小样本目标检测研究综述

史燕燕[1)] 史殿习[2),3)] 乔子腾[2),3)] 张 轶[2)] 刘洋洋[1),3)] 杨绍武[1)]

1 国防科技大学 计算机学院 长沙 410073

2 军事科学院国防科技创新研究院 北京 100071

3 天津(滨海)人工智能创新中心 天津 300457

摘　要 数据驱动下的深度学习技术在计算机视觉领域取得重大突破，但模型的高性能严重依赖于大量标注样本的训练。然而在实际场景当中，大规模数据的获取和高质量的标注十分困难，限制了其在特定应用领域的进一步推广。近年来小样本学习在目标检测领域的发展，为解决上述问题提供了新的研究思路。小样本目标检测旨在通过少量标注样本实现对图像中目标的分类和定位。本文从任务和问题、学习策略、检测方法、数据集与实验评估等角度出发，对当前小样本目标检测的研究成果加以梳理和总结。首先，系统性地阐述了小样本目标检测的任务定义及核心问题，并讨论了当前方法采用的学习策略。其次，从工作原理角度出发，将现有检测方法归纳总结为四类，对这四类检测方法的核心思想、特点、优势及存在的不足进行了系统性的阐述，为不同场景下选择不同的方法提供了依据。之后，本文对目前小样本目标检测采用的典型数据集、实验设计及性能评估指标进行了深入分析，进而对四类典型方法在数据集上的实验结果进行概括总结，尤其是对部分典型方法的检测性能进行了系统性对比分析。最后，立足于现有方法的优势和劣势，我们指出当前方法面临的挑战，并对下一阶段小样本目标检测技术未来的发展趋势提出了见解，期望为该领域的后续研究提供参考。

关键词　深度学习；目标检测；小样本学习；小样本目标检测

中图法分类号 TP18

A Survey on Recent Advances in Few-shot Object Detection

SHI Yan-yan[1)] SHI Dian-xi[2),3)] QIAO Zi-teng[2),3)] ZHANG Yi[2)] Liu Yang-yang[1),3)] Yang Shao-wu[1)]

1) (College of Computer, National University of Defense Technology. Changsha, 410073)

2) (Artificial Intelligence Research Center (AIRC), National Innovation Institute of Defense Technology (NIIDT), Beijing 100071)

3) (Tianjin Artificial Intelligence Innovation Center, Tianjin 300457)

Abstract In recent years, with the substantial progress in large data sets and hardware technologies and the tremendous continuous breakthroughs of deep learning models in various fields, several fundamental computer vision tasks based on deep neural network models have gradually matured. Traditional supervised machine learning models must be trained with large-scale labeled data, while visual data in the real world presents a significant long-tail effect. Data-rich categories occupy the majority of the total categories. However, in practical application scenarios, scarce categories may make data acquisition and labeling difficult due to privacy, security, high labeling cost, and

本课题得到科技部科技创新2030—重大项目(No.2020AAA0104802)、国家自然科学基金集成项目“基于群体智能机器人操作系统的集成与创新”(No.91948303)资助. **史燕燕**，博士研究生，主要研究领域为计算机视觉、小样本学习、小样本目标检测等. E-mail: yany_shi@163.com. **史殿习**（**通信作者**），博士，博士生导师，研究员，中国计算机学会(CCF)会员(14099M)，主要研究领域为人工智能、分布式计算等. E-mail: dxshi@nudt.edu.cn. **乔子腾**，博士研究生，主要研究领域为计算机视觉、域适应目标检测等. E-mail: ztqiao99@163.com. **张　轶**，博士，助理研究员，主要研究领域为信息安全、人工智能安全等. E-mail: jxnzdl@126.com. **刘洋洋**，硕士研究生，主要研究领域为人工智能、目标检测等. E-mail: liuyangyang@nudt.edu.cn. **杨绍武**，博士，副研究员，主要研究领域为人工智能、SLAM等. E-mail: shaowu.yang@nudt.edu.cn.

other factors. Accessing large-scale data and high-quality annotated samples is often challenging. In few-shot learning scenarios, the traditional deep learning algorithm cannot be fully trained, which makes the deep neural network easy to overfit, and the generalization ability of the model is seriously affected. The recent deep learning techniques cannot meet the needs of scenarios with fewer labeled training samples. Unlike deep neural networks, the visual system of humans can exhibit a remarkable ability to learn novel concepts from a few examples quickly. Such data-efficient ability is precisely what the practical application needs. The universality and generalization capabilities of existing data-driven deep learning technology are far from reaching the level of human cognitive learning. Inspired by the human learning mode, few-shot learning is gradually gaining attention in the academic field. With the deepening of the research, developing few-shot learning in object detection provided a new research idea for solving the above problems. Few-shot object detection aims to classify and locate objects in images by a small number of labeled samples. In the scenario of data scarcity, how to exploit a few labeled samples to learn, design a detection model with good generalization ability, and extend it to new tasks, is an urgent problem to be solved in few-shot object detection. In this paper, we sort out the research findings of few-shot object detection from the perspectives of tasks and problems, learning strategies, detection methods, datasets, and experimental evaluation. First, the task definition and core problem of few-shot object detection are systematically described, and learning strategies are discussed. Second, from the principle perspective, the existing few-shot object detection methods are summarized into four categories, including meta-learning based, transfer-learning based, data augmentation based, and metric-learning based methods. The core ideas, characteristics, advantages, and shortcomings of the four detection methods are systematically elaborated, providing a basis for choosing different methods in different scenarios. After that, this paper provides an in-depth analysis of the typical datasets, experimental design, and performance evaluation indexes currently used for few-shot object detection. Then we summarize the experimental results of four types of typical methods on datasets and systemically compare and analyze the detection performance of some typical methods. Finally, we point out the challenges of the current methods and provide insights on the future development trends of few-shot object detection based on the advantages and disadvantages of the existing methods. It is expected to provide references for subsequent research works in this field.

Key words deep learning; object detection; few-shot learning; few-shot object detection

1. 引言

得益于大规模数据集和硬件技术的发展，基于深度学习的模型在基础的计算机视觉任务中取得了令人瞩目的成就[1]。如：图像和视频分类[2][3]、目标检测[4]、语义分割[5]和图像生成[6]等。传统监督式的机器学习模型需要借助大规模带标注的数据进行训练，而现实世界中的视觉数据呈现显著的长尾效应，数据丰富的类别占据总类别的大多数，在某些特定应用场景下，一些稀缺的类别可能由于隐私、安全和高标记成本等因素使得数据的获取和标注十分困难，例如军事遥感检测[7]、疾病诊断[8]及工业生产中的残次品检测[9]等，这为计算机视觉领域的进一步发展带来了挑战。在有限的训练样本条件下，传统的深度学习算法无法得到充分的训练，使得深度神经网络模型易发生过拟合，导致模型的泛化能力受到严重影响，仅依靠当前深度学习技术难以满足样本较少的场景和需求[10]。

机器学习与人类智能之间存在显著差异，不同于深度神经网络，人类擅长从极少的样本示例中学习认识新事物，并做出准确的预测与评估，这种高效的数据利用能力正是当前机器学习模型在实际应用中所需要的。目前数据驱动下的深度学习模型，其通用性和泛化能力还远不能达到人类认知学习的水平，弥补这种差距是迈向更高机器感知能力的关键一步[11]。受人类学习模式的启发，为解决因训练样本数量较少而带来的模型过拟合问题，李飞飞等人[12]在2003年首次提出小样本学习的概念，认为计算机视觉模型的学习应该利用已获得的先验知识和少量的训练样本学习识

别新类别的模型。作为一种新的理论方法，当前小样本学习主要用于图像分类任务，并取得了突破性的进展[13][14]。随着研究的不断深入，小样本目标检测(Few-shot Object Detection, FSOD)逐渐引起学术界的关注，其核心思想是通过少量的标注样本的训练来对图像中的目标进行分类和定位[15]，通过设计合理的训练方法、模型结构和损失函数，获得具有一定泛化能力的检测模型，实现复杂环境下对小样本目标的有效检测，在数据获取和标注困难的场景下具有重要的价值和意义。相比于小样本图像分类任务，小样本目标检测更具挑战，其原因在于在识别目标类标签的基础上，还需进一步定位每个目标在图像中的位置[16]，因而对模型的数据利用能力提出了更高的要求。因此，在数据稀缺场景下，如何利用极少的标注样本进行学习，设计具有良好泛化能力的检测模型，并推广至新任务上，成为小样本目标检测亟待解决的问题。

随着小样本学习关注度的增加，小样本目标检测技术迅速发展，已成为热门研究方向。目前，已有四篇关于小样本目标检测的综述文献[17][18][19][20]。潘等人[17]主要对该领域发展初期的工作进行综述，将小样本目标检测方法分为三类，但是，该文献发表于2019年，未涉及之后的研究成果。目前，小样本目标检测技术的发展日新月异，2019年之后涌现大量新的检测方法，本文则全面、系统性对当前小样本目标检测技术进行梳理和总结。张等人[18]从小样本目标检测的问题定义、主要方法和实验设计等方面进行阐述。然而，该文献仅选取几种小样本检测模型在PASCAL VOC[21]数据集上进行对比分析，缺乏对其他数据集的详细论述与评估，本文则对数据集、实验设计及典型方法的检测效果等内容进行全面、完整地评估与分析。刘等人从数据、模型和算法三个角度阐述了小样本目标检测的解决方案与存在的难点。Leng等人[20]则根据数据稀缺程度将小样本目标检测分为有限监督、半监督和弱监督三种场景设置，基于这三种场景讨论面临的挑战和解决方法。但是，这两项工作均缺乏对较新的研究成果与数据集及实验的归纳总结。随着时代的发展，小样本目标检测技术突飞猛进，各种新理论、新思想、新方法的研究不断涌现，以上综述已不能满足该领域蓬勃发展的需求，使得初学者难以窥其全貌。

不同于现有研究综述，本文的主要贡献可总结如下：

（1）系统性梳理了小样本目标检测技术，涵盖了现有的小样本目标检测方法以及目前最新的研究成果。本文通过梳理小样本目标检测方法的发展脉络，从任务和问题、学习策略、检测方法及数据集和实验等角度出发，对当前的研究成果进行了全面、细致地归纳和梳理，系统性总结了小样本目标检测任务定义及三个核心问题，讨论了现有方法采用的两种学习策略，涵盖了较新的研究成果。

（2）分类角度独特，方法间的对比直观清晰。本文按照工作原理将当前检测方法分为四类，根据研究思路对每类方法进行更为精细的归类，分类角度更加合理，使读者能够快速了解每类检测方法的工作原理。同时，以表格的形式对四类方法采用的学习策略、优缺点及适用场景进行了总结，使读者能够根据不同的场景选择和使用不同的小样本目标检测方法。

（3）对数据集、实验设计及典型方法的检测效果等内容进行全面、完整地梳理与概括。本文详细论述了当前小样本目标检测常用的四个数据集、实验设计细节、性能评估指标及典型方法性能对比等内容，以图表的方式对不同方法在四个数据集上的实验结果进行了系统性评估，使读者能够快速掌握该领域当前的研究热点。本文旨在为该领域的研究者提供一个包含最新方法的综述，加深对小样本目标检测研究的理解，进一步促进小样本目标检测技术的发展。

本文的组织结构如图 1所示，第2节给出了小样本目标检测的任务定义及三个关键问题；第3节讨论了现有方法的两种学习策略；第4节从工作原理、问题及存在的不足等方面对现有的小样本目标检测方法进行分类阐述；第5节对数据集与实验等内容进行系统性地归纳总结，对四类典型方法在数据集上的实验结果进行了对比分析；第6节梳理了小样本目标检测面临的挑战，并指出了一些潜在的发展方向，供更多相关研究者参考和借鉴；最后第7节总结全文。

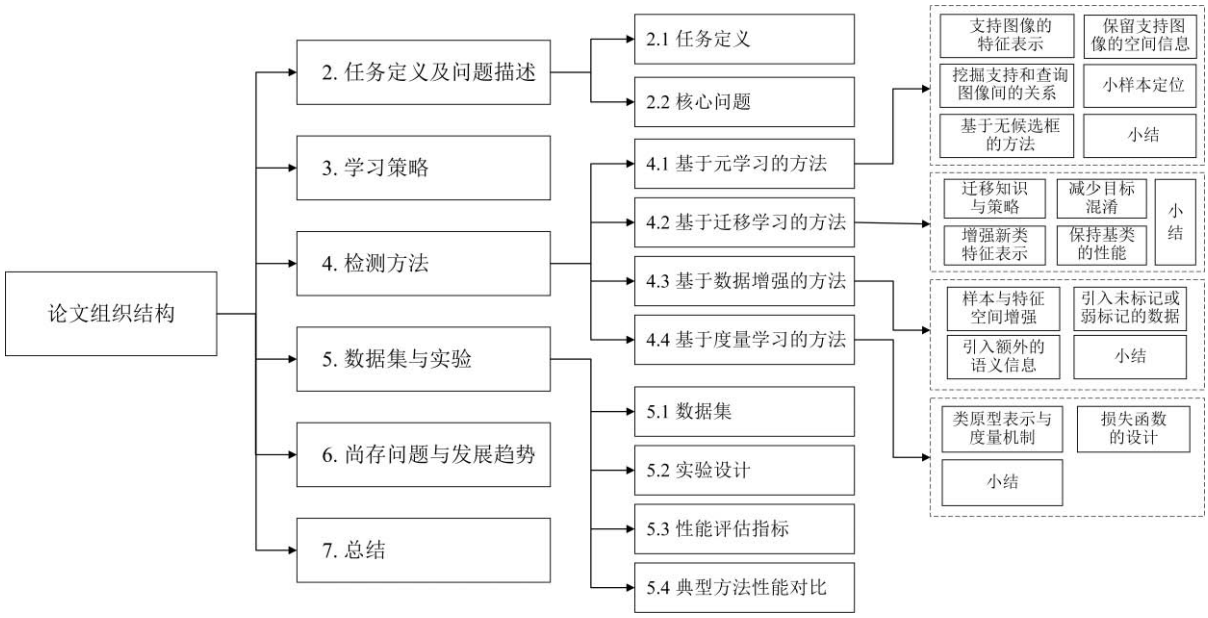

图 1 本文的组织结构图

Fig.1 The organizational structure of this paper

2. 任务定义及问题描述

2.1 任务定义

小样本目标检测任务旨在通过少量标注样本的训练来对图像中的目标进行分类和准确定位，以此得到具有良好泛化能力的检测模型。该任务可描述为：给定数据集D_{base}和D_{novel}，D_{base}表示基类数据集，每个类别有充足的标注训练样本，D_{novel}表示新类数据集，每类仅有少量标注样本（通常少于10个）。基类和新类中的类别不重叠，即$C_{base} \cap C_{novel} = \emptyset$。给定测试图像$x$，预测$x$中的$N$个目标的类别$\{cls_i\}_{i=1}^{N} \in C_{base} \cup C_{novel}$和边界框坐标$\{box_i\}_{i=1}^{N}$，小样本目标检测的目标是借助在丰富注释的基类中学习的先验知识和少量的新类训练样本实现对测试图像中目标的预测。

从概念上来讲，小样本目标检测是指在带有大量注释信息的基类数据集上训练得到基类检测模型，仅利用极少标注的新类数据集和基类模型提供的先验知识实现对新类的检测，如图 2所示。与开放世界的目标检测(Open World Object Detection, OWOD)和增量小样本目标检测(Incremental Few-Shot Object Detection, iFSOD)不同的是，前者针对新类样本数充足、类别数未知的场景[22]，后者在连续的少量数据流场景下实现对新类和基类的检测[23]，本文中的FSOD则面向极少标注数据场景，且新类类别数已知。尤其是，小样本目标检测任务更注重对新类别的检测性能，广义小样本目标检测任务则要求检测新类别的同时保持基类类别的性能。

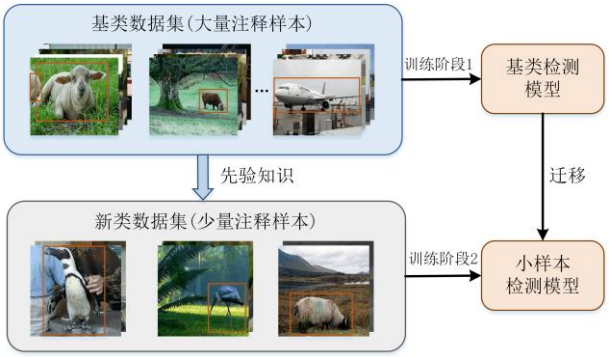

图 2 小样本目标检测示意图

Fig.2 Illustration of few-shot object detection

2.2 核心问题

小样本学习的根本问题在于因目标域样本稀缺，导致难以训练出鲁棒的小样本模型。小样本目标检测是指在给定少量训练样本的条件下，如何训练一个能够有效识别并准确定位这些目标的检测模型。现有的小样本目标检测模型均是基于深度学习的检测网络而设计的，然而在有限的数

据下对模型进行训练将会面临诸多挑战。为此，本文从模型和数据两个角度出发，梳理出三个核心问题。

过拟合。小样本目标检测的核心问题之一是过拟合。当新类数据与基类数据属于同域，且新类别仅有少量的训练样本可用，同时还需考虑目标的分类和定位任务时，在训练深度检测模型时极易造成模型过拟合，使训练良好的检测模型在新类数据集上性能较差，从而导致模型的泛化能力不足和鲁棒性差等问题。换言之，小样本数据集与模型复杂度间的高度不匹配导致了模型训练问题，因此，如何在小样本条件下进行模型训练，降低模型的学习难度，进一步增强模型的泛化性能成为当前小样本检测技术发展的难点之一。

域偏移。目前，小样本目标检测方法通常是借助大规模基类数据集来学习通用知识，同时将这些知识迁移至新任务的学习中。然而，当源域和目标域数据具有不同的数据分布时，可能出现域偏移问题。域偏移是指源域训练的模型在应用于具有不同统计量的目标域时表现不佳，属于异构迁移学习的范畴[24]。具体而言，当源域的基类与目标域的新类数据间存在较大的域差异，且二者共享的知识较少时，将基类训练的模型作为知识迁移至新类时很可能出现负迁移，从而导致模型对新任务的检测性能不佳，这就是通常所说的域偏移问题。因此，如何利用先验知识弥补样本数据量不足问题，是当前研究面临的巨大挑战之一。与此同时，构建小样本下的检测模型，需综合考虑合适的先验知识和迁移策略，因此，如何有效地将源域知识迁移并泛化至目标域有待进一步探索。

数据及分布偏差。数据集本质上是从数据分布中观察到的样本集合。然而，当训练样本数量不充足时，数据的多样性降低，导致数据偏差及分布偏差等问题。与大规模的数据集相比，有限的训练数据会放大数据集中的噪声，造成数据偏差，比如对于相同类别的图像存在较大的类内变化，不同类别的图像间的距离较小等等。而且，因目标域样本极其有限，无法准确地表征目标域的真实数据分布，导致目标域类别间及类别与背景间相互混淆，从而影响模型的检测精度。因此，如何提升训练数据的多样性，降低分类混淆，进而保证小样本检测模型的稳定性具有很大的研究空间。

3. 学习策略

针对小样本下的模型训练问题，当前的小样本目标检测方法通常采用两种学习策略，即：基于任务的episode训练策略[25]和基于数据驱动的训练策略[26]。前者以任务为基本单元，每个任务的数据集分为支持集和查询集，其目标是从大量训练任务中获取先验知识，从而能够通过少量数据在新任务中更快地学习。整个训练流程可分为元训练和元测试两个阶段：在元训练阶段，通过组合不同的训练集构建不同的元任务，使得模型学习独立于任务的泛化能力；在元测试阶段，模型不需要重新训练或仅需少量迭代次数即可学习新任务，最终实现“学会学习”。后者采用“预训练-微调”的训练范式[27]，直接针对数据集进行训练，在具有大量注释的基类数据集上进行预训练获得基类检测模型，在小样本数据集上进行微调泛化至新类。

基于任务的 episode 训练策略。训练数据表现为集合的形式，每个小样本任务被设定为 N-way K-shot 任务，包含支持集和查询集，从训练集中随机抽取N个类别，每个类别包含K个样本，构成单个元任务的支持集，同时随机选出少量图像作为查询集，目标是使模型从$N \times K$个样本中学会识别N个类别。图 3 展示了 2-way 3-shot 任务的训练范式。其中，不同任务间的类别不尽相同，每个任务称为一个 episode，在多个 episode 任务上进行元训练，每个 episode 通过给定的小样本支持集图像来预测查询集图像的类别，经过不同任务的元训练后，使得模型能够学习泛化能力强的初始参数，最终快速适应至新的元测试任务，从而掌握学会学习的能力。当执行元测试时，新任务的设置与元训练的任务设置相同，根据给定的少量支持集样本来预测查询集图像的类别，最终让模型对未知的任务具备良好的判别和泛化能力。

基于数据驱动的训练策略。训练集表现为批量数据的形式，直接作为模型的输入进行训练。图 4 展示了基于数据驱动的训练范式，通过在大规模带标注的数据集上进行预训练，在小的数据集上通过微调实现模型的可迁移性，使其泛化至

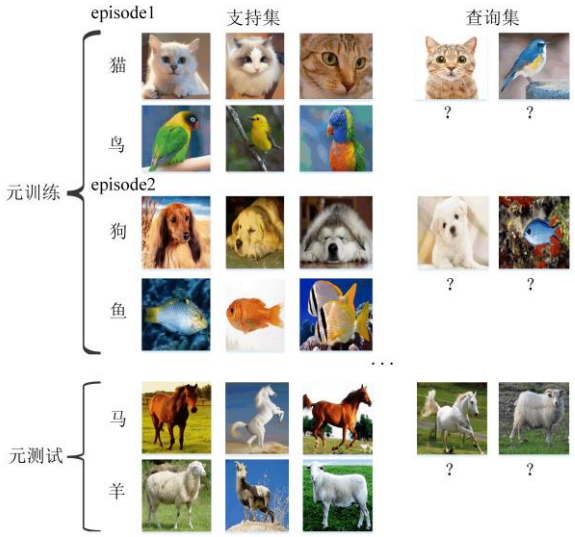

图 3 基于任务的episode训练范式

Fig.3 Task-based episode training paradigm

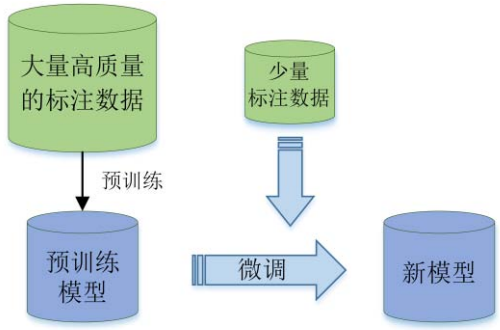

图 4 基于数据驱动的训练范式

Fig.4 Data-driven training paradigm

新任务。该过程不需要任务具备强关联性，一般而言，任务的关联性越强，微调至新任务的效果就越好。

当前的小样本目标检测方法均采用以上两种训练范式，两种训练范式各有优劣。基于任务的episode训练策略在极少样本中表现良好，不需要微调就能快速泛化至新任务。然而，该学习策略要求所有任务满足同分布[28]，其任务的设计可能限制了模型的学习能力，原因在于单个训练任务由多种数据构成，默认网络对所有类数据一视同仁，但是在小样本学习设定中，可能出现某类数据比其他数据更难被区分等情况。而且，当新任务的训练样本数较多或类别数较多时，模型处理速度变慢，导致性能下降。基于数据驱动的学习策略较为简单，不需要人为构建任务，但是基类学习的先验知识迁移什么，怎么迁移至新类是一个值得思考和深入研究的问题。

4. 检测方法

目前，针对小样本目标检测问题，国内外学者提出了一系列方法，主要是借助现有的成熟的检测框架和小样本学习方法，构建面向样本稀缺下的检测模型。早期研究阶段通常采用较少标记数据的半监督方法和不完全匹配标记数据的弱监督方法，核心是通过收集额外的易注释标签的训练示例来缓解目标检测中注释困难的问题[29]。因缺乏对训练图像充分的监督及复杂的模型设计，难以泛化至标记数据较少的新类上，从而导致新类的检测性能较差。

近年来，小样本目标检测研究取得了重大突破，从工作原理的角度出发，我们将小样本目标检测方法分为基于元学习的方法、基于迁移学习的方法、基于数据增强的方法以及基于度量学习的方法四类。表 1对这四类方法进行了简要地概括和对比。

4.1 基于元学习的方法

目前，针对小样本学习问题的元学习方法，被广泛应用于小样本检测中，其核心思想是通过模拟一系列相似的小样本任务，将先验知识从注释丰富的基类迁移至数据匮乏的新类之上，以应对样本数量不足的问题[50][51][52]。元学习方法以任务为单元进行训练，通过任务和数据的双重采样来设计不同的小样本任务，使其能够利用少量的支持集样本快速更新模型参数，最终在特定任务下仅需少量迭代即可快速泛化至新任务，不需要进一步微调。

基于元学习方法的思想，研究人员提出一系列卓有成效的小样本目标检测方法，且大多数方法采用两阶段检测模型Faster RCNN[53]来实现。该模型的工作原理如下：首先，通过候选区域生成网络(Region Proposal Network, RPN)生成感兴趣区域边界框(Region of Interests, RoIs)，并判断其是前景还是背景区域；然后，采用RoI池化操作将大小不同感兴趣区域边界框处理为相同的大小；最后，将获得的RoIs特征进行边界框的分类和回归。

表 1 四种小样本目标检测方法算法的对比分析

Table 1 Comparisons of the four methods for few-shot object detection algorithms

方法	代表方法	学习策略	优势	劣势	适用场景
基于元学习的方法	FSRW[30] Meta-RCNN[31] Meta FR-CNN[32] SQMG[33] DRL[34] MetaDet[35] Meta-DETR[36]	基于任务的episode训练策略	快速适应新任务； 仅需少量迭代即可更新模型参数；	模型设计困难，不易收敛； 类别或样本数量较多时效率低下； 要求相似任务；	样本量极少
基于迁移学习的方法	LSTD[37] TFA[38] FSCE[39] FSOD-SR[40] SRR-FSD[41]	基于数据驱动的训练策略	检测精度相对较高； 学习策略简单；	易出现过拟合； 不适用实时检测场景；	对速度、精度要求较高
基于数据增强的方法	TIP[42] HallucFsDet[43] MPSR[44] FADI[45]	基于数据驱动的训练策略	实现容易； 可叠加不同的增强方法提升性能；	性能提升有限； 计算量较大	样本量极少
基于度量学习的方法	RepMet[46] NP-RepMet[47] PNPDet[48] CME[49]	基于任务的episode训练策略	简单易操作； 易添加新类别； 易实现增量学习；	计算量较大； 占用内存较高； 定位精度相对较差；	定位精度要求低、样本量极少

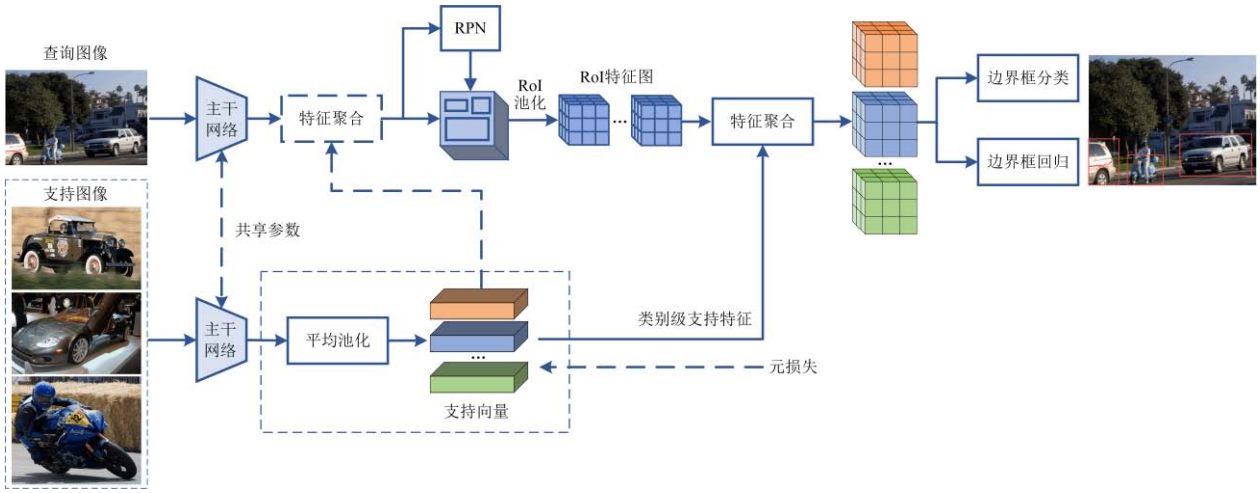

图 5 基于元学习的两阶段小样本目标检测框架

Fig.5 Meta-learning based framework for two-stage few-shot object detection

为描述基于元学习的小样本目标检测方法的基本工作原理，本文将Faster RCNN作为基础检测模型，以3-way 1-shot任务为例，构建了基于元学习的小样本目标检测框架。如图 5所示，该框架通常采用并行结构，整个流程包括元训练和元测试两个阶段，其工作原理如下：在元训练过程中，包括支持和查询两个分支，首先利用孪生主干网络同时从支持图像和查询图像中提取特征，支持图像经平均池化操作得到支持向量作为类别支持特征，同时查询图像特征经RPN和RoI处理得到RoI特征图；然后，通过特征聚合机制将支持集图像每类的特征与查询图像特征融合，以此来检测查询图像的目标。在上述过程中，边界框分类任务通常采用交叉熵损失，定位任务采用Smooth L1损失；为避免相似的支持特征造成歧义，部分研究工作[31][54]增加了额外的元损失以鼓励支持图像特征属于对应目标的类别。而且，为了提升查询图像边界框的生成质量，一些工作[33][55][56][57]在RPN之前进行了特征融合操作。在元测试过程中，利用少量的新类图像对所有类别计算支持特征，故不需要支持分支。

为详细阐述当前基于元学习的小样本目标检测方法的原理、性能及存在的不足，本文从解决思路入手，从支持图像的特征表示、保留支持图像的空间信息、挖掘支持和查询图像间的关系及小样本定位等四个方面，对当前工作进行概括总结。同时，本文还梳理了近期基于无候选框的方法。

4.1.1 支持图像的特征表示

早期的小样本目标检测方法在元训练阶段引入元学习者模型，获取任务级的元知识，在不同任务间实现元知识的迁移和共享，从而提升模型在新任务上的快速泛化能力。元知识通常被定义为特征表示或模型的权重参数，元学习者则为学习元知识的模型。因训练样本极其有限，支持集图像的特征表示和学习成为衡量查询集图像预测结果的一个关键因素，其特征表示作为支持向量来编码特定类别信息，以指导模型能够在查询图像上检测属于支持集中类别的目标。基于此，Kang等人[30]提出了一种特征重加权的小样本目标检测方法(Few-shot object detection via feature re-weighting, FSRW)，该方法基于YOLO v2检测模型，设计了特征学习、特征重加权和预测三种机制。其中，通过大规模注释的基类数据集训练特征提取器，特征学习机制学习可泛化的元特征；特征重加权机制学习支持集中每个目标类别的全局特征，将其与查询图像的元特征进行通道级融合，得到加权后的特征图，以调整查询图像的元特征；预测机制将调整后的元特征送入检测模块，以预测查询图像的类别和边界框。FSRW方法采用整个查询图像的特征图进行预测，考虑到查询图像中可能含有多个目标和背景，在整个图像上进行元学习并不是最优方案，为此，Yan等人[31]设计了Meta-RCNN模型，引入预测模型网络(Predictor-head Remodeling Network, PRN)对支持集图像中的所有类目标推断类注意向量，并将其作为元知识与查询图像经RPN提取的感兴趣区域特征图进行通道级融合，最终得到对应的检测图，与FSRW方法的性能相比提升了5.3%。同样地，Xiao等人[54]对支持集中图像进行实例裁剪，计算样本均值作为支持实例的类注意向量，将类注意向量与查询图像生成的RoI特征图进行聚合，通过联合特征嵌入机制聚合支持特征和查询特征，以此来衡量特征间的相似性，该方法不仅提升了小样本检测模型的性能，还降低了因支持集的随机采样产生的影响。上述方法均将小样本目标检测视为小样本分类问题来解决，通过设计元学习者模型，将少量带注释的支持集图像作为该模型的输入，提取类别级的特征表示作为元知识来调整查询图像的特征表示，从而对查询图像中的目标实现更好的预测。

4.1.2 保留支持图像的空间信息

尽管早期的小样本目标检测方法能够有效提升小样本新类的检测性能，但支持集图像表示均采用全局池化操作映射成一维向量，与查询特征图聚合时，可能导致支持集图像的空间信息和局部上下文信息严重缺失，从而造成误分类和漏检。而且，由于支持集图像中背景噪声的存在，使得生成的区域边界框并不能保证与真实目标完全对齐，可能只包含部分目标或较大的背景区域，从而导致空间错位问题。尤其是，部分研究工作[30][31][54]直接将支持集中同类目标的平均特征作为特定类的表示，由于存在多个语义信息导致信息纠缠问题。总之，支持图像可能存在背景噪声，其全局表示忽略了空间信息，严重影响了小样本目标检测模型的性能。

针对上述提到的误分类和漏检问题，Hu等人[55]从上下文信息的角度出发，提出了一种基于上下文感知聚合的稠密关系蒸馏方法(Dense Relation Distillation with Context-aware Aggregation, DCNet)，包含稠密关系蒸馏(Dense Relation Distillation, DRD)与上下文感知聚合(Context-aware Feature Aggregation, CFA)两个机制。当图像出现外观变化或遮挡时，局部细节特征在匹配查询图像候选目标和支持图像时占据主要地位。因此，利用DRD机制提取查询图像的细粒度特征，同时设计浓缩蒸馏机制，实现了查询图像和支持图像的像素级匹配。CFA机制自适应地利用来自不同尺度特征的语义信息，获得更全面的尺度感知特征，从而更好地捕捉全局和局部特征，以缓解小样本下的尺度变化问题，并在小样本目标检测数据集上取得了优异的性能。针对支持集中存在的背景噪声导致的空间错位问题，Han等人[32]提出了一种将空间特征对齐机制(Spatial Alignment Module,

SAM)和前景注意机制(Foreground Attention Module, FAM)相结合的细粒度原型匹配网络Meta-Classifier，利用前景注意掩码强调图像中的目标区域，通过度量查询图像特征图的每个空间位置与支持类类原型间的相似性，联合优化小样本边界框生成和分类，最终获得具有竞争力的检测性能。然而，该方法仅适合新类数量小于20的情况，如果新类数量过多，则需要大量的计算资源，导致模型推理速度缓慢，效率低下。针对空间错位及信息纠缠问题，Chen等人[56]在充分考虑支持和查询图像间成对空间关系的情况下，提出了一种双重感知注意力(Dual-Awareness Attention, DAnA)方法，设计了背景衰减的注意力(Background Attenuation, BA)和跨图像空间注意的注意力(Cross-image Spatial Attention, CISA)两个机制。一方面，利用BA机制学习支持图像中目标的语义表示，增强支持特征信息，同时抑制复杂背景和不相关的前景目标；另一方面，CISA机制通过卷积级的注意力捕获支持和查询特征间的成对空间关系，从而有效地打破了空间对齐的物理限制。最后，通过实验验证了减少支持集中的背景噪声和空间变化能够有效提升小样本目标检测的性能。可以看出，上述方法均强调了图像中的目标区域，通过抑制背景噪声来增强支持图像中目标的语义信息，以促进查询图像中目标的预测。

4.1.3 挖掘支持和查询图像间的关系

在基于元学习的小样本目标检测方法中，将查询图像的感兴趣区域与支持图像中的目标特征表示相结合，实现分类和回归的预测。因此，支持集图像和查询图像间的相关性对小样本检测模型的性能有很大的影响。然而，先前工作[32][55][56]仅利用单一类别的支持集图像作为注意力来指导查询图像的预测，并未考虑支持图像和查询图像的相关性。针对这一问题，近期工作从增强特征表示和生成更多与查询图像中类别相关的候选框两个角度出发，提出了新的研究思路。

增强特征表示。因新类训练样本有限，检测模型很难学习到泛化的特征表示，从而导致错误分类。因此，研究人员利用支持集图像和查询集图像特征间的关系，通过增强小样本新类的特征表示，使得所学特征更具区分性。Liu等人[34]探索了所有的支持图像和查询图像RoI特征间的相关性，提出了一种动态相关学习(Dynamic Relevance Learning, DRL)模型，利用所有支持图像和查询图像上RoI特征间的依赖性关系，构造了一个动态图卷积网络(Graph Convolutional Network, GCN)，采用该网络的输出调整基类检测模型的预测分布，指导检测模型隐式地改进类表示，从而使得模型更具区分性。该工作首次将图神经网络引入至小样本目标检测中，实现了新类特征表示的增强。类似地，Han等人[57]提出查询自适应的小样本目标检测方法(Query Adaptive Few-Shot Object Detection, QA-FewDet)，引入一种异质图捕获支持类原型和查询边界框特征间的关系。该异质图模型分为三种类型的边和子图。其中，类与类间的边模拟基类和新类间关系，利用其它相似类的原型来增强新类原型；边界框与类间的边使得边界框特征与类原型相互适应，获得了查询自适应的多类别增强原型表示，从而减少两个特征分布的差异性；边界框与边界框间的边则提供局部与全局上下文信息，获得了上下文感知的边界框特征，从而促进成对匹配，最终实现边界框的分类和定位。该方法学习了支持类与查询目标间的所有关系，在不同的评估指标下均获得了更高的检测性能。同时，Zhang等人[32]提出了一种支持查询相互引导的方法(Support-Query Mutual Guidance, SQMG)，包括支持引导的查询增强(Support-guided Query Enhancement, SQE)和查询引导的支持权重(Query-guided Support Weighting, QSW)两个机制。SQE机制通过动态核卷积增强查询图像的特征，QSW机制则利用可学习的权重操作聚合不同的支持图像生成支持特征，同时增强了查询图像和支持集的特征表示，并在PASCAL VOC数据集的第二种分割下取得了先进的性能。

候选框生成。因候选区域生成网络RPN提取所有目标区域的边界框，若不借助支持集图像的相关信息，则导致RPN漫无目的地在查询集图像中寻找存在目标概率大的区域，从而对后续的边界框分类造成困难。因此，一些工作针对如何提高生成边界框的质量展开研究，旨在生成更多与查询图像相关的边界框，过滤掉大多数背景框和不匹配类别的边界框，从而提升边界框分类的效率。Fan等人[58]在边界框生成和检测头部分利用了支持集和查询集间的匹配关系，提出了一种基于注意力的边界框生成网络(Attention-Based Region

Proposal Network, Attention RPN)，以支持图像的特征为卷积核，在查询图像上滑动，使其利用支持图像信息指导网络生成更多与查询图像相关的候选框，抑制大多数背景框和其它无关类的边界框，进而生成更高质量的新类候选框，以此缓解RPN检测新类时可能错过新类目标或在背景上产生大量误检等问题。该方法能够在线检测新类别，无需再训练和微调。同样地，Zhang等人[32]提出一种支持引导的边界框生成机制(Support-guided Proposal Generation, SPG)，利用支持图像中目标的信息，通过动态卷积的方法增强查询图像的RoI特征图，以此生成更多与查询图像类别相关的边界框，过滤掉大多数背景框和不匹配的类别，从而生成更精确的新类目标边界框。

不难看出，支持集图像的信息对查询图像的特征表示和边界框生成具有一定的指导作用，如何充分利用支持集信息是提升小样本检测模型性能的关键因素之一。

4.1.4 小样本定位

上述小样本检测方法均集中在分类任务，仅仅将小样本分类方法引入小样本检测模型中，并未考虑小样本定位任务，导致检测性能不佳。为此，Wang等人[35]以一种统一的方式同时考虑小样本分类和定位问题，从解耦参数学习的角度出发，构建了一个基于元学习的MetaDet框架，分离类别不可知部分和特定类别部分的学习，通过引入一种参数化的权重预测元模型(Weight Prediction Meta-model, WPM)来学习元知识，使其从少量样本中预测特定类别参数，在域内、跨域和长尾等三种少量样本的场景中，验证了该方法对新类的检测性能有显著提升。2021年，Zhang等人[36]摒弃了区域边界框的方法，充分利用分类和回归任务的互补关系，将近年来流行的Transformer模型[59]与元学习结合，构建了一个Meta-DETR框架，将支持集和查询集图像经编码器生成特定类的特征，采用类无关的解码器对特定的类别进行预测，最终在图像级层面上实现了目标的分类和定位。该方法无需RPN操作，通过大量的实验证实了Transformer模型在小样本检测任务上的有效性。

4.1.5 基于无候选框的方法

目前，虽然大多数小样本检测方法基于两阶段检测模型 Faster RCNN 实现，但是，这些方法需要进一步处理不准确的区域边界框，而无锚框的检测模型则无需考虑该问题，且已有部分研究工作将无锚框的模型作为基础检测模型，应用至小样本检测任务中[30][56]。例如：早期的 FSRW 方法采用 YOLO v2 模型[60]，DAnA 方法则采用 RetinaNet[61]和 Faster RCNN 两种模型。

近年来，随着 Transformer 模型在计算机视觉任务中的成功应用，研究人员逐渐将其应用至小样本目标检测中。Chen 等人[62]提出一种自适应的图像 Transformer 模型(Adaptive Image Transformer, AIT)，将其用于单样本目标检测任务中，该模型通过模拟语言翻译的过程，设计了一种多头注意机制(Multi-head Co-Attention, MCA)来关联支持和查询图像的特征，利用 AIT 提取每个支持查询对的共同语义信息，自适应地翻译每个目标边界框特征，以更好地关联给定的查询特征，从而区分支持查询对之间的类相似性；最后，提出一种选择性的通道注意力机制(Selective Channel Attention, SCA)对多头信息进行有效融合，提升具有较高相似性通道的重要性。进一步，Zhang等人[36]摒弃传统的两阶段检测模型，基于 Deformable DETR 模型设计了 Meta-DETR，在主干网络之后直接聚合支持特征和查询特征，实现了图像级的目标检测。尤其是，Meta-DETR 通过其设计的相关聚合模块(Correlational Aggregation Module, CAM)，进行查询特征与支持集之间的特征和编码匹配。Meta-DETR 不仅能够在一个前馈中同时处理多个支持类别，而且能够有效地捕捉不同类别之间的相关性，从而大大地降低相似类间的误分类，增强类别间的泛化能力。从上述两项研究可以看出，将 Transformer 模型应用于小样本目标检测任务中，不仅能够消除不准确的新类边界框带来的限制，而且还能充分利用视觉和语言信息融合的优势，有效缓解样本缺失导致的局限性。

4.1.6 小结

概括起来，基于元学习的小样本目标检测方法，能够有效利用少量的支持样本实现新类的检测，具备快速适应新任务的能力，但仍然存在以下缺陷和不足：（1）模型只能在固定任务上进行预训练和迁移；（2）基于任务的训练需要对每个

类别执行一次前馈计算，时间复杂度较高，消耗大量的计算资源，导致效率低下，一旦类别数量过多，模型的速度变得极其缓慢；（3）基于任务的训练方式将单个任务中的样本看作一个训练批次，类别数固定，需人为构造任务中的支持集；（4）虽然元知识可表示为参数、网络结构、优化器、注意力、梯度等等。然而小样本目标检测中的元知识难以被准确表征，使得元学习者模型的设计较为困难，且在学习迭代过程中易出现不收敛问题[63]。

尽管基于元学习的小样本目标检测方法存在上述缺陷和问题，但仍然是一种有效的检测方法。未来需根据任务需求合理设计采样方式，对检测效率问题深入研究和探索，进而推动小样本目标检测技术的发展。

4.2 基于迁移学习的方法

与弱监督或半弱监督方法相比，迁移学习方法不需要额外的数据收集，训练策略简单，可作为一种有效的小样本目标检测方法[64]。微调(Fine-tuning, FT)方法作为迁移学习的一种方法，其核心思想是：首先，在大规模基类标注的数据集上预训练源域模型；然后，基于少量的目标域训练样本对模型的参数进行微调。该方法在解决小样本目标检测问题上展现出巨大的潜力[65]。

为阐述基于迁移学习的小样本检测方法的工作原理，本文以两阶段检测模型为基础，构建了基于迁移学习的小样本目标检测框架。如图 6所示，该框架分为基类训练和小样本微调两个阶段：图 6(a) 展示了基类训练阶段，该阶段在大量的基类标注数据集上进行训练得到基类模型；图 6(b) 展示了小样本微调阶段，该阶段将基类模型的部分权重作为小样本微调模型的初始化，同时将基类和新类的小样本平衡集作为小样本微调模型的输入，以统一的方式实现边界框分类和回归。整个模型采用交叉熵和Smooth L1损失训练，且不同的迁移策略意味着固定和微调的参数不同，图 6(b) 中展示了TFA经典算法的迁移策略，固定主干网络、RPN和RoI三个部分的参数，仅微调最后一层边界框分类和回归部分的参数。

近年来，基于迁移学习的小样本目标检测方法因训练方式简单，获得了研究人员的广泛关注。相比于元学习方法，基于迁移学习的小样本目标检测方法不需要设计训练任务，通过微调的方式将基类训练的检测模型迁移至新类。该方法不需要任务间存在很强的关联性，且更强调在迁移的新任务上的性能，但依然存在诸多挑战与难点。因此，本文从迁移知识与策略、减少目标混淆、增强新类特征表示及保持基类的性能等四个方面入手，详细地阐述当前方法的工作原理。

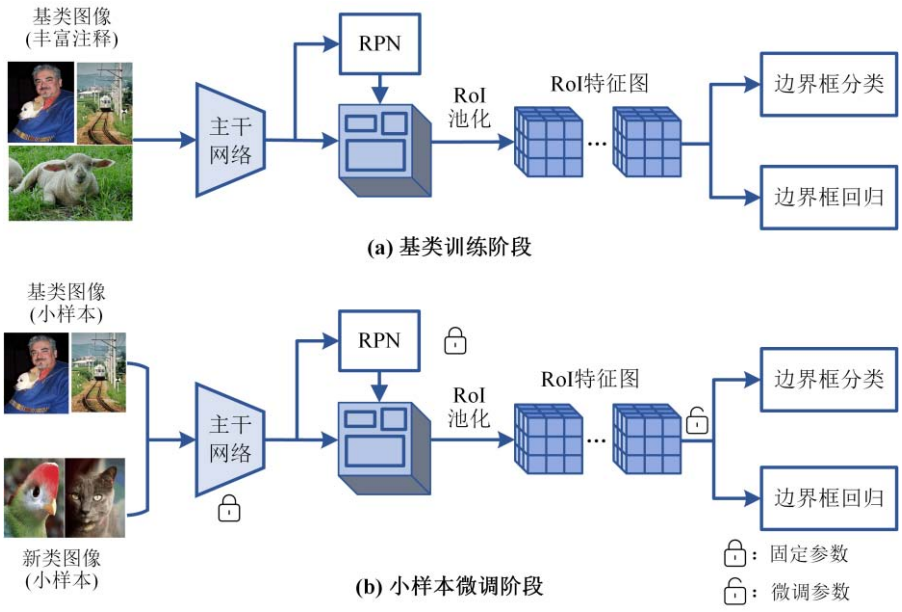

图 6 基于迁移学习的两阶段小样本检测框架

Fig.6 Transfer learning based framework for two-stage few-shot object detection

4.2.1 迁移知识与策略

与基于元学习的方法不同，基于迁移学习的小样本目标检测工作更关注迁移的知识及迁移策略，希望源域迁移的知识能够泛化至少量样本的目标域。基于此，Chen等人[66]重点研究和分析了现有元学习方法和微调方法的性能，在设定相同的实验场景和实验条件下，通过实验证实，在小样本分类任务中，基于微调的方法比基于元学习的方法更具优势，因而进一步引起学术界对微调方法的关注。然而，当新类训练样本数量不足时，深度检测模型比分类模型更容易出现过拟合问题，直接进行微调可能存在域偏移，从而降低可迁移性[67]。因此，针对迁移过程中存在的域偏移问题，Chen等人[37]结合单阶段检测模型SSD[68]和两阶段检测模型Faster RCNN的优势，提出了一种小样本迁移检测器(Low-shot Transfer Detector, LSTD)。首先，在包含大量标注数据的源域数据集上训练学习源域模型，用该模型参数初始化目标域模型；然后，基于少量的目标域数据进行微调。同时，LSTD设计了背景抑制正则化(Background-Depression, BD)和知识迁移正则化(Transfer-Knowledge, TK)两个机制，使其在迁移学习过程中能够聚焦于前景目标，降低语义混淆对模型精度的影响，更好地利用源域知识来增强对少量目标图像的微调，从而学习到更多与目标类别相关的知识。同样地，Wang等人[38]指出，因评估准则的不可靠性，基于微调的小样本目标检测方法的性能，超越了很多现有的基于元学习的方法。在此基础上，他们提出了一种两阶段微调方法(Two-stage Fine-tuning Approach, TFA)。TFA在进行新类微调时，冻结模型的主干网络、RPN和RoI特征提取三个部分，在基类和新类的平衡小样本集合上，微调检测模型最后一层的边界框分类器和回归器，有效提升了模型的迁移能力。同时，为减少类内方差，将实例级边界框特征进行归一化处理，将其输入到基于余弦相似性的边界框分类器当中进行分类；通过对比实验发现，简单的微调方法比基于元学习的方法的性能提高了2%~20%，在单样本下的精度甚至提升一倍。然而，由于在简单的微调过程中，有可能忽略了来自源域和目标域的重要知识[37]，因此简单的微调可能导致可迁移性的降低。Khandelwal等人[69]提出了一种统一的半监督学习框架(Unified semi-supervised framework, UniT)，在基类训练阶段，该框架充分利用图像级注释数据来训练弱检测模型；在小样本微调阶段，将新类的分类器和回归器的权重表示为基类对应项的加权线性组合，并利用基类和新类间语言及视觉的多模态相似性度量确定组合权重，实现有效的知识迁移和适应。

概括起来，上述方法大多从基类训练的模型中获取先验知识，作为新类模型的初始化参数或设计更好的参数初始化算法，其优势是不需要引入额外的参数，但劣势是通用的检测框架可能不适合小样本场景，从而影响模型的检测性能。

4.2.2 减少目标混淆

在小样本目标检测过程中，因缺乏足够的样本来获得可区分的特征，导致检测模型对高度相似的类别存在目标混淆问题。自2021年以来，涌现出一些新的研究工作，这些工作从难样本挖掘、类间差异、上下文信息和分类校正等四个不同的角度，对目标混淆问题展开研究[39][70][71][72]。Li等人[70]提出了一种解耦的小样本分类网络(Few-Shot Classification Network, FSCN)，设计了一种小样本分类细化机制，其核心是分割出目标边界框区域，增加额外的类别区分信息，进而提升检测模型的语义可区分性，以此来缓解类别混淆问题。同时，针对因注释稀缺导致的干扰样本问题，将干扰样本建模为半监督问题，通过设计干扰利用损失(Distractor Utilization Loss, DUL)提升数据稀缺类别的利用率。该方法能够在不增加注释成本的前提下促进对小样本新类的学习能力。Sun等人[39]从类间差异的新角度考虑，首次将对比学习应用至小样本目标检测中，提出了一种对比边界框编码(Contrastive Proposal Encoding, CPE)损失，利用对比分支来增强目标RoI特征，减少来自同一类别的目标候选嵌入差异，同时分离不同类别的实例，进而提升实例级的类内紧凑性和类间差异，达到减少目标混淆的目的，该方法在标准的PASCAL VOC数据集基准上的性能提升了8.8%，在COCO[73]数据集基准上的性能提升了2.7%。与上述工作不同，Yang等人[71]发现将基类训练的检测模型迁移至新类时，虽然在定位任务上表现良好，但在分类任务上性能不佳。为此，提出了一种深度迁移框架

Context-Transformer，包括相似性发现(Affinity Discovery, AD)和上下文聚合(Context Aggregation, CA)两部分，前者从周围环境中寻找与待识别目标相关联的目标作为线索，计算相关性分数；后者则将相关性分数作为权重向量，将其叠加在待识别目标的特征之上，达到增强特征的目的，从而进一步提升了检测模型的分辨能力。该方法通过基类和新类的上下文关联关系，有效避免了小样本场景中的目标混淆问题。Qiao等人[72]认为Faster RCNN模型本身不适合数据稀缺场景，因而从多阶段(RPN vs RCNN)和多任务(分类 vs 定位)两个正交的角度出发，分析了该模型的潜在矛盾，提出了Decoupled Faster R-CNN框架。该框架通过设计梯度解耦层(Gradient Decopled Layer, GDL)解耦分类和定位任务，提出一种原型校准模块(Prototypical Calibration Block, PCB)，通过额外的成对评分对原始分类分数进行校正，以提升检测模型的分类能力，在多个基准上的广泛实验表明，Decoupled Faster R-CNN模型的检测性能明显优于其他现有方法。

4.2.3 增强特征表示

因新类样本有限，当基类训练的模型迁移至新类时，新类特征可能与相似的基类特征产生混淆现象。为更好地区分混淆类别，在有限的新类训练样本条件下，增强新类特征表示成为应对小样本目标检测问题的一种有效方法。目前，研究人员从注意力图、上下文信息及基类和新类目标间关系等三个方面来增强特征表示。由于小样本下的检测模型很难找到检测目标的所在正确区域，Chen等人[74]采用视觉显著性注意图模拟人类观察图像中的感兴趣位置，提出了一种基于注意力的小样本检测网络(Attentive Few-shot object Detection Network, AttFDNet)，通过全局上下文(Global Context, GC)机制捕获场景中全局表示的远程依赖关系，为空间位置分配不同的权重，使得检测模型能够充分利用重要位置的特征进行目标检测。Kim等人[40]充分利用上下文信息，提出了一种空间推理(Spatial Reasoning, SR)框架，该框架采用图卷积网络(Graph Convolutional Network, GCN)编码基类和新类RoI特征间的关系，以RoI特征为节点，通过传播每个节点的信息，增强感兴趣区域的辨别能力，并利用目标间的关系来增强新类的特征表示。Zhu等人[41]受Transformer的启发，通过学习一个由图像数据驱动的动态关系图，建模基类和新类间的语义关系，并利用所学习的图执行关系推理，减小视觉和语言域间的差异。其中，图采用自注意结构[75]实现，动态地从词向量中生成，无需重新定义新图进行训练，这种图结构提供了更灵活的图生成范式，能够轻松地适应新类别，从而更好地捕获基类和新类的相互关系。该方法在1-shot或2-shot下表现良好，证实了通过建模目标间的关系，能够有效提升新类特征表示，进而提升对新类目标的泛化能力。

4.2.4 保持基类的性能

尽管上述基于迁移学习的方法在小样本新类上取得了较好的检测性能，然而，Wang等人[38]发现现有的小样本检测方法仅关注新类的表现，忽略了基类的性能下降问题，从而影响了网络的整体检测性能。因此，研究人员认为小样本目标检测方法应关注模型在所有类别上的检测性能，并提出了广义小样本目标检测和增量小样本目标检测两种任务。前者不仅需要保证模型能够检测新类，而且还要关注模型对基类的检测性能；后者属于持续学习的范畴，旨在连续的新类数据流中随时检测新类，且不忘记基类。

广义小样本目标检测任务。广义小样本目标检测任务要求小样本检测模型检测新类别的同时保持对基类的检测性能。Fan等人[76]重点研究了广义小样本目标检测任务，并认为现有的检测模型在基类训练阶段因缺乏新类的注释信息，将新类目标边界框当作背景类来学习，导致基类训练的RPN可能偏向基类样本，从而进一步加剧了新类数据稀缺问题。针对该问题，Fan等人[76]提出了一种小样本检测模型Retentive R-CNN。其中，偏置平衡的边界框生成器(Bias-Balanced Region Proposal Network, Bias-Balanced RPN)由基类RPN和微调的RPN构成，前者作为预训练的基类边界框生成器，后者则微调RPN的最后一层，为新类提供充足的边界框；重检测器(Re-detector)由基于全连接层的基类检测头和基于余弦相似性分类器的所有类检测头构成，目的是平衡特征的多样性。同时，通过辅助一致性损失(Auxiliary Consistency Loss, ACL)来调整适应过程，最终使得检

测模型在不降低基类检测性能的前提下，又能有效提升新类的检测性能，与TFA方法[38]相比，基类检测的性能有明显的提升。

增量小样本目标检测任务。增量小样本目标检测任务旨在面向连续的小样本数据流的场景，要求检测模型不断学习新任务，在该场景下实现对所有类别的检测。其中，新类训练数据满足小样本设定，新类图像中可以包含来自属于旧类和新类的目标，仅提供新类目标的注释，旧类目标不进行注释。而且，每次增量过程中，仅有少量的新类数据可用。现有的小样本目标检测方法在扩展至新类时，需要对旧类样本进行随机采样，存储和重放一些旧样本。然而，这样可能导致模型需重新训练，阻碍了其在现实场景中的应用。为此，Li等人[77]提出了一种LEAST方法(Less forgetting, fEwer training resources, And Stronger Transfer capability, LEAST)，通过一种基于聚类的样本选择算法，保留以前学习的更多有区别的特征，期望利用一些样本代表性地捕捉基类的分布和类内方差。通过特定的迁移策略，将类别敏感的特征提取模型从整个检测模型中分离，获得更强的迁移能力，减少不必要的权重调整；同时，采用知识蒸馏技术[78]缓解灾难性遗忘问题。该方法不仅能够从少量带注释的新类样本中增量学习新类，而且不会对先前学习的类别产生灾难性遗忘。为进一步提升基类的检测性能，Li等人[79]基于TFA方法提出了一种增量小样本目标检测方法(Incremental Few-Shot Object Detection, iFSOD)。一方面，该方法通过一种双分支框架(Double-Branch Framework, DBF)解耦基类和新类的特征表示，有利于旧知识的保持和新知识的适应；另一方面，通过一种稳定匹配规则(Stable Moment Matching, SMM)，将参数搜索空间限制在一个围绕旧类检测的最佳局部区域，以防止旧类的遗忘。最后，在保留旧类的RoI特征基础上，通过基于间隔的正则化损失来校正对旧类的误分类，在实现检测新类的同时不忘记基类。更进一步，Chen等人[80]针对持续增量小样本目标检测问题，也就是当大量的新任务连续出现，每个新任务仅有少量注释的训练样本时，提出了一种解耦记忆蒸馏方法(Disentangling Imprinting Distilling, DID)，该方法在学习新任务时，采用知识蒸馏机制对前一个检测任务的知识进行蒸馏，以减少先前知识的遗忘，使得最终的模型不仅能够检测新类，而且还能同时保持基类的检测性能。可以看出，目前小样本目标检测问题已扩展至更具挑战的增量小样本目标检测问题，设计的模型主要解决基类的灾难性遗忘问题，但其检测性能还有待提升。因此，该方向可作为未来的一个重点研究方向，将其应用至开放世界的检测任务中，对该任务中相关技术的发展具有重要的现实意义和价值。

4.2.5 小结

综上所述，基于迁移学习的小样本目标检测方法主要是通过微调的方式实现对新类的检测。在具有一定数量样本条件下，获得了较好的检测性能，适合对速度要求较高的检测场景。然而，就目前的研究现状而言，还存在以下不足之处：（1）当样本量极少时，难以准确表征整个类别的特征分布，使得模型存在严重的过拟合问题，导致其泛化能力较差；（2）因微调时需重新训练，不能适应实时检测的场景；（3）如何相对准确地区分类别相关和类别无关参数，以及如何选择合适的超参数，尚未有相应的理论研究和实验支撑，还有待进一步探究。

4.3 基于数据增强的方法

小样本数据集因样本数量有限，不足以让模型捕捉到具有丰富变化的数据分布[81]。针对训练样本不足的问题，通过数据增强来提升数据及分布的多样性是最朴素直接的做法。对于目标检测任务而言，可以采用两种数据增强方式：一是不改变标注框的增强方式，如色彩变换、高斯噪声以及弹性变换等等；二是改变标注框的增强方式，如裁剪变换、旋转变换以及镜像变换等[82]。

基于数据增强的小样本目标检测方法，旨在通过生成更多新类样本、增强新类特征表示或为模型提供其他信息等方式，克服训练样本短缺的局限性。目前，一些工作从如何生成更多示例的角度展开研究，而另一些工作则从未标记/弱标记的图像或语义信息中，如何为模型提供额外的先验知识的角度出发展开研究，提出了一系列创新性方法，并取得了令人可喜的研究成果。基于Faster RCNN模型，图 7展示了基于数据增强的小

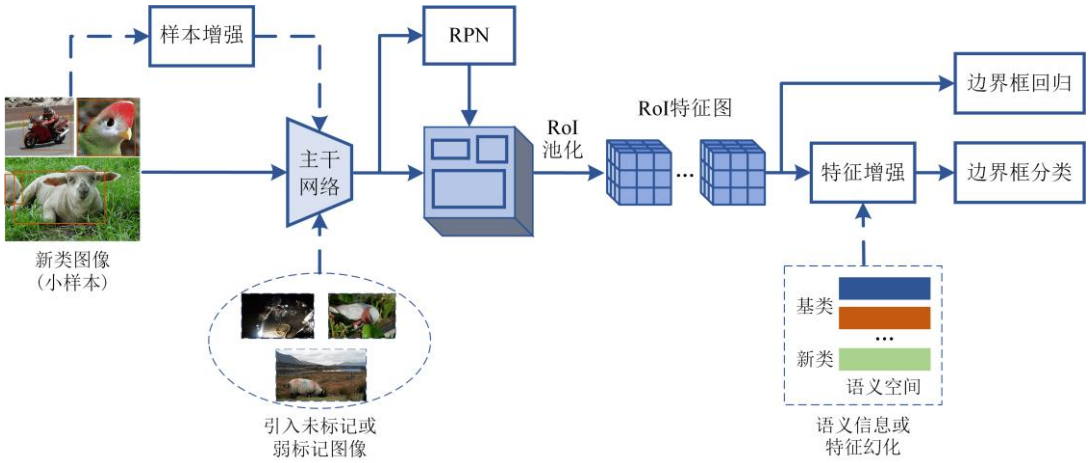

图 7 基于数据增强的小样本检测方法

Fig.7 Few-shot object detection methods based data augmentation

样本目标检测方法。可以看出，目前的方法主要从生成更多的图像级或特征级示例、引入未标记/弱标记的图像或引入语义信息为模型提供额外的先验知识等三个方面展开研究。因此，我们将当前方法分为三类：样本与特征空间增强、引入未标记或弱标记数据及引入额外的语义信息。下面我们将从这三种增强方式讨论当前基于数据增强的小样本目标检测方法。

4.3.1 样本与特征空间增强

基于数据增强的方法试图丰富监督信息，以缓解过拟合问题，从而获得可靠的检测模型。本节从样本空间增强和特征空间增强两个角度概括总结现有工作，前者对新类样本进行扩充，将小样本目标检测问题转化成监督目标检测问题，再利用已有的检测框架进行训练和测试；后者则针对图像经各种变换后的表示不变性这一特性，增强少量样本的特征表示，基于增强的特征表示完成检测过程。

样本空间增强。Kim等人[40]提出了一种空间数据增强方法(Spatial Data Augmentation, SDA)，在图像级增加训练示例，根据图像中目标的数量及其增强大小指数来放大训练示例。结合图像增强和实例增强两种方式丰富了新类的训练样本，在小样本数据集的基准上表现出优异的性能。Li等人[42]首次从样本扩增的角度入手，针对图像经遮挡、噪声等各种变换后的特征表示不一致问题，提出了一种简单有效的变换不变性方法(Transformation Invariant Principle, TIP)，在对各种变换图像的预测上，引入了一致性正则化来学习特征表示，通过学习设计变换一致性损失(Transformed Guidance Consistency, TGC)，计算原始特征表示和变换图像特征表示间的损失，使得图像经不同变换后的表示具有一致性，进而提高变换图像的泛化能力。该方法能够处理未标记数据，灵活用于各种元学习模型，而且能够扩展至更具挑战的半监督小样本目标检测场景中。

特征空间增强。Wu等人[83]利用目标特征表示在不同的视觉变化下具有不变性这一特性，提出了一种通用原型(Universal Prototype, UP)方法，用来表示不同类别的通用目标特征(Generalized Object Features, GOF)，这些通用原型用来捕获在不同视觉变化下的不变的目标内在特征，有效增强了原始的目标特征。同时，通过一致性损失函数(Consistency Loss, CL)来学习目标特征的不变性，进而提升目标特征学习的泛化能力；最终通过实验验证了该方法能够提高小样本检测模型的泛化能力。另一方面，训练数据缺乏变化亦是影响小样本目标检测的一个关键因素，其原因在于当单个样本的差异性极小时，无法训练新类的分类器。针对这一问题，Zhang等人[43]采用从极少的样本实例中建立具有外观变化的模型的方法，通过从基类迁移共享的类内变化的方式，为新类建立一个良好的变化模型；通过引入幻觉网络(Hallucinator Network, HN)，在RoI特征空间中生成额外的特征级训练示例，将类内差异的共享模式迁移至新类之上，从而有效提升小样本目标检测性能。该方首次证明了丰富样本变化对小样本

目标检测的有效性，而且，幻觉策略能够与其他方法相结合，从而进一步提升检测性能。然而，该文献仅考虑了新类的检测性能，未考虑基类的检测性能。

样本和特征空间增强相结合。Wu等人[44]首次发现了小样本下的尺度稀疏问题，提出了一种多尺度正样本细化方法(Multi-scale Positive Sample Refinement, MPSR)来丰富少量样本的目标尺度，通过构建目标金字塔形成多个尺度的正样本来增强样本的多样性，进一步利用特征金字塔(Feature Pyramid Net, FPN)提取多尺度样本的特征，达到增强样本特征表示的目的，并作为辅助分支集成至Faster RCNN网络中进行训练，通过实验证实，该方法在扩大正样本尺度的同时亦限制了不适当的负样本，从而进一步提升了小样本目标检测方法的精度。

4.3.2 引入未标记或弱标记的数据

通常而言，未标记或图像级标记数据容易获得，但是实例级注释却耗时耗力。因此，半监督或弱监督的小样本目标检测方法应运而生。Dong等人[15]提出一种自学习和多模型学习的半监督目标检测框架(Multi-modal Self-Paced Learning for Detection, MSPLD)，每个类别仅需2~4个注释示例。其核心是：首先为一些无注释的图像生成一系列伪标签，再利用网络预测出无标签图像的伪标签，进而选择置信度高的伪标签图像对检测模型进行训练，反复交替迭代训练，逐步提升模型的鲁棒性。Siddhesh等人[69]则设计一种统一的半监督学习框架UniT，在基类上学习一个从弱监督到全监督的检测模型映射，并利用新类和基类间视觉和语言相似性，迁移所学的映射至新类检测模型。整个模型采用端到端的训练方式，高度灵活，甚至在零样本条件下也能获得良好的性能。Hu等人[84]针对少量样本下的弱监督目标检测问题，利用图网络和空间相似度机制，搜索给定图像间的空间共性，并在没有边界框注释的查询图像中实现了公共目标的定位，但是该方法仅适用1-way 5-shot场景。更进一步，当前小样本下的弱监督检测方法已扩展至实际应用场景，Li等人[85]针对无人机导航过程中如何准确定位建筑物问题，提出了一种无人机图像检测框架(Hybrid Convolutional-Transformer, HCT)。该框架借助于目标建筑物的卫星图和街景图，将无人机视角的图像视为查询图像，对应的卫星视角和街道视角图像视为支持集。在支持集的基础上，仅利用图像级别的注释来训练网络，从输入图像的提取特征图中获取目标建筑的大致位置信息，从而实现无人机对目标的准确定位。

4.3.3 引入额外的语义信息

由于目标外观、形状和大小的不同，使得每个类别的视觉信息多变，导致类内差异较大。然而，每个类别的语义信息是不变的。因此，通过增加额外的语义信息来构建类间关系，以增强数据分布的多样性。在小样本分类任务中，Rahman和Chen等人[86][87]通过语义嵌入增强方式进行数据扩充，丰富样本的语义信息，以提升分类性能。受此启发，在小样本检测任务中，Cao等人[45]提出了一种基于联系和区分的小样本目标检测方法(Few-shot object detection via Association and DIscrimination, FADI)，通过一个大型英文词汇图WordNet[88]将每个新类与基类联系起来，结合丰富的词汇知识，对类间相似性关系进行描述，并利用Lin相似性[89]进行度量。同时，通过区分机制分离基类和新类的分类分支，以此减少类内间距离，增大类间距离，从而获得更好的检测性能。Zhu等人[41]通过引入语义信息以补充小样本下视觉信息的稀缺，利用基类和新类的语义关系不变这一特性，提出了一种语义关系推理小样本检测模型(Semantic Relation Reasoning Few-Shot Detector, SRR-FSD)，其核心是从视觉信息和语义关系中同时学习新类目标，采用词嵌入方式来构建语义空间，将来自视觉空间的目标投影到语义空间，实现视觉特征与类语义嵌入的对齐，获得增强的特征表示，进而提升小样本目标检测的性能。该方法在极少的样本下性能优越，证实了语义信息对小样本下图像信息的促进作用。Huang等人[90]认为少量的新类样本存在显著的类内差异，导致单个视觉信息无法准确地表征目标本身。针对这一问题，提出了一种语义增强机制(Semantic Enhancement, SE)，通过从大型文本语料库中学习词嵌入模型的方式获得类语义表示，以增强新类目标的视觉表示，证实了语义表示能够补充有限训练样本目标的视觉特征。同时，通过语义信息来定义语义原型的对比损失(Semantic

Prototype Contrastive, SPC)，使学习的特征具有良好的类内紧凑性和类间区分度，进而有效提升小样本目标检测性能。

4.3.4 小结

基于数据增强的小样本目标检测方法较为简单，具有较强的可解释性。上述工作表明，数据增强作为一种通用方法，能够与其他方法结合来提高小样本下的检测性能。然而，因标注样本数量不足，单纯的对小样本进行样本和特征空间的增强，仅能带来有限的性能提升，难以从根本上解决小样本目标检测问题。利用未标记数据或引入语义信息有助于提升数据分布的多样性，但这两类方法与标准的FSOD方法相比缺乏公平性。在未来的研究中，如何增加数据分布多样性，捕捉类内不变性，在不损失语义信息的前提下，实现样本或特征空间的增强是值得深入研究的一个方向。

4.4 基于度量学习的方法

基于度量学习的小样本目标检测方法采用学习比较的思想，在小样本图像分类任务下性能良好[91][92][93]。通常，由于检测模型中包含数以万计的参数，直接采用传统分类器的交叉熵损失进行学习，可能导致严重的模型过拟合问题。因此，基于度量学习的小样本目标检测方法借鉴传统非参数化的思想，在获取潜在目标边界框的特征之后，将查询图像的目标区域特征与支持集图像特征映射至相同的嵌入空间，并通过距离度量或者相似度计算实现对潜在目标区域的分类，进而实现对图像中目标的检测，并将其推广至具有少量标记训练示例的新类上实现对图像中不同目标的检测。图 8给出了基于度量学习的小样本目标检测框架，其工作原理如下：主干网络用于提取查询图像和支持图像的特征，RPN和RoI池化生成查询图像的感兴趣区域，编码网络将查询图像的感兴趣区域和支持集特征映射至嵌入空间，再通过度量模块进行相似度计算，检测查询图像中的不同目标。度量模块的结构如图 9所示，其核心是通过距离或者相似性度量方法对查询图像的RoI特征向量与支持类的原型进行比较，进而预测该向量的类别。

借助于以上思想，研究人员提出了一系列基于度量学习的方法，该类方法将小样本目标检测视为小样本分类问题，其侧重点在于边界框的分类部分，但不能简单地将小样本分类中的度量方法直接应用于小样本目标检测中，其原因在于检测模型需要知道潜在目标区域才能进行比较。目前，基于度量学习的小样本目标检测方法，主要从支持集图像的类原型表示、度量机制的实现以及损失函数设计等三个角度分别进行改进，下面我们就这三个方面进行归纳与总结。

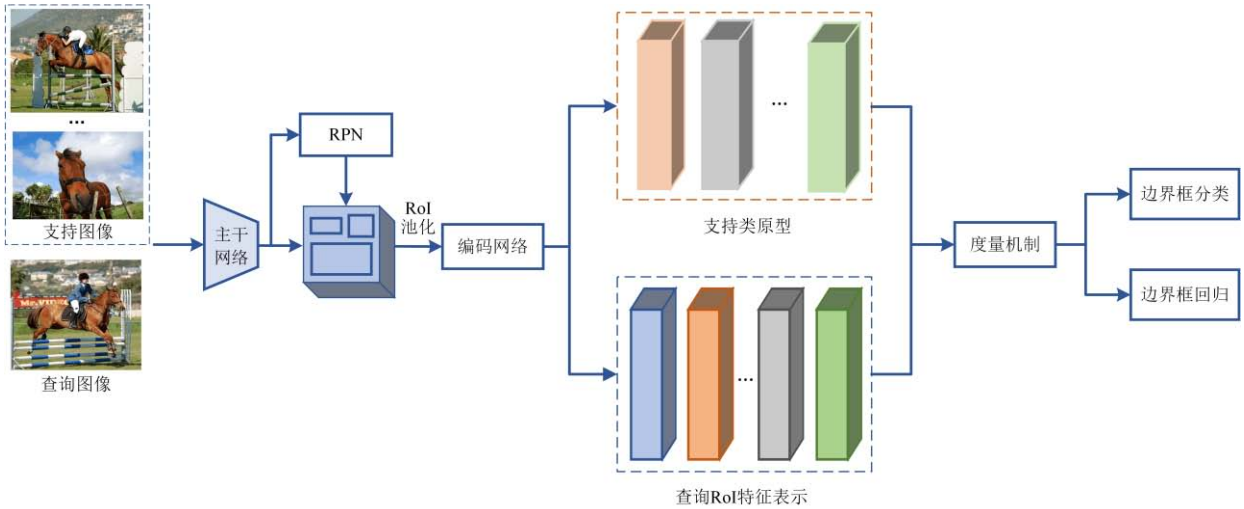

图 8 基于度量学习的小样本目标检测框架

Fig.8 Metric-learning based framework for few-shot object detection

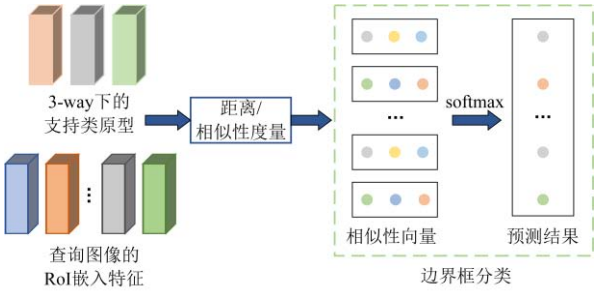

图 9 边界框分类的度量模块实现

Fig.9 Implementation of metric module for bounding box classification

4.4.1 类原型表示与度量机制

支持图像集中的类原型表示和度量机制的实现是影响查询图像预测的两个关键因素。其中，原型的概念始于原型网络[93]，其核心是从支持集中学习每个类的先验知识作为类原型表示，与查询图像的边界框特征进行比较，从而实现查询边界框特征的预测。基于此，Li等人[94]结合元学习和度量学习的思想，提出了一种小样本检测框架(Meta and Metric integrated Few-Shot Object Detection, MM-FSOD)。该框架包含元表示机制(Meta-Representation, MR)和皮尔逊度量机制(Pearson Metric, PR)两个部分。其中，MR机制通过学习不同类别支持实例的类内均值原型，对查询图像的RoI特征进行重构；采用PR机制计算重构后的RoI特征与支持原型之间的距离，进一步减少因背景信息的存在而造成类内方差较大的影响。MM-FSOD方法在不需要微调时能够快速学习新类别的表征。但是，该方法中的支持图像仅有单个实例块，不能充分表征每个类别的分布。借鉴深度度量学习的思想，Karlinsky等人[46]设计了一种基于表示的距离度量学习模型(Representative-based metric learning, RepMet)，采用具有多模态的混合模型表示每个类别，将模态中心作为类表示向量，同时学习主干网络参数、空间嵌入以及表示向量，在此基础上，进一步提出了一种子网结构来代替传统的边界框分类头，分别应用到单/小样本目标检测和小样本图像分类任务中，通过实验证实了该方法的有效性。概括起来，上述两种方法均是通过模拟小样本图像分类场景，仅利用图像中目标类别中正样本的边界框特征来实现分类，忽略了负样本尤其是困难样本的特征信息，可能导致训练样本利用不完善的问题。因此，Yang等人[47]提出一种度量学习框架NP-RepMet，结合正负样本的边界框特征信息学习嵌入空间，通过计算每个候选框的正/负样本特征的最小距离，以及与学习到的每个类别正/负样本表示的最小距离，将二者距离进行结合，再将距离通过概率度量机制转换为分类概率，从而能够有效地对每个边界框进行分类。与RepMet模型的性能相比，NP-RepMet模型在PASCAL VOC 数据集的5-way 1-shot实验下的性能提升了10%以上，表明了充分利用样本信息的重要性。为更好地学习类原型表示，Zhang等人[48]提出了一种基于深度度量学习的PNPDet方法，基于CenterNet网络[95]引入自适应余弦头(Adaptive Cosine Similarity Comparison Head, Adaptive CosHead)，通过为每个新类学习一个比例因子的方式，归一化特征表示，从而有效减小新类的类内方差。同时，采用余弦相似性距离作为度量方式，持续提升了所有类的检测性能。该方法简单、灵活且高效，可在提升新类检测性能的同时，保证所有类的检测性能。上述方法均采用固定的度量方式实现边界框的分类，受小样本学习中关系网络[92]的启发，Zhang等人[96]提出了一种单样本条件下的目标检测框架(One-Shot Conditional Object Detection, OSCD)。首先，采用孪生网络分别提取查询图像和目标图像的特征，作为边际概率用于特征空间的度量；其次，通过引入贝叶斯条件概率理论，在候选区域网络和分类网络设计可学习的度量机制获得条件概率估计，用于比较查询与目标图像特征，进而得到检测结果。但是，该方法仅关注单样本目标检测问题，尚未应用至少量样本的检测场景中。Lee等人[97]则将每个支持图像视为单个原型，通过类内支持注意力模块(Intra-Support Attention Module, ISAM)提取支持集原型间的共性信息，引入查询支持注意力模块(Query-Support Attention Module, QSAM)将其与查询特征进行比较，其特点是充分利用了支持图像间的信息，无需每类采用一个原型进行比较，有效提升了模型的分类效率。

综上所述，不难发现，类原型表示的设计旨在学习更好的类别表征，而度量机制逐渐从人工设计的距离度量向可学习的度量方式发展，能够进一步提升模型的效率。

4.4.2 损失函数设计

因数据及分布偏差，使得少量样本难以表征目标域的特征分布，造成类内距离较大，类间距离较小，可能出现类间重叠现象。因此，在基于度量学习的方法中，通常可以通过设计良好的损失函数来解决该问题，其目的是通过优化特征空间的方式，增加类间差异和减少类内差异，使得同类样本特征尽可能相似，不同类样本特征尽可能远离，从而减少类间重叠。为此，Hsieh等人[98]提出了一种基于边距的排序损失函数(Margin-based Ranking Loss, MRL)，用于隐式学习一种度量来预测区域边界框与查询图像特征间的相似性，在单样本目标检测数据集上实现了最先进的性能。在此基础上，Li等人[49]从特征空间优化角度入手，提出了一种类间距均衡方法(Class Margin Equilibrium, CME)，将小样本目标检测视为小样本分类问题，通过重新优化特征空间划分和新类的特征表示，并引入类间隔损失为新类保留足够的间隔空间，同时以对抗的最小最大化方式，扰动新类实例特征来达到间隔平衡，进而提升特征的可辨别性和表示能力，最后通过实验，验证了CME方法能够在小样本环境下对新类进行区分，进而减少类间重叠。概括起来，该类方法有助于学习更多的判别特征，能够更好地区分不同的类别，从而有效提升检测性能。

4.4.3 小结

综上所述，基于度量学习的小样本目标检测方法从支持图像中的类原型表示、边界框的度量机制以及损失函数设计三个方面进行改进，首先设计良好的支持集图像的类原型表示，与查询集图像的特征映射至同一嵌入空间，再通过特征优化和合适的度量机制实现边界框的分类。该类方法的优势是易实现增量学习，即模型在基类数据集上训练之后，能够直接用于新类别的检测；但在数据规模较大、特征维度较高时，导致边界框比较的计算时间长、消耗内存大等问题，因而不能有效满足对速度和实时性有较高要求的检测场景。在未来的研究中，如何设计更好的表示类别分布的方法以及如何设计更加合理的度量方式是值得研究的方向。

5. 数据集与实验

5.1 数据集

当前小样本目标检测的数据集主要包括PASCAL VOC[21]、MSCOCO[73]、LVIS[99]和FSOD[58]等四个数据集。

PASCAL VOC数据集。为创建小样本数据集，研究人员对PASCAL VOC数据集的20个类别进行划分，设置了三种不同的类别分割，每种分割设置15个类别作为基类，剩余5个类别作为新类。另外，VOC07+12训练验证集用于训练，VOC07测试集用于评估。

MSCOCO数据集。对于MSCOCO数据集中的80个类别，选择与VOC类别相同的20类作为新类，其余的60个类别作为基类。

LVIS数据集。LVIS数据集是一个满足长尾分布的大规模数据集，COCO图像被重新标注新的类别，超过1000个目标类别。其中，LVIS中的类别分为频繁类(出现在100多张图片中)、常见类(出现在10-100张图片中)和稀少类(出现在少于10张图片中)。为测试小样本检测方法在LVIS数据集下的检测性能，Wang等人[38]对该数据集进行K-shot分割，将频繁类和常见类作为基类，将稀少类作为新类。

FSOD数据集。2020年腾讯优图实验室专门为小样本目标检测构建的数据集。该数据集来源于Image Net[100]和Open Image[101]两个公开数据集，涵盖了1000个类别，共66502张图像和182000个标注框。其中，531个类别来自于Image Net，469个类别来自于Open Image。Fan等人[58]从中选择200类作为新类，其余800类作为基类。另外，该数据集由于其类别的高度多样性及边界框大小和长宽比的较大差异而极具挑战性。

5.2 实验设计

就训练样本的输入而言，基类提供全部图像和注释信息，新类则按照不同的实验设定选择相应图像和注释信息。PASCAL VOC数据集的新类从PASCAL VOC2007、PASCAL VOC2012中按照K=1、2、3、5、10进行实例抽样。MSCOCO数据集的新类从COCO2014中按照K=10、30进行实例抽样。LVIS数据集的新类按照K=10进行实例抽

样。FSOD数据集的新类则按照K=1、5进行实例抽样。

就训练策略而言，当前小样本目标检测方法包括基于任务的episode训练策略和基于数据驱动的训练策略两种方式。前者通过设计不同的小样本学习任务，在训练和测试阶段人为构建episode任务，每个任务包含支持集$S=\{(I_s,M_s)\}$和查询集$Q=\{(I_q,M_q)\}$，其中，支持集满足N-way K-shot设定，I_s和I_q分别表示支持集和查询集图像，M_s和M_q分别表示对应图像的边界框注释；对N个类别，每类包含K个实例的训练样本，查询集的每张图像可能包含多个类别目标，而支持集每张图像只包含一个有效目标，通过边界框裁剪或掩码分割得到。在元训练阶段，I_s和I_q均从基类数据集中采样；在元测试阶段，I_s和I_q从所有数据集中采样。后者采用“预训练-微调”的训练范式，在预训练阶段，基类数据集具有大量的训练样本，若某张基类训练图像包含新类目标，则被设置为背景；在微调阶段，利用新类或基类和新类的小样本平衡集(从基类和新类数据中随机采样K个实例级样本)对模型进行微调。

5.3 性能评估指标

对目标检测任务而言，模型的平均精度值(Average Precision, AP)表示模型对某一类目标的检测性能，基于准确率-召回率(Precision-Recall, PR)曲线来计算。其中，准确率表示分类正确的样本数占所有样本数的比例，召回率表示预测正确的正样本数与所有正样本数的比值。召回率作为x轴，准确率作为y轴绘制PR曲线，PR曲线下的面积即为AP值。实际应用中，并不直接对PR曲线进行计算，而是对PR曲线采用公式(1)进行平滑处理，再用公式(2)计算AP的值，计算公式如下：

$$p_{smooth}(r)=\max_{r'\geq r}p(r') \tag{1}$$

$$AP=\int_0^1 p(r)dr \tag{2}$$

一般而言，准确率与召回率不能同时达到最优，单独采用准确率或召回率不能较好地反映模型的检测性能。因此，通常采用模型的平均精度均值(mean Average Precision, mAP)来衡量检测精度，表示为各类别AP的均值，计算如下式(3)所示，其中，AP_i表示第i个类别的平均精度值，其值越大，表示模型对某一类目标的检测性能越好。通过设置不同的IoU阈值可以得到不同的mAP指标，如AP50和AP75分别表示IoU阈值等于0.5和0.75时的mAP值。

$$mAP=\frac{1}{N}\sum_{i=1}^{N}AP_i \tag{3}$$

FSOD方法更多考虑检测精度指标，早期的FSOD方法仅着眼于新类的性能。之后Wang等人[38]指出，不仅应该报告新类的检测性能，还应关注基类的表现，以提供更全面的评估。因此，FSOD的性能评估指标包含基类平均精度值(AP-base)和新类平均精度值(APnovel)，通过带入不同基类或新类的AP来计算。

PASCAL VOC数据集根据不同的基类/新类分割设定，计算不同基类/新类下的平均精度值。MSCOCO数据集通常采用标准的COCO指标进行评价，最常用的两个指标是AP和AP50，部分方法还用到了AP75、APs、APm、APl等评估指标。其中，平均精度和平均召回率用于小型、中型和大型目标。LVIS数据集中则还包括APr、APc及APf等评估指标，分别表示稀少类、常见类和频繁类的检测性能。

5.4 典型方法性能对比

5.4.1 PASCAL VOC数据集的实验结果

表 2列出了在三种不同的基类/新类类别分割设定下，小样本目标检测典型方法在PASCAL VOC数据集下的新类检测性能。表中红色加粗/蓝色加粗分别表示性能最优/次优结果。其中，FRCN V-16表示采用Faster RCNN检测模型，主干网络为VGG-16，FRCN R-101表示主干网络为ResNet-101，FRCN R-34则表示主干网络为ResNet-34，*表示基于Anchor-free的检测模型。对比的性能指标为新类别的检测精度。另外，我们将原始论文中的Meta Faster R-CNN[32]缩写为Meta FR-CNN、cos-FRCN+CGDP+FRCN[70]缩写为cos-FRCN-C、Retentive R-CNN[76]缩写为RR-CNN。

由表 2可知，大多数方法采用Faster RCNN作为基础检测模型，目前性能最好的方法是UniT，该方法融合了迁移学习和数据增强的优势，引入图像级的标注数据，在基类数据上学习了一个从

表 2 PASCAL VOC 数据集中对新类的小样本检测性能

Table 2 Few-shot detection performance on the PASCAL VOC novel set.

分类	方法	检测框架	会议(期刊)/时间	新类分割1					新类分割2					新类分割3				
				1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
基于元学习的方法	FSRW[30]	YOLOv2	ICCV 19	14.8	15.5	26.7	33.9	47.2	15.7	15.3	22.7	30.1	40.5	21.3	25.6	28.4	42.8	45.9
	Meta-RCNN[31]	FRCN R-101	ICCV 19	19.9	25.5	35.0	45.7	51.5	10.4	19.4	29.6	34.8	45.4	14.3	18.2	27.5	41.2	48.1
	Meta FR-CNN[32]	FRCN R-50	AAAI 22	41.8	46.7	52.7	59.6	62.3	26.1	33.6	43.8	47.8	50.1	35.6	42.1	45.8	53.4	52.3
	SQMG[33]	FRCN R-101	CVPR 21	48.6	51.1	52.0	53.7	54.3	**41.6**	**45.4**	45.8	46.3	48.0	46.1	**51.7**	**52.6**	54.1	55.0
	DRL[34]	FRCN R-101	TMM 21	28.0	40.5	49.4	49.9	59.4	22.9	33.4	36.4	36.1	52.7	28.0	32.0	40.4	46.7	53.5
	MetaDet[35]	FRCN V-16	ICCV 19	17.1	19.1	28.9	35.0	48.8	18.2	20.6	25.9	30.6	41.5	20.1	22.3	27.9	41.9	42.9
	Meta-DETR[36]	DETR* R-101	TPAMI 22	40.6	51.4	58.0	59.2	**63.6**	37.0	36.6	43.7	49.1	**54.6**	41.6	45.9	52.7	**58.9**	**60.6**
	FSDetView[54]	FRCN R-50	ECCV 20	24.2	35.3	42.2	49.1	57.4	21.6	24.6	31.9	37.0	45.7	21.2	30.0	37.2	43.8	49.6
	DCNet[55]	FRCN R-101	CVPR 21	33.9	37.4	43.7	51.1	59.6	23.2	24.8	30.6	36.7	46.6	32.3	34.9	39.7	42.6	50.7
	QA-FewDet[57]	FRCN R-101	ICCV 21	42.4	51.9	55.7	62.6	63.4	25.9	37.8	46.6	48.9	51.1	35.2	42.9	47.8	54.8	53.5
基于迁移学习的方法	LSTD[37]	FRCN V-16	AAAI 18	8.2	11.0	12.4	29.1	38.5	11.4	3.8	5.0	15.7	31.0	12.6	8.5	15.0	27.3	36.3
	TFA[38]	FRCN R-101	ICML 20	39.8	36.1	44.7	55.7	56.0	23.5	26.9	34.1	35.1	39.1	30.8	34.8	42.8	49.5	49.8
	FSCE[39]	FRCN R-101	CVPR 21	44.2	43.8	51.4	61.9	63.4	27.3	29.5	43.5	44.2	50.2	37.2	41.9	47.5	54.6	58.5
	FSOD-SR[40]	FRCN R-50	PR 21	50.1	54.4	56.2	60.0	62.4	29.5	39.9	43.5	44.6	48.1	43.6	46.6	53.4	53.4	59.5
	SRR-FSD[41]	FRCN R-101	CVPR 21	47.8	50.5	51.3	55.2	56.8	32.5	35.3	39.1	40.8	43.8	40.1	41.5	44.3	46.9	46.4
	UniT[69]	FRCN R-101	CVPR 21	**75.7**	**75.8**	**75.9**	**76.1**	**76.7**	**57.2**	**57.4**	**57.9**	**58.2**	**63.0**	**67.6**	**68.1**	**68.2**	**68.6**	**70.0**
	cos-FRCN-C[70]	FRCN R-50	CVPR 21	40.7	45.1	46.5	57.4	62.4	27.3	31.4	40.8	42.7	46.3	31.2	36.4	43.7	50.1	55.6
	DeFRCN[72]	FRCN R-101	ICCV 21	**53.6**	**57.5**	**61.5**	**64.1**	60.8	30.1	38.1	**47.0**	**53.3**	47.9	**48.4**	50.9	52.3	54.9	57.4
	AttFDNet[74]	SSD V-16	arXiv 20	29.6	34.9	35.1	-	-	16.0	20.7	22.1	-	-	22.6	29.1	32.0	-	-
	RR-CNN[76]	FRCN R-101	CVPR 21	42.4	45.8	45.9	53.7	56.1	21.7	27.8	35.2	37.0	40.3	30.2	37.6	43.0	49.7	50.1
基于数据增强的方法	TIP[42]	FRCN R-101	CVPR 21	27.7	36.5	43.3	50.2	59.6	22.7	30.1	33.8	40.9	46.9	21.7	30.6	38.1	44.5	50.9
	HallucFsDet[43]	FRCN R-101	CVPR 21	47.0	44.9	46.5	54.7	54.7	26.3	31.8	37.4	37.4	41.2	40.4	42.1	43.3	51.4	49.6
	MPSR[44]	FRCN R-101	ECCV 20	41.7	-	51.4	55.2	61.8	24.4	-	39.2	39.9	47.8	35.6	-	42.3	48.0	49.7
	FADI[45]	FRCN R-101	NIPS 21	50.3	54.8	54.2	59.3	63.2	30.6	35.0	40.3	42.8	48.0	45.7	49.7	49.1	55.0	59.6
	FSOD-UP[83]	FRCN R-101	ICCV 21	43.8	47.8	50.3	55.4	61.7	31.2	30.5	41.2	42.2	48.3	35.5	39.7	43.9	50.6	53.5
基于度量学习的方法	RepMet[46]	FRCN R-101	CVPR 19	26.1	32.9	34.4	38.6	41.3	17.2	22.1	23.4	28.3	35.8	27.5	31.1	31.5	34.4	37.2
	NP-RepMet[47]	FRCN R-101	NIPS 20	37.8	40.3	41.7	47.3	49.4	**41.6**	43.0	43.4	47.4	49.1	33.3	38.0	39.8	41.5	44.8
	PNPDet[48]	CenterNet*	WACV 21	18.2	-	27.3	-	41.0	16.6	-	26.5	-	36.4	18.9	-	27.2	-	36.2
	CME[49]	FRCN R-101	CVPR 21	41.5	47.5	50.4	58.2	60.9	27.2	30.2	41.4	42.5	46.8	34.3	39.6	45.1	48.3	51.5
	MM-FSOD[94]	FRCN R-34	CVIU 20	50.0	-	55.9	57.9	60.9	37.3	-	45.7	46.5	48.2	35.6	-	43.3	44.1	45.4

弱监督到全监督的检测模型映射，通过新类和基类间视觉和语言相似性，迁移所学的映射至新类检测模型，使得新类检测模型充分学习到基类的先验知识。然而，因额外图像的引入使得对比缺乏公平性。性能次之的方法是基于迁移学习的 DeFRCN 方法，该方法首次考虑了传统 Faster RCNN 模型不适合数据稀缺场景，设计了梯度解耦层调整各个模块间的耦合程度，将基类和新类的检测进行分离，从而在多个基准数据集上获得较高的性能。除此之外，基于迁移学习的 FSOD-SR 方法、基于数据增强的 FADI 方法以及基于度量学习的 MM-FSOD 方法在训练样本为 1-shot 时

的检测精度高达 50%。从表 2 中可以看出，基于迁移学习方法的性能在大多数任务下达到 SOTA，基于元学习方法的性能仅次于基于迁移学习方法，基于数据增强和度量学习方法的性能较差，表明在 PASCAL VOC 数据集上还有较大的提升空间。

针对以上四类小样本目标检测方法，每类分别选出三种经典算法进行检测性能的对比。如图 10 所示，每种算法中 5-shot、10-shot 的检测性能均比 1-shot、2-shot 的高，表明相同的算法在学习类别数相同的条件下，训练样本数越多，检测性能就越高。其中，整体性能最优/次优的方法分别是基于迁移学习的 FSOD-SR 方法和基于数据增强的 FADI 方法，且二者在极少样本(1-shot)下的检测效果最好。FSOD-SR 方法结合了迁移学习和数据增强的优势，通过编码基类和新类 RoI 特征间的关系来增强新类特征表示，同时融入空间数据增强方法以扩充图像级训练示例，从而获得更高的检测性能，证实了不同方法的融合能够进一步提升模型的检测性能。

此外，小样本目标检测中的部分方法重点关注广义小样本目标检测任务。这些方法在比较新类检测性能的同时，还关注基类的检测性能。表 3 给出了第一次分割下，四类经典方法对基类遗忘性能的比较。其中，nAP50表示IoU阈值等于0.5时的新类检测性能，bAP50表示IoU阈值等于0.5时的基类检测性能。从表 3的实验结果中可看出，基于迁移学习的两种方法TFA和SRR-FSD分别在基类和新类的检测精度指标上达到了最新水平。基于数据增强的方法FADI在新类别的检测性能上具有优势，而且能够保证对基类的检测性能不受影响，而基于元学习的方法对基类的检测性能较差。这侧面说明了元学习对新类具备较强的快速适应能力，但是对基类的性能保持能力稍弱于基于迁移学习的方法。除此之外，仅有个别基于数据增强和度量学习的方法给出了基类的性能指标。

图 10 四类经典方法在VOC数据集下新类别的小样本检测性能

Fig.10 Novel categories few-shot detection performance of four classical methods on the VOC dataset.

表 3 PASCAL VOC数据集下基类和新类的小样本检测性能

Table 3 Few-shot detection performance of Base classes and novel classes on the PASCAL VOC

分类	方法	nAP50		bAP50	
		3 shot	10 shot	3 shot	10 shot
基于元学习的方法	FSRW[30]	26.7	47.2	64.8	63.6
	Meta-RCNN[31]	35.0	51.5	64.8	67.9
	Meta-DETR[36]	53.2	**62.0**	70.0	73.0
	FSDetView[54]	42.2	57.4	69.4	71.1
基于迁移学习的方法	LSTD[37]	12.4	38.5	66.3	66.3
	TFA[38]	44.7	56.0	**79.1**	**78.4**
	FSCE[39]	46.8	59.7	73.7	75.2
	FSOD-SR[40]	**56.2**	**62.4**	77.4	77.4
	SRR-FSD[41]	51.3	56.8	78.2	**78.2**
基于数据增强的方法	MPSR[44]	51.4	61.8	67.8	71.8
	FADI[45]	**54.2**	-	**78.9**	-
基于度量学习的方法	PNPDet[48]	27.3	41.0	75.5	75.5

表 4 四类经典方法在COCO数据集10/30个样本下的新类检测性能

Table 4 Novel classes detection performance of four classical methods on MSCOCO in 10/30-shot case.

分类	方法	nAP		nAP50		nAP75		nAPs		nAPm		nAPl	
		10 shot	30 shot	10 shot	30 shot	10 shot	30 shot	10 shot	30 shot	10 shot	30 shot	10 shot	30 shot
基于元学习的方法	FSRW[30]	5.6	9.1	12.3	19.0	4.6	7.6	0.9	0.8	3.5	4.9	10.5	16.8
	Meta-RCNN[31]	8.7	12.4	19.1	25.3	6.6	10.8	2.3	2.8	7.7	11.6	14.0	19.0
	Meta FR-CNN[32]	12.7	16.6	25.7	31.8	10.8	15.8	-	-	-	-	-	-
	SQMG[33]	13.9	-	29.5	-	11.7	-	**7.6**	-	15.2	-	19.0	-
	DRL[34]	10.9	15.0	25.2	31.7	7.0	11.8	3.6	4.8	11.2	15.9	16.0	23.1
	MetaDet[35]	7.1	11.3	14.6	21.7	6.1	8.1	1.0	1.1	4.1	6.2	12.2	17.3
	Meta-DETR[36]	**19.0**	**22.2**	30.5	35.0	**19.7**	**22.8**	-	-	-	-	-	-
	FSDetView[54]	12.5	14.7	27.3	30.6	9.8	12.2	2.5	3.2	13.8	15.2	19.9	23.8
	DCNet[55]	12.8	18.6	23.4	32.6	11.2	17.5	4.3	6.9	13.8	16.5	21.0	27.4
	DAnA[56]	18.6	21.6	-	-	17.2	20.3	-	-	-	-	-	-
	QA-FewDet[57]	11.6	16.5	23.9	31.9	9.8	15.5	-	-	-	-	-	-
基于迁移学习的方法	LSTD[37]	3.2	6.7	8.1	15.8								
	TFA[38]	10.0	13.7	19.1	24.9	9.3	13.4	4.5	5.9	8.8	12.2	15.8	21.3
	FSCE[39]	11.9	16.4	-	-	10.5	16.2	-	-	-	-	-	-
	FSOD-SR[40]	11.6	15.2	21.7	27.5	10.4	14.6	4.6	**14.5**	10.5	14.5	17.2	24.7
	SRR-FSD[41]	11.3	14.7	23.0	29.2	9.8	13.5	-	-	-	-	-	-
	UniT[69]	**21.7**	**23.1**	**40.8**	**43.0**	**20.6**	**21.6**	**9.1**	**9.8**	**23.8**	**25.3**	**31.3**	**33.8**
	cos-FRCN-C[70]	11.3	15.1	20.3	29.4	-	-	-	-	-	-	-	-
	DeFRCN[72]	16.8	21.2	-	-	-	-	-	-	-	-	-	-
	AttFDNet[74]	12.9	16.3	19.5	24.6	13.9	17.3	-	-	-	-	-	-
	RR-CNN[76]	10.5	13.8	-	-	-	-	-	-	-	-	-	-
基于数据增强的方法	TIP[42]	16.3	18.3	**33.2**	**35.9**	14.1	16.9	5.4	6.0	17.5	**19.3**	25.8	**29.2**
	MPSR[44]	9.8	14.1	17.9	25.4	9.7	14.2	3.3	4.0	9.2	12.9	16.1	23.0
	FADI[45]	12.2	16.1	22.7	29.1	11.9	15.8	-	-	-	-	-	-
	FSOD-UP[83]	11.0	15.6	-	-	10.7	15.7	4.5	4.7	11.2	15.1	17.3	25.1
基于度量学习的方法	PNPDet[48]	5.5	-	-	-	-	-	-	-	-	-	-	-
	CME[49]	15.1	16.9	24.6	28.0	16.4	17.8	4.6	4.6	**16.6**	18.0	**26.0**	**29.2**
	MM-FSOD[94]	8.2	-	19.2	-	8.0	-	-	-	-	-	-	-

5.4.2 MSCOCO数据集的实验结果

表 4列举了四类经典的小样本目标检测方法在COCO数据集上的新类检测性能对比。其中，每类采用10个样本和30个样本，nAP表示新类别的平均检测性能，nAP50表示IoU阈值等于0.5时的新类别检测性能。从表 4的实验结果中可以看出，对于检测难度较大的COCO数据集而言，基于小样本的检测性能还有较大的提升空间。

与在PASCAL VOC数据集上的实验结果相同，UniT方法达到了当前的SOTA性能。除此之外，就nAP和nAP75指标而言，基于元学习的方法Meta-DETR在10-shot和30-shot下均获得最好的性能。就nAP50指标而言，基于数据增强的方法TIP在10-shot和30-shot下获得了最好的性能。对比nAP50和nAP75的结果可知，从nAP50到nAP75，检测性能大幅度下降，说明现有的FSOD方法仍然存在定位不准确的问题。对比nAPs、nAPm及nAPl的结果可知，对于不同尺度的目标而言，在小目标上的检测性能与大目标相比差距较大。以基于元学习的方法DCNet为例，30-shot下的nAPs性能比nAPl性能低20.5%。因此，现有的FSOD方法还需进一步解决通用目标检测中常见的尺度变化问题，提升在小目标上的检测性能。结合表 2和表 4可知，相同方法在两个基准测试集上的性能有所不同，在一个基准测试集上性能较好的方法，在另一个基准测试集上并不能保证仍然优于其他方法。例如：基于数据增强的方法TIP在PASCAL VOC数据集上性能较差，但在MSCOCO数据集上

表 5 两种方法在LVIS数据集10个样本下的新类检测性能

Table 5 Novel classes detection performance of two methods on LVIS in 10-shot case.

方法	主干网络	重复采样	AP	AP50	AP75	APs	APm	APl	APr	APc	APf
联合训练[99]			19.8	33.6	20.4	17.1	25.9	33.2	2.1	18.5	**28.5**
TFA fc[38]	FRCN R-50		22.3	37.8	22.2	18.5	28.2	36.6	14.3	21.1	27.0
TFA cos[38]			22.7	37.2	23.9	18.8	27.7	37.1	15.4	20.5	28.4
联合训练[99]			23.1	38.4	24.3	18.1	28.3	36.0	13.0	22.0	28.4
TFA fc[38]	FRCN R-50	√	24.1	39.9	25.4	19.5	29.1	36.7	14.9	23.9	27.9
TFA cos[38]			**24.4**	**40.0**	**26.1**	**19.9**	**29.5**	**38.2**	**16.9**	**24.3**	27.7
联合训练[99]			21.9	35.8	23.0	18.8	28.0	36.2	3.0	20.8	**30.8**
TFA fc[38]	FRCN R-101		23.9	39.3	25.3	19.5	29.5	38.6	16.2	22.3	28.9
TFA cos[38]			24.3	39.3	25.8	20.1	30.2	39.5	**18.1**	21.8	29.8
联合训练[99]			24.7	40.5	26.0	19.0	30.3	38.0	13.4	24.0	30.1
TFA fc[38]	FRCN R-101	√	25.4	**41.8**	27.0	19.8	31.1	39.2	15.5	26.0	28.6
TFA cos[38]			**26.2**	**41.8**	**27.5**	**20.2**	**32.0**	**39.9**	17.3	**26.4**	29.6

表现良好。总之，当前方法在PASCAL VOC数据集上的检测性能较高，在MSCOCO数据集上的检测性能有待提升。

5.4.3 LVIS数据集的实验结果

到目前为止，仅有TFA方法对LVIS数据集进行小样本下的性能评估。表 5列举了TFA方法与联合训练方法在LVIS数据集10-shot下的新类检测性能。其中，Gupta等人[99]采用一种加权采样方案来解决数据不均衡问题。当无重复采样时，对于稀少类别，TFA方法能够超越基线方法，将稀少类别的AP提升了约13%~15%，常见类的AP则提升了2%左右。有重复采样时，TFA方法将稀少类的AP提升了约4%，将常见类的AP提升了约1%~2%，证实了TFA方法不仅能够有效缓解严重的数据失衡问题，而且无需重复采样。

5.4.4 FSOD数据集的实验结果

表 6列举了四种小样本目标检测方法在FSOD数据集上的新类别检测性能对比。当前仅有少数方法采用该数据集展开研究。为公平比较，以下方法均采用5-shot设定，每类采用5个样本，在带有200个类别的测试数据集上进行测试，对比新类的检测精度。值得注意的是，Yes和No表示是否在FSOD数据集上预训练或微调，nAP50表示IoU阈值等于0.5时新类别的平均检测性能，nAP75表示IoU阈值等于0.75时新类别的平均检测性能。从表 6的实验结果中可以看出，基于元学习的方法FSOD和MM-FSOD不需要后续的微调步骤就能检测新类别，且MM-FSOD方法的性能与FSOD方法相比平均高出17.95%，而基于迁移学习的LSTD方法则需要进一步的微调。

表 6 四种方法在FSOD数据集5个样本下的新类检测性能

Table 6 Novel classes detection performance of four methods on FSOD in 5-shot case.

方法	FSOD 预训练	FSOD 微调	nAP50	nAP75
LSTD[37]	Yes	Yes	24.2	13.5
FRCNN[53]	No	Yes	11.8	6.7
FRCNN[53]	Yes	Yes	23.0	12.9
FSOD[58]	Yes	No	27.5	19.4
MM-FSOD[94]	Yes	No	51.7	31.1

6. 尚存问题与发展趋势

6.1 当前面临的挑战

当前小样本目标检测发展迅速，已引起研究者们的持续关注。然而，解决该问题的技术尚处于学术探索的阶段，其性能与大规模数据集下的目标检测技术存在很大差距。就目前的研究现状而言，仍面临如下几个方面的问题与挑战，下面我们进行分析并探讨相关应对思路。

(1) 小样本目标检测中的过拟合问题

当数据量较少时，深度检测模型更易发生过拟合现象，当前小样本目标检测模型的泛化能力在实际应用中还存在很大的差距[30][37]。与大规模数据训练的目标检测方法相比，小样本目标检测

方法的性能还有待提升。针对过拟合问题，可以从两个方面解决：一是从模型设计的角度出发，提出一种良好的迁移策略来降低模型过拟合程度，如正则化策略等[37]；二是从数据的角度出发，进行数据扩增或数据增强来提升模型的性能[42][102]。

(2) 小样本目标检测中的定位问题

当前小样本检测方法重点研究分类任务，将先验知识用于检测模型的边界框分类上，而对于定位任务，一般采用候选区域与类别无关的思想[94]。但在实际检测过程中，同类目标产生的候选区域可能被判定为背景，导致现有的小样本检测方法与大数据下的目标检测方法存在较大差距，错检漏检问题严重，使得模型的泛化能力受到严重的影响[30][31][46]。在样本数据缺乏的情况下如何更有效的提取目标的位置信息，融入类别信息至候选区域网络，产生更有效的候选区域，从而提升复杂场景下的定位准确度是未来需要解决的问题。

(3) 增量小样本目标检测问题

由于有限的计算资源或数据隐私问题，可能导致基类的训练数据访问受限，而当前小样本目标检测方法在很大程度上限制了在该场景下的可扩展性和适用性[103][104]，尤其是在开放式或机器人学习环境下的实际部署中基本上无法扩展。因此，增量小样本目标检测应运而生。鉴于当前研究还处于起步阶段，仅有少量文献对其进行研究，目前缺乏行之有效的方法。在增量小样本目标检测中，仅用少量新类数据进行训练，达到检测所有可见类别的目标；该场景存在以下两方面问题：一是因新类数据较少导致的过拟合问题；二是因缺乏基类数据导致的灾难性遗忘问题。针对以上问题，可以借鉴类增量小样本学习中的相关方法，引入知识蒸馏方法来解决遗忘问题，采用少量新类样本实现真正意义上的增量小样本检测[105][106]。

(4) 弱监督小样本目标检测问题

与小样本目标检测相比，弱监督小样本目标检测的新类样本仅提供少量的目标图像及类别标签信息，缺乏边界框标注信息，该场景更具挑战性[107]。其中，Yang 等人[108]基于度量学习框架提出了一种纳米监督的目标检测方法(Nano-Supervised Object Detection, NSOD)，该方法挖掘了所有支持图像的原型和查询图像的RoI特征间关系，利用支持类原型为每个查询RoI特征赋予一个伪标签，然后通过学习一个老师模型来完善查询RoI 特征的伪标签。该方法利用少量带注释的图像将图像级标签迁移至未标记图像中实例级的边界框。但是该方法计算的类原型可能包含噪声，影响最终的检测效果。Amirreza 等人[109]在预训练的嵌入空间中采用冯·米塞斯分布(von Mises distribution)，从而更好地捕捉语义信息正态分布；在新类别定位时，学习了一个线性外观模型来检测新图像中的新类别，并在大规模的弱监督目标检测和少量样本下的弱监督目标检测任务中均取得了较好的性能。Karlinsky 等人[110]提出了一种星状网络 StartNet，利用图像间的几何匹配关系进行定位和分类，实现了弱监督下的小样本目标检测。以上工作相对独立，相关研究不够完善，没有统一的数据集和算法性能评估标准，这些问题亟待突破。

(5) 域适应小样本目标检测问题

当前小样本目标检测方法主要针对同域不同类问题，即基类和新类数据的类别标签不同，但属于同一数据域，而域适应小样本目标检测主要关注同类不同域问题，即基类和新类数据标签相同，但属于不同的域，如基类数据来自于PASCAL VOC 或 MSCOCO 等自然景物图像数据集，而新类数据可能是线条简笔画或卡通图像等[111][112][113]。

域适应小样本目标检测是面向具体场景开展研究的，与小样本目标检测存在互补关系。但是域适应小样本目标检测仍处于起步阶段，当前研究还不成熟，未来可以考虑将域适应小样本目标检测和小样本目标检测进行融合，解决同类不同域和不同类不同域下的小样本目标检测问题[114]。

6.2 小样本目标检测的发展趋势

基于以上分析，基于现有的小样本检测方法可衍生三个未来值得深入研究的方向。

(1) 持续增量小样本目标检测

现实场景中的单个检测任务仅有少量带注释的训练图像，而大量的新任务会依次出现。这样的场景一般用于无人机探测任务[115]、无人驾驶[116]和机器人应用[117]等。大量的已有工作仅仅是面对单个的小样本检测任务，如何从小样本检测中继承迁移学习的思想，快速吸收新的检测任务

而不忘记旧的检测任务，实现连续的小样本目标检测仍然需要进行深入研究。未来可以将元学习与增量小样本目标检测方法相结合，探索相似任务的关联性，并应用至具体场景，以较小的推理代价快速适应新的任务[118][119]。

(2) 弱监督或域适应小样本目标检测

目标域图像仅有少量图像级标签注释，不需要很大的数据集，不需要边界框注释，仅需图像级标签，适合仅给定图像的目标分类和检测[120]。与监督小样本目标检测方法相比，该设定不需要手动标记注释框，更加现实，用于机器人技术[121]或视频目标分割[122]等应用场景中学习新类的检测模型。然而，当前弱监督小样本目标检测的研究还处于起步阶段，其模型性能与全监督下的小样本目标检测性能相比还有很大差距，未来还需进一步研究和探索。域适应小样本目标检测作为小样本目标检测的一种特殊应用，未来可扩展至复杂情况下的域适应小样本目标检测场景，如标签空间不一致和复杂目标域情况下的域适应小样本目标检测等场景[123]。

(3) 小样本目标检测的应用

结合学术研究和实际需求，开发面向用户的智能检测系统，仅需少量样本需要同时识别出未知类别和已知类别，并不断地进行增量学习。其中，主动学习主要是针对数据标签较少或打标签“代价”较高这一场景而设计的[124][125]，选择数据集上信息量较大的样本来降低标记成本，以尽可能少的标注样本实现高检测性能。未来可以将小样本目标检测与主动学习相结合，选择信息量较大的少量标注样本，让模型利用较少的标记数据获得较好性能。除了计算机视觉任务，其他领域的小样本目标检测同样值得探索，未来希望能够扩展至机器人中的视觉导航、医学领域中罕见药物的发现以及时间序列异常检测等诸多领域。

7. 总 结

小样本目标检测是目标检测领域新兴的问题，具有很大的应用价值。目前，关于系统性介绍小样本目标检测的中文文献相对缺失，不利于研究人员对该领域快速且深入地了解。鉴于此，本文对当前小样本目标检测的相关文献进行了分类整理和概括总结，对该领域的研究进展进行了广泛且全面的描述。首先，描述了小样本目标检测的任务定义及核心问题。然后，针对小样本下的模型训练问题详细地阐述了两种学习策略。按照工作原理将现有的小样本检测方法分为元学习、迁移学习、数据增强和度量学习四种学习范式，对这四类算法进行了系统、全面的总结，重点梳理了不同方法的核心思想、适用范围及优劣势，并对每类方法的下一步研究提出了见解，我们发现源域到目标域的知识迁移、目标域少量样本的数据或特征表示增强和边界框的分类是小样本目标检测的关键。尤其是，本文系统性地总结了当前小样本目标检测的四个数据集，对四类方法在数据集上的实验结果进行了详细的对比分析，最后归纳了小样本目标检测所面临的挑战与困境，并对未来可能的三个发展趋势进行了总结和探讨。

总体而言，小样本目标检测取得了很大进展，但各类方法都有各自的适用场景和局限性，现有算法仅停留在有限监督下的小样本目标检测场景，较难实现复杂场景下少量样本的目标检测，如：持续增量学习、弱监督或域适应等小样本目标检测场景。其中，面向无人机和机器人场景的持续增量小样本检测研究偏少，缺乏相应的试验性验证。弱监督或域适应小样本目标检测尚处于起步阶段，还需要根据特定的领域知识和任务特性设计有针对性的小样本检测算法。同时，算法的实际应用也存在挑战，在多种人工智能领域中具有广泛的应用价值和研究意义。因此，这些复杂场景下的小样本目标检测方法及其应用仍然是需要努力的目标。最后，希望本文能够促进其他学者对该领域的进一步探索，为该领域的未来研究提供有力的参考和帮助。

参 考 文 献

[1] Hu W C, Chen C H, Chen T Y, et al. Moving object detection and tracking from video captured by moving camera. Journal of Visual Communication and Image Representation, 2015, 30(C): 164-180

[2] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, USA, 2016: 770-778

[3] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84-90

[4] Zou Z, Shi Z, Guo Y, et al. Object detection in 20 years: A survey. arXiv preprint arXiv:1905.05055, 2019

[5] Zhang Y, Qiu Z, Yao T, et al. Fully convolutional adaptation networks for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 6810-6818

[6] Wu X, Xu K, Hall P. A survey of image synthesis and editing with generative adversarial networks. Tsinghua Science and Technology, 2017, 22(6): 660-674

[7] Huang X, He B, Tong M, et al. Few-shot object detection on remote sensing images via shared attention module and balanced fine-tuning strategy. Remote Sensing, 2021, 13(19): 3816

[8] Mahajan K, Sharma M, Vig L. Meta-dermdiagnosis: few-shot skin disease identification using meta-learning//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 730-731

[9] Wang T, Chen Y, Qiao M, et al. A fast and robust convolutional neural network-based defect detection model in product quality control. The International Journal of Advanced Manufacturing Technology, 2018, 94(9): 3465-3471

[10] Wang Y, Yao Q, Kwok J T, et al. Generalizing from a few examples: a survey on few-shot learning. ACM Computing Surveys, 2020, 53(3): 1-34

[11] Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. Science, 2015, 350(6266):1332–1338

[12] Li F F, Fergus R, Perona P. A bayesian approach to unsupervised one-shot learning of object categories//Proceedings Ninth IEEE International Conference on Computer Vision. Nice, France, 2003: 1134-1141

[13] Wang X, Yu F, Wang R, et al. Tafe-net: task-aware feature embeddings for low shot learning//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA, 2019: 1831-1840

[14] Huang S, Zeng X, Wu S, et al. Behavior regularized prototypical networks for semi-supervised few-shot image classification. Pattern Recognition, 2021, 112: 107765

[15] Dong X, Zheng L, Ma F, et al. Few-example object detection with model communication. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(7): 1641-1654

[16] Dong, X; Zheng L, Ma, F, Yang, Y. and Meng, D. Few-shot object detection. arXiv preprint arXiv:1706.08249. 2017

[17] Pan Xing-Jia, Zhang Xu-Long, Dong Wei-Ming, Yao Han-Xing, Xu Chang-Sheng. A survey of few-shot object detection. Journal of Nanjing University of Information Science and Technology (Natural Science Edition), 2019, 11(6):698-705 (in Chinese)
(潘兴甲, 张旭龙, 董未名,等. 小样本目标检测的研究现状. 南京信息工程大学学报：自然科学版, 2019, 11(6):698-705)

[18] Zhang Zhen-Wei, Hao Jian-Guo, Huang Jian, Pan Chong-Yu. Review of few-shot object detection. Computer Engineering and Applications, 2022, 58(05): 1-11 (in Chinese)
(张振伟, 郝建国, 黄健, 潘崇煜. 小样本图像目标检测研究综述. 计算机工程与应用, 2022, 58(05): 1-11)

[19] Liu Hao-Yu, Wang Xiang-Jun. A survey of few-shot object detection[J]. NAVIGATION AND CONTROL, 2021, 20(01): 1-14 (in Chinese)
(刘浩宇, 王向军. 小样本目标检测综述. 导航与控制, 2021, 20(01): 1-14)

[20] Jiaxu L, Taiyue C, Xinbo G, et al. A comparative review of recent few-shot object detection algorithms. arXiv preprint arXiv:2111.00201, 2021.

[21] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge. International journal of computer vision, 2010, 88(2): 303-338

[22] Joseph K J, Khan S, Khan F S, et al. Towards open world object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 5830-5840

[23] Perez-Rua J M, Zhu X, Hospedales T M, et al. Incremental few-shot object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 13846-13855

[24] Wang J, Cheng M M, Jiang J. Domain shift preservation for zero-shot domain adaptation. IEEE Transactions on Image Processing, 2021, 30: 5505-5517.

[25] Li Fan-Zhang, Liu Yang, Wu Peng-Xiang, et al. A survey on recent advances in meta-learning. Chinese Journal of Computers, 2021, 44(2): 25 (in Chinese)
(李凡长, 刘洋, 吴鹏翔,等. 元学习研究综述. 计算机学报, 2021, 44(2): 25)

[26] Zhou D, Zhou X, Zhang H, et al. Cheaper pre-training lunch: an efficient paradigm for object detection// Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020: 258-274

[27] Li Y, Zhang H, Zhang Y. Rethinking training from scratch for object detection. arXiv preprint arXiv:2106.03112, 2021

[28] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks//Proceedings of the IEEE Conference on International Conference on Machine Learning. Sydney, Australia, 2017: 1126-1135

[29] Diba A, Sharma V, Pazandeh A, et al. Weakly supervised cascaded convolutional networks//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 914-922

[30] Kang B, Liu Z, Wang X, et al. Few-shot object detection via feature reweighting//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea, 2019: 8420-8429

[31] Yan X, Chen Z, Xu A, et al. Meta r-cnn: towards general solver for instance-level low-shot learning//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea, 2019: 9577-9586

[32] Han G, Huang S, Ma J, et al. Meta faster r-cnn: towards accurate few-shot object detection with attentive feature alignment. arXiv preprint arXiv:2104.07719, 2021

[33] Zhang L, Zhou S, Guan J, et al. Accurate few-shot object detection with support-query mutual guidance and hybrid loss//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 14424-14432

[34] Liu W, Li H, Yu S, et al. Dynamic relevance learning for few-shot object detection. arXiv preprint arXiv:2108.02235, 2021

[35] Wang Y X, Ramanan D, Hebert M. Meta-learning to detect rare objects//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea, 2019: 9925-9934

[36] Zhang G, Luo Z, Cui K, et al. Meta-detr: image-level few-shot object detection with inter-class correlation exploitation. arXiv preprint arXiv:2103.11731, 2021, 2(6)

[37] Chen H, Wang Y, Wang G, et al. Lstd: A low-shot transfer detector for object detection//Proceedings of the AAAI conference on Artificial Intelligence. Palo Alto, USA, 2018, 32(1):2836-2843

[38] Xin Wang, Thomas E. Huang, Trevor Darrell, Joseph E Gonzalez, and Fisher Yu. Frustratingly simple few-shot object detection//Proceedings of the IEEE conference on International Conference on Machine Learning, 2020. 6

[39] Sun B, Li B, Cai S, et al. Fsce: few-shot object detection via contrastive proposal encoding//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA. 2021: 7352-7362

[40] Kim G, Jung H G, Lee S W. Spatial reasoning for few-shot object detection. Pattern Recognition, 2021, 120: 108118

[41] Zhu C, Chen F, Ahmed U, et al. Semantic relation reasoning for shot-stable few-shot object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 8782-8791

[42] Li A, Li Z. Transformation invariant few-shot object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 3094-3102

[43] Zhang W, Wang Y X. Hallucination improves few-shot object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 13008-13017

[44] Wu J, Liu S, Huang D, et al. Multi-scale positive sample refinement for few-shot object detection//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020: 456-472

[45] Y. Cao, J. Wang, Y. Jin, T. Wu, K. Chen, Z. Liu, and D. Lin. Few-shot object detection via association and discrimination//Proceedings of the IEEE/CVF Conference on Neural Information Processing Systems. 2021: 16570-16581

[46] Karlinsky L, Shtok J, Harary S, et al. Repmet: representative-based metric learning for classification and few-shot object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA, 2019: 5197-5206

[47] Yang Y, Wei F, Shi M, et al. Restoring negative information in few-shot object detection//Proceedings of the Neural Information Processing Systems. Bangkok, Thailand, 2020: 3521-3532

[48] Zhang G, Cui K, Wu R, et al. PNPDet: efficient few-shot detection without forgetting via plug-and-play sub-networks//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA, 2021: 3823-3832

[49] Li B, Yang B, Liu C, et al. Beyond max-margin: class margin equilibrium for few-shot object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 7363-7372

[50] Qiao S, Liu C, Shen W, et al. Few-shot image recognition by predicting parameters from activations//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 7229-7238

[51] Rusu A A, Rao D, Sygnowski J, et al. Meta-learning with latent embedding optimization. arXiv preprint arXiv:1807.05960, 2018

[52] Lee Y, Choi S. Gradient-based meta-learning with learned layerwise metric and subspace//Proceedings of the IEEE Conference on International Conference on Machine Learning. Stockholm, Sweden, 2018: 2927-2936

[53] Ren S, He K, Girshick R, et al. Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149

[54] Xiao Y, Marlet R. Few-shot object detection and viewpoint estimation for objects in the wild//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020: 192-210

[55] Hu H, Bai S, Li A, et al. Dense relation distillation with context-aware aggregation for few-shot object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 10185-10194

[56] Chen T I, Liu Y C, Su H T, et al. Dual-awareness attention for few-shot object detection. IEEE Transactions on Multimedia, 2021,25:291-301

[57] Han G, He Y, Huang S, et al. Query adaptive few-shot object detection with heterogeneous graph convolutional networks//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 3263-3272

[58] Fan Q, Zhuo W, Tang C K, et al. Few-shot object detection with attention-RPN and multi-relation detector//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 4013-4022

[59] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers//Proceedings of the IEEE/CVF European Conference on Computer Vision. Glasgow, US, 2020: 213-229

[60] J. Redmon and A. Farhadi. Yolo9000: better, faster, stronger//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 6517–6525

[61] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 2980-2988

[62] Chen D J, Hsieh H Y, Liu T L. Adaptive image transformer for one-shot object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 12247-12256

[63] Wang D, Cheng Y, Yu M, et al. A hybrid approach with optimization-based and metric-based meta-learner for few-shot learning. Neurocomputing, 2019, 349: 202-211

[64] Pan S J, Yang Q. A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 2009, 22(10): 1345-1359

[65] Nakamura A, harada T. Revisiting fine-tuning for few-shot learning. arXiv preprint arXiv:1910.00216, 2019

[66] Chen W Y, Liu Y C, Kira Z, et al. A closer look at few-shot classification. arXiv preprint arXiv:1904.04232, 2019

[67] Chen Y, Li W, Sakaridis C, et al. Domain adaptive faster r-cnn for object detection in the wild//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 3339-3348

[68] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: single shot multibox detector//Proceedings of the IEEE/CVF European Conference on Computer Vision. Amsterdam, Netherlands, 2016: 21–37

[69] Khandelwal S, Goyal R, Sigal L. UniT: unified knowledge transfer for any-shot object detection and segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021:5951-5961

[70] Li Y, Zhu H, Cheng Y, et al. Few-shot object detection via classification refinement and distractor retreatment//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 15395-15403

[71] Yang Z, Wang Y, Chen X, et al. Context-transformer: tackling object confusion for few-shot detection//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020: 12653-12660

[72] Qiao L, Zhao Y, Li Z, et al. Defrcn: decoupled faster r-cnn for few-shot object detection//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 8681-8690

[73] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014: 740-755

[74] Chen X, Jiang M, Zhao Q. Leveraging bottom-up and top-down attention for few-shot object detection. arXiv preprint arXiv:2007.12104, 2020

[75] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in neural information processing systems, 2017, 30

[76] Fan Z, Ma Y, Li Z, et al. Generalized few-shot object detection without forgetting//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 4527-4536

[77] Li Y, Cui H, et al. Class-incremental few-shot object detection. arXiv preprint arXiv:2105.07637, 2021

[78] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. Computer Science, 2015, 14(7):38-39

[79] Li Y, Zhu H, Ma J, et al. Towards generalized and incremental few-shot object detection. arXiv preprint arXiv:2109.11336, 2021

[80] Chen X, Wang Y, Liu J, et al. DID: Disentangling-imprinting-distilling for continuous low-shot detection. IEEE Transactions on Image Processing, 2020, 29:7765-7778

[81] Allen K, Shelhamer E, Shin H, et al. Infinite mixture prototypes for few-shot learning//Proceedings of the International Conference on Machine Learning. Los Angeles, USA, 2019: 232-241

[82] Riou K, Zhu J, Ling S, et al. Few-shot object detection in real life: case study on auto-harvest//Proceedings of the IEEE 22nd International Workshop on Multimedia Signal Processing. Tampere, Finland, 2020: 1-6

[83] Wu A, Han Y, Zhu L, et al. Universal-prototype enhancing for few-shot object detection//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 9567-9576

[84] Hu T, Mettes P, Huang J H, et al. Silco: show a few images, localize the common object//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea, 2019: 5067-5076

[85] Li S, Xue L, Feng L, et al. Hybrid convolutional-transformer framework for drone-based few-shot weakly supervised object detection. Computers and Electrical Engineering, 2022, 102: 108154

[86] Rahman S, Khan S, Barnes N, et al. Any-shot object detection//Proceedings of the Asian Conference on Computer Vision. Kyoto, Japan, 2020: 89-106

[87] Chen Z, Fu Y, Zhang Y, et al. Semantic feature augmentation in few-shot learning. arXiv preprint arXiv:1804.05298, 2018, 86: 89

[88] Miller, George A. WordNet: a lexical database for English. Communications of the Acm, 1995, 38(11):39-41

[89] D. Lin et al. An information-theoretic definition of similarity//Proceedings of the International Conference on Machine Learning. Madison, USA, 1998: 296-304

[90] Huang L, Dai S, He Z. Few-shot object detection with semantic enhancement and semantic prototype contrastive learning. Knowledge-Based Systems, 2022, 252: 109411

[91] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning. Advances in neural information processing systems, 2016, 29: 3630-3638

[92] Sung F, Yang Y, Zhang L, et al. Learning to compare: relation network for few-shot learning//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 1199-1208

[93] Snell J, Swersky K, Zemel R S. Prototypical networks for few-shot learning. arXiv preprint arXiv:1703.05175, 2017

[94] Li Y, Feng W, Lyu S, et al. MM-FSOD: meta and metric integrated few-shot object detection. arXiv preprint arXiv:2012.15159, 2020

[95] Duan K, Bai S, Xie L, et al. Centernet: keypoint triplets for object detection//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea, 2019: 6569-6578

[96] Zhang T, Zhang Y, Sun X, et al. Comparison network for one-shot conditional object detection. arXiv preprint arXiv:1904.02317, 2019

[97] Lee H, Lee M, Kwak N. Few-shot object detection by attending to per-sample-prototype//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa Beach Marriott Resort in Waikoloa, Hawaii,USA, 2022: 2445-2454

[98] Hsieh T I, Lo Y C, Chen H T, et al. One-shot object detection with co-attention and co-excitation//Proceedings of the Neural Information Processing Systems. Vancouver, Canada, 2019, 32

[99] Gupta, A., Dollar, P., and Girshick, R. Lvis: a dataset for large vocabulary instance segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019, 5356–5364

[100] Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 248-255

[101] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Tom Duerig, and Vittorio Ferrari. The open images dataset v4: unified image classification, object detection, and visual relationship detection at scale. arXiv:1811.00982, 2018

[102] Hariharan B, Girshick R. Low-shot visual recognition by shrinking and hallucinating features//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 3018-3027

[103] Dong N, Zhang Y, Ding M, et al. Incremental-detr: incremental few-shot object detection via self-supervised learning. arXiv preprint arXiv:2205.04042, 2022

[104] Cheng M, Wang H, Long Y. Meta-learning based incremental few-shot object detection. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(4):2158-2169

[105] Shmelkov K, Schmid C, Alahari K. Incremental learning of object detectors without catastrophic forgetting//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 3400-3409

[106] Peng C, Zhao K, Lovell B C. Faster ilod: incremental learning for object detectors based on faster rcnn. Pattern Recognition Letters, 2020, 140: 109-115

[107] Choe J, Oh S J, Lee S, et al. Evaluating weakly supervised object localization methods right//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 3133-3142

[108] Yang Z, Shi M, Xu C, et al. Training object detectors from few weakly-labeled and many unlabeled images. Pattern Recognition, 2021, 120: 108164

[109] Shaban A, Rahimi A, Ajanthan T, et al. Few-shot weakly-supervised object detection via directional statistics//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA, 2022: 3920-3929

[110] Karlinsky L, Shtok J, Alfassy A, et al. Starnet: towards weakly supervised few-shot object detection//Proceedings of the AAAI conference on Artificial Intelligence. Vancouver, Canada, 2021: 1743-1753

[111] Inoue N, Furuta R, Yamasaki T, et al. Cross-domain weakly-supervised object detection through progressive domain adaptation//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 5001-5009

[112] RoyChowdhury A, Chakrabarty P, Singh A, et al. Automatic adaptation of object detectors to new domains using self-training//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 780-790

[113] Xu J, Ramos S, Vázquez D, et al. Domain adaptation of deformable part-based models. IEEE transactions on pattern analysis and machine intelligence, 2014, 36(12): 2367-2380

[114] Gao Y, Yang L, Huang Y, et al. Acrofod: an adaptive method for cross-domain few-shot object detection//Proceedings of the IEEE/CVF European Conference on Computer Vision. Tel Aviv, Israel, 2022: 673-690

[115] Gao Y, Hou R, Gao Q, et al. A fast and accurate few-shot detector for objects with fewer pixels in drone image. Electronics, 2021, 10(7): 783

[116] Tambwekar A, Agrawal K, Majee A, et al. Few-shot batch incremental road object detection via detector fusion//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada, 2021: 3070-3077

[117] Lesort T, Lomonaco V, Stoian A, et al. Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. Information fusion, 2020, 58: 52-68

[118] Joseph K J, Rajasegaran J, Khan S, et al. Incremental object detection via meta-learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021: 0162-8828

[119] Liu Q, Majumder O, Achille A, et al. Incremental few-shot meta-learning via indirect discriminant alignment//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020: 685-701

[120] Wang T, Zhang X, Yuan L, et al. Few-shot adaptive faster r-cnn//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 7173-7182

[121] Kim D, Lee G, Jeong J, et al. Tell me what they're holding: weakly-supervised object detection with transferable knowledge from human-object interaction//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020: 11246-11253

[122] Lu X, Wang W, Ma C, et al. See more, know more: unsupervised video object segmentation with co-attention siamese networks//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 3623-3632

[123] Shi Y, Li J, Li Y, et al. Sensor-independent hyperspectral target detection with semi-supervised domain adaptive few-shot learning. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(8): 6894-6906

[124] Choi J, Elezi I, Lee H J, et al. Active learning for deep object detection via probabilistic modeling//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10264-10273

[125] Boney R, Ilin A. Semi-supervised and active few-shot learning with prototypical networks. arXiv preprint arXiv:1711.10856, 2017

SHI Yan-yan, Ph.D. candidate. Her current research interests include computer vision, few-shot learning, few-shot object detection etc.

SHI Dian-xi, Ph.D., researcher. His research interests include artificial intelligence, distributed computing etc.

QIAO Zi-teng, Ph.D. candidate. His research interests include computer vision and domain adaptive object detection.

ZHANG Yi, Ph.D., assistant researcher. His research interests include information security and artificial intelligence security.

LIU Yang-yang, M.S. candidate. His research interests include artificial intelligence, object detection etc.

YANG Shao-wu, Ph.D., associate researcher. His current research interests include artificial intelligence, SLAM etc.

Background

Few-shot object detection (FSOD) is currently a research hotspot that aims to solve the problems of the traditional object detection model, which requires a large amount of training data and iterative steps and performs poorly on less sample training data. It aims to classify and locate objects in images by a small number of labeled samples and is more challenging. Therefore, a higher requirement is put forward for the data utilization ability of the model. In the scenario of data scarcity, how to use a few labeled samples to learn, design a detection model with good generalization ability, and extend it to new tasks, is an urgent problem to be solved in FSOD. FSOD aims to design an object detection model that can classify and locate objects in images through a small number of labeled samples.

Generally, few-shot object detection has made significant progress, but all kinds of methods have their applicable scenarios and limitations. Existing algorithms only stay in few-shot object detection scenarios of limited supervision. It is difficult to achieve the object detection of a small number of samples in complex scenarios, such as continuous incremental learning, weak supervision, or domain adaptation. The studies of continuous incremental few-shot object detection for UAV and robot scenarios are little and lack corresponding experimental verification. Weak supervision or domain adaptive few-shot object detection is still in its infancy, and it is necessary to design detection algorithms according to specific domain knowledge and task characteristics. Therefore, the few-shot object detection method and its application in complex scenes remain to be studied.

There is a relative lack of Chinese literature on the systematic introduction of few-shot object detection. This is not conducive to researchers gaining a rapid and in-depth understanding of the field. Given this, this paper systematically summarizes the main methods of few-shot object detection, including meta-learning based, transfer-learning based, data augmentation based, and metric-learning based methods. In detail, we present the task definition and core problems, learning strategies, research methods, and experimental design of FSOD. The performance of the current few-shot object detection methods is compared. Finally, we outlook future challenges and possible developing directions of FSOD. This paper aims to review the latest methods for researchers in the few-shot object detection field and further promote the development of FSOD technology. We hope this paper inspires subsequent research works.

This paper is supported by the Science and Technology Innovation 2030 Major Project (No.2020AAA0104802) and the Integrated Program of National Natural Science Foundation of China (No.91948303).