WILEY | Hindawi

Research Article

Conventional Neural Network-Based Radio Frequency Fingerprint Identification Using Raw I/Q Data

Tian Yang [ID], Su Hu [ID], Weiwei Wu [ID], Lixin Niu, Di Lin [ID], and Jiabei Song

University of Electronic Science and Technology of China, China

Correspondence should be addressed to Su Hu; husu@uestc.edu.cn

Radio frequency (RF) fingerprint identification is a nonpassword authentication method based on the physical layer of communication devices. Deep learning methods have thrown new light on RF fingerprint identification. In this paper, a conventional neural network- (CNN-) based RF identification model is proposed. The CNN models are designed to be lightweight. Raw data that reflects the characteristics of the I channel, the Q channel, and the 2-dimensional $I + Q$ data is successively fed into a CNN model. Therefore, three submodels are generated. The final predictive labels are determined by the results of the three submodels through a voting scheme. Experimental results have demonstrated that in the SNR setting at 5 dB, the final recognition accuracy of four transmit devices could achieve as high as 97.25%, while the identification accuracies based on the I channel data, Q channel data, and $I + Q$ channel data are 94.5%, 95%, and 94.5%, respectively. The training time for the 4 devices is around 30 seconds.

1. Introduction

By analysing the subtle differences of transmitters' RF fingerprints, RF fingerprint-based learning models could distinguish varieties of devices, thereby being difficult to clone and fake [1–4]. These subtle differences of transmitters are normally hard to be identified, and artificial intelligence (AI) methods could be appropriate to mitigate this problem [5–7].

AI models are divided into two main categories, including machine learning and deep learning. Machine learning (ML) needs first to extract the features of data samples and then select an appropriate model to train the samples to get model parameters. The ML methods include decision tree, naive Bayesian, K-nearest neighbor, and support vector machine. For communication signals, the extracted features include carrier frequency, phase noise, constellation, high-order moments, power spectrum, and fractal features. In [8], Hu et al. extracted primary features, such as information dimension, constellation feature, and phase noise spectrum of the transmitted signal. The ML models, including the sup-

port vector machine, bagged tree, and weighted K-nearest neighbor, were used for RF fingerprinting recognition.

The feature extraction process requires professional knowledge. Moreover, specific information in original data might be lost after feature extraction. To mitigate the issues, unlike ML models, deep learning (DL) models could directly process raw data, transform the original data into higher-level representations, and automatically learn better feature representations. The automatic learning process could replace feature extraction, thus avoiding the procedure of feature engineering that may be at the cost of extracomputing resources. Examples of DL methods include feedforward neural networks [9], CNN, and recurrent neural networks [10, 11]. Particularly, the CNN has widely been applied in signal processing. In [12], Merchant et al. developed a framework for training a CNN algorithm using the time-domain complex baseband error signal. Building on seven 2.4 GHz commercial Zigbee devices, the experimental results demonstrated 92.29% identification accuracy. The robustness of the method over a wide range of signal-to-noise ratios (SNRs) was also illustrated. In [13], raw I/Q data

was used as a 2-dimensional input dataset into the CNN model to perform modulation classifications. In [14], raw I/Q data was used as a complex dataset and an RF-based deep complex residual network model that effectively extracts the I/Q-related information in the electromagnetic signal waveform was proposed. The authors showed that the recognition accuracy of the proposed method was 99.56%, compared with the contour stellar-based method 90.4%, and the deep complex CNN-based method 94.8%. However, the complex-valued method is complicated and not widely generalized.

In our paper, the raw data directly sampled from the I channel and the Q channel and a new 2-dimensional $I + Q$ data are fed into a CNN model. The final identification result depends on the aggregated results of the three submodels. Specifically, if two or three predictive results of the three submodels are the same, the final result is determined by the two or three submodels. On the other hand, if all the results from the three submodels are different, the final result is determined purely by the result that is obtained from $I + Q$ data.

The main contributions of this paper are as follows:

(i) The raw sampled data from the I channel and the Q channel and a new 2-dimensional $I + Q$ data are trained by a typical CNN structure like LeNet-5 [15]. The identification accuracies for 4 transmit devices by the three CNN submodel are compared under 0 dB, 5 dB, and 10 dB SNRs

(ii) The CNN submodel is designed to be lightweight. A small number (32 or 64) of sample points are used for training. Only two conventional layers are included in the CNN structure. The output size of each layer is decreased to a half of the input size. The training time for the 4 devices is about 30 seconds. This is useful to extract the RF fingerprint in real time

(iii) A voting scheme is proposed to synthesize the above three identifications, which further improves the identification accuracy

2. CNN-Based RF Fingerprint Identification Method

2.1. RF Fingerprint Identification and CNN Introduction. The RF fingerprints refer to the hardware features of wireless devices based on the signals transmitted by the devices. The hardware features for RF fingerprint identification attribute to the differences between wireless devices due to the tolerances of electronic components [16]. The CNN is a feedforward neural network that includes convolutional computation and has a deep structure. A CNN model can classify the categories of input information, which is also called "a translation-invariant artificial neural network." The relevant research began from the 1980s. With the introduction of deep learning theory and the improvement of computing equipment, CNN models have been rapidly developed and widely used in computer vision, natural language processing, and other fields.

2.2. CNN-Based RF Fingerprint Identification Method. The CNN-based RF fingerprint identification method is shown in Figure 1. A small slice is selected from the original raw I/Q data to reduce computation time. For each transmitter, 200 out of 10000 transmission records and 32 or 64 out of 8000 sample points are extracted for the following signal processing. The selected samples are then normalized to eliminate the amplitude differences due to the distances between the transmitter and receiver. The normalization method is illustrated in the following equations.

$$
\begin{aligned}
X_{Imn} &= \frac{N \bullet x_{Imn}}{\sqrt{\sum_{n=1}^{N} \left(x_{Imn}^2 + x_{Qmn}^2 \right)}}, \\
X_{Qmn} &= \frac{N \bullet x_{Qmn}}{\sqrt{\sum_{n=1}^{N} \left(x_{Imn}^2 + x_{Qmn}^2 \right)}},
\end{aligned}
\tag{1}
$$

where x_{Imn} is the sample data at the I channel of the nth sample point of the mth transmission, x_{Qmn} is the sample data at the Q channel of the nth sample point of the mth transmission, X_{Imn} is the normalized result of x_{Imn}, X_{Qmn} is the normalized result of x_{Qmn}, and N is the total sample points, which equals to 32 or 64.

Additive White Gaussian noise is then added to the normalized sample data by employing MATLAB function "awgn." For each case, 200 transmissions are equally split into a training dataset and a testing dataset. The training process is then carried out by CNN models. Three predicted results by the I channel data, Q channel data, and $I + Q$ channel data are obtained and denoted as pI, pQ, and pIQ, respectively. The final identification is voted by the three predicted results, which means that the final predicted device is the one receiving the most votes among the three predicted results. For example, if the predicted results of RF fingerprint identification by the I channel data and the Q channel data are both device 1, while the predicted result by the $I + Q$ channel data is device 2, that means that device 1 receives 2 notes, device 2 receives 1 note, and device 3 and device 4 receive 0 notes, so the voted result, denoted as pV, is device 1. If all the three predicted results are different from each other, the final identification is the same as the predicted result by the $I + Q$ channel data, because $I + Q$ channel data contains more information. For example, if the predicted result by the I channel data is device 1, the predicted result by the Q channel data is device 2 and the predicted result by the $I + Q$ channel data is device 3; then, the final identification is device 3. The voting scheme can be simply summarized as Figure 2.

A CNN structure like LeNet-5 is applied to carry out the RF fingerprint identification. The structure consists of 10 layers. The CNN model is designed to be lightweight. Only two conventional layers are included in the CNN structure. The output size of each layer is decreased to a half of the input size. The parameters of each layer are given as follows.
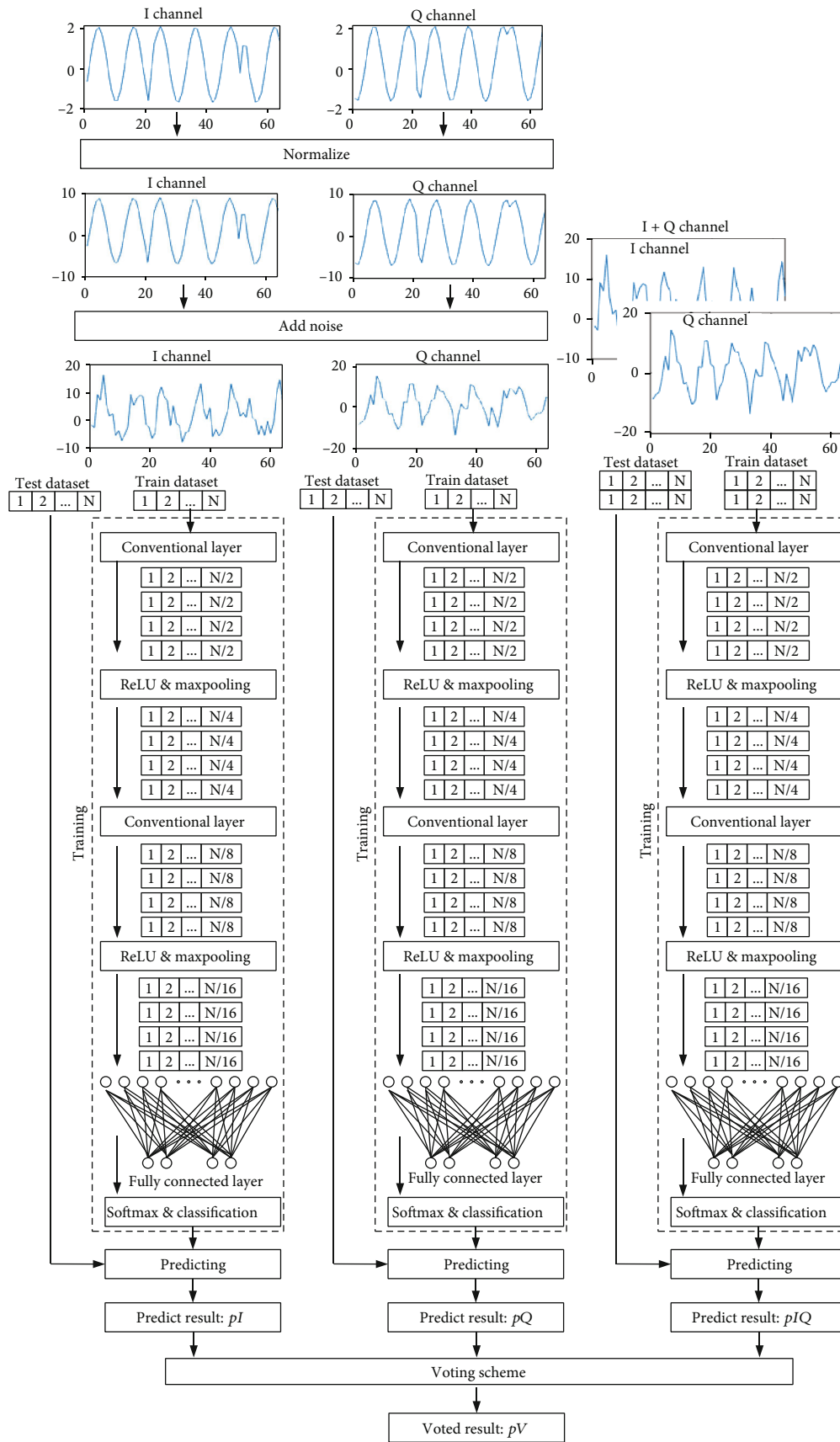
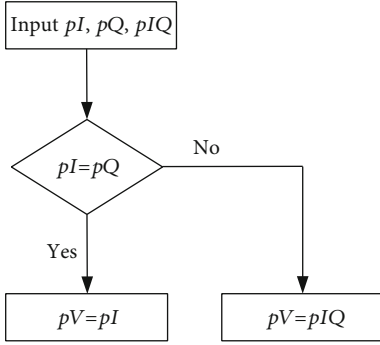FIGURE 1: A radio frequency fingerprint identification method based on CNN.

Figure 2: Algorithm flow of voting scheme.

The first layer is the data input layer. The size is $[L \times N \times 1]$, where $L = 1$ for the I channel data or the Q channel data, $L = 2$ for the $I + Q$ channel data, and N is the number of sample points that equals 32 or 64, which denotes that a segment of 32 data points can represent one signal transmission. The second layer is a 2D convolutional layer which is the major building block in the CNN. The convolutional layer performs a convolutional operation that involves the multiplication of a set of weights with the input to simulate a traditional neural network. The filter size is $[L \times 2 \times 1]$. The number of filters is 4. The stride is set as 2. The output data size is $[4 \times N/2 \times 1]$. The third layer is a rectified linear unit (ReLU) layer that performs a threshold operation to each element of the input, where any value less than zero is set to zero. The fourth layer is a 2D max pooling layer that manipulates downsampling by dividing the input into rectangular pooling regions and computing the maximum of each region. The pool size is $[1 \times 2]$. The stride is also set as 2. The output data size is $[4 \times N/4 \times 1]$. The fifth layer is a 2D convolutional layer. The filter size is $[1 \times 2 \times 1]$. The number of filters is 16. The stride is still set as 2. The output data size is $[4 \times N/8 \times 1]$. The sixth layer is a ReLU layer. The seventh layer is a 2D max pooling layer. The pool size is $[1 \times 2]$. The stride is 2. The output data size is $[4 \times N/16 \times 1]$. The eighth layer is a fully connected layer that compiles the data extracted by previous layers to form the final output. The output size is flattened to be 4, corresponding to the classifications of 4 devices. The ninth layer is a softmax layer that turns arbitrary real values into probabilities. The tenth layer is a classification layer that computes the cross-entropy loss for classification. A flow chart of the CNN structure is shown in Figure 3.

The CNN training models are implemented by the MATLAB deep learning toolbox. The MATLAB program is run by the Intel Xeon CPU E5-2678 v3 @ 2.5 GHz of a Dell PowerEdge T430 computer with a RAM size of 128 GB.

3. Results and Discussion

An experimental hardware platform consisting of four transmitters and one receiver is built and shown in Figure 4. Three low-cost USRP RIO-1082 (universal software radio peripheral-radio reconfigurable input/output) devices are used. Each USRP RIO-1082 device contains 2 radio fre-

quency transmitting channels and 2 radio frequency receiving channels. The four transmitters are comprised of two USRP RIO-1082 devices and four antennas. The receiver is comprised of a USRP RIO-1082 device and an antenna. An NI-PXIe 1085 device is used to collect and store the signal data. The QPSK-modulated signals are transmitted with the carrier frequency of 2 GHz and symbol rate 10 kbps. The sample rate of the receiver is 1 MHz.

During the experiment, each transmitter separately transmitted 10000 times. Each transmission was sampled at the I channel and Q channel simultaneously. After an synchronization, 8000 sample points of each transmission were stored. Therefore, we obtained four $10000 \times 8000 \times 2$ arrays of raw I/Q data corresponding to the four transmitters.

The influences of the sample size and SNR on the accuracy of RF fingerprint identification have been observed. Figure 5 shows the training progresses of the CNN with the I channel data, Q channel data, and $I + Q$ channel data under 0 dB SNR and 32 sample points. After 1400 iterations, the accuracies of the I channel data and $I + Q$ channel data are both about 90%, while the accuracy of the Q channel data is about 80%.

Figure 6 shows the training progresses of the CNN with the I channel data, Q channel data, and $I + Q$ channel data under 5 dB SNR and 32 sample points. After 1400 iterations, the accuracies of the I channel data and Q channel data are both about 95%, while the accuracy of the $I + Q$ channel data is about 97%.

Figure 7 shows the training progresses of the CNN with the I channel data, Q channel data, and $I + Q$ channel data under 10 dB SNR and 32 sample points. The accuracy of the I channel is about 95% after 1400 iterations, while the accuracies of the Q channel data and $I + Q$ channel are nearly 100% after 1000 iterations, which means that the CNN model is overfitting during the late stages of training.

Figure 8 shows the training progresses of the CNN with the I channel data, Q channel data, and $I + Q$ channel data under 0 dB SNR and 64 sample points. After 1400 iterations, the accuracies of the I channel data, Q channel data, and $I + Q$ channel data are about 97%, 95%, and 99%, respectively.

Figure 9 shows the training progresses of the CNN with the I channel data, Q channel data, and $I + Q$ channel data under 5 dB SNR and 64 sample points. The accuracy of the I channel is about 95% after 1400 iterations, while the accuracies of the Q channel data and $I + Q$ channel are nearly 100% after 1000 iterations, which means that the CNN model is overfitting during the late stages of training.

Figure 10 shows the training progresses of the CNN with the I channel data, Q channel data, and $I + Q$ channel data under 10 dB SNR and 64 sample points. The accuracies of the three CNN models are all near 100% after 600 iterations, which means that the CNN model is overfitting during the late stages of training.

Table 1 shows the elapsed time of the training progress when the I channel data, Q channel data, and the $I + Q$ channel data are input into the CNN model under 0 dB, 5 dB, and 10 dB SNRs, respectively, and 32 and 64 sample
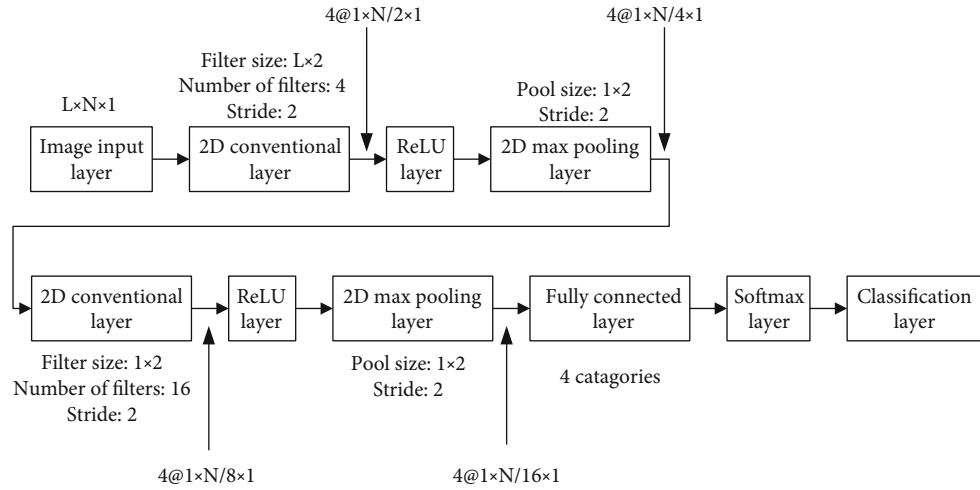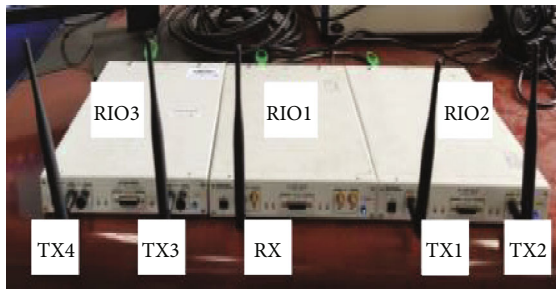
FIGURE 3: A flow chart of the CNN structure.

FIGURE 4: Experimental setup: four transmitters and one receiver.
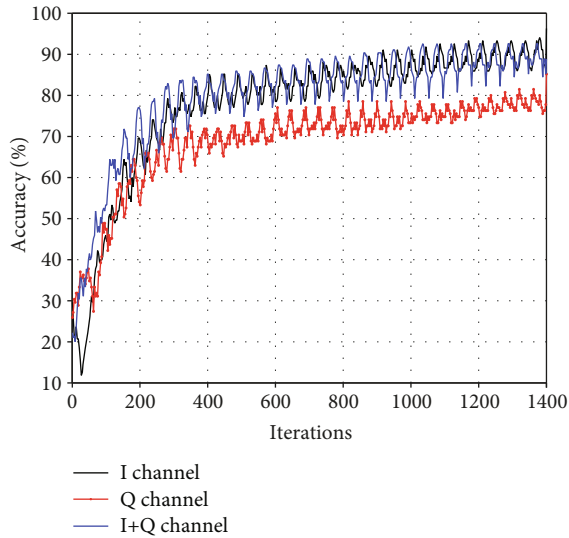

FIGURE 5: The training progresses of the CNN under 0 dB SNR and 32 sample points.

FIGURE 6: The training progresses of the CNN under 5 dB SNR and 32 sample points.

points. The elapsed time is between 24 s and 35 s, which is capable to extract the RF fingerprint in real time.

Table 2 shows the model accuracy rates of the submodel based on the I channel data, the submodel based on the Q channel, the submodel based on the $I + Q$ dat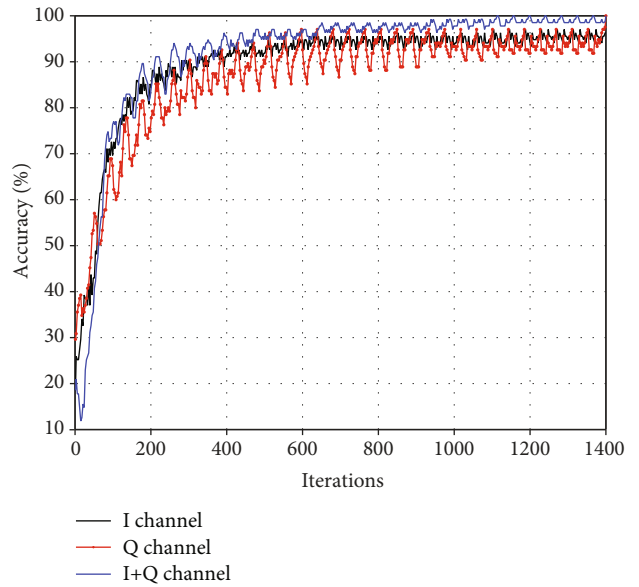a, and the final aggregated model. Generally, it can be seen that the identification accuracy of the final model is higher than that of all the other three ones. Under the same SNR, the model accuracy with 64 sample points is higher than with 32 sample points. When the sample points are set as 32, the model accuracy rates are 87.25%, 93.75%, and 98.25% under SNR 0 dB, 5 dB, and 10 dB, respectively. When the sample points are set as 64, the model accuracy rates are 93%, 97.25%, and 99.5% under SNR 0 dB, 5 dB, and 10 dB, respectively.

Tables 3–8 show the confusion matrices of the voted results for the four device identification problems in consideration of different sample sizes and SNRs. The sum of every row in the confusion matrices is 100, which is the total number of running epochs. The diagonal numbers are the numbers of iterations for correct predictions, while the other numbers are the numbers of iterations for wrong predictions.
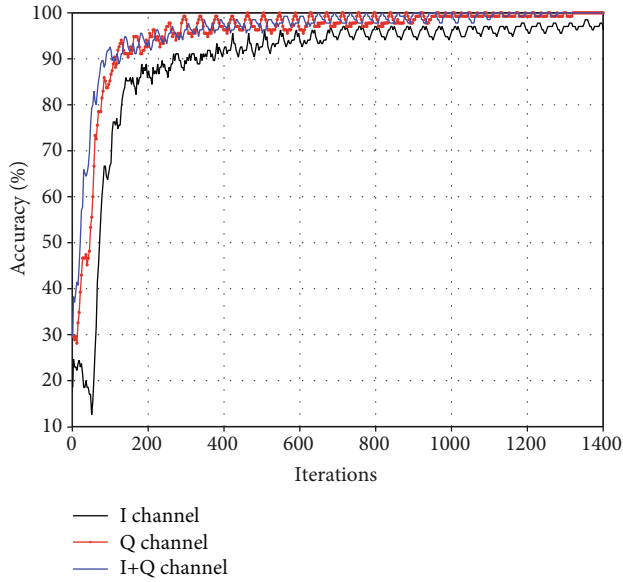
Figure 7: The training progresses of the CNN under 10 dB SNR and 32 sample points.
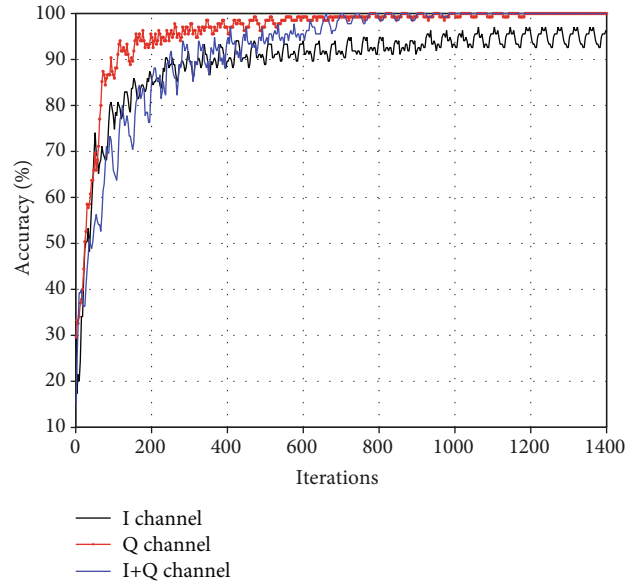

Figure 9: The training progresses of the CNN under 5 dB SNR and 64 sample points.
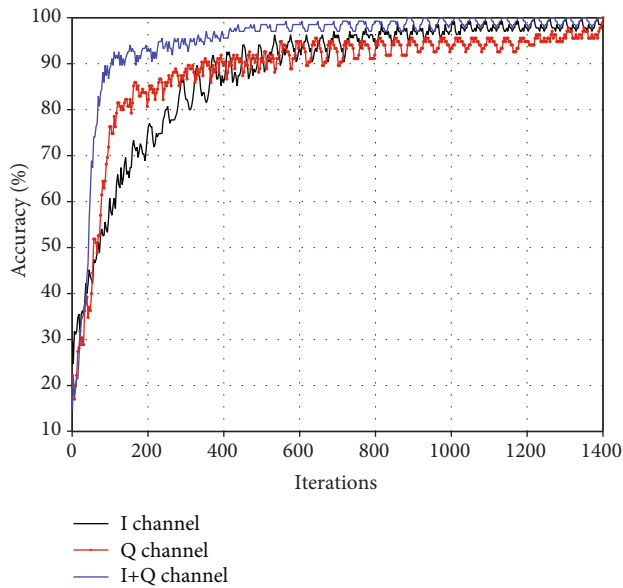

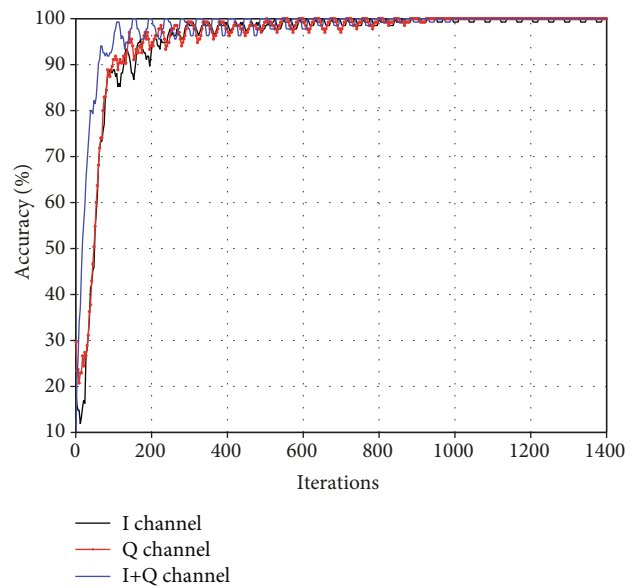Figure 8: The training progresses of the CNN under 0 dB SNR and 64 sample points.

Figure 10: The training progresses of the CNN under 10 dB SNR and 64 sample points.

Table 3 shows the confusion matrix when the number of sample points is 32 and the SNR is 0 dB. Out of 100 epochs, device 1 is predicted to be device 1 for 89 times, device 2 for one time, and device 3 for 10 times. Device 2 is predicted to be device 1 for one time, device 2 for 98 times, and device 4 for one time. Device 3 is predicted to be device 1 for 5 times, device 2 for 2 times, device 3 for 74 times, and device 4 for 19 times. Device 4 is predicted to be device 3 for 12 times and device 4 for 88 times.

Table 4 shows the confusion matrix when the number of sample points is 32 and the SNR is 5 dB. Out of 100 epochs, device 1 is predicted to be device 1 for 98 times and device 3

for 2 times. Device 2 is predicted to be device 1 for one time and device 2 for 99 times. Device 3 is predicted to be device 1 for 3 times, device 3 for 81 times, and device 4 for 16 times. Device 4 is predicted to be device 3 for 3 times and device 4 for 97 times.

Table 5 shows the confusion matrix when the number of sample points is 32 and the SNR is 10 dB. Device 1, device 2, and device 4 are correctly predicted for all 100 times. Out of 100 epochs, device 3 is predicted to be device 1 for one time, device 3 for 93 times, and device 4 for 6 times.

Table 6 shows the confusion matrix when the number of sample points is 64 and the SNR is 0 dB. Out of 100 epochs,

TABLE 1: Elapsed time of the training progress.

	$N = 32$, SNR = 0 dB	$N = 32$, SNR = 5 dB	$N = 32$, SNR = 10 dB	$N = 64$, SNR = 0 dB	$N = 64$, SNR = 5 dB	$N = 64$, SNR = 10 dB
I channel	28 s	27 s	33 s	30s	33 s	26 s
Q channel	29 s	25 s	30s	29 s	34 s	26 s
$I + Q$ channel	27 s	25 s	33 s	28 s	35 s	24 s

TABLE 2: Accuracy of RF fingerprint identification under different conditions.

	$N = 32$, SNR = 0 dB	$N = 32$, SNR = 5 dB	$N = 32$, SNR = 10 dB	$N = 64$, SNR = 0 dB	$N = 64$, SNR = 5 dB	$N = 64$, SSNR = 10 dB
I channel	0.79	0.8675	0.965	0.8	0.945	0.985
Q channel	0.7725	0.9225	0.965	0.8025	0.95	0.98
$I + Q$ channel	0.7875	0.93	0.965	0.92	0.945	0.9825
Voted results	0.8725	0.9375	0.9825	0.93	0.9725	0.995

TABLE 3: Confusion matrix of the voted RF fingerprint identification. $N = 32$, SNR = 0 dB.

		Predicted devices			
		1	2	3	4
Actual devices	1	89	1	10	0
	2	1	98	0	1
	3	5	2	74	19
	4	0	0	12	88

TABLE 6: Confusion matrix of the voted RF fingerprint identification. $N = 64$, SNR = 0 dB.

		Predicted devices			
		1	2	3	4
Actual devices	1	94	0	6	0
	2	1	98	0	1
	3	9	0	85	6
	4	0	0	5	95

TABLE 4: Confusion matrix of the voted RF fingerprint identification. $N = 32$, SNR = 5 dB.

		Predicted devices			
		1	2	3	4
Actual devices	1	98	0	2	0
	2	1	99	0	0
	3	3	0	81	16
	4	0	0	3	97

TABLE 7: Confusion matrix of the voted RF fingerprint identification. $N = 64$, SNR = 5 dB.

		Predicted devices			
		1	2	3	4
Actual devices	1	99	0	1	0
	2	0	100	0	0
	3	4	0	93	3
	4	0	0	3	97

TABLE 5: Confusion matrix of the voted RF fingerprint identification. $N = 32$, SNR = 10 dB.

		Predicted devices			
		1	2	3	4
Actual devices	1	100	0	0	0
	2	0	100	0	0
	3	1	0	93	6
	4	0	0	0	100

TABLE 8: Confusion matrix of the voted RF fingerprint identification. $N = 64$, SNR = 10 dB.

		Predicted devices			
		1	2	3	4
Actual devices	1	100	0	0	0
	2	0	100	0	0
	3	2	0	98	0
	4	0	0	0	100

device 1 is predicted to be device 1 for 94 times and device 3 for 6 times. Device 2 is predicted to be device 1 for one time, device 2 for 98 times, and device 4 for one time. Device 3 is predicted to be device 1 for 9 times, device 3 for 85 times, and device 4 for 6 times. Device 4 is predicted to be device 3 for 5 times and device 4 for 95 times.

Table 7 shows the confusion matrix when the number of sample points is 64 and the SNR is 5 dB. Out of 100 epochs,

device 1 is predicted to be device 1 for 99 times and device 3 for 1 time. Device 2 is correctly predicted for all 100 times. Device 3 is predicted to be device 1 for 4 times, device 3 for 93 times, and device 4 for 3 times. Device 4 is predicted to be device 3 for 3 times and device 4 for 97 times.

Table 8 shows that the confusion matrix when the number of sample points is 64 and the SNR is 10 dB. Device 1, device 2, and device 4 are correctly predicted for all 100 times. Out of 100 epochs, device 3 is predicted to be device 1 for 2 times and device 3 for 98 times.

4. Conclusions

The CNN-based RF identification models on the I channel-related data, the Q channel-related data, and the $I + Q$ -related data are used to recognize four transmitters. The final identification labels are determined by a voting scheme of the three submodels. It has been shown that the accuracy of the final model is higher than those of all the three submodels. Additionally, experimental results illustrate that the recognition accuracy with 64 sample points is higher than that of 32 sample points in the same settings of SNRs. The model accuracy is more than 90% under 32 sample points in the setting of SNR at 5 dB. The elapsed time of training the CNN model is between 24 s and 35 s, which proves that the model is capable to extract the RF fingerprint in real time.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] G. Shen, J. Zhang, A. Marshall, and J. R. Cavallaro, "Towards scalable and channel-robust radio frequency fingerprint identification for LoRa," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 774–787, 2022.

[2] A. Al-Shawabka, F. Restuccia, S. D'Oro et al., "Exposing the fingerprint: dissecting the impact of the wireless channel on radio fingerprinting," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, pp. 646–655, Toronto, ON, Canada, 2020.

[3] K. Youssef, L. Bouchard, K. Haigh, J. Silovsky, B. Thapa, and C. V. Valk, "Machine learning approach to RF transmitter identification," *IEEE Journal of Radio Frequency Identification*, vol. 2, no. 4, pp. 197–205, 2018.

[4] A. Jagannath, J. Jagannath, and P. S. P. V. Kumar, "A comprehensive survey on radio frequency (RF) fingerprinting-traditional approaches, deep learning, and open challenges," 2022, http://arxiv.org/abs/2201.00680.

[5] M. Liu, Z. Liu, L. Weidang, Y. Chen, X. Gao, and 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, 2022.

[6] M. Liu, B. Li, Y. Chen et al., "Location parameter estimation of moving aerial target in space–air–ground-integrated networks-based IoV," *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 5696–5707, 2022.

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

[8] S. Hu, P. Wang, Y. Peng et al., "Machine learning for RF fingerprinting extraction and identification of soft defined radio devices," in *Artificial Intelligence in China*, Lecture Notes in Electrical Engineering, Springer, Singapore, 2020.

[9] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," *IEEE Transactions On Neural Networks And Learning Systems*, vol. 23, no. 4, pp. 541–551, 2012.

[10] Q. Wu, C. Feres, D. Kuzmenko et al., "Deep learning based RF fingerprinting for device identification and wireless security," *Electronics Letters*, vol. 54, no. 24, pp. 1405–1407, 2018.

[11] W. Lee, S. Y. Baek, and S. H. Kim, "Deep-learning-aided RF fingerprinting for NFC security," *IEEE Communications Magazine*, vol. 59, no. 5, pp. 96–101, 2021.

[12] K. Merchant, S. Revay, G. Stantchev, and B. Nousain, "Deep learning for RF device fingerprinting in cognitive communication networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 160–167, 2018.

[13] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, 2018.

[14] S. Wang, H. Jiang, X. Fang, Y. Ying, J. Li, and B. Zhang, "Radio frequency fingerprint identification based on deep complex residual network," *IEEE Access*, vol. 8, pp. 204417–204424, 2020.

[15] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[16] V. R. Pagán and T. E. Murphy, "Electrooptic millimeter-wave downconversion and vector demodulation using phase-modulation and optical filtering," in *2014 IEEE Avionics, Fiber-Optics and Photonics Technology Conference (AVFOP)*, Atlanta, GA, USA, Nov.2014.