Segmentation and reconstruction of wireless signals under interference using the FoT-UNet neural network

Nguyen Duy Thai*

Institute of Information Technology - Electronics, Academy of Military Science and Technology, 17 Hoang Sam, Nghia Do, Hanoi, Vietnam.
*Corresponding author: ndthai03@gmail.com
Received 10 Nov. 2025; Revised 06 Jan. 2026; Accepted 26 Feb. 2026; Published 25 Apr. 2026.
DOI: https://doi.org/10.54939/1859-1043.j.mst.110.2026.45-54

ABSTRACT

The increasing prevalence of intentional radio frequency (RF) jamming poses significant challenges to spectrum monitoring and signal reconstruction in complex interference environments. Accordingly, this paper proposes FoT-UNet, a two-stage deep learning framework for time–frequency spectrum segmentation and reconstruction of jammed wireless signals. In the first stage, FoT-UNet-Seg performs RF interference segmentation on 256×256×1 short-time Fourier transform spectrograms, employing the Focal-Tversky loss to enhance sensitivity to small jamming regions. In the second stage, FoT-UNet-Recon, leverages the cleaned spectral features to reconstruct the original spectrum, optimized using a mean absolute error (MAE) objective to minimize amplitude distortion. Experiments on a dataset of 30,000 samples demonstrate that FoT-UNet-Seg achieves IoU = 0.9933, F1 = 0.9966, and Precision = 0.9971, with an average processing time of 13.9 ms per image and 10.9 million parameters. The reconstruction network attains MAE = 0.0801, MSE = 0.0125, PSNR = 20.04 dB, and SSIM = 0.7357, indicating accurate recovery of spectral patterns even under broadband burst interference. In comparison, our approach outperforms several other works in the same task.

Keywords: RF jamming; Spectrum segmentation; Unet; Signal reconstruction; Time–frequency analysis; Deep learning.

1. INTRODUCTION

In modern wireless communications, radio frequency interference (RFI) has become a serious challenge [1], causing a significant degradation in signal quality, spectral efficiency, and receiver sensitivity. The increasing density of wireless devices, along with the coexistence and spectral overlap among various communication standards (such as Wi-Fi, LTE, and 5G-NR), makes radio systems more susceptible to diverse and complex interference sources. These interferences may include both unintentional and intentional jamming [2], such as burst, chirp [3], multi-tone [4], or continuous-wave [5] signals, which distort the time–frequency structure of the received signal. Consequently, accurate detection, localization, and suppression of RF interference are essential for maintaining signal integrity and ensuring stable operation of next-generation wireless networks. Traditional signal-processing approaches, such as adaptive filtering [6] or Wiener filtering [7], have been extensively investigated for interference mitigation. However, their effectiveness decreases markedly under nonlinear, dynamic, or multi-source interference conditions, where the interference can vary rapidly or overlap across multiple frequency bands. In particular, in modern wireless environments characterized by nonstationary and fast-changing interference, conventional analytical methods struggle to fully recover the time–frequency structure and the original signal amplitude.

In recent years, deep learning has demonstrated outstanding capability in extracting complex features and modeling nonlinear relationships between signals and interference [8]. In particular, convolutional neural networks (CNNs) have shown high effectiveness in processing signal and spectral data, enabling direct modeling from time–frequency representations without the need for handcrafted feature engineering. Among these, the U-Net architecture stands out due to its

symmetric encoder–decoder design with skip connections, allowing precise localization and detailed feature recovery during reconstruction [9]. Originally developed for biomedical image segmentation, U-Net has since been successfully extended to various tasks such as audio denoising, radar imaging, and wireless signal processing. Building on these characteristics, this study introduces FoT-UNet, a deep learning architecture specifically designed for RF interference segmentation and spectrum reconstruction under complex jamming conditions. Unlike conventional RF denoising methods that focus only on filtering or spectral restoration, FoT-UNet formulates the problem as a two-stage learning process with explicit guidance. The first stage accurately segments interference regions within the STFT domain, enabling localized spectral masking. The second stage, a symmetric 106 layer U-Net, reconstructs the clean spectrum using an L1 based loss to ensure stable amplitude recovery and smooth transitions. Importantly, the segmentation output is not treated as an auxiliary result; it serves as an explicit time–frequency prior that guides the reconstruction stage by indicating where the spectrum is corrupted. This segmentation guided design enables targeted suppression of interference dominated regions while preserving the underlying clean spectral structures, thereby differentiating FoT-UNet from prior works that perform blind reconstruction or segmentation-only processing. By leveraging multi scale time–frequency feature learning, FoT-UNet shows strong potential as a deep learning solution for detection, analysis, and mitigation of RF interference in modern wireless communication systems.

2. DATA PROCESSING AND MODEL TRAINING

2.1. Dataset and preprocessing

This study employs a real-world wireless communication dataset, which is publicly available on the Kaggle platform [10]. This dataset is one of the largest real-world collections of wireless signal measurements captured directly from standard communication protocols such as Wi-Fi (IEEE 802.11ax), LTE, and 5G-NR. The data were acquired using software defined radio (SDR) devices operating in realistic environments and include various frame types (e.g., uplink, downlink, beacon, control channel, etc.) along with metadata such as center frequency, transmit power, and timestamp. Each raw data file is stored in binary format, containing complex signal samples represented as:

$$x[n] = I[n] + jQ[n] \tag{1}$$

To facilitate deep network training, each data file is transformed and normalized into the time-frequency domain using the Short-Time Fourier Transform (STFT), which is defined as:

$$S_x(k,m) = \sum_{n=0}^{N-1} x[n]w[n-mH]e^{-j2\pi kn/N} \tag{2}$$

where $w[.]$ denotes the Hanning window, H is the hop size between adjacent frames, and N is the FFT length. The indices k and m correspond to the frequency bin and time frame, respectively. The time–frequency representation obtained by the short time Fourier transform provides a physically interpretable description of wireless signals, where spectral energy distributions reflect modulation characteristics, channel effects, and interference behavior. In practical radio environments, interference typically appears as localized or structured energy components in the time–frequency plane, depending on its bandwidth and temporal duration. This representation therefore provides a natural basis for interference-aware spectrum analysis and reconstruction. After applying the STFT, the magnitude spectrum is mapped into a log-amplitude scale and normalized to the range [0,1] as follows:

$$\tilde{X} = \mathrm{clip}\left(\frac{\log\left(1+|S_x|\right)-P_1}{P_{99}-P_1+\varepsilon}, 0, 1\right) \tag{3}$$

where P_q represents the q^{th} percentile of the amplitude distribution. As shown in Figure 1, the normalized spectrograms with dimensions 256×256×1 are used as unified inputs for both training stages of the proposed network.

From the normalized spectrograms, the segmentation dataset is synthesized by simulating various types of interference in the time–frequency domain. The interference masks $M \in [0,1]^{256\times256}$ are randomly generated to represent different jamming patterns, including continuous wave (CW), multi tone, chirp, burst, and broadband types, combined with morphological operations to ensure smooth mask boundaries. The corresponding noisy spectrograms are then created as follows:

$$I_{noisy} = \text{clip}\left(A\left(X_{norm}\right) + \psi M + \eta, 0, 1\right) \tag{4}$$

where $A(\cdot)$ includes several augmentation operations such as horizontal or vertical flipping (fliplr/flipud), time shifting, brightness contrast adjustment, and mild blurring. In the interference model, the amplitude factor ψ is sampled from a uniform distribution $U(0.3,1.0)$ to vary the jamming intensity across samples. Additionally, minor stochastic variations are introduced via additive white Gaussian noise $\eta \sim N(0,\sigma^2)$. This methodology yields a diverse set of paired samples (I_{noisy}, M) which are used to train the FoT-UNet-Seg branch for RF interference segmentation.

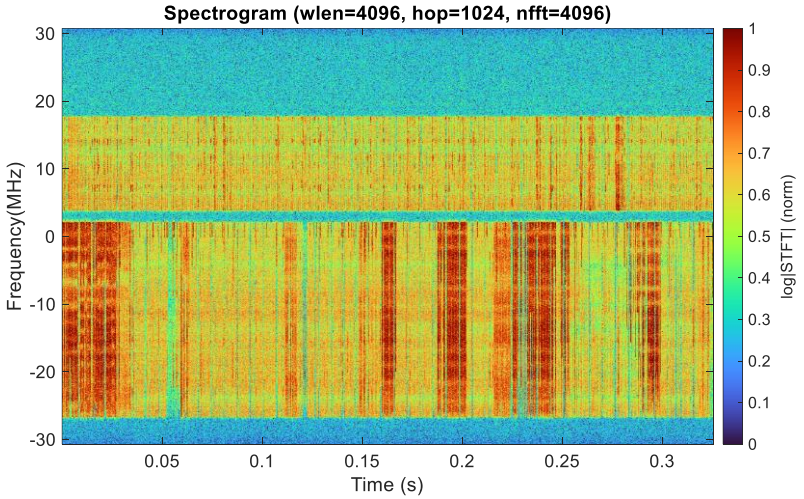


Figure 1. Time–frequency spectrogram of the signal after normalization

The FoT-UNet-Recon branch is trained using paired noisy and clean spectrograms (X_{in},Y_{tar}). The clean signal x_{clean} is artificially mixed with various types of interference $j[n]$ (e.g., CW, multi-tone, chirp, and burst) according to a specified amplitude ratio JSR_{amp} ranging from 0 dB to 15 dB, defined as follows:

$$JSR_{amp} = 20\log_{10}\frac{RMS\left(x_{jam}\right)}{RMS\left(x_{clean}\right)} \tag{5}$$

The noisy waveform is generated as:

$$x_{noisy}[n] = x_{clean}[n] + \psi x_{jam}[n], \quad \psi = \frac{RMS\left(x_{clean}\right)}{RMS\left(x_{jam}\right)} 10^{\left(JSR_{amp}/20\right)} \tag{6}$$

ensuring that the jammer power matches the desired JSR level.

The resulting signal x_{noisy} is then transformed into the time–frequency domain using the short-time Fourier transform (STFT), and its magnitude is converted to a logarithmic scale. The STFT operation yields $S_{\text{x,noisy}}$ and $S_{\text{x,clean}}$, representing the noisy and clean spectrograms, respectively.

The corresponding input–target pair for training the reconstruction network is finally defined as:

$$X_{in} = \log\left(1+\left|S_{x,noisy}\right|\right), Y_{tar} = \log\left(1+\left|S_{x,clean}\right|\right) \tag{7}$$

The jammer mask is then derived from the jammer spectrogram $S_{\mathrm{x,jam}}$ using the 92% threshold as follows:

$$M(k,m) = \begin{cases} 1, & \left|\mathrm{S}_{\mathrm{x,jam}}(k,m)\right| \geq \tau_{92\%} \\ 0 \end{cases}, \tau_{92\%} = P_{92}\left(\left|\mathrm{S}_{\mathrm{x,jam}}\right|\right) \tag{8}$$

The paired samples (X_{in},Y_{tar}, M) are divided into training, validation, and test sets in a 70/15/15 ratio for training the FoT-UNet-Recon model.

2.2. Proposed FoT-UNet model

The proposed FoT-UNet model is designed based on a symmetric UNet architecture as illustrated in Figure 2, optimized for both segmentation and spectrum reconstruction tasks in the time–frequency domain. The architecture consists of five encoder–decoder levels, with feature map channel sizes of 16, 32, 64, 128, 256, and 512, ensuring multi scale feature extraction from fine local spectral details to global contextual information.

In the encoder branch, each level consists of three 2D convolutional layers with 3×3 kernels, combined with Instance Normalization and the LeakyReLU activation function (α=0.01). This is followed by a 2×2 MaxPooling layer to further reduce the spatial resolution. *Instance Normalization* is selected instead of *Batch Normalization* to enhance stability when the spectral power distribution varies significantly across samples, whereas *LeakyReLU* helps maintain stable gradient propagation in low energy regions. The bottleneck employs 512 feature channels with a dropout rate of 0.5 to mitigate overfitting and improve generalization. In the decoder branch, feature maps are progressively reconstructed through a sequence of 2×2 Transposed Convolutions, combined via Depth Concatenation with the corresponding encoder features through skip connections. This symmetric architecture allows precise preservation of spectral details during reconstruction, while maintaining deep contextual representations across higher level abstractions.

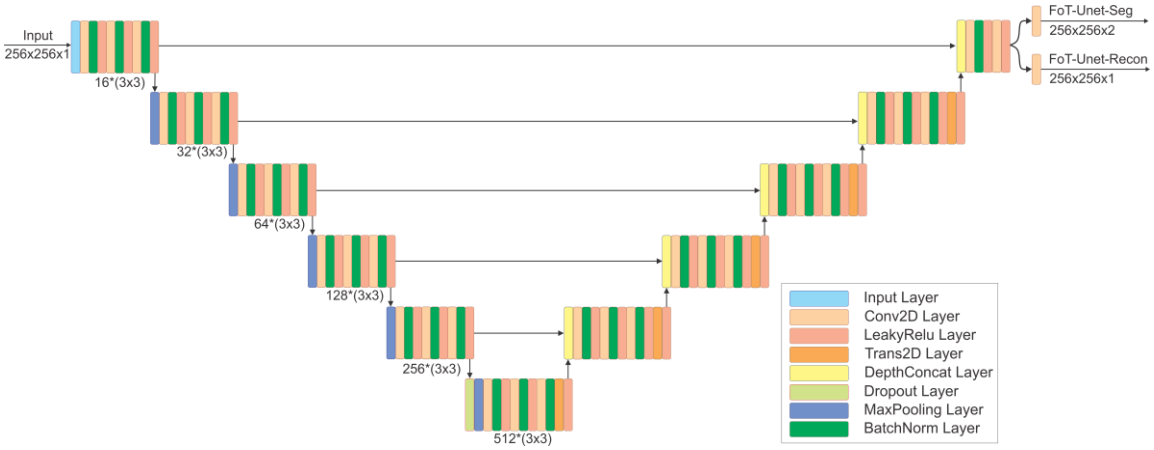


Figure 2. *Architecture of the proposed FoT-UNet model.*

The segmentation branch ends with a Conv2D (1×1) layer producing two output channels that represent the pixel wise probabilities of interference and background. This branch is trained using the Focal-Tversky loss function to enhance sensitivity to small jamming regions. The reconstruction branch (Recon-Head) uses a Conv2D (1×1) layer with a single output channel (clean spectrum), trained with the L1 (MAE) loss to ensure accurate spectral amplitude restoration. Overall, the FoT-UNet consists of 106 layers and performs end-to-end learning on spectrograms

of size 256×256×1. This architecture inherits the strong encoding–decoding capability of UNet while being specifically adapted for RF spectral characteristics, enabling the model to achieve high performance in both RF interference segmentation and clean signal reconstruction tasks.

2.3. Loss function and training hyperparameters

The FoT-UNet model is trained through two independent branches corresponding to two distinct tasks: RF interference segmentation (FoT-UNet-Seg) and clean spectrum reconstruction (FoT-UNet-Recon). Each branch employs a dedicated loss function tailored to its specific objective, along with optimized hyperparameters to ensure convergence stability and strong generalization capability. The RF interference segmentation task is trained using the Focal-Tversky Loss [11], a generalized variant of the Dice Loss designed to address class imbalance between the interference region (minority pixels) and the background (majority pixels), which conventional Dice Loss and Cross-Entropy Loss cannot effectively handle. The loss function is defined as follows:

$$L_{FTL} = \sum_{c=1}^{C} \left(1 - \frac{TP_c}{TP_c + \alpha FP_c + \beta FN_c} \right)^{\gamma} \tag{9}$$

where TPc, FPc, FNc represent the numbers of true-positive, false-positive, and false-negative pixels for class c, respectively. The weighting parameters are set to α=0.7, β=0.3, γ=1.33 to improve sensitivity to small and thin interference regions. This property is particularly important for RF spectrograms, where interference often appears as narrowband or short duration structures in the time-frequency domain, and missing such regions may significantly affect subsequent spectrum reconstruction. This loss formulation allows the model to focus more effectively on difficult-to-classify areas while mitigating gradient vanishing issues under strong class imbalance conditions.

The clean spectrum reconstruction task is formulated as an image regression problem, where the network output Ypred is expected to approximate the ground-truth clean spectrum Ytar to preserve both spectral amplitude and fine structural details, the L1 loss (Mean Absolute Error) is employed, which helps reduce oversmoothing effects in regions with strong interference. The loss function is expressed as:

$$L_1 = \frac{1}{N} \sum_{i=1}^{N} \left| \hat{Y}^{(i)} - Y_{tar}^{(i)} \right| \tag{10}$$

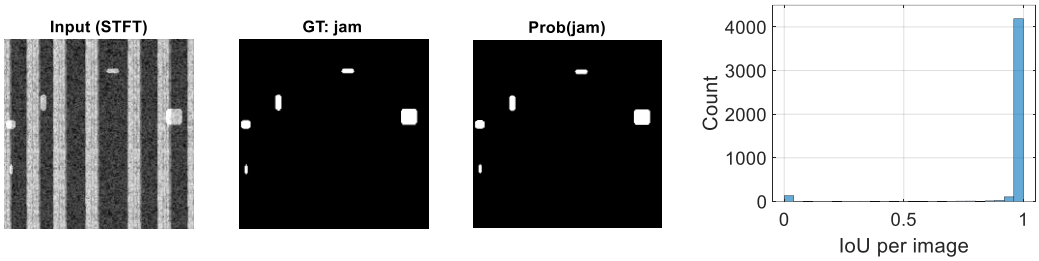
In the FoT-UNet-Recon branch, spectrum reconstruction is conducted in the log-amplitude time–frequency domain, where preserving sharp spectral transitions and narrowband components is essential. The L1 loss is adopted to emphasize the preservation of localized amplitude variations and fine spectral structures, particularly at the boundaries between clean signal components and interference. In the log-amplitude representation, L1 provides a balanced optimization objective across low and high energy regions, which contributes to stable spectrum reconstruction and improved fidelity of the recovered spectral patterns. Both branches are trained using the Adam optimizer with an initial learning rate of lr = 10−4, a batch size of B = 16, and 30 training epochs per branch. Gradients are globally normalized and clipped at 1.0 to prevent gradient explosion, while Dropout (p = 0.5) is applied at the deepest layers to reduce overfitting. The models are monitored using the validation loss (ValLoss), and the best-performing checkpoints are saved automatically when the lowest validation value is reached. All training procedures

are automated in MATLAB R2024a and executed on an NVIDIA RTX 3060 GPU to accelerate computation. The final dataset contains over 30,000 spectrogram samples (256×256), normalized and clearly labeled for both segmentation and reconstruction tasks, ensuring high diversity in interference types, JSR ratios, and real-world noise levels.

3. RESULTS AND DISCUSSION

3.1. Results of RFI segmentation

The qualitative results of the FoT-UNet-Seg model are illustrated in Figure 3a, demonstrating its capability to accurately segment RF interference regions in the input time-frequency spectrograms.



a) *b)*

Figure 3. *Qualitative results of the FoT-UNet-Seg model.*
a) RF interference segmentation; b) IoU distribution for all test samples.

The model effectively reconstructs the interference boundaries and locations relative to the ground truth (GT), while preserving the overall spectral background structure with minimal distortion. The output probability maps indicate that FoT-UNet focuses strongly on regions with abnormal energy while suppressing background noise and avoiding false activations, thereby confirming the network's strong ability to capture discriminative time-frequency representations of interference patterns. The IoU distribution for individual test samples (Figure 3b) shows that nearly all images achieve IoU > 0.9, indicating that the model performs consistently well across the entire dataset, regardless of interference type or spectral intensity. This result confirms the strong generalization capability of FoT-UNet-Seg, demonstrating that the combination of Instance Normalization, LeakyReLU, and Focal-Tversky Loss enables the network to achieve near perfect accuracy in the RF interference segmentation task. From a physical perspective, the segmented interference regions correspond to distinct spectral energy patterns associated with different interference mechanisms. Narrowband interference appears as frequency localized components, whereas burst or broadband interference manifests as short duration, wideband structures in the time–frequency domain. These patterns are consistent with the physical characteristics of wireless interference, enabling the reconstruction stage to selectively suppress interference while preserving signal components shaped by modulation and channel effects.

Table 1. *Quantitative performance of the FoT-UNet-Seg model on the test set.*

Evaluation Mode	**Threshold** (T)	**IoU**	**Precision** (P)	**F1-Score**	**Speed** (ms/image)	**Parameters** (M)
Default	0.500	0.9932	0.9965	0.9966	13.9	10.929
Best F1	0.889	0.9933	0.9971	0.9966	13.9	10.929

The quantitative evaluation on the entire test set demonstrates that the proposed model achieves outstanding performance. Table 1 summarizes the key metrics under two evaluation modes: the default threshold (T = 0.5) and the optimal threshold determined by the best F1 score. In both settings, the model achieves IoU ≈ 0.993, Precision = 0.9965–0.9971, and F1 = 0.9966, reflecting a strong balance between detection accuracy and false rejection. With 10.9 million parameters and

an average processing speed of approximately 13.9 ms per image, FoT-UNet-Seg satisfies the real-time deployment requirements for intelligent wireless communication systems. With an average inference time of approximately 13.9 ms per 256×256 spectrogram, the proposed FoT-UNet enables near real time processing in time–frequency spectrum analysis. This level of computational efficiency supports continuous spectrum monitoring and interference aware processing without introducing excessive latency in intelligent wireless communication systems.

These results demonstrate that FoT-UNet-Seg achieves high pixel level accuracy and robust generalization, meanwhile maintaining fast inference (≈13.9 ms per image) and stable operation under complex real-world interference conditions, confirming its suitability for real time intelligent wireless applications.

3.2. Results of spectrum reconstruction

The spectrum reconstruction results of the FoT-UNet-Recon model are illustrated in Figure 4, clearly demonstrating the model's ability to recover clean wireless signals from noisy spectrograms. The reconstructed spectra closely resemble the ground-truth clean spectra, confirming the model's effectiveness in restoring both amplitude and structural details of the original signals. The comparison between the input (noisy) and target (clean) spectrograms shows that the input spectrum is heavily affected by high amplitude interference, particularly in narrow and rapidly varying frequency regions. After processing by the model, the reconstructed (Predicted) spectrogram accurately restores the signal's energy structure while effectively removing most interference components. The absolute difference map (|Pred − Target|) reveals that the majority of residual errors are concentrated in low-energy or boundary regions of the spectrum, confirming that the network successfully learns the time–frequency correlations between the signal and interference.

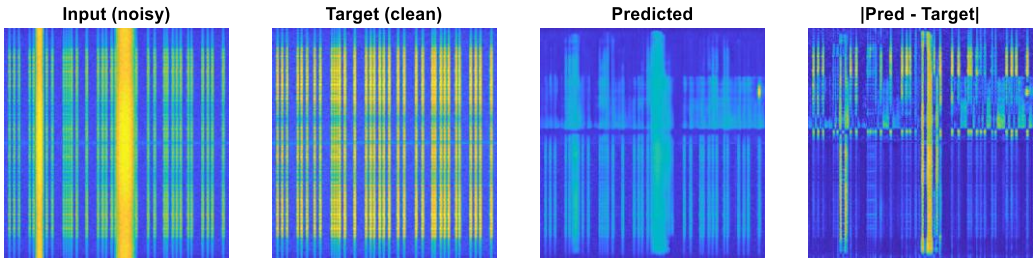
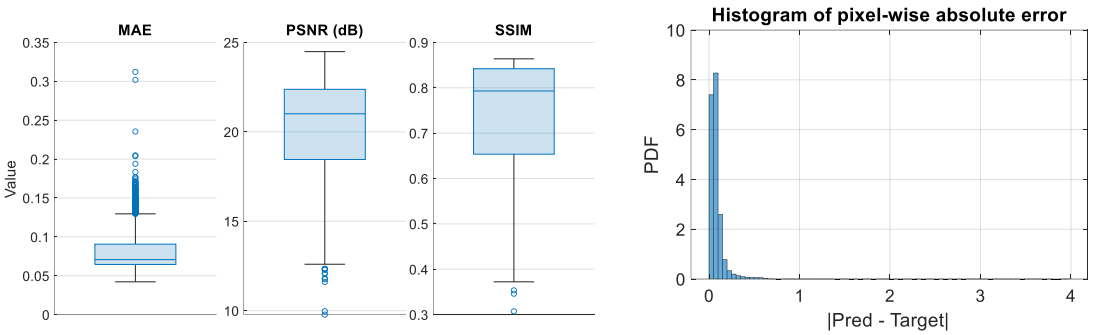


Figure 4. *Spectrum reconstruction results of the FoT-UNet-Recon model.*

Figure 5a illustrates the distributions of MAE, PSNR, and SSIM metrics on the test dataset. The PSNR values mainly fall within the range of 18–24 dB, indicating that the model achieves high spectral fidelity during reconstruction. The SSIM values concentrate around 0.65–0.85, reflecting the model's ability to preserve clear spectral structures, while the MAE remains below 0.1 for most samples, confirming that the model maintains very low absolute error and stable performance across different interference scenarios. On average, the model achieves MAE = 0.0801, MSE = 0.0125, PSNR = 20.04 dB, and SSIM = 0.7357, while the corresponding median values are MAE = 0.0707, MSE = 0.0079, PSNR = 21.01 dB, and SSIM = 0.7930. These results indicate that most reconstructed samples exhibit high spectral fidelity, low reconstruction error, and strong structural similarity compared with the original clean spectra.

The distribution of pixel-wise absolute error is shown in Figure 5b. The histogram indicates that most |Pred−Target| values are concentrated below 0.1, based on more than 36.9 million pixels analyzed across the dataset. The probability density decreases sharply in high-error regions, demonstrating that the model effectively learns the nonlinear and cross-domain relationships between interference and clean signals, thereby significantly reducing local amplitude deviations

across the entire spectrum. The reported MAE, PSNR, and SSIM values indicate that the proposed FoT-UNet-Recon achieves accurate amplitude recovery while preserving the structural characteristics of the clean spectrum. In particular, the low MAE reflects reduced amplitude deviation, whereas the PSNR and SSIM values confirm effective preservation of spectral patterns under strong interference conditions. These results highlight the benefit of the segmentation guided reconstruction strategy, which leverages explicit interference localization to support stable and reliable spectrum recovery, in contrast to blind reconstruction approaches that do not exploit such prior information.



a) *b)*

Figure 5. *Evaluation metrics of the FoT-UNet-Recon model:*
a) Distribution of quantitative evaluation metrics on the test set;
b) Probability distribution of pixel-wise absolute error across the test set.

Overall, both qualitative and quantitative results confirm that FoT-UNet-Recon achieves accurate and stable spectrum reconstruction even under complex interference conditions. Benefiting from its symmetric U-Net architecture combined with Instance Normalization, LeakyReLU, and the L1 loss function, the model effectively removes interference components while preserving the spectral amplitude and energy distribution of the original signal.

3.3. Comparison with other models

To objectively evaluate the performance of the proposed model, this study compares FoT-UNet with several recently published deep-learning models, including AC-UNet [12], SISNet [13], LDNet [14], and EMCA-UNet [15]. These models represent modern approaches to RFI processing in the time-frequency domain, primarily focusing on interference detection or segmentation at the pixel level. In contrast, FoT-UNet is designed as a two-stage architecture, enabling the network to both accurately segment interference and reconstruct clean signal spectra, thereby providing a more comprehensive processing solution compared with previous methods.

Table 2. *Comparison of deep-learning models for RFI segmentation.*

Model	IoU/mIoU	F1-Score	Speed (ms/image)	Parameters (M)
AC-UNet [12]	-	0.925	16.5	13.5
SISNet [13]	0.875	-	6.1	-
LDNet [14]	0.853	0.912	4.46	0.02
EMCA-UNet [15]	0.961	0.960	14.6	10.2
FoT-UNet	0.993	0.996	13.9	10.9

The results presented in table 2 demonstrate that FoT-UNet achieves superior segmentation accuracy, with IoU = 0.993 and F1 = 0.996, outperforming other models such as EMCA-UNet (IoU = 0.961, F1 = 0.960) and LDNet (IoU = 0.853, F1 = 0.912). Although FoT-UNet contains a larger number of parameters (10.9 M), the model still maintains a fast inference speed of approximately 13.9 ms per image, satisfying real-time processing requirements. A distinctive

advantage of FoT-UNet lies in its ability not only to segment interference but also to reconstruct clean signal spectra. The combination of Instance Normalization, LeakyReLU, and Focal-Tversky Loss enables stable gradient propagation and effective learning of complex nonlinear time–frequency features. Consequently, FoT-UNet achieves high-fidelity spectrum recovery with PSNR ≈ 20 dB, SSIM ≈ 0.74, and MAE ≈ 0.08, an aspect that previous models have not addressed. These comparison results confirm that FoT-UNet delivers excellent overall performance in terms of accuracy, efficiency, and reconstruction capability, demonstrating its potential for practical deployment in intelligent spectrum monitoring and anti-jamming communication systems.

4. CONCLUSIONS

This study proposed FoT-UNet, a unified deep learning framework for radio-frequency interference segmentation and wireless signal reconstruction in complex and dynamic spectral environments. The framework consists of two stages: FoT-UNet-Seg, which accurately identifies interference regions in time-frequency spectrograms using the Focal-Tversky loss to address class imbalance, and FoT-UNet-Recon, which reconstructs clean spectra based on a symmetric 106-layer U-Net, optimized with the L1 loss to preserve amplitude and ensure smooth spectral recovery. A key architectural contribution of this work is the explicit coupling between these two stages: the segmentation results provide direct time frequency guidance that informs the reconstruction process, enabling interference aware restoration rather than blind denoising. Experimental results show that FoT-UNet achieves IoU = 0.993, F1 = 0.9966, and reconstruction performance of MAE = 0.0801, PSNR ≈ 20 dB, and SSIM ≈ 0.74, outperforming existing approaches in both accuracy and robustness. These results confirm the model's strong generalization capability and stability when applied to real-world RF datasets. It is important to highlight that the core contribution of this study resides in the proposed two stage segmentation–reconstruction paradigm with explicit time–frequency guidance, rather than in the introduction of a novel network topology. This design effectively embeds interference awareness into the reconstruction process, making it particularly suitable for spectrum processing under complex jamming conditions. Moreover, although synthetic interference is employed during training, this choice facilitates controlled experimentation and reliable quantitative benchmarking. Future extensions incorporating more diverse real world data are expected to further improve the robustness and generalization of the proposed framework.

Furthermore, the integration of Instance Normalization, LeakyReLU, and adaptive loss functions enables the network to converge stably under diverse interference conditions. In the future, research will focus on expanding the dataset across multiple frequency bands and device types, integrating attention mechanisms or time-frequency Transformer architectures to enhance contextual understanding, and adopting multi-task learning to jointly perform segmentation, denoising, and reconstruction within a unified framework. Additionally, developing a lightweight version of FoT-UNet with fast inference capability will facilitate real-time deployment on edge devices, further advancing its applicability in intelligent and adaptive communication systems. The achieved inference latency further confirms the practical feasibility of the proposed framework for near real time spectrum monitoring and anti jamming applications in intelligent wireless systems.

REFERENCES

[1]. Malik, T. A. Alghamdi, M. A. Al-Harthi, and H. Alshahrani. *"Radio Frequency Interference, Its Mitigation and Its Propagation in GNSS"*. Electronics, vol. 14, no. 12, p. 2483, (2025). DOI: 10.3390/electronics14122483.

[2]. H. Pirayesh and H. Zeng. *"Jamming Attacks and Anti-Jamming Strategies in Wireless Networks: A Comprehensive Survey"*. IEEE Communications Surveys & Tutorials, vol. 24, no. 2, pp. 767–809, (2022). DOI: 10.1109/COMST.2022.3155640.

[3]. A. S. Aldhaeebi, M. A. Altahrawi, and H. M. Alghamdi. *"Mitigation of Chirp Jamming with Three Types of Neural Networks"*. Proc. Int. Conf. on Computer Engineering and Systems, (2021).
[4]. S. H. Kiani, N. S. Khan, and M. Qureshi. *"Taxonomy of Physical Layer Jamming Techniques"*. TechRxiv preprint, (2023).
[5]. H. Calatrava, E. Fernandez, and M. Martin. *"Robust Interference Mitigation in GNSS Snapshot Receivers"*. Navigation: Journal of the Institute of Navigation, vol. 72, no. 2, pp. 201–216, (2025). DOI: 10.33012/navi.699.
[6]. Z. Guan, J. Zhang, and H. Wang. *"Retrospection of Nonlinear Adaptive Algorithm-Based Signal Enhancement"*. Sensors, vol. 22, no. 5, p. 1885, (2022). DOI: 10.3390/s22051885.
[7]. C.-L. Liao and J.-J. Ding. *"Enhanced Adaptive Wiener Filtering for Frequency-Varying Noise"*. Applied Sciences, vol. 9, no. 2, p. 47, (2025). DOI: 10.3390/app9020047.
[8]. J. Akeret, C. Chang, A. Lucchi, and A. Refregier. *"Radio Frequency Interference Mitigation Using Deep Convolutional Neural Networks"*. Astronomy and Computing, vol. 18, pp. 35–39, (2017). DOI: 10.1016/j.ascom.2016.10.001.
[9]. A. Vafaei Sadr, A. Sadr, M. Heydari, and A. Refregier. *"Deep Learning Improves Identification of Radio Frequency Interference"*. arXiv preprint, arXiv:2005.08992, (2020).
[10]. S. Siddharth. *"Real-World Wireless Communication Dataset"*. Kaggle Open Data Platform, [Online] Available: https://www.kaggle.com/datasets/siddharthsahoo/real-world-wireless-communication-dataset, (2023).
[11]. N. Abraham and N. Mefraz Khan. *"A Novel Focal Tversky Loss Function with Improved Attention U-Net for Lesion Segmentation"*. Proc. IEEE Int. Symp. on Biomedical Imaging (ISBI), pp. 683–687, (2019). DOI: 10.1109/ISBI.2019.8759329.
[12]. R. Q. Yan, C. Dai, W. Liu, J. X. Li, S. Y. Chen, X. C. Yu, S. F. Zuo, and X. L. Chen. *"Radio-frequency interference detection based on the AC-U-Net network model with atrous convolution"*. Research in Astronomy and Astrophysics, vol. 21, no. 5, pp. 1–13, (2021). DOI: 10.1088/1674-4527/21/5/119.
[13]. Z. Wang, J. Zhao, and X. Li. *"Smart Interference Segmentation Network for Space-Borne SAR Images"*. Remote Sensing, vol. 15, no. 23, article 5462, (2023). DOI: 10.3390/rs15235462.
[14]. F. Zheng, Z. Zhang, and D. Zhang. *"Lightweight deep neural network for radio frequency interference detection and segmentation in synthetic aperture radar"*. Scientific Reports, vol. 14, article 20685, (2024). DOI: 10.1038/s41598-024-71775-8.
[15]. F. Gu, L. Hao, B. Liang, S. Feng, S. Wei, W. Dai, Y. Xu, Z. Li, and Y. Dao. *"Radio Frequency Interference Detection Using Efficient Multi-Scale Convolutional Attention UNet"*. arXiv preprint arXiv:2404.00277, (2024), available: https://arxiv.org/abs/2404.00277.

TÓM TẮT

Phân đoạn và tái tạo tín hiệu vô tuyến khi có nhiễu sử dụng mạng nơ ron FoT-Unet

Sự phổ biến ngày càng tăng của gây nhiễu vô tuyến cố ý đặt ra những thách thức đáng kể cho việc giám sát phổ tần và tái tạo tín hiệu trong các môi trường can nhiễu phức tạp. Bài báo này đề xuất FoT-UNet, một mô hình học sâu hai giai đoạn để phân đoạn phổ thời gian-tần số và tái tạo các tín hiệu bị gây nhiễu. Trong giai đoạn đầu tiên, FoT-UNet-Seg thực hiện phân đoạn nhiễu vô tuyến trên các ảnh phổ biến đổi Fourier thời gian ngắn (STFT) kích thước 256×256×1, sử dụng hàm mất mát Focal-Tversky để tăng cường độ nhạy với các vùng bị gây nhiễu nhỏ. Giai đoạn thứ hai, FoT-UNet-Recon, tận dụng các đặc trưng phổ đã được làm sạch để tái tạo phổ gốc, được tối ưu hóa bằng cách sử dụng lỗi tuyệt đối trung bình (MAE) nhằm giảm thiểu méo biên độ. Các thử nghiệm trên tập dữ liệu 30.000 mẫu cho thấy FoT-UNet-Seg đạt IoU = 0,9933, F1 = 0,9966 và Precision = 0,9971, với thời gian xử lý trung bình khoảng 13,9 ms mỗi ảnh và có 10,9 triệu tham số. Mạng tái tạo đạt MAE = 0,0801, MSE = 0,0125, PSNR = 20,04 dB và SSIM = 0,7357, cho thấy khả năng khôi phục chính xác các mẫu phổ ngay cả trong điều kiện nhiễu xung (burst) băng rộng. Khi so sánh, phương pháp đề xuất của chúng tôi cho thấy hiệu năng tốt hơn một số phương pháp khác khi thực hiện cùng một nhiệm vụ.

Từ khoá: Gây nhiễu RF; Phân đoạn phổ; UNet; Tái tạo tín hiệu; Phân tích thời gian-tần số; Học sâu.