# Blind Detection Techniques for Non-Cooperative Communication Signals Based on Deep Learning

**DA KE [ID], ZHITAO HUANG, XIANG WANG, AND XUEQIONG LI [ID]**

State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, National University of Defense Technology, Changsha 410073, China

Corresponding author: Da Ke (1747884404@qq.com)

**ABSTRACT** The performance of existing signal detection methods depends heavily on the amount of prior information acquired by the sensor of interest. Therefore, to improve cognitive radio-based detection in low-signal-to-noise (SNR) environments, we propose a deep learning method-based passive signal detection. A convolution neural network (CNN) and the long short-term memory (LSTM) approach are used to extract the frequency and time domain features of the signal. Our method can detect signal when little to none prior information exists. The simulation experiments verify the probability of detection for our method. The results show that our method is about 4.5–5.5 dB better than a traditional blind detection algorithm under different SNR environments.

**INDEX TERMS** Cognitive radio, deep learning, signal detection.

## I. INTRODUCTION

As 5G communication and Internet of Things (IoT) technologies grow and become more sophisticated, the number of wireless network devices will only increase more dramatically than before. However, available spectrum resources are extremely limited, and many allocated spectrum resources are not fully exploited, which results in unbalanced utilization of RF spectrum [1]. Cognitive Radio (CR) is regarded as a promising technology that can improve the efficient utilization of spectrum resources. Sensors that detect available spectrum resources is a key technology in CR communication, particularly since it can detect available resources in any region without any prior information. A spectrum is vacant if it can be utilized by a secondary user (SU) without interfering with the primary user (PU).

Many spectrum-sensing algorithms have been proposed and studied. Energy detection(ED) [2]–[6] is the most widely used method because of its low complexity and also because no prior information is required [7], [8]. However, this method has a poor detection performance in low signal-to-noise (SNR) environments. A more advanced method, cyclostationary feature detection (CFD) [9]–[11] performs better than the simple ED method. CFD can distinguish between signals and noise by analyzing spectral correlation functions of the signal. However, this requires certain prior knowledge about the PU's signals. Matched filter [12]–[15] techniques are optimal in the sense that they require only a few samples to achieve good performance detection, but the PUs waveforms and channels need to be determined prior to filtering [16]. In non-cooperative communication, a signal is transmitted over an unknown channel, so methods that rely on prior knowledge, are difficult to adapt to CR.

Recently, some researchers have proposed machine learning based methods to improve CR communication. Thomas and Brunskill [17] propose reinforcement learning based algorithms to find better spectrum sensing modalities. Cui *et al.* [18], Huang *et al.* [19], and Ramon *et al.* [20] have designed a support vector machine (SVM)-based learning systems to classify PUs and determine wireless communication parameters. However, developing a method leverages deep learning to determine spectrum availability efficiently and effectively is still a challenging problem.

To improve detection performance in non-cooperative communication, radio signals are artificially generated according to theoretical information, which can be expressed in frequency-domain and time-domain form; however, noise does not have such information. Therefore, a deep learning

---

The associate editor coordinating the review of this manuscript and approving it for publication was Chengpeng Hao.

method is used in this work to extract and integrate theoretical models to distinguish between signal and noise. Convolution, long short-term memory and fully-connected deep neural networks (CLDNN) [21]–[23] are successfully integrated into natural language processing (NLP). This integration is composed of a convolution neural network (CNN), long short-term memory (LSTM) and fully-connected deep neural networks (DNN). CNN is suitable for extracting local features, which are related to frequency-domain information. LSTM is a recurrent neural network, that can efficiently model time-domain features. DNN can map signal features to classifiable intervals. In this paper, we propose a method based on CLDNN networks to model additional features and characteristics of a random signal. To better describe the time-domain correlation of signals, the input data of the network will contain past and future signals, so that the whole system can be regarded as an advanced finite impulse-response-based smoothing function. LSTM has more parameters than general networks; therefore, the method will require more training time. To overcome this additional computational cost, CNNs can also reduce the dimension of input data [23].

The paper is organized as follows. In section II, we describe a signal detection model and CLDNN structure in detail. Simulation results are given in in section III, and we finally discuss our conclusion in section IV.

## II. SYSTEM MODEL

### A. BINARY HYPOTHESIS

In our work, we propose a model based on CLDNN that can be utilized for spectrum sensing. Initially, we assume that there is only one antenna that receives a single PU's signal. We set $x(\mathbf{t})$ as the received signal, which is transmitted into unknown channel. Therefore, the received signal sample is characterized by

$$x[n] = x\left(nT_s\right), \tag{1}$$

where $T_s$ is the sampling period.

Generally, a signal detection problem can be regarded as a binary hypothesis test. Where $H_0$ represents the case when only the noise exists and $H_1$ represents the case when both the signal and noise exist. The receive signal samples, of the two hypotheses are expressed as

$$\begin{cases} H_0 : x[n] = \omega[n] \\ H_1 : x[n] = s[n] + \omega[n] \end{cases}, \tag{2}$$

where $\omega[n]$ is additive white Gaussian noise with a zeromean and a variance of $\sigma^2$. Each $\omega[n]$ sample is assumed to be independent and identically distributed (i.i.d.). The term s[n] represents the PUs signal. There are two probabilities used to evaluate the performance of the method. When detecting the signal at hypothesis $H_1$, it is called the probability of detection $P_d$, and when detecting the signal at hypothesis $H_0$, it is characterized as the probability of false alarm $P_{fa}$. Our purpose is to obtain a high $P_d$ and a low $P_{fa}$.
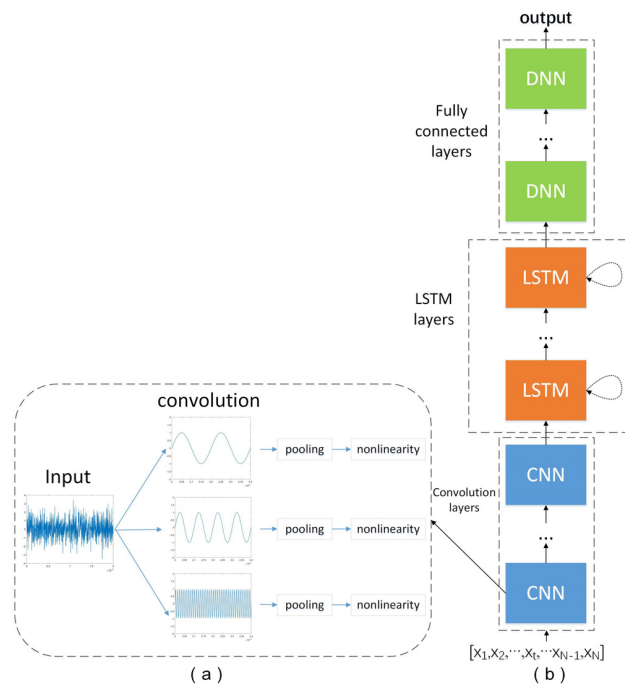


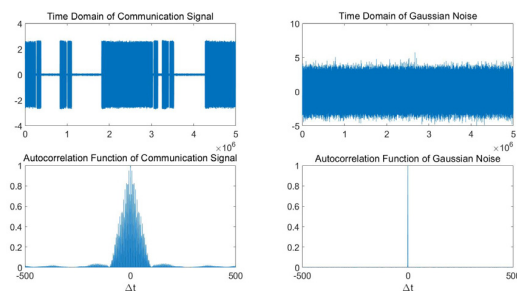**FIGURE 1.** The general structure of CLDNN.



**FIGURE 2.** The correlation function of signal and noise.

### B. PROPOSED CLDNN STRUCTURE FOR SIGNAL DETECTION

The general structure of CLDNN is shown in Fig. 1(b). First, we pass the input data through several convolutional layers. As Fig.1(a) shows, radio signals enjoy different locality characteristics along the frequency axis, which means that the signal has energy concentrations in different local bands along the frequency axis, but the noise fills up the entire frequency band instead of a local band. These local energy concentrations become critical clues to distinguish between signal and noise. After frequency-domain-based modeling is performed, we pass the CNN output to the LSTM layers; this is an appropriate first step to model signal in the time domain. As shown in Fig.2, there is a correlation between $s(t)$ and $s(t + \Delta t)$ in communication signals, but there is no correlation between noise terms. The correlation function is a simple expression and can be easily mathematically formulate it:

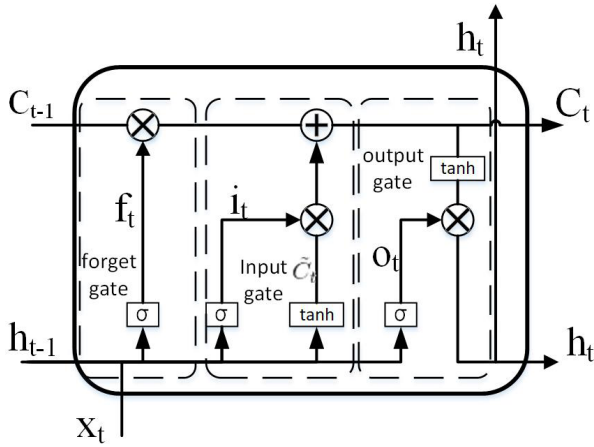$$r_x(\Delta t) = E[x_t^* x_{t+\Delta t}], \tag{3}$$

**FIGURE 3.** Structure of a LSTM cell.

TABLE 1. Configuration of our structure.

| layer | configuration |
|---|---|
| Convolution1 | Filters number:64 length:10 activation:ReLU  kernel regularier:l2 |
| Pooling1 | Pooling size:2 |
| Convolution2 | Filters number:128 length:10 activation:ReLU  kernel regularier:l2 |
| Pooling2 | Pooling size:2 |
| Convolution3 | Filters number:256 length:10 activation:ReLU  kernel regularier:l2 |
| Pooling3 | Pooling size:2 |
| Convolution4 | Filters number:256 length:20 activation:ReLU  kernel regularier:l2 |
| Pooling4 | Pooling size:2 |
| Convolution5 | Filters number:512 length:20 activation:ReLU  kernel regularier:l2 |
| Pooling5 | Pooling size:2 |
| LSTM1 | Length:128 kernel  regularier:l2 |
| LSTM2 | Length:128 kernel  regularier:l2 |
| DNN1 | Neurons number:128 activation:ReLU Kernel  regularier:l1 |
| DNN2 | Neurons number:32 activation:ReLU Kernel  regularier:l1 |
| DNN3 | Neurons number:1 activation:sigmoid |

This correlation function is used in existing signal detection methods. However, there are still some valuable features hidden in the signal that we can not easily mathematically model, but that can help the signal detection process significantly. Therefore, we propose modeling these unknown features via LSTM to improve signal detection performance.

LSTM was initially proposed as an extension of the traditional recurrent neural networks (RNN). Although RNN can, in theoretically, deal with temporally correlated data, it cannot handle long-term time dependency in practice. LSTM addresses this problem by implementing a memory cell, with transmission gates, as shown in Fig. 3. Each gate type is mathematically formulated as follows.

Forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \qquad (4)$$

where $\sigma$ is a sigmoid function, $W_f$ is the weight matrix connecting the input vector $x_t$ and the forget gate. The $\sigma$ function will read the $h_{t-1}$ and $x_t$, and output a value between 0 and 1 to decide how much information of the previous state $C_{t-1}$ will be preserved. Therefore, the input gate is

Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \qquad (5)$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_c), \qquad (6)$$

where $W_i$ and $W_C$ are the weight matrices connecting the input vector $x_t$ and input gate, respectively. The tanh function will output a cell state vector $C_t$, and the cell will update its status based on the $i_t$ and $C_t$ terms as shown in the (7). Therefore, the cell state $C_t$, output gate $o_t$ and $h_t$ function are characterized by

Cell state:

$$C_t = f_t * C_{t-1} + i_t * C_t, \qquad (7)$$

Output gate:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \qquad (8)$$

$$h_t = o_t * tanh(C_t), \qquad (9)$$

Finally, the cell will output the state of the current moment $C_t$ to the next cell, which will then output $h_t$, filtered by $o_t$, to next layer.

After performing frequency and temporal modeling, the output of the LSTM passes through several fully connected DNN layers. These layers can produce a higher-order feature representation that can be more easily separable into the different classes we want to discriminate.

The detailed configuration of the network is listed in Tab. 1. The network is composed of 5 convolution layers and max-pooling layers, 2 LSTM layers, and 3 fully connected layers. The length of the convolution filter is the size of the receptive field. The receptive field of the first two layers has a smaller size and has less filter numbers, which can select small scale features. The next three convolution layers have more filter numbers but a larger receptive field, which can identify more large scale features, because as the data go through more layers, the original signal information is gradually lost (especially with the max pooling layer after each convolution layer), more original signal information could be retained based on a smalled size of receptive field. The receptive fields are increased in the latter convolution layers as all the layers are cascaded, because a lager receptive field means more features can be integrated. After convolution feature modeling, we pass the CNN output to two LSTM layers, where each LSTM layer has 128 cells. Finally, we pass the output of the LSTM to 3 fully connected DNN layers such that the last DNN layer has only one unit and the activation function is a sigmoid function, which will output a value between 0 and 1. The value of the output characterized that a PU signal is near. The extreme cases are ''0'', which represents only noise exists, and ''1'', which represents both signal and noise exists. Given these extreme values, we can decide wether a signal is present or not.
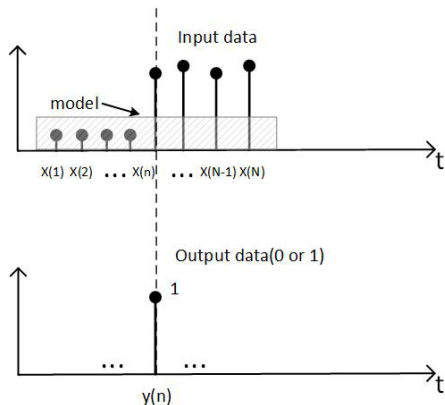
**FIGURE 4.** Input and output data.

## C. INPUT AND OUTPUT

Let $\vec{X}_N = [x(1), x(2) \cdots x(\text{n}) \cdots x(N-1), x(N)]$ be the input data, and W the network weight matrix. Then $y(\text{n}) = W[\vec{X}]$ is the output data. As we have observed, information about future and past help to better predict the current sample [23], the input data is delayed by $N/2 - 1$ samples. As shown in Fig. 4, our model can be thought of as a finite impulse-response smoothing function followed by a nonlinearity [23].

## III. SIMULATION

### A. COMPARISON WITH OTHER ALGORITHM IN DIFFERENT SNR

This section provides the simulation results of our method and its analysis. Since this paper focuses on a blind detection algorithm for non-cooperative communication signals, we will compare our the proposed deep-learning (DL) method to the energy detection (ED) algorithm, which is the most widely used blind detection algorithm. The ED method assumes the signal follows a zero-mean Gaussian distribution with covariance matrix $\sigma_s^2 I$. The noise is also characterized by a zero-mean Gaussian distribution with covariance matrix $\sigma^2 I$.

$$s \sim N(0, \sigma_s^2 I), \tag{10}$$

$$\omega \sim N(0, \sigma^2 I), \tag{11}$$

Both $\sigma_s^2$ and $\sigma^2$ are known. Thus, the ED method computes the energy of the samples as the test statistic $T_{ED}$ and compares the result to a threshold. The hypothesis $H_1$ is true if

$$T_{ED} = \sum_{n=0}^{N-1} x^2[\text{n}] > \gamma_{ED}, \tag{12}$$

The dataset was collected from a real world communication station. The experiment equipment is shown in Fig. 5. It comprises PCs, Anykey AKDS700 radios, a digital receiver, and an oscilloscope. Two radios are linked to two computers respectively to form a transmitter and a receiver. Then, real time wireless communication is performed between the transmitting side and the receiving side.
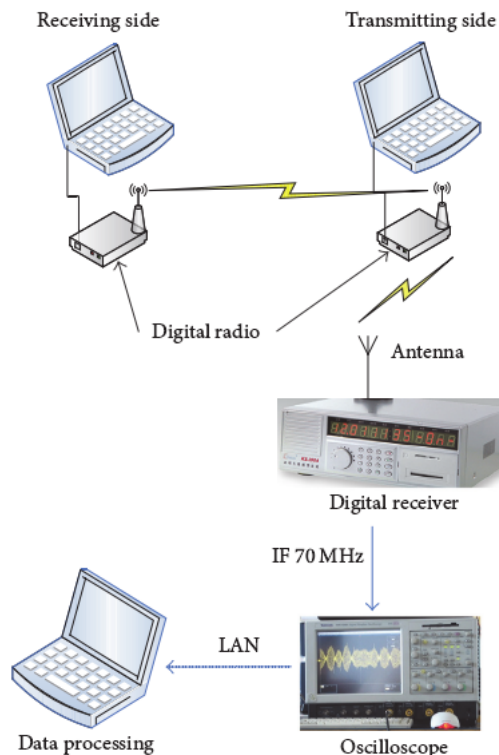


**FIGURE 5.** Schematic diagram of experimental equipment.

Following this, a digital oscilloscope is utilized to collect the radio's RF signals. The frequency of the signal is 763 MHz and the code rate is 5-10 MHz. The modulations of signal include DQPSK, QPSK, BPSK, 16QAM. The sampling rate of the digital receiver is 250 MHz, and the intermediate frequency (IF) of the digital receiver is 70 MHz, and the instantaneous bandwidth is approximate 65-75 MHz. We set the digital receiver is close to the transmitting side, therefore, the SNR of the received signal is high enough to ensure there is no noise in the received signal. We can add white Gaussian noise to obtain various SNR levels to evaluate the algorithms at different SNRs.

The simulation results of the detection performance, in terms of number of inputs N (where $N = 1000$), are shown in Fig. 6. When false alarm probability is 0.1, the detection probability of DL method is the same as that of ED method, and it grows faster than DL method with the increase of SNR. The performance of DL method is up to 5.5dB better than ED method under the same detection probability(90%).

When false alarm probability is 0.05, DL method performs worse than ED method at first, but its performance also grows faster. When SNR is higher than $-12$dB, DL method starts to work better than ED method. Especially, when detection probability is 90%, DL method performs about 4.5dB better than ED method.

Next, several experimental trials were run to analyze the influence of the number of input samples. Fig. 7 shows the relationship between $P_d$ and the length of input sample. We notice that as the input samples increase, the performance
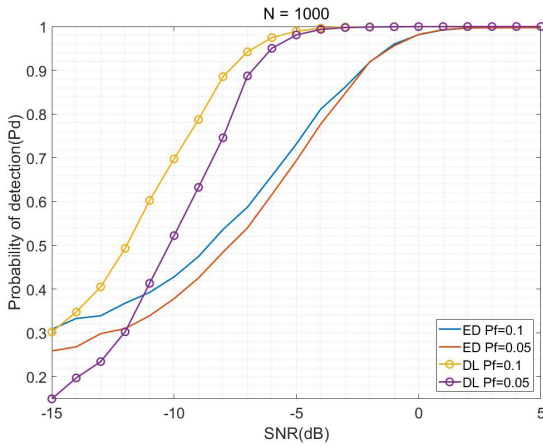
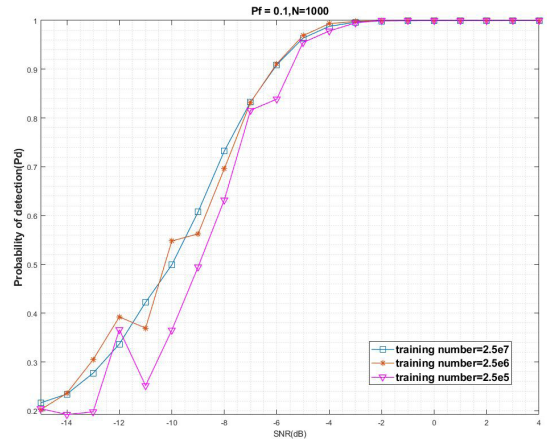**FIGURE 6.** Detection performance of DL method and ED method.



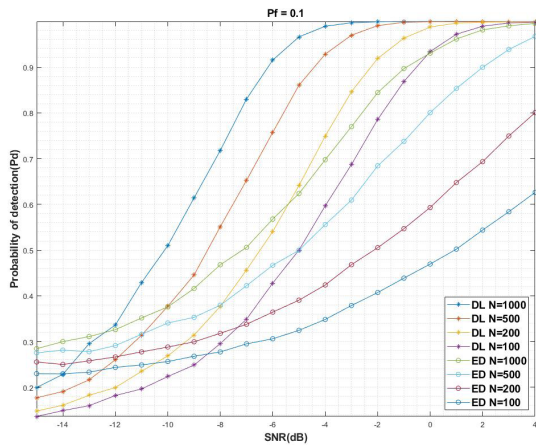**FIGURE 8.** Probability of detection of ED and DL methods under different amount of training data.



**FIGURE 7.** Probability of detection of ED and DL methods under different input size.



**FIGURE 9.** The comparison of CNN + LSTM networks and pure CNN networks.

of both the DL and ED method improve. When the number of input samples decrease from 1000 to 500, the performance of DL deteriorates by approximately 4 dB, while ED method deteriorates by approximately 6dB. Therefore, we surmise that the DL method is not insensitive to input length. The result for the ED method illustrates that the DL method utilizes more knowledge about the signal than the ED method at the same length of input.

The influence of the number of training sample is also tested and the result is shown in Fig. 8. The number of training samples is from $2.5 \times 10^5$ to $2.5 \times 10^7$ samples. We notice that the number of training samples has a small effect on performance. However, if we continue to reduce the number of training samples by $2.5 \times 10^6$, the performance dose deteriorate rapidly and starts to fluctuate. Thus, there is a relationship between sample size and performance that slowly affects performance until a threshold is reached where the performance deteriorates sharply. Thus we determine that the minimum number of training samples for our model is about $2.5 \times 10^7$.

Besides, the performance of CNN + LSTM networks and pure CNN networks was compared in this paper. The result is shown in Fig. 9. There is a similar performance between the
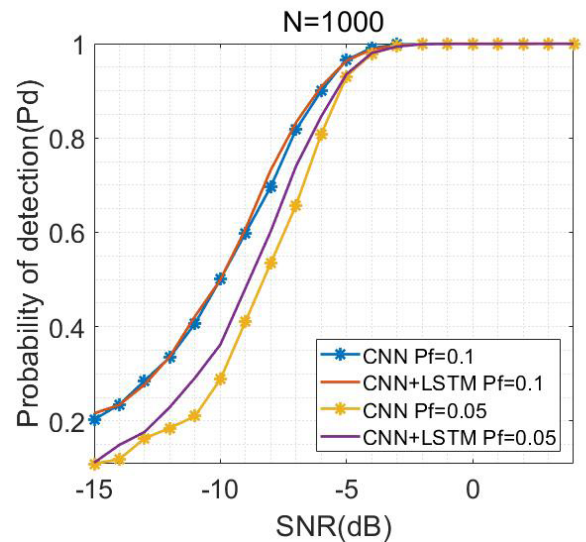
CNN + LSTM networks and CNN networks when $P_f = 0.1$. However, the CNN + LSTM networks perform 0.5dB-1dB better than the CNN networks when $P_f = 0.05$. Our experiment also found that the training time of CNN + LSTM networks is 10 times that of CNN networks, so pure CNN networks is also efficient if a little performance degradation is acceptable. Pure LSTM networks are generally not used, since the training complexity is unacceptable.

We also analyzed the influence of different length of CNN kernels, because we believe that the size of receptive field can impact the performance of the whole networks, especially those of the first few layers. We compared performance of two networks with different hyper parameter settings in terms of the field. The hyper parameter settings are shown in Tab. 2. The only difference between the two is the length of the CNN kernels. The result can be seen in Fig. 10, and it gives the information that networks with a smaller receptive field in the first two layers perform about 1dB better than those with a lager receptive field do when $P_f = 0.05$, but performance

**TABLE 2.** Hyper parameter settings of two kinds of network.

| layer | configuration |
|---|---|
| Convolution1 | Filters number:64 length:10 |
| Convolution2 | Filters number:128 length:10 |
| Convolution3 | Filters number:256 length:10 |
| Convolution4 | Filters number:256 length:20 |
| Convolution5 | Filters number:512 length:20 |
| **(a)** | **Hyper parameter1** |

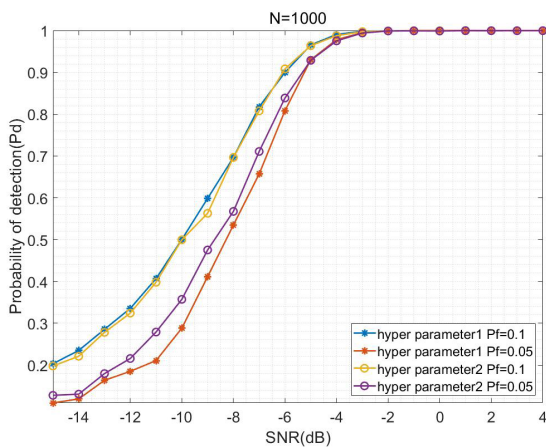| layer | configuration |
|---|---|
| Convolution1 | Filters number:64 length:20 |
| Convolution2 | Filters number:128 length:20 |
| Convolution3 | Filters number:256 length:10 |
| Convolution4 | Filters number:256 length:10 |
| Convolution5 | Filters number:512 length:10 |
| **(b)** | **Hyper parameter2** |



**FIGURE 10.** Performance under different hyper parameters.



**FIGURE 11.** The spectrum of input signal.



**FIGURE 12.** FFT of the First CNN layer's output.

of the both two different hyper parameter settings is similar. The result proves that as the data goes through more layers, the original signal information will be gradually lost and a smaller receptive field in the first few layers can keep more original signal information, especially with the max pooling layer after each convolution layer.

### B. ANALYSIS AND DISCUSSION

We found that CLDNN-based method significantly improves the performance of signal detection in a poor SNR environment. We also analyzed the structure of the CLDNN and found that there are three features of CLDNN that help the signal detection problem.

Locality: the communication signal has locality characteristics along the frequency axis. As a result, filters that work on a local frequency region will provide an efficient way to represent these local structures and their combinations along the entire frequency axis, which may be eventually used to distinguish between signal and noise. CNN is capable of modeling these local frequency structures by allowing each neuron of the convolutional layer to receive input only from features representing a limited bandwidth of the whole spectrum. The spectrum of input signal with
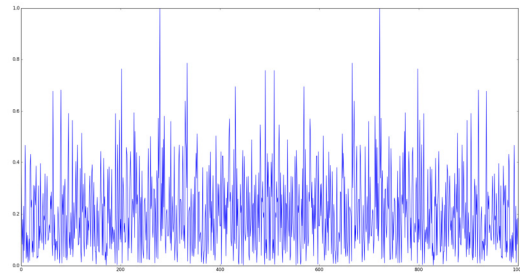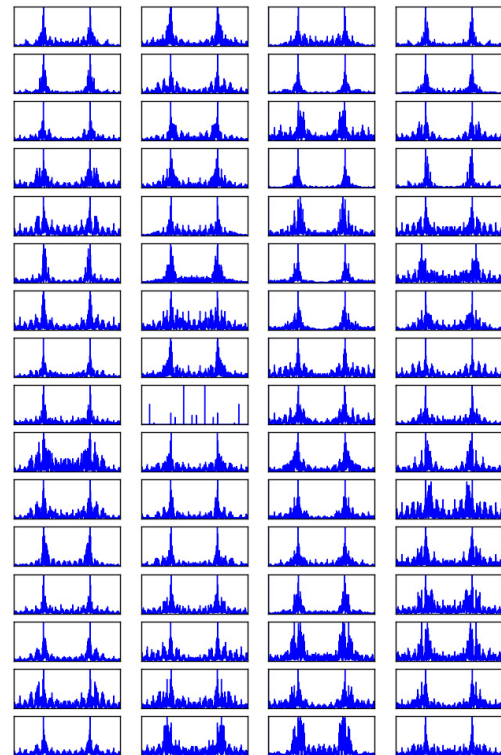
SNR $= -15$dB is shown in Fig. 11, the spectral peak of target signal is not obvious enough. After being fed into CNN layer, the outputs of the first CNN layer are shown in Fig. 12. These two figures illustrate that the output features of CNN are more likely to concentrate in a local area of the whole spectrum.

Max-pooling: communication signals have many local features; these features are distributed on the frequency axis, where every feature centers around a particular frequency that varies in a limited range. Thus, there are a lot of shape changes and displacements due to errors caused by transmitting and intercepting signals. To deal with the problem of variability in CNN, max-pooling layers are introduced into the network structure. As shown in Fig. 13, the input of every max-pooling layer is the output of a previous convolution layer, and the output of three different signals are the same after convolution and max-pooling, which is called translation invariance.
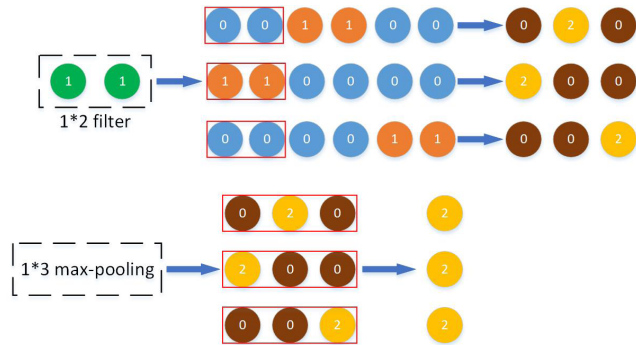
**FIGURE 13.** Translation invariance of max-pooling.

Translation invariance enables convolution layers to extract useful information of signal within a limited error.

Integration of temporal information: after modeling the local features of the original signal via CNN, LSTM then uses its own feedback characteristics to conduct time-domain feature modeling for the output features of CNN. Finally, the information contained in the continuous phase of a communication signal is discovered.

## IV. CONCLUSIONS

In this paper, we proposed a CLDNN based method for signal detection in low-SNR environment. The results show that using CLDNN significantly improves the detection performance. Simulation results show that the DL method utilizes more information of than the ED method, which has a 4.5dB-5dB performance increase under the similar conditions. Thus, DL methods need little-to-no data preprocessing, except for some labels, since features can be automatically extracted by the neural networks. Future work includes research about our method's ability to be generalized. We will also examine the detection performance for cooperative communication.

## REFERENCES

[1] B. Razavi, "Cognitive radio design challenges and techniques," *IEEE J. Solid-State Circuits*, vol. 45, no. 8, pp. 1542–1553, Aug. 2010.

[2] A. Ranjan, Anurag, and B. Singh, "Design and analysis of spectrum sensing in cognitive radio based on energy detection," in *Proc. Int. Conf. Signal Inf. Process.*, Oct. 2016, pp. 1–5.

[3] D. M. M. Plata and Á. G. A. Reátiga, "Evaluation of energy detection for spectrum sensing based on the dynamic selection of detection-threshold," *Procedia Eng.*, vol. 35, no. 35, pp. 135–143, 2012.

[4] K. Srisomboon, A. Prayote, and W. Lee, "Double constraints adaptive energy detection for spectrum sensing in cognitive radio networks," in *Proc. 8th Int. Conf. Mobile Comput. Ubiquitous Netw.*, Jan. 2015, pp. 76–77.

[5] Y. Arjoune, Z. E. Mrabet, H. E. Ghazi, and A. Tamtaoui, "Spectrum sensing: Enhanced energy detection technique based on noise measurement," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2018, pp. 828–834.

[6] D. R. Joshi, D. C. Popescu, and O. A. Dobre, "Adaptive spectrum sensing with noise variance estimation for dynamic cognitive radio systems," in *Proc. 44th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2010, pp. 1–5.

[7] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 4–17, Feb. 2008.

[8] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 21–24, Jan. 2007.

[9] P. S. Yawada and J. W. An, "Cyclostationary detection based on non-cooperative spectrum sensing in cognitive radio network," in *Proc. IEEE Int. Conf. Cyber Technol. Automat., Control, Intell. Syst. (CYBER)*, Jun. 2016, pp. 184–187.

[10] D. Cohen and Y. C. Eldar, "Compressed cyclostationary detection for cognitive radio," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 3509–3513.

[11] S. K. Sharma, T. E. Bógale, S. Chatzinotas, B. L. Le, X. Wang, and B. Ottersten, "Improving robustness of cyclostationary detectors to cyclic frequency mismatch using Slepian basis," in *Proc. IEEE 26th Annu. Int.Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Aug./Sep. 2015, pp. 456–460.

[12] F. Salahdine, H. E. Ghazi, N. Kaabouch, and W. F. Fihri, "Matched filter detection with dynamic threshold for cognitive radio networks," in *Proc. Int. Conf. Wireless Netw. Mobile Commun. (WINCOM)*, Oct. 2016, pp. 1–6.

[13] X. Zhang, R. Chai, and F. Gao, "Matched filter based spectrum sensing and power level detection for cognitive radio network," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2014, pp. 1267–1270.

[14] C. Jiang, Y. Li, W. Bai, Y. Yang, and J. Hu, "Statistical matched filter based robust spectrum sensing in noise uncertainty environment," in *Proc. IEEE 14th Int. Conf. Commun. Technol.*, Nov. 2012, pp. 1209–1213.

[15] Q. Lv and F. Gao, "Matched filter based spectrum sensing and power level recognition with multiple antennas," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process. (ChinaSIP)*, Jul. 2015, pp. 305–309.

[16] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1784–1793, Jun. 2009.

[17] S. P. Thomas and E. Brunskill, "Policy gradient methods for reinforcement learning with function approximation and actiondependent baselines," Jun. 2017, *arXiv:1706.06643*. [Online]. Available: https://arxiv.org/abs/1706.06643

[18] Y. Cui, X. J. Jing, and S. Sun, X. Wang, D. Cheng, and H. Huang, "Deep learning based primary user classification in cognitive radios," in *Proc. 15th Int. Symp. Commun. Inf. Technol. (ISCIT)*, Oct. 2015, pp. 165–168.

[19] Y. Huang, H. Jiang, H. Hu, and Y. Yao, "Design of learning engine based on support vector machine in cognitive radio," in *Proc. Int. Conf. Comput. Intell. Softw. Eng.*, Dec. 2009, pp. 1–4.

[20] M. M. Ramon, T. Atwood, S. Barbin, and C. G. Christodoulou, "Signal classification with an SVM-FFT approach for feature extraction in cognitive radio," in *Proc. SBMO/IEEE MTT-S Int. Microw. Optoelectron. Conf.*, Nov. 2009, pp. 286–289.

[21] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," *Comput. Sci.*, pp. 338–342, Feb. 2014. [Online]. Available: https://arxiv.org/abs/1402.1128

[22] Y. Miao, M. Gowayyed, and F. Metze, "EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand. (ASRU)*, Dec. 2015, pp. 167–174.

[23] T. N. Sainath, R. J. Weiss, A. W. Senior, K. W. Wilson, and O. Vinyals, "Learning the speech front-end with raw waveform CLDNNs," in *Proc. 16th Annu. Conf. Int. Speech Commun. Assoc.*, Sep. 2015, pp. 1–5.

**DA KE** was born in 1994. He received the B.S. degree in electronic science and engineering from the National University of Defense Technology, Changsha, Hunan, China, in 2017, where he is currently pursuing the master's degree. His research interest includes blind signal processing in radar and communication applications.

**ZHITAO HUANG** was born in 1976. He received the B.S. and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha, Hunan, China, in 1998 and 2003, respectively, where he is currently a Professor with the College of Electronic Science and Engineering. His research interests include radar and communication signal processing, and array signal processing.

**XUEQIONG LI** received the B.S. degree from the College of Transportation, Southeast University, Nanjing, China, in 2013, and the M.S. degree from the College of Electronic Science and Engineering, National University of Defense Technology, Changsha, China, in 2015, where she is currently pursuing the Ph.D. degree. Her current research interests include signal processing and machine learning.

● ● ●

**XIANG WANG** was born in 1985. He received the bachelor's and Ph.D. degrees in electronic science and engineering from the National University of Defense Technology, Changsha, Hunan, China, in 2007 and 2013, respectively, where he is currently a Lecturer with the College of Electronic Science and Engineering. His research interest includes blind signal processing in radar and communication applications.