基于深度学习的小目标检测方法综述

员娇娇, 胡永利, 孙艳丰, 尹宝才
(北京工业大学信息学部, 北京 100124)

摘 要: 小目标检测一直是目标检测领域中的热点和难点,其主要挑战是小目标像素少,难以提取有效的特征信息. 近年来,随着深度学习理论和技术的快速发展,基于深度学习的小目标检测取得了较大进展,研究者从网络结构、训练策略、数据处理等方面入手,提出了一系列用于提高小目标检测性能的方法. 该文对基于深度学习的小目标检测方法进行详细综述,按照方法原理将现有的小目标检测方法分为基于多尺度预测、基于数据增强技术、基于提高特征分辨率、基于上下文信息,以及基于新的主干网络和训练策略等5类方法,全面分析总结基于深度学习的小目标检测方法的研究现状和最新进展,对比分析这些方法的特点和性能,并介绍常用的小目标检测数据集. 在总体梳理小目标检测方法的研究进展的基础上,对未来的研究方向进行展望.

关键词: 深度学习; 目标检测; 小目标检测; 特征金字塔; 上下文; 数据增强
中图分类号: U 461; TP 308 **文献标志码**: A **文章编号**: 0254 - 0037(2021)03 - 0293 - 10
doi: 10.11936/bjutxb2020090019

Survey of Small Object Detection Methods Based on Deep Learning

YUAN Jiaojiao, HU Yongli, SUN Yanfeng, YIN Baocai
(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: Small object detection has always been a difficult problem in the research area of object detection. In recent years, with the rapid development of deep learning, the research on small object detection has made great progress. Researchers have studied and proposed a series of methods to improve the performance of small object detection from the aspects of network structure, training strategy and data processing. This paper provides a detailed overview of small object detection methods based on deep learning. According to the principle of the methods, the existing small object detection methods were divided into multi-scale prediction, data enhancement technology, feature resolution enhancement, context information, new backbone network and training strategy, a more detailed comparison of these methods was done, and the commonly used small object detection datasets were introduced. Finally, summary and prospect were made combined with the development of the existing technology of small object detection.

Key words: deep learning; object detection; small object detection; feature pyramid networks; context; data enhancement

随着深度学习的发展,基于深度学习的目标检测技术取得了巨大的进展,但小目标由于像素少,难

收稿日期: 2020-09-28
基金项目: 国家自然科学基金资助项目(61672071, U1811463, U19B2039)
作者简介: 员娇娇(1990—),女,博士研究生,主要从事目标检测、深度学习方面的研究,E-mail: yjj@emails.bjut.edu.cn

以提取有效信息,造成小目标的检测面临着巨大的困难和挑战. 为了提高小目标的检测性能,研究人员从网络结构、训练策略、数据处理等方面展开了大量的研究,并取得了一定的进展. 然而,与大、中目标检测相比,目前小目标的检测性能依然存在着较大的差距.

目标尺度是影响目标检测性能的重要因素之一. 目前,无论在公开数据集还是现实世界采集的图像中,小目标的检测精度远远低于大目标和中等尺度目标,并经常出现漏检和误检. 但小目标检测在许多实际场景中具有重要的应用,甚至是很多智能设备能否有效安全运行的关键所在. 例如,在无人驾驶系统中,当交通信号灯或行人等目标比较小时,仍然要求无人车能准确识别这些目标并做出相应的动作;在卫星图像的分析中,需要检测汽车、船舶等之类的目标,但这些目标往往由于尺度过小造成检测困难. 因此,研究小目标检测的有效方法、提高小目标的检测性能,是当前目标检测领域非常重要和迫切的研究课题.

小目标的定义主要有 2 种:第 1 种是绝对小物体,COCO 数据集中指明,当物体的像素点数小于 32 ×32 时,此物体即可被看作是小物体;第 2 种是相对小物体,当目标尺寸小于原图尺寸的 0.1 时可认为是相对小物体[1]. 在传统的基于机器学习的目标检测中,主要通过构建图像金字塔以求在金字塔的底部检测出小目标. 这种方式需要在不同分辨率的图像上分别提取特征,对于人工设计的特征,计算量尚在可接受范围内;但是对于深度学习提取的特征,这种方式会由于计算量大而无法满足实时性的要求.

随着深度学习的出现和发展,利用图像金字塔来检测不同尺度物体的方法逐渐被深度卷积神经网络(convolutional neural network, CNN)替代. 深度卷积神经网络通过对物体形成多层次的、丰富的特征表达,有效提高了不同尺度物体的检测性能. 在深度卷积网络中,底层特征含有丰富的细节信息,有利于小目标的检测;高层特征含有丰富的语义信息,有利于大目标的检测. 随着研究的不断深入,小目标的检测性能得到了较大的提升,但和大、中目标的检测性能相比仍然存在着一定差距.

关于小目标检测研究的进展,文献[2]较早进行了综述,对主流的方法和网络模型进行了分析对比. 文献[3]也从应用的角度对小目标检测的方法进行了讨论. 除此之外,国内相关学者也对小目标检测的研究现状进行了综述. 文献[1]按照网络结构将小目标检测技术分为一阶段、两阶段、多阶段共 3 种方法,并介绍了相关的小目标检测数据集;文献[4]介绍了使用多尺度预测和增强特征图的分辨率来提升小目标检测性能的方法;文献[5]介绍了一些基于深度学习的小目标检测模型和常用的小目标检测数据集. 然而,由于小目标检测研究进展很快,尤其基于深度学习的小目标检测新方法不断出现,现有的综述对一些新方法介绍不多,特别是对数据增强的小目标检测方法、利用上下文信息的小目标检测方法以及使用新主干网络和训练策略的小目标检测方法的讨论不够充分,例如文献[4]缺少对基于数据增强的小目标检测方法的介绍.

针对上述情况,为了更加清晰地阐述基于深度学习的小目标检测方法的研究思路,本文首先按照原理的不同将这些方法分成 5 类,介绍了每一类的典型模型,并对现有的方法进行了比较,然后介绍了小目标检测常用的数据集,最后结合当前小目标检测的研究现状给出了相应的结论和思考.

1 小目标检测方法

目前,基于深度学习的目标检测方法可分为 2 类,一类是两阶段的目标检测方法,即先生成候选区域,然后再对候选区域进行分类和回归,例如 Faster R-CNN[6];另一类是一阶段的目标检测方法,这类方法直接从图像中回归出物体的类别和坐标,无须生成候选框,代表性的方法有 YOLO[7]、SSD[8]等. 无论是一阶段的目标检测方法,还是两阶段的目标检测方法,都面临着小目标检测困难的情况. 具体地,小目标检测主要面临以下几个方面的挑战:

1) 底层特征缺乏语义信息. 在现有的目标检测模型中,一般使用主干网络的底层特征检测小目标,但底层特征缺乏语义信息,给小目标的检测带来了一定的困难.

2) 小目标的训练样本数据量较少. 目前,主流的目标检测算法广泛使用的数据集(PASCAL VOC、COCO)中小目标的训练样本较少,这种情况使得在模型训练的过程中小目标得不到充分的学习.

3) 检测模型使用的主干网络与检测任务的差异. 现有的目标检测模型的主干网络都是在分类数据集上进行训练的,但是分类数据集中目标的尺度分布与检测数据集中目标的尺度分布存在一定的差异.

现有的基于深度学习的小目标检测方法都是在主流的目标检测模型上做改进来提高小目标的检测性能. 按照改进思路的不同,小目标检测方法可分为基于多尺度预测、基于提高特征分辨率、基于上下文信息、基于数据增强技术、基于新的主干网络和训练策略共 5 种方法.

1.1　基于多尺度预测的小目标检测方法

多尺度预测指的是在多个不同尺度的特征图上分别对物体的类别和坐标进行预测. 在目标检测模型发展的早期,代表性的算法如 YOLO、Faster R-CNN 等,只使用主干网络的最后一层特征进行目标检测,造成对小目标的检测性能不够好;SSD 中首次采用了多尺度预测的方式,改善了小目标的检测性能. 目前,采用多尺度预测的方式已经成为提升小目标检测性能的基本操作.

1.1.1　基于图像金字塔的多尺度目标检测

在基于机器学习的目标检测阶段,图像金字塔是构建多尺度特征的主流方法,在 CNN 发展的早期,这种方法也得到了一定的应用. 该方法首先将图像缩放到不同分辨率,通过在不同分辨率的图像上分别提取特征来形成多尺度的表达,然后在每个分辨率图像上分别利用基于滑动窗口的方法进行目标检测,以求在金字塔底部检测出小目标. MTCNN[9] 就利用了这种思想,首先构建图像金字塔,然后在每层图像上利用 CNN 提取人脸特征,从而检测出不同分辨率的人脸目标. 这种方式在每一层分辨率图像上提取的特征都含有丰富的语义,有利于小目标的检测,但是由于需要对多种分辨率图像分别提取特征,严重增加了推理时间,限制了该方法在实时性要求比较高的条件下的应用.

1.1.2　DSSD 算法

随着深度学习的发展,利用 CNN 提取多尺度特征基本替代了图像金字塔的方式. 在多个不同尺度的特征图上分别进行预测有利于小目标检测性能的提升. SSD 通过在多个不同尺度的特征图上分别对目标的类别和坐标进行预测,在一定程度上提高了小目标的检测效果,但对小目标的检测仍然不理想. 在 DSSD[10] 中,作者认为造成上述现象的原因在于 SSD 中用于检测小目标的特征层含有的语义信息不丰富,较低的语义信息会造成一定的分类错误或置信度较低,给小目标的检测带来误检和漏检.

针对上述情况,DSSD 从两方面提出改进. 首先,使用 ResNet-101 代替 VGG16 作为提取特征的主干网络,前者相比于后者网络层次更深,特征表达能力更强;其次,将高层特征的语义信息融入底层特征,提高底层特征的表达能力,具体操作为:将 SSD 额外添加的卷积层中第 n 层的特征图进行反卷积扩大到和 $n-1$ 层同样的分辨率,然后再将扩大后的特征图和第 $n-1$ 层的特征图进行元素级别的乘积操作得到最终的用于检测的特征图,记为 $(n-1)'$,后续的多尺度预测在 $(n-1)'$ 上进行. 经过这样的操作,得到的底层特征相比于 SSD 中的底层特征具有更加丰富的语义信息,更利于小目标的检测.

1.1.3　特征金字塔(feature pyramid networks,FPN)算法

与 DSSD 的思想类似,文献[11]提出了基于 FPN 的方法来提升目标检测算法中底层特征的语义信息. FPN 算法的框架如图 1 所示,FPN 一共包括 2 个分支,自底向上的分支用于产生多尺度的特征,自顶向下的分支用于将高层含有的丰富语义信息传递到底层. 具体地,首先,高层特征进行 2 倍的上采样得到和相邻底层一样的分辨率,然后底层特征经过 1×1 的卷积和上采样之后的高层特征进行元素级别的相加后,再经过 3×3 的卷积得到最终的特征图. FPN 充分融合了高层特征和底层特征,使得用于检测的每一层特征都具有丰富的语义信息,利于小目标的检测. 目前,FPN 在基于深度学习的目标检测算法中得到了广泛的应用,这种结构是端到端可训练的,可无缝嵌入到现有的目标检测模型中,提高目标检测算法的性能. 通过将 FPN 嵌入到 Faster R-CNN 的区域候选网络(region proposal network,RPN),使得 Faster R-CNN 在 COCO 数据集上的小目标平均精度(average precision of small objects,APs)指标提高到了 17.5%,比之前的 COCO 数据集上的最优结果提升了 5% 左右. 目前,FPN 基本已成为目标检测算法的一个标准配置,有很多基于 FPN 的优化工作也相继涌现出来.

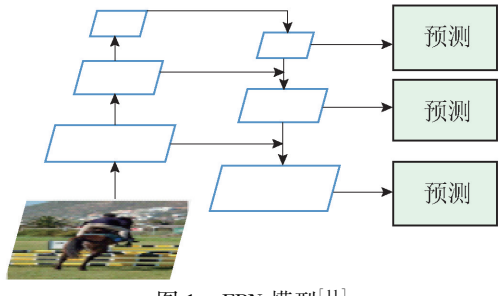

图 1　FPN 模型[11]

Fig. 1　Model of FPN[11]

1.1.4 PANet 算法

PANet[12] 在 FPN 的基础上进行了改进,更加充分地融合了高层特征和底层特征的信息,将其应用在目标检测和实例分割模型中,分别获得了 COCO 2017 目标检测算法的第二名和实例分割比赛的第一名. 该方法的结构如图 2 所示,作者在 FPN 的基础上又增加了一个自底向上的路径增强分支. 作者认为底层特征对于检测和分割至关重要,有助于进行更精确的定位. 但在 FPN 中,高层特征与低层特征之间路径较长(红色的虚线),造成在金字塔的顶部含有的底层信息较少. 为了解决这个问题,PANet 使用较少数量的卷积层构建了路径增强模块(见图 2(b)),尽可能多地保留底层信息;同时,又增加了自适应的特征池化模块,使得感兴趣区域(region of interest,ROI)中包含多层特征,而不是单层特征,进行了进一步的特征融合. 经过这样的操作,将 COCO 2017 目标检测比赛中的 AP 指标提高了 3 个百分点左右.

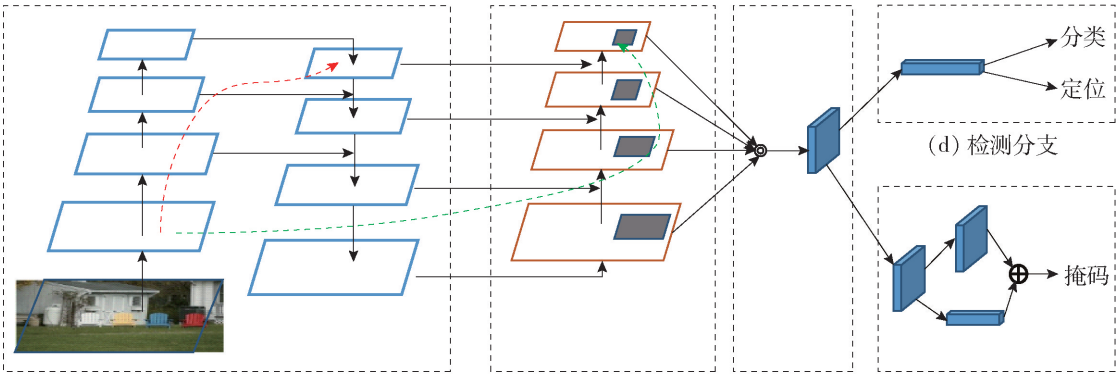

图 2 PANet 模型[12]

Fig. 2 Model of PANet[12]

1.1.5 ASFF 算法

FPN 的多层不同特征尺度之间存在不一致问题,即大特征图检测小目标,小特征图检测大目标. 当某个目标在某一层被当作正类时,在其他层可能会被当作负类,这样在特征金字塔的某一层单独检测时候就会引入其他层的矛盾信息. 为了解决上述问题,ASFF[13] 对 FPN 的特征融合方式做了改进,提出了一种自适应的空间特征融合方法. 该方法在 FPN 的基础上,通过学习权重参数的方式将不同层的特征融合到一起,得到融合之后的特征图用于最终的预测.

在论文中,作者将 ASFF 应用到 YOLO3 中,为了验证 ASFF 的有效性,首先在 YOLO3 应用了一系列的技巧,对 YOLO3 进行优化,将其在 COCO 2017 验证集上的 APs 指标由 18.3 提升到 24.6,将优化之后的 YOLO3 作为一个强的基线. 然后,在此基础上加入 ASFF, APs 指标由 24.6 提升到 27.5,提升了将近 3 个百分点,由此可见 ASFF 对于小目标检测的有效性.

1.1.6 Libra R-CNN 算法

关于如何更好地融合特征金字塔中的多尺度特征,Libra R-CNN[14] 给出了相应的优化方案. 改进的方法分别提取了 4 个级别的多尺度特征{C2,C3,C4,C5},然后将{C2,C3,C5}缩放到和 C4 同样大小,进行集成操作,也就是将这 4 个尺度的特征进行求和取平均得到集成之后的特征,再将得到的特征送入设计的增强模块中进行一个加强操作,最后再将加强后的特征和{C2,C3,C4,C5}相加,增强原特征. 这一操作将模型在 COCO 2017 验证集上的 APs 指标提高了 1.2 个百分点.

1.1.7 AugFPN 算法

经过对 FPN 的分析,AugFPN[15] 的作者认为 FPN 主要存在以下 3 个缺点:1) 特征融合前没有考虑不同层次的特征之间的语义差异;2) 在自上而下的特征融合过程中,高层特征存在丢失;3) 每层的 ROI 没有结合其他层次的有用信息. 对此,作者分别做出了改进,在特征融合之前,对每一层的特征都添加了相同的监督信息,减少了它们之间的语义差异;然后,对于最高层特征信息丢失的问题,采用了残差结构将其他层的特征加入到最高层特征中,增强它的上下文信息;最后,将候选框在不同层池化后的特征进行融合. 经过这样的操作,AugFPN 将

RetinaNet 在 COCO 2017 验证集上的 APs 指标提高了 2.7 个百分点.

1.1.8　SNIP 算法

当前的目标检测模型中,用于提取特征的主干网络都是在 ImageNet 数据集上预训练得到的. SNIP[16]认为 ImageNet 数据集中的目标尺度和检测用的 COCO 数据集中的目标尺度的分布相差较大,这种差异造成了小目标的检测性能不够好. 基于以上发现,作者提出了 SNIP 算法,该算法从两方面做出改进:在训练的过程中只对那些和 ImageNet 数据集中的目标尺度接近的 ROI 计算梯度,减少 ImageNet 数据集和 COCO 数据集中目标的尺度差异;利用图像金字塔得到多尺度的高分辨率目标信息. 该方法在 COCO 数据集上的 APs 指标达到了 31.4%,比之前的算法有较大提升,获得了 COCO 2017 挑战赛的最佳学生奖.

1.1.9　SNIPER 算法

SNIPER[17]是 SNIP 算法的改进版本,针对 SNIP 计算量大的问题,作者提出不再将整张图像作为网络的输入,而是从图像中得到低分辨率的图像块,将包含前景的图像块作为正样本,使用一个精度不是很高的 RPN 网络生成一些不太准确的候选框作为负样本的图像块. 将正负样本图像块作为网络的输入,进行多尺度的训练.

1.1.10　TridentNet[18]算法

针对目标检测中的尺度变化问题,TridentNet 从感受野的角度探讨了感受野对不同尺度物体检测的影响. 作者发现感受野和物体尺度呈正相关:感受野越大,对于大目标的检测就越好;感受野越小,对于小目标的检测就越好. 算法通过控制空洞卷积的参数来控制感受野的大小,生成了 3 个并行的卷积层,这 3 个卷积层有不同的感受野,用来检测不同尺度的目标,通过融合 3 个卷积层的优势来提高检测算法的性能.

1.2　基于提高特征分辨率的小目标检测方法

该方法的主要思想是:通过增大高层特征图的分辨率或通过生成对抗网络(generative adversarial network,GAN)的方式将小目标的特征表达转化为和大、中目标一样或近似的特征表达来提高小目标的检测精度.

1.2.1　STDN[19]算法

目标检测中,小目标的检测需要大分辨率的特征图来提供更精细的特征和更密集的采样,但是往往大分辨的特征图包含的语义信息不够充分. 为了解决上述问题,作者采用 DenseNet[20]作为提取特征的主干网络.

由于 DenseNet 每一层特征具有同样的分辨率,因此对浅层特征进行池化操作,起到扩大感受野的作用,用于检测大目标;对深层特征采用了一个尺度迁移模块,即:将特征图的分辨率按一定的比例放大,例如将 $H \times W \times C$ 的特征图变为 $RH \times RW \times (C/R^2)$,其中 H 代表特征图的高度,W 代表特征图的宽度,C 代表特征图的通道数,R 为特征图放大的比例. 显然,在深层特征上应用尺度迁移模块,可以保证特征图在包含足够语义信息的情况下扩大分辨率,从而提高小目标的检测效果.

1.2.2　PGAN 算法

PGAN[21]是第一个用 GAN 来提升小目标检测性能的算法,类似的工作还有 SOD-MTGAN[22]. 算法的主要思想是:鉴于小目标和大目标在 CNN 的高层生成的特征表达存在着明显的差异,作者希望通过 GAN 将小目标的特征表达转化为和大目标一样的超分辨特征表达,从而达到提升小目标检测性能的目的. 模型包含两部分:生成器和判别器. 在生成网络中,首先将第 1 层卷积之后的特征(该特征含有丰富的有利于小目标检测的底层信息)送入生成器中得到大目标和小目标之间的残差表示,用于增强 ROI 输出的小目标特征表达;在判别网络中共包含 2 个分支,一个是对抗分支,用于判别高分辨率的特征表达是来自生成的还是大目标的,另一个是用于检测的感知分支,用于判别小目标的检测精度是否从生成的高分辨特征中受益. 具体地,在训练的时候,首先用大目标的实例训练判别网络的 2 个分支,然后再用大目标和小目标的实例集合迭代训练生成器、判别器. 模型在 Tsinghua-Tencent 100K 和 Caltech 这 2 个小目标的数据集上进行了验证,可有效提升小目标的检测效果.

1.3　基于上下文信息的小目标检测方法

上下文信息指的是:在图像中,单个像素或单个目标并不是单独存在的,而是和周围像素、目标存在某种联系. 挖掘并利用物体与物体之间的关系即上下文信息将有利于小目标检测.

为了检测不同尺度的人脸,文献[23]同样引入了上下文信息. 首先,提出了一种基于先验框的上下文辅助方法,对于一个目标人脸,会存在一系列和人脸相关的先验框,这些先验框在感受野大的特征图上包含更多上下文信息,比如头、身体等,利用这些先验框作为辅助信息将有利于监督学习尺度小、

模糊和部分遮挡人脸的上下文特征的监督信息. 其次,文章还设计了一个上下文敏感的预测模块,该模块可在不同的特征上预测前景、背景,以及面部、头部和身体等. 该方法在 2018 年的 WIDER FACE 人脸检测比赛中获得冠军,算法重点解决难度检测大的人脸,尤其是小尺度人脸. 文献[24]认为在传统的 ROI 中每个目标都是单独进行检测的,没有考虑到目标之间的关系,但这种关系对于目标检测是有用的. 因此,作者提出了一种关系模块,用于提取不同物体之间的关联关系. 通过将每个物体的特征分为外形特征(大小、颜色、形状等)和几何特征(位置和大小等),每个关系模块都将所有前景目标的 2 个特征作为输入,得到不同物体之间的关系特征之后再做拼接,然后和物体原来的特征信息融合,作为物体检测的最后特征. 文献[25]提出了 2 种上下文信息来帮助提高目标检测:一种是图像级别的上下文信息,主要描述目标和整幅图像的关系;另一种是目标级别的上下文信息,主要描述目标与目标之间的关系. 通过在目标检测中引入这些上下文信息,将 Faster R-CNN 在 COCO 数据集上的小目标的平均召回率提升了 0.7 个百分点. CoupleNet[26] 通过将 ROI 对应的特征图往外扩大 1 倍的方式获取和物体相关的上下文信息,作者提出之所以这样做是因为尽管深度神经网络的高层由于感受野较大,可以涉及到物体周围的空间背景信息,但实际的感受野要比理论的感受野小得多. 因此,有必要明确地收集周围信息,以减少误识别的机会. 通过这样的操作,CoupleNet 在 COCO 2015 测试集上的 APs 指标比 R-FCN 提高了 2.6 个百分点.

1.4 基于数据增强技术的小目标检测方法

数据增强指的是通过重采样、旋转、平移等方式增加训练样本的数量供神经网络学习. 无论在公开数据集还是现实世界采集的数据集中,小目标的样本数量普遍较少. 因此,通过数据增强的手段增加小目标的样本数量将有助于提高检测性能.

采用数据增强策略是提高小目标检测效果的有效手段之一. 文献[27]提出了针对小目标的数据增强策略. 一方面对包含小目标的图像进行过采样,增强其数量;另一方面对图像中的小目标进行复制粘贴(不与其他图像中的其他目标重叠),增加单张图像中小目标的数量. 经过这些数据增强策略,将 Mask R-CNN[28] 在 COCO 数据集上小目标的检测指标提升了 7.1%. Zoph 等[29] 提出了一种基于神经网络搜索的目标检测数据增强算法. 在传统算法中,使用哪些增强算法,每个算法用几次以及其先后顺序都是人为定义的,这种人工设计的数据增强算法也许并不是最优的. 近年来,随着神经架构搜索(neural architecture search, NAS)[30] 的兴起,这种基于神经网络搜索的技术被用到越来越多的领域. Zoph 等将 NAS 应用到目标检测算法的数据增强领域,利用神经网络搜索出最优的数据增强策略. 作者首先定义了 22 个数据增强算法,包括对颜色、框位置变化、光照等方面进行增强,然后在这 22 个算法的基础上构建搜索空间. 作者将整个搜索空间离散化为 K 个子策略,每个子策略包含 N 个数据增强运算,每个运算又包含 2 个超参数:被用到的概率 P 和次数 M. 训练的时候作者选用 RetinaNet 作为基础检测模型,将文章提出的数据增强算法应用到 RetinaNet 网络上,从 COCO 数据集中抽样出数据量为 5 000 ~ 23 000 的小子集作为训练集,文中提出的方法在小目标上的平均精度得到了 1.3% ~ 2.8% 不同程度的提升.

1.5 基于新的主干网络和训练策略的小目标检测方法

目前,在目标检测模型中,主干网络都是在分类数据集上预训练得到的,然而分类数据集中目标尺度的分布与检测数据集中目标尺度的分布存在一定的差异,这就造成了小目标的检测性能不佳. 因此,有研究者提出设计专门的针对目标检测任务的主干网络和训练策略来提升小目标的检测性能.

He 等[31] 提出从零开始训练检测模型将有助于精确定位. 文献[32]提出一种新的目标检测训练方法和模型,将预训练模型与从零开始训练的方法结合起来,使用一个预训练的 SSD 网络作为主干网络,同时采用了一个从零开始训练的轻量级的辅助网络 LSN. LSN 的作用是弥补主干网络在提取特征过程中的损失,提取更加准确的中底层特征,为目标检测提供更加精细的轮廓信息. 具体地,对于输入图像,首先采用一个大的下采样网络将图片大小调整到 SSD 中第 1 层的输入大小,然后使用设计的卷积网络提取特征,需要强调的是,LSN 的参数是随机初始化的. 同时,针对 FPN 中信息只能从高层往底层单向传播的情况,文中提出了一种双向 FPN,实现了底层特征和高层特征的双向传播. 在 COCO 数据集上,该方法在和 SSD 同样使用 VGG16 作为主干网络的情况下,与 SSD 相比,将 APs 的指标提高了 2 倍以上. 文献[33]针对现有的主干网络与目标检测任务之间的矛盾,设计了一种专门用于目标检测的

主干网络，根据目标检测任务的特点，通过精心设计网络中用于预测的特征层的数量，同时兼顾空间分辨率和感受野，相比于 ResNet-50，文中提出的 DetNet-59 网络在发现小目标方面更为强大，在交并比（intersection over union，IOU）为 0.5 的情况下，小目标的平均召回率提高了 6 个百分点. 文献[34]中，作者借鉴了 Faster R-CNN 和 Cascade R-CNN 的做法，在主干网络的设计中也应用了将信息重复利用的思想，提出了 DetectoRS 算法. 首先，算法将 FPN 层的信息反馈到自下而上的主干网络，这样递归的结构相当于将图像信息重复使用了 2 遍；其次，引入了可切换的空洞卷积，经过这样的操作，目标检测的性能得到了极大的提升.

2　方法对比

前面详细介绍了现有的提高小目标检测性能的方法，本节对一些性能表现突出的小目标检测算法进行对比，结果如表 1 所示. 评价的指标是 APs，即在小目标上的平均精度；表中显示的结果是算法在 COCO 数据上的 APs 指标情况. 从中可以看出，目前主干网络结合特征金字塔的结构已经成为主流的框架，基于特征金字塔的网络模型注重通过增强底层特征的语义信息来提高小目标的检测性能，对特征金字塔中特征融合的方式不断优化，不断提高特征的表达能力将有利于小目标检测性能的提升，PANet 和 DetectoRS 的结果明确地验证了这一点. 此外，DetectoRS 的结果也说明了对信息重复利用的重要性. 从 TridentNet 的结果可以发现，感受野和目标尺度之间有着密切的关系，怎样合理利用感受野来优化小目标的检测将大大提升小目标检测的效果. 除了对网络结构做改变之外，数据增强技术对小目标的检测同样重要.

表 1　小目标检测算法对比

Table 1　Comparison of small object detection algorithms

算法	小目标检测方法					主干网络			APs
	多尺度预测	数据增强技术	提高特征分辨率	基于上下文信息	新的主干网络和训练策略	ResNet-101	ResNeXt-101	DetNet-59	
SSD513[8]	√	√				√			10.2
DSSD513[10]	√		√			√			13.0
FPN[11]	√					√			18.2
PANet[12]	√						√		30.1
Libra R-CNN[14]	√						√		25.3
SNIP[16]	√					√			27.3
SNIPER[17]	√					√			29.6
TridentNet[18]	√					√			31.8
CoupleNet[16]				√		√			13.4
DetNet[33]					√			√	23.6
DetectoRS[34]	√						√		37.4

3　小目标检测的数据集

当前，基于深度学习的目标检测算法都是基于数据驱动的，然而在早期的目标检测数据集中，大、中目标的数量远远高于小目标的数量，这也是限制小目标检测技术发展的一个重要因素. 为了促进小目标检测技术的发展，近年来，很多针对小目标检测的数据集相继被提出和发布.

1）COCO[35]：常用的目标检测数据集，包含了大量的小目标. 一共包含了 91 类目标，有 328 000 张图像和 250 万个标注框.

2）Tsinghua-Tencent100K[36]：一个大型交通标志数据集，提供了 10 万张图像，包含了 3 万个交通标志实例.

3）Tiny Person[37]：中国科学院大学提出的数据集，只包含人这一个类别. 其中，训练集包含 794 张

图像,测试集包含 816 张图像.

4) DOTA[38]:遥感图像目标检测领域的数据集,一共包含 15 个种类,共 2 806 张图像.

5) UCAS-AOD[39]:遥感图像目标检测数据集,只包含汽车、飞机 2 类目标. 其中,共有飞机样本 7 482 个,汽车样本 7 114 个.

6) NWPU VHR-10[40]:西北工业大学标注的航天遥感目标检测数据集,包括飞机、舰船、车辆等 10 个类别. 该数据集共有 800 张图像,其中包含目标的图像有 650 张,背景图 150 张.

7) RSOD-Dataset[41]:武汉大学标注的航空遥感图像,包括飞机、操场、立交桥、油桶 4 类目标. 其中,飞机类有 446 张图像,操场类有 189 张图像,立交桥类有 176 张图像,油桶类有 165 张图像.

8) INRIA aerial image dataset[42]:一个专用于城市建筑物检测的遥感图像数据集,训练集包含 180 张图像.

9) URPC 2018[43]:水下图像,包含大量的小目标,类别包括海参、扇贝、海胆、海星. 该数据集共包含 2 897 张训练图像和 797 张测试图像.

4 结论与展望

目前基于深度学习的小目标检测研究的核心问题是,如何提高小目标的特征表达使其含有丰富的语义信息,这也是提升小目标检测性能的关键. 现有的大部分研究工作也是围绕小目标的特征表达展开的,包括前文综述的基于多尺度预测的方法、基于上下文信息的方法和基于提高特征分辨率的方法. 其中,基于多尺度预测的方法研究取得了很大进展,该类方法主要以 FPN 网络模型为基础进行优化和改进,将高层特征含有的丰富语义信息充分融合到底层特征中以提高小目标的特征表达. 该类方法在 COCO 数据集上取得了小目标检测的最优结果.

然而对比大、中目标的检测性能,目前小目标检测的性能依然存在很大的差距,未来如何探索更优的、更充分的多尺度特征融合策略来进一步提升小目标的检测性能,还有很多问题需要解决. 本文结合现有的小目标检测方法,对未来的几个有潜力的研究方向进行展望:

1) 基于 FPN 优化的多尺度预测. 目前基于 FPN 的多尺度预测框架已经成为主流,很多研究者对 FPN 中的特征融合的方式进行了深入的探索和优化,使得小目标的精度得到了一定的提升. DetectoRS 就是在 FPN 的基础上改进并在 COCO 数据集上获得最高 APs 指标的检测算法. 未来,如何提出更优的、基于 FPN 的特征融合方式,充分挖掘并利用不同层次的特征,将对小目标检测的提升有帮助.

2) 从头开始训练网络的方法. He 等[31]提出了从头开始训练检测网络将有利于减少检测和分类任务之间的差异,开辟了新的道路和方向. 目前,这方面的工作还比较少,但是还有很多工作值得探索.

3) 探索感受野对小目标检测的影响. 目标的尺度和感受野呈正相关的关系,如何设置合理的感受野用于检测不同大小的目标是一个有意思的方向. TridentNet 在 COCO 数据集上取得的效果充分证明了这种方式的有效性.

参考文献:

[1] 刘颖, 刘红燕, 范九伦, 等. 基于深度学习的小目标检测研究与应用综述[J]. 电子学报, 2020, 48(3): 590-601.
LIU Y, LIU H Y, FAN J L, et al. Overview of research and application of small target detection based on deep learning[J]. Chinese Journal of Electronics, 2020, 48(3): 590-601. (in Chinese)

[2] LIU L, OUYANG W L, WANG X G, et al. Deep learning for generic object detection: a survey[J]. International Journal of Computer Vision, 2020, 128(2): 261-318.

[3] ZHAO Z Q, ZHENG P, XU S T, et al. Object detection with deep learning: a review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232.

[4] 张新, 郭福亮, 梁英杰, 等. 基于深度学习的小目标检测算法综述[J]. 软件导刊, 2020, 211(5): 282-286.
ZHANG X, GUO F L, LIANG Y J, et al. A review of small target detection algorithms based on deep learning[J]. Software Guide, 2020, 211(5): 282-286. (in Chinese)

[5] 刘晓楠, 王正平, 贺云涛, 等. 基于深度学习的小目标检测研究综述[J]. 战术导弹技术, 2019, 193(1): 106-113.
LIU X N, WANG Z P, HE Y T, et al. Review of small object detection based on deep learning[J]. Tactical Missile Technology, 2019, 193(1): 106-113. (in Chinese)

[6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only

look once: unified, real-time object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 779-788.

[8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // European Conference on Computer Vision. Zurich: ECVA, 2016: 21-37.

[9] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23 (10): 1499-1503.

[10] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector[J]. ArXiv Preprint ArXiv, 2017: 1701. 06659.

[11] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 936-944.

[12] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8759-8768.

[13] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection[J]. ArXiv Preprint ArXiv, 2019: 1911. 09516.

[14] PANG J, CHEN K, SHI J, et al. Libra R-CNN: towards balanced learning for object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 821-830.

[15] GUO C, FAN B, ZHANG Q, et al. AugFPN: improving multi-scale feature learning for object detection[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 12592-12601.

[16] SINGH B, DAVIS L S. An analysis of scale invariance in object detection- SNIP [C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3578-3587.

[17] SINGH B, NAJIBI M, DAVIS L S. SNIPER: efficient multi-scale training[C] // Conference and Workshop on Neural Information Processing Systems. Nevada: Massaehusetts Institute of Technology Press, 2018: 9333-9343.

[18] LI Y, CEHN Y, WANG N, et al. Scale-aware trident networks for object detection [C] // International Conference on Computer Vision. Piscataway: IEEE, 2019: 6053-6062.

[19] ZHOU P, NI B B, GENG C, et al. STDN: scale-transferrable object detection[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 528-537.

[20] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2261-2269.

[21] LI J, LIANG X, WEI Y, et al. Perceptual generative adversarial networks for small object detection[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1951-1959.

[22] BAI Y C, ZAHNG Y Q, DING M L. SOD-MTGAN: small object detection via multi-task generative adversarial network[C] // European Conference on Computer Vision. Zurich: ECVA, 2018: 210-226.

[23] TANG X, DU D K, HE Z, et al. PyramidBox: a context-assisted single shot face detector [C] // European Conference on Computer Vision. Zurich: ECVA, 2018: 812-828.

[24] HU H, GU J, ZHANG Z, et al. Relation networks for object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3588-3597.

[25] CHEN X, GUPTA A. Spatial memory for context reasoning in object detection [C] // International Conference on Computer Vision. Piscataway: IEEE, 2017: 4106-4116.

[26] ZHU Y, ZHAO C, WANG J, et al. CoupleNet: coupling global structure with local parts for object detection[C] // International Conference on Computer Vision. Piscataway: IEEE, 2017: 4146-4154.

[27] KISANTAL M, WOJNA Z, MURAWSKI J, et al. Augmentation for small object detection [J]. ArXiv Preprint ArXiv, 2019: 1902. 07296.

[28] HE K M, GEOGIA G, PIOTR D, et al. Mask R-CNN [C] // International Conference on Computer Vision. Piscataway: IEEE, 2017: 2980-2988.

[29] ZOPH B, CUBUK E D, GHIASI G, et al. Learning data augmentation strategies for object detection[J]. ArXiv Preprint ArXiv, 2019: 1906. 11172.

[30] WANG N, GAO Y, CEHN H, et al. NAS-FCOS: fast neural architecture search for object detection[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 11940-11948.

[31] HE K M, GIRSHICK R, DOLLAR P. Rethinking ImageNet pre-training[C] // International Conference on Computer Vision. Piscataway: IEEE, 2019: 4917-4926.

[32] WANG T, ANWER R M, CHOLAKKAL H, et al. Learning rich features at high-speed for single-shot object detection [C] // International Conference on Computer

Vision. Piscataway: IEEE, 2019: 1971-1980.
[33] LI Z M, PENG C, YU G, DetNet: a backbone network for object detection[C]// ArXiv Preprint ArXiv, 2018: 1804. 06215.
[34] QIAO S, CEHN L C, YUILLE A. DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution[J]. ArXiv Preprint ArXiv, 2020: 2006. 02334.
[35] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]// European Conference on Computer Vision. Zurich: ECVA, 2014: 740-755.
[36] ZHU Z, LIANG D, ZHANG S, et al. Traffic-sign detection and classification in the wild[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 2110-2118.
[37] YU X, GHONG Y, JIANG N, et al. Scale match for tiny person detection[C]//Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2020: 1246-1254.
[38] XIA G S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3974-3983.
[39] ZHU H, CEHN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]// International Conference on Image Processing. Piscataway: IEEE, 2015: 3735-3739.
[40] CHENG G, HAN J W, ZHOU P C, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98(1): 119-132.
[41] XIAO Z F, LIU Q, TANG G F, et al. Elliptic fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. International Journal of Remote Sensing, 2015, 36(2): 618 -644.
[42] MAGGIORI E, TARABALKA Y, CHARPIAT G, et al. Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark[C]// IEEE International Geoscience & Remote Sensing Symposium. Piscataway: IEEE, 2017: 3226-3229.
[43] ZHANG L, YANG X, LIU Z, et al. Single shot feature aggregation network for underwater object detection[C]// International Conference on Pattern Recognition. Piscataway: IEEE, 2018: 1906-1911.

(责任编辑 吕小红)

(上接第 215 页)

[10] SHIMADA K, MATSUO Y. A new float-polishing technique with large clearance untillsing magnetic compound fluid[J]. International Journal of Abrasive Technology, 2008, 1(3/4): 302-305.
[11] WU Y, SATO T, LIN W, et al. Mirror surface finishing of acrylic resin using MCF-based polishing liquid[J]. Int J Abras Technol, 2010, 3(1): 11-24.
[12] GUO H, WU Y. Behaviors of MCF (magnetic compound fluid) slurry and its mechanical characteristics: normal and shearing forces under a dynamic magnetic field[J]. J Jpn Soc Exp Mech, 2012, 12(4): 369-374.
[13] WU Y, WANG Y, MASAKAZU F. Nano-precision polishing of CVD SiC using MCF (magnetic compound fluid) slurry[J]. J Korean Soc Manuf Technol Eng, 2014, 23(6): 547-554.
[14] DEGROOTE J E, MARINO A E, WILSON J P, et al. Removal rate model for magnetorheological finishing of glass[J]. Applied Optics, 2007, 46(32): 7927-41.
[15] SHIMADA K, FUJITA T, OKA H, et al. Hydrodynamic and magnetized characteristics of MCF (magnetic compound fluid)[J]. Transactions of the Japan Society of Mechanical Engineers, 2001, 67(664): 3034-3040.
[16] WU Y, SATO T, LIN W, et al. MCF (magnetic compound fluid) polishing process for free-formed resin device using robotic arm[C]// International Conference on Advances in Materials and Processing Technologies. Paris: American Institute of Physics, 2011: 979-984.
[17] GUO H R, WU Y B, LU D, et al. Effects of pressure and shear stress on material removal rate in ultra-fine polishing of optical glass with magnetic compound fluid slurry[J]. Journal of Materials Processing Technology, 2014, 214(11): 2759-2769.

(责任编辑 张 蕾)