Densely-Accumulated Convolutional Network for Accurate LPI Radar Waveform Recognition

Thien Huynh-The[1], Quoc-Viet Pham[2], Toan-Van Nguyen[3],Van-Sang Doan[4], Nhan Thanh Nguyen[5], Daniel Benevides da Costa[6], and Dong-Seong Kim[1]
[1]ICT Convergence Research Center, Kumoh National Institute of Technology, Korea
[2]Korean Southeast Center for the 4th Industrial Revolution Leader Education, Pusan National University, Korea
[3]Department of Electronics and Computer Engineering, Hongik University, Korea
[4]Faculty of Communication and Radar, Naval Academy, Vietnam
[5]Centre for Wireless Communications, University of Oulu, Finland
[6]Future Technology Research Center, National Yunlin University of Science and Technology, Taiwan, R.O.C.
and with the Dept. of Computer Engineering, Federal University of Ceará, Sobral 62010-560, CE, Brazil

***Abstract*—This paper presents a deep learning-based method to automatically recognize low probability of intercept (LPI) radar waveforms against diversified jamming attacks. Concretely, an efficient convolutional neural network (CNN) architecture, namely Densely-Accumulated Network (DANet), is introduced to learn the time-frequency representation transformed by the Wigner-Ville distribution. Such an architecture has several novel densely-accumulated connection modules specified by various symmetric and asymmetric convolutional layers to enrich diversified features at multiple representational maps. Besides, the skip-connection and dense-connection are leveraged to improve feature learning efficiency and prevent the vanishing gradient when the network goes deeper. Some image processing techniques (e.g., global thresholding and digital filtering) are adopted to enhance the quality of time-frequency image. Relying on simulations, we benchmark the proposed method on a synthetic 13-waveform dataset and also investigate the influence of hyper-parameters (such as image size, number of modules, training data size) on the overall recognition performance. Remarkably, with average accuracy of 98.2% at 0 dB signal-to-noise ratio (SNR), DANet outperforms several backbone CNNs and state-of-the-art networks of LPI waveform recognition while keeping a cost-efficient model.**

***Index Terms*—LPI radar, waveform recognition, convolutional neural network, densely-accumulated connection.**

I. Introduction

One of the most effective counter-jamming techniques commonly deployed in automotive radar systems is the utilization of low probability of intercept (LPI) radar waveform [1]. This urgently requires the development of a high-performance method to automatically recognize LPI waveform signals. Numerous existing waveform recognition approaches have exploited several principal time-frequency analysis (TFA) techniques [2], such as short-time Fourier transforms, Wigner-Ville distribution (WVD), and Choi-William distribution (CWD), to extract the radio characteristics in the time and frequency domains [3], [4]. However, the traditional machine learning algorithms adopted by conventional approaches cannot learn high-level representational features in time-frequency image (TFI) to discriminate many waveforms explicitly [5].

Recently, the emergence of deep learning (DL) with convolutional neural network (CNN) architectures [6], [7] has been inspiring to radio signal processing in communications [8], [9]. By deploying multiple stacks of convolutional kernels (so-called filters), convolutional networks are able to automatically capture spatial features from high-dimensional data at multi-scale representations. In [10], Wang *et al.* proposed a radar waveform recognition framework with a cascade-connected CNN to learn five waveform patterns from the visual features of WVD-TFIs. Zhang *et al.* [11] and Kong *et al.* [12] designated several fully connected layers in the network architecture to precisely recognize a large number of waveform classes from CWD-TFIs. Transfer learning, one of the widely used techniques to take advantage of learned rich representational features, was leveraged in [13], [14] to dramatically save training time and computing resources. Although existing DL-based methods can improve the recognition performance by exploiting the superior learning ability of CNNs when compared with traditional classification algorithms, two challenging issues remain: (i) the low quality of time-frequency representation is caused by channel impairments and (ii) the learning efficiency is embarrassed by plain-designed network architectures without considering the overfitting and vanishing gradient problems [15].

To overcome these above-mentioned issues, we propose a high-performance LPI radar waveform recognition approach. A densely-accumulated convolutional neural network (DANet) is introduced to learn the highly discriminative features of TFIs converted by the smoothed pseudo Wigner-Ville distribution (SPWVD). With several advanced-designed modules to densely accumulate features at multi-resolution maps, our proposed CNN achieves a high accuracy of 13-class classification under the channel impairment with multipath fading propagation and additive noise. The main contributions of this

This research was financially supported by National Research Foundation of Korea (NRF) through Creativity Challenge Research-based Project (2019R1I1A1A01063781), and in part by the Priority Research Centers Program through the NRF funded by the Ministry of Education, Science and Technology (2018R1A6A1A03024003).

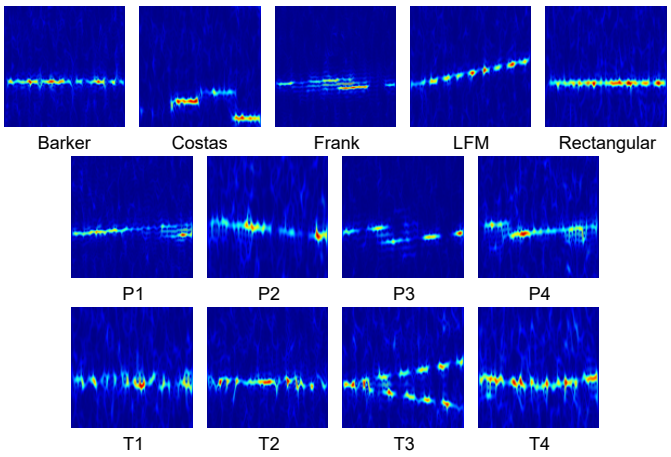


Fig. 1. Example of various LPI radar waveform signals analyzed by Wigner-Ville distribution.

work are summarized as follows:

- We propose an LPI radar waveform recognition method by exploiting CNNs to learn representational features of SPWVD-TFIs.
- We design DANet having a series of densely-accumulated connection modules, where each module is specified by multiple convolutional layers with different symmetric and asymmetric kernels in an association of skip-connection and dense-connection to improve learning efficiency and prevent the network from vanishing gradient problem.
- We investigate the influence of various hyper-parameters on the overall recognition performance and compare DANet with other deep networks, where our proposed CNN demonstrates the superiority in terms of accuracy and cost efficiency.

II. TFA for Signal Representation

A. Smoothed Pseudo Wigner–Ville Distribution

WVD is one of the most widely used TFA techniques to provide a high-resolution time-frequency representation of a radio signal. Given an auto-correlation expression $y\left(t+\frac{\tau}{2}\right)y^{*}\left(t-\frac{\tau}{2}\right)$ of a continuous signal $y\left(t\right)$, the WVD can be defined as follows:

$$\mathsf{WVD}_{y}(t,f)=\int_{-\infty}^{+\infty}y\left(t+\frac{\tau}{2}\right)y^{*}\left(t-\frac{\tau}{2}\right)e^{-j2\pi f\tau}d\tau, \quad (1)$$

where t is the time stamp, f is the ordinary frequency, and τ is the time lag. For a discrete signal with K samples, the WVD is written by

$$\mathsf{WVD}_{y}(n,k)=\sum_{m=-N}^{N}y\left(n+\frac{m}{2}\right)y^{*}\left(n-\frac{m}{2}\right)e^{-j2\pi km/K}. \quad (2)$$

For the multiple components in a synthetic signal like the intercepted noisy radar signal $y\left(t\right)=x\left(t\right)+n\left(t\right)$, this distribution can be formulated by

$$\begin{aligned}\mathsf{WVD}_{y}\left(t,f\right)=&\mathsf{WVD}_{x+n}\left(t,f\right),\\ =&\mathsf{WVD}_{x}\left(t,f\right)+\mathsf{WVD}_{n}\left(t,f\right)\\ &+2\Re\left\{\mathsf{WVD}_{xn}\left(t,f\right)\right\},\end{aligned} \quad (3)$$

where WVD_{x} and WVD_{n} are the auto-correlation terms of free-noise signal and additive noise, respectively, and $2\Re\left\{\mathsf{WVD}_{xn}\right\}$ is the disturbing cross-term (or interference term) that corrupts the auto-terms when visualizing WVD distribution. To suppress the cross-term interference components, the Cohen's class is considered for the bilinear transformation of multiple single-frequency signals. Definitely, SPWVD allows smoothing the distribution in time and frequency by utilizing the adjustable sampling windows independently, which can be written as follows:

$$\begin{aligned}\mathsf{SPWVD}_{y}^{\mathcal{G},\mathcal{H}}\left(t,f\right)=&\int_{-\infty}^{+\infty}\mathcal{G}\left(t\right)\mathcal{H}\left(f\right)\\ &y\left(t+\frac{\tau}{2}\right)y^{*}\left(t-\frac{\tau}{2}\right)e^{-j2\pi f\tau}d\tau,\end{aligned} \quad (4)$$

where $\mathcal{G}\left(t\right)$ and $\mathcal{H}\left(f\right)$ are the window functions in time and frequency domains, respectively.

B. LPI Signal Model

In a typical radar system, the complex sample of an intercepted LPI signal is disturbed by additive noise at the receiver

$$y\left(k\right)=x\left(k\right)+n\left(k\right)=Ae^{j\varphi(k)}+n\left(k\right), \quad (5)$$

where $x\left(k\right)$ is the noise-free signal specified by the amplitude $A=1$ and the instantaneous phase φ; $n\left(k\right)$ is the complex additive white Gaussian noise (AWGN). The instantaneous phase is determined by instantaneous frequency and phase offset [12]. The index k is termed for every interval T_s with the sampling rate f_s. The pulse period $\gamma_{pp}=1/f_{pp}$ is assumed to be larger than the signal pulse width γ_{pw}, where f_{pp} denotes the pulse repetition frequency. Thus, the duty cycle $\rho=\gamma_{pw}/\gamma_{pp}$ is less than 1. If $y\left(k\right)$ contains only the real component, the Hilbert transformation, denoted H, is applied to return a complex output as follows:

$$y'\left(k\right)=y\left(k\right)+j\mathsf{H}\left[y\left(k\right)\right]. \quad (6)$$

C. Data Generation with Image Processing

In this research, we comprehensive recognize 13 LPI radar waveform fashions: linear frequency modulation (LFM), rectangular (Rect), Costas code, Barker code, five polyphase codes (Frank, P1, P2, P3, and P4), and four polytime codes (T1, T2, T3, and T4). The detailed waveform parameters are summarized in Table I, where $f_s=100$ MHz is the sampling frequency, f_c is the center frequency, f_m is the fundamental frequency, B is the bandwidth, N is the number of samples, FH is the frequency hop, L_c is the code length, c_{pp} is the number of cycles per phase code. In addition, other parameters M, n_s, n_g, and n_p are the numbers of frequency steps, sub-codes, segments, and phase states, respectively. For performance evaluation, we produce a challenging dataset of LPI radar waveforms bearing a synthetic channel impairment, including multipath Rayleigh fading channels and AWGN, where signal-to-noise ratio (SNR) varies from -20 dB to $+10$

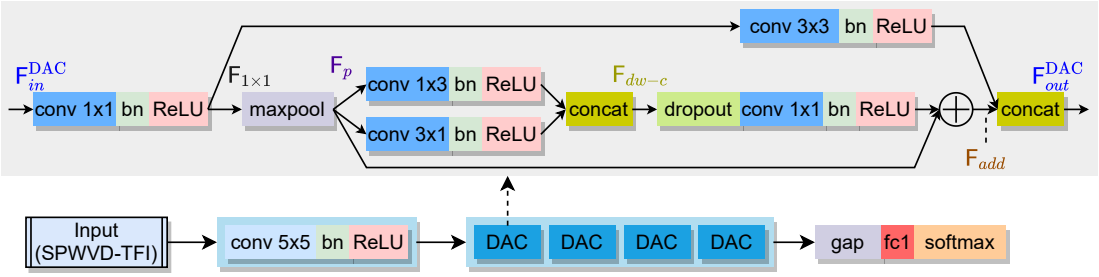


Fig. 2. Densely-accumulated convolutional neural network (DANet) with the detailed structure of DAC module.

TABLE I
WAVEFORM PARAMETER CONFIGURATION.

Waveforms	Parameters	Range of value
LFM	N	$[512, 1024]$
	B	$U\,(f_s/20, f_s/15)$
Rect	N	$[512, 1024]$
Costas	N	$[512, 1024]$
	FH	$\{3, 4, 5, 6\}$
	f_m	$U\,(f_s/32, f_s/25)$
Barker	L_c	$\{7, 11, 13\}$
	c_{pp}	$[2, 5]$
Frank	c_{pp}	$[3, 5]$
	M	$\{6, 7, 8\}$
P1, P2	c_{pp}	$[3, 5]$
	M	$\{6, 8\}$
P3, P4	c_{pp}	$[3, 5]$
	n_s	$\{36, 64\}$
T1, T2	n_p	2
	n_g	$\{4, 5, 6\}$
	N	$[512, 1024]$
T3, T4	n_p	2
	n_g	$\{4, 5, 6\}$
	N	$[512, 1024]$
	B	$U\,(f_s/20, f_s/15)$

Note: All waveforms are configured with $f_c = U\,(f_s/6, f_s/5)$.

dB with the step size of 2.0 dB. The multipath fading channel is configured as follows: the path delay $\tau = U\,(1, 1000)$ ns, the path gain $G = U\,(-20, 0)$ dB, and the maximum Doppler shift $f_{\mathrm{D}} = U\,(10, 1000)$ Hz. From radio signal to time-frequency representation, we produce 1,000 TFIs (corresponding 1,000 signals) per LPI waveform per SNR, that means, the whole dataset has 208,000 images for training model and evaluating performance.

With the SPWVD-based TFA technique, we transform raw LPI signals to color TFIs having the size of 112×112, where the TFIs of different radar waveforms of interest are illustrated in Fig. 1. To increase the performance of recognition model, we improve the quality of TFIs by adopting some digital image processing techniques for denoising, including median filtering and Otsu-based global thresholding. The median filtering process, which calculates each output pixel as the median value in a 3×3 region around the corresponding pixel in the input image, aims to reduce salt and pepper noise [16]. The thresholding process determines a global threshold by the Otsu's method to separate the background (containing noise) and the foreground (a.k.a., useful information) in an image. Consequently, the noise in SPWVD-TFI is reduced with the preservation of meaningful spectrum information.

III. DANET: CNN FOR TFA-BASED WAVEFORM RECOGNITION

In this section, we describe our proposed DANet for TFA-based waveform recognition, where the overall architecture is shown in Fig. 2. At the beginning, DANet is specified with the input size of 112×112 identical to the resolution of color TFIs. Subsequently, a processing unit comprising a convolutional (conv) layer with the kernels of size 5×5, a batch-normalization (bn) layer, and a rectified linear unit (ReLU) layer is specified to capture coarse representational features via a large receptive field.

As the core of network architecture, DANet is particularized by four feature extraction modules, namely densely accumulated connection (DAC), to effectively learn the meaningful information at multi-scale visual feature representations. As illustrated in Fig. 2, each DAC module is cleverly configured with multiple conv layers having different filter sizes, where the outcomes of these conv layers are associated via skip-connection (so-called residual connection) [17] and dense-connection [18]. In particular, each DAC module begins with a unit processing block (comprising one conv layer with the filter size of 1×1 followed by a bn layer, and a ReLU layer). A max pooling (maxpool) layer with the pool size of 3×3 is then arranged to reduce the computational cost of subsequent layers. With the stride of $(2, 2)$, the maxpool layer divides the spatial dimension of feature maps in half. To learn the time-frequency characteristics in the horizontal and vertical dimensions individually, two processing units, which are principally specified by two conv layers leveraging one-dimensional (1D) asymmetric filters of sizes 1×3 and 3×1, are organized in a parallel connection. Their output feature maps are stacked along the third dimension (or channel dimension) via a depth-wise concatenation (concat) layer. A dropout layer with the dropout probability of 0.5 is configured in each DAC module to prevent the network from overfitting. When the network goes deeper, the vanishing gradient problem arises in the training stage to reduce the efficiency of pattern learning. To handle this kind of problem, we adopt the skip connection using an element-wise addition layer, where the gradient identity from the maxpool layer is utilized to strengthen the gradient flow after a series of nonlinear operations. The

TABLE II
NETWORK CONFIGURATION.

DAMod	Output Size	No. filters (all conv layers)
First module	$56 \times 56 \times 32$	16
Second module	$28 \times 28 \times 64$	32
Third module	$14 \times 14 \times 128$	64
Fourth module	$7 \times 7 \times 256$	128

module is finalized with a concat layer to enrich the diversified features by gathering the outcome of the skip connection and the output of a processing block with a convolutional (conv) layer specified by the kernels of size 3×3 and the stride of $(2, 2)$. It is worth noting that two unit processing blocks with 1×1 conv layers are arranged in the structure to pool features across channels and to align the number of feature maps.

For a better description, we now present the feature flow in each the DAC module. Given an input of the module $\mathbf{F}_{in}^{\mathrm{DAC}} \in \mathbb{R}^{H \times W \times D}$ (where the size of the input is defined by the height H, the width W, and the depth D), the outcome of the first unit processing block, denoted as $\mathbf{F}_{1\times1}$, can be written by

$$\mathbf{F}_{1\times1} = \mathcal{F}_{1\times1}\left(\mathbf{F}_{in}^{\mathrm{DAC}}\right), \tag{7}$$

where $\mathcal{F}_{a\times b}$ denotes the unified operation of a convolution operation (with the filter size of $a \times b$), a normalization, and an activation. Then, the pooled output can be obtained by computing the maximum of each 3×3 region as follows:

$$\mathbf{F}_p = \mathcal{P}(\mathbf{F}_{1\times1}), \tag{8}$$

where $\mathcal{P}$ denotes the max pool operation which reduces the spatial size (i.e., height $\times$ width) of $\mathbf{F}_p \in \mathbb{R}^{(H/2)\times(W/2)\times D}$. It should be noticed that there is no parameter to learn in pooling layers and the subsequent feature maps have the same spatial size with $\mathbf{F}_p$. Subsequently, the first concat layer is executed to stack the feature maps as follows:

$$\mathbf{F}_{dw-c} = \mathcal{C}(\mathcal{F}_{1\times3}\left(\mathbf{F}_p\right), \mathcal{F}_{3\times1}\left(\mathbf{F}_p\right)), \tag{9}$$

where $\mathcal{C}$ represents the depth-wise feature maps concatenation. The skip connection with the add layer can be written as

$$\mathbf{F}_{add} = \mathbf{F}_p + \mathcal{F}_{1\times1}(\mathbf{F}_{dw-c}), \tag{10}$$

where $\mathbf{F}_{add}$ is the output of skip connection. The DAC module is finalized with the second concat layer, where its output can be represented as follows:

$$\mathbf{F}_{out}^{\mathrm{DAC}} = \mathcal{C}(\mathbf{F}_{add}, \mathcal{F}_{3\times3}\left(\mathbf{F}_{1\times1}\right)). \tag{11}$$

The above-described processing flow with notations can be referred in Fig. 2. It is observed that the DAC module halves the spatial size of feature maps, which allows the network to extract the intrinsic features at multi-scale representations and reduce the computational cost of subsequent conv layers. The detailed configuration of conv layers in DAC modules is summarized in Table II, in which the number of kernels and the size of output are provided sufficiently.

At the end of architecture, the network is finalized with a global average pooling (gap) layer, a fully connected (fc) layer and a softmax layer for classification. The number of hidden neurons of the fc layer is identical to the number of waveform classes in a given dataset. Regarding the registered architecture, DANet has around 447K trainable parameters, including weights and biases. Some primary network training properties are configured as follows: the optimizer is Adam with the gradient decay factor of 0.9, the maximum number of epochs is 90, the initial learning rate is 0.001 (drop 90% after 45 epochs), and the mini-batch size is 64. The network is evaluated by using a single GeForce GTX 1080Ti GPU. The generated dataset in Section II-C is split with 70% for training and 30% for testing.

IV. SIMULATION RESULTS AND DISCUSSIONS

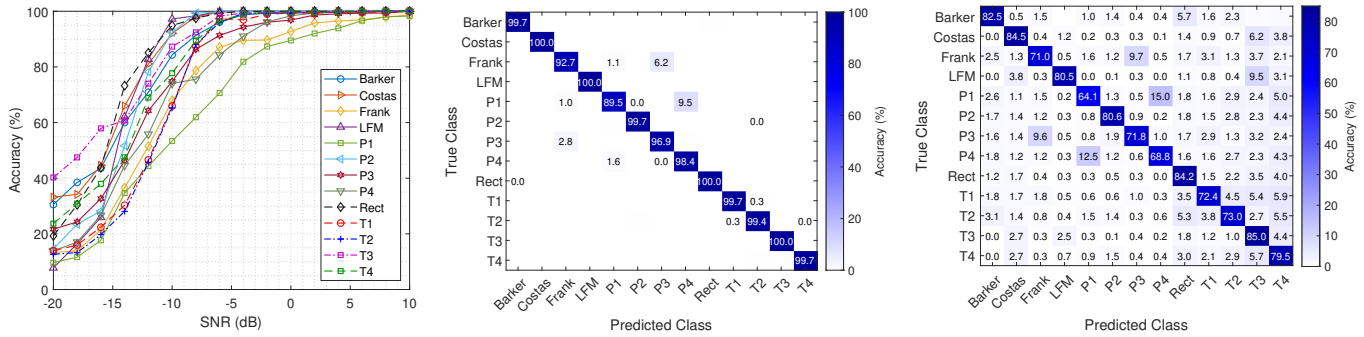
In this section, the performance of the proposed DL-based LPI radar waveform recognition method is evaluated on the dataset generated with 13 waveform types, in which we comprehensively produce four experiments: deep model robustness, ablation study and parameter investigation, and method comparison: accuracy and complexity. We measure the average accuracy (a.k.a., the percentage of correct classification) on the test set with several times, meanwhile, the complexity is assessed via the network size based on the number of parameters and the computational cost using floating point operations (FLOPs).

A. Model Robustness

In Fig. 3(a), we present the recognition results of 13 LPI waveforms on the test set, where the accuracy increases along the increment of SNR. Remarkably, some waveforms, such as Costas, LFM, Rect, and P2, obtain the recognition rate of over 92.0% at -10 dB SNR. Moreover, we report the waveform classification accuracy of DANet on the test set by the confusion matrices. From the confusion matrix at 0 dB SNR as shown in Fig. 3(b), the network achieves the average accuracy of 98.20%, in which many LPI waveforms are recognized with the accuracy greater than 96%. Remarkably, some such waveforms as Costas, LFM, Rect, and T3 reach 100% recognition rate, whereas some confusions occur with Frank and P1 (e.g., they are mostly confused with P3 by 6.2% and P4 by 9.5%). With respect to all SNR levels distributed in a range $[-20 : 10]$ dB, where the confusion matrix is given in Fig. 3(c), DANet presents the overall accuracy of 76.6%. Due to the multipath fading propagation channel coupled with the additive noise at low SNR, many waveforms are misclassified by DANet (for example, P1 and P4 are critically confused together by around $12.5 - 15.0\%$).

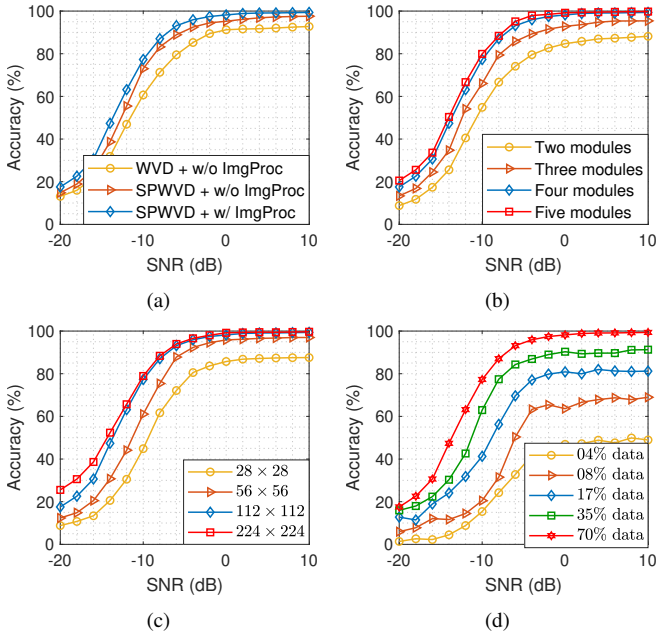
B. Ablation Study and Parameter Investigation

Fig. 4 shows the recognition results of ablation study and hyper-parameter investigation. We first analyze the effectiveness of SPWVD for time-frequency analyse and image processing for denoising via an ablation study, where the proposed strategy is compared with two variants (i.e., WVD without image processing and SPWVD without image processing) in Fig. 4(a). In general, without image processing, SPWVD is better than WVD by the overall accuracy of 6.14% while the accuracy increases by around 3.75% by applying image



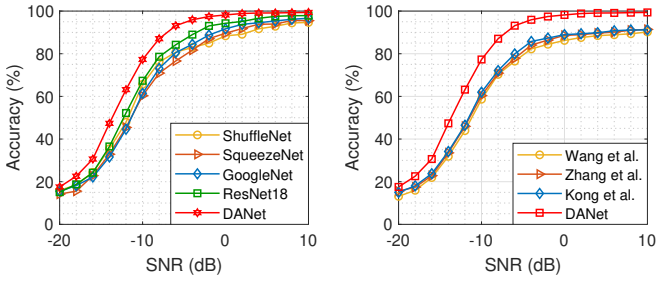
(a) (b) (c)

Fig. 3. The LPI radar waveform recognition accuracy: (a) detailed accuracy along SNR values, (b) confusion matrix at 0 dB SNR with the average accuracy of 98.20% and (c) confusion matrix at all SNR levels with the average accuracy of 76.63%.



(a) (b) (c) (d)

Fig. 4. Ablation study and hyper-parameters investigation with: (a) the effectiveness of SPWVD and image processing, (b) various number of DAC modules, (c) different image sizes, and (d) different training set sizes.



(a) (b)

Fig. 5. Method comparison: (a) with backbone networks and (b) with other CNNs specifically designed for TFI-based radar waveform recognition.

denoising for SPWVD-TFIs. We then investigate the influence of different hyper-parameters (such as the number of DAC modules, the image size of SPWVD-TFIs, and training data size) on the overall performance of our proposed method. In Fig. 4(b), the accuracy increases along the increment of the number of DAC modules configured in DANet because more diversified features are extracted at multiple deeper levels to gain a more discriminative pattern recognition. From the numerical results of DANet with different image sizes in Fig. 4(c), encoding higher resolution images for LPI signal representation can improve recognition accuracy thanks to the more details of time-frequency analysis. Notably, increasing the number of DAC modules and the size of SPWVD-TFIs will boost the computational cost and the memory resource for deep feature calculation and data storage, respectively. Experimentally, we select four modules and specify the image size of 112×112 to keep a nice balance between accuracy and complexity. In data-centric supervised learning methods, the dataset often has an crucial impact on the quality of classification model, hence we consider the influence of the training data size (i.e., the total number of training signals) in Fig. 4(d). With 4% data of 208,000 TFIs for training model, DANet presents the overall accuracy of 29.3% and improves by $9.68 - 13.92\%$ for each time of doubling the size up to 70% (around 145,600 signals). The numerical results illustrate that having sufficient training data is important to improve the performance of deep learning models.

C. Method Comparison: Accuracy and Complexity

In the last experiment, we compare the recognition accuracy between DANet with several backbone networks (which are originally introduced for the image classification task) and other CNNs (which are specifically designed for TFI-based radar waveform recognition). In Fig. 5(a), DANet outperforms ShuffleNet [19], SqueezeNet [20], GoogleNet [21], and ResNet18 [17] by the overall accuracy of around $5.49-9.89\%$. Notably, the accuracy of DANet is greater than that of ResNet (as the second best model of this test) by $6.96 - 10.85\%$ at $[-14, -4]$ dB SNR. Compared with the shufflenet unit in ShuffleNet, the fire module in SqueezeNet, the inception module in GoogleNet, and the residual block in ResNet18, the DAC module in our DANet learns intrinsic features more effectively. All comparison backbone CNNs are adapted with

the input size of 112×112 and trained with the same condition. In Fig. 5(b), there are big accuracy gaps between DANet with other CNNs proposed by Wang *et al.* [10], Zhang *et al.* [11], and Kong *et al.* [12], where DANet is considerably better by $10.01 - 12.27\%$. With the plain architecture designs (i.e., adopting a cascade structures of multiple conv layers and activation layers without skip-connection and dense-connection), these CNNs cannot acquire high-relevant features diversely to optimize learning efficiency.

Besides recognition accuracy, we also compare the deep networks in terms of computational complexity. Concretely, we measure the number of trainable parameters and the number of floating point operations (FLOPs). Fig. 6 shows the relationship between accuracy and computational cost, where the area of each bubble is proportional to the number of parameters. Despite producing the highest accuracy, DANet has a lightweight architecture with approximately 447K parameters and 83 megaFLOPs. In [10]–[12], deep CNNs were regularly designed with two-dimensional filter size of 3×3 and without the global average pooling layer, which increase the FLOPs of conv layers and the number of parameters of fc layers. Compared with ResNet18 and GoogleNet, ShuffleNet and SqueezeNet are more low-efficient by leveraging some advanced structures and techniques, such as grouped conv layers, unit conv layers with 1×1 replacing 3×3, and channel reduction of input activation.

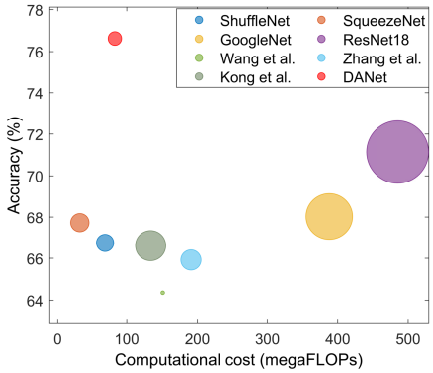


Fig. 6. Network comparison in terms of computational complexity with computational cost (megaFLOPs) and network size.

V. Conclusions

In this paper, we developed a high-performance LPI radar waveform recognition approach in automotive sensing systems. A novel CNN, namely DANet, was designated with specific modules to densely accumulate relevant features of SPWVD-TFIs at multiple representational maps. Furthermore, image processing techniques were leveraged to enhance image quality critically reduced by additive noise. In the simulations, DANet achieved the high accuracy of 13-LPI recognition and further outperformed several state-of-the-art CNNs in term of accuracy and computing expense. In future, we intend to improve the recognition performance of DANet with attention connection to selectively collect high relevant features and grouped convolution to reduce network complexity.

References

[1] F. Uysal, "Phase-coded FMCW automotive radar: System design and interference mitigation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 270–281, Jan. 2020.

[2] X. Zhang, L. Zuo, D. Yang, and J. Guo, "Coherent-like integration for PD radar target detection based on short-time fourier transform," *IET Radar, Sonar Navig.*, vol. 14, no. 1, pp. 156–166, Feb. 2020.

[3] T. Huynh-The, C.-H. Hua, V.-S. Doan, Q.-V. Pham, and D.-S. Kim, "Accurate deep CNN-based waveform recognition for intelligent radar systems," *IEEE Commun. Lett.*, pp. 1–1, 2021.

[4] T. Huynh-The, V.-S. Doan, C.-H. Hua, Q.-V. Pham, T.-V. Nguyen, and D.-S. Kim, "Accurate LPI radar waveform recognition with CWD-TFA for deep convolutional network," *IEEE Wirel. Commun. Lett.*, vol. 10, no. 8, pp. 1638–1642, 2021.

[5] T. Ravi Kishore and K. D. Rao, "Automatic intrapulse modulation classification of advanced LPI radar waveforms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 901–914, Feb. 2017.

[6] T. Huynh-The, C. Hua, and D. Kim, "Encoding pose features to images with data augmentation for 3-D action recognition," *IEEE Trans. Ind. Inform.*, vol. 16, no. 5, pp. 3100–3111, 2020.

[7] C. Hua, T. Huynh-The, and S. Lee, "Convolutional networks with bracket-style decoder for semantic scene segmentation," in *Proc. 2018 IEEE Int. Conf. Syst. Man Cybern. (SMC)*, 2018, pp. 2980–2985.

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

[9] 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.

[10] C. Wang, J. Wang, and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," in *Proc. 2017 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2017, pp. 2437–2441.

[11] M. Zhang, M. Diao, and L. Guo, "Convolutional neural networks for automatic cognitive radio waveform recognition," *IEEE Access*, vol. 5, pp. 11 074–11 082, Jun. 2017.

[12] S. Kong, M. Kim, L. M. Hoang, and E. Kim, "Automatic LPI radar waveform recognition using CNN," *IEEE Access*, vol. 6, pp. 4207–4219, Jan. 2018.

[13] L. Gao, X. Zhang, J. Gao, and S. You, "Fusion image based radar signal feature extraction and modulation recognition," *IEEE Access*, vol. 7, pp. 13 135–13 148, Jan. 2019.

[14] 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 Process. Lett.*, vol. 27, pp. 1275–1279, Jul. 2020.

[15] S. Zhang, A. Ahmed, and Y. D. Zhang, "Sparsity-based time-frequency analysis for automatic radar waveform recognition," in *Proc. 2020 IEEE Int. Radar Conf. (RADAR)*, Apr. 2020, pp. 548–553.

[16] T. Huynh-The, O. Banos, S. Lee, B. H. Kang, E. Kim, and T. Le-Tien, "NIC: A robust background extraction algorithm for foreground detection in dynamic scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 7, pp. 1478–1490, July 2017.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. 2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2261–2269.

[19] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," *arXiv preprint arXiv: 1707.01083*, 2017.

[20] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and $<$ 0.5MB model size," *arXiv preprint arXiv: 1602.07360*, 2016.

[21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *arXiv preprint arXiv: 1409.4842*, 2014.