Improved Waveform Classification for Integrated Radar-Communication 6G Systems via Convolutional Neural Networks

Thien Huynh-The, *Senior Member, IEEE*, Nguyen Cong Luong, Hoc Phan,
Daniel Benevides da Costa, *Senior Member, IEEE*, and Quoc-Viet Pham, *Senior Member, IEEE*

***Abstract*—To overcome the spectrum congestion problem in next-generation wireless networks, an integrated radar-communication system with spectrum sharing becomes a promising solution, wherein radar and communication signals can be discriminated by means of modulated waveforms. This letter presents an efficient radar-communication waveform classification method by taking advantage of the combination of smooth pseudo Wigner-Ville distribution-based time-frequency analysis and deep learning to achieve a good trade-off between complexity and accuracy. To this end, a high-performance convolutional network, namely the radar-communication waveform recognition network (RadComNet), is designed with multiple cutting-edge techniques and advanced structures, including depthwise convolution for complexity reduction and residual connection and multi-level attention mechanisms for learning efficiency enhancement. Relying on the simulation results acquired on a synthetic signal dataset of 12 radar and communication waveform types with the presence of channel impairments, our proposed method shows superiority over other classification approaches and deep models in terms of accuracy and complexity.**

***Index Terms*—Deep learning, radar-communication coexistence systems, time-frequency analysis, waveform classification.**

I. Introduction

In recent years, spectrum congestion has soared and affected to radar and communication systems. Both demand wide bandwidths for high-performance next-generation wireless networks, leading to competition for limited resources [1]. Integrated radar-communication systems with waveform-based classification offer promising solutions across diverse scenarios (such as dynamic spectrum sharing, interference mitigation, and detection of communication anomalies) and applications. In the context that multiple technologies in 5G and 6G compete for limited spectrum resources, waveform classification plays a crucial role in enabling efficient cooperative spectrum sharing. By accurately identifying the types of signals occupying a specific frequency band, network operators can dynamically allocate spectrum resources to different users, thus maximizing spectrum utilization and network performance. In autonomous vehicles, waveform recognition can differentiate crucial radar signals for obstacle detection and collision avoidance from communication signals used for vehicle-to-vehicle communication [2]. For sensing and communication performance, waveform classification can improve target detection accuracy, facilitate higher data rate, reinforce channel estimation, and optimize networking resources [3].

Numerous waveform-based signal classification methods, wherein signals transmitted from different sources with different modulation modes can be discriminated, have been built on a conventional machine learning framework, but most were developed separately for radar and communication categories. A few years ago, deep learning (DL) was exploited for both radar waveform recognition and communication modulation classification to deal with large-messy-confusing signal datasets under various channel impairments, in which the input of deep networks can be the raw data of in-phase and quadrature components and the time-frequency representation of signals. Lin *et al.* [4] developed a regular convolutional neural network (CNN) to classify waveform-based radar signals by jointly learning representative features of Choi-William distribution (CWD) and short-time autocorrelation. The work [5] extracted representative spectrum features by Fourier synchrosqueezing transform and then learned by a CNN-based hierarchical classifier, but a plain-designed architecture limits feature learning efficiency. Huang *et al.* [6] improved the radar waveform classification accuracy by fusing various visual features acquired by different signal decomposition techniques, such as Wigner-Ville distribution (WVD), CWD, and wavelet transform. Another work [7] designed a CNN with residual blocks to enhance the learning efficiency of CWD features and preserve the network from vanishing gradient, but this large-sized network caused high memory usage. Some methods have designed cutting-edge architectures to improve accuracy, however they are inapplicable for resource-limited devices due to high computation and large memory consumption [8]. Recently, a few radar-communication waveform identification methods were introduced with DL to address spectrum resource exhaustion problems. Three CNN-based classifiers were deployed for single-carrier radar waveforms, single-carrier communication waveforms, and multi-carrier waveforms [9]; and an advanced CNN was contrived to extract highly distinguishable features [10]. However, both methods suffer from high memory consumption despite their high classification accuracy.

This paper proposes a high-performance waveform classification method for integrated radar-communication systems as shown in Fig. 1, in which a deep CNN is designed to achieve a

This work is supported by Ho Chi Minh City University of Technology and Education (HCMUTE) under Grant No. T2023-61. Thien Huynh-The and Hoc Phan are with Department of Computer and Communications Engineering, Ho Chi Minh City University of Technology and Education, Ho Chi Minh City 70000, Vietnam (e-mail: thienht@hcmute.edu.vn, phanhoc@hcmute.edu.vn). Nguyen Cong Luong is with the Faculty of Computer Science, Phenikaa University, Vietnam (e-mail: luong.nguyencong@phenikaa-uni.edu.vn). Daniel Benevides da Costa is with the Department of Electrical Engineering, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia (e-mail: danielbcosta@ieee.org). Quoc-Viet Pham is with the School of Computer Science and Statistics, Trinity College Dublin, D02 PN40, Ireland (e-mail: viet.pham@tcd.ie).

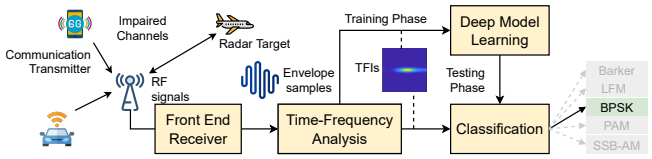


Fig. 1. The processing framework of our proposed DL-based waveform recognition in radar-communication systems. The framework consists of two key steps: time-frequency analysis (TFA) extracts features representing not only the overall frequency spectrum but also its temporal variations and deep model learning (with offline training and online classification). The output of TFA is then passed to a deep network to train a classification model and predict waveform types with the trained model.

good trade-off between the practical implementation complexity and the classification accuracy under channel impairments. The main contributions are summarized as follows:

- We realize the smooth pseudo WVD (SPWVD) to decompose a signal into a high-resolution time-frequency representation, where the interference problem in conventional WVD and CWD is alleviated significantly.
- To classify waveform-based radar and communication signals, we design a deep CNN with cutting-edge architecture, wherein some modern connection topologies and structures (such as depthwise convolution, residual connection, and attention mechanism) are aptly coordinated to reduce complexity without sacrificing accuracy.
- For performance evaluation, we generate a synthetic signal dataset comprising 12 radar and communication waveform types under different impairments. Based on a multifarious simulation, the proposed deep model demonstrates superiority in complexity reduction and accuracy improvement compared to other deep learning models.

II. Signal Model and TFA Technique

A. Signal Model

This work considers a typically integrated radar-communication system, in which the receiver can receive pulsed radar signals and modulated communication signals from different transmission sources in Fig. 1. The complex envelope of a received signal is disturbed by additive noise as

$$y(k) = x(k) \otimes h(k) + n(k), \tag{1}$$

where $x(k)$ is the transmitted radio signal that is already modulated at the transmitter, $h(k)$ denotes the transmission channel coefficient, $n(k)$ symbolizes the complex additive white Gaussian noise (AWGN). Notably, the waveforms of pulsed radar signals and modulated communication signals should be recognized without the channel state information at the receiver. The propagation channel properties and signal configurations are described hereafter in dataset generation.

B. Time-Frequency Analysis: Smooth Pseudo Wigner-Ville Distribution

Compared with direct processing raw waveform envelope data, performing time-frequency analysis (TFA) allows to exhibit time-varying frequency characteristics that are effectively learned by deep CNN architectures for such modulation classification and waveform identification tasks. WVD has been realized as one of the most effective time-frequency signal analysis and processing techniques thanks to capably producing a high-resolution representation in the time and frequency of non-stationary signals. Besides deriving the total energy in a time-frequency plane, WVD alleviates the time and frequency marginals in terms of instantaneous power in time as well as the energy spectrum in frequency. Given a continuous signal $s(t)$, its WVD can be written as follows:

$$\mathrm{WVD}_s(t,\omega) = \int_{-\infty}^{\infty} s\left(t+\frac{\tau}{2}\right) s^*\left(t-\frac{\tau}{2}\right) e^{-j\omega\tau} d\tau, \tag{2}$$

Although the WVD performs TFA better than some standard transform techniques in many radio signal processing tasks, it contains interference terms (a.k.a., cross-term interference) that may confound interpretation. For instance, the WVD of a signal involving multiple components, i.e., $y(t) = s(t) + n(t)$ with $s(t) = x(t) \cdot h(t)$ in (1), is given as

$$\begin{aligned}\mathrm{WVD}_y(t,\omega) =& \mathrm{WVD}_s(t,\omega) + \mathrm{WVD}_n(t,\omega) \\ &+ 2\Re\left\{\mathrm{WVD}_{s,n}(t,\omega)\right\},\end{aligned} \tag{3}$$

where WVD_s and WVD_n, respectively, are the auto-terms of the free-noise signal $s(t)$ and the additive noise $n(t)$. The $2\Re\{\mathrm{WVD}_{s,n}\}$ is the so-called cross-term that bothers the auto-terms in visualizing a high-resolution time-frequency representation. Although the cross-term has no physical interpretation, its magnitude is double that of the auto-term. The negative effect of the cross-term can be alleviated by sharpening the distribution in time and frequency dimensions via the SPWVD technique - an upgrade of the regular WVD with a smoothing function $\Xi(t,\omega) = g(t)H(\omega)$. Notably, two independent low-pass windows of $g(t)$ and $H(\omega)$ separately perform time and frequency filtering, respectively. In particular, SPWVD sharpens the distribution as follows:

$$\begin{aligned}\mathrm{SPW}_s(t,\omega;\Xi) = \int_{-\infty}^{\infty} & g(t)H(\omega)\, s\left(t+\frac{\tau}{2}\right) \\ & s^*\left(t-\frac{\tau}{2}\right) e^{-j\omega\tau} d\tau.\end{aligned} \tag{4}$$

To establish the time and frequency windows, we adopt Kaiser windows, whose coefficients are calculated as

$$c(n) = \mathcal{B}\left(\gamma\sqrt{1-\left(\frac{n-N/2}{N/2}\right)^2}\right)\mathcal{B}^{-1}(\gamma),\ 0 \le n \le N, \tag{5}$$

where the zeroth-order altered Bassel function $\mathcal{B}$ has the length of $N+1$ and γ is the shape factor. The samples of 12 radar and communication waveforms are displayed in Fig. 3.

III. RadComNet: Robust and Cost-Efficient ConvNet for Waveform Classification

This section presents the detailed architecture of the proposed deep ConvNet, namely Radar-Communication Waveform Recognition Network (RadComNet) in Fig. 2, for learning waveform patterns of radio signals effectively in the training stage and identifying the waveform fashions of received signals automatically in the prediction stage. RadComNet is aptly designed with a sophisticated architecture involving several advanced structures of layer connection that can perform waveform classification of radar and communication signals

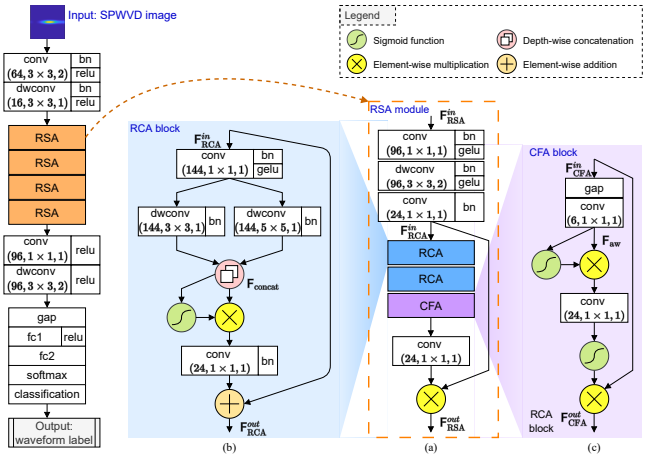


Fig. 2. Architecture of RadComNet for waveform classification in radar-communication systems has four RSA modules. As detailed in (a), each RSA has two RCA blocks and one CFA block, where their structures are clarified in (b) and (c), respectively. RCA has different filter sizes and residual connections. Remarkably, the features extracted inside RCA as well as the output derived by RSA are emphasized by the multi-level attention mechanism. Notably, $(k, n \times n, s)$ indicates k $n \times n$ filters with the stride s in regular conv layers and k filter groups in dwcon layers.

robustly under different channel impairments while using fewer computing resources. Beginning with an input layer that is specified by the size 224×224 identical to the size of SPWVD images, the architecture continues with one regular convolutional (conv) layer and one depth-wise convolutional (dwconv) layer. Two convolutional layers are followed by batch normalization layers and activation rectified linear unit (relu) layers. To extract and learn underlying visual features from SPWVD images, RadComNet has four modules, namely Rational Spectrum Association (RSA), to generally extract and learn highly discriminative features.

RSA: As showed in Fig. 2(a), RSA module begins with an operation sequence of conv, dwconv, bn, and relu layers. The output of this sequential processing procedure then inputs to the series connection of two Residual Cost-Efficient Accumulation (RCA) blocks and one Channel-wise Feature Awareness (CFA) block. Given an input $\mathbf{F}^{in}_{\text{RSA}}$, we obtain the output of this procedure as well as the input of RCA as follows:

$$\mathbf{F}^{in}_{\text{RCA}} = \mathcal{C}_{1\times1}\left(\mathcal{G}_{3\times3}\left(\mathcal{F}_{1\times1}\left(\mathbf{F}^{in}_{\text{RSA}}\right)\right)\right), \tag{6}$$

where $\mathcal{F}_{n\times n}$ represents an operation sequence conv-bn-relu with $n \times n$ filters, $\mathcal{C}$ denotes the regular convolution operation, and $\mathcal{G}$ symbolizes an operation sequence with dwconv layer instead of conv layer. RSA finalizes with a conv layer for feature synthesis and an element-wise multiplication layer for selective feature emphasis

$$\mathbf{F}^{out}_{\text{RSA}} = \mathbf{F}^{in}_{\text{RCA}} \circ \mathcal{C}_{1\times1}\left(\text{CFA}\left(\text{RCA}\left(\text{RCA}\left(\mathbf{F}^{in}_{\text{RCA}}\right)\right)\right)\right), \tag{7}$$

where RCA and CFA denote the synthetic operations of RCA and CFA blocks, respectively, and $\circ$ symbolizes the element-wise multiplication operation.

RCA: As the core blocks in RSA that is responsible for enhancing the feature diversity, RCAs first extract intrinsic features using different filter sizes simultaneously. As the structural design, RCA has two dwconv layers with 3×3 and 5×5 filters arranged in parallel, where those outputs are combined in the depth dimension via a depth-wise concatenation layer. For a better description, we track the feature flow in RCA by coordinating the notations in Fig. 2(b)

$$\begin{aligned}
\mathbf{F}_{\text{concat}} &= \langle \mathbf{F}_{3\times3}, \mathbf{F}_{5\times5} \rangle, \\
\mathbf{F}_{3\times3} &= \mathcal{G}^{-}_{3\times3}\left(\mathcal{F}_{1\times1}\left(\mathbf{F}^{in}_{\text{RCA}}\right)\right), \\
\mathbf{F}_{5\times5} &= \mathcal{G}^{-}_{5\times5}\left(\mathcal{F}_{1\times1}\left(\mathbf{F}^{in}_{\text{RCA}}\right)\right),
\end{aligned} \tag{8}$$

where $\langle\cdot\rangle$ symbolizes the depthwise concatenation operation and $\mathcal{G}^{-}$ denotes the depthwise convolution operation only. The gathered output then passes through an attention mechanism to strengthen relevant features and a residual connection to preserve gradients from preceding layers.

$$\mathbf{F}^{out}_{\text{RCA}} = \mathbf{F}^{in}_{\text{RCA}} + \mathcal{F}^{-}_{1\times1}\left(\mathbf{F}_{\text{concat}} \circ \left(\sigma\left(\mathbf{F}_{\text{concat}}\right)\right)\right), \tag{9}$$

where σ denotes the sigmoid activation.

CFA: A two-RCA-block sequence is followed by CFA in Fig. 2(c), a post-processing block that can refine the extracted features essentially, i.e., reinforcing the more relevant features and weakening the less relevant features. To this end, CFA is functioned as an attention mechanism [11]. In particular, the feature refining can be obtained by scaling features with attention values as follows:

$$\mathbf{F}^{out}_{\text{CFA}} = \mathbf{F}^{in}_{\text{CFA}} \circ \sigma\left(\mathcal{F}^{-}_{1\times1}\left(\mathbf{F}_{\text{att}} \circ \left(\sigma\left(\mathbf{F}_{\text{att}}\right)\right)\right)\right), \tag{10}$$

where

$$\mathbf{F}_{\text{att}} = \mathcal{C}_{1\times1}\left(\text{gap}\left(\mathbf{F}^{in}_{\text{CFA}}\right),\right), \tag{11}$$

where $\mathcal{C}$ denotes only the regular convolution operation and gap represents the global average pooling. Many 1×1 convolutional layers (a.k.a., pointwise convolution) are arranged in RSA modules for output channel alignment (i.e., reducing the number of feature maps while preserving salient features). In summary, RCAs extract intrinsic features at multiple scales of SPWVD-based time-frequency representations using multiple filter sizes and attention mechanisms, whereas CFA refines these features by selectively amplifying the more relevant ones. Therefore, the combination of RCA and CFA improves classification accuracy and robustness against channel impairments, thus making RadComNet an effective solution for integrated radar-communication waveform classification.

The spatial dimensions of feature maps in RSA decrease two times by a dwconv layer with stride 2 in (6), enabling the network to learn highly-discriminative spectrum features at multi-scale representations. In Fig. 2, the fourth RSA module is followed by a group of layers to intensively reveal more fundamental visual features. Finally, RadComNet is completed with a series of multiple layers: a global average pooling layer, a fully connected (fc) layer, a softmax layer, and an output layer. The number of neurons in the fc layer equals the number of waveform types to classify. The waveform class probability distribution is calculated by the softmax function, where the class taking the highest probability offers the classification results. The network is trained with the adaptive moment estimation (adam) optimizer and other setups: the maximum number of epochs is 60, the initial learning rate is 0.001 with a piecewise constant decay schedule (reduce 5 times for every 20 epochs), and the mini-batch size is 64.

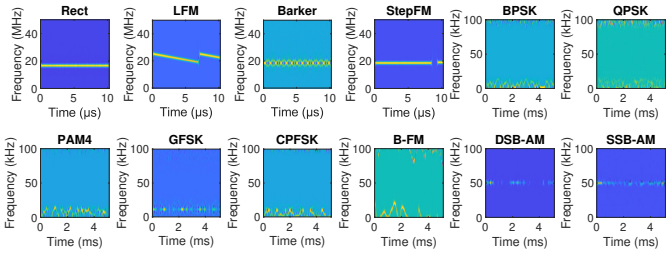


Fig. 3. SPWVD images of radar and communication waveform types.

TABLE I
DATASET INFORMATION OF 12 WAVEFORM TYPES.

Category	Waveform type	Abbr	No. signals
Radar	Rectangular	Rect	32,768 signals per type
	Linear frequency modulation	LFM	
	Barker Code	Barker	
	Step frequency modulation	StepFM	
Communication	Binary phase-shift keying	BPSK	
	Quadrature phase-shift keying	QPSK	
	Pulse amplitude modulation 4-level	PAM4	
	Gaussian frequency shift keying	GFSK	
	Continuous phase frequency shift keying	CPFSK	
	Broadcast frequency modulation	B-FM	
	Double sideband amplitude modulation	DSB-AM	
	Single sideband amplitude modulation	SSB-AM	

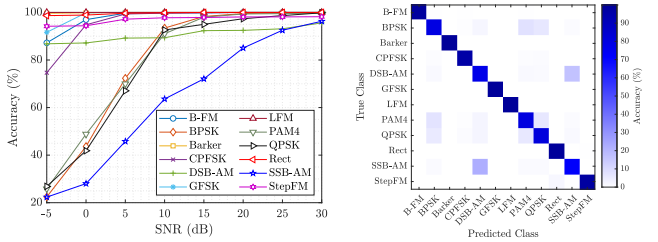
IV. Dataset, Results, and Discussions

A. Dataset Generation and Simulation Setup

Considering some realistic challenges (e.g., cost and privacy) to capture real-world data, we generate a synthetic radio signal dataset with 12 waveform types (detailed information is reported in Table I) for performance evaluation. Each signal is randomly generated with specific characteristics and augmented with various channel impairments to make it more realistic. For example, the pulse width, sweep width, and chip width of radar signals are randomly chosen. Moreover, multipath fading channels are used to model real-world phenomena in wireless communications: multipath scattering effects with the path delays $[0, 9, 17] \times 10^{-6}$ seconds corresponding to the average path gains of $[0, -2, -10]$ dB, AWGN with signal-to-noise ratio (SNR) in the range $[-5 : 5 : 30]$ dB, the maximum Doppler shift $f_{\mathrm{D}} = 70$ Hz, and the maximum clock offset 5 ppm (parts per million). The dataset has $393,216$ signals to transform into spectrogram images with the size of 224×224, which is randomly divided into the training, validation, and test sets with a ratio of $0.7 - 0.15 - 0.15$. The deep network-based waveform classifier is evaluated on a computing platform using $3.80-$GHz CPUs, 64GB RAM, and a single NVIDIA GTX 3060Ti GPU. Multiple independent running times of training, validation, and testing are to measure the overall classification accuracy on the test set.

B. Simulation Results and Discussions

In Fig. 4(a), we first evaluate the detailed classification results of 12 waveforms versus SNR. It observes that the classification rates increase along with the increase of SNR, where all the waveforms attain greater than 90% at 25 dB. The network recognizes Rect, Barker, LFM, StepAM, B-FM, GFSK, and CPFSK outstandingly with over 95% at



(a) (b)

Fig. 4. RadComNet's performance: (a) accuracy of for 12 radar and communication waveform types versus SNR and (b) confusion matrix of classification.

5 dB. Interestingly, the combination of ConvNet and TFA with SPWVD is more advantageous to radar signals than communication ones. With SPWVD-TFA, many digital and analog signals are confused together at low SNR regimes, such as QPSK versus PAM4 and SSB-AM versus DSB-AM. These phenomena are relatively demonstrated via the confusion matrix of waveform classification in Fig. 4(b).

Secondly, we investigate the feature extraction and learning efficiency of RSA by varying the number of modules in the network architecture. Fig. 5(a) plots the correct classification rate of RadComNet versus SNR, where a greater number of modules yields a higher accuracy. The improvements are significant with small numbers of modules, e.g., RadComNet achieves 82.85% with two modules and improves by 4.00% and 7.02% with three and four modules, respectively. When the number of RSA modules is larger than four, the increment is trivial and inappropriate for the growth of complexity (assessed over the number of trainable parameters and the computation cost). In Fig. 5(b), we clarify the effectiveness of RSA's sophisticated-designed structure via an ablation study, in which Sim-RadComNet indicates RadComNet with ordinary layer connection (no residual connection and attention mechanism), Reg-RadComNet refers to RadComNet with regular conv layers, and w/o-CFA-RadComNet means RadComNet with without CFA blocks in RSA modules. Indeed, RadComNet classifies more accurately than SimpleNet by 2.4% while it considerably reduces the model size of Reg-RadComNet (approximately $3\times$ times smaller). Without CFA, the RadComNet's accuracy slightly reduces but still being better than Sim-RadComNet. In Fig. 5(c), we examine the Doppler influence to the overall accuracy by configuring the channel model with different maximum Doppler shift $f_{\mathrm{D}} = \{5, 70, 300\}$ Hz. The higher the Doppler frequency is tuned, the stronger the time-frequency presentation is scratched in the frequency dimension, thus resulting in the accuracy reduction at multiple SNR regimes. Furthermore, Fig. 5(d) investigates the effect of SPWVD image resolution on the overall classification accuracy, in which higher resolution leads to better accuracy due to the presence of detailed radio signatures in the time-frequency domain. However, the computational cost of feature learning in convolutional layers grows significantly.

In the last simulation, we conduct a comparison of the proposed method with existing TFA-aided signal identification methods (including Lin *et al.* [4], Shen *et al.* [7], and Huynh-

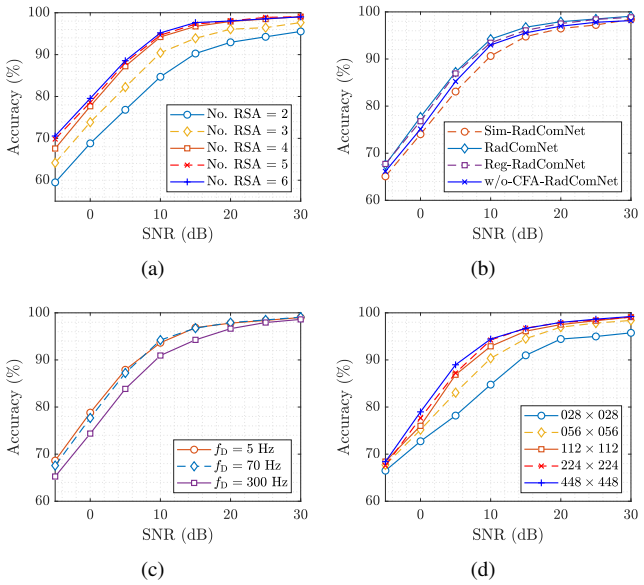


Fig. 5. Accuracy of RadComNet with (a) different numbers of RSA module, (b) alternative versions, (c) different values of maximum Doppler shift f_D, and (d) different image resolutions.

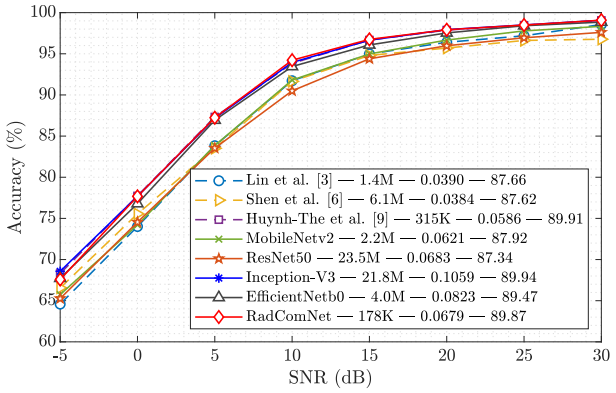


Fig. 6. Performance comparison between RadComNet with other existing methods and various backbone networks. In the legends, the format is as follows: method – network size (as number of parameters) – prediction speed (milliseconds) – overall classification accuracy (%).

The *et al.* [10]) and various deep backbone networks (such as MobileNetv2 [12], ResNet50 [13], Inception-V3 [14], and EfficientNetb0 [15]) with respect to classification accuracy and complexity. In Fig. 6, the proposed method with SPWVD-TFA and RadComNet achieves the overall accuracy of 89.87%, much better than Lin *et al.*, Shen *et al.*, ResNet50, and MobileNetv2 by $1.9 - 2.5\%$. The works [4] and [7] designed networks with simple architectures without any advanced connections and sophisticated structures, thus causing inefficient feature learning and waveform discrimination. Despite being worse than Huynh-The *et al.* and Inception-V3 slightly, RadComNet gains significant cost efficiency (a.k.a., network size and average prediction speed). As reported in Fig. 6, RadComNet is the smallest network when compared with several high-accuracy models. For instance, our network has 178K parameters and consumes 0.0679 milliseconds (ms), that is approximately 130× smaller and 1.6× faster than Inception-V3. Obviously, the combination of depthwise convolution and various advanced techniques (residual connection and multi-level attention mechanism) comes along with dual benefits, i.e., accuracy improvement and complexity reduction.

V. Conclusions

We have presented an efficient radar-communication waveform classification method for coexistence systems to alleviate the spectrum congestion problem by learning a deep ConvNet with SPWVD representation. RadComNet, a cost-efficient deep network, was proposed with a cutting-edge architecture with multiple cutting-edge techniques and advanced topologies to reduce complexity without sacrificing accuracy. RadComNet attained classification robustness in simulations on an impaired signal dataset of 12 radar and communication waveform types. Compared with other deep models, RadComNet reserved a low-cost complexity while yielding a comparable accuracy (same accuracy but 130× smaller and 1.6× faster than Inception-V3) for potential of resource-constrained devices.

References

[1] P. Gong, K. Xu, Y. Wu, J. Zhang, and H. C. So, "Optimization of LPI-FDA-MIMO radar and MIMO communication for spectrum coexistence," *IEEE Wirel. Commun. Lett.*, 2023, Early Access.

[2] A. Venon, Y. Dupuis, P. Vasseur, and P. Merriaux, "Millimeter wave FMCW RADARs for perception, recognition and localization in automotive applications: A survey," *IEEE Trans. Intell. Veh.*, vol. 7, no. 3, pp. 533–555, Sep. 2022.

[3] S. Zheng, S. Chen, P. Qi, H. Zhou, and X. Yang, "Spectrum sensing based on deep learning classification for cognitive radios," *China Commun.*, vol. 17, no. 2, pp. 138–148, Feb. 2020.

[4] A. Lin, Z. Ma, Z. Huang, Y. Xia, and W. Yu, "Unknown radar waveform recognition based on transferred deep learning," *IEEE Access*, vol. 8, pp. 184 793–184 807, 2020.

[5] G. Kong, M. Jung, and V. Koivunen, "Waveform recognition in multipath fading using autoencoder and CNN with fourier synchrosqueezing transform," in *Proc. IEEE RADAR*, 2020, pp. 612–617.

[6] H. Huang, Y. Li, J. Liu, D. Shen, G. Chen, E. Blasch, and K. Pham, "LPI waveform recognition using adaptive feature construction and convolutional neural networks," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 38, no. 4, pp. 14–26, Apr. 2023.

[7] L. Shen, D. Quan, X. Wang, X. Jin, and N. Jin, "Radar waveform recognition based on time-frequency images under deteriorating channel conditions," in *Proc. IEEE ICCC*, China, Dec. 2022, pp. 188–192.

[8] I. Guven, C. Yagmur, B. Karadas, and M. Parlak, "Classifying LPI radar waveforms with time-frequency transformations using multi-stage CNN system," in *Proc. 23rd IRS*, Gdansk, Poland, Sep. 2022, pp. 501–506.

[9] G. Kong, M. Jung, and V. Koivunen, "Waveform classification in radar-communications coexistence scenarios," in *Proc. IEEE GLOBECOM*, Taipei, Taiwan, Dec. 2020, pp. 1–6.

[10] 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. IEEE ICC*, Seoul, South Korea, May 2022, pp. 1–6.

[11] D. Soydaner, "Attention mechanism in neural networks: where it comes and where it goes," *Neural Computing and Applications*, vol. 34, no. 16, pp. 13 371–13 385, 2022.

[12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE CVPR*, USA, Jun. 2018, pp. 4510–4520.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, USA, Jun. 2016, pp. 770–778.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE CVPR*, USA, Jun. 2016, pp. 2818–2826.

[15] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. ICML*. PMLR, 2019, pp. 6105–6114.