Separable Attention Capsule Network for Signal Classification

SHAOQING LIU\textsuperscript{1}, HUILING LIU\textsuperscript{2}, CHEN YANG\textsuperscript{2}, SHUYUAN YANG\textsuperscript{2}, (Member, IEEE), AND MIN WANG\textsuperscript{3}

\textsuperscript{1}Department of Basic Courses, Shandong University of Science and Technology, Jinan 250031, China
\textsuperscript{2}School of Artificial Intelligence, Xidian University, Xi’an 710071, China
\textsuperscript{3}Key Laboratory of Radar Signal Processing, Xidian University, Xi’an 710071, China

Corresponding author: Shuyuan Yang (syyang2009@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 91438103, Grant 61771376, and Grant 61771380.

ABSTRACT In this paper, a new Separable Attention Capsule Network (SACN) is proposed for signal classification. SACN is a light-weight network composed of multi-channel separable convolution layer, attention module and classification layer. First, depth-wise convolution is employed to extract features of signals in a low-complexity manner, and the multi-channel network structure is designed to increase the network width to improve the diversity of features of signals. Then a channel attention module is followed by a capsule network whose element contains a group of neurons. This attention module can explore the interdependence among channels to use global information to selectively strengthen some important channels, thus achieving the improvement of generalization ability of SACN. Some experiments are taken on several datasets with communication and radar signals, and the comparison results prove the efficiency of SACN and the superiority to its counterparts.

INDEX TERMS Separable convolution, multi-channel, capsule network, channel attention.

I. INTRODUCTION

With the rapid development of radio technology, there are increasing types of radio signals and their classification has been an important topic in the field of signal processing [1], [2]. For example, early radar devices are relatively simple, and the signal classification mainly relies on some empirical parameters, such as Time of Arrival (TOA) and Pulse Repetition Frequency (PRF). However, with the increasing complexity of the electromagnetic environment, the statistical characteristics of received signals have been remarkably destroyed. Thus empirical parameters could not distinguish various kinds of signals accurately. Some time-frequency features are then proposed for classification [3], [4], [6]–[8]. For example, the cyclic stationary features are explored to identify radar signals [5]. The second order cyclic stationary of multi-phase keying signals is estimated and used for the subsequent classification [7].

In recent years, more and more machine learning models have been used for signal classification, via the empirical features (such as intra-pulse-modulation features) [9]–[14]. With the development of deep learning technology, Deep Neural Networks (DNNs) have been proposed to classify the time-frequency maps of signals [15]–[17]. DNN not only avoids the tedious “feature engineering”, but also brings a new paradigm for “end-to-end” classification of signals.

Although deep learning has shown its effectiveness in signal classification, most of the available methods have the following limitations: 1) They first extract the time-frequency features and then input into DNNs, which can not well explore the capability of deep learning; 2) They often exhibit low classification accuracy when the Signal-Noise Ratio (SNR) is low, especially for negative SNR; 3) They can only classify signals with large differences, such as signals with different modulation types. However, it is difficult to distinguish signals with different carrier frequencies and encoding modes; 4) They often have large amount of network parameters and require a large training dataset, which limits their application in mobile and embedded devices with limited memory and resources.

In this paper, a new Separable Attention Capsule Network (SACN) is proposed for signal classification. Different with traditional CNNs that employ a large number of samples to learn discriminative features of signals, SACN uses capsule
composed by a group of neurons to replace a single neuron. The dynamic routing algorithm is used to replace the pooling operation. Borrowing from the depth-wise convolution [20] and MobileNet [19], we construct a separable convolution layer using $1 \times 1$ convolution kernels. This convolution can separate the channel convolution to reduce the number of addition and multiplication operations. Thus SACN is of light-weight and the features can be extracted with a small number of samples. Moreover, inspired by the idea of parallel convolution kernel, the multi-channel structure is designed to increase the network width to improve the diversity of signal features, and the capsule network is used to improve the generalization ability of the model. Finally a channel attention module follows a capsule network, to explore the interdependence among channels to use global information to selectively strengthen some important channels.

Some experiments are taken on several datasets with communication and radar signals, to validate our proposed SACN, and the results are compared with its counterparts. Three signal datasets include the modulation and coding mode classification, and the task of frequency parameter identification. The experimental results show that SACN has an overall accuracy of 94.5%, 96.1% and 98.2% on the three datasets, which proves that SACN can achieve accurate and robust classification. Moreover, the network complexity is also analyzed and compared with the ordinary and separable convolution. The determination of the network parameter is also analyzed.

The remainder of this paper is organized as follows. In section II, a detailed description of the proposed SACN is given. In section III, several experiments are taken to demonstrate the effectiveness of SACN. Section IV draws the conclusion of this paper.

II. SEPARABLE ATTENTION CAPSULE NETWORK

This section introduces the structure of the proposed network, including the separable convolution block, the attention block, the network architecture and the learning algorithm, along with a complexity analysis of the convolution module.

A. SEPARABLE CONVOLUTION

In the traditional convolution, all the feature maps from different channels are involved in the operation. Inspired by the $1 \times 1$ cross-channel convolution, we design a separable convolution module (named as SConv-module in this paper). The module structure is shown in Fig.1, where the separable convolution operation is named as SConv.

In Fig.1, SConv1 and SConv2 represent separable convolution layers, and Batch Normalization (BN) operation and ReLU activation function are also adopted. As shown in Fig.2, the SConv contains a layer-by-layer operation $\tilde{C}_1$ and a pixel-by-pixel operation $\tilde{C}_2$, as follows:

$$\tilde{C}_1 : F_c = S_c \cdot K_c$$  \hspace{1cm} (1)

$$\tilde{C}_2 : F_u = \sum_{c=1}^{C} F_c \ast \text{ker}_u$$  \hspace{1cm} (2)

where $c$ indicates the $c$-th channel, $S_c$ represents the feature matrix corresponding to the $c$-th channel of the input, $\ast$ is the inner product operation and $K_c$ is the convolution kernel of the $c$-th channel. $C$ and $U$ are the number of input and output channels respectively. $F_c$ represents the feature matrix corresponding to the $c$-th channel of the output and $\ast$ is the convolution operation. $F_u$ represents the output feature of the $u$-th channel and $\text{ker}_u$ represents $1 \times 1$ convolution kernel used in the $u$-th channel.

The input of the separable convolution module is the signal $X$, and the output is the feature matrix $B$. In the first layer of the separable convolution, the signal is calculated using (1) and (2) to obtain the primary feature $X^1$. The mean $\mu$ and variance $\sigma^2$ of $X^1$ are calculated as,

$$\mu = \frac{1}{m} \sum_{i=1}^{m} X^1_i$$  \hspace{1cm} (3)

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (X^1_i - \mu)^2$$  \hspace{1cm} (4)

where $m$ is the number of channels of the primary feature $X^1$. In the SConv-module, the number of convolution kernels of SConv1 is 8, and the size of the convolution kernel is $1 \times 1$. The primary feature $X^1$ is then processed by BN operation to make data follow the Gaussian distribution with mean 0 and variance 1. It is beneficial to enhance the back-propagation gradient and speed up the convergence in the network training. The BN operation is as follows:

$$y_i = \gamma \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta$$  \hspace{1cm} (5)

where $y_i$ represents the $i$-th element of the normalized feature, $\beta$ and $\epsilon$ are parameters to be learned in the network, and the mean and variance are calculated from the formulas (3) and (4). After the normalization by BN layer, the normalized
feature can be obtained. Then the primary non-linear feature is obtained by using the ReLU function for nonlinear mapping of the normalized feature. The activation function is as follows,

\[ X^2 = \max(0, y) \]  

(6)

In the second layer of separate convolution, the primary non-linear feature \( X^2 \) is also calculated by using formulas (1) and (2), and then the quadratic feature is obtained. The number of convolution kernels of SConv2 is 16, and the size of convolution kernel is \( 1 \times 2 \). Then the primary feature is normalized by using Batch Normalization operation. The output feature matrix \( B \) of this separable convolution module is then obtained.

**B. MULTI-CHANNEL SEPARABLE CONVOLUTION MODULE**

Taking the above \( SConv-module \) as a fundamental unit, we construct a multi-channel separable convolution module (\( M-SConv-module \)) that has multiple \( SConv-modules \). As shown in Fig.3, six-way \( SConv-module \) is used to extract features of signals. The outputs of multiple \( SConv-module \) are contacted to combine the features together.

**FIGURE 3.** M-SConv module.

Assuming that \( SConv-module-k \) represents the \( k \)-th \( SConv-module \) of the \( M-SConv-module \). Denote \( X \) as the input of \( SConv-module-1 \) to \( SConv-module-6 \), the feature extraction is carried out in sequence as described in section II.A to obtain the output feature of the six separable convolution modules, which are denoted as \( B_i, i = 1, \ldots, 6 \) respectively. The six groups of output feature are contacted to obtain multi-dimensional features \( Mul_B \),

\[ Mul_B = B_1 \oplus B_2 \oplus B_3 \oplus B_4 \oplus B_5 \oplus B_6 \]  

(7)

Next, the convolution layer (Conv1) is used to extract multi-dimensional feature in a hierarchical manner. The convolution of this layer uses 32 convolution kernels with the size of \( 1 \times 1 \), followed by Batch Normalization layer. The dropout ratio value is set as 0.5. Then another convolution layer (Conv2) is performed, which is composed of 32 convolution kernels with the size of \( 1 \times 2 \), and the Batch Normalization operation is carried out to normalize the feature. After extracting feature through the M-SConv-module, we can obtain the convolution feature \( F \).

**C. ATTENTION MODULE**

In this section, the constructed multi-channel separable convolution module is combined with a channel attention module and a spatial attention module. Similar to “Squeeze-and-Excitation” (SE) in SENet [25], this attention module can explore the interdependence among channels to explore the global information of signals to selectively strengthen some important channels, thus improving the generalization ability of the network. The schematic diagram of the channel attention is shown in Fig.4.

**FIGURE 4.** Scheme of the channel attention.

In our work we employ the structure in Fig.4 and the convolution feature \( F \) is taken as the input of the channel attention block. The compression operation \( F_{sq} \) averages the information of each channel. The compressed feature \( z \) with dimensionality \( 1 \times 1 \times C \) is obtained. An excitation operation \( F_{ex} \) follows \( F_{sq} \), where \( z \) is mapped to a middle feature with size \( 1 \times 1 \times (Cr) \) \((r) \) is the compression ratio) using the first fully connected layer and activation function. Then the second fully connected layer and activation function are used to restore the middle feature to a feature \( z \) with size \( 1 \times 1 \times C \). Finally, we carry out a channel attention calibration operation \( F_{scale} \) to obtain the feature \( \tilde{X} \). \( \tilde{X} \) is taken as the input of the capsule block, which is introduced in section II.D, and obtain the output vector \( v \). In capsule network, eight 8-dimensional primary capsule layers and eight 12-dimensional signal capsule layers are chosen. The spatial attention can perform feature weighting and enhances local feature by capturing the global information of signals.

**D. NETWORK ARCHITECTURE**

The capsule network [23] is proposed by Hinton, where the input vector is recombined to obtain the base capsule layer \( u \). The dynamic routing algorithm is used for selective connection between \( u \) and signal capsule layer \( T \), and the connection weight is \( W \). The fully connected neural network is similar to the selective connection of the capsule network, but it adds the coupling coefficient \( c \) in the summation. The output of the \( j \)-th signal capsule is:

\[ T_j = \sum_i c_{ij}W_{ij}\tilde{u}_i \]  

(8)

where \( w_{ij} \) is the connection weight between the \( i \)-th base capsule and the \( j \)-th signal capsule, \( u_i \) is the vector represented by the \( i \)-th base capsule and \( c_{ij} \) is the coupling coefficient between the \( i \)-th base capsule and the \( j \)-th signal capsule.
The coupling coefficient $c$ is calculated according to the following formulas (the initial value of $b_{ij}$ is set as zero):

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}$$ (9)

$$b_{ij} \leftarrow b_{ij} + W_i u_j v_j$$ (10)

**FIGURE 5.** The overall structure of the separable attention capsule network.

The overall network structure is shown in Fig. 5, where the capsule output is denoted as $V$, and $\| \cdot \|$ is used to represent the norm of the vector. Then the capsule output of the $j$-th capsule can be written as,

$$V_j = \frac{\| T_j \|}{1 + \| T_j \|} \cdot T_j$$ (11)

The spatial attention block is then used to improve the features [24]. As shown in Fig. 5, the capsule output is taken as the input of a learnable system $g(\cdot)$, from which the intermediate weight of spatial attention can be obtained. Then the probability vector can be generated by using the intermediate weight of attention.

After the feature weighting by the spatial attention block, a weighted vector $A_{ij}$ is then obtained. Then the FC Layer is used for feature mapping of $A_j$, and the output dimensionality of the fully connected layer is equal to the number of categories. Finally, the SoftMax classifier is used for signals classification.

**E. COMPUTATION COMPLEXITY**

CNN requires a large number of samples to be trained for feature extraction, while the capsule network can use relatively few samples to obtain comparable generalization ability. The proposed SACN is of light-weight for employing the separable convolution. In this section, the number of parameters of separable convolution and ordinary convolution are analyzed. Assuming that the size of the convolution kernel is $D_1 \times D_2$, $C$ is the number of input channels and $U$ is the number of output channels. The depthwise separable convolution uses $C$ kernels of size $D_1 \times D_2$ to obtain $C$ feature maps. Next, $1 \times 1$ convolution kernels are used to fuse the $C$ feature maps in a pointwise manner. The number of parameters of this separable convolution can be calculated by the following formula:

$$(D_1 \times D_2) \times C + (1 \times 1) \times U = D_1 D_2 C + U$$ (12)

However, the number of parameters of ordinary convolution can be calculated by the following formula:

$$(D_1 \times D_2) \times C \times U = D_1 D_2 C U$$ (13)

Therefore, the reduction ratio of separable convolution to ordinary convolution can be calculated,

$$\frac{D_1 D_2 C + U}{D_1 D_2 C U} = \frac{1}{U} + \frac{1}{D_1 D_2 C}$$ (14)

**III. SIMULATIONAL RESULTS**

In this section, three radio signal datasets are used to validate the performance of our proposed SACN. The parameters setting and the classification results are detailed, and the results are compared with its counterparts.

**A. DATASETS**

Dataset 1: The joint modulation and coding dataset JCCM. The dataset contains short-wave frequency band signals, which include signals with three types of coding modes and three types of modulation modes. The three coding modes are Hamming code, 216 non-systematic convolutional code with one-half code rate, and 432 non-systematic convolutional code with three-quarter code rate. Three modulation modes are four-phase shift keying QPSK, eight-phase shift keying 8PSK, and frequency shift keying FSK. In each class, 40,000 samples are used for the training.

Dataset 2: The modulation dataset ELS. It has signals with 12 types of modulation modes, including LFM, NLFM, BPSK, QPSK, 8PSK, FSK, AM, FM, QAM16, GFSK, CPFSK and WBFM. The signal carrier frequency is 250MHz and the sampling frequency is 2GHz. Each signal includes 10,000 samples with length 2048. Totally there are 120,000 samples for the network training.

Dataset 3: The radar signal dataset RSS. The dataset contains 12 kinds of radar signals with four modulation modes and multiple frequency parameters. Four modulation modes are LFM, BPSK and QPSK, complex modulation (Complex) and single frequency modulation CW signals. In LFM, the frequency parameters have different directions to generate four modes of signals. Each type of signals has 10,000 samples, with the length 2048. A detailed description of the dataset is shown in Table 1.

**B. PARAMETER SETTING**

In the training, the batch size is set as 128 and the maximum number of iterations is set as 100. The loss function of the network is softmax cross entropy loss, and the optimization algorithm is based on the adaptive moment estimation [26]. In the network training, the value of loss function and accuracy on the training set and the verification set, are recorded. Three indexes are used to evaluate the performance of SACN: Overall Accuracy (OA), Average Accuracy (AA), and Kappa.
TABLE 1. Description of the RSS Dataset.

| Signal   | Frequency parameter | Modulation |
|----------|---------------------|------------|
| LFMI     |                     | 1kHz–50MHz |
| LFMI     | -                   | 1kHz–50MHz |
| LFMI     | +                   | 50MHz–500MHz |
| CW1      | Carrier Frequency   | CW         |
| CW1      | (5.0GHz–5.4GHz)     |            |
| CW2      | (5.4GHz–5.7GHz)     |            |
| CW3      | (5.7GHz–6.0GHz)     |            |
| Complex1 | Mode of FM          | Complex    |
| Complex2 | Hyperbolic Sine FM  |            |
| Complex3 | Exponential FM      |            |
| complex3 | Tangent FM          |            |
| BPSK     | 2PSK                |            |
| QPSK     | 4PSK                |            |

C. THE EVALUATION OF THE PROPOSED SACN

In this section, we use the above parameter setting to obtain a trained network and then test it. Under the same experimental conditions, 30 independent experiments are taken and the average results are calculated. The calculated numerical results of the three datasets are shown in Table 2.

TABLE 2. Performance of SACN on The three datasets.

| Dataset | OA(%)  | AA(%)  | Kappa   |
|---------|--------|--------|---------|
| JCCM    | 94.5 ± 0.38 | 94.5 ± 0.31 | 0.944 ± 0.009 |
| RSS     | 96.2 ± 0.14 | 96.1 ± 0.19 | 0.960 ± 0.012 |
| ELS     | 98.3 ± 0.17 | 98.2 ± 0.17 | 0.982 ± 0.007 |

From the table we can observe that the proposed SACN can achieve high classification accuracy (including OA, AA and Kappa) on the three datasets. Next, several related algorithms are used to compare with SACN, including LSTM [27], Inception [22], CLDNN [28], Densenet [29], Cov-Net [30], Resnet [30], Random Forest (RF) [31] and Support Vector Machine (SVM). Among them, CLDNN combines CNN with LSTM, thus achieving better results than CNN and LSTM. The Inception network expands the network width through the decomposition of convolution kernels. Cov-Net is one of the earlier deep learning based automatic modulation classification approaches, which contains two convolutional layers and two fully connected layers. It. By analyzing the signal entropy as the feature, three features are selected and then random forest and SVM is used for the classification. The Radial Basis Function (RBF) is adopted in SVM, whose width is determined using ten-fold cross validation in the range of [0.0001, 0.001, 0.01, 0.1, 1, 10, 100]. ResNet-50 is also used for a comparison, with multiple convolutional blocks and identity blocks. The classification results of different methods are shown in Fig.6, where the horizontal axis represents different algorithms while the vertical axis represents the overall accuracy of different algorithms on each dataset. It can be observed from Fig.6 that the performance of SACN is better than that of the other six machine learning based methods.

Next, we calculate the number of network parameters of SACN. The separable convolution in the network is replaced by the ordinary convolution, but the network structure remains unchanged, so obtaining a comparative network, OACN. The number of network parameters and the classification results on the JCCM dataset (with the ratio of training samples being 0.4) are shown in Table 3. From it we can observe that SACN has a remarkable reduction of network parameters by employing the separable convolution. Under the same hierarchical structure, the parameter number of SACN is only 45% of OACN. Therefore, SACN is easy to be implemented on mobile devices.

TABLE 3. The number of parameters for different networks.

| Network structure | SACN   | OACN   |
|-------------------|--------|--------|
| OA(%)             | 90.3 ± 0.39 | 90.2 ± 0.31 |
| #Network parameters | 728,628 | 1,584,564 |

Moreover, the influence of the capsule block on the network performance is also analyzed, by deleting the capsule block in SACN to obtain a corresponding SAN. In SAN, we remove the capsule network structure in the SACN, and use the spatial attention block to directly weight the features processed by the channel attention block. The classification block of SAN is the same with that of SACN. Then we vary the ratio of training samples to test SACN and compare it with SAN, and the results are shown in Table 4.

It can be observed that under the same number of training samples, the performance of SACN is better than that of SAN, which verifies the effectiveness of our constructed network. In order to achieve the same classification accuracy,
TABLE 4. Network performance under different training sample sizes.

| Network | dataset | OA(%) | AA(%) | Kappa | Ratio |
|---------|---------|-------|-------|-------|-------|
| SACN    | RSS     | 92.9 ± 0.54 | 92.8 ± 0.58 | 0.928 | 0.35  |
| SACN    | RSS     | 96.2 ± 0.47  | 96.1 ± 0.42  | 0.960 | 0.50  |
| SAN     | RSS     | 92.5 ± 0.72  | 92.5 ± 0.80  | 0.924 | 0.50  |
| SAN     | RSS     | 96.1 ± 0.59  | 96.0 ± 0.58  | 0.960 | 0.80  |
| SACN    | ICCM    | 94.5 ± 0.35  | 94.5 ± 0.36  | 0.944 | 0.50  |
| SAN     | ICCM    | 91.3 ± 0.44  | 91.3 ± 0.40  | 0.912 | 0.50  |
| SACN    | ELS     | 98.3 ± 0.29  | 98.2 ± 0.33  | 0.982 | 0.50  |
| SAN     | ELS     | 92.6 ± 0.39  | 92.5 ± 0.27  | 0.925 | 0.50  |

the number of samples required by SACN is less than that of SAN, indicating that the capsule module can enhance the network generalization performance and reduce the number of training samples. Also, the robustness of SACN is illustrated by analyzing its performance under different Signal-Noise Ratios (SNRs). The number of separable convolution layers in the SConv-module used in SACN is 2. From the results we can observe an improvement of SACN over SAN, for all the three datasets and three guidelines.

For the RSS dataset, the influence of the number of separable convolution layers in SConv-module on the performance of the SACN is analyzed. We use SConv-j to indicate that the SConv-module in SACN contains j separable convolutional layers. The performance of the networks with different separable convolutional layers is investigated, and the results are shown in Fig.7. It can be seen that SConv-2 has the highest overall accuracy, and the training time is less than SConv-3. Therefore, we set the number of separable convolutional layers in SConv-module as 2.

FIGURE 7. The overall accuracy of SACN with different number of separable convolutional layers.

The number of samples required by SACN is less than that of SAN, indicating that the capsule module can enhance the network generalization performance and reduce the number of training samples. Also, the robustness of SACN is illustrated by analyzing its performance under different Signal-Noise Ratios (SNRs). The number of separable convolution layers in the SConv-module used in SACN is 2. From the results we can observe an improvement of SACN over SAN, for all the three datasets and three guidelines.

accuracies of 12 signals is above 90%, which reflects the effectiveness and robustness of the proposed SACN.

IV. CONCLUSION AND FUTURE WORK
In this paper, a deep multi-channel separable attention capsule network is proposed for signal classification. A separable module is designed to reduce the number of parameters in the network. The channel and spatial attention are explored separately for more comprehensive feature extraction. The network width is enlarged by using the parallel convolution module, and the capsule network is used to improve the generalization ability of the model. Some experiments are taken on several datasets to validate our proposed method, and the results are compared with its counterparts. The experimental results on three datasets show that the network can recognize the coding, frequency parameter and modulation accurately, and reduce the complexity of the network to some extent.

REFERENCES
[1] D. J. Milojević and B. M. Popović, “Improved algorithm for the deinterleaving of radar pulses,” IEE Proc. F (Radar Signal Process.), vol. 139, no. 1, pp. 98–104, 1992.
[2] J. Liu, J. P. Y. Lee, L. Li, Z.-Q. Luo, and K. M. Wong, “Online clustering algorithms for radar emitter classification,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1185–1196, Aug. 2005.
[3] D. R. Wilkinson and A. W. Watson, “Use of metric techniques in ESM data processing,” IEEE Proc. F-Commun., Radar Signal Process., vol. 132, no. 4, pp. 229–232, Jul. 1985.
[4] C.-S. Shieh and C.-T. Lin, “A vector neural network for emitter identification,” IEEE Trans. Antennas Propag., vol. 50, no. 8, pp. 1120–1127, Aug. 2002.
[5] L. Li and H. Ji, “Radar emitter recognition based on cyclostationary signatures and sequential iterative least-square estimation,” Expert Syst. Appl., vol. 38, no. 3, pp. 2140–2147, Mar. 2011.
[6] S. Qian and D. Chen, “Joint time-frequency analysis,” IEEE Signal Process. Mag., vol. 16, no. 2, pp. 52–67, Mar. 1999.
[7] L. Cohen, “Time-frequency distributions-a review,” Proc. IEEE, vol. 77, no. 7, pp. 941–981, Jul. 1989.
[8] G. X. Zhang and H. N. Rong, “Improved quantum-inspired genetic algorithm based time-frequency analysis of radar emitter signals,” in Proc. 2nd Int. Conf. Rough Sets Knowl. Technol., Toronto, ON, Canada, 2007, pp. 484–491.
[9] G. Liu, G. Y. Zhang, and Y. Yu, “Intra-pulse modulation recognition of radar signal based on random forest,” Telecom Sci., vol. 32, no. 5, pp. 69–76, 2016.
[10] Z. L. Tang, X. N. Yang, and J. D. Li, “Study on fractal features of modulated radio signal,” Acta Phys. Sinica, vol. 60, no. 5, pp. 550–556, 2011.
[11] X. W. Wang, “A new method for time and frequency analysis of radar signal,” Ship Electron. Countermeasures, vol. 29, no. 6, pp. 67–69, 2006.
[12] L. W. Wu, Y. Q. Zhao, Z. Wang, F. Y. O. Abdalla, and G. Ren, “Specific emitter identification using fractal features based on box-counting dimension and variance dimension,” in Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT), Bilbao, Spain, 2017, pp. 226–231.
[13] G. X. Zhang, L. Z. Hu, and W. D. Jin, “Analysis of radar source signals and pulse-inside characteristics,” J. Infr. Millim. Waves, vol. 23, no. 6, pp. 477–480, 2004.
[14] N. Petrov, I. Jordanov, and J. Roe, “Radar emitter signals recognition and classification with feedforward networks,” Procedia Comput. Sci., vol. 22, pp. 1192–1200, 2013.
[15] A. Ali and F. Yangyu, “Automatic modulation classification using deep learning based on sparse autoencoders with nonnegativity constraints,” IEEE Signal Process. Lett., vol. 24, no. 11, pp. 1626–1630, Nov. 2017.
[16] C. Wang, J. Wang, and X. Zhang, “Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2017, pp. 2437–2441.
[17] 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, Apr. 2017.
[18] 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 <0.5 MB model size,” 2016, arXiv:1602.07360. [Online]. Available: http://arxiv.org/abs/1602.07360
[19] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” 2017, arXiv:1704.04861. [Online]. Available: http://arxiv.org/abs/1704.04861
[20] L. Sifre and S. Mallat, “Rigid-motion scattering for texture classification,” Comput. Sci., vol. 3559, pp. 501–515, 2014.