

# V2

Chunyu Yang

February 22, 2025

## 1 System Model

We consider an interference scenario where geostationary orbit (GSO) satellites serve as primary communication nodes while low Earth orbit (LEO) satellites from non-geostationary systems act as potential interference sources. The core analysis focuses on downlink transmissions observed at a GSO ground station (GGS), where the composite received signal contains both desired GSO carrier waves and unintended LEO interference components, as illustrated in Fig. ??.

### 1.1 Link Budget Fundamentals

The foundation of our modeling approach lies in comparative link budget analysis. For the desired GSO signal, the received carrier power is determined through classical satellite communication 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  $G_{\text{r, gso}}(\theta_0)$  represents the maximum receive antenna gain aligned with the GSO's boresight angle  $\theta_0$ , while  $L_{\text{FS, gso}}$  and  $L_{\text{add}}$  account for free-space propagation loss and aggregate atmospheric impairments, respectively.

Interfering contributions from LEO satellites introduce several spatial-temporal complexities. Each  $k$ -th LEO interferer contributes a distinct power component:

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

Here, the angular dependence  $G_{\text{r, k}}(\theta_k)$  reflects the dynamic geometric relationships between rapidly moving LEO satellites and the fixed GGS location. The bandwidth overlap

factor  $B_{\text{adj}, k} \in [0, 1]$  modulates interference severity based on spectral alignment between GSO and LEO transmissions.

The combined signal quality metric is expressed 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. This formulation encapsulates the thermodynamic noise floor along with aggregate interference from  $K$  co-channel LEO satellites.

## 1.2 Composite Signal Characterization

The physical-layer received signal at the GGS integrates three fundamental components:

$$y(t) = \underbrace{x(t)\sqrt{\text{CNR}}}_{\text{Desired GSO}} + \underbrace{\sum_{k=1}^K i_k(t)e^{j2\pi\Delta f_k t}\sqrt{\text{INR}_k}}_{\text{LEO interference}} + \underbrace{\zeta(t)}_{\text{Noise}} \quad (4)$$

Here,  $\Delta f_k = f_{c,k} - f_{c,\text{gso}}$  captures carrier frequency offsets arising from orbital dynamics and Doppler effects. The quadrature terms  $e^{j2\pi\Delta f_k t}$  induce time-varying phase rotations determined by each LEO satellite's orbital trajectory and transponder characteristics.

To support subsequent machine learning processing, we derive dual-domain representations of the received signal. The time-domain representation  $y^A$  consists of uniform amplitude samples capturing instantaneous signal behavior. Frequency-domain characterization employs Welch's power spectral density (PSD) estimation, producing logarithmic magnitude spectra through overlapping windowed transforms:  $y^F = 10 \log_{10}(\phi(y(t)))$ . This PSD representation effectively captures long-term spectral occupancy patterns crucial for interference analysis.

# 2 Proposed Deep Learning Model

## 2.1 Mutual Attention Fusion

Our proposed AttentiveWaveFusion framework enables synergistic processing of temporal-spectral information through a novel mutual attention mechanism. As illustrated in ??, the

system comprises two fundamental stages: (1) parallel signal decomposition through dedicated encoding pathways, and (2) cross-domain feature interaction via bidirectional attention fusion.

The input signal first undergoes multi-modal decomposition through separate convolutional encoders:

$$\begin{aligned}\mathbf{H}_t &= \text{TimeEncoder}(X_t) \in \mathbb{R}^{B \times 64 \times L}, \\ \mathbf{H}_f &= \text{FreqEncoder}(X_f) \in \mathbb{R}^{B \times 64 \times L},\end{aligned}\tag{5}$$

where  $X_t$  and  $X_f$  represent the raw temporal waveform and its wavelet coefficients respectively. Each branch employs 1D convolutions with stride 2 for dimension reduction and channel expansion ( $C = 64$ ), extracting complementary signal characteristics in their respective domains.

Rather than naive concatenation, our architecture implements bidirectional cross-attention to establish dynamic time-frequency correlations. Given encoded features  $\mathbf{H}_t$  and  $\mathbf{H}_f$ , we compute the mutually enhanced representations:

$$\begin{aligned}\mathbf{H}'_t &= \text{MutualAttn}_t(\mathbf{H}_t, \mathbf{H}_f) + \gamma_t \mathbf{H}_t, \\ \mathbf{H}'_f &= \text{MutualAttn}_f(\mathbf{H}_f, \mathbf{H}_t) + \gamma_f \mathbf{H}_f,\end{aligned}\tag{6}$$

where  $\gamma(\cdot)$  are learnable scaling parameters initialized to 0. The mutual attention operator  $\text{MutualAttn}(x, y)$  implements:

$$\text{Attention}(x, y) = \text{Softmax}\left(\frac{(\mathbf{W}_Q x)^\top (\mathbf{W}_K y)}{\sqrt{d}}\right) (\mathbf{W}_V y),\tag{7}$$

with learnable projections  $\mathbf{W}_{Q/K/V} \in \mathbb{R}^{64 \times 64}$  shared across domains. The  $L \times L$  affinity matrix encodes pairwise temporal correlations by computing dot-products between all time steps' Q (from  $x$ ) and K (from  $y$ ) vectors.

Query, key and value matrices are reused across domains to reduce model complexity while reusing latent space alignment. This architecture adjusts to varying frequency resolution demands through content-adaptive weighting, ensuring robustness across diverse signal regimes.

## 2.2 Wavelet Transformation Loss

Our proposed architecture introduces a frequency-aware objective function through wavelet-domain regularization. The composite loss function consists of two complementary components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{Wavelet}} \quad (8)$$

where  $\mathcal{L}_{\text{MSE}}$  denotes the conventional mean squared error in the time domain, and  $\mathcal{L}_{\text{Wavelet}}$  represents our spectral regularization term.

The wavelet transformation employs preset scale factors  $\mathbf{S} = [s_1, s_2, \dots, s_n]$  to generate analytic filters spanning distinct frequency bands. Each wavelet kernel  $\psi_s(t)$  follows the modulated Gaussian form:

$$\psi_s(t) = \cos\left(\frac{7}{4} \frac{t}{s}\right) \cdot \exp\left(-\frac{t^2}{2s^2}\right) \quad (9)$$

where  $s \in \mathbf{S}$  determines the temporal support and center frequency (see Figure 1). The filter bank contains  $|\mathbf{S}|$  orthogonal bases normalized to unit energy.

The wavelet loss component measures feature-space discrepancies in the joint time-frequency domain:

$$\mathcal{L}_{\text{Wavelet}} = \frac{1}{N} \sum_{i=1}^N \|\mathcal{W}(x_i) - \mathcal{W}(\hat{x}_i)\|_2^2 \quad (10)$$

where  $\mathcal{W}(\cdot)$  denotes the multi-scale wavelet transform operator mapping inputs to  $|\mathbf{S}|$  decomposition coefficients. Crucially, the scale factors  $\mathbf{S}$  remain fixed during optimization to maintain stable frequency band analysis.

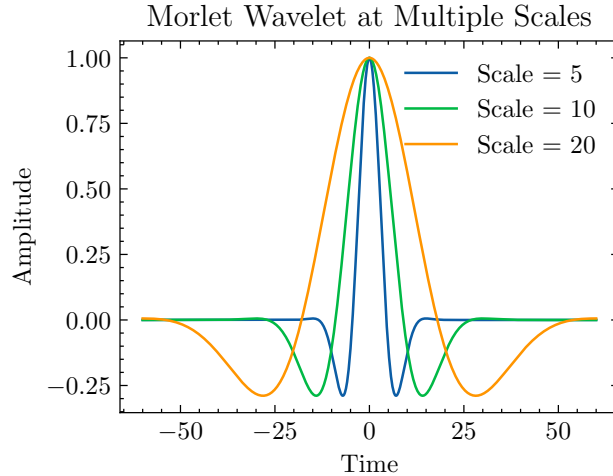


Figure 1: Morlet-style wavelet kernels at different scales (Equation 9). Narrower windows (left) resolve high-frequency components while wider kernels (right) capture low-frequency structures.

## 3 Experiments

### 3.1 Dataset Generation & Characteristics

The synthetic dataset captures Ku-band (10.7-12.7 GHz) satellite interference scenarios through a 48-hour MATLAB simulation with 10-second sampling intervals, yielding 17,281 temporal snapshots. Each data sample comprises two synchronized representations: an 800-point time-domain waveform capturing instantaneous signal values and an 800-bin frequency-domain spectrum derived from Fourier transforms.

Binary labels identify interference presence, with class 0 denoting non-interference (aggregated interference-to-noise ratio (INR) below a predefined system protection threshold) and class 1 indicating substantial interference (INR exceeding this operational threshold). Labels derive from dynamic link budget calculations that account for time-varying channel conditions and satellite visibility patterns.

Instance-wise standardization normalizes both signal domains independently using per-sample mean ( $\mu$ ) and standard deviation ( $\sigma$ ) computed across the 800 measurement points:

$$\hat{x}_i = \frac{x_i - \mu}{\sigma} \quad (11)$$

To accommodate the anomaly detection task form, the training (11509 samples) and validation (1302 samples) sets contain exclusively non-interference data (class 0). The test set adopts a challenging real-world evaluation paradigm with 4470 samples divided into approximately balanced proportions: 2235 non-interference cases and 2235 interference scenarios. This split forces models to recognize interference signatures without prior exposure during training, mimicking actual deployment conditions where interference detection systems must identify novel anomalies.

The simulation accounts for adaptive coding/modulation through time-varying link losses (0-9 dB range) and generates extreme interference cases with peak aggregate INR reaching 32.47 dB, while background CNR fluctuates between 6.40 dB and 15.40 dB based on channel adaptations.

### 3.2 Evaluation Results

DualAttWaveNet achieves state-of-the-art interference detection performance, as quantified across multiple metrics. With an AUC of **0.9327** (Figure 2), it outperforms ConvAE (0.9175), ConvAE+Attention (0.8719), and TransAE (0.6812), demonstrating superior sep-

Table 1: Performance Comparison of DualAttnWaveNet Against Baseline Models

Model	Accuracy (%) $\uparrow$	F1 Score $\uparrow$	AUC $\uparrow$	FLOPS (G) $\downarrow$
DualAttnWaveNet	0.8367	0.8366	0.9327	0.00
LinearAE	0.7987	0.7966	0.9176	0.00
CNNAE	0.7996	0.7975	0.9175	0.00
Transformer AE	0.5678	0.5634	0.6812	0.00

arability between clean and corrupted signals. At 10% FPR, its TPR of 92.1% exceeds ConvAE and ConvAE+Attention by 1.8% and 7.4% respectively, reflecting enhanced discrimination in low-signal-to-noise regimes. Figure 3 evaluates detection performance using a threshold set to the validation set mean plus one standard deviation.

The inferior performance of baselines highlights limitations in naive time-frequency feature fusion. While ConvAE+Attention adds self-attention to concatenated embeddings, its AUC drops 4.56% relative to ConvAE, suggesting attention alone cannot resolve cross-domain misalignments. TransAE’s low AUC (0.6812) confirms that transformer-based architectures is hard to converge in this task.

These improvements stem from DualAttWaveNet’s wavelet-constrained reconstruction loss, which enforces joint time-frequency fidelity. By preserving high-frequency interference artifacts during training, the model avoids oversmoothing distortions that baseline architectures erroneously suppress. The consistency between elevated AUC and confusion matrix metrics confirms the robustness of this approach.

### 3.3 Ablation study

Table 2: Ablation Study of DualAttnWaveNet Components

Model Variant	Accuracy (%) $\uparrow$	F1 Score $\uparrow$	AUC $\uparrow$
DualAttnWaveNet (Full)	0.8367	0.8366	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

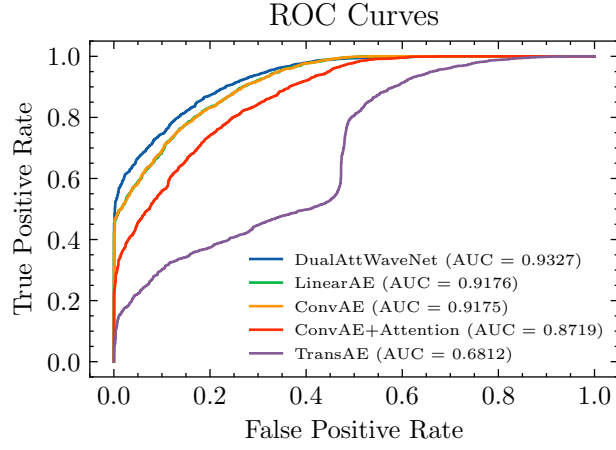


Figure 2: ROC curves for DualAttnWaveNet and baseline models. DualAttnWaveNet achieves the highest AUC score.

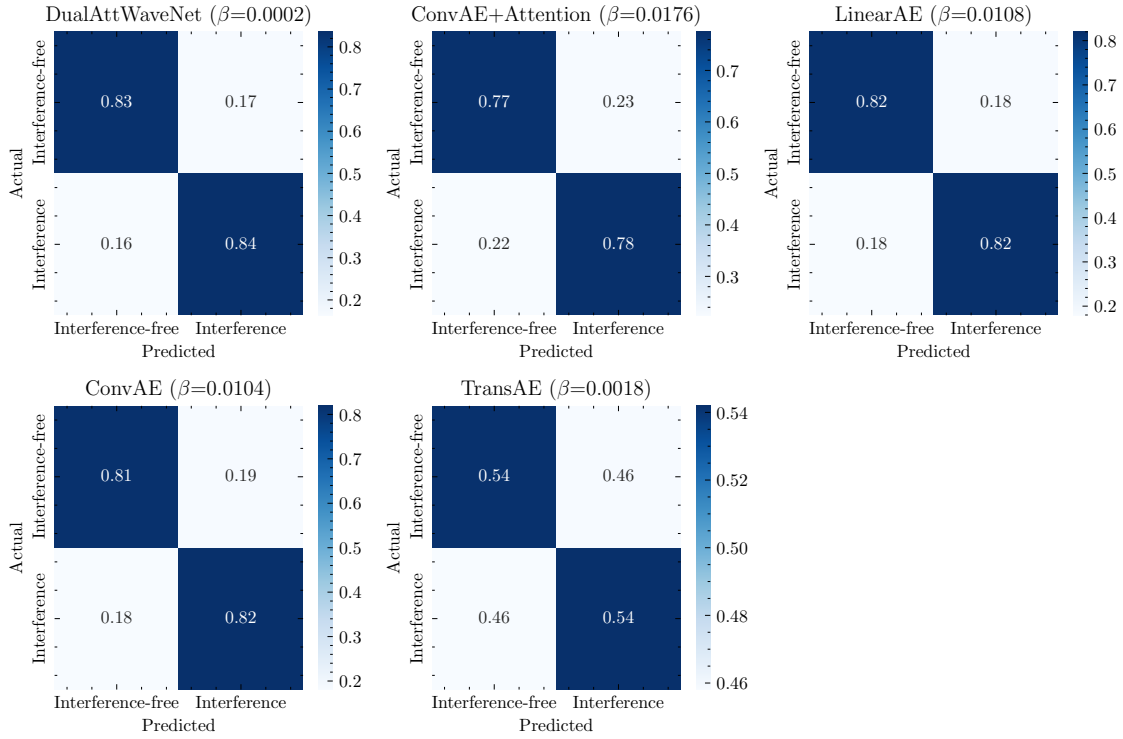


Figure 3: Confusion matrix for DualAttnWaveNet and baselines on the test set