



(12) 发明专利

(10) 授权公告号 CN 117216308 B

(45) 授权公告日 2024. 04. 26

(21) 申请号 202311481448.5

(22) 申请日 2023.11.09

(65) 同一申请的已公布的文献号
申请公布号 CN 117216308 A

(43) 申请公布日 2023.12.12

(73) 专利权人 天津华来科技股份有限公司
地址 300110 天津市南开区雅安道金平路
10号华侨创业大厦8楼(科技园)

(72) 发明人 郭阳 王静 罗瑞 冯啸 刘斌

(74) 专利代理机构 中科专利商标代理有限责任
公司 11021
专利代理师 李鹏宇

(51) Int. Cl.
G06F 16/58 (2019.01)

(56) 对比文件

- CN 112765387 A, 2021.05.07
- CN 114625897 A, 2022.06.14
- CN 116467607 A, 2023.07.21
- CN 112541091 A, 2021.03.23
- CN 113821669 A, 2021.12.21
- CN 105518678 A, 2016.04.20
- CN 103530649 A, 2014.01.22
- CN 104965921 A, 2015.10.07
- CN 116229095 A, 2023.06.06
- US 2019163699 A1, 2019.05.30

审查员 任树杰

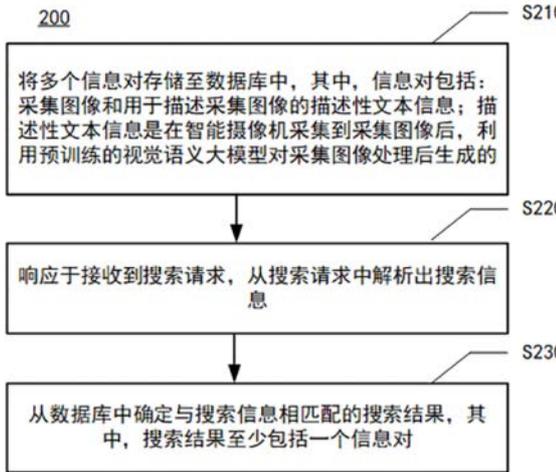
权利要求书3页 说明书13页 附图4页

(54) 发明名称

基于大模型的搜索方法、系统、设备及介质

(57) 摘要

本发明提供了基于大模型的搜索方法、系统、设备及介质,应用于人工智能以及智能摄像机技术领域。该方法包括:将多个信息对存储至数据库中,其中,信息对包括:采集图像和用于描述采集图像的描述性文本信息;描述性文本信息是在智能摄像机采集到采集图像后,利用预训练的视觉语义大模型对采集图像处理后生成的;响应于接收到搜索请求,从搜索请求中解析出搜索信息;从数据库中确定与搜索信息相匹配的搜索结果,其中,搜索结果至少包括一个信息对。本发明提供的基于大模型的搜索方法能够提高搜索效率,提升用户体验。



1. 一种基于大模型的搜索方法,其特征在于,所述方法包括:

将多个信息对存储至数据库中,其中,所述信息对包括:采集图像和用于描述所述采集图像的描述性文本信息;所述描述性文本信息是在智能摄像机采集到所述采集图像后,利用预训练的视觉语义大模型对所述采集图像处理后生成的,所述采集图像是所述智能摄像机通过实时监控场景,在确定检测到目标对象发生移动的情况下捕获当前场景下的图像后得到的,所述预训练的视觉语义大模型是预先通过对初始视觉语义大模型进行训练得到的,所述初始视觉语义大模型包括计算机视觉模型和自然语言处理模型,所述训练的方法包括:获取样本图像和用于描述所述样本图像的样本描述性文本信息;将所述样本图像输入所述计算机视觉模型,输出样本图像特征,其中,所述样本图像特征是基于样本对象的属性和不同样本对象之间的关系属性确定的;将所述样本图像特征输入所述自然语言处理模型,输出初始描述性文本信息;基于所述初始描述性文本信息和所述样本描述性文本信息,调整所述计算机视觉模型的参数和所述自然语言处理模型的参数,得到所述预训练的视觉语义大模型;

响应于接收到搜索请求,从所述搜索请求中解析出搜索信息,其中,用户在允许查看通过所述智能摄像机录制的画面以及需要查看特定画面的情况下,向搜索界面输入所述搜索信息,搜索引擎根据所述搜索信息发起所述搜索请求;

从所述数据库中确定与所述搜索信息相匹配的搜索结果,其中,所述搜索结果至少包括一个所述信息对;

其中,所述搜索信息包括搜索内容信息,所述从所述数据库中确定与所述搜索信息相匹配的搜索结果,包括:根据所述搜索内容信息与所述信息对中的所述采集图像和/或所述用于描述所述采集图像的描述性文本信息匹配,得到所述搜索结果;

其中,所述描述性文本信息满足预设字符数限制条件,所述搜索内容信息包括搜索文本信息;

所述方法还包括:

提取所述数据库中所有描述性文本信息的字符特征信息;

根据所述字符特征信息,生成字符数提示信息,以使用户根据所述字符数提示信息确定所述搜索文本信息的字符数;

所述描述性文本信息满足预设描述形式限制条件,

所述方法还包括:

提取所述数据库中所有描述性文本信息的描述特征信息;

根据所述描述特征信息,生成描述提示信息,以使用户根据所述描述提示信息确定所述搜索文本信息的描述形式。

2. 根据权利要求1所述的方法,其特征在于,所述搜索信息还包括搜索标识信息;

其中,所述从所述数据库中确定与所述搜索信息相匹配的搜索结果,还包括:

从所述多个信息对中确定与所述搜索标识信息相关联的多个关联信息对;

基于所述搜索内容信息和所述多个关联信息对,确定所述搜索结果。

3. 根据权利要求2所述的方法,其特征在于,所述关联信息对包括:关联采集图像和用于描述所述关联采集图像的关联描述性文本信息;

其中,所述基于所述搜索内容信息和所述多个关联信息对,确定所述搜索结果,包括:

将所述搜索内容信息、所述关联采集图像以及所述关联描述性文本信息,分别转化为搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息;

基于所述搜索内容的向量信息、所述关联采集图像的向量信息以及所述关联描述性文本的向量信息,确定所述搜索结果。

4. 根据权利要求3所述的方法,其特征在于,所述搜索结果包括目标采集图像和目标描述性文本信息;

其中,所述基于所述搜索内容的向量信息、所述关联采集图像的向量信息以及所述关联描述性文本的向量信息,确定所述搜索结果,包括:

在确定所述搜索内容信息为搜索文本信息的情况下,确定所述搜索文本信息对应的向量信息与所述关联描述性文本的向量信息之间的相似度值;

将所述相似度值超过阈值的关联描述性文本的向量信息对应的关联描述性文本信息,确定为所述目标描述性文本信息;

根据所述目标描述性文本信息,从所述多个关联信息对中确定所述目标采集图像;或者

在确定所述搜索内容信息为搜索图像信息的情况下,确定所述搜索图像信息对应的向量信息与所述关联采集图像的向量信息之间的相似度值;

将所述相似度值超过阈值的关联采集图像的向量信息对应的关联采集图像,确定为所述目标采集图像;

根据所述目标采集图像,从所述多个关联信息对中确定所述目标描述性文本信息。

5. 根据权利要求1所述的方法,其特征在于,搜索标识信息包括用户标识信息;

所述方法还包括:

从所述用户标识信息中提取用户属性特征;

根据所述用户属性特征,确定所述搜索结果的展示形式,其中,所述展示形式包括以下至少之一:图像形式、视频形式、语音播报形式。

6. 一种基于大模型的搜索系统,其特征在于,所述系统包括:

存储模块,用于利用与信息对相关智能摄像机的设备标识,对多个所述信息对打标后再存储至数据库中,其中,所述信息对包括:采集图像和用于描述所述采集图像的描述性文本信息;所述描述性文本信息是在所述智能摄像机采集到所述采集图像后,利用预训练的视觉语义大模型对所述采集图像处理后生成的,所述采集图像是所述智能摄像机通过实时监控场景,在确定检测到目标对象发生移动的情况下捕获当前场景下的图像后得到的,所述预训练的视觉语义大模型是预先通过对初始视觉语义大模型进行训练得到的,所述初始视觉语义大模型包括计算机视觉模型和自然语言处理模型,所述训练的方法包括:获取样本图像和用于描述所述样本图像的样本描述性文本信息;将所述样本图像输入所述计算机视觉模型,输出样本图像特征,其中,所述样本图像特征是基于样本对象的属性和不同样本对象之间的关系属性确定的;将所述样本图像特征输入所述自然语言处理模型,输出初始描述性文本信息;基于所述初始描述性文本信息和所述样本描述性文本信息,调整所述计算机视觉模型的参数和所述自然语言处理模型的参数,得到所述预训练的视觉语义大模型;

解析模块,用于响应于接收到搜索请求,从所述搜索请求中解析出搜索信息,其中,用

户在允许查看通过所述智能摄像机录制的画面以及需要查看特定画面的情况下,向搜索界面输入所述搜索信息,搜索引擎根据所述搜索信息发起所述搜索请求;

确定模块,用于从所述数据库中确定与所述搜索信息相匹配的搜索结果,其中,所述搜索结果至少包括一个所述信息对;

其中,所述搜索信息包括搜索内容信息,所述从所述数据库中确定与所述搜索信息相匹配的搜索结果,包括:根据所述搜索内容信息与所述数据库中存储的所述信息对中的所述采集图像和/或所述用于描述所述采集图像的描述性文本信息匹配,得到所述搜索结果;

其中,所述描述性文本信息满足预设字符数限制条件,所述搜索内容信息包括搜索文本信息;

所述系统还包括:

第一提取模块,用于提取所述数据库中所有描述性文本信息的字符特征信息;

第一生成模块,用于根据所述字符特征信息,生成字符数提示信息,以使用户根据所述字符数提示信息确定所述搜索文本信息的字符数;

所述描述性文本信息满足预设描述形式限制条件,

所述系统还包括:

第二提取模块,用于提取所述数据库中所有描述性文本信息的描述特征信息;

第二生成模块,用于根据所述描述特征信息,生成描述提示信息,以使用户根据所述描述提示信息确定所述搜索文本信息的描述形式。

7.一种电子设备,其特征在于,所述电子设备包括:

一个或多个处理器;

存储器,用于存储一个或多个程序,

其中,当所述一个或多个程序被所述一个或多个处理器执行时,使得所述一个或多个处理器实现权利要求1至5中任一项所述的方法。

8.一种计算机可读存储介质,其上存储有可执行指令,该指令被处理器执行时使处理器实现权利要求1至5中任一项所述的方法。

基于大模型的搜索方法、系统、设备及介质

技术领域

[0001] 本发明涉及人工智能以及智能摄像机技术领域,尤其涉及基于大模型的搜索方法、系统、设备及介质。

背景技术

[0002] 在相关技术中,在允许监控且符合相关规定的情况下,用户在允许查看通过摄像机录制的画面以及需要查看特定画面的情况下,通常需要直接手动的通过监控程序查看画面。在查看的过程中,仅能以单一的方式,从头到尾播放查看或者按时间查看,查询效率低,影响用户体验。此外,在现有搜索技术中,利用特定接口调用第三方提供者的数据,可能涉及数据隐私问题、数据依赖性问题以及投入成本等问题。

发明内容

[0003] 鉴于上述问题,本发明提供了一种基于大模型的搜索方法、系统、设备、介质和程序产品。

[0004] 根据本发明的第一个方面,提供了一种基于大模型的搜索方法,包括:

[0005] 将多个信息对,存储至数据库中,其中,信息对包括:采集图像和用于描述采集图像的描述性文本信息;描述性文本信息是在智能摄像机采集到采集图像后,利用预训练的视觉语义大模型对采集图像处理后生成的;

[0006] 响应于接收到搜索请求,从搜索请求中解析出搜索信息;

[0007] 从数据库中确定与搜索信息相匹配的搜索结果,其中,搜索结果至少包括一个信息对。

[0008] 根据本发明的实施例,搜索信息包括搜索内容信息和搜索标识信息;

[0009] 其中,从数据库中确定与搜索信息相匹配的搜索结果,包括:

[0010] 从多个信息对中确定与搜索标识信息相关联的多个关联信息对;

[0011] 基于搜索内容信息和多个关联信息对,确定搜索结果。

[0012] 根据本发明的实施例,关联信息对包括:关联采集图像和用于描述关联采集图像的关联描述性文本信息;

[0013] 其中,基于搜索内容信息和多个关联信息对,确定搜索结果,包括:

[0014] 将搜索内容信息、关联采集图像以及关联描述性文本信息,分别转化为搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息;

[0015] 基于搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息,确定搜索结果。

[0016] 根据本发明的实施例,搜索结果包括目标采集图像和目标描述性文本信息;

[0017] 其中,基于搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息,确定搜索结果,包括:

[0018] 在确定搜索内容信息为搜索文本信息的情况下,确定搜索文本信息对应的向量信

息与关联描述性文本的向量信息之间的相似度值；

[0019] 将相似度值超过阈值的关联描述性文本的向量信息对应的关联描述性文本信息，确定为目标描述性文本信息；

[0020] 根据目标描述性文本信息，从多个关联信息对中确定目标采集图像；或者

[0021] 在确定搜索内容信息为搜索图像信息的情况下，确定搜索图像信息对应的向量信息与关联采集图像的向量信息之间的相似度值；

[0022] 将相似度值超过阈值的关联采集图像的向量信息对应的关联采集图像，确定为目标采集图像；

[0023] 根据目标采集图像，从多个关联信息对中确定目标描述性文本信息。

[0024] 根据本发明的实施例，描述性文本信息满足预设字符数限制条件，搜索信息包括搜索文本信息；

[0025] 基于大模型的搜索方法还包括：

[0026] 提取数据库中所有描述性文本信息的字符特征信息；

[0027] 根据字符特征信息，生成字符数提示信息，以使用户根据字符数提示信息确定搜索文本信息的字符数。

[0028] 根据本发明的实施例，描述性文本信息满足预设描述形式限制条件，搜索信息包括搜索文本信息；

[0029] 基于大模型的搜索方法还包括：

[0030] 提取数据库中所有描述性文本信息的描述特征信息；

[0031] 根据描述特征信息，生成描述提示信息，以使用户根据描述提示信息确定搜索文本信息。

[0032] 根据本发明的实施例，搜索标识信息包括用户标识信息；

[0033] 基于大模型的搜索方法还包括：

[0034] 从用户标识信息中提取用户属性特征；

[0035] 根据用户属性特征，确定搜索结果的展示形式，其中，展示形式包括以下至少之一：图像形式、视频形式、语音播报形式。

[0036] 本发明的第二方面提供了一种基于大模型的搜索系统，包括：

[0037] 存储模块，用于将多个信息对，存储至数据库中，其中，信息对包括：采集图像和用于描述采集图像的描述性文本信息；描述性文本信息是在智能摄像机采集到采集图像后，利用预训练的视觉语义大模型对采集图像处理后生成的；

[0038] 解析模块，用于响应于接收到搜索请求，从搜索请求中解析出搜索信息；

[0039] 确定模块，用于从数据库中确定与搜索信息相匹配的搜索结果，其中，搜索结果至少包括一个信息对。

[0040] 本发明的第三方面提供了一种电子设备，包括：一个或多个处理器；存储器，用于存储一个或多个程序，其中，当一个或多个程序被一个或多个处理器执行时，使得一个或多个处理器执行上述基于大模型的搜索方法。

[0041] 本发明的第四方面还提供了一种计算机可读存储介质，其上存储有可执行指令，该指令被处理器执行时使处理器执行上述基于大模型的搜索方法。

[0042] 本发明的第五方面还提供了一种计算机程序产品，包括计算机程序，该计算机程

序被处理器执行时实现上述基于大模型的搜索方法。

[0043] 根据本发明的实施例,智能摄像机采集图像,并利用预训练的视觉语义大模型对采集图像处理后生成用于描述采集图像的描述性文本信息,将这些数据预先存储于数据库,在用户需要搜索时,从数据库中匹配搜索结果,能够提高搜索效率;由于视觉语义大模型借助大模型较强的捕捉特征性能、泛化能力、图像细节处理能力以及对输入噪声或变化更加鲁棒的特点,能够准确得到用于描述采集图像的描述性文本信息,有利于用户根据描述性文本信息快速知晓特定画面发生的事件以及查看特定画面,提升了用户体验。此外,本发明提供的基于大模型的搜索方法,在搜索的过程中,搜索数据是智能摄像机在检测的过程中采集到的,无需借助第三方提供者的数据,至少部分解决现有搜索技术中,利用特定接口调用第三方提供者的数据,涉及数据依赖性问题以及投入成本等问题。

附图说明

[0044] 通过以下参照附图对本发明实施例的描述,本发明的上述内容以及其他目的、特征和优点将更为清楚,在附图中:

[0045] 图1示出了根据本发明实施例的基于大模型的搜索方法、系统、设备、介质和程序产品的应用场景图;

[0046] 图2示出了根据本发明实施例的基于大模型的搜索方法的流程图;

[0047] 图3示出了根据本发明另一实施例的基于大模型的搜索方法的流程图;

[0048] 图4示出了根据本发明实施例的基于大模型的搜索系统的结构框图;以及

[0049] 图5示出了根据本发明实施例的适于实现基于大模型的搜索方法的电子设备的方框图。

具体实施方式

[0050] 以下,将参照附图来描述本发明的实施例。但是应该理解,这些描述只是示例性的,而并非要限制本发明的范围。在下面的详细描述中,为便于解释,阐述了许多具体的细节以提供对本发明实施例的全面理解。然而,明显地,一个或多个实施例在没有这些具体细节的情况下也可以被实施。此外,在以下说明中,省略了对公知结构和技术的描述,以避免不必要地混淆本发明的概念。

[0051] 在此使用的术语仅仅是为了描述具体实施例,而并非意在限制本发明。在此使用的术语“包括”、“包含”等表明了特征、步骤、操作和/或部件的存在,但是并不排除存在或添加一个或多个其他特征、步骤、操作或部件。

[0052] 在此使用的所有术语(包括技术和科学术语)具有本领域技术人员通常所理解的含义,除非另外定义。应注意,这里使用的术语应解释为具有与本说明书的上下文相一致的含义,而不应以理想化或过于刻板的方式来解释。

[0053] 在使用类似于“A、B和C等中至少一个”这样的表述的情况下,一般来说应该按照本领域技术人员通常理解该表述的含义来予以解释(例如,“具有A、B和C中至少一个的系统”应包括但不限于单独具有A、单独具有B、单独具有C、具有A和B、具有A和C、具有B和C、和/或具有A、B、C的系统等)。

[0054] 在本发明的技术方案中,所涉及的用户信息(包括但不限于用户个人信息、用户图

像信息、用户设备信息,例如位置信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等),均为经用户授权或者经过各方充分授权的信息和数据,并且相关数据的收集、存储、使用、加工、传输、提供、公开和应用等处理,均遵守相关国家和地区的相关法律法规和标准,采取了必要保密措施,不违背公序良俗,并提供有相应的操作入口,供用户选择授权或者拒绝。

[0055] 在本发明实施例的技术方案中,在获取或采集用户个人信息之前,均获取了用户的授权或同意。

[0056] 在相关技术中,在允许监控且符合相关规定的情况下,用户在允许查看通过摄像机录制的画面,且需要查看特定画面的情况下,通常需要直接手动的通过监控程序查看画面。在查看的过程中,仅能以单一的方式,从头到尾播放查看或者按时间查看,查询效率低,影响用户体验。此外,在现有搜索技术中,利用特定接口调用第三方提供者的数据,可能涉及数据隐私问题、数据依赖性问题以及投入成本等问题。例如,在完全依赖第三方提供者的数据的情况下,如果更改接口或者第三方提供者停止服务,可能均会影响用户体验。如果使用开源工具和库,自建搜索系统,例如搜索服务器Elasticsearch,虽然可以存储图像的元数据和描述,然后基于这些数据进行搜索,但是,需要大量的时间、资源和技术投资,甚至需要不断地更新和维护系统,仍存在用户搜索效率低,准确度低的问题。

[0057] 本发明的实施例提供了基于大模型的搜索方法、系统、设备及介质。该方法包括:将多个信息对,存储至数据库中,其中,信息对包括:采集图像和用于描述采集图像的描述性文本信息;描述性文本信息是在智能摄像机采集到采集图像后,利用预训练的视觉语义大模型对采集图像处理后生成的;响应于接收到搜索请求,从搜索请求中解析出搜索信息;从数据库中确定与搜索信息相匹配的搜索结果,其中,搜索结果至少包括一个信息对。

[0058] 图1示出了根据本发明实施例的基于大模型的搜索方法、系统、设备、介质和程序产品的应用场景图。

[0059] 如图1所示,根据该实施例的应用场景100可以包括第一终端设备101、第二终端设备102、第三终端设备103、网络104以及服务器105。网络104用以在第一终端设备101、第二终端设备102、第三终端设备103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0060] 用户可以使用第一终端设备101、第二终端设备102、第三终端设备103中的至少一个通过网络104与云端服务器105交互,以接收或发送消息等。第一终端设备101、第二终端设备102、第三终端设备103上可以安装有智能摄像机以及与智能摄像机相关联的客户端应用。

[0061] 第一终端设备101、第二终端设备102、第三终端设备103可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。

[0062] 服务器105可以是提供各种服务的服务器,例如对用户利用第一终端设备101、第二终端设备102、第三终端设备103上智能摄像机所检测的目标对象发生移动的情况提供支持的后台管理服务器(仅为示例)。后台管理服务器可以对接收到的采集图像等进行分析等处理,并将处理结果(例如根据采集图像生成用于描述场景图像的描述性文本信息等)反馈给终端设备。

[0063] 应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器的。

[0064] 以下将基于图1描述的场景,通过图2~图3对公开实施例的基于大模型的搜索方法进行详细描述。

[0065] 图2示出了根据本发明实施例的基于大模型的搜索方法的流程图。

[0066] 如图2所示,该实施例的基于大模型的搜索方法200包括操作S210~操作S230。

[0067] 在操作S210,将多个信息对存储至数据库中,其中,信息对包括:采集图像和用于描述采集图像的描述性文本信息;描述性文本信息是在智能摄像机采集到采集图像后,利用预训练的视觉语义大模型对采集图像处理后生成的。

[0068] 需要说明的是,智能摄像机的检测,以及采集图像时,所涉及的信息(包括但不限于用户个人信息、用户图像信息、用户设备信息,例如位置信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等),均为经用户授权或者经过各方充分授权的信息和数据,并且相关数据的收集、存储、使用、加工、传输、提供、公开和应用等处理,均遵守相关国家和地区的相关法律法规和标准,采取了必要保密措施,不违背公序良俗,并提供有相应的操作入口,供用户选择授权或者拒绝。

[0069] 根据本发明的实施例,智能摄像机可以通过图像采集功能采集图像,将其上传至预先部署了预训练的视觉语义大模型的云端服务器,以使得预训练的视觉语义大模型处理采集图像,并输出用于描述采集图像的描述性文本信息。将采集图像和描述性文本信息作为信息对存储于数据库,还可以利用与信息对相关联的智能摄像机的设备标识,对信息对打标后再存储于数据库。

[0070] 其中,视觉语义大模型是基于样本对象的属性和不同样本对象之间的关系属性,对初始视觉语义模型进行预先训练得到的,初始视觉语义模型通过计算机视觉模型和自然语言处理模型共同构建得到。其中,样本对象可以是具有移动能力的对象,例如,可以是车辆、行人、动物等。

[0071] 在操作S220,响应于接收到搜索请求,从搜索请求中解析出搜索信息。

[0072] 根据本发明的实施例,用户可以将搜索信息输入搜索界面,搜索引擎根据搜索信息发起搜索请求,搜索系统在接收到搜索请求后解析出搜索信息。

[0073] 在操作S230,从数据库中确定与搜索信息相匹配的搜索结果,其中,搜索结果至少包括一个信息对。

[0074] 根据本发明的实施例,搜索信息可以包括搜索内容信息。可以根据搜索内容信息与数据库中存储的信息对中的采集图像和/或用于描述采集图像的描述性文本信息匹配,得到搜索结果。搜索信息也可以包括搜索内容信息和搜索标识信息,可以根据搜索标识信息查找与其相关联的信息对,然后再从该信息对中,根据搜索内容信息匹配搜索结果。

[0075] 根据本发明实施例,智能摄像机采集图像,并利用预训练的视觉语义大模型对采集图像处理后生成用于描述采集图像的描述性文本信息,将这些数据预先存储于数据库,在用户需要搜索时,从数据库中匹配搜索结果,能够提高搜索效率;由于视觉语义大模型借助大模型较强的捕捉特征性能、泛化能力、图像细节处理能力以及对输入噪声或变化更加鲁棒的特点,能够准确得到用于描述采集图像的描述性文本信息,有利于用户根据描述性文本信息快速知晓特定画面发生的事件以及查看特定画面,提升了用户体验。此外,本发明

提供的基于大模型的搜索方法,在搜索的过程中,搜索数据是智能摄像机在检测的过程中采集到的,无需借助第三方提供者的数据,至少部分解决现有搜索技术中,利用特定接口调用第三方提供者的数据,涉及数据依赖性问题以及投入成本等问题。

[0076] 根据本发明的实施例,搜索信息可以包括搜索内容信息和搜索标识信息。

[0077] 其中,从数据库中确定与搜索信息相匹配的搜索结果,可以包括:

[0078] 从多个信息对中确定与搜索标识信息相关联的多个关联信息对;基于搜索内容信息和多个关联信息对,确定搜索结果。

[0079] 根据本发明的实施例,搜索标识信息可以是用户的标识信息。可以根据生成信息对的设备标识信息,确定与用户的标识信息相关的关联设备标识信息。根据关联设备标识信息,从多个信息对中确定多个关联信息对。可以将搜索内容信息与多个关联信息对中的关联采集图像和/或用于描述关联采集图像的关联描述性文本信息进行匹配,将相匹配的关联采集图像和/或用于描述关联采集图像的关联描述性文本信息确定为搜索结果。

[0080] 根据本发明实施例,通过搜索标识信息先确定关联信息对,再从关联信息对中确定搜索结果,能够降低搜索计算量,提高搜索效率。此外,在与搜索标识相关联的关联信息对中搜索,避免涉及其他不相关信息对的隐私安全性问题。

[0081] 根据本发明的实施例,关联信息对可以包括:关联采集图像和用于描述关联采集图像的关联描述性文本信息。

[0082] 其中,基于搜索内容信息和多个关联信息对,确定搜索结果,可以包括:将搜索内容信息、关联采集图像以及关联描述性文本信息,分别转化为搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息;基于搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息,确定搜索结果。

[0083] 根据本发明的实施例,搜索内容信息可以是搜索文本信息或者搜索图像信息。可以对搜索内容信息、关联采集图像以及关联描述性文本信息分别进行向量化处理。

[0084] 根据本发明实施例,通过向量处理后,便于利用向量间的相似性计算,从而准确确定搜索结果。

[0085] 根据本发明的实施例,搜索结果可以包括目标采集图像和目标描述性文本信息。

[0086] 其中,基于搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息,确定搜索结果,可以包括:在确定搜索内容信息为搜索文本信息的情况下,确定搜索文本信息对应的向量信息与关联描述性文本的向量信息之间的相似度值;将相似度值超过阈值的关联描述性文本的向量信息对应的关联描述性文本信息,确定为目标描述性文本信息;根据目标描述性文本信息,从多个关联信息对中确定目标采集图像。

[0087] 其中,基于搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息,确定搜索结果,可以包括:在确定搜索内容信息为搜索图像信息的情况下,确定搜索图像信息对应的向量信息与关联采集图像的向量信息之间的相似度值;将相似度值超过阈值的关联采集图像的向量信息对应的关联采集图像,确定为目标采集图像;根据目标采集图像,从多个关联信息对中确定目标描述性文本信息。

[0088] 根据本发明的实施例,阈值可以是根据实际所需搜索精确度确定。相似度值可以根据相似度计算函数确定。其中,相似度计算函数本发明不作具体限定。

[0089] 根据本发明实施例,通过计算向量信息之间的相似度值,能够快速确定搜索结果,

提高搜索效率,提升用户体验。

[0090] 根据本发明的实施例,描述性文本信息满足预设字符数限制条件,搜索信息包括搜索文本信息。

[0091] 基于大模型的搜索方法还可以包括:

[0092] 提取数据库中所有描述性文本信息的字符特征信息;根据字符特征信息,生成字符数提示信息,以使用户根据字符数提示信息确定搜索文本信息的字符数。

[0093] 根据本发明的实施例,预设字符数限制条件可以根据实际对图像进行描述的情况以及用户阅读文本信息的情况等确定。例如,预设字符数限制条件可以是用于描述图像的文本信息中的描述长度的字符数不超过预设数量的字符数等。字符特征信息用于表征字符数的数量特征。可以统计数据库中所有描述性文本信息的字符特征信息,然后将所有字符特征信息均值化后,得到目标字符特征信息。根据目标字符特征信息,生成字符数提示信息。在用户进行搜索时,可以根据字符数提示信息,输入搜索文本信息。

[0094] 在实现本发明实施例的过程中发现,在进行搜索时,输入过少字符数的搜索文本信息,会导致搜索的准确度降低;输入过多字符数的搜索文本信息,会增加用户搜索的繁琐,用户搜索体验感差,甚至搜索耗时较长,降低搜索效率。本发明通过描述性文本信息的字符特征信息生成字符数提示信息,便于用户快速输入适当字符数的搜索文本信息,提高了搜索准确度,增强了用户体验,提升了搜索效率。

[0095] 根据本发明的实施例,描述性文本信息满足预设描述形式限制条件,搜索信息包括搜索文本信息。

[0096] 基于大模型的搜索方法还可以包括:

[0097] 提取数据库中所有描述性文本信息的描述特征信息;根据描述特征信息,生成描述提示信息,以使用户根据描述提示信息确定搜索文本信息的描述形式。

[0098] 根据本发明的实施例,预设描述形式限制条件可以是根据用户的阅读习惯、语言描述特点等确定的。描述特征信息可以用于表征文本描述形式的特征。可以统计数据库中所有描述性文本信息的描述特征信息,然后筛选出现次数最多的描述特征信息作为目标描述特征信息。根据描述特征信息,生成描述特征信息。在用户进行搜索时,可以根据描述提示信息,输入搜索文本信息。

[0099] 根据本发明实施例,通过满足预设描述形式限制条件的描述性文本信息的描述特征信息生成描述提示信息,有助于用户输入的搜索文本信息的语言表达清晰且有条理,并且与数据库中的描述性文本信息形式相似,能够提高搜索的准确度的同时,还提升搜索效率。

[0100] 根据本发明的实施例,搜索标识信息可以包括用户标识信息。

[0101] 基于大模型的搜索方法还可以包括:

[0102] 从用户标识信息中提取用户属性特征;根据用户属性特征,确定搜索结果的展示形式,其中,展示形式包括以下至少之一:图像形式、视频形式、语音播报形式。

[0103] 根据本发明的实施例,用户属性特征可以包括用户的年龄特征。可以根据不同年龄特征,设定不同的搜索结果展示形式。

[0104] 例如,用户的年龄特征可以包括青少年、中年、老年等。针对老年,可以以语音播报形式展示。

[0105] 根据本发明实施例,根据用户属性特征,确定搜索结果的展示形式,增加了搜索趣味性,提高了用户体验感,尤其对于老年人而言,通过语音播报形式展示搜索结果,有助于快速得知搜索结果。

[0106] 根据本发明实施例,在搜索结果包括多个信息时,可以通过相关性算法对结果进行排序,确定展示搜索结果的顺序。

[0107] 根据本发明实施例,在从数据库中确定与搜索信息相匹配的搜索结果后,可以将搜索结果功能化,使其在返回用户时,可以具有交互功能,如缩放、预览、分享或下载选项,使用户可以更方便地处理搜索结果。

[0108] 根据本发明实施例,可以利用预处理算法,提取数据库中所有描述性文本信息的关键词,作为搜索关键词,根据搜索关键词生成提示关键词,以使用户根据提示关键词确定搜索文本信息,避免用户输入的错误,方便用户使用搜索功能,确保搜索的质量。

[0109] 图3示出了根据本发明另一实施例的基于大模型的搜索方法的流程图。

[0110] 如图3所示,该实施例的基于大模型的搜索方法300包括操作S310~操作S370。

[0111] 在操作S310,获取智能摄像机检测到目标对象发生移动时采集的采集图像。

[0112] 在操作S320,利用预训练的视觉语义大模型对采集图像处理后生成描述性文本信息。

[0113] 在操作S330,将采集图像和描述性文本信息向量化后存储至数据库中。

[0114] 在操作S340,响应于接收到搜索请求,从搜索请求中解析出搜索信息。

[0115] 在操作S350,将搜索信息向量化。

[0116] 在操作S360,将向量化后的搜索信息与存储于数据库中的向量化后的采集图像和描述性文本信息进行相似度计算,得到相似度值。

[0117] 在操作S370,根据相似度值,从数据库中确定与搜索信息相匹配的搜索结果。

[0118] 根据本发明的实施例,智能摄像机可以通过内置的移动检测算法实时监控场景,一旦检测到目标对象发生移动,会捕获当前场景下的图像,及时将采集图像接入云端服务器部署的视觉语义大模型,得到用于描述采集图像的描述性文本信息,并存储至数据库中。

[0119] 根据本发明的实施例,视觉语义大模型是预先通过对初始视觉语义模型进行训练得到的。训练方法可以包括:获取样本图像和用于描述样本图像的样本描述性文本信息;将样本图像输入计算机视觉模型,输出样本图像特征,其中,样本图像特征是基于样本对象的属性和不同样本对象之间的关系属性确定的;将样本图像特征输入自然语言处理模型,输出初始描述性文本信息;基于初始描述性文本信息和样本描述性文本信息,调整计算机视觉模型的参数和自然语言处理模型的参数,得到视觉语义大模型。其中,计算机视觉模型可以由多层卷积神经网络构成,可以利用多层卷积层和池化层识别样本图像中的样本图像特征。自然语言处理模型可以是一个循环神经网络、长短时记忆网络或Transformer网络。其中,样本图像特征可以包括:样本图像中样本对象的属性特征和不同样本对象之间的关系属性特征。其中,样本对象可以是样本图像中所有物,例如样本图像可以是某一时刻针对客厅的拍摄图像,则样本对象可以是墙、人物、动物、灯、沙发、茶几、电视等。样本对象的属性可以是样本对象的基本属性和样本对象在样本图像中的位置属性等。不同样本对象之间的关系属性可以是不同样本对象之间的位置关系属性和不同样本对象之间存在关联的关联属性等,例如,人坐在沙发上,人与沙发存在关联等。

[0120] 根据本发明实施例,计算机视觉模型基于样本对象的属性和不同样本对象之间的关系属性确定的样本图像特征,能够更加准确地提取到用于描述样本图像的特征,有利于对样本对象的识别,以及有利于自然语言处理模型根据输入的样本图像特征,准确生成可以描述样本图像的文本信息,基于此训练得到的视觉语义大模型能够对输入图像进行精准处理后,能够输出针对图像的重要描述性文本。

[0121] 根据本发明的实施例,在相似度值超过阈值的情况下,确定采集图像和描述性文本信息为搜索结果。

[0122] 根据本发明实施例,视觉语义大模型能够捕捉文本描述和图像之间的深层次关系,不仅仅是基于关键词或明显的标签。例如,用户搜索“一个悠闲的下午时光”,视觉语义大模型可能会返回一个人在吊床上休息或在户外读书的图像,即使这些图像没有明确的“悠闲”或“下午”标签。

[0123] 根据本发明实施例,传统的图像搜索系统通常依赖于大量的标注数据。而视觉语义大模型可以在少量标签数据或甚至无标签数据上表现得很好。由于模型的预训练过程通常涉及大量的数据,它们在处理不同类型、风格和来源的图像时可能更加鲁棒。对于噪声、低分辨率或不常见的图像内容,视觉语义大模型可能仍然能够提供有意义的搜索结果。视觉语义大模型还可以通过在线学习或迁移学习继续微调模型,使其适应特定的数据集或应用场景,从而持续提高搜索的准确性和相关性。

[0124] 基于上述基于大模型的搜索方法,本发明还提供了一种基于大模型的搜索系统。以下将结合图4对该系统进行详细描述。

[0125] 图4示出了根据本发明实施例的基于大模型的搜索系统的结构框图。

[0126] 如图4所示,该实施例的基于大模型的搜索系统400包括存储模块410、解析模块420和确定模块430。

[0127] 存储模块410用于将多个信息对存储至数据库中,其中,信息对包括:采集图像和用于描述采集图像的描述性文本信息;描述性文本信息是在智能摄像机采集到采集图像后,利用预训练的视觉语义大模型对采集图像处理后生成的。在一实施例中,存储模块410可以用于执行前文描述的操作S210,在此不再赘述。

[0128] 解析模块420用于响应于接收到搜索请求,从搜索请求中解析出搜索信息。在一实施例中,解析模块420可以用于执行前文描述的操作S220,在此不再赘述。

[0129] 确定模块430用于从数据库中确定与搜索信息相匹配的搜索结果,其中,搜索结果至少包括一个信息对。在一实施例中,确定模块430可以用于执行前文描述的操作S230,在此不再赘述。

[0130] 根据本发明的实施例,搜索信息包括搜索内容信息和搜索标识信息。其中,确定模块430包括:第一子确定单元和第二子确定单元。

[0131] 第一子确定单元用于从多个信息对中确定与搜索标识信息相关联的多个关联信息对。

[0132] 第二子确定单元用于基于搜索内容信息和多个关联信息对,确定搜索结果。

[0133] 根据本发明的实施例,关联信息对包括:关联采集图像和用于描述关联采集图像的关联描述性文本信息。其中,基于搜索内容信息和多个关联信息对,确定搜索结果,包括:将搜索内容信息、关联采集图像以及关联描述性文本信息,分别转化为搜索内容的向量信

息、关联采集图像的向量信息以及关联描述性文本的向量信息；基于搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息，确定搜索结果。

[0134] 根据本发明的实施例，搜索结果包括目标采集图像和目标描述性文本信息。其中，基于搜索内容的向量信息、关联采集图像的向量信息以及关联描述性文本的向量信息，确定搜索结果，包括：在确定搜索内容信息为搜索文本信息的情况下，确定搜索文本信息对应的向量信息与关联描述性文本的向量信息之间的相似度值；将相似度值超过阈值的关联描述性文本的向量信息对应的关联描述性文本信息，确定为目标描述性文本信息；根据目标描述性文本信息，从多个关联信息对中确定目标采集图像；或者

[0135] 在确定搜索内容信息为搜索图像信息的情况下，确定搜索图像信息对应的向量信息与关联采集图像的向量信息之间的相似度值；将相似度值超过阈值的关联采集图像的向量信息对应的关联采集图像，确定为目标采集图像；根据目标采集图像，从多个关联信息对中确定目标描述性文本信息。

[0136] 根据本发明的实施例，描述性文本信息满足预设字符数限制条件，搜索信息包括搜索文本信息。

[0137] 基于大模型的搜索系统400还可以包括：第一提取模块和第一生成模块。

[0138] 第一提取模块用于提取数据库中所有描述性文本信息的字符特征信息。

[0139] 生成模块用于根据字符特征信息，生成字符数提示信息，以使用户根据字符数提示信息确定搜索文本信息的字符数。

[0140] 根据本发明的实施例，描述性文本信息满足预设描述形式限制条件，搜索信息包括搜索文本信息。

[0141] 基于大模型的搜索系统400还可以包括：第二提取模块和第二生成模块。

[0142] 第二提取模块用于提取数据库中所有描述性文本信息的描述特征信息。

[0143] 第二生成模块用于根据描述特征信息，生成描述提示信息，以使用户根据描述提示信息确定搜索文本信息的描述形式。

[0144] 根据本发明的实施例，搜索标识信息包括用户标识信息。

[0145] 基于大模型的搜索系统400还可以包括：第三提取模块和展示形式确定模块。

[0146] 第三提取模块用于从用户标识信息中提取用户属性特征。

[0147] 展示形式确定模块用于根据用户属性特征，确定搜索结果的展示形式，其中，展示形式包括以下至少之一：图像形式、视频形式、语音播报形式。

[0148] 根据本发明的实施例，存储模块410、解析模块420和确定模块430中的任意多个模块可以合并在一个模块中实现，或者其中的任意一个模块可以被拆分成多个模块。或者，这些模块中的一个或多个模块的至少部分功能可以与其他模块的至少部分功能相结合，并在一个模块中实现。根据本发明的实施例，存储模块410、解析模块420和确定模块430中的至少一个可以至少被部分地实现为硬件电路，例如现场可编程门阵列(FPGA)、可编程逻辑阵列(PLA)、片上系统、基板上的系统、封装上的系统、专用集成电路(ASIC)，或可以通过对电路进行集成或封装的任何其他的合理方式等硬件或固件来实现，或以软件、硬件以及固件三种实现方式中任意一种或以其中任意几种的适当组合来实现。或者，存储模块410、解析模块420和确定模块430中的至少一个可以至少被部分地实现为计算机程序模块，当该计算机程序模块被运行时，可以执行相应的功能。

[0149] 根据本发明的实施例,本发明的基于大模型的搜索系统,可以根据用户的反馈进行在线学习。例如,如果用户标记某个搜索结果为不相关或低质量,系统可以利用这些信息来微调模型,从而在未来提供更准确的结果。除了基本的文搜图或视频功能外,还支持其他的搜索模式。例如,用户可以上传一个图像,然后搜索与之相关的文本描述或其他图像。这增加了系统的灵活性,使其可以满足各种不同的搜索需求。

[0150] 根据本发明的实施例,本发明的基于大模型的搜索系统的可以使用户交互的前端部分的功能可以预先配置为不仅仅是接收用户的文本查询,还包括一系列的界面设计和算法,使用户可以更容易、直观地与系统交互。在界面设计中采用简洁、直观的界面设计。为了简化用户的搜索体验,可以是除了传统的搜索框,还提供了一些高级搜索选项,如使用自然语言描述、语音输入或基于情境的建议查询等。

[0151] 图5示出了根据本发明实施例的适于实现基于大模型的搜索方法的电子设备的方框图。

[0152] 如图5所示,根据本发明实施例的电子设备500包括处理器501,其可以根据存储在只读存储器(ROM) 502中的程序或者从存储部分508加载到随机访问存储器(RAM) 503中的程序而执行各种适当的动作和处理。处理器501例如可以包括通用微处理器(例如CPU)、指令集处理器和/或相关芯片组和/或专用微处理器(例如,专用集成电路(ASIC))等等。处理器501还可以包括用于缓存用途的板载存储器。处理器501可以包括用于执行根据本发明实施例的方法流程的不同动作的单一处理单元或者是多个处理单元。

[0153] 在RAM 503中,存储有电子设备500操作所需的各种程序和数据。处理器 501、ROM 502以及RAM 503通过总线504彼此相连。处理器501通过执行ROM 502和/或RAM 503中的程序来执行根据本发明实施例的方法流程的各种操作。需要注意,程序也可以存储在除ROM 502和RAM 503以外的一个或多个存储器中。处理器501也可以通过执行存储在一个或多个存储器中的程序来执行根据本发明实施例的方法流程的各种操作。

[0154] 根据本发明的实施例,电子设备500还可以包括输入/输出(I/O)接口505,输入/输出(I/O)接口505也连接至总线504。电子设备500还可以包括连接至I/O接口505的以下部件中的一项或多项:包括键盘、鼠标等的输入部分506;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分507;包括硬盘等的存储部分508;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分509。通信部分509经由诸如因特网的网络执行通信处理。驱动器510也根据需要连接至I/O接口505。可拆卸介质511,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器510上,以便于从其上读出的计算机程序根据需要被安装入存储部分508。

[0155] 本发明还提供了一种计算机可读存储介质,该计算机可读存储介质可以是上述实施例中描述的设备/装置/系统中所包含的;也可以是单独存在,而未装配入该设备/装置/系统中。上述计算机可读存储介质承载有一个或者多个程序,当上述一个或者多个程序被执行时,实现根据本发明实施例的方法。

[0156] 根据本发明的实施例,计算机可读存储介质可以是非易失性的计算机可读存储介质,例如可以包括但不限于:便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本发明中,计算机可读存储介质可

以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。例如,根据本发明的实施例,计算机可读存储介质可以包括上文描述的ROM 502和/或RAM 503和/或ROM 502和RAM 503以外的一个或多个存储器。

[0157] 本发明的实施例还包括一种计算机程序产品,其包括计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。当计算机程序产品在计算机系统中运行时,该程序代码用于使计算机系统实现本发明实施例所提供的方法。

[0158] 在该计算机程序被处理器501执行时执行本发明实施例的系统/装置中限定的上述功能。根据本发明的实施例,上文描述的系统、装置、模块、单元等可以通过计算机程序模块来实现。

[0159] 在一种实施例中,该计算机程序可以依托于光存储器件、磁存储器件等有形存储介质。在另一种实施例中,该计算机程序也可以在网络介质上以信号的形式进行传输、分发,并通过通信部分509被下载和安装,和/或从可拆卸介质511被安装。该计算机程序包含的程序代码可以用任何适当的网络介质传输,包括但不限于:无线、有线等等,或者上述的任意合适的组合。

[0160] 在这样的实施例中,该计算机程序可以通过通信部分509从网络上被下载和安装,和/或从可拆卸介质511被安装。在该计算机程序被处理器501执行时,执行本发明实施例的系统中限定的上述功能。根据本发明的实施例,上文描述的系统、设备、装置、模块、单元等可以通过计算机程序模块来实现。

[0161] 根据本发明的实施例,可以以一种或多种程序设计语言的任意组合来编写用于执行本发明实施例提供的计算机程序的程序代码,具体地,可以利用高级过程和/或面向对象的编程语言、和/或汇编/机器语言来实施这些计算程序。程序设计语言包括但不限于诸如Java,C++,python,“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中,远程计算设备可以通过任意种类的网络,包括局域网(LAN)或广域网(WAN),连接到用户计算设备,或者,可以连接到外部计算设备(例如利用因特网服务提供商来通过因特网连接)。

[0162] 附图中的流程图和框图,图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0163] 本领域技术人员可以理解,本发明的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合,即使这样的组合或结合没有明确记载于本发明中。特别地,在不脱离本发明精神和教导的情况下,本发明的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合。所有这些组合和/或结合均落入本发明的范围。

[0164] 以上对本发明的实施例进行了描述。但是,这些实施例仅仅是为了说明的目的,而并非为了限制本发明的范围。尽管在以上分别描述了各实施例,但是这并不意味着各个实施例中的措施不能有利地结合使用。本发明的范围由所附权利要求及其等同物限定。不脱离本发明的范围,本领域技术人员可以做出多种替代和修改,这些替代和修改都应落在本发明的范围之内。

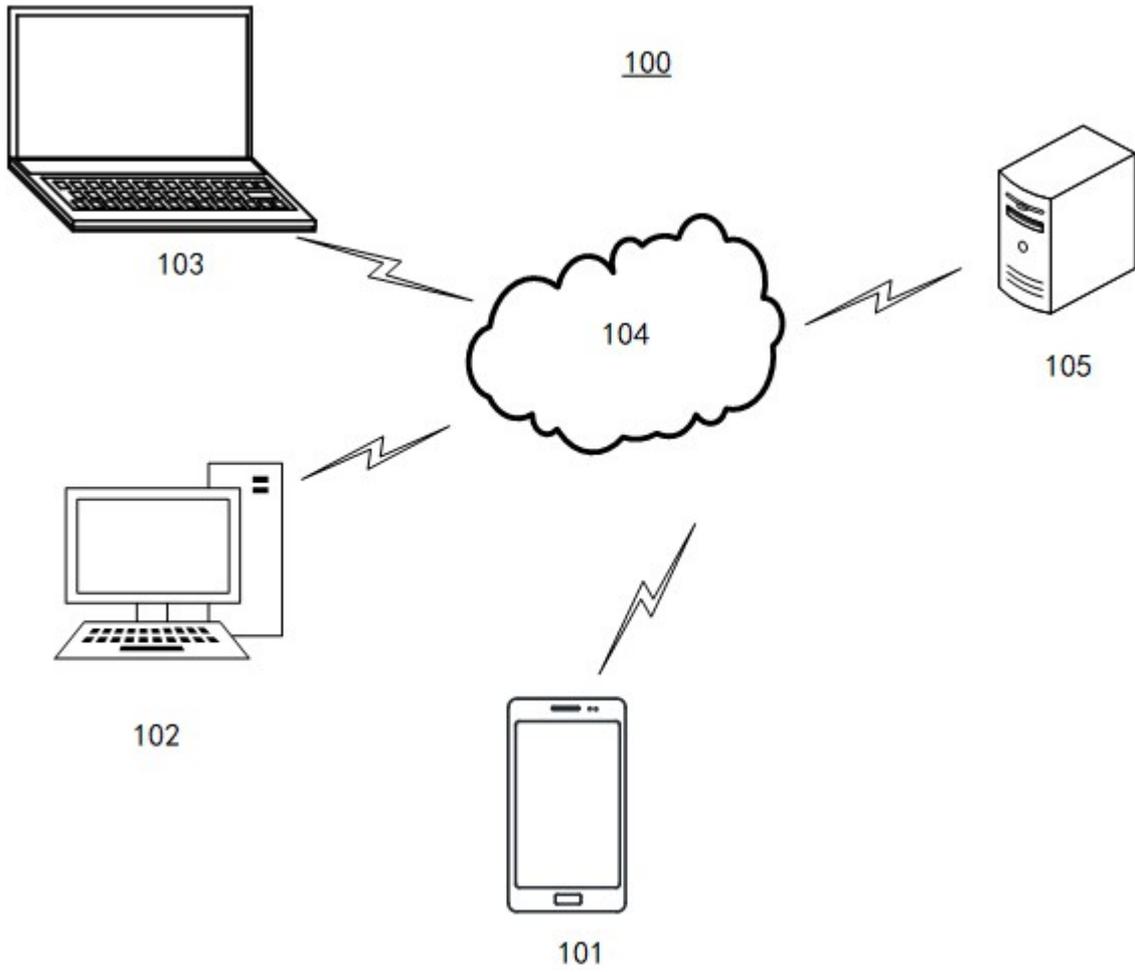


图 1

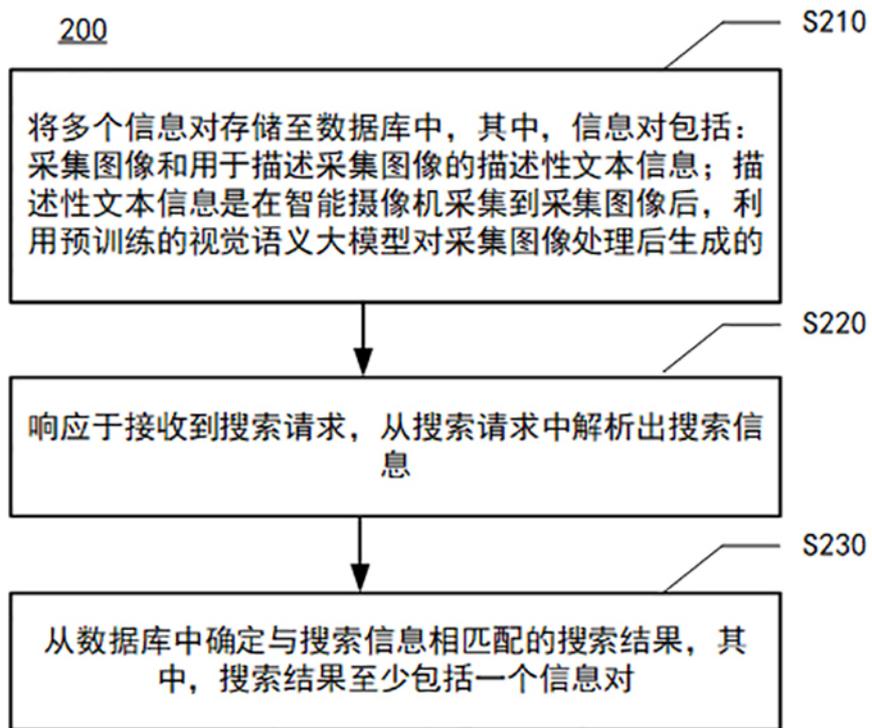


图 2

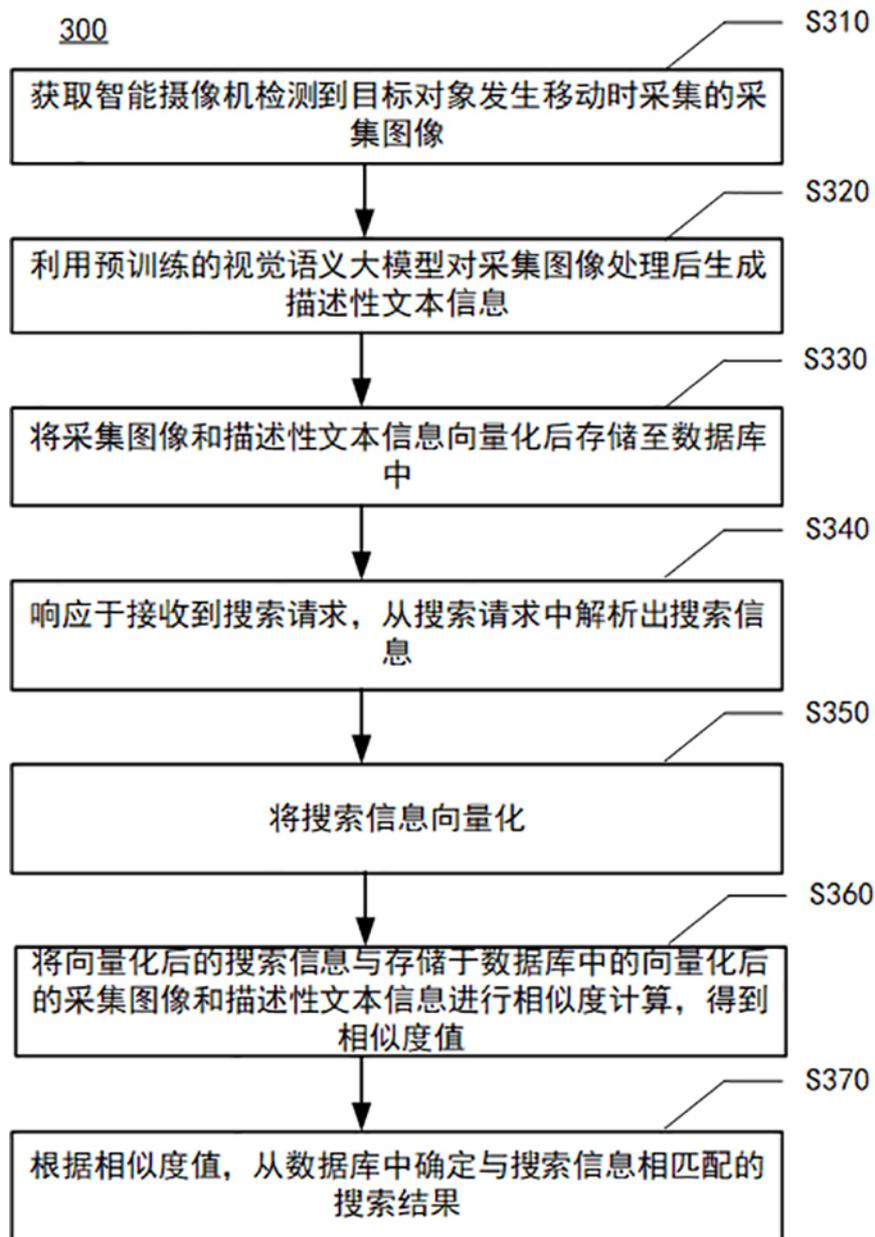


图 3

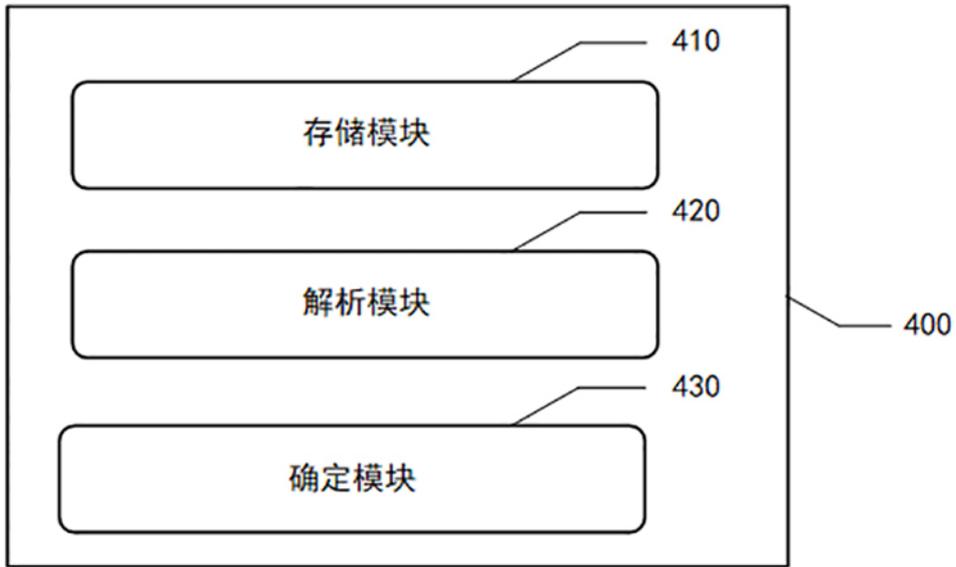


图 4

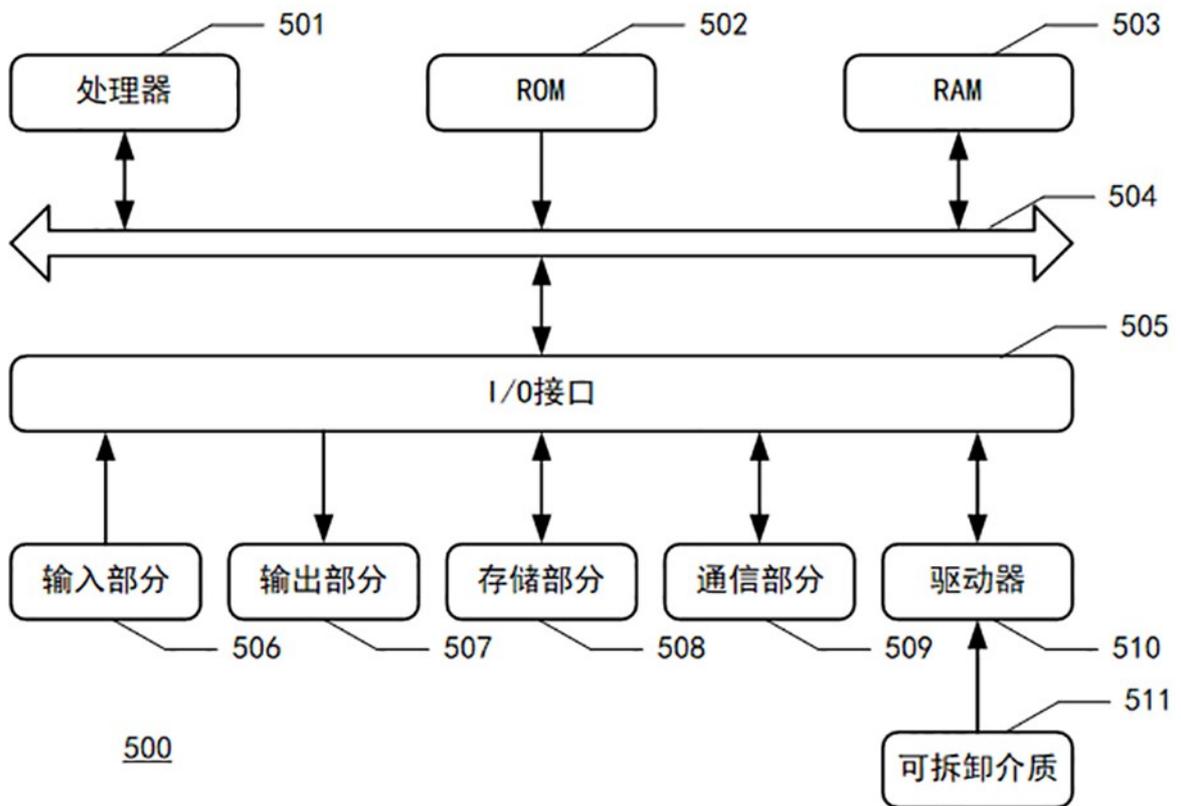


图 5