Radio Frequency Fingerprint Collaborative Intelligent Blind Identification for Green Radios

Mingqian Liu, *Member, IEEE*, Chunheng Liu, Yunfei Chen, *Senior Member, IEEE*, Zhiwen Yan and Nan Zhao, *Senior Member, IEEE*

***Abstract*—Radio frequency fingerprint identification (RFFI) technology identifies the emitter by extracting one or more unintentional features of the signal from the emitter. To solve the problem that the traditional deep learning network is not highly adaptable for the contour features extracted from the signal, this paper proposes a novel RFFI method based on a deformable convolutional network. This network makes the convolution operation more biased towards the useful information content in the feature map with higher energy, and ignores part of the background noise information. Moreover, a distributed federated learning system is used to solve the problem of insufficient number of local training samples for a multi-party joint training model without exchanging the original data of the samples. The federated learning center receives the network parameters uploaded by all local models for aggregation, and feeds the aggregated parameters back to each local model for a global update. The proposed blind identification method requires less information and no training sequences and pilots. Thus, it achieves energy-efficiency and spectrum-efficiency. Simulation verifies that the proposed method can achieve better recognition performance and is beneficial for green radios.**

***Index Terms*—Blind identification, deformable convolutional network, federated learning, radio frequency fingerprint.**

I. Introduction

RADIO frequency fingerprint identification (RFFI) refers to the analysis and extraction of the characteristics in the emitted signals to identify their frequency fingerprint [1]. The characteristics of the frequency fingerprint can be divided into two types: unintentional modulation and intentional modulation. Common intentional modulation includes frequency modulation, amplitude modulation, and phase modulation. Unintentional modulation comes from the inevitable subtle difference in the performances of oscillators, amplifiers, digital-to-analog converters, pulse modulators, and other electronic equipment in a radio transmitter caused by the hardware. Therefore, these features from the signal represent different radio equipment and can be used to identify them.

Green communications is to reduce public's exposure to radiation to a minimum level which is compatible with a good quality of service. Blind identification requires less information and no training sequences and pilots for energy and spectrum efficient radio communications [2]. The traditional blind identification method uses information, such as signal power, modulation mode, center frequency, and symbol rate, to identify RFFI [3]-[4]. However, with the increase in the number of radiation source devices and the need for accuracy in identification, these traditional features no longer meet the requirements for more accurate classification of individual radiation sources. Therefore, more subtle information is being used to describe the hardware differences in the RFFI from the signal [5]. Duan *et al*. proposed a method of using short-time Fourier transform to extract the time-frequency distribution of the signal as a fingerprint feature recognition method. This method was more suitable for non-stationary signals, such as chirp signals, but it was less effective for complex and irregular signals [6]. Reference [7] proposed an identification method based on the Wegener-Well distribution, but this method had the disadvantage of the existence of cross-interference terms. Reference [8] proposed an individual identification method based on empirical mode decomposition, which unfortunately suffered from modal aliasing. Zhu *et al*. extracted the fractal dimension, pulse rise, fall time, kurtosis and other characteristics to identify RFFI, but this method was sensitive to noise [9]. Hilbert-Huang transform was employed in [10] to extract instantaneous frequency and instantaneous amplitude by combining empirical mode decomposition and Hilbert transform.

On the other hand, deep learning methods are widely used in RFFI. Most traditional convolutional neural network-based methods for the identification of RFFI use network structures directly from the field of image processing. In image processing, natural pictures are used so that the natural picture is rich in detail. The pixel information of each position of the picture may help to understand the content of the whole picture. In this case, using traditional convolution to uniformly sample the data gives good results. However, the fixed structure of traditional convolution has problems when processing contour-type features extracted from radio signals [11]. The content shown in the frequency domain or the time-frequency characteristic of a specific emitter signal is divided into a few places of higher energy, which occupy only a small part of the whole picture and correspond to a kind of contour

This work was supported by the National Natural Science Foundation of China under Grant 62071364, in part by the Aeronautical Science Foundation of China under Grant 2020Z073081001, in part by the Fundamental Research Funds for the Central Universities under Grant JB210104, and in part by the 111 Project under Grant B08038. *(Corresponding author: Chunheng Liu.)*

M. Liu, C. Liu and Z. Yan are with the State Key Laboratory of Integrated Service Networks, Xidian University, Shaanxi, Xi'an 710071, China (e-mail: mqliu@mail.xidian.edu.cn, liuchunheng@126.com, y1223257251@163.com).

Y. Chen is with the School of Engineering, University of Warwick, Coventry, U.K. (e-mail: yunfei.chen@warwick.ac.uk).

N. Zhao is with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, P. R. China. (email: zhaonan@dlut.edu.cn).

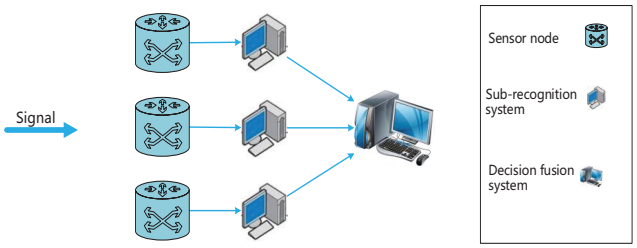

Fig. 1. Radio frequency fingerprint collaborative intelligent identification system model.

information. Meanwhile, most of the background information in the receptive domain of the convolutional kernel is noise, which is of little help to RFFI identification.

As well, deep learning-based identification methods greatly improve the accuracy of RFFI, but its training process requires a large number of samples. It is difficult to have a large training data sample for radio frequency fingerprint. Thus, it is important to break the phenomenon of "data islands" without directly exchanging the original sample data for multiple data owners to participate in training to achieve efficient learning.

In this paper, an intelligent blind RFFI identification method based on deformable convolutional neural network is proposed. The main contributions of this work are summarized as follows:

- We introduce an intelligent representation of Choi-Williams distribution (CWD), ambiguity function (AF) and bi-spectrum for efficient processing of radio frequency fingerprint. These features fully represent the various features of the radio frequency fingerprint for distributed sensor system.
- Our proposed method replaces part of the convolutional layer in the traditional convolutional network with a deformable convolutional layer, and uses the idea of deep separable convolution to perform convolution.
- We employ federated learning for learning and recognition. The multi-sensor distributed learning realizes efficient radio frequency fingerprint identification while ensuring the information security of all sensors.
- Compared with the existing methods, our experiments indicate that the proposed method has better identification performance. Moreover, the proposed method requires only a small amount of information without training sequences and pilots.

The remainder of this paper is organized as follows. We introduce the system model in Section II. In Section III, we present an intelligent representation of radio frequency fingerprint. Intelligent recognition of radio frequency fingerprint based on deformable convolutional network and federated learning are discussed in Section IV. Simulation studies are given in Section V. Section VI summarizes the whole paper.

II. System Model

In this paper, we consider a transmission network with multiple sub-nodes and distributed federated learning structure as the radio frequency fingerprint collaborative intelligent blind identification system, which is shown in Fig. 1. The sub-recognition system blindly identities RFFI using signals from each sensor and deformable convolutional network. A central node is selected from all edge nodes as the fusion center for parameter fusion and output coordination to complete federated learning. In the system, the received signal can be written as [1]:

$$X(t) = HS(t) + N(t), \tag{1}$$

where $S(t)$ is the automatic dependent surveillance CE broadcast (ADS-B) signal, H stands for the transmission channel and $N(t)$ represents the additive Gaussian noise. As a common radiation source, ADS-B signal is often used to verify the RFFI identification technology of radiation source. Therefore, this paper adopts ADS-B signals as input.

III. Intelligent Representation of Radio Frequency Fingerprint

In order to fully exploit of a deep learning network, the radio frequency fingerprint needs to be intelligently characterised before being fed into the learning network. The proposed method extracts the time-frequency feature, ambiguity function (AF) and bi-spectrum of the radio frequency fingerprint as intelligent representation. The three intelligent representation are used to convert one-dimensional signals into two-dimensional features, similar to images, which have a good fit for deep learning networks.

A. Choi-Williams Distribution

Choi-Williams distribution (CWD) describes the time-frequency characteristics, which belongs to the Cohen-like time-frequency distribution. CWD is obtained by two-dimensional convolution of the Wigner-Ville distribution with a kernel function, which is an improved method for the defects of the Wigner-Ville distribution as [12]

$$CW_s(t,\omega)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\sqrt{\frac{\sigma}{4\pi\tau^2}}\exp\left(-\frac{(t-u)^2}{4\tau^2/\sigma}\right)r_s(t,\tau)e^{-j\omega t}d\tau du, \tag{2}$$

where σ is the scaling factor that plays an important role in localizing the time-frequency energy marginals of a signal. $r_s(t,\tau)$ is called the kernel function, representative of particular distribution function, u and τ give the time lag and the frequency lag, respectively. CWD has a good cross-term suppression effect and can well present the distribution of signal frequency information in the time plane.

B. Ambiguity Function

AF is a secondary time-frequency distribution that depicts the internal structure of the signal. The AF $AF_s(\tau,\xi)$ is the joint representation of the signal on the time delay τ and the Doppler frequency shift ξ, which can be given by [13]

$$AF_s(\tau,\xi) = \int_{-\infty}^{+\infty} s(t)\, s^*(t+\tau)\, e^{j2\pi\xi t}dt. \tag{3}$$

C. Bi-spectrum

Bi-spectrum is a signal obtained by two-dimensional Fourier transform, and contains phase information that is not available in the second-order statistics. It reflects the relationship between every two frequency points of the signal in the frequency domain [14]. Because the high-order statistics of the Gaussian signal equal to zero, the bi-spectrum can suppress the Gaussian noise. The third-order cyclic cumulant of the received signal $s(t)$ can be written as

$$C_{3s}(\tau_1, \tau_2) = E\{s^*(t)\, x(t+\tau_1)\, x(t+\tau_2)\}, \quad (4)$$

where $C_{3s}(\tau_1, \tau_2)$ is the third-order cyclic cumulant of each signal. The bi-spectrum $B_s(\omega_1, \omega_2)$ can be obtained from the third-order cumulant of the signal as

$$B_s(\omega_1, \omega_2) = \sum_{\tau_1} \sum_{\tau_2} C_{3s}(\tau_1, \tau_2)\, e^{-j(\omega_1 \tau_1 + \omega_2 \tau_2)}. \quad (5)$$

The above three intelligent representations can describe the internal information of the radio frequency fingerprint. Therefore, they will be merged and input from the three channels into the deformable convolutional network to be discussed below.

IV. Radio Frequency Fingerprint Collaborative Blind Identification

A. Deformable Convolutional Network

Convolutional neural network is not only a machine learning architecture which contains a large number of neuron nodes, but also is a deep learning network structure that simulates the neural model of the visual cortex of the biological brain. Compared with the traditional feature extractor, the convolutional neural network does not need to manually design the feature extraction section and it is carried out along with the training. Moreover, the feature extraction of the convolutional neural network is composed of a neural network whose weights are obtained through training. This transforms manual analysis into automatic processing to improve the quality of features [15].

The geometric structure of the receptive field in the traditional convolutional neural network model was fixed. The input feature processing will be limited by the size of the convolution kernel and the input feature can only be sampled at a fixed position. For traditional image processing, the pixel information at each position of the picture may contribute to the understanding of the content of the entire picture due to the rich details of natural pictures. In this case, using traditional convolution to uniformly sample the data can get good results. However, the fixed structure of traditional convolution will cause problems when processing features extracted from radio signals.

If this normalized rectangular receptive field can be broken, the network will be more inclined to useful information content during convolution operations and ignoring part of the background noise information, which will be more conducive to the network's learning of signal features for better classification results. Deformable convolution adds a two-dimensional offset to the sampling position of traditional convolution, which can freely deform the size of the convolution kernel and focus on the area of interest. Fig. 2 (a) shows the sampling position of the traditional convolution while Fig. 2 (b) and Fig. 2 (c) depict the sampling position with the offset in the deformable convolution. In Fig. 2 (b) and Fig. 2 (c), it is shown that the size, aspect ratio, and rotation angle of the sampling window in the deformable convolution can be changed. These offsets are learned from the input feature by adding a new convolutional layer. Therefore, the size of the sampling window depends on the input features [16].

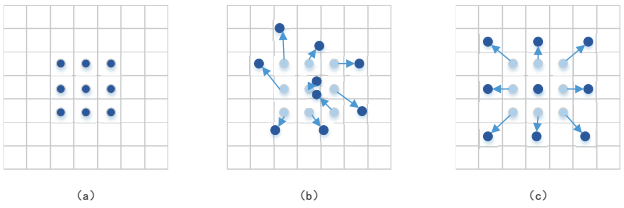

Fig. 2. Sampling positions of traditional convolution and deformable convolution.

In order to achieve deformable convolution, it is necessary to create an additional convolution layer to model the offset and use the conventional convolution kernel to sample on the input feature. For the output feature y, the position is obtained as

$$y(p_0) = \sum_{p_i \in F} w_i \cdot x(p_0 + p_i), \quad (6)$$

$$F = \{(-1,-), (-1,0), ..., (0,1), (1,1)\}, \quad (7)$$

where F represents the receptive field that can be sampled by the convolution kernel, w_i is the weighted value at position p_i, $y(p)$ and $x(p)$ refer to the feature value at p, and p_i is the preset offset of the conventional convolution kernel. After adding a new offset Δp_i, $y(p_0)$ becomes

$$y(p_0) = \sum_{p_i \in F} w_i \cdot x(p_0 + p_i + \Delta p_i). \quad (8)$$

Thus, the sampling position of the convolution kernel becomes an irregular offset position $p_i + \Delta p_i$. Since the offset Δp_i is usually a decimal, (8) needs to be implemented by bilinear interpolation as

$$x(p) = \sum_{q} G(q, p) \cdot x(q), \quad (9)$$

where $p = p_0 + p_n + \Delta p_n$ represents the decimal position, q traverses all positions in the input feature x, $G(\cdot, \cdot)$ denotes a two-dimensional bilinear interpolation kernel, which consists of two one-dimensional spatial kernels

$$G(q, p) = g(q_x, p_x) \cdot (q_y, p_y), \quad (10)$$

where $g(q, p) = \max(0, 1 - |q - p|)$.

Fig. 3 shows the principle of deformable convolution. The offset is obtained by creating a new convolution layer on the same input feature map. The channel size $2N$ corresponds to N two-dimensional offsets, and the deformed convolution kernel has the same spatial resolution and expansion as the

current convolution layer. During the training, the convolution kernel and the offset are learned at the same time to generate the output feature, where the offset is back-propagated through the bilinear operation in (10) to learn the gradient descent.

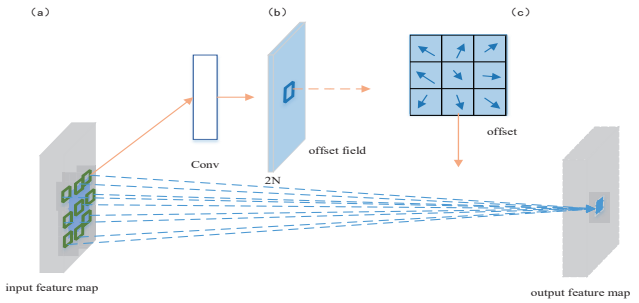

Fig. 3. Deformable convolution principle.

The structure of the deformable convolutional network is similar to that of the convolutional neural network, except that the convolutional layer is replaced with a deformable convolutional layer. Considering that the feature map becomes smaller and the feature contour becomes blurred as the number of layers deepens, it is only necessary to replace the first few convolutional layers with deformable convolutional layers.

The input feature map and the convolution in traditional convolutional neural networks are both three-dimensional, while deformable convolution adds an offset to the two-dimensional space to process data. The proposed method has three channels in the input layer, as the time-frequency diagram, ambiguity function and bi-spectrum. Different from the three channels of image processing, the energy distribution and contours of these three feature maps are quite different, and the offset of the convolution kernel when using deformable convolution is also completely different. Therefore, each channel needs to be processed separately in two dimensions during the convolution operation. Depth separable convolution as a light-weight modal method can not only operate on the input feature map of each channel individually, but also reduce the computational burden of the network.

B. Grouped Convolution and Pointwise Convolution

Grouped convolution was first proposed by Google. The idea is to perform channel-by-channel convolution. Each channel will only be convolved by one convolution kernel, and each convolution kernel is only responsible for one channel. Grouped convolution separately collect the characteristics of the data on each channel, in line with the requirement of inputting three different characteristics in the three channels in this paper.

For example, an input of $\mathrm{H}\times\mathrm{W}\times\mathrm{C}$, where H and W are the length and width of the feature map, and C is the number of channels, is divided into C groups, and each group contains the data on one channel. Then performs $\mathrm{K}\times\mathrm{K}$ two-dimensional convolution for each group. The number of convolution kernels is the same as the number of channels. Fig. 4 shows the principle of grouped convolution.

The idea of point-wise convolution is to mix the information at the same position of each channel. For example, it is also

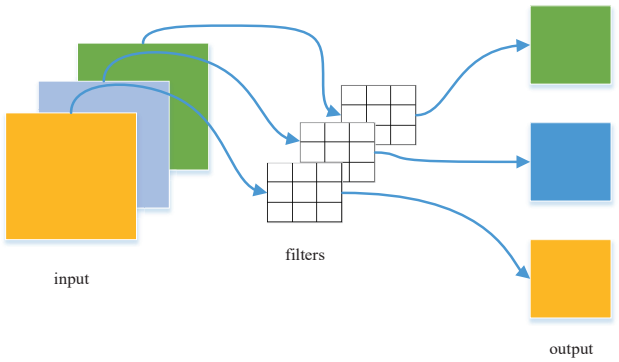

Fig. 4. Grouped Convolution principle.

an input of $\mathrm{H}\times\mathrm{W}\times\mathrm{C}$, doing $\mathrm{H}\times\mathrm{W}$ three-dimensional convolution. Each position on the input feature map is weighted in the depth direction, and there are as many output feature maps as there are convolution kernels. Fig.4. shows the principle of point-wise convolution.

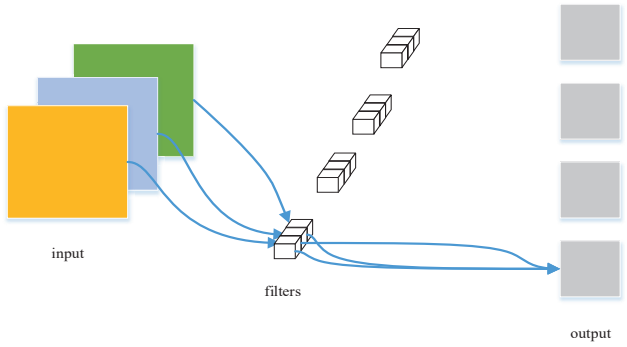

Fig. 5. Point-wise convolution principle.

C. Distributed Federated Learning

Data silos seriously hinder the development of big data and artificial intelligence, but the data they own is a valuable asset, and even involves a lot of privacy for any data owner. Therefore, how to realize resource sharing and improve the identification ability of the network under the premise of ensuring that private data is not leaked, federated learning has been widely concerned. Federated learning is an emerging technology that combines big data processing and artificial intelligence. The idea of federated learning is to break the data islands to ensure the information security of all parties [17]-[18].

Federated learning defines a new framework. Under this framework, each data owner does not need to upload its own data, but only needs to train the model locally, encrypt the network weight and loss function through the sensor and upload them to the aggregation center. After the aggregation center receives and decrypts the network weight and loss function, it calculates the global weight and the global loss function according to the aggregation criteria and feeds them back to each local model. All sensors perform the next step of

gradient optimization based on the data from the aggregation center until the network model converges. Only the loss function information during network training is uploaded for security. From the above, federated learning has the following excellent characteristics:

- Data information isolation: the original data at the input end is only locally processed and will not be leaked to the outside, which meets the needs of data owners for information security and privacy protection;
- Every sensor has the same status and can enjoy the network model finally trained, so that data owners can participate in cooperation in a fair and equitable manner;
- The aggregation center feeds back global information, and all sample information participates in the training to ensure the better performance and quality of the joint training model than the model trained by either sensor;
- The data owners of all sensors exchange model parameters during local distributed training to jointly improve the quality of the model.

Federated learning consists of two elements: data source and federated learning network. The relationship between them is shown in Fig. 6. This structure can ensure that no one except the original owner of the data can access their original data, even at the server. The server can only obtain local network weight and loss function, which ensures that everyone can safely share the global training model [19]-21].

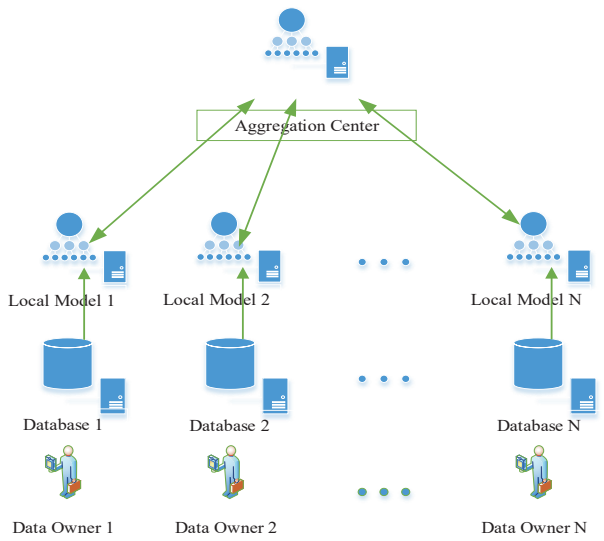

Fig. 6. Distributed federated Learning Structure.

The entire work-flow of federated learning can be divided into several steps. Firstly, all training participants use their own data and initialization parameters to train the network model locally. Each local model updates the network weight according to its own loss function. When the number of updates reaches a specified number of times, the network parameters are encrypted and uploaded to the aggregation center through the sensor node. The aggregation center obtains the data from all the sensors, decrypts the data, calculates the global information according to the aggregation criteria, and transmits it back to each sensor node. After each local model receives the global information from the aggregation center, it continues to train the network according to the global information, until a global convergence is achieved.

The purpose of training the network is to keep the network loss function smaller and make the predicted value of the network closer to the true label value. For federated learning, the purpose of training is to minimize the global loss function. In federated learning, the weight update process is divided into local update and global update. When the local model is trained, the batch size is set to b_i. After training b_i samples, a network weight w will be obtained and the loss function $f_i(w)$ will be calculated once, and the network weight will be locally updated through the loss function; When a global update is required, all local sensor nodes will encrypt and upload $f_i(w)$, w and b_i to the federated learning aggregation center. After the aggregator receives and decrypts the data from N sensors, it calculates the weighted average of these data to obtain the global loss function as

$$F(w) = \frac{\sum_{i=1}^{N} f_i(w) \times b_i}{\sum_{i=1}^{N} b_i}. \tag{11}$$

The ultimate goal is to minimize the global loss function or

$$w^* = \arg\min F(w), \tag{12}$$

where w^* is the optimal weight. Due to the inherent complexity of most machine learning models, it is usually impossible to find a closed-form solution to (12). Therefore, the gradient descent method is usually used to solve (12). Here, a typical distributed gradient descent algorithm is used to solve (12). At present, this algorithm is widely used in federated learning.

Each local model has its own network weight $w_i(t)$. when $t = 0$, the weights of all local models are initialized to the same value. When $t > 0$, the network weight will be updated by the gradient descent method according to the last weight along the negative direction of the loss function, and the new value of $w_i(t)$ will be calculated. After the local model undergoes a specified number of local updates, it performs a global aggregation. After global aggregation, the network weights on each local model are usually changed. For convenience, use $\widetilde{w}_i(t)$ to represent the network weight of the local model i after global aggregation. If no aggregation operation has been performed at the t-th update, then $\widetilde{w}_i(t) = w_i(t)$. If the aggregation operation is performed in the t-th iteration, then $\widetilde{w}_i(t) \neq w_i(t)$, set $\widetilde{w}_i(t) = w(t)$, where $w(t)$ denotes the weighted average of $w_i(t)$ and can be expressed as

$$w(t) = \frac{\sum_{i=1}^{N} w_i(t) \times b_i}{\sum_{i=1}^{N} b_i}. \tag{13}$$

The change of the above weight update is shown in Fig. 7.

After global aggregation, the network will feed back $w(t)$ to each local model, and each local model will update the network weight to the global network weight $w(t)$. This step

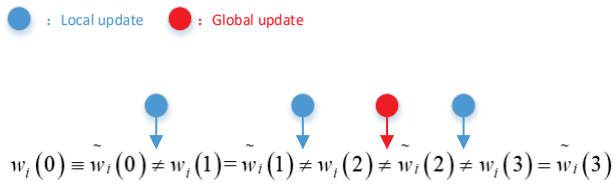

Fig. 7. Network weight update process of the local model.

is the global update. In order to ensure that the loss function of the update weight is the smallest when the global update is performed, a new parameter w^f is set and used to record the network weight during the last global aggregation. Calculate the corresponding loss functions of w^f and $w(t)$ according to (13), and make the weight w^f corresponding to the smallest loss function be

$$w^f = \arg\min F(w), w \in \left\{w^f, w(t)\right\}. \tag{14}$$

Feedback w^f to each local model for global update. Assuming that the local model performs a global update after τ local updates are performed. From the above, the steps of the distributed gradient descent algorithm are summarized in **Algorithm 1**.

Algorithm 1 Distributed gradient descent algorithm

1: Initialize w^f, $w_i(0)$ and $\widetilde{w_i}(0)$;
2: Partially update each local model;
3: When the total number of updates T is an integer multiple of τ, each local model uploads data to the aggregation center to calculate the weighted average of all network weights $w(t)$. Update w^f by comparing the loss function, and pass the global weight w^f back to each local model;
4: Repeat step 2 and step 3 until the network converges.

According to the deformable convolutional network and the distributed federated learning system introduced in the previous section, the specific process of the intelligent identification method of RFFI based on the deformable convolutional network is introduced as follows: Firstly, each training participant performs the intelligent representation of radio frequency fingerprint, and inputs the intelligent representation of the radio frequency fingerprint into their respective local deformable convolutional networks, and then obtain their respective network weights after partial updates based on their respective data. All participants upload the network weights to the aggregation center. After aggregation, the aggregation center feeds back the aggregated network weights to each local model for global update. The process of local update and global update is repeated until the network converges.

V. Numerical Results and Discussion

The simulation was performed on an Intel core i9-9920X with GPU RTX2080Ti. The simulation signal uses an S-mode transponder to expand the message (1090es, The ADS-B signal of 1090 MHz mode s extended squitter) has a

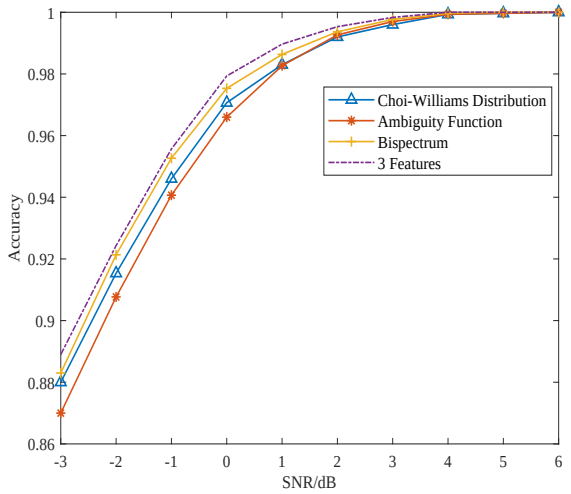

Fig. 8. Identification performance based on deformable convolutional network with different features.

working frequency of 1090 MHz, a data rate of 1 Mbps, and a modulation mode of PPM and 2ASK. The period of the signal is 120 μs, the leading pulse lasts 8 μs, and the duration is 0.5 μs. The starting time of the pulses is at 0.1 μs, 3.5 μs, and 4.5 μs respectively. The information pulse occupies 112 μs and transmits 112 bits of data. One bit of data represents a message, including the position, altitude, speed, heading, identification number, and other information of the aircraft. The 01 and 10 binary data used after PPM modulation represents each message. We intercept the data of 5 μs length of ADS-B signal's leading pulse, use the frequency of 600MHz to sample, and set three different individuals in total, each of which generates five different signals. Table I shows the target parameter settings. The channel has Gaussian white noise, and the signal-to-noise ratio (SNR) ranges from -3dB to 6dB. Each individual has 800 sample data for network training and 200 sample data for testing for each SNR. Therefore, the total number of training samples is 40,000, and the total number of test set samples is 10,000. The learning framework is the Keras framework, and GPU computing is used when training the network model. During training, intelligently characterize all the signals in the sample set and input the deformable convolutional neural network for training. The stochastic gradient descent optimization method is used in the training process, the loss function is the cross-entropy loss function, and Table II shows the deformable convolutional network structure parameter settings.

The identification accuracy when inputting different features in the input layer of the deformable convolutional network is shown in Fig. 8. The deformable convolutional layer in the network is used in the first two layers. From Fig. 8, it can be seen that the recognition rate of several input features is more than 99% when the SNR is 3dB, but when the SNR is lower than 3dB, the performance of inputting three features at the same time is obviously better than that when only one of the features is used.

When the number of deformable convolutional layers in the

TABLE I
TARGET PARAMETER SETTINGS.

Individual	Parameter	Parameter Settings				
Target 1	Modulation frequency of the phase noise(MHz)	4	6	7	10	15
	Phase modulation coefficient	0.16	0.27	0.32	0.15	0.25
	Harmonic component	1	0.5	0.3	0.2	0.1
Target 2	Modulation frequency of the phase noise(MHz)	2	5	9	11	13
	Phase modulation coefficient	0.21	0.32	0.15	0.24	0.28
	Harmonic component	1	0.8	0.6	0.4	0.2
Target 3	Modulation frequency of the phase noise(MHz)	3	5	6	8	12
	Phase modulation coefficient	0.34	0.3	0.23	0.21	0.26
	Harmonic component	1	0.1	0.08	0.05	0.03
Target 4	Modulation frequency of the phase noise(MHz)	2	5	7	10	15
	Phase modulation coefficient	0.21	0.32	0.32	0.15	0.25
	Harmonic component	1	0.8	0.3	0.2	0.1
Target 5	Modulation frequency of the phase noise(MHz)	2	5	9	8	13
	Phase modulation coefficient	0.21	0.3	0.15	0.21	0.28
	Harmonic component	1	0.1	0.15	0.05	0.28

TABLE II
STRUCTURE PARAMETER SETTINGS OF DEFORMABLE CONVOLUTIONAL NETWORK.

Structure Type	Size	Step Size	Number of Filters	Activation Function
Input layer	$128 \times 128 \times 3$	-	-	-
(Deformable) convolutional layers	5×5	1	96	ReLU
batch normalization layer	-	-	-	-
Max pooling layer	3×3	3	-	-
(Deformable) convolutional layers	3×3	1	256	ReLU
batch normalization layer	-	-	-	-
Max pooling layer	3×3	3	-	-
(Deformable) convolutional layers	3×3	1	384	ReLU
Convolutional layers	3×3	1	384	ReLU
Convolutional layers	3×3	1	256	ReLU
Max pooling layer	3×3	2	-	-
Dense layer	1024	-	-	ReLU
Dense layer	1024	-	-	ReLU
Dense layer	1024	-	-	ReLU
Dense layer	5	-	-	softmax

deformable convolutional network is set to one, two and three, the RFFI identification accuracy in the three cases are shown in Fig. 9. It has been found that setting two deformable roll layers has a better performance than setting one layer. However, the contour features of the feature map after the deformable convolution operation are not as obvious as when it was just input, The identification performance is not satisfactory when increasing the number of deformable convolution layers.

Fig. 10 compares the identification performance of the deformable convolutional network and the traditional convolutional neural network. The deformable convolutional network has three deformable convolutional layers. From Fig. 10, it can be seen that the deformable convolutional network has a better identification performance than the traditional convolutional neural network.The deformable convolutional network has a better learning effect than the traditional convolutional neural network in processing this contour-type signal feature. The identification performance of the deformable convolutional network for each of the five individuals when the three-layer deformable convolutional layer is set in Fig. 11. When the SNR is greater than 3dB, the identification rate is more than 99%.

In the case of setting the same parameters, the deformable convolutional network requires an additional convolution layer to calculate the offset of the convolution kernel, which consumes more computing power, and therefore requires a longer training time. Under the same data and hardware platform, the training time of the convolutional neural network is 1023s, while the deformable convolutional network needs 2564s, which increases the burden of computing power while improving the recognition performance.

Fig. 12 shows the effects of different Doppler frequency shifts and ground-to-air channel on the identification performance caused by the movement of the aerial radiation. In this simulation, the flight speed of the aircraft is set as the general speed of a civil airliner (800-1000 km/h), the flying height is

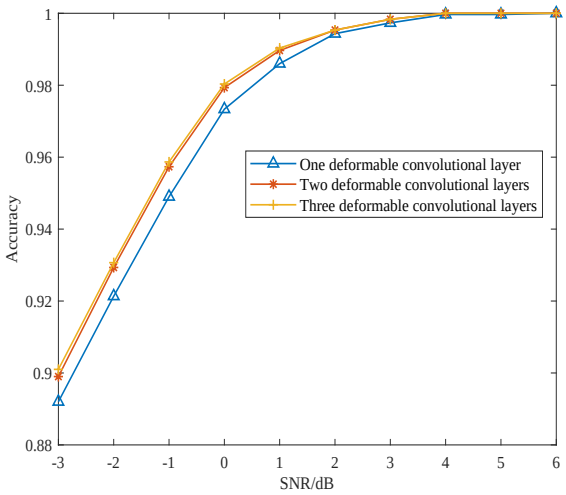

Fig. 9. Identification performance with different deformable convolutional layers.

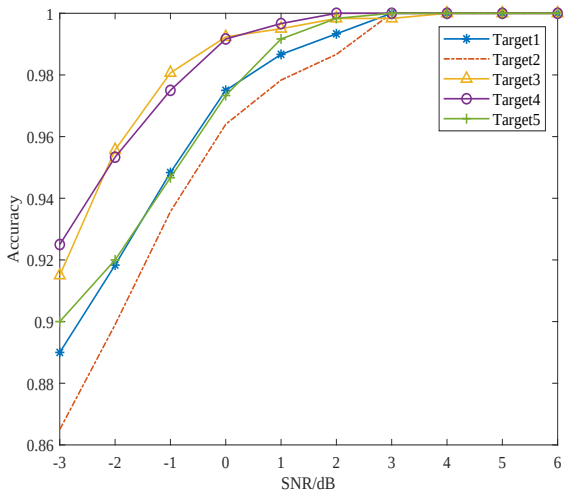

Fig. 11. Each individual identification performance with the deformable convolutional network.

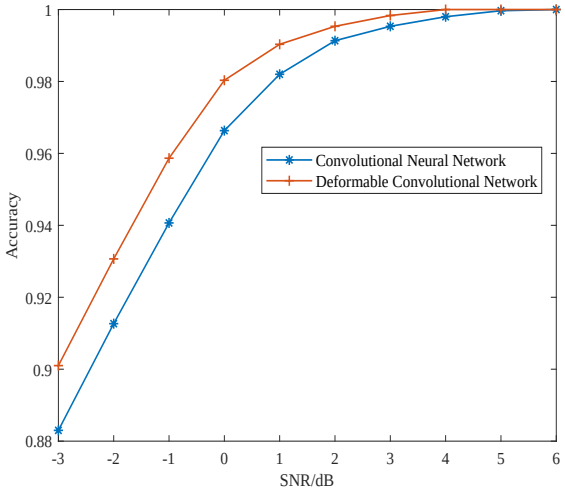

Fig. 10. Identification performance compared with the different networks.

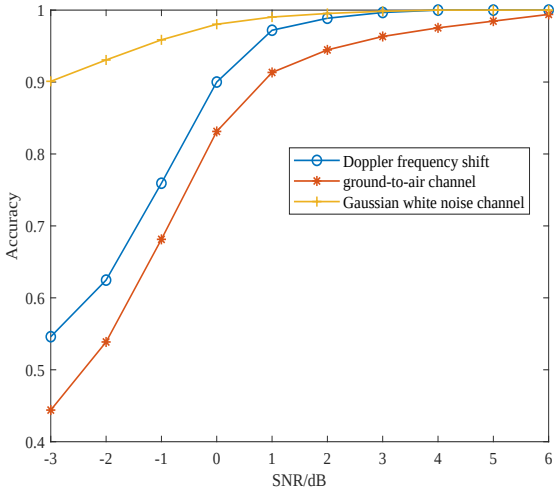

Fig. 12. Identification performance with Doppler frequency shift over ground-to-air channel.

8000-12000 meters, the signal receiving antenna height is 3 meters, and the ground reflection coefficient is 0.4-0.6 [22]. It can be seen from Fig.12 that the frequency shift and the fading caused have a greater impact on the identification performance in a low SNR environment. Compared with the Gaussian white noise channel, the identification accuracy rate is lower. When the SNR is greater than 0dB, the identification rate exceeds 90%. When the SNR is close to 6dB, it reaches more than 99% in the ground-to-air channel environment.

Simulations are conducted to verify the effect of distributed federated learning system on network training. The network architecture with deformable convolutional network as the local model is compared with the federated learning and local model training alone. Three local models are set up in the simulation, and each local model corresponds to a separate database. The signals in the three databases are the ADS-B signals of the five different individuals, and the channel is AWGN. The SNR is from -3dB to 6dB. Database 1 has 100 sample data with each individual under each SNR for training, database 2 has 120 sample data with each individual under each SNR for training, while database 3 has 150 sample data for each individual under each SNR. During the training, each local model is uploaded to the aggregation center for global update every time it is locally updated.

Fig. 13 shows the identification accuracy of distributed federated learning and individual training of each local model. From Fig. 13, it can be seen that distributed federated learning effectively implements multi-location distributed training, and the identification performance is significantly better than the individual training of each local model.

Database 1 only has signals in the SNR environment of -3dB to -1dB, and 800 samples for each individual under each SNR are used for training; Database 2 only has signals in the 0dB to 2dB SNR environment, and 800 samples data for each individual under each SNR are used for training; Database 3

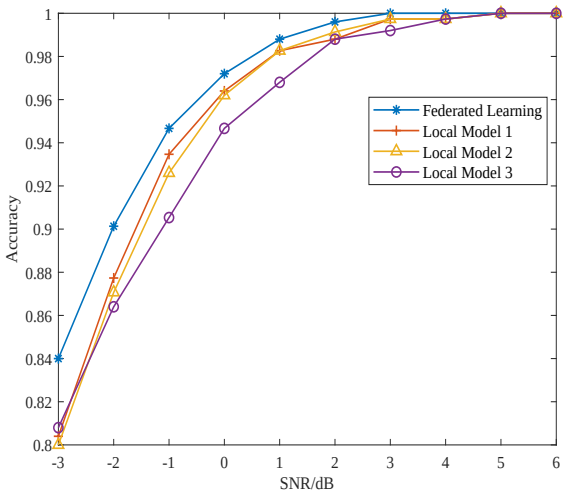

Fig. 13. Identification performance with the insufficient samples.

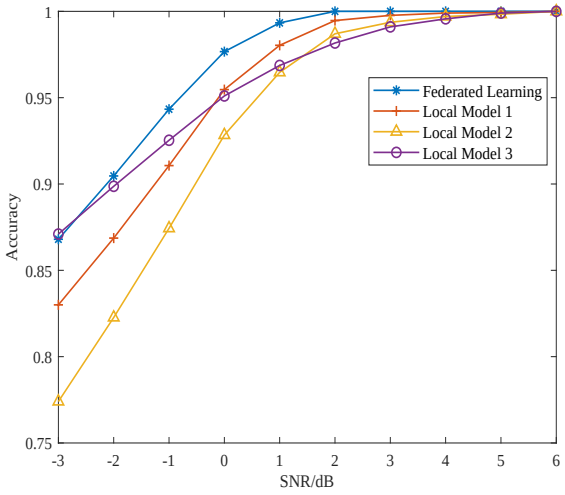

Fig. 14. Identification performance with insufficient SNR information.

only has signals in the SNR environment of 3dB to 6dB, and 800 samples data for each individual under each SNR are used for training.

Fig. 14 shows the identification accuracy of distributed federated learning and individual training of each local model in the case of insufficient SNR information. It can be seen from Fig. 14 that distributed federated learning effectively implements multi-location distributed training, and the identification performance is significantly better than that of each model training separately.

TABLE III
PERFORMANCE COMPARISON WITH THE DIFFERENT METHODS

Methods	-3dB	0dB	3dB	6dB
Proposed	84.0%	97.6%	99.7%	99.9%
Method 1	79.6%	93.4%	99.5%	99.8%
Method 2	73.0%	85.2%	92.6%	96%

Under the same parameter settings and simulation experiments, the signals all use the ADS-B signals of the five individuals used in the previous experiment, and compare the different individual identification methods. Each of these methods sets 3 local models. When the SNR information is sufficient, the number of samples for each individual under each SNR in each local database is 100, 80, and 50, respectively. The results are listed in Table III. In Table III, the proposed method are the deformable convolution network and federated learning method, method 1 is deep belief networks (DBN) [23] with federated learning, and method 2 is Multilayer Perceptron (MLP) [24] with federated learning. Compared with deep belief networks and MLP, the deformable convolutional network proposed in this paper achieves identification accuracy more than 99%, when SNR is 3dB. Therefor, the proposed method has a better identification performance. Under the same parameters, operating platform and environment, the deformable convolutional network needs to consume more computing power because it requires an additional convolutional layer to calculate the offset of the convolution kernel, so it requires a longer training time. The training time of the proposed method in this paper is 8230s, the method 1 training takes 3256s, and the method 2 training takes 2586s.

VI. CONCLUSION

We have proposed an blind radio frequency fingerprint identification method based on deformable convolutional neural network, and the effectiveness of deformable convolutional network was verified through simulation. In view of the lack of local training sample data, it is necessary to realize the problem of multi-party joint training model without exchanging sample original data. In addition, a distributed federated learning system has been proposed to improve identification performance. Simulation results have shown that the distributed federated learning system can effectively realize the multi-party joint training model, and that the proposed method can achieve better identification performance.

REFERENCES

[1] M. Liu, J. Wang, N. Zhao, Y. Chen, H. Song, R. Yu. “Radio frequency fingerprint collaborative intelligent identification using incremental learning," *IEEE Transactions on Network Science and Engineering,* 2021, DOI: 10.1109/TNSE.2021.3103805.

[2] M. Liu, N. Nan, B. Shang, et al. “Energy and spectrum efficient blind equalization with unknown constellation for air-to-ground multipath UAV communications," *IEEE Transactions on Green Communications and Networking,* vol. 5, no. 3, pp. 2473-2400, Sept. 2021.

[3] K. Talbot, P. Duley, M. Hyatt, “Specific emitter identification and verification," *Technology Review Journal,* pp. 113-133, 2003.

[4] L. Wu, Y. Zhao, Z. Wang, et al. “ Specific emitter identification using fractal features based on box-counting dimension and variance dimension," in *Proc. 2017 IEEE International Symposium on Signal Processing and Information Technology,* 2017, pp. 226-231.

[5] S. Rehman, K. Sowerby, C.Coghill, “ RF fingerprint extraction from the energy envelope of an instantaneous transient signal," in *Proc. 2012 Australian Communications Theory Workshop,* 2012, pp. 90-95.

[6] F. Duan, M. Corsar, D. Mba, “Using empirical mode decomposition scheme for helicopter main gearbox bearing defect identification," in *Proc. Prognostics & System Health Management Conference,* 2017, pp. 1-4.

[7] H. Hang, “Time-frequency DOA estimate algorithm based on SPWVD," in *Proc. IEEE International Symposium on Microwave,* 2005, pp. 1253-1256.

[8] S. Guo, R. White, M. Low, “A comparison study of radar emitter identification based on signal transients," in *Proc. 2018 IEEE Radar Conference,* 2018, pp. 1-7.

[9] M. Zhu, X. Zhang, Y. Qi, et al. “Compressed sensing mask feature in time-frequency domain for civil flight radar emitter recognition," in *Proc. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing,* 2018, pp. 15-20.

[10] H. Wang, T. Zhang, “Specific emitter identification based on fractal and wavelet theories," in *Proc. Advanced Information Technology, Electronic & Automation Control Conference,* 2017, pp. 1613-1617.

[11] M. Liu, Z. Liu, W. Lu, Y. Chen, X. Gao, N. Zhao, “Distributed few-shot learning for intelligent recognition of communication jamming," *IEEE Journal of Selected Topics in Signal Processing,* vol. 16, no. 3, pp. 395-405, Apr. 2022.

[12] T. Huynh-The, V. Doan, C. Hua, Q. Pham, T. Nguyen, D. Kim, “Accurate LPI radar waveform recognition with CWD-TFA for deep convolutional network," *IEEE Wireless Communications Letters,* vol. 10, no. 8, pp. 1638-1642, Aug. 2021.

[13] D. Dash, V. Jayaraman, “Ambiguity function analysis for orthogonal-LFM waveform based multistatic radar," *IEEE Sensors Letters,* vol. 5, no. 12, pp. 2475-1472, Dec. 2021.

[14] A. Ducout, F. Bouchet, S. Colombi, D. Pogosyan, S. Prunet, “Non-gaussianity and minkowski functionals: forecasts for planck," *Monthly Notices of the Royal Astronomical Society,* vol. 429, no. 3, pp. 2104-2126, Mar. 2013.

[15] J. Dai, H. Qi, Y. Xiong, et al. “Deformable convolutional networks," in *Proc. Proceedings of the IEEE international conference on computer vision,* 2017, pp. 764-773.

[16] F. Chen, F. Wu, J. Xu, et al. “Adaptive deformable convolutional network," *Neurocomputing,* vol. 453, pp. 853-864, Sept. 2021.

[17] F. Chen, P. Li, T. Miyazaki, C. Wu, “FedGraph: federated graph learning with intelligent sampling," *IEEE Transactions on Parallel and Distributed Systems,* vol. 33, no. 8, pp. 1775-1786, Aug. 2021.

[18] Z. Wei, Q. Pei, N. Zhang, X. Liu, C. Wu, A. Taherkordi, “Lightweight federated learning for large-scale IoT devices with privacy guarantee," *IEEE Internet of Things Journal,* 2021. DOI: 10.1109/JIOT.2021.3127886.

[19] C. Wu, Z. Liu, F. Liu, T. Yoshinaga, Y. Ji, and J. Li, “Collaborative learning of communication routes in edge-enabled multi-access vehicular environment," *IEEE Transactions on Cognitive Communications and Networking,* vol.6, no.4, pp. 1155-1165, Dec. 2020.

[20] R. Yin, Y. Shen, H. Zhu, X. Chen, and C. Wu, “Time-critical tasks implementation in MEC based multi-robot cooperation systems," *China Communications,* vol. 19, no. 4, pp. 199 - 215, Apr. 2022.

[21] X. Chen, C. Wu, Z. Liu, N. Zhang, and Y. Ji, “Computation offloading in beyond5G networks: a distributed learning framework and applications," *IEEE Wireless Communications,* vol.28, no.2, pp.56-62, Apr. 2021.

[22] M. Liu, C. Liu, M. Li, Y. Chen, S. Zheng, N. Zhao, “Intelligent passive detection of aerial target in space-air-ground integrated networks," *China Communications,* vol. 19, no. 1, pp. 52-63, Jan. 2022.

[23] J. Liu, N. Wu, Y. Qiao, Z. Li, “Short-term traffic flow forecasting using ensemble approach based on deep belief networks," *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 1, pp. 404-417, Jan. 2022.

[24] J. Zhu, J. Qin, F. Yin, Z. Ren, J. Qi, J. Zhang, R. Wang, “An APMLP deep learning model for bathymetry retrieval using adjacent pixels," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 15, pp. 235- 246, Dec. 2021.