Dual Information Purification for Lightweight SAR Object Detection

Xi Yang, Jiachen Sun, Songsong Duan, De Cheng*

State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China
{yangx, dcheng}@xidian.edu.cn, {sunjc, ssduan}@stu.xidian.edu.cn

Abstract

Synthetic aperture radar (SAR) object detection requires accurate identification and localization of targets at various scales within SAR images. However, background clutter and speckle noise can obscure key features and mislead the knowledge distillation process. To address these challenges, we introduce the Dual Information Purification Knowledge Distillation (DIPKD) method, which improves the performance of the student model through three key strategies: denoising, enrichment, and decoupling. First, our Selective Noise Suppression (SNS) technique reduces speckle noise in global features by minimizing misleading information from the teacher model. Second, the Knowledge Level Decoupling (KLD) module separates features into target and non-target knowledge, balancing feature mapping and reducing background noise to enhance the extraction of critical information for the student model. Finally, the Reverse Information Transfer (RIT) module refines intermediate features in the student model, compensating for the loss of detailed local information. Experimental results demonstrate that DIPKD significantly outperforms existing distillation techniques in SAR object detection, achieving 60.2% and 51.4% mAP scores on the SSDD and HRSID datasets, respectively. Additionally, the student model shows performance improvements of 1.3% and 2.9% over the teacher model, highlighting the effectiveness of the information purification approach.

Introduction

Synthetic aperture radar (SAR) imagery is characterized by its all-day, all-weather capabilities, making it a crucial tool for earth observation. SAR object detection aims to identify and locate specific targets, such as ships or aircraft, across vast SAR images, playing a vital role in monitoring and safeguarding marine and air domains. Although large-scale deep learning models perform well in SAR object detection, they still fall short of meeting real-time processing requirements. Therefore, balance between accuracy and efficiency has become a crucial and significant research for the lightweight object detection. Knowledge Distillation (KD) (Hinton, Vinyals, and Dean 2015) is one of the effective methods to address this issue.

*Corresponding Author

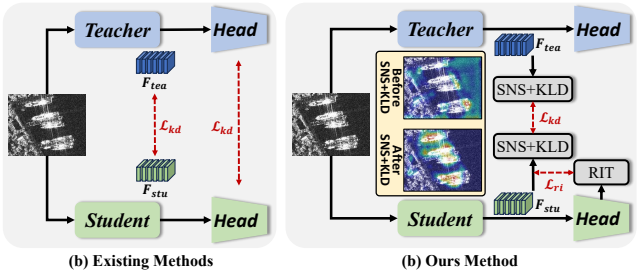


Figure 1: Comparison of existing methods with ours: (a) Existing methods often overlook noise management, leading to the unfiltered distillation of misleading information from teacher models to student models. (b) Our approach refines knowledge with precision and supplements intermediate features with detailed output information.

As illustraged in Figure 1(a), most existing KD methods typically guide the student model to learn the predictions (Wang et al. 2024) (Chen et al. 2021), intermediate features (Zhixing et al. 2021) (Li et al. 2023) (He et al. 2024) (Ni et al. 2023), and attention maps (Zagoruyko and Komodakis 2022) of teacher model via knowledge transfer. In order to overcome the issue of misleading noisy information, various decoupling strategies have been proposed (Guo et al. 2021) (Yang et al. 2022c). These strategies distinguish context-related foregrounds from context-unrelated backgrounds by generating imitation masks. However, these approaches have three limitations. First, unlike natural images, SAR images contain unique speckle noise (Singh and Shree 2016), which introduces a significant amount of misleading information into teacher models. Second, the extreme background ratio in remote sensing images makes it challenging for the student model to learn effective knowledge. Finally, the limited semantic detail in SAR images hinders the student model's ability to extract effective local information, making it challenging to achieve accurate detection results by relying solely on intermediate features.

To address these problems, this paper introduces the Dual Information Purification Knowledge Distillation (DIPKD) method, illustrated in Figure 1(b). The strategy focuses on three key elements: (1) Mitigating speckle noise in global features through selective wavelet domain noise re-

duction, thereby minimizing misleading information in the teacher's knowledge; (2) Addressing the imbalance between target-level and non-target-level features by employing a knowledge-level decoupling strategy, which enhances the interaction of local features between the teacher and student models; (3) Supplementing the student model's intermediate features with detailed local information through reverse information transfer.

Firstly, the Selective Noise Suppression (SNS) module employs spatial and channel attention-guided wavelet technology to reduce noise in key feature areas. This approach enhances feature representation while preserving the original information, enabling the model to effectively filter out speckle noise interference. Secondly, the Knowledge Level Decoupling (KLD) module separates transmission features, reducing conflicts between target and non-target features during the learning process. This decoupling minimizes background noise interference, allowing the student model to better learn and retain critical information. Finally, the Reverse Information Transfer (RIT) branch uses the predicted scores of student model to constrain intermediate features, subtly reducing the dominance of the teacher model. This approach encourages the student model to explore and develop its potential more fully. The DIPKD method enhances the purity of knowledge transmission by selectively de-noising and decoupling information while also supplementing features in detail. This approach strengthens the effectiveness of feature transfer, ensuring more refined and accurate knowledge transmission, which enhances the clarity of transmitted information and significantly improves the performance of knowledge distillation in SAR object detection tasks.

The main contributions of this work can be summarized as follows:

- We propose the DIPKD method to address the issue of noise in SAR images compromising knowledge purity during distillation. Specially, Selective Noise Suppression (SNS) module effectively reduces speckle noise interference in the global features.
- We propose the Knowledge Level Decoupling (KLD) and Reverse Information Transfer (RIT) modules, which effectively address the imbalance in knowledge levels and the lack of detailed local feature information in the student model.
- Extensive experiments demonstrate that our approach achieves state-of-the-art results in both single- and two-stage detector tests, significantly enhancing the student model's performance and even surpassing the teacher model on certain metrics.

Related Works

Object Detection in Remote Sensing Images

Recent advancements in object detection have led to significant improvements in accuracy. Mainstream deep learning-based algorithms can be divided into single-stage (Redmon et al. 2016) (Redmon and Farhadi 2017) (Tian et al. 2020) and two-stage detectors (Cai and Vasconcelos 2018) (He et al. 2017) (Xie et al. 2021). Remote sensing images, with their complex backgrounds, multi-scale targets, varying aspect ratios, and inconsistent orientations (Wan et al. 2023), pose unique challenges, especially for lightweight detectors. Additionally, speckle noise in SAR images complicates feature extraction, disrupting key information transmission. To overcome these challenges, researchers have developed specialized algorithms, often building on existing architectures like Faster-RCNN, SSD (Liu et al. 2016), and RetinaNet. Notable advancements include Li et al.'s rotation-sensitive RPN using polygonal anchors (Li et al. 2017), Wang et al.'s Adaptive Spatial Feature Pyramid for small targets (Wang et al. 2019a), and Xu et al.'s feature alignment detection method to address scale variations (Xu et al. 2021).

Knowledge Distillation

Knowledge distillation, introduced by Hinton (Hinton, Vinyals, and Dean 2015), is a model compression technique that enhances student model performance by transferring knowledge from a teacher model. Existing methods are categorized into logit-based, feature-based (Heo et al. 2019) (Romero et al. 2014), and relation-based approaches (Liu et al. 2019) (Park et al. 2019) (Tung and Mori 2019). However, its effectiveness diminishes in complex tasks like object detection. Chen et al. first applied KD to object detection by having the student imitate the teacher's features and regression results (Chen et al. 2017). Li et al. addressed foreground-background imbalance by distinguishing positive and negative regions using RPN features (Li, Jin, and Yan 2017). Wang et al. simulated fine-grained features through cross-location differences (Wang et al. 2019b), and Zhang et al. used attention-guided feature extraction with simulated masks (Zhang and Ma 2020a). Yang et al. introduced a KD method for remote sensing image detection (Yang et al. 2022b), but it neglected the relationship between different instances. Our method, instead, employs an information fidelity distillation strategy that enhances key knowledge transfer to the student while focusing on refining feature and minimizing noise interference.

Our Method

Problem Definition

In the object detection task, the model processes a set of input images and labels information $\{\mathbf{X}, \mathbf{Y}\}$ during training, where labels contain category and location data. The model divides the images into regions and performs classification and regression tasks within those regions. Specifically, the images are fed into the model to generate category probabilities and location information, which are then compared with the label values to calculate the loss:

$$L_{gt} = L(f_m(x), y), X \in \mathbf{X}, Y \in \mathbf{Y}, \tag{1}$$

where x and y denote the image and corresponding label respectively. $f_m(\cdot)$ denotes the the process of extracting features from the model backbone or neck. $L(\cdot)$ denotes cross entropy loss function or smooth L_1 loss function.

Knowledge distillation involves a pre-trained teacher model and a student model to be trained. The student's loss

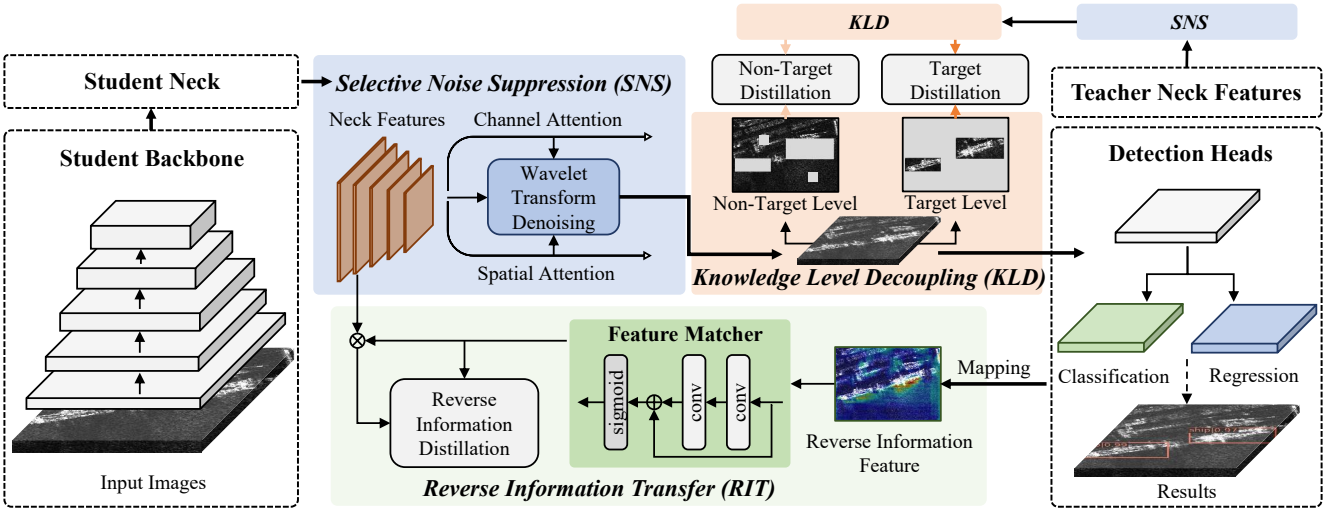


Figure 2: Overview of our DIPKD framework. The teacher and student model features undergo selective wavelet denoising and feature knowledge decoupling through the SNS and KLD modules, allowing the student model to focus on the most critical information from the teacher model. The RIT further mitigates the student model's over-reliance on the teacher model.

function typically has two parts: the loss between ground truth labels and the student's outputs, and the loss between the features or predictions of the teacher and student models. Since feature distillation offers better constraints than prediction distillation, we use feature distillation as our base method (Romero et al. 2014). Most detectors leverage multi-scale semantic information through Feature Pyramid Networks (FPN) (Lin et al. 2017a), which integrate semantic information from different levels of the backbone network for final predictions. The features produced by FPN play a decisive role in prediction outcomes. Generally, feature distillation is formulated as:

$$L_{fea} = \frac{1}{CHW} \sum_{k=1}^{C} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(F^T_{k,i,j} - f_{adp}(F^S_{k,i,j}) \right)^2, \quad (2)$$

where F^T and F^S denote the FPN features from the teacher and student, respectively, and $f_{adp}(\cdot)$ is the adaptation layer to reshape the F^S to the same dimension as F^T. H, W denote the height and width of the feature and C is the number of channels.

Overall Framework

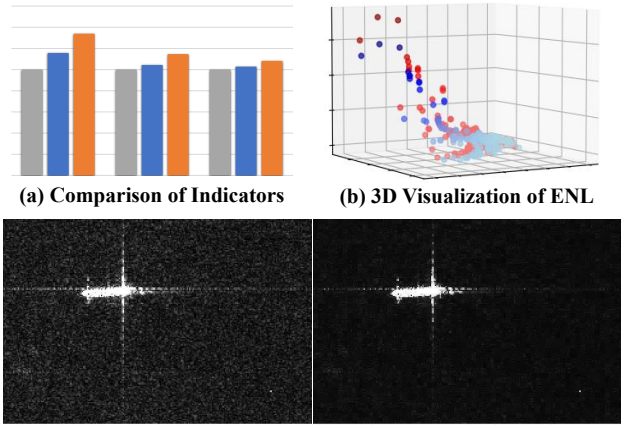
The inherent noise in SAR images compromises the purity of knowledge in teacher models, diminishing the effectiveness of traditional knowledge distillation techniques. To address this, we propose the Dual Information Purification Knowledge Distillation (DIPKD) method. As illustrated in Figure 2, our approach involves three key steps: the Selective Noise Suppression (SNS) module minimizes feature noise, the Knowledge Level Decoupling (KLD) module separates and refines knowledge, and the Reverse Information Transfer (RIT) module enriches the student's focal information. This process significantly improves the purity and efficacy of the transferred knowledge.

We applied our method to both the two-stage detector Faster-RCNN (Ren et al. 2015) and the single-stage detector RetinaNet (Lin et al. 2017b). Faster-RCNN uses a Region Proposal Network (RPN) to generate candidate regions (ROIs), followed by classification and bounding box regression. In contrast, RetinaNet generates numerous anchors across each feature map scale, directly performing object classification and bounding box adjustment through its classification and regression subnetworks. Our KD method is applied to the features outputted by the FPN layer.

Selective Noise Suppression

Speckle noise, a common issue in SAR images, significantly complicates effective knowledge distillation. To address this, we introduce the Selective Noise Suppression (SNS) module, which targets and denoises critical local features and channels. The SNS module employs feature attention to guide wavelet transform-based denoising. As illustrated in Figure 3(a), the SNS module's impact on feature processing is evident, with its effectiveness in suppressing speckle noise demonstrated through comparisons of Equivalent Number of Looks (ENL), Speckle Suppression Index (SSI), and Speckle Suppression and Mean Preservation Index (SMPI) against the fast Fourier transform method. The results clearly show the superiority of the SNS module. Figure 3(b) provides a 3D comparison of features before and after applying the SNS module, highlighting the enhancements in feature quality. Additionally, Figure 3(c) illustrates that directly applying wavelet transform denoising to SAR images can also yield improved results.

Previous works, like SENet (Hu, Shen, and Sun 2018) and CBAM (Woo et al. 2018), have shown that emphasizing key pixels and channels can significantly enhance the performance of CNN-based models. In a similar vein, AT demonstrated that a straightforward method for generating



(a) Comparison of Indicators (b) 3D Visualization of ENL

(c) Comparison of Image denoising

Figure 3: The speckle noise suppression effect of the SNS module: (a) Comparison of denoising performance among the original feature (gray), FFT denoising (blue), and SNS denoising (orange), presenting three evaluation metrics (from left to right: ENL, SSI^{-1}, and SMPI^{-1}). (b) 3D visual comparison of the original feature (blue) and the SNS-denoised feature (red) using the ENL index (z-axis). (c) Application of SNS module's wavelet transform denoising on SAR image.

a spatial attention mask can improve distillation outcomes (Zagoruyko and Komodakis 2022). Based on these insights, our approach employs a similar attention generation approach to precisely identify and focus on basic pixels and channels, and then generates corresponding attention maps to guide the denoising process.

We calculate the absolute mean on different pixels and different channels:

$$V_{i,j}^{S}(F) = \frac{1}{C} \cdot \sum_{k=1}^{C} |F_{k,i,j}|, \tag{3}$$

$$V_{k}^{C}(F) = \frac{1}{H \cdot W} \cdot \sum_{i=1}^{H} \sum_{j=1}^{W} |F_{k,i,j}|, \tag{4}$$

where H, W, C denote the feature's height, width, and channel. $V_{i,j}^{S}$ and V_{k}^{C} are the spatial and channel attention values. $|\cdot|$ denotes absolute value calculation. Then the attention map can be formulated as:

$$A_{i,j}^{S}(F) = H \cdot W \cdot \mathrm{softmax}\big(V_{i,j}^{S}(F)/\tau\big), \tag{5}$$

$$A_{k}^{C}(F) = C \cdot \mathrm{softmax}\big(V_{k}^{C}(F)/\tau\big), \tag{6}$$

where τ is the temperature hyperparameter proposed by Hinton et al. to regulate the distribution.

Then, using the extracted attention map to guide the feature denoising by wavelet transform, the features after denoising are as follows:

$$F' = f_{wld}\big(f_{dim}(A_{k}^{C}) \cdot f_{dim}(A_{i,j}^{S}) \cdot F_{k,i,j}\big), \tag{7}$$

where F denote input features, A_{k}^{C}, $A_{i,j}^{S}$ are the channel and spatial attention map corresponding to the input features. $f_{wld}(\cdot)$ is the wavelet transform denoising function. (Choi and Jeong 2019) $f_{dim}(\cdot)$ is the dimension matching function which unifies the dimensions of the input tensor with the dimensions of F.

Knowledge Level Decoupling

The DeFeat method (Guo et al. 2021) has demonstrated that non-target regions within a feature can offer valuable supplementary information to target regions. However, the high background-to-target ratio in remote sensing images can negatively impact target-level distillation. To address this challenge, we propose the Knowledge Level Decoupling (KLD) module. KLD effectively decouples features into target level and non-target level categories, allowing for separate distillation of each. To further mitigate noise from complex backgrounds, a random binary mask is applied to the non-target-level features. This approach ensures more precise knowledge transfer, enhancing the overall effectiveness of the distillation process.

First, we set a binary mask M to separate the background from the foreground:

$$M_{i,j} = \begin{cases} 1, & \text{if}(i,j) \in R \\ n, & \text{Otherwise} \end{cases}, \tag{8}$$

where R denotes the ground-truth boxes, n denotes elements in the set $\{0, 1\}$. i and j are the horizontal and vertical coordinates of the feature map, respectively. If (i, j) falls within the ground truth, then $M_{i,j} = 1$, otherwise, it is 0 or 1.

Larger-scale targets will occupy more loss because they cover more pixels, which can negatively impact the distillation of smaller targets. Additionally, the ratios of foreground to background vary significantly across different SAR remote sensing images. Therefore, to ensure equal treatment of different targets and to balance the loss between foreground and background, we set a scale mask S as follows:

$$S_{i,j} = \begin{cases} \frac{1}{H_r \cdot W_r}, & \text{if}(i,j) \in R \\ \frac{1}{N_{bg}}, & \text{Otherwise} \end{cases}, \tag{9}$$

$$N_{bg} = \sum_{i=1}^{H} \sum_{j=1}^{W} (1 - M_{i,j}), \tag{10}$$

where H_r and W_r denote the height and width of the ground-truth box R.

During the training process, the neck features of the teacher and student models are separated into target level features and non-target level features for distillation after the SNS module and KLD module, respectively. Then the target level feature loss L_{tar} and the non-target level feature loss L_{ntar} are given as:

$$L_{tar} = \sum_{k=1}^{C} \sum_{i=1}^{H} \sum_{j=1}^{W} M'_{k,i,j} \cdot S'_{k,i,j} \cdot \big(F^{T\prime}_{k,i,j} - f(F^{S\prime}_{k,i,j})\big)^2, \tag{11}$$

$$L_{ntar} = \sum_{k=1}^{C}\sum_{i=1}^{H}\sum_{j=1}^{W} \overline{M'_{k,i,j}} \cdot S'_{k,i,j} \cdot \left(F^{T\prime}_{k,i,j} - f(F^{S\prime}_{k,i,j})\right)^2, \tag{12}$$

$$\begin{aligned} M'_{k,i,j} = f_{dim}(M_{i,j}), S'_{k,i,j} = f_{dim}(S_{i,j}), \\ \overline{M'_{k,i,j}} = E_{k,i,j} - M'_{k,i,j}, \end{aligned} \tag{13}$$

where E is the tensor whose values are all 1. $F^{T\prime}$ and $F^{S\prime}$ denote the feature map of the teacher detector and the student detector through the SNS module, respectively. Therefore, the feature loss of distillation L_{fea} is:

$$L_{fea} = \alpha L_{tar} + \beta L_{ntar}. \tag{14}$$

Reverse Information Transfer

Existing knowledge distillation methods usually constrain the student model using only the teacher model's output or features. However, inherent differences between the models can cause inconsistent feature perception, making the student model miss critical details and limiting its predictive potential. To address this, we propose a Reverse Information Transfer (RIT) branch that uses the student model's predicted output to enhance local feature knowledge guided by the teacher model. This reduces the teacher model's dominance, helping the student model better utilize its predictive power for more accurate results.

We feed student's multi-scale features $F^S_{i,j}$ from FPN and their segmented regions into the classification head to obtain class probability information. p_c denotes the probability that the feature region is an object of class c:

$$p_c = P(c \mid F), F \in \mathcal{F}, \tag{15}$$

where F denotes feature region and $\mathcal{F}$ denotes a set of feature regions.

For the classification information of each feature region, the class probability is aggregated using the following maximum function:

$$P^F = \max_{1 \le c \le Cls} p_c = \max_{1 \le c \le Cls} P\left(c \mid F\right), \tag{16}$$

where Cls denotes the category of the object and $Cls > 1$. P^F denotes the probability that there is a target in the feature region F.

Next, we construct a feature that propagates reverse information and map the class probabilities onto that feature:

$$F^{ri}_{i,j} \Leftarrow \sum_{F \in \mathcal{F}} P^F, \tag{17}$$

where $\Leftarrow$ denotes the the mapping function, which aggregates the classification information on all regions $F \in \mathcal{F}$ on the reverse information feature F^{ri}. Specifically, the P^F of each feature region in the inverse information feature are superimposed, and finally the locally enhanced feature is obtained.

In addition, we use feature matcher to optimize the extracted reverse information feature:

$$F^{fm}_{i,j} = \mathrm{sigmoid}(F^{ri}_{i,j} + \mathrm{conv}(F^{ri}_{i,j})), \tag{18}$$

where $\mathrm{conv}(\cdot)$ denotes feature through the convolution layer.

The reverse information loss is given as follows:

$$L_{rit} = \sum_{i=1}^{H}\sum_{j=1}^{W} \left(\mathrm{sigmoid}(F^{fm}_{i,j} \cdot F^{S}_{i,j}) - F^{fm}_{i,j}\right)^2. \tag{19}$$

To sum up, we train the student detector with the total loss is given as follows:

$$L = L_{gt} + L_{fea} + \gamma \cdot L_{rit}, \tag{20}$$

where L_{gt} is the original training loss of the student detector.

The distillation loss is calculated just on feature maps, which can be obtained from the neck of the detectors. So it can be easily applied to different detectors with FPN-like network structures.

Experiments

Datasets and Implement Details

We evaluate the proposed method on the SSDD (Zhang et al. 2021) and HRSID (Wei et al. 2020) SAR datasets. For evaluation, the standard COCO-style measurement, i.e., Average Precision (AP), is used. We also report mAP with IoU thresholds of 0.5 and 0.75, as well as AP for small, medium, and large objects.

Our proposed method, DIPKD, is implemented under the MMDetection (Chen et al. 2019) framework in Python. Unless otherwise stated, all the hyper-parameters follow the default settings of the corresponding student model for both training and testing. DIPKD uses α, β, γ to balance the loss of target and non-target in Eq.(14), reverse information loss in Eq.(19), respectively. $\tau = 0.5$ is used to adjust the attention distribution and mask rate is 25% for all the experiments. We adapt the hyper-parameters $\alpha = 5 \times 10^{-5}$, $\beta = 3.5 \times 10^{-5}$ and $\gamma = 4.5 \times 10^{-7}$ for all the experiments. Feature loss and inverse information loss are not normalized, but choose the form of unified hyperparameter which reduces the quantization effect. This kind of parameter setting balances the feature loss, the reverse information loss and the original loss well, and makes them exert their own constraint effect.

Performance on Different Detection Frameworks

We validate the effectiveness of the proposed DIPKD across multiple detection frameworks, including anchor-based single-stage detectors (RetinaNet) and two-stage detectors (Faster-RCNN). In these experiments, the teacher models use ResNet50 as the backbone, while the student model employs ResNet18 as the backbone.

In this research, we rigorously validated our DIPKD method across both the two-stage Faster R-CNN and single-stage RetinaNet frameworks. As detailed in Table 1, our experimental results demonstrate that the DIPKD strategy significantly boosts the performance of student detectors. Remarkably, in some cases, the student models even surpassed their teacher models. For instance, within the SSDD dataset, a student detector utilizing ResNet18 as its feature extractor achieved an mAP of 60.2%, outperforming the

	Detector	Backbone	mAP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	mAR	AR_S	AR_M	AR_L
SSDD	Faster-RCNN	ResNet50(tea)	58.9	94.5	65.6	53.9	66.3	60.9	64.3	60.7	**71.1**	60.5
		ResNet18(stu)	57.4	92.4	65.0	54.3	63.9	53.1	63.4	60.4	69.4	55.5
		ours	**60.2**	**95.3**	**70.2**	**56.6**	**66.5**	**60.2**	**65.3**	**62.2**	70.8	**63.5**
	Ritinanet	ResNet50(tea)	48.5	84.7	53.3	42.1	59.4	56.1	57.6	51.6	67.5	63.5
		ResNet18(stu)	44.9	82.4	43.9	38.2	57.3	45.4	54.5	48.0	65.7	55.0
		ours	**51.4**	**89.0**	**55.1**	**46.3**	**61.0**	**59.1**	**59.6**	**54.1**	**68.7**	**66.5**
HRSID	Faster-RCNN	ResNet50(tea)	**58.9**	78.8	**67.6**	59.3	**67.2**	**31.3**	62.7	61.1	**76.0**	**52.2**
		ResNet18(stu)	57.4	78.3	66.5	58.1	63.0	12.3	61.8	60.4	75.4	45.3
		ours	58.8	**80.5**	**67.6**	**59.4**	64.0	15.1	**63.0**	**61.5**	75.8	48.9
	Ritinanet	ResNet50(tea)	**50.9**	76.7	**56.4**	51.5	**59.1**	**18.4**	56.7	55.0	**71.0**	46.4
		ResNet18(stu)	44.8	70.2	49.4	45.3	51.6	8.2	50.6	48.7	65.7	44.1
		ours	50.6	**78.2**	55.1	**51.7**	56.2	12.4	**56.8**	**55.4**	68.7	**47.1**

Table 1: Results using different detection methods in the SSDD and HRSID dataset.

Method	**Ref**	SSDD		HRSID	
		mAP	AP_{50}	mAP	AP_{50}
Student		57.4	92.4	57.4	78.3
FKD	ICLR20	57.0	94.2	57.5	77.0
FRS	NIPS21	57.7	92.8	57.8	78.2
FGD	CVPR22	56.5	93.5	57.1	76.9
MGD	ECCV22	58.1	93.3	57.6	76.9
ARSD	TGRS22	57.9	93.7	58.2	79.1
TWA	TGRS24	59.0	95.2	58.1	79.3
IFKD	MM24	59.7	94.8	58.0	79.0
DIPKD		**60.2**	**95.3**	**58.8**	**80.5**

Table 2: Comparison with sate-of-the-art methods using different datasets on Faster-RCNN ResNet50-ResNet18.

SNS	KLD	RIT	mAP	AP_S	AP_M	AP_L
-	-	-	57.4	54.3	63.9	53.1
✓	-	-	59.4	56.2	65.9	56.8
-	✓	-	58.2	55.1	64.3	54.1
-	-	✓	59.7	55.8	65.8	59.6
✓	✓	-	59.8	56.6	66.2	56.6
✓	-	✓	60.1	56.4	66.1	56.0
-	✓	✓	59.8	56.1	65.7	57.5
✓	✓	✓	**60.2**	**56.6**	**66.5**	**60.2**

Table 3: Faster-RCNN ResNet50-ResNet18 ablation study of different modules on SSDD dataset.

teacher model's 58.5%. These findings highlight the substantial enhancement potential of student models when properly guided and suggest that the limits of performance can be further expanded.

Comparison with SOTA Methods

We compared our DIPKD method with several state-of-the-art knowledge distillation techniques, including FKD (Zhang and Ma 2020b), FGD (Yang et al. 2022c), MGD (Yang et al. 2022d), FRS (Zhixing et al. 2021), ARSD (Yang et al. 2022a), TWA (Yang, Zhang, and Yang 2024), and IFKD (Yang and Zhang 2024). FKD addresses foreground-background pixel imbalance and inter-pixel relationship extraction through guided and non-local distillation strategies. FRS, FGD, and MGD use masking techniques to highlight specific features. TWA and ARSD are designed for knowledge distillation in remote sensing images, while IFKD employs multi-information fusion. For a fair comparison, we implemented these methods using the hyperparameters specified in their original papers and evaluated them under the same experimental conditions and datasets.

As shown in Table 2, on the SSDD dataset, other distillation methods overlooked the purity or imbalance of the teacher model's knowledge, leading to a decline in the student model's performance after distillation. In contrast, our method effectively suppressed noise in the global features while refining and enhancing the local features, thereby improving the purity of the teacher's knowledge. This approach increased the student's mAP by 2.8%, demonstrating its effectiveness and surpassing the teacher model's performance on several key metrics.

Ablation Study

We conducted ablation experiments to evaluate each module's impact on the knowledge distillation process. Table 3 shows that all modules independently improved the student model's performance, with the RIT module contributing the most (+2.3 mAP), followed by SNS (+2.0 mAP) and KLD (+0.8 mAP). Combined, they achieved the highest improvement of +2.8 mAP, demonstrating the effectiveness of integrating noise suppression, feature decoupling, and supplementary knowledge.

τ	0.25	0.5	0.75	1.0	1.25
mAP	59.9	**60.2**	60.1	59.9	59.8
mAR	65.1	**65.3**	65.2	65.1	65.0

Table 4: Sensitivity study of temperature hyperparameters τ on SSDD dataset by Faster-RCNN ResNet50-ResNet18.

Mask Rate	0%	25%	50%	75%	100%
mAP	60.0	**60.2**	60.1	59.7	58.1
mAR	64.8	**65.3**	65.0	64.5	62.9

Table 5: Sensitivity study of mask rate hyperparameters on SSDD dataset by Faster-RCNN ResNet50-ResNet18.

Sensitivity to Hyperparameters

In Eq.(5) and Eq.(6), the temperature hyperparameter τ adjusts the distribution of pixels and channels in the feature map. The gap widens when $\tau < 1$ and narrows when $\tau > 1$. We experimented with different τ values to assess their impact. As shown in Table 4, $\tau = 0.5$ led to a 0.3 mAP and 0.2 mAR improvement compared to $\tau = 1$, which does not adjust the distribution. At $\tau = 0.5$, high-value pixels and channels are emphasized, helping the student detector focus on crucial parts and perform better. The worst result was only a 0.3 mAP drop from the best, indicating our method's robustness to τ.

In Eq.(8), we introduce a random variable n to adjust non-target level feature knowledge, with a default mask rate of 25%. Experiments were conducted to assess the sensitivity of this parameter. As shown in Table 5, a 25% mask rate led to a 0.2 mAP and 0.5 mAR improvement over a 0% mask rate. Randomly discarding background information helps suppress complex noise and reduces detector interference. However, when the mask rate is 100%, meaning only target-level features are distilled, performance declines sharply due to the loss of supplementary information from non-target level features.

Visualization Analysis

The MGD method uses mask generation, the FGD method focuses on local-global feature distinction, and the IFKD method incorporates multi-information fusion. Figure 4's heatmap visualizations compare detectors, showing that the undistilled student detector assigns higher weights to non-target areas, affected by speckle noise and background clutter in SAR images. While FGD, MGD, and IFKD improve performance, they still face background interference. In contrast, our DIPKD method effectively suppresses non-target weights and enhances feature focus in target-rich areas, like the center.

As illustrated in Figure 5, MGD's results are affected by background complexity and speckle noise. FGD and IFKD reduce background impact but remain sensitive to noise. In contrast, our DIPKD method handles both issues effectively, performing well on offshore, near-shore, and small target images.

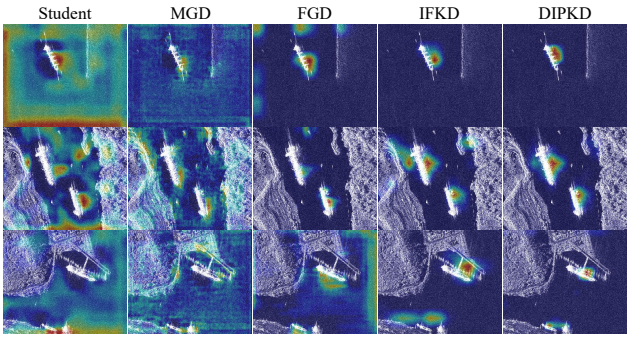


Figure 4: Visual feature heat maps results of the MGD, FGD, IFKD method and our proposed DIPKD method on the SSDD datasets.

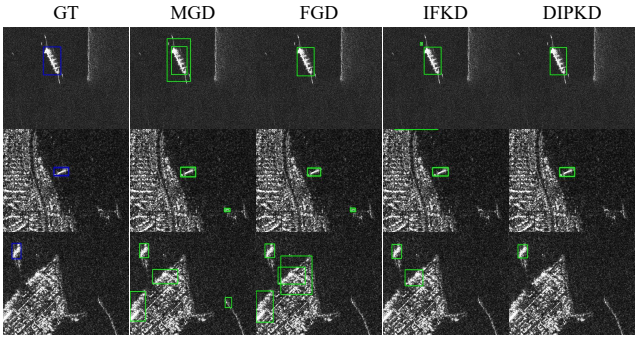


Figure 5: Visual detection results of images from the SSDD dataset using Faster-RCNN with ResNet18. The Ground truth, MGD, FGD, IFKD and ours DIPKD method are compared respectively.

Conclusion

This paper identifies the degradation of knowledge purity in SAR images as a key challenge for student model performance. To address this, we introduce the Dual Information Purification Knowledge Distillation (DIPKD) method, utilizing Selective Noise Suppression (SNS), Knowledge Level Decoupling (KLD), and Reverse Information Transfer (RIT) modules. Experiments on two detectors demonstrate our method's simplicity and effectiveness. Since it's based on FPN features, it applies across multiple detectors. The analysis shows that students not only grasp core knowledge from teachers but also surpass them in some metrics, though optimizing knowledge acquisition remains a future focus.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62372348 and 62176198, in part by the Key Research and Development Program of Shaanxi under Grant 2024GX-ZDCYL-02-10, in part by Shaanxi Outstanding Youth Science Fund Project under Grant 2023-JC-JQ-53, in part by Shaanxi Province Natural Science Basic Research Program under Grant 2023-JC-QN-0381.

References

Cai, Z.; and Vasconcelos, N. 2018. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6154–6162.

Chen, G.; Choi, W.; Yu, X.; Han, T.; and Chandraker, M. 2017. Learning efficient object detection models with knowledge distillation. *Advances in Neural Information Processing Systems*.

Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J.; et al. 2019. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*.

Chen, P.; Liu, S.; Zhao, H.; and Jia, J. 2021. Distilling knowledge via knowledge review. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5008–5017.

Choi, H.; and Jeong, J. 2019. Speckle noise reduction technique for SAR images using statistical characteristics of speckle noise and discrete wavelet transform. *Remote Sensing*.

Guo, J.; Han, K.; Wang, Y.; Wu, H.; Chen, X.; Xu, C.; and Xu, C. 2021. Distilling object detectors via decoupled features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2154–2164.

He, K.; Gkioxari, G.; Dollár, P.; and Girshick, R. 2017. Mask r-cnn. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2961–2969.

He, L.; Cheng, D.; Wang, N.; and Gao, X. 2024. Exploring Homogeneous and Heterogeneous Consistent Label Associations for Unsupervised Visible-Infrared Person ReID. *International Journal of Computer Vision*.

Heo, B.; Lee, M.; Yun, S.; and Choi, J. Y. 2019. Knowledge transfer via distillation of activation boundaries formed by hidden neurons. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 3779–3787.

Hinton, G.; Vinyals, O.; and Dean, J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE/CVF Conference on Computer vision and Pattern Recognition*, 7132–7141.

Li, C.; Cheng, G.; Wang, G.; Zhou, P.; and Han, J. 2023. Instance-aware distillation for efficient object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 1–11.

Li, K.; Cheng, G.; Bu, S.; and You, X. 2017. Rotation-insensitive and context-augmented object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2337–2348.

Li, Q.; Jin, S.; and Yan, J. 2017. Mimicking very efficient network for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6356–6364.

Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017a. Feature pyramid networks for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2117–2125.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017b. Focal loss for dense object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2980–2988.

Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; and Berg, A. C. 2016. Ssd: Single shot multibox detector. In *Proceedings of the IEEE/CVF European Conference on Computer Vision*, 21–37.

Liu, Y.; Cao, J.; Li, B.; Yuan, C.; Hu, W.; Li, Y.; and Duan, Y. 2019. Knowledge distillation via instance relationship graph. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7096–7104.

Ni, Z.; Yang, F.; Wen, S.; and Zhang, G. 2023. Dual relation knowledge distillation for object detection. *arXiv preprint arXiv:2302.05637*.

Park, W.; Kim, D.; Lu, Y.; and Cho, M. 2019. Relational knowledge distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3967–3976.

Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 779–788.

Redmon, J.; and Farhadi, A. 2017. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7263–7271.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*.

Romero, A.; Ballas, N.; Kahou, S. E.; Chassang, A.; Gatta, C.; and Bengio, Y. 2014. Fitnets: Hints for thin deep nets. *arXiv preprint arXiv:1412.6550*.

Singh, P.; and Shree, R. 2016. Analysis and effects of speckle noise in SAR images. In *International Conference on Advances in Computing, Communication, Automation*, 1–5.

Tian, Z.; Shen, C.; Chen, H.; and He, T. 2020. FCOS: A simple and strong anchor-free object detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1922–1933.

Tung, F.; and Mori, G. 2019. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1365–1374.

Wan, H.; Chen, J.; Huang, Z.; Du, W.; Xu, F.; Wang, F.; and Wu, B. 2023. Orientation Detector for Small Ship Targets in SAR Images Based on Semantic Flow Feature Alignment and Gaussian Label Matching. *IEEE Transactions on Geoscience and Remote Sensing*.

Wang, J.; Chen, Y.; Zheng, Z.; Li, X.; Cheng, M.-M.; and Hou, Q. 2024. CrossKD: Cross-head knowledge distillation for object detection. In *Proceedings of the IEEE/CVF*

Conference on Computer Vision and Pattern Recognition, 16520–16530.

Wang, P.; Sun, X.; Diao, W.; and Fu, K. 2019a. FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 3377–3390.

Wang, T.; Yuan, L.; Zhang, X.; and Feng, J. 2019b. Distilling object detectors with fine-grained feature imitation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4933–4942.

Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; and Shi, J. 2020. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation. *IEEE Access*, 120234–120254.

Woo, S.; Park, J.; Lee, J.-Y.; and Kweon, I. S. 2018. Cbam: Convolutional block attention module. In *Proceedings of the IEEE/CVF European Conference on Computer Vision*, 3–19.

Xie, X.; Cheng, G.; Wang, J.; Yao, X.; and Han, J. 2021. Oriented R-CNN for object detection. In *Proceedings of the IEEE/CVF international Conference on Computer Vision*, 3520–3529.

Xu, T.; Sun, X.; Diao, W.; Zhao, L.; Fu, K.; and Wang, H. 2021. ASSD: Feature aligned single-shot detection for multiscale objects in aerial imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 1–17.

Yang, X.; and Zhang, S. 2024. Information Fusion with Knowledge Distillation for Fine-grained Remote Sensing Object Detection. In *ACM Multimedia 2024*.

Yang, X.; Zhang, S.; and Yang, W. 2024. Two-Way Assistant: A Knowledge Distillation Object Detection Method for Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*.

Yang, Y.; Sun, X.; Diao, W.; Li, H.; Wu, Y.; Li, X.; and Fu, K. 2022a. Adaptive knowledge distillation for lightweight remote sensing object detectors optimizing. *IEEE Transactions on Geoscience and Remote Sensing*, 1–15.

Yang, Y.; Sun, X.; Diao, W.; Yin, D.; Yang, Z.; and Li, X. 2022b. Statistical sample selection and multivariate knowledge mining for lightweight detectors in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 1–14.

Yang, Z.; Li, Z.; Jiang, X.; Gong, Y.; Yuan, Z.; Zhao, D.; and Yuan, C. 2022c. Focal and global knowledge distillation for detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4643–4652.

Yang, Z.; Li, Z.; Shao, M.; Shi, D.; Yuan, Z.; and Yuan, C. 2022d. Masked generative distillation. In *Proceedings of the IEEE/CVF European Conference on Computer Vision*, 53–69.

Zagoruyko, S.; and Komodakis, N. 2022. Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer. In *Proceedings of the IEEE/CVF International Conference on Learning Representations*.

Zhang, L.; and Ma, K. 2020a. Improve object detection with feature-based knowledge distillation: Towards accurate and efficient detectors. In *Proceedings of the IEEE/CVF International Conference on Learning Representations*.

Zhang, L.; and Ma, K. 2020b. Improve object detection with feature-based knowledge distillation: Towards accurate and efficient detectors. In *Proceedings of the IEEE/CVF International Conference on Learning Representations*.

Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. 2021. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sensing*.

Zhixing, D.; Zhang, R.; Chang, M.; Liu, S.; Chen, T.; Chen, Y.; et al. 2021. Distilling object detectors with feature richness. *Advances in Neural Information Processing Systems*, 5213–5224.