remote sensing

MDPI

Article

Intelligent Detection Method for Satellite TT&C Signals under Restricted Conditions Based on TATR

Yu Li [1], Xiaoran Shi [1,*], Xiaoning Wang [1], Yongqiang Lu [2], Peipei Cheng [1] and Feng Zhou [1]

[1] Key Laboratory of Electronic Information Countermeasure and Simulation Technology, Ministry of Education, Xidian University , Xi'an 710071, China; yli_1999@stu.xidian.edu.cn (Y.L.); 21021211157@stu.xidian.edu.cn (X.W.); ppei.cheng@stu.xidian.edu.cn (P.C.); fzhou@mail.xidian.edu.cn (F.Z.)
[2] State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
* Correspondence: xrshi@xidian.edu.cn

Abstract: In complex electromagnetic environments, satellite telemetry, tracking, and command (TT&C) signals often become submerged in background noise. Traditional TT&C signal detection algorithms suffer a significant performance degradation or can even be difficult to execute when phase information is absent. Currently, deep-learning-based detection algorithms often rely on expert-experience-driven post-processing steps, failing to achieve end-to-end signal detection. To address the aforementioned limitations of existing algorithms, we propose an intelligent satellite TT&C signal detection method based on triplet attention and Transformer (TATR). TATR introduces the residual triplet attention (ResTA) backbone network, which effectively combines spectral feature channels, frequency, and amplitude dimensions almost without introducing additional parameters. In signal detection, TATR employs a multi-head self-attention mechanism to effectively address the long-range dependency issue in spectral information. Moreover, the prediction-box-matching module based on the Hungarian algorithm eliminates the need for non-maximum suppression (NMS) post-processing steps, transforming the signal detection problem into a set prediction problem and enabling parallel output of the detection results. TATR combines the global attention capability of ResTA with the local self-attention capability of Transformer. Experimental results demonstrate that utilizing only the signal spectrum amplitude information, TATR achieves accurate detection of weak TT&C signals with signal-to-noise ratios (SNRs) of −15 dB and above (mAP@0.5 > 90%), with parameter estimation errors below 3%, which outperforms typical target detection methods.

Keywords: satellite TT&C signal detection; complex electromagnetic environment; attention mechanism; Transformer; Hungarian algorithm

check for updates

Citation: Li, Y.; Shi, X.; Wang, X.; Lu, Y.; Cheng, P.; Zhou, F. Intelligent Detection Method for Satellite TT&C Signals under Restricted Conditions Based on TATR. *Remote Sens.* **2024**, *16*, 1008. https://doi.org/10.3390/rs16061008

Academic Editors: Benoit Vozel and Stephan Havemann

Received: 29 January 2024
Revised: 9 March 2024
Accepted: 10 March 2024
Published: 13 March 2024

Copyright: © 2024 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) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, with advancements in wireless communication technology, satellite communication has gained widespread application in both military and civilian sectors due to its abundant spectrum resources, wide coverage, and freedom from geographical constraints [1,2]. However, the continuous development of communication frequency bands, the emergence of new modulation techniques, and the proliferation of unidentified systems and proprietary protocols have resulted in an increasing number of signals. Furthermore, complex terrestrial and celestial electromagnetic environments present various challenges such as frequency and time-selective fading, dynamic noise, and interference [3,4]. These challenging channel conditions pose severe difficulties for the detection [5] and analysis of satellite telemetry, tracking, and command (TT&C) signals under non-cooperative conditions.

Under the premise of receiving complete signals, signal restoration can be achieved through techniques such as signal adaptive reconstruction [6], time-domain equalization [7], etc., enabling high-precision detection. However, in practical reception, the equipment side necessitates swift signal transmission for real-time analysis. Due to limitations in

non-cooperative information or constraints on the data volume of intermediate frequency signals, real-time processing often only obtains the spectral amplitude of the signal, with missing phase information. Under the condition of utilizing solely spectral amplitude information, signal processing and restoration methods based on time-domain sequences are almost infeasible. Existing solutions addressing this challenge remain limited, as traditional time-domain processing methods exhibit significant limitations under severe restrictions on phase information.

For classical detectors based on deep learning, the framework adjustment from general target detection to signal detection primarily involves the adaptation of prior anchors' sizes to accommodate signals through statistical or clustering methods. However, the aspect ratio variations in signals in spectrograms are substantial, making it challenging for even well-designed prior anchors to match all signals. When predicting the signal positions for candidate regions, the model generates multiple prediction boxes, which inevitably raises the question of how to eliminate redundant predictions. Most approaches rely on post-processing steps, such as non-maximum suppression (NMS) [8] or its related improved algorithms [9], to filter and retain the most probable box among multiple potential predictions for the same target by setting a threshold. However, there is significant variation in the aspect ratios of signals in the spectrogram, making it challenging for well-designed prior anchors to match different signals. Furthermore, incorporating the NMS data post-processing step complicates the detector and hinders the optimization and fine-tuning of the entire model. Even a well-performing trained model faces hardware compatibility issues with the NMS algorithm during practical deployment, resulting in poor transferability and usability. The above defects result in a decrease in the performance of classical detectors when used for signal detection.

Therefore, this paper aims to address the task of satellite TT&C signal detection under constrained conditions by proposing a TATR-based intelligent satellite signal detection algorithm. In the scenario where only the signal spectrum amplitude information is available, we first convert the 1D signal amplitude sequence into a 2D spectrogram image to facilitate the correlation of signal frequency and spatial features. Based on Triplet Attention [10] and the ResNet50 network, we designed the ResTA backbone for extracting spectrum spatial features. To enhance the interaction capability of spectral context, we introduce the Transformer network to associate the local frequency information of signal spectra. Meanwhile, the ResTA backbone, while adaptively correlating the global spatial features of spectra, also contributes to the Transformer's capture of local contextual information, achieving lightweight transformation of the Transformer. Finally, by utilizing the Hungarian algorithm [11], we transform signal detection into a set prediction problem, eliminating the need for reference box setting based on prior knowledge and the post-processing step of NMS. TATR is capable of directly and parallelly outputting the coordinates and signal types of predicted boxes, without the need for expert feature information and manual threshold setting.

In conclusion, this paper makes the following key contributions:

1. For signal detection challenges under restricted conditions with incomplete phase information. In contrast to traditional detection methods employing 1D signal sequences as inputs, we transform the 1D sequences into 2D spectrogram images, providing a visual representation of the distinctions between the amplitude envelopes of telemetry signal spectra and background signals. This conversion shifts the frequency prediction problem based on sequences into an object detection problem based on images.
2. We design a ResTA backbone based on residual structure and triplet attention, which can correlate the features of the channel, frequency, and amplitude dimensions within the spectrogram. ResTA enhances the capability of extracting spectral features almost without introducing additional parameters.
3. We propose a novel signal detection model TATR based on ResTA and Transformer. TATR combines the global attention capability of ResTA and the local attention mechanism of self-attention in the Transformer to capture both global and local features of

the spectrogram. Furthermore, it reduces the number of parameters in the Transformer model while adaptively selecting optimal spectral features of TT&C signals.

4. Setting fixed anchor points in advance is not suitable for the dynamic nature of TT&C signals' electromagnetic environments. Therefore, we employ bipartite graph matching in the detection phase, eliminating the necessity for setting anchor points based on prior knowledge, which converts the signal detection problem into a set prediction problem, and the Hungarian algorithm is applied to achieve anchor-free signal detection.

2. Related Works

2.1. Traditional Signal Detection

Traditional signal detection methods can be broadly categorized into four types: feature-based detection, energy-based algorithms, matched filtering, and cyclostationary detection. Energy-based algorithms [12,13] quantify the overall energy of the received signal over a specific time interval, estimate the noise variance, and compare it to a predetermined decision threshold to ascertain signal presence. Although the energy-based method operates without requiring prior knowledge and exhibits rapid detection, it falls short in accurately detecting signals amid substantial noise or interference. The basic idea of feature-based detection [14] is to calculate the eigenvalues of the signal's covariance matrix and use the ratio of the maximum and minimum eigenvalues as the threshold of the test statistic for signal detection. The accurate determination of the threshold is crucial for the success of feature-based detection algorithms. Matched filtering [15,16] constructs a matched filter based on the amplitude–frequency response of the signal to be detected. In cases where prior information about the signal is accessible, and the background noise adheres to a Gaussian distribution, matched filtering frequently demonstrates enhanced detection capabilities. Cyclostationary detection [17] utilizes the cyclostationary characteristics of a signal, such as symbol rate, carrier frequency, and sampling frequency, to detect the presence of a signal. It can also estimate certain modulation parameters [18]. However, this method has high computational complexity and poor real-time capabilities.

Feature-based detection, matched filtering, and cyclostationary detection methods achieve pleasing results under conditions of complete information. However, the performance of these time-domain signal-processing-based methods sharply declines under restricted conditions. Although energy detection can alleviate this problem under certain conditions [19,20]. As the variety of interference and clutter types grows, devising appropriate threshold levels becomes intricate, especially when the signal energy is weaker than those of interference and noise. Consequently, there is a pressing necessity to investigate intelligent detection methods.

2.2. Deep-Learning-Based Signal Detection

Currently, deep learning models for object detection in computer vision can be broadly categorized into two types. The first is a two-stage model based on region recommendation, exemplified by region-based convolutional neural networks (R-CNNs) [21], fast region-based convolutional neural network (Fast-RCNN) [22], faster region-based convolutional neural network (Faster-RCNN) [23], etc. These algorithms follow a two-step process: firstly, they extract candidate regions in the global sliding window and determine whether these regions are foreground or background; secondly, they classify the targets within the region. The second type consists of regression-based one-stage models like Single Shot Multi-Box Detector (SSD) [24], You Only Look Once (YOLO) [25], YOLOv2 [26], and subsequent YOLO versions [27,28].

The rapid advancement in deep learning has injected new vitality into signal detection and recognition [29–31]. Ke et al. [32] employ convolutional long short-term deep neural networks (CLDNNs) cascaded with convolutional neural networks (CNNs) and long short-term memory (LSTM) to extract time-domain and frequency-domain features from input signal sequences. which gives better performance than traditional energy detection

algorithms. This type of algorithm can directly extract features from preset candidate regions, with fast detection speed but relatively low accuracy. Prasad et al. [33] utilize the Fast-RCNN object detection model and achieve an average signal detection rate of 90% based on broadband signal time–frequency maps. However, due to the two-stage process involved in this type of algorithm, its real-time performance is suboptimal. Li et al. [29] divide the entire signal frequency band into broadband segments, obtain time–frequency maps using short-time Fourier transform, and propose a CNN-based broadband signal detection method, successfully detecting and recognizing six types of signals. Additionally, Xu [34] introduces a deep learning method based on YOLO to automatically and accurately pick out tweek signals from VLF measurements. The aforementioned deep-learning-based methods typically require time-domain IQ sequences or their transformed counterparts, such as time–frequency spectrograms, as inputs to deep neural networks. However, this is difficult to meet under the condition of limited phase information.

Most object detectors locate objects through internal pixels in the image, and it is difficult for the receptive field of CNN to cover signals of a long duration. Transformers [35] have undergone significant development, emerging as a popular research topic. Initially employed in natural language processing (NLP), they have proven their ability to capture local information based on self-attention. The success of the Vision Transformer (ViT) [36] in image recognition tasks has demonstrated the substantial potential of Transformers in computer vision (CV). With ongoing research, numerous Transformer models tailored for object detection tasks have emerged. For instance, Carion et al. [37] defined object detection as a set prediction problem and introduced a novel framework named Detection with Transformer (DETR). Subsequently, Zhu et al. [38] proposed Deformable DETR, incorporating locally sparse deformable attention modules to address the challenge of poor performance in detecting small objects. Wang et al. [39] presented an anchor-decoupled attention mechanism, referred to as anchor DETR. However, Transformer-based models face challenges of slow algorithm convergence and high computational resource consumption. This is primarily attributed to the need for multiple layers of encoders and decoders to capture global attention, introducing a significant number of parameters. Some researchers have alleviated this issue from the perspective of optimizing data parallelism [40,41], but a more effective approach involves lightweight and optimized modifications to the network itself [42].

3. Problem Definition

Satellite TT&C signals are transmitted by satellites to monitor and transmit data, playing a crucial role in various fields such as aerospace, meteorology, and Earth remote sensing observations. TT&C signals wirelessly transmit information from sensors and monitoring parameters to ground stations or other receiving devices. Detecting TT&C signals is a prerequisite for signal reception and estimation. However, the electromagnetic space is cluttered with various types of noise and interference signals of different frequency bands and mechanisms, intertwining satellite and background signals. Figure 1 illustrates the transmission scenario of the satellite downlink, where the satellite communicates with the ground station through the downlink channel. The ground station, while receiving signals, also encounters intentional or unintentional interference from ground-based sources. Due to the extended transmission distance and the presence of shading, severe channel attenuation occurs, resulting in a diminished signal-to-noise ratio (SNR). Consequently, the signal is often submerged in complex backgrounds, exhibiting extremely low spectral density characteristics. The blind detection problem of non-cooperative satellite signals, determining the presence of satellite downlink signals from the received raw data, essentially represents a binary hypothesis testing problem:

$$\begin{cases} H_0 : y(t) = \sum_{k=1}^{M} b_k(t) + n(t) \\ H_1 : y(t) = \sum_{i=1}^{N} x_i(t) + \sum_{k=1}^{M} b_k(t) + n(t) \end{cases} \tag{1}$$

where $x_i(t)$ is the TT&C signal, N is the number of signals contained in the original data, $b_i(t)$ is the background signal or interference signal present in the environment, and M is the number of background signals or interference signals present in the environment. $n(t)$is the noise introduced by the channel environment. In this paper, we model it as Gaussian white noise with a distribution function of $f_n(\varsigma) = 1/\sqrt{2\pi{\sigma_n}^2}\exp(-\varsigma^2/2{\sigma_n}^2)$. H_0 indicates that the satellite signal does not exist, with H_1 indicating that the satellite signal is present.

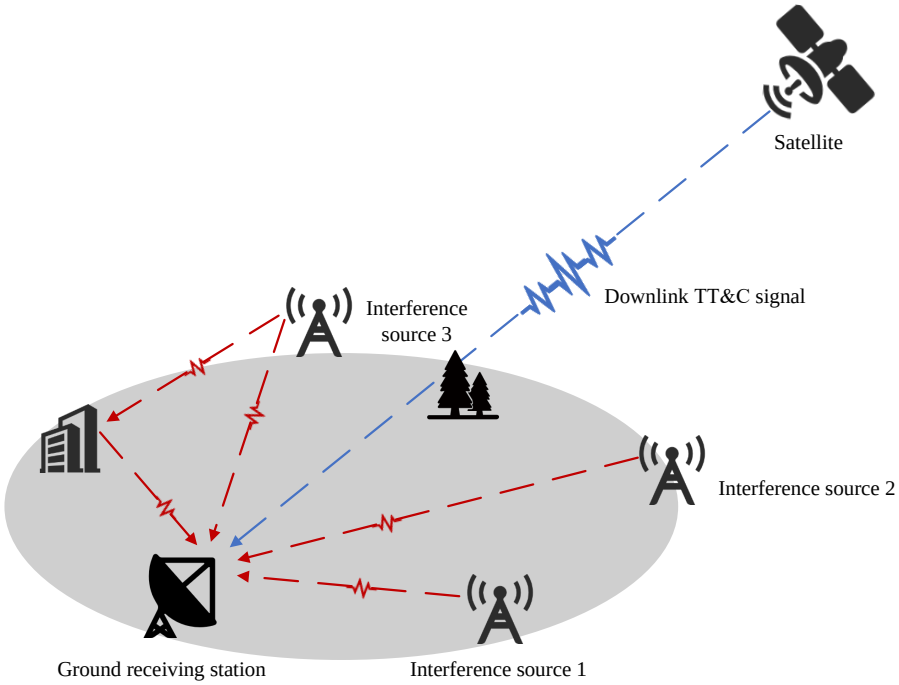

Figure 1. The transmission scenario of the satellite TT&C signal downlink. Ground receiving stations, while receiving downlink TT&C signals, may receive intentional or unintentional interference signals and multipath clutter from ground-based stations. Furthermore, there may be obstacles such as trees and buildings which lead to the submergence of TT&C signals within the ambient signal environment.

Traditional time-domain detection methods construct test statistics based on observed values of $y(t)$, seeking differences between $x(t)$ and $b(t)$ in various representation domains, setting thresholds to make decisions and obtain results. However, due to limitations in non-collaborative information or constraints on the data volume of intermediate frequency signals, practical processing or real-time monitoring often only provides amplitude information of the signal spectrum, lacking phase information. Therefore, under restricted conditions, based on the amplitude sequence of the spectrum of $y(t)$, the detection problem can be modeled as

$$\hat{H} = \arg\max_{j\in\{0,1\}} P\left(H_j \mid |DFT(y(t))|\right) \tag{2}$$

where $\hat{H}$ represents the detection result, and $||$ represents the modulus operation. This problem is easily solvable using deep learning methods. Deep neural networks, as data-driven models, are well-suited for handling complex, nonlinear relationships and patterns. The black box model $\mathcal{F}$ within deep neural networks can be seen as a mapping function between inputs and outputs:

$$\hat{H} = \mathcal{F}(|DFT(y(t))|) \quad \mathcal{F}: \mathbb{R}^d \to \mathbb{R}. \tag{3}$$

By training on a large amount of labeled data, deep neural networks capture the relationship between the received signal spectrum and the TT&C signal spectrum fea-

tures, uncovering the distinguishability between spectrum envelopes (discussed in the next section), thereby determining the presence and number of signals and subsequently identifying the class of each TT&C signal $x_i(t)$.

4. Signal Model

TT&C signals in satellite downlinks often employ composite modulation types, undergoing multi-level modulation of information. Among them, PCM-BPSK-PM is a commonly used type of command and measurement signal [43]. The initial information undergoes sequential pulse code modulation (PCM) encoding, binary phase-shift keying (BPSK) modulation, and outer phase modulation (PM), resulting in the following expression:

$$s_{PCM-BPSK-PM}(t) = A_c e^{j[2\pi f_c t + K_p s_{BPSK}(t)]} \tag{4}$$

where A_c represents the main carrier amplitude, f_c is the main carrier frequency, and K_p is the phase modulation sensitivity, indicating the phase shift in the PM signal caused by unit modulation signal amplitude.

s_{BPSK} represents the inner BPSK signal, where BPSK modulation adds the original information to the phase, and the carried information is represented through phase changes. The time-domain expression of a BPSK signal is as follows:

$$s_{BPSK}(t) = A_B \cos[2\pi f_B t + \varphi] \tag{5}$$

The PCM-BPSK-PM signal has PM modulation as the main carrier and BPSK modulation as the subcarrier. The main carrier is primarily used for detection, while the subcarrier is mainly employed to convey information for communication. Figure 2 depicts the spectrograms of the satellite downlink PCM-BPSK-PM signal under different phase modulation sensitivities. The envelope illustrates the spectral characteristics of the inner modulation signal. Whether these features are distinct is closely related to the phase modulation sensitivity. When the modulation sensitivity is small, the spectrogram shows only a particularly strong carrier component. When the modulation sensitivity is large, the spectrogram exhibits both the inner modulation signal and the carrier frequency of the inner modulation signal, along with symbol rate information.

Ground receiving equipment needs to choose a suitable data transmission method based on the requirements, balancing data volume, real-time performance, and the level of detail in the signal information. The original data quantity of intermediate-frequency signals is substantial, making real-time transmission impractical. Therefore, in scenarios requiring fast detection and analysis, typically only the spectrum amplitude information of TT&C signals can be utilized.

In addition, the downlink channel, often subject to losses like shadowing and multipath fading, results in a decline in SNR. Background interference signals, whether intentional or unintentional, may possess significantly higher energy than the satellite TT&C signal when close to the ground receiving station. Figure 3 illustrates a schematic spectrum received by a ground station in a complex scenario, displaying a wideband spectrum of signals, where the frequency range is from 2100 MHz to 2400 MHz. The scene includes persistent signals from multiple sources, such as 3G and 4G base stations, as shown in Table A1 of Appendix A, making the TT&C signal nearly submerged in the spectrogram. Notably, 3G and 4G signals commonly employ code-division multiple access (CDMA) or orthogonal frequency-division multiple access (OFDMA) [44] technologies, with a wide bandwidth ranging from a few to tens of megahertz. These 3G and 4G signals are reflected in the spectrogram as continuous multi-carrier spectral bands. In contrast, TT&C signals have a narrower bandwidth and often appear as discrete, concentrated spectral bands in specific frequency ranges. As a result, TT&C signals demonstrate modulation characteristics that distinguish them from background signals in the spectrogram. Effective detection can be achieved by observing the spectrogram of the received signal and extracting envelope features.

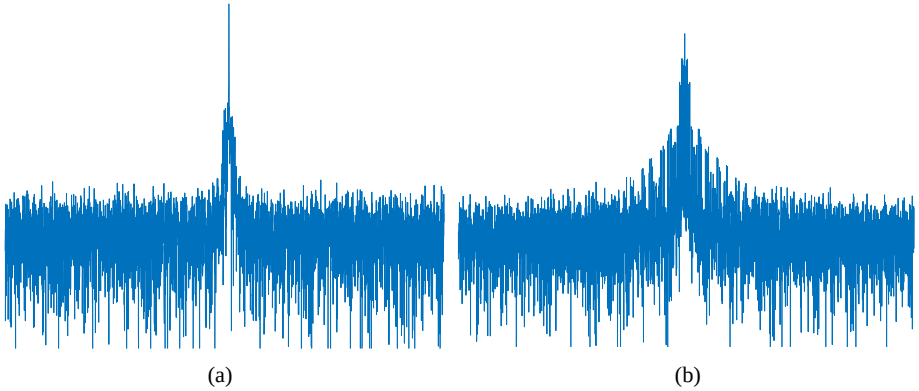

Figure 2. PCM-BPSK-PM signal spectrogram characteristics of satellite TT&C signals under different modulation sensitivities. (**a**) $K_p = 0.1$. (**b**) $K_p = 1$. The spectral characteristics of the inner modulation signal hinge on the phase modulation sensitivity. Low sensitivity yields a spectrum dominated by a strong carrier component, while high sensitivity reveals the inner modulation signal, its carrier frequency, and symbol rate information.

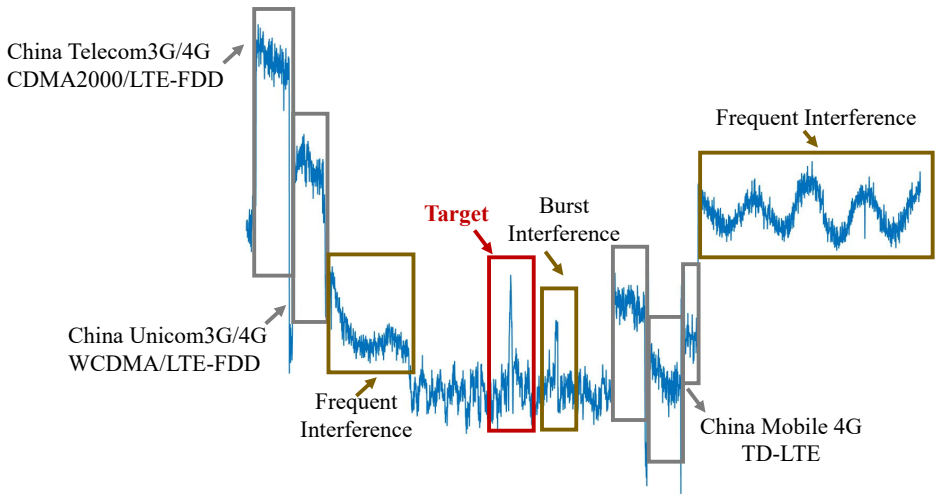

Figure 3. Spectrogram of satellite downlink signal reception in complex scenarios, which contain diverse 3G/4G signals as well as burst and frequent interference signals.

5. Methods

As depicted in Figure 2, TT&C signals exhibit modulation characteristics on the spectrogram amplitude envelope in a manner distinct from background signals. Under conditions where phase information is constrained, careful utilization and exploration of these features contribute to signal detection and identification. Therefore, departing from traditional signal detection algorithms, we transform the 1D amplitude sequence into a 2D spectrogram, which provides an intuitive representation of envelope variations. Such 1D spectrum amplitude sequences contain only signal frequency information, making it challenging to identify envelope changes. In contrast, 2D spectrogram images not only incorporate frequency contextual information but also reveal spatial information about the amplitude envelope. This spatial information is crucial for uncovering modulation features on the envelope.

The overall framework of the proposed TT&C signal detection network TATR under restricted conditions is illustrated in Figure 4. Initially, we design the ResTA backbone based on a residual structure and triplet attention mechanism. By extracting deep features, the ResTA backbone captures global spectral attention through a multi-level triplet attention. Subsequently, drawing inspiration from the architecture of DETR [37], we integrate signal position embedding with the features extracted by the backbone, and feed them into the

TATR encoder–decoder module. This module employs a multi-head attention mechanism to correlate contextual local spectral information. The decoder module establishes pairing relationships between the queried signal objects and the encoded features, producing features of the detected signal set of interest. Finally, employing the Hungarian algorithm, our signal detection module is designed as an anchor-free framework. Through bipartite graph matching, we obtain signal detection results.

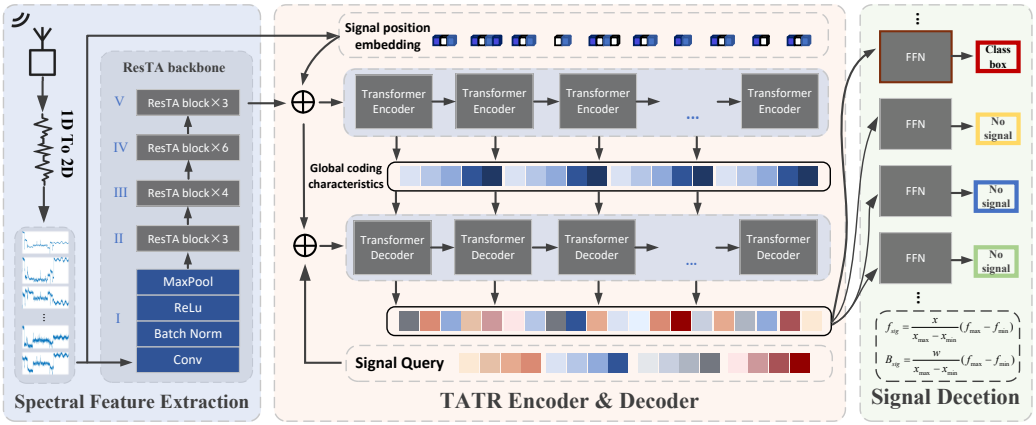

Figure 4. The overall network structure of TATR. This network consists of three parts: ResTA backbone network for spectral feature extraction; neck with multilayer encoders and decoders; and signal detection head block, including class loss and bounding box loss. Firstly, the 1D spectrum amplitude sequence is transformed into a 2D spectrogram and fed into the ResTA backbone. Subsequently, the position embedding of the spectrogram is jointly utilized as input for the TATR encoder and decoder. Finally, the output of the TATR decoder undergoes FFN mapping to derive the positional coordinates and parameter information of the signal.

5.1. ResTA Spectrum Feature Extraction Network

Detection models based on deep learning typically utilize CNNs as a backbone. However, CNNs, constrained by the size of convolutional kernels, often face challenges in correlating global information, which is detrimental for tasks requiring full-spectrum detection. To address this, we integrate the triplet attention mechanism [10] into the ResNet50 [45] and propose the ResTA backbone, which serves as the spectral feature extraction network for TATR. ResTA globally correlates the height, width, and channel dimensions of features, divided into five parts. The first part, without residual blocks, primarily performs convolution, normalization, activation, and max-pooling operations on the input. The second, third, fourth, and fifth parts feature stacked ResTA blocks. The ResTA block structure, illustrated in Figure 5, involves three convolutional layers followed by the triplet attention module to capture spectral attention. Additionally, it incorporates a residual connection mechanism across layers to mitigate the impact of gradient vanishing.

The triplet attention module, as illustrated on the right-hand side of Figure 5, computes attention weights by capturing cross-dimensional interactions using a three-branch structure. For an input tensor ϑ, the triplet attention module establishes dependencies between dimensions through rotation operations and residual transformations, encoding information across channels and spatial dimensions with negligible computational overhead. Specifically, triplet attention consists of three parallel branches. The first branch is responsible for capturing cross-dimensional interactions between channel C and spatial dimension H, achieving this by rotating the input spectral feature $\vartheta \in R^{C\times H\times W}$ counterclockwise by 90° along the H-axis, resulting in tensor $\vartheta_r \in R^{W\times H\times C}$, which undergoes a Z-Pool operation to transform into $\hat{\vartheta}_r \in R^{2\times H\times C}$. The Z-Pool operation is represented as

$$Z - Pool(\vartheta_r) = [MaxPool(\vartheta_r), AvgPool(\vartheta_r)] \quad (6)$$

Subsequently, a standard convolution layer with a kernel size of $k \times k$ is applied to the Z-Pool to reduce its W dimension to 1. Finally, the attention weights are generated through a sigmoid activation, and the resulting attention feature $\vartheta_{att-1} \in R^{C \times H \times W}$ from the first branch is obtained by multiplying it by the input spectral feature tensor.

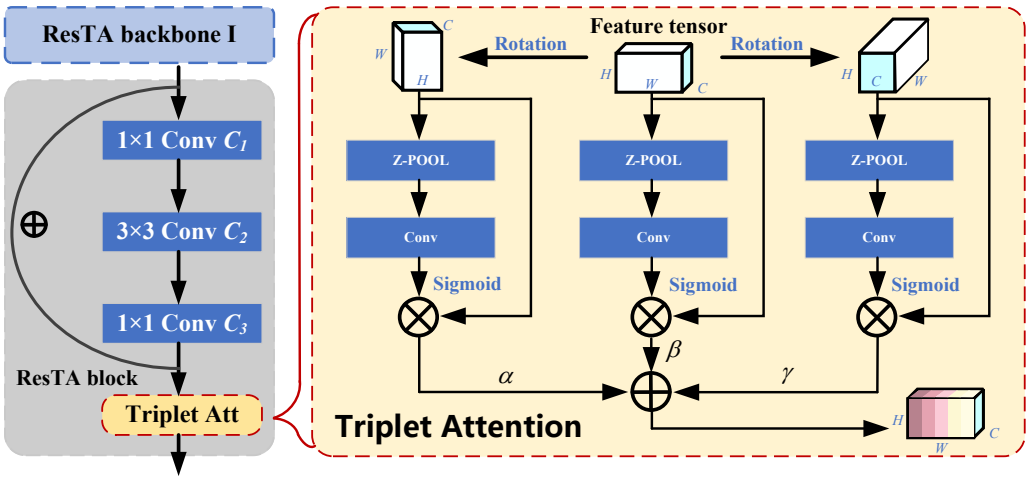

Figure 5. ResNet with triplet attention module (ResTA block). This structure divides the output of each block in ResNet into three channels. These channels then undergo rotation, attention calculation, and stacking to associate signal characteristics across channels, amplitude, and frequency dimensions.

The second branch is responsible for capturing cross-dimensional interactions between channel C and spatial dimension W, while the third branch captures spatial attention between the spatial dimensions H and W. They go through Z-Pool operations, standard convolution layers, and sigmoid activations to generate attention weights. The resulting attention features, denoted as $\vartheta_{att-2} \in R^{C \times H \times W}$ and $\vartheta_{att-3} \in R^{C \times H \times W}$, are obtained by multiplying them by the corresponding input tensors.

The output of the triplet attention module is obtained by weighting the attention tensors from the three branches and is expressed as

$$\vartheta_{att} = \alpha \vartheta_{att-1} + \beta \vartheta_{att-2} + \gamma \vartheta_{att-3} \tag{7}$$

The weighted aggregation of outputs from the three branches is conducted, introducing attention mechanisms into the network. The multilayer stacking of triplet attention enables ResTA to adaptively optimize features, expanding the receptive field with almost no introduction of additional parameters. ResTA correlates the three dimensions of channels, frequencies, and amplitudes in the spectrogram, selectively processing spectrogram features. Features with attention will be beneficial for subsequent signal detection and parameter estimation tasks.

5.2. TATR Encoder and Decoder

The TATR encoder and decoder is established upon the encoder and decoder architecture of the Transformer model [35]. This structure entirely eliminates CNNs and RNNs, opting for self-attention mechanisms for both encoding and decoding to establish connections among local features in the temporal sequence of signal spectrograms. The architecture of the TATR encoder and decoder is depicted in Figure 6.

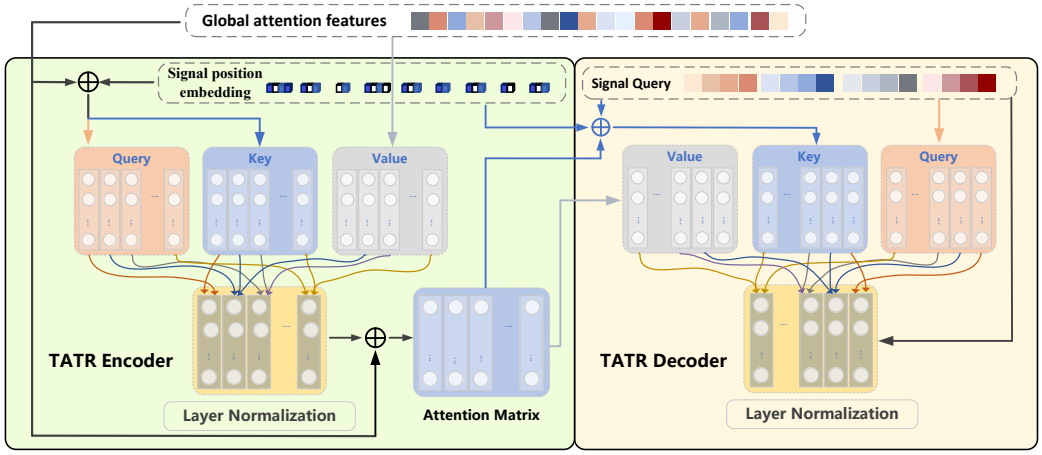

Figure 6. TATR encoder and decoder structure. Structure of the encoder (**left**); structure of the decoder (**right**). The output of the encoder jointly uses the query volume as the input for the decoder.

The TATR encoder module is composed of stacked identical encoders. Each encoder consists of multiple self-attention layers and two feed forward network (FFN) layers. The spectrogram features obtained from the ResTA spectrogram feature extraction network are first processed by a convolutional layer with a kernel size of 1×1, reducing the channel dimension from C to C_1. The spatial dimensions are then folded into a single dimension to obtain a new feature map $F_{1f-img} \in R^{C_1 \times (H_1 \times W_1)}$, which serves as part of the input to the encoder.

However, unlike 1D sequences, the envelope variation information of 2D spectrograms requires guidance on frequency positions. Therefore, we apply positional embedding to the signal to help the network learn the correlation between different frequency positions of the signal and integrate them with global attention features as inputs to the TATR encoder. These inputs are transformed into the required query, keys, and values vectors using three weight matrices: W^Q, W^K, and W^V. Then, a self-attention operation is performed:

$$attention(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{8}$$

where softmax denotes the normalized exponential function, Q represents the query vector, K is the keys vector, V stands for the values vector, and the dot product of Q and the transpose of K yields the attention scores for each word vector. d_k is the dimensionality of the keys vector.

As every vector in the input signal features undergoes self-attention, forming a direct connection between any two vectors, the self-attention mechanism can learn the correlation between frequency points in TT&C signals in a wideband reception scenario. The overall architecture is similar to the encoder, with the difference that the decoder adds a masked-multi head self-attention layer, where the mask ensures that the prediction at position i depends only on the outputs before position i. In signal detection, the decoder outputs results through an autoregressive process. The structure of the TATR decoder is depicted on the right-hand side of Figure 6.

The decoder receives self-attention vectors from the encoder as the values vector for the decoder layer. The combination of satellite signal position encoding and self-attention vectors serves as the keys vector, and the encoded satellite signal detection box acts as the query vector. The self-attention operation on the query, keys, and values vectors is expressed in Equation (8). Similar to the encoder, the result goes through a residual connection, followed by layer normalization, which calculates the mean and variance across different channels for each sample to address the issues of gradient vanishing and weight matrix degradation. The multi-head self-attention mechanism facilitates the network in automatically selecting relevant contextual information in latent space for

detection purposes (this will be empirically demonstrated and extensively discussed in the experimental section). In addition, the inclusion of position-encoded spectral features aids the subsequent decoder in accurately locating the signal position.

5.3. Signal Detection

The features obtained from the decoder module undergo global and local adaptive feature selection, mapping through the feed forward network (FFN) to a higher-dimensional feature space. Class labels and the coordinates of detection boxes are predicted through an FFN layer with shared weights. The output dimension of the FFN layer is $D \times 5$, providing four coordinate predictions and one class prediction, where D represents the number of set signal query vectors. The traditional post-processing step, based on the NMS algorithm, relies on manually set anchor points [46] and retains the box with the highest probability among D results using a fixed threshold [47]. This approach is unsuitable for complex and dynamic signal broadband reception scenarios. In this paper, the Hungarian algorithm [11] is employed for bipartite graph matching of signal detection boxes, as illustrated in Figure 7. Where $\mathcal{O}_i$ represents the i-th signal detection box predicted by TATR, $\mathcal{T}_i$ is the i-th ground truth detection box, and $\omega_{ij} \in 0, 1$ indicates whether the predicted box $\mathcal{O}_i$ and the true box match. The Hungarian algorithm is used to establish a one-to-one correspondence between the predicted results and the labels to minimize the matching loss. The optimal matching of signal prediction boxes is found by recursively changing the corresponding relationship. The calculation of the matching loss is shown in Equation (9):

$$\begin{aligned} \hat{\sigma} &= \arg\min_{\sigma \in D} \sum_{i}^{D} L_m(\mathcal{O}_i, \mathcal{T}_{\sigma(i)}) \\ L_m(\mathcal{O}_i, \mathcal{T}_{\sigma(i)}) &= -1_{\{c_i \neq \varnothing\}} \hat{p}_{\sigma(i)}(c_i) + 1_{\{c_i \neq \varnothing\}} L_{box}(s_i, \hat{s}_{\sigma(i)}) \end{aligned} \tag{9}$$

where $\hat{\sigma}$ represents the paired matching indices obtained by minimizing the matching loss through the Hungarian algorithm, and $L_m(\mathcal{O}_i, \mathcal{T}_{\sigma(i)})$ represents the matching loss between each predicted signal detection box and the padding detection box of the true label. The loss consists of two parts: the bounding box loss $1_{\{c_i \neq \varnothing\}} L_{box}(s_i, \hat{s}_{\sigma(i)})$ minus the category loss $1_{\{c_i \neq \varnothing\}} \hat{p}_{\sigma(i)}(c_i)$.

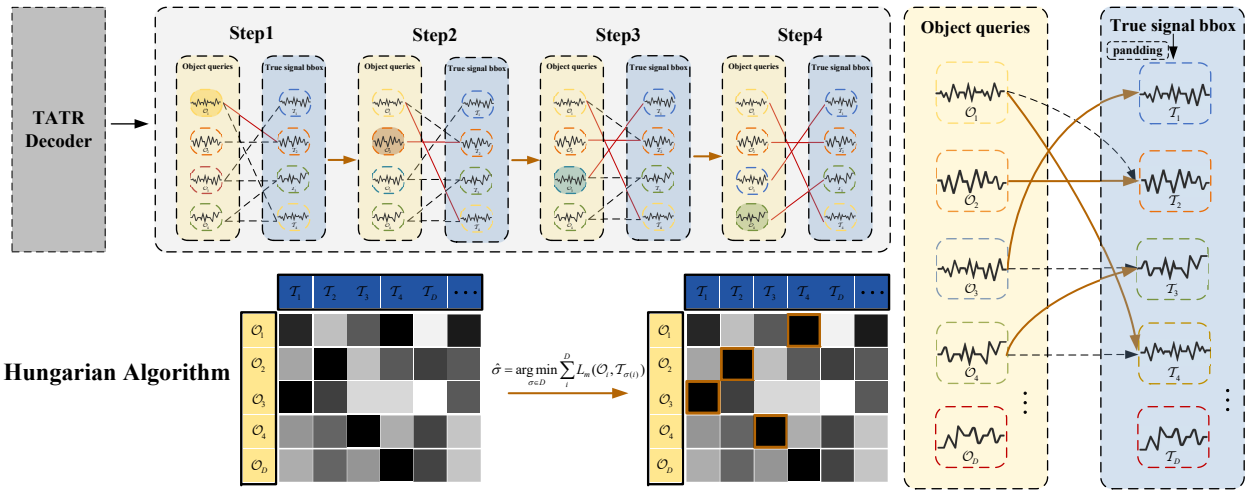

Figure 7. Signal prediction box matching using Hungarian algorithm. The features output by TATR decoder are processed through FFN to obtain the predicted signal box, which is then matched with the actual detection box padding results using the Hungarian algorithm to obtain the minimum matching loss result.

The matching cost considers both class prediction and the similarity between predicted boxes and true boxes, where c_i is the signal class label, $c_i = \varnothing$ represents the case where the

predicted signal class is not considered as background, $\hat{p}_{\sigma(i)}(c_i)$ represents the predicted probability for each signal class c_i, and $s_i \in [0,1]^{d=4}$ represents the center coordinates of the true signal box ($\mathcal{O}_i^x$ and $\mathcal{O}_i^y$), as well as its height h_i and width w_i.

After obtaining the optimal match $\hat{\sigma}$ for the signal prediction box, the loss function of the TATR model is calculated and consists of two parts: the negative log-likelihood loss for signal category prediction L_{pre} and the bounding box loss L_{box}.

The negative log-likelihood loss for signal category prediction is the same as the cross-entropy loss. It measures the goodness of fit between the model and the training data, maximizing the likelihood estimation to obtain a model that is closest to the data distribution, which is defined as

$$L_{pre} = -\log \hat{p}_{\hat{\sigma}(i)}(c_i) \tag{10}$$

The bounding box loss measures the difference between predicted bounding boxes and true bounding boxes. To alleviate the sensitivity of the l_1 loss to different predicted bounding box sizes, the bounding box loss is a weighted sum of the l_1 loss and the generalized intersection over union (GIoU) loss [48] function. The bounding box loss is defined as

$$L_{box}\left(s_i, \hat{s}_{\sigma(i)}\right) = \lambda_{iou} L_{iou}\left(s_i, \hat{s}_{\sigma(i)}\right) + \lambda_{L1}\left\|s_i, \hat{s}_{\sigma(i)}\right\|_1 \tag{11}$$

where λ_{iou} and $\lambda_{L1} \in R$ are the weights for the GIoU loss function and l_1 loss function, respectively. s_i represents the true labels of the bounding boxes, and $\hat{s}_{\sigma(i)}$ corresponds to the predicted box results under optimal matching conditions. The bounding box loss L_{iou} using the GIoU loss function is defined as

$$L_{iou}\left(s_i, \hat{s}_{\sigma(i)}\right) = 1 - \left(\frac{\left|\hat{A}_{\sigma(i)} \cap A_i\right|}{\left|\hat{A}_{\sigma(i)} \cup A_i\right|} - \frac{\left|B\left(s_i, \hat{s}_{\sigma(i)}\right) \backslash \hat{A}_{\sigma(i)} \cup A_i\right|}{\left|B\left(s_i, \hat{s}_{\sigma(i)}\right)\right|}\right) \tag{12}$$

where $\hat{A}_{\sigma(i)}$ and A_i represent the boundaries of the predicted and true bounding boxes, respectively, and $B(\cdot,\cdot)$ denotes the area of the minimum bounding rectangle of the two rectangles. The smaller the overlap area of the two bounding boxes, the larger the GIoU loss.

The set results predicted by TATR need to undergo a reverse mapping operation to restore the parameters of the TT&C signal for detection. The mapping from the FFN prediction results to signal parameters is given by

$$\begin{aligned} f_s &= \frac{\mathcal{O}^x}{\mathcal{O}^x_{\max} - \mathcal{O}^x_{\min}}(f_{\max} - f_{\min}) \\ B_s &= \frac{w}{\mathcal{O}^x_{\max} - \mathcal{O}^x_{\min}}(f_{\max} - f_{\min}) \end{aligned} \tag{13}$$

where f_s and B_s represent the center frequency and bandwidth of the satellite signal to be detected, x and w are the normalized center abscissa and width from the FFN-predicted detection box results. Different from previous detection methods, using bipartite graph matching allows us to train the entire network end-to-end without the need for pre-setting anchors for non-maximum suppression, simplifying the signal detection process.

6. Experiment and Results

6.1. Sat_SD2023 Dataset

The dataset Sat_SD2023 is a satellite signal detection dataset proposed by mapping the simulated environment of the actual background based on the recently received satellite data in a certain location; it comprises 2100 simulated signal spectrograms. The distribution range of the SNR is −15 dB to 15 dB, with an interval of 5 dB. The simulation experiments in the MATLAB environment utilized parameters consistent with those observed in the real environment. The constant presence of wireless 3G and 4G signals in the background

environment was simulated using the MATLAB Communications Toolbox. The channel environment considered Gaussian noise and Rayleigh fading. The parameters for the dataset simulation and experiments are detailed in Table 1.

Table 1. Dataset and experimental parameters.

TT&C Signal Simulation Parameters			
Signal modulation type	BPSK, QPSK, PCM-BPSK-PM, PCM-QPSK-PM		
Signal center frequency (MHz)	2200–2240		
Bandwidth (MHz)	3–30		
Channel	Gaussian Noise and Rayleigh Fading		
SNR (dB)	−15–15		
Number of satellite telemetry signals	0–1		
Background Signal Simulation Parameters			
Spectrum size	3500 × 2625		
Frequency range (MHz)	2100–2400		
Background environmental signal	3G/4G signal, etc.		
3G signal type	WCDMA	CDMA2000	TD-SCDMA
Chip rate (Mchip/s)	3.8	3.68	1.28
Frame length (ms)	10	25	10
Time slot	15	15	15
Data modulation	QPSK	QPSK	QPSK
Channel width (MHz)	5	5	1.6
4G signal	TD-LTE		
Carrier bandwidth (MHz)	20		
Subcarrier bandwidth (MHz)	15		
Frame length (ms)	10		
Time slot configuration	10:2:2		
Data modulation	OFDM		
Neural Network Hyperparameters			
Training epoch	200		
Batch size	200		
Initial learning rate	1×10^{-4}		
Learning rate adjustment strategy	StepLR		
Optimizer	Adam		

For Sat_SD2023 datasets, we manually drew the ground truth bounding boxes using Labelme v5.1.0 (available at https://gitcode.com/wkentaro/labelme/overview, accessed on 15 November 2022). The annotated files were then converted into the PASCAL VOC format [49] for ease of validation. Our experiments divided the entire dataset into three parts, allocating 60% for the training dataset, 20% for the testing dataset, and 20% for the validation set used during inference detection. The training and testing sets were used to train the model, while the validation set was used to evaluate the performance of the model. As for the validation set, the spectrograms were not labeled but manually inspected in order to evaluate the effectiveness of our method. During training, the annotated spectrogram was input into the TATR network, and the network parameters were updated through backpropagation until the training round was reached. The experiments were conducted in the Windows environment, utilizing the PyTorch open-source framework for both training and testing. The GPU was RTX3090 (24 G) from NVIDIA Corporation in Santa Clara, California.

6.2. Evolution and Indicator

To objectively assess the detection performance of the proposed method, we evaluate the algorithm's effectiveness from common evaluation metrics in the field of object

detection, including average recall (AR), mean average precision (mAP), frames per second (FPS), and the estimation errors of signal parameters, denoted as M_s.

Table 2 presents a confusion matrix, elucidating the concepts of true positives (TPs), true negatives (TNs), false negatives (FNs), and false positives (FPs).

Table 2. Signal detection confusion matrix.

Ground Truth\Predicted Value	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

Average recall [50], employed to assess the network's miss detection rate, is the probability of correctly detecting a target (TP) among all prospective targets to be detected (TP and FN). The calculation formula is given by

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{14}$$

Precision [50] is the probability that the correctly detected target (TP) accounts for all detected targets (TP and FP). The calculation formula is

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{15}$$

Average precision [51], the area under the precision–recall (PR) curve, as shown in (16), is calculated from a curve plotted by precision and recall. Recall serves as the horizontal axis, while precision serves as the vertical axis. A larger area under the PR curve, denoted as a higher AP, indicates higher precision and recall, hence a better overall detection performance of the model.

$$\mathrm{AP} = \int_0^1 \mathrm{P_d}(\mathrm{r})\mathrm{dr} \tag{16}$$

where $\mathrm{P_d}(\mathrm{r})$ represents the PR curve of the detection results.

FPS is used to assess the speed of object detection, representing the number of images that can be processed in one second.

The estimation error in signal parameters M_s is the average normalized estimation error percentage for parameters such as the center frequency f_s and bandwidth B_s of the detected satellite signal. The calculation formula is

$$M_s = \frac{1}{k}\sum_{i}^{k}\left|\frac{\overline{\varepsilon_i} - \varepsilon_i}{\varepsilon_i}\right| \tag{17}$$

where $\bar{\varepsilon}_i$ is the estimated value of each signal parameter, ε_i is the true value, and k is the number of parameters to be estimated.

6.3. Experimental Results and Analysis

To validate the strength of our proposed method, we used the Sat_SD2023 dataset to perform performance validation on the proposed TATR model to demonstrate the effectiveness and superiority of the algorithm proposed in this paper. By evaluating the loss value of the model, the convergence and accuracy of the model during the training process can be determined. We first validate the effectiveness of the anchor-free framework.

6.3.1. Validity Analysis and Ablation Experiments

In the case of training and testing TATR across all SNRs, the bounding box loss and signal classification losses for each detection box are illustrated in Figure 8a. The losses decrease continuously with an increase in training epochs, ultimately converging, which indicates that the Hungarian algorithm is capable of matching the predicted detection

boxes with the ground truth boxes, validating the effectiveness of our proposed anchor-free signal detection framework.

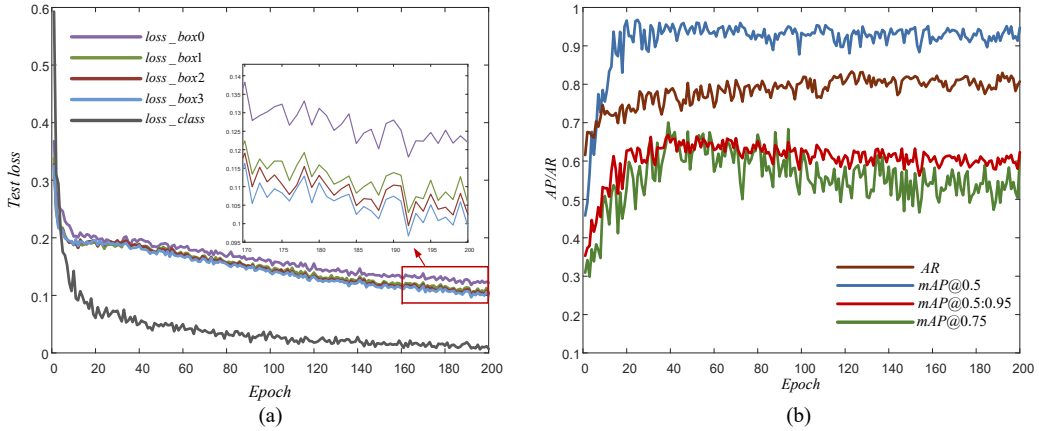

Figure 8. Performance of TATR effectiveness analysis experiment: (**a**) testing loss varies with training rounds, (**b**) testing AP/AR indicators vary with training rounds.

The AR and multiple mAP metrics values show a significant increase with the training epochs, as evident in Figure 8b. TATR achieves an mAP@0.5 surpassing 90% and AR exceeding 80% across all SNRs. Later, we present an ablation analysis to delve into the performance of TATR across various SNRs and the impact of different ratios of labeled training samples.

The performance of TATR under different SNRs is illustrated in Figure 9a, where it is visually apparent that as the SNR increases, both the AR and mAP metrics exhibit continuous improvement. This trend indicates the robustness of our model to changes in SNR, with a notable detection performance of mAP@0.5 (>95%) even at -10 dB, attributed to the global and local attention mechanisms aiding the network in learning robust spectral features. Notably, a sharp decline is observed in mAP@0.75 and mAP@0.5:0.95 when the SNR drops below -10 dB, while the mAP@0.5 metric shows relatively minor fluctuations, suggesting that at -15 dB the network can identify the signal occurrence but struggles to obtain precise bounding box coordinates.

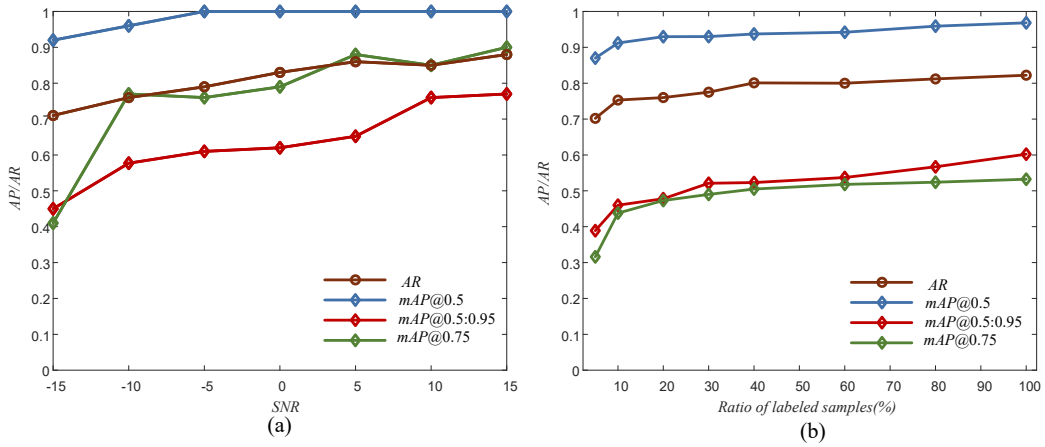

Figure 9. Performance of TATR effectiveness analysis experiment: (**a**) AP/AR metrics vary with SNRs, (**b**) AP/AR metrics vary with the ratio of labeled samples.

Furthermore, the quantity of labeled samples is considered a crucial factor in evaluating performance, as it is challenging to obtain under non-collaborative conditions. We investigated the performance of TATR across all SNRs with a proportion of labeled samples in the training dataset, only $\chi = [5, 10, 20, 30, 40, 60, 80, 100(\%)]$ samples have labels in the

training set. As depicted in Figure 9b, the model's performance improves as the quantity of available labeled samples increases. The performance of the model did not show a significant decrease until $\chi = 5$, indicating that TATR can maintain good performance under different labeled sample sizes.

Subsequently, we conducted ablation experiments on the global and local attention mechanisms of the TATR model, comparing the performance with and without the triplet attention mechanism, as well as altering the number of Transformer encoder and decoder layers (2, 4, and 6 layers). The results are summarized in Table 3. It is evident that the detection performance of TATR improves with an increasing number of encoder and decoder layers. However, along with the increase in encoder and decoder layers, the total parameters and inference time of the network also increase simultaneously, because the multi-head self-attention mechanism reloads keys and values repeatedly during the inference phase, posing a balance issue between speed and accuracy when stacking too many layers.

Table 3. Performance of indicators for global and local attention module ablation in TATR (mAP@0.5 represents average precision at IoU threshold of 0.5, TA Block column represents whether triplet attention (TA) module is added, ✘ represents TA module added, ✔ represents TA module not been added).

Model	TA Block	mAP@0.5 (%)	mAP@0.5:0.95 (%)	mAP@0.75 (%)	AR (%)	Parameters (M)	FPS
TR-2Layer	✘	27.93	19.05	11.24	53.57	29.70	39
TATR-2Layer	✔	35.84	26.86	20.04	62.45	29.70	39
TR-4Layer	✘	92.21	47.37	44.98	73.62	35.49	30
TATR-4Layer	✔	95.78	56.82	50.21	79.67	35.49	30
TR-6Layer	✘	95.81	56.45	51.23	79.81	41.28	19
TATR-6Layer	✔	**96.84**	**60.21**	**53.23**	**82.23**	41.28	19

The ResTA block effectively mitigates this issue. It is evident that, with layers 4 and 6, the results of adding the triplet attention module consistently outperform those relying solely on encoder and decoder for local attention in associating spectrum features. This is attributed to the triplet attention module capturing global attention features during the spectrum feature extraction phase, alleviating the subsequent encoder and decoder's local attention learning overhead.

Notably, the ResTA block introduces almost no additional learning parameters. The 4-layer TATR with the triplet attention module and the 6-layer TATR without the triplet attention module show similar detection performance. However, the former exhibits a reduction of almost 14% in model parameters and a 36% improvement in inference speed. Therefore, the ResTA block in the proposed TATR model can associate global channel, frequency, and amplitude information, enhancing the detection performance of the model across multiple metrics while reducing the model's parameter count.

Furthermore, we analyze the performance of the detection indicators of the TATR model at different SNRs, as shown in Figure 10. Figure 10a–d, respectively, display the curves of the detection metrics mAP@0.5, mAP@0.5:0.95, mAP@0.75, and AR with varying SNRs. Overall, TATR-6Layer achieves nearly the best performance at various SNRs. When the SNR is greater than -5 dB, mAP@0.5 reaches close to 100%, and the AR exceeds 80%, demonstrating the robustness of the proposed TATR model across different SNRs. It is noteworthy that the ResTA block shows a more significant improvement in the model with two and four encoder and decoder layers, with mAP@0.5:0.95 increasing by 4.8–14.7% and 4–15%, respectively. In contrast, on the TR-6Layer model, the improvement is only -1–9.3%, indicating that the global attention capability of the ResTA backbone can extract more effective features, aiding subsequent self-attention modules in obtaining spectrum information and signal position detection. However, the performance of TATR-2Layer is

poor and insufficient to support the capture of spectral features. Therefore, TATR-4Layer and TATR-6Layer may be considered as two model versions for practical deployment.

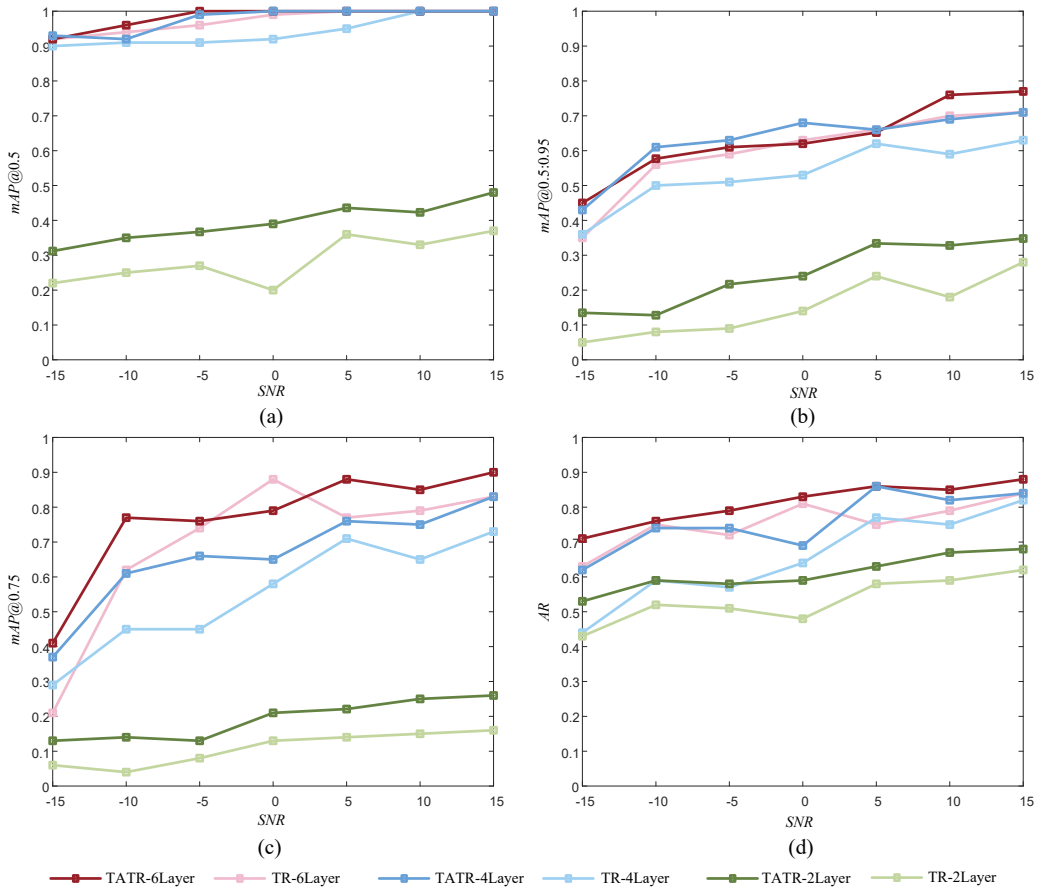

Figure 10. Performance of various indicators for TATR at different SNRs: (**a**) mAP@0.5, (**b**) mAP@0.5:0.95, (**c**) mAP@0.75, (**d**) AR. TR represents the TATR model without the triplet attention (TA) module.

6.3.2. Visualization of Attention Positions

In a detailed analysis of the impact of the global and local attention mechanisms on signal detection, we visualize the attention features of the TATR model during the inference stage. Specifically, we display the feature maps of the last layer of the ResTA backbone and the weighted summation of the last self-attention weights in the encoder and decoder, leveraging position encoding, on the spectrogram. The results are shown in Figure 11. At higher SNRs, the TATR network effectively concentrates attention on the positions where TT&C signals appear, focusing on both the peak positions and the signal envelope within the bandwidth. When the SNR is lower, such as at −15 dB, signal features are almost overwhelmed. TATR manages to capture useful contextual information around the signal spectrum to learn relevant knowledge for detection. By visualizing the abstract-level feature representations learned by the network on the spectrogram, we robustly demonstrate that the proposed TATR, leveraging the global attention capability of triplet attention and the contextual local attention of Transformer, achieves high-precision detection of TT&C signals, enhancing the interpretability of the network.

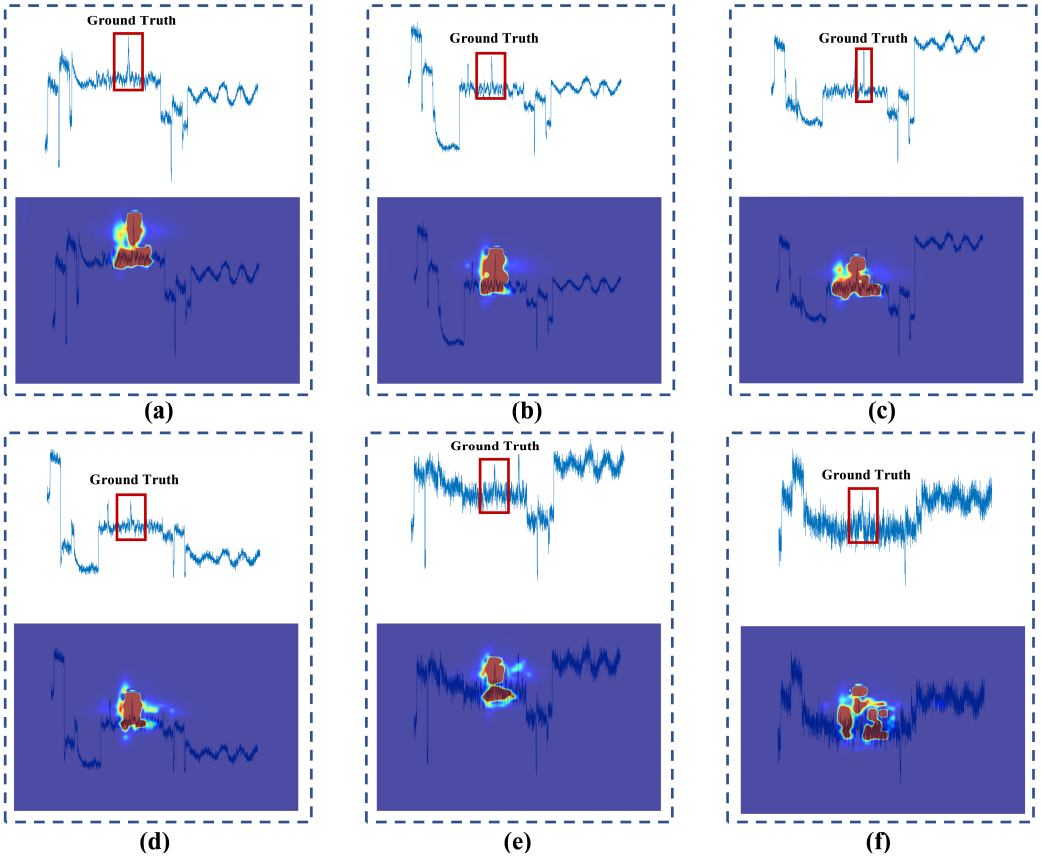

Figure 11. The feature visualization results under various SNRs of TATR: (**a**–**f**) show SNRs from 10 dB to −15 dB in 5 dB decrements, respectively. The red box indicates the true signal location, while the bright areas in the heat map show the key focus of the network after training. At high SNRs, it targets the signal peak, then the network shifts attention to sidelobes and the context envelope in low SNR conditions.

6.3.3. Comparative Experiment

Additionally, to validate the superiority of the proposed method, we compared the detection performance of the TATR with four traditional object detection methods, including Faster-RCNN [23], YOLOv5 [26], YOLOv7 [28], and DETR [37]. The quantitative comparison of performance indicators is presented in Table 4.

According to Table 4, the signal detection performance of the TATR model excels, achieving the best results. YOLOv5s, YOLOv5m, and YOLOv5l represent different versions of the YOLOv5 model, while similarly, YOLOv7, YOLOv7w, and YOLOv7x denote varying versions of the YOLOv7 model, with an increasing number of parameters leading to an overall enhancement in detection accuracy. In comparison to the classical Faster-RCNN (R50) model, the TATR-6layer model exhibits remarkable improvements of 14.03% and 17.74% in the mAP@0.5 and mAP@0.5:0.95 metrics, respectively. Furthermore, compared to DETR, TATR shows a 13.46% enhancement in mAP@0.5:0.95, a 12.98% improvement over the YOLOv5l model, and a 6.59% boost over the YOLOv7x model. The significant performance enhancement primarily manifests in the mAP@0.5:0.95 metric, underscoring the superior precision of the proposed approach in detecting signals with the presence of accurate detections being crucial for signal parameter estimation, thereby substantiating the superiority of the algorithm proposed in this study for signal detection tasks.

Table 4. Comparison of detection performance between TATR and traditional detection models. (MV2 represents the backbone feature extraction network using MobileNetv2, R50 represents the backbone feature extraction network using ResNet50. ↑ represents the improvement of TATR on each model, and ↓ represents the gap of TATR on each model.)

Model	mAP@0.5 (%)	mAP@0.5:0.95 (%)	mAP@0.75 (%)	AR (%)	M_s (%)	Parameters (M)	FPS
Faster-RCNN (MV2)	$86.23_{(\uparrow 10.61)}$	$43.67_{(\uparrow 16.54)}$	$41.29_{(\uparrow 11.94)}$	$80.21_{(\uparrow 2.02)}$	0.049	82.3	17
Faster-RCNN (R50)	$82.81_{(\uparrow 14.03)}$	$42.47_{(\uparrow 17.74)}$	$39.34_{(\uparrow 13.89)}$	$79.65_{(\uparrow 2.58)}$	0.054	41.7	20
YOLOv5s	$94.21_{(\uparrow 2.63)}$	$45.52_{(\uparrow 14.69)}$	$45.14_{(\uparrow 8.09)}$	$81.15_{(\uparrow 1.08)}$	0.044	7.2	53
YOLOv5m	$94.28_{(\uparrow 2.56)}$	$46.63_{(\uparrow 13.58)}$	$45.38_{(\uparrow 7.85)}$	$83.23_{(\downarrow 1.00)}$	0.041	21.2	43
YOLOv5l	$95.19_{(\uparrow 1.65)}$	$47.23_{(\uparrow 12.98)}$	$46.67_{(\uparrow 6.56)}$	$85.24_{(\downarrow 3.01)}$	0.034	46.5	37
DETR	$94.21_{(\uparrow 2.63)}$	$46.75_{(\uparrow 13.46)}$	$45.16_{(\uparrow 8.07)}$	$80.03_{(\uparrow 2.2)}$	0.032	41.3	18
YOLOv7	$95.64_{(\uparrow 1.2)}$	$42.57_{(\uparrow 17.64)}$	$43.25_{(\uparrow 9.98)}$	$83.51_{(\downarrow 1.28)}$	0.046	36.5	14
YOLOv7w	$96.85_{(\downarrow 0.01)}$	$47.85_{(\uparrow 12.36)}$	$46.81_{(\uparrow 6.42)}$	$86.11_{(\downarrow 3.88)}$	0.035	69.8	8
YOLOv7x	$97.12_{(\downarrow 0.28)}$	$53.62_{(\uparrow 6.59)}$	$46.92_{(\uparrow 6.31)}$	$87.80_{(\downarrow 5.57)}$	0.032	70.8	7
TATR-4layer	$95.78_{(\uparrow 1.06)}$	$56.82_{(\uparrow 3.39)}$	$50.21_{(\uparrow 3.02)}$	$79.67_{(\uparrow 2.56)}$	0.031	35.5	30
TATR-6layer	96.84	60.21	53.23	82.23	0.029	41.2	19

It is noteworthy that in terms of the AR, TATR slightly underperforms compared to the YOLO series models. This discrepancy may stem from the inappropriate preset anchor points in the YOLO series models, leading to partial false alarms during detection. Notably, in the mAP@0.75 metric, TATR exhibits a 6.56% and 6.31% improvement over the larger-scale versions of YOLOv5l and YOLOv7x, respectively, indicating an increased intersection between predicted signal boxes and ground truth signal boxes, resulting in higher IoU values for the detected boxes. Regarding signal parameter estimation error, the TATR algorithm demonstrates a reduction in estimation error ranging from 2.5% to 0.3% in comparison to other algorithms, highlighting that the integration of global and local attention mechanisms in the proposed approach is more conducive to capturing signal spectral features and facilitating more precise signal detection. The outcomes of parameter estimation will further impact non-collaborative signal analysis, with enhanced parameter estimation accuracy contributing to the seamless integration of TATR into signal analysis systems. Additionally, the proposed model boasts a reasonable number of parameters, facilitating practical deployment of the model.

To further validate the effectiveness of our proposed method, we visually compare the detection results of TATR with the models mentioned above. The results, as shown in Figure 12, clearly indicate that our method achieves the best detection performance across various SNRs compared to the ground truth. Other detection algorithms exhibit false positives or low IoU with detection boxes, which can be fatal for TT&C signal detection and parameter estimation tasks, as detection boxes that are either too large or too small can introduce errors in parameter estimation. Moreover, when similar background signals or interference signals appear around the TT&C signal, our network demonstrates more accurate recognition and localization. For example, in the results displayed at 0 dB and 10 dB, other models exhibit cases where the detection results include other signals. In the -10 dB result, the YOLO series models incorrectly detect background signals as TT&C signals. Finally, in terms of detection confidence, TATR's detection results have the highest confidence compared to other methods, which further proves TATR has the advantage of a low miss rate and high detection accuracy in TT&C signal detection with complex backgrounds.

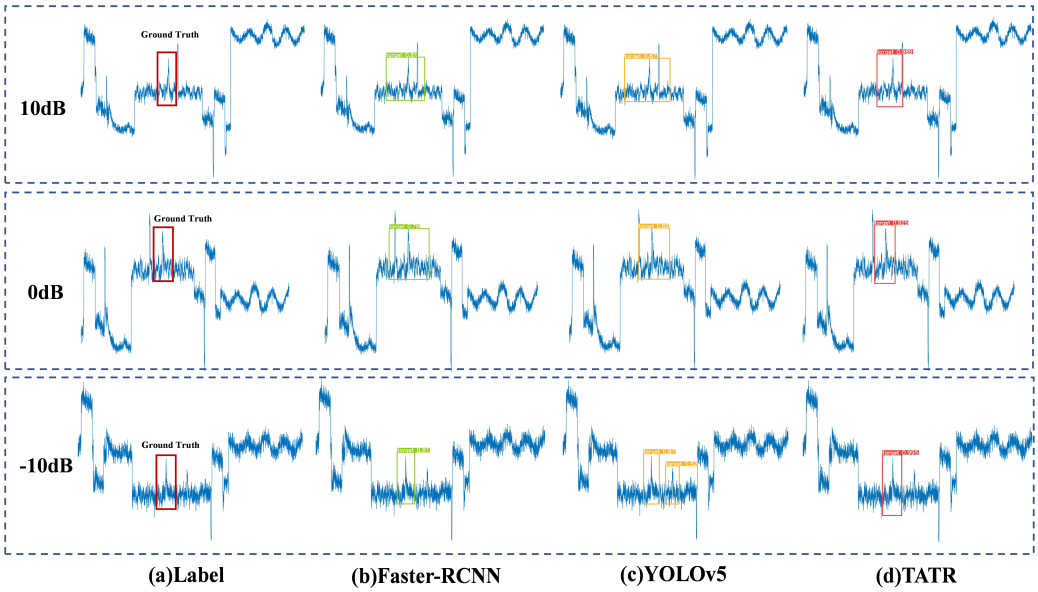

Figure 12. Comparison of signal detection results among different models under different SNRs. (**a**) Represents the ground truth; (**b**) represents the detection result of the Faster-RCNN model; (**c**) represents the detection result of the YOLOv5 model; (**d**) represents the detection result of the TATR model.

7. Discussion

In previous studies, traditional time-domain processing methods often experience performance degradation or even failure in the absence of phase information. Deep learning approaches for TT&C signal spectrum detection are limited currently. In non-collaborative scenarios with constrained receiver conditions, reconnaissance parties face challenges in acquiring prior knowledge and complete signal information, leading to difficulties in detecting TT&C signals amidst strong background signal interference in complex electromagnetic environments. This paper introduces an end-to-end TT&C signal detection model, TATR. As demonstrated in the empirical results shown in Figure 8, with an increase in training epochs, the losses in detection box coordinates and categories consistently decrease while detection AP continually improves. This clearly indicates that our proposed framework achieves effective TT&C signal detection under constrained conditions.

From Table 4, it is evident that the proposed TATR model exhibits significant improvements in the mAP@0.5:0.95 compared to the YOLO series models and Faster-RCNN. This can be attributed to TATR being an anchor-free detection framework that eliminates the post-processing step of non-maximum suppression. Inappropriate anchor settings can significantly degrade signal detection performance based on our years of experience in satellite signal observation and processing. Our network transforms the signal detection box filtering problem into a bipartite graph matching problem, utilizing the Hungarian algorithm to find the optimal matching results.

It is worth noting that in Table 4, TATR slightly underperforms compared to the YOLO series models in AR, which is attributed to fixed anchors causing YOLO to sacrifice detection accuracy slightly to boost AR, resulting in some false alarms, as depicted in Figure 12. A comparison between TATR-4Layer and TATR-6Layer in AR performance reveals that increasing the layers of the encoder and decoder contributes to addressing this issue. However, balancing the high AR with the introduced computational overhead is a crucial focus for future optimization of TATR.

Visualizing the detection box results, as shown in Figure 12, demonstrates that TATR provides more accurate detection localization and higher confidence, owing to its integration of triplet attention and multi-head self-attention mechanisms to capture global and local features of signal spectra, respectively. As indicated in Table 3, with similar signal

detection performance, compared to TATR-6Layer, the parameter count of TATR-4Layer model is reduced by 5.7 M. This reduction is attributed to triplet attention correlating three-dimensional information of signal spectra—amplitude, frequency, and channel—enhancing the feature extraction capability without introducing additional learnable parameters (the parameter quantity only increases by 0.003 M). This alleviates the issue of excessive model complexity caused by stacking multiple encoder and decoder layers. However, due to the need for the decoder architecture to use the inference from the previous inference as the input for the next inference in the inference stage, real-time detection inference performance is insufficient. This will be a key focus for our future research to optimize the TATR framework.

8. Conclusions

In this paper, we propose an end-to-end anchor-free detection framework for TT&C signal detection and parameter estimation under constrained conditions. This framework addresses the issue of performance degradation in traditional methods under phase-restricted conditions. By combining the global attention mechanism of triplet attention with the local contextual correlation ability of the Transformer, TATR can effectively detect satellite signals with an SNR not lower than -10 dB in the presence of strong background interference. Compared to representative object detection networks, TATR not only achieves superior results in mAP metrics and parameter estimation errors but also eliminates the need for the post-processing step of NMS that heavily relies on prior knowledge, making it more suitable for TT&C signal detection and parameter estimation.

In future work, we plan to integrate and deploy TATR into existing satellite spectrum monitoring systems, which will impose further requirements on the real-time performance of the model. Expanding TATR to handle dynamic frames in real-time, such as waterfall plots, may represent a potential research direction in this field. Additionally, given the diversity of TT&C signals, enhancing parameter estimation accuracy and integrating temporal information or modulation features for further demodulation and interpretation of TT&C signals hold significant research significance.

Author Contributions: Conceptualization, Y.L. (Yu Li) and P.C.; methodology, Y.L. (Yu Li) and X.S.; data curation, Y.L. (Yu Li), X.W. and Y.L. (Yongqiang Lu); validation, Y.L. (Yu Li); formal analysis, F.Z. and Y.L. (Yongqiang Lu); investigation, X.W.; resources, F.Z.; writing—original draft preparation, Y.L. (Yu Li); writing—review and editing, Y.L. (Yu Li), X.S. and P.C.; visualization, Y.L. (Yu Li); supervision, X.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 62001350); China Postdoctoral Science Foundation (No. 2016M602775); Postdoctoral Science Research Projects of Shaanxi Province (No. 2018BSHEDZZ39); Joint Fund of Ministry of Education (No. 6141A02022367); Fundamental Research Funds for the Central Universities (No. XJS210210).

Data Availability Statement: Data are available on request due to privacy.

Acknowledgments: We are grateful to the editor and anonymous reviewers for their assistance in evaluating this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Chinese mainland frequency band usage by various operators.

Frequency Band (MHz)	Operator	Format	
825–840/869–885	China Telecom	CDMA	2G/4G
890–909/935–954	China Mobile	GSM900	2G/4G
909–915/954–960	China Unicom	GSM900	2G
1710–1725/1805–1820	China Mobile	DCS1800	2G

Table A1. *Cont.*

Frequency Band (MHz)	Operator	Format	
1745–1755/1840–1850	China Unicom	DCS1800	2G
1755–1765/1850–1860	China Unicom	FDD-LTE	4G
1765–1780/1860–1875	China Telecom	FDD-LTE	4G
1885–1905	China Mobile	TD-LTE	4G
1920–1935/2110–2125	China Telecom	CDMA2000/LTE-FDD	3G/4G/5G
1940–1955/2130–2145	China Unicom	WCDMA/LTE-FDD	3G/4G/5G
2010–2025	China Mobile	TD-SCDMA/ TD-LTE	3G/4G
2300–2320	China Unicom	TD-LTE	4G
2320–2370	China Mobile	TD-LTE	4G
2370–2390	China Telecom	TD-LTE	4G
2555–2575	China Unicom	TD-LTE	4G
2575–2635	China Mobile	TD-LTE	4G
2635–2655	China Telecom	TD-LTE	4G

References

1. Wu, Y.; Pan, J. Detecting Changes in Impervious Surfaces Using Multi-Sensor Satellite Imagery and Machine Learning Methodology in a Metropolitan Area. *Remote Sens.* **2023**, *15*, 5387. [CrossRef] [CrossRef]
2. Li, W.; Sun, Y.; Bai, W.; Du, Q.; Wang, X.; Wang, D.; Liu, C.; Li, F.; Kang, S.; Song, H. A Novel Approach to Evaluate GNSS-RO Signal Receiver Performance in Terms of Ground-Based Atmospheric Occultation Simulation System. *Remote Sens.* **2024**, *16*, 87. [CrossRef] [CrossRef]
3. Feng, S.; Ji, K.; Wang, F.; Zhang, L.; Ma, X.; Kuang, G. Electromagnetic Scattering Feature (ESF) Module Embedded Network Based on ASC Model for Robust and Interpretable SAR ATR. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5235415. [CrossRef] [CrossRef]
4. Feng, S.; Ji, K.; Wang, F.; Zhang, L.; Ma, X.; Kuang, G. PAN: Part Attention Network Integrating Electromagnetic Characteristics for Interpretable SAR Vehicle Target Recognition. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5204617. [CrossRef] [CrossRef]
5. Chen, D.; Shi, S.; Gu, X.; Shim, B. Weak Signal Frequency Detection Using Chaos Theory: A Comprehensive Analysis. *IEEE Trans. Veh. Technol.* **2021**, *70*, 8950–8963. [CrossRef] [CrossRef]
6. Sun, J.; Wang, Y.; Shen, Y.; Lu, S. High-Precision Trajectory Data Reconstruction for TT&C Systems Using LS B-Spline Approximation. *IEEE Signal Process. Lett.* **2020**, *27*, 895–899. [CrossRef]
7. Zhao, Y.; Yang, P.; Xiao, Y.; Dong, B.; Xiang, W. Soft-Feedback Time-Domain Turbo Equalization for Single-Carrier Generalized Spatial Modulation. *IEEE Trans. Veh. Technol.* **2018**, *67*, 9421–9434. [CrossRef] [CrossRef]
8. Wang, D.; Chen, X.; Yi, H.; Zhao, F. Improvement of Non-Maximum Suppression in RGB-D Object Detection. *IEEE Access* **2019**, *7*, 144134–144143. [CrossRef] [CrossRef]
9. Symeonidis, C.; Mademlis, I.; Pitas, I.; Nikolaidis, N. Neural Attention-Driven Non-Maximum Suppression for Person Detection. *IEEE Trans. Image Process.* **2023**, *32*, 2454–2467. [CrossRef] [CrossRef]
10. Misra, D.; Nalamada, T.; Arasanipalai, A.U.; Hou, Q. Rotate to Attend: Convolutional Triplet Attention Module. In Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 5–9 January 2021; pp. 3138–3147.
11. Stewart, R.; Andriluka, M.; Ng, A.Y. End-To-End People Detection in Crowded Scenes. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26–30 June 2016; pp. 2325–2333.
12. Oh, H.; Nam, H. Energy Detection Scheme in the Presence of Burst Signals. *IEEE Signal Process. Lett.* **2019**, *26*, 582–586. [CrossRef] [CrossRef]
13. Shui, P.L.; Bao, Z.; Su, H.T. Nonparametric Detection of FM Signals Using Time-Frequency Ridge Energy. *IEEE Trans. Signal Process.* **2008**, *56*, 1749–1760. [CrossRef] [CrossRef]
14. Liu, C.; Wang, J.; Liu, X.; Liang, Y.C. Maximum Eigenvalue-Based Goodness-of-Fit Detection for Spectrum Sensing in Cognitive Radio. *IEEE Trans. Veh. Technol.* **2019**, *68*, 7747–7760. [CrossRef] [CrossRef]
15. Akhter, M.A.; Heylen, R.; Scheunders, P. A Geometric Matched Filter for Hyperspectral Target Detection and Partial Unmixing. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 661–665. [CrossRef] [CrossRef]
16. Theiler, J.; Foy, B. Effect of Signal Contamination in Matched-filter Detection of the Signal on a Cluttered Background. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 98–102. [CrossRef] [CrossRef]
17. Lunden, J.; Kassam, S.A.; Koivunen, V. Robust Nonparametric Cyclic Correlation-Based Spectrum Sensing for Cognitive Radio. *IEEE Trans. Signal Process.* **2010**, *58*, 38–52. [CrossRef] [CrossRef]
18. Hong, S.; Li, Y.; He, Y.C.; Wang, G.; Jin, M. A Cyclic Correlation-Based Blind SINR Estimation for OFDM Systems. *IEEE Commun. Lett.* **2012**, *16*, 1832–1835. [CrossRef] [CrossRef]
19. Ishihara, S.; Umebayashi, K.; Lehtomäki, J.J. Energy Detection for M-QAM Signals. *IEEE Access* **2023**, *11*, 6305–6319. [CrossRef] [CrossRef]

20. Zheng, L.; Yang, C.; Deng, X.; Ge, W. Linearized Model for MIMO-MFSK Systems with Energy Detection. *IEEE Commun. Lett.* **2022**, *26*, 1408–1412. [CrossRef] [CrossRef]
21. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 24–27 June 2014; pp. 580–587.
22. Girshick, R. Fast R-CNN. *arXiv* **2015**, arXiv:1504.08083.
23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [CrossRef]
24. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. *arXiv* **2016**, arXiv:1512.02325.
25. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2016**, arXiv:1506.02640.
26. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
27. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
28. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-freebies Sets New State-of-the-art for Real-time Object Detectors. *arXiv* **2022**, arXiv:2207.02696.
29. Li, W.; Wang, K.; You, L.; Huang, Z. A New Deep Learning Framework for HF Signal Detection in Wideband Spectrogram. *IEEE Signal Process. Lett.* **2022**, *29*, 1342–1346. [CrossRef] [CrossRef]
30. Li, Y.; Shi, X.; Yang, X.; Zhou, F. Unsupervised Modulation Recognition Method Based on Multi-Domain Representation Contrastive Learning. In Proceedings of the 2023 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Zhengzhou, China, 14–17 November 2023; pp. 1–6. [CrossRef]
31. Zhang, Z.; Shi, X.; Zhou, F. An Incremental Recognition Method for MFR Working Modes Based on Deep Feature Extension in Dynamic Observation Scenarios. *IEEE Sens. J.* **2023**, *23*, 21574–21587. [CrossRef] [CrossRef]
32. Ke, D.; Huang, Z.; Wang, X.; Li, X. Blind Detection Techniques for Non-Cooperative Communication Signals Based on Deep Learning. *IEEE Access* **2019**, *7*, 89218–89225. [CrossRef] [CrossRef]
33. Prasad, K.N.R.S.V.; Dsouza, K.B.; Bhargava, V.K.; Mallick, S.; Boostanimehr, H. A Deep Learning Framework for Blind Time-Frequency Localization in Wideband Systems. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020; pp. 1–6.
34. Xu, W.; Ma, W.; Wang, S.; Gu, X.; Ni, B.; Cheng, W.; Feng, J.; Wang, Q.; Hu, M. Automatic Detection of VLF Tweek Signals Based on the YOLO Model. *Remote Sens.* **2023**, *15*, 5019. [CrossRef] [CrossRef]
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762.
36. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.
37. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In Proceedings of the 2020 European Conference on Computer Vision (ECCV), Online, 23–28 August 2020; pp. 213–229.
38. Zhu, X.; Su, W.; Lu, L.; Li, B.; Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. *arXiv* **2020**, arXiv:2010.04159.
39. Wang, Y.; Zhang, X.; Yang, T.; Sun, J. Anchor DETR: Query Design for Transformer-Based Detector. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 2567–2575. [CrossRef] [CrossRef]
40. Jiang, Y.; Gu, H.; Lu, Y.; Yu, X. 2D-HRA: Two-Dimensional Hierarchical Ring-Based All-Reduce Algorithm in Large-Scale Distributed Machine Learning. *IEEE Access* **2020**, *8*, 183488–183494. [CrossRef] [CrossRef]
41. Jiang, Y.; Fu, F.; Miao, X.; Nie, X.; Cui, B. OSDP: Optimal Sharded Data Parallel for Distributed Deep Learning. *arXiv* **2023**, arXiv:2209.13258.
42. Xu, Z.; Zhu, J.; Geng, J.; Deng, X.; Jiang, W. Triplet Attention Feature Fusion Network for SAR and Optical Image Land Cover Classification. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4256–4259.
43. Modenini, A.; Ripani, B. A Tutorial on the Tracking, Telemetry, and Command (TT&C) for Space Missions. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 1510–1542. [CrossRef]
44. Zhang, T.; Zhang, X.; Yang, Q. Passive Location for 5G OFDM Radiation Sources Based on Virtual Synthetic Aperture. *Remote Sens.* **2023**, *15*, 1695. [CrossRef] [CrossRef]
45. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26–30 June 2016; pp. 770–778.
46. Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

47. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
48. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 658–666.
49. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef] [CrossRef]
50. Yu, C.; Feng, Z.; Wu, Z.; Wei, R.; Song, B.; Cao, C. HB-YOLO: An Improved YOLOv7 Algorithm for Dim-Object Tracking in Satellite Remote Sensing Videos. *Remote Sens.* **2023**, *15*, 3551. [CrossRef] [CrossRef]
51. Everingham, M.; Eslami, S.M.A.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef] [CrossRef]

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.