Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 7646527, 17 pages
https://doi.org/10.1155/2020/7646527

Research Article

Specific Emitter Identification via Bispectrum-Radon Transform and Hybrid Deep Model

Yipeng Zhou,[1] **Xing Wang**,[1] **You Chen**,[1] **and Yuanrong Tian**[2]

[1]*Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038, China*
[2]*College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China*

Correspondence should be addressed to Yipeng Zhou; zhou_yipeng@163.com

Received 27 September 2019; Revised 10 December 2019; Accepted 19 December 2019; Published 13 January 2020

Academic Editor: Alessandro Lo Schiavo

Specific emitter identification is a technique that distinguishes different emitters using radio fingerprints. Feature extraction and classifier selection are critical factors affecting SEI performance. In this paper, we propose an SEI method using the Bispectrum-Radon transform (BRT) and a hybrid deep model. We propose BRT to characterize the unintentional modulation of pulses due to the superiority of bispectrum distributions in characterizing nonlinear features of signals. We then apply a hybrid deep model based on denoising autoencoders and a deep belief network to perform further deep feature extraction and discriminative identification. We design an automatic dependent surveillance-broadcast signal acquisition system to capture signals and to build dataset for validating our proposed SEI method. Theoretical analysis and experimental results show that the BRT feature outperformed traditional features in characterizing UMOP, and our proposed SEI method outperformed other feature and classifier combination methods.

1. Introduction

Specific emitter identification (SEI) is a technique that distinguishes different emitters using radio frequency fingerprints. The technique has attracted much attention in recent decades, meeting many application requirements in electronic intelligence, cognitive radio, wireless network security, and Internet of things. To realize SEI, there are two key problems to be solved: effective feature extraction and classifier selection.

The RF fingerprint, also known as the unintentional modulation on pulses (UMOPs), has characteristics of universality, uniqueness, and stability. Thus, extraction of unintentional modulation features from signals makes it possible to identify specific emitters. Various methods have been proposed to characterize UMOP [1–16]. These methods fall into categories of waveform-, transform-, and transmitter-based approaches. Waveform-based methods extract features from pulse waveforms directly, usually from time and frequency domains. Features obtainable with waveform-based methods include instantaneous frequency [1–3] and amplitude features [4, 5]. These features have the advantage of lower computational costs, but they are easily affected by environmental noise. They also perform poorly when analyzing non-Gaussian and nonstationary processes. For these reasons, researchers have proposed other transforms to extract features in the transform domain, including the ambiguity function [6], the wavelet transform [7], the Hilbert–Huang transform (HHT) [8–10], and short-time Fourier transform [11, 12]. These transforms are appropriate for analyzing nonlinear and nonstationary signals. However, they are time-consuming when used as input features for SEI due to their large dimensions. Furthermore, the choice of kernel and basis functions influences their performance considerably. Other researchers have explored modeling the imperfections and distortions of some transmitter components to realize SEI, including the imperfections and distortions of oscillators, receivers, and transmitters [13–16]. However, this kind of method offers only incomplete results because UMOP is a systematic result caused by all components of the transmitter, not just those chosen for modeling.

Since the theory of higher-order spectral analysis (HOSA) was developed in the late 1980s, research has shown its superiority in analyzing non-Gaussian and nonstationary

processes, such as modeling of nonlinear systems [17, 18] and signal classification [19, 20]. The higher-order spectrum refers to the multidimensional Fourier transform of higher-order cumulants. Compared with feature extraction methods in the time and frequency domains, the higher-order spectrum has three characteristics: (a) it contains amplitude and phase information simultaneously, (b) it includes the nonlinear component of signals, and (c) the higher-order cumulant of Gaussian noise equals to zero. Owing to these characteristics, it is possible to extract UMOP features using the higher-order spectrum.

Deep learning has garnered much attention because of its superiority for feature learning and feature representation. Deep learning stems from the simulation of the information processing process of human brain. Humans often describe the input information in hierarchical ways, with multiple levels of abstraction. The brain also appears to process information through multiple stages of transformation and representation [21]. To simulate the information processing process in human brain, deep learning methods extract abstract feature in hierarchical ways. Deep learning models are built using the basic modules including convolutional neural network (CNN), restricted Boltzmann machine (RBM), and autoencoder (AE) and its variants. By combining these simple but nonlinear modules, deep learning methods are able to transfer the feature representation at one level into a higher and more abstract level. By combining enough such transformations, very complex and nonlinear features can be learned [22]. According to the architectures and techniques intended for use, deep learning models can be categorized into three main classes: generative deep architectures, discriminative deep architectures, and hybrid deep architectures. Among the three architectures, the hybrid deep architectures stack several generative and discriminative modules, making them superior in some fields [23]. Hybrid deep models have been used extensively to perform image classification [24, 25] and video classification [26, 27].

Motivated by the preceding discussion, we propose a feature extraction method called the Bispectrum-Radon transform (BRT). Furthermore, we propose an SEI method based on the BRT and a hybrid deep model. In the proposed SEI method, the BRT extracts UMOP features while the hybrid deep learning model performs further feature learning and identification. We summarize the main contributions of this paper as follows:

(i) We propose a feature extraction method based on bispectrum theory and used it to characterize UMOP.

(ii) We propose an SEI method. The method extracts the UMOP features via BRT and uses a hybrid deep model to perform deep feature extraction and discriminative identification.

(iii) We design a signal acquisition system to receive the automatic dependent surveillance-broadcast (ADS-B) signals of civil aircraft and use the received signals to verify the identification performance of the proposed method.

(iv) The thorough experiments and analysis show the effectiveness and superiority of the proposed method.

The rest of this paper is organized as follows: in Section 2, we analyze the UMOP generation mechanism. We introduce the BRT in Section 3. We present the hybrid deep learning model and the proposed SEI method in Section 4. In Section 5, we present our experimental results and discussions. Section 6 concludes the paper.

2. Generation Mechanism of UMOP

The UMOP is caused by various components of the transmitter circuit, including oscillators, power amplifiers, and modulators. Figure 1 [28] shows the general structure of a radio transmitter.

As shown in Figure 1, before being transmitted by the antenna, the signal is processed by a series of electronic devices, including many nonlinear electronic devices, such as power amplifiers, intermediate frequency (IF) filters, and radio frequency (RF) filters. This process causes the unintentional generation of various nonlinear components that remain in the transmitted signals as UMOP.

UMOP has the characteristics of universality, uniqueness, and stability. First, UMOP is ubiquitous because differences in the performance of electronic devices and circuits are inevitable, even when for identical devices are operating under the same conditions. Meanwhile, the uniqueness of UMOP is attributable to differences in device manufacturing processes, component tolerances, material properties, and environmental factors [29]. Finally, UMOP is stable because these differences between devices and circuits are subtle but consistent over a considerable period. Owing to the characteristics mentioned above, UMOP is extremely useful in solving SEI problems, making the choice of features for characterizing UMOP a significant research focus for SEI.

3. Bispectrum-Radon Transform

In this section, we introduce our proposed BRT to characterize UMOP. We first introduce the bispectrum distribution and the Radon transform. We then present the BRT and the algorithm for extracting the bispectrum projection vectors via the BRT. Finally, we analyze the discrimination performance of the BRT feature.

3.1. Bispectrum Feature of UMOP. Based on HOSA theory, a bispectrum distribution is the two-dimensional Fourier transform of the third-order cumulants. It has the lowest order and the fewest calculations among higher-order spectrums. We express the bispectrum distribution of a pulse $x(t)$ as

$$B(w_1, w_2) = G_{3,x}(w_1, w_2) = \sum_{\tau_1} \sum_{\tau_2} C_{3x}(\tau_1, \tau_2) e^{-j(w_1\tau_1 + w_2\tau_2)}, \quad (1)$$

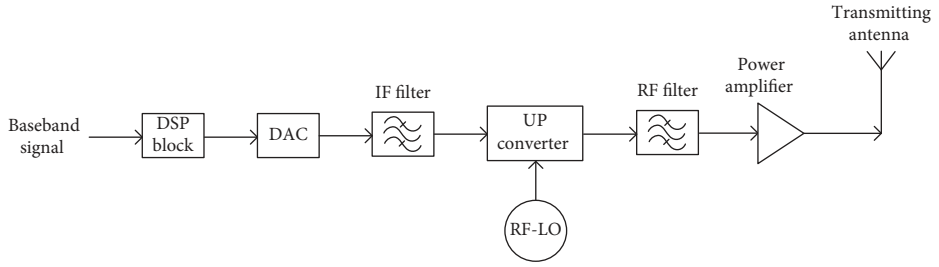

Figure 1: General structure of a radio transmitter.

where w_1 and w_2 are frequencies in the two-dimensional Fourier transform. $C_{3x}(\tau_1, \tau_2)$ is a third-order autocorrelation function calculated as

$$C_{3x}(\tau_1, \tau_2) = E\{x^*(t)x(t+\tau_1)x(t+\tau_2)\}. \quad (2)$$

Figure 2 shows bispectrum distributions of a theoretical pulse and a received ADS-B pulse. The subtle UMOP features are visible, demonstrating the effectiveness of the bispectrum distribution in describing UMOP.

3.2. Bispectrum-Radon Transform. As presented in Figure 2, the bispectrum distribution shows strong potential for characterizing UMOP. However, it is time-consuming to use two-dimensional bispectrum matrices to perform SEI directly. Hence, we introduce the Radon transform to project the two-dimensional bispectrum matrices into one-dimensional vectors.

In a two-dimensional Euclidean space, a matrix can be projected in a certain direction by applying the Radon transform. Figure 3 shows the geometrical interpretation.

For a line $f(w_1, w_2)$ in a plane (w_1, w_2), the Radon transform is given by

$$R(\theta, p) = \iint_D f(w_1, w_2)\delta(p - w_1\cos\theta - w_2\sin\theta)\mathrm{d}w_1\mathrm{d}w_2, \quad (3)$$

where θ is the given angle and p is the distance from the origin to the line $f(w_1, w_2)$. We define the delta function $\delta(x)$ as

$$\delta(x) = \begin{cases} 0, & x \neq 0, \\ 1, & x = 0. \end{cases} \quad (4)$$

Thus, we can express the delta function in equation (3) as

$$\delta(p - w_1\cos\theta - w_2\sin\theta) = \begin{cases} 0, & p - w_1\cos\theta - w_2\sin\theta \neq 0, \\ 1, & p - w_1\cos\theta - w_2\sin\theta = 0. \end{cases} \quad (5)$$

Therefore, we can understand the Radon transform as a projection process, where each point in $R(\theta, p)$ corresponds to a line in the plane.

Based on the bispectrum theory and the Radon transform, we propose a feature extraction method called Bispectrum-Radon transform (BRT). BRT projects the bispectrum matrix into a projection vector, $\mathbf{BR}(\theta)$, using the Radon transform. $\mathbf{BR}(\theta)$ is the linear integral of $\mathbf{B}(w_1, w_2)$ with θ, written as

$$\mathbf{BR}(\theta) = \iint_D \mathbf{B}_1(w_1, w_2)\delta(p - w_1\cos\theta - w_2\sin\theta)\mathrm{d}w_1\mathrm{d}w_2. \quad (6)$$

Thus, the process for extracting bispectrum projection vectors using the BRT decomposes into two processes: (a) the estimation of the bispectrum distribution and (b) the projection using the Radon transform. We use an indirect method for bispectrum estimation. Algorithm 1 presents the whole algorithm. Figure 4 shows the process of extracting bispectrum projection vectors, and Figure 5 shows the bispectrum projection vectors from actual ADS-B signals of three aircraft. Subtle and stable differences are clearly visible in Figure 5, proving the usefulness of the projection vectors for characterizing UMOP, distinguishing different emitters and accomplishing SEI.

4. Hybrid Deep Model

In this section, we present an SEI method with a hybrid deep model performing identification using the bispectrum projection vectors as input. We first introduce the basic modules of the hybrid model and their hierarchical feature learning mechanism. Then we present the proposed SEI method.

4.1. RBM and DBN. First proposed by Hinton et al. [30], DBN is a class of the generative model constructed from stacked restricted Boltzmann machines (RBMs). It learns hierarchical feature representations from complex input data. RBMs are probabilistic generative models able to learn a joint probability distribution of the training data [31]. They are good at extracting hidden and abstract features from training data. Le Roux and Bengio [32] have given the theoretical proof that RBMs are universal approximators of discrete distributions. The proof means that RBMs are excellent for modeling complex distributions.

As Figure 6 shows, an RBM consists of a visible layer and a hidden layer. The visible layer processes the input data, and the hidden layer detects the feature. There are full and undirected connections between visible and hidden units but none between pairs of visible or pairs of hidden units [33]. In Figure 6, the visible units $v = [v_1, v_2, \ldots, v_n]^T$ and the hidden units $h = [h_1, h_2, \ldots, h_m]^T$ are in binary states (i.e., for

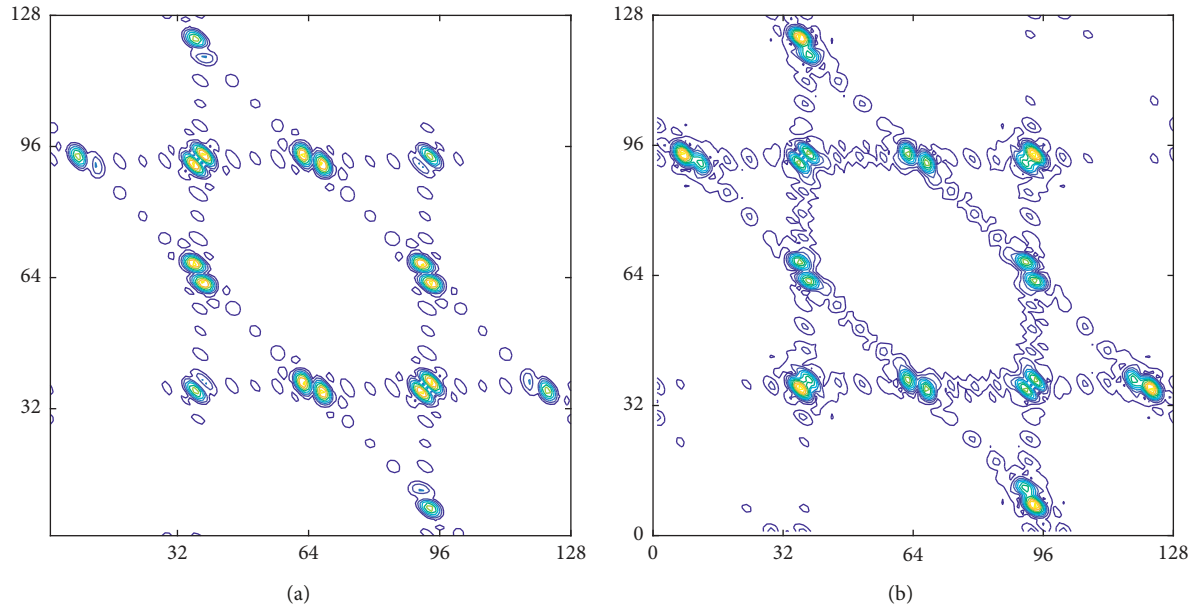

Figure 2: The bispectrum distributions of (a) a theoretical and (b) a received ADS-B pulse.

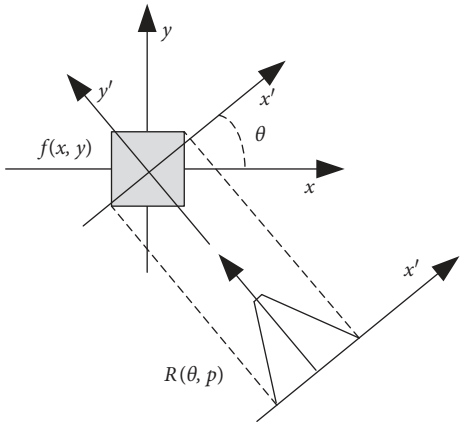

Figure 3: Geometrical interpretation of the Radon transform.

arbitraries i and j, $v_i \in \{0, 1\}$ and $h_j \in \{0, 1\}$). w_{ij} is the weight between unit v_i and h_j.

To describe the state of an RBM, the energy function of $E(v, h)$ is defined as

$$E(v,h) = -\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}v_i h_j - \sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j, \tag{7}$$

where a_i and b_j are the biases of units v_i and h_j, respectively. $E(v, h)$ is determined by the weights and biases under a given v and h. According to the given energy function, a joint probability distribution $p(v, h)$ can be obtained for all of the possible pairs of visible and hidden vectors:

$$p(v,h) = \frac{1}{Z}\exp(-E(v,h)) = \frac{\exp(-E(v,h))}{\sum_{v,h}\exp(-E(v,h))}, \tag{8}$$

where Z is a normalization term and $Z = \sum_{v,h}\exp(-E(v, h))$. The probability distribution $p(v)$ of visible layer v is the marginal distribution of $p(v, h)$. Thus, we can express $p(v)$ as

$$p(v) = \frac{1}{Z}\sum_{h}\exp(-E(v,h)). \tag{9}$$

As mentioned above, there are no visible-visible or hidden-hidden connections in RBM. Hence, when given a visible layer v, the state of each hidden unit is conditionally independent, and vice versa. Therefore, the conditional distributions of a visible unit and a hidden unit take the forms

$$p(h_j = 1 \mid v) = \sigma\left(b_j + \sum_i v_i w_{ij}\right), \tag{10}$$

$$p(v_i = 1 \mid h) = \sigma\left(a_i + \sum_j h_j w_{ij}\right), \tag{11}$$

where $\sigma = (1 + e^{-x})^{-1}$. Hinton [34] proposed a learning algorithm called the contrastive divergence (CD) algorithm to learn the optimal distribution for RBM. In the CD algorithm, the original state of the visible layer is set to an arbitrary training sample with the binary state of each hidden layer deduced by equation (10). After obtaining all of the states of hidden units, the probability of each visible unit can be deduced by equation (11). Therefore, a reconstructed visible layer is obtained, and a reconstruction error exists between the reconstructed visible layer and the initial input visible layer. The application of a stochastic gradient ascent algorithm optimizes the reconstruction error. With sufficient training, the reconstruction error gradually decreases and the RBM parameters stabilize.

As noted, a DBN is composed of several stacked RBMs. To effectively train a DBN, Hinton et al. [30] proposed a

input: S: signal with length N; M: the segment size of signal; K: the number of segments, $K = N/M$; θ: the projection angle in Radon transform

output: **BR**: the bispectrum projection vector

(1) Divide the signal S into K segments with M points in each segment;

(2) Let $x_k(0), x_k(1), \ldots, x_k(M-1)$ be the M points of k-th segment. Calculate the third-order cumulants of K segments $c_k(i,j) = 1/M \sum_{n=-M_1}^{M_2} x_k(n) x_k(n+i) x_k(n+j), k = 1, \ldots, K$,
where i and j are the indexes of the third-order cumulants.
$M_1 = \max(0, -i, -j)$ and $M_2 = \min(M-1, M-1-i, M-1-j)$;

(3) The average of the third-order cumulants of all segments, $\hat{c}(i,j)$, is calculated and is taken as the estimation of the third-order cumulant of the signal S
$\hat{c}(i,j) = 1/K \sum_{k=1}^{K} c_k(i,j)$

(4) Calculate the estimation of the bispectrum distribution
$\hat{B}(w_1, w_2) = \sum_{i=-L}^{L} \sum_{j=-L}^{L} \hat{c}(i,j) w(i,l) \exp[-j(w_1 i + w_2 l)]$
where $L < M-1$. $w(i,l)$ is the two-dimensional lag window function;

(5) Calculate the bispectrum projection vector $\mathbf{BR}(\theta)$ by conducting (5).

Algorithm 1: Extraction of bispectrum projection vector via the Bispectrum-Radon transform.

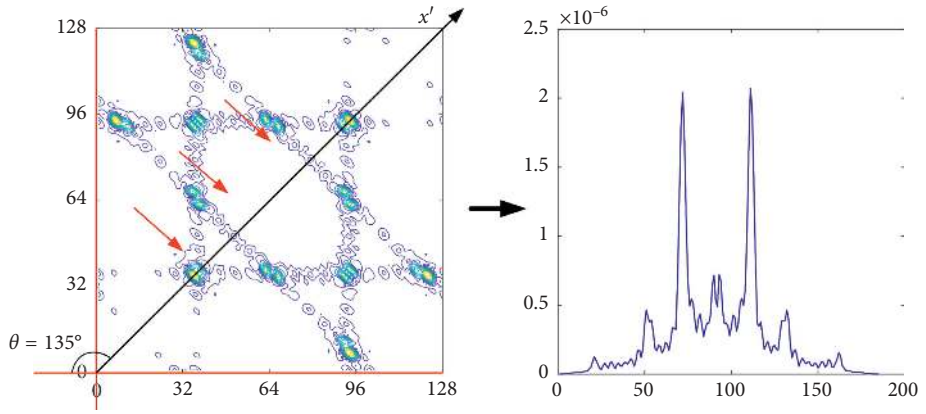

Figure 4: Process of the Bispectrum-Radon transform.

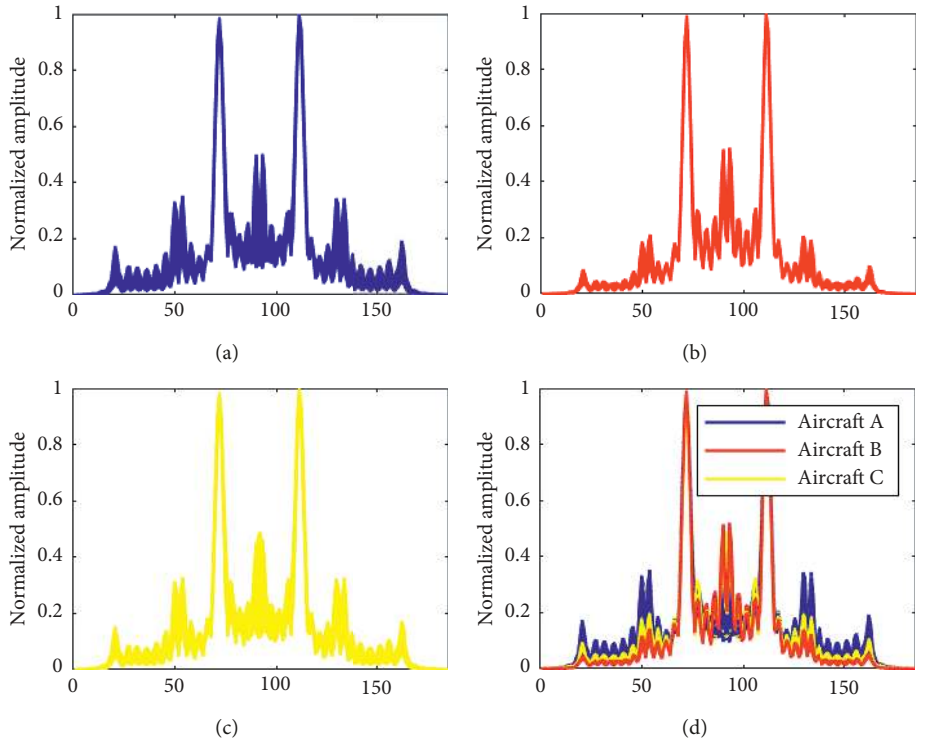

Figure 5: Bispectrum-Radon transform features of three aircraft. (a) Aircraft A. (b) Aircraft B. (c) Aircraft C. (d) Comparison of three aircraft.

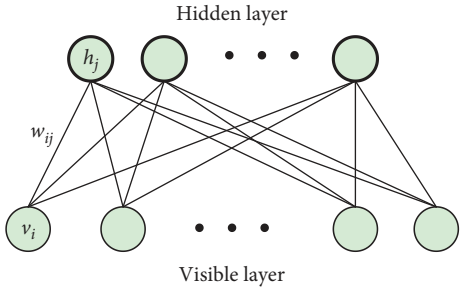

Figure 6: RBM architecture.

greedy layer-wise unsupervised learning method. The training process has two stages: pretraining and fine-tuning. In the first stage, the RBMs are trained from the bottom-up layer by layer. The CD algorithm is used to train and optimize the RBM parameters. In the second stage, the accumulated error in the layers after pretraining leaves the parameters less than optimal, so the backpropagation is then used to fine-tune the DBN parameters. The fine-tuning stage optimizes the entire DBN.

4.2. AE and DAE. Motivated by the training strategy of DBN, Bengio et al. [35] proposed replacing the RBM modules with autoencoders to build the stacked autoencoder (SAE) model. SAE is also a generative model, and it could be used for unsupervised feature learning. An AE consists of an encoder and a decoder that perform the encoding and decoding process separately. Figure 7 shows the basic architecture of an AE.

In the encoding process, the hidden layer effectively extracts the intrinsic expression of the input data. The encoding function can be described as

$$h = f(x) = s_f(Wx + p), \tag{12}$$

where x is the input data and h is the code obtained. W represents the weight matrix between the input layer and the hidden layer. p is an offset vector. $s_f(\,)$ is the activation function of the encoder. In this paper, we choose the sigmoid function as the activation function:

$$s_f(z) = \frac{1}{1 + e^{-z}}. \tag{13}$$

In the decoding process, the output layer reconstructs the input data based on the hidden layer. The decoding function can be described as

$$r = g(h) = s_g\left(W^T h + q\right), \tag{14}$$

where W^T represents the weight matrix between the hidden layer and the output layer. q is an offset vector. $s_g(\,)$ is the activation function of the decoder. Hence, the learning task of AE can be abstracted into a convex optimization problem that minimizes the reconstruction error between the reconstruction and the input data. We describe the optimization problem as

$$\min_{\theta} \left\| s_g\left(W^T s_f(Wx + p) + q\right) - x \right\|_2, \tag{15}$$

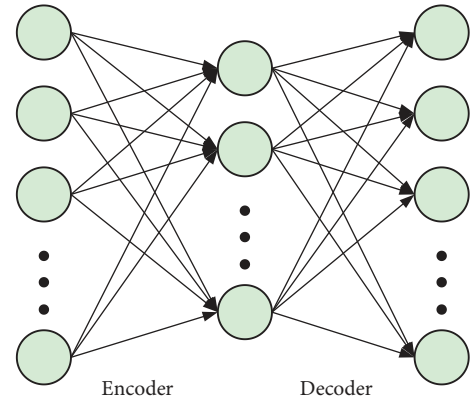

Figure 7: AE architecture.

where θ is the parameter set and $\theta = \{W, W^T, p, q\}$. Equation (15) shows that the AE attempts to learn an identity function such that the reconstruction is as close as possible to the input data.

Based on the basic AE model, some variants are proposed with different constraints, such as sparse autoencoders (SAE), denoising autoencoders (DAE), contractive autoencoders, and convolutional autoencoders. With different constraints, the AE variants can be applied for data denoising, dimensionality reduction and feature extraction. Among the variants of AE, DAE reconstructs the original input data from the noise-contaminated input data by minimizing the reconstruction errors. In DAE, part of the input data is randomly replaced with noise. In practice, they are usually set to 0. Hence, the added noise allows the autoencoder to extract more expressive features. Therefore, the optimization problem of DAE can be described as

$$\min_{\theta} \left\| s_g\left(W^T s_f(W\widehat{x} + p) + q\right) - x \right\|_2, \tag{16}$$

where $\widehat{x}$ is the noise-contaminated input data.

4.3. Hybrid Deep Model. In this part, we propose a hybrid deep model to realize deeper feature extraction and expression in SEI process. The hybrid deep model is built based on RBM and DAE. Figure 8 shows its architecture. A DAE is used as the first layer, two RBMs as the middle layers, and a logistic layer acts as a classification layer. The hybrid model combines the advantages of DAE and DBN both. First, the structure of the AE model and its variants maintain the recoverability of the input data, which is very important for the extraction of primary feature. Furthermore, the DAE maps the data into a high-dimensional space, which effectively increases the separability of the data. In this process, the recoverability of DAE ensures the stability of the projection result. Moreover, DAE also extracts abstract features better than the basic AE model. Two stacked RBMs make up a two-tier DBN that used to extract deep feature from the output of the DAE layer. Therefore, in the hybrid deep model, the DAE maps the input data into a high-dimensional space, and then two RBM layers project the data into the low-dimension space and extract deep features layer by

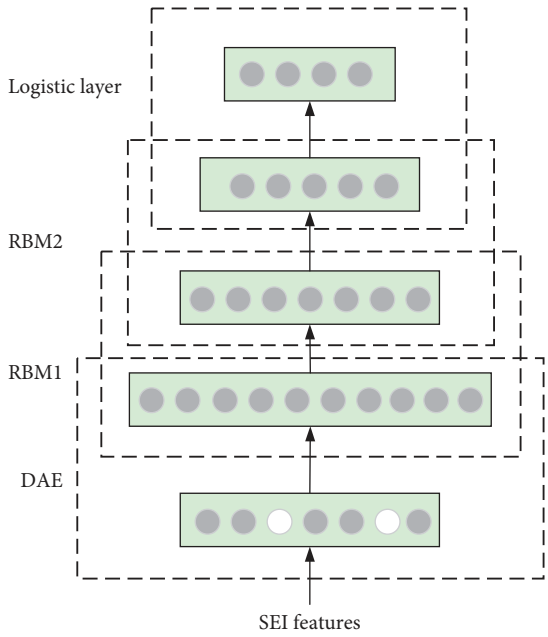

Figure 8: Architecture of the hybrid deep model.

layer. Hence, the hybrid model is able to further extract the intrinsic feature of the input data hierarchically and is resistant to noise within the input data.

4.4. SEI Method via BRT and the Hybrid Deep Model. We propose an SEI method using the BRT and the hybrid deep model. The BRT is applied to extract bispectrum projection vectors of emitters' signals, and then the hybrid model is applied to realize hierarchical feature detection and classification. The training of the hybrid deep model has two phases: the pretraining phase and the parameter fine-tuning phase. In the pretraining phase, we choose different training methods according to the modules of each layer. The RBM layers are trained using the CD algorithm, and the DAE layer is trained using the method proposed by Vincent et al. [36]. In the parameter fine-tuning phase, we use backpropagation to fine-tune the parameters of the entire model. Hence, the proposed SEI method includes the following four steps:

Step 1 (feature extraction): calculate the bispectrum projection vectors of the received signal via the proposed BRT method. Then, normalize the bispectrum projection vectors and divide them into training and testing sets.

Step 2 (pretraining): use the unsupervised training method mentioned above to train the hybrid deep model layer by layer and obtain the initial parameters of the model.

Step 3 (fine-tuning): use the backpropagation to fine-tune the parameters of the hybrid deep model based on the expected results and the output.

Step 4 (individual identification): use the testing set as the input of the hybrid deep model, and obtain the identification result. Compare the identification results with the original labels, and then calculate the identification accuracy.

5. Experiments and Discussion

5.1. Experimental Settings. In this section, we first introduce the ADS-B signals and our signal acquisition system. Then, the preprocessing method of the received signals and the parameter settings are introduced.

5.1.1. Signal Acquisition. To evaluate the performance of the proposed SEI model, we conducted a signal acquisition experiment and the received signals were used to test the performance of our SEI model.

We chose ADS-B transmitters for civil aircraft as the objects to be identified. ADS-B is a type of surveillance technology using satellite and data link technology. Aircraft equipped with ADS-B systems periodically and automatically broadcast their real-time flight information, such as speed, altitude, heading, and climb rate, into the surrounding airspace. Ground stations monitor air targets in the airspace by receiving and decoding ADS-B signals. The most widely used data link in ADS-B systems is the 1090 MHz extended squitter Model S (1090ES) data link, using pulse position modulation (PPM) recommended by the International Civil Aviation Organization (ICAO). ICAO has drafted general technical standards for ADS-B transmitters and stipulates that each aircraft has a unique aircraft registered number (ARN). Therefore, it is possible to obtain the ARN of an ADS-B pulse train by decoding the received pulse train according to the general standard. After obtaining the ADS-B signals and its ARNs, we can build a dataset to test the SEI method. Figure 9 shows a received ADS-B pulse train.

We designed a signal acquisition system, shown in Figure 10, to receive and store ADS-B signals. Figure 11 shows the structure of the system. The system includes a 1090 MHz antenna, a Tektronix DPO7054 oscilloscope, and an ADS-B RF front end. The antenna receives ADS-B signals transmitted by aircraft in surrounding airspace in real time, then the RF front end's power amplifiers and bandpass filters process the signals, amplifying the ADS-B signals and filtering out environmental noise and other signals. After that, an oscilloscope shows and stores the ADS-B signals synchronously. The SEI of the ADS-B transmitters is subsequently performed offline on a computer. The sampling rate and the bandwidth of the signal acquisition system were 5 GHz and 500 MHz separately.

We carried out the experiment in Xi'an, China, over two weeks. We collected pulse trains from six aircraft to create a dataset for our SEI experiments. There are three types of aircraft within the six aircraft: Boeing 737-89P (A and B), Airbus A321-214 (C and D), and Airbus A320-232 (E and F). Figure 12 shows the number of pulses from the six aircraft.

5.1.2. Parameter Settings. According to the general technical standards, the pulse width of ADS-B signals is $0.5 \pm 0.05\,\mu$s. Considering the differences in device manufacturing processes, component tolerance, and environmental factors, the duration length of a single pulse is set to be $0.8\,\mu$s, which is

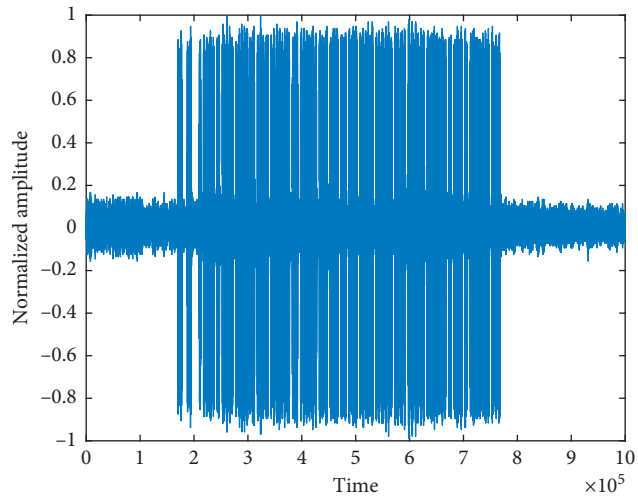

Figure 9: A received ADS-B pulse train.

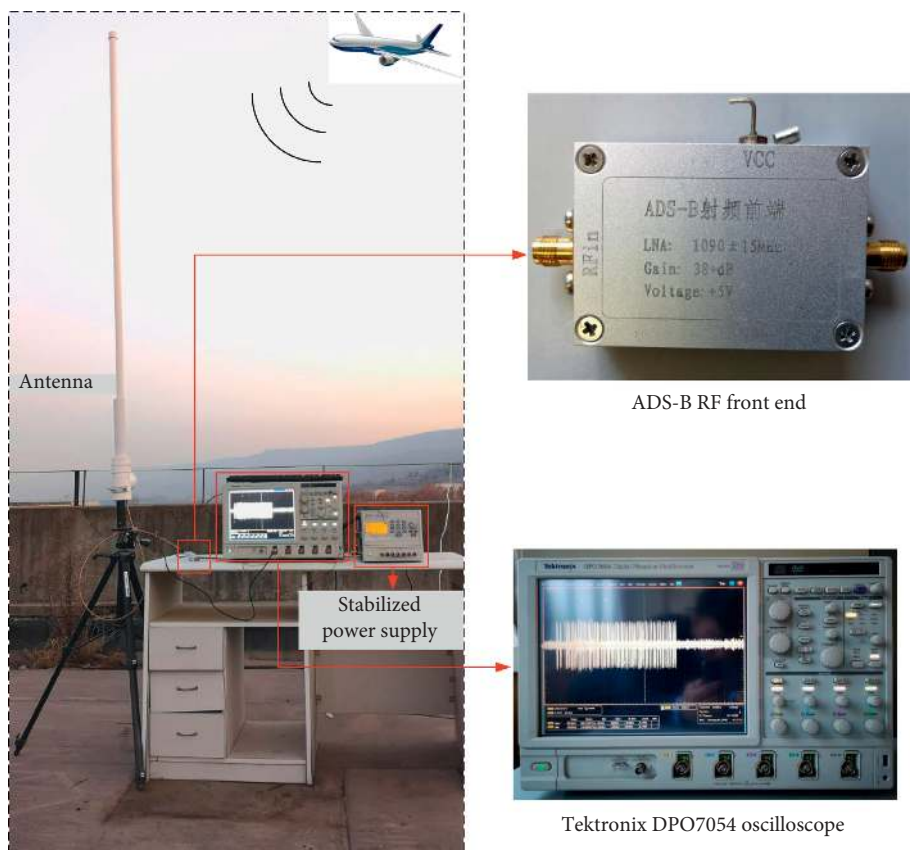

Figure 10: The ADS-B acquisition system.

4000 points. As shown in Figure 13, we consider setting the pulse detection threshold to detect the start of the pulse. The following methods are applied to determine the start of each pulse and obtain the rising edge of the pulses.

First, we detect the maximum amplitude of the pulse, $A_{\max}$. Then the position of the point with an amplitude threshold $\alpha A_{\max}$ could be searched. According to the pulse width L which is specified in the general standard of the ADS-B system, the start and the end points of the pulse could be determined based on β_1 and β_2. Finally, an ADS-B pulse with the length of $L_s = (\beta_1 + \beta_2)L$ is obtained. In our experiment, $\alpha = 0.4$, $L = 2500$, $\beta_1 = 0.2$, and $\beta_2 = 1.4$. We set the duration of the rising edge to be 0.3 μs, so the first 1500 points were intercepted as the rising edge of the pulse.

Using the received signals, we built a dataset to evaluate the SEI method. The dataset consisted of the bispectrum projection vectors and their labels. The label of each ADS-B signal, i.e., the ARN of the aircraft, was deduced by decoding

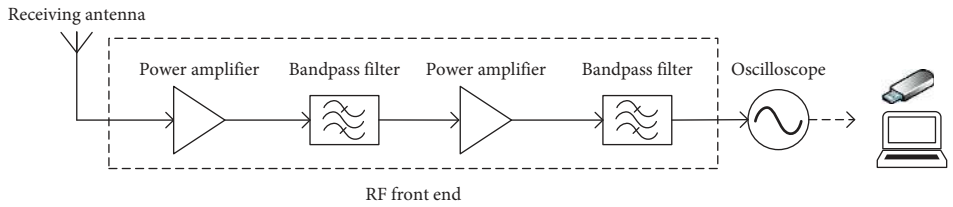

Figure 11: Structure of the ADS-B acquisition system.

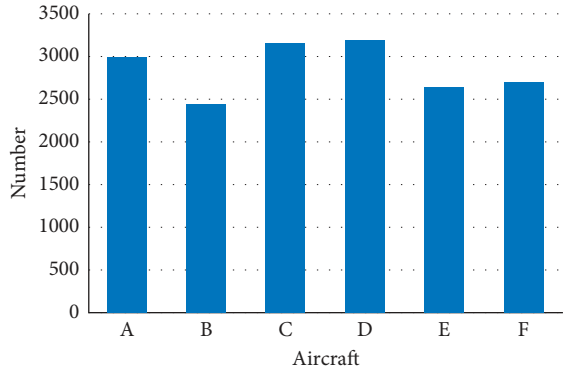

Figure 12: The composition of the data set.

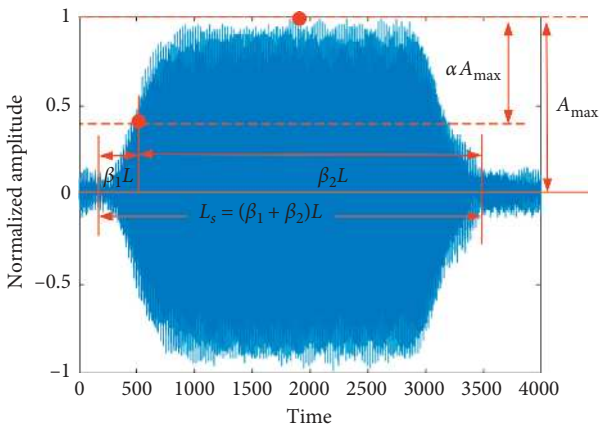

Figure 13: Acquisition method of a pulse.

ADS-B pulse trains. Therefore, signals with the same ARN were categorized together. Then the bispectrum projection vectors of the rising edges of the pulses were obtained using BRT. We divided the dataset into training and testing sets, putting 50% into the training set.

We used the indirect method with a 128-point fast Fourier transform to compute the bispectrum distribution of the rising edges. The projection angle in the Radon transform is a key parameter affecting the performance of BRT features, so we sampled at intervals of 5° between 0° and 180°. The performance of BRT features with different projection angles are shown in Figure 14. Finally, we set θ to 45°, which achieved the highest identification performance. In addition, before the bispectrum projection vectors input into the hybrid deep model, we applied a normalization process.

The performance of the hybrid deep model is highly dependent on the amount of available training samples, the

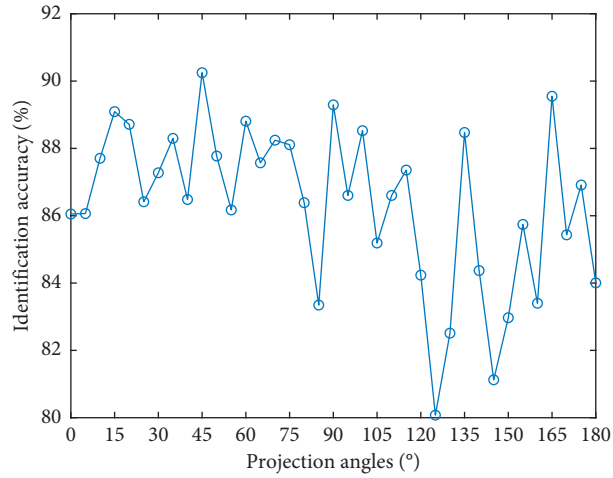

Figure 14: Comparison of different projection angles.

architecture, and the hyperparameters. The dimension of the bispectrum projection vectors is 185 in the experiment, so we used 185 units in the input layer. We applied a DAE layer to map the data to a dimension space three times as large as the original dimension, so there are 555 units in the output of the DAE. After that, the DBN consists of two RBM layers that were applied to compress and project the data to 300 and 100 dimensions. Finally, we used a logistic layer to determine the category of the input data. Therefore, the architecture of our model is (185, 555, 300, 100).

Among the hyperparameters, we found the number of the epochs, the minibatch size, the learning rate, the zero-masked fraction, and the momentum have a significant influence on the performance of the model. We therefore applied a gird search method to tune and select the hyperparameters of the hybrid deep model. Finally, in the pretraining stage, we used a learning rate of 0.03, a minibatch size of 25, an epoch number of 25, and a momentum of 0.9. The zero-masked fraction is 10%, which means that 10% of the input units were randomly set to zero. Besides, in the fine-tuning stage, we used a learning rate of 0.3, a minibatch size of 25, and an epoch number of 200.

5.2. Experiments. In this part, we performed comparative experiments to evaluate the feature extraction and identification performance of the proposed SEI method. First, we presented the output of each layer to observe the processing of the hybrid deep model. Second, we compared the feature extraction and identification performance of the existing method and the proposed model. Furthermore, we analyzed

the SEI ability of the proposed model based on the confusion matrix. Finally, we compared the performance of the features under different SNRs.

5.2.1. Validation Experiment. To observe the classification process of the hybrid deep model, the output of each layer was projected onto the 2-D plane via the linear discriminant analysis (LDA) method. Figures 15(a)–15(d) show the 2-D reduction of the input and output data of each layer. Figure 15(a) shows that although the input data overlapped, they were still distinguishable and discriminative. It means that the BRT features are effective for extracting the UMOP features. Figures 15(b)–15(d) show that the data are gradually separated and increasingly differentiated layer by layer. It shows that the hybrid deep model successfully performed further extraction of the input data hierarchically.

5.2.2. Comparative Experiment. The comparative experiments evaluate the performance of the proposed model from two aspects: feature extraction and identification.

In comparison with feature extraction methods, we compared our method with traditional bispectrum-based methods and other state-of-the-art methods. These methods were axially integral bispectrum (AIB), surrounding-line integral bispectrum (SIB), radially integral bispectrum (RIB), circularly integral bispectrum (CIB), bispectrum diagonal slice (BDS), RF distinct native attribute (RF-DNA) [29, 37, 38], Hilbert–Huang transform- (HHT-) based postprocessing features [39], and fractal features [40].

AIB, SIB, RIB, and CIB are integral bispectrum features with different integration paths. Figure 16 shows the integral paths of AIB, SIB, RIB, and CIB. As mentioned earlier, bispectrum has shown good performance in extracting UMOP features. However, the existing bispectrum-based features have some limitations. First, the integration paths of AIB and SIB could contain all the bispectrum points, but subtle energies caused by UMOP are easily drowned in integral calculation. Second, the comprehensiveness of RIB and CIB is related to the density of integration paths. The integration paths of RIB and CIB make it difficult to achieve uniform sampling in the process of discrete integral summation. As shown in Figure 16, if the integration paths are sparse, some points will be missed, or, if they are dense, some points will be reused in the adjacent path. Finally, the BDS features contain only the energy on the diagonal line, so a considerable amount of energies scattered in other regions is lost.

RF-DNA features are statistical RF fingerprint features that extracted from instantaneous amplitude, phase, and frequency responses [29, 37, 38]. In [29, 37, 38], the RF-DNA sequences are extracted by (a) segmenting the instantaneous amplitude, phase, and frequency of the received signals, then centering and normalizing the segments; (b) calculating the metrics (i.e., variance, skewness, and kurtosis) of each segment and the entire region; and (c) arranging the metrics in a vector form. In our experiments, the received signals were segmented into 10 segments, so finally we obtained a full-dimensional $N = 99$ RF-DNA feature set.

HHT is the well-known technique to deal with nonlinear and nonstationary signals [39]. In [39], an SEI feature set named VMD-SF is proposed based on variational mode decomposition (VMD) and spectral features. The VMD-SF feature set includes temporal and spectral features. The temporal features are extracted from the Hilbert spectrum of the decomposed VMD modes, and the spectral features are extracted from the spectral modes obtained using VMD. In our experiments, signals were decomposed into three modes using VMD, so we obtained a full-dimensional $N = 18$ VMD-SF feature set.

Fractal features have potential in characterizing complexity and irregularity of signals. In [40], fractal features based on box-counting dimension and variance dimension are used to distinguish phase noises between emitters. In our experiments, we applied the method proposed in [40] to obtain the fractal features of the received signals. The received signals were segmented into 10 segments, so we obtained a full-dimensional $N = 20$ fractal feature set.

Figure 17 shows the features extracted from actual ADS-B signals of six aircraft.

Second, to evaluate the identification ability of the hybrid deep model, six classifiers were applied for a comparison: DAE, DBN, conventional neural network (CNN), k-nearest neighbor (KNN), support vector machine (SVM), and LDA. The DAE model used in our experiment has two stacked autoencoders and a backpropagation layer. The DAE model has an architecture of (185, 100, 60). The DBN model consisted of two stacked RBMs with an architecture of (185, 120, 60). The LeNet-5 used in this paper consists of two sets of conventional and average pooling layers, followed by two fully connected layers and finally a softmax classifier. The size of the input image in LeNet-5 is 28×28, so we extended the one-dimensional feature to 1×784 using cubic spline interpolation and further reshaped to 28×28. In addition, we reduced the dimension of input data via the PCA method before inputting the data into the SVM and LDA. The ratio of energy loss in the PCA method is 0.01. A gird search method was used to tune and select the hyperparameters of classifiers. The parameter settings of the classifiers are listed in Table 1, and the results of the comparative experiments are shown in Table 2.

Table 2 shows the performance of different features with different classifiers. Among all the combinations of features and classifiers, the proposed method based on BRT features and the hybrid deep model performs best.

As mentioned above, a bispectrum is a type of higher-order spectrum that is superior in expressing intrinsic nonlinear features of signals and reducing environmental noise. The bispectrum distribution contains amplitude and phase information simultaneously and accurately characterizes nonlinear components within signals. Gaussian noise can be suppressed in the bispectrum calculation because the higher-order cumulant of Gaussian noise is equal to zero. Among the traditional bispectrum-based features, the RIB and CIB features have a problem of nonuniform sampling that introduces additional errors. The AIB and SIB have fixed integration paths, which means it is difficult to maximize the UMOP difference of different emitters' signals. Furthermore, the UMOP features are

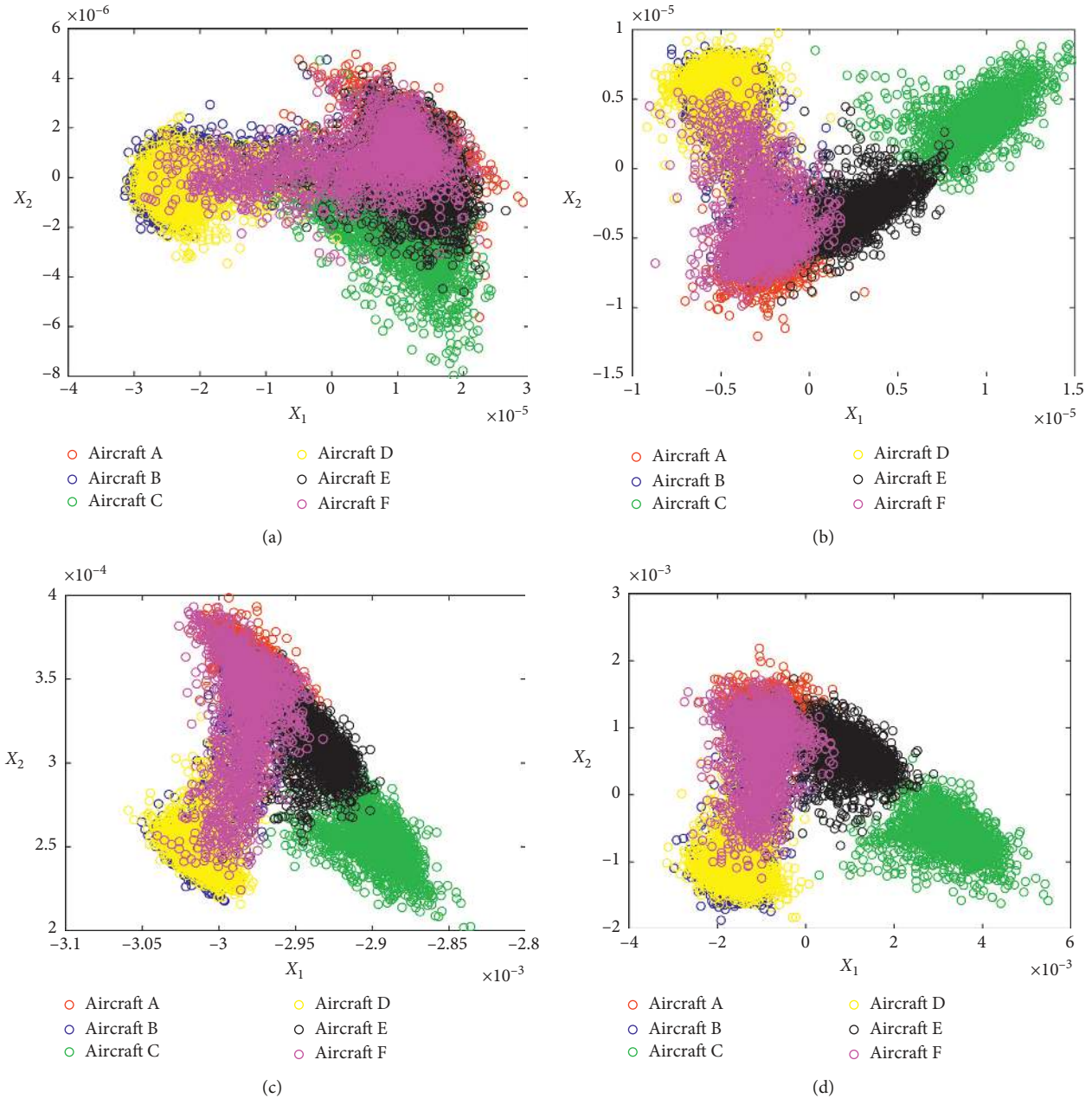

Figure 15: 2-D reduction of the input data and output of each layer. (a) 2-D reduction for the input data, (b) 2-D reduction for output of DAE, (c) 2-D reduction for output of RBM1, and (d) 2-D reduction for output of RBM2.

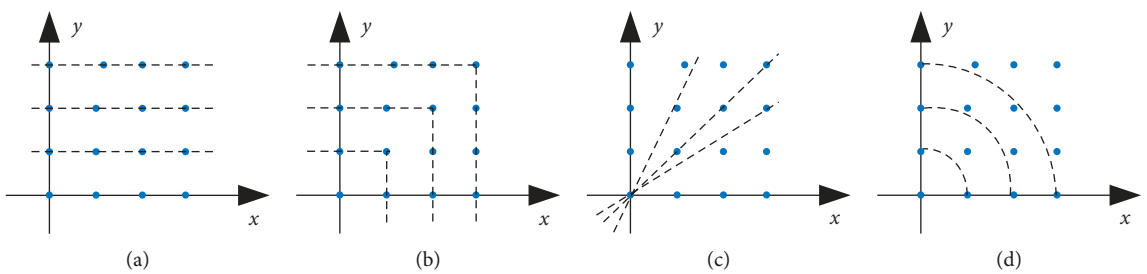

Figure 16: Geometric interpretation of integral bispectrum features. (a) AIB, (b) SIB, (c) RIB, and (d) CIB.

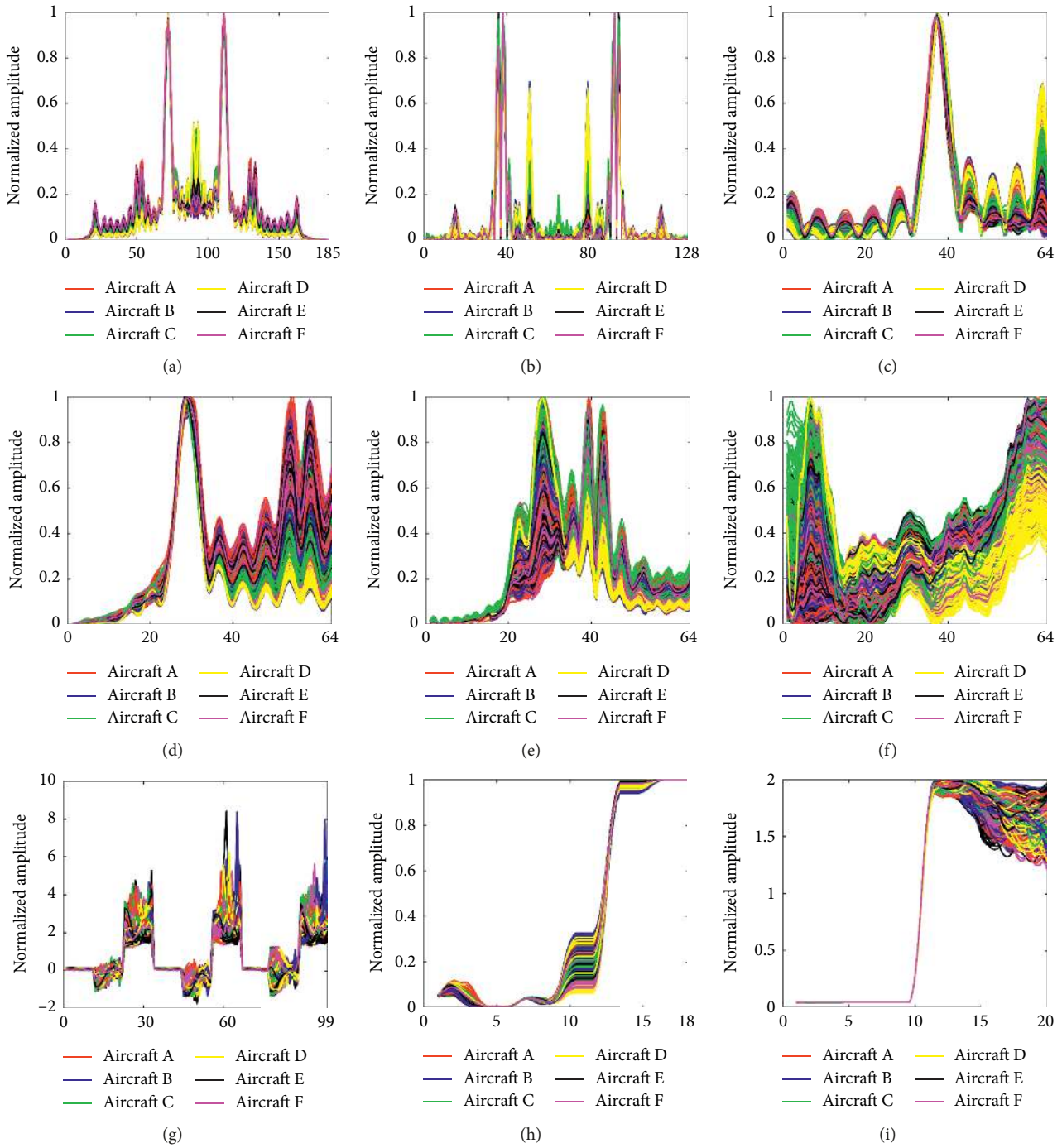

Figure 17: Extracted features of aircraft. (a) BRT, (b) BDS, (c) AIB, (d) SIB, (e) CIB, (f) RIB, (g) RF-DNA, (h) VMD-SF, and (i) fractal.

easily drowned in integral calculation. Our previous studies [41, 42] have proven that the BDS is useful for identifying signal modulations. However, although a considerable amount of energy is accumulated on the diagonal line, there are still significant scattered energies gathered in other regions. These scattered energies contain abundant individual information, which are lost in the BDS feature. It should be noted that the six bispectrum-based features are continuous features, while the RF-DNA, VMD-SF, and fractal features are discrete features consisting of different metrics. Discrete features such as RF-DNA, VMD-SF, and fractal features characterize UMOP through different metrics. But its characterization ability depends on the validity of the selected metrics. In the experiments, we found that the variance metrics, within RF-DNA, of different aircraft were not distinguishable so does the spectral roll-off within VMD-SF and the box-counting dimension within the fractal features. For SEI problems like the identification of ADS-B emitters, the ultimate goal is to identify individual emitters discriminatively, even if these emitters operate on the same frequency and have the same modulation.

Table 1: The parameter settings of classifiers.

Classifiers	Parameters setting	
Hybrid deep model	DAE layer Learning rate 0.03; momentum 0.9; Number of epochs 25; size of minibatch 25	Fine-tuning stage Learning rate 0.3; Number of epochs 200; size of minibatch 25
DBN	Pretraining stage Learning rate 0.01; momentum 0.5; Number of epochs 20; size of minibatch 25	Fine-tuning stage Learning rate 0.5; Number of epochs 300; size of minibatch 25
DAE	The autoencoder layers Learning rate of autoencoder 1 0.1; Learning rate of autoencoder 2 0.1; Size of minibatch 25; number of epochs 5; Activation function "sigmoid"	The backpropagation layer Learning rate 0.5; Number of epochs 60; size of minibatch 25; Zero-masked fraction 10%
CNN (LeNet-5)	Shape of convolutional layer C1 [1, 5, 5, 6]; Size of subsampling layer S2 $2 * 2$; Shape of convolutional layer C3 [5, 5, 6, 12]; Size of subsampling layer S4 $2 * 2$	Learning rate 0.5 Number of epochs 20 size of minibatch 5
SVM	Kernel function "RBF"; Gamma in kernel function 0.07	Cost parameter 1
KNN	Number of neighbors 10	Energy loss ratio in PCA method 0.01
LDA	Energy loss ratio in PCA method 0.01	

Table 2: The results of the comparative experiments (identification accuracy in %).

Features	Classifiers						
	Hybrid deep model	DBN	SAE	CNN (LeNet-5)	SVM	KNN	LDA
BRT	**90.25**	87.57	86.39	81.06	86.37	83.23	82.16
RIB	83.84	81.89	81.06	76.76	83.52	79.92	77.18
CIB	87.07	85.18	83.16	76.29	84.44	79.97	81.12
SIB	85.59	82.86	81.00	74.16	83.39	78.15	76.36
AIB	84.23	82.45	82.19	72.40	86.09	81.57	81.40
BDS	87.08	85.53	84.03	66.16	85.42	77.67	81.76
RF-DNA	62.36	60.58	85.87	73.26	86.72	87.57	86.88
VMD-SF	65.62	62.34	66.38	62.85	80.37	82.97	77.28
Fractal features	78.25	76.68	63.48	62.29	62.45	80.07	72.34

Hence, to achieve stable and effective individual recognition performance, it is necessary to extract subtle UMOP features adequately and maximize the distinction between UMOP features of different emitters. The proposed BRT extracts features using the projection of the bispectrum distribution and characterizes the UMOP within the signals. It not only contains the amplitude, the phase, and the spatial information of the distribution but also contains all of the distribution details. Hence, it offers significantly more complete extraction of UMOP features. In addition, the difference in different emitters' signals is maximized by selecting the optimal projection angle. Therefore, the BRT extracted features contain more individual signal characteristics and enable the best individual recognition performance.

Among all the classifiers, the hybrid deep model performed best. Actually, KNN, LDA, and SVM are shallow learning methods, and it is difficult to use them to extract the deep structural features of high-dimensional input data. KNN and LDA are traditional linear classifiers. The SVM classifier has some nonlinear learning ability by applying kernel functions, but it remains a shallow learning algorithm when compared with the DBN and DAE. Compared with traditional classifiers, DBN, DAE, and CNN better perform deep learning of high-dimensional input data. CNN is more suitable for processing images and may not take full advantage of the BRT features. The hybrid model combines the advantages of DAE and DBN. The experimental results demonstrate that the hybrid deep architecture performed better than the single deep architecture. In the hybrid model, the application of DAE mapped the data into a high-dimensional space and increased the data separability. In this process, DAE's recoverability ensured the stability of the projection result. Then, based on the output of the DAE layer, the two-tier DBN further extracted intrinsic and nonlinear features hierarchically. Furthermore, the pretraining process provided reasonable initial parameters for the model, and the backpropagation mechanism improved the discrimination performance.

Consequently, theoretical analysis and experimental results show that the proposed model achieved satisfactory SEI performance.

5.2.3. Analysis of Confusion Matrix. The experimental results in Table 2 show that the proposed model performed satisfactorily in identifying ADS-B signals of aircraft. To

further evaluate and analyze the individual identification ability of the method, especially the identification ability of the same type of emitters, we analyzed the identification result in detail.

As noted previously, there were three types of aircraft within the six aircraft in our experiment. Figure 18 shows the normalized confusion matrix of the identification result of the proposed model. It shows that the proposed model satisfactorily performed SEI. The identification accuracy ranged between 85.63% and 94.29%. It is worth noting that the identification error rate of aircraft of the same type is higher than that between different types of aircraft, which agrees with our expectations. Based on the confusion matrix, we use four evaluation metrics to analyze the performance of the model further: precision, recall, specificity, and F1 score. These metrics are defined based on the concepts of true positive (TP), false negative (FN), false positive (FP), and true negative (TN), as illustrated in Figure 19.

The definitions of the metrics are as given below:

$$\begin{aligned} \text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \\ \text{F1 score} &= \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \end{aligned} \tag{17}$$

where precision refers to the proportion of the correctly predicted samples in the positive-predicted samples. Recall refers to the proportion of the correctly predicted samples in the positive-labeled samples. Specificity refers to the proportion of the correctly predicted samples in the negative-labeled samples. The F1 score is a comprehensive indicator that combines the results of precision and recall. These metrics value ranges from 0 to 1, with 1 representing the best output of the model. These are commonly presented as percentages between 0 and 100. The metrics of the six aircraft are shown in Table 3. All the metrics indicate good success, demonstrating that the proposed method exhibited satisfactory performance for individual identification of the six aircraft.

5.2.4. Further Analysis. We considered evaluating the performance of different features by superimposing Gaussian noise on the received signals. The function *awgn.m* in Matlab was applied to add white Gaussian noise to the received signals. Its calling format is $y = \text{awgn}(x, snr, \text{“measured”})$, where x is the received signals, snr is the signal-to-noise ratio, and the parameter “measured” means that the signal level of x is computed to determine the appropriate noise level based on the value of snr. However, it should be noted that the SNR of the processed signals is lower than the true SNR of the ADS-B signals due to the noise carried in the receiving process. Therefore, we define the snr in *awgn.m* function as RSNR, which means the received-signal-to-noise ratio (RSNR). Figure 20 shows the identification accuracies of features under different RSNRs. The classifier is the hybrid deep model. It can be seen from Figure 20 that the performance of the features decreased as the RSNR decreases. Compared with other features, the BRT feature maintained higher identification accuracy than other features under different RSNRs.

5.2.5. Further Analysis. Feature extraction and classifier selection are two critical factors affecting SEI performance. It is easy to see in Table 2 that the combination of excellent individual features and classifiers resulted in satisfactory SEI performance. But it is worth noting that the difference in

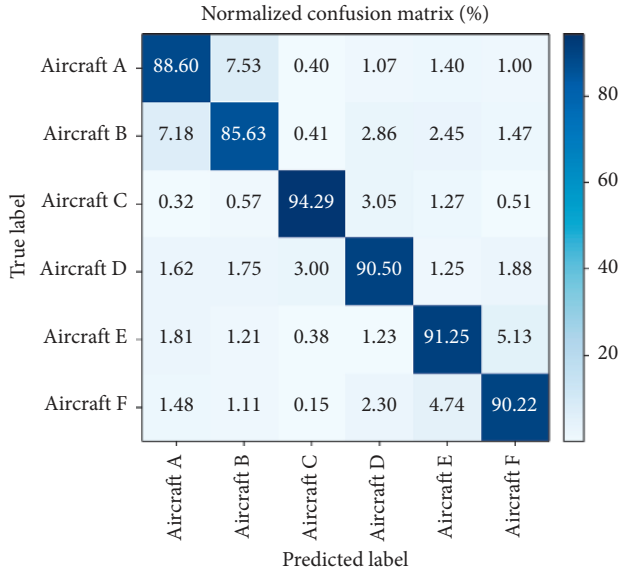

Figure 18: The normalized confusion matrix of the experimental result.

Confusion matrix		Predicted label	
		Positive	Negative
True label	Positive	TP	FN
	Negative	FP	TN

Figure 19: The illustration of the confusion matrix.

Table 3: The confusion matrix metrics (%).

Aircraft	Metric			
	Precision	Recall	Specificity	F1 score
A	89.08	88.67	97.70	88.87
B	85.35	85.63	97.55	85.49
C	95.74	94.29	99.06	95.01
D	91.59	90.50	98.09	91.04
E	88.64	91.25	97.86	89.92
F	89.76	90.22	98.08	89.99

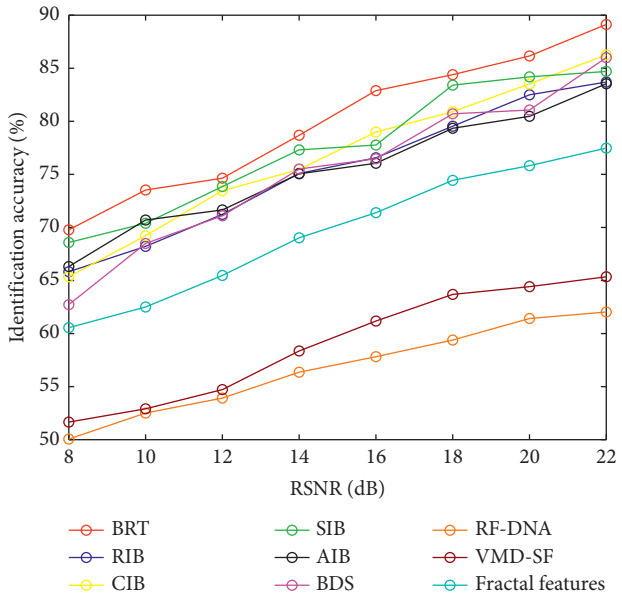

Figure 20: The performance of features under different RSNRs.

performance between features was greater than the difference between classifiers. This phenomenon indicates that the selection of UMOP features is the more influential factor for SEI performance. The discriminative UMOP features could provide a good start for the identification of classifiers. Excellent UMOP features combined with different classifiers achieved satisfactory SEI performance, while the performance using poor feature expression always remained low. In future research, we will consider further improvements to the BRT feature extraction method.

Furthermore, the experimental results from Figure 18 and Table 2 show that the individual identification accuracy of the proposed method averaged 90.25% using BRT features of a single ADS-B pulse. Therefore, a significant advantage of the proposed method is that the individual feature of the aircraft can be obtained based on a single pulse without decoding the entire ADS-B pulse sequence. If some ADS-B pulses are missing or overlapping, the proposed method can still identify the airplane while the traditional method based on PPM decoding may not be able to do so. Furthermore, the receiving process can accept multiple ADS-B pulses received continuously. Therefore, it is possible to improve the confidence of the identification results by fusing the identification results of several continuous pulses. This is part of our future work.

Our proposed method is not only applicable to identifying ADS-B emitters in civil aviation but is also usable for other SEI application, such as the identification of airborne fire-control radar and shipborne warning radar. Many other emitters behave differently from ADS-B transmitters. Their individual information is not contained in their intentional modulation, so the identifying information of these transmitters cannot be determined solely from their modulation modes. However, with sufficient training samples, our proposed method can successfully identify these emitters.

6. Conclusions

Feature extraction and classifier selection are critical factors for realizing SEI. In this paper, we propose an SEI method using bispectrum theory and deep learning models. The proposed method extracts the UMOP features via BRT and uses a hybrid deep model to perform deep feature extraction and discriminative identification. Experimental results show that BRT was superior in characterizing UMOP and that our proposed method achieved satisfactory SEI performance with ADS-B emitters. Owing to the satisfactory performance of the proposed method, we have considered further investigations improve the confidence of the identification results by fusing the identification results of several continuous pulses. We also plan to investigate the performance of our proposed method in other SEI applications.

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 no conflicts of interest.

Acknowledgments

This work was supported by the Aeronautical Science Foundation of China (Grant no. 20152096019) and the National Natural Science Foundation of China (Grant no. 61701524).

References

[1] F. Digne, A. Baussare, C. Cornu, and D. Jahan, "Classification of radar pulses in a naval warfare context using Bezier curve modeling of the instantaneous frequency law," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, Article ID 2671578, pp. 1469–1480, 2017.

[2] Y. Zhao, L. Wu, J. Zhang, and Y. Li, "Specific emitter identification using geometric features of frequency drift curve," *Bulletin of the Polish Academy of Science Technical Sciences*, vol. 66, no. 1, pp. 99–108, 2018.

[3] X. Ru, Z. Huang, Z. Liu, and W. Jiang, "Frequency-domain distribution and band-width of unintentional modulation on pulse," *Electronics Letters*, vol. 52, no. 22, pp. 1853–1855, 2016.

[4] G. Huang, Y. Yuan, X. Wang, and Z. Huang, "Specific emitter identification based on nonlinear dynamical characteristics," *Canadian Journal of Electrical and Computer Engineering*, vol. 39, no. 1, pp. 34–41, 2016.

[5] D. Stefano, "Specific emitter identification based on amplitude features," in *Proceedings of the 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, Malaysia, October 2015.

[6] Y. Shi and H. Ji, "Kernel canonical correlation analysis for specific radar emitter identification," *Electronics Letters*, vol. 50, no. 18, pp. 1318–1320, 2014.

[7] C. Bertoncini, K. Rudd, B. Nousain, and M. Hinders, "Wavelet fingerprinting of radio-frequency identification (RFID) tags," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 12, pp. 4843–4850, 2012.

[8] J. Han, T. Zhang, Z. Qiu, and X. Zheng, "Communication emitter individual identification via 3D-Hilbert energy spectrum-based multiscale segmentation features," *International Journal of Communication Systems*, vol. 32, no. 1, Article ID e3833, 2018.

[9] J. Zhang, F. Wang, O. A. Dobre, and Z. Zhong, "Specific emitter identification via Hilbert-huang transform in single-hop and relaying scenarios," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 6, pp. 1192–1205, 2016.

[10] Y. Yuan, H. Wu, X. Wang, and Z. Huang, "Specific emitter identification based on Hilbert-Huang transform-based time-frequency-energy distribution features," *IET Communications*, vol. 8, no. 13, pp. 2404–2412, 2014.

[11] X. Wang, G. Huang, Z. Zhou, W. Tian, J. Yao, and J. Gao, "Radar emitter recognition based on the energy cumulant of short time fourier transform and reinforced deep belief network," *Sensors*, vol. 18, no. 9, p. 3103, 2018.

[12] Z. Yang, W. Qiu, H. Sun, and A. Nallanathan, "Robust radar emitter recognition based on the three-dimensional distribution feature and transfer learning," *Sensors*, vol. 16, no. 3, p. 289, 2016.

[13] A. C. Polak, S. Dolatshahi, and D. L. Goeckel, "Identifying wireless users via transmitter imperfections," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 7, pp. 1469–1479, 2011.

[14] A. C. Polak and D. L. Goeckel, "Wireless device identification based on RF oscillator imperfections," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 12, pp. 2492–2501, 2015.

[15] O. Gungor and C. E. Koksal, "On the basic limits of RF-fingerprint based authentication," *IEEE Transactions on Information Theory*, vol. 62, no. 8, pp. 4523–4543, 2016.

[16] Y. Huang and H. Zheng, "Theoretical performance analysis of radio frequency fingerprinting under receiver distortions," *Wireless Communications and Mobile Computing*, vol. 15, no. 5, pp. 823–833, 2015.

[17] D. Zhang and W. B. Wu, "Asymptotic theory for estimators of high-order statistics of stationary processes," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4907–4922, 2018.

[18] Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, "Musical-noise analysis in methods of integrating microphone array and spectral subtraction based on higher-order statistics," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, Article ID 431347, 2010.

[19] M. Nasrolahzadeh, Z. Mohammadpoory, and J. Haddadnia, "Higher-order spectral analysis of spontaneous speech signals in alzheimer's disease," *Cognitive Neurodynamics*, vol. 12, no. 6, 2018.

[20] L. Saidi, J. Ben Ali, and F. Fnaiech, "Application of higher order spectral features and support vector machines for bearing faults classification," *ISA Transactions*, vol. 54, pp. 193–206, 2015.

[21] Y. Bengio, "Learning deep learning architectures for AI," *Foundations and Trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[22] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015.

[23] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014.

[24] A. T. Lopes, E. D. Aguiar, A. F. D. Souza, and T. Oliveira-Santos, "Facial expression recognition with conventional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.

[25] Y. Sun, X. Wang, and X. Tang, "Hybrid deep learning for face verification," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1489–1496, Sydney, Australia, December 2013.

[26] Y. Jiang, Z. Wu, J. Tang, Z. Li, X. Xue, and S. Chang, "Modeling spatial-temporal clues in a hybrid deep learning framework for video classification," *IEEE Transactions on Multimedia*, vol. 14, no. 8, 2017.

[27] S. Zhang, S. Zhang, T. Huang, W. Gao, and Q. Tian, "Learning affective features with a hybrid deep model for audio-visual emotion recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 3030–3043, 2018.

[28] S. Deng, Z. Huang, X. Wang, and G. Huang, "Radio frequency fingerprint extraction based on multidimension permutation entropy," *International Journal of Antennas and Propagation*, vol. 2017, Article ID 1538728, 6 pages, 2017.

[29] R. W. Klein, M. A. Temple, and M. J. Mendenhall, "Application of wavelet-based RF fingerprinting to enhance wireless network security," *Journal of Communications and Networks*, vol. 11, no. 6, pp. 544–555, 2009.

[30] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[31] L. F. Polania and K. E. Barner, "Exploiting restricted Boltzmann machines and deep belief networks in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4538–4550, 2017.

[32] N. Le Roux and Y. Bengio, "Representational power of restricted boltzman machines and deep belief networks," *Neural Computation*, vol. 20, no. 6, pp. 1631–1649, 2008.

[33] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[34] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[35] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," *Advances in Neural Information Processing Systems*, vol. 19, pp. 153–160, 2007.

[36] P. Vincent, H. Larochelle, Y. Bengio, and P. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the Twenty-Fifth International Conference (ICML2008)*, pp. 1096–1103, Helsinki, Finland, July 2008.

[37] C. Dubendorfer, B. Ramsey, and M. Temple, "ZigBee device verification for securing inductrial control and building automation systems," in *Proceedings of the International Conference of Critical Infrastructure Protection*, vol. 417, pp. 47–62, Washington, DC, USA, 2013.

[38] T. J. Bihl, K. W. Bauer, and M. A. Temple, "Feature selection for RF fingerprinting with multiple discriminant analysis and using ZigBee device emissions," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1862–1874, 2016.

[39] 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 Transactions on Information Forensics and Security*, vol. 14, no. 3, pp. 581–591, 2019.

[40] L. Wu, Y. Zhao, Z. Wang, F. Y. O. Abdalla, and G. Ren, "Specific emitter identification using fractal features based on box-counting dimension and variance dimension," in *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, pp. 226–231, Bilbao, Spain, December 2017.

[41] X. Wang, Y. Zhou, D. Zhou, Z. Chen, and Y. Tian, "Research on low probability of intercept radar signal recognition using deep belief network and bispectra diagonal slice," *Journal of Electronics and Information Technology*, vol. 38, no. 11, pp. 2972–2976, 2016.

[42] X. Wang, P. Guo, Y. Tian, and Y. Wang, "LPI radar signal recognition based on BDS-GD," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 44, no. 3, pp. 583–592, 2018.