A Mutual-Attention Guided Feature Extraction and Adaptative Decision Fusion Framework for Fine-Grained Dual-Band Radar Target Classification

Ao Zhong, Shiguo Chen, *Member, IEEE*, Tianyu Wang, Yao Ma, and Yawei Zhang

***Abstract*— Fine-grained radar target classification based on single-band, such as wideband or narrowband, poses challenges even when utilizing deep learning methods. Since different bands reflect distinct characteristics of the targets, we focus on the fine-grained classification of radar aircraft benefits from dual-band data. However, the selection of complementary dual-band features and the fusion of decisions from two bands are the key issues that need to be addressed. In order to tackle these issues, we propose a framework for fine-grained radar target classification called mutual-attention guided feature extraction and adaptative decision fusion. In this article, we propose a mutual attention selection mechanism to explore the complementary feature information between wideband and narrowband data. Furthermore, the wideband and narrowband features make distinct contributions to determine the class types. In order to address the uncertainty associated with the contribution of the wideband and narrowband data, we propose a new adaptive decision fusion strategy that adaptively assigns different weights to model the contribution uncertainty. We conducted extensive experiments on a homebrew simulated dual-band fine-grained aircraft dataset, which includes the high resolution range profile signal and the jet engine modulation signal. Compared with other classification methods, the proposed approach exhibits a remarkable classification accuracy of 95.5% in our homebrew dataset and maintains an impressive accuracy of 87.4% even in challenging environments with a 5 dB SNRs. Moreover, it achieves exceptional inference speeds of up to 3073 data pairs per second on the GPU: RTX3090. The results demonstrate the robustness and efficiency of the proposed method.**

***Index Terms*—Adaptive decision fusion (DF), deep learning, dual-band, fine-grained classification, mutual-attention selection.**

I. Introduction

DUE to the good penetration and wide detection range of electromagnetic waves, radar signals possess the advantage of working all day in all weather conditions. Consequently, radar automatic target recognition (RATR) technology is extensively utilized in both military and civilian domains. Radar detects the distance, position, size, and other target information by transmitting electromagnetic waves and receiving the target echo. Depending on their bandwidth, radar echo signals can be divided into wideband and narrowband signals. Commonly used wideband signals include high-resolution range profile (HRRP) [1], [2], synthetic aperture radar (SAR) [3], [4], [5], and inverse synthetic aperture radar (ISAR) [6], [7], [8], etc. Narrow jet engine modulation (JEM) signals [9] can effectively capture the micromotion characteristics of long-range aircraft targets. This includes the periodic motion of the moving parts of the aircraft, such as propellers or rotating blades of jet engines, and rotor blades of helicopters [10]. HRRP not only provides information about the intensity distribution of the scatterer along the radar line of sight, but also reveals the size of the target from a specific radar perspective [11]. Numerous studies have concentrated on the coarse-level classification of long-range aircraft targets based on radar echo signals. This includes categorizing them by size (large, medium, small) [12] or power type (propeller, jet, helicopter, etc.) [13]. However, the fine-grained classification of long-range aircraft targets plays a crucial role in guiding battlefield situational estimation, assessing the threat level of aircraft targets, and evaluating the operational performance of aircraft targets. Fine-grained classification of radar data is highly challenging due to its sensitivity to radar data orientation, translation, and intensity. In addition, information obtained from a single radar band is limited in detecting long-range aircraft targets. There is limited research on the fine-grained classification of long-range aircraft targets based on radar echo signals. In this article, we use wideband HRRP and narrowband JEM signals for fine-grained classification of aircraft.

Traditional manually-designed features may not always provide enough discrimination for radar data in fine-grained aircraft classification tasks [14]. Most traditional radar target classification methods only use single-band radar data to extract manually-designed features with shallow classification networks. Commonly used features can be categorized into physical and transform domain features. The physical features include the location and magnitude of peak, the scattering center location, the intensity magnitude, the profile, the perimeter and

Manuscript received 31 December 2023; revised 7 February 2024; accepted 5 March 2024. Date of publication 12 March 2024; date of current version 10 April 2024. This work was supported by the National Natural Science Foundation of China, under Grant 62173265. *(Corresponding author: Shiguo Chen.)*

Ao Zhong, Tianyu Wang, Yao Ma, and Yawei Zhang are with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: aozhong@stu.xidian.edu.cn; tywang_456852@stu.xidian.edu.cn; mayao@stu.xidian.edu.cn; ywzhang_007@stu.xidian.edu.cn).

Shiguo Chen is with the School of Physics, Xidian University, Xi'an 710071, China (e-mail: sgchen@xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3375806

area of the target, the shape, the micromotion characteristics, etc. The transform domain features mean the features obtained by Fourier transform [15], [16], wavelet transform [17], [18], nonnegative matrix factorization [19], [20], radon transform [21], or sparse representation [22], [23],. Then, these features are classified by K-nearest neighbor classifier [24], Bayesian classifier [25], AdaBoosting Classifier [26], [27], support vector machine (SVM) [28], or hidden Markov model [29]. Deep learning has recently demonstrated strong capabilities in representing features, allowing for data-driven feature extraction. Commonly used deep architectures encompass various models, such as deep belief network [30], autoencoders (AE) [31], bidirectional recurrent neural networks (BiRNN) [32], and convolutional neural networks (CNN) [33].

The majority of current RATR methods rely on single-band signals, which provide limited information for fine-grained target classification. Fine-grained target classification methods using both wideband and narrowband signals have received limited attention in the existing literature. Moreover, multimodality classification involves feature selection, feature association, and decision fusion (DF). Distributed array radar captures data from multiple bands, including narrowband JEM, broadband HRRP, SAR, and ISAR, among others [34] [35]. The complementary nature of multimodal data can enhance the representation of target features [36], [37], thus, potentially improving the accuracy of classification algorithms. Studying the correlation among radar multimodal data is crucial for improving algorithm accuracy [38], [39], [40]. Multimodal data fusion can be broadly categorized into three types: signal fusion, feature fusion, and DF. Signal fusion involves concatenating and summing the original data from two modalities at the data level to obtain the fused signal. Feature fusion aims to extract features from different modalities and combine them using mutual attention mechanisms, concatenation, or dot product operations. DF aims to obtain fused classification results based on the classification results from each individual modality.

In this manuscript, we conduct a large number of ablation and comparison experiments on the simulated dataset, which verify the superiority of MAAF and the effectiveness of each module. In addition, to verify the robustness of the model [41], we add random Gaussian noise with different signal-to-noise ratios (SNRs) to the simulated data to evaluate the fine-grained recognition accuracy of MAAF under various SNRs conditions.

The main contribution of this article is that we propose a framework for fine-grained radar target classification called mutual-attention guided feature extraction and adaptative decision fusion (MAAF). Specifically, it lies in three folds.

1) Since the narrowband JEM and wideband HRRP echos reflect target properties from different aspects, we can extract complementary features of the target from them. However, the features from JEM and HRRP echos are not equally important to classification. Thus, there are strong and weak modalities in the process of fusion. To address this issue, we propose a novel mutual-attention selection mechanism, which enhances weak modality features while maintaining the effectiveness of the strong modality to improve the classification accuracy of radar target.
2) Further, we propose a novel adaptive DF strategy to model the imbalance of JEM and HRRP signals. It adaptively assigns different weights to the feature of each modality to estimate their uncertainty. This strategy enhances radar target classification accuracy by adjusting the impact of strong and weak modalities in DF.
3) We conduct extensive experiments on a custom simulated dual-band fine-grained aircraft dataset, which includes wideband signal HRRP and the narrowband signal JEM. The superior classification performance of our method compared with popular radar target classification methods showcases its effectiveness and robustness.

II. Related Works

A. Bimodal Fusion Classification for Remote Sensing

Though the research on crossmodal fine-grained aircraft classification methods are rare currently, the crossmodal methods have been explored in the context of remote sensing fields, such as ship identification, ground object classification, radar jamming signal classification, and so on. Zhang et al. [42] employed deep CNN features from SAR and integrated them with traditional HOG features to perform feature fusion, thereby enhancing the accuracy of ship classification in SAR images. Geng et al. [43] utilized dual-modal data from SAR images and multispectral images to perform feature fusion at conclusion of the network, leading to a significant enhancement in the accuracy of ground target classification. Shao et al. [44] initially preprocessed the radar jamming signal to acquire the filtered jamming signal and its corresponding time-frequency diagrams, which are referred to as two modal data. Subsequently, they extract depth features from these two modalities using a one-dimensional convolutional neural network (1-D-CNN) and a two-dimensional convolutional neural network (2-D-CNN). Finally, the depth features from both modalities are combined and input into the fully connected (FC) layer for radar jamming signal classification. Zhang et al. [45] introduced a feature fusion algorithm that combines electromagnetic scattering features with deep CNN features. This approach utilizes the extraction of scattering centers and deep learning to enhance SAR characterization, resulting in the improved classification accuracy. Sun et al. [46] presented a multifeature DF method for classifying aircraft targets using wideband and narrowband data. This method utilizes the traditional approach to classify aircraft based on wideband and narrowband signals. Subsequently, the classification results from different modalities are combined using Dempster–Shafer (DS) evidence theory. The aircraft type is determined through a voting process, which can significantly enhance classification accuracy compared with using single-modal data alone. These methods effectively showcase the benefits of incorporating multiple modal data in RATR. However, they do not emphasize feature fusion and DF, and the architectural designs of these methods still face limitations in extracting more robust features.

In this article, we propose a fine-grained classification algorithm for aircraft based on mutual attention in feature selection and adaptive wideband and narrowband DF for

classification, which achieves the complementary information of HRRP and JEM signals at both feature and decision levels.

B. HRRP Signal-Based Radar Object Classification

HRRP reflects the distribution and size of scattered intensity from target scatterers along the radar line-of-sight in a specific radar view. Neural networks have recently gained widespread usage in HRRP target classification. Fu et al. [47] demonstrated that CNN outperforms AE in HRRP target classification. Similarly, Ding and Chen [48] showed that CNN effectively extracts features from time-frequency maps of HRRP. Wan et al. [49] also computed the time-frequency maps from HRRP. They introduced an attention module that assigns different weights to time-frequency maps with varying resolutions. This approach led to a significant improvement in HRRP classification performance using CNN. Wan et al. [50] combined the strengths of CNN and BiRNN with an attention mechanism. They first utilized CNN to capture the spatial correlation within the original HRRP data. Subsequently, they employed BiRNN to fully account for the temporal dependencies among distance units. Finally, the attention mechanism directed the focus of the overall network toward the target region of interest. These methods effectively demonstrate the robust feature extraction capability of the CNN facilitated by the attention mechanism. However, their attention mechanism solely emphasizes the spatial positioning of features and does not account for the varying importance of different channels.

In this article, our HRRP feature extraction branch employs a CNN network with a channelwise attention mechanism to represent the weights of various channels in HRRP features. This approach effectively captures classification information that reflects the scale of the target.

C. JEM Signal-Based Radar Object Classification

JEM signal reflects the micromovement characteristics of aircraft through the periodic motion of its moving parts, such as propellers, rotating blades of jet engines, and rotor blades of helicopters. While few methods directly utilize JEM for aircraft target classification, it effectively captures the micromotion characteristics of targets, typically applied to the classification of small targets. Kim et al. [51] merged time domain and frequency domain maps of microDoppler signals. Subsequently, the drone classification CNN is trained using the combined maps, showcasing its robust classification performance. Meanwhile, Wang et al. [52] computed the time-frequency maps of radar echoes, capturing the micromotion characteristics of the target. The CNN classification is used to determine the micromotion category of the target. In a similar approach, Wang et al. [53] first computed the time-frequency maps that reflect the microDoppler of the target, and then employ CNN to classify the three motion states of the rotor target: hover, ascent, and descent. The existing research on JEM classification is limited, with most classifications being based on time-frequency transformation.

TABLE I
PARAMETERS OF THE FEATURE EXTRACTION NETWORK

Modules	Layers	Kernel	Channel	Stride	BN
JEMConv1	Conv	63	128	1	True
-	MaxPooling	3		3	-
JEMConv2	Conv	31	256	1	True
-	MaxPooling	2		2	-
JEMConv3	Conv	15	512	1	True
-	MaxPooling	2		2	-
JEMConv4	Conv	7	1024	1	True
-	MaxPooling	2		2	-
JEMConv5	Conv	3	1024	1	True
-	MaxPooling	2		2	-
HRRPConv1	Conv	31	128	1	True
-	MaxPooling	3		3	-
HRRPConv2	Conv	15	256	1	True
-	MaxPooling	2		2	-
HRRPConv3	Conv	7	512	1	True
-	MaxPooling	2		2	-
HRRPConv4	Conv	3	1024	1	True
-	MaxPooling	2		2	-
Dense FC	FC 4096-DropOut-4096-DropOut-1000				
Classification	FC 1000-7				

In this article, our JEM feature extraction branch utilizes a CNN network with an attention mechanism to effectively extract classification information that reflects the micromovement characteristics of the target.

III. METHODOLOGY

In order to fully exploit the relevance and complementary information from the features of HRRP and JEM, we explore an effective DF strategy for strong and weak modalities. This approach aims to enhance the fine-grained classification accuracy of aircraft targets. As a solution, we present a fine-grained classification network for aircraft that is based on mutual attention and adaptive wideband and narrowband fusion. The overall architecture is shown in Fig 1. The HRRP and JEM of each individual aircraft are input into a dual-branch network. In addition, we employ a channel mutual-attention mechanism to integrate the two modal data at the feature level. Subsequently, the outcomes from different branches are directed into the same classification layer to acquire evidence for each modality. These pieces of evidence are then input into the adaptive DF network to obtain the final evidence, which discriminates the target class based on its confidence.

A. Feature Selection

In this article, we propose a dual-branch CNN with multiscale channel mutual attention for feature extraction. Drawing inspiration from the squeeze-and-excitation networks (SE-Net) [54] module, we incorporate self-attention within each branch and mutual attention between the two branches. The structure is illustrated in Fig. 1, and the parameters of the dual-branch network are detailed in Table I. The CNN comprises three fundamental components: the convolutional layer, pooling layer, and FC

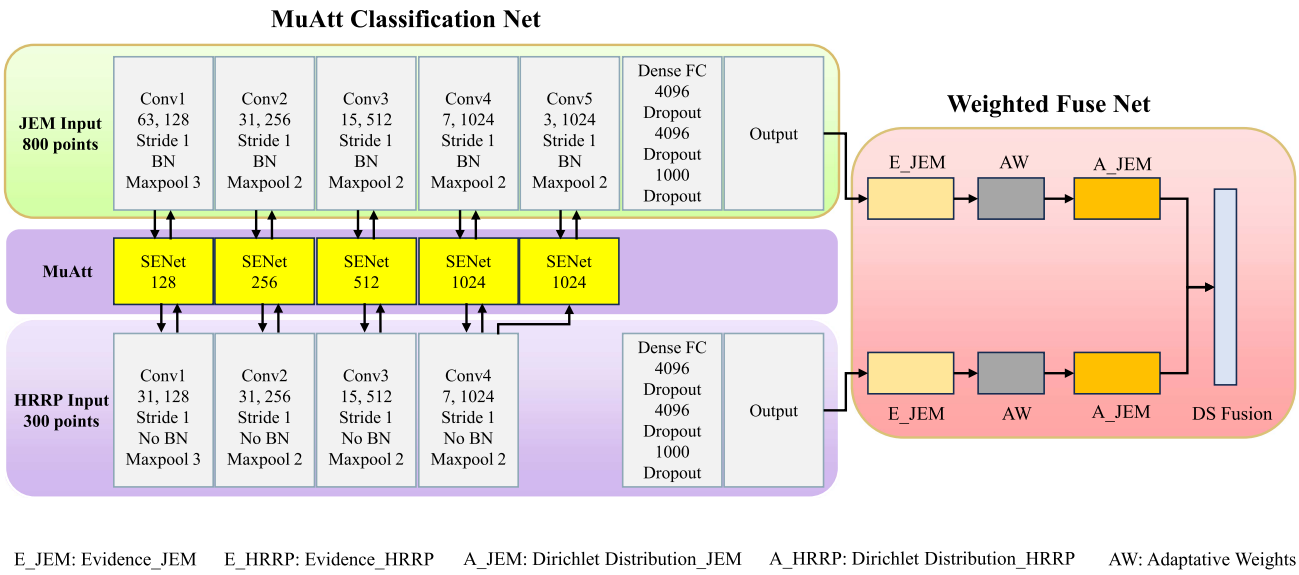

Fig. 1. Framework of the proposed MAAF framework for fine-grained dual-band radar target classification. It is comprised of the MuAtt classification net and the weighted fuse net. In the first stage, we trained the features of the classification network. In the second stage, we adaptively trained the weights corresponding to the HRRP and JEM.

layer. Specifically, the convolutional layer is employed to extract features, thereby augmenting the perceptual field of features. The pooling layer reduces the scale of the feature map, thereby increasing the sensitivity of the network to small changes in the data. Subsequently, the FC layer integrates all features learned by the network. In each convolutional layer of CNN, spatial and channel information can be combined within the local perceptual field. The number of convolutional kernels corresponds to the number of channels. Given that this article deals with 1-D data lacking spatial information, our primary emphasis lies in the channel information of HRRP and JEM across various CNN layers. SE-Net can effectively model the significance of various channels in the convolution process through compression and excitation, utilizing the channel information from the convolutional layer with minimal computational resources. The channelwise features are then rescaled based on attention weighting, thereby enhancing the feature representation. Since features extracted by convolution have multiple channel dimensions, SE-Net applies to the HRRP and JEM signals in this article.

The backbone employs a combination of 5+4 networks with dense FC. Due to the high number of 0 elements in the original HRRP data, the addition of a batch normalization (BN) layer would actually lead to a significant performance degradation of the network. Consequently, we only incorporate BN layers in the JEM branch to normalize various features and achieve optimal network performance. In the 5+4 CNN, the number of channels remains the same across different branches. A mutual attention mechanism is incorporated between different branches, allowing us to place the JEM features into the attention module in order to obtain their respective weights. Subsequently, these weights are added to the HRRP features, followed by inputting the HRRP features into the same attention module to obtain the weights of different channels. Similarly, the weights are added to the JEM features. This sequence of operations results in the implementation of a layer of self-attentive and mutual-attentive fusion within the network. As a result, we have constructed a feature extraction network based on multiscale channel mutual attention by incorporating mutual attention into each layer of the network.

B. Adaptive DF Network

HRRP data reflects the distribution and size of target scattering points, whereas JEM data reflects the micromovement characteristics of the target. The fusion of these two modal data has the potential to enhance the accuracy of fine-grained aircraft target classification. However, a challenge arises in the aircraft target fine-grained classification process due to the disparity in strength between the JEM and HRRP modalities. To address this issue, we propose a weighted adaptive DF strategy based on DS evidence theory to effectively combine the detection characteristics of two modalities, as inspired by the literature [55]. The flowchart of the DF process is depicted in Fig. 2.

Through the dual-branch CNN feature extraction network based on multiscale channel mutual attention, we can obtain the classification confidence of both HRRP and JEM branches, which is called HRRP and JEM classification evidence, denoted by $\bar{e}_K^h = [\bar{e}_1^h, \bar{e}_2^h, \ldots \bar{e}_K^h]$ and $\bar{e}_K^j = [\bar{e}_1^j, \bar{e}_2^j, \ldots \bar{e}_K^j]$, K represents the number of categories and the superscripts j and h represent the confidence parameters of JEM and HRRP data, respectively. After adding adaptive weights to the classification evidence $\bar{e}_K^h$ and $\bar{e}_K^j$, we can obtain the final evidence of HRRP and JEM.

$$e_K^j = w_j \cdot \bar{e}_K^j, \quad e_K^h = w_h \cdot \bar{e}_K^h \tag{1}$$

where w_j and w_h represent the weights of JEM and HRRP, respectively, and satisfy $w_j + w_h = 1$.

For each modal evidences $e_K = [e_1, e_2, \ldots e_K]$, the corresponding opinions $\alpha_K = [\alpha_1, \alpha_2, \ldots \alpha_K]$, belief quality b_K, Dirichlet strength S, and uncertainty u are inferred according to

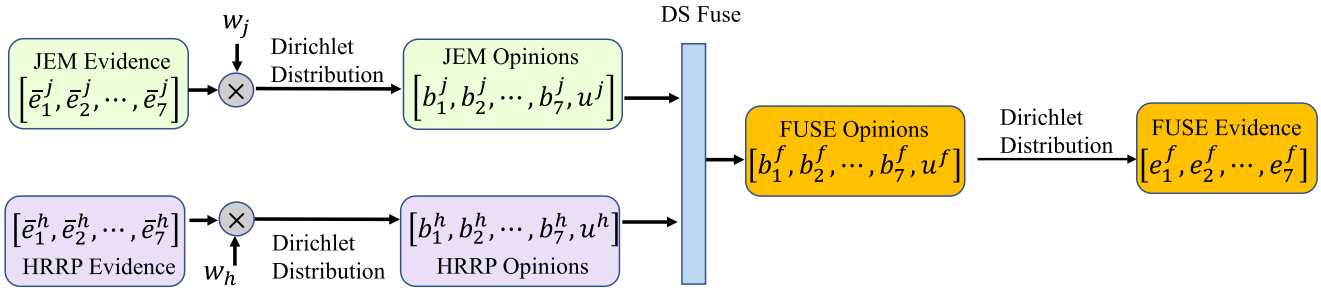

Fig. 2. Architecture of adaptive DF network.

the Dirichlet distribution, respectively.

$$\alpha_K = e_K + 1, \quad S = \sum_{i=1}^{K} \alpha_i, \quad b_K = \frac{e_K}{S}, \quad u = \frac{K}{S}. \tag{2}$$

Based on $e_K^j = [e_1^j, e_2^j, \ldots e_K^j]$ and $e_K^h = [e_1^h, e_2^h, \ldots e_K^h]$, we derive parameter groups $p^j = [\alpha_K^j, S^j, b_K^j, u^j]$ and $p^h = [\alpha_K^h, S^h, b_K^h, u^h]$ as determined by (2), for the two modalities of JEM and HRRP using the Dirichlet distribution. The fusion mass $M^f = \{\{b_k^f\}_{k=1}^K, u^f\}$ is calculated from the JEM and HRRP masses $M^j = \{\{b_k^j\}_{k=1}^K, u^j\}$ and $M^h = \{\{b_k^h\}_{k=1}^K, u^h\}$ in following manner:

$$M^f = M^j \oplus M^h \tag{3}$$

where $\oplus$ means DS fusion.The fused belief mass b_K^f with uncertainty u^f is calculated as follows:

$$b_K^f = \frac{1}{1-C}\left(b_K^j b_K^h + b_K^j u^h + b_K^h u^j\right), \quad u^f = \frac{1}{1-C} u^j u^h \tag{4}$$

where $C = \sum_{m \neq n} b_m^j b_n^h$ is a measure of the amount of conflict between the two modal belief mass, and $1/1-C$ is used as scale factors to normalize the belief quality and uncertainty after fusion.

After obtaining the fused belief quality b_K^f and uncertainty u^f, the fused Dirichlet strength S^f and the final fused evidence $e_K^f = [e_1^f, e_2^f, \ldots e_K^f]$ are calculated by (2). The final classification result is inferred by e_K^f.

C. Training Strategy and Loss Function

For traditional classification tasks, most methods usually use the softmax function in the last layer of the network to obtain the probability of the corresponding category p_{ij}. Then, the crossentropy loss function is calculated between p_{ij} and one-hot ground truth y_{ij} for back propagation, which is defined as follows:

$$L_{\text{ce}} = -\sum_{j=1}^{K} y_{ij} \log(p_{ij}). \tag{5}$$

It has been shown that using a softmax output as confidence often leads to high confidence values, even for erroneous predictions since the largest softmax output is used for the final prediction [56].

In addition, we use the parameters of Dirichlet distribution to measure the confidence of the category. Thus, we adjust the crossentropy loss based on the Dirichlet distribution as follows:

$$\begin{aligned} L_{\text{ace}} &= \int \left[\sum_{j=1}^{K} -y_{ij} \log(p_{ij})\right] \frac{1}{B(\alpha_i)} \prod_{j=1}^{K} {p_{ij}}^{\alpha_{ij}-1} dp_i \\ &= \sum_{j=1}^{K} y_{ij}(\psi(S_i) - \psi(\alpha_{ij})) \end{aligned} \tag{6}$$

where y denotes the ground truth label, $\psi(\bullet)$ is the digamma function, $B(\bullet)$ is the multinomial beta function, and the entire equation is the mean of the crossentropy loss determined by α_i alone. The above loss function ensures that the correct category for each sample generates large categorical evidence than the other categories, but does not suppress evidence generation for incorrect labels.

To ensure a more reasonable Dirichlet distribution and suppress the generation of mislabeled evidence, we introduce the KL loss function, which is formulated as follows:

$$\begin{aligned} \text{KL}\left[D(p_i|\tilde{\alpha}_i)||D(p_i|1)\right] = \log &\left(\frac{\Gamma\left(\sum_{k=1}^{K} \tilde{\alpha}_{ik}\right)}{\Gamma(K)\prod_{k=1}^{K} \tilde{\alpha}_{ik}}\right) \\ &+ \sum_{k=1}^{K} (\tilde{\alpha}_{ik} - 1)\left[\psi(\tilde{\alpha}_{ik})\right. \\ &\left. - \psi\left(\sum_{j=1}^{K} \tilde{\alpha}_{ij}\right)\right] \end{aligned} \tag{7}$$

where $\tilde{\alpha}_i = y_i + (1 - y_i) \odot \alpha_i$ is the parameter-adjusted Dirichlet distribution, which avoids penalizing the evidence of the ground truth class to 0, and $\Gamma(\bullet)$ is the gamma function. Therefore, the total loss of the Dirichlet distribution α_i is defined as follows:

$$L(\alpha_i) = L_{\text{ace}}(\alpha_i) + \lambda_t \text{KL}\left[D(p_i \mid \tilde{\alpha}_i) \| D(p_i \mid 1)\right] \tag{8}$$

where $\lambda_t > 0$ is a time-dependent factor that balances the expected classification error and KL regularization. In practice, we gradually increase the value of λ_t to prevent the network from focusing too much on the KL part at the initial stage of training, which may cause the network hard to converge.

In this article, we employ a two-stage training strategy. In the first stage, to ensure that each modality gets reasonable evidence, we first fix the weights w_h and w_j of both modalities to be 0.5, which can prevent excessive shifts of modal weights due to

different convergence rates of two-branch networks in the initial training stage.

The feature extraction network is trained first to ensure that the dual-branch feature extraction network can generate robust features. The loss function for training is defined as follows:

$$L_{\text{stage\_1}} = L({\alpha_i}^f) + \sum_{v=1}^{V} L({\alpha_i}^v) \tag{9}$$

where $L({\alpha_i}^f)$ denotes the loss function of the opinions after the fusion of JEM and HRRP, and $L({\alpha_i}^v)$ denotes the loss function of each modal opinions.

In the second stage, to solve the modal strength imbalance issue and achieve the adaptive weight (AW) change strategy of HRRP and JEM, we fix the parameters of the CNN feature extraction network. After training the two-branch feature extraction network and obtaining good features, the networks JEM and HRRP are separately trained based on the modal weights w_h and w_j, leading to further optimization of the overall network performance.

As the network ultimately utilizes the fused multiple opinions for prediction, stage 2 exclusively supervises the fused opinions, and the loss function is as follows:

$$L_{\text{stage\_2}} = L(\alpha^f) \tag{10}$$

where α^f is the fused multiple opinions after DF. By supervising the fusion results, the problem of strong and weak modal imbalance can be effectively solved.

In the first stage, we obtained reasonable evidence from each modality. Subsequently, in the second stage, we effectively addressed the issue of modal strength imbalance and implemented an AW change strategy for HRRP and JEM. Extensive experiments validate the effectiveness and superiority of our training strategy.

IV. Experiments and Analysis

A. Homebrew Simulated Dataset

We follow the dataset construction method of [57] to develop our homebrew multimode HRRP and JEM dataset, which is designed for radar target classification. The experimental data consist of HRRP and JEM data from a specific band of dual-base radar simulating seven types of aircraft targets, labeled P1 to P7, conducted by a research institute. There are 2520 HRRP and JEM data points for both radars. The data of the original radar for both modalities include 630 items, with 90 items for each aircraft class. The data of the other radar comprise 1890 items for both modalities, with 270 items for each aircraft class. The training dataset comprises 630 HRRP and JEM data points from the original radar, whereas the testing dataset includes 1890 HRRP and JEM data points from the other radar. During the data preprocessing stage, normalization is applied to both the HRRP and JEM data.

We initially disordered the HRRP and JEM signals of each aircraft class based on azimuth. Subsequently, we matched HRRP and JEM data from each aircraft class to create 90 pairs of training data, and the rest 270 pairs as testing data.

B. Ablation Experiments

We set HRRP+CNN and JEM+CNN as baseline. In addition, we incorporated the DF module, the AW adjustment module, and three mutual attention modules: the HRRP branch adds mutual attention to the JEM branch (H2JAtt), the JEM branch adds attention to the HRRP branch (J2HAtt), and both the HRRP and JEM branches add attention to each other (MuAtt). In our experiments, we concurrently classified and tested the HRRP evidence, JEM evidence, and FUSE evidence to determine their respective accuracy rates. The ablation analysis of MAAF is presented in Table II.

In Experiment 1, the classification accuracy of single-branch CNN achieved is 69.6% for HRRP and 90.3% for JEM modal. Compared with Experiment 1, Experiment 2 achieved 23.5% improvement in the accuracy of HRRP and 2.8% improvement in the accuracy of JEM, which indicates that the DF module can effectively integrate the classification evidence of the two modalities by modeling their uncertainties. Compared with the baseline result of single modality in Experiment 1, the classification accuracy of FUSE is significantly improved. However, in Experiment 2 FUSE result decreased by 0.7% compared with the result of strong modality JEM at 93.8%, which is caused by the imbalance between the strong and weak modalities. The FUSE result is dragged down by the HRRP modality. To address the unbalanced modality in DF, we propose the AW module, which can adaptively adjust the weights of HRRP and JEM in DF module by network learning, so as to achieve better classification results. Experiment 3 demonstrates the effectiveness of AW module. Compared with Experiment 2, the FUSE accuracy can be improved by 1.2% by adjusting the weights of HRRP and JEM. Furthermore, the FUSE outcome surpasses the accuracy of the strong modality JEM. Thus, the proposed AW module effectively resolves the imbalance issue between the strong and weak modalities.

In addition, we investigate the feature correlation between HRRP and JEM using various feature interactions. Experiments 4, 6, and 8 confirmed the impact of H2JAtt, J2HAtt, and MuAtt mechanisms on the classification results, respectively. Based on the results of Experiment 2, the H2JAtt module can enhance JEM accuracy by 1.0% and FUSE accuracy by 1.0%. The J2HAtt module enhances HRRP accuracy by 35.3% and FUSE accuracy by 1.2%. The MuAtt module achieves accuracy improvements of 36.7%, 1.1%, and 2.2% on HRRP, JEM, and FUSE, respectively. The aforementioned experiments demonstrate that the H2JAtt module utilizes HRRP information to enhance JEM learning. The J2HAtt module leverages JEM information to improve HRRP learning. The MuAtt module establishes feature-level connections between HRRP and JEM, resulting in effective learning improvements for both modalities. All those three attention mechanisms can improve FUSE classification results. MuAtt is the only attention module among the three that enhances the classification accuracy of HRRP, JEM, and FUSE. It improves both weak modality HRRP and strong modality JEM. As a result, we select MuAtt as our final attention mechanism, which combined with the AW module, contributes to an additional 0.2% improvement in accuracy. As a result, the final MAAF achieves an amazing classification accuracy of 95.5%.

TABLE II
ABLATION EXPERIMENTS

Exp.	Baseline	DF	AW	H2JAtt	J2HAtt	MuAtt	HRRP Acc(%)	JEM Acc(%)	FUSE Acc(%)	Weights(HRRP, JEM)
1	√						69.6	90.3	-	-
2	√	√					56.7	93.8	93.1	(0.500, 0.500)
3	√	√	√				56.7	93.8	94.3	(0.174, 0.826)
4	√	√		√			55.0	94.8	94.1	(0.500, 0.500)
5	√	√	√	√			55.0	94.8	95.0	(0.087, 0.913)
6	√	√			√		92.0	93.5	94.3	(0.500, 0.500)
7	√	√	√		√		92.0	93.5	94.3	(0.591, 0.409)
8	√	√				√	93.4	94.9	95.3	(0.500, 0.500)
9	√	√	√			√	**93.4**	**94.9**	**95.5**	(0.267, 0.733)

Bold indicates the highest accuracy.

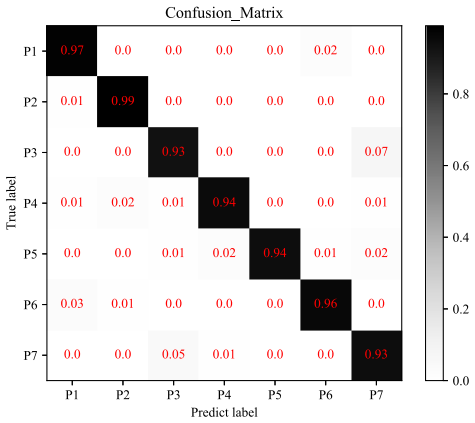

Fig. 3. Confusion matrix of MAAF.

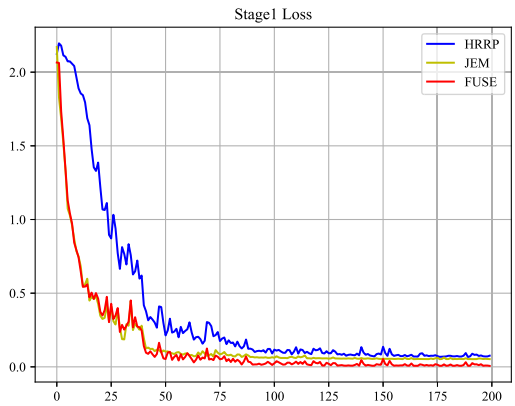

Fig. 4. Loss curves of MAAF in Stage 1.

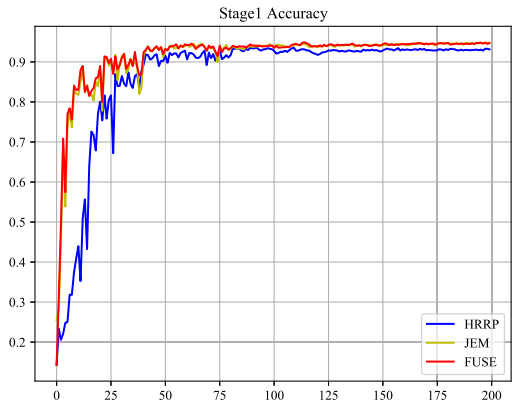

Fig. 5. Accuracy curves of MAAF in Stage 1.

From above analysis, we obtain the following conclusions.

1) The DF module enhances network learning by modeling modality uncertainty, surpassing the baseline with single branch. However, it is hard for DF to handle the unbalanced modality.
2) The AW module adjusts the modal weights adaptively in the DF module through network learning even when it confronts the unbalanced modalities. It does not alter the parameters in the backbone network.
3) The H2JAtt module leverages HRRP information to enhance the learning on JEM modal, whereas the J2HAtt module utilizes JEM information to improve the learning on HRRP modal.

The MuAtt module establishes connections between HRRP and JEM features, effectively enhancing the learning of both HRRP and JEM modals. It helps the model to extract complementary features even when it confronts with the unbalanced modalities.

C. Convergence and Robustness Experiments and Analysis

The confusion matrix clearly displays the classification accuracy and misclassification of each aircraft type. Fig. 3 illustrates the confusion matrix generated by our method.

The classification accuracy for each aircraft category exceeds 90%, demonstrating the strong performance of our method in classifying the seven aircraft types. We present the training and testing loss and accuracy for each iteration, and depict the convergence loss curve and variation accuracy curve in Figs. 4 and 5.

Fig. 4 illustrates the gradual convergence of training loss for HRRP, JEM, and FUSE, reaching its optimal value at 50 epochs. Furthermore, Fig. 5 indicates that the accuracy of HRRP, JEM, and FUSE shows no significant increase after 50 epochs, indicating the convergence and stability of our model after approximately 50 epochs.

To assess noise immunity and robustness, random Gaussian noise with average SNRs of 5, 10, 15, and 20 dB is added to both

TABLE III
RESULTS OF MAAF ON DATA WITH DIFFERENT SNRS

SNRs(dB)	Accuracy(%)
5	87.4
10	91.5
15	93.5
20	94.7

TABLE IV
COMPARATIVE EXPERIMENTS ON RADAR TARGET CLASSIFICATION

Modality	Method	Accuracy(%)
HRRP	LDA+SVM	35.8
	PCA+SVM	40.7
	RNN	46.5
	LSTM	48.5
	BiGRU	52.2
	CNN	69.6
JEM	LDA+SVM	18.2
	PCA+SVM	53.1
	RNN	43.2
	LSTM	53.0
	BiGRU	66.8
	CNN	90.3
FUSE	DS Fusion	93.1
	MAAF	**95.5**

Bold indicates the highest accuracy.

the training and testing sets to simulate the classification accuracy and robustness of our method in a real-world environment. The experimental results are presented in Table III.

The experiments demonstrate that our algorithm maintains strong performance at SNRs of 15 and 20 dB, whereas exhibiting a slight decrease at SNRs of 5 and 10 dB. This provides full verification of the noise immunity and robustness of our method. Besides, the model has 121.857 M parameters and 0.817 G FLOPs. On RTX3090, the recognition speed reaches 3073 items per second. The performance meets the requirements for radar target recognition in real-world scenarios.

D. Comparative Experiments

To comprehensively evaluate the performance of our algorithm, we conducted comparative experiments with several classical algorithms, including LDA+SVM [58], PCA+SVM [59], RNN [60], LSTM [61], BiGRU [32], and CNN [62]. The traditional algorithms PCA and LDA are used to reduce the dimensionality of features and then classify with them using SVM. For PCA, we found through experiments that for HRRP and JEM modals, the best results are achieved when the cumulative contribution degree reaches 80% and 77% in dimensionality reduction. The deep learning algorithms RNN, LSTM, and Bi-GRU take raw data as input into the network after applying a sliding window. The FUSE in the table refers to fusion algorithm that utilizes CNN on HRRP and JEM signals. The experiments were conducted separately on HRRP and JEM signals.

The CNN algorithm used in this article achieves the best classification accuracy for both HRRP and JEM modalities on the homebrew dataset. MAAF demonstrates significantly superior performance compared with other advanced classification methods. Specifically, on our homebrew multimodal simulated dataset, the classification accuracy of MAAF reaches 95.5%, representing 35.9% and 5.2% improvements over the baseline. This outstanding performance and substantial improvement proves the effectiveness of our proposed framework.

V. CONCLUSION

In this article, we propose a fine-grained classification algorithm for aircraft based on mutual attention and adaptive wideband and narrowband fusion called MAAF. We employ a novel dual-branch CNN network structure for feature extraction, integrating a mutual channel attention mechanism that supports self-supervision and mutual supervision at the feature level. By employing adaptive weighting decision fusion, MAAF addresses the challenge of strong and weak modal imbalance, further optimizing the network and enhancing its performance. Our algorithm fully leverages the inherent relationship between HRRP and JEM, effectively utilizing information from these modalities to significantly enhance the classification performance of radar aircraft targets compared with traditional methods and single-modal CNNs. In addition, our algorithm demonstrates strong robustness and maintains high performance even at low SNR. The nonaligned azimuth of the HRRP and JEM data pairs used in this study enhances their compatibility with real-world radar target identification scenarios. We plan to conduct future research on real-world data as in [63], which focuses on the issues such as uneven data quality and cross-scene challenges. Drawing from the spectral foundation large model described in [63], we aim to develop our own foundational large model for RATR to address the challenges encountered in real-world data.

REFERENCES

[1] Y. Xu et al., "A contrastive-based adversarial training algorithm for HRRP target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 3507105.
[2] Y. Song, Q. Zhou, W. Yang, Y. Wang, C. Hu, and X. Hu, "Multi-view HRRP generation with aspect-directed attention GAN," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7643–7656, 2022.
[3] X. Ma, K. Ji, L. Zhang, S. Feng, B. Xiong, and G. Kuang, "SAR target open-set recognition based on joint training of class-specific subdictionary learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 3500805.
[4] X. Zhou, C. Luo, P. Ren, and B. Zhang, "Multiscale complex-valued feature attention convolutional neural network for SAR automatic target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 2052–2066, 2024.
[5] H. Chen et al., "A multi-input channel U-Net landslide detection method fusing SAR multisource remote sensing data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1215–1232, 2024.
[6] S. Shao, H. Liu, and J. Wei, "GEO targets ISAR imaging with joint intra-pulse and inter-pulse high-order motion compensation and sub-aperture image fusion at ULCPI," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5200515.
[7] F. Santi, I. Pisciottano, D. Cristallini, and D. Pastina, "Impact of rotational motion estimation errors on passive bistatic ISAR imaging via back-projection algorithm," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 3345–3365, 2024.
[8] Z. Yang, X. Tan, W. Tian, X. Dong, and C. Cui, "ISAR imaging for NONCO-operative targets based on sharpness criterion under low SNR," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7690–7703, 2023.

[9] A. L. Lee, P. Ly, M. Jackson, E. S.-Pierre, P. P. T. Ho, and D. Wilson, "Jet engine modulation recognition with deep learning," in *Proc. IEEE Conf. Artif. Intell.*, 2023, pp. 1–4.

[10] J. Sun and J. Yu, "Robust and fast image processing method for feature extraction of jet engine modulation signals," in *Proc. IET Int. Radar Conf.*, 2020, vol. 2020, pp. 62–67.

[11] Y.-P. Zhang, L. Zhang, L. Kang, H. Wang, Y. Luo, and Q. Zhang, "Space target classification with corrupted HRRP sequences based on temporal–spatial feature aggregation network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5100618.

[12] H. Zhang, W. Wang, W. Jiang, Z. Chen, and Z. Zhuang, "Aircraft target classification based on registration information for low-resolution radar," *Syst. Eng. Electron.*, vol. 26, no. 4, pp. 480–490, 2004.

[13] M. Li, J. Wu, L. Zuo, W. Song, and H. Liu, "Aircraft target classification and recognition algorithm based on measured data," *J. Electron. Inf. Technol.*, vol. 40, no. 11, pp. 2606–2613, 2018.

[14] J. Wan, B. Chen, B. Xu, H. Liu, and L. Jin, "Convolutional neural networks for radar HRRP target recognition and rejection," *EURASIP J. Adv. Signal Process.*, no. 5, pp. 1–17, 2019.

[15] P. Alexey, P. Olga, and S. Natalia, "2D discrete fast Fourier transform with variable parameters," in *Proc. 24th Int. Conf. Digit. Signal Process. Appl.*, 2022, pp. 1–8.

[16] H. Zhang, T. Shan, S. Liu, and R. Tao, "Parameter optimization of sparse Fourier transform for radar target detection," in *Proc. IEEE Radar Conf.*, 2020, pp. 1–5.

[17] X. Liu and W. Fa, "A fully automatic algorithm for reflector detection in radargrams based on continuous wavelet transform and minimum spanning tree," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4601620.

[18] A. Kenge, M. Panse, S. Choudhari, H. Pethkar, and R. Pinto, "Wavelet transform based compression of raw SAR data," in *Proc. 12th Int. Conf. Comput. Commun. Netw. Technol.*, 2021, pp. 1–5.

[19] Z. Yang, N. Liang, W. Yan, Z. Li, and S. Xie, "Uniform distribution non-negative matrix factorization for multiview clustering," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3249–3262, Jun. 2021.

[20] S. Şahin and A. K. Hocaoğlu, "Non-negative matrix factorization method for ground penetrating radar images," in *Proc. 27th Signal Process. Commun. Appl. Conf.*, 2019, pp. 1–4.

[21] H. Chang, R. Smith, S. Paisey, R. Boutchko, and D. Mitra, "Deep learning image transformation under radon transform," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, 2020, pp. 1–3.

[22] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, Jan. 2016.

[23] P. Wang, Y. Meng, W. Zhu, X. Song, J. Ding, and H. Liu, "Robust variant target recognition based on structured sparse representation for radar HRRP data," in *Proc. Int. Radar Conf.*, 2019, pp. 1–5.

[24] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[25] J. Tao, X. Yue, K. Tang, Y. Yan, and Y. Wang, "Identification of defects in ultrasonic inspection of overhead distribution lines based on Bayesian classification algorithm," in *Proc. Int. Conf. Adv. Elect. Equip. Reliable Operation*, 2021, pp. 1–5.

[26] P. Wang, W. Qu, H. Jiang, Y. Dong, and T. Gao, "A fusion classifier algorithm based on Adaboost," in *Proc. 2nd Int. Seminar Artif. Intell., Netw. Inf. Technol.*, 2021, pp. 47–51.

[27] X. Hu, P. Zhang, and Q. Zhang, "A novel framework of CNN integrated with Adaboost for remote sensing scene classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2643–2646.

[28] V. Cherkassky, "The nature of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 8, no. 6, pp. 1564–1564, Nov. 1997.

[29] J. Chen, L. Du, Y. Li, and Y. Liu, "Hybrid hidden Markov modeling for target recognition based on high-resolution radar signal," in *Proc. IET Int. Radar Conf.*, 2020, vol. 2020, pp. 1606–1610.

[30] C. Qin, X. Song, and H. Chen, "Radar behavior classification based on DBN," in *Proc. IEEE 9th Joint Int. Inf. Technol. Artif. Intell. Conf.*, 2020, vol. 9, pp. 1169–1173.

[31] N. Ferreira and M. Silveira, "Ship detection in SAR images using convolutional variational autoencoders," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2503–2506.

[32] J. Liu, B. Chen, and X. Jie, "Radar high-resolution range profile target recognition based on attention mechanism and bidirectional gated recurrent," *J. Radars*, vol. 8, no. 5, pp. 589–597, 2019.

[33] H. Bi, J. Deng, T. Yang, J. Wang, and L. Wang, "CNN-based target detection and classification when sparse SAR image dataset is available," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6815–6826, 2021.

[34] S. Feng, K. Ji, F. Wang, L. Zhang, X. Ma, and G. Kuang, "Electromagnetic scattering feature (ESF) module embedded network based on ASC model for robust and interpretable SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5235415.

[35] S. Feng, K. Ji, L. Zhang, X. Ma, and G. Kuang, "SAR target classification based on integration of ASC parts model and deep learning algorithm," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10 213–10 225, 2021.

[36] H. Zhou et al., "Position-aware relation learning for RGB-Thermal salient object detection," *IEEE Trans. Image Process.*, vol. 32, pp. 2593–2607, 2023.

[37] H. Zhou, C. Tian, Z. Zhang, C. Li, Y. Xie, and Z. Li, "Frequency-aware feature aggregation network with dual-task consistency for RGB-T salient object detection," *Pattern Recognit.*, vol. 146, 2023, Art. no. 110043.

[38] C. Li, L. Du, Y. Li, and J. Song, "A novel SAR target recognition method combining electromagnetic scattering information and GCN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4508705.

[39] X. Zhang, S. Feng, C. Zhao, Z. Sun, S. Zhang, and K. Ji, "MGSFA-Net: Multi-scale global scattering feature association network for SAR ship target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 4611–4625, 2024.

[40] S. Feng, K. Ji, F. Wang, L. Zhang, X. Ma, and G. Kuang, "PAN: Part attention network integrating electromagnetic characteristics for interpretable SAR vehicle target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5204617.

[41] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[42] T. Zhang, X. Zhang, J. Shi, and S. Wei, "A hog feature fusion method to improve CNN-based SAR ship classification accuracy," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 5311–5314.

[43] J. Geng, H. Wang, J. Fan, and X. Ma, "Classification of fusing SAR and multispectral image via deep bimodal autoencoders," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 823–826.

[44] G. Shao, Y. Chen, and Y. Wei, "Deep fusion for radar jamming signal classification based on CNN," *IEEE Access*, vol. 8, pp. 117 236–117 244, 2020.

[45] J. Zhang, M. Xing, and Y. Xie, "FEC: A feature fusion framework for SAR target recognition based on electromagnetic scattering features and deep CNN features," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2174–2187, Mar. 2021.

[46] J. Sun, Y. Chen, C. Li, and S. Xie, "Fusion recognition of air target based on DS evidence theory with wideband and narrowband characteristics," *Modern Radar*, vol. 37, no. 7, pp. 15–19, 2015.

[47] Z. Fu, S. Li, X. Li, B. Dan, and X. Wang, "A performance analysis of neural network models in HRRP target recognition," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process.*, 2019, pp. 1–4.

[48] B. Ding and P. Chen, "HRRP feature extraction and recognition method of radar ground target using convolutional neural network," in *Proc. Int. Conf. Electromagn. Adv. Appl.*, 2019, pp. 0658–0661.

[49] J. Wan, B. Chen, Y. Yuan, H. Liu, and L. Jin, "Radar HRRP recognition using attentional CNN with multi-resolution spectrograms," in *Proc. Int. Radar Conf.*, 2019, pp. 1–4.

[50] J. Wan, B. Chen, Y. Liu, Y. Yuan, H. Liu, and L. Jin, "Recognizing the HRRP by combining CNN and BiRNN with attention mechanism," *IEEE Access*, vol. 8, pp. 20 828–20 837, 2020.

[51] B. K. Kim, H.-S. Kang, and S.-O. Park, "Drone classification using convolutional neural networks with merged Doppler images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 38–42, Jan. 2017.

[52] Y. Wang, C. Feng, Y. Zhang, and Q. Ge, "Classification of space targets with micro-motion based on deep CNN," in *Proc. IEEE 2nd Int. Conf. Electron. Inf. Commun. Technol.*, 2019, pp. 557–561.

[53] W. Wang et al., "Motion states classification of rotor target based on micro-Doppler features using CNN," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1390–1393.

[54] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[55] Z. Han, C. Zhang, H. Fu, and J. T. Zhou, "Trusted multi-view classification with dynamic evidential fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2551–2566, Feb. 2023.

[56] J. Moon, J. Kim, Y. Shin, and S. Hwang, "Confidence-aware learning for deep neural networks," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, vol. 119, pp. 7034–7044.

[57] D. Hong et al., "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sens. Environ.*, vol. 299, 2023, Art. no. 113856.
[58] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data–with application to face recognition," *Pattern Recognit.*, vol. 34, no. 10, pp. 2067–2070, 2001.
[59] L. Du, H. Liu, Z. Bao, and J. Zhang, "Radar automatic target recognition using complex high-resolution range profiles," *IET Radar, Sonar Navigation*, vol. 1, no. 1, pp. 18–26, 2007.
[60] W. Chen et al., "Tensor RNN with Bayesian nonparametric mixture for radar HRRP modeling and target recognition," *IEEE Trans. Signal Process.*, vol. 69, pp. 1995–2009, 2021.
[61] L. Sun, J. Liu, Y. Liu, and B. Li, "HRRP target recognition based on soft-boundary deep SVDD with LSTM," in *Proc. Int. Conf. Control, Automat. Inf. Sci.*, 2021, pp. 1047–1052.
[62] J. Wan, B. Chen, B. Xu, H. Liu, and L. Jin, "Convolutional neural networks for radar HRRP target recognition and rejection," *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, pp. 1–17, 2019.
[63] D. Hong et al., "SpectralGPT: Spectral foundation model," 2023, *arXiv:2311.07113*.

Ao Zhong received the B.S. degree in electronic information engineering in 2021 from Xidian University, Xi'an, China, where he is working toward the M.S. degree in electronic science and technology.

His research interests include multimodal data classification, object tracking, and deep learning.

Shiguo Chen (Member, IEEE) received the B.S. degree in mechanical design, manufacturing, and automation from Xidian University, Xi'an, China, in 2004, and the M.S. degree in mechanical engineering from the National University of Defense Technology, Changsha, China, in 2006, and the Ph.D. degree in aeronautical and astronautical science and technology from Air Force Engineering University, Xi'an, in 2015.

He is currently a Professor with School of Physics, Xidian University. He is a leading Professor and Doctoral Supervisor of Huashan Scholars with Xidian University of Electronic Science and Technology. His research interests include infrared scene simulation and intelligent learning algorithm.

Tianyu Wang received the B.S. degree in communications engineering from Nanjing University of Information Science and Technology, Nanjing, China, in 2022. He is working toward the M.S. degree in electronic science and technology with Xidian University, Xi'an, China.

His research interests include incremental learning, radar automatic target recognition, and deep learning.

Yao Ma received the B.S. degree in electronic information engineering in 2021 from Xidian University, Xi'an, China, where she is working toward the M.S. degree in electronic science and technology.

Her research interests include radar automatic target recognition, incremental learning, and deep learning.

Yawei Zhang received the B.S. degree in electronic information engineering in 2021 from Xidian University, Xi'an, China, where he is working toward the M.S. degree in electronic science and technology.

His research interests include object detection, transfer learning, and deep learning.