



(12) 发明专利

(10) 授权公告号 CN 118196397 B

(45) 授权公告日 2024.07.19

(21) 申请号 202410584964.9

G06N 3/0464 (2023.01)

(22) 申请日 2024.05.13

G06V 10/764 (2022.01)

G06V 10/82 (2022.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 118196397 A

(43) 申请公布日 2024.06.14

(73) 专利权人 华南理工大学

地址 510641 广东省广州市天河区五山路
381号

(56) 对比文件

Xing Xi et al..KTCN:Enhancing open-world detection with knowledge transfer and class-awareness neutralization.《互联网档案馆》.2024,论文简介以及其对应的Github代码库.

(72)发明人 席星 黄阳阳 罗荣华

审查员 李炯

(74) 专利代理机构 北京慕达星云知识产权代理
事务所(特殊普通合伙)

11465

专利代理师 郭米琪

(51) Int.Cl.

G06V 10/25 (2022.01)

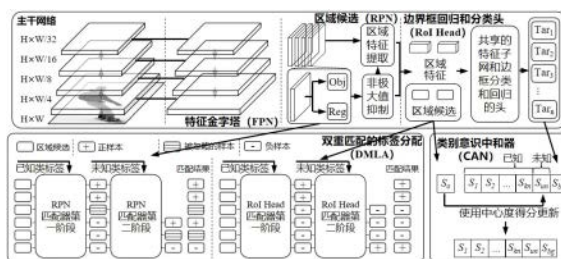
权利要求书2页 说明书10页 附图3页

(54) 发明名称

一种图像目标检测方法及系统

(57) 摘要

本发明公开了一种图像目标检测方法及系统,属于图像处理技术领域,本发明方法中借助大型视觉模型的知识,为未知类别提供监督。使用SAM为潜在物体生成初步伪标签,并通过交并比和最短边界框边缘长度进行精炼。本发明中提出双重匹配标签分配策略,以区分已知和未知类别的标签。此外,还使用类别意识中和器,以减少模型对已知类别的偏见。在开放世界目标检测基准测试上的评估结果表明,本发明方法在未知类别召回率上比先前的SOTA方法具有显著优势;此外,本发明方法不增加任何额外参数,保持了Faster R-CNN的推理速度优势,效率较高。



1. 一种图像目标检测方法,其特征在于,该方法包括:

采用Faster R-CNN作为图像检测的基准模型;

使用SAM对输入图像进行分割,生成潜在目标的伪标签;

在训练阶段,在Faster R-CNN的区域提议网络RPN和边界框微调网络RoI head中采用双重匹配标签分配策略,以区分已知和未知类别的标签;

在推理阶段,利用类别意识中和器调整未知类别的置信度;

所述双重匹配标签分配策略,包括:对于模型的预测结果,优先匹配已知类别的标签;将已知类别中匹配失败的预测结果对伪标签进行二次匹配;再合并两次匹配的结果;

所述类别意识中和器调整未知类别的置信度,具体为:使用区域提议网络RPN的类不可知的中心度得分调整更新模型对未知类别的置信度;将未知类别的置信度更新为未知类别的置信度与候选分数的乘积的算术平方根,公式为:

$$S = \{S_i \in \{S_1, \dots, S_{kn}, S_{un}\} | \sqrt{S_i \cdot S_o}\} \cup \{S_{bg}\}$$

其中, S_o 为RPN的候选分数, S 为预测置信度; S_1-S_{kn} 表示已知类别的置信度, S_{un} 和 S_{bg} 分别表示未知类别的置信度和背景类别的置信度。

2. 根据权利要求1所述的一种图像目标检测方法,其特征在于,所述使用SAM对输入图像进行分割,生成潜在目标的伪标签,具体为:

使用SAM对输入图像进行分割,获得每个对象的类不可知掩码;针对每个掩码计算最大和最小的x、y坐标,获得物体的边界框,经过筛选过滤后,生成潜在目标的伪标签。

3. 根据权利要求2所述的一种图像目标检测方法,其特征在于,利用IOU和最短边过滤的方法筛选过滤边界框;最终保留下来的边界框的类别被标记为未知,生成伪标签。

4. 根据权利要求1所述的一种图像目标检测方法,其特征在于,所述双重匹配标签分配策略中,采用Max IOU匹配已知类别的标签。

5. 根据权利要求1所述的一种图像目标检测方法,其特征在于,在所述边界框微调网络RoIhead中,第一次和第二次匹配均不生成忽略样本集。

6. 根据权利要求1所述的一种图像目标检测方法,其特征在于,使用目标检测指标mAP,评估模型在已知类别上的性能;在未知类别上,使用未知类别的召回率,评估模型对没有标注的对象的性能。

7. 一种图像目标检测系统,其特征在于,应用时执行如权利要求1-6中任一项所述的一种图像目标检测方法,进行图像目标检测,该系统包括:模型选择模块、图像分割模块、标签分配模块和置信度调整模块;其中:

所述模型选择模块,采用Faster R-CNN作为图像检测的基准模型;

所述图像分割模块,使用SAM对输入图像进行分割,生成潜在目标的伪标签;

所述标签分配模块,用于在训练阶段,在Faster R-CNN的区域提议网络RPN和边界框微调网络RoI head中采用双重匹配标签分配策略,以区分已知和未知类别的标签;所述双重匹配标签分配策略,包括:对于模型的预测结果,优先匹配已知类别的标签;将已知类别中匹配失败的预测结果对伪标签进行二次匹配;再合并两次匹配的结果;

所述置信度调整模块,用于在推理阶段,利用类别意识中和器调整未知类别的置信度;具体为:使用区域提议网络RPN的类不可知的中心度得分调整更新模型对未知类别的置信

度;将未知类别的置信度更新为未知类别的置信度与候选分数的乘积的算术平方根,公式:

$$S = \{S_i \in \{S_1, \dots, S_{kn}, S_{un}\} | \sqrt{S_i \cdot S_o}\} \cup \{S_{bg}\}$$

其中, S_o 为RPN的候选分数, S 为预测置信度; S_1 - S_{kn} 表示已知类别的置信度, S_{un} 和 S_{bg} 分别表示未知类别的置信度和背景类别的置信度。

一种图像目标检测方法及系统

技术领域

[0001] 本发明涉及图像处理技术领域,更具体的说是涉及一种图像目标检测方法及系统。

背景技术

[0002] 目标检测作为计算机视觉的基础任务,在很多现实任务上得到应用,如:自动驾驶、医学诊断和工业安全。

[0003] 传统的检测任务在封闭的数据集中进行训练,并且被期望检测出数据集中已经标注的物体。然而,在实际环境中标注所有物体是不可能的。因此,传统的检测模型无法检测出未标注的物体,这限制了其在真实环境中的应用。为解决这一问题,开放世界目标检测(OWOD)任务被提出。开放世界目标检测由于其能召回没有标注物体的特性受到广泛的关注,现有的工作通过启发式先验为模型生成伪标签,但这限制了模型的性能。OWOD任务包括两个主要部分:潜在物体的识别和增量学习。潜在物体的识别要求检测模型发现潜在的物体,而增量学习则在已检测的物体中标记一个子集,并将该子集加入到训练样本中。因此,潜在物体的识别要求模型尽可能多地召回潜在的物体,而增量学习则期望模型在已有知识的基础上进行微调,以尽可能少的代价检测新加入的物体。目前,开放世界目标检测主要存在以下问题:

[0004] 1、如何确定潜在的未知对象。2、如何更好的引入对这些物体的监督而不影响模型在已知类别中的表现。3、如何处理模型对已知物体的偏见。

发明内容

[0005] 有鉴于此,本发明提供至少解决上述部分技术问题的一种图像目标检测方法及系统,本发明采用Faster R-CNN作为基准模型,利用SAM对输入图像进行分割,生成潜在目标的伪标签,为模型提供对潜在未知物体的监督;使用双重匹配标签分配策略,以区分已知和未知类别的标签;使用类别意识中和器,以减少模型对已知类别的偏见;在开放世界目标检测基准测试上的评估结果表明,本发明方法在未知类别召回率上更高;此外,本发明方法不增加任何额外参数,保持了Faster R-CNN的推理速度优势。

[0006] 为实现上述目的,本发明采取的技术方案为:

[0007] 第一方面,本发明提供一种图像目标检测方法,该方法包括:

[0008] 采用Faster R-CNN作为图像检测的基准模型;

[0009] 使用SAM对输入图像进行分割,生成潜在目标的伪标签;

[0010] 在训练阶段,在Faster R-CNN的区域提议网络RPN和边界框微调网络RoIhead中采用双重匹配标签分配策略,以区分已知和未知类别的标签;

[0011] 在推理阶段,利用类别意识中和器调整未知类别的置信度;

[0012] 所述双重匹配标签分配策略,包括:对于模型的预测结果,优先匹配已知类别的标注;将已知类别中匹配失败的部分对伪标签进行二次匹配;再合并两次匹配的结果;

[0013] 所述类别意识中和器调整未知类别的置信度,具体为:使用区域提议网络RPN的类不可知的中心度得分调整更新模型对未知类别的置信度。

[0014] 进一步地,所述使用SAM对输入图像进行分割,生成潜在目标的伪标签,具体为:

[0015] 使用SAM对输入图像进行分割,获得每个对象的类不可知掩码;针对每个掩码计算最大和最小的x、y坐标,获得物体的边界框,经过筛选过滤后,生成潜在目标的伪标签。

[0016] 进一步地,利用IOU和最短边过滤的方法筛选过滤边界框;最终保留下来的边界框的类别被标记为未知,生成伪标签。

[0017] 进一步地,所述双重匹配标签分配策略中,采用Max IOU匹配已知类别的标注。

[0018] 进一步地,在所述边界框微调网络RoI head中,第一次和第二次匹配均不生成忽略样本集。

[0019] 进一步地,使用目标检测指标mAP,评估模型在已知类别上的性能;在未知类别上,使用未知类别的召回率,评估模型对没有标注的对象的性能。

[0020] 第二方面,本发明还提供一种图像目标检测系统,应用于上述的一种图像目标检测方法,进行图像目标检测,该系统包括:模型选择模块、图像分割模块、标签分配模块和置信度调整模块;其中:

[0021] 所述模型选择模块,采用Faster R-CNN作为图像检测的基准模型;

[0022] 所述图像分割模块,使用SAM对输入图像进行分割,生成潜在目标的伪标签;

[0023] 所述标签分配模块,用于在训练阶段,在Faster R-CNN的区域提议网络RPN和边界框微调网络RoI head中采用双重匹配标签分配策略,以区分已知和未知类别的标签;所述双重匹配标签分配策略,包括:对于模型的预测结果,优先匹配已知类别的标注;将已知类别中匹配失败的部分对伪标签进行二次匹配;再合并两次匹配的结果;

[0024] 所述置信度调整模块,用于在推理阶段,利用类别意识中和器调整未知类别的置信度;具体为:使用区域提议网络RPN的类不可知的中心度得分调整更新模型对未知类别的置信度。

[0025] 与现有技术相比,本发明至少具有如下有益技术效果:

[0026] 本发明提供了一种图像目标检测方法及系统,本发明中采用Faster R-CNN作为基准模型,利用SAM对输入图像进行分割,生成潜在目标的伪标签,为模型提供对潜在未知物体的监督;使用双重匹配标签分配策略,以区分已知和未知类别的标签;使用类别意识中和器,以减少模型对已知类别的偏见;在开放世界目标检测基准测试上的评估结果表明,本发明方法在未知类别召回率上具有显著优势;此外,本发明方法不增加任何额外参数,保持了Faster R-CNN的推理速度优势,效率较高。

[0027] 本发明的其它特征和优点将在随后的说明书中阐述,并且,部分地从说明书中变得显而易见,或者通过实施本发明而了解。本发明的目的和其他优点可通过在所写的说明书以及附图中所特别指出的结构来实现和获得。

[0028] 下面通过附图和实施例,对本发明的技术方案做进一步的详细描述。

附图说明

[0029] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍,显而易见地,下面描述中的附图是本申

请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0030] 附图用来提供对本发明的进一步理解,并且构成说明书的一部分,与本发明的实施例一起用于解释本发明,并不构成对本发明的限制。

[0031] 图1为本发明实施例提供的标准的Faster R-CNN结构示意图。

[0032] 图2为本发明实施例提供的ORE方法建立能量模型和基于能量模型更新类别标签的流程示意图。

[0033] 图3为本发明实施例提供的图像目标检测方法的原理示意图。

[0034] 图4为本发明实施例提供的基于Faster R-CNN构建的图像检测网络模型总体结构示意图。

[0035] 图5为本发明实施例提供的伪标签过滤提纯原理示意图。

[0036] 图6为本发明实施例提供的实验结果对比示意图。

[0037] 图7为本发明实施例提供的图像目标检测系统结构示意图。

[0038] 图8为本发明实施例提供的电子设备结构示意图。

具体实施方式

[0039] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。

[0040] 在本发明的描述中,需要说明的是:在本申请说明书及附图等描述的一些流程中,包含了按照特定顺序出现的多个操作,但是应该清楚了解,这些操作可以不按照其在本文中出现的顺序来执行或并行执行。此外,各种序号等仅用于描述目的,而不能理解为指示或暗示相对重要性。

[0041] 因此,以下对在附图中提供的本发明的实施例的详细描述并非旨在限制要求保护的本发明的范围,而是仅仅表示本发明的选定实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0042] 实施例1:

[0043] 本发明实施例提供了一种图像目标检测方法,该方法主要包括:

[0044] (1) 采用Faster R-CNN作为图像检测的基准模型;

[0045] (2) 使用SAM(segment anything model)对输入图像进行分割,生成潜在目标的伪标签;

[0046] (3) 在训练阶段,在Faster R-CNN的区域提议网络RPN(Region Proposal Network)和边界框微调网络RoI(Region of interest)head中采用双重匹配标签分配策略,以区分已知和未知类别的标签;

[0047] (4) 在推理阶段,利用类别意识中和器调整未知类别的置信度。

[0048] 在本实施例中,双重匹配标签分配策略,包括:对于模型的预测结果,优先匹配已知类别的标注;将已知类别中匹配失败的部分对伪标签进行二次匹配;再合并两次匹配的结果。

[0049] 在本实施例中,类别意识中和器调整未知类别的置信度,具体为:使用区域提议网络RPN的类不可知的中心度得分调整更新模型对未知类别的置信度。

[0050] 一、下面对本发明方法涉及的相关原理进行介绍:

[0051] 开放世界目标检测任务的目的是召回那些在数据集中出现的但是没有标注的物体,它最早由ORE提出。为了实现潜在物体的召回,现有研究的重点是如何对潜在的物体生成伪标签。ORE将具有高中心度得分并且与已经标注的物体没有重合的预测标记为潜在的候选样本。UC-OWOD在RPN中匹配失败的候选样本中标记潜在候选样本。OW-DETR基于注意力机制计算中心度得分,挑选那些Top-ku的样本作为伪标签。RE-OWOD和CAT使用选择性搜索生成伪标签。除此之外,一些方法尝试对中心度进行估计。2B-OCD使用第二个基于IOU (Intersection Over Union)的中心度分支去估计中心度,在推理时使用它去确认检测到的已知和未知的实例。PROB使用概率的中心度去调整模型的置信度输出。与上述的方法不同,在本发明中,提出迁移大视觉模型的知识来指导模型对潜在的未知样本的学习。并且,提出了双重匹配标签分配(Dual Matching Label Assignment)策略,以区分已知和未知类别的标签。此外,还提出了类别意识中和器(Class-Awareness Neutralizer),使用类不可知的中心度分数去校准未知类别的置信度,以减少模型对已知类别的偏见。

[0052] 如图1所示,Faster R-CNN是一个传统的检测模型,由三个部分组成:在上游任务(图像分类)中预训练的Backbone(主干网络)、区域提议网络RPN和边界框微调网络RoI head。传统的检测任务在闭集上训练,并且假设图片中不存在任何潜在的物体。因此,训练好的模型不具备检测出未知物体的能力。为了解决这一问题,参见图2所示,ORE利用RPN的类不可知能力标记那些具有高的objectness分数,并且与任何标签样本没有重合的样本为潜在的候选样本。除此之外,为了更好的区分已知样本和未知样本,ORE将已知和未知样本预测的置信度得分保存,并建立一个能量模型。因为,他们假设:模型对已知类别的能量大于未知类别的能量。推理时,如果一个样本被预测为已知类别,但其对应的未知类别的能量大于已知类别的能量,则将其类别更新为未知类别。然而,建立能量模型需要用到完全标记的数据,这导致模型必须接触到潜在实例的标注,造成数据泄露。后续的工作意识到了这个问题,使用deformable detr作为基准模型,并使用注意力机制(OW-DETR)、自适应伪标记机制(CAT)和概率的objectness挖掘潜在的物体样本。然而,Faster R-CNN对未知类别的检测具有天然的优势,它已经作为许多开放词汇或世界目标检测的基线网络。因此,在本发明实施例中,采用Faster R-CNN作为基准模型,构建图像检测网络模型。

[0053] 本实施例中,对开放世界目标检测的研究围绕三个问题展开。1、如何确定潜在的未知对象。2、如何更好的引入对这些物体的监督而不影响模型在已知类别中的表现。3、如何处理模型对已知物体的偏见。

[0054] 为了解决第一个问题,本实施例中利用SAM模型的分割任何物体的能力为图像中的所有物体生成掩码,并将这些掩码作为伪标签,为模型提供对潜在未知物体的监督。如图3所示,对图像中的所有对象进行分割,获得每个对象的类不可知掩码。通过寻找最大和最小的x、y坐标,可以获得物体的包围框。然后,经过筛选后,生成潜在目标的伪标签。

[0055] 第二个问题是由于已知类别和未知类别的竞争导致的。为了引入对潜在类别的监督,使用SAM生成伪标签。然而,这些标签与人工的注释相比是粗糙的。除此之外,当标签的数量变多,一对多标签分配存在边界不清晰的问题,在匹配时不可避免的出现竞争。许多预

测本能被匹配到已知实例却被分配到伪标签；即一个预测可能同时被匹配到多个真值标签。在标签数量较少时，该问题对模型的性能影响较小。然而，SAM对图片中的每一个对象进行分割，生成了大量的候选，这加重了边界模糊的问题；这样的竞争影响模型的性能。为了改善这一现象，本实施例中，提出了双重匹配标签分配 (Dual Matching Label Assignment, DMLA) 策略，该DMLA策略方法通过区别对待伪标签和真值标签解决已知和未知标签的匹配模糊问题。具体来说，本发明认为真值标签错误较少，并且模型在已知类别上的表现完全依赖于对已知类别的监督。而已知类别的监督完全由真值样本的匹配所决定。因此，伪标签的匹配必须不影响真值标签的匹配。对于模型的预测结果，优先匹配已知类别的标注；将已知类别中匹配失败的部分对伪标签进行二次匹配；最后，合并两次匹配的结果。根据上述设计，DMLA能够解决已知类别和未知类别之间的边界冲突问题，提升模型在已知类别上的表现，同时为未知类别提充足的监督信号。

[0056] 进一步地，模型的偏见问题是一个重要的问题，在OWOD任务中，模型最后的输出是类可知的，这导致模型在预测时偏向已知类别。为了解决这一问题，本发明实施例中，使用类别意识中和器 (Class-Awareness Neutralizer, CAN)，以减少模型对已知类别的偏见；即利用RPN的类不可知的中心度得分去解决模型对已知类别的偏见。由于RPN是类不可知的预测，具有高中心度得分的对象有更大的可能是需要检测的物体。因此，本发明中使用中心度得分去更新模型对未知物体的预测置信度。值得注意的是，在训练过程中更新未知类别的置信度并不能缓解偏见问题。因为，模型的优化也是基于类别的标签，这只会导致RPN丧失类不可知的能力。因此，本发明实施例中仅在推理过程中更新未知类别的置信度。

[0057] 在上述方法的优化下，本发明在潜在物体的识别和增量学习的对比实验中以显著的优势超过了之前的SOTA (行业最佳水平state-of-the-art)。由于本发明中的网络是纯的卷积神经网络，没有密集的注意力机制，在运行速度上，模型具有优势。除此之外，本发明中没有引入任何额外的参数，保留了传统的Faster R-CNN的简洁性。

[0058] 二、下面对本发明方法的具体实施方式进行更为详细的介绍：

[0059] 参见图4所示，本发明基于标准的Faster R-CNN with FPN (Feature Pyramid Network) 构建的图像检测网络模型。在训练阶段，在RPN和ROI Head采用所提出的双重匹配标签分配策略，用于解决已知实例和未知实例的边界模糊。在推理阶段，采用类别意识中和器，用于调整未知实例的得分以中和模型对已知类的偏见。

[0060] 在本实施例中，假设模型的输入为 $\{X, Y\}$ 。其中 X 来自于数据集 D 的图像集合 D_x ，而 Y 代表其对应的标签集合。标签集合中含有 n 个标签 y ， $Y = \{y_1, y_2, \dots, y_n\}$ 。标签 y_i 包含类别 l_i 和对应的边界框标注 $b_i = [x_i, y_i, w_i, h_i]$ ，其中 w_i 和 h_i 分别表示 b_i 的宽和高；则有 $y_i = [l_i, b_i]$ 。每一个类别 l_i 都属于类别集合 K ， $l_i \in K$ 。在传统的目标检测任务中，训练集和测试集共享相同的类别集合 K 。而在开放世界目标检测任务设置中，它们是不同的。OWOD由四个任务组成 $T = \{T_1, T_2, T_3, T_4\}$ 。每一个任务的输入图像集合 $D_X^{T_i}$ 均来自于 D ， $D_X^{T_i} \in D$ ，它们之间存在公共部分但不互相包含，即 $D_X^{T_i} \cap D_X^{T_j} \neq \emptyset$ ， $D_X^{T_i} \cup D_X^{T_j} \neq D_X^{T_i}$ 或 $D_X^{T_j}$ ， $i \neq j$ ， $i, j \in [1, 2, 3, 4]$ 。每一个任务标签的类别集合 K_{T_i} 满足前一个任务是后一个任务的真子集， $K_{T_{i-1}} \subset K_{T_i}$ 。所有的任务都共享测试集。测试集的类别集合包含所有训练集的类别集合， $K_{\text{test}} = K = (K_{T_1} \cup K_{T_2} \cup K_{T_3} \cup K_{T_4})$ 。当模型在任务 T_i 上进行训练时，满足 $l \in K_{\text{test}}$ 且 $l \notin K_{T_i}$ 的类别将被视为未知类别。任务

T_1 评估当数据集标注不完整时,模型对潜在物体的召回能力,即潜在物体发现的问题。而 T_2 - T_4 模拟增量学习问题:当标注者标注了新的类别,模型在已有的知识上微调的能力。

[0061] 1.大视觉模型的知识迁移

[0062] 参见图5所示,是伪标签提纯的原理展示。本发明中优选的同时使用IOU和最短边过滤所生成的候选。IOU用于过滤已知类别的标注,而最短边用于过滤噪声。

[0063] OWO任务的键是如何确定潜在的目标并加以监督(ORE, OW-DETR, CAT),因为SAM具有强大的零样本能力,所以本发明迁移SAM的知识为未标记的目标提供伪标签。

[0064] 在本实施例中,采用了一种基于结果等级的知识迁移方法。首先,使用SAM模型对整张图像I进行分割,以获得类别不可知的掩码Masks。表示为 $SAM(I) = \{(x_1, y_1), (x_2, y_2), (x_3, y_3) \dots\}, \dots\}$ 。随后,针对每个掩码计算最大的x和y坐标,以获得每个掩码的包围框集合 $\mathcal{B}_{SAM} = \{b_1, b_2, \dots, b_{n_{sam}}\}$ 。这些包围框可能包含了真值标签(Ground Truth)信息以及噪声。

[0065] 为了避免模糊的标签,利用IOU和最短边过滤的方法生成伪标签。对于每个伪标签,计算其与已知标注的IOU并计算其最大值。若IOU大于阈值 Tr_{IOU} ,则将该标签视为重复标注。对于剩余的包围框,计算其长宽的最小值。若该最小值小于阈值 TR_{sz} ,则将该标签视为噪声,公式为:

$\mathcal{B}_1 = \{b_j \in \mathcal{B}_{SAM} | \left(\max_{i \in [1, n_{gt}]} IOU(b_j, gt_i) \right) < Tr_{IOU}\}$,其中,gt和 n_{gt} 分别代表已知类别的标注和数量。最终,为每个生成的包围框添加了类别标签“unknown”,从而为当前图像中的潜在物体进行标注,即 $Y_{pl} = \{(unknown, b_1), \dots\}$ 。

[0066] 在增量学习的训练阶段,仅计算新加入的标注和伪标签之间的IOU。在微调阶段,对所有类别的重复标注进行了过滤。

[0067] 2.训练流程

[0068] 在本实施例中,图4展示了本发明方法的整体工作原理。该方法基于标准的Faster R-CNN with FPN架构。首先,输入图像经过主干网络和特征金字塔网络(FPN)得到多尺度的特征图。接下来,利用类无关的区域提议网络(RPN)生成候选边界框。根据初始设定,基于RPN生成的物体置信度分数选择前TOP_K个候选框,并对其进行非极大值抑制。剩余的候选框送入边界框微调网络ROI Head进行微调和特定类别的分类。为了迁移SAM的知识,使用SAM对输入图像进行分割并计算所生出掩码的边界框。此外,使用IOU和最短边同时过滤这些边界框。最终,保留下来的边界框的类别将被标记为未知,生成伪标签。为了解决真值样本和伪标签之间的边界问题,RPN和ROI Head中采用了双重匹配标签分配策略进行标签分配。标准的分类损失和边界框损失用于监督匹配结果。

[0069] 双重匹配标签分配策略:

[0070] 标签分配是目标检测任务中的关键步骤,它用于确定模型的预测P和标签Y之间的关系。对于Faster R-CNN这类的两阶段检测器,分类和定位的损失函数由RPN和RoI Head两部分损失的和构成。动态的一对多标签分配存在边界模糊问题,即模型的一个预测 p_i 同时分配给多个标签。尽管手工定义的规则,如MAX IOU和Min Area能缓解这一问题,但伪标签和Ground Truths之间仍然存在竞争。许多样本本能被匹配到已知实例,但由于加入了粗糙的伪标签并且在这样的规则约束下,这些预测被分配到了伪标签。这样的竞争损害了模型的表现。

[0071] 为了应对这些,参见图4左下部分所示,本发明提出双重匹配标签分配策略。首先,对已知类别遵循标准的Max IOU匹配。首先,计算Ground Truth labels(真值标签)与所有预测之间的IOU,并生成IOU矩阵 $\mathbf{M}_{Y_{gt}}^P = \text{IOU}(p_i, gt_j) \in \mathbb{R}^{n_{gt} \times n_p}$, $p_i \in P, gt_j \in Y_{gt}$, n_p 为P的个数;然后,计算每一个预测与已知类别标签的最大值,

$\hat{\mathbf{M}}_{Y_{gt}}^P = \max_{y_j \in Y_{gt}} \text{IOU}(p_i, gt_j) \in \mathbb{R}^{n_p}$, $p_i \in P$;这意味着在的第一维上取最大值。在RPN匹配时,如果这个最大值大于阈值 Tr_h ,那么这个预测将被对待为正样本,小于阈值 Tr_l ,将被对待为负样本,而位于 Tr_l 和 Tr_h 之间的预测将被忽略。最后,根据所得的匹配关系,生成正样本有序对集合 $\mathbf{R}_{\pi_1^+}$,忽略样本有序对集合 $\mathbf{R}_{\pi_1^\sim}$ 和负样本有序对集合 $\mathbf{R}_{\pi_1^-}$ (为了清晰,省略采样和生成额外的匹配过程)。然后收集 $\mathbf{R}_{\pi_1^-}$ 的所有左投影,生成集合 P^- 。进一步地,使用 P^- 和 Y_{pl} 生成对应的正、忽略和负样本集合 $\mathbf{R}_{\pi_1^+}$ 、 $\mathbf{R}_{\pi_1^\sim}$ 和 $\mathbf{R}_{\pi_1^-}$ 。最终,DMLA所生成的有序对匹配集合 $\mathbf{R}_{rpn} = \mathbf{R}_{\pi_1^+} \cup \mathbf{R}_{\pi_1^\sim} \cup \mathbf{R}_{\pi_2^+} \cup \mathbf{R}_{\pi_2^\sim} \cup \mathbf{R}_{\pi_2^-}$ 。在RoI Head阶段,DMLA执行相同的匹配流程。不同的是,第一次和第二次匹配均不生成忽略样本集。因此,DMLA在RoI Head中所产生的有序对集合为 $\mathbf{R}_{roi} = \mathbf{R}_{\pi_1^+} \cup \mathbf{R}_{\pi_2^+} \cup \mathbf{R}_{\pi_2^-}$ 。

[0072] 3.推理流程

[0073] 在本实施例中,与标准的Faster R-CNN一致,在推理时,利用RPN生成区域候选,然后将这些区域候选送入RoI Head微调和分类,最后,非极大值抑制NMS(Non-Maximum-Suppression)将被用于过滤微调后的包围框中重复的预测。与标准的推理流程的不同之处在于,在NMS之前,使用本发明所提出的类别意识中和器去调整未知类别的置信度。

[0074] 类别意识中和器:

[0075] 模型存在的偏见问题表现为对某些类别具有更高的预测可能性。在开放词汇目标检测任务中,模型在训练过程中仅接触到基本的类别,这导致模型对这些基础类别具有更高的置信度。对于开放世界目标检测的任务,模型没有接触未知类别的标注,偏见问题仍然存在于OWOD任务的模型中。尽管引入了SAM来为潜在的类别生成伪标签,但这些标注与已知类别的人工标注相比,其精度较低。因此,模型不可避免的对已知类别预测更大的置信度。为了解决这一问题,提出类别意识中和器,参见图4右下角部分,它利用RPN的类不可知能力去调整未知实例的置信度。

[0076] 具体的,由于RPN的预测是类不可知的形式,所以认为在RPN中objectness分数更高的更可能是一个潜在的物体。假如RPN的候选分数为 $S_o \in [0, 1]$,与它对应的RoI Head的预测置信度 S 。CAN将未知类别的置信度更新为未知类别的置信度与候选分数的乘积的算术平方根,如公式所示: $S = \{S_i \in \{S_1, \dots, S_{kn}, S_{un}\} | \sqrt{S_i \cdot S_o}\} \cup \{S_{bg}\}$, 其中 $S_1 - S_{kn}$ 表示已知类别的置信度,而 S_{un} 和 S_{bg} 则分别对应未知和背景类。与中心度的乘积会降低未知类别的置信度,但同时也会提升中心度高的预测在所有未知样本的绝对排序,这使得它在NMS中有更大的可能被保留。而算术平方根在保留顺序时能够提升这些预测的分数,防止它们由于较低的置信度被过滤掉。

[0077] 三、实验

[0078] 1.数据集

[0079] 在本实施例中,本发明使用VOC和COCO数据集构建了一个混合的数据集,所有的训练和评估都在这个数据集上进行。与先前工作,例如:OW-DETR和CAT一致,遵循OWOD的原始设置,将混合数据集划分为4个任务,每个任务分别的训练图片和训练实例个数为: T_1 (16551, 47223), T_2 (45520, 113741), T_3 (39402, 114452), T_4 (40260, 138996)。任务一包含VOC的类别,然后每个任务在之前任务的基础上添加20种新的类别进行增量实验。测试集包含所有的类别包含10246张图片和61707个测试实例。评估时,所有的已经见过的类别视为已知类别(包括新加入的类别),而剩余的则被视为未知类别。

[0080] 2. 评估指标

[0081] 在本实施例中,本发明遵循标准的OWOD任务评估指标。使用目标检测指标mAP评估模型在已知类别上的性能。在未知类别上,使用未知类别的召回率评估模型对没有标注的对象的性能。在增量学习任务(T_2 - T_4)中,已知类别的性能被分为两部分:之前已经见过的类别和当前任务引入的类别。之前见过的类别用于评估在增量学习任务中新加入类别对已有知识的影响,而当前引入的类别则用于评估模型在新类别上的性能。

[0082] 3. 参数设置

[0083] 实验是基于Detectron2构建的。设定Batch size为8,初始学习率为0.005。为了避免训练初期的不稳定性,在前4000次迭代中将模型的学习率线性增长至初始学习率。最大迭代次数设定为120000,当迭代次数达到60000或100000时,将模型当前的学习率减少到原来的十分之一。在多尺度训练中,图片的最短边缩放范围为220到1088,最长边不超过1333。对于增量学习,设定学习率为0.0005,预热迭代次数为500,最大迭代次数为80000,且不减少学习率。除了 T_1 任务,其他所有任务在训练完成后都会使用replay来回顾已经学习到的知识。

[0084] 4. 与其他SOTA模型比较

[0085] 参见图6所示,本发明方法与其他模型进行了实验比较,包括ORE、ORE-EBUI、UC-OWOD、OCPL、2B-OCPL、OW-DETR、CAT、PROB等,每一列中表现最好的结果已用粗体字表示。

[0086] ORE建立了一个能量模型,该模型将检测出的已知类别更新为未知类别。然而,这个能量模型可能会导致数据泄露,因为它使用了完全标注的类别。因此,我们移除了ORE中的能量模型(被标记为ORE-EBUI),并评估了其性能。尽管CAT通过使用级联的编码器实现了更高的效果,但这会带来大量的计算消耗。因此,我们添加了一个没有级联编码器的版本,被标记为CAT+。

[0087] 参见图6所示,模型对潜在物体发现能力(任务一):在已知类别的表现上,本发明仅落后于CAT 0.4 mAP,但超过了CAT+ 0.1mAP。这微小的差距是可以接受的,因为本发明召回了更多的潜在物体,而且没有使用级联的方法。在没有标注物体的表现上,实现了最高的未知类别召回率40.9。与此最接近的结果是使用级联编码器的CAT,但本发明仍然以17.2召回率的优势大幅度领先。对于CAT+,差距则达到了19.1。对于最新的研究PROB,本发明实现了两倍以上提升,在未知召回率的差距达到了21.5。

[0088] 模型在已有知识上微调的能力(任务二-任务四):在任务二-任务三,本发明的方法实现了全面领先,无论是在未知的召回率还是在已经见过类别上的mAP。在未知召回率上,本发明实现了与CAT的最大差距19.2 (38.4 vs 19.2),而最小的差距仍然达到了15 (39.4 vs 24.4)。在任务四中,本发明的方法超过了CAT,但以0.7 mAP和0.1 mAP的微弱差距

落后于PRON。如果我去掉SAM知识迁移,能实现更好的结果,由于需要预测的实例变少。但这减少了模型对潜在物体发现的能力,这在任务四中没有指标能体现,因为所有的类别都已经接触。因此,本发明仍然保留SAM的伪标签生成,尽管会带来微弱的差距。

[0089] 由上述实施例的描述,本领域技术人员可知:本发明实施例提供了一种图像目标检测方法,具有以下优势:

[0090] ①引入了大规模视觉模型SAM,通过生成伪标签,迁移大模型的知识,加强了模型对潜在的物体的监督,为模型提供了对无标注类别的监督;

[0091] ②为了充分利用大型语言模型的知识,提出了双重匹配标签分配策略(DMLA),以解决一对多标签分配中的边界模糊问题;

[0092] ③为了解决模型的偏见问题,提出类别意识中和器(CAN);以消除模型对已知类别的偏见,该方法无需参与训练,仅在推理时发挥作用;

[0093] ④在Pascal VOC和MS COCO数据集上进行的OWOD任务评估显示,本发明方法具有显著优势,在未知类别的召回率和推理速度上实现了新的行业最佳水平;从而实现了效率和速度上的新SOTA。

[0094] 实施例2:

[0095] 参照图7所示,本发明实施例还提供了一种图像目标检测系统,应用于上述实施例1所描述的一种图像目标检测方法,进行图像目标检测,该系统包括:

[0096] 模型选择模块,用于采用Faster R-CNN作为图像检测的基准模型;

[0097] 图像分割模块,用于使用SAM对输入图像进行分割,生成潜在目标的伪标签;

[0098] 标签分配模块,用于在训练阶段,在Faster R-CNN的区域提议网络RPN和边界框微调网络RoI head中采用双重匹配标签分配策略,以区分已知和未知类别的标签;双重匹配标签分配策略包括:对于模型的预测结果,优先匹配已知类别的标注;将已知类别中匹配失败的部分对伪标签进行二次匹配;再合并两次匹配的结果;

[0099] 置信度调整模块,用于在推理阶段,利用类别意识中和器调整未知类别的置信度;具体为:使用区域提议网络RPN的类不可知的中心度得分调整更新模型对未知类别的置信度。

[0100] 本发明实施例所提供的一种图像目标检测系统,其实现原理及产生的技术效果和前述方法实施例相同,为简要描述,该实施例部分未提及之处,可参考前述方法实施例中相应内容,在此不再赘述。

[0101] 实施例3

[0102] 参照图8所示,本发明实施例还提供了一种用于图像目标检测的电子设备,该电子设备可以包括处理器10、存储器11、通信总线12以及通信接口13,还可以包括存储在存储器11中并可在处理器10上运行的计算机程序。

[0103] 其中,处理器10在一些实施例中可以由集成电路组成,例如可以由单个封装的集成电路所组成,也可以是由多个相同功能或不同功能封装的集成电路所组成,包括一个或者多个中央处理器(Central Processing unit,CPU)、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。处理器10是电子设备的控制核心,利用各种接口和线路连接整个电子设备的各个部件,通过运行或执行存储在存储器11内的程序或者模块,以及调用存储在所述存储器11内的数据,以执行电子设备的各种功能和处理数据。

[0104] 实施例4:

[0105] 本发明实施例还提供一种存储介质,其上存储有计算设备可读的一个或多个程序,一个或多个程序包括指令,指令当由计算设备执行时,使得计算设备执行实施例1中的一种图像目标检测方法。

[0106] 这里的计算机可读存储介质的示例包括:只读存储器(ROM)、随机存取可编程只读存储器(PROM)、电可擦除可编程只读存储器(EEPROM)、随机存取存储器(RAM)、动态随机存取存储器(DRAM)、静态随机存取存储器(SRAM)、闪存、非易失性存储器、CD-ROM、DVD-ROM、蓝光或光盘存储器、硬盘驱动器(HDD)、固态硬盘(SSD)、卡式存储器(诸如,多媒体卡、安全数字(SD)卡或极速数字(XD)卡)、磁带、软盘、磁光数据存储装置、光学数据存储装置、硬盘、固态硬盘以及任何其他装置,上述任何其他装置被配置为以非暂时性方式存储计算机程序以及任何相关联的数据、数据文件和数据结构并将所述计算机程序以及任何相关联的数据、数据文件和数据结构提供给处理器或计算机使得处理器或计算机能执行所述计算机程序。上述计算机可读存储介质中的计算机程序可在诸如客户端、主机、代理装置、服务器等计算机设备中部署的环境中运行,此外,在一个示例中,计算机程序以及任何相关联的数据、数据文件和数据结构分布在联网的计算机系统上,使得计算机程序以及任何相关联的数据、数据文件和数据结构通过一个或多个处理器或计算机以分布式方式存储、访问和执行。

[0107] 本领域内的技术人员应明白,本发明的实施例可提供为方法、系统或计算机程序产品等。因此,本发明可采用完全硬件实施例、完全软件实施例,或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0108] 应当注意的是,词语“包含”不排除存在未列在权利要求中的部件或步骤。位于部件之前的词语“一”或“一个”不排除存在多个这样的部件。本发明可以借助于包括有若干不同部件的硬件以及借助于适当编程的计算机来实现。

[0109] 本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。

[0110] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

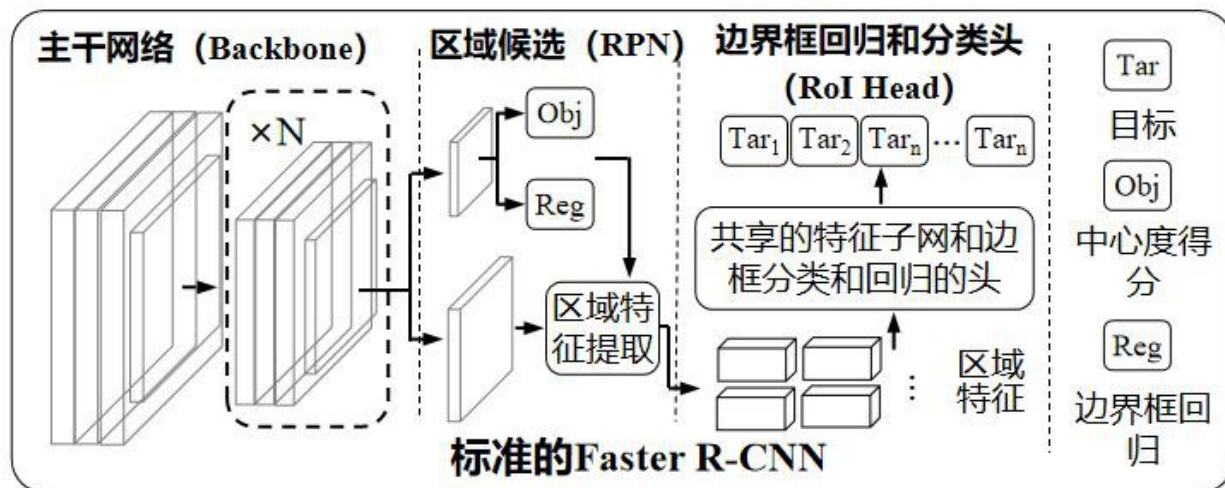


图 1

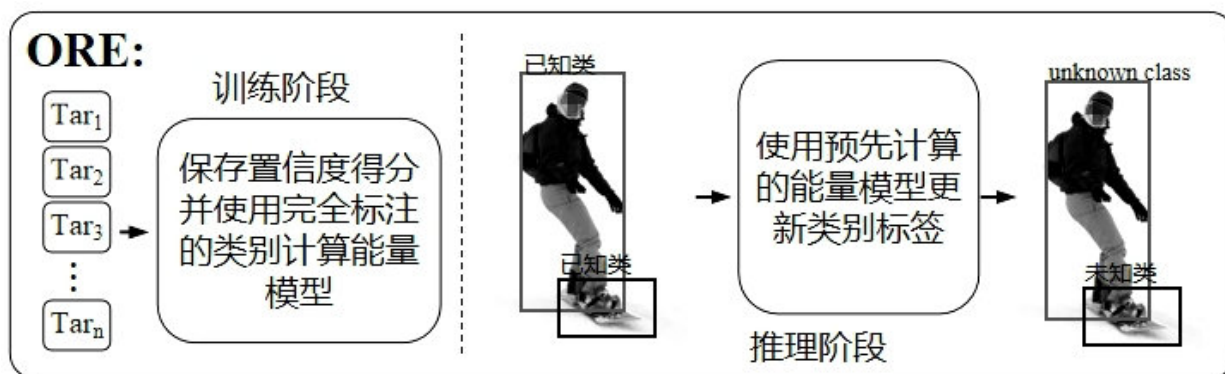


图 2

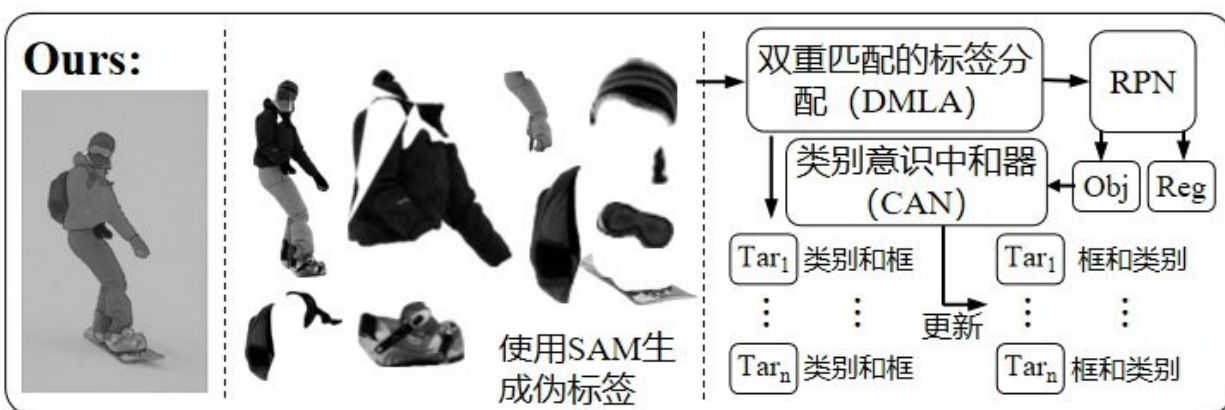


图 3

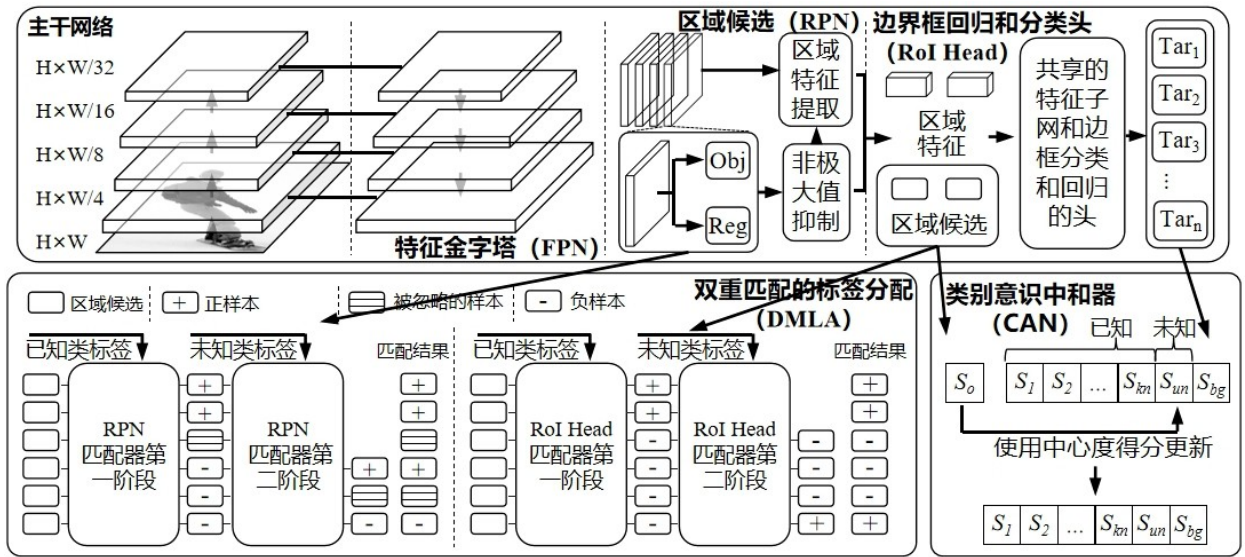


图 4



图 5

任务ID (→)	任务一		任务二				任务三				任务四		
	未知召回率 (↑)	精度 (↑) 当前已知类别	未知召回率 (↑)	之前已知类别	精度 (↑) 当前已知类别	所有已知类别	未知召回率 (↑)	之前已知类别	精度 (↑) 当前已知类别	所有已知类别	之前已知类别	精度 (↑) 当前已知类别	所有已知类别
ORE	8.6	56.3	6.7	52.0	25.2	38.6	6.9	37.9	12.4	29.4	30.1	13.1	25.8
ORE-EBUI	4.9	56.0	2.9	52.7	26.0	39.4	3.9	38.2	12.7	29.7	29.6	12.4	25.3
UC-OWOD	2.4	50.7	3.4	33.1	30.5	31.8	8.7	28.8	16.3	24.6	25.6	15.9	23.2
OCPL	8.3	56.6	7.7	50.6	27.5	39.1	11.9	38.7	14.7	30.7	30.7	14.4	26.7
2B-OCOD	12.1	56.4	9.4	51.6	25.3	38.5	11.6	37.2	13.2	29.2	30.0	13.3	25.8
OW-DETR	7.5	59.2	6.2	53.6	33.5	42.9	5.7	38.3	15.8	30.8	31.4	17.1	27.8
CAT [†]	21.8	59.5	19.2	54.6	32.6	43.6	24.4	42.3	18.9	34.5	34.4	16.6	29.9
CAT	23.7	60.0	19.1	55.5	32.7	40.1	24.4	42.8	18.7	34.8	34.4	16.6	29.9
PROB	19.4	59.5	17.4	55.7	32.2	44.0	19.6	43.0	22.2	36.0	35.7	18.9	31.5
Ours	40.9	59.6	38.4	56.1	36.0	46.0	39.4	43.2	22.9	36.5	35.8	18.2	31.4

图 6

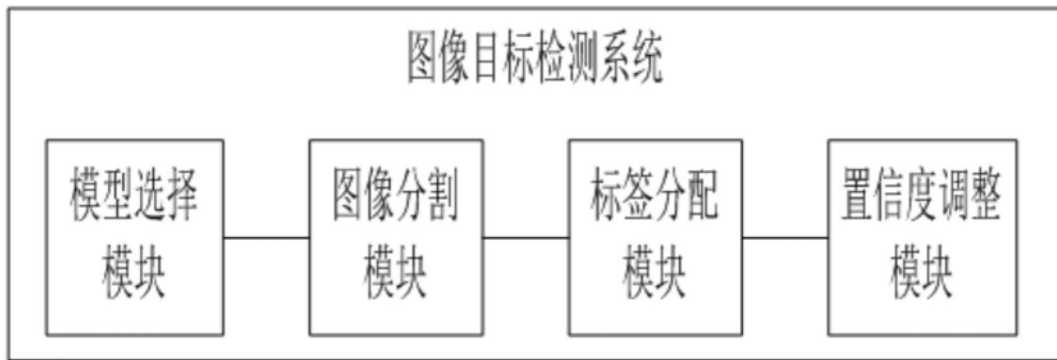


图 7

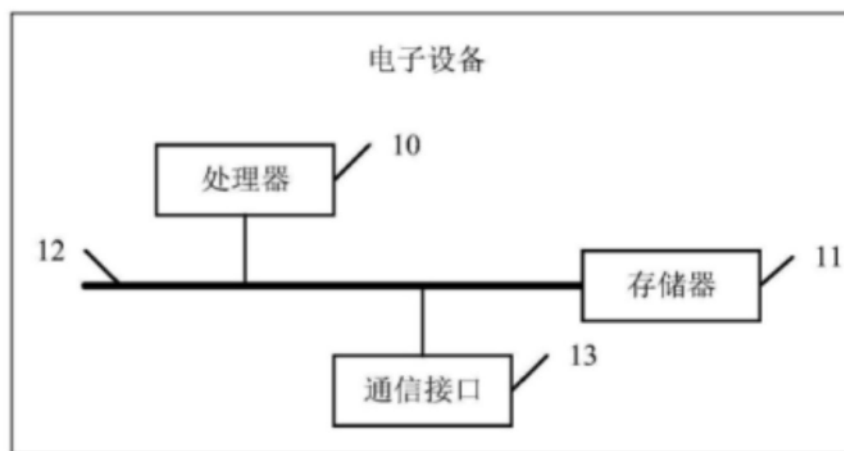


图 8