

Fine-Tuned YOLOv26 for Automated Detection of Radio Signals in Spectrograms

Eliezer S. Flores, Paulo R. B. da Silva, Rodrigo Florindo, Rubem V. Pacelli, Victor I. A. de Medeiros, Eric M. Delgado, Fabio S. Lobão, Marcelo L. Nunes, Dimas I. Alves, Felix Antreich, and Renato Machado

Abstract—This paper investigates a YOLOv26 model for automated detection of radio signals in spectrogram representations. The proposed method formulates the problem as a single-class time-frequency domain object detection, enabling the recognition of emissions in the RF spectrum. Raw IQ data from the CSRD2025 dataset are converted into spectrograms and used to fine-tune a YOLOv26 model. Experimental results are promising, with a probability of detection of 90.87% and a precision of 97.71%. The method shows robustness in multi-signal scenarios, although performance degrades at low signal-to-noise ratios and in situations with strong overlap among signals. These findings highlight the potential of YOLO-based models for spectrum monitoring applications.

Keywords—radio signal detection, spectrum monitoring, YOLO, transfer learning.

I. INTRODUCTION

There has been growing demand for automation in wireless communication and spectrum-monitoring systems [1]. An important initial step in automated spectrum analysis is to detect radio emissions in the time-frequency domain, particularly in wideband and multi-signal scenarios [2]. In non-cooperative scenarios, such as spectrum monitoring, cognitive radio, and electronic warfare applications, this task becomes particularly challenging, especially under adverse propagation conditions, low Signal-to-Noise Ratio (SNR), and signal overlap.

In recent years, deep learning methods [3] have demonstrated remarkable performance in automated spectrum-monitoring tasks due to their ability to learn discriminative representations directly from raw or transformed signal data. Convolutional Neural Networks (CNNs), for example, have been widely employed to process radio-frequency signals represented either in the IQ domain or in time-frequency representations such as spectrograms (waterfalls) [4]. Early studies demonstrate that CNN-based approaches can outperform traditional handcrafted-feature pipelines for spectrum

sensing. At the same time, subsequent works reinforce the effectiveness of deep architectures under challenging operating conditions, including low Signal-to-Noise Ratio (SNR) scenarios and complex channel effects (e.g., fading) [5]. These advances have contributed to consolidating deep learning as a relevant and powerful alternative for spectrum-monitoring applications in complex electromagnetic environments.

In this context, object detection has emerged as an attractive paradigm for time-frequency signal analysis [6]. Instead of formulating this analysis exclusively as a conventional classification problem, it could be modeled as an object-detection problem, i.e., signal patterns can be modeled as objects to be first detected in spectrograms, enabling subsequent analysis. This perspective is particularly advantageous in wideband scenarios and in situations involving multiple emissions, including cases where signals share the same frequency band over time (i.e., time-sharing), since these conditions often involve several signals competing for spectral resources. As a result, interference regions become more frequent and harder to distinguish from isolated signal patterns in the time-frequency plane.

Among modern object detection approaches, the YOLO family has gained considerable prominence due to its favorable balance between detection accuracy and computational efficiency [7]. Recent studies have explored the application of YOLO-based models to radio-frequency signal analysis and spectrogram-based detection tasks, demonstrating their effectiveness in detecting signal components in complex spectral environments, such as wideband spectrum monitoring in scenarios involving overlapping or time-varying emissions [8].

Within this family, YOLOv26, the most recent model introduced by Ultralytics, was released on January 14, 2026, and was designed with emphasis on end-to-end inference and deployment efficiency, especially in edge and resource-constrained environments. Recent analyses have highlighted the architectural refinements introduced in this version, as well as its suitability for efficient detection pipelines. These characteristics make YOLOv26 particularly appealing for spectrogram-based signal analysis, in which both detection performance and computational cost are relevant considerations.

Motivated by these advances, this work investigates the use of YOLOv26 for automated detection of radio emissions in spectrogram-based representations. More specifically, the problem is approached from an object-detection perspective, enabling signal patterns to be detected directly in the time-frequency domain. In this study, the medium variant of the

Eliezer S. Flores and Victor I. A. de Medeiros are with the Federal University of Pampa (UNIPAMPA), Alegrete-RS, Brazil. E-mails: {eliezerflores, victormedeiros.aluno}@unipampa.edu.br

Eliezer S. Flores, Paulo R. B. da Silva, Rodrigo Florindo, Rubem V. Pacelli, Dimas I. Alves, Felix Antreich, and Renato Machado are with the Aeronautics Institute of Technology (ITA), São José dos Campos-SP, Brazil. E-mails: {eliezer, pbranco, florindo, rubemrvm, dimasirion, fean, rmachado}@ita.br

Eric M. Delgado, Fabio S. Lobão, and Marcelo L. Nunes are with the National Telecommunications Agency (ANATEL), Brazil. E-mails: {eric, lobao, mlnunes}@anatel.gov.br

This work was supported in part by the National Institute of Science and Technology for Signals (INCT Signals), with funding from the National Council for Scientific and Technological Development (CNPq), under Grant 406517/2022-3 and 305064/2025-8 (PQ-C), the National Telecommunications Agency (ANATEL), and the São Paulo Research Foundation (FAPESP), under Grant 20/09838-0 (BIOS – Brazilian Institute of Data Science).

model, namely YOLOv26m, is adopted due to its more balanced trade-off between detection performance and computational cost, and is referred to simply as YOLOv26 hereafter.

The remainder of this manuscript is organized as follows. Section II presents the YOLOv26 architecture adopted in this paper. Section III describes the dataset and the methodology of the proposed approach, including the YOLOv26 fine-tuning procedure. Section IV reports and discusses the experimental results. Finally, Section V concludes the manuscript and outlines possible directions for future research.

II. YOLOv26 ARCHITECTURE

Object detection is a central topic in computer vision [6], demanding models that combine accuracy and computational efficiency, with YOLOv26 (You Only Look Once) [9] emerging as a recent and prominent architecture for this task. As the latest member of the YOLO family, YOLOv26 has drawn attention due to its architectural refinements and efficient deployment characteristics. This version introduces improvements that enhance object detection performance through the integration of distinct architectural components and optimization strategies. The choice of this architecture was also motivated by the potential of methods of this nature for embedded applications and scenarios with computational constraints, especially considering that YOLOv26 was designed with optimized edge deployment in mind.

Figure 1 presents a schematic overview of the YOLOv26 architecture. In general, the model is organized into three main components, namely, the Backbone, the Neck, and the Head. The Backbone operates as a feature extractor, capturing hierarchical representations from the input data, ranging from low-level local patterns to higher-level semantic structures, which is essential for robust object characterization across different scales and spectral conditions. The Neck is responsible for aggregating and refining these features across different scales, enhancing contextual information. Finally, the Head predicts object classes and bounding boxes.

Within this framework, convolutional blocks combining 2D convolution [10], batch normalization [11], and non-linear activation functions such as SiLU [12] contribute to feature extraction and model execution. In addition, bottleneck structures with shortcut connections [13] improve training efficiency and predictive performance, while feature aggregation mechanisms support the propagation and fusion of feature maps across different stages of the network. Modules inspired by Spatial Pyramid Pooling [14], adapted for faster execution, reinforce the model's ability to handle objects at varying scales while preserving relevant spatial relationships in the extracted feature maps.

In the YOLOv26 framework, several model variants are provided to meet different performance and computational requirements: YOLOv26n (nano), YOLOv26s (small), YOLOv26m (medium), YOLOv26l (large), and YOLOv26x (extra large). Each variant is designed to balance detection accuracy and inference speed, enabling applicability across a wide range of devices, from those with limited computational capabilities, such as embedded and

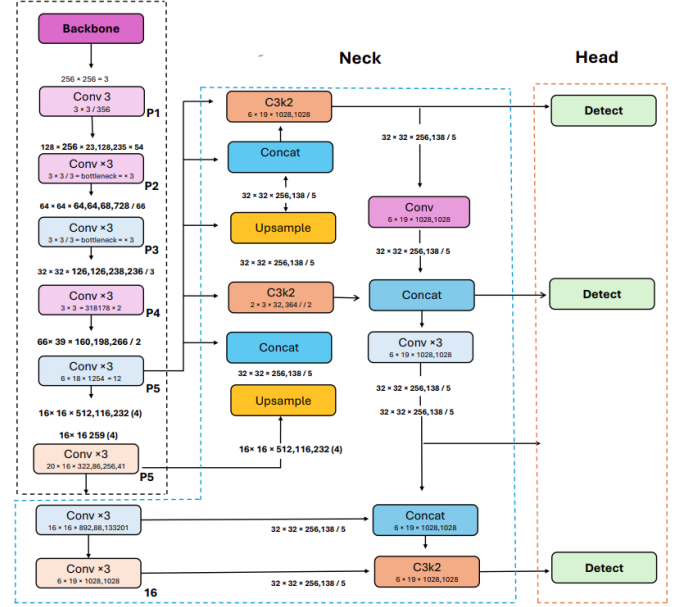


Fig. 1: Core architecture diagram of the Ultralytics YOLOv26 framework, adapted from [9].

mobile systems, to high-performance computing platforms. YOLOv26n is the lightest and fastest configuration, being particularly suitable for applications under strict resource constraints. Progressively, YOLOv26s, YOLOv26m, YOLOv26l, and YOLOv26x provide increasing detection accuracy at the cost of higher computational demand. In this work, the YOLOv26m variant is adopted and is referred to simply as YOLOv26, as it provides the best trade-off between detection performance and computational cost for the target task.

This version adopts an end-to-end, anchor-free detection mechanism, where predictions are made directly over grid-based representations without predefined anchor boxes. In addition, YOLOv26 incorporates a Non-Maximum Suppression (NMS)-free inference design [15], which simplifies the post-processing pipeline and is particularly advantageous for efficient deployment in edge environments. The detection module comprises parallel branches responsible for bounding box regression and class prediction, leveraging sequences of convolutional layers followed by a final prediction layer. This design supports efficient signal detection and bounding-box prediction while simplifying the overall inference framework.

Overall, the streamlined integration among the Backbone, the Neck, and the Head, together with multi-scale feature aggregation and efficient end-to-end prediction, establishes YOLOv26 as a strong candidate for addressing complex object detection problems.

III. MATERIALS AND METHODS

This section describes the dataset and the proposed method used in this work. It first introduces the CSRD2025 dataset and then details the proposed method.

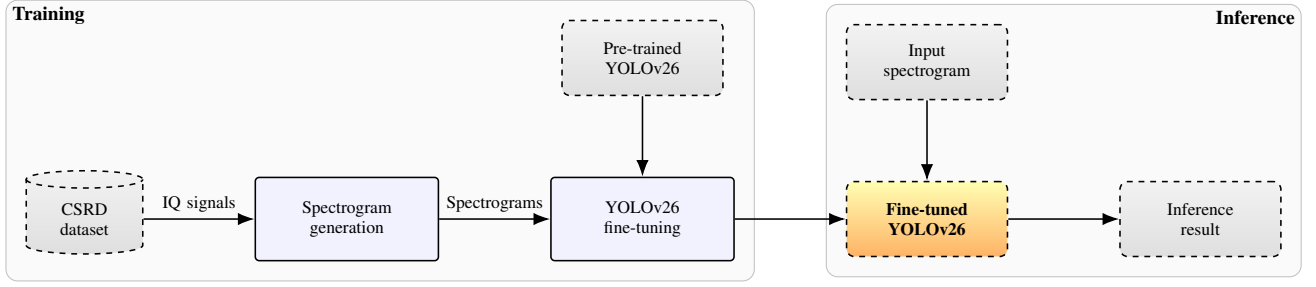


Fig. 2: Overview of the proposed pipeline.

A. Dataset

The experiments conducted in this work are based on the CSRD2025 dataset [16].¹ CSRD2025 is a large-scale benchmark dataset comprising over 25,000,000 frames (approximately 200 TB), specifically designed to provide diverse and realistic RF data for AI-driven spectrum sensing.

TABLE I: Confusion matrix for signal detection on the test set (global percentages).

Predicted Class	Actual Class	
	Signal	Background
Signal	8588 (88.98%)	201 (2.08%)
Background	863 (8.94%)	–

The dataset consists of synthetically generated radio-frequency signals in the form of raw complex baseband IQ samples, obtained from an end-to-end simulation of the transmission and reception process. These signals represent realistic wireless communication scenarios under a wide range of operating conditions, including variations in SNR, carrier frequencies, bandwidths, modulation schemes, and the presence of multiple simultaneous transmissions. The dataset also includes detailed annotations describing the temporal and spectral occupancy of each signal component.

In this study, a reduced subset of CSRD2025 was generated to enable tractable local experimentation on the available computational platform. More specifically, the simulation was limited to 10,000 frames. Since each simulated frame may produce multiple receiver-specific observations, the data effectively used in this work were not the frames themselves, but the corresponding spectrogram representations derived from the receiver outputs, as the same transmission may be observed through substantially different propagation channels depending on the relative transmitter–receiver geometry. This procedure resulted in 17,158 spectrogram images and 94,323 annotated signal objects.

B. Proposed Method

An overview of the proposed method is illustrated in Figure 2. The pipeline is organized into two main stages: (i) Spectrogram Generation, and (ii) Fine-Tuning of YOLOv26. These stages are detailed next.

¹This dataset is publicly available at <https://github.com/Singingkettle/ChangShuoRadioData>.

1) *Spectrogram Generation*: The raw IQ signals are converted into time-frequency representations using the Short-Time Fourier Transform (STFT) with a Hamming window. Let $x[n] = I[n] + jQ[n]$, $n \in \{0, 1, \dots, N-1\}$, denote the discrete-time samples of the received complex baseband signal. The corresponding spectrogram is defined as

$$S[m, k] = \left| \sum_{\tau=0}^{L-1} x[mH + \tau] h[\tau] e^{-j\frac{2\pi}{K}k\tau} \right|^2, \quad (1)$$

where H is the hop size between consecutive frames, and $m \in \{0, 1, \dots, M-1\}$ and $k \in \{0, 1, \dots, K-1\}$ denote the time-frame and frequency-bin indices, respectively, with m restricted such that $mH + L \leq N$. The Hamming window is given by

$$h[\tau] = 0.54 - 0.46 \cos\left(\frac{2\pi\tau}{L-1}\right), \quad (2)$$

with $\tau \in \{0, 1, \dots, L-1\}$.

Finally, the spectrograms are normalized via min-max scaling and represented as 8-bit grayscale images with values in the range $[0, 255]$. These images are subsequently resized to a fixed resolution of 640×640 to ensure compatibility with the input requirements of the YOLOv26 model.

2) *YOLOv26 Fine-Tuning*: Transfer learning with fine-tuning was employed to adapt YOLOv26, originally pre-trained on the COCO dataset of natural images, to the specific spectrogram context. The network was fine-tuned using the spectrogram dataset derived from CSRD2025.

As the objective of this study is to detect the presence and time-frequency occupancy of radio emissions independently of their modulation format, all annotated instances were treated as belonging to a single class, denoted as `signal`. The fine-tuning procedure was carried out for 100 epochs with a batch size of 16, using the original YOLOv26 training objective. The AdamW optimizer was adopted with an initial learning rate of 0.001. In addition, automatic mixed precision was disabled, while mosaic and mixup data augmentation were enabled with factors of 1.0 and 0.1, respectively.

IV. EXPERIMENTAL RESULTS

In this section, the experiments and corresponding results are presented and discussed. All experiments were conducted on a workstation equipped with an AMD Ryzen 9 5900X processor, an NVIDIA GeForce RTX 4090 GPU with 24 GB

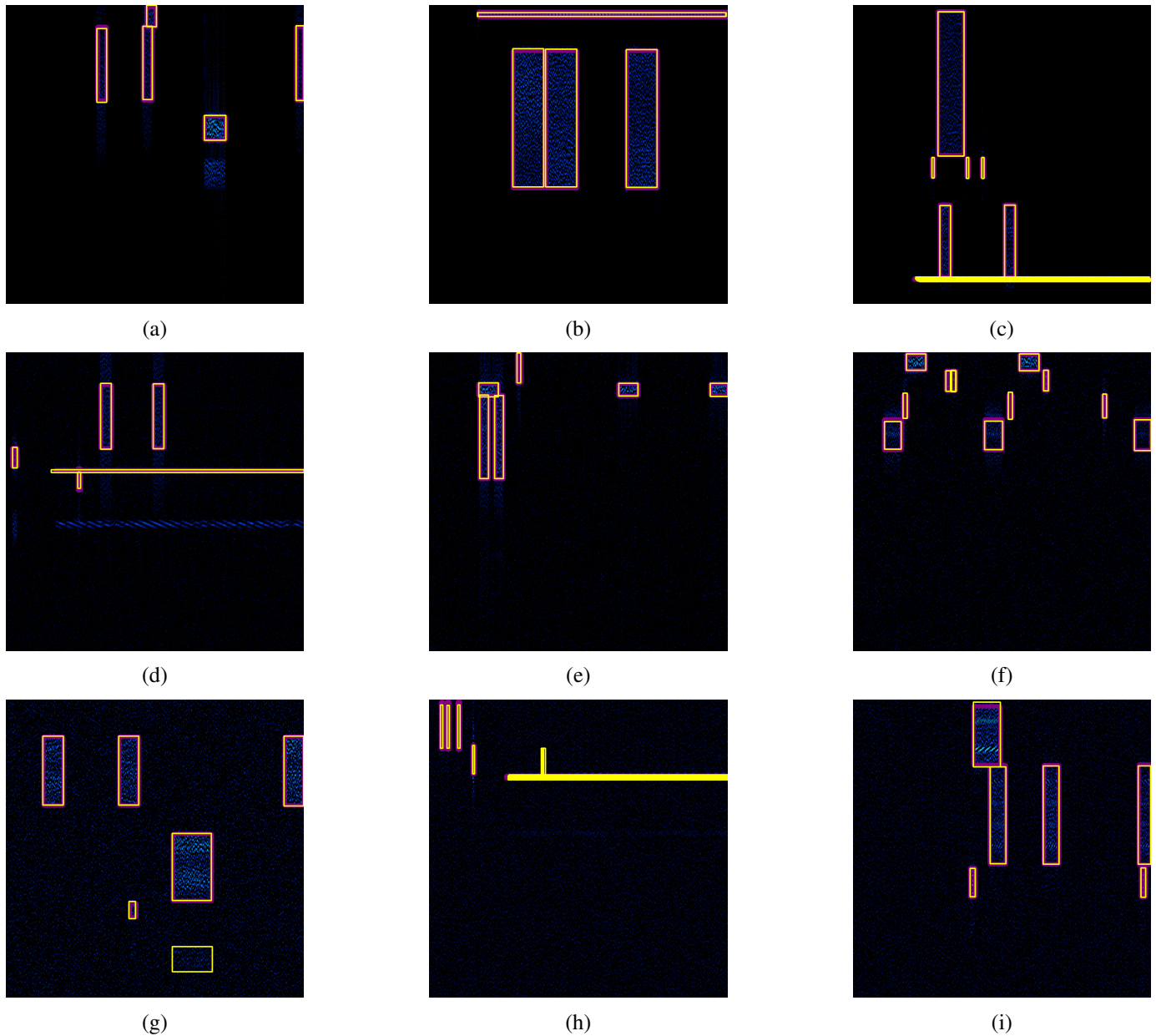


Fig. 3: Representative qualitative results obtained by the proposed method on spectrogram representations. The predicted bounding boxes are shown in yellow, while the ground-truth annotations are shown in purple. The first row presents successful detections in scenarios with more uniform backgrounds. The second row shows successful detections in more challenging conditions, including noisier backgrounds and greater variability in signal appearance. The third row illustrates representative failure cases, where weak signal components, low-contrast patterns, or adverse background conditions lead to missed detections or inaccurate bounding-box predictions.

of VRAM, 64 GB of DDR4 RAM, and a 4 TB M.2 SSD. For reproducibility purposes, all source code developed in this work is publicly available at <https://github.com/eliezersflores/ChangShuoRadioData>.

For the experimental evaluation carried out in this work, the dataset was partitioned into training, validation, and test subsets comprising 80%, 10%, and 10% of the samples, respectively. The training set was used to fine-tune the model, the validation set was used to monitor generalization performance during training, and the test set was reserved exclusively for the final performance evaluation. The detailed analysis that fol-

lows combines qualitative inspection of detection results with quantitative evidence, providing a comprehensive assessment of the model behavior.

Figure 3 displays representative qualitative results obtained by the proposed method on spectrograms from the test set, with predicted bounding boxes shown in yellow and ground-truth annotations in purple. The first row shows successful detections in scenarios with relatively uniform backgrounds. Even under these comparatively cleaner conditions, the examples exhibit noticeable variability in signal appearance, including differences in time-frequency occupancy and intensity,

as well as partially overlapping emissions. Still, the predicted bounding boxes remain well aligned with the ground-truth annotations, indicating that the proposed method can detect heterogeneous emission patterns and properly separate adjacent or partially overlapping emissions.

The second row shows successful detections in more challenging conditions, characterized by noisier and less uniform backgrounds. In these cases, the lower SNR makes the distinction between actual emissions and background fluctuations less evident. Even so, the method correctly identifies the relevant emissions without confusing background clutter with signal components. The third row illustrates representative failure cases, including false alarms, in which portions of the background are incorrectly detected as emissions, and bounding-box prediction inaccuracies, in which the predicted bounding boxes extend slightly beyond the corresponding ground-truth annotations.

Table I presents the confusion matrix associated with the detection results. Although the problem is formulated as a single-class detection task, this analysis still provides useful insights into detection consistency and error patterns, particularly with respect to the distinction between signal and background. The results show that the model correctly identifies most signal instances, while the observed errors are predominantly associated with missed detections under challenging conditions, such as low SNR.

Quantitative evidence derived from Table I further supports the effectiveness of the proposed method. More specifically, the proposed method achieved a probability of detection P_D of 90.87%, correctly identifying 8588 out of 9451 ($= 8588 + 863$) annotated signal boxes, and a precision of 97.71%, corresponding to 8588 correct detections out of 8789 ($= 8588 + 201$) predicted boxes, indicating that most predicted detections indeed correspond to true signal occurrences. It is worth noting that the false-alarm probability P_F , in the classical detection-theoretic sense, could not be explicitly computed in this study. Although false-positive detections can be counted under the adopted object-detection protocol, the dataset and evaluation procedure do not define an explicit negative hypothesis space, such as a fixed set of non-target trials, background cells, or negative candidate regions from which a probability of false alarm could be estimated.

V. CONCLUSION

This paper investigated the use of a YOLOv26 model for the automated analysis of radio signals in spectrogram representations. By formulating the problem from an object-detection perspective, the proposed method enables simultaneous detection and bounding-box estimation of signal patterns in the time-frequency domain, providing a more informative interpretation than conventional classification-based strategies.

The experimental results were promising, with a probability of detection of 90.87% and a precision of 97.71%, indicating that most predicted detections correspond to actual signal occurrences. These results highlight the effectiveness of combining spectrogram-based representations with a fine-tuned YOLOv26 architecture, particularly in scenarios involv-

ing multiple simultaneous transmissions and diverse signal conditions.

Despite these promising results, performance degradation was observed under challenging conditions, such as low signal-to-noise ratio. These limitations suggest that the current approach may benefit from further improvements, particularly in enhancing robustness to weak or ambiguous spectral patterns.

Future work may explore hyperparameter optimization, multi-scale training schemes, and the integration of complementary models to improve detection performance in adverse scenarios. Additionally, extending the proposed framework to multi-class detection, enabling the discrimination of different modulation families, represents a natural and relevant direction for further investigation.

REFERENCES

- [1] A. Gbenga-Ilori, A. L. Imoize, K. Noor, and P. O. Adebolu-Olode, "Artificial intelligence empowering dynamic spectrum access in advanced wireless communications: A comprehensive overview," *AI*, vol. 6, no. 6, p. 126, 2025.
- [2] R. Zhao, Y. Ruan, and Y. Li, "Cooperative time-frequency localization for wideband spectrum sensing with a lightweight detector," *IEEE Communications Letters*, vol. 27, no. 7, pp. 1844–1848, 2023.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [4] S. N. Syed, P. I. Lazaridis, F. A. Khan, Q. Z. Ahmed, M. Hafeez, A. Ivanov, V. Poulkov, and Z. D. Zaharis, "Deep neural networks for spectrum sensing: A review," *IEEE Access*, vol. 11, pp. 89 591–89 615, 2023.
- [5] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.
- [6] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.
- [7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.
- [8] A. Vagollari, M. Hirschbeck, and W. Gerstaecker, "An end-to-end deep learning framework for wideband signal recognition," *IEEE Access*, vol. 11, pp. 52 899–52 922, 2023.
- [9] R. Sapkota, R. H. Cheppally, A. Sharda, and M. Karkee, "YOLO26: Key architectural enhancements and performance benchmarking for real-time object detection," *arXiv preprint arXiv:2509.25164*, 2025, <https://arxiv.org/abs/2509.25164>.
- [10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [11] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, 2015, pp. 448–456.
- [12] S. Elfving, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Networks*, vol. 107, pp. 3–11, 2018.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [14] —, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014, pp. 346–361.
- [15] S. Chakrabarty, "YOLO26: An analysis of NMS-free end to end framework for real-time object detection," *arXiv preprint arXiv:2601.12882*, 2026, <https://arxiv.org/abs/2601.12882>.
- [16] S. Chang, R. Sun, J. He, S. Huang, K. Yu, and Z. Feng, "CSR2025: A large-scale synthetic radio dataset for spectrum sensing in wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 44, pp. 2378–2392, 2026.