

1<sup>st</sup> Chunyu Yang*School of Computer Science**Fudan University*

Shanghai, China

22307140114@m.fudan.edu.cn

**Abstract**—The increase of low Earth orbit (LEO) satellite constellations has intensified interference risks with geostationary (GSO) networks, demanding scalable, real-time detection solutions. Existing deep learning approaches either sacrifice accuracy for efficiency or fail to adapt to dynamic spectral overlaps. We propose DualAttWaveNet, a multimodal autoencoder integrating bidirectional cross-domain attention and multi-scale wavelet regularization to address these challenges. Our framework processes synchronized time-domain and frequency-domain signal, fusing them via a lightweight mutual attention mechanism without computational bottlenecks. Wavelet transform regularization enforces reconstruction fidelity across temporal and spectral resolutions.

**Index Terms**—interference detection, multimodal fusion, bidirectional attention, wavelet transform

## I. INTRODUCTION

The rapid increase of non-geostationary orbit (NGSO) satellite constellations, fueled by demand for low-latency broadband and advancements in cost-effective launch technologies, has intensified orbital congestion in low-Earth orbit (LEO). With thousands of satellites now competing for limited spectrum resources, the risk of cross-system interference between NGSO and geostationary orbit (GSO) networks has reached critical levels. Spectral overlap between these systems disrupts service reliability, as even minor deviations in orbital positioning or beam alignment can generate harmful interference-to-noise ratios ( $I/N$ ) [1]. Mitigating these effects requires real-time, adaptive detection mechanisms capable of discerning interference within dynamic, multi-operator environments. However, there has been an absence of scalable solutions for modern satellite coexistence scenarios.

Standardized regulatory frameworks define preventive interference thresholds and aggregation methodologies to limit harm to critical satellite services. The International Telecommunication Union (ITU) mandates a  $-10$  dB  $I/N$  ceiling over 99.9% of operational intervals to protect GSO receiver integrity [2]. Central to compliance is the Equivalent Power Flux Density (EPFD) metric, a tool codified in ITU Radio Regulations to address distributed interference risks. By aggregating and bounding the cumulative power flux density from all non-geostationary (NGSO) transmitters, EPFD ensures NGSO systems restrict their aggregate emissions at GSO terminals before operations [3].

Traditional interference detection operates reactively through signal analysis after interference occurs. For example, energy detection (ED) identifies anomalies by summing squared signal amplitudes  $E_n = \sum y_n[m]^2$  over a

fixed time window and flagging energy exceeding a predefined threshold [4]. While ED is computationally efficient, its static thresholding fails to distinguish weak interference from noise in low-SNR regimes, leading to missed detections [5]. Enhanced techniques like cyclostationary feature detection improve robustness by exploiting periodic signal properties [6], yet they cause significant computational costs when applied to LEO systems. Two-step hybrid approaches integrating pilot cancellation and hardware-based filtering have demonstrated limited success but lack the adaptability to support evolving satellite architectures [7].

Recent advances in machine learning (ML) seek to mitigate the shortcomings of traditional detection frameworks. Initial efforts utilized deep neural networks, such as LSTM-based architectures [8], to identify interference by processing in-phase/quadrature (IQ) signal components or time-domain amplitude signal. Contemporary methodologies view the detection task as an anomaly discrimination problem, training models to isolate interference from nominal transmissions by learning their intrinsic signatures. Autoencoder-based strategies [9], for instance, reconstruct idealized interference-free signals from raw inputs; significant energy deviations between the input and reconstructed waveforms are flagged as potential interference. Transformers have further pushed detection boundaries, attaining state-of-the-art area-under-curve (AUC) scores by modeling long-range spectral and temporal correlations in signal [5]. Nevertheless, these architectures impose substantial computational burdens during training, making them impractical for latency-sensitive satellite operations.

To overcome these barriers, we propose DualAttWaveNet, a multimodal architecture that unifies time-frequency signal representations, achieving an AUC of 0.9327 while reducing GFLOPS to 5.19. The core contributions include:

- 1) **Bidirectional Cross-Domain Attention:** A lightweight mutual attention mechanism that dynamically fuses temporal waveforms and spectral features, learning interdependencies between domains without concatenation bottlenecks.
- 2) **Multi-Scale Wavelet Regularization:** A loss function enforcing reconstruction fidelity across four predefined wavelet scales, aligning interference patterns with broader spectral trends for robust detection.
- 3) **Our detector is validated using the system model in [5], which synthesizes realistic GSO/LEO coexistence scenarios. This ensures evaluation under interference**

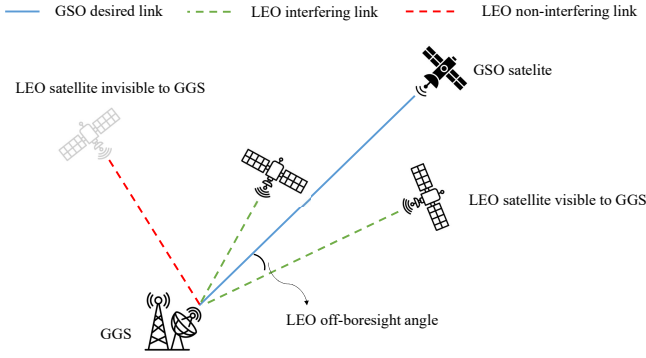


Fig. 1. Interference scenario between GSO and LEO satellite systems.

conditions mirroring actual orbital dynamics and spectral constraints.

## II. SYSTEM MODEL

### A. Interference Scenario Configuration

An interference scenario between geostationary orbit (GSO) and low Earth orbit (LEO) satellite systems is considered, as conceptually illustrated in Fig. 1. The GSO satellite operates as the primary communication node, while multiple LEO satellites from non-geostationary systems introduce unintended interference components observed at a GSO ground station (GGS). The composite received waveform contains both desired carrier signals and interference.

### B. Link Budget Analysis

The received GSO carrier power is determined through classical satellite link relationships:

$$C = \frac{\text{EIRP}_{\text{gso}} \cdot G_{\text{r, gso}}(\theta_0)}{L_{\text{FS, gso}} \cdot L_{\text{add}}} \quad (1)$$

where  $G_{\text{r, gso}}(\theta_0)$  denotes the maximum receive antenna gain at boresight angle  $\theta_0$ ,  $L_{\text{FS, gso}}$  represents free-space path loss, and  $L_{\text{add}}$  accounts for aggregate atmospheric impairments.

Interference contributions from  $K$  LEO satellites are modeled as individual components:

$$I_k = \frac{\text{EIRP}_k \cdot G_{\text{r, k}}(\theta_k) \cdot B_{\text{adj, k}}}{L_{\text{FS, k}} \cdot L_{\text{add}}} \quad (2)$$

The angular gain term  $G_{\text{r, k}}(\theta_k)$  reflects spatial relationships caused by LEO orbital motion, while  $B_{\text{adj, k}} \in [0, 1]$  is the spectral overlap between GSO and LEO transmissions.

The composite signal quality is characterized through the carrier-to-interference-plus-noise ratio:

$$\text{CINR} = \frac{C}{\sum_{k=1}^K I_k + k_B T B} \quad (3)$$

where  $k_B$  denotes the Boltzmann constant,  $T$  the system noise temperature, and  $B$  the operational bandwidth.

### C. Signal Composition

The signal received by physical layer at the GGS has three components:

$$\begin{aligned} y(t) = & x(t) \sqrt{\text{CNR}} \text{ (Desired GSO)} \\ & + \sum_{k=1}^K i_k(t) e^{j2\pi \Delta f_k t} \sqrt{\text{INR}_k} \text{ (LEO interference)} \\ & + \zeta(t) \text{ (Thermal noise)} \end{aligned} \quad (4)$$

where  $\Delta f_k = f_{c,k} - f_{c,\text{gso}}$  captures carrier frequency offsets from Doppler effects and orbital dynamics. The exponential terms induce time-varying phase rotations proportional to relative satellite motion.

Dual signal representations are derived for machine learning processing:

- Time-domain:  $y^A$  captures instantaneous amplitude variations through uniform sampling
- Frequency-domain: Welch's power spectral density estimation generates logarithmic magnitude spectra via overlapping windowed transforms:  $y^F = 10 \log_{10}(\phi(y(t)))$

## III. PROPOSED DEEP LEARNING MODEL

We propose DualAttWaveNet, an autoencoder that takes both time and frequency domain signal as input, and try to reconstruct both of them. The model consists of separate encoder and decoder for both domains, and a fusion module that utilizes bidirectional attention before concatenating. The loss function is regularized with wavelet transform to enforce reconstruction fidelity across multiple scales.

### A. Bidirectional Attention

We propose a parameter-efficient mutual attention mechanism for cross-modal feature fusion. Unlike conventional multi-head attention in [10], our design employs single-head dot-product attention with spatial reduction to minimize computational overhead while maintaining inter-domain alignment capacity, as show in Fig. 2.

Given input feature maps  $\mathbf{X} \in \mathbb{R}^{B \times C \times L}$  (temporal domain) and  $\mathbf{Y} \in \mathbb{R}^{B \times C \times L}$  (spectral domain), the mutual attention operator computes:

$$\text{MutualAttn}(\mathbf{X}, \mathbf{Y}) = \mathbf{X} + \gamma \cdot \text{AttentionGate}(\mathbf{X}, \mathbf{Y})$$

$$\text{AttentionGate}(\mathbf{X}, \mathbf{Y}) = \mathbf{V}_y \cdot \text{Softmax} \left( \frac{\mathbf{Q}_x \mathbf{K}_y^\top}{\sqrt{d}} \right) \quad (5)$$

where  $\gamma$  is a learnable scalar initialized to 0,  $\mathbf{Q}_x = \mathcal{W}_Q(\mathbf{X}) \in \mathbb{R}^{B \times L \times \frac{C}{8}}$ ,  $\mathbf{K}_y = \mathcal{W}_K(\mathbf{Y}) \in \mathbb{R}^{B \times \frac{C}{8} \times L}$ , and  $\mathbf{V}_y = \mathcal{W}_V(\mathbf{Y}) \in \mathbb{R}^{B \times C \times L}$ . The projection matrices  $\mathcal{W}_{Q,K,V}$  implement 1D convolutions with kernel size 1, reducing channel dimensions by  $8\times$  for queries/keys to optimize memory footprint.

The compatibility scores between temporal ( $\mathbf{X}$ ) and spectral ( $\mathbf{Y}$ ) features are computed through matrix multiplication of

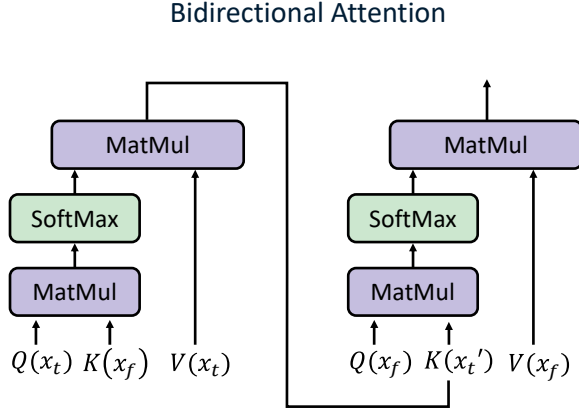


Fig. 2. DualAttnWaveNet’s bidirectional attention mechanism. Both attention phase share the same parameter  $Q, K$  and  $V$ . To decrease channel numbers, they are implemented as 1d-Convolution.

the reduced-channel representations. This produces an  $L \times L$  affinity matrix that represents position-wise cross-domain correlations before applying row-wise softmax normalization.

To preserve original feature stability during early training, the residual connection is initially dampened ( $\gamma = 0$ ) and progressively strengthened through learning. The symmetric architecture applies identical attention operations in both temporal $\rightarrow$ spectral and spectral $\rightarrow$ temporal directions, in sequential form:

$$\begin{aligned}\hat{\mathbf{X}} &= \text{MutualAttn}(\mathbf{X}, \mathbf{Y}) \\ \hat{\mathbf{Y}} &= \text{MutualAttn}(\mathbf{Y}, \mathbf{X})\end{aligned}\quad (6)$$

This bidirectional design shares parameters across both call of attention, enables joint refinement of both modalities without major computational overhead.

### B. Wavelet-Domain Spectral Regularization

Our temporal-spectral analysis framework employs parameterized Morlet wavelets with momentum-optimized scale distributions. Given an input tensor sequence  $\mathbf{x} \in \mathbb{R}^{B \times C \times L}$  where  $B, C$ , and  $L$  denote batch size, channels, and temporal length respectively, we construct a learnable filter bank  $\mathcal{F} \in \mathbb{R}^{S \times 1 \times K}$  through the following parameterization:

$$\mathcal{W}_s(\tau) = \text{Norm} \left( \cos \left( \frac{1.75}{s} \tau \right) \odot \mathcal{G}(\tau, s) \right) \quad (7)$$

where  $\mathcal{G}(\tau, s) = \exp(-\tau^2/(2s^2))$  represents the Gaussian envelope function,  $s \in \mathbb{S}$  enumerates the wavelet scale parameters, and  $K = 4s_{\max}$  determines the kernel size from the maximum scale  $s_{\max}$  to ensure sufficient spatial support. Each filter undergoes  $L_2$ -normalization to preserve energy consistency across multiple scales.

The discrete wavelet transform is implemented as asymmetric depth-wise separable convolution with spectral preservation properties:

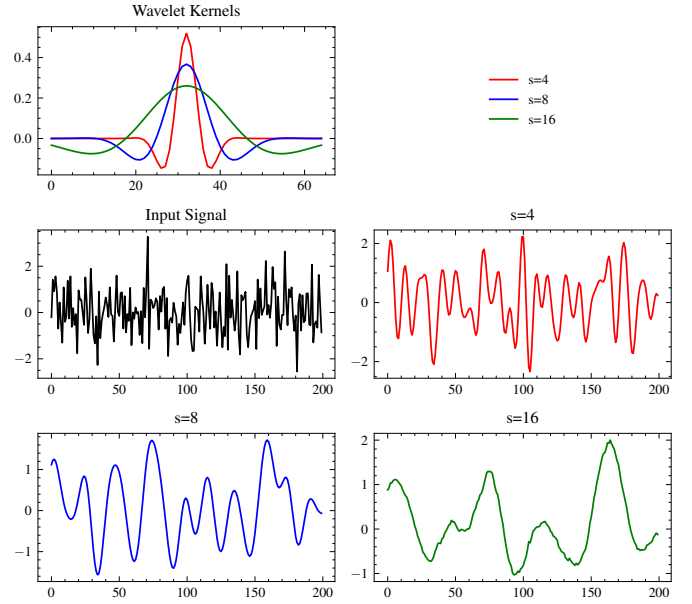


Fig. 3. (Top) Normalized Morlet-style wavelet kernels at scales [4, 8, 16]. (Bottom) Input–output relationships: Gaussian white noise input (top–left) and its filtered components using the respective wavelets (red:  $s = 4$ , blue:  $s = 8$ , green:  $s = 16$ ). Responses exhibit progressive high–frequency suppression and temporal smoothing with increasing scale. The visualizer implements bank normalization and preserves temporal resolution via same–padding convolutions (using a  $\beta$ -spline differentiable implementation).

$$\begin{aligned}\mathbf{X}_{\text{wave}} &= \text{Conv1D}(\mathbf{X}, \mathcal{F}) \\ &= \bigcup_{s \in \mathbb{S}} \mathbf{X} * \mathcal{W}_s \in \mathbb{R}^{B \times C \times S \times L}\end{aligned}\quad (8)$$

Reflection padding strategy mitigates boundary artifacts while maintaining temporal resolution through stride-1 convolution across  $S$  parallel scale dimensions. Fig. 3 illustrates the kernel morphology and signal transformation characteristics.

Our joint optimization framework combines both signal-space and wavelet-domain consistency through an adaptive loss weighting mechanism:

$$\mathcal{L} = \lambda_1 \|\hat{\mathbf{X}} - \mathbf{X}\|_{2, \Omega}^2 + \lambda_2 \sum_{s \in \mathbb{S}} \|\hat{\mathbf{X}}_{\text{wave}}^{(s)} - \mathbf{X}_{\text{wave}}^{(s)}\|_{\text{F}}^2 \quad (9)$$

where  $\lambda_1 = 1$  and  $\lambda_2 = 0.5$  represent empirically optimized weighting and  $\|\cdot\|_{\text{F}}$  denotes the norm over wavelet subbands. This dual-domain regularization ensures simultaneous preservation of temporal features and spectral distribution.

## IV. EXPERIMENTS

### A. Dataset Generation

The synthetic dataset is generated through a 48-hour MATLAB simulation sampling Ku-band (10.7-12.7 GHz) interference scenarios at 10-second intervals, producing 17,281 temporal snapshots following [5]. Each instance contains synchronized time-domain and frequency-domain representations: an 800-point waveform captures signal amplitudes, while an 800-bin spectral magnitude is derived via FFT processing.

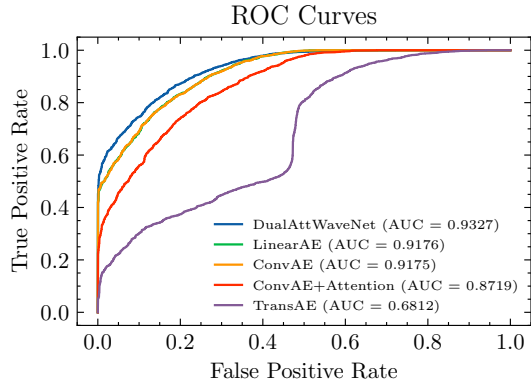


Fig. 4. ROC curves for DualAttWaveNet and baseline models. DualAttWaveNet achieves the highest AUC score.

Binary classification labels are assigned through link budget analysis, where class 0 denotes non-interference scenarios ( $\text{INR} < \Gamma_{\text{th}}$ ) below the system protection threshold, and class 1 indicates substantial interference ( $\text{INR} \geq \Gamma_{\text{th}}$ ) exceeding operational limits. We normalize the input signals in both domains separately to zero mean and unit variance.

The dataset is partitioned under anomaly detection constraints, with training (11,509 samples) and validation (1,302 samples) sets containing exclusively non-interference data (class 0). The test set comprises balanced proportions of 2,235 class 0 and 2,235 class 1 instances. The simulation incorporates time-varying link losses with 0-9 dB range, extreme interference cases reaching peak aggregate INR of 32.47 dB, with background CNR fluctuations between 6.40-15.40 dB.

### B. Evaluation Results

As summarized in Table I, DualAttWaveNet achieves the highest performance across all metrics. Regarding classification accuracy, the model reaches 83.51%, outperforming LinearAE (81.49%), CNNAE (81.70%), CNNAE+Attention (76.95%), and Transformer AE (68.12%). Its F1 score (0.8351) slightly surpasses TrID (0.8321), indicating a comparable balance between precision and recall. The most significant advantage lies in its AUC value (0.9327), which is 1.52 percentage points higher than LinearAE (0.9176) and CNNAE (0.9175), and 10.09 percentage points above TrID (0.8318). In computational efficiency, DualAttWaveNet requires only 5.19 GFLOPS, which is 6 times fewer operations than Transformer AE (26.43 GFLOPS) and 90% fewer than TrID (50 GFLOPS), despite TrID's slightly lower accuracy (83.18%). Transformer AE performs weakest overall, with both accuracy (68.12%) and F1 score (0.5921) falling to near-random levels.

The results validate three critical advantages of DualAttWaveNet. First, its cross-modal fusion efficacy is underscored by AUC dominance (0.9327), a 1.5% increase over CNNAE (0.9175) and a 10% leap over TrID (0.8318), directly attributable to dual-attention mechanisms that synchronize time-frequency features better than TrID's isolated domain processing. Second, precision-efficiency trade-offs are

TABLE I  
PERFORMANCE COMPARISON OF DUALATTWAVENET AGAINST  
BASELINE MODELS

Model	Accuracy (%) $\uparrow$	F1 Score $\uparrow$	AUC $\uparrow$	FLOPS (G)
DualAttWaveNet	0.8351	0.8351	0.9327	5.19
LinearAE	0.8149	0.8149	0.9176	0.49
CNNAE	0.8170	0.8169	0.9175	0.50
CNNAE+Attention	0.7695	0.7691	0.8718	3.01
TrID	0.8318	0.8321	0.8318	50.00
Transformer AE	0.6812	0.5921	0.6812	26.43

TABLE II  
ABLATION STUDY OF DUALATTWAVENET COMPONENTS

Model Variant	Accuracy (%) $\uparrow$	F1 Score $\uparrow$	AUC $\uparrow$
DualAttWaveNet (Full)	0.8351	0.8351	0.9327
w/o Mutual Attention	0.8289	0.8288	0.9294
w/o Wavelet Loss	0.8273	0.8273	0.9283
Vanilla Implementation	0.7995	0.7975	0.9175

uniquely optimized: while LinearAE and CNNAE require minimal compute (0.49–0.50 GFLOPS), they sacrifice 1.5–2.0% accuracy and 1.5% AUC relative to DualAttWaveNet. Conversely, TrID's 50 GFLOPS expenditure yields diminishing returns (83.18% accuracy vs. 83.51%). Third, the model's operational robustness manifests in stable F1 parity (0.8351) between accuracy and harmonic mean—avoiding the severe F1 degradation seen in Transformer AE (0.5921). ROC analysis in Fig. 4 corroborates these findings, demonstrating less overlap between interference/normal clusters compared to baselines.

For the confusion matrix analysis, the detection threshold was determined as the sum of the mean and standard deviation of the reconstruction error computed on the validation set. Reconstruction error over this threshold is considered to be in class 1. As shown in Fig. 5, DualAttWaveNet achieves a balanced precision-recall tradeoff (F1: 83.51%), outperforming all baselines. Notably, LinearAE and ConvAE exhibit identical precision (0.82) and recall (0.81) values, yielding comparable F1 scores (81.5%). While their performance is relatively stable.

DualAttWaveNet processes 5,000 batched samples (batch size: 128, 800 timesteps/sample) in under one second. While its FLOPs (5.19G) exceed those of lighter models (Table I), it significantly outperforms TrID (50.00G) and Transformer AE (26.43G) in computational efficiency. This is due it only uses one dot-product attention, instead of multi-head attention implemented in TrID and TransAE.

### C. Ablation Study

To validate the architectural design of DualAttWaveNet, systematic component ablation experiments were conducted (Table II). The full configuration achieves peak performance, while removing the mutual attention module reduces AUC to 0.9294 (-0.33%), demonstrating its necessity for temporal-spectral feature alignment. Further ablation of the wavelet coherence loss degrades performance to 0.9283 AUC (-0.44%

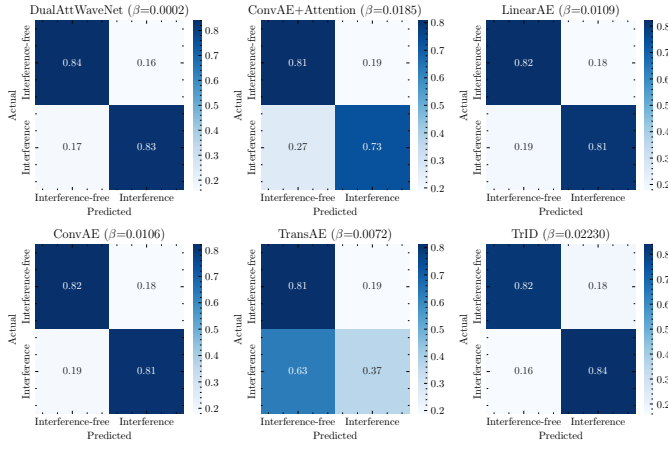


Fig. 5. Confusion matrix for DualAttnWaveNet and baselines on the test set

from full model), confirming the loss's role in preserving time-frequency structure. The baseline implementation (no mutual attention or wavelet loss) yields the lowest metrics (AUC: 0.9175, Accuracy: 79.95%), underperforming all proposed variants.

## V. CONCLUSION

This paper introduces DualAttnWaveNet, a multimodal interference detection framework tailored for GSO/LEO co-existence scenarios. By integrating time-domain waveforms and spectral representations through a bidirectional attention mechanism, the model achieves effective cross-domain feature alignment while minimizing computational overhead. Multi-scale wavelet regularization is employed to enforce reconstruction error across temporal and spectral resolutions. Evaluation in synthesized Ku-band interference scenarios demonstrated strong performance, with the model attaining an area under the curve (AUC) of 0.9327 and an accuracy of 83.51%, outperforming state-of-the-art baselines. Testing on a dataset of 5,000 samples yielded an average latency of 97.2 milliseconds per sample when deployed on a single NVIDIA 3050Ti GPU. Subsequent research will examine the model's performance across varying overlap conditions and investigate alternative wavelet families to refine regularization strategies.

## REFERENCES

- [1] *Radio Regulations*, International Telecommunication Union (ITU) Std., 2020.
- [2] *Protection Criteria Related to the Operation of Data Relay Satellite Systems*, International Telecommunication Union (ITU) Recommendation ITU-R SA.1155-2, 2017.
- [3] *Methodologies for Calculating Aggregate Downlink Equivalent Power Flux-Density Produced by Multiple Non-Geostationary Fixed-Satellite Service Systems into a Geostationary Fixed-Satellite Service Network*, International Telecommunication Union (ITU-R) Report ITU-R S.1588-0, 2002.
- [4] S. M. Kay, *Fundamentals of Statistical Processing, Volume 2: Detection Theory*. Pearson Education, 2009.
- [5] A. Saifaldawla, F. Ortiz, E. Lagunas, A. B. M. Adam, and S. Chatzinotas, "Genai-based models for ngso satellites interference detection," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 904–924, 2024.

- [6] F. Dimc, G. Baldini, and S. Kandeepan, "Experimental detection of mobile satellite transmissions with cyclostationary features," *International Journal of Satellite Communications and Networking*, vol. 33, no. 2, pp. 163–183, 2015.
- [7] T. Wang, W. Li, and Y. Li, "Co-frequency interference analysis between large-scale ngso constellations and gso systems," in *Proc. 2020 International Conference on Wireless Communications and Signal Processing (WCSP)*, 2020, pp. 679–684.
- [8] L. Pellaco, N. Singh, and J. Jaldén, "Spectrum prediction and interference detection for satellite communications," in *Proc. 37th International Communications Satellite Systems Conference (ICSSC-2019)*, 2019, pp. 1–18.
- [9] A. Saifaldawla, F. Ortiz, E. Lagunas, and S. Chatzinotas, "Convolutional autoencoders for non-geostationary satellite interference detection," in *Proc. 2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, Denver, CO, USA, 2024, pp. 1334–1339.
- [10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. ukasz Kaiser, and I. Polosukhin, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.