




## Article

# The Research of Intra-Pulse Modulated Signal Recognition of Radar Emitter under Few-Shot Learning Condition Based on Multimodal Fusion

Yunhao Liu <sup>1</sup>, Sicun Han <sup>2</sup>, Chengjun Guo <sup>3,\*</sup>, Jiangyan Chen <sup>1</sup> and Qing Zhao <sup>4</sup>

<sup>1</sup> Research Institute of Electronic Science and Technology, University of Electronic Science and Technology of China, Chengdu 611731, China; 202222230106@std.uestc.edu.cn (Y.L.); 202222230124@std.uestc.edu.cn (J.C.)

<sup>2</sup> Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180, USA; hans9@rpi.edu

<sup>3</sup> School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

<sup>4</sup> Southwest China Institute of Electronic Technology, Chengdu 610036, China; zhaoqingwork@outlook.com

\* Correspondence: johnsonguo@uestc.edu.cn; Tel.: +86-18908070306

**Abstract:** Radar radiation source recognition is critical for the reliable operation of radar communication systems. However, in increasingly complex electromagnetic environments, traditional identification methods face significant limitations. These methods often struggle with high noise levels and diverse modulation types, making it difficult to maintain accuracy, especially when the Signal-to-Noise Ratio (SNR) is low or the available training data are limited. These difficulties are further intensified by the necessity to generalize in environments characterized by a substantial quantity of noisy, low-quality signal samples while being constrained by a limited number of desirable high-quality training samples. To more effectively address these issues, this paper proposes a novel approach utilizing Model-Agnostic Meta-Learning (MAML) to enhance model adaptability in few-shot learning scenarios, allowing the model to quickly learn with limited data and optimize parameters effectively. Furthermore, a multimodal fusion neural network, DCFANet, is designed, incorporating residual blocks, squeeze and excitation blocks, and a multi-scale CNN, to fuse I/Q waveform data and time–frequency image data for more comprehensive feature extraction. Our model enables more robust signal recognition, even when the signal quality is severely degraded by noise or when only a few examples of a signal type are available. Testing on 13 intra-pulse modulated signals in an Additive White Gaussian Noise (AWGN) environment across SNRs ranging from  $-20$  to  $10$  dB demonstrated the approach’s effectiveness. Particularly, under a  $5 - way 5 - shot$  setting, the model achieves high classification accuracy even at  $-10$  dB SNR. Our research underscores the model’s ability to address the key challenges of radar emitter signal recognition in low-SNR and data-scarce conditions, demonstrating its strong adaptability and effectiveness in complex, real-world electromagnetic environments.

**Keywords:** radar emitter; few-shot learning; multimodal fusion; intra-pulse modulation



**Citation:** Liu, Y.; Han, S.; Guo, C.; Chen, J.; Zhao, Q. The Research of Intra-Pulse Modulated Signal Recognition of Radar Emitter under Few-Shot Learning Condition Based on Multimodal Fusion. *Electronics* **2024**, *13*, 4045. <https://doi.org/10.3390/electronics13204045>

Academic Editor: Silvia Liberata Ullò

Received: 23 September 2024

Revised: 12 October 2024

Accepted: 13 October 2024

Published: 14 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Radar radiant source identification is a key technology in radar communication systems, and the classification of radiant source intra-pulse modulation can provide radar systems with more powerful analytical capabilities. This has already become an indispensable element in the electronic information system. The traditional intra-pulse modulation classification method requires manual signal feature extraction. For example, Han et al. [1] proposed a new modulation recognition algorithm for digital communication signals based on higher-order cumulants (HOCs) and Support Vector Machine (SVM) to realize the automatic classification of modulation signals. Nuaimi et al. [2] applied signal statistical characteristics and maximum likelihood estimation to the automatic classification of modulated and identified signals; Park et al. [3] used the wavelet transform characteristic and

SVM to automatically classify and recognize eight different digitally modulated signals. These traditional methods not only require a lot of prior knowledge but also spend a lot of manpower and material resources to learn. In addition, the obtained signal has very poor performance in the case of low SNR, making it impractical to reach the index for actual application.

In recent years, with the development of technology, deep learning algorithms have been widely used in the radar emitter signal recognition field. Convolutional Neural Network (CNN) is one of the most classic models in deep learning. With the idea of weight sharing and local perception, CNN greatly reduced the complexity and computing cost of the network. In addition, it could also effectively process multi-dimensional images and multi-dimensional sequences, especially in two-dimensional image processing. In the field of radar, CNN has been widely used in classification and recognition problems, such as high-resolution synthetic aperture radar image classification [4], radar one-dimensional range target recognition [5] and radar gesture recognition [6]. For the intra-pulse modulated signal of the radar emitter, existing researchers mainly uses CNN to extract the signal features. Therefore, it is possible to directly transform the radar emitter signal and use the two-dimensional image extracted after signal transformation as the input of the CNN network, which will automatically extract the deep hidden features of the radar radiation source signal and classify and recognize it. With CNN's excellent feature extraction capability and huge data processing capability, Zheng et al. [7] designed an image feature deep learning network based on AlexNet [8], which used large sample full pulse data to form an image feature representation of intra-pulse parameter changes, and it revealed the hidden waveform design mechanism of radar emitter from a macroperspective to realize radar emitter recognition. Since then, in addition to using CNN to extract the two-dimensional features of the signal [9–12], some scholars have also used CNN to extract the features of the signal sequence of the emitter for classification and recognition [13,14]. Especially, Niu et al. [15] proposed a one-dimensional convolutional neural network to learn and recognize the effective features of the signal sequence of the radar emitter. In addition to the feature extraction method of CNN, Qu et al. [16] carried out research on pulse repetition interval modulation recognition combining a Recurrent Neural Network (RNN) and Long Short-Term Memory Neural Network (LSTM).

With the development of jamming and anti-jamming technologies, the electromagnetic environment has become increasingly complex and variable, making the acquisition of radar emitter signals more challenging, particularly in obtaining high-quality and complete emitter signals that can be used as training data [17]. Therefore, researching the intelligent recognition of intra-pulse modulation of radar emitters under small-sample conditions holds significant practical and applied value. In recent years, some scholars have studied the recognition of radar emitters in small-sample scenarios. For example, Ding et al. [18] proposed incorporating Generative Adversarial Networks (GANs) into CNNs, either by adding a GAN module after the Sparse Autoencoder (SAE) module or after the feature enhancement module to enhance features, achieving emitter recognition with few-shot learning. Li et al. [19] utilized the learning capabilities of deep convolutional networks and GANs to generate time–frequency images with noise reduction and data enhancement effects on the basis of the original training set. These images were then used as a supplementary training set, which was later used to assist the Visual Geometry Group (VGG) in recognizing radar emitters in the original training set. Wang [20], addressing the issue of data imbalance, built a model for expanding few-shot radar emitter signals using the commonly used SMOTE and Borderline SMOTE algorithms, which is followed by signal classification using decision trees, random forests, and KNN classifiers.

At present, most research on radar emitter recognition under few-shot learning conditions focuses on methods such as expanding the target data samples or enhancing their features, as mentioned above [21]. While these approaches are feasible in scenarios with a high SNR, they encounter significant limitations in low-SNR environments. In such cases, noise tends to overwhelm the actual radar emitter signals, resulting in insufficient diversity

among the expanded samples. Additionally, generative algorithms like GANs lack strong robustness and may even lead to model instability or failure [22], especially in high-SNR scenarios. Moreover, it can be challenging to quantify when high-quality samples are generated or whether the model has converged [23].

To address these shortcomings, this article proposes several significant contributions to the recognition of intra-pulse modulated signals of radar emitters under few-shot learning conditions. Inspired by the work in [24], building a model that can quickly adapt to environments with limited samples is more robust and generalized than simply training a model using data augmentation. The Model-Agnostic Meta-Learning (MAML) method, which we are going to introduce, can achieve this goal. Despite the advantages of the MAML optimization method in developing a model that can rapidly adapt under few-shot conditions, recognizing radiation source signals at low SNRs continues to pose a substantial challenge. Acknowledging that signals exhibit diverse dimensional features, we conduct an analysis from various modalities and integrate these insights to create a multimodal fusion model, further enhancing the accuracy of signal recognition.

In the following text, we will first conduct a mechanism analysis and modeling of the radar signal model used in this article, and then we provide a detailed explanation of the signal preprocessing methods, such as Short-Time Fourier Transform (STFT). Subsequently, we are about to introduce the MAML method to optimize model parameters, enhancing adaptability to few-shot conditions and improving the overall training process. Apart from this, we design a multimodal fusion neural network, DCFANet, which integrates information from both I/Q waveform and time–frequency image domains. This integration facilitates comprehensive feature extraction and improves recognition capabilities. Additionally, we implement residual blocks and squeeze and excitation blocks within the architecture to enhance feature representation and channel attention, maintaining the uniqueness of each modality. Furthermore, we introduce a joint loss function that supports collaborative training across multiple modalities, contributing to the robustness of the recognition process. Lastly, we will detail the experimental environment and settings, conduct a comparative analysis of various methods and their associated variables through ablation experiments, and present the significant results obtained from the experiments.

Our experiments generate 13 distinct intra-pulse modulated signals under the AWGN environment, and the trained model is evaluated against existing advanced network models. The results demonstrate that our multimodal fusion model with MAML achieves high classification accuracy. Moreover, under challenging conditions involving doppler shift and multipath fading, the model continues to perform as well as we expect.

Based on experimental studies, we summarize the following key contributions and improvements. (1) The Base-CNN-MAML model constructed using the MAML method achieved an approximately 30% higher accuracy under few-shot conditions compared to traditional algorithms' Base-CNN, and it performs on par with ResNet under traditional algorithms, verifying the effectiveness of MAML. (2) We also investigated the impact of sample size, further confirming the feasibility of the MAML method under few-shot conditions and determining the optimal sample size parameters for this study. (3) Both the proposed SE-ResNet-MAML and 1D-MSCNN-MAML models demonstrated superior recognition performance for radar emitter signals in their respective modalities under few-shot conditions. Particularly, our proposed multimodal fusion model DCFANet, when combined with the MAML method, exhibited superior noise resistance in the sequential modality and strong spatial feature capture and representation capabilities in the image modality. Its recognition accuracy under both low and high SNR conditions surpassed the other two unimodal models, achieving an overall recognition accuracy of 89.3% across global SNR levels. (4) Finally, we further explored DCFANet-MAML through visualization analysis, generating a confusion matrix and projecting high-dimensional vectors to low-dimensional space before classification, which strongly confirmed the effectiveness of the proposed method.

## 2. Radar Signal Model and Signal Preprocessing

We are going to analyze and model the mechanism of the radar signal in this section and present the preprocessing operations performed on the signal from different dimensions.

In order to simulate the radar radiation source signals more closely to the real environment, we will consider the noise and echo time delay to fit our study. On this basis, multipath fading, doppler shift and time delay are added to the radar signals under a multipath effect to correspond to the few-shot scenarios caused by complex electromagnetic environments.

In order to achieve the purpose of analyzing and processing the signal from different dimensions, we implement time–frequency transformation, which is most widely used in signal processing, to transform the radar signal into a two-dimensional graph domain for analysis and processing; besides, vector normalization is used for the time domain waveform for analysis and processing.

### 2.1. Radar Signal Model

In the condition of white Gaussian noise, the signal model received by the radar receiver can be simply expressed as

$$x(t) = s(t) + n(t) \quad (1)$$

In Equation (1),  $x(t)$  represents the radar signal received,  $s(t)$  represents the radar modulated signal, and  $n(t)$  represents the white Gaussian noise.

Radar modulated signals can also be expressed as

$$s(t) = Ae^{j[2\pi f_0 t + \varphi(t) + \varphi_0]} \quad (2)$$

In Equation (2),  $A$  represents the amplitude of the radar emitter signal,  $f_0$  represents the central frequency of the emitter signal,  $\varphi(t)$  represents the instantaneous phase shift of the signal, and  $\varphi_0$  represents the initial phase of the signal. The primary distinctions in the characteristics of the intra-pulse modulation information for radar emitter signals are mainly reflected in  $\varphi(t)$ . This study mainly considers 13 common types of intra-pulse modulation signals for radar emitters, including the conventional pulse signal (NS), binary phase-shift keying (BPSK) signal, which uses Barker codes, multi-phase shift keying (MPSK) signals, including Frank, P1, P2, P3, and P4 codes, a linear frequency modulation (LFM) signal, a frequency-coded signal (Costas), and multi-time coding signals, including T1, T2, T3, and T4 codes.

However, during signal propagation, environmental factors often cause refraction, reflection, and other phenomena, which in turn lead to the occurrence of multipath effects [25]. Considering a real-world situation, the received signal model affected by multipath effects can be expressed as

$$X = AS + N \quad (3)$$

where  $A = [A_0, A_1, A_2, \dots, A_{n-1}]$ ,  $A_k$  represents the amplitude fading caused by the signal traveling through path  $k$  in multipath propagation;  $S = [s_0, s_1, s_2, \dots, s_{n-1}]$ , where  $s_k = s(t - d_k)$ ,  $s_k$  represents the signal on path  $k$ , and  $d_k$  represents the time delay of path  $k$ ;  $X = [x_0, x_1, x_2, \dots, x_{n-1}]$ , where  $x_k$  represents the final signal received passing through path  $k$ ;  $N = [n_0, n_1, n_2, \dots, n_{n-1}]$ , where  $n_k = n(t)$  represents the white Gaussian noise on path  $k$  referred to [26].

For specific modulation mode or the radar signal received at the specific time  $t$ , the expression is

$$x(t) = \sum_{k=0}^{n-1} A_k e^{j[2\pi f_0(t-d_k) + \varphi(t-d_k) + \varphi_0]} + n(t) \quad (4)$$

## 2.2. Signal Time–Frequency Image Domain Analysis and Processing

Some articles presents traditional signal processing as [27] does, so that the analysis can be typically conducted in either the time domain or by transforming the signal into the frequency domain. The time domain represents the variation of a signal with respect to time, while the frequency domain characterizes the spectral components of the signal over its entire duration. However, neither of these approaches can effectively capture the relationship between the signal and frequency variation over time. Time–frequency analysis, which integrates the advantages of both time-domain and frequency-domain analysis for non-stationary signals, transforms a one-dimensional signal into a two-dimensional time–frequency domain. This approach simultaneously presents the signal’s characteristics from two perspectives, providing a clear and intuitive display of how frequency changes over time. Such visualization aids in extracting more features during signal recognition [28]. Common methods of time–frequency analysis include (1) linear time–frequency distributions, such as Short-Time Fourier Transform (STFT) and wavelet transform; and (2) quadratic time–frequency distributions, such as Wigner–Ville Distribution (WVD) and Choi–Williams Distribution (CWD).

The Wigner–Ville Distribution (WVD) is a typical nonlinear quadratic transformation, characterized by the absence of window function constraints, which grants it superior resolution. The WVD of signal  $x(t)$  is expressed as follows:

$$w_x(t, f) = \int_{-\infty}^{\infty} x(t + \tau/2)x^*(t - \tau/2)e^{-j2\pi f\tau} d\tau \quad (5)$$

After the WVD transformation, the signal exhibits strong time–frequency concentration, making it easy to distinguish single-component signals. However, for multi-component signals, the WVD transformation might introduce numerous interference terms. To address this, Cohen improved the method by incorporating different kernel functions into the analysis [29]. The resulting collection of time–frequency distributions is defined as Cohen’s class of time–frequency distributions:

$$C_x(t, w) = \int_{-\infty}^{+\infty} R(t, \tau)e^{-j\omega\tau} d\tau = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} A_x(\tau, v)\phi(\tau, v)e^{-j(vt+\omega\tau)} d\tau dv \quad (6)$$

In the function,  $A_x(\tau, v)$  is the fuzzy function of the signal  $x(t)$ , while  $\phi(\tau, v)$  represents a different filter function. The Cohen class time–frequency distribution can also be equivalent to

$$C_x(t, \omega) = \iiint x\left(u + \frac{\tau}{2}\right)x^*\left(u - \frac{\tau}{2}\right)\phi(t - u, \tau)e^{-j\omega\tau - j\omega t} dv du d\tau \quad (7)$$

A lot of researchers, such as in [30], have worked on improving the WVD by modifying Cohen’s class of time–frequency distributions through various kernel functions. These modifications aim to both effectively suppress the interference terms generated during the time–frequency transformation process and preserve the original time–frequency characteristics of the signal. Experimental results have demonstrated that the exponential kernel function provides the best suppression of interference terms. Consequently, the exponential kernel function was incorporated into Cohen’s class distribution, leading to the Choi–Williams Distribution (CWD). The CWD offers excellent time–frequency resolution and effectively eliminates interference terms. Its mathematical representation is as follows:

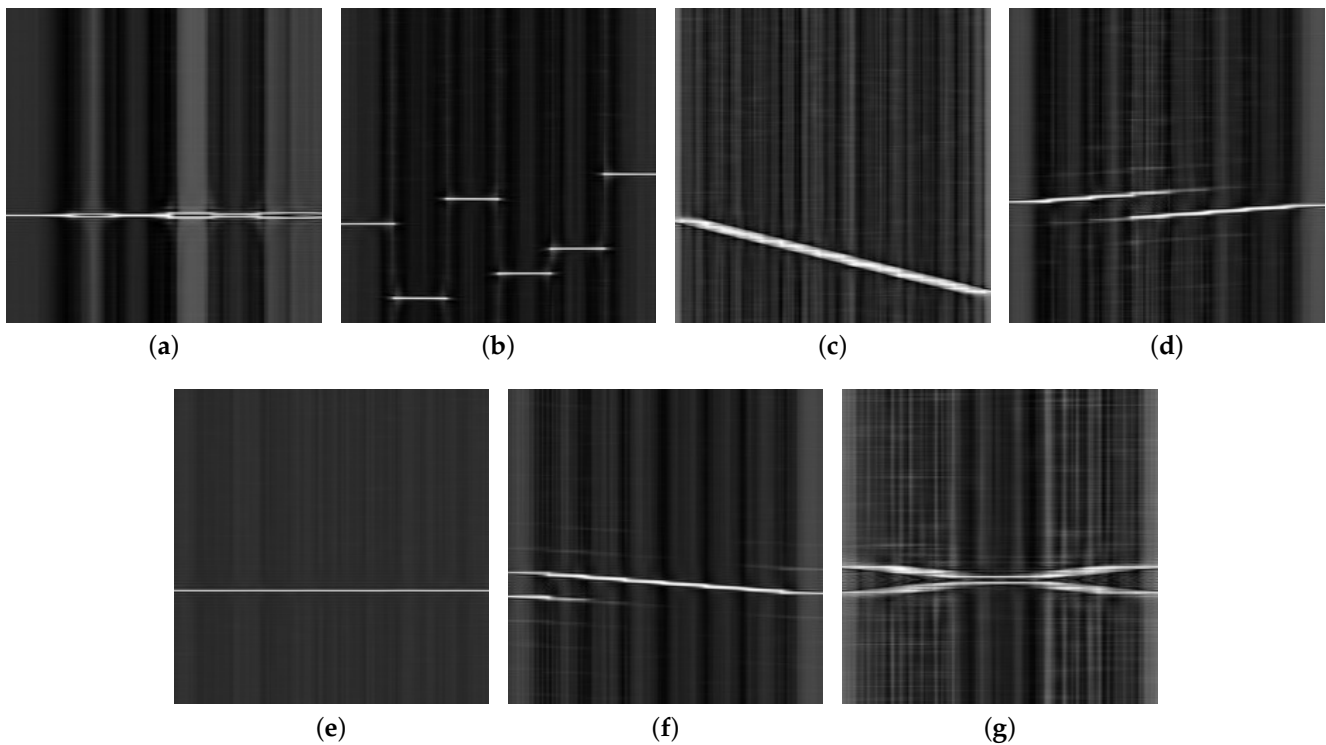
$$CWD_x(t, w) = \iiint \sqrt{\frac{\pi\sigma}{\tau^2}} x\left(u + \frac{\tau}{2}\right)x^*\left(u - \frac{\tau}{2}\right) \exp\left[-\frac{-\pi^2\sigma(t - u)}{4\tau^2}\right] e^{-j\omega\tau - j\omega t} dv du d\tau \quad (8)$$

In the function,  $\varphi(\tau, v) = e^{\frac{-v^2\tau^2}{2}}$  is the exponential kernel function introduced for CWD.  $\sigma$  is the attenuation coefficient, which normally has the range of [0.1, 10].

In this article, we used Choi–Williams Distribution (CWD) to perform time–frequency transformations on radar emitter signals to generate time–frequency images. Since the



primary features extracted from these images pertain to the edges in the time–frequency domain rather than the texture, using three-channel time–frequency images would significantly increase computational complexity. Therefore, the time–frequency images are converted into grayscale images to facilitate subsequent processing steps. (Further experiments demonstrated that the neural network does not encounter a feature bottleneck during training due to the use of grayscale images). When the SNR is 0 dB, the time–frequency distribution images of seven types of intra-pulse modulation signals were extracted, as shown in Figure 1.



**Figure 1.** Part of the time–frequency image of the pulse-modulated signal: (a) Barker; (b) Costas; (c) LFM; (d) Frank; (e) NS; (f) P2; (g) T4.

### 2.3. Analysis and Processing of Signal I/Q Waveform Domain

After sampling, the accepted radar signal  $x(t)$  becomes a discrete signal  $x[n]$ , which is expressed as the following formula.

$$x[n] = x^I[n] + j \cdot x^O[n] \tag{9}$$

where  $x^I[n] \in R$  represents the in-phase component of the signal, and  $x^O[n] \in R$  represents the orthogonal component of the signal. If the number of sampling points in a sampling period is  $N$ , the  $k_{th}$  signal data vector can be expressed as

$$x_k = [x_k[0], x_k[1], x_k[2], \dots, x_k[N - 1]]^T \tag{10}$$

The accepted  $x[n]$  is the plural discrete signal, which consists of real parts and imaginary parts. The vector of the signal's  $l_{th}$  sampling point  $x_l$  is defined as

$$x_k^{I/O}[l + 1] = [x_k^O[l + 1], x_k^I[l + 1]]^T \tag{11}$$

where  $x_k^I[l + 1] \in \text{Re}[x[n]]$ ,  $x_k^O[l + 1] \in \text{Im}[x[n]]$ . In order to improve the subsequent model training speed and prevent the gradient from disappearing or exploding, this article then

normalizes the real and imaginary parts of the complex sequence; as follows,  $\overset{\Delta}{x}_k [l + 1]$  is the processed signal vector:

$$\begin{aligned} \overset{\Delta}{x}_k^I [l + 1] &= \frac{x_k^I [l + 1] - \min(x_k^I)}{\max(x_k^I) - \min(x_k^I)} \\ \overset{\Delta}{x}_k^Q [l + 1] &= \frac{x_k^Q [l + 1] - \min(x_k^Q)}{\max(x_k^Q) - \min(x_k^Q)} \end{aligned} \quad (12)$$

### 3. Intra-Pulse Modulated Signal Recognition of Radar Emitter under Few-Shot Learning Condition Based on Multimodal Fusion

Generally speaking, the initial parameters of the traditional model are randomly initialized and need to be updated in multiple steps to achieve better results, while under few-shot conditions, the insufficient number of samples can lead to underfitting or overfitting after multiple steps. Even though, as mentioned in the introduction, some scholars have focused on using data augmentation methods such as GAN to increase the number of training samples, these autonomously generated “fake samples” are still very different from the radar source signals in the real environment. To ensure that the model maintains good generalization ability under the few-shot conditions, this paper adopts a meta-learning strategy Model-Agnostic Meta-Learning (MAML) to optimize and improve the traditional model training process. Different from the traditional optimization algorithms, the MAML approach is adopted to obtain a better initial value, and only after one step of updating can it obtain a better parameter for the current task, so as to train a model that can achieve fast adaptation.

For the image modality in the 2D time–frequency domain of the signal, the use of a conventional CNN will result in gradient vanishing and gradient decreasing due to the increase in depth resulting in the model not being able to learn, so the residual module is introduced to solve this problem. At the same time, in order to lighten the network, the redundant features are weighted with the useful features, and the squeeze and excitation attention mechanism (SE) module is introduced. Finally, a network based on residual modules and a channel attention mechanism is used as the feature extractor. For the sequential modality in the one-dimensional I/Q waveform domain of the signal, due to the difficulty of identifying the feature scales of the sequence, and inspired by [31], utilizing convolution kernels with different scales will increase the fault tolerance of the feature extraction, and in summary, a Multi-Scale deep Neural Network (MSCNN) is proposed for feature extraction.

#### 3.1. Training Strategy of MAML Method Based on Meta-Learning

Addressing the issue of very limited trainable sample sizes for radar emitter intra-pulse modulation signals, this article proposed the use of a meta-learning strategy. Differing from traditional data-based learning methods, meta-learning operates with tasks as the fundamental unit rather than individual data samples. In meta-learning, both the training set and the validation set take tasks as elementary units rather than data units in traditional deep learning. Furthermore, each training and validation task consists of a support set and a query set. Through the training process, a meta-learning model learns from various tasks, enabling the model to develop cross-task feature extraction and recognition capabilities. During the test process, the model can accurately recognize new tasks, including those with completely novel categories or only a few labeled samples. In meta-learning, if the support set for a given task contains  $N$  categories and each category has  $K$  samples, the task is referred to as an  $N$  – way  $K$  – shot task; details can also be found in [24].

The core concept of the Model-Agnostic Meta-Learning (MAML) method is to obtain an optimal initialization of parameters, denoted as  $\theta$ , which does not necessarily need to perform well on any specific task but should be adaptable to a variety of similar tasks. The

objective is to achieve strong generalization capabilities. The overall algorithmic framework is illustrated in Figure 2.

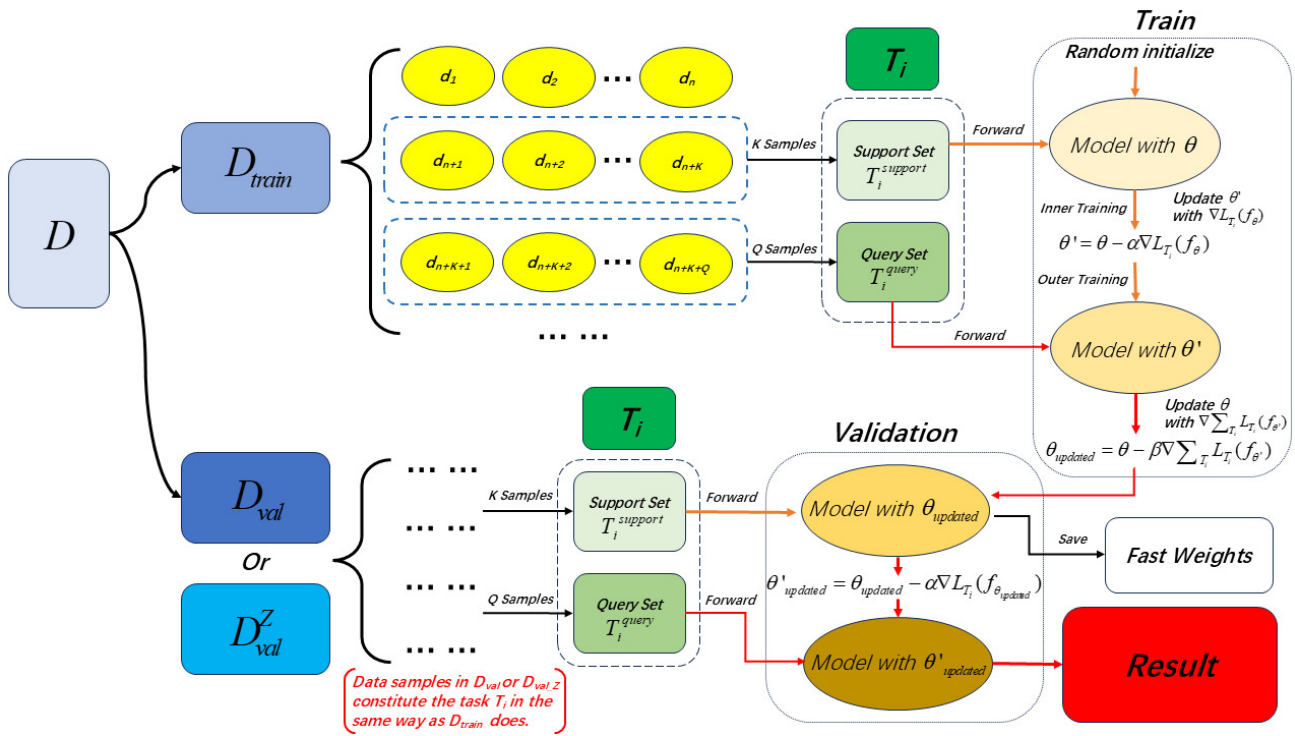


Figure 2. MAML optimization algorithm framework flow.

As shown in Figure 2, MAML is divided into two layers of training, the learning rate of the inner layer and the outer layer of training are, respectively, specified as  $\alpha$  and  $\beta$ . First, we randomly initialized the weight parameters of the model  $\theta$ , and the dataset  $D$  is divided into training dataset  $D_{train}$  and validation dataset  $D_{val}$ . Moreover, we include an additional validation set  $D_{val}^Z$  that is distinct from the regular validation set  $D_{val}$ , which will be described in detail in subsequent articles and experiments. And then several data samples in both the training dataset  $D_{train}$  and validation set are extracted to constitute a training task unit  $T_i$ , all of which obey a distribution  $P(T)$ .

- (1) **Inner layer training:** Training and learning are carried out inside each task  $T_i$ ,  $K$  data are extracted from each task  $T_i = \{d_{n+1}, d_{n+2}, \dots, d_{n+K}, d_{n+K+1}, \dots, d_{n+K+Q}\}$  to obtain the support set  $T_i^{support} = \{d_{n+1}, d_{n+2}, \dots, d_{n+K}\}$ , and the remaining  $Q$  data constitute the query set  $T_i^{query} = \{d_{n+K+1}, d_{n+K+2}, \dots, d_{n+K+Q}\}$ . The corresponding loss function is used to calculate the loss value  $L_{T_i}(f_{\theta})$  on  $T_i^{support}$  and its gradient  $\nabla$  on  $\theta$ . The submodel parameter  $\theta'$  is updated by the gradient descent of Equation (13) on this task with the calculated gradient  $\nabla$ :

$$\theta' = \theta - \alpha \nabla L_{T_i}(f_{\theta}) \tag{13}$$

After the parameter is updated, we randomly extracted  $Q$  data points. Then, we processed forward propagation and loss function, calculating and obtaining  $L_{T_i}(f_{\theta'})$ , but we did not update  $\theta'$ .

- (2) **Outer layer training:** The second layer is the outer layer training, which is the meta-learning process. First, we extracted  $batch_{train}$  tasks from the training dataset  $D_{train}$ , and each task  $T_i$  was obtained by distributed  $P(T)$  sampling. MAML sums all the loss  $L_{T_i}(f_{\theta'})$  values that were calculated in the inner training with  $\theta'$  in  $T_i^{query}$  and carries out gradient descent after calculating the gradient in order to update  $\theta$ , the weight of the real model, to obtain the updated model weights  $\theta_{updated}$ , as shown in Equation (14).



$$\theta_{\text{updated}} = \theta - \beta \nabla \sum_{T_i} L_{T_i}(f_{\theta'}) \quad (14)$$

After the end of the inner and outer training process, the  $\theta_{\text{updated}}$  obtained after updating during the training process is saved as fast weights. In the validation set,  $\theta_{\text{updated}}$  is used as the initial weight of the model, and only the inner gradient descent is carried out without updating the weight parameter of the whole model.  $\theta_{\text{updated}}$  is obtained as  $\theta'_{\text{updated}}$  after gradient descent, and the validation set  $D_{\text{val}}$  or  $D_{\text{val}}^Z$  is input into the model with the weight parameter  $\theta'_{\text{updated}}$  to output the final predicted classification results.

### 3.2. Feature Extraction of Modulation Signal in the Image Domain Based on Residual and Attention Mechanisms

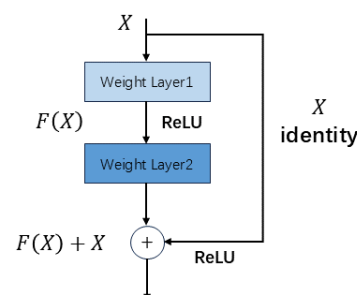
Since the introduction of AlexNet, the era of deep learning has officially begun, with Convolutional Neural Networks (CNNs) becoming increasingly deeper and their fitting capabilities becoming stronger. However, as the depth of the network increases, the risk of overfitting also rises. Numerous scholars have studied this phenomenon and identified challenges in optimizing deep networks. It has been observed that valuable information is difficult to directly utilize, leading to feature bottlenecks. Moreover, due to the nature of activation functions and the multiplicative nature of backpropagation during gradient computation, deep networks are prone to issues such as vanishing gradients or exploding gradients.

Moreover, during the training process of neural networks, an increase in the number of model parameters leads to greater information storage capacity and stronger representational power. However, this also introduces the issue of information overload. The emergence of attention mechanisms enables neural network models to focus on more critical information within the training data. The squeeze and excitation (SE) module incorporates attention mechanisms along the channel dimension, reducing the model's attention to non-essential channels and improving the overall efficiency and accuracy of the learning process.

In this article, a neural network architecture combining residual networks and attention mechanisms is designed to perform time–frequency feature extraction and classification on the signal images obtained from the preprocessing in Section 2.2. This approach effectively avoids information redundancy and enhances both the efficiency and accuracy of the model.

#### 3.2.1. Deep Residual Neural Network

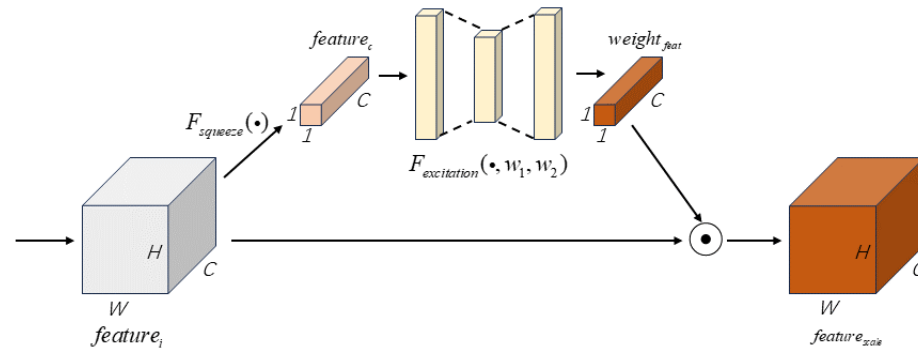
The Deep Residual Neural Network, also known as ResNet [32], is designed based on a residual structure, utilizing shortcut connections through skip paths in the residual network. The structure of the residual module is illustrated in Figure 3, where the residual connection is represented by a skip line, with  $X$  as the input and  $F(X)$  as the residual function. ResNet deepens the network by stacking residual modules, which mitigates the performance degradation issues that often arise when networks become excessively deep.



**Figure 3.** Basic structure of residual network.

### 3.2.2. Squeeze and Excitation Attention Mechanism

The squeeze and excitation attention mechanism (SE) module applies a series of transformations to generate a weight matrix [33], which is then used to reconstruct the original features and extract key feature information. By automatically learning the importance of each feature channel, the SE module assigns corresponding weights to each channel, enhancing the utilization of feature channels that are more useful for the task, while simultaneously suppressing the non-essential feature channels. The structure of the SE module is illustrated in Figure 4.



**Figure 4.** Principle of SE channel attention mechanism, where  $H$ ,  $W$  and  $C$  represent the height, width, and channel dimensions of the input feature image  $feature_i$ , respectively.

The channel attention mechanism consists of three processes: squeeze, excitation, and regulation.

**Squeeze Process:**  $F_{squeeze}(\cdot)$ , the Global Average Pooling (GAP), is applied to the feature map  $feature_i$  with dimensions  $H \times W \times C$ , transforming  $feature_i$  along the channel direction to generate a  $1 \times 1 \times C$  feature channel descriptor  $feature_c$ , as described in Equation (15), where  $feature_i(m, n)$  is the eigenvalue of the feature graph  $feature_i$  at the position  $(m, n)$ .

$$F_{squeeze}(feature_i) = \frac{1}{H \times W} \sum_{m=1}^H \sum_{n=1}^W feature_i(m, n) \tag{15}$$

**Excitation Process:** Two fully connected layers  $F_{excitation}(\cdot, w_1, w_2)$  are used to learn the importance of channel features, thus obtaining the channel attention weight of features  $weight_{feat} \in R^{1 \times 1 \times C}$ , as shown in Equation (16).

$$F_{excitation}(feature_i, w_1, w_2) = \sigma(w_2 \cdot \delta(w_1 \cdot feature_i)) \tag{16}$$

where  $w_1$  and  $w_2$  are the weight parameters of the two fully connected layers,  $\sigma$  indicates the ReLu activation function, and  $\delta$  indicates the Sigmoid activation function.

**Regulation Process:** The attention weight  $weight_{feat}$  is multiplied by the input feature graph  $feature_i$ , so we can obtain the output feature graph  $feature_{scale}$  which incorporates the channel weight.

### 3.2.3. SE-ResNet Network Architecture

By embedding the SE module into the ResNet residual module, SE-ResNet facilitates the fitting of correlations between network channels and effectively reduces the computational complexity of the model, thus decreasing its overall size. In this article, SE-ResNet is used as the feature extractor for the two-dimensional time–frequency image modality of the signal. The SE-ResNet architecture consists of one convolutional module, eight residual attention modules, and one fully connected layer. The input to the network is a single-channel time–frequency image of size  $224 \times 224 \times 1$ . The structure of the SE-ResNet network is shown in Figure 5.

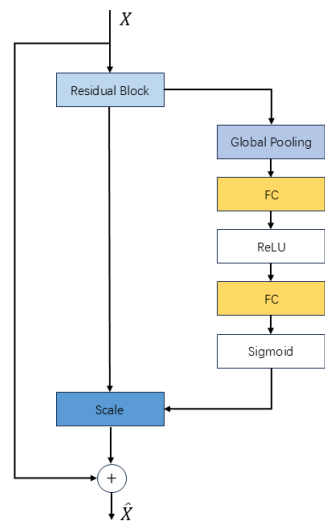


Figure 5. SE-ResNet module.

### 3.3. Feature Extraction of Modulation Signal in the I/Q Waveform Domain Based on Multi-Scale 1D Convolution

For the complex radar emitter signals in the form of I/Q waveforms, the preprocessing outlined in Section 2.3 is used to convert the data into a format suitable for input into a One-Dimensional Convolutional Neural Network (1D CNN). However, standard convolutional networks typically utilize convolution kernels of a single scale, which limits the range of features that can be extracted. The quality of the model’s classification results is somewhat dependent on the amount of spatial information available in the signal. Yet, in few-shot learning conditions, the limited spatial information in the signals becomes a significant challenge. By increasing the multi-scale span of the model, the richness of the signal’s spatial information can be enhanced [34].

To address the problems above, the article proposes a One-Dimensional Multi-Scale Convolutional Neural Network (1D MSCNN) with different convolution kernel sizes to capture various types of features. These complementary features improve the accuracy of classification. The input to the network is a one-dimensional complex I/Q waveform sequence, and the network consists of an input layer, 10 convolutional modules, a fully connected layer, and an output layer, which are all cascaded sequentially. Each convolutional module comprises a multi-scale convolution unit, a max-pooling layer, and a batch normalization layer, which are also arranged sequentially. The structure of 1D MSCNN is shown in Figure 6.

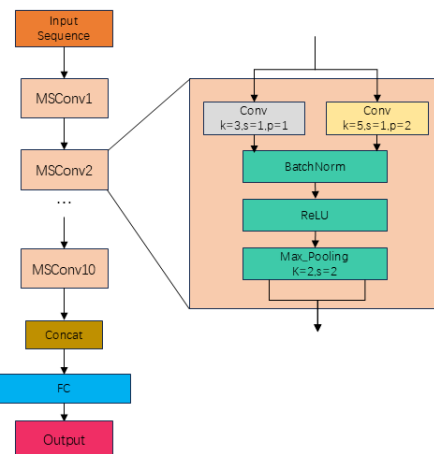


Figure 6. The 1D MSCNN structure flow.

### 3.4. Intra-Pulse Modulated Signal Recognition of Radar Emitter Based on Multimodal Fusion

Currently, most signal modulation recognition approaches only consider single modal information from a single dimension when selecting signal features. This becomes particularly challenging in scenarios with few-shot learning conditions, where such single modal features are insufficient to provide a robust representation. By leveraging the diversity and complementarity between multiple modalities, multimodal fusion allows for the extraction of adequate features even when sample sizes are limited. Therefore, this article proposes a multimodal fusion network, Double Channel Fusion Attention Neural Network (DC-FANet), combined with a meta-learning strategy for feature extraction and the classification of radar emitter signals.

First, the fusion model is constructed using SE-ResNet and 1D MSCNN, which are used to extract features from the image domain and the I/Q waveform domain, respectively. The deep features from these two modalities are then fused at the feature level according to a specific ratio, reflecting the relative importance of each modality. The combined features, along with the deep features from the individual modalities, are input into the SoftMax layer for separate predictions. Finally, decision fusion is performed by linearly weighting and summing the three prediction results, which are then passed through a SoftMax layer to obtain the final multimodal output.

Moreover, to enable the model to learn across tasks under few-shot learning conditions, the MAML algorithm is used in place of traditional optimization algorithms. The model weights obtained from training on the support set are saved as fast weights, which are fine-tuned on the query set without being updated to produce predictions. The technical flow of the multimodal fusion model is shown in Figure 7.

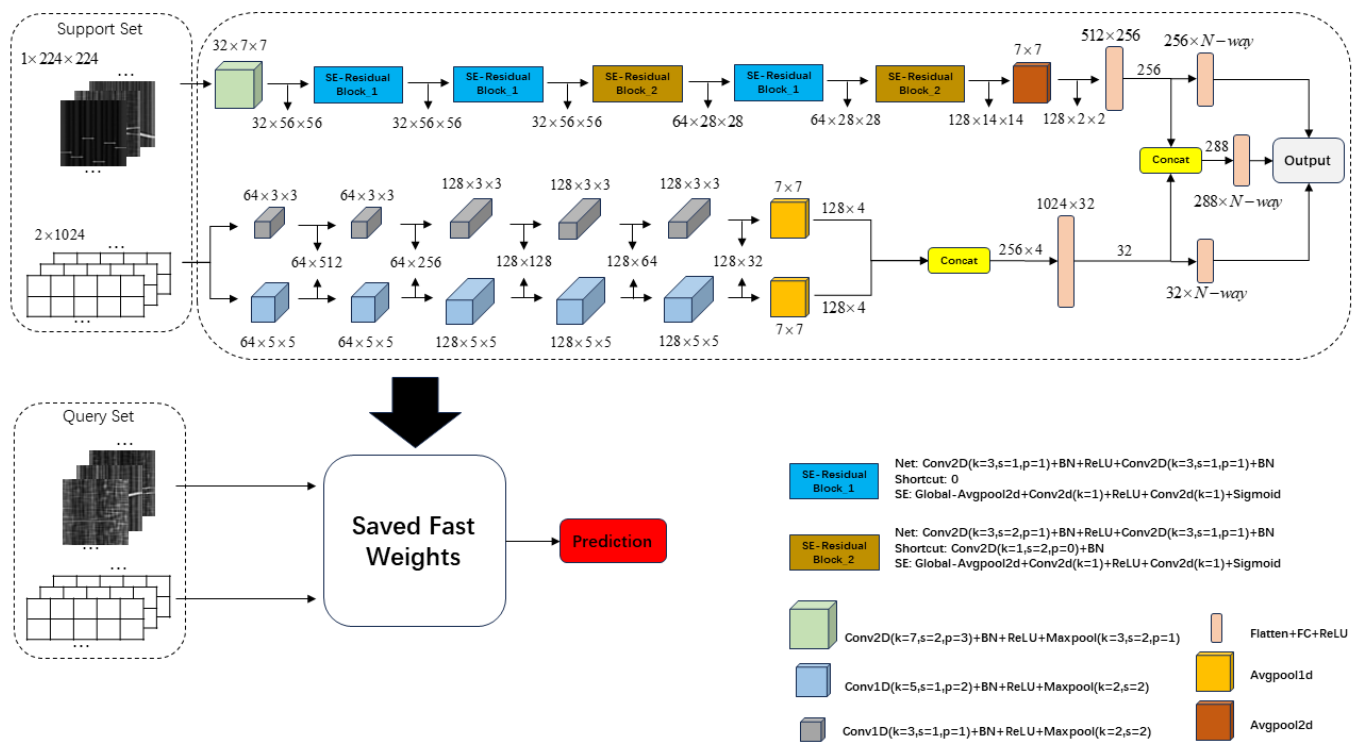


Figure 7. Technical process framework of multimodal fusion model.

In this article, we use cross-entropy to calculate the loss between the predicted results and the true label distribution. Let  $x_i$  represent the  $i$ / $text{th}$  sample,  $p(x_i)$  be the true label, and  $q(x_i)$  represent the predicted label. The cross-entropy loss function is defined as shown in Equation (17).

$$E_{\text{cross}}(p, q) = - \sum_{i=1}^n p(x_i) \log[q(x_i)] \quad (17)$$

Since all three predicted results will influence the final decision, and we aim to minimize the distance between predictions with the same label while maximizing the distance between predictions with different labels, we propose a joint loss function as shown in Equation (18) to achieve consistency across multimodal objectives.

$$E_{\text{joint-cross}}(p, q_i, q_s, q_j) = \frac{4}{5} \sum_{i=1}^n p(x_i) \log[q_i(x_i)] - \frac{1}{5} \sum_{i=1}^n p(x_i) \log[q_s(x_i)] - \sum_{i=1}^n p(x_i) \log[q_j(x_i)] - \frac{1}{n} \left\{ \frac{4}{5} \sum_{i=1}^n q(x_i) \log[q_i(x_i)] + \frac{1}{5} \sum_{i=1}^n q(x_i) \log[q_s(x_i)] \right\} \quad (18)$$

where  $q_i(x_i)$  and  $q_s(x_i)$  represent the unimodal prediction results from the time–frequency image and the I/Q waveform sequence, respectively.  $q_j(x_i)$  is the prediction result from the joint modal features. Based on the ratio of the joint features to the individual unimodal features, which is 320:256:64=5:4:1, the ratio of the unimodal feature vector to the joint feature vector is used as a coefficient.

#### 4. Simulation Experiment and Analysis

In this section, the data used in this paper and the proposed method will be described and experimented.

Since there is no publicly produced dataset for radar radiation source signals, in order to simulate the real complex electromagnetic environment as much as possible, the radar pulse signal parameter settings in this paper refer to some books and publicly published papers [10,11,35,36]. In this paper, 13 types of major radar radiation source signals are considered, and the detailed radar signal parameters are shown in Table 1.

**Table 1.** Radar signal parameters for simulation.

Signal Type	Signal Parameter	Parameter Range
NS	Sampling Frequency $F_s$	100 MHz
	Center Frequency $F_c$	$[F_s/6 - F_s/5]$
Barker	Sampling Frequency $F_s$	100 MHz
	Center Frequency $F_c$	$[F_s/6 - F_s/5]$
	Code Length	[3,4,5,7,11]
	Code Width	[40 ns–290 ns]
LFM	Sampling Frequency $F_s$	100 MHz
	Center Frequency $F_c$	$[F_s/6 - F_s/5]$
	Sweep Width $\Delta f$	$[F_s/20 - F_s/16]$
	Sweep Direction	[Up, Down]
MPSK	Sampling Frequency $F_s$	100 MHz
	Center Frequency $F_c$	$[F_s/6 - F_s/5]$
	Frank Code Length	[4,9,16]
	P1 Code Length	[4,9,16]
	P2 Code Length	[4,9,36]
	P3 Code Length	[8,10,16]
P4 Code Length	[8,10,16]	
Code Width	[40 ns–290 ns]	
Costas	Sampling Frequency $F_s$	100 MHz
	Center Frequency $F_c$	$[F_s/6 - F_s/5]$
	Frequency-Hopping $F_H$	[3,4,5,6]
	Fundamental Frequency	$[F_s/32 - F_s/25]$



**Table 1.** *Cont.*

Signal Type	Signal Parameter	Parameter Range
T1,T2	Sampling Frequency $F_s$	100 MHz
	Center Frequency $F_c$	$[F_s/6-F_s/5]$
	Phase State Number $N_c$	2
	Segments Number $k$	$[4,5,6]$
T3,T4	Sampling Frequency $F_s$	100 MHz
	Center Frequency $F_c$	$[F_s/6-F_s/5]$
	Phase State Number $N_c$	2
	Segments Number $k$	$[4,5,6]$
	Sweep Width $\Delta f$	$[F_s/20-F_s/15]$

Subsequently, we are about to describe the experimental environment and construct the multimodal dataset by simulation modeling of the radar signals strictly according to the parameters in Table 1 using the processing methods mentioned in Section 2. The detailed composition of the multimodal dataset will be clarified in this section.

In addition, the effectiveness of the MAML method in few-shot conditions will be validated in this section, and the effect of the number of samples on the model performance will be explored. In particular, we will conduct comparison and ablation experiments between the model under multimodal fusion and the model under each unimodal modality as a way to highlight the superiority of our proposed multimodal model incorporating MAML.

#### 4.1. Experimental Environment and Settings

The experiments in this article are conducted using the deep learning framework PyTorch with the network model built and tested on a personal computer. The system configuration includes Windows 11 as the operating system, 16 GB of RAM, an Intel i5-13600KF CPU, and an NVIDIA 4080 GPU with 16 GB of VRAM. The various methods utilized in this experiment are accompanied by hyperparameter settings, which are listed in Table 2 for detailed reference.

**Table 2.** Hyperparameter settings for different methods.

Method	Task Num	Batchsize	Epochs	Learning Rate (LR)	Inner LR	Outer LR
Base-CNN	-	128	100	0.001	-	-
ResNet18	-	256	100	0.001	-	-
Base-CNN-MAML	10	-	150	-	0.02	0.01
ResNet18-MAML	10	-	180	-	0.02	0.003
SE-ResNet-MAML	10	-	180	-	0.02	0.003
1d-VGG-MAML	10	-	200	-	0.01	0.005
1d-MSCNN-MAML	10	-	200	-	0.01	0.005
DCFANet-MAML	8	-	150	-	0.02	0.01

#### 4.2. Dataset Description

In this article, MATLAB 2022a is used to simulate the 13 common radar emitter modulation signals described in Section 2.1, including NS, Barker, Frank, P1, P2, P3, P4, LFM, Costas, T1, T2, T3, and T4. The simulation is conducted in an Additive White Gaussian Noise (AWGN) environment, and the number of time-domain signal sampling points is set to 1024. To construct a multimodal dataset, each signal sample undergoes the time–frequency transformation and waveform processing described in Sections 2.2 and 2.3, resulting in both a time–frequency image dataset and a waveform dataset. The structure of the entire multimodal dataset is as follows.

##### Image Domain Dataset:

- Sample format: CWD time–frequency image, file storage format is .png;
- Sample dimensions:  $[224, 224, 1]$ , 224 indicates the length and width of the image and 1 is the number of RGB channels, since the time–frequency plots in this article are in grayscale;

- SNR range:  $-20$  to  $10$  dB, with a step size of  $2$  dB;
- Number of signal samples per single modulation type at a specific SNR:  $150$ .

#### Waveform Domain Dataset:

- Sample format: I/Q waveform sequence, file storage format is `.mat`;
- Sample dimensions:  $[1024, 2]$ ,  $1024$  is the length of the sequence, and  $2$  corroborates that the characteristic dimensions of the sequence are the real and imaginary parts of the I/Q waveforms;
- SNR range:  $-20$  to  $10$  dB with a step size of  $2$  dB;
- Number of signal samples per single modulation type at a specific SNR:  $150$ .

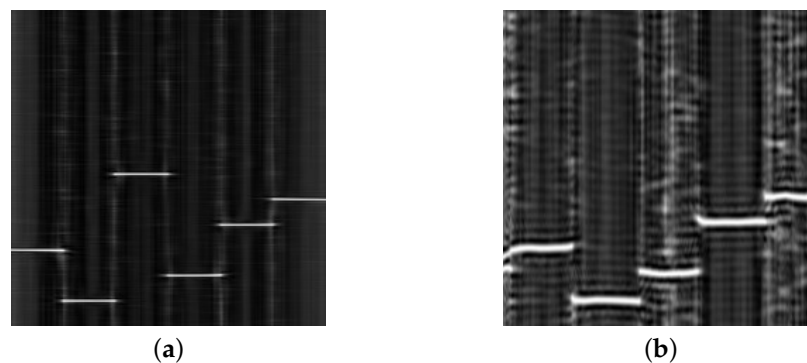
In this article, the dataset  $D$  was randomly divided into a training set  $D_{\text{train}}$  and a validation set  $D_{\text{val}}$  at a ratio of  $4 : 1$ , resulting in a total of  $24,960$  training samples for each individual modality. To ensure the data alignment across different modalities, the data distribution was kept consistent across all modality datasets.

To further verify the generalization ability and cross-task learning capability of the model, an additional validation set  $D_{\text{val}}^Z$ , distinguished from the standard validation set  $D_{\text{val}}$ , is introduced. The model was trained using the standard training dataset  $D_{\text{train}}$ , and it was validated with  $D_{\text{val}}^Z$ .

The differences between  $D_{\text{val}}^Z$  and  $D_{\text{val}}$  are outlined below:

1. The signals in the regular dataset  $D$  are simulated using a Rician channel, while  $D_{\text{val}}^Z$  uses a Rayleigh channel;
2. In  $D$ , the signal propagation does not account for multipath fading, time delay, or doppler shift. However, in  $D_{\text{val}}^Z$ , these three influencing factors are incorporated into the signal.

The visual differences in the time–frequency images are illustrated in Figure 8.



**Figure 8.** Costas time–frequency images at  $-2$  dB from two different validation sets: (a) Costas time–frequency image from  $D$ ; (b) Costas time–frequency image from  $D_{\text{val}}^Z$ .

It is essential to emphasize that the training set  $D_{\text{train}}$  does not partition samples based on the SNR. Specifically, within each category of radiation source signals, there are samples corresponding to  $16$  different SNR levels (ranging from  $-20$  to  $10$  dB with increments of  $2$  dB). This approach is designed to enable MAML to search for optimal initial parameters across the entire SNR spectrum rather than being limited to a specific SNR level. Conversely, the validation sets  $D_{\text{val}}$  and  $D_{\text{val}}^Z$  are meticulously stratified by SNR, facilitating a clear and comprehensive evaluation of the model’s performance at each SNR level during the subsequent analysis and testing phases.

### 4.3. Algorithm’s Performance Analysis

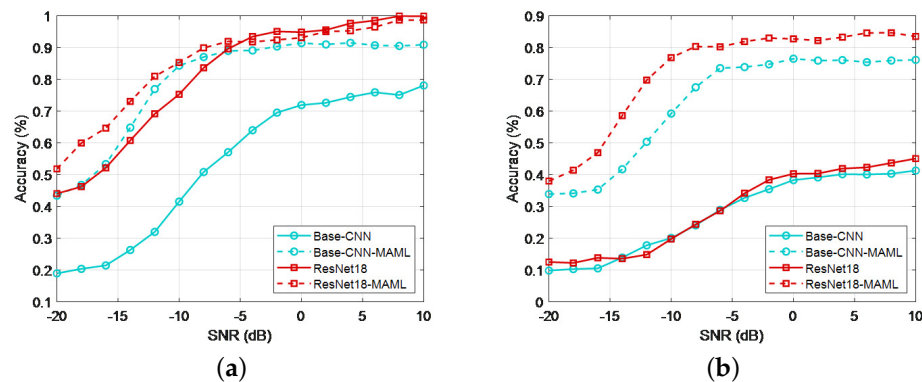
#### 4.3.1. Comparison between MAML Algorithm and Traditional Optimization Algorithm

In this section, experiments are conducted to explore the recognition performance of models using traditional optimization algorithms and the MAML algorithm under few-shot learning conditions. A CNN with  $5$  convolutional layers and  $1$  linear layer is

used as the Baseline model, which we call Base-CNN. Then, we train Base-CNN on the two-dimensional image dataset along with the advanced model ResNet18. Both models are tested on the time–frequency images of  $D_{\text{val}}$  and  $D_{\text{val}}^Z$ .

For the MAML optimization, the gradient update steps are set to 5, and the meta-learning task is set to 5 – way 2 – shot. The inner and outer learning rates are set to 0.005 and 0.02, respectively. The traditional optimization method uses the Adam optimizer with a learning rate of 0.001. The comparison results, showing the accuracy curves, are presented in Figure 9.

From Figure 9a, we can observe that Base-CNN optimized using MAML achieves an approximately 30% improvement in accuracy across the entire SNR range compared to the traditionally optimized Base-CNN. Furthermore, Base-CNN-MAML’s performance on the few-shot dataset is comparable to that of ResNet18 despite the increase in training time. However, Base-CNN-MAML has only 0.520M parameters, which is nearly 1/40th of ResNet18’s parameter count (11.6 M). As shown in Figure 9b, models using traditional optimization algorithms exhibit poor generalization ability in few-shot learning conditions, failing to achieve the expected recognition accuracy even at a high SNR. In contrast, both Base-CNN-MAML and ResNet-MAML not only perform well on the original  $D_{\text{val}}$  but also demonstrate excellent generalization performance on  $D_{\text{val}}^Z$ .



**Figure 9.** Part of the time–frequency image of the pulse-modulated signal: (a) the accuracy rate on  $D_{\text{val}}$  under each SNR; (b) the accuracy rate on  $D_{\text{val}}^Z$  under each SNR.

Experiments have demonstrated that on few-shot datasets, the MAML optimization algorithm significantly enhances the model’s generalization performance and recognition accuracy compared to traditional optimization algorithms. This validates the effectiveness of using the MAML optimization algorithm under few-shot learning conditions.

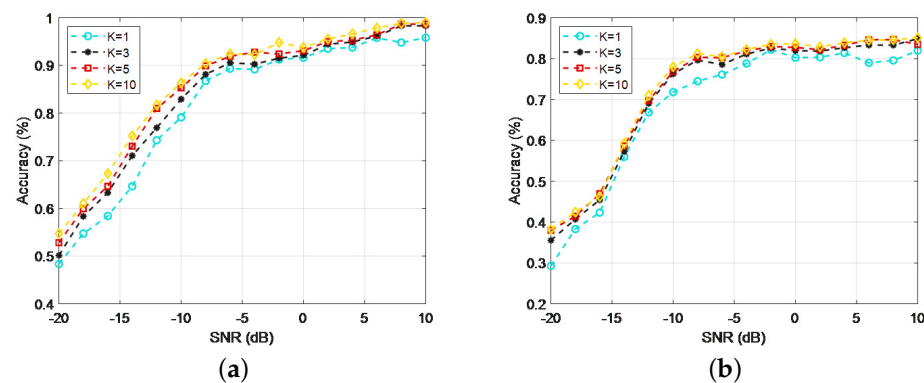
#### 4.3.2. Influence of Sample Size on Model Recognition Performance

Building on the performance advantage of the MAML optimization algorithm demonstrated in the previous section, this section continues the experimental analysis of MAML’s recognition performance under few-shot learning conditions. Since the MAML algorithm adopts a meta-learning strategy, the task settings  $N$  – way  $K$  – shot also influence the model’s performance. For a 5 – way radar emitter signal recognition task, the recognition performance is analyzed and compared for different numbers of samples per class in the task’s support set with  $K = 1, 3, 5, 10$ .

In the experiment, ResNet18 is used as the feature extraction network. The variation in recognition accuracy of radar emitter signals on the  $D_{\text{val}}$  and  $D_{\text{val}}^Z$  with respect to SNR is shown in Figure 10. In each recognition task, 5 samples per class are selected to form the query set, which indicates  $Q = 5$ .

The results show that at an SNR of  $-2$  dB, the recognition accuracy on  $D_{\text{val}}$  for  $K = 1$  exceeds 90%, further proving the feasibility of the MAML optimization algorithm under few-shot conditions. As the sample size increases, the overall recognition accuracy on both

$D_{\text{val}}$  and  $D_{\text{val}}^Z$  continues to improve. Notably, on  $D_{\text{val}}$  at  $-10$  dB, the accuracy improves by 3.81%, 6.15%, and 7.23% for  $K = 3, 5, 10$  compared to  $K = 1$ , respectively.



**Figure 10.** Influence of MAML on model recognition accuracy under different sample sizes: (a) recognition accuracy rate of ResNet18-MAML on  $D_{\text{val}}$  under different sample sizes; (b) recognition accuracy rate of ResNet18-MAML on  $D_{\text{val}}^Z$  under different sample sizes.

However, when  $K > 5$ , the improvement in overall recognition accuracy begins to plateau, and increasing the sample size significantly raises both training and inference time. Therefore, a compromise is adopted, and in subsequent experiments, the recognition task is set to 5 – way 5 – shot.

#### 4.3.3. Comparative Analysis of Model Recognition Results under Different Unimodal and Multimodal Fusion Approaches

Based on the experimental results from the previous two sections, it is evident that current advanced models still struggle to achieve high recognition rates at low SNR. Therefore, in the following analysis, multiple models will be used to extract features and perform classification for radar emitter signals across different unimodal modalities. The results will then be compared and analyzed to assess performance improvements and limitations.

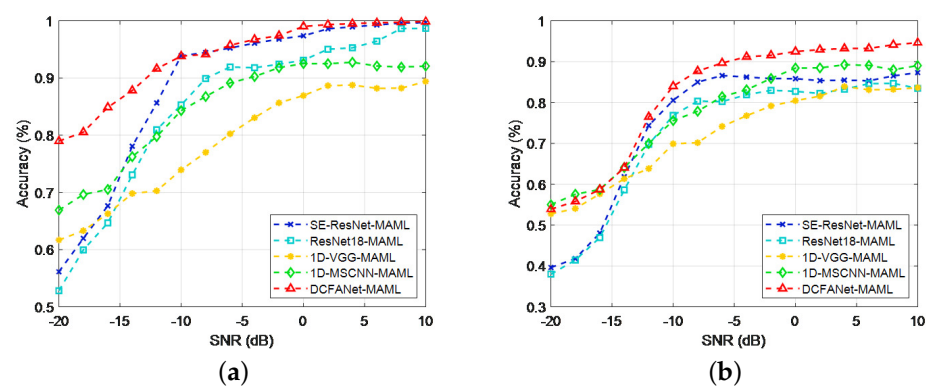
In this article, different models are adopted to extract features and classify radar emitter signals based on their respective modalities. For the one-dimensional I/Q waveform sequence modality, the 1D VGG model, constructed by analogy to the VGG architecture, and the 1D MSCNN model proposed in Section 3.3, are used. For the two-dimensional time–frequency image modality, in addition to ResNet18 as previously mentioned, the SE-ResNet model introduced in Section 3.2 is also utilized for feature extraction.

Experimental results demonstrate that the performance of different signal modalities varies depending on the model applied to the corresponding modality with each modality exhibiting its own strengths and weaknesses. To leverage the diversity and complementarity of features from different modalities, this article proposes fusing the features from multiple modalities to form a multimodal representation. Considering the potential differences in decision making between the fused features and unimodal features, decision-level fusion is employed to unify the outputs from both unimodal and multimodal models. This approach is implemented in the multimodal fusion model DCFANet as previously introduced.

In this article, the trained multimodal fusion model, along with the models from each unimodal modality, is experimentally validated on both  $D_{\text{val}}$  and  $D_{\text{val}}^Z$ . The recognition accuracy curves for the different unimodal models and the multimodal model as a function of SNR are presented in Figure 11.

As shown in Figure 11a, at a low SNR, SE-ResNet-MAML and ResNet18-MAML perform less effectively in extracting and recognizing the two-dimensional time–frequency image modality compared to 1D-VGG-MAML and 1D-MSCNN-MAML, which focus on extracting and recognizing the one-dimensional I/Q waveform sequence modality. Specifically, the feature extraction and recognition models for the sequence modality achieve an average recognition accuracy 11.53% higher than the image modality models at  $-20$  dB.

In contrast, at a higher SNR, the models focused on the two-dimensional image modality outperform those focused on the sequence modality. For instance, at 4 dB, the image modality models show an average improvement in recognition accuracy of 9.79% compared to the sequence modality models. The DCFANet-MAML model, which combines the features from both modalities, demonstrates superior performance. It benefits from the sequence modality's excellent noise resistance while also capturing and representing spatial features from the image modality. As a result, its recognition accuracy is higher than the two unimodal models at both low and high SNRs. The overall recognition accuracy of DCFANet-MAML across the entire SNR range reaches 89.3%, with an accuracy above 90% at  $-12$  dB, and the recognition rate approaches 100% when the SNR exceeds 2 dB. This confirms that the proposed multimodal fusion method for few-shot learning effectively complements the information from different modalities, resulting in more comprehensive joint features that significantly enhance recognition accuracy and model robustness.



**Figure 11.** The relationship between recognition accuracy and SNR of different models in different modes. (a) recognition accuracy of each method on  $D_{val}$ . (b) recognition accuracy rate of each method on  $D_{val}^Z$ .

Subsequently, focusing on Figure 11b, we analyze the generalization performance and cross-task learning capability of DCFANet-MAML. The model demonstrates impressive results on  $D_{val}^Z$  as well. At both low and high SNRs, it effectively leverages the fused features, maintaining an accuracy of 83.9% at  $-10$  dB, even when faced with signals affected by various factors. The fact that the knowledge learned from the training set  $D_{train}$  can be transferred to  $D_{val}^Z$  while maintaining robustness further validates the cross-task learning characteristics of DCFANet-MAML. This robust performance under few-shot conditions underscores the effectiveness of the proposed DCFANet-MAML model.

The parameter and calculation amount of the models involved in the experiments are shown in Table 3. This comparison further illustrates that the multimodal fusion method, DCFANet-MAML, maintains a lightweight structure while still achieving excellent performance under few-shot learning conditions.

**Table 3.** Comparison of parameters and calculation amount of different models.

Model	Params	FLOPs
ResNet18-MAML	11.7 M	7.4 M
SE-ResNet-MAML	11.2 M	6.2 M
1D-VGG-MAML	0.85 M	7.1 M
1D-MSCNN-MAML	0.83 M	4.3 M
DCFANet-MAML	0.94 M	5.5 M

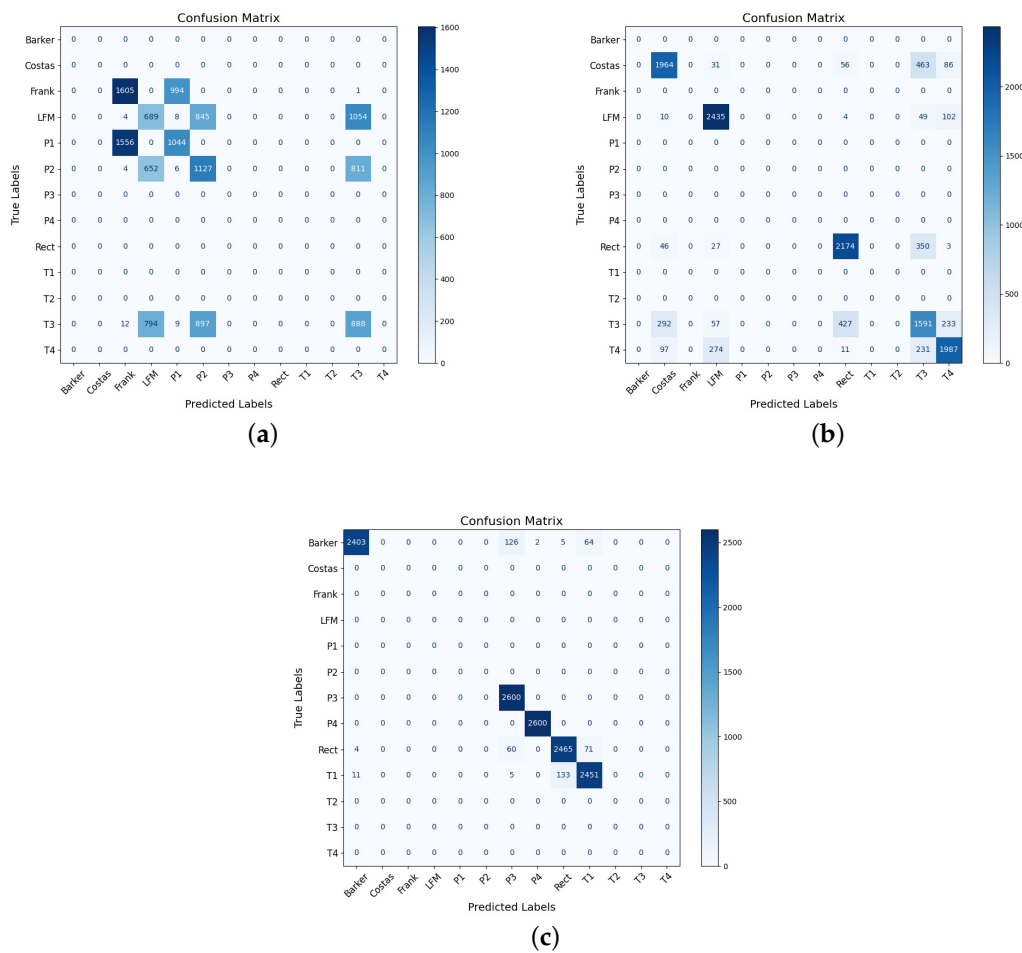
#### 4.3.4. Confusion Matrix and Feature Visualization Analysis of the Multimodal Fusion Method

In this section, we delve deeper into the recognition performance of the proposed multimodal fusion method, DCFANet-MAML, under few-shot learning conditions. The



evaluation is conducted through the use of a confusion matrix and feature visualization on  $D_{val}^Z$ , providing a comprehensive assessment of the model’s performance in terms of classification accuracy and feature representation.

Figure 12 presented the confusion matrices for the DCFANet-MAML model under three different SNR conditions:  $-20$  dB,  $-10$  dB, and  $0$  dB. These matrices represent the classification performance on three meta-validation tasks with randomly sampled categories. The categories are as follows: LFM, P1, P2, T3, and Frank at  $-20$  dB; Costas, Rect, LFM, T3, and T4 at  $-10$  dB; and Barker, Rect, T1, P3, and P4 at  $0$  dB. Each meta-validation task consists of 2600 samples. From the results, it is evident that at  $-20$  dB, the model exhibits significant errors across all sampled categories, although it still manages to extract subtle features from the signals. At  $-10$  dB, the main errors occur between the categories T3 and T4, as their two-dimensional and one-dimensional features, are highly similar. At  $0$  dB, while some errors persist, the F1 scores for the sampled categories are calculated as 0.9578, 0.9475, 0.9452, 0.9646, and 0.9996, indicating that the proposed method can achieve high-quality and highly reliable predictions on randomly sampled meta-tasks at  $0$  dB.



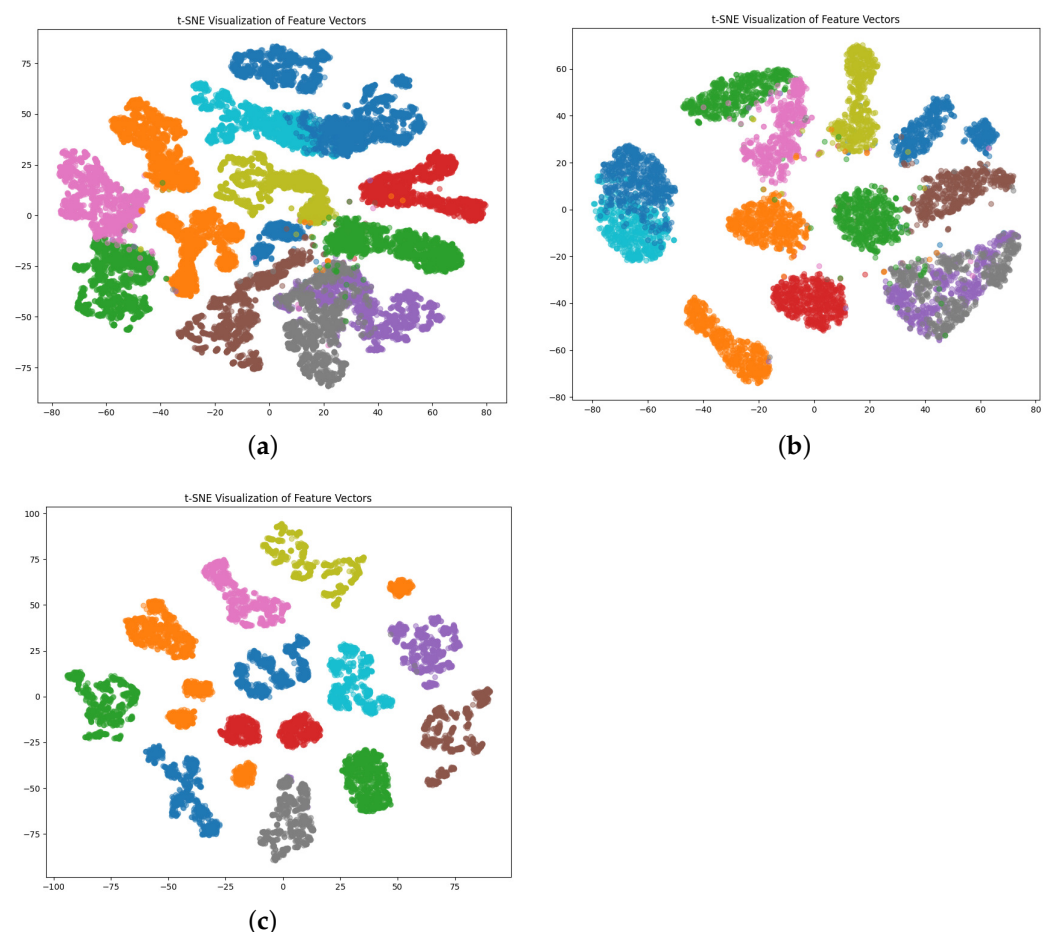
**Figure 12.** Meta-validation classification confusion matrix on  $D_{val}^Z$  for different SNRs: (a) Meta-validation classification under SNR =  $-20$  dB, sampling categories: LFM, P1, P2, T3, Frank; (b) Meta-validation classification under SNR =  $-10$  dB, sampling categories: Costas, Rect, LFM, T3, T4; (c) Meta-validation classification under SNR =  $0$  dB, sampling categories: Barker, Rect, T1, P3, P4.

Next, the focus shifts to the joint feature vectors obtained from DCFANet-MAML under few-shot conditions on  $D_{val}^Z$ , which are extracted and concatenated. These feature vectors are then flattened and projected onto a two-dimensional plane using t-SNE for visualization,

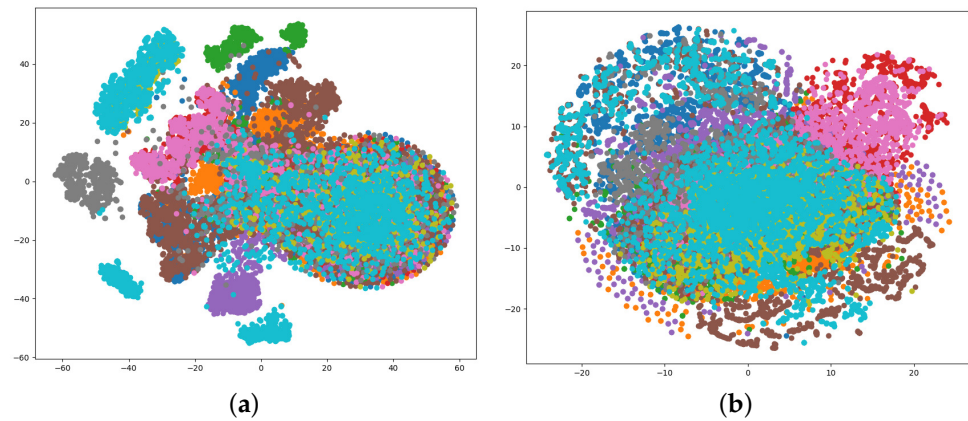
as shown in Figure 13. Additionally, t-SNE is also applied to the raw, unprocessed data from  $D_{val}^Z$  to provide a clear before-and-after contrast, which is presented in Figure 14, highlighting the improvement in feature separability and representation achieved by the model.

As shown in Figure 14, both the image and sequence modalities of  $D_{val}^Z$  exhibit significant overlap with many regions where features from different categories are indistinguishable. Yet, some categories still show larger inter-class separations, particularly where a high SNR enhances the clarity of the features, whereas at low SNRs, the features become more difficult to distinguish. This visualization result aligns well with real-world conditions, reflecting the challenge of feature extraction at lower SNR.

After feature extraction using the DCFANet-MAML method, features across all SNRs tend to become more dispersed, as shown in Figure 13a. However, some categories still exhibit small inter-class distances, and some overlapping feature points persist. Further analysis of the mapped feature vectors below  $-10$  dB and above  $2$  dB SNR is conducted, as illustrated in Figure 13b,c. These figures show that the SNR significantly influences the inter-class distances of the features, affecting feature discriminability. Nonetheless, DCFANet-MAML can still distinguish the majority of categories at low SNR below  $-10$  dB. At higher SNR, the inter-class distances are relatively larger, making the features easier to differentiate.



**Figure 13.** High-dimensional feature vector mapping of DCFANet: (a) feature vector mapping across all SNRs; (b) feature vector mapping for SNRs under  $-10$  dB; (c) feature vector mapping for SNRs above  $-2$  dB.



**Figure 14.** t-SNE visualization of  $D_{val}^Z$ : (a) t-SNE visualization of the time–frequency image modality; (b) t-SNE visualization of the I/Q waveform sequence modality.

In summary, the proposed method demonstrates the ability to obtain richer and more distinguishable features, improving the model’s performance and recognition accuracy.

#### 4.4. Method Limitations and Future Work

Despite the promising results demonstrated by the proposed model in this study, several limitations must be acknowledged, and it is also essential to provide avenues for further exploration in future research.

Firstly, due to the absence of publicly accessible datasets and the challenges associated with hardware limitations in acquiring real radar emitter signals, we are constrained to generating a trainable signal dataset solely within a simulated environment. While this allows for controlled experimentation, the performance of the model may differ when applied to real-world radar emitter signals, which are subject to more complex noise characteristics, signal distortions, and environmental factors. To successfully incorporate real radar signal data into the model, it is imperative to maintain consistency between the number of sampling points and the length of the signal sequences with the specifications outlined in our experimental design. This congruence ensures that all data transformations align precisely with those described in the article. Furthermore, adjustments to the real radar signal characteristics and model parameters can be facilitated by fine tuning variables such as STFT parameters and the hyperparameters of the model. The potential discrepancies between simulated data and real-world radar signals may significantly impact the model’s generalization performance. Therefore, future research will aim to explore additional methodologies for obtaining empirical radar data as well as investigate the model’s adaptability to the inherent variations present in real-world radar environments.

Additionally, the study only considered radar emitter signals under the AWGN environment. In more complex electromagnetic environments, a variety of noise types are likely to occur, including interference from other signals, atmospheric noise, and jamming techniques. The model’s robustness to these more intricate noise conditions remains unexplored and represents an area of ongoing research. Future work will focus on evaluating and enhancing the model’s performance in more realistic and diverse channel environments.

Moreover, the signals used in this study are limited to those that are known and present in the existing databases. The few-shot conditions represent only one of the challenges arising from a complex electromagnetic environment, but in real-world scenarios, electromagnetic environments often contain unknown or previously unencountered radar emitter signals. The current approach may not be suitable for detecting or recognizing such unknown signals. With the superior generalization capabilities provided by MAML, future research will extend the work by investigating the detection and recognition of unknown radar emitter signals with the aim of adapting and improving the proposed model to handle previously unseen or unidentified signals in complex environments.

## 5. Conclusions

In this article, we first propose the MAML method to optimize the model parameters and improve the training process, acclimatizing the training process to the few-shot environment. By introducing residual block, SE block, and 1D MSCNN, we design a multimodal fusion neural network, DCFANet, based on deep learning. The function of this network is to fuse the multimodal information from the I/Q waveform domain and the time–frequency image domain of the intra-pulse modulated signal of the emitter in order to realize the recognition of the intra-pulse modulated signal of the radar emitter. Moreover, designing the multimodal collaborative converged infrastructure and introducing joint loss function are also benefit to the maintaining the uniqueness and exclusivity of each mode. The experimental simulation generated 13 different kinds of intra-pulse modulated signals. We then tested the trained MAML and the proposed multimodal fusion neural network to recognize the modulated signal and at last compared the results with most existing advanced network models. Experiments show that the proposed multimodal fusion method has a good result when classifying 13 different intra-pulse modulated signals of a radar emitter under few-shot learning conditions. In particular, when we set the experimental setting to be  $N - way = 5$   $K - shot = 5$ , the global average accuracy on the validation set  $D_{val}$  reaches 95.7%, and the classification accuracy reaches 93.8% when the SNR is  $-10$  dB; the global average accuracy on the validation set  $Z D_{val}^Z$ , of which signals are added in doppler shift, multipath fading and time delay, reaches 87.3%, and the classification accuracy reaches 83.5% under the condition of SNR  $-10$  dB.

**Author Contributions:** Conceptualization, Y.L.; methodology, Y.L., Q.Z. and C.G.; software, Y.L.; validation, Y.L., S.H. and J.C.; formal analysis, Y.L.; investigation, Y.L.; resources, C.G.; data curation, Y.L. and S.H.; writing—original draft preparation, Y.L. and S.H.; writing—review and editing, Y.L. and J.C.; visualization, Y.L.; supervision, C.G. and Q.Z.; project administration, C.G.; funding acquisition, C.G. and Q.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding

**Data Availability Statement:** The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author. Moreover, the authors have uploaded the code for generating the radar intra-pulse modulated signal dataset used in this paper to the author’s GitHub profile; the URL is at <https://github.com/Liuyh0308/Radar-Intra-Pulse-Modulation-Signal-Simulation>(accessed on 18 September 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Gang, H.; Jiandong, L.; Donghua, L. Study of modulation recognition based on HOCs and SVM. In Proceedings of the 2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring (IEEE Cat. No. 04CH37514), Milan, Italy, 17–19 May 2004 ; IEEE: Piscataway, NJ, USA, 2004; Volume 2, pp. 898–902.
2. Al-Nuaimi, D.H.; Hashim, I.A.; Zainal Abidin, I.S.; Salman, L.B.; Mat Isa, N.A. Performance of feature-based techniques for automatic digital modulation recognition and classification—A review. *Electronics* **2019**, *8*, 1407. [[CrossRef](#)]
3. Park, C.S.; Choi, J.H.; Nah, S.P.; Jang, W.; Kim, D.Y. Automatic modulation recognition of digital signals using wavelet features and SVM. In Proceedings of the 2008 10th International Conference on Advanced Communication Technology, Gangwon, Republic of Korea, 17–20 February 2008; IEEE: Piscataway, NJ, USA, 2008; Volume 1, pp. 387–390.
4. Fuxuan, C. High-Resolution SAR Image Classification Based on Deep Feature Extraction with Conditional Random Fields. Master’s Thesis, Xi’an University of Electronic Science and Technology, Xi’an, China, 2019.
5. Jie, Z. Research on Radar One-Dimensional Distance Image Recognition Algorithm Based on Self-Supervised Capsule Network. Master’s Thesis, Hangzhou University of Electronic Science and Technology, Hangzhou, China, 2022.
6. Yong, W.; Jinjun, W.; Zengshan, T.; Mu, Z.; Wang, S. Multidimensional Parametric Gesture Recognition Algorithm Based on FMCW Radar. *J. Electron. Inf. Sci.* **2019**, *41*, 822–829.
7. Zheng, C.; Wu, H.; Hao, Y.; Liu, Z. Radar radiation source identification based on deep learning of intervein features. *J. Signal Process.* **2020**, *36*, 8.
8. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]

9. Kong, M.; Zhang, J.; Liu, W.; Zhang, G. Radar emitter identification based on deep convolutional neural network. In Proceedings of the 2018 International Conference on Control, Automation and Information Sciences (ICCAIS), Hangzhou, China, 24–27 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 309–314.
10. Dudczyk, J. Radar emission sources identification based on hierarchical agglomerative clustering for large data sets. *J. Sensors* **2016**, *2016*, 1879327. [[CrossRef](#)]
11. Li, S.; Sun, J.; Guo, J.; Niu, Y.; Liu, B. Design of radar radiation source identification method based on Random Convolution Kernel and XGboost. In Proceedings of the International Conference on Electronic Materials and Information Engineering (EMIE 2023), Guangzhou, China, 14–16 July 2023; SPIE: Bellingham, DC, USA, 2023; Volume 12919, pp. 210–216.
12. Xiao, Z.; Yan, Z. Radar emitter identification based on feedforward neural networks. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; IEEE: Piscataway, NJ, USA, 2020; Volume 1, pp. 555–558.
13. Li, P. Research on radar signal recognition based on automatic machine learning. *Neural Comput. Appl.* **2020**, *32*, 1959–1969. [[CrossRef](#)]
14. Xiao, Y.; Zhang Wei, X. Specific emitter identification of radar based on one dimensional convolution neural network. *Proc. J. Phys. Conf. Ser.* **2020**, *1550*, 032114. [[CrossRef](#)]
15. Hao-nan, N.; Wen-can, W.; Qing-bo, L. Radar Radiation Source Identification Based on Convolution Neural Network. *Mod. Def. Technol.* **2021**, *49*, 130.
16. Qu, W.; Yao, G.; Meng, L. Research on Radar PRI Modulation Pattern Recognition Based on Recurrent Neural Network. In Proceedings of the 2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL), Zhuhai, China, 12–14 May 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 250–254.
17. Kishore, T.R.; Rao, K.D. Automatic intrapulse modulation classification of advanced LPI radar waveforms. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 901–914. [[CrossRef](#)]
18. Chenwei, D. Research on Individual Identification Technology of Radar Radiation Source Based on Deep Learning. Master's Thesis, Hangzhou University of Electronic Science and Technology, Hangzhou, China, 2020.
19. Kun, L.; Weigang, Z. A deep learning method for radar radiation source identification. *Electron. Des. Eng.* **2020**, *28*, 99–104.
20. Xinyue, W. Research and Application of Small-Sample Radar Radiation Source Signal Recognition Technology. Master's Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2021.
21. Jing, Z.; Li, P.; Wu, B.; Yuan, S.; Chen, Y. An adaptive focal loss function based on transfer learning for few-shot radar signal intra-pulse modulation classification. *Remote. Sens.* **2022**, *14*, 1950. [[CrossRef](#)]
22. Truong, T.; Yanushkevich, S. Generative adversarial network for radar signal synthesis. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–7.
23. Parnami, A.; Lee, M. Learning from few examples: A summary of approaches to few-shot learning. *arXiv* **2022**, arXiv:2203.04291.
24. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.
25. Basholli, F.; Daberdini, A. Influence of terrain and atmospheric conditions on the propagation of radio waves. In Proceedings of the 1st International Scientific Conference on Mathematics and Physics, and Their Applications (1st ISCMPA), ISCMPA University Elbasan “Aleksandër Xhuvani”, Elbasan, Albania, 3–4 November 2022.
26. Liu, C.; Liu, S.; Zhang, C.; Huang, Y.; Wang, H. Multipath propagation analysis and ghost target removal for FMCW automotive radars. In Proceedings of the IET International Radar Conference (IET IRC 2020), IET, Online, 4–6 November 2020; Volume 2020, pp. 330–334.
27. Zhang, J.A.; Liu, F.; Masouros, C.; Heath, R.W.; Feng, Z.; Zheng, L.; Petropulu, A. An overview of signal processing techniques for joint communication and radar sensing. *IEEE J. Sel. Top. Signal Process.* **2021**, *15*, 1295–1315. [[CrossRef](#)]
28. Ahmad, A.A.; Ahmad, S.A.; Muhammad, A.L. Recent developments in the use of time-frequency analysis for radar-based applications. In Proceedings of the AFRICON 2015, Addis Ababa, Ethiopia, 14–17 September 2015; pp. 1–5.
29. Huynh-The, T.; Doan, V.S.; Hua, C.H.; Pham, Q.V.; Nguyen, T.V.; Kim, D.S. Accurate LPI radar waveform recognition with CWD-TFA for deep convolutional network. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1638–1642. [[CrossRef](#)]
30. Wu, Z.L.; Huang, X.X.; Du, M.; Xu, X.S.; Bi, D.; Pan, J.F. Intra-Pulse Recognition of Radar Signals via Bicubic Interpolation WVD. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 8668–8680. [[CrossRef](#)]
31. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
33. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
34. Ye, Y.; Pan, C.; Wu, Y.; Wang, S.; Xia, Y. MFI-Net: Multiscale feature interaction network for retinal vessel segmentation. *IEEE J. Biomed. Health Informatics* **2022**, *26*, 4551–4562. [[CrossRef](#)] [[PubMed](#)]



35. Ruban, I.; Khudov, H.; Lishchenko, V.; Pukhovyi, O.; Popov, S.; Kolos, R.; Kravets, T.; Shamrai, N.; Solomonenko, Y.; Yuzova, I. Assessing the detection zones of radar stations with the additional use of radiation from external sources. *Boston-Eur. J. Technol.* **2020**, *6*, 6–17. [[CrossRef](#)]
36. Yoo, J.; Turnes, C.; Nakamura, E.B.; Le, C.K.; Becker, S.; Sovero, E.A.; Wakin, M.B.; Grant, M.C.; Romberg, J.; Emami-Neyestanak, A.; et al. A compressed sensing parameter extraction platform for radar pulse signal acquisition. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2012**, *2*, 626–638. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.