



(12) 发明专利申请

(10) 申请公布号 CN 115546566 A

(43) 申请公布日 2022. 12. 30

(21) 申请号 202211483266.7

(22) 申请日 2022.11.24

(71) 申请人 杭州心识宇宙科技有限公司  
地址 310000 浙江省杭州市余杭区仓前街  
道仓兴路1号22幢3楼302号

(72) 发明人 陈阳 石翔飞 舒会 顾宇鑫  
鲍方毅

(51) Int. Cl.  
G06V 10/764 (2022.01)  
G06V 10/74 (2022.01)  
G06V 10/82 (2022.01)

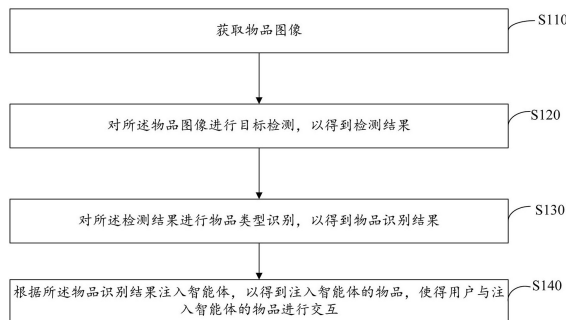
权利要求书2页 说明书9页 附图4页

(54) 发明名称

基于物品识别的智能体交互方法、装置、设备及存储介质

(57) 摘要

本发明实施例公开了基于物品识别的智能体交互方法、装置、设备及存储介质。所述方法包括：获取物品图像；对所述物品图像进行目标检测，以得到检测结果；对所述检测结果进行物品类型识别，以得到物品识别结果；根据所述物品识别结果注入智能体，以得到注入智能体的物品，使得用户与注入智能体的物品进行交互。通过实施本发明实施例的方法可实现物品真正站在物品的视角与用户进行交互。



1. 基于物品识别的智能体交互方法,其特征在于,包括:  
获取物品图像;  
对所述物品图像进行目标检测,以得到检测结果;  
对所述检测结果进行物品类型识别,以得到物品识别结果;  
根据所述物品识别结果注入智能体,以得到注入智能体的物品,使得用户与注入智能体的物品进行交互。
2. 根据权利要求1所述的基于物品识别的智能体交互方法,其特征在于,所述对所述物品图像进行目标检测,以得到检测结果,包括:  
采用目标检测模型对所述物品图像进行目标检测,以得到检测结果。
3. 根据权利要求2所述的基于物品识别的智能体交互方法,其特征在于,所述目标检测模型是以YOLOV5模型为基础,通过模型剪枝、通道剪枝、减少Anchor的方式,结合非极大值抑制算法实现的。
4. 根据权利要求1所述的基于物品识别的智能体交互方法,其特征在于,所述对所述检测结果进行物品类型识别,以得到物品识别结果,包括:  
对所述检测结果进行物品视觉特征提取,以得到提取结果;  
根据所述提取结果进行特征检索,以得到检索结果;  
根据所述检索结果确定物品类型,以得到物品识别结果。
5. 根据权利要求4所述的基于物品识别的智能体交互方法,其特征在于,所述对所述检测结果进行物品视觉特征提取,以得到提取结果,包括:  
对所述检测结果采用基于MobileNetv2的识别网络进行物品特征提取,以得到提取结果。
6. 根据权利要求4所述的基于物品识别的智能体交互方法,其特征在于,所述根据所述提取结果进行特征检索,以得到检索结果,包括:  
通过开源的FAISS框架建立特征检索引擎,并选择乘积量化优化索引,检索与所述提取结果相似的目标物品,以得到检索结果。
7. 根据权利要求4所述的基于物品识别的智能体交互方法,其特征在于,所述根据所述检索结果确定物品类型,以得到物品识别结果,包括:  
筛选符合要求的检索结果,以得到筛选结果;  
对所述筛选结果采用非极大值抑制算法消除边界框,以得到物品识别结果。
8. 基于物品识别的智能体交互装置,其特征在于,包括:  
图像获取单元,用于获取物品图像;  
目标检测单元,用于对所述物品图像进行目标检测,以得到检测结果;  
识别单元,用于对所述检测结果进行物品类型识别,以得到物品识别结果;  
注入单元,用于根据所述物品识别结果注入智能体,以得到注入智能体的物品,使得用户与注入智能体的物品进行交互。
9. 一种计算机设备,其特征在于,所述计算机设备包括存储器及处理器,所述存储器上存储有计算机程序,所述处理器执行所述计算机程序时实现如权利要求1至7中任一项所述的方法。
10. 一种存储介质,其特征在于,所述存储介质存储有计算机程序,所述计算机程序被

处理器执行时实现如权利要求1至7中任一项所述的方法。

## 基于物品识别的智能体交互方法、装置、设备及存储介质

### 技术领域

[0001] 本发明涉及AR技术,更具体地说是指基于物品识别的智能体交互方法、装置、设备及存储介质。

### 背景技术

[0002] 在动画电影中,物品能记得你们共同参与过的活动、你的喜好,在你犹豫不决的时候给你建议,在你难过的时候用它们的方式宽慰你。在现实生活中,要实现这个功能,则需要解决两个关键问题,一是如何借助现有硬件,识别现实中的物品并与之交互;二是如何让物品真正站在物品的视角与人们进行开放域对话,即不局限在特定的专业领域的对话,现有的物品可参与到实际生活中的例子为客服机器人等交互智能体。

[0003] 目前常见的交互智能体如客服机器人一般直接嵌入手机应用、网页中与用户交互。客服机器人通常采用知识图谱、基于意图识别的自然语言处理技术等方式实现,这让客服机器人在限定知识范围内能做出应答,而超出限定知识范围后无法应答,也就是无法进行开放域对话。

[0004] 因此,有必要设计一种新的方法,实现物品真正站在物品的视角与用户进行交互。

### 发明内容

[0005] 本发明的目的在于克服现有技术的缺陷,提供基于物品识别的智能体交互方法、装置、设备及存储介质。

[0006] 为实现上述目的,本发明采用以下技术方案:基于物品识别的智能体交互方法,包括:

获取物品图像;

对所述物品图像进行目标检测,以得到检测结果;

对所述检测结果进行物品类型识别,以得到物品识别结果;

根据所述物品识别结果注入智能体,以得到注入智能体的物品,使得用户与注入智能体的物品进行交互。

[0007] 其进一步技术方案为:所述对所述物品图像进行目标检测,以得到检测结果,包括:

采用目标检测模型对所述物品图像进行目标检测,以得到检测结果。

[0008] 其进一步技术方案为:所述目标检测模型是以YOLOV5模型为基础,通过模型剪枝、通道剪枝、减少Anchor的方式,结合非极大值抑制算法实现的。

[0009] 其进一步技术方案为:所述对所述检测结果进行物品类型识别,以得到物品识别结果,包括:

对所述检测结果进行物品视觉特征提取,以得到提取结果;

根据所述提取结果进行特征检索,以得到检索结果;

根据所述检索结果确定物品类型,以得到物品识别结果。

[0010] 其进一步技术方案为:所述对所述检测结果进行物品视觉特征提取,以得到提取结果,包括:

对所述检测结果采用基于MobileNetv2的识别网络进行物品特征提取,以得到提取结果。

[0011] 其进一步技术方案为:所述根据所述提取结果进行特征检索,以得到检索结果,包括:

通过开源的FAISS框架建立特征检索引擎,并选择乘积量化优化索引,检索与所述提取结果相似的目标物品,以得到检索结果。

[0012] 其进一步技术方案为:所述根据所述检索结果确定物品类型,以得到物品识别结果,包括:

筛选符合要求的检索结果,以得到筛选结果;

对所述筛选结果采用非极大值抑制算法消除边界框,以得到物品识别结果。

[0013] 本发明还提供了基于物品识别的智能体交互装置,包括:

图像获取单元,用于获取物品图像;

目标检测单元,用于对所述物品图像进行目标检测,以得到检测结果;

识别单元,用于对所述检测结果进行物品类型识别,以得到物品识别结果;

注入单元,用于根据所述物品识别结果注入智能体,以得到注入智能体的物品,使得用户与注入智能体的物品进行交互。

[0014] 本发明还提供了一种计算机设备,所述计算机设备包括存储器及处理器,所述存储器上存储有计算机程序,所述处理器执行所述计算机程序时实现上述的方法。

[0015] 本发明还提供了一种存储介质,所述存储介质存储有计算机程序,所述计算机程序被处理器执行时实现上述的方法。

[0016] 本发明与现有技术相比的有益效果是:本发明通过对物品图像进行目标检测和物品类型识别,根据识别所得的类型注入智能体,以使得物品具备智能体的功能,实现物品真正站在物品的视角与用户进行交互。

[0017] 下面结合附图和具体实施例对本发明作进一步描述。

## 附图说明

[0018] 为了更清楚地说明本发明实施例技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,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。

[0019] 图1为本发明实施例提供的基于物品识别的智能体交互方法的应用场景示意图;

图2为本发明实施例提供的基于物品识别的智能体交互方法的流程示意图;

图3为本发明实施例提供的基于物品识别的智能体交互方法的子流程示意图;

图4为本发明实施例提供的基于物品识别的智能体交互方法的子流程示意图;

图5为本发明实施例提供的目标检测模型的流程示意图;

图6为本发明实施例提供的基于物品识别的智能体交互装置的示意性框图;

图7为本发明实施例提供的基于物品识别的智能体交互装置的识别单元的示意性框图;

图8为本发明实施例提供的基于物品识别的智能体交互装置的确定子单元的示意性框图；

图9为本发明实施例提供的计算机设备的示意性框图。

### 具体实施方式

[0020] 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。基于本发明中的实施例,本领域普通技术人员在没有做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。

[0021] 应当理解,当在本说明书和所附权利要求书中使用时,术语“包括”和“包含”指示所描述特征、整体、步骤、操作、元素和/或组件的存在,但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

[0022] 还应当理解,在此本发明说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本发明。如在本发明说明书和所附权利要求书中所使用的那样,除非上下文清楚地指明其它情况,否则单数形式的“一”、“一个”及“该”意在包括复数形式。

[0023] 还应当进一步理解,在本发明说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合,并且包括这些组合。

[0024] 请参阅图1和图2,图1为本发明实施例提供的基于物品识别的智能体交互方法的应用场景示意图。图2为本发明实施例提供的基于物品识别的智能体交互方法的示意性流程图。该基于物品识别的智能体交互方法应用于服务器中。该服务器与终端进行数据交互,通过终端拍摄物品的图像,对图像进行物品检测和类型识别,根据识别的类型注入对应的智能体,从而实现物品真正站在物品的视角与用户进行交互。

[0025] 图2是本发明实施例提供的基于物品识别的智能体交互方法的流程示意图。如图2所示,该方法包括以下步骤S110至S140。

[0026] S110、获取物品图像。

[0027] 在本实施例中,物品图像是指使用终端或者带有摄像头的设备拍摄需要进行交互的物品所形成的图像。

[0028] S120、对所述物品图像进行目标检测,以得到检测结果。

[0029] 在本实施例中,检测结果是针对物品图像进行物品目标所在的位置进行检测所形成的边界框以及位置信息。

[0030] 具体地,采用目标检测模型对所述物品图像进行目标检测,以得到检测结果。

[0031] 具体地,所述目标检测模型是以YOLOV5模型为基础,通过模型剪枝、通道剪枝、减少Anchor的方式,结合非极大值抑制算法实现的。

[0032] 在本实施例中,如图5所示,输入图片通过模型主干网络得到最终预测检测框,具体地,经过模型颈部模块处理后再进行检测框预测。实例中选择以YOLOV5作为基准的单阶段目标检测模型;在YOLOV5基础上通过模型剪枝+通道剪枝+减少Anchor(候选检测框)的方式,基于MobileNetv2(一种轻量级识别网络)从通道深度与宽度两个尺度分别采用0.5和0.33作为裁剪比例进行通道剪枝,从而在尽量减少性能丢失的前提下大幅度降低模型推理的时间消耗。由于在产品所在的应用场景中,待检测的物体往往是常见尺寸的自然物体,出

现的大尺寸物体或者小尺度的物体的概率较低,结合使用场景,对网络结构做了适当的裁剪,仅保留原YOLOV5输出的三个不同尺寸特征的中间特征,裁剪特征金字塔中最大与最小特征图上的Anchor,通过Anchor裁剪有效提升模型推理效率。在保证精度的前提下大幅提高模型推理速度最后使用常用的NMS(非极大值抑制算法,Non Max Suppression)作为后处理方式,消除冗余的边界框,找到最合适、置信度最高的物品。

[0033] S130、对所述检测结果进行物品类型识别,以得到物品识别结果。

[0034] 在本实施例中,物品识别结果是指物品的类型识别的结果。

[0035] 在一实施例中,请参阅图3,上述的步骤S130可包括步骤S131~S133。

[0036] S131、对所述检测结果进行物品视觉特征提取,以得到提取结果。

[0037] 在本实施例中,所有提取结果均为物品的视觉特征。

[0038] 具体的,对所述检测结果均采用MobileNetv2识别网络进行物体视觉特征提取,并得到提取结果。MobileNetv2是业界常用的设备级目标识别与检测任务非常有效的特征提取器,在参数较少的前提下仍能保持较优的模型性能。

[0039] 在本次实施例中,基于COCO数据集与自建数据集,自建数据集经过数据增强操作(旋转、场景贴图、噪声、镜像),以MobileNetv2网络结构作为backbone(模型主干网络)训练了一个物体识别网络,从而得到一个推理性能较快,且拥有较强表征能力的特征提取模型。

[0040] 基于MobileNetv2的网络结构,针对网络结构做了适当的修正和优化,实现物体类别识别任务。

[0041] S132、根据所述提取结果进行特征检索,以得到检索结果。

[0042] 在本实施例中,检索结果是指根据提取结果与各个类型的对应的特征进行比对形成的结果。

[0043] 具体地,通过开源的FAISS框架建立特征检索引擎,并选择乘积量化优化索引,检索与所述提取结果相似的目标物品,以得到检索结果。

[0044] 具体地,检索结果包括若干个特征与提取结果相近似的带类型的物品。

[0045] 基于MobileNetv2网络结构提取得到的视觉特征簇,为了兼顾了检索速度和准确率的要求,选择了倒排表+乘积量化作为索引策略,通过开源的FAISS框架构建特征检索引擎,在检索阶段通过计算特征相似度查询topK(得分前K个)的候选分类结果,选择在K中出现次数最多的物品类别作为分类结果;若出现次数相同,选择其中相似度最高的类别作为结果。

[0046] FAISS是Facebook AI团队开源的针对聚类和相似性搜索库,提供高效相似度搜索和聚类的向量搜索能力。其中主要用到了乘积量化的优化查询方法,其原理是将原始向量分解成若干个低维向量的笛卡尔积,并对分解得到的低维特征做量化,这种方式在性能与效率上都远胜过聚类、直接量化索引。

[0047] S133、根据所述检索结果确定物品类型,以得到物品识别结果。

[0048] 在一实施例中,请参阅图4,上述的步骤S133可包括步骤S1331~S1332。

[0049] S1331、筛选符合要求的检索结果,以得到筛选结果。

[0050] 在本实施例中,筛选结果是指特征相似度排列在前N名的检索结果,其中,N可根据实际情况确定。

[0051] S1332、对所述筛选结果采用非极大值抑制算法消除边界框,以得到物品识别结果。

[0052] 在本实施例中,通过物品类别和根类别的关系设计topK排序规则,得到最相似的图片的标签作为最终的识别结果,其次,使用非极大值抑制算法消除冗余的边界框,找到最合适、置信度最高的物品识别结果。

[0053] S140、根据所述物品识别结果注入智能体,以得到注入智能体的物品,使得用户与注入智能体的物品进行交互。

[0054] 在本实施例中,根据物品的类型给物品注入自我认知,让物品站在自己的视角跟用户对话。基于脑启发心识框架和基座模型(Foundation Model)。基座模型是一个大型的人工智能模型,它在大量的无标签数据上进行大规模的训练,从而形成一个可以适应广泛的下游任务的模型。基础模型的早期例子是大型预训练的语言模型,包括BERT(Bidirectional Encoder Representations from Transformers,是2018年10月由Google AI研究院提出的一种预训练模型,与GPT-3(Generative Pre-trained Transformer 3,由Open AI团队提出的一种自回归语言模型。随后,一些多模态的基础模型被生产出来),包括DALL-E(即2021年1月Open AI提出的一种图片生成模型)、Flamingo(即2022年4月DeepMind团队提出的一种基于小样本学习的视觉语言模型)和Florence(即微软在2021年11月最新提出的视觉基础模型)。脑启发心识框架将基座模型作为通用计算单元,结合了脑区(包括记忆、感知、认知等)的设计,可以灵活地赋予智能体不同的“自我认知”。

[0055] 具体地,基座模型是一个大型的人工智能模型,它在大量的无标签数据上进行大规模的训练,从而形成一个可以适应广泛的下游任务的模型。基础模型的早期例子是大型预训练的语言模型,包括BERT和GPT-3。随后,一些多模态的基础模型被生产出来,包括DALL-E、Flamingo和Florence。基座模型使用了decoder only(仅使用Transformer编码器模块)的transformer模型结构,基于开源的互联网数据等进行自回归的预训练。脑启发心识框架将基座模型作为通用计算单元,构建起了多个脑区:记忆脑区、感知脑区、认知脑区。其中记忆脑区负责智能体的知识、设定、交互过程等的存储与使用。包含长记忆模块与短记忆模块。每个模块中,又包含了关系记忆,情节记忆,语义记忆等。通过基座模型与提示句的融合,记忆模块可以支持从用户与智能体交互的过程中沉淀记忆,并从长期记忆中召回与对话相关的记忆。

[0056] 感知脑区主要负责接收视觉信息的输入,并对视觉信息加以理解,以供后续的信息整合处理作为内容物料。具体的,感知脑区首先通过Clip模型的图像编码器,将视觉空间映射到统一语义空间,然后通过多层感知机得到能够被后续流程接收的符号化结果。认知脑区将上述符号化信息进行接收,结合由前文中得到的物品类别识别结果,形成复合提示句。

[0057] 此外,智能体还可以具备物品相关的知识,比如价格、购买地点等。用户可以通过与智能体交互,询问商品相关信息。

[0058] 在交互时,可以通过语音或文字的方式与智能体交互。用户可以对摄像机拍摄的任意物品进行检测和识别。物品识别的结果将影响智能体的自我认知,智能体可以“物品”视角跟用户交互,给用户带来新奇的对话体验;借助物品识别,商家可以通过后端系统给物品注入关键的商品信息,在用户与该物品交互时,可以给用户传达关键的商品信息,引导用

户购买。让用户可以与身边的任意物品进行对话交互，且物品能站在物品的视角与用户进行开放域对话。

[0059] 对于物品注入智能体与用户的实际交互过程，举个例子，第一步，AR扫描：在AR扫描页面中，小程序本地的AI目标检测模型会时刻检测相机窗口中的主体物体。当用户点击“连接异次元”的按钮后，可开启会话，与AI虚拟心识进行聊天。

[0060] 第二步：开启会话，上个步骤中位于黄色探测框中的物品，会经过后台的抠图服务，去除杂乱背景，将主体显著图精准地抠出来，展示在小程序前端。同时，会对黄色探测框中的物品图像进行细分类，如此例中，物品图像被准确识别为“鼠标”。

[0061] 开启会话时，服务后端会根据后台配置的规则配置AI虚拟心识的个性类别（此例中为“做作”）、稀有度（此例中为“R”）和此次会话的背景图片、虚拟心识的五官表情（眼睛、眉毛、嘴）。一切资源准备就绪后开启会话。

[0062] 开启会话后，AI系统会基于物品分类（在此例中为“鼠标”）和随机来进行开场白。用户可以打字或者语音输入和虚拟心识聊天。

[0063] 第三步：分享，对话页面中截屏，App会根据当前会话内容自动生成海报，供用户分享。此图片中右下角的二维码支持其他用户在微信App扫描后进入万物总动员字节小程序，和被分享的虚拟心识进行会话。

[0064] 第四步：探索记录，在AR扫描页面的右下角，点击“探索记录”按钮，进入“探索记录”页。探索记录页会记录最近四次的心识聊天记录。点击对应的物品图片，重新开启与此虚拟心识的会话。

[0065] 上述的基于物品识别的智能体交互方法，通过对物品图像进行目标检测和物品类型识别，根据识别所得的类型注入智能体，以使得物品具备智能体的功能，实现物品真正站在物品的视角与用户进行交互。

[0066] 图6是本发明实施例提供的一种基于物品识别的智能体交互装置300的示意性框图。如图6所示，对应于以上基于物品识别的智能体交互方法，本发明还提供一种基于物品识别的智能体交互装置300。该基于物品识别的智能体交互装置300包括用于执行上述基于物品识别的智能体交互方法的单元，该装置可以被配置于服务器中。具体地，请参阅图6，该基于物品识别的智能体交互装置300包括图像获取单元301、目标检测单元302、识别单元303以及注入单元304。

[0067] 图像获取单元301，用于获取物品图像；目标检测单元302，用于对所述物品图像进行目标检测，以得到检测结果；识别单元303，用于对所述检测结果进行物品类型识别，以得到物品识别结果；注入单元304，用于根据所述物品识别结果注入智能体，以得到注入智能体的物品，使得用户与注入智能体的物品进行交互。

[0068] 在一实施例中，所述目标检测单元302，用于采用目标检测模型对所述物品图像进行目标检测，以得到检测结果。

[0069] 在一实施例中，如图7所示，所述识别单元303包括特征提取子单元3031、特征检索子单元3032以及确定子单元3033。

[0070] 特征提取子单元3031，用于对所述检测结果进行物品视觉特征提取，以得到提取结果；特征检索子单元3032，用于根据所述提取结果进行特征检索，以得到检索结果；确定子单元3033，用于根据所述检索结果确定物品类型，以得到物品识别结果。

[0071] 在一实施例中,特征提取子单元3031,用于对所述检测结果采用基于MobileNetv2的识别网络进行物品特征提取,以得到提取结果。

[0072] 所述特征检索子单元3032,用于通过开源的FAISS框架建立特征检索引擎,并选择乘积量化优化索引,检索与所述提取结果相似的目标物品,以得到检索结果。

[0073] 在一实施例中,如图8所示,所述确定子单元3033包括筛选模块30331以及消除模块30332。

[0074] 筛选模块30331,用于筛选符合要求的检索结果,以得到筛选结果;消除模块30332,用于对所述筛选结果采用非极大值抑制算法消除边界框,以得到物品识别结果。

[0075] 需要说明的是,所属领域的技术人员可以清楚地了解到,上述基于物品识别的智能体交互装置300和各单元的具体实现过程,可以参考前述方法实施例中的相应描述,为了描述的方便和简洁,在此不再赘述。

[0076] 上述基于物品识别的智能体交互装置300可以实现为一种计算机程序的形式,该计算机程序可以在如图9所示的计算机设备上运行。

[0077] 请参阅图9,图9是本申请实施例提供的一种计算机设备的示意性框图。该计算机设备500可以是服务器,其中,服务器可以是独立的服务器,也可以是多个服务器组成的服务器集群。

[0078] 参阅图9,该计算机设备500包括通过系统总线501连接的处理器502、存储器和网络接口505,其中,存储器可以包括非易失性存储介质503和内存储器504。

[0079] 该非易失性存储介质503可存储操作系统5031和计算机程序5032。该计算机程序5032包括程序指令,该程序指令被执行时,可使得处理器502执行一种基于物品识别的智能体交互方法。

[0080] 该处理器502用于提供计算和控制能力,以支撑整个计算机设备500的运行。

[0081] 该内存储器504为非易失性存储介质503中的计算机程序5032的运行提供环境,该计算机程序5032被处理器502执行时,可使得处理器502执行一种基于物品识别的智能体交互方法。

[0082] 该网络接口505用于与其它设备进行网络通信。本领域技术人员可以理解,图9中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备500的限定,具体的计算机设备500可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0083] 其中,所述处理器502用于运行存储在存储器中的计算机程序5032,以实现如下步骤:

获取物品图像;对所述物品图像进行目标检测,以得到检测结果;对所述检测结果进行物品类型识别,以得到物品识别结果;根据所述物品识别结果注入智能体,以得到注入智能体的物品,使得用户与注入智能体的物品进行交互。

[0084] 在一实施例中,处理器502在实现所述对所述物品图像进行目标检测,以得到检测结果步骤时,具体实现如下步骤:

采用目标检测模型对所述物品图像进行目标检测,以得到检测结果。

[0085] 其中,所述目标检测模型是以YOLOV5模型为基础,通过模型剪枝、通道剪枝、减少Anchor的方式,结合非极大值抑制算法实现的。

[0086] 在一实施例中,处理器502在实现所述对所述检测结果进行物品类型识别,以得到物品识别结果步骤时,具体实现如下步骤:

对所述检测结果进行物品视觉特征提取,以得到提取结果;根据所述提取结果进行特征检索,以得到检索结果;根据所述检索结果确定物品类型,以得到物品识别结果。

[0087] 在一实施例中,处理器502在实现所述对所述检测结果进行物品视觉特征提取,以得到提取结果步骤时,具体实现如下步骤:

对所述检测结果采用基于MobileNetv2的识别网络进行物品特征提取,以得到提取结果。

[0088] 在一实施例中,处理器502在实现所述根据所述提取结果进行特征检索,以得到检索结果步骤时,具体实现如下步骤:

通过开源的FAISS框架建立特征检索引擎,并选择乘积量化优化索引,检索与所述提取结果相似的目标物品,以得到检索结果。

[0089] 在一实施例中,处理器502在实现所述根据所述检索结果确定物品类型,以得到物品识别结果步骤时,具体实现如下步骤:

筛选符合要求的检索结果,以得到筛选结果;对所述筛选结果采用非极大值抑制算法消除边界框,以得到物品识别结果。

[0090] 应当理解,在本申请实施例中,处理器502可以是中央处理单元(Central Processing Unit,CPU),该处理器502还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。其中,通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0091] 本领域普通技术人员可以理解的是实现上述实施例的方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成。该计算机程序包括程序指令,计算机程序可存储于一存储介质中,该存储介质为计算机可读存储介质。该程序指令被该计算机系统至少一个处理器执行,以实现上述方法的实施例的流程步骤。

[0092] 因此,本发明还提供一种存储介质。该存储介质可以为计算机可读存储介质。该存储介质存储有计算机程序,其中该计算机程序被处理器执行时使处理器执行如下步骤:

获取物品图像;对所述物品图像进行目标检测,以得到检测结果;对所述检测结果进行物品类型识别,以得到物品识别结果;根据所述物品识别结果注入智能体,以得到注入智能体的物品,使得用户与注入智能体的物品进行交互。

[0093] 在一实施例中,所述处理器在执行所述计算机程序而实现所述对所述物品图像进行目标检测,以得到检测结果步骤时,具体实现如下步骤:

采用目标检测模型对所述物品图像进行目标检测,以得到检测结果。

[0094] 其中,所述目标检测模型是以YOLOV5模型为基础,通过模型剪枝、通道剪枝、减少Anchor的方式,结合非极大值抑制算法实现的。

[0095] 在一实施例中,所述处理器在执行所述计算机程序而实现所述对所述检测结果进行物品类型识别,以得到物品识别结果步骤时,具体实现如下步骤:

对所述检测结果进行物品视觉特征提取,以得到提取结果;根据所述提取结果进

行特征检索,以得到检索结果;根据所述检索结果确定物品类型,以得到物品识别结果。

[0096] 在一实施例中,所述处理器在执行所述计算机程序而实现所述对所述检测结果进行物品视觉特征提取,以得到提取结果步骤时,具体实现如下步骤:

对所述检测结果采用基于MobileNetv2的识别网络进行物品特征提取,以得到提取结果。

[0097] 在一实施例中,所述处理器在执行所述计算机程序而实现所述根据所述提取结果进行特征检索,以得到检索结果步骤时,具体实现如下步骤:

通过开源的FAISS框架建立特征检索引擎,并选择乘积量化优化索引,检索与所述提取结果相似的目标物品,以得到检索结果。

[0098] 在一实施例中,所述处理器在执行所述计算机程序而实现所述根据所述检索结果确定物品类型,以得到物品识别结果步骤时,具体实现如下步骤:

筛选符合要求的检索结果,以得到筛选结果;对所述筛选结果采用非极大值抑制算法消除边界框,以得到物品识别结果。

[0099] 所述存储介质可以是U盘、移动硬盘、只读存储器(Read-Only Memory,ROM)、磁碟或者光盘等各种可以存储程序代码的计算机可读存储介质。

[0100] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0101] 在本发明所提供的几个实施例中,应该理解到,所揭露的装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的。例如,各个单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。

[0102] 本发明实施例方法中的步骤可以根据实际需要进行顺序调整、合并和删减。本发明实施例装置中的单元可以根据实际需要进行合并、划分和删减。另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以是两个或两个以上单元集成在一个单元中。

[0103] 该集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分,或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,终端,或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。

[0104] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到各种等效的修改或替换,这些修改或替换都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求要求的保护范围为准。

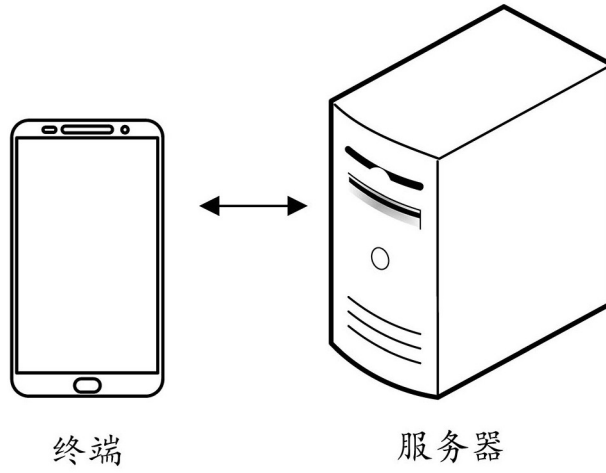


图1

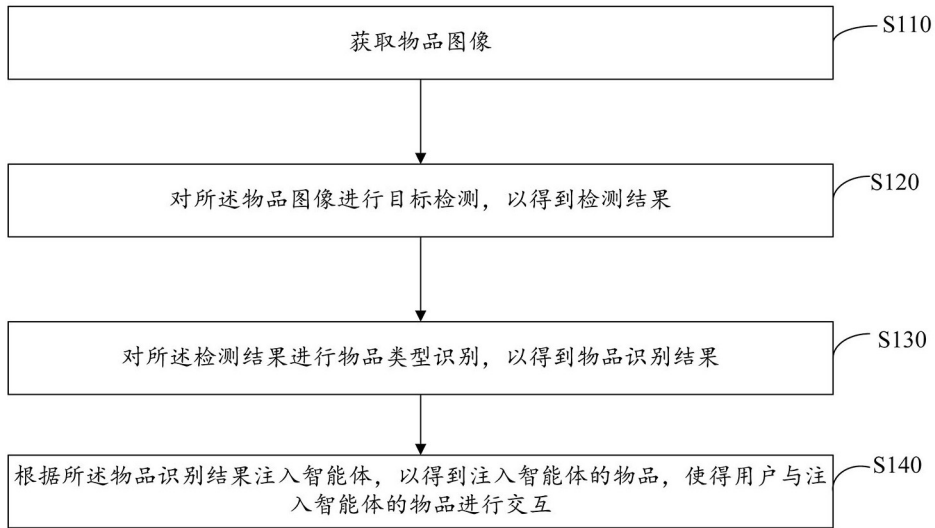


图2

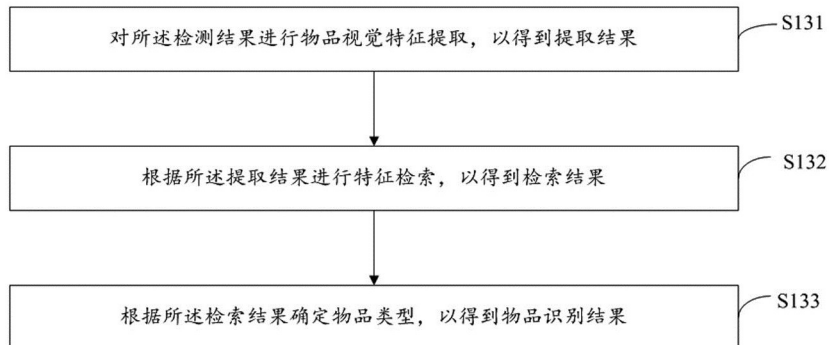


图3

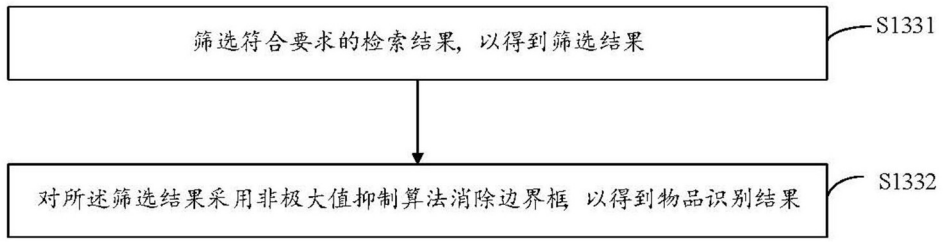


图4

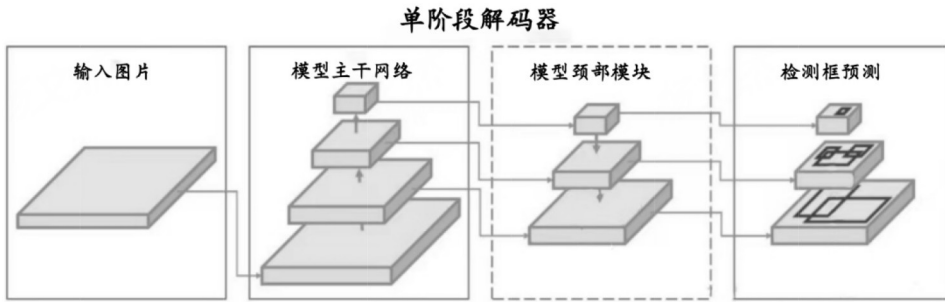


图5

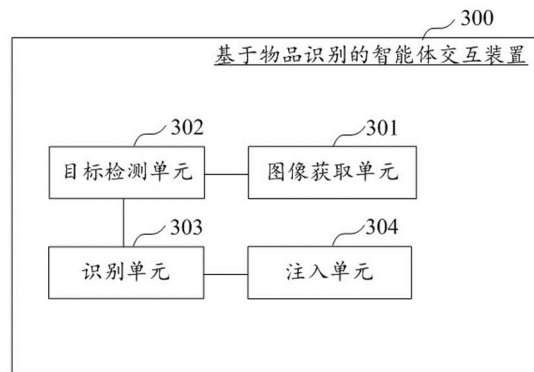


图6

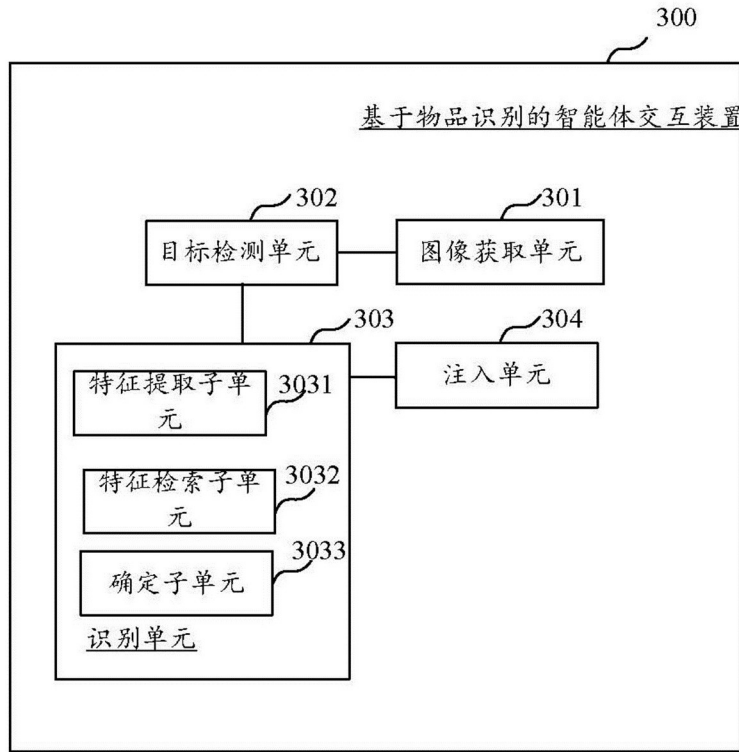


图7

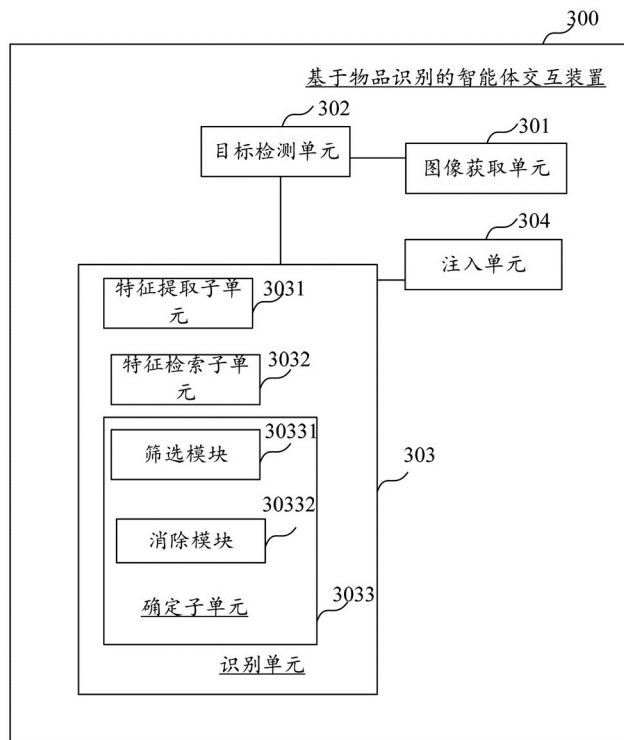


图8

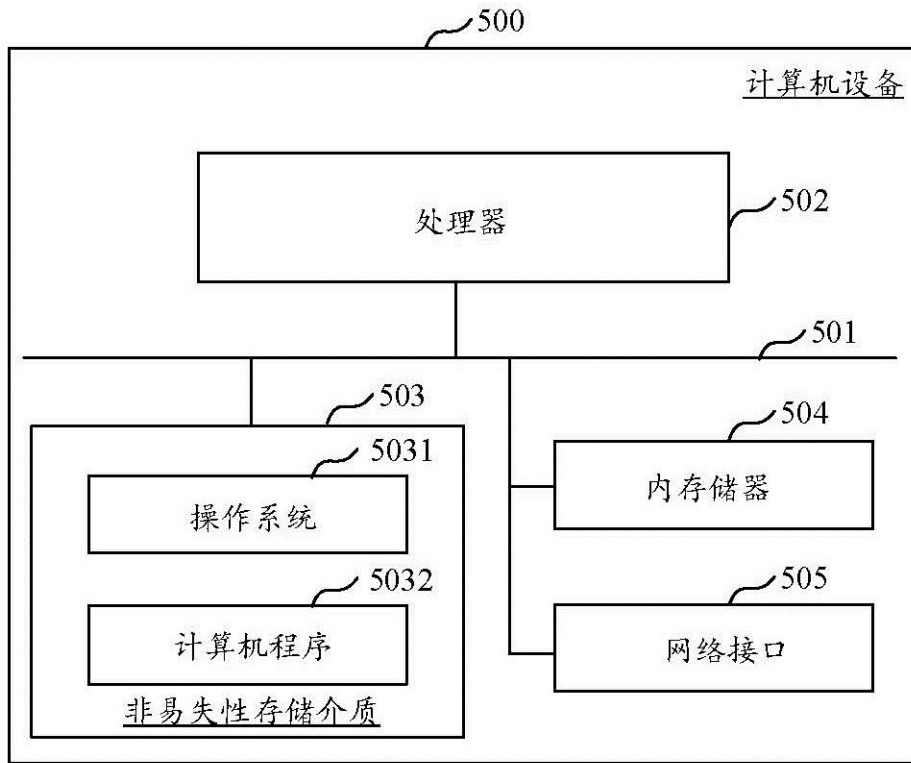


图9