Intelligent Spectrum Sensing with ConvNet for 5G and LTE Signals Identification

Thien Huynh-The[1, 5], Quoc-Viet Pham[2], Thai-Hoc Vu[3], Daniel Benevides da Costa[4], and Van-Phuc Hoang[5]
[1] Department of Computer and Communication Engineering, HCMC University of Technology and Education, Vietnam
[2] School of Computer Science and Statistics, Trinity College Dublin, Ireland
[3] Department of Electrical, Electronic and Computer Engineering, University of Ulsan, Republic of Korea
[4] Technology Innovation Institute, 9639 Masdar City, Abu Dhabi, United Arab Emirates
[5] Institute of System Integration, Le Quy Don Technical University, Vietnam
Emails: thienht@hcmute.edu.vn, viet.pham@tcd.ie, vuthaihoc1995@gmail.com,
daniel.costa@tii.ae, phuchv@lqdtu.edu.vn

2023 IEEE Statistical Signal Processing Workshop (SSP) | 978-1-6654-5245-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/SSP53291.2023.10208054

***Abstract*—The paper presents an intelligent spectrum sensing approach for next-generation wireless networks by exploiting deep learning, in which we develop a deep convolutional network (ConvNet) to automatically identify Fifth Generation New Radio (5G NR) and Long-Term Evolution (LTE) signals under standards-specified channel models with diversified RF impairments. In particular, we design a semantic segmentation ConvNet to detect and localize the spectral content of 5G NR and LTE in a synthetic signal featured by spectrum occupancy. A received signal is first converted by a short-time Fourier transform and represented as a wideband spectrogram image which is then passed through the ConvNet, incorporated by DeepLabv3+ and ResNet18 to improve the accuracy of pixel-wise segmentation to further increase the accuracy of signal identification. In the simulations, our ConvNet achieves around 95% mean accuracy and 91% mean intersection-over-union (IoU) at medium SNR level and demonstrates robustness under various practical channel impairments.**

***Index Terms*—5G NR, cognitive ration, deep learning, encoder-decoder architecture, signal identification, spectrum sensing.**

I. Introduction

Current wireless communication systems have witnessed an unprecedented increase in data transmission rate and volume caused by the growth of mobile users and the demand for high-quality services [1]. It is definitely a critical concern for wireless communication networks to manage and utilize the limited spectrum resource as efficiently and effectively as possible to meet all requirements from both users and service providers. Several solutions have been introduced, developed, and deployed to achieve the ultimate goal in next-generation wireless networks beyond the fifth generation (5G), including small cell optimization, mmWave band utilization, dynamic resource allocation, massive multi-input multi-output (MIMO), and cognitive radio technologies. As a form of wireless communication designed with intelligent communication channel detection and allocation, cognitive radio allows the dynamic sharing of spectrum among users (licensed and unlicensed) based on some advanced techniques like spectrum sensing [2] and signal identification [3]. In 6G, efficient spectrum monitoring, management, and allocation issues are more crucial than in previous network generations; for example, dynamic spectrum sharing enables parallel operation of 5G New Radio (NR) and Long-Term Evolution (LTE) in the same band. In this context, signal identification could be realized with the help of artificial intelligence (AI) [4], including conventional machine learning algorithms and modern deep learning (DL) architectures, without channel state information.

Thien Huynh-The was funded by the Postdoctoral Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2022.STS.04.

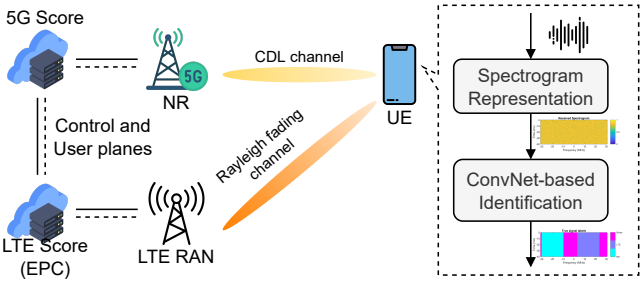


Fig. 1. A general LTE-NR wireless network architecture with an intelligent spectrum sensing.

In the last decades, several approaches have been introduced for intelligent spectrum sensing in cognitive radio networks, in which one of the most useful signal characteristics for signal identification is the modulation type [5]. Indeed, many recent modulation classification works were proposed with DL-based classifiers to discriminate received signals accurately based on their modulation types, including analog and digital fashions [6]. In [7], a ConvNet, namely MCNet, was proposed for modulation-based baseband signal identification, in which the authors exploited asymmetric convolution to learn spatiotemporal signal correlations. Some other works addressed the modulation classification problem in multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) systems with data reconstruction and three-dimensional convolution [8], [9]. As not only communication signals but also radar signals were studied, Pan et al. [10] revised spectrum sensing in cognitive radio with low-

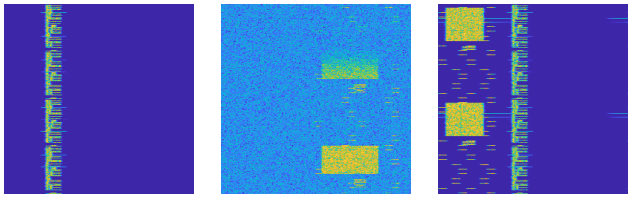
probability-of-intercept (LPI) waveform recognition to automatically identify radar sources using a multi-label deep network. Another work [11] optimized spectrum sensing in radar-communication coexistence systems by leveraging a high-performance waveform recognition method that accurately identifies simultaneously radar and communication signals based on their waveform. A combination of time-frequency analysis and deep learning [12] was introduced for spectrum sensing in cognitive radio, but the approach has been restricted to a binary classification of spectrum sensing detection.

Many previous spectrum sensing works have been introduced for simplified baseband signal models, but most of them do not really match the current realistic wireless communication systems (e.g., LTE and 5G NR). Consequently, this work introduces an intelligent spectrum sensing approach that enables the receiver to identify 5G NR and LTE signals at once in a wideband spectrogram by exploiting the advantages of deep learning in relevant feature extraction and high pattern discrimination. We consider extracting the underlying radio characteristics of a received signal in the time-frequency domain with a short-time Fourier transform (STFT) to generate its spectrogram image instead of processing it with the primitive complex envelope data. Inspired by semantic image segmentation in computer vision, we study a ConvNet to detect and localize multiple objects as the spectral contents of 5G NR and LTE in the wideband spectrogram of synthetic signals. Our semantic segmentation ConvNet is designed by incorporating DeepLabv3+ (an efficient encode-decoder architecture for pixel-level classification) and ResNet18 (as a backbone network for feature extraction). In the simulations, the performance of the designed ConvNet is evaluated on a spectrogram image dataset of 5G NR and LTE signals generated in the presence of RF channel impairments.

II. Methodology

In this work, our primary purpose is to present an efficient spectrum sensing method for identifying 5G NR and LTE signals over spectral content in a wideband spectrogram by exploiting deep learning techniques. To this end, we build a deep convolutional neural network that is capable of characterizing spectrum occupancy in the next generation of wireless communication. Inspired by semantic image segmentation to identify and localize visual objects in images or video in computer vision, our approach familiarizes the signal identification problem in wireless signal processing as the discrimination of frequency and time occupied by 5G NR and LTE signals in the physical layer. As the general system model shown in Fig. 1, an intelligent spectrum sensing method is developed to discriminate the 5G NR and LTE spectral elements in time and frequency in synthetic signals. In detail, the method has two modules: spectrogram representation and ConvNet-based identification.

Spectrogram representation: We adopt STFT to calculate the spectrogram of a baseband signal using the Hann window with overlapping between adjoining segments. The magnitude squared of STFT, realized as the spectrogram time-frequency representation of a signal, should be scaled to the interval $[0, 1]$ for pixel intensity normalization and then converted to a color image. Some examples of spectrogram images of 5G NR and LTE signals are illustrated in Fig. 2.



(a) (b) (c)

Fig. 2. Spectrogram images of: (a) LTE signal, (b) 5G NR signal, and (c) synthetic LTE-NR signal.

ConvNet-based identification: To perform spectral-based signal identification, we intend to build a semantic segmentation neural network with an encoder-decoder architecture based on DeepLabv3+ [13] with ResNet18 [14] featured as the backbone network. As shown in Fig. 3, the deep network involves two parts: the encoder is to extract relevant-meaningful visual information with the spatial dimension reduction of feature maps, and the decoder aims to recover object details and spatial dimension via an upsampling mechanism. Remarkably, in the encoding flow, there is one Atrous Spatial Pyramid Pooling (ASPP) [15], which adopts parallel atrous convolution layers (a.k.a., dilated convolution) with different rates (i.e., dilation factors) to capture multi-scale information, corresponding to different sizes of the receptive field without increasing trainable parameters and computation. Regular convolution is an exceptional case of atrous convolution with a rate of 1. Inspired by the operation efficiency of the previous spatial pyramid pooling [16], ASPP is able to appraise features at different scaling representations and classify competently at any arbitrary scale. In the structural perspective, ASPP is arranged by four convolution layers, specialized by the kernels of size 1×1 and 3×3 with the rates of $\{1, 6, 12, 18\}$, followed by batch normalization layers and activation layers, for parallel processing. The outputs of ASPP are then gathered via a depth-wise concatenation layer that can be expressed as follows:

$$\mathbf{F}_{\mathrm{ASPP}} = \left\langle \mathcal{A}^{1}_{1\times1}\left(\mathbf{F}\right), \mathcal{A}^{6}_{3\times3}\left(\mathbf{F}\right), \mathcal{A}^{12}_{3\times3}\left(\mathbf{F}\right), \mathcal{A}^{18}_{3\times3}\left(\mathbf{F}\right)\right\rangle, \quad (1)$$

where $\mathcal{A}^{\alpha}_{n\times n}$ denotes an operation sequence of atrous convolution-batch normalization-relu with the kernels of size $n \times n$ and the rate α, $\mathbf{F}$ is the input of ASPP, and $\langle\cdot\,\rangle$ symbolizes the depth-wise concatenation. $\mathbf{F}_{\mathrm{ASPP}}$ is then passed through a point-wise convolution layer before upsampling.

In the encoder, we analyze and extract features with an output stride of 16, therefore, they should be upsampled bilinearly (i.e., sliding filters horizontally and vertically) by a factor of 16 in the decoder to generate the segmented output with a size identical to that of the input. In particular, there are two upsampling processes characterized by two transposed convolution layers, in which each upsampling should be accomplished with a factor of 4. The first upsampling output is concatenated with the low-level features acquired from the

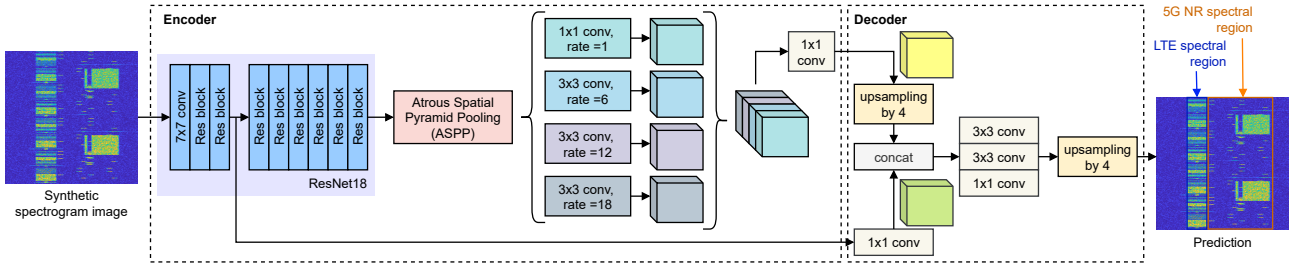


Fig. 3. Deep semantic segmentation network for spectrogram-based signal identification using DeepLabv3+ with ResNetv18 as the backbone network in the encoder and ASPP in the decoder.

TABLE I
SEGMENTATION ACCURACY VERSUS SNR

SNR (dB)	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
0	0.3820	0.3368	0.1647	0.1875	0.2955
20	0.3989	0.3401	0.1677	0.1960	0.3002
40	0.7503	0.7134	0.5762	0.5989	0.4666
60	0.9562	0.9481	0.9132	0.9163	0.8027
80	0.9722	0.9693	0.9460	0.9460	0.8585
100	0.9733	0.9715	0.9487	0.9480	0.8618
≥ 40	0.9120	0.8984	0.8334	0.8391	0.7452
Average	0.7340	0.7077	0.5677	0.5783	0.5926

network backbone, and passed through a 1×1 convolution. Continuously, we adopted two 3×3 convolutions to sharpen features and one 1×1 convolution to align the depth dimension (i.e., to be identical to the depth dimension of the network input) before upsampling by a factor of 4. The network is finalized with a softmax layer and a pixel-wise classification layer that allow the yielding of a segmentation mask localizing the spectral regions of noise, 5G NR, and LTE for signal identification.

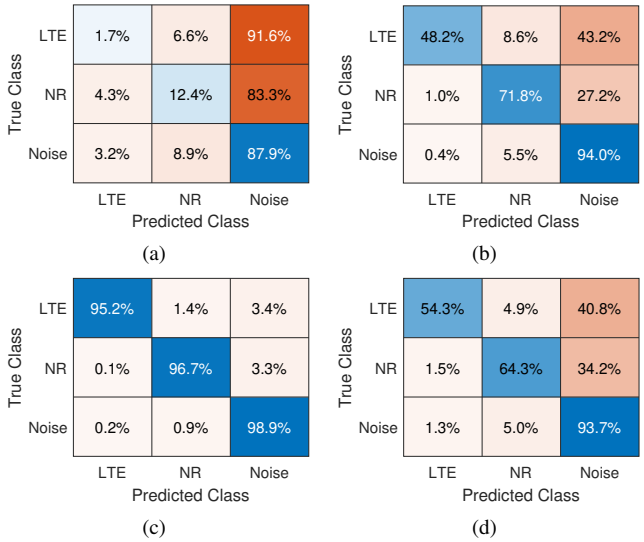


Fig. 4. Confusion matrices of spectral content segmentation at various SNR regimes: (a) 20 dB, (b) 40 dB, (c) 80 dB, and (d) all dB levels.

III. Simulation Results

A. Dataset and Simulation Setup

To evaluate the performance of the proposed signal identification, we generated spectrogram image datasets of 5G NR and LTE signals with highly reliable channel and RF impairment configurations. In detail, we generate two datasets: the training set, where signal frames contain either 5G NR or LTE signals (each frame has 40 subframes corresponding to 40 ms) and the test set, where frames encompass 5G NR and LTE signals on randomly distinctive bands without a band of interest. The sampling rate is set to 61.44 MHz as high as enough to process the latest standard signals and various low-cost software-defined radio systems (i.e., supplying up to 50 MHz of useful bandwidth). Some common 5G NR signal parameters include the bandwidth $[10, 15, 20, 25, 30, 40, 50]$, the sub-carrier spacing $[15, 30]$, synchronization signal block (SSB) period 20 ms, and SSB `Case A` and `Case B` patterns [17]. For LTE signals, we tune some parameters as follows: the reference channel $[\mathrm{R}.2, \mathrm{R}.6, \mathrm{R}.8, \mathrm{R}.9]$, the bandwidth $[5, 10, 15, 20]$, and the frequency-division duplexing transmission mode [18]. To simulate the real-world phenomena having diversified channel impairments in wireless communications, we pass 5G NR signals through a clustered delay line (CDL) multi-input multi-output (MIMO) link-level fading channel and LTE signals through a MIMO multipath Rayleigh fading channel [19], where SNR and Doppler are configured in the ranges of $[0, 20, 40, 60, 80, 100]$ dB and $[0, 70, 750]$ Hz, respectively. We then apply STFT with the FFT length of $4,096$ to compute spectrogram images with the size of 256×256 from complex baseband signals (it can vary image size depending on the computation capacity of the system). We generate $4,000$ frames of 5G NR only and LTE for training and validation with a split ratio of $80/20$, and $2,000$ frames of 5G NR-LTE coexistence for performance measurement in the test phase.

For deep network training configurations, we set the optimizer with the stochastic gradient descent with momentum (SGDM) optimization algorithm, the maximum number of epochs 40, the mini-batch size 40, the learning rate 0.02 with a reduction of 10 times after 20 epochs. Due to the imbalance of the training set (i.e., 5G NR signals usually have a larger bandwidth than LTE signals, and noise fills the background (see samples in Fig. 2), we leverage class weighting to mitigate the learning bias issue caused by observation imbalance.

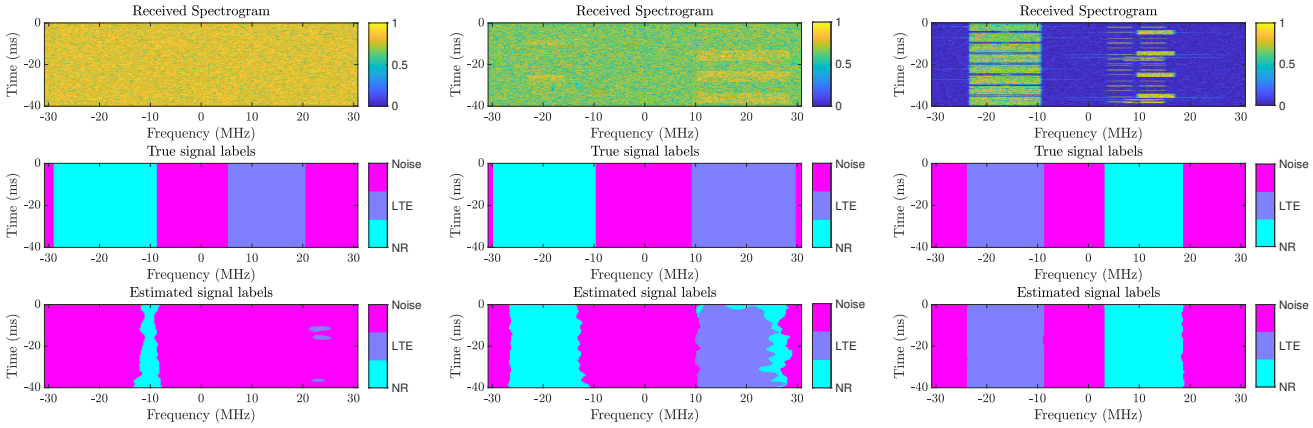


Fig. 5. Visualization of the received spectrum (of a synthesized signal), true labels, and predicted labels at various SNR regimes: (a) 20 dB, (b) 40 dB, and (c) 80 dB.

B. Results and Discussions

We measure five metrics, including global accuracy, mean accuracy, mean intersection-over-union (IoU), weighted IoU, and mean boundary F1 (BF) score, which are widely used to evaluate the accuracy of semantic segmentation. Table I shows the segmentation accuracy versus SNR of the deep network on the test set. It can be observed that the accuracy metrics, IoU metrics, and BF score increase along with the growth of SNR. Remarkably, it witnesses a shocking accuracy increase for all metrics, from 20 dB to 40 dB and 60 dB, for example, around 0.351 and 0.206 for global accuracy. For the whole range of SNR, the network obtains an overall accuracy of 0.734 because of high misclassification at low SNR regimes. In the case of $\mathrm{SNR} \geq 40$ dB, the global accuracy can reach more than 0.91.

Fig. 4 presents the confusion matrices of multi-class (i.e., noise, LTE, and 5G NR) pixel-level image segmentation at different SNRs. At 20 dB, the correct classification results of 5G NR and LTE are extremely poor due to the significant signal quality distortion caused by RF impairments and additive noise. In Fig. 4(a), a massive number of LTE and 5G NR pixels in spectrogram images are misclassified in the noise category. At 40 dB and 80 dB in Fig. 4(b) and 4(c), respectively, the classification accuracy of LTE and 5G NR improved comprehensively (e.g., the percentage of correct segmentation of LTE increases by two times). Notably, Fig. 4(d) reports the overall results of the whole test set. For visualization results, we show three examples of received spectrograms, along with true signal labeling mask and an estimated signal labeling mask done by the deep network, at different SNR regimes in Fig. 5. At low SNR, the meaningful spectral content of 5G NR and LTE in the received spectrogram cannot be identified because of the rude effects of additive noise and other RF impairments (e.g., clustered delay and multipath Rayleigh fading); thus, our deep model conducts many pixel-level classification errors in the estimated labeling mask. When increasing SNR, the segmentation quality of output masks is greatly enhanced, along with the improvement of quantitative metrics.

IV. Conclusions

In this paper, we have presented an intelligent spectrum sensing approach to identify 5G NR and LTE signals in next-generation wireless networks by exploiting the advancement of deep learning techniques. To discriminate the types of received signals that are characterized by spectrum occupancy at the receiver, our approach transforms the raw signal from the complex envelope form to a spectrogram image using STFT before feeding it to a deep semantic segmentation network. The deep network is designed by incorporating DeepLabv3+ and ResNet18 with ASPP as a sophisticated-designed module to improve pixel-level classification of spectrogram images. Relying on simulation results, the approach demonstrated the efficiency of the deep network in spectrogram-based signal identification and its robustness under different RF impairments. In the future, we will attempt to: (i) investigate the influence of networks' hyper-parameters on the overall accuracy; (ii) measure the complexity and optimize it for resource-constrained devices; and (iii) extend the pool of RF signals (such as satellite and radar) for discrimination.

References

[1] S. D. Liyanaarachchi, T. Riihonen, C. B. Barneto, and M. Valkama, "Optimized waveforms for 5g–6g communication with sensing: Theory, simulations and experiments," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 8301–8315, Dec. 2021.

[2] J. Gao, X. Yi, C. Zhong, X. Chen, and Z. Zhang, "Deep learning for spectrum sensing," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1727–1730, Dec. 2019.

[3] S. Haykin and P. Setoodeh, "Cognitive radio networks: The spectrum supply chain paradigm," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 1, pp. 3–28, Mar. 2015.

[4] H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, and K. Wu, "Artificial-intelligence-enabled intelligent 6g networks," *IEEE Network*, vol. 34, no. 6, pp. 272–280, Nov./Dec. 2020.

[5] Q.-V. Pham, N. T. Nguyen, T. Huynh-The, L. Bao Le, K. Lee, and W.-J. Hwang, "Intelligent radio signal processing: A survey," *IEEE Access*, vol. 9, pp. 83 818–83 850, 2021.

[6] T. Huynh-The, Q.-V. Pham, T.-V. Nguyen, T. T. Nguyen, R. Ruby, M. Zeng, and D.-S. Kim, "Automatic modulation classification: A deep architecture survey," *IEEE Access*, vol. 9, pp. 142 950–142 971, 2021.

[7] T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "MCNet: an efficient CNN architecture for robust automatic modulation classification," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 811–815, 2020.

[8] T. Huynh-The, T.-V. Nguyen, Q.-V. Pham, D. B. da Costa, and D.-S. Kim, "MIMO-OFDM modulation classification using three-dimensional convolutional network," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6738–6743, Jun. 2022.

[9] T. Huynh-The, T.-V. Nguyen, Q.-V. Pham, D. B. da Costa, G.-H. Kwon, and D.-S. Kim, "Efficient convolutional networks for robust automatic modulation classification in OFDM-based wireless systems," *IEEE Systems Journal*, vol. 17, no. 1, pp. 964–975, Mar. 2023.

[10] Z. Pan, S. Wang, M. Zhu, and Y. Li, "Automatic waveform recognition of overlapping LPI radar signals based on multi-instance multi-label learning," *IEEE Signal Processing Letters*, vol. 27, pp. 1275–1279, Jul. 2020.

[11] T. Huynh-The, Q.-V. Pham, T.-V. Nguyen, D. B. da Costa, and D.-S. Kim, "Racomnet: High-performance deep network for waveform recognition in coexistence radar-communication systems," in *Proc. ICC 2022 - IEEE International Conference on Communications*, Seoul, South Korea, May 2022, pp. 1–6.

[12] Z. Chen, Y.-Q. Xu, H. Wang, and D. Guo, "Deep STFT-CNN for spectrum sensing in cognitive radio," *IEEE Communications Letters*, vol. 25, no. 3, pp. 864–868, Mar. 2021.

[13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. European conference on computer vision (ECCV)*, Munich, Germany, Sep. 2018, pp. 801–818.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[15] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[17] 3GPP TS 38.101-1, "User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone," 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, Technical Specification (TS), Dec. 2019, version 15.8.0.

[18] 3GPP TS 36.101, "Evolved Universal Terrestrial Radio Access (E-UTRA);User Equipment (UE) radio transmission and reception," 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, Technical Specification (TS), Sep. 2019, version 15.8.0.

[19] 3GPP TS 38.901, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, Technical Specification (TS), Sep. 2019, version 15.1.0.