SARDet-100K: 迈向大规模 SAR 目标检测的开源数据集基准及工具

李宇轩 [1] **李翔** [1,2,†] **李玮杰** [3] **侯淇彬** [1,2] **刘丽** [3]
程明明 [1,2] **杨健** [1,†]

[1] PCA Lab, VCIP, CS, 南开大学 [2]NKIARI, 福田, 深圳
[3] 湖南先进技术研究院
[†] 通讯作者
yuxuan.li.17@ucl.ac.uk, {xiang.li.implus,houqb,cmm,csjyang}@nankai.edu.cn

摘要

合成孔径雷达 (SAR) 目标检测近年来因其不可替代的全天候成像能力而备受关注。然而，该研究领域受限于公共数据集的匮乏（大多数数据集仅包含少于 2 千张图像，且目标类别单一）以及源代码不可访问。为了应对这些挑战，我们建立了一个新的基准数据集和一个开源方法，用于大规模 SAR 目标检测。我们的数据集 SARDet-100K 是通过深入调研、收集和标准化 10 个现有的 SAR 检测数据集而构建的，为研究目的提供了一个大规模且多样化的数据集。据我们所知，SARDet-100K 是首个 COCO 级别的大规模多类别 SAR 目标检测数据集。借助这个高质量的数据集，我们进行了全面的实验，并揭示了 SAR 目标检测中的一个关键挑战：即在 RGB 数据集上预训练与在 SAR 数据集上微调之间存在显著差异，这种差异体现在数据领域和模型结构两个方面。为了弥合这些差距，我们提出了一种新颖的多阶段滤波器增强 (MSFA) 预训练框架，该框架从数据输入、领域过渡和模型迁移的角度解决这些问题。所提出的 MSFA 方法显著提升了 SAR 目标检测模型的性能，同时展现了跨多种模型的卓越泛化性和灵活性。这项工作旨在为进一步推动 SAR 目标检测的发展铺平道路。数据集和代码公开在`https://github.com/zcablii/SARDet_100K`。

1 简介

合成孔径雷达 (Synthetic Aperture Radar，SAR) [57; 60] 是遥感领域的一项关键技术，相较于传统光学传感器，它具有诸多优势。值得注意的是，SAR 具备在任何天气条件下获取地理图像的能力，不受阳光、地表覆盖或某些类型的伪装等因素的影响，如图 1(a) 所示。得益于这些优势，SAR 已在关键领域得到广泛应用，包括国防 [48]、人道主义救援 [3; 68]、伪装检测 [19] 和地质勘探 [51; 24]。

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

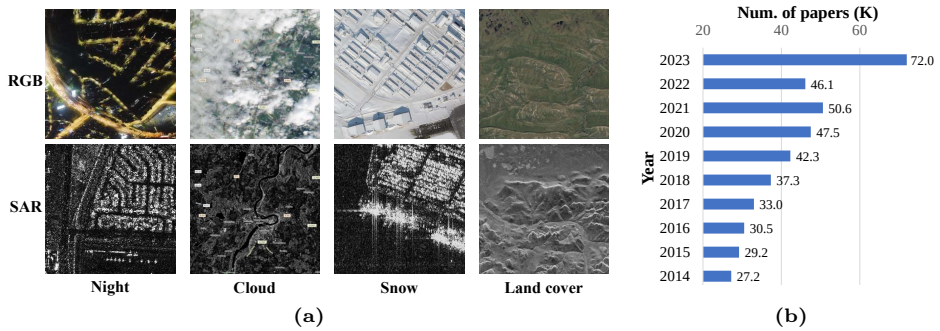


图 1: **(a)** SAR 图像的优势：不受天气条件、阳光和地表覆盖影响。**(b)** 从谷歌学术检索，关键词为“SAR detection”的论文数量 (千篇)。

凭借其宝贵的优势，SAR 目标检测领域日益受到关注。近年来，专注于该领域的研究论文数量大幅增长，如图 1(b) 所示。尽管影响力日益提升，但该研究领域仍然面临着重大挑战，包括资源有限和迁移性差距。

资源有限 高分辨率 SAR 图像目标检测的一个重要障碍是 SAR 图像的敏感性，加之标注这些图像的高成本。这严重限制了公共数据集的可用性。现有的数据集，例如 SAR-AIRcraft [90]、Air-SARShip [76]、SSDD [84] 和 HRSID [71]，通常包含单一类型的物体，背景也较为简单。此外，这些数据集的规模通常也有限，在评估不同方法时，可能会引入偏差。另外，推进 SAR 目标检测研究的一个显著障碍是缺乏公开可访问的源代码，这使得重现先前的研究成果、进行公平比较或在现有工作基础上进行扩展变得具有挑战性。

为了解决这个问题，我们合并了最公开可用的 SAR 检测数据集。这项工作包括全面回顾当前公开的 SAR 检测资源，随后进行收集并将这些数据集标准化为统一的格式，从而创建一个统一的大规模多类别 SAR 目标检测数据集，命名为 SARDet-100k。这个数据集包含约 11.7 万张图像和 24.6 万个物体实例，分布在六个不同的类别中。据我们所知，SARDet-100k 是该研究领域中首个 COCO 规模量级的数据集。它通过提供丰富的资源，显著地有助于克服前面提到的局限性，为 SAR 目标检测模型的开发和评估提供丰富的资源。此外，数据集和源代码将公开。

迁移性差距 通过我们的实证研究和详细分析，我们发现 SAR 目标检测中的一个主要障碍是，当把在自然 RGB 数据集（例如，ImageNet [17]）上预训练的骨干网络迁移到 SAR 图像的检测网络时，会遇到显著的领域差距和模型差距。领域差距源于 RGB 和 SAR 图像之间鲜明的视觉差异，而模型差距则源于预训练的骨干网络与下游任务中采用的整个检测框架之间的模型结构差异。

为了减小上述领域差距和模型差距，我们提出了一种新颖的多阶段滤波器增强 (Multi-Stage with Filter Augmentation，MSFA) 预训练框架，以弥合这些差距。该框架从多个角度应对挑战：数据输入、领域过渡和模型迁移，每个角度都针对 SAR 图像检测任务的独特属性量身定制。对于数据输入：为了解决预训练和微调数据集之间的输入领域差距，我们采用了传统的、手工设计的特征描述符。这些描述符有效地将输入数据从像素空间转换到特征空间，该特征空间不仅对噪声具有鲁棒性，而且在统计上缩小了来自 RGB 和 SAR 模态的数据之间的差距（见图 2(a)），从而增强了预训练知识的可迁移性。对于领域过渡：我们提出了一个利用光学遥感检测数据集的领域过渡桥梁。这个桥梁通过光学相关性连接自然 RGB 图像，并通过目标相关性连接 SAR 图像，建立了一种分层预训练方法，有效地弥合了 RGB 和 SAR 图像之

间的领域差距（见图 2(b)）。对于模型迁移：为了保证整个检测框架的充分训练，并促进用于微调的完整模型迁移，我们在整个多阶段预训练过程中采用完整的检测器作为桥接模型。

MSFA 框架在减少预训练和微调阶段之间通常遇到的显著领域差距和模型差距方面，展示了卓越的有效性。MSFA 不仅有效，而且具有通用性和适用性，可以应用于各种现代深度神经网络。

我们对 SAR 目标检测领域的贡献可以总结为以下四点：

- 引入了首个 COCO 级别的大规模 SAR 多类别目标检测数据集。
- 识别了传统模型预训练和微调方法在 SAR 目标检测中存在的关键差距。
- 提出了多阶段滤波器增强 (MSFA) 预训练框架，该框架展示了卓越的有效性，以及在各种深度网络模型中优秀的通用性和灵活性。
- 通过发布与我们研究相关的数据集和代码，建立了一个新的 SAR 目标检测基准。这一贡献有望促进该领域进一步的发展和进步。

2 相关工作

2.1 SAR 图像和手工特征

由于乘性斑点噪声和伪影的影响，SAR 成像通常受到图像质量不佳的困扰 [57; 47]。为了缓解这个问题，人们开发或改进了许多传统的手工特征描述符，以便从 SAR 图像中提取更易于辨别的特征。这些描述符包括方向梯度直方图 [15] (Histogram of Oriented Gradients，HOG)、Canny 边缘检测器 [5]、比率梯度边缘 (Gradient by Ratio Edge，GRE)[29]、Haar-like[62] 特征描述符和小波散射变换 [45] (Wavelet Scattering Transform，WST)。早期工作采用了传统算法，例如 HOG 用于 SAR 目标识别 [56; 49]，Canny [28; 38] 用于边缘检测。然而，近年来，SAR 图像分析领域已在很大程度上被深度学习方法所主导。

虽然最近的研究主要集中在与低级处理 [69; 86]、分类 [83; 85; 93; 26; 50; 78] 和预训练 [29; 29] 相关的任务，但它们尝试将经典的手工特征集成到现代神经网络中，以实现鲁棒的 SAR 图像特征提取和精细化。与此相反，我们的工作并非简单地将这些手工特征注入网络，而是探索手工特征在现代深度神经网络下，在领域自适应和 SAR 目标检测中的优势和潜力。这个研究领域在很大程度上仍未被探索，而我们的工作旨在弥合这一差距。

2.2 SAR 目标检测

各种流行的基于深度学习的目标检测框架，包括 RetinaNet [33]、FCOS [59]、GFL [30]、RCNN 系列 [53; 4]、YOLO 系列 [52; 10] 和 DETR [6]，在通用目标检测领域展现出了卓越的通用性。此外，诸如 ConvNext [42]、VAN [22]、LSKNet [31] 和 Swin Transformer [41] 等现代骨干网络旨在高效且有效地建模视觉特征。然而，由于 SAR 图像固有的诸如小目标尺寸、斑点噪声和稀疏信息等因素，SAR 图像目标检测提出了独特的挑战。因此，最近用于 SAR 目标检测的深度学习方法主要集中于网络和模块设计，以应对这些挑战。诸如 MGCAN [9]、MSSDNet [91] 和 SEFEPNet [79] 等方法通过多尺度特征融合来增强目标特征。Quad-FPN [82] 结合了四个不同的特征金字塔网络，用于全面的多尺度特征交互，以减轻噪声干扰和多尺度目标特征错位。PADN [88] 和 EWFAN [65] 采用注意力机制来增强存在 SAR 斑点噪声情况下的目标特征。CenterNet++[21] 是 CenterNet[92] 的扩展，它融入了特征增强、多尺度融合和头部细化

表 1: SARDet-100K 数据集的图像和实例级别统计信息。*: 原始数据集被裁剪成 512 × 512 的图像块。

数据集	图片				实例				Ins/Img
	Train	Val	Test	ALL	Train	Val	Test	ALL	
AIR_SARShip 1* [76]	438	23	40	501	816	33	209	1,058	2.11
AIR_SARShip 2 [76]	270	15	15	300	1,819	127	94	2,040	6.80
HRSID [71]	3,642	981	981	5,604	11,047	2,975	2,947	16,969	3.03
MSAR* [75]	27,159	1,479	1,520	30,158	58,988	3,091	3,123	65,202	2.16
SADD [80]	795	44	44	883	6,891	448	496	7,835	8.87
SAR-AIRcraft* [90]	13,976	1,923	2,989	18,888	27,848	4,631	5,996	38,475	2.04
ShipDataset [67]	31,784	3,973	3,972	39,729	40,761	5,080	5,044	50,885	1.28
SSDD [84]	928	116	116	1,160	2,041	252	294	2,587	2.23
OGSOD [63]	14,664	1,834	1,833	18,331	38,975	4,844	4,770	48,589	2.65
SIVED [35]	837	104	103	1,044	9,561	1,222	1,230	12,013	11.51
SARDet-100k	94,493	10,492	11,613	116,598	198,747	22,703	24,023	245,653	2.11

模块，以提高检测器针对 SAR 图像的鲁棒性。此外，CRTransSar [74] 构建于高性能 Swin transformer [41] 之上，利用上下文表示学习来增强目标特征。

虽然大多数现有工作专注于通过改进网络结构来减轻 SAR 斑点噪声干扰，但很少有尝试在输入数据层面解决这个问题。此外，大多数研究利用 ImageNet 预训练的骨干网络作为检测框架的初始化，忽略了预训练的自然场景数据集与微调的 SAR 数据集之间存在的巨大领域差距，以及骨干网络与整个检测框架之间的模型差距。与此相反，我们力求通过精心设计的预训练策略来应对这些独特的挑战。

3 用于 SAR 目标检测的新基准数据集

3.1 当前现状

SAR 图像通常由卫星捕获，并且有大量的低分辨率 SAR 图像可用，通常地面采样距离 (GSD) 为 10 米 × 10 米或更大。诸如 Sentinel-1 [12] 等平台提供了对这些图像的访问，这些图像提供了各种地球物理场所（如城市、山脉、河流和耕地）的宏观视图。这使得它们对于场景分类任务特别有利。然而，这些图像固有的低分辨率限制了它们描绘较小物体的精细细节的能力，例如船舶、汽车和飞机。相反，高分辨率 SAR 图像提供更详细的信息，但需要大量的硬件资源。此外，这些高清图像通常包含敏感信息，因此不适合公开发布。而且，获取高分辨率 SAR 数据集可能非常昂贵，对其可访问性构成重大挑战。

许多研究团队经常遇到预算限制，这限制了他们获取大量且多样化的高分辨率 SAR 数据集的能力。这些财务约束不仅限制了可以覆盖的地理区域范围，还会影响可以访问的数据源的多样性。因此，这些团队提供的数据集通常缺乏多样性，尤其是在光谱波段、极化和分辨率等方面。从研究人员的角度来看，在如此小而同质的数据集上评估模型可能会引入偏差，并导致不公平的性能比较。

3.2 SARDet-100K

为了应对上述挑战，我们对 SAR 目标检测数据集进行了全面的调研。因此，我们精心收集了总共 **10** 个公开可用的高质量数据集，这些数据集不仅具有多样性，而且没有冲突的目标类别。这些数据由中国科研部门、欧洲空间部门和美国军事部门等不同国家和机构发布或收

表 2: SARDet-100K 数据集来源信息。GF-3: 高分三号, S-1: Sentinel-1。目标类别 S: 船舶, A: 飞机, C: 汽车, B: 桥梁, H: 港口, T: 坦克。

数据集	目标	分辨率. (m)	波段	极化	卫星	协议
AIR_SARShip [76]	S	1,3m	C	VV	GF-3	-
HRSID [71]	S	0.5∼3m	C/X	HH, HV, VH, VV	S-1B,TerraSAR-X,TanDEMX	GNU General Public
MSAR [75]	A, T, B, S	$\leq$ 1m	C	HH, HV, VH, VV	HISEA-1	CC BY-NC 4.0
SADD [80]	A	0.5∼3m	X	HH	TerraSAR-X	-
SAR-AIRcraft [90]	A	1m	C	Uni-polar	GF-3	CC BY-NC 4.0
ShipDataset [67]	S	3∼25m	C	HH, VV, VH, HV	S-1,GF-3	-
SSDD [84]	S	1∼15m	C/X	HH, VV, VH, HV	S-1,RadarSat-2,TerraSAR-X	Apache2.0
OGSOD [63]	B, H, T	3m	C	VV/VH	GF-3	-
SIVED [35]	C	0.1,0.3m	Ka,Ku,X	VV/HH	Airborne SAR synthetic slice	-

集。收集的数据集的详细信息如表. 2 所示。为了确保收集的数据集之间的一致性，我们投入了大量时间和精力进行严格的数据集标准化。这包括解决训练集-验证集-测试集划分状态、图像分辨率和标注格式的差异。有关数据收集和标准化的更多详细信息，请参见附录。

表 1 展示了 SARDet-100K 的标准化子数据集及其相应的统计信息，其中包括图像级别和实例级别的统计信息。SARDet-100K 数据集总共包含 116,598 张图像和 245,653 个实例，分布在六个类别中：飞机、船舶、汽车、桥梁、坦克和港口。SARDet-100K 数据集是第一个大规模 SAR 目标检测数据集，其规模可与广泛使用的 COCO[34] 数据集（11.8 万张图像）相媲美，而 COCO 数据集是通用目标检测的标准基准。SARDet-100K 数据集的规模和多样性 有效地模拟了跨多个数据源的 SAR 目标检测模型应用中遇到的真实场景。 SARDet-100K 为研究人员提供了稳健的训练和评估，以推进 SAR 目标检测算法和技术，从而促进该领域 SOTA 模型的发展。

4 带有滤波增强的预训练框架

最近的一些研究 [26; 78; 74; 21] 已经证明了成熟的手工特征和专门的网络模块设计在提高 SAR 目标检测性能方面的有效性。然而，这些工作中的大多数都依赖于默认的 ImageNet 预训练方法，因此忽略了预训练的自然场景数据集与微调的 SAR 数据集之间存在的巨大领域差距。此外，他们也未能解决骨干网络和整个检测框架之间存在的模型差距。为了解决这些局限性，我们提出了一种名为带有滤波增强的多阶段 (MSFA) 预训练框架的新框架。我们的框架从数据输入、领域过渡和模型迁移的角度应对挑战。MSFA 包含两个核心设计：滤波增强输入和多阶段预训练策略。

4.1 滤波增强输入

正如 相关工作部分所讨论的，许多现有的手工特征描述符利用精心设计的滤波器来提取特征。这些特征具有鲁棒性和丰富的信息，充当从原始图像导出的增强信息。因此，我们建议使用此类特征作为原始像素数据的辅助信息。数据 x 的滤波增强特征 M 可以概括地定义为:

$$M_i^x = T_i(x), i \in \{HOG, Canny, Haar, WST, GRE\}. \tag{1}$$

其中，T_i 是预定义的变换。从 ResNet [23] 中的信息残差设计中汲取灵感， 我们构建了检测模型的滤波增强输入 Inp，方法是将原始灰度 SAR 图像 x 与生成的滤波增强特征 M_i^x 连接

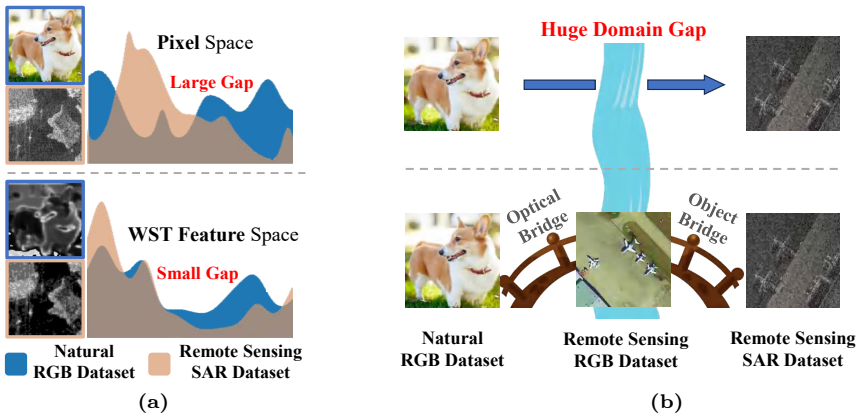


图 2: 图示了自然 RGB 数据集和遥感 SAR 数据集之间存在的显著领域差距。**(a)** 展示了 WST 特征空间显著缩小了领域差距。**(b)** 证明了遥感 RGB 数据集充当了有效的领域过渡桥梁，促进更平滑的领域转移。

起来，如下所示：

$$Inp = \text{concat}(x, M_i^x). \tag{2}$$

通过将原始数据输入从异构像素空间转换为同构滤波增强特征空间，可以大大缩小不同图像域之间的领域差距，如图 2(a) 所示。

4.1.1 多阶段预训练

我们将传统的预训练模式公式化为：

$$B = \text{Train}_{cls}(B_\theta)(D_{IN}), \tag{3}$$

$$A = \text{Train}_{det}(A_B)(D_{SAR}). \tag{4}$$

函数 $\text{Train}_t(a)(b)$ 表示使用任务 t 在数据集 b 上训练模型 a，并返回训练后的模型。其中 t 是训练任务，$t \in \{cls, det\}$，cls 代表分类，det 代表检测。B 表示骨干模型，A 是完整的检测模型。传统上，预训练阶段会随机初始化骨干模型 B_θ，并在 ImageNet 数据集 D_{IN} 上进行训练（如公式 (3) 所示）。然后使用从检测模型 A_B 初始化预训练骨干网络的检测模型在 SAR 数据集 D_{SAR} 上进行微调（如公式 (4) 所示）。

我们提出的多阶段预训练策略，作为替代方案，可以如图公式 (3)、(5)、(6) 所示。

$$A' = \text{Train}_{det}(A_B)(D_{RS}). \tag{5}$$

$$A = \text{Train}_{det}(A_{A'})(D_{SAR}). \tag{6}$$

其中在公式 (5) 中增加了一个额外的第二阶段预训练。

我们建议利用大规模光学遥感数据集 D_{RS} 作为检测预训练以用于领域过渡。该数据集由光学模态图像组成，这些图像在下游 SAR 数据集中也共享相似的物体形状、尺度和类别。这一特性充当了 ImageNet 中自然图像的光学分布与 SAR 遥感图像中物体分布之间有价值的桥梁。通过利用这种第二阶段预训练，可以有效地最小化领域差距，如图 2(b) 所示。

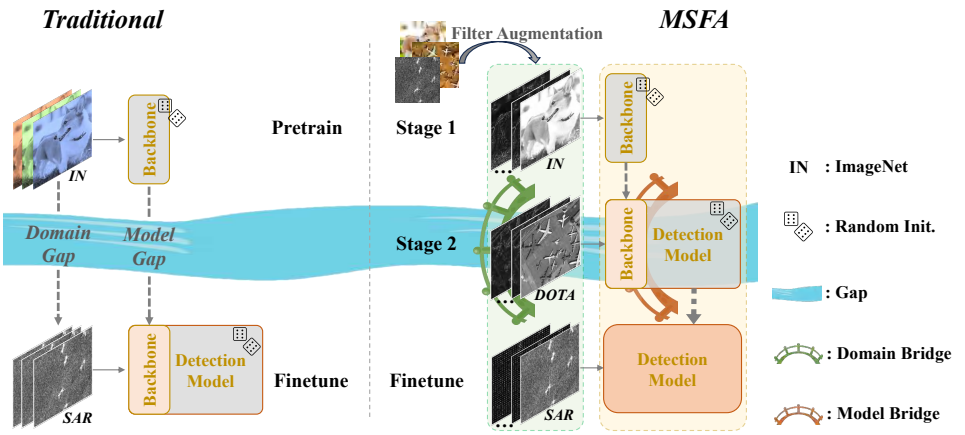


图 3: 传统 ImageNet 预训练和我们提出的带有滤波增强的多阶段 (MSFA) 预训练框架的概念图。

4.1.2 MSFA

最后，我们提出的 MSFA 框架集成了滤波增强输入与多阶段预训练，如图 3 所示。我们的 MSFA 框架有效地弥合了在自然图像上进行预训练和在 SAR 图像检测上进行微调之间存在的巨大领域和模态差距。

通过引入滤波增强输入，我们利用成熟的手工特征描述符来提取对噪声鲁棒的特征。这也使我们能够有效地将预训练和微调图像的异构图像域转换为同构特征域。通过将输入数据统一到一致的特征域中，我们解决了不同类型图像之间存在的差异。因此，它增强了跨领域知识的对齐和可迁移性。此外，多阶段训练的结合包括利用额外的、大规模的光学遥感数据集进行检测预训练。该数据集充当领域桥梁，连接了 ImageNet 自然图像的领域与 SAR 遥感图像的领域。因此，它进一步缩小了领域差距，促进了两个领域之间更平滑的过渡。更重要的是，MSFA 框架第二阶段的检测预训练也可以充当模型桥梁。它允许对整个检测框架进行全面训练，而不仅仅是专注于骨干网络，使得整个检测框架得到良好的初始化，从而在 SAR 检测微调中实现最佳性能。

5 实验与分析

5.1 滤波增强输入

Input	**mAP** ↑	mAP_{50} ↑
SAR (as RGB)	50.2	83.0
SAR+Canny	50.7	83.6
SAR+Hog	50.7	83.5
SAR+Haar	50.6	83.4
SAR+WST	**51.1**	<u>83.9</u>
SAR+GRE	50.6	83.8
SAR+Hog+Haar+WST	**51.1**	**84.0**

表 3: 使用 Faster R-CNN 和 ResNet50 作为检测模型时，不同滤波增强输入的比较。

Domain	PCC ↑
Pixel Space	0.394
Canny Space	0.992
Hog Space	0.995
Haar Space	0.990
WST Space	**0.996**
GRE Space	0.984

表 4: ImageNet 和 SARDet-100k 在 RGB 和手工特征空间上的皮尔逊相关系数 (PCC)。

在我们提出的 MSFA 方法框架内，为了研究和评估提出的滤波增强输入的影响，我们对 相关工作中讨论的每种传统特征描述符进行了实验。表 3 中详细介绍的调查结果表明，结合这些手工特征显著提高了检测器的性能。此外，我们的分析表明，将图像像素转换为手工特征空间可以显著缩小 ImageNet 和 SARDet-100K 数据集之间的分布差距。这在 ImageNet 和 SARDet-100K 数据集输入之间的皮尔逊相关系数 (PCC) 中尤为明显，如表 4 所示。这 了所提出的方法在弥合自然图像和 SAR 图像之间的领域差距方面的有效性，从而提高了从预训练过程进行知识转移的效率。

值得注意的是，小波散射变换 (WST) 特征以其卓越的性能脱颖而出。这种优越性不仅可以归因于其在显著缩小领域差距方面的作用，还可以归因于其提取丰富的多尺度信息的能力。这些信息通过减轻噪声和保留与对象相关的细节，充当了强大的辅助特征。然而，我们也发现使用多种滤波增强特征不会带来进一步显著的性能提升。可能是现有的 WST 已经捕获了有效对象检测所需的必要信息，而结合额外的特征并不会提供大量额外的有益信息。

由于 WST 的出色性能，我们在本文的剩余部分中，将其用作 MSFA 方法中的默认滤波增强输入。

5.2 多阶段预训练

为了评估所提出的多阶段预训练方法的有效性，我们进行了实验，在这些实验中，我们保持输入模态一致，并使用各种预训练策略在 SARDet-100K 数据集上微调检测模型。作为基线，实验 1 采用单通道 SAR 数据作为输入，在 ImageNet 上预训练骨干网络模型 100 个 epoch，然后直接在 SARDet-100K 数据集上微调检测器（遵循广泛使用的默认设置）。除了基线之外，我们还针对光学遥感数据集（例如 DOTA [73] 或 DIOR [27]）进行专门用于目标检测的第二阶段预训练。(有关 DOTA 和 DIOR 数据集的详细信息，请参见附录)。在第二阶段预训练之后，我们仅在骨干网络或整个框架上微调模型。

表 5 中实验 2、4、6 和 8 的结果证明了两阶段预训练方法的显着优势。值得注意的是，即使是规模相对较小的 DIOR 数据集也显示出比基线（实验 1 和 5）明显的性能提升。这一观察结果 了在 SAR 检测的预训练阶段减少领域差距的重要性。

表 5: 使用 Faster-RCNN 和 ResNet50 作为检测模型时，不同预训练策略的比较。

ID	Model Input	Pretrain			mAP ↑
		Multi-stage	Dataset	Component	
1	SAR (Raw Pixels)	✗	ImageNet	Backbone	49.0
2		✓	ImageNet + DIOR	Framework	49.5
3		✓	ImageNet + DOTA	Backbone	49.3
4				Framework	**50.2**
5	SAR+WST (Filter Augmented)	✗	ImageNet	Backbone	49.2
6		✓	ImageNet + DIOR	Framework	50.1
7		✓	ImageNet + DOTA	Backbone	49.6
8				Framework	**51.1**

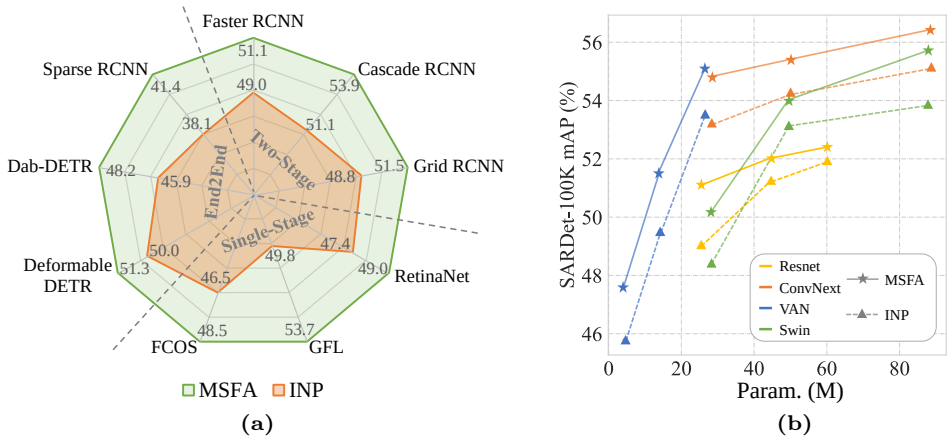


图 4: MSFA 在不同检测框架 **(a)** 和不同骨干网络 **(b)** 上的泛化性。模型在 SARDet-100K 数据集上进行了微调和测试。 INP：仅在骨干网络上进行传统的 ImageNet 预训练。

然而，DIOR 数据集预训练不如更大规模的 DOTA 数据集有效（实验 2 与 4，实验 6 与 8）。这种比较强调了预训练规模在获得最佳结果方面的重要性。DOTA 数据集具有更大的规模，并且平均实例面积与 SARDet-100K 相似，因此提供了更全面和信息更丰富的预训练，从而提高了后续微调阶段的性能。

实验 3 和 4 与实验 7 和 8 之间的比较，证明了预训练整个框架优于仅预训练骨干网络，突出了模型差距对 SAR 检测性能的显着影响。

总之，我们提出的 MSFA 中的多阶段预训练策略缓解了预训练和下游模型之间的数据领域差距和模型差距，从而显着提高了 SAR 检测性能。详细的实验结果和可视化效果可在附录中找到。

5.3 MSFA 的泛化性

为了评估所提出的 MSFA 的有效性和泛化性，我们使用各种检测器和骨干网络进行了实验，如图 4(a) 和 4(b) 所示。在不同的框架（包括单阶段 [53; 4; 43]、两阶段 [33; 30; 59] 和端到端 [94; 58; 39]）和各种骨干网络（包括 ResNets [23]、ConvNexts [42]、VANs [22] 和 Swin-Transformer [41] 网络）中，都观察到了显著的性能提升。这为我们提出的方法的有效性和广泛适用性提供了强有力的证据。此外，如图 4(b) 所示，我们观察到随着骨干网络规模的扩大，性能也稳定提升，这表明我们提出的方法具有良好的可扩展性。

表 6: 提出的 MSFA 与先前最先进方法在 SSDD 和 HRSID 数据集上的比较。

检测器		开源	年份	mAP_{50} ↑	
				SSDD	HRSID
常规检测器	Grid R-CNN [43]	✓	2019	88.9	79.4
	Faster R-CNN [53]	✓	2015	89.7	80.7
	Cascade R-CNN [4]	✓	2019	90.5	81.3
	Free-Anchor [87]	✓	2019	91.0	81.8
	Double-Head R-CNN [72]	✓	2020	91.1	82.1
	PANET [40]	✓	2018	91.2	81.6
	DCN [14]	✓	2017	92.3	82.1
SAR 检测器	NNAM [7]	✗	2019	79.8	-
	DCMSNM [25]	✗	2018	89.6	-
	ARPN [89]	✗	2020	89.9	81.8
	DAPN [13]	✗	2019	90.6	81.8
	HR-SDNet [70]	✗	2020	90.8	82.5
	SER Faster R-CNN [37]	✗	2018	91.5	81.5
	FBR-Net [20]	✗	2020	94.1	-
	NRENet [44]	✗	2024	94.6	75.6
	CenterNet++ [21]	✗	2021	95.1	-
	CRTransSar [74]	✗	2022	97.0	-
	SARATR-X [77]	✗	2024	97.3	80.3
Faster R-CNN + VAN-B		✓	2023	92.9	81.8
MSFA (Faster R-CNN + VAN-B)		✓	2024	**97.9**(+5.0)	**83.7**(+1.9)

值得注意的是，我们 MSFA 方法的设计在开发时就考虑了灵活性、泛化性和广泛的适用性。因此，该方法可以无缝集成到大多数现有模型中，而无需进行任何修改。

5.4 与 SOTAs 方法的比较

我们比较了各种 SOTA 方法，包括通用目标检测模型 [43; 53; 14; 64; 4; 87; 72] 以及 SAR 目标检测模型 [7; 25; 89; 13; 70; 40; 37; 74; 21; 20]。我们评估了它们在 SSDD 和 HRSID 数据集上的性能，这些数据集是常用的 SAR 目标检测基准。为了利用 VAN [22] 骨干网络卓越的效率和性能（如图 4(b) 所示），我们采用经典的 Faster R-CNN 检测框架，并使用轻量级的 VAN-B (参数量 26.6M) 骨干网络作为我们的检测模型。表 6 中展示的结果表明，我们的 MSFA 方法明显优于所有比较的方法。具体而言，MSFA 在 SSDD 数据集上实现了 97.9% 的 mAP@50，在 HRSID 数据集上实现了 83.7% 的 mAP@50，创造了新的最先进水平的结果。值得注意的是，在我们比较的 SAR 检测 SOTA 方法中，我们的方法是唯一开源的方法。

6 局限性和未来工作

本文的范围仅限于有监督的预训练。然而，考虑到大量未标注的 SAR 图像的可用性，探索半监督、弱监督或无监督学习方法在 SAR 目标检测中进行领域迁移的潜力将非常有价值。

虽然本文旨在提出一种简单、实用、有效和通用的方法，但并未深入研究具体设计的细节。未来的工作可以扩展到更深入地探索上述方向，结合复杂和专门的设计，以增强 SAR 目标检测的性能和能力。

7 结论

本文提出了一个用于大规模 SAR 目标检测的新基准，介绍了 SARDet-100k 数据集和多阶段滤波器增强 (MSFA) 预训练方法。我们的 SARDet-100k 数据集包含超过 11.6 万张图像，涵盖 6 个类别，为开展 SAR 目标检测研究提供了大型且多样化的数据集。为了弥合 SAR 目标检测中预训练和微调阶段之间的领域和模型差距，我们提出了 MSFA 预训练框架。MSFA 显著提高了 SAR 目标检测模型的性能，并在之前的基准数据集上取得了新的最先进的性能水平。此外，MSFA 在各种模型中都表现出卓越的泛化性和灵活性。我们的研究致力于克服当前 SAR 目标检测中普遍存在的障碍。我们预计我们的贡献将为该领域未来的研究和创新铺平道路。

我们的研究致力于克服当前 SAR 目标检测中普遍存在的障碍。我们预计我们的贡献将为该领域未来的研究和创新铺平道路。

8 致谢

我们向以下研究人员致以最深切的感谢，他们按名字的首字母顺序排列：Hong Zhang、Runfan Xia、Shunjun Wei、Tianwen Zhang、Xian Sun、Xiaofang Zhu、Xiaoling Zhang 以及其他做出贡献的研究人员，感谢他们允许我们整合他们的数据集。他们的贡献极大地推进和促进了该领域的研究。

本研究由国家自然科学基金 (62361166670, 62276145, 62176130, 62276134) 和中央高校基本科研业务费 (南开大学: 070-63233084, 070-63243142) 资助。计算工作由南开大学高性能计算中心 (NKSC) 支持。

参考文献

[1] Ai, J., Tian, R., Luo, Q., Jin, J., Tang, B.: Multi-scale rotation-invariant haar-like feature integrated cnn-based ship detection algorithm of multiple-target environment in sar imagery. TGRS (2019)

[2] Besnassi, M., Neggaz, N., Benyettou, A.: Face detection based on evolutionary haar filter. Pattern Analysis and Applications (2020)

[3] Braun, A.: Radar satellite imagery for humanitarian response. Bridging the gap between technology and application. Ph.D. thesis, Universität Tübingen (2019)

[4] Cai, Z., Vasconcelos, N.: Cascade R-CNN: High quality object detection and instance segmentation. TPAMI (2019)

[5] Canny, J.: A computational approach to edge detection. TPAMI (1986)

[6] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: ECCV (2020)

[7] Chen, C., He, C., Hu, C., Pei, H., Jiao, L.: A deep neural network based on an attention mechanism for sar ship detection in multiscale and complex scenarios. IEEE Access (2019)

[8] Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., et al.: Mmdetection: Open mmlab detection toolbox and benchmark. arXiv (2019)

[9] Chen, L., Luo, R., Xing, J., Li, Z., Yuan, Z., Cai, X.: Geospatial transformer is what you need for aircraft detection in sar imagery. TGRS (2022)

[10] Chen, Y., Yuan, X., Wu, R., Wang, J., Hou, Q., Cheng, M.M.: YOLO-MS: Rethinking multi-scale representation learning for real-time object detection. arXiv (2023)

[11] Contributors, M.: Openmmlab' s pre-training toolbox and benchmark. https://github.com/openmmlab/mmpretrain (2024)

[12] Contributors, S..: Sentinel-1 - missions. https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1 (2024)

[13] Cui, Z., Li, Q., Cao, Z., Liu, N.: Dense attention pyramid networks for multi-scale ship detection in sar images. TGRS (2019)

[14] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y.: Deformable convolutional networks. In: ICCV. pp. 764–773 (2017)

[15] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)

[16] Dellinger, F., Delon, J., Gousseau, Y., Michel, J., Tupin, F.: SAR-SIFT: a SIFT-like algorithm for SAR images. TGRS (2014)

[17] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: CVPR (2009)

[18] Feng, Y., Han, B., Wang, X., Shen, J., Guan, X., Ding, H.: Self-supervised transformers for unsupervised sar complex interference detection using canny edge detector. Remote Sensing (2024)

[19] Frolind, P.O., Gustavsson, A., Lundberg, M., Ulander, L.M.: Circular-aperture vhf-band synthetic aperture radar for detection of vehicles in forest concealment. IEEE Transactions on Geoscience and Remote Sensing (2011)

[20] Fu, J., Sun, X., Wang, Z., Fu, K.: An anchor-free method based on feature balancing and refinement network for multiscale ship detection in sar images. TGRS (2020)

[21] Guo, H., Yang, X., Wang, N., Gao, X.: A centernet++ model for ship detection in sar images. Pattern Recognition (2021)

[22] Guo, M.H., Lu, C., Liu, Z.N., Cheng, M.M., Hu, S.: Visual attention network. Computational Visual Media (2022)

[23] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

[24] Ivanov, A.Y., Gerivani, H., Evtushenko, N.V.: Characterization of natural hydrocarbon seepage in the south caspian sea off iran using satellite sar and geological data. Marine Georesources & Geotechnology (2020)

[25] Jiao, J., Zhang, Y., Sun, H., Yang, X., Gao, X., Hong, W., Fu, K., Sun, X.: A densely connected end-to-end neural network for multiscale and multiscene sar ship detection. IEEE Access (2018)

[26] Jin, Y., Duan, Y.: Wavelet scattering network-based machine learning for ground penetrating radar imaging: Application in pipeline identification. Remote Sensing (2020)

[27] Li, K., Wan, G., Cheng, G., Meng, L., Han, J.: Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS (2020)

[28] Li, W., He, M., Zhang, S.: A new canny-based edge detector for sar image. In: Congress on Image and Signal Processing (2008)

[29] Li, W., Wei, Y., Liu, T., Hou, Y., Liu, Y., Liu, L.: Self-supervised learning for sar atr with a knowledge-guided predictive architecture. arXiv (2023)

[30] Li, X., Lv, C., Wang, W., Li, G., Yang, L., Yang, J.: Generalized focal loss: Towards efficient representation learning for dense object detection. TPAMI (2022)

[31] Li, Y., Hou, Q., Zheng, Z., Cheng, M.M., Yang, J., Li, X.: Large selective kernel network for remote sensing object detection. In: ICCV (2023)

[32] Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. In: International Conference on Image Processing (2002)

[33] Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV (2017)

[34] Lin, T.Y., Maire, M., Belongie, S.J., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: ECCV (2014)

[35] Lin, X., Zhang, B., Wu, F., Wang, C., Yang, Y., Chen, H.: Sived: A sar image dataset for vehicle detection based on rotatable bounding box. Remote Sensing (2023)

[36] Lin, Y.N., Hsieh, T.Y., Huang, J.J., Yang, C.Y., Shen, V.R., Bui, H.H.: Fast iris localization using haar-like features and adaboost algorithm. Multimedia Tools and Applications (2020)

[37] Lin, Z., Ji, K., Leng, X., Kuang, G.: Squeeze and excitation rank faster r-cnn for ship detection in sar images. IEEE Geoscience and Remote Sensing Letters (2018)

[38] Liu, H., Jezek, K.: Automated extraction of coastline from satellite imagery by integrating canny edge detection and locally adaptive thresholding methods. International journal of remote sensing (2004)

[39] Liu, S., Li, F., Zhang, H., Yang, X., Qi, X., Su, H., Zhu, J., Zhang, L.: DAB-DETR: Dynamic anchor boxes are better queries for DETR. In: ICLR (2022)

[40] Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: CVPR (2018)

[41] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: CVPR (2021)

[42] Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: CVPR (2022)

[43] Lu, X., Li, B., Yue, Y., Li, Q., Yan, J.: Grid r-cnn. In: CVPR. pp. 7363–7372 (2019)

[44] Ma, W., Yang, X., Zhu, H., Wang, X., Yi, X., Wu, Y., Hou, B., Jiao, L.: Nrenet: Neighborhood removal-and-emphasis network for ship detection in sar images. International Journal of Applied Earth Observation and Geoinformation (2024)

[45] Mallat, S.: Group invariant scattering. Communications on Pure and Applied Mathematics (2012)

[46] Mita, T., Kaneko, T., Hori, O.: Joint haar-like features for face detection. In: ICCV (2005)

[47] Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K.P.: A tutorial on synthetic aperture radar. IEEE Geoscience and remote sensing magazine (2013)

[48] Peng, B., Peng, B., Zhou, J., Xie, J., Liu, L.: Scattering model guided adversarial examples for sar target recognition: Attack and defense. IEEE Transactions on Geoscience and Remote Sensing (2022)

[49] Qi, S., Ma, J., Lin, J., Li, Y., Tian, J.: Unsupervised ship detection based on saliency and s-hog descriptor from optical satellite images. IEEE geoscience and remote sensing letters (2015)

[50] Qin, R., Fu, X., Chang, J., Lang, P.: Multilevel wavelet-srnet for sar target recognition. IEEE Geoscience and Remote Sensing Letters (2021)

[51] Ramadan, T.M., Onsi, H.M.: Use of ers-2 sar and landsat tm images for geological mapping and mineral exploration of sol hamid area, south eastern desert, egypt. Egyptian Journal of Remote Sensing and Space Sciences (2003)

[52] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: CVPR (2016)

[53] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: NeurIPS (2015)

[54] Schwegmann, C.P., Kleynhans, W., Salmon, B.P.: Ship detection in south african oceans using sar, cfar and a haar-like feature classifier. In: IEEE Geoscience and Remote Sensing Symposium (2014)

[55] Schwegmann, C.P., Kleynhans, W., Salmon, B.P.: Synthetic aperture radar ship detection using haar-like features. IEEE Geoscience and Remote Sensing Letters (2016)

[56] Song, S., Xu, B., Yang, J.: Sar target recognition via supervised discriminative dictionary learning and sparse representation of the sar-hog feature. Remote Sensing (2016)

[57] Sun, G.C., Liu, Y., Xiang, J., Liu, W., Xing, M., Chen, J.: Spaceborne synthetic aperture radar imaging algorithms: An overview. IEEE Geoscience and Remote Sensing Magazine (2021)

[58] Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., Tomizuka, M., Li, L., Yuan, Z., Wang, C., et al.: Sparse r-cnn: End-to-end object detection with learnable proposals. In: CVPR (2021)

[59] Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: ICCV (2019)

[60] Tsokas, A., Rysz, M., Pardalos, P.M., Dipple, K.: Sar data applications in earth observation: An overview. Expert Systems with Applications (2022)

[61] Vidal-Pantaleoni, A., Marti, D.: Comparison of different speckle-reduction techniques in sar images using wavelet transform. International Journal of Remote Sensing (2004)

[62] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR (2001)

[63] Wang, C., Ruan, R., Zhao, Z., Li, C., Tang, J.: Category-oriented localization distillation for sar object detection and a unified benchmark. IEEE Transactions on Geoscience and Remote Sensing (2023)

[64] Wang, J., Chen, K., Yang, S., Loy, C.C., Lin, D.: Region proposal by guided anchoring. In: CVPR (2019)

[65] Wang, J., Xiao, H., Chen, L., Xing, J., Pan, Z., Luo, R., Cai, X.: Integrating weighted feature fusion and the spatial attention module with convolutional neural networks for automatic aircraft detection from sar images. Remote Sensing (2021)

[66] Wang, Y., Hernández, H.H., Albrecht, C.M., Zhu, X.X.: Feature guided masked autoencoder for self-supervised learning in remote sensing. arXiv (2023)

[67] Wang, Y., Wang, C., Zhang, H., Dong, Y., Wei, S.: A sar dataset of ship detection for deep learning under complex backgrounds. remote sensing **11**(7), 765 (2019)

[68] Wegmuller, U., Wiesmann, A., Strozzi, T., Werner, C.: Envisat asar in disaster management and humanitarian relief. In: IEEE International Geoscience and Remote Sensing Symposium (2002)

[69] Wei, Q.R., Wang, Y.K., Xie, P.Y.: Sar edge detector with high localization accuracy. In: International Geoscience and Remote Sensing Symposium (2019)

[70] Wei, S., Su, H., Ming, J., Wang, C., Yan, M., Kumar, D., Shi, J., Zhang, X.: Precise and robust ship detection for high-resolution sar imagery based on hr-sdnet. Remote Sensing (2020)

[71] Wei, S., Zeng, X., Qu, Q., Wang, M., Su, H., Shi, J.: Hrsid: A high-resolution sar images dataset for ship detection and instance segmentation. IEEE Access (2020)

[72] Wu, Y., Chen, Y., Yuan, L., Liu, Z., Wang, L., Li, H., Fu, Y.: Rethinking classification and localization for object detection. In: CVPR (2020)

[73] Xia, G.S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L.: DOTA: A large-scale dataset for object detection in aerial images. In: CVPR (2018)

[74] Xia, R., Chen, J., Huang, Z., Wan, H., Wu, B., Sun, L., Yao, B., Xiang, H., Xing, M.: Crtranssar: A visual transformer based on contextual joint representation learning for sar ship detection. Remote Sensing (2022)

[75] Xia, R., Chen, J., Huang, Z., Wan, H., Wu, B., Sun, L., Yao, B., Xiang, H., Xing, M.: Crtranssar: A visual transformer based on contextual joint representation learning for sar ship detection. Remote Sensing (2022)

[76] Xian, S., Zhirui, W., Yuanrui, S., Wenhui, D., Yue, Z., Kun, F.: Air-sarship-1.0: High-resolution sar ship detection dataset. J. Radars (2019)

[77] Yang, W., Hou, Y., Liu, L., Liu, Y., Li, X., et al.: Saratr-x: A foundation model for synthetic aperture radar images target recognition. arXiv (2024)

[78] Zhang, J., Xing, M., Xie, Y.: Fec: A feature fusion framework for sar target recognition based on electromagnetic scattering features and deep cnn features. TGRS (2020)

[79] Zhang, P., Xu, H., Tian, T., Gao, P., Li, L., Zhao, T., Zhang, N., Tian, J.: Sefepnet: Scale expansion and feature enhancement pyramid network for sar aircraft detection with small sample dataset. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2022)

[80] Zhang, P., Xu, H., Tian, T., Gao, P., Li, L., Zhao, T., Zhang, N., Tian, J.: Sefepnet: Scale expansion and feature enhancement pyramid network for sar aircraft detection with small sample dataset. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2022)

[81] Zhang, S., Bauckhage, C., Cremers, A.B.: Informed haar-like features improve pedestrian detection. In: CVPR (2014)

[82] Zhang, T., Zhang, X., Ke, X.: Quad-fpn: A novel quad feature pyramid network for sar ship detection. Remote Sensing (2021)

[83] Zhang, T., Zhang, X., Ke, X., Liu, C., Xu, X., Zhan, X., Wang, C., Ahmad, I., Zhou, Y., Pan, D., et al.: Hog-shipclsnet: A novel deep learning network with hog feature fusion for sar ship classification. TGRS (2021)

[84] Zhang, T., Zhang, X., Li, J., Xu, X., Wang, B., Zhan, X., Xu, Y., Ke, X., Zeng, T., Su, H., et al.: Sar ship detection dataset (ssdd): Official release and comprehensive data analysis. Remote Sensing (2021)

[85] Zhang, T., Zhang, X., Shi, J., Wei, S.: A hog feature fusion method to improve cnn-based sar ship classification accuracy. In: IEEE International Geoscience and Remote Sensing Symposium (2021)

[86] Zhang, W.: Combination of sift and canny edge detection for registration between sar and optical images. IEEE Geoscience and Remote Sensing Letters (2020)

[87] Zhang, X., Wan, F., Liu, C., Ji, R., Ye, Q.: Freeanchor: Learning to match anchors for visual object detection. NeurIPS (2019)

[88] Zhao, Y., Zhao, L., Li, C., Kuang, G.: Pyramid attention dilated network for aircraft detection in sar images. IEEE Geoscience and Remote Sensing Letters (2020)

[89] Zhao, Y., Zhao, L., Xiong, B., Kuang, G.: Attention receptive pyramid network for ship detection in sar images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2020)

[90] Zhirui, W., Yuzhuo, K., Xuan, Z., Yuelei, W., Ting, Z., Xian, S.: Sar-aircraft-1.0: High-resolution sar aircraft detection and recognition dataset. J. Radars (2023)

[91] Zhou, K., Zhang, M., Wang, H., Tan, J.: Ship detection in sar images based on multi-scale feature extraction and adaptive feature fusion. Remote Sensing (2022)

[92] Zhou, X., Wang, D., Krähenbühl, P.: Objects as points. arXiv (2019)

[93] Zhu, H., Wong, T., Lin, N., Lung, H., Li, Z., Thedoridis, S.: A new target classification method for synthetic aperture radar images based on wavelet scattering transform. In: ICSPCC (2020)

[94] Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159 (2020)

A 附录

Fig. S5 以可视化方式展示了提议的 SARDet-100K 数据集中的样本图像。突出了每个类别的代表性样本，包括 Ship（舰船）、Tank（坦克）、Bridge（桥梁）、Harbour（港口）、Aircraft（飞机）和 Car（车辆）。

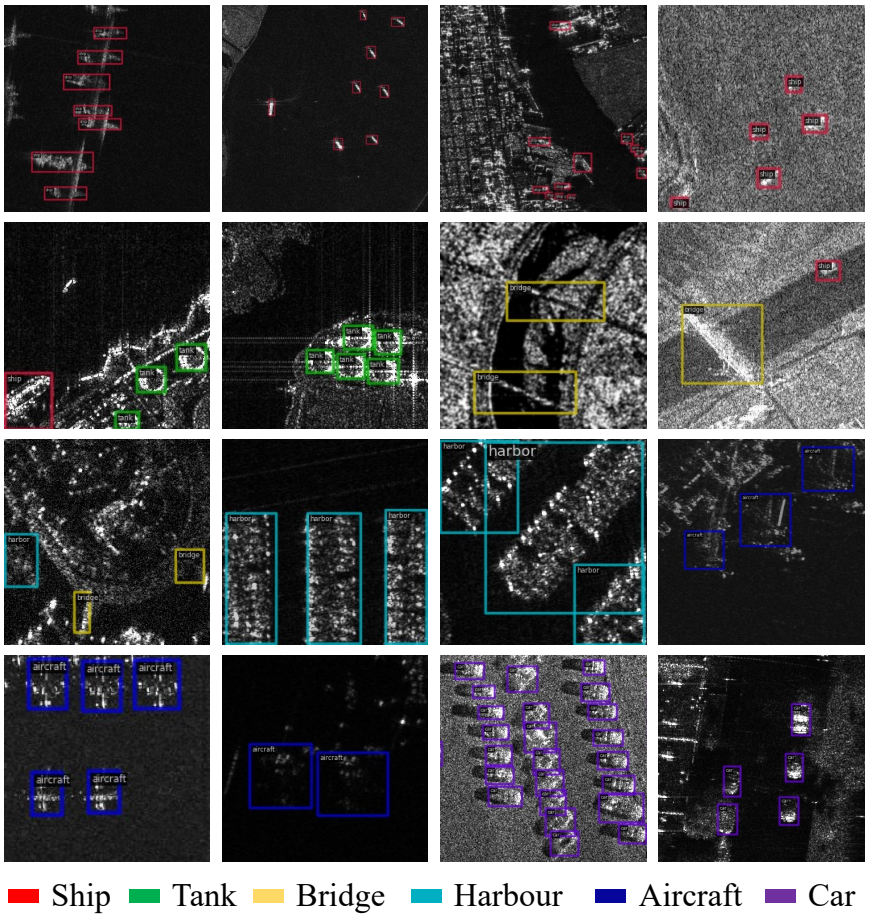


图 S5: 提议的 SARDet-100K 数据集样本图像可视化。

A.1 标准化

我们总共收集了 **10** 个公开可用的高质量数据集，这些数据集不仅多样化，而且没有冲突的对象类别。这些数据由中国科研部门、欧洲航天部门和美国军事部门等不同国家和机构发布或收集。为了确保收集的数据集之间的一致性，有必要进行标准化处理。这包括解决训练集-验证集-测试集划分状态、图像分辨率和注释格式的差异。SARDet-100K 数据集处理流程的概述如图 S6(a) 所示。

如果源数据集已经提供了预定义的训练集、验证集和测试集划分，我们则采用其划分设置。否则，我们执行划分以确保训练集、验证集和测试集的比例分别为 8:1:1。

此外，我们还解决了某些收集的数据集中图像分辨率较高的问题。出现这个问题的原因是，在将这些图像传递给模型之前调整其大小可能会导致目标变得非常小。我们对所有包含大于 1000×1000 分辨率图像的数据集执行图像切片。具体而言，对于 AIR SARShip 1、MSAR 和

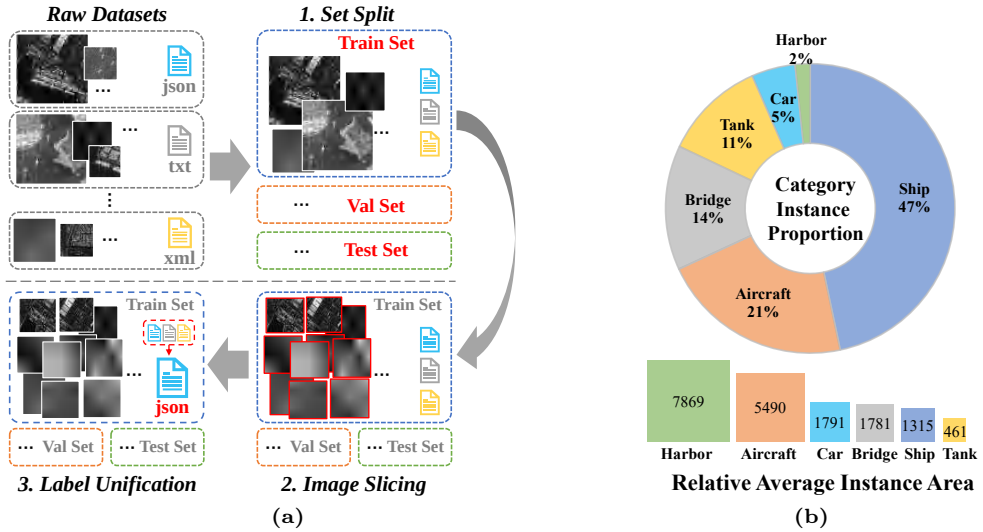


图 S6: **(a)** SARDet-100K 数据集标准化过程，包括数据集划分、大图像切片和标签注释格式统一。**(b)** SARDet-100K 中每个类别的实例百分比和平均实例面积（像素为单位）。

SAR-AIRcraft 数据集，我们将每张图像裁剪成大小为 512 × 512 的图像块，图像块重叠为 200。

此外，我们将所有数据集注释转换为 COCO 注释格式 [34]。此步骤确保了不同数据集之间的一致性和兼容性。因此，合并后的数据集 SARDet-100K 也以 COCO 格式标准化，该格式与流行的开源检测代码框架 readily 兼容，无需额外的手动数据预处理。图 S6(b) 概述了 SARDet-100K 数据集的类别级统计信息。

A.2 手工特征描述符

方向梯度直方图 [15] (HOG) HOG 是一种在图像处理和计算机视觉中广泛使用的局部特征描述符。它通过分析梯度方向的分布来捕获和表示图像的局部结构和形状信息。HOG 被证明对 SAR 图像分类和目标检测任务有效，因为它对 SAR 图像中的随机噪声具有不变性。早期工作采用 HOG 进行 SAR 目标识别 [56; 49]，最近的研究表明 HOG 特征对于 SAR 图像分类 [83; 85] 以及现代神经网络中的模型预训练 [66] 的有效性。

Canny 边缘检测器 [5] Canny 是一种广泛使用的边缘检测算法，旨在识别图像中的显著边缘，同时最大限度地减少噪声和虚假响应。该算法利用高斯平滑、像素方向梯度幅值和非极大值抑制。

早期研究 [28; 38] 认识到 Canny 在 SAR 图像处理中的优势。最近的工作也验证了 Canny 特征在 SAR 图像边缘检测 [69]、干扰检测 [18] 和图像配准 [86] 中的有效性。除 Canny 边缘检测器外，梯度比率边缘 (GRE) [29] 是一种最近提出的边缘检测器，它利用 SAR-HOG [16] 和 SAR-SIFT [56] 在 SAR 图像中实现有效的边缘检测。

Haar-like [62] 特征描述符 Haar-like 特征常用于人脸检测 [46; 2]、行人检测 [81] 和其他目标检测任务 [32; 36]。它们通过利用预定义的特征模板来描述图像的特征。

Haar-like 特征可以捕获线性特征、边缘特征、点特征和对角线特征。早期研究 [54; 55] 证明了 Haar-like 特征对于 SAR 目标检测的鲁棒性。一种名为 MSRIHL [1] 的最新方法将低级 Haar-like 特征集成到深度学习模型中，以实现准确的 SAR 目标检测，突显了其潜在的有效性。

小波散射变换 [45] (WST) WST 是一种强大的信号处理技术，广泛应用于图像处理。它旨在同时提取鲁棒且具有区分性的低级和高级特征。通过同时捕获高频和低频信息，它提供了局部和全局图像特征的丰富表示。小波散射变换提供的分层表示能够实现不同尺度和分辨率的特征提取，这对于捕获 SAR 图像中小物体的精细细节以及鲁棒地处理高频噪声特别有用。Vidal 等人 [61] 进行了一项全面的研究，证明了小波变换在 SAR 图像去噪方面的巨大潜力。许多最近的工作 [93; 26; 50; 78] 利用小波变换或 WST 进行鲁棒的特征提取，并与 CNN 网络集成以进行目标识别。这些研究验证了将 WST 特征融入现代 CNN 模型的可行性。

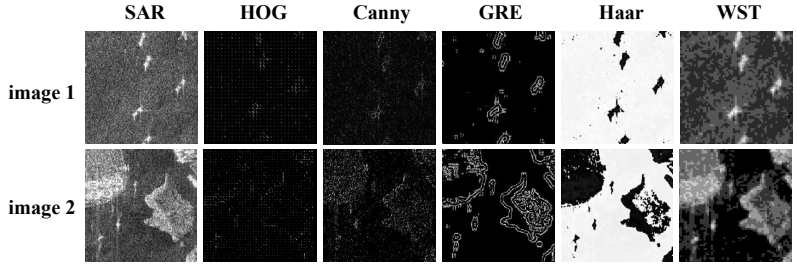


图 S7: SAR 图像上手工特征的可视化。（为了便于可视化，这些特征经过平均池化并表示为单通道。）

A.3 多阶段滤波器增强

图 S8 提供了所提出的滤波器增强数据输入的直观图示。它涉及将原始单通道灰度合成孔径雷达 (SAR) 图像（表示为 x）与滤波器增强表示 M_i^x 连接起来。

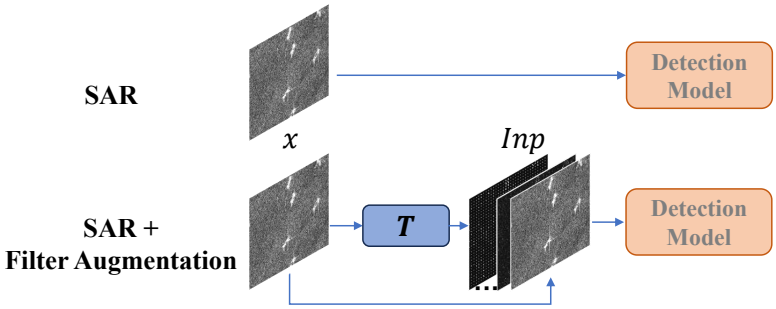


图 S8: 滤波器增强输入的图示。

表 S7: 使用 Faster R-CNN 和 Resnet-50 作为检测模型，对不同滤波器增强输入的比较。

方法	**mAP**	@50	@75	@s	@m	@l
SAR (as RGB)	50.2	83.0	54.8	44.8	61.6	58.6
SAR+Canny	50.7	83.6	55.0	45.3	62.0	57.1
SAR+Hog	50.7	83.5	55.2	45.1	61.4	**58.4**
SAR+Haar	50.6	83.4	54.7	45.4	61.6	58.0
SAR+WST	**51.1**	83.9	54.7	45.2	**62.3**	57.5
SAR+GRE	50.6	83.8	54.7	44.8	61.7	57.6
SAR+Hog+Haar+WST	**51.1**	**84.0**	**55.9**	**45.7**	62.0	58.2

A.4 实验结果

A.4.1 SARDet-100K 上的主要结果.

详细的实验结果呈现在表 S7、表 S8 和表 S9 中，其中提供了全面、细粒度的测试指标。这些指标包括 AP@50、AP@75、AP@small (AP@s)、AP@medium (AP@m) 和 AP@large (AP@l)，为模型评估提供了更深入的见解和稳健的结果。

表 S7 是主论文中表 3 的扩展，展示了使用 Faster R-CNN 和 ResNet-50 作为检测模型，对各种滤波器增强输入的比较。

表 S8 扩展了主论文的图 6(a)，表 S9 扩展了图 6(b)。这些表格主要关注探索 MSFA 框架在不同检测框架和骨干网络上的通用性的实验。值得注意的是，在各种框架（包括单阶段、两阶段和基于查询的框架）以及各种骨干网络架构（从 ResNets 和现代设计的 ConvNets 到 Vision Attention Networks (VANs) 和基于 Vision Transformer (ViT) 的 Swin 网络）中，都观察到性能的显著提升。这些结果为我们提出的方法的有效性和广泛适用性提供了令人信服的证据。

为了确保公平的比较，我们在相似的计算预算下评估 MSFA 和原始检测模型。在表 S11 中，我们训练了 Faster-RCNN 模型（使用 ResNet-50 ImageNet-1K 预训练骨干网络），在 DOTA 数据集上进行了 12 个 epoch 的 MSFA 预训练，然后在 SARDet-100K 数据集上进行了微调。这与直接在 SARDet-100K 数据集上对模型进行 36 个 epoch 微调进行了比较。在相同的总训练 epoch 数和略少的总迭代次数下，所提出的 MSFA 在 SARDet-100K 测试集上实现了显著更高的 mAP 结果（高 1.7%）。MSFA 预训练结合了来自自然和遥感数据集的通用知识，这使得高效且有效的微调成为可能，并有助于减轻下游任务中的过拟合。同样重要的是要注意，

表 S8: MSFA 在不同检测框架上的泛化性。INP：仅在骨干网络上进行传统的 ImageNet 预训练。

框架		预训练	测试					
			mAP	@50	@75	@s	@m	@l
两阶段	Faster RCNN [53]	INP	49.0	82.2	52.9	43.5	60.6	55.0
		MSFA	51.1 (+2.1)	83.9	54.7	45.2	62.3	57.5
	Cascade RCNN [4]	INP	51.1	81.9	55.8	44.9	62.9	60.3
		MSFA	53.9 (+2.8)	83.4	59.8	47.2	66.1	63.2
	Grid RCNN [4]	INP	48.8	79.1	52.9	42.4	61.9	55.5
		MSFA	**51.5** (+2.7)	81.7	56.3	45.1	64.1	60.0
单阶段	RetinaNet [33]	INP	47.4	79.3	49.7	40.0	59.2	57.5
		MSFA	49.0 (+1.6)	80.1	52.6	41.3	61.1	59.4
	GFL [30]	INP	49.8	80.9	53.3	42.3	62.4	58.1
		MSFA	53.7 (+3.9)	84.2	57.8	47.8	66.2	59.5
	FCOS [59]	INP	46.5	80.9	49.0	41.1	59.2	50.4
		MSFA	48.5 (+2.0)	82.1	51.4	42.9	60.4	56.0
端到端	DETR [6]	INP	31.8	62.3	30.0	22.2	44.9	41.1
		MSFA	47.2 (+15.4)	77.5	49.8	37.9	62.9	58.2
	Deformable DETR [94]	INP	50.0	85.1	51.7	44.0	65.1	61.2
		MSFA	51.3 (+1.3)	85.3	54.0	44.9	65.6	61.7
	Sparse RCNN [58]	INP	38.1	68.8	38.8	29.0	51.3	48.7
		MSFA	41.4 (+3.3)	74.1	41.8	33.6	53.9	53.4
	Dab-DETR [39]	INP	45.9	79.0	47.9	38.0	61.1	55.0
		MSFA	48.2 (+2.3)	81.1	51.0	41.2	63.1	55.4

MSFA 预训练是一次性的工作。预训练的 MSFA 模型可以重复用于微调不同的 SAR 检测数据集。

A.4.2 SARDet-100K 与其他数据集的比较

为了评估所提出的 SARDet-100K 数据集作为大规模 SAR 目标检测基准的质量，我们评估了该数据集上的不同模型，并将结果与其他流行的基准数据集（例如 SSDD [84] 和 HRSID [71]）的结果进行了比较。结果如图 S9 所示。这些结果表明，现代模型在我们数据集上的性能尚未达到饱和。在各种评估的模型中，最弱模型和最强模型之间存在 8.4% 的性能差距，而对于 SSDD 和 HRSID，这一差距分别仅为 4.1% 和 4.3%。这表明 SSDD 和 HRSID 对于大多数现有模型来说相对简单，导致这些较小数据集上的性能饱和。

此外，我们观察到在这些较小的数据集上，参数数量相对较多的模型往往会因过拟合而导致性能下降。例如，ResNet-152 在 SSDD 上的性能不如 ResNet-101，而 ResNet-101 在 HRSID 上的性能不如 ResNet-50 和 ResNet-18。然而，在大型 SARDet-100K 上，这个问题不会出现。在我们更大的数据集上，模型继续从模型尺寸的增加中获益，这表明我们提出的数据集适用于开发用于大规模 SAR 目标检测的相对较大的模型。

表 S9: MSFA 在不同检测骨干网络上的泛化性。INP：仅在骨干网络上进行传统的 ImageNet 预训练。

架构	#P(M)	预训练	测试					
			mAP	@50	@75	@s	@m	@l
R50 [23]	25.6	INP	49.0	82.2	52.9	43.5	60.6	55.0
		MSFA	51.1 (+2.1)	83.9	54.7	45.2	62.3	57.5
R101 [23]	44.7	INP	51.2	84.1	55.6	45.9	61.9	56.3
		MSFA	52.0 (+0.8)	84.6	56.6	46.6	63.4	57.7
R152 [23]	60.2	INP	51.9	85.2	55.9	46.4	62.5	57.9
		MSFA	52.4 (+0.5)	85.4	57.2	47.4	63.3	58.7
ConvNext-T [42]	28.6	INP	53.2	86.3	58.1	47.2	65.2	59.6
		MSFA	54.8 (+1.6)	87.1	59.8	48.8	66.7	62.1
ConvNext-S [42]	50.1	INP	54.2	87.8	59.2	49.2	65.8	59.8
		MSFA	55.4 (+1.2)	87.6	60.7	50.1	67.1	61.3
ConvNext-B [42]	88.6	INP	55.1	87.8	59.5	48.9	66.9	61.1
		MSFA	56.4 (+1.3)	88.2	61.5	51.1	68.3	62.4
VAN-T [22]	4.1	INP	45.8	79.8	48.0	38.6	57.9	53.3
		MSFA	47.6 (+1.8)	81.4	50.6	40.5	59.4	56.7
VAN-S [22]	13.9	INP	49.5	83.8	52.8	43.2	61.6	56.4
		MSFA	51.5 (+2.0)	85.0	55.6	44.8	63.4	60.4
VAN-B [22]	26.6	INP	53.5	86.8	58.0	47.3	65.5	60.6
		MSFA	55.1 (+1.6)	87.7	60.2	48.8	67.3	62.2
Swin-T [41]	28.3	INP	48.4	83.5	50.8	42.8	59.7	55.7
		MSFA	50.2 (+1.8)	84.1	53.9	44.1	61.3	58.8
Swin-S [41]	49.6	INP	53.1	87.3	57.8	47.4	63.9	60.6
		MSFA	54.0 (+0.9)	87.0	59.2	48.2	64.5	61.9
Swin-B [41]	87.8	INP	53.8	87.8	59.0	49.1	64.6	60.0
		MSFA	55.7 (+1.9)	87.8	61.4	50.5	66.5	62.5

表 S10: DOTA 和 DIOR 之间的数据集统计比较。*：多尺度预处理。

数据集	图片	实例	分类	图片大小	平均实例区域
DOTA*	68,324	1,058,641	15	1024*1024	5,021
DIOR	23,463	192,518	20	800*800	12,726

表 S11: 在相似的计算预算下，MSFA 和微调性能的比较。INP：骨干网络在 ImageNet 上进行了预训练。

模型	INP	MSFA Epoch	Finetune Epochs	Total Epochs	Total Iterations	mAP
Faster-RCNN [53]	✓	12	24	36	16.1k	54.5
Faster-RCNN [53]	✓	0	36	36	17.7k	52.8

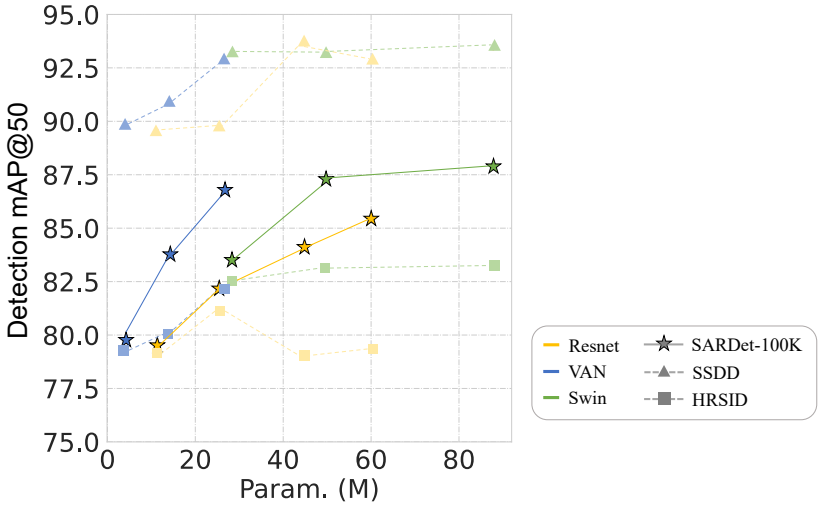


图 S9: 在 SARDet-100K 数据集和其他先前流行的基准（SSDD [84] 和 HRSID [71]）上评估不同的骨干模型。骨干网络插入到 Faster-RCNN [53] 检测架构。

A.5 实现细节

对于 ImageNet 预训练，我们在 Imagenet-1K 上采用 100 个 epoch 的骨干网络预训练策略，并在 MMPretrain [11] 配置中使用其默认训练策略。

对于所提出的多阶段预训练的第二阶段，我们选择大规模光学遥感 DOTA 数据集作为主要数据集。我们还在 DIOR 数据集上进行了比较实验，以找出第二阶段预训练数据集对下游性能的影响因素。它们的详细信息如表 S10 所示。

对于 DOTA 数据集，我们执行额外的数据集预处理，因为 DOTA 数据集图像具有很大范围的不同图像分辨率。原始图像仅包含 1,411 张训练图像。为了获得更多的图像和多尺度实例以进行有效的训练，我们遵循 [31] 采用多尺度数据集分割策略，将原始高分辨率图像重新缩放到三个不同的尺度 (x0.5, x1.0, x1.5)，然后将每个缩放后的图像裁剪成 1024 × 1024 的 patches，每个 patch 的重叠像素为 500 像素，以避免在 patch 边界上对实例进行破坏性划分。使用预处理的数据集，加载调整大小为 1024 × 1024，RandomFlip 概率为 0.5。对于 DIOR 数据集，我们通过将图像大小调整为 800*800 来训练模型，RandomFlip 概率为 0.5。

为了在 SARDet-100k、SSDD 和 HRSID 上进行微调，我们通过将图像大小调整为 800 × 800 进行训练，RandomFlip 概率为 0.5。我们通过在训练集上训练模型 12 个 epoch，并使用第 12 个 epoch 的检查点在测试集上测试模型。

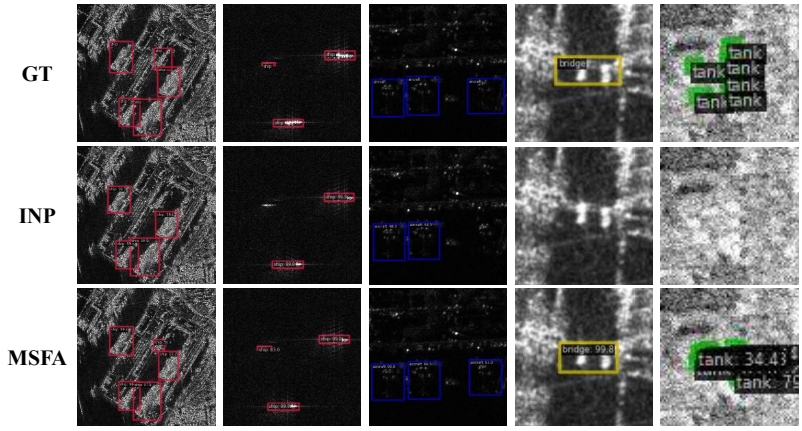
我们主要使用 MMPretrain [11] 和 MMDetection [8] 框架在 8 个 RTX-3090 GPU (24G) 上进行实验。有关超参数和训练设置的详细信息，请参阅表 S12。

表 S12: 预训练和微调设置的超参数。Cls.：分类，Det.：检测，B.S.：batch size，L.R.：学习率。

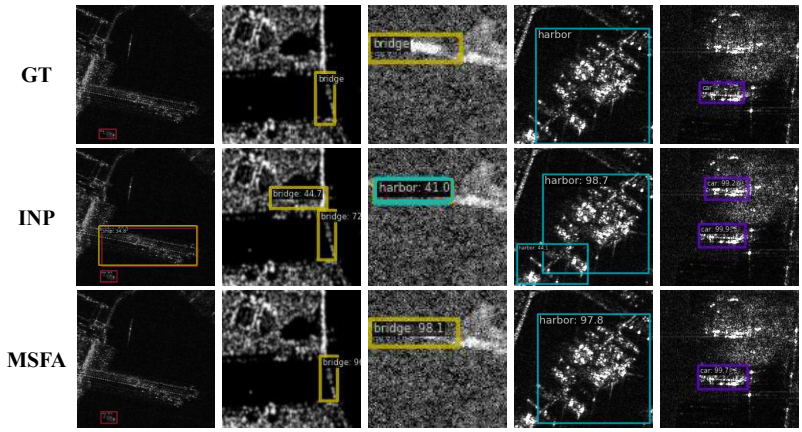
任务 / 模型	数据集	Optim.	B.S.	L.R	Epochs
Cls. Pretrain	ImageNet	AdamW	512	1e-8	100
Det. Pretrain	DOTA	AdamW	16	1e-4	12
Det. Pretrain	DIOR	AdamW	16	1e-4	12
Det. Finetune	SARDet-100k	AdamW	16	1e-4	12
Det. Finetune	SSDD	AdamW	32	2.5e-4	12
Det. Finetune	HRSID	AdamW	32	2.5e-4	12
DETR	DOTA/SARDet-100k	AdamW	16	1e-4	150
Deformable-DETR	DOTA/SARDet-100k	AdamW	16	2e-4	50
Dab-DETR	DOTA/SARDet-100k	AdamW	16	1e-4	50
Sparse-RCNN	DOTA/SARDet-100k	AdamW	16	2.5e-5	12

A.6 检测结果可视化

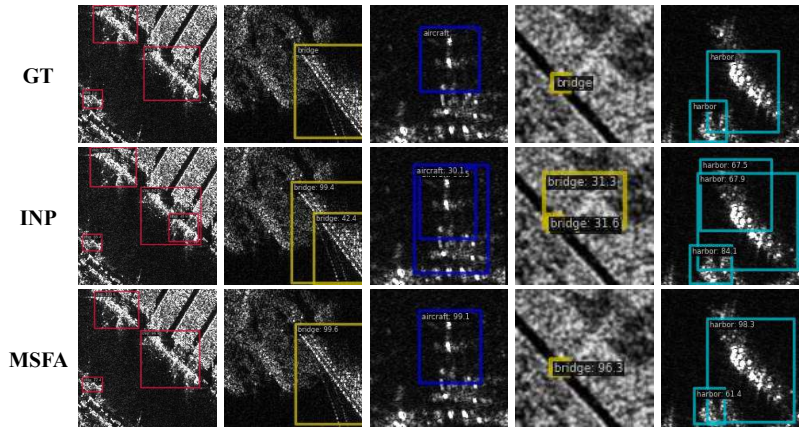
检测可视化结果（将 MSFA 框架与传统的 ImageNet 预训练方法进行比较）如图 S10 所示。这些结果表明，在减少漏检、误检和提高定位精度方面，MSFA 的性能优于传统的 ImageNet 骨干网络预训练。



(a) Missing detection



(b) False detection



(c) Inaccurate localization

图 S10: MSFA 优于传统的 ImageNet 骨干网络预训练，体现在 (a) 漏检，(b) 误检和 (c) 不准确的定位

A.7 失败场景

然而，当前的模型并非没有缺点和不足之处。图 S11 突出显示了几个失败的场景。当输入的 SAR 图像缺少可识别的细节或上下文信息时，可能会导致不正确的分类。当 SAR 图像包含小的且密集堆积的物体时，模型可能无法检测到其中一些物体。以褪色、模糊或整体低分辨

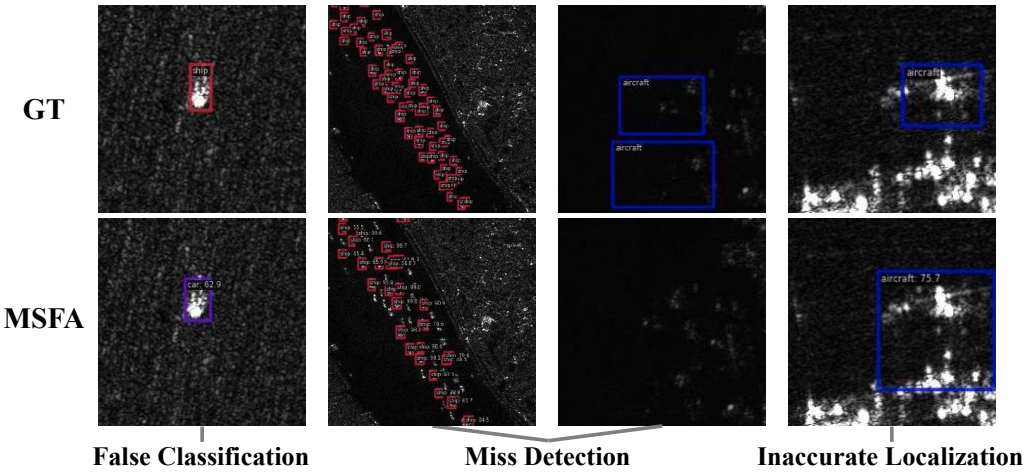


图 S11: 失败案例的可视化：错误分类、漏检、不准确的定位。

表 S13: ConvNext-B MSFA

分类	mAP	@50	@75	@s	@m	@l
ship	0.669	0.923	0.783	0.653	0.714	0.555
aircraft	0.455	0.753	0.466	0.419	0.458	0.47
car	0.655	0.985	0.792	0.553	0.674	n/a
tank	0.454	0.766	0.427	0.429	0.891	n/a
bridge	0.441	0.885	0.368	0.411	0.609	0.714
harbor	0.711	0.979	0.858	0.601	0.751	0.756
Average	0.564	0.882	0.616	0.511	0.733	0.624

率为特征的图像质量不佳，进一步加剧了漏检的风险。在具有挑战性的情况下，检测器也可能难以进行准确定位。

关于细粒度类别检测性能，表 S13 展示了使用 ConvNext-B [42] 和 MSFA 预训练的 Faster-RCNN [53] 的检测结果。值得注意的是，对于以下物体，该模型表现出相对较低的性能：

- 小尺寸的物体，例如 Tank（平均面积为 461 像素）
- 长宽比大的物体，例如 Bridge
- 外观变化性高的物体，例如 Airplane

然而，这项工作的主要重点是解决 SAR 目标检测预训练和微调中存在的领域和模型差距。我们的方法证明了与大多数现有深度网络的出色兼容性，并且可以与专门为解决上述具有挑战性的场景而设计的模型无缝集成。