MSDDS: Multivariate Time Series Anomaly Detection via Multi-Scale Decomposition with Fusion and Dual-Stream Channel Transformer

Yuying Ma[1], Jiong Yu[1*], Shu Li[1], Yong Hu[2]
[1]School of Computer Science and Technology, Xinjiang University, Urumqi, China
[2]School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China
{mayuying, lsmiao}@stu.xju.edu.cn, yujiong@xju.edu.cn, 13627223892@163.com

***Abstract*—Multivariate time series anomaly detection is widely used in applications such as industrial monitoring, financial risk control, and cyber-physical system security. However, real-world multivariate time series usually exhibit non-stationary temporal dynamics and complex channel correlations, making anomaly detection particularly challenging. Surpassing existing time-frequency integrated approaches that typically process dual-domain features in parallel, we revisit anomaly detection from a structure-enhanced perspective, where temporal dynamics are first explicitly refined through multi-scale decomposition to serve as a robust basis for subsequent frequency-domain modeling. Accordingly, we propose MSDDS, which integrates a Multi-Scale Decomposition with Fusion (MSDF) module that leverages a dynamic adaptive strategy to integrate diverse temporal patterns, and a Dual-Stream Channel Transformer (DSCT) for modeling complementary global and local channel dependencies in the frequency domain. Extensive experiments on seven widely used benchmark datasets demonstrate that MSDDS achieves state-of-the-art performance on most benchmarks, validating its effectiveness and robustness in unsupervised multivariate time series anomaly detection.**

***Index Terms*—anomaly detection, multi-scale decomposition, channel attention, frequency-domain analysis, multivariate time series**

I. Introduction

With the rapid advancement of Industry 4.0 and cyber-physical systems, large-scale sensor networks generate massive multivariate time series data [1], [2], [12]. Timely detection of anomalies in these measurements is crucial for ensuring system reliability and safety, yet remains challenging due to the non-stationary dynamics and diverse anomaly patterns. While point anomalies (extreme values) are relatively easy to detect, *subsequence anomalies*—manifesting as subtle shape distortions, seasonal disruptions, or trend drifts—are substantially more difficult to identify, as their individual observations often fall within normal statistical ranges [3], [4].

Early research relied on statistical models such as ARIMA [5] and classical detectors like OC-SVM [6], which identify anomalies via probabilistic deviations or geometric boundaries. However, these shallow models struggle with non-stationary dynamics and high dimensionality. Consequently, deep reconstruction approaches, encompassing Autoencoders [7]–[9] and Transformers [10], [11], have emerged as the dominant methodology, significantly improving the modeling of temporal dependencies.

Despite these advancements, many recent anomaly detection methods, such as TimesNet [40], still operate in the time domain and struggle with subsequence anomalies, where local shapes are distorted but statistical properties remain within normal ranges. To address this, recent works incorporate time-frequency analysis via FFT or wavelet transforms [22], [42], leveraging spectral signatures to expose periodic disruptions and shape anomalies masked in the time domain. However, two key challenges remain: (1) frequency-domain analysis is often applied directly to raw sequences without explicit decomposition of non-stationary components, limiting interpretability; (2) effective cross-channel dependency modeling remains a dilemma, as channel-independent methods [23] ignore correlations while channel-dependent approaches [25] are computationally expensive and noise-sensitive.

To address the challenges of non-stationary temporal dynamics and complex channel dependencies, we propose Multi-Scale Decomposition with Fusion and Dual-Stream Channel Transformer (MSDDS), a time-frequency collaborative framework for multivariate time series anomaly detection. MSDDS enhances temporal structures via multi-scale decomposition and models inter-channel dependencies in the frequency domain, enabling robust detection of diverse subsequence anomalies.

Our main contributions are summarized as follows:

- We propose MSDDS, a unified time-frequency anomaly detection framework that enhances temporal structures via multi-scale decomposition before frequency-domain modeling.
- We design a Multi-Scale Decomposition with Fusion (MSDF) module with adaptive weighting to fuse multi-scale trend and seasonal components.
- We introduce a Dual-Stream Channel Transformer (DSCT) with a Dual-Stream Channel Attention (DSCA) mechanism to model both global and local channel depen-

* Corresponding author: Jiong Yu.

dencies, achieving state-of-the-art performance on seven benchmarks.

II. Related Works

A. Multivariate Time Series Anomaly Detection

Early anomaly detection methods relied on statistical models such as ARIMA [5] and classical detectors such as OC-SVM [6], which struggle with complex and non-linear multivariate dynamics. Deep learning-based methods have since been widely adopted, including forecasting models such as LSTM-NDT [13] and DeepAnT [14], reconstruction-based approaches such as OmniAnomaly [8], Donut [15], and USAD [16], and Transformer-based architectures such as Anomaly Transformer [10] and TranAD [11]. Despite their success, most methods still treat time series as generic sequences and fail to explicitly model multi-scale temporal structures and inter-variable dependencies.

B. Decomposition-Based Anomaly Detection Models

Time series decomposition aims to disentangle complex signals into interpretable components such as trends and seasonality. Early methods such as STL decomposition and Hodrick–Prescott filtering [17], [18] are typically static preprocessing steps with limited interaction with downstream models. Recent works integrate decomposition into neural architectures to capture multi-scale temporal variations, as exemplified by models such as Autoformer [19] and FEDformer [20]. While decomposition enhances sensitivity to trend shifts and seasonal disruptions, most existing anomaly detection approaches rely on fixed or single-scale schemes and utilize decomposed components in an implicit, entangled manner, limiting their adaptability to diverse anomaly patterns, especially under dynamic distribution shifts in multivariate time series.

C. Frequency-domain and Channel Dependency Modeling

Frequency-domain analysis has been widely used to uncover subsequence anomalies such as periodic fluctuations that are difficult to detect in the time domain. Methods such as SR-CNN, TFAD and FreqTAD [21], [22], [47] leverage spectral representations to enhance detection performance. However, these approaches often treat multivariate time series as independent univariate sequences or rely on simple concatenation, failing to capture inter-channel correlations.

Modeling channel dependencies is essential for characterizing system-wide anomalies. Existing approaches include Channel-Independent (CI) methods such as PatchTST [23], which ignore inter-variable interactions, and Channel-Dependent (CD) methods such as iTransformer and Crossformer [24], [25], which capture cross-channel correlations but incur high computational costs. Moreover, channel modeling is predominantly conducted in the time domain, while its potential in the frequency domain remains under-explored. Existing frequency-based methods lack dedicated mechanisms to decouple global channel correlations from local topological contexts, limiting their ability to model complex multivariate anomalies.

III. Methodology

A. Problem Formulation

We consider unsupervised multivariate time series anomaly detection. Given a multivariate time series

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}, \tag{1}$$

where $\mathbf{x}_t \in \mathbb{R}^M$ denotes the observation at timestamp t, we aim to identify anomalous timestamps or segments. To capture temporal dependencies, a sliding window is applied:

$$\mathbf{W}_t = \{\mathbf{x}_{t-L+1}, \ldots, \mathbf{x}_t\} \in \mathbb{R}^{L \times M}, \tag{2}$$

where L is the window size.

The anomaly score is computed based on the reconstruction error between $\mathbf{x}_t$ and its reconstruction $\hat{\mathbf{x}}_t$, followed by thresholding for anomaly detection.

B. MSDDS Overall Framework

The overall framework of MSDDS is illustrated in Fig. 1. MSDDS adopts a time-frequency collaborative design where temporal structures are enhanced before frequency-domain dependency modeling. The input time series is first processed in the time domain via multi-scale decomposition and adaptive fusion to separate trend and seasonal components across different scales, producing a structure-enhanced representation while preserving the channel dimensionality. Frequency-domain modeling is then applied to capture inter-channel dependencies through a dual-stream design that models global channel correlations and local topological contexts. Detailed designs are presented in the following subsections.

C. Time-domain Multi-Scale Decomposition with Fusion

MSDDS introduces a Time-domain Multi-Scale Decomposition with Fusion (MSDF) module to explicitly decompose mixed temporal patterns before frequency-domain modeling. Given an input window $\mathbf{X} \in \mathbb{R}^{L \times M}$, where L is the window length and M is the number of variables, MSDF performs decomposition under multiple smoothing horizons and adaptively fuses the resulting components with a residual pathway. For clarity, we omit the batch dimension in most notations and include it only when necessary.

Multi-scale moving-average decomposition Industrial multivariate time series often exhibit mixed temporal patterns, including slowly varying trends and locally varying seasonal fluctuations. A single smoothing horizon is insufficient to capture such behaviors [19]. Therefore, MSDF adopts a set of kernel sizes $\{k_1, k_2, \ldots, k_S\}$ to extract temporal patterns at different scales. For each scale k_i, replication padding is applied along the temporal dimension to preserve sequence length, and the trend and seasonal components are computed as

$$\mathbf{T}_i = \mathrm{AvgPool}(\mathbf{X}; k_i), \tag{3}$$

$$\mathbf{S}_i = \mathbf{X} - \mathbf{T}_i \tag{4}$$

where $\mathbf{T}_i, \mathbf{S}_i \in \mathbb{R}^{L \times M}$ denote the trend and seasonal components at scale k_i, and $\mathrm{AvgPool}(\cdot; k_i)$ is a one-dimensional average pooling operator with kernel size k_i and stride 1.

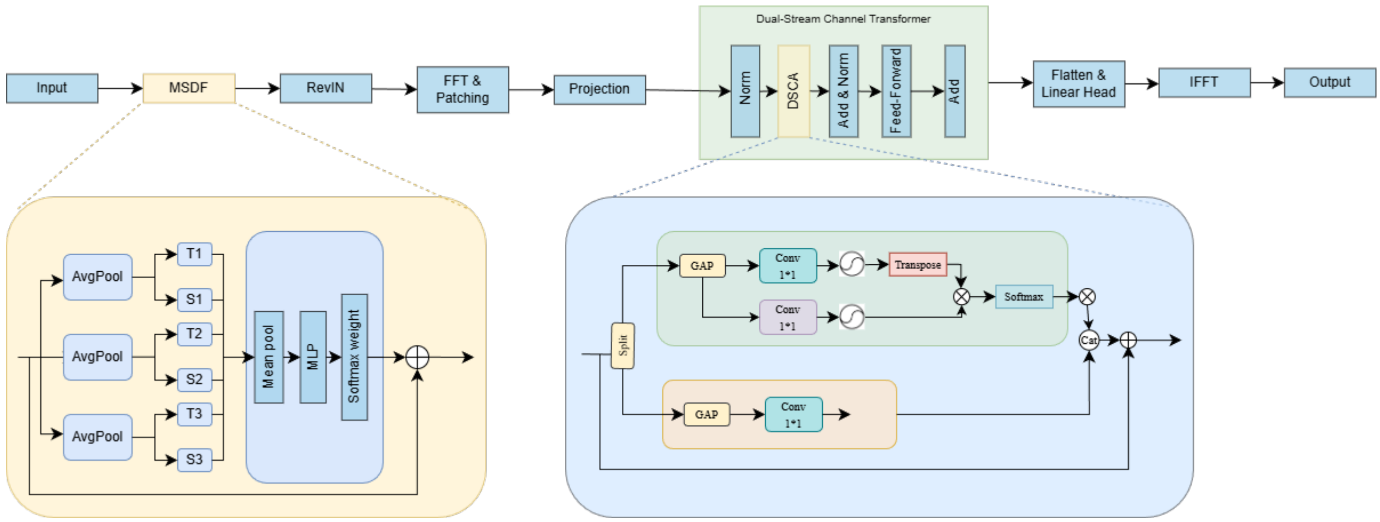


Fig. 1. Overall framework of MSDDS.

Larger kernels produce smoother $\mathbf{T}_i$ capturing long-term trends, while smaller kernels preserve local variations in $\mathbf{S}_i$. By decomposing the series across multiple scales, MSDF captures both slow trend drifts and fast seasonal perturbations, providing a richer temporal representation for subsequent frequency-domain modeling. All components are collected as

$$\mathcal{C} = \{\mathbf{T}_1, \ldots, \mathbf{T}_S, \mathbf{S}_1, \ldots, \mathbf{S}_S\}, \tag{5}$$

where $|\mathcal{C}| = 2S$ and each $\mathbf{C}_j \in \mathbb{R}^{L\times M}$.

Adaptive residual fusion Although $\mathcal{C}$ provides a comprehensive temporal basis, directly concatenating all components introduces redundancy and noise [27]. To address this, we adopt an adaptive residual fusion strategy that dynamically aggregates components based on their global relevance. The fusion is formulated as

$$\mathbf{u}_j = \frac{1}{L}\sum_{t=1}^{L} \mathbf{C}_j(t,:) \in \mathbb{R}^{B\times M}, \tag{6}$$

$$\boldsymbol{\alpha} = \mathrm{Softmax}([\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{2S}]) \in \mathbb{R}^{B\times 2S\times 1\times 1}, \tag{7}$$

$$\mathbf{X}^{\mathrm{MSDF}} = \sum_{j=1}^{2S} \alpha_j \odot \mathbf{C}_j + \mathbf{X}, \quad \mathbf{C}_j \in \mathbb{R}^{B\times L\times M}. \tag{8}$$

where $\mathbf{u}_j$ is a temporal summary, $\mathbf{a}_j = \phi(\mathbf{u}_j)$ is a scalar importance coefficient generated by a lightweight MLP, and α_j is a per-sample weight broadcast across temporal and channel dimensions. The residual connection with $\mathbf{X}$ stabilizes optimization by learning deviations from the raw signal [26].

D. Frequency-domain Dual-Stream Channel Transformer

The Dual-Stream Channel Transformer (DSCT) models inter-variable dependencies in the frequency domain. Unlike time-domain self-attention, DSCT decouples global spectral correlations and local channel-specific features, enabling robust modeling in high-dimensional sensor systems.

FFT and Patching Following MSDF, the input is transformed into the frequency domain via Fast Fourier Transform (FFT) [28], retaining both real and imaginary parts. Frequency patching is then applied to segment the spectra into local windows with patch size p and stride s:

$$\mathbf{F} = \mathcal{F}(\mathbf{X}^{\mathrm{MSDF}}) \in \mathbb{C}^{B\times C\times L}, \tag{9}$$

$$\mathbf{P} = \mathrm{Concat}(\mathrm{Unfold}(\mathbf{F}_r), \mathrm{Unfold}(\mathbf{F}_i)) \in \mathbb{R}^{B\times C\times N\times 2p}. \tag{10}$$

The patches are reshaped into a token sequence $\mathbf{H} \in \mathbb{R}^{B\times N\times C}$, serving as input to the dual-stream channel modeling.

Dual-Stream Channel Attention (DSCA) DSCA decouples local and global inter-channel dependencies in the frequency domain [29]. Given $\mathbf{H} \in \mathbb{R}^{B\times N\times C}$, the channel dimension is split into two subspaces $\mathbf{H}_1$ and $\mathbf{H}_2$.

The local branch captures short-range channel dependencies via a gating mechanism. Global average pooling is applied along the token dimension, followed by a 1D convolution and Sigmoid activation to produce channel-wise weights $\mathbf{w} \in \mathbb{R}^{B\times C}$:

$$\mathbf{A}_1 = \mathbf{H}_1 \odot \mathbf{w}. \tag{11}$$

The global branch models holistic channel dependencies. $\mathbf{H}_2$ is first pooled along the token dimension to obtain $\bar{\mathbf{H}}_2 \in \mathbb{R}^{B\times C}$, and projected to Query and Key vectors $\mathbf{q}, \mathbf{k} \in \mathbb{R}^{B\times C}$. The channel affinity matrix is computed as

$$\mathbf{M} = \mathrm{Softmax}(\mathbf{q}\mathbf{k}^{\top}) \in \mathbb{R}^{B\times C\times C}. \tag{12}$$

The affinity matrix is then used to aggregate channel responses into a global context vector $\mathbf{C} \in \mathbb{R}^{B\times C}$, which is broadcast to modulate $\mathbf{H}_2$:

$$\mathbf{A}_2 = \mathbf{H}_2 \odot \mathbf{C} + \mathbf{H}_2. \tag{13}$$

Finally, the two branches are fused with a balancing coefficient α and combined with a residual connection:

$$\mathbf{H}_{\mathrm{out}} = \mathrm{Concat}(\alpha\mathbf{A}_1, (1-\alpha)\mathbf{A}_2) + \mathbf{H}. \tag{14}$$

E. Optimization and Anomaly Scoring

Time-Frequency Reconstruction Process Following DSCT, the refined representations are reconstructed to the time domain. The patch-wise frequency representations are flattened and projected to recover the real and imaginary spectra, followed by inverse FFT (iFFT):

$$\mathbf{X}'_{\text{real}} = \text{Projection}_{\text{R}}(\text{Flatten}(\mathbf{H}_{\text{real}})), \tag{15}$$

$$\mathbf{X}'_{\text{imag}} = \text{Projection}_{\text{I}}(\text{Flatten}(\mathbf{H}_{\text{imag}})), \tag{16}$$

$$\mathbf{X}' = \text{iFFT}(\mathbf{X}'_{\text{real}}, \mathbf{X}'_{\text{imag}}), \tag{17}$$

where $\mathbf{X}'_{\text{real}}, \mathbf{X}'_{\text{imag}}, \mathbf{X}' \in \mathbb{R}^{B\times L\times C}$. Reconstruction losses are defined in both time and frequency domains:

$$\mathcal{L}_{\text{time}} = ||\mathbf{X} - \mathbf{X}'||_F^2, \tag{18}$$

$$\mathcal{L}_{\text{freq}} = ||\mathcal{F}_r(\mathbf{X}) - \mathbf{X}'_{\text{real}}||_1 + ||\mathcal{F}_i(\mathbf{X}) - \mathbf{X}'_{\text{imag}}||_1. \tag{19}$$

The 2-norm is used in the time domain and the 1-norm in the frequency domain due to their distinct characteristics [30].

Optimization Strategy The overall loss is defined as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{time}} + \lambda_{\text{train}} \cdot \mathcal{L}_{\text{freq}}, \tag{20}$$

where $\mathcal{L}_{\text{time}}$ enforces point-wise fidelity and $\mathcal{L}_{\text{freq}}$ constrains spectral consistency.

Anomaly Scoring Anomaly scores are computed from both temporal and spectral residuals. The time-domain component captures point-wise deviations, while the frequency-domain component reflects spectral discrepancies over local contexts [21], [22], [31]. The final score is

$$\mathcal{S} = \mathcal{S}_{\text{time}} + \lambda_{\text{score}} \cdot \mathcal{S}_{\text{freq}}, \tag{21}$$

where λ_{score} controls the relative sensitivity to point-wise outliers and subsequence anomalies. Incorporating frequency-domain residuals into point-wise scoring improves anomaly localization and enhances sensitivity to subsequence anomalies with subtle amplitude changes but structural distortions [4], [32]. A dynamic threshold is then determined based on validation-set percentiles to produce binary predictions.

IV. Experiments

We evaluate the proposed MSDDS on multiple public benchmarks for multivariate time-series anomaly detection and compare it with a wide range of state-of-the-art baselines, including both classical and deep learning-based methods.

Datasets We evaluate MSDDS on seven public datasets: SMD [8], MSL [13], PSM [33], CICIDS [34], CalIt2 [35], Creditcard [36], and GECCO [37]. These datasets cover diverse domains, including system monitoring, space telemetry, network security, finance, and industrial sensing, and involve various anomaly types such as point anomalies, subsequence anomalies, and long-duration deviations. Transactional datasets (CICIDS and Creditcard) are formulated as multivariate time series to evaluate model robustness under the high noise and severe imbalance characteristic of non-sensor domains. Dataset statistics are summarized in Table I.

TABLE I
Dataset Statistics.

Dataset	Total Size	Test Size	Dimensions	Anomaly Ratio (%)
SMD	1,416,825	708,420	38	2.08
MSL	132,046	73,729	55	5.88
PSM	220,322	87,841	25	11.07
CICIDS	170,231	85,116	72	1.28
CalIt2	5,040	2,520	2	4.09
Creditcard	284,807	142,404	29	0.17
GECCO	138,521	69,261	9	1.25

Baselines We compare MSDDS with representative anomaly detection methods, including classical models Isolation Forest [38] and One-Class SVM [6], and recent deep learning approaches such as AutoEncoder [7], Anomaly Transformer [10], DCDetector [39], TimesNet [40], DLinear [41], PatchTST [23], DualTF [42], iTransformer [24], ModernTCN [43], and CATCH [44]. For fair comparison, we re-implemented all baselines under a unified protocol and verified that the results are consistent with the CATCH benchmark [44]. Thus, we directly report benchmark results for these baselines. CATCH is re-evaluated under the same setup as MSDDS to ensure strict comparability.

Implementation Details Models are implemented in PyTorch and optimized using Adam with an initial learning rate of 10^{-4} and a OneCycleLR scheduler. Following [44], the sliding window length is fixed to 192 for all datasets. Early stopping is applied based on the validation set (20% of training data). For fair comparison, we evaluate our proposed MSDDS alongside a re-run of CATCH under a unified protocol, other baseline results are taken from the CATCH benchmark [44].

Evaluation Metrics To ensure consistency with prior work, we adopt Aff-F (Affiliation-F1) [45] and AUC-ROC [46] as primary metrics. Aff-F measures the overlap between predicted and ground-truth segments, making it suitable for evaluating subsequence anomalies, while AUC-ROC assesses the ranking quality of anomaly scores independent of thresholds. Accuracy (Acc) is also reported for completeness.

A. Main Results

Table II summarizes the performance of MSDDS against state-of-the-art baselines on seven datasets. MSDDS consistently achieves the best Aff-F, demonstrating strong capability in detecting contiguous anomaly segments, particularly on challenging datasets such as GECCO and CalIt2. It also outperforms recent Transformer-based models, including iTransformer, PatchTST, and ModernTCN, indicating the importance of jointly modeling temporal and spectral structures. Moreover, MSDDS attains the highest AUC-ROC across all datasets, reflecting the effectiveness of the proposed dual-domain scoring mechanism.

Although some baselines, such as IF and AE, occasionally achieve higher Accuracy, their significantly lower Aff-F suggests a bias toward majority classes. In contrast, MSDDS maintains a better balance between precision and recall, enabling more reliable anomaly detection.

TABLE II
PERFORMANCE COMPARISON OF MSDDS WITH BASELINE METHODS ON SEVEN REAL-WORLD DATASETS. AFF-F: AFFILIATION-F1, ACC: ACCURACY, AUC-ROC: AREA UNDER THE ROC CURVE. BEST AND SECOND-BEST RESULTS ARE HIGHLIGHTED IN **BOLD** AND UNDERLINED, RESPECTIVELY. RESULTS OF MSDDS ARE OBTAINED FROM OUR EXPERIMENTS. BASELINE RESULTS FOLLOW THE CATCH BENCHMARK UNDER THE SAME EVALUATION PROTOCOL.

Dataset	Metric	IF	OCSVM	AE	DLinear	PatchTST	TimesNet	DC	Atrans	DualTF	iTrans	Modern	CATCH	Ours
SMD	Aff-F	0.626	0.742	0.439	0.841	0.845	0.831	0.675	0.724	0.679	0.827	0.840	0.847	**0.853**
	Acc	0.958	0.865	**0.959**	0.943	0.934	0.931	0.774	0.952	0.863	0.914	0.931	0.946	0.952
	AUC-ROC	0.664	0.602	0.774	0.728	0.736	0.727	0.502	0.508	0.631	0.745	0.722	0.811	**0.832**
MSL	Aff-F	0.584	0.641	0.625	0.725	0.724	0.734	0.694	0.692	0.588	0.710	0.726	0.710	**0.735**
	Acc	0.890	0.818	0.891	0.858	0.873	0.855	0.892	0.891	0.891	0.854	0.857	0.928	**0.937**
	AUC-ROC	0.524	0.524	0.562	0.624	0.637	0.613	0.507	0.508	0.576	0.611	0.633	0.664	**0.665**
PSM	Aff-F	0.620	0.531	0.707	0.831	0.831	0.842	0.682	0.710	0.725	0.846	0.825	0.831	**0.851**
	Acc	0.723	0.737	0.715	0.726	0.729	0.726	0.615	0.719	0.697	0.728	0.729	0.984	**0.986**
	AUC-ROC	0.542	0.619	0.650	0.580	0.586	0.592	0.501	0.514	0.600	0.593	0.593	0.645	**0.651**
CICIDS	Aff-F	0.604	0.693	0.243	0.669	0.660	0.657	0.664	0.560	0.692	0.708	0.654	0.772	**0.791**
	Acc	0.852	0.082	**0.994**	0.757	0.751	0.750	0.750	0.745	0.565	0.762	0.760	0.760	0.763
	AUC-ROC	0.787	0.537	0.629	0.751	0.716	0.732	0.638	0.528	0.603	0.692	0.697	0.791	**0.803**
CalIt2	Aff-F	0.402	0.783	0.587	0.793	0.808	0.771	0.527	0.533	0.574	0.791	0.676	0.835	**0.843**
	Acc	**0.968**	0.867	0.960	0.887	0.889	0.946	0.841	0.944	0.910	0.930	0.889	0.962	0.964
	AUC-ROC	0.775	0.804	0.767	0.752	0.808	0.771	0.527	0.533	0.574	0.791	0.676	0.838	**0.846**
Credit	Aff-F	0.634	0.714	0.561	0.738	0.746	0.744	0.632	0.650	0.663	0.713	0.744	0.750	**0.758**
	Acc	**0.996**	0.893	0.994	0.981	0.980	0.981	0.716	0.847	0.688	0.951	0.981	0.981	0.990
	AUC-ROC	0.860	0.953	0.909	0.954	0.957	0.957	0.504	0.552	0.703	0.934	0.957	0.958	**0.965**
GECCO	Aff-F	0.424	0.666	0.823	0.893	0.906	0.894	0.687	0.782	0.701	0.839	0.893	0.908	**0.920**
	Acc	0.989	0.055	0.888	0.981	**0.989**	0.984	0.979	0.983	0.612	0.981	0.984	0.986	0.986
	AUC-ROC	0.619	0.804	0.769	0.947	0.949	0.954	0.555	0.516	0.714	0.795	0.952	0.970	**0.976**

TABLE III
ABLATION STUDY ON FOUR DATASETS (AFF-F & AUC). THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Method	SMD		MSL		PSM		CalIt2	
	Aff-F	AUC	Aff-F	AUC	Aff-F	AUC	Aff-F	AUC
w/o MSDF & DSCA	0.831	0.778	0.716	0.656	0.824	0.641	0.794	0.819
w/o MSDF	0.840	0.789	0.721	0.658	0.840	0.646	0.811	0.822
w/o DSCA	0.846	0.790	0.724	0.657	0.829	0.640	0.810	0.821
w/o frequency loss	0.824	0.765	0.714	0.642	0.813	0.635	0.803	0.812
MSDDS (Ours)	**0.853**	**0.832**	**0.735**	**0.665**	**0.851**	**0.651**	**0.843**	**0.846**

B. Model Analysis

Ablation Study We conduct ablation studies on four representative datasets (SMD, MSL, PSM, and CalIt2) to evaluate the effectiveness of key components across diverse data characteristics (Table I). These benchmarks represent large-scale, high-dimensional, low-dimensional, and high anomaly-ratio scenarios, respectively. We compare the full model with four variants: (1) **w/o MSDF**, removing the multi-scale decomposition module; (2) **w/o DSCA**, replacing the dual-stream channel attention; (3) **w/o MSDF & DSCA**, removing both components; (4) **w/o freq loss**, excluding the frequency-domain reconstruction objective.

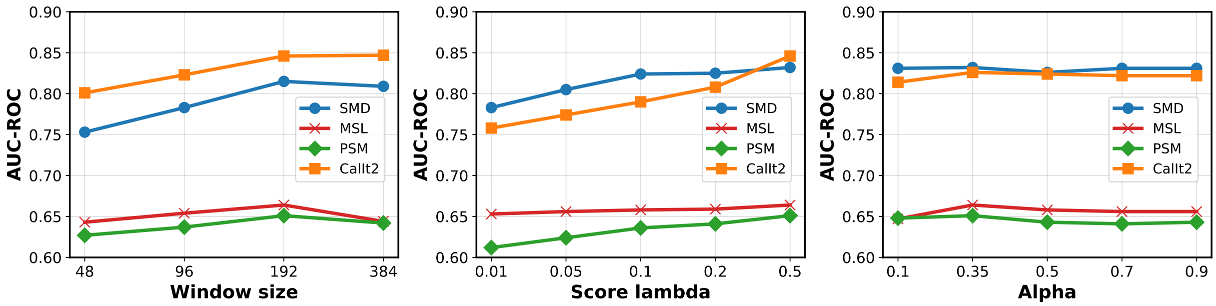
Results in Table III show that the full MSDDS consistently outperforms all variants. Removing MSDF results in a significant performance drop, highlighting the importance of multi-scale temporal modeling. Similarly, the performance decline in w/o DSCA and w/o freq loss variants confirms the necessity of inter-channel dependency modeling and spectral supervision.

Parameter Sensitivity Figure 2 analyzes the sensitivity of MSDDS to key hyperparameters. For the input window size L, performance improves from 48 to 192 and then plateaus or slightly degrades at 384, suggesting that overly long contexts may introduce noise; thus, $L = 192$ is adopted. For the anomaly score weight λ_{score}, performance increases on complex datasets such as SMD, indicating the importance of frequency-domain signals, with optimal values typically in [0.1, 0.5]. Finally, performance remains stable across fusion coefficients $\alpha \in [0.1, 0.9]$, demonstrating robustness and the complementary roles of local and global channel dependencies in DSCA.

Efficiency Analysis We further evaluate the computational efficiency of MSDDS in terms of model size, inference latency, and peak GPU memory usage. All measurements are conducted on an NVIDIA RTX 4090 GPU with batch size $B{=}1$ and input window $(L{=}192, C{=}25)$. Despite the relatively large parameter count, the low inference latency is mainly attributed to the shallow architecture, short input length, and the use of lightweight channel-wise operations and FFT-based transforms instead of quadratic token-wise self-attention. Table IV reports the efficiency comparison between MSDDS and its backbone CATCH. As shown, MSDDS achieves lower inference latency while maintaining almost identical parameter count and peak memory usage. This indicates that the performance improvements of MSDDS are not obtained at the cost of increased computational overhead.

V. CONCLUSION

We propose MSDDS, a novel framework for multivariate time series anomaly detection that jointly models multi-scale temporal structures and frequency-domain channel depen-



(a) Window size (b) λ_{score} (c) α

Fig. 2. Parameter sensitivity analysis on window size L, score weight λ_{score}, and fusion coefficient α. By spanning the entire page width, the trends and labels are clearly visible.

TABLE IV
EFFICIENCY COMPARISON BETWEEN MSDDS AND ITS BACKBONE CATCH ON RTX 4090 WITH INPUT (B=1, L=192, C=25).

Method	Params (M) ↓	Inference (ms) ↓	Peak Mem (MB) ↓
CATCH	210.88	3.87	816.5
MSDDS (Ours)	**210.78**	**3.19**	**816.3**

dencies. By integrating Multi-Scale Decomposition with Fusion (MSDF) and Dual-Stream Channel Transformer (DSCT), MSDDS effectively captures non-stationary dynamics and complex inter-variable correlations. Extensive experiments on seven real-world benchmarks demonstrate that MSDDS consistently outperforms state-of-the-art methods, particularly in detecting subtle subsequence anomalies. These results highlight the critical importance of synergistic time-frequency modeling for robust anomaly detection. In future work, we aim to explore the potential of MSDDS within a unified pre-training paradigm to further enhance its generalization capability across unseen domains.

ACKNOWLEDGMENT

This work was supported by the Academy Member (Academician) Major Science and Technology Innovation Project under the "Two Zones" Science and Technology Development Program Grant 2024LQ02001; and the National Natural Science Foundation of China Project under Grant 62262064, 62466057; and the Key R&D projects of Xinjiang Uygur Autonomous Region, grant number 2022295358; Xinjiang University doctor postgraduate innovation project (Grant No.XJDX2025YJS084, XJDX2025YJS086).

REFERENCES

[1] T. Pan, J. Chen, J. Xie, Z. Zhou, and S. He, "Deep feature generating network: A new method for intelligent fault detection of mechanical systems under class imbalance," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6282–6293, Sept. 2021.

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

[3] J. Paparrizos, Y. Kang, P. Boniol, R. S. Tsay, T. Palpanas, and M. J. Franklin, "TSB-UAD: An end-to-end benchmark suite for univariate time-series anomaly detection," *Proc. VLDB Endow.*, vol. 15, no. 8, pp. 1697–1711, 2022.

[4] Y. Nam *et al.*, "Breaking the time-frequency granularity discrepancy in time-series anomaly detection," in *Proc. ACM Web Conf. (WWW)*, 2024, pp. 4204–4215.

[5] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2015.

[6] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 12, pp. 582–588, 1999.

[7] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proc. MLSDA Workshop Mach. Learn. Sensory Data Anal.*, 2014, pp. 4–11.

[8] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural networks," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2019, pp. 2828–2837.

[9] V. Kumar, V. Srivastava, S. Mahjabin, A. Pal, S. Klüttermann, and E. Müller, "Autoencoder optimization for anomaly detection: A comparative study with shallow algorithms," in *Proc. Int. Jt. Conf. Neural Netw. (IJCNN)*, 2024, pp. 1–8.

[10] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly Transformer: Time series anomaly detection with association discrepancy," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2022.

[11] S. Tuli, G. Casale, and N. R. Jennings, "TranAD: Deep transformer networks for anomaly detection in multivariate time series data," *Proc. VLDB Endow.*, vol. 15, no. 6, pp. 1201–1214, 2022.

[12] S. Wang, J. Cao, and P. S. Yu, "Deep learning for spatio-temporal data mining: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3681–3700, 2022.

[13] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding," in *Proc. ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2018, pp. 387–395.

[14] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "DeepAnT: A deep learning approach for unsupervised anomaly detection in time series," *IEEE Access*, vol. 7, pp. 1991–2005, 2019.

[15] H. Xu *et al.*, "Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications," in *Proc. Int. Conf. World Wide Web (WWW)*, 2018, pp. 187–196.

[16] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "USAD: Unsupervised anomaly detection on multivariate time series," in *Proc. ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2020, pp. 3395–3404.

[17] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A seasonal-trend decomposition procedure based on loess," *J. Off. Stat.*, vol. 6, no. 1, pp. 3–73, 1990.

[18] R. J. Hodrick and E. C. Prescott, "Postwar US business cycles: An empirical investigation," *J. Money, Credit Banking*, vol. 29, no. 1, pp. 1–16, 1997.

[19] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 34, pp. 22419–22430, 2021.

[20] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2022, pp. 27268–27286.

[21] H. Ren *et al.*, "Time-series anomaly detection service at Microsoft," in *Proc. ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2019, pp. 3009–3017.

[22] C. Zhang, T. Zhou, Q. Wen, and L. Sun, "TFAD: A decomposition time series anomaly detection architecture with time-frequency analysis," in *Proc. ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2022, pp. 2497–2506.

[23] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Jayaram, "A time series is worth 64 words: Long-term forecasting with transformers," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2023.

[24] Y. Liu *et al.*, "iTransformer: Inverted transformers are effective for time series forecasting," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2024.

[25] Y. Zhang and J. Yan, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2023.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[27] 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," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2024.

[28] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, no. 90, pp. 297–301, 1965.

[29] C. H. Song, H. J. Han, and Y. Avrithis, "All the attention you need: Global-local, spatial-channel attention for image retrieval," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, 2022, pp. 2754–2763.

[30] H. Wang *et al.*, "FreDF: Learning to forecast in the frequency domain," *arXiv preprint arXiv:2402.02399*, 2024.

[31] Y. Park, J.-G. Jang, and U. Kang, "Fast and accurate partial Fourier transform for time series data," in *Proc. ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2021, pp. 1309–1318.

[32] S. Schmidl, P. Wenig, and T. Papenbrock, "Anomaly detection in time series: A comprehensive evaluation," *Proc. VLDB Endow.*, vol. 15, no. 9, pp. 1779–1797, 2022.

[33] A. Abdulaal, Z. Liu, and T. Lancewicki, "Practical approach to asynchronous multivariate time series anomaly detection and localization," in *Proc. ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2021, pp. 2485–2494.

[34] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. Int. Conf. Inf. Syst. Secur. Priv. (ICISSP)*, 2018, pp. 108–116.

[35] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[36] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, 2015, pp. 159–166.

[37] S. Moritz *et al.*, "GECCO industrial challenge 2018 dataset: A water quality dataset for the 'Internet of Things: Online anomaly detection for drinking water quality' competition," in *Proc. Genet. Evol. Comput. Conf. Companion (GECCO)*, 2018, pp. 1823–1830.

[38] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. IEEE Int. Conf. Data Min. (ICDM)*, 2008, pp. 413–422.

[39] Y. Yang, C. Zhang, T. Zhou, Q. Wen, and L. Sun, "DCdetector: Dual attention contrastive learning for time series anomaly detection," in *Proc. ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2023, pp. 3033–3045.

[40] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "TimesNet: Temporal 2D-variation modeling for general time series analysis," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2023.

[41] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are Transformers effective for time series forecasting?" in *Proc. AAAI Conf. Artif. Intell.*, 2023.

[42] Y. Chen, Z. Liu, and H. Wang, "Dual-stream time-frequency anomaly detection," *J. Time Ser. Anal.*, vol. 44, no. 2, pp. 150–165, 2023.

[43] L. Dong *et al.*, "ModernTCN: A modern pure convolution structure for general time series analysis," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2024.

[44] X. Wu *et al.*, "CATCH: Channel-aware multivariate time series anomaly detection via frequency patching," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2025.

[45] A. Huet, B. Lafabregue, and G. Forestier, "Local affiliation: A robust metric for time series anomaly detection evaluation," in *Proc. ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2022, pp. 605–614.

[46] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.

[47] C. Li, R. Liu, Z. Zhao, H. Zhou, and Q. Zeng, "FreqTAD: Multi-scale frequency encoding and time-frequency attention for anomaly detection in dynamic graphs," in *Proc. AAAI Conf. Artif. Intell.*, 2026.