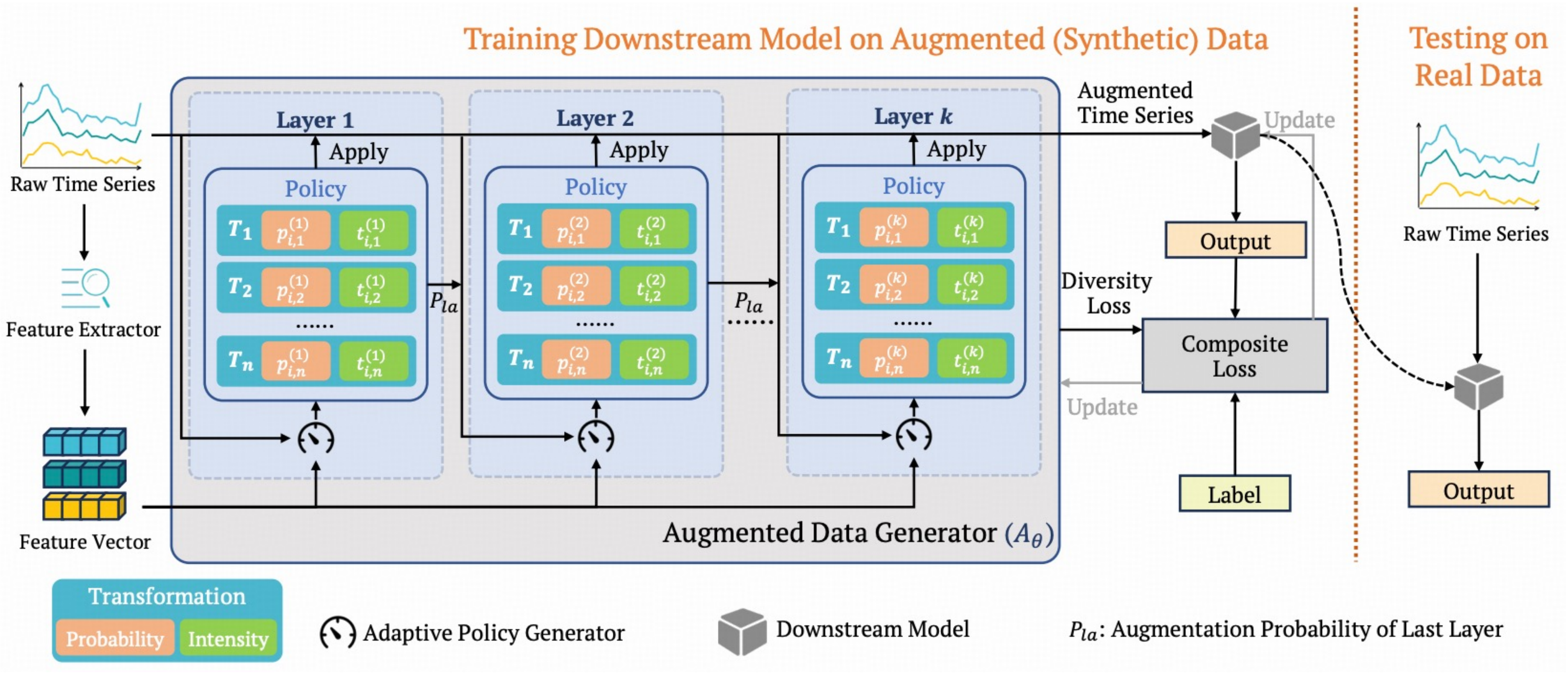
Methodology

- A time series *feature-aware* augmented data generator (denoted as $\mathcal{A}_\theta$) is composed of *multiple stacked Augmentation Layers* $\mathcal{A}_{\theta_k}^{(k)}$, each of which is responsible for selecting and applying one of the available transformations in the set $\mathcal{T} = \{T_1, T_2, \cdots, T_n\}$. The k-th augmentation layer generates *an augmentation policy* consisting of (i) a series of *probability* $p_{i,j}^{(k)}$ indicating the likelihood of choosing transformation T_j and (ii) a series of *intensity* $t_{i,j}^{(k)}$ to apply a chosen transformation.



Methodology

- By stacking these augmentation layers, the framework can explore a variety of transformation sequences, allowing for more diverse and potentially useful augmented data. The final output augmented time series is used to train a single downstream model in a *single-stage*, *end-to-end manner*, with a *composite loss* to update the parameters in *the augmented data generator* together with *the downstream model*.



Implementation

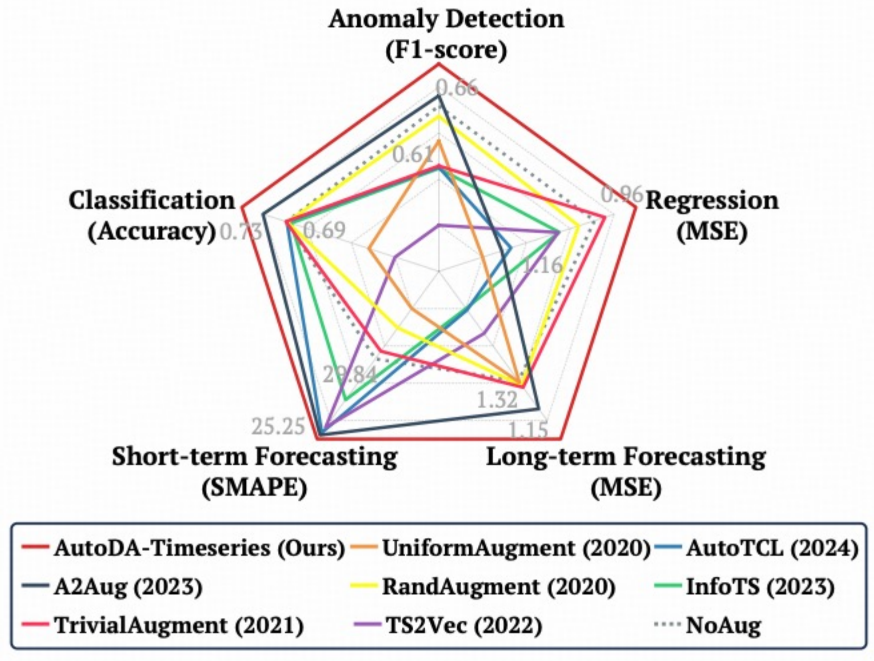
- We evaluate on representative downstream models and extend the scope by incorporating both *classical* and *advanced architectures*, covering *convolutional*, *recurrent*, *Transformer-based*, and *generative* paradigms

Table 1: Summary of benchmarks, evaluation metrics, and representative downstream models.

Tasks	Benchmarks	Metrics	Downstream Models
Classification	UEA (26 subsets)	Accuracy	TCN, ROCKET
Forecasting	**Long-term**: ETT (4 subsets), Exchange, Weather	MSE, MAE	RNN, Autoformer
	Short-term: M4 (6 subsets)	SMAPE, MASE, OWA	
Regression	UEA & UCR (6 subsets)	MSE, MAE	CNN, MLP
Anomaly Detection	MSL, SMAP, SMD	Precision, Recall, F1-score	UNet, VAE

Overall Performance

- NoAug: as a control group, which does not apply any augmentation
- Representation Learning:
  - TS2Vec: AAAI’ 2022
  - InfoTS: AAAI’ 2023
  - AutoTCL: ICLR’ 2024
- Automated Data Augmentation:
  - RandAugment: CVPR’ 2020
  - UniformAugment: arXiv’ 2020
  - TrivialAugment: ICCV’ 2021
  - A2Aug: CVPR’ 2023



Adaptive Augmentation Policy Visualization

- The results reveal a *clear layer-wise differences*. Layer 0 rapidly converges to a few operators (e.g., Raw augmentation), reflecting deterministic exploitation, while Layer 1 and Layer 2 maintain higher entropy and more diverse strategies
- Overall, lower layers *stabilize* the augmentation policies, whereas upper layer remain *adaptive*, forming *a complementary balance between stability and diversity*

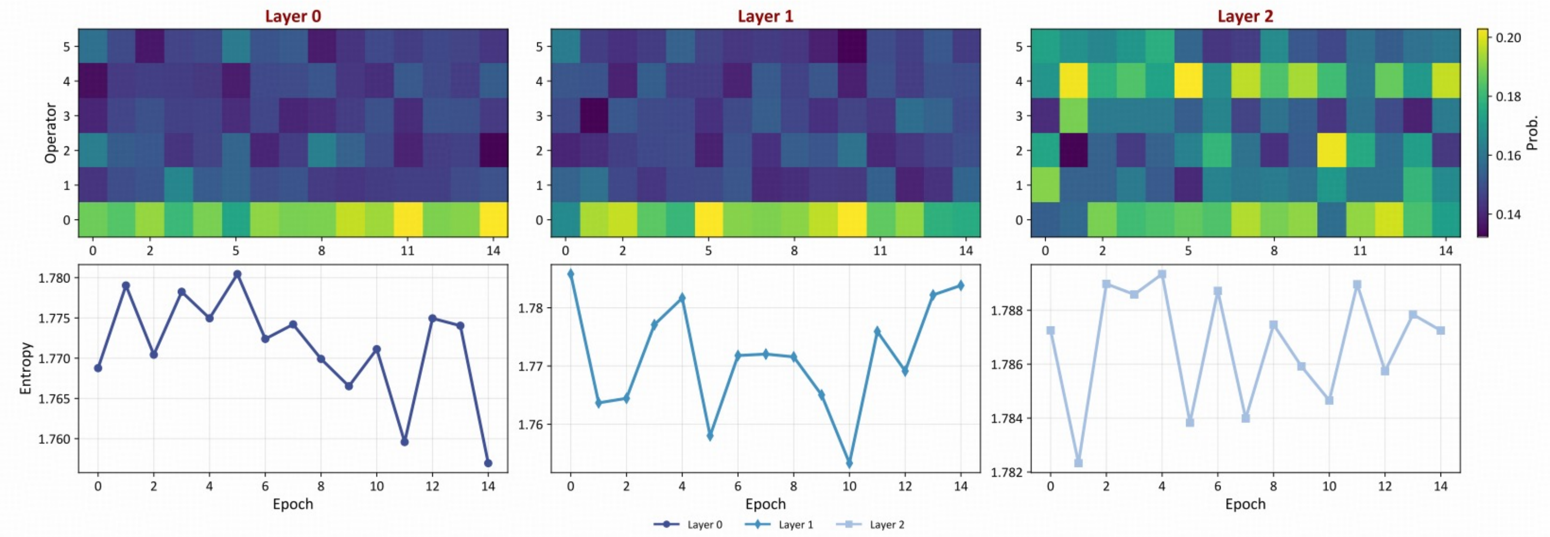


Figure 5: Adaptive augmentation policy. Top: operator distribution over training epochs. Bottom: entropy dynamics showing convergence in lower layers and diversity in higher layers.

Generalization Across Datasets

- We train the downstream model together with augmentation policies on *ETTh1* and directly evaluate the trained model on *ETTh2* and *ETTm2*
- AutoDA-Timeseries not only enhances performance within a single dataset but also exhibits strong potential for *cross-dataset generalization*, validating its generalization and applicability in real-world scenarios

Table 13: Generalization performance of AutoDA-Timeseries on RNN under cross-dataset transfer settings.

Settings		NoAug		UniformAugment		AutoDA-Timeseries	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1 → ETTh2	96	0.4761	0.4602	0.6486	0.5577	0.4431	0.4409
	192	0.5418	0.4944	0.7172	0.5882	0.5146	0.4788
	336	0.5566	0.5116	0.7366	0.6091	0.5374	0.4996
	720	0.5416	0.5115	0.7727	0.6315	0.5354	0.5052
	Avg	0.5290	0.4944	0.7188	0.5966	**0.5076**	**0.4811**
ETTh1 → ETTm2	96	0.8668	0.6702	1.1370	0.7578	0.8663	0.6724
	192	1.0198	0.7361	1.3273	0.8305	1.0087	0.7338
	336	1.2997	0.8369	1.6327	0.9313	1.2791	0.8317
	720	1.7294	0.9623	2.1243	1.0623	1.7105	0.9576
	Avg	1.2289	0.8014	1.5553	0.8955	**1.2162**	**0.7989**

Robustness under Limited Training Samples

- Evaluated model robustness with *10% - 100%* training data
- AutoDA-Timeseries consistently outperforms NoAug across all data ratios
- The advantage is more significant in *low-data regimes (10% - 50%)*, showing strong robustness

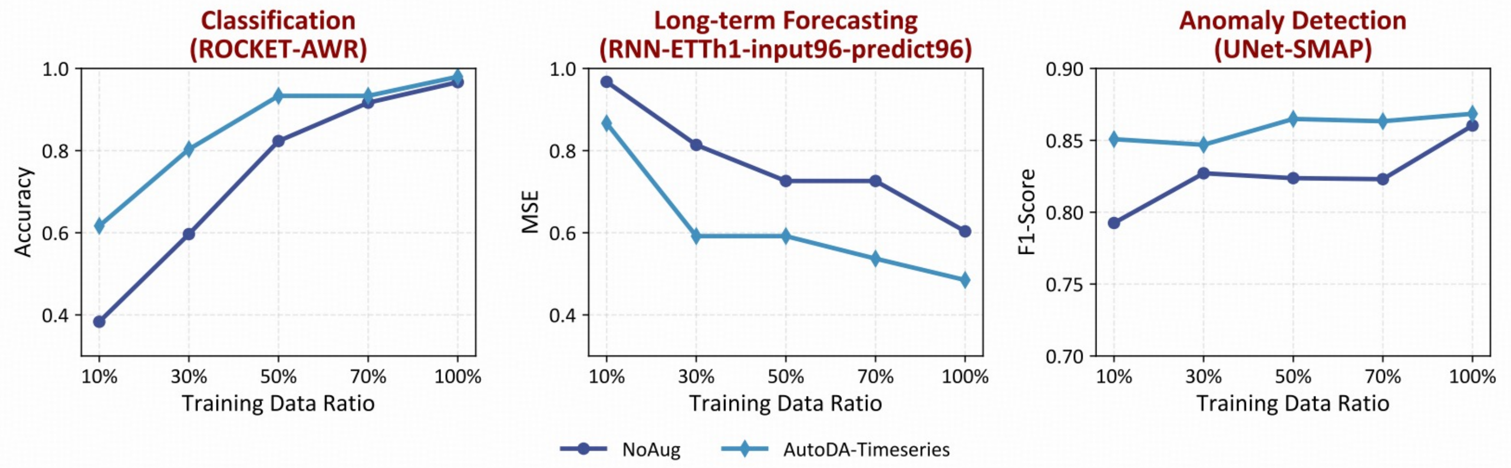
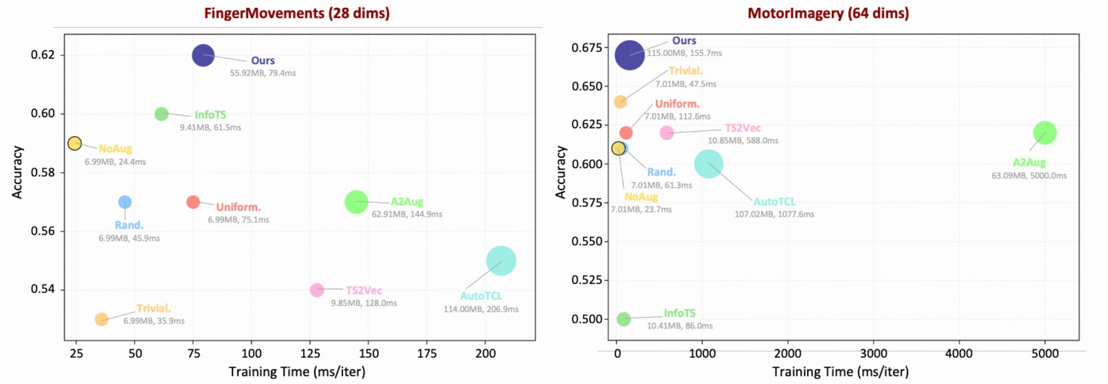


Figure 10: Performance comparison between NoAug and AutoDA-Timeseries under varying training data ratios (10%, 30%, 50%, 70%, 100%) across three representative tasks: classification (Accuracy), long-term forecasting (MSE), and anomaly detection (F1-score).

Model Efficiency

- Three factors: ***parameter size***, ***training time*** (ms/iter), and ***accuracy***
- Outperforms complex baselines (A2Aug, AutoTCL) while keeping ***lower computational cost***
- Compared with simple baselines (Rand., Trivial., Uniform.), it achieves significantly ***higher accuracy***



(a) Efficiency comparison on FingerMovements dataset (28 dimensions) (b) Efficiency comparison on MotorImagery dataset (64 dimensions)

Figure 9: Model efficiency comparison across datasets. The x-axis represents training time per iteration, the y-axis shows accuracy, and the bubble size reflects model parameter size.

Summary

Domain Data

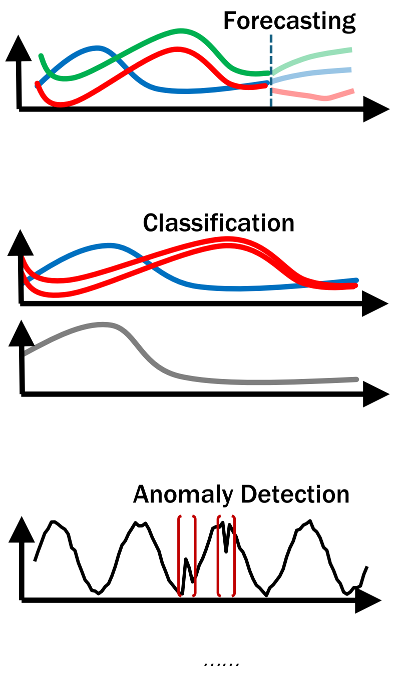
- Energy
- Healthcare
- AIOps
- Weather
- Economy
- ......

AutoDA-Timeseries

- Feature-Driven
- Adaptive
- Plug-and-Play
- Physically Meaningful
- Efficient
- Interpretable

A general-purpose, multi-task automated data augmentation framework for time series

Downstream Tasks



实际落地价值

01 全流程部署测试

在阿里云仿真环境进行了部署测试，打通了各类故障预警、定位、诊断的数据增强模型及工具实时拉取数据并进行推理的流程。

02 集成验证

相关诊断模型及工具已经交付阿里云，正在进行集成验证。

评估指标	本方案结果	对比现有最优方案提升幅度
A@1	0.776	+64%
A@3	0.898	+28%
A@5	0.959	+21%

模型	AIOps 挑战赛数据集			
	A@1	A@3	A@5	Avg@5
Ours	**0.776**	**0.898**	**0.959**	**0.878**
DéjàVu	0.473	0.701	0.793	0.670
Eadro	0.310	0.446	0.484	0.413
DiagFusion	0.333	0.500	0.648	0.493
MicroHECL	0.091	0.232	0.386	0.236
MicroRank	0.144	0.218	0.259	0.209
AutoMAP	0.279	0.574	0.729	0.531
TraceRCA	0.243	0.310	0.338	0.302
Microscope	0.074	0.113	0.227	0.127
RCD	0.095	0.124	0.174	0.128

本工作得到国家重点研发计划（No. 2024YFB4505903）的资助

Tsinghua University

ICLR 2026

AutoDA-Timeseries: Automated Data Augmentation for Time Series

Zijun Dou[1], Zhenhe Yao, Zhe Xie, Xidao Wen, Tong Xiao, Dan Pei

窦梓峻
北京 海淀

扫一扫上面的二维码图案，加我为朋友。

1. Presenter. Email: douzj25@mails.tsinghua.edu.cn
Tsinghua NetMan Lab

Research Interests:
Agent, Time Series Analysis