ISSN: 2664-9640

A Lightweight SAR Ship Classification Method based on VV/VH Dual-Polarization Fusion and Knowledge Distillation

Xuanyu Bai[a]

School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing, China

[a]E-mail: bai_xuanyu@126.com

Abstract

Synthetic Aperture Radar (SAR) can acquire clear and reliable radar images under complex weather conditions such as rain and fog, offering significant research and application value in maritime surveillance, port monitoring, and ship target recognition. However, existing SAR ship target classification models still face challenges of large parameter counts and low inference efficiency in edge computing scenarios, and the classification performance for certain target categories remains unsatisfactory in practical situations involving limited samples and imbalanced data. To address these issues, this paper adopts a lightweight SAR ship target classification approach that combines VV-VH dual-polarization feature fusion with model compression. Using OpenSARShip_uint8 as the experimental dataset, the method first constructs dual-channel inputs that leverage the complementary scattering response information from VV and VH polarization channels to enhance the model's ability to represent ship structural and textural features, thereby improving classification stability. Second, to build a high-quality teacher model, this paper employs VGG16 pre-trained on ImageNet and designs a two-stage training strategy: in the first stage, the feature extraction layers are frozen and only the classifier is trained; in the second stage, all network parameters are unfrozen, and a differentiated learning rate strategy is applied for fine-tuning the entire network-a smaller learning rate is used for the feature extraction layers to preserve pre-trained features, while a relatively larger learning rate is used for the classifier to accelerate adaptation, combined with cosine annealing scheduling to gradually decay the learning rate. This strategy helps improve the model's capacity to represent fine-grained features in SAR images and mitigates, to a certain extent, the model bias caused by data imbalance. Furthermore, knowledge distillation is employed to transfer the feature learning capability of the teacher model to a lightweight student model, significantly reducing the model size while preserving recognition accuracy as much as possible. Experimental results show that the teacher model achieves a test accuracy of 74.08%, while the distilled student model attains a test accuracy of 73.86% with only a 0.22% accuracy degradation. Despite the moderate absolute classification accuracy limited by speckle noise, inter-class scattering ambiguity and dataset imbalance, the proposed distillation framework drastically shrinks the model scale: parameters drop from 18.92M to 1.24M, storage footprint falls from 72.17MB to 4.74MB, and single-image inference latency decreases from 1.93ms to 0.71ms.

Keywords

SAR; Machine Learning; Knowledge Distillation; Ship Classification.

1. Introduction

Synthetic Aperture Radar (SAR) possesses all-day, all-weather imaging capabilities and exhibits strong robustness against adverse meteorological conditions such as rain, fog, and complex sea

ISSN: 2664-9640

states, offering significant application value in maritime vessel monitoring, port security surveillance, and intelligent maritime applications. In the SAR image interpretation pipeline, ship target classification is a core task, and its recognition accuracy and model lightweight level directly determine the practical deployment feasibility and engineering effectiveness of related systems in edge computing scenarios.

In the field of SAR ship target recognition, current technical research still faces dual challenges from both the data level and the neural network model level. On the data side, constrained by the imaging principles of radar, SAR images exhibit inherent speckle noise, which causes blurred target edges and unstable scattering characteristics [1], thereby interfering with the neural network's feature extraction, leading to model classification errors and unsatisfactory recall rates for certain ship targets [2]. Although the OpenSARShip public dataset already provides VV (vertical transmit, vertical receive) and VH (vertical transmit, horizontal receive) dual-polarization data, VV and VH polarization carry distinct scattering signatures of ship hulls and superstructures, and their complementary information can effectively suppress speckle noise interference [3]. Nevertheless, how to efficiently fuse dual-polarization information and fully exploit their complementary advantages still requires more targeted network design. On the model side, traditional machine learning methods such as support vector machines and random forests rely on manually designed features and have limited representation capability in complex scenarios; while deep learning models represented by VGG16 possess stronger feature extraction capabilities, they generally suffer from large parameter counts, high computational complexity, and slow inference speed. In practical maritime applications, device computational resources, storage capacity, and power consumption are all strictly constrained. Therefore, achieving lightweight SAR ship classification models while preserving recognition accuracy as much as possible has become a key research direction for engineering deployment in this field.

In existing SAR ship target recognition research, although traditional deep learning models achieve relatively high classification accuracy, they generally suffer from large parameter counts and slow inference speed, making them difficult to adapt to resource-constrained edge computing scenarios such as embedded platforms and low-power terminals. In terms of feature utilization, some methods rely solely on single-polarization SAR images and fail to fully exploit the complementary characteristics of VV and VH dual-polarization channels, resulting in limited effectiveness in suppressing speckle noise [4]. Some methods introduce dual-polarization feature enhancement networks to improve ship feature representation capability, but without incorporating model compression and knowledge distillation, they cannot achieve lightweight deployment [2]. At the data level, limited samples and class imbalance issues tend to cause model bias, leading to low recognition accuracy for minority class targets [5]. Meanwhile, existing lightweight schemes mostly apply direct pruning or network simplification, making it difficult to maintain high accuracy while compressing the model [6]. Therefore, a SAR ship recognition method that simultaneously achieves "efficient dual-polarization feature fusion, small-sample robustness improvement, and high-accuracy model lightweight deployment" still presents a notable research gap.

To address the aforementioned limitations, this paper proposes a lightweight SAR ship target recognition method that integrates dual-polarization features with knowledge distillation. Using the OpenSARShip dataset as the experimental subject, a dual-channel input is constructed from VV+VH dual-polarization data, leveraging polarimetric complementarity to suppress speckle noise and enhance the representation of ship structural features. On this basis, a high-accuracy teacher model is constructed based on a VGG16 network pre-trained on ImageNet, employing a two-stage training strategy to improve fine-grained feature learning capability and mitigate model bias induced by class imbalance. To meet the deployment requirements of edge computing scenarios, a lightweight student network is further designed, and knowledge

ISSN: 2664-9640

distillation is employed to efficiently transfer both feature-level knowledge and discriminative knowledge from the teacher model to the student network, thereby significantly reducing model parameters and computational overhead while maximally preserving the original recognition accuracy. The overall method achieves a favorable balance among recognition accuracy, computational efficiency, and engineering practicability.

2. Related Works

The classification of ship targets in SAR imagery has undergone an evolution from traditional statistical models to data-driven deep learning models. Existing research can be broadly categorized into traditional classification algorithms and deep learning-based classification algorithms.

Traditional SAR ship classification methods often rely on the statistical distribution of background clutter and Constant False Alarm Rate (CFAR) techniques. For instance, Hou et al. [7] proposed a multi-layer CFAR ship detection algorithm incorporating a log-normal distribution model. Such methods do not require large volumes of annotated data; however, they typically depend on manually predefined statistical assumptions and post-processing rules. In scenarios involving complex coastal backgrounds, strong scattering from man-made structures, speckle noise, and ship targets with significant scale variations, the background distribution becomes difficult to model accurately, thereby leading to elevated false alarm rates or missed detections.

In recent years, deep learning-based SAR ship classification methods have achieved remarkable progress. Compared with traditional CFAR approaches, deep learning models are capable of automatically learning high-level semantic features of ship targets from data, demonstrating superior robustness in complex backgrounds, small target classification, and multi-scale target recognition. Li et al. [8] proposed an improved Faster R-CNN-based SAR ship detection method, demonstrating that employing VGG as the backbone network in conjunction with feature fusion strategies can significantly enhance classification performance for small targets. Subsequently, detection frameworks such as YOLO and RetinaNet have been extensively applied to SAR ship detection tasks. Zhi et al. [6] proposed MBNet based on YOLOv5, incorporating multi-branch convolution, attention mechanisms, and deformable convolution to enhance multi-scale ship features while suppressing complex coastal background interference. To address the high cost of annotation and the insufficiency of labeled samples in SAR imagery, Yang et al. [5] proposed a Dynamic Data Augmentation Method (DDAM), which improves classification performance under limited annotation conditions through iterative pseudo-label updating. Overall, although existing deep learning methods have effectively improved SAR ship classification accuracy, the model architectures have become increasingly complex, with growing numbers of parameters and computational costs, posing considerable challenges for real-time deployment on resource-constrained platforms such as spaceborne and airborne systems.

Although the complex deep learning models mentioned above have achieved remarkably high classification accuracy, their substantial parameter counts and high floating-point operations (FLOPs) introduce severe computational latency. On edge computing platforms such as spaceborne or airborne radar systems, hardware resources and power consumption are subject to stringent constraints. The VGG network, owing to its clean and regular architecture and superior feature extraction capability, remains a widely adopted classical model in the field of target classification; however, its parameter redundancy issue limits practical deployment.To address the growing complexity of deep neural network architectures, the research community has developed several model compression paradigms centered on pruning, quantization, and knowledge distillation.[9]

ISSN: 2664-9640

Pruning serves as a fundamental technique for model lightweighting by eliminating low-contribution weights or channels to reduce model redundancy. Existing approaches can be broadly categorized into global unstructured weight pruning and structured channel- or layer-level pruning[10].The L1-norm filter pruning adopted in this paper is a classic structured pruning scheme with excellent hardware adaptability. Letaifa et al. [11] conducted a fine-grained layer-wise pruning analysis on Transformer-based speech recognition models, pointing out that conventional one-size-fits-all pruning strategies—whether global or local—fail to account for the distinct parameter distributions and weight characteristics of heterogeneous modules such as convolutional layers, multi-head attention, and feed-forward networks. Indiscriminately pruning shallow convolutional layers with relatively few parameters was shown to severely degrade recognition accuracy, whereas targeted pruning applied exclusively to parameter-intensive linear layers, such as feed-forward and attention modules, can achieve a favorable trade-off between compression ratio and model performance. This layer-aware differentiated pruning strategy offers an important reference for CNN lightweighting. Rui et al. [12] proposed a joint channel and layer pruning scheme targeting edge-deployed inspection devices, employing Deep Deterministic Policy Gradient (DDPG)-based deep reinforcement learning to adaptively determine the optimal pruning ratio for each convolutional layer. They further introduced an adaptive weight clustering algorithm that combines BIRCH and K-meansII to perform secondary compression on the sparse floating-point weights remaining after pruning, forming a two-stage "pruning–clustering" compression pipeline. This approach achieves 24× model compression on object detection tasks with acceptable accuracy degradation; however, the scheme does not incorporate knowledge distillation to compensate for the feature representation degradation induced by pruning. Wu et al. [13] further established a synergistic pipeline that bridges pruning and knowledge distillation, proposing the use of a pruned pretrained large model as the student network in a distillation framework, where soft label transfer from the teacher network compensates for discriminative feature information removed via pruning. Experiments conducted on VGG-series networks demonstrate that, compared to pruning alone, the combined pruning-and-distillation approach achieves up to 5.81% improvement in classification accuracy on the CIFAR-100 few-shot dataset, confirming the significant complementary gains attainable through the integration of pruning and distillation.

Therefore, to achieve a balance between classification accuracy and inference speed, this paper constructs a VV/VH dual-polarization SAR ship classification input scheme tailored for the OpenSARShip dataset, leveraging dual-channel fusion to fully exploit the complementary scattering information from different polarization modes, thereby enhancing the model's representational capacity for inter-class discrimination among ship categories. A teacher model is built upon ImageNet-pretrained VGG16 with Batch Normalization (VGG16-BN), and a two-stage training strategy is employed-first freezing the feature extraction layers to pretrain the classification head, then fine-tuning the entire network with discriminative learning rates-to improve the transfer learning effectiveness for few-shot SAR ship classification tasks. Concurrently, a lightweight student network is designed, and knowledge distillation is utilized to transfer the soft-label knowledge from the teacher model to the student model, thereby maintaining classification accuracy comparable to that of the teacher model while significantly reducing model parameters, model size, and inference time.

3. Method

3.1. Original Model

To provide a high-accuracy, highly generalizable knowledge source for subsequent structured pruning and knowledge distillation, this work adopts VGG16-BN pre-trained on ImageNet as

ISSN: 2664-9640

the backbone network architecture. In conjunction with the imaging characteristics, noise properties, and dataset scale and class distribution of SAR dual-polarization ship imagery, a task-specific original model architecture and a layer-wise fine-grained training strategy are devised for SAR ship classification.

The core design objective of the original model is to maximize classification accuracy and fine-grained feature representation capability while fully exploiting the complementary features of VV/VH dual-polarization channels, thereby producing predictive probability distributions with strong discriminability and high confidence. This, in turn, furnishes high-quality, highly reliable supervisory signals for knowledge transfer and model compression of the subsequent lightweight student model.

3.1.1. Original Model Structure

The original model employs VGG16-BN as its backbone network. This network features a well-organized stacked convolutional architecture, stable gradient propagation enabled by batch normalization, and mature pre-trained weights, conferring strong adaptability and transferability for image feature extraction tasks. As the original VGG16-BN was designed for 3-channel RGB images, it cannot directly handle the 2-channel VV/VH dual-polarization SAR data used in this study. Therefore, targeted modifications to the input layer and classification head are required.

In terms of input layer adaptation, the number of input channels in the first convolutional layer is modified from 3 to 2, specifically changing Conv2d(3, 64, 3, padding=1) to Conv2d(2, 64, 3, padding=1), while the spatial dimensions of the convolution kernels, stride, and padding remain unchanged. This ensures that the network's receptive field and the feature propagation logic of subsequent layers remain fully consistent. For weight initialization, to avoid random initialization disrupting the effectiveness of the pre-trained features, the original 3-channel convolutional kernel weights are averaged along the channel dimension to obtain a single-channel template of shape [64, 1, 3, 3]. This template is then replicated across the two input channels (VV and VH), as shown in the following Equation (1):

$$W_{NEW}^{(C)} = \frac{1}{3}\sum_{k=1}^{3} W_{PRE}^{(k)}, c = 1,2 \quad (1)$$

Here, W(k) PRE denotes the convolutional kernel weights for the k-th input channel in the pre-trained first convolutional layer, and W(C) NEW represents the convolutional kernel weights for the c-th input channel after adaptation. This operation enables the network to accommodate dual-polarization SAR image inputs while preserving the transferable features from ImageNet. Since the modification affects only the number of input channels in the first convolutional layer, the number of output channels remains to be 64, and the structures and pre-trained weights of all subsequent convolutional layers require no adjustments.

In the feature extraction stage, the full 13 convolutional layers of VGG16-BN, along with their corresponding batch normalization layers and five max-pooling layers, are retained without pruning or modification to ensure that the capacity for extracting deep semantic features remains unaffected. In the classification stage, the original fully connected layers designed for 1000 classes are replaced with a structure tailored for the SAR ship three-classification task: the feature maps are first pooled to a uniform spatial size of 4×4, then flattened into an 8192-dimensional vector (512×4×4), which is sequentially passed through a fully connected layer reducing 8192→512 dimensions, followed by a ReLU activation, a Dropout layer with a dropout probability of 0.5, and finally a 512→3 output layer to obtain the prediction logits for the three ship classes. The schematic diagram of the network architecture is shown in Figure 1.

ISSN: 2664-9640

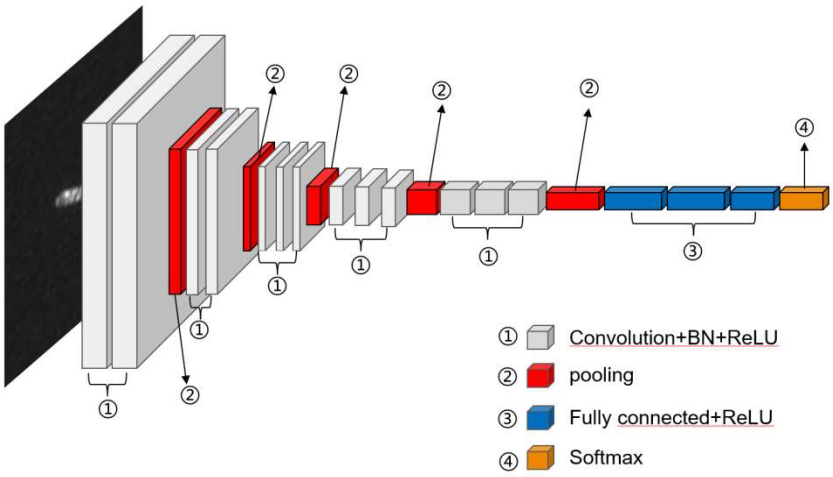


Figure 1. Structure of VGG16

Through the above structural modification, the teacher model achieves end-to-end feature extraction and classification of dual-polarization SAR ship images without compromising the advantages of pretrained feature representations.

3.1.2. Training Strategy of Original Model

Directly performing full-parameter fine-tuning on a pretrained model carries a significant risk: the newly initialized classification head can generate large and unstable gradient signals during the early stages of training. When these gradients are backpropagated to the convolutional layers, they may severely disrupt or even destroy the effective feature representations encoded in the pretrained weights, leading to catastrophic forgetting. This risk is particularly pronounced when there is a substantial domain shift between the source and target datasets. In the present study, SAR images are acquired via microwave coherent imaging, whose scattering texture characteristics differ fundamentally from natural optical images in ImageNet. Moreover, the OpenSARShip_uint8 dataset is relatively small, making full-parameter fine-tuning highly susceptible to rapid degradation of the pretrained general features.

To mitigate this issue, a two-stage hierarchical training strategy [14] was employed during model construction. In the first stage, all convolutional feature layers of VGG16-BN were frozen, and the newly initialized classification head was trained with a learning rate of 3×10-4. This allows the head to quickly converge to a reasonable parameter region within a stable feature space, avoiding severe gradient fluctuations that could impact the pretrained weights. In the second stage, all parameters were unfrozen, and the entire network was fine-tuned with a differential learning rate: the feature extraction layers used 0.1 of the base learning rate to protect low-level general features, while the classification head used 0.5 of the base learning rate to allow high-level semantic representations to adapt fully to the SAR ship classification task. Additionally, a cosine annealing learning rate schedule was applied to smoothly decay the learning rate from its initial value to 10-6 over a maximum of 100 epochs, ensuring stable convergence in later training stages. To prevent overfitting, an early stopping mechanism was introduced, terminating training if the validation loss did not improve for 20 consecutive epochs. Furthermore, gradient clipping with a maximum norm of 2.0 was applied throughout training to prevent gradient explosion and maintain numerical stability.

It should be honestly noted that the final 74.08% test accuracy of this VGG16-BN teacher model is constrained by three inherent drawbacks of the experimental setup: (1) severe speckle noise and highly similar scattering signatures among Cargo, Tanker and Fishing vessels; (2) severe intra-dataset class imbalance with insufficient minority-class Fishing samples; (3) large domain gap between ImageNet optical pre-training weights and microwave SAR imagery. The two-

ISSN: 2664-9640

stage learning rate schedule only mitigates training bias but cannot break this accuracy bottleneck. Subsequent experiments will analyze whether lightweight compression is a reasonable research target under this moderate baseline, and clarify scenarios where improving teacher feature extraction should take precedence over model compression.

Table 1. Training Hyperparameter Configuration of the Original Model

Parameters	Stage 1	Stage 2
Trainable params	Classification Head Only	All params
Optimizer	AdamW	AdamW
LR of Feature Layer	Frozen	3×10^{-5}
LR of Classification Head	3×10^{-4}	1.5×10^{-4}
Weight Decay	1×10^{-4}	1×10^{-4}
LR Schedule	Fixed	Cosine Annealing ($\eta_{min}=10^{-6}$)
Training Epochs	10	100(Early Stopping)
Early Stopping Patience	--	20
Gradient Clipping	2.0	2.0

3.2. Model Compression

3.2.1. Structured Pruning

After obtaining a high-performing baseline model, this study further investigates model pruning by exploring the compression of redundant filters within the teacher model. Pruning is a widely adopted technique in the field of neural network model compression. Its fundamental objective is to identify parameters or structural units that contribute relatively little to the network's performance and weaken or remove them accordingly. In doing so, the redundancy of the model can be reduced and computational efficiency improved while maintaining an acceptable level of classification accuracy degradation.

In this study, pruning is primarily employed to examine whether a substantial amount of parameter redundancy exists within VGG16-BN and to serve as a comparative baseline for subsequent lightweight model design approaches.

(1) Pruning Strategy

This study employs L1-norm-based filter-level structured pruning to sparsify the convolutional layers of the VGG16 backbone network. It should be noted that the adopted pruning strategy does not physically remove convolutional channels. Instead, it utilizes the pruning utilities provided by PyTorch to zero out entire convolutional filters with relatively low contributions along the output-channel dimension. Therefore, in terms of the granularity of the pruning operation, this approach belongs to filter-level structured pruning. However, from the perspective of architectural modification, it can be categorized as soft pruning or mask-based pruning, since the network topology and the shapes of the weight tensors remain unchanged after pruning.

For a two-dimensional convolutional layer in the model, its weight tensor can be represented as Equation (2):

$$W \in R^{C_{out}\times C_{in}\times K_h\times K_W} \tag{2}$$

Here, C_{out} denotes the number of output channels, which corresponds to the number of convolutional filters, while C_{in} represents the number of input channels.Kw and Kh denote the

ISSN: 2664-9640

width and height of the convolutional kernel, respectively. For the i-th output filter, its weights are defined as Equation (3):

$$W \in R^{C_{in} \times K_h \times K_W} \tag{3}$$

In this study, the L1 norm metric proposed in [10] is employed to quantitatively assess the importance of each filter, and is defined as Equation (4):

$$S_i = |W_i|_1 = \sum_{c=1}^{C_{in}} \sum_{h=1}^{K_h} \sum_{w=1}^{K_w} \left|W_{i,c,h,w}\right| \tag{4}$$

Here, S_i denotes how important the i-th convolutional filter is. This metric represents the sum of absolute values of all weights within the i-th filter from a physical perspective. If a filter yields a small S_i, it generally produces weak activation responses on input feature maps and contributes marginally to subsequent feature extraction as well as the final classification prediction. Accordingly, we preferentially prune filters with small L1 norms in our work.

In this study, the pruning ratio was set to 50%. This ratio was chosen as a compromise between model compression strength and accuracy retention. A lower pruning ratio may fail to sufficiently reflect the redundancy in the model, whereas an excessively high pruning ratio may severely weaken the feature extraction capability, making it difficult for the model to recover its performance through subsequent fine-tuning. In the experiments, the aforementioned L1-norm filter-level pruning operation was applied to all convolutional layers in the VGG16-BN backbone network, such that approximately half of the output filters in each convolutional layer were set to zero, thereby enabling an analysis of the degree of filter redundancy in the model.

(2) Fine-tuning after Pruning

Zeroing out a large proportion of convolutional filters directly impairs the feature representation capabilities learned by the original model. As a result, the performance of the pruned model typically degrades noticeably. The experimental results also indicate that, without fine-tuning, the classification accuracy of the pruned model on the test set is lower than that of the original teacher model. This outcome is expected, since setting 50% of the convolutional filters to zero substantially weakens the representational capacity of many intermediate feature channels within the network.

To restore the classification performance of the pruned model, the pruned network was fine-tuned. During the fine-tuning phase, the cross-entropy loss function was used, and model parameters were updated using the AdamW optimizer. The learning rate was set to 0.1 of the base learning rate used in the original training, while the weight decay coefficient of the optimizer remained the same as in the original training phase. The fine-tuning process was conducted for a total of 40 epochs.

Since the fine-tuning is performed on the sparsified version of the original teacher model rather than training from random initialization, the pruned model still retains a large number of effective parameters learned during the original training. Therefore, a learning rate of 0.1 times the initial value was chosen for fine-tuning, which facilitates adaptive adjustment of the remaining effective parameters while avoiding severe disruption of the previously learned feature representations, thereby gradually compensating for the performance loss caused by zeroing out 50% of the filters.

During the fine-tuning process, the classification accuracy of the model on the validation set was monitored, and the model achieving the highest validation accuracy was saved as the final pruned model. After fine-tuning, the model weights corresponding to the best validation

ISSN: 2664-9640

performance were reloaded for subsequent evaluation on the test set and comparison with other models.

3.2.2. Knowledge Distillation

Following the completion of the original VGG16 backbone network training and structured pruning analysis, this study further introduces the knowledge distillation method to explore the subject of model compression and lightweighting. Compared with training a compact network using the original SAR image dataset, the core advantage of knowledge distillation lies in the abundance of supervisory information it provides. The distillation process not only requires the student model to fit the ground-truth class labels but also constrains its output distribution to remain consistent with that of the teacher model, enabling the student model to learn finer-grained knowledge encoded by the teacher model through its comprehensive training process.

Specifically, such knowledge is typically manifested in the inter-class similarity relationships, sample uncertainty reflected in prediction confidence scores, and more discriminative feature response distributions. Consequently, compared with training approaches that rely solely on hard labels, knowledge distillation provides supervision signals with higher information density to the lightweight student network, allowing it to achieve significantly reduced parameter scale and computational cost while simultaneously approaching the generalization performance and classification accuracy of the teacher model.

(1) Distillation Framework and Loss Function

Knowledge distillation typically consists of three key components: teacher network, student network, and distillation loss. The teacher network first acquires robust classification capability through a complete training procedure and maintains frozen parameters during the distillation stage, serving solely as a knowledge provider. In contrast, the student network undergoes training under the supervision of the teacher, with the objective of simultaneously fitting the ground-truth labels and the soft target distribution produced by the teacher network.

Let the output logits of the teacher network for input sample (x) be denoted as Zt, and the output logits of the student network be denoted as Z_s. Then, under the temperature coefficient (T), the soft probability distributions of the teacher and student networks are respectively defined as following Equation (5)

$$p_t^{(T)} = \text{softmax}\left(\frac{z_t}{T}\right), p_s^{(T)} = \text{softmax}\left(\frac{z_s}{T}\right) \tag{5}$$

Specifically, the temperature coefficient T > 1 serves to smooth the class distribution, allowing the relative probability information of non-target classes to be more fully revealed. Compared with standard softmax at temperature T = 1, the output distribution at higher temperatures can provide richer "dark knowledge" concerning inter-class similarity and decision boundaries, which is particularly crucial for enabling the student model to learn the discriminative patterns encoded within the teacher model.

The distillation loss employed in this study comprises two components: one is the distillation loss derived from the teacher network's soft output distribution, and the other is the supervised loss based on ground-truth labels. The overall objective function is defined as following Equation (6):

$$\mathcal{L}_{KD} = \alpha \mathcal{L}_{soft} + (1 - \alpha)\mathcal{L}_{hard} \tag{6}$$

ISSN: 2664-9640

Where α denotes the distillation loss weight, which serves to balance the relative importance of hard supervision and soft supervision. $\mathcal{L}_{\text{soft}}$ represents the distribution matching loss between the teacher and student networks, while $\mathcal{L}_{\text{hard}}$ represents the cross-entropy loss between the student output and the ground-truth labels. Specifically, this study employs a distillation term based on Kullback-Leibler (KL) divergence, whose expression is as following Equation (7) and (8):

$$\mathcal{L}_{\text{distill}} = \mathrm{T}^2 \cdot \mathrm{D}_{\mathrm{KL}}(\mathrm{p}_{\mathrm{T}} \parallel \mathrm{p}_{\mathrm{S}}) \tag{7}$$

$$\mathrm{D}_{\mathrm{KL}}(\mathrm{p}_{\mathrm{T}} \parallel \mathrm{p}_{\mathrm{S}}) = \sum_{\mathrm{i}} \mathrm{p}_{\mathrm{T}}(\mathrm{i}) \log \frac{\mathrm{p}_{\mathrm{T}}(\mathrm{i})}{\mathrm{p}_{\mathrm{S}}(\mathrm{i})} \tag{8}$$

T^2 is a common scaling factor in knowledge distillation that serves to compensate for the attenuation of gradient magnitude induced by the softening process, while simultaneously preserving the relative differences embedded within the dark knowledge.

Hard supervision adopts standard cross-entropy loss, its expression can be shown as the Equation (9) below.

$$\mathcal{L}_{hard} = \mathrm{CE}(\mathrm{y}, \mathrm{z}_{\mathrm{s}}) \tag{9}$$

where y denotes the ground-truth class label.

In the experiments presented in this study, the temperature coefficient is set to T = 4.0, and the distillation loss weight is set to α = 0.7. This hyperparameter configuration indicates that the distillation process places greater emphasis on the soft knowledge provided by the teacher model's output distribution, while simultaneously retaining a substantial proportion of supervision from ground-truth labels. This design ensures that the student model can achieve stable convergence during the early stages of training and mitigates the risk of over-fitting to the biases inherent in the teacher model.

(2) Student Model Structure and Training Strategy

This study constructs a compact convolutional network as the student model to receive discriminative knowledge transferred from the teacher model. Compared with the VGG16-BN teacher model, the student network significantly reduces network depth, convolutional channel width, and the parameter scale of fully connected layers, while maintaining fundamental convolutional feature extraction capability, thereby lowering both model storage requirements and inference computational load.

The input tensor channel number of the student model is consistent with that of the teacher model, both being 2, to accommodate the VV-VH dual-polarized SAR image classification task presented in this work. The feature extraction module is composed of five sequentially stacked convolutional blocks, each containing a two-dimensional convolution layer (Conv2d), batch normalization layer (BatchNorm), and a nonlinear activation function (ReLU). Following each convolutional block, a max-pooling layer is applied for spatial downsampling, progressively enlarging the receptive field and compressing the spatial dimensions of the feature maps. The evolution of network channel numbers is as follows:

$$2 \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 256$$

In particular, the first convolutional layer maps the 2-channel dual-polarization input to 32 feature channels, which are then progressively increased to 64, 128, and 256 channels, with the

ISSN: 2664-9640

final convolutional block maintaining 256 output channels. This design preserves multi-level feature abstraction capability at a relatively low parameter scale, enabling the student model to learn local scattering structures, ship contour textures, and the complementary features between the VV and VH polarization channels in SAR imagery.

After the feature extraction module, the student network employs an adaptive average pooling layer, AdaptiveAvgPool2d((2, 2)), to compress the output feature maps into a fixed spatial size. The pooled features are then flattened into a 256 × 2 × 2 vector and fed into the classification head for final discrimination. The classification head consists of a fully connected layer, a ReLU activation function, a Dropout regularization layer, and a final output layer, with the following structure:

$$\text{Linear}(256\times2\times2,256)\rightarrow\text{ReLU}\rightarrow\text{Dropout}(0.3)\rightarrow\text{Linear}(256,3)$$

The output dimension of the final layer is set to 3, corresponding to the three ship categories in the SAR image classification task considered in this study: Cargo, Tanker, and Fishing. A Dropout rate of 0.3 is adopted to mitigate the risk of overfitting in the student model when trained on a small-scale SAR dataset.

Regarding the training strategy, the distillation student model follows the same dataset split and preprocessing procedure as the teacher model, with the training, validation, and test sets partitioned in a 70% / 15% / 15% ratio, and employing the same VV/VH dual-channel input format. The distillation training is conducted for 100 epochs, using the AdamW optimizer with an initial learning rate of 3×10^{-4} and a weight decay coefficient of 10^{-4}. To improve convergence smoothness in the later stages of training, a cosine annealing learning rate scheduler is applied, gradually reducing the learning rate to 10^{-6}. During backpropagation, to ensure numerical stability in student model training, gradient clipping with a maximum norm of 2.0 is applied, mitigating the impact of abnormally large gradients on the model update process.

4. Experiment

4.1. Dataset

4.1.1. Data Loading

This study conducts ship target classification experiments on SAR imagery using the publicly available OpenSARShip_uint8 dataset. Three ship categories in the dataset-Cargo, Tanker, and Fishing-are selected as the classification targets. Each ship sample in the dataset provides SAR imagery acquired under both VV and VH polarization modes. To improve data processing efficiency, this study uses the VV images as the indexing reference and matches the corresponding VH images according to the filenames provided in the dataset. Furthermore, only valid samples containing both VV and VH polarization images are retained to ensure a standardized dataset format and facilitate subsequent model training.

4.1.2. Dataset Partition and Data Augmentation

In this experiment, the dataset is randomly split into training, validation, and test sets with proportions of 70%, 15%, and 15%, respectively, using a random seed of 42 to ensure the reproducibility of the experiment. The training set is used for learning model parameters, the validation set is employed for model selection and monitoring the training process, and the test set is reserved exclusively for the final performance evaluation.

To improve the robustness of the model against variations in ship pose, imaging discrepancies, and local occlusions, a data augmentation strategy was applied to the training set. The adopted augmentation techniques included random resized cropping (scaling range: 0.8–1.0; output size: 128 × 128), horizontal flipping (probability: 50%), vertical flipping (probability: 50%),

ISSN: 2664-9640

random rotation (rotation angle within ±180°), Gaussian blur (kernel size: 3, $\sigma \in [0.1, 1.0]$), and random erasing (probability: 30%, erasing area ratio ranging from 2% to 15%). No random augmentation was applied to the validation or test sets; instead, all images were resized to a uniform scale, followed by tensor conversion and normalization. Specifically, after conversion to tensors, all images were normalized using a mean of 0.5 and a standard deviation of 0.5, thereby mapping pixel values to the range of $[-1, 1]$.

Since images acquired under the VV and VH polarization modes essentially represent observations of the same target, special care was taken to preserve the spatial correspondence between the dual-polarization channels during augmentation. To avoid disrupting the alignment between the VV and VH images, identical random transformation parameters were applied to both polarization modalities for the same target, ensuring pixel-level alignment after stochastic operations such as cropping, rotation, and flipping.

In addition, considering the potential imbalance in sample numbers across different ship categories within the dataset, a weighted random sampling strategy based on the inverse of class frequencies was employed during the training phase. This approach assigns higher sampling probabilities to minority-class samples, thereby mitigating the impact of class imbalance on model training.

4.2. Model Training Result

4.2.1. Convergence Analysis During Training

Figure 2 illustrates the loss variations of the four categories of models during the training process, while Figure 3 presents the changes in training accuracy over epochs. From the overall trends, the training losses of all models gradually decreased as the number of training epochs increased, whereas the training accuracy exhibited an overall upward trend. These results indicate that the models were capable of learning effective discriminative features for ship target recognition from dual-polarization SAR images.

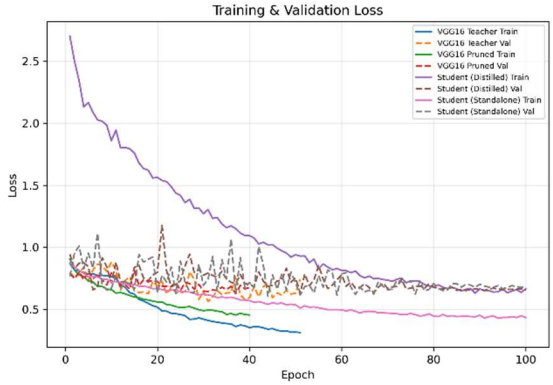


Figure 2. Loss variation of models with training epochs

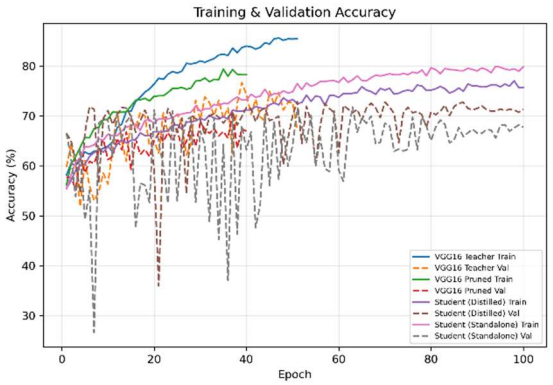


Figure 3. Accuracy of models with training epochs

ISSN: 2664-9640

The two training curves of the VGG16 Teacher (original model) indicate that the training loss decreases steadily while the training accuracy continuously improves throughout the training process, ultimately reaching a high level. This demonstrates that the teacher model, built upon the pretrained VGG16-BN architecture, possesses strong feature-learning and fitting capabilities. The validation accuracy remains generally above 70%, although certain fluctuations can be observed during training. These fluctuations can be attributed, on the one hand, to the inherent characteristics of SAR imagery, including speckle noise, variations in target pose, and similarities in scattering signatures among different classes. On the other hand, they are also related to the class imbalance present in the dataset. In particular, the Fishing class contains relatively few samples, resulting in a limited number of minority-class instances in the validation and test sets. Consequently, the validation accuracy becomes more sensitive to the prediction outcome of individual samples.

The training curves of the pruned model primarily reflect the fine-tuning process following pruning. Since a filter-zeroing operation with a pruning ratio of 50% was applied to all convolutional layers, a portion of the feature channels previously learned by the original teacher model was disrupted, leading to a noticeable degradation in model performance immediately after pruning. However, as fine-tuning progresses, both the training accuracy and validation accuracy gradually recover with increasing training epochs and ultimately remain only slightly lower than those of the original model. This indicates that the filters retained after pruning still preserve most of the model's effective feature representation capability. Furthermore, the fine-tuning process enables the model parameters to be re-optimized, thereby compensating for the performance loss caused by pruning.

For the student models, both the distilled student model and the non-distilled student model exhibit stable convergence during training. It should be noted that the training loss of the distilled student model consists of both the hard-label cross-entropy loss and the Kullback–Leibler (KL) divergence loss derived from the teacher's soft targets. Moreover, the loss value is scaled by the temperature factor T^2. Therefore, its numerical loss values cannot be directly compared with those of the non-distilled model, which is trained solely using the cross-entropy loss. Nevertheless, the overall loss trend and the final test accuracy can still provide valuable insights into the training process. The results show that, under the same network architecture and parameter budget, the distilled student model achieves a higher test accuracy than its non-distilled counterpart. In addition, the fluctuations in both the loss and accuracy curves of the distilled student model are substantially smaller throughout training. These findings suggest that the soft-target distribution produced by the teacher model provides the student model with additional information regarding inter-class relationships, thereby enhancing the generalization capability of the lightweight model.

4.2.2. Comprehensive Performance Comparison of Models

Table 2. Comprehensive performance comparison of different models

Model	Params/M	Size/MB	Test Accuracy/%	Inference Time/ms
VGG16 Teacher	18.92	72.17	74.08	1.93
VGG16 Pruned	18.92	72.17	70.12	1.73
Student Distilled	1.24	4.74	73.86	0.71
Student Standalone	1.24	4.74	71.80	0.71

As shown in the table 2 above, the VGG16 Teacher model achieves the highest test accuracy of 74.08%. This result demonstrates that the VGG16-BN network, based on ImageNet pretrained weights and further adapted through dual-channel input adaptation and two-stage fine-tuning, can be effectively transferred to the VV/VH dual-polarization SAR ship classification task. Benefiting from its deep convolutional architecture and large parameter capacity, VGG16-BN

ISSN: 2664-9640

possesses strong feature representation capability and consequently attains the highest overall classification accuracy among the four models.

As shown in Table 2, the VGG16 Teacher model achieves a moderate test accuracy of 74.08%. This limited absolute performance arises from the dataset's inherent scattering ambiguity, unbalanced sample distribution and optical-SAR domain shift, even though the dual-channel adaptation and two-stage fine-tuning fully exploit the transferable feature capacity of ImageNet pre-trained VGG16-BN. Among all compared models, the uncompressed teacher obtains the best discriminative power but suffers prohibitive overhead for edge deployment.

However, the teacher model contains 18.92M parameters, occupies 72.17 MB of storage, and requires 1.93 ms to infer a single image. Although this model achieves high accuracy, its large scale imposes considerable demands on storage and computational resources, making it unsuitable for deployment on edge computing devices, embedded platforms, or low-power maritime monitoring terminals. Therefore, relying solely on the VGG16 Teacher can hardly meet the requirements of lightweight design and real-time performance.

The VGG16 Pruned model is derived from the teacher model by applying L1-norm-based filter-level soft pruning at a 50% pruning ratio to all convolutional layers, followed by fine-tuning for 40 epochs. Experimental results show that the pruned and fine-tuned model attains a test accuracy of 70.12%, which is only 3.96% lower than the teacher model's 74.08%. This soft pruning experiment demonstrates that, after zeroing out 50% of the convolutional filters and performing fine-tuning, the model still retains strong representational capacity. Such behavior indicates that VGG16-BN possesses a certain degree of filter redundancy in this task and holds potential for further compression and sparsification. Because mask-based soft pruning was adopted in this work, the shapes of the model's parameter tensors remain unchanged after pruning; consequently, the number of parameters and the storage size are not reduced. This experiment is primarily intended to verify the filter redundancy of VGG16-BN in the SAR ship classification task and to provide a basis for the subsequent design of a lightweight student model.

4.2.3. Performance Analysis of Distilled Student Network

Compared with the soft-pruned model, the student model trained via knowledge distillation reduces parameter quantity and computational complexity from the perspective of network architecture design, thereby achieving superior lightweight performance. The distilled student model only has 1.24 M parameters with a model size of 4.74 MB, and the inference latency for a single image is 0.71 ms. Compared with the VGG16 Teacher model, its parameter reduction ratio is calculated as Equation (10) below:

$$\frac{18.92-1.24}{18.97} \times 100\% = 93.45\% \tag{10}$$

The reduction percentage of model size can be calculated by Equation (11):

$$\frac{72.17-4.74}{72.17} \times 100\% = 93.43\% \tag{11}$$

The improvement multiple of inference speed is calculated as Equation (12):

$$\frac{1.93}{0.71} \approx 2.72 \tag{12}$$

ISSN: 2664-9640

That is, when the number of parameters and model size of the distilled student model are both reduced by approximately 93%, its inference speed is increased by about 2.72 times. Meanwhile, its test accuracy reaches 73.86%. This result demonstrates that knowledge distillation can effectively transfer the discriminative knowledge learned by the teacher model to a lightweight student model, enabling the student model to maintain high classification accuracy while significantly reducing model scale and inference time.

To verify the effectiveness and necessity of knowledge distillation, this study also trains a Student Standalone model that shares the exact same architecture as the distilled student model but is trained without distillation supervision. This model also has 1.24M parameters, a model size of 4.74MB, and an inference time of 0.71ms, yet its test accuracy is 71.80%. Compared with the non-distilled student model, the accuracy improvement of the distilled student model is calculated by Equation (13):

$$73.86\% - 71.80\% = 2.06\% \tag{13}$$

Since the two student models share identical architectures, with the same number of parameters, model size, and inference time, the difference in their accuracy mainly arises from the distinct training supervision methods. The distilled student model learns not only the ground-truth labels but also the soft probability distribution output by the teacher model. These soft labels contain information such as inter-class similarities, model prediction confidence, and sample uncertainty, thereby providing the student model with richer supervision signals than hard labels. Consequently, knowledge distillation can improve the generalization performance of lightweight models without increasing inference overhead. Overall, the Student Distilled model achieves a favorable balance among accuracy, number of parameters, model size, and inference speed.

4.3. Ablation Study

4.3.1. Effectiveness of Dual-Polarization VV-VH Feature Fusion

(1) Quantitative Accuracy Comparison

To verify the effectiveness of dual-polarization two-channel input, three comparative experiments are conducted with identical network architectures, training hyperparameters and datasets. The best validation accuracy (Best Val Acc) and final validation accuracy at the end of training (Final Val Acc) are summarized in Table 3.

Table 3. Validation accuracy under different polarization input schemes

Input Scheme	Best Val ACC / %	Final Val ACC / %
VV Only	72.48	70.20
VH Only	72.64	71.01
VV-VH Dual-Pol	76.38	71.88

As illustrated in Table 1, the proposed VV-VH dual-polarization scheme attains a peak validation accuracy of 76.38%, exceeding the VV-only and VH-only single-polarization baselines by 3.90% and 3.74% respectively; the two single-polarization setups deliver similar performance ceilings, with VH-only marginally outperforming VV-only by merely 0.16%.

In terms of SAR scattering physics, VV polarization mainly captures global ship contours via hull-sea single-bounce scattering but suffers from speckle noise and sea clutter interference, while VH cross-polarization encodes delicate local textural information from superstructure multi-bounce scattering; duplicating a single polarization across dual input channels confines the model to one-dimensional scattering feature learning with obvious performance limitations,

ISSN: 2664-9640

whereas two-channel VV-VH fusion leverages their intrinsic scattering complementarity to jointly extract global and fine-grained ship features, mitigate inter-class scattering confusion and raise the overall classification upper bound.

(2) Analysis of Training Convergence and Generalization

Comparisons of training loss and accuracy curves shown in Figure 4 reveal distinct convergence characteristics among the three models: the VV-Only and VH-Only single-polarization baselines show much more violent fluctuations in validation accuracy due to their single-dimensional feature representations, which make them susceptible to hard samples and noise and cause unstable predictions, whereas complementary dual-polarization features supply extra discriminative information and strengthen the model's anti-noise ability; all models achieve similar accuracy at the initial training stage, yet the validation accuracy of single-polarization models rapidly saturates in later epochs, while the dual-polarization model keeps continuous accuracy growth owing to its powerful feature extraction capability; although slight overfitting exists in all groups, the dual-polarization input narrows the accuracy gap between training and validation sets, as sufficient complementary polarization features prevent the network from memorizing noise in training data and improve generalization performance.

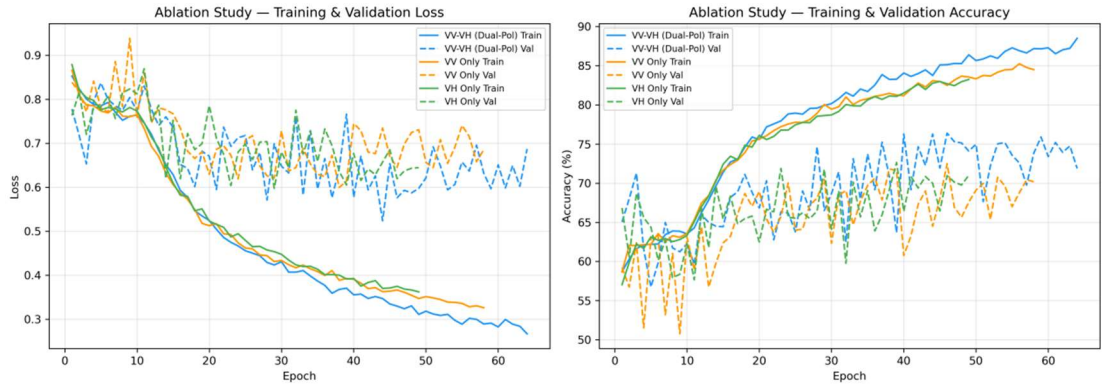


Figure 4. Comparisons of training loss and accuracy curves

(3) conclusion

This ablation study verifies the necessity and superiority of the VV-VH dual-polarization two-channel input strategy. Relying merely on duplicated single VV or VH polarization leads to limited classification accuracy, severe validation oscillation and inferior generalization due to insufficient scattering information. By contrast, the fusion of complementary VV and VH polarization signatures significantly boosts peak classification accuracy, stabilizes training convergence and alleviates overfitting, which serves as a fundamental guarantee for the outstanding performance of the proposed SAR ship classification model.

4.3.2. Comparison of Training Paradigms: Two-Stage vs. End-to-End

This ablation study investigates the effectiveness of the proposed two-stage progressive fine-tuning strategy, by comparing end-to-end full-network training with a staged training pipeline: we first freeze all convolutional layers of pretrained VGG16 and only optimize the classification head for 10 epochs, then unfreeze the entire backbone for joint fine-tuning with layer-wise learning rates. In the first stage, fixed ImageNet pre-trained low-level texture features are preserved to avoid drastic gradient disturbance at early training epochs, while only the task-specific classifier adapted to dual-polarization SAR inputs is optimized. After initial convergence of the classification branch, all feature extraction layers are unlocked and fine-tuned together with a cosine annealing learning scheduler. The training loss and accuracy curves of the two training paradigms are visualized in Figure 5, and the quantitative accuracy

ISSN: 2664-9640

comparison under best validation checkpoint, standard test set and TTA-augmented test set is presented in Figure 6.

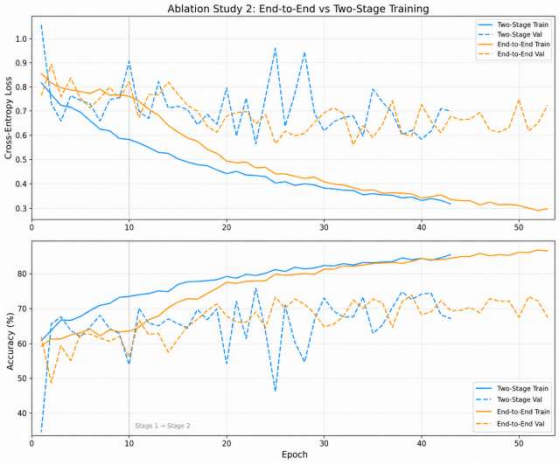


Figure 5. Training loss and accuracy curves: two-stage training versus end-to-end training

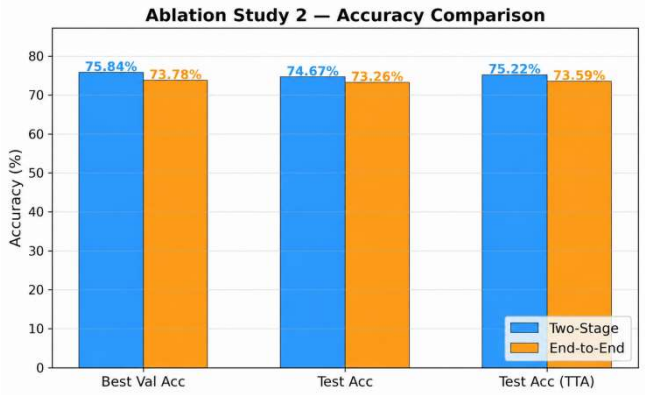


Figure 6. Accuracy comparison between two-stage training and end-to-end training

The training curves shown in Figure 5 demonstrate that the two-stage scheme achieves faster convergence and less volatile validation loss, and obtains higher final test accuracy than the end-to-end baseline without layer freezing. The single-stage full fine-tuning suffers from catastrophic forgetting of general visual priors caused by the domain gap between natural images and SAR ship imagery, accompanied by severe overfitting risks. In contrast, staged freezing first stabilizes category decision boundaries and then adaptively refines feature representations specialized for SAR vessel data, which balances inherited generic visual knowledge and task-aware remote sensing feature learning. The results verify that the two-stage progressive fine-tuning pipeline delivers indispensable performance improvements for the dual-polarization SAR ship classification task.

4.3.3. Ablating the Knowledge Distillation Branch

Ablation experiments are conducted to evaluate the effectiveness of the knowledge distillation (KD) module. We compare a standalone student model with the KD-based student trained using teacher soft labels, and the corresponding training curves are illustrated in Figure 6. Although the standalone model achieves higher training accuracy (79.4% vs. 76.1%), it exhibits a larger generalization gap (11.7%) and lower validation/test accuracy (67.7% and 70.28%). In contrast, the KD-based model achieves improved validation and test performance (+3.5% and +2.77%) with a significantly reduced generalization gap (4.9%), together with more stable convergence behavior in the validation loss. These improvements may be attributed to the KL-

ISSN: 2664-9640

divergence constraint, which introduces soft-label supervision and encodes inter-class similarity information, thereby acting as an implicit regularizer that mitigates overfitting to training noise.

It should be noted that due to the different definitions of loss functions between the distillation model (with soft supervision from the teacher model) and the standalone model, a direct numerical comparison of the loss functions of the two models is meaningless, but the oscillation amplitude and general variation trend of the loss functions still have reference and analytical value.

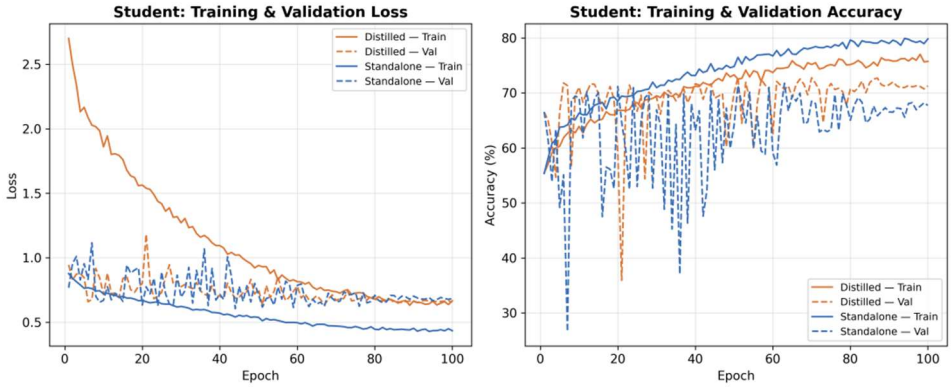


Figure 7. Training and validation curves of standalone student and distilled student models

4.3.4. Sensitivity Analysis of Distillation Hyperparameters

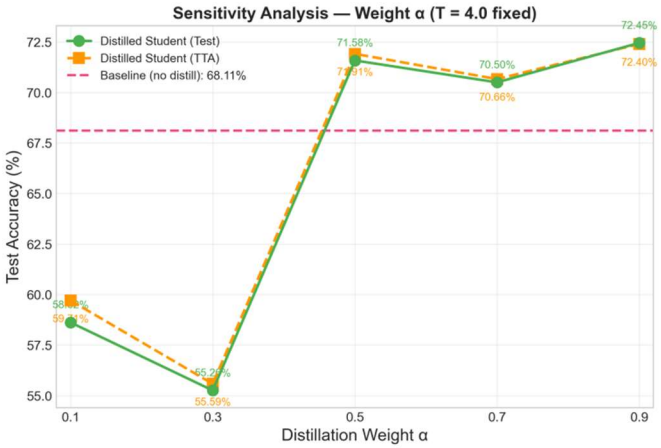


Figure 8. Sensitivity analysis of loss weight α(fixed temperature T = 4.0)

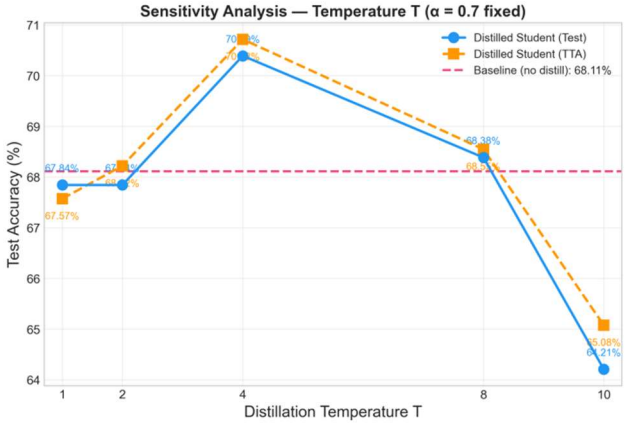


Figure 9. Sensitivity analysis of distillation temperature T (fixed weight α=0.7)

ISSN: 2664-9640

We conduct ablation studies to analyze the sensitivity of knowledge distillation to the loss weight coefficient α and temperature T, with quantitative results visualized in Figure 8 and Figure 9, respectively. When fixing T=4.0 and sweeping α from 0.1 to 0.9, the model yields anomalously low test and TTA accuracy of 55.26% and 55.59% at α=0.3, which is far below the non-distilled baseline of 68.11%. The performance at α=0.1 also underperforms the baseline, while all settings with α≥0.5 bring substantial improvements over the baseline, and the optimal test accuracy of 72.45% (TTA 72.40%) is achieved at α=0.9. For the second group with fixed α=0.7 and varying T, low temperatures (T=1,2) fail to transfer inter-class implicit knowledge from soft labels, leading to slightly inferior accuracy compared with the baseline. The best distillation performance appears at T=4, with test accuracy of 70.32% and TTA accuracy of 70.39%. Further increasing T to 8 and 10 over-flattens the teacher probability distribution and dilutes discriminative supervision signals, resulting in gradual degradation; the test accuracy drops sharply to 64.21% at T=10. These controlled experiments demonstrate that distillation performance is highly sensitive to both hyperparameters: insufficient soft-loss weight or extreme temperature severely hinders knowledge transfer.

5. Conclusion and Prospect

This paper targets edge-deployment bottlenecks of SAR ship classification, yet we explicitly acknowledge a core limitation of the proposed framework: the VGG16-BN teacher only reaches a moderate 74.08% test accuracy, restricted by dataset noise, inter-class similarity and cross-domain pre-training gaps, which in turn caps the distilled student's accuracy at 73.86%. To tackle model bloat and slow inference while retaining nearly all discriminative knowledge from the limited baseline, this study proposes a lightweight classification framework integrating VV/VH dual-polarization input, two-stage transfer learning, soft pruning ablation and knowledge distillation.

Experimental results indicate that the fine-tuned VGG16 teacher model achieves a moderate test accuracy of 74.08%, which verifies the feasibility of transfer learning for dual-polarization SAR ship classification yet exposes an obvious accuracy bottleneck limited by speckle noise, inter-class scattering confusion and unbalanced samples. Soft pruning experiments based on the L1 norm revealed that the VGG16-BN model contains a certain degree of parameter redundancy; even with a 50% pruning ratio, the pruned model still retains 70.12% accuracy, which validates the space for model compression. Furthermore, the knowledge-distilled student model, with only 1.24M parameters and a model size of 4.74 MB, attained a test accuracy of 73.86%. Compared with the teacher model, the student model reduces parameter count and storage requirements by approximately 93% and increased inference speed by about 2.72 times, while incurring with only a 0.22% reduction in accuracy. These results indicate that knowledge distillation can substantially enhance model lightweighting without significant loss of classification performance, making it more suitable for deployment in resource-constrained edge computing and maritime intelligent monitoring scenarios.

It is critical to clarify the applicable boundary of this lightweight scheme. For shore-based centralized maritime platforms with sufficient computing resources, boosting the teacher backbone's absolute classification accuracy is the primary target while model compression serves as minor refinement. For satellite/UAV on-board edge terminals restricted by memory, power and latency, lightweight knowledge distillation is mandatory for real-time ship identification even with the 74% baseline accuracy ceiling. Targeting edge deployment scenarios, this paper verifies that knowledge distillation achieves compact models with negligible accuracy loss, and the distilled student fits resource-limited real-time maritime monitoring systems well.

ISSN: 2664-9640

Despite promising lightweight performance, the proposed framework suffers from several constraints, with the moderate teacher baseline accuracy as its dominant defect. Limited by severe speckle noise, inter-class scattering similarity and dataset imbalance, the VGG16-BN teacher only reaches 74.08% accuracy, and all lightweight optimizations cannot fundamentally lift this fixed performance ceiling. Besides, experiments are limited to three ship classes within OpenSARShip_uint8, leaving the model's generalization across multi-category datasets and complex marine conditions unvalidated. The shallow VV/VH fusion module fails to extract deep complementary polarization features, and the adopted soft pruning merely masks redundant filters without physical channel removal, failing to deliver hardware-efficient structured compression. Moreover, the lightweight student network still retains redundant layers and channels for further structural optimization.

Future research prioritizes baseline accuracy improvement to address the core limitation: we will adopt SAR-specialized backbones, polarization-guided attention fusion and expanded multi-condition datasets to break the ~74% accuracy bottleneck. Subsequent work includes designing deep dual-polarization fusion modules, exploring hard structured channel pruning for genuine hardware compression, and combining native lightweight networks with advanced feature and attention distillation to balance inference speed and classification precision. Such improvements enable SAR ship classification models to simultaneously support high-precision maritime survey and low-power edge real-time monitoring for practical engineering systems.

References

[1] Xiao, M., Wang, H., Zeng, Z., et al. (2021). SAR image segmentation and target detection based on mathematical morphology. In 2021 2nd China International SAR Symposium (pp. 1–5). Shanghai, China.

[2] Li, Y., Liu, J., Li, X., et al. (2023). A dual-polarization feature enhancement network for SAR ship detection. In 2023 International Conference on Machine Vision, Image Processing and Imaging Technology (pp. 38–42). Hangzhou, China.

[3] Zhang, Y., Chen, L., Wang, S., Zhao, Y., & Wei, J. (2025). PGAF-Net: An adaptive fusion network with polarization-guided hybrid attention for dual-polarized SAR ship classification. Multimedia Systems, 31, Article 399.

[4] Guo, R., Zhang, Z., Gao, R., et al. (2021). Ocean target investigation using spaceborne SAR under dual-polarization strip-map mode. In 2021 2nd China International SAR Symposium (pp. 1–5). Shanghai, China.

[5] Yang, T., Liu, Z.-G., Wen, Z.-D., et al. (2021). A new ship detection method with limited labeled data in SAR imagery. In Proceedings of the IEEE Radar Conference (pp. 2123–2127).

[6] Zhi, Y., Wang, Y., Zhang, F., et al. (2024). A multi-branch network for SAR ship detection. In CISS 2024 - 5th China International SAR Symposium. Xi'an, China.

[7] Hou, B., Chen, X., & Jiao, L. (2015). Multilayer CFAR detection of ship targets in very high resolution SAR images. IEEE Geoscience and Remote Sensing Letters, 12(4), 811–815.

[8] Li, J., Qu, C., & Shao, J. (2017). Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of 2017 SAR in Big Data Era: Models, Methods and Applications, BIGSARDATA 2017 (pp. 1–6).

[9] Shao, R., Liu, Y., Zhang, W., & Wang, J. (2022). A survey of knowledge distillation in deep learning. Chinese Journal of Computers, 45(8), 1638–1673.

[10] Li, H., Kadav, A., Durdanovic, I., Samet, H., & Graf, H. P. (2016). Pruning filters for efficient convnets.

[11] Letaifa, L. B., Rouas, J. L. (2022). Fine-grained analysis of the transformer model for efficient pruning. In 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 897–902).

ISSN: 2664-9640

[12] Rui, L., Yang, S., Chen, S., Yang, Y., & Gao, Z. (2022). Smart network maintenance in an edge cloud computing environment: An adaptive model compression algorithm based on model pruning and model clustering. IEEE Transactions on Network and Service Management, 19, 4165–4175.

[13] Wu, M., Ma, W., Li, Y., & Zhao, X. (2020). The optimization method of knowledge distillation based on model pruning. In 2020 Chinese Automation Congress (CAC) (pp. 1386–1390). Shanghai, China.

[14] Rise, B., Uney, M., & Huang, X. (2025). Two-stage transfer learning for airborne multi-spectral image classifiers. Signal Processing, Article 110358.