

## 基于YOLO模型合并的SAR图像人造目标检测

何适存 王欣超 陈思伟\*  
(国防科技大学电子科学学院 长沙 410073)

**摘要:** 高价值人造目标检测是合成孔径雷达(SAR)的重要应用。现有基于深度学习的检测方法,对模型中的参数化知识的有效复用不足,在不同数据集上常表现出泛化性能弱的问题,导致实际迁移应用困难。针对这一问题,该文从向模型学习(LFM)的角度出发,提出一种基于YOLO模型合并的SAR图像人造目标检测方法。该方法的核心思想是以多个同构模型为知识来源,将源模型中已形成的特征提取、多尺度特征融合和目标判别能力整合到统一检测框架中,实现已有模型知识向检测性能增益的转化。该方法以共享预训练骨干网络为基础,首先将多源模型的颈部网络参数迁移至合并模型。随后,采用多尺度通道拼接并结合点卷积的方式,实现多源模型的多尺度特征融合。同时,在各特征尺度并行拓展检测头分支,以保持每个源模型的判别能力。基于SADD, SSDD和HRSID 3个公开SAR数据集的实验结果表明,所提方法可有效提升SAR图像飞机、舰船等人造目标的检测性能。在Recall, mAP50和mAP50-95指标上均实现了稳定的性能增益,尤其在高交并比阈值和少样本场景下,展现出更优的检测鲁棒性与泛化能力。

**关键词:** 合成孔径雷达; 人造目标检测; 模型合并; 特征融合; 参数化知识

**中图分类号:** TN958

**文献标识码:** A

**文章编号:** 2095-283X(2026)x-0001-19

**DOI:** 10.12000/JR26080

**CSTR:** 32380.14.JR26080

**引用格式:** 何适存, 王欣超, 陈思伟. 基于YOLO模型合并的SAR图像人造目标检测[J]. 雷达学报(中英文), 待出版. doi: 10.12000/JR26080.

**Reference format:** HE Shicun, WANG Xinchao, and CHEN Siwei. SAR image artificial target detection based on YOLO model merging[J]. *Journal of Radars*, in press. doi: 10.12000/JR26080.

## SAR Image Artificial Target Detection Based on YOLO Model Merging

HE Shicun WANG Xinchao CHEN Siwei\*

(College of Electronic Science and Technology, National University of Defense Technology,  
Changsha 410073, China)

**Abstract:** High-value artificial target detection is a critical application of synthetic aperture radar (SAR). Existing deep learning-based detection methods often fail to adequately reuse parameterized knowledge from trained models and demonstrate limited generalization across different datasets. This limitation hinders their practical application in new scenarios. To address this issue, this study proposes a SAR image artificial target detection method based on YOLO model merging, adopting a learning from models approach. The core idea is to use multiple homogeneous models as knowledge sources. The method integrates the feature extraction, multiscale feature fusion, and target discrimination capabilities developed in these source models into a unified detection framework, thereby transforming existing model knowledge into improved detection performance. Specifically, based on a shared pretrained backbone, the neck network parameters of multiple source models are

收稿日期: 2026-04-29; 改回日期: ; 网络出版: 2026-xx-xx

\*通信作者: 陈思伟 [chenswnudt@163.com](mailto:chenswnudt@163.com) \*Corresponding Author: CHEN Siwei, [chenswnudt@163.com](mailto:chenswnudt@163.com)

基金项目: 国家自然科学基金(U24B20189, 62122091), 湖南省省级科技计划(2024RC1040)

Foundation Items: The National Natural Science Foundation of China (U24B20189, 62122091), the Science and Technology Innovation Program of Hunan Province (2024RC1040)

责任主编: 高贵 Corresponding Editor: GAO Gui

©The Author(s) 2026. This is an open access article under the CC-BY 4.0 License

(<https://creativecommons.org/licenses/by/4.0/>)

first transferred to the merged model. Multiscale feature fusion is then achieved through channel concatenation and point-wise convolution. In addition, parallel detection head branches are introduced at each feature scale to preserve the discriminative capabilities of the source models. Experimental results on three public SAR datasets (SADD, SSDD, and HRSID) demonstrate that the proposed method effectively improves the detection performance of man-made targets in SAR images. Consistent improvements are observed in Recall, mAP50, and mAP50-95, particularly under high intersection-over-union thresholds and limited-sample scenarios, where the proposed method exhibits superior robustness and generalization ability.

**Key words:** Synthetic Aperture Radar; Artificial target detection; Model merging; Feature fusion; Parameterized knowledge

## 1 引言

合成孔径雷达(Synthetic Aperture Radar, SAR)具有全天时、全天候对地观测能力,能够在不受光照条件和气象环境影响的情况下获取高分辨率雷达图像,已在地形测绘、海洋监测、灾害评估等领域得到广泛应用<sup>[1-4]</sup>。同时,研究人员相继构建并公开了一系列高质量SAR图像数据集,如AIR-SARShip-Dataset<sup>[5]</sup>,SSDD<sup>[6]</sup>和SAR-AIRCRAFT1.0<sup>[7]</sup>等,这些数据集具有多样化的成像分辨率、目标类型覆盖和场景复杂度,为SAR图像目标智能检测与识别研究提供了坚实的数据支撑。

得益于丰富的SAR数据资源与人工智能技术的快速革新,以卷积神经网络(Convolutional Neural Network, CNN)为代表的深度学习技术已被广泛应用于SAR图像目标检测领域。此外,多尺度特征融合策略和注意力机制等关键技术的引入,有效提升了SAR目标检测性能<sup>[8]</sup>。早期工作直接将光学目标检测框架迁移至SAR领域<sup>[9-11]</sup>,后续逐步针对SAR图像特有的散射特性<sup>[12, 13]</sup>,相干斑现象<sup>[14, 15]</sup>,以及小样本情形<sup>[16]</sup>等进行优化。现有研究中常用的检测模型主要包括3类。一是单阶段检测模型,代表方法有YOLO系列<sup>[17]</sup>、SSD<sup>[18]</sup>、FCOSR<sup>[19]</sup>等,其特点是端到端、实时性强和便于部署。二是双阶段检测模型,代表方法有Faster R-CNN<sup>[20]</sup>和Cascade R-CNN<sup>[21]</sup>等。这类方法通常在检测精度方面表现更优,但计算复杂度较高。三是基于Transformer架构的Swin-Transformer<sup>[22]</sup>、DETR<sup>[23]</sup>、ViT<sup>[24]</sup>等。基于上述模型架构,研究人员围绕SAR成像特性开展了大量分析与改进工作,积累了多样化的深度学习检测模型,并在多种应用场景下表现出良好的检测性能。然而,在实际应用中,仍普遍面临着模型知识难以高效整合、训练成本较高以及模型泛化能力受限等问题<sup>[25]</sup>。

值得注意的是,原始数据在训练过程中已被高度抽象并内化为以模型权重为载体的参数化知识。考虑到模型重训练所需的时间成本与算力资源,复

用模型相较于复用数据更加经济高效。向模型学习(Learning from Models, LFM)正是基于这一理念提出的新兴研究方法<sup>[26]</sup>,包括模型微调、编辑与合并等多种技术。模型合并(Model Merging)<sup>[27-30]</sup>作为LFM的重要组成部分,为高效利用模型参数化知识提供了一条切实可行的路径。该技术通过在参数或结构空间对多任务知识进行融合,能够在较低训练开销条件下,有效整合并扩展源模型能力,规避高昂的重训练成本。

现有研究表明,通过模型合并技术,构建的新模型能够在一定程度上保留源模型的能力,并促进来自不同数据和任务的知识快速整合。现有模型合并研究已形成多种技术路线,既包括权重平均、任务向量叠加、参数插值等直接作用于模型权重的参数合并方法<sup>[26]</sup>,也包括基于模块组合、层级融合或数据流搜索的结构合并方法<sup>[30]</sup>。代表性工作包括:Ilharco等人<sup>[31]</sup>提出了一种模型参数的合并方法,将训练过程视作在模型参数向量上叠加任务向量,并通过对任务向量执行简单加法操作来合并多任务模型。Roberto等人<sup>[32]</sup>提出了一种层级融合模型合并方法,在保持任务特定最终层的独特性的同时,统一主要负责特征提取的初始层,实现多任务单模型的构建。Takuya等人<sup>[33]</sup>提出了一种利用进化算法的模型合并方法,在参数空间和数据流空间中同时有效组合模型,实现了跨领域的知识融合。这些研究证明了模型合并技术在高效整合模型参数化知识和提升模型性能的可行性。

一方面,SAR图像目标检测受限于成像机理与场景特性,长期面临目标尺度差异显著、背景复杂以及跨数据集泛化能力不足等严峻挑战。另一方面,随着海量遥感数据与检测模型的不断积累,在不同数据集上独立训练得到的检测模型,在多尺度特征提取、目标定位精度及类别判别上存在一定的互补性,蕴含着丰富的融合利用价值。然而,现有模型合并研究主要集中于大语言模型和基础模型,针对SAR目标检测任务,尤其是对具有相同结

构，广泛使用的目标检测模型，如何有效融合多个源模型中的参数化知识仍缺乏针对性研究。

由于大语言模型与目标检测模型在参数空间的几何特性与结构约束不同，二者在模型合并后的表现存在差异。大语言模型具备大量的参数冗余，不同任务微调产生的参数任务向量在高维空间中往往呈现近似正交的特性<sup>[34]</sup>，这使得参数合并造成的冲突区域较小，可以高效实现多模型能力组合。而目标检测模型，其特征提取网络和检测头依赖跨尺度的精密特征对齐，不同任务在共享参数上易发生符号与方向冲突，这种冲突会直接破坏特征提取或定位能力，从而导致参数合并方法在检测任务上的失效。通用的模型参数合并方法直接迁移应用到目标检测模型存在困难，因此，目前相较于直接合并参数，从网络结构和特征融合路径出发，探索结构层面的模型合并方法可行性更强。

YOLO作为实时目标检测领域的标杆架构，凭借其卓越的速度与精度平衡，在SAR目标检测领域具有极高的代表性。其从v1至v11持续迭代的技术演进路线，构建了包含海量预训练模型与多样化变体的庞大开源生态，这种丰富的“参数化知识库”与标准化的网络设计，天然契合模型合并技术对于高效复用现有参数、跨模型融合特征的核心诉求，使其成为验证模型合并策略的理想载体之一。

基于上述背景，为高效利用现有SAR检测模型中蕴含的参数化知识，本文提出了一种基于YOLO系列模型结构合并的SAR图像目标检测方法。本文的创新点在于，面向SAR目标检测任务，将多个源模型中积累的参数化知识高效转化为合并模型的性能增益。所提方法以同构源模型为知识载体，通过整合其在特征提取、多尺度表征和目标判别等方面的差异化检测能力，将源模型中分散存在的参数化知识转化为合并模型的互补检测优势，从而提升复杂SAR场景下的人造目标检测性能。在3组公开的SAR目标检测数据集上的测试结果表明，构建的合并模型相比源模型具有更强的人造目标检测能力。

## 2 方法设计

为避免概念混淆，本文首先对相关术语进行说明。本文中的“源域”遵循域自适应<sup>[25]</sup>领域中的基本定义，指具有标注样本并用于训练源模型的数据域，例如SSDD、HRSID等数据集。“源模型”指在对应源域上训练得到的检测模型。“原始数据”指源模型训练阶段所使用的原始标注样本集合，在后续优化微调过程中，这些数据继续用于监督融合层的端到端优化。需要说明的是，本文中的“异源”

并不完全等同于传统域自适应中的源域—目标域设定，而是强调不同源域数据在成像传感器、采集平台、成像分辨率或场景来源等方面存在差异，即不同源模型所学习的参数化知识来自具有物理来源差异的数据分布。

本文所提方法基于YOLO系列目标检测架构进行设计。源模型均采用由骨干网络、颈部网络和检测头组成的三段式结构。其中，骨干网络主要用于提取底层通用特征，颈部网络负责多尺度特征融合，检测头用于输出目标类别及其边界框预测结果。所提方法的核心目标在于有效复用并融合多个源模型中蕴含的参数化知识，从而在较低训练成本下构建具备更强检测能力的新模型。针对来源于同一预训练基座、模型结构一致但在不同数据集上完成微调的目标检测模型，本文在保持骨干网络参数冻结与结构不变的条件下，重点对颈部网络与检测头进行结构级重构与拓展。具体而言，通过引入双分支特征金字塔结构，保留各源模型对应颈部网络的原始参数配置，并采用通道拼接与点卷积的方式实现多尺度特征的融合，该设计使得不同源模型在特征提取阶段形成的语义信息能够在统一的特征空间中得到有效整合互补。同时，对检测头部分进行尺度一致的并行拓展，使其与重构后的颈部网络有效耦合，并通过后处理阶段对来自不同尺度检测头的预测结果进行统一融合。该方法使新模型充分继承了多个源模型在特征提取与目标定位上的优势，使合并模型具备多模型的检测能力，从而高效地提升整体检测性能与泛化能力。本文所提的基于YOLO模型合并的SAR图像人造目标检测方法框图如图1所示。

### 2.1 模型合并函数设计

本文采用的同组源模型来源于同一预训练基座，网络结构完全一致，在不同的训练数据集上进行了微调，因此其参数在层次结构和排列顺序上天然对齐，无需额外进行参数置换或匹配操作。本文中重点考虑通过模型结构的重新设计以实现模型参数的有效合并。表1总结了YOLO架构中对模型性能具有影响的主要参数类型。模型中的可学习参数主要包括卷积层权重，其次为批归一化(Batch Normalization, BN)层参数。其中，BN层中的缓冲区参数不参与反向传播更新，属于不可学习参数，但其统计量会在训练前向传播过程中持续更新，并对中间特征分布产生影响，进而影响模型的检测性能。因此，在模型合并过程中，将对表中所列参数进行整体迁移。

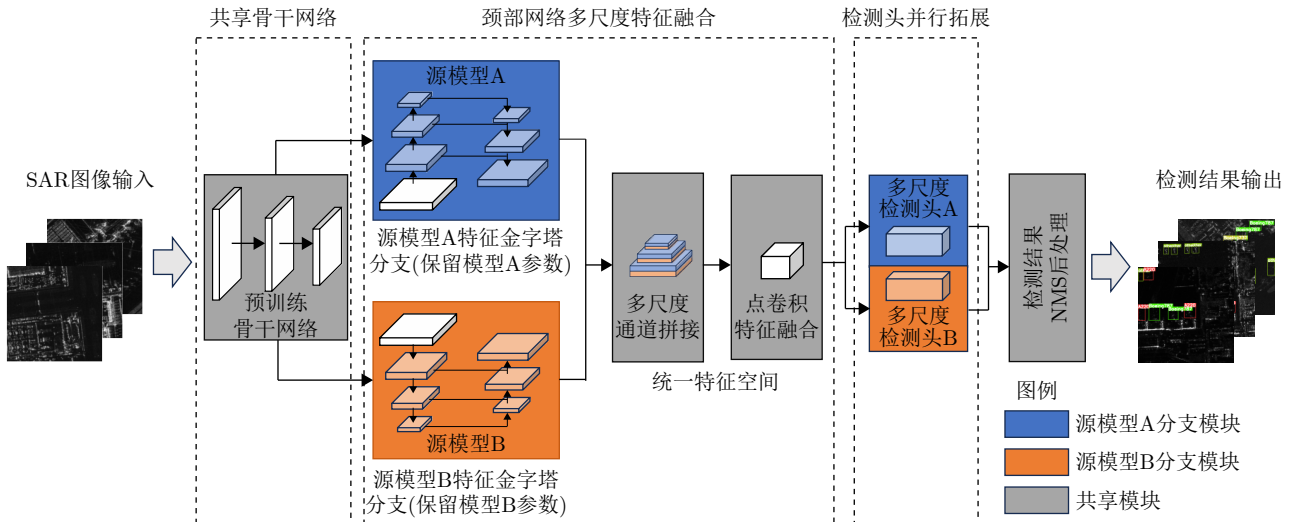


图1 基于YOLO模型合并的SAR图像人造目标检测方法框图

Fig. 1 Block diagram of SAR image artificial target detection method based on YOLO model merging

表1 目标检测模型参数组成

Tab. 1 Parameter composition of target detection models

参数类型	参数命名	可学习参数
卷积权重	*.conv.weight	是
BN缩放	*.bn.weight	是
BN平移	*.bn.bias	是
其它权重	*.weight	是
其它偏置	*.bias	是
	*.bn.running_mean	否
BN缓冲区	*.bn.running_var	否
	*.bn.num_batches_tracked	否

注: \*.指参数所属的具体模块

下面对模型合并过程建模, 设有 $K$ 个网络结构完全一致的模型集合 $\{M_i\}_{i=1}^K$ ,  $M_i$ 模型参数向量定义为 $\theta^{(i)}$ , 基座模型为 $M_0$ 。模型合并的核心是构建一个作用于 $\{M_i\}_{i=1}^K$ 的合并函数 $f$ , 得到合并模型的参数向量 $\theta^{(M)}$ , 描述公式如下:

$$\theta^{(M)} = f(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(K)}) \quad (1)$$

对于YOLO单阶段目标检测网络结构, 模型预测的前向过程可写为

$$\begin{cases} \mathbf{z} = \text{Backbone}(x; \theta_B) \\ \mathbf{F}_N = \text{Neck}(\mathbf{z}; \theta_N) \\ Y = \text{Head}(\mathbf{F}_N; \theta_H) \end{cases} \quad (2)$$

其中,  $x$ 表示输入的原始图像,  $\theta_B$ 表示模型的骨干网络的卷积特征提取层,  $\theta_N$ 表示颈部网络的多尺度特征金字塔层,  $\theta_H$ 则表示任务目标检测头。Backbone( $\cdot$ ;  $\theta_B$ )输出底层特征图 $\mathbf{z}$ , Neck( $\cdot$ ;  $\theta_N$ )输出金字塔特征张量 $\mathbf{F}_N$ , Head( $\cdot$ ;  $\theta_H$ )对不同尺度的特征图输出检测结果 $Y$ 。

本文中, 使用两个相同架构, 在不同数据集上训练的YOLO目标检测模型, 分别记为 $M^{(1)}$ 和 $M^{(2)}$ , 统一的网络结构为骨干网络-颈部网络-检测头(Backbone-Neck-Head), 对应的参数集合记为

$$M^{(k)} = (\theta_B^{(k)}, \theta_N^{(k)}, \theta_H^{(k)}), k = 1, 2 \quad (3)$$

其中,  $\theta_B$ 是骨干网络中卷积特征提取层的参数向量,  $\theta_N$ 是颈部网络中特征金字塔与多尺度融合层的可学习参数向量,  $\theta_H$ 是负责类别预测与边界框回归的检测头参数。

合并后的模型各部分参数向量分别记作 $\theta_B^{(M)}$ ,  $\theta_N^{(M)}$ ,  $\theta_H^{(M)}$ , 下面按模型结构介绍本文设计的模型合并方法。

## 2.2 基座骨干网络共享

骨干网络主要负责进行通用特征提取。在预训练的基座模型权重中, 这部分的网络参数已经在COCO和ImageNet等大规模数据集上进行了充分训练, 具有相对稳健的对图像中目标进行特征提取的能力。在进行小规模数据集下的模型训练时可以考虑冻结, 这能有效防止模型过拟合, 既不会造成较大的精度损失, 也便于模型的合并<sup>[34, 35]</sup>。因此, 本文方法的骨干网络部分在模型合并的流程中采用共享参数的方法, 即合并后模型骨干网络使用预训练基座模型的权重参数。

为了维持模型底层特征的正常输出, 骨干网络参数冻结, 由于源模型基于相同的预训练权重微调, 故源模型与合并模型的参数保持一致, 即

$$\theta_B^{(M)} = \theta_B^{(1)} = \theta_B^{(2)} \quad (4)$$

其中,  $\theta_B^{(1)}$ 和 $\theta_B^{(2)}$ 分别为 $M^{(1)}$ 和 $M^{(2)}$ 的骨干网络参

数，该操作保持骨干网络的通道数，特征图尺寸不变，保留基座模型捕获底层纹理、边缘以及高层语义信息的能力。

### 2.3 颈部网络参数迁移与多尺度特征融合

颈部网络的核心作用是融合不同尺度的特征，高分辨率特征负责小目标，低分辨率特征负责大目标，网络结构由多个功能明确的模块构成。以YOLO<sub>v11</sub>系列的颈部网络为例，网络包含卷积(Conv)、上采样(Upsample)、拼接(Concat)和残差模块(C3k2)，最终输出的特征同时具备目标定位细节和语义信息。图2展示了单个源模型的模块组成以及各模块的层间连接关系。

Upsample模块可以将低分辨率特征图放大到高分辨率，让语义信息能传递到小目标检测层。Concat模块将对齐后的不同尺度的特征图在通道维度拼接，合并多尺度信息，如高分辨率的细节与低分辨率的语义。C3k2模块是由经典的C3模块改进，k2代表使用了两个并行的卷积层，有较强的局部特征建模能力。其中，由于Upsample与Concat模块并没有可变参数，在进行模型合并的过程中只需要对Conv和C3k2模块进行源模型参数的迁移。

在此基础上，本文基于模型合并的原理，对颈部网络部分的结构进行了重新设计，具体来讲，采用了双分支的特征金字塔(Dual-Branch FPN/PAN)结构。为了充分利用不同模型对深层特征提取的差异性互补能力，我们保留两组源模型中独立的颈部

参数 $\theta_N^{(1)}$ 和 $\theta_N^{(2)}$ 。底层特征经过颈部网络得到的深层特征张量分别记作 $\mathbf{F}_N^{(1)}, \mathbf{F}_N^{(2)} \in \mathbb{R}^{C \times H \times W}$ ，这里的 $C$ 为通道数， $H$ 和 $W$ 为特征金字塔的空间尺寸。为了实现特征层的融合，我们采用“通道拼接+点卷积”的方法，特征金字塔张量具有3个不同尺度，将两个分支的张量在相同尺度上进行处理且不同尺度的处理方法一致。每个尺度的合并公式均可写为

$$\mathbf{F}_N^{(M)} = \mathbf{C}_{1 \times 1} \left( \text{Concat} \left( \mathbf{F}_N^{(1)}, \mathbf{F}_N^{(2)} \right); \mathbf{W}_M \right) \quad (5)$$

其中， $\text{Concat}(\cdot)$ 表示在通道维的拼接操作，生成的中间特征张量维度为 $2C \times H \times W$ 。 $\mathbf{C}_{1 \times 1}$ 表示卷积核大小为 $1 \times 1$ 的卷积层，用于将通道数降维回到 $C$ ，并使之与检测头匹配。将该卷积层的权重矩阵 $\mathbf{W}_M$ 初始化为对角元素为1，其余元素为0的对角分块矩阵，可表示为

$$\mathbf{W}_M = [\mathbf{I}_C \mid \mathbf{I}_C] \quad (6)$$

其中， $\mathbf{I}_C$ 表示维度为 $C \times C$ 的单位矩阵。

上述操作的核心在于，先将两组特征金字塔张量映射至统一特征空间，再完成线性组合。具体而言，通过训练优化权重矩阵 $\mathbf{W}_M$ 参数，能够隐式调节两个源模型特征在合并模型中的融合方式及其贡献比例，同时缓解通道拼接可能引入的冗余信息对特征表示的不利影响。同时，对角分块矩阵初始化使融合层在训练初始阶段能够近似保留各源模型原有特征，避免随机初始化造成异源特征的无序混合和有效表征破坏。该初始化方式为融合层提供了较

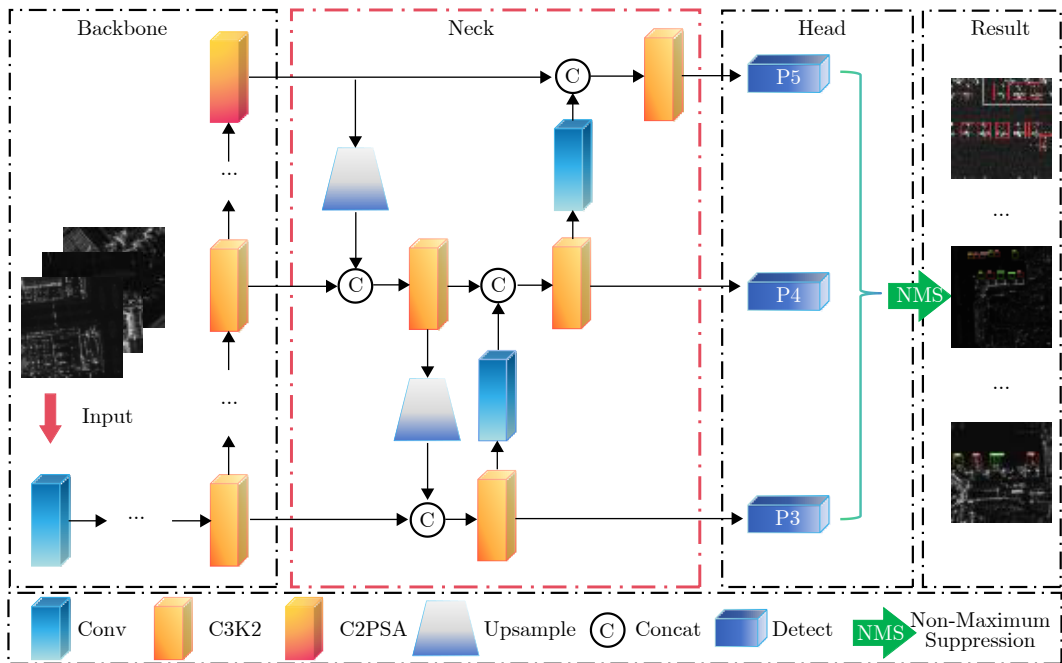


图2 源模型网络结构

Fig. 2 Network architecture of the source model

稳定的优化起点,使模型可以在源模型已有特征表达的基础上进一步学习通道重组与互补融合关系,从而降低训练初期的特征扰动,并有助于提升融合层微调过程的收敛稳定性。

在融合策略上,采用“通道拼接+1×1卷积”的方式。其中,通道拼接能够完整保留两个源模型的原始表征,避免直接相加可能造成的信息抵消。1×1卷积则通过可训练参数对跨通道信息进行线性重组。模型通过可训练的权重矩阵对不同来源的通道响应进行重加权,从而增强与任务相关的融合表征。融合后的特征 $\mathbf{F}_N^{(M)}$ 能够同时兼顾两个源模型的高层语义信息与底层定位细节,并尽量保留两个源模型中对下游检测任务有价值的响应模式。

在融合层训练阶段,本文在保持源模型主体结构冻结的条件下,仅对卷积融合层参数进行优化。设卷积融合层参数为 $\theta_f$ ,冻结的骨干网络、特征金字塔分支及检测头参数为 $\theta_{\text{fix}}$ ,用于微调的训练数据集为 $D$ ,则融合层优化目标可表示为

$$\theta_f^* = \arg \min_{\theta_f} \mathcal{L}_{\text{det}}(D; \theta_{\text{fix}}, \theta_f) \quad (7)$$

不额外采用特定的采样策略或加权损失函数,使用两个源模型对应的原始训练样本,并基于标准YOLO检测损失对融合层进行端到端优化。源模型训练与融合层优化使用一致的损失函数,以YOLOv8的损失函数为例,检测损失主要由边界框回归损失、分类损失和分布焦点损失组成,可表示为

$$\mathcal{L}_{\text{det}} = \lambda_{\text{box}} \mathcal{L}_{\text{box}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}} + \lambda_{\text{dfl}} \mathcal{L}_{\text{dfl}} \quad (8)$$

其中,3项损失权重遵循原文中的基准设置<sup>[17]</sup>,边界框回归损失 $\lambda_{\text{box}} = 7.5$ ,分类损失 $\lambda_{\text{cls}} = 0.5$ ,分布焦点损失 $\lambda_{\text{dfl}} = 1.5$ 。

在融合层微调过程中,本文保持检测损失形式及权重设置与源模型训练阶段一致,而不额外引入新的损失项或针对不同源模型分支重新设置损失权重。该设置可以保证融合层微调阶段与源模型训练阶段具有一致的检测优化目标,避免由于损失重加权引入额外变量,使性能变化能够更明确地反映融合层对异源特征对齐与重组的作用。同时,较大的边界框回归权重 $\lambda_{\text{box}}$ 有助于强化SAR目标定位能力,适中的分布焦点损失权重 $\lambda_{\text{dfl}}$ 有利于提升边界框精细回归效果,而相对较小的分类损失权重 $\lambda_{\text{cls}}$ 则可以避免在类别数量有限的SAR检测任务中过度强化分类项。训练过程中仅更新轻量化卷积融合层,其余网络参数均保持冻结,从而在较低适配成本下实现异源特征的稳定融合。

## 2.4 检测头分支拓展

对于检测头部分,为保持数据流与原始网络一

致,同时使两个模型中的检测头对不同目标的差异化检测能力保持完整,合并模型每个检测尺度均添加一个检测头。基于以上方法,合并模型融合了源模型中可学习的参数化知识,同时保留了不可学习的模块参数。

对于合并后的金字塔特征 $\mathbf{F}_N^{(M)}$ ,每套检测头负责输出3路并行分支(分类、边界框回归、置信度),对于两套模型,检测头的参数记为 $\theta_H^{(k)}$ ,输出 $Y^{(k)}$ 为

$$\begin{aligned} Y^{(k)} &= \text{Head} \left( \mathbf{F}_N^{(M)}; \theta_H^{(k)} \right) \\ &= \left\{ \left( \mathbf{b}_i^{(k)}, s_i^{(k)}, c_i^{(k)} \right) \mid i = 1, 2, \dots, N_k, k = 1, 2 \right\} \end{aligned} \quad (9)$$

其中, $\mathbf{b}_i^{(k)} \in \mathbb{R}^4$ ,为检测框中心点坐标 $(x, y)$ 与宽高 $(w, h)$ , $s_i^{(k)} \in [0, 1]$ 为置信度, $c_i^{(k)} \in \{1, 2, \dots, C_{\text{cls}}\}$ 为类别标签。

通过合并并行输出结果与NMS(非极大值抑制)处理,最终的检测结果表示为

$$Y^{(M)} = \text{NMS}_{\tau_c, \tau_{\text{IoU}}} \left( Y^{(1)} \cup Y^{(2)} \right) \quad (10)$$

其中,NMS算子表示保留置信度最高的且互不重叠的集合, $\tau_c$ 表示置信度阈值, $\tau_{\text{IoU}}$ 表示交并比阈值。

图3中展示了基于v11版本设计的合并模型,其余版本的设计方法类似,图中序号1和2标注的模块迁移使用源模型1和2的模型参数。

## 3 实验结果与分析

### 3.1 实验平台

本实验基于Windows 10操作系统与CUDA 11.8开发环境搭建,采用PyTorch框架编写算法代码,通过单张NVIDIA A800 GPU完成训练加速。实验选取Ultralytics团队开源的YOLOv5、YOLOv8、YOLOv11 3个版本的网络架构,基于各版本预训练模型,在选定数据集上完成微调训练。所有模型均选用m参数量级版本,每个YOLO版本均训练2个差异化源模型,再按照本文所提方法构建合并模型,最终分别测试源模型与对应合并模型的检测性能。

### 3.2 实验数据集介绍与实验设置

实验采用公开发布的SAR人造目标检测数据集,选取飞机与舰船两类典型高价值人造目标作为研究对象。飞机目标由于结构复杂、散射机制多样且对成像方位角高度敏感,在SAR图像中常呈现为离散散射中心与不完整轮廓,类内差异显著,具有较高的检测与识别难度<sup>[36]</sup>。因此,本文选取SADD (SAR Aircraft Detection Dataset)飞机目标数据集

开展同源实验，以验证所提出模型合并方法在复杂目标结构条件下的有效性与稳定性。

相较于飞机目标，舰船目标通常呈现为细长矩形或条带状结构，整体几何结构上更规则、稳定，类内结构差异相对较小，不同舰船之间的外观相似性较强，但容易受复杂的近岸场景干扰<sup>[37]</sup>。主要考虑到不同成像分辨率、观测条件及场景复杂度带来的显著图像分布差异，本文选用SSDD (SAR Ship Detection Dataset)与HRSID (High-Resolution SAR Images Dataset)两个舰船目标数据集设计异源实验。二者的差异主要来源于采集平台、成像分辨率和采样区域的不同，这些因素会影响SAR图像的灰度分布、目标尺度、边缘细节及背景杂波特征，从而形成一定的跨域分布偏移。表2展示了实验数据集的数量分布。其中，SADD数据集仅展示了8:2划分时的数量分布，具体划分方式详见图4。

### 3.2.1 SAR飞机目标数据集

SADD是2022年公开的SAR飞机目标检测数据集<sup>[16]</sup>，共有2966张不重叠的像素切片，包含7835个结构、轮廓及主要部件清晰的飞机目标。数据集中

的飞机目标具有多尺度特性，且存在大量小尺寸目标。数据集背景环境相对复杂，涵盖机场跑道、停机坪、民用航空机场等多种场景，负样本数量较多，适合用于检验模型的泛化性能。

SADD数据集实验数据划分方式如图4所示，首先，将SADD中的全部样本等分为两个互斥子集，分别记为Subset-A和Subset-B。随后，在每个子集内部按照多种训练/测试比例随机划分，得到Train-A, Test-A以及Train-B, Test-B。其中，Train-A和Train-B分别用于训练源模型A和源模型B。Test-A和Test-B分别表示对应子集下的测试集。最终测试集Test-M由Test-A与Test-B的并集构成，即 $Test-M = Test-A \cup Test-B$ 。由于Subset-A与Subset-B严格互斥，且各子集内部训练集与测试集无重叠，因此可有效避免训练样本与测试样本之间的交叉。

### 3.2.2 SAR舰船目标数据集

SSDD是2017年公开发布的SAR舰船检测基准数据集，本文中采用的是更新后的BBBox版本<sup>[6]</sup>，共有1160幅图像，包含2456个舰船实例。目标普遍呈

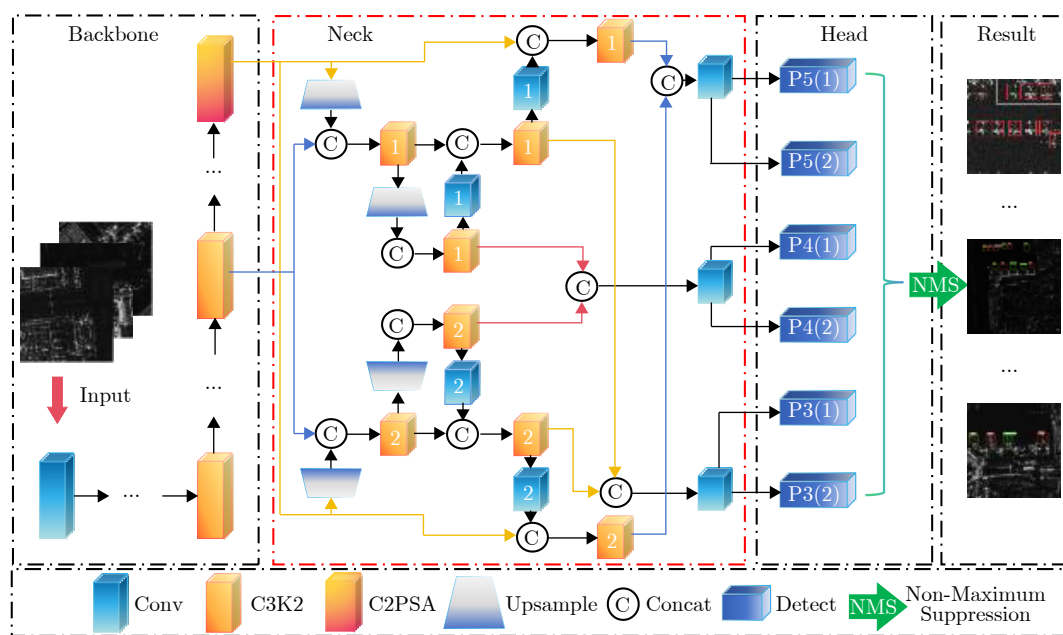


图 3 设计的合并模型网络结构

Fig. 3 Network architecture of the designed merged model

表 2 实验数据集数量分布

Tab. 2 Number distribution of experimental datasets

目标类型	数据集	训练样本数/张	验证样本数/张	平均目标数/(个/张)
飞机	SADD	2373	593	2.64
舰船	SSDD	928	232	2.12
舰船	HRSID	3643	1961	3.02

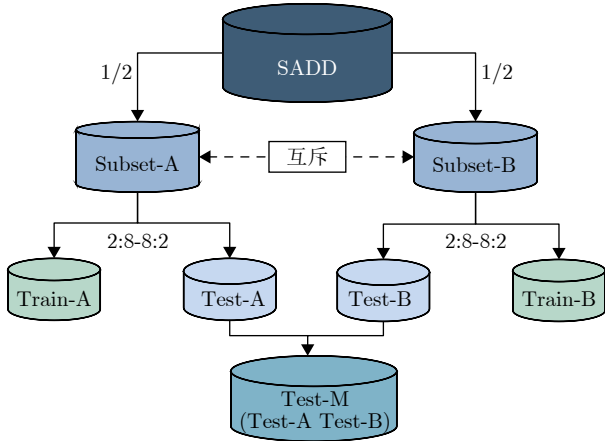


图4 SADD目标检测实验数据集划分方式

Fig. 4 Dataset partitioning strategy for SADD target detection experiments

现小尺寸特征,且港口区域存在明显杂波与密集停泊现象,具有小目标占比高、背景复杂等特点。

HRSID是2020年公开发布的高分辨率SAR舰船检测与实例分割数据集<sup>[37]</sup>,共有5604张样本,包含16951艘舰船目标,实验按照原文中的65%训练集与35%测试集进行划分。数据集中包含大量近岸港口密集舰船、小尺度弱散射目标以及复杂背景干扰。

为系统评估所提出模型合并方法在异源场景下的泛化能力,本文在SSDD与HRSID两个具有显著差异的SAR舰船检测数据集之间开展交叉验证实验。这两个数据集在空间分辨率、场景复杂度以及目标尺度分布方面存在明显差异,为跨域检测提供了具有挑战性的评测环境。具体而言,采用相同的网络结构与训练策略,分别在SSDD与HRSID上独立训练两个源检测模型,基于第2节所提模型合并方法构建一个合并模型。

在评测阶段,实验采用双向交叉测试,即分别在SSDD-test与HRSID-test上评估所有模型性能,得到同域与跨域检测结果。该评估方式能够显式刻画模型在分布偏移条件下的泛化鲁棒性,而非仅反映其在混合测试分布上的平均性能。该评测设置能够公平、全面地比较单源模型与合并模型在异源条件下的表现差异,从而验证所提出方法在缓解跨域性能退化方面的有效性。

### 3.3 实验评价指标

在深度学习目标检测领域,对目标检测模型的主要评价指标包括交并比(IoU)、精确率(Precision)、召回率(Recall)等。计算公式为

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (11)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

其中,A和B表示不同检测框的框定区域,IoU取值越接近1,表明两个框重叠度越高,定位越精准。TP(True Positives)指正确检测的目标,FP(False Positives)指误将背景或其他物体识别为目标的情况,即虚警,FN(False Negatives)指真实目标未被检测出的漏检情况,即漏警。精确率反映预测结果的准确性,召回率反映预测结果的完整性。

综合评估指标主要包括平均精度(AP)及均值平均精度(mAP)等。平均精度AP是单个类别的综合评价指标,计为该类别Precision-Recall曲线下的面积,值越高说明该类别检测性能越稳定,是对Precision与Recall两个指标的综合评估。mAP是所有类别的AP平均值,计算公式为

$$\text{AP} = \int_0^1 P(R) dR \quad (14)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (15)$$

mAP50指的是IoU阈值为0.5时,模型的平均精度,mAP50-95衡量的是IoU阈值从0.5到0.95范围内的平均精度。

所有实验均选取Precision, Recall, mAP50及mAP50-95这4个检测指标以综合衡量模型性能。评估指标能够全面比较单源模型、合并模型在同域与异域场景下的检测效果,为验证所提出方法的有效性与泛化优势提供可靠依据。

通常采用参数量(Params)、浮点运算量(FLOPs)和每秒处理帧数(FPS)作为衡量模型复杂度与推理效率的统计指标。其中,参数量表示模型中可学习参数的总数,能够反映模型规模及存储开销。浮点运算量表示模型完成一次前向传播所需的浮点运算次数,用于衡量模型的理论计算复杂度。每秒处理帧数表示模型在单位时间内能够处理的图像帧数,用于衡量模型的推理速度与实时性能。一般而言,参数量越小、FLOPs越低,说明模型越轻量,计算开销越小。FPS越高,则表明模型具有更好的实时处理能力。

### 3.4 对比实验研究

#### 3.4.1 飞机目标检测对比实验

为验证所提基于模型合并的人造目标检测方法的有效性与稳定性,根据3.2节中的实验设置,在

同源数据集上训练得到3组共6个源模型，在测试集Test-M上进行了测试，同组源模型测试结果取均值。表3对比了在训练-测试集8:2的常规划分比例下，3个版本模型合并前后的性能，v5-avg指的是在Subset-A与Subset-B训练出的两组v5源模型在相同测试集Test-M测试结果的均值，v5-Merging指的是两组v5源模型构造的合并模型，其余命名依此类推。评估指标为Precision, Recall, mAP50和mAP50-95。集成学习与模型合并核心思想都旨在利用多个源模型的已有知识，但二者融合层次存在差别。集成学习保留多个完整模型，在推理阶段对并行输出结果进行择优或后处理，属于结果级融合。本文所提模型合并方法则在输出前整合多模型参数化知识，对中间特征进行对齐与重组，形成一个统一模型，更侧重内部表征能力融合和效率提升。考虑到目前SAR目标检测领域还没有成熟的模型合并对比方法，因此，本文选取了3种目标检测领域中经典的集成学习方法作为对比实验，分别是NMS集成<sup>[38]</sup>、投票法(Voting)<sup>[39]</sup>和加权框融合(Weighted Boxes Fusion, WBF)<sup>[40]</sup>。

由表3可见，3组模型中模型合并策略均取得了最优的综合检测性能。其中，v5, v8和v11的mAP50分别达到0.968, 0.971和0.968，mAP50-95分别达到0.593, 0.688和0.656，整体均优于对应的源模型平均结果。尤其在v8与v11上，合并方法在保持较高

Precision的同时，进一步提升了mAP50-95，表明其不仅改善了检测召回，还提升了定位质量。相比之下，Voting和WBF在v5上虽提高了Recall，但Precision明显下降，最终mAP50与mAP50-95反而低于源模型平均值，说明简单投票、检测框级集成并不必然带来性能增益，其效果受源模型之间的预测差异、互补程度以及所采用集成方式是否适合当前任务的共同影响。已有研究指出，集成性能并不随模型数量或融合操作的引入而必然提升<sup>[41, 42]</sup>。Voting和WBF等集成学习方法是否有效，主要取决于不同检测模型的预测是否具有互补性。若各模型在误检、漏检或定位偏差上高度相似，集成的方法难以实现纠错，反而可能引入更多冗余或低质量检测结果。投票法的收益受个体精度与分类器间依赖关系约束，而WBF也依赖预测框匹配质量和置信度一致性，当条件不满足时可能出现Recall上升而Precision受损的现象。

进一步分析，合并的优势主要体现在两方面：首先，合并并非仅在后处理阶段对预测结果做简单平均或投票，而是在模型层面实现参数或表示空间的有效整合，因此能够在保留源模型判别能力的同时减少冗余冲突，减弱Voting和WBF等方法中常见的误检累积问题。其次，合并3组实验中均保持了更稳定的Precision-Recall平衡，说明该策略能够更充分地利用不同源模型的有效信息，而不是单纯以增加候选预测为代价换取召回提升。就本部分实验而言，v5的Voting和WBF之所以未取得效果，本质上是因为这两类方法更多体现为“结果层互补”，当源模型输出相似、错误模式相关时，新增预测难以形成有效纠错，反而会放大低质量框或错误决策。而合并策略则表现出更强的鲁棒性和更稳定的收益，说明其更适合作为此类任务的策略。

所提合并模型在保留多模型知识复用能力的同时，将多个模型的信息有效整合到单一网络结构中，从而避免了在存储与计算上的重复开销。该方法在参数量更小、计算复杂度更低、推理速度更快的条件下，仍能够获得优于集成学习方法的检测性能。

表4给出了不同版本下，源模型、3种集成学习方法以及合并模型在参数量(Params)、浮点运算量(FLOPs)和推理速度(FPS)指标的统计结果。整体来看，所提合并模型在模型复杂度和推理效率之间取得了较好的平衡，与源模型相比，由于在原结构上引入了新的双分支特征金字塔等结构。因此，模型参数量规模和计算开销均有所增加，而推理速度则相应下降，但就实时性而言，仍保持在可接受的范围内。

表 3 源模型、集成学习方法及合并模型在 SADD 数据集上的检测结果

Tab. 3 Detection results of source models, ensemble learning methods, and the merged model on the SADD dataset

模型版本	方法	Precision	Recall	mAP50	mAP50-95
v5	avg(基线)	<b>0.933</b>	0.900	0.954	0.576
	NMS集成	0.928	0.922	0.967	0.592
	Voting	0.869	0.932	0.946	0.567
	WBF	0.875	0.931	0.947	0.570
	Merging	0.918	<b>0.939</b>	<b>0.968</b>	<b>0.593</b>
v8	avg(基线)	0.925	0.902	0.955	0.609
	NMS集成	<b>0.953</b>	0.901	0.960	0.656
	Voting	0.872	0.947	0.958	0.661
	WBF	0.879	0.953	0.957	0.663
	Merging	0.938	<b>0.955</b>	<b>0.971</b>	<b>0.688</b>
v11	avg(基线)	0.937	0.895	0.953	0.604
	NMS集成	0.892	0.938	0.960	0.630
	Voting	0.885	0.949	0.958	0.638
	WBF	0.884	<b>0.945</b>	0.957	0.637
	Merging	<b>0.950</b>	0.906	<b>0.968</b>	<b>0.656</b>

注：表中加粗数值表示最优结果。

表4 模型参数量、复杂度与推理效率统计

Tab. 4 Model parameter, complexity, and inference efficiency statistics

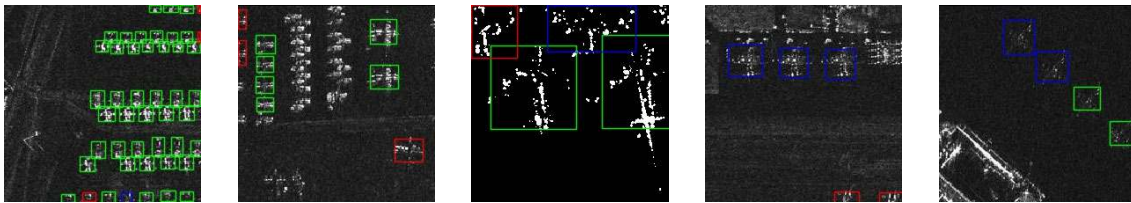
模型版本	方法	Params (M)	$\Delta$ Params (M)	FLOPs (G)	$\Delta$ FLOPs (G)	FPS (fps)
v5	avg(基线)	20.8	—	47.9	—	122.1
	NMS集成	41.6	+20.8	95.8	+47.9	42.6
	Voting	41.6	+20.8	95.8	+47.9	41.5
	WBF	41.6	+20.8	95.8	+47.9	41.5
	Merging	29.5	+8.7	64.7	+16.8	83.9
v8	avg(基线)	23.2	—	67.4	—	91.2
	NMS集成	46.4	+23.2	134.8	+67.4	25.1
	Voting	46.4	+23.2	134.8	+67.4	24.9
	WBF	46.4	+23.2	134.8	+67.4	29.0
	Merging	34.6	+11.4	96.4	+29.0	63.9
v11	avg(基线)	20.0	—	67.6	—	76.4
	NMS集成	40.0	+20.0	135.2	+67.6	26.3
	Voting	40.0	+20.0	135.2	+67.6	31.6
	WBF	40.0	+20.0	135.2	+67.6	25.7
	Merging	29.7	+9.7	96.0	+28.4	53.2

与NMS集成、Voting和WBF 3种集成学习方法相比, 合并模型在资源消耗和推理效率方面均表

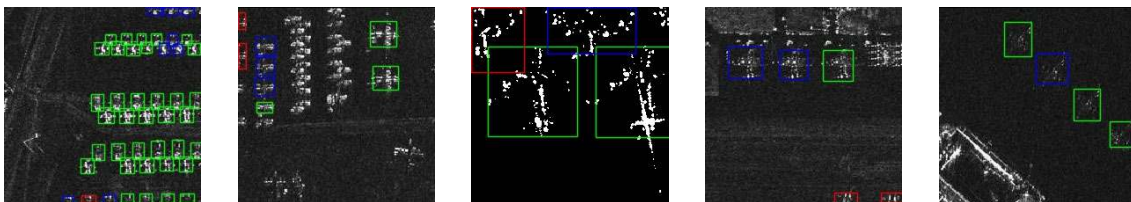
现出明显优势。实验采用两模型进行集成, 因此集成学习方法的Params和FLOPs均为源模型的两倍。以v5为例, 3种集成方法的Params和FLOPs均为41.6 M和95.8 G, 而合并模型分别仅为29.5 M和64.7 G, 分别降低约29.1%和32.5%。该版本下的合并模型的FPS则达到83.9, 显著高于集成方法。类似地, 在v8和v11版本中, 合并模型的Params分别由集成方法的46.4 M和40.0 M降低至34.6 M和29.7 M, FLOPs由134.8 G和135.2 G降低至96.4 G和96.0 G, 而FPS则分别为63.9和53.2, 也均明显优于集成方法。由此可见, 集成策略虽然能够有效融合多个模型的预测结果, 但同时需要保留并调用多个完整模型, 且在模型推理完毕后还需对结果进行繁琐的后处理, 导致参数冗余明显、计算成本较高且推理速度受限。

所提合并模型在保留多模型知识复用能力的同时, 将多个模型的信息有效整合到单一网络结构中, 从而避免了在存储与计算上的重复开销。该方法在参数量更小、计算复杂度更低、推理速度更快的条件下, 仍能够获得优于集成学习方法的检测性能。

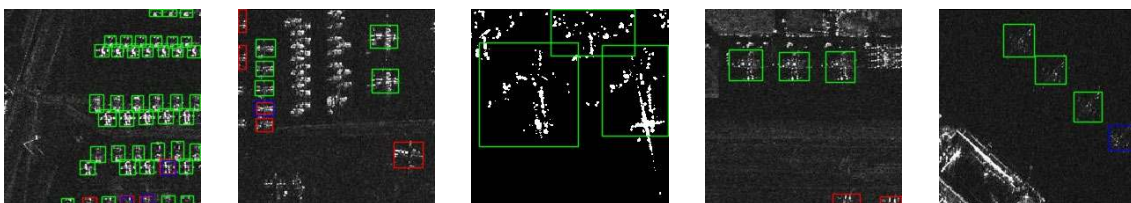
图5展示了相同条件下YOLOv8的两个源模型、3种集成学习方法以及构造的合并模型在SADD数据集上的部分检测结果。图中绿色框为正确检测的



(a) YOLOv8源模型A的检测结果  
(a) Detection results of YOLOv8 source model A



(b) YOLOv8源模型B的检测结果  
(b) Detection results of YOLOv8 source model B



(c) NMS集成的检测结果  
(c) Detection results of NMS ensemble

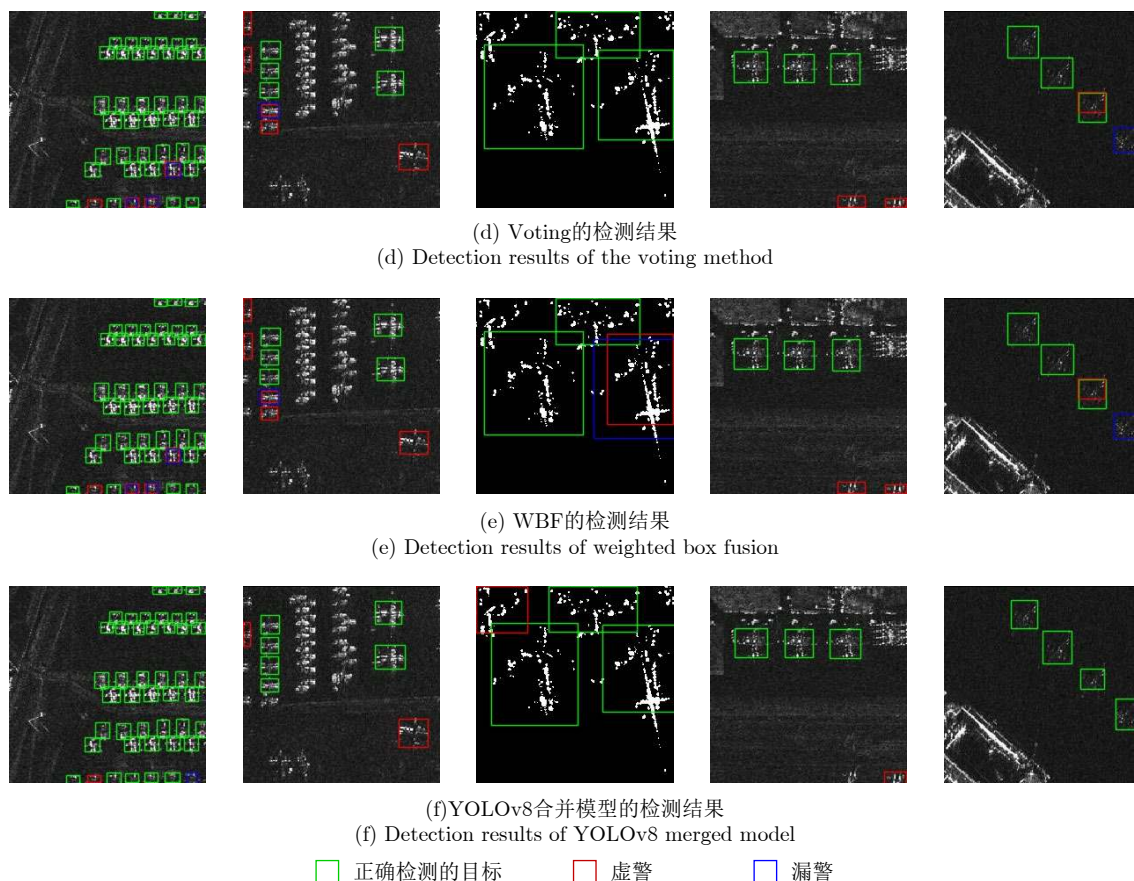


图 5 源模型、集成学习方法与合并模型在SADD数据集上的检测结果对比

Fig. 5 Comparison of detection results among source models, ensemble learning methods, and the merged model on the SADD dataset

目标，红色框为虚警，蓝色框为漏警。可视化结果显示，所提方法有效降低了漏警和虚警，相较源模型提升明显。与NMS集成、Voting及WBF等集成学习方法相比，合并模型在SADD数据集上的检测结果更优。集成方法容易引入较多重叠的冗余框，造成误检的产生，而合并模型的冗余预测更少，结果更准确。同时，合并模型在部分复杂场景下能有效补充源模型的漏检目标，在提升检测完整性的同时，保持了较低的漏警，体现出较好的平衡。

为验证所提方法在不同数据规模下的有效性与稳定性，基于v8架构，在SADD数据集上设置多组训练-测试集划分比例(2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2)进行对比实验。分别训练源模型并构建对应的合并模型，在统一的测试集上进行综合性能评估。同时还加入对不同集成学习方法在相同条件下的对比。评估指标包括Precision, Recall, mAP50, mAP50-95。

图6给出了在不同训练-测试集划分比例下，各方法相对于源模型在Precision, Recall, mAP50及mAP50-95指标上的差值变化情况。从整体结果来看，所提的合并方法在大多数数据划分条件下均表

现出更为稳定且显著的性能增益，尤其在mAP50与mAP50-95指标上优势更为突出。

从mAP50与mAP50-95指标，图6(c)与图6(d)可以看出，合并方法在所有训练-测试划分比例下均取得了正向提升，且整体增幅高于WBF, Voting及NMS集成等方法。其中，mAP50-95指标的提升幅度尤为显著，表明该方法在更严格的IoU评价标准下依然能够有效提升检测性能。这说明所提出的策略不仅提升了目标的定位精度与边界框质量，同时在不同的样本规模下具有更强的鲁棒性。相比之下，其他方法虽然在部分比例下也能带来一定提升，但整体增益较为有限。

从Recall指标即图6(b)来看，合并方法在各划分比例下均保持稳定提升，且在低训练样本比例(如2:8, 3:7)条件下优势更为明显。这表明该方法能够有效整合来自不同模型的判别信息，在样本不足的情况下仍能提升目标覆盖能力，降低漏检率。与此同时，Voting方法在部分比例下也表现出较高的Recall，但其在其他指标上的不稳定性限制了整体性能提升。

从图6(a)可见，各方法在Precision上存在不同

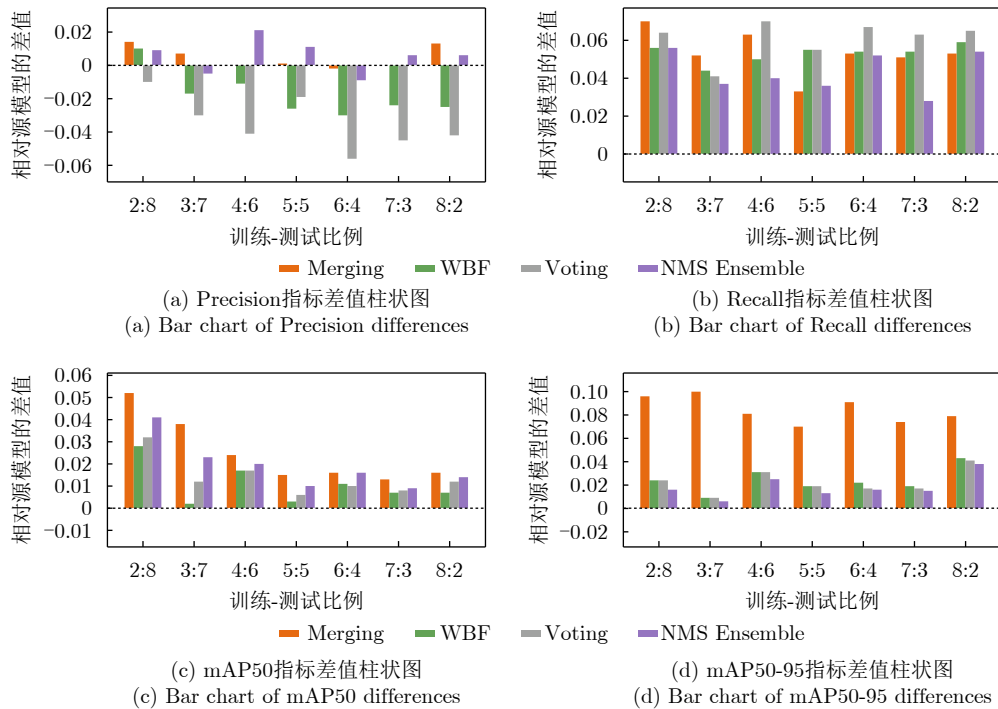


图6 不同比例训练-测试集下的性能差值柱状图(相对源模型)

Fig. 6 Bar chart of performance differences under different train-test split ratios (relative to the source model)

程度波动,部分方法甚至体现出负增益,其主要原因在于多模型的集成或合并过程会引入额外候选框,导致误检率的增加。其中,Voting与WBF方法更容易保留位置不一致或质量较低的预测,因此Precision下降更明显。NMS能够抑制部分冗余框,下降幅度相对较小。相比之下,Merging方法整体波动较小,说明其通过特征层融合与权重矩阵的重加权,在提升Recall的同时较好地控制了误检。

综合Precision与Recall的变化可以看出,Merging方法在提升Recall的同时较好地控制了Precision的下降幅度,体现出更优的平衡能力。这种现象表明,该方法并非单纯依赖增加候选框数量来提升召回,而是通过更有效的特征融合与响应重加权,实现了对有效目标信息的增强与冗余信息的抑制。

进一步结合4项指标可以得出,所提出的融合方法在不同训练数据规模下均表现出稳定的综合性能优势,尤其在高IoU阈值(mAP50-95)条件下表现更为突出。这表明该方法能够在统一特征空间内实现更有效的异源特征对齐,有效增强模型对目标结构与位置信息的刻画能力。同时,该优势在不同数据划分比例下均保持一致,验证了所提方法良好的泛化能力与鲁棒性。

### 3.4.2 舰船目标检测对比实验

为验证所提出合并方法在异源数据条件下的泛

化能力,按照第3.2节中的实验设置,分别在SSDD和HRSID两个数据集上,对YOLOv5,YOLOv8和YOLOv11 3种模型进行独立训练,从而得到3组共6个源模型。在此基础上,通过所提出的合并策略构建对应的合并模型,并在SSDD-test与HRSID-test上进行交叉验证,以系统评估模型在同域与跨域场景下的检测性能。实验采用Precision,Recall,mAP50和mAP50-95作为评价指标,从检测精度、召回能力及整体性能等多个维度对不同模型进行对比分析。

表5给出了在SSDD-test与HRSID-test上的交叉验证结果。可以观察到,单源训练模型在跨数据集测试时普遍存在明显的性能退化,例如在SSDD-test上,基于HRSID训练的模型在mAP50与mAP50-95指标上显著低于对应的model-SSDD,而在HRSID-test上,model-SSDD亦表现出类似的泛化不足。这表明不同SAR舰船数据集在成像条件与目标分布方面存在较大的域差异,单一数据源所构建的模型难以支撑稳定的跨域检测性能,泛化性能不强。相比之下,采用合并策略得到的model-Merge在两种测试集及3个不同版本的检测模型上均表现出更加稳定且均衡的性能,其mAP50-95在多数设置下明显优于跨域测试的单源模型,并在部分情况下接近甚至略优于同域最优模型。Precision与Recall的结果来看,model-Merge能够在保持较高检测精度的同时兼顾召回能力,未表现出通过牺

牲某一指标换取整体提升的现象。需要指出的是，合并模型并非在所有情况下都显著超越单源最优模型，其优势主要体现在缓解跨域性能退化、提升多源数据条件下的检测稳定性与泛化能力，这一特性使其在实际多数据源SAR船舶检测场景中具有更高的应用价值。

图7和图8分别展示了YOLOv8源模型和合并模型在SSDD和HRSID数据集上的部分检测结果，图

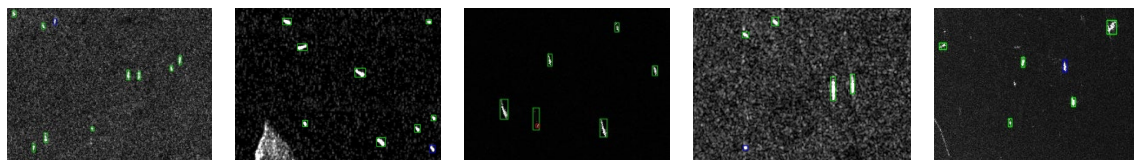
中绿色框为正确检测的目标，红色框为虚警，蓝色框为漏警。可以清楚地观察到，各源模型在源数据集上表现较好，而跨数据集的泛化能力较弱，出现了较多的漏警和虚警。合并模型在两个数据集上的表现均较好。

为直观分析异源模型合并前后的特征响应的变化情况，对模型特征融合前后层的特征热力图响应可视化。图9展示了SSDD-HRSID异源实验中真

表 5 SSDD与HRSID数据集交叉验证实验  
Tab. 5 Cross-validation experiment between SSDD and HRSID datasets

模型版本	模型来源	SSDD-test				HRSID-test			
		Precision	Recall	mAP50	mAP50-95	Precision	Recall	mAP50	mAP50-95
v5	SSDD	<b>0.968</b>	0.907	0.978	<b>0.734</b>	0.838	0.575	0.706	0.481
	HRSID	0.720	0.560	0.640	0.353	0.899	0.794	0.891	0.634
	Merge	0.930	<b>0.951</b>	<b>0.983</b>	0.721	<b>0.901</b>	<b>0.810</b>	<b>0.902</b>	<b>0.662</b>
v8	SSDD	0.942	0.929	0.977	0.705	0.844	0.573	0.699	0.444
	HRSID	0.795	0.570	0.697	0.397	0.905	0.773	0.882	0.615
	Merge	<b>0.957</b>	<b>0.949</b>	<b>0.981</b>	<b>0.736</b>	<b>0.915</b>	<b>0.822</b>	<b>0.913</b>	<b>0.679</b>
v11	SSDD	<b>0.959</b>	0.912	0.976	<b>0.730</b>	0.811	0.575	0.711	0.482
	HRSID	0.819	0.547	0.672	0.365	0.904	0.802	0.899	0.655
	Merge	0.929	<b>0.951</b>	<b>0.982</b>	0.718	<b>0.908</b>	<b>0.809</b>	<b>0.901</b>	<b>0.670</b>

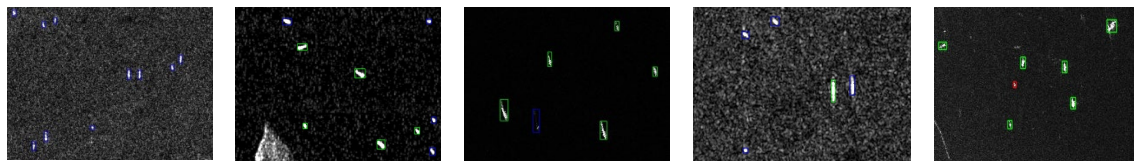
注：表中加粗数值表示最优结果。



□ 正确检测的目标 □ 虚警 □ 漏警

(a) SSDD源模型在SSDD数据集上的检测结果

(a) Detection results of the SSDD source model on the SSDD dataset



□ 正确检测的目标 □ 虚警 □ 漏警

(b) HRSID源模型在SSDD数据集上的检测结果

(b) Detection results of the HRSID source model on the SSDD dataset



□ 正确检测的目标 □ 虚警 □ 漏警

(c) 合并模型在SSDD数据集上的检测结果

(c) Detection results of the merged model on the SSDD dataset

图 7 源模型与合并模型在SSDD数据集上的检测结果对比

Fig. 7 Comparison of detection results between the source models and the merged model on the SSDD dataset

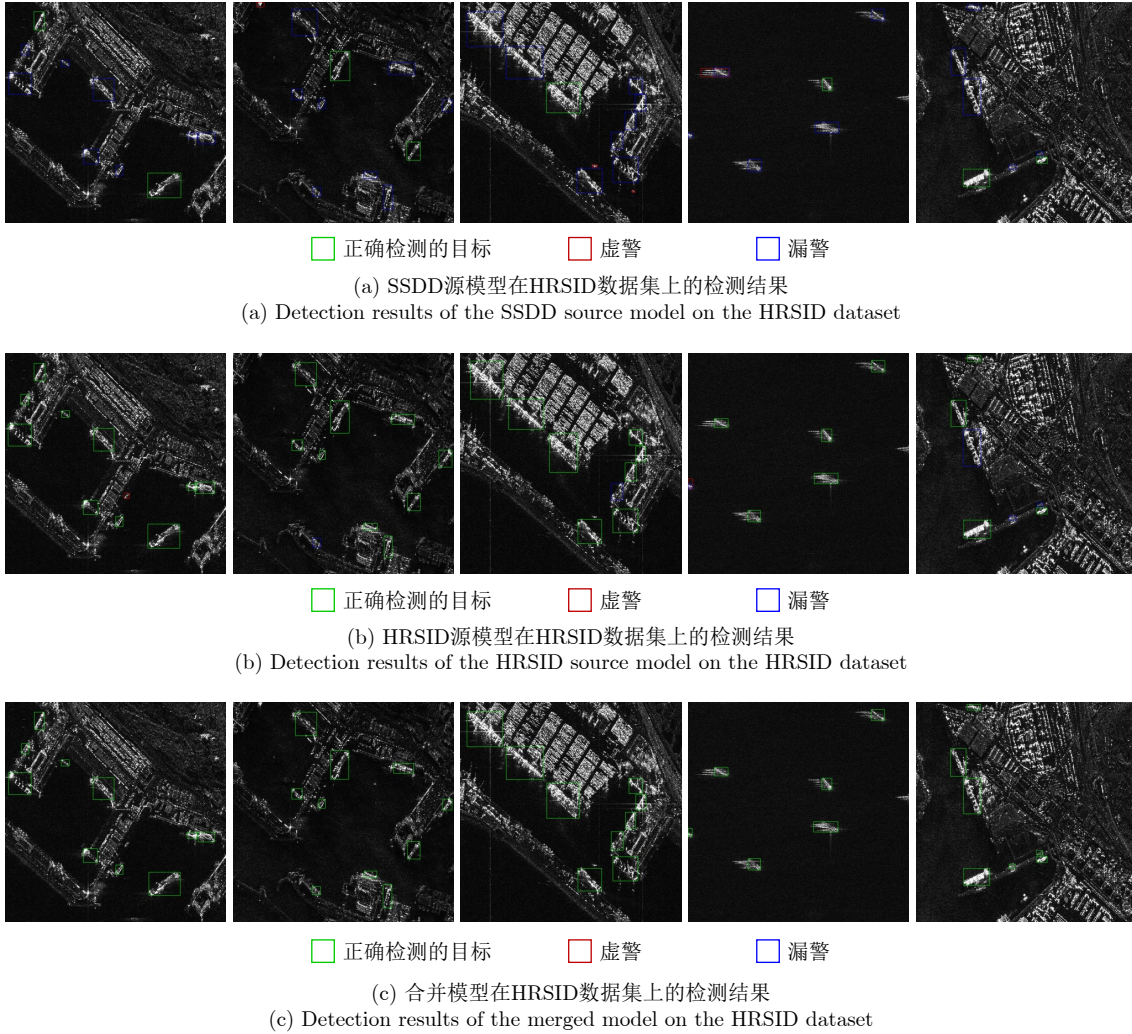


图8 源模型与合并模型在HRSID数据集上的检测结果对比

Fig. 8 Comparison of detection results between the source models and the merged model on the HRSID dataset

值标注与融合层前后特征响应的可视化结果。前两列为SSDD样本，后两列为HRSID样本。其中，图9(a)为输入SAR图像及其真值标注，绿色框表示舰船目标位置。图9(b)和图9(c)分别为SSDD源模型分支与HRSID源模型分支在融合前的特征响应。图9(d)为经融合后的特征响应。热力图通过结合中间特征激活与检测输出梯度，反映不同空间区域对检测结果的判别性贡献，颜色越亮表示该区域对模型检测输出的正向贡献越强。对比图9(a)—图9(c)可以看出，两个源模型分支在不同SAR场景下的响应分布存在差异。对于海面背景较简单的样本，源分支能够在部分舰船目标区域形成较明显响应。而在港口、近岸等复杂背景样本中，码头边缘、岸线结构和强散射区域也会产生一定激活，说明异源数据间的场景复杂度和目标尺度差异会影响源模型的关注区域。进一步对比图9(b)—图9(d)可以发现，融合后特征响应在多数舰船目标区域更加集中，并

与真值框位置具有更好的一致性。这表明，融合层能够通过跨通道线性重组，对两个源分支的特征进行重新加权，使融合特征更有利于突出舰船目标区域。

为验证所提方法在SAR人造目标检测任务中的整体先进性，本文在SSDD与HRSID舰船目标检测数据集上进行全量数据对比实验。具体而言，将SSDD训练集与HRSID训练集合并构成联合训练集，并分别在SSDD-test和HRSID-test上测试。对比方法选取近年来目标检测领域具有代表性的模型，包括单阶段检测方法FCOS<sup>[19]</sup>、两阶段检测方法Faster R-CNN<sup>[20]</sup>、基于Transformer架构的Swin-Transformer<sup>[22]</sup>。此外，为验证本文方法的性能提升是否仅来源于训练数据增加，本文额外设置一组YOLOv8(Union)，即在与合并模型相同的联合训练数据上训练单一YOLOv8模型，并与本文YOLOv8合并模型进行对比。

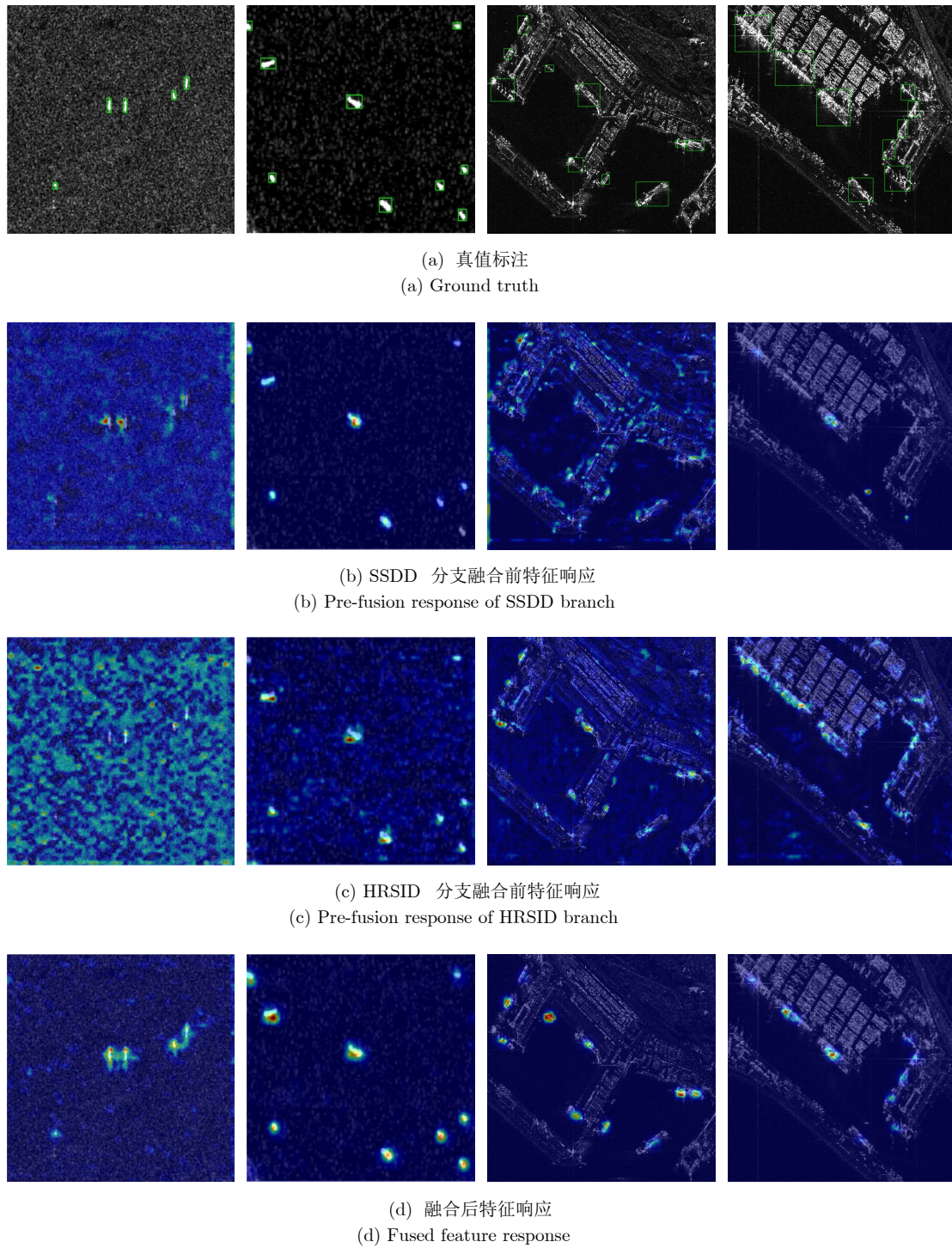


图 9 SSDD-HRSID异源实验中真值标注与融合前后特征响应对比

Fig. 9 Comparison of ground-truth annotations and feature responses before and after fusion in the SSDD-HRSID heterogeneous experiment

对于FCOS, Faster R-CNN, Swin-Transformer和YOLOv8 (Union), 均采用相同的训练轮次、数据划分和评价指标等进行训练与评估。本文选取YOLOv8合并模型作为所提方法的代表结果。所有方法均采用Precision, Recall, mAP50和mAP50-95作为评价指标, 以综合评估模型的检测精度、召回能力和高IoU阈值下的定位性能。

表6给出了全量训练数据条件下不同方法在SSDD-test和HRSID-test上的检测结果。可以看出, 各类代表性检测方法在SAR舰船目标检测任务中均取得了一定效果, 但不同方法在检测完整性和定位精度上存在差异。相比FCOS, Faster R-CNN和Swin-Transformer, 本文YOLOv8 (Merge)在SSDD-test上4项指标均取得最优结果, Precision, Recall,

表6 代表性检测方法全量数据对比实验  
Tab. 6 Full-data comparison of representative detection methods

方法	SSDD-test				HRSID-test			
	Precision	Recall	mAP50	mAP50-95	Precision	Recall	mAP50	mAP50-95
FCOS	0.843	0.923	0.886	0.598	0.843	0.816	0.781	0.573
Faster R-CNN	0.890	0.938	0.955	0.684	0.794	<b>0.833</b>	0.866	0.614
Swin-Transformer	0.889	0.927	0.956	0.680	0.843	0.831	0.850	0.621
YOLOv8(Union)	0.927	0.894	0.954	0.706	0.895	0.766	0.881	0.676
YOLOv8 (Merge)	<b>0.957</b>	<b>0.949</b>	<b>0.981</b>	<b>0.736</b>	<b>0.915</b>	0.822	<b>0.913</b>	<b>0.679</b>

注：表中加粗数值表示最优结果。

mAP50和mAP50-95分别达到0.957, 0.949, 0.981和0.736。在HRSID-test上, 本文方法的Precision, mAP50和mAP50-95分别达到0.915, 0.913和0.679, 均为最优结果。虽然其Recall略低于Faster R-CNN和Swin-Transformer, 但整体mAP指标更优, 说明合并模型在保持较好目标检出能力的同时, 具有更高的预测准确性和边界框定位质量。

为排除训练数据规模增加对性能提升的影响, 本文进一步设置YOLOv8 (Union)作为控制实验。该模型使用与合并模型相同的SSDD与HRSID联合训练数据, 但不进行模型合并。实验结果表明, 在训练数据范围一致条件下, YOLOv8 (Merge)仍优于YOLOv8 (Union)。其中, 在SSDD-test上, YOLOv8 (Merge)相较YOLOv8 (Union)的Precision, Recall, mAP50和mAP50-95分别提升3.0, 5.5, 2.7和3.0个百分点; 在HRSID-test上分别提升2.0, 5.6, 3.2和0.3个百分点。该结果说明, 本文方法的性能增益并非单纯来自训练样本数量增加, 而主要得益于对不同源模型参数化知识的结构化复用与互补融合。

表7给出了代表性检测方法的参数量与计算复杂度统计结果。可以看出, YOLOv8 (Merge)的FLOPs和Params分别为96.4 G和34.6 M, 相较YOLOv8 (Union)有所增加, 主要原因在于合并模型引入了双分支特征金字塔和并行检测头结构, 用于保留并融合不同源模型的检测能力。尽管如此,

表7 代表性检测方法计算复杂度与参数量对比

Tab. 7 Comparison of FLOPs and Params among representative detection methods

方法	FLOPs(G)	Params(M)
FCOS	88.6	32.1
Faster R-CNN	112.8	41.4
Swin-Transformer	94.3	27.5
YOLOv8 (Union)	67.4	23.2
YOLOv8 (Merge)	96.4	34.6

结合表6可知, 该额外计算开销带来了更稳定的Recall, mAP50和mAP50-95提升。同时, 与Faster R-CNN等检测方法相比, 本文方法在参数量相当的情况下, 计算复杂度仍保持较优水平。

总体来看, 本文方法相较于对比方法与联合训练单模型, 均表现出较好的检测完整性、定位精度和跨数据集稳定性, 进一步验证了所提YOLO模型合并方法在SAR人造目标检测任务中的有效性与先进性。

## 4 结语

本文针对SAR图像人造目标检测模型的参数化知识未充分利用, 导致模型迁移应用难的问题, 提出了一种基于YOLO模型合并的SAR图像人造目标检测方法。该方法通过在结构层面对同构目标检测模型进行合并, 在共享预训练骨干网络的基础上, 融合不同源模型在特征提取与目标定位方面的互补能力, 从而高效构建具备多模型检测能力的合并模型。实验结果表明, 在SADD, SSDD和HRSID多个SAR数据集上, 提出的模型合并方法在多种YOLO网络架构和不同数据规模条件下均能够稳定提升检测性能, 尤其可在高IoU阈值和较少样本场景下增强模型的鲁棒性与泛化能力。此外, 合并模型能够有效降低漏检率, 并在一定程度上缓解单一训练数据带来的类别覆盖不足问题。所提方法为SAR图像目标检测中参数化知识的高效复用提供了一种可行的技术路径, 对降低模型训练成本、提升检测性能具有一定的工程应用价值。未来工作将进一步探索异构模型之间的合并策略, 拓展其在车辆等更多典型SAR人造目标检测任务中的适用性, 并研究模型合并方法与SAR特定物理成像机理的结合方式, 以深入挖掘其在SAR图像智能解译和复杂场景目标检测中的应用潜力。

**利益冲突** 所有作者均声明不存在利益冲突

**Conflict of Interests** The authors declare that there is no conflict of interests

## 参 考 文 献

- [1] 陈思伟. 成像雷达极化旋转域解译: 理论与应用[M]. 北京: 科学出版社, 2024: 1–251.  
CHEN Siwei. Imaging Radar Polarimetric Rotation Domain Interpretation: Theory and Application[M]. Beijing: Science Press, 2024: 1–251.(in Chinese).
- [2] CHEN Siwei, WANG Xuesong, XIAO Shunping, *et al.* Target Scattering Mechanism in Polarimetric Synthetic Aperture Radar: Interpretation and Application[M]. Singapore: Springer, 2018: 1–225. doi: [10.1007/978-981-10-7269-7](https://doi.org/10.1007/978-981-10-7269-7).
- [3] LI Haoliang, LIU Shenwen, and CHEN Siwei. PolSAR ship characterization and robust detection at different grazing angles with polarimetric roll-invariant features[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5225818. doi: [10.1109/TGRS.2024.3474702](https://doi.org/10.1109/TGRS.2024.3474702).
- [4] 徐丰, 金亚秋. 微波视觉与SAR图像智能解译[J]. 雷达学报(中英文), 2024, 13(2): 285–306. doi: [10.12000/JR23225](https://doi.org/10.12000/JR23225).  
XU Feng and JIN Yaqiu. Microwave vision and intelligent perception of radar imagery[J]. *Journal of Radars*, 2024, 13(2): 285–306. doi: [10.12000/JR23225](https://doi.org/10.12000/JR23225).
- [5] 孙显, 王智睿, 孙元睿, 等. AIR-SARShip-1.0: 高分辨率SAR舰船检测数据集[J]. 雷达学报, 2019, 8(6): 852–862. doi: [10.12000/JR19097](https://doi.org/10.12000/JR19097).  
SUN Xian, WANG Zhirui, SUN Yuanrui, *et al.* AIR-SARShip-1.0: High-resolution SAR ship detection dataset[J]. *Journal of Radars*, 2019, 8(6): 852–862. doi: [10.12000/JR19097](https://doi.org/10.12000/JR19097).
- [6] ZHANG Tianwen, ZHANG Xiaoling, LI Jianwei, *et al.* SAR Ship Detection Dataset (SSDD): Official release and comprehensive data analysis[J]. *Remote Sensing*, 2021, 13(18): 3690. doi: [10.3390/rs13183690](https://doi.org/10.3390/rs13183690).
- [7] 王智睿, 康玉卓, 曾璇, 等. SAR-AIRcraft-1.0: 高分辨率SAR飞机检测识别数据集[J]. 雷达学报, 2023, 12(4): 906–922. doi: [10.12000/JR23043](https://doi.org/10.12000/JR23043).  
WANG Zhirui, KANG Yuzhuo, ZENG Xuan, *et al.* SAR-AIRcraft-1.0: High-resolution SAR aircraft detection and recognition dataset[J]. *Journal of Radars*, 2023, 12(4): 906–922. doi: [10.12000/JR23043](https://doi.org/10.12000/JR23043).
- [8] LANG Ping, FU Xiongjun, DONG Jian, *et al.* Recent advances in deep-learning-based SAR image target detection and recognition[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, 18: 6884–6915. doi: [10.1109/JSTARS.2025.3543531](https://doi.org/10.1109/JSTARS.2025.3543531).
- [9] GUO Yuchen, DU Lan, and LYU Guoxin. SAR target detection based on domain adaptive faster R-CNN with small training data size[J]. *Remote Sensing*, 2021, 13(21): 4202. doi: [10.3390/rs13214202](https://doi.org/10.3390/rs13214202).
- [10] SHANG Ronghua, WANG Jiaming, JIAO Licheng, *et al.* SAR targets classification based on deep memory convolution neural networks and transfer parameters[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(8): 2834–2846. doi: [10.1109/JSTARS.2018.2836909](https://doi.org/10.1109/JSTARS.2018.2836909).
- [11] WAGNER S A. SAR ATR by a combination of convolutional neural network and support vector machines[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2016, 52(6): 2861–2872. doi: [10.1109/TAES.2016.160061](https://doi.org/10.1109/TAES.2016.160061).
- [12] CHEN Junyi, SHEN Yanyun, LIANG Yinyu, *et al.* YOLO-SAD: An efficient SAR aircraft detection network[J]. *Applied Sciences*, 2024, 14(7): 3025. doi: [10.3390/app14073025](https://doi.org/10.3390/app14073025).
- [13] LI Haoliang and CHEN Siwei. General polarimetric correlation pattern: A visualization and characterization tool for target joint-domain scattering mechanisms investigation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2026, 64: 5200417. doi: [10.1109/TGRS.2025.3647123](https://doi.org/10.1109/TGRS.2025.3647123).
- [14] ZHOU Jie, XIAO Chao, PENG Bo, *et al.* DiffDet4SAR: Diffusion-based aircraft target detection network for SAR images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 4007905. doi: [10.1109/LGRS.2024.3386020](https://doi.org/10.1109/LGRS.2024.3386020).
- [15] DAI Linyu and CHEN Siwei. Context2Context: A zero-shot SAR image speckle filter[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2026, 64: 5204911. doi: [10.1109/TGRS.2026.3669951](https://doi.org/10.1109/TGRS.2026.3669951).
- [16] ZHANG Peng, XU Hao, TIAN Tian, *et al.* SEFEPNet: Scale expansion and feature enhancement pyramid network for SAR aircraft detection with small sample dataset[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 3365–3375. doi: [10.1109/JSTARS.2022.3169339](https://doi.org/10.1109/JSTARS.2022.3169339).
- [17] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: Unified, real-time object detection[C]. The 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 779–788. doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).
- [18] LIU Wei, ANGUELOV D, ERHAN D, *et al.* SSD: Single shot MultiBox detector[C]. 14th European Conference on Computer Vision – ECCV 2016, Amsterdam, The Netherlands, 2016: 21–37. doi: [10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [19] LI Zhonghua, HOU Biao, WU Zitong, *et al.* FCOSR: A simple anchor-free rotated detector for aerial object detection[J]. *Remote Sensing*, 2023, 15(23): 5499. doi: [10.3390/rs15235499](https://doi.org/10.3390/rs15235499).
- [20] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al.* Faster R-

- CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [21] CAI Zhaowei and VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]. The 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 6154–6162. doi: [10.1109/CVPR.2018.00644](https://doi.org/10.1109/CVPR.2018.00644).
- [22] LIU Ze, LIN Yutong, CAO Yue, *et al.* Swin transformer: Hierarchical vision transformer using shifted windows[C]. The 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 9992–10002. doi: [10.1109/ICCV48922.2021.00986](https://doi.org/10.1109/ICCV48922.2021.00986).
- [23] CARION N, MASSA F, SYNNAEVE G, *et al.* End-to-end object detection with transformers[C]. 16th European Conference on Computer Vision – ECCV 2020, Glasgow, UK, 2020: 213–229. doi: [10.1007/978-3-030-58452-8\\_13](https://doi.org/10.1007/978-3-030-58452-8_13).
- [24] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale[C]. 9th International Conference on Learning Representations, Virtual Event, Austria, 2021: 1–22.
- [25] 周正, 赵凌君, 何奇山, 等. 基于多源信息跨域学习的SAR图像目标检测技术研究进展与展望[J]. 雷达学报(中英文), 2026, 15(2): 387–408. doi: [10.12000/JR25205](https://doi.org/10.12000/JR25205).  
ZHOU Zheng, ZHAO Lingjun, HE Qishan, *et al.* Research progress and prospects of SAR image target detection based on multi-source information cross-domain learning[J]. *Journal of Radars*, 2026, 15(2): 387–408. doi: [10.12000/JR25205](https://doi.org/10.12000/JR25205).
- [26] ZHENG Hongling, SHEN Li, TANG Anke, *et al.* Learning from models beyond fine-tuning[J]. *Nature Machine Intelligence*, 2025, 7(1): 6–17. doi: [10.1038/s42256-024-00961-0](https://doi.org/10.1038/s42256-024-00961-0).
- [27] YADAV P, RAFFEL C, MUQEETH M, *et al.* A survey on model MoErging: Recycling and routing among specialized experts for collaborative learning[J]. *Transactions on Machine Learning Research*, 2025: 1–32. (期刊无卷号).
- [28] LU Jinliang, PANG Ziliang, XIAO Min, *et al.* Merge, ensemble, and cooperate! A survey on collaborative strategies in the era of large language models[EB/OL]. <http://arxiv.org/abs/2407.06089>, 2024.
- [29] YANG Enneng, SHEN Li, GUO Guibing, *et al.* Model merging in LLMs, MLLMs, and beyond: Methods, theories, applications, and opportunities[J]. *ACM Computing Surveys*, 2026, 58(8): 216. doi: [10.1145/3787849](https://doi.org/10.1145/3787849).
- [30] LI Weishi, PENG Yong, ZHANG Miao, *et al.* Deep model fusion: A survey[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2026, 37(5): 2008–2024. doi: [10.1109/TNNLS.2025.3628666](https://doi.org/10.1109/TNNLS.2025.3628666).
- [31] ILHARCO G, RIBEIRO M T, WORTSMAN M, *et al.* Editing models with task arithmetic[C]. The Eleventh International Conference on Learning Representations, Kigali, Rwanda, 2023: 1–31.
- [32] ALCOVER-COUSO R, SANMIGUEL J C, ESCUDERO-VIÑOLO M, *et al.* Layer-wise model merging for unsupervised domain adaptation in segmentation tasks[J]. *The Visual Computer*, 2025, 41(10): 7867–7882. doi: [10.1007/s00371-025-03843-7](https://doi.org/10.1007/s00371-025-03843-7).
- [33] AKIBA T, SHING M, TANG Yujin, *et al.* Evolutionary optimization of model merging recipes[J]. *Nature Machine Intelligence*, 2025, 7(2): 195–204. doi: [10.1038/s42256-024-00975-8](https://doi.org/10.1038/s42256-024-00975-8).
- [34] HE Kaiming, GIRSHICK R, and DOLLAR P. Rethinking ImageNet pre-training[C]. The 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 4917–4926. doi: [10.1109/ICCV.2019.00502](https://doi.org/10.1109/ICCV.2019.00502).
- [35] DOBRZYCKI A D, BERNARDOS A M, and CASAR J R. An analysis of layer-freezing strategies for enhanced transfer learning in YOLO architectures[J]. *Mathematics*, 2025, 13(15): 2539. doi: [10.3390/math13152539](https://doi.org/10.3390/math13152539).
- [36] 罗汝, 赵凌君, 何奇山, 等. SAR图像飞机目标智能检测识别技术研究进展与展望[J]. 雷达学报(中英文), 2024, 13(2): 307–330. doi: [10.12000/JR23056](https://doi.org/10.12000/JR23056).  
LUO Ru, ZHAO Lingjun, HE Qishan, *et al.* Intelligent technology for aircraft detection and recognition through SAR imagery: Advancements and prospects[J]. *Journal of Radars*, 2024, 13(2): 307–330. doi: [10.12000/JR23056](https://doi.org/10.12000/JR23056).
- [37] WEI Shunjun, ZENG Xiangfeng, QU Qizhe, *et al.* HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation[J]. *IEEE Access*, 2020, 8: 120234–120254. doi: [10.1109/ACCESS.2020.3005861](https://doi.org/10.1109/ACCESS.2020.3005861).
- [38] JOCHER G. Ultralytics YOLOv5[EB/OL]. <https://github.com/ultralytics/yolov5>, 2020.
- [39] CASADO-GARCÍA Á and HERAS J. Ensemble methods for object detection[C]. 24th European Conference on Artificial Intelligence, Santiago de Compostela, Spain, 2020: 2688–2695. doi: [10.3233/FAIA200407](https://doi.org/10.3233/FAIA200407).
- [40] SOLOVYEV R, WANG Weimin, and GABRUSEVA T. Weighted boxes fusion: Ensembling boxes from different object detection models[J]. *Image and Vision Computing*, 2021, 107: 104117. doi: [10.1016/j.imavis.2021.104117](https://doi.org/10.1016/j.imavis.2021.104117).
- [41] MATTEI P A and GARREAU D. Are ensembles getting better all the time?[J]. *Journal of Machine Learning Research*, 2025, 26(201): 1–46. (未查询到doi).
- [42] WOOD D, MU Tingting, WEBB A M, *et al.* A unified theory of diversity in ensemble learning[J]. *Journal of Machine Learning Research*, 2023, 24(359): 359. doi: [10.5555/3618408.3618767](https://doi.org/10.5555/3618408.3618767).

### 作者简介

何适存，硕士生，主要研究方向为深度学习、SAR图像目标检测识别。

陈思伟，教授，主要研究方向为极化雷达成像、目标识别、机器学习、电子对抗等。

王欣超，博士生，主要研究方向为深度学习、遥感基础模型。

(责任编辑：于青)