

## Research Article

# Real-Time Multisignal Detection and Identification in Known and Unknown HF Channels: A Deep Learning Method

Huang Lin  and Xuchu Dai 

The CAS Key Laboratory of Wireless-Optical Communications, University of Science and Technology of China, Hefei 230026, China

Correspondence should be addressed to Huang Lin; [linhuang@mail.ustc.edu.cn](mailto:linhuang@mail.ustc.edu.cn)

Received 24 June 2022; Revised 24 August 2022; Accepted 1 September 2022; Published 15 September 2022

Academic Editor: Lei Zhang

Copyright © 2022 Huang Lin and Xuchu Dai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The high-frequency (HF) signal detection and identification plays an important role in HF communications, and it is challenging since the HF environment randomly varies. Due to the success of deep learning (DL) methods in the fields of computer vision and natural language processing, some researchers adopt DL-based object detection methods to detect and identify signals in wideband spectrograms and achieve good performance. However, the existing DL-based methods are not suitable for real-time HF signal detection, and their performance will be significantly degraded when these methods are applied to an unknown HF environment. In this paper, we design a novel multiresolution signal detection and identification network for real-time HF signal detection and identification and propose a domain adaptation method to adapt the network to unknown environments. The experimental results show that the running speed and accuracy of our designed network are superior to ones of the existing DL-based networks in different HF environments, and the proposed domain adaptation method can achieve obvious performance improvement in unknown environments.

## 1. Introduction

The HF communication in the frequency range of 2 MHz to 30 MHz is widely used in military and civilian life due to its flexibility and long-distance transmission capability [1]. Currently, the HF channel is crowded and the HF environment randomly varies, where there are multipath delay, Doppler frequency shift, fading, and serious interference [2]. The urgent demand for accurate and real-time HF signal detection and identification in the military field and civilian field has attracted increasing attention from researchers, which requires detecting and identifying the signals of interest in the HF band as soon as they appear. However, there are still few research works on the real-time signal detection and identification in known and unknown HF environments.

The traditional HF signal detection and identification methods divide the above demand into two tasks: signal detection [3] and modulation identification [4]. Signal detection methods, including threshold-based methods, nonthreshold-based methods, and DL-based methods [5,

6], usually perform in a narrow band to detect the presence of signals. Modulation identification methods mainly use feature-based (FB) methods. The popular features for the modulation identification include instantaneous time features [7], statistical features [8], transform features [9], and deep learning features [10].

DL-based methods have made great successes in computer vision and natural language processing for different tasks, such as object detection [11, 12], human emotions detection [13], facial expression recognition [14], and domain adaptation [15]. Recently, some researchers have introduced the DL-based object detection methods into signal detection and identification [16–18]. In [16], a DL-based spectrum sensing approach was presented for cognitive radio communication. In [17], researchers regarded the time-frequency spectrum as an image and used SSD for signal detection and identification. In [18], researchers modified the CenterNet for HF signal detection and identification in wideband spectrograms and achieved excellent performance. These works have shown the great potential of DL-based methods in HF communication fields

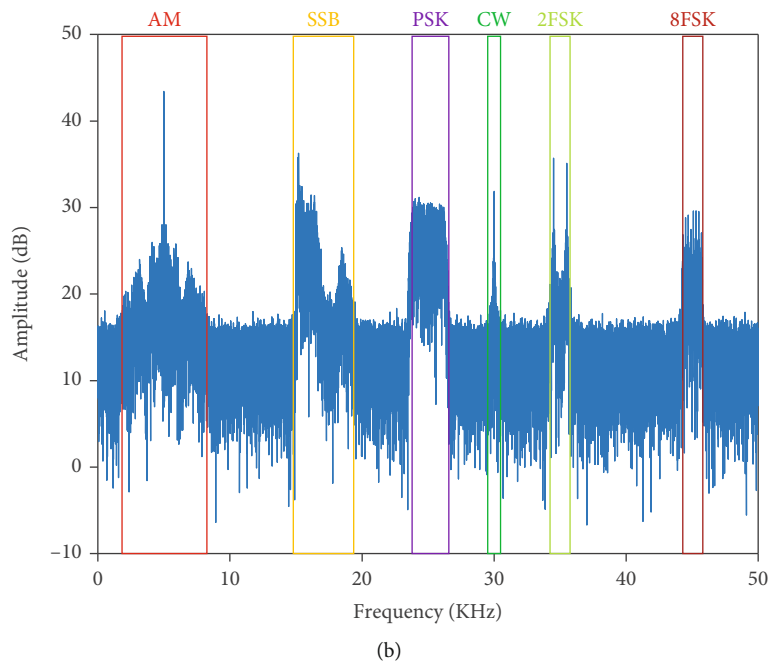
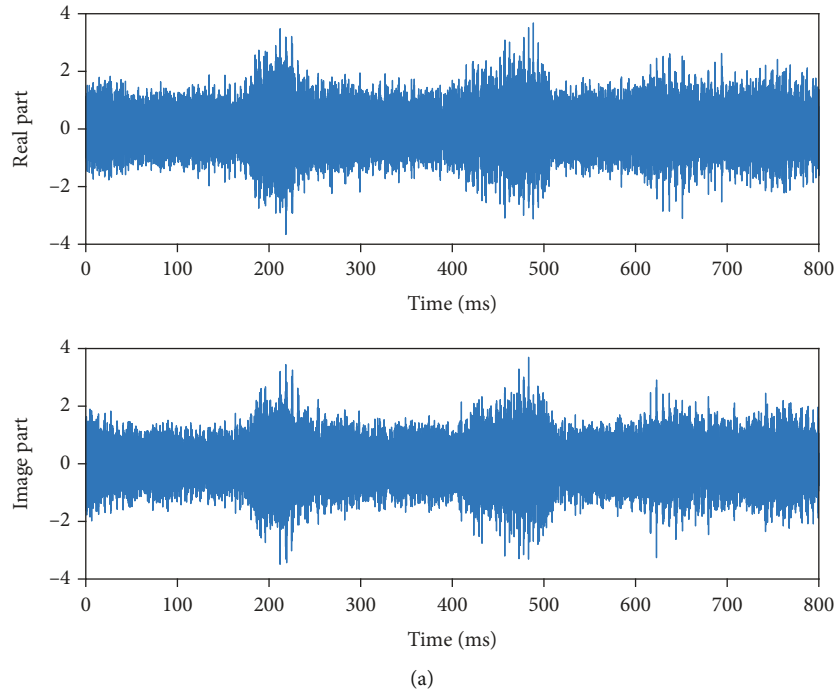
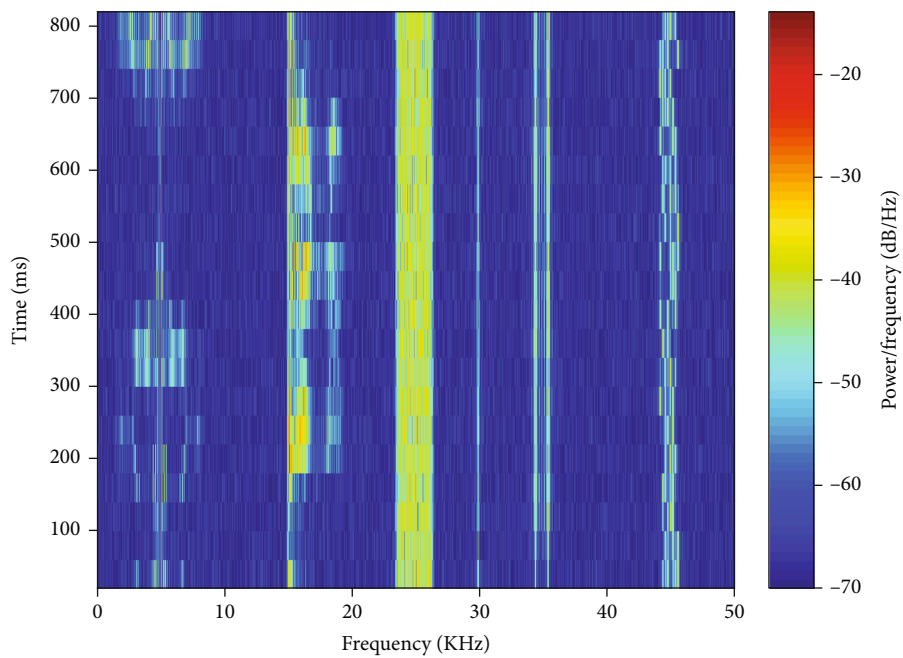
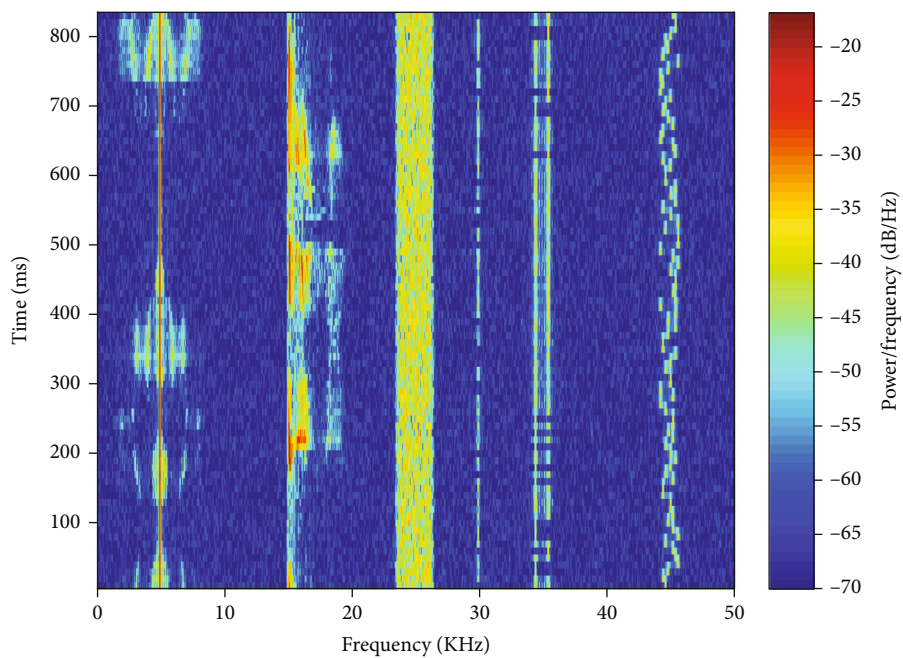


FIGURE 1: Continued.



(c)



(d)

FIGURE 1: Continued.

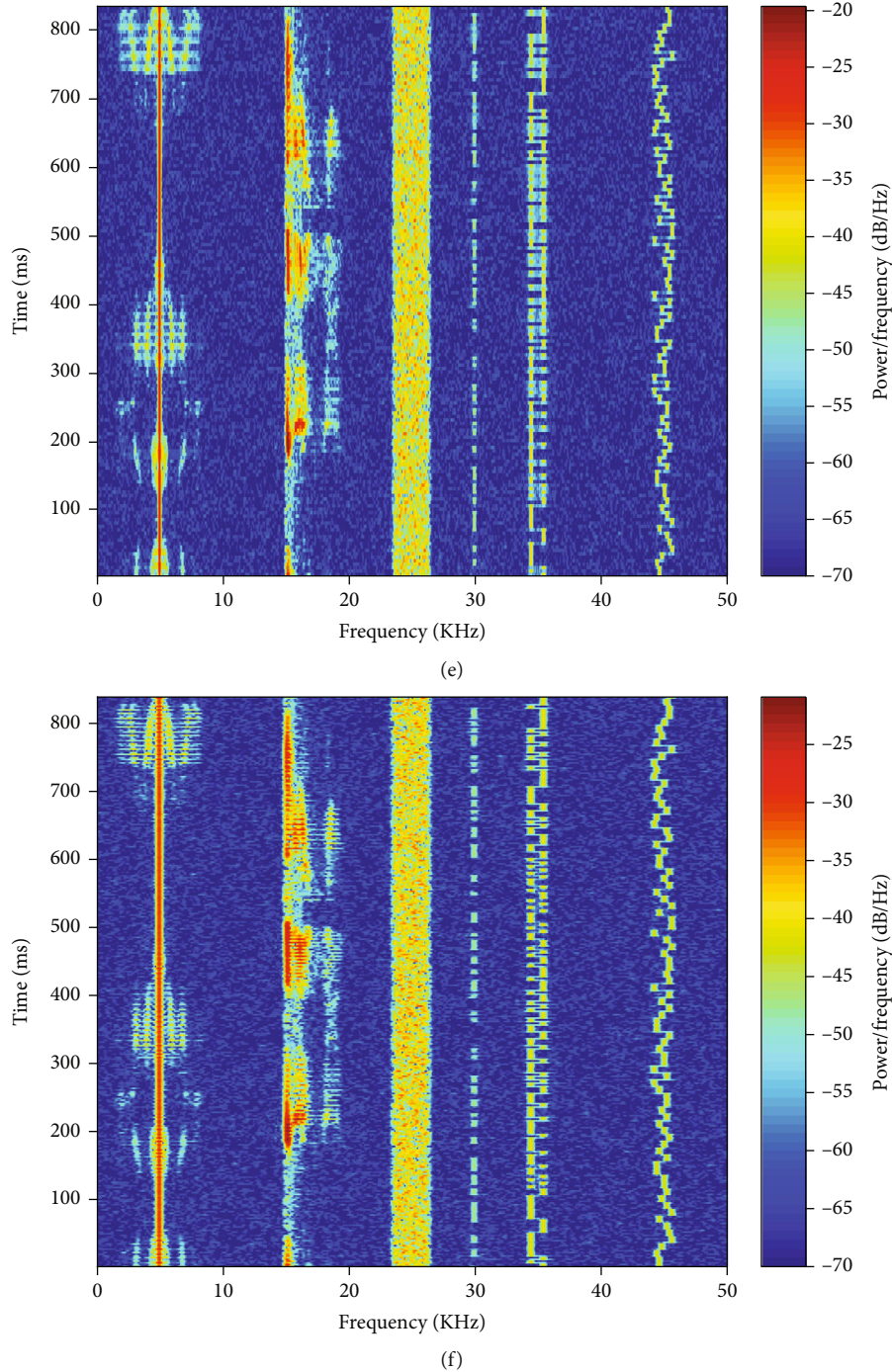


FIGURE 1: The time-frequency information of  $r(t)$  under AWGN channels. (a) The time sequence of  $r(t)$ . (b) The frequency spectrum and label of  $r(t)$ . (c)  $P_1(t, f)$  with  $\Delta t_1 = 40$  ms and  $\Delta f_1 = 25/4\pi$  Hz. (d)  $P_2(t, f)$  with  $\Delta t_2 = 10$  ms and  $\Delta f_1 = 100/4\pi$  Hz. (e)  $P_3(t, f)$  with  $\Delta t_3 = 5$  ms and  $\Delta f_1 = 200/4\pi$  Hz. (f)  $P_4(t, f)$  with  $\Delta t_4 = 2.5$  ms and  $\Delta f_4 = 400/4\pi$  Hz.

and achieved state-of-the-art performance in known environments.

However, there may be some disadvantages in applying the above-mentioned works to real-time HF signal detection and identification. First, these works use long-time signals for the detection and identification, which cannot meet requirements of real time. Second, these works treat the time-frequency spectrum of wideband signals as an image

and only use one time-frequency spectrum as input, so the details of different transmitted signals may be lost. Third, these works assume that the signals used for training and testing the networks are gathered from the same environment without considering how to ensure the performance of signal detection and identification in unknown environments. However, HF environment can be significantly different in time, frequency, weather, and place, and it is also

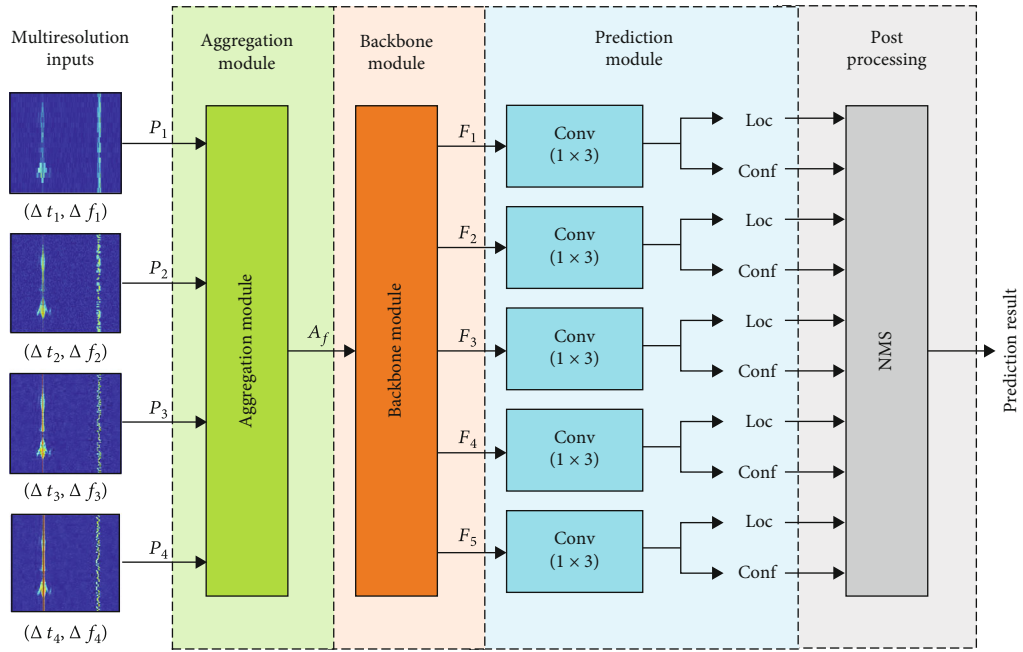


FIGURE 2: The overall network structure of MSDIN.

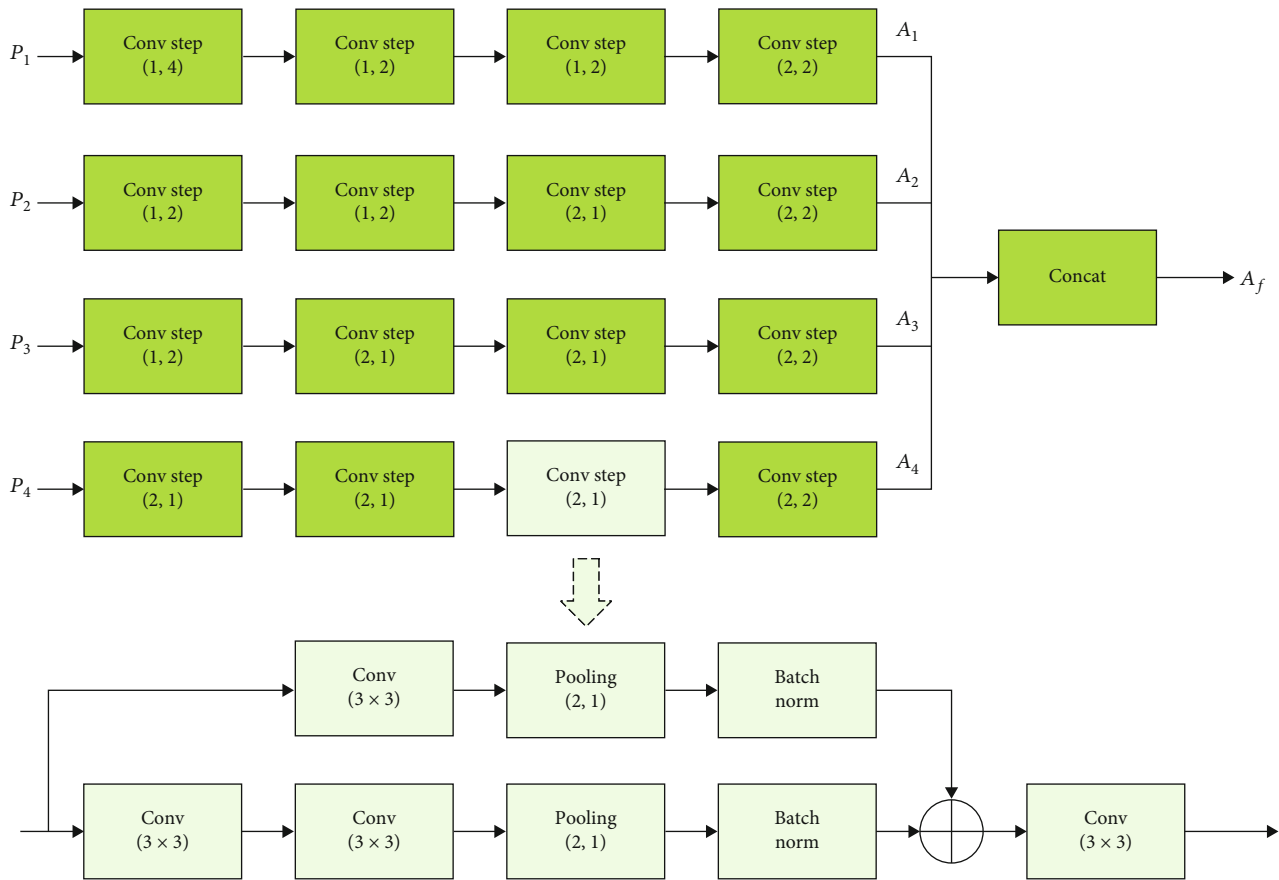


FIGURE 3: The structure of the aggregation module.

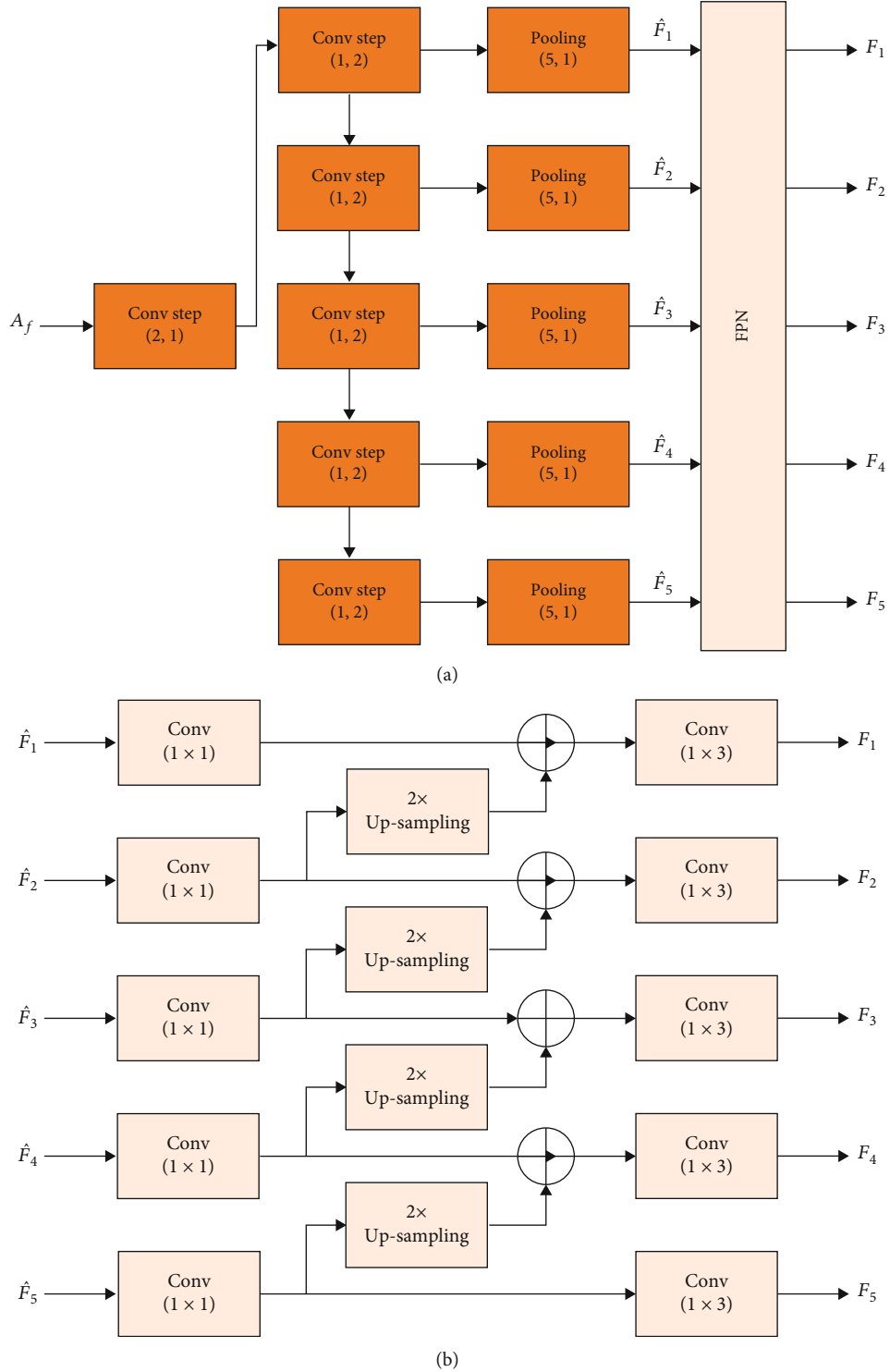


FIGURE 4: The structure of the backbone module and FPN module. (a) The structure of the backbone module. (b) The structure of the FPN module.

impossible to obtain sufficient labeled signals for all possible HF environments. Thus, the problem of how to ensure the performance of HF signal detection and identification in unknown environments naturally appears.

The main contributions of our work are as follows. (1) We adopt the multiresolution time-frequency spectra as fea-

ture representation and design an efficient network for HF signal detection and identification, and the experimental results show that the proposed network has distinct improvements in running speed and accuracy. (2) We adapt the proposed network to unknown HF environments by using the domain adaptation method, and the experimental



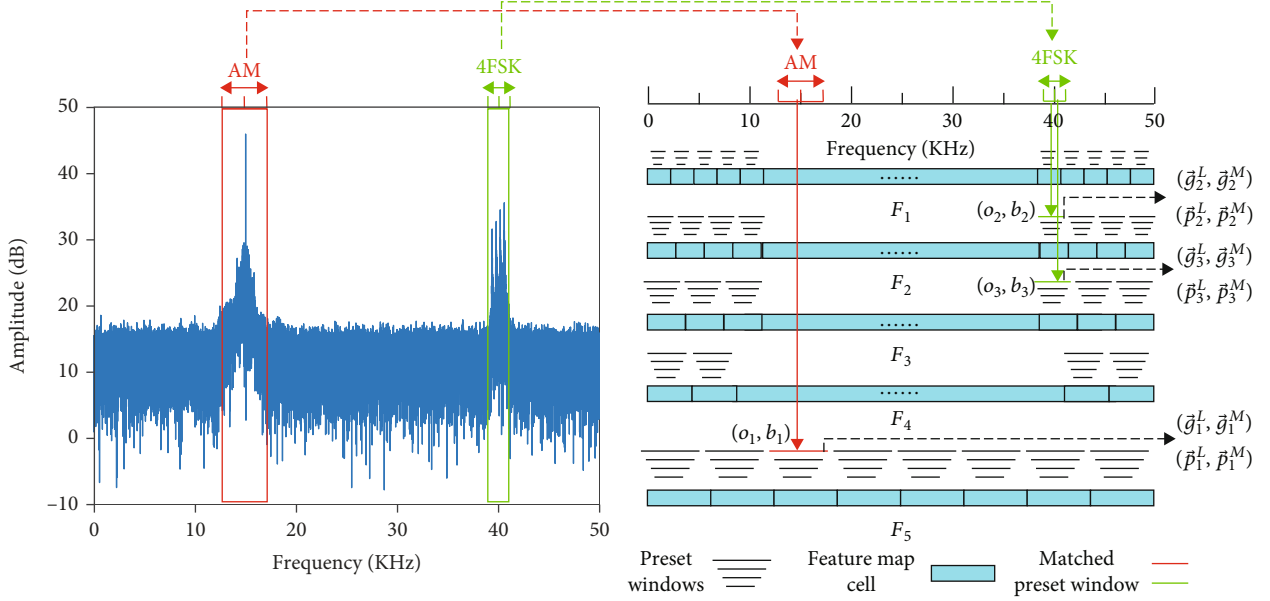


FIGURE 5: The matching process between the transmitted signals and preset windows.

results demonstrate the effectiveness of our proposed method. (3) We evaluate the robustness of our proposed network and methods in various environment conditions and analyze and compare the performance of the network under different environment conditions.

## 2. Signal Model and Feature Representation

**2.1. Signal Model.** Consider a wideband HF receiver with bandwidth  $B$ , and assume the receiver captures  $K$  different signals emitted by HF transmitters. Without losing generality, the baseband model of the captured signal can be formulated as

$$r(t) = \sum_{i=1}^K h_i(t) * s_i(t) + n(t) = \sum_{i=1}^K r_i(t) + n(t), \quad (1)$$

where  $t \in [0, T]$  is the observation time range of  $r(t)$ ,  $s_i(t)$  is the signal emitted by the  $i$ -th HF transmitter,  $h_i(t)$  is the impulse response of the transmitting channel from the  $i$ -th transmitter to the receiver,  $r_i(t) = h_i(t) * s_i(t)$  is the received signal corresponding to  $s_i(t)$ , “ $*$ ” denotes linear convolution, and  $n(t)$  is the additional noise. The typical channel models in HF environment include the additive white Gaussian noise (AWGN) channel, Watterson channel, and Rayleigh fading channel [19]. Generally, the transmitted signal  $s_i(t)$  is a modulated signal, which can be modeled as

$$s_i(t) = a_i(t) e^{-j(2\pi f_i t + \theta_i(t))}, \quad (2)$$

where  $a_i(t)$ ,  $f_i$ , and  $\theta_i(t)$  are the instantaneous envelope, carrier frequency, and instantaneous phase of  $s_i(t)$ . The modulation type of  $s_i(t)$  is denoted as  $m_i$ . The modulation types of transmitted signals widely used in HF communications include amplitude modulation (AM), frequency modulation (FM), single side band (SSB), continuous wave (CW), frequency-shift keying (FSK), phase-shift keying (PSK), amplitude-shift keying (ASK), and Gaussian minimum shift keying (GMSK).

Moreover, the following constraint conditions are assumed in expressions (1) and (2).

- (1)  $m_i \in \mathcal{M}$ , where  $\mathcal{M}$  is the set of possible modulation types of transmitted signals. In this paper, we set  $\mathcal{M} = \{\text{AM, FM, SSB, CW, 2FSK, 4FSK, 8FSK, PSK, ASK, GMSK}\}$  and  $|\mathcal{M}| = M$
- (2)  $F_i = [f_i^l, f_i^u]$  and  $B_i = f_i^u - f_i^l \ll B$ , where  $F_i$  and  $B_i$  are the effective frequency range and bandwidth of  $s_i(t)$ , respectively
- (3) The  $K$  received transmitted signals hardly overlap each other in the frequency domain

The aim of real-time HF signal detection and identification is to estimate the effective frequency range and modulation type of each transmitted signal  $s_i(t)$  by only utilizing the captured signal  $r(t)$  without any other prior information as fast as possible.

**2.2. Feature Representation.** According to the above-mentioned signal model, the  $K$  received transmitted signals are completely mixed in the time domain, while they can be easily separated from each other in the frequency domain. Transmitted signals with different modulation types and modulation parameters usually possess different time-varying characteristics and frequency characteristics, and the time-frequency information of the captured signal can clearly display these characteristics. Therefore, we utilize the time-frequency information of captured signals as a feature representation for HF signal detection and identification.

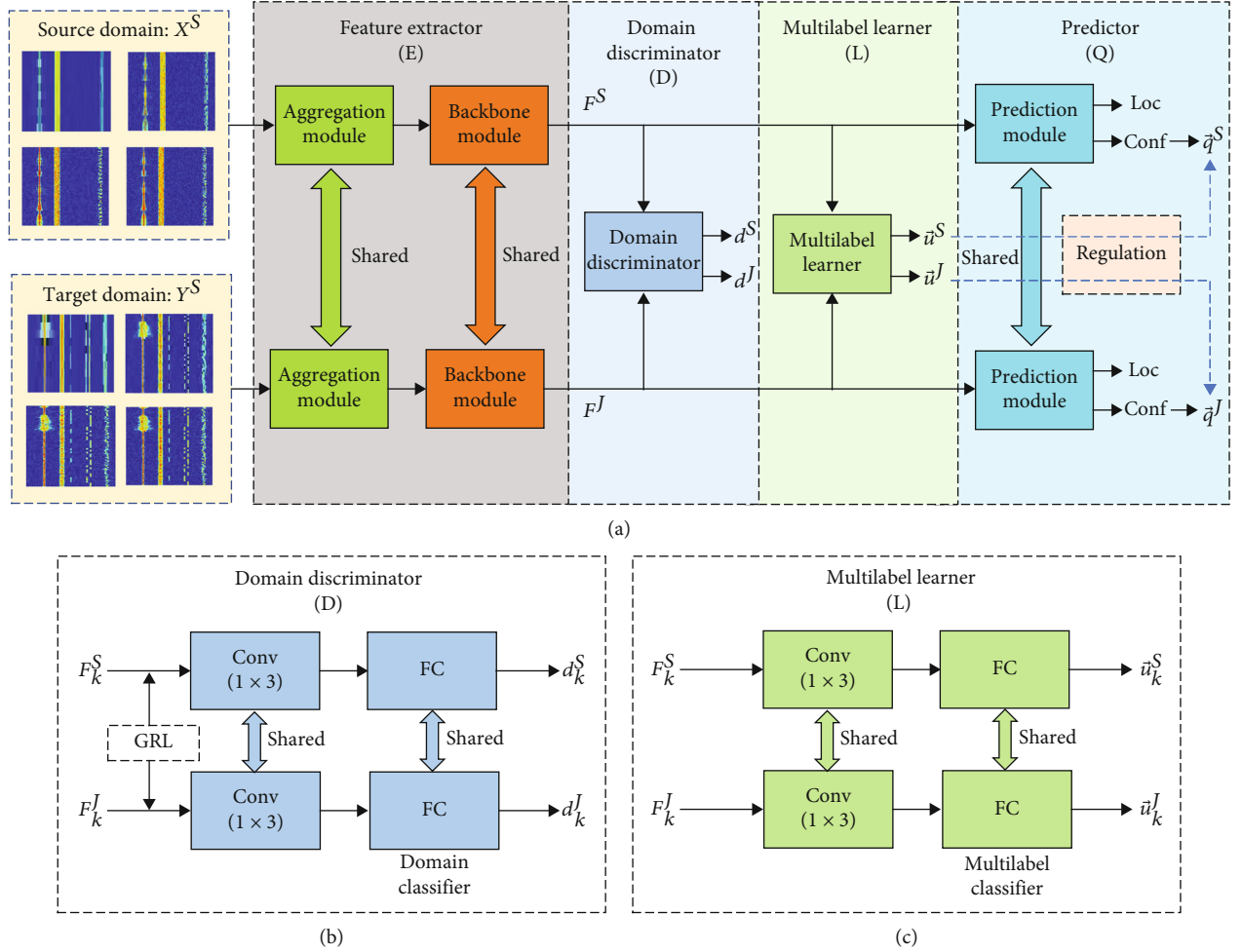


FIGURE 6: The overall structure of the MAAN. (a) The structure of the MAAN. (b) The structure of the domain discriminator. (c) The structure of the multilabel learner.

The short-time Fourier transform (STFT) is a commonly used method for time-frequency analysis. The STFT of the captured signal  $r(t)$  is

$$S(t, f) = \int_{-\infty}^{\infty} [r(\tau)w(\tau - t)]e^{-j2\pi f\tau} d\tau, \quad (3)$$

$$P(t, f) = |S(t, f)|^2, \quad (4)$$

where  $w(t)$  is the time window function,  $P(t, f)$  is the time-frequency spectrum of  $r(t)$ , and  $t \in [0, T]$  and  $f \in [0, B]$  are the time range and frequency range of  $P(t, f)$ , respectively. Denote  $\Delta t$  and  $\Delta f$  as the time resolution and frequency resolution of  $P(t, f)$ , respectively. According to [20], the time-frequency resolution of  $P(t, f)$  is determined by the time window function. In particular, the following relationship must be satisfied:

$$\Delta t \cdot \Delta f \geq 1/4\pi, \quad (5)$$

and the equality holds when the time window function is a Gaussian window function. From expression (5), it is obvious that  $\Delta t$  and  $\Delta f$  are a pair of contradictions.

Figure 1 shows the time-frequency information of a captured signal  $r(t)$  comprising six different transmitted signals, AM, SSB, PSK, CW, 2FSK, and 8FSK, where the observation time  $T = 800$  ms and bandwidth  $B = 50$  kHz. Figures 1(a) and 1(b) show the time sequence and amplitude of the frequency spectrum of  $r(t)$ , respectively. The effective frequency range and modulation type of each transmitted signal are labeled in Figure 1(b). Figures 1(c)–1(f) depict the time-frequency spectrum of  $r(t)$  with different time-frequency resolutions. It should be noted that from Figures 1(c) to 1(f), the frequency resolution gradually declines, while the time resolution gradually increases. Specifically, the time resolutions of these time-frequency spectra range from 40 ms to 2.5 ms, and the time-frequency resolutions satisfy with  $\Delta t \cdot \Delta f = 1/4\pi$ . These time-frequency spectra show the contradictions between  $\Delta t$  and  $\Delta f$ . For example, the time-frequency spectrum in Figure 1(c) has the highest frequency resolution and the lowest time resolution, where the frequency characteristics of 2FSK and 8FSK



TABLE 1: The details of different datasets.

Available dataset	Transmitted channels	$B$ [kHz]	$T$ [ms]	MDFS [Hz]	SNR [dB]	Training samples
AWGN dataset	AWGN	50	800	0	[-10,4]	20000
Watterson dataset	Watterson	50	800	100	[-10,4]	20000
Rayleigh dataset	Rayleigh fading	50	800	100	[-10,4]	20000
Time dataset	Rayleigh fading	50	[160,1600]	100	[-10,4]	20000
MDFS dataset	Rayleigh fading	50	800	[0,400]	[-10,4]	20000

are obvious, but the time-varying characteristics of AM, SSB, CW, 2FSK, and 8FSK are fuzzy. In contrast, the time-frequency spectrum in Figure 1(f) has the lowest frequency resolution and the highest time resolution, where the time-varying characteristics of AM, SSB, CW, and 2FSK are distinct, but the frequency characteristics of 8FSK are fuzzy.

The above example shows that it is difficult to represent the frequency characteristics and time-varying characteristics of various transmitted signals by utilizing only one time-frequency spectrum with the fixed time-frequency resolution. Therefore, it is necessary to utilize multiple time-frequency spectra with different time-frequency resolutions for robust HF signal detection and identification.

Let  $P_i$  express the time-frequency spectrum with the time resolution  $\Delta t_i$  and frequency resolution  $\Delta f_i$  in expression (4), and denote the feature representation of  $r(t)$  as

$$I = \{P_i, i = 1, \dots, N_f\}, \quad (6)$$

where  $N_f$  is the number of time-frequency spectra. Specifically, four different time-frequency spectra, as shown in Figures 1(c)–1(f), are adopted as inputs for HF signal detection and identification in this paper.

### 3. Deep Signal Detection and Identification

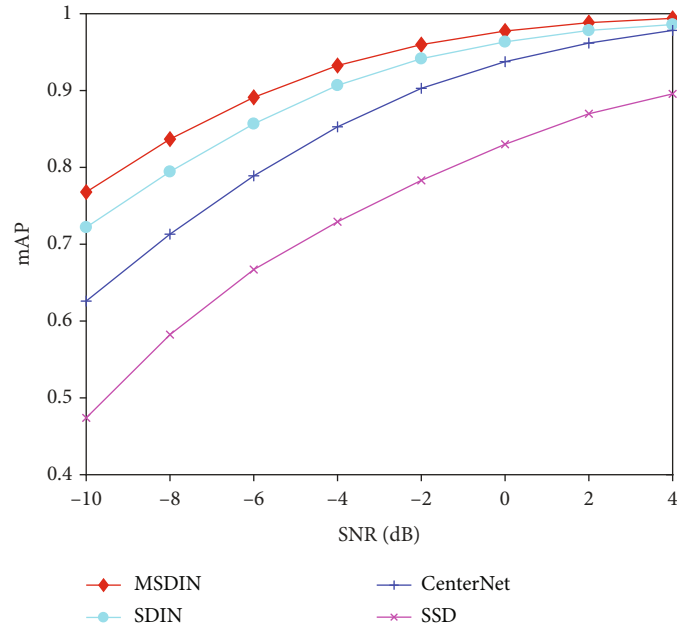
In this paper, the training signals with sufficient labels are called source domain signals, and the testing signals with no labels are called target domain signals. We denote the source domain and target domain as  $\mathcal{S}$  and  $\mathcal{T}$ , respectively. Specifically, the source domain signals and target domain signals are gathered from the same HF environment in the task of signal detection and identification in known environment, and the source domain signals and target domain signals are gathered from the different HF environments in the task of signal detection and identification in unknown environment.

*3.1. Signal Detection and Identification in Known Environment.* In this subsection, we present a multiresolution signal detection and identification network (MSDIN) for HF signal detection and identification in known environment. Assuming that the training signals and testing signals are gathered from the same environment, the task of HF signal detection and identification in known environment can be treated as a data-driven DL-based object detection task. The network model and training process of the MSDIN are shown in the following.

*3.1.1. Network Model.* The overall network structure of the MSDIN is shown in Figure 2, which is specially designed for HF signal detection and identification. To address the multiresolution inputs, we present an aggregation module to align and integrate the deep features of the inputs with different resolutions and obtain a comprehensive aggregation feature for the following detection. The backbone module generates multiscale features to address the signals with different modulation types and bandwidths. The prediction module is used to predict the modulation type and locations of each signal, and the post-processing deals with the outputs of prediction module to obtain the final results. The details of different modules are described in the following.

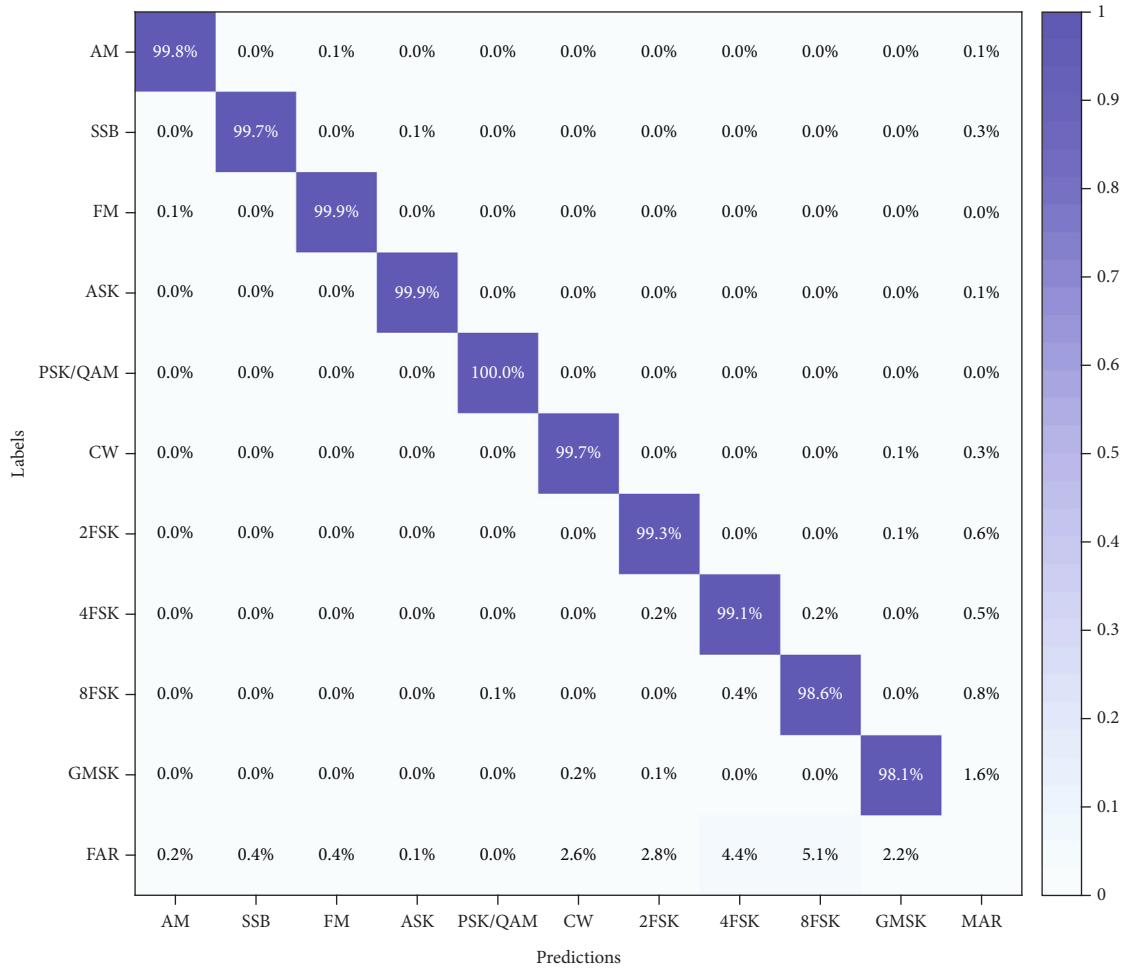
(1) *Aggregation Module.* The aggregation module is used to align and integrate the deep features of all multiresolution input to obtain a comprehensive aggregation feature for the following detection. Figure 3 shows the structure of aggregation module, which consists of multiple paths with different convolutional steps (Conv Step) and a concatenation layer (Concat). The structure of convolutional step is designed with the reference to ResNet [21]. As show in Figure 3, the deep input feature  $A_i$  is extracted from the input  $P_i$  through the  $i$ -th path of the aggregation module. Due to the special design of each path, all input features  $\{A_i, i = 1, \dots, N_f\}$  have the same size. Then, these features are concatenated in the concatenation layer, and a comprehensive aggregation feature  $A_f$  is obtained.

(2) *Backbone Module.* The backbone module accepts the aggregation feature  $A_f$  as input and generates a set of multiscale features to detect the signal with different modulation types and bandwidths. These multiscale features allow the MSDIN to predict the transmitted signals at different frequency sizes rather than at a single size. The structure of the backbone module is shown in Figure 4(a), which consists of a sequence of convolutional steps, pooling layers and a feature pyramid network (FPN) [22]. These convolutional steps down-sample the input in the frequency domain step by step, followed by the pooling layers with full-time pooling to obtain a set of full-time multiscale features  $\mathcal{F} = \{\hat{F}_i, i = 1, \dots, M_f\}$ , where  $M_f$  is the number of multiscale features. The frequency size of these features is decreasing step by step, while the time size is fixed to one. As shown in Figure 4(b), the FPN module is added at the end of the backbone module, which can enhance the representation ability and robustness of multiscale features and generate a set of enhanced full-time multiscale features  $\mathcal{F} = \{F_i, i = 1, \dots, M_f\}$ . In addition,  $\hat{F}_i$  and  $F_i$  have the same time-frequency size.



(a)

AWGN dataset at SNR = 4 dB



(b)

FIGURE 7: Continued.



(c)  
FIGURE 7: Continued.



FIGURE 7: The performance of different methods on the AWGN dataset. (a) The mAP vs SNR curves of different methods. (b) The confusion matrix of MSDIN at SNR = 4 dB. (c) The confusion matrix of MSDIN at SNR = 0 dB. (d) The confusion matrix of MSDIN at SNR = -4 dB.

(3) *Preset Windows*. Inspired by the anchor box used in SSD [11], we associate a set of preset windows with each cell of enhanced full-time multiscale features  $\mathcal{F} = \{F_i, i = 1, \dots, M_f\}$ , as shown in Figure 5. The preset windows bind the cell in a convolutional manner so that the default center of each preset window relative to its corresponding cell is fixed. The preset windows bind the cell in a convolutional manner so that the default center of each preset window relative to its corresponding cell is fixed. Specifically, for a feature with  $K$  cells, assume that each cell is bound with  $N_d$  different preset windows so that there are a total of  $KN_d$  preset windows applied in this feature. For each preset window, we compute  $(M + 1)$  modulation type confidence (Conf) and 2 frequency location offsets (Loc) relative to the preset window. This results in a total of  $(M + 3)N_d$  convolutional filters that are

applied around each cell in this feature, yielding  $(M + 3)KN_d$  outputs. Moreover, for each of multiscale features, the default number and bandwidth of bound preset windows can be set according to the frequency scale of each feature. Generally, the cells of large-scale feature are bound to wide preset windows, and vice versa.

(4) *Prediction Module*. The prediction module utilizes the enhanced full-time multiscale features  $\mathcal{F} = \{F_i, i = 1, \dots, M_f\}$  as input to predict the frequency location offset and the modulation type confidence of each preset window. Specifically, the predictors of each enhanced feature are the convolutional layers with kernel size  $(1 \times 3)$ , as shown in Figure 2. The center and bandwidth of preset windows can be fine-tuned by the prediction location offset to obtain an

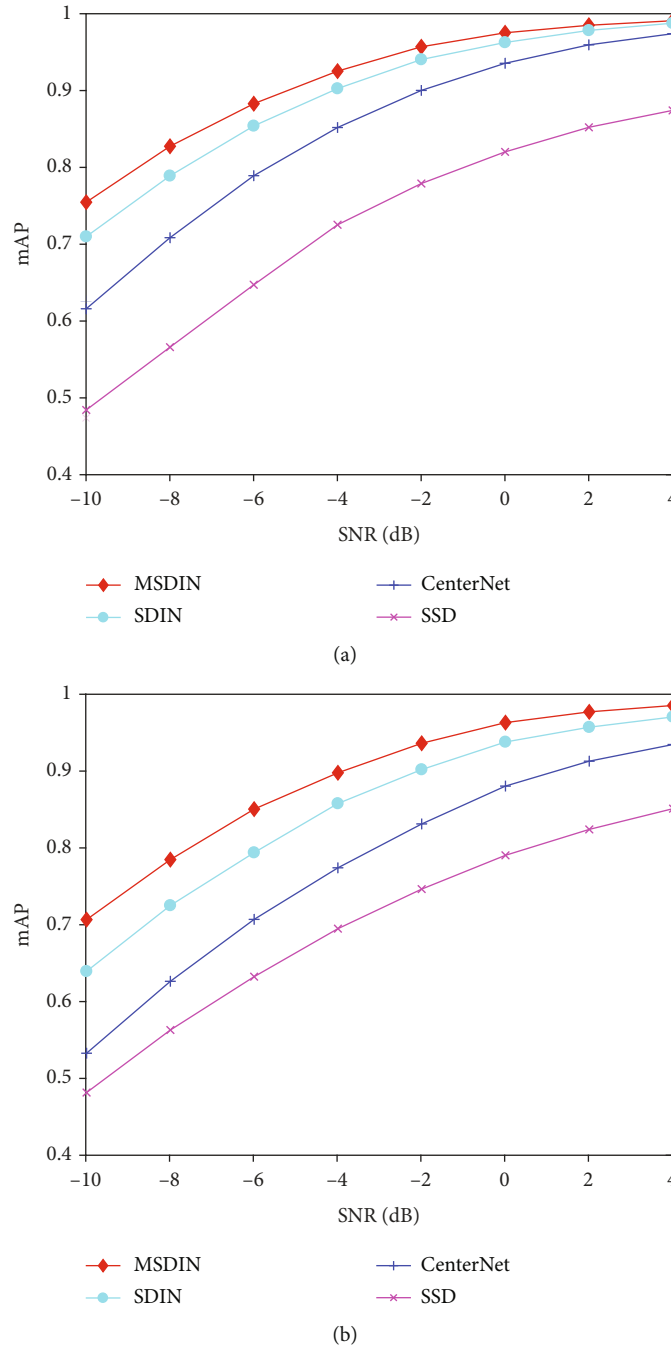


FIGURE 8: The performance of different methods on the Watterson dataset and Rayleigh dataset. (a) The mAP vs SNR curves on the Watterson dataset. (b) The mAP vs SNR curves on the Rayleigh dataset.

accurate prediction frequency location. These fine-tuned prediction windows allow the network to better match the transmitted signals with different center frequencies and bandwidths.

(5) *Post Processing Module.* The post processing module is used in the testing phase to address the repeated prediction windows obtained from the prediction module. As shown in Figure 5, each transmitted signal may match multiple preset windows, so we need to remove the repeated prediction windows to produce the final prediction results. It is realized

by nonmaximum suppression (NMS) algorithm [23], which can keep the windows with maximum confidence and remove the repeated windows.

3.1.2. *Training Process.* During the training phase, we need to assign each transmitted signal to specific preset windows and predict the location offset and confidence of each preset window. Once the assignment is decided, the loss function and backpropagation are applied to update the parameters of the network. The matching strategy and loss function

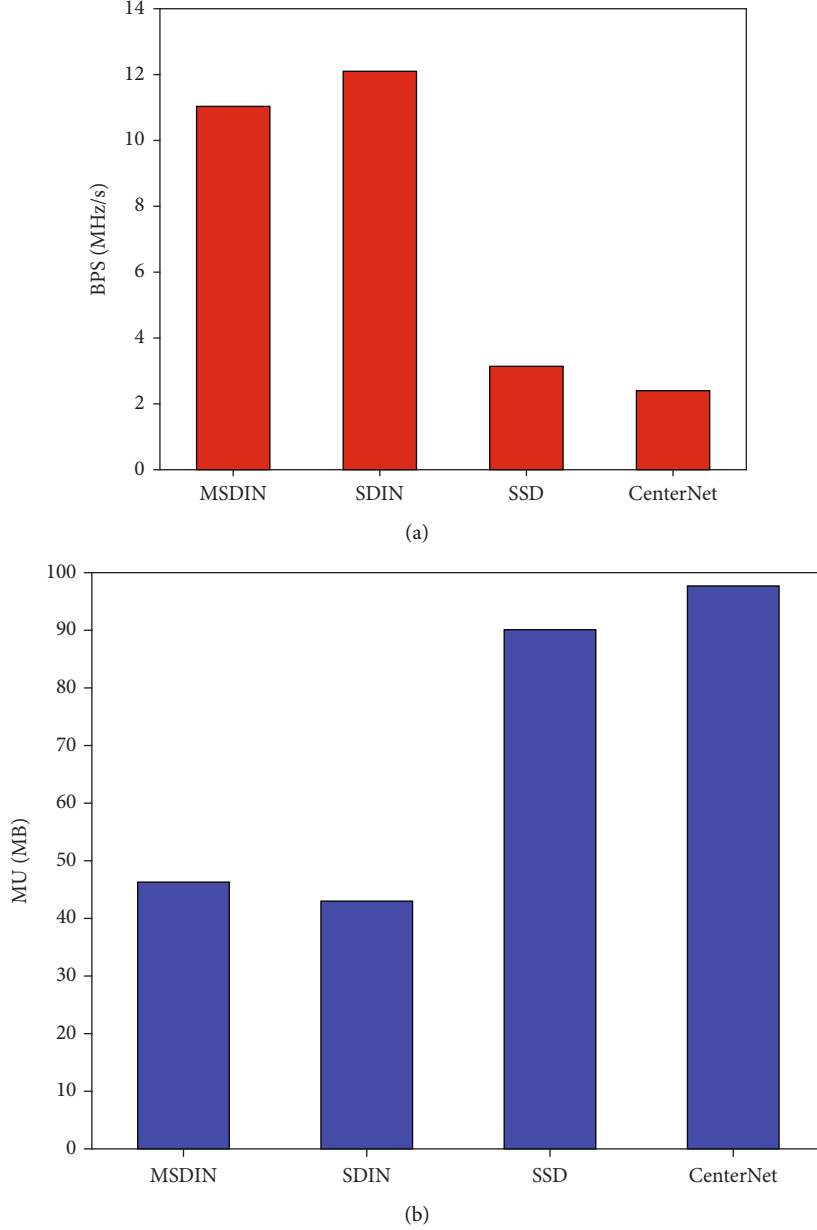


FIGURE 9: The running speed and model size comparison of different methods. (a) Running speed comparison. (b) Model size comparison.

are the keys to training the network, which will be described in detail in the following.

(1) *Matching Strategy*. During the training phase, we need to determine which preset window corresponds to a transmitted signal and then train the network accordingly. Inspired by the matching strategy of SSD, we first match each of the transmitted signals to the preset window with the maximum overlap and then match the transmitted signals to the preset windows with high overlap. In experiments, the overlapping threshold is set to 0.5. This matching strategy can simplify the learning process, which allows the network to predict high confidences for multiple preset windows with high overlap, rather than only picking the window with maximum overlap.

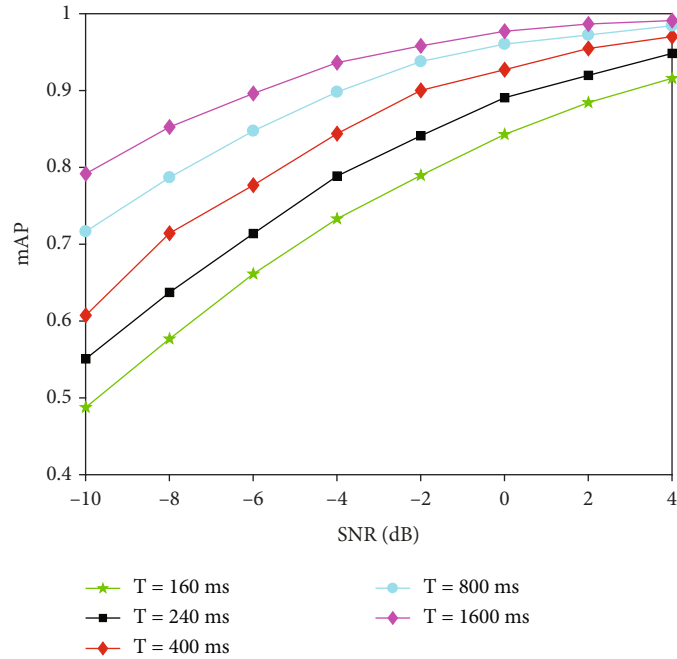
The matching process is shown in Figure 5. The frequency spectrum and ground truth of the captured signal are shown on the left of Figure 5, which captures two transmitted signals, AM and 4FSK. Following the above-mentioned matching strategy, the matched preset windows are painted with the color of the corresponding transmitted signal, while the color of unmatched preset windows remains unchanged. The matching result is shown on the right of Figure 5.

Denote all preset windows as

$$\mathcal{W} = \{\mathcal{W}_{mp} \cup \mathcal{W}_{ump}\}, \quad (7)$$

where  $\mathcal{W}_{mp}$  and  $\mathcal{W}_{ump}$  are the set of matched preset windows and unmatched preset windows, respectively. The





(a)

Rayleigh dataset at SNR = 0 dB, T = 1600 ms



(b)

FIGURE 10: Continued.



(c)



(d)

FIGURE 10: The performance of MSDIN with different duration times on the time dataset. (a) The mAP vs SNR curves of MSDIN with different duration times. (b) The confusion matrix of MSDIN at SNR = 0 dB, T = 1600 ms. (c) The confusion matrix of MSDIN at SNR = 0 dB, T = 800 ms. (d) The confusion matrix of MSDIN at SNR = 0 dB, T = 400 ms.

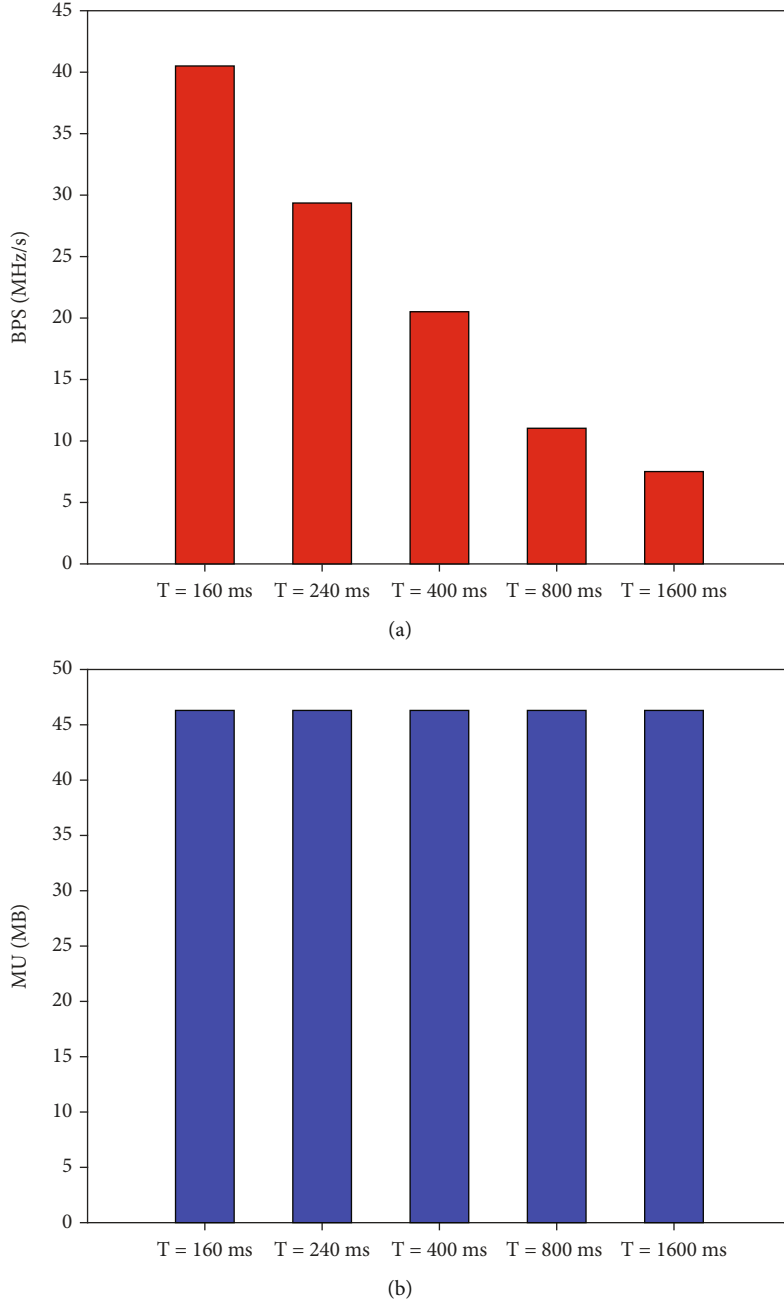


FIGURE 11: The running speed and model size comparison of MSDIN on the time dataset. (a) Running speed comparison with different duration times. (b) Model size comparison with different duration times.

ground truths of matched preset windows can be expressed as

$$G = \left\{ \left( \vec{g}_i^L, \vec{g}_i^M \right), w_i \in \mathcal{W}_{\text{mp}} \right\}, \quad (8)$$

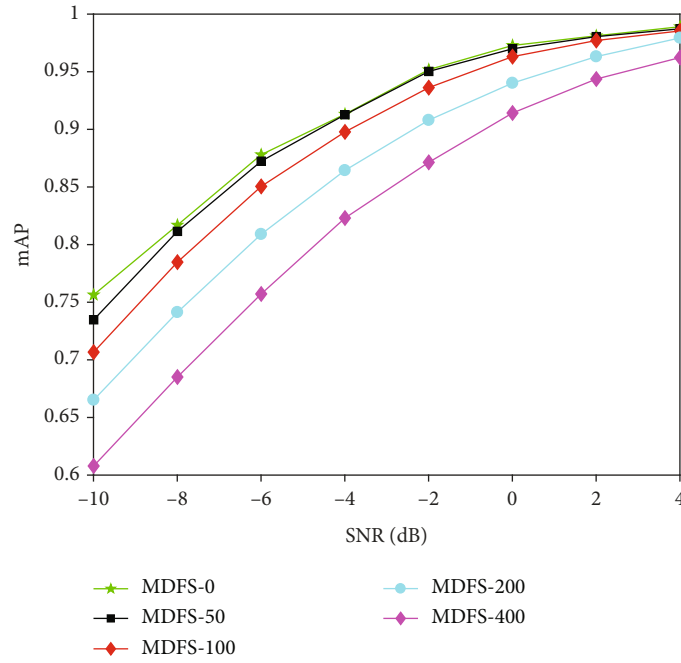
where  $w_i$  is the  $i$ -th matched preset window;  $\vec{g}_i^L = (g_i^o, g_i^b)$  is the frequency location of the transmitted signal matched to  $w_i$ ;  $g_i^o$  and  $g_i^b$  are ground truth center frequency and band-

width of the signal, respectively; and  $\vec{g}_i^M$  is the modulation type of the signal. In addition, the default center and bandwidth of  $w_i$  are denoted as  $o_i$  and  $b_i$ , respectively.

Denote the prediction results of all preset windows as

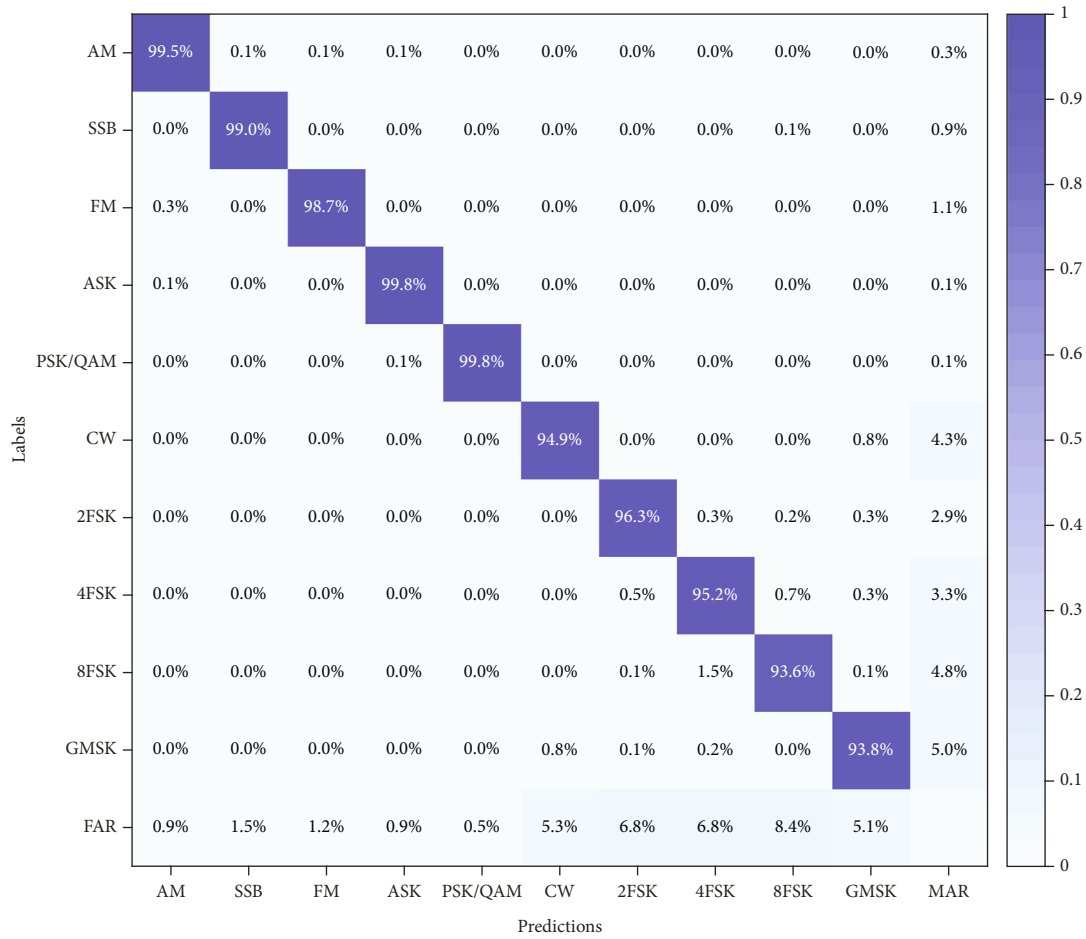
$$P_{\mathcal{W}} = \{P_{\text{mp}}, P_{\text{ump}}\}, \quad (9)$$

where  $P_{\text{mp}}$  and  $P_{\text{ump}}$  are the prediction results of matched preset windows and unmatched preset windows,



(a)

Rayleigh dataset at SNR = 0 dB, MDFS = 50 Hz



(b)

FIGURE 12: Continued.



(c)



(d)

FIGURE 12: The performance of MSDIN on the MDFS dataset. (a) The mAP vs SNR curves of MSDIN with different MDFS. (b) The confusion matrix of MSDIN at SNR = 0 dB, MDFS = 50 Hz. (c) The confusion matrix of MSDIN at SNR = 0 dB, MDFS = 100 Hz. (d) The confusion matrix of MSDIN at SNR = 0 dB, MDFS = 200 Hz.

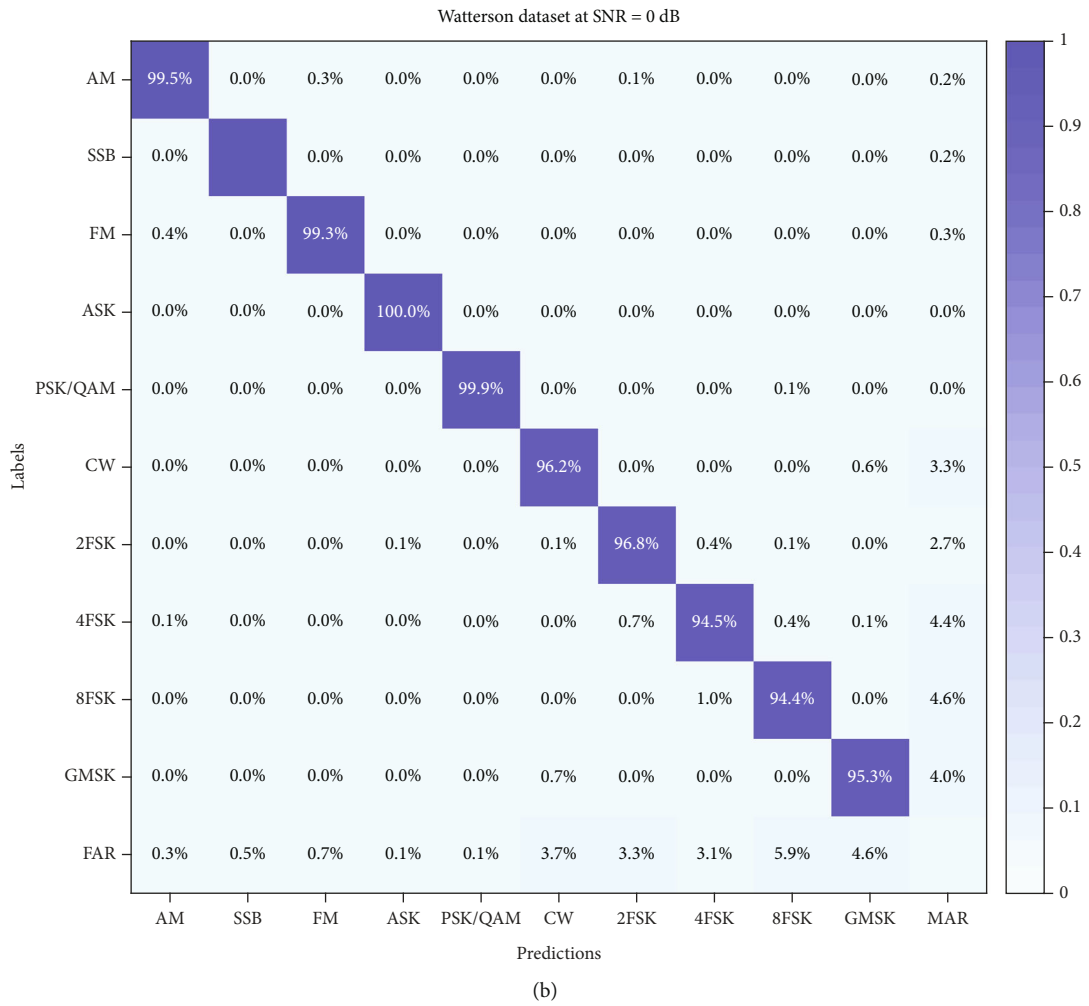
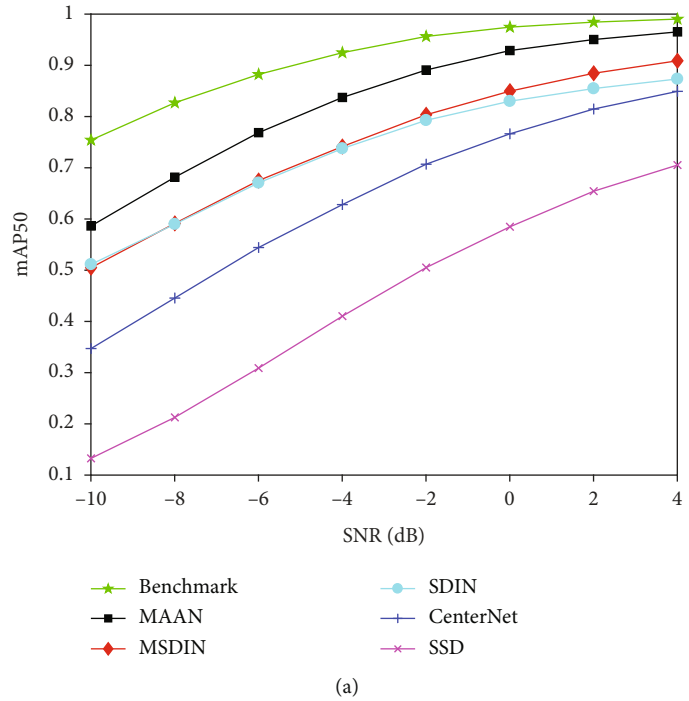
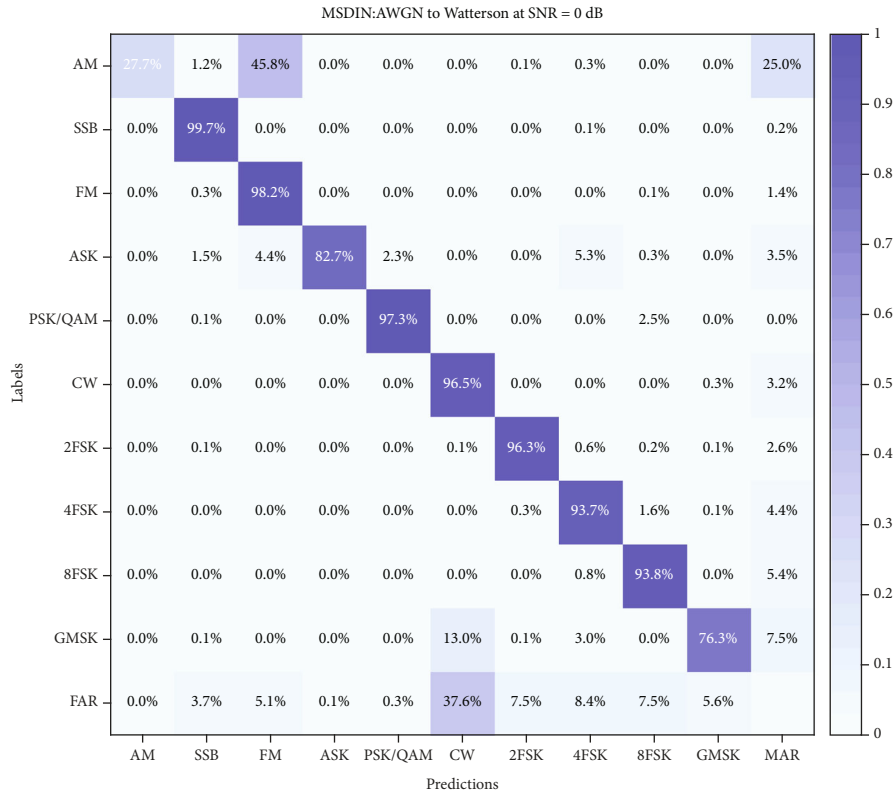
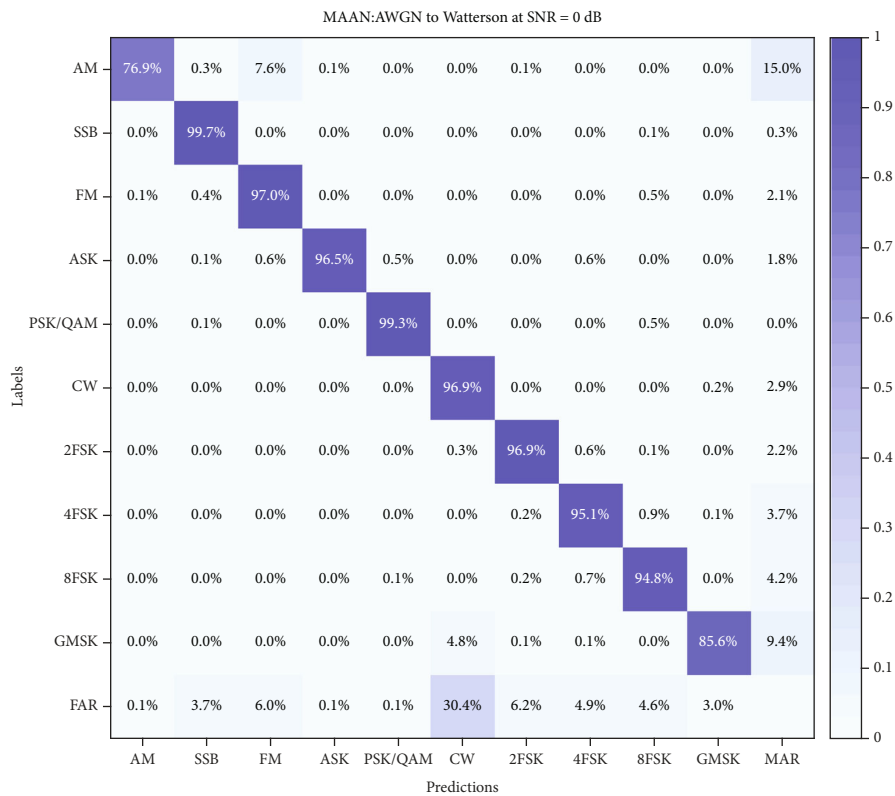


FIGURE 13: Continued.



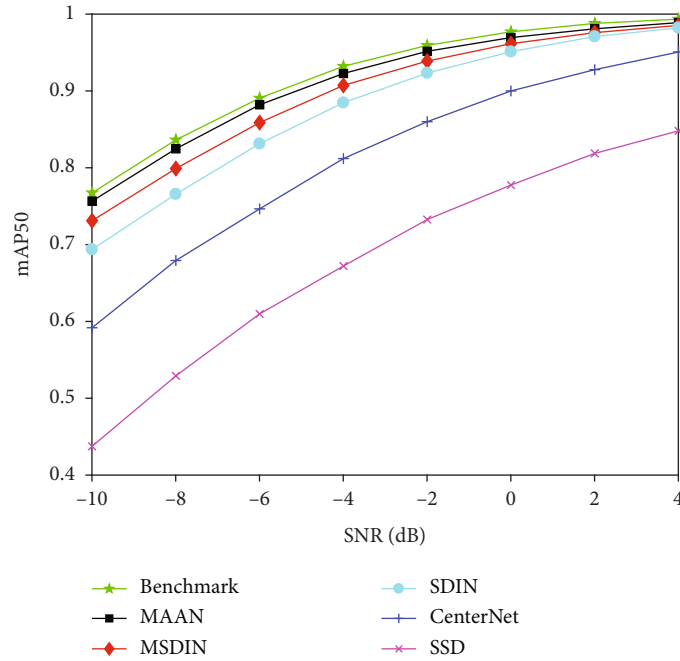


(c)



(d)

FIGURE 13: The performance of different methods in case 1. (a) The mAP vs SNR curves of different methods. (b) The confusion matrix of the benchmark at SNR = 0 dB. (c) The confusion matrix of the MSDIN at SNR = 0 dB. (d) The confusion matrix of the MAAN at SNR = 0 dB.



(a)

AWGN dataset at SNR = 0 dB



(b)

FIGURE 14: Continued.



FIGURE 14: Continued.



FIGURE 14: The performance of different methods in case 2. (a) The mAP vs SNR curves of different methods in an unknown environment. (b) The confusion matrix of the benchmark at SNR = 0 dB. (c) The confusion matrix of the MSDIN at SNR = 0 dB. (d) The confusion matrix of the MAAN at SNR = 0 dB.

respectively. The prediction results of matched preset windows can be expressed as

$$P_{\text{mp}} = \left\{ \left( \overset{\rightarrow L}{p}_i, \overset{\rightarrow M}{p}_i \right), w_i \in \mathcal{W}_{\text{mp}} \right\}, \quad (10)$$

where  $\overset{\rightarrow L}{p}_i = (p_i^o, p_i^b)$  is the prediction frequency location offsets of  $w_i$ ;  $p_i^o$  and  $p_i^b$  are the prediction center offset and bandwidth offset of  $w_i$ , respectively; and  $\overset{\rightarrow M}{p}_i$  is the modulation type confidence of  $w_i$ . Similarly, the prediction results of unmatched preset windows can be expressed as

$$P_{\text{ump}} = \left\{ \left( \overset{\rightarrow L}{p}_j, \overset{\rightarrow M}{p}_j \right), \tilde{w}_j \in \mathcal{W}_{\text{ump}} \right\}, \quad (11)$$

where  $\tilde{w}_j$  is the  $j$ -th unmatched preset window and  $\overset{\rightarrow L}{p}_j = (p_j^o, p_j^b)$  and  $\overset{\rightarrow M}{p}_j$  are the prediction location offsets and modulation type modulation type confidence of  $\tilde{w}_j$ , respectively.

(2) *Loss Function.* The overall signal detection and identification loss ( $L_{\text{det}}$ ) is a weighted sum of the frequency location loss ( $L_{\text{loc}}$ ) and modulation type confidence loss ( $L_{\text{conf}}$ ):

$$L_{\text{det}}(G, P_{\mathcal{W}}) = L_{\text{loc}}(G, P_{\text{mp}}) + \lambda_1 L_{\text{conf}}(G, P_{\mathcal{W}}), \quad (12)$$

where  $\lambda_1$  is the balance weight term. The frequency location loss is a smooth  $L_1$  loss [24] between the ground truth location ( $\overset{\rightarrow L}{g}_i$ ) and prediction location ( $\overset{\rightarrow L}{p}_i$ ) of the matched preset

windows.

$$L_{\text{loc}}(G, P_{\text{mp}}) = \frac{1}{N_{\text{mp}}} \sum_{w_i \in \mathcal{W}_{\text{mp}}} \sum_{h \in \{a, b\}} \text{Smooth}_{L_1}(\hat{g}_i^h - p_i^h),$$

$$\hat{g}_i^o = \frac{(g_i^o - o_i)}{b_i, \hat{g}_i^b} = \log \frac{g_i^b}{b_i},$$
(13)

where  $N_{\text{mp}}$  is the number of matched preset windows and  $\hat{g}_i^o$  and  $\hat{g}_i^b$  are the normalized ground truth center offset and bandwidth offset between  $w_i$  and the corresponding transmitted signal, respectively. If there are no transmitted signals in the captured signal,  $L_{\text{loc}} = 0$ . With this frequency location loss, the MSDIN learns to refine the center and bandwidth of matched preset windows to better match the transmitted signals to obtain an accurate prediction location.

The modulation type confidence loss is a cross-entropy loss between the prediction confidence ( $\vec{p}_i^M$  and  $\vec{p}_j^M$ ) and ground truth modulation type ( $\vec{g}_i^M$ ) of all preset windows.

$$L_{\text{conf}}(G, P_{\mathcal{W}}) = -\frac{1}{N_{\text{mp}}} \sum_{w_i \in \mathcal{W}_{\text{mp}}} \sum_{m=1}^M g_i^m \log(p_i^m) - \frac{1}{N_{\text{ump}}} \sum_{\tilde{w}_j \in \mathcal{W}_{\text{ump}}} \log(p_j^0),$$
(14)

where  $N_{\text{ump}}$  is the number of unmatched preset windows,  $g_i^m$  is the  $m$ -th item of  $\vec{g}_i^M$ , and  $p_i^m$  is the  $m$ -th item of  $\vec{p}_i^M$ . Specifically,  $g_i^m$  and  $p_i^m$  present the ground truth and prediction probability of modulation type  $m$  of  $w_i$ , respectively, and  $p_j^0$  is the prediction probability of background (noise). With this modulation type confidence loss, MSDIN learns to identify the modulation type of transmitted signals.

The training process also involves other training strategies of DL-based object detection, such as hard-negative mining, data augmentation, and data balance. We use the Adam optimization algorithm to perform training with an initial learning rate of  $10^{-3}$  in our experiments.

For readers interested in the mathematical justification or derivation of loss function in expressions (12)–(14), please refer to [25] and references therein.

**3.2. Signal Detection and Identification in Unknown Environment.** In this subsection, we present a multilabel alignment adversarial network (MAAN) for HF signal detection and identification in an unknown environment. It is assumed that the source domain signals and the target domain signals are gathered from different environments. The source domain signals are fully labeled, while the target domain signals are entirely unlabeled. Let  $X^{\mathcal{S}} = \{I_i^{\mathcal{S}}, G_i^{\mathcal{S}}\}_{i=1}^{n^{\mathcal{S}}}$  denote the set of labeled data in the source domain, where  $I_i^{\mathcal{S}}$  and  $G_i^{\mathcal{S}}$  are the feature representation and ground truth of the  $i$ -th source domain signal, respectively. Let  $X^{\mathcal{T}} =$

$\{I_j^{\mathcal{T}}\}_{j=1}^{n^{\mathcal{T}}}$  denote the set of unlabeled data in target domain, where  $I_j^{\mathcal{T}}$  is the feature representation of the  $j$ -th target domain signal.

The main idea behind MAAN is to utilize domain adaptation and signal identification as auxiliary tasks to perform conditional adversarial cross-domain feature alignment and prediction consistency regularization for signal detection and identification in an unknown environment [26]. As shown in Figure 6(a), MAAN utilizes MSDIN as the basic signal detector and adds a domain discriminator and a multilabel learner for conditional adversarial training and multilabel learner training. The details of the MAAN will be described in the following.

**3.2.1. Domain Discriminator.** The popular generative adversarial network (GAN) [27] has shown that two domain datasets with different distributions can be aligned by using a domain discriminator to play a minimax two-player game. Therefore, we utilize the domain discriminator to perform feature alignment between source domain features ( $F_k^{\mathcal{S}}$ ,  $k = 1, \dots, M_f$ ) and target domain features ( $F_k^{\mathcal{T}}$ ,  $k = 1, \dots, M_f$ ). The domain discriminator predicts the domain of each input feature, with class “1” indicating the source domain and class “0” indicating the target domain. The domain discriminator consists of a convolutional layer and a domain classifier, where the domain classifier is a fully connected layer (FC) as shown in Figure 6(b). Specifically, the domain discriminator utilizes the source domain feature ( $F_k^{\mathcal{S}}$ ) and target domain features ( $F_k^{\mathcal{T}}$ ) as input to predict the domain probabilities ( $d_k^{\mathcal{S}}$  and  $d_k^{\mathcal{T}}$ ). For the domain discriminator training, we adopt a focal loss [28], which uses the prediction confidence deficiency score to weight each instance to give more weights to hard-to-classify examples. As mentioned in [28], the domain discriminator can be trained by optimizing the following equations:

$$\min_{\mathbf{E}} \max_{\mathbf{D}} L_{\text{adv}} = -\frac{1}{2(L_{\text{adv}}^{\mathcal{S}} + L_{\text{adv}}^{\mathcal{T}})},$$

$$L_{\text{adv}}^{\mathcal{S}} = -\sum_{k=1}^{M_f} (1 - (d_k^{\mathcal{S}})^{\gamma}) \log(d_k^{\mathcal{S}}), L_{\text{adv}}^{\mathcal{T}} = -\sum_{k=1}^{M_f} (d_k^{\mathcal{T}})^{\gamma} \log(1 - d_k^{\mathcal{T}}),$$
(15)

where  $\mathbf{E}$  and  $\mathbf{D}$  represent the parameters of the domain discriminator and feature extractor, respectively;  $d_k^{\mathcal{S}}$  and  $d_k^{\mathcal{T}}$  are the domain prediction probabilities of source domain feature  $F_k^{\mathcal{S}}$  and target domain feature  $F_k^{\mathcal{T}}$ , respectively; and  $\gamma$  is the regulatory factor of focus loss. If  $\gamma = 1$ , focal loss degenerates into cross-entropy loss. With this adversary loss, the domain discriminator  $\mathbf{D}$  aims to maximally separate multiscale features, while the feature extractor  $\mathbf{E}$  attempts to confuse the domain discriminator  $\mathbf{D}$ . As a result, the multiscale features of the two domains are gradually indistinguishable. The domain discriminator is expected to bridge the domain distribution gaps and improve the adaptation of the target domain.

**3.2.2. Multilabel Learner.** The signal detection and identification task include signal detection and corresponding modulation type identification, which is much more difficult than the signal identification task. We find that the features with excellent signal identification performance are also informative for signal detection and identification. Therefore, we use the signal identification task as an auxiliary task for signal detection and identification and add a multilabel learner to learn this task. The multilabel learner consists of a convolutional layer and a multilabel classifier, where multilabel classifier is a fully connected layer (FC) as shown in Figure 6(c). The multilabel learner utilizes the source domain feature ( $F_k^\delta$ ) to predict the signal probability ( $\vec{u}_k$ ). The signal identification label ( $\vec{y}^\delta$ ) can be obtained from the ground truth label ( $\vec{g}_i^M$ ) as

$$y^m = \max_{w_i \in \mathcal{M}} g_i^m, m = 1, \dots, M, \quad (16)$$

where  $y^m$  is the  $m$ -th item of  $\vec{y}^\delta$ . Specifically,  $y^m$  represents the presence of modulation type  $m$ . For multilabel learner training, we adopt the cross-entropy loss for multilabel classification. The multilabel learner loss,  $L_{ML}$ , is

$$L_{ML} = - \sum_{k=1}^{M_f} \left[ \vec{y}^\delta \left( \log \left( \vec{u}_k^\delta \right) \right)^\dagger + \left( \vec{1} - \vec{y}^\delta \right) \left( \log \left( \vec{1} - \vec{u}_k^\delta \right) \right)^\dagger \right], \quad (17)$$

where  $\vec{u}_k^\delta$  is the signal prediction probability vector of multilabel learner of the source domain feature  $F_k^\delta$ , “ $\dagger$ ” denotes the vector transposition, and  $\vec{1}$  is an all-one vector.

**3.2.3. Consistency Regularization.** We find that the prediction location errors in the signal detector may be accumulated to signal identification errors so that multilabel learner can produce a more accurate signal identification prediction in an unknown environment. Based on this observation, we propose a prediction consistency regularization mechanism between the signal identification prediction probabilities of multilabel learner and signal detector. The signal prediction probability vector of the signal detector ( $\vec{q}$ ) can be obtained from the prediction results of all preset windows ( $P_{\mathcal{W}}$ ) as

$$q^m = \max_{w_i \in \mathcal{W}} p_i^m, m = 1, \dots, M, \quad (18)$$

where  $q^m$  is the  $m$ -th item of  $\vec{q}$ . Specifically,  $q^m$  represents the signal prediction probability of modulation type  $m$  of the signal detector. For consistency regularization training, we adopt the Kullback-Leibler (KL) divergence to enforce the consistency between the predictions produced by the signal detector and multilabel learner. The consistency regularization loss,  $L_{KL}$ , is

$$L_{KL} = L_{KL}^\delta + L_{KL}^\mathcal{T},$$

$$L_{KL}^\delta = \frac{1}{2} \sum_{k=1}^{M_f} \left( \text{KL} \left( \vec{u}_k^\delta, \vec{q}^\delta \right) + \text{KL} \left( \vec{q}^\delta, \vec{u}_k^\delta \right) \right), \quad (19)$$

$$L_{KL}^\mathcal{T} = \frac{1}{2} \sum_{k=1}^{M_f} \left( \text{KL} \left( \vec{u}_k^\mathcal{T}, \vec{q}^\mathcal{T} \right) + \text{KL} \left( \vec{q}^\mathcal{T}, \vec{u}_k^\mathcal{T} \right) \right),$$

where  $\vec{u}_k^\delta$  and  $\vec{u}_k^\mathcal{T}$  are the signal prediction probability vectors of the multilabel learner of source domain signals and target domain signal, respectively, and  $\text{KL}(\cdot, \cdot)$  is the function of KL divergence. With this consistency regularization loss, we expect the multilabel learner to assist the signal detector in achieving a better signal identification performance in an unknown environment through unified mutual learning.

**3.2.4. Overall Loss.** The overall loss of the MAAN is a sum of the detection loss ( $L_{\text{det}}$ ), conditional adversarial loss ( $L_{\text{adv}}$ ), multilabel learning loss ( $L_{ML}$ ), and consistency regularization loss ( $L_{ML}$ ):

$$L_{\text{all}} = L_{\text{det}} + \mu_1 L_{\text{adv}} + \mu_2 L_{ML} + \mu_3 L_{KL}, \quad (20)$$

$$\min_{\mathbf{L}, \mathbf{Q}, \mathbf{E}} \max_{\mathbf{D}} L_{\text{all}}$$

where  $\mathbf{L}$ ,  $\mathbf{Q}$ , and  $\mathbf{E}$  represent the parameters of the multilabel learner, predictor, and feature extractor, respectively, and  $\mu_1$ ,  $\mu_2$ , and  $\mu_3$  are the trade-off weights that balance the multiple loss terms. We use the SGD optimization algorithm to perform training with an initial learning rate of  $10^{-4}$ . The initial parameters of the MAAN are obtained from the corresponding MSDIN. In addition, the optimization of minmax operation is achieved by the gradient inversion layer (GRL) as described in [29].

## 4. Experiments

In this section, we show the experimental results of the proposed methods and take some existing methods as baselines for comparison to demonstrate the robustness and effectiveness of our methods. The details of the datasets utilized in the experiments are described in Subsection 4.1, the baselines are introduced in Subsection 4.2, the comparison metrics are given in Subsection 4.3, and the experimental results are presented in Subsection 4.4. All of the experiments are conducted on GTX 1080ti. The source code for experiments is freely available from Github at <https://github.com/huanglin123136/Real-Time-HF-signal-Detection-and-Identification-in-Known-and-Unknown-HF-Channels>.

**4.1. Dataset.** In this section, we generate several different datasets, including the AWGN dataset, Watterson dataset, Rayleigh dataset, time dataset, and MDFS (maximum Doppler frequency shifts) dataset. The signals in these datasets are transmitted through the AWGN channels, Watterson channels, and Rayleigh fading channels, respectively. The details of these datasets are shown in Table 1, where  $B$  is the



bandwidth of the signals,  $T$  is duration time of the signals, MDFS is the maximum Doppler frequency shifts of the transmitted channels, SNR is the range of signal noise ratio of the training samples, and training sample is the number of the signals in the training dataset.

**4.2. Baseline for Comparison.** We set the following typical methods used in the previous works as the baselines for comparison.

- (1) SSD [11]: SSD is a representative one-stage object detection method that uses anchor boxes to predict objects. The backbone network of SSD used in our experiments is VGG-16
- (2) CenterNet [12]: CenterNet is a representative anchor-free object detection method that regards each object as a point of the bounding box for detection. The backbone network of CenterNet used in our experiments is ResNet-50
- (3) SDIN: SDIN is a single-resolution input version of MSDIN. The network structure of the SDIN is similar to that of the MSDIN, but only one path of the aggregation module is reserved

**4.3. Metrics.** In this paper, we compare different methods in three aspects: detection precision, running speed, and model size. The mean average precision (mAP) [30] is used to evaluate the detection precision of different methods, which is a widely used precision metric in objection detection tasks. AP is a comprehensive metric of the prediction precision (Precision) and recall (Recall), as shown in the following equations:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (21)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (22)$$

$$\text{AP}_m = \int_0^1 P_m(r) dr, \quad (23)$$

where TP, FP, and FN are the number of correctly detected transmitted signals, missing alarmed transmitted signals and missing detected transmitted signals, respectively,  $P_m(r)$  is the precision-recall curve of modulation type  $m$ , and  $\text{AP}_m$  is the area size of  $P_m(r)$ . The mAP is the mean of  $\text{AP}_m$  of different modulation types

$$\text{mAP} = \frac{\sum_{m=1}^M \text{AP}_m}{M}, \quad (24)$$

where  $M$  is the number of modulation types. In addition, the signal detection and identification task usually adopt the false alarm rate (FAR) and missing alarm rate (MAR) to evaluate the performance

$$\text{FAR} = 1 - \text{Recall}, \quad (25)$$

$$\text{MAR} = 1 - \text{Precision}, \quad (26)$$

The processing bandwidth per second (BPS) is used to evaluate the running speed of different methods.

$$\text{BPS} = \frac{B_{\text{process}}}{T_{\text{process}}}, \quad (27)$$

where  $T_{\text{process}}$  and  $B_{\text{process}}$  are the processing time and processing bandwidth of the network, respectively. The unit for running speed is MHz/s.

The memory usage (MU) of model parameters is used to evaluate the model size of different methods. The unit for model size is MB.

#### 4.4. Experimental Results and Analysis

**4.4.1. Performance Comparison with Baseline Methods.** In this subsection, we compare the performance of MSDIN with ones of the baseline methods on the different datasets in known environment. Figure 7(a) shows the mAP vs SNR curves of different methods on the AWGN dataset. It can be seen that MSDIN achieves the best performance at all SNRs, and SDIN is better than SSD and CenterNet. Figures 7(b)–7(d) show the confusion matrixes of MSDIN on the AWGN dataset with different SNRs, which indicate that there are more errors happens in low SNR and most of errors are false alarms and missing alarms, rather than the confusion between different modulation types. Similarly, Figure 8 shows the performance of different methods on the Watterson dataset and Rayleigh dataset, respectively. It is again seen that the MSDIN achieves the best performance at all SNRs on these datasets (see Figures S1-S3 in the Supplementary Material for comprehensive performance comparisons at different SNRs).

Figure 9 shows the running speed and model size comparison of different methods. It is obvious that MSDIN and SDIN are significantly faster and smaller than SSD and CenterNet. Compared with SDIN, MSDIN achieves better performance, especially at low SNR, with little cost increment of running speed and model size. These experimental results indicate that our proposed network is not only superior to the existing DL-based networks in running speed and accurate, but also robust to different HF environments.

**4.4.2. Performance Comparison with Different Duration Times.** In this subsection, we present the experimental results of MSDIN on the time dataset. Figure 10(a) shows the mAP vs SNR curves of MSDIN with different duration times on the time dataset. It is shown that the long duration time can achieve better signal detection performance. Compared with  $T = 800$  ms, the performance when  $T = 1600$  ms has an approximately 2 dB gain. Figures 10(b)–10(d) show the confusion matrixes of MSDIN with different duration times. It can be seen that the shorter duration time brings more false alarms and missing alarms. Figure 11 presents the running speed and model size of the MSDIN with different duration times. It is obvious that the running speed of the proposed network is increasing with the duration time and that the longer duration time can obtain better performance. Therefore, the duration time is a key parameter

affecting the trade-off between real time and accuracy and should be selected according to the demands of real applications.

**4.4.3. Performance Comparison with Different MDFS.** In this subsection, we present the experimental results of the MSDIN on the MDFS dataset with Rayleigh Fading channels. Figure 12(a) shows the mAP vs SNR curves of MSDIN with different MDFS. It is indicated that the Rayleigh Fading channels with lower MDFS can obtain better detection performance. Figures 12(b)–12(d) show the confusion matrixes of MSDIN with different MDFS, where it can be seen that Rayleigh Fading channels with larger MDFS may suffer more modulation type confusion errors. The experimental results show that the signals in the Rayleigh Fading channels with larger Doppler frequency shifts are difficult to detect and identify. The running speed and model size of MSDIN with different MDFS are same as those in Figure 9.

**4.4.4. Performance Comparison in Unknown Environments.** In this subsection, we present the experimental results of different methods in unknown environments, where the source domain signals and target domain signals are gathered from different environments (channels). Specially, we consider two different situations, case 1, the source domain signals are gathered from the AWGN channels, and the target domain signals are gathered from the Watterson channels, and case 2, the source domain signals are gathered from the Watterson channels, and the target domain signals are gathered from the AWGN channels.

Figure 13(a) shows the mAP vs SNR curves of different methods in case 1. The “Benchmark” presents the mAP vs SNR curve of MSDIN in a known environment, i.e., source domain signals and target domain signals are both from the Watterson dataset. It is shown that the mAP performance of all methods is degraded over all SNR, and the MAAN achieves better performance than other methods. Figures 13(b)–13(d) show the confusion matrixes of different methods in case 1, where it can be seen that the MSDIN suffers more modulation type confusion errors than the benchmark, and MAAN can correct most of them. However, the MAAN is not very effective for errors of false alarms and missing alarms, which indicates that there may still be some room for improvement of our method, especially at low SNR. Similarly, Figure 14 shows the performance of different methods in case 2, where it is shown that MAAN achieves better performance than MSDIN and is very close to the benchmark. These experimental results show that the MAAN can achieve obvious performance improvement in unknown environments, and it is robust for the different situations (see Figures S4–S7 in the Supplementary Material for comprehensive performance comparisons in unknown environments).

In addition, it should be noticed that the domain discriminator and multilabel learner of the MAAN only work in the training phase to assist the signal detector in bridging the domain shift between the source domain signals and target domain signals. Therefore, the running speed and model

size of the MAAN are the same as MSDIN as shown in Figure 9.

## 5. Conclusions

In this paper, we analyze the characteristics of different transmitted signals and set forth the validity to utilize multi-resolution time-frequency spectra for HF signal detection and identification. Then, we design a novel multiresolution signal detection and identification network for real-time HF signal detection and identification. Finally, we propose a domain adaptation method to adapt the proposed network to unknown environments.

We have demonstrated, by a series of simulation experiments, the effectiveness of our works on the different transmitted environmental conditions (channels), SNRs, duration times, and maximum Doppler frequency shifts. These experiment conditions and parameters are typical enough for most of HF channels. Specially, the experimental results show that the running speed and accuracy of our proposed network is superior to ones of the existing DL-based networks in different HF environments, and the proposed domain adaptation method can achieve obvious performance improvement in unknown environments.

In future researches, we will enrich the datasets of different environments, add more modulation types, explore more comprehensive features, and further improve the performance of real-time signal detection and identification in known and unknown environments.

## Data Availability

The datasets used in this work are available from the corresponding author on reasonable request.

## Conflicts of Interest

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

## Acknowledgments

The authors acknowledge the support from Dr. Chongfa Wang and Weixing Jie. This work was supported in part by the National Natural Science Foundation of China under grant 61971391.

## Supplementary Materials

Figure S1: The confusion matrixes of MSDIN on the AWGN dataset at different SNRs. Figure S2: The confusion matrixes of MSDIN on the Watterson dataset at different SNRs. Figure S3: The confusion matrixes of MSDIN on the Rayleigh dataset at different SNRs. Figure S4: The confusion matrixes of MSDIN training on the AWGN dataset and testing on the Watterson dataset at different SNRs. Figure S5: The confusion matrixes of MAAN at different SNRs, where the source domain is the AWGN dataset and target domain is the Watterson dataset. Figure S6: The confusion matrixes of MSDIN training on the Watterson dataset and testing on the AWGN

dataset at different SNRs. Figure S7: The confusion matrixes of MAAN at different SNRs, where the source domain is the Watterson dataset and the target domain is the AWGN dataset. (*Supplementary Materials*)

## References

- [1] J. Wang, G. Ding, and H. Wang, "HF communications: past, present, and future," *China Communications*, vol. 15, no. 9, pp. 1–9, 2018.
- [2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.
- [3] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.
- [4] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *IET Communications*, vol. 1, no. 2, pp. 137–156, 2007.
- [5] D. Ke, Z. Huang, X. Wang, and X. Li, "Blind detection techniques for non-cooperative communication signals based on deep learning," *IEEE Access*, vol. 7, pp. 89218–89225, 2019.
- [6] L. Depeng, X. Jianhua, X. Changbo, and Y. Mingyang, "A single frame spectrum signal detection approach based on deep learning," in *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pp. 726–731, Chongqing, China, 2020.
- [7] A. K. Nandi and E. E. Azzouz, "Algorithms for automatic modulation recognition of communication signals," *IEEE Transactions on Communications*, vol. 46, no. 4, pp. 431–436, 1998.
- [8] D. C. Chang and P. K. Shih, "Cumulants-based modulation classification technique in multipath fading channels," *IET Communications*, vol. 9, no. 6, pp. 828–835, 2015.
- [9] K. C. Ho, W. Prokopiw, and Y. T. Chan, "Modulation identification of digital signals by the wavelet transform," *IEE Proceedings Radar, Sonar and Navigation*, vol. 147, no. 4, pp. 169–176, 2000.
- [10] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proceedings of the International Conference on Engineering Applications of Neural Networks*, vol. 629, pp. 213–226, Cham, 2016.
- [11] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector," in *2016 European conference on computer vision (ECCV)*, pp. 21–37, Amsterdam, Netherlands, 2016.
- [12] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: keypoint triplets for object detection," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6568–6577, Seoul, Korea (South), 2019.
- [13] P. Naveen and P. Sivakumar, "Human emotions detection using kernel nonlinear collaborative discriminant regression classifier: human emotions detection using KNCDRC," in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 1807–1812, Trichy, India, 2021.
- [14] P. Naveen and P. Sivakumar, "Deep convolution neural network for facial expression recognition," *Journal of current Science and Technology*, vol. 11, no. 3, pp. 402–410, 2021.
- [15] Y. H. Tsai, K. Sohn, S. Schuler, and M. Chandraker, "Domain adaptation for structured output via discriminative patch representations," in *2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1456–1465, Long Beach, CA, USA, 2019.
- [16] Z. Xu, I. Petrunin, A. Tsourdos, and S. Ayub, "Learning based spectrum hole detection for cognitive radio communication," in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*, pp. 1–7, San Diego, CA, USA, 2019.
- [17] X. Zha, H. Peng, X. Qin, G. Li, and S. Yang, "A deep learning framework for signal detection and modulation classification," *Sensors*, vol. 19, no. 18, pp. 4042–4043, 2019.
- [18] W. Li, K. Wang, L. You, and Z. Huang, "A new deep learning framework for HF signal detection in wideband spectrogram," *IEEE Signal Processing Letters*, vol. 29, pp. 1342–1346, 2022.
- [19] C. Watterson, J. Juroshek, and W. Bensema, "Experimental confirmation of an hf channel model," *IEEE Transactions on Communication Technology*, vol. 18, no. 6, pp. 792–803, 1970.
- [20] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [22] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2117–2125, Las Vegas, NV, USA, 2017.
- [23] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3, pp. 850–855, Hong Kong, China, 2006.
- [24] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago, Chile, 2015.
- [25] L. Deng, *Deep Learning: Methods and Applications*, Now Publishers Inc, 2014.
- [26] Z. Zhao, Y. Guo, H. Shen, and J. Ye, "Adaptive object detection with dual multi-label prediction," in *2020 European Conference on Computer Vision (ECCV)*, pp. 54–69, Cham, 2020.
- [27] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," *2014 Advances in neural information processing systems (NIPS)*, pp. 2672–2680, 2014.
- [28] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.
- [29] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *2015 Proceedings of Machine Learning Research (PMLR)*, pp. 1180–1189, Lille, France, 2015.
- [30] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: a retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, 2015.