

Article

DL-MFFSSnet: A Multi-Feature Fusion-Based Dynamic Collaborative Spectrum Sensing Method in a Satellite–Terrestrial Converged System

Chao Tang ¹, Yueyun Chen ^{1,2,*}, Guang Chen ^{1,2}, Liping Du ¹, Zhen Wang ¹ and Huan Liu ¹

¹ School of Computer and Communication Engineering, University of Science and Technology Beijing (USTB), Beijing 100083, China

² Shunde Innovation School, University of Science and Technology Beijing (USTB), Foshan 528000, China

* Correspondence: chenyy@ustb.edu.cn

Abstract

Satellite–terrestrial spectrum sensing plays a crucial role in enhancing spectrum efficiency through reusing spectra. However, in a satellite–terrestrial converged system, the large SNR range, non-Gaussian signal characteristics and noise uncertainty pose significant challenges for spectrum sensing. In this paper, we investigate a downlink spectrum sensing framework where multi-terrestrial BSs act as a secondary system to sense idle satellite spectra through a multi-domain feature-level sensing signal fusion. To enhance the characterization of signal/noise features, we provide a fusion strategy of multi-features including energy, power spectral density, cyclic autocorrelation function, higher-order moments, sparse ratio, and I/Q samples, constructing two feature tensors of statistical features and an I/Q component. Then, we propose a deep-learning-enabled multi-feature fusion spectrum sensing method (DL-MFFSSnet) based on a dual-branch deep neural network architecture with the constructed two feature tensors as inputs. In the statistical feature processing branch, CNN and channel self-attention are incorporated to capture intra-channel correlations and inter-channel relative contributions of different feature modalities. In the I/Q branch, multi-scale dilated convolutions and spatial self-attention are introduced to analyze dependencies across different temporal positions and multi-scale spatial features. The feature map extracted from both branches passed through fully connected layers for deepwise feature fusion, achieving accurate spectrum sensing. Extensive simulation results demonstrate that the DL-MFFSSnet method outperforms the existing state-of-the-art algorithms.

Keywords: satellite–terrestrial converged system; spectrum sensing; spectrum detection probability; multi-scale dilated convolution; channel/spatial self-attention; deep learning



Academic Editor: Silvia Liberata Ullo

Received: 15 January 2026

Revised: 16 February 2026

Accepted: 19 February 2026

Published: 23 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Motivation

With the rapid development of terrestrial and satellite networks, the convergence of satellite and terrestrial systems has become an inevitable development trend and will usher in a new upsurge for wireless communications [1]. In satellite–terrestrial converged networks (STCNs), efforts are being exerted to achieve high-throughput broadband communication, wide-area massive smart connectivity, and global ubiquitous seamless access [2]. However, under the ambitious vision, the explosive growth of wireless devices and services has drastically increased spectrum demand, further exacerbating the scarcity of spectrum resources [3].

To alleviate the spectrum constraints limiting the development of STCNs, there is an urgent need to expand or reuse available frequency bands [4]. Satellite–terrestrial dynamic spectrum sharing is regarded as a promising strategy. It allows satellite and terrestrial systems access to the same spectrum, dynamically adjusting their usage based on the real-time conditions and demands [5]. Specifically, through effective spectrum sensing and interference management, the satellite–terrestrial system facilitates the reuse of existing frequency bands, such as leveraging underutilized satellite spectra in urban areas and terrestrial spectra in rural regions, thereby improving spectral efficiency [6].

The effectiveness of spectrum sharing ultimately depends on the ability to accurately and efficiently identify available spectra in real time [7]. Spectrum sensing serves as a key enabler in satellite–terrestrial spectrum sharing by providing timely awareness of spectrum occupancy, assisting secondary systems in adopting overlay or underlay strategies to share spectra [8]. Classical spectrum sensing methods, including energy detection [9], matched filtering [10], cyclostationary analysis [11], etc., are valued for their analytical tractability and low complexity. However, these methods assume favorable conditions, such as known noise power, stationary channels, and homogeneous interference. In STCNs, the channel variability and noise uncertainty disrupts the above-mentioned assumptions, leading to significant performance degradation and limiting practical applicability [12].

To overcome the limitations of traditional spectrum sensing methods, deep learning (DL)-based spectrum sensing has made rapid progress in recent years [13]. Capitalizing on the powerful representation capability of DNN models, DL spectrum sensing methods implicitly learn the geometric correlation of the signal in the frequency/time/temporal domain through sensing signal features, including energy [14], power spectral density (PSD) [15], short-time fourier transform (STFT)-based time–frequency features [16], covariance matrices (CMs) [17], in-phase/quadrature (I/Q) data [18], etc., as DNN model inputs. To achieve efficient spectrum sensing under a dynamic environment with noise uncertainty, in our previous work, we have proposed a generalized likelihood ratio (GLR)-based spectrum detection method [19], and developed a Deep-CRnet detection framework based on deep learning [20], which used I/Q sampling data as input, providing a significant foundation for dynamic spectrum sensing in STCNs. Although proposed DL-based spectrum sensing methods have demonstrated improved detection performance under noise and signal uncertainty conditions, they are still challenged by the unique constraints posed by the characteristics of STCNs.

In particular, achieving both low false alarm probability and high detection probability in STCNs presents significantly more stringent challenges than in cellular networks [21]. (i) Highly dynamic sensing environment: the rapid movement of LEO satellites affects communication distances and introduces significant Doppler shift, atmospheric fading and intermittent shadowing, leading to dynamic changes of the sensing and sharing environment. (ii) Multi-modulation signal and noise uncertainty: Highly variable satellite downlink conditions result in the coexistence of multiple modulation types and uncertain noise characteristics, making accurate spectrum sensing more challenging. (iii) Low SNR and non-Gaussian signal characteristics: High propagation losses, shadowing and atmospheric fading lead to low SNR conditions and large variations in SNR over time, causing the received signals to deviate from the Gaussian assumption. Therefore, designing robust and intelligent spectrum sensing methods is particularly important to reduce missed and false detection that result in low spectrum utilization and harmful co-channel interference, particularly in STCNs with changing parameters of different signal modulation types, low SNR at the receiver, varying noise density, and no satellite state assumptions at run-time.

1.2. Related Works

In STCNs, spectrum sensing is crucial to identifying available idle spectra and improving spectral efficiency, which has attracted intense attention from both academia and industry. The spectrum sensing strategies are mainly divided into statistical-model-based traditional methods and DL/ML-based intelligent spectrum sensing methods.

Traditional spectrum sensing methods are mainly based on statistical decision theory, including energy detection (ED) [9], matched filter detection (MFD) [10], and cyclostationary feature detection (CFD) [11]. Among them, ED is widely favored for its simplicity and independence from prior signal knowledge, but the performance is significantly degraded under noise uncertainty and low SNR conditions. The MFD-based detection method, while optimal when full knowledge of the signal and channel is available, suffers from poor generalization in practice due to the unavailability of such information. The CFD method exploits the periodic statistical properties of modulated signals, providing resilience against noise power uncertainty, but the performance is limited by the requirement for cyclic features and accurate synchronization.

To overcome the limitations, improved statistical detection schemes have been proposed. In [14], the authors introduced a multi-slot cooperative ED approach to mitigate the effects of time-varying channels, and [22] utilized phased-array antennas to enhance ED-based detection accuracy in satellite scenarios. To address the low detection accuracy caused by channel dynamics and noise uncertainty, ref. [19] employed generalized likelihood ratio testing (GLRT) based on noise and channel estimation to enhance detection accuracy, and [23] introduced higher-order moments to handle noise uncertainty. Furthermore, in [24], the authors proposed a robust multi-stage detector combining energy, a singular value, and autocorrelation features for both high/low SNR environments. Although the aforementioned methods incorporate targeted improvements over traditional methods, they still exhibit limited adaptability in non-stationary environments, constrained scalability, and a heavy reliance on single handcrafted statistical features, leading to suboptimal performance.

Deep learning-based spectrum sensing has emerged as another powerful solution for improving detection accuracy in complex STCNs, particularly by eliminating reliance on prior signal knowledge [25]. Recent studies have focused on combining spectrum sensing with deep learning to explore and identify multi-dimensional characteristics of signals in the time/frequency/spatial/power domain, improving sensing performance while retaining interpretability and low computational complexity [15–18,26,27].

Considering the frequency characteristics of the signal, power spectral density (PSD) is an important feature. Ref. [15] constructed a multi-band PSD vector and introduced a self-attention module based on the DL framework to learn cross-band and intra-band spectral correlations, thereby achieving spectrum awareness. To make full use of historical spectrum occupancy information, ref. [26] adopted a long short-term memory (LSTM) model to achieve the sensing and prediction of future spectrum occupancy using both current and historical PSD data. Incorporating the temporal continuity of signals into frequency-domain feature analysis can enhance the accuracy of signal characterization. Ref. [16] proposed the STFT-CNN method, which utilized the STFT method to construct time–frequency feature maps for convolutional neural network (CNN)-based spectrum detection, and [28] adopted a sliding window to obtain segmented power spectral density sequences, and leveraged transformer networks to extract multidimensional features for enhanced sensing accuracy. However, PSD-based models are often sensitive to spectral leakage and resolution limitations, and may lose critical spatial-domain characteristics of the signal under fast-changing environments.

In multi-user or multi-antenna systems, the spatial correlation of signals can further enhance spectrum sensing accuracy. Covariance matrix (CM)-based methods provide robust feature extraction for spectrum sensing by capturing second-order statistical characteristics. In [17], the authors utilized user multi-antenna reception perception signal sequences to construct a covariance matrix and proposed a sensing method based on multi-band CNN networks. To make full use of historical spectrum occupancy information, refs. [29,30] proposed dual-branch neural networks of CNN and CNN-LSTM to process present sensing data and historical sensing data. To further study the interpretability of CNN-based networks, ref. [31] formulated the CM-CNN model as a non-linear function, and theoretically proved that the test statistic of the CM-CNN method is equivalent to the optimal estimator–correlator test statistic when the signal samples are independent and identically distributed. These CM-based spectrum detection methods exhibit strong noise robustness and typically assume that the received signals follow a complex Gaussian distribution, an assumption often violated due to non-Gaussian characteristics introduced by shadowed fading, line-of-sight fluctuations, and multi-interference effects. Moreover, CM-based approaches mainly rely on second-order statistics and may lose fine-grained temporal structures and higher-order signal characteristics, which are particularly important in low-SNR and highly dynamic satellite–terrestrial environments. Their representation capability is therefore constrained when dealing with diverse modulation schemes and rapidly time-varying channel conditions.

In contrast, an increasing number of studies directly utilizes raw I/Q sampling data as deep learning inputs, aiming to fully exploit the temporal and structural characteristics of the original signal without relying on handcrafted feature extraction. Ref. [18] combined CNNs and transformers to capture spatial–temporal patterns from I/Q data, while [27,32] proposed hybrid CNN-LSTM models that incorporate channel attention and cooperative fusion strategies across users with I/Q data. These I/Q-based approaches could capture fine-grained, real-time signal temporal and spectral characterization; however, they typically process a single-domain representation and focus primarily on local temporal dependencies. They often overlook complementary statistical-domain features such as sparsity, cyclic structure, and higher-order moments, which are critical for robust detection under a non-Gaussian signal, noise uncertainty and shadowed fading. In addition, directly learning from raw I/Q samples may require large-scale data to generalize well across wide SNR ranges and diverse satellite channel conditions.

In summary, while traditional statistical methods and deep learning approaches based on PSD, CM and I/Q features have made significant progress in spectrum sensing, they are often limited by their reliance on single-domain representations, without explicitly modeling the complementary relationships between statistical-domain features and waveform-domain representations. Such single-modality designs may limit robustness and generalization capability in complex satellite–terrestrial converged systems characterized by dynamic channels and heterogeneous signal conditions. Therefore, further research is needed to develop intelligent spectrum sensing methods that integrate complementary multi-domain features, enabling joint exploitation of statistical-domain and waveform-domain representations and systematic modeling of temporal, spatial, and cross-feature dependencies to enhance robustness and detection performance in dynamic satellite–terrestrial environments characterized by diverse modulation types, low SNR, non-Gaussian signal characteristics, noise uncertainty, and short sensing durations.

1.3. Contribution

In this paper, we investigate a collaborative satellite–terrestrial downlink spectrum sensing architecture that enables terrestrial systems to cooperatively perform the accurate

sensing and detection of idle licensed LEO satellite spectra, assisting terrestrial users to opportunistically access satellite spectra to alleviate spectrum scarcity. To enable efficient and reliable spectrum sensing in dynamic environments, we propose a DL-enabled spectrum sensing method, DL-MFFSSnet, which employs a modality-aware dual-branch neural network to jointly exploit complementary statistical-domain and waveform-domain representations collected from multiple sensing nodes, thereby improving spectrum sensing accuracy and robustness. The main contributions are as follows.

- We propose a collaborative downlink spectrum sensing framework in a satellite–terrestrial converged system, where multiple terrestrial BSs act as secondary systems to sense and detect idle spectra of licensed LEO satellites through multi-domain feature-level sensing signal fusion, assisting terrestrial users to opportunistically access satellite spectra and improving spectrum efficiency.
- To tackle the challenge of signal feature submergence in low SNR conditions at terrestrial sensing nodes caused by time-varying satellite–terrestrial channels, noise uncertainty, and multiple fading effects in STCNs, based on multi-perspective signal characteristics, we provide a fusion strategy of multi-feature including energy, power, higher-order moments, PSD, CAF, sparse ratio, as well as I/Q samples of sensing signal, constructing two feature tensors of statistical features and an I/Q component, thereby enhancing the representational capability of satellite signals.
- To further improve spectrum detection probability in dynamic and complex STCNs characterized by low SNR, diverse signal modulation types, non-Gaussian signal characteristics and noise uncertainty, based on the two constructed tensors of statistical features and the I/Q component, we propose a DL-enabled multi-feature fusion spectrum sensing method, named DL-MFFSSnet, which adopts a dual-branch deep neural network architecture to process the two-perspective features in parallel.
- In the statistical feature tensor processing branch, CNN and channel self-attention are incorporated to capture the intra-channel correlations and inter-channel relative contributions of different feature modalities. Meanwhile, in the I/Q tensor processing branch, multi-scale dilated convolutions and spatial self-attention mechanisms are introduced to analyze dependencies across different temporal positions and multi-scale spatial features. In addition, residual networks are incorporated into both branches to mitigate the vanishing gradient problem and to promote the reuse of shallow and deep features.
- Extensive simulation results demonstrate that DL-MFFSSnet achieves superior detection performance compared to state-of-the-art spectrum sensing methods in low SNR STCNs. At an SNR of -24 dB, the DL-MFFSSnet is 14.3% higher than the APASS strategy [29], 21.1% higher than Deep-CRnet [20] and 37.0% higher than the Deep STFT-CNN model [16] in spectrum detection probability, and when the SNR is -15 dB, the improvements are 4.7%, 8.5%, and 23.2%, respectively.

The remainder of this paper is organized as follows. Section 2 details the system model and problem formulation. A dynamic intelligent spectrum sensing method, DL-MFFSSnet, is presented in Section 3. Simulation results and discussions are provided in Section 6, followed by conclusions in Section 7.

2. System Model

As shown in Figure 1, we consider a downlink spectrum sensing and sharing scenario between LEO satellites and multi-terrestrial cellular base stations. Due to the demand for massive high-rate terminal connections, the spectrum resources of the terrestrial system are insufficient to support all-device access. To address the limitation of scarce spectrum resources, the terrestrial system, functioning as a secondary system, aims to sense and

detect unoccupied subbands of the LEO satellite system, allowing terrestrial users to opportunistically utilize the idle spectrum of the satellite system for communication.

In STCNs, the LEO satellite sa serves multiple satellite users ($u_{sa} \in \mathcal{U}_{sa}$) and a non-terrestrial network station e during the LEO visibility time. The large beam radius of the LEO satellite covers multiple terrestrial cellular base stations ($u_{te} \in \mathcal{U}_{te}$), which adopts a hexagonal structure to achieve seamless coverage. The sharing bandwidth B of the satellite system contains K orthogonal subbands, and the corresponding bandwidth of subband k is B_k . We assume that the number of satellite users is less than the number of subbands, supporting satellite spectrum reuse. The terrestrial base station, called the sensing node, performs feature-level cooperative spectrum sensing to tackle the challenges associated with the multiple modulation types of the signal, noise uncertainty and low SNR at the receiver. In order to realize information interaction, the different terrestrial stations use an X_2 interface [33], and each terrestrial sensing node shares and feeds back signal feature and detection results.

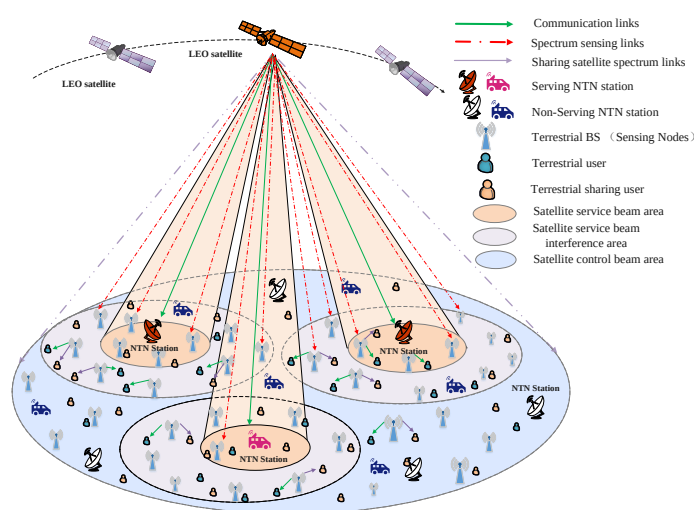


Figure 1. Cognitive-radio-enabled STCN scenario.

To better align with real-world LEO satellite downlink communication scenarios, according to ITU [34] and 3GPP standards [35], the satellite LOS channel models of K -factor NTN-TDL-C are considered, which combine path loss, shadow fading and Rician fading with Doppler shift and multi-path effects. Therefore, the channel model of LEO sa to satellite users u_{sa} or to the terrestrial sensing node u_{te} with the variable of time t and delay τ is $h_{sa,i}(t, \tau) = \sqrt{L_{sa,i}} g_{sa,i}(t, \tau)$, $i \in \{u_{sa}, u_{te}\}$, where $L_{sa,i}(t) = 10^{-PL_{total}/10}$ denotes the large-scale fading gain. The total large-scale path loss is given by $PL_{total} = PL_b + PL_g + PL_s$, with basic path loss $PL_b = FSPL(f, d) + CL(f, \varphi) + SF$ (including free-space path loss $FSPL(f, d)$, clutter loss $CL(f, \varphi)$, and shadow fading $SF \sim \mathcal{N}(0, \sigma_{SF}^2)$), and atmospheric gas attenuation PL_g and cloud/fog attenuation PL_s [34]. The small-scale impulse response is $g_{sa,i}(t, \tau) = \sum_{l=1}^L a_l(t) \cdot e^{j2\pi f_d t} \cdot \delta(\tau - \tau_l)$, where $a_l(t)$ is the K -factor, L is the total number of paths, and τ_l are tapped delays.

To characterize the variation in signal propagation distance $d_{sa,i}(t)$ and the resulting Doppler shift $f_d(t)$ caused by the movement of LEO satellites in time t , we define $\psi(t)$ to represent the angle between the Earth's center and the terrestrial sensing node, and the Earth's center and the LEO satellite in time t , which is varying with LEO satellite motion. Let $\theta(t)$ denote angle between the direction of v and the satellite downlink, and $\varphi(t)$ is the satellite elevation angle. The relationships between $\varphi(t)$, $\theta(t)$ and $\psi(t)$ are [36]:

$$\varphi(t) = \arctan \left[(\cos \psi(t) - 1/\lambda) / \sqrt{1 - \cos^2 \psi(t)} \right], \quad (1)$$

$$\theta(t) = \arccos\left(\frac{\sin \psi(t)}{\sqrt{1 + \lambda^2 - 2\lambda \cos \psi(t)}}\right), \quad (2)$$

where $\lambda = \frac{R_a + h}{R_a}$, R_a is the Earth's radius, and h is the satellite orbital height. Based on ephemeris information, the distance $d_{sa,i}(t)$ between LEO sa and user $i \in \{u_{sa}, u_{te}\}$ is

$$d_{sa,i}(t) = \sqrt{R_a^2 \sin^2 \varphi(t) + h^2 + 2R_a h - R_a \sin \varphi(t)}. \quad (3)$$

The Doppler shift $f_d(t)$ is

$$f_d(t) = \frac{v}{c} \times \left(\frac{R}{R+h} \cos \varphi(t)\right) \times f, \quad (4)$$

where f is the LEO carrier frequency, v is the satellite speed and c denotes the speed of electromagnetic wave propagation.

Therefore, after the LEO satellite downlink communication signal passes through the NTN-TDL-C channel, the signal received $y_{u_{te},t}$ by the terrestrial sensing node u_{te} during the t -th can be expressed as

$$y_{u_{te},t} = \begin{cases} h_{sa,u_{te}} * x_{sa,t} + w_t, & \mathcal{H}_1, \\ w_t, & \mathcal{H}_0, \end{cases} \quad (5)$$

where $x_{sa,t}$ denotes the satellite signal, $*$ is a convolution operation, and $\mathcal{H}_1$ and $\mathcal{H}_0$ represent the binary hypotheses to indicate whether the spectrum is occupied or idle, respectively. The term $h_{sa,u_{te}}$ is the channel impulse response. During a particular sensing period, $h_{sa,u_{te}}$ is considered constant, as the sensing duration is shorter than the channel coherence time. The w_t is modeled as additive white Gaussian noise.

In each spectrum occupancy or idle time T , it is divided into three parts of sensing time t_s , decision time t_j and utilization time t_{sh} . The duration of sensing time t_s and spectrum detection accuracy p_d directly influence spectrum efficiency. Based on (5), a total of N_{sy} symbol durations are chosen for user spectrum sensing; each terrestrial sensing node u_{te} conducts spectrum sensing and gathers $N = N_{sy} \times f_s \times T_{sy}$ observations, $y_{u_{te},t} = \{y_{u_{te},t}(1), \dots, y_{u_{te},t}(n), \dots, y_{u_{te},t}(N)\}$, $n = 1, 2, \dots, N$, where f_s is the sampling rate, and T_{sy} is the duration of each symbol. Therefore, the whole collaborative spectrum sensing matrix $\{\mathbf{Y}_t\}_{u_{te} \times N}$ in the t -th time slot is expressed as

$$\mathbf{Y}_t = \begin{pmatrix} y_{1,t}(1) & y_{1,t}(2) & \cdots & y_{1,t}(N) \\ \vdots & \vdots & \ddots & \vdots \\ y_{u_{te},t}(1) & y_{u_{te},t}(2) & \ddots & y_{u_{te},t}(N) \\ \vdots & \vdots & \ddots & \vdots \\ y_{U_{te},t}(1) & y_{U_{te},t}(2) & \cdots & y_{U_{te},t}(N) \end{pmatrix}. \quad (6)$$

Based on the observation vectors $\mathbf{Y}_t$, a test statistic T_y can be designed to facilitate decision-making in spectrum sensing with a preset detection threshold γ_{th} . $T_y > \gamma_{th}$ indicates that the primary satellite system is currently occupying the spectrum resource being detected; otherwise, the spectrum resource is considered idle. $\mathcal{F}$ is the mapping between statistic T_y and $\mathbf{Y}_t$ of signal and noise. According to the Neyman–Pearson (NP)

criterion [31], the key is to design a test statistic to maximize the detection probability P_d for a given false alarm probability P_f , which can be formulated as

$$\begin{aligned} \text{P1: } \max_{\mathcal{F}, T_y} P_d &= \int_{\gamma}^{\infty} f_{T_y|H_1}(t) dt \\ \text{s.t. } P_f &= \int_{\gamma}^{\infty} f_{T_y|H_0}(t) dt = \tau_{th} \end{aligned} \quad (7)$$

where $P_d = P(T_y > \gamma_{th}|H_1)$ and $P_f = P(T_y > \gamma_{th}|H_0)$ denote the detection probability and false alarm probability, respectively. $T_y|H_j$ is a test statistic under hypothesis H_j , $j \in \{0, 1\}$, $f_{T_y|H_j}(\cdot)$ is the probability density function of $T_y|H_j$, and τ_{th} denotes required false alarm probability P_f .

Aiming to achieve both a low false alarm probability and a high detection probability of spectrum sensing in the satellite–terrestrial converged system, significantly more stringent challenges arise compared to terrestrial cellular. (i) High transmission losses, atmospheric/cloud/rain attenuation, shadowing and noise uncertainty bring about a low SNR at the receiver; (ii) the rapid movement of LEO leads to highly dynamic sensing environments and large variations in Doppler shift and SNR over time; (iii) spectrum switching occurs frequently, leaving only a short duration available for sensing; and (iv) the variability of signal modulation types and the complexity of the dynamic sensing environment necessitate algorithms with robust generalization and resilience capabilities.

To address above challenges, a feasible solution is to design an intelligent algorithm that integrates both data-driven and model-driven paradigms to enable high detection probability under dynamic, low SNR and noise uncertainty environments. Considering the difficulty of deriving an explicit closed-form test statistic T_y in complex satellite–terrestrial scenarios, the nonlinear mapping $\mathcal{F}(\mathbf{Y}_t)$ can be learned in a data-driven manner to approximate the optimal detection rule under the Neyman–Pearson criterion formulated in problem P1. By leveraging deep neural networks, the nonlinear relationship between the collaborative observation matrix and detection decision can be effectively captured, thereby enhancing detection performance in complex satellite–terrestrial sensing scenarios.

3. DL-Enabled Multiple Feature Fusion Spectrum Sensing Framework Design

Faced with the inherent challenges in the satellite–terrestrial spectrum sensing and sharing system, such as dynamic sensing environments, sensing signal with diverse modulation types, low SNR and non-Gaussian characteristics, noise uncertainty, and short sensing durations, traditional model-driven methods that rely solely on single-domain signal statistics often exhibit limited sensing accuracy and robustness.

From a representation perspective, heterogeneous feature domains capture different aspects of the received signal. Statistical-domain features provide relatively noise-robust global descriptors, reflecting long-term distribution characteristics, whereas waveform-domain I/Q representations preserve fine-grained temporal dynamics and phase continuity. In satellite–terrestrial environments with severe path loss and fading, neither domain alone is sufficient to ensure stable sensing performance. By jointly modeling both domains within a dual-branch architecture, the proposed framework enables cross-domain interaction and complementary feature learning. Moreover, multi-domain fusion reduces representation variance while maintaining discriminative power, which is particularly important for improving robustness and generalization under dynamic and low-SNR conditions.

As shown in Figure 2, we propose a dynamic spectrum sensing method DL-MFFSSnet based on integrating both data-driven and model-driven paradigms. In the DL-MFFSSnet model, a dual-branch neural network architecture is employed to process both statistical

features of the signal and raw I/Q sampling data, based on multi-scale dilation convolution, a channel/spatial self-attention mechanism, and a deep residual network to capture multiple dimensions' signal characteristics and enhance feature interaction across heterogeneous domains.

Specifically, the DL-MFFSSnet framework consists of three main parts in Figure 2. Part A: Multi-feature extraction and fusion module includes signal/noise multi-feature tensor construction, as well as data fusion and dataset preparation. Part B: Offline training for the DL-MFFSSnet module incorporates the two constructed feature tensors of statistical features and the I/Q component as dual-branch inputs, with spectrum occupancy as the output. The cross-entropy function is employed as a loss function to guide model training. Part C: Online spectrum detection module enables accurate and rapid detection of spectrum occupancy based on real-time sensed signals.

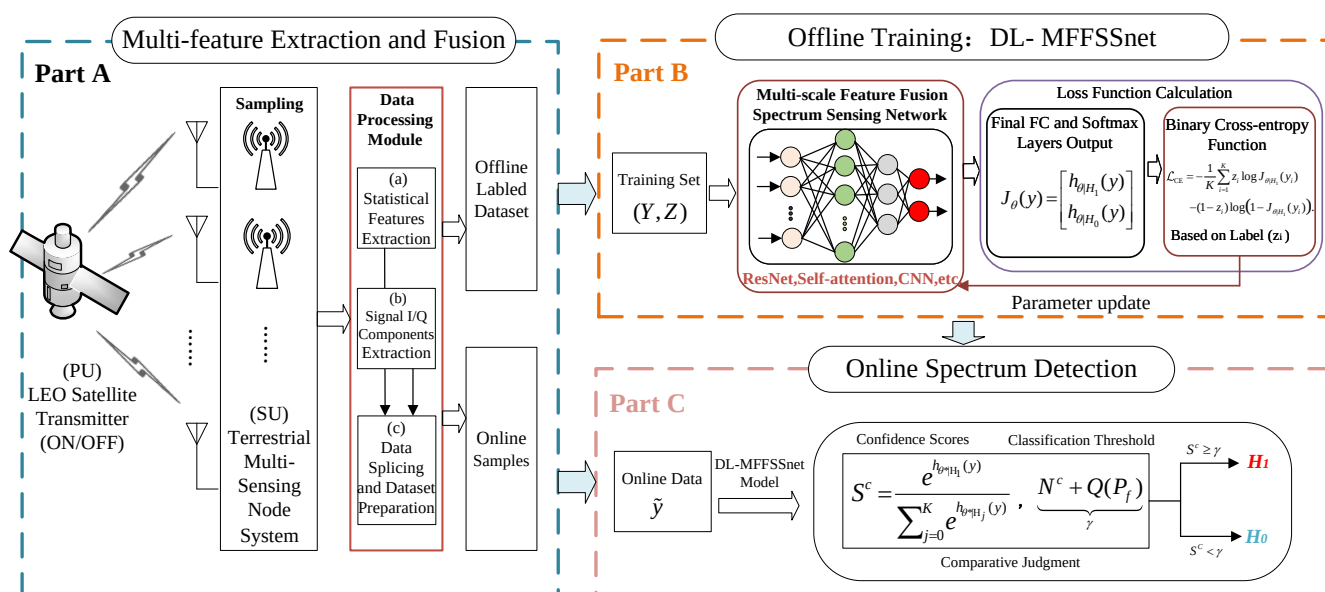


Figure 2. DL-enabled multi-feature fusion spectrum sensing framework.

4. Multi-Feature Extraction and Fusion Strategy

In the dynamic satellite–terrestrial downlink scenarios, the spectrum sensing performance relies on analyzing and extracting the characteristics of satellite signals and noise under various conditions, including signals with different modulation types, low SNR and noise uncertainty. Considering the differences between signal and noise in terms of strength, cyclic properties, sparsity and time–frequency domain distribution, we proposed a multi-feature fusion strategy that constructs two feature tensors of the statistical feature $f_{sf} \in \mathbb{R}^{N_s \times U_{te} \times N_u}$ and I/Q component $f_{iq} \in \mathbb{R}^{U_{te} \times N_{iq} \times N}$, providing strong support for DL-MFFSSnet to realize fast and accurate spectrum detection, where N_s , U_{te} and N_u are the number of features, sensing users and sliding windows for feature processing. Let N_{iq} denote the number of I/Q branches, and N denote the total number of samples. As shown in Figure 3, the specific processes include signal/noise generation and sampling, feature extraction, and data fusion and dataset preparation.

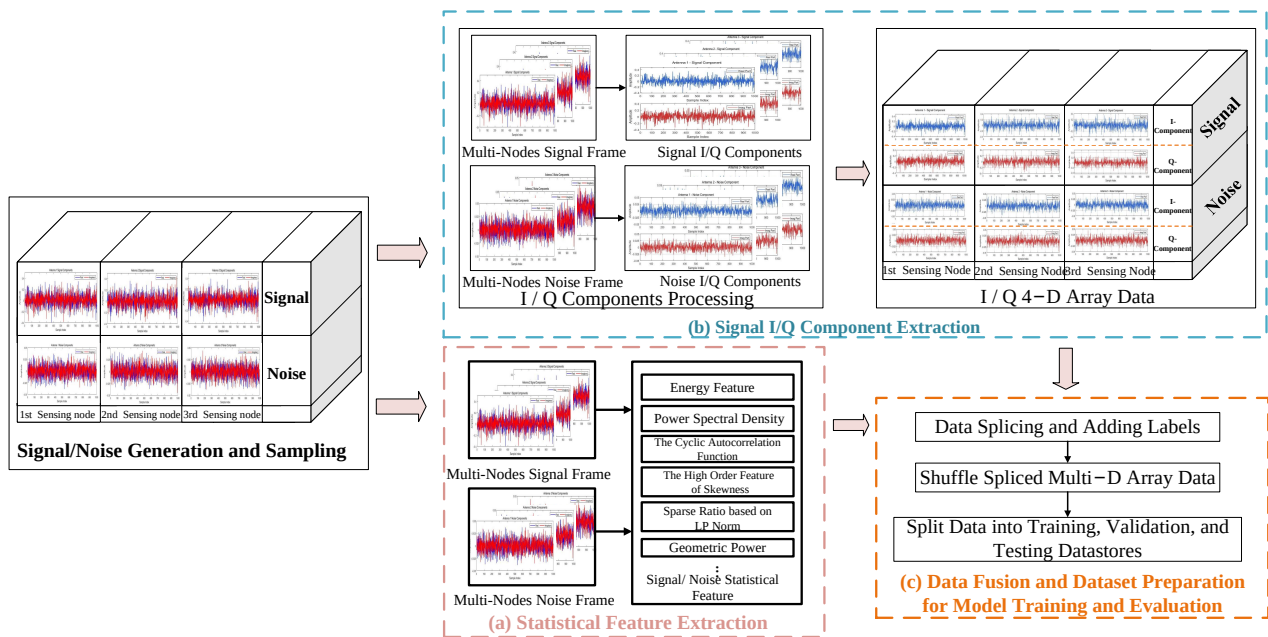


Figure 3. Signal/noise multi-feature extraction and fusion diagram.

4.1. Feature Extraction Phase

As shown in block (a) in Figure 3, $f_{sf} = \{E, G, Skew, PSD, CAF, SR\}^T$, and $f_{sf} \in \mathbb{R}^{N_s \times U_{te} \times N_u}$ is the statistical feature tensor that is constructed by multiple features including energy E , geometric power G , skewness $Skew$, power spectral density PSD , cyclic autocorrelation function CAF , and sparse ratio SR , which comprehensively characterize the signal from multiple perspectives, thereby enhancing the model's capability to detect spectrum occupancy. The feature extraction process is identical across different sensing nodes and time slots, and the $y_{u_{te},t}$ is abbreviated as y for simplicity. Therefore, based on the received signal sequence with discrete value $y_{u_{te},t} = [y_{u_{te},t}(0), y_{u_{te},t}(1), \dots, y_{u_{te},t}(N-1)]$ and $\mathbf{Y}_t = [y_{1,t}, y_{2,t}, \dots, y_{u_{te},t}]^T$, the statistical features can be described as follows.

- **Energy feature:** The energy feature is commonly used to analyze the total energy of signals, directly reflecting the signal amplitude and duration. To extract the energy feature, $y = [y(0), y(1), \dots, y(N-1)]$ is divided into 64 overlapping blocks via a sliding window mechanism, and N_u is the number of samples of each part u where each window sequentially traverses the data with a stride equal to the block size minus the overlap region, and the average energy of each part is expressed as

$$E_u = \frac{1}{N_u} \sum_{l=1}^{N_u} |y(l)|^2. \quad (8)$$

- **Geometric power:** Geometric power represents the geometric mean of the squared magnitudes of the signal, which provides a non-linear representation of signal intensity. The geometric power of $y(l)$ is denoted as

$$G_u = \exp \frac{1}{N_u} \sum_{l=1}^{N_u} \log |y(l)|, \quad (9)$$

where $\exp(\cdot)$ is an exponential operation.

- Higher-order moments of the skewness feature: The skewness captures the asymmetry of the signal amplitude distribution, reflecting non-Gaussian characteristics prevalent in STCNs. The skewness $Skew_u$ is expressed as

$$Skew_u = \frac{\frac{1}{N_u} \sum_{l \in N_u} (y(l) - \mu_u)^3}{\sigma_u^3}, \quad (10)$$

where μ_u and σ_u are the mean and standard deviation of the windowed signal.

- Power spectral density (PSD): The PSD reflects the distribution of a signal's power in the frequency domain, illustrating how the power varies with frequency, which represents the average power within a unit frequency bandwidth. The $PSD_y(k)$ is shown as

$$PSD_y(k) = \frac{1}{K} |Y(k)|^2, k = 0, 1, \dots, K-1, \quad (11)$$

where $PSD_y(k)$ represents the PSD of the k -th frequency component. $Y(k) = FFT(y(l)) = \sum_{l=0}^{N-1} y(l) e^{-j \frac{2\pi k l}{N}}$, and $FFT(\cdot)$ represents the fast fourier transform, and N is the length of the block (FFT size).

- Cyclic autocorrelation function (CAF): Compared to Gaussian white noise, modulated signals exhibit periodic autocorrelation properties, which are periodic with respect to T_0 in time t . The CAF is defined as

$$CAF_y(\tau) = \frac{1}{N} \sum_{n=1}^{L-\tau} y(l) y^H(l + \tau), \quad (12)$$

where τ is the constant hysteresis, $(\cdot)^H$ is a conjugate operation, and L is the signal length.

- Sparse ratio based on Lebesgue p-norm: The sparse ratio feature quantifies the signal sparsity, useful for detecting sparse modulation types. In this paper, we select $SR_u = LP_1 / LP_2$ ratio to measure the sparsity of a signal. The L_p norm calculation method is

$$L_p = \left(\frac{1}{N_u} \sum_{l=1}^{N_u} |y(l)|^p \right)^{\frac{1}{p}}, \quad 0 < p \leq 2, \quad (13)$$

As shown in block (b) in Figure 3, the I/Q component tensor $f_{sf} = \{I_N, Q_N\}^T$, and $f_{iq} \in \mathbb{R}^{U_{ie} \times N_{iq} \times N}$ is constructed by the in-phase (I) and quadrature (Q) sampling points. These I/Q components capture the raw complex signal characteristics, including amplitude, phase and modulation patterns, providing rich information for distinguishing spectral states. By leveraging the temporal and spectral dependencies inherent in the I/Q samples, the proposed model can effectively learn fine-grained representations of the signal structure. In each sensing slot, the input data has a dimension of $2 \times 3 \times 1024$, where 2 denotes the number of channels representing the I and Q data, and 3×1024 indicates the sampled points of three sensing nodes.

4.2. Sensing Data Fusion and Dataset Preparation Phase

As shown in block (c) in Figure 3, the dataset preparation phase includes three parts: data splicing and adding labels, shuffling spliced multi-D array data, and splitting data into training, validation, and testing datastores. The extracted I/Q components and corresponding statistical features are integrated to form a unified feature set. Feature labels, denoted as [1, 0] for signal and [0, 1] for noise, are added accordingly. In single/noise input features, the data of the dual-branch network's dimensions are $6 \times 3 \times 64$ and $2 \times 3 \times 1024$, where 6 denotes the number of statistic features, and 3×64 indicates the number of feature

samples for three sensing nodes under a single feature. The overall dataset contains 60,000 samples and is divided into training, validation, and test sets using a split ratio of 8:1:1. Specifically, the training set consists of 48,000 samples, the validation set includes 6000 samples, and the test set contains 6000 samples. Each set maintains an equal number of signal and noise frames.

5. DL-MFFSSnet Spectrum Sensing Network Framework

To address the challenges of dynamic and low-SNR spectrum sensing environments in STCNs, we propose a deep-learning-based multi-scale feature fusion spectrum sensing algorithm. As shown in Figure 4, the DL-MFFSSnet takes two types of input data: (i) the statistical features tensor, and (ii) the raw I/Q data tensor. The inputs of the signal/noise multi-feature extraction and fusion module are processed and aligned, as shown in Figure 3.

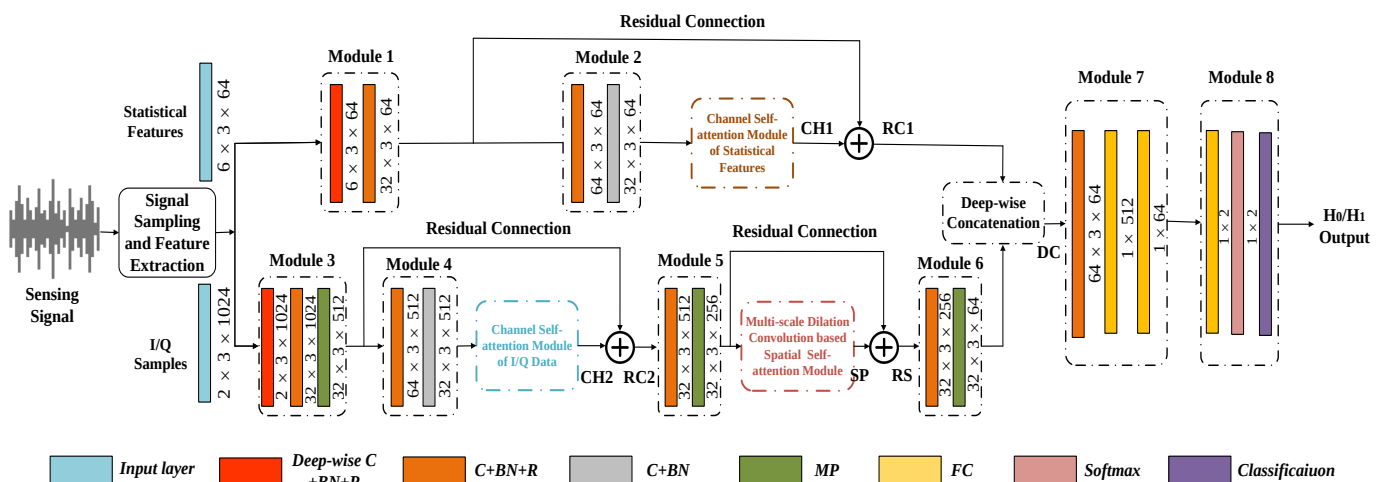


Figure 4. DL-MFFSSnet network structure.

Based on the constructed multi-feature tensors of sensed signals, the DL-MFFSSnet adopts a dual-branch deep neural network architecture to detect a spectrum. Specifically, in the statistical feature processing branch, CNN and channel self-attention (Ch_sa) are incorporated to capture the intra-channel correlations and inter-channel relative contributions of different feature modalities. In the I/Q data processing branch, multi-scale dilated convolutions (MSDC) and spatial self-attention (Sp_sa) mechanisms are introduced to analyze dependencies across different temporal positions and multi-scale spatial features. In addition, residual networks (Resnet) are incorporated into both branches to mitigate the vanishing gradient problem in deep networks and to promote the reuse of shallow and deep features. Finally, the features extracted from the two branches are flattened and concatenated, followed by a fully connected layer to obtain the final spectrum sensing result.

For notational simplicity in Figure 4, let C denote standard convolution, and deep-wise C indicate depthwise convolution. BN , R , MP and FC denote batch normalization, the activation function, max pooling and the fully connected layer. The network configurations of each module in DL-MFFSSnet are shown in Table 1. The main modules are described below.

Table 1. Network configurations of DL-MFFSSNet.

Set	I/P Size	Operation	[S, P, Fit]	O/P Size
I/P	Complex Signal	Feature Extraction	-	$2 \times 3 \times 1024$ $6 \times 3 \times 64$
M1	$6 \times 3 \times 64$ $6 \times 3 \times 64$	[DC1 + BN + R] [C2 + BN + R]	[(1,1), (1,2), 6] [(1,1), (1,1), 32]	$6 \times 3 \times 64$ $32 \times 3 \times 64$
M2	$32 \times 3 \times 64$ $64 \times 3 \times 64$	[C2 + BN + R] [C2 + BN]	[(1,1), (1,2), 64] [(1,1), (1,1), 32]	$64 \times 3 \times 64$ $32 \times 3 \times 64$
CH1	$32 \times 3 \times 64$ M2, CH1	Equation (14) $\otimes$	- -	$32 \times 1 \times 1$ $32 \times 3 \times 64$
RC1	M1, CH1	C3(M1), $\oplus$	[(1,1), -, 32]	$32 \times 3 \times 64$
M3	$2 \times 3 \times 1024$ $2 \times 3 \times 1024$ $32 \times 3 \times 1024$	[DC1 + BN + R] [C2 + BN + R] MP	[(1,1), (1,2), 2] [(1,1), (1,1), 32] [(1,2), (1,2), -]	$2 \times 3 \times 1024$ $32 \times 3 \times 1024$ $32 \times 3 \times 512$
M4	$32 \times 3 \times 512$ $64 \times 3 \times 512$	[C2 + BN + R] [C2 + BN]	[(1,1), (1,1), 64] [(1,1), (1,1), 32]	$64 \times 3 \times 512$ $32 \times 3 \times 512$
CH2	$32 \times 3 \times 512$ M4, CH2	Equation (14) $\otimes$	- -	$32 \times 1 \times 1$ $32 \times 3 \times 512$
RC2	M3, CH2	C3(M3), $\oplus$	[(1,1), -, 32]	$32 \times 3 \times 512$
M5	$32 \times 3 \times 512$ $32 \times 3 \times 512$	[C2 + BN + R] MP	[(1,1), (1,1), 32] [(1,2), (1,2), -]	$32 \times 3 \times 512$ $32 \times 3 \times 256$
SP	$32 \times 3 \times 256$ $128 \times 3 \times 256$ $32 \times 3 \times 256$ MSDC, SP	[MSDC] $\times 4$ C4 Equation (20) $\odot$	- [(1,1), (4,4), 32] - -	$128 \times 3 \times 256$ $32 \times 3 \times 256$ $1 \times 3 \times 256$ $32 \times 3 \times 256$
RS	M5, SP	$\oplus$	-	$32 \times 3 \times 256$
M6	$32 \times 3 \times 256$ $32 \times 3 \times 256$	[C2 + BN + R] MP	[(1,1), (1,2), 32] [(1,4), (1,4), -]	$32 \times 3 \times 256$ $32 \times 3 \times 64$
DCon	RC1, M6	Concat	-	$64 \times 3 \times 64$
M7	$64 \times 3 \times 64$ $64 \times 3 \times 64$ 1×512	[C1 + BN + R] Flatten, FC1 FC2	[(1,1), (1,2), 64] - -	$64 \times 3 \times 64$ 1×512 1×64
M8	1×64 1×2 1×2	FC3 Softmax Classification	- - -	1×2 1×2 -

I/P = Input, O/P = Output, S = Stride, P = Padding, Fit = fitters, M = Module, CH = Channel self-attention, SP = Spatial self-attention, RC= Residual connection, and DC = deep-wise Concatenation. DC1-C4 and MSDC denote 2×5 (depthwise), 2×3 , 1×1 , 7×7 , and 1×6 kernels used for convolution. FC1, FC2, and FC3 denote FC networks with 512, 64 and 2 neurons. MSDC denotes atrous convolutions with dilation factors of 1, 2, 4, and 6. $\odot$ and $\otimes$ denote element-wise multiplication and channel-wise multiplication, and $\oplus$ is element-wise addition.

5.1. Channel Self-Attention-Residual Block

The channel attention mechanism is designed to strengthen the representational capability of deep neural networks by dynamically adjusting the importance of feature channels. The channel attention approach is based on the intuition that different channels encapsulate varying degrees of task-relevant information and should therefore be selectively emphasized.

In DL-MFFSSNet, the residual channel attention mechanism is composed of two parts: a residual connection and a channel attention branch. In the channel attention branch, a squeeze-and-excitation strategy is employed. Specifically, the output feature of module 1, named M_1^O , is first processed by two convolutional layers in module 2 to produce an intermediate feature map M_2^O . Subsequently, M_2^O undergoes both global max pooling (GAP) and global average pooling (GMP) operations along spatial dimensions for each channel,

where the corresponding feature map is $F_{max}^{ch} = GMP(M_2^O)$ and $F_{avg}^{ch} = GAP(M_2^O)$, and $F_{avg}^{ch}, F_{max}^{ch} \in \mathbb{R}^{D_{M_2^O} \times 1 \times 1}$, where D is the number of channels. These operations generate global context descriptors by compressing feature dimensions from $D_{M_2^O} \times H_{M_2^O} \times W_{M_2^O}$ to $D_{M_2^O} \times 1 \times 1$, where H and W indicate the height and width of the feature map, thereby effectively capturing global information while reducing spatial redundancy. The channel feature F_{ch} is denoted as

$$F_{ch} = \sigma\{MLP[F_{avg}^{ch}] + MLP[F_{max}^{ch}]\}, \quad (14)$$

where $MLP(x) = W_1(\text{ReLU}(W_0(x)))$, which consists of two fully connected layers and a nonlinear ReLU activation function, to generate a set of channel-wise attention weights. Meanwhile, two different features share the same MLP to ensure that both descriptors are mapped into a common feature space, enabling the consistent modeling of channel-wise dependencies regardless of the type of pooling operation.

The weights are subsequently used to rescale the original feature map through channel-wise multiplication ($\otimes$), enabling the network to emphasize informative features while suppressing irrelevant or noisy ones. $F_{ch} \in \mathbb{R}^{D_{M_2^O} \times 1 \times 1}$, $M_2^O \in \mathbb{R}^{D_{M_2^O} \times H_{M_2^O} \times W_{M_2^O}}$, and the output of the channel attention module M_{ch}^O is

$$M_{ch}^O = F_{ch} \otimes M_2^O. \quad (15)$$

Finally, the output of the channel attention module is fused with residual connection by element-wise addition ($\oplus$) to produce the final feature representation M_{rc}^O , denoted as

$$M_{rc}^O = M_{ch}^O \oplus M_1^O. \quad (16)$$

The channel attention mechanism enables the model to focus on and exploit inter-channel feature relationships, while the residual connection helps avoid gradient vanishing and facilitates the fusion of shallow and deep features.

5.2. Multi-Scale Dilated Convolution-Based Spatial Self-Attention-Residual Block

The I/Q data inherently exhibit temporal continuity and local correlations. To effectively model multi-scale temporal-spatial dependencies, a module combining parallel dilated convolutions with varying dilation rates and spatial attention is designed, enabling an adaptive focus on informative regions while capturing diverse contextual features.

As shown in Figure 5, we choose multi-scale dilation convolution containing multiple dilation rates (1, 2, 4, and 6), which represent different receptive fields to capture both local details and global contexts simultaneously. The output feature map $M_5^O \in \mathbb{R}^{D_{M_5^O} \times H_{M_5^O} \times W_{M_5^O}}$ of module 5 is an input feature map to the MSDC module, where D , H and W represent the number of channels and the height and width of the output feature map, and each dilated convolution branch is defined as

$$F_{MSDC}^{d_i} = \text{Conv}_{d_i}(M_5^O), \quad d_i \in \mathbf{Dr}, \quad (17)$$

where $\mathbf{Dr} = \{d_1, d_2, \dots, d_{N_d}\}$ denotes the set of dilation rates, and Conv_{d_i} represents a convolution with dilation rate d_i and padding $p = d_i$ to maintain spatial resolution.

The multi-scale features are then concatenated along the channel dimension and compressed via a 1×1 convolution to obtain the fused feature map with the feature dimensions from $N_d D_{M_5^O} \times H_{M_5^O} \times W_{M_5^O}$ to $C_{M_5^O} \times H_{M_5^O} \times W_{M_5^O}$,

$$F_{MSDC} = \text{Conv}_{1 \times 1} \left(\text{Concat}(F_{MSDC}^{d_1}, \dots, F_{MSDC}^{d_{N_d}}) \right). \quad (18)$$

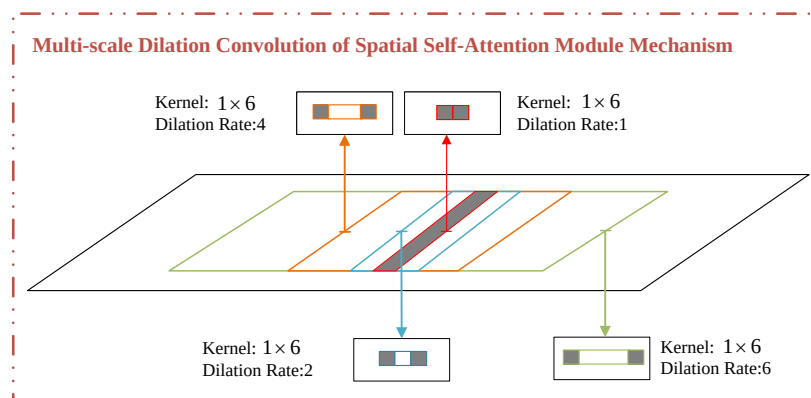


Figure 5. Multi-scale dilated convolution structure illustration.

Subsequently, a spatial attention map is generated to highlight informative regions. The spatial attention is

$$F_{sp} = \sigma(\text{Conv}(\text{Concat}(\text{GAP}_{ch}[F_{MSDC}], \text{GMP}_{ch}[F_{MSDC}]))), \quad (19)$$

where $\sigma(\cdot)$ is the activation function, $\text{GAP}_{ch}(\cdot)$ and $\text{GMP}_{ch}(\cdot)$ represent average-pooling and max-pooling operations across the channel dimension $D_{M_5^O}$, and $F_{sp} \in \mathbb{R}^{1 \times H_{M_5^O} \times W_{M_5^O}}$.

The spatially attended feature map M_{sp}^O is obtained by element-wise multiplication ($\odot$), and the M_{sp}^O is expressed as

$$M_{sp}^O = F_{sp} \odot F_{MSDC}. \quad (20)$$

Finally, a residual connection is employed to facilitate feature reuse and gradient propagation, and the feature vector is

$$M_6^I = M_{sp}^O \oplus M_5^O. \quad (21)$$

The integration of multi-scale dilated convolution, spatial attention, and residual connections enables the network to effectively capture spatial features across different receptive fields, focus on salient regions, and maintain stable feature propagation for improved representation learning.

5.3. Detection Mechanism of DL-MFFSSnet

In DL-MFFSSnet, the spectrum detection mechanism is based on Module 8, which comprises full connection, softmax and classification layers to output the sensed spectrum

as H_0 or H_1 using calculated probability for both hypotheses. The fully connected layer maps extracted deep features into probability space $h_\theta(y)$ corresponding to H_0 or H_1 ,

$$h_\theta(y) = \begin{bmatrix} h_{\theta|H_1}(y) \\ h_{\theta|H_0}(y) \end{bmatrix}. \quad (22)$$

Based on the softmax function, the confidence score of the signal and noise is $h_\theta^*(y) = [S^c, N^c]^T$, and S^c and N^c is [31]:

$$S^c = \frac{e^{h_{\theta|H_1}(y)}}{\sum_{j=0}^K e^{h_{\theta|H_j}(y)}}, \quad N^c = \frac{e^{h_{\theta|H_0}(y)}}{\sum_{j=0}^K e^{h_{\theta|H_j}(y)}}. \quad (23)$$

In the classification layer, the DL classifier compares the predicted class probabilities against a threshold γ . If the probability exceeds the threshold, the output is determined as 1; otherwise, it is determined as 0. Typically, a threshold of 0.5 is used in DL-based classifiers. In our work, we introduce the following improvements to the classification layer,

$$S^c = \frac{e^{h_{\theta|H_1}(y)}}{\sum_{j=0}^K e^{h_{\theta|H_j}(y)}} \begin{matrix} H_1 \\ \geq \\ H_0 \end{matrix} \begin{matrix} N^c + Q(P_f), \\ \underbrace{\gamma} \end{matrix} \quad (24)$$

where $Q(\cdot)$ is the standard Gaussian Q-function, and γ is the detection threshold. $Q(P_f)$ is a deterministic bias term derived from the target false alarm probability P_f . It is not learned during training, but analytically computed to adjust the decision threshold. Compared with the conventional fixed threshold of 0.5, the proposed formulation explicitly incorporates the false alarm constraint into the decision rule. By increasing P_f , the effective threshold γ decreases, leading to higher detection probability P_d ; conversely, decreasing P_f results in a more conservative decision boundary and lower false alarm rate.

Therefore, the proposed threshold adjustment mechanism enables explicit control over the P_d – P_f trade-off and shifts the operating point along the ROC curve without modifying the trained network parameters.

5.4. DL-MFFSSnet Training and Online Detection

5.4.1. Offline Training Stage:

In this stage, 80% of the labeled data from the dataset is randomly selected to construct the training set Ω , consisting of feature matrices associated with various SNR levels, and can be represented as

$$\Omega = \{(f_1, z_1), (f_2, z_2), \dots, (f_K, z_K)\}, \quad (25)$$

where K is the size of the training set, (f_k, z_k) represents the feature matrix and label of the k -th training sample, and z_k indicates the label of sensing data, and is denoted as

$$z_k = \begin{cases} [1, 0]^T, & H_1, \\ [0, 1]^T, & H_0. \end{cases} \quad (26)$$

By feeding f_k into DL-MFFSSnet, the output is a normalized 2×1 class score vector, and $J(f_k; \theta)$ is denoted as

$$J(f_k; \theta) = [H_0(f_k), H_1(f_k)]^T, \quad (27)$$

where J denotes the mapping of DL-MFFSSnet outputs to inputs, and θ represents the parameter set of DL-MFFSSnet. $H_\theta(f_k)$ represents the likelihood value $P(z_k = 0|f_k; \theta)$ of f_k

being classified as H_0 , and similarly, $H_\theta(y_k) = P(z_k = 1|f_k; \theta)$. Both $H_0(f_k)$ and $H_1(f_k)$ are values in the range $[0, 1]$. It must satisfy $H_0(f_k) + H_1(f_k) = 1$.

In spectrum sensing, the Binary Cross-Entropy (BCE) loss emerges as a ubiquitous optimization criterion. To evaluate the classification accuracy of DL-MFFSSnet, we define the following loss function to minimize the error between the predicted labels and true labels; the likelihood is $L(\theta) = P(Z|F; \theta) = \prod_{k=1}^K \left(J_{\theta|H_1}(f_k) \right)^{z_k} \left(J_{\theta|H_0}(f_k) \right)^{1-z_k}$. Based on $L(\theta)$, the binary cross-entropy loss function is adopted to optimize the spectrum sensing network. It penalizes the deviation between the predicted probability $H(f_i)$ and the truth z_i , and is defined as

$$\mathcal{L}_{\text{CE}} = -\frac{1}{K} \sum_{i=1}^K [z_i \log J_{\theta|H_1}(f_i) - (1 - z_i) \log (1 - J_{\theta|H_1}(f_i))]. \quad (28)$$

The goal of DL-MFFSSnet spectrum sensing model training is to obtain the optimal θ to maximize the posteriori probability $P(Z|F)$, which is denoted as

$$\theta^* = \arg \max_{\theta} P(Z|F; \theta), \quad (29)$$

where θ^* denotes the optimal θ under the maximum a posteriori (MAP) criterion.

5.4.2. Online Detection

After offline training, the well-trained DL-MFFSSnet is deployed for online spectrum sensing. During the online detection phase, real-time received signals are first preprocessed to extract the corresponding features f_k , which are then fed into the network to infer the probability of spectrum occupancy. Based on the output probabilities $h_\theta^*(f_k)$ and the predetermined decision threshold γ , the system determines the current spectrum state, classifying it as either idle (H_0 , if $h_\theta|H_1(f_k) < \gamma$) or occupied (H_1 , if $h_\theta|H_1(f_k) \geq \gamma$). This online detection process enables rapid and adaptive spectrum sensing under dynamic environmental conditions without the need for retraining. For clarity, we summarize the spectrum sensing strategy based on multi-feature fusion in Algorithm 1.

Algorithm 1 DL-MFFSSnet: DL-enabled Multiscale Feature Fusion Spectrum Sensing Algorithm

```

1: Step-I: Feature Extraction phase
2: Based on signal  $y(l)$ , obtain I/Q feature tensor, and generate statistical feature tensor with (8) to (13).
3: Label the corresponding feature matrix  $f(k)$  with  $z_k$  and split into training set  $\Omega_{\text{train}}$  and validation set  $\Omega_{\text{val}}$ ;
4: Step-II: Offline Training phase
5: Initialize DL-MFFSSnet parameters  $\theta$ , best validation loss  $L_{\text{best}} \leftarrow \infty$ , maximum epochs  $T_{\text{max}}$ ;
6: for  $t = 1$  to  $T_{\text{max}}$  do
7:   Shuffle and divide  $\Omega_{\text{train}}$  into mini-batches;
8:   for each mini-batch  $(\mathbf{f}, z)$  in  $\Omega_{\text{train}}$  do
9:     Forward propagate:  $\hat{z} \leftarrow J_{\theta}(\mathbf{f})$ ;
10:    Compute loss:  $\mathcal{L}_{\text{CE}} \leftarrow \mathcal{L}_{\text{CE}}(\hat{z}, z)$  with (28);
11:    Backpropagate and update  $\theta$ ;
12:   end for
13:   Compute validation loss  $\mathcal{L}_{\text{val}}$  on  $\Omega_{\text{val}}$ ;
14:   if  $\mathcal{L}_{\text{val}} < L_{\text{best}}$  then
15:      $L_{\text{best}} \leftarrow \mathcal{L}_{\text{val}}$ ,  $\theta^* \leftarrow \theta$ ;
16:   end if
17: end for
18: return  $\theta^*$ 
19: Step-III: Online Sensing phase
20: Input  $f_i$  into the trained network to obtain output vector  $J(\theta, \hat{f}_i) = [H_0(f_i), H_1(f_i)]^T$ ;
21: Determine the final spectrum state.
  
```

5.5. Computational Complexity Analysis

In the proposed DL-MFFSSnet, the main computational cost of each component lies in feature extraction, including energy, LP norm, and geometric power with a complexity of $O(N)$, power spectral density with $O(N \log N)$, and sample covariance matrix estimation with $O(M^2 N)$, where N denotes the number of sampling points. Each extra calculation of the neural network in online and offline processes $O\left(\sum_{i=1}^D c_{i-1} \cdot c_i \cdot s_i^2 \cdot L_i^2\right)$, where D denotes the total number of convolutional layers, c_{i-1} and c_i represent the input and output channels of the i -th layer, s_i is the kernel size, and L_i denotes the spatial resolution of the output feature map. The computational complexity of channel and spatial attention are approximately $O\left(CHW + \frac{C^2}{r}\right)$ and $O(CHW)$, where C , H , and W denote the number of channels and the height and width of the input, respectively, and r is the reduction ratio. So, the computational complexity of DL-MFFSSnet includes data preprocessing $O(N \log N + 5N)$, statistical feature processing $O\left(\sum_{i=1}^3 c_{i-1} \cdot c_i \cdot s_i^2 \cdot L_i^2 + c_3 L_3^2 + \frac{c_3^2}{r}\right)$, I/Q feature processing $O\left(\sum_{j=1}^5 c_{j-1} \cdot c_j \cdot s_j^2 \cdot L_j^2 + (c_{3,IQ} L_{3,IQ}^2 + \frac{c_{3,IQ}^2}{r} + c_{4,IQ} L_{4,IQ}^2)\right)$, and the detection process $O\left(\sum_{t=1}^3 L_{t-1} \cdot L_t\right)$.

6. Performance Evaluation

In this section, we evaluate the performance of the DL-MFFSSnet algorithm through simulations and compare it with state-of-the-art spectrum sensing strategies.

6.1. Experiment Setup and Dataset Generation

In the satellite–terrestrial converged downlink scenario, we consider LEO satellites operating at an orbital height of 600 km. Each satellite beam covers multiple terrestrial BSs, which are capable of performing collaborative spectrum sensing on received satellite

signals to assess the utilization of satellite spectrum resources. To better align with real-world LEO satellite downlink communication scenarios, according to ITU [34] and 3GPP standards [35], the training data is generated in a satellite S-band (commonly known as 2.1 GHz; uplink 1980–2100 MHz, downlink 2170–2200 MHz), and the satellite channel model is considered a NTN-TDL-C model, which combines path loss, shadow fading and Rician fading with Doppler shift and multi-path effects in satellite communications [35]. The first tap of NTN-TDL-C follows a Rician distribution with a K-factor of $K_1 = 10.224$ dB and a mean power of 0 dB [37], and the normalized delay profile is $[0, 0.3.2433 \times 10^{-8}, 8.5248 \times 10^{-8}]$ [38]. The free loss fading is $Loss = 92.45 + 20 \log_{10}(f) + 20 \log_{10}(d)$, where d is distance (km) and f is frequency (GHz) [36]. The satellite elevation angle φ is (10,90], and the gaseous/cloud/fog absorption with the rain attenuation coefficient $k = 0.0001$, $\alpha = 1.2$, the cloud attenuation coefficient $K_l = 0.05$ dB/km/(kg/m²), and the gaseous attenuation mean $\mu_g = 0.01$, $\sigma_g = 0.003$ [34]. Meanwhile, to closely emulate real-world satellite signal transmission, the sensed signals account for various signal modulation schemes (BPSK, QPSK, 8-PSK, 16QAM, 64QAM), as well as the wide range of SNR (−25 dB to 25 dB) resulting from different LEO satellite orbital positions and noise uncertainties. The other signal/noise generation parameters are as shown in Table 2.

Specifically, the received sensing samples are generated according to the hypotheses H_0 and H_1 defined in Section 2. Under H_0 , the received signal at each sensing node consists of additive noise only. Under H_1 , the received signal is modeled as the superposition of the satellite downlink OFDM signal and channel-affected noise, where the satellite–terrestrial channel follows the NTN-TDL-C model described above. For each time slot, the collaborative observation matrix $\mathbf{Y}_t$ is constructed using the received samples collected from multiple sensing nodes. The signal and noise datasets are constructed as follows.

- Signal dataset: Each ‘signal’ data indicates that the current orthogonal narrowband spectrum resource B_k is occupied. Considering the satellite downlink signals under different modulation schemes passing through the NTN-TDL-C channel and affected by uncertain noise, complex ‘signal’ data are generated for five modulation types, with 6000 sensed ‘signals’ per modulation type, totaling 30,000 signals, and the SNR ranges from −25 dB to 24 dB, with an interval of 1 dB. Specifically, the signals are generated with an OFDM waveform structure comprising 128 subcarriers, each spaced 15 kHz apart. A raised cosine filter is applied for pulse shaping. To further address the low SNR at the receiver, an oversampling technique is employed with a sampling rate of 7.68 MHz, corresponding to 1024 sampling points of two symbol per signal. Afterward, all ‘signals’ are stored in a signal datastore with a $[1, 0]^T$ label having $30,000 \times 2 \times 3 \times 1024$ dimensions, where $2 \times 3 \times 1024$ represents the length of a single complex frame (containing real and imaginary components) of three sensing nodes.
- Noise dataset: To effectively simulate the dynamic, complex, low SNR and noise uncertainties of satellite–terrestrial scenarios, the complex noise data are generated based on additive white Gaussian noise, with noise power levels ranging from −145 dBm to −50 dBm in increments of 5 dBm. For each noise power level, 1500 data samples are generated, resulting in a total of 30,000 datasets under varying noise conditions. Similar to the signal data, each noise sample consists of 1024 complex-valued samples. Similarly, the generated noise is stored in a separate noise datastore with a $[0, 1]^T$ label and $30,000 \times 2 \times 3 \times 1024$ dimensions.

Table 2. Simulation parameters.

Parameter	Value
LEO satellite altitude	600 km
LEO satellite speed	7.562 km/s
LEO satellite cycle	96.72 min
The satellite transmit power	46 dBm
The satellite antenna gain	35 dBi
The satellite elevation angle	(10,90]
The modulation types	BPSK, QPSK, 8PSK, 16QAM, 64QAM
Satellite sharing frequency	S-band (2.1 GHz)
The channel model	NTN-TDL-C model
NTN-TDL-C of K -factor	10.224 dB
The subcarrier interval	15 kHz
The symbol duration	66.66 μ s
The number of subcarriers	128
Sampling rate (S_r)	3.84 MHz
The samples of each signal	1024
SNR	[−25,24] dB, with 1 dB increment
The number of sensing signal	30,000
Noise model	AWGN
Noise power	[−145,−50] dBm, with 5 dBm increment
The number of noise sets	30,000
The size of training set	48,000
The size of validation set	6000
The size of test set	6000

A total of 60,000 complex waveforms is considered to train, validate and test the performance of DL-MFFSSnet with a split ratio of 8:1:1.

Each sample in the signal and noise datasets corresponds to one practical spectrum sensing process. Based on the received signal or noise samples, the sensing nodes construct the collaborative observation matrix $\mathbf{Y}_t$ according to the formulation in Section 2, and multi-domain features including energy, power spectral density (PSD), cyclic autocorrelation function (CAF), higher-order moments, sparse ratio, and I/Q samples are extracted to form the statistical feature tensor and the I/Q tensor. These tensors are then used as input samples for the supervised training and testing of DL-MFFSSnet. The data-driven detection framework DL-MFFSSnet can be interpreted as learning the nonlinear mapping $\mathcal{F}(\mathbf{Y}_t)$, which approximates the optimal test statistic under the Neyman–Pearson criterion formulated in Problem P1.

6.2. Performance Metrics

Spectrum sensing and detection is fundamentally a binary classification task, aiming to distinguish between two hypotheses: the presence (H_1) or absence (H_0) of a signal within a frequency band. True positives (T_p) occur when the system correctly detects the presence of a signal H_1 , while false positives (F_p) refer to it incorrectly labeling hypothesis H_0 as H_1 . Conversely, true negatives (T_n) denote correct identifications of signal absence, and false negatives (F_n) reflect missed detections of existing signals.

From the fundamentals of T_p , F_p , T_n and F_n , the key performance metrics of accuracy (Acc) evaluates overall correctness but may mislead in imbalanced scenarios, and precision (Pre) quantifies reliability in accurately predicting positive results,

$$\text{Acc} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}, \quad \text{Pre} = \frac{T_p}{T_p + F_p}. \quad (30)$$

True positive rate (TPR) or probability of detection (P_d) quantifies the detector's ability to avoid missed opportunities. The true negative rate (TNR) measures the ability to correctly identify the idle spectrum,

$$\text{TPR} = P_d = \frac{T_p}{T_p + F_n}, \quad \text{TNR} = \frac{T_n}{T_n + F_p}. \quad (31)$$

To characterize error risks, the false positive rate (FPR) reflects the likelihood of falsely alarming in negative cases, and the false negative rate (FNR) captures the probability of missing true positives,

$$\text{FPR} = P_f = \frac{F_p}{T_n + F_p}, \quad \text{FNR} = \frac{F_n}{T_p + F_n}. \quad (32)$$

The above metrics satisfy $\text{TPR} + \text{FNR} = 1$ and $\text{TNR} + \text{FPR} = 1$. Meanwhile, optimizing TPR minimizes wasted spectrum resources, whereas controlling F_p reduces interference risks.

6.3. Performance Comparison and Analysis

We provide the performance comparison and analysis of DL-MFFSSnet against the state-of-the-art methods, including:

- Deep STFT-CNN model [16]: This method constructs time-frequency feature maps using the STFT and subsequently employs a CNN model for spectrum detection.
- PSD-CT method [28]: This method takes signal power spectral density (PSD) as input features and employs the transformer network for spectrum sensing.
- APASS strategy [29]: It constructs two CM matrices: one for the current frame and another incorporating past CMs, and employs a CNN to jointly determine PU activity.
- Deep-CRnet [20]: Integrating a deep-CNN with a residual network to effectively extract features from I/Q data, enabling intelligent spectrum sensing.
- E-CNN-LSTM method [32]: The energy features of multi-sensing periods is the network input, and spectrum sensing is performed using a hybrid architecture combining CNN and LSTM networks.
- ED method [9]: The energy feature is employed as the primary basis for distinguishing signals from noise.

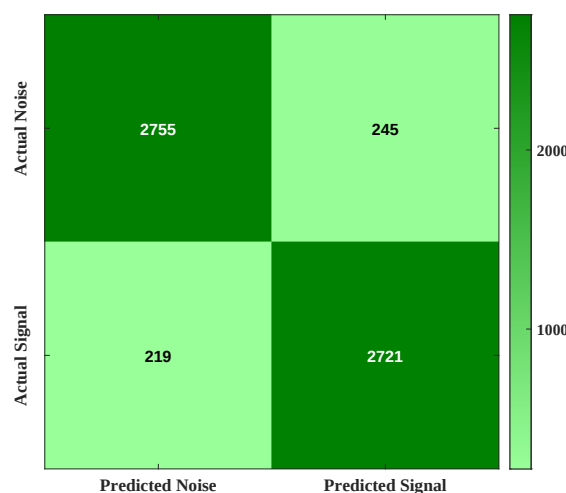
6.3.1. Prediction and Classification Performance Analysis Based on Confusion Matrices

As shown in Figure 6, the performance of DL-MFFSSnet in prediction and classification tasks was evaluated using the test dataset. The proposed model performs well under low SNR conditions and is capable of accurately distinguishing between signal and noise in scenarios involving multiple modulation types. Out of 6000 test samples (3000 signal and 3000 noise samples), the model correctly identifies 2721 signals, and only 219 are misclassified as noise, mainly within the SNR range of -25 dB to -15 dB. Simultaneously, 2744 noise samples are correctly classified, with only 245 false positives, all of which occurred under high noise power conditions of -50 dBm to -90 dBm.

Furthermore, Table 3 provides a detailed performance breakdown of typical spectrum sensing algorithms. The proposed DL-MFFSSnet achieves an accuracy of 92.19%, precision of 91.74%, TPR of 92.55%, TNR of 91.83%, FPR of 8.17%, and FNR of 7.45%. The prediction accuracy of the algorithm is limited to 91.74% due to the complexity of the sensing data, which involves multiple modulation types, extremely low SNR conditions down to -25 dB, and uncertain noise distributions. Despite these challenges, the proposed method outperforms deep CRnet [20] and APASS [29] in terms of detection accuracy and robustness where received satellite downlink signals deviate from complex Gaussian distribution.

Table 3. Performance metrics analysis.

Metric/Algorithm	STFT-CNN	APASS	Deep-CRnet	DL-MFFSSnet
Accuracy	0.8692	0.9022	0.8915	0.9219
Precision	0.8728	0.9015	0.8951	0.9174
TPR	0.8643	0.9027	0.8870	0.9255
TNR	0.8740	0.9017	0.8960	0.9183
FPR	0.1260	0.0983	0.1040	0.0817
FNR	0.1357	0.0973	0.1130	0.0745

**Figure 6.** Validation of prediction and classification accuracy.

6.3.2. Comparison of Detection Probability in Different Spectrum Sensing Methods

As illustrated in Figures 7 and 8, we compare the detection probability (P_d) in different spectrum sensing algorithms over a range of SNR values from -25 dB to 25 dB. In accordance with the IEEE 802.22 standard, the maximum acceptable false alarm probability is set to $P_f = 0.1$. Figure 8 shows that the detection performance of all algorithms improves as SNR increases. Notably, DL-MFFSSnet consistently outperforms other methods across the entire SNR range. Specifically, at a SNR of -15 dB, the detection probabilities of DL-MFFSSnet and APASS achieve above 90%, and their performance approaches 1 at SNR = -10 dB, indicating strong robustness under low-SNR conditions. Furthermore, when the SNR reaches 0 dB, the detection probability P_d exceeds 90% across all intelligent methods.

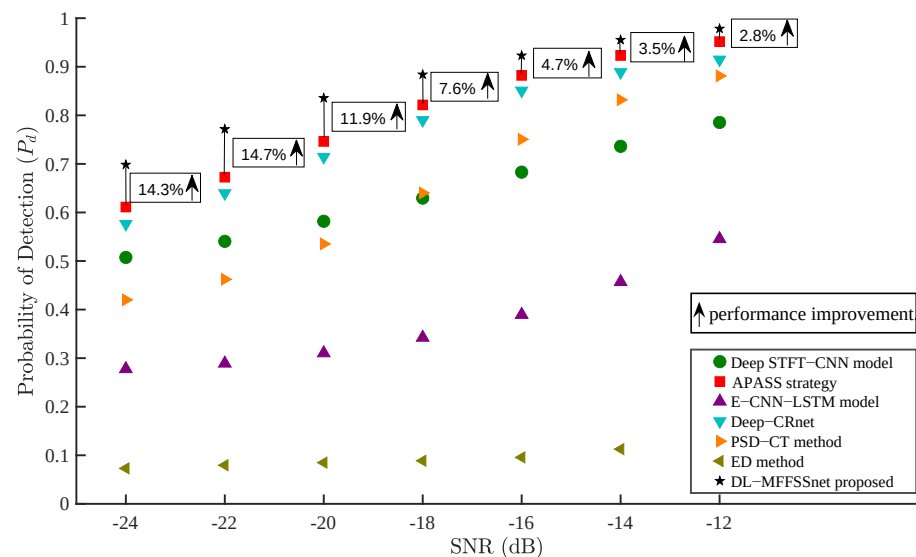


Figure 7. Comparative analysis of detection probability p_d in a low SNR, where $p_f = 0.1$.

The superior performance of DL-MFFSSnet, especially in the low SNR regime, can be attributed to its multi-feature fusion and modality-aware dual-branch architecture. Unlike single-domain approaches (e.g., PSD-based or CM-based models), the proposed framework jointly exploits complementary statistical-domain features and raw I/Q waveform representations. Statistical features provide global and noise-robust descriptors, while I/Q components preserve fine-grained temporal structures and modulation-specific characteristics. The fusion mechanism enables cross-domain interaction, allowing the model to compensate for information loss in one domain using discriminative cues from another. This complementary representation significantly enhances robustness under severe noise uncertainty and channel impairments.

To provide a more intuitive understanding of detection performance under low SNR conditions, as shown in Figure 7, at the SNRs of -24 dB and -20 dB, the detection probability of the proposed DL-MFFSSnet surpasses the APASS method by approximately 14.3% and 14.7%, respectively. The primary reason lies in the characteristics of satellite downlink transmission, where the received signals experience significant attenuation due to long-distance propagation. Additionally, impairments such as cloud attenuation, rain fading and shadowing effects lead to deviations from the complex Gaussian assumption, thereby degrading the performance of CM-based spectrum sensing models.

The relative performance ordering among different algorithms in Figure 7 can also be explained by their underlying modeling mechanisms. Traditional ED and PSD-based methods rely on energy statistics and frequency-domain features, which are highly sensitive to noise power uncertainty, leading to rapid degradation at a low SNR. CM-based approaches such as APASS improve robustness by exploiting second-order spatial correlations; however, they are fundamentally limited to covariance-domain information and may lose higher-order and temporal structural features. CNN-based models operating on single-domain inputs (e.g., STFT-CNN or Deep-CRNet) can learn nonlinear mappings but still process homogeneous feature modalities. In contrast, DL-MFFSSnet benefits from residual connections for stable deep feature propagation, spatial self-attention for modeling inter-feature dependencies, and channel self-attention for adaptively weighting heterogeneous feature contributions. These mechanisms collectively improve feature discrimination and generalization capability, resulting in consistently higher detection probability under severe low-SNR conditions.

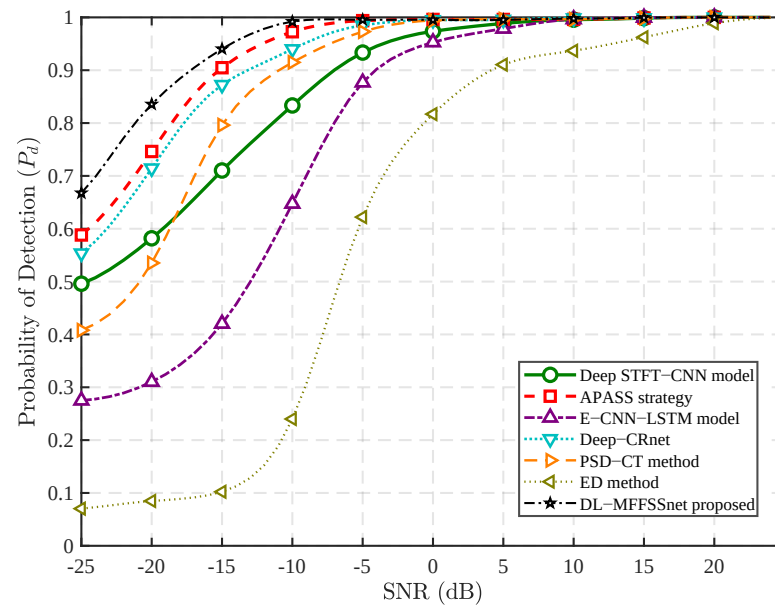


Figure 8. PD versus SNR curves for comparative analysis where $p_f = 0.1$, and the curves are Gaussian-smoothed.

6.3.3. Comparison of Detection Probability in Different False Alarm Probabilities

The ROC curves of detection probability (P_d) and false alarm probability (P_f) under Monte Carlo experiments are plotted in Figure 9a,b, where the P_f varies from 0 to 1 with a step-size of 0.01, and the SNR = −10 dB. The solid black curve as observed from Figure 9a indicates that the proposed DL-MFFSSnet significantly outperforms other detectors, including Deep STFT-CNN [16], PSD-CT [28], APASS [29], Deep-CRnet [20], and E-CNN-LSTM [32], in terms of correct detection probability under the same false alarm probability. Figure 9b presents the ROC curves on a logarithmic scale, highlighting the variation of P_f between 0.01 and 0.1. The proposed spectrum sensing method exhibits faster convergence within this range compared to other methods, with the detection probability approaching 1 when the false alarm probability exceeds 0.1.

6.3.4. Comparison of Spectrum Sensing Performance Under Different Numbers of Samples

In satellite–terrestrial spectrum sensing and sharing, the time required for spectrum sensing and the accuracy of detection directly impact the efficiency of spectrum sharing. The number of signal samples is directly related to the sensing duration, and the performance of the algorithm is compared under different sample lengths.

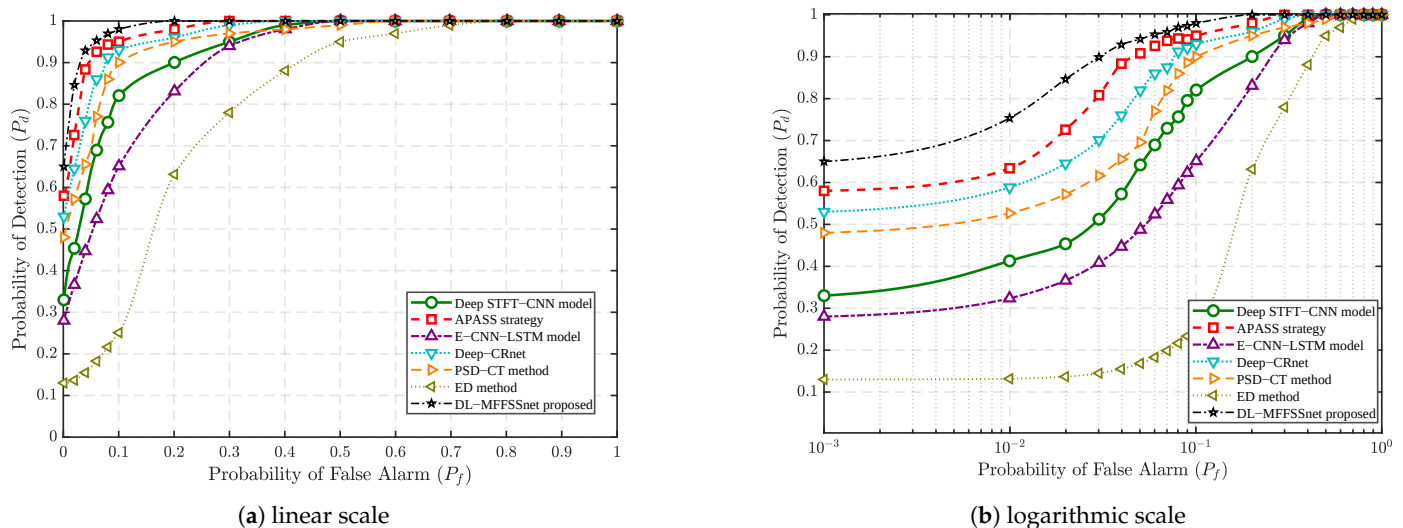


Figure 9. Comparison of p_d and p_f ROC performance curves, where SNR = -10 dB.

As shown in Figures 10 and 11, we respectively select 256, 512, and 1024 sampled I/Q data points as inputs for the model, and $p_f = 0.1$ and SNR = -15 dB. It can be clearly observed that the detection probability steadily improves with an increase in sensing data, particularly demonstrating superior detection performance in low-SNR scenarios. In Figure 10, at SNR = -20 dB, the proposed DL-MFFSSnet algorithm achieves detection probabilities of 0.82 and 0.68 for $N = 1024$ and $N = 512$, respectively. In comparison, the APASS strategy yields 0.75 and 0.62, while deep-CRnet achieves only 0.71 and 0.58, respectively. Evidently, the proposed algorithm exhibits superior performance even with fewer sampling points. Figure 11 presents the ROC curves under different sensing durations, from which it can clearly be observed that the proposed algorithm achieves superior performance, enabling better detection capability even at lower false alarm probabilities. The main reason behind the performance gain in Figures 10 and 11 is that, beyond relying solely on statistical features, the proposed method incorporates residual networks along with spatial and channel self-attention mechanisms, thereby more effectively capturing inherent correlations among consecutive data points.

6.3.5. Ablation Study of DL-MFFSSnet

As shown in Figures 12 and 13, we analyzed the ablation experiment performance of DL-MFFSSnet. We mainly studied the impact of different modules of DL-MFFSSnet and different feature inputs on the overall sensing performance.

Figure 12 analyzes the impact of different input features on perception performance, including single-branch inputs under different statistical features, single-branch inputs of I/Q data, and dual-branch multiple feature inputs combining statistical features and I/Q features.

From the simulation results shown in Figure 12, a clear performance progression can be observed as additional feature components are incrementally introduced. When only energy and power-based statistics (E+G) are used, the sensing performance degrades significantly at low SNR below -10 dB, with a detection probability of 0.25 at -20 dB, since energy detection relies solely on second-order power characteristics and is highly sensitive to noise power uncertainty. The inclusion of PSD improves detection probability by incorporating frequency-domain distribution information, which captures spectral structure beyond total energy, increasing the detection probability at -20 dB to 0.3. Further introducing CAF yields more pronounced gains, as cyclostationary features exploit periodic

signal properties that are inherently absent in stationary Gaussian noise, thereby enhancing robustness under severe low SNR conditions. In contrast, skewness and sparsity features provide only limited additional improvement, since higher-order amplitude asymmetry becomes less distinctive in satellite–terrestrial fading environments where attenuation and multipath effects tend to approximate Gaussian statistics. As a result, purely statistical-domain sensing remains insufficient, with detection probability around 0.45 at -20 dB where signals are deeply submerged in noise.

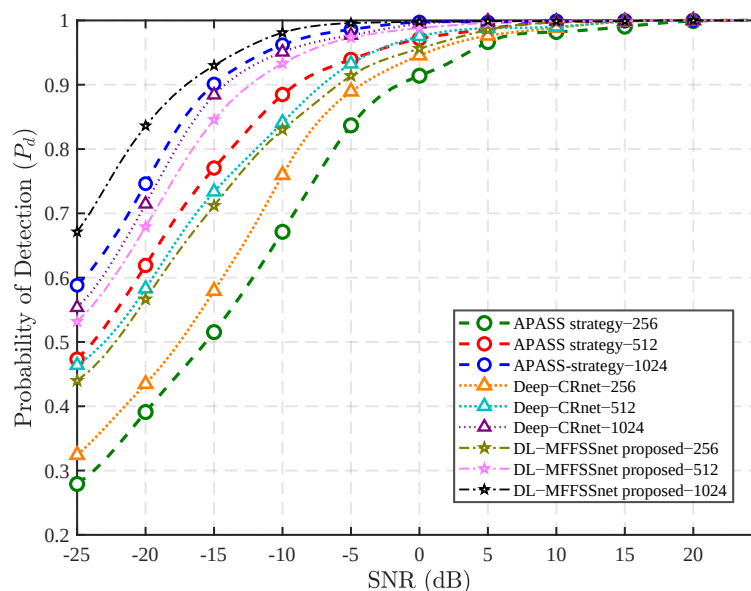


Figure 10. Comparison of p_d and SNR with different samples; $N = \{256, 512, 1024\}$, $p_f = 0.1$. (Gaussian-smoothed curves.)

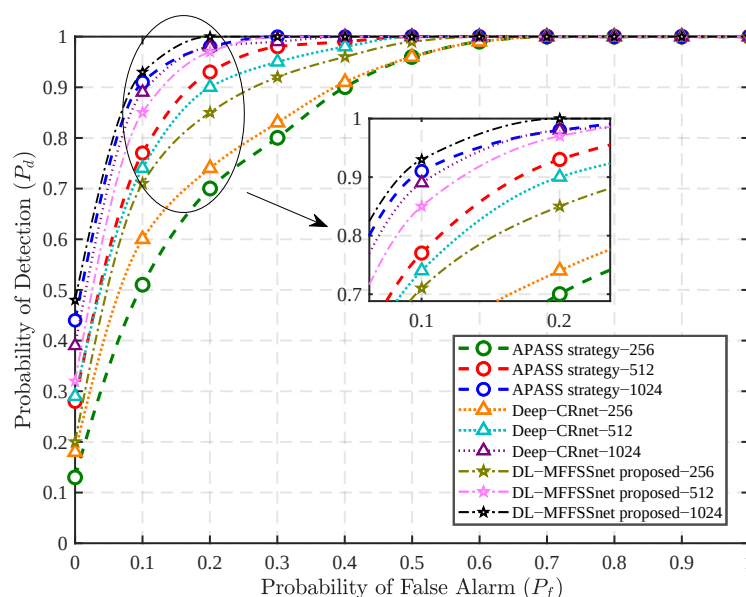


Figure 11. Comparison of ROC curves with different samples where $\text{SNR} = -15$ dB, and the curves are Gaussian-smoothed.

Notably, the single-branch I/Q feature tensor outperforms all purely statistical configurations in the low SNR regime. This is because raw I/Q samples preserve fine-grained temporal dynamics, phase continuity, and local correlation structures that are not fully captured by aggregated statistical descriptors. Even when global energy differences are weak, convolutional modeling on waveform-level data can extract nonlinear discriminative

patterns embedded in modulation transitions and instantaneous phase variations. However, waveform-only modeling may still exhibit sensitivity to severe noise distortions due to the absence of global statistical constraints.

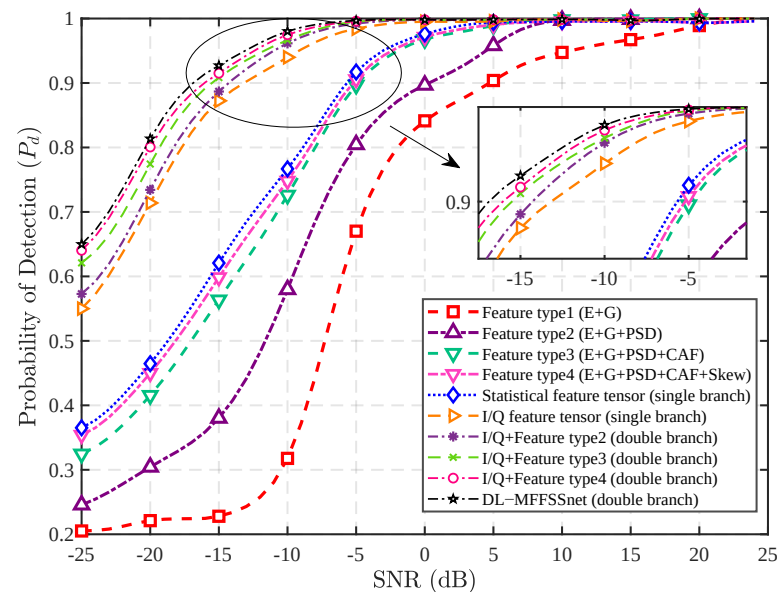


Figure 12. Ablation experiments of different features, where $p_f = 0.1$ and Gaussian smoothed.

By jointly integrating statistical-domain descriptors and waveform-domain representations within the dual-branch architecture, DL-MFFSSnet effectively combines global noise-robust characteristics with detailed temporal structures. This cross-domain complementarity enhances discriminative capability and generalization stability, thereby explaining the consistent performance superiority of the proposed multi-feature fusion framework across the entire low SNR range.

Figure 13 shows a comparison of the performance importance of each module, including the multi-scale dilated convolution module (MCD), residual network module (ResNet), channel self-attention module (Ch-sa), and spatial self-attention module (Sp-sa).

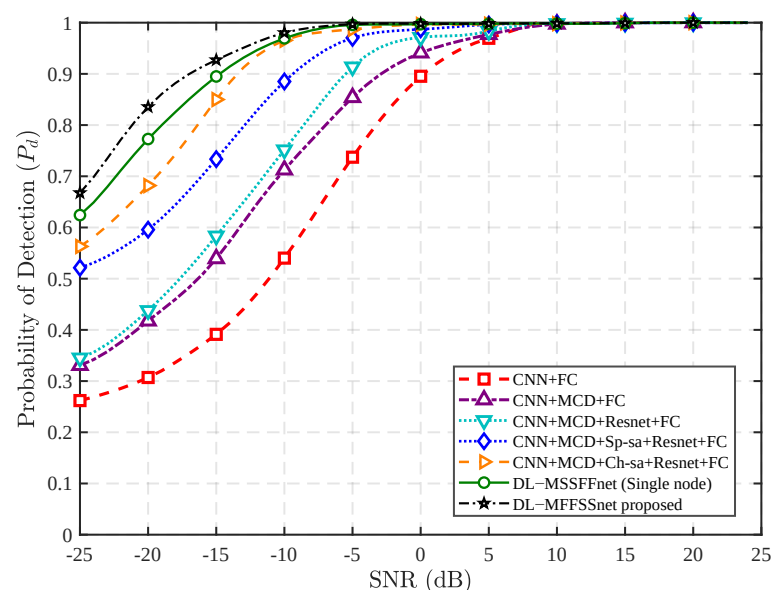


Figure 13. Ablation experiments of different modules, where $p_f = 0.1$ and the curves are Gaussian-smoothed.

The proposed DL-MFFSSnet algorithm, which incorporates MCD, self-attention mechanisms, and residual connections on top of traditional CNN and FC networks, achieves a detection probability of 0.82 under the condition of SNR = −20 dB. The channel and spatial self-attention mechanisms effectively capture the trend of changes in consecutive sampling points, significantly improving perception performance in low SNR environments. Figure 13 validates the importance of each module in DL-MFFSSnet to improve the sensing performance.

More specifically, the multi-scale dilated convolution (MCD) module enhances the receptive field without increasing parameter complexity, enabling the model to capture long-range temporal dependencies that are easily submerged in noise. The residual connections stabilize gradient propagation and allow deeper feature abstraction, which improves robustness but contributes moderately compared with attention mechanisms. The spatial self-attention (Sp-sa) module explicitly models correlations across sampling positions, strengthening global context modeling under low SNR. Meanwhile, the channel self-attention (Ch-sa) module dynamically re-weights heterogeneous feature channels, suppressing noise-dominant channels and emphasizing more reliable discriminative components. Therefore, modules that explicitly model cross-position and cross-channel dependencies (Sp-sa and Ch-sa) show larger performance gains than purely structural enhancements (e.g., residual connections). This explains the observed performance ordering among different configurations in Figure 13.

6.3.6. Complexity Analysis and Comparison

Although the proposed DL-MFFSSnet method presents satisfactory performance, we should be aware of the additional complexity, and the results are shown in Table 4.

Table 4. Runtime and Model Complexity Evaluation.

Algorithm	FLOPs	Total Parameters	Detection (ms)
Deep STFT-CNN [16]	656,441,856	16,868,610	2.56
APASS [29]	31,353,472	5,589,042	0.77
E-CNN-LSTM [32]	748,971	342,332	0.31
Deep-CRnet [20]	25,768,671	422,188	1.03
PSD-CT [28]	6,576,689,152	4,880,138	8.26
DL-MFFSSnet	95,468,800	6,442,698	1.26

As shown in Table 4, the proposed DL-MFFSSnet requires 95,468,800 FLOPs and 6,442,698 parameters, resulting in a detection time of 1.26 ms per sensing instance. Compared with APASS (31,353,472 FLOPs, 0.77 ms) and Deep-CRnet (25,768,671 FLOPs, 1.03 ms), DL-MFFSSnet introduces moderately higher computational cost. This increase mainly stems from the enlarged input dimensionality caused by multi-feature fusion, as well as the dual-branch architecture incorporating multi-scale dilated convolutions and channel/spatial self-attention mechanisms.

However, the detection latency remains at the millisecond level and is still well within the practical sensing slot duration in satellite–terrestrial dynamic spectrum access systems. More importantly, DL-MFFSSnet achieves higher detection probability under low-SNR conditions compared to both APASS and Deep-CRnet, as demonstrated in the previous experiments. Therefore, the proposed method attains a favorable trade-off between improved sensing performance and acceptable computational overhead, without compromising real-time sensing capability.

7. Conclusions

In this paper, we have investigated a collaborative satellite–terrestrial downlink spectrum sensing architecture in which multiple terrestrial sensing nodes cooperatively sense and detect idle satellite spectrum through multi-domain feature-level fusion, aiming to improve spectrum utilization efficiency in STCNs. To enhance detection probability under low-SNR, non-Gaussian signal and noise uncertainty, a multi-feature fusion strategy was developed to jointly characterize signal and noise from statistical-domain and waveform-domain perspectives. Based on constructed feature tensors, we proposed a dynamic spectrum sensing model, DL-MFFSSnet, based on channel/spatial self-attention and multi-scale dilation convolutions to capture inter-channel relative, inter-channel relative contributions of different feature modalities and multi-scale spatial features, achieving accurate and fast spectrum detection. Extensive simulations demonstrated that the proposed method achieves consistently higher detection probability and robustness across diverse modulation types and wide SNR ranges compared with state-of-the-art sensing approaches. At an SNR of -24 dB, the DL-MFFSSnet is 14.3% higher than the APASS strategy, 21.1% higher than Deep-CRnet and 37.0% higher than the Deep STFT-CNN model in spectrum detection probability, and when SNR is -15 dB, the improvements are 4.7%, 8.5%, and 23.2%, respectively.

Despite these promising results, several limitations remain. The current study assumes a predefined channel model and fixed sensing-node configuration, and the generalization performance under unseen channel conditions, varying interference scenarios, and scalable network sizes requires further investigation. Moreover, practical deployment challenges such as real-time processing latency, computational overhead, synchronization errors, hardware impairments, and large-scale data acquisition in real satellite–terrestrial environments need to be carefully addressed. Future work will focus on enhancing cross-condition generalization capability, adaptive online learning strategies, and validation using real measurement data. In addition, extending the framework to wideband spectrum sensing, dynamic spectrum access optimization, and multi-level power recognition in large-scale STCN deployments constitutes an important research direction.

Author Contributions: methodology, C.T. and Y.C.; software, C.T., G.C. and Z.W.; validation, C.T., H.L. and Z.W.; investigation, L.D.; resources, Y.C.; writing—original draft preparation, C.T. and Y.C.; writing—review and editing, L.D. and G.C.; visualization, C.T.; project administration, Y.C.; and funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Science and Technology Major Project of China (2025ZD1304200).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be provided by the corresponding author upon reasonable request by the reader.

Conflicts of Interest: No conflict of interest exists in the submission of this manuscript.

References

1. Saad, W.; Bennis, M.; Chen, M. A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Netw.* **2019**, *34*, 134–142. <https://doi.org/10.1109/MNET.001.1900287>.
2. Xiao, Y.; Ye, Z.; Wu, M.; Li, H.; Xiao, M.; Alouini, M.S.; Al-Hourani, A.; Cioni, S. Space-Air-Ground Integrated Wireless Networks for 6G: Basics, Key Technologies, and Future Trends. *IEEE J. Sel. Areas Commun.* **2024**, *42*, 3327–3354. <https://doi.org/10.1109/JSAC.2024.3492720>.

3. Lee, H.W.; Medles, A.; Chen, C.C. Feasibility and Opportunities of Terrestrial Network and Non-Terrestrial Network Spectrum Sharing. *IEEE Wirel. Commun.* **2023**, *30*, 36–42. <https://doi.org/10.1109/MWC.001.2300209>.
4. Yue, P.; An, J.; Zhang, J.; Ye, J.; Pan, G.; Wang, S.; Xiao, P.; Hanzo, L. Low Earth Orbit Satellite Security and Reliability: Issues, Solutions, and the Road Ahead. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 1604–1652. <https://doi.org/10.1109/COMST.2023.3296160>.
5. Tang, C.; Chen, Y.; Chen, G.; Du, L.; Liu, H. A Dynamic and Collaborative Spectrum Sharing Strategy Based on Multi-Agent DRL in Satellite-Terrestrial Converged Networks. *IEEE Trans. Veh. Technol.* **2025**, *74*, 7969–7984. <https://doi.org/10.1109/TVT.2024.3523331>.
6. Zhao, B.; Wang, M.; Xing, Z.; Ren, G.; Su, J. Integrated Sensing and Communication Aided Dynamic Resource Allocation for Random Access in Satellite Terrestrial Relay Networks. *IEEE Commun. Lett.* **2023**, *27*, 661–665. <https://doi.org/10.1109/LCOMM.2022.3227594>.
7. Liu, X.; Lam, K.Y.; Li, F.; Zhao, J.; Wang, L.; Durrani, T.S. Spectrum Sharing for 6G Integrated Satellite-Terrestrial Communication Networks Based on NOMA and CR. *IEEE Netw.* **2021**, *35*, 28–34. <https://doi.org/10.1109/MNET.011.2100021>.
8. Rong, B.; Rutagemwa, H. Leveraging Large Language Models for Intelligent Control of 6G Integrated TN-NTN With IoT Service. *IEEE Netw.* **2024**, *38*, 136–142. <https://doi.org/10.1109/MNET.2024.3384013>.
9. Digham, F.F.; Alouini, M.S.; Simon, M.K. On the Energy Detection of Unknown Signals Over Fading Channels. *IEEE Trans. Commun.* **2007**, *55*, 21–24. <https://doi.org/10.1109/TCOMM.2006.887483>.
10. Axell, E.; Leus, G.; Larsson, E.G.; Poor, H.V. Spectrum Sensing for Cognitive Radio: State-of-the-Art and Recent Advances. *IEEE Signal Process. Mag.* **2012**, *29*, 101–116. <https://doi.org/10.1109/MSP.2012.2183771>.
11. Cohen, D.; Eldar, Y.C. Sub-Nyquist Cyclostationary Detection for Cognitive Radio. *IEEE Trans. Signal Process.* **2017**, *65*, 3004–3019. <https://doi.org/10.1109/TSP.2017.2684743>.
12. Heydarishahreza, N.; Han, T. Spectrum Sharing and Interference Management for 6G LEO Satellite-Terrestrial Network Integration. *IEEE Commun. Surv. Tutor.* **2025**, *27*, 2794–2825. <https://doi.org/10.1109/COMST.2024.3507019>.
13. Mahboob, S.; Liu, L. Revolutionizing Future Connectivity: A Contemporary Survey on AI-Empowered Satellite-Based Non-Terrestrial Networks in 6G. *IEEE Commun. Surv. Tutor.* **2024**, *26*, 1279–1321. <https://doi.org/10.1109/COMST.2023.3347145>.
14. Hu, J.; Bian, D.; Tian, X. Sensing-Based Spectrum Sharing Method in Satellite-Terrestrial Network. *Space-Integr.-Ground Inf. Netw.* **2021**, *2*, 43–52.
15. Zhang, W.; Wang, Y.; Chen, X.; Cai, Z.; Tian, Z. Spectrum Transformer: An Attention-Based Wideband Spectrum Detector. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 12343–12353. <https://doi.org/10.1109/TWC.2024.3391515>.
16. Chen, Z.; Xu, Y.Q.; Wang, H.; Guo, D. Deep STFT-CNN for Spectrum Sensing in Cognitive Radio. *IEEE Commun. Lett.* **2021**, *25*, 864–868. <https://doi.org/10.1109/LCOMM.2020.3037273>.
17. Zhang, J.; He, Z.Q.; Rui, H.; Xu, X. Multiband Joint Spectrum Sensing via Covariance Matrix-Aware Convolutional Neural Network. *IEEE Commun. Lett.* **2022**, *26*, 1578–1582. <https://doi.org/10.1109/LCOMM.2022.3163841>.
18. Vijay, E.V.; Aparna, K. Deep Learning-CT based spectrum sensing for cognitive radio for proficient data transmission in Wireless Sensor Networks. *e-Prime-Adv. Electr. Eng. Electron. Energy* **2024**, *9*, 100659–100668. <https://doi.org/https://doi.org/10.1016/j.prime.2024.100659>.
19. Ahmed, R.; Chen, Y.; Hassan, B. Optimal Spectrum Sensing in MIMO-based Cognitive Radio Wireless Sensor Network Using GLRT with Noise Uncertainty at Low SNR. *AEU-Int. J. Electron. Commun.* **2021**, *136*, 153741–153759. <https://doi.org/10.1016/j.aeue.2021.153741>.
20. Ahmed, R.; Chen, Y.; Hassan, B. Deep learning-driven opportunistic spectrum access (OSA) framework for cognitive 5G and beyond 5G (B5G) networks. *Ad Hoc Netw.* **2021**, *123*, 102632–102646. <https://doi.org/10.1016/j.adhoc.2021.102632>.
21. He, P.; Lei, H.; Wu, D.; Wang, R.; Cui, Y.; Ying, Z. Non-Terrestrial Network Technologies: Applications and Future Prospects. *IEEE Internet Things J.* **2025**, *12*, 6275–6299. <https://doi.org/10.1109/JIOT.2024.3522912>.
22. Tian, Q.; Wu, Y.; Shen, F.; Zhou, F.; Wu, Q.; Dobre, O.A. ED-based Spectrum Sensing for the Satellite Communication Networks Using Phased-array Antennas. *IEEE Commun. Lett.* **2023**, *27*, 1834–1838.
23. Benedetto, F.; Giunta, G.; Pallotta, L. Cognitive Satellite Communications Spectrum Sensing Based on Higher Order Moments. *IEEE Commun. Lett.* **2021**, *25*, 574–578. <https://doi.org/10.1109/LCOMM.2020.3029091>.
24. Mourougayane, K.; Amgothu, B.; Bhagat, S.; Srikanth, S. A robust multistage spectrum sensing model for cognitive radio applications. *AEU-Int. J. Electron. Commun.* **2019**, *110*, 152876–152887. <https://doi.org/10.1016/j.aeue.2019.152876>.
25. Syed, S.N.; Lazaridis, P.I.; Khan, F.A. Deep Learning Approaches for Spectrum Sensing in Cognitive Radio Networks. In Proceedings of the 2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC), Herning, Denmark, 30 October–2 November 2022; pp. 480–485. <https://doi.org/10.1109/WPMC55625.2022.10014805>.
26. Ding, X.; Lv, Q.; Zou, Y.; Zhang, G. Spectrum Prediction for Satellite based Spectrum-Sensing Systems Using Deep Learning. In Proceedings of the 2022 IEEE Global Communications Conference (GLOBECOM), Rio de Janeiro, Brazil, 4–8 December 2022; pp. 3472–3477. <https://doi.org/10.1109/GLOBECOM48099.2022.10000832>.

27. Bai, W.; Zheng, G.; Mu, Y.; Ma, H.; Xue, Y. Cooperative spectrum sensing method based on channel attention and parallel CNN-LSTM. *Digit. Signal Process.* **2025**, *158*, 104963–104974. <https://doi.org/https://doi.org/10.1016/j.dsp.2024.104963>.
28. Liu, S.; Xu, D.; Liu, D.; Hu, G.; Kang, K. A Study on a Spectrum Sensing Method Based on Feature-Fusion Transformer. *J. China Acad. Electron. Inf. Technol.* **2024**, *19*, 658–666.
29. Xie, J.; Liu, C.; Liang, Y.C.; Fang, J. Activity Pattern Aware Spectrum Sensing: A CNN-Based Deep Learning Approach. *IEEE Commun. Lett.* **2019**, *23*, 1025–1028. <https://doi.org/10.1109/LCOMM.2019.2910176>.
30. Xie, J.; Fang, J.; Liu, C.; Li, X. Deep Learning-Based Spectrum Sensing in Cognitive Radio: A CNN-LSTM Approach. *IEEE Commun. Lett.* **2020**, *24*, 2196–2200. <https://doi.org/10.1109/LCOMM.2020.3002073>.
31. Liu, C.; Wang, J.; Liu, X.; Liang, Y.C. Deep CM-CNN for Spectrum Sensing in Cognitive Radio. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2306–2321. <https://doi.org/10.1109/JSAC.2019.2933892>.
32. Ding, X.; Ni, T.; Zou, Y.; Zhang, G. Deep Learning for Satellites Based Spectrum Sensing Systems: A Low Computational Complexity Perspective. *IEEE Trans. Veh. Technol.* **2023**, *72*, 1366–1371. <https://doi.org/10.1109/TVT.2022.3211845>.
33. 3GPP. *Solutions for NR to Support Non-Terrestrial Networks (NTN): Non-Terrestrial Networks (NTN) Related RF and Co-Existence Aspects*; Technical Report TR 38.863, V17.2.0; 3GPP: Sophia Antipolis, France, 2023. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3926> (accessed on 14 January 2026).
34. ITU. *Attenuation by Atmospheric Gases and Related Effects*. Technical Report P.676-12; ITU-R: Geneva, Switzerland, 2019. Available online: <https://www.itu.int/rec/R-REC-P.616/en> (accessed on 14 January 2026).
35. 3GPP. *Solutions for NR to Support Non-Terrestrial Networks (NTN)*; Technical Report TR 38.821, V16.2.0; 3GPP: Sophia Antipolis, France, 2023. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3525> (accessed on 14 January 2026).
36. 3GPP. *Study on New Radio (NR) to Support Non-Terrestrial Networks*; Technical Report TR 38.811, V15.4.0; 3GPP: Sophia Antipolis, France, 2020. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3524> (accessed on 14 January 2026).
37. 3GPP. *Study on Channel Model for Frequencies from 0.5 to 100 GHz*; Technical Report TR 38.901, V18.0.0; 3GPP: Sophia Antipolis, France, 2024. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3173> (accessed on 14 January 2026).
38. ITU. *Propagation Data Required for the Design Systems in the Land Mobile-Satellite Service*; Technical Report P.681-10; ITU-R: Geneva, Switzerland, 2019. Available online: <https://www.itu.int/rec/R-REC-P.681/en> (accessed on 14 January 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.