

基于改进 CenterNet 少样本条件的增量化目标检测算法

李 航, 贾 可, 金治成, 李涵鑫, 许昌源

(成都信息工程大学 计算机学院, 四川 成都 610000)

摘 要: 深度学习模型需要大量标注样本进行训练,而在实际场景中,往往很难获得大规模且高质量的标注样本,而没有足够多的训练样本会导致模型性能不佳。目前面向少样本条件下的目标检测算法通常由于训练样本不足会出现漏检和错检的情况,不能有效生成目标的检测框。针对以上问题,以 CenterNet 方法为框架,在 ONCE 网络模型的基础上提出一种面向少样本条件的增量化目标检测算法。该检测器主要由结合注意力机制与特征融合的特征提取器、分解注意力编码器、自适应高斯核以及定位器组成,采用两阶段训练策略,通过增量化策略注册新类。其中,结合注意力机制与特征融合的特征提取器能增强对基类和新类特征的提取和利用能力;自适应高斯核与分解注意力方法能使检测器获得更准确的目标中心点,从而使生成的目标框更加精准。在 COCO 数据集和 PASCAL VOC 数据集上的验证实验结果表明,与其他先进方法相比,所提方法在 COCO 数据集上相较 ONCE 模型在新类、基类、全类的 AP 值分别平均提高了 7.7、2.5 和 1.3,AR 值分别平均提高了 15.0、21.7 和 17.6;在 PASCAL VOC 数据集上相较其他方法的 AP 值和 AR 值分别平均提高了 2.4 和 21.2,具有更好的检测性能。

关键词: 注意力机制;特征融合;自适应高斯核;分解注意力

DOI: 10.11907/rjdk.241203

中图分类号: TP391

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)005-0200-07



An Incremental Object Detection Algorithm Based on Enhanced CenterNet under Few-shot Scenarios

LI Hang, JIA Ke, JIN Zhicheng, LI Hanxin, XU Changyuan

(School of Computer Science, Chengdu University of Information Technology, Chengdu 610000, China)

Abstract: Deep learning models require a large number of annotated samples for training, but in practical scenarios, it is often difficult to obtain large-scale and high-quality annotated samples, and not having enough training samples can lead to poor model performance. At present, object detection algorithms targeting few sample conditions often suffer from missed and false detections due to insufficient training samples, which cannot effectively generate detection boxes for the target. To address the above issues, a few sample condition oriented incremental object detection algorithm is proposed based on the ONCE network model using the CenterNet method as a framework. The detector mainly consists of a feature extractor that combines attention mechanism and feature fusion, a decomposed attention encoder, an adaptive Gaussian kernel, and a locator. It adopts a two-stage training strategy and registers new classes through an incremental strategy. Among them, the feature extractor that combines attention mechanism and feature fusion can enhance the ability to extract and utilize features of base classes and new classes; The adaptive Gaussian kernel and decomposition attention method can enable the detector to obtain more accurate target center points, thereby making the generated target boxes more precise. The validation experiments on the COCO dataset and PASCAL VOC dataset showed that compared with other advanced methods, the proposed method improved the AP values of the new class, base class, and full class

收稿日期:2024-03-04

扫描二维码阅读全文:



基金项目:四川省科技计划项目(2023YFG0305);成都信息工程大学科研基金项目(KYTZ202156)

作者简介: 李航(1998-),男,成都信息工程大学计算机学院硕士研究生,研究方向为机器学习;贾可(1983-),男,博士,成都信息工程大学计算机学院副教授、硕士生导师,研究方向为人工智能、图像处理;金治成(1998-),男,成都信息工程大学计算机学院硕士研究生,研究方向为机器学习;李涵鑫(1999-),男,成都信息工程大学计算机学院硕士研究生,研究方向为机器学习;许昌源(2000-),男,成都信息工程大学计算机学院硕士研究生,研究方向为机器学习。本文通讯作者:贾可。

by an average of 7.7, 2.5, and 1.3, respectively, and the AR values by an average of 15.0, 21.7, and 17.6, respectively, on the COCO dataset compared to the ONCE model; On the PASCAL VOC dataset, the average AP and AR values were improved by 2.4 and 21.2, respectively, compared to other methods, indicating better detection performance.

Key Words: self-attention mechanism; feature fusion; adaptive Gaussian kernel; decomposed attention

0 引言

目标检测一直以来都是计算机视觉领域研究的热点之一。目前的主流目标检测算法均基于深度学习模型,包括双阶段(Two-Stage)和单阶段(One-Stage)两大类。双阶段目标检测算法首先通过选择性搜索等算法对输入图像进行候选区域选择,然后对候选区域进行分类和定位,从而得到检测结果^[1]。典型的双阶段目标检测算法包括R-CNN (Region-based Convolutional Neural Networks)^[2]、Fast R-CNN^[3]、Faster R-CNN^[4]、SPP-Net^[5]等。在Faster R-CNN的基础上,Cai等^[6]提出Cascade R-CNN,He等^[7]提出Mask R-CNN。单阶段目标检测算法中没有候选区域选择这一阶段,直接对目标进行分类和定位,代表性算法包括YOLO系列^[8]、SSD^[9]、CornerNet^[10]等。其中,YOLO系列算法的独特之处在于通过单个神经网络模型实现目标检测和定位,具有快速、简单、端到端的特点。YOLOv1将目标检测视为一个回归问题,通过单个神经网络模型同时预测目标的类别和位置信息,具有较高的检测速度,但在定位精度和小目标检测精度方面存在一定缺陷。YOLOv2在一定程度上改善了对小目标的检测效果,但在处理密集目标时仍有不足。YOLOv3由于引入了一些技术和优化策略,导致模型的计算复杂度有所增加,检测速度有所下降。YOLOv4引入了更多技术和优化策略,进一步提升了检测精度,但模型的训练和部署复杂度有所增加,需要更多计算资源支持。

增量式目标检测任务是指在模型已经学习了一定数量的目标类别后,能够逐步增加新的目标类别,同时保持对原有类别检测能力的任务。这种学习方法允许模型在不断变化的环境中逐步学习新知识,而不会忘记旧知识,从而适应新的检测需求。该任务的主要挑战在于防止“灾难性遗忘”,即当模型使用新数据进行训练时,如果不采取特殊措施,模型在原有类别上的表现会显著下降。因此,少样本条件下的增量化目标检测目的是模型可以通过极少的新类样本来添加新的可检测类别,并且模型对新类和基类的检测准确率都尽可能要高。为此,Perez-Rua等^[11]提出ONCE (Open-ended CenterNet)检测框架,首先使用CenterNet算法基于基础数据训练通用类别特征提取模块,然后加入一个采用元学习方法训练的类别编码生成器,通过反向传播的方式注册新的类别;Joseph等^[12]提出一个新的计算机视觉问题——开放世界目标检测(Towards Open World Object Detection),模型的任务是在没有明确的监督下能够将没有见过的目标识别为未知类,当逐步获得对应

的标签后能够渐进地学习它们,且不遗忘以前学过的那些类别。目前,面向少样本条件的增量化目标检测算法通常为双阶段式,其复杂的设计和基于非极大值抑制的后处理机制很难灵活检测新类。此外,由于检测器对样本特征提取能力不足,一些有用的特征信息在提取过程中产生损耗,模型会出现错检和漏检的情况,而且最终生成的检测框效果亦不佳。

为解决上述问题,本文以单阶段目标检测算法CenterNet为框架,在ONCE网络模型的基础上提出一种面向少样本条件的增量化目标检测算法,通过增量化策略注册新类。该算法主要由结合注意力机制与特征融合的特征提取器、注意力编码器和定位器组成,采用两阶段训练策略。其中,特征提取器下采样部分在ResNet50主干网络的每个残差块中添加通道注意力和空间注意力机制,以提高模型对特征图上关键点的表达能力。为有效降低特征在采样过程中的损失,充分利用不同特征层的信息,本文使用特征金字塔网络融合下采样和上采样过程中的不同层次特征,增强模型对基类和新类特征的提取与利用能力。为更好地预测中心点和目标框大小偏移量,本文使用两个注意力编码器分别用于定位和尺寸回归,对应任务的特征图和代码进行卷积操作后送入定位器,一个用于生成热力图确定中心点和类别,一个用于预测中心点和目标框大小偏移量,从而有效提升目标框生成质量。针对定位问题,本文在原有框架的基础上优化高斯核,考虑到物体的长宽因素,采用自适应形状的高斯核进行编码,使检测器获得更准确的目标中心点。与对照方法的比较实验结果表明,所提算法在新类检测准确率方面明显提升。

1 模型构建

所提模型采用两阶段训练策略。第一阶段主要是在基类数据集上训练两个特征提取器,分别得到用于中心点定位的特征图和尺寸回归的特征图;第二阶段主要是在新类数据集上训练注意力编码器和定位器。定位器由简洁轻量化的卷积神经网络组成,并且在网络后部添加3个不同的预测分支头,分别生成预测热力图(其峰值为目标中心点)、中心点坐标和中心点偏移量,并对目标进行定位。模型整体结构如图1所示。

1.1 结合注意力机制与特征融合的特征提取器

标准的CenterNet网络使用ResNet作为Backbone提取特征,由于训练样本不足,经常会出现错检和漏检的情况。为解决这一问题,本文使用嵌入注意力模块的残差网络使特征提取器能更好地提取样本中的特征信息,并对不同层

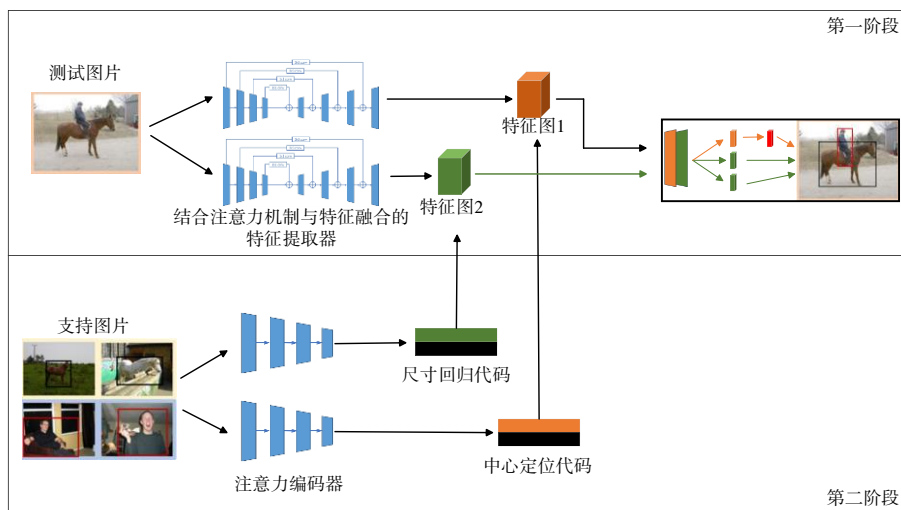


Fig. 1 Overall structure of the model

图1 模型整体结构

次特征进行融合复用,以减少训练过程中的特征损耗。结合注意力机制与特征融合的特征提取器主要分为下采样和上采样两部分。下采样部分使用 ResNet-50 作为主干网络,并且在每个残差块中嵌入卷积注意力模块 (Convolutional Block Attention Module, CBAM)^[13-14]。结构如图 2 所

示。在每个残差块中,将通道注意力模块和空间注意力模块依次串联在 1×1 卷积之后。同时借鉴特征金字塔网络的思想,对下采样和上采样过程中提取的不同层次特征进行融合,充分利用不同特征层的信息,有效降低了采样过程中的特征损失^[15]。

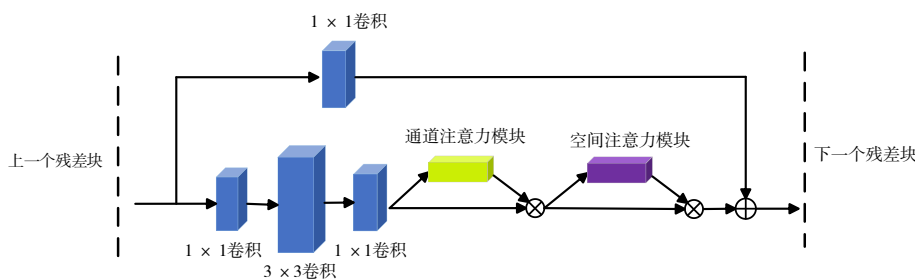


Fig. 2 Residual block embedded with CBAM structure

图2 嵌入CBAM的残差块结构

1.2 自适应高斯核

在标准的 CenterNet^[16]训练过程中,圆形的高斯核将中心点编码到热图上,能够在目标中心附近产生较高的激活值,但是目标中心点附近的激活值大致相同,预测中心点在 x 轴和 y 轴的偏移量也相同。然而,被检测目标通常不是正方形,长和宽往往不一样,甚至有些目标的长宽差距很大,那么预测中心点在 x 轴和 y 轴的偏移量也应因对应长宽的不同而不同。为解决这一问题,本文采用自适应高斯核替代 CenterNet 中的圆形高斯核,充分考虑长宽因素,生成与目标物体大小形状一致的中心点热力图,从而获得更准确的中心点位置。两种高斯核编码示意如图 3 所示。

1.3 分解注意力编码器

在 ONCE 的增量化目标检测过程中,第二阶段使用类别代码生成器分别生成中心点定位代码和尺寸回归代码,对应代码与特征提取器所得到的特征图进行卷积操作用于生成检测结果。该类代码需要在一张特征图上激活不同区域用于后续下游任务。然而,一张特征图包含的信息

有限,类代码需要对应多个下游任务,在特征图上所注意的区域也不同,导致分配给不同任务的注意程度不同,影响后续任务性能。为此,本文提出将检测任务分解为中心点定位和尺寸回归两个部分,针对不同任务分别建立任务通道。同时采用两个注意力编码器分别生成中心点定位代码和尺寸回归代码,对应任务的特征图和代码进行卷积操作后送入定位器,一个用于生成热力图确定中心点和类别,一个用于预测中心点偏移量和目标框大小偏移量。注意力编码器结构与结合注意力机制与特征融合的特征提取器的下采样部分完全一致。

1.4 元学习策略

模型使用两阶段元学习策略。第一阶段是在具有丰富训练数据的基类上执行常规监督学习,主要用于训练两个结合注意力机制与特征融合的特征提取器,学习所得的两个特征提取器下采样部分权重参数分别用于初始化第二阶段对应任务的注意力编码器;第二阶段采用训练数据模拟少样本场景下的任务,用于训练产生中心定位代码和尺寸回归代码的两个注意力编码器和定位器。

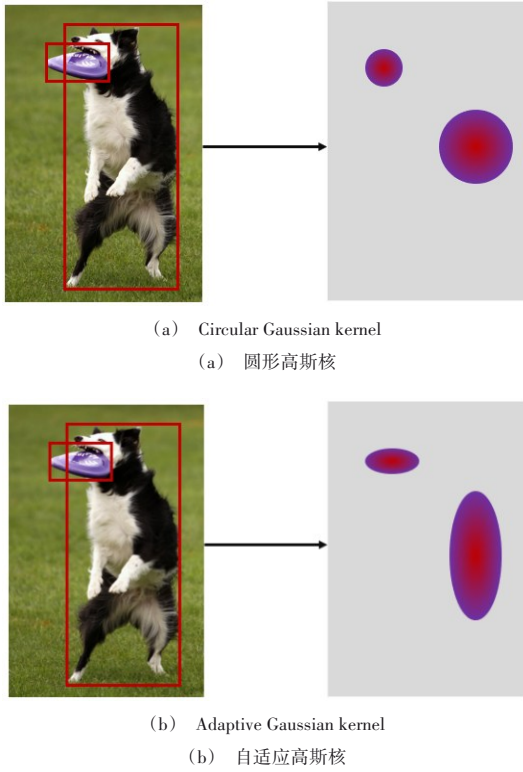


Fig. 3 Two kinds of Gaussian Kernel Coding

图3 两种高斯核编码示意

1.4.1 第一阶段训练

设输入图像 I 的高度和宽度分别为 h 和 w , 则 $I \in R^{h \times w \times 3}$. 将图片 I 分别送入两个结合注意力机制与特征融合的特征提取器 f_1 和 f_2 中, 得到特征图 m_1 和 m_2 , 即 $f_1(I) = m_1, f_2(I) = m_2$, 且 $m_1, m_2 \in R^{r \times \frac{h}{r} \times \frac{w}{r} \times c}$, 其中 r 为输出步长, 本实验中 $r=4$; c 为关键点个数, 即检测类别个数, 第一阶段取 $c=60$. 将通过学习得到的中心定位卷积核 $c_{1,k} \in R^{1 \times 1 \times c}$ 和尺寸回归卷积核 $c_{2,k} \in R^{1 \times 1 \times c}$ 与特征图 m_1 和 m_2 进行卷积运算后得到 k_1 和 k_2 . 计算公式为:

$$m_1 \odot c_1 = k_1, m_2 \odot c_2 = k_2 \quad (1)$$

式中: $\odot$ 表示卷积运算。

将 k_1 送入定位器, 生成预测热力图 $\hat{Y}_k \in [0, 1]^{\frac{h}{r} \times \frac{w}{r}}$ 对中心点定位, 其中 k 表示每个被检测的种类, $k \in \{1, 2, \dots, 60\}$. 当 $\hat{Y}_{x,y,c} = 1$ 时, 表示检测到特定类别的关键点; 当 $\hat{Y}_{x,y,c} = 0$ 时, 表示没有检测到关键点。将 k_2 送入定位器用于回归偏移量。

对于每个类别的真实关键点 $p = ((x_1 + x_2)/2, (y_1 + y_2)/2)$, 通过下采样计算一个低分辨率等效点 $\tilde{p}$. 采用自适应高斯核 Y_{xye} 将真实关键点标记到热力图 $Y_k \in [0, 1]^{\frac{h}{r} \times \frac{w}{r} \times c}$ 上。高斯核计算公式为:

$$Y_{xye} = \exp \left(-\frac{(x - \tilde{p}_x)^2}{2\sigma_x^2} - \frac{(y - \tilde{p}_y)^2}{2\sigma_y^2} \right) \quad (2)$$

式中: $\sigma_x = \alpha w, \sigma_y = \alpha h$, w 和 h 分别为目标物体的长和宽, α 为超参数, 用于限制得到的热力图大小。如果同一类的两个高斯分布发生重叠, 则选择元素最大的那个。

第一阶段训练总损失主要由3个部分组成, 分别为关键点预测损失 L_k 、中心点偏移损失 L_{off} 和目标框大小损失 L_{size} . 表示为:

$$L_{det} = L_k + \lambda_{off} L_{off} + \lambda_{size} L_{size} \quad (3)$$

式中: λ_{off} 和 λ_{size} 为超参数, 本实验中分别设定为1和0.1。

1.4.2 第二阶段训练

在训练过程中, 查询集中的图片 I 输入到第一阶段训练好的特征提取器 f_1 和 f_2 中, 分别得到特征图 m_1 和 m_2 . 支持集中的图片则输入到两个注意力编码器中生成中心定位代码 $\tilde{c}_{1,k}$ 和尺寸回归代码 $\tilde{c}_{2,k}$, 对于不同类别生成两个集合 $\{\tilde{c}_{1,k}\}$ 和 $\{\tilde{c}_{2,k}\}$. 将中心定位卷积核 $c_{1,k} \in R^{1 \times 1 \times c}$ 和尺寸回归卷积核 $c_{2,k} \in R^{1 \times 1 \times c}$ 与特征图 m_1 和 m_2 进行卷积运算后得到 k_1 和 k_2 , k_1 送入定位器通过热力图预测分支头生成对应类别的热力图 $\tilde{Y}$, k_2 送入定位器用于回归偏移量。

第二阶段训练总损失由3个部分组成, 其中中心点偏移损失 L_{off} 和目标框大小损失 L_{size} 的计算公式和超参数设定与第一阶段一致, 同时使用热力图预测损失代替第一阶段的关键点预测损失。热力图预测损失计算公式为:

$$L_{hm} = |\tilde{Y} - Z| \quad (4)$$

式中: $\tilde{Y}$ 表示预测热力图; Z 表示采用自适应形状高斯核生成的更精准的真实热力图。

2 实验方法与结果分析

2.1 数据集

COCO 数据集是微软公司发布的一个大型图像数据集, 专门为目标检测、语义分割而设计。该数据集拥有超过 330 K 张图片, 其中超过 200 K 张图片具有标注, 共包含 150 万个目标、80 个目标种类和 91 个材料种类^[17]。

PASCAL VOC 数据集是在 PASCAL VOC 挑战赛中使用的数据集, 包含 20 个目标种类, 分别为人、自行车、汽车、摩托车、飞机、巴士、火车、船、鸟、猫、狗、马、羊、牛、瓶子、椅子、沙发、盆栽、餐桌、电视^[18]。

由于 COCO 数据集中有 20 个种类与 PASCAL VOC 数据集完全一致, 本文设计两种测试方法。首先使用 COCO 数据集中与 PASCAL VOC 数据集中不一样的 60 个种类作为基类进行训练, 第一种测试方法为同数据集测试, 使用 COCO 数据集中与 PASCAL VOC 数据集中相同的 20 个种类作为新类进行测试; 另一种方法为跨数据集测试, 使用 PASCAL VOC 数据集中与 COCO 数据集相同的 20 个种类作为新类测试。

2.2 训练策略

在元学习的第一个训练阶段, 模型在基类上只训练 30

个 epoches,这是由于对于少样本条件下的增量化目标检测任务而言,如果不提前停止训练,这些特征将过度适用于已知类别,而不适用于新的类别。学习率(Learning rate)设置为 0.000 25,批处理大小(Batch size)设置为 30。在第二个训练阶段,模型在新类上训练 140 个 epoches,学习率设置为 0.000 1,批处理大小设置为 20。两个训练阶段均使用 Adam 优化器,均采用通过实验得出的最优参数。

2.3 评价指标

平均精度(Average Precision, AP)主要用于衡量目标检测模型的准确性。其是考虑了不同置信度阈值下的精度计算出的一个平均值。精度的计算公式为:

$$Precision=TP/(TP+FP) \quad (5)$$

式中:TP 为真阳性,表示正例被识别为正例;FP 为假阳性,表示负例被识别为正例。

平均召回率(Average Recall, AR)是指在多个阈值下召回率的平均值,用于衡量模型在不同条件下的整体性能。召回率的计算公式为:

$$Recall=TP/(TP+FN) \quad (6)$$

式中:TP 为真阳性,表示正例被识别为正例;FN 为假阳性,表示正例被识别为负例。

2.4 实验结果分析

为验证所提方法的有效性,首先,在 COCO 数据集上进行测试。将所提方法与目前主流的 4 种少样本条件下的目标检测模型进行比较,分别为微调方法(Fine-Tuning)、模型无关元学习(Model Agnostic Meta Learning, MAML)^[19-20]、Feature-Reweight^[21]、ONCE。结果如表 1 所示。可以看出,本文方法在新类上的 AP 值和 AR 值均达到最优,ONCE 方法次之。在测试样本数量 shot=1 的情况下,与 ONCE 方法相比,本文方法的新类 AP 值提升了 2.4,基类 AP 值提升了 2.2,全类 AP 值提升了 1.2;当 shot=5 和 shot=10 时,与 ONCE 方法相比,本文方法的新类 AP 值分别提升了 2.8 和 2.5,基类 AP 值分别提升了 2.3 和 2.2,全类 AP 值分别提升了 2.7 和 1.6。在 shot=1、5 和 10 的情况下,与 ONCE 方法相比,本文方法的新类 AR 值平均提升了 15.0,基类 AR 值平均提升了 21.7,全类 AR 值平均提升了 17.6。

本文方法与 ONCE 方法在 COCO 数据集新类上的可视化测试结果如图 4 所示。可以看出,本文方法对新类生成的检测框更优质,能完全覆盖被检测目标。

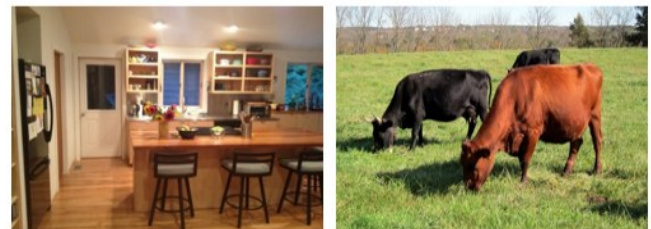
其次,在 PASCAL VOC 数据集上进行测试。首先选取 COCO 数据集中与 PASCAL VOC 数据集不一样的 60 个种类作为基类进行训练;然后选取 PASCAL VOC 数据集中的 20 个种类作为新类进行测试。在 shot=5、10 的情况下将本文方法与 Fine-Tuning、MAML、ONCE 3 种方法进行比较实验。结果如表 2 所示。可以看出,当 shot=5 和 10 时,本文方法均表现最优,在新类上的 AP 值和 AR 值上均较次优方法 ONCE 有接近 2 倍或 3 倍的提升。

本文方法与 ONCE 方法在 PASCAL VOC 数据集新类

Table 1 Performance comparison of different methods on COCO dataset

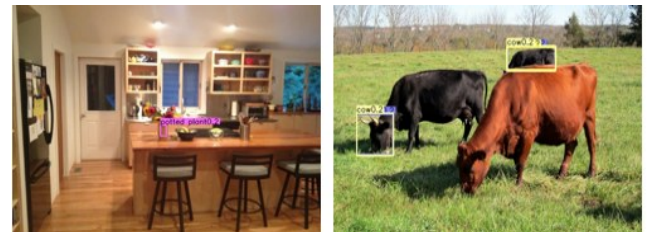
表 1 不同方法在 COCO 数据集上的性能比较

shot	方法	新类		基类		全类	
		AP	AR	AP	AR	AP	AR
1	Fine-Tuning	0.0	0.0	1.1	1.8	0.8	1.4
	MAML	0.1	0.5	N/A	N/A	N/A	N/A
	Feature-Reweight	0.1	0.3	2.5	4.3	1.9	3.3
	ONCE	0.7	6.3	17.9	19.5	13.6	16.2
	本文方法	3.1	19.7	20.1	40.9	14.4	33.5
5	Fine-Tuning	0.2	3.5	2.6	7.4	2.0	6.4
	MAML	0.4	3.9	N/A	N/A	N/A	N/A
	Feature-Reweight	0.8	5.1	3.3	8.2	2.6	7.4
	ONCE	1.0	7.4	17.9	19.5	13.7	16.4
	本文方法	3.8	23.7	20.2	41.2	15	34.1
10	Fine-Tuning	0.6	4.2	2.8	8	2.3	7.0
	MAML	0.8	4.9	N/A	N/A	N/A	N/A
	Feature-Reweight	1.5	8.3	3.7	8.9	3.1	8.7
	ONCE	1.2	7.6	17.9	19.5	13.7	16.5
	本文方法	3.7	22.9	21.1	41.5	15.3	34.5



(a) Original images

(a) 原图



(b) ONCE method

(b) ONCE 方法



(c) The proposed method

(c) 本文方法

Fig. 4 Visual detection results of new classes of COCO dataset using different methods

图 4 不同方法对 COCO 数据集新类的可视化检测结果

上的可视化测试结果如图 5 所示。可以看出,本文方法对新类进行检测时生成的检测框更优质,能完全覆盖被检测目标。

Table 2 Performance comparison of different methods on PASCAL VOC DATASET

表 2 不同方法在 PASCAL VOC 数据集上的性能比较			
shot	方法	AP	AR
5	Fine-Tuning	0.1	1.9
	MAML	0.6	2.2
	ONCE	2.4	12.2
	本文方法	4.8	32.5
	Fine-Tuning	0.3	2.8
10	MAML	1.0	3.2
	ONCE	2.6	11.6
	本文方法	5.1	33.6



Fig. 5 Visual detection results of new classes in PASCAL VOC dataset using different methods

图 5 不同方法对 PASCAL VOC 数据集新类的可视化检测结果

2.5 消融实验

为验证所提模型各改进模块的有效性,设计了消融实验。在第一阶段,仅使用普通的残差网络 ResNet 特征提取器来提取特征,然后使用嵌入注意力模块的残差网络 (ResNet+CBAM) 特征提取器来提取特征,最后在嵌入注意力模块的残差网络基础上结合特征融合方法 (ResNet+CBAM+Feature Fusion), 记为 IFSOD-AAFF (Incremental Few Shot Object Detection-Attention And Feature Fusion)。

在第二阶段,在结合注意力机制与特征融合的特征提取器模型 IFSOD-AAFF 基础上加上自适应高斯核方法,记为 IFSOD-SAGK (Incremental Few Shot Object Detection-Self Adaption Gaussian Kernel)。在 IFSOD-SAGK 的基础上加上分解注意力方法,记为 IFSOD-AD (Incremental Few Shot Object Detection-Attention Decomposed)。由于 COCO 数据集比 PASCAL VOC 数据集更具有挑战性,实验结果更具代表性,本文实验只展示在 COCO 数据集上的消融实验结果。具体如表 3 所示。

可以看出,在 shot=1、5、10 的 3 种情况下,使用嵌入注意力模块的残差网络 ResNet+CBAM 比只使用残差网络 ResNet 的 AP 值平均提升了 1.3, AR 值平均提升了 8.0; IFSOD-AAFF 能使模型检测性能达到最优,尤其是在 shot=5 的情况下,其对新类的 AP 值相较普通残差网络 ResNet 提升了 50%,表明使用 ResNet+CBAM 能有效提升模型的特征提取能力,而特征融合方法能有效利用不同层次的特征,二者结合具有协同作用。自适应高斯核方法 IFSOD-SAGK 相较 ResNet+CBAM 在新类上的 AP 和 AR 平均提升了 1.2 倍。在 IFSOD-SAGK 的基础上采用分解注意力方法后,IFSOD-AD 在新类上的 AP 值和 AR 值平均分别相较 IFSOD-SAGK 提升了 2.0 倍和 1.1 倍,达到最优检测效果。在 shot=1 时,IFSOD-AD 的 AP 相较 IFSOD-AAFF 提升了 3.1 倍,相较 IFSOD-SAGK 提升了接近 2.4 倍;在 shot=5 和 shot=10 时,IFSOD-AD 的 AP 相较 IFSOD-SAGK 均提升了 2.5 倍。当 shot=1、5、10 时,IFSOD-AD 方法在新类上的 AR 值相较 IFSOD-AAFF 平均提升了 1.4 倍,相较 IFSOD-SAGK 平均提升了接近 1.1 倍。IFSOD-AAFF 和 IFSOD-SAGK 模型第二阶段仅使用单独的注意力编码器同时用于中心点定位和尺寸回归,IFSOD-AD 则使用两个独立的注意力编码器分别用于中心点定位和尺寸回归。消融实验结果表明,使用分解注意力的方法能有效提升模型对新类的检测性能。

Table 3 Ablations experiment results

表 3 消融实验结果

shot	方法	COCO 数据集	
		AP	AR
1	ResNet	0.7	6.3
	ResNet+CBAM	0.8	12.9
	IFSOD-AAFF	1.0	13.0
	IFSOD-SAGK	1.3	16.7
	IFSOD-AD	3.1	19.7
5	ResNet	1.0	7.4
	ResNet+CBAM	1.2	15.1
	IFSOD-AAFF	1.5	17.9
	IFSOD-SAGK	1.9	20.7
	IFSOD-AD	3.8	23.7
10	ResNet	1.2	7.6
	ResNet+CBAM	1.3	17.3
	IFSOD-AAFF	1.5	17.3
	IFSOD-SAGK	1.8	21.0
	IFSOD-AD	3.7	22.9

3 结语

本文针对传统检测器对少量样本特征提取能力不足、特征在提取过程中有损耗、中心定位框不准确以及目标框大小偏移等问题,以 CenterNet 为基础框架,提出一种少样本条件下的增量化目标检测算法。该算法结合注意力机制与特征融合的特征提取器提高对样本特征的提取能力;同时建立两个独立的任务通道分别用于中心点定位和尺寸回归,采用自适应高斯核获取更准确的目标中心点。所提算法在公开数据集上的检测性能相对对照算法有所提升。然而,本文模型是在公开数据集上进行训练与测试的,所用样本数量和类别相对固定。如果应用于其他复杂场景或真实世界,模型对新类的检测效果可能不尽如人意,泛化能力需进一步提升。后续可使用更多样化的数据训练模型,提高其对未知数据的处理能力。

参考文献:

- [1] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.
- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [3] GIRSHICK R. Fast R-CNN[C]//Proceedings of IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceeding of International Conference on Neural Information Processing System, 2015: 91-99.
- [5] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//Proceedings of the European Conference on Computer Vision, 2014: 346-361.
- [6] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [7] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//Proceedings of IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision, 2016: 21-37.
- [10] LAW H, DENG J. Cornernet: detecting objects as paired keypoints[C]//Proceedings of the European Conference on Computer Vision, 2018: 734-750.
- [11] PEREZ-RUA J M, ZHU X T, HOSPEDALES T, et al. Incremental few-shot object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 13846-13855.
- [12] JOSEPH K J, KHAN S, KHAN F S, et al. Towards open world object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021: 5830-5840.
- [13] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [14] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision, 2018: 3-19.
- [15] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the European Conference on Computer Vision, 2017: 2117-2125.
- [16] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as point [DB/OL]. <https://arxiv.org/abs/1904.07850>.
- [17] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//Proceedings of the European Conference on Computer Vision, 2014: 740-755.
- [18] MARK E, ALI E, LUC V G, et al. The PASCAL visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111: 98-136.
- [19] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//International Conference on Machine Learning, 2017: 1126-1135.
- [20] NICHOL A, ACHIAM J, SCHULMAN J. On first-order meta-learning algorithms[DB/OL]. <https://arxiv.org/abs/1803.02999>.
- [21] KANG B, LIU Z, WANG X, et al. Few-shot object detection via feature reweighting[C]//Proceedings of IEEE International Conference on Computer Vision, 2019: 8420-8429.

(责任编辑:尹晨茹)