

文章编号: 1008-1542(2022)06-0643-08

基于孪生神经网络的小样本目标检测综述

冯 珺¹, 彭梁英¹, 赵 帅¹, 潘司晨¹, 郭雪强²

(1. 国网浙江省电力有限公司信息通信分公司, 浙江杭州 310014; 2. 浙江大学控制科学与工程学院, 浙江杭州 310027)

摘要: 目标检测是计算机视觉的基础任务之一, 其主要任务是对图像中的目标进行分类和定位。小样本目标检测的目的就是利用极少数的训练样本实现对目标的检测, 从而减少繁杂的标注工作, 并实现在只有少量样本场景下的应用。现有的小样本目标检测方法主要包括基于孪生神经网络的方法和基于微调的方法, 这些方法通过利用现有的包含大量样本的基类数据集和包含少量样本的小样本数据集的训练, 使模型实现对小样本类别的分类和定位。重点调研了基于孪生神经网络的双分支小样本目标检测方法, 简要介绍了基于微调的小样本目标检测方案, 分析了这些方案的优缺点, 指出现有的小样本目标检测方案虽不成熟, 模型精度有待提升, 性能评估方案也有待完善, 但却有着十分广阔的应用前景, 未来若能通过深入研究解决小样本目标检测的现有问题, 其精度必将赶超传统目标检测。

关键词: 计算机感知; 目标检测; 小样本学习; 孪生神经网络; 图像

中图分类号: TN958.98

文献标识码: A

DOI: 10.7535/hbkd.2022yx06009

A survey of few-shot object detection based on siamese neural networks

FENG Jun¹, PENG Liangying¹, ZHAO Shuai¹, PAN Sichen¹, GUO Xueqiang²

(1. State Grid Zhejiang Electric Power Company Information and Communication Branch, Hangzhou, Zhejiang 310014, China; 2. College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China)

Abstract: Object detection is one of the basic tasks of computer vision, and its main task is to classify and locate the targets in the image. The purpose of few-shot object detection is to use a very small number of training samples to achieve the detection ability of the objects, so as to reduce the complicated annotation work, and realize the application in the scenarios with only a small number of samples. The existing methods for few-shot object detection mainly include siamese neural network-based methods and fine-tuning-based methods, which enable models to achieve the classification and localization ability of few-shot categories by using the existing base-class datasets containing a large number of samples and few-shot datasets containing a small number of samples. The two-branch few-shot object detection method based on siamese neural

收稿日期: 2022-11-09; 修回日期: 2022-11-28; 责任编辑: 张士莹

基金项目: 国网浙江省电力有限公司科技项目(B311XT210082)

第一作者简介: 冯 珺(1991—), 女, 湖北武汉人, 高级工程师, 博士, 主要从事人工智能技术在电力领域科研推进和产业落地等方面的研究。

E-mail: fengjun@zj.sgcc.com.cn

冯珺, 彭梁英, 赵帅, 等. 基于孪生神经网络的小样本目标检测综述[J]. 河北科技大学学报, 2022, 43(6): 643-650.

FENG Jun, PENG Liangying, ZHAO Shuai, et al. A survey of few-shot object detection based on siamese neural networks[J]. Journal of Hebei University of Science and Technology, 2022, 43(6): 643-650.

networks was focused on, and fine-tuning-based few-shot object detection schemes were briefly introduced. The advantages and disadvantages of these schemes were analyzed. It is pointed out that the existing small-sample target detection scheme is not mature, the precision of the model needs to be improved and the performance evaluation scheme needs to be improved. However, the small-sample target detection scheme has a very broad application prospect, and in the future, the existing problems of few-shot object detection will be solved by in-depth research, so that its accuracy can catch up with the traditional target detection.

Keywords: computer awareness; object detection; few-shot learning; siamese neural network; image

目标检测是对场景中的目标进行分类和定位,即在图片中用方框把待检测目标框出并标示出目标所属的类别。目标检测是计算机视觉的基础任务之一,广泛应用于人们的日常生活,比如人脸检测^[1]、目标跟踪^[2]、视频理解^[3]等。随着深度学习的兴起,基于深度学习的方法成为目标检测的主流方案。基于深度学习的目标检测方案主要包含2大类:单阶段目标检测器和双阶段目标检测器。单阶段目标检测器主要指YOLO(you only look once,你只需要看一次)^[4]系列,不需要生成候选框,直接使用端到端的卷积神经网络进行分类和定位。双阶段目标检测器以Faster R-CNN(faster region convolutional neural network,更快的区域卷积神经网络)^[5]及其变体为代表,首先通过RPN(region proposal network,候选框生成网络)生成一系列的候选框并对其进行简单分类和定位,然后通过后续的微调网络进行精细分类和定位。与单阶段目标检测器相比,双阶段目标检测器通常具有更高的精确度,因为通过RPN模块可以过滤掉很多的负样本,从而使检测器更容易识别和定位目标,因此小样本目标检测器大多使用Faster R-CNN作为基准网络。

小样本学习的目标是通过少量的训练样本就能够对新的概念进行识别。基于元学习的方法是近些年小样本分类任务的主流方法,通过训练单个任务的元模型来适应小样本的任务,即学习如何学习。基于度量学习的方法在最近的小样本分类任务中取得了很好的效果,它通过匹配网络,将查询图像和支持图像的特征编码为深度神经特征,并对它们执行加权最近邻匹配,通过两者的相似度对查询图像进行分类。最直观的度量包括余弦相似度、到类中心的欧氏距离和图距离等。目前,关于小样本分类任务的研究报道较多,但是针对更复杂的小样本目标检测任务的研究还很少。

传统的目标检测任务需要大量标注的样本,这需要耗费大量的人力;此外,还有一些类别的样本很难获得,比如稀有动植物等,这些都限制了目标检测任务的发展。相比于神经网络而言,人类中的小孩子也能根据一两张图片来记住一个类别。为了缩小神经网络与人类之间的差距,小样本目标检测任务提出在已有目标检测数据集和少量新类数据集训练下,使神经网络能够实现对新类别的分类和定位。

现有的小样本目标检测任务的研究主要包括2种方案:基于孪生神经网络(元学习)的方案和基于微调的方案。基于孪生神经网络的方案先在包含大量样本的基类上对目标检测网络进行训练,然后将小样本类别的图像分为支持集和询问集,其中支持集是同时包含图像和标注的图片,询问集是待检测的小样本图像,之后通过共享权重的特征提取网络得到两者的特征图,再通过特征聚合方式利用支持集的特征来调整询问集的识别网络,并通过损失函数使识别网络收敛,从而实现对小样本目标的检测。基于微调的方案同样是先在包含大量样本的基类上对目标检测网络进行训练,之后冻结一部分网络参数,在同时包含基类和新类(小样本类别)的数据集上微调另一部分的网络参数。现有的基于元学习的小样本目标检测方案在支持图像和询问图像的特征聚合时虽然使用的方式多种多样,但是仍没有很好地解决支持图像特征和询问图像特征之间的关系,很多情况下支持图像的加入反而会对原有的询问图像特征造成破坏,而且当支持集包含多张支持图像时,普遍采用的方法是求平均值的方式,这会损失很多信息。此外,基于微调的小样本目标检测方案在对于图像尺度和类别之间语义关系的处理上仍然没有有效的解决方案。因此,现有的小样本目标检测方法的精度与传统目标检测方法相比仍然有较大差距。

基于以上情况,本文结合当前小样本目标检测的研究情况,重点阐述基于孪生神经网络的小样本目标检测方法,并对基于微调的小样本目标检测方案进行简要概括。此外,为了方便对比,介绍了常用的数据集及其评估标准,在这些数据集和评估标准的基础下比较了一些经典方案的性能。最后,通过总结这些方案的优缺点,指出当前小样本目标检测存在的问题及面临的挑战。

1 小样本目标检测任务定义及方法分类

1.1 定义

小样本目标检测任务通常是将训练数据集 $D = D_{\text{base}} \cup D_{\text{novel}}$ 分为2个数据集 D_{base} 和 D_{novel} ,小样本目

标检测任务的目标就是在包含大量基类 C_{base} 样本的基类数据集 D_{base} 和包含少量小样本类别 C_{novel} 样本的小样本数据集 D_{novel} 的训练下,使神经网络能够实现对基类和新类别的分类和定位,其中 C_{base} 和 C_{novel} 没有交集。在小样本目标检测中,目标检测器 $f(x; \theta)$ 利用基类数据 D_{base} 和新类数据 D_{novel} 训练模型来获得对新类的检测能力,之后在测试集上进行测试以检验模型的性能。包含 N 个小样本类别且每个小样本类别包含 K 个样本的小样本目标检测问题称为 N 类别 K 样本的小样本目标检测问题,对于包含较少类别的小型数据集,每个小样本类别的数目一般不超过 10 个,对于包含较多类别的大型数据集,每个小样本类别的数目一般不超过 30 个。值得注意的是,小样本的数量 K 不一定与图像数量相对应,因为一个图像可能包含多个实例。

1.2 方法分类

1.2.1 基于孪生神经网络的小样本目标检测方法

基于元学习的小样本目标检测方案,即基于孪生神经网络的方案,是一种双分支的小样本目标检测方案。基于元学习的方案是小样本目标检测的主流方案,与此相关的方法有很多。如图 1 所示,该方法先在包含大量样本的基类数据集上对目标检测网络进行训练,然后将小样本类别的图像分为支持集和询问集,其中支持集同时包含已有的小样本图像及其标签,询问集是待检测的小样本图像,之后通过共享权重的特征提取网络(孪生神经网络)得到两者的特征图,再通过对两者特征图做特征聚合的方式,利用支持集的特征来调整询问集的识别网络,通过损失函数使识别网络收敛,从而实现对小样本目标的检测。基于元学习的小样本目标检测方法的区别主要在于特征聚合的方式以及位置的不同。

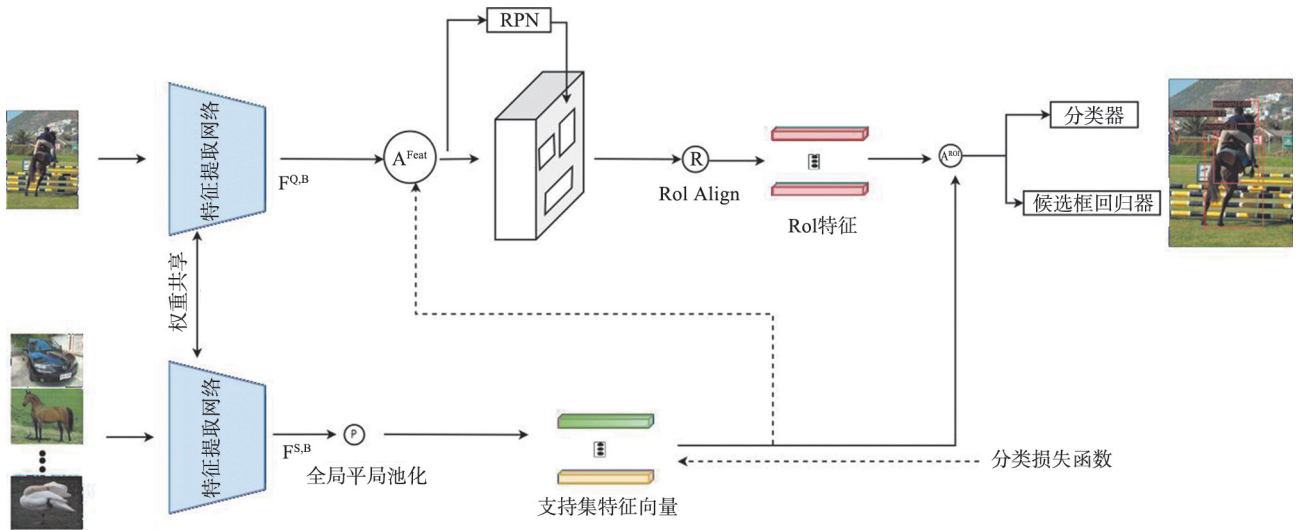


图 1 基于孪生神经网络的小样本目标检测

Fig.1 Few-shot object detection based on meta-learning

有学者提出了一种基于 YOLOv2^[6] 的轻量级元学习小样本目标检测模型,通过共享权重的特征提取网络支持集和询问集的特征,之后用支持集特征图生成的注意力向量加权询问类特征图对两者进行特征聚合,从而调整这些特征^[7]。Meta R-CNN^[8] 是一种基于 Faster R-CNN^[5] 的小样本目标检测框架,通过逐元素相乘的方式利用支持集的特征图来调整询问集的 ROI(region of interest,感兴趣的区域)池化模块,使网络对小样本类别更敏感。FSOD^[9] 也是一种基于 Faster R-CNN^[5] 的小样本目标检测模型,它利用支持集特征调整询问集的 RPN 生成候选框,从而过滤掉大部分的背景框和不匹配类别的候选框。同样是基于 Faster R-CNN,DCNet^[10] 则是利用支持集的特征图,通过知识蒸馏的方式来和询问集特征图进行聚合,之后进行分类和定位。Meta-DETR^[11] 将 transformer^[12] 的方法引入到基于元学习的小样本目标检测中,将支持集和询问集特征进行 transformer 编码后进行聚合,之后再进行 transformer 解码,大大提升了检测精度。Meta Faster R-CNN^[13] 在支持图像特征图和询问图像特征图逐元素相乘、相减和相连接之后,将三者分别输入 3 个包含激活函数的小型卷积网络中,再将卷积网络输出的结果联结在一起作为聚合结果,调整询问图像特征。SQMG^[14] 通过动态卷积方式,利用支持图像特征调整询问图像特征,该方法首先将支持图像特征图输入到一个核生成器中生成动态卷积的权重,然后利用该动态卷积的权重对询问图像特征图进行卷积操作。

在基于元学习的小样本目标检测方案的基础上,人们提出一种新的训练策略——Attention-RPN^[9],它在训练过程中对逐个类别的图像进行训练,因此每次训练时只输入一个类别的支持图像,并在网络最后对Faster R-CNN^[5]输出的RoI候选框区域和支持图像做一个特征匹配,从而挑选出匹配度最高的区域。QA-FewDet^[15]是在Attention-RPN的基础上对匹配方式做了改进,使用异构图(heterogeneous graph^[16])卷积方式对询问图像的候选框和支持图像的特征图进行度量。FCT^[17]则是将特征提取网络从ResNet^[18]改为了PVTv2^[19],并使用cross-transformer的方式进行特征聚合。

基于元学习的方案能够利用已有小样本目标的信息通过特征聚合方式调整目标检测模型,从而使模型具备对小样本目标的检测能力。但是已有的小样本目标检测方法并没有充分利用小样本信息,对于大于一个样本的小样本都是采用取平均值的方式,这会造成一定程度的信息浪费,而且都是在一个位置进行特征聚合,使得小样本信息不能对网络产生整体效果。

1.2.2 基于微调的小样本目标检测方法

基于微调的小样本目标检测方案是一种新兴方法,经过近两年的发展,取得了不俗的效果。基于微调的方案先在包含大量样本的基类上对目标检测网络进行训练,之后冻结一部分网络参数,在同时包含基类和新类(小样本类别)的数据集上微调另一部分的网络参数。TFA^[20]的结构如图2所示,它是一种基于Faster R-CNN^[5]的小样本检测框架,也是第1种基于微调的小样本目标检测方案,仅仅通过微调方法就实现了较高的小样本识别精度。TFA先用基类训练Faster R-CNN^[5]网络,之后用同时包含新类和基类的数据微调分类头和定位头,而其他部分的参数被冻结。FSCE^[21]对TFA做了进一步改进,通过加入聚类的思想使微调范围扩大,模型的识别精度进一步得到了提升。DeFRCN^[22]提出了一种解耦的Faster R-CNN^[5]模型,通过对模型梯度传播的解耦来微调参数,实现了小样本目标检测识别精度的进一步提升。MPSR^[23]同样是一种基于Faster R-CNN的小样本目标检测框架,该框架提出多尺度正样本精炼(multi-scale positive sample refinement, MPSR)方法,通过一个辅助分支来补偿多尺度信息,同时利用巧妙的损失函数设计,通过支持集调整询问集的分类头和定位头,实现小样本目标检测。FSOD-UP^[24]使用类似精炼方法,但是加入了差分损失函数,2个分支的特征相似性更高。SRR-FSD^[25]将每一个类别的特征映射到一个语义空间,通过新类和基类之间的语义关系来辅助网络对新类进行检测。基于微调的小样本目标检测方案同样有一些缺陷,因为基类信息和新类信息会有一定的差别,在基类上训练的参数并不能完全适应对新类的检测,因此只微调部分参数的方式并不能实现小样本信息的充分利用。

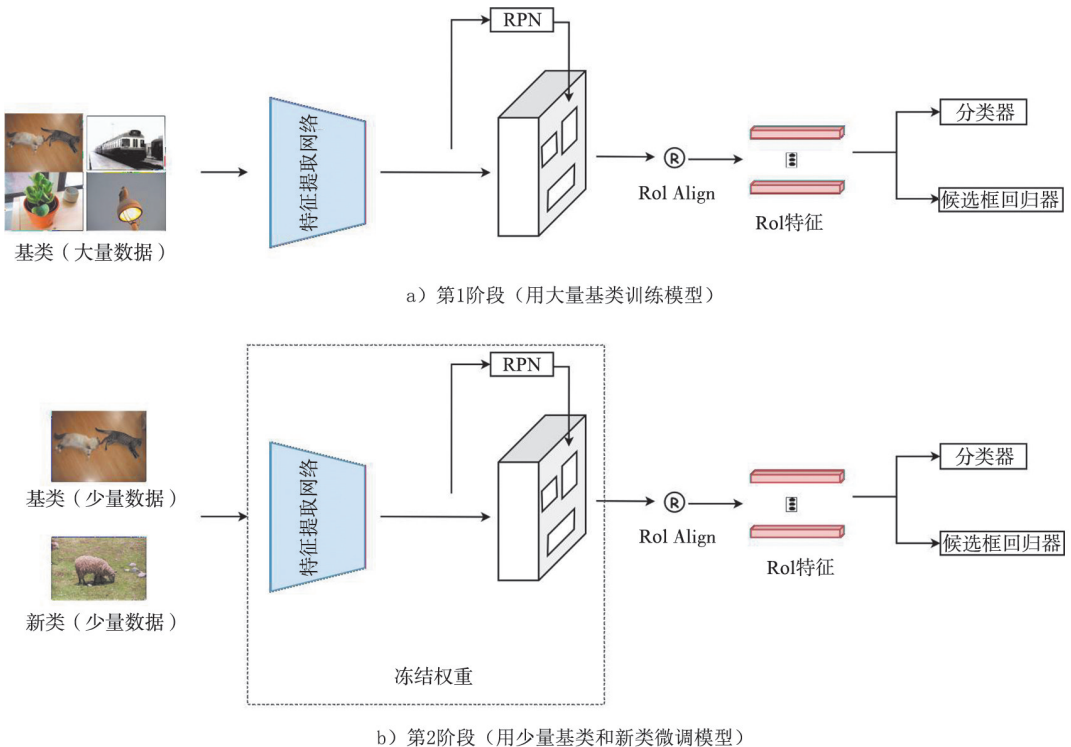


图2 基于微调的小样本目标检测

Fig.2 Few-shot object detection based fine-tuning

2 常用数据集及评估标准

2.1 数据集

大部分的小样本目标检测方法使用 PASCAL VOC^[26]和 Microsoft COCO^[27]的拆分数据集来进行实验和测试,即将数据集的类别划分为基类和小样本类别,然后根据 N 类别 K 样本的方式确定小样本类别的种类和每个类别包含的样本数。PASCAL VOC 和 Microsoft COCO 都是开源的已标注的目标检测数据集,在目标检测和实例分割等任务中有着十分广泛的应用。

2.1.1 PASCAL VOC 数据集

PASCAL VOC 数据集包含 20 个类别,通常由 VOC2007 和 VOC2012 组成,训练时使用 VOC07+12,测试时使用 VOC07。在小样本目标检测中,根据 Meta R-CNN^[8]设定,通常使用其中的 5 个类别作为新类,其余的 15 个类别作为基类,根据基类和新类的不同设定,将数据集分成 3 个子集。同时,针对 K -shot 问题,每个新类包含 K 个样本,通常取 $K=1,2,3,5,10$ 。

2.1.2 Microsoft COCO 数据集

Microsoft COCO 是比 PASCAL VOC 更大的数据集,因此识别难度也更大。Microsoft COCO 包含 80 个类别,通常将与 PASCAL VOC 相同的 20 个类别作为新类,而剩余的 60 个类别作为基类来划分数据集。同时,针对 K -shot 问题,通常取 $K=10,30$,在某些极少样本的情况下也会取 $K=1-5$ 。

2.2 评估标准

与传统目标检测的技术指标相同,小样本目标检测的技术指标主要包括 AP, AP50 和 AP75。AP, AP50 和 AP75 指的是在 IoU(交并比)分别取 $[0.5:0.05:0.95]$ 、0.5 和 0.75 时各个类别识别准确率的均值,其中 $[0.5:0.05:0.95]$ 指的是 IoU 从 0.5 开始每隔 0.05 一直取值到 0.95,然后取均值。其中 IoU 指的是识别出的检测框和目标实际的检测框(指手动标注的检测框)之间的交集面积和并集面积的比值。mAP, mAP50 和 mAP75 分别表示 AP, AP50 和 AP75 在所有类别上求均值。PASCAL VOC 数据集一般用 mAP50 来作为评估标准,为了使结果更有说服力,同时减少因为样本选取而带来的结果不稳定性,TFA^[9]提出了修订的评估标准,其中的结果平均了 30 次,每次使用随机抽取的小样本目标来作为小样本数据集;此外, TFA 还报告了因为忽略基类的性能可能会导致潜在的性能下降,因此会同时评估模型在基类和新类上的性能。Microsoft COCO 数据集一般使用 mAP, mAP50 和 mAP75 作为评估标准, mAP75 更严格,因为当检测框与真实框大于 0.75 时,检测才视为正确。某些方法还报告了平均召回率 AR1, AR100 和 AR1 000,它们分别为 1, 100 或 1 000 个图像检测。中小型、中型和大物体(APS, APM, APL, ARS, ARM, ARL)也可以使用平均精度和平均召回率。

3 基于元学习的小样本目标检测典型方法的结果对比与分析

为了充分对比基于孪生神经网络的小样本目标检测方法,本文使用经典的 PASCAL VOC 数据集和 Microsoft COCO 数据集上的结果来进行对比,基类和新类的划分也采用前边提到的经典划分方式。此外,对每种方法的基准检测网络做了备注。所有的结果均来自于论文结果和作者开源代码复现结果,表 1 是 PASCAL VOC 上的结果对比,它包含 3 种不同的新类划分,以及 5 种样本数划分^[26];表 2 是 Microsoft COCO 上的结果对比,包含 10 个样本和 30 个样本 2 种样本数的结果^[27]。

表 1 模型在 PASCAL VOC 上的准确率对比
Tab.1 Comparison of model accuracy on PASCAL VOC

单位:%

模 型	基准网络	分组 1					分组 2					分组 3				
		1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
FSRW ^[6]	YOLOv2	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 R-CNN ^[8]	FRCN-r101	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
FsDetView ^[28]	FRCN-r101	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 ^[10]	FRCN-r101	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
Meta-DETR ^[11]	DETR-r101	35.1	49.0	53.2	57.4	62.0	27.9	32.3	38.4	43.2	51.8	34.9	41.8	47.1	54.1	58.2

表2 模型在 Microsoft COCO 上的准确率对比
Tab.2 Comparison of model accuracy on Microsoft COCO

单位: %

Shot	模型	AP	AP 50	AP 75	APS	APM	APL	AR 1	AR 10	AR 100	ARS	ARM	ARL
10	FSRW ^[6]	5.6	12.3	4.6	0.9	3.5	10.5	10.1	14.3	14.4	1.5	8.4	28.2
	Meta RCNN ^[8]	8.7	19.1	6.6	2.3	7.7	14.0	12.6	17.8	17.9	7.8	15.6	27.2
	FsDetView ^[28]	12.5	27.3	9.8	2.5	13.8	19.9	20.0	25.5	25.7	7.5	27.6	38.9
	DCNet ^[10]	12.8	23.4	11.2	4.3	13.8	21.0	18.1	26.7	25.6	7.9	24.5	36.7
	Meta-DETR ^[11]	17.8	28.8	18.5	3.3	14.0	29.3	21.0	32.2	34.1	7.9	29.9	56.0
30	FSRW ^[6]	9.1	19.0	7.6	0.8	4.9	16.8	13.2	17.7	17.8	1.5	10.4	33.5
	Meta RCNN ^[8]	12.4	25.3	10.8	2.8	11.6	19.0	15.0	21.4	21.7	8.6	20.0	32.1
	FsDetView ^[28]	14.7	30.6	12.2	3.2	15.2	23.8	22.0	28.2	28.4	8.3	30.3	42.1
	DCNet ^[10]	18.6	32.6	17.5	6.9	16.5	27.4	22.8	27.6	28.6	8.4	25.6	43.4
	Meta-DETR ^[11]	22.9	35.8	23.8	4.7	20.9	36.5	23.3	36.0	38.4	12.5	36.0	59.9

尽管大部分的小样本目标检测模型都是基于双阶段的目标检测器 Faster R-CNN^[5],但是双阶段目标检测器比较复杂,且生成的候选框很多都是背景,因此在哪个位置聚合询问集和支持集的特征就比较重要。相比之下,单阶段目标检测器不需要生成候选框,因此只需直接在特征提取网络之后、分类和定位网络之前进行特征聚合,FSRW^[6]便是一种基于 YOLOv2^[7]的小样本目标检测框架,其特点是轻量化,但受限于基准检测器,模型精度不是很高。类似的,Meta DETR^[11]使用的基准框架是基于 Transformer^[12]的目标检测器 DETR^[27],同样是在共享权重的孪生神经网络之后进行询问集和支持集的特征聚合,因为基准检测器 DETR 准确率很高,因此 Meta DETR 也具有更好的识别精度,但其缺点是模型大,耗费的训练时间和训练资源比较多。

Meta R-CNN^[8]是早期以 Faster R-CNN^[5]为基准的基于孪生神经网络的小样本目标检测框架,主要是在 RPN 网络之后对支持集和询问集进行特征聚合,采用的特征聚合方式是向量相同位置元素相乘,如式(1)所示。FsDetView^[28]是在 Meta R-CNN 的基础上改进了特征聚合方式,如式(2)所示,除了矩阵相乘外,还加入了询问集和支持集的差值以及询问集特征,使模型精度有了较大提升。DCNet^[10]同样是在 Meta R-CNN 的基础上进行改进,由于在 RPN 网络之后进行特征聚合会使一部分的新类候选框在 RPN 网络中被当作负样本过滤掉,因此将特征聚合操作放在了 RPN 网络之前,共享权重的特征提取网络之后对支持集和询问集的特征进行聚合,采用知识蒸馏的聚合方式,增强了图像中待检测目标的关键信息,同时加入了多尺度聚合模块对不同尺度的特征进行聚合,提升了对不同尺度目标的检测能力。

$$\Phi(f^Q, f^S) = f^Q \otimes f^S, \quad (1)$$

$$\Phi(f^Q, f^S) = [f^Q \otimes f^S, f^Q - f^S, f^Q]. \quad (2)$$

4 当前面临的挑战

4.1 模型精度有待提高

虽然目前小样本目标检测已经取得了较大发展,但相比于传统目标检测,在精度上仍有较大差距,尤其是在极少数样本的情况下,更没有达到人类的水平。而较低的精度又会限制小样本目标检测的实际应用,其仍处于研究阶段,落地比较困难。但是针对极少数样本的情况,目前也有一些小样本目标检测的衍生任务,比如单样本检测,只需针对单个样本的任务进行模型设计,不需要考虑多个样本之间的差异性,虽在单个样本的任务中取得了很好的结果,但仍不能达到落地要求^[29]。

4.2 模型性能的评估标准有待完善

由于小样本任务的特殊性,使得样本选择对结果的影响很大,不同的样本和样本组合可能会得到差距很大的结果。因此 TFA 提出利用多次随机选择样本进行模型训练,之后取平均值的评估方式。此外,关于小样本任务的定义仍不是很清晰,比如在 Microsoft COCO 上的训练,在基类图片中就包括一些新类的未标注样本,这种情况下新类的样本数是多于指定的任务的。此外,现有的评估方案基本都是针对 PASCAL VOC 和 Microsoft COCO 2 个数据集的,并没有针对特定任务来进行数据集和评估标准的设计,因此也没有合适的对比方案。同时,大多数方法都使用 ImageNet^[30]预训练的模型权重,而 ImageNet 中包含大部分的新类。

尽管这对于传统目标检测很常见,但它对于小样本目标检测具有负面影响:新类不再是真正的新类,因为该模型可能已经看到了此类别的图像。但是,省略 ImageNet 预训练,即使对于基类,也会导致性能较差。针对这些问题,目前仍没有较好的解决方案,有些方法提出在预训练和基类训练的过程中剔除包含新类的图片,但是这需要针对每一个小样本数据集都进行一次预训练,显然需要耗费大量的时间和精力。

5 结 语

本文基于小样本目标检测的任务和目标,介绍了基于孪生神经网络的小样本目标检测方案,并简要介绍了其他一些小样本目标检测方法,在此基础上,对几种经典方法进行了对比,分析了各自的优缺点。

本文是第 1 篇系统概括基于孪生神经网络小样本目标检测的中文综述文章,可为小样本目标检测的深入研究以及落地应用提供参考。由于篇幅有限,本文仅对基于孪生神经网络的小样本目标检测模型的实现方式与效果进行了概括。虽然现有的小样本目标检测方案仍不成熟,模型精度有待提升,性能评估方案也有待完善,但小样本目标检测有着十分广阔的应用前景,未来必定有更多方法来解决小样本目标检测的现有问题,使其精度达到要求。

参考文献/References:

- [1] SUN Yi, LIANG Ding, WANG Xiaogang, et al. Deepid3: Face Recognition with very Deep Neural Networks[DB/OL]. <https://arxiv.org/abs/1502.00873>, 2015-02-03.
- [2] 杨文焕, 翟雨, 殷亚萍, 等. 一种多目标检测跟踪算法研究[J]. 河北科技大学学报, 2022, 43(2): 127-136.
YANG Wenhuan, ZHAI Yu, YIN Yaping, et al. Research on a multi-target detection and tracking algorithm[J]. Journal of Hebei University of Science and Technology, 2022, 43(2): 127-136.
- [3] 杨丽娟, 刘教民, 王震洲, 等. 基于分块帧差的视频图像运动检测[J]. 河北科技大学学报, 2006, 27(1): 89-92.
YANG Lijuan, LIU Jiaomin, WANG Zhenzhou, et al. Video motion detection based on blocking and frame difference[J]. Journal of Hebei University of Science and Technology, 2006, 27(1): 89-92.
- [4] CHEN Qiang, WANG Yingming, YANG Tong, et al. You only look one-level feature[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 13034-13043.
- [5] REN S Q, HE K M, GIRSHICK R B, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015. Montreal: Curran Associates Inc, 2015: 91-99.
- [6] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6517-6525.
- [7] KANG Bingyi, LIU Zhuang, WANG Xin, et al. Few-shot object detection via feature reweighting[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 8419-8428.
- [8] YAN Xiaopeng, CHEN Ziliang, XU Anni, et al. Meta R-CNN: Towards general solver for instance-level low-shot learning[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 9576-9585.
- [9] FAN Q, ZHUO W, TANG C K, et al. Few-shot object detection with attention-RPN and multi-relation detector[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 4012-4021.
- [10] HU Hanzhe, BAI Shuai, LI Aoxue, et al. Dense relation distillation with context-aware aggregation for few-shot object detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 10180-10189.
- [11] ZHANG Gongjie, LUO Z, CUI K, et al. Meta-detr: Few-shot Object Detection Via Unified Image-level Meta-learning[DB/OL]. <https://arxiv.org/abs/2103.11731v1>, 2021-09-20.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc, 2017: 6000-6010.
- [13] HAN Guangxing, HUANG Shiyuan, MA Jiawei, et al. Meta faster R-CNN: Towards accurate few-shot object detection with attentive feature alignment[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(1): 780-789.
- [14] ZHANG Lu, ZHOU Shuigeng, GUAN Jihong, et al. Accurate few-shot object detection with support-query mutual guidance and hybrid loss[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 14419-14427.
- [15] HAN Guangxing, HE Yicheng, HUANG Shiyuan, et al. Query adaptive few-shot object detection with heterogeneous graph convolutional networks[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 3243-3252.
- [16] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[C]//5th International Conference on Learning Representations. Toulon: ICLR, 2017: 1-17.
- [17] HAN Guangxing, MA Jiawei, HUANG Shiyuan, et al. Few-shot object detection with fully cross-transformer[C]//2022 IEEE/CVF Con-

ference on Computer Vision and Pattern Recognition (CVPR).New Orleans:IEEE,2022;5311-5320.

- [18] HE Kaiming,ZHANG Xiangyu,REN Shaoqing,et al.Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Las Vegas:IEEE,2016;770-778.
- [19] WANG Wenhai,XIE Enze,LI Xiang,et al.PVT v2:Improved baselines with pyramid vision transformer[J].Computational Visual Media, 2022,8(3):415-424.
- [20] WANG Xin,HUANG T E,DARRELL T,et al.Frustratingly Simple Few-shot Object Detection[DB/OL].<https://arxiv.org/abs/2003.06957>,2020-03-16.
- [21] SUN Bo,LI Banghuai,CAI Shengcai,et al.FSCE:Few-shot object detection via contrastive proposal encoding[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Nashville:IEEE,2021;7348-7358.
- [22] QIAO Limeng,ZHAO Yuxuan,LI Zhiyuan,et al.DeFRCN:Decoupled faster R-CNN for few-shot object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV).Montreal:IEEE,2021;8661-8670.
- [23] WU Jiayi,LIU Songtao,HUANG Di,et al.Multi-scale positive sample refinement for few-shot object detection[C]//Computer Vision-ECCV 2020.Cham:Springer,2020;456-472.
- [24] WU Aming,HAN Yahong,ZHU Linchao,et al.Universal-prototype enhancing for few-shot object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV).Montreal:IEEE,2021;9547-9556.
- [25] ZHU Chenchen,CHEN Fangyi,AHMED U,et al.Semantic relation reasoning for shot-stable few-shot object detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Nashville:IEEE,2021;8778-8787.
- [26] EVERINGHAM M,VAN GOOL L,WILLIAMS C K I,et al.The pascal visual object classes (VOC) challenge[J].International Journal of Computer Vision,2010,88(2):303-338.
- [27] CARION N,MASSA F,SYNNAEVE G,et al.End-to-end object detection with transformers[C]//Computer Vision-ECCV 2020.Cham:Springer,2020;213-229.
- [28] XIAO Yang,MARLET R.Few-shot object detection and viewpoint estimation for objects in the wild[C]//Computer Vision-ECCV 2020. Cham:Springer,2020;192-210.
- [29] WU X W,SAHOO D,HOI S C H.Recent advances in deep learning for object detection[J].Neurocomputing,2020,396:39-64.
- [30] KRIZHEVSKY A,SUTSKEVER I,HINTON G E.Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems 25:26th Annual Conference on Neural Information Processing Systems 2012.Lake Tahoe:Curran Associates Inc,2012,25:1106-1114.