E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
©中国图象图形学报版权所有

中图法分类号:TP18;TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(2026)09-3120-24

论文引用格式:Chen B Y, Guo C R, Shan C, Zhang L and Wei W. 2026. Research progress of cross-domain few-shot object detection in aerial remote sensing scenes. Journal of Image and Graphics, 31(9):3120-3143(陈布衣, 郭宸瑞, 单淳, 张磊, 魏巍. 2026. 航空遥感场景跨域少样本目标检测进展研究. 中国图象图形学报, 31(9):3120-3143)[DOI:10.11834/jig.260206]

航空遥感场景跨域少样本目标检测进展研究

陈布衣[1],郭宸瑞[2,3*],单淳[1],张磊[1],魏巍[1]

1. 西北工业大学计算机学院,西安　710000; 2. 宇航智能控制技术全国重点实验室,北京　100854;
3. 北京航天自动控制研究所,北京　100854

摘　要:航空遥感目标检测在国土资源调查、灾害监测、交通监管和军事侦察等领域具有重要应用价值。然而,真实航空遥感场景中常同时存在目标域标注样本稀缺和源域与目标域分布差异显著的问题,导致传统依赖大规模同分布标注数据的检测方法难以获得稳定性能。跨域少样本目标检测(cross-domain few-shot object detection,CD-FSOD)期望在目标域仅有少量标注样本且存在明显域偏移的条件下实现对目标域目标的有效分类与定位,已成为航空遥感智能解译中的重要研究方向。本文围绕航空遥感场景下的CD-FSOD问题,梳理其基本定义、任务边界及其与少样本目标检测、跨域目标检测和遥感目标检测等相关任务的区别,并分析航空遥感场景中由成像平台、传感器类型、空间分辨率、地理区域、环境条件及目标分布差异所引起的复杂域偏移表现。在此基础上,对现有代表性方法进行分类综述,重点总结基于迁移与域对齐、基于元学习与度量学习、基于生成式数据增强与扩充、基于视觉语言对齐以及基于大模型的方法特点与研究进展。同时,归纳相关数据集、任务设置、常用评测指标及航空遥感目标检测的特殊性,并进一步讨论复杂域偏移统一建模、少样本条件下检测稳定性、数据集与真实应用脱节、多源异构数据适配以及大模型迁移等方面的主要问题与发展趋势。总体来看,CD-FSOD正由单一迁移或少样本学习策略,逐步发展为融合域适应、特征增强、语义建模和基础模型迁移的综合性研究方向。随着遥感专用数据资源的丰富和多模态基础模型的发展,该方向有望为航空遥感智能感知提供更高效、更稳健的技术支撑。本文同步汇总代表性算法在航空遥感跨域少样本情境下的性能表现,并给出所涉及数据集、算法的开放资源访问链接:https://github.com/Farenweh/CD-FSOD-Links(Github)或https://gitee.com/cby1241385936/cd-fsod-links(Gitee)。

关键词:航空遥感;目标检测;跨域少样本目标检测(CD-FSOD);视觉语言对齐;大模型

Research progress of cross-domain few-shot object detection in aerial remote sensing scenes

Chen Buyi[1], Guo Chenrui[2,3*], Shan Chun[1], Zhang Lei[1], Wei Wei[1]

1. School of Computer Science, Northwestern Polytechnical University, Xi'an　710000, China; 2. State Key Laboratory of Aerospace Intelligent Control Technology, Beijing　100854, China; 3. Beijing Aerospace Automatic Control Institute, Beijing　100854, China

Abstract: Cross-domain few-shot object detection (CD-FSOD) in aerial remote sensing aims to detect target objects when the target domain contains only a few annotated instances and differs substantially from the source domain. This setting is important for various applications (e. g., land resource investigation, disaster monitoring, traffic supervision, maritime

收稿日期:2026-04-10;**修回日期**:2026-06-06;**预印本日期**:2026-06-13

*通信作者:郭宸瑞　g-ch-r@163.com

基金项目:国家自然科学基金项目(62372379, 62472359)

Supported by: National Natural Science Foundation of China (62372379, 62472359)

and airport surveillance, emergency response, and military reconnaissance), because many operational remote sensing systems must be deployed in new regions, seasons, sensors, or platforms before large-scale annotation is available. Typically, conventional object detectors assume abundant labeled data and a relatively stable training-test distribution. In contrast, aerial images are influenced by platform altitude, satellite or unmanned aerial vehicle (UAV) imaging geometry, optical or synthetic aperture radar (SAR) sensing mechanisms, spatial resolution, illumination, weather, seasonal background, geographical scene composition, and category occurrence frequency. These factors change target scale, texture, orientation, density, and contextual appearance at the same time. Consequently, detectors trained on a well-annotated source domain may suffer from severe classification confusion, localization drift, false alarms, and missed detections after being transferred to a new target domain. Related to this, the current paper provides a systematic survey of CD-FSOD for aerial remote sensing scenes. First, the task definition and research boundary are clarified. In particular, CD-FSOD is distinguished from classic object detection, few-shot object detection (FSOD), cross-domain object detection, and general remote sensing object detection. Classic object detection emphasizes classification and localization under roughly independent and identically distributed data. FSOD focuses on new-category detection with limited examples but often assumes that the base and novel domains are close. Meanwhile, cross-domain object detection addresses domain shift but usually does not impose an extreme target-domain annotation constraint. Finally, remote sensing object detection describes the application scenario and its characteristic difficulties, including small objects, dense object distributions, arbitrary orientations, and complex backgrounds. CD-FSOD integrates the last two difficulties in which the target domain is both data-scarce and distributionally shifted. Second, existing methods are organized into five technical routes. Transfer and domain-alignment methods (i. e. , domain-adaptive faster region-based convolutional neural network (DA Faster R-CNN), strong-weak distribution alignment (SWDA), spatial attention pyramid network (SAPNet), and adaptive teacher frameworks) reduce the discrepancy between source and target features by image-level, feature-level, instance-level, or teacher-student constraints. Meta-learning and metric-learning methods, including meta region-based convolutional neural network (Meta R-CNN), attention region proposal network (Attention-RPN), few-shot object detection via contrastive proposal encoding (FSCE), Detect everything with vision Transformer (DE-ViT), and CD-ViTO, all improve the use of support samples through episodic learning, class prototypes, relation modeling, and contrastive representation learning. Meanwhile, generative augmentation methods, such as hallucination-based feature synthesis, multiperspective data augmentation (MPAD), AeroGen, Control Copy-Paste, and Domain-RAG, expand the target-domain sample distribution by generating features, foreground instances, target layouts, or domain-consistent backgrounds. To compensate for insufficient visual examples, vision-language alignment methods, including RegionCLIP, PromptDet, Grounding DINO, and rich-text cross-domain multimodal FSOD, introduce category names, attributes, scene descriptions, and professional textual semantics. Furthermore, large-model methods exploit visual foundation models, detection transformer (DETR) variants, and multimodal large models (e. g. , Rex-Omni) to provide broader open-vocabulary recognition and stronger initial representations. Third, the paper summarizes datasets, task protocols, and evaluation metrics. The reviewed benchmarks include multidomain few-shot object detection (MoFSOD), CD-FSOD, object detection in optical remote sensing images (DIOR), dataset for object detection in aerial images (DOTA), xView, vehicle detection in aerial imagery (VEDAI), and the Northwestern Polytechnical University very-high-resolution 10-class dataset (NWPU VHR-10). Their differences can be detected in terms of image number, category number, bounding-box type, scene coverage, object scale, and whether the protocol stresses a balanced K-shot instance setting. Common metrics include precision (P), recall (R), average precision (AP), mean average precision (mAP), AP50, AP75, mAP@50:95, novel-class mAP, and intersection over union (IoU). Instead of listing formulae for routine metrics, this survey discusses their specific meanings and how they should be used to evaluate target-domain generalization, shot sensitivity, robustness across domains, and localization quality for small dense objects. Fourth, this review identifies several open problems. Notably, complex domain shift should be modeled at global-scene, local-object, and category-conditional levels rather than by a single overall feature alignment. Few-shot robustness requires support-sample quality assessment, stable prototype learning, multiscale representation, uncertainty estimation, and joint optimization of classification and localization. Current datasets and evaluation protocols remain partly disconnected from real remote sensing applications, in which categories are long-tailed, annotations are incomplete,

sensors are heterogeneous, and deployment constraints are strict. To date, multisource adaptations across optical, infrared, multispectral, and SAR data remain underexplored. While foundation models and multimodal large models offer stronger semantic priors, they also introduce problems, such as high computational cost, insufficient remote sensing domain knowledge, and limited performance on small, dense, rotated, or sensor-specific targets. Therefore, to improve reuse and citation of the survey materials, a curated performance summary of representative algorithms under the aerial remote sensing CD-FSOD setting is provided in the manuscript. Overall, CD-FSOD in aerial remote sensing is moving from isolated fine-tuning or domain-alignment strategies toward integrated frameworks that combine domain adaptation, few-shot discrimination, data generation, vision-language semantics, and foundation-model transfer. In terms of practical use, it is emphasized that CD-FSOD should not be evaluated only by a single accuracy score obtained under a fixed split. Instead, a more informative protocol should report the stability of each method under different shots, target domains, object scales, and sensor conditions. This is because an algorithm that performs well on one aerial dataset may fail when the dominant background, ground sampling distance, or object density changes. Therefore, this survey recommends that future benchmark construction record the source of domain shift explicitly and separate the effects of category novelty, image style, spatial resolution, and annotation scarcity. Such analysis is especially necessary for aerial remote sensing, in which missed small targets and inaccurate localization may be more damaging than a modest decrease in overall mAP. Finally, future progress depends on remote-sensing-specific pretraining resources, reliable benchmark construction, efficient adaptation for high-resolution imagery, and stronger integration between precise detectors and general-purpose foundation models. This is maintained through the open resource page found in https://github.com/Farenweh/CD-FSOD-Links(Github) or https://gitee.com/cby1241385936/cd-fsod-links(Gitee).

Key words: aerial remote sensing; object detection; cross-domain few-shot object detection(CD-FSOD); vision-language alignment; large model

0 引 言

目标检测是航空遥感智能解译中的基础任务之一,其核心任务是在航空遥感图像中定位和识别感兴趣目标,并为后续信息解译与推理提供基础支撑,相关技术在资源勘探、灾害救援、目标监视和情报侦察等领域具有重要应用价值。与自然场景图像相比,航空遥感图像通常具有视角俯视、目标尺度小、背景复杂、目标密集分布以及方向变化显著等特点(石争浩 等,2023),对目标检测算法的特征表征和定位能力提出了更高要求。

近年来,深度学习方法显著推动了目标检测技术的发展,以快速区域卷积神经网络(faster region-based convolutional neural network,Faster R-CNN)为代表的两阶段检测框架通过区域建议网络与检测网络的联合优化,有效提升了目标定位和分类的性能(Ren等,2017)。但现有主流检测模型通常依赖大规模标注数据和相对一致的数据分布,例如通用对象上下文(common objects in context,COCO)等通用目标检测数据集包含大量实例级标注以支撑面向自然场景图像的检测模型学习(Lin等,2014)。在真实航空遥感应用中往往面临样本获取困难、标注成本高以及训练测试分布不一致等问题(袁翔 等,2023),导致模型泛化性能受限(Chen等,2018)。

在实际航空遥感任务中,数据分布差异常由多种因素共同造成。例如,不同成像平台、传感器类型、空间分辨率、地理区域、季节气候和光照条件都会引起显著的域偏移,而域偏移会使源域训练得到的目标检测模型在目标域中出现明显性能退化(Chen等,2018);同时,一些重点目标类别在新场景中仅能获得少量标注样本,甚至只能提供极少的支持样本进行快速适配,这与少样本目标检测所关注的低样本新类别检测问题密切相关(Yan等,2019)。在此背景下,跨域少样本目标检测(cross-domain few-shot object detection,CD-FSOD)逐渐成为航空遥感目标检测研究中的重要方向。该任务旨在同时应对“域偏移”和“样本稀缺”两类挑战,在源域数据较为充足、目标域标注样本有限的条件下,实现对目标域目标的有效检测(Xiong,2023)。因此,跨域少样本目标检测不仅是对传统目标检测方法的延伸,也是提升航空遥感智能感知系统实用性与鲁棒性的关键途径。

围绕上述问题,有研究从迁移学习、域对齐、元

学习、度量学习、数据增强、视觉语言对齐以及大模型迁移等多角度展开探索。基于迁移与域对齐的方法主要通过缩小源域与目标域之间的分布差异提升模型跨场景泛化能力，该方向具有代表性的早期工作包括域自适应 Faster R-CNN（domain-adaptive Faster R-CNN，DA Faster R-CNN）（Chen 等，2018）等直接在标准检测器中应用域自适应技术的方法；基于元学习与度量学习的方法强调从少量样本中快速建立类别判别能力，例如元学习区域卷积神经网络（meta region-based convolutional neural network，Meta R-CNN）将元学习机制引入实例级少样本目标检测任务并提升了模型对新类别的适应能力（Yan 等，2019），也有工作通过增强候选区域特征的类内紧致性和类间可分性，缓解少样本条件下新类别易混淆的问题（Sun 等，2021）；基于生成式数据增强与扩充的方法尝试利用样本合成、风格迁移或特征扩增等手段缓解目标域样本不足问题（Zhang 和 Wang，2021）；近年来，视觉语言模型和大规模预训练模型的发展又为 CD-FSOD 提供了新的通用表征和语义先验，少样本目标检测任务逐渐从传统检测器微调转向基础模型适配（Madan 等，2024）。总体来看，相关方法不断演进，但在复杂域偏移建模、遥感场景专用知识注入以及真实应用适配等方面仍存在不少挑战。

基于此，本文聚焦航空遥感场景下的跨域少样本目标检测问题，重点综述该方向的任务定义、关键方法、数据集与评测方式，并分析其在真实应用中的难点与发展趋势。将现有方法大致归纳为 5 类：基于迁移与域对齐的方法、基于元学习与度量学习的方法、基于生成式数据增强与扩充的方法、基于视觉语言对齐的方法，以及基于大模型的方法。通过对不同方法核心思想、适用条件和局限性的对比分析，本文力图为航空遥感目标检测研究提供一条较为清晰的技术脉络。

1　跨域少样本目标检测的任务定义与问题特点

跨域少样本目标检测（CD-FSOD）面向目标域样本稀缺且源域与目标域存在显著分布差异的应用场景，是少样本学习与跨域迁移在目标检测任务中的交叉延伸。该任务需要在有限监督条件下完成目标分类与定位，还需应对成像平台、传感器类型、空间分辨率及场景背景变化所带来的域偏移问题。本节明确其基本概念、任务边界及航空遥感场景中的域差异表现。

1.1　基本定义

跨域少样本目标检测是少样本目标检测与跨域目标检测相结合的一类目标检测任务，其核心目标是在目标域标注样本极少且源域与目标域存在明显分布差异的条件下，实现对目标域目标的有效检测。与传统目标检测依赖大规模同分布标注数据不同，该任务同时面临“样本稀缺”和“域偏移”两方面挑战，因此对模型的迁移能力、适应能力和泛化能力提出了更高要求。

少样本目标检测通常是指在目标类别仅有少量标注样本的条件下完成检测任务。现有研究中，常将样本数限定为每类 K 个标注样本，即所谓 K-shot 设置；当类别数为 N 时，表述为 N-way K-shot 任务。该类方法重点关注如何从有限样本中学习具有判别性的类别表示，以缓解检测器对大量标注数据的依赖。

跨域目标检测则强调训练数据与测试数据之间存在域分布差异。一般将标注较充分的数据分布称为源域，将待适应的数据分布称为目标域。所谓“域”，是由成像平台、传感器类型、空间分辨率、拍摄视角、地理区域、天气光照和场景背景等因素共同决定的数据分布。在航空遥感场景中，这种域差异往往表现得更加明显，因此直接将源域训练得到的检测模型应用于目标域时，经常会出现检测精度显著下降的问题。图 1 从特征分布层面展示了跨域如何影响检测器性能。

综合来看，跨域少样本目标检测可定义为：在源域具有相对充足标注数据、目标域仅有少量标注样本，且两域之间存在显著分布偏移的条件下，利用源域知识和目标域有限样本，实现对目标域目标类别的准确分类与定位。该任务既要求模型能够有效迁移源域已有知识，又要求其能够在有限监督下快速适应新域分布，是航空遥感智能感知中具有代表性的低资源目标检测问题。

1.2　与相关任务的区别

跨域少样本目标检测并不是对现有目标检测任务的简单叠加，而是在少样本学习、跨域迁移与目标检测三者交叉基础上形成的一类更具挑战性的任

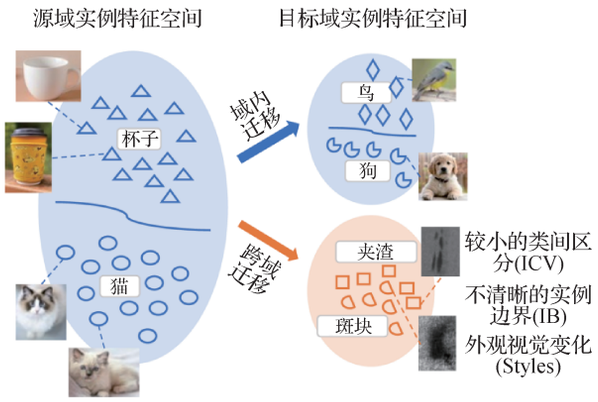


图1　Li等人(2025)所提跨域少样本目标检测定义与挑战

Fig. 1　Definition and challenges of cross-domain few-shot object detection given by Li et al. (2025)

务。为避免概念混淆,有必要将其与传统目标检测、少样本目标检测、跨域目标检测以及航空遥感目标检测等相关任务加以区分。

传统目标检测通常假设训练集与测试集满足近似独立同分布条件,并依赖较大规模的标注样本进行模型训练,其关注重点主要在于提升分类与定位精度。而少样本目标检测将研究重点转向样本稀缺条件下的检测能力,强调如何利用极少标注样本完成新类别检测,但通常默认训练域与测试域之间不存在明显分布偏移。跨域目标检测则主要面向源域与目标域之间的数据分布差异问题,重点在于通过域适应、域对齐或知识迁移减弱域偏移带来的性能退化,但多数方法仍假设目标域具有相对充分的数据支持。相比之下,跨域少样本目标检测需要同时处理"目标域样本极少"和"源域与目标域分布不一致"两类问题,其任务难度和方法设计复杂度均高于上述单一设定。

此外,航空遥感目标检测是一个应用场景层面的概念,强调检测对象来自航空遥感或卫星遥感图像,主要关注小目标、密集目标、旋转目标和复杂背景等遥感特有问题。跨域少样本目标检测则是一个更具体的任务设定,其研究重点是遥感场景中由于平台、分辨率、区域和成像条件变化所引起的域偏移,以及由标注困难带来的少样本约束。

为便于比较,表1和图2对跨域少样本目标检测与相关任务的区别进行了总结。

总体而言,跨域少样本目标检测的核心特征在于双重困难耦合:一方面,目标域可用监督信息极为有限;另一方面,源域知识又难以直接迁移到目标域。正因如此,该任务不能简单套用少样本目标检测或跨域目标检测的已有方法,而需要在域不变表征学习、快速适应机制和有限样本利用策略之间进行联合设计。

1.3　航空遥感场景中的域差异表现

与自然场景图像相比,航空遥感数据的获取方式更加复杂,成像条件和场景属性变化更加显著,因此"域差异"在该类任务中表现得更为突出。所谓域差异,本质上是指不同数据来源在成像特征、目标外观、背景分布和标注方式等方面存在系统性偏移,这种偏移会导致在源域上训练得到的检测模型难以直接适用于目标域。对于跨域少样本目标检测而言,准确识别航空遥感场景中的域差异来源,是理解任务难点和设计有效方法的基础。

首先,成像平台与传感器差异是最常见的域偏移来源。航空遥感数据可能来自卫星、无人机或机载平台等不同载体。不同平台在飞行高度、拍摄角度、成像范围和稳定性方面存在明显差异;同时,可见光、红外、多光谱以及合成孔径雷达(synthetic aperture radar, SAR)等不同传感器在纹理、光谱响应和成像机理上也不相同,其中SAR等主动成像传感器尤为典型。这类差异会直接改变目标的外观特征和背景分布,从而影响模型已学习到的表征能力。

表1　跨域少样本目标检测与相关任务的区别

Table 1　Differences between cross-domain few-shot object detection and related tasks

任务类型	域偏移	仅有少量新类样本	主要目标	典型关注点
经典目标检测	否	否	在同分布数据上实现高精度检测	分类与定位精度、检测效率
少样本目标检测	否	是	利用少量样本完成新类别检测	类别表征、快速适配、样本效率
跨域目标检测	是	否	缓解源域与目标域分布差异带来的性能下降	域对齐、特征迁移、分布适应
航空遥感目标检测	可能	否	面向航空遥感图像中的目标检测	小目标、密集目标、旋转目标
跨域少样本目标检测	是	是	在域偏移与样本稀缺双重约束下实现有效检测	域迁移、少样本适配、泛化能力

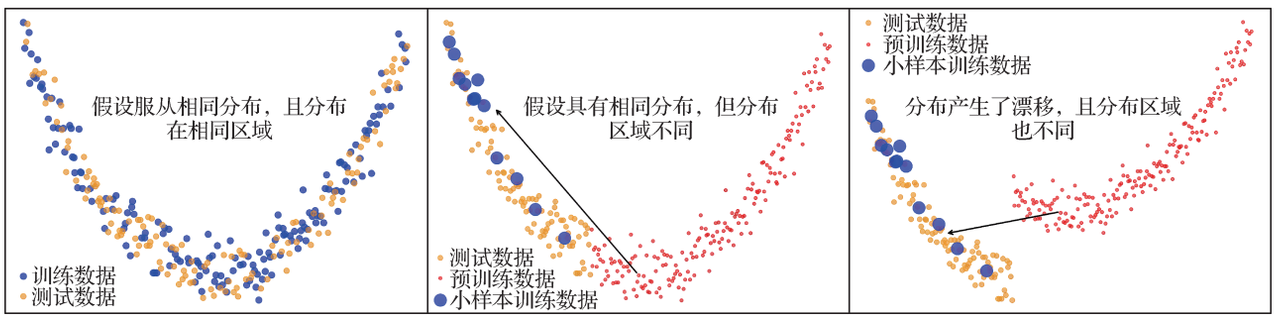


图 2　经典深度学习目标检测、少样本目标检测和跨域少样本目标检测的分布区别

Fig. 2　The distributional differences among classic deep learning object detection, few-shot object detection, and cross-domain few-shot object detection

其次，空间分辨率与目标尺度差异在航空遥感检测中尤为突出。由于成像设备和拍摄高度不同，同一类目标在不同数据域中可能呈现出完全不同的尺寸、清晰度和细节水平。高分辨率影像中可辨识的局部结构，在低分辨率场景中可能仅表现为模糊斑块；而小目标在跨域迁移过程中更容易受到尺度变化的影响，导致检测器在分类和定位上均出现性能下降。

再次，地理区域与场景背景差异也是重要因素。航空遥感图像通常覆盖城市、港口、机场、山区、海岸和农田等多种场景，不同区域在地表覆盖类型、目标分布密度和背景纹理模式上差异明显。例如，同一类飞行器、车辆或舰船目标在不同区域可能处于截然不同的背景环境中，使得模型容易过度依赖源域背景先验，而在目标域中产生误检或漏检。

此外，时间条件与环境因素差异同样会引入显著域偏移。季节变化、天气条件、太阳高度角、阴影分布以及成像时段等因素都会对图像亮度、对比度、颜色风格和可见性产生影响。在航空遥感场景下，这类变化往往不是简单的光照扰动，而是会进一步影响目标边缘、纹理细节和背景可分性，使模型的域泛化能力受到挑战。

最后，目标自身形态与标注分布差异也不容忽视。同一类别目标在不同域中可能表现出朝向变化、结构差异、密集排列程度不同以及遮挡程度不一致等特点；同时，不同数据集在类别划分标准、标注精度和样本均衡性方面也常存在差异。这不仅增加了跨域迁移的难度，也会在少样本条件下进一步放大类别表示不充分的问题。

综上，航空遥感场景中的域差异并非单一因素导致，而是由成像平台、传感器、分辨率、场景背景、环境条件和目标分布等多种因素共同作用形成的复合偏移。正是这种复杂而多源的域差异，使得跨域少样本目标检测相比通用视觉场景更具挑战性，也决定了相关方法需要同时兼顾域不变表征学习与有限样本适配能力。

2　跨域少样本目标检测方法综述

针对域偏移显著、目标域标注样本有限等问题，现有跨域少样本目标检测方法主要围绕知识迁移、快速适配、样本扩充和语义增强等思路展开。总体来看，不同方法虽然侧重点各异，但其核心目标均在于提升检测模型在有限目标域监督条件下的跨域泛化能力。为更清晰地梳理该方向的技术脉络，本文将相关方法概括为基于迁移与域对齐的方法、基于元学习与度量学习的方法、基于生成式数据增强与扩充的方法、基于视觉语言对齐的方法以及基于大模型的方法，并分别进行综述与分析。从技术作用机制看，现有方法大体可对应 5 个层面：迁移与域对齐方法主要解决源域知识如何迁移到目标域的问题；元学习与度量学习方法主要解决目标域少量样本如何得到有效利用的问题；生成式增强方法从数据层面扩展目标域样本分布；视觉语言对齐方法通过文本语义补充视觉样本不足；大模型方法则利用大规模预训练获得更强的开放泛化能力。上述技术路线分别从特征、样本、数据、语义和模型能力层面缓解跨域少样本目标检测中的核心矛盾。

2.1　基于迁移与域对齐的方法

基于迁移与域对齐的方法是跨域少样本目标检测中最直接、也最具代表性的一类思路。其基本出发点是：先利用源域较充足的标注数据训练或预训

练检测器，再结合少量目标域样本对模型进行迁移适配；同时，通过显式或隐式的域对齐机制缩小源域与目标域之间的分布差异，以提升检测器在目标域中的泛化能力。与单纯少样本目标检测相比，这类方法更加关注“已有检测知识如何跨域迁移”；与传统跨域目标检测相比，则更强调在目标域标注极少的条件下抑制过拟合并保持类别判别能力。

从迁移学习角度看，预训练—微调仍然是这一方向的重要基线。Lee等人（2022）在多领域少样本目标检测（multi-domain few-shot object detection，MoFSOD）基准上指出，在跨领域设定下，简单微调（fine-tuning）本身就是一个很强的基线，其性能甚至优于部分已有少样本目标检测方法，这说明源域预训练特征与合理的参数迁移策略对跨域适配具有重要作用。随后，Xiong（2023）在其所提出的跨域少样本目标检测基准上进一步发现，现有不少元学习式或常规少样本目标检测方法在跨域场景下会明显退化，甚至不如朴素微调；为此，其提出了一种基于自蒸馏师生互益的强基线，以缓解目标域少样本条件下的过拟合问题。上述研究表明，在跨域少样本场景中，如何稳定地继承源域检测知识、控制微调过程中的偏移和遗忘，是迁移类方法的核心问题。

从域对齐角度看，相关方法通常通过图像级、特征级和实例级等不同粒度的对齐来减弱域偏移影响。Chen等人（2018）提出的DA Faster R-CNN是该方向的代表性工作之一，其在Faster R-CNN框架（Ren等，2017）中分别针对图像级偏移和实例级偏移引入域分类器，并通过一致性正则化学习域不变的候选区域表示。Saito等人（2019）提出的强弱分布对齐（strong-weak distribution alignment，SWDA）进一步指出，目标检测中的全局分布不宜被强行完全对齐，因此采用“局部强对齐、全局弱对齐”的策略，以避免由于场景布局差异过大而带来的负迁移。Li等人（2020a）提出的空间注意金字塔网络（spatial attention pyramid network，SAPNet）则从多尺度特征出发，利用空间注意金字塔整合全局与局部语义信息，在对齐过程中更加关注具有判别性的区域特征。这些方法共同构成了后续跨域少样本目标检测中“迁移 + 对齐”范式的重要基础。

在此基础上，部分研究开始将域对齐与少样本适配进一步结合。Li等人（2022b）提出的Adaptive Teacher采用教师—学生框架，在学生分支中引入特征级对抗对齐，同时结合weak-strong augmentation与互学习机制，以提高跨域伪标签的质量并缓解源域偏置。该类方法虽然主要面向跨域目标检测，但其“源域预训练—目标域伪监督—跨域一致性约束”的思想，对跨域少样本目标检测同样具有较强借鉴意义，尤其适用于航空遥感中跨平台、跨区域和跨传感器场景下的快速迁移。

总体来看，基于迁移与域对齐的方法是跨域少样本目标检测中较为基础且实用的一类技术路线。其优势在于能够充分利用源域中较丰富的标注数据和已有检测器的表征能力，并通过微调、对抗对齐、教师—学生学习或跨域一致性约束等方式，缓解源域与目标域之间的分布差异。这类方法实现路径清晰，通常较容易与Faster R-CNN、检测Transformer（detection transformer，DETR）及其变体等主流检测框架结合，因而适合作为跨域少样本目标检测的基础基线。

然而，该类方法的效果通常依赖源域与目标域之间是否存在可迁移的共享结构。当两域差异较大时，简单微调容易造成过拟合，强制域对齐也可能削弱类别判别信息，甚至引入负迁移。在航空遥感场景中，平台、传感器、空间分辨率、地理区域和背景纹理往往同时变化，域偏移具有明显的复合性，仅依赖图像级或特征级整体对齐难以充分兼顾小目标检测、细粒度类别区分和复杂背景抑制。因此，迁移与域对齐方法更适合作为跨域适配的基础模块，后续通常需要结合少样本判别学习、数据增强或语义先验建模，以提升目标域检测稳定性。

2.2 基于元学习与度量学习的方法

基于元学习与度量学习的方法主要关注模型在极少样本条件下的快速适配能力。其核心思想并不是直接依赖目标域大量标注进行再训练，而是通过“支持集—查询集”式任务构造、类别原型建模、相似性度量或对比表示学习，使检测器能够从少量样本中迅速建立稳定的类别判别边界。与前述迁移与域对齐方法相比，这类方法更强调“如何学会利用少量样本”，因此在跨域少样本目标检测中通常承担着提升新类别适应能力和细粒度判别能力的重要作用。

早期代表性工作大多沿着元学习检测框架展开。Yan等人（2019）提出的Meta R-CNN将元学习机制引入Faster R-CNN，在感兴趣区域（region of

interest，RoI）特征层面而非整幅图像层面进行任务建模，并通过 predictor-head remodeling network 利用少量支持样本生成类别注意向量，对检测头进行动态调制，从而提升对新类别的快速适应能力。Fan 等人（2020）提出的注意力区域建议网络（attention region proposal network，Attention-RPN）与多关系检测器（multi-relation detector）则进一步从支持图像与查询图像之间学习匹配关系，在候选框生成和目标判别阶段同时引入关系建模，使模型能够依据少量支持样本完成类条件检测。这类方法为后续跨域少样本目标检测奠定了基本范式，即通过元任务训练学习可迁移的检测经验，而非仅学习固定类别的分类器参数。图 3 展示了基于元学习的检测器的一般结构。

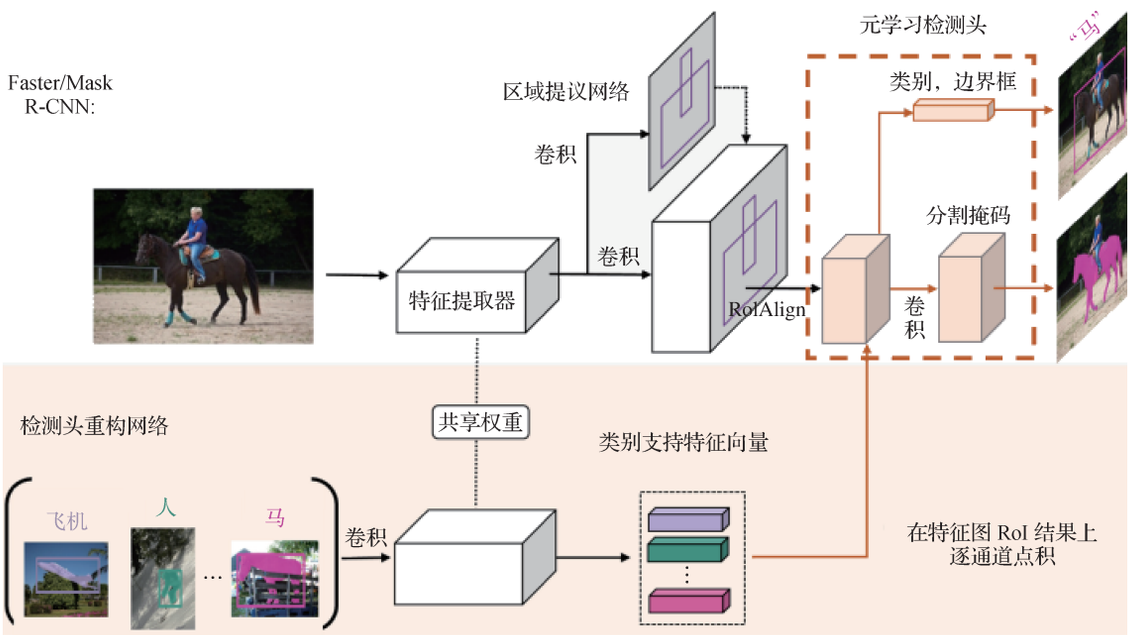


图 3　Yan 等人（2019）提出的典型元学习检测器

Fig. 3　The typical meta-learning detector proposed by Yan et al.（2019）

随着少样本目标检测研究的发展，度量学习思想进一步用于增强类别可分性，其理念如图 4 所示。

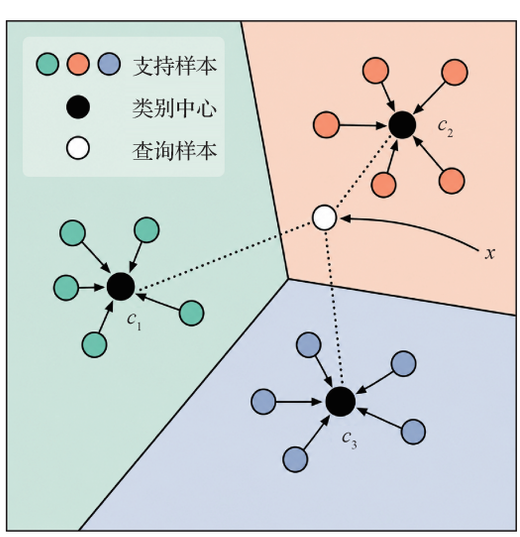


图 4　度量学习理念图示

Fig. 4　Schematic diagram of metric learning concept

Sun 等人（2021）提出的少样本目标检测候选区域对比编码方法（few-shot object detection via contrastive proposal encoding，FSCE）通过候选区域对比编码损失（contrastive proposal encoding loss）强化候选区域表征的类内紧致性与类间可分性，从而缓解少样本条件下新类别易被混淆的问题。此类方法虽然主要在同域少样本设定下取得效果，但其通过原型、关系或对比约束学习稳健特征空间的思路，对跨域少样本目标检测同样具有重要启发意义：当目标域样本极少时，检测器往往难以学习到充分的类别边界，而度量学习能够在有限监督下增强支持样本与查询目标之间的匹配稳定性。

近年来，基于视觉 Transformer（vision Transformer，ViT）的样例条件（example-conditioned）检测方法进一步拓展了这一技术路线。Zhang 等人（2025）提出的 DE-ViT（detect everything with vision

Transformer)不再依赖传统少样本目标检测中的复杂微调流程,而是利用预训练ViT构建一种无需微调的少样本目标检测器:一方面通过区域传播机制完成目标定位,另一方面利用支持样本原型将ViT特征投影到对基础类过拟合更不敏感的子空间,从而显著提升对新类别的开放式检测能力。由于其本质上仍依赖支持样本与查询目标之间的原型匹配和相似性判断,因而从广义上看也属于度量学习范式在Transformer检测中的一种延伸。

在跨域少样本目标检测场景下,Fu等人(2024)进一步分析发现,尽管DE-ViT在传统少样本目标检测中表现优异,但在存在明显域差异时仍会出现明显性能退化。为此,其构建了CD-FSOD基准,并提出CD-ViTO方法,对DE-ViT进行面向跨域场景的增强。该方法指出跨域退化主要与3类因素相关,即风格差异(style)、类间方差减小(inter-class variance, ICV)以及类别边界模糊(indefinable boundaries, IB)。在此基础上,CD-ViTO引入learnable instance features、instance reweighting和domain prompter等模块,以增强实例特征的可区分性、抑制低质量实例干扰,并提升模型对风格变化的鲁棒性。相较于一般的元学习或度量学习方法,CD-ViTO更进一步体现了“少样本适配+跨域稳健表征”联合建模的思路,是该方向较具代表性的工作之一。

总体来看,基于元学习与度量学习的方法主要从如何有效利用少量目标域样本的角度提升跨域少样本目标检测能力。元学习通过支持集—查询集任务构造,使模型在训练阶段学习快速适配新类别或新任务的能力;度量学习则通过原型匹配、关系建模或对比约束增强特征空间中的类内紧致性与类间可分性。与迁移和域对齐方法相比,这类方法更加关注少样本条件下类别判别边界的建立,因而在目标域标注极少、新类别样本有限的情况下具有较强适用性。

但该类方法也存在明显限制。首先,其性能往往对支持样本质量较为敏感,当支持样本存在遮挡、尺度过小、背景干扰或标注偏差时,类别原型和相似性度量容易失真;其次,在跨域场景中,源域学习到的度量空间未必能够直接适用于目标域,目标域样本过少又使模型难以充分校正这种原型偏移。对于航空遥感图像而言,小目标、密集排列、方向任意和类间外观相似等问题会进一步放大度量不稳定性。因此,元学习与度量学习方法虽然能够增强少样本适配能力,但若要在复杂遥感跨域场景中稳定发挥作用,仍需与域鲁棒表征、支持样本筛选、多尺度特征增强或基础模型表征相结合。

2.3 基于生成式数据增强与扩充的方法

总体来看,在跨域少样本目标检测中,目标域标注样本极少且域偏移显著,容易导致检测器在微调阶段出现过拟合、类别边界不稳定以及背景偏置等问题。基于生成式数据增强与扩充的方法,核心思路是通过特征生成、实例合成或图像级合成来扩展目标域可用样本的数量与多样性,从而缓解少样本约束,并尽可能提升模型对目标域分布的覆盖能力。与前两类方法相比,这类方法并不直接改变检测器的基本结构,而是从“数据侧”增强训练信号,因此具有较强的通用性和可移植性。其主要方法如图5所示。

从方法演进看,早期代表性工作主要集中在特征空间增强。Zhang和Wang(2021)提出的hallucination improves few-shot object detection在RoI特征空间中引入hallucinator network,用于生成额外的、有用的训练样本,从而提升极少样本条件下检测器的鲁棒性。该方法的优势在于不必直接生成高分辨率图像,而是在更紧凑的特征空间中扩充新类别的表征分布,因此实现相对简洁,也为后续通过生成机制补足少样本监督的研究提供了重要启发。

随后,研究开始从特征级生成逐步转向图像级与实例级合成。Lin等人(2023)系统讨论了文本生成图像在少样本目标检测中的可用性,并设计了基于显著性目标提取与copy-paste的合成流程,同时结合对比语言—图像预训练模型(contrastive language-image pre-training, CLIP)过滤高误报样本,以提升合成数据对少样本目标检测的实际价值。进一步地,Vu等人(2025)提出多视角数据增强(multi-perspective data augmentation, MPAD),利用扩散模型从前景—前景关系和前景—背景关系两个角度生成更具代表性与“困难性”的样本,以增强少样本目标检测器对类别边界和背景变化的适应能力。上述研究说明,生成式增强已不再局限于简单复制或随机扰动,而是开始强调样本语义、布局合理性和背景匹配度。

在遥感相关场景中,这一思路也得到进一步发展。Tang等人(2025)提出的AeroGen面向航空遥感目标检测构建了layout-conditional diffusion model,

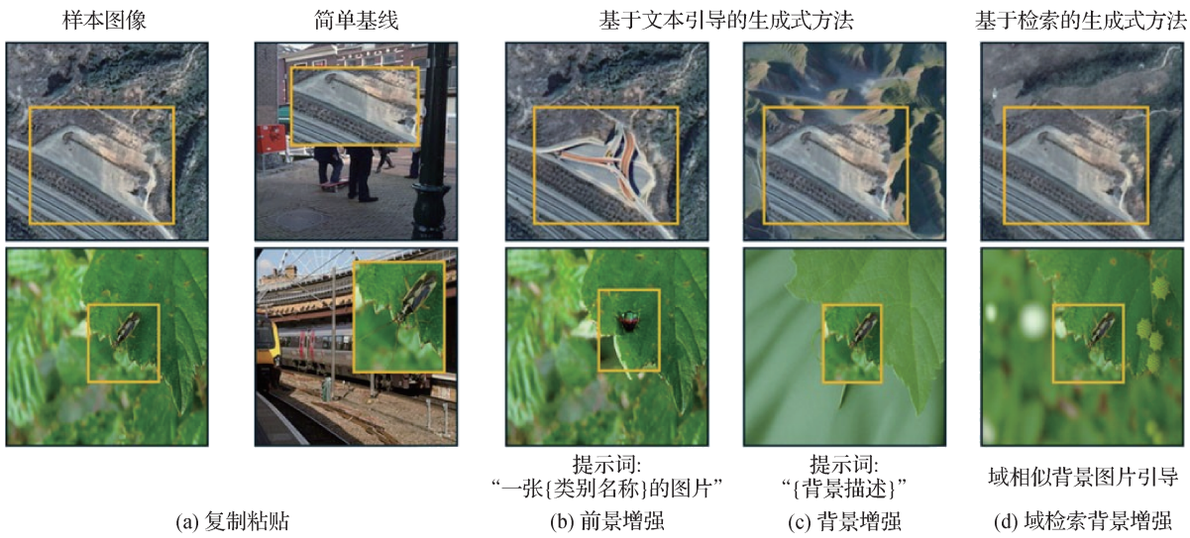


图 5　基于 Li 等人（2025）整理的生成式数据增强与扩充方法概览

Fig. 5　Overview of methods based on generative data augmentation and expansion summarized from Li et al.（2025）
（（a）copy-paste；（b）text-based foreground enhancement；（c）text-based background enhancement；
（d）domain retrieval-based background enhancement）

通过布局约束的数据生成和双准则筛选机制提升合成样本对下游检测的有效性。与此同时，Liu 等人（2025b）提出的 Control Copy-Paste 针对光学遥感少样本目标检测，借助可控扩散模型提升新类别实例的多样性与场景真实性。这些工作虽然并不都直接针对跨域少样本目标检测设定，但其共同特点在于：开始显式关注遥感场景中的目标尺度、布局稀疏性和背景复杂性，从而使生成样本更接近航空遥感任务的真实分布。

在跨域少样本目标检测中，具有代表性的工作是 Li 等人（2025）提出的域检索增强生成方法（Domain-RAG）。该方法指出，已有增强策略如 copy-paste 或文本到图像生成，往往难以同时保证目标类别正确性与背景域一致性，因而难以直接适用于跨域少样本目标检测。为此，Domain-RAG 提出一种 training-free 的检索引导组合式图像生成框架，整体包括 3 个阶段：域感知背景检索、域引导背景生成以及前景—背景组合。具体而言，该方法先将输入图像分解为前景与背景区域，再检索在语义和风格上相近的图像作为引导，条件化生成与目标域更一致的新背景，最后保留原有前景并与新背景进行组合，得到兼顾目标语义和域一致性的增强样本。论文报告表明，Domain-RAG 在跨域少样本目标检测、遥感少样本目标检测以及伪装目标少样本目标检测等任务上均能稳定提升强基线表现，是当前生成式增强方法中较具代表性的跨域方案。

总体来看，基于生成式数据增强与扩充的方法从数据层面缓解跨域少样本目标检测中的样本稀缺问题。与直接改进检测器结构不同，该类方法通过特征生成、实例合成、图像级生成、背景替换或前景—背景组合等方式，扩展目标域样本的外观、尺度、姿态、背景和风格多样性，从而为检测器提供更丰富的训练信号。其优势在于具有较强的通用性，可与迁移学习、少样本微调和视觉语言模型等方法结合使用，尤其适合目标域标注样本极少但需要快速扩充训练数据的场景。

不过，生成式增强方法的核心难点在于如何同时保证样本的类别一致性、域一致性和标注可靠性。如果合成样本的目标语义不准确、边界框与目标区域不匹配，或者前景与背景之间存在明显不协调，增强数据不仅不能提升性能，反而可能向检测器引入噪声监督。在航空遥感场景中，这一问题更加突出：遥感目标通常尺度小、纹理弱、视角特殊，背景又具有较强复杂性和场景依赖性，因此生成真实且检测友好的样本比自然图像场景更困难。未来该类方法需要进一步加强生成样本质量评估、跨域背景一致性约束、目标实例几何合理性控制以及自动化样本筛选机制，以提高其在航空遥感跨域少样本目标检测中的实用性。

2.4 基于视觉语言对齐的方法

基于视觉语言对齐的方法，核心在于引入文本语义作为辅助先验，通过建立区域特征—语言描述之间的对应关系，增强检测器对新类别和新域的理解能力。相较于仅依赖视觉支持样本的方法，语言信息能够提供更稳定的类别语义约束，尤其在目标域样本极少时，有助于缓解类别表征不足和支持样本噪声带来的影响。因此，这类方法近年来逐渐成为跨域少样本目标检测中的一个重要研究方向。与此同时，开放词汇目标检测中的研究也表明，大规模视觉语言预训练模型具有较强的语义迁移能力，但若直接将图像级视觉语言模型用于区域检测，往往会受到“整图—区域”分布不匹配的限制，因此需要专门的区域级视觉语言对齐机制。图6展示了基于视觉语言对齐的检测器的一般结构。

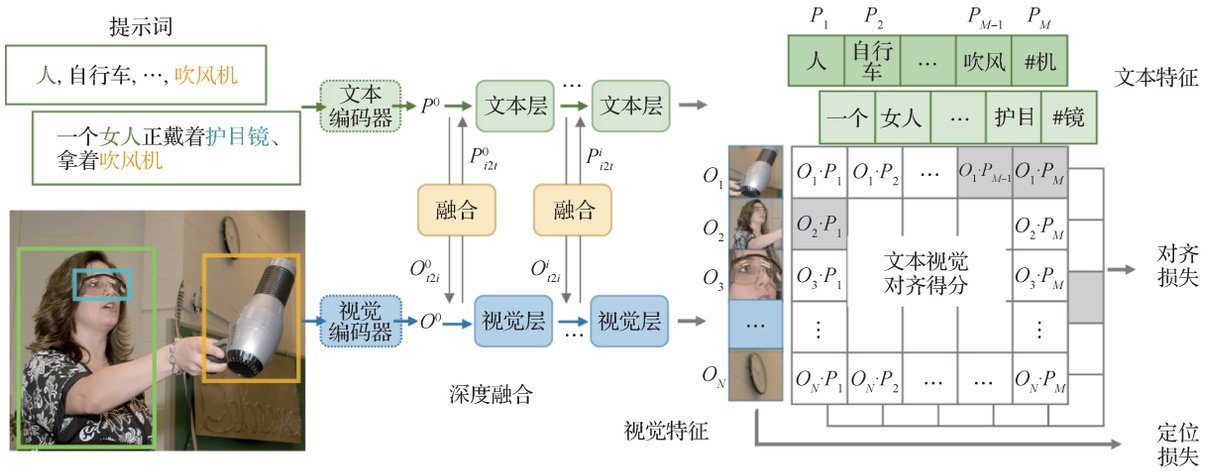


图6 Li等人(2022a)提出的基于视觉语言对齐的检测器一般架构

Fig. 6 General architecture of detectors based on vision-language alignment proposed by Li et al. (2022a)

从发展脉络看，这类方法与开放词汇检测和基础视觉语言模型密切相关。以Zhong等人(2022)提出的区域级对比语言—图像预训练(RegionCLIP)为代表的工作通过构造区域—文本对齐预训练，将CLIP的图文对齐能力扩展到候选区域层面，从而提升模型对新类别的识别能力；Feng等人(2022)提出的提示检测器(PromptDet)等方法则进一步利用提示学习和未标注图像扩展检测器词汇表，说明语言描述能够为检测任务提供超越有限标注类别的外部知识。虽然这些方法并非专门针对跨域少样本目标检测提出，但它们验证了一个关键事实：文本语义能够为目标检测提供更强的类别可迁移性和概念泛化能力，这为后续将语言模态引入跨域少样本目标检测奠定了基础。

在少样本目标检测背景下，视觉语言模型的潜力也得到进一步强调。Madan等人(2024)指出，诸如Grounding DINO(grounding detection Transformer with improved denoising anchor boxes)(Liu等，2024)等基础视觉语言模型在零样本设定下已能在部分少样本测评上超过传统少样本目标检测器，说明语言先验和大规模跨模态预训练确实能够显著提升对新类别的识别能力。这一结果进一步说明，在跨域少样本场景中，若能将目标域少量视觉样本与更丰富的语言语义有效结合，模型有望在极低监督条件下获得更强的跨域泛化性。

针对跨域少样本目标检测，具有代表性的工作是Shangguan等人(2025)提出的基于丰富文本的跨域多模态少样本目标检测(cross-domain multi-modal few-shot object detection via rich text)。该方法认为，现有多模态目标检测方法在面对显著域偏移时容易退化，而更丰富的文本描述有助于模型建立视觉实例与语言语义之间更稳健的知识关联。为此，该方法引入丰富文本(rich text)作为辅助模态，在Meta-DETR(Zhang等，2023a)基线上通过双向文本特征生成进一步强化视觉—语言对齐，并提升模型对语言描述的理解能力。与仅使用类别名称或简短文本提示的方法相比，rich text中包含更具体的属性、语义关系乃至专业术语，因此在跨域场景下能更有效地弥补视觉样本不足带来的信息缺失。实验结果表明，该方法在多个跨域少样本测评上优于已有少样

本目标检测方法，体现了语言增强对抗域偏移的实际价值。

总体来看，基于视觉语言对齐的方法为跨域少样本目标检测提供了从语义层面增强模型泛化能力的技术路径。该类方法通过引入类别名称、属性描述、场景语义或更丰富的文本提示，将视觉区域特征与语言语义表示进行对齐，从而在目标域视觉样本不足时补充类别先验信息。相比仅依赖少量支持图像的方法，文本语义具有更强的抽象性和可迁移性，有助于缓解少样本条件下类别表征不完整、支持样本偶然性强以及跨域视觉特征不稳定等问题。

然而，视觉语言对齐方法的效果高度依赖文本描述质量和区域级对齐能力。若文本提示过于简单，可能无法表达遥感目标的细粒度外观差异；若描述过于宽泛或与实际遥感成像特征不一致，也可能引入语义噪声。此外，航空遥感目标多以俯视角呈现，许多类别在图像中缺少自然场景下明显的纹理和结构特征，仅依靠通用语言模型中的类别语义难以充分区分相似目标。因此，视觉语言对齐方法在遥感场景中的关键问题不仅是如何引入文本，更在于如何构造符合遥感成像特点的专业文本描述、如何实现区域级视觉特征与遥感语义知识的精确对齐，以及如何避免语言先验与目标域视觉证据之间的不匹配。

2.5　基于大模型的方法

随着视觉基础模型和多模态大模型的快速发展，跨域少样本目标检测开始从“面向特定任务设计检测器”转向“利用大规模预训练模型的通用表征能力”。

这一类方法的基本思路是：借助在海量图文数据或大规模检测数据上预训练得到的基础模型（即大模型），为目标检测提供更强的类别语义先验、跨域表征能力和开放词汇泛化能力，再通过少量目标域样本进行轻量适配或提示式对齐，以缓解传统少样本目标检测器在新域上的性能退化。Madan 等人（2024）提出的 Foundational FSOD 研究表明，基于大规模数据预训练的视觉语言模型已显著改变少样本目标检测的研究范式。例如，Grounding DINO 的零样本结果在其通用对象上下文（COCO）（Lin 等，2014）对比实验中已明显超过传统少样本目标检测基线，这说明大模型预训练所带来的通用知识已成为少样本目标检测中不可忽视的重要基础。

图 7 展示了基于视觉语言大模型的检测器的一般结构。在跨域少样本目标检测中，较具代表性的一类工作是直接将视觉基础模型引入检测框架。Pan 等人（2025b）提出的 ETS（enhance then search）指出，诸如 Grounding DINO 和 LAE-DINO（locate anything on earth with DINO，LAE-DINO）（Pan 等，2025a）这类在大规模数据上预训练的基础模型，在跨域少样本目标检测任务中已表现出较强潜力；在此基础上，该方法结合图像增强与网格化子域搜索策略，在不依赖复杂重训练的前提下进一步提升基础模型对目标域的适应能力。其核心思想并不是重新构造检测器，而是充分利用基础模型已有的跨域表征能力，并通过少样本条件下的策略搜索找到更合适的域内配置，从而提高在新域中的检测性能。

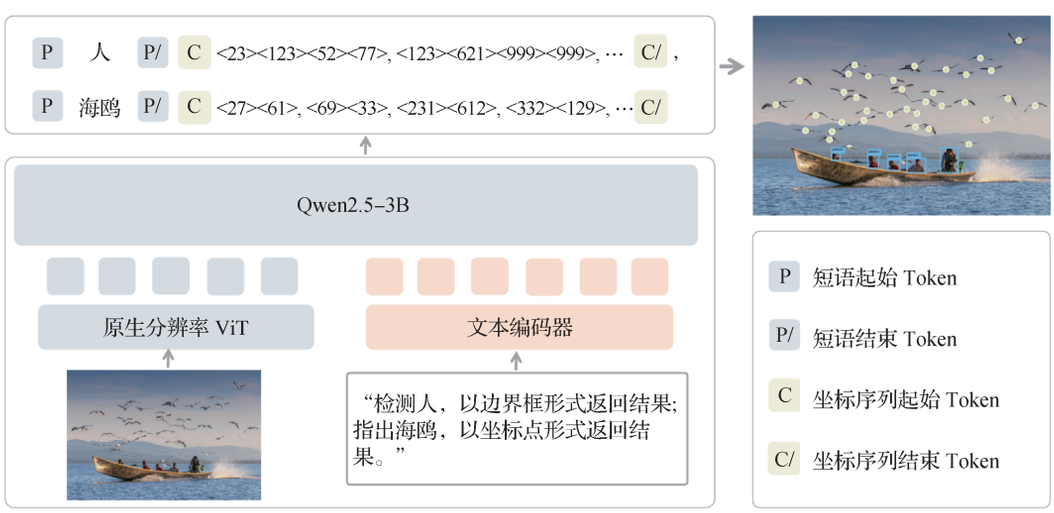


图 7　Jiang 等人（2026）提出的基于视觉语言大模型的检测器一般架构

Fig. 7　General architecture of a detector based on a large vision-language model proposed by Jiang et al.（2026）

另一类代表性思路是将传统检测器与视觉基础模型进行结构级融合。Liu等人(2025a)认为,现有成熟检测器通常具有较强的定位能力,但泛化性不足;而视觉基础模型往往具备更强的跨域表征能力,却缺乏精确定位能力。为此,该方法构建了基于Mixture-of-Experts的融合框架,通过Expert-wise Router和Region-wise Router选择并强化与目标域更相关的基础模型特征,同时利用共享与私有投影模块缩小检测器特征空间与基础模型特征空间之间的表示差异。该方法的特点在于无需重新在基础类上进行完整重训练,而是通过"检测器定位能力+基础模型泛化能力"的互补融合来实现对未见领域的适配。

在更广义的大模型检测方向上,近期出现的多模态大模型也开始展现出对目标检测任务的直接建模能力。Jiang等人(2026)提出的Rex-Omni是一个3B规模的多模态大模型,其将目标检测重新表述为"下一点预测"问题,利用离散化坐标token、针对grounding/referring/pointing构建的数据引擎,以及"监督微调+基于GRPO的强化后训练"两阶段训练流程,提高了坐标预测精度并缓解了重复检测等问题。论文报告表明,Rex-Omni在COCO和大规模词汇实例分割数据集(large vocabulary instance segmentation, LVIS)(Gupta等,2019)等基准上可在零样本条件下达到与DINO(Zhang等,2023b)、Grounding DINO等回归式检测器相当甚至更优的性能。尽管Rex-Omni并非专门针对跨域少样本目标检测提出,但它表明:当检测能力被直接整合进大模型的统一视觉—语言—定位框架后,模型有可能在更少监督、更开放类别和更复杂域变化条件下展现出更强的泛化潜力。

总体来看,基于大模型的方法正在推动跨域少样本目标检测从任务特定模型设计转向通用基础能力迁移。视觉基础模型、视觉语言模型和多模态大模型通过大规模预训练获得了更强的通用表征能力、开放词汇识别能力和跨域泛化潜力,能够在目标域样本极少甚至零样本条件下提供较好的初始检测能力。与传统微调或元学习方法相比,该类方法不再完全依赖少量目标域样本学习类别特征,而是利用预训练阶段获得的语义知识和视觉表征作为少样本目标检测的基础支撑,因而在新类别、新场景和开放词汇检测任务中具有较大发展潜力。

但该类方法在航空遥感场景中的应用仍面临现实限制。首先,大模型通常计算和存储开销较高,而航空遥感图像分辨率大、目标尺度小,直接输入完整图像往往难以兼顾效率与检测精度;其次,通用大模型主要从自然图像或通用图文数据中学习先验,其语义知识未必能够充分覆盖遥感场景中的专业类别、俯视视角、小目标结构和复杂地物背景;再次,部分基础模型虽然具备较强的类别泛化能力,但定位精度、密集目标处理能力和跨传感器适配能力仍需进一步提升。因此,未来基于大模型的跨域少样本遥感检测研究,应更加关注遥感专用数据预训练、轻量化适配、高分辨率图像分块推理、区域级语义对齐以及检测器定位能力与基础模型泛化能力的互补融合。

2.6 方法对比与小结

前文从迁移与域对齐、元学习与度量学习、生成式数据增强与扩充、视觉语言对齐以及大模型等5个方面,对跨域少样本目标检测的代表性方法进行了梳理。总体来看,这5类方法分别从不同层面缓解跨域少样本目标检测中的核心矛盾:迁移与域对齐方法主要关注源域知识如何稳定迁移到目标域;元学习与度量学习方法强调如何从少量支持样本中建立有效的类别判别机制;生成式数据增强方法通过扩展目标域样本分布缓解标注不足问题;视觉语言对齐方法借助文本语义先验增强类别表示与跨域泛化能力;大模型方法则进一步依托大规模预训练获得更强的通用表征和开放词汇迁移能力。

为更直观地比较不同技术路线的特点,表2总结了5类代表性方法的核心思想、主要优势、局限性及适用条件。从方法关系看,上述5类技术路线关注的核心问题各不相同,具有明显的互补性。

迁移与域对齐方法为跨域适配提供了基本框架,其重点在于减小源域与目标域之间的分布差异,但在目标域样本极少时,单纯域对齐往往难以形成稳定的类别判别边界。元学习与度量学习方法正好从少样本判别角度补足这一不足,通过支持集—查询集建模、类别原型学习或对比约束提升模型对少量样本的利用效率。然而,当源域与目标域存在明显成像差异时,原型表示和相似性度量也可能发生偏移,因此仍需要与域鲁棒表征或迁移适配机制结合。

生成式数据增强方法则从数据层面提供另一种

补充思路。它通过特征生成、实例合成、图像生成或前景—背景组合等方式扩充目标域样本，使检测器能够接触到更多目标外观、尺度、姿态和背景变化。该类方法能够缓解少样本约束，但其有效性依赖生成样本的类别一致性、域一致性和标注可靠性。视觉语言对齐方法进一步从语义层面增强模型泛化能力，利用类别名称、属性描述或丰富文本提示弥补视觉样本不足的问题。相较于单纯视觉特征，语言语义具有更强的抽象性和可迁移性，但在航空遥感场景中，通用文本描述未必能够准确表达俯视视角、小目标结构和专业类别差异，因此仍需面向遥感任务构造更具领域性的文本先验。

表 2　跨域少样本目标检测代表方法比较

Table 2　Comparison of representative methods for cross-domain few-shot object detection

方法类别	代表方法	核心思想	主要优势	主要局限
迁移与域对齐	DA Faster R-CNN SWDA Adaptive Teacher	通过特征对齐、教师学生学习或微调减小源域与目标域差异	易与现有检测器结合，利用充分源域知识	强域偏移下可能负迁移，新类别边界不稳定
元学习与度量学习	Meta R-CNN FSCE DE-ViT CD-ViTO	通过支持集、原型、关系建模或对比学习实现少样本快速适配	样本效率较高，适合新类别快速检测	对支持样本质量敏感，跨域时原型易偏移
生成式增强	Hallucination MPAD AeroGen Domain-RAG	通过特征生成、图像合成或背景组合扩展目标域样本	可缓解样本不足，提高数据多样性	合成样本可能引入噪声，标注可靠性难保证
视觉语言对齐	RegionCLIP PromptDet RT CD-MFSOD	利用文本语义与视觉区域对齐，增强类别语义表达	可引入外部语义先验，提升开放类别泛化	依赖文本质量，遥感专业语义表达不足
大模型方法	Grounding DINO ETS Rex-Omni	利用基础模型或多模态大模型的通用表征和开放词汇能力	跨域泛化潜力强，低样本适配能力较好	计算成本高，遥感高分辨率图像适配困难

近年来，大模型方法正在重塑跨域少样本目标检测的研究范式。与传统方法主要依赖任务内训练数据不同，视觉基础模型、视觉语言模型和多模态大模型能够从大规模预训练中获得更丰富的视觉语义知识，为新域、新类别和少样本目标检测提供更强的初始能力。大模型方法与前述技术路线也存在天然结合空间：其通用表征可作为域对齐和元学习的特征基础，其开放词汇能力可增强视觉语言对齐，其生成能力和语义理解能力也可服务于目标域样本扩充与筛选。然而，航空遥感图像通常具有高分辨率、小目标密集、背景复杂和传感器差异显著等特点，通用大模型若缺乏遥感专用适配，仍可能面临定位精度不足、计算开销过高和领域语义不匹配等问题。

从适用条件看，不同方法各有侧重。迁移与域对齐方法更适用于源域标注较充分且源域与目标域具有一定相关性的场景；元学习与度量学习方法更适用于目标域支持样本数量极少但样本质量较高的设定；生成式增强方法适合目标域样本不足且需要扩展背景、风格或目标外观多样性的场景；视觉语言对齐方法更依赖类别语义清晰、文本描述可获得或可构造的任务；大模型方法则更适合开放词汇、新类别快速适配和跨域泛化要求较高的场景。对于航空遥感任务而言，实际应用往往同时包含平台差异、传感器差异、分辨率差异、地理区域变化和样本稀缺等多重因素，因此单一技术路线通常难以充分应对全部挑战。

综上，跨域少样本目标检测已经从早期依赖简单微调或单一域对齐的研究阶段，逐渐发展为融合迁移适配、少样本判别、数据增强、语义对齐和基础模型迁移的综合性研究方向。未来更具潜力的方案可能不是某一类方法对其他方法的完全替代，而是在统一框架下实现多种机制的协同设计。例如，可利用基础模型提供稳健初始表征，通过域对齐减弱源域与目标域分布差异，结合元学习或度量学习提升少样本判别能力，再借助生成式增强和视觉语言先验扩展目标域信息。对于航空遥感场景，如何在复杂域偏移、小目标密集分布、目标域标注稀缺和高分辨率图像处理需求之间取得平衡，仍是后续研究需要重点解决的问题。

3 跨域少样本目标检测数据集、评测方式与应用特点

跨域少样本目标检测的研究不仅依赖于方法设计，还高度依赖于合理的数据集构建、任务设置和评测方式。由于该任务同时涉及域偏移与样本稀缺两类因素，不同数据集及实验设定往往会直接影响方法性能比较的公平性与结论的可靠性。与此同时，航空遥感场景在成像条件、目标分布和应用需求方面具有明显区别于自然场景的特点，本节从数据支撑、评测标准和任务属性3个层面，对该方向的研究基础进行进一步梳理。

3.1 相关数据集与任务设置

本文所称航空遥感场景主要涵盖航空/航天平台获取的俯视遥感图像，相关数据集包括航空与卫星遥感目标检测基准。现有跨域少样本目标检测研究所使用的数据集大致可分为两类：一类是面向通用视觉场景构建的跨域少样本目标检测基准，用于系统评估方法在不同域之间的迁移能力；另一类是航空遥感目标检测数据集，通常进一步改造为少样本或跨域实验设定，用于验证方法在航空遥感场景中的实际适用性。前者强调“基准统一”和“跨域可比性”；后者更强调目标尺度、成像条件和场景复杂性对检测性能的影响。图8展示了一个代表性的跨域少样本目标检测测评—— CD-FSOD benchmark，其下游目标域任务中包含了航空领域的目标检测数据集DIOR。

在通用基准方面，较有代表性的工作包括MoFSOD和CD-FSOD。其中，MoFSOD由Lee等人（2022）提出，包含来自多种图像域的10个数据集，主要用于考察少样本目标检测方法在多领域场景下的泛化表现；CD-FSOD由Fu等人（2024）专门面向跨域少样本目标检测任务构建，强调源域与目标域之间存在更明显的数据分布差异。值得注意的是，这两类基准在样本定义上也存在差别：MoFSOD更偏向以每类K幅图像进行少样本采样，而CD-FSOD更强调每类K个实例的平衡设定，因此后者通常更贴近标准少样本目标检测中的K-shot评测方式。

在航空遥感场景中，常用于目标检测及其少样本扩展研究的数据集包括光学遥感图像目标检测数据集（object detection in optical remote sensing images，DIOR）（Li等，2020b）、航空图像目标检测数据集（dataset for object detection in aerial images，DOTA）（Xia等，2018）、xView（Lam等，2018）和航空图像车辆检测数据集（vehicle detection in aerial imagery，VEDAI）（Razakarivony和Jurie，2016）等。DIOR是一个较大规模的光学航空遥感目标检测基准，包含23 463幅图像、192 472个目标实例和20个类别，具有明显的尺度变化、成像条件变化以及较强的类间相似性与类内差异性；DOTA则由多传感器、多平台航空影像构成，包含2 806幅大尺寸图像、15个类别和188 282个实例，并采用任意四边形标注，特别适合评测旋转目标与多尺度目标检测；xView由0.3 m分辨率WorldView-3卫星图像构成，覆盖60个类别和超过100万个目标实例，具有较强的区域与场景多样性；VEDAI则更聚焦于航空视角下的小目标车辆检测，并提供不同分辨率与光谱条件下的数据。近年来，面向遥感少样本目标检测的工作也常直接在DIOR和西北工业大学高分辨率10类遥感目标数据集（Northwestern Polytechnical University very-high-resolution 10-class dataset，NWPU VHR-10）（Cheng等，2014）上构建实验设定，以评估模型在遥感场景中的低样本学习能力。各数据集间对比如表3所示。

从任务设置看，现有研究通常沿用少样本目标检测的两阶段范式：首先在源域基础类数据上进行充分预训练，然后在目标域新类别的少量样本上进行适配。对于元学习方法，常采用支持集—查询集的元训练方式，而对于微调类方法，则通常在基础类模型基础上对新类别或目标域样本进行少量迭代更



图 8　基于 Fu 等人（2024）整理的代表性 CD-FSOD 测评基准

Fig. 8　A representative CD-FSOD evaluation benchmark summarized from Fu et al.（2024）

新。在跨域设定下，实验重点不再仅比较同域新类别检测精度，而是同时考察模型在域偏移条件下对新类别的检测能力。因此，就本文关注的航空遥感场景而言，更合理的数据与任务设置通常应满足两个条件：1）源域与目标域在平台、分辨率、区域或传感器层面存在可观差异；2）目标域仅提供极少量标注实例，用以模拟真实应用中新场景、少样本的检测需求。

3.2　常用评测指标

跨域少样本目标检测的评测通常仍以目标检测任务中的精度指标为基础，同时结合少样本设定和跨域设定的特点，对模型在目标域中的检测效果进行综合评价。由于该任务既关注目标分类与定位的准确性，也关注模型在有限样本和域偏移条件下的适应能力，因此评测指标不仅要反映整体检测性能，还应体现模型对新类别和目标域分布的泛化能力。

在目标检测领域，精确度（precision，P）和召回率（recall，R）是两个基本指标：前者反映检测结果中正确目标所占比例，后者反映真实目标被成功检出

表3　跨域少样本目标检测基准与航空遥感场景目标检测数据集

Table 3　Cross-domain few-shot object detection benchmarks and aerial remote sensing object detection datasets

数据集	场景类型	图像数量	类别数	标注	主要特点
MoFSOD	跨域少样本目标检测基准	由10个目标域构成，涉及30 272幅测试图像	未规定源域，目标域共462类	水平框	面向域差别不大的多领域少样本检测，强调跨数据域泛化
CD-FSOD	跨域少样本目标检测基准	由1个源域和6个目标域构成，涉及8 658幅测试图像	源域80个类别，目标域共57类	水平框	强调源域与目标域之间的显著分布差异
DIOR	卫星航空遥感	23 463幅	20类	水平框	类别丰富，尺度变化和场景变化明显
DOTA	卫星航空遥感	2 806幅	15类	任意四边形	大尺寸图像，旋转目标和密集目标突出
xView	卫星航空遥感	1 127幅	60类	水平框	类别多，实例数量大，区域覆盖广
VEDAI	无人机航空遥感	1 210幅	9类	水平框/旋转框	聚焦小目标车辆检测
NWPU VHR-10	无人机航空遥感	800幅	10类	水平框	数据规模较小，常用于遥感少样本实验

的比例。平均精度（average precision，AP）用于综合衡量单一类别在不同置信度阈值下的检测效果，通常理解为精确率—召回率曲线的整体表现；平均精度均值（mean average precision，mAP）则对多个类别的AP取平均，用于评价检测器整体性能。

实际评测中，AP通常结合不同交并比（intersection over union，IoU）阈值计算。AP50表示IoU阈值为0.5时的平均精度，AP75表示IoU阈值为0.75时的平均精度，mAP@50:95表示在0.50到0.95多个IoU阈值下求平均，可更全面反映分类与定位质量。常用错误统计量包括真正例（true positive，TP）、假正例（false positive，FP）和假负例（false negative，FN）。

在少样本目标检测中，研究者通常进一步关注新类别检测性能。这是因为少样本设定的核心目标在于提升模型对新类别的检测能力，因此除了报告整体mAP外，还常单独报告新类别上的AP或mAP，以体现模型在有限监督条件下对新类别的适应效果。部分研究还会同时给出基础类别与新类别的检测结果，用于分析模型在迁移适配过程中是否出现对基础类知识的遗忘或对新类别学习不足的问题。

在跨域设定下，评测重点则进一步扩展为目标域泛化性能。具体而言，模型通常在源域上完成训练或预训练，再在目标域少量样本条件下进行适配，并最终在目标域测试集上评估检测精度。因此，目标域上的mAP、AP50或novel mAP往往是衡量方法有效性的关键指标。若方法面向多个目标域进行验证，还可通过比较不同目标域上的平均表现，分析模型对不同域偏移类型的鲁棒性和稳定性。因此，跨域少样本目标检测的评测不应仅关注单一mAP数值，还应结合目标域整体性能、新颖类别性能、不同shot设置下的性能变化以及不同目标域之间的方差进行综合分析。尤其在航空遥感场景中，不同目标域可能对应不同平台、传感器或地理区域，若模型在某一目标域表现较好而在另一目标域明显退化，则说明其跨域迁移稳定性仍然不足。

此外，在部分研究中还会结合样本规模敏感性分析进行评估，例如比较1-shot、3-shot、5-shot或10-shot等不同设定下的检测性能变化，以观察模型在样本逐渐增加时的收益趋势。这类分析有助于揭示不同方法对样本规模的依赖程度，也更贴近真实应用中目标域标注资源有限但可逐步补充的情形。

总体来看，跨域少样本目标检测的评测指标体系仍以mAP和AP50等标准检测指标为核心，并通过新类别性能、目标域测试结果以及不同shot数下的性能变化，进一步反映模型在少样本与跨域双重约束下的检测能力。对于航空遥感场景而言，合理选择和统一这些评测指标，不仅有助于公平比较不同方法的性能，也有助于更准确地判断模型在真实复杂环境中的应用潜力。

为便于读者复用和后续补充实验结果，表4进一步汇总本文涉及方法在COCO→DIOR航空遥感跨域少样本设定下的代表性性能。表中列出相关方

法的发表期刊或会议、是否需要在COCO→DIOR设定下进行目标域微调，以及1-shot、5-shot、10-shot和3种设置的平均性能。

3.3 航空遥感目标检测的特殊性

与自然场景目标检测相比，航空遥感目标检测在数据特征、任务需求和应用环境等方面均具有明显特殊性，这也是跨域少样本目标检测在该场景下更具挑战性的根本原因。首先，航空遥感图像通常采用俯视视角成像，目标在图像中的表观特征与自然场景下存在较大差异。大量目标以顶视或斜视形式出现，缺乏明显的侧面纹理和结构细节，使得类别区分更加依赖整体轮廓、局部几何形状以及上下文环境，而这类特征在少样本条件下往往更难稳定学习。

其次，小目标多、尺度变化大是航空遥感目标检测中最典型的问题之一。由于成像高度和覆盖范围较大，车辆、舰船以及飞机等目标在图像中通常只占据较少像素，部分目标甚至仅呈现为模糊斑点；同时，不同平台和分辨率条件下，同一类别目标又可能表现出明显的尺度差异。这使得检测器不仅需要具备较强的多尺度表征能力，还要在极少样本条件下保持对小目标的稳定识别与定位能力。

再次，背景复杂且目标分布密集进一步增加了检测难度。航空遥感图像往往覆盖机场、港口、道路、城区和农田等大尺度复杂场景，地物种类丰富，背景纹理多样，容易与目标形成视觉混淆。此外，许多遥感目标呈现密集排列、局部遮挡或方向交叉分布的特点，导致候选区域生成、边界框回归和类别判别均面临更大压力。在这种情况下，模型若过度依赖背景先验，便更容易在跨域迁移过程中产生误检与漏检。

同时，航空遥感目标还具有较强的方向任意性和形态变化性。许多目标在图像中不存在固定朝向，且在不同拍摄高度、不同区域场景下，其外观形态、排列方式和尺度比例都可能发生明显变化。这种变化不仅增加了模型学习稳定类别表示的难度，

表4 本文所涉及算法在航空遥感跨域少样本情境下性能

Table 4 Performance of reviewed methods in aerial remote sensing CD-FSOD scenarios

方法	发表期刊/会议	微调	mAP/%			
			1-shot	5-shot	10-shot	平均
Meta R-CNN	ICCV2019	√	7.8	17.7	20.6	15.4
TFA w/cos	arXiv2020	√	8.0	18.1	20.5	15.5
FSCE	CVPR2021	√	8.6	18.7	21.9	16.4
DeFRCN	ICCV2021	√	9.3	18.9	22.9	17.0
Distill-CDFSOD	ICASSP2023	√	10.5	19.1	26.5	18.7
ViTDeT-FT	ECCV2022	√	12.9	23.3	29.4	21.9
Detic	ECCV2022	×	0.1	0.1	0.1	0.1
Detic-FT	ECCV2022	√	4.1	12.1	15.4	10.5
DE-ViT	CoRL2024	×	2.7	7.8	8.4	6.3
DE-ViT-FT	CoRL2024	√	14.7	23.4	25.6	21.2
CD-ViTO	ECCV2024	√	17.8	26.9	30.8	25.2
Meta-DETR	TPAMI2023	√	11.1	19.4	19.9	16.8
Grounding DINO	ECCV2024	√	14.8	29.6	37.2	27.2
ETS	CVPRW2025	√	12.7	29.3	37.5	26.5
Domain-RAG	NeurIPS2025	√	18.0	**31.5**	**39.0**	**29.5**
RT CD-MFSOD	WACV2025	√	14.3	26.9	31.4	24.2
StyleProto	AAAI2026	√	**22.6**	27.7	32.1	27.5

注：加粗字体表示各列最优结果。“√”、“×”分别表示COCO→DIOR设定下是或否在目标域微调。

也使得传统基于水平框和固定外观模式的检测方法在遥感场景下受到限制。

此外，航空遥感数据通常具有更加显著的域差异复合性。平台更换、传感器变化、空间分辨率差异、地理区域变化以及季节和光照条件变化，常常同时作用于同一数据域中，使航空遥感目标检测面临的域偏移远比通常自然图像更复杂。对于跨域少样本目标检测而言，这意味着模型不仅要解决样本少的问题，还必须具备较强的跨平台、跨区域和跨成像条件适应能力。

最后，从应用需求看，航空遥感目标检测往往服务于灾害监测、交通监管、资源调查和军事侦察等实际任务，对模型的鲁棒性、实时性和可部署性提出了更高要求。真实应用中，目标域数据往往难以及时获得充分标注，且新场景切换频繁，因此模型需要在极少样本甚至弱监督条件下快速完成适配。这种面向真实复杂环境的应用导向，也决定了航空遥感目标检测不能简单沿用自然场景中的标准设定，而需要结合遥感数据特征与任务需求进行针对性设计。

上述特点不仅影响检测模型设计，也会直接影响数据集构建和评测方式。小目标和密集目标要求数据集具有较高标注精度，并要求评测指标能够反映定位误差对检测结果的影响；方向任意和形态变化使水平框标注难以完全描述目标几何特征，因此旋转框或任意四边形标注在部分遥感数据集中更为重要；复杂背景和多源域差异则要求评测协议不仅比较同域检测精度，还应考察模型在跨平台、跨区域和跨传感器条件下的迁移稳定性。因此，航空遥感跨域少样本目标检测的数据集与评测体系，应同时兼顾检测精度、样本效率和跨域鲁棒性。

综上，航空遥感目标检测的特殊性主要体现在俯视成像、小目标密集分布、背景复杂、方向任意、域差异复合以及应用条件苛刻等多个方面。这些特点共同决定了该场景下的跨域少样本目标检测比通用视觉任务更具挑战性，也为相关方法设计与评测体系构建提出了更高要求。

3.4 本节小结

本节从数据集、任务设置、评测指标和航空遥感应用特点4个方面，对跨域少样本目标检测的研究基础进行了梳理。现有通用跨域少样本检测基准为方法比较提供了统一协议，航空遥感数据集则体现了小目标、复杂背景、尺度变化和方向任意等实际挑战。在评测方面，mAP、AP50和novel mAP等指标仍是主要依据，但仅依赖单一精度指标难以充分反映模型在复杂域偏移和少样本约束下的稳定性。因此，面向航空遥感场景的跨域少样本目标检测评测体系，应进一步关注目标域泛化性能、不同shot设置下的样本效率、多目标域条件下的鲁棒性以及多源异构数据适配能力。这些问题也为后续研究中的瓶颈分析和发展趋势讨论提供了基础。

4 现存问题与发展趋势

尽管跨域少样本目标检测已取得一定进展，但对于航空遥感场景而言，现有研究仍处于持续探索阶段。一方面，航空遥感数据具有平台多样、传感器异构、空间分辨率差异显著、场景背景复杂和目标尺度变化大等特点，使模型在不同数据域之间的迁移稳定性面临较大挑战；另一方面，真实应用中目标域标注样本通常极为有限，检测模型不仅需要在少样本条件下完成类别判别，还需要保持稳定的目标定位能力和复杂背景抑制能力。与此同时，视觉语言模型和多模态大模型的发展为该方向提供了新的技术机遇，但其在航空遥感场景中的专用适配、计算效率和可信迁移能力仍有待进一步研究。因此，有必要围绕跨域迁移稳定性、少样本鲁棒检测、多源异构数据适配、数据集与评测协议以及大模型迁移等方面，对该方向的现存问题与发展趋势进行进一步分析。

4.1 复杂域偏移与跨域迁移稳定性

航空遥感场景中的域偏移通常并非由单一因素引起，而是平台、传感器、分辨率、地理区域、季节光照和背景环境等多种因素共同作用形成的复合偏移。

相比自然图像中的风格变化或场景变化，航空遥感图像中的域偏移往往同时影响目标外观、尺度分布、背景纹理和类别先验，使源域训练得到的检测模型在目标域中容易出现分类置信度下降、定位偏差增大以及误检漏检增多等问题。

现有方法通常从特征对齐、风格迁移、实例级适配或教师—学生学习等角度缓解域偏移，但多数方法仍主要针对某一类分布差异进行建模。在实际航空遥感任务中，平台差异、分辨率差异和区域背景差

异往往同时存在，单一粒度的域对齐可能难以充分刻画复杂域偏移；若对齐约束过强，还可能削弱类别判别信息，导致负迁移。尤其在目标域样本极少的情况下，模型难以准确估计目标域真实分布，使跨域适配过程更容易受到少量样本偏差的影响。

未来研究需要更加关注跨域迁移的稳定性。一方面，应构建能够同时描述全局场景差异、局部目标差异和类别条件分布差异的统一域建模框架，避免仅依赖整体特征分布对齐；另一方面，可结合因果表征学习、域泛化、测试时适配和不确定性估计等思想，增强模型对未知目标域的稳健适应能力。此外，面向航空遥感场景，还需要进一步研究跨平台、跨区域、跨分辨率和跨传感器条件下的迁移稳定性评价机制，从而更准确地判断模型在真实复杂环境中的应用可靠性。

4.2 少样本条件下的鲁棒检测能力

少样本条件下的检测稳定性是跨域少样本目标检测面临的另一核心问题。与图像分类任务相比，目标检测不仅需要判断目标类别，还需要完成候选区域生成、目标定位和边界框回归，因此对样本数量、样本质量和特征表达的要求更高。当目标域仅有少量标注实例时，检测器容易出现过拟合、类别边界模糊、定位回归不稳定和基础类知识遗忘等问题。

在航空遥感场景中，这一问题更加突出。首先，遥感图像中大量目标具有小尺度、低纹理和弱边缘等特点，少量支持样本往往难以充分覆盖目标外观变化。其次，车辆、舰船和飞机等类别在不同区域和不同成像条件下可能呈现较强类内差异，而部分类别之间又具有较高外观相似性，导致少样本类别表征容易不稳定。再次，复杂背景和密集分布会增加误检和漏检风险，使模型在少样本适配阶段更容易学习到偶然背景先验，而非稳定的目标判别特征。

未来研究可从3个方面提升少样本鲁棒检测能力。1）应加强支持样本质量评估与选择机制，降低异常样本、遮挡样本和背景干扰样本对模型适配的影响。2）应构建更稳健的类别表征学习方法，例如结合类别原型、对比学习、多尺度特征增强和不确定性建模，提高少样本条件下类别边界的稳定性。3）应进一步关注分类与定位的协同优化问题，避免模型只提升类别识别能力而忽视边界框定位精度。对于航空遥感目标检测而言，鲁棒少样本检测不仅要求模型“能识别”，还要求模型在小目标、密集目标和复杂背景下“能稳定定位”。

4.3 数据集与真实应用脱节问题

合理的数据集和评测协议是推动跨域少样本目标检测研究发展的基础。现有研究中，部分跨域少样本目标检测基准为方法比较提供了统一实验环境，但这些基准多由已有数据集重新划分或组合构建，其域差异、类别分布和样本设置往往与真实航空遥感应用仍存在一定距离。真实任务中，数据来源更加复杂，目标类别更加不均衡，成像条件变化更加多样，且目标域标注样本的获取通常具有明显随机性和不完整性。

在航空遥感场景中，公开数据集虽然覆盖了机场、港口、城市道路、车辆目标、舰船目标和飞机目标等典型场景，但在跨平台、跨传感器、多时相、多区域和多任务联合评测方面仍存在不足。许多实验设定主要关注固定标注样本数下的检测精度，而对目标域样本选择差异、类别长尾分布、标注噪声、跨域泛化方差和实际部署约束关注不足。因此，部分方法在标准基准上取得较好结果，并不必然意味着其能够稳定适用于真实遥感应用。

未来需要构建更贴近实际任务需求的跨域少样本遥感检测数据集与评测协议。一方面，应在数据层面增加跨平台、跨区域、跨分辨率、跨季节和跨传感器的样本覆盖，以更真实地反映航空遥感场景中的复合域偏移；另一方面，应在评测层面不仅报告单一 mAP 或 AP50 结果，还应进一步分析不同标注样本数、不同目标域、不同类别频次和不同目标尺度下的性能变化。此外，还可引入跨域稳定性、样本效率、标注成本和推理效率等指标，使评测结果更能反映方法的实际应用价值。

4.4 多源异构数据适配问题

航空遥感数据具有明显的多源异构特征。不同任务可能涉及光学图像、红外图像、多光谱图像、SAR 图像以及不同高度、不同视角和不同分辨率下获取的数据。这些数据不仅在颜色、纹理和空间结构上存在差异，在成像机理、噪声模式和目标可见性方面也可能完全不同。例如，光学图像更依赖颜色和纹理信息，SAR 图像则反映目标的电磁散射特性，红外图像更突出热辐射差异。不同模态之间的特征空间差异，使跨域少样本检测面临更复杂的适配问题。

目前，多数跨域少样本目标检测研究仍主要集

中于光学图像之间的域迁移，对光学—SAR、光学—红外、多光谱—可见光等异构模态之间的少样本适配关注不足。在目标域样本极少的条件下，模型不仅难以学习目标类别的稳定特征，也难以建立不同模态之间的可靠对应关系。若直接套用常规域对齐或微调策略，可能难以有效处理不同成像机理导致的结构性差异。

未来研究需要进一步探索面向多源异构遥感数据的统一表征学习与跨模态适配方法。一方面，可通过多模态预训练、跨模态对比学习和语义一致性约束，学习不同遥感模态之间共享的目标语义表示；另一方面，可结合视觉语言模型或遥感领域知识，将类别语义、目标属性和成像机理信息引入跨模态检测过程，提升模型对异构数据的理解能力。此外，还需要研究轻量化跨模态适配策略，使模型能够在少量目标域样本条件下快速适应新传感器和新成像条件，从而提高航空遥感检测系统在复杂数据来源下的实用性。

4.5 大模型迁移的机遇与挑战

视觉基础模型、视觉语言模型和多模态大模型的发展，为跨域少样本目标检测提供了新的研究机遇。与传统检测模型相比，大模型通常依托大规模图像、文本或图文数据预训练，具备更强的通用表征能力、语义理解能力和开放词汇泛化能力。在目标域标注样本极少的情况下，大模型可以为检测任务提供更好的初始特征、更丰富的类别语义先验以及更灵活的提示式适配方式，从而在一定程度上缓解传统少样本检测模型对目标域标注数据的依赖。

然而，大模型在航空遥感场景中的应用仍面临多方面挑战。首先，航空遥感图像通常分辨率较高，而目标尺度较小，直接将整幅图像输入大模型会带来较高计算和存储开销，也可能导致小目标特征在下采样过程中被削弱；其次，通用大模型主要基于自然图像或通用图文数据进行预训练，其学习到的语义先验未必能够充分覆盖遥感场景中的专业目标类别、俯视视角和复杂地物背景；再次，大模型虽然具有较强识别和语义理解能力，但在密集小目标定位、旋转目标检测和跨传感器适配方面仍可能存在不足。

未来基于大模型的跨域少样本遥感检测研究，应更加关注遥感专用适配与高效部署。一方面，可构建遥感领域预训练数据和遥感语义描述体系，使基础模型获得更符合航空遥感场景的视觉与语言知识；另一方面，可采用参数高效微调、提示学习、专家路由、分块推理和多尺度特征融合等策略，降低大模型在高分辨率遥感图像上的应用成本。同时，还可探索大模型与传统检测器之间的互补融合：利用传统检测器的精确定位能力弥补大模型定位不足，利用大模型的语义泛化能力提升检测器对新域和新类别的适应能力。

4.6 本节小结

总体来看，航空遥感场景下的跨域少样本目标检测仍面临多重挑战。复杂域偏移使模型跨域迁移稳定性难以保证，少样本约束使类别判别和目标定位更容易受到样本偏差影响，数据集与真实应用之间的差距限制了方法评测结论的外推能力，多源异构数据进一步增加了跨模态适配难度，而大模型迁移虽然带来了新的技术可能，也对遥感专用知识注入、高分辨率图像处理和轻量化部署提出了更高要求。

未来，该方向需要从单一方法改进转向系统性研究。一方面，应构建更贴近真实应用的跨域少样本遥感检测数据集和评测协议，增强方法比较的可信度；另一方面，应融合域适应、少样本学习、多模态语义建模、生成式增强和基础模型迁移等技术，形成面向复杂航空遥感场景的综合检测框架。随着遥感专用数据资源、多模态基础模型和高效适配技术的持续发展，跨域少样本目标检测有望在航空遥感智能感知、应急监测和复杂目标解译等任务中发挥更大作用。

5 结　语

航空遥感场景下的跨域少样本目标检测，是面向真实复杂环境目标感知需求提出的一类重要研究问题。该任务同时受到目标域样本稀缺和源域与目标域分布差异显著的双重制约，较传统目标检测、少样本目标检测和跨域目标检测具有更高的研究难度，也更贴近航空遥感智能解译中的实际应用需求。

本文围绕该问题进行了系统综述。首先，从任务定义、相关概念和航空遥感场景中的域差异表现出发，明确了跨域少样本目标检测与传统目标检测、少样本目标检测、跨域目标检测和遥感目标检测之间的区别；其次，按照技术路线将现有方法归纳为基

于迁移与域对齐、基于元学习与度量学习、基于生成式数据增强与扩充、基于视觉语言对齐以及基于大模型的方法，并从核心思想、适用条件、主要优势和局限性等方面进行了比较分析；在此基础上，进一步总结了相关数据集、任务设置、常用评测指标以及航空遥感目标检测的小目标、密集分布、复杂背景和方向任意等特点，梳理了该方向的方法基础与评测依据。

综合已有研究可以看出，跨域少样本目标检测正由单一的微调、域对齐或少样本适配策略，逐步发展为融合域适应、少样本判别、数据增强、语义建模和基础模型迁移的综合性研究方向。不同技术路线之间并非相互独立，而是在特征迁移、样本利用、数据扩充、语义先验和通用表征等层面形成互补关系。对于航空遥感场景而言，未来研究仍需重点解决复杂域偏移下的迁移稳定性、少样本条件下的鲁棒检测能力、数据集与真实应用之间的差距、多源异构数据适配以及大模型遥感专用迁移等问题。

总体而言，跨域少样本目标检测为航空遥感智能解译提供了一种面向少样本、高变化和复杂场景的有效研究范式。随着遥感专用数据资源的持续完善、多模态语义建模能力的增强以及视觉基础模型和多模态大模型的进一步发展，该方向有望在灾害监测、交通监管、资源调查、目标监视和应急响应等航空遥感应用中发挥更重要的作用，并为构建高效、稳健和可迁移的遥感智能感知系统提供技术支撑。

参考文献（References）

Chen Y H, Li W, Sakaridis C, Dai D X and Van Gool L. 2018. Domain adaptive faster R-CNN for object detection in the wild//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3339-3348 [DOI: 10.1109/CVPR.2018.00352]

Cheng G, Han J W, Zhou P C and Guo L. 2014. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. ISPRS Journal of Photogrammetry and Remote Sensing, 98: 119-132 [DOI: 10.1016/j. isprsjprs. 2014. 10.002]

Fan Q, Zhuo W, Tang C K and Tai Y W. 2020. Few-shot object detection with attention-RPN and multi-relation detector//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4012-4021 [DOI: 10.1109/CVPR42600.2020.00407]

Feng C J, Zhong Y J, Jie Z Q, Chu X X, Ren H B, Wei X L, et al. 2022. PromptDet: towards open-vocabulary detection using uncurated images//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 701-717 [DOI: 10.1007/978-3-031-20077-9_41]

Fu Y Q, Wang Y, Pan Y X, Huai L, Qiu X Y, Shangguan Z Y, et al. 2024. Cross-domain few-shot object detection via enhanced open-set object detector//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 247-264 [DOI: 10.1007/978-3-031-73636-0_15]

Gupta A, Dollár P and Girshick R. 2019. LVIS: a dataset for large vocabulary instance segmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5351-5359 [DOI: 10.1109/CVPR. 2019.00550]

Jiang Q, Huo J, Chen X, Xiong Y, Zeng Z, Chen Y, et al. 2026. Detect anything via next point prediction//Proceedings of 2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Denver, USA: IEEE: 25472-25483

Lam D, Kuzma R, McGee K, Dooley S, Laielli M, Klaric M, et al. 2018. xView: objects in context in overhead imagery//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, UT, USA: Computer Vision Foundation: 1-10

Lee K, Yang H, Chakraborty S, Cai Z W, Swaminathan G, Ravichandran A, et al. 2022. Rethinking few-shot object detection on a multi-domain benchmark//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 366-382 [DOI: 10.1007/978-3-031-20044-1_21]

Li C C, Du D W, Zhang L B, Wen L Y, Luo T J, Wu Y J, et al. 2020a. Spatial attention pyramid network for unsupervised domain adaptation//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 481-497 [DOI: 10.1007/978-3-030-58601-0_29]

Li K, Wan G, Cheng G, Meng L Q and Han J W. 2020b. Object detection in optical remote sensing images: a survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 159: 296-307 [DOI: 10.1016/j.isprsjprs.2019.11.023]

Li L H, Zhang P C, Zhang H T, Yang J W, Li C Y, Zhong Y W, et al. 2022a. Grounded language-image pre-training//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 10955-10965 [DOI: 10.1109/CVPR52688.2022.01069]

Li Y, Qiu X Y, Fu Y Q, Chen J, Qian T W, Zheng X, et al. 2025. Domain-RAG: retrieval-guided compositional image generation for cross-domain few-shot object detection//Proceedings of the 39th Annual Conference on Neural Information Processing Systems. San Diego, USA: Curran Associates, Inc.

Li Y J, Dai X L, Ma C Y, Liu Y C, Chen K, Wu B C, et al. 2022b.

Cross-domain adaptive teacher for object detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 7571-7580 [DOI: 10.1109/CVPR52688.2022.00743]

Lin S B, Wang K, Zeng X Y and Zhao R. 2023. Explore the power of synthetic data on few-shot object detection//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Vancouver, Canada: IEEE: 638-647 [DOI: 10.1109/CVPRW59228.2023.00071]

Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]

Liu C H, Xiang X Z, Duan Z X, Li W B, Fan Q and Gao Y. 2025a. Don't need retraining: a mixture of DETR and vision foundation models for cross-domain few-shot object detection//Proceedings of the 39th Annual Conference on Neural Information Processing Systems. San Diego, USA: Neural Information Processing Systems Foundation

Liu S L, Zeng Z Y, Ren T H, Li F, Zhang H, Yang J, et al. 2024. Grounding DINO: marrying DINO with grounded pre-training for open-set object detection//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 38-55 [DOI: 10.1007/978-3-031-72970-6_3]

Liu Y X, Pan J C and Zhang B C. 2025b. Control copy-paste: controllable diffusion-based augmentation method for remote sensing few-shot object detection//2025 IEEE International Geoscience and Remote Sensing Symposium. Brisbane, Australia: IEEE: 7954-7958 [DOI: 10.1109/IGARSS55030.2025.11242535]

Madan A, Peri N, Kong S and Ramanan D. 2024. Revisiting few-shot object detection with vision-language models//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #617 [DOI: 10.52202/079017-0617]

Pan J C, Liu Y X, He X, Peng L, Li J H, Sun Y Z, et al. 2025b. Enhance then search: an augmentation-search strategy with foundation models for cross-domain few-shot object detection//Proceedings of 2015 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Nashville, USA: IEEE: 1539-1547 [DOI: 10.1109/CVPRW67362.2025.00143]

Pan J C, Liu Y X, Fu Y Q, Ma M Y, Li J H, Paudel D P, et al. 2025a. Locate anything on earth: advancing open-vocabulary object detection for remote sensing community//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI Press: 6281-6289 [DOI: 10.1609/aaai.v39i6.32672]

Razakarivony S and Jurie F. 2016. Vehicle detection in aerial imagery: a small target detection benchmark. Journal of Visual Communication and Image Representation, 34: 187-203 [DOI: 10.1016/j.jvcir.2015.11.002]

Ren S, He K, Girshick R and Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [EB/OL]. [2026-04-16]. https://arxiv.org/pdf/1506.01497.pdf

Saito K, Ushiku Y, Harada T and Saenko K. 2019. Strong-weak distribution alignment for adaptive object detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 6949-6958 [DOI: 10.1109/CVPR.2019.00712]

Shangguan Z Y, Seita D and Rostami M. 2025. Cross-domain multimodal few-shot object detection via rich text//Proceedings of 2025 IEEE/CVF Winter Conference on Applications of Computer Vision. Tucson, USA: IEEE: 6570-6580 [DOI: 10.1109/WACV61041.2025.00640]

Shi Z H, Wu C W, Li C J, You Z Z, Wang Q and Ma C C. 2023. Object detection techniques based on deep learning for aerial remote sensing images: a survey. Journal of Image and Graphics, 28(9): 2616-2643 (石争浩, 仵晨伟, 李成建, 尤珍臻, 王泉, 马城城. 2023. 航空遥感图像深度学习目标检测技术研究进展. 中国图象图形学报, 28(9): 2616-2643) [DOI: 10.11834/jig.221085]

Sun B, Li B H, Cai S C, Yuan Y and Zhang C. 2021. FSCE: few-shot object detection via contrastive proposal encoding//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 7348-7358 [DOI: 10.1109/CVPR46437.2021.00727]

Tang D T, Cao X Y, Wu X, Li J L, Yao J, Bai X R, et al. 2025. AeroGen: enhancing remote sensing object detection with diffusion-driven data generation//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 3614-3624 [DOI: 10.1109/CVPR52734.2025.00342]

Vu A K N, Truong Q T, Nguyen V T, Ngo T D, Do T T and Nguyen T V. 2025. Multi-perspective data augmentation for few-shot object detection//Proceedings of the 13th International Conference on Learning Representations. Singapore, Singapore: OpenReview.net

Xia G S, Bai X, Ding J, Zhu Z, Belongie S, Luo J B, et al. 2018. DOTA: a large-scale dataset for object detection in aerial images//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3974-3983 [DOI: 10.1109/CVPR.2018.00418]

Xiong W T. 2023. CD-FSOD: a benchmark for cross-domain few-shot object detection//Proceedings of 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes, Greece: IEEE: 1-5 [DOI: 10.1109/ICASSP49357.2023.10096216]

Yan X P, Chen Z L, Xu A N, Wang X X, Liang X D and Lin L. 2019. Meta R-CNN: towards general solver for instance-level low-shot learning//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9576-9585 [DOI: 10.1109/ICCV.2019.00967]

Yuan X, Cheng G, Li G, Dai W, Yin W X, Feng Y C, et al. 2023.

Progress in small object detection for remote sensing images. Journal of Image and Graphics, 28(6): 1662-1684 (袁翔, 程塨, 李戈, 戴威, 尹文昕, 冯瑛超, 等. 2023. 遥感影像小目标检测研究进展. 中国图象图形学报, 28(6): 1662-1684) [DOI: 10.11834/jig.221202]

Zhang G J, Luo Z P, Cui K W, Lu S J and Xing E P. 2023a. Meta-DETR: image-level few-shot detection with inter-class correlation exploitation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(11): 12832-12843 [DOI: 10.1109/TPAMI.2022.3195735]

Zhang H, Li F, Liu S L, Zhang L, Su H, Zhu J, et al. 2023b. DINO: DETR with improved DeNoising anchor boxes for end-to-end object detection//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net

Zhang W L and Wang Y X. 2021. Hallucination improves few-shot object detection//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 13003-13012 [DOI: 10.1109/CVPR46437.2021.01281]

Zhang X Y, Liu Y H, Wang Y T and Boularias A. 2025. Detect everything with few examples//Proceedings of the 8th Conference on Robot Learning. Munich, Gemany: PMLR: 3986-4004

Zhong Y W, Yang J W, Zhang P C, Li C Y, Codella N, Li L H, et al. 2022. RegionCLIP: region-based language-image pretraining//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 16772-16782 [DOI: 10.1109/CVPR52688.2022.01629]

作者简介

陈布衣，男，博士研究生，主要研究方向为跨域少样本学习、深度学习、计算机视觉。
E-mail：hermite_buyi@mail.nwpu.edu.cn
郭宸瑞，通信作者，男，工程师，主要研究方向为目标检测、多模态大模型和雷达光学系统设计。E-mail：g-ch-r@163.com
单淳，男，本科生，主要研究方向为视觉语言导航。
E-mail：shanchun@mail.nwpu.edu.cn
张磊，男，教授，主要研究方向为深度学习、计算机视觉。
E-mail：nwpuzhanglei@nwpu.edu.cn
魏巍，男，教授，主要研究方向为深度学习、计算机视觉。
E-mail：weiweinwpu@nwpu.edu.cn