ICML 2026 WORKSHOP: FORECASTING AS A NEW FRONTIER OF INTELLIGENCE (AI FORECASTING WORKSHOP)

Semantics-Enhanced Retrieval-Augmented Time Series Forecasting

Shiqiao Zhou[1], Zipeng Wu[1], Holger Schöner[2], Edouard Fouché[2], IAG Wilson[1], Shuo Wang[1]

[1]University of Birmingham [2]Siemens AG

Contact: Shiqiao Zhou | sxz363@student.bham.ac.uk

UNIVERSITY OF BIRMINGHAM

SIEMENS

1. MOTIVATION

Why semantics?

Numerically similar histories are not always the best guides under non-stationarity.

- Useful future patterns can repeat across different local shapes or scales.
- Pure TS retrieval may miss season, trend, and volatility matches.
- External text resources add cost and reduce scalability.

What SERAF adds

S — **Semantic retrieval**
Search by generated temporal attributes.

T — **No external text**
Descriptions come from the time series itself.

G — **Adaptive fusion**
Learns how much retrieval should guide the forecast.

Self-generated semantic view

Multivariate input segment

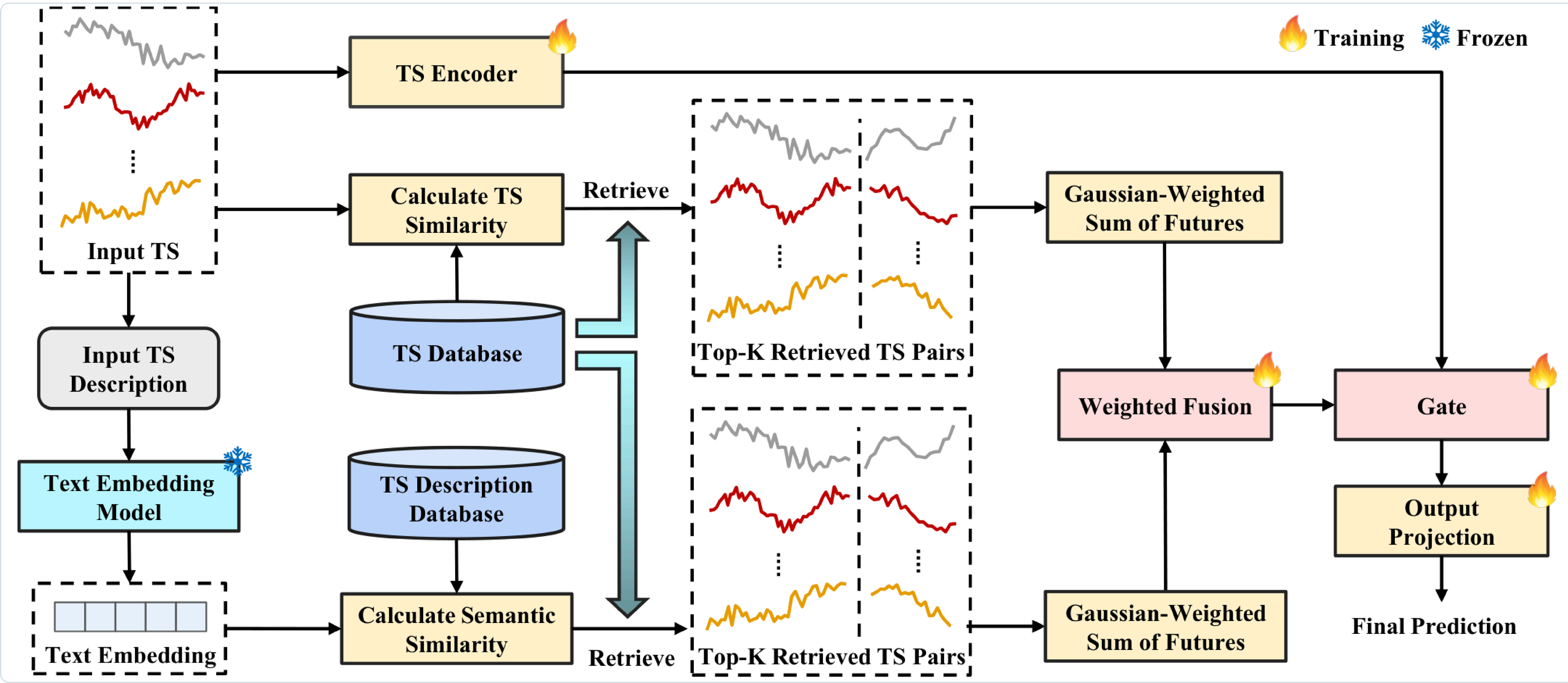
Complete description template

Time period: from Thursday, January 4, 00:00 to Friday, February 2, 23:00;
Season: Winter;
Main trend: upward;
Main volatility: high.

2. METHOD OVERVIEW

SERAF retrieves futures twice, then learns how to fuse them.

Historical futures are retrieved from both raw time-series similarity and self-generated semantic descriptions.



Temporal path
Find neighbors by raw shape similarity.

Semantic path
Retrieve by trend, season, and volatility.

Weighted futures
Gaussian weights aggregate retrieved futures.

Gate
Fuse retrieval with the backbone forecast.

1 **Describe**
Generate compact text for each segment.

2 **Retrieve**
Find top-K futures through TS and text similarity.

3 **Fuse**
Gate retrieved futures with the backbone prediction.

TAKE-HOME **SERAF makes retrieval more than shape matching.**
The semantic channel expands retrieval beyond raw shape matching by searching historical futures through generated temporal descriptions.

3. EXPERIMENTAL RESULTS — ACCURACY GAINS ACROSS DATASETS

SETTING Input length 720, seven multivariate datasets, four forecast horizons (96, 192, 336, 720), and RAFT as the retrieval baseline.

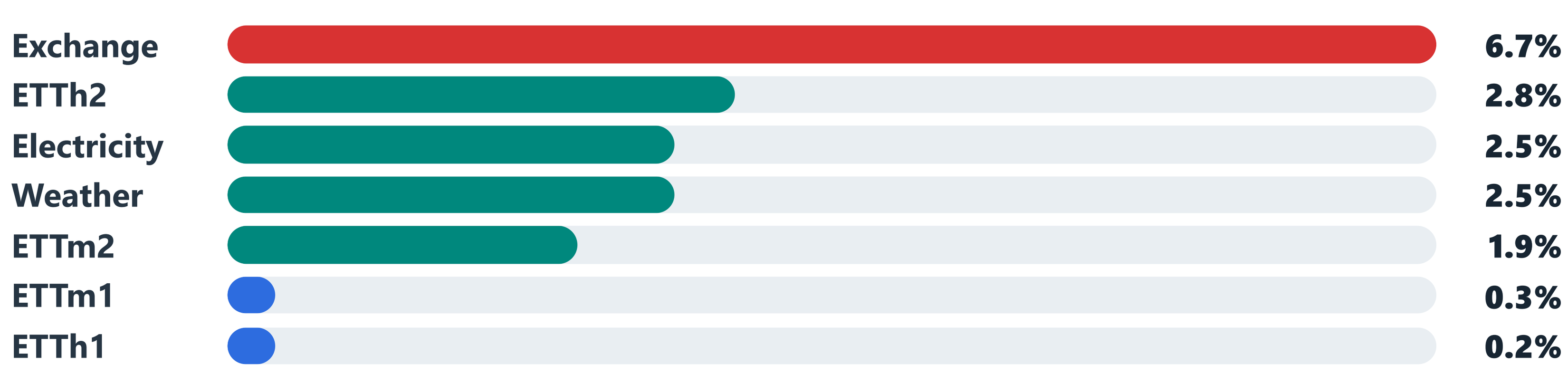
5/7 datasets where SERAF has best MSE

0.088M SERAF backbone parameters

-2.56% average MSE vs RAFT

-1.42% average MAE vs RAFT

MSE REDUCTION VS RAFT



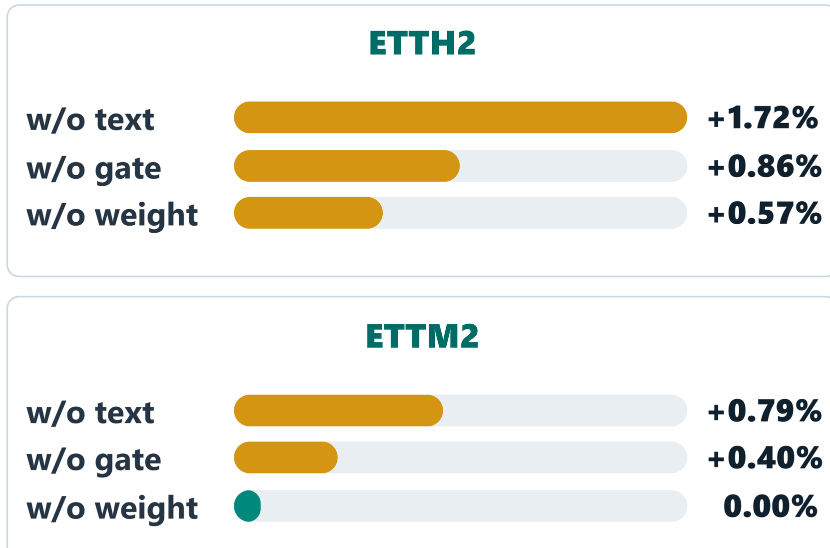
ACTUAL MSE / MAE RESULTS VS RAFT

Dataset	SERAF MSE / MAE	RAFT baseline MSE / MAE	MSE gain
Exchange	0.419 / 0.430	0.449 / 0.444	-6.7%
ETTh2	0.348 / 0.401	0.358 / 0.408	-2.8%
Electricity	0.156 / 0.257	0.160 / 0.259	-2.5%
Weather	0.235 / 0.280	0.241 / 0.287	-2.5%
ETTm2	0.252 / 0.317	0.257 / 0.321	-1.9%
ETTm1	0.346 / 0.377	0.347 / 0.377	-0.3%
ETTh1	0.417 / 0.432	0.418 / 0.434	-0.2%

Why the improvement matters
Semantic retrieval makes historical futures searchable by meaning, not only by local numeric shape. This is useful under scale shifts, trend changes, and weak periodicity.

4. ANALYSIS — WHICH PIECES MATTER?

ABLATION EVIDENCE



Relative MSE increase vs Full SERAF. Text removal is the largest drop on both ETTh2 and ETTm2.

5. EFFICIENCY

INFERENCE TIME (S) **PARAMETERS (M)**

0.34 MiB SERAF memory

3.1x faster than RAFT

Smallest forecasting backbone

6. INTERPRETATION

Best fit
Semantic retrieval helps when trend, season, or volatility carries future-relevant structure.

Hard cases
Noisier Weather and Exchange dynamics may need richer semantic templates.

Why ETT gains
Repeated operating regimes make coarse descriptions good retrieval keys.

Practical use
Use semantics as lightweight retrieval metadata, without external text sources.

Shape + Meaning
Raw similarity finds local matches, while semantic retrieval preserves future-relevant temporal attributes.

7. NEXT STEPS

C — **Richer templates**
Event-like or weakly periodic dynamics.

K — **Retrieval sensitivity**
Top-K and bandwidth analysis.

D — **Dataset diagnosis**
When semantic retrieval helps most.