Specific Emitter Identification Based on Deep Residual Networks

YIWEI PAN [ID], **SIHAN YANG, HUA PENG, (Member, IEEE), TIANYUN LI, AND WENYA WANG**

National Digital Switching System Engineering and Technological Research Center, Zhengzhou 450002, China

Corresponding author: Yiwei Pan (novakd@163.com)

ABSTRACT Specific emitter identification (SEI) enables the discrimination of individual radio emitters with the external features carried by the received waveforms. This identification technique has been widely adopted in military and civil applications. However, many previous methods based on hand-crafted features are subject to the present expertise. To remedy these shortcomings, this paper presents a novel SEI algorithm using deep learning architecture. First, we perform Hilbert-Huang transform on the received signal and convert the resulting Hilbert spectrum into a grayscale image. As a signal representation, the Hilbert spectrum image has high information integrity and can provide abundant information about the nonlinear and non-stationary characteristics of signals for identifying emitters. Thereafter, we construct a deep residual network for learning the visual differences reflected in the Hilbert spectrum images. By using the residual architectures, we effectively address the degradation problem, which improves efficiency and generalization. From our analysis, the proposed approach combines high information integrity with low complexity, which outperforms previous studies in the literature. The simulation results validate that the Hilbert spectrum image is a successful signal representation, and also demonstrate that the fingerprints extracted from raw images using deep learning are more effective and robust than the expert ones. Furthermore, our method has the capability of adapting to signals collected under various conditions.

INDEX TERMS Deep residual network, Hilbert spectrum grayscale image, information integrity, Rayleigh fading, relay, specific emitter identification.

I. INTRODUCTION

Specific emitter identification (SEI) is a technique to identify individual radio emitters by extracting external features from a given signal [1]. The external features, namely radio frequency (RF) fingerprinting, are primarily the result of hardware variability in the device's analog components. Given that RF fingerprinting is unique for each emitter and difficult to counterfeit, the SEI technique plays an increasingly important role in military and civil fields, such as battlefield spectrum management [2], and wireless network security [3].

The key of SEI is a set of features that renders identification possible. In prior open research, features may be predefined or inferred [4]. Predefined features relate to the well-understood signal characteristics known in advance prior to signal recording. For example, Brik creatively proposes a passive radiometric device identification system (PARADIS) that differentiates 138 wireless devices with accuracy of

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

over 99% [5]. The characteristics are imperfections in the modulation domain, such as frequency offset, I/Q origin offset, error vector magnitude, and magnitude and phase errors. Based on PARADIS, several imperfections reflected in the modulation domain are further utilized, thereby achieving superior identification performance [6], [7]. Additionally, Liu identifies wireless devices by estimating the nonlinearity of the power amplifier (PA) [8]. The nonlinear PA model is described by a Taylor series expansion, and the power-series coefficients derived by observing the spectral regrowth are used as the RF fingerprints. In [9], the characteristics of the phase noise originating from the RF oscillator imperfections are used as the unique device tags.

Differently from the predefined features, we say that features are inferred when they are extracted from signals by means of some spectral transformations. Numerous signal characteristics have been investigated using fast Fourier transform, including power spectrum density [10] and spectral correlation function [11]. However, the performances of these methods are far from satisfactory because the subtle

differences between emitters are difficult to manifest in the power spectrum. Therefore, other complicated transformations have attracted the attentions of scholars. In [12], features are extracted using discrete wavelet transform to identify the individual RF emitters. Another common approach to RF fingerprinting involves developing the bispectrum characteristics. Square integral bispectra provide useful identification features [13]. Han further improves identification accuracy by utilizing the distribution of the bispectrum phase [14]. Furthermore, methods based on Hilbert-Huang transform (HHT) have been demonstrated for successful RF fingerprinting [15]–[17]. Specifically, the EM^2 algorithm uses energy entropy and color moments as identification fingerprints [15], [16], which are extracted from the Hilbert spectrum of the received signal. These features can differentiate emitters by evaluating the uniformity of the Hilbert spectrum. Consequently, several different spectral features, including spectral flatness, spectral brightness, and spectral roll-off, are further explored for improving identification accuracy [17].

However, predefined and inferred features are largely subject to the present expertise. The former is limited to prior knowledge of the types and characteristics of signals, whereas the latter relies on the existing tools of signal processing.

In recent years, deep learning has achieved a series of breakthroughs in machine vision and speech recognition. Motivated by such success, scholars have used deep neural networks in modulation recognition [18], [19] and radar waveform recognition [20], [21]. As the differences between the individual emitters are subtle, SEI using deep learning is still in a nascent stage. A convolutional neural network (CNN) operating on the time-domain complex baseband error signal has achieved 92.29% identification accuracy on a set of seven commercial ZigBee devices [22]. It should be noted that the architecture of CNN is more suitable for extracting features from images than waveforms. This implies that the method in [22] cannot maximize the powerful self-learning capabilities of CNN, which degrades identification accuracy. Another pioneer has demonstrated that CNNs trained on a compressed bispectrum considerably outperform conventional methods based on handcrafted features [23]. However, we notice that identification performance using the bispectrum is far from satisfactory [15], [16], [23]. Therefore, using a bispectrum image as the data representation of the received signal is not an ideal choice. In addition, dimensionality reduction may lead to the loss of important details while removing the redundant information.

To remedy these shortcomings, this paper presents a novel SEI approach using deep learning architecture. First, we perform the HHT on the received signal and convert the resulting Hilbert spectrum into a grayscale image. Thereafter, we construct a deep residual network for learning the visual differences reflected in the Hilbert spectrum images in order to discern the devices.

The main contributions are summarized as follows:
1) The Hilbert spectrum image is used as the data representation of the received signal, which lays the foundation for RF fingerprinting extraction using deep learning architectures. To the best of our knowledge, this attempt is the first time that the Hilbert spectrum image is adopted as a signal representation.
2) We apply a deep residual network to the RF fingerprinting problem. Deep residual network, which effectively addresses the degradation problem, can ease the training of deep networks.
3) We further investigate the performance of our approach under fading channels in single-hop and relaying scenarios. Simulation results demonstrate that the proposed approach is robust to complex communication systems and wireless channels.

The remainder of this paper is organized as follows. Section II shows a concise description of the system model. Section III introduces the novel deep learning methodology for SEI, and provides the performance and complexity analysis. Section IV presents a series of simulations and discusses the experimental results. Finally, Section V concludes the paper.

As a general convention, we use the following notations throughout the paper: real part of complex number, $\Re\{\cdot\}$; maximum value, $\max\{\cdot\}$; round toward negative infinity, $\lfloor\cdot\rfloor$; and mathematical expectation, $E\{\cdot\}$.

II. SYSTEM MODEL
A. NONLINEAR MODEL
RF fingerprinting is possible due to hardware imperfections in the analog circuitry introduced at the manufacturing process. Here, we concentrate our attention on the PA, which is the final stage prior to transmission.

As a memoryless nonlinear system, PA may cause distortions of the frequency and amplitude characteristics, including the amplitude/amplitude and amplitude/phase conversions [24]. In general, the system response of PA is described in a Taylor polynomial form. Let $s_0(t)$ represent the baseband signal and f_c be the carrier frequency. With the input $s(t) = \Re\left\{s_0(t)e^{j2\pi f_c t}\right\}$, the output signal of PA can be given as
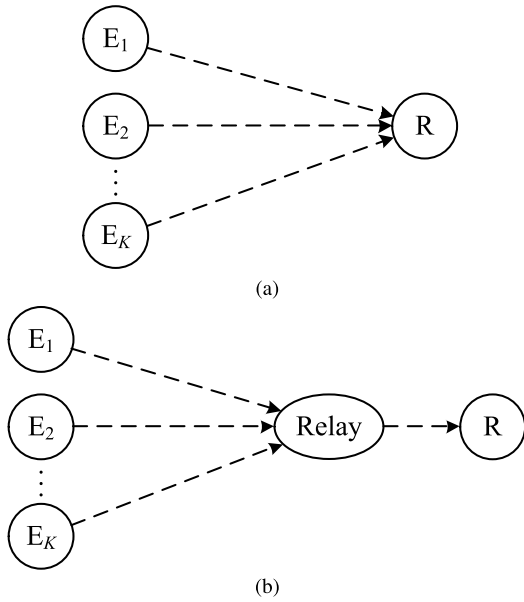
$$x(t) = \sum_{i=1}^{R} \lambda_i (s(t))^i, \qquad (1)$$

where R is the Taylor polynomial order and $\{\lambda_1, \lambda_2, \cdots, \lambda_R\}$ denotes the nonlinearity coefficients, which can be regarded as the RF fingerprinting. As a convention, $\lambda_1 = 1$.

B. SINGLE-HOP SCENARIO
Fig. 1(a) illustrates the typical communication system of the single-hop scenario. The received signal transmitted by emitter k can be expressed as

$$r(t) = \alpha \cdot x(t) + \nu(t)$$
$$= \alpha \sum_{i=1}^{Re} \lambda_{k,i}^e (s(t))^i + \nu(t), \qquad (2)$$

FIGURE 1. Typical communication systems: (a) single-hop, and (b) relaying.

where α is the fading coefficient, Re indicates the Taylor polynomial order, $\lambda_{k,i}^e$ represents the i-th nonlinearity coefficient of the k-th emitter, and $v(t)$ denotes the white Gaussian noise process corrupting the signal. When $\alpha = 1$, it refers to the additive white Gaussian noise (AWGN) channel, whereas when α is Rayleigh-distributed, it is the Rayleigh fading channel.

C. RELAYING SCENARIO

Now, we consider the relaying scenario shown in Fig. 1(b). The signal is amplified and retransmitted at the relay before being collected by the receiver. Hence, for specific emitter k, the received signal at the relay can be given by

$$r'(t) = \alpha \sum_{i=1}^{Re} \lambda_{k,i}^e (s(t))^i + v(t), \qquad (3)$$

where $\{\lambda_{k,1}^e, \lambda_{k,2}^e, \cdots, \lambda_{k,Re}^e\}$ are the nonlinearity coefficients of emitter k, α indicates the fading coefficient, and $v(t)$ represents the noise of the channel from the emitters to the relay. Similarly, the received signal at the receiver can be written as

$$\begin{aligned} r(t) &= \beta \cdot r'(t) + \upsilon(t) \\ &= \beta \sum_{m=1}^{Rr} \lambda_m^r \left(\alpha \sum_{n=1}^{Re} \lambda_{k,n}^e (s(t))^n + v(t) \right)^m + \upsilon(t), \quad (4) \end{aligned}$$

where $\{\lambda_1^r, \lambda_2^r, \cdots, \lambda_{Rr}^r\}$ refer to the nonlinearity coefficients of the relay, Rr denotes its Taylor polynomial order, β is the fading coefficient, and $\upsilon(t)$ indicates the channel noise from the relay to the receiver. The fingerprinting of emitters, carried by $r(t)$ in (4), is contaminated by the relay characteristics, thereby increasing the difficulty of discerning the individual emitters.

III. METHOD

In essence, SEI is a classification problem. The key to address the problem lies in the effectiveness and reliability of the features used. Previous studies based on hand-crafted features have been subject to the present expertise. Even experts can hardly understand and extract complex high-dimensional features, but such capability is the advantage of deep learning.

A. SIGNAL REPRESENTATION

Inspired by the excellent work of [15], we represent the signal data using HHT, which is a powerful method for analyzing nonlinear and non-stationary data [25]. The key technique is empirical mode decomposition (EMD), with which complicated data can be decomposed into a finite number of intrinsic mode functions (IMFs) that admit well-behaved Hilbert transforms. With the Hilbert transform, the IMFs yield instantaneous frequencies as functions of time that give the results as an energy-frequency-time distribution.

EMD is like a sifting process. Local maxima and minima are identified to produce the upper and lower envelopes by fitting. Designate the mean of envelopes as m_1, and the the difference between the data $r(t)$ and m_1 is the first component, i.e. $h_1(t) = r(t) - m_1$. Then, the second component is $h_{11}(t) = h_1(t) - m_{11}$, where $h_1(t)$ is treated as the data. We repeat this sifting process M times until $h_{1M}(t)$ is an IMF. The process can be stopped by limiting the size of the standard deviation $\varepsilon = \sum_{t=0}^T \frac{|h_{1(M-1)}(t) - h_{1M}(t)|^2}{h_{1(M-1)}^2(t)}$. Let $c_1(t) \triangleq h_{1M}(t)$ be the first IMF. Once $c_1(t)$ is obtained, we can separate $c_1(t)$ from the rest of the data by

$$r(t) - c_1(t) = e_1(t), \qquad (5)$$

Since the residue $e_1(t)$ still contains IMF components, the procedure can be repeated as

$$e_1(t) - c_2(t) = e_2(t), \cdots, e_{N-1}(t) - c_N(t) = e_N(t), \quad (6)$$

The procedure is stopped by the predetermined criterion [25]. By summing up (5) and (6), the received signal is decomposed as

$$r(t) = \sum_{i=1}^N c_i(t) + e_N(t), \qquad (7)$$

where N represents the number of IMF components, and $e_N(t)$ is the final residue.

By performing the Hilbert transform on each IMF component, we can express the signal in the following form

$$r(t) = \Re \left\{ \sum_{i=1}^N a_i(t) e^{j \int \omega_i(t) \, dt} \right\}, \qquad (8)$$

where $a_i(t) = \sqrt{c_i^2(t) + \hat{c}_i^2(t)}$ and $\omega_i(t) = \frac{d\theta_i(t)}{dt}$ indicate the instantaneous amplitude and frequency of $c_i(t)$, respectively. $\theta_i(t) = \arctan \frac{c_i(t)}{c_i(t)}$, $\hat{c}_i(t)$ is the Hilbert transform of $c_i(t)$. Equation (8) enables us to represent $a_i(t)$ and $\omega_i(t)$ as functions of time in a three-dimensional plot. The frequency-time distribution of the amplitude is designated as the Hilbert spectrum $H(\omega, t)$.
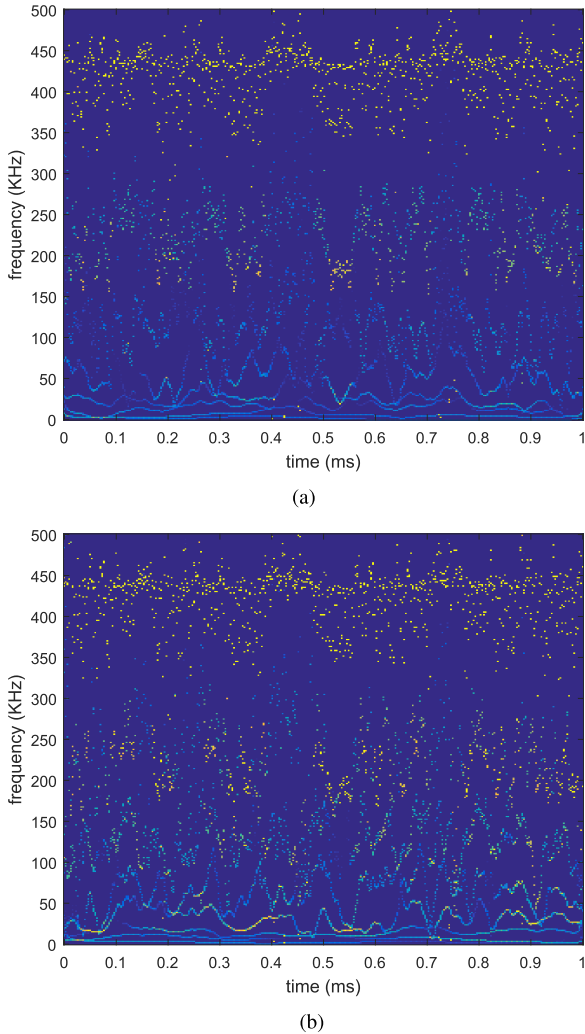
FIGURE 2. Hilbert spectrum images for different emitters with PA nonlinearity. The emitter in (b) has further serious nonlinear distortion.

To represent the signal data, we convert the Hilbert spectrum into a grayscale image. Let $H_{i,j}$ be the (i, j)-th amplitude value of $H(\omega, t)$ and $G_{i,j}$ be the (i, j)-th pixel value of the greyscale image. Hence, the ζ-bit Hilbert spectrum image can written as

$$G_{i,j} = \left\lfloor \left(2^{\zeta} - 1\right) \frac{H_{i,j}}{\max\left\{H_{i,j}\right\}} \right\rfloor, \tag{9}$$

where $\lfloor \cdot \rfloor$ stands for the round toward negative infinity. The Hilbert spectrum image is mapped to an interval of $\left[0, 2^{\zeta} - 1\right]$.

To form some intuitions, Fig. 2 visualizes the Hilbert spectrum images for different emitters with PA nonlinearity as described in Section II, where the emitter in Fig. 2(b) has further serious nonlinear distortion. As is seen, the energy of the received signal is mainly concentrated in the high-frequency part (i.e., carrier frequency). In the intermediate-/low-frequency parts, abundant frequency components are present, which may provide the discriminative features

for identification. Furthermore, the higher the nonlinearity of the PA, the more the frequency components that leak into the intermediate-/low-frequency parts, which leads to more chaotic distribution in the Hilbert spectrum.

As mentioned, the algorithms proposed in [15]–[17] extract the features from the Hilbert spectrum. It is worth noticing that feature extraction is a process of information loss. Each feature in [15]–[17] only describes the characteristics of the Hilbert spectrum from a certain perspective. Due to limited knowledge, it is difficult to provide enough features to obtain a panoramic view of the Hilbert spectrum. This leads to amount of information being lost inevitably. Hence, we turn to deep learning.

B. NETWORK ARCHITECTURE

Deep neural networks naturally integrate low-/mid-/high-level features, which promises to overcome the limitations of expert features. Recent evidence reveals the importance of network depth. However, deeper neural networks are more difficult to train. One obstacle is the notorious problem of vanishing/exploding gradients [26], which hampers convergence from the beginning.
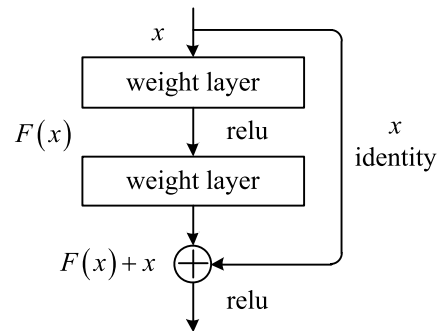
FIGURE 3. Residual unit.

To address the degradation problem, we apply the deep residual network [27] to extract the RF fingerprinting. The residual unit with a shortcut connection shown in Fig. 3 is heavily used in a deep residual network, which allows the features to operate flexibly at multiple scales and depths through the network. Denoting the desired underlying mapping as $H(x)$, we let the stacked nonlinear layers fit another mapping of $F(x) = H(x) - x$. The original mapping is recast into $F(x) + x$. As the residual mapping is easier to optimize than the original one, the residual unit leads to considerable improvements in efficiency and generalization.

Here, we feed the Hilbert spectrum images into the deep neural network to distinguish individual emitters. The input images have a unified size of 300×300. We do not perform any expert feature extraction, thereby allowing the neural network to learn overall features directly from the raw images. Fig. 4 depicts the structure of the used deep residual network, in which four residual units are utilized. In practice, the number of residual units is selected through simulations (see Section IV-B). The parametric rectified
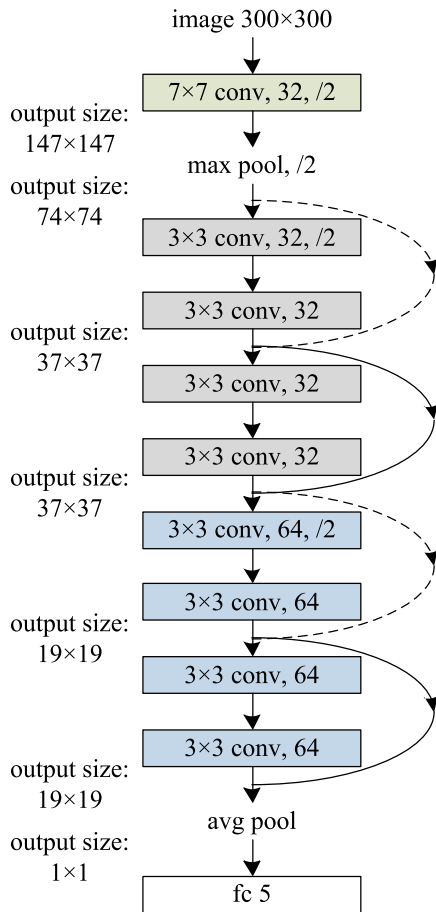
image 300×300

output size:
147×147

7×7 conv, 32, /2

output size:
74×74

max pool, /2

3×3 conv, 32, /2

3×3 conv, 32

output size:
37×37

3×3 conv, 32

3×3 conv, 32

output size:
37×37

3×3 conv, 64, /2

3×3 conv, 64

output size:
19×19

3×3 conv, 64

3×3 conv, 64

output size:
19×19

avg pool

output size:
1×1

fc 5

FIGURE 4. Layout of the used deep residual network.

linear unit [28] is selected as the activation function for all applicable layers except for the output layer where Softmax is utilized. We adopt batch normalization [29] right after each convolution and before activation. The network has a total of 167,781 trainable parameters.

C. PERFORMANCE AND COMPLEXITY ANALYSIS

Now, we analyze the identification performance of the proposed approach in comparison with the EM^2 algorithm [15] and the CNN-based one [23]. As we mentioned earlier, information carried by the received signal is corrupted or lost during feature extraction. Indeed, the integrity of information determines the identification performance. Here, we provide the performance analysis from a perspective of information integrity.

Using (5) and (6), the received signal is decomposed as a sum of several IMFs $c_i(t)$ and the final residue $e_N(t)$. It is apparent that the decomposition in (7) does not corrupt any information. From (8), the spectrum $H(\omega, t)$ is derived by performing the Hilbert transform on each IMF component. Due to the reversibility of the transform, information carried by the received signal is completely preserved in $H(\omega, t)$. Note that, with an appropriate ζ, error introduced by image mapping in (9) is negligible. Therefore, the image fed into

the network has considerably high information integrity. This lays the foundation for deep learning, thereby guaranteeing excellent identification performance.

Unlike the proposed approach, three hand-crafted features are extracted from the Hilbert spectrum image in the EM^2 algorithm. Suppose that $G_{i,j}$ represents the (i, j)-th pixel value of the greyscale image, N_r and N_c are the number of rows and columns, respectively. The energy entropy, mean and standard deviation are given by

$$I = -\sum_{i=1}^{N_r}\sum_{j=1}^{N_c} G_{i,j} \log G_{i,j}, \tag{10}$$

$$\mu = \frac{1}{N_r \cdot N_c}\sum_{i=1}^{N_r}\sum_{j=1}^{N_c} G_{i,j}, \tag{11}$$

$$\sigma = \sqrt{\frac{1}{N_r \cdot N_c}\sum_{i=1}^{N_r}\sum_{j=1}^{N_c} \left(G_{i,j} - \mu\right)^2}, \tag{12}$$

It is easily checked that these fingerprints can reflect some differences of the individual emitters. Nevertheless, more characteristics beyond our knowledge remain undiscovered. Clearly, present expertise limits the identification performance.

For the CNN-based algorithm, compressed bispectrum image is adopted as the signal representation. Assuming that received signal is divided into P segments, with Q samples per segment. Then, the estimated bispectrum is expressed as

$$\hat{B}(\omega_1, \omega_2) = \sum_{\tau_1=-\delta}^{\delta}\sum_{\tau_2=-\delta}^{\delta} \hat{c}(\tau_1, \tau_2)\omega(\tau_1, \tau_2) e^{-j(\omega_1\tau_1+\omega_2\tau_2)}, \tag{13}$$

where $\delta < Q - 1$, $\hat{c}(\tau_1, \tau_2)$ is the estimated third-order cumulant, $\omega(\tau_1, \tau_2)$ is the hexagonal window function. As a statistic, $c(\tau_1, \tau_2)$ can retain only part of the information carried by received signal. This means that bispectrum is not a successful data representation. Unfortunately, dimensionality reduction leads to a loss of important details, which further worsens the identification performance. It is clear that the CNN is powerless when the feeding images have poor information integrity.

Next, we concentrate on the computational complexity. As the training can be offline, we take no consideration of its cost. The computational load of the proposed algorithm involves with Hilbert spectrum estimation and greyscale image mapping. As the signal decomposition in (9) is empirical, the complexity of the first part can be roughly expressed as $O(MNL + MN \log L + NL \log L)$ [15], where M refers to the sifting number to obtain an IMF, N represents the number of IMFs (no more than 10, generally) and L is the signal length. Generally speaking, M and N are small numbers. The cost of the second part is linear with the Hilbert spectrum size $S_H = N_r \cdot N_c$, i.e. $O(S_H)$. Hence, the computational complexity of the proposed approach is

TABLE 1. Computational complexity of the algorithms.

Algorithms	Complexity
Proposed algorithm	$O\left(MNL + MN \log L + NL \log L + S_H\right)$
EM^2 algorithm	$O\left(MNL + MN \log L + NL \log L + 7S_H\right)$
CNN-based algorithm	$O\left(PQ \log Q + 2Q^2 + S_B^2\right)$

$O\left(MNL + MN \log L + NL \log L + S_H\right)$. Differently from the proposed method, the EM^2 algorithm performs artificial feature extraction on the Hilbert spectrum image. This leads to an additional complexity of $O(6S_H)$. For the CNN-based algorithm, the main computational load comes from bispectrum estimation and dimensionality reduction, whose complexities are $O\left(PQ \log Q + 2Q^2\right)$ and $O\left(S_B^2\right)$, respectively. (S_B is the bispectrum size.) It is worth stressing that the complexity of the CNN-based algorithm is approximately quadratic. This may become a burden provided that $S_B = S_H$. For the sake of clarity, Table 1 presents the comparison of the computational complexity.

IV. RESULTS

In this section, we investigate the performance of the proposed method through simulations. To assess the signal-to-noise ratio (SNR) under different channels, we define the average SNR as

$$\bar{\gamma} = E_s/N_0 \cdot E\left\{\alpha^2\right\}, \tag{14}$$

where E_s/N_0 refers to the ratio of symbol energy to noise power spectral density, and $E\left\{\alpha^2\right\}$ is simply the average value of α^2. The case of Rayleigh fading channel corresponds to when α has a Rayleigh distribution, whereas $\alpha = 1$ characterizes the AWGN channel.

In the following simulations, $K = 5$ radio emitters are to be identified. All devices operate at the 420 KHz center frequency, and the received signals are digitized at a sampling rate of 1 MHz. Specifically, QPSK modulation, 200 Kbps data rate, and the shaping filter is the raised cosine filter with a roll-off factor of 0.35. Signals are generated with the impairment parameters [15], [16] shown in Table 2.

TABLE 2. Impairment parameters of the emitters.

| Parameters	E$_1$	E$_2$	E$_3$	E$_4$	E$_5$	Realy
λ_2	0.5	0.08	0.01	0.01	0.6	0.1
λ_3	0.3	0.6	0.01	0.4	0.08	0.1

For each emitter, $N = 5000$ segments per SNR and $M = 100$ symbols per segment are present. Each segment is converted to one 8-bit Hilbert spectrum image with a resolution of 300×300 pixels. The dataset is randomly partitioned into 60% training data and 40% testing data. We train the network for 6000 epochs each with a batch size of 300, and each batch is randomly shuffled at each epoch during training. The Glorot uniform initialization [30] is used for the kernel initialization of all convolutional and dense layers.

Categorical cross-entropy is used to compute the loss. The Adam optimizer with a learning rate $\alpha = 0.05$ is utilized to ensure convergence [31].

Four NVIDIA Titan Xp GPUs are used to train and test the networks. All data pre-processing is performed by MATLAB R2015a. The deep residual networks are implemented in Python with TensorFlow 1.3.0 as the backend.

A. IDENTIFICATION PERFORMANCE BY SHORTCUT CONNECTION

We first investigate the effect of shortcut connection on identification performance. Experiments are conducted to compare the residual network shown in Fig. 4 with the plain one. Here, the plain network has the same number of parameters, depth, and width, except for the shortcut connection. In this simulation, signals are collected under the AWGN channel in the single-hop scenario.
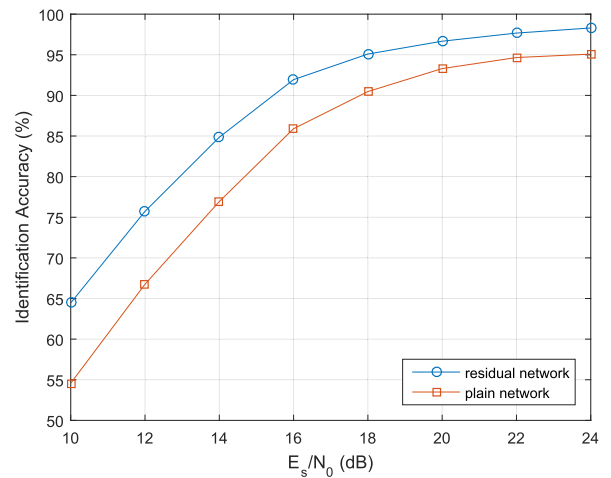
FIGURE 5. Identification performance affected by shortcut connection.

Fig. 5 depicts the identification performance affected by the shortcut connection. As is seen, the residual network with the shortcut connection outperforms the plain network without the shortcut connection. Much information is being lost when data pass through the plain network. The residual network solves this problem to some extent by directly passing input information to the output, which protects information integrity. As the shortcut connection introduces neither extra parameter nor computation complexity, it is attractive for us to use the residual network with the shortcut connection in the following simulations.

B. IDENTIFICATION PERFORMANCE BY DEPTH

The number of residual units related to depth may considerably affect the capability to extract complex features accurately. Here, we consider the influence of the number of residual units on identification performance. In this simulation, the number of residual units used in the network is set to 2, 4, 6, 8, and 10, respectively. Table 3 gives the structures of the different networks. Here, signals are collected under the AWGN channel in the single-hop scenario.

TABLE 3. Layouts of the networks, along with the number of parameters and training time spent on a single batch.

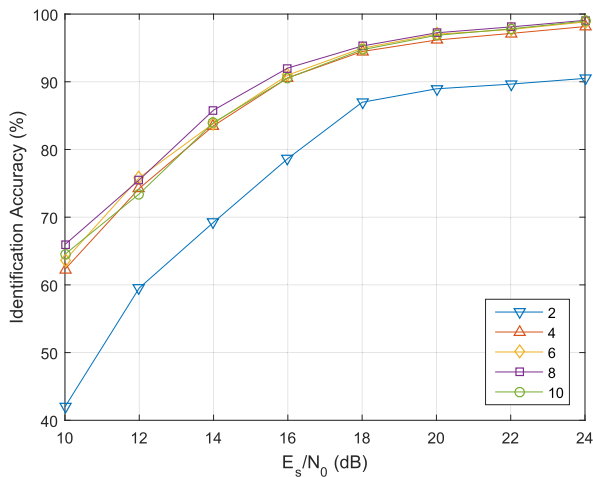| Layer	Output Size	2	4	6	8	10
conv1	147×147			7×7, 32, stride2		
max pool	74×74			3×3, stride2		
conv2_x	37×37	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$
conv3_x	19×19	—	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv4_x	10×10	—	—	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv5_x	5×5	—	—	—	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv6_x	3×3	—	—	—	—	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
avg pool	1×1			5-d fc, softmax		
parameters		3.9×10^4	1.7×10^5	6.8×10^5	2.7×10^6	1.1×10^7
time (s)		0.4891	0.5516	0.5900	0.6173	0.6596

FIGURE 6. Identification performance affected by depth.

FIGURE 7. Identification performances in the single-hop scenario under the AWGN channel.

Fig. 6 presents the identification performance affected by the number of residual units. It is apparent that the network with two residual units achieves the lowest accuracy. This implies that the network is not deep enough to extract complex RF fingerprints from the images. When the number of residual units is not less than four, no substantial gain on identification accuracy occurs with the increase in depth. However, as shown in the bottom of Table 3, the number of parameters and training time spent on a single batch increase with the deepening network. Hence, we use four residual units in the following simulations for simplicity.

C. IDENTIFICATION PERFORMANCE UNDER THE AWGN CHANNEL

In this subsection, we examine the performance of the algorithm proposed in Section III (denoted as HilSpec-ResNet)
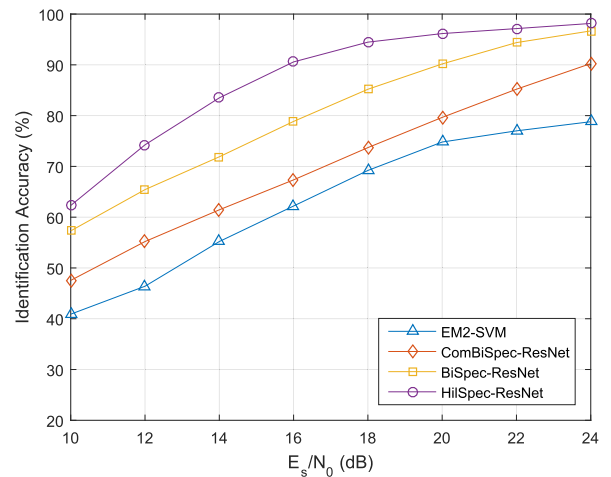
under the AWGN channel. Experiments are performed for comparison with the three other algorithms: (1) ComBiSpec-ResNet representing the signal with the compressed bispectrum proposed in [23], which uses the residual network introduced in Section III-B; (2) BiSpec-ResNet representing the signal with the uncompressed bispectrum image, which also uses the residual network introduced in Section III-B; and (3) EM^2-SVM, which involves the conventional algorithm EM^2 [15] by using a support vector machine for classification.

To begin, we consider the single-hop scenario. Fig. 7 illustrates the identification performances in the single-hop scenario under the AWGN channel. Clearly, the proposed algorithm achieves a remarkable improvement on identification accuracy compared with the EM^2 algorithm.

This validates the foregoing discussion that deep learning can avoid the limitations of hand-crafted features and extract RF fingerprinting from raw images. The proposed algorithm also outperforms BiSpec-ResNet, thereby corroborating the Hilbert spectrum image is a successful signal representation with high information integrity. This is why the HHT-based algorithms are superior to those based on bispectrum, as shown in [15], [16], [23]. By comparing BiSpec-ResNet with ComBiSpec-ResNet, an evident gap in identification accuracy is revealed. This is expected, because dimensionality reduction causes loss of important details while removing the redundant information.

Similarly, we consider the relaying scenario. The masked RF fingerprints are further difficult to discover because signals are amplified and retransmitted at the relay. As shown in Fig. 8, the performances of all four algorithms are affected by the relay to some degree. However, the proposed algorithm remains superior to the other three. This demonstrates the effectiveness and robustness of the proposed algorithm in the relaying scenario.
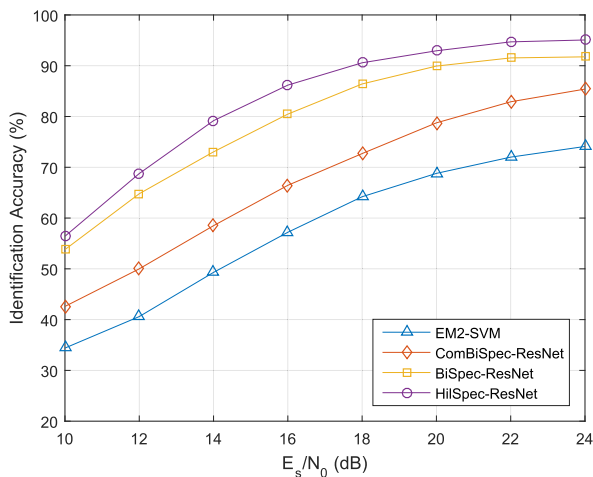

FIGURE 8. Identification performances in the relaying scenario under the AWGN channel.

Here, there are $K = 5$ radio emitters to be identified. With a large set of devices, the performances of all SEI algorithms will be degraded, inevitably. Within the reasonable range of PA nonlinearity, the larger set of devices, the smaller difference between emitters. This increases the difficulty of discerning the devices. Nevertheless, the deep learning architecture will have less performance degradation compared with the hand-crafted features, because the fingerprints extracted by deep neural network will be more intrinsic and more resilient.

D. IDENTIFICATION PERFORMANCE UNDER THE RAYLEIGH FADING CHANNEL

In this subsection, we further examine the performance of the proposed algorithm under the Rayleigh fading channel. The residual network is trained with the signals collected under
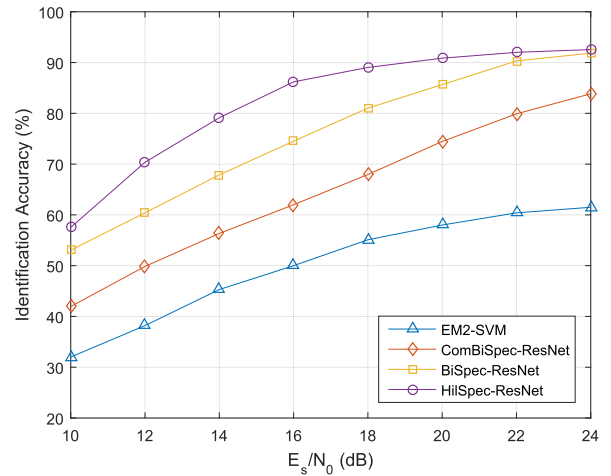

FIGURE 9. Identification performance in the single-hop scenario under the Rayleigh fading channel.
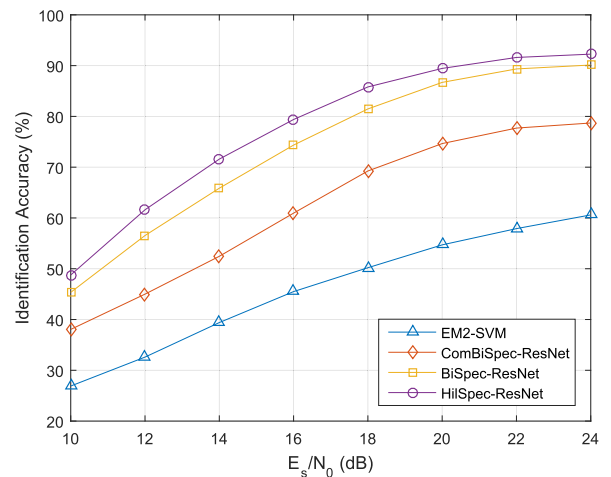

FIGURE 10. Identification performance in the relaying scenario under the Rayleigh fading channel.

such channel. Figs. 9 and 10 present the identification performances in single-hop and relaying scenarios, respectively.

Comparing the results in Figs. 7 and 8 with those in Figs. 9 and 10, we can easily observe that a severe penalty in SNR is paid as a consequence of the fading characteristics of the received signal. Despite the penalty, the proposed algorithm has substantially less loss in identification accuracy than EM2-SVM. The improved robustness further validates the advantages of deep learning. Furthermore, the proposed algorithm outperforms BiSpec-ResNet and ComBiSpec-ResNet in Figs. 7–10, which demonstrates that the Hilbert spectrum image is a better signal representation than the bispectrum image used in [23].

E. IDENTIFICATION PERFORMANCE UNDER THE FREQUENCY-SELECTIVE CHANNEL

In this subsection, we examine the performance of the proposed algorithm under the frequency-selective channel. We apply an empirical multipath channel model, $h(t) = \delta(t) + 0.6\delta(t - T_0) + 0.3\delta(t - 2T_0)$, where T_0 is the symbol period. The residual network is trained with the signals
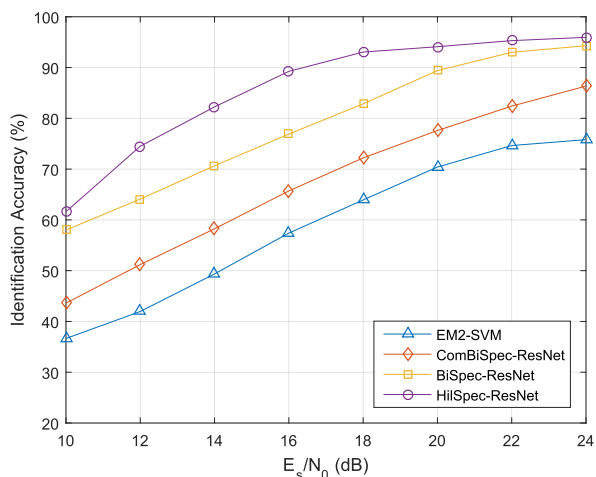
FIGURE 11. Identification performance in the single-hop scenario under the frequency-selective channel.
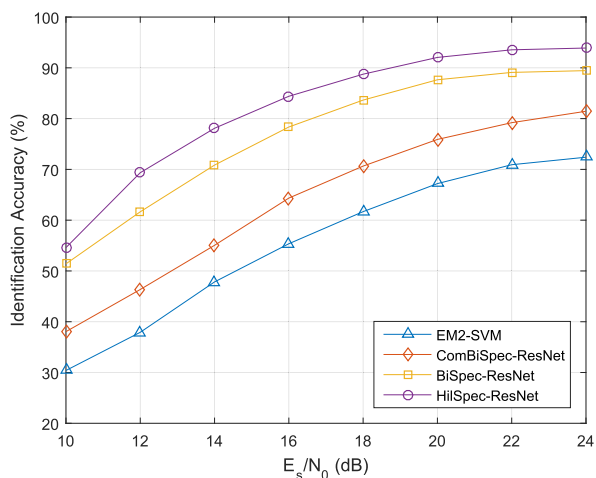

FIGURE 12. Identification performance in the relaying scenario under the frequency-selective channel.

collected under such channel. Figs. 11 and 12 present the identification performances in single-hop and relaying scenarios, respectively. As is seen, the proposed algorithm performs better than BiSpec-ResNet, ComBiSpec-ResNet, and EM²-SVM under the frequency-selective channel in both the single-hop and relaying scenarios. It is worth stressing that the performances under the multipath channel approximate those under the AWGN channel, even better than those under the Rayleigh fading channel. This implies that inter-symbol interference introduced by the multipath channel has a similar effect on the distribution in the spectrum images of different emitters; whereas the fading coefficient α of the Rayleigh channel leads to more randomness, which blurs the difference between emitters.

V. CONCLUSION

Many previous studies based on hand-crafted features are largely subject to the present expertise. In this paper, we propose a novel SEI algorithm based on the deep residual network. Our network operates on Hilbert spectrum grayscale images and enables learning the complicated

high-dimensionality features from raw images. The Hilbert spectrum image that can provide rich information is validated as a good choice of signal representation. The residual units with shortcut connection effectively address the degradation problem, thereby improving efficiency and generalization. Simulation results demonstrate that our approach outperforms those of previous studies. Furthermore, it can adapt to signals collected under various conditions. Our future work involves increasing the robustness of the network architectures to allow the extension of our approach to include data captured from a large set of devices.

REFERENCES

[1] K. I. Talbot, P. R. Duley, and M. H. Hyatt, "Specific emitter identification and verification," *Technol. Rev. J.*, pp. 113–133, Jan. 2003.
[2] A. E. Spezio, "Electronic warfare systems," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 3, pp. 633–644, Mar. 2002.
[3] O. Ureten and N. Serinken, "Wireless security through RF fingerprinting," *Can. J. Elect. Comput. Eng.*, vol. 32, no. 1, pp. 27–33, May 2007.
[4] B. Danev, D. Zanetti, and S. Capkun, "On physical-layer identification of wireless devices," *ACM Comput. Surv.*, vol. 45, no. 1, Nov. 2012, Art. no. 6.
[5] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc. 14th ACM Int. Conf. Mobile Comput. Netw.*, Sep. 2008, pp. 116–127.
[6] A. Candore, O. Kocabas, and F. Koushanfar, "Robust stable radiometric fingerprinting for wireless devices," in *Proc. 2nd Int. Workshop Hardw.-Oriented Secur. Trust*, Jul. 2009, pp. 43–49.
[7] Y. Huang and H. Zheng, "Radio frequency fingerprinting based on the constellation errors," in *Proc. 18th Asia–Pacific Conf. Commun.*, Oct. 2012, pp. 900–905.
[8] M.-W. Liu and J. F. Doherty, "Specific emitter identification using nonlinear device estimation," in *Proc. IEEE Sarnoff Symp.*, Princeton, NJ, USA, Apr. 2008, pp. 1–5.
[9] A. C. Polak and D. L. Goeckel, "Wireless device identification based on RF oscillator imperfections," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2492–2501, Dec. 2015.
[10] K. A. Remley *et al.*, "Electromagnetic signatures of WLAN cards and network security," in *Proc. 5th Int. Symp. Signal Process. Inf. Technol.*, Dec. 2005, pp. 484–488.
[11] K. Kim, C. M. Spooner, I. Akbar, and J. H. Reed, "Specific emitter identification for cognitive radio with application to IEEE 802.11," in *Proc. 22nd IEEE Global Telecommun.*, Nov./Dec. 2008, pp. 1–5.
[12] C. Bertoncini, K. Rudd, B. Nousain, and M. Hinders, "Wavelet fingerprinting of radio-frequency identification (RFID) tags," *IEEE Trans. Ind. Electron.*, vol. 59, no. 12, pp. 4843–4850, Dec. 2012.
[13] S. Xu, B. Huang, L. Xu, and Z. Xu, "Radio transmitter classification using a new method of stray features analysis combined with PCA," in *Proc. 26th IEEE Mil. Commun. Conf.*, Oct. 2007, pp. 1–5.
[14] J. Han, T. Zhang, D. Ren, and X. Zheng, "Communication emitter identification based on distribution of bispectrum amplitude and phase," *IET Sci., Meas. Technol.*, vol. 11, no. 8, pp. 1104–1112, Aug. 2017.
[15] J. Zhang, F. Wang, Z. Zhong, and O. Dobre, "Novel Hilbert spectrum-based specific emitter identification for single-hop and relaying scenarios," in *Proc. 28th IEEE Global Commun. Conf.*, Dec. 2015, pp. 1–6.
[16] J. Zhang, F. Wang, O. A. Dobre, and Z. Zhong, "Specific emitter identification via Hilbert–Huang transform in single-hop and relaying scenarios," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 6, pp. 1192–1205, Jun. 2016.
[17] U. Satija, N. Trivedi, G. Biswal, and B. Ramkumar, "Specific emitter identification based on variational mode decomposition and spectral features in single hop and relaying scenarios," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 581–591, Mar. 2019.
[18] 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.
[19] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications," *IEEE Access*, vol. 6, pp. 18484–18501, 2018.

[20] C. Wang, J. Wang, and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," in *Proc. 42nd IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2017, pp. 2437–2441.

[21] Z. Zhou, G. Huang, H. Chen, and J. Gao, "Automatic radar waveform recognition based on deep convolutional denoising auto-encoders," *Circuits, Syst., Signal Process.*, vol. 37, no. 9, pp. 4034–4048, Jan. 2018.

[22] 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.

[23] L. Ding, S. Wang, F. Wang, and W. Zhang, "Specific emitter identification via convolutional neural networks," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2591–2594, Dec. 2018.

[24] P. B. Kenington, *High Linearity RF Amplifier Design*. Boston, MA, USA: Artech House, 2000, pp. 123–135.

[25] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math. Phys. Eng. Sci.*, vol. 454, pp. 903–995, Mar. 1998.

[26] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. 24th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5353–5360.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 25th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. 15th IEEE Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[29] S. Ioffe and C. Szegedy. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167

[30] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. of 13th Int. Conf. Artif. Intell. Statist.*, May 2010, pp. 249–256.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 2nd Int. Conf. Learn. Represent.*, San Diego, CA, USA, May 2015, pp. 1–15.
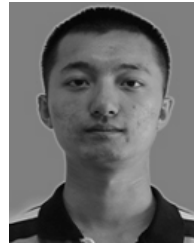
SIHAN YANG received the B.S. degree in information engineering from the Zhengzhou Institute of Information Science and Technology, in 2013. He is currently pursuing the M.S. degree in information and communication engineering with the National Digital Switching System Engineering and Technological Research Center. His main research interests include artificial intelligence and computer vision.

HUA PENG received the Ph.D. degree in signal processing technology from the Zhengzhou Institute of Information Science and Technology, in 2002. Since 2014, he has been a Professor with the National Digital Switching System Engineering and Technological Research Center. His research interests include wireless communication theory, signal processing, and cognitive radio.

TIANYUN LI received the Ph.D. degree in signal processing technology from the Zhengzhou Institute of Information Science and Technology, in 2012. He is currently an Associate Professor with the National Digital Switching System Engineering and Technological Research Center. His research interests include wireless communication theory, signal processing, and cognitive radio.

YIWEI PAN was born in Qingdao, China, in 1990. He received the B.S. degree in information engineering from the Zhengzhou Institute of Information Science and Technology, Zhengzhou, China, in 2013, and the M.S. degree in information and communication engineering from the National Digital Switching System Engineering and Technological Research Center (NDSC), Zhengzhou, China, in 2016. He is currently pursuing the Ph.D. degree in information and communication engineering with NDSC. His main research interests include signal processing, artificial intelligence, and specific emitter identification.

WENYA WANG received the B.S. degree from the Wuhan University of Technology, Wuhan, China, in 2013. She is currently pursuing the M.S. degree in information and communication engineering with the National Digital Switching System Engineering and Technological Research Center. Her main research interests include wireless communication theory, visible light communications, and signal processing.

• • •