A Deep Ensemble Semantic Segmentation Framework for Efficient Spectrum Sensing in Cognitive Radio Networks

Md. Minhajul Islam Arnab
Department of Electronics and Communication Engineering, Khulna University of Engineering & Technology (KUET), Bangladesh
arnab@ece.kuet.ac.bd

Sk. Shariful Alam
Department of Electronics and Communication Engineering, Khulna University of Engineering & Technology (KUET), Bangladesh
ssalm@ece.kuet.ac.bd (corresponding author)

Fariha Alam Rafa
Department of Electronics and Communication Engineering, Khulna University of Engineering & Technology (KUET), Bangladesh
farihaalamrafa@gmail.com

Rubaiyat Hasan Ucchwas
Department of Electronics and Communication Engineering, Khulna University of Engineering & Technology (KUET), Bangladesh
rhkhan099@gmail.com

Received: 28 December 2025 | Revised: 11 February 2026 | Accepted: 14 February 2026

 DOI: https://doi.org/10.48084/etasr.17221

ABSTRACT

With 5G networks transitioning to congested frequency bands, distinguishing Long Term Evolution (LTE) and New Radio (NR) signals from noise becomes critical for effective interference management. However, simultaneously locating local signal boundaries and establishing a global context can be difficult when traditional semantic segmentation techniques are utilized. To address this issue, we propose a robust ensemble deep learning framework that integrates two distinct architectures. The first is DeepLabV3+ with Atrous Spatial Pyramid Pooling (ASPP) to acquire multi-scale contextual features. The second is a U-Net architecture integrated with the spatial and channel Squeeze-and-Excitation (scSE) attention enhancement mechanism. The integration of the scSE attention enhancement mechanism with the U-Net model helps adjust the feature maps when needed, reducing insignificant areas and revealing valuable spectral features. In order to ensure model generalizability, the training process adopted large-scale data augmentation and used median frequency balancing to correct the effect of class imbalance. Evaluations demonstrate that our dual-stream ensemble approach is much more effective than baseline models and previous studies. The proposed ensemble framework achieved an excellent mean Intersection over Union (mIoU) score of 0.9878. The average mIoUs of common architectures such as PixelMLP, SimpleFCN and SimpleSegNet were 0.2962, 0.4998, and 0.5931, respectively. These findings demonstrate that combining attention-based feature refinement with global context extraction is a highly dependable and precise technique for automating 5G spectrum detection.

Keywords-DeepLabV3+; ensemble learning; scSE attention; semantic segmentation; spectrum sensing; U-Net

I. INTRODUCTION

All wireless communication technologies are rapidly evolving with the advent of the Internet of Things (IoT), along with the emergence of 5G and beyond-5G networks. IoT refers to a network of physical objects equipped with sensors and software that enable them to exchange data over the network. These devices generate vast amounts of data that can be used

for monitoring, automation, and improving services, making a wide range of smart applications feasible in day-to-day life, such as autonomous traffic management, efficient hospital management with real-time remote patient monitoring, automated transportation systems, advanced security solutions, and so on. These smart applications serve as the fundamental building blocks of smart homes, smart cities, and digital smart hospitals. Ultimately, Radio Frequency (RF) spectrum is primarily responsible for the seamless functionalities of these smart applications in real life. However, current spectrum management regulations have had a hard time keeping up with modern demand. While allocating the bandwidths amongst users, often the licensed services receive the allocation of large segments of bandwidth [1]. Studies show that a severe underutilization of the available spectrum exists, as a significant portion of the allocated spectrum sits idle most of the time. This creates "artificial scarcity," which may appear as a bottleneck for the advancement of future wireless systems [2].

Dynamic Spectrum Access (DSA), which is implemented using cognitive radio, is being utilized extensively to solve this problem. DSA enables unlicensed secondary users to use a certain portion of the spectrum temporarily when no primary licensed users are occupying it. As a result, sensing the spectrum to detect whether it is occupied, even in real-world noisy environments, has become an important critical requirement today. Traditional signal processing methods, such as energy detection and matched filtering, cause unavoidable problems in modern applications, resulting in performance degradation and increased complexity. Although energy detection is a straightforward and simple method to detect signals, it drastically fails when the signal is buried in extreme noise, known as the Signal-to-Noise Ratio (SNR) wall [3]. This misdetection occurs because the method often confuses low-power signals with noise. In contrast, although matched filtering can accurately detect signals even in noisy channels, it requires complete prior knowledge of the primary signal's characteristics (e.g. modulation type, carrier frequency, bandwidth, pulse duration), which is highly impractical in dynamic and heterogeneous networks [4].

Recent developments in spectrum sensing are increasingly embracing semantic segmentation to utilize the rich time–frequency structure of spectrograms. A few studies have also suggested dedicated architectures, which provide greater segmentation accuracy due to an effective balance between local and global information. Indicatively, authors in [5] propose an attention-based Spectrum Transformer that uses multi-head self-attention to learn inner-band features and inter-band spectrum correlations in detecting wideband spectrum. However, high computational complexity makes it difficult to execute in lightweight or resource-constrained environments. Likewise, the proposed PRMNet in [6] adopts a resolution-preserving architecture that avoids aggressive downsampling of input spectrograms and incorporates a multi-scale convolutional feature extraction module to capture both fine and coarse spectral patterns.

In addition to models based on spectrograms, other models consider fusion of features or alternative input representations. Combining various radio-domain characteristics (In-phase and Quadrature (I/Q), amplitude, phase) with a Convolutional Neural Network–Long Short-Term Memory (CNN–LSTM) hybrid architecture, authors in [7] improve robustness at low SNR conditions, but do not conduct pixel-wise segmentation. Such a variety of features is also utilized in ensemble-based segmentation and works like U-Net RGB/HSV bagging [8] and the ensemble-based analysis of [9] prove that ensemble models are more robust when they utilize complementary features. However, these gains typically come with a tradeoff of increased computational cost and rely on fixed or heuristic fusion strategies rather than adaptive attention-based feature fusion.

Recent research in RF sensing has performed segmentation to I/Q or sequence representations directly. The approach of SSDSS [10] performs sequence semantic segmentation on Signal-Wise Differential Features (SWDF) using recurrent and attention-based networks to resolve co-channel signals, although the model relies on specialized preprocessing. Seek and Classify [11] performs a joint narrowband signal segmentation and modulation classification in 1D frequency-band sequences to reduce covariate shifts but operates on 1D frequency sequences rather than utilizing the 2D spectrogram structure. Another data-centric model, Stitching the Spectrum [12], uses a 1D U-Net with non-local attention on stitched wideband I/Q samples to separate multi-signals with low-latency, although the system relies on a handcrafted data generation and augmentation pipeline and assumes a fixed set of signal classes.

Similar advances in the computer vision field can shed more light on segmentation model construction. A reduced kernel Principal Component Analysis (PCA) model based on orthogonal linear projections was proposed in [13], noting its sensitivity to kernel parameter selection. In [14], noise uncertainty at low SNR in cognitive radio networks was addressed by integrating kernel PCA with a dynamic-threshold energy detection technique. Transformer-style frameworks like Mask2Former [15], SegFormer [16], and SegViT v2 [17] demonstrate the power of using attention and multi-scale feature integration, but they are computationally intensive and typically rely on large-scale natural-image pretraining when applied to RF domains. Innovations using CNNs such as PIDNet [18] and InternImage [19] focus on efficiently balancing details and context, and enabling adaptive spatial aggregation, yet are still primarily concerned with the natural-image domain. In addition, generalization is shown to be impressive using universal models such as Segment Anything (SAM) [20], which uses a ViT-based image encoder within a promptable segmentation framework. However, its size and natural-image bias make it impractical for resource-constrained spectrum sensing tasks.

Compared to these more general solutions, SRNet [21] is much more focused on spectrogram-based spectrum sensing and uses an encoder-decoder CNN with attention mechanisms to identify pixel-level spectral occupancy. Although SRNet has demonstrated potential for semantic segmentation of 5G/LTE occupancy mapping, it is hindered by performance degradation at low SNR and relies on a single-model architecture. This

encourages more powerful feature fusion methods, such as the Ensemble DeepLabV3+ and U-Net framework with spatial and channel Squeeze-and-Excitation (scSE) attention examined in this paper.

So, in this study, an ensemble framework is proposed that is capable of efficiently identifying spectral content in wideband spectrograms, as shown in Figure 1. The key contributions of this work are as follows:

- Inverse class weighting via frequency balancing addresses the severe class imbalance between background noise and active signals to guarantee convergence during training.
- A dual-stream ensemble integrates DeepLabV3+ and scSE-enhanced U-Net, combining Atrous Spatial Pyramid Pooling (ASPP) for global context with attention mechanisms for local boundary refinement.
- A soft-voting fusion strategy averages softmax probabilities from the constituent models, effectively minimizing prediction variance and improving segmentation accuracy.

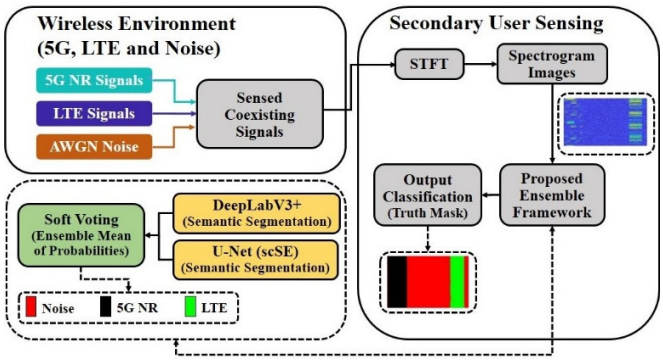


Fig. 1. 5G–LTE network architecture with deep learning-based spectrum sensing.

II. SIGNAL MODEL AND DATASET REPRESENTATION

A. Signal Model

In the context of wireless communication systems requiring spectrum sensing, the sensed or received signal can be represented as a complex-valued signal, consisting of 5G New Radio (NR) and Long Term Evolution (LTE) signals. These transmitted signals often get affected by noise as well as different channel impairments, representing real-world imperfect channel conditions. The signal received at the receiver $y(t)$ can be mathematically expressed as:

$$y(t) = x(t) * h(t) + n(t) \quad (1)$$

where $x(t)$ signifies the noiseless message signal transmitted over the channel, $h(t)$ signifies the channel impulse response representing its physical characteristics, and $n(t)$ is the noise added to the signal. In order to create a realistic situation, imperfect channel conditions such as Doppler shifts and link-level fading are considered to strictly test how well the spectrum sensing framework performs.

B. Dataset Representation and Transformation

To test the proposed ensemble structure, the typical wideband spectrogram dataset provided by authors in [21] in their work on SRNet, and available on Github [22], is used. It is a collection of artificially produced 1D time-domain 5G NR and LTE signals, which are generated through the 5G and LTE Toolboxes of MATLAB. These signals, as mentioned in (1), are subjected to real-world channel impairments such as Doppler shifts and fading to create rigorous testing conditions.

The Short-Time Fourier Transform (STFT) converts these 1D time-domain signals into the time–frequency domain that serve as input to the proposed semantic segmentation network. STFT of signal $x(n)$ in the discrete time domain is given as:

$$X(t,f) = \sum_{n=0}^{N-1} x(n)w(n-t)e^{-i2\pi\frac{fn}{N}} \quad (2)$$

where $w(n-t)$ is the window function, and N is the Fast Fourier transform (FFT) length (4096). This transformation results in a dataset of 2D RGB spectrogram images with a 256 × 256 pixel resolution and corresponding pixel-wise ground truth masks. These masks classify the time–frequency grid into three semantic classes: background noise (Class 0), 5G NR (Class 1), and LTE (Class 2). Table I summarizes the detailed signal configurations and channel parameters that characterize the dataset. Figure 2 demonstrates the spectrogram images and their corresponding ground truth masks at 20 dB and 60 dB SNR.

III. METHODOLOGY

A. Dataset Preprocessing

Before training, spectrogram images undergo a systematic preprocessing pipeline to make them compatible with deep learning models and to resolve inherent challenges of the dataset, including extreme class imbalance.

- Label remapping: The original ground truth mask pixel values are mapped to discrete class indices that help to compute cross-entropy losses: background noise (0), 5G NR (1), and LTE (2).
- Inverse weight calculation through median frequency balancing: To mitigate the effect of severe class imbalance of majority class (background noise), median frequency balancing is employed to calculate inverse class weights. Each class weight is calculated as the median of all the class frequencies divided by the class frequency:

$$f_c = \frac{N_c}{\sum_{i=0}^{C-1} N_i} \quad (3)$$

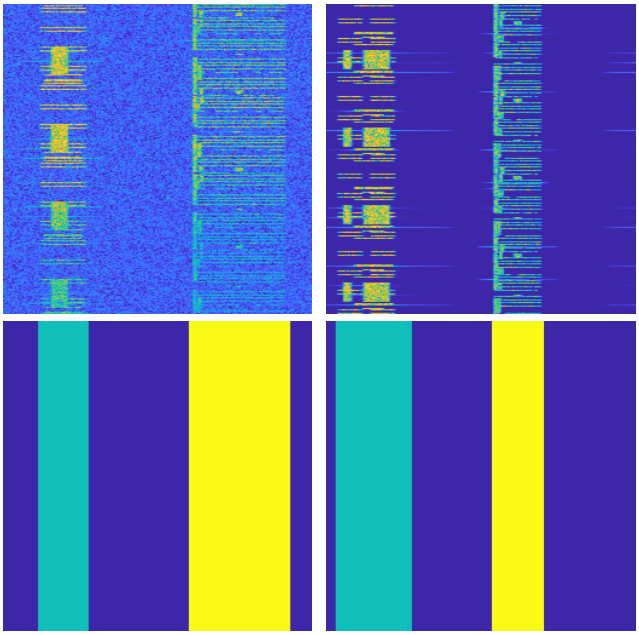
where N_c represents the total number of pixels belonging to class c across the entire training set, and C is the total number of classes. The weight w_c for each class is then derived using the median of these frequencies:

$$w_c = \frac{\text{median}(\{f_0, f_1, f_2, \ldots, f_{C-1}\})}{f_c} \quad (4)$$

Such weights are incorporated into the weighted cross-entropy loss, which ensures that the misclassification of minor classes (5G/LTE) is punished more heavily.

TABLE I. SIGNAL CONFIGURATIONS AND CHANNEL CONDITIONS [21]

Category	Parameter	Range of values
Channel parameters	SNR	{20, 30, 40, 50} dB
	Doppler	{0, 70, 750} Hz
	Center frequency	{1500, 2600, 3500, 4000} MHz
5G parameters	Subcarrier spacing	{15, 30} kHz
	Bandwidth	{10, 15, 20, 25, 30, 40, 50} MHz
	SSB period	{5, 10, 20} ms
	Modulation	{QPSK, 16-QAM, 64-QAM}
LTE parameters	Reference channel	{R.2, R.3, R.5, R.6, R.8, R.9}
	Bandwidth	{1.5, 10, 20} MHz
	Modulation	{QPSK, 16-QAM, 64-QAM}



(a) (b)

Fig. 2. Visualization of the RGB spectrogram image and ground truth masks at SNR: (a) 20dB, and (b) 60 dB.

B. Proposed Ensemble Framework

To enhance the robustness of spectrum sensing, this study proposes an ensemble framework consisting of two segmentation architectures that take into consideration complementary spectral characteristics. The ensemble comprises:

- DeepLabV3+: The DeepLabV3+ architecture can effectively capture broader multi-scale spectral patterns by utilizing the ASPP module. ASPP utilizes a dilated convolution mechanism and thereby possesses the capability of analyzing feature maps across multiple receptive fields without sacrificing resolution. Subsequently, the decoder module upsamples the ASPP output and fuses it with low-level encoder features, resulting in clearer reconstruction of signal boundaries.
- U-Net with scSE: The traditional U-Net encoder-decoder, when integrated with the scSE module, gains the ability to recalibrate feature maps to emphasize informative spatial locations and channel dependencies. This characteristic is crucial for detecting weak signals even in noisy environments.
- Ensemble inference (soft voting): The average value of the softmax probabilities of the two models is calculated to provide the final prediction for the sensed signal classification. This soft voting strategy not only minimizes the variance of the predictions of individual models but also enhances overall segmentation accuracy. The ensemble probability can be calculated as:

$$P_{ensemble}(X) = \frac{1}{K}\sum_{k=0}^{K} P_k(X) \quad (5)$$

where K signifies the total number of models in the ensemble framework.

Algorithm 1 outlines the procedural steps of the proposed ensemble framework, and its flowchart is illustrated in Figure 3.

Algorithm 1: Deep Ensemble Semantic Segmentation Framework for Efficient Spectrum Sensing in Cognitive Radio Networks

Input:

D: Wideband spectrogram dataset containing pairs of RGB images (I) and ground-truth masks (M).

C: Set of semantic classes {Noise(0), 5G NR (1), LTE (2)}.

M: Ensemble models { $M_{DeepLab}$, M_{UNet} }.

Output:

$\hat{Y}$: Final segmentation masks for test images.

Data Preprocessing:

1. Split dataset D into training, validation, and test sets with a 54:13:33 ratio, yielding D_{train}, D_{val}, and D_{test} sets.
2. For each training mask Msk_i in D_{train}:
   - Compute global pixel count N_c for each class c.
   - Compute inverse class weights using median frequency balancing:
     $f_c \leftarrow \frac{N_c}{\sum_k N_k}$, $W_c \leftarrow \frac{\text{median}(\{f_0, f_1, f_2, \ldots, f_{C-1}\})}{f_c}$
3. Define augmentation pipeline A: horizontal flip, vertical flip.

Independent Model Training:

1. For each model m in M:
   - Initialize m with backbone f_θ (ResNet-18, pretrained on ImageNet).
   - Initialize optimizer AdamW ($\eta \leftarrow$ 1e-4).
   - Define weighted cross-entropy loss $L(y, \hat{y}) \leftarrow -\sum W_c * y\log(\hat{y})$
2. While epoch e < $E_{\max}$:
   - For each batch (x_b, y_b) in D_{train}:

- Map RGB pixel values to discrete class indices $c \in \{0,1,2\}$ of each training mask Msk_i in D_{train}.
- Apply augmentations: $x_b' \leftarrow A(x_b)$.
- Forward pass: $\hat{y}_b \leftarrow m(x_b')$.
- Compute loss: $J \leftarrow L(y_b, \hat{y}_b)$
- Update parameters using AdamW.

- Validate on D_{val} and checkpoint best weights.

Ensemble Inference:

1. Load best weights for $M_{DeepLab}$ and M_{UNet}.
2. For each test image x_{test} in D_{test}:
   - $P_1 \leftarrow \text{Softmax}\left(M_{DeepLab}(x_{test})\right)$
   - $P_2 \leftarrow \text{Softmax}\left(M_{UNet}(x_{test})\right)$
   - Ensemble fusion: $P_{ensemble} \leftarrow (P_1 + P_2)/2$
   - Generate mask: $\hat{Y}_i \leftarrow \text{argmax}(P_{ensemble})$
3. Return $\hat{Y}_i$.

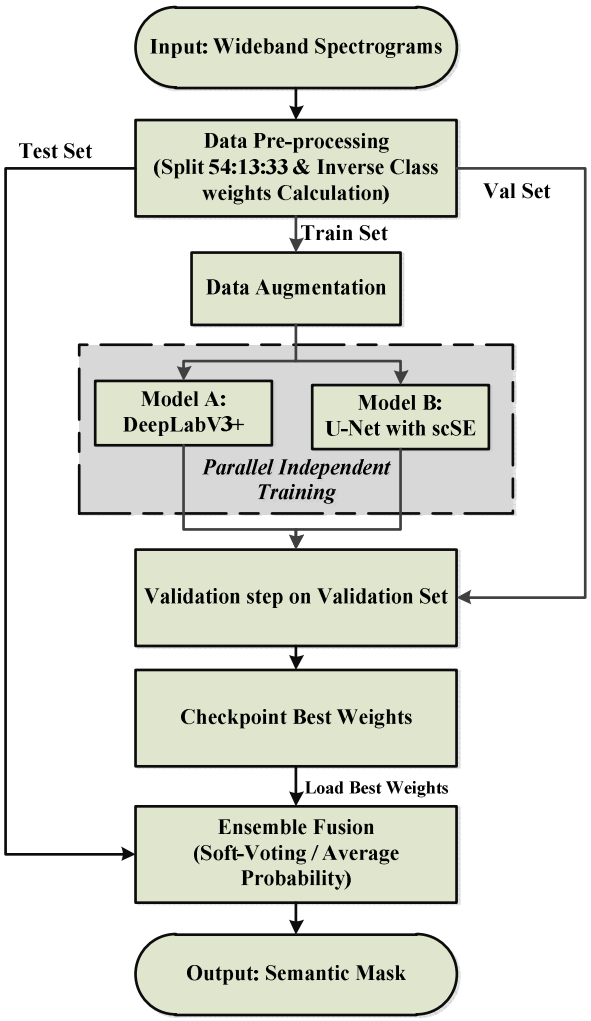


Fig. 3. Proposed deep ensemble semantic segmentation framework for efficient spectrum sensing.

IV. PERFORMANCE EVALUATION

In this research, we propose an ensemble framework comprising DeepLabV3+ and U-Net with scSE attention, aimed at identifying coexisting 5G NR and LTE signals through semantic segmentation of the spectrogram images. The performance of the framework was rigorously evaluated via extensive simulations on the Kaggle computational platform. The experimental setup utilized an NVIDIA Tesla T4 GPU accelerator. To determine its efficiency in detecting 5G NR and LTE signals, the proposed model was compared against three baseline architectures: PixelMLP, SimpleFCN, and SimpleSegNet. An identical experimental setup was maintained to ensure fair and consistent comparison. This involved uniform data preprocessing steps, such as remapping of labels, calculation of inverse class weights to mitigate the effect of class imbalance, and a consistent training-validation-testing splitting maintaining a ratio of 54:13:33.

Evaluation outcomes demonstrate that the proposed ensemble framework surpasses the baseline models on all major measures. Notably, the ensemble framework is exceptionally robust to different SNR conditions. Although baseline models can hardly detect signal boundaries in the presence of strong background noise, the proposed framework greatly reduces false positives and maintains the structural integrity of 5G NR and LTE signals, leading to a much higher mean Intersection over Union (mIoU). Figures 4(a) and 4(b) demonstrate the training stability and convergence behavior. These plots show that the DeepLabV3+ and U-Net (scSE) models efficiently converge to a high-accuracy state, avoiding the fluctuations observed in the baseline architectures.

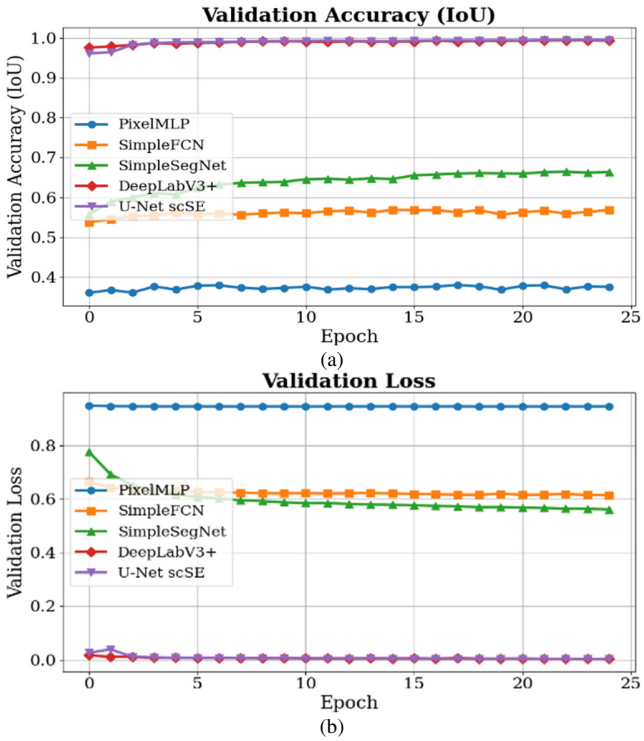
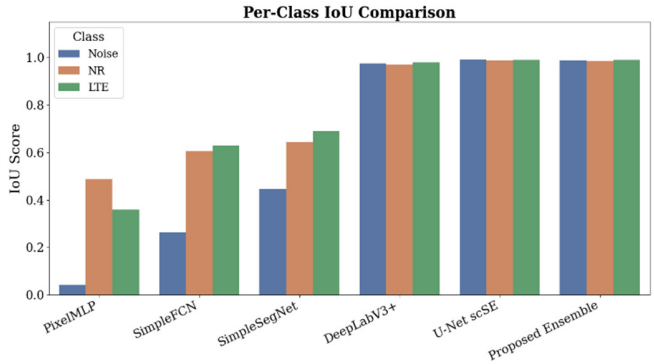
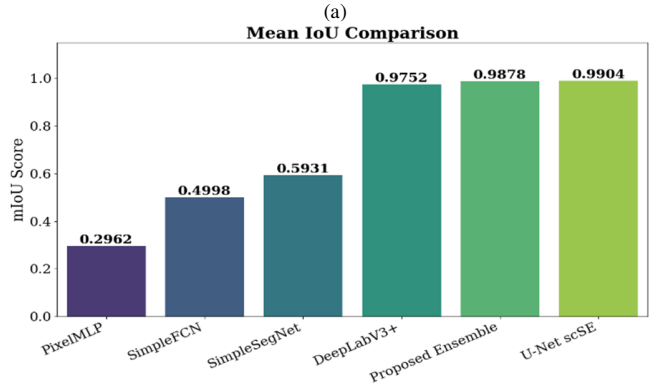


Fig. 4. Validation performance versus training epochs: (a) validation accuracy, and (b) validation loss.

Figure 5 illustrates the quantitative comparison, and it can be observed that the proposed framework achieves a much higher mIoU of 0.9878. Moreover, the class-wise analysis confirms the exceptional accuracy, with IoU scores of 0.9895 for 5G NR and 0.9853 for LTE, significantly higher than the baseline models. Although the U-Net enhanced with scSE performed better in terms of individual quantitative measures (mIoU = 0.9904), the proposed ensemble architecture (mIoU = 0.9878) was chosen as the ultimate inference engine to prioritize generalization and robustness. The ensemble approach mitigates the risk of architecture-specific overfitting by combining the multi-scale global context features of DeepLabV3+ with the fine-grained local attention mechanisms of U-Net with scSE. This marginal tradeoff in mIoU (0.0026) ensures greater reliability and decision stability by averaging diverse feature extraction pathways.



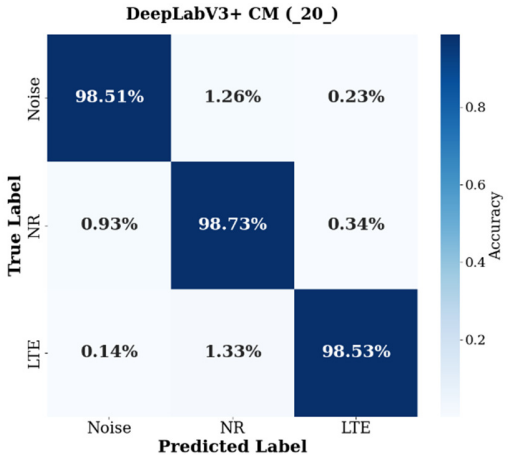
(a)



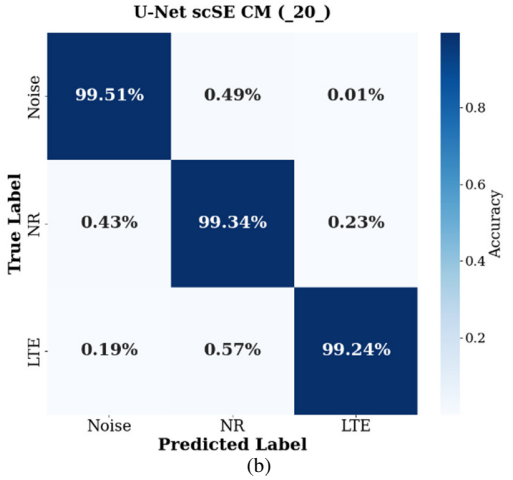
(b)

Fig. 5. Comparison of semantic segmentation performance of the proposed ensemble, its constituent models, and the baselines: (a) class-wise IoU, and (b) mean IoU (mIoU).

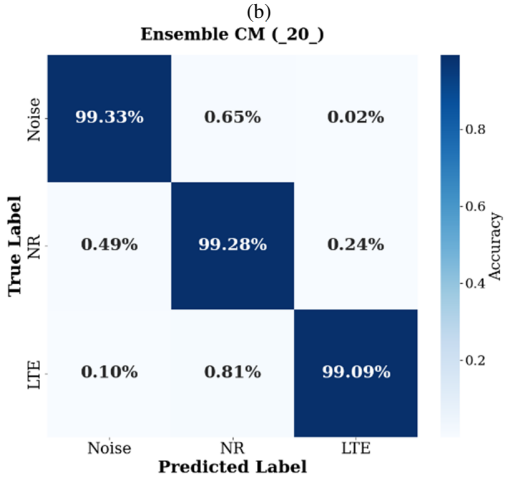
Figure 6 shows the confusion matrices of DeepLabV3+, U-Net using scSE attention, and the proposed ensemble at an SNR of 20 dB. The three architectures demonstrate strong diagonal dominance, outperforming SRNet [21], which typically suffers from high inter-class ambiguity. Notably, the ensemble model is highly efficient in detecting signals in noisy environments: it mitigates false positives by suppressing background noise while simultaneously reducing false negatives by accurately recovering faint 5G NR and LTE signals. The proposed ensemble framework correctly predicts 99.33% of the noise samples at 20 dB SNR, whereas SRNet [21] achieves an accuracy of 84.1% under the same conditions. The ensemble framework also correctly identifies 99.28% of 5G NR signals and 99.09% of LTE signals at 20 dB SNR, whereas SRNet [21] is limited to 25.9% and 11.1%, respectively.



(a)



(b)



(c)

Fig. 6. Confusion matrices illustrating classification performance of spectral content at 20 dB SNR: (a) DeepLabV3+, (b) U-Net with scSE, and (c) proposed ensemble framework.

A rigorous comparison between the proposed ensemble framework and existing state-of-the-art methods is provided in Table II. The proposed model clearly outperforms recent semantic segmentation methods, with 99.37% global accuracy and 99.39% mean F1-score. Therefore, the statistical robustness of the framework in noise-dominated environments stems from the combination of DeepLabV3+ ASPP for capturing global context and U-Net for recovering local boundaries. In addition, soft-voting fusion reduces prediction variance by averaging out stochastic errors, and inverse-class weighting reduces bias towards background noise. However, this comes with a slight increase in processing time compared to DPSegnet [5], which yielded 97.25% global accuracy with a processing time of 2.36 ms. So, the primary limitation of this study is the increased computational overhead and sensing time due to the model's complexity (26.6 M parameters), which may restrict deployment on resource-constrained edge nodes.

Figure 7 presents a qualitative comparison of segmentation results, displaying the input spectrogram alongside the ground truth and the proposed ensemble's prediction. The visualizations demonstrate high fidelity in distinguishing LTE, NR, and background noise. Notably, the predicted masks align precisely with the ground truth, effectively recovering signal blocks and suppressing artifacts even in complex, noisy environments.

TABLE II. PERFORMANCE COMPARISON OF PROPOSED ENSEMBLE AND STATE-OF-THE-ART MODELS

Model	Global acc (%)	Mean IoU (%)	Mean F1-score (%)	Params (M)	Time (ms)
BiSeNet [23]	72.74	53.73	52.29	49.12	7.63
DPSegnet [5]	97.25	94.76	-	7.0	2.36
SegFormer[16]	72.92	54.23	53.41	24.98	9.88
PRMNet [6]	92.44	87.60	93.35	14.5	16.5
SRNet [21]	72.96	55.05	53.85	19.32	5.41
Proposed ensemble framework	99.37	98.78	99.39	26.66	6.64

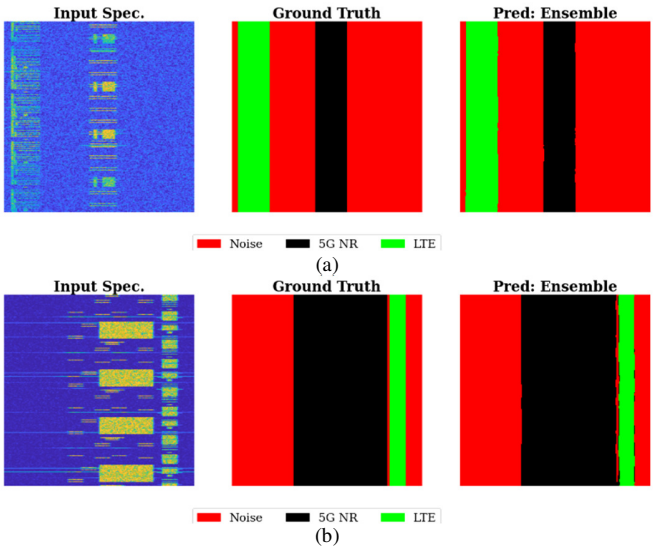


Fig. 7. Visualization of the spectrogram images, ground truth masks, and predicted masks of the proposed model at: (a) 20 dB SNR, and (b) 40 dB SNR.

V. CONCLUSION

This study proposes an ensemble framework based on a ResNet-18 backbone that effectively performs spectrum sensing in heterogeneous 5G networks. The primary novelty of this work lies in the synergistic integration of DeepLabV3+ and spatial and channel Squeeze-and-Excitation (scSE)–enhanced U-Net into a unified framework that formulates spectrum sensing as a semantic segmentation task. Unlike traditional approaches relying on single-stream architectures, this study presents a heterogeneous soft-voting ensemble designed to reconcile the trade-off between the multi-scale feature extraction of DeepLabV3+'s Atrous Spatial Pyramid Pooling (ASPP) and the fine-grained localization provided by U-Net's channel-wise attention. In terms of specific contributions, this framework establishes a variance-reducing soft-voting fusion strategy and validates the effectiveness of median frequency balancing to overcome severe class imbalance in Radio Frequency (RF) spectrograms. Rigorous evaluations demonstrate that the proposed ensemble framework has the capability to identify coexisting 5G New Radio (NR) and Long Term Evolution (LTE) signals more efficiently even in noisy environments, outperforming the baseline segmentation models. Quantitatively, the proposed ensemble framework achieves a mean Intersection over Union (mIoU) score of 0.9878, demonstrating significant performance improvements compared to the baseline models as well as state-of-the-art architectures from recent studies such as BiSeNet, DPSegnet, SegFormer, PRMNet, and SRNet.

In the future, we plan to validate the robustness of the framework by considering more complicated conditions such as higher-order modulation schemes and real-world hardware impairments. Additionally, we intend to evaluate its performance in real-world scenarios by implementing the ensemble framework on edge devices.

REFERENCES

[1] K. Kimani and M. Njiraine, "Cognitive Radio Spectrum Sensing Mechanisms in TV White Spaces: A Survey," *Engineering, Technology & Applied Science Research*, vol. 8, no. 6, pp. 3673–3680, Dec. 2018, https://doi.org/10.48084/etasr.2442.

[2] P. Venkatapathi, H. Khan, S. S. Rao, and G. Immadi, "Cooperative Spectrum Sensing Performance Assessment using Machine Learning in Cognitive Radio Sensor Networks," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12875–12879, Feb. 2024, https://doi.org/10.48084/etasr.6639.

[3] M. U. Muzaffar and R. Sharqi, "A review of spectrum sensing in modern cognitive radio networks," *Telecommunication Systems*, vol. 85, no. 2, pp. 347–363, Feb. 2024, https://doi.org/10.1007/s11235-023-01079-1.

[4] A. A. Raji and T. O. Olwal, "Spectrum Sensing in Cognitive Radio Internet of Things: State-of-the-Art, Applications, Challenges, and Future Prospects," *Journal of Sensor and Actuator Networks*, vol. 14, no. 6, Nov. 2025, Art. no. 109, https://doi.org/10.3390/jsan14060109.

[5] T.-T. Le, D. B. da Costa, and T. Huynh-The, "Efficient Spectrum Sensing via a Multi-Scale Dual-Path Segmentation Network," *IEEE Wireless Communications Letters*, vol. 14, no. 7, pp. 2134–2138, July 2025, https://doi.org/10.1109/LWC.2025.3564151.

[6] H.-T. Nguyen, H.-T. Phuoc Dang, Q.-V. Pham, and T. Huynh-The, "Resolution-Preserving Multi-Scale Network for 5G-LTE Spectrogram-Based Spectrum Sensing," *IEEE Wireless Communications Letters*, vol. 14, no. 6, pp. 1673–1677, June 2025, https://doi.org/10.1109/LWC.2025.3552193.

[7] Y. Zhang and Z. Luo, "A Deep-Learning-Based Method for Spectrum Sensing with Multiple Feature Combination," *Electronics*, vol. 13, no. 14, July 2024, Art. no. 2705, https://doi.org/10.3390/electronics13142705.

[8] L. Fu and S. Li, "A New Semantic Segmentation Framework Based on UNet," *Sensors*, vol. 23, no. 19, Sept. 2023, Art. no. 8123, https://doi.org/10.3390/s23198123.

[9] L. Nanni, A. Lumini, and C. Fantozzi, "Exploring the Potential of Ensembles of Deep Learning Networks for Image Segmentation," *Information*, vol. 14, no. 12, Dec. 2023, Art. no. 657, https://doi.org/10.3390/info14120657.

[10] W. Deng, X. Wang, and Z. Huang, "Semantic-Segmentation-Based Deep Spectrum Sensing for Cochannel Signals," *IEEE Internet of Things Journal*, vol. 11, no. 23, pp. 37645–37660, Dec. 2024, https://doi.org/10.1109/JIOT.2024.3436683.

[11] P. Subedi, S. Shin, and M. C. Vuran, "Seek and Classify: End-to-end Joint Spectrum Segmentation and Classification for Multi-signal Wideband Spectrum Sensing," in *2024 IEEE 49th Conference on Local Computer Networks*, Normandy, France, 2024, pp. 1–9, https://doi.org/10.1109/LCN60385.2024.10639826.

[12] D. Uvaydov, M. Zhang, C. P. Robinson, S. D'Oro, T. Melodia, and F. Restuccia, "Stitching the Spectrum: Semantic Spectrum Segmentation with Wideband Signal Stitching," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications*, Vancouver, Canada, 2024, pp. 2219–2228, https://doi.org/10.1109/INFOCOM52122.2024.10621332.

[13] V. Pallam, H. Khan, S. R. Surampudi, and G. Immadi, "Reduced Kernel PCA Model for Nonlinear Spectrum Sensing in Cognitive Radio Network," *Journal of The Institution of Engineers (India): Series B*, vol. 106, no. 1, pp. 181–187, Feb. 2025, https://doi.org/10.1007/s40031-024-01089-w.

[14] P. Venkatapathi, H. Khan, and S. S. Rao, "Performance Analysis of Spectrum Sensing in Cognitive Radio under Low SNR and Noise Floor," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 2, pp. 2655–2661, Dec. 2019, https://doi.org/10.35940/ijeat.F8703.129219.

[15] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention Mask Transformer for Universal Image Segmentation," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2022, pp. 1280–1289, https://doi.org/10.1109/CVPR52688.2022.00135.

[16] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: simple and efficient design for semantic segmentation with transformers," in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, Online, 2021, pp. 12077–12090.

[17] B. Zhang, L. Liu, M. H. Phan, Z. Tian, C. Shen, and Y. Liu, "SegViT v2: Exploring Efficient and Continual Semantic Segmentation with Plain Vision Transformers," *International Journal of Computer Vision*, vol. 132, no. 4, pp. 1126–1147, Apr. 2024, https://doi.org/10.1007/s11263-023-01894-8.

[18] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "PIDNet: A Real-time Semantic Segmentation Network Inspired by PID Controllers," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, Canada, 2023, pp. 19529–19539, https://doi.org/10.1109/CVPR52729.2023.01871.

[19] W. Wang *et al.*, "InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, Canada, 2023, pp. 14408–14419, https://doi.org/10.1109/CVPR52729.2023.01385.

[20] A. Kirillov *et al.*, "Segment Anything," in *2023 IEEE/CVF International Conference on Computer Vision*, Paris, France, 2023, pp. 3992–4003, https://doi.org/10.1109/ICCV51070.2023.00371.

[21] T. Huynh-The, G.-V. Nguyen, T.-H. Vu, D. Benevides da Costa, and Q.-V. Pham, "SRNet: Deep Semantic Segmentation Network for Spectrum Sensing in Wireless Communications," *IEEE Wireless Communications Letters*, vol. 14, no. 2, pp. 355–359, Feb. 2025, https://doi.org/10.1109/LWC.2024.3502003.

[22] T. Huynh-The. "ThienHuynhThe/SpectrumSensing5G." GitHub. https://github.com/ThienHuynhThe/SpectrumSensing5G.

[23] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation," in *Computer Vision – ECCV 2018: 15th European Conference*, Munich, Germany, 2018, pp. 334–349, https://doi.org/10.1007/978-3-030-01261-8_20.