LATF-AD: LEARNING LOCALLY ADAPTIVE TIME-FREQUENCY NORMALITY FOR TIME SERIES ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Time series anomaly detection (TSAD) is essential for reliable monitoring of complex systems. Although frequency-aware modeling can reveal latent anomalies, global spectral representations often obscure localized transients, while a single fixed transformation inadequately captures heterogeneous temporal dynamics across varying resolutions. To address these limitations, we propose LATF-AD, a novel framework for locally adaptive time-frequency normality modeling. LATF-AD constructs an ordered bank of heterogeneous time-frequency views and learns content-dependent, locally adaptive routing to accommodate dynamic spectral patterns via a weighted consensus. It further feeds the consensus into dual prediction and reconstruction heads, while incorporating a prototype-guided raw-domain branch to modulate anomaly scoring against time-frequency artifacts. Extensive experiments on the comprehensive TSB-AD Benchmark demonstrate that LATF-AD consistently achieves state-of-the-art performance across both point-wise and range-based metrics.

1 INTRODUCTION

Time Series Anomaly Detection (TSAD) is a fundamental task for reliable monitoring of complex systems, with broad applications in industrial control, cloud services, cyber-physical systems, healthcare, and scientific sensing. Given a series of observations, the goal is to identify timestamps whose behaviors deviate from normal temporal patterns. In many practical scenarios, anomalous labels are scarce or unavailable during training, making unsupervised normality modeling the dominant paradigm. Recent deep models have made substantial progress by reconstructing input series, forecasting future observations, modeling long-range dependencies with attention, learning contrastive representations, or building patch-level embeddings (Su et al., 2019; Xu et al., 2022; Tuli et al., 2022; Yang et al., 2023; Park & Kang, 2026). However, real-world anomalies remain difficult to detect because they are often local, non-stationary, and expressed through changes in dynamic structures rather than large point-wise deviations.

Recent studies have established the value of incorporating frequency-domain information into TSAD (Zhang et al., 2022; Nam et al., 2024; Fang et al., 2024; Wu et al., 2025; Ekambaram et al., 2026). However, existing methods typically rely on fixed transformation parameters or static time-frequency representations, failing to construct a bank of heterogeneous time-frequency views across varying temporal and spectral resolutions. Consequently, they struggle to simultaneously isolate short-lived transient bursts and subtle long-term rhythmic anomalies, ultimately degrading overall anomaly detection accuracy (Cohen, 1994).

To address these limitations, we introduce LATF-AD, a novel framework that learns **L**ocally **A**daptive **T**ime-**F**requency normality for Time Series **A**nomaly **D**etection. To resolve the resolution trade-off, LATF-AD explicitly constructs an ordered multi-view time-frequency bank encompassing short- and long-window Short Time Fourier Transform (STFT), continuous wavelet transform (CWT), and the Wigner–Ville distribution (WVD). Nevertheless, effectively integrating these heterogeneous views poses a non-trivial challenge: determining their relative importance at every timestamp is inherently dynamic. Because different anomalies exhibit distinct spectral signatures—some manifesting as localized high-frequency transients while others disrupt stationary low-frequency

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

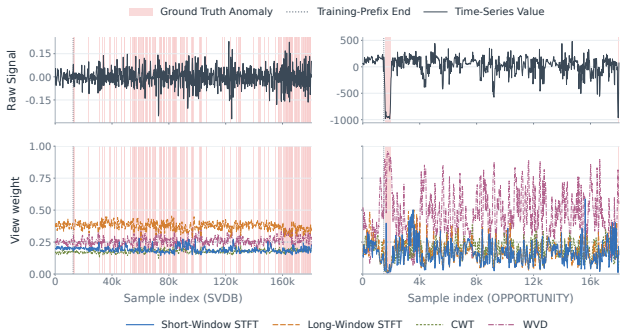

Figure 1: Learned local view weights on representative SVDB and OPPORTUNITY series. Top: raw signals with ground-truth anomaly bands (shaded) and training prefixes (dotted). Bottom: dynamic view routing weights. The dominant view shifts both across datasets and over time.

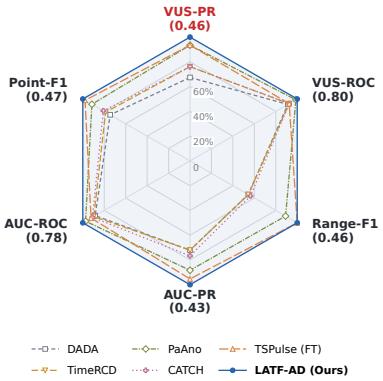

Figure 2: Anomaly detection performance of LATF-AD and recent methods on TSB-AD-M. LATF-AD (ours) achieves the best performance across six evaluation metrics.

patterns—a static or uniform aggregation fails to capture timestamp-specific view reliability. To overcome this hurdle, we introduce a local adaptive routing mechanism. Guided by temporal dynamics, energy distributions, and cross-view discrepancy statistics, the router dynamically assigns localized reliability weights across representations, yielding a localized consensus representation fed into dual reconstruction and prediction heads.

Furthermore, time-frequency consensus alone can remain ambiguous when localized transients induce artifacts across views. LATF-AD remedies this by incorporating a complementary, self-supervised raw patch normality branch. By clustering time-domain representations against a prototype memory bank, this branch provides an independent local sanity check that collaborates with the time-frequency evidence via a robust dual-calibration scoring function.

As shown in Figure 1, the optimal time-frequency perspective dynamically shifts across temporal regimes and domains: while long-window STFT provides stable dominance in capturing rhythmic patterns on SVDB, WVD and CWT exhibit rapid, timestamp-sensitive weight spikes around transient anomalies on OPPORTUNITY, confirming that a globally fixed fusion is sub-optimal. Moreover, as shown in Figure 2, LATF-AD substantially outperforms competitive recent baselines across all six evaluation metrics on the comprehensive TSB-AD-M Benchmark (Liu & Paparrizos, 2024), achieving leading scores in both point-wise and range-based evaluations, which corroborates the effectiveness of our locally adaptive multi-view modeling.

The main contributions of this work are summarized as follows:

- We introduce LATF-AD, a novel framework that addresses the fundamental time-frequency resolution trade-off by explicitly constructing an ordered multi-view bank, capturing both localized transients and non-stationary rhythmic dynamics.
- We devise a local adaptive router that dynamically allocates timestamp-specific reliability weights across different time-frequency views based on temporal and spectral dynamics. To resolve residual time-frequency ambiguities, we further incorporate a self-supervised raw-domain patch branch with a prototype memory bank, unifying the evidence via a robust dual-calibration scoring mechanism.
- We evaluate LATF-AD on both univariate and multivariate TSAD settings using the TSB-AD Benchmark (Liu & Paparrizos, 2024), comparing against representative baselines and demonstrating both state-of-the-art performance and the effectiveness of locally adaptive time-frequency modeling.

2 Related Work

2.1 Deep TSAD Paradigms

Unsupervised TSAD primarily models temporal normality directly from raw observations. Classical methods capture sequential dynamics via stochastic latent spaces (Su et al., 2019), self-conditioned adversarial reconstruction (Tuli et al., 2022), or dual-attention contrastive representations (Yang et al., 2023). More recently, patch-based architectures like PaAno model patch-level normality patterns (Park & Kang, 2026). While these paradigms highlight the mutual benefits of reconstruction, forecasting, and local patch comparisons, they operate predominantly in the raw time domain or a single latent space, struggling to decouple non-stationary dynamics from subtle rhythmic anomalies. LATF-AD inherits the synergy of dual prediction–reconstruction heads and patch-level modeling, but advances them by operating upon a dynamically routed time-frequency consensus complemented by a raw-domain prototype memory.

2.2 Frequency and Time-Frequency TSAD

Incorporating frequency representations effectively isolates periodic shifts and harmonic disruptions obscure in the time domain. TFAD couples trend-seasonal decomposition with spectral filtering (Zhang et al., 2022), while DualTF and TF-MAE employ dual-domain or masked time-frequency reconstruction objectives to expose spectral-temporal anomalies (Nam et al., 2024; Fang et al., 2024). CATCH partitions the frequency spectrum into patches and models inter-channel relationships via masked attention (Wu et al., 2025). Recently, TSPulse triangulates anomaly evidence from time reconstruction, frequency reconstruction, and short-horizon forecasting (Ekambaram et al., 2026).

Nevertheless, these methods remain constrained by the fundamental Heisenberg–Gabor uncertainty principle (Cohen, 1994): a single transformation or fixed window cannot simultaneously provide fine temporal and spectral localization. Narrow windows resolve transient bursts at the cost of spectral resolution, whereas wide windows blur localized changes. Although approaches like TSPulse combine dual-space forecasting, they apply a shared patching scheme without resolving timestamp-specific view reliability over diverse transformations. In contrast, LATF-AD explicitly constructs an ordered multi-view bank spanning short-window STFT, long-window STFT, multiscale CWT, and quadratic WVD, systematically breaking the single-resolution barrier.

2.3 Multi-View Fusion and Adaptive Routing

Mixture-of-Experts and multi-view architectures aggregate specialized representations via gating networks, sparse token dispatching (Shazeer et al., 2017; Fedus et al., 2022), or late fusion. In TSAD, dynamic routing is critical yet difficult: view informativeness varies across regimes, unsupervised settings lack selection labels, and transformations like CWT/WVD introduce energy artifacts. LATF-AD overcomes this via a lightweight content-aware router guided by unsupervised salience and corruption rejection. The resulting consensus $\mathbf{z}_\ell$ drives dual normality heads, while an auxiliary raw patch branch provides artifact-free temporal verification, unified via a robust multiplicative-additive scoring function.

3 Problem Formulation

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T] \in \mathbb{R}^{T \times C}$ denote a time series of length T with C variables, where $\mathbf{x}_t \in \mathbb{R}^C$ represents the observation vector at timestamp t. We assume a training segment $\mathbf{X}^{\mathrm{tr}} \in \mathbb{R}^{T_{\mathrm{tr}} \times C}$ containing only normal observations and an unlabelled test segment $\mathbf{X}^{\mathrm{te}} \in \mathbb{R}^{T_{\mathrm{te}} \times C}$ that may contain anomalies. The objective of TSAD is to learn an anomaly scoring function using only $\mathbf{X}^{\mathrm{tr}}$ without anomaly supervision. During inference, an anomaly score $S_t \in \mathbb{R}_+$ is assigned to each timestamp t in $\mathbf{X}^{\mathrm{te}}$, where a larger score indicates a higher likelihood of an anomaly.

In this paper, we cast the core challenge of TSAD as learning a locally adaptive multi-view time-frequency fusion model. Formally, given a bank of V deterministic heterogeneous transformations $\mathcal{T} = \{\mathcal{T}_v\}_{v=1}^V$ spanning complementary temporal and spectral resolutions, each segment $\mathbf{X}$ is transformed into multi-view representations $\mathbf{M} = \{\mathbf{M}_v\}_{v=1}^V$, where $\mathbf{M}_v = \mathcal{T}_v(\mathbf{X}) \in \mathbb{R}^{D_v \times L}$ captures view-specific dynamics with L temporal frames and feature dimension D_v.

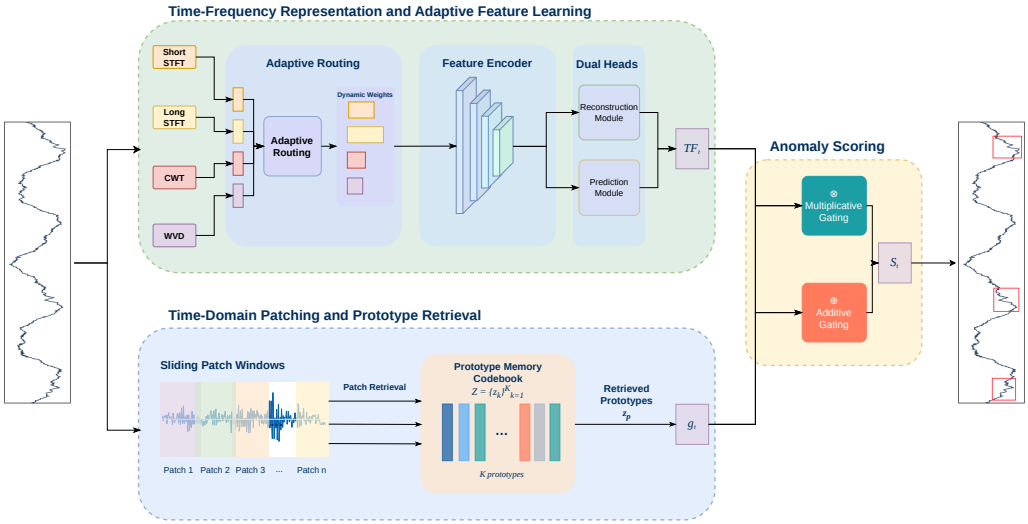

Figure 3: Architecture of LATF-AD, which combines an adaptive multi-view time-frequency stream and a prototype-anchored time-domain stream to output collaborative anomaly scores.

Under this multi-view paradigm, our problem is formulated as simultaneously learning:

1. **A local adaptive router** $\mathcal{R}_\theta : \mathbf{M} \to \Delta_\ell^{V-1}$, which estimates time-varying reliability weights $\mathbf{w}_\ell = [w_{1,\ell}, \ldots, w_{V,\ell}]^\top$ across all views at each frame ℓ, subject to $\sum_{v=1}^{V} w_{v,\ell} = 1$ and $w_{v,\ell} \geq 0$, to construct a weighted consensus representation:

$$\mathbf{z}_\ell = \sum_{v=1}^{V} w_{v,\ell} \mathbf{h}_{v,\ell}, \quad \text{where } \mathbf{h}_{v,\ell} = \mathcal{E}_{\phi_v}(\mathbf{M}_{v,:,:,\ell});$$

2. **A dual-normality estimation head** $\mathcal{D}_\psi(\mathbf{z}_\ell)$ operating on the routed consensus $\mathbf{z}_\ell$ to capture multi-view reconstruction and temporal transition dynamics of normal patterns;
3. **A time-domain verification branch** $\mathcal{V}_\omega(\mathbf{X})$ to mitigate spectral artifacts and ambiguity.

Given the normal training set $\mathbf{X}^{\text{tr}}$, the goal is to optimize the parameter set $\Theta = \{\theta, \{\phi_v\}_{v=1}^{V}, \psi, \omega\}$:

$$\Theta^* = \arg\min_{\Theta} \mathcal{L}\left(\mathcal{R}_\theta, \{\mathcal{E}_{\phi_v}\}_{v=1}^{V}, \mathcal{D}_\psi, \mathcal{V}_\omega; \mathbf{X}^{\text{tr}}\right),$$

such that for any test sample $\mathbf{X}^{\text{te}}$, the composite scoring function $S_t = \mathcal{S}(\mathbf{z}_{\ell(t)}, \mathcal{V}_\omega(\mathbf{X}^{\text{te}})_t; \Theta^*)$ produces calibrated low values for normal dynamics and high deviations when anomalous structural or transient patterns occur.

4 Methodology

4.1 Framework Overview

As illustrated in Figure 3, LATF-AD couples an adaptive multi-view time-frequency (TF) modeling branch with a complementary raw-domain verification branch. Given a time series $\mathbf{X} \in \mathbb{R}^{T \times C}$ and a normal training prefix $\mathbf{X}^{\text{tr}} = \mathbf{X}_{1:n}$, normal statistics are computed strictly on $\mathbf{X}^{\text{tr}}$ to standardize the series into $\widetilde{\mathbf{X}}$, preventing test anomalies from contaminating the normalization scale.

The standardized series $\widetilde{\mathbf{X}}$ is concurrently processed through two collaborative paths: (1) a multi-view time-frequency branch that constructs $V = 4$ heterogeneous spectral views, dynamically weighs their timestamp-level reliability via a local router, and optimizes dual reconstruction-prediction normality heads; and (2) an auxiliary raw patch branch that clusters temporal representations against a normal prototype memory to mitigate spectral artifact interference. Finally, the

residual evidences from both branches are calibrated via robust statistics to produce a composite anomaly score S_t. The end-to-end training, calibration, and inference procedures are summarized in Algorithm 1 (Appendix A).

4.2 Heterogeneous Multi-View Time-Frequency Bank

Owing to the fundamental Heisenberg–Gabor uncertainty principle, a single time-frequency transformation is inherently constrained by the trade-off between temporal and spectral resolution. A short analysis window isolates rapid transient bursts with high temporal fidelity but blurs low-frequency rhythmic oscillations, whereas a wide window resolves subtle spectral harmonics at the expense of temporal smearing. To overcome this limitation, LATF-AD explicitly constructs an ordered multi-view bank comprising $V = 4$ complementary transformations:

$$\mathcal{T} = \{\mathcal{T}_{\text{ss}}, \mathcal{T}_{\text{ls}}, \mathcal{T}_{\text{cwt}}, \mathcal{T}_{\text{wvd}}\},$$

which represent short-window STFT, long-window STFT, Continuous Wavelet Transform (CWT), and the Smoothed Pseudo Wigner–Ville Distribution (WVD), respectively.

Specifically, for each variable c of the standardized series $\widetilde{\mathbf{X}}$, we compute:

- **Short- and Long-Window STFT ($\mathbf{M}_{\text{ss}}, \mathbf{M}_{\text{ls}}$):** Complex STFTs are computed with Hann windows of lengths $W_{\text{len}} \in \{48, 192\}$ and frame hop $H = 8$. The log-magnitude spectra $A^{(W_{\text{len}})}_{c,k,\ell} = \log(1 + |S^{(W_{\text{len}})}_{c,k,\ell}|)$ are bilinearly resized to $F \times L$, where $L = \lceil T/H \rceil$ and $F = 16$. Here, $r = 48$ delivers sharp temporal localization for sudden spikes, while $r = 192$ provides fine frequency bins to expose long-term cyclic drift.
- **CWT View ($\mathbf{M}_{\text{cwt}}$):** We employ a Morlet wavelet over $S = 16$ logarithmically spaced scales $a_s \in [1.5, 36]$ to obtain coefficients $W_{c,s,\ell}$, resized to $F \times L$. CWT provides multi-scale dyadic tiling, naturally adapting temporal resolution at high frequencies and spectral resolution at low frequencies.
- **WVD View ($\mathbf{M}_{\text{wvd}}$):** To circumvent the linear Fourier resolution limit, we apply temporal lag window smoothing to the instantaneous correlation of the analytic signal $z_c(t)$, followed by log-magnitude compression and resizing to $F \times L$. This quadratic representation yields optimal instantaneous energy concentration for transient anomalies while suppressing cross-term interference.

By concatenating the variable and frequency axes into $D = CF$, the resulting feature tensors $\mathbf{M}_v \in \mathbb{R}^{D \times L}$ are standardized across the prefix frames $L_n = \lfloor n/H \rfloor$ and clipped to $[-12, 12]$. This ordered construction guarantees that diverse anomaly topologies—ranging from abrupt impulses to subtle stationary deviations—are distinctly exposed in at least one specialized coordinate space. Further discrete formulations and smoothing schemes for WVD and CWT are detailed in Appendix B.1.

4.3 Locally Adaptive Time-Frequency Routing and Consensus

While constructing multiple views guarantees anomaly coverage, assigning them uniform or static weights is suboptimal: views sensitive to transient shocks (e.g., CWT or WVD) may introduce high variance during stationary regimes, whereas stable views (e.g., long STFT) tend to dilute sharp localized spikes. To dynamically direct model capacity toward the most informative representation at every frame, independent view encoders $\mathcal{E}_{\phi_v}$ first project each input $\mathbf{M}_v$ into hidden representations:

$$\mathbf{h}_{v,\ell} = \mathcal{E}_{\phi_v}(\mathbf{M}_{v,:,\ell}) \in \mathbb{R}^d, \tag{1}$$

where $d = 48$, and each encoder utilizes depthwise and pointwise convolutions to decouple spatial-frequency mixing from local temporal modeling without forcing parameter sharing across heterogeneous views.

To evaluate timestamp-level view reliability without label supervision, a shared router network $\mathcal{R}_\theta$ extracts a 7-dimensional statistical descriptor $\mathbf{u}_{v,\ell}$ for each view at frame ℓ:

$$\mathbf{u}_{v,\ell} = [\mu_{v,\ell}, e_{v,\ell}, \Delta_{v,\ell}, \bar{e}_{v,\ell}, \delta_{v,\ell}, \mathbf{p}_v]^\top \in \mathbb{R}^7, \tag{2}$$

which encapsulates the temporal mean $\mu_{v,\ell} = D^{-1}\sum_d M_{v,d,\ell}$, root energy $e_{v,\ell}$, frame innovation $\Delta_{v,\ell} = D^{-1}\sum_d |M_{v,d,\ell} - M_{v,d,\ell-1}|$, 5-frame average-pooled energy $\bar{e}_{v,\ell}$, cross-view energy discrepancy $\delta_{v,\ell} = |e_{v,\ell} - V^{-1}\sum_u e_{u,\ell}|$, and a learned 2D view embedding $\mathbf{p}_v$. The router then estimates normalized reliability weights across views:

$$\pi_{v,\ell} = \frac{\mathcal{R}_\theta(\mathbf{u}_{v,1:L})_\ell}{\tau}, \qquad w_{v,\ell} = \frac{\exp(\pi_{v,\ell})}{\sum_{u=1}^{V}\exp(\pi_{u,\ell})}, \tag{3}$$

where $\tau > 0$ is a temperature hyperparameter. The localized consensus representation is formed by:

$$\mathbf{z}_\ell = \sum_{v=1}^{V} w_{v,\ell}\mathbf{h}_{v,\ell} \in \mathbb{R}^d. \tag{4}$$

By explicitly conditioning the weights on local frame-to-frame innovation $\Delta_{v,\ell}$ and cross-view discrepancy $\delta_{v,\ell}$, this formulation guides the network to dynamically amplify views capturing coherent transient or rhythmic changes while suppressing non-informative or noisy representations, preventing the detector from committing to a globally dominant view.

4.4 Dual Normality Heads

Pure reconstruction can erroneously reconstruct slow-evolving anomalies, while pure forecasting compounds errors over extended anomalies. LATF-AD resolves this by feeding the consensus $\mathbf{z}_\ell$ into two decoupled heads:

$$\widehat{\mathbf{M}}^{\text{rec}}_\ell = \mathcal{R}_{\psi_{\text{rec}}}(\mathbf{z}_\ell) \in \mathbb{R}^{V\times D}, \qquad \widehat{\mathbf{M}}^{\text{pred}}_\ell = \mathcal{P}_{\psi_{\text{pred}}}(\mathbf{z}_{1:\ell-1}) \in \mathbb{R}^{V\times D}, \tag{5}$$

where $\mathcal{R}_{\psi_{\text{rec}}}$ is a pointwise linear mapping capturing cross-spectral structures, and $\mathcal{P}_{\psi_{\text{pred}}}$ consists of causal dilated convolutional blocks forecasting one-step-ahead dynamics to penalize transition violations.

To train the router to discount corruptions unsupervisedly, we inject Gaussian noise into a random view during training: $\widetilde{\mathbf{M}}_{b,i_b} = \mathbf{M}_{b,i_b} + \lambda_b\boldsymbol{\epsilon}_b$. The total TF loss is:

$$\mathcal{L}_{\text{TF}} = \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{pred}} + 0.1\mathcal{L}_{\text{route}} + 0.01\mathcal{L}_{\text{rej}}, \tag{6}$$

where $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{pred}}$ denote MSE against clean targets $\mathbf{M}$ over batch size B and context length K. $\mathcal{L}_{\text{route}} = -(BVK)^{-1}\sum y_{b,v,\ell}\log w^{\text{clean}}_{b,v,\ell}$ aligns weights with dynamic innovation targets $y_{b,v,\ell}$ (Appendix B.2), while $\mathcal{L}_{\text{rej}} = -(BK)^{-1}\sum\log(1 - w^{\text{corr}}_{b,i_b,\ell} + 10^{-5})$ suppresses corrupted view weights.

4.5 Raw Patch Normality Branch

Nonlinear time-frequency transformations can produce spectral energy leakage and cross-term ripples when localized transients occur, occasionally causing multiple TF views to exhibit plausible but deceptive energy. To eliminate such ambiguity, LATF-AD deploys an independent raw-domain branch directly on length-P patches $\mathbf{U}_i = (\widetilde{\mathbf{X}}_{i:i+P-1,:})^\top \in \mathbb{R}^{C\times P}$ ($P = 128$). Here, $\overline{\mathbf{U}}_i$ denotes the patch normalized by its instance-level channel-wise mean and variance. A 1D-CNN encoder extracts normalized embeddings $\mathbf{e}_i = \mathcal{E}_\omega(\overline{\mathbf{U}}_i) \in \mathbb{R}^{64}$, trained via a triplet contrastive loss $\mathcal{L}_{\text{tri}}$ to pull temporally adjacent patches together and push distant patches apart, complemented by an adjacent patch ordering loss $\mathcal{L}_{\text{adj}}$.

After fitting on the normal prefix, all training embeddings are clustered via MiniBatchKMeans into a compact prototype memory bank $\mathcal{M} = \{\mathbf{m}_k\}_{k=1}^{K_m}$. During inference, each test patch $\mathbf{e}_i$ receives an anomaly distance defined by its cosine deviation from the top-3 nearest normal prototypes:

$$d_i = 1 - \frac{1}{3}\sum_{\mathbf{m}_k\in\text{Top3}} \mathbf{e}_i^\top\mathbf{m}_k, \tag{7}$$

which is aligned to a continuous pointwise time-domain score r_t by averaging overlapping patch estimates. Operating strictly in the physical time domain, this branch acts as an uncorrupted sanity check, invariant to spectral artifacts.

4.6 Robust Residual Calibration and Scoring

Because prediction residuals, reconstruction residuals, and prototype distances operate on fundamentally disparate numerical scales, naive linear combinations can easily be dominated by coordinate-wise variance. To address this scale discrepancy, squared residuals $E^{\mathrm{pred}}_{v,d,\ell}$ and $E^{\mathrm{rec}}_{v,d,\ell}$ are first normalized by their coordinate-wise normal prefix averages to yield scale-free evidences $e_{\mathrm{pred}}(\ell)$ and $e_{\mathrm{rec}}(\ell)$, which are linearly interpolated back to the original time resolution T.

Let $q^{\mathrm{tr}} = a\,\mathrm{Rob}(e^{\mathrm{tr}}_{\mathrm{pred}}; e^{\mathrm{tr}}_{\mathrm{pred}}) + b\,\mathrm{Rob}(e^{\mathrm{tr}}_{\mathrm{rec}}; e^{\mathrm{tr}}_{\mathrm{rec}})$ denote the composite training prefix evidence. The calibrated time-frequency evidence is obtained as:

$$\mathrm{TF}_t = \mathrm{Rob}\left(a\,\mathrm{Rob}(e_{\mathrm{pred}}; e^{\mathrm{tr}}_{\mathrm{pred}}) + b\,\mathrm{Rob}(e_{\mathrm{rec}}; e^{\mathrm{tr}}_{\mathrm{rec}}); q^{\mathrm{tr}}\right)_t. \tag{8}$$

Similarly, the raw evidence is calibrated and bounded as $g_t = \min(\mathrm{Rob}(r_t; r^{\mathrm{tr}}), 10)$. The final anomaly score is computed through a joint multiplicative-additive modulation:

$$S_t = \mathrm{TF}_t \cdot (1 + \alpha g_t) + \beta g_t, \tag{9}$$

where α and β are positive balancing hyperparameters. Here, the multiplicative term $(1 + \alpha g_t)$ allows genuine localized raw spikes to amplify borderline time-frequency anomalies, while the additive term βg_t ensures robust detection whenever spectral views fail to reach consensus due to severe energy diffusion, establishing a mutually reinforcing dual-domain defense. A asymptotic analysis of this scoring mechanism across typical anomaly cases is provided in Appendix B.3.

5 Experiments

5.1 Experimental Setups

Datasets. We evaluate LATF-AD on the TSB-AD Benchmark (Liu & Paparrizos, 2024), which provides a large-scale and standardized evaluation setting for TSAD. We use both TSB-AD-U and TSB-AD-M to cover univariate and multivariate detection scenarios. This benchmark is suitable for our study because it aggregates diverse real-world domains, contains heterogeneous anomaly types, and defines unified train/test splits and threshold-free evaluation. Following the benchmark protocol, models are trained on the provided training segments without anomaly labels and evaluated on the corresponding test series. Dataset statistics are provided in Appendix C.1.

Metrics. For a comprehensive evaluation, we adopt both range-wise metrics including VUS-PR (V-P), VUS-ROC (V-R) (Paparrizos et al., 2022), and Range-F1 (R-F) (Tatbul et al., 2018)) and point-wise metrics including AUC-PR (A-P), AUC-ROC (A-R) (Fawcett, 2006; Davis & Goadrich, 2006), and Point-F1 (P-F) (Rijsbergen, 1979). Following TSB-AD, VUS-PR is regarded as the primary measure. Detailed metric definitions are given in Appendix C.2.

Baselines. We evaluate LATF-AD against 14 representative baselines: (1) *Time-domain methods*, including AutoEncoder (Sakurada & Yairi, 2014), Donut (Xu et al., 2018), OmniAnomaly (Su et al., 2019), USAD (Audibert et al., 2020), AnomalyTransformer (Xu et al., 2022), TranAD (Tuli et al., 2022), PatchTST (Nie et al., 2023), DADA (Shentu et al., 2025), CrossAD (Li et al., 2026), PaAno (Park & Kang, 2026), and TimeRCD (Lan et al., 2026); and (2) *Frequency and time-frequency methods*, including Dual-TF (Nam et al., 2024), TFMAE (Fang et al., 2024), CATCH (Wu et al., 2025), and TSPulse (Ekambaram et al., 2026) (under both zero-shot and fine-tuned settings). All baselines follow the standard TSB-AD benchmark protocol and are configured with their officially recommended optimal hyperparameters; for CATCH on univariate series, the channel-wise awareness module is omitted.

Implementation Details. All methods are evaluated from timestamp-level anomaly scores under the same benchmark protocol. For all baseline models, we adopt the optimal hyperparameters reported in their respective original papers. We conduct each experiment 10 times and report the average results. Additional LATF-AD implementation settings are provided in Appendix C.3.

Table 1: Main results on TSB-AD. For all metrics, higher is better. Best results are in **bold**, and second-best are underlined.

Method	TSB-AD-U						TSB-AD-M					
	V-P	V-R	R-F	A-P	A-R	P-F	V-P	V-R	R-F	A-P	A-R	P-F
AutoEncoder	0.26	0.69	0.28	0.19	0.63	0.25	0.30	0.69	0.28	0.30	0.67	0.34
Donut	0.20	0.68	0.20	0.14	0.56	0.20	0.26	0.71	0.21	0.20	0.64	0.28
OmniAnomaly	0.29	0.72	0.29	0.27	0.65	0.31	0.31	0.69	0.37	0.27	0.65	0.32
USAD	0.36	0.71	0.40	0.32	0.66	0.37	0.30	0.68	0.37	0.26	0.64	0.31
AnomalyTransformer	0.12	0.56	0.14	0.08	0.50	0.12	0.12	0.57	0.14	0.07	0.52	0.12
TranAD	0.26	0.68	0.25	0.20	0.57	0.25	0.18	0.65	0.21	0.14	0.59	0.21
PatchTST	0.30	0.76	0.25	0.23	0.65	0.27	0.30	0.71	0.22	0.28	0.67	0.33
DADA	0.31	0.77	0.31	0.29	0.71	0.33	0.31	0.73	0.25	0.31	0.69	0.35
CrossAD	0.42	0.82	0.39	0.40	0.77	0.43	0.34	0.72	0.36	0.34	0.70	0.39
PaAno	0.53	0.89	0.49	0.47	0.87	0.52	0.43	0.79	0.41	0.38	0.76	0.43
TimeRCD	0.46	0.85	0.41	0.44	0.82	0.49	0.35	0.74	0.25	0.31	0.70	0.37
Dual-TF	0.31	0.76	0.30	0.27	0.69	0.33	0.32	0.70	0.33	0.29	0.66	0.35
TFMAE	0.10	0.51	0.10	0.05	0.40	0.10	0.10	0.50	0.09	0.06	0.45	0.11
CATCH	0.43	0.85	0.40	0.42	0.82	0.45	0.35	0.74	0.26	0.33	0.71	0.38
TSPulse (ZS)	0.47	0.80	0.48	0.42	0.77	0.48	0.40	0.71	0.45	0.38	0.69	0.44
TSPulse (FT)	0.54	0.87	**0.51**	0.48	0.84	0.54	0.43	0.74	**0.46**	0.41	0.72	0.46
LATF-AD (Ours)	**0.60**	**0.93**	**0.51**	**0.50**	**0.91**	**0.56**	**0.46**	**0.80**	**0.46**	**0.43**	**0.78**	**0.47**

Table 2: Ablation study on the core components of LATF-AD.

	Ablation Variant	VUS-PR	VUS-ROC	Range-F1	AUC-PR	AUC-ROC	Point-F1
TSB-AD-U	w/ Uniform TF Fusion	0.56	0.88	0.48	0.47	0.85	0.52
	Worst Leave-One-Out (w/o WVD)	0.56	0.85	0.48	0.46	0.88	0.53
	Best Single View (Short STFT)	0.53	0.83	0.45	0.44	0.85	0.50
	Worst Single View (WVD)	0.51	0.83	0.36	0.40	0.83	0.45
	w/o Raw Patch Normality Branch	0.56	0.91	0.49	0.48	0.88	0.54
	w/ Reconstruction Head Only	0.56	0.91	0.46	0.45	0.89	0.52
	Full LATF-AD	**0.60**	**0.93**	**0.51**	**0.50**	**0.91**	**0.56**
TSB-AD-M	w/ Uniform TF Fusion	0.43	0.77	0.42	0.39	0.75	0.45
	Worst Leave-One-Out (w/o WVD)	0.45	0.78	0.44	0.41	0.75	0.46
	Best Single View (Short STFT)	0.42	0.77	0.38	0.37	0.72	0.44
	Worst Single View (WVD)	0.40	0.75	0.38	0.37	0.72	0.42
	w/o Raw Patch Normality Branch	0.44	0.78	0.42	0.39	0.75	0.44
	w/ Reconstruction Head Only	0.45	0.79	0.42	0.41	0.77	0.46
	Full LATF-AD	**0.46**	**0.80**	**0.46**	**0.43**	**0.78**	**0.47**

5.2 Main Results

Table 1 reports the quantitative comparison between LATF-AD and 15 competitive baselines on both TSB-AD-U and TSB-AD-M benchmarks. Overall, LATF-AD achieves state-of-the-art detection across all evaluation metrics, attaining a primary score of 0.60 VUS-PR and 0.93 VUS-ROC on univariate series, and 0.46 VUS-PR and 0.80 VUS-ROC on multivariate series, consistently outperforming both fine-tuned foundation models (e.g., TSPulse) and advanced domain-dedicated architectures (e.g., PaAno). By dynamically routing heterogeneous time-frequency representations (STFT, CWT, WVD) while anchoring raw patch prototypes, LATF-AD successfully resolves the Heisenberg uncertainty dilemma, thereby yielding sharp anomaly segment boundaries and suppressing false alarms through robust multiplicative-additive calibration without requiring massive external pre-training. Detailed per-dataset evaluations are reported in Appendix C.4, and computational efficiency analyses are provided in Appendix C.5.

5.3 Ablation Study

We conducted a comprehensive ablation study to assess the contribution of each key component in LATF-AD across both TSB-AD-U and TSB-AD-M benchmarks, as summarized in Table 2. Across both settings, replacing dynamic routing with uniform fusion or isolating single views results in noticeable drops across range-wise and point-wise metrics. Removing the raw patch normality

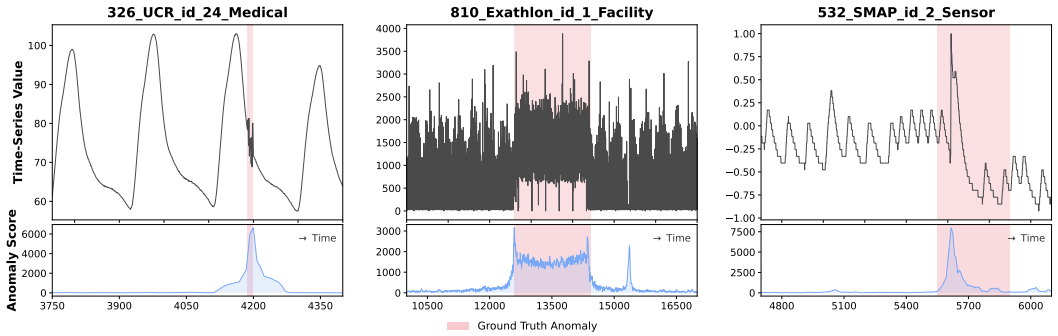

Figure 4: Qualitative anomaly detection visualizations on representative TSB-AD-U segments. Shaded regions denote ground-truth anomalies.

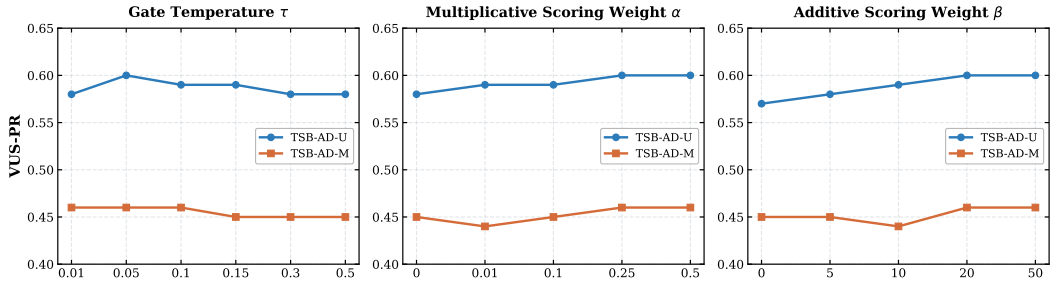

Figure 5: Hyperparameter sensitivity of gate temperature τ, multiplicative weight α, and additive weight β in terms of VUS-PR on TSB-AD-U and TSB-AD-M.

branch further degrades performance, confirming the necessity of time-domain anchoring. Detailed full-scale ablation results are provided in Appendix C.6.

5.4 Model Analysis

Qualitative visualization. Figure 4 illustrates the anomaly score profiles of LATF-AD across three representative scenarios: high-frequency disturbances (Medical), sustained shifts under noise (Facility), and transient spikes (Sensor). Across all cases, LATF-AD exhibits high detection fidelity: scores remain suppressed near zero during nominal phases while escalating sharply strictly within ground-truth intervals.

Sensitivity analysis. Figure 5 illustrates that LATF-AD achieves remarkably stable VUS-PR across broad parameter ranges on both benchmarks. Specifically, a moderate temperature ($\tau \in [0.05, 0.15]$) yields optimal results by encouraging decisive view routing over uniform weighting. Moreover, positive scoring weights ($\alpha \in [0.25, 0.5]$, $\beta \in [20, 50]$) steadily boost performance to peak scores (0.60 and 0.46), confirming the synergy of cross-domain calibration.

6 Conclusion

We proposed LATF-AD, an unsupervised TSAD framework with locally adaptive multi-view time-frequency modeling. LATF-AD constructs a heterogeneous spectral bank spanning STFT, CWT, and WVD, adaptively dispatching timestamp-specific reliability weights to drive dual reconstruction and prediction heads. To guard against spectral energy leakage and cross-term artifacts, an auxiliary time-domain prototype memory provides independent local grounding, unified through a robust multiplicative-additive calibration mechanism. Empirical results across the comprehensive TSB-AD-U and TSB-AD-M benchmarks demonstrate that LATF-AD achieves state-of-the-art performance across all six point-wise and range-based evaluation metrics.

AI USE STATEMENT

In this work, we used generative AI tools (such as large language models) for three main purposes: (1) refining and polishing the academic writing, (2) assisting in literature retrieval to survey relevant works in time-series anomaly detection, and (3) facilitating code debugging and preliminary brainstorming during the experimental execution. We have not used AI tools to draft full sections of the paper, generate synthetic datasets, or prove mathematical claims. All AI-assisted suggestions and outputs were manually verified, tested, and thoroughly checked by the authors. We take full responsibility for the entire content, technical claims, and artifacts presented in this paper.

ETHICS STATEMENT

This work focuses on developing algorithms for time-series anomaly detection. All experiments are conducted using publicly available benchmark datasets. This research does not involve human subjects, personal data, or sensitive information, and we foresee no significant negative ethical impacts or potential risks of malicious misuse.

REPRODUCIBILITY STATEMENT

Our code and default hyperparameter configurations are fully included in the Supplementary Material. In addition, detailed descriptions and source links for all public datasets used in our experiments are elaborated in Appendix C.1. Full implementation details, training schedules, and evaluation protocols are further documented in Appendix C.3.

REFERENCES

Julien Audibert, Pietro Michiardi, Frédéric Guyard, Sébastien Marti, and Maria A Zuluaga. USAD: Unsupervised anomaly detection on multivariate time series. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3395–3404, 2020.

Leon Cohen. The uncertainty principle in signal analysis. In *Proceedings of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 182–185. IEEE, 1994.

Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pp. 233–240. Association for Computing Machinery, 2006. ISBN 1595933832.

Vijay Ekambaram, Subodh Kumar, Arindam Jati, Sumanta Mukherjee, Tomoya Sakai, Pankaj Dayama, Wesley M. Gifford, and Jayant Kalagnanam. Tspulse: Dual space tiny pre-trained models for rapid time-series analysis, 2026.

Yuchen Fang, Jiandong Xie, Yan Zhao, Lu Chen, Yunjun Gao, and Kai Zheng. Temporal-frequency masked autoencoders for time series anomaly detection. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pp. 1228–1241. IEEE, 2024.

Tom Fawcett. An introduction to roc analysis. *Pattern Recogn. Lett.*, 27(8):861–874, June 2006. ISSN 0167-8655. doi: 10.1016/j.patrec.2005.10.010. URL https://doi.org/10.1016/j.patrec.2005.10.010.

William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

Tian Lan, Hao Duong Le, Jinbo Li, Wenjun He, Meng Wang, Chenghao Liu, and Chen Zhang. Towards foundation models for zero-shot time series anomaly detection: Leveraging synthetic data and relative context discrepancy. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026.

Beibu Li, Qichao Shentu, Yang Shu, Hui Zhang, Ming Li, Ning Jin, Bin Yang, and Chenjuan Guo. CrossAD: Time series anomaly detection with cross-scale associations and cross-window modeling. *Advances in Neural Information Processing Systems*, 38:137959–137985, 2026.

Qinghua Liu and John Paparrizos. The elephant in the room: Towards a reliable time-series anomaly detection benchmark. *Advances in Neural Information Processing Systems*, 37:108231–108261, 2024.

Youngeun Nam, Susik Yoon, Yooju Shin, Minyoung Bae, Hwanjun Song, Jae-Gil Lee, and Byung Suk Lee. Breaking the time-frequency granularity discrepancy in time-series anomaly detection. In *Proceedings of the ACM Web Conference 2024*, pp. 4204–4215, 2024.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

John Paparrizos, Paul Boniol, Themis Palpanas, Ruey S Tsay, Aaron J Elmore, and Michael J Franklin. Volume under the surface: A new accuracy evaluation measure for time-series anomaly detection. *Proc. VLDB Endow.*, 15(11):2774–2787, 2022.

Jinju Park and Seokho Kang. PaAno: Patch-based representation learning for time series anomaly detection. In *International Conference on Learning Representations*, 2026.

CJ van Rijsbergen. *Information retrieval*. Butterworth-heinemann, 1979.

Mayu Sakurada and Takehisa Yairi. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, pp. 4–11, 2014.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017.

Qichao Shentu, Beibu Li, Kai Zhao, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. Towards a general time series anomaly detector with adaptive bottlenecks and dual adversarial decoders. In *International Conference on Learning Representations*, 2025.

Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data mining*, pp. 2828–2837, 2019.

Nesime Tatbul, Tae Jun Lee, Stan Zdonik, Mejbah Alam, and Justin Gottschlich. Precision and recall for time series. *Advances in Neural Information Processing Systems*, 31, 2018.

Shreshth Tuli, Giuliano Casale, and Nicholas R. Jennings. TranAD: deep transformer networks for anomaly detection in multivariate time series data. *Proc. VLDB Endow.*, 15(6):1201–1214, February 2022. ISSN 2150–8097. doi: 10.14778/3514061.3514067. URL `https://doi.org/10.14778/3514061.3514067`.

Xingjian Wu, Xiangfei Qiu, Zhengyu Li, Yihang Wang, Jilin Hu, Chenjuan Guo, Hui Xiong, and Bin Yang. CATCH: Channel-aware multivariate time series anomaly detection via frequency patching. In *International Conference on Learning Representations*, volume 2025, pp. 17017–17045, 2025.

Haowen Xu, Wenxiao Chen, Nengwen Zhao, Zeyan Li, Jiahao Bu, Zhihan Li, Ying Liu, Youjian Zhao, Dan Pei, Yang Feng, et al. Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. In *Proceedings of the 2018 World Wide Web Conference*, pp. 187–196, 2018.

Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly transformer: Time series anomaly detection with association discrepancy. In *International Conference on Learning Representations*, 2022.

Yiyuan Yang, Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun. DCdetector: Dual attention contrastive representation learning for time series anomaly detection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, pp. 3033–3045. ACM, August 2023.

Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun. TFAD: A decomposition time series anomaly detection architecture with time-frequency analysis. In *Proceedings of the 31st ACM International Conference on Information amp; Knowledge Management*, CIKM '22, pp. 2497–2507. ACM, October 2022.

A Algorithmic Summary

Algorithm 1 provides the algorithmic pipeline of LATF-AD, formalizing the two-phase workflow: (1) prefix-only unsupervised training and calibration, and (2) offline dual-domain inference.

Algorithm 1 End-to-End Training, Calibration, and Inference of LATF-AD

Require: Series $\mathbf{X} \in \mathbb{R}^{T\times C}$, prefix length n, TF operators $\{\mathcal{T}_v\}_{v=1}^{V}$, hyperparameters a, b, α, β.
Ensure: Pointwise calibrated anomaly scores $\{S_t\}_{t=1}^{T}$.

// Phase 1: Prefix Calibration and Representation Extraction
Fit mean $\boldsymbol{\mu}$ and std $\boldsymbol{\sigma}$ on $\mathbf{X}^{\mathrm{tr}} = \mathbf{X}_{1:n}$; compute $\widetilde{\mathbf{X}} = (\mathbf{X} - \boldsymbol{\mu})/\boldsymbol{\sigma}$.
Extract $V = 4$ standardized multi-view feature tensors $\mathbf{M}_v = \mathcal{T}_v(\widetilde{\mathbf{X}}) \in \mathbb{R}^{D\times L}$.
Fit feature normalization statistics on prefix frames $L_n = \lfloor n/H \rfloor$ to produce inputs $\mathbf{M}_v$.
// Phase 2: Unsupervised Model Training
Sample normal windows from prefix features; inject synthetic noise to form corrupted views $\widetilde{\mathbf{M}}$.
Optimize TF branch parameters $\Theta_{\mathrm{TF}} = \{\theta, \{\phi_v\}_{v=1}^{V}, \psi_{\mathrm{rec}}, \psi_{\mathrm{pred}}\}$ via $\mathcal{L}_{\mathrm{TF}}$ (Eq. 6) to convergence.
Segment $\widetilde{\mathbf{X}}_{1:n}$ into patches $\mathbf{U}_i \in \mathbb{R}^{C\times P}$; optimize raw patch encoder $\mathcal{E}_\omega$ via $\mathcal{L}_{\mathrm{tri}} + \lambda_{\mathrm{adj}}\mathcal{L}_{\mathrm{adj}}$.
// Phase 3: Offline Memory and Robust Reference Fitting
Cluster prefix patch representations $\{\mathbf{e}_i\}_{i=1}^{N}$ via MiniBatchKMeans into prototype memory $\mathcal{M} = \{\mathbf{m}_k\}_{k=1}^{K_m}$.
Evaluate trained TF heads on normal prefix to establish coordinate residual scales $a_{v,d}^{\mathrm{pred}}, a_{v,d}^{\mathrm{rec}}$.
Compute prefix baseline evidences $e_{\mathrm{pred}}^{\mathrm{tr}}, e_{\mathrm{rec}}^{\mathrm{tr}}, q^{\mathrm{tr}}, r^{\mathrm{tr}}$, then fit median and MAD scaling references $\mathrm{Rob}(\cdot)$.
// Phase 4: Full-Series Test Inference
For entire $\widetilde{\mathbf{X}}$, perform dynamic routing $\mathbf{z}_\ell = \sum_{v=1}^{V} w_{v,\ell}\mathcal{E}_{\phi_v}(\mathbf{M}_{v,:,\ell})$ and obtain dual predictions $\widehat{\mathbf{M}}^{\mathrm{rec}}, \widehat{\mathbf{M}}^{\mathrm{pred}}$.
Compute scaled residual evidences $e_{\mathrm{pred}}, e_{\mathrm{rec}}$ and linearly interpolate to time length T.
Evaluate patch distances d_i against $\mathcal{M}$ (Eq. 7); compute overlap-averaged time evidence r_t.
Compute robustly calibrated scores TF_t (Eq. 8) and $g_t = \min(\mathrm{Rob}(r_t; r^{\mathrm{tr}}), 10)$.

return Final composite anomaly scores $S_t = \mathrm{TF}_t \cdot (1 + \alpha g_t) + \beta g_t$ for $t = 1, \ldots, T$.

B Method Details

B.1 Discrete Formulation of the Heterogeneous Time-Frequency Bank

This subsection details the mathematical discretization and algorithmic filtering implemented in the $V = 4$ heterogeneous time-frequency operators $\mathcal{T} = \{\mathcal{T}_{\mathrm{ss}}, \mathcal{T}_{\mathrm{ls}}, \mathcal{T}_{\mathrm{cwt}}, \mathcal{T}_{\mathrm{wvd}}\}$.

Smoothed Pseudo Wigner–Ville Distribution (SPWVD). For a real-valued time series, naive quadratic representations suffer heavily from low-frequency cross-terms generated by negative frequency components. To eliminate interference between positive and negative frequencies, we first transform each standardized channel $\widetilde{X}_{:,c}$ into its complex analytic counterpart:

$$z_c(t) = \widetilde{X}_{t,c} + \mathrm{i} \cdot \mathcal{H}\{\widetilde{X}_{:,c}\}(t), \tag{10}$$

where $\mathcal{H}\{\cdot\}$ denotes the Hilbert transform computed via Fast Fourier Transform (FFT). For frame index ℓ (sampling time $t = \ell H$) and frequency lag index $k \in \{0, \ldots, 2R\}$, the smoothed discrete pseudo Wigner–Ville distribution is formulated as:

$$W_{c,\ell,k}^{\mathrm{wvd}} = \sum_{r=-R}^{R} h(r) \left[\sum_{m=-M}^{M} g(m) z_c(\ell H + m + r) z_c^*(\ell H + m - r) \right] \exp\left(-\mathrm{i} \frac{2\pi k r}{2R+1} \right), \tag{11}$$

where $R = 18$ denotes the frequency-direction lag window radius, $h(r)$ is a 7-point Hann smoothing window across frequency lag, and $g(m)$ is an independent temporal smoothing window. In our discrete implementation, separable convolution over time and lag enables $\mathcal{O}(R \log R)$ per-frame evaluation, suppressing the prominent oscillatory artifacts characteristic of classical bilinear distributions while preserving optimal instantaneous energy concentration.

Continuous Wavelet Transform (CWT) Scale Tiling. The continuous wavelet transform maps the time series across continuous translation τ and dilation scale a. In our discrete bank, we employ the complex Morlet wavelet:

$$\psi(t) = \pi^{-1/4} \exp(\mathrm{i}\omega_0 t) \exp\left(-\frac{t^2}{2}\right), \tag{12}$$

with central angular frequency $\omega_0 = 5$. To balance computational budget and spectral span, we sample $S = 16$ discrete scales logarithmically:

$$a_s = a_{\min} \cdot \left(\frac{a_{\max}}{a_{\min}}\right)^{\frac{s-1}{S-1}}, \quad s \in \{1, \ldots, S\}, \tag{13}$$

where $a_{\min} = 1.5$ and $a_{\max} = 36$. This geometric progression satisfies dyadic multiscale tiling, yielding high temporal precision for rapid high-frequency bursts and expansive support for low-frequency oscillations.

B.2 Construction of the Unsupervised Dynamic Salience Target

In Section 4.4, the routing loss $\mathcal{L}_{\text{route}}$ guides the network without ground-truth anomaly supervision. We elaborate on the statistical design of the dynamic salience target $y_{b,v,\ell}$.

For a clean feature window $A \in \mathbb{R}^{B \times V \times D \times K}$, the local structural activity of view v at frame ℓ is characterized by the squared frame innovation:

$$\Delta_{b,v,\ell} = \frac{1}{D} \sum_{d=1}^{D} (A_{b,v,d,\ell} - A_{b,v,d,\ell-1})^2, \tag{14}$$

with the boundary condition $\Delta_{b,v,1} = \Delta_{b,v,2}$. Because the absolute variance scales naturally vary across heterogeneous transforms (e.g., STFT log-power vs. CWT modulus), direct magnitude comparison across views introduces systematic bias. To resolve this, we first normalize the innovation dynamically within each view's context window:

$$\widetilde{\Delta}_{b,v,\ell} = \frac{\Delta_{b,v,\ell}}{\frac{1}{K}\sum_{j=1}^{K} \Delta_{b,v,j} + 10^{-5}}. \tag{15}$$

To evaluate relative local informativeness across the heterogeneous views, we compute cross-view zero-centered salience deviations:

$$c_{b,v,\ell} = r_{b,v,\ell} - \frac{1}{V}\sum_{u=1}^{V} r_{b,u,\ell}, \qquad s_{b,\ell} = \max\left(\sqrt{\frac{1}{V}\sum_{v=1}^{V} c_{b,v,\ell}^2}, 0.1\right). \tag{16}$$

The pseudo-ground-truth probability distribution $y_{b,v,\ell}$ over the simplex Δ^{V-1} is defined by:

$$y_{b,v,\ell} = \frac{\exp\left(\mathrm{clip}(c_{b,v,\ell}/s_{b,\ell}, -4, 4)/\kappa\right)}{\sum_{u=1}^{V} \exp\left(\mathrm{clip}(c_{b,u,\ell}/s_{b,\ell}, -4, 4)/\kappa\right)}, \tag{17}$$

where $\kappa = 0.5$ is a target temperature. This normalization guarantees that if a localized disturbance causes one view to undergo dynamic structural reorganization significantly above the multi-view background, $y_{b,v,\ell}$ decisively assigns high target mass to that view.

B.3 Mathematical Analysis of the Collaborative Scoring Function

The composite anomaly score defined in Eq. 9 couples time-frequency evidence TF_t and raw patch evidence g_t via $S_t = \mathrm{TF}_t \cdot (1 + \alpha g_t) + \beta g_t$. We provide an asymptotic case analysis demonstrating how this formulation provides robust cross-domain defense.

Case 1: Coincident Anomalous Detection ($\mathrm{TF}_t \gg 0$ and $g_t \gg 0$). When a true anomalous event exhibits both temporal distortion and spectral deviation:

$$S_t \approx \alpha \cdot \mathrm{TF}_t \cdot g_t + \mathrm{TF}_t + \beta g_t \approx \mathcal{O}(\mathrm{TF}_t \cdot g_t). \tag{18}$$

The multiplicative interaction dominates, inducing a super-linear score surge. This drastically elevates the signal-to-noise ratio around true anomalies, creating sharp boundaries that facilitate thresholding and range-based detection metrics.

Case 2: Spectral Diffusion or Multi-View Consensus Breakdown ($\mathrm{TF}_t \approx 0$ and $g_t \gg 0$). When non-stationary noise spreads energy diffusely across multiple TF views, the router may form a smooth but flawed consensus that yields low reconstruction/prediction errors ($\mathrm{TF}_t \approx 0$). In this regime:

$$S_t \approx \beta g_t. \tag{19}$$

The additive term acts as an unconditional safety fallback, ensuring that any genuine physical-domain deviation detected by the prototype memory directly triggers an alert, preventing false negatives.

Case 3: Spurious Spectral Artifacts or Cross-Term Ripples ($\mathrm{TF}_t \gg 0$ and $g_t \approx 0$). When complex nonlinear transformations (e.g., WVD or CWT) generate mathematical artifact peaks in the absence of physical time-domain anomalies ($g_t \approx 0$):

$$S_t = \mathrm{TF}_t \cdot (1 + 0) + 0 = \mathrm{TF}_t. \tag{20}$$

Because the multiplier $(1 + \alpha g_t)$ remains bounded at unity and the additive offset vanishes, the unconfirmed artifact is denied exponential amplification, effectively containing false alarm rates.

C Additional Experimental Details

C.1 Dataset Statistics

Table 3 summarizes the TSB-AD Benchmark splits used in our experiments. TSB-AD-U contains univariate series, while TSB-AD-M contains multivariate series. The benchmark is publicly available at `https://thedatumorg.github.io/TSB-AD/`, and the TSB-AD code repository is released under the Apache-2.0 license at `https://github.com/TheDatumOrg/TSB-AD`. The tuning split is used for model selection and hyperparameter tuning, and the evaluation split is used for final reporting.

Table 3: Statistics of the TSB-AD Benchmark.

Category	# Time-Series Collected	Average Length	Average Anomaly Ratio
TSB-AD-U	350	51,886.75	4.47%
TSB-AD-M	180	108,826.76	5.01%

C.2 Metric Definitions

Let $y_t \in \{0, 1\}$ be the ground-truth label and $s_t \in \mathbb{R}_+$ be the anomaly score at timestamp t. For a threshold θ, the binary prediction is $\hat{y}_t(\theta) = \mathbb{I}[s_t \geq \theta]$. The true positive rate, false positive rate, precision, and recall are

$$\mathrm{TPR}(\theta) = \frac{\mathrm{TP}(\theta)}{\mathrm{TP}(\theta) + \mathrm{FN}(\theta)}, \quad \mathrm{FPR}(\theta) = \frac{\mathrm{FP}(\theta)}{\mathrm{FP}(\theta) + \mathrm{TN}(\theta)}, \tag{21}$$

$$\mathrm{Prec}(\theta) = \frac{\mathrm{TP}(\theta)}{\mathrm{TP}(\theta) + \mathrm{FP}(\theta)}, \quad \mathrm{Rec}(\theta) = \mathrm{TPR}(\theta). \tag{22}$$

AUC-ROC is the area under the ROC curve:

$$\text{AUC-ROC} = \int_0^1 \mathrm{TPR}(u)\, du, \quad u = \mathrm{FPR}\,. \tag{23}$$

AUC-PR is computed as average precision over the score-induced ranking:

$$\text{AUC-PR} = \sum_i (\text{Rec}_i - \text{Rec}_{i-1})\,\text{Prec}_i\,. \tag{24}$$

VUS metrics follow the range-aware evaluation protocol established by Paparrizos et al. (2022) and standardized in TSB-AD. For a given buffer window size r, a range-based AUC metric $\text{Range-AUC}(r)$ (either Range-AUC-ROC or Range-AUC-PR) is computed by integrating the range-aware curve over all decision thresholds while relaxing the ground-truth anomaly boundaries with temporal tolerance r. To eliminate sensitivity to the choice of buffer size, VUS computes the volume under the two-dimensional surface formed across both thresholds and buffer sizes $r \in [0, W]$, where W is the maximum tolerance window (set to twice the median length of ground-truth anomaly segments following the benchmark standard). In practice, VUS-ROC and VUS-PR are evaluated by averaging the range-based AUC values across discrete buffer sizes:

$$\text{VUS-ROC} = \frac{1}{W+1}\sum_{r=0}^{W} \text{Range-AUC-ROC}(r), \tag{25}$$

$$\text{VUS-PR} = \frac{1}{W+1}\sum_{r=0}^{W} \text{Range-AUC-PR}(r). \tag{26}$$

The standard point-wise F1 score (Point-F1) is defined as the maximum F1 score over all possible detection thresholds θ:

$$\text{Point-F1} = \max_\theta \frac{2\cdot \text{Prec}(\theta)\cdot \text{Rec}(\theta)}{\text{Prec}(\theta)+\text{Rec}(\theta)}. \tag{27}$$

To evaluate segment-level detection performance, an anomaly segment is defined as a contiguous subsequence of timestamps labeled as an anomaly. Let $\mathcal{A} = \{A_1, \ldots, A_M\}$ denote the set of ground-truth anomaly segments, where $A_j = \{t \mid a_j \le t \le b_j, y_t = 1\}$. For a given threshold θ, let $\hat{\mathcal{A}}(\theta) = \{\hat{A}_1, \ldots, \hat{A}_{N_\theta}\}$ denote the set of predicted anomaly segments, where $\hat{A}_i = \{t \mid \hat{a}_i \le t \le \hat{b}_i, \hat{y}_t(\theta) = 1\}$. The range-based precision and recall at threshold θ are defined as:

$$\text{Range-Prec}(\theta) = \frac{\left|\left\{\hat{A}_i \in \hat{\mathcal{A}}(\theta) \mid \exists A_j \in \mathcal{A},\ \hat{A}_i \cap A_j \neq \emptyset\right\}\right|}{|\hat{\mathcal{A}}(\theta)|}, \tag{28}$$

$$\text{Range-Rec}(\theta) = \frac{\left|\left\{A_j \in \mathcal{A} \mid \exists \hat{A}_i \in \hat{\mathcal{A}}(\theta),\ A_j \cap \hat{A}_i \neq \emptyset\right\}\right|}{|\mathcal{A}|}. \tag{29}$$

The optimal range-based F1 score is then defined by maximizing across all thresholds:

$$\text{Range-F1} = \max_\theta \frac{2\cdot \text{Range-Prec}(\theta)\cdot \text{Range-Rec}(\theta)}{\text{Range-Prec}(\theta)+\text{Range-Rec}(\theta)}. \tag{30}$$

C.3 Implementation Details

We evaluate LATF-AD on the TSB-AD univariate and multivariate benchmarks, which contain 350 TSB-AD-U files and 180 TSB-AD-M files, respectively. Each file is trained independently using the normal prefix specified by its filename; anomaly labels are never used during preprocessing, model fitting, prototype construction, or score calibration. The input is standardized using prefix statistics and transformed into four time-frequency views: short- and long-window STFT with window lengths 48 and 192, CWT with 16 logarithmically spaced scales from 1.5 to 36, and WVD with maximum lag 18 and smoothing length 7. All views are represented with 16 frequency bins and a frame hop of 8. The TF branch uses four independent encoders with hidden width 48 and GroupNorm, followed by two temporal blocks for TSB-AD-U and three for TSB-AD-M, with the routing temperature set to $\tau = 0.05$. It is optimized with AdamW for 100 steps using a batch size of 256, 32-frame contexts, learning rate 10^{-3}, weight decay 10^{-4}, and gradient clipping at 1. Checkpoints are selected using four fixed normal validation batches evaluated every 10 steps. The Raw Patch branch uses length-128 patches with stride one, a CNN with widths $(128, 256, 128, 64)$, and a batch

Table 4: Per-dataset results of LATF-AD on TSB-AD-U and TSB-AD-M.

	Name of Subset	# Time-Series	Avg. Dim	VUS-PR	VUS-ROC	Range-F1	AUC-PR	AUC-ROC	Point-F1
TSB-AD-U	UCR	70	1	0.4590 ± 0.0243	0.9448 ± 0.0269	0.4362 ± 0.0099	0.4120 ± 0.0037	0.9323 ± 0.0072	0.4671 ± 0.0131
	YAHOO	30	1	0.6087 ± 0.0040	0.9649 ± 0.0105	0.3502 ± 0.0283	0.2514 ± 0.0100	0.9578 ± 0.0158	0.3544 ± 0.0212
	WSD	20	1	0.6331 ± 0.0246	0.9795 ± 0.0259	0.6921 ± 0.0007	0.5993 ± 0.0156	0.9757 ± 0.0128	0.6601 ± 0.0071
	CATSv2	1	1	0.3989 ± 0.0115	0.7645 ± 0.0285	0.2306 ± 0.0221	0.5071 ± 0.0182	0.7470 ± 0.0051	0.5826 ± 0.0051
	Daphnet	1	1	0.5185 ± 0.0022	0.9561 ± 0.0261	0.3951 ± 0.0182	0.4972 ± 0.0214	0.9514 ± 0.0011	0.5176 ± 0.0291
	Exathlon	30	1	0.9406 ± 0.0251	0.9954 ± 0.0068	0.6827 ± 0.0059	0.9343 ± 0.0059	0.9953 ± 0.0095	0.9587 ± 0.0160
	IOPS	15	1	0.4867 ± 0.0132	0.9711 ± 0.0091	0.4780 ± 0.0185	0.3045 ± 0.0046	0.9566 ± 0.0091	0.4266 ± 0.0113
	LTDB	8	1	0.6177 ± 0.0140	0.7972 ± 0.0237	0.6174 ± 0.0064	0.5628 ± 0.0157	0.7558 ± 0.0180	0.6247 ± 0.0019
	MGAB	8	1	0.1094 ± 0.0184	0.9068 ± 0.0055	0.2815 ± 0.0024	0.1115 ± 0.0285	0.8369 ± 0.0290	0.2987 ± 0.0243
	MITDB	7	1	0.2664 ± 0.0095	0.8125 ± 0.0034	0.2750 ± 0.0207	0.2483 ± 0.0135	0.7784 ± 0.0041	0.3071 ± 0.0151
	MSL	7	1	0.6203 ± 0.0015	0.9249 ± 0.0273	0.5703 ± 0.0081	0.5439 ± 0.0200	0.8846 ± 0.0097	0.5794 ± 0.0158
	NAB	23	1	0.5696 ± 0.0166	0.8211 ± 0.0060	0.5499 ± 0.0291	0.5536 ± 0.0234	0.8043 ± 0.0282	0.5539 ± 0.0269
	NEK	8	1	0.7484 ± 0.0181	0.8909 ± 0.0277	0.7699 ± 0.0031	0.7601 ± 0.0063	0.9128 ± 0.0018	0.7417 ± 0.0101
	OPPORTUNITY	27	1	0.3585 ± 0.0120	0.8359 ± 0.0085	0.2852 ± 0.0249	0.3319 ± 0.0110	0.8254 ± 0.0088	0.4729 ± 0.0165
	Power	1	1	0.1418 ± 0.0047	0.6233 ± 0.0242	0.1486 ± 0.0027	0.1397 ± 0.0296	0.6047 ± 0.0233	0.2302 ± 0.0064
	SED	2	1	0.9837 ± 0.0007	0.9992 ± 0.0246	0.8207 ± 0.0214	0.8947 ± 0.0220	0.9950 ± 0.0233	0.8301 ± 0.0027
	SMAP	17	1	0.8810 ± 0.0111	0.9727 ± 0.0039	0.7949 ± 0.0260	0.8524 ± 0.0189	0.9703 ± 0.0103	0.8185 ± 0.0024
	SMD	33	1	0.7508 ± 0.0097	0.9719 ± 0.0101	0.6974 ± 0.0220	0.6786 ± 0.0193	0.9692 ± 0.0267	0.7348 ± 0.0144
	Stock	8	1	0.6367 ± 0.0040	0.7775 ± 0.0215	0.3253 ± 0.0229	0.1171 ± 0.0171	0.5952 ± 0.0232	0.1859 ± 0.0151
	SVDB	18	1	0.5743 ± 0.0159	0.9664 ± 0.0131	0.4922 ± 0.0012	0.5510 ± 0.0037	0.9487 ± 0.0014	0.5741 ± 0.0193
	SWaT	1	1	0.2269 ± 0.0098	0.6955 ± 0.0155	0.0941 ± 0.0273	0.2361 ± 0.0079	0.7217 ± 0.0126	0.3859 ± 0.0228
	TAO	2	1	0.9373 ± 0.0072	0.9548 ± 0.0028	0.2717 ± 0.0090	0.1750 ± 0.0053	0.6089 ± 0.0279	0.2529 ± 0.0243
	TODS	13	1	0.7409 ± 0.0192	0.9086 ± 0.0262	0.3011 ± 0.0242	0.2315 ± 0.0060	0.7768 ± 0.0268	0.3352 ± 0.0164
TSB-AD-U Average				$\mathbf{0.5980 \pm 0.0155}$	$\mathbf{0.9270 \pm 0.0157}$	$\mathbf{0.5060 \pm 0.0168}$	$\mathbf{0.4970 \pm 0.0118}$	$\mathbf{0.9053 \pm 0.0145}$	$\mathbf{0.5570 \pm 0.0149}$
TSB-AD-M	CATSv2	5	17	0.4006 ± 0.0112	0.8232 ± 0.0292	0.4073 ± 0.0289	0.3823 ± 0.0079	0.7967 ± 0.0152	0.4560 ± 0.0094
	CreditCard	1	29	0.0907 ± 0.0089	0.7658 ± 0.0016	0.0746 ± 0.0185	0.0312 ± 0.0153	0.7796 ± 0.0020	0.1018 ± 0.0087
	Daphnet	1	9	0.4203 ± 0.0273	0.9416 ± 0.0076	0.3547 ± 0.0048	0.4136 ± 0.0149	0.9376 ± 0.0296	0.4555 ± 0.0076
	Exathlon	25	20.16	0.9008 ± 0.0203	0.9474 ± 0.0230	0.8039 ± 0.0075	0.9246 ± 0.0220	0.9646 ± 0.0113	0.9346 ± 0.0192
	GECCO	1	9	0.6028 ± 0.0192	0.9912 ± 0.0163	0.4897 ± 0.0032	0.5533 ± 0.0251	0.9896 ± 0.0100	0.5389 ± 0.0060
	Genesis	1	18	0.7018 ± 0.0017	0.9991 ± 0.0179	0.4644 ± 0.0205	0.6518 ± 0.0010	0.9983 ± 0.0156	0.5818 ± 0.0072
	GHL	23	19	0.0091 ± 0.0195	0.3420 ± 0.0056	0.0190 ± 0.0209	0.0109 ± 0.0119	0.3163 ± 0.0281	0.0283 ± 0.0046
	LTDB	4	2.25	0.4457 ± 0.0106	0.7628 ± 0.0038	0.5200 ± 0.0278	0.4125 ± 0.0264	0.7273 ± 0.0081	0.5176 ± 0.0200
	MITDB	11	2	0.3260 ± 0.0246	0.8907 ± 0.0169	0.3934 ± 0.0161	0.3709 ± 0.0076	0.8732 ± 0.0032	0.4271 ± 0.0270
	MSL	14	55	0.3749 ± 0.0271	0.8278 ± 0.0192	0.5059 ± 0.0105	0.3422 ± 0.0108	0.8125 ± 0.0219	0.4266 ± 0.0270
	OPPORTUNITY	7	248	0.1072 ± 0.0267	0.3420 ± 0.0235	0.2827 ± 0.0194	0.1068 ± 0.0030	0.3399 ± 0.0053	0.2016 ± 0.0270
	PSM	1	25	0.1898 ± 0.0184	0.6580 ± 0.0008	0.2710 ± 0.0035	0.1689 ± 0.0201	0.6260 ± 0.0006	0.2479 ± 0.0052
	SMAP	25	25	0.5988 ± 0.0167	0.9185 ± 0.0209	0.5607 ± 0.0197	0.5722 ± 0.0071	0.9328 ± 0.0215	0.6154 ± 0.0075
	SMD	20	38	0.4363 ± 0.0101	0.8930 ± 0.0225	0.5340 ± 0.0197	0.4578 ± 0.0256	0.9026 ± 0.0199	0.5086 ± 0.0173
	SVDB	28	2	0.4585 ± 0.0033	0.9015 ± 0.0113	0.4764 ± 0.0083	0.4603 ± 0.0077	0.8736 ± 0.0292	0.4828 ± 0.0121
	SWaT	2	58.5	0.3006 ± 0.0268	0.7009 ± 0.0191	0.2985 ± 0.0239	0.2829 ± 0.0153	0.7782 ± 0.0175	0.4493 ± 0.0150
	TAO	11	3	0.7562 ± 0.0063	0.8662 ± 0.0218	0.3738 ± 0.0088	0.1271 ± 0.0012	0.6060 ± 0.0195	0.1953 ± 0.0057
TSB-AD-M Average				$\mathbf{0.4642 \pm 0.0134}$	$\mathbf{0.7992 \pm 0.0187}$	$\mathbf{0.4580 \pm 0.0149}$	$\mathbf{0.4260 \pm 0.0112}$	$\mathbf{0.7779 \pm 0.0178}$	$\mathbf{0.4694 \pm 0.0146}$

Table 5: Computational overhead and resource consumption of the proposed method on TSB-AD-U and TSB-AD-M benchmarks. All latency and memory statistics are reported per time-series file.

Benchmark	Avg. Parameters	Avg. Train Time (s)	Avg. Inference Time (s)	Avg. Peak VRAM (MiB)	Max Peak VRAM (MiB)
TSB-AD-U	412,017	6.971	0.2889	993.30	1,053.37
TSB-AD-M	869,352	8.211	0.7526	824.10	1,938.68

size of 512. It is trained independently for 100 AdamW steps with an initial learning rate of 10^{-4} and cosine decay to 10% of the initial rate. Its normal prototype memory contains approximately 10% of the normal patches, bounded between 500 and 2048 prototypes, and anomaly distances are computed from the three nearest prototypes. Inference is performed offline on the complete series, while all normalization, residual scaling, prototype construction, and robust median–MAD calibration use only the normal prefix. We report AUC-ROC, AUC-PR, VUS-ROC, VUS-PR, Range-based F1, and Point-wise F1 over the full series; the VUS metrics use 250 thresholds, and the F1 metrics follow the official oracle-threshold evaluation protocol.

C.4 Per-Dataset Results

Table 4 reports the per-dataset performance of LATF-AD on TSB-AD-U and TSB-AD-M.

C.5 Model Efficiency Analysis

This section reports the computational overhead and resource consumption of our proposed framework across the univariate benchmark (TSB-AD-U) and multivariate benchmark (TSB-AD-M). All measurements were conducted on a single server equipped with an NVIDIA GPU under uniform environment configurations.

We systematically profile five key metrics on a per-file basis: average parameter count, average training latency, average inference latency, average peak allocated GPU memory (VRAM), and the maximum peak GPU memory encountered. The detailed results are summarized in Table 5.

Table 6: Full ablation results across all architectural variants on both TSB-AD-U and TSB-AD-M.

Variant	TSB-AD-U						TSB-AD-M					
	V-P	V-R	R-F	A-P	A-R	P-F	V-P	V-R	R-F	A-P	A-R	P-F
(a) Fusion & Dynamic Routing												
U0: Uniform TF Fusion	0.56	0.88	0.48	0.47	0.85	0.52	0.43	0.77	0.42	0.39	0.75	0.45
(b) Leave-One-View-Out Evaluations												
V1: w/o Short STFT	0.57	0.87	0.48	0.48	0.87	0.53	0.43	0.78	0.43	0.42	0.77	0.45
V2: w/o Long STFT	0.56	0.87	0.46	0.47	0.88	0.52	0.44	0.79	0.42	0.40	0.75	0.45
V3: w/o CWT	0.56	0.87	0.47	0.46	0.89	0.53	0.45	0.78	0.45	0.42	0.76	0.46
V4: w/o WVD	0.56	0.85	0.48	0.46	0.88	0.53	0.45	0.78	0.44	0.41	0.75	0.46
V5: w/o Both STFT Views	0.55	0.85	0.46	0.46	0.87	0.52	0.43	0.78	0.42	0.40	0.75	0.45
(c) Single-View Evaluations												
R1: Only Short STFT	0.53	0.83	0.45	0.44	0.85	0.50	0.42	0.77	0.38	0.37	0.72	0.44
R2: Only Long STFT	0.50	0.83	0.46	0.43	0.84	0.49	0.41	0.76	0.42	0.38	0.73	0.44
R3: Only CWT	0.51	0.84	0.45	0.46	0.84	0.50	0.41	0.74	0.39	0.38	0.74	0.43
R4: Only WVD	0.51	0.83	0.36	0.40	0.83	0.45	0.40	0.75	0.38	0.37	0.72	0.42
(d) Temporal Branch & Output Heads												
B0: w/o Raw Patch Normality Branch	0.56	0.91	0.49	0.48	0.88	0.54	0.44	0.78	0.42	0.39	0.75	0.44
B1: Prediction Head Only	0.56	0.92	0.49	0.48	0.89	0.54	0.45	0.79	0.44	0.41	0.77	0.46
B2: Reconstruction Head Only	0.56	0.91	0.46	0.45	0.89	0.52	0.45	0.79	0.42	0.41	0.77	0.46
Full LATF-AD	**0.60**	**0.93**	**0.51**	**0.50**	**0.91**	**0.56**	**0.46**	**0.80**	**0.46**	**0.43**	**0.78**	**0.47**

Resource Profile and Observations The empirical measurements in Table 5 characterize the computational requirements of our method across the univariate and multivariate settings:

- **Model Parameters:** On the univariate benchmark (TSB-AD-U), the model contains an average of 412,017 parameters. On the multivariate benchmark (TSB-AD-M), the average parameter count scales with input channels to 869,352.
- **Training Latency:** The average training time per time-series file is 6.971 seconds on TSB-AD-U and 8.211 seconds on TSB-AD-M under standard training iterations.
- **Inference Latency:** During evaluation, the average execution time required to process a time-series file is 0.2889 seconds for TSB-AD-U and 0.7526 seconds for TSB-AD-M.
- **GPU Memory Footprint:** The average peak allocated GPU memory is 993.30 MiB on TSB-AD-U and 824.10 MiB on TSB-AD-M. The maximum peak allocated memory across all evaluated files reaches 1,053.37 MiB on TSB-AD-U and 1,938.68 MiB on TSB-AD-M.

Summary In summary, the empirical profiling confirms that our framework maintains a manageable computational footprint across both benchmark settings. Specifically, the model requires fewer than 0.9×10^6 parameters and bounds peak GPU memory consumption within 2.0 GiB even in high-dimensional scenarios. In terms of latency, the framework completes per-file training in 7–9 seconds and achieves sub-second inference throughput (< 0.8 seconds per file), demonstrating its computational feasibility for standard hardware environments and latency-sensitive deployments.

C.6 Full Ablation Studies

Table 6 provides the complete quantitative evaluation across all 13 ablation variants on both TSB-AD-U and TSB-AD-M benchmarks. We observe consistent empirical trends across both benchmarks:

- **Severe Bottlenecks of Single Views**: Restricting the model to any single time-frequency view (R1–R4) substantially undermines detection power, with VUS-PR dropping to 0.50–0.53 on TSB-AD-U and 0.40–0.42 on TSB-AD-M. In particular, relying solely on WVD (R4) collapses Range-F1 to 0.36 on TSB-AD-U due to unmitigated quadratic cross-term interference, confirming that isolated operators cannot break the time-frequency trade-off.
- **Necessity of Each Time-Frequency View**: In the leave-one-view-out evaluations (V1–V4), excluding any single spectral representation consistently impairs detection accuracy compared to the full model. Specifically, removing WVD (V4) degrades VUS-ROC to 0.85 on TSB-AD-U and AUC-PR to 0.41 on TSB-AD-M, while dropping Short STFT (V1) or

Long STFT (V2) leads to pronounced drops across range-wise and point-wise metrics. Concurrently removing both STFT operators (V5) precipitates the steepest drop in primary VUS-PR (0.55 on U and 0.43 on M), demonstrating that multi-scale linear transforms and high-resolution quadratic distributions provide indispensable, complementary spectral evidence.

- **Dual Heads and Temporal Prototype Grounding**: Isolating either the predictive or reconstructive head (B1, B2) or eliminating the raw patch prototype memory (B0) leads to consistent degradation across both univariate and multivariate settings, corroborating that cross-domain verification and multi-objective supervision are mutually reinforcing.