

Article

An Improved Recognition Technique for Ship Targets Based on Dual-Path Cooperative Fusion Mechanism

Sifan Su [†], Wei Yang ^{*,†}, Shiwen Lei , Xiaozhang Zhu , Jing Tian  and Haoquan Hu

School of Electronic Science and Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China; 202421021221@std.uestc.edu.cn (S.S.); swlei@uestc.edu.cn (S.L.); xzzhu@uestc.edu.cn (X.Z.); j.tian@uestc.edu.cn (J.T.); huhq@uestc.edu.cn (H.H.)

* Correspondence: wyang@uestc.edu.cn

[†] These authors contributed equally to this work.

Highlights

What are the main findings?

- A mechanism-data-physics trinity recognition framework is developed to extract deep discriminative features from raw HRRP sequences, RCS sequences, and HRRP statistical features for the ship target recognition problem.
- A dual-path cooperative fusion mechanism is designed to realize deep complementarity and enhancement of multi-modal information.

What is the implication of the main finding?

- The dual-path cooperative method can achieve explicit inter-modal correlation modeling through a cross-attention module and dynamically learn the importance of each modality through an adaptive weight fusion layer.
- The proposed method outperforms existing single-source and fixed-weight fusion methods under various signal-to-noise ratios and polarization conditions.

Abstract

With the increasing complexity of the electromagnetic environment, traditional radar target recognition methods face severe challenges. High-resolution range profile (HRRP) and Radar Cross Section (RCS), as two important radar features, each has its own advantages in target recognition but also exhibits limitations. To enhance radar target recognition performance in complex scenarios such as low signal-to-noise ratio (SNR), this paper proposes a recognition method based on heterogeneous multi-modal feature fusion. The proposed method constructs a three-channel parallel encoding network, which utilizes Convolutional Long Short-Term Memory (ConvLSTM), One-Dimensional Convolutional Gated Recurrent Unit (Conv1D-GRU), and Gated Recurrent Unit (GRU) to extract deep discriminative features from raw HRRP sequences, RCS sequences, and HRRP statistical features, respectively. Furthermore, it innovatively designs a dual-path cooperative fusion mechanism, achieving explicit inter-modal correlation modeling through a cross-attention module and dynamically learning the importance of each modality through an adaptive weight fusion layer, thereby realizing deep complementarity and enhancement of multi-modal information. Experimental results demonstrate that under various signal-to-noise ratios and polarization conditions, the proposed method achieves a maximum average recognition accuracy of over 99% for 6 ship targets. Compared with the traditional three-channel fixed-weight fusion method, the recognition accuracy of the proposed method increases from 90.01% to 99.42% under co-polarization, and from 81.14% to 98.15% under cross-polarization, fully validating the effectiveness and superiority of the dual-path fusion mechanism.



Academic Editor: Dusan Gleich

Received: 9 April 2026

Revised: 8 July 2026

Accepted: 13 July 2026

Published: 24 July 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.

Keywords: radar target recognition; HRRP; RCS; multi-modal fusion; cross-attention mechanism

1. Introduction

Driven by the complex electromagnetic environment and diverse target recognition requirements in modern battlefields, high-precision and high-robustness target recognition technologies have become critical research topics [1]. Traditionally, recognition methods based on optical or infrared images can achieve effective identification under ideal conditions [2–4]. However, these methods suffer from severe degradation under adverse weather or night-time conditions, making them unsuitable for all-weather operations.

In contrast, radar systems provide reliable information sources for robust target perception. While synthetic aperture radar (SAR) image-based ship recognition methods have been extensively studied [5–7], they often require large-scale annotated datasets and may overlook underlying electromagnetic scattering mechanisms. Radar-based approaches using high-resolution range profile (HRRP) and Radar Cross Section (RCS) sequences operate on lower-dimensional, physically meaningful data representations that can be effectively learned from relatively smaller datasets. Although some research has been conducted on HRRP-based recognition [8,9], studies based on RCS are still relatively scarce [10–15]. With the development of AI technologies, scattering-data-driven methods can adaptively learn deep discriminative features from raw radar data, overcoming the limitations of traditional handcrafted features [16–19].

In HRRP recognition, Ding and Chen (2019) proposed a CNN-based method [20], while Zhang et al. (2019) combined CNN with extreme learning machine [21]. Subsequent research explored more expressive temporal architectures [22,23], such as the recurrent gamma belief network by Guo et al. (2020) [24] and the attention-enhanced convolutional GRU network by Chu and Guo (2021) [25]. Recently, Yang et al. (2024) proposed a ConvLSTM-based multi-input GRU fusion method, achieving 93.32% accuracy in six-class ship target recognition (full polarization, 20 dB SNR) [26]. However, this method employs fixed parameter combinations for feature fusion and lacks adaptive weight adjustment, limiting its generalization capability.

For RCS recognition, Zhu et al. (2024) proposed a flight target recognition method based on RCS sequences [27], while Wang et al. (2023) utilized bidirectional LSTM to model RCS statistical features, improving low-SNR performance [28]. F. Xueman et al. (2016) explored statistical description and classifier fusion based on RCS frequency characteristics [29]. Notably, some methods convert RCS data into images before recognition [30,31], which increases computational complexity and may lose original electromagnetic information.

Although the aforementioned methods have made progress, most still focus on single modalities or rely on image conversion, lacking in-depth mining of multi-modal correlations and adaptive fusion mechanisms. Particularly, designing effective fusion mechanisms that explicitly model deep semantic relationships between modalities and adaptively measure each modality's contribution remains key to enhancing recognition performance in complex environments.

To address these challenges, this paper proposes a three-channel network architecture based on heterogeneous multi-modal feature fusion, constructing a complete “mechanism-data-physics” trinity recognition framework. This framework utilizes raw HRRP sequences to provide the data foundation of target structural details, combines RCS sequences that reflect electromagnetic scattering physical mechanisms, and introduces statistically signifi-

cant features extracted from HRRP as physical prior knowledge. The core innovations of our method are twofold:

- Fine-grained interactive enhancement at the feature level: A cross-attention module explicitly models deep semantic relationships between heterogeneous modalities by dynamically establishing bidirectional association mappings, treating features from different modalities as Query, Key, and Value, thereby effectively capturing complementary information among HRRP structural details, RCS scattering characteristics, and statistical prior knowledge.
- Adaptive weighting of modality contributions at the decision level: An adaptive weight fusion layer is incorporated to dynamically learn the contribution of each modality in a data-driven manner.

The two innovations enable full utilization of multi-modal information in complex scenarios. Furthermore, the proposed framework demonstrates strong generalization capability even under small-sample conditions. The performance of the method is systematically evaluated under various SNR and polarization conditions, and the effectiveness of each module is validated through ablation studies and visual analysis.

The paper is structured as follows: Section 2 presents the three-channel feature mechanism analysis, detailing the physical meanings and complementarity of HRRP, RCS, and statistical features. Section 3 introduces the dual-path collaborative fusion network with cross-attention and adaptive weight fusion. Section 4 provides experimental results and analysis, including ablation studies, performance comparison, and visualizations. Section 6 concludes the work and outlines future directions.

2. Data Selection and Three-Channel Feature Mechanism Analysis

RCS compresses the target's scattering characteristics into a frequency-dependent scalar, macroscopically characterizing the overall electromagnetic scattering mechanism [32]. HRRP, obtained via IFFT-based wideband synthesis, provides microscopic structural details as the vector sum of scattering center projections [33]. Statistical features extracted from HRRP serve as physical prior knowledge, offering noise-robust representations through higher-order moments and entropy. Based on the complementary characteristics of these three sources, a “mechanism-data-physics” information system is constructed. As shown in Figure 1, the entire network architecture revolves around this core concept. Their systematic integration constitutes a complete multi-modal information system for robust target recognition.

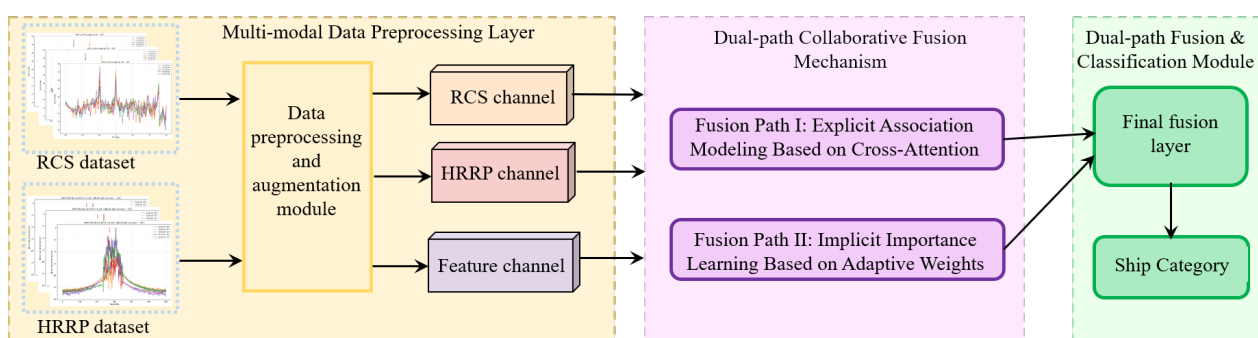


Figure 1. Heterogeneous three-channel parallel encoding and dual-path collaborative fusion network.

2.1. Unified Framework by Mechanism-Data-Physics

Existing radar target recognition methods mostly focus on data-driven approaches, typically relying on a single source of scattering data for feature extraction and classification. However, different types of scattering data physically characterize different dimensions of

target attributes. As shown in Table 1, RCS describes the target's macroscopic scattering capability, HRRP captures the fine-grained structural distribution of scattering centers, and statistical features characterize the distributional regularities of the scattering mechanisms. These three correspond to coarse, fine, and medium granularity information dimensions, respectively, possessing natural complementarity.

Table 1. Comparison of physical characteristics of three types of scattering data sources.

Comparison Dimension	RCS	HRRP	Statistical Features
Core target attributes reflected	Overall target size, shape contour, and material reflectivity	Radial distribution of scattering centers, geometric structure layout, and fine-scale dimensions of the target	Scattering distribution shape (skewness/kurtosis), energy concentration (entropy), structural complexity, and spectral characteristics
Information dimension and representation granularity	Zero-dimensional scalar (forms a 1D sequence with angle/frequency); coarse granularity—global description	One-dimensional vector (range bin amplitude sequence); fine granularity—local structure	Multi-dimensional scalar set (16 handcrafted statistical features); medium granularity—statistically robust

Based on the above analysis, multi-source scattering data fusion serves as the fundamental starting point of this paper. To quantitatively characterize the trust superposition of complementary information from different data sources, a mathematical description of the fusion framework is established from both Bayesian decision theory and deep learning perspectives.

From the viewpoint of Bayesian decision theory, the core problem of radar target recognition is to find the target category y that maximizes the posterior probability given the multi-source observation data $\mathcal{X}_{\text{RCS}}$, $\mathcal{X}_{\text{HRRP}}$, $\mathcal{X}_{\text{STAT}}$:

$$\hat{y} = \arg \max_{y \in \mathcal{Y}} P(y \mid \mathcal{X}_{\text{RCS}}, \mathcal{X}_{\text{HRRP}}, \mathcal{X}_{\text{STAT}}) \quad (1)$$

where $\mathcal{X}_{\text{RCS}}$ denotes the RCS sequence, $\mathcal{X}_{\text{HRRP}}$ denotes the high-resolution range profile, and $\mathcal{X}_{\text{STAT}}$ denotes the statistical feature vector.

Given the target category, the three types of observation data are determined by different physical mechanisms and approximately satisfy conditional independence. Then the above expression can be factorized as

$$P(y \mid \mathcal{X}_{\text{RCS}}, \mathcal{X}_{\text{HRRP}}, \mathcal{X}_{\text{STAT}}) \propto P(\mathcal{X}_{\text{RCS}} \mid y) \cdot P(\mathcal{X}_{\text{HRRP}} \mid y) \cdot P(\mathcal{X}_{\text{STAT}} \mid y) \cdot P(y) \quad (2)$$

Equation (2) constitutes the unified theoretical framework of this paper: the trust superposition of multi-source information is achieved by multiplying the conditional probabilities of different data sources. When the information quality of one data source degrades under specific conditions, the conditional likelihoods of the other data sources can still provide effective discriminative information, thereby reducing the overall decision uncertainty.

Under the deep learning framework, the neural network directly fits the posterior probability distribution described above. The recognition process is decomposed into three cascaded operators: feature encoding E , feature fusion $\mathcal{F}$, and classification decision $\mathcal{C}$:

$$\hat{y} = \mathcal{C}(\mathcal{F}(E_{\text{RCS}}(\mathcal{X}_{\text{RCS}}), E_{\text{HRRP}}(\mathcal{X}_{\text{HRRP}}), E_{\text{STAT}}(\mathcal{X}_{\text{STAT}}))) \quad (3)$$

The encoding operator E maps heterogeneous data into a unified semantic space, eliminating dimensional differences. The fusion operator $\mathcal{F}$ adopts a dual-path cooperative strategy: the cross-attention path realizes cross-source semantic correspondence at the feature level, while the adaptive weighting path learns the optimal contribution ratio of

each data source at the decision level. The classification operator $\mathcal{C}$ maps the fused features to class probabilities. Equations (1)–(3) together form the unified theoretical framework for collaborative fusion of multi-source scattering data in this paper. Based on this framework, the following subsections detail the three encoding pathways for RCS, HRRP, and statistical features, respectively. The specific network implementation is detailed in Section 3.

2.2. Encoding Pathway by RCS Scattering Mechanism

In the RCS channel, since RCS data exhibit both local patterns and global dependencies, a hybrid feature extraction branch composed of a one-dimensional convolutional layer (Conv1D) and a gated recurrent unit (GRU) is employed [34–36]. This architecture efficiently extracts local scattering features in the frequency domain through the Conv1D layer, and then models the long-range temporal dependencies between these features using the GRU layer, thereby generating a high-level representation of an RCS sequence. The specific network framework is shown in Figure 2.

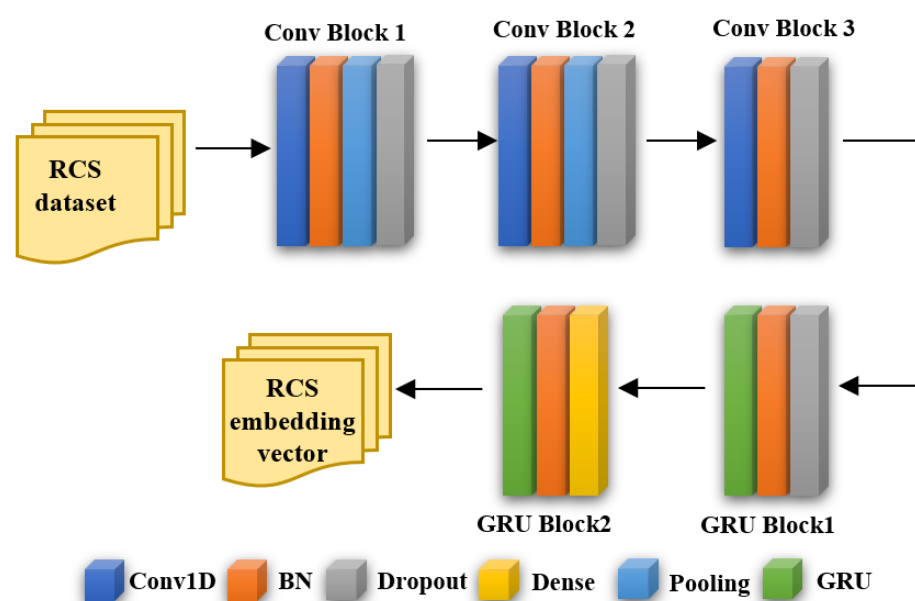


Figure 2. RCS channel architecture from input to embedding vector.

The RCS processing channel implements a hierarchical feature extraction pipeline:

- Shallow Layers: 64 filters with a kernel size of 3 capture fine-grained local fluctuations.
- Middle Layers: 128 filters with a kernel size of 5 capture broader context.
- Deep Layers: 256 filters with a kernel size of 3 produce high-level representations.

The RCS data processing channel constructs a hierarchical feature extraction architecture. Its core consists of a one-dimensional convolutional layer (Conv1D) in tandem with a gated recurrent unit (GRU). Through multi-layer stacking of convolutional and pooling operations, the Conv1D achieves hierarchical abstraction and dimensionality reduction of the local scattering patterns in the original signals, generating a high-level feature sequence. The GRU module models the temporal dynamics of the feature sequence, captures the long-range dependencies through its gating mechanism, and ultimately encodes the variable-length sequence into a context-aware vector with fixed dimensions. This “local-aware-global-encoding” collaborative architecture effectively solves the problem of co-modeling local patterns and global context in the RCS spectrum and generates highly discriminative feature embeddings for subsequent fusion and classification tasks.

2.3. Encoding Pathway by HRRP Data

Convolutional LSTM (ConvLSTM) is an extension of the LSTM architecture, with the core idea of replacing fully connected operations with convolutions, thereby enabling effective modeling of high-dimensional spatiotemporal sequence data [37–39]. The model maintains the input, output and internal states as 3D tensor structures to preserve the spatial topological properties; its gating mechanisms (including input, forget and output gates) are all implemented by convolution, which combines both spatial feature extraction and temporal dynamics modeling, and allows structured spatio-temporal modulation of hidden and cellular states simultaneously. The key equations are:

$$\begin{aligned} i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\ C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\ o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \\ H_t &= o_t \circ \tanh(C_t) \end{aligned} \quad (4)$$

where $*$ denotes convolution, and $\circ$ denotes the Hadamard product.

The HRRP data processing channel constructs a deep spatio-temporal feature extraction architecture, the core of which consists of a Convolutional Long Short-Term Memory (ConvLSTM) network, the specific network framework of which is shown in Figure 3. The ConvLSTM module synergistically captures spatial structural properties and underlying temporal evolution patterns in the one-dimensional range profile through its unique convolutional gating mechanism. Its convolution operation explicitly extracts the local spatial correlations between adjacent scattering units in the distance dimension, while the internal state transfer implicitly models the long-range dependencies between multiple scattering centers and their dynamic trends. The architecture ultimately encodes the input high-dimensional range profile sequence into a highly condensed fixed-dimensional feature vector. This integrated design of “spatial perception-temporal context encoding” effectively unifies the spatial and temporal representations of HRRP data, providing robust and discriminative feature embeddings for downstream cross-modal fusion and high-precision classification.

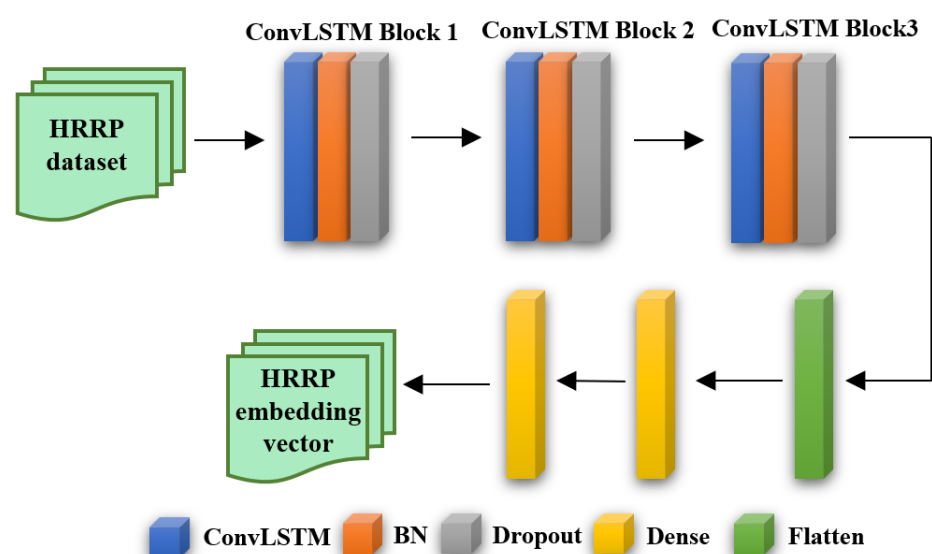


Figure 3. HRRP channel architecture from input to embedding vector.

2.4. Encoding Pathway by Statistical Physical Priors

Statistical feature channels combine prior knowledge from signal processing with the representation power of deep learning. The operational process is as follows: firstly, four categories of 16 manual statistical features are extracted from the original HRRP sequences, including: (1) Basic Amplitude Statistics—mean (μ), maximum ($x_{\max}$), minimum ($x_{\min}$), range (R), standard deviation (σ), variance (σ^2), and root mean square (x_{rms}); (2) Shape Distribution—skewness (γ_1) and kurtosis (γ_2); (3) Energy and Information—energy (E), average energy (E_{avg}), and entropy (H); (4) Peak and Spectral—number of peaks (N_p), average peak height (h_{avg}), spectral centroid (C_s), and spectral bandwidth (B_s). These features are then encoded at a higher order by a deep GRU network, which finally generates 128-dimensional feature embedding vectors, and the specific network framework is shown in Figure 4.

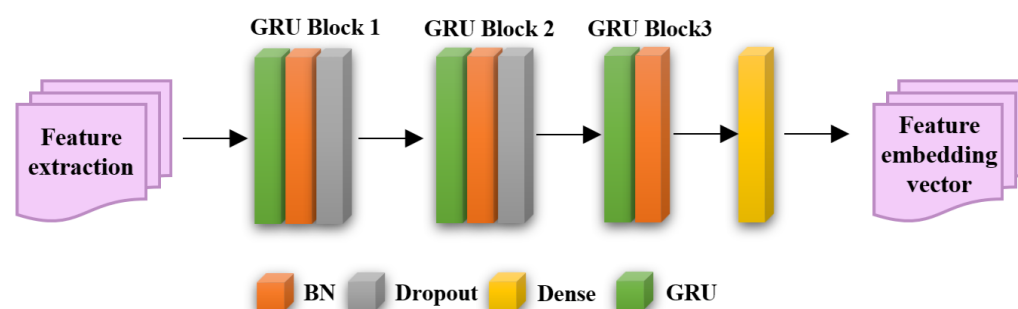


Figure 4. Statistical feature channel architecture from input to embedding vector.

3. The Proposed Dual-Path Cooperative Fusion Network

The core innovation of this network architecture is the dual-path collaborative fusion mechanism, which addresses two key challenges: modeling deep semantic correlations between modalities and adaptively weighting each modality's contribution in the final decision. It first preprocesses raw data through three heterogeneous encoding channels to generate high-order features, then focuses entirely on two parallel and complementary pathways: cross-attention fusion and adaptive weight fusion, ultimately completing decision making through collaborative integration.

3.1. Fusion Path I: Explicit Association Modeling Based on Cross-Attention

By introducing a cross-attention mechanism, this pathway actively establishes fine-grained semantic associations between modalities in the feature space, achieving information interaction and enhancement. The mechanism facilitates cross-modal information interaction and feature enhancement by using features from one modality as Query and features from another modality as Key and Value [40–42]. This mechanism can dynamically establish association mappings between modalities in the feature space, highlighting key information and suppressing redundant features, thereby enhancing the discriminative power and interpretability of fused representations [43]. This paper constructs a ternary parallel cross-attention architecture to ensure that any two modalities can interact deeply with each other.

3.1.1. Theoretical Framework of Cross-Attention Block

A complete Cross-Attention Block consists of Layer Normalization, Multi-Head Attention, Residual Connections, and a Feed-Forward Network (FFN). Its computational process is described below.

(1) Input Definition: Assume there are two modality features requiring interaction.

Query modality feature tensor: $\mathbf{X}_Q \in \mathbb{R}^{L_Q \times D}$;

Key-Value modality feature tensor: $\mathbf{X}_{KV} \in \mathbb{R}^{L_{KV} \times D}$;

Where L_Q and L_{KV} are the sequence lengths, and D is the feature dimension.

(2) Layer Normalization: First, normalize the input features to stabilize the training process.

$$\mathbf{X}'_Q = \text{LayerNorm}(\mathbf{X}_Q) \quad (5)$$

$$\mathbf{X}'_{KV} = \text{LayerNorm}(\mathbf{X}_{KV}) \quad (6)$$

(3) Core Computation of Multi-Head Attention (MHA): The core of the attention mechanism is to compute the similarity between Query and Key, and use this as weights for a weighted sum of the Value.

Scaled Dot-Product Attention:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (7)$$

where d_k is the dimension of the Key vectors.

Multi-Head Attention: Linearly project $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ and split them into h heads, compute attention for each head separately, then concatenate the results.

$$\text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)\mathbf{W}^O \quad (8)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (9)$$

where the projection matrices $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V, \mathbf{W}^O$ are learnable parameters.

In cross-attention, the Query comes from one modality, while the Key and Value come from the other modality:

$$\mathbf{H}_{\text{attn}} = \text{MHA}(\mathbf{X}'_Q, \mathbf{X}'_{KV}, \mathbf{X}'_{KV}) \quad (10)$$

(4) Residual Connection and Layer Normalization: Add the output of the attention layer to the original Query (residual connection) and apply normalization again to mitigate gradient vanishing in deep networks.

$$\mathbf{H}_{\text{res1}} = \text{LayerNorm}(\mathbf{X}_Q + \mathbf{H}_{\text{attn}}) \quad (11)$$

(5) Feed-Forward Network (FFN) and Second Residual Connection: The FFN introduces non-linear transformation and spatial transformation to further enhance the model's representational capacity. It typically consists of two linear layers and an activation function.

$$\mathbf{H}_{\text{ffn}} = \text{FFN}(\mathbf{H}_{\text{res1}}) = \text{ReLU}(\mathbf{H}_{\text{res1}}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \quad (12)$$

$$\mathbf{Y} = \text{LayerNorm}(\mathbf{H}_{\text{res1}} + \mathbf{H}_{\text{ffn}}) \quad (13)$$

The final output $\mathbf{Y}$ is the feature representation enhanced through cross-modal interaction.

3.1.2. Implementation of Ternary Parallel Interaction

In the three-channel fusion model, the cross-attention mechanism is strategically applied three times, constructing a hierarchical, bidirectional cross-modal information interaction network. This design aims to fully exploit the deep correlations and complementarity between HRRP, RCS, and statistical features, going beyond simple feature concatenation or weighted averaging.

- $\mathbf{H}_{\text{hrrp}}$: High-dimensional feature vector output by the HRRP branch;
- $\mathbf{H}_{\text{rsc}}$: High-dimensional feature vector output by the RCS branch;
- $\mathbf{H}_{\text{stat}}$: High-dimensional feature vector output by the statistical feature branch.

HRRP-RCS Modality Interaction Layer: The HRRP sequence reflects the distribution of scattering points of the target in the radial direction, while the RCS sequence reflects the scattering intensity of the target at different azimuth angles. This attention mechanism allows HRRP features to actively query RCS features, aiming to establish a fine-grained mapping between the radial distribution and azimuthal scattering intensity, thereby enhancing the model's unified representation of the target's electromagnetic characteristics. Its computation can be expressed as

$$\mathbf{H}_{\text{HRRP-RCS}} = \text{CrossAttention}(\mathbf{H}_{\text{hrrp}}, \mathbf{H}_{\text{rcs}}) \quad (14)$$

HRRP-Statistical Features Interaction Layer: Handcrafted statistical features (such as mean, variance, entropy, etc.) are macroscopic, low-dimensional descriptors with physical meaning extracted from HRRP sequences. This attention mechanism allows the original HRRP detailed features to query their corresponding macroscopic statistical features, aiming to supplement contextual semantic information for the detailed features, mitigate noise interference in the original sequence, and enhance feature robustness:

$$\mathbf{H}_{\text{HRRP-Stat}} = \text{CrossAttention}(\mathbf{H}_{\text{hrrp}}, \mathbf{H}_{\text{stat}}) \quad (15)$$

RCS-Statistical Features Interaction Layer: Similarly, RCS sequences can also benefit from statistical features. This attention mechanism ensures the symmetry and completeness of information interaction, preventing modeling bias. By querying statistical features, RCS features can enhance their perception of macroscopic characteristics such as target fluctuation and stability:

$$\mathbf{H}_{\text{RCS-Stat}} = \text{CrossAttention}(\mathbf{H}_{\text{rcs}}, \mathbf{H}_{\text{stat}}) \quad (16)$$

Finally, the outputs of the three interaction layers are concatenated and fused, forming a joint representation rich in multi-level cross-modal information:

$$\mathbf{H}_{\text{cross}} = \text{Concat}(\mathbf{H}_{\text{HRRP-RCS}}, \mathbf{H}_{\text{HRRP-Stat}}, \mathbf{H}_{\text{RCS-Stat}}) \quad (17)$$

This representation, together with the output from the adaptive weight fusion layer, is fed into the subsequent classifier for final decision-making. Figure 5 details the overall architecture and information flow of this mechanism.

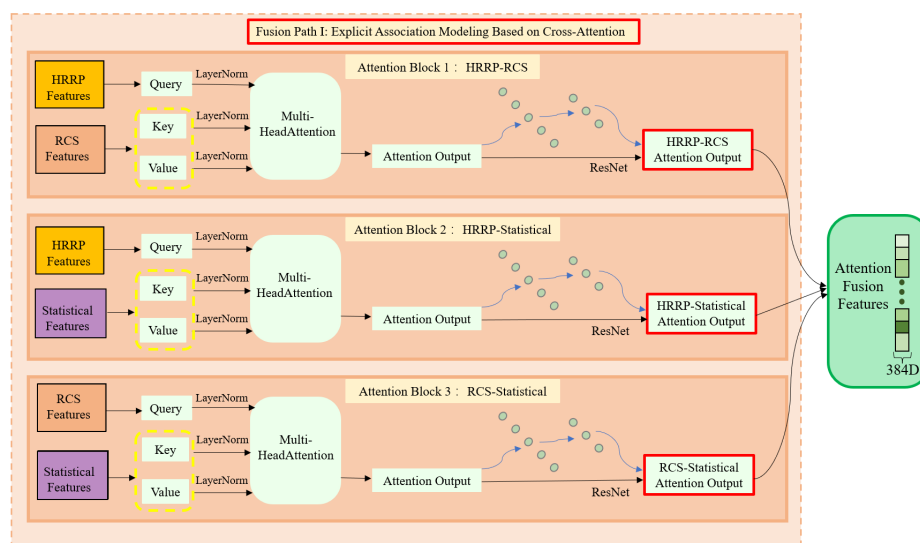


Figure 5. Overall architecture and information flow of the cross-attention mechanism. The yellow, orange, and purple blocks represent the input features from the HRRP, RCS, and Statistical branches,

respectively. Black arrows denote forward propagation paths. The yellow dashed boxes distinguish the roles of different input sources within the cross-attention mechanism: features from one branch serve as Query, while features from another branch serve as Key and Value. Three parallel attention blocks (Block 1–3) extract features from HRRP-RCS, HRRP-Statistical, and RCS-Statistical pairs, respectively, followed by residual connections. The outputs are concatenated into a 384-dimensional attention fusion feature vector.

3.2. Fusion Path II: Implicit Importance Learning Based on Adaptive Weights

This pathway dynamically optimizes each modality's contribution, replacing traditional fixed weights or simple concatenation. By designing a learnable adaptive weight fusion layer, it implicitly learns the global importance of each modality in a data-driven manner. The overall architecture is shown in Figure 6.

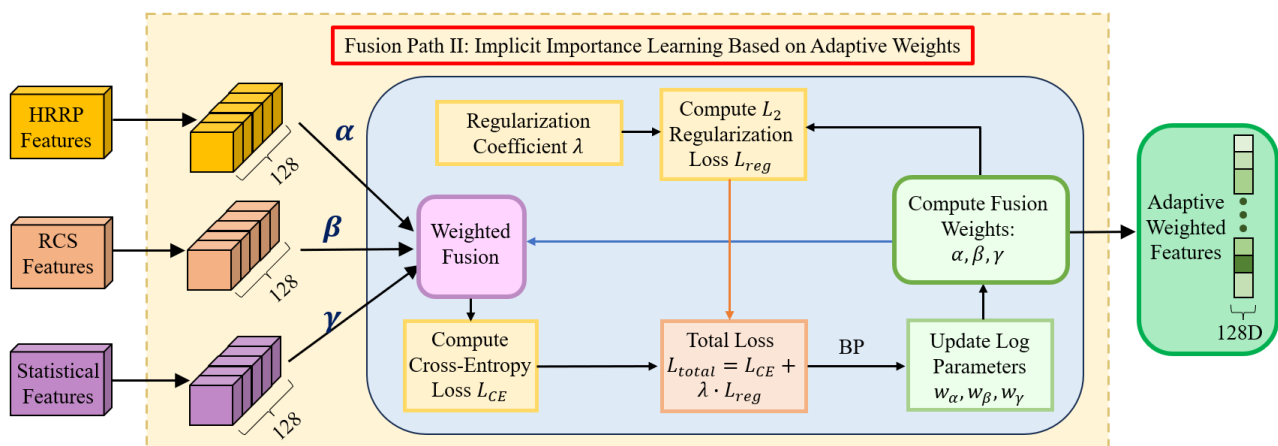


Figure 6. Adaptive weight channel architecture. The yellow, orange, and purple blocks represent the 128-dimensional feature vectors from the HRRP, RCS, and Statistical branches, respectively. Unlabeled black arrows indicate forward propagation paths; black arrows labeled “BP” indicate backpropagation paths. The three 128-dimensional feature vectors are fused through a learnable weighting layer to produce a 128-dimensional adaptively weighted feature vector. The weighting layer contains trainable log parameters $w_\alpha, w_\beta, w_\gamma$, normalized via Softmax to fusion weights α, β, γ . The learning is driven by the total loss $\mathcal{L}_{total} = \mathcal{L}_{CE} + L_{reg}$, where $L_{reg} = \lambda \cdot (\alpha^2 + \beta^2 + \gamma^2)$. Backpropagation updates the log parameters to iteratively compute the fusion weights.

The traditional fixed-weight fusion approach can be expressed as

$$\mathbf{h}_{\text{fusion}} = \alpha \cdot \mathbf{h}_{\text{hrrp}} + \beta \cdot \mathbf{h}_{\text{rsc}} + \gamma \cdot \mathbf{h}_{\text{stat}} \quad (18)$$

where α, β, γ are pre-defined scalar values, and $\alpha + \beta + \gamma = 1$.

The proposed adaptive fusion mechanism generalizes this to optimizable weight parameters. To achieve this, we introduce three trainable log-scale initial parameters (logits) $w_\alpha, w_\beta, w_\gamma$, and map them to normalized fusion weights via the Softmax function:

$$\alpha = \frac{\exp(w_\alpha)}{\exp(w_\alpha) + \exp(w_\beta) + \exp(w_\gamma)} \quad (19)$$

$$\beta = \frac{\exp(w_\beta)}{\exp(w_\alpha) + \exp(w_\beta) + \exp(w_\gamma)} \quad (20)$$

$$\gamma = \frac{\exp(w_\gamma)}{\exp(w_\alpha) + \exp(w_\beta) + \exp(w_\gamma)} \quad (21)$$

The final multimodal fusion feature $\mathbf{H}_{\text{fusion}}$ is computed as follows:

$$\mathbf{H}_{\text{fusion}} = \alpha \cdot \mathbf{H}_{\text{hrrp}} + \beta \cdot \mathbf{H}_{\text{rsc}} + \gamma \cdot \mathbf{H}_{\text{stat}} \quad (22)$$

The initial log parameters $w_\alpha, w_\beta, w_\gamma$ are set as trainable variables of the model. During the initialization phase, these parameters are configured based on prior knowledge of modality importance:

$$w_\alpha = \log(\alpha_{\text{init}} + \varepsilon) \quad (23)$$

$$w_\beta = \log(\beta_{\text{init}} + \varepsilon) \quad (24)$$

$$w_\gamma = \log(\gamma_{\text{init}} + \varepsilon) \quad (25)$$

where $\alpha_{\text{init}}, \beta_{\text{init}}, \gamma_{\text{init}}$ represent the initial importance weights for HRRP, RCS, and statistical features respectively, and ε is a small positive constant (10^{-8}) included for numerical stability.

During training, $w_\alpha, w_\beta, w_\gamma$ are optimized together with other network parameters through backpropagation and gradient descent. The gradient of the loss function with respect to $w_\alpha, w_\beta, w_\gamma$ clearly reflects the contribution of each modal feature to the final classification task, thereby driving the model to assign higher weights to more important modalities.

To prevent any single modality from dominating the fusion output and to encourage balanced utilization of the three complementary sources, an L2 regularization term is introduced on the normalized fusion weights:

$$\mathcal{L}_{\text{reg}} = \lambda \cdot (\alpha^2 + \beta^2 + \gamma^2) \quad (26)$$

where α, β , and γ are the fusion weights after softmax normalization, and λ is the regularization coefficient. This penalty term discourages extreme weight distributions, thereby promoting a more uniform and robust combination of the HRRP, RCS, and statistical feature modalities. It complements the adaptive weight learning mechanism by preventing over-reliance on any single data source, enhancing the model's generalization capability across varying SNR and polarization conditions.

3.3. Dual-Path Cooperation and Classification Decision

The cross-attention pathway and adaptive weight pathway form a hierarchical complementarity in function, collectively constructing the cooperative fusion architecture of this method and achieving a complete chain from feature interaction to decision optimization. The specific results are shown in Figure 7.

The cross-attention pathway achieves interactive enhancement at the feature level. By establishing cross-modal attention mappings, this pathway performs semantic alignment and information filtering in the feature space, generating enhanced feature representations F_{cross} rich in inter-modal correlations. This process can be regarded as a content-based dynamic feature refinement, effectively improving feature discriminability.

The adaptive weight pathway achieves feature selection at the decision level. Through learnable weight parameters, this pathway evaluates the contribution of each modality's features to the global classification decision in a data-driven manner, performing implicit feature importance weighting to form optimized weighted feature representations F_{weight} . This mechanism enables the model to adapt to different polarization conditions, enhancing the system's scenario robustness.

Finally, the output features of the two pathways are integrated through concatenation, forming a comprehensive joint feature representation $H_{\text{final}} = [H_{\text{cross}}; H_{\text{weight}}]$. This

feature preserves both the deep cross-modal correlation structure and reflects the global importance weights of each modality, ultimately being fed into the classifier for end-to-end target recognition.

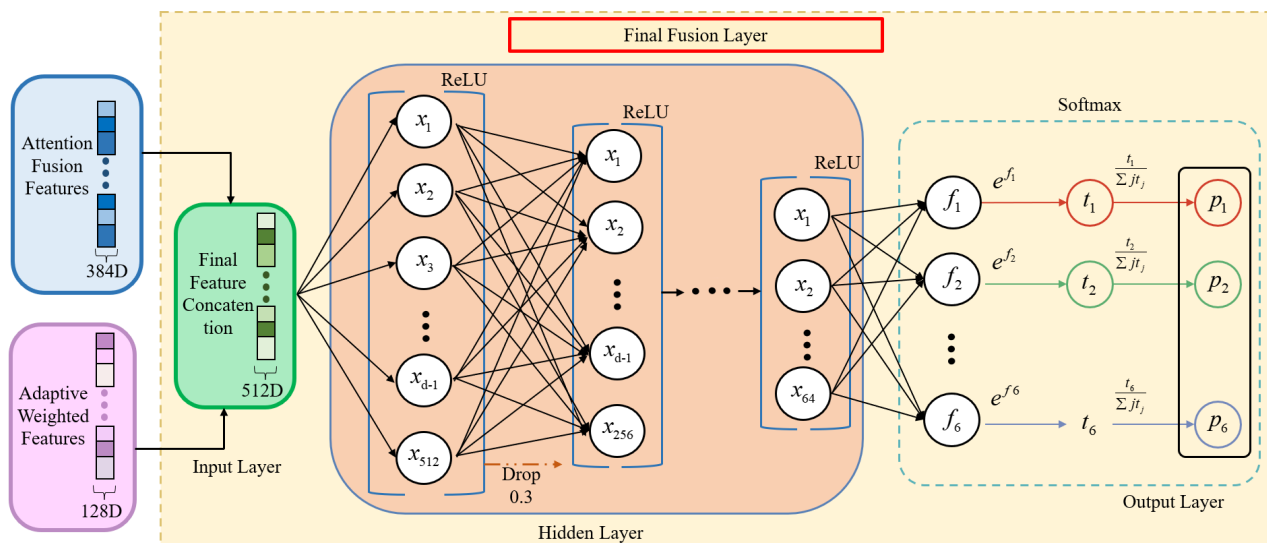


Figure 7. Final fusion structure. The blue and purple modules represent the 384-dimensional attention-fused features from Path I and the 128-dimensional adaptively weighted features from Path II, respectively. The green module denotes the 512-dimensional concatenated feature vector. The orange and cyan dashed regions indicate the hidden layers and the output layer (with Softmax), respectively. Black solid arrows denote feature forward propagation and fully connected operations. The ellipses inside feature vectors and neurons represent omitted intermediate elements, and ellipses between hidden layers indicate omitted transitional layers.

This cooperative mechanism of “explicit association at the feature level” and “implicit weighting at the decision level” jointly optimizes the model’s multi-modal information fusion capability from both representation learning and feature selection dimensions, providing a reliable technical approach for high-precision target recognition in complex electromagnetic environments.

4. Experiments and Analysis of Results

4.1. Experimental Configuration and Dataset Description

The radar system adopts a stepped-frequency signal scheme with the following core parameters: center frequency (f_c) of 18 GHz, bandwidth (B) of 300 MHz, and 601 frequency points. To simulate the terminal attack process of an anti-ship missile, the missile trajectory is discretized into a series of observation positions corresponding to different radar viewing angles. The elevation angle θ covers the range from 30° to 90° in steps of 5° , while the azimuth angle ϕ covers the range from -60° to 60° in steps of 1° . Based on these settings, a total of 1573 observation position samples are generated, forming the data foundation for subsequent research. The specific ship model is shown in Figure 8.

The dataset contains high-resolution range profile (HRRP) and Radar Cross Section (RCS) data for six different types of ship targets. The electromagnetic scattering calculations for the targets were simulated using the Shooting and Bouncing Rays-Physical Optics (SBR-PO) method to accurately obtain their echo information under various polarization combinations [44,45]. The dataset comprises two co-polarization modes (HH and VV) and two cross-polarization modes (HV and VH), enabling comprehensive exploration of target scattering characteristics in the polarization dimension [46]. The dataset was randomly di-

vided into training and testing sets in an 8:2 ratio, ensuring a consistent sample distribution across all categories to guarantee balanced training and reliable evaluation results.

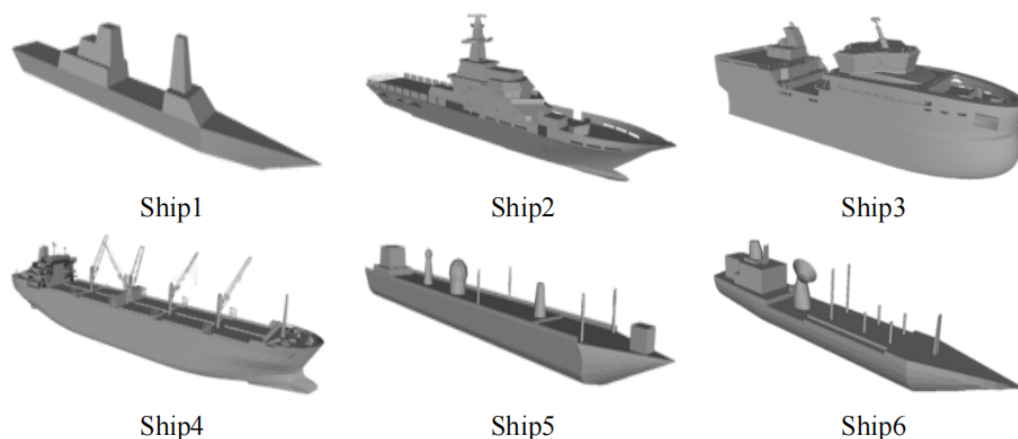


Figure 8. Ship model diagram.

4.2. Performance Evaluation Methods

4.2.1. Loss Function

The loss function is used to measure the degree of discrepancy between the model's predicted values and the true values. It is a non-negative real-valued function, typically denoted by L , where a smaller L value indicates better model robustness.

The loss function operates during the model training phase. After each batch of samples undergoes forward propagation to obtain predicted outputs, the loss function computes the discrepancy between the predicted values and the true labels, yielding the loss value. Subsequently, the model utilizes the backpropagation algorithm to update the trainable parameters based on this loss value, aiming to minimize the loss and drive the predicted distribution closer to the true distribution, thereby enabling model learning.

The training process of the model aims to minimize the difference between the predicted probability distribution and the true label distribution. Given that this task is a multi-class classification problem, we employ the Categorical Cross-Entropy Loss as the optimization objective function for the model. Consider a classification task with C classes. For a given sample, its true label can be represented as a one-hot encoded vector $\mathbf{y}^T = [y_1, y_2, \dots, y_C]^T$, where only the element corresponding to the true class is 1, and the others are 0. Simultaneously, the model's output layer generates a probability distribution vector $\hat{\mathbf{y}}^T = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_C]^T$ through the softmax function, where $\hat{y}_i$ represents the model's predicted probability that the sample belongs to the i -th class.

The cross-entropy loss $\mathcal{L}_{CE}$ for this sample is defined as

$$\mathcal{L}_{CE}(\mathbf{y}, \hat{\mathbf{y}}) = - \sum_{i=1}^C y_i \cdot \log(\hat{y}_i) \quad (27)$$

Since $\mathbf{y}$ is a one-hot vector, the above equation can be simplified to

$$\mathcal{L}_{CE} = -\log(\hat{y}_k) \quad (28)$$

where k is the true class index of the sample.

For a training mini-batch containing N samples, the total loss is the average of all sample losses:

$$\mathcal{L}_{\text{Batch}} = -\frac{1}{N} \sum_{j=1}^N \log(\hat{y}_k^{(j)}) \quad (29)$$

where $\hat{y}_k^{(j)}$ is the predicted probability of the j -th sample for its true class.

Cross-entropy loss measures the “distance” between two probability distributions. When the predicted probability $\hat{\mathbf{y}}$ exactly matches the true distribution $\mathbf{y}$, the loss is 0. Lower prediction confidence (smaller $\hat{y}_k$) results in a larger loss value, thereby imposing a stronger penalty on the model’s incorrect predictions.

4.2.2. Evaluation Metrics

To comprehensively and fairly evaluate the model’s performance, a series of evaluation metrics widely used in classification tasks are employed for analysis.

(1) Accuracy

Accuracy is a fundamental metric that measures the correctness of model predictions, representing the proportion of correctly predicted samples to the total number of samples. It intuitively reflects the model’s prediction accuracy on the overall dataset.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (30)$$

where TP represents True Positive (number of samples correctly predicted as positive), TN represents True Negative (number of samples correctly predicted as negative), FP represents False Positive (number of samples incorrectly predicted as positive), and FN represents False Negative (number of samples incorrectly predicted as negative).

In multi-class classification problems, the concepts of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are extended to each class. Accuracy can also be understood as the sum of correctly predicted samples across all classes divided by the total number of samples.

(2) Precision, Recall and F1-Score

Accuracy can be misleading for imbalanced datasets. Therefore, metrics based on the Confusion Matrix are introduced to evaluate performance at a finer granularity.

Precision measures the accuracy of positive predictions, defined as the proportion of true positive samples among all samples predicted as positive. It measures prediction accuracy; higher Precision indicates better detection accuracy and a lower false positive rate.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (31)$$

Recall is a key metric that measures the model’s ability to identify positive samples, defined as the proportion of correctly detected positive samples to all actual positive samples. Higher Recall indicates fewer missed detections.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (32)$$

F1-Score achieves a balance between Precision and Recall by calculating their harmonic mean. It integrates the performance of both and is a more balanced metric, particularly suitable for scenarios with imbalanced class distributions.

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (33)$$

For multi-class problems, Precision, Recall, and F1-Score are typically computed for each class, and then the overall performance is reported through Macro-Average. Macro-Average simply calculates the metrics for each class and then takes the arithmetic mean, giving equal weight to each class, which better reflects the model’s performance on minority classes.

$$\text{Precision}_{\text{macro}} = \frac{1}{C} \sum_{i=1}^C \text{Precision}_i \quad (34)$$

$$\text{Recall}_{\text{macro}} = \frac{1}{C} \sum_{i=1}^C \text{Recall}_i \quad (35)$$

$$F_{1\text{macro}} = \frac{1}{C} \sum_{i=1}^C F_{1i} \quad (36)$$

4.3. Ablation Studies: Deconstructing the Dual-Path Fusion Mechanism

To validate the effectiveness of each key module, seven ablation variants are designed: three single-source baselines (Only-HRRP, Only-RCS, Only-Statistic), a fixed equal-weight fusion of three sources (Fixed Weight), an adaptive weight fusion channel without cross-attention (Weight Fusion Only), a cross-attention interaction channel without adaptive fusion (Cross-Attention Only), and the full dual-path cooperative fusion model (Proposed Method). All experiments were conducted under identical dataset and training conditions to ensure comparability of results.

4.3.1. Performance Comparison Under Different Polarization Modes

To evaluate the impact of polarization modes on model performance, recognition accuracy of seven networks under five polarization conditions is compared under a high signal-to-noise ratio condition (30 dB). The results are shown in Table 2. Single-data-source variants exhibit polarization-dependent performance: co-polarization (HH, VV) yields higher accuracy than cross-polarization (HV, VH). For example, Only-HRRP achieves 0.9004 and 0.9078 under HH and VV, respectively, but drops to 0.8215 and 0.7950 under HV and VH. This is because co-polarized echoes are dominated by direct backscattering with stronger energy, while cross-polarized echoes mainly originate from multiple scattering and depolarization effects, leading to lower signal quality. Under full polarization, all single-source variants degrade significantly, as full-polarization data contain multiple polarization modes, increasing internal diversity beyond a single modality's capacity. Fixed-weight fusion (1/3) achieves only 0.8049, showing no substantial improvement over the best single source, indicating that simply stacking multi-source data without effective fusion yields limited gains.

Table 2. Recognition accuracy comparison of seven networks under different polarization modes.

Polarization Mode	Only HRRP	Only RCS	Only Statistic	Fixed Weight (1/3)	Weight Fusion Only	Cross-Attention Only	Proposed Method
HH	0.9004	0.9248	0.8400	0.9001	0.9433	0.9677	0.9942
HV	0.8215	0.8013	0.7457	0.8114	0.8416	0.9105	0.9815
VH	0.7950	0.7918	0.7929	0.8366	0.8490	0.9026	0.9841
VV	0.9078	0.9057	0.8400	0.9208	0.9444	0.9825	0.9952
Full Polarization	0.7586	0.8212	0.7640	0.8049	0.8845	0.9556	0.9939

Among the three fusion variants, Weight Fusion Only outperforms Fixed Weight across all polarizations, confirming that adaptive weight assignment is superior to equal fusion—the contributions of different data sources are not balanced and correlate with polarization mode. Cross-Attention Only shows a more significant advantage, reaching 0.9556 under full polarization, about 7 percentage points higher than Weight Fusion Only, indicating that feature-level cross-source interaction utilizes multi-source information more effectively than decision-level weighting alone. When both mechanisms work together,

the full model reaches 0.9939, suggesting a synergistic effect: adaptive fusion balances contributions at a global level, while cross-attention establishes semantic correspondences at the feature level. The full model's performance under full polarization is on par with the average of single-polarization results, whereas other variants degrade, confirming the dual-path design's robustness to polarization diversity.

4.3.2. Robustness Analysis Under Different Signal-to-Noise Ratio Conditions

The recognition accuracy under HH polarization for seven SNR levels (0–30 dB) is shown in Figure 9. Single-source variants exhibit distinct noise sensitivities: Only-HRRP saturates above 15 dB, with a performance drop to 0.5720 at 0 dB. Only-RCS improves slowly with increasing SNR, reflecting that RCS as a zero-dimensional scalar carries limited independent discriminative information. Only-Statistic performs poorly at low SNR but recovers to 0.8379 at 30 dB, because statistical estimates depend on the HRRP amplitude distribution; low SNR distorts the distribution, while cleaner signals restore estimation accuracy. Fixed-weight fusion achieves only 0.2119 at 0 dB, even lower than the best single source, indicating that without a learning mechanism, noise from low-quality sources may harm fusion results.

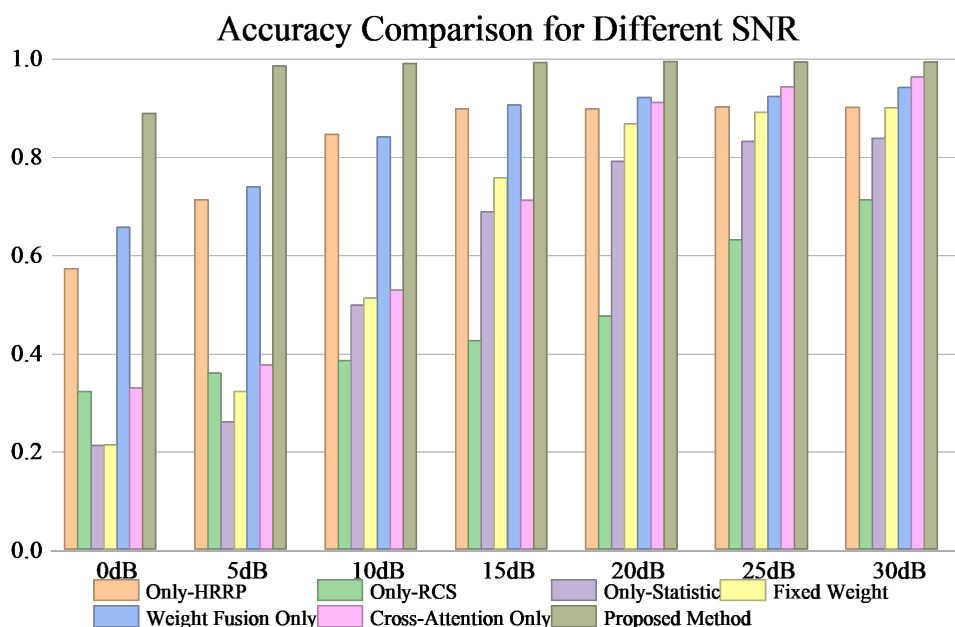


Figure 9. Recognition accuracy comparison of seven networks under different signal-to-noise ratios (HH polarization).

Comparing the three variants, Fixed Weight, Weight Fusion Only, and Cross-Attention Only, we can separate the respective contributions of the two fusion mechanisms under different SNRs. Weight Fusion Only consistently outperforms Fixed Weight across all SNR levels. For example, at 0 dB, Weight Fusion Only achieves 0.6562 accuracy, whereas Fixed Weight attains only 0.2119. This demonstrates that trainable weights can estimate relative source reliability even under extreme noise. Cross-Attention Only performs relatively weakly at low SNRs but catches up quickly as SNR increases, exceeding Weight Fusion Only above 25 dB. This suggests that adaptive weight fusion provides robust global balancing when noise is severe, while cross-source interaction plays a stronger complementary role when signal quality is good. When both mechanisms work together, the full model maintains 0.8888 at 0 dB and stabilizes above 0.98 for $\text{SNR} \geq 5$ dB, significantly outperforming any single-mechanism variant at all SNRs. This suggests a complementary relationship: adaptive fusion performs noise-adaptive contribution assignment at the global level, while

cross-attention establishes semantic correspondences at the feature level. Together they form a layered response to information quality variations, ensuring performance stability over a wide SNR range.

4.4. Overall Performance Comparison Experiments

To validate the effectiveness of the proposed method, it is compared with six representative baseline methods mentioned in the Introduction. These methods include: ResNet18 [20] based on residual networks, CNN_ELM [21] combining convolutional neural networks and extreme learning machine, AC_GRU [25] based on attention-enhanced convolutional gated recurrent units, ConvLSTM_GRU [26] based on ConvLSTM and multi-input gated recurrent units, BiLSTM [28] based on bidirectional long short-term memory networks, and LMBP [29] based on classifier fusion. Among them, ResNet18, CNN_ELM, and AC_GRU use only HRRP data, while ConvLSTM_GRU uses both HRRP raw data and its statistical features; BiLSTM and LMBP use only RCS data. The proposed method simultaneously utilizes HRRP, RCS, and statistical features for multi-modal fusion. All experiments are based on HH polarization data and are conducted under three SNR conditions. Recall, Precision, F1-Score, and Accuracy are adopted as evaluation metrics.

As shown in Table 3, the proposed three-channel dual-path cooperative fusion network achieves the best performance under all SNR conditions. Particularly in a low-SNR environment, the proposed method attains an accuracy of 98.41%, with Recall, Precision, and F1-Score all exceeding 0.984, significantly outperforming the other compared methods.

Table 3. Performance comparison of different methods under various signal-to-noise ratio conditions.

Network	Data Source	SNR	Recall	Precision	F1	Acc
ResNet18 [20]	HRRP	5 dB	0.7021	0.6114	0.6430	0.7023
		10 dB	0.8112	0.8889	0.7709	0.8114
		15 dB	0.9820	0.9825	0.9820	0.9820
CNN_ELM [21]	HRRP	5 dB	0.2164	0.4710	0.1331	0.2161
		10 dB	0.4397	0.4644	0.3829	0.4396
		15 dB	0.6547	0.7370	0.6852	0.7373
AC_GRU [25]	HRRP	5 dB	0.3637	0.2931	0.3186	0.3856
		10 dB	0.6563	0.6834	0.6553	0.7050
		15 dB	0.8809	0.8931	0.8811	0.8910
ConvLSTM_GRU [26]	HRRP, Statistical	5 dB	0.7581	0.7585	0.7176	0.7845
		10 dB	0.8382	0.8174	0.8860	0.8365
		15 dB	0.8858	0.8179	0.8811	0.8850
BiLSTM [28]	RCS	5 dB	0.4382	0.4200	0.4252	0.4380
		10 dB	0.5266	0.5597	0.5224	0.5265
		15 dB	0.6225	0.6318	0.6223	0.6224
LMBP [29]	RCS	5 dB	0.4073	0.4003	0.3978	0.4072
		10 dB	0.4895	0.4821	0.4828	0.4895
		15 dB	0.5757	0.5709	0.5713	0.5757
Proposed	HRRP, RCS, Statistical	5 dB	0.9841	0.9846	0.9842	0.9841
		10 dB	0.9915	0.9916	0.9915	0.9915
		15 dB	0.9942	0.9942	0.9942	0.9942

Comparing the experimental results of different data source combinations under various SNR levels, a regular pattern of recognition performance improvement with the richness of data sources can be observed. At 5 dB, the accuracy of ResNet18 using only a single HRRP data source is 70.23%; adding statistical features as a dual-data-source input, ConvLSTM_GRU [26] achieves 78.45%; the proposed method, with joint input of HRRP,

RCS, and statistical features, reaches 98.41%. The contribution of statistical features exhibits an SNR-dependent characteristic: ConvLSTM_GRU performs similarly to AC_GRU under high SNR, but its advantage expands to about 40 percentage points under low SNR. This trend indicates that the distribution-level information provided by statistical features yields limited marginal gain when the HRRP signal quality is high, but provides effective complementary discriminative information when HRRP is severely affected by noise. The addition of RCS further expands the available information dimension—although RCS alone has limited recognition capability, its macroscopic electromagnetic scattering characteristics are physically independent of the microscopic structural information of HRRP.

When SNR decreases from 15 dB to 5 dB, single-source methods degrade by 28–52 percentage points; ConvLSTM_GRU degrades by about 10 points; while the proposed method degrades by only about 1 point. These differences stem from three levels: different noise response mechanisms at the input, cross-source attention at the feature interaction level, and adaptive fusion weights at the decision level. Together, they form a multi-layer buffering mechanism, explaining the stable performance of the proposed method under low-SNR conditions.

4.5. Visualization and Interpretability Analysis

4.5.1. Feature Heatmap Analysis

To elucidate the discriminative basis of the model at the feature level and validate the effectiveness of the multimodal fusion mechanism, activation intensity visualization is conducted on feature tensors from individual modalities and their fused representations. Figure 10 presents the analysis results for sample No. 662 of Ship1.

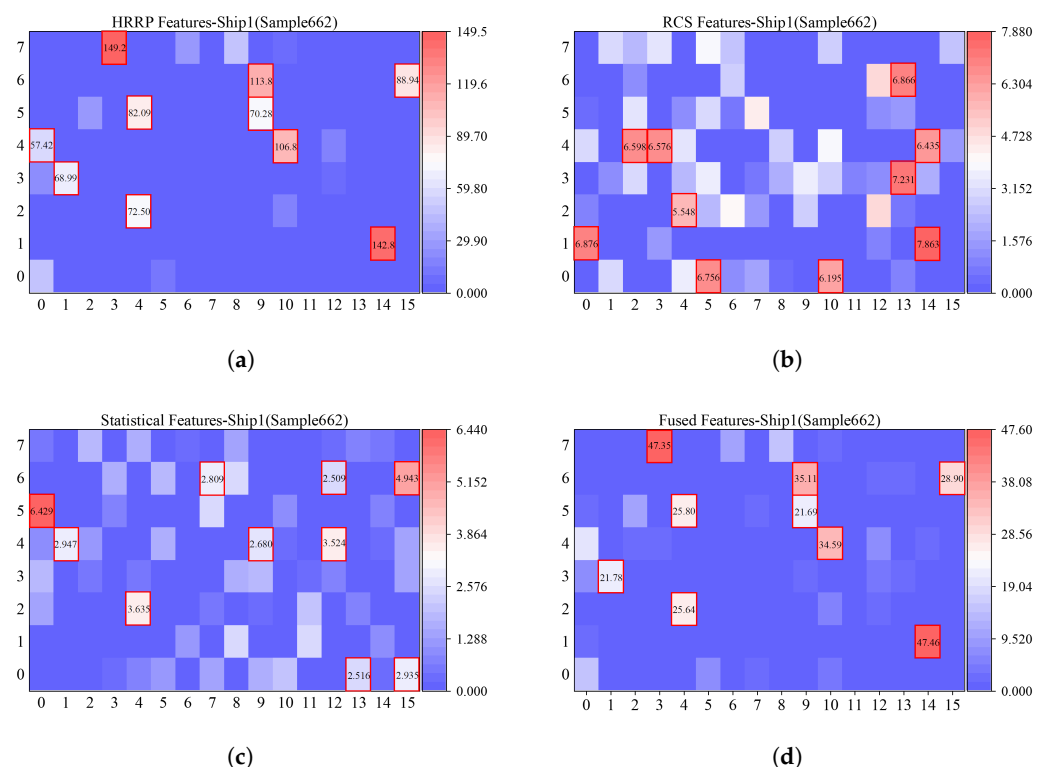


Figure 10. Feature heatmaps: (a) HRRP, (b) RCS, (c) Statistical, (d) Fused.

The heatmap of HRRP features (a) demonstrates a prominent discrete peak activation pattern. Multiple isolated strong activation units are distributed across the front and middle regions of the feature dimension, whose morphology highly corresponds to the distribution

of strong scattering centers of the target in the radar radial direction. This indicates that the HRRP channel primarily focuses on local fine structures.

The heatmap of RCS features (b) exhibits continuous block-like activation characteristics. Its activated regions are more widely distributed, forming relatively uniform response bands in the middle to latter segments, reflecting the macroscopic distribution of electromagnetic scattering intensity across different observation angles.

The heatmap of statistical features (c) displays global and smooth characteristics, with relatively uniform activation throughout the entire feature dimension without drastic peak fluctuations. This verifies that this modality originates from higher-order encoding of macroscopic statistical measures, possessing stronger anti-interference robustness.

The heatmap of fused features (d) reveals a novel distribution resulting from integration and reconstruction. It is not simply an overlay of the three previous heatmaps: critical discrete peak information from HRRP is preserved and enhanced, while continuous distribution characteristics of RCS and global stability of statistical features are incorporated, ultimately forming a more concentrated and discriminative core activation region in the middle of the feature space.

4.5.2. Cross-Attention Mechanism Visualization

To deeply investigate cross-modal interactions within the dual-path fusion mechanism, a visual analysis of the ternary parallel cross-attention module is conducted. Figure 11 illustrates the dynamic semantic relationships using sample No. 107 of Ship1.

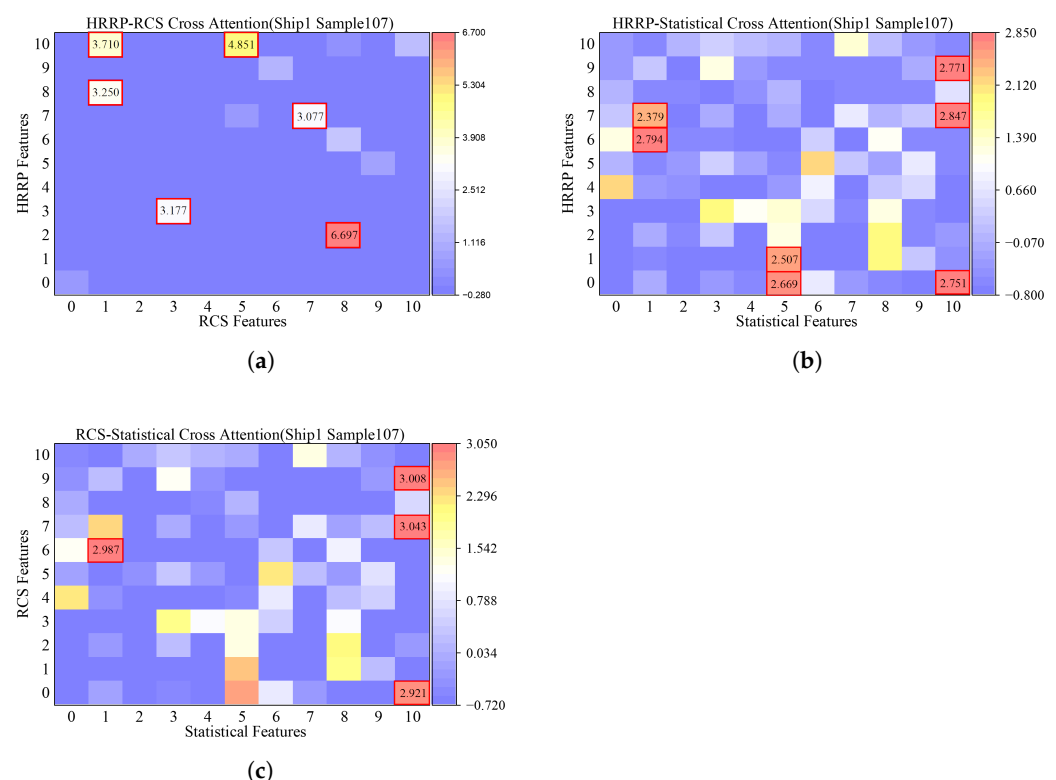


Figure 11. Cross-attention heatmaps: (a) HRRP-RCS, (b) HRRP-Statistical, (c) RCS-Statistical.

The three sub-modules exhibit significantly distinct interaction patterns. HRRP-RCS attention (a) demonstrates concentrated and high-intensity focusing, with a few prominent local blocks, indicating strong semantic binding between specific structural units in HRRP and specific angular scattering intensity units in RCS. HRRP-statistical attention (b) displays relatively dispersed but extensive interactions, with more uniform attention weight distribution and multiple medium-strength connection regions, reflecting multi-

scale contextual supplementation between raw data and derived priors. Statistical features serve as prior knowledge, providing robust background for HRRP's detailed features, while HRRP features in turn refine and validate the statistical descriptions. RCS-statistical attention (c) exhibits medium-strength block-wise associations, reflecting their synergy at the macroscopic electromagnetic attribute level, helping the model stably perceive overall electromagnetic properties such as scattering stability and fluctuation patterns.

4.5.3. Adaptive Fusion Weight Visualization

Figure 12 shows the adaptive fusion weight evolution under different polarization conditions. Under full polarization, weights stabilize at HRRP (0.147), RCS (0.391), and statistical features (0.462). Under HH polarization, the HRRP weight increases to 0.315 with faster convergence. Under HV polarization, the HRRP weight further increases to 0.333 while the RCS weight reduces to 0.383, compensating for degraded feature quality. These differentiated strategies reflect the model's adaptive capability based on electromagnetic scattering physics.

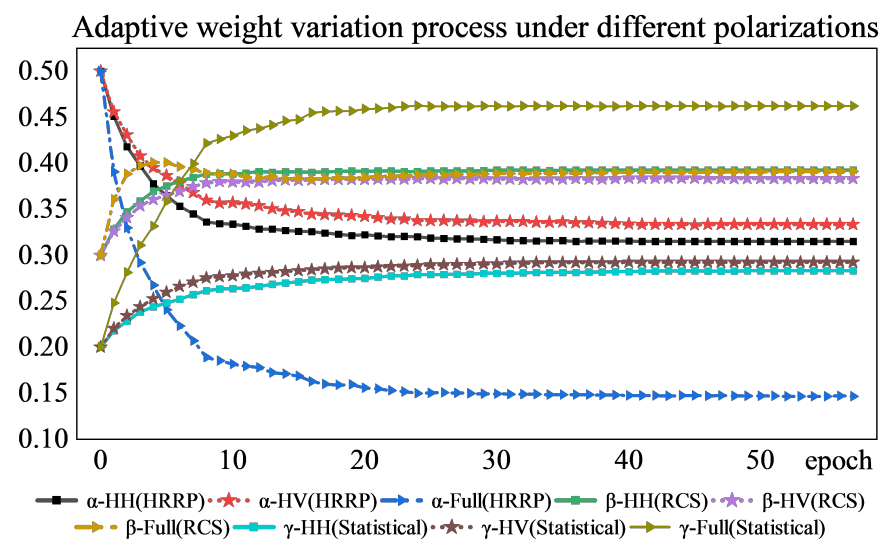


Figure 12. Adaptive fusion weight evolution.

4.5.4. t-SNE Dimensionality Reduction Visualization

Using t-SNE for dimensionality reduction, the clustering distribution of six ship targets in the feature space is observed (Figure 13). Dimensionality reduction analysis based on t-SNE feature distribution reveals significant differences in the feature representation capabilities of each branch [47]. Single-modality features exhibit limited discriminability: HRRP features (a) show mixed sample points with blurred inter-class boundaries; RCS features (b) present some aggregation tendency but still significant overlap; statistical features (c) are relatively better, but confusion remains serious. In contrast, the fused features (d) exhibit optimal clustering performance: samples within each category are highly aggregated, while clear separation boundaries are maintained between clusters. This strongly proves that fusing HRRP, RCS, and statistical features effectively integrates and complements characteristic information from different sources, significantly enhancing feature discriminability.

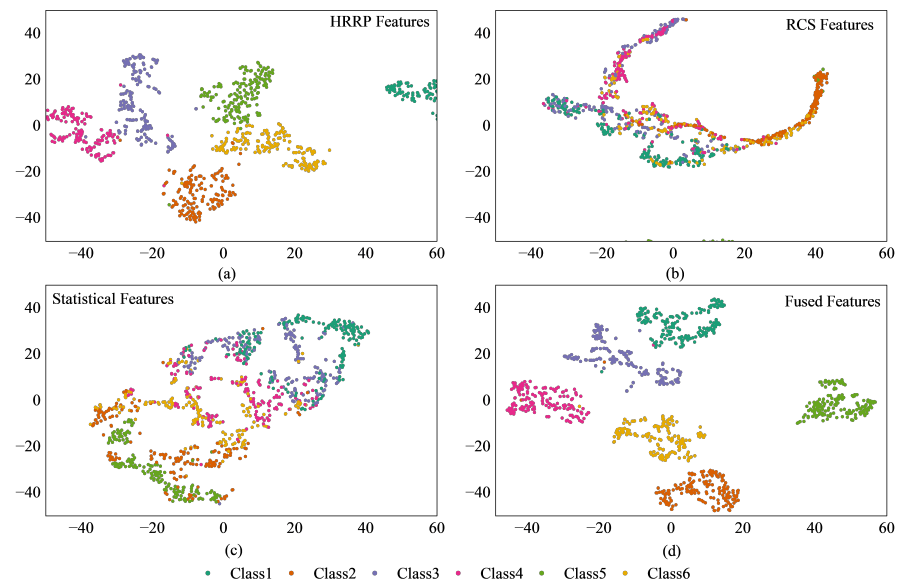


Figure 13. t-SNE feature space visualization: (a) HRRP, (b) RCS, (c) Statistical, (d) Fused.

5. Discussion

Through electromagnetic simulation, HRRP and RCS data for six ship targets were acquired, constructing an experimental dataset with varying SNR conditions. The research employs a three-channel network architecture to process HRRP sequences, RCS sequences, and statistical features separately, achieving effective feature integration through a dual-path fusion mechanism.

Compared with existing methods, the proposed approach addresses the issues of relying on a single data source and lacking adaptability in multimodal fusion, which are unable to model deep semantic relationships. Even though the ConvLSTM_GRU [26] explores multimodal fusion, it relies on fixed weights and cannot adapt to varying conditions. Instead, the adaptive weight layer in our method can allocate importance to each modality in a data-driven manner, and the cross-attention module establishes inter-modal correlations, leading to the best performance for the ship target recognition problem, as discussed in Table 3. Notably, under full-polarization conditions, accuracy improves from 80.49% to 99.39%, validating the superiority of the dual-path cooperative fusion mechanism.

Future research will focus on extending this approach to a broader range of ship types, validating it with real measured radar data, and addressing the more challenging composite scattering problem of ships on sea surfaces. These efforts are expected to further demonstrate the generalizability and robustness of the proposed framework in increasingly complex and practical scenarios.

6. Conclusions

This paper addresses the challenges of radar target recognition in complex electromagnetic environments by proposing a novel recognition method based on a “mechanism-data-physics” trinity fusion paradigm.

Experimental results demonstrate that the proposed method achieves 99.42% accuracy under HH polarization at 15 dB SNR and 98.41% at 5 dB SNR, significantly outperforming comparative methods. Ablation studies confirm the effectiveness of both the cross-attention module and the adaptive weight fusion layer, showing substantial improvements over the fixed-weight baseline (1/3).

The core contribution of this work lies in proposing a methodological framework for ship target recognition.

Author Contributions: Conceptualization: S.S., W.Y., S.L., X.Z., J.T. and H.H.; methodology: S.S. and W.Y.; Validation: X.Z. and S.L.; writing—original draft preparation: S.S. and W.Y.; writing—review and editing: W.Y., H.H. and X.Z.; supervision: J.T. and H.H.; project administration: H.H.; funding acquisition: H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62471108, U25A20412) and the Natural Science Foundation of Sichuan Province (2025ZNSFSC0530).

Data Availability Statement: The data presented in this study are available upon request from the corresponding author due to privacy restrictions.

Acknowledgments: The authors would like to thank the reviewers for their valuable comments and suggestions. During the preparation of this manuscript, the authors used DeepSeek-V3 for language polishing and grammatical refinement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

References

1. Zhao, X.; Zhu, Y.; Chen, Z.; Xu, D. Marine target detection and recognition method based on YOLO neural network in embedded system. In Proceedings of the IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), Haikou, China, 18–20 August 2023; pp. 229–235.
2. Zhang, H.; Wang, W.; Deng, J.; Guo, Y.; Liu, S.; Zhang, J. MASFF-Net: Multiazimuth scattering feature fusion network for SAR target recognition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 19425–19440. [\[CrossRef\]](#)
3. Wang, H.; Liu, C.; Yu, L.; Zhao, J. Research on target detection and recognition algorithm based on deep learning. In Proceedings of the Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 8483–8487.
4. Li, Y.; Zhang, D.; Fan, L.; Ma, H.; Xu, Z. Performance analysis of ship target recognition in multi-band infrared images based on deep learning. In Proceedings of the 7th International Conference on Transportation Information and Safety (ICTIS), Xi'an, China, 4–6 August 2023; pp. 432–436.
5. Pan, Y.; Liu, X.; Zhang, J.; Wang, Q.; Chen, L. Fine-grained recognition framework for ship targets in SAR images based on component architecture. In Proceedings of the IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Zhuhai, China, 22–24 November 2024; pp. 1–5.
6. He, J.; Chang, W.; Wang, F.; Wang, Q.; Li, Y.; Gan, Y. Polarization matters: On bilinear convolutional neural networks for ship classification from synthetic aperture radar images. In Proceedings of the 4th International Conference on Natural Language Processing (ICNLP), Xi'an, China, 25–27 March 2022; pp. 315–319.
7. Song, Y.; Wang, X.; Zhang, L.; Liu, Z. Analysis of BiSAR images characteristics: From an electromagnetic scattering perspective. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *62*, 2852–2869. [\[CrossRef\]](#)
8. Guo, Y.; Xiao, H. Pruning support vector data description method for HRRP-based radar target recognition. In Proceedings of the International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 10–11 December 2016; pp. 330–333.
9. Zhu, L.; Xiong, G.; Yu, W. Radar HRRP group-target recognition based on combined methods in the background of sea clutter. In Proceedings of the International Conference on Radar (RADAR), Brisbane, QLD, Australia, 27–31 August 2018; pp. 1–6.
10. Tang, W.; Yu, L.; Wei, Y.; Tong, P. Radar target recognition of ballistic missile in complex scene. In Proceedings of the IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–6.
11. Sehgal, B.; Shekhawat, H.S.; Jana, S.K. Automatic target recognition using recurrent neural networks. In Proceedings of the International Conference on Range Technology (ICORT), Balasore, India, 15–17 February 2019; pp. 1–5.
12. Rawat, A.; Sharma, A.; Awasthi, A. Machine learning based non-cooperative target recognition with dynamic RCS data. In Proceedings of the IEEE Wireless Antenna and Microwave Symposium (WAMS), Ahmedabad, India, 7–10 June 2023; pp. 1–5.
13. Mansukhani, J.; Penchalaiah, D.; Bhattacharyya, A. RCS based target classification using deep learning methods. In Proceedings of the 2nd International Conference on Range Technology (ICORT), Chandipur, Balasore, India, 5–6 August 2021; pp. 1–5.
14. Chen, Y.; Luo, N. A ship radar cross section recognition method based on bi-directional gated recurrent unit net. In Proceedings of the IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 8–10 December 2023; pp. 512–515.
15. Bai, C.; Peng, N.; Yu, Y.; Li, Z.; Wang, X. Space targets classification using broadband RCS sequences. In Proceedings of the International Conference on Microwave and Millimeter Wave Technology (ICMMT), Beijing, China, 16–19 May 2024; pp. 1–3.

16. Shi, G.; Da, L.; Hu, J. Feature extraction study of an automatic target recognition. In Proceedings of the 2nd International Conference on Information Science and Engineering (ICISE), Hangzhou, China, 4–6 December 2010; pp. 5117–5120.
17. Wang, C.; Zhang, W.; Wang, J.; Wu, Y. Few-samples target recognition based on dual-generation network data augmentation and feature fusion. *IEEE Signal Process. Lett.* **2025**, *32*, 3984–3988. [[CrossRef](#)]
18. Fu, Z.; Li, S.; Li, X.; Dan, B.; Wang, X. A performance analysis of neural network models in HRRP target recognition. In Proceedings of the IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–4.
19. Yang, X.; Cheng, G.; Yang, L.; Huang, Z. High-resolution range profile target recognition based on GWO-RBF-BP neural network. In Proceedings of the IEEE 11th Asia-Pacific Conference on Antennas and Propagation (APCAP), Guangzhou, China, 19–22 November 2023; pp. 1–2.
20. Ding, B.; Chen, P. HRRP feature extraction and recognition method of radar ground target using convolutional neural network. In Proceedings of the International Conference on Electromagnetics in Advanced Applications (ICEAA), Granada, Spain, 9–13 September 2019; pp. 658–661.
21. Zhang, Q.; Lu, J.; Liu, T.; Deng, J.; Guo, Y. Ship HRRP target recognition based on CNN and ELM. In Proceedings of the 4th International Conference on Electromechanical Control Technology and Transportation (ICECTT), Guiyang, China, 26–28 April 2019; pp. 1–5.
22. Chen, S.; Huang, X. Radar HRRP open set target recognition based on supervised contrastive learning. In Proceedings of the 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE), Xi'an, China, 19–21 January 2024; pp. 1–5.
23. Huang, P.; Li, S.; Zheng, M.; Wang, Y.; Zhao, L. Noise-robust HRRP target recognition based on residual scattering network. In Proceedings of the 2024 International Conference on Radar Signal Processing (ICRSP), Nanjing, China, 12–14 July 2024; pp. 1–5.
24. Guo, D.; Chen, B.; Chen, W.; He, F.; Liu, H. Variational temporal deep generative model for radar HRRP target recognition. *IEEE Trans. Signal Process.* **2020**, *68*, 5795–5809. [[CrossRef](#)]
25. Chu, Y.; Guo, Z. Attention enhanced spatial temporal neural network for HRRP recognition. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 1–5.
26. Yang, W.; Chen, T.; Lei, S.; Zhao, Z.; Hu, H.; Hu, J. Radar HRRP feature fusion recognition method based on ConvLSTM network with multi-input gate recurrent unit. *Remote Sens.* **2024**, *16*, 4533. [[CrossRef](#)]
27. Zhu, M.; Qiao, X.; Jiang, J.; Wang, L.; Zhang, Y. Flight target recognition method based on RCS sequence. In Proceedings of the 8th International Conference on Electrical, Mechanical and Computer Engineering (ICEMCE), Shanghai, China, 25–27 October 2024; pp. 1–5.
28. Wang, Y.; Long, B.; Wang, F. RCS statistical feature extraction for space target recognition based on bi-LSTM. In Proceedings of the IEEE 6th International Conference on Electronics, Information and Communication Technology (ICEICT), Nanjing, China, 21–24 July 2023; pp. 1–5.
29. Xueman, F.; Shengliang, H.; Jingbo, H. Research on target recognition of low resolution anti-ship missile based on classifier combination. In Proceedings of the 3rd International Conference on Systems and Informatics (ICSAI), Shanghai, China, 19–21 November 2016; pp. 1–5.
30. Zhang, L.; Xue, B.; Zhou, Y.; Chen, W.; Liu, Z. Intelligent classification of spatial targets based on RCS sequences. In Proceedings of the Applied Computational Electromagnetics Society Symposium (ACES), Monterey, CA, USA, 26–30 March 2023; pp. 1–5.
31. Zhu, C.; Liu, Q.; Gu, D.; Yang, Y.; Wu, L. Fine-grained vehicle target recognition in SAR images integrating RCS distribution feature. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Athens, Greece, 7–12 July 2024; pp. 1–4.
32. Rius, J.M.; Ferrando, M.; Jofre, L. High-frequency RCS of complex radar targets in real-time. *IEEE Trans. Antennas Propag.* **1993**, *41*, 1308–1319. [[CrossRef](#)]
33. Xu, B.; Chen, B.; Liu, J.; Du, C. Gaussian mixture model-tensor recurrent neural network for HRRP target recognition. In Proceedings of the International Radar Conference (RADAR), Toulon, France, 23–27 September 2019; pp. 1–6.
34. Wensheng, L.; Hongshi, Z.; Kun, W.; Zhen, W.; Weilong, C. Short-term net load forecasting model based on GRU neural network optimized by improved arithmetic optimization algorithm. In Proceedings of the IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China, 27–29 September 2024; pp. 52–56.
35. Zhang, Y. Prediction model of UAV damage efficiency based on GRU neural network. In Proceedings of the 10th International Conference on Mechanical and Electronics Engineering (ICMEE), Xi'an, China, 27–29 December 2024; pp. 334–338.
36. Zhang, S.; Li, G.; Gu, Z.; Li, J.; Ye, X.; Ren, Y. Research on demand response strategy of electric vehicles based on GRU neural network. In Proceedings of the China International Conference on Electricity Distribution (CICED), Changsha, China, 7–8 September 2022; pp. 92–96.

37. Wang, Y.; Chang, J.; Lu, Z.; Yu, F.; Wei, J.; Xu, Y. Channel estimation of 5G OFDM system based on ConvLSTM network. In Proceedings of the 7th International Conference on Communication, Image and Signal Processing (CCISP), Chengdu, China, 18–20 November 2022; pp. 62–66.
38. Bo, M.; Ning, Y.; Chenggang, C.; Jing, C.; Peifeng, X. Cloud position forecasting based on ConvLSTM network. In Proceedings of the 5th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 12–14 September 2020; pp. 562–565.
39. Chen, W.; Zhang, L.; Xi, Y.; Wang, Y.; Li, Y. Polarimetric HRRP target recognition based on ConvLSTM. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Yokohama, Japan, 28 July–2 August 2019; pp. 334–337.
40. Zhong, J.; Zhou, Y.; Shen, Y.; Yuan, J.; Wu, X. ChanCrossFormer: A ballistic trajectory prediction model integrating channel attention and cross-attention mechanisms. In Proceedings of the 4th Conference on Fully Actuated Systems: Theory and Applications (FASTA), Nanjing, China, 4–6 July 2025; pp. 1444–1449.
41. Yijun, L.; Jianmei, S.; Haoping, S. An improved Siamese tracking network based on self-attention and cross-attention. In Proceedings of the 35th Chinese Control and Decision Conference (CCDC), Yichang, China, 20–22 May 2023; pp. 466–470.
42. Wu, D.; Wang, Y.; Wu, X.; Qu, T. Cross-attention inspired selective state space models for target sound extraction. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hyderabad, India, 6–11 April 2025; pp. 1–5.
43. Jhong, S.-Y.; Hsu, C.-W.; Fang, W.-H.; Chen, Y.-C. Camera-radar-based adaptive cross attention fusion for multimodal object detection. In Proceedings of the International Automatic Control Conference (CACS), Taoyuan, Taiwan, 31 October–3 November 2024; pp. 1–6.
44. Yang, W.; Liao, C.; Zhu, X.; Qi, C.; Lei, S.; Hu, H. An effective high-frequency method for the near-field scattering from an electrically large ship illuminated by a Hertzian dipole. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3507805. [[CrossRef](#)]
45. Zhang, C.-W.; Zhao, Z.-Q.; Yang, W.; Zhou, L.-L.; Zhu, H.-Y. Electromagnetic scattering and imaging simulation of extremely large-scale sea-ship scene based on GPU parallel technology. *J. Electron. Sci. Technol.* **2024**, *22*, 100257. [[CrossRef](#)]
46. Baden, J.M.; Tripp, V.K. Ray reversal in SBR RCS calculations. In Proceedings of the 31st International Review of Progress in Applied Computational Electromagnetics (ACES), Williamsburg, VA, USA, 22–26 March 2015; pp. 1–2.
47. Takamatsu, K.; Oyamada, M.; Kato, Y.; Aoki, S. A new way of visualizing curricula using competencies: Cosine similarity and t-SNE. In Proceedings of the 7th International Congress on Advanced Applied Informatics (IIAI-AAI), Yonago, Japan, 8–12 July 2018; pp. 390–395.

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.