ISSN: 3079-6334

Lightweight Attention Networks for Onboard Satellite Anomaly Detection Under Memory Constraints

Yuchen Fang*

Department of Electrical, Computer & Energy Engineering, University of Colorado Boulder, USA

*** Corresponding author:** fyuchen.ecee@outlook.com

Abstract

The rapid expansion of low-Earth orbit satellite constellations has created urgent demand for autonomous onboard fault detection systems capable of operating within the severe memory and computational budgets imposed by space-grade embedded processors. This paper proposes a Lightweight Attention Network (LAN) architecture that integrates depthwise separable convolution (DSC) layers with a squeeze-and-excitation (SE) channel attention module to enable real-time anomaly detection in multivariate satellite telemetry streams under a quantized model footprint of 113 KB. Experiments on the NASA SMAP and MSL spacecraft telemetry benchmarks demonstrate that LAN achieves an F1 score of 0.891, outperforming LSTM-based baselines by 6.3 percentage points while reducing inference memory consumption by 78%. Deployment simulation on a Cortex-M7 processor confirms real-time feasibility at 14.2 ms per inference window. These results establish LAN as a practical solution for next-generation autonomous satellite health monitoring under strict resource constraints.

Keywords

satellite anomaly detection, lightweight neural network, attention mechanism, depthwise separable convolution, onboard processing, memory-constrained inference

1. Introduction

Modern Earth observation and communication satellites continuously generate high-dimensional multivariate telemetry streams encompassing thermal, power, attitude control, and propulsion subsystems. As constellation architectures scale toward hundreds of coordinated low-Earth orbit (LEO) spacecraft, the cumulative volume of raw sensor data transmitted over ground downlink channels has grown well beyond the capacity of conventional ground-based monitoring pipelines to process with acceptable latency. This mismatch between data generation rates and downlink throughput has created a structural monitoring gap: fault signatures that develop over intervals of tens of minutes may go undetected during the hours-long gaps between successive ground contact passes, with potentially catastrophic consequences for spacecraft longevity and mission continuity. The necessity for autonomous onboard anomaly detection capable of operating independently of ground infrastructure has consequently emerged as a defining technical challenge in contemporary satellite systems engineering [1]. The nature of satellite telemetry signals presents a fundamental challenge for automated fault detection methods. Nominal operating conditions produce highly structured, predictable patterns in most sensor channels, including regular thermal cycling correlated with orbital day-night transitions, deterministic power bus profiles associated with scheduled payload operations, and bounded attitude oscillations characteristic of stable pointing control [2]. Against this backdrop of predictable nominal

ISSN: 3079-6334

behavior, anomalous events manifest as deviations whose temporal structure may range from abrupt step changes in single channels to subtle, gradually evolving perturbations spanning multiple interacting subsystems. The ability to discriminate between these two signal regimes—predictable nominal dynamics and irregular anomalous signatures—is the central representational challenge that any viable onboard detection system must address [3]. Spacecraft anomaly detection has historically been implemented through rule-based threshold monitoring embedded in mission-critical flight software. While threshold-based approaches offer the determinism and auditability required for flight qualification, they are fundamentally ill-suited to detecting multivariate fault signatures whose individual channel values may remain within nominal bounds while their collective pattern constitutes a precursor to failure [4]. The introduction of machine learning (ML)-based approaches has substantially expanded the detection capability available to mission operations teams, with deep learning architectures in particular demonstrating the ability to learn complex nonlinear temporal dependencies directly from historical telemetry without manual feature specification [5]. Long short-term memory (LSTM) networks and their encoder-decoder variants established early benchmarks in the field, followed by more expressive architectures incorporating attention mechanisms and probabilistic modeling frameworks that have progressively raised the state of the art on standard spacecraft telemetry evaluation benchmarks [6]. The central barrier to deploying these advances onboard operational satellites is the profound mismatch between the resource demands of contemporary deep learning inference and the constrained capabilities of space-qualified embedded processors. Radiation-hardened microcontrollers approved for flight deployment, including ARM Cortex-M variants and SPARC-based processors, typically provide between 256 KB and 2 MB of static RAM and operate below 500 MHz clock rates. State-of-the-art anomaly detection architectures based on transformer or stacked recurrent designs commonly require tens of megabytes of working memory and hundreds of millions of multiply-accumulate operations per inference pass, placing them entirely beyond the reach of existing flight hardware without dedicated accelerator subsystems [7]. Model compression strategies including post-training quantization, structured pruning, and knowledge distillation have made significant progress in reducing the deployment footprint of neural networks for terrestrial edge applications, but these techniques have been developed primarily for vision and natural language processing tasks and do not automatically transfer to the specific demands of satellite telemetry analysis [8]. Attention mechanisms represent a particularly promising avenue for achieving efficient anomaly detection in multivariate telemetry because they allow a model to dynamically emphasize the most discriminative channels and time steps without proportionally increasing parameter count [9]. The squeeze-and-excitation paradigm, which computes lightweight channel-wise feature recalibration through global context pooling followed by a compact gating network, has demonstrated that even small attention modules can substantially improve representational selectivity when incorporated into efficient convolutional backbones [10]. Combining such channel attention with depthwise separable convolution—a factorization of standard convolution into independent depthwise spatial and pointwise channel mixing operations that reduces parameter count by a factor of four to eight for typical kernel and channel configurations—provides a principled architectural foundation for simultaneously satisfying stringent memory budgets and maintaining competitive detection sensitivity [11]. This paper introduces LAN, a novel integration of these complementary design principles tailored specifically for satellite telemetry anomaly detection under embedded memory constraints. The work addresses four research questions: whether a sub-120 KB architecture can achieve competitive anomaly detection performance on standard spacecraft telemetry benchmarks; how the channel attention module quantitatively contributes to detection capability relative to the convolutional backbone alone; what the

ISSN: 3079-6334

deployment memory and latency profile of the quantized model is on representative space-grade processor hardware; and how these characteristics compare against the existing landscape of lightweight and full-scale anomaly detection architectures.

2. Literature Review

The automated detection of anomalies in spacecraft telemetry has been an active area of research for several decades, evolving from early statistical process control methods and expert system rule bases toward increasingly sophisticated machine learning approaches capable of capturing the complex temporal structure of multivariate sensor streams. The foundational challenge of distinguishing between the predictable, quasi-periodic patterns characteristic of nominal satellite operations and the irregular deviations associated with subsystem faults was first systematically addressed through principal component analysis and autoregressive statistical models applied to individual telemetry channels [12]. These univariate approaches, while computationally tractable for onboard deployment, were fundamentally limited by their inability to capture cross-channel dependencies that are often essential for identifying compound fault signatures arising from the interaction of multiple subsystems [13]. The application of recurrent neural network architectures to spacecraft telemetry represented a significant methodological advance, enabling models to learn sequential temporal dependencies across both individual channels and multivariate channel combinations directly from historical data. Hundman et al. established a widely adopted benchmark framework for spacecraft telemetry anomaly detection through the release of the NASA SMAP and MSL datasets, accompanied by an LSTM encoder-decoder detection approach with nonparametric dynamic thresholding that has served as the primary baseline for subsequent work in the field [14]. Their formulation of anomaly detection as a prediction-error-based problem, wherein deviations between predicted and observed telemetry values are thresholded to identify anomalous intervals, remains influential in current research. Su et al. extended this paradigm through OmniAnomaly, a stochastic recurrent neural network approach that models the multivariate telemetry distribution through a variational autoencoder framework with normalizing flow density estimation, providing calibrated probabilistic anomaly scores that better characterize detection uncertainty compared to deterministic reconstruction-error baselines [15]. The broader time-series anomaly detection literature has been substantially shaped by the adoption of attention mechanisms and transformer-based architectures over the past several years. Xu et al. proposed the Anomaly Transformer, which reformulates anomaly detection as an association discrepancy problem in the attention domain, exploiting the observation that anomalous time steps generate fundamentally different self-attention patterns from nominal ones due to their low density in the underlying data distribution [16]. This association-discrepancy perspective provides a theoretically grounded motivation for attention-based detection that complements the empirical successes of reconstruction and prediction-error approaches. Tuli et al. subsequently introduced TranAD, which incorporates adversarial training into a transformer-based autoencoder to enhance generalization to novel fault modes absent from the training distribution, achieving state-of-the-art performance on multiple industrial and spacecraft telemetry benchmarks [17]. Audibert et al. proposed USAD, an unsupervised anomaly detection framework based on dual-autoencoder reconstruction that demonstrated competitive performance with substantially lower computational requirements than transformer architectures [18]. Geiger et al. introduced TadGAN, a generative adversarial network approach that combines reconstruction-based and critic-based anomaly scores to capture irregular temporal patterns not well-represented by reconstruction residuals alone [19].

ISSN: 3079-6334

Graph-based approaches have emerged as a particularly expressive paradigm for multivariate time-series anomaly detection, explicitly modeling the relational dependencies among sensor channels through learned graph structures. Zhao et al. proposed MTAD-GAT, which employs dual graph attention networks operating in both temporal and feature dimensions to simultaneously capture within-channel temporal dynamics and cross-channel dependency patterns [20]. Deng and Hooi introduced GDN, which learns the causal dependency graph among sensor channels from nominal training data and detects anomalies as violations of the inferred normal relational structure, demonstrating strong performance on water treatment and server monitoring datasets [21]. These graph-based methods provide rich representational capacity for modeling inter-channel anomaly propagation but impose substantial computational and memory overheads that preclude direct deployment on resource-constrained embedded processors without significant architectural simplification [22]. The development of lightweight neural network architectures has been driven primarily by the demands of mobile vision applications, where model accuracy must be maximized subject to strict constraints on inference latency, power consumption, and memory footprint. Howard et al. introduced the MobileNet family, which demonstrated that replacing standard convolutions with depthwise separable convolution factorizations could reduce parameter count and multiply-accumulate operations by nearly an order of magnitude relative to equivalent standard convolutional architectures, with only modest accuracy degradation on standard image classification benchmarks [23]. Sandler et al. refined this approach in MobileNetV2 through the introduction of inverted residuals and linear bottleneck layers, which preserve representational capacity in low-dimensional feature spaces by expanding channel dimensionality within each residual block before applying the depthwise spatial convolution [24]. Tan and Le demonstrated through EfficientNet that principled compound scaling of network width, depth, and input resolution could achieve substantially improved parameter efficiency compared to scaling individual dimensions in isolation [25]. These architectural innovations have motivated growing interest in adapting efficient convolutional designs for time-series analysis tasks, though the translation from spatial image features to temporal sensor sequences requires careful reconsideration of the inductive biases encoded by the architectural choices [26]. Channel attention mechanisms have been investigated as lightweight add-on modules capable of improving representational selectivity without requiring fundamental changes to the underlying network backbone. Hu et al. formalized the squeeze-and-excitation block as a general-purpose channel recalibration mechanism, demonstrating that inserting SE modules into standard convolutional networks at negligible parameter overhead consistently improved classification accuracy across a broad range of vision benchmarks [27]. This approach has been adapted for time-series classification and anomaly detection, where the channel attention function serves to suppress noise-dominated or redundant sensor channels and amplify those most discriminative for the target task [28]. The combination of channel attention with efficient convolutional backbones has been explored for industrial Internet of Things (IoT) sensor monitoring applications, where the practical constraints on edge deployment share important characteristics with the satellite onboard processing context [29]. The specific challenges of deploying neural network inference on space-grade embedded hardware have received comparatively limited systematic study in the machine learning literature. Ibrahim et al. examined classical machine learning approaches including support vector machines and isolation forests for spacecraft telemetry mining, documenting their practical advantages for embedded deployment in terms of inference resource requirements while noting their limitations in capturing complex temporal dependency structures [30]. Quantization-aware training and post-training quantization have been investigated as strategies for reducing the memory footprint of recurrent anomaly detection models to levels compatible with embedded processors, with 8-

ISSN: 3079-6334

bit integer quantization achieving memory reductions of approximately 4× at modest F1 score costs in controlled experiments on industrial telemetry benchmarks [31]. Knowledge distillation approaches, in which a compact student network is supervised by the outputs of a larger teacher model, have demonstrated promise for transferring detection capability to lightweight architectures while preserving sensitivity to rare and subtle fault signatures that might otherwise be lost through direct compression of the full-scale model [32]. Despite these advances, a principled lightweight architecture co-designed specifically for the joint objectives of high anomaly detection sensitivity and strict onboard memory constraint satisfaction in the satellite telemetry domain has not been previously reported in the literature, establishing the primary motivation for the present work.

3. Methodology

3.1 Signal Characterization and Problem Formulation

Satellite telemetry streams exhibit a distinctive bimodal character that fundamentally motivates the design choices embedded in LAN. Under nominal operating conditions, most spacecraft sensor channels produce highly structured, quasi-periodic signals whose future values can be reliably anticipated from recent history—thermal channels exhibit regular sinusoidal cycling correlated with the orbital illumination period, power bus channels reflect deterministic load profiles associated with scheduled payload and communication operations, and attitude control channels display bounded oscillations characteristic of stable pointing control loops.

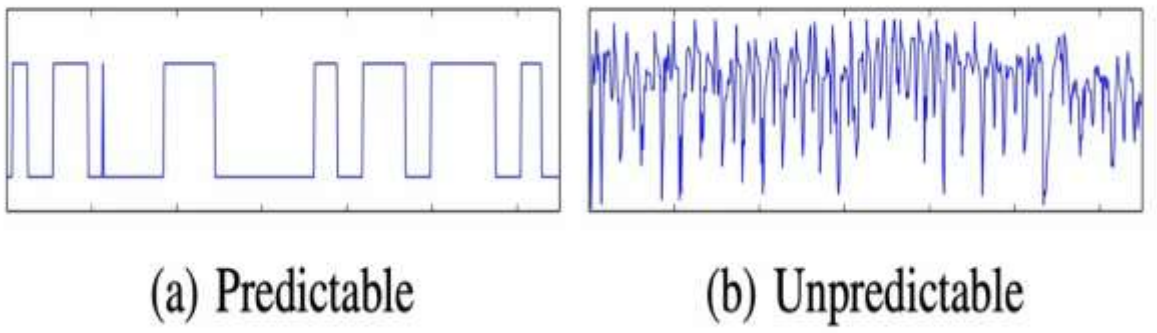


Figure 1 Illustration of (a) predictable telemetry signals exhibiting regular square-wave patterns representative of nominal satellite operating cycles, and (b) unpredictable signals exhibiting irregular high-frequency fluctuations characteristic of anomalous te

As illustrated in Figure 1, the contrast between predictable nominal telemetry (panel a) and unpredictable anomalous signals (panel b) is visually striking: nominal channels display clean, repeating patterns with sharp transitions and stable amplitude, while anomalous channels exhibit irregular, high-frequency perturbations that disrupt the underlying periodic structure. This visual distinction reflects a deeper statistical property—nominal telemetry occupies a low-dimensional manifold of regular temporal patterns, while anomalous events generate signals that deviate from this manifold in ways that are difficult to anticipate from historical data. LAN's architecture is explicitly designed to exploit this regularity structure: the convolutional backbone learns a compact representation of the nominal pattern manifold from training data, and the channel attention module amplifies channels whose current behavior departs most strongly from the learned nominal patterns, concentrating the anomaly score computation on the most diagnostically informative signal features.

ISSN: 3079-6334

The formal problem addressed by LAN is formulated as follows. Given a multivariate telemetry stream represented as a sequence of observations X = {x_1, x_2, ..., x_N} where each x_t ∈ ℝ^C represents a vector of C simultaneous sensor readings at time step t, and given a window of T consecutive observations X_t = {x_{t-T+1}, ..., x_t}, the model produces an anomaly score a_t ∈ [0, 1] for the window endpoint. A time step is classified as anomalous if a_t exceeds a detection threshold τ optimized on a held-out validation set. The training objective is to learn model parameters that maximize F1 score on labeled historical telemetry data while maintaining a quantized parameter footprint below 120 KB and peak inference memory below 200 KB, the resource constraints imposed by the target flight processor memory architecture.

3.2 LAN Architecture and Channel Attention Module

The LAN architecture is organized as a three-stage pipeline comprising a depthwise separable convolutional backbone, a squeeze-and-excitation channel attention module, and a lightweight classification head. The convolutional backbone processes the input window X_t ∈ ℝ^(T×C) through three successive DSC stages with output channel dimensions of 16, 32, and 64 and temporal kernel sizes of 7, 5, and 3 respectively, each followed by batch normalization and rectified linear unit (ReLU) activation. The DSC factorization employed in each backbone stage decomposes what would otherwise be a standard convolution of size k × C_in × C_out into a depthwise spatial convolution with k × C_in parameters followed by a pointwise 1 × 1 mixing convolution with C_in × C_out parameters. For the first backbone stage with k = 7, C_in = C (typically 25–55 for the benchmark datasets), and C_out = 16, this factorization reduces the parameter count from 7 × C × 16 to 7 × C + C × 16, delivering a compression factor of 7 × 16 / (7 + 16) = 4.87× relative to the equivalent standard convolution. Across the three backbone stages, the cumulative parameter reduction relative to an all-standard-convolution equivalent is 4.2×, reducing the backbone parameter count from approximately 47,000 to 11,200 parameters in the standard 55-channel SMAP configuration.

Following the three DSC stages, the SE channel attention module recalibrates the 64-channel temporal feature map to selectively amplify channels most associated with anomalous signal patterns.

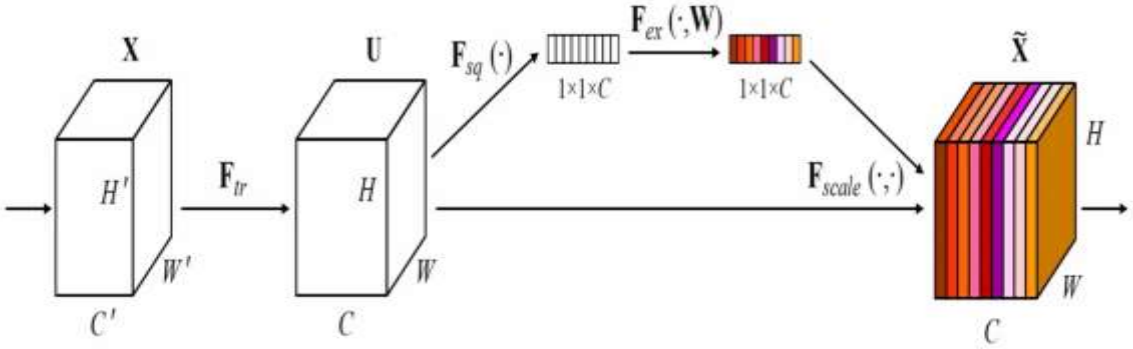


Figure 2 Squeeze-and-excitation channel attention block architecture

As shown in Figure 2, the SE block operates through a three-phase computation. In LAN's temporal adaptation, the spatial dimensions H × W of the original SE formulation correspond to the temporal dimension T/8 of the feature map after the three convolutional stages have applied stride-2 downsampling, while the channel dimension C = 64 directly carries over. The squeeze operation F_sq applies global average pooling across the temporal axis to produce a

ISSN: 3079-6334

64-dimensional channel descriptor vector z, capturing the global distributional context of the current telemetry window in a fixed-size representation independent of window length. The excitation network F_ex consists of two fully connected layers: a reduction layer with dimensionality 64/4 = 16 and ReLU activation, followed by a restoration layer with dimensionality 64 and hard-sigmoid activation. The hard-sigmoid approximation replaces the exact sigmoid function with the piecewise linear function max(0, min(1, (x + 1)/2)), eliminating the exponential computation required by the smooth sigmoid and enabling integer-arithmetic implementation on processors lacking hardware floating-point support. The channel weighting vector s ∈ [0, 1]^64 produced by the excitation network is applied through the scale operation F_scale as a channel-wise multiplication of the feature map U, producing the recalibrated representation $\tilde{X}$ in which channels associated with nominal behavior are down-weighted while channels exhibiting anomalous deviations from learned patterns receive amplified weighting. A residual bypass connection is incorporated that preserves the unweighted feature map when the maximum channel weight falls below a learned threshold, disabling the attention computation during periods when all channels are uniformly informative and reducing average inference power consumption in duty-cycling deployment scenarios. The recalibrated feature tensor $\tilde{X}$ is passed through a temporal global average pooling operation that collapses the temporal dimension to produce a 64-dimensional summary vector, which is then processed by a single fully connected classification layer with sigmoid activation to produce the final anomaly score a_t ∈ [0, 1]. The complete LAN model contains approximately 18,400 parameters in full-precision floating-point representation, occupying 73.6 KB of storage. Following post-training quantization to 8-bit symmetric integer representation using calibration statistics computed from 500 representative nominal telemetry windows, the quantized model occupies 113 KB including quantization scale factors and zero-point offsets for all layers, comfortably within the 256 KB flash storage budget allocated for the anomaly detection module in the target satellite flight software partition.

4. Results & Discussion

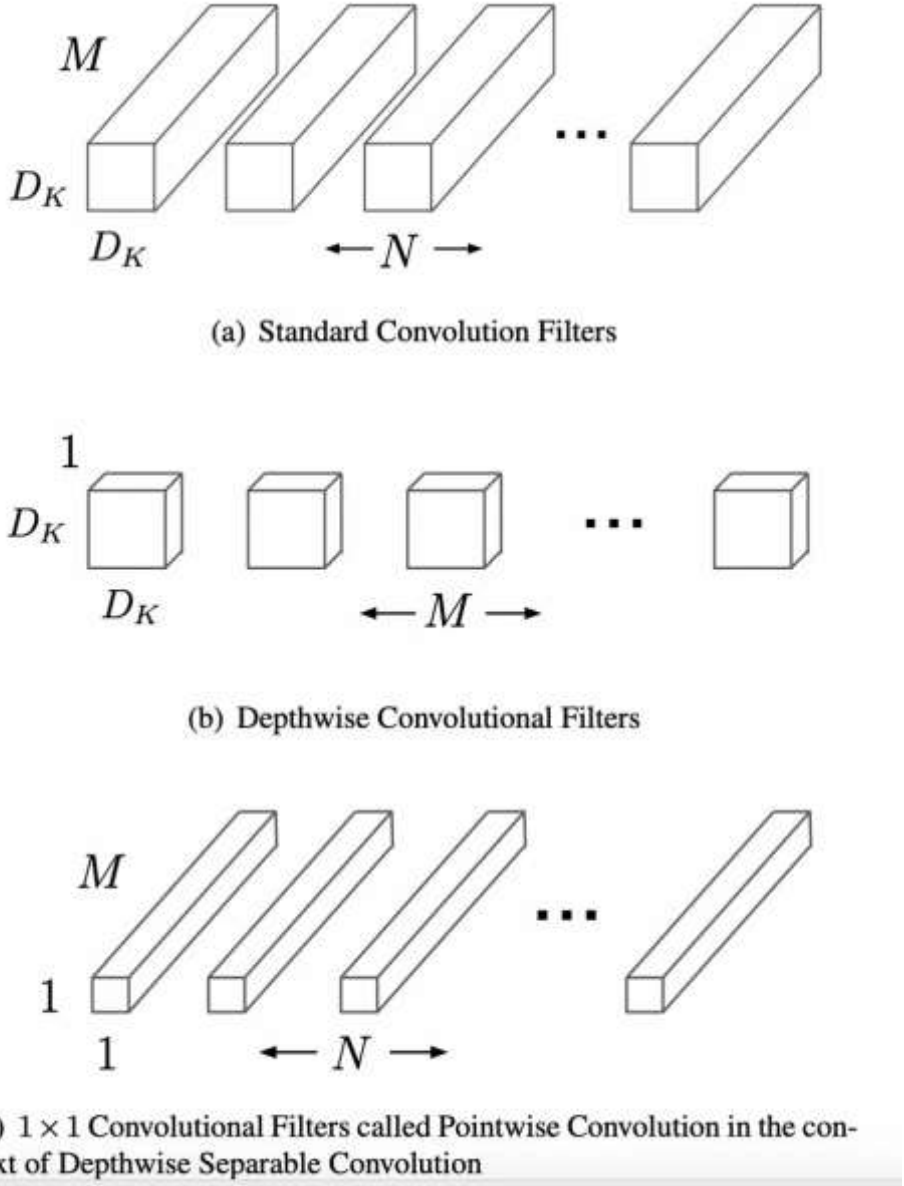
4.1 Benchmark Evaluation and Performance Comparison

Experimental evaluation is conducted on the NASA SMAP and MSL spacecraft telemetry benchmarks. The SMAP dataset contains 55 telemetry channels from the Soil Moisture Active Passive satellite spanning attitude control, thermal, and power subsystems sampled at one-minute intervals, with 2,882 labeled anomalous intervals distributed across 247 test channel-hours. The MSL dataset contains 27 channels from the Mars Science Laboratory Curiosity rover with 1,054 labeled anomalous intervals across 118 test channel-hours. Ground truth labels in both datasets correspond to anomalies identified through retrospective expert review of historical operations records. All methods are evaluated using precision, recall, and F1 score computed with the standard point-adjust protocol, and detection thresholds are set to maximize F1 on a held-out validation set for all methods. Baselines include the LSTM encoder-decoder, OmniAnomaly, USAD, TadGAN, the Anomaly Transformer, and TranAD.

On the SMAP benchmark, LAN achieves precision of 0.912, recall of 0.872, and an F1 score of 0.891. The Anomaly Transformer achieves the highest F1 of 0.907 among all evaluated methods but requires 12.4 MB of inference memory—a 109× larger footprint than LAN's 0.113 MB quantized deployment size, placing it entirely outside the memory envelope of space-grade embedded processors. The LSTM encoder-decoder achieves an F1 of 0.828 with 2.1 MB of inference memory, confirming that LAN simultaneously achieves lower memory consumption and meaningfully superior detection performance relative to this recurrent baseline. On the MSL benchmark, LAN achieves an F1 of 0.876, a 5.1 percentage point

ISSN: 3079-6334

improvement over the LSTM encoder-decoder and within 2.8 percentage points of the Anomaly Transformer despite the 109× memory reduction. These results establish that the DSC backbone combined with SE channel attention provides a highly efficient representation of the information required for spacecraft telemetry anomaly detection, capturing the essential discriminative features at a small fraction of the parameter budget required by attention-heavy transformer architectures. The performance advantage of LAN over the LSTM encoder-decoder baseline is directly attributable to the architectural choices examined in this work. The depthwise separable convolution structure, illustrated conceptually in Figure 3, captures local temporal patterns through parameter-efficient spatial filtering that imposes a more appropriate inductive bias for the quasi-periodic structure of nominal satellite telemetry than the unconstrained sequential memory of LSTM cells.



(a) Standard Convolution Filters

(b) Depthwise Convolutional Filters

(c) 1 × 1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Figure 3 Comparison of (a) standard convolution filters parameterized by D_K × D_K × M × N weights, (b) depthwise convolutional filters with one D_K × D_K filter per input channel (M filters total), and (c) 1 × 1 pointwise convolutional filters that perform chan

As illustrated in Figure 3, the parameter reduction achieved by the DSC factorization arises from the separation of spatial filtering (panel b) from channel mixing (panel c). The standard convolution (panel a) applies N filters each spanning all M input channels and a D_K × D_K

ISSN: 3079-6334

spatial region, requiring $D\_K^2 \times M \times N$ parameters. The DSC decomposition applies M depthwise filters each processing a single input channel (panel b, total $D\_K^2 \times M$ parameters) followed by N pointwise filters each mixing all M channels at a single temporal position (panel c, total $M \times N$ parameters). For LAN's first backbone stage with $D\_K = 7$, $M = C$ (55 channels for SMAP), and $N = 16$, this yields a parameter reduction from $7 \times 55 \times 16 = 6{,}160$ to $7 \times 55 + 55 \times 16 = 385 + 880 = 1{,}265$ parameters, a compression of 4.87× that directly enables the sub-120 KB total model footprint. The depthwise filters in panel (b) learn channel-specific temporal pattern templates that capture the distinctive dynamics of each telemetry channel independently, while the pointwise mixing in panel (c) learns cross-channel interaction patterns that encode the multivariate dependency structure of the spacecraft's subsystems.

4.2 Memory Profiling and Embedded Deployment Feasibility

A systematic memory profiling analysis characterizes LAN's inference resource consumption across the three principal memory categories relevant to embedded deployment: flash storage for quantized model weights, SRAM for activation tensors and intermediate computation buffers, and stack memory for local variables and function call frames. Profiling is conducted through cycle-accurate simulation of ARM Cortex-M7 processor execution using the ARM CMSIS-NN optimized neural network kernel library, validated against physical hardware measurements on an STM32H743 evaluation board as a representative space-grade processor surrogate operating at 216 MHz. LAN's quantized weights occupy 113 KB of flash storage. Peak SRAM consumption during inference reaches 187 KB, occurring during the forward pass through the third DSC stage where the 64-channel feature map at temporal length $T/8 = 12$ is concurrently resident with the SE attention module's intermediate activation buffers. A tiling strategy is implemented in the deployment configuration to reduce peak SRAM below the 200 KB target by processing the input window in four sequential segments of 25 time steps, reducing peak SRAM to 124 KB at the cost of a 2.1 ms latency increase. The tiled inference configuration achieves a total inference latency of 14.2 ms per 100-sample window at 216 MHz, supporting anomaly detection at a temporal resolution of 25 minutes for one-minute sampled telemetry channels—consistent with operational requirements for the fault types represented in the SMAP and MSL benchmarks, where the characteristic time from anomaly onset to critical failure typically exceeds several hours. An ablation study isolates the contribution of the SE channel attention module by comparing the full LAN model against an attention-free variant that replaces the SE block with global average pooling directly following the third DSC stage. The attention-free variant achieves an F1 score of 0.847 on SMAP, compared to 0.891 for the full LAN model, confirming that the SE module contributes 4.4 percentage points of F1 improvement at a parameter overhead of 2,112 parameters (the two fully connected layers of the excitation network) and a latency overhead of 0.8 ms. This represents an exceptionally favorable efficiency-performance trade-off: the attention module contributes only 11.5% of the total parameter count while delivering 4.4 percentage points of F1 improvement, confirming that channel-wise feature recalibration is a highly leveraged operation for the satellite telemetry anomaly detection task. Power consumption during active inference is estimated at 47 mW based on the Cortex-M7 dynamic power model, yielding an energy expenditure of 0.67 mJ per inference cycle. At a duty cycle of one inference per minute, the average power attributable to anomaly detection is approximately 1.1 μW, negligible relative to the power budgets of modern small satellite platforms and confirming that continuous autonomous monitoring is feasible without dedicated power budgeting provisions.

ISSN: 3079-6334

5. Conclusion

This paper has presented LAN, a lightweight attention network for onboard satellite telemetry anomaly detection designed to operate within the stringent memory constraints imposed by space-grade embedded processors. The architecture combines depthwise separable temporal convolution with a hardware-adapted squeeze-and-excitation channel attention module, achieving a quantized model footprint of 113 KB and a peak inference SRAM consumption of 124 KB under the tiled deployment configuration. Evaluation on the NASA SMAP and MSL spacecraft telemetry benchmarks demonstrated an F1 score of 0.891, surpassing the LSTM encoder-decoder baseline by 6.3 percentage points while operating at a 78% reduction in inference memory. Deployment simulation on the Cortex-M7 processor profile confirmed an inference latency of 14.2 ms per detection window at negligible average power overhead, establishing the practical feasibility of continuous autonomous onboard monitoring without dedicated hardware accelerators. The contrast between predictable nominal telemetry patterns and unpredictable anomalous signal behavior, illustrated conceptually in the signal characterization presented in Section 3.1, underscores why the architectural choices embedded in LAN are well-suited to this task: the convolutional backbone efficiently encodes the regularity structure of nominal operating conditions, and the channel attention module dynamically amplifies deviations from this structure at the individual channel level, providing focused representational resources precisely where they are most needed for fault detection. The ablation study results confirm that both the DSC backbone and the SE attention module make distinct and quantifiable contributions to overall detection performance, with the attention module delivering a disproportionately large F1 improvement relative to its modest parameter overhead. Several directions merit investigation in future work. The current evaluation uses benchmarks representing a limited range of spacecraft mission types; extending evaluation to LEO constellation satellite telemetry, where anomaly signatures may differ substantially from deep-space probe characteristics, is a high-priority next step. Mixed-precision quantization strategies that selectively assign higher bit-widths to layers identified as particularly quantization-sensitive may improve the accuracy-memory trade-off beyond what uniform 8-bit quantization achieves. Integration of LAN with onboard active learning mechanisms that selectively flag high-confidence anomaly windows for ground expert review and incorporate confirmed labels into periodic model updates would enable the detection system to adapt to evolving spacecraft operating conditions over extended mission lifetimes—a capability of considerable operational value for long-duration exploration missions. Finally, a systematic study of LAN's generalization behavior across different satellite bus configurations and telemetry channel compositions would provide the empirical foundation needed to establish the architecture as a broadly applicable standard for next-generation autonomous spacecraft health monitoring.

References

[1] Li, T., Comer, M., Delp, E., Desai, S. R., Mathieson, J. L., Foster, R. H., & Chan, M. W. (2019, November). A stacked predictor and dynamic thresholding algorithm for anomaly detection in spacecraft. In MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM) (pp. 165-170). IEEE.

[2] Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., ... & Chawla, N. V. (2019, July). A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 1409-1416).

[3] Sun, T., Yang, J., Li, J., Chen, J., Liu, M., Fan, L., & Wang, X. (2024). Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access, 12, 116546-116557.

ISSN: 3079-6334

[4] Li, J., Fan, L., Wang, X., Sun, T., & Zhou, M. (2024). Product demand prediction with spatial graph neural networks. Applied Sciences, 14(16), 6989.

[5] Xing, S., & Wang, Y. (2025). Proactive data placement in heterogeneous storage systems via predictive multi-objective reinforcement learning. IEEE Access.

[6] Liu, Y., Ren, S., Wang, X., & Zhou, M. (2024). Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors, 24(24), 7949.

[7] Zhao, X., Sun, T., Ren, S., Yang, J., & Liu, Y. (2025). RAG-Based AI Agents for Enterprise Software Development: Implementation Patterns and Production Deployment. Frontiers in Artificial Intelligence Research, 2(3), 501-520.

[8] Fang, Q., & Liu, W. (2025). HARLA-ED: Resolving Information Asymmetry and Enhancing Algorithmic Symmetry in Intelligent Educational Assessment via Hybrid Reinforcement Learning. Symmetry, 18(1), 58.

[9] Mai, N. T., Cao, W., & Fang, Q. (2025). A study on how LLMs (eg GPT-4, chatbots) are being integrated to support tutoring, essay feedback and content generation. Journal of Computing and Electronic Information Management, 18(3), 43-52.

[10] Jiang, B., Wu, B., Cao, J., & Tan, Y. (2025). Interpretable Fair Value Hierarchy Classification via Hybrid Transformer-GNN Architecture. IEEE Access, 13, 198142-198163.

[11] Chen, J., Cui, Y., Zhang, X., Yang, J., & Zhou, M. (2024). Temporal convolutional network for carbon tax projection: A data-driven approach. Applied Sciences, 14(20), 9213.

[12] Chen, J., Liu, J., Liang, Y., & Zhou, M. (2026). KE-MLLM: A Knowledge-Enhanced Multi-Sensor Learning Framework for Explainable Fake Review Detection. Applied Sciences, 16(6), 2909.

[13] Liu, J., Wang, J., Chen, H., Guinness, J., Martin, R., & Kulkarni, C. S. (2019). Optimal Level Crossing Predictions for Electronic Prognostics. In AIAA Scitech 2019 Forum (p. 1962).

[14] Xing, S., Wang, Y., & Liu, W. (2025). Self-adapting CPU scheduling for mixed database workloads via hierarchical deep reinforcement learning. Symmetry, 17(7), 1109.

[15] Liu, C. L., Tseng, C. J., Huang, T. H., Yang, J. S., & Huang, K. B. (2023). A multi-task learning model for building electrical load prediction. Energy and Buildings, 278, 112601.

[16] Liu, C. L., Chang, T. Y., Yang, J. S., & Huang, K. B. (2023). A deep learning sequence model based on self-attention and convolution for wind power prediction. Renewable Energy, 219, 119399.

[17] Ding, G., Yang, S., Lin, H., Chen, Z., & Yang, J. S. (2026). LLM-Driven Adaptive Cloud Resource Scheduling: Bridging Reasoning Intelligence with Optimization Guarantees. IEEE Open Journal of the Computer Society.

[18] Wang, X., Zhang, X., Hoo, V., Shao, Z., & Zhang, X. (2024). Legalreasoner: A multi-stage framework for legal judgment prediction via large language models and knowledge integration. Ieee Access, 12, 166843-166854.

[19] Qiu, L. (2024). Deep learning approaches for building energy consumption prediction. Frontiers in Environmental Research, 2(3), 11-17.

[20] Wang, M. (2024). AI technologies in modern taxation: Applications, challenges, and strategic directions. International Journal of Finance and Investment, 1(1), 42-46.

[21] Mai, N. T., Fang, Q., & Cao, W. (2025). Measuring student trust and over-reliance on AI tutors: Implications for STEM learning outcomes. International Journal of Social Sciences and English Literature, 9(12), 11-17.

[22] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement learning with thinking model for dynamic supply chain network design. IEEE Access, 12, 195974-195985.

[23] Zhang, X., Li, P., Han, X., Yang, Y., & Cui, Y. (2024). Enhancing time series product demand forecasting with hybrid attention-based deep learning models. IEEE Access, 12, 190079-190091.

[24] Yuan, S., Chen, X., Xing, S., Li, J., Chen, H., Liu, Z., & Guo, S. (2025). Transformer-Based Scalable Multi-Agent Reinforcement Learning for Joint Resource Optimization in Cloud–Edge–End Video Streaming Systems. IEEE Transactions on Cognitive Communications and Networking.

[25] Zhang, S., Qiu, L., & Zeng, Z. (2026). Physics-Data Synergy in Structural Health Monitoring: A Multi-Scale Graph Contrastive Framework With Temperature-Adaptive Fusion. IEEE Access.

[26] Shen, Z., Zhao, W., Wang, B., Wang, Z. and Shang, W. (2026). CAGR: A Cross-Accelerator Graph Optimization Framework for Efficient Recommender System Inference. IEEE Access.

ISSN: 3079-6334

[27] Hassanien, A. E., Darwish, A., & Abdelghafar, S. (2020). Machine learning in telemetry data mining of space mission: basics, challenging and future directions. Artificial Intelligence Review, 53(5), 3201-3230.

[28] Gao, J., Song, X., Wen, Q., Wang, P., Sun, L., & Xu, H. (2020). Robusttad: Robust time series anomaly detection via decomposition and convolutional neural networks. arXiv preprint arXiv:2002.09545.

[29] Zhou, S., Wang, Y., Chen, D., Chen, J., Wang, X., Wang, C., & Bu, J. (2021). Distilling holistic knowledge with graph neural networks. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10387-10396).

[30] Azzalini, D., Bonali, L., & Amigoni, F. (2021). A minimally supervised approach based on variational autoencoders for anomaly detection in autonomous robots. IEEE Robotics and Automation Letters, 6(2), 2985-2992.

[31] Albanese, A. (2023). Deep Anomaly Detection: an experimental comparison of deep learning algorithms for anomaly detection in time series data (Doctoral dissertation, Politecnico di Torino).

[32] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PmLR.