Semi-Global Attention and Inter-Class Multi-constraint Enforced Open-Set Radar Emitter Individual Identification

Jinxin XU [1], Jin HU [1], Wenqiang JIANG [1], Zhiqiang ZHANG [1], Mingyuan YANG [1], Zheng ZOU [2]

[1] 724th Research Institute of CSIC, Changqin Street 30, 211153 Nanjing, China
[2] Southeast University, Southeast University Street 2, 2111110 Nanjing, China

2019377631@qq.com

Submitted October 24, 2025 / Accepted February 27, 2026 / Online first March 12, 2026

Abstract. *This paper proposes an open-set radar emitter individual identification algorithm based on semi-global attention and inter-class multi-constraint to address the problem of inadequately adaptive feature representation and discrimination difficulty with out-of-distribution samples. By introducing semi-global attention mechanism combined with dual-path convolutional neural network, individual feature representation ability can be enhanced. The mechanism combines semi-global geometric similarity and learnable similarity. Through adaptively adjusting the attention score and retaining long-distance feature association, more accurate feature details and higher efficiency can be obtained. Furthermore, inter-class multi-constraint based loss function is presented in open-set scenario. By constructing multiple loss functions to constrain the distribution of known class samples in feature space, discriminated feature difference between known and unknown class samples can be achieved. In addition, class-based adaptive threshold selection strategy is proposed to achieve accurate rejection of open-set samples. Experimental results show that the proposed algorithm achieves identification accuracy of 97.55% in closed-set radar emitter data. In open-set scenarios, when maintaining a closed-set recognition rate of 91.78%, the rejection rate achieves 87.64%, which represents the proposed algorithm can effectively discriminate known individuals and unknown ones.*

Keywords

Adaptive threshold, inter-class multi-constrain, open-set identification, radar emitter individual identification, semi-global attention mechanism

1. Introduction

Specific emitter identification (SEI) is a technique that extracts the "fingerprint features" of the emitter signal and then uniquely identifies the emitter. SEI can effectively distinguish the signals of different emitters, thereby providing key intelligence support for related fields, which is highly significant for specific radar emitter individual identification research [1], [2]. This technique is widely studied because of strong adaptability, high security, and relatively low computational overhead [3–6].

Signal feature extraction and utilization is one key portion. The primary objective of feature extraction is to provide useful information for subsequent identification tasks through linear or nonlinear transformations. Effective features can not only fully represent emitter data but also discriminate the differences between different emitter individuals.

Traditional feature extraction method can be mainly divided into three classes, including the time-domain class, the frequency-domain class and the transform class. 1) Time-domain class, such as the feature extraction of rising edges, falling edges and envelope [7], UAMOP (Unintentional Amplitude MOP) curve feature extraction [8], UPMOP (Unintentional Phase MOP) curve feature extraction [9], etc. 2) Frequency-domain class, such as fast Fourier transform [10], power spectrum [11], bispectrum-radon transform [12], high-order statistics [13], etc. 3) Transform class, such as empirical mode decomposition (EMD) [14], variational mode decomposition (VMD) [15], short-time Fourier transform [16], wavelet packet transform [17], spectral quotient transform (SQ) [18], etc. Although these feature extraction methods achieve good performance, most of them suffer from some limitations. First, part of the separable features may be lost during the process, resulting in incomplete key discriminant information. Second, the performance is highly dependent on the prior knowledge and experience accumulation of domain experts with radar emitter data. The dependence causes the methods to show significant limitations when confronted with unknown or new emitter signals.

With the continuous development of deep learning in the field of SEI, emitter individual identification methods that utilize neural networks for feature extraction and iden-

DOI: 10.13164/re.2026.0147

tification show significant advantages. Potential features may be lost when radar emitter data is converted into images. Therefore, researchers generally extract subtle features and perform identification on baseband signals [19], [20]. For example, in [19], time-domain in-phase/quadrature (I/Q) data are directly input into convolutional neural network to identify emitter devices, then the effectiveness of the approach is demonstrated through theoretical analysis and experiments. To deal with long-term dependency problem of network, one-dimensional convolutional neural network (CNN) and long short-term memory network (LSTM) are combined to directly extract features from baseband I/Q sequence data in [20]. Identification performance can be enhanced by the algorithm, which reduces the loss of signal features and model complexity. However, the two methods mentioned still suffer from insufficient feature extraction capability for emitter individuals. The representation may be confused, which results in extracting discriminative individual features difficultly.

The need to resolve these representation ambiguities has led to the adoption of attention mechanisms, which offer a dynamic way to prioritize informative signal regions. By integrating channel, spatial, or self-attention modules, researchers can recalibrate the importance of features in both time and frequency domains [21], [22]. This selective focus effectively amplifies the subtle, idiosyncratic traits of emitters while suppressing redundant noise, thereby overcoming the limitations of static extraction methods. In [23], channel attention mechanism is introduced into a multi-scale CNN, which improves the accuracy of individual identification and reduces the training time. In [24], channel and spatial attention mechanisms are combined for individual identification of communication emitters, demonstrating robust stability while significantly improving identification accuracy. In order to obtain more global features, self-attention mechanism and gated recurrent unit (GRU) are combined to extract comprehensive features to increase the accuracy of individual identification in [25]. In addition, to address the challenges of scarce labeled data and changing environments in real-world scenarios, related research are conducted in the fields of few-shot learning, weakly supervised learning, domain adaptation, and meta-learning on SEI [26 ,27, 28].

Despite the advancements in attention-based models, translating these techniques into robust radar emitter identification remains hampered by several bottlenecks. 1) Lack of comprehensive preprocess mechanism to align non-fixed arrival time, pulse width and amplitude of radar emitter data is a challenge, which increases the impact of such factors on feature extraction. 2) The feature extraction capability for emitter individuals is insufficient, which will compromise identification performance. The features extracted by the attention mechanism under global weight calculation are usually redundant and have high computational complexity. 3) In addition, most of existing algorithms for radar emitter individual identification are based on the assumption of closed-set scenario. These algorithms often cannot adapt to open-set samples in actual tasks, resulting in serious decision errors.

In recent years, a lot of research are conducted on the open-set identification problem. However, there are few studies in the field of radar emitter individual identification [29]. OpenMax based extreme value theory is used to identify high-resolution radar range profiles in the work of [30], which obtains a superior performance improvement compared with traditional open-set identification model. Subsequent research has focused on refining decision boundaries and feature representation for better separation between known and unknown classes. For instance, fixed threshold classifiers and OpenMax classifiers are combined to avoid overlap in feature space to improve the rate of open-set specific emitter identification (OS-SEI) tasks in [31]. The work of [32] presents an open-set identification algorithm based on prototype networks and extreme value theory to identify I/Q signals for emitters, which can effectively improve the rate of open-set identification. Auxiliary classifiers are used in adversarial networks to generate outlier samples, followed by the introduction of OpenMax to obtain calibrated activation vectors. Thresholds are then applied to class confidence scores to accomplish open-set identification in the work of [2]. However, radar emitter data based existing algorithms have limited loss functions and unclear rejection threshold methods, which leads to the aliasing of known and unknown class samples.

Based on the above discussion, this paper introduces an open-set radar emitter individual identification algorithm. By employing semi-global attention and inter-class multi-constraint, the algorithm accurately identifies individuals of known classes while effectively rejecting individuals of unknown classes. The main contributions of this work are as follows:

1) In order to solve the problem of incomplete radar emitter data cleaning mechanism, data preprocessing method based on sample alignment is used to align the arrival time, pulse width and amplitude of each sample. Semi-global attention mechanism is introduced, which combines the semi-global geometric similarity and learnable similarity. By adaptively adjusting the attention score and retaining long-distance feature associations, more accurate feature details and higher computational efficiency can be obtained. The algorithm demonstrates robust closed-set identification performance, which is further introduced into the open-set identification issue in this paper.

2) Inter-class multi-constraint based loss function is presented. By constructing multiple loss functions, namely cross entropy loss, inter-class distance loss, intra-class dot product loss and intra-class distance loss, the distribution of known class samples in the feature space is constrained. The correct classification of known classes is ensured and the distribution of intra-class samples in feature space is made more compact. Then space for open-set samples can be reserved to reduce the overlapping of known and un-

known samples in the feature space. Moreover, class-based adaptive threshold selection strategy is proposed. The strategy selects the feature boundaries of each class based on the classification scores of known class samples, and then sets the open-set rejection threshold. Therefore, identification of known class samples and rejection of unknown class samples can be achieved.

3) This paper conducts experiments based on radar emitter data, which includes three individual radars, each with four frequencies and two pulse widths. In the experiments, analysis of the generalization capability cross-frequency points is first conducted to explore boundary capabilities, which is rarely discussed in existing literature. Second, effectiveness of the proposed algorithm is verified by conducting closed-set and open-set comparative experiments with state-of-the-art identification algorithms. The results show that the proposed algorithm has the best generalization performance and achieves the best identification performance in closed-set scenario. Besides, the algorithm can effectively identify known class individuals and reject unknown class individuals in open-set scenario.

In the remainder of this paper, the illustration of the proposed algorithm is first described in Sec. 2. In Sec. 3, lots of comparative experiments are conducted to verify the effectiveness of the algorithm in this paper. Finally, the conclusions and future prospect of this paper are listed in Sec. 4.

2. Construction of Radar Emitter Individual Identification Model

The working principle diagram of the open-set radar emitter individual identification algorithm proposed in this paper is shown in Fig. 1. The algorithm is divided into two phases: training phase and test phase. In the training phase, features from preprocessed samples are first extracted through the semi-global attention mechanism. The extracted features are highly discriminative, which can solve the problem of insufficient sample feature representation. Second, the known samples are correctly identified and the discrimination among known classes is improved through a variety of loss functions. As a result, discriminated feature difference between known and unknown class samples can be achieved, which is conducive to the identification of unknown and known classes. Thereby the classification risk of known samples and the open space risk of unknown samples can be reduced. Finally, the feature boundary of each class is determined by the classification scores of known samples based on the proposed class-based adaptive threshold selection strategy, thereby establishing the open-set rejection threshold.

In the test phase, the test data is first preprocessed. Second, the classification scores of test samples are calculated by loading the model parameters. Finally, the correct identification of known class samples and the rejection of open-set samples can be achieved by comparing the scores with open-set rejection threshold. In terms of individual identification in closed-set scenario, the open-set identification module is replaced with a global average pooling (GAP) layer, a fully connected (FC) layer, etc. Moreover, the prediction result is directly used as the identification result in the test phase without the need for secondary discrimination through the open-set rejection threshold.

2.1 Sample Alignment Based Data Preprocessing

Guard interval is reserved around each pulse edge during the detection, acquisition, and storage of intermediate frequency signal pulses, which results in non-uniform distribution of various features of acquired samples. Therefore, this subsection proposes a data preprocessing method

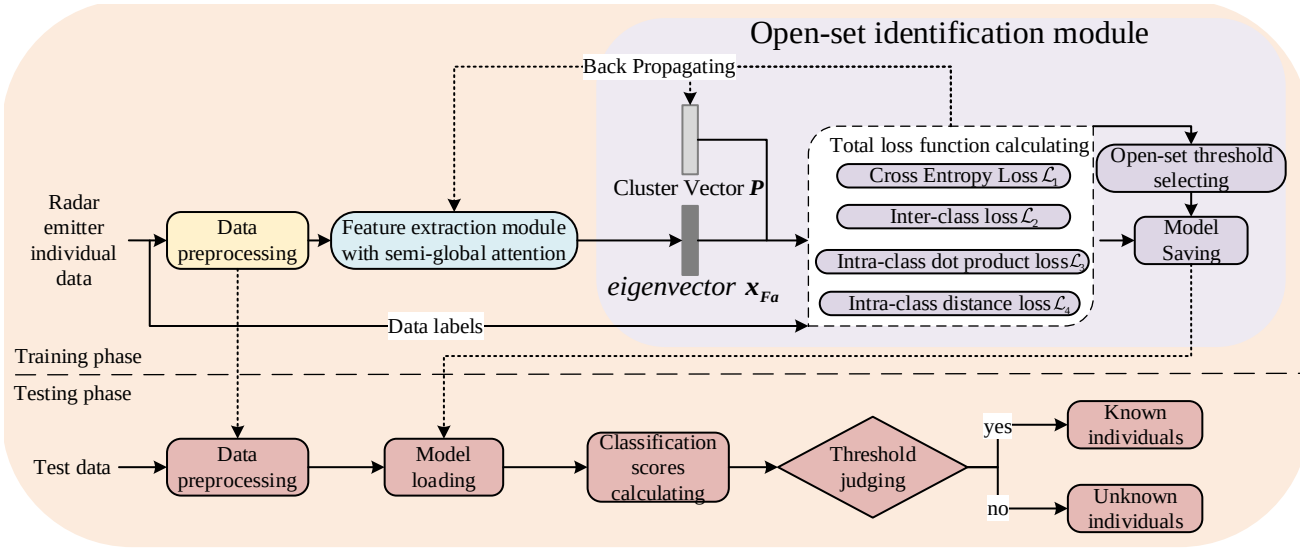


Fig. 1. Illustration of the proposed algorithm in this paper.

based on sample alignment. This method can eliminate the impact of non-uniform signal distribution on subsequent identification, ensuring the accuracy and reliability of the identification process.

First, the received intermediate frequency signal can be defined as $\boldsymbol{x}_{iq} \in R^{1\times N}$, N is the length of the signal, then the amplitude of $\boldsymbol{x}_{iq}$ can be expressed as

$$\boldsymbol{x}_{\text{shipingValue}} = \sqrt{\text{Re}(\boldsymbol{x}_{iq}) \circ \text{Re}(\boldsymbol{x}_{iq}) + \text{Im}(\boldsymbol{x}_{iq}) \circ \text{Im}(\boldsymbol{x}_{iq})} \quad (1)$$

where Re(·) denotes the real part of the signal, Im(·) represents the imaginary part and ∘ denotes element-wise product.

In order to obtain the effective pulse part and align the time among samples, the double thresholds T_1 and T_2 are set:

$$\begin{cases} T_1 = \beta_1 \text{mean}(\boldsymbol{x}_{\text{shipingValue}}) \\ T_2 = \beta_2 \text{mean}(\boldsymbol{x}_{\text{shipingValue}}(\boldsymbol{e}_1)) \end{cases} \quad (2)$$

where mean(·) is signal mean operation. The index of the position in $\boldsymbol{x}_{\text{shipingValue}}$ that is greater than T_1 is recorded as $\boldsymbol{e}_1 = \text{find}(\boldsymbol{x}_{\text{shipingValue}} > T_1)$, where find(·) means finding the pulse position that meets the conditions. Specifically, $\boldsymbol{x}_{\text{shipingValue}}(\boldsymbol{e}_1)$ denotes the extraction of the values corresponding to the index set $\boldsymbol{e}_1$ from $\boldsymbol{x}_{\text{shipingValue}}$, where β_1 and β_2 are weighing parameters.

Then, the pulse position selected by the double threshold is recorded as $\boldsymbol{e}_2 = \text{find}(\boldsymbol{x}_{\text{shipingValue}} > T_2)$. To ensure the samples are aligned in time, $\boldsymbol{e}_2(1) - L_1$ and $\boldsymbol{e}_2(e) + L_1$ are used as the starting and ending points of the selected intermediate frequency signal $\boldsymbol{x}_{iqs}$, respectively, where e represents the index of the last element in $\boldsymbol{e}_2$. Accordingly, $\boldsymbol{x}_{iqs}$ can be expressed as

$$\boldsymbol{x}_{iqs} = \boldsymbol{x}_{iq}(\boldsymbol{l}_1) \quad (3)$$

where $\boldsymbol{l}_1 = \boldsymbol{e}_2(1) - L_1, \ldots, \boldsymbol{e}_2(e) + L_1$. The offset $L_1 = M_g T_S/T_g$, which has pulse protection effect. In this expression, T_S, M_g, T_g denote the sampling interval, the proportional coefficient and the reference time, respectively.

Second, in order to adapt to the input of the deep learning network, the pulse width of the sample $\boldsymbol{x}_{iqs}$ needs to be processed so that the dimension of the sample remains consistent. The real part $\boldsymbol{x}_{\text{real\_select}}$ and imaginary part $\boldsymbol{x}_{\text{imag\_select}}$ of processed sample can be expressed as

$$\begin{cases} \boldsymbol{x}_{\text{real\_select}} = \boldsymbol{x}_{iqs\_\text{real}}(\boldsymbol{l}_2), \boldsymbol{x}_{\text{imag\_select}} = \boldsymbol{x}_{iqs\_\text{imag}}(\boldsymbol{l}_2) & , L_2 \le \text{size}(\boldsymbol{x}_{iqs\_\text{real}}) \\ \boldsymbol{x}_{\text{real\_select}} = [\boldsymbol{x}_{iqs\_\text{real}}, \boldsymbol{0}], \boldsymbol{x}_{\text{imag\_select}} = [\boldsymbol{x}_{iqs\_\text{imag}}, \boldsymbol{0}] & , L_2 > \text{size}(\boldsymbol{x}_{iqs\_\text{real}}) \end{cases} \quad (4)$$

where $\boldsymbol{l}_2 = 1, \ldots, L_2$. $\boldsymbol{x}_{iqs\_\text{real}}$ and $\boldsymbol{x}_{iqs\_\text{imag}}$ are the real and imaginary parts of $\boldsymbol{x}_{iqs}$, respectively. In this context, $\boldsymbol{0} \in R^{1\times(L_2 - \text{size}(\boldsymbol{x}_{iqs\_\text{real}}))}$ denotes a zero vector, where size(·) represents the length of the data. The parameter $L_2 = N_g f_s/f_g$, where f_s, f_g, N_g represent the sampling rate, the reference frequency and the proportional coefficient, respectively.

Finally, $\boldsymbol{x}_{\text{real\_select}}$ and $\boldsymbol{x}_{\text{imag\_select}}$ are normalized respectively and concatenated to get the preprocessed data $\boldsymbol{x}$, expressed as

$$\boldsymbol{x} = \left[\boldsymbol{x}_{\text{real\_select}} / \sqrt{\boldsymbol{x}_{\text{real\_select}} \circ \boldsymbol{x}_{\text{real\_select}}}, \boldsymbol{x}_{\text{imag\_select}} / \sqrt{\boldsymbol{x}_{\text{imag\_select}} \circ \boldsymbol{x}_{\text{imag\_select}}} \right] \quad (5)$$

where $\boldsymbol{x} \in R^{1\times N_1}$ and the data length $N_1 = 2L_2$.

2.2 Semi-Global Attention Mechanism Based Feature Extraction

Illustration of semi-global attention mechanism based feature extraction is shown in Fig. 2. The framework mainly consists of two parts: a dual-path convolution layer and a semi-global attention mechanism layer. The dual-path convolution layer can extract multi-scale features and the

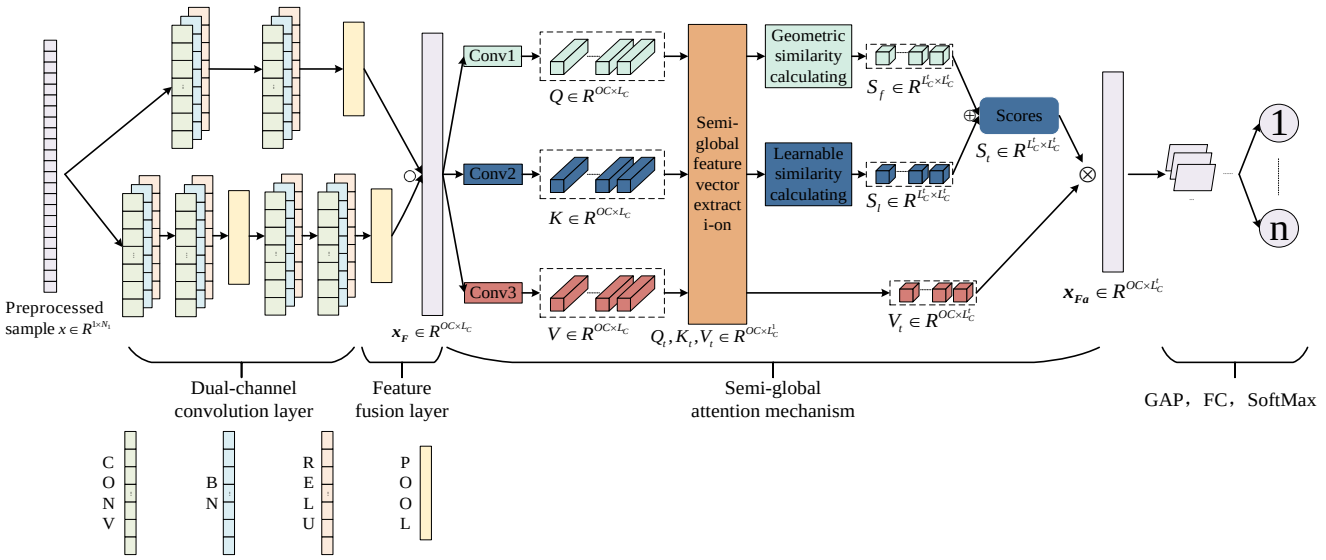


Fig. 2. Illustration of semi-global attention mechanism based feature extraction.

semi-global attention mechanism layer can extract highly discriminative individual features of radar emitters. First, a multi-scale feature space is constructed through a dual-path convolution layer. Then feature enhancement is achieved using the semi-global attention mechanism. Finally, individual identification can be performed through extracted discriminative features.

2.2.1 Dual-Path Convolution Layer

The network contains two independent feature extraction paths which extract feature information from input samples. Specifically, the first path employs smaller convolution kernels combined with nonlinear activation functions to extract local detail features. On the other path, larger convolution kernels are used to extract features which can enhance the robustness to noise and retain the long-range features within samples.

First, the *l*-th convolution operation in the *level*-th path is defined as follows

$$\boldsymbol{F}^{level,l} = \boldsymbol{x}^{level,l} \otimes \boldsymbol{Kernel}^{level,l} + \boldsymbol{b}_K \tag{6}$$

where $level \in \{1,2\}$ represents the first or second path. The terms $\boldsymbol{F}^{level,l}$ and $\boldsymbol{Kernel}^{level,l}$ represent the output features and convolution kernels of the convolution layer in the *level*-th path, respectively. Additionally, $\boldsymbol{x}^{level,l}$ is the input of the *l*-th convolution layer in the *level*-th path, and $\boldsymbol{b}_K$ denotes the bias term of the corresponding convolution kernel. Specifically, the dimensions of the output features are given by $\boldsymbol{F}^{level,l} \in R^{OC^{level,l} \times L_C^{level,l}}$, where $OC^{level,l}$ and $L_C^{level,l}$ indicate the number of output channels and the output length of the corresponding convolution layer, respectively.

Second, the normalized features $\tilde{\boldsymbol{F}}^{level,l}$ can be obtained through batch normalization (BN) layer and activation function layer:

$$\tilde{\boldsymbol{F}}^{level,l}(h,m) = \sigma\left(\frac{\gamma^k(\boldsymbol{F}^{level,l}(h,m) - \mathrm{E}(\boldsymbol{F}^{level,l}(h,m)))}{\sqrt{\mathrm{Var}(\boldsymbol{F}^{level,l}(h,m)) + \varepsilon}} + \beta^k\right) \tag{7}$$

where *h*, *m* represent the *m*-th element of the *h*-th channel in $\boldsymbol{F}^{level,l}$, respectively. The variables γ^k and β^k are the learnable parameters in the *h*-th channel, and ε is a constant. Additionally, E(·) and Var(·) represents the mean and variance, respectively, while $\sigma(\cdot)$ is ReLU activation function.

The outputs of the two paths are $\boldsymbol{x}_1 = \tilde{\boldsymbol{F}}^{1,end1}$, $\boldsymbol{x}_2 = \tilde{\boldsymbol{F}}^{2,end2}$ where $\boldsymbol{x}_1, \boldsymbol{x}_2 \in R^{OC \times L_C}$, while OC and L_C represent the number of output channels and length, respectively. Furthermore, *end*1 and *end*2 represent the last operation in the corresponding path.

Finally, the two output vectors are directly fused through element-wise product, the feature vector obtained by dual-path convolution layer can be expressed as

$$\boldsymbol{x}_{\boldsymbol{F}} = \boldsymbol{x}_1 \circ \boldsymbol{x}_2 . \tag{8}$$

2.2.2 Semi-Global Attention Mechanism

The attention mechanism mainly includes two variants, global attention and local attention [33], [34]. Global attention mechanism assigns weights to each output position by focusing on all positions in the input sequence. The approach can capture all information from input, but computational complexity is prohibitively high. In contrast, local attention mechanism such as channel attention only focuses on a local window or sparse position in the input sequence, which reduces computational complexity but may lose global information.

In order to balance processing efficiency and feature expression ability, this paper adopts semi-global attention mechanism. The mechanism is a hybrid attention mechanism of global attention and local attention. The core idea of this mechanism is to divide the input features into blocks according to rules and perform attention calculations within each block, so that long-range feature associations are captured and computational efficiency can be maintained. The proposed algorithm uses semi-global geometric similarity and learnable similarity functions as attention weights to adaptively adjust the attention score and effectively preserve long-distance feature associations.

The structure of the semi-global attention mechanism is shown in Fig. 2. The output $\boldsymbol{x}_{\boldsymbol{F}} \in R^{OC \times L_C}$ of the dual-path convolutional layer undergoes three convolution operation to generate query features $\boldsymbol{Q}$, dynamic bias features $\boldsymbol{K}$ and value features $\boldsymbol{V}$ respectively, which can be expressed as

$$\begin{cases} \boldsymbol{q}_t = \mathrm{Conv}_q(\boldsymbol{x}_{\boldsymbol{F}}{}^t) \\ \boldsymbol{k}_t = \mathrm{Conv}_k(\boldsymbol{x}_{\boldsymbol{F}}{}^t) \\ \boldsymbol{v}_t = \mathrm{Conv}_v(\boldsymbol{x}_{\boldsymbol{F}}{}^t) \end{cases} \tag{9}$$

where $\boldsymbol{x}_{\boldsymbol{F}}{}^t$ represents the *t*-th column vector of $\boldsymbol{x}_{\boldsymbol{F}}$, and Conv(·) denotes the convolution operation. The matrices $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in R^{OC \times L_C}$, $\boldsymbol{q}_t, \boldsymbol{k}_t, \boldsymbol{v}_t \in R^{OC}$ is the channel-wise vector of $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}$ in L_C dimension, $t \in L_C$.

First, orthogonal vector group is defined as $\boldsymbol{M}_a \in R^{b \times OC}$, where *b* represents the dimension parameter. After projecting onto $\boldsymbol{M}_a$ to obtain $\boldsymbol{q}'_t = \boldsymbol{M}_a \boldsymbol{q}_t$, the index corresponding to the maximum value in $\boldsymbol{q}'_t$ can be expressed as $\mathrm{Index}(\boldsymbol{q}'_t) = \mathrm{argmax}(\boldsymbol{q}'_t)$. Along the L_C dimension, the *s*-th and *r*-th eigenvectors of $\boldsymbol{Q}$ are recorded as $\boldsymbol{q}_s$ and $\boldsymbol{q}_r$. Then, the set of eigenvectors $\boldsymbol{q}_s$ with the same maximum projection position as $\boldsymbol{q}_t$ can be expressed as $\boldsymbol{Q}_t$

$$\boldsymbol{Q}_t = \left\{\boldsymbol{q}_s \mid \mathrm{Index}(\boldsymbol{q}_t^{'}) = \mathrm{Index}(\boldsymbol{q}_s^{'})\right\} . \tag{10}$$

Second, the index set δ can be generated based on $\boldsymbol{q}_s$ that meets the above conditions of having the maximum projection position. Then $\boldsymbol{K}_t$, $\boldsymbol{V}_t$ can be expressed as

$$\begin{aligned} \boldsymbol{K}_t &= \boldsymbol{K}_{:,\boldsymbol{\delta}}, \\ \boldsymbol{V}_t &= \boldsymbol{V}_{:,\boldsymbol{\delta}} \end{aligned} \tag{11}$$

where $\boldsymbol{K}_t$, $\boldsymbol{V}_t \in R^{OC \times L_C{}^t}$, and $L_C{}^t$ is the length of index set δ. The vector along dimension $L_C{}^t$ in $\boldsymbol{K}_t$, $\boldsymbol{V}_t$ are denoted by $\boldsymbol{k}_s \in \boldsymbol{K}_t$, $\boldsymbol{v}_s \in \boldsymbol{V}_t$, respectively.

Only the local feature similarity in $\boldsymbol{Q}_t$, $\boldsymbol{K}_t$, $\boldsymbol{V}_t$ need to be calculated after division. As illustrated in Fig. 2, $s_f(\boldsymbol{q}_t,\boldsymbol{q}_s)$ and $s_l(\boldsymbol{k}_s)$ represents the geometric similarity score and the learnable similarity score, respectively. They can be calculated as follows:

$$s_f(\boldsymbol{q}_t,\boldsymbol{q}_s) = \boldsymbol{q}_t{}^{\mathrm{T}}\boldsymbol{q}_s, \tag{12}$$

$$s_l(\boldsymbol{k}_s) = \boldsymbol{W}_2\sigma_a(\boldsymbol{W}_1\boldsymbol{k}_s + \boldsymbol{b}_1) + \boldsymbol{b}_2 \tag{13}$$

where $(\cdot)^{\mathrm{T}}$ and $\sigma_a(\cdot)$ represent the transpose operation and the one-dimensional adaptive activation function, respectively. The matrices $\boldsymbol{W}_1 \in R^{L_C{}^t \times OC}$, $\boldsymbol{W}_2 \in R^{L_C{}^t \times L_C{}^t}$ are learnable weight coefficients, while $\boldsymbol{b}_1 \in R^{L_C{}^t}$, $\boldsymbol{b}_2 \in R^{L_C{}^t}$ are learnable offsets.

Then, by fusing and normalizing $s_f(\boldsymbol{q}_t,\boldsymbol{q}_s)$ and $s_l(\boldsymbol{k}_s)$, the adaptively adjustable attention weight can be obtained:

$$S_t = \mathrm{softmax}(s_f(\boldsymbol{q}_t,\boldsymbol{q}_s) + s_l(\boldsymbol{k}_s)). \tag{14}$$

Finally, the weighted sum of S_t and V_t are used to obtain the enhanced feature representation, the feature vector after the semi-global attention module can be expressed as

$$\boldsymbol{x}_{\boldsymbol{Fa}}{}^t = \sum\nolimits_{\boldsymbol{q}_s \in \boldsymbol{Q}_t} \frac{\exp\left(s_f(\boldsymbol{q}_t,\boldsymbol{q}_s) + s_l(\boldsymbol{k}_s)\right)}{\sum_{\boldsymbol{q}_r \in \boldsymbol{Q}_t}\exp\left(s_f(\boldsymbol{q}_t,\boldsymbol{q}_r) + s_l(\boldsymbol{k}_s)\right)}\boldsymbol{v}_s. \tag{15}$$

2.2.3 One-Dimensional Adaptive Activation Function

In the training phase of neural networks, activation functions play a critical role. Static activation functions (such as Sigmoid, Tanh, ReLU) maintain fixed parameters and shapes during training. The performance of even the generally used ReLU is highly dependent on the specific task and is not a universal solution for all datasets or network architectures.

To enhance the flexibility of activation functions and improve ability of generalization, this paper incorporates a one-dimensional adaptive activation function, AReLU [35], into (13). AReLU dynamically adapts to different tasks based on the complexity of input data, thereby helping to avoid local minima, accelerate convergence. Then the accuracy, efficiency, and generalization can be improved. AReLU is a parameterized generalization of the standard ReLU function. By introducing four learnable parameters, it can adaptively adjust the activation slope and threshold. The AReLU function $\sigma_a(a,b,c,d,z)$ is defined as follows

$$\sigma_a(a,b,c,d,z) = b\mathrm{ReLU}(az + c) + d \tag{16}$$

where a,b,c,d represent four learnable parameters and z is the input vector $\boldsymbol{W}_1\boldsymbol{k}_s + \boldsymbol{b}_1$. Specifically, when $a = 1$, $b = 1$, $c = 0$, and $d = 0$, AReLU degrades to the standard ReLU function.

To better understand the internal mechanism, equation (16) can be divided into the internal linear equation $f_{\mathrm{in}}(a,c,z)$ and the external linear equation $f_{\mathrm{ex}}(b,d)$, which can be expressed as

$$\begin{aligned} f_{\mathrm{in}}(a,c,z) &= az + c, \\ f_{\mathrm{ex}}(b,d) &= b\mathrm{ReLU}(f_{\mathrm{in}}) + d \end{aligned} \tag{17}$$

where $f_{\mathrm{in}}(a,c,z)$ is the input to the ReLU function. Then equation (16) can be rewritten as

$$\begin{aligned} \sigma_a(a,b,c,d,z) &= f_{\mathrm{ex}}(b,d) \\ &= b\mathrm{ReLU}(f_{\mathrm{in}}(a,c,z)) + d. \end{aligned} \tag{18}$$

According to above analysis, the adaptive activation function AReLU only needs four parameters. The parameter quantity and calculation amount can be negligible compared with the whole network.

2.3 Inter-Class Multi-Constraint Based Loss Function

Currently, there are few open-set identification algorithms for radar emitter individual. Existing algorithms have limited loss functions and unclear rejection threshold methods, which leads to the entanglement of known and unknown class samples in feature space, making unknown class samples difficult to reject. To address the above problems, inter-class multi-constraint based loss function and adaptive rejection threshold are presented in this section. First, the total loss function can be calculated according to the extracted feature vector, the cluster center vector and data label information. Then, parameters are adjusted until the model is fitted. The threshold range can be determined based on the classification scores of known class samples in the training set. Finally, known-class identification and unknown-class rejection are performed simultaneously on test samples.

2.3.1 Loss Function Construction for Open-Set Scenario

A radar emitter dataset of P samples with K known classes is given as $\mathcal{D}_{\mathcal{L}} = \{(x_1,y_1), (x_2,y_2), \ldots, (x_p,y_p),\}$, where $y_p \in \{1,\ldots,K\}$ is the label of the known class.

In [36], each known class is distinguished from the corresponding reciprocal point by combining Euclidean distance and vector dot product. The algorithm can discriminate the sample class and exhibit certain open-set identification capability. However, the algorithm primarily discriminates among samples from different classes, which leads to entanglement between samples of different classes and the inability to reserve space for unknown samples. In addition, a large number of unknown samples usually exists in open scenario, which makes it difficult to effectively limit the risk of open space by constraining the open space. Therefore, to effectively measure the discrepancy between extracted features and the cluster center vectors of corresponding class, this paper proposes a distance metric inte-

grating distance similarity reflected by Euclidean distance and geometric similarity captured by vector dot product.

The metric $d(x_{Fa}, \mathcal{P}^k)$ between sample x and cluster center vector is defined as

$$\begin{aligned} d(\boldsymbol{x}_{Fa}, \mathcal{P}^k) &= d_{\mathrm{e}}(\boldsymbol{x}_{Fa}, \mathcal{P}^k) - d_{\mathrm{d}}(\boldsymbol{x}_{Fa}, \mathcal{P}^k), \\ d_{\mathrm{e}}(\boldsymbol{x}_{Fa}, \mathcal{P}^k) &= \frac{1}{OC} \cdot \left\| \boldsymbol{x}_{Fa} - \mathcal{P}^k \right\|_2^2, \\ d_{\mathrm{d}}(\boldsymbol{x}_{Fa}, \mathcal{P}^k) &= \boldsymbol{x}_{Fa} \cdot \mathcal{P}^k \end{aligned} \tag{19}$$

where $d_{\mathrm{e}}(\cdot)$ and $d_{\mathrm{d}}(\cdot)$ represent normalized Euclidean distance calculation and vector dot product calculation, respectively. The vector $\mathcal{P}^k \in R^{OC}$ represents the cluster center vector of class k, $k \in K$, while $\boldsymbol{x}_{Fa} \in R^{OC}$ is the extracted feature vectors of sample x. In order to ensure the identification performance of known class samples, the metric $d(x_{Fa}, \mathcal{P}^k)$ is used as the classification score. The scores of P samples can be recorded as $lots \in R^{P \times K}$, where each row represents the score of a sample for K cluster center vectors.

The cross entropy loss $\mathcal{L}_1$ is defined as

$$\begin{aligned} \mathcal{L}_1 &= -\log p(y = k \mid x, \boldsymbol{x}_{Fa}, \mathcal{P}), \\ p(y = k \mid x, \boldsymbol{x}_{Fa}, \mathcal{P}) &= \frac{\mathrm{e}^{\gamma d(\boldsymbol{x}_{Fa}, \mathcal{P}^k)}}{\sum_{i=1}^{N} \mathrm{e}^{\gamma d(\boldsymbol{x}_{Fa}, \mathcal{P}^i)}} \end{aligned} \tag{20}$$

where γ is a hyperparameter that controls the distance-probability conversion, and $p(y = k \mid x, \boldsymbol{x}_{Fa}, \mathcal{P})$ represents the normalized probability function of the correct identification of the k-th class sample.

The score between each sample and the corresponding cluster center vector is maximized by minimizing $\mathcal{L}_1$. The cluster center corresponding to the maximum score represents the sample class, indicating that the cluster center of corresponding class can reversely characterize sample x. Therefore, the score distribution between extracted feature $\boldsymbol{x}_{Fa}$ and each known class cluster center vector determines the sample class, enabling correct identification of known class samples.

Furthermore, to reduce open space risk, this paper further proposes three loss functions, inter-class distance loss $\mathcal{L}_2$, intra-class dot product loss $\mathcal{L}_3$, and intra-class distance loss $\mathcal{L}_4$, which are defined as

$$\begin{aligned} \mathcal{L}_2 &= \sum_{i=1}^{K} \sum_{\substack{j=1 \\ i \neq j}}^{K} \left(\frac{1}{OC} \left\| \mathcal{P}^i - \mathcal{P}^j \right\|_2^2 - D_1 \right)^2 / \left(K^2 - K \right), \\ \mathcal{L}_3 &= (\boldsymbol{x}_{Fa} \mathcal{P}^k - D_2)^2, \\ \mathcal{L}_4 &= \max\left(\frac{1}{OC} \left\| \boldsymbol{x}_{Fa} - \mathcal{P}^k \right\|_2^2 - R, 0 \right) \end{aligned} \tag{21}$$

where D_1, D_2 and R represent a fixed value, a learnable parameter and a learnable boundary constraint, respectively.

The above three equations further constrain the distribution of known class samples in the feature space from three aspects, which reserves enough space for the distribution of unknown class samples by improving the feature discriminability between known classes. The main function of $\mathcal{L}_2$ is to constrain the distance between the center of each cluster. The main function of $\mathcal{L}_3$ and $\mathcal{L}_4$ serves to compact the distribution of intra-class samples in the feature space and reserves adequate space for unknown class samples.

Then the total loss function $\mathcal{L}_0$ can be expressed as

$$\mathcal{L}_0 = \mathcal{L}_1 + \partial_1 \mathcal{L}_2 + \partial_2 \mathcal{L}_3 + \partial_3 \mathcal{L}_4 \tag{22}$$

where ∂_1, ∂_2, ∂_3 are loss weight coefficients. Different constraint effects can be achieved by adjusting corresponding parameters.

2.3.2 Class-based Adaptive Threshold Selection Strategy

In the open-set identification task, another core challenge is to set suitable thresholds for distinguishing between known and unknown classes. The selection of thresholds needs to ensure both accurate identification of known class samples and the ability to reject unknown class samples. In addition, the occurrence probability of unknown class samples in real scenario is uncertain, which makes the open-set identification algorithm based on fixed threshold comparison usually not applicable. Therefore, this paper proposes an adaptive threshold selection strategy based on known classes, which ensures the accuracy of known class identification and quantitatively evaluates the identification ability for unknown samples.

According to (19) and (20), this paper uses $d(x_{Fa}, \mathcal{P}^k)$ as the score to classify known class samples. In the training phase, the distribution of each cluster center and known class sample in feature space is compact and separable after model fitting. Therefore, the distance score can be used as the threshold to reject unknown class samples. The criterion for sample rejection is the distance between the sample and the cluster center. Then the identification rate of the known class and the rejection rate of the unknown class are inversely proportional. The higher the identification rate of the known class, the lower the rejection rate of the unknown class will be. Balanced threshold should be selected to improve the open-set rejection rate without compromising the closed-set identification rate. The selection process of thresholds is shown in Algorithm 1, where the 5th and 95th percentile values were selected to ensure a closed-set recognition rate of over 90 %.

In the test phase, the scores between each input sample and cluster center can be calculated by (19). The index corresponding to the maximum score represents the predicted class, denoted as k. By comparing the maximum score against the threshold range (T^k_{start}, T^k_{end}) corresponding to predicted class k, the identification result can be obtained. If the maximum score falls within threshold range, the sample will be identified as a known class and

Algorithm 1 Class-based adaptive threshold selection strategy

Input: Training set $\mathcal{D}_L$, model parameters.

Output: Class-based adaptive threshold range (T^k_{start}, T^k_{end}).

1: Apply (19) to obtain the model output scores *lots*.

2: Select the maximum value x_l and the corresponding index denotes as *preds* along the second dimension of $lots \neq x_l$, $preds \in R^{P\times 1}$, x_l is the maximum score combination, *preds* is predicted class.

3: scor $\leftarrow \{\ \}$ # scor is to store the scores corresponding to the samples predicted as class k

4: for $k = 1,\ldots,K$

5: $u \leftarrow 1$

6: for $j = 1,\cdots,P$

7 if $preds(j) == k$ do

8 $\text{scor}(k,u) \leftarrow x_l(k)$

9 $u \leftarrow u+1$

10: end

11: end

12: Sort $\text{scor}_{k,:}$ in ascending order along the second dimension

13: $T^k_{\text{start}} \leftarrow$ the 5-th percentile value of scor(k,:)

14: $T^k_{\text{endt}} \leftarrow$ the 95-th percentile value of scor(k,:)

15: end

output class k. Otherwise, the sample is rejected as an unknown class.

2.4 Open-Set Identification Training and Testing Process

The training and test process of the proposed algorithm are shown in Algorithm 2 and Algorithm 3. In the training phase, the original radar emitter individual data first preprocessed and a deep neural network is constructed with parameters randomly initialized. Then the network parameters are updated through back propagation. Finally, the network parameters and the selected threshold range are saved.

Algorithm 2 Open-set identification training process

Input: Radar emitter samples $\boldsymbol{x}_{iq}$. Randomly initialize parameter sets Θ and learning rate l_r.

Output: Model parameters and rejection threshold

1: Apply (1)–(6) to align each sample $\boldsymbol{x}_{iq}$, obtain dataset $\mathcal{D}_L$.

2: For each epoch:

3: Shuffle the dataset and divide it into E batches

4: Begin the batch loop:

5: Calculate features extracted by semi-global attention mechanism according to (7)–(14).

6: Calculate loss function according to (22).

7: Obtain threshold range (T^k_{start}, T^k_{end}) according to algorithm 1.

8: Calculate the stochastic gradient of each parameter in the parameter set to update parameters according to $\Theta \leftarrow \Theta - l_r \cdot \nabla_\Theta \mathcal{L}_0$, where $\nabla_\Theta \mathcal{L}_0$ is the stochastic gradient of the loss function.

9: End the batch loop.

10: End epoch loop

Algorithm 3 Open-set identification test process

Input: Test sample $\boldsymbol{x}_{\text{test}}$, pre-trained model and threshold range.

Output: Identification result.

1: Apply (1)–(6) to align each sample $\boldsymbol{x}_{\text{test}}$.

2: The processed samples are input into the model and the classification score distribution is recorded as *lots*.

3: Obtain the maximum value x_l and the index k according to *lots*, select the threshold range (T^k_{start}, T^k_{end}) corresponding to k.

4: if $T^k_{\text{start}} < x_l < T^k_{\text{end}}$ do

5: identify as a known class and output the classification result k.

6: else

7: reject as unknown

8: end

In the test phase, the same preprocessing operation is performed on each test sample, then the pre-trained model is loaded for feature extraction. If only individual identification is performed in a closed-set scenario, the features extracted by the semi-global attention mechanism are directly passed through GAP, FC, and SoftMax to obtain the identification result. Otherwise, the distance measurement between the test sample and the center of each cluster is calculated, accurate identification of known classes and rejection of unknown classes can be achieved based on the preset threshold range.

3. Experimental Results

3.1 Dataset and Experimental Settings

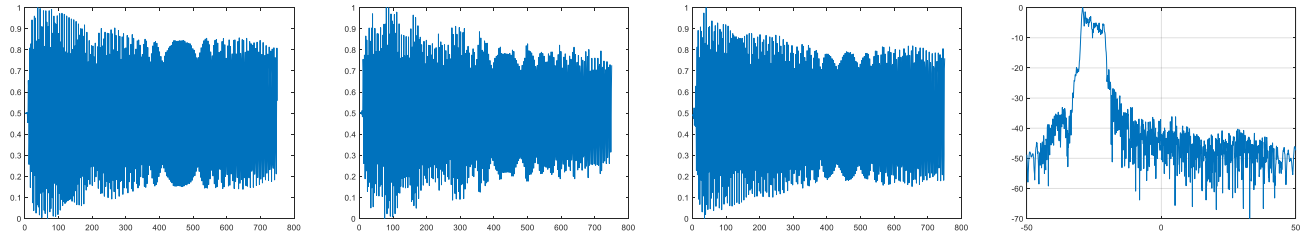
In this section, individual identification experiments are carried out based on radar emitter data, aiming to study the robustness and generalization ability of proposed algorithm. The parameters of the dataset are shown in Tab. 1. The dataset contains three individual radars, recorded as individual 1, individual 2 and individual 3. The pulse repetition interval (PRI) for all is 20 μs. The sampling rate is uniformly 100 MHz. The pulse width of individual 1 and individual 2 is 9.8 μs and the pulse width of individual 3 is 8.7 μs. Each radar is configured with four frequency points of 15800, 15920, 16220 and 16340 MHz. The sample size for each individual radar at each frequency point is 30,000. For convenience of subsequent description, these four frequency points are denoted as frequency point A, frequency point B, frequency point C and frequency point D, respectively.

In order to analyze the differences among different individuals in the time-domain and frequency-domain, the

time-frequency diagram and spectrum diagram of random samples of three individuals at frequency point A and frequency point B are shown in Fig. 3 and Fig. 4. The spectra of different individuals at the same frequency point are consistent, so that only the spectrum diagram of individual 1 is shown in Fig. 3. It can be seen that although time-domain waveforms at the same frequency point show differences across different individuals, these differences are insufficient to serve as discriminative features. Obviously, the same individual at different frequencies shows different time-domain waveforms, due to the different propagation characteristics of electromagnetic waves at different frequencies. It is necessary to divide the dataset into independent data groups according to different frequency points for separate individual identification experiments. Moreover, the spectrums from frequency point A and frequency point B exhibit differences under the same individual due to channelization process in the data processing stage. In this condition, the difference can be used to verify the generalization ability of the proposed algorithm.

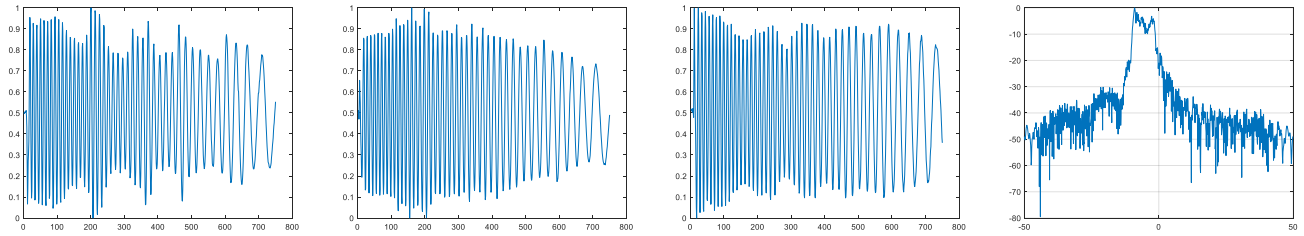
Individual\Parameter	Frequency point type (MHz)	PW (μs)	PRI (μs)	Sampling rate (MHz)	Sample size
Individual 1	15800/15920/16220/16340	9.8	20	100	30000
Individual 2	15800/15920/16220/16340	9.8	20	100	30000
Individual 3	15800/15920/16220/16340	8.7	20	100	30000

Tab. 1. Dataset parameters.



（a）Time-domain diagram of individual 1　（b）Time-domain diagram of individual 2　（c）Time-domain diagram of individual 3　（d）Spectrum of individual 1

Fig. 3. Time-domain and spectrum diagram of three individuals at frequency point A.



（a）Time-domain diagram of individual 1　（b）Time-domain diagram of individual 2　（c）Time-domain diagram of individual 3　（d）Spectrum of individual 1

Fig. 4. Time-domain and spectrum diagram of three individuals at frequency point B.

Data preprocessing module	β_1	0.5
	β_2	0.6
	M_g	10
	N_g	1000
	T_g	0.01 μs
	f_g	100 MHz
Semi-global attention mechanism module	b	128
Loss function construction	D_1	15
	$\partial_1, \partial_2, \partial_3$	0.1

Tab. 2. Model parameters setting.

In this experiment, the dataset is divided into four independent groups based on different frequency points. Each group contains only three types of individual samples at one frequency point. The sample size of each individual at each frequency point is 30,000. Training set and test set are randomly divided in a ratio of 8:2. The experimental hardware platform is configured as NVIDIA RTX 4060, which is implemented based on the Pytorch framework. The network uses the Adam optimizer for training with a batch size of 64. The weight decay coefficient is 1×10^{-4}, the initial learning rate is set to 0.01, which is reduced by 0.1 times every 30 epochs during training. Other parameters are shown in Tab. 2.

3.2 Comparison of Different Feature Extraction Algorithms

In order to evaluate the feature extraction capabilities of different algorithms, this section conducts comparative experiments with two algorithms, Intrinsic Mode Function Distinct Native Attribute (IMF-DNA) [37] and bispectral contour integral [38] on three individuals at each frequency point. IMF-DNA is a feature extraction algorithm based on time-frequency domain analysis. This approach extracts fingerprint and fractal features from the intrinsic properties of intrinsic mode functions, establishing a foundation for subsequent multi-domain feature representation of individuals. The feature extraction process is to expand the signal into multiple IMFs in the frequency-domain and time-domain, then the first three components are taken for extraction. Bispectral contour integral is a feature extraction algorithm based on the frequency domain. The algorithm quantifies bispectral mean features along contour integrals using an entropy-like function to characterize the subtle differences between emitters. The feature extraction process adopts Fisher class-separability to select optimal bispectral contour integrals as the extracted feature. Obviously, segmentation process is applied to the preprocessed data in both two traditional feature extraction algorithms. In the experiment, traditional algorithms need manually extraction of time-frequency features from preprocessed I/Q data for model input, while the proposed algorithm directly uses the raw preprocessed I/Q data itself as input. The results are shown in Tab. 3.

It can be seen that the identification results at four frequency points, the average identification rate of proposed algorithm reached 97.55%, which is 8.37% and 10.95% higher than the IMF-DNA algorithm and the bispectral contour integral algorithm, respectively. Obviously, the identification performance of IMF-DNA is better than that of bispectral contour integral. It is because IMF-DNA retains more complete subtle features than bispectral contour integral through multi-scale decomposition. Furthermore, the proposed algorithm is superior to the two traditional algorithms at different frequency points, especially at frequency point B, where the performance advantage is more obvious. IMF-DNA relies on EMD decomposition, which is easy to cause modal aliasing, resulting in a decrease in the quality of extracted features. In addition, the segmentation process may affect the generalization performance. The bispectral contour integral is limited by the choice of integral path, which makes improper paths significantly affect the discriminative ability of the extracted features. In contrast, this paper uses the preprocessed I/Q signal as input, which abandons the limitations of traditional feature extraction and learns highly discriminative feature representation from the raw preprocessed data. The algorithm can effectively avoid the information loss problem caused by segmentation process, thereby showing significant advantages in feature representation ability and identification performance.

Frequency type	IMF-DNA [37]	Bispectral contour integral [38]	The proposed algorithm
Frequency A	88.89	86.32	**97.08**
Frequency B	89.54	86.98	**98.59**
Frequency C	90.56	85.52	**97.33**
Frequency D	89.53	87.56	**97.19**
Average identification rate	89.63	86.60	**97.55**

Tab. 3. Identification results of different feature extraction algorithms (%).

3.3 Comparative Methods for Closed-Set and Open-Set Identification

The comparison method for closed-set experiments are as follows.

a) SVM: Linear support vector machines (SVMs) [39] use linear hyperplanes to segment data and are suitable for scenario where data in feature space is linearly separable. Nonlinear SVMs [40] employ nonlinear kernel functions to project data into a high-dimensional space. Then an optimal classification hyperplane is constructed in the space to achieve effective separation for data of different classes.

b) MSFKSPP [41]: MSFKSPP is a sparse-preserving projection algorithm that extracts local detail information through small-scale features and captures global contour features through large-scale features. Then effective extraction of target features is achieved to identify individuals based on the maximization criterion of inter-class separability in the reproducing kernel Hilbert space.

c) Stacked LSTM [42]: The stacked long short-term memory (LSTM) network is composed of multiple layers of classic LSTM units, each layer containing several memory modules. The architecture can extract hierarchical features in time series data and achieve the modeling of long-term dependencies through a gating mechanism.

d) DE-CNN [43]: The algorithm is a convolutional neural network framework based on demodulation embedding, which can directly identify radar emitter individual from the raw real signal. The algorithm enhances the identification ability of modulation features in complex electromagnetic environments by coupling signal demodulation with deep feature learning.

e) DCA-BIGRU [44]: DCA-BIGRU integrates dual-path convolution, attention mechanism and bidirectional gated recurrent unit. The algorithm captures spatial features through parallel convolution paths and combines attention-weighted bidirectional time series modeling, significantly improving the utilization efficiency of contextual information in time series data.

The comparison method for experiments in the open-set scenario are as follows.

a) SOFTMAX [45]: Hard threshold is introduced to determine prediction confidence of the classifier. If the

sample confidence is less than the threshold, the sample will be judged as an unknown one.

b) Generalized Convolutional Prototype Learning [46]: This algorithm is a convolutional prototype learning (CPL) framework that uses prototype loss as a regularization constraint to improve the intra-class compactness of feature representation, thereby improving model performance. This section adds a softmax module based on the CPL to complete the open-set experiment.

c) Adversarial reciprocal point learning (ARPL) [36]: ARPL proposes a learning framework based on reciprocal points, which adopts adversarial edge constraints and adversarial enhancement methods. Open space risks can be reduced by restricting the distribution of reciprocal points and generating diverse confusion samples.

3.4 Comparison of Existing Individual Identification Algorithms

In order to verify the effectiveness of the individual identification algorithm in this paper, this section designs three experiments. The first experiment is a comparative experiment of existing individual identification algorithms, which is conducted on three individuals at each frequency point. The second experiment is a comparative experiment of various algorithms on different training sample sizes, which is used to verify the impact of different training sample sizes on the identification performance. The third experiment is a generalization experiment at different frequency points, which is used to analyze the generalization ability of each algorithm.

In order to systematically compare the identification performance of the proposed algorithm with existing algorithms, this experiment conducted identification experiments on three individuals at each frequency point, as shown in Tab. 4. Compared with the three traditional classification algorithms, MSFKSPP shows significant performance advantages, which indicates that the sparseness-preserving projection algorithm proposed by MSFKSPP has the multi-scale fusion kernel sparsity of data. However, loss of available features in the classification process leads to an average identification rate of only 89.21% for MSFKSPP. The performance of nonlinear SVM is significantly better than that of linear SVM. Linear SVM is limited by inherent characteristics of linear kernel function, which is only applicable to ideal scenario where the feature space is linearly separable. The data in the experiment presents high-dimensional nonlinear distribution characteristics in feature space, which enables the nonlinear SVM to effectively construct complex decision boundaries. However, the nonlinear SVM is sensitive to the choice of kernel function, especially dealing with time series data. The nonlinear correlation features in data are difficult to be extracted effectively, resulting in average identification rate of only 83.75%.

Compared with three existing deep learning-based algorithms, DCA-BIGRU has the best identification performance. The result shows that DCA-BIGRU has good feature extraction ability for this time series data by integrating dual-path convolution, attention mechanism and bidirectional gated recurrent unit. DE-CNN has poor identification performance, indicating that the demodulation layer of the network embedding is not suitable for this dataset and performs poorly. Although stacked the average identification rate of LSTM is 90.72%, which may be due to the lack of explicit feature selection capabilities and insufficient focus on key time steps. In summary, the proposed algorithm shows the best identification performance at each frequency point, which is 2.52% higher than DCA-BIGRU algorithm. The results show that feature extraction algorithm of the proposed semi-global attention mechanism has significant advantages in feature extraction and identification and can more effectively utilize data features to improve identification performance.

To present the identification of each individual in detail, the confusion matrix of each algorithm is given under the conditions of 20,000 training samples and frequency point B, as shown in Fig. 5. The three traditional algorithms, including linear SVM, nonlinear SVM and MSFKSPP, show large differences in identification rates for each individual, while the deep learning algorithms show relatively balanced identification performance for each individual. The proposed algorithm has the most balanced identification effect on the three individuals and the rate of three individuals is above 98%.

To evaluate the impact of different training sample sizes on the identification performance of various algorithms, the average identification rates of the above algo-

Frequency point type	Linear SVM [39]	Nonlinear SVM [40]	MSFKSPP [41]	Stacked LSTM [42]	DE-CNN [43]	DCA-BIGRU [44]	The proposed algorithm
Frequency A	78.85	84.32	89.68	90.94	87.28	95.12	**97.08**
Frequency B	76.52	82.15	88.31	91.50	85.25	94.35	**98.59**
Frequency C	77.95	83.87	88.34	90.32	87.05	95.05	**97.33**
Frequency D	79.28	84.65	90.52	90.12	88.65	95.58	**97.19**
Average identification rate	78.15	83.75	89.21	90.72	87.06	95.03	**97.55**

Tab. 4. Identification results of different algorithms (%).

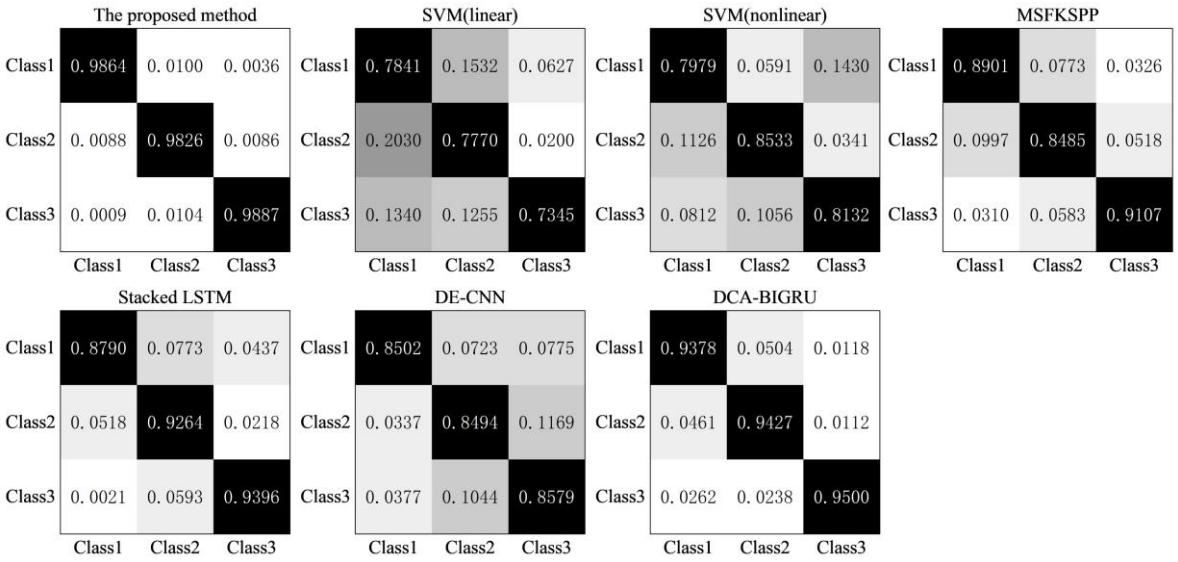


Fig. 5. Confusion matrix of each algorithm.

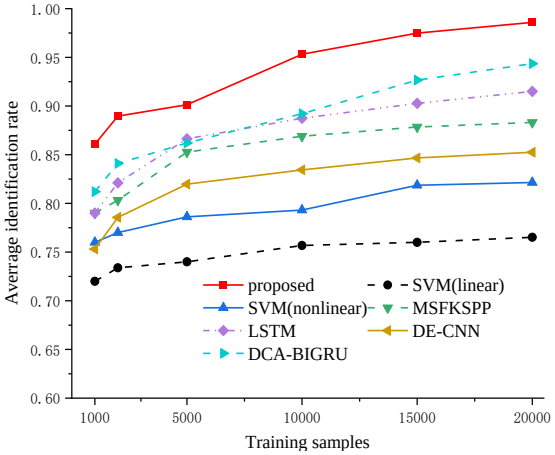


Fig. 6. Average identification rates on different training sample sizes.

Algorithm	Frequency point B → Frequency point A	Frequency point B → Frequency point B	Frequency point B → Frequency point C	Frequency point B → Frequency point D
Nonlinear SVM	52.34	82.15	71.05	55.74
MSFKSPP	64.19	88.31	85.62	62.45
DCA-BIGRU	59.88	94.35	83.57	67.63
The proposed algorithm	**65.12**	**98.59**	**90.36**	**68.23**

Tab. 5. Generalization experiment at different frequency points (%).

rithms are shown in Fig. 6. The training sample set sizes used in the experiments are 1000, 2000, 5000, 10000, 15000, and 20000 at frequency point B. It can be seen that the identification accuracy of these algorithms has been improved with the increase in the number of training samples. Obviously, LSTM has the most obvious performance improvement, indicating that LSTM is the most sensitive to the reduction of the number of training samples. The performance of the proposed algorithm is the most stable, reaching identification rate of 86.08% when the number of training samples is 1000. It is clear that the proposed algorithm achieves the best performance under various sample numbers.

In addition, samples of other frequency types are difficult to process under training with a single frequency point data group. Therefore, the generalization ability of various algorithms for individual samples at different frequencies can be verified by performing cross-frequency generalization tests on individual samples at other frequency points. In order to evaluate the generalization performance of each algorithm under frequency changes, this section uses four algorithms, including nonlinear SVM, MSFKSPP, DCA-BIGRU, and the proposed algorithm for comparative experiments, as shown in Tab. 5. In the experiment, identification model is constructed based on the training set of three individuals at frequency point B. The generalization rate in Tab. 5 represents the average probability that individual samples at frequency points different from those used in training are correctly classified into the corresponding categories.

The results show that the model trained on the samples of frequency point B has good generalization ability for samples of frequency point C. The proposed algorithm has the best generalization performance of 90.36%. The generalization performance of nonlinear SVM, MSFKSPP, and DCA-BIGRU are 71.05%, 85.62%, and 83.57%, respectively, which possesses certain generalization ability. However, the generalization rates of all algorithms are relatively low for individuals of frequency points A and D. It is because that the two specific frequency points of B and C show consistent frequency deviations, while the frequency deviations of other frequency points are significantly different. Therefore, in terms of the generalization performance for cross-frequency point samples, if the frequency deviations are the same, the corresponding samples

can be better identified. If the frequency deviations are different, although the generalization of the proposed algorithm is poor, it still shows the best generalization performance. The generalization of different frequency deviations cross-frequency points will be further studied in the future.

3.5 Comparison of Open-Set Identification Algorithms

In order to evaluate effectiveness of the proposed open-set identification algorithm, this section conducts open-set identification comparison experiments with existing algorithms. In the experiment, the closed-set samples consist of individual 1 and individual 2, the samples of individual 3 are used as open-set samples.

Based on the frequency point type, the training set and test set are divided into four groups, as shown in Tab. 6. For example, training set I contains samples of individual 1 and individual 2 at frequency A, with a sample size of 10,000 each.

In the experiment, the trained models of each algorithm are evaluated by using the test set. The closed-set identification rate and open-set rejection rate are shown in Tab. 7. The average open-set rejection rate of SOTFMAX algorithm based on fixed confidence is only 66.78% under five training sets. The limitations are as follows. First, SOTFMAX algorithm does not incorporate the concept of unknown and only designed to classify known class samples. As a result, the distributions of known classes in the feature space will be overlapped and space for unknown-class samples is not reserved appropriately. Second, the linear decision boundary based on fixed confidence lacks flexibility and cannot effectively adapt to the nonlinear distribution characteristics of feature space, which causes the confidence scores of some unknown class samples to exceed the preset threshold. Finally, although lower confidence threshold may improve the open-set rejection rate, it will lead to decrease in the closed-set identification accuracy. Therefore, threshold setting needs to balance the open-set rejection and closed-set rate, which is a complex process to adjust parameters and lacks universality.

Comparing the two algorithms of CPL and ARPL, the average open-set rejection rate of CPL is 76.64% is lower than that of ARPL. It is because that the features of each class are compressed into a single prototype point and the distribution diversity of intra-class samples is ignored in

	Training set			Test set		
Number	Frequency type	Individual class	Sample size	Frequency type	Individual class	Sample size
I	Frequency point A	Individual 1, individual 2	10,000 each	Frequency point A	Individual 1, individual 2, individual 3	The sample size of individual 1 and individual 2 is 2000, and the sample size of individual 3 is 20000
II	Frequency point B			Frequency point B		
III	Frequency point C			Frequency point C		
IV	Frequency point D			Frequency point D		

Tab. 6. Training set and test set for open-set experiments.

Dataset	SOFTMAX [45]		GCPL [46]		ARPL [36]		The proposed algorithm	
	known	unknown	known	unknown	known	unknown	known	unknown
I	90.01	70.32	90.63	78.53	90.26	87.68	89.80	**95.38**
II	90.17	62.78	90.07	79.50	90.56	82.64	89.81	**84.89**
III	89.98	57.23	90.10	69.56	89.89	72.85	93.80	**76.94**
IV	90.06	76.80	89.89	78.97	90.69	83.23	93.70	**93.36**
Average rejection and identification rate	90.06	66.78	90.17	76.64	90.35	81.60	**91.78**	**87.64**

Tab. 7. Closed-set identification rate and open-set rejection rate of each algorithm (%).

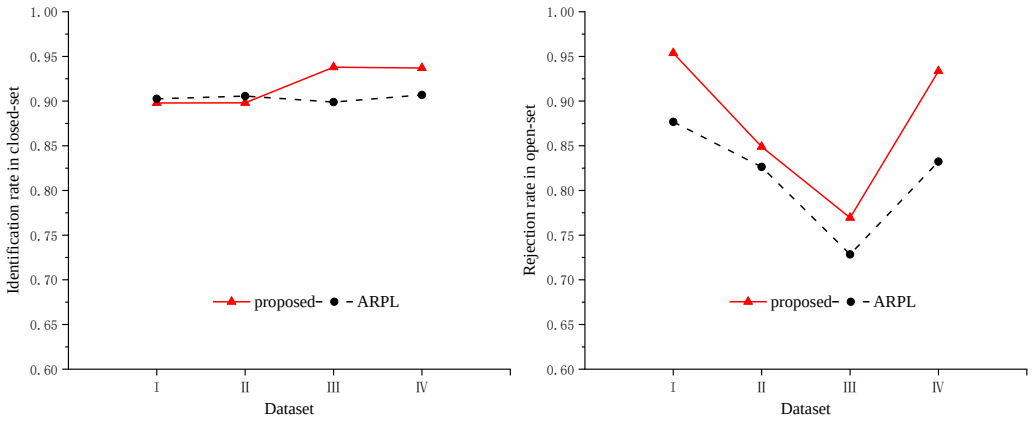


Fig. 7. Closed-set identification rate and open-set rejection rate of the two algorithms.

CPL. This leads to excessive contraction of the intra-class features, thereby unknown class samples are easy to fall into the vicinity of known class prototypes. By contrast, the advantage of ARPL lies in effectively reserving discriminative space for unknown samples by pushing the feature distributions of different class samples and adopting a generative adversarial strategy to reduce the risk of open space. However, ARPL still has some limitations. The generated confusing samples are difficult to construct reasonably, the rejection performance may be reduced due to unreasonable generated samples. Moreover, the parameters tuning process are relatively complicated. Compared with above algorithms, the average rejection rate of the proposed algorithm is 87.64%, representing the best performance. The proposed algorithm constrains the feature distribution of known class samples through inter-class multi-constraint losses and uses learnable parameters to leave a certain margin, which not only optimizes the identification performance of known class samples but also reserves appropriate space for unknown class samples. The proposed class-based adaptive threshold selection strategy abandons the fixed threshold strategy. The strategy not only generates corresponding thresholds based on the distribution of known class samples, but also sets differentiated threshold intervals for each class by quantitative analysis, leaving a large amount of threshold space for open-set samples.

To illustrate the identification of known and unknown individuals in detail, ARPL and the proposed algorithm are compared in Fig. 7. The results show that the proposed algorithm consistently achieves a higher open-set rejection rate than ARPL across all training sets when the closed-set accuracy reaches around 90%, which reflects superior open-set identification performance.

3.6 Ablation Study

This section conducts ablation experiments by removing or replacing different modules to evaluate the contribution of each module in the proposed closed-set and open-set algorithms.

Based on the closed-set algorithm proposed in this paper, this experiment removes the sample alignment, semi-global attention mechanism and dual-path convolution modules to conduct ablation experiments on three individual samples at frequency point B. In this ablation study, the selection combination of different modules and identification results are shown in Tab. 8 and Fig. 8.

It can be seen that after removing the sample alignment module, the identification rate of the model is reduced to 86.57%, indicating that more complete data can be obtained after the sample alignment module, which greatly improves the identification rate and generalization ability of the model. After changing the dual-path convolution layer to single-path convolution layer, the average identification rate of the model is reduced by 2.35% to

Number	Sample alignment module	Dual-path convolution module	Attention mechanism module	
			Semi-global attention	Spatial attention
a	√	√	√	
b		√	√	
c	√		√	
d	√	√		√

Tab. 8. Selection and combination of different modules under closed-set algorithm.

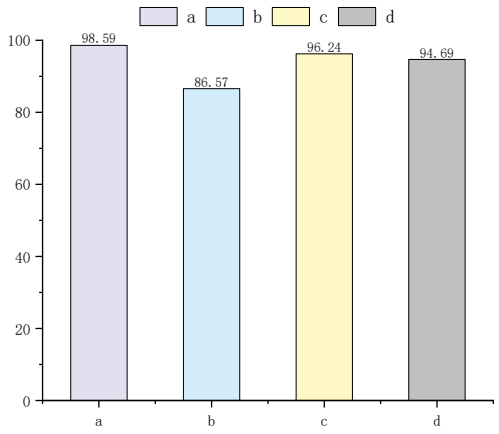


Fig. 8. Average identification rate of different module combinations.

96.24%. Compared with the single-path convolution, the dual-path convolution can improve the feature representation ability of the model by extracting features from two independent paths and combining high-level semantic features with low-level detail features. When the semi-global attention mechanism module becomes the spatial attention mechanism [47], the average identification rate of the model is reduced by 3.90% to 94.69%. The result indicates that the semi-global attention mechanism not only considers global information, but also combines local features. Owing to the balance between global and local information, the model can better capture the key features of the target class to improve the identification performance.

For the loss function based on inter-class multi-constraints presented in this paper, this experiment adopts different loss function combinations to conduct ablation experiments on training and test set I in Tab. 6. In this ablation study, the selection combination of different losses and identification results is shown in Tab. 9 and Fig. 9.

It can be seen that when the inter-class distance loss $\mathcal{L}_2$ is removed, the rejection rate of the model is reduced to 90.26%. When only the inter-class distance loss $\mathcal{L}_3$ is removed, the rejection rate of the model is reduced to 91.87%. After removing the inter-class distance loss $\mathcal{L}_4$, the rejection rate of the model is reduced to 91.83%. Furthermore, when all three losses are removed, the rejection

Number	Loss function			
	Cross entropy loss $\mathcal{L}_1$	Inter-class distance loss $\mathcal{L}_2$	Intra-class dot product loss $\mathcal{L}_3$	Intra-class distance loss $\mathcal{L}_4$
a	√	√	√	√
b	√		√	√
c	√	√		√
d	√	√	√	
e	√			

Tab. 9. Selection and combination of different modules under open-set algorithm.

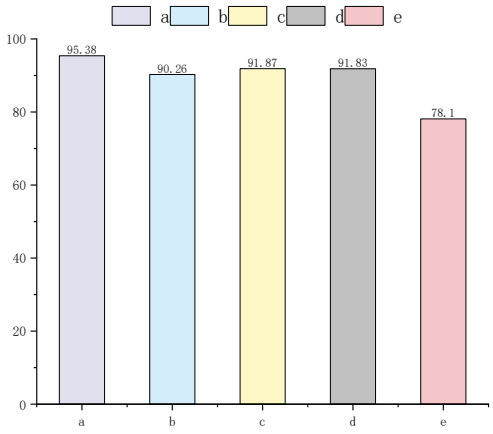


Fig. 9. Average rejection rate of different module combinations.

rate of the model is reduced to 78.10%. The results show that the removal of $\mathcal{L}_2$, $\mathcal{L}_3$ and $\mathcal{L}_4$ will lead to the decrease of open-set rejection rate. Among these, inter-class distance loss $\mathcal{L}_2$ demonstrates the most substantial impact on open-set identification performance.

4. Conclusion

This paper proposes an open-set identification algorithm for radar emitters based on semi-global attention and inter-class multi-constraint. First, by introducing the semi-global attention mechanism, the attention score is adaptively adjusted and long-distance feature associations are retained. More accurate feature details and higher computational efficiency can be obtained. Second, inter-class multi-constraint based loss function is presented. Multiple losses are used to constrain the distribution of known class samples in the feature space, and the threshold is set based on the mean score of known class samples. Finally, according to the proposed class-based adaptive threshold selection strategy, known-class identification and unknown-class rejection can be performed simultaneously. Experimental results show that the proposed algorithm exhibits strong generalization performance for individual samples in closed-set scenario, especially with same frequency offset. Moreover, in open-set scenario the proposed algorithm can effectively identify known class individuals and reject unknown class individuals, which shows good performance and high engineering practical value. Future work will focus on exploring open-set individual identification methods with higher interpretability and generalization to different frequency offsets.

References

[1] SHI, Y., ZHANG, W. B., ZHU, Z. M., et al. Specific radar emitter identification: A comprehensive review (in Chinese). *Journal of Electronics & Information Technology*, 2022, vol. 44, no. 6, p. 2216–2229. DOI: 10.11999/JEIT210161

[2] TAN, K. W., YAN, W. J., ZHANG, L. M., et al. Specific emitter identification based on software-defined radio and decision fusion. *IEEE Access*, 2021, vol. 9, p. 86217–86229. DOI: 10.1109/ACCESS.2021.3088542

[3] XIE, N., LI, Z., TAN, H. A survey of physical-layer authentication in wireless communications. *IEEE Communications Surveys & Tutorials*, 2021, vol. 23, no. 1, p. 282–310. DOI: 10.1109/COMST.2020.3042188

[4] GUO, L. T., LIU, C., LIU, Y. C., et al. Toward open-set specific emitter identification using auxiliary classifier generative adversarial network and OpenMax. *IEEE Transactions on Cognitive Communications and Networking*, 2024, vol. 10, no. 6, p. 2019–2028. DOI: 10.1109/TCCN.2024.3408417

[5] HUANG, S. N., GUO, L. T., FU, X., et al. Open-set specific emitter identification leveraging enhanced metric denoising autoencoders. *IEEE Internet of Things Journal*, 2025, vol. 12, no. 4, p. 3453–3462. DOI: 10.1109/JIOT.2024.3404042

[6] WANG, J. Z., WANG, H. H., WANG, W., et al. OS-SEI: Open-set specific emitter identification based on outlier exposure and label smoothing. *IEEE Internet of Things Journal*, 2025, vol. 12, no. 21, p. 44128–44140. DOI: 10.1109/JIOT.2025.3560233

[7] KAWALEC, A., OWCZAREK, R. Specific emitter identification using intrapulse data. In *First European Radar Conference*. Amsterdam (Netherlands), 2004, p. 249–252. ISBN: 1-58053-993-9

[8] RU, X. H., YE, H. H., LIU, Z., et al. An experimental study on secondary radar transponder UMOP characteristics. In *The 13th European Radar Conference*. London (UK), 2016, p. 314–317. ISBN: 978-1-5090-1513-9

[9] LIU, J., LEE, J. P. Y., LI, L., et al. Online clustering algorithms for radar emitter classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, vol. 27, no. 8, p. 1185–1196. DOI: 10.1109/TPAMI.2005.166

[10] RU, X. H., LIU, Z., HUANG, Z. T., et al. Evaluation of unintentional modulation for pulse compression signals based on spectrum asymmetry. *IET Radar, Sonar & Navigation*, 2017, vol. 11, no. 4, p. 656–663. DOI: 10.1049/iet-rsn.2016.0248

[11] RU, X. H., HUANG, Z. T., LIU, Z., et al. Frequency domain distribution and band-width of unintentional modulation on pulse. *Electronics Letters*, 2016, vol. 52, no. 22, p. 1853–1855. DOI: 10.1049/el.2016.0733

[12] ZOU, Y. P., WANG, X., CHEN, Y., et al. Specific emitter identification via bispectrum-radon transform and hybrid deep model. *Mathematical Problems in Engineering*, 2020, no. 1, p. 1–17. DOI: 10.1155/2020/7646527

[13] LI, L., JI, H. B. Radar emitter recognition based on cyclostationary signatures and sequential iterative least-square estimation. *Expert*

Systems with Applications, 2011, vol. 38, no. 3, p. 2140–2147. DOI: 10.1016/j.eswa.2010.07.155

[14] ZHU, B., JIN, W. D. Radar emitter signal recognition based on EMD and neural network. *Journal of Computers*, 2012, vol. 7, no. 6, p. 1413–1420. DOI: 10.4304/jcp.7.6.1413-1420

[15] GOK, G., ALP, Y. K., ARIKAN, O. A new method for specific emitter identification with results on real radar measurements. *IEEE Transactions on Information Forensics and Security*, 2020, vol. 15, p. 3335–3346. DOI: 10.1109/TIFS.2020.2988558

[16] BHATTI, S. G., BHATTI, A. I. Radar signals intrapulse modulation recognition using phase-based STFT and BiLSTM. *IEEE Access*, 2022, vol. 10, p. 80184–80194. DOI: 10.1109/ACCESS.2022.3195273

[17] LIU, C. C., ZHOU, F., WANG, F. Fault diagnosis of commutation failure using wavelet transform and wavelet neural network in HVDC transmission system. *IEEE Transactions on Instrumentation and Measurement*, 2021, vol. 70, p. 1–8. DOI: 10.1109/TIM.2021.3115574

[18] HE, J. S., HUANG, S., CHANG, S., et al. Radio frequency fingerprint identification with hybrid time-varying distortions. *IEEE Transactions on Wireless Communications*, 2023, vol. 22, no. 10, p. 6724–6736. DOI: 10.1109/TWC.2023.3245070

[19] WONG, L. J., HEADLEY, W. C., MICHAELS, A. J. Specific emitter identification using convolutional neural network-based IQ imbalance estimators. *IEEE Access*, 2019, vol. 7, p. 33544–33555. DOI: 10.1109/ACCESS.2019.2903444

[20] LIU, Y. H., XU, H., QI, Z. S., et al. Specific emitter identification against unreliable features interference based on time-series classification network structure. *IEEE Access*, 2020, vol. 8, p. 200194–200208. DOI: 10.1109/ACCESS.2020.3035813

[21] LIAO, L. L., LI, H. Z., CAO, L. Z., et al. Fast Fourier transform with multihead attention for specific emitter identification. *IEEE Transactions on Instrumentation and Measurement*, 2024, vol. 73, p. 1–12. DOI: 10.1109/TIM.2023.3338706

[22] DENG, P. F., HONG, S. H., QI, J., et al. A lightweight transformer-based approach of specific emitter identification for the automatic identification system. *IEEE Transactions on Information Forensics and Security*, 2023, vol. 18, p. 2303–2317. DOI: 10.1109/TIFS.2023.3266627

[23] QIAN, Y. H., QI, J., KUAI, X. Y., et al. Specific emitter identification based on multi-level sparse representation in automatic identification system. *IEEE Transactions on Information Forensics and Security*, 2021, vol. 16, p. 2872–2884. DOI: 10.1109/TIFS.2021.3068010

[24] QU, L. Z., YANG, J. A., LIU, H., et al. Method for individual identification of communication radiation source embedded in attention mechanism (in Chinese). *Systems Engineering and Electronics*, 2022, vol. 44, no. 1, p. 20–27. DOI: 10.12305/j.issn.1001-506X.2022.01.03

[25] ZHANG, W. X., ZHANG, F. S., ZHAO, Z. K., et al. Radar specific emitter identification via the Attention-GRU model. *Digital Signal Processing*, 2023, vol. 142, p. 1–12. DOI: 10.1016/j.dsp.2023.104198

[26] SUN, M. H., TENG, J. Z., LIU, X. Y., et al. Few-shot specific emitter identification: A knowledge, data, and model-driven fusion framework. *IEEE Transactions on Information Forensics and Security*, 2025, vol. 20, p. 3247–3259. DOI: 10.1109/TIFS.2025.3550080

[27] WANG, H., WANG, Q., FU, X., et al. VC-SEI: Robust variable-channel specific emitter identification method using semi-supervised domain adaptation. *IEEE Transactions on Wireless Communications*, 2024, vol. 23, no. 12, p. 18228–18239. DOI: 10.1109/TWC.2024.3463740

[28] LI, M. F., DOU, Z., JIANG, H., et al. Adversarial attack generation based on meta learning in specific emitter identification. *IEEE Wireless Communications Letters*, 2025, vol. 14, no. 2, p. 285–289. DOI: 10.1109/LWC.2024.3495677

[29] TYLER, H. J., FADUL, M. K. M., REISING, D. R. Considerations, advances, and challenges associated with the use of specific emitter identification in the security of Internet of Things deployments: A survey. *Information*, 2023, vol. 14, no. 9, p. 1–49. DOI: 10.3390/info14090479

[30] CHEN, W., WANG, Y., SONG, J., et al. Open-set HRRP recognition based on convolutional neural network. *The Journal of Engineering*, 2019, no. 21, p. 7701–7704. DOI: 10.1049/joe.2019.0706

[31] ZHAO, Y., WANG, X., LIN, Z., et al. Multi-classifier fusion for open-set specific emitter identification. *Remote Sensing*, 2022, vol. 14, no. 9, p. 1–22. DOI: 10.3390/rs14092226

[32] WANG, C., WANG, Y., ZHANG, Y., et al. Open-set specific emitter identification based on prototypical networks and extreme value theory. *Applied Sciences*, 2023, vol. 13, no. 6, p. 1–20. DOI: 10.3390/app13063878

[33] MOU, C., ZHANG, J., FAN, X. P., et al. COLA-Net: Collaborative attention network for image restoration. *IEEE Transactions on Multimedia*, 2022, vol. 24, p. 1366–1377. DOI: 10.1109/TMM.2021.3063916

[34] SU, J. N., GAN, M., CHEN, G. Y., et al. Global learnable attention for single image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, vol. 45, no. 7, p. 8453 to 8465. DOI: 10.1109/TPAMI.2022.3229689

[35] HU, H. G., LIU, A. Z., GUAN, Q. Adaptively customizing activation functions for various layers. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, vol. 34, no. 9, p. 6096–6107. DOI: 10.1109/TNNLS.2021.3133263

[36] CHEN, G. Y., PENG, P. X., WANG, X. Q., et al. Adversarial reciprocal points learning for open-set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, vol. 44, no. 11, p. 8065–8081. DOI: 10.1109/TPAMI.2021.3106743

[37] WU, L. W., ZHAO, Y., FENG, M., et al. Specific emitter identification using IMF-DNA with a joint feature selection algorithm. *Electronics*, 2019, vol. 8, no. 9, p. 1–16. DOI: 10.3390/electronics8090934

[38] CHEN, T. W., JIN, W. D., LI, J. Feature extraction using surrounding-line integral bispectrum for radar emitter signal. In *IEEE International Joint Conference on Neural Networks*. Hong Kong, 2008, p. 294–298. DOI: 10.1109/IJCNN.2008.4633806

[39] CHANG, C. C., LIN, C. J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, vol. 2, no. 3, p. 1–27. DOI: 10.1145/1961189.1961199

[40] WU, J. Y., WU, B., NIU, H. N., et al. A novel support vector machine based radar individual identification algorithm under inconsistent noise condition. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2020)*. Waikoloa (HI, USA), 2020, p. 2847–2850. DOI: 10.1109/IGARSS39084.2020.9323850

[41] YANG, X. L., ZHANG, G., SONG, H. Ship recognition based on HRRP via multi-scale sparse preserving method. *Journal of Systems Engineering and Electronics*, 2024, vol. 35, no. 3, p. 599 to 608. DOI: 10.23919/JSEE.2023.000136

[42] AVOLA, D., CASCIO, M., CINQUE, L., et al. 2-D skeleton-based action recognition via two-branch stacked LSTM-RNNs. *IEEE Transactions on Multimedia*, 2020, vol. 22, no. 10, p. 2481–2496. DOI: 10.1109/TMM.2019.2960588

[43] LI, H. B., ZHAO, J., ZHAO, Y. Q., et al. Specific emitter identification through demodulation embedding in convolutional neural networks using raw real signals. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2024)*. Athens (Greece), 2024, p. 10422–10425. DOI: 10.1109/IGARSS53475.2024.10642058

[44] XIN, Z., CHAO, H., LU, Y. P., et al. Fault diagnosis for small samples based on attention mechanism. *Measurement*, 2022, vol. 187, p. 1–14. DOI: 10.1016/j.measurement.2021.110242

[45] HENDRYCKS, D., GIMPEL, K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *5th International Conference on Learning Representations*. Toulon (France), 2017, p. 1–12. DOI: 10.48550/arXiv.1610.02136

[46] YANG, H. M., ZHANG, X. Y., YIN, F., et al. Robust classification with convolutional prototype learning. In *IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City (UT, USA), 2018, p. 3474–3482. DOI: 10.1109/CVPR.2018.00366

[47] YANG, H., YUAN, C. F., ZHANG, L., et al. STA-CNN: Convolutional spatial-temporal attention learning for action identification. *IEEE Transactions on Image Processing*, 2020, vol. 29, p. 5783–5793. DOI: 10.1109/TIP.2020.2984904

About the Authors ...

Jinxin XU (corresponding author) was born in Nantong, China. He received his Ph.D. degree in Electronics and Communication Engineering from Hohai University, Nanjing, China, in 2021. His research interests include image processing, radar signal processing, and machine learning.

Jin HU was born in Wuhan, China. He received his Ph.D. degree in Electronics and Communication Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2010. His research interests include radar signal processing, and machine learning.

Wenqiang JIANG was born in Xiamen, China. He received his M.SC. degree in Electronics and Communication Engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2013. His research interests include radar signal processing, and machine learning.

Zhiqiang ZHANG was born in Huanggang, China. He received his B.Sc. degree from the South China University of Technology, Guangzhou, China, in 2023. He is pursuing the master's degree in Electronic Science and Technology in the 724th Research Institute of CSIC. His major research interests are radar signal processing and deep learning.

Mingyuan YANG was born in Suzhou, China. He received his M.Sc. degree in Electronics and Communication Engineering from Soochow University, Suzhou, China, in 2004. His research interests include radar signal processing.

Zheng ZOU was born in Nanjing, China. He is pursuing the Ph.D. degree in Electronics and Communication Engineering from the Southeast University. His major research interests are radar signal processing and deep learning.