(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 116884003 B
(45)授权公告日 2024.03.22

(21)申请号 202310880629.9
(22)申请日 2023.07.18
(65)同一申请的已公布的文献号
申请公布号 CN 116884003 A
(43)申请公布日 2023.10.13
(73)专利权人 南京领行科技股份有限公司
地址 211100 江苏省南京市江宁区苏源大道19号九龙湖国际企业总部园B4栋2层(江宁开发区)
(72)发明人 谢奔 朱亮 陈炜
(74)专利代理机构 北京润泽恒知识产权代理有限公司 11319
专利代理师 苏培华
(51)Int.Cl.
G06V 20/70(2022.01)
G06V 10/26(2022.01)
G06V 10/764(2022.01)

(56)对比文件
JP 2022177242 A,2022.11.30
CN 116320524 A,2023.06.23
CN 112258504 A,2021.01.22
CN 113095338 A,2021.07.09
CN 114092707 A,2022.02.25
CN 114463197 A,2022.05.10
CN 114913525 A,2022.08.16
CN 115049817 A,2022.09.13
CN 115129848 A,2022.09.30
CN 115272828 A,2022.11.01
CN 115546630 A,2022.12.30
CN 115983322 A,2023.04.18
CN 116363212 A,2023.06.30
WO 2022121766 A1,2022.06.16
张文丽.基于深度学习的交通标志检测方法研究.《中国优秀硕士学位论文全文数据库 工程科技Ⅱ辑》.2023,(第06期),第C034-415页.

审查员 仁艳秋

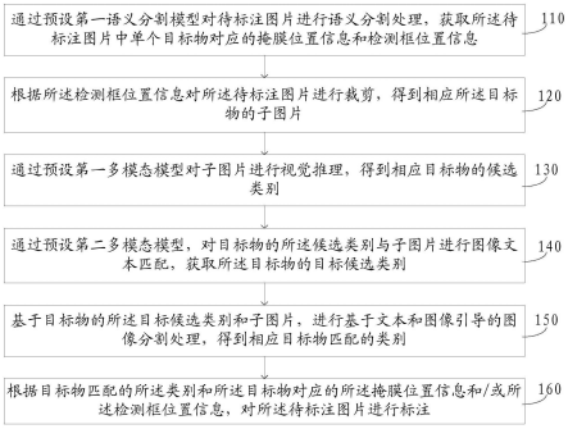
权利要求书3页 说明书11页 附图5页

(54)发明名称

图片自动标注方法、装置、电子设备及存储介质

(57)摘要

本申请公开了一种图片自动标注方法、装置,属于图像处理技术领域。所述方法包括:通过预设第一语义分割模型对待标注图片进行语义分割处理,获取待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息;根据检测框位置信息对待标注图片进行裁剪,得到相应目标物的子图片;通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别;通过预设第二多模态模型,对目标物的候选类别与子图片进行图像文本匹配,获取目标物的目标候选类别;基于目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应标物匹配的类别,用于结合掩膜位置信息和/或检测框位置信息对待标注图片进行标注。本方法提升了图片标注效率。



CN 116884003 B

1.一种图片自动标注方法,其特征在于,所述方法包括:

通过预设第一语义分割模型对待标注图片进行语义分割处理,获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息;

根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片;

通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别;

通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别;

基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别;

根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息,对所述待标注图片进行标注;

其中,所述预设第一多模态模型为:推理语言—图像预训练模型,所述通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别,包括:

基于预设问答提示,通过所述推理语言—图像预训练模型对子图片进行视觉推理,得到相应目标物的候选类别;

所述预设第二多模态模型包括:对比语言—图像预训练模型,所述通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别,包括:

将目标物的所述候选类别的集合与包括所述目标物的子图片输入至所述对比语言—图像预训练模型,获取所述对比语言—图像预训练模型输出的所述子图片与所述集合中各所述候选类别的匹配概率;

选择所述匹配概率最高的最多预设数量所述候选类别,作为所述目标物的目标候选类别。

2.根据权利要求1所述的方法,其特征在于,所述通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别之前,还包括:

通过预设闭集语义分割模型对所述待标注图片进行分割识别,获取所述目标物的补充候选类别;

通过所述补充候选类别对所述候选类别进行扩展,得到扩展后的候选类别。

3.根据权利要求1所述的方法,其特征在于,所述根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片,包括:

根据所述检测框位置信息对所述待标注图片进行多尺度裁剪,得到相应所述目标物的不同尺度的第一子图片、第二子图片和第三子图片;

所述通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别,包括:

通过预设第一多模态模型对第一子图片进行视觉推理,得到相应目标物的候选类别;

所述通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别,包括:

通过预设第二多模态模型,对目标物的所述候选类别与第二子图片进行图像文本匹配,获取所述目标物的目标候选类别;

所述基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别,包括:

基于目标物的所述目标候选类别和第三子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别。

4.根据权利要求2所述的方法,其特征在于,所述通过预设闭集语义分割模型对所述待标注图片进行分割识别,获取所述目标物的补充候选类别,包括:

通过预设闭集语义分割模型对所述待标注图片进行分割处理,得到所述待标注图片中包括的分割区域和所述分割区域对应的目标物类别;

根据所述分割区域和所述掩膜位置信息的对应关系,将所述分割区域对应的目标物类别,作为所述分割区域对应的所述掩膜位置信息所属目标物的补充候选类别。

5.根据权利要求1所述的方法,其特征在于,所述基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别,包括:

将所述目标物的所述目标候选类别和子图片输入至预先训练的基于文本和图像引导的多任务分割模型中进行图像分割处理,得到每个所述目标候选类别匹配的所述子图片中的像素点;

将匹配最多数量像素点的所述目标候选类别,作为相应目标物匹配的类别。

6.一种图片自动标注装置,其特征在于,所述装置包括:

掩膜位置信息获取模块,用于通过预设第一语义分割模型对待标注图片进行语义分割处理,获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息;

子图片获取模块,用于根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片;

候选类别获取模块,用于通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别;

目标候选类别获取模块,用于通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别;

类别和分割信息获取模块,用于基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别;

图片标注模块,用于根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息,对所述待标注图片进行标注;

其中,所述预设第一多模态模型为:推理语言—图像预训练模型,所述候选类别获取模块,还用于基于预设问答提示,通过所述推理语言—图像预训练模型对子图片进行视觉推理,得到相应目标物的候选类别;

所述预设第二多模态模型包括:对比语言—图像预训练模型,所述目标候选类别获取模块,还用于将目标物的所述候选类别的集合与包括所述目标物的子图片输入至所述对比语言—图像预训练模型,获取所述对比语言—图像预训练模型输出的所述子图片与所述集合中各所述候选类别的匹配概率;选择所述匹配概率最高的最多预设数量所述候选类别,作为所述目标物的目标候选类别。

7.根据权利要求6所述的装置,其特征在于,所述装置还包括:

候选类别扩展模块,用于通过预设闭集语义分割模型对所述待标注图片进行分割识别,获取所述目标物的补充候选类别;

所述候选类别扩展模块,还用于通过所述补充候选类别对所述候选类别进行扩展,得到扩展后的候选类别。

8.根据权利要求6所述的装置,其特征在于,所述子图片获取模块,进一步用于:

根据所述检测框位置信息对所述待标注图片进行多尺度裁剪,得到相应所述目标物的不同尺度的第一子图片、第二子图片和第三子图片;

所述候选类别获取模块,进一步用于:

通过预设第一多模态模型对第一子图片进行视觉推理,得到相应目标物的候选类别;

所述目标候选类别获取模块,进一步用于:

通过预设第二多模态模型,对目标物的所述候选类别与第二子图片进行图像文本匹配,获取所述目标物的目标候选类别;

所述类别和分割信息获取模块,进一步用于:

基于目标物的所述目标候选类别和第三子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别。

9.一种电子设备,包括存储器、处理器及存储在所述存储器上并可在处理器上运行的程序代码,其特征在于,所述处理器执行所述程序代码时实现权利要求1至5任意一项所述的图片自动标注方法。

10.一种计算机可读存储介质,其上存储有程序代码,其特征在于,该程序代码被处理器执行时实现权利要求1至5任意一项所述的图片自动标注方法的步骤。

图片自动标注方法、装置、电子设备及存储介质

技术领域

[0001] 本申请涉及图像处理技术领域,特别是涉及图片自动标注方法、装置、电子设备,以及计算机可读存储介质。

背景技术

[0002] 2D(Two Dimensional,二维)语义分割算法广泛应用于图像识别、目标检测等领域。例如,在自动驾驶场景下,通常需要借用相机采集视场内图像,并通过2D语义分割、2D目标检测等算法检测图像中的信息。而2D语义分割、2D目标检测算法需要大量标签数据进行训练。现有技术中,通常是通过人工针对特定任务对图片进行标注。例如,对于2D语义分割任务,需要对图像的每个像素进行目标位置标注与分类。又例如,对于2D目标检测任务,需要对图像中的每个目标物进行检测框标注与分类。通过人工对图片进行标注的方式费时费力,标注成本大。同时,标注速度慢,不利于模型的快速迭代升级。

[0003] 可见,现有技术中的图片标注方法仍需要改进。

发明内容

[0004] 本申请实施例提供一种图片自动标注方法及装置、电子设备及存储介质,能够提升图片标注效率,降低图片标注成本。

[0005] 第一方面,本申请实施例提供了一种图片自动标注方法,包括:

[0006] 通过预设第一语义分割模型对待标注图片进行语义分割处理,获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息;

[0007] 根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片;

[0008] 通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别;

[0009] 通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别;

[0010] 基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别;

[0011] 根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息,对所述待标注图片进行标注。

[0012] 第二方面,本申请实施例提供了一种图片自动标注装置,包括:

[0013] 掩膜位置信息获取模块,用于通过预设第一语义分割模型对待标注图片进行语义分割处理,获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息;

[0014] 子图片获取模块,用于根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片;

[0015] 候选类别获取模块,用于通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别;

[0016] 目标候选类别获取模块,用于通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别;

[0017] 类别和分割信息获取模块,用于基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别;

[0018] 图片标注模块,用于根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息,对所述待标注图片进行标注。

[0019] 第三方面,本申请实施例还公开了一种电子设备,包括存储器、处理器及存储在所述存储器上并可在处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现本申请实施例所述的图片自动标注方法。

[0020] 第四方面,本申请实施例提供了一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时本申请实施例公开的图片自动标注方法的步骤。

[0021] 本申请实施例公开的图片自动标注方法,通过预设第一语义分割模型对待标注图片进行语义分割处理,获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息;根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片;通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别;通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别;基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别;根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息,对所述待标注图片进行标注,实现了对待标注图片进行自动标注,提升了应用于二维图像分割和/或目标物检测场景的图片的标注效率。

[0022] 上述说明仅是本申请技术方案的概述,为了能够更清楚了解本申请的技术手段,而可依照说明书的内容予以实施,并且为了让本申请的上述和其它目的、特征和优点能够更明显易懂,以下特举本申请的具体实施方式。

附图说明

[0023] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合本申请实施例中的附图,对本申请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申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,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,都属于本申请保护的范围。

[0024] 图1是本申请实施例公开的图片自动标注方法流程图之一;

[0025] 图2是本申请实施例公开的图片自动标注方法中图像语义分割结果示意图;

[0026] 图3是本申请实施例公开的图片自动标注方法流程图之二;

[0027] 图4是本申请实施例公开的图片自动标注方法中子图片图像分割结果示意图;

[0028] 图5是本申请实施例公开的图片自动标注方法中标注结果示意图;

[0029] 图6是本申请实施例公开的图片自动标注装置的结构示意图之一;

[0030] 图7是本申请实施例公开的图片自动标注装置的结构示意图之二;

[0031] 图8示意性地示出了用于执行根据本申请的方法的电子设备的框图;以及

[0032] 图9示意性地示出了用于保持或者携带实现根据本申请的方法的程序代码的存储

单元。

具体实施方式

[0033] 下面将结合本申请实施例中的附图,对本申请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申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,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,都属于本申请保护的范围。

[0034] 本申请实施例公开的一种图片自动标注方法,如图1所示,所述方法包括:步骤110至步骤160。

[0035] 步骤110,通过预设第一语义分割模型对待标注图片进行语义分割处理,获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息。

[0036] 所述预设第一语义分割模型可以为通用的语义分割模型。例如,所述预设第一语义分割模型可以为SAM(Segment Anything Model,一类处理图像分割任务的通用模型)模型。

[0037] 对于一张大小H×W的待标注图片P,将该待标注图片P输入至预设第一语义分割模型中,所述预设第一语义分割模型将输出待标注图片P中包括的各个物体(本申请实施例中记为“目标物”)的掩膜位置信息。可选的,预设第一语义分割模型将分割出的单个目标物的掩膜位置信息通过二值掩码表示。

[0038] 进一步的,可以根据所述掩膜位置信息确定掩膜位置的最大外接矩形区域的位置信息,即目标物的最大外接矩形的位置信息,本申请实施例中记为“检测框位置信息”。可选的,最大外接矩形区域的位置信息可以包括:最大外接矩形区域的左上角坐标和右下角坐标。

[0039] 以图2中所示的像素尺寸为H为10,W为10,且包括一个目标物的待标注图片为例,通过预设第一语义分割模型对该待标注图片进行语义分割,可以得到如图2中标记为数字“1”的单个目标物的掩膜位置。相应的,标记为数字“0”的区域表示该部分区域不属于目标物,例如为背景区域。图2中的矩形框210表示单个目标物的最大外接矩形,即目标物的检测框。例如,对于图2中所示的待标注图片,可以得到目标物的最大外接矩形的位置信息为:左上角d_1的像素坐标和右下角d_2的像素坐标,其中,$d_1=(1,2)$,$d_2=(7,9)$。

[0040] 步骤120,根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片。

[0041] 接下来,可以根据待标注图片中包括的每个目标物的所述检测框位置信息对所述待标注图片分别进行裁剪,得到每个所述目标物各自的子图片。例如,对于可以按照图2中所示的目标物的最大外接矩形区域(即检测框)对图2中所示的待标注图片P进行裁剪,得到最大外接矩形区域的图片内容,作为目标物的子图片。

[0042] 本申请的一些实施例中,为了在后续各步骤基于子图片对目标物进行处理时,获取到目标物更多的上下文信息,可以对各目标物的掩膜位置的最大外接矩形区域,即检测框,进行不同比例的外扩,将单个目标物周边的图片内容囊括到目标物的子图片中,进而帮助模型进行准确的识别。

[0043] 可选的,所述根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所

述目标物的子图片,包括:根据所述检测框位置信息对所述待标注图片进行多尺度裁剪,得到相应所述目标物的不同尺度的第一子图片、第二子图片和第三子图片。通过对待标注图片中目标物所在的图片区域进行多尺度裁剪,可以在得到的不同子图片中包括不同的上下文信息,从而帮助后续步骤中采用的多模态模型识别目标物对应类别。具体举例而言,如果语义分割出来的是一个车道线,单从分割结果来看只能看出来是一个矩形区域,但是结合道路信息就可以促使多模态模型正确判定为车道线。

[0044] 可选的,所述检测框位置信息包括:掩膜位置的最大外接矩形区域的位置信息,所述根据所述检测框位置信息对所述待标注图片进行多尺度裁剪,得到相应所述目标物的不同尺度的第一子图片、第二子图片和第三子图片,包括:根据所述最大外接矩形区域的位置信息,获取对所述最大外接矩形区域外扩不同比例后的三个矩形区域的位置信息;按照三个矩形区域的位置信息对所述待标注图片进行分别裁剪,得到每个所述矩形区域的图片,分别作为相应所述目标物的第一子图片、第二子图片和第三子图片。

[0045] 例如,对于待标注图片中的某个目标物,可以对该目标物的所述掩膜位置信息描述的掩膜位置的最大外接矩形区域,在中心点不变的情况下,按照比例r_1、r_2和r_3分别对原面积进行放大,得到大于最大外接矩形区域的三个不同大小的矩形区域。之后,分别将所述待标注图片中每个矩形区域对应图片区域裁剪出来,得到三个矩形图片,分别作为第一子图片、第二子图片和第三子图片。

[0046] 可选的,比例r_1、r_2和r_3为大于1的数值,例如,r_1取值为1.6、r_2取值为1.2、r_3取值为3.0。对所述最大外接矩形区域外扩的不同比例根据待标注图片对应的应用场景确定。

[0047] 步骤130,通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别。

[0048] 接下来,对于进行语义分割得到的每个目标物的一个子图片,通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别。

[0049] 相应的,如果前述步骤裁剪得到了每个目标物的不同尺度的多张子图片,可以选择每个目标物的一张较小尺度的子图片输入至预设第一多模态模型,通过预设第一多模态模型对输入的子图片进行视觉推理,得到相应目标物的候选类别。例如,所述通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别,包括:通过预设第一多模态模型对第一子图片进行视觉推理,得到相应目标物的候选类别。

[0050] 可选的,所述预设第一多模态模型可以为:推理语言—图像预训练模型,所述通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别,包括:基于预设问答提示,通过所述推理语言—图像预训练模型对子图片进行视觉推理,得到相应目标物的候选类别。

[0051] 本申请的一些实施例中,所述推理语言—图像预训练模型可以采用BLIP—2模型。BLIP—2(Bootstrapping Language—Image Pre—training—2)模型,是一种推理语言—图像预训练模型,可以实现视觉问答、视觉推理等功能。在使用过程中,通过将子图片(如第一子图片)和合适的prompt(即问答提示),比如“Question:Which category might this object in the picture belong to?Answer:”输入至BLIP—2模型,BLIP—2模型就会输出对于输入子图片中包括的目标物的与所述问答提示相关的候选类别的集合,例如[“car”,”person”]。

[0052] 需要说明的是,预设问答提示prompt可以根据实际场景进行调整。例如,对于特定推理任务,可能只需要分特定的类别比如人和车,相应的,预设问答提示prompt可以在前面加类别(category)约束,例如["car","person"],这样,推理语言—图像预训练模型就会输出设置的类别集合中的类别。

[0053] 本申请的一些实施例中,为了提升图片分割准确度,在通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别之前,还可以对步骤130中获取的各目标物的候选类别进行进一步补充。

[0054] 如图3所示,所述方法还包括:步骤135和步骤136。

[0055] 步骤135,通过预设闭集语义分割模型对所述待标注图片进行分割识别,获取所述目标物的补充候选类别;

[0056] 步骤136,通过所述补充候选类别对所述候选类别进行扩展,得到扩展后的候选类别。

[0057] 可选的,所述通过预设闭集语义分割模型对所述待标注图片进行分割识别,获取所述目标物的补充候选类别,包括:通过预设闭集语义分割模型对所述待标注图片进行分割处理,得到所述待标注图片中包括的分割区域和所述分割区域对应的目标物类别;根据所述分割区域和所述掩膜位置信息的对应关系,将所述分割区域对应的目标物类别,作为所述分割区域对应的所述掩膜位置信息所属目标物的补充候选类别。

[0058] 为了得到对单个目标物更多的候选类别,本申请的一些实施例中,可以将待标注将图片输入至一个预设闭集语义分割模型,通过该预设闭集语义分割模型对待标注将图片中包括的各目标物进行语义分割和分类,得到一个或多个分割区域,以及每个分割区域匹配的目标物类别。

[0059] 之后,针对每个目标物,将通过第一语义分割模型进行语义分割得到的该目标物的掩膜位置,与通过预设闭集语义分割模型进行语义分割得到的分割区域进行匹配,确定每个目标物对应的分割区域。之后,将每个分割区域对应的目标物类别,作为该分割区域对应的目标物的候选类别,记为"补充候选类别"。

[0060] 可选的,所述预设闭集语义分割模型可以为基于公开的语义分割数据集,例如cityscapes、ade20k等训练得到的通用闭集语义分割模型。可选的,所述预设闭集语义分割模型不限于以下任意一种:oneformer、segformer、Mseg等。

[0061] 之后,将所述补充候选类别补充至所述候选类别的集合中,对所述候选类别进行扩展,得到扩展后的候选类别。具体举例而言,通过设置问答提示,预设第一多模态模型可以识别出包括某个目标物的子图片中该目标物的一个或多个特定的候选类别。而通过预设闭集语义分割模型则可以识别出该目标物的在更大范围的闭集中的候选类别。综合通过上述两种方式得到的该目标物的候选类别,作为该目标物的目标候选类别,用于后续图片和文本匹配,可以扩大图片和文本的匹配范围,从而提升匹配准确度,进而,提升对目标物进行分类的准确度。

[0062] 步骤140,通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别。

[0063] 之后,对于进行语义分割得到的每个目标物的一个子图片,通过预设第二多模态模型对该子图片与前述步骤得到的该目标物的候选类别,进行图像文本匹配,从候选类别

中选择与该子图片匹配的候选类别,作为该子图片中目标物的目标候选类别。

[0064] 本申请的一些实施例中,在未执行对候选类别进行补充的步骤135和步骤136的情况下,通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配时采用的候选类别为执行步骤130后获取的候选类别;在执行对候选类别进行补充的步骤135和步骤136的情况下,通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配时采用的候选类别为执行步骤136后得到扩展后的候选类别。

[0065] 相应的,如果前述步骤裁剪得到了每个目标物的不同尺度的多张子图片,可以选择每个目标物的一张最小尺度的子图片输入至预设第二多模态模型,以通过预设第二多模态模型对输入的子图片和候选类别集合中的候选类别进行图像文本匹配。例如,所述通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别,包括:通过预设第二多模态模型,对目标物的所述候选类别与第二子图片进行图像文本匹配,获取所述目标物的目标候选类别。

[0066] 可选的,所述预设第二多模态模型包括:对比语言-图像预训练模型,所述通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别,包括:将目标物的所述候选类别的集合与包括所述目标物的子图片输入至所述对比语言-图像预训练模型,获取所述对比语言-图像预训练模型输出的所述子图片与所述集合中各所述候选类别的匹配概率;选择所述匹配概率最高的最多预设数量所述候选类别,作为所述目标物的目标候选类别。

[0067] 本申请的一些实施例中,所述对比语言-图像预训练模型可以为CLIP模型。CLIP(Contrastive Language—Image Pre—Training)模型,是一个用于匹配图像和文本的预训练神经网络模型。当将目标物的所述候选类别的集合(例如记为“D1”)与包括所述目标物的子图片(如第二子图片)输入至CLIP模型时,CLIP模型将输出输入至该CLIP模型的子图片中目标物属于集合D1中的各候选类别的匹配概率。

[0068] 本申请的一些实施例中,后续步骤进行图像分割处理时,能够处理的目标候选类别的数量并非无限多,而是预设数量。因此,在获取到所述对比语言-图像预训练模型(如CLIP模型)输出的所述子图片与所述集合中各所述候选类别的匹配概率之后,可以选择所述匹配概率最高的前预设数量所述候选类别,作为所述目标物的目标候选类别。

[0069] 其中,所述预设数量根据应用场景的具体需求确定。例如,所述预设数量可以为3。

[0070] 本申请的一些实施例中,当所述集合中候选类别的数量大于或等于预设数量时,可以选择所述匹配概率最高的前预设数量所述候选类别,作为所述目标物的目标候选类别;当所述集合中候选类别的数量小于预设数量时,可以将所述集合中的所有所述候选类别,作为所述目标物的目标候选类别。

[0071] 步骤150,基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别。

[0072] 之后,对于进行语义分割得到的每个目标物,基于该目标物的一个子图片,以及该目标物的各目标候选类别,进行基于文本(如目标候选类别)和图像(如该目标物的一张子图片)引导的图像分割处理,得到该目标物匹配的图像区域位置信息和类别。

[0073] 相应的,如果前述步骤裁剪得到了每个目标物的不同尺度的多张子图片,可以选择每个目标物的一张最大尺度的子图片,进行基于文本和图像引导的图像分割处理。例如,

所述基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别,包括:基于目标物的所述目标候选类别和第三子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别。

[0074] 可选的,所述基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别,包括:将所述目标物的所述目标候选类别和子图片输入至预先训练的基于文本和图像引导的多任务分割模型中进行图像分割处理,得到每个所述目标候选类别匹配的所述子图片中的像素点;将匹配最多数量像素点的所述目标候选类别,作为相应目标物匹配的类别。

[0075] 本申请的一些实施例中,可以采用CLIPSeg模型,基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理。CLIPSeg模型是哥廷根大学提出的一个使用文本和图像问答提示,能同时作三个分割任务的模型。

[0076] 在应用过程中,对于某个目标物,将该目标物的子图片(如第三子图片)和目标候选类别(如["car","road","sky"])作为问答提示(即prompt),输入至CLIPSeg模型,输入的子图片和目标候选类别将引导CLIPSeg模型中的图像分割任务,对子图片进行图像分割,并输出子图片中存在各目标候选类别的区域位置。当输入至CLIPSeg模型的文本包括3个目标候选类别时,CLIPSeg模型得到的图像分割结果中将最多包括3种类别像素点的区域位置。如图4所示,CLIPSeg模型得到的图像分割结果中,包括数字"1"、"2"和"3"标记的三种类别的像素点,并确定了每种类别的像素点的位置。

[0077] 最后需要计算每种类别的像素点的数量,并取包括像素点数量最多的类别作为相应的单个目标物的类别。如图4所示,数字"1"标记的目标候选类别包括的像素点数量为26,数字"2"标记的目标候选类别包括的像素点数量为12,数字"3"标记的目标候选类别包括的像素点数量为4,因此,该单个目标物属于数字"1"标记的目标候选类别。

[0078] 本领域技术人员应当理解,当采用包括不同数量分割任务的多任务分割模型时,输入的目标候选类别的数量有所不同,多任务分割模型的分割结果中像素点所属类别的最大数量也相应变化。

[0079] 另外,需要说明的是,本申请实施例中对于输入至不同多模态模型的子图片的尺度大小关系不做限制。如果采用目标物的不同尺度的子图片输入至不同多模态模型,将可以获得目标物更加丰富的上下文信息,取得更好的识别和分割效果。

[0080] 步骤160,根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息,对所述待标注图片进行标注。

[0081] 按照上述步骤120至步骤150,可以分别获取到待标注图片中包括的每个目标物匹配的类别。之后,根据步骤110中确定的每个目标物在待标注图片中的所述掩膜位置信息,对待标注图片中所述掩膜位置信息对应的像素点,采用相应目标物匹配的类别进行标注,将不属于任何目标物的像素点标注为背景,从而完成对待标注图片的自动标注。例如,可以得到如图5所示的标注结果。图5中,数字"0"表示背景区域的像素点,数字"1"至数字"5"表示目标物所在区域的像素点,不同数字表示该像素点为不同类别的目标物的掩膜位置。

[0082] 本申请的一些实施例中,还可以根据步骤110中确定的每个目标物在待标注图片中的所述检测框位置信息和相应目标物匹配的类别,对待标注图片进行标注。例如,根据各目标物的类别和检测框(如最大外接矩形)的左上角坐标和右下角坐标,待标注图片中包括

的各目标物的类别和检测框位置,从而完成对待标注图片的目标物检测信息的标注。

[0083] 本申请的一些实施例中,还可以对待标注图片同时标注掩膜位置信息及类别,和检测框位置信息及类别。

[0084] 本申请实施例公开的图片自动标注方法,通过预设第一语义分割模型对待标注图片进行语义分割处理,获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息;根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片;通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别;通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别;基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别;根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息,对所述待标注图片进行标注,实现了对待标注图片进行自动标注,提升了应用于二维图像分割和/或二维目标物检测场景的图片的标注效率。

[0085] 采用本申请实施例公开的图片自动标注方法,通过自动标注待标注图片中目标物的类别和掩膜位置信息,使得标注后的图片可以应用于二维图像分割场景,提升分割模型样本图片的标注效率,有助于分割模型快速迭代;通过自动标注待标注图片中目标物的类别和检测框位置信息,使得标注后的图片可以应用于二维目标检测场景,有效提升了二维目标检测场景中图片标注的效率,有助于二维目标检测模型快速迭代。

[0086] 本申请实施例公开的图片自动标注方法的一个具体实施例中,对于一张待标注图片,首先经过一个通用强大的语义分割模型(如SAM模型),得到待标注图片中所存在的物体目标物的掩膜位置信息和检测框位置信息,其次,分别裁剪出每个目标物的图片,得到各目标物子图片,输入至预设第一多模态模型(如BLIP—2模型),通过设置合适的prompt(即问答提示),预测子图片中目标物所归属的候选类别。另一方面,为了得到对该目标物更多可能类别,将待标注图片输入至一个通用闭集的语义分割模型中得到该子图片中目标物的补充候选类别,并与预设第一多模态模型输出的候选类别合并,作为最终候选类别。接下来,再将子图片和最终候选类别输入至预设第二多模态模型(如CLIP模型),得到最有可能的前预设数量类别。最后,将最有可能的前预设数量类别和子图片输入至基于文本和图像引导的多任务分割模型(如CLIPSeg模型),预测子图片中属于这前预设数量类别的图片区域,并取覆盖像素点数量最多的类别作为目标物的最终类别。依次遍历所有分割出的目标物,即可完成一张待标注图片的自动标注。

[0087] 本申请实施例公开的图片自动标注方法,可以有效提升图片标注速度,提升标注效率,降低图片标注成本。另一方面,本申请实施例公开的图片自动标注方法,通过综合预设第一多模态模型和通用闭集语义分割模型分别获取的目标物的候选类别,通过增加候选类别的数量,提升类别和图片匹配的范围和精度,从而提升图像分割和/或目标物检测的准确度。

[0088] 本申请实施例还公开了一种图片自动标注装置,如图6所示,所述装置包括:

[0089] 掩膜位置信息获取模块610,用于通过预设第一语义分割模型对待标注图片进行语义分割处理,获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息;

[0090] 子图片获取模块620,用于根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片;

[0091] 候选类别获取模块630,用于通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别;

[0092] 目标候选类别获取模块640,用于通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别;

[0093] 类别和分割信息获取模块650,用于基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配类别;

[0094] 图片标注模块660,用于根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息,对所述待标注图片进行标注。

[0095] 可选的,如图7所示,所述装置还包括:

[0096] 候选类别扩展模块635,用于通过预设闭集语义分割模型对所述待标注图片进行分割识别,获取所述目标物的补充候选类别;

[0097] 所述候选类别扩展模块635,还用于通过所述补充候选类别对所述候选类别进行扩展,得到扩展后的候选类别。

[0098] 可选的,所述子图片获取模块620,进一步用于:

[0099] 根据所述检测框位置信息对所述待标注图片进行多尺度裁剪,得到相应所述目标物的不同尺度的第一子图片、第二子图片和第三子图片;

[0100] 所述候选类别获取模块630,进一步用于:

[0101] 通过预设第一多模态模型对第一子图片进行视觉推理,得到相应目标物的候选类别;

[0102] 所述目标候选类别获取模块640,进一步用于:

[0103] 通过预设第二多模态模型,对目标物的所述候选类别与第二子图片进行图像文本匹配,获取所述目标物的目标候选类别;

[0104] 所述类别和分割信息获取模块650,进一步用于:

[0105] 基于目标物的所述目标候选类别和第三子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别。

[0106] 可选的,所述通过预设闭集语义分割模型对所述待标注图片进行分割识别,获取所述目标物的补充候选类别,包括:

[0107] 通过预设闭集语义分割模型对所述待标注图片进行分割处理,得到所述待标注图片中包括的分割区域和所述分割区域对应的目标物类别;

[0108] 根据所述分割区域和所述掩膜位置信息的对应关系,将所述分割区域对应的目标物类别,作为所述分割区域对应的所述掩膜位置信息所属目标物的补充候选类别。

[0109] 可选的,所述类别和分割信息获取模块650,进一步用于:

[0110] 将所述目标物的所述目标候选类别和子图片输入至预先训练的基于文本和图像引导的多任务分割模型中进行图像分割处理,得到每个所述目标候选类别匹配的所述子图片中的像素点;

[0111] 将匹配最多数量像素点的所述目标候选类别,作为相应目标物匹配的类别。

[0112] 本申请实施例公开的图片自动标注装置,用于实现本申请实施例中所述的图片自

动标注方法,装置的各模块的具体实施方式不再赘述,可参见方法实施例相应步骤的具体实施方式。

[0113] 本申请实施例公开的图片自动标注装置,通过预设第一语义分割模型对待标注图片进行语义分割处理,获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息;根据所述检测框位置信息对所述待标注图片进行裁剪,得到相应所述目标物的子图片;通过预设第一多模态模型对子图片进行视觉推理,得到相应目标物的候选类别;通过预设第二多模态模型,对目标物的所述候选类别与子图片进行图像文本匹配,获取所述目标物的目标候选类别;基于目标物的所述目标候选类别和子图片,进行基于文本和图像引导的图像分割处理,得到相应目标物匹配的类别;根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息,对所述待标注图片进行标注,实现了对待标注图片进行自动标注,提升了应用于二维图像分割和/或二维目标物检测场景的图片的标注效率,降低了人工标注成本。

[0114] 另一方面,本申请实施例公开的图片自动标注装置,通过综合预设第一多模态模型和通用闭集语义分割模型分别获取的目标物的候选类别,通过增加候选类别的数量,提升类别和图片匹配的范围和精度,从而提升图像分割的准确度。

[0115] 本说明书中的各个实施例均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的部分互相参见即可。对于装置实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0116] 以上对本申请提供的一种图片自动标注方法及装置进行了详细介绍,本文中应用了具体个例对本申请的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本申请的方法及其一种核心思想;同时,对于本领域的一般技术人员,依据本申请的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本申请的限制。

[0117] 以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0118] 本申请的各个部件实施例可以以硬件实现,或者以在一个或者多个处理器上运行的软件模块实现,或者以它们的组合实现。本领域的技术人员应当理解,可以在实践中使用微处理器或者数字信号处理器(DSP)来实现根据本申请实施例的电子设备中的一些或者全部部件的一些或者全部功能。本申请还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序(例如,计算机程序和计算机程序产品)。这样的实现本申请的程序可以存储在计算机可读介质上,或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到,或者在载体信号上提供,或者以任何其他形式提供。

[0119] 例如,图8示出了可以实现根据本申请的方法的电子设备。所述电子设备可以为PC机、移动终端、个人数字助理、平板电脑等。该电子设备传统上包括处理器810和存储器820及存储在所述存储器820上并可在处理器810上运行的程序代码830,所述处理器810执行所

述程序代码830时实现上述实施例中所述的方法。所述存储器820可以为计算机程序产品或者计算机可读介质。存储器820可以是诸如闪存、EEPROM(电可擦除可编程只读存储器)、EPROM、硬盘或者ROM之类的电子存储器。存储器820具有用于执行上述方法中的任何方法步骤的计算机程序的程序代码830的存储空间8201。例如,用于程序代码830的存储空间8201可以包括分别用于实现上面的方法中的各种步骤的各个计算机程序。所述程序代码830为计算机可读代码。这些计算机程序可以从一个或者多个计算机程序产品中读出或者写入到这一个或者多个计算机程序产品中。这些计算机程序产品包括诸如硬盘,紧致盘(CD)、存储卡或者软盘之类的程序代码载体。所述计算机程序包括计算机可读代码,当所述计算机可读代码在电子设备上运行时,导致所述电子设备执行根据上述实施例的方法。

[0120] 本申请实施例还公开了一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现如本申请实施例所述的图片自动标注方法的步骤。

[0121] 这样的计算机程序产品可以为计算机可读存储介质,该计算机可读存储介质可以具有与图8所示的电子设备中的存储器820类似布置的存储段、存储空间等。程序代码可以例如以适当形式进行压缩存储在所述计算机可读存储介质中。所述计算机可读存储介质通常为如参考图9所述的便携式或者固定存储单元。通常,存储单元包括计算机可读代码830',所述计算机可读代码830'为由处理器读取的代码,这些代码被处理器执行时,实现上面所描述的方法中的各个步骤。

[0122] 本文中所称的“一个实施例”、“实施例”或者“一个或者多个实施例”意味着,结合实施例描述的特定特征、结构或者特性包括在本申请的至少一个实施例中。此外,请注意,这里“在一个实施例中”的词语例子不一定全指同一个实施例。

[0123] 在此处所提供的说明书中,说明了大量具体细节。然而,能够理解,本申请的实施例可以在没有这些具体细节的情况下被实践。在一些实例中,并未详细示出公知的方法、结构和技术,以便不模糊对本说明书的理解。

[0124] 在权利要求中,不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本申请可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中,这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

[0125] 最后应说明的是:以上实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。

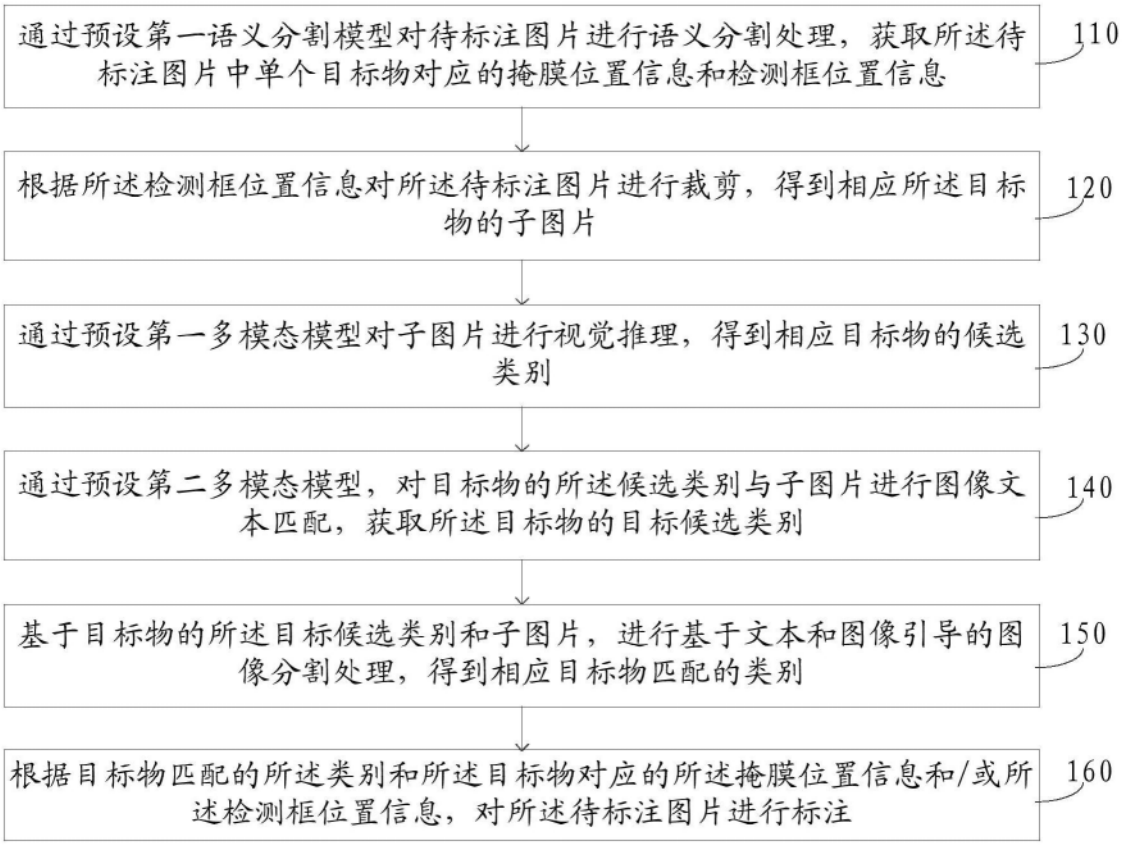

通过预设第一语义分割模型对待标注图片进行语义分割处理，获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息
110
根据所述检测框位置信息对所述待标注图片进行裁剪，得到相应所述目标物的子图片
120
通过预设第一多模态模型对子图片进行视觉推理，得到相应目标物的候选类别
130
通过预设第二多模态模型，对目标物的所述候选类别与子图片进行图像文本匹配，获取所述目标物的目标候选类别
140
基于目标物的所述目标候选类别和子图片，进行基于文本和图像引导的图像分割处理，得到相应目标物匹配的类别
150
根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息，对所述待标注图片进行标注
160

图1

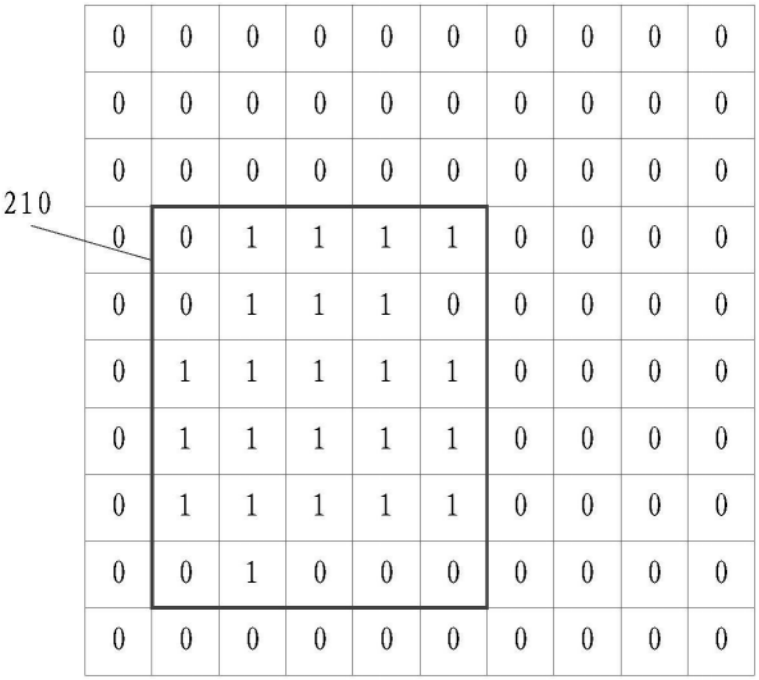

210
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 1 1 1 1 0 0 0 0
0 0 1 1 1 0 0 0 0 0
0 1 1 1 1 1 0 0 0 0
0 1 1 1 1 1 0 0 0 0
0 1 1 1 1 1 0 0 0 0
0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0

图2

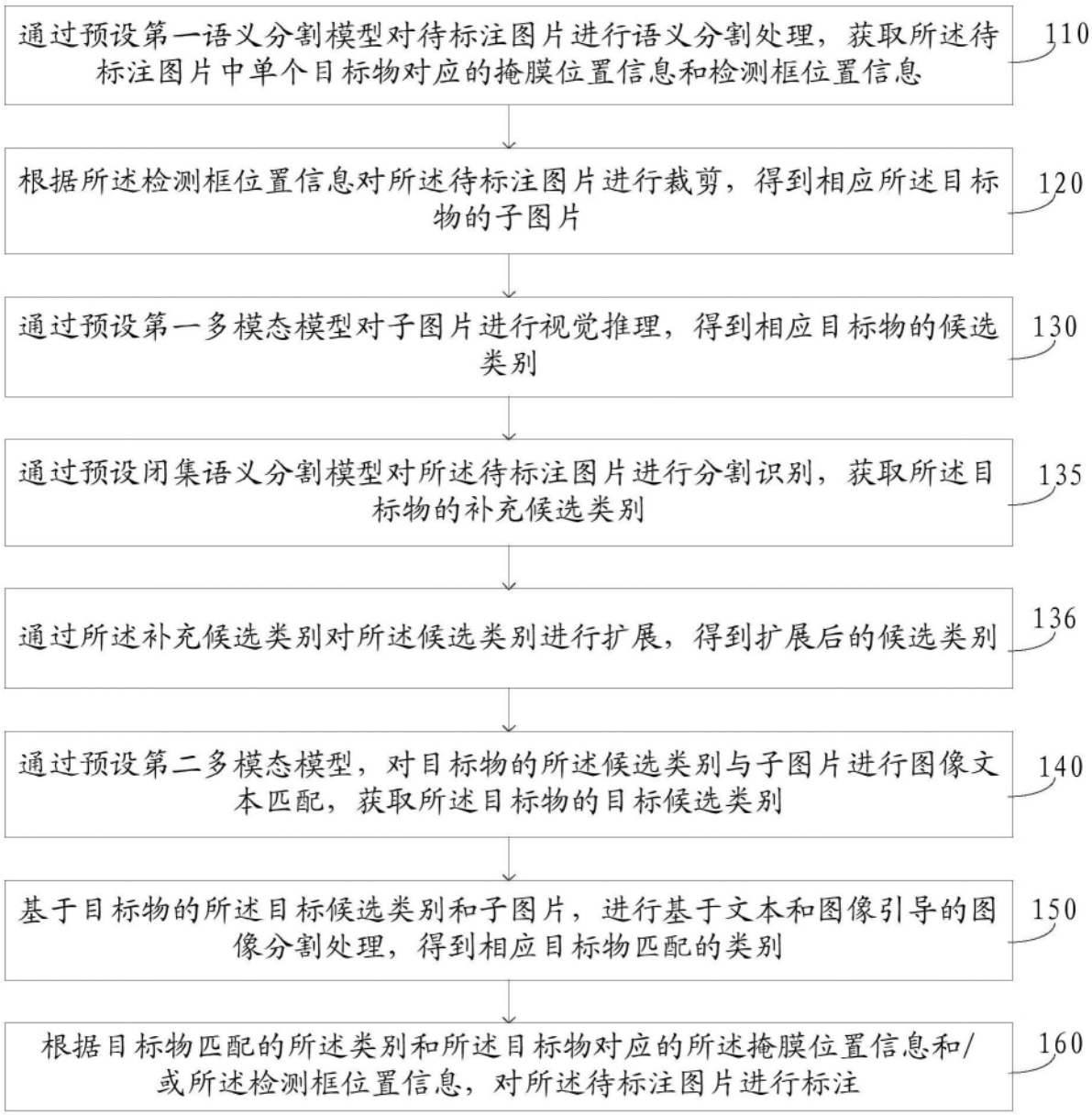

通过预设第一语义分割模型对待标注图片进行语义分割处理，获取所述待标注图片中单个目标物对应的掩膜位置信息和检测框位置信息
110
根据所述检测框位置信息对所述待标注图片进行裁剪，得到相应所述目标物的子图片
120
通过预设第一多模态模型对子图片进行视觉推理，得到相应目标物的候选类别
130
通过预设闭集语义分割模型对所述待标注图片进行分割识别，获取所述目标物的补充候选类别
135
通过所述补充候选类别对所述候选类别进行扩展，得到扩展后的候选类别
136
通过预设第二多模态模型，对目标物的所述候选类别与子图片进行图像文本匹配，获取所述目标物的目标候选类别
140
基于目标物的所述目标候选类别和子图片，进行基于文本和图像引导的图像分割处理，得到相应目标物匹配的类别
150
根据目标物匹配的所述类别和所述目标物对应的所述掩膜位置信息和/或所述检测框位置信息，对所述待标注图片进行标注
160

图3

2	1	1	1	1
2	1	1	1	2
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
2	1	3	3	3

图4

2	2	2	2	2	2	4	4	4	0
2	2	2	2	2	2	4	4	4	4
2	2	2	2	2	2	4	4	4	4
2	2	1	1	1	1	4	4	4	4
2	2	1	1	1	2	2	2	2	4
0	1	1	1	1	1	2	2	2	2
0	1	1	1	1	1	2	2	2	2
2	1	1	1	1	1	3	5	5	5
2	2	1	3	3	3	3	5	5	5
0	0	0	3	3	3	3	5	5	0

图5

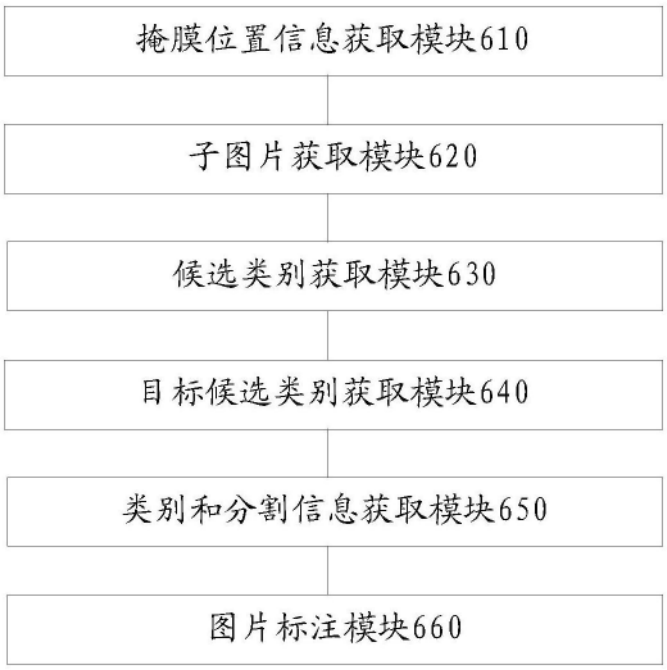

掩膜位置信息获取模块610
子图片获取模块620
候选类别获取模块630
目标候选类别获取模块640
类别和分割信息获取模块650
图片标注模块660

图6

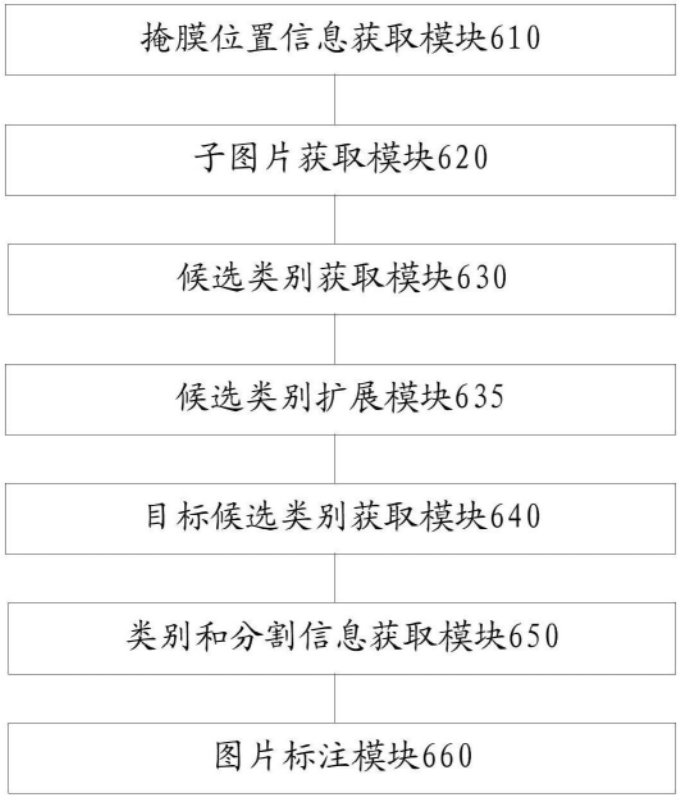

掩膜位置信息获取模块610
子图片获取模块620
候选类别获取模块630
候选类别扩展模块635
目标候选类别获取模块640
类别和分割信息获取模块650
图片标注模块660

图7

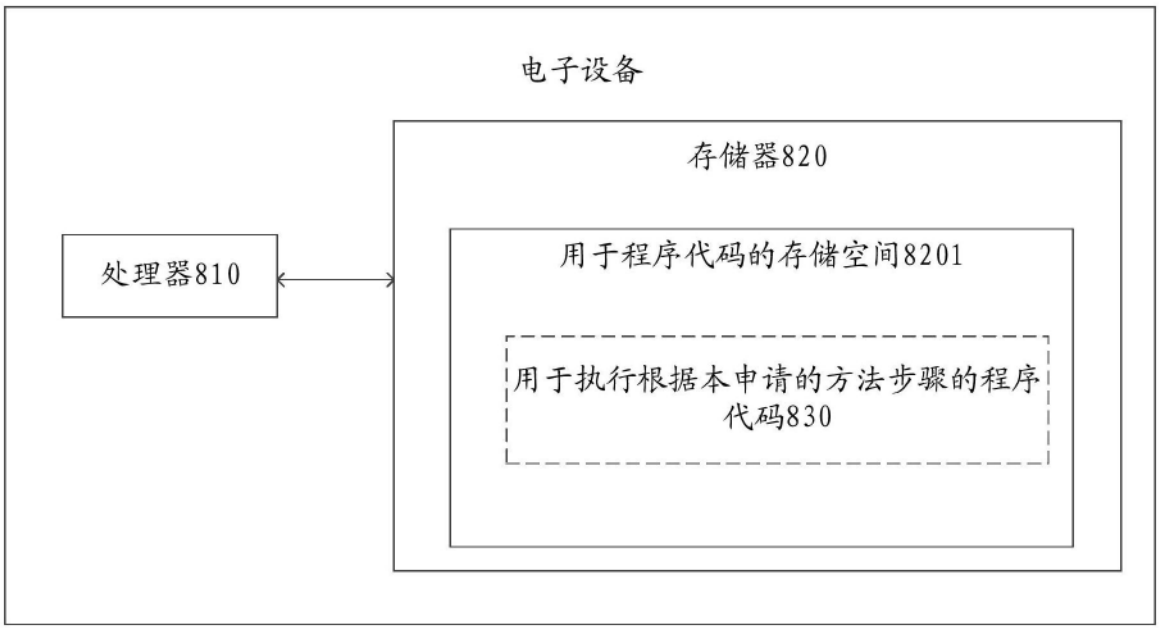

电子设备
存储器820
处理器810
用于程序代码的存储空间8201
用于执行根据本申请的方法步骤的程序代码830

图8


存储单元
用于执行根据本申请的方法步骤的程序代码830′

图9