IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 20, 2017, accepted June 10, 2017, date of publication June 15, 2017, date of current version July 3, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2716191

Convolutional Neural Networks for Automatic Cognitive Radio Waveform Recognition

MING ZHANG, MING DIAO, AND LIMIN GUO
College of Information and Telecommunication, Harbin Engineering University, Harbin 150001, China
Corresponding author: Limin Guo (guolimin@hrbeu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61201410 and Grant 61571149, and in part by the Fundamental Research Funds for the Central Universities under Grant HEUCF170802.

ABSTRACT Cognitive radio technology is an important branch in the field of wireless communication, and automatic identification is a major part of cognitive radio technology. Convolutional neural network (CNN) is an advanced neural network, which is the forefront of application in the digital image recognition area. In this paper, we explore CNN in an automatic system to recognize the cognitive radio waveforms. Excitedly, it is a more effective model with high ratio of successful recognition (RSR) under high power background noise. The system can identify eight kinds of signals, including binary phase shift keying (Barker codes modulation) linear frequency modulation, Costas codes, Frank code, and polytime codes (T1, T2, T3, and T4). The recognition part includes a CNN classifier. First, we determine the appropriate architecture to make CNN effective for proposed system. Specifically, we focus on how many convolutional layers are needed, what appropriate number of hidden units is, and what the best pooling strategy is. Second, we research how to obtain the image features into CNN that based on Choi–Williams time-frequency distribution. Finally, by means of the simulations, the results of classification are demonstrated. Simulation results show the overall RSR is 93.7% when the signal-to-noise ratio is −2dB.

INDEX TERMS Cognitive radio, radar countermeasures, waveform recognition, time-frequency distribution, convolutional neural network.

I. INTRODUCTION

Cognitive radio is an agile way to perceive the electromagnetic spectrum, the wireless communicated environment and recognition of the waveforms etc. It is crucial to make efficient and effective operation for the future communication. Many scholars have devoted themselves in automatic recognition area, because of its important applications in cognitive radio. They have been looking for a practical and effective method to make the recognition system more intelligent and more robust. However, modern communication and radar systems usually have low instantaneous power. And some classification approaches that rely on instantaneous power are no longer applicable. A second-order cyclostationarity-based algorithm is utilized for the recognition of communication signals, such as spatial multiplexing (SM) orthogonal frequency-division multiplexing (OFDM) signals, Alamouti (AL)-coded OFDM signals and SC-FDMA signals [1], [2]. The experimental results show a good identification performance with phase noise, channel conditions etc. Time-frequency techniques can gather the power of signals in time-frequency domain [3]. In [4] and [5], atomic decomposition (AD) is investigated in the frame of complex radar signal detection and classification, and a digital channelized receiver is presented for the interception of a wide variety of signals (LFM, PSK, FSK and continuous wave). The RSR is more than 90% at SNR of 0dB by using the short-time Fourier transform (STFT). J. Lundén introduces a supervised classification system based on features that is extracted from the intercepted pulse compression waveforms [6]. The intercepted waveforms are classified into eight classes based on Wigner time-frequency distributions (WD) and Choi-Williams time-frequency distribution (CWD). The overall correct classification rate achieves 98% at SNR of 6 dB. Ming improves the method of Lundén and achieves great results [7]. Zeng D. uses the Rihaczek distribution (RD) and Hough transform (HT) to concentrate the energy on time-frequency plane and then derives two new characteristic features [8]. The algorithm is better than the majority of algorithms just through ambiguity function. In [9], a radar waveform

recognition algorithm which is based on random projections and sparse classification (SC) is presented. The algorithm can improve the information completeness, efficiency and robustness of noise. Three kinds of waveforms (LFM, FSK and PSK) are classified and the ratio of successful recognition (RSR) is over 90% at SNR of 0dB.

In fact, convolutional neural network (CNN) has been proposed in image recognition and computer vision fields, offering improvements over classical neural network on many tasks [10], [11]. Recently, CNN has been addressed for speech recognition [12], [13], computer vision [14], handwritten recognition [15] etc. Specifically, Abdel-Hamid introduces a novel framework of model spectral correlations where convolutional weights are shared over limited frequency regions. It is a technique known as limited weight sharing (LWS) [12]. After that, he discusses how to reduce the further error rate by using CNNs [13]. Experimental results show that CNNs reduce the error rate by 6%-10% compared with deep neural networks (DNNs) on the speech recognition tasks. X. Niu and C. Suen present a hybrid model of two superior classifiers CNN and Support Vector Machine (SVM), which have proven results in recognizing different types of patterns [15]. In their model, CNN works as a trainable feature extractor and SVM performs as a recognizer. Experiments show the RSR achieves more than 99.81% without rejection.

In this paper, we explore a widely automatic cognitive radio waveform recognition system, which can identify 8 kinds of waveforms (BPSK, LFM, Costas codes, Frank code and T1-T4) by using CNN classifier. We employ CWD as a key signal processing method to improve SNR with detected signals. The detected signals will become a 2-D image about time and frequency after CWD. We process the 2-D image into a binary image by using image binarization, image opening operation etc. As is known to all, CNN is an effective classifier especially in image recognition. The excellent structure gives us a high RSR in the low SNR environments. Experimental results show that when SNR ≥ -2dB, the total RSR of recognition system will be more than 93%.

Major contributions of our work can be summarized as follows: (1) Propose the recognition scheme based on CWD and CNN. (2) Explore an adaptive structure of CNN for the scheme. (3) Our recognition system can identify as many as 8 kinds of waveforms. Previous systems can seldom discuss such kinds of waveforms together, especially classification of polytime codes. (4) The system performs well without any priori information.

The paper is organized as follows. The signal model and basic algorithm about CWD are proposed in Section II. Section III describes the structure of the system. Section IV designs a suitable CNN for the cognitive radio waveform recognition system. Section V explores the procedures about how to process the detected signal, how to obtain the binary image and how to remove the image noise etc. Section VI creates the simulation data and discusses the simulation results. Finally, Section VII draws the conclusions.

II. THEORETICAL BACKGROUND

In this section, we review recent approaches related to this paper. Artificial neural network (ANN) has developed rapidly in the last few years. And its development has greatly promoted the automation of the classification system. ANN types vary from one or two layers of single direction logic to many complicated multi-input directional feedback loops and layers [16], [17].

A. SIGNAL MODEL

We assume the amplitude of signal has no connection with time. At the same time, the detected signal transmits in additive white Gaussian noise (AWGN) environment. And the SNR is defined as SNR $= 10\log_{10}(\sigma_s^2)/(\sigma_\varepsilon^2)$, where σ_s^2 and σ_ε^2 are the variances of signal and the noise, respectively. In summary, the signal model is given by

$$x(nT) = s(nT) + m(nT) = Ae^{j\phi(nT)} + m(nT) \tag{1}$$

where, T is sampling interval, n is integer, $x(nT)$ is the complex of detected signal, $m(nT)$ is complex AWGN with the variance σ_ε^2. $s(nT)$ is complex signal, and A is amplitude. For simplicity, we assume $A = 1$. ϕ is the instantaneous phase. However, the Hilbert transform will be applied if the detected signal $x(nT)$ is real. For more details, see [18].

$$y(k) = x(k) + j\mathrm{H}[x(k)] \tag{2}$$

where, $y(k)$ is complex and $x(k)$ is real. H[·] is Hilbert transformation.

B. Choi-Williams DISTRIBUTION (CWD)

Choi-Williams distribution is a kind of time-frequency transformation which is an effective approach to prevent the cross terms [19].

$$C(\omega, t) = \iiint_{\infty} f(\xi, \tau)e^{j2\pi\xi(s-t)} \cdot x(s+\tau/2)x^*(s-\tau/2)e^{-j\omega\tau}d\xi ds d\tau \tag{3}$$

where, $C(\omega, t)$ is the time-frequency result. ω and t are frequency and time axes respectively. $f(\xi, \tau)$ is the kernel function as follows

$$f(\xi, \tau) = \exp\left[\frac{(\pi\xi\tau)^2}{2\sigma}\right]. \tag{4}$$

The kernel is a low-pass filter in two-dimensional space, which removes the cross terms interference. σ is controllable factor, which decides the bandwidth of the filter. Once σ is bigger, the cross terms follows. In this paper, we assume $\sigma = 1$ to balance the cross terms and resolution. In Fig. 1, 8 types of CWD transformation results are presented. We have found the approach to improve the calculation speed of CWD showed in [20]. 1024×1024 points of CWD are chosen in this paper.

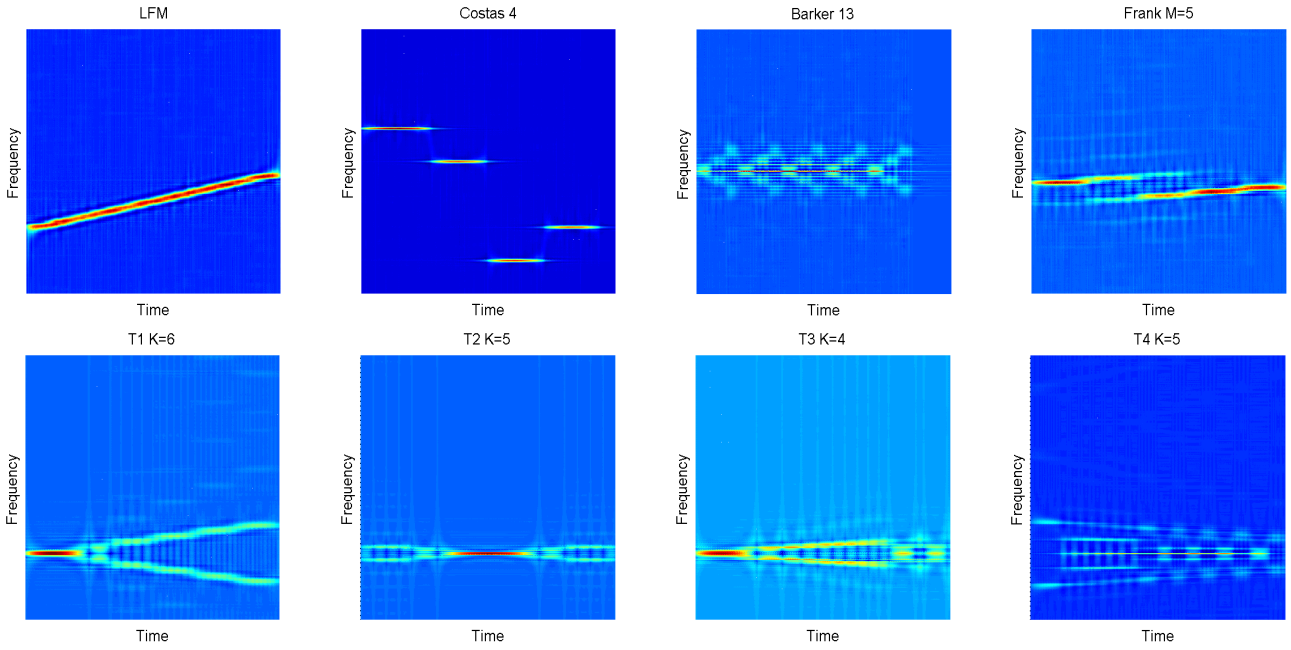


FIGURE 1. In this figure, different waveform classes are shown, including LFM, Costas codes, BPSK, Frank, T1, T2, T3 and T4. There are significant differences among the CWD images.

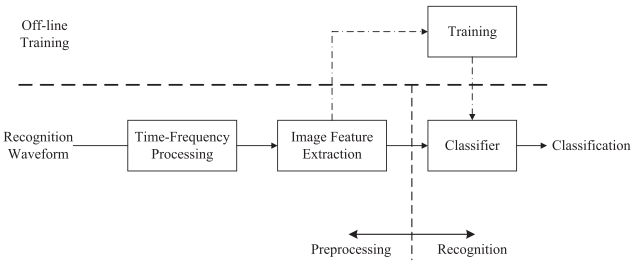


FIGURE 2. The figure shows the system components. The system consists of three parts, time-frequency processing, image feature extraction and classifier. In the first part, the signal is converted into 2-D image. Further, binary images are extracted in the second part for training and testing. The classifier can discriminate between 8 different categories of cognitive radio waveforms into different classes.

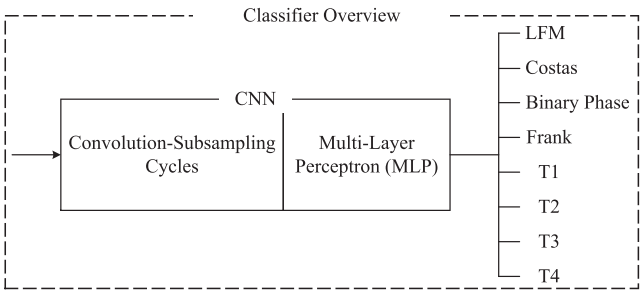


FIGURE 3. There is the structure of classifier. Through the two cycles of convolution and subsampling, the input image would become a features vector for MLP. And then, the features vector is classified into one of 8 types of waveforms directly.

III. SYSTEM OVERVIEW

In this section, we describe the recognition scheme in details. As Fig. 2. shown, the entire identification system mainly consists of three components: time-frequency processing, image feature extraction and classifier. All waveforms go through the time-frequency processing part. In this part, signals will become a 2-D image about time and frequency by using CWD transformation. Different kinds of waveforms usually have different 2-D images. Now the 2-D images are ''colorful'', because the signals in the 2-D images have different values in different time or frequency. However, we pay more attention to what the shape of signal is, rather than what the value is or where the largest value locates. At the second part, 2-D images are processed to binary images by utilizing image binarization and image denoising algorithm. Then, the differences between different kinds of waveforms are more significant, thus CNN can identify them easily. After the first and second part, the system collects a great deal of information about different kinds of waveforms, and all waveforms are classified in the third part.

The classifier can recognize 8 different kinds of cognitive radio waveforms. They are LFM, BPSK, Costas codes, Frank code, and T1-T4. At the output of S2, we determine the appropriate convolution kernel in S2 layer to make the convolution results of the input image into a features vector rather than a features map. N1 and N2 layers constitute a complete MLP. And the features vector acts as the input data for it, then the MLP tells us which waveform has the biggest possibility in current input. As a full connected network, MLP is considered as an output part of CNN, with robustness and short training time. The excellent structure gives us a high RSR in the low SNR environments. For more details, see Fig. 3.

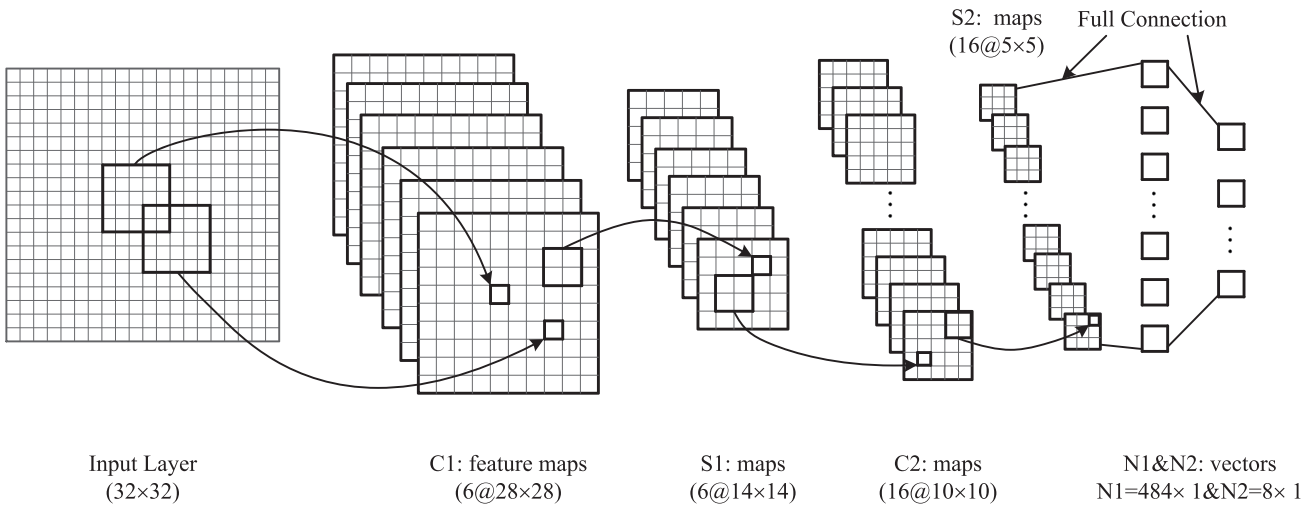


FIGURE 4. It is the structure of CNN. It has 7 layers for different functions. The formula 6@28×28 means, there are 6 feature maps and the size of every map is 28×28.

IV. CNN DESIGN

CNN has special structure for feature extraction [11]. The input of CNN is a 2-D feature (image), and the classification results are expressed in the form of probability. CNN is not a full connected neural network. It is a great innovation, in which the network can focus more on the extraction of features, rather than fitting input data. As a simple example, when we focus on a picture, the picture will be converted into electrical signals by retina, and be transmitted to the visual areas of the cerebral cortex. Further, neurons are also full connected, which allows us to pay more attention to the picture, but ignores the feature that is far from the focus automatically. CNN is more similar to human in visual processing, and it has a powerful performance in complex image recognition. The architecture of the CNN model is shown in Fig. 4. As CNN contains a large number of parameters in the hierarchical architecture, it would be quite lengthy if we use general constants to denote the parameters. Hence, we describe the neural architecture as follows.

a) The input data is an image of the normalized and centralized pattern with size 32×32. The detected signal, after CWD transformation and image binarization processes, will become a binary image. However, the size of binary image is too large to train CNN. We attempt the nearest neighbor interpolation algorithm to match the CNN input size, the advantage of which is to reduce the computer load and survive the key features of image.

b) The first hidden layer C_1 is a convolutional layer with 6 feature maps. With different feature maps, we select different kernels. Each kernel in each feature map is connected to a 5×5 neighborhood of the input. All the neurons in one feature map share the same kernel and connecting weights [21]. We indicate $C_1(i, j, k)$ as the value of kth feature map at position (i, j) in the C_1 layer.

c) The second hidden layer S_1 is a down-sampling layer with 6 feature maps by using mean operation. Such an operation introduces some local translation invariance to the model. According to the same rule of C_1, we denote $S_1(i, j, k)$. Further, we have

$$\begin{aligned} S_1(i,j,k) = \text{mean}(&C_1(2i-1, 2j-1, k), \\ &C_1(2i-1, 2j, k), C_1(2i, 2j-1, k), \\ &C_1(2i, 2j, k)). \end{aligned} \tag{5}$$

The size of S_1 reduces to $1/4$ compared with C_1.

d) Next layer C_2 is also a convolutional layer with 16 different kernels. It is not full connected with S_1 layer. The connection detail is described in Table 1. The reason why not utilize a complete connection is described in [22]. We also use $C_2(i, j, k)$ to show this layer. For the αth column in Table 1, we mark row indices by $\beta_{\alpha,0}, \beta_{\alpha,1}, \cdots, \beta_{\alpha,p-1}$. For example, if $\alpha = 7$, then we get $p = 4$, $\beta_{7,0} = 1$, $\beta_{7,1} = 2$, $\beta_{7,2} = 3$, $\beta_{7,3} = 4$. Further, the size of convolution kernel is $p \times 5 \times 5$, we denote the αth kernel by K_α. And we have

$$\begin{aligned} C_2(i,j,\alpha) = \sum_{r=0}^{p-1} \sum_{i_0=0}^{4} \sum_{j_0=0}^{4} [\, &S_1(i+i_0, j+j_0, \beta_{\alpha,r}) \\ &\times K_\alpha(5-i_0, 5-j_0, p-1-r)\,]. \end{aligned} \tag{6}$$

For instance, for the 0th column, $p = 3$, $\alpha_{00} = 0$, $\alpha_{01} = 1$, $\alpha_{02} = 2$, and we also have

$$\begin{aligned} C_2(i,j,0) = \sum_{r=0}^{p-1} \sum_{i_0=0}^{4} \sum_{j_0=0}^{4} [\, &S_1(i+i_0, j+j_0, \beta_{0,r}) \\ &\times K_0(5-i_0, 5-j_0, p-1-r)\,]. \end{aligned} \tag{7}$$

Finally, the size of receptive field is 10×10, which is suitable for training and recognizing.

e) Similarly, to match with C_2 layer, this layer is a down-sampling layer, called S_2. S_2 has 16 feature maps, and we follow a similar notation as in Equ. 5, we donate

$$S_2(i,j,k) = \text{mean}(C_2(2i-1,2j-1,k), C_2(2i-1,2j,k), C_2(2i,2j-1,k), C_2(2i,2j,k)). \quad (8)$$

f) After that, S_2 and C_3 are full connected. That is to say, each kernel in C_3 will be connected with all feature maps in S_2. And the size of kernel is 5×5, which means, the feature map will become the size of 1×1, after convolution. In this layer, there are 120 kernels and the output is a column vector with the size of 120×1. We define $C_3(\lambda)$ as the λth feature map of C_3, and K_λ as the λth kernel. So, we have

$$C_3(\lambda) = \sum_{r=0}^{15}\sum_{i_0=0}^{4}\sum_{j_0=0}^{4} \times [\, S_2(i_0,j_0,r)\times K_\lambda(5-i_0,5-j_0,15-r)\,]. \quad (9)$$

g) Finally, a fully connected multi-layer perceptron (MLP) is connected with C_3, which has a hidden layer N_1 and an output layer N_2. For N_1, a method to estimate the suitable neurons number of hidden layer H_{num} is introduced in [23]. In this paper, an empirical formula to decide the neurons number is shown as

$$H_{num} = \frac{X \times C + 0.5 \times (X^2 + X) \times C - 1}{X + C} \quad (10)$$

where, X is dimension of features vector, C is the number of classification. In this paper, $X = 120$, $C = 8$. It is just an empirical formula that can not completely determine the best suitable number of hidden layer. We may fine-tune the number to search the value for proposed network. An effective approach is to take the training data as the testing data and put them into well-trained networks. The current correct recognition rate should be beyond 99%, if not, the number needs to be revised. The original number $H_{num} = 462$ is calculated from Equ. 10, through the fine-tune process, the suitable number $H_{num} = 484$ in the paper. The number is determined before the experiments and is not changed. Finally, the network is trained by using the gradient descent method [24]. The size of N_2 is a 8×1 column vector with one-hot code (defined by classes we want to classify).

V. IMAGE PROCESSING DESIGN

In this section, we process time-frequency images as suitable ones for classifier. The quality of images (for classifier) influences RSR and the robustness of system directly. The section is organized as follows. First, we obtain the 2-D image, by applying CWD transformation. Then, 2-D images are processed into binary images by using digital image processing. For every detected signal, there are one 2-D image and one binary image to match it.

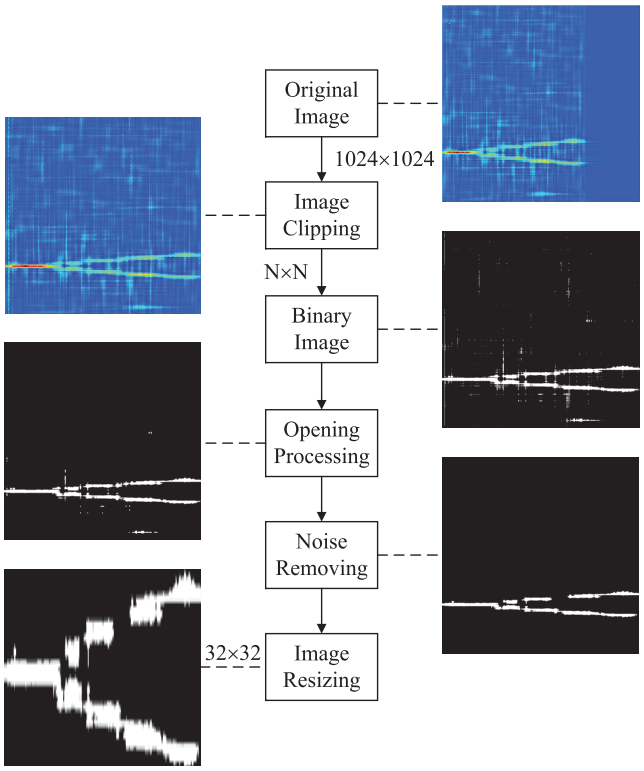


FIGURE 5. In this figure, T1 code acts as an example with SNR of −6dB. Original 2-D image (1024×1024) are clipped by 720×720 (decided by the length of original samples), and the bicubic interpolation is adopted. By utilizing the opening operation and topological approach, the noises are removed effectively.

A. IMAGE CLIPPING

In the following part, the original images are processed into 2-D images. First, we acquire the basic CWD image from the real parts of CWD transformation. In fact, when we calculate CWD by computer, FFT kernel is used for many times. In order to improve the calculation speed, 1024×1024 points is selected. However, the length of signals is $N < 1024$ in general, and zero padding is utilized. Further, the size of CWD image is 1024×1024. Then, in order to reduce the computation load and adapt to the approaches, the CWD image will be resized to $N \times N$ (for more details, see Fig. 5.).

B. BINARY IMAGE

On the basis of global thresholding algorithm, the resized image is changed into a binary image [7], [25]. The algorithm is organized as follows:

a) Normalize the resized image with the range of [0, 1], and record as $H(x, y)$.

b) Estimate the initial threshold T, and the image is divided into two parts H_1 and H_2 by using T

$$T = (\max_{x,y}(H(x,y)) + \min_{x,y}(H(x,y)))/2;$$

TABLE 1. Connection detail about S_1 and C_2.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	✓				✓	✓	✓			✓	✓	✓	✓		✓	✓
1	✓	✓				✓	✓	✓			✓	✓	✓	✓		✓
2	✓	✓	✓				✓	✓	✓			✓		✓	✓	✓
3		✓	✓	✓			✓	✓	✓	✓			✓		✓	✓
4			✓	✓	✓			✓	✓	✓	✓		✓	✓		✓
5				✓	✓	✓			✓	✓	✓	✓		✓	✓	✓

c) H_1 includes the pixels in which the values $> T$, and H_2 includes others;
d) Calculate the average value μ_1 and μ_2 of H_1 and H_2, respectively;
e) Update T as follows

$$T := (\mu_1 + \mu_2)/2;$$

f) Repeat (b)-(e), until the difference between the two calculation results of T is less than 0.001;
g) Binarize the $H(x, y)$ and record as $B(x, y)$,

$$B(x, y) = \begin{cases} 1 & H(x, y) \geq T \\ 0 & others; \end{cases}$$

h) Output $B(x, y)$.

C. NOISE REMOVING

After the image binarization, however, some processing noises and isolated noises can be found in the binary image. Isolated noises are from external environment. They are natural properties of signals. In the binary image, they are many groups of pixels in the random positions. Processing noises are created in the signal processing. For example, the kernel of CWD $f(\xi, \tau)$ influences the processing noises. The shape of the noises is a straight line. Depending on the parameter of CWD's kernel $\sigma = 1$ and the binarization algorithm, the lines are long but thin. The length of most lines is longer than half of the image length and some of them are even as long as the image. But the width is less than 3 pixels. Image morphology algorithm is appropriate to remove both two types of noise. Generally speaking, in 2-D image, signal is the main component. In other words, the number of pixels of signal is the largest, and the amount of each group of isolated noise pixels is much smaller than the signal's. We can employ the topological structure (such as "connected" or "disconnected") of signals and noise to count the number of pixels respectively. And remove the group of pixels which size is much smaller than the largest one. In this paper, the size of groups that is smaller than 10% of the largest group are considered as isolated noises and will be removed. We also utilize opening operation with the kernel of 3×3 to remove the processing noises. The process is introduced in Fig. 5.

D. IMAGE RESIZING

Finally, we will resize the binary image to adapt the input of CNN. The size of 32×32 is suitable for the network. It is small but it has enough pixels to characterize the waveform. The nearest resizing algorithm is explored in this paper. It is the nearest-neighbor interpolation process and the output pixel assigns the value of the pixel that the point falls within. For more details, see [25].

VI. SIMULATION AND DISCUSSION

In this section, we create different kinds of signals as simulation data to test the recognition system. All the generated data and experiments are simulated in the MATLAB 2016a. Each waveform has different parameters that need to be set. U(·) denotes a uniform distribution based on the sample rate. For instance, there is an initial frequency $f_0 = 2000$Hz, and the sample rate is $f_s = 8000$Hz. The uniform result is $f_0 = \mathrm{U}(f_0/f_s) = \mathrm{U}(1/4)$. U(1/16, 1/8) means the parameter is a random number and the range is $[f_s/16, f_s/8] = [500, 1000]$. The parameter setting becomes simple and clear by using this method. For BPSK, the Barker codes length is any of 7, 11 and 13. And the carrier frequency f_c is from U(1/8) to U(1/4) randomly. The cycles per phase code *cpp* and code periods number N_p are range of [1, 5] and [100, 300] respectively. For LFM, the length of signal is range of [500, 1024]. The remaining parameters, initial frequency f_0 and frequency bandwidth Δf, all of which have the same set. Both the ranges are U(1/16, 1/8). For Costas codes, the fundamental frequency f_{min} is a random number from U(1/24) to U(1/20). And the number change N_c is a random integer from 3 to 6. For example, the parameters of Costas codes are $f_{min} = f_s/20$ and $N_c = 4$ respectively. Then, a no-repeating sequence {4, 1, 3, 2} is generated. After that, the frequency sequence $\{4f_{min}, f_{min}, 3f_{min}, 2f_{min}\}$ is created. For Frank code, the carrier frequency f_c and the number of cycles per phase code *cpp* are similar to BPSK. The samples per frequency steps M is also a random integer with the interval [4, 8]. For T1-T4 polytime codes, they are two phase states in the code sequence, and the number of segments k and overall code duration T are range of [4, 6] and [0.07, 0.1] respectively. For more details, see Table 2. There are 1000 labels provided for each waveform class in the same SNR condition. The production

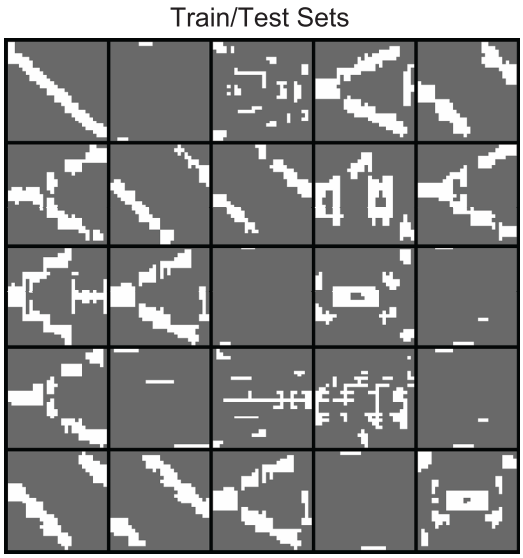


FIGURE 6. In the figure, 25 binary images are selected randomly from the train/test sets of SNR = 0dB. All 8 kinds of waveforms are included in the figure.

TABLE 2. List of simulation parameters.

Radar waveforms	Parameter	Ranges
-	Sampling rate f_s	1
BPSK	Barker codes	$\{7, 11, 13\}$
	Carrier frequency f_c	$\mathrm{U}(1/8, 1/4)$
	Cycles per phase code cpp	$[1, 5]$
	Number of code periods N_p	$[100, 300]$
LFM	Number of samples N	$[500, 1024]$
	Bandwidth Δf	$\mathrm{U}(1/16, 1/8)$
	Initial frequency f_0	$\mathrm{U}(1/16, 1/8)$
Costas codes	Number change N_c	$[3, 6]$
	Fundamental frequency f_{min}	$\mathrm{U}(1/24, 1/20)$
	N	$[512, 1024]$
Frank	f_c	$\mathrm{U}(1/8, 1/4)$
	cpp	$[1, 5]$
	Samples of frequency step M	$[4, 8]$
T1-T4 codes	Number of segments k	$[4, 6]$
	Overall code duration T	$[0.07, 0.1]$

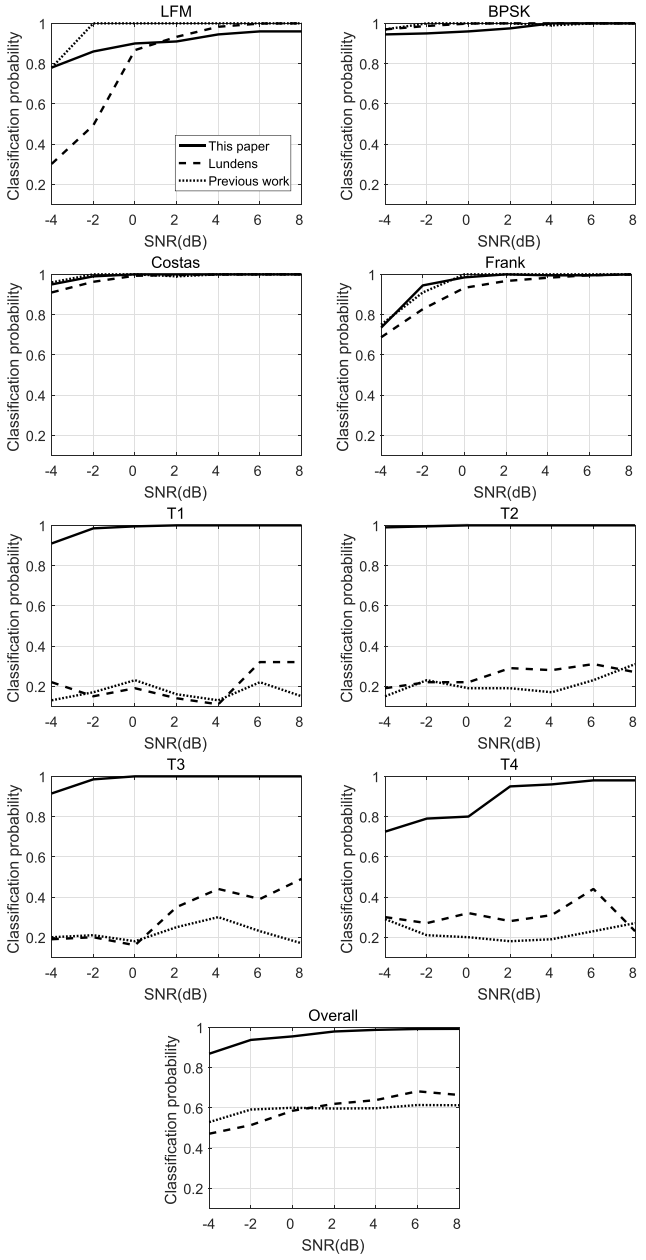


FIGURE 7. Classification performance as a function of the SNR.

of labels repeats for 7 times, in which the SNR is increased from −4dB to 8dB at a step of 2dB. We divide the labels into two parts, 80% labels for training and 20% for testing. We compare our recognition system with J. Lundén's [6] and our previous work [7]. They are the most widely recognition systems at present and they classify the similar waveforms with proposed system. We repeat the two methods and test them by using proposed data. The simulation results are shown as Fig. 7.

Fig. 7 plots RSR as a function of the SNR. Each waveform has its own relation between SNR and RSR and the overall probabilities also be provided. In our system, all waveforms have positive correlation between SNR and RSR. When SNR $<$ 2dB, RSR is increased significantly with the improvement of SNR, and when SNR $\geq$ 2dB, RSR increases more slowly. RSRs of majority waveforms are more than 90%, and some of them even reach 98%. At SNR of −2dB, the overall probabilities is still more than 90%. In considering that the detected signals often has a low SNR, it will be more meaningful that better RSR are achieved in such environment. For LFM, proposed method is worse than previous work but better than Lundén's. Because the time-frequency images are similar to Frank code. In terms of BPSK and Costas codes, the RSRs of three methods have a high level. For Frank code, the results of proposed method and previous work are analogical, and better than Lundén's. For poly-time codes, the performance of proposed method is the best.

TABLE 3. Confusion matrix for the system at SNR of −2*dB* (actual/predicted). the overall RSR is 93.7%.

	LFM	BPSK	Costas	T1	T2	T3	T4	Frank
LFM	100/86	0/0	0/0	0/0	0/0	0/0	0/0	0/5.5
BPSK	0/0	100/95	0/0	0/0	0/0.5	0/0	0/7.5	0/0
Costas	0/0	0/0	100/99	0/0	0/0	0/0	0/4	0/0
T1	0/0	0/0	0/0	100/98.5	0/0	0/0	0/0	0/0
T2	0/0	0/0	0/0	0/0	100/99.5	0/0	0/3.5	0/0
T3	0/0	0/0	0/0	0/0	0/0	100/98.5	0/6	0/0
T4	0/0	0/5	0/1	0/1.5	0/0	0/1.5	100/79	0/0
Frank	0/14	0/0	0/0	0/0	0/0	0/0	0/0	100/94.5

Table 3 exhibits the detail of classification at SNR of −2dB and the overall RSR is more than 93%. The successful recognition about T1-T3 and Costas codes are in a high level. The error of LFM and T4 are higher than others. Some of LFM waveforms are identified as Frank, while, Frank is identified as LFM by the system. In fact, the LFM and Frank code are much similar, which is shown in Fig. 1. T4 has the most complex CWD image in all waveforms. In other words, it has the lowest RSR in all types, and other types of waveforms are also easily identified as T4 at the same time.

VII. CONCLUSION

In this paper, an automatic cognitive radio waveform recognition system based on CWD transformation and CNN classifier is explored. The system can classify 8 kinds of cognitive radio waveforms (including BPSK (Barker modulation), LFM, Costas codes, Frank code, and T1-T4) in a high noisy environment. Through the application of CWD transformation, we can pay more attention to the structure of waveforms. It is a reliable method to classify them. The simulation results show the RSR is more than 93.7% when SNR $\geq$ −2dB. According to the waveforms classification, we can detect, track and locate the radiation sources effectively. It is useful for the wireless communication and radar countermeasure system. However, there are many waveforms in the airspace in each moment. Proposed method is suitable for the single sample classification but not good at multiple samples. How to realize the classification of complex multiple samples is our future work.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the reviewers for their comments and suggestions, which helped to improve the paper.

REFERENCES

[1] W. A. Jerjawi, Y. A. Eldemerdash, and O. A. Dobre, "Blind recognition of SC-FDMA signals using second-order cyclostationarity," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2014, pp. 1163–1166.

[2] E. Karami and O. A. Dobre, "Identification of SM-OFDM and AL-OFDM signals based on their second-order cyclostationarity," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 942–953, Mar. 2015.

[3] L. Cohen, "Time-frequency distributions—A review," *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, Jul. 1989.

[4] G. López-Risueño, J. Grajal, and A. Sanz-Osorio, "Digital channelized receiver based on time-frequency analysis for signal interception," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 879–898, Jul. 2005.

[5] G. López-Risueño, J. Grajal, and O. Yeste-Ojeda, "Atomic decomposition-based radar complex signal interception," *IEE Proc.-Radar, Sonar Navigat.*, vol. 150, no. 4, pp. 323–331, Aug. 2003.

[6] J. Lundén and V. Koivunen, "Automatic radar waveform recognition," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 124–136, Jun. 2007.

[7] Z. Ming, L. Lutao, and D. Ming, "LPI radar waveform recognition based on time-frequency distribution," *Sensors*, vol. 16, no. 10, p. 1682, 2016.

[8] D. Zeng, X. Zeng, H. Cheng, and B. Tang, "Automatic modulation classification of radar signals using the Rihaczek distribution and Hough transform," *IET Radar, Sonar Navigat.*, vol. 6, no. 5, pp. 322–331, Jun. 2012.

[9] J. Ma, G. Huang, W. Zuo, X. Wu, and J. Gao, "Robust radar waveform recognition algorithm based on random projections and sparse classification," *IET Radar, Sonar Navigat.*, vol. 8, no. 4, pp. 290–296, Apr. 2013.

[10] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Jan. 1997.

[11] Y. LeCun, F. J. Huang, and L. Bottou, "Learning methods for generic object recognition with invariance to pose and lighting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Jun. 2004, pp. II-97–II-104.

[12] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, and G. Penn, "Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 4277–4280.

[13] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2014.

[14] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1449–1457.

[15] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN–SVM classifier for recognizing handwritten digits," *Pattern Recognit.*, vol. 45, no. 4, pp. 1318–1325, Apr. 2012.

[16] A. E. W. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. A. Clifton, and G. D. Clifford, "Machine learning and decision support in critical care," *Proc. IEEE*, vol. 104, no. 2, pp. 444–466, Feb. 2016.

[17] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

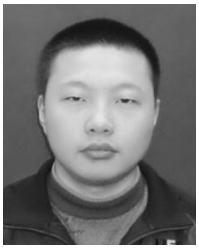
[18] B. Xu, L. Sun, L. Xu, and G. Xu, "Improvement of the Hilbert method via ESPRIT for detecting rotor fault in induction motors at low slip," *IEEE Trans. Energy Convers.*, vol. 28, no. 1, pp. 225–233, Mar. 2013.

[19] Z. Feng, M. Liang, and F. Chu, "Recent advances in timeŰfrequency analysis methods for machinery fault diagnosis: A review with application examples," *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 165–205, Jul. 2013.

[20] P. E. Pace, *Detecting and Classifying Low Probability of Intercept Radar*, 2nd ed. Norwood, MA, USA: Artech House, 2009.

[21] F. Lauer, C. Y. Suen, and G. Bloch, "A trainable feature extractor for handwritten digit recognition," *Pattern Recognit.*, vol. 40, no. 6, pp. 1816–1824, Jun. 2007.

[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[23] K. G. Sheela and S. Deepa, "Review on methods to fix number of hidden neurons in neural networks," *Math. Problems Eng.*, vol. 2013, 2013, Art. no. 425740, doi: 10.1155/2013/425740.
[24] J. Bouvrie, "Notes on convolutional neural networks," pp. 38–44, 2006.
[25] R. C. Gonzalez, *Digital Image Processing*. Upper Saddle River, NJ, USA: Pearson Education, 2009.



MING ZHANG received the bachelor's and master's degrees from the Harbin Engineering University of China, where he is currently pursuing the Ph.D. degree. His research interests include machine learning, image processing, and spectrum sensing for cognitive radio and communication.

MING DIAO received the bachelor's, master's, and Ph.D. degrees from the Harbin Engineering University of China. He is currently a Professor with the College of Information and Telecommunication, Harbin Engineering University. His research interests include wideband signal processing, pattern recognition, machine learning, and telecommunication.

LIMIN GUO received the bachelor's, master's, and Ph.D. degrees from the Harbin Engineering University of China. He is currently an Associate Professor with the College of Information and Telecommunication, Harbin Engineering University. His research interests include statistical, telecommunication, and sensor array signal processing.