2022 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom) | 978-1-6654-9749-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/BLACKSEACOM54372.2022.9858132

Application of Object Detection Approaches on the Wideband Sensing Problem

Sefa KAYRAKLIK*†, Yusuf ALAGÖZ*, Ahmet Faruk COŞKUN*
*Informatics and Information Security Research Center (BİLGEM), TÜBİTAK, Kocaeli, Turkey
†Department of Electrical and Electronics Engineering, Koç University, İstanbul, Turkey
Emails: {sefa.kayraklik,yusuf.alagoz,ahmet.coskun}@tubitak.gov.tr

***Abstract*—Wideband spectrum sensing (WBS) has been a critical issue for communication system designers and specialists to monitor and regulate the wireless spectrum. Detecting and identifying the existing signals in a continuous manner enable orchestrating signals through all controllable dimensions and enhancing resource usage efficiency. This paper presents an investigation on the application of deep learning (DL)-based algorithms within the WBS problem while also providing comparisons to the conventional recursive thresholding-based solution. For this purpose, two prominent object detectors, You Only Learn One Representation (YOLOR) and Detectron2, are implemented and fine-tuned to complete these tasks for WBS. The power spectral densities (PSDs) belonging to over-the-air (OTA) collected signals within the wide frequency range are recorded as images that constitute the signal signatures (i.e., the objects of interest) and are fed through the input of the above-mentioned learning and evaluation processes. The main signal types of interest are determined as the cellular and broadcast types (i.e., GSM, UMTS, LTE and Analogue TV) and the single-tone. With a limited amount of captured OTA data, the DL-based approaches YOLOR and Detectron2 are seen to achieve a classification rate of 100% and detection rates of 85% and 69%, respectively, for a nonzero intersection over union threshold. The preliminary results of our investigation clearly show that both object detectors are promising to take on the task of wideband signal detection and identification, especially after an extended data collection campaign.**

***Index Terms*—Pervasive Artificial Intelligence, Wideband Spectrum Sensing, Object Detection and Classification Methods, YOLOR, Detectron2**

I. Introduction

Thanks to the developing and expanding nature of communication technologies, many civil and military applications have been implemented in smarter, more accurate, and more practical manners. In line with the different requirements of applications such as data rate, accuracy, and privacy, the physical layer designs of communication systems are shaped using different center frequencies, bandwidths, and waveforms. The aforementioned differences in requirements result in enormous heterogeneity in the overall signal spectrum that should be seriously handled both by a communication system designer who would pay attention to predicting adjacent channel effects and to provide designs robust to these effects and for communication intelligence and spectrum allocation specialists who would prioritize monitoring and even inspecting activities in the entire spectrum. In terms of serving this purpose, the concept of wideband sensing (WBS) allows the detection of signals in the entire frequency spectrum in a simple sense and classification in addition to detection in an extended sense. In this respect, the WBS approach has turned into a tool used in many civilian and military applications, as it has been met with great interest by both the cognitive spectrum sensing (CSS) and electronic warfare (EW) communities.

As the leading part of the CSS and EW-oriented applications, WBS approaches aim to achieve accurate detection of multiple signals within the entire signal spectrum of the interest, to either feed the signal recognition steps with the most proper inputs or to perform the ultimate classification. The detection of multiple signals in a wideband reception scenario has been formerly introduced by several conventional signal processing methods such as the energy-detection [1], the cyclostationary feature extraction [2], matched filtering [3], and the recursive signal identification with adaptive thresholding [4]. As already mentioned in [5], [6] and the referring research articles, the mentioned signal detection methods have been shown to suffer from several defects such as the strict exigence on the priori signal information, the limited usage at low signal-to-noise ratios (SNR), and the computational complexity. On the other hand, as the third-wave communications technologies (i.e., the fifth generation and beyond) become more prevalent through massive machine-type communications and Internet of Things concepts, network terminals with limited transmit power and signal processing capability would be stimulated to sense the existing frequency spectrum (i.e., the detection and recognition of the existing signals) as fast and accurate as possible. At this point, as in many application areas, the artificial intelligence and especially deep-learning (DL)-based approaches are seen to provide reliable and evolving solutions in severe cases where conventional tools have dramatically failed. Hence, the DL-based methods are expected to be the indispensable components of 5G and beyond communication systems.

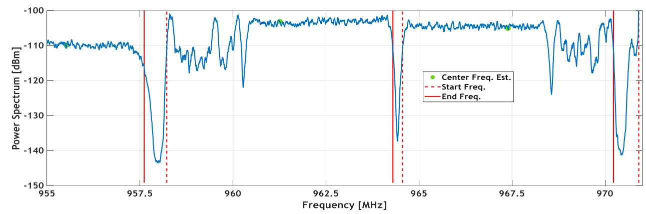
The literature on DL-based WBS consists of a few research articles and investigations provided by [7]- [14]. The examination of [7] has focused on the identification of binary frequency-shift keying (BFSK), 4FSK, phase-shift keying (PSK), quadrature amplitude modulation (QAM), Morse, speech and resident noise by using the short-time Fourier transform (STFT) image generated for a duration of 5 s and the maximum frequency of 125 kHz. The comparisons between centerline-based network (CLN), Faster R-CNN [15], and single-shot multi-box detector (SSD) [16] are also provided that clearly exhibit the superiority of the SSD in terms of processing speed while it achieves classification accuracy closer to those of CLN and Faster R-CNN. The same modulation identification case has been investigated by the authors of [8] by using the STFT images belonging to a different modulation set consisting of BPSK, QPSK, Offset QPSK, 8PSK, 16QAM, 64QAM, and amplitude-phase shift keying signals of levels 16 and 64 (i.e., 16APSK and 64APSK). The reference [9] has provided a framework for the MATLAB-based generation of IEEE 802.11n high throughput (also known as Wi-Fi HT protocol) signals operating at 2.45 GHz at the ISM band with a duration of 630 ms, a wideband bandwidth of 56 MHz, and an SNR drawn from the set 0, 10, 20, 30 dB, and has performed blind localization in the time-frequency grid with the help of STFT imaging feeding a Faster R-CNN model. The authors of [9] have extended their analyses by including the interference effects caused by the Bluetooth Low-Energy (BLE) signals and microwave ovens [10]. A DL-based approach aiming to jointly detect, localize and classify the radio-frequency (RF) signals has been introduced in [11], which utilizes the wideband spectrograms as the input of the state-of-the-art object detector YOLOv3 [17]. In [11], by constructing the STFT images with the time and frequency ranges of 40 ms and 25 MHz belonging to a synthetically generated dataset consisting of both the analogue modulations SSB-AM, DSB-AM, WB-FM, NB-FM, and the digital ones 2-FSK, 4-FSK, 8-FSK, 16-FSK, GMSK (i.e., Gaussian minimum shift keying), and complex Pulse Amplitude Modulation (PAM), joint signal detection, time-frequency localization, and classification has been performed in addition to the comparisons to a former DL approach Faster R-CNN. The paper has emphasized the efficiency of SSD-type object detectors (e.g., YOLOv3) by exhibiting remarkable average precision (AP) and intersection over union (IoU) rates. Besides, the authors have pointed out the necessity of an investigation on the real-world RF data by promising a future study in Conclusions section. To the best of the authors' knowledge, the literature lacks the investigation on WBS from both detection-only and detection&classification perspectives by using the state-of-art object detection schemes on real-world over-the-air (OTA) signal recordings of Global System for Mobile communications (GSM), Universal Mobile Telecommunications System (UMTS), Long-Term Evolution (LTE), Analogue television and single-tone signals. By also employing the conventional recursive and adaptive thresholding-based detection method proposed in [4] as a benchmark algorithm, the authors have exhibited the efficiency of You Only Learn One Representation (YOLOR) and Detectron2 approaches in the WBS problem.

II. Conventional Wideband Sensing Approaches

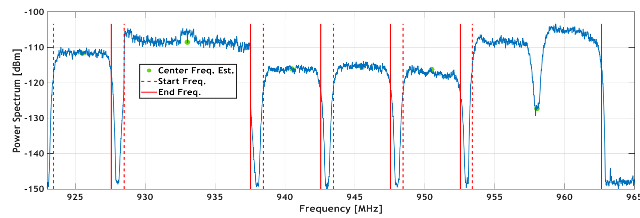
As mentioned in Introduction, WBS approaches have been serving several design perspectives, such as cognitive radio (CR), spectrum management, and communications intelligence (COMINT). For all cases, the conventional signal detection approaches have basically focused on the binary hypothesis for a specified frequency band given as

$$\mathcal{I}_{S(f)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 \tag{1}$$

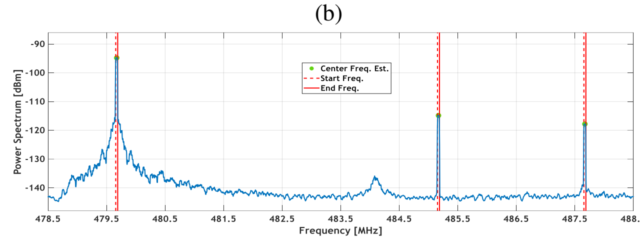
where $\mathcal{H}_0$, $\mathcal{H}_1$, and η_0 denote the absence and presence of the signal for the given PSD $S(f)$ and the detection threshold, respectively. Several thresholding-based approaches have formulated various solutions to these parameters. Nevertheless,



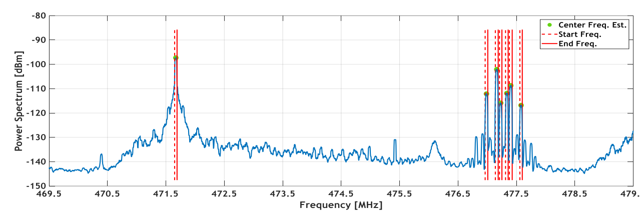
(a)



(b)



(c)



(d)

Fig. 1: Examples of detection impairments on (a) LTE, (b) UMTS, and (c)(d) Analogue TV signals

the conventional WBS approaches have suffered from several flaws in certain conditions that are sketched in Fig. 1. Each sub-figure illustrates different kinds of detection flaws that have been faced by using the conventional recursive thresholding (CRT) method on the collected test dataset. Here, the red dashed and solid lines represent the start and end frequencies of the detected signal, and the green circled marker shows the center frequency estimation. Fig. 1-(a) depicts the PSD characteristics of the signals existing within 955 MHz - 971 MHz, which is occupied by commonly GSM, UMTS and LTE signals. As can be seen from two signal estimates corresponding to the frequency intervals of $[958.1, 964.2]$ MHz and $[964.4, 970.2]$ MHz, the CRT method has mistakenly resulted in a combined signal spectrum estimate in both cases due to the neighboring leakage and the disability of recognizing the PSD discrepancies between GSM signals ($[958.1, 960.3]$ MHz), and LTE signals ($[960.4, 964.2]$ MHz). The adaptive thresholding mechanism has encountered another combined estimate problem for UMTS signals ($[953, 963]$ MHz), as shown in Fig. 1-(b). Another sub-figures sketch two different

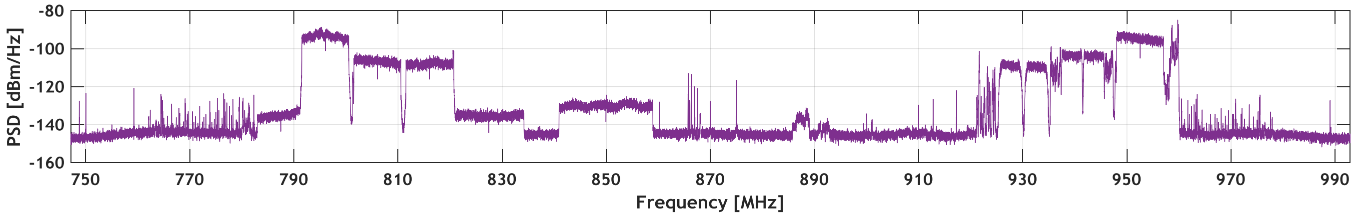


Fig. 2: An example of a wideband spectrum at a middle frequency of 870 MHz through 250 MHz bandwidth

estimation imperfections faced while using the CRT method on Analogue TV signals (in PAL norms), which mainly overlay on a maximally 8-MHz frequency span and consist of a video message part followed by a chrominance data part and an audio message part. The detection estimates demonstrated in Fig. 1-(c) clearly show that the conventional technique captures the video and audio carrier peaks however lacks to point out the chrominance data peak at 484 MHz. Fig. 1-(d) depicts another signal spectrum and detection estimates of an Analogue TV broadcasting scenario where the chrominance data peak is not intercepted, and the FM-modulated audio band yields multiple nearby peaks. As seen in 1-(c) and (d), one can easily realize that apart from the fact that the chrominance data peaks are left out by the detection process, a complementary association algorithm that might be employed at the detector output to check the detected peaks to fit within the PAL format would even substantially fail. After examining the defects of the CRT method for GSM, UMTS, LTE, and Analogue TV signals and doing further reasoning on the possible insufficiencies of similar thresholding-based approaches, it would be clear that the WBS problem should be taken into account with the help of state-of-art pattern recognition tools.

III. State-of-art Object Detectors: YOLOR and Detectron2

Object detection, whose aims are to classify and localize objects in a given image, is one of the vigorous fields in computer vision area. This field has a huge variety of applications, from detecting cancer cells in magnetic resonance imaging (MRI) results to classifying and locating enemy threads in synthetic aperture radar (SAR) images whereas, within the scope of this paper, the main objective is the detection, classification, and localization of OTA signals on the images that demonstrate the wideband PSD recordings.

With the help of DL-based structures, object detection algorithms have been more successful in the applications of complex scenarios, and the detection speed is assured to meet the real-time applications without sacrificing accuracy [18]. The most prominent object detector solutions are YOLOR [19], which is a modified version of YOLO [20], and Detectron2 [21], which is an open-source library developed by Facebook AI Research (FAIR).

YOLOR is an improved version of YOLO, whose primary purpose is to handle the whole image as a network input and directly provide the class probabilities and locations of the bounding boxes. YOLO contains a unified architecture of a single end-to-end neural network, which divides each input image into $S \times S$ grids whose responsibility is to detect an object if its center falls into that grid cell. YOLOR introduces the concept of implicit and explicit knowledge to the YOLO architecture, which enables the trained model to learn a general representation. In YOLOR, the ways to model implicit knowledge can be characterized by a vector z, neural network Wz, or matrix factorization $Z^T c$. Through the general representation and multitask learning applied by combining implicit and explicit knowledge, YOLOR achieves better performance and generalization capability.

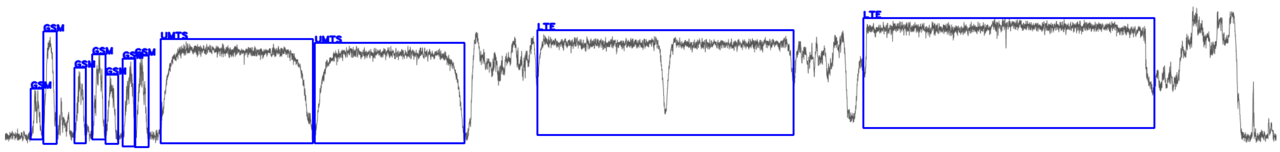
Detectron2 is FAIR's next-generation open-source platform for object detection and segmentation, and it is provided in a modular form to assist rapid implementation and evaluation of future works on computer vision research. There exist implementations for many object detection algorithms in Detectron2, such as Mask R-CNN, RetinaNet, Faster R-CNN, TensorMask, etc.

IV. The Proposed DL-Based WBS

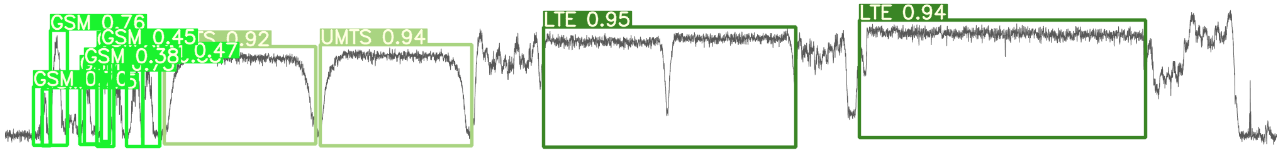
The comparative study provided in this paper aims to substitute the DL-based object detectors (i.e., YOLOR and Detectron2), which had already proven their efficiency in miscellaneous detection& identification problems into the WBS scenarios, and to provide insights for further research. Time domain samples of the OTA signals are collected by using a commercial off-the-shelf (COTS) receiver hardware with 250 MHz instantaneous bandwidth, and their Power Spectral Density (PSD)s are calculated in order to represent the signals in the frequency domain and used to detect and classify the OTA signals. Fig. 2 illustrates an example of the PSD images corresponding to a signal recording at 870 MHz. Each signal of interest is labeled as an object in the PSD images via its bounding box. The PSDs of the wideband spectrum are saved as images to be fed to the object detection algorithms.

A. Data Collection and Representation

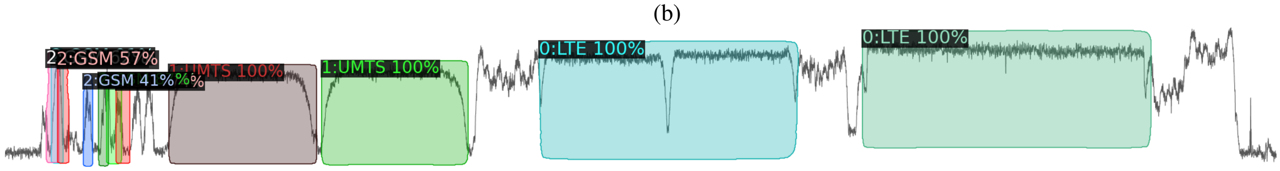
Time domain samples of OTA signals are collected in a wooded suburban area of the Gebze campus of TÜBİTAK. Firstly, time domain samples of signals within four different center frequencies of $\{470, 870, 1850, 2100\}$ MHz through 250 MHz bandwidth are collected via a COTS receiver unit. The OTA signals are captured during a 2.2 ms observation period with the sampling rate of $F_s \triangleq 250$ MHz which corresponds to a 557056-length signal instance for both the inphase and quadrature (IQ) dimensions after the baseband conversion. Then, the corresponding PSD is calculated by using Welch's method [22] with an FFT size of $N = 2^{16} = 65536$.



(a)



(b)



(c)

Fig. 3: (a) The labels, (b) YOLOR estimations, and (c) Detectron2 estimations related to an OTA record at 870 MHz

The recorded signals are processed by a Hanning window with a size equal to the FFT size and a normalized overlapping ratio of 0.5. Hence, the resulting frequency resolution is 250 MHz/65536 = 3.8147 kHz. Finally, each PSD plot is saved as a grayscale image in a PNG file with 12000×300 pixels. The dataset of time domain IQ samples and labeled PSD images are available from the authors upon request.

B. Signals of Interest

In the communication frequency bands, the investigation has been focused on five different signals of interest: Analogue television (Analogue TV) broadcasts in PAL norms in 470 MHz band; LTE with 10 and 20 MHz bandwidth in 870, 1850, and 2100 MHz bands; UMTS in 870, and 2100 MHz bands; GSM in 870 MHz band; single tone signals along the whole spectrum. PSD images with the above signals are labeled using LabelImg tool[1], and the part from 920 MHz to 960 MHz of the labeled version of the given PSD example in Fig. 2 and an example of AnalogueTV and single tone signals at frequencies from 438 MHz to 502 MHz are shown in Fig. 4.

The dataset to train and test the object detection models, constructed by YOLOR and Detectron2, consists of totally 287 wideband signal recordings that disintegrate into 63, 70, 77, and 77 many PSD images corresponding to the above-mentioned frequency bands, respectively. In a random manner, the overall signal recordings are split into 258 and 29 PSD images for train&validation (with proportions of 80% and 20%) and test processes, respectively. Table I summarizes the data splitting step and gives the number of PSD images for all types of signals.

[1]Available at: https://github.com/tzutalin/labelImg

TABLE I: The numbers of signals in train and test datasets

-	Train	Test
LTE	973	104
UMTS	542	62
GSM	760	84
Single Tone	873	101
AnalogueTV	392	49
Total	3540 (258 images)	400 (29 images)

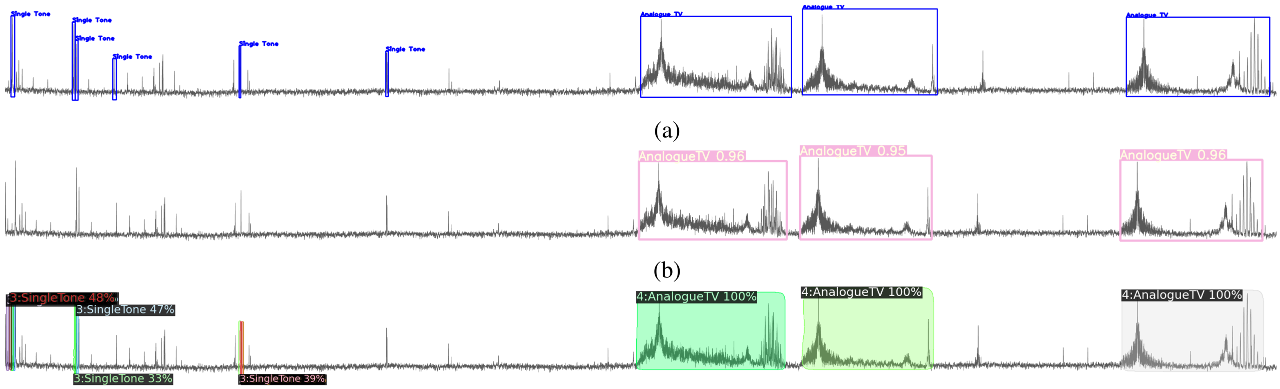
C. Models' Parameters

The open-source codes of both YOLOR[2] and Detectron2[3] model architectures have been implemented and revised according to the WBS problem. The model training processes are simplified by utilizing the concept of transfer learning. Without training all the network layers from scratch, we use the pre-trained weights, which are trained for the Common Objects in Context (COCO) dataset, for both model architectures. The models are trained and tested on a computer with a graphics processing unit of Nvidia RTX Quadro P6000 24GB, a central processor unit of Intel Xeon W-2145, and a random access memory of 64GB.

The YOLOR's model is initiated with the pre-trained model of *yolor_w6*[2]; additionally, the epoch size and batch size of the model are set as 200 and 1, respectively, where the image size through the training process is selected as 3584.

This paper considers the mask R-CNN network of Detectron2 started with the pre-trained model of *mask_rcnn_X_101_32x8d_FPN_3x*[3]. The model is trained with an iteration number of 4000 (every 20 iterations corresponds to the standard definition of an epoch), batch size per image of 512, image per batch of 8, and learning rate of 0.00025.

[2]Available at: https://github.com/WongKinYiu/yolor
[3]Available at: https://github.com/facebookresearch/detectron2

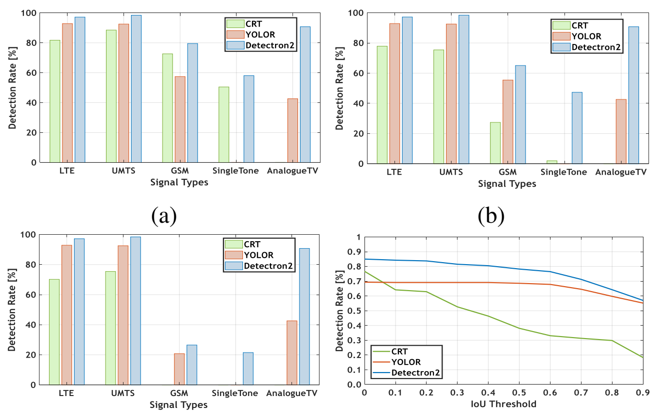


(a)

(b)

(c)

Fig. 4: (a) The labels, (b) YOLOR estimations, and (c) Detectron2 estimations related to an OTA record at 470 MHz



(a) (b)

(c) (d)

Fig. 5: Comparison of the average detection rates for signal types ((a), (b), and (c) α_{IoU} of 0.2, 0.5, and 0.8, respectively), (d) Total average detection rates for varying α_{IoU}

V. Performance Analysis

After the training periods of both object detectors, their detection and classification performances are evaluated by using the test dataset. Fig. 4 and Fig. 3 show two examples of YOLOR and Detectron2 detection results at the frequency bands between $[438, 502]$ MHz and $[920, 960]$ MHz, respectively. The figures illustrate the detected signals' boundaries, their estimated classes, and confidence scores, which indicate the conditional class probabilities. Here, the normalized boundaries (i.e., x_{min} and x_{max}) of each detected signal might be transformed to the center frequency and bandwidth estimates ($\widehat{f_c}$ and $\widehat{BW}$) via the equations:

$$\widehat{f_c} = f_0 + \frac{F_s}{2}(x_{min} + x_{max} - 1) \tag{2}$$

$$\widehat{BW} = F_s(x_{max} - x_{min}) \tag{3}$$

where f_0 denotes the middle frequency value of the recorded wideband. The performances of the models are evaluated according to their detection rates, and their classification rates. For the detection performance, the average detection rate is defined as the percentage ratio between the number of detected signals which intersect with the signals' ground truth frequency interval with a threshold of α_{IoU} and the total number of signals in the ground truth data. The average detection rates of the YOLOR and Detectron2 are depicted with and compared to those of the conventional WBS scheme (i.e., recursive thresholding [4]) in Fig. 5.

There exist three main observations from the detection performance results: (i) YOLOR and Detectron2 detection rates for wideband signals (i.e., LTE, UMTS, and AnalogueTV) are insensitive to the detection threshold α_{IoU}. For increasing values of this threshold, the detection rates for narrowband signals (i.e., GSM and SingleTone) decrease due to the frequency resolution degradation faced due to PSD plot to image conversion. In fact, for the same reason, YOLOR is not able to detect any SingleTone signal. (ii) The CRT methods' detection rates are seen to be significantly fragile to the variation of α_{IoU}. This method fails to detect AnalogueTV signals since it isn't capable of associating groups of signal peaks (as stated in Section II). (iii) Lastly, Detectron2 is seen to outperform both YOLOR and recursive thresholding method. YOLOR and Detectron2 achieve detection rates of 69% and 85%, respectively, for a nonzero α_{IoU}.

For examining the classification performance, the confusion matrices of detected signals for YOLOR and Detectron2 are provided in Fig. 6-(a) and Fig. 6-(d). As seen, all detected signals could be correctly classified by both object detectors because the measurement dataset consists of a limited number of 287 PSD images with a total of 3940 signals. Also, YOLOR's confusion matrix does not include any values related to the SingleTone signals since YOLOR fails to detect these signal types as mentioned above. To further elaborate on the classification performances, for YOLOR, the mean average precision (mAP) of $\alpha_{\text{IoU}} = 0.5$ versus epochs graph, and for Detectron2, the accuracy versus iterations graph through learning processes are provided in Fig. 6-(b) and Fig. 6-(e), respectively. In addition, Fig. 6-(c) and Fig. 6-(f) show the loss plots versus epochs and iterations for YOLOR and Detectron2, respectively. Briefly, Detectron2 achieves a better detection probability than YOLOR, and both detectors

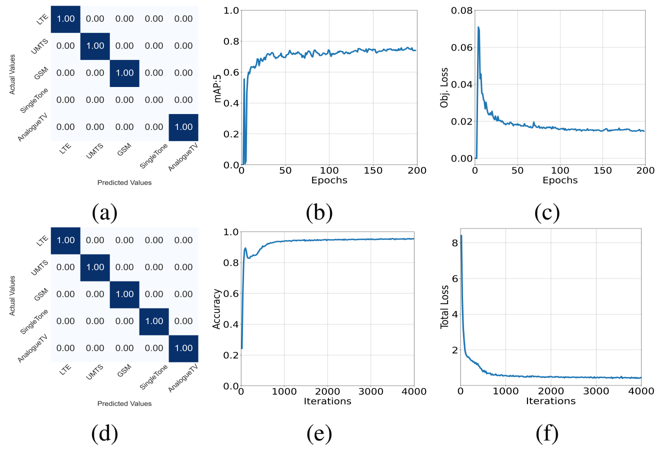


Fig. 6: YOLOR model's (a) confusion matrix, (b) mAP:0.5, (c) objection loss; Detectron2 model's (d) confusion matrix, (e) accuracy, and (f) total loss.

classify the signals accurately in case of successful detection.

VI. Conclusion

In this study, two prominent state-of-art DL-based object detectors has been utilized in order to address the WBS problem, whose objectives are to detect, localize and classify the signals in the frequency domain. Besides, the object detectors, YOLOR and Detectron2, have been compared to the CRT method. The detectors' networks are trained and evaluated with the PSD images, which contain real-world OTA signals recorded within the middle frequencies of $\{470, 870, 1850, 2100\}$ MHz through 250 MHz bandwidth. The detectors accomplish the tasks of localization and identification of the interested signals on the wideband spectrum and with limited dataset, achieve 100% classification accuracy, 85% and 69% detection rates of Detectron2 and YOLOR, respectively for a nonzero IoU threshold. The results show that Detectron2 has better detection performance. Moreover, DL-based detectors outperform the CRT method's detection capability. The results also indicate that the detection performance of the detectors is relatively poor for narrowband signals (i.e., GSM and SingleTone); in fact, YOLOR cannot detect any SingleTone signal. The authors think that the comparative study presented within this paper would achieve enhanced detection&identification scores and provide more useful insights after two main future works: (i) Handling the PSD images part by part in a smarter way that narrowband signals would not be overlooked while converting the evaluated PSD plot to PSD images; as a result, the issue of the frequency resolution degradation on PSD images will be resolved. (ii) An extensive measurement study, including a diversified data recording campaign, is thought to be beneficial to produce a more general and accurate detector and/or classifier tool.

Acknowledgment

This study has been carried out by using the signal capturing capabilities of the Multi-Dimensional Wireless Communication Signal Analysis System (KAŞİF), which was developed within the scope of the Communications and Signal Processing Research (HİSAR) Laboratory in TÜBİTAK BİLGEM. The authors would like to thank all BİLGEM personnel who contributed to the development of the KAŞİF system. Additionally, the authors gratefully acknowledge the financial support of European Union's Horizon 2020 research and innovation programme regarding the project StorAIge (Embedded storage elements on next MCU generation ready for AI on the edge) with the grant agreement number 101007321.

References

[1] H. Urkowitz, "Energy detection of unknown deterministic signals", *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523 - 531, April 1967.

[2] W. A. Gardner, "Signal Interception: A Unifying Theoretical Framework for Feature Detection", *IEEE Trans. on Communications*, vol. 36, no. 8, pp. 897 - 906, August 1988.

[3] A. Sahai, N. Hoven, R. Tandra, "Some Fundamental Limits on Cognitive Radio", *Proc. of Allerton Conference*, Monticello, Oct 2004.

[4] A Görçin, K. A. Q, H. Çelebi, H. Arslan, "An Adaptive Threshold Method for Spectrum Sensing in Multi Channel Cognitive Radio Networks", *IEEE 17th Int. Conf. on Telecom. (ICT)*, Doha, Qatar, pp. 425-429, April 2010.

[5] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," *38th Asilomar Conf. Signals, Syst. Comput.*, vol. 1. pp. 772–776, Nov. 2004.

[6] A. Muralidharan, P. Venkateswaran, S. G. Ajay, D. Arun Prakash, M. Arora, and S. Kirthiga, "An adaptive threshold method for energy based spectrum sensing in Cognitive Radio Networks," *Int. Conf. on Control, Instrum., Commun. and Computat. Techno.*, pp. 8–11, 2015.

[7] W. Li, K. Wang, and L. You, "A deep convolutional network for multitype signal detection and classification in spectrogram," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9797302, 16 pages, 2020.

[8] X. Zha, H. Peng, X. Qin, G. Li, and S. Yang, "A deep learning framework for signal detection and modulation classification," *Sensors*, vol. 19, no. 18, p. 4042, Sep. 2019.

[9] K. N. R. S. V. Prasad, K. B. D'Souza, V. K. Bhargava, H. Boostanimehr and S. Mallick, "A deep learning framework for blind time-frequency localization in wideband systems", *IEEE Veh. Tech. Conf. (VTC-Spring)*, May 2020.

[10] K. N. R. S. V. Prasad, K. B. D'Souza, and V. K. Bhargava, "A Downscaled Faster-RCNN Framework for Signal Detection and Time-Frequency Localization in Wideband RF Systems", *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4847 – 4862, July 2020.

[11] A. Vagollari, V. Schram, W. Wicke, M. Hirschbeck and W. Gerstacker, "Joint Detection and Classification of RF Signals Using Deep Learning," *IEEE Veh. Tech. Conf. (VTC-Spring)*, pp. 1-7, April 2021.

[12] J. Wang, Z. Sun, Z. Wu, "Labeling of Radio Signal by Jointly Using Feature Engineering and Unsupervised Learning", *6th Int. Commun. Image and Signal Proces. Conf.*, pp. 403-407, 2021.

[13] H. Franco, C. C.-Kroenke, S. Welch, and M. Graciarena, "Wideband spectral monitoring using deep learning", *2nd ACM Workshop on Wirel. Secur. and Mach. Learn.*, New York, NY, USA, pp. 19 – 24, 2020.

[14] K. Tekbıyık, Ö. Akbunar, A. R. Ekti, A. Görçin, G. K. Kurt and K. A. Qaraqe, "Spectrum Sensing and Signal Identification With Deep Learning Based on Spectral Correlation Function," *in IEEE Trans. Vehic. Techno.*, vol. 70, no. 10, pp. 10514-10527, Oct. 2021.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Info. Proces. Sys.*, pp. 91–99, Montreal, Canada, December 2015.

[16] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," *14th European Conf. ECCV 2016*, pp. 21–37, Amsterdam, Netherlands, Oct. 2016.

[17] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv*, preprint: 1804.02767, 2018.

[18] Xiao, Youzi, et al. "A review of object detection based on deep learning." *Multimedia Tools and Applications* 79.33, pp. 23729-23791, 2020.

[19] Wang, C., Yeh, I. & Liao, H. You Only Learn One Representation: Unified Network for Multiple Tasks. *Arxiv abs/2105.04206*, 2021.

[20] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016.

[21] W. Y. Yuxin, A. Kirillov, F. Massa, L. Wan-Yen, and R. Girshick, Detectron2, 2019, [Online]. Available: https://github.com/facebookresearch/detectron2

[22] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Tran. Audio Elec.*, vol. 15, no. 2, pp. 70–73, 1967.