



Understanding and Mitigating Structural Forgetting

in Fine-Tuned Time Series Foundation Models

Zeyu Shi, Yirong Xue, Xin Xue, Ziming Wang, Jintao Wu, Shiqi Gao, Haoyi Zhou
SKLCCSE, School of Computer Science and Engineering, Beihang University
School of Software, Beihang University
Xiaomi Corporation

Problem / Diagnosis / Mechanism / NeST / Results

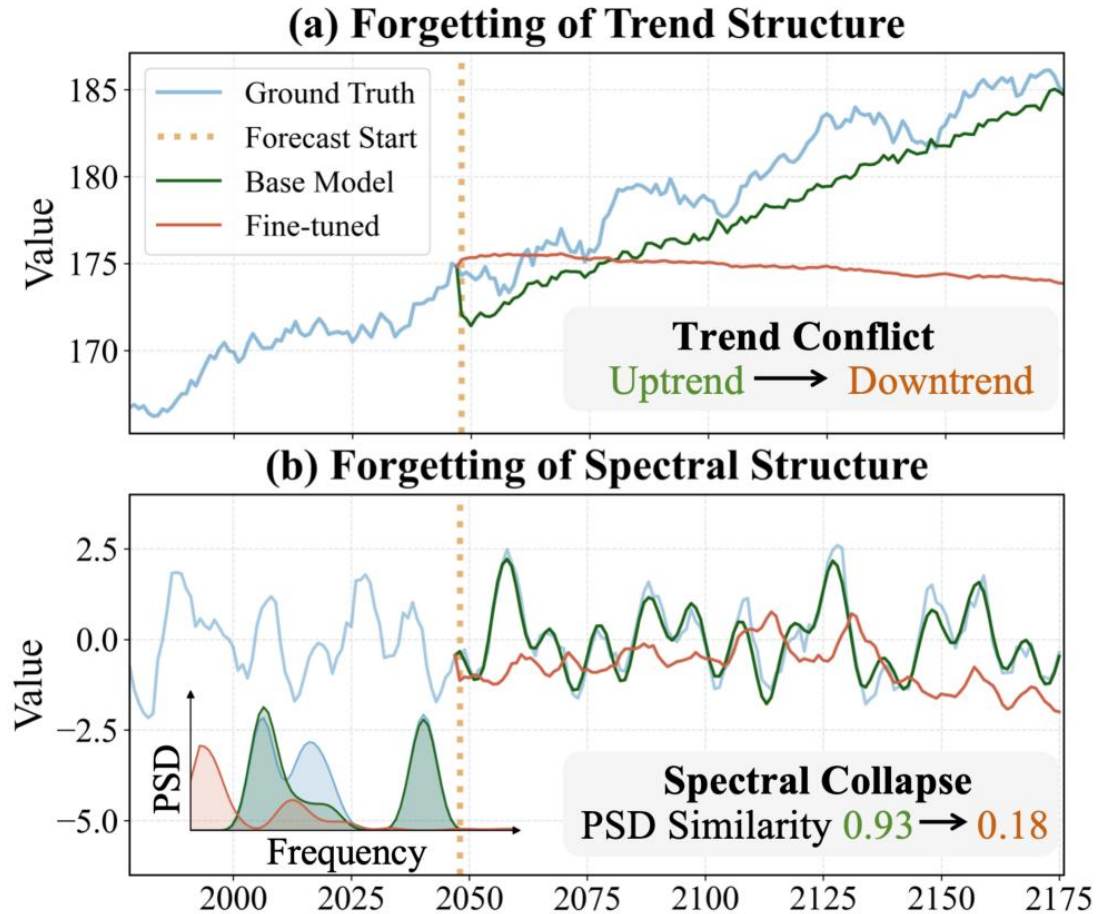
Structural forgetting after fine-tuning

Pre-training learns universal temporal patterns.

Fine-tuning can improve task fit while erasing those patterns.

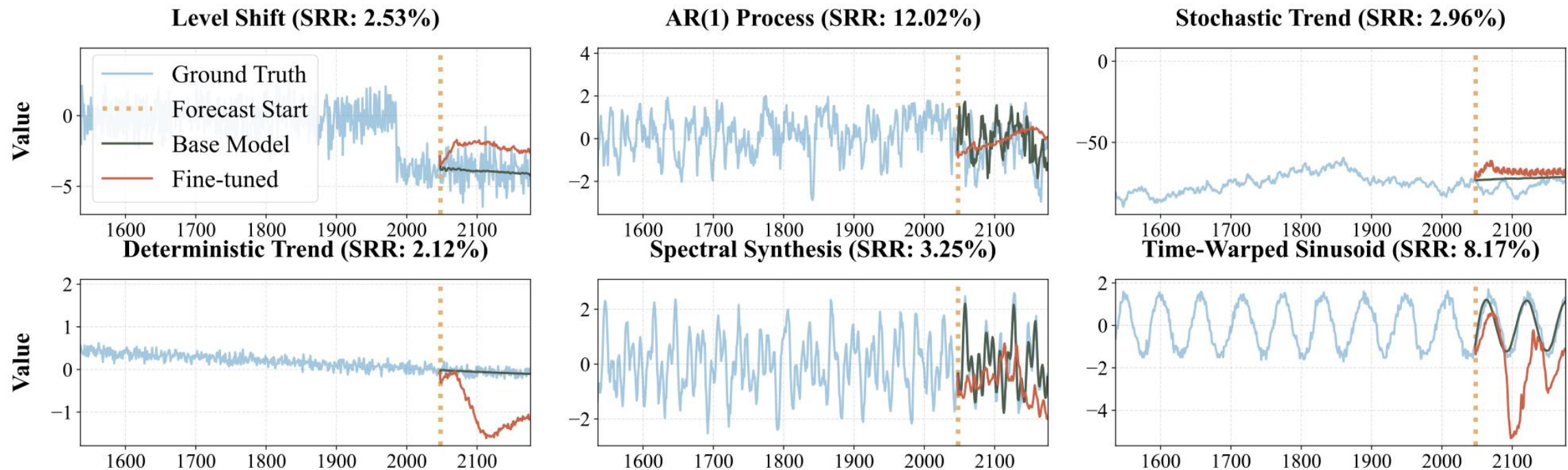
Why it matters

Trend-based decisions and periodic scheduling depend on structure.



Synthetic probes isolate structural ability

Six evaluated patterns across shifts, temporal dynamics and periodicity



Trend: slope error. Periodicity: PSD similarity. Shifts: reversion bias.

Task gains coexist with low structural retention

Structural Retention Rate (SRR)

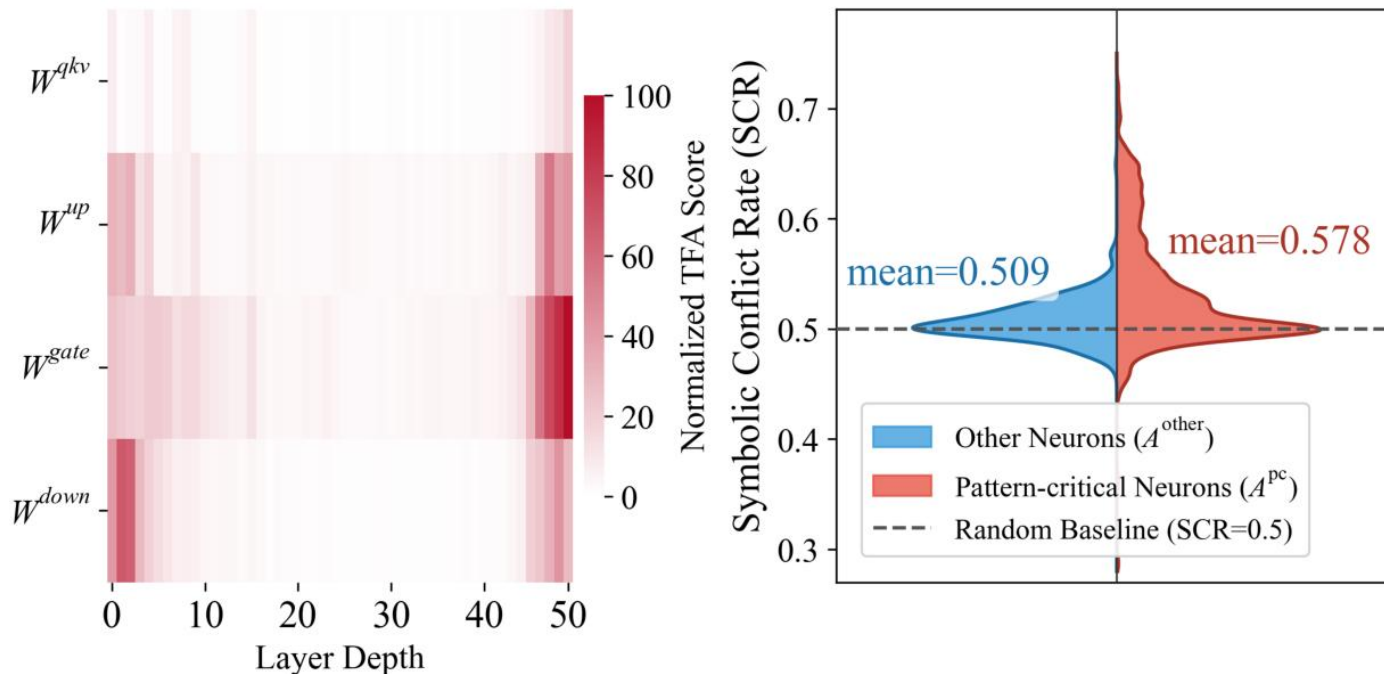
SRR = 1: preserved SRR < 1: forgetting

Higher-is-better metric: post / pre. Lower-is-better metric: pre / post.

Fine-tuning dataset	MSE gain (%)	Mean SRR (%)
Crude Oil	+7.11	29.32
Exchange Rate	+17.35	30.80
Gold	+18.80	32.34
MSFT	+12.78	26.04
SPX500	+8.99	47.22

Mean SRR across these tasks: 33.14%

Task gradients conflict with pattern-critical neurons



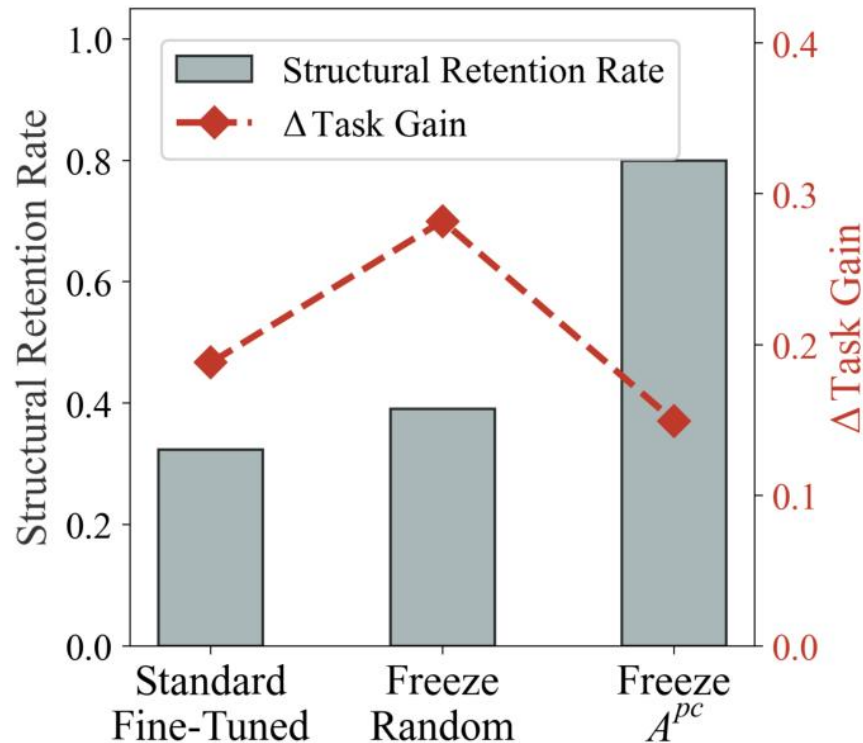
Localized structural importance

Temporal error +
DFT-amplitude error

TFA-Fisher scores identify
pattern-critical neurons.

**Mean sign conflict ratio
0.578 critical vs 0.509 other**

Much of the structural damage is avoidable



Freeze pattern-critical neurons

Structural retention
32.3% to 79.9%

MSE gain
18.8% to 14.9%

The remaining 3.9-point gap motivates restoration with task compensation.

NeST restores structure and compensates task behavior

1 Locate

Use TFA-Fisher importance and layer displacement to select pattern-critical parameters.

2 Restore

Revert selected parameters to their pre-trained values.

$$\mathbf{W}_{\text{res}} = \mathbf{W}_{\text{ft}} \odot (1 - \mathbf{M}) + \mathbf{W}_{\text{pre}} \odot \mathbf{M}$$

3 Compensate

Adjust the remaining parameters with OBS-based compensation to preserve downstream activations.

Post-hoc correction with small calibration sets, without retraining

NeST recovers structure with near-SFT task gains

Method	MSE gain (%)	Mean SRR (%)
Vanilla SFT	18.80	32.34
LoRA	9.33	59.29
Rehearsal	12.96	60.42
EWC	11.71	50.99
LwF	17.45	51.84
Model Merging	6.25	51.64
NeST	18.23	84.39

84.39%

Structural
retention

18.23%

MSE gain
SFT: 18.80%

NeST mean SRR: 82.36% on MSFT and 91.23% on SPX500

Both targeted restoration and compensation matter

Restoration / compensation	MSE gain (%)	SRR (%)
Random selection + compensation	17.02	47.11
Targeted restoration only	11.70	86.24
Full NeST	18.23	84.39

Targeted restoration recovers structural ability.

Compensation recovers much of the task-side performance.

Restored structure reduces persistent false alarms

Lost level-shift awareness prevents the model from re-anchoring after an anomaly, leaving a tail of false alarms.

Model	AUC ↑	FAR ↓	PFAR ↓
Base model	61.02	10.87	21.32
Fine-tuned	48.22	13.25	25.28
NeST	59.35	10.64	19.06

PFAR measures false alarms after the true anomaly ends.

Structural preservation is a distinct adaptation goal

Problem

Task accuracy can improve while universal temporal patterns disappear.

Mechanism

Task updates disrupt a sparse set of pattern-critical neurons.

Solution

NeST restores those parameters and compensates the rest.

Remaining challenge: stochastic-trend retention is 43.47% on Gold.