Journal of Systems Engineering and Electronics
Vol. PP, No. 99, May 2023, pp.1 – 17

RFFsNet-SEI: a multidimensional balanced-RFFs deep neural network framework for specific emitter identification

FAN Rong[1,2], SI Chengke[2], HAN Yi[2], and WAN Qun[1,*]

1. School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 2. Institute of Electronic and Electrical Engineering, Civil Aviation Flight University of China, Guanghan 618307, China

Abstract: Existing specific emitter identification (SEI) methods based on hand-crafted features have drawbacks of losing feature information and involving multiple processing stages, which reduce the identification accuracy of emitters and complicate the procedures of identification. In this paper, we propose a deep SEI approach via multidimensional feature extraction for radio frequency fingerprints (RFFs), namely, RFFsNet-SEI. Particularly, we extract multidimensional physical RFFs from the received signal by virtue of variational mode decomposition (VMD) and Hilbert transform (HT). The physical RFFs and I-Q data are formed into the balanced-RFFs, which are then used to train RFFsNet-SEI. As introducing model-aided RFFs into neural network, the hybrid-driven scheme including physical features and I-Q data is constructed. It improves physical interpretability of RFFsNet-SEI. Meanwhile, since RFFsNet-SEI identifies individual of emitters from received raw data in end-to-end, it accelerates SEI implementation and simplifies procedures of identification. Moreover, as the temporal features and spectral features of the received signal are both extracted by RFFsNet-SEI, identification accuracy is improved. Finally, we compare RFFsNet-SEI with the counterparts in terms of identification accuracy, computational complexity, and prediction speed. Experimental results illustrate that the proposed method outperforms the counterparts on the basis of simulation dataset and real dataset collected in the anechoic chamber.

Keywords: specific emitter identification (SEI), deep learning (DL), radio frequency fingerprint (RFF), multidimensional feature extraction (MFE), variational mode decomposition (VMD).

DOI: 10.23919/JSEE.2023.000069

1. Introduction

As one of the cases of high-dimensional signal processing, specific emitter identification (SEI) refers to a process of distinguishing individual emitters by comparing radio frequency fingerprints (RFFs) originated from the nonlinearity of power amplifiers (PAs) [1]. Nowadays, SEI has become a critical issue in multidimensional domain deep learning [2], Internet of Things (IoT) [3], communication countermeasures [4], etc.

Conventionally, SEI is performed through three steps such as preprocessing, feature extraction, and classifier designing, in which the feature extraction is a core step. Up to now, various feature extraction techniques based on transient or steady-state RFFs including the amplitude, phase, frequency, energy envelope and the coefficients of wavelet are proposed [5−7]. However, these methods have the problem of RFF information loss, which reduces identification accuracy in SEI.

In addition, emitters from the same manufacturer may be so much alike that some emitters are completely indistinguishable from each other by existing methods. For these emitters, more subtle features need to be extracted. Several state-of-the-art (SOTA) approaches include short-time Fourier transform (STFT) [8], Wigner-Ville [9], and Choi-Williams distribution [10]. In these approaches, the procedures include calculating the time-frequency distribution of the received signal, applying morphological processing on the generated time-frequency images and finally classifying emitters.

It should be pointed out that all of the above methods are usually limited by the uncertain principle, which results in the contradiction between time resolution and frequency resolution. Moreover, the difficulty with the quadratic time-frequency representations is the inevitable cross-terms problem. As one of self-adaptive analysis methods for non-stationary signals, the Hilbert-Huang transform (HHT) was proposed in [11]. In this method, the received signal is decomposed into various empirical

---

Manuscript received August 10, 2022.

*Corresponding author.

The work was supported by the National Natural Science Foundation of China (62061003), Sichuan Science and Technology Program (2021YFG0192), and the Research Foundation of the Civil Aviation Flight University of China (ZJ2020-04; J2020-033).

modes, but it suffers from the mode mixing problem [12]. Recently, although variational mode decomposition (VMD) is proposed to decompose the received signal into temporal-and-spectral modes in order to extract more nonstationary RFFs, the problem of mode mixing is eliminated at the cost of high computational complexity [13]. Another SEI method based on multi-dimension approximate entropy (MApEn) was proposed in [14], where experimental results illustrate that when the length of samples is less than one hundred, the identification accuracy of this method deteriorates dramatically.

In fact, SEI can be regarded as a fitting problem of a nonlinear function mathematically. The universal approximation theorem in deep learning claims that it is feasible to use deep neural network (DNN) to fit any multidimensional (or high-dimensional) nonlinear function [15]. Therefore, several deep SEI methods have been proposed.

A deep SEI method based on complex-valued neural network was proposed in [16]. A multi-domain feature fusion SEI method via integrated learning (MDFFIL) was addressed in [17]. They both confirm superiority to the previously proposed SEI methods based on convolutional neural network (CNN) in both identification accuracy and convergence. Meanwhile, a deep SEI approach based on bi-spectrum feature was addressed in [18], in which the compressed spectrum of the received signal is used so that the feature information cannot be utilized sufficiently. In [19], a deep SEI method based on one-dimensional CNN (CNN-1D) was proposed, which resolves shortcomings of traditional methods of feature extraction and low identification accuracy. Moreover, another deep SEI method by virtue of deep residual network (DRN) was proposed in [20]. Based on ADS-B signal real data sets, it shows that DRN can achieve excellent recognition accuracy. In addition, an unsupervised SEI method using STFT and K-means algorithm was reported in [21]. In addition, a multi-feature fusion SEI method via deep ensemble learning (MFFDEL-SEI) can be found in [22].

Although these above methods can achieve good performance and have robustness in the case of low signal-to-noise ratio (SNR), the time-frequency spectrograms of emitters must be calculated in advance even in the identification stage. An open-set SEI method based on multi-classifier fusion is proposed to deal with unknown situations which have not been learned by the models during training [23]. It should be pointed out that the SEI issue on close-set is different from the case of open-set SEI, we do not address it in this paper.

Although SEI methods based on hand-crafted features have explicit physical meaning, the recognition processes of which are complicated. The error imposed in any step will affect the effectiveness of subsequent processing. In addition, the existing deep SEI methods suffer from weak physical interpretability about general features of emitters, which lead to poor stability and scalability of such methods. Compared with model-driven SEI methods, the deep SEI methods with high dimensional data representation ability have advantages in terms of recognition accuracy, environmental generalization, as well as fitting accuracy.

Motivated by these reasons, we propose a deep SEI method based on DNN. To be specific, VMD technique is firstly used to extract intrinsic mode functions (IMFs) of the received signal. And then, high dimensional features including temporal (i.e., the entropy feature (EF), the first order moment (FOM), and the second order moments (SOM)) and spectral (i.e., spectral flatness and spectral brightness) and some other effective features of the received are formed into RFFs for identification. The main contributions of this paper are as follows:

(i) A DNN framework is designed for learning maps from raw I-Q data to classifier of emitters, which realizes an end-to-end implementation of SEI. By virtue of RFFs-embedding, the combined loss function with multistage physical constraints is developed to train the network. The mode decomposition loss is used to decompose time modes and spectral modes of the received signal. Meanwhile, the feature extraction loss extracts time features and spectral features from previous modes. Moreover, the classification loss function learns classifier of emitters from features to the individual of emitters. As the RFFs of emitters are extracted by neural network automatically, it improves the prediction speed by an order of magnitude compared with baseline methods.

(ii) In the proposed framework, we extract the auxiliary modes of the received signal by virtue of auxiliary paths. Based on the auxiliary modes, the auxiliary features are also extracted. In this way, the hybrid-driven mechanism including data and physical RFFs is developed. As subtle features of emitters are utilized, the proposed framework improves the physical interpretability. Meanwhile, we analyze the computational complexity and convergence of RFFsNet-SEI. Based on the simulation dataset and real dataset, the merits of RFFsNet-SEI are illustrated in terms of identification accuracy, computational complexity, and prediction speed.

The remainder of this paper is organized as follows. Section 2 accounts for the system model of SEI. Section 3 introduces the methodology for SEI, and provides an accelerated-and-enhanced SEI implementation. Section 4 presents numerical experiments. Finally, conclusions of the paper are given in Section 5.

2. System model

A typical received signal contains channel noise followed by the transient signal and the steady-state signal. Transient signals contain the unintentionally modulated characteristics during the on-off switch or mode transformation. And the steady-state signal is defined when the emitter is transmitting between the start and end of the transients over the entire signal [24]. In an additive white Gaussian noise (AWGN) channel, the discrete-time received signal at time n can be expressed as follows:

$$r(n)=\begin{cases}\varepsilon(n), & 1\leqslant n\leqslant N_0\\ s_t(n)+\varepsilon(n), & N_0+1\leqslant n\leqslant N_1\\ s_s(n)+\varepsilon(n), & N_1+1\leqslant n\leqslant N\end{cases} \tag{1}$$

where $r(n)$ is the data sample at time instant n. N is the length of all sample points. N_0 is the starting point of the transient signal. N_1 is the endpoint. $\varepsilon(n)$ is channel noise. $s_t(n)$ is the transient signal. When $n\leqslant N_0$, $s_t(n)=0$, and $s_s(n)$ is the stable signal. When $n<N_1+1$, $s_s(n)=0$. At the receiver, RFFs extracted from $r(n)$ are used to identify different emitters. A typical captured signal including transient state and stable-state from two real PAs is shown in Fig. 1. The bottom subplot is the local zoom in version of the top subplot.

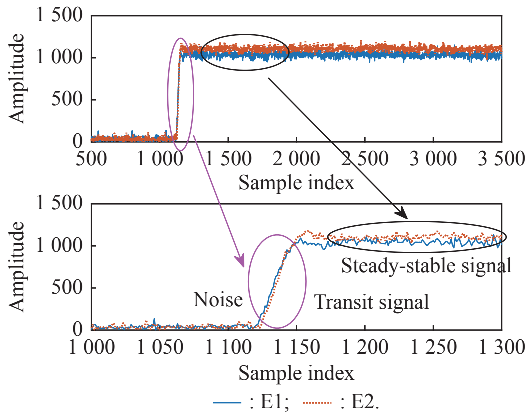


Fig. 1 **Typical captured step change signal from two power amplifiers**

Finally, the transient and steady-state attributes involved in the received signal, such as shown in Fig. 1 for an instance, are utilized to identify individual of emitters from $r(n)$ with existing methods [11,12,14,17−20,22].

3. Problem solutions

In traditional SEI approaches, three stages are involved, namely preprocessing raw I-Q data, extracting features of signal in the time or transform domain, and designing a classifier for identifying features. A typical SEI flowchart based on VMD is shown on the left side of Fig. 2, which has been proposed in [14]. Loss1(·), Loss2(·), and Loss3(·) are mode decomposition loss, feature extraction loss, and classification loss, respectively.

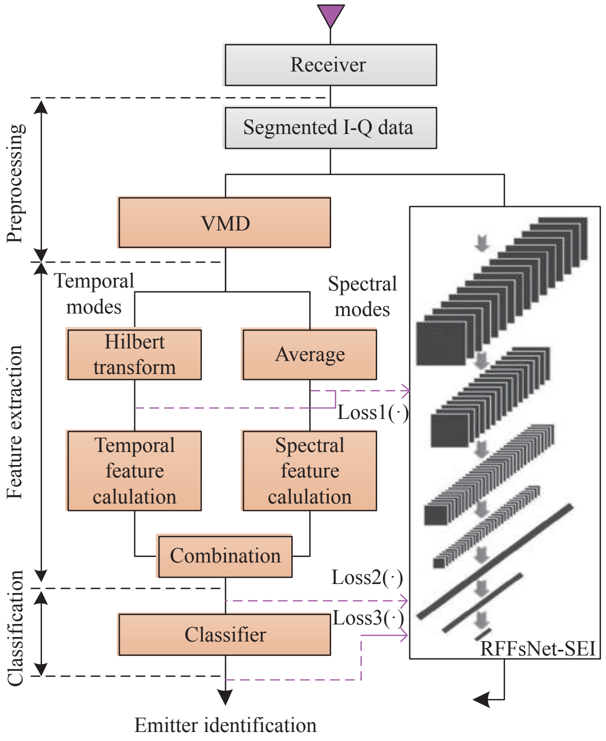


Fig. 2 **The proposed SEI flowchart**

Mathematically, identification procedures can be modeled as a series of linear and nonlinear nested functions from the left part of Fig. 2. As the universal approximation theorem confirms that any nonlinear function can be fitted by DNN, we can use neural networks to fit these functions. In the proposed method, a segmented neural network framework composed of three parts is developed to fit typical processing flow of the traditional SEI.

In the process of network development, we should pay attention to two factors including network parameters and computational complexity. The former affects the deployment cost of network. The larger the network parameters are, the larger the training data set is required. If the network size does not match the training data set, over fitting or under fitting problem may occur. As far as the computational complexity is concerned, it affects training time and system resource.

Another issue is the selection of hyper-parameters. Two-dimensional parameter space including network parameters and their feasible values are developed. The greedy searching strategy can be used to solve the optimal hyper-parameters in the two-dimensional parameter space.

Under the above considerations, we present the diagram based on neural network on the right side of Fig. 2. In the training stage, by virtue of minimizing the loss function including mode decomposition loss, feature extraction loss and classification loss, it enables the neural network to fit the SEI procedures. Therefore, the well trained neural network can directly output the identification results based on I-Q data in the prediction stage. The end-to-end recognition scheme simplifies procedures of identification.

3.1 Physical RFFs and model-aided SEI procedures

In this subsection, we first review the selection about hand-crafted RFFs and model-aided SEI procedures as shown on the left side of Fig. 2.

In the preprocess stage, VMD is used to decompose the received signal. Through steps of modes update, center frequencies update, and dual ascent update, the received signal is decomposed into a specified number of temporal and spectral modes with its center frequency. Without loss of generality, let q be the number of decomposed modes, then the output of VMD can be written as

$$[\boldsymbol{v}_q,\hat{\boldsymbol{v}}_q,\pi_q] = \mathrm{VMD}\{\boldsymbol{x}\},\quad q = 1,2,\cdots,Q \tag{2}$$

where Q is the number of IMFs. $\boldsymbol{v}_q$ is the qth temporal mode. $\hat{\boldsymbol{v}}_q$ is the qth spectral mode and π_q is the center frequency of the qth mode. The detailed procedures of VMD are summarized in [24].

In the feature extraction stage, both temporal features and spectral features are extracted from temporal mode $\boldsymbol{v}_q$ and spectral mode $\hat{\boldsymbol{v}}_q$, respectively. As shown in Fig. 2, temporal features are extracted from the Hilbert spectrum. The Hilbert spectrum for each mode of original signal $\boldsymbol{x}$ can be calculated as

$$H_q(\omega,m) = \begin{cases} \sqrt{v_q^2(m)+\tilde{v}_q^2(m)}, & \omega = \dfrac{\mathrm{d}\theta_q(m)}{\mathrm{d}m} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $\tilde{v}_q(m)$ denotes the Hilbert transform of $v_q(m)$. ω is the instantaneous angular frequency and $\theta_q(m)$ denotes the instantaneous phase for the qth mode. With the calculated time-frequency distribution $H_q(\omega,m)$, the EF, FOM, and SOM of the modes are extracted as follows:

$$\mathrm{EF} = -\sum_{i=1}^{F_m}\sum_{j=1}^{F_\omega}\left(\varepsilon_{i,j}\Big/\sum\nolimits_i\sum\nolimits_j\varepsilon_{i,j}\right)\lg\left(\varepsilon_{i,j}\Big/\sum\nolimits_i\sum\nolimits_j\varepsilon_{i,j}\right) \tag{4}$$

where F_m and F_ω are the number of time frames with time resolution δ_m and frequency frames with frequency resolution δ_ω respectively. The energy in each time-frequency frame is defined [25] as

$$\varepsilon_{i,j} = \int_{(i-1)\delta_m}^{i\delta_m}\int_{(j-1)\delta_\omega}^{j\delta_\omega} H_{i,j}(\omega,m)\mathrm{d}\omega\mathrm{d}m. \tag{5}$$

In (5), $H_{i,j}(\omega,m)$ is the (i,j)th time-frequency frame for $i=1,2,\cdots,F_m$ and $j=1,2,\cdots,F_\omega$.

Before computing FOM and SOM features, the Hilbert transform is converted into K-bit greyscale image [25] as

$$G_{i,j} = \left\lfloor (2^K-1)\times\frac{H_{i,j}}{\max\limits_{i,j} H_{i,j}}\right\rfloor \tag{6}$$

where $\lfloor\cdot\rfloor$ is the floor operation. K denotes the bit value in an image. $G_{i,j}$ is the (i,j)th pixel value of the greyscale image $\boldsymbol{G}$. And then FOM and SOM are computed by

$$\mathrm{FOM} = \frac{1}{I\cdot J}\sum_{i=1}^{I}\sum_{j=1}^{J} G_{i,j}, \tag{7}$$

$$\mathrm{SOM} = \sqrt{\frac{1}{I\cdot J}\sum_{i=1}^{I}\sum_{j=1}^{J}(G_{i,j}-\mathrm{FOM})^2}, \tag{8}$$

where I and J are the dimensions of rows and columns of matrix $\boldsymbol{G}$.

In Fig. 2, spectral features including spectral flatness (SF), spectral brightness (SB), and spectral roll-off (SR) are also extracted from spectral modes. We denote each spectral mode $\hat{v}_q(m)$ and N_t denotes the number of points in Fourier transform. And SF and SB are calculated by

$$\mathrm{SF} = \frac{\sqrt[N_t/2]{\prod\limits_{m=0}^{N_t/2}\left|\dfrac{1}{Q}\sum\limits_{q=1}^{Q}\hat{v}_q(m)\right|}}{\dfrac{2}{N_t}\sum\limits_{m=0}^{N_t/2}\left|\dfrac{1}{Q}\sum\limits_{q=1}^{Q}\hat{v}_q(m)\right|}, \tag{9}$$

$$\mathrm{SB} = \frac{\sum\limits_{m=B}^{N_t/2}\left|\sum\limits_{q=1}^{Q}\hat{v}_q(m)\right|}{\sum\limits_{m=0}^{N_t/2}\left|\sum\limits_{q=1}^{Q}\hat{v}_q(m)\right|}, \tag{10}$$

where B is the bin corresponding to the frequency $f(B)$.

The SR is defined as the frequency corresponding to bin B, below which certain percentage (for example, 80%) of total spectral energy is concentrated. Then SR is calculated [26] as

$$\mathrm{SR} = \max f(B): \sum_{j=0}^{B}\left|\sum_{q=1}^{Q}\hat{v}_q(j)\right| \leqslant 0.8\sum_{j=0}^{N_t/2}\left|\sum_{q=1}^{Q}\hat{v}_q(j)\right|. \tag{11}$$

Once temporal and spectral features of training sequences are extracted, the feature vector of the training sequence can be utilized to construct the training set. With the training set, any classifier such as k-NN can be trained and emitter can be identified using the well trained classifier as the same as in [13].

In the following, we illustrate the discriminability of the selected features by virtue of boxplot representation. The boxplot representation of each feature for eight PAs are denoted by E1, E2, ⋯ , E8, respectively. The details of eight PAs are addressed in Section 4. The boxplot representations of EF, FOM, SOM, SF, SB and SR for different emitters are shown in Fig. 3−Fig. 8. The boxplots of the six features depict the median value with the central line for different emitters. The median accounts for the concentration trend of each feature. Top and bottom edges indicate the maximum and minimum value, respectively.

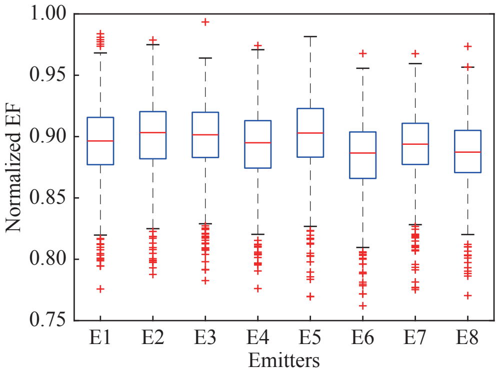


Fig. 3 Boxplot representation of EF value for different emitters

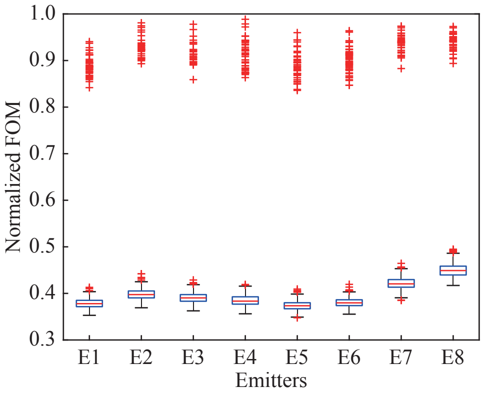


Fig. 4 Boxplot representation of FOM value for different emitters

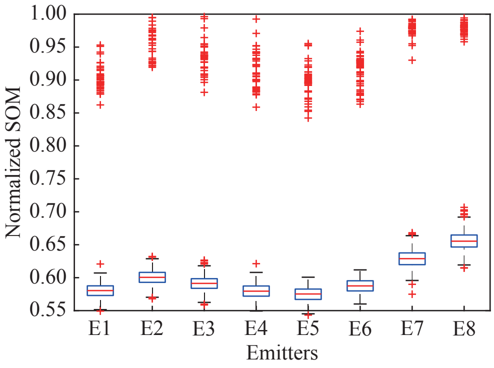


Fig. 5 Boxplot representation of SOM value for different emitters

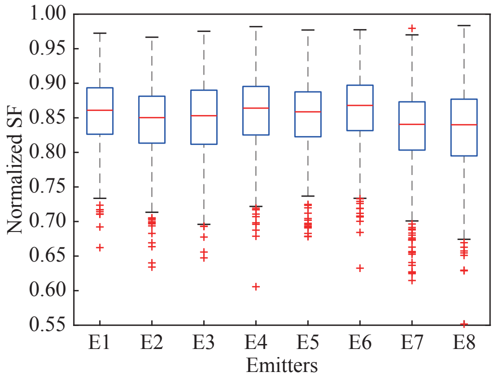


Fig. 6 Boxplot representation of SF value for different emitters

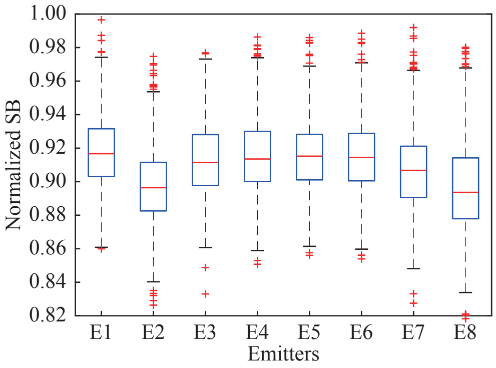


Fig. 7 Boxplot representation of SB value for different emitters

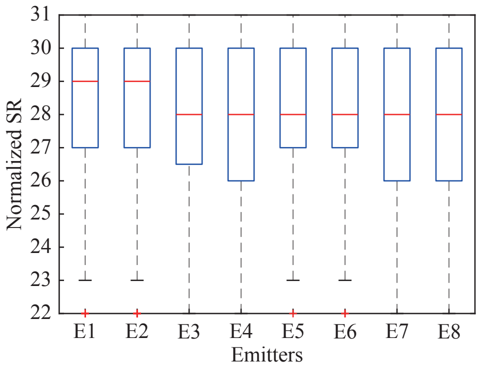


Fig. 8 Boxplot representation of SR value for different emitters

It can be seen from Fig. 3 that the median entropy value of E1 and E4 are both 0.895, while the median entropy value of the other six PAs are distinguishable. Though an overlap in the values can be seen in the boxplot of entropy value for E1 and E4, they can be discriminated based on the 25th percentile and the 75th percentile of entropy value. Thus it shows that entropy feature is effective for identifying individual of emitters.

Fig. 4 and Fig. 5 illustrate the variation in FOM value and SOM value for eight emitters, respectively. It can be observed that emitters have different median values. Similarly, though an overlap in the values can be seen from the boxplot of FOM value for E3 and E4 in Fig. 4, emitters can be discriminated based on the other median value such as EF and SOM.

In the same way, SF and SB plots have been illustrated in Fig. 6 and Fig. 7, respectively. Meanwhile, the 25th percentile and the 75th percentile based on the median value are also shown in Fig. 6 and Fig. 7, respectively. We can see that both SF and SB features provide discriminative feature for emitters.

From the results of Fig. 8, we can see that the SR features based on the emitters are difficult to distinguish, because the median value for E1−E2 and the 75th percentile of SR for E1−E8 are same constants. We found that SR feature should be discarded through experiments. Thus we do not use SR feature in our proposed method.

If limited physical features are utilized, there is the problem of feature information loss. Extracting more physical features may improve the performance of identification. For instance, the additional features including high-order cumulants, entropy spectrum, fractional Fourier transform and other features can also be utilized. However, they increase the computational complexity in the SEI. More importantly, experimental results show that the increase of feature types cannot always imply the identification accuracy. Hence, an alternated solution is developed by using limited physical features and deep features from I-Q data. It realizes the balance between the identification accuracy and the number of RFFs types.

3.2 An accelerated-and-enhanced deep SEI implementation based on balanced-RFFs

As shown on the left part of Fig. 2, the SEI has carried out a series of transformations including VMD, Hilbert transform, feature extraction, and classification. As any linear and nonlinear transforms can be fitted by DNN, without loss of generality, the nonlinear function g is parameterized by a CNN of L layers, i.e.,

$$g(\boldsymbol{x}) = g_L(g_{L-1}(\cdots g_1(\boldsymbol{x}))) = \boldsymbol{y} \tag{12}$$

where $\boldsymbol{x}$ is the input data of the neural network. $\boldsymbol{y}$ is the corresponding one-hot label for emitter. In order to fit the nonlinear function $g(\boldsymbol{x})$, we propose the RFFsNet-SEI framework based on CNN architecture in Fig. 9, which is composed of convolution-batch-normalization-Leaky-ReLU (CBL) blocks, residual unit (ResUnit) blocks, and convolution-batch-normalization-fusion (CBF) blocks. With these blocks, signal modes decomposition, signal features extraction, and emitter classification are realized in sequence.

In Fig. 9, the input layer of RFFsNet-SEI is configured according to the received signal model $\boldsymbol{x} \in \mathbf{R}^{P\times1\times1}$. Symbol ‘C’ is the channel combination operation. Symbols ‘i’ and ‘o’ are the number of input channels and output channels, respectively. Q is the number of IMFs. N_{aux} is the number of auxiliary channels. In RFFsNet-SEI, Q+1+N_{aux}=8. Q and N_{aux} are chosen by experiments.

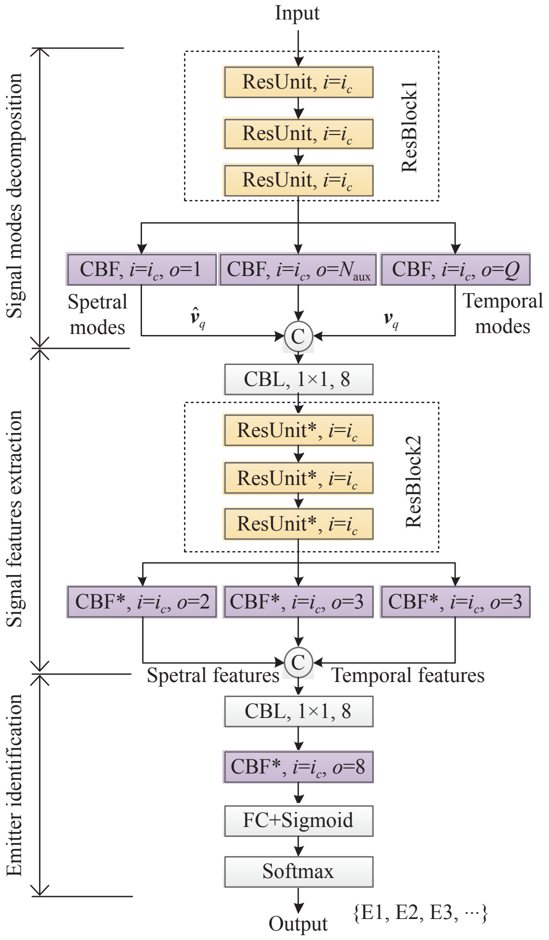


Fig. 9 Diagram of RFFsNet-SEI

We present the workflow of the framework in the following.

Firstly, that signal modes decomposition module contains three cascaded residual units and three parallel CBF blocks. These CBF blocks share the output feature maps of three cascaded ResUnits, namely ResBlock1. This sharing scheme is helpful because the temporal modes $\boldsymbol{v}_q$ and the spectral modes $\hat{\boldsymbol{v}}_q$ of the received signal cannot perform orthogonal decompose completely by neural network. In order to reduce the errors of decomposed modes, an additional CBF block is used to extract the auxiliary modes. Based on the advanced scheme, the temporal modes and spectral modes as well as the auxiliary modes of the received signal are decomposed from the outputs of ResBlock1.

Secondly, the temporal modes and spectral modes as well as the auxiliary modes (i.e., the other unknown but existed modes except for the temporal modes and spectral modes) are combined in the feature extraction module, which are fused by the CBL block. In the next, by using three cascaded residual units (ResBlock2), the features including EF, FOM, SOM, SF, and SB are extracted. As five kinds of physical features cannot cover all the information of the original data (viz., feature information is lossy), similar with the auxiliary mode scheme, the auxiliary feature extracted by CBF is used to increase the information of features.

Thirdly, the temporal features, spectral features, and the auxiliary features are combined in the emitter identification module and nonlinear transformation is performed by the CBL block at first. And then, the CBF block is imposed to extract the fused features from outputs of CBL block. A full connection (FC) layer followed by CBF block reshapes outputs of the CBF block into a vector. Finally, the emitter identification is performed by a Sigmoid layer, which applies Sigmoid function elementwise to the values of previous layer and returns values in (0, 1). The selection of Sigmoid function over the softmax is due to the presence of labels, which can receive a value equal or close to 1 independently.

There are three merits in the proposed framework. Firstly, the manually selected empirical features and the other potential features are used by RFFsNet-SEI, which improve information utilization of the received signal. Moreover, the nonlinear mappings from raw data to emitter identification are learned by network, so as to realize an end-to-end identification. RFFsNet-SEI improves the identification accuracy because both transient and stable-state of the received signal are utilized.

Given by workflow of the framework, we further introduce the specific blocks shown in Fig. 10, viz., CBL, ResUnit, and CBF block. Symbol ‘+’ is addition operation. Both ‘Conv’ and ‘Conv*’ are the convolutional layers with the kernel size of $d\times 1$, wherein the strides are 1×1 and 2×1, respectively. In ‘CBL*’, the convolutional layers are constructed by ‘Conv*’ instead of ‘Conv’.

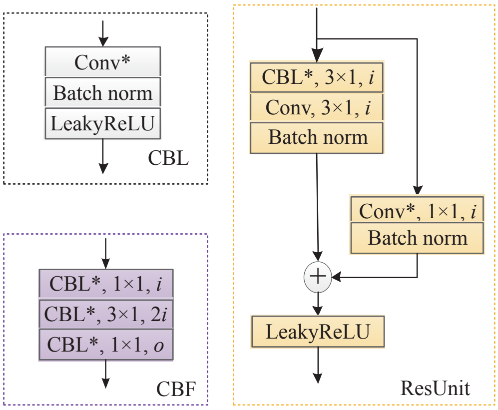


Fig. 10 Structure of CBL, ResUnit, and CBF block

In each CBL block, the convolutional layer has n_C channels and is followed by a batch normalization layer and a LeakyReLU layer. The batch normalization layer accelerates network convergence in the training and reduces the initialization sensitivity by performing zero mean and unit variance. As ReLU suffers from slow learning or even dead neurons, LeakyReLU is used as the activation function. For each one of the n_C channels, mathematical expression of the convolution operation with input $\boldsymbol{x}\in\mathbf{R}^{P\times 1\times 1}$ and the kernel $\boldsymbol{\kappa}\in\mathbf{R}^{n_{\text{ker}}\times 1\times n_C}$ is a vector of dimension Δ as

$$(\boldsymbol{x} * \boldsymbol{\kappa})_\sigma = \sum_{i=1}^{n_{\text{ker}}} \sum_{c=1}^{n_C} \kappa_{i,1,c}\, x_{s(\sigma-1)+i,1,c} + b_c \tag{13}$$

where $\sigma = 1,2,\cdots,\varDelta$, $\varDelta = \lfloor (N-n_{\text{ker}})/s \rfloor + 1$, n_{ker} is the width of kernel, and s is the stride. b_c is the bias for the cth channel.

$$F_{\text{out}}^{\text{CBL}} = F_{\text{LeakyReLU}}(F_{\text{BN}}(F_{\text{Conv}}(F_{\text{in}}))) \tag{14}$$

where F_{in} is the input of CBL. The σth entry of $F_{\text{Conv}}(F_{\text{in}})$ equals $(\boldsymbol{x} * \boldsymbol{\kappa})_\sigma$. $F_{\text{BN}}(\cdot)$ is the batch normalization function [27]. $F_{\text{leakyReLU}}(\cdot)$ is the Leaky ReLU function [28].

Deepening the neural network layers may saturate or degrade the performance due to the back propagation gradient disappearing, the residual connection is designed in RFFsNet-SEI, so the ResUnit block involves two branches in Fig. 10. The left branch has the CBL block with the kernel size of $d\times 1$, wherein the stride is 1×1 without down sampling or 2×1 with down sampling carried out. The convolutional layer followed by a batch normalization layer is cascaded to the output of CBL block. Mathematically, given the input of ResUnit $F_{\text{in}}^{\text{ResUnit}}$, the output feature of ResUnit is performed as

$$F_{\text{out}}^{\text{left}} = F_{\text{CBL}}^{*}\left(F_{\text{BN}}\left(F_{\text{Conv}}\left(F_{\text{in}}^{\text{ResUnit}}\right)\right)\right), \tag{15}$$

$$F_{\text{out}}^{\text{right}} = F_{\text{BN}}\left(F_{\text{Conv}}^{*}\left(F_{\text{in}}^{\text{ResUnit}}\right)\right). \tag{16}$$

where $F_{\text{CBL}}^{*}(\cdot)$ is the CBL function wherein the down sampling is adopted.

The right branch of ResUnit is composed of a convolutional layer with the size of 1×1 and cascaded by a batch normalization layer. Finally, the features from different layers are merged. The outcome of residual connectivity with LeakyReLU activation operation is

$$F_{\text{out}}^{\text{ResUnit}} = F_{\text{LeakyReLU}}\left(F_{\text{out}}^{\text{left}} + F_{\text{out}}^{\text{right}}\right). \tag{17}$$

In order to improve the expression ability of features in different receptive fields, three cascaded CBL blocks with different channel sizes are involved in CBF block:

$$F_{\text{out}}^{\text{CBF}} = F_{1\times 1}^{\text{CBL}}\left(F_{3\times 1}^{\text{CBL}}\left(F_{1\times 1}^{\text{CBL}}\right)\right). \tag{18}$$

Base on the diagram of RFFsNet-SEI, we present its explicated implementation in Table 1. The ‘Output size’ is the size of the output tensor of current component. In CBL, Conv, and the right sub-branch of ResUnit (#Res), the kernel size of filter, the number of channels and the stride are written in the form of ‘d×1, n_C/s’. It should be pointed out that the architecture details of RFFsNet-SEI in Table 1 correspond to the order in Fig. 9.

Table 1 The architecture details of RFFsNet-SEI

Name	Component	Output size	CBL	Conv	#Res
ResBlock1	ResUnit(top)	64×1×16	3×1,16/1	3×1,16/1	1×1,16/1
	ResUnit(mid.)	64×1×16	3×1,16/1	3×1,16/1	1×1,16/1
	ResUnit(bot.)	64×1×16	3×1,16/1	3×1,16/1	1×1,16/1
CBFs	Component	Output size	CBL(top)	CBL(mid.)	CBL(bot.)
	CBF(left)	64×1×1	1×1,16/1	3×1,32/1	1×1,1/1
	CBF(mid.)	64×1×3	1×1,16/1	3×1,32/1	1×1,3/1
	CBF(right)	64×1×4	1×1,16/1	3×1,32/1	1×1,4/1
	CBL	64×1×8	1×1,8/1	–	–
ResBlock2	Component	Output size	CBL	Conv	#Res
	ResUnit(top)	32×1×16	3×1,16/2	3×1,16/1	1×1,16/2
	ResUnit(mid.)	16×1×16	3×1,16/2	3×1,16/1	1×1,16/2
	ResUnit(bot.)	8×1×16	3×1,16/2	3×1,16/1	1×1,16/2
CBF*s	Component	Output size	CBL(top)	CBL(mid.)	CBL(bot.)
	CBF*(left)	1×1×2	1×1,16/2	3×1,32/2	1×1,2/2
	CBF*(mid.)	1×1×3	1×1,16/2	3×1,32/2	1×1,3/2
	CBF*(right)	1×1×3	1×1,16/2	3×1,32/2	1×1,3/2
	CBL	1×1×8	1×1,8/1	–	–
CBF	Component	Output size	CBL(top)	CBL(mid.)	CBL(bot.)
	CBF	1×1×8	1×1,16/1	3×1,32/1	1×1,8/1
	FC	10×1	–	–	–

3.3 Loss function

In this paper, RFFsNet-SEI conducts offline training in supervised manner over the training dataset. The loss function is composed of three parts, namely, the mode decomposition loss including temporal IMF modes and spectral IMF modes, the feature extraction loss including temporal and spectral features, and the classification loss. It is explicitly expressed as

$$\min_{\vartheta} \mathrm{Loss}(\vartheta) = \min_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3} \mathrm{Loss}(\boldsymbol{\theta}_0,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3) = \underbrace{\sum_{p=1}^{P} \|\boldsymbol{\Psi}_p - \boldsymbol{G}(\boldsymbol{x}_p,\boldsymbol{\theta}_0)\|_F + \sum_{p=1}^{N}\|\boldsymbol{v}_p - g_1(\boldsymbol{x}_p,\boldsymbol{\theta}_1)\|_2 +}_{\text{mode decomposition loss}} \underbrace{\sum_{p=1}^{N}\|\boldsymbol{\lambda}_p - g_2(\widehat{\boldsymbol{x}}_p,\boldsymbol{\theta}_2)\|_2}_{\text{feature extraction loss}} + \underbrace{\sum_{p=1}^{N}\|\boldsymbol{y}_p - g_3(\tilde{\boldsymbol{x}}_p,\boldsymbol{\theta}_3)\|_2}_{\text{classification loss}} \tag{19}$$

where $\vartheta = \{\boldsymbol{\theta}_0,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3\}$. $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ are the network parameters for fitting temporal IMF modes and spectral IMF modes and fitting temporal features and spectral features, respectively. $\boldsymbol{\theta}_3$ is the network parameter of emitter identification module. $\boldsymbol{\Psi}_p$ and $\boldsymbol{G}(\boldsymbol{x}_p,\boldsymbol{\theta}_0)$ are temporal IMF modes matrix and the corresponding prediction about the pth training sample $\boldsymbol{x}_p$, respectively. $\boldsymbol{v}_p$ and $g_1(\boldsymbol{x}_p,\boldsymbol{\theta}_1)$ are the averaged spectral IMF modes and the corresponding prediction version about $\boldsymbol{x}_p$, respectively. $\boldsymbol{\lambda}_p$ is composed of the temporal features and spectral features, viz., $\boldsymbol{\lambda}_p = [\mathrm{EF}_p, \mathrm{FOM}_p, \mathrm{SOM}_p, \mathrm{SF}_p, \mathrm{SB}_p]^{\mathrm{T}}$. $g_2(\widehat{\boldsymbol{x}}_p,\boldsymbol{\theta}_2)$ is the temporal features and spectral features about the pth outputs of the signal decomposed module (viz., $\widehat{\boldsymbol{x}}_p$). $\boldsymbol{y}_p$ is the classifying label. $g_3(\tilde{\boldsymbol{x}}_p,\boldsymbol{\theta}_3)$ is the predicted emitter identifier about the pth outputs of the signal feature extraction component (viz., $\tilde{\boldsymbol{x}}_p$).

We optimize the set of all trainable parameters whose updates are carried out via back-propagation by minimizing the reconstruction error. For simplicity, we let $\boldsymbol{x}^{[l-1]}$ be the input in the lth layer of RFFsNet-SEI. Operator $\boldsymbol{\varphi}^{[l]}(\cdot)$ is used in the convolution map for the input of the lth layer. $\boldsymbol{W}^{[l]}$ is the weight matrix of the lth layer for

$2 \leqslant l \leqslant L$. $\rho(\cdot)$ is the activation function. γ is a scaling factor to normalize the input of each layer. Thus, the forward calculation can be defined recursively as

$$\begin{cases} \boldsymbol{x}^{[0]} = \boldsymbol{x} \\ \boldsymbol{x}^{[l]} = \sqrt{\gamma}\rho\left(\boldsymbol{W}^{[l]}\boldsymbol{\varphi}^{[l]}(\boldsymbol{x}^{[l-1]})\right),\ 1 \leqslant l \leqslant L \end{cases}. \tag{20}$$

The training of RFFsNet-SEI can be achieved via a randomly initialized gradient descent algorithm to find the global optimum of the loss function. The random initialization scheme adopted in this paper is xavier. For any layer, each entry in $\boldsymbol{W}^{[l]}$ is sampled from a standard Gaussian distribution with mean 0 and unit variance. Then the back propagation update can be formulated as

$$\boldsymbol{W}^{[l]}(k) = \boldsymbol{W}^{[l]}(k-1) - \eta\frac{\partial \mathrm{Loss}(\vartheta(k-1))}{\partial(\boldsymbol{W}^{[l]}(k-1))} \tag{21}$$

where $\eta > 0$ is the learning rate.

Let $\Omega(\cdot)$ and $O(\cdot)$ be Big-Ω and Big-O notations and poly(n) represents a polynomial about the kernel number n. According to the results of [29], we know that if the number of hidden nodes per layer and the step size satisfy $\Omega(\mathrm{poly}(n)2^{O(L)})$and $\Omega(\mathrm{poly}(n)/2^{O(L)})$, then we have

$$\mathrm{Loss}(\vartheta(k)) \leqslant (1-\eta\cdot c)^k \mathrm{Loss}(\vartheta(0)) \tag{22}$$

In(22), $k = 1, 2, \cdots$ and c is a constant value related to a certain fixed matrix, which only depends on the input training data and the architecture of neural network and has no dependence on weight parameters.

Obviously, (22) implies that the loss function is convergent. In other words, once the training dataset is given, RFFsNet-SEI can learn the data distribution by minimizing the loss function$\mathrm{Loss}(\vartheta)$. And the nonlinear mappings from the received signal to individual of emitters are established, which can be used to predict individual of emitters in the prediction stage.

In deep SEI methods, the optimal weights of the network are learned by forward gradient descent searching and back forward error propagation. As the back propagation algorithm usually converges to the local solution, a reasonable loss function is needed. In this paper, the classification loss function (Loss 3) is developed for the classification of abstract features. In addition to Loss 3, mode decomposition loss (Loss 1) and feature extraction loss (Loss 2) are also imposed in the developed loss function. As Loss 1 and Loss 2 have explicit physical meaning, they enhance distinguishability of emitters. The final loss is devised on the basis of Loss 1, Loss 2 and Loss 3. In this way, the probability of the optimal gradient descending and convergence speed of network are both improved by the loss. Besides, the identification accuracy, physical interpretability, and generalization of network are also enhanced.

3.4 Computation complexity of RFFsNet-SEI

Next, we assess the computation complexity of the RFFsNet-SEI in terms of network size and the computational cost. The network size is characterized by the number of trainable parameters and the computational cost is measured via mega floating-point operations (MFLOPs). The input feature maps, the convolution kernel size, and the output channels of the lth layer are denote by $C_{\text{in}}^{[l]}$, $S^{[l]} \times 1$, and $C_{\text{out}}^{[l]}$, respectively. The learnable parameters of the lth convolution layer equal $(C_{\text{in}}^{[l]}S^{[l]} + 1)C_{\text{out}}^{[l]}$. Thus the learnable parameters of RFFsNet-SEI can be calculated as

$$P_{\text{size}} = \sum_{l=1}^{L}(C_{\text{in}}^{[l]}S^{[l]} + 1)C_{\text{out}}^{[l]}. \tag{23}$$

The FLOPs measurements include the height H=1 and width W of the input channel of each convolution layer and the non-linearity function is calculated for free. The FLOPs of the l th convolution layer is $2 \times W^{[l]}(C_{\text{in}}^{[l]}S^{[l]} + 1)C_{\text{out}}^{[l]}$. And the FLOPs of RFFsNet-SEI can be calculated as

$$F_{\text{conv}}^{[l]} = \sum_{l=1}^{L} 2 \times W^{[l]}(C_{\text{in}}^{[l]}S^{[l]} + 1)C_{\text{out}}^{[l]}. \tag{24}$$

Let I_{fc} and O_{fc} refer to the input and output dimensionality of the fully connected layer, then the FLOPs for fully connected layer equals $2(I_{\text{fc}} - 1)O_{\text{fc}}$. Therefore, the whole computational cost of the proposed architecture is

$$F_{\text{FLOPs}} = \sum_{l=1}^{L} 2 \times W^{[l]}(C_{\text{in}}^{[l]}S^{[l]} + 1)C_{\text{out}}^{[l]} + 2(I_{\text{fc}} - 1)O_{\text{fc}}. \tag{25}$$

Equations (23) and (25) imply that the computation complexity of RFFsNet-SEI are mainly affected by the input feature maps, the convolution kernel size, the output channels, the input and output dimensionality of the fully connected layer and each convolution layer, as well as the network depth. Therefore, we design the explicated network structure of RFFsNet-SEI in Table 1 based on these factors, the effectiveness of which is confirmed by virtue of experiment in Section 4.

With the increase of network depth, there will be gradient disappearance and gradient explosion, which degrades the recognition performance. Therefore, we introduce residual connection used in the proposed network to overcome this problem, which has been successfully used in our previous work [30].

3.5 Dataset setup

The pseudo code procedures for dataset setup are given as follows.

Input parameters: Fix the length of the segmented signal Z, the number of classes of emitters M, the dataset size P, and the number of IMFs Q.

(i) Compute temporal $\boldsymbol{v}_q$ and spectral modes $\hat{\boldsymbol{v}}_q$ of all received signal sequences for q=1,2,$\cdots$,Q using (2). The details procedures of VMD are calculated with Algorithm 1 in [13].

(ii) Compute Hilbert spectral $H_q(\omega, m)$ from the decomposed VMD temporal modes $\boldsymbol{v}_q$ using (3) for q=1,2,$\cdots$,Q.

(iii) Compute temporal features (EF, FOM, and SOM) from the decomposed VMD modes using (5), (7) and (8) for all the signal sequences.

(iv) Compute averaged spectral IMF modes from $\boldsymbol{v}_q$ using

$$\hat{\boldsymbol{v}} = \frac{1}{Q}\sum_{q=1}^{Q}\hat{\boldsymbol{v}}_q.$$

(v) Compute spectral features (SR and SB) from the decomposed VMD modes using (9) and (10) for all the signal sequences.

(vi) Combine the feature vector of training sequence $\boldsymbol{\lambda}_p$= [EF$_p$, FOM$_p$, SOM$_p$, SF$_p$, SB$_p$]$^{\mathrm{T}}$.

(vii) Let $\{[\boldsymbol{x}_p, \hat{\boldsymbol{v}}_p, \boldsymbol{v}_p, \boldsymbol{\lambda}_p], \boldsymbol{y}_p\}$ be the pth training sample, where $\boldsymbol{y}_p$ denotes the class label of emitters.

After repeating Step (i) to Step (vii), the dataset is constructed. Finally, the dataset is randomly split into training (80%) and test sets (20%) to train and test RFFsNet-SEI, respectively.

4. Numerical results

In this section, we conduct comparison for the proposed RFFsNet-SEI and the others including RFFs-SVM [11], VMD-KNN [13], MApEn [14], MDFFIL [17], BiSpectral-CNN [18], CNN-1D [19], DRN [20], as well as MFFDEL-SEI [22]. Experiments are verified on two datasets, namely, the simulation dataset and the real dataset. As typical transmitter RFFs are originated from the power amplifying process of analog elements, so we just concentrate our attention on PA in this paper.

4.1 Data generation

(i) Simulation data: To verify the adaptability of the proposed method to the environment, a simulation dataset consisting of monotone signals is generated under different SNR levels. We use the Taylor polynomial to model different power amplifiers. The system response of PA is described in a memory polynomial model [31] as follows:

$$r(n) = \begin{cases} \varepsilon(n), \ 1 \leqslant n \leqslant N_0 \\ \displaystyle\sum_{v=0}^{V}\sum_{u=1}^{U} h_v \chi_{2u-1} |x(n-v)|^{2(u-1)} x(n-v), \\ \qquad N_0+1 \leqslant n \leqslant N_0+Z \end{cases} \tag{26}$$

where $r(n)$ is the data sample at time instant n, and Z is the number of I-Q points for each sample (Z=64 in simulation). N_0 is the starting point of the transient signal, which is the rise edge start point of the pulse signal in simulation. In (26), h_v, χ_{2u-1}, V, and U are the memory coefficients, nonlinear coefficients, memory depth and order of the polynomial, respectively. In this paper, V and U are set as 2 and 3, respectively. The details of h_v and χ_{2u-1} for different PA coefficients are shown in Table 2.

Table 2 Details of power amplifier coefficients

Emitter	V=2			U=3		
	h_0	h_1	h_2	χ_1	χ_3	χ_5
E1	1.00	−0.36	−0.36	1.00	−0.36	−0.36
E2	1.00	−0.27	−0.27	1.00	−0.27	−0.27
E3	1.00	−0.18	−0.18	1.00	−0.18	−0.18
E4	1.00	−0.09	−0.09	1.00	−0.09	−0.09
E5	1.00	0.00	0.00	1.00	0.00	0.00
E6	1.00	0.09	0.09	1.00	0.09	0.09
E7	1.00	0.18	0.18	1.00	0.18	0.18
E8	1.00	0.27	0.27	1.00	0.27	0.27
E9	1.00	0.36	0.36	1.00	0.36	0.36
E10	1.00	0.45	0.45	1.00	0.45	0.45

The input signal to the power amplifier is written as

$$x(n) = \mathrm{rect}(nT_s/T)\mathrm{e}^{\mathrm{j}2\pi nfT_s}. \tag{27}$$

where rect($\cdot$) denotes the rectangle baseband modulated signal with width T, T_s is the sampling period, and f=1 575 MHz is the carrier frequency. Additionally, the signal passes through various SNRs from 0 dB to 30 dB with a step size of 3 dB. Each signal is sampled as 125 MHz.

We have simulated a total of 10 emitters, which are marked as E1−E10, respectively. The number of training and test samples is set to 24 000 and 6 000 for each emitter. For each sample, the first five I-Q points are noise, the subsequent ten I-Q points are transient signal, and residual points are steady-state signals. Therefore, both transient signal and steady signal are fed into the network for each sample.

(ii) Real-world data: A real-data acquisition system is deployed in the anechoic chamber. The interconnection diagram is shown in Fig.11, which is designed to collect signals from eight PAs cascaded by the same vector signal generator. Two rod antennas are used for transmitters and receivers and eight emitters are marked as E1−E8, respectively. In order to reduce the difference of power amplifiers, all these PAs come from the same manufacturer, factory batch, and working mode act as objective emitters. an RF receiver connected to a rod antenna converts the RF signals to the intermediate frequency (IF) and a digital acquisition card is used to collect the zero-IF I-Q data. Then the captured signals are stored via 1 000 Base-T Ethernet in the digital format for further processing. For each PA, the number of training and test samples is also set to 24 000 and 6 000, respectively. The real-world data generation and acquisition system configurations are listed in Table 3−Table 6, respectively.

The performance of RFFsNet-SEI is analyzed in terms of probability of correct identification. Let $A_{cc}(i'/i)$ denote the probability of identifying a received signal as a signal from the i'th emitter, when the signal from the ith emitter has been transmitted. Here, i and i' belong to the emitter set {E1, E2,⋯,EM}. The $M \times M$ matrix is formed where diagonal element represents the probability of correct classification $A_{cc}(i'/i)$ for the ith emitter. Considering all the M emitters to be equiprobable, the aggregate probability of correct classification can be given as

$$A_{cc} = \frac{1}{M}\sum_{i=1}^{M} A_{cc}(i'/i). \tag{28}$$

In our experiments, the RFFsNet-SEI is implemented within the Tensorflow1.12 framework and the network is trained on NVIDIA GeForce GTX 2080 GPU using the Adam optimizer, and the learning rate is 0.001. The total number of epochs is 100. In the prediction, all experiments are conducted on the 128GB RAM and two Intel Xeon E5-2678v3@2.50 GHz CPUs.

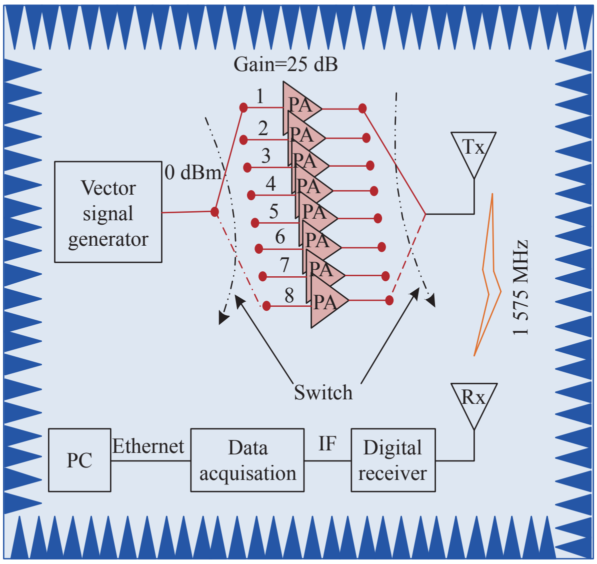


Fig. 11 Signal acquisition system setup for SEI

Table 3 Vector signal generator specifications

Type	Power output/dBm	Waves type	Pulse repetition interval/μs	Duty ratio	RF output/MHz
E4438C	0	Pulse wide modulation	50	26/50	1 575

Table 4 PA specifications

Frequency range	Power supply voltage	Maximum power output/dBm	Gain	System impedance/Ω
0.1 MHz~2 GHz	6-12 VDC	+10	25 dB @ 1500 MHz	50

Table 5 RF-receiver specifications

Frequency range	Tuning frequency/MHz	Bandwidth/MHz	Attenuation/dB	Work mode	IF value/MHz
30 MHz~3 GHz	1 575	4	0	normal	70

Table 6 Digital acquisition card specifications

Sampling rate	Quantization bit	DDC value	Sampling duration/s	I-Q rates/MHz
250 MSPS	16	2	5	125

4.2 Identification experiments

4.2.1 Effectiveness verification of RFFsNet-SEI

In this experiment, both simulation dataset (given by SNR=15 dB) and real-data dataset are exploited to validate the performance of RFFsNet-SEI. The trainable parameters of neural network, the computational cost, and the prediction speed are denoted by Para., MFLOPs (million floating-point operations per second), and FPS (frame per second), respectively. In order to illustrate the effectiveness of the proposed neural network, we show the design process of RFFsNet-SEI in an incremental manner.

S1: For simplicity, we denote the simplified structure of RFFsNet-SEI without ResBlock1, ResBlcok2 and two auxiliary CBL blocks in Fig. 3 as the RFFsNet-SEI-Basic. In RFFsNet-SEI-Basic, the number of channel n_C is 16. The performance illustrates that it is feasible to take four CBF blocks to fit the functions of VMD and Hilbert transform and obtain the features including EF, FOM, SOM and SF. However, the drawback of RFFsNet-SEI-Basic is the poor identification accuracy. The identification accuracy A_{cc} on simulation dataset for 10 emitters is only 49.1% while A_{cc} on real dataset is 21%, which indicates that the individual differences of eight PAs are more subtle than the simulated version. The prediction speed of emitter identification reaches 15366 FPS, which indicates that RFFsNet-SEI-Basic can meet the needs of high real-time processing.

S1→S2: Noted that temporal modes and spectral modes of signal by using VMD are separately fitted in RFFsNet-SEI-Basic, however, the VMD map capacity is limited with two parallel paths. To enhance the feature extraction ability of RFFsNet-SEI-Basic, the ResBlock1 is deployed in front of temporal modes and spectral modes extracting blocks. The identification accuracy of RFFsNet-SEI-Basic including ResBlock1 is improved as much as 9.2% on simulation dataset and 18.2% on real dataset, which illustrates that the deployed ResBlock1 is effective in the neural network. However, the negative effect is that both the trainable parameters of neural network and the computational cost increase, and the prediction speed decreases.

S2→S3: Similarly, temporal features (EF, FOM, SOM) and spectral features (SF, SB) of VMD modes cannot be fitted perfectly in RFFsNet-SEI-Basic plus ResBlock1. Therefore, the ResBlock2 is deployed in front of temporal features and spectral features extracting blocks. The identification accuracy of RFFsNet-SEI-Basic including ResBlock1 and ResBlock2 is further improved as much as 18.9% on simulation dataset and 38.9% on real dataset, which illustrates that ResBlock2 is effective in the RFFsNet-SEI. It shows that the information fusion mechanism of ResBlock2 can improve the feature extraction. It also illustrates that Loss $2(\cdot)$ function in (19) plays a significant effect on improving the identification accuracy.

S3→S4: Although ResBlock1 improves the fitting ability about the VMD modes of signal, fitting errors still occur. Meanwhile, although ResBlock2 enhances the representation ability of features, feature information is still losing with limited features. Therefore, two auxiliary CBFs are additionally deployed in the RFFsNet-SEI, which decreases the fitting residual of signal modes and improves the utilization of feature information. By imposing auxiliary CBFs, the identification accuracy is improved by 2.5% on the simulation dataset and 4.2% on the real dataset, respectively. Here it must be pointed out that the results in terms of identification accuracy seem to violate the common sense because the accuracy tested on simulation dataset is worse than that of real dataset. A possible reason is that ten emitters are considered in the simulation dataset while only eight PAs are tested in the real dataset.

S5→S6: Additionally, an appropriate number of channels is effective to balance the network size and the computational complexity. Therefore, we evaluate the identification accuracy under different channel numbers in this step. It is found that when the number of channels is eight, the computational complexity will be reduced, but the accuracy will also deteriorate. On the contrary, we find that even if the number of channels increases by 32, the identification accuracy is almost unchanged, but the network size increases significantly. Therefore, the number of channels of RFFsNet-SEI-Basic n_C is set to 16.

Besides, in the RFFsNet-SEI designing, we verify the identification accuracy of different numbers of ResUnit in ResBlock1 and ResBlock2 on the simulation dataset. When the number of ResUnit increases from 1 to 5, the recognition accuracy is 60.2%, 70.3%, 79.7%, 79.9% and 79.8%, respectively. It can be seen that in order to balance the network complexity and prediction accuracy, it is appropriate to set the number of ResUnit to 3. In addition, we also verify the identification accuracy of other skills such as sequeze-and-excitation block [32], label smooth [33], and dropout [34]. It is found that the improvement of accuracy is the minimal. Therefore, considering the tradeoff between computational complexity of network and its identification accuracy, the final RFFsNet-SEI network only adopts the design skills in Table 7.

Table 7 Effectiveness verification of RFFsNet-SEI

Step	Note	MFLOPs	Para.	FPS	Simulation-data A_{cc}/%	Real-data A_{cc}/%
S1	RFFsNet-SEI-Basic	601.5	12976	15366	49.1	21.0
S2	+ ResBlcoks1	1756.5	18384	10380	58.3	39.2
S3	+ResBlcoks2	1993.4	23872	7135	77.2	78.1
S4	+auxiliary CBL blocks	2627.2	26730	5645	79.7	82.3

Continued

Step	Note	MFLOPs	Para.	FPS	Simulation-data A_{cc}/%	Real-data A_{cc}/%
S5	n_C:16→8	692.0	7514	9851	76.0	78.1
S6	n_C:16→32	10266.8	100874	2687	79.6	82.3

4.2.2 Comparative analysis to baselines under different SNR cases

(i) Identification accuracy of different methods

This experiment is implemented under the same simulation dataset for the case of ten emitters for all methods. The identification accuracy A_{cc} of different methods against SNR values is shown in Fig. 12.

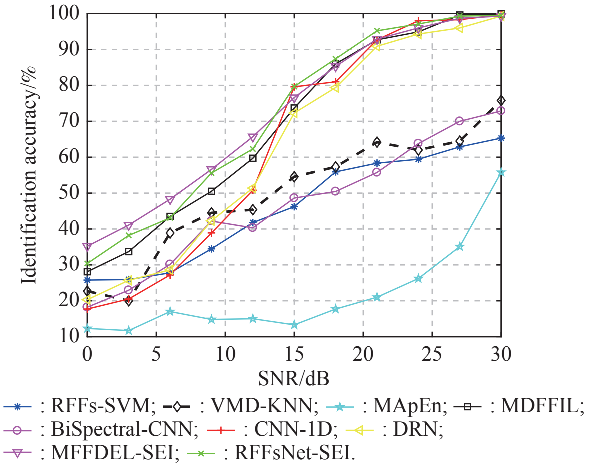


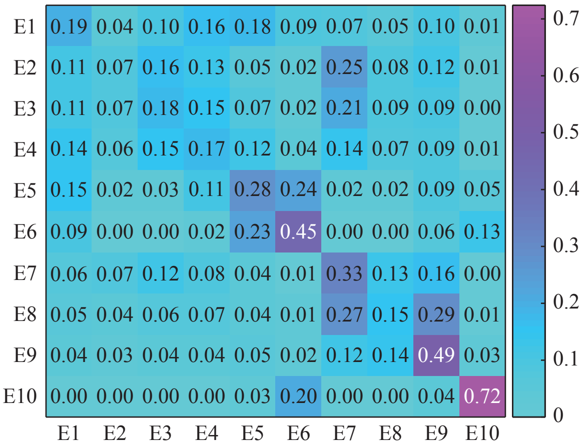
Fig. 12 Identification accuracy versus SNRs

As shown in Fig.12, the recognition accuracy of all methods shows an overall upward trend with the increase of SNR. For RFFs-based methods, the accuracy of identification is less than 60% even if SNR reaches 18 dB. And the identification rates of VMD-KNN, Bi-Spectral CNN, and RFFs-SVM are close to 75.9%, 72.9%, and 65.4% respectively for SNR=30 dB and far behind the identification accuracy of RFFsNet-SEI method, CNN-1D method, DRN method, MDFFIL method, and MFFDEL-SEI method, which indicates the advantages of the data-driven methods. Regardless of SNR, the MApEn method is the worst. The reason is that MApEn requires a large number of I-Q samples (typically thousands of I-Q samples) is needed for good performance. However, only 64 I-Q points for each sample is used in the experiment.

In addition, although the CNN-1D method and DRN method, the MDFFIL method, the and MFFDEL-SEI method have approximated accuracy with the RFFSNet-SEI method when the SNR exceeds 20 dB, their recognition accuracy is far lower than that of the proposed method except for the MFFDEL-SEI method when SNR is small (for example, SNR<15 dB). The reason is that RFFs-SEI has stronger feature identification ability than the other three methods in the low SNR scenario because of the explainable physical features. To some extent, this experiment shows that our proposed neural network is more effective than other data-driven methods (i.e., MDFFIL method [17], CNN-1D method [19] and DRN method [20]).

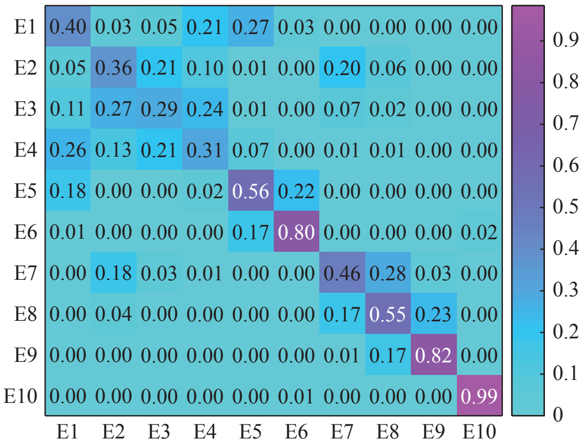
(ii) The classification confusion matrix of RFFsNet-SEI

On the simulation dataset of 10 emitters, this experiment analyzes the identification of RFFsNet-SEI through the classification confusion matrix, whose diagonal elements represent the correct recognition rate A_{cc} (i'/i) of each class. The details of confusion matrix at SNR=0 dB, 9 dB, 18 dB, and 27 dB are shown in Fig. 13−Fig. 16, respectively. In Fig.13, the SNR is 0 dB. The accuracy of the other nine emitters except E10 is less than 50%, especially for E2, the accuracy is only 7%, which shows that our method has the ability of preliminary feature extraction, rather than distinguishing emitters through random guess.



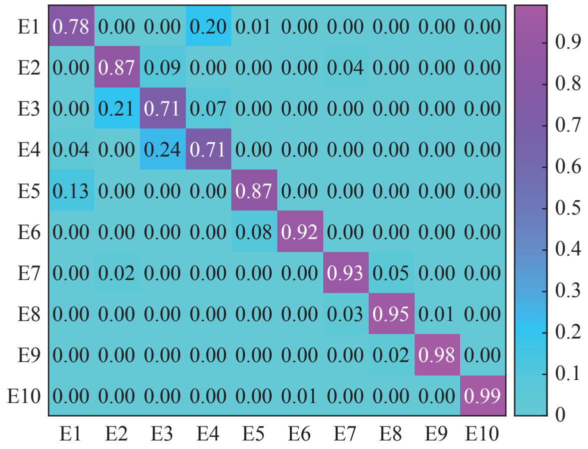
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
E1	0.19	0.04	0.10	0.16	0.18	0.09	0.07	0.05	0.10	0.01
E2	0.11	0.07	0.16	0.13	0.05	0.02	0.25	0.08	0.12	0.01
E3	0.11	0.07	0.18	0.15	0.07	0.02	0.21	0.09	0.09	0.00
E4	0.14	0.06	0.15	0.17	0.12	0.04	0.14	0.07	0.09	0.01
E5	0.15	0.02	0.03	0.11	0.28	0.24	0.02	0.02	0.09	0.05
E6	0.09	0.00	0.00	0.02	0.23	0.45	0.00	0.00	0.06	0.13
E7	0.06	0.07	0.12	0.08	0.04	0.01	0.33	0.13	0.16	0.00
E8	0.05	0.04	0.06	0.07	0.04	0.01	0.27	0.15	0.29	0.01
E9	0.04	0.03	0.04	0.04	0.05	0.02	0.12	0.14	0.49	0.03
E10	0.00	0.00	0.00	0.00	0.03	0.20	0.00	0.00	0.04	0.72

Fig. 13 Classification confusion matrix of RFFsNet-SEI at SNR=0 dB



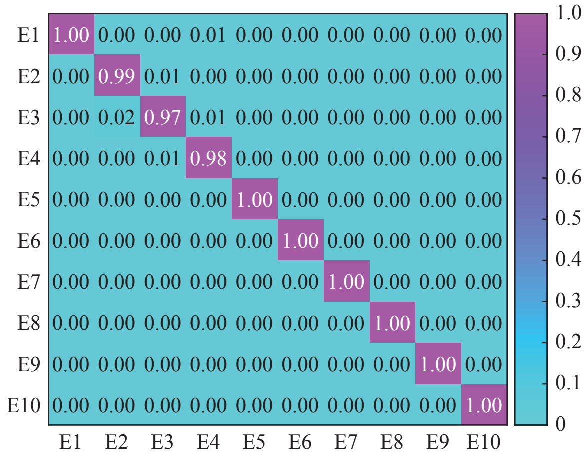
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
E1	0.40	0.03	0.05	0.21	0.27	0.03	0.00	0.00	0.00	0.00
E2	0.05	0.36	0.21	0.10	0.01	0.00	0.20	0.06	0.00	0.00
E3	0.11	0.27	0.29	0.24	0.01	0.00	0.07	0.02	0.00	0.00
E4	0.26	0.13	0.21	0.31	0.07	0.00	0.01	0.01	0.00	0.00
E5	0.18	0.00	0.00	0.02	0.56	0.22	0.00	0.00	0.00	0.00
E6	0.01	0.00	0.00	0.00	0.17	0.80	0.00	0.00	0.00	0.02
E7	0.00	0.18	0.03	0.01	0.00	0.00	0.46	0.28	0.03	0.00
E8	0.00	0.04	0.00	0.00	0.00	0.00	0.17	0.55	0.23	0.00
E9	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.17	0.82	0.00
E10	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.99

Fig. 14 Classification confusion matrix of RFFsNet-SEI at SNR=9 dB



	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
E1	0.78	0.00	0.00	0.20	0.01	0.00	0.00	0.00	0.00	0.00
E2	0.00	0.87	0.09	0.00	0.00	0.00	0.04	0.00	0.00	0.00
E3	0.00	0.21	0.71	0.07	0.00	0.00	0.00	0.00	0.00	0.00
E4	0.04	0.00	0.24	0.71	0.00	0.00	0.00	0.00	0.00	0.00
E5	0.13	0.00	0.00	0.00	0.87	0.00	0.00	0.00	0.00	0.00
E6	0.00	0.00	0.00	0.00	0.08	0.92	0.00	0.00	0.00	0.00
E7	0.00	0.02	0.00	0.00	0.00	0.00	0.93	0.05	0.00	0.00
E8	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.95	0.01	0.00
E9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.98	0.00
E10	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.99

Fig. 15 Classification confusion matrix of RFFsNet-SEI at SNR=18 dB



	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
E1	1.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
E2	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E3	0.00	0.02	0.97	0.01	0.00	0.00	0.00	0.00	0.00	0.00
E4	0.00	0.00	0.01	0.98	0.00	0.00	0.00	0.00	0.00	0.00
E5	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
E6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
E7	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
E8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
E9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
E10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Fig. 16 Classification confusion matrix of RFFsNet-SEI at SNR=27 dB

When the SNR is increased to 9 dB, the overall recognition accuracy of 10 emitters is increased to 55%, especially the recognition accuracy of E10 is up to 99%, which indicates that E10 has obvious characteristic differences with the other nine emitters. When the SNR is increased to 18 dB, the emitters recognition accuracy of E6−E10 is more than 90%. Meanwhile, there is still a ratio of recognition errors between E1−E4, and the overall recognition accuracy is greatly improved to 87%. When the SNR reaches 27 dB, except for a few E2 and E3 recognition errors, the overall recognition accuracy exceeds 99.4%, which indicates that the proposed method can realize high-precision recognition for emitters.

4.2.3 Comparative analysis to the baseline methods under different emitter numbers on the real dataset

This experiment evaluates the recognition accuracy of the proposed method and the comparison methods under different number of emitters. The experimental results are shown in Fig.17 . Except for the RFFsNet-SEI method, the identification performance of all methods decreases significantly with the increase of the number of emitters. The RFFs-SVM method is better than the BiSpectral-CNN method and the VMD-KNN method, but far behind the RFFsNet-SEI method, the MDFFIL method, and the DRN method, which also shows the effectiveness of the proposed method.

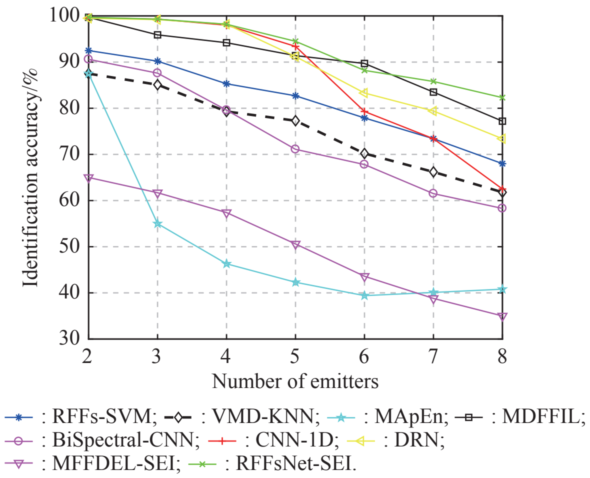


Fig. 17 Identification accuracy versus different number of emitters

Meanwhile, we can see that the accuracy of all data-driven methods seams declining obviously when the number of emitters exceeds four. The reason is that the identification accuracy in the experiment is calculated by the average value of multiple emitters. If an emitter with low identification accuracy exists in the test dataset, and the individual difference between the selected one and the rest emitters in the test dataset is very small, the newly selected emitter will decrease the recognition accuracy of all tested emitters. It ultimately leads to a significant decline in terms of identification accuracy.

Moreover, we can also see from Fig.17 that when the number of emitters exceeds four, recognition accuracy of the RFFsNet-SEI method is the best. The identification accuracy of the RFFsNet-SEI method achieves 82.3% or 99.7% when identifying 8 PAs or 2 PAs, respectively. However, the identification accuracy of the other model-driven methods fluctuates within 40%−92.5% when the number of emitters ranges from 2 to 8. Regardless of the number of emitters, the MApEn method is the worst. There are two reasons for the phenomenon. On the one hand, MApEn cannot utilize enough steady-state information to identify emitters because each sample is formed from 64 I-Q data points. On the other hand, since the sampling rate of data acquisition system is not high enough, which makes the transient information of received signals cannot be exploited by MApEn. Meanwhile, according to the results of Fig.17, we can see that the MFFDEL-SEI method has poor accuracy on the real dataset, which implies that the data-driven method with good recognition accuracy on the simulation dataset cannot be directly extended to the real dataset.

Besides, as far as the RFFsNet-SEI method, the CNN-1D method, and the DRN method are concerned, when the number of transmitters exceeds four, the identification accuracy of the CNN-1D method is the worst and the RFFsNet-SEI method has the best identification accuracy, which shows the effectiveness of the proposed method.

4.3 Computational time analysis

In this experiment, the computational complexity of all SEI methods is evaluated on the real dataset. We firstly fix the selected number of IFMs according to the identification accuracy A_{cc} and prediction speed (FPS) between RFFsNet-SEI and VMD-KNN method, comparisons of which are shown in Table 8. The 1st entry and the 2nd entry in brackets is the FPS and recognition accuracy, respectively.

Table 8 Identification accuracy and prediction speed comparisons versus different number of IFMs

Number of IMFs	RFFSNET-SEI	VMD-KNN
2	(6 002, 78.7%)	(99, 56.7%)
3	(5 875, 79.1%)	(91, 59.2%)
4	(5 645, 79.7%)	(84, 61.5%)
5	(5 284, 79.4%)	(79, 60.9%)

It shows from Table 8 that the highest identification accuracy of the RFFsNet-SEI method is achieved when four IMFs are used. Therefore, four IFMs are chosen to evaluate the computational time complexity of RFFsNet-SEI and VMD-KNN methods in this paper. In this case, the proposed method consumes just about 1/67 of the prediction time required by the VMD-KNN method.

Based on above settings, we compare the identification accuracy, prediction speed, computational complexity for model-driven methods, and the number of parameters for data-driven methods, which are listed in Table 9. It can be seen from Table 6 that the identification speeds of data-driven SEI methods are much faster than those of model-driven SEI methods. That is because the latter (i.e. model-driven SEI methods) require complex feature extraction process, while the former directly extract features from I-Q data. Meanwhile, as far as the recognition accuracy is concerned, the data-driven methods perform SEI by learning the distribution of data sets, which have stronger adaptability to the environment compared with the model-driven methods. Thus the recognition accuracy of data-driven methods is better than that of model-driven methods. According to the results of Table 9, we can see that although the RFFsNet-SEI method is not the best in terms of prediction speed and amount of computation, it has optimal identification accuracy with nice FPS. Combined with the results of Fig.12 and Fig.17, our proposed method is the optimal choice for all compared SEI methods.

Table 9 Comparison in terms of prediction speed, accuracy, network size, and complexity of different methods

Criterion	RFFS-SVM [11]	VMD-KNN [13]	MAPEN [14]	MDFFIL [17]	BISPECTRAL-CNN [18]	CNN-1D [19]	DRN [20]	MFFDEL-SEI [22]	RFFSNET-SEI (Proposed)
FPS	171	84	144	413	372	35643	8969	1073	5 645
A_{cc}/%	64.0	61.5	13.3	73.7	45.3	79.6	72.3	76.6	79.7
MFLOPs	–	–	–	14318	77536	54	1684	2160	2627
Parameters	–	–	–	719066	1246314	28112	105884	258856	26730

5. Conclusions

In this paper, an accelerated emitter identification approach is proposed based on the hybrid driven scheme. A new RFF-embedding framework is designed for SEI as an alternative to the traditional method based on fixed-RFF, which not only has a high identification accuracy, but also achieves fast efficiency. The experimental results illustrate that the new mechanism exploiting the expertise RFFs and other subtle features of the received signal outperformsthestate-of-the-artmanually-designedmethods. The identification accuracy, adaptive capability of environment, and prediction speed of the proposed method have also been validated quantitatively. One of our future works will practically implement the RFFsNet-SEI on the embedding platform like FPGAs or software defined radio (SDR) platforms. Another potential future work includes analyzing other RFFs (besides these RFFs adopted in this paper) to distinguish more PA devices and introducing the generative adversarial network to solve the small sample problem in SEI. The source codes of RFFsNet-SEI and supporting datasets are available at https://github.com/ fanrongca/RFFsNet-SEI.

References

[1] XU S H. On the identification technique of individual transmitter based on signal prints. Wuhan: Huazhong University

of Science and Technology, 2007. (in Chinese)

[2] LI D, LIANG Q H, LIU H Q, et al. A novel multidimensional domain deep learning network for SAR ship detection. IEEE Trans. on Geoscience and Remote Sensing, 2022, 60: 5203213.

[3] SANKHE K, BELGIOVINE M, ZHOU F, et al. No radio left behind: radio fingerprinting through deep learning of physical-layer hardware impairments. IEEE Trans. on Cognitive Communications and Networking, 2020, 6(1): 165–178.

[4] CHEN H, YANG J N, LIU H. Communication transmitter individual identification based on deep residual adaptation network. Systems Engineering and Electronics, 2021, 43(3): 603–609. (in Chinese)

[5] YUAN H L, HU A Q. Preamble-based detection of Wi-Fi transmitter RF fingerprints. Electronics Letters, 2010, 46(1): 1165–1167.

[6] SUN L T, WANG X, HUANG Z T, Unintentional modulation microstructure enlargement. Journal of Systems Engineering and Electronics, 2022, 33(3): 522−533.

[7] BERTONCINI C, RUDD K, NOUSAIN B, et al. Wavelet fingerprinting of radio-frequency identification (RFID) tags. IEEE Trans. on Industrial Electronics, 2005, 59(12): 4843–4850.

[8] LOPEZ G, GRAJAL J, OSORIO A, et al. Digital channelized receiver based on time-frequency analysis for signal interception. IEEE Trans. on Aerospace and Electronic Systems, 2005, 41(3): 879–898.

[9] LUNDEN J, KOIVUNEN V. Automatic radar waveform recognition. IEEE Journal of Selected Topics in Signal Processing, 2007, 1(1): 124−136.

[10] LIU Z G, ZHANG Q G. An approach to recognize the transient disturbances with spectral kurtosis. IEEE Trans. on Instrumentation and Measurement, 2014, 63(1): 46–55.

[11] YUAN Y J, HUANG Z T, WU H, et al. Specific emitter identification based on hilbert-huang transform-based time-frequency-energy distribution features. IET Communication, 2014, 8(13): 2404–2412.

[12] HE B, WANG F G. Cooperative specific emitter identification via multiple distorted receivers. IEEE Trans. on Information Forensics and Security, 2016, 15: 3791–3806.

[13] SATIJA U, TRIVEDI N, BISWAL G, et al. Specific emitter identification based on variational mode decomposition and spectral features in single hop and relaying scenarios. IEEE Trans. on Information Forensics and Security, 2019, 14(3): 581–591.

[14] SUN L T, WANG X, YANG A F, et al. Radio frequency fingerprint extraction based on multi-dimension approximate entropy. IEEE Signal Processing Letters, 2020, 27: 471–475.

[15] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning: adaptive computation and machine learning series. Cambridge: MIT Press, 2016.

[16] WANG Y, GUI G, GACANIN H, et al. An efficient specific emitter identification method based on complex-valued neural networks and network compression. IEEE Journal on Selected Areas in Communications, 2021, 39(8): 2305–2317.

[17] QU L Z, LIU H, HUANG K J. Specific emitter identification based on multi-domain feature fusion and integrated learning. Symmetry, 2021, 13(8): 1481.

[18] DING L, WANG S L, WANG F G, et al. Specific emitter identification via convolutional neural networks. IEEE Communications Letters, 2018, 22(12): 2591–2594.

[19] XIAO Y, WEI X Z. Specific emitter identification of radar based on one dimensional convolution neural network. Journal of Physics: Conference Series, 2020, 1550: 032114.

[20] WENG L T R, PENG J L, HE Y, et al. Specific emitter identification of ADS-B signal based on deep residual network. Aero Weaponry, 2021, 28(4): 24–29.

[21] XIAO Z L, YAN Z Y. Radar emitter identification based on novel time-frequency spectrum and convolutional neural network. IEEE Communications Letters, 2021, 25(8): 2634–2638.

[22] LIU Z M. Multi-feature fusion for specific emitter identification via deep ensemble learning. Digital Signal Processing, 2021, 110: 102939.

[23] ZHAO Y R, WANG X, LIN Z Y, et al. Multi-classifier fusion for open-set specific emitter identification. Remote Sensing, 2022, 14: 2226.

[24] ZHUO F, HUANG Y L, CHEN J. Specific emitter identification based on linear polynomial fitting of the energy envelope. Proc. of the IEEE 6th International Conference on Electronics Information and Emergency Communication, 2016: 278−281.

[25] ZHANG J, WANG F, DOBRE O A. Specific emitter identification via Hilbert-Huang transform in single-hop and relaying scenarios. IEEE Trans. on Information Forensics Security, 2016, 11(6): 1192–1205.

[26] WONG L J, HEADLEY S, ANDREWS R , et al. Clustering learned CNN features from raw I/Q data for emitter identification. Proc. of the IEEE Military Communications Conference, 2018: 26−33.

[27] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proc. of the International Conference on Machine Learning, 2015: 448−456.

[28] FAN R, ZHU X Y, TANG W B, et al. MFFNet: multi-path features fusion network for source enumeration. IEEE Communications Letters, 2022, 26(3): 572–576.

[29] GONG J L, XU X D, LEI Y K. Unsupervised specific emitter identification method using radio-frequency fingerprint embedded InfoGAN. IEEE Trans. on Information Forensics and Security, 2020, 15: 2898–2913.

[30] FAN R, SI C K, GUO H S, et al. Direction finding for coherent sources with deep hybrid neural networks. International Journal of Electronics, 2022, 109(5): 811–833.

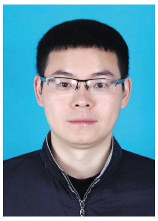
[31] KENINGTON P B. High linearity RF amplifier design. Boston: Artech House, 2000.

[32] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011–2023.

[33] ZHANG C B, JIANG P T, HOU Q B, et al. Delving deep into label smoothing. IEEE Trans. on Image Processing, 2021, 30: 5984–5996.

[34] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014, 15(1): 1929–1958.

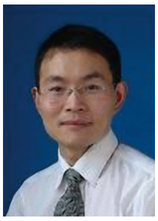
Biographies



FAN Rong was born in 1984. He received his B.S. degree in information engineering from Chengdu University of Technique, Chengdu, China, in 2007, and M.S. and Ph.D. degrees in electronic engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2010 and 2014, respectively. In 2014, he joined the 10th Research Institute of China Electronics Technology Group Corporation. In 2018, he joined the Institute of Electronics and Electrical Engineering, Civil Aviation Flight University of China, Guanghan, China. Since 2020, he has been a postdoctor in the UESTC. His research interests include passive detection, direction finding, and specific emitter identification for air traffic control signals.
E-mail: fanrong@cafuc.edu.cn

SI Chengke was born in 1988. He received his B.S. and M.S. degrees in computer science from University of Science and Technology of China, Hefei, China, in 2011 and 2014, respectively. In 2014, he joined the 10th Research Institute of China Electronics Technology Group Corporation. In 2019, he joined the Institute of Electronics and Electrical Engineering, Civil Aviation Flight University of China, Guanghan, China. His research interests include deep learning and intelligent sensing of radio signals.
E-mail: sichengke@cafuc.edu.cn

HAN Yi was born in 1998. She received her B.S. degree in the Internet of things engineering from Jiangsu University of Technology, Jiangsu, in 2020. She is currently working towards her M.S degree in Institute of Electrical and Electronics Engineering, Civil Aviation Flight University of China, Guanghan, China. Her current research interest is intelligent recognition of UAV radio signals.
E-mail: Han.Holly@Outlook.com



WAN Qun was born in 1971. He received his B.S. degree in electronic engineering from Nanjing University, Nanjing, China, in 1993, and M.S. and Ph.D. degrees in electronic engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1996 and 2001, respectively. From 2001 to 2003, he was a postdoctor with the Department of Electronic Engineering, Tsinghua University. Since 2004, he has been a professor with the Department of Electronic Engineering, UESTC. He is currently the Director of Joint Research Lab of Array Signal Processing and the Deputy Dean of School of Electronic Engineering. His research interests include direction finding, radio localization, and signal processing based on information criterion.
E-mail: wanqun@uestc.edu.cn