IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 24, 2018, accepted September 18, 2018, date of publication September 28, 2018, date of current version October 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2872769

Big Data Processing Architecture for Radio Signals Empowered by Deep Learning: Concept, Experiment, Applications and Challenges

SHILIAN ZHENG, (Member, IEEE), SHICHUAN CHEN, LIFENG YANG, JIAWEI ZHU, ZHENXING LUO, JUNJIE HU, AND XIAONIU YANG

Science and Technology on Communication Information Security Control Laboratory, Jiaxing 314033, China

Corresponding author: Xiaoniu Yang (yxn2117@126.com)

ABSTRACT In modern society, the demand for radio spectrum resources is increasing. As the information carriers of wireless transmission data, radio signals exhibit the characteristics of big data in terms of volume, variety, value, and velocity. How to uniformly handle these radio signals and obtain value from them is a problem that needs to be studied. In this paper, a big data processing architecture for radio signals is presented and a new approach of end-to-end signal processing based on deep learning is discussed in detail. The radio signal intelligent search engine is used as an example to verify the architecture, and the system components and experimental results are introduced. In addition, the applications of the architecture in cognitive radio, spectrum monitoring, and cyberspace security are introduced. Finally, challenges are discussed, such as unified representation of radio signal features, distortionless compression of wideband sampled data, and deep neural networks for radio signals.

INDEX TERMS Radio signals, big data, deep learning, neural networks, search engine, cognitive radio, cyberspace.

I. INTRODUCTION

The Big Bang created the world we live in, and the Big Data explosion is creating a whole new digital universe. In 2006, individual users were just entering the TB era and about 180 EBs of data were globally generated. In 2011, this number reached 1800 EB. It is expected that this number will increase to 35 by 2020, which is 40 times that of 2009 [1]. From B, KB, MB, GB, TB, to PB, EB, ZB, and YB, the boundaries of data are constantly expanding. With such rapid data growth, we have stepped into the ''big data'' era [1]–[4].

In the field of wireless spectrum, the rapid growth of mobile devices [5] and the emergence of the Internet of things [6], [7] have led to a surge in wireless application data. The demand for wireless spectrum in modern society is increasing, and the rapid growth of various radio applications such as wireless communication, navigation, radar, and broadcasting has led to wider radio frequency (RF) bands being used day by day. As information carriers of these data, radio signals are also characterized by big data in terms of large volume, variety, low value density and high velocity. How to uniformly and efficiently process the radio signals is a complex task because it involves efficiently extracting radio signals from a large number of high-speed sampled data and performing efficient analysis on these signals. Nonetheless, big radio signal data processing is important for lots of applications.

One of the most important applications is cognitive radio. Traditional radio signal processing only needs to be designed for specific applications. For example, the broadcast receiver only needs to receive the frequency modulation (FM) signal at the corresponding frequency band to demodulate the sound/music for the user to listen to; the mobile phone receiver only needs to receive the signal of the specific system on the specific channel and recover the information for the user to use. However, in recent years, new intelligent communication methods such as cognitive radios (CRs) [8]–[10] require radio receivers to have wideband spectrum sensing [11], [12] capabilities. Unlike conventional radio receivers, in addition to processing the radio signals transmitted by its communication counterpart, the cognitive radio also needs to analyze various radio signals in a wide frequency band in order to achieve detection and identification of primary users, and to better understand the spectrum environment for use in spectrum decision making [13]. Therefore, CRs need to have a certain degree of big data processing capabilities in order to better understand the outside environment, optimize the communication system and network resources and improve communication performance.

In addition, wireless systems are rapidly evolving toward next-generation mobile communications (5G) [14] driven by wireless data and diverse applications [15]. In order to accommodate more user data, cell base stations are expected to be deployed densely. The application of massive MIMO and full-duplex technology makes the unintentional interference between radio signals generated by the terminal equipment and the base stations more serious, resulting in a more complex radio environment. Radio signal analysis in this context is also more challenging. For mobile network operators, through real-time monitoring and analysis of radio signals, signal distribution in multiple dimensions such as time, frequency and location can be obtained, which can be used to guide cell optimization and frequency allocation to adapt to dynamic change of mobile data change and optimize mobile network performance.

Based on the above-mentioned requirements for radio signal analysis applications, this paper makes a new interpretation of radio signals from the perspective of big data. We present a big data processing architecture to efficiently process, identify, search and analyze radio signals. In the signal processing method, we take deep learning as the core driving force of big data processing for radio signals. Deep learning [16], [17] has made remarkable progress in recent years, and its application covers almost all industries and research fields [18], including speech recognition [19], computer vision [20] and natural language processing [21]. In the field of radio signal processing, the application of deep learning is still in its infancy, and recent applications include radio signal classification (e.g., modulation classification [22]–[24], RF fingerprinting [25], and radio burst classification [26]) and channel estimation [27]–[29]. In these applications, deep learning has shown excellent performance. In this paper, we use deep learning as an enabling technology for big data processing of radio signals. We discuss the new paradigm of end-to-end processing of radio signals based on deep learning, making signal processing from careful feature and algorithm design to unified deep neural network processing.

The rest of this paper is organized as follows. In Section II we discuss the big data characteristics of radio signals. Then we present the architecture of big data processing for radio signals in Section III, where we also discuss in detail the signal processing paradigm based on deep learning. Experimental system and results are given in Section IV. Applications and challenges are presented in Section V and Section VI, respectively. Finally, we summarize the paper in Section VII.

II. BIG DATA CHARACTERISTICS OF RADIO SIGNALS

There are various definitions of big data. One of the most widely used definitions is the one given by IDC in a report in 2011 [30]: "Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis." This definition leads to the four characteristics of big data, namely volume, variety, velocity, and value [31]. As a result, the "4Vs" terminology has been used widely to characterize big data. Radio signals also exhibit such "4Vs" characteristics of big data.

A. VOLUME

With the growing demand for full-spectrum sensing, the frequency band to be sensed and processed is becoming wider, from tens of hertz to several hundred of megahertz, or to several gigahertz or even terahertz. As a result, the signal sampling rate is increasing and thus the amount of data acquired for radio signal processing.

For example, in order to deal with a bandwidth of 1 GHz, we assume the sampling rate to be 2.5 Gsps. To store each sample in double bytes, the amount of data per second is about 4.66 GB. The amount of data generated in one day is about 392.9 TB. If we want to cover spectral range of 3 kHz - 300 GHz with a total sampling rate of 750 Gsps, then the amount of data collected will increase by hundreds of times. The amount of data collected in one day will be close to 115 PB. Moreover, if the number of sensor nodes increases, the data volume will further increase linearly. Fig. 1 shows the amount of data collected in one day. It can be seen that as the bandwidth and the number of sensing nodes increase, the amount of data collected increases linearly and is quite astonishing.

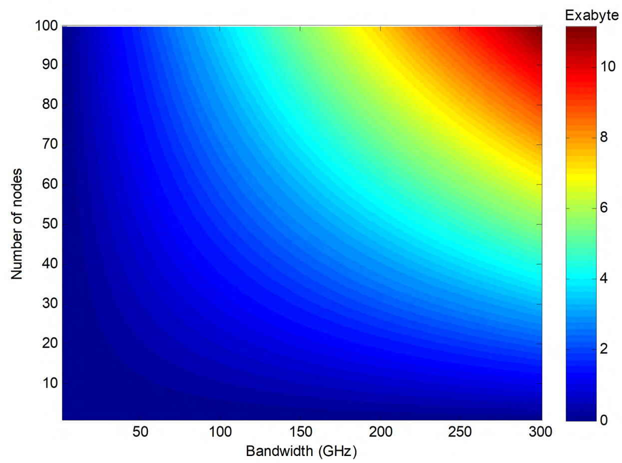


FIGURE 1. The amount of data collected in one day. Sampling rate = 2.5*bandwidth.

B. VARIETY

There are many kinds of radio signals and waveforms. These radio signals can be classified into communication signals, radar signals, navigation signals, etc. according to the services. Taking communications signals for example, according to multiple access method, there are TDMA, CDMA, and OFDMA signal; according to the modulation type, there are AM, FM, PSK, FSK, ASK, QAM, etc. Table 1 shows the categories/ranges of different attributes of radio signals. Such variant signal types increase the difficulty of radio signal analysis.

TABLE 1. Variety of radio signals.

Item	Categories or Ranges
Service	Communications, radar, navigation, broadcasting, etc.
Frequency band	VLF, LF, MF, HF, VHF, UHF, SHF, EHF, etc.
Propagation	Surface wave, ground wave, sky wave, tropospheric scattering, etc.
Multiple access	FDMA, TDMA, CDMA, OFDMA, etc.
Modulation types	AM, FM, BPSK, QPSK, 8 PSK, OQPSK, π/4DQPSK, 2/4/8 FSK, MSK, 2/4/8 ASK, 8/16/32/64/128/256 QAM, 2/4/8 CPM, 16/32 APSK, AM-FSK, FM-FSK, etc.
Frequency usage	Fixed frequency, frequency hopping, direct sequence spread spectrum
Bandwidth	Several kHz (e.g., HF communication) to several GHz (e.g., UWB)
Symbol rate	Several ksps (e.g., HF communication) to several hundred of Msps (e.g., UWB)
Roll-off factor	Raised cosine: 0-1
Digital communications systems in VHF/UHF band	ACARS, ADS-B, AIS, AMPS, APCO-25, ARDIS, ATC-S, ATIS, Bluetooth, CCITT, CDMA2000, CDPD, DAB, D-AMPS, DCSS, DECT, DGPS, DSTAR, DVB-T, EEA, EPIRB, EPLRS, ERMES, EUROSIGNAL, FLEX, FM-S BOS, GSM, Have Quick, INMARSAT, JT2, JT44, JT6M, LINK 4A, LINK 11, LINK 22, MDC-4800, NATEL, ORBCOMM, TACAN, VDEW, etc. [32]

C. VALUE

The purpose of radio signal analysis is to extract useful information from the massive data collected. However, in the massive data collected, in addition to the signals of interest, it also contains a large amount of noise or interference, as well as various signals we are not interested in, so it is necessary to extract useful information in the dense signals, which is of great value but with low density. Fig. 2 shows the real-time spectrum of a certain frequency band. Strong sparsity can be observed in the time-frequency dimension. If the user is concerned with only one type of signal, then the proportion of the data of potential value is further reduced.

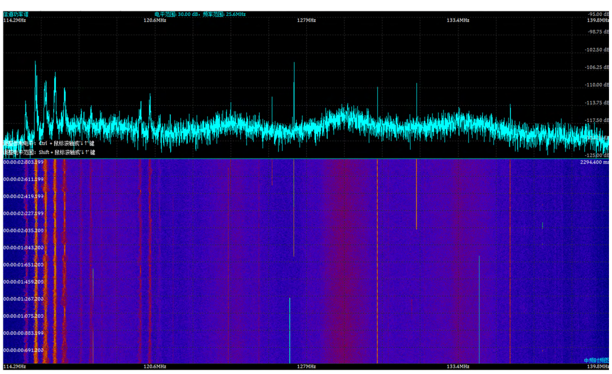


FIGURE 2. Instantaneous spectral graph (the upper graph) and time-frequency waterfall graph (the lower graph) of frequency band 114MHz-140MHz. We can see that there is only noise on a large amount of time and frequency. High energy radio signals are very sparse.

D. VELOCITY

Radio signals are developing towards broadband and high speed. For example, mobile communications signals have evolved from 1G to 4G and 5G. The bandwidth is becoming wider (from 30kHz in 1G to 60GHz in 5G) [14], and the information rate is getting higher (from 2.4kbps in 1G to expected 10-50Gbps in 5G) [14]. The data of radio signals is obtained by sampling the real-time analog input. In order to sense the complete bandwidth, the sampling rate is also increasing, and the subsequent preprocessing and signal analysis tasks must meet the real-time requirements to obtain real-time signal content. Thus, the system is required to have a high processing speed.

III. BIG DATA PROCESSING ARCHITECTURE FOR RADIO SIGNALS

A. BASIC CONCEPT

Big data processing for radio signals not only needs to process the current signal, but also has to manage and make full use of the historical data in order to extract the value of big data. Fig. 3 shows the core conceptual framework of big data processing for radio signals. The framework focuses on the unified representation, management, and rapid search of radio signals. The *signal "crawler"* extracts each signal (the extracted signal is represented in in-phase (I) and quadrature phase (Q) data format, i.e., IQ format) in the original sampled data and acquires the external features of the signal. These external features along with the internal features of the signal learned by deep learning are then stored in the feature database. The user input is designed to search information. There are two ways for user input, one is "text input" similar to the Internet search engine, and the other is "waveform input" based on the real-time spectrogram. Regardless of the input, the user input needs to be converted into a signal feature vector that can be used to search for matches. After the feature vector is obtained, it is matched with the stored signal feature vectors in the signal feature database, and the matching result is returned and displayed to the user in a specific manner.

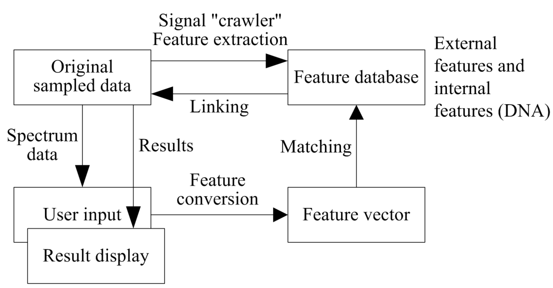


FIGURE 3. Conceptual diagram of big data processing for radio signals.

The signal feature representation is the core of the architecture. The features of radio signals can be referred to two parts: external features and internal features. The external features are obtained from the signal ''crawlers'', and used to characterize the radio signals in four dimensions: time, frequency, space, and energy. For example, the start time and duration time of the signal are to represent time occupation of the signal; the center frequency and bandwidth of the signal are to represent frequency occupation of the signal; the direction and position of emitter are for characterizing the spatial properties of the signal; the signal energy level and signal-to-noise ratio (SNR) are defined as the energy of the signal. Internal features reflect the essential characteristics of the signal that distinguish it from other signals. We call these internal features *signal DNA* in this paper. Signal DNAs are learned by the learning algorithm based on the signal data and are not artificially designed. All historical signals and current signal features are stored in the signal feature database. In addition, the feature database also contains indexes of signals to support rapid acquisition of the corresponding signal IQ data according to the feature matching results. The acquired signal IQ data are processed and displayed to users using visualization technology.

We should note that big data processing for radio signals is not just limited to above signal search function, but the conceptual block diagram shown in Fig. 3 provides the most basic and most essential components that the big data processing tasks for radio signals depend on. Signal correlation, prediction, and other functions can all be performed based on the signal IQ data acquired by the signal ''crawler'' and the signal feature representation. The physical architecture we'll present in the following section focuses on the unified implementation architecture of big data processing for radio signals in all aspects.

B. PHYSICAL ARCHITECTURE

The physical architecture of big data processing for radio signals presented in this paper is shown in Fig. 4, which includes a signal acquisition end, a signal processing end, and a signal application end. The data exchange between each end is implemented through a high-speed data exchange network. The signal acquisition end collects radio signals in a specific frequency band as required, and submits the sampling data to a high-speed data exchange network. The signal processing end is the core of the architecture. It analyzes, stores, and learns the data obtained by the collector. The analysis results of different tasks are stored in the result data storage center according to the user's requirements. The learned models are also stored and can be provided to the application node for inference based on user needs. The signal application end is oriented to the user application. It acquires and displays signal data analysis results to the user. The high-speed data exchange network completes the data exchange between each end. The following sections describe each end in detail.

1) ACQUISITION END

The signal acquisition end consists of a series of acquisition nodes. Each acquisition node acquires radio signal sampling data within a specific bandwidth according to the task requirements. Analog-to-digital converter (ADC) digitizes analog signals of a specific bandwidth to complete the conversion from RF signals to digital sampling streams. Sampling can be achieved by low-pass sampling or band-pass sampling. According to Nyquist-Shannon's sampling theorem, the sampling rate is at least twice the bandwidth. Fig. 5 shows two common analog signal sampling structures. Regardless of the sampling structure employed, the final obtained data are digital sampling streams sampled within a specific bandwidth.

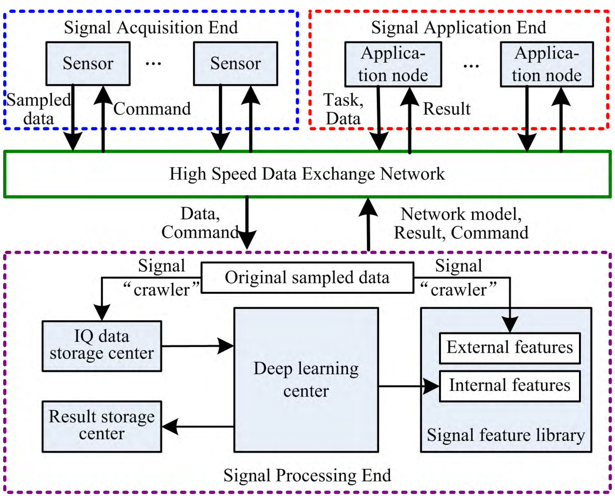


FIGURE 4. Physical architecture of big data processing for radio signals, which include signal acquisition end, signal processing end, signal application end, and high speed data exchange network.

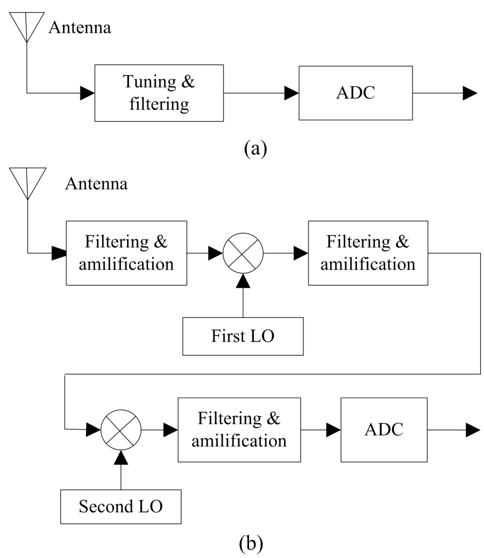


FIGURE 5. Analog signal sampling structure. (a) Direct RF sampling structure. The desired frequency band is selected by a tuning filter, and then sampled by the ADC. (b) Superheterodyne reception sampling structure. Through multi-stage tuning, the frequency band to be sampled is moved to a fixed intermediate frequency, and then sampled by the ADC.

2) PROCESSING END

The signal processing end is the core of the big data architecture. It focuses on the core functions of radio signal "crawler", data storage, learning, and inference.

As pointed out in Section II, radio signals in a wide frequency band exhibit strong sparseness in the time-frequency space. Therefore, it is possible to extract the contents of each signal contained therein and discard other noise components, thereby reducing data transmission stress. One of the tasks of the signal "crawler" is to complete this function, which completes the transformation from the digital sampling sequence to the individual complex envelope radio signals. From the point of view of time, a single radio signal can be categorized into two types: continuous signals and burst signals, in which continuous signals exist for a long time, while burst signals occupy a limited time period. Continuous signals can be regarded as a special form of burst signals with an infinite duration. The IQ representation of a single radio signal contains its complete information content. In addition to obtaining signal IQ data, the signal "crawler" also extracts external features corresponding to the signal: start time, time duration, center frequency, bandwidth, direction, position, power level, and signal-to-noise ratio. The deep learning center uses signal IQ data to learn internal features of the signal. Because the dimension of the external features and internal features of the signal can be fixed, they can be stored in a structured database.

In addition to the feature database, the stored data includes signal IQ data, signal processing result data, and trained deep learning model. The signal IQ data has variant length and can be stored in various files of different sizes. The signal result data includes signal feature parameters related to a specific task (such as modulation type and channel coding method) other than external features corresponding to the signal described above and analysis result data of indefinite length (such as demodulated bit stream and recovered original speech). The application end can directly output the analysis result for the user to use. The learning model data is a learnt model corresponding to each task (for example, neural network model parameters), and the signal application end can acquire these models for inference.

Learning and inference are the core functions of the processing end. Based on the historical data and the current data, a learning model (such as a trained neural network model) is obtained. Then the model is used to analyze and infer the newly emerged signal, and the inference result is obtained. Although many learning methods in the field of machine learning can be used to solve some problems in the field of radio signal processing, in view of the superiority of end-to-end processing, this paper focuses on deep learning to complete big data processing tasks for radio signals, which will be discussed in Section *C*. In order to adapt to the characteristics of radio big data, learning methods need to be able to adapt to massive data, various types of signals, multiple tasks, and incremental learning.

3) APPLICATION END

The signal application end is directly oriented to user applications and includes two application types. The first type of application is instruction-oriented application. After the processing end receives the instruction (and data), it analyzes the data and return the processing results to the user for use. The second type of application is inference-oriented application. This type of application is mainly targeted at users who are sensitive to data privacy and requires the application node to have inference capabilities. The inference-oriented node acquires a trained model (e.g., deep neural network model) from the processing end and then uses the model to analyze and process the data to obtain the desired result. The inference-oriented application node can also have certain transfer learning ability according to needs. It can utilize a widely-trained general learning model acquired from the processing end, performs tailoring and fine tuning for specific tasks, and thus solves the specific task the user needs to solve.

4) HIGH-SPEED DATA EXCHANGE NETWORK

The high-speed data exchange network completes the data exchange between the acquisition end, the processing end, and the application end. The most important task is to transfer the sampling data acquired by the acquisition end to the signal processing end for storage, learning, and analysis. High-speed data exchange networks can use 10GbE, IB or other fiber switching networks.

C. DEEP LEARNING: THE CORE DRIVER OF BIG DATA PROCESSING FOR RADIO SIGNALS

1) DEEP LEARNING BASICS

Machine learning methods can be divided into supervised learning, unsupervised learning and reinforcement learning [33]. Deep learning has achieved good performance in all three aspects. Considering the potential application in radio signal processing problems, this section introduces deep feedforward networks, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and auto-encoders (AEs).

a: DEEP FEEDFORWARD NETWORKS

Deep feedforward network is a basic neural network structure. The basic unit is a neuron, the operation of which is shown in Fig. 6. The mathematical expression is

$$y = \varphi\left(\sum_{j=1}^{m} w_j x_j + b\right), \tag{1}$$

where $x_1, x_2, \ldots, x_m$ are input signal, $w_1, w_2, \ldots, w_m$ are weights, b is the bias, $\varphi(\cdot)$ is the activation function, and y is the output. The activation function increases the nonlinear expression ability of the entire network. The commonly used activation functions are Sigmoid function, hyperbolic tangent function, rectified linear unit (ReLU) [34], and various improvements of ReLU (e.g, Leaky ReLU [35], parameterized ReLU [36], randomized ReLU, and exponential linear unit (ELU) [37]).

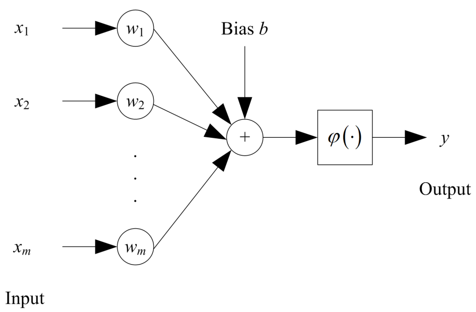


FIGURE 6. Basic model of a neuron.

Deep feedforward networks have many hidden layers. Each hidden layer consists of multiple neurons. the output unit can be divided into two categories: softmax units for classification tasks and linear units for regression tasks. The purpose of learning is to optimize an objective function. For the classification task, the commonly used objective function is the cross-entropy loss function. For the regression task, the objective function commonly used is the mean square error function. The network can be trained with the popular Back-Propagation (BP) algorithm [38] with stochastic gradient descent (SGD) [39].

b: CONVOLUTIONAL NEURAL NETWORKS

CNNs are a special type of neural network originally designed for computer vision applications [40]. Unlike deep feedforward networks, CNN has introduced new layers such as convolution layer and pooling layer.

In the convolution layer, the feature map of the previous layer is convolutionally processed with the current convolution kernel to obtain the convolution output. After the nonlinear activation, the output feature map of the convolution layer is obtained. Denote the input tensor of the convolution layer l as $x^l \in {}^{H^l \times W^l \times D^l}$, the convolution kernel as $f^l \in {}^{H \times W \times D^l}$, if there are D convolution kernels (the number of channels output by the $l+1$ layer D^{l+1} is equal to D), the convolution output is

$$y_{i^{l+1},j^{l+1},d} = \sum_{i=0}^{H}\sum_{j=0}^{W}\sum_{d^l=0}^{D^l} f_{i,j,d^l,d} \times x^l_{i^{l+1}+i,j^{l+1}+j,d^l}, \quad (2)$$

where (i^{l+1}, j^{l+1}) is the index of the convolution result, satisfying

$$0 \leq i^{l+1} < H^l - H + 1 = H^{l+1}, \quad (3)$$
$$0 \leq j^{l+1} < W^l - W + 1 = W^{l+1}. \quad (4)$$

The coefficients of the convolution kernel can be regarded as the learned weights. These weights are the same for all inputs in different positions. This is the "weight sharing" characteristic of the convolution layer.

The pooling layer mainly completes the task of downsampling, and the number of input channels and output channels are equal, but the size of the output feature map is often smaller than the input feature map, depending on the size of the pooling kernel and the step size. The commonly used pooling operations are average pooling and max-pooling, and the average (maximum) value of all values covered by the pooling kernel is used as the pooling result. Because the pooling layer is concerned with the existence of certain features rather than the specific location of the features, it provides a certain degree of translation invariance. In addition, due to the reduced size of output feature map, the computational complexity of subsequent networks is reduced.

In addition to the convolution layer and the pooling layer, CNNs can also include a fully connected layer. The output layer can be either a classification layer or a regression layer, depending on the specific task. The commonly used CNNs include LeNet (1998) [40], AlexNet (2012) [41], VGGNET [42], ResNet [43], DenseNet [44], GoogLeNet (i.e., Inception-v1) [45], Inception-v2 [46], Inception-v3 [47], Inception-v4 [48], Inception-ResNet [48].

c: RECURRENT NEURAL NETWORKS

An RNN is a neural network with memory that is suitable for processing sequence as input [49]. In order to solve the vanishing gradient problem of RNN, several solutions have been proposed, and the long short term memory (LSTM) network [50], [51] is one of them. The key idea of LSTM is the cell state. As shown in Fig. 7, LSTM adds and removes information to the cell state called input gate, forget gate, and output gate [33]:

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right), \quad (5)$$
$$i_t = \sigma\left(W_i[h_{t-1}, x_t] + b_i\right), \quad (6)$$
$$\tilde{C}_t = \tanh\left(W_C[h_{C-1}, x_t] + b_C\right), \quad (7)$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad (8)$$
$$O_t = \sigma\left(W_0[h_{t-1}, x_t] + b_0\right), \quad (9)$$
$$h_t = O_t \odot \tanh(C_t), \quad (10)$$

where $\odot$ denotes element-wise product.

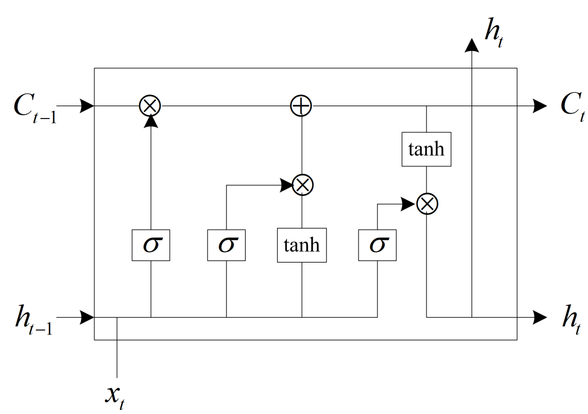


FIGURE 7. Diagram for LSTM.

The gated recurrent unit (GRU) [52] has made some improvements to the LSTM. It merges the forget gate and the input gate into a single "update gate" and merges the cell state with the hidden state. Its main advantages are lower computational complexity and simpler models. In addition to GRU, there are other improvements to LSTM, such as convolutional LSTM [53].

d: AUTO-ENCODER

AE is a deep neural network method for unsupervised feature learning [54], [55]. AE contains two parts: the encoding part and the decoding part, as shown in Fig. 8. The input data of the coding part is mapped to a low-dimensional feature space, and the decoding part reconstructs low-dimensional features back to the original input. The AE training goal is to make the reconstruction error as small as possible, i.e.,

$$\ell\left(x,\hat{x}\right)=\left\|x-\hat{x}\right\|^{2}, \tag{11}$$

where x is input, $\hat{x}$ is the output of the decoder. Deep AEs are obtained by extending the encoder and decoder into multiple hidden layers. It should be noted that the layers in the AE are not necessarily limited to the fully connected layer, but may also be convolutional layers or LSTM layers. In general, the dimension of the features is lower than the dimension of the input signal, so the learned features can be used for signal compression.

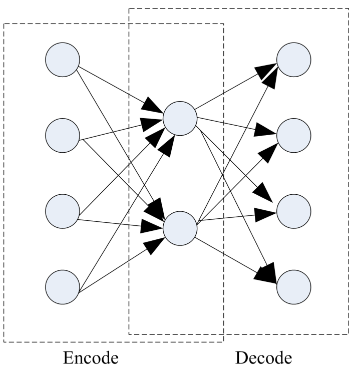


FIGURE 8. Diagram for AE.

2) NEW PROCESSING PARADIGM: FROM SIGNAL TO RESULT

An important idea of deep learning is the end-to-end learning paradigm, which is the most important aspect that distinguishes it from other machine learning algorithms. For the deep model, the input data is in the form of original samples, and the features used to solve specific tasks are automatically learned through training the network. In the traditional radio signal processing field, due to the specificity of each task, the traditional signal processing algorithms are specifically designed. The features of radio signals used to solve tasks such as signal recognition also need to be extracted through feature engineering. In the face of big radio signal data scenarios, these over-designed methods lack flexibilities and are difficult to deal with the demands of variant signal inputs and multi-task outputs that big data analysis faces. This paper uses deep learning approaches for radio signal processing to obtain a new "signal to result" processing paradigm, as shown in Fig. 9.

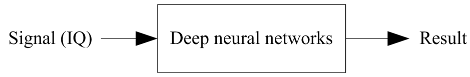


FIGURE 9. Radio signal processing based on deep learning.

Traditional radio signal processing includes tasks such as signal detection, parameter estimation, and information recovery. These tasks all can be accomplished with deep learning methods. In addition, for big data analysis of radio signals, tasks such as signal prediction, signal correlation, and signal representation also need to be completed. Table 2 lists some signal processing tasks based on deep learning. Among them, the network input used for signal prediction can be simplified as the input of signal parameter sequence (such as the time sequence of the signal to predict its future appearance time) without inputting the complete signal IQ data. In addition, unlike other tasks, the signal correlation task needs to analyze whether there is association among multiple signals. Therefore, the input is not a simple single signal but multiple signal bursts. The length of the input signal depends on the specific task and the specific signal. There are two circumstances: fixed length and variable length. The CNNs can be used to process fixed-length input, while the RNNs can handle variable-length input. In Table 2, except that the signal interpretation tasks may output a sequence of varying length, the network output of other tasks is mostly of fixed length.

TABLE 2. Signal processing tasks based on deep learning.

Task	Specific tasks	Input-Output length	Neural networks
Signal recognition	Recognition of modulation, channel coding, protocol, RF emitter, etc.	Fixed/Variant - Fixed	CNN/RNN
Parameter estimation	Estimation of carrier frequency, bandwidth, symbol rate, SNR, location of the emitter, etc.	Fixed/Variant - Fixed	CNN/RNN
Signal interpretation	Demodulation, channel decoding, information recovery	Fixed - Fixed, Variant - Variant	CNN/RNN
Prediction	Prediction of time, frequency, time duration, power, location, etc. of the signal in the future	Variant - Fixed	RNN
Anomaly detection	Detection of jammers, transmissions at higher power than permitted power levels, unauthorized transmissions, etc.	Fixed/Variant - Fixed	CNN/RNN
Signal correlation	Clustering, signal searching	Fixed/Variant - Fixed	CNN/RNN
Signal representation	Signal compression	Fixed-Fixed	AE

Using deep learning techniques will bring a series of advantages to radio signal processing, including:

- Massive data learning. First, deep learning can benefit from large amounts of data [56], [57]. The performance of traditional machine learning methods is no longer significantly enhanced as the amount of data reaches a certain amount. However, the performance of deep neural networks can be further enhanced as the amount of data increases, and massive amounts of data can help avoid over-fitting of deep network models. In terms of computation, the forward and backward propagations can be implemented effectively in parallel [58]. The use of processing units such as GPU, TPU [59] and FPGA [60] greatly improves the efficiency of deep learning training and inference, making it possible to learn large amounts of data.

- High-speed data learning. Radio signal data is generated at high speed in real time, and deep learning can adapt to continuous learning of high-speed data streams. Deep learning mostly uses the SGD for optimization, in which at each iteration one sample or one mini-batch of samples are used to update the model parameters. This training mechanism can be easily adapted to online training of newly generated data [56]. In addition, as the real-time data stream of radio signals is continuously generated, the environment may also change. Therefore, it is necessary to adopt a lifelong learning mechanism [61] to absorb the features of new data without forgetting the features that have been learned before. Deep lifelong learning [62]–[64] provides an effective mechanism for high-speed data stream learning.
- Automatic feature extraction. A major advantage of deep learning is the ability to automatically learn features from the data. Traditional signal processing methods (including methods using traditional machine learning) require careful research of radio signal features. Feature engineering is time consuming and it is not always possible to extract signal features that are most useful to a particular task. In addition, radio signals propagate through the wireless channel, and various factors such as multipath, noise, and interference cause distortion of the received radio signals. The distortions are dynamically changing in time and space. Artificially designed features can not adapt to the impact of various non-ideal factors. Rather, deep learning methods learn features from the received radio signal data itself, which not only simplifies the process of feature extraction, but also improves the robustness.
- Multi-task adaptability. The features learned via deep neural networks have certain adaptability to multiple tasks. Therefore, a single model can be used to solve multiple tasks. For big data processing for radio signals, we often face with multiple tasks. Although a dedicated learning model can be used for each task, this will inevitably lead to too many models and complicate the implementation. Using the same model to deal with multiple tasks is a simpler and more economical implementation. Furthermore, the deep neural network model learned for a certain task can also solve new tasks via transfer learning [65], which helps rapid learning and processing of new tasks.

D. BIG DATA BRINGS NEW FEATURES TO RADIO SIGNAL PROCESSING

Compared to traditional radio signal processing methods, the big data processing architecture brings some new features to signal processing.

- Memorize the history. Since the architecture stores all historical radio signal data, the system has a history memory function. According to the historical data, the processing system can learn the inherent features of the data, and then based on the learned model, the new received data can be inferred and analyzed, so as to be able to deal with various new environments. In addition to this, it is also possible to trace back the history of signals of current interest, and to obtain the evolution of the signals from the history to the current events, and to provide possibilities for future state prediction.
- Predict the future. Based on the learning of historical data, combined with the current data, the future trend of the signal can be predicted. Future state predictions are of great value for many applications. For example, for cognitive radios, prediction of the spectrum occupancy of the primary user may provide support for the decision of spectrum access strategy. Another example is for wireless network security. Through the prediction of interference signals, corresponding anti-jamming measures can be activated in advance to ensure the security of the wireless network to the greatest extent.
- Associate one another. Big data processing can analyze the correlation between signals. Traditional signal processing methods mostly deal with the signals they are concerned with, and very limited attention is paid to whether there is a certain degree of correlation between signals. The ''beer-diapers'' association provides good value for commercial applications and the ''signal A-signal B'' association can provide potential value for radio applications. For example, for interference classification, the interference signal which is highly correlated with the communications signal in time is likely to be tracking interference, and the interference signal highly correlated with the communications waveform is likely to be forward-type interference. For another example, through online correlation analysis of radio signals transmitted by wireless ad hoc networks, real-time network node connectivity and network topology can be obtained, thereby weak links of the network in operation can be analyzed, and corresponding measures to improve its robustness can be taken in advance.

IV. EXPERIMENTAL SYSTEM AND RESULTS

A. EXPERIMENTAL SYSTEM

The experimental system adopts a simplified structure of a single station and is used for demonstration of one of big data application of radio signals, i.e., radio signal search engine. The physical architecture is shown in Fig. 10. The acquisition end contains one sensor node with maximum instantaneous

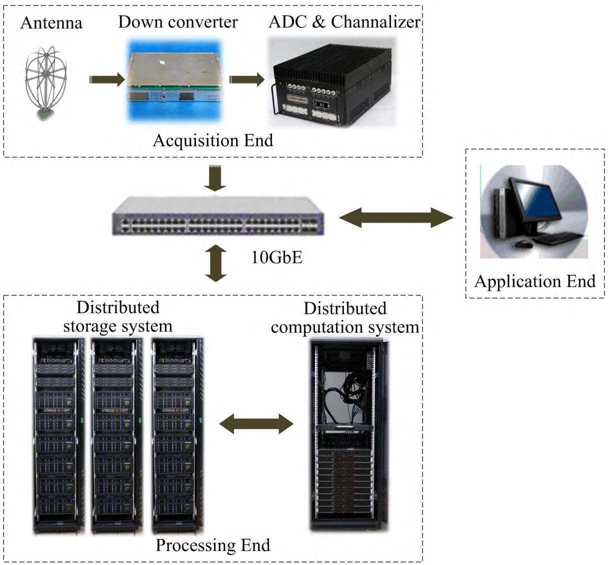


FIGURE 10. Experimental system.

acquisition bandwidth of 300 MHz which works in the very high frequency (VHF) band. The high-speed switching network is 10GbE. The processing end includes a 3PB distributed memory system (data read/write rate of 6.4 GB/s) and a 10-node CPU/GPU distributed computing platform. The application end is a user terminal (computer) for input of user search and display of results.

There are two ways to search for signals, as shown in Fig. 11. The first way is text input: the user inputs the signal parameters to be searched in the text box provided by the interface. The input may be a single parameter (such as "frequency 30MHz-200MHz", "modulation type BPSK"), or may be multi-parameters (e.g., "frequency 30MHz-200MHz, modulation type BPSK", or "frequency about 88MHz, modulation type is not BPSK"). The system interprets the user input, converts the content to be a feature vector, computes correlation with the stored feature vectors and then returns the search results. The matching degree is displayed one by one from the highest to the lowest. The second way is signal input: the user can select the signal he wants to search in the spectrum graph, or load the signal sampling data file to be searched, and the system extracts the signal feature vector and matches it with the stored feature vectors and returns the matching result.

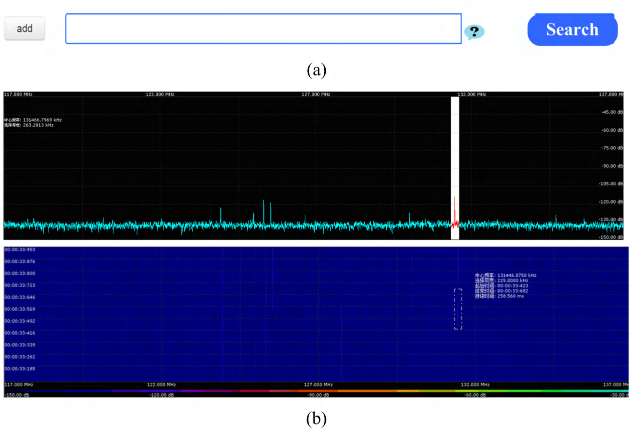


FIGURE 11. Two ways for signal search input. (a) Text input; (b) Signal input. We can select a signal on the instantaneous spectrogram (the upper graph) or on time-frequency waterfall graph (the lower graph).

B. EXPERIMENTAL RESULTS

1) SIGNAL CLASSIFICATION EXAMPLE: CNN-BASED ACARS IDENTIFICATION

Radio signal search involves the recognition of received signals. This section uses Aircraft Communication Addressing and Reporting System (ACARS) [66] emitter identification as an example to introduce a deep learning-based recognition method implemented in the experimental system. ACARS is a digital data link system that transmits short messages (messages) between an aircraft and a ground station via radio or satellite. ACARS has been widely used in the current civil aviation system. In order to verify the powerful end-to-end processing function of deep-learning, we do not perform classical processing operations such as demodulation, decoding, and information recovery, but directly use CNN to obtain the classification result based on the input IQ. The structure of the used CNN is shown in Fig. 12. The network input is a two-dimensional matrix composed of IQ components of the signal.

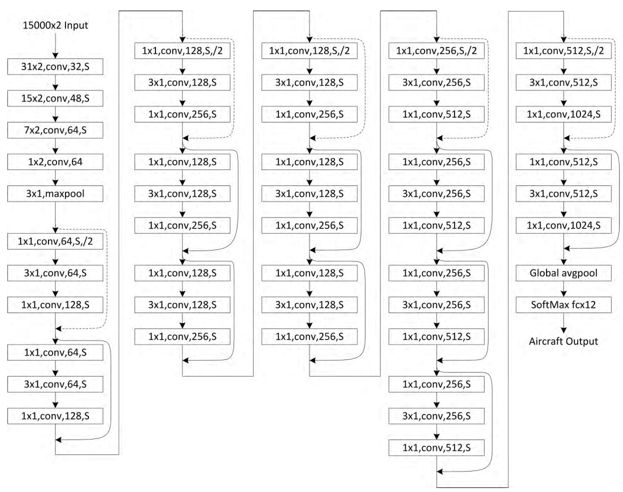


FIGURE 12. Residual CNN network structure. "S" in the figure indicates that the convolution contains padding to make the input and output the same size; "/2" indicates that the downsampling factor is 2, making the output size reduce to half of the input size; "maxpool" means maximum pooling; "Global avgpool" means global average pooling; "conv" stands for the convolutional layer; the number before "conv" indicates the size of the convolution kernel; the number following "conv" indicates the number of convolution kernels; "SoftMax" indicates the SoftMax layer; the output is the category; the class label uses One-Hot encoding; the dotted arrow indicates there is pooling in the process. All activation functions are ReLU. There is also a batch normalization layer between each convolutional layer and activation layer, which is not shown in the figure for the sake of simplicity.

We acquired samples of ACARS signals transmitted over the air as the training samples. We conducted experiments on the identification performance. ACARS signals of 2016 aircrafts were used. Each aircraft had 20 signal samples for training and 10 signal samples for verification. Fig. 13 shows the network training and validation curves. The training method is SGD with momentum. The learning rate is gradually reduced, the mini-batch size is 64, and a validation is performed every 2000 iterations. With the progress of training, the recognition rate is getting higher and higher. After the training stops, the total test accuracy rate is 99.67%, which validates the effectiveness of the method.

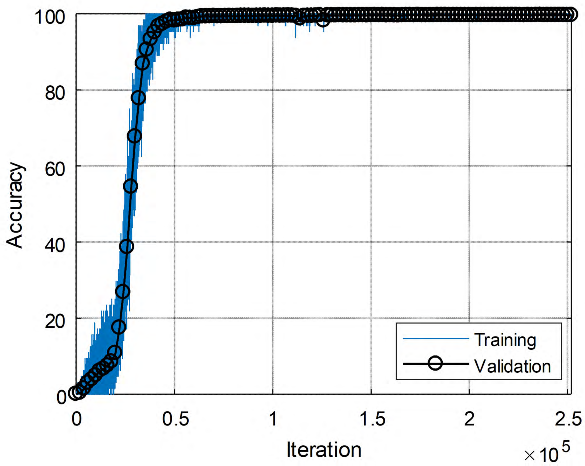


FIGURE 13. **Classification rate of the network.**

2) SIGNAL SEARCH RESULTS

For ease of analysis, the experiment was conducted on a 1.6 MHz bandwidth. The antenna was mounted in a place in Jiaxing, Zhejiang, China. Following gives the experimental results obtained by different search inputs.

(a) Text input "2016-12-30 to 2017-01-02 all signals."

The search returned 1,265,184 signal bursts. The time-frequency occupancy of all these signals is shown in Fig. 14. It can be seen that at some frequencies, the signal is very dense, but overall it shows sparseness.

FIGURE 14. **Signal heat map. The horizontal coordinate is time and the vertical coordinate is frequency.**

(b) Text input "2016-12-30 to 2017-01-02, ACARS signals."

The search returned all ACARS signals from 2016-12-30 to 2017-01-02 with a total of 83,812 records. The time and frequency distributions are shown in Fig. 15. It can be seen that the signals are concentrated in three frequency points and show strong regularity in time. The number of signals in the daytime is large and the number of signals in the evening is small. This is closely related to the aircraft flight planning.

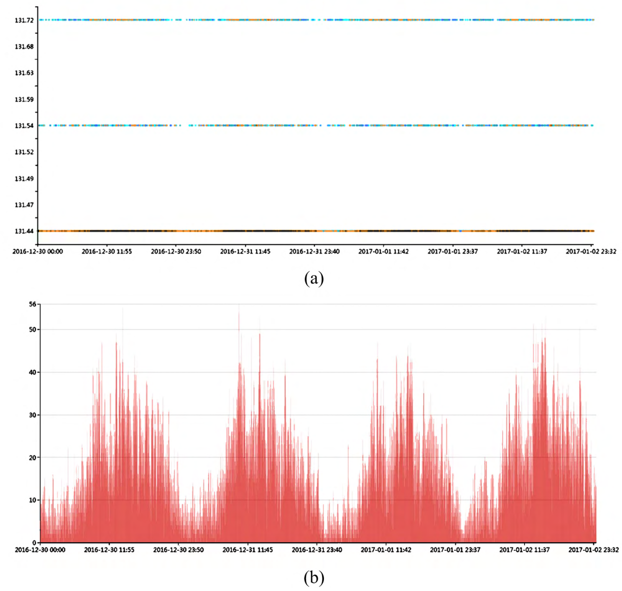


FIGURE 15. **ACARS search result. (a) Time-frequency distribution. The horizontal coordinate is time and the vertical coordinate is frequency. (b) Time distribution of the signal. The horizontal coordinate is time and the vertical coordinate is number of signals.**

(c) Text input "2016-12-30 to 2017-01-02, Aircraft .B-6341."

Signals corresponding to the Aircraft .B-6341 were searched. Note that .B-6341 is the aircraft registration number of this aircraft. Fig. 16 shows the time distribution of these signals. It can be seen that the Aircraft flew over the place where the system antenna was located four times a day. Based on the time interval between any consecutive two flights, we can argue that the Aircraft took off and landed at one of the airports located near the antenna (for instance, Xiaoshan Airport of Hangzhou, or Pudong Airport of Shanghai) with high probability.

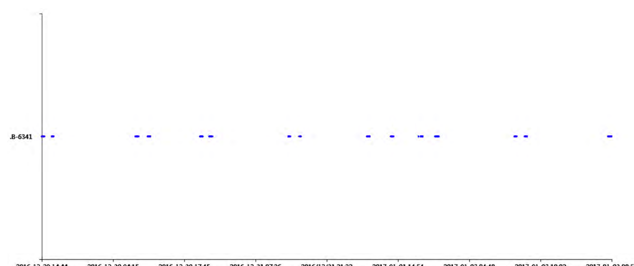


FIGURE 16. **Time distribution of signals from Aircraft .B-6341. The horizontal coordinate is time and the vertical coordinate is aircraft registration number.**

(d) Waveform input.

In the experiment, a signal near 131.45 MHz was selected based on the real-time spectrum graph, as shown in Fig. 11(b). The system first converts the selected signal into a feature vector, and then correlates the feature vector with the feature vectors of other signals in the feature database, and returns results of other signals that match the signal. Fig. 17 shows the matching degree between the returned signals and the selected signal (in time period from 2016-12-30 to 2017-01-02). It can be seen that the returned signals and the selected signal (further analysis shows that it comes from Aircraft .B-7196) mostly belonged to the same aircraft, which illustrates the effectiveness of the matching method.

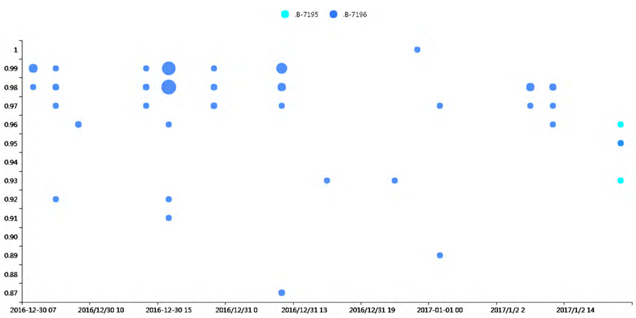


FIGURE 17. Matching degree. The horizontal coordinate is time and the vertical coordinate is the matching degree.

V. APPLICATIONS

The big data processing architecture for radio signals proposed in this paper can be applied in the fields of cognitive radio, spectrum monitoring, cyberspace security and so on.

A. COGNITIVE RADIO

Wireless spectrum is a valuable natural resource. The fixed spectrum allocation mechanism designated by the government regulatory agency leads to an unbalanced utilization of spectrum. At the same time, the overall average utilization rate of the spectrum is also low. In order to improve the spectrum utilization, CR technology has been proposed to dynamically utilize the spectrum [8]. The secondary user (SU) can access the corresponding frequency band only when the primary user (PU) is idle, and once the PU reuses the frequency band, the SU needs to vacant the frequency band and look for other idle channels to continue communication.

Dynamic spectrum access is an autonomous, dynamic, and efficient method of spectrum use, which is of great significance for alleviating spectrum resource shortage and improving spectrum utilization. However, the realization of dynamic spectrum access needs to solve a series of problems, including spectrum sensing, frequency rendezvous, spectrum handoff and so on [9]. The big data processing for radio signals proposed in this paper can be used to sense the primary user signals. It can not only acquire the frequency and time occupancy of the spectrum, but also can obtain the signal parameters (such as center frequency, bandwidth and symbol rate) as well as the signal types and activity patterns of primary user. Based on the spectrum information obtained by radio signal processing, cognitive radio can make spectrum decisions and optimize resources to achieve the purpose of efficient use of wireless spectrum resources.

B. SPECTRUM MONITORING

At present, with the continuous increase in the number of radio emitters used, the continuous increase in spectrum coverage density, and the further deterioration of the electromagnetic environment, spectrum regulation is becoming more and more important. In order to be able to accurately find and enforce interference sources to ensure the communication, spectrum regulation department needs to find and identify interference sources, illegal stations, and malicious base stations through spectrum monitoring, that is, detect various abnormal radio behaviors. There are mainly three types of these abnormal behaviors.

- Users with equipment failure or selfish users who in order to maximize their own communication performance, utilizing the spectrum in a way beyond regulatory constraints, such as transmitting at a power higher than the permitted levels of operation. This abnormal behavior can be detected by estimating the parameters of the received signal (such as estimating the transmit power).
- Illegal users using waveforms different from those used by authorized systems, such as jammers that maliciously disrupt communications. This illegal behavior can be detected by estimating the parameters or identifying the protocols of the received signals.
- Illegal users imitating the authorized user system, such as illegal broadcast stations, fake base stations, and primary user emulation attackers [67], [68] in cognitive radio. For such illegal users, simply identifying the signal parameters/protocols cannot distinguish them from the authorized user signals. By learning and extracting RF or/and wireless channel fingerprinting features, it is helpful to identify such illegal users.

C. CYBERSPACE SECURITY

With the increasing dependence of human social life on cyberspace, cyberspace security [69], [70] has become a highly-regarded concept. Wireless communication networks based on radio signals are an important part of cyberspace and their security is equally important. The electromagnetic spectrum is a completely open space. Therefore, the wireless network cannot form an island of information exchange through physical isolation, and there is a risk of being attacked on the transmission medium itself.

For designing high-security wireless networks, in addition to using encryption and other mechanisms to ensure information security at the information level, it is also necessary to detect, identify, and respond to abnormal attacks at the signal level. Due to the openness of the radio frequency spectrum environment, the jamming environment faced by the wireless network is also a complicated environment composed of various radio signals. The detection and identification of wireless network jamming itself is a problem of big data analysis of radio signals. There are multiple types of jamming, such as single-tone jamming for interfering a single channel, wideband noise jamming for attacking the entire operating frequency band, tracking jamming which tracks and interferes the frequency of frequency hopping system, and fraudulent jamming that attempts to access and deceive a communications network.

By identifying the type of interference, the wireless system/network can take appropriate measures to deal with it. For example, for broadband noise jamming, the communication performance can be enhanced by switching the operating frequency band, or increasing the transmission power. For fraudulent jamming, the spoofing signal is identified and its information is not responded, thereby minimizing the possibility that the spoofing source accesses the network and disturbs the normal operation of the network. For tracking jamming, the difficulty of being tracked can be reduced by reducing the transmission power, increasing the hopping speed, or increasing the hopping bandwidth. For forward-type jamming, since the interference signal is a past communication signal, the interference signal can be utilized to enhance communication performance, such as anti-fragile communication technology.

VI. CHALLENGES

A. UNIFORM REPRESENTATION OF RADIO SIGNAL FEATURES

In order to facilitate search for and correlation analysis of radio signals, it is necessary to uniformly represent radio signals, such as signal feature vectors given in Section III of this paper. The difficulty in the unified representation of radio signals lies in the uniform description of various radio signals with feature vectors of the smallest dimension without loss of information contained in the signals. The representation needs to be unique, universal, robust, concise, and time-sensitive.

- Uniqueness: The representation vector of a signal can uniquely represent this signal, and there will be no case where two different signals have the same representation.
- Universality: The representation can be applied to radio signals in a general sense, not just to some specific types of radio signals.
- Robustness: The noise disturbance of the input signal should not cause a significant change in its representation. The representation should be robust to noise.
- Conciseness: the representation dimension should be as low as possible and the dimensions are required to be fixed. Principal component analysis (PCA), independent component analysis (ICA), or nonnegative matrix factorization (NMF) [71] may be used to reduce the dimensionality of the feature vector.
- Timeliness: the computational complexity of extraction for each feature element should also be as low as possible to meet the real-time requirements of the application.

In this paper, the feature representation of radio signals is divided into two parts: external features and internal features. The features obtained from deep learning are used as internal features of signals (we call it "signal DNA" in this paper). This idea provides a preliminary approach for the unified representation of radio signals. However, at present, we cannot theoretically analyze the performance of this "signal DNA" representation. Therefore, signal DNA is still an open issue.

B. DISTORTIONLESS COMPRESSION OF WIDEBAND SAMPLED DATA

Big data processing for radio signals is faced with a wide frequency band processing environment. With Nyquist-Shannon's sampling theorem, the sampling rate is at least twice the sampling bandwidth. Therefore, the amount of data after wideband sampling is very large. In order to reduce the pressure of high-speed data exchange, it is necessary to study the distortionless compression method for wideband sampled data.

In the previous of this paper, it has been pointed out that radio signals exhibit sparseness in the time-frequency domain. Therefore, there is information redundancy in sampled data acquired under Nyquist sampling framework, which provides a prerequisite for its compression. Compressed sensing is a signal sampling method that uses signal sparsity (sampling and compression are combined into one), which mainly includes sub-Nyquist sampling and sparse recovery. A variety of compression sampling methods for radio signals have been proposed, such as random demodulator sampling (RD) [72], modulated wideband converter sampling (MWC) [73], and multi-rate Nyquist sampling (MRSS) [74]. These methods have carried out a series of theoretical explorations and experimental analyses on sub-Nyquist sampling. Sparse recovery methods such as matching pursuit, convex relaxation, Bayesian methods and even CNNs can be used to reconstruct the original sparse signal [75]–[77]. However, these sub-Nyquist methods all make certain assumptions on the input signal model. In a general sense, compressive sampling methods applicable to arbitrary radio signals are still a problem. In addition, it is also possible to consider the use of AEs [78] to compress radio signals. How to design and train AEs adapted to the actual spectrum environment is also a problem that requires subsequent research.

C. DEEP NEURAL NETWORKS FOR RADIO SIGNALS

At present, the research on deep neural networks mainly focuses on the fields of image, speech, and language processing. The neural network structures adopted also focus on these areas. For example, CNNs are used for image recognition and video analysis, and RNNs are used for speech recognition and machine translation. Although CNNs and

RNNs can be used for certain tasks to handle radio signals and obtain good performance. However, radio signals have their own particularities. The deep neural network structure that can adapt to general radio signal processing is still an open problem to be solved.

Firstly, radio signals are usually represented by the complex baseband envelopes, and the samples are complex numbers. Currently, deep learning-based radio signal classification often extracts IQ components of the signal and the IQ components are used as the input of the neural network. The computations are performed in the real-number domain. However, whether the direct use of a complex neural network [79] to solve radio signal identification problems has advantages in performance still requires future research.

Secondly, in the field of image processing, CNN's hierarchical feature extraction process has a certain physical correspondence with human visual cognition system. However, for radio signal processing, the physical meaning of the activated features will affect the design of network structure. Taking the communication signal as an example, pulse shaping on the physical layer and transmission distortions in the channel transmission process can be described by filtering (convolution). In order to learn these convolution processes, CNNs are regarded as a good candidate. In addition, processes such as differential encoding in the modulation process will lead to correlation of consecutive symbols. In order to learn these correlations, RNNs may be a better choice. The generation process of communication signals contains a large number of convolution processes and also introduces temporal correlation. Therefore, the hybrid structure of CNN and RNN may be more suitable.

Finally, big data processing for radio signals is faced with a multi-task scenario, such as signal parameter estimation, signal type identification, anomaly detection and so on. A simple approach is to design a specialized neural network for each task, which will inevitably lead to too many network models, increasing the storage complexity. Another approach is to design a common network structure to solve various tasks. At present, there have been many advances in deep learning for multi-task learning. In particular, [80] has designed a unified structure to solve multi-domain tasks such as image recognition and machine translation, but for radio signal big data processing, universal deep neural network structure is still a problem to be solved.

D. DEEP OPEN SET LEARNING METHOD FOR RADIO SIGNALS

Radio signal recognition is a main task in big radio data processing. Existing radio signal recognition tasks mainly considered the closed set recognition scenario, but in reality, we often face with the open set scenario, that is, the signal to be identified may not belong to any class in the trained classifier. It is necessary to determine whether the signal belongs to an unknown class. A natural approach is to thresholding on probability/confidence to reject unknown classes because one might hope that for an unknown input all classes would have low probability. However, the existence of "fooling" samples [81], [82] shows that thresholding confidence alone is not sufficient to determine what is unknown. In image recognition, Bendale and Boult [83] and Ge *et al.* [84] introduced OpenMax layer to estimate the probability of an input being from an unknown class and obtained better performance than the existing methods. This idea needs to be tested in the field of radio signal open set recognition. Non-Gaussian feature modeling [85] may need to be combined in the distribution estimation of the activated features to improve the performance. Other open set recognition methods based on deep learning still need to be studied.

VII. CONCLUSION

In this paper, we have presented a big data processing architecture for radio signals and discussed in detail the new signal-to-result (end-to-end) signal processing method based on deep learning. As a verification of the architecture, we used the radio signal intelligent search engine as an example to introduce the system composition and experimental results, which validates the effectiveness of the signal big data processing method based on deep learning. In addition to radio signal search engine, we have introduced the potential application of this architecture in cognitive radio, spectrum monitoring, and cyberspace security. In the future work, challenges such as unified representation of radio signal features, distortionless compression of wideband sampled data, deep neural networks for radio signals, and deep open set recognition of radio signals need to be studied and solved. In addition, signal processing tasks such as demodulation and decoding need to be tested on the experimental system to show the generation and robustness of deep learning-based signal-to-result processing paradigm.

REFERENCES

[1] J. Gantz and D. Reinsel. (2010). *The Digital Universe Decade—Are You Ready*. [Online]. Available: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-201005.pdf

[2] IBM. *What Is Big Data: Bring Big Data to the Enterprise*. Accessed: Mar. 3, 2016. [Online]. Available: http://www-01.ibm.com/software/data/bigdata

[3] J. Gantz and D. Reinsel, *Extracting Value From Chaos*. Hopkinton, MA, USA: EMC, Jun. 2011.

[4] Z. Ma *et al.*, "The role of data analysis in the development of intelligent energy networks," *IEEE Netw.*, vol. 31, no. 5, pp. 88–95, Sep. 2017.

[5] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 190–199, Oct. 2015.

[6] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[7] Q. Wu *et al.*, "Cognitive Internet of Things: A new paradigm beyond connection," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 129–143, Apr. 2014.

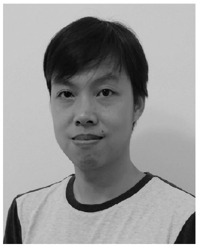
[8] S. Haykin and P. Setoodeh, "Cognitive radio networks: The spectrum supply chain paradigm," *IEEE Trans. Cogn. Commun. Netw.*, vol. 1, no. 1, pp. 3–28, Mar. 2015.

[9] V. Ramani and S. K. Sharma, "Cognitive radios: A survey on spectrum sensing, security and spectrum handoff," *China Commun.*, vol. 14, no. 11, pp. 185–208, Nov. 2017.

[10] F. Zhou, Y. Wu, Y.-C. Liang, Z. Li, Y. Wang, and K.-K. Wong, "State of the art, taxonomy, and open issues on cognitive radio networks with NOMA," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 100–108, Apr. 2018.

[11] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116–130, 1st Quart., 2009.
[12] H. Sun, A. Nallanathan, C.-X. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: A survey," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 74–81, Apr. 2013.
[13] M. T. Masonta, M. Mzyece, and N. Ntlatlapa, "Spectrum decision in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1088–1107, 3rd Quart., 2013.
[14] A. Gupta and E. R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, Jul. 2015.
[15] N. Zhang, P. Yang, J. Ren, D. Chen, Y. Li, and X. Shen, "Synergy of big data and 5G wireless networks: Opportunities, approaches, and challenges," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 12–18, Feb. 2018.
[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[18] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.
[19] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
[20] A. Ioannidou, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, "Deep learning advances in computer vision with 3D data: A survey," *ACM Comput. Surv.*, vol. 50, no. 2, 2017, Art. no. 20.
[21] R. Socher, Y. Bengio, and C. D. Manning. (2012). *Deep Learning for NLP (Without Magic)*. [Online]. Available: https://nlp.stanford.edu/~socherr/SocherBengioManning-DeepLearning-ACL2012-20120707-NoMargin.pdf
[22] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 2016, pp. 213–226.
[23] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.
[24] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Distributed deep learning models for wireless signal classification with low-cost spectrum sensors," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, 2018.
[25] K. Merchant, S. Revay, G. Stantchev, and B. Nousain, "Deep learning for RF device fingerprinting in cognitive communication networks," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 160–167, Feb. 2018.
[26] L. Connor and J. van Leeuwen. (2018). "Applying deep learning to fast radio burst classification." [Online]. Available: https://arxiv.org/abs/1803.03084
[27] N. Samuel, T. Diskin, and A. Wiesel. (2017). "Deep MIMO detection." [Online]. Available: https://arxiv.org/abs/1706.01151
[28] X. Yan, F. Long, J. Wang, N. Fu, W. Qu, and B. Liu, "Signal detection of MIMO-OFDM system based on auto encoder and extreme learning machine," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1602–1606.
[29] C. Zhang, P. Patras, and H. Haddadi. (2018). "Deep learning in mobile and wireless networking: A survey." [Online]. Available: https://arxiv.org/abs/1803.04311
[30] J. Gantz and D. Reinsel, "Extracting value from chaos," IDC iView, 2011, pp. 1–12. [Online]. Available: https://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf
[31] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.
[32] *Technical Handbook for Radio Monitoring VHF/UHF*, Demand GmbH, Nordersedt, Germany, 2017.
[33] M. Z. Alom *et al.* (2018). "The history began from AlexNet: A comprehensive survey on deep learning approaches." [Online]. Available: https://arxiv.org/abs/1803.01164
[34] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[35] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1–6.
[36] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.
[37] W. Chen, J. T. Wilson, S. Tyree, K. Q. Weinberger, and Y. Chen, "Compressing neural networks with the hashing trick," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2285–2294.
[38] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.
[39] L. Bottou, "Stochastic gradient descent tricks," in *Neural Networks, Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 421–436.
[40] 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.
[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.
[42] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[44] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. (2016). "Densely connected convolutional networks." [Online]. Available: https://arxiv.org/abs/1608.06993
[45] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
[46] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[47] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. (2015). "Rethinking the inception architecture for computer vision." [Online]. Available: https://arxiv.org/abs/1512.00567
[48] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. (2016). "Inception-v4, Inception-ResNet and the impact of residual connections on learning." [Online]. Available: https://arxiv.org/abs/1602.07261
[49] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[51] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *J. Mach. Learn. Res.*, vol. 3, pp. 115–143, Aug. 2002.
[52] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. (2014). "Empirical evaluation of gated recurrent neural networks on sequence modeling." [Online]. Available: https://arxiv.org/abs/1412.3555
[53] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process.*, 2015, pp. 802–810.
[54] G. E. Hinton and R. S. Zemel, "Autoencoders, minimum description length and helmholtz free energy," in *Proc. NIPS*, 1993, pp. 3–10.
[55] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[56] X.-W. Chen and X. Lin "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014.
[57] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, p. 1, Feb. 2015.
[58] J. Dean *et al.*, "Large scale distributed deep networks," in *Proc. Adv. NIPS*, 2012, pp. 1223–1231.
[59] N. P. Jouppi *et al.*, "In-datacenter performance analysis of a tensor processing unit," in *Proc. 44th Annu. Int. Symp. Comput. Archit.*, 2017, pp. 1–12.
[60] K. Abdelouahab, M. Pelcat, J. Serot, and F. Berry. (2018). "Accelerating CNN inference on FPGAs: A survey." [Online]. Available: https://arxiv.org/abs/1806.01683
[61] Z. Chen and B. Liu, *Lifelong Machine Learning* (Synthesis Lectures on Artificial Intelligence and Machine Learning). San Rafael, CA, USA: Morgan & Claypool, 2016.

[62] S.-W. Lee, C.-Y. Lee, D.-H. Kwak, J. Kim, J. Kim, and B.-T. Zhang, "Dual-memory deep learning architectures for lifelong learning of everyday human behaviors," in *Proc. IJCAI*, 2016, pp. 1669–1675.
[63] G. I. Parisi, J. Tani, C. Weber, and S. Wermter, "Lifelong learning of human actions with deep neural network self-organization," *Neural Netw.*, vol. 96, pp. 137–149, Dec. 2017.
[64] C. Tessler, S. Givony, T. Zahavy, D. J. Mankowitz, and S. Mannor, "A deep hierarchical approach to lifelong learning in minecraft," in *Proc. AAAI*, 2017, pp. 1553–1561.
[65] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[66] A. Roy, "Secure aircraft communications addressing and reporting system (ACARS)," in *Proc. 20th Digit. Avionics Syst. Conf. (DASC)*, Oct. 2001, pp. 7A2/1–7A2/11.
[67] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxylakis, "A survey on security threats and detection techniques in cognitive radio networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 428–445, 1st Quart., 2013.
[68] J. Soto, S. Queiroz, and M. Nogueira, "Taxonomy, flexibility, and open issues on PUE attack defenses in cognitive radio networks," *IEEE Wireless Commun.*, vol. 20, no. 6, pp. 59–65, Dec. 2013.
[69] S. Liu and B. Cheng, "Cyberattacks: Why, what, who, and how," *IT Prof.*, vol. 11, no. 3, pp. 14–21, May 2009.
[70] L.-X. Yang, P. Li, X. Yang, and Y. Y. Tang, "Security evaluation of the cyber networks under advanced persistent threats," *IEEE Access*, vol. 5, pp. 20111–20123, 2017.
[71] Z. Ma, A. E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian matrix factorization for bounded support data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 4, pp. 876–889, Apr. 2015.
[72] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.
[73] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.
[74] H. Sun, W.-Y. Chiu, J. Jiang, A. Nallanathan, and H. V. Poor, "Wideband spectrum sensing with sub-Nyquist sampling in cognitive radios," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 6068–6073, Nov. 2012.
[75] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, 3rd Quart., 2017.
[76] L. Stanković, E. Sejdić, S. Stanković, M. Daković, and I. Orović, "A tutorial on sparse signal reconstruction and its applications in signal processing," *Circuits, Syst., Signal Process.*, pp. 1–58, 2018, doi: 10.1007/s00034-018-0909-2.
[77] S. Lohit, K. Kulkarni, R. Kerviche, P. Turaga, and A. Ashok, "Convolutional neural networks for noniterative reconstruction of compressively sensed images," *IEEE Trans. Comput. Imag.*, vol. 4, no. 3, pp. 326–340, Sep. 2018.
[78] S. Wu *et al.* (2018). "The sparse recovery autoencoder." [Online]. Available: https://arxiv.org/abs/1806.10175
[79] A. Hirose, *Complex-Valued Neural Networks*. Heidelberg, Germany: Springer, 2006.
[80] L. Kaiser *et al.* (2017). "One model to learn them all." [Online]. Available: https://arxiv.org/abs/1706.05137
[81] I. Goodfellow *et al.*, "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[82] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 427–436.
[83] A. Bendale and T. Boult. (2015). "Towards open set deep networks." [Online]. Available: https://arxiv.org/abs/1511.06233
[84] Z. Ge, S. Demyanov, Z. Chen, and R. Garnavi. (2017). "Generative OpenMax for multi-class open set classification." [Online]. Available: https://arxiv.org/abs/1707.07418
[85] Z. Ma, Y. Lai, W. B. Kleijn, Y.-Z. Song, L. Wang, and J. Guo, "Variational Bayesian learning for Dirichlet process mixture of inverted Dirichlet distributions in non-Gaussian image feature modeling," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2018.2844399.

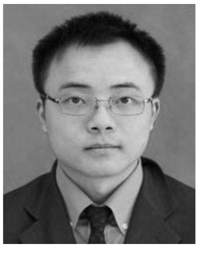


SHILIAN ZHENG (M'11) was born in Wenzhou, Zhejiang, China, in 1984. He received the B.S. degree in telecommunication engineering and the M.S. degree in signal and information processing from Hangzhou Dianzi University, Hangzhou, China, in 2005 and 2008, respectively, and the Ph.D. degree in communication and information system from Xidian University, Xi'an, China, in 2014.

He is currently an Associate Researcher with the Science and Technology on Communication Information Security Control Laboratory, Jiaxing, China. He holds more than 10 patents. His research interests include cognitive radio, spectrum management, compressed sensing, and deep learning-based radio signal processing.

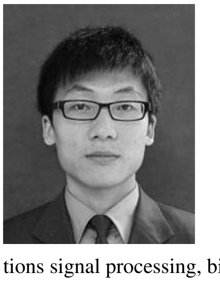
SHICHUAN CHEN was born in Jinhua, Zhejiang, China, in 1976. He received the B.S. degree in electronic circuits and systems from the University of Electronic Science and Technology of China, Chengdu, China, in 1998.

He is currently an Associate Researcher with the Science and Technology on Communication Information Security Control Laboratory, Jiaxing, China. He holds seven patents. His research interests include radio signal processing, analysis, recognition, big data for radio signals, and deep learning algorithms for radio signals.



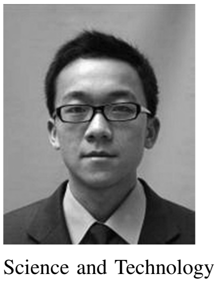
LIFENG YANG received the B.S. and M.S. degrees in computer science and technology from Tianjin University, Tianjin, China, in 2013 and 2016, respectively.

He is currently a Research Assistant with the Science and Technology on Communication Information Security Control Laboratory, Jiaxing, China. His research interests include radio signal identification with deep neural networks.



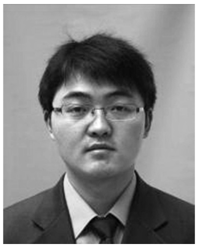
JIAWEI ZHU received the B.S. degree in electronics engineering and the M.S. degree in information and communications engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2009 and 2012, respectively.

He is currently a Senior Engineer with the Science and Technology on Communication Information Security Control Laboratory, Jiaxing, China. His current research interests include communications signal processing, big data analytics, and transmitter localization.



ZHENXING LUO was born in Shaoxing, Zhejiang, China, in 1983. He received the B.S. degree in telecommunication engineering and the M.S. degree in signal and information processing from Hangzhou Dianzi University, Hangzhou, China, in 2006 and 2009, respectively. He is currently pursuing the Ph.D. degree in communication and information system with Xidian University, Xi'an, China.

He is currently an Associate Researcher with the Science and Technology on Communication Information Security Control Laboratory, Jiaxing, China. He holds two patents. His research interests include the development and fundamental study of signal processing and machine learning techniques used in spectrum management.



JUNJIE HU received the B.S. degree in telecommunication engineering and the M.S. degree in communications and information system from Xidian University, Xi'an, China, in 2006 and 2009, respectively.

He is currently a Senior Engineer with the Science and Technology on Communication Information Security Control Laboratory, Jiaxing, China. His current research interests include field-programmable gate array-based design, software-defined radio, and machine learning.

XIAONIU YANG is currently a Chief Scientist with the Science and Technology on Communication Information Security Control Laboratory, Jiaxing, China. He is a fellow of the Chinese Academy of Engineering, a fellow of the Chinese Institute of Electronics, an Adjunct Professor of Zhejiang University, Hangzhou, China, and a Ph.D. Supervisor of Xidian University, Xi'an, China. He has published the first software radio book in China along with C. Lou and J. Xu, *Software Radio Principles and Applications* [Publishing House of Electronics Industry, 2001 (in Chinese)]. His current research interests are software-defined satellite, big data for radio signals, and deep learning-based signal processing.