remote sensing

MDPI

Article

SDRnet: A Deep Fusion Network for ISAR Ship Target Recognition Based on Feature Separation and Weighted Decision

Jie Deng and Fulin Su *

School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China; 22s105183@stu.hit.edu.cn

* Correspondence: franklin_su@hit.edu.cn

Abstract: Existing methods for inverse synthetic aperture radar (ISAR) target recognition typically rely on a single high-resolution radar signal type, such as ISAR images or high-resolution range profiles (HRRPs). However, ISAR images and HRRP data offer representations of targets across different aspects, each containing valuable information crucial for radar target recognition. Moreover, the process of generating ISAR images inherently facilitates the acquisition of HRRP data, ensuring timely data collection. Therefore, to fully leverage the different information from both HRRP data and ISAR images and enhance ISAR ship target recognition performance, we propose a novel deep fusion network named the Separation-Decision Recognition network (SDRnet). First, our approach employs a convolutional neural network (CNN) to extract initial feature vectors from ISAR images and HRRP data. Subsequently, a feature separation module is employed to derive a more robust target representation. Finally, we introduce a weighted decision module to enhance overall predictive performance. We validate our method using simulated and measured data containing ten categories of ship targets. The experimental results confirm the effectiveness of our approach in improving ISAR ship target recognition.

Keywords: target recognition; inverse synthetic aperture radar; high-resolution range profile; deep fusion; feature separation; weighted decision

Citation: Deng, J.; Su, F. SDRnet: A Deep Fusion Network for ISAR Ship Target Recognition Based on Feature Separation and Weighted Decision. *Remote Sens.* **2024**, *16*, 1920. https://doi.org/10.3390/rs16111920

Academic Editor: Kevin Tansey

Received: 20 March 2024
Revised: 22 May 2024
Accepted: 24 May 2024
Published: 27 May 2024

1. Introduction

Inverse synthetic aperture radar (ISAR) target recognition, an extension of radar automatic target recognition (RATR), holds growing importance in both military and civilian applications owing to its capability to function under diverse weather conditions and at any time of the day [1,2]. As a result, ISAR target recognition has attracted significant international attention.

High-resolution range profile (HRRP) data, extensively employed in RATR [3–7], serves as a valuable source for discerning target size and the distribution of scattering points along a single dimension. Its acquisition is straightforward, and its quality remains unaltered by focusing algorithms [1,2]. Notably, during the process of ISAR imaging, HRRP data can be concurrently captured. Figure 1 illustrates the ISAR imaging framework utilizing classic range-Doppler (RD) algorithms. The initial step involves range compression of received ISAR echoes, followed by the extraction of HRRP data through a modulo operation. Subsequently, translational motion compensation (TMC) is employed to mitigate the translational component of the target, involving stages of range alignment and phase adjustment. If significant rotational motion occurs within the coherent processing interval (CPI), resulting in a blurred ISAR image due to a rotation-induced time-varying Doppler spectrum, rotational motion compensation (RMC) is necessary to rectify rotational errors. Once translational and rotational motion components are removed, the ISAR image is generated through Fourier transformation along the pulse direction. ISAR images effectively capture the distribution of scattering points along the azimuth direction, providing a clear

depiction of the target's structural information, thereby enhancing its visual interpretability and comprehension.

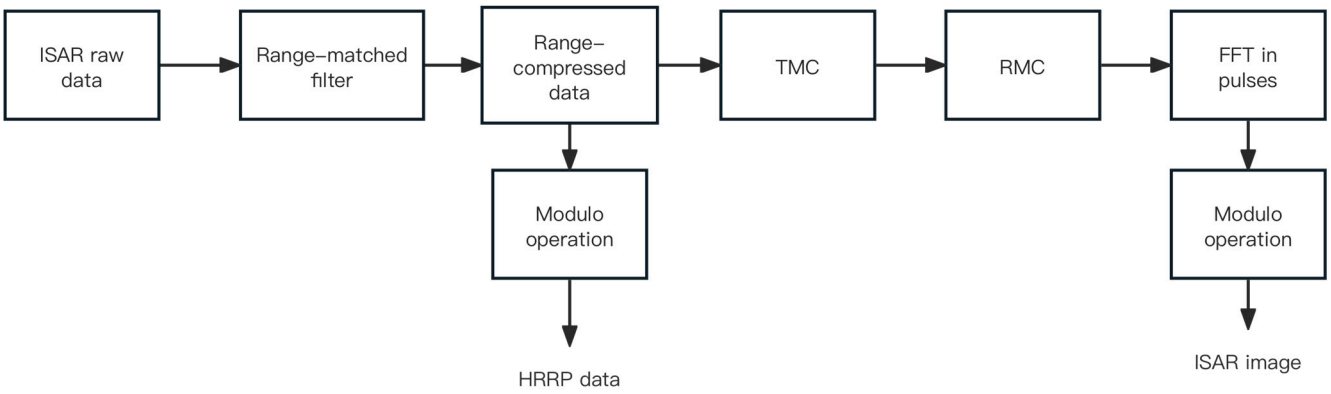


Figure 1. Range-Doppler algorithm for ISAR imaging.

Feature extraction plays a pivotal role in target recognition, directly influencing the performance and accuracy of recognition systems. The evolution of feature extraction, transitioning from manual methods to deep feature extraction utilizing deep networks for both HRRP data and ISAR images, has notably enhanced the performance of RATR systems [1,2,8–11]. In the realm of HRRP target recognition, numerous studies have leveraged deep neural network methodologies for analyzing HRRP data. For instance, a stacked corrective autoencoder (SCAE) was employed to extract features from HRRP, with the average profile serving as the correction term [9]. An enhanced variational autoencoder (VAE) was introduced to acquire probabilistic latent features [12]. A deep belief network (DBN) was employed to extract discriminative features, complemented by t-distributed stochastic neighbor embedding (t-SNE) to enhance HRRP segmentation across different target-aspect sectors [13]. Furthermore, the efficacy of a fully connected network (FCN) was evaluated on HRRP data [14]. Recognizing the successful deployment of convolutional networks in image processing, CNNs were applied to radar HRRP target recognition, with structural features learned from multiple layers being visualized [15]. In the domain of ISAR image target recognition, there also has been a notable evolution from manual feature extraction techniques [16–20] to the adoption of neural networks for feature extraction [1,2,21–23]. In deep neural network-based methodologies, CNNs serve as the primary feature extractors. For instance, a spatial transformer network model was utilized to address the unknown deformations of ISAR images resulting from changes in the attitude of targets [21]. Zhao et al. [22] proposed a pre-trained CNN tailored for small datasets. CNNs are widely favored in image recognition tasks due to their capacity to abstractly represent target structures through iterative learning processes, wherein structural information plays a pivotal role in accurate recognition. Leveraging the advantages offered by CNNs, our approach involves employing a renowned CNN model in the realm of image recognition, specifically AlexNet [24], to serve as the feature extraction network for ISAR images.

However, existing RATR methods predominantly rely on a single modality, either HRRP data or ISAR images [11]. Figure 1 illustrates the process where HRRP data are obtained before using imaging algorithms to generate ISAR images, ensuring timely data acquisition and real-time system performance. Combining both modalities for target recognition has the potential to enhance RATR performance. Some may question the need for combining the two modalities, given that ISAR images are derived from HRRP data and presumably contain the same information. However, the distinct generation mechanisms of HRRP data and ISAR images imply that their information contents are not identical, as they represent the target in different aspects. By integrating HRRP data with ISAR images, we are able to acquire a more comprehensive understanding of the target. Therefore, to effectively integrate the two modalities for target recognition, this paper

presents the Separation-Decision Recognition network (SDRnet), aimed at recognizing ISAR ship targets. Initially, in the feature extraction phase, we employ convolutional neural networks (CNNs) to extract features from both ISAR images and the average HRRP of the target. These extracted features serve as the initial features for subsequent modules. Subsequently, we introduce a feature separation module leveraging multi-kernel maximum mean discrepancy (MK-MMD) to explicitly disentangle each modal feature into shared and private components. We posit that shared components offer a more abstract representation, while private components could retain valuable complementary information about the target. Following the feature separation module, the weighted decision fusion module is implemented. We create three sub-classifiers and use the private features of the two modalities after feature separation and the private features after integration as inputs, with maximum class probability (MCP) used to set the weights of the three sub-classifiers. Finally, the weighted decision vector is passed into a softmax layer to predict the classification results. The primary contributions of this work can be summarized as follows:

1. We propose a deep feature fusion method based on ISAR image and HRRP data for target recognition. This method can fully exploit feature information about the target, thereby achieving satisfying recognition performance.
2. We used a feature separation module based on MK-MMD to effectively exploit shared and private information contained in HRRP data and ISAR images for robust target recognition. The module facilitates thorough consideration of correlation and complementarity between the two modalities to obtain a more robust representation of the target.
3. We designed a weighted decision fusion module to fit the feature separation module. We used it to further improve the accuracy and reliability of prediction. We verified the robustness and effectiveness of the proposed method on simulated and measured datasets. Moreover, the proposed method could achieve a higher recognition rate than the traditional fusion methods.

The subsequent sections of this paper are structured as follows: Section 2 provides an overview of related works focusing on RATR based on HRRP data and ISAR images. In Section 3, we introduce the proposed deep fusion network. Section 4 assesses the performance of the proposed method through a series of experiments and ablation studies conducted on both simulated and measured data. Finally, Section 5 presents the conclusions drawn from our research findings.

2. Related Works

2.1. Information Fusion

In recent years, advancements in sensor technology have led to a notable increase in both the diversity and complexity of available information. This evolution highlights the need for more advanced information fusion technologies to effectively address these changes. Information fusion involves the integration, aggregation, and processing of data from diverse sources, formats, and levels to produce comprehensive, accurate, and actionable insights [25].

Information fusion can be categorized into three main types based on the level of abstraction of the information being fused: data-level fusion [26], feature-level fusion [27], and decision-level fusion [28].

Data-level fusion is the simplest fusion method, but its improvement in model performance is very limited. Decision-level fusion can integrate the decision or prediction results of multiple classifiers, thus introducing diversity and helping to improve the reliability of model decisions. However, since decision fusion operates at the highest abstraction level, a single decision fusion will inevitably lose some important details or information. In addition, determining the weight of each sub-classifier is also a key issue in decision fusion. A decision-level fusion method that combines ISAR images and range profile (RP) data were used for target recognition [29]. Feature-level fusion is considered a more efficient approach of fusion and is frequently used to improve model performance. Several

studies concentrating on image segmentation or classification employ feature-level fusion to integrate multi-level features [30–38]. However, these studies typically fuse features derived from the same data at varying levels. In contrast, our approach involves fusing features extracted from distinct data through their respective feature extraction networks. Based on the gated recurrent unit (GRU) method, an extended-GRU feature-level fusion module that combines SAR images and average HRRP was used for target recognition. This method adaptively learns the weight of each modal feature to distinguish their contribution to sample discrimination [11]. However, this method does not consider the deep correlation and complementarity between the two modalities when performing feature-level fusion, although it achieves good results. Hence, this paper introduces feature separation technology at the feature level to fuse the features of both modalities, fully considering their correlation and complementarity, thereby resulting in a more robust representation of the target. Furthermore, this article incorporates a decision-making fusion module subsequent to feature separation. This hybrid fusion method enhances the reliability of the model's decision-making process, compensating for the shortcomings of decision-level fusion, which may result in the loss of crucial information.

2.2. Feature Separation

Feature separation technology involves explicitly partitioning the latent representation of each modality to enhance the understanding and processing of multimodal data, primarily utilized across various tasks in the field of computer vision [39–43]. To the best of our knowledge, its application in combining ISAR images and HRRP data for radar target recognition remains unexplored. A feature separation method based on deep networks was previously employed in the domain separation network (DNS) for unsupervised adaptation [39]. DNS employs a shared-weight encoder to capture domain-shared features from input samples, utilizing auxiliary loss to facilitate the convergence of shared representations. These shared representations from the source domain are then utilized to train the network for the task at hand. However, the approach solely relies on modality-shared representations to accomplish the task. Recognizing this, both modality-shared and modality-private feature representations were considered for action prediction [43]. In this work, the author explicitly separates the latent space of each modality into shared and private feature spaces to enhance recognition robustness. While the shared feature space is obtained using an auxiliary similarity loss, the private feature space is derived without employing additional processes. Our approach diverges from this previous work in several key aspects. Firstly, to obtain two more complementary private feature spaces, auxiliary loss is employed in this study to guide their formation. Secondly, in our task, we make an assumption regarding the correlation between the initial modal features, and we investigate the impact of this assumption on correlating the shared features of the two modalities.

3. The Proposed Method

The structure of the SDRnet for ISAR target recognition is shown in Figure 2. As shown in Figure 2, the framework of the proposed method can be briefly summarized in the following parts:

1. Initial feature extraction: As shown in Figure 1, during the ISAR imaging process, we can obtain the ISAR image of the target and the corresponding HRRP data simultaneously. For the HRRP data, we use the average HRRP obtained after preprocessing. Then, the ISAR image and the average HRRP are fed into the CNNs shown in Figure 3 for training to obtain their corresponding initial features. These initial features serve as input to the subsequent fusion process.
2. Feature separation: This section introduces feature separation technology into RATR, aiming to explicitly partition the initial feature space of each modality into shared feature space and private feature space. Private features play a crucial role: when disturbances affect one modality's private features or critical information is lost, the other modality's private features can offer valuable support for target differentiation,

thereby enhancing the recognition system's robustness. Moreover, shared features are expected to provide a more abstract common representation of the two modalities, reducing overfitting to specific modes and improving robustness. Shared feature information is obtained by maximizing the similarity between the features of the average HRRP data and ISAR images, while private feature information is derived by maximizing the difference between the two. However, in our task, we found that the acquired shared feature representations may not consistently enhance sample discriminability. Consequently, we decided to forego the shared feature branches and only retain the private feature components to enhance the sample discriminability, making the model more robust and stable.

3. Weighted decision fusion: We constructed three sub-classifiers for weighted decision fusion. We used the private features of the two modalities after feature separation and the private features after integration as inputs of these three sub-classifiers, with MCP used to set their weights. The purpose of using this module is to further improve the accuracy and reliability of decision making.
4. Finally, the decision vector obtained by integrating the outputs of the three sub-classifiers is fed into a softmax layer to classify the target.

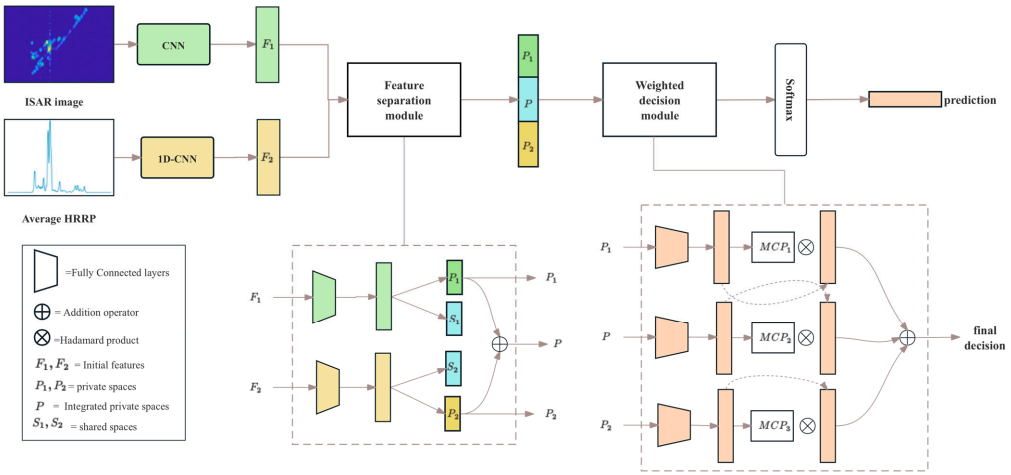
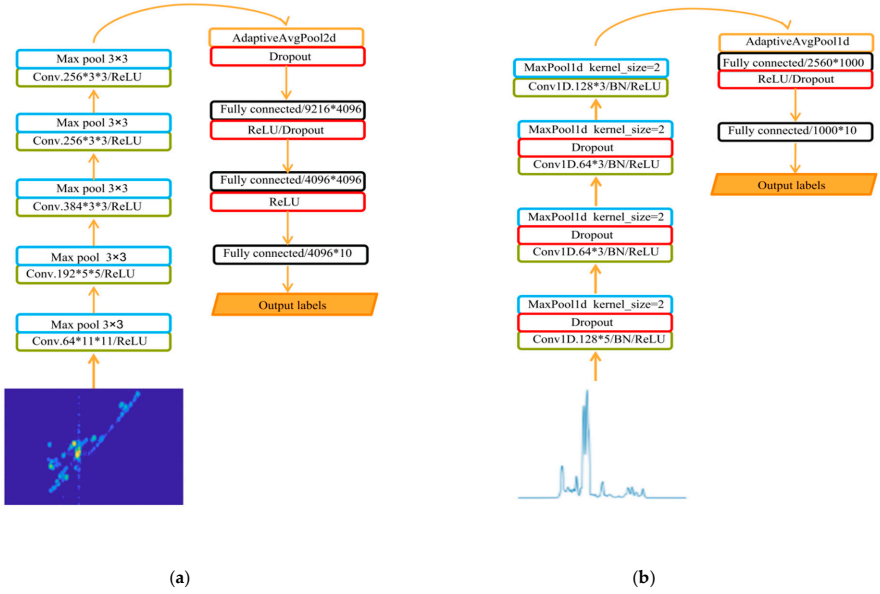


Figure 2. Structure of the SDRnet.



(**a**) (**b**)

Figure 3. (**a**) CNN for ISAR image; (**b**) 1D-CNN for HRRP data.

3.1. Initial Feature Extraction

CNNs have characteristics such as local connection, weight sharing, pooling and multi-layer structures, which can effectively reduce network complexity and improve generalization ability [15].

For HRRP-based target recognition, several practical considerations arise, including translation sensitivity, aspect sensitivity, and amplitude-scale sensitivity. CNNs employ convolution kernels for feature extraction and filter spatial position information through pooling operations, endowing the model with spatial transformation invariance. This characteristic of CNNs effectively mitigates translation sensitivity inherent in HRRP data. Considering the aspect sensitivity of HRRPs, research suggests that the average HRRP exhibits a smoother and more concise signal shape compared to individual HRRP profiles, potentially enhancing the capture of the target's scattering property in specific aspect frames [7]. From a signal processing standpoint, the average profile offers a stable representation of the target's physical structure within a frame, effectively reducing the speckle impact of HRRPs and mitigating the effects of noise spikes and amplitude fluctuations. Amplitude-scale sensitivity can be addressed through normalization techniques such as L_1 and L_2 normalization. In this study, prior to CNN-based initial feature extraction from the HRRP data, we first acquire the average profile of the target, and then, perform L_1 normalization on it.

Based on the pertinent literature [7], the average HRRP is defined as follows:

$$\mathbf{x}^{AP} = \left[\frac{1}{M} \sum_{i=1}^{M} x_{i1}, \frac{1}{M} \sum_{i=1}^{M} x_{i2}, \ldots, \frac{1}{M} \sum_{i=1}^{M} x_{ir} \right] \tag{1}$$

where $\{\mathbf{x}_i\}_{i=1}^{M}$ represents a real-value HRRP sequence following envelope alignment, with the ith HRRP sample $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ir}]$, and r is the dimension of HRRP samples. The average profile after normalization by L_1 norm is

$$\bar{\mathbf{x}}^{AP} = \frac{\mathbf{x}^{AP}}{\|\mathbf{x}^{AP}\|_1} \tag{2}$$

CNNs also play a vital role in leveraging ISAR images for target recognition, as they can abstractly extract target structure information from two-dimensional images through layer-by-layer learning. This structural information serves as a critical component in image recognition.

Therefore, the deep fusion network proposed in this paper extracts the initial features of the two modalities, namely, the average profiles and ISAR images, through CNNs. It is important to note that ISAR images are two-dimensional images, while the average profile of the target is one-dimensional. Consequently, the initial feature extraction network of ISAR images in this article uses a general two-dimensional convolutional neural network, while a one-dimensional convolutional neural network is designed to extract the initial features of average HRRP data.

Two detailed CNN structures are illustrated in Figure 3. Taking the CNN for ISAR images as an example, "Conc.64*11*11/ReLU" signifies that there are 64 feature maps, each with a kernel size of 11 × 11, followed by the rectified linear unit (ReLU) activation function. Additionally, "Max pool 3 × 3" denotes max pooling with a pooling size of 3 × 3, while "Fully connected 4096*10" indicates that the fully connected layer has 4096 input nodes and 10 output nodes. Finally, the softmax classifier generates the predicted label. The CNN for the HRRP data follows a similar structure, with the convolution and pooling operations applied in one dimension instead of two.

For the two CNN initial feature extraction networks, their inputs are ISAR images and normalized average profiles, respectively, and the outputs of the penultimate fully connected layer are used as the initial features. Therefore, the initial feature dimensions are 4096 and 1000 for ISAR images and average HRRP, respectively. Before proceeding, it is necessary to standardize them to the same dimensions.

3.2. Feature Separation

Considering our recognition task, where ISAR images and HRRP data depict target information from different aspects, there exists complementarity between the two modalities. By incorporating this complementarity, we can further enhance the model's prediction performance. In neural network architectures, two common feature-level fusion methods are the direct concatenation and addition, as shown in Figure 4. Concatenation combines different features to generate a mixed feature, resulting in a fused feature dimension equal to the sum of the original feature dimensions. Addition adds different features based on their element positions, maintaining the original feature dimensions. While these methods ensure that fused features contain all information from pre-fusion features, thus enhancing model robustness to some extent, they overlook the deep correlation and complementarity between the initial multimodal features.

To accomplish this objective, our method incorporates feature separation technology to construct a more resilient representation of the target. Initially, we address the disparity in initial feature lengths between the two modalities by standardizing them to the same dimension through a fully connected layer. Subsequently, we explicitly divide these features into private and shared feature spaces. We aim for the private features to be more complementary compared to the initial features: in instances where one modality's private feature is corrupted by noise or lacks essential information, the complementary private feature from another modality can effectively enhance target discriminability, thereby bolstering the recognition system's robustness. Moreover, we also expect that shared features will exhibit a higher correlation compared to the initial ones, which can furnish more abstract and informative representations common to both modalities [42]. In the feature separation part, the core idea is that we explicitly separate the feature of each modality into shared and private feature spaces. We regard the shared (private) features of the two modalities as samples of two distributions, so that the feature separation problem is transformed into a problem of constraining the distance between the distributions. During the training process, we obtain the shared feature spaces by minimizing the distribution distance between the two modal features to maximize the similarity. Similarly, the acquisition of private features relies on maximizing the distance to maximize the difference. Therefore, we use MK-MMD to access these similarities, as in [43], and we also propose to use it to measure the difference to obtain more complementary private feature information.

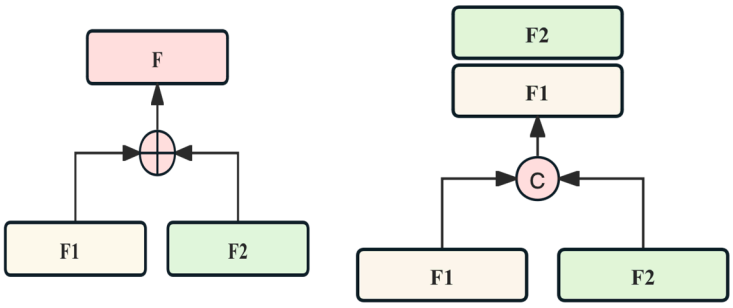


Figure 4. Two traditional feature-level fusion methods. (**a**) Element-wise addition; (**b**) direct concatenation operation.

Before introducing MK-MMD, we need to briefly introduce MMD. In the field of transfer learning, MMD loss [44] is commonly used to measure the distance between multi-domain feature distributions and demonstrates excellent performance. By minimizing the distribution difference of multi-domain features, the feature distributions of the source domain and the target domain become as similar as possible.

The general definition of MMD is as follows: given two probability distributions p and q, it is defined as

$$MMD(p,q) = \sup_{f \in \mathcal{F}} \left(E_p[f(x)] - E_q[f(y)] \right) \tag{3}$$

where f is a mapping function from feature space to real numbers, $\mathcal{F}$ is a set of feasible mapping functions, x is the source domain sample, y is the target domain sample, their distributions are, respectively, p and q, and $f(x)$ and $f(y)$ are the mapped values of the source domain and target domain samples, respectively. What this formula means is that the maximum value of the difference between the expected values of distribution p and distribution q under the mapping is called the MMD value. The specific form of the MMD formula is as follows:

$$MMD(p,q) = \| \frac{1}{n} \sum_{i=1}^{n} \phi(x_i) - \frac{1}{m} \sum_{j=1}^{m} \phi(y_j) \|_{\mathcal{H}} \tag{4}$$

where n and m represent the numbers of samples in the source and target domains, respectively. Determining an appropriate $\phi(x)$ to serve as the mapping function is critical to MMD. Nevertheless, the nature of this mapping function may vary across tasks, and it could potentially operate in a high-dimensional space, making its definition or selection challenging.

Fortunately, we can solve this problem with the help of the idea of the kernel function. We square the MMD, yet we continue to denote the label as MMD. We can expand it to obtain the following form:

$$\begin{aligned} MMD(p,q) &= \| \tfrac{1}{n} \sum_{i=1}^{n} \phi(x_i) - \tfrac{1}{m} \sum_{j=1}^{m} \phi(y_j) \|_{\mathcal{H}}^2 \\ &= \| \tfrac{1}{n^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} \phi(x_i)\phi(x'_i) - \tfrac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \phi(x_i)\phi(y_j) + \tfrac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} \phi(y_j)\phi(y'_i) \|_{\mathcal{H}} \end{aligned} \tag{5}$$

As shown in Formula (5), the MMD loss contains the inner product after feature mapping, which exactly meets the definition of the kernel function. By replacing all these inner products with the kernel function, the following formula can be obtained:

$$MMD(p,q) = \| \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} k(x_i, x'_i) - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} k(x_i, y_j) + \frac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} k(y_j, y'_i) \|_{\mathcal{H}} \tag{6}$$

In this way, with the help of the kernel function, we can skip the calculation of ϕ and directly use the kernel function to calculate the MMD loss function, making the calculation of the MMD feasible. Generally, different kernel functions, such as linear kernel and Gaussian kernel, can be used to define the MMD measuring method. Unlike MMD, MK-MMD uses multiple sets of mapping functions at different scales. This means that when calculating MMD, features at multiple different scales will be considered to better describe the difference between the two distributions. These functions at different scales can be tuned with parameters to achieve optimal performance on different tasks and datasets. One of the advantages of MK-MMD is its flexibility, and different kernel functions and scale parameters can be selected according to the specific problem. In summary, MK-MMD is an extension of MMD that allows the use of multiple kernel functions of different scales when measuring distribution differences to improve its performance and adaptability. MK-MMD expands the kernel function in Formula (6) into the sum of multiple kernel functions.

$$MK-MMD(p,q) = \sum_{l=1}^{L} \| \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} k_l(x_i, x'_i) - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} k_l(x_i, y_j) + \frac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} k_l(y_j, y'_i) \|_{\mathcal{H}} \tag{7}$$

where k_l is the lth kernel function, and L represents the number of kernel functions.

In our task, our objective is to maximize the similarity between the shared feature distributions s_1 and s_2 of the two modalities, while ensuring that the modality-private feature distributions p_1 and p_2 are distinct from each other. Therefore, during the training process we minimize $MK-MMD(s_1, s_2)$ and simultaneously maximize $MK-MMD(p_1, p_2)$. For simplicity, we write $MK-MMD(s_1, s_2)$ and $MK-MMD(p_1, p_2)$ as $d(s_1, s_2)$ and $d(p_1, p_2)$, respectively. The loss function of our feature separation module is as follows:

$$L_{fs} = d(s_1, s_2) - d(p_1, p_2) \tag{8}$$

After completing feature separation, we aim for both the shared and private components to consistently contribute to the discriminative support of the samples. However, in our task, it intuitively seems that there is no practically significant shared information between the two modalities; instead, each contains information about different aspects of the target. Therefore, we assume their correlation is low. When feature correlation is minimal, they may occupy disjoint distribution areas or have very limited overlap in the feature space, potentially resulting in distant distributions even for features of the same category. Additionally, the two modalities from different categories might exhibit similar feature distributions and the MMD loss function, used for global alignment, may inadvertently induce a "misalignment" phenomenon when extracting shared features, as depicted in Figure 5. This phenomenon may result in the fused features retaining similar information from different classes, making it challenging to distinguish differences between various categories. In our research, we refer to this phenomenon as "feature ambiguity". In such cases, incorporating shared feature components in predictive features could detrimentally affect model performance. Conversely, due to the global nature of the method, we utilize MK-MMD to maximize the distance between the feature distributions of the two modalities, typically resulting in the dispersion of the two private feature distributions across the feature space. The obtained private features serve as personalized representations for each modality, and their fusion can yield more distinct representations for each category. Hence, we opt to remove the shared feature branches post-feature separation, retaining solely the private feature components of both modalities for subsequent processing. This approach ensures a consistently positive impact on prediction performance.

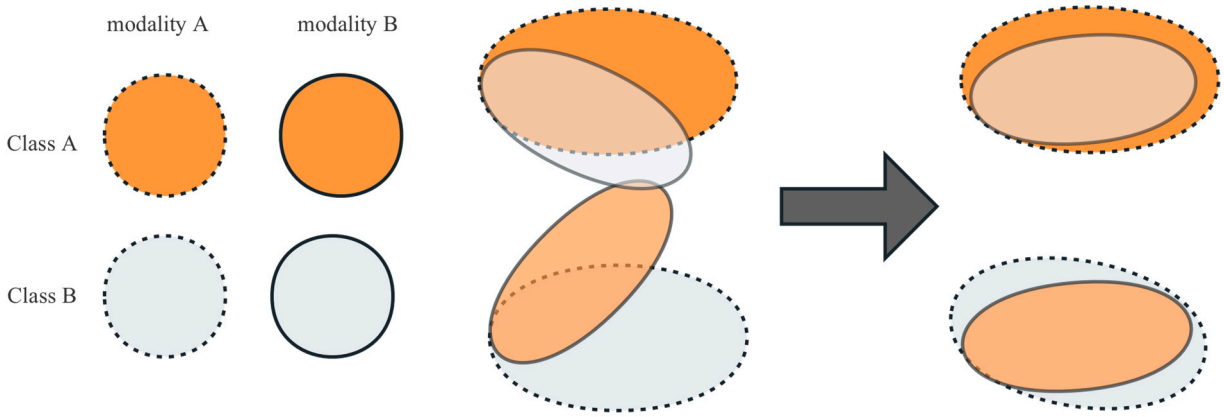


Figure 5. Misalignment when extracting shared features.

3.3. Weighted Decision Fusion

Decision-level fusion belongs to the category of intelligent information processing and is a high-level information fusion method. It analyzes and integrates multiple decision vectors to fully utilize a variety of relevant information about the target, resulting in more reliable and accurate decision results than those obtained from a single decision maker.

In the ISAR target recognition task, decisions based on a single classifier are heavily dependent on the performance of the current classifier. In complicated real-world circumstances, classifiers are easily influenced by external factors, leading to significant risks

and uncertainties in target recognition. Decision fusion can analyze and process multiple prediction vectors, alleviating misjudgments caused by relying on a single classifier to some extent. Therefore, the decision-level fusion method plays an important role in improving the reliability of recognition systems [45].

Weighted decision fusion methods offer increased flexibility in considering the contributions of different classifiers, thereby enhancing overall performance. In this study, we employ this strategy and advocate for using the MCP to determine the weight of each sub-classifier during weighted decision fusion. This weight determination approach is based on an assumption: when a sample category is misclassified by a sub-classifier, its decision vector likely exhibits a high probability similar to that of its misclassified category. Consequently, during misclassification, the maximum probability across the decision vector may be low, resulting in a smaller weight being assigned to the sub-classifier. This rationale aligns with practical scenarios, where an incorrectly classified classifier should indeed carry less influence. To achieve this objective, each sub-classifier undergoes training with a cross-entropy loss function, aiming to minimize the Kullback–Leibler divergence between the predicted distribution and the true distribution. This optimization process maximizes the classification performance of each sub-classifier, laying essential groundwork for the subsequent decision-making integration.

To obtain the classification confidence of different classifiers, assume that M classifiers $f^m : x_n^m \rightarrow y_n$ are constructed. Classifier f^m can be regarded as a probability model that converts sample x^m into probability distribution $\mathbf{p}^m(y|x^m) = \left(p_1^m \ldots p_k^m\right)$, and k represents the number of categories of the classification task. Therefore, the loss function of the weighted decision fusion module is as follows:

$$L_{wd} = -\sum_{m=1}^{M}\sum_{k=1}^{K} y_k \log(p_k^m) \tag{9}$$

where y_k is the true label, p_k^m is the softmax probability of classifier f^n for the kth class, and M represents the number of classifiers and is equal to three in our design. Assume that the prediction vectors of the t hree sub-classifiers are $\mathbf{d}^m = \left(p_1^m, p_2^m, \ldots, p_k^m\right), m = 1, 2, 3$; the weight of each sub-classifier in the weighted decision is derived from its prediction vector.

$$mcp^m = \max(p_1^m, p_2^m, \ldots, p_k^m) \tag{10}$$

Therefore, the final decision vector obtained by the MCP weighted decision is as follows.

$$\mathbf{d} = \sum_{i=1}^{3} mcp^i \mathbf{d}^i \tag{11}$$

3.4. Overall Loss Function

This article includes a total of three parts in the loss function: the first part is the loss function used by the feature separation module; the second part is the loss function used by the weighted decision fusion module; the last part is the cross-entropy loss function, required for the final decision vector. Therefore, the total loss function used in this article is as follows:

$$L = \alpha_{fs} L_{fs} + \alpha_{wd} L_{wd} + CE_{final} \tag{12}$$

where α_{fs}, α_{wd} are hyperparameters to balance the three terms. In our method, the transformed feature dimension of each modal feature is 2048 with the sizes of shared and private feature both being 1024. For MK-MMD, we use a linear combination of multiple Gaussian kernels. We set the number of Gaussian kernels to 5 to better form the shared and private features. The coefficients in the total loss function are all set to 1 without tuning. We trained the fusion model for 50 epochs using the Adam optimizer, with a learning rate 0.0001 and a batch size 32.

4. Experiments and Results

4.1. Simulated Data

For the simulation datasets, 3D ship models are used to generate ISAR images and corresponding HRRP data using computational electromagnetics software for recognition processing. The simulation radar parameters include a center frequency of 8.075 GHz, a bandwidth of 150 MHz, and a pulse repetition frequency (PRF) of 200 Hz, as shown in Table 1. Ten distinct 3D ship models are designed to construct the dataset. Notably, the ISAR images of each target type include top-view images at two pitch angles and side-view images at two azimuth angles. The HRRP data represent the average profiles after amplitude normalization during the imaging process. For top-view images, the ten target types are imaged at pitch angles of 80 degrees and 85 degrees, and for each pitch angle, we conducted two sets of experiments, each with its own yaw motion speed. Consequently, top-view images captured at each pitch angle exhibit two distinct cross-range resolutions. The experiment begins with the azimuth angle set to 5 degrees. ISAR top-view images are then generated at a fixed azimuth angle interval. Different azimuth angle intervals are configured for the two different yaw motion speeds, resulting in the generation of 25 ISAR images each. Thus, a total of 50 ISAR images are produced across both modes of movement at each pitch angle. Similarly, side-view ISAR images include imaging outcomes at two azimuth angles (10 degrees and 15 degrees), with the initial pitch angles set at 40 degrees, and 50 images captured at each azimuth angle under two different pitch motion modes. The imaging details are shown in Table 2. Figure 6 illustrates typical side-view ISAR images of ten ship targets at an azimuth angle of 10 degrees under ideal conditions, accompanied by their corresponding average HRRP shown in Figure 7, where T1 denotes target 1. The geometric relationship between pitch angle and azimuth angle is depicted in Figure 8, where the angle with the positive z-axis is denoted as the pitch angle θ, and the angle between the xOy-plane projection and the positive x-axis is marked as the azimuth angle φ, with the ship's bow facing the positive x-axis direction. As depicted in Figure 9, varying azimuth angles induce severe angle glints and diverse occlusion scenarios in ISAR images [46,47]. Under such conditions, relying solely on ISAR images for target recognition becomes challenging when missing angles occur.

Table 1. Settings of radar parameters for simulated data.

Parameter	Value
Center frequency	8.075 GHz
Bandwidth	150 MHz
PRF	200 Hz
Observation time	0.32 s

Table 2. The imaging details for top-view and side-view ISAR images.

Top-View			Side-View		
Target	T1	T2–T10	Target	T1	T2–T10
Pitch angle (θ)	80°/85°	80°/85°	Azimuth angle (φ)	10°/15°	10°/15°
Initial azimuth angle (φ)	5°	5°	Initial pitch angle (θ)	40°	40°
Azimuth motion 1	0.04°/s	0.27°/s	Pitch motion 1	0.08°/s	0.51°/s
Azimuth angle interval 1	0.02°	0.132°	Pitch angle interval 1	0.04°	0.240°
Azimuth motion 2	0.08°/s	0.54°/s	Pitch motion 2	0.16°/s	1.01°/s
Azimuth angle interval 2	0.04°	0.211°	Pitch angle interval 2	0.08°	0.384°

To verify the effectiveness and superiority of the algorithm under different signal-to-noise ratios (SNRs), we added three levels of Gaussian noise of 10 dB, 5 dB, and 3 dB to the original echo data. Meanwhile, to fully verify the robustness of the fusion method proposed in this article in the absence of angles, for each SNR data, we can obtain a total of four ways

to divide the training set and the test set, which we call missing_aspect15_pitch85, missing_aspect15_pitch80, missing_aspect10_pitch85, and missing_aspect10_pitch80. Since the four situations are similar, we take the case of missing_aspect15_pitch85 as an example. It means that our training set consists of side-view ISAR images with an azimuth angle of 10 degrees and top-view images with a pitch angle of 80 degrees, as well as their corresponding average profiles. The side-view ISAR image at an azimuth angle of 15 degrees and the top-view images at a pitch angle of 85 degrees and their corresponding average profiles are used as the test set. This allows us to fully verify the robustness of the fusion method proposed in this article in the absence of angles. For the generality of the results and the simplicity of the presentation of the results, in experiments with simulated data we use the average of the recognition rates in the four cases under each SNR as the display results in our experiments.

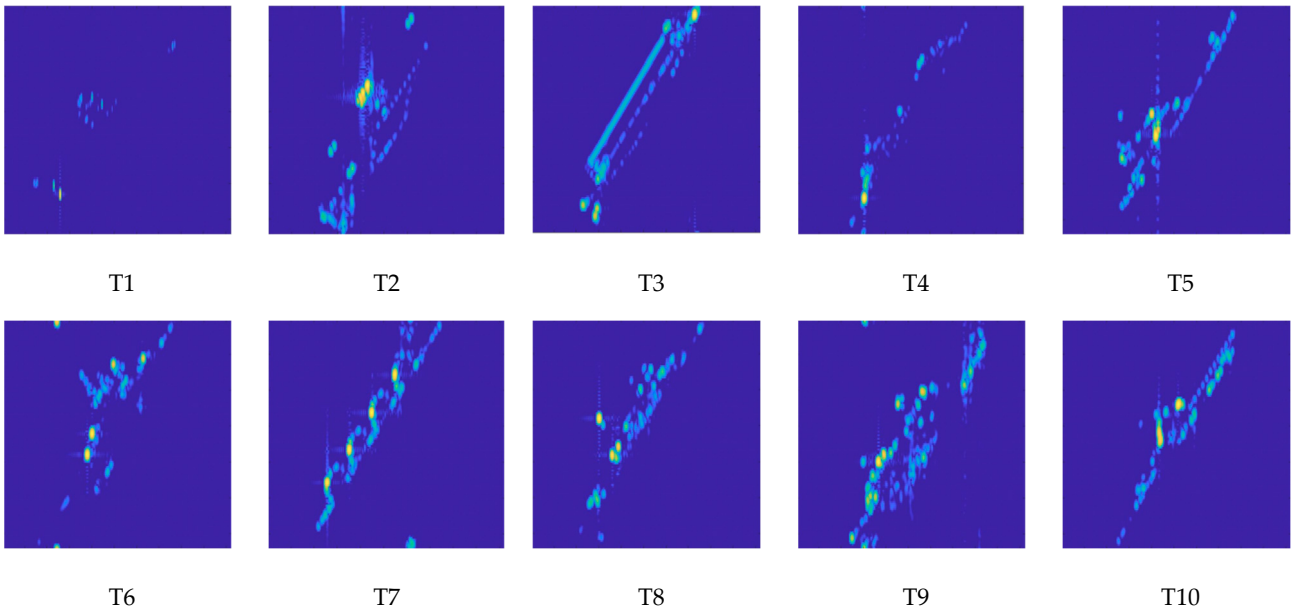


Figure 6. Typical ISAR images of targets in the simulated dataset.

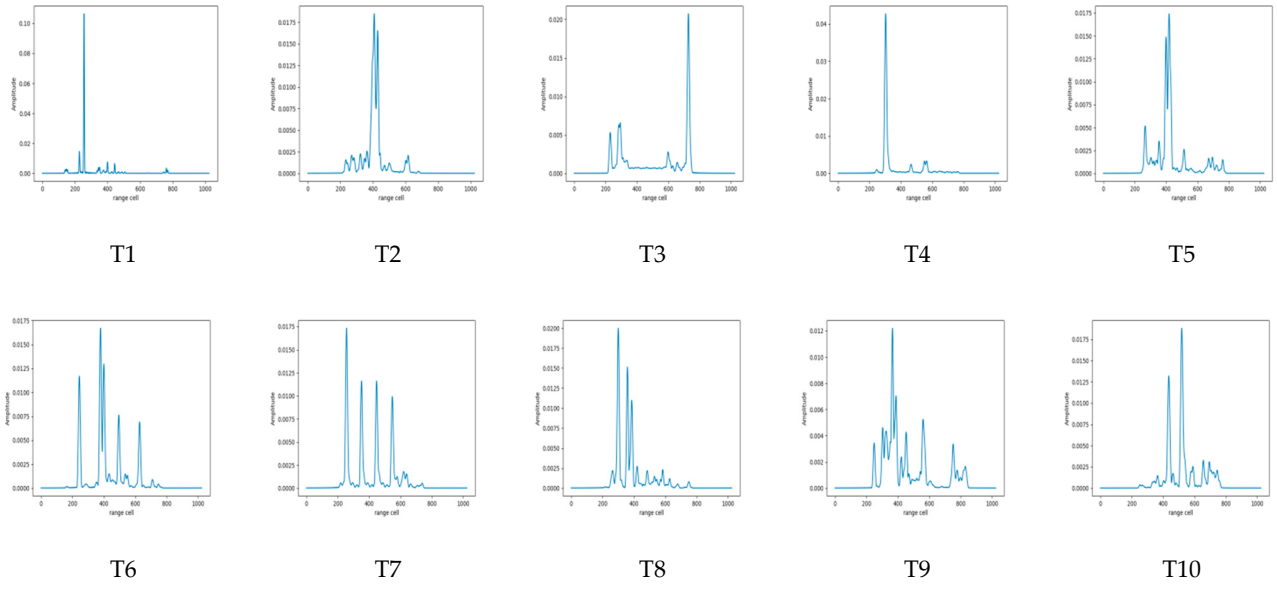


Figure 7. Average HRRPs of targets in the simulated dataset.

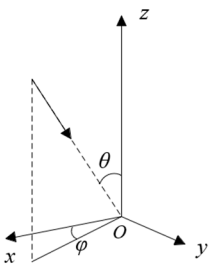


Figure 8. Geometric relations.

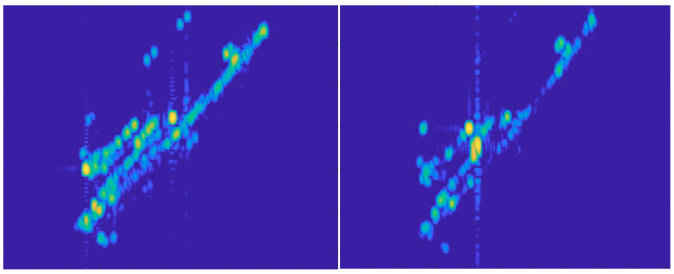


Figure 9. ISAR images with different azimuth angles.

The recognition accuracies of the proposed method are shown in Table 3. In order to verify the generality of the results and the simplicity in presenting the results, for the experimental part on the simulated data the results shown are the average value of the recognition rates under four data division conditions unless otherwise specified.

Table 3. The recognition accuracy with different SNRs and division methods for the proposed method.

	Missing_ Aspect15_Pitch85	Missing_ Aspect15_Pitch80	Missing_ Aspect10_Pitch85	Missing_ Aspect10_Pitch80	Average
3 dB	91.34%	97.06%	91.90%	96.08%	94.09%
5 dB	92.53%	97.33%	94.56%	96.15%	95.14%
10 dB	94.45%	97.61%	95.67%	97.53%	96.32%

To showcase the efficacy of our proposed approach, we conducted comparative experiments employing various algorithms. Ensuring fairness, we standardized the initial features across different fusion methods by transforming the initial features of each modality into identical dimensions. We evaluated several fusion strategies, including early fusion, where the ISAR image and its corresponding average profile are concatenated and fed directly into a CNN. Additionally, we explored two common feature-level fusion methods [48,49]: the first method involved concatenating the transformed initial features of both modalities before feeding them into the prediction layer. The second method utilized element-wise addition for feature fusion, the extended-GRU (Ex-GRU) [11] fusion method, which adaptively learns the weight of each modal feature to distinguish the contribution to sample discrimination; the research content of this article is similar to our research content, and the fusion method in the article has achieved better results than many other methods. Additionally, we utilized a mask to assess the impact of shared feature branches and private feature branches after feature separation on sample discriminability. Our findings revealed that not all branches consistently enhanced prediction performance. When prediction features included shared components, there was a modest decrease in prediction accuracy. We conducted comparative experiments across various SNRs and data partitions, with the accuracy of each method presented in Table 4. The "mask shared" approach involves masking shared feature branches, retaining only private feature parts after separation, and then, processing these private features using element-wise addition before feeding them to the prediction layer. Similarly, "mask private" follows a comparable procedure.

Notably, "no mask" initially performs an element-wise addition operation on the shared feature branches of the two modalities, concatenates them with private features [44], and subsequently feeds the concatenated features to the prediction layer.

Table 4. Comparisons of other fusion methods with the proposed method on simulated data with different SNRs.

	3 dB	5 dB	10 dB
Early fusion	77.24%	82.53%	85.28%
Concatenation	89.17%	91.40%	91.98%
Addition	88.39%	90.70%	90.88%
Ex-GRU [11]	87.48%	90.85%	91.13%
Mask private	91.03%	92.30%	92.53%
Mask shared	93.20%	94.15%	95.51%
No mask	92.43%	93.36%	93.95%
SDRnet (Ours)	**94.09%**	**95.14%**	**96.32%**

Figure 10 visually demonstrates the superior robustness of the method proposed in this article compared to previous methods, particularly concerning angle missing and noise. In our experiments evaluating the effectiveness of shared and private features, we observed a modest decrease in overall prediction performance when the prediction features included shared components. This decline may stem from the inferred ambiguity of the shared features. To substantiate our inference, we present the recognition rates of "mask private", "mask shared", and "no mask" under the "missing_aspect15_pitch85" condition, as outlined in Table 5. Furthermore, we offer corresponding visual feature distribution maps using t-distributed stochastic neighbor embedding (t-SNE). We specifically chose this case as the ambiguity of shared features becomes particularly evident under this partition condition.

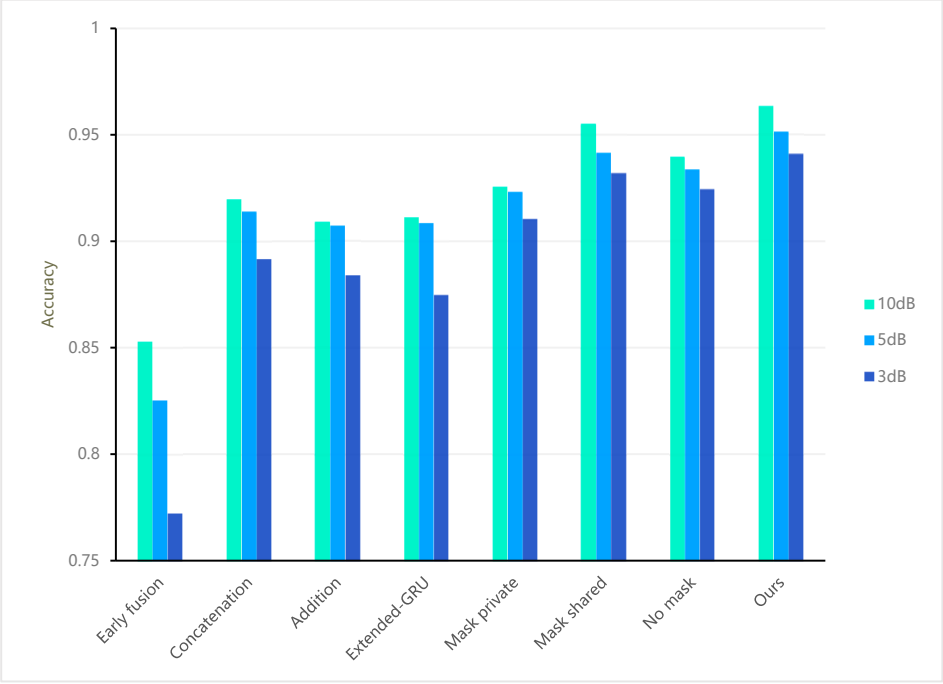


Figure 10. Comparisons of other fusion methods with the proposed method on simulated data with different SNRs.

Table 5. The recognition accuracy in the case of missing_aspect15_pitch85.

	3 dB	5 dB	10 dB
Mask private	84.31%	86.63%	88.92%
Mask shared	**89.35%**	**90.20%**	**93.15%**
No mask	87.50%	87.85%	89.26%

As can be seen from Figure 11, in our recognition task, the shared features obtained through MK-MMD auxiliary function constraints exhibit feature ambiguity in some categories, confirming our hypothesis. In this case, only private features with complementary properties would be used as predictive features to enhance the prediction performance of the model rather than using private and shared features at the same time. Therefore, to achieve better prediction performance, our method chooses to mask the shared feature branches and only retain the private feature parts after feature separation.

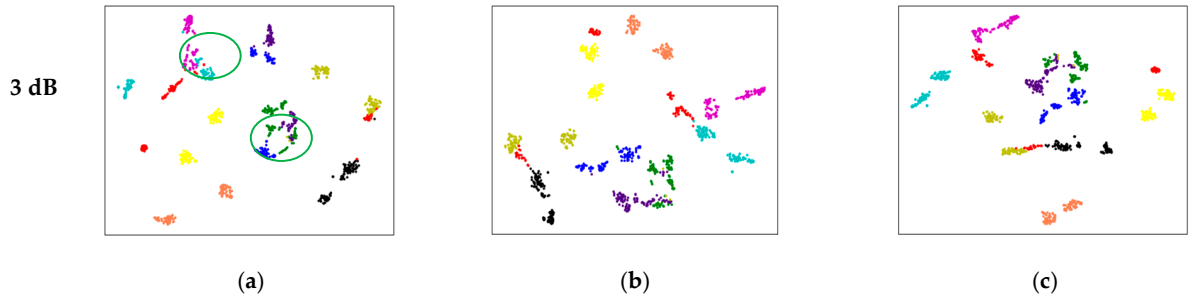


Figure 11. *Cont.*

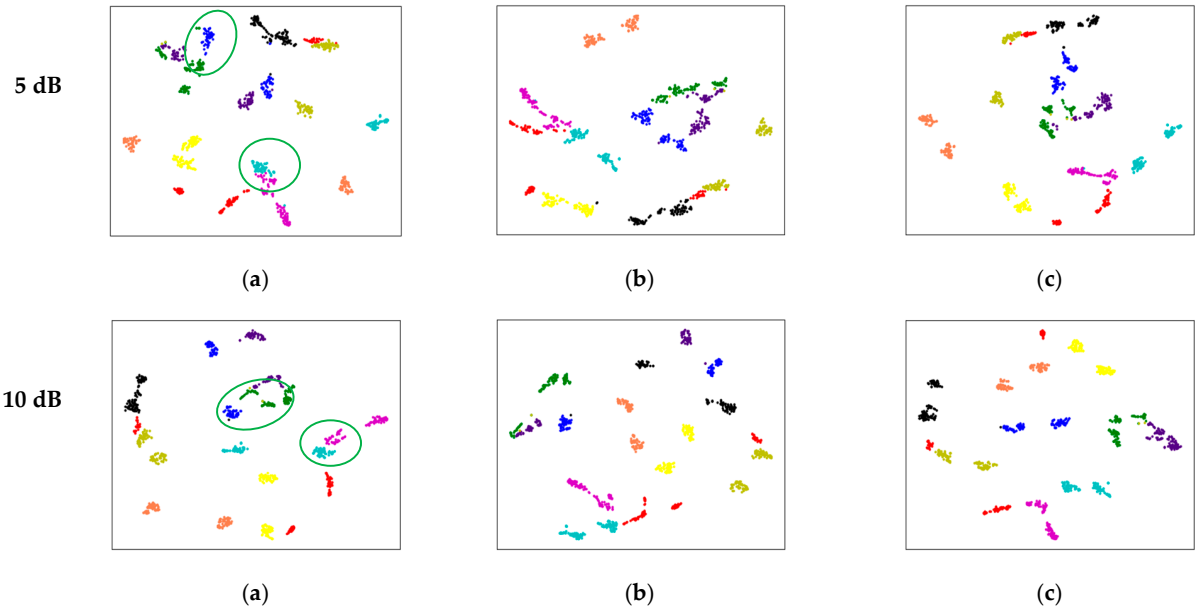


Figure 11. T-SNE visualization in the case of missing_aspect15_pitch85 when performing (**a**) mask private; (**b**) no mask; (**c**) mask shared. Different colors represent different sample categories.

4.2. Measured Data

We further verified the effectiveness of the proposed method using measured data. The measured data were acquired using an X-band radar. The measured data also contain ten categories of target data, each category having different imaging time periods. Typical ISAR images of the ten types of targets are shown in Figure 12, and the corresponding

average profiles are listed in Figure 13. In this experiment, part of the data in each category were used as the training set, and the rest were used as the test set. The number of training and test sets for the ten categories of targets is shown in Table 6, where T1 represents Target 1.

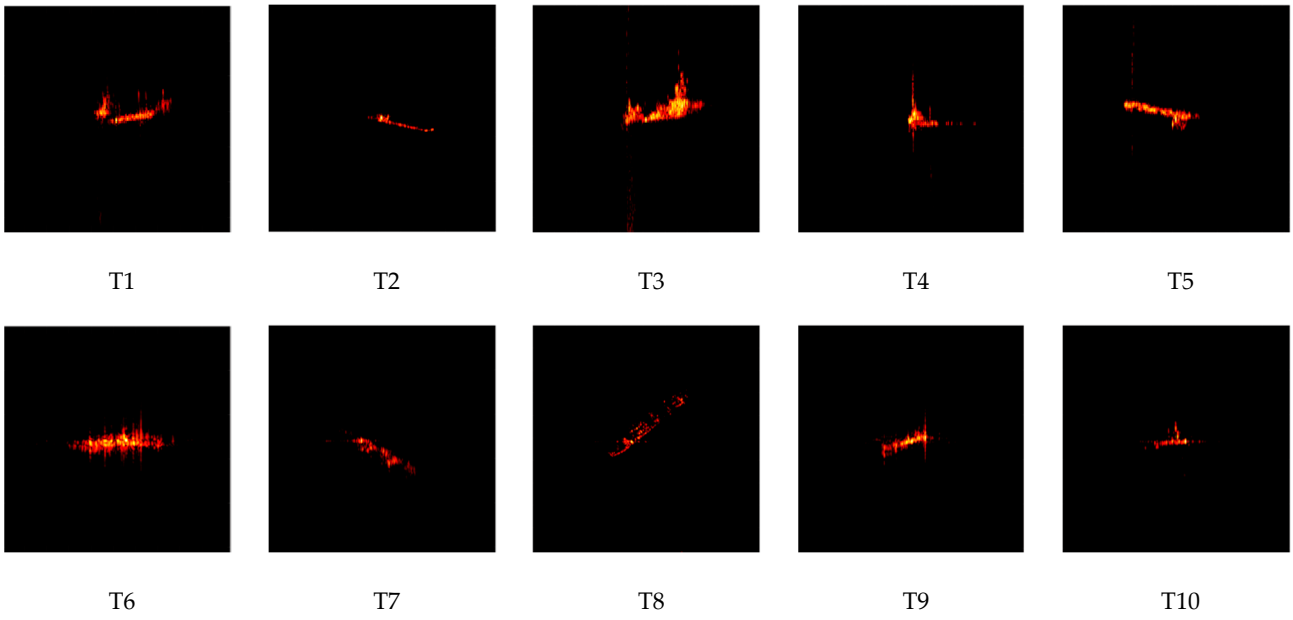


Figure 12. ISAR images of targets in the measured dataset.

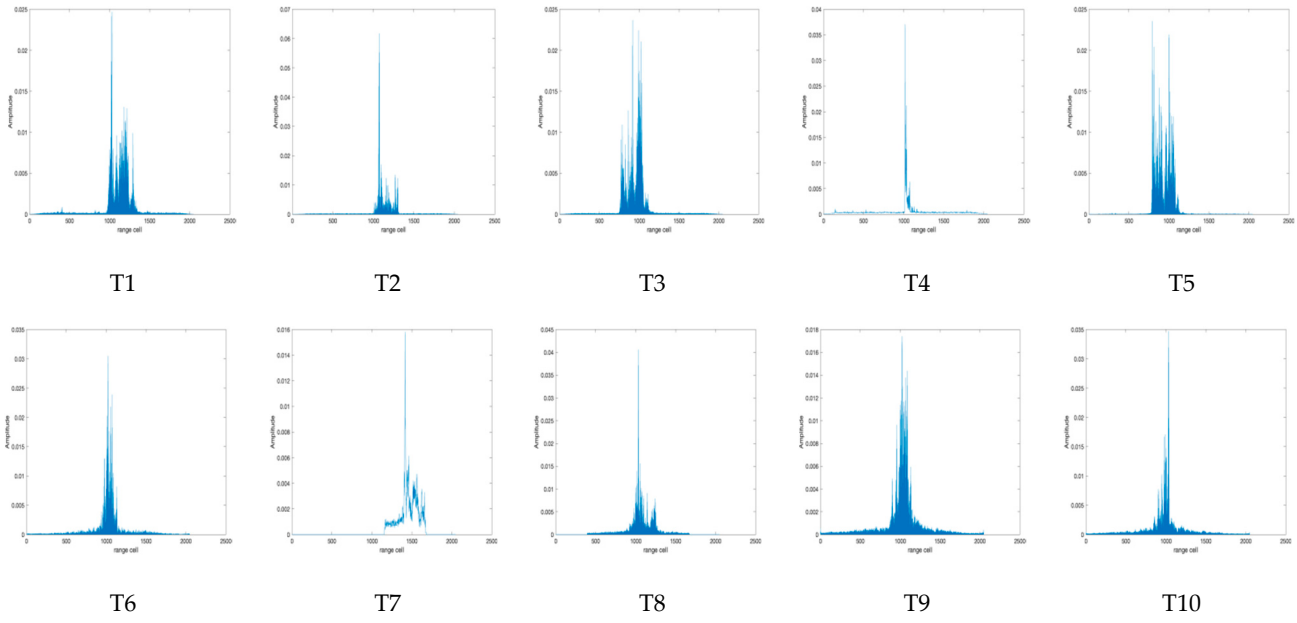


Figure 13. Average HRRPs of the targets in the measured dataset.

Table 6. Number of training and test samples for targets in the measured dataset.

Dataset	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Training Samples	150	91	91	200	127	73	88	49	80	93
Test Samples	350	90	189	285	222	97	180	50	21	32

Similar to the analysis conducted on the simulated dataset, the effectiveness of our proposed method was also evaluated using measured data. Consistent with the methodology applied to the simulated data, the comparison included similar methods. The recognition outcomes are detailed in Table 7. The superior recognition accuracy attained by our method further underscores its efficacy compared to alternative approaches.

Table 7. Comparisons of other fusion methods with the proposed method on measured data.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	Accuracy (%)
Early fusion	90.00	100	95.76	99.65	87.84	100	93.88	96.00	47.62	84.38	93.40
Concatenation	82.57	100	98.94	100	95.05	100	91.67	100	76.19	68.75	93.14
Addition	80.29	100	98.94	100	94.59	100	91.67	92.00	61.90	68.75	92.08
Ex-GRU [11]	81.43	100	96.83	100	98.20	100	91.67	94.00	85.71	56.25	92.74
Mask private	85.14	97.78	100	100	98.65	100	91.67	84.00	66.67	78.13	93.80
Mask shared	89.71	100	100	100	98.20	100	91.67	94.00	66.67	65.62	94.85
No mask	87.43	100	96.30	100	98.65	98.97	91.11	96.00	85.71	71.88	94.39
SDRnet (ours)	92.00	96.67	100	100	99.55	100	90.56	94.00	90.48	68.75	**95.78**

4.3. Ablation Study

To gain deeper insight into the network's functionality and validate the advantages of fusing HRRP data and ISAR images for ISAR target recognition, an ablation study can be employed. In this section, we aim to investigate the efficacy and superiority of integrating two modalities for target recognition in contrast to single-modal target recognition. We conducted ablation experiments on both measured data and simulated data across three SNRs. We employed a CNN for classifying the ISAR images of targets and utilized a 1D-CNN for identifying the average profile. The experimental results are shown in Table 8. It is evident that both simulated and measured data yield low recognition accuracy rates when utilizing only a single modality for target recognition. Whereas, our fusion method, which combines the two modalities, achieves higher recognition accuracy.

Table 8. Ablation study.

	Only Image	Only HRRP	Proposed Fusion Method	Accuracy (%)	
3 dB	√	×	×	82.48	Simulated
	×	√	×	83.03	
	√	√	√	**94.09**	
5 dB	√	×	×	83.48	
	×	√	×	86.23	
	√	√	√	**95.14**	
10 dB	√	×	×	86.33	
	×	√	×	87.18	
	√	√	√	**96.32**	
	√	×	×	90.22	Measured
	×	√	×	89.12	
	√	√	√	**95.78**	

5. Discussion

5.1. Comparison

Table 4 and Figure 10 present the identification performance of our proposed method compared to other methods on a simulated dataset, while Table 7 shows the performance on a measured dataset. Our method consistently achieves higher identification accuracy relative to other approaches. Firstly, after feature separation, our model obtains more robust representations of the target, thereby improving identification accuracy. This enhancement is evident in the third row and the fifth to seventh rows of Tables 4 and 7. Secondly, with

the incorporation of weighted decision making, the model can integrate the decisions or prediction results of multiple classifiers, enhancing the reliability of the model's decisions. This further improves the model's prediction performance, as demonstrated in the sixth and eighth rows of Tables 4 and 7.

5.2. Feature Ambiguity

Table 5 and Figure 11 illustrate the influence of shared feature ambiguity that may occur after separation. In our task, it intuitively appears that there is no practically significant shared information between the two modalities; instead, each modality contains information about different aspects of the target. Therefore, we assume that their correlation is low. When feature correlation is very low, the features may occupy disjointed distribution areas or have very limited overlap in the feature space, potentially resulting in distant distributions even for features of the same category. Additionally, the two modalities from different categories might exhibit similar feature distributions. The MMD loss function, utilized for global alignment, may inadvertently induce a "misalignment" phenomenon, as shown in Figure 5. This phenomenon may result in the fused features retaining similar information from different classes, thereby making it challenging to distinguish differences between various categories. In such cases, incorporating shared feature components in predictive features could detrimentally affect model performance.

To address this, we utilized a mask to assess the impact of shared and private feature branches on sample discriminability after feature separation. As shown in rows 5 to 7 of Tables 4 and 7, as well as in Table 5, the "mask private" method consistently results in the lowest accuracy among the three methods. To understand this, we employed T-SNE to visualize the distribution of shared features. As depicted in Figure 11, a feature ambiguity phenomenon indeed exists, confirming our hypothesis. Consequently, our method opts to mask the shared feature branches and only retain the private feature parts after feature separation to acquire a stable prediction performance.

5.3. Ablation

A series of ablation experiments were conducted on both simulated and measured datasets to demonstrate that combining two modalities for target identification yields higher accuracy compared to single-modality identification. The results of these ablation experiments are presented in Table 8.

5.4. Future Work

Non-target areas inevitably exist in both ISAR imagery and HRRP data. When these non-target areas contain noise, it may affect recognition accuracy, making it a significant issue to investigate. Currently, our method does not consider the impact of redundant information in non-target areas on target recognition. Therefore, our future research will focus on addressing this aspect.

6. Conclusions

In this study, we introduced the SDRnet, a novel deep fusion network designed for ISAR ship target recognition by leveraging feature separation and weighted decision. By acknowledging the inherent differences between ISAR images and HRRP data, our proposed method aims to provide a more robust representation of the target by considering the deep correlation and complementarity between the features of these two modalities. Through experimental evaluation on both simulated and measured datasets, our results consistently demonstrate the superior performance of our fusion method compared to conventional approaches. These findings underscore the efficacy of our approach in ship target recognition, highlighting its potential for practical applications in real-world scenarios.

Author Contributions: Conceptualization, J.D and F.S.; methodology, J.D.; software, J.D.; validation, J.D.; formal analysis, J.D.; investigation, J.D.; resources, F.S.; data curation, J.D.; writing—original

draft preparation, J.D.; writing—review and editing, J.D.; visualization, J.D.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are not publicly available due to the request of the data owner.

Acknowledgments: The authors would like to thank the editors and anonymous reviewers for their competent comments and suggestions to improve this article. And special thanks are given to Xinfei Jin for the help with the simulated dataset.

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

References

1. Xue, R.; Bai, X.; Zhou, F. SAISAR-Net: A robust sequential adjustment ISAR image classification network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [CrossRef]
2. Ni, P.; Liu, Y.; Pei, H.; Du, H.; Li, H.; Xu, G. Clisar-net: A deformation-robust isar image classification network using contrastive learning. *Remote Sens.* **2022**, *15*, 33. [CrossRef]
3. Yan, H.; Zhang, Z.; Xiong, G.; Yu, W. Radar HRRP recognition based on sparse denoising autoencoder and multi-layer perceptron deep model. In Proceedings of the 2016 Fourth International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS), Shanghai, China, 2–4 November 2016; pp. 283–288.
4. Du, C.; Chen, B.; Xu, B.; Guo, D.; Liu, H. Factorized discriminative conditional variational auto-encoder for radar HRRP target recognition. *Signal Process.* **2019**, *158*, 176–189. [CrossRef]
5. Du, L.; Liu, H.; Bao, Z.; Zhang, J. Radar automatic target recognition using complex high-resolution range profiles. *IET Radar Sonar Navig.* **2007**, *1*, 18–26. [CrossRef]
6. Du, L.; Liu, H.; Wang, P.; Feng, B.; Pan, M.; Bao, Z. Noise robust radar HRRP target recognition based on multitask factor analysis with small training data size. *IEEE Trans. Signal Process.* **2012**, *60*, 3546–3559.
7. Xing, M.; Bao, Z.; Pei, B. Properties of high-resolution range profiles. *Opt. Eng.* **2002**, *41*, 493–504. [CrossRef]
8. Pan, M.; Liu, A.; Yu, Y.; Wang, P.; Li, J.; Liu, Y.; Lv, S.; Zhu, H. Radar HRRP target recognition model based on a stacked CNN–Bi-RNN with attention mechanism. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
9. Feng, B.; Chen, B.; Liu, H. Radar HRRP target recognition with deep networks. *Pattern Recognit.* **2017**, *61*, 379–393. [CrossRef]
10. Chen, J.; Du, L.; Guo, G.; Yin, L.; Wei, D. Target-attentional CNN for radar automatic target recognition with HRRP. *Signal Process.* **2022**, *196*, 108497. [CrossRef]
11. Du, L.; Li, L.; Guo, Y.; Wang, Y.; Ren, K.; Chen, J. Two-stream deep fusion network based on VAE and CNN for synthetic aperture radar target recognition. *Remote Sens.* **2021**, *13*, 4021. [CrossRef]
12. Liao, L.; Du, L.; Chen, J. Class factorized complex variational auto-encoder for HRR radar target recognition. *Signal Process.* **2021**, *182*, 107932. [CrossRef]
13. Pan, M.; Jiang, J.; Kong, Q.; Shi, J.; Sheng, Q.; Zhou, T. Radar HRRP target recognition based on t-SNE segmentation and discriminant deep belief network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1609–1613. [CrossRef]
14. Chen, J.; Du, L.; Liao, L. Discriminative mixture variational autoencoder for semisupervised classification. *IEEE Trans. Cybern.* **2020**, *52*, 3032–3046. [CrossRef] [PubMed]
15. Wan, J.; Chen, B.; Xu, B.; Liu, H.; Jin, L. Convolutional neural networks for radar HRRP target recognition and rejection. *EURASIP J. Adv. Signal Process.* **2019**, *2019*, 5. [CrossRef]
16. Sathyendra, H.M.; Stephan, B.D. Data fusion analysis for maritime automatic target recognition with designation confidence metrics. In Proceedings of the 2015 IEEE Radar Conference (RadarCon), Arlington, VA, USA, 10–15 May 2015; pp. 0062–0067.
17. Manno-Kovacs, A.; Giusti, E.; Berizzi, F.; Kovács, L. Automatic target classification in passive ISAR range-crossrange images. In Proceedings of the 2018 IEEE Radar Conference (RadarConf18), Oklahoma, OK, USA, 23–27 April 2018; pp. 0206–0211.
18. Jarabo-Amores, P.; Giusti, E.; Rosa-Zurera, M.; Bacci, A.; Capria, A.; Mata-Moya, D. Target classification using passive radar ISAR imagery. In Proceedings of the 2017 European Radar Conference (EURAD), Nuremberg, Germany, 11–13 October 2017; pp. 155–158.
19. Kurowska, A.; Kulpa, J.S.; Giusti, E.; Conti, M. Classification results of ISAR sea targets based on their two features. In Proceedings of the 2017 Signal Processing Symposium (SPSympo), Jachranka, Poland, 12–14 September 2017; pp. 1–6.
20. Kawahara, T.; Toda, S.; Mikami, A.; Tanabe, M. Automatic ship recognition robust against aspect angle changes and occlusions. In Proceedings of the 2012 IEEE Radar Conference, Atlanta, GA, USA, 7–11 May 2012; pp. 0864–0869.
21. Bai, X.; Zhou, X.; Zhang, F.; Wang, L.; Xue, R.; Zhou, F. Robust pol-ISAR target recognition based on ST-MC-DCNN. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9912–9927. [CrossRef]
22. Zhao, W.; Heng, A.; Rosenberg, L.; Nguyen, S.T.; Hamey, L.; Orgun, M. ISAR ship classification using transfer learning. In Proceedings of the 2022 IEEE Radar Conference (RadarConf22), New York, NY, USA, 21–25 March 2022; pp. 1–6.
23. Lu, W.; Zhang, Y.; Yin, C.; Lin, C.; Xu, C.; Zhang, X. A deformation robust ISAR image satellite target recognition method based on PT-CCNN. *IEEE Access* **2021**, *9*, 23432–23453. [CrossRef]

24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*. [CrossRef]
25. Stiller, C.; Leon, F.P.; Kruse, M. Information fusion for automotive applications–An overview. *Inf. Fusion* **2011**, *12*, 244–252. [CrossRef]
26. Jiang, L.; Yan, L.; Xia, Y.; Guo, Q.; Fu, M.; Lu, K. Asynchronous multirate multisensor data fusion over unreliable measurements with correlated noise. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 2427–2437. [CrossRef]
27. Rasti, B.; Ghamisi, P.; Plaza, J.; Plaza, A. Fusion of hyperspectral and LiDAR data using sparse and low-rank component analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6354–6365. [CrossRef]
28. Bassford, M.; Painter, B. Intelligent bio-environments: Exploring fuzzy logic approaches to the honeybee crisis. In Proceedings of the 2016 12th International Conference on Intelligent Environments (IE), London, UK, 14–16 September 2016; pp. 202–205.
29. Choi, I.O.; Jung, J.H.; Kim, S.H.; Kim, K.T.; Park, S.H. Classification of targets improved by fusion of the range profile and the inverse synthetic aperture radar image. *Prog. Electromagn. Res.* **2014**, *144*, 23–31. [CrossRef]
30. Wang, L.; Weng, L.; Xia, M.; Liu, J.; Lin, H. Multi-resolution supervision network with an adaptive weighted loss for desert segmentation. *Remote Sens.* **2021**, *13*, 2054. [CrossRef]
31. Guan, R.; Li, Z.; Tu, W.; Wang, J.; Liu, Y.; Li, X.; Tang, C.; Feng, R. Contrastive Multi-view Subspace Clustering of Hyperspectral Images based on Graph Convolutional Networks. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14.
32. Guan, R.; Li, Z.; Li, X.; Tang, C. Pixel-Superpixel Contrastive Learning and Pseudo-Label Correction for Hyperspectral Image Clustering. *arXiv* **2023**, arXiv:2312.09630.
33. Shang, R.; Zhang, J.; Jiao, L.; Li, Y.; Marturi, N.; Stolkin, R. Multi-scale adaptive feature fusion network for semantic segmentation in remote sensing images. *Remote Sens.* **2020**, *12*, 872. [CrossRef]
34. Guan, R.; Li, Z.; Li, T.; Li, X.; Yang, J.; Chen, W. Classification of heterogeneous mining areas based on rescapsnet and gaofen-5 imagery. *Remote Sens.* **2022**, *14*, 3216. [CrossRef]
35. Chen, J.; He, F.; Zhang, Y.; Sun, G.; Deng, M. SPMF-Net: Weakly supervised building segmentation by combining superpixel pooling and multi-scale feature fusion. *Remote Sens.* **2020**, *12*, 1049. [CrossRef]
36. Liu, J.; Guan, R.; Li, Z.; Zhang, J.; Hu, Y.; Wang, X. Adaptive multi-feature fusion graph convolutional network for hyperspectral image classification. *Remote Sens.* **2023**, *15*, 5483. [CrossRef]
37. Li, X.; Ran, J.; Wen, Y.; Wei, S.; Yang, W. MVFRnet: A Novel High-Accuracy Network for ISAR Air-Target Recognition via Multi-View Fusion. *Remote Sens.* **2023**, *15*, 3052. [CrossRef]
38. Li, R.; Hu, Y.; Li, L.; Guan, R.; Yang, R.; Zhan, J.; Cai, W.; Wang, Y.; Xu, H.; Li, L. SMWE-GFPNNet: A high-precision and robust method for forest fire smoke detection. *Knowl.-Based Syst.* **2024**, *289*, 111528. [CrossRef]
39. Bousmalis, K.; Trigeorgis, G.; Silberman, N.; Krishnan, D.; Erhan, D. Domain separation networks. *Adv. Neural Inf. Process. Syst.* **2016**, *29*.
40. Lee, M.; Pavlovic, V. Private-shared disentangled multimodal vae for learning of latent representations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 1692–1700.
41. Wu, F.; Jing, X.Y.; Wu, Z.; Ji, Y.; Dong, X.; Luo, X.; Huang, Q.; Wang, R. Modality-specific and shared generative adversarial network for cross-modal retrieval. *Pattern Recognit.* **2020**, *104*, 107335. [CrossRef]
42. Wang, J.; Wang, Z.; Tao, D.; See, S.; Wang, G. Learning common and specific features for RGB-D semantic segmentation with deconvolutional networks. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part V 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 664–679.
43. van Amsterdam, B.; Kadkhodamohammadi, A.; Luengo, I.; Stoyanov, D. Aspnet: Action segmentation with shared-private representation of multiple data sources. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 2384–2393.
44. Long, M.; Cao, Y.; Wang, J.; Jordan, M. Learning transferable features with deep adaptation networks. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 97–105.
45. Glodek, M.; Tschechne, S.; Layher, G.; Schels, M.; Brosch, T.; Scherer, S.; Kächele, M.; Schmidt, M.; Neumann, H.; Palm, G.; et al. Multiple classifier systems for the classification of audio-visual emotional states. In Proceedings of the Affective Computing and Intelligent Interaction: Fourth International Conference, ACII 2011, Memphis, TN, USA, 9–12 October 2011; Proceedings, Part II. Springer: Berlin/Heidelberg, Germany, 2011; pp. 359–368.
46. Jin, X.; Su, F. Aircraft Recognition Using ISAR Image Based on Quadrangle-points Affine Transform. In Proceedings of the 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 5–7 November 2022; pp. 1–6.
47. Jin, X.; Su, F.; Li, H.; Xu, Z.; Deng, J. Automatic ISAR Ship Detection Using Triangle-Points Affine Transform Reconstruction Algorithm. *Remote Sens.* **2023**, *15*, 2507. [CrossRef]
48. Zadeh, A.; Chen, M.; Poria, S.; Cambria, E.; Morency, L.P. Tensor fusion network for multimodal sentiment analysis. *arXiv* **2017**, arXiv:1707.07250.
49. Liu, Z.; Shen, Y.; Lakshminarasimhan, V.B.; Liang, P.P.; Zadeh, A.; Morency, L.P. Efficient low-rank multimodal fusion with modality-specific factors. *arXiv* **2018**, arXiv:1806.00064.