FSA-AD: Frequency-Semantic Anomaly Detection with Adaptive Patching and Orthogonal Prototypes

Limin Liu, Yiran Qian, Qianqian Qi, Jun Long, and Yueyi Luo*
Central South University, Changsha, China

***Abstract*—Abnormal patterns in time series often manifest heterogeneously across frequency components: low-frequency signals encode long-term trends and regime shifts, whereas high-frequency signals capture short-term fluctuations and local structural variations. However, prevailing unsupervised anomaly detection methods typically employ uniform modeling granularities across frequency components, leading to suboptimal feature representation and semantic entanglement. To address these limitations, we propose FSA-AD, a Frequency-Semantic Adaptive framework. Central to our approach is a frequency-adaptive patching mechanism that explicitly aligns modeling resolution with spectral densities, enabling the simultaneous capture of global regimes and fine-grained morphological transients. Furthermore, we introduce a dual-stream orthogonal prototype memory to enforce a disjoint information bottleneck. This design effectively mitigates the "identity mapping" paradox and prevents semantic leakage between frequency components. Extensive experiments on real-world and synthetic benchmarks demonstrate that FSA-AD achieves highly competitive performance by effectively disentangling frequency-specific anomaly patterns.**

***Index Terms*—Time Series Anomaly Detection;Frequency-Semantic Modeling; Adaptive Patching; Prototype Learning**

I. Introduction

Time series anomaly detection (TSAD) is a fundamental task in a wide range of real-world applications, including industrial monitoring, energy systems, finance, and cybersecurity [1]. With the exponential growth of sensor networks, these systems generate massive streams of high-dimensional data characterized by intricate temporal dependencies, rendering manual monitoring infeasible. Consequently, given the prohibitive cost of obtaining precise anomaly labels, unsupervised learning has emerged as the dominant paradigm for automating the detection of deviations from normal system dynamics.

However, reliable detection remains elusive due to a pathological phenomenon known as over-reconstruction, where deep models undesirably generalize to anomalous patterns. We argue that the root cause of this failure lies in the overlooked spectral bias inherent in time series data. As illustrated in Fig. 1, signal energy follows a long-tailed distribution: optimization objectives are disproportionately dominated by high-amplitude low-frequency components, while high-frequency details contribute negligibly to the total loss [2]. This imbalance creates a critical optimization hazard: to minimize global error, the model prioritizes fitting the dominant trends and tends to allocate its remaining capacity to trivially memorize high-frequency inputs via an "identity mapping" rather than learning their

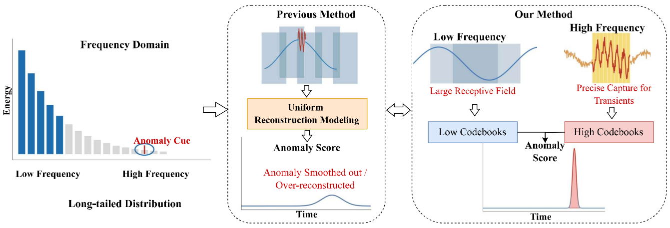


Fig. 1: Motivation: Limitations of Uniform Modeling under Long-Tailed Frequency Distributions.

underlying generative laws. Consequently, since the model acts as a simple "copy-paste" function for high-frequency bands, it can reconstruct subtle structural anomalies just as well as normal patterns [3], rendering the reconstruction error indistinguishable and leading to significant detection failures.

To dismantle this spectral bias and mitigate the identity mapping paradox, we propose **FSA-AD** (Frequency-Semantic Adaptive Anomaly Detection). The core philosophy of FSA-AD is to disentangle the modeling process. Specifically, we introduce a **Frequency-Adaptive Patching** mechanism that dynamically aligns modeling granularity with spectral density—assigning large receptive fields to capture global trends and fine-grained resolutions to scrutinize local transients. Furthermore, to explicitly prevent the model from degenerating into an identity function, we design a **Dual-Stream Orthogonal Prototype Memory**. This module enforces a disjoint information bottleneck, restricting the feature space to a finite set of normal orthogonal patterns, thereby suppressing the reconstruction of anomalous behaviors that fall outside the memory manifold.

Our main contributions are summarized as follows: **(1)** We identify frequency-semantic heterogeneity as a fundamental challenge in time series anomaly detection and show that uniform modeling across frequencies inherently limits detection sensitivity. **(2)** We propose FSA-AD, a frequency-semantic-aware framework that aligns temporal modeling strategies with the intrinsic roles of low- and high-frequency components. **(3)** We introduce dual-stream orthogonal prototype codebooks that enforce a disjoint reconstruction bottleneck, effectively mitigating over-reconstruction of anomalies. **(4)** Extensive experiments on real-world benchmarks demonstrate that FSA-AD achieves competitive and robust performance across diverse anomaly types.

*Corresponding author. E-mail: luoyueyi@csu.edu.cn.

II. Related Work

A. Reconstruction-based and Memory-augmented Methods

Reconstruction-based methods form the dominant paradigm for unsupervised time series anomaly detection. Representative approaches, such as OmniAnomaly [4], USAD [5], Anomaly Transformer [6], TimesNet [7], and ModernTCN [8], leverage deep architectures to capture long-range dependencies. However, they are prone to over-reconstruction [9], where anomalies are reconstructed with high fidelity. This is especially true for pattern-wise anomalies, which exhibit normal individual observations but abnormal collective temporal structures. Memory-augmented approaches like MemAE [10] attempt to mitigate this by constraining reconstruction through learned normal prototypes. Yet, most rely on shared latent representations, remaining susceptible to semantic entanglement across temporal scales and limiting robustness against structurally subtle anomalies.

B. Spectral and Multi-scale Analysis

To overcome the limitations of uniform temporal modeling, recent studies explore spectral and multi-scale representations. Frequency-based methods (e.g., TFAD [11], FITS [12]) and multi-scale architectures (e.g., Pyraformer [13], TimeMixerr [14]) model temporal dependencies at different resolutions. While effective for periodic patterns, most adopt fixed and uniform resolutions across temporal or frequency components. This implicitly assumes homogeneous anomaly characteristics across scales, causing dominant low-frequency structures to overshadow localized or high-frequency deviations. This mismatch between anomaly characteristics and modeling granularity remains a key limitation that our FSA-AD framework aims to resolve.

III. Methodology

In the context of time series anomaly detection, $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\} \in \mathbb{R}^{T\times D}$ denote a multivariate time series of length T with D variables, where $\mathbf{x}_t \in \mathbb{R}^D$ represents the observation at timestamp t. Time series anomaly detection can be defined as: given input time series sequence $\mathcal{X}$, for another unknown test time series $\mathbf{X}_{test}$ of length T' with the same modality as the training sequence, we aim to output $\hat{\mathbf{y}}_{test} = (y_1, y_2, \dots, y_{T'})$. Here $y_t \in \{0, 1\}$ denotes whether the observation x_t at time t in the test time series is an anomaly or not.

A. Overall Architecture

As illustrated in Fig. 2, FSA-AD follows a reconstruction-based paradigm, fundamentally designed to dismantle spectral bias via frequency-semantic disentanglement. The proposed architecture is composed of three core components: (1) Frequency-Adaptive Disentanglement: Utilizing Discrete Wavelet Transform (DWT) [15], we bifurcate raw signals into low-frequency approximation and high-frequency detail components, which are then tokenized using granularity-specific patching resolutions to align with their respective spectral densities. (2) Cross-Frequency Contextual Interaction: A bidirectional cross-attention mechanism facilitates semantic calibration, ensuring that local transients are validated against global trends. (3) Dual-Stream Orthogonal Prototype Codebooks: To mitigate the identity mapping paradox, features are reconstructed via sparse addressing over disjoint, orthogonal codebooks, compelling the model to represent inputs as combinations of learned normal patterns.

B. Frequency-Adaptive Disentanglement

To resolve the scale-resolution mismatch, FSA-AD adopts a physics-aware decomposition strategy that aligns data representation with the spectral properties of time series. Given a multivariate input $\mathbf{X} \in \mathbb{R}^{B\times D\times T}$, where B is the batch size, D is the number of variables, and T is the sequence length, we first utilize the Discrete Wavelet Transform (DWT) to bifurcate the signal into approximation (low-frequency) and detail (high-frequency) coefficients. Due to the downsampling inherent in DWT, this yields condensed components $\mathbf{X}_{\text{low}}, \mathbf{X}_{\text{high}} \in \mathbb{R}^{B\times D\times \frac{T}{2}}$, where $\mathbf{X}_{\text{low}}$ encapsulates global trends (spectral "head") and $\mathbf{X}_{\text{high}}$ preserves transient fluctuations (spectral "tail").

To capture these distinct dynamics efficiently, we apply a **Frequency-Adaptive Patching** mechanism. Specifically, we assign a larger patch size P_l with stride S_l to $\mathbf{X}_{\text{low}}$ for semantic context, and a smaller patch size P_h (with $P_h < P_l$) with stride S_h to $\mathbf{X}_{\text{high}}$ for local precision. Consequently, the number of patches N_k for each frequency component $k \in \{\text{low}, \text{high}\}$ is determined by $N_k = \left\lfloor \frac{T/2 - P_k}{S_k} \right\rfloor + 1$. Through sliding window operations and dimension permutation, the frequency components are transformed into structured patch sequences $\mathcal{P}_{\text{low}} \in \mathbb{R}^{B\times N_l \times D\times P_l}$ and $\mathcal{P}_{\text{high}} \in \mathbb{R}^{B\times N_h\times D\times P_h}$. Finally, frequency-specific linear projections ϕ are employed to map these heterogeneous granularities into a unified latent space:

$$\mathbf{Z}_{\text{low}} = \phi_{\text{low}}(\mathcal{P}_{\text{low}}) \tag{1}$$

$$\mathbf{Z}_{\text{high}} = \phi_{\text{high}}(\mathcal{P}_{\text{high}}). \tag{2}$$

The resulting embeddings $\mathbf{Z}_{low} \in \mathbb{R}^{B\times N_l\times D\times d}$ and $\mathbf{Z}_{high} \in \mathbb{R}^{B\times N_h\times D\times d}$ serve as frequency-aware patch embeddings for subsequent cross-frequency interaction and prototype-based reconstruction.

C. Cross-Frequency Contextual Interaction

Following patch-wise embedding, the extracted representations are independently processed to capture intrinsic spatio-temporal dependencies within each frequency domain. Let $\mathbf{Z}_k \in \mathbb{R}^{B\times N_k\times D\times d}$ denote the input feature tensor for frequency stream $k \in \{low, high\}$, where N_k is the number of patches (with $N_l \neq N_h$). We employ stream-specific Gated Mixer Encoders to refine these features via two orthogonal mixing steps:

To capture the inter-sensor correlations, we first perform mixing along the multivariate dimension D. We transpose the tensor to align D with the projection axis and apply a

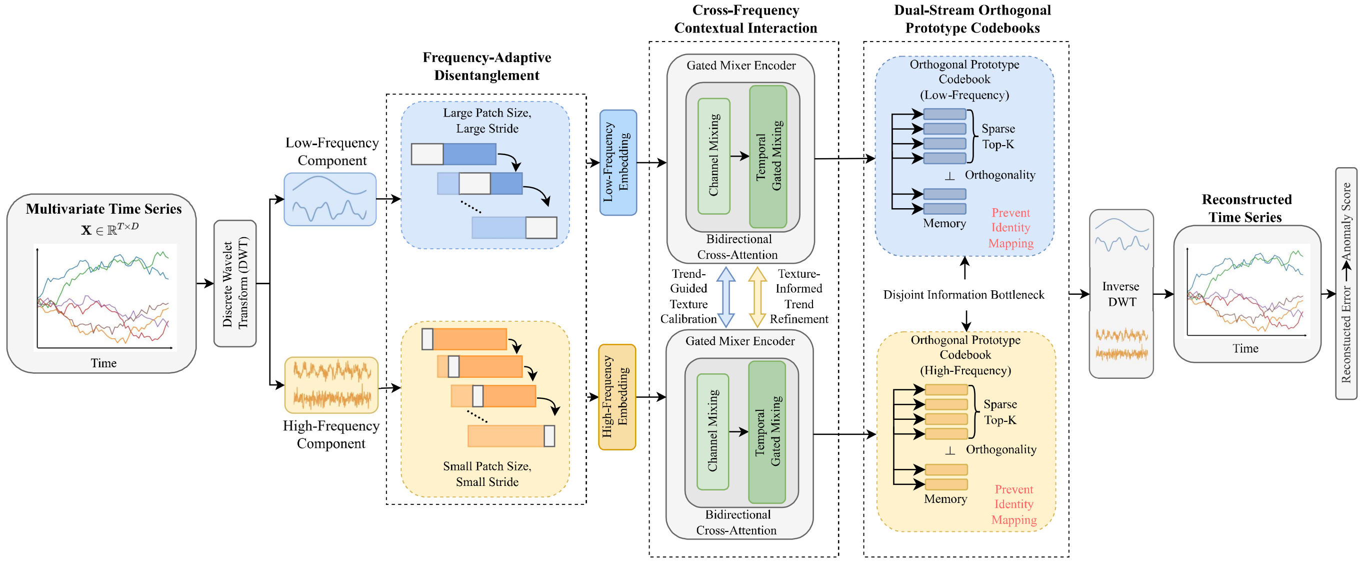


Fig. 2: Overview of FSA-AD with detailed views of Frequency-Adaptive Patching and Orthogonal Prototype Memory.

shared MLP. For an input $\mathbf{Z}_k$, the channel-mixed output $\mathbf{Z}_k^{(c)} \in \mathbb{R}^{B\times N_k \times D\times d}$ is computed as:

$$\mathbf{Z}_k^{(c)} = \mathbf{Z}_k + \text{MLP}_{\text{channel}}\left(\mathbf{Z}_k.\text{permute}(\dots, M)\right) \quad (3)$$

where $\text{MLP}_{\text{channel}}$ is applied independently to each patch and mixes information along the variable dimension C.

Subsequently, to capture the evolutionary dynamics, we perform mixing along the patch dimension N_k. Unlike standard linear mixing, we employ a Gated MLP mechanism to dynamically highlight informative temporal segments. The tensor is transposed to operate on the sequence length N_k, producing the refined intra-frequency representation $\mathbf{Z}_k^{(t)} \in \mathbb{R}^{B\times N_k\times D\times d}$:

$$\mathbf{Z}_k^{(t)} = \mathbf{Z}_k^{(c)} + \left(\mathbf{W}_{val}\mathbf{Z}_k^{(c)}\right) \odot \sigma\left(\mathbf{W}_{gate}\mathbf{Z}_k^{(c)}\right) \quad (4)$$

where $\mathbf{W}_{val}, \mathbf{W}_{gate}$ are learnable linear projections, σ denotes GELU activation, and $\odot$ represents element-wise multiplication.

Although the Gated Mixers effectively refine spatio-temporal features within each stream, the low-frequency trends and high-frequency textures remain semantically isolated. However, anomalous events often manifest as contextual discrepancies—such as a high-frequency spike that is considered normal during a volatile regime but anomalous during a stable trend. To bridge this semantic gap, we introduce a Cross-Frequency Contextual Interaction module.

Prior to interaction, we adopt a channel-independent strategy to focus exclusively on temporal semantic alignment. We reshape the refined features $\mathbf{Z}_{low}^{(t)}$ and $\mathbf{Z}_{high}^{(t)}$ by folding the multivariate dimension D into the batch dimension B. This yields the flattened representations $\hat{\mathbf{Z}}_k \in \mathbb{R}^{(B\cdot D)\times N_k\times d}$, allowing the attention mechanism to verify temporal consistency for each variable independently while sharing valid interaction patterns across all sensors.

Bidirectional Semantic Calibration: We utilize a dual Cross-Attention mechanism to mutually calibrate the heterogeneous frequency streams. Crucially, this mechanism naturally resolves the scale mismatch between the coarse-grained trends (N_{low}) and fine-grained textures (N_{high}) via the learnable alignment matrix. Mathematically, let $\text{Attn}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{Softmax}(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}})\mathbf{V}$ denote the standard scaled dot-product attention. The calibrated representations $\mathbf{H}_k \in \mathbb{R}^{(B\cdot D)\times N_k\times d}$ are computed as:

$$\mathbf{H}high = \hat{\mathbf{Z}}high + \text{Attn}(\hat{\mathbf{Z}}high, \hat{\mathbf{Z}}low, \hat{\mathbf{Z}}low) \quad (5)$$

$$\mathbf{H}low = \hat{\mathbf{Z}}low + \text{Attn}(\hat{\mathbf{Z}}low, \hat{\mathbf{Z}}high, \hat{\mathbf{Z}}high) \quad (6)$$

Through this bidirectional calibration, $\mathbf{H}_{high}$ suppresses noise that aligns with the trend (contextual verification), while $\mathbf{H}_{low}$ absorbs structural evidence. Finally, the outputs are reshaped back to $\mathbb{R}^{B\times N_k\times D\times d}$ to restore the multivariate structure, serving as the input for the subsequent orthogonal memory module.

D. Dual-Stream Orthogonal Prototype Codebooks

While the previous module aligns features contextually, the reconstruction network may still degenerate into a trivial identity mapping, generalizing well to anomalous patterns due to the powerful expressiveness of deep networks. To impose a strict information bottleneck, we introduce **Dual-Stream Orthogonal Prototype Codebooks**. By maintaining disjoint memory subspaces for low-frequency trends and high-frequency textures, we ensure that anomalies—which differ structurally or semantically from normal patterns—cannot be accurately reconstructed, thereby amplifying the detection signal.

For each frequency stream, we maintain an independent prototype codebook

$$\mathcal{P}_k = \left\{\mathbf{p}_k^{(1)}, \dots, \mathbf{p}_k^{(K)}\right\}, \quad \mathbf{p}_k^{(i)} \in \mathbb{R}^d, \quad (7)$$

where K is the number of prototypes. To enhance diversity at initialization, the prototype vectors are orthogonally initialized. During inference, each patch-level feature vector is reconstructed via sparse addressing over the corresponding codebook.

Specifically, we flatten the input features across batch, patch, and variable dimensions:

$$\mathbf{q}_k = \mathrm{reshape}(\mathbf{H}_k) \in \mathbb{R}^{(BN_kD)\times d}, \tag{8}$$

and compute cosine similarity between normalized queries and prototypes:

$$\mathbf{S}_k = \frac{\mathbf{q}_k}{\|\mathbf{q}_k\|} \cdot \frac{\mathcal{P}_k^\top}{\|\mathcal{P}_k\|} \in \mathbb{R}^{(BN_kD)\times K}, \tag{9}$$

To enforce sparse reconstruction and reduce the risk of anomaly interpolation, we apply **Top-**K' hard selection on similarity scores and retain only the most relevant prototypes. The resulting attention weights are sharpened via temperature-scaled softmax:

$$\boldsymbol{\alpha}_k = \mathrm{Softmax}\left(\frac{\mathrm{TopK}(\mathbf{S}_k)}{\tau}\right), \tag{10}$$

where τ is a small temperature controlling sparsity. The reconstructed features are then obtained as a weighted combination of selected prototypes:

$$\hat{\mathbf{q}}_k = \boldsymbol{\alpha}_k \mathcal{P}_k, \quad \hat{\mathbf{H}}_k = \mathrm{reshape}(\hat{\mathbf{q}}_k). \tag{11}$$

To ensure the effectiveness and stability of the prototype memory, we introduce three auxiliary regularization terms:

Diversity Loss: To maximize the expressiveness of the codebook and prevent redundant prototypes, we force the prototype vectors to be orthogonal to each other:

$$\mathcal{L}_{\mathrm{div}} = \left\| \frac{\mathcal{P}_k}{\|\mathcal{P}_k\|} \cdot \frac{\mathcal{P}_k^\top}{\|\mathcal{P}_k\|} - \mathbf{I} \right\|_F^2 \tag{12}$$

Commitment Loss: To minimize the gap between the encoder output and the stored normal patterns:

$$\mathcal{L}_{\mathrm{com}} = \|\mathbf{q}_k - \mathrm{sg}(\hat{\mathbf{q}}_k)\|_2^2 + \beta\|\mathrm{sg}(\mathbf{q}_k) - \hat{\mathbf{q}}_k\|_2^2 \tag{13}$$

where $\mathrm{sg}(\cdot)$ denotes the stop-gradient operator.

Entropy Regularization: To encourage near one-hot prototype selection:

$$\mathcal{L}_{\mathrm{ent}} = -\mathbb{E}\left[\sum_{i=1}^{K} \alpha_k^{(i)} \log(\alpha_k^{(i)} + \epsilon)\right] \tag{14}$$

The overall memory regularization loss is defined as:

$$\mathcal{L}_{\mathrm{mem}} = \mathcal{L}_{\mathrm{div}} + \lambda_{\mathrm{com}}\mathcal{L}_{\mathrm{com}} + \lambda_{\mathrm{ent}}\mathcal{L}_{\mathrm{ent}}. \tag{15}$$

By enforcing orthogonality, the memory bank learns a set of independent basis vectors for normal patterns. The Top-K sparsity and Entropy loss ensure that any input must be represented by a very small number of these bases. Consequently, an anomalous input, which lies outside the span of these normal orthogonal bases, results in a high reconstruction error, effectively functioning as an anomaly indicator.

TABLE I: Statistics of the Benchmark Datasets

Datasets	Applications	Channels	Anomalies (%)
CalIt2	Visitors flowrate	2	4.09
GECCO	Water treatment	9	1.25
PSM	Server Machine	25	11.07
SWAT	Water treatment	51	5.78
TODS	Synthetic	5	6.35

IV. Experiments

A. Experimental Settings

Datasets: We evaluate FSA-AD on a diverse benchmark including real-world multivariate time series datasets—CalIt2 [16], GECCO [17], PSM [18], and SWaT [19]—covering human flow monitoring, industrial control systems, server metrics, and critical infrastructure. In addition, we adopt synthetic datasets generated by the TODS framework [20], which inject controlled anomalies of different types (e.g., gloal, contextual, shapelet, seasonal, and trend anomalies) to enable systematic evaluation across anomaly morphologies. Detailed statistics and descriptions of these benchmark datasets are provided in Table I.

Baselines: We benchmark FSA-AD against eight representative methods, including the latest state-of-the-art (SOTA) models.These baselines feature the 2024 SOTA ModernTCN [8], along with the 2023 SOTA Anomaly Transformer [6], TimesNet [7], PatchTST [21], DLinear [22]. Additionally, we include non-learning methods such as One-Class SVM (OCSVM) [23], Principal Component Analysis (PCA) [24], and Histogram-based Outlier Score (HBOS) [25]. All baselines are evaluated using official or widely adopted implementations under consistent settings.

Metrics: To avoid the performance inflation caused by the "Point Adjustment" protocol [6], [26], we explicitly adopt Affiliation F1 (Aff-F1) [27] and VUS-ROC [28] as our primary metrics. Aff-F1 assesses the quality of interval-level alignment between predictions and ground truth , while VUS-ROC evaluates the intrinsic separability of anomaly scores across varying window sizes, ensuring a rigorous and unbiased assessment.

B. Main Results

As presented in Table II, FSA-AD demonstrates superior generalization capability, achieves highly competitive performance across diverse real-world and synthetic benchmarks. Notably, on the complex GECCO dataset, our model achieves a commanding VUS-ROC of 0.982, exhibiting exceptional discriminative power compared to the Transformer-based baseline (0.503), thereby validating the robustness of the proposed orthogonal memory against diverse signal dynamics. In terms of precise fault localization, FSA-AD attains near-perfect Aff-F1 scores on Shapelet (0.987) and Seasonal (0.999) anomalies; this empirical evidence corroborates that our frequency-adaptive patching mechanism effectively resolves the resolution-mismatch problem, capturing fine-grained morphological transients that are often smoothed out by uniform modeling strategies. Furthermore, distinct from deep reconstruction baselines like ModernTCN,

TABLE II: Aff-F1 (Affiliated-F1) and VUS-ROC accuracy measures for 4 real-world datasets and 5 synthetic datasets of different types of anomalies. The best results are highlighted in **bold**, and the second-best results are underlined.

Dataset	Metric	FSA-AD	PCA	OCSVM	HBOS	DLinear	TimesNet	PatchTST	ModernTCN	ATransformer
CalIt2	Aff-F1	**0.794**	0.768	0.783	0.756	0.793	**0.794**	0.793	0.780	0.729
	VUS-ROC	0.813	0.786	0.828	**0.831**	0.781	0.796	0.824	0.716	0.517
GECCO	Aff-F1	0.844	0.785	0.666	0.708	0.893	0.894	**0.906**	0.893	0.782
	VUS-ROC	**0.982**	0.595	0.756	0.503	**0.982**	0.974	0.979	0.975	0.503
PSM	Aff-F1	0.834	0.702	0.531	0.658	0.831	**0.842**	0.831	0.825	0.710
	VUS-ROC	0.584	0.585	0.532	0.575	0.579	**0.593**	0.585	0.589	0.451
SWaT	Aff-F1	0.763	0.708	0.703	0.650	0.736	**0.793**	0.716	0.728	0.718
	VUS-ROC	0.656	0.687	0.618	**0.784**	0.581	0.392	0.346	0.348	0.434
Global	Aff-F1	0.932	0.704	0.849	0.528	0.928	0.909	**0.939**	0.748	0.656
	VUS-ROC	0.967	0.653	**0.970**	0.489	0.939	0.962	0.966	0.772	0.430
Contextual	Aff-F1	0.659	0.474	0.686	0.447	**0.770**	0.665	0.761	0.616	0.577
	VUS-ROC	**0.847**	0.401	0.563	0.351	0.567	0.819	0.746	0.442	0.406
Shapelet	Aff-F1	**0.987**	0.680	0.751	0.635	0.943	0.917	0.925	0.673	0.658
	VUS-ROC	**0.891**	0.601	0.727	0.485	0.712	0.875	0.844	0.599	0.518
Seasonal	Aff-F1	**0.999**	0.606	0.790	0.661	0.993	0.990	0.984	0.679	0.718
	VUS-ROC	0.929	0.544	0.838	0.602	0.839	**0.932**	0.929	0.596	0.567
Trend	Aff-F1	0.850	0.671	0.668	0.665	0.724	0.832	**0.855**	0.827	0.640
	VUS-ROC	0.860	0.460	0.458	0.582	0.818	0.883	**0.895**	0.843	0.504

TABLE III: Component-wise ablation study of FSA-AD on three representative datasets with VUS-ROC.

Variant	GECCO	SWaT	Shapelet
FSA-AD (Full)	0.982	0.656	0.891
w/o Adaptive Patching	0.912	0.556	0.770
w/o Cross-Freq Interaction	0.898	0.588	0.863
w/o Orthogonal Prototypes	0.826	0.517	0.786

which suffer significant performance degradation on Contextual anomalies (VUS-ROC 0.442) due to the "identity mapping" paradox, FSA-AD maintains a robust score of 0.847. This sharp contrast underscores the critical role of the disjoint information bottleneck in enforcing semantic consistency, ensuring that contextually incongruent patterns yield high reconstruction errors rather than being trivially memorized.

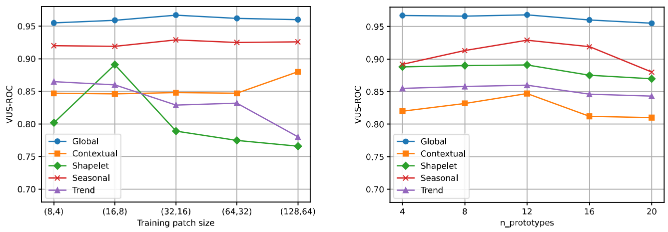
C. Model Analysis

Ablation study: To validate the contribution of each module, we performed an ablation study on the representative GECCO, SWaT, and Shapelet datasets (Table III). The removal of adaptive patching caused the most significant degradation on the Shapelet dataset, confirming that aligning patch resolution with spectral semantics is critical for capturing fine-grained morphological details. Furthermore, eliminating cross-frequency interaction consistently reduced detection robustness, while discarding orthogonal prototypes led to severe performance drops on GECCO and SWaT, validating the necessity of disjoint memory constraints in preventing reconstruction overgeneralization. The superior performance of the full model confirms the synergistic efficacy of these components.

Model efficiency: Table IV reports the efficiency comparison among different methods in terms of model size and inference cost. Overall, FSA-AD achieves competitive efficiency while delivering superior detection performance. Despite incorporating frequency-adaptive modeling and prototype-based constraints,

TABLE IV: Comparison of various methods with respect to training time, inference time and VUS-ROC metric result.

Method	Training Time (s)	Inference Time (s)	VUS-ROC
ModernTCN	**22.75**	0.57	0.975
TimesNet	342.79	1.68	0.974
ATrans	270.49	15.05	0.503
FSA-AD	259.11	**0.12**	**0.982**



(a) Impact of Patch Size (b) Impact of Prototypes Number

Fig. 3: Parameter sensitivity analysis on the TODS dataset. (a) Impact of patch size (P_{low}, P_{high}). (b) Impact of prototype number K.

FSA-AD maintains a compact parameter footprint, remaining comparable to lightweight baselines and significantly more efficient than attention-based models. This efficiency is primarily attributed to the use of MLP-based mixers and patch-wise processing, which avoid the quadratic complexity associated with self-attention mechanisms.These results suggest that FSA-AD is well suited for practical anomaly detection scenarios with resource or latency constraints.

Parameter Sensitivity: We evaluate the impact of patch resolution (P_{low}, P_{high}) and prototype count (K) on the TODS datasets, as shown in Fig. 3. The model remains stable across a wide range of patch configurations when low- and high-frequency components are assigned different granularities, with moderate patch sizes achieving consistently strong performance. Smaller patches favor localized anomalies such as Shapelet and Contextual, while larger low-frequency patches better capture long-term structures in Global and Trend anomalies; overly large patches may slightly degrade performance on localized patterns. Increasing the number of prototypes improves performance up to a moderate scale, after which gains saturate or slightly decline, indicating that excessive prototypes weaken the reconstruction bottleneck.

V. Conclusion

In this work, we addressed the limitations of uniform temporal modeling and feature entanglement in unsupervised time series anomaly detection by proposing FSA-AD, a Frequency-Semantic Adaptive framework. By introducing a frequency-adaptive patching mechanism, FSA-AD explicitly aligns modeling granularity with spectral densities, allowing for the simultaneous capture of global trends and fine-grained morphological transients. Furthermore, we designed a dual-stream orthogonal prototype memory to enforce a disjoint information bottleneck, effectively mitigating the "identity mapping" paradox and preventing semantic leakage between frequency components. Extensive experiments on multiple real-world and synthetic benchmarks demonstrate that FSA-AD achieves highly competitive performance. Future work will explore extending this frequency-adaptive paradigm to self-supervised pre-training on large-scale time series and investigating dynamic mechanisms for automatic frequency band segmentation.

Acknowledgment

This work was supported by the project "Research on the Construction of Full-Process Moisture Chain Model and Precision Technology" under Grant KY2025ZB0008.

References

[1] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM computing surveys (CSUR)*, vol. 54, no. 3, pp. 1–33, 2021.

[2] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, "On the spectral bias of neural networks," in *International conference on machine learning*, pp. 5301–5310, PMLR, 2019.

[3] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi, "Deep learning for time series anomaly detection: A survey," *ACM Computing Surveys*, vol. 57, no. 1, pp. 1–42, 2024.

[4] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. 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.

[5] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. 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.

[6] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," *arXiv preprint arXiv:2110.02642*, 2021.

[7] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," *arXiv preprint arXiv:2210.02186*, 2022.

[8] D. Luo and X. Wang, "Moderntcn: A modern pure convolution structure for general time series analysis," in *The twelfth international conference on learning representations*, pp. 1–43, 2024.

[9] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.

[10] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1705–1714, 2019.

[11] C. Zhang, T. Zhou, Q. Wen, and L. Sun, "Tfad: A decomposition time series anomaly detection architecture with time-frequency analysis," in *Proceedings of the 31st ACM international conference on information & knowledge management*, pp. 2497–2507, 2022.

[12] Z. Xu, A. Zeng, and Q. Xu, "Fits: Modeling time series with $10k$ parameters," *arXiv preprint arXiv:2307.03756*, 2023.

[13] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu, and S. Dustdar, "Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting," in *# PLACEHOLDER_PARENT_METADATA_VALUE#*, 2022.

[14] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, "Timemixer: Decomposable multiscale mixing for time series forecasting," *arXiv preprint arXiv:2405.14616*, 2024.

[15] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 674–693, 2002.

[16] A. Ihler, J. Hutchins, and P. Smyth, "Adaptive event detection with time-varying poisson processes," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 207–216, 2006.

[17] V. H. A. Ribeiro and G. Reynoso-Meza, "Online anomaly detection for drinking water quality using a multi-objective machine learning approach," in *Proceedings of the genetic and evolutionary computation conference companion*, pp. 1–2, 2018.

[18] A. Abdulaal, Z. Liu, and T. Lancewicki, "Practical approach to asynchronous multivariate time series anomaly detection and localization," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 2485–2494, 2021.

[19] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A dataset to support research in the design of secure water treatment systems," in *International conference on critical information infrastructures security*, pp. 88–99, Springer, 2016.

[20] K.-H. Lai, D. Zha, G. Wang, J. Xu, Y. Zhao, D. Kumar, Y. Chen, P. Zumkhawaka, M. Wan, D. Martinez, *et al.*, "Tods: An automated time series outlier detection system," in *Proceedings of the aaai conference on artificial intelligence*, vol. 35, pp. 16060–16062, 2021.

[21] Y. Nie, "A time series is worth 64words: Long-term forecasting with transformers," *arXiv preprint arXiv:2211.14730*, 2022.

[22] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, pp. 11121–11128, 2023.

[23] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," *Advances in neural information processing systems*, vol. 12, 1999.

[24] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang, "A novel anomaly detection scheme based on principal component classifier," 2003.

[25] M. Goldstein and A. Dengel, "Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm," *KI-2012: poster and demo track*, vol. 1, pp. 59–63, 2012.

[26] Y. Yang, C. Zhang, T. Zhou, Q. Wen, and L. 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*, pp. 3033–3045, 2023.

[27] A. Huet, J. M. Navarro, and D. Rossi, "Local evaluation of time series anomaly detection algorithms," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 635–645, 2022.

[28] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin, "Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection," *Proceedings of the VLDB Endowment*, vol. 15, no. 11, pp. 2774–2787, 2022.