Differential Contrastive Representation Learning with Frequency-Domain Attention for Anomaly Detection in Aerospace Cyber-Physical Systems

Jiandun Li, Sha Zhou, Shuo Zhang*
School of Electronic Information Engineering, Shanghai Dianji University, China

***Abstract*—Anomaly detection in aerospace cyber-physical systems (CPS) is crucial for ensuring spacecraft safety and mission reliability. However, multivariate telemetry data exhibit strong cross-variable coupling, pronounced non-stationarity, and severe environmental noise, which significantly challenge existing unsupervised detection methods. To address these issues, we propose DCRFA, a differential contrastive representation learning anomaly detection framework with frequency-domain attention. Our differential enhancement module explicitly separates high-frequency dynamics from low-frequency drifts, mitigating the influence of non-stationary trends. Meanwhile, the frequency-domain attention mechanism operates in the spectral domain, providing a global receptive field to capture periodic patterns and transient energy variations while suppressing local noise and improving computational efficiency. Furthermore, we introduce a KL-divergence-based consistency contrastive objective with a stop-gradient strategy to stabilize the learning of nominal physical representations in latent space. Extensive experiments on multiple aerospace telemetry datasets demonstrate that DCRFA consistently outperforms nine representative baseline methods, validating its effectiveness and robustness in complex aerospace anomaly detection scenarios.**

***Index Terms*—Aerospace CPS, anomaly detection, frequency-domain attention, contrastive learning, differential representation**

I. Introduction

Identifying anomalies in aerospace cyber-physical systems (CPS) is a critical research area that has attracted widespread attention due to its profound impact on spacecraft safety and mission reliability [1]–[4]. Driven by the rapid development of the aerospace industry and the increasing complexity of CPS, anomaly detection tasks have evolved from traditional statistical techniques to deep learning algorithms, with the demand for robust solutions continuing to grow at a rapid pace [3], [5]–[7].

The integration of computing and physical processes in aerospace CPS introduces unique challenges, as multivariate telemetry data inherently exhibit severe environmental noise, pronounced non-stationarity, and strong cross-variable coupling. To address these challenges, various deep learning approaches have been proposed. For instance, recurrent models such as LSTM/BiGRU have been utilized to capture long-term temporal dependencies [8], while RmsAnomaly proposes a reconstruction-based convolutional autoencoder for multisensor modeling [9]. Furthermore, digital twin-assisted GANs [10] and advanced Transformer architectures like PAFormer [11] and VVT [12] have been developed to enhance the modeling of complex inter-variable relationships. Despite these encouraging advancements, the high reliance of deep learning models on uncontaminated or labeled training data remains a significant obstacle, particularly in aerospace CPS environments where authentic fault labels are exceedingly scarce and normal data are often corrupted.

Although existing unsupervised and reconstruction-based paradigms provide compelling solutions when labeled data is scarce, they face critical domain-specific bottlenecks. Unlike image or speech domains [13]–[17], aerospace telemetry is highly dynamic and extremely sensitive to environmental perturbations. Standard Transformer architectures tend to disproportionately overfit high-frequency noise spikes, leading to misallocated attention and degraded modeling of true physical modes [18]. Furthermore, prevalent reconstruction methods (e.g., ACUDL [19] and CPS-GUARD [20]) heavily rely on the assumption of pristine training data; under strong distributional shifts or non-ideal training conditions, they often absorb intermittent anomalies into normal patterns, resulting in ambiguous decision boundaries and higher false negative rates. Unlike these approaches, our proposed DCRFA effectively mitigates these modeling limitations by explicitly disentangling high-frequency dynamics from non-stationary drifts.

In our study, we propose Differential Contrastive Representation Learning with Frequency-Domain Attention (DCRFA), an unsupervised framework designed for aerospace CPS multivariate time-series anomaly detection. DCRFA leverages contrastive learning and focuses on robust spatiotemporal and spectral representations. First, it employs a differential enhancement strategy to explicitly decouple non-stationary background drifts from raw telemetry signals. Second, it utilizes a frequency-domain attention mechanism that captures global periodic operational modes while efficiently suppressing high-frequency noise. Finally, to address training data contamination, DCRFA introduces a negative-free contrastive learning paradigm based on symmetrized Kullback-Leibler (KL) divergence with a stop-gradient strategy to stabilize the learning of nominal physical manifolds. Extensive experiments on three real-world aerospace benchmarks (MSL, SMAP, and OPSSAT) demonstrate that DCRFA consistently outperforms nine representative baseline methods in both detection accuracy and inference efficiency.

*Corresponding Author.

II. METHOD

A. Problem Formulation

Given a multivariate time series $S \in \mathbb{R}^{T\times d}$, unsupervised anomaly detection identifies deviations from nominal system dynamics. Assuming a predominantly healthy training set S_{train}, our objective is to learn robust representations of regular temporal evolution, ultimately constructing a binary function $f : S_{\text{test}} \to \{0, 1\}$ to detect anomalies in unseen sequences.

B. Framework

Figure 1 shows that the DCRFA comprises three core stages: (1) dual-view embedding of the original and differential sequences; (2) a dual-path spectral-spatial encoder that jointly fuses frequency-domain temporal dynamics with spatial inter-sensor dependencies; and (3) an asymmetric contrastive learning module equipped with a stop-gradient mechanism to align heterogeneous representations and prevent feature collapse.

C. Time Series Differencing

Standard Transformer architectures are prone to attention bias when processing non-stationary telemetry data due to large-magnitude trend components. To decouple system dynamics from background drift, we introduce a first-order differencing mechanism, acting as a high-pass filter to explicitly extract incremental features reflecting spacecraft state evolution. Formally, let the differencing operator Δ act on a multivariate time series. For any time step t, the dynamic change vector is defined in Eq. (1):

$$\Delta x_t = x_{t+1} - x_t, \quad \forall t \in \{1, \ldots, T-1\}. \tag{1}$$

Consequently, the original observation matrix $X \in \mathbb{R}^{T\times d}$ is mapped to the differential view $X_{\text{diff}} = \{\Delta x_1, \ldots, \Delta x_{T-1}\} \in \mathbb{R}^{(T-1)\times d}$. This view removes low-frequency trends induced by orbital cycles and environmental factors, thereby effectively compelling the model to focus on high-frequency dynamics intrinsic to the system itself.

D. Frequency-domain Temporal Dependency Learning

To address the point-wise noise sensitivity and computational overhead of standard Transformers when processing high-noise aerospace telemetry, we introduce a *Frequency-Domain Global Attention* (FDGA) mechanism. This attention is applied along the temporal path to capture long-range dependencies. According to the convolution theorem, global circular convolution in the time domain is equivalent to a point-wise multiplication in the frequency domain. Therefore, this module projects the input features into an orthogonal spectral space via the Discrete Fourier Transform (DFT) and exploits the frequency sparsity inherent in spacecraft telemetry to capture subtle latent orbital periodicities and long-range temporal dependencies more effectively.

Formally, let the input feature tensor of the l-th encoder layer be $\mathbf{Z}^{(l)} \in \mathbb{R}^{T\times d_{model}}$, where T denotes the temporal window length. First, we apply the Fast Fourier Transform (FFT) along the temporal dimension to map the input features into the frequency domain, thereby explicitly yielding the complex-valued spectrum representation:

$$\mathcal{Z}^{(l)}_{freq} = \mathcal{F}(\mathbf{Z}^{(l)}) \in \mathbb{C}^{F\times d_{model}}, \tag{2}$$

where $\mathcal{F}(\cdot)$ denotes the FFT operator, and $F = \lfloor T/2 \rfloor + 1$ represents the number of effective frequency components due to the conjugate symmetry property.

To adaptively enhance dominant frequency components reflecting normal system dynamics (e.g., orbital and rotational periods) while suppressing high-frequency noise, we introduce a learnable complex weight matrix $\mathbf{W}^{(l)}_{spec} \in \mathbb{C}^{F\times d_{model}}$ as a global spectral filter. The filter modulates the magnitude and phase of frequency features via Hadamard product. The modulated spectrum is then transformed back to the time domain using the inverse FFT, and combined with the input through LayerNorm [21] and Dropout:

$$\tilde{\mathbf{Z}}^{(l)}_{time} = \text{LN}\left(\mathbf{Z}^{(l)} + \text{Dropout}\big(\mathcal{F}^{-1}(\mathcal{Z}^{(l)}_{freq} \odot \mathbf{W}^{(l)}_{spec})\big)\right), \tag{3}$$

where $\mathcal{F}^{-1}(\cdot)$ denotes the inverse FFT and $\odot$ represents element-wise complex multiplication. This operation achieves global information aggregation with $O(T \log T)$ complexity.

To align the feature space before fusion, the temporal branch output is transformed via FFT:

$$Z^{(l)}_{time} = \mathcal{F}(\tilde{Z}^{(l)}_{time}) \tag{4}$$

E. Spatiotemporal Dependency Learning

To overcome standard self-attention's vulnerability to high-frequency noise and its inadequacy in modeling complex spatiotemporal couplings, we propose a dual-view Spectral-Spatial encoder. This module parallelly processes original and differential sequences, leveraging a temporal branch for frequency-domain evolution and a spatial branch for intra-step, cross-channel dependencies.

a) Temporal-wise Transformer: The temporal-wise branch is designed to uncover latent temporal evolution patterns in telemetry sequences while efficiently capturing long-range dependencies. Formally, the input features to the l-th encoder layer are denoted as $\mathbf{Z}^{(l)} \in \mathbb{R}^{T\times d_{model}}$, where T is the temporal window length and d_{model} is the hidden feature dimension.

In this branch, the previously introduced FDGA is applied to model temporal dependencies. To unify residual connections and normalization, we define the standard Add & Norm operator as:

$$\mathcal{R}(\mathbf{x}, f) = \text{LayerNorm}\big(\mathbf{x} + f(\mathbf{x})\big). \tag{5}$$

The state update of a single layer in the temporal-wise Transformer is then expressed as:

$$\mathbf{Z}^{(l)}_{time} = \mathcal{R}\Big(\mathcal{R}\big(\mathbf{Z}^{(l)}, \Phi_{\text{spec}}\big), \Phi_{\text{ffn}}\Big), \tag{6}$$

where $\Phi_{\text{spec}}(\cdot)$ denotes the FDGA mapping, and $\Phi_{\text{ffn}}(\cdot)$ denotes the position-wise feedforward network. This design mitigates attention bias caused by local anomalies along the temporal dimension and enables global dependency modeling with $O(T \log T)$ complexity.

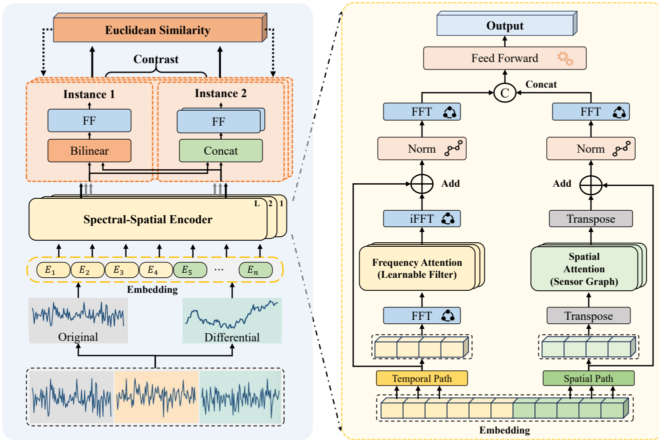

Fig. 1. Overview of the proposed framework, which jointly models temporal-frequency dynamics and spatial dependencies via a dual-path spectral–spatial encoder with contrastive representation learning.

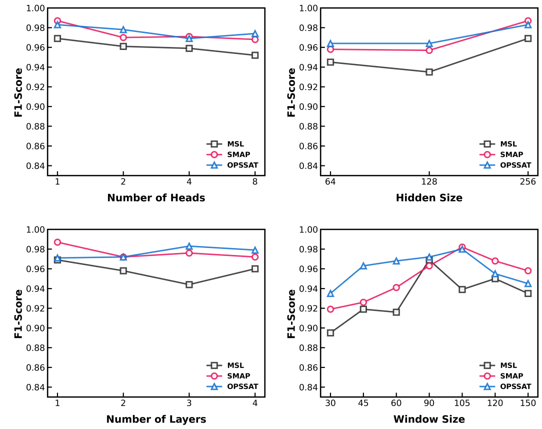

Fig. 2. Sensitivity analysis of key hyperparameters, including the number of heads, hidden dimension, network depth, and window size.

b) Spatial-wise Transformer: The spatial-wise branch is designed to capture implicit topological dependencies and nonlinear interactions among sensor variables at a fixed time step. Notably, the sensitivity of the self-attention mechanism to anomalous perturbations is primarily manifested along the temporal dimension, whereas the inter-variable correlations at a given time slice are generally more stable. Therefore, a standard Transformer encoder is employed in the spatial branch to model dependencies among variables.

Specifically, given the input features at layer l, $\mathbf{Z}^{(l)}$, the variable dimension is treated as a sequence by transposition, and a standard Transformer encoder is applied to model inter-variable dependencies:

$$\mathbf{Z}_{spa}^{(l)} = \left(\mathcal{T}_{std}\left(\mathbf{Z}^{(l)\top}\right)\right)^{\top}, \tag{7}$$

where $(\cdot)^{\top}$ denotes the matrix transpose, and $\mathcal{T}_{std}(\cdot)$ represents a standard Transformer encoder consisting of multi-head self-attention and a feedforward network.

c) Spatio-Temporal Feature Fusion: Features from both branches are mapped to the frequency domain via FFT, concatenated, and then passed through a shared Feed-Forward Network (FFN):

$$Z^{(l+1)} = \mathrm{FFN}\left(W_{proj}\left[\mathcal{F}(\tilde{Z}_{time}^{(l)}) \parallel \mathcal{F}(\tilde{Z}_{spa}^{(l)})\right] + b_{proj}\right) \tag{8}$$

where $[\cdot \parallel \cdot]$ denotes concatenation along the feature dimension, and $\mathbf{W}_{proj}$ and $\mathbf{b}_{proj}$ are learnable parameters. This fusion mechanism ensures that the model captures both long-range temporal evolution patterns and multivariate spatial dependencies, providing a robust representation foundation for subsequent anomaly detection.

F. Representation Discrepancy

Due to the heterogeneity between temporal representations and differential dynamics, direct point-wise comparison at the encoder outputs is ill-posed and difficult to optimize. We therefore adopt an asymmetric dual-projection scheme that maps heterogeneous features into a shared metric space for contrastive alignment. Let $H_t \in \mathbb{R}^{L\times d}$ and $H_d \in \mathbb{R}^{(L-1)\times d}$ denote the temporal and differential encoder outputs, respectively. To construct the first contrastive instance, a parameterized bilinear transformation is applied to explicitly model second-order interactions between state and trend. The projection process is formalized as follows:

$$Z_{i1} = S\Big(W_{p1}\cdot\sigma(H_d W_{bi} H_t + b_{bi}) + b_{p1}\Big) \tag{9}$$

where W_{bi} denotes the bilinear tensor weights, and W_{p1}, b_{p1} correspond to the fully connected projection layer parameters. $\sigma(\cdot)$ is the LeakyReLU activation, preserving gradient flow in the negative region, and $S(\cdot)$ is the Sigmoid, constraining the output features Z_{i1} within the $(0,1)$ probability range.

The second contrastive instance adopts a feature concatenation strategy to fuse dual-stream information, followed by a deep linear mapping to extract higher-order semantic features:

$$Z_{i2} = S\Big(W_{p2}\cdot\sigma(W_{cat}[H_d||H_t] + b_{cat}) + b_{p2}\Big) \tag{10}$$

where $[\cdot||\cdot]$ denotes concatenation along the feature dimension, and W_{cat} represents the fusion layer weight matrix. Through these asymmetric transformations, the heterogeneous representations H_t and H_d are mapped into isomorphic views Z_{i1} and Z_{i2} in a shared latent space.

To delineate the decision boundary of normal patterns in the absence of negative samples, we adopt a positive-only alignment contrastive learning paradigm. The underlying assumption is that heterogeneous views of the same system state should remain distributionally consistent under nominal conditions, while anomalous perturbations disrupt such alignment. To capture this discrepancy and avoid trivial collapse, we employ an asymmetric consistency loss with a stop-gradient operator ($\perp$), where a symmetrized Kullback-Leibler (KL) divergence measures the latent distribution mismatch between H_{i1} and H_{i2}. By blocking gradient propagation through the target view, the online view is compelled to predict the target distribution, enabling robust feature disentanglement with minimal computational overhead.

The overall consistency objective $\mathcal{L}_{total}$ is reformulated as the expectation over the sample set:

$$\mathcal{L}_{total} = \mathbb{E}_{x \in \mathcal{X}} \left[\mathcal{D}_{sym}(H_{i1}, H_{i2}) + \mathcal{D}_{sym}(H_{i1}, H_{i2}) \right] \quad (11)$$

where $\mathcal{D}_{sym}(\cdot, \cdot)$ denotes a bidirectional KL divergence with gradient constraints:

$$\mathcal{D}_{sym}(P, Q) = D_{KL}(P \parallel \perp [Q]) + D_{KL}(\perp [Q] \parallel P) \quad (12)$$

In this formulation, $\perp [\cdot]$ indicates that its operand is treated as a constant during backpropagation, and $D_{KL}(\cdot \parallel \cdot)$ represents the standard KL divergence.

Conventional reconstruction-based anomaly detection methods are often fragile in heterogeneous feature spaces, as they rely on high-quality prior representations that are difficult to obtain under strong distributional shifts and temporal dynamics. Reconstruction objectives further emphasize input fidelity rather than discriminative dynamics, leading to unstable performance. In contrast, DCRFA dispenses with reconstruction and adopts a purely self-supervised representation discrepancy learning paradigm. By directly modeling cross-view semantic differences, it learns more discriminative and robust representations for anomaly detection in complex temporal settings.

G. Anomaly Criterion

Leveraging the inductive bias of self-supervised contrastive learning, DCRFA enforces consistency between the interaction view H_{i1} and the reference view H_{i2} exclusively along the nominal data manifold. During inference, deviations from normal physical modes disrupt this alignment, causing a marked increase in distributional divergence. Therefore, we define the anomaly score $\mathcal{A}(x_t)$ at time t as the cumulative bidirectional KL divergence:

$$\mathcal{A}(x_t) = D_{KL}(H_{i1}^{(t)} \parallel \perp H_{i2}^{(t)}) + D_{KL}(H_{i2}^{(t)} \parallel \perp H_{i1}^{(t)}) \quad (13)$$

Because anomalies induce distributional shifts that manifest as long-tail high values in $\mathcal{A}(x_t)$, the final binary anomaly label $\mathcal{Y}_t \in \{0, 1\}$ can be concisely determined by applying an empirical threshold ξ, i.e., $\mathcal{Y}_t = \mathbb{I}(\mathcal{A}(x_t) \geq \xi)$.

III. Experiments

A. Experimental Setup

Datasets We evaluate our method on three publicly available datasets: MSL [22], SMAP [22], and OPSSAT [23], which are highly relevant and representative for aerospace CPS time-series anomaly detection. MSL and SMAP are real telemetry benchmarks released by NASA, containing anomalies derived from spacecraft monitoring system event reports. MSL consists of 27-dimensional sensor telemetry streams, while SMAP includes 55-dimensional satellite monitoring streams. Their training subsets are unlabeled and may contain unknown noise or minor nominal drifts, whereas the test subsets provide precisely annotated anomalous events. OPSSAT, released by the European Space Agency, comprises thousands of time steps of in-orbit satellite telemetry, providing a realistic setting for anomaly detection research.

Implementation details Experiments are conducted using the PyTorch 3.9.24 framework on an NVIDIA A40 GPU. For DCRFA, the hidden dimension d_{model} is set to 256, and the optimizer is initialized with a learning rate of 10^{-4}. Considering the physical characteristics of the aerospace datasets, we configure the sliding window T, number of encoder layers L, and anomaly threshold ξ as follows: MSL: $(90, 1, 1.1)$; SMAP: $(105, 3, 0.8)$; OPSSAT: $(105, 3, 1.2)$. Evaluation follows the point-adjusted strategy [24], where any segment containing at least one anomalous point is classified as anomalous.

Baselines We compare our method with nine representative baselines, including LOF [25], OmniAnomaly [26], InterFusion [27], THOC [28], TranAD [29], Anomaly-Trans [24], TS2Vec [30], D^3R [21], and DACAD [31]. The selected baselines cover anomaly detection paradigms commonly adopted in multivariate time-series analysis. LOF represents classical distance-based anomaly detection methods, while OmniAnomaly, InterFusion, and THOC exemplify probabilistic and generative reconstruction-based approaches. TranAD and Anomaly Transformer correspond to state-of-the-art Transformer-based architectures for multivariate time-series modeling. In addition, DACAD, TS2Vec, and D^3R reflect recent advances in self-supervised representation learning frameworks for time-series data.

B. Performance

Table I reports the quantitative comparison results on the MSL, SMAP, and OPSSAT datasets using Precision (P), Recall (R), and F1-score (F1). Overall, DCRFA achieves the best or near-best performance across all three datasets, particularly in terms of the F1-score, indicating a more balanced detection capability under diverse anomaly characteristics. On OPSSAT, DCRFA obtains the highest Precision and F1-score while maintaining competitive Recall, demonstrating its robustness in modeling complex satellite telemetry data. Similar performance trends are observed on MSL and SMAP, where DCRFA consistently achieves the top F1-score, despite certain baselines exhibiting superior performance on individual metrics. For example, TranAD reaches a Recall of 99.99% on both OPSSAT and SMAP; however, its comparatively lower Precision results in inferior F1-score performance, highlighting the imbalance caused by optimizing a single metric.

The performance gains of DCRFA arise from the joint effect of differential representation learning and spatiotemporal modeling. Differential representations mitigate biased attention allocation in Transformer architectures, while spatiotemporal modeling enhances the capture of latent temporal patterns and inter-variable dependencies. As a result, DCRFA improves upon Transformer-based baselines by achieving more stable and balanced performance across datasets.

C. Ablation study

To evaluate the contribution of core components in the DCRFA framework, we designed two ablation variants: (1)

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT MODELS (ALL RESULTS ARE IN %.)

Dataset	Metric	LOF	OmniAnomaly	InterFusion	THOC	TranAD	AnomalyTrans	TS2Vec	D^3R	DACAD	DCRFA
OPSSAT	P	56.42	66.06	87.02	46.22	83.33	97.92	81.43	55.31	53.04	**99.99**
	R	55.57	45.8	85.43	92.19	**99.99**	96.11	64.34	**99.99**	87.32	96.66
	F1	55.65	85.6	86.22	61.57	90.91	97.01	71.88	71.23	65.99	**98.30**
MSL	P	47.72	89.02	81.28	88.07	90.38	91.92	92.07	87.05	20.81	**94.04**
	R	85.25	86.37	92.70	92.39	95.78	96.03	**99.99**	92.54	44.66	**99.99**
	F1	61.18	87.67	86.62	90.18	93.04	93.93	95.87	89.71	28.39	**96.93**
SMAP	P	58.93	92.49	89.77	91.51	80.43	93.59	76.61	83.45	99.99	**97.61**
	R	56.33	81.99	88.52	90.40	**99.99**	99.41	**99.99**	94.19	72.93	**99.99**
	F1	57.60	86.92	89.14	90.05	89.15	96.41	86.76	88.50	84.35	**98.79**

Temporal-only, retaining only the temporal association module to assess detection using global frequency-domain features, and (2) Spatial-only, retaining only the spatial association module. Results are summarized in Table II. Both variants exhibit notable performance degradation compared with full DCRFA, with the Temporal-only variant showing the most significant drop, highlighting the effectiveness of the frequency-domain attention mechanism in capturing long-range periodic patterns and suppressing high-frequency noise.

Furthermore, Table III validates the necessity of the bidirectional stop-gradient strategy. Removing the operator $\Phi(\cdot)$ completely induces severe feature collapse and significant F1-score degradation across all datasets. While a single-view constraint yields partial recovery, the bidirectional mechanism proves essential for achieving optimal performance, ensuring stable convergence and robust nominal pattern learning within the positive-only contrastive framework.

TABLE II
THE ABLATION EXPERIMENTS OF DCRFA ON MSL AND SMAP (ALL RESULTS ARE IN %. THE BEST RESULTS ARE IN BOLD).

Dataset	MSL			SMAP		
	P	R	F1	P	R	F1
$DCRFA_{time}$	91.21	96.01	93.65	92.76	93.52	93.14
$DCRFA_{spa}$	90.42	94.03	92.19	91.34	97.63	94.38
DCRFA	91.68	99.30	**95.34**	94.50	98.79	**96.60**

TABLE III
EFFECT OF THE STOP-GRADIENT STRATEGY ON DCRFA PERFORMANCE (%, BEST IN BOLD).

view1	view2	MSL			SMAP			OPSSAT		
		P	R	F1	P	R	F1	P	R	F1
×	×	90.31	98.34	86.14	79.96	98.13	88.87	98.46	92.35	96.58
✓	×	92.32	97.34	91.61	79.32	95.47	88.40	92.01	95.43	96.64
×	✓	79.31	98.11	88.42	79.87	93.29	88.81	98.16	92.37	95.62
✓	✓	94.04	99.30	**96.93**	97.61	99.79	**98.79**	99.88	96.66	**98.30**

D. Sensitivity analysis

To systematically evaluate DCRFA's robustness under varying hyperparameters, we conducted sensitivity experiments on the number of frequency attention heads (H), hidden dimension (d_{model}), sliding window size (T), and encoder depth (L), as shown in Fig. 2. The model achieved optimal performance with a single attention head ($H = 1$), while additional heads degraded accuracy. Performance varies with d_{model} reaching the highest value at $d_{model} = 256$. In contrast, the model was less sensitive to encoder depth, as increasing layers ($L > 1$) offered no gains and even slightly degraded results, further confirming the effectiveness of a shallow frequency-domain structure in practice.

E. Model Validation

To validate the proposed positive-only contrastive paradigm, DCRFA was benchmarked against mainstream contrastive frameworks, SimCLR [32] and MoCo [33], on the MSL, SMAP, and OPSSAT datasets. Unlike these baselines, which rely on intra-batch negative samples, DCRFA employs an asymmetric stop-gradient SimSiam approach, aligning only positive pairs to achieve feature disentanglement. As shown in Fig. 4, DCRFA consistently outperforms all baselines, demonstrating robustness to sparse anomalies. Crucially, DCRFA also achieves superior computational efficiency.

By eliminating negative-sample computation and leveraging frequency-domain attention to reduce temporal dependency modeling to $O(T \log T)$, the framework simultaneously lowers computational cost and enhances detection accuracy and inference speed under practical deployment constraints.

To evaluate the effectiveness of the symmetric KL divergence in the contrastive objective, we compared it with Jensen-Shannon (JS) divergence and asymmetric KL divergence, as shown in Fig. 3. Results indicate that asymmetric KL yields suboptimal alignment, while JS divergence leads to severe performance deterioration. This highlights the advantage of symmetric KL: its bidirectional constraint provides stronger regularization than one-way or smoothed metrics, enabling the model to more effectively capture discriminative dependencies in the latent space during end-to-end training.

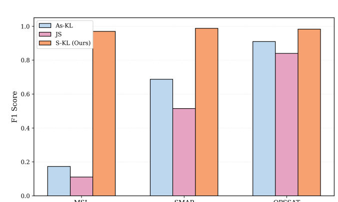

Fig. 3. Empirical evaluation of KL-based divergence variants in DCRFA on three datasets.

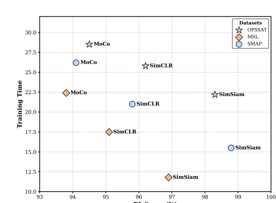

Fig. 4. Comparative evaluation of different contrastive learning paradigms on three datasets.

IV. CONCLUSION

We propose DCRFA, a spectral-spatial self-supervised anomaly detection framework for aerospace cyber-physical systems (CPS). To address non-stationary trends and strong environmental noise in telemetry data, it employs a spectral-spatial encoder that captures global periodic patterns via FFT and learnable frequency filters, while modeling inter-sensor dependencies in the spatial domain. This design decouples high-frequency dynamics from low-frequency drifts and mitigates noise sensitivity under $O(T \log T)$ complexity. Leveraging an asymmetric contrastive learning paradigm with stop-gradient, DCRFA achieves robust alignment of nominal physical manifolds without negative samples. Extensive evaluations on MSL, SMAP, and OPSSAT demonstrate superior performance over nine baselines. Future work will explore pruning, quantization, and knowledge distillation to enable lightweight deployment on resource-constrained platforms.

References

[1] S. Chen, G. Jin, and X. Long, "On-orbit satellite hierarchical anomaly detection using causal structure learning," *Advances in Space Research*, vol. 75, no. 1, pp. 718–736, 2025.

[2] J. Jiang, Y. Zhang, and S. Zhang, "Deep vision transformer with association divergence for image anomaly detection and localization," in *International Conference on Neural Information Processing*. Springer, 2024, pp. 1–15.

[3] S. Zhang, X. Hu, and J. Liu, "Causal fusion of convolutional neural network and vision transformer for image anomaly detection and localization," in *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2024, pp. 1–6.

[4] S. Zhang and J. Liu, "Anomalous signal detection for cyber-physical systems using interpretable causal neural network," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[5] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM Computing Surveys*, vol. 54, no. 3, pp. 56:1–56:33, 2022.

[6] S. Zhang, X. Hu, and J. Liu, "Tranbf: Deep transformer networks and bayesian filtering for time series anomalous signal detection in cyber-physical systems," in *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2024, pp. 1–6.

[7] S. Zhang and J. Liu, "Feature-constrained and attention-conditioned distillation learning for visual anomaly detection," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 2945–2949.

[8] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. Yao, "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities," *ACM Computing Surveys*, vol. 54, no. 5, pp. 1–36, 2021.

[9] Z. Dong, K. Liu, D. Han, Y. Cao, and Y. Xia, "Reconstruction-based multi-scale anomaly detection for cyber-physical systems," in *Proceedings of the International Conference on Industrial Artificial Intelligence (IAI)*, 2022, pp. 1–6.

[10] S. Liu, B. Zhou, Q. Ding, B. Hooi, Z. Zhang, H. Shen, and X. Cheng, "Time series anomaly detection with adversarial reconstruction networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 4293–4306, 2023.

[11] N. Bai, X. Wang, R. Han, Q. Wang, and Z. Liu, "Paformer: Anomaly detection of time series with parallel-attention transformer," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 2, pp. 3315–3328, 2025.

[12] H. Kang and P. Kang, "Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism," *Knowledge-Based Systems*, vol. 290, p. 110000, 2024.

[13] S. Zhang, J. Huang, W. Tang, Y. Wu, T. Hu, X. Xu, and J. Liu, "Dimsod: A diffusion-based framework for multi-modal salient object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 10, 2025, pp. 10 103–10 111.

[14] S. Zhang, J. Huang, W. Tang, L. Tian, Y. Wei, and J. Liu, "Multi-modal salient object detection via a unified diffusion model," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[15] S. Zhang, J. Huang, Y. Wu, T. Hu, W. Tang, and J. Liu, "Seg-diffusion: Text-to-image diffusion model for open-vocabulary semantic segmentation," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[16] S. Zhang, J. Huang, S. Chen, Y. Wu, T. Hu, and J. Liu, "Sod-diffusion: Salient object detection via diffusion-based image generators," in *Computer Graphics Forum*, vol. 43, no. 7. Wiley Online Library, 2024, p. e15251.

[17] S. Zhang, J. Huang, W. Tang, J. Liu, L. HAN, J. Li, H. Yuan, and Z. Fan, "Simplediffusion: A lightweight and efficient conditional diffusion model for multi-modal salient object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 15, 2026, pp. 12 753–12 761.

[18] X. Piao, Z. Chen, T. Murayama, Y. Matsubara, and Y. Sakurai, "Fredformer: Frequency debiased transformer for time series forecasting," *arXiv:2406.09009*, 2024.

[19] H. Xu, Y. Wang, S. Jian, Q. Liao, Y. Wang, and G. Pang, "Calibrated one-class classification for unsupervised time series anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 11, pp. 5723–5736, 2024.

[20] M. Catillo, A. Pecchia, and U. Villano, "Cps-guard: Intrusion detection for cyber-physical systems and iot devices using outlier-aware deep autoencoders," *Computers & Security*, vol. 129, p. 103210, 2023.

[21] C. Wang, Z. Zhuang, Q. Qi, J. Wang, X. Wang, H. Sun, and J. Liao, "Drift doesn't matter: Dynamic decomposition with diffusion reconstruction for unstable multivariate time series anomaly detection," in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 2023, pp. 10 758–10 774.

[22] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 387–395.

[23] B. Ruszczak, K. Kotowski, D. Evans, and J. Nalepa, "The ops-sat benchmark for detecting anomalies in satellite telemetry," *Scientific Data*, vol. 12, p. 710, 2025.

[24] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," in *Proceedings of the International Conference on Learning Representations*, 2022, pp. 1–20.

[25] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers," *ACM SIGMOD Record*, vol. 29, no. 2, pp. 93–104, 2000.

[26] 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 and Data Mining*, 2019, pp. 2828–2837.

[27] Z. Li, Y. Zhao, J. Han, Y. Su, R. Jiao, X. Wen, and D. Pei, "Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2021, pp. 3220–3230.

[28] L. Shen, Z. Li, and J. T. Kwok, "Time-series anomaly detection using temporal hierarchical one-class network," in *Advances in Neural Information Processing Systems*, 2020, pp. 13 016–13 026.

[29] S. Tuli, G. Casale, and N. R. Jennings, "Tranad: Deep transformer networks for anomaly detection in multivariate time series data," *Proceedings of the VLDB Endowment*, vol. 15, no. 6, pp. 1201–1214, 2022.

[30] Z. Yue, J. Wang, J. Sun, and Chen, "Ts2vec: Towards universal representation of time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022, pp. 8980–8987.

[31] Z. Z. Darban, Y. Yang, G. I. Webb, C. C. Aggarwal, Q. Wen, S. Pan, and M. Salehi, "Dacad: Domain adaptation contrastive learning for anomaly detection in multivariate time series," *IEEE Transactions on Knowledge and Data Engineering*, vol. 37, no. 8, pp. 4485–4496, 2025.

[32] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proceedings of the International Conference on Machine Learning*, 2020, pp. 1597–1607.

[33] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9729–9738.