

## Article

# TSAR-Net: Transfer-Enhanced Local-to-Global Representation Learning for Micro-Doppler Classification of Small Aerial Targets

Yuxian Sun, Yuanshuai Li, Jiaxiang Zhang, Huayu Fan and Yuanyuan Song \*

Radar Research Lab, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China; 3120235605@bit.edu.cn (Y.S.); 3120235604@bit.edu.cn (Y.L.); 3220215121@bit.edu.cn (J.Z.); fanhuayu@bit.edu.cn (H.F.)

\* Correspondence: beral@bit.edu.cn

## Abstract

Radar remote sensing provides an effective means of analyzing target characteristics for low-altitude surveillance, and the rapid growth and misuse of drones highlight the need for efficient classification of small aerial targets such as drones and birds. Micro-motions, such as rotor rotation or wing flapping, induce characteristic frequency shifts in radar echoes known as micro-Doppler effects. These radar-derived features provide essential structural and kinematic information for deep-learning-based target classification. However, in practical applications, existing methods still suffer from inadequate micro-Doppler feature extraction capability and rarely account for small-sample scenarios, both of which limit their classification performance. To address these challenges, this article proposes a transfer learning and self-attention enhanced ResNet50 (TSAR-Net), which combines transferred hierarchical convolutional features with multi-head self-attention to jointly exploit local and global time–frequency characteristics. Furthermore, to alleviate the small-sample classification problem, we introduce a projection-based TSAR-Net (PTSAR-Net), which maps the 2048-dimensional pooled feature through a 512-dimensional layer to a compact 128-dimensional representation for limited-label classification. On the measured DIAT- $\mu$ SAT dataset, TSAR-Net achieves 99.59% accuracy, compared with 96.71% for VGG19, an improvement of 2.88 percentage points. With five training samples per class, PTSAR-Net achieves 81.14% accuracy, compared with 70.72% for TSAR-Net, an improvement of 10.42 percentage points.

**Keywords:** micro-Doppler; small aerial target; radar target classification; representation learning; transfer learning; limited-label classification



Academic Editor: Dusan Gleich

Received: 29 August 2026

Revised: 28 September 2026

Accepted: 5 October 2026

Published: 8 October 2026

**Copyright:** © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

## 1. Introduction

Micro-motions refer to minor movements of a target or its components beyond overall translational motion, such as rotations, vibrations, or oscillations. These features can be detected and analyzed using advanced sensors and techniques capable of capturing subtle changes in the target's motion. Among these technologies, radar systems are particularly notable for their all-time, all-weather, and long-range detection capabilities, making them a focal point of research in micro-motion feature perception and extraction [1,2]. Related micro-Doppler signatures have also been investigated using other sensing modalities [3]. In radar remote sensing, deep-learning-based interpretation of micro-Doppler information provides a useful route for analyzing target characteristics in low-altitude surveillance.

Micro-motions induce frequency modulation in radar echoes, generating sidebands around the central Doppler frequency caused by the target's translational motion. This phenomenon, known as the micro-Doppler effect, has been extensively investigated [4–10]. By extracting micro-Doppler features from radar echoes, researchers can capture the structural and kinematic information of targets, enabling improved target identification and classification [11–16].

With advancements in air target recognition technologies, micro-Doppler feature-based classification has become a key focus of radar-based research. Air targets often exhibit distinctive micro-Doppler signatures in their motion characteristics, such as the wing flapping of birds and the rotor blades of drones [17,18]. Accurate classification based on these features is critical for improving radar identification accuracy and enhancing the reliability of air target recognition systems. To address these challenges, various approaches have been explored, including empirical mode decomposition and synchrosqueezing phase analysis for feature extraction and classification [19,20], as well as the integration of micro-Doppler features with convolutional neural networks (CNNs) or dictionary learning for multi-target scenarios [21,22]. Techniques such as the short-time Fourier transform (STFT), principal component analysis (PCA), and long short-term memory (LSTM) neural networks have also been employed for micro-motion classification [23]. Related CNN-based studies have examined classification of similar micromotion targets under low-SNR conditions [24]. Furthermore, transfer learning with VGG19 has been applied for multi-target classification tasks [25,26], which also led to the release of the DIAT- $\mu$ SAT dataset, the first publicly available micro-Doppler dataset dedicated to small aerial targets. Recent studies have also investigated enhanced Doppler spectrograms and cascaded denoising-classification pipelines for radar target recognition [16,27]. Despite these advancements, two challenges remain. Conventional CNNs mainly aggregate local spatial information through finite receptive fields, which may limit explicit modeling of long-range time–frequency dependencies. In addition, limited labeled samples can constrain classifier training.

The small-sample problem is a critical factor limiting the practical application of radar-based air micro-motion target classification. Due to the complexity of target types, acquiring large amounts of sample data for various targets incurs high costs. Furthermore, obtaining micro-Doppler features requires high-precision radar observations under controlled experimental conditions, imposing stringent requirements on experimental equipment and environments. Most existing studies primarily depend on abundant sample data for feature extraction and classifier design [28–30]. Recent low-altitude radar work has also investigated target recognition with limited labeled samples [31]. Therefore, there is a need to explore efficient classification algorithms tailored to small sample sizes, enabling full utilization of micro-Doppler features from aerial targets under such conditions.

To address these challenges, this article proposes a transfer learning and self-attention enhanced ResNet50 (TSAR-Net). The motivation for adopting transfer learning stems from the difficulty and high cost of collecting large-scale labeled radar datasets. Transfer learning uses ImageNet-pretrained ResNet50 weights to improve training efficiency, while multi-head self-attention (MHSA) models dependencies among spatially separated time–frequency features. TSAR-Net therefore combines transferred hierarchical convolutional features with global dependency modeling to form a transfer-enhanced local-to-global representation framework. To further improve performance in small-sample scenarios, the projection-based TSAR-Net (PTSAR-Net) maps the 2048-dimensional pooled feature through a 512-dimensional layer to a compact 128-dimensional embedding before classification.

The contributions of this article are summarized as follows.

- (1) A transfer-enhanced local-to-global representation framework, TSAR-Net, is proposed for micro-Doppler classification of small aerial targets. It combines transfer learning for hierarchical convolutional representation with MHSA for global dependency modeling.
- (2) PTSAR-Net introduces a compact representation strategy for limited-label classification by mapping the pooled feature to a lower-dimensional embedding before classification.
- (3) The proposed methods are validated on the measured DIAT- $\mu$ SAT dataset under conventional supervision and with 5, 10, 20, and 50 training samples per class.

Beyond micro-Doppler signatures, full-polarimetric information may provide complementary cues on target scattering characteristics for radar target recognition [32,33]. Future extensions could explore its joint use with micro-Doppler representations to characterize both target motion and scattering properties.

The remainder of this article is organized as follows. Section 2 describes the radar signal model, the DIAT- $\mu$ SAT dataset, the TSAR-Net and PTSAR-Net architectures, and the experimental protocol. Section 3 presents the measured data verification and analysis. Section 4 discusses the main findings and future directions, and Section 5 concludes the article.

## 2. Dataset and Methods

### 2.1. Micro-Doppler Signal Model and Time–Frequency Representation

To illustrate the translational and rotational contributions to the radar echo, this subsection considers a rotating blade on a translating platform, as illustrated in Figure 1. The platform excluding the blade is represented by scattering point  $A$ , about which the blade rotates with angular velocity  $\omega$ .

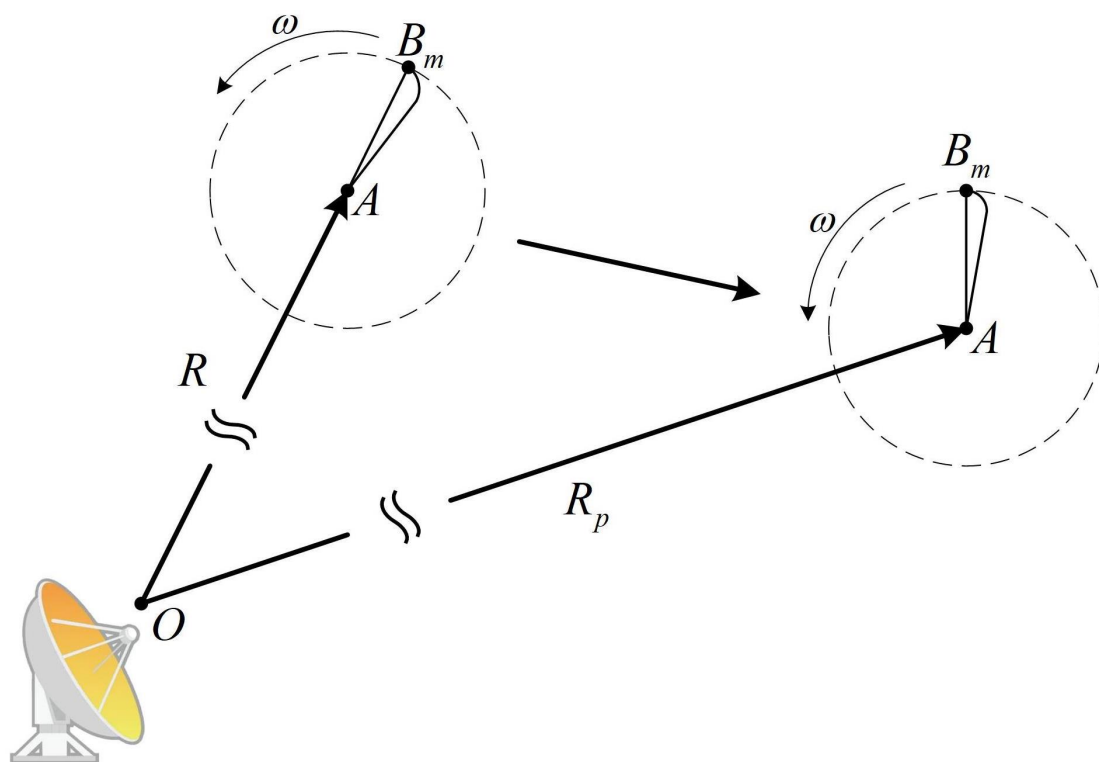


**Figure 1.** Geometry used to illustrate translational and rotational contributions to the micro-Doppler response.

For this illustrative model, the transmitted linear frequency-modulated (LFM) signal is given by

$$s(t) = \exp(j2\pi f_0 t) \exp(jk\pi t^2) \quad (1)$$

where  $t$  represents time,  $f_0$  is the carrier frequency,  $k = B/T_p$  is the frequency-modulation rate,  $B$  is the bandwidth, and  $T_p$  is the chirp duration.

The platform's initial radial distance is  $R$ , with the direction away from the radar defined as positive. Its translational range variation is described by the polynomial model

$$R_p(t) = R + \sum_{n=1}^N \frac{\alpha_n t^n}{n!} \quad (2)$$

where  $\alpha_1$  and  $\alpha_2$  represent the initial velocity and initial acceleration of the target, respectively.  $N$  denotes the dimensionality of the motion parameters in the polynomial model.

The blade contains  $M$  strong scattering points. For the  $m$ -th point  $B_m$ , the rotation radius and initial phase are denoted by  $r_m$  and  $\varphi_m$ , respectively, where  $m = 1, 2, \dots, M$ . The rotational range component relative to  $A$  and the total range between the radar and  $B_m$  are given by

$$R_{bm}(t) = r_m \cos(\omega t + \varphi_m) \quad (3)$$

$$R_m(t) = R_p(t) + R_{bm}(t) = R + \sum_{n=1}^N \frac{\alpha_n t^n}{n!} + r_m \cos(\omega t + \varphi_m) \quad (4)$$

With  $K_m$  denoting the reflection coefficient of the scattering point  $B_m$ , the down-converted baseband echo can be expressed as

$$s_r(t) = \sum_{m=1}^M K_m \exp\left(-j\frac{4\pi f_0}{c} R_m(t)\right) \exp\left(j\pi k \left(t - \frac{2R_m(t)}{c}\right)^2\right) \quad (5)$$

where  $c$  is the speed of light.

Considering the range-dependent carrier phase in (5), the translational Doppler component and the rotational Doppler component of the  $m$ -th scattering point are expressed as follows.

$$f_p(t) = -\frac{2f_0}{c} \sum_{n=1}^N \frac{\alpha_n t^{n-1}}{(n-1)!} \quad (6)$$

$$f_{b,m}(t) = \frac{2f_0}{c} r_m \omega \sin(\omega t + \varphi_m), \quad m = 1, \dots, M. \quad (7)$$

Here,  $f_p(t)$  is the common translational contribution, and  $f_{b,m}(t)$  is the rotational contribution of the  $m$ -th scattering point. The received echo is the superposition of the individual scattering contributions in Equation (5). The translational contribution is constant when the radial translational velocity is constant. The STFT is then applied to the baseband echo to obtain its time–frequency representation

$$\text{STFT}\{s_r(t)\}(t, f) = S(t, f) = \int_{-\infty}^{\infty} s_r(\tau) h(\tau - t) e^{-j2\pi f \tau} d\tau \quad (8)$$

where  $h(t)$  is the window function,  $f$  represents frequency, and  $S(t, f)$  is the joint representation of the signal in the time–frequency domain. A schematic diagram of the STFT principle is shown in Figure 2.

For classification, this study uses the measured STFT spectrograms from the DIAT- $\mu$ SAT dataset described in Section 2.2.

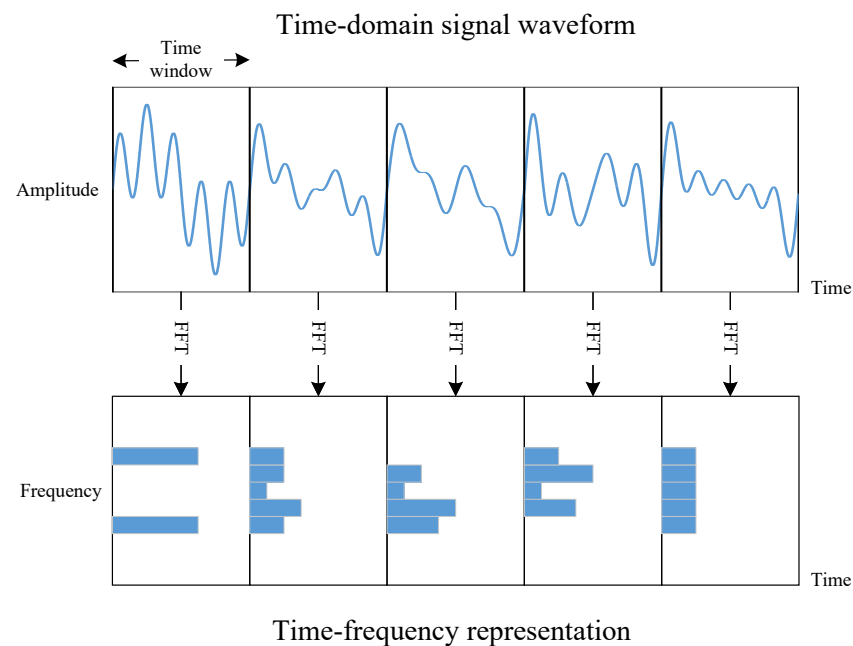


**Figure 2.** STFT processing used to obtain the time–frequency spectrogram.

## 2.2. Measured DIAT- $\mu$ SAT Dataset

The DIAT- $\mu$ SAT dataset comprises 4849 STFT time–frequency spectrograms from five types of small aerial targets, grouped into six categories: two-blade rotor, three-short-blade rotor, three-long-blade rotor, quadcopter, bionic bird, and a mixed class of bionic bird and two-blade rotor. The data were collected using an X-band continuous-wave radar [25]. Each spectrogram was generated from 30,000 time-domain points using a 1024-point discrete Fourier transform, a Hamming window, and an overlap of 200 points. Figure 3 presents representative examples with the coordinate axes removed. Accordingly, the following analysis focuses on visual characteristics, including the relative spacing of recurrent structures, the apparent spread along the frequency dimension, and distinctive patterns in the mixed-target class.

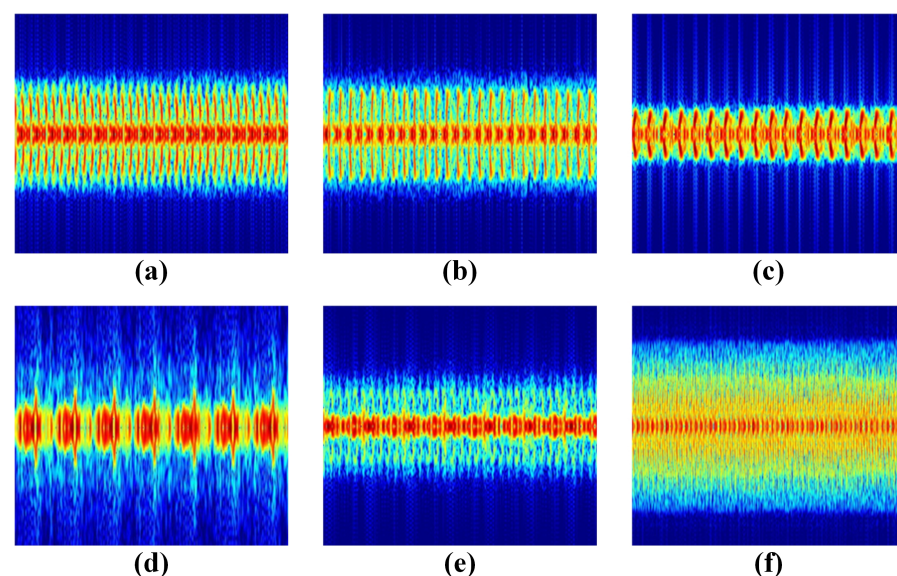


**Figure 3.** Representative measured STFT spectrograms from the six DIAT- $\mu$ SAT categories. (a) Two-blade rotor. (b) Three-short-blade rotor. (c) Three-long-blade rotor. (d) Quadcopter. (e) Bionic bird. (f) Bionic bird and two-blade rotor.

In the representative rotor examples, the two-blade rotor and three-short-blade rotor in Figure 3a,b exhibit closely spaced repetitive structures around a strong central band. The three-long-blade rotor in Figure 3c also shows recurring curved structures, but with a relatively narrower apparent spread along the frequency dimension. These examples share recurrent micro-Doppler structures while exhibiting differences in their visible spacing and extent.

The quadcopter example in Figure 3d exhibits several strong central patches accompanied by irregular surrounding structures. The bionic bird in Figure 3e shows small recurrent structures around the central band, whereas the mixed bionic bird and two-blade rotor example in Figure 3f presents a more densely structured central region. These observations indicate that the spectrograms contain both local time–frequency textures and repeated structures over wider regions of the spectrogram. This motivates jointly modeling local features and broader time–frequency dependencies.

### 2.3. TSAR-Net for Conventional Supervised Classification

ResNet provides a residual convolutional backbone for hierarchical feature extraction [34]. Transfer learning uses pretrained model parameters as an initialization for adaptation to the target task. In this study, an ImageNet-pretrained ResNet50 was adopted, with the first 40 layers frozen during training and the remaining layers fine-tuned for STFT spectrogram classification.

A multi-head self-attention mechanism is introduced into the transferred ResNet50 to model dependencies among spatially separated time–frequency features [35]. The self-attention mechanism can be expressed as

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (9)$$

Here,  $\mathbf{Q}$ ,  $\mathbf{K}$ , and  $\mathbf{V}$  represent the query vector, key vector, and value vector, respectively, while  $d_k$  is the dimension of the key vectors. The query, key, and value vectors are obtained by separate linear projections of the input features.

Multi-head self-attention models feature relationships in parallel across multiple heads

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_q)W^O \quad (10)$$

where  $q$  represents the number of heads. The outputs from each head are concatenated and then fused into the final result through the linear layer  $W^O$ . The multi-head design enables the model to capture correlated information at both local and global scales.

In TSAR-Net, the MHSA module is inserted between Conv4\_x and Conv5\_x to aggregate global dependencies before the final convolutional stage.

Since sample difficulty varies with observation angles in the measured dataset, this study uses focal loss [36] to down-weight easily classified samples and emphasize challenging cases. The focal loss is defined as

$$\text{FL}(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (11)$$

where  $(1 - p_t)^\gamma$  is the modulation factor, which reduces the loss weight for easily classified samples.  $\gamma \geq 0$  is the focusing parameter, controlling the degree of emphasis on hard-to-classify samples.  $p_t$  represents the predicted probability of the target class by the model and is defined as

$$p_t = \frac{\exp(z_y)}{\sum_{c=1}^6 \exp(z_c)}, \quad y \in \{1, \dots, 6\} \quad (12)$$

where  $z_c$  denotes the logit for class  $c$ , and  $p_t$  is the softmax probability assigned to the ground-truth class  $y$  in the six-class task.

In summary, the TSAR-Net architecture diagram is shown in Figure 4. Conv1 refers to the initial convolutional layer, while Conv2\_x, Conv3\_x, Conv4\_x, and Conv5\_x consist of 3, 4, 6, and 3 bottleneck blocks, respectively. Each bottleneck block is made up of three consecutive convolutional layers: a  $1 \times 1$  layer, a  $3 \times 3$  layer, and another  $1 \times 1$  layer. The first  $1 \times 1$  layer aims to reduce the number of channels in the feature maps, thereby lessening computational complexity. The following  $3 \times 3$  layer captures the spatial information from the input feature maps. The final  $1 \times 1$  layer returns the number of channels back to their original size. MHSA denotes the multi-head self-attention layer, Avgpool refers to the average pooling layer, and FC represents the fully connected layer.

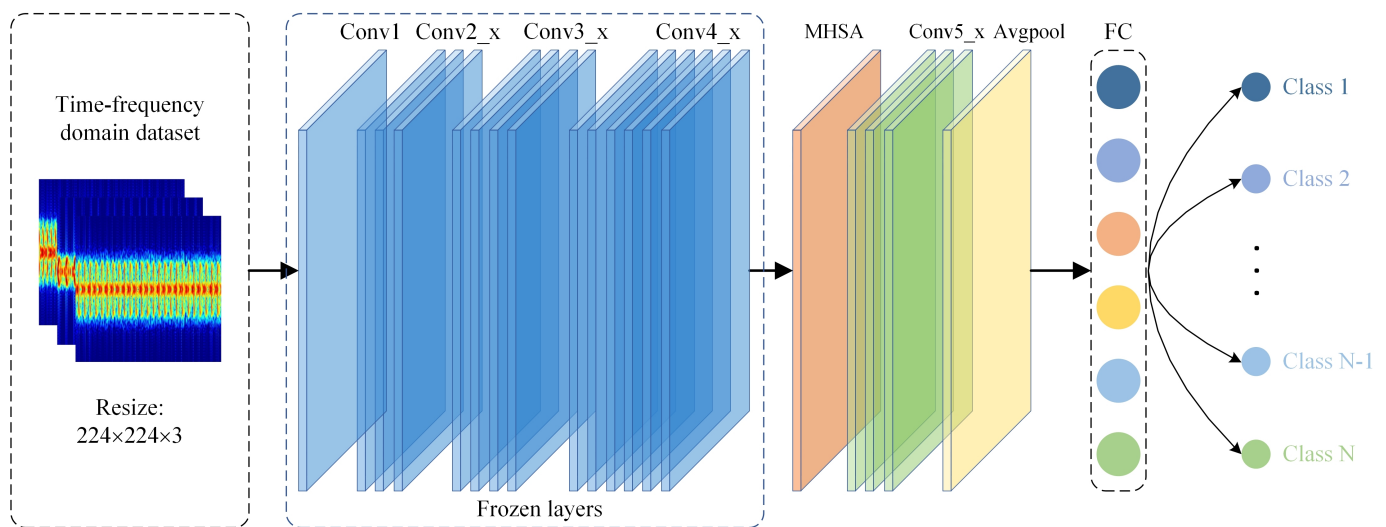


**Figure 4.** TSAR-Net architecture for transfer-enhanced local-to-global representation learning.

#### 2.4. PTSAR-Net for Limited-Label Classification

The output feature dimension of the average pooling layer in TSAR-Net is 2048. When the sample size is insufficient, the training data cover only a small portion of the feature space, limiting the learning of discriminative representations. High-dimensional features may also retain irrelevant or weakly discriminative information. In small-sample scenarios, the model may memorize specific features of the training data instead of capturing the overall distribution pattern, leading to overfitting.

To address the challenge of small-sample micro-motion target classification, this article introduces a projection head after the average pooling layer in TSAR-Net. The projection head compresses the 2048-dimensional average-pooling output into a lower-dimensional embedding for classification under limited-label settings.

The structure of the projection head is shown in Figure 5 and consists of three layers. The first layer is a linear layer that maps the 2048-dimensional input features to a 512-dimensional intermediate space. The second layer is an activation layer, where the ReLU function introduces a nonlinear transformation. The third layer is another linear layer that maps the 512-dimensional intermediate features to the final 128-dimensional embedding space.

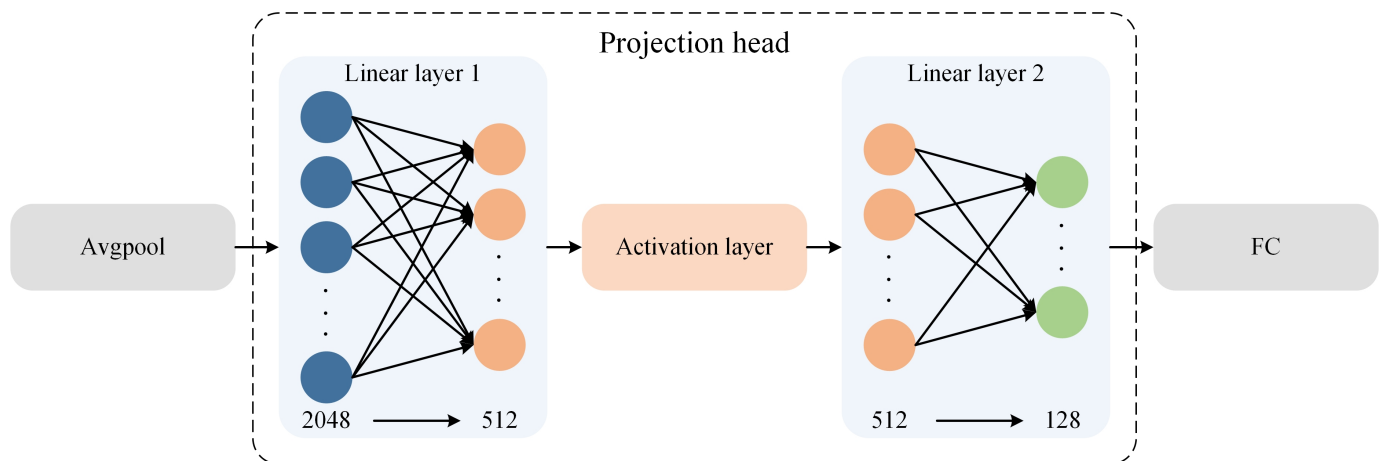


**Figure 5.** Compact projection stage used by PTSAR-Net under limited-label training.

### 2.5. Experimental Protocol

The hardware configuration of the workstation used in this experiment included an AMD EPYC 7742 processor, an NVIDIA RTX A6000 GPU with 48 GB of VRAM, and 256 GB of RAM. The software environment included CUDA v11.0.194 and PyTorch 1.7.1. All hardware and software configurations described in this article are based on publicly available commercial or open-source products. During the experiment, the images were resized to  $224 \times 224$  pixels. The multi-head self-attention layer was configured with 32 attention heads. Focal loss was configured with parameter  $\gamma = 2$ . The Adam optimizer was used with an initial learning rate of 0.001. The training process utilized a batch size of 256 and was conducted for 50 epochs.

## 3. Measured Data Verification and Analysis

### 3.1. Performance of TSAR-Net

#### 3.1.1. Ablation Study on the TSAR-Net Architecture

This section explores the practical roles of the self-attention mechanism and transfer learning in TSAR-Net. To evaluate the contributions of these two mechanisms to the model's performance, ablation studies are conducted on the measured DIAT- $\mu$ SAT dataset.

For the standard supervised experiments, the DIAT- $\mu$ SAT dataset was partitioned into 3838 training samples, 524 validation samples, and 487 test samples. The exact numbers of samples from each class in the three subsets are provided in Table 1.

**Table 1.** Class-wise numbers of training, validation, and test samples in the standard supervised experiments on DIAT- $\mu$ SAT.

| Class                         | Training | Validation | Test | Total |
|-------------------------------|----------|------------|------|-------|
| Two-blade rotor               | 634      | 86         | 80   | 800   |
| Three-long-blade rotor        | 632      | 86         | 81   | 799   |
| Three-short-blade rotor       | 633      | 87         | 80   | 800   |
| Bionic bird                   | 633      | 87         | 80   | 800   |
| Bionic bird & two-blade rotor | 645      | 88         | 82   | 815   |
| Quadcopter                    | 661      | 90         | 84   | 835   |
| Total                         | 3838     | 524        | 487  | 4849  |

For the self-attention mechanism, the ablation study removes the multi-head self-attention layer, using the transfer learning-based ResNet50 (TR-Net) as the control. In the transfer learning setup, the layer-freezing configuration is kept consistent with TSAR-Net, where the weights of the first 40 layers are not updated during training. For transfer

learning, the ablation study evaluates the performance of the SAR-Net model trained without pre-trained parameters and with no frozen layers. All models in the ablation study were trained with consistent batch-size and optimizer settings.

The classification results for TR-Net, SAR-Net, and TSAR-Net on the DIAT- $\mu$ SAT dataset are presented in Table 2, with accuracy variations shown in Figure 6. Precision is the ratio of true positive predictions to the total number of positive predictions made by the model. Recall is the ratio of true positive predictions to the total number of actual positive samples. F1-score, the harmonic mean of precision and recall, is used to evaluate both metrics simultaneously. Accuracy is the ratio of correctly predicted samples to the total number of samples.

**Table 2.** Ablation results on the measured DIAT- $\mu$ SAT dataset. Bold values indicate the uniquely highest results for each metric among the compared methods.

|                               | SAR-Net       |        |          | TR-Net    |        |          | TSAR-Net      |               |               |
|-------------------------------|---------------|--------|----------|-----------|--------|----------|---------------|---------------|---------------|
|                               | Precision     | Recall | F1-Score | Precision | Recall | F1-Score | Precision     | Recall        | F1-Score      |
| Two-blade rotor               | 1             | 0.9875 | 0.9937   | 0.9877    | 1      | 0.9938   | 1             | 1             | <b>1</b>      |
| Three-long-blade rotor        | <b>0.9875</b> | 0.9753 | 0.9814   | 0.9759    | 1      | 0.9878   | 0.9759        | 1             | 0.9878        |
| Three-short-blade rotor       | 0.9405        | 0.9875 | 0.9634   | 1         | 0.9750 | 0.9873   | 1             | 0.9875        | <b>0.9937</b> |
| Bionic bird                   | 1             | 1      | 1        | 1         | 1      | 1        | 1             | 1             | 1             |
| Bionic bird & two-blade rotor | 1             | 0.9756 | 0.9877   | 1         | 0.9878 | 0.9939   | 1             | 0.9878        | 0.9939        |
| Quadcopter                    | 1             | 1      | 1        | 1         | 1      | 1        | 1             | 1             | 1             |
| Macro average                 | 0.9880        | 0.9877 | 0.9877   | 0.9939    | 0.9938 | 0.9938   | <b>0.9960</b> | <b>0.9959</b> | <b>0.9959</b> |
| Weighted average              | 0.9881        | 0.9877 | 0.9878   | 0.9940    | 0.9938 | 0.9938   | <b>0.9960</b> | <b>0.9959</b> | <b>0.9959</b> |
| Accuracy                      |               | 0.9877 |          |           | 0.9938 |          |               | <b>0.9959</b> |               |

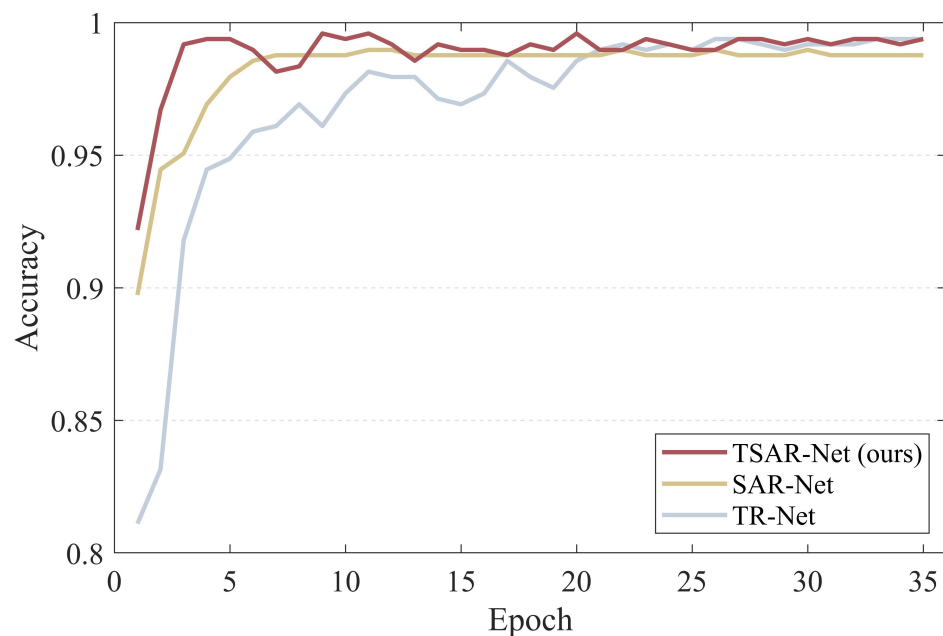


**Figure 6.** Validation accuracy for the transfer and MHSA ablation configurations on DIAT- $\mu$ SAT.

The ablation results show that transfer learning improves the model's learning efficiency and classification accuracy. The introduction of the multi-head self-attention mechanism further improves the model's ability to aggregate features from different locations and capture long-range dependencies between internal features, which benefits the classification of the three-short-blade rotor class. Overall, the ablation results support the integration of transfer learning and self-attention for micro-motion target classification.

### 3.1.2. Comparison with Representative Baselines

The comparison includes VGG19 [25] and MobileNetV2 [37] as CNN-dominant baselines, ViT-Tiny [38] as a Transformer-based baseline, and HOG-RBF-SVM [39] as a conventional feature-engineered machine-learning baseline. Table 3 summarizes overall accuracy and macro-averaged precision, recall, and F1-score on DIAT- $\mu$ SAT. Each macro metric is the unweighted mean of the corresponding class-wise values across the six classes.

**Table 3.** Classification performance of different methods on the measured DIAT- $\mu$ SAT dataset. Bold values indicate the uniquely highest results for each metric among the compared methods.

| Method      | Type           | Accuracy      | Macro Precision | Macro Recall  | Macro F1      |
|-------------|----------------|---------------|-----------------|---------------|---------------|
| VGG19       | CNN            | 0.9671        | 0.9691          | 0.9669        | 0.9671        |
| MobileNetV2 | CNN            | 0.9897        | 0.9897          | 0.9897        | 0.9897        |
| ViT-Tiny    | Transformer    | 0.9856        | 0.9856          | 0.9856        | 0.9855        |
| HOG-RBF-SVM | Traditional ML | 0.9138        | 0.9136          | 0.9132        | 0.9130        |
| TSAR-Net    | CNN + MHSA     | <b>0.9959</b> | <b>0.9960</b>   | <b>0.9959</b> | <b>0.9959</b> |

TSAR-Net achieves the highest accuracy of 99.59% and the highest scores across all three macro-averaged metrics. Its accuracy exceeds that of MobileNetV2, the best-performing external baseline at 98.97%, by 0.62 percentage points. The corresponding gains over ViT-Tiny, VGG19, and HOG-RBF-SVM are 1.03, 2.88, and 8.21 percentage points, respectively.

The confusion matrices in Figure 7a–e show that misclassifications are concentrated mainly among the rotor classes and the mixed bionic bird and two-blade rotor class. Compared with VGG19, MobileNetV2 and ViT-Tiny reduce the number of misclassified samples from the three-short-blade rotor class from seven to one and three, respectively, and reduce the mixed-class errors from five to two for both models. Both models misclassify two samples from the three-long-blade rotor class as three-short-blade rotor. HOG-RBF-SVM exhibits more dispersed confusion, including misclassifications of rotor and mixed-class samples as quadcopters. TSAR-Net makes only two errors: one three-short-blade rotor sample and one mixed-class sample are classified as three-long-blade rotors. All five methods correctly classify all bionic bird and quadcopter samples. Overall, the confusion matrices further confirm the lower misclassification rate of TSAR-Net on this dataset.

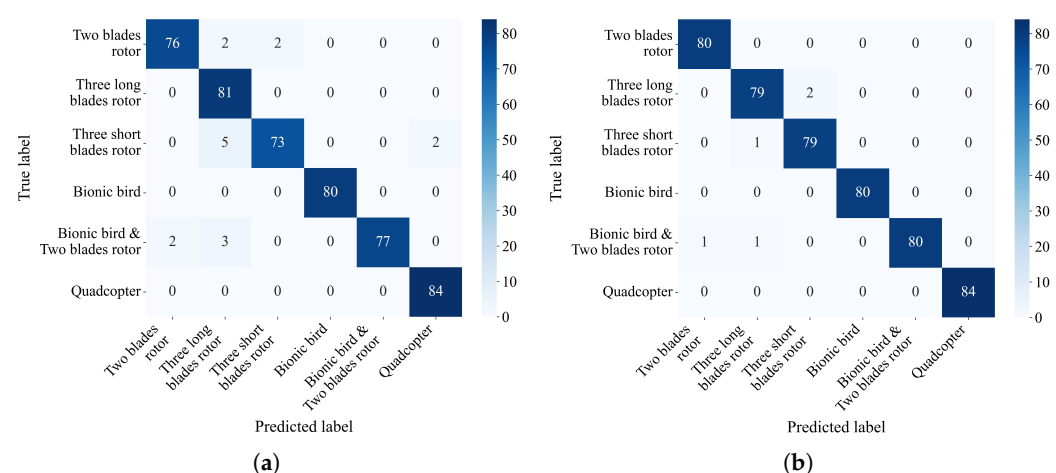


**Figure 7.** Cont.

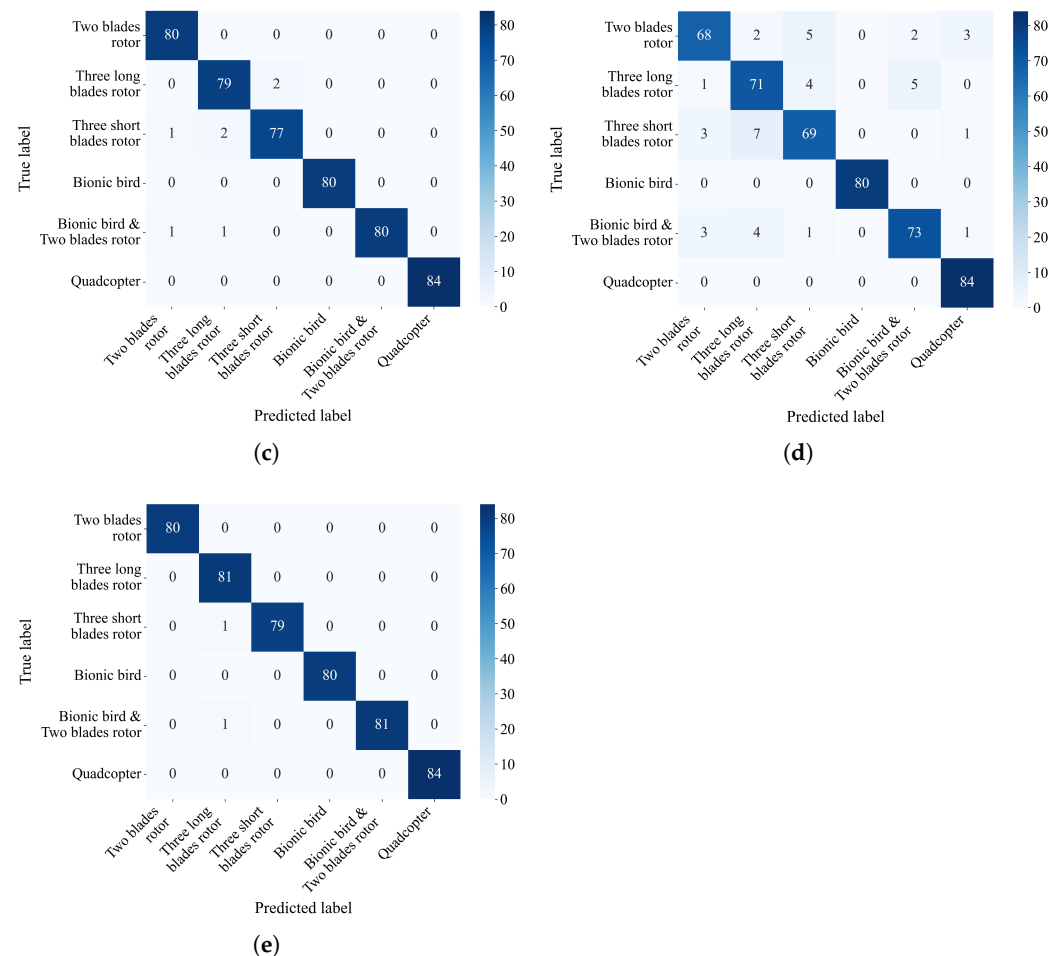


**Figure 7.** Confusion matrices of different methods on the measured DIAT- $\mu$ SAT test set. (a) VGG19. (b) MobileNetV2. (c) ViT-Tiny. (d) HOG-RBF-SVM. (e) TSAR-Net.

### 3.2. Performance of PTSAR-Net

To evaluate classification under limited-label conditions, 5, 10, 20, and 50 samples are randomly selected from each class of DIAT- $\mu$ SAT for training, while the remaining samples are used for testing. Table 4 reports the classification accuracies of all six methods under these four settings.

PTSAR-Net achieves the highest accuracy in every setting. With five samples per class, it reaches 81.14%, exceeding TSAR-Net, the second-best method, by 10.42 percentage points. With 10 samples per class, TSAR-Net again achieves the second-highest accuracy, while PTSAR-Net leads by 2.97 percentage points. With 20 and 50 samples per class, MobileNetV2 is the next-best method: PTSAR-Net achieves 87.88% and 93.93%, exceeding MobileNetV2 by 0.93 and 1.23 percentage points, respectively. PTSAR-Net also exceeds TSAR-Net by 2.83 and 3.71 percentage points under these two settings. The substantially larger gain in the five-sample setting is consistent with the role of the compact projection in limited-label learning. When only a few samples are available, the high-dimensional pooled representation of TSAR-Net is more susceptible to insufficient feature-space coverage and overfitting, whereas the lower-dimensional embedding of PTSAR-Net provides a more compact representation for classification. As the number of training samples increases, the feature space is better covered by the training data, allowing TSAR-Net to learn more stable decision boundaries and thereby reducing the relative benefit of the compact projection.

**Table 4.** Classification accuracy under limited-label training on DIAT- $\mu$ SAT. Bold values indicate the highest accuracy for each training sample size among the compared methods.

| Method      | Training Samples per Class |               |               |               |
|-------------|----------------------------|---------------|---------------|---------------|
|             | 5                          | 10            | 20            | 50            |
| VGG19       | 0.6323                     | 0.7797        | 0.8458        | 0.9253        |
| MobileNetV2 | 0.6867                     | 0.7622        | 0.8695        | 0.9270        |
| ViT-Tiny    | 0.6016                     | 0.5617        | 0.6441        | 0.7852        |
| HOG-RBF-SVM | 0.5460                     | 0.6289        | 0.6957        | 0.7624        |
| TSAR-Net    | 0.7072                     | 0.8177        | 0.8505        | 0.9022        |
| PTSAR-Net   | <b>0.8114</b>              | <b>0.8474</b> | <b>0.8788</b> | <b>0.9393</b> |

#### 4. Discussion

On the measured DIAT- $\mu$ SAT dataset, TSAR-Net achieves higher accuracy than its ablation variants and the CNN-based, Transformer-based, and handcrafted-feature-based baselines evaluated in the standard supervised setting. PTSAR-Net achieves the highest accuracy among the six methods in all four limited-label settings, with its largest margin over the next-best method at five training samples per class. Together, these results support the effectiveness of the proposed framework for the evaluated standard supervised and limited-label radar classification tasks.

The current evaluation is based on the measured DIAT- $\mu$ SAT dataset, and the reported results reflect the acquisition conditions represented in this dataset. The use of STFT spectrograms and local-to-global feature modeling provides a basis for adapting the framework to other radar datasets through suitable preprocessing and model fine-tuning. Extending the evaluation to additional radar configurations and datasets is a natural direction for future work.

The micro-Doppler time–frequency representations used in this study primarily characterize motion-related signatures. Full-polarimetric information can provide complementary information on polarization-dependent scattering characteristics [32,33]. These complementary cues could be integrated with micro-Doppler features at the feature level to form a richer target representation [40–42]. Future work could therefore explore combining scattering information across polarization channels with micro-Doppler features to jointly characterize target motion and scattering. Such a representation may help distinguish small aerial targets with similar motion signatures but different scattering responses. This direction would complement the present focus on local-to-global micro-Doppler representation learning and limited-label classification.

#### 5. Conclusions

This article presents TSAR-Net, a transfer-enhanced local-to-global representation framework for micro-motion target classification using micro-Doppler features extracted from radar echoes. By integrating transfer learning and multi-head self-attention, TSAR-Net extracts hierarchical local features and models global time–frequency dependencies. On the measured DIAT- $\mu$ SAT dataset, TSAR-Net achieves 99.59% accuracy, the highest among the methods compared in the standard supervised setting, exceeding the closest external baseline by 0.62 percentage points. To address limited training samples, PTSAR-Net incorporates a projection head that maps the 2048-dimensional pooled feature through 512 dimensions to a compact 128-dimensional embedding. PTSAR-Net achieves the highest accuracy in all four limited-label settings, including 81.14% with five training samples per class, which is 10.42 percentage points higher than TSAR-Net. These results show that TSAR-Net and PTSAR-Net offer effective approaches for the evaluated conventional and limited-label radar classification settings.

**Author Contributions:** Conceptualization, Y.S. (Yuxian Sun) and Y.L.; Methodology, Y.S. (Yuxian Sun); Software, Y.S. (Yuxian Sun) and J.Z.; Validation, Y.S. (Yuanyuan Song); Formal Analysis, Y.S. (Yuxian Sun); Resources, Y.S. (Yuanyuan Song); Data Curation, Y.S. (Yuanyuan Song) and H.F.; Writing—Original draft, Y.S. (Yuxian Sun); Writing—Review and Editing, Y.S. (Yuanyuan Song), H.F. and J.Z.; Visualization, Y.S. (Yuxian Sun), Y.L. and J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant 62371046 and Grant 62201048.

**Data Availability Statement:** The data analyzed in this study are available in the publicly reported DIAT- $\mu$ SAT dataset described in Reference [25].

**Acknowledgments:** The authors thank the contributors of the DIAT- $\mu$ SAT dataset [25]. During the preparation of this manuscript, generative AI tools were used solely for English language editing. No generative AI tools were used for study design, methodology development, data generation, data analysis, or figure generation. All scientific content was developed and verified by the authors.

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

## References

- Berndt, R.J.; Abdul Gaffar, M.Y.; Nel, W.A.J.; O'Hagan, D.W. Radar detectability of light aircraft micro-Doppler modulation. *IET Radar Sonar Navig.* **2024**, *18*, 1750–1766. [\[CrossRef\]](#)
- Mao, T.; Li, Z.; Zhu, K.; Zhang, Y.; Sun, H. Radar backscattering modelling and micro-motion parameter estimation method for quadcopter. *IET Radar Sonar Navig.* **2022**, *16*, 161–169. [\[CrossRef\]](#)
- Saffari, A.; Zahiri, S.H.; Khishe, M. Automatic recognition of sonar targets using feature selection in micro-Doppler signature. *Def. Technol.* **2023**, *20*, 58–71. [\[CrossRef\]](#)
- Arabpour, F.; Sebt, M.A. Enhancing vulnerable road user classification through micro-Doppler and deep learning: The impact of time window. *IET Radar Sonar Navig.* **2025**, *19*, e70065. [\[CrossRef\]](#)
- Hasan, M.M.; Chakraborty, M.; Raj, A.A.B. A hyper-parameters-tuned R-PCA+ SVM technique for sUAV targets classification using the range-/micro-Doppler signatures. *IEEE Trans. Radar Syst.* **2023**, *1*, 623–631. [\[CrossRef\]](#)
- Zhao, Y.; Su, Y. Estimation of micro-Doppler parameters with combined null space pursuit methods for the identification of LSS UAVs. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5102811. [\[CrossRef\]](#)
- Zhang, R.; Wang, Y.; Yeh, C.; Lu, X. Precession parameter estimation of warhead with fins based on micro-Doppler effect and radar network. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 443–459. [\[CrossRef\]](#)
- Wu, Z.; Liu, H.; Ma, C.; Liu, Z. Aerial target recognition with enhanced micro-Doppler dynamic features based on frequency-modulated continuous-wave radar. *IEEE Sens. J.* **2023**, *23*, 23119–23132. [\[CrossRef\]](#)
- Xu, Z.; Wu, Q.; Ai, X.; Liu, X.; Wu, J.; Zhao, F. Micromotion frequency estimation of multiple space targets based on RCD spectrum. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *61*, 14019–14030. [\[CrossRef\]](#)
- Zhu, Y.; Xu, Z.; Ai, X.; Wu, Q.; Xie, X. A novel estimation method for multiple micromotion periods based on iterative CVD filtering. *IEEE Sens. J.* **2025**, *25*, 8481–8492. [\[CrossRef\]](#)
- Djurović, I.; Popović-Bugarin, V.; Simeunović, M. The STFT-based estimator of micro-Doppler parameters. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 1273–1283. [\[CrossRef\]](#)
- Seyfioglu, M.S.; Erol, B.; Gurbuz, S.Z.; Amin, M.G. DNN transfer learning from diversified micro-Doppler for motion classification. *IEEE Trans. Aerosp. Electron. Syst.* **2019**, *55*, 2164–2180. [\[CrossRef\]](#)
- Rojhani, N.; Passafiume, M.; Sadeghibakhi, M.; Collodi, G.; Cidronali, A. Model-based data augmentation applied to deep learning networks for classification of micro-Doppler signatures using FMCW radar. *IEEE Trans. Microw. Theory Techn.* **2023**, *71*, 2222–2236. [\[CrossRef\]](#)
- Mignone, L.; Ilioudis, C.; Clemente, C.; Ullo, S. Enhancing micro-Doppler classification using superlet-based time-frequency distribution. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 9831–9838. [\[CrossRef\]](#)
- Park, J.; Park, J.S. Classification of small drones using low-uncertainty micro-Doppler signature images and ultra-lightweight convolutional neural network. *IEEE Trans. Image Process.* **2024**, *33*, 2979–2994. [\[CrossRef\]](#) [\[PubMed\]](#)
- Song, Q.; Huang, S.; Zhang, Y.; Chen, X.; Chen, Z.; Zhou, X.; Deng, Z. Radar target classification using enhanced Doppler spectrograms with ResNet34\_CA in ubiquitous radar. *Remote Sens.* **2024**, *16*, 2860. [\[CrossRef\]](#)

17. Fang, X.; Xiao, G. Rotor blades micro-Doppler feature analysis and extraction of small unmanned rotorcraft. *IEEE Sens. J.* **2021**, *21*, 3592–3601. [[CrossRef](#)]
18. Xu, X.; Feng, C. Micro-Doppler extraction of radar targets with translational motion based on spatial transformer network. *IEEE Signal Process. Lett.* **2022**, *29*, 2348–2352. [[CrossRef](#)]
19. Oh, B.S.; Guo, X.; Wan, F.; Toh, K.A.; Lin, Z. Micro-Doppler mini-UAV classification using empirical-mode decomposition features. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 227–231. [[CrossRef](#)]
20. Zhao, Y.; Su, Y. Synchrosqueezing phase analysis on micro-Doppler parameters for small UAVs identification with multichannel radar. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 411–415. [[CrossRef](#)]
21. Kang, H.; Kim, B.K.; Park, J.S.; Suh, J.S.; Park, S.O. Drone elevation angle classification based on convolutional neural network with micro-Doppler of multipolarization. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3501205. [[CrossRef](#)]
22. Zhang, W.; Li, G.; Baker, C. Radar recognition of multiple micro-drones based on their micro-Doppler signatures via dictionary learning. *IET Radar Sonar Navig.* **2020**, *14*, 1310–1318. [[CrossRef](#)]
23. Sun, Y.; Abeywickrama, S.; Jayasinghe, L.; Yuen, C.; Chen, J.; Zhang, M. Micro-Doppler signature-based detection, classification, and localization of small UAV with long short-term memory neural network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6285–6300. [[CrossRef](#)]
24. Wang, Y.; Feng, C.; Hu, X.; Zhang, Y. Classification of space micromotion targets with similar shapes at low SNR. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3504305. [[CrossRef](#)]
25. Kumawat, H.C.; Chakraborty, M.; Raj, A.A.B.; Dhavale, S.V. DIAT- $\mu$ SAT: Small aerial targets' micro-Doppler signatures and their classification using CNN. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6004005. [[CrossRef](#)]
26. Kumawat, H.C.; Chakraborty, M.; Raj, A.A.B. DIAT-RadSATNet—A novel lightweight DCNN architecture for micro-Doppler-based small unmanned aerial vehicle (SUAV) targets' detection and classification. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 8504011. [[CrossRef](#)]
27. Ma, B.; Chen, B. A 1D cascaded denoising and classification framework for micro-Doppler-based radar target recognition. *Remote Sens.* **2025**, *17*, 1515. [[CrossRef](#)]
28. White, D.; Jahangir, M.; Baker, C.J.; Antoniou, M. Urban bird-drone classification with synthetic micro-Doppler spectrograms. *IEEE Trans. Radar Syst.* **2024**, *2*, 167–179. [[CrossRef](#)]
29. Oh, B.S.; Lin, Z. Extraction of global and local micro-Doppler signature features from FMCW radar returns for UAV detection. *IEEE Trans. Aerosp. Electron. Syst.* **2021**, *57*, 1351–1360. [[CrossRef](#)]
30. Kim, B.K.; Kang, H.S.; Lee, S.; Park, S.O. Improved drone classification using polarimetric merged-Doppler images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 1946–1950. [[CrossRef](#)]
31. Tu, J.; Tu, J.; Feng, W.; Liu, Z. Physics-informed semantic prompt learning for few-shot low-altitude radar target recognition in remote sensing. *Remote Sens.* **2026**, *18*, 2316. [[CrossRef](#)]
32. Zhu, Y.; Xu, Z.; Ai, X.; Chen, Z.; Yang, Z.; Xiao, S. Subterahertz full-polarimetric measurement and ISAR imaging of scaled space targets. *IEEE Trans. Antennas Propag.* **2026**, *74*, 7982–7996. [[CrossRef](#)]
33. Xu, Z.; Zheng, Y.; Zeng, Y.; Yang, Z.; Zhang, Y.; Wu, Q.; Ai, X. Measurement and calibration methods for full-polarization characteristics of radar targets at subterahertz frequencies. *IEEE Trans. Antennas Propag.* **2026**, *74*, 3471–3479. [[CrossRef](#)]
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2016; pp. 770–778. [[CrossRef](#)]
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 5998–6008.
36. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2017; pp. 2980–2988. [[CrossRef](#)]
37. Ayaz, F.; Alhumaily, B.; Hussain, S.; Imran, M.A.; Arshad, K.; Assaleh, K.; Zoha, A. Radar signal processing and its impact on deep learning-driven human activity recognition. *Sensors* **2025**, *25*, 724. [[CrossRef](#)] [[PubMed](#)]
38. Furusawa, T.; Premachandra, C.; Kihara, S.; Takizawa, M.; Katsuragi, H.; Hashimura, S.; Hosoya, N. AI-based bolt/nut looseness sensing using spectrogram images of shaft vibrations. *IEEE Sens. J.* **2025**, *25*, 15882–15892. [[CrossRef](#)]
39. Guendel, R.G.; Fioranelli, F.; Yarovoy, A. Phase-based classification for arm gesture and gross-motor activities using histogram of oriented gradients. *IEEE Sens. J.* **2021**, *21*, 7918–7927. [[CrossRef](#)]
40. Zhang, R.; Xu, L.; Yu, Z.; Shi, Y.; Mu, C.; Xu, M. Deep-IRTarget: An automatic target detector in infrared imagery using dual-domain feature extraction and allocation. *IEEE Trans. Multimed.* **2022**, *24*, 1735–1749. [[CrossRef](#)]

41. Zhang, R.; Li, L.; Zhang, Q.; Zhang, J.; Xu, L.; Zhang, B.; Wang, B. Differential feature awareness network within antagonistic learning for infrared-visible object detection. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 6735–6748. [[CrossRef](#)]
42. Sun, H.; Chen, X.; Cao, Z.; Yuan, W.; Zhang, R. DIHARNet: A temporal-Doppler-spatial domain fusion method for direction insensitive human activity recognition based on MMWave radar. *IEEE Trans. Mob. Comput.* **2026**, *25*, 21872–21883. [[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.