

结合 Graph-FPN 与稳健优化的开放世界目标检测

谢斌红, 张鹏举⁺, 张 睿

太原科技大学 计算机科学与技术学院, 太原 030024

+ 通信作者 E-mail: S20202011071@stu.tyust.edu.cn

摘要: 开放世界目标检测(OWOD)要求检测图像中所有已知和未知的目标类别,同时模型必须逐步学习新的类别以自适应更新知识。针对 ORE 方法存在的未知目标召回率低以及增量学习的灾难性遗忘等问题,提出一种基于图特征金字塔的稳健优化开放世界目标检测方法(GARO-ORE)。首先,利用 Graph-FPN 中的超像素图像结构以及上下文层和层次层的分层设计,获取丰富的语义信息并帮助模型准确定位未知目标;之后,利用稳健优化方法对不确定性综合考量,提出了基于平坦极小值的基类学习策略,极大地保证模型在学习新类别的同时避免遗忘先前学习到的类别知识;最后,采用基于知识迁移的新增类别权值初始化方法提高模型对新类别的适应性。在 OWOD 数据集上的实验结果表明,GARO-ORE 在未知类别召回率上取得较优的检测结果,在 10+10、15+5、19+1 三种增量目标检测(iOD)任务中,其 mAP 指标分别提升了 1.38、1.42 和 1.44 个百分点。可以看出,GARO-ORE 能够较好地提高未知目标检测的召回率,并且在有效缓解旧任务灾难性遗忘问题的同时促进后续任务的学习。

关键词: 开放世界目标检测(OWOD);图特征金字塔网络;平坦极小值;知识迁移

文献标志码:A 中图分类号:TP391

Open World Object Detection Combining Graph-FPN and Robust Optimization

XIE Binhong, ZHANG Pengju⁺, ZHANG Rui

College of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China

Abstract: Open world object detection (OWOD) requires detecting all known and unknown object categories in the image, and the model must gradually learn new categories to adaptively update knowledge. Aiming at the problems of low recall rate of unknown objects and catastrophic forgetting of incremental learning in ORE (open world object detection) method, this paper proposes adjustable robust optimization of ORE based on graph feature pyramid (GARO-ORE). Firstly, using the superpixel image structure in Graph-FPN and the hierarchical design of context layer and hierarchical layer, rich semantic information can be obtained and the model can accurately locate unknown object. Then, using the robust optimization method to comprehensively consider the uncertainty, a base class learning strategy based on flat minimum is proposed, which greatly ensures that the model avoids forgetting the previously learnt category knowledge while learning new categories. Finally, the classification weights initialization method based on knowledge transfer is used to improve the adaptability of the model to new classes.

基金项目: 山西省重点研发计划项目(201803D121048, 201803D121055);山西省基础研究计划项目(面上)(2021030212321);吕梁市引进高层次科技人才重点研发项目(2022RC08)。

This work was supported by the Key Research and Development Program of Shanxi Province (201803D121048, 201803D121055), the Basic Research Project of Shanxi Province (General Program) (2021030212321), and the Key Research and Development Projects of Introducing High-level Scientific and Technological Talents in Lvliang City (2022RC08).

收稿日期: 2022-11-17 **修回日期:** 2023-03-16

Experimental results on the OWOD dataset show that GARO-ORE achieves better detection results on the recall rate of unknown categories. In the three types of incremental object detection tasks of 10 + 10, 15 + 5, and 19 + 1, the mAP is increased by 1.38, 1.42 and 1.44 percentage points, respectively. It can be seen that GARO-ORE can improve the recall rate of unknown object detection, and promote the learning of subsequent tasks while effectively alleviating the catastrophic forgetting problem of old tasks.

Key words: open world object detection (OWOD); graph feature pyramid network; flat minimum; knowledge transfer

目标检测^[1]是计算机视觉领域的重要研究方向,其旨在从图像中定位感兴趣的目标,并准确判断每个目标的类别。随着计算机技术的发展和计算机视觉技术的广泛应用,目标检测算法在准确性和实时性方面已经取得了出色的表现,并广泛应用于智慧安防、无人驾驶以及远程医疗等领域。然而,常规的目标检测方法^[1-8]通常设定在封闭世界的假设中,即在特定的数据集上进行训练,学习固定数量的目标类别,并应用于特定场景,这很大程度上限制了目标检测技术的发展和應用。开放世界目标检测突破了主流基准中的封闭假设,将目标检测模型应用到开放领域中。它于2021年由Joseph等^[9]首次提出,其主要任务是:(1)图像中可能包含未知类别的目标对象,在无监督信息的情况下,能将未参与训练的类别目标对象识别为“未知”;(2)当未知目标有可用标签信息时,模型可以在不遗忘已知类别也不需要再在已知类上重新训练的情况下,能够实现未知类别的目标检测。在现实世界中,由于新目标类别的数量不断增加,并且动态变化,探索OWOD(open world object detection)对机器人、自动驾驶、医疗保健监测等领域具有重要的应用价值和现实意义。

与常规目标检测任务不同,OWOD在从背景中分离未知目标以及模型持续学习等方面提出了重大挑战。ORE(open world object detection)^[9]是Joseph等^[9]提出的一种开放世界目标检测的解决方案,它以两阶段的目标检测模型Faster R-CNN(faster region-based convolutional neural networks)作为基础模型。首先使用自动标注未知类别的RPN网络(auto-labelling unknowns with region proposal network, ALU-RPN)标记出图像中潜在的目标区域,将预测框中对象性分数高且与已知类目标不重叠的目标直接归类为未知目标,并将提取的特征加入到相应的已知和未知队列中。然后在ROI Head中利用对比聚类对特征队列进行特征分离,帮助基于能量的分类头更有效地区分已知和未知实例。之后使用基于能量的分类器

(energy based classification models, EBM)对已知类别和未知类别的能量值分布进行Weibull建模,根据目标特征的已知和未知的能量值将其进行分类。最后,采用两阶段微调策略(two-stage fine-tuning approach, TFA)^[10]帮助模型在开放世界目标检测的场景下实现增量学习。

ORE作为首次提出应对具有挑战性的开放世界目标检测任务的解决方案,仍然存在一些问题:(1)RPN依赖单一尺度的Feature Map容易造成未知目标的边界模糊和细节丢失,难以对未知目标进行精确定位,导致未知目标召回率较低。(2)新增类别样本较少导致模型难以很好地学习新任务,并且微调策略不足以克服旧类别的灾难性遗忘,使得模型在增量学习过程中性能严重下降。

针对上述问题,提出一种基于图特征金字塔网络的稳健优化增量学习方法(adjustable robust optimization of ORE based on graph feature pyramid, GARO-ORE),本文的主要贡献是:

(1)采用Graph-FPN作为特征提取网络,通过其超像素结构和分层设计促使多尺度特征进行交互,获得了丰富的语义信息,优化了未知目标的定位问题。

(2)提出基于平坦极小值的基类学习策略,提升模型的泛化能力,有效缓解了增量学习过程中灾难性遗忘的影响,并利用轻量级锐度感知训练算法进一步降低计算开销。

(3)提出基于知识迁移的分类权值初始化策略,利用旧类的分类权值及新旧类别的特征相似度,对新增类别分类权值进行初始化,帮助网络快速适应新类别。

(4)在OWOD数据集上进行的大量实验验证了GARO-ORE方法的有效性。

1 相关工作

Joseph等^[9]发现人类具有识别环境中未知对象实例的自然本能,并根据当前开集识别和增量学习等

技术提出了开放世界目标检测(OWOD)这样一个新的计算机视觉任务。由于OWOD的设定比现有的封闭世界静态学习设置更贴近实际场景,受到了研究学者的广泛关注。下面将从开集识别和增量学习两方面对相关工作进行阐述。

1.1 开集识别

开集识别(open set recognition, OSR)^[11-14]认为通过训练集获得的知识是不完整的,即在现实应用过程中可能会遇到新的未知类。针对该问题,相关学者研究探索了利用自监督学习^[15]和基于重构的无监督学习^[16]方法用于开集识别。虽然这些方法可以识别未知类别,但这些网络在训练过程中不能以增量方式动态更新。Bendale等^[17]提出开放世界设置中的图像分类任务,这是首个基于深度学习的开集OSR方法,并能够实现分类网络中的增量学习方法。本文的研究内容与此方法^[17]更加贴近,不同的是,此方法^[17]应用于图像分类任务,不需要对未知目标定位,而本文的研究内容以端到端的方式在开放世界场景下对图像中的已知和未知目标进行检测。

1.2 增量学习

增量学习^[18-20]是一种通过从新数据中学习来增加现有知识的范式。然而,增量学习长期存在一个严重的问题,即灾难性遗忘,它是指由于当前训练中无法访问旧数据,从而导致旧类别的测试性能急剧下降。随着近年来深度学习的发展,提出了许多增量学习的方法用以解决灾难性遗忘问题。当前主要包括三类方法:基于样本回放的方法、基于参数隔离的方法和基于正则化的方法。其中,基于样本回放的方法^[21],是根据一系列筛选标准,从旧数据集中抽取一些具有代表性的原始样本,将它们和新任务的样本进行联合训练来达到记忆目的;基于参数隔离的方法^[22]通常是在新任务上扩大旧模型,并针对不同任务分配不同的模型参数,进行不同程度的隔离,以防止后续任务对之前学习的知识产生干扰;基于正则化的方法^[23]通过在损失函数上引入额外的正则项来对权值调整进行约束。而本文提出了基于平坦极小值的基类学习策略,有效缓解增量学习过程中灾难遗忘问题。

2 基于稳健优化的开放世界目标检测

本章主要针对ORE算法存在的未知类召回率低,以及增量学习中灾难性遗忘等问题提出了改进后的GARO-ORE模型。

针对ORE模型未知类召回率低的问题,在2.3节中提出了利用Graph-FPN优化特征提取,帮助模型提高未知类的召回率。首先,由于OWOD任务中无法利用监督信息对未知目标进行回归处理,对图像进行超像素分割,并将其作为Graph-FPN的输入,以更强的语义表征目标边界,从而有助于获取更高质量的候选区域。其次,利用部分-整体的层次结构(part-whole hierarchies)弥合了像素与对象之间的语义差距,增强了目标的语义特征。

为了有效缓解模型在增量学习过程中造成的灾难遗忘问题,在2.4节中提出了基于平坦极小值的基类学习策略,更有效地保留旧类别检测性能。此外,为了在模型容量固定条件下,提高增量学习的持久性,在2.5节中提出了基于知识迁移的新增类别权值初始化策略,更有效地提升模型的持续学习能力。

2.1 任务定义

在Joseph等^[9]定义的OWOD中,模型 M_t 在 t 时刻除了需要检测已知类 $K_t = \{1, 2, \dots, C\}$ 之外,还需要将先前未学习到的目标实例检测为未知类 $U = \{C+1, C+2, \dots\}$ (由标签0表示)。然后,用户可以选择性地标记 n 个感兴趣的新类并标注相应未知实例训练模型,将这组新类别添加到已知类别中,即 $K_{t+1} = K_t + \{C+1, C+2, \dots, C+n\}$ 。最后,模型 M_t 在 K_{t+1} 上进行增量学习,不需要在整个数据集上进行训练,便可检测 K_{t+1} 中所有的目标类别。这种循环在模型的生命周期中持续进行,促使模型自适应更新。

2.2 GARO-ORE网络结构

GARO-ORE网络结构如图1所示,包含Backbone、ALU-RPN、ROI Pooling和分类回归网络四部分。GARO-ORE首先通过Graph-FPN提取图像特征。然后,利用ALU-RPN获取已知和未知目标的潜在区域,并且将对象性得分 Top_k 的背景区域建议框作为候选未知对象。之后,在ROI Pooling层将ALU-RPN的输出进行池化,生成大小一致的特征图。最后,在分类回归网络中,为了更好地区分已知和未知对象,在潜在空间中对池化后的特征进行对比聚类。利用基于能量的分类头学习已知和未知类的能量分布,促进模型更有效识别未知目标。分别预测目标边界框和目标类别。

此外,在本文中,为了缓解基类的遗忘问题,采用基于平坦极小值的基类学习方法对分类回归网络进行训练,并在后续的增量学习过程中钳制模型参数处于平坦区域(flat region)内,使其在增量学习过

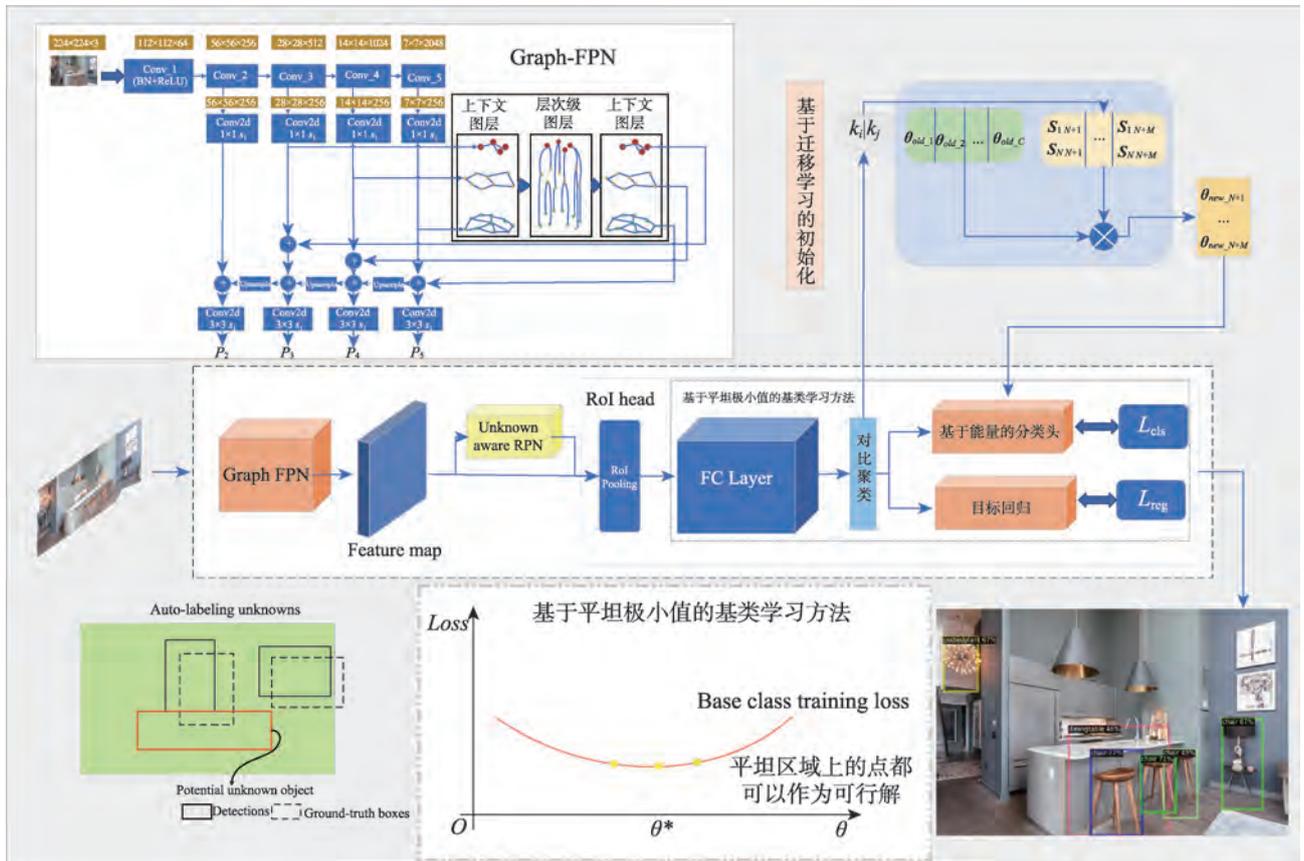


图1 GARO-ORE网络结构

Fig.1 GARO-ORE network structure

程中保持旧类别性能的稳定性。另外,在学习新增类别时,采用基于知识迁移的新增类别权值初始化策略,更有效地发掘模型的持续学习能力。

2.3 图特征金字塔网络

ORE算法使用 ResNet50 作为模型特征提取网络,虽然 ResNet 具有较强的特征表达和逐层学习能力,但由于其容易造成目标的特征边界模糊和细节丢失,导致模型对未知目标定位不准确,目标框存在偏大或偏小等情况,并且单一的分辨率通常会丢失部分图像信息,继而影响后续目标检测和增量学习性能。针对上述问题,本文在 Backbone 模块采用 Graph-FPN^[24]作为特征提取网络,其结构如图2所示。首先利用基于全卷积编解码器网络的轮廓检测算法(convolutional oriented boundaries, COB)^[25]将原始图像分割成符合目标物体边界语义区域的超像素图像结构,并在图神经网络中利用部分-整体层次结构判断二者之间对象特征语义一致性,进一步提高每个显著对象微观层面的完整性,从而帮助模型对目标准确定位,并快速生成定位质量较高的候选区

域。之后,利用上下文层和层次层的分层设计,将多尺度特征进行跨空间和跨尺度的特征交互,并引入局部通道注意力机制增强目标特征的表达能力,为目标检测和新类别学习提供更加丰富的语义信息,从而在有效提升未知目标的召回率的同时,也提升了模型的学习能力。

具体来讲,如图2所示,Graph-FPN包含4个步骤进行特征提取:

步骤1 如图3所示,首先利用COB超像素算法对图像进行分割,这种超像素图像不仅能防止背景和目标之间的特征混合,而且能有效保留目标的边界信息。然后利用分层分割技术获取图2中超像素图像的金字塔层级结构 $S = \{S_0, S_1, \dots, S_4\}$,以此来表达图像的局部到整体的层次结构,其中 S_{i+1} 的超像素数目是 S_i 中的 $1/4$ 。

步骤2 利用 S_2, S_3, S_4 分别构建图特征金字塔(Graph-FPN)。其中, S_i 中的每个超像素点与 Graph-FPN 中的节点对应,并为每个节点构建上下文交互边和层次交互边,通过图神经网络对同一尺度内和不

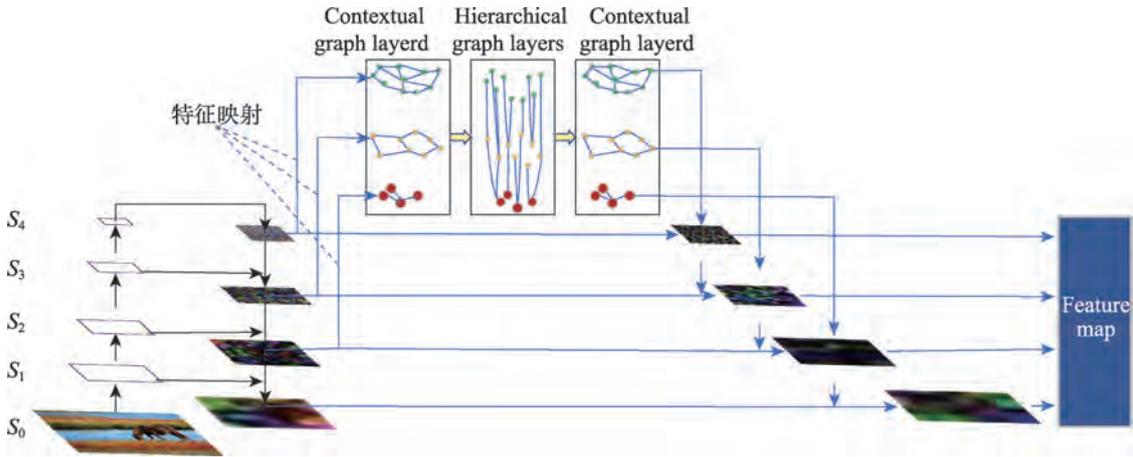


图2 Graph-FPN 特征提取网络

Fig.2 Graph-FPN feature extraction network



图3 COB 算法生成超像素

Fig.3 Superpixel generation by COB algorithm

同尺度之间的特征进行融合。

步骤3 Graph-FPN通过空间和通道注意力机制进一步增强特征的表达能力。其中,空间注意力采用式(1)对图中的节点特征进行更新:

$$h'_i = M(h_i, \{h_j\}_{j \in N_i}) \quad (1)$$

其中, M 表示图注意力机制^[26], $h_{j \in N_i}$ 是节点 i 的邻居节点特征向量, h_i 和 h'_i 分别是节点 i 更新前后的特征向量。

通道注意力是由平均池化的局部通道注意力模块和局部通道自注意力模块组成。平均池化的局部通道注意力模块特征更新如式(2)所示:

$$h_i'' = \sigma(W_1 a_i') \odot h_i' \quad (2)$$

其中, σ 表示 sigmoid 函数, $W_1 \in \mathbb{R}^{C \times C}$ 为全连通层可学习权重矩阵。 $a_i' \in \mathbb{R}^C$ 表示节点 i 及其邻居的特征向量进行平均得到特征向量。局部通道自注意力模块特征更新如式(3)所示:

$$h_i''' = \beta X h_i'' + h_i'' \quad (3)$$

其中, β 表示初始化为0的可学习权值。 $X = A^T * A$ 表示通道相似矩阵, A 表示节点 i 及其邻居的特征向量集合。

步骤4 将经过融合后的特征由图神经网络映射到特征金字塔网络。

总之,利用 Graph-FPN 作为特征提取网络,不仅优化了模型的特征表达能力,提高了模型在开放世界环境中的未知目标的检测精度,同时也能帮助模型更好地完成后续的目标检测和增量学习等任务。本文的对比实验也验证了该网络在提高对未知目标检测和增量学习的有效性。

2.4 基于平坦极小值的基类学习策略

灾难性遗忘是增量学习长期存在的一个严重问题,如相关工作中所述,解决该问题对开放世界目标检测任务尤为重要。众所周知,深度神经网络的泛化能力一般与其极小值附近的平坦度^[27]有关,并且较为平坦的最优解通常具有更好的鲁棒性,因此,本文提出了基于平坦极小值的基类学习策略,即:在平坦极小值附近对模型参数随机扰动,使其损失移动更加稳健,以此能够更好地解决模型的灾难性遗忘问题。

基于上述动机,本文利用式(5)计算模型的平坦极小值 θ^* 。

$$L_s(f_\theta) = \frac{1}{N_{cls}} \sum_i L_{cls}(K_i, K_i^*) + \frac{\lambda}{N_{reg}} \sum_i P_i^* L_{reg}(P_i, P_i^*) \quad (4)$$

$$\theta^* = \arg \min \{R_s(f_\theta) + L_s(f_\theta)\} \quad (5)$$

其中,式(4)表示 Faster R-CNN 的分类和回归损失之和。 $L_{cls}(K_i, K_i^*)$ 、 $L_{reg}(P_i, P_i^*)$ 分别为模型的分类和回归损失, K_i 和 P_i 分别为模型预测的类别和目标框位置, K_i^* 和 P_i^* 分别为训练样本的类别和目标框位置, N_{cls} 和 N_{reg} 表示分类和回归的样本数。式(5)中 $R_s(f_\theta)$

表示模型的锐度度量损失,近似于 $L_s(f_{\theta+\varepsilon})-L_s(f_\theta)$ 的最大值, s 表示训练集数据, θ 表示 GARO-ORE 的模型参数, ε 是将 $L_s(f_{\theta+\varepsilon})$ 的一阶泰勒逼近在 θ 附近最大化问题的一个近似解,即 $\varepsilon: \|\varepsilon\|_2 \leq \rho$, 其中 ρ 表示约束邻域半径的预定义常数。

使用基于平坦极小值的基类学习策略时,首先需要在 $t=1$ 的训练阶段,按式(5)求解模型的平坦局部极小值点 θ^* 作为模型的最优解,然后在 $t \geq 2$ 的持续学习过程中,在平坦区域内钳制模型参数(Fast R-CNN 网络的全连接层参数)进行微调,即 $\theta^* - \rho \leq \theta \leq \theta^* + \rho$ 。如图 4 所示为现有技术(尖锐极小值)和本文方法(平坦极小值)的解释说明^[28],其中图 4(a)表示模型在引入新类别参数后直接微调模型导致旧类别的检测性能严重下降,图 4(b)表示通过引入平坦极小值微调参数可以有效地缓解灾难性遗忘。

2.4.1 寻找基类训练阶段的平坦局部极小值

本文采用添加随机噪声的方法来定位目标函数的平坦局部极小值。具体来讲,在 Faster R-CNN 全连接层网络参数中通过多次添加随机噪声 ε , 获取相似但不同的损失函数,并将这些损失函数联合优化以定位平坦局部极小值,具体计算如式(5)所示。由于额外的 $R_s(f_\theta)$ 计算增加了模型的计算成本,从而导致在 COCO2017、VOC2007 等大型数据集训练时极大降低了模型训练速度。针对该问题,本文采用一种计算开销较小的轻量级锐度感知算法(sharpness-aware training for free, SAF)^[29]替换锐度损失 $R_s(f_\theta)$ 来寻找基类训练阶段的局部极小值,SAF算法具体计算过程如下。

首先按式(6)计算锐度度量损失 $R_s(f_\theta)$ 的一阶泰

勒展开式。

$$R_s(f_\theta) = L_s(f_{\theta+\varepsilon}) - L_s(f_\theta) \approx L_s(f_\theta) + \varepsilon \nabla_\theta L_s(f_\theta) - L_s(f_\theta) = \rho \frac{\nabla_\theta L_s(f_\theta)^T}{\|\nabla_\theta L_s(f_\theta)\|_2} \nabla_\theta L_s(f_\theta) = \rho \|\nabla_\theta L_s(f_\theta)\|_2 \quad (6)$$

如式(6)所示,最小化锐度损失 $R_s(f_\theta)$ 等价于最小化梯度 $\nabla_\theta L_s(f_\theta)$ 的 L_2 范数,这与最小化 $L_s(f_\theta)$ 的梯度相同。另外,由于常规训练(SGD 优化器)的学习率 μ 一般小于 ρ ,模型在两个连续小批量(mini-batch)数据集迭代中更新权值的训练损失变化如式(7)所示:

$$L_s(f_\theta) - L_s(f_{\theta - \mu \nabla_\theta L_s(f_\theta)}) \approx \mu (\|\nabla_\theta L_s(f_\theta)\|_2)^2 = \frac{\mu}{\rho^2} R_s(f_\theta)^2 \quad (7)$$

由此可知,权重更新后训练损失的变化与 $R_s(f_\theta)^2$ 成正比,因此,最小化损失差近似于最小化锐度损失,于是本文引入了一种新的轨迹损失,即用标准训练过程中学习到的权值轨迹代替锐度损失,以降低锐度度量损失的计算开销。

如图 5 所示,红色线条表示常规的训练损失 $L_s(f_\theta)$,蓝色箭头表示模型训练轨迹。可以看出,尖锐的局部极小值往往具有较大的轨迹损失,而平坦极小值具有较小的轨迹损失。其中 L_s^{tr} 表示 E 个阶段的轨迹损失大小。

在计算权值轨迹时,首先用 $\Theta = \{\theta_1, \theta_2, \dots, \theta_t\}$ 表示迭代过程中的权值轨迹,其中 θ_t 表示第 t 次迭代权值,则 θ_{t+1} 的迭代权值可表示为 $\theta_{t+1} = \theta_t - \mu \nabla_{\theta_t} L_{S_t}(f_\theta)$ 。对于 t 时刻的锐度损失,可以由式(8)表示:

$$\arg \min_{\theta_t} R_s(f_{\theta_t}) = \arg \min_{\theta_t, \theta_i \sim \text{Unif}(\Theta)} E [r_i R_s(f_{\theta_t}) R_s(f_{\theta_i})] \quad (8)$$

其中, $\gamma_i = \frac{\mu_i}{\rho^2} \cos \varphi_i$, φ_i 表示小批量训练集 S_i 和 S_t 的梯度角度之差, θ_i 是均匀分布在集合 Θ 上的迭代权

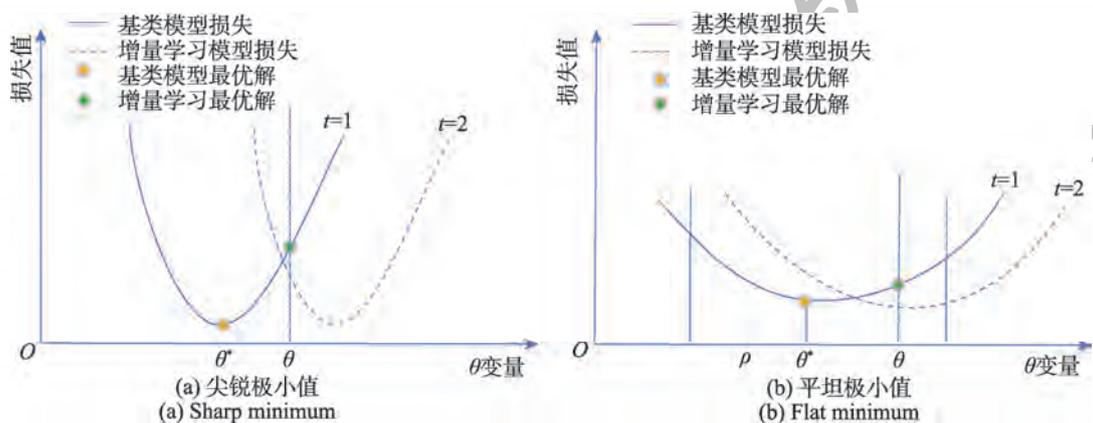


图4 尖锐极小值与平坦极小值的权值变化

Fig.4 Weight change of sharp minimum and flat minimum

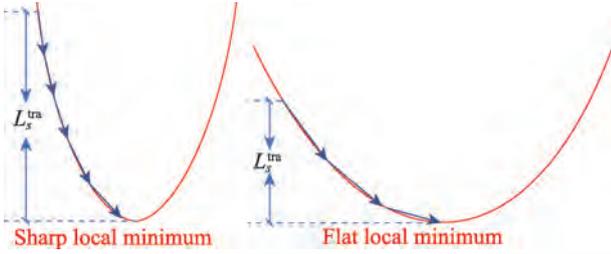


图5 尖锐及平坦极小值的轨迹损失

Fig.5 Trajectory loss of sharp and flat minimum

值。最后利用式(9)进行轨迹损失计算。

$$\begin{aligned} E_{\theta_i \sim \text{Unif}(\Theta)} [\mathbf{r}_i \mathbf{R}_{s_i}(f_{\theta_i}) \mathbf{R}_{s_i}(f_{\theta_i})] = \\ E_{\theta_i \sim \text{Unif}(\Theta)} [\mu_i \cos(\varphi_i) \|\nabla_{\theta_i} \mathbf{L}_{s_i}(f_{\theta_i})\| \|\nabla_{\theta_i} \mathbf{L}_{s_i}(f_{\theta_i})\|] = \\ E_{\theta_i \sim \text{Unif}(\Theta)} [\mu_i \nabla_{\theta_i} \mathbf{L}_{s_i}(f_{\theta_i})^T \nabla_{\theta_i} \mathbf{L}_{s_i}(f_{\theta_i})] \approx \\ E_{\theta_i \sim \text{Unif}(\Theta)} [\mathbf{L}_{s_i}(f_{\theta_i}) - \mathbf{L}_{s_i}(f_{\theta_{i+1}})] = \frac{1}{t-1} [\mathbf{L}_{s_i}(f_{\theta_i}) - \mathbf{L}_{s_i}(f_{\theta_i})] \quad (9) \end{aligned}$$

由式(6)可知,最小化损失差 $\mathbf{R}_{s_i}(f_{\theta_i})$ 等价于 $[\mathbf{L}_{s_i}(f_{\theta_i}) - \mathbf{L}_{s_i}(f_{\theta_{i+1}})]$,表明轨迹损失可以替代锐度度量损失 $\mathbf{R}_{s_i}(f_{\theta_i})$,并且能极大降低训练损失计算开销。

另外,为了进一步精确估计锐度损失,只考虑过去 E 次迭代的模型更新轨迹,其中 E 为训练的超参数。然而,当同时最小化分类损失和锐度度量损失时,式(9)中的 $\mathbf{L}_{s_i}(f_{\theta_{i+1}})$ 交叉熵损失与式(4)中的 $L_{\text{cls}}(\mathbf{K}_i, \mathbf{K}_i^*)$ 交叉熵损失相抵消。因此,本文将交叉熵损失替换为KL散度损失以解耦回归损失。另外,受知识蒸馏的启发,本文还使用温度 τ 来软化KL散度损失。因此,第 e 次迭代的轨迹损失定义如下:

$$\begin{aligned} L_s^{\text{tra}}(f_{\theta}, Y^{(e-E)}) = \frac{\lambda}{|S|} \sum_{\substack{x_i \in S, y_i \\ (e-E) \in Y^{(e-E)}}} \text{KL}\left(\frac{1}{\tau} y_i^{(e-E)}, \frac{1}{\tau} f_{\theta}(x_i)\right) \\ \text{where: } Y^{(e-E)} = \{y_i^{(e-E)} = f_{\theta}^{(e-E)}(x_i); x_i \in S\} \quad (10) \end{aligned}$$

其中, $y_i^{(e-E)}$ 是实例 x_i 在 E 次迭代之前的网络输出值。在此过程中,轨迹损失降低了模型训练损失的变化率,并且帮助模型收敛到平坦的局部极小值。

2.4.2 持续学习阶段模型微调

在进行增量学习时,对平坦区域内微调分类权值和定位权值以学习新类。值得注意的是,虽然平坦区域可能相对较小,但是足以很好地完成增量学习。

更新全连接层网络参数后,通过对参数进行钳制,以确保其位于平坦区域内,即 $\theta^* - \rho \leq \theta \leq \theta^* + \rho$ 。

2.5 基于知识迁移的新增类别权值初始化

在增量学习时,模型首先需要为每个新类别的权值进行初始化。ORE使用了一组随机值进行初始

化,该方法容易导致模型以较慢速度收敛于尖锐极小值,从而往往无法达到期望的学习效率。因此,为了解决ORE中随机初始化参数方法的缺陷,提出了一种基于知识迁移的新增类别权值初始化方法,其核心思想是利用特征相似度从旧类中学习新类的特定信息,实现新旧类之间知识迁移,从而为分类头提供有效的初始值,并有利于使用少量数据进行微调。

具体来讲,首先采用类别的特征分布表示知识,并通过使用ORE论文中的对比聚类方法将每个基类的高维特征聚合成多个特征簇,然后将每个特征簇中特征均值作为该类别的知识,如式(11)所示:

$$k_c = \frac{1}{N_i} \sum_{j=1}^{N_i} f_{\theta}(x_k) \quad (11)$$

其中, k_c 表示类别 c 的特征均值, N_i 表示该簇内的特征总数, $f_{\theta}(x_k)$ 表示该簇内的第 k 个样本的特征向量。然后,利用余弦相似度计算新增类别特征均值与旧类别的特征均值相似度并归一化处理,具体计算如式(12)、式(13)所示,其中 $k_i, i \in \{1, 2, \dots, C\}$, 与 $k_j, j \in \{C+1, C+2, \dots\}$ 分别代表 n 维空间中旧类别与新类别两个特征簇的特征均值。

$$\cos_{(k_i, k_j)} \theta = \frac{k_i \cdot k_j}{|k_i| \cdot |k_j|} \quad (12)$$

$$s_{ij} = \frac{\cos_{(k_i, k_j)} \theta}{\sum_{i=1}^c \cos_{(k_i, k_j)} \theta} \quad (13)$$

最后,利用式(14)获取分类权值的初始值,利用得到的初始化权值对模型进行训练。

$$\theta_{\text{new}_j} = [\theta_{\text{old}_1}, \theta_{\text{old}_2}, \dots, \theta_{\text{old}_C}] [s_{1j}, s_{2j}, \dots, s_{Cj}]^T \quad (14)$$

其中, $\theta_{\text{old}_i}, i \in \{1, 2, \dots, C\}$ 表示模型学习的旧类别参数, $s_{ij}, i \in \{1, 2, \dots, C\}$ 表示新类别与旧类别的归一化的相似度, θ_{new_j} 表示经过基于知识迁移的新增类别权值初始化得到的新类别参数。通过知识迁移的分类权值初始化,可以有效地将知识从旧类传播到新类,为分类头提供有效的初始值,帮助模型更有效地对新类别建模,使得模型更快地适应新类别,提高模型的检测精度。其实现过程如图6所示。

通过上述基于知识迁移的新增类别权值初始化,结合在平坦区域内钳制模型参数微调的方法,模型能够有效地避免灾难性遗忘,同时增强其对新类别检测的适应能力。

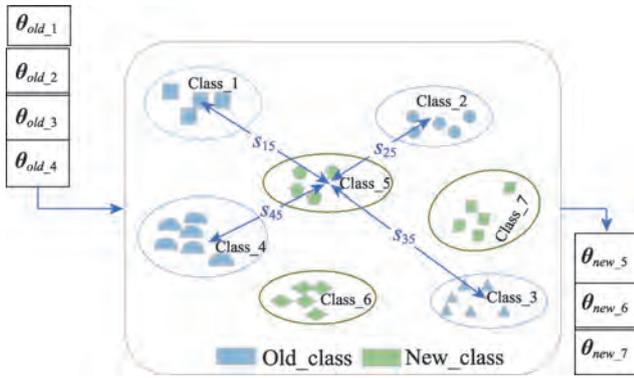


图6 基于知识迁移的权值初始化

Fig.6 Weight initialization based on knowledge transfer

3 实验

3.1 数据集和评价指标

为了验证 GARO-ORE 模型的可行性和有效性,以及比较的公平性,本文遵照 ORE^[9]实验中的设定,选用 OWOD 数据集对模型进行评估。OWOD 数据集由标准数据集 VOC2007 和 COCO2017 数据集合并而成,共包含 80 种目标类别。同时参考 ORE 的实验设置,将这 80 个类别划分为 4 组不重叠的数据集 $\{D_1, D_2, D_3, D_4\}$ 进行实验分析。如表 1 所示,每个数据集含有 20 个不同类别,其中,数据集 D_1 包含 VOC-2007 数据集所有类别及其相应类别的训练及测试数据, D_2, D_3, D_4 数据集是将剩余 60 个类别对应的训练和测试数据划分成具有语义漂移的连续数据集。

表1 实验数据集

Table 1 Experimental dataset

数据集 ID	Classes	Training images	Test images	Training instances	Test instances
D_1	VOC Classes	16 551	4 952	47 223	14 976
D_2	Outdoor Accessories Appliance Truck	45 520	1 914	113 741	4 966
D_3	Sports Food	39 402	1 642	114 452	4 826
D_4	Electronic Indoor Kitchen Furniture	40 260	1 738	138 996	6 039

本文采用了标准平均精度 (mean average precision, mAP) 和未知类别召回率 (U-Recall) 两个指标验证模型在 OWOD 任务上的性能。对于每个指标,

其值越大表示性能越好。

本实验使用 $mAP@0.5$ (IoU 阈值取大于 0.5) 作为已知类别目标检测的度量指标,其准确率 (Precision)、召回率 (Recall) 和平均准确率 AP 值和平均准确率均值 mAP 的计算过程如下式所示:

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

$$AP = \int_0^1 Precision(Recall) d(Recall) \quad (17)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (18)$$

其中, TP 为真正例, FP 为假正例, FN 为假负例,准确率 (Precision) 表示模型识别出的正样本占真正样本的比值,召回率 (Recall) 表示模型正确识别为正样本的比值, AP 值表示 Precision 和 Recall 曲线下的面积。 mAP 表示所有类别 $AP@0.5$ 的平均值。

未知类别召回率 $U-Recall$ 作为未知对象检测的评估指标,表示模型预测的未知类别的正确目标占总未知目标的比例。 $U-Recall$ 计算如式 (19) 所示:

$$U-Recall = \frac{U_TP}{U_TP + U_FN} \quad (19)$$

其中, U_TP 表示未知类别预测正确的实例数量, U_FN 为未预测出的未知实例数量。

3.2 参数设置

在 Graph-FPN 模块中,为了提高其性能,首先利用 ImageNet 数据集对该网络进行监督预训练。然后,在 GARO-ORE 的训练过程中利用图卷积神经网络对 Conv3 到 Conv5 提取的多尺度特征充分融合获取更丰富的语义信息。

在 Unknown Aware RPN 模块中,本文参考 ORE 的实验设置,设置伪标签 (未知目标) 的个数最大为 5;在基于平坦极小值的基类学习策略中将式 (4) 中的平衡系数 λ 从 $\{0.1, 1, 10, 100\}$ 中选择 10,将式 (6) 中的超参数 ρ 设置为 0.05,将式 (10) 中的超参数 E 和温度 τ 分别设置为 5 来提高轨迹损失的精度。

在优化方面,使用动量为 0.9 的随机梯度下降 (stochastic gradient descent, SGD),其初始学习率设置为 0.02,随后降低到 0.000 2。mini batch 是从 $\{8, 16, 32\}$ 中选择的 16, NMS (non-maximum suppression) 阈值设置为 0.4。本文在 $\{D_1, D_2, D_3, D_4\}$ 数据集中对每个任务进行 100 000 次迭代训练。在 1 000 次迭代后启

动聚类和执行分类权值初始化方法对增量学习参数进行初始化(基类学习不进行权值初始化),每3 000次迭代之后更新聚类原型。

3.3 实验结果与分析

实验环境: Intel® Xeon® Silver 4310 CPU@2.10 GHz, 64 GB 内存, CentOS7, GPU 处理器为4块 NVIDIA TESLA T4 的独立显卡, 本文在 detectron2 框架下完成对 GARO-ORE 网络模型训练和验证。为了验证 GARO-ORE 模型的有效性, 首先与当前性能最优的 ORE 方法进行了对比实验, 然后进行了3组增量学习实验和时间代价实验, 最后进行消融实验, 以全面考察本文方法对模型性能贡献, 具体实验结果如下。

(1) OWOD 对比实验

本文在 $\{D_1, D_2, D_3, D_4\}$ 数据集上与 ORE 及其他算法进行了对比实验, 其中任务 T_1 表示 GARO-ORE 模型基类学习 D_1 数据集上的所有类别, T_2 、 T_3 和 T_4 分别表示将 D_2 、 D_3 、 D_4 数据以增量学习的形式引入模型中, 实验结果如表2所示。

从表2的实验结果可以看出, 本文方法在 OWOD 任务上取得了最好的结果, 无论在旧类别性能的保留还是新类别的检测性能上都有较好的提升。特别地, 在数据集 $\{D_1, D_2, D_3\}$ 上, GARO-ORE 比最先进的

ORE 在 U-Recall 指标上分别提升了0.84个百分点、0.66个百分点和0.64个百分点, 在 $\{D_1, D_2, D_3, D_4\}$ 数据集上, GARO-ORE 在 mAP 指标上分别获得0.65个百分点、1.74个百分点、1.15个百分点和1.59个百分点提升。这说明本文方法不仅能有效增强模型对未知目标的检测效果, 并且还可以持续高效地学习到新类别知识, 同时保持较好的已知类别识别性能。

(2) 增量学习对比实验

本文在 D_1 数据集上模拟增量学习场景。根据开放世界目标检测的设定, 模型在 $t=1$ 时学习基类 $Class_{base}$, 在 $t=2$ 时模型以增量学习方式引入其余类别 $Class_{next}$ 。在实验中本文按照难度递减的方式设置了3种不同的增量学习实验: ①在 $t=1$ 时刻, 模型学习 D_1 数据集上前10个类别, $t=2$ 时刻学习剩余10个类别。②在 $t=1$ 时刻, 模型学习 D_1 数据集上前15个类别, $t=2$ 时刻学习剩余5个类别。③在 $t=1$ 时刻, 模型学习 D_1 数据集上前19个类别, $t=2$ 时刻学习剩余的1个类别。通过上述3种不同的实验对模型的目标检测增量学习(incremental learning for object detectors, iOD)能力进行评估, 实验结果如表3~表5所示。

表3~表5的实验结果表明, GARO-ORE 在增量目标检测(iOD)任务中表现良好, 在大多数旧类别和

表2 GARO-ORE 在 OWOD 设定下的对比实验结果

Table 2 Comparative experimental results of GARO-ORE under OWOD setting

单位: %

Model	T_1			T_2			T_3			T_4			
	U-Recall	mAP (\uparrow)	U-Recall	mAP (\uparrow)		Both	U-Recall	mAP (\uparrow)		mAP (\uparrow)			
	(\uparrow)	Current known	(\uparrow)	Previously known	Current known		(\uparrow)	Previously known	Current known	Both	Previously known	Current known	Both
Faster R-CNN ^[2]	—	56.16	—	4.076	25.74	14.91	—	6.96	13.48	9.14	2.04	13.68	4.95
FasterRCNN+ Finetuning ^[30]	—	—	—	51.09	23.84	37.47	—	35.69	11.53	27.64	29.53	12.78	25.34
ORE	5.75	56.34	3.66	52.58	25.58	38.98	3.52	37.77	12.41	29.32	30.01	13.44	26.66
GARO-ORE	6.59	56.99	4.32	53.55	27.20	40.72	4.16	38.38	14.02	30.47	31.78	12.93	28.25

表3 GARO-ORE 增量学习的对比实验(10+10)

Table 3 Comparative experiment of GARO-ORE on incremental learning (10+10)

单位: %

Model	aero	cycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	bike	person	plant	sheep	sofa	train	tv	mAP
ILOD ^[16]	69.9	70.4	69.4	54.3	48.0	68.7	78.9	68.4	45.5	58.1	59.7	72.7	73.5	73.2	66.3	29.5	63.4	61.6	69.3	62.2	63.15
ILOD+Faster R-CNN	70.5	75.6	68.9	59.1	56.6	67.6	78.6	75.4	50.3	70.8	43.2	68.1	66.2	65.1	66.5	24.3	61.3	46.6	58.1	49.9	61.14
Faster ILOD ^[15]	72.8	75.7	71.2	60.5	61.7	70.4	83.3	76.6	53.1	72.3	36.7	70.9	66.8	67.6	66.1	24.7	63.1	48.1	57.1	43.6	62.16
ORE-(CC+EBUI) ^[9]	53.3	69.2	62.4	51.8	52.9	73.6	83.7	71.7	42.8	66.8	46.8	59.9	65.5	66.1	68.6	29.8	55.1	51.6	65.3	51.5	59.42
ORE ^[9]	63.5	70.9	58.9	42.9	34.1	76.2	80.7	76.3	34.1	66.1	56.1	70.4	80.2	72.3	81.8	42.7	71.6	68.1	77.0	67.7	64.58
GARO-ORE	62.8	70.1	64.7	44.2	48.3	77.3	79.4	77.6	35.3	70.5	56.8	75.3	81.2	73.4	81.5	43.1	73.8	69.3	75.6	69.1	66.46

表4 GARO-ORE 增量学习的对比实验(15+5)

Table 4 Comparative experiment of GARO-ORE on incremental learning (15+5)

单位: %

Model	aero	cycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	bike	person	plant	sheep	sofa	train	tv	mAP
ILOD ^[16]	70.5	79.2	68.8	59.1	53.2	75.4	79.4	78.8	46.6	59.4	59.0	75.8	71.8	78.6	69.6	33.7	61.5	63.1	71.7	62.2	65.87
ILOD+Faster R-CNN	63.5	76.3	70.7	53.1	55.8	67.1	81.5	80.3	49.6	73.8	62.1	77.1	79.7	74.2	73.9	37.1	59.1	61.7	68.6	61.3	66.35
Faster ILOD ^[15]	66.5	78.1	71.8	54.6	61.4	68.4	82.6	82.7	52.1	74.3	63.1	78.6	80.5	78.4	80.4	36.7	61.7	59.3	67.9	59.1	67.94
ORE-(CC+EBUI) ^[9]	65.1	74.6	57.9	39.5	36.7	75.1	80.0	73.3	37.1	69.8	48.8	69.0	77.5	72.8	76.5	34.4	62.6	56.5	80.3	65.7	62.66
ORE ^[9]	75.4	81.0	67.1	51.9	55.7	77.2	85.6	81.7	46.1	76.2	55.4	76.7	86.2	78.5	82.1	32.8	63.6	54.7	77.7	64.6	68.51
GARO-ORE	76.1	82.5	68.2	51.7	57.4	79.8	85.2	81.9	48.3	80.6	60.1	77.0	86.1	79.9	84.2	33.0	64.5	59.0	78.2	65	69.93

表5 GARO-ORE 增量学习的对比实验(19+1)

Table 5 Comparative experiment of GARO-ORE on incremental learning (19+1)

单位: %

Model	aero	cycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	bike	person	plant	sheep	sofa	train	tv	mAP
ILOD ^[16]	69.4	79.3	69.5	57.4	45.4	78.4	79.1	80.5	45.7	76.3	64.8	77.2	80.8	77.5	70.1	42.3	67.5	64.4	76.7	62.7	68.25
ILOD+Faster R-CNN	60.9	74.6	70.8	56.0	51.3	70.7	81.7	81.5	49.5	78.3	58.3	79.5	79.1	74.8	75.7	42.8	74.7	61.2	67.2	65.1	67.72
Faster ILOD ^[15]	64.2	74.7	73.2	55.5	53.7	70.8	82.9	82.6	51.6	79.7	58.7	78.8	81.8	75.3	77.4	43.1	73.8	61.7	69.8	61.1	68.56
OR-(CC+EBUI) ^[9]	60.7	78.6	61.8	45.0	43.2	75.1	82.5	75.5	42.4	75.1	56.7	72.9	80.8	75.4	77.7	37.8	72.3	64.5	70.7	49.9	64.93
ORE ^[9]	67.3	76.8	60.0	48.4	58.8	81.1	86.5	75.8	41.5	79.6	54.6	72.8	85.9	81.7	82.4	44.8	75.8	68.2	75.7	60.1	68.89
GARO-ORE	69.5	79.1	68.9	49.0	57.6	80.0	88.9	79.8	42.2	80.6	53.5	75.5	87.5	83.0	81.0	45.3	77.6	69.3	77.4	61.0	70.33

新类别的 AP 值有较大的提升,并且在 10+10、15+5、19+1 的增量学习任务中,其 mAP 指标分别提升了 1.38、1.42 和 1.44 个百分点。从旧类别的 AP 值可以看出,绝大多数旧类别 AP 值都有一定的提升,验证了 GARO-ORE 通过求解基类模型平坦极小值,并在其附近的平坦区域内微调参数可有效缓解灾难性遗忘。此外,通过学习的新类别的 AP 值结果可以看出,引入基于知识迁移的新增类别权值初始化的方法,能够有效增强对新增目标类别的建模能力,帮助模型有效保留旧类别性能的同时还能促进后续任务的学习。

(3) 时间代价实验

为了验证 GARO-ORE 算法的检测效率,本文与 Faster R-CNN、ORE 两个基线模型进行对比,测试图像尺寸统一为 800×800,其参数量和推理速度测试结果如表 6 所示。

表6 参数量及推理速度对比

Table 6 Comparison of parameters and reasoning speed

id	Model	Parameters/MB ↓	FPS ↑
1	Faster R-CNN ^[2]	252	14.00
2	ORE ^[9]	257	9.65
3	GARO-ORE	314	8.48

从表 6 实验结果表明, GARO-ORE 相对于 ORE 模型推理速度由原 9.65 FPS 下降到 8.48 FPS。其主要原因是为了优化未知目标的检测和增量学习,本

文引入 Graph-FPN 作为特征提取网络,导致模型参数增加,推理速度下降。相较于 Faster R-CNN 以及 ORE,尽管 GARO-ORE 模型的推理速度有所下降,但是该模型在处理开放世界目标检测问题上实现了较大的性能提升。

(4) 消融实验

为了验证 GARO-ORE 每个关键模块的有效性,本文基于数据集 $\{D_1, D_2, D_3, D_4\}$ 设计了一组消融实验,以详细考察各模块对 OWORD 的影响。实验结果如表 7 所示,其中“√”表示引入相应模块,“×”表示暂不使用相应模块。其中 BLFM (baseclass learning strategy based on flat minima) 表示基于平坦极小值的基类学习策略, WIKT (weight initialization based on knowledge transfer) 表示基于知识迁移的新增类别权值初始化, Pre_Known 为先前类别的 mAP 值, C_Known 为新增类别的 mAP 值。

从表 7 实验结果可以看出, GARO-ORE 在 $\{D_1, D_2, D_3, D_4\}$ 数据集上均获得了较好的性能,其余不完整的模型的性能有不同程度的下降,证明了每个关键模块的重要性,表明所有关键模块都对整体性能有着积极的贡献。以具有代表性数据集 D_1 为例介绍模型各模块实验效果,通过对比实验 1 和实验 2 可以看出,使用 Graph-FPN 代替 ResNet 网络能够有效优化未知目标的检测问题。在基类学习过程中,其未知目标的召回率有 0.56 个百分点的精度提升,并且在

表7 消融实验结果

Table 7 Ablation experimental results

单位:%

id	G-FPN	BLFM	WIKT	U-Recall				Pre_Known				C_Known				Both_mAP			
				D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4
1	×	×	×	6.25	6.98	6.56	6.42	71.78	72.15	71.17	69.45	58.68	59.28	57.42	56.42	68.51	68.93	67.73	66.19
2	√	×	×	6.81	7.42	6.72	6.92	71.95	72.42	71.62	69.13	59.30	59.52	57.31	57.12	68.78	69.20	68.04	66.13
3	×	√	×	6.29	6.96	6.46	6.37	72.80	72.93	72.26	70.84	58.52	59.17	57.12	56.27	69.22	69.49	68.47	67.19
4	×	×	√	6.21	6.91	6.50	6.31	71.59	72.11	71.12	69.21	60.12	60.21	58.12	56.92	68.76	69.13	67.87	66.14
5	×	√	√	6.27	6.95	6.54	6.41	73.09	73.05	72.14	70.42	59.86	60.15	58.04	57.12	69.84	69.82	68.61	67.10
6	√	√	√	6.92	7.36	6.77	6.89	72.96	73.25	72.43	70.41	59.96	60.42	58.09	57.51	69.93	70.04	68.46	67.19

新增类别目标检测中实现了0.62个百分点的mAP精度提升,很好促进模型进行了增量学习。对比实验1和实验3可以看出,通过引入基于平坦极小值的基类学习策略能够较好地保留旧类别的检测性能,相较于ORE,在旧类别的检测中mAP指标提升了1.02个百分点。对比实验1和实验4可以看出,该方法能够有效帮助模型进行增量学习。在增量学习中,新类别的mAP指标提升了1.44个百分点。实验6结果表明本文方法达到了最佳性能。在未知目标召回率达到了6.92%,通过增量学习后的目标检测mAP达到了69.93%。

3.4 可视化分析

为了更直观地展示算法的实验效果,如图7所示,本文选取了三组典型场景图像来对ORE和GARO-ORE做定性比较。

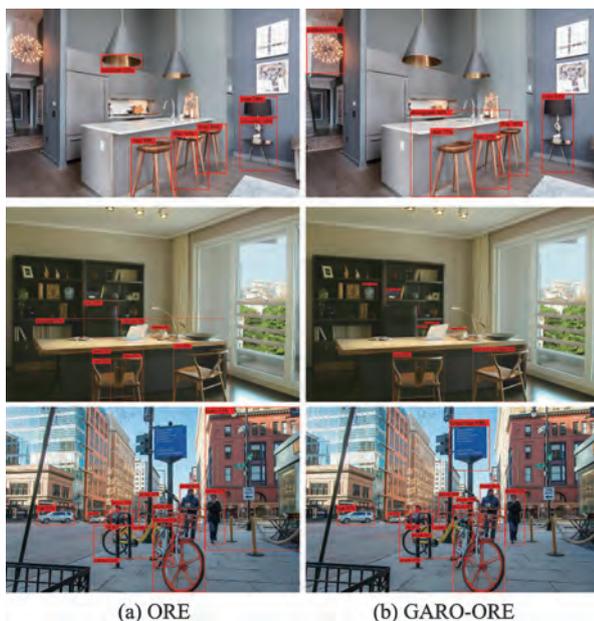


图7 ORE与GARO-ORE检测效果对比

Fig.7 Comparison of detection results between ORE and GARO-ORE

从第一组实验结果中可以看出,与ORE算法相比,GARO-ORE检测到了图像左侧的吊灯和中间的餐桌,并正确识别出右侧的椅子。在第二组实验结果中显示,GARO-ORE将书架中间位置未学习的罐子和书架识别为未知物体。在第三组对比中,ORE错误地将楼层检测成火车,而GARO-ORE并没有出现类似的误判目标。这些结果表明,GARO-ORE模型在OWOD任务上取得了良好的性能,验证了模型的可行性,达到了预期效果。

4 结束语

针对开放世界目标检测任务,提出了一种提升OWOD检测性能的模型GARO-ORE,该模型由包括Graph-FPN、基于平坦极小值的基类学习策略和基于知识迁移的新增类别权值初始化方法等专用组件组成,用于处理开放世界设置下的目标检测,提升了ORE模型在OWOD设定中的检测性能。本文在OWOD数据集上进行了大量实验,对于该数据集上的所有任务设置,GARO-ORE均优于当前性能最优的模型ORE。然而,GARO-ORE提升性能的同时,也导致模型结构更加复杂、庞大。因此在下一步工作中,考虑设计一种轻量化模型进一步提高检测效率和性能。

参考文献:

- [1] 蒋弘毅,王永娟,康锦煜.目标检测模型及其优化方法综述[J].自动化学报,2021,47(6):1232-1255.
JIANG H Y, WANG Y J, KANG J Y. A survey of object detection models and its optimization methods[J]. Acta Automatica Sinica, 2021, 47(6): 1232-1255.
- [2] GIRSHICK R B. Fast R-CNN[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Dec 7-13, 2015. Washington: IEEE Computer Society, 2015:

- 1440-1448.
- [3] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 386-397.
- [4] SINGH B, NAJIBI M, DAVIS L S. SNIPER: efficient multi-scale training[C]//*Advances in Neural Information Processing Systems* 31, Montréal, Dec 3-8, 2018: 9333-9343.
- [5] NIE J, ANWER R M, CHOLAKKAL H, et al. Enriched feature guided refinement network for object detection[C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*, Seoul, Oct 27- Nov 2, 2019. Piscataway: IEEE, 2019: 9536-9545.
- [6] PANG Y, WANG T, ANWER R M, et al. Efficient feature-based image pyramid network for single shot detector[C]//*Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, Jun 16- 20, 2019. Piscataway: IEEE, 2019: 7336-7344.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. *arXiv: 2004.10934*, 2020.
- [8] 李青援, 邓赵红, 罗晓清, 等. 注意力与跨尺度融合的 SSD 目标检测算法[J]. *计算机科学与探索*, 2022, 16(11): 2575-2586.
- LI Q Y, DENG Z H, LUO X Q, et al. SSD object detection algorithm with attention and cross-scale fusion[J]. *Journal of Frontiers of Computer Science and Technology*, 2022, 16(11): 2575-2586.
- [9] JOSEPH K J, KHAN S, KHAN F S, et al. Towards open world object detection[C]//*Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition*, Jun 19-25, 2021. Piscataway: IEEE, 2021: 5830-5840.
- [10] WANG X, HUANG T E, DARRELI T, et al. Frustratingly simple few-shot object detection[J]. *arXiv:2003.06957*, 2020.
- [11] HALL D, DAYOUB F, SKINNER J, et al. Probabilistic object detection: definition and evaluation[C]//*Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision*, Snowmass Village, Mar 1-5, 2020. Piscataway: IEEE, 2020: 1020-1029.
- [12] 高菲, 杨柳, 李晖. 开放集识别研究综述[J]. *南京大学学报(自然科学版)*, 2022, 58(1): 115-134.
- GAO F, YANG L, LI H. A survey on open set recognition [J]. *Journal of Nanjing University (Natural Science)*, 2022, 58(1): 115-134.
- [13] MILLER D, DAYOUB F, MILFORD M, et al. Evaluating merging strategies for sampling-based uncertainty techniques in object detection[C]//*Proceedings of the 2019 International Conference on Robotics and Automation*, Montreal, May 20-24, 2019. Piscataway: IEEE, 2019: 2348-2354.
- [14] DHAMIJA A, GUNTHER M, VENTURA J, et al. The overlooked elephant of object detection: open set[C]//*Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision*, Snowmass Village, Mar 1-5, 2020. Piscataway: IEEE, 2020: 1010-1019.
- [15] PERERA P, MORARIU V I, JAIN R, et al. Generative-discriminative feature representations for open-set recognition[C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, Jun 13-19, 2020. Piscataway: IEEE, 2020: 11811-11820.
- [16] YOSHIHASHI R, SHAO W, KAWAKAMI R, et al. Classification-reconstruction learning for open-set recognition [C]//*Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 4016-4025.
- [17] BENDALE A, BOULT T. Towards open world recognition [C]//*Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, Jun 7-12, 2015. Washington: IEEE Computer Society, 2015: 1893-1902.
- [18] PENG C, ZHAO K, LOVELL B C. Faster ILOD: incremental learning for object detectors based on faster RCNN [J]. *Pattern Recognition Letters*, 2020, 140: 109-115.
- [19] SHMELKOV K, SCHMID C, ALAHARI K. Incremental learning of object detectors without catastrophic forgetting [C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 3420-3429.
- [20] 刘冰瑶, 刘进锋. 增量学习研究综述[J]. *现代计算机*, 2022, 28(13): 72-75.
- LIU B Y, LIU J F. Literature review of incremental learning [J]. *Modern Computer*, 2022, 28(13): 72-75.
- [21] REBUFFI S A, KOLESNIKOV A, SPERL G, et al. iCaRL: incremental classifier and representation learning[C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 5533-5542.
- [22] KANAKIS M, BRUGGEMANN D, SAHA S, et al. Reparameterizing convolutions for incremental multi-task learning without task interference[C]//*LNCS 12365: Proceedings of the 16th European Conference on Computer Vision*, Glasgow, Aug 23-28, 2020. Cham: Springer, 2020: 689-707.
- [23] KIRKPATRICK J, PASCANU R, RABINOWITZ N, et al. Overcoming catastrophic forgetting in neural networks[J]. *Proceedings of the National Academy of Sciences*, 2017,

114(13): 3521-3526.

- [24] ZHAO G M, GE W F, YU Y Z. GraphFPN: graph feature pyramid network for object detection[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Oct 10-17, 2021. Piscataway: IEEE, 2021: 2743-2752.
- [25] MANINIS K K, PONT-TUSET J, ARBELÁEZ P, et al. Convolutional oriented boundaries: from image segmentation to high-level tasks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 819-833.
- [26] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv:1710.10903, 2017.
- [27] CHAUDHARI P, CHOROMANSKA A, SOATTO S, et al. Entropy-SGD: biasing gradient descent into wide valleys [J]. Journal of Statistical Mechanics: Theory and Experiment, 2019(12): 124018.
- [28] SHI G, CHEN J, ZHANG W, et al. Overcoming catastrophic forgetting in incremental few-shot learning by finding flat minima[C]//Advances in Neural Information Processing Systems 34, Dec 6-14, 2021: 6747-6761.
- [29] DU J, ZHOU D, FENG J, et al. Sharpness-aware training for free[C]//Advances in Neural Information Processing Systems 35, New Orleans, Nov 28-Dec 9, 2022: 23439-23451.
- [30] DONG N, ZHANG Y, DING M, et al. Open world DETR:

transformer based open world object detection[J]. arXiv: 2212.02969, 2022.



谢斌红 (1971—), 男, 山西运城人, 硕士, 副教授, CCF 会员, 主要研究方向为智能化软件工程、机器学习。

XIE Binhong, born in 1971, M.S., associate professor, member of CCF. His research interests include intelligent software engineering and machine learning.



张鹏举 (1997—), 男, 山西朔州人, 硕士研究生, 主要研究方向为目标检测、增量学习。

ZHANG Pengju, born in 1997, M.S. candidate. His research interests include object detection and incremental learning.



张睿 (1987—), 男, 山西太原人, 博士, 副教授, CCF 高级会员, 主要研究方向为智能信息处理。

ZHANG Rui, born in 1987, Ph.D., associate professor, senior member of CCF. His research interest is intelligence information processing.