



(12) 发明专利申请

(10) 申请公布号 CN 113569740 A

(43) 申请公布日 2021. 10. 29

(21) 申请号 202110860912.6

(22) 申请日 2021.07.27

(71) 申请人 合肥综合性国家科学中心人工智能  
研究院(安徽省人工智能实验室)

地址 230094 安徽省合肥市高新区望江西  
路5089号

(72) 发明人 于灵云 方鸣骐 谢洪涛 张勇东

(74) 专利代理机构 中科专利商标代理有限责任  
公司 11021

代理人 鄢功军

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/32 (2006.01)

G06K 9/62 (2006.01)

G10L 25/03 (2013.01)

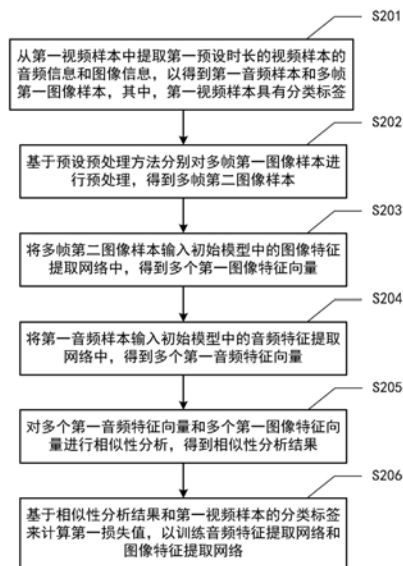
权利要求书3页 说明书14页 附图7页

(54) 发明名称

视频识别模型训练方法与装置、视频识别方法  
与装置

(57) 摘要

本公开提供了一种视频识别模型训练方法,包括:从第一视频样本中提取第一预设时长的视频样本的音频信息和图像信息,以得到第一音频样本和多帧第一图像样本,第一视频样本具有分类标签;基于预设预处理方法分别对多帧第一图像样本进行预处理,得到多帧第二图像样本;将多帧第二图像样本输入初始模型中的图像特征提取网络中,得到多个第一图像特征向量;将第一音频样本输入初始模型中的音频特征提取网络中,得到多个第一音频特征向量;对多个第一音频特征向量和多个第一图像特征向量进行相似性分析,得到相似性分析结果;基于相似性分析结果和第一视频样本的分类标签来计算第一损失值,以训练音频特征提取网络和图像特征提取网络。



1. 一种视频识别模型训练方法,包括:

从第一视频样本中提取第一预设时长的视频样本的音频信息和图像信息,以得到第一音频样本和多帧第一图像样本,其中,所述第一视频样本具有分类标签;

基于预设预处理方法分别对多帧所述第一图像样本进行预处理,得到多帧第二图像样本;

将多帧所述第二图像样本输入初始模型中的图像特征提取网络中,得到多个第一图像特征向量;

将所述第一音频样本输入所述初始模型中的音频特征提取网络中,得到多个第一音频特征向量,其中,所述第一音频特征向量和所述第一图像特征向量的数量和维度均相同;

对多个所述第一音频特征向量和多个所述第一图像特征向量进行相似性分析,得到相似性分析结果;以及

基于所述相似性分析结果和所述第一视频样本的分类标签来计算第一损失值,以训练所述音频特征提取网络和所述图像特征提取网络。

2. 根据权利要求1所述的方法,还包括:

使用第二视频样本对初始音频特征提取网络进行预训练,以得到所述音频特征提取网络;以及

使用第三视频样本对初始图像特征提取网络进行预训练,以得到所述图像特征提取网络。

3. 根据权利要求2所述的方法,其中,所述使用第二视频样本对初始音频特征提取网络进行预训练,包括:

从所述第二视频样本中获取第二音频样本,其中,所述第二音频样本具有第一文本标签;

将所述第二音频样本输入所述初始音频特征提取网络中,得到多个第二音频特征向量;

将多个所述第二音频特征向量输入第一时序网络中,输出得到第一文本信息;以及

基于所述第一文本标签和所述第一文本信息来计算第二损失值,以训练所述初始音频特征提取网络和所述第一时序网络。

4. 根据权利要求2所述的方法,其中,所述使用第三视频样本对初始图像特征提取网络进行预训练,包括:

从所述第三视频样本中获取多帧第三图像样本,其中,所述多帧第三图像样本具有第二文本标签;

基于所述预设预处理方法分别对多帧所述第三图像样本进行预处理,得到多帧第四图像样本;

将多帧所述第四图像样本输入所述初始图像特征提取网络中,得到多个第二图像特征向量;

将多个所述第二图像特征向量输入第二时序网络中,输出得到第二文本信息;以及

基于所述第二文本标签和所述第二文本信息来计算第三损失值,以训练所述初始图像特征提取网络和所述第二时序网络。

5. 根据权利要求1所述的方法,其中,所述对多个所述第一音频特征向量和多个所述第

一图像特征向量进行相似性分析,得到相似性分析结果,包括:

将多个所述第一音频特征向量和多个所述第一图像特征向量分为多组特征向量,其中,每一组所述特征向量包括一个所述第一音频特征向量和一个所述第一图像特征向量;

分别对每一组所述特征向量中所述第一音频特征向量和所述第一图像特征向量进行相似性分析,得到多个中间分析结果;以及

计算所述多个中间分析结果的均值,得到所述相似性分析结果。

6. 根据权利要求1~5中任一项所述的方法,其中,所述预设预处理方法包括:

确定所述第一图像样本或第三图像样本的目标区域;以及

以所述目标区域为中心,将所述第一图像样本或第三图像样本裁减为固定尺寸。

7. 一种使用根据权利要求1~6中任一项所述的方法训练得到的视频识别模型实现的视频识别方法,包括:

从待识别视频中提取第二预设时长的视频片段中的音频数据和多帧视频图像;

基于预设预处理方法分别对多帧视频图像进行预处理,得到多帧处理后的视频图像;

将多帧所述处理后的视频图像输入所述视频识别模型中的图像特征提取网络中,得到多个第三图像特征向量;

将所述音频数据输入所述视频识别模型中的音频特征提取网络中,得到多个第三音频特征向量;

对多个所述第三音频特征向量和多个所述第三图像特征向量进行相似性分析,得到分析结果;以及

基于所述分析结果和预设阈值的大小关系,确定所述待识别视频的真假。

8. 根据权利要求7所述的方法,其中,所述基于所述分析结果和预设阈值的大小关系,确定所述待识别视频的真假,包括:

在所述分析结果小于所述预设阈值的情况下,确定所述待识别视频为虚假视频;以及

在所述分析结果大于或等于所述预设阈值的情况下,确定所述待识别视频为真实视频。

9. 一种视频识别模型训练装置,包括:

第一提取模块,用于从第一视频样本中提取第一预设时长的视频样本的音频信息和图像信息,以得到第一音频样本和多帧第一图像样本,其中,所述第一视频样本具有分类标签;

第一预处理模块,用于基于预设预处理方法分别对多帧所述第一图像样本进行预处理,得到多帧第二图像样本;

第一特征提取模块,用于将多帧所述第二图像样本输入初始模型中的图像特征提取网络中,得到多个第一图像特征向量;

第二特征提取模块,用于将所述第一音频样本输入所述初始模型中的音频特征提取网络中,得到多个第一音频特征向量,其中,所述第一音频特征向量的数量和所述第一图像特征向量的数量相同;

第一分析模块,用于对多个所述第一音频特征向量和多个所述第一图像特征向量进行相似性分析,得到相似性分析结果;以及

训练模块,用于基于所述相似性分析结果和所述第一视频样本的分类标签来计算第一

损失值,以训练所述音频特征提取网络和所述图像特征提取网络。

10.一种使用根据权利要求1~6中任一项所述的方法训练得到的视频识别模型实现的视频识别装置,包括:

第二提取模块,用于从待识别视频中提取第二预设时长的视频片段中的音频数据和多帧视频图像;

第二预处理模块,用于基于预设预处理方法分别对多帧视频图像进行预处理,得到多帧处理后的视频图像;

第三特征提取模块,用于将多帧所述处理后的视频图像输入所述视频识别模型中的图像特征提取网络中,得到多个第三图像特征向量;

第四特征提取模块,用于将所述音频数据输入所述视频识别模型中的音频特征提取网络中,得到多个第三音频特征向量;

第二分析模块,用于对多个所述第三音频特征向量和多个所述第三图像特征向量进行相似性分析,得到分析结果;以及

判断模块,用于基于所述分析结果和预设阈值的大小关系,确定所述待识别视频的真假。

## 视频识别模型训练方法与装置、视频识别方法与装置

### 技术领域

[0001] 本公开涉及人工智能领域,具体地,涉及视频识别技术领域,更具体地,涉及一种视频识别模型训练方法、一种视频识别方法、一种视频识别模型训练装置和一种视频识别装置。

### 背景技术

[0002] 随着互联网的普及,人们习惯于从互联网上获取信息,其中,视频是互联网中信息的重要载体。然而,随着人工智能技术的发展,互联网上开始出现大量虚假视频,例如,通过换脸操作改变视频中人物的身份而得到的虚假视频,或者通过修改视频的音频而得到的虚假视频等。互联网上的大量虚假视频严重威胁着互联网环境安全。

[0003] 在实现本公开构思的过程中发明人发现,相关技术中的视频识别方法使用的音频特征和视频特征的关联性不强,识别精度较低。

### 发明内容

[0004] 有鉴于此,本公开提供了一种视频识别模型训练方法、一种视频识别方法、一种视频识别模型训练装置和一种视频识别装置。

[0005] 本公开的一个方面提供了一种视频识别模型训练方法,包括:从第一视频样本中提取第一预设时长的视频样本的音频信息和图像信息,以得到第一音频样本和多帧第一图像样本,其中,上述第一视频样本具有分类标签;基于预设预处理方法分别对多帧上述第一图像样本进行预处理,得到多帧第二图像样本;将多帧上述第二图像样本输入初始模型中的图像特征提取网络中,得到多个第一图像特征向量;将上述第一音频样本输入上述初始模型中的音频特征提取网络中,得到多个第一音频特征向量,其中,上述第一音频特征向量和上述第一图像特征向量的数量和维度均相同;对多个上述第一音频特征向量和多个上述第一图像特征向量进行相似性分析,得到相似性分析结果;以及基于上述相似性分析结果和上述第一视频样本的分类标签来计算第一损失值,以训练上述音频特征提取网络和上述图像特征提取网络。

[0006] 根据本公开的实施例,上述方法还包括:使用第二视频样本对初始音频特征提取网络进行预训练,以得到上述音频特征提取网络;以及使用第三视频样本对初始图像特征提取网络进行预训练,以得到上述图像特征提取网络。

[0007] 根据本公开的实施例,上述使用第二视频样本对初始音频特征提取网络进行预训练,包括:从上述第二视频样本中获取第二音频样本,其中,上述第二音频样本具有第一文本标签;将上述第二音频样本输入上述初始音频特征提取网络中,得到多个第二音频特征向量;将多个上述第二音频特征向量输入第一时序网络中,输出得到第一文本信息;以及基于上述第一文本标签和上述第一文本信息来计算第二损失值,以训练上述初始音频特征提取网络和上述第一时序网络。

[0008] 根据本公开的实施例,上述使用第三视频样本对初始图像特征提取网络进行预训

练,包括:从上述第三视频样本中获取多帧第三图像样本,其中,上述多帧第三图像样本具有第二文本标签;基于上述预设预处理方法分别对多帧上述第三图像样本进行预处理,得到多帧第四图像样本;将多帧上述第四图像样本输入上述初始图像特征提取网络中,得到多个第二图像特征向量;将多个上述第二图像特征向量输入第二时序网络中,输出得到第二文本信息;以及基于上述第二文本标签和上述第二文本信息来计算第三损失值,以训练上述初始图像特征提取网络和上述第二时序网络。

[0009] 根据本公开的实施例,上述对多个上述第一音频特征向量和多个上述第一图像特征向量进行相似性分析,得到相似性分析结果,包括:将多个上述第一音频特征向量和多个上述第一图像特征向量分为多组特征向量,其中,每一组上述特征向量包括一个上述第一音频特征向量和一个上述第一图像特征向量;分别对每一组上述特征向量中上述第一音频特征向量和上述第一图像特征向量进行相似性分析,得到多个中间分析结果;以及计算上述多个中间分析结果的均值,得到上述相似性分析结果。

[0010] 根据本公开的实施例,上述预设预处理方法包括:确定上述第一图像样本或第三图像样本的目标区域;以及以上述目标区域为中心,将上述第一图像样本或第三图像样本裁减为固定尺寸。

[0011] 本公开的另一个方面提供了一种视频识别方法,包括:从待识别视频中提取第二预设时长的视频片段中的音频数据和多帧视频图像;基于预设预处理方法分别对多帧视频图像进行预处理,得到多帧处理后的视频图像;将多帧上述处理后的视频图像输入上述视频识别模型中的图像特征提取网络中,得到多个第三图像特征向量;将上述音频数据输入上述视频识别模型中的音频特征提取网络中,得到多个第三音频特征向量;对多个上述第三音频特征向量和多个上述第三图像特征向量进行相似性分析,得到视频识别结果;以及基于上述分析结果和预设阈值的大小关系,确定上述待识别视频的真假。

[0012] 根据本公开的实施例,上述基于上述分析结果和预设阈值的大小关系,确定上述待识别视频的真假,包括:在上述分析结果小于上述预设阈值的情况下,确定上述待识别视频为虚假视频;以及在上述分析结果大于或等于上述预设阈值的情况下,确定上述待识别视频为真实视频。

[0013] 本公开的另一个方面提供了一种视频识别模型训练装置,包括第一提取模块、第一预处理模块、第一特征提取模块、第二特征提取模块、第一分析模块和训练模块。其中,第一提取模块,用于从第一视频样本中提取第一预设时长的视频样本的音频信息和图像信息,以得到第一音频样本和多帧第一图像样本,其中,上述第一视频样本具有分类标签;第一预处理模块,用于基于预设预处理方法分别对多帧上述第一图像样本进行预处理,得到多帧第二图像样本;第一特征提取模块,用于将多帧上述第二图像样本输入初始模型中的图像特征提取网络中,得到多个第一图像特征向量;第二特征提取模块,用于将上述第一音频样本输入上述初始模型中的音频特征提取网络中,得到多个第一音频特征向量,其中,上述第一音频特征向量的数量和上述第一图像特征向量的数量相同;第一分析模块,用于对多个上述第一音频特征向量和多个上述第一图像特征向量进行相似性分析,得到相似性分析结果;以及训练模块,用于基于上述相似性分析结果和上述第一视频样本的分类标签来计算第一损失值,以训练上述音频特征提取网络和上述图像特征提取网络。

[0014] 本公开的另一个方面提供了一种视频识别装置,包括第二提取模块、第二预处理

模块、第三特征提取模块、第四特征提取模块、第二分析模块和识别模块。其中，第二提取模块，用于从待识别视频中提取第二预设时长的视频片段中的音频数据和多帧视频图像；第二预处理模块，用于基于预设预处理方法分别对多帧视频图像进行预处理，得到多帧处理后的视频图像；第三特征提取模块，用于将多帧上述处理后的视频图像输入上述视频识别模型中的图像特征提取网络中，得到多个第三图像特征向量；第四特征提取模块，用于将上述音频数据输入上述视频识别模型中的音频特征提取网络中，得到多个第三音频特征向量；第二分析模块，用于对多个上述第三音频特征向量和多个上述第三图像特征向量进行相似性分析，得到视频识别结果；以及判断模块，用于基于上述分析结果和预设阈值的大小关系，确定上述待识别视频的真假。

[0015] 根据本公开的实施例，使用多组第一预设时长的视频样本中的图像信息和音频信息对初始模型中的图像特征提取网络和音频特征提取网络进行训练；在训练过程中，可以对图像信息包含的多帧第一图像样本进行预处理，将预处理后的第二图像样本输入图像特征提取网络，得到第一图像特征向量；将第一音频样本输入音频特征提取网络，得到第一音频特征向量；之后，可以计算第一图像特征向量和第一音频特征向量的相似度，并基于该相似度和分类标签计算得到的损失值来对初始模型的模型参数进行修改，实现对初始模型的训练。通过上述技术手段，至少部分地克服了相关技术中视频特征和音频特征之间关联性不强的技术问题，从而有效提高了训练得到的模型的识别精度。

## 附图说明

[0016] 通过以下参照附图对本公开实施例的描述，本公开的上述以及其他目的、特征和优点将更为清楚，在附图中：

[0017] 图1示意性示出了根据本公开实施例的可以应用视频识别模型训练方法或视频识别方法的示例性系统架构。

[0018] 图2示意性示出了根据本公开实施例的视频识别模型训练方法的流程图。

[0019] 图3示意性示出了根据本公开另一实施例的视频识别模型训练方法的示意图。

[0020] 图4示意性示出了根据本公开的实施例的视频识别方法的流程图。

[0021] 图5示意性示出了根据本公开的实施例的视频识别模型训练装置的框图。

[0022] 图6示意性示出了根据本公开的实施例的视频识别装置的框图。

[0023] 图7示意性示出了根据本公开实施例的适于实现视频识别模型训练方法或视频识别方法的电子设备的框图。

## 具体实施方式

[0024] 以下，将参照附图来描述本公开的实施例。但是应该理解，这些描述只是示例性的，而并非要限制本公开的范围。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公开实施例的全面理解。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公开的概念。

[0025] 在此使用的术语仅仅是为了描述具体实施例，而并非意在限制本公开。在此使用的术语“包括”、“包含”等表明了所述特征、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。

[0026] 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。

[0027] 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)。在使用类似于“A、B或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或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)。

[0028] 相关技术中,视频识别方法大多关注于单一的视觉特征,通过寻找视频图像中的修改伪影、图像模糊等修改痕迹来完成视频识别任务。然而随着虚假视频生成技术的不断发展,生成的虚假视频也越来越逼真,仅仅依靠单一视觉模态的信息已经很难确定视频的真假。

[0029] 此外,另一些视频识别方法立足与虚假视频中的视听不一致的现象,使用音频和视频两个模态的特征来完成视频识别任务。然而,相关技术中的方法的音频和视频两个模态的特征之间的关联性不强,易受到一些干扰信息的影响,例如音频中的背景噪声、图像中的人物肤色、光照等,从而导致模型的识别精度较低。

[0030] 有鉴于此,本公开实施例提供了一种视频识别模型训练方法,训练得到的模型可以有效地提取音、视频两个模态的语义特征,在语义层面对视频进行视听不一致性检测,从而有效地完成视频识别任务。

[0031] 具体地,本公开的实施例提供了一种视频识别模型训练方法、一种视频识别方法、一种视频识别模型训练装置和一种视频识别装置。其中,视频识别模型训练方法包括:从第一视频样本中提取第一预设时长的视频样本的音频信息和图像信息,以得到第一音频样本和多帧第一图像样本,其中,第一视频样本具有分类标签;基于预设预处理方法分别对多帧第一图像样本进行预处理,得到多帧第二图像样本;将多帧第二图像样本输入初始模型中的图像特征提取网络中,得到多个第一图像特征向量;将第一音频样本输入初始模型中的音频特征提取网络中,得到多个第一音频特征向量,其中,第一音频特征向量和第一图像特征向量的数量和维度均相同;对多个第一音频特征向量和多个第一图像特征向量进行相似性分析,得到相似性分析结果;以及基于相似性分析结果和第一视频样本的分类标签来计算第一损失值,以训练音频特征提取网络和图像特征提取网络。

[0032] 图1示意性示出了根据本公开实施例的可以应用视频识别模型训练方法或视频识别方法的示例性系统架构。需要注意的是,图1所示仅为可以应用本公开实施例的系统架构的示例,以帮助本领域技术人员理解本公开的技术内容,但并不意味着本公开实施例不可以用于其他设备、系统、环境或场景。

[0033] 如图1所示,根据该实施例的系统架构100可以包括终端设备101、102、103,网络104和服务器105。网络104用以在终端设备101、102、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,例如有线和/或无线通信链路等等。



[0034] 用户可以使用终端设备101、102、103通过网络104与服务器105交互,以接收或发送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用,例如购物类应用、网页浏览器应用、搜索类应用、即时通信工具、邮箱客户端和/或社交平台软件等(仅为示例)。

[0035] 终端设备101、102、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。

[0036] 服务器105可以是提供各种服务的服务器,例如对用户利用终端设备101、102、103所浏览的网站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用户请求等数据进行分析等处理,并将处理结果(例如根据用户请求获取或生成的网页、信息、或数据等)反馈给终端设备。

[0037] 需要说明的是,本公开实施例所提供的视频识别模型训练方法或视频识别方法一般可以由服务器105执行。相应地,本公开实施例所提供的视频识别模型训练装置或视频识别装置一般可以设置于服务器105中。或者,本公开实施例所提供的视频识别模型训练方法或视频识别方法也可以由终端设备101、102、或103执行。相应地,本公开实施例所提供的视频识别模型训练装置或视频识别装置也可以设置于终端设备101、102、或103中。本公开实施例所提供的视频识别模型训练方法或视频识别方法还可以由能够与终端设备101、102、103和/或服务器105通信的终端设备、服务器或服务器集群执行。相应地,本公开实施例所提供的视频识别模型训练装置或视频识别装置还可以设置于能够与终端设备101、102、103和/或服务器105通信的终端设备、服务器或服务器集群中。

[0038] 例如,本公开实施例所提供的视频识别方法或装置可以部署于网络音视频服务提供商的后台服务器上,或者部署在用户的电子设备上,在虚假视频的传播过程中对其进行检测和拦截;再或者,也可以应用于网络安全部门,用来对虚假视频进行鉴别。

[0039] 再例如,视频样本可以存储在终端设备101、102、103中的任意一个上,也可以存储在服务器105上,还可以存储在网络104中的其他终端设备或服务器上。任意终端设备或者服务器均可执行本公开实施例所提供的视频识别模型训练方法,从本地或者网络104获取视频样本来实现模型的训练。

[0040] 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。

[0041] 图2示意性示出了根据本公开实施例的视频识别模型训练方法的流程图。

[0042] 如图2所示,视频识别模型训练方法包括操作S201~S206。

[0043] 需要说明的是,本公开实施例中的流程图所示的操作除非明确说明不同操作之间存在执行的先后顺序,或者不同操作在技术实现上存在执行的先后顺序,否则,多个操作之间的执行顺序可以不分先后,多个操作也可以同时执行。

[0044] 在操作S201,从第一视频样本中提取第一预设时长的视频样本的音频信息和图像信息,以得到第一音频样本和多帧第一图像样本,其中,第一视频样本具有分类标签。

[0045] 在操作S202,基于预设预处理方法分别对多帧第一图像样本进行预处理,得到多帧第二图像样本。

[0046] 在操作S203,将多帧第二图像样本输入初始模型中的图像特征提取网络中,得到多个第一图像特征向量。

[0047] 在操作S204,将第一音频样本输入初始模型中的音频特征提取网络中,得到多个

第一音频特征向量。

[0048] 在操作S205,对多个第一音频特征向量和多个第一图像特征向量进行相似性分析,得到相似性分析结果。

[0049] 在操作S206,基于相似性分析结果和第一视频样本的分类标签来计算第一损失值,以训练音频特征提取网络和图像特征提取网络。

[0050] 根据本公开的实施例,第一视频样本可以是具有明确的真伪标签的任意视频,包括但不限于共享视频数据库中的视频、用户通过相关视频生成方法生成的虚假视频等。

[0051] 根据本公开的实施例,第一视频样本可以有多个。

[0052] 根据本公开的实施例,第一预设时长可以任意设置,例如可以设置为1s、1min等。

[0053] 根据本公开的实施例,第一音频样本可以是未处理的音频文件,表现为一个一维序列。

[0054] 根据本公开的实施例,每一帧图像样本可以是第一预设时长的视频样本中的一帧图像,多帧第一图像样本可以是在时间上连续的多帧图像。

[0055] 根据本公开的实施例,预设预处理方法至少可以包括对图像大小的处理过程和将图像映射到颜色空间的过程,例如,预设预处理方法可以是将图像以唇部为中心裁剪成 $96 \times 96$ 大小的图像,并裁减后的图像作灰度化处理。

[0056] 根据本公开的实施例,第一图像样本映射到颜色空间后,可以表示为一个二维矩阵或一个三维矩阵。例如,将第一图像样本映射到灰度空间后,可以得到以二维矩阵表示的第二图像样本。再例如,将第一图像样本映射到RGB空间后,可以得到一个三维矩阵表示的第一图像样本,该三维矩阵包括三个二维矩阵;在此情况下,还需要对该三维矩阵进行拆分,以得到三个第二图像样本,其中,每一个第二图像样本可以表示一个二维矩阵。

[0057] 根据本公开的实施例,可以处理后得到的多帧第二图像样本作为一个整体,输入到图像特征提取网络中。

[0058] 根据本公开的实施例,第一音频特征向量和第一图像特征向量的数量和维度可以均相同。

[0059] 根据本公开的实施例,可以使用的相似性分析方法不受限制,例如,距离相似性分析方法、余弦相似性分析方法等。例如,余弦相似性分析方法可如公式(1)所示:

$$D = \frac{f_v^T \cdot f_a}{|f_v| \cdot |f_a|} \quad (1)$$

[0061] 其中,D表示余弦相似度; $f_v$ 表示第一图像特征向量; $f_a$ 表示第一音频特征向量。D越大,表示第一图像特征向量和第一音频特征向量的相似性越高。

[0062] 根据本公开的实施例,可以采用任意策略实现对多个第一音频特征向量和多个第一图像特征向量的相似性分析。例如,可以将多个第一音频特征向量和多个第一图像特征向量分为多组,每一组中存在一个第一音频特征向量和一个第一图像特征向量;再利用相似性分析的相关方法,计算得到每一组特征向量的相似度,取所有相似度的均值作为相似性分析结果。

[0063] 根据本公开的实施例,根据相似性分析方法的不同,第一视频样本的分类标签的值也存在差异。例如,采用的相似性分析方法为余弦相关性分析的情况下,分类标签可以设置为-1表示虚假视频,1表示真实视频。再例如,采用的相似性分析方法为距离相关性分析

的情况下,分类标签可以设置为0表示真实视频,1表示虚假视频;在这种情况下,训练时计算得到的相似性分析结果还需要作归一化处理。

[0064] 根据本公开的实施例,可以采用任意方法来计算第一损失值,包括但不限于均方损失、log损失等。

[0065] 根据本公开的实施例,可以使用任意方法基于一个或多个第一损失值进行模型参数的修改,包括但不限于随机梯度下降法等。

[0066] 根据本公开的实施例,使用多组第一预设时长的视频样本中的图像信息和音频信息对初始模型中的图像特征提取网络和音频特征提取网络进行训练;在训练过程中,可以对图像信息包含的多帧第一图像样本进行预处理,将预处理后的第二图像样本输入图像特征提取网络,得到第一图像特征向量;将第一音频样本输入音频特征提取网络,得到第一音频特征向量;之后,可以计算第一图像特征向量和第一音频特征向量的相似度,并基于该相似度和分类标签计算得到的损失值来对初始模型的模型参数进行修改,实现对初始模型的训练。通过上述技术手段,至少部分地克服了相关技术中视频特征和音频特征之间关联性不强的技术问题,从而有效提高了训练得到的模型的识别精度。

[0067] 下面参考图3,结合具体实施例对图2所示的方法做进一步说明。

[0068] 图3示意性示出了根据本公开另一实施例的视频识别模型训练方法的示意图。

[0069] 如图3所示,视频识别模型训练方法可以包括基于语义特征的预训练的过程,以及针对同一视频样本中的音视频信息的模型训练过程。

[0070] 根据本公开的实施例,预训练过程可以包括使用第三视频样本302对初始图像特征提取网络301进行预训练以得到图像特征提取网络310的第一预训练过程,以及使用第二视频样本312对初始音频特征提取网络311进行预训练以得到音频特征提取网络320的第二预训练过程。

[0071] 根据本公开的实施例,第一预训练过程可以通过如下操作实现:

[0072] 首先,从第三视频样本302中可以获取多帧第三图像样本303作为训练样本。

[0073] 例如,第三视频样本302中可以有a个视频样本,从每一个视频中获取连续的m帧图像作为第三图像样本303进行训练。

[0074] 根据本公开的实施例,第三视频样本302中的所有视频样本可以均为真实样本。

[0075] 根据本公开的实施例,多帧第三图像样本303可以具有第二文本标签304,第二文本标签304可以是与连续的m帧图像对应的时长内视频中人物发出的声音所生成的一维序列。

[0076] 之后,可以对第三图像样本303进行预处理,得到多帧第四图像样本305。

[0077] 根据本公开的实施例,预处理的过程可以是确定第一图像样本或第三图像样本的目标区域;以及以目标区域为中心,将第一图像样本或第三图像样本裁减为固定尺寸。

[0078] 再将多帧第四图像样本305输入到初始图像特征提取网络301中,可以得到多个第二图像特征向量306。

[0079] 例如,将m帧第四图像样本305输入到初始图像特征提取网络301中,可以得到m个n维的第二图像特征向量306。

[0080] 然后,可以将多个第二图像特征向量306输入第二时序网络307中,得到输出的第二文本信息308。

[0081] 根据本公开的实施例,第二文本信息308可以表示为一维序列,且第二文本信息308的长度和第二文本标签304的长度相等。

[0082] 最后,可以基于第二文本信息308和第二文本标签304计算第三损失值309,并利用第三损失值309来调整初始图像特征提取网络301和第二时序网络307的模型参数,以最终得到图像特征提取网络310。

[0083] 根据本公开的实施例,第三损失值309可以通过任意损失函数计算得到,在此并不受限。

[0084] 根据本公开的实施例,可以通过任意方法实现初始视频特征提取网络301和第二时序网络307的模型参数的调整,在此并不受限。

[0085] 根据本公开的实施例,每一次调整初始图像特征提取网络301和第二时序网络307的模型参数时,可以使用一个或多个第三损失值309。

[0086] 根据本公开的实施例,第二预训练过程可以通过如下操作实现:

[0087] 首先,从第二视频样本312中可以获取第二音频样本313作为训练样本。

[0088] 根据本公开的实施例,第二视频样本312中的视频样本可以与第三视频样本302中的视频样本不同。

[0089] 根据本公开的实施例,第二视频样本312中的视频样本可以都是真实视频。

[0090] 例如,第二视频样本312中可以有b个视频样本,从每一个视频样本中截取一段音频作为第二音频样本313进行训练。

[0091] 根据本公开的实施例,第二音频样本313可以表示为一维序列。

[0092] 根据本公开的实施例,第二音频样本313可以具有第一文本标签314。第一文本标签314可以是作为第二音频样本313的音频中人物发出的声音所生成的一维序列。

[0093] 之后,可以将第二音频样本313输入初始音频特征提取网络311中,得到输出的多个第二音频特征向量315。

[0094] 根据本公开的实施例,初始音频特征提取网络311的最后一层可以是均值池化层,例如,将一个第二音频样本313输入初始音频特征提取网络311中,可以得到m个n维的第二音频特征向量315。

[0095] 然后,可以将多个第二音频特征向量315输入第一时序网络316中,得到输出的第一文本信息317。

[0096] 根据本公开的实施例,第一文本信息317可以表示为一维序列,且第一文本信息317的长度和第一文本标签314的长度相等。

[0097] 最后,可以基于第一文本标签314和第一文本信息317来计算第二损失值318,并利用第二损失值318来调整初始音频特征提取网络311和第一时序网络316的模型参数,以最终得到音频特征提取网络320。

[0098] 根据本公开的实施例,第二损失值318可以通过任意损失函数计算得到,在此并不受限。

[0099] 根据本公开的实施例,可以通过任意方法实现初始音频特征提取网络311和第一时序网络316的模型参数的调整,在此并不受限。

[0100] 根据本公开的实施例,每一次调整初始音频特征提取网络311和第一时序网络316的模型参数时,可以使用一个或多个第二损失值318。

[0101] 根据本公开的实施例,通过使用文本标签分别对初始图像特征提取网络301和初始音频特征提取网络311进行预训练,使得训练得到的图像特征提取网络310和音频特征提取网络320更为关注视频和音频两个模态的语义特征,有效避免了无关信息的影响,进而有效提高了视频识别模型的识别精度。

[0102] 根据本公开的实施例,在得到经过预训练的图像特征提取网络310和音频特征提取网络320之后,可以使用第一视频样本321对图像特征提取网络310和音频特征提取网络320进行再次的训练。

[0103] 根据本公开的实施例,第一视频样本321可以具有分类标签322,从第一视频样本321中可以获取多帧第一图像样本323和第一音频样本324。多帧第一图像样本323需要进行预处理,可以得到多帧第二图像样本325。之后,可以将多帧第二图像样本325输入到图像特征提取网络310中,输出得到多个第一图像特征向量326;可以将第一音频样本324输入到音频特征提取网络320中,输出得到多个第一音频特征向量327。

[0104] 根据本公开的实施例,多个第一图像特征向量326和多个第一音频特征向量327的相似性分析结果的获取可以通过如下操作实现:首先将多个第一图像特征向量326和多个第一音频特征向量327分为多组特征向量,其中,每一组特征向量包括一个第一图像特征向量326和一个第一音频特征向量327;然后,分别对每一组特征向量中第一图像特征向量326和第一音频特征向量327进行相似性分析,得到多个中间分析结果;再计算多个中间分析结果的均值,得到相似性分析结果328。

[0105] 根据本公开的实施例,根据相似性分析结构328和分类标签322,可以计算得到第一损失值329,并根据第一损失值329调整图像特征提取网络310和音频特征提取网络320的模型参数。

[0106] 根据本公开的实施例,使用第一视频样本321对图像特征提取网络310和音频特征提取网络320进行再次的训练具体可以参见操作S201~S206的方法,在此不再赘述。

[0107] 图4示意性示出了根据本公开的实施例的视频识别方法的流程图。

[0108] 如图4所示,视频识别方法包括操作S401~S406。

[0109] 在操作S401,从待识别视频中提取第二预设时长的视频片段中的音频数据和多帧视频图像。

[0110] 在操作S402,基于预设预处理方法分别对多帧视频图像进行预处理,得到多帧处理后的视频图像。

[0111] 在操作S403,将多帧处理后的视频图像输入视频识别模型中的图像特征提取网络中,得到多个第三图像特征向量。

[0112] 在操作S404,将音频数据输入视频识别模型中的音频特征提取网络中,得到多个第三音频特征向量。

[0113] 在操作S405,对多个第三音频特征向量和多个第三图像特征向量进行相似性分析,得到分析结果。

[0114] 在操作S406,基于分析结果和预设阈值的大小关系,确定待识别视频的真假。

[0115] 根据本公开的实施例,视频识别方法中的数据处理过程可以参考视频识别训练方法中的数据处理过程,在此不再赘述。

[0116] 根据本公开的实施例,根据选择的相似性分析方法的不同,根据分析结果和预设

阈值的大小来判断待识别视频的真假的结果也有所不同。例如,采用的相似性分析方法为余弦相似性分析方法,则在分析结果小于预设阈值的情况下,确定待识别视频为虚假视频;在分析结果大于或等于预设阈值的情况下,确定待识别视频为真实视频。

[0117] 根据本公开的实施例,预设阈值可以是视频识别模型的一个超参,也可以利用一组视频作为验证集来设置。例如,将真实视频输入视频识别模型后得到的分析结果为*i*,将虚假视频输入视频识别模型后得到的分析结果为*j*,则可以设置视频识别模型的预设阈值为 $(i+j)/2$ 。

[0118] 图5示意性示出了根据本公开的实施例的视频识别模型训练装置的框图。

[0119] 如图5所示,视频识别模型训练装置包括第一提取模块510、第一预处理模块520、第一特征提取模块530、第二特征提取模块540、第一分析模块550和训练模块560。

[0120] 第一提取模块510,用于从第一视频样本中提取第一预设时长的视频样本的音频信息和图像信息,以得到第一音频样本和多帧第一图像样本,其中,第一视频样本具有分类标签。

[0121] 第一预处理模块520,用于基于预设预处理方法分别对多帧第一图像样本进行预处理,得到多帧第二图像样本。

[0122] 第一特征提取模块530,用于将多帧第二图像样本输入初始模型中的图像特征提取网络中,得到多个第一图像特征向量。

[0123] 第二特征提取模块540,用于将第一音频样本输入初始模型中的音频特征提取网络中,得到多个第一音频特征向量,其中,第一音频特征向量的数量和第一图像特征向量的数量相同。

[0124] 第一分析模块550,用于对多个第一音频特征向量和多个第一图像特征向量进行相似性分析,得到相似性分析结果。

[0125] 训练模块560,用于基于相似性分析结果和第一视频样本的分类标签来计算第一损失值,以训练音频特征提取网络和图像特征提取网络。

[0126] 根据本公开的实施例,使用多组第一预设时长的视频样本中的图像信息和音频信息对初始模型中的图像特征提取网络和音频特征提取网络进行训练;在训练过程中,可以对图像信息包含的多帧第一图像样本进行预处理,将预处理后的第二图像样本输入图像特征提取网络,得到第一图像特征向量;将第一音频样本输入音频特征提取网络,得到第一音频特征向量;之后,可以计算第一图像特征向量和第一音频特征向量的相似度,并基于该相似度和分类标签计算得到的损失值来对初始模型的模型参数进行修改,实现对初始模型的训练。通过上述技术手段,至少部分地克服了相关技术中视频特征和音频特征之间关联性不强的技术问题,从而有效提高了训练得到的模型的识别精度。

[0127] 根据本公开的实施例,视频识别模型训练装置还包括第一预训练模块和第二预训练模块。

[0128] 第一预训练模块,用于使用第二视频样本对初始音频特征提取网络进行预训练,以得到音频特征提取网络。

[0129] 第二预训练模块,用于使用第三视频样本对初始图像特征提取网络进行预训练,以得到图像特征提取网络。

[0130] 根据本公开的实施例,第一预训练模块包括第一预训练单元、第二预训练单元、第

三预训练单元和第四预训练单元。

[0131] 第一预训练单元,用于从第二视频样本中获取第二音频样本,其中,第二音频样本具有第一文本标签。

[0132] 第二预训练单元,用于将第二音频样本输入初始音频特征提取网络中,得到多个第二音频特征向量。

[0133] 第三预训练单元,用于将多个第二音频特征向量输入第一时序网络中,输出得到第一文本信息。

[0134] 第四预训练单元,用于基于第一文本标签和第一文本信息来计算第二损失值,以训练初始音频特征提取网络和第一时序网络。

[0135] 根据本公开的实施例,第二预训练模块包括第五预训练单元、第六预训练单元、第七预训练单元、第八预训练单元和第九预训练单元。

[0136] 第五预训练单元,用于从第三视频样本中获取多帧第三图像样本,其中,多帧第三图像样本具有第二文本标签。

[0137] 第六预训练单元,用于基于预设预处理方法分别对多帧第三图像样本进行预处理,得到多帧第四图像样本。

[0138] 第七预训练单元,用于将多帧第四图像样本输入初始图像特征提取网络中,得到多个第二图像特征向量。

[0139] 第八预训练单元,用于将多个第二图像特征向量输入第二时序网络中,输出得到第二文本信息。

[0140] 第九预训练单元,用于基于第二文本标签和第二文本信息来计算第三损失值,以训练初始图像特征提取网络和第二时序网络。

[0141] 根据本公开的实施例,第一分析模块550包括第一分析单元、第二分析单元和第三分析单元。

[0142] 第一分析单元,用于将多个第一音频特征向量和多个第一图像特征向量分为多组特征向量,其中,每一组特征向量包括一个第一音频特征向量和一个第一图像特征向量。

[0143] 第二分析单元,用于分别对每一组特征向量中第一音频特征向量和第一图像特征向量进行相似性分析,得到多个中间分析结果。

[0144] 第三分析单元,用于计算多个中间分析结果的均值,得到相似性分析结果。

[0145] 图6示意性示出了根据本公开的实施例的视频识别装置的框图。

[0146] 如图6所示,视频识别装置包括第二提取模块610、第二预处理模块620、第三特征提取模块630、第四特征提取模块640、第二分析模块650和判断模块660。

[0147] 第二提取模块610,用于从待识别视频中提取第二预设时长的视频片段中的音频数据和多帧视频图像;

[0148] 第二预处理模块620,用于基于预设预处理方法分别对多帧视频图像进行预处理,得到多帧处理后的视频图像;

[0149] 第三特征提取模块630,用于将多帧处理后的视频图像输入视频识别模型中的图像特征提取网络中,得到多个第三图像特征向量;

[0150] 第四特征提取模块640,用于将音频数据输入视频识别模型中的音频特征提取网络中,得到多个第三音频特征向量;

[0151] 第二分析模块650,用于对多个第三音频特征向量和多个第三图像特征向量进行相似性分析,得到分析结果;以及

[0152] 判断模块660,用于基于分析结果和预设阈值的大小关系,确定待识别视频的真假。

[0153] 根据本公开的实施例,判断模块660包括第一判断单元和第二判断单元。

[0154] 第一判断单元,用于在分析结果小于预设阈值的情况下,确定待识别视频为虚假视频。

[0155] 第二判断单元,用于在分析结果大于或等于预设阈值的情况下,确定待识别视频为真实视频。

[0156] 根据本公开的实施例的模块、单元、子单元中的任意多个、或其中任意多个的至少部分功能可以在一个模块中实现。根据本公开实施例的模块、单元、子单元中的任意一个或多个可以被拆分成多个模块来实现。根据本公开实施例的模块、单元、子单元中的任意一个或多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的硬件或固件来实现,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,根据本公开实施例的模块、单元、子单元中的一个或多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。

[0157] 例如,第一提取模块510、第一预处理模块520、第一特征提取模块530、第二特征提取模块540、第一分析模块550和训练模块560和/或视频识别装置包括第二提取模块610、第二预处理模块620、第三特征提取模块630、第四特征提取模块640、第二分析模块650和判断模块660中的任意多个可以合并在一个模块/单元/子单元中实现,或者其中的任意一个模块/单元/子单元可以被拆分成多个模块/单元/子单元。或者,这些模块/单元/子单元中的一个或多个模块/单元/子单元的至少部分功能可以与其他模块/单元/子单元的至少部分功能相结合,并在一个模块/单元/子单元中实现。根据本公开的实施例,第一提取模块510、第一预处理模块520、第一特征提取模块530、第二特征提取模块540、第一分析模块550和训练模块560和/或视频识别装置包括第二提取模块610、第二预处理模块620、第三特征提取模块630、第四特征提取模块640、第二分析模块650和判断模块660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,第一提取模块510、第一预处理模块520、第一特征提取模块530、第二特征提取模块540、第一分析模块550和训练模块560和/或视频识别装置包括第二提取模块610、第二预处理模块620、第三特征提取模块630、第四特征提取模块640、第二分析模块650和判断模块660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。

[0158] 需要说明的是,本公开的实施例中视频识别模型训练装置部分与本公开的实施例中视频识别模型训练方法部分是相对应的,本公开的实施例中视频识别装置部分与本公开的实施例中视频识别方法部分是相对应的,视频识别模型训练装置部分和视频识别装置部



分的描述具体参考视频识别模型训练方法和视频识别方法部分,在此不再赘述。

[0159] 图7示意性示出了根据本公开实施例的适于实现视频识别模型训练方法或视频识别方法的电子设备的框图。图7示出的电子设备仅仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。

[0160] 如图7所示,根据本公开实施例的计算机电子设备700包括处理器701,其可以根据存储在只读存储器 (ROM) 702中的程序或者从存储部分708加载到随机访问存储器 (RAM) 703中的程序而执行各种适当的动作和处理。处理器701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701还可以包括用于缓存用途的板载存储器。处理器701可以包括用于执行根据本公开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。

[0161] 在RAM 703中,存储有电子设备700操作所需的各种程序和数据。处理器701、ROM 702以及RAM 703通过总线704彼此相连。处理器701通过执行ROM 702和/或RAM 703中的程序来执行根据本公开实施例的方法流程的各种操作。需要注意,所述程序也可以存储在除ROM702和RAM 703以外的一个或多个存储器中。处理器701也可以通过执行存储在所述一个或多个存储器中的程序来执行根据本公开实施例的方法流程的各种操作。

[0162] 根据本公开的实施例,电子设备700还可以包括输入/输出 (I/O) 接口705,输入/输出 (I/O) 接口705也连接至总线704。电子设备700还可以包括连接至I/O接口705的以下部件中的一项或多项:包括键盘、鼠标等的输入部分706;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707;包括硬盘等的存储部分708;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709。通信部分709经由诸如因特网的网络执行通信处理。驱动器710也根据需要连接至I/O接口705。可拆卸介质711,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710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708。

[0163] 根据本公开的实施例,根据本公开实施例的方法流程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读存储介质上的计算机程序,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。在这样的实施例中,该计算机程序可以通过通信部分709从网络上被下载和安装,和/或从可拆卸介质711被安装。在该计算机程序被处理器701执行时,执行本公开实施例的系统中限定的上述功能。根据本公开的实施例,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。

[0164] 本公开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,实现根据本公开实施例的方法。

[0165] 根据本公开的实施例,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。在本公开中,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。

[0166] 例如,根据本公开的实施例,计算机可读存储介质可以包括上文描述的ROM 702和/或RAM 703和/或ROM 702和RAM 703以外的一个或多个存储器。

[0167] 本公开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本公开实施例所提供的方法的程序代码,当计算机程序产品在电子设备上运行时,该程序代码用于使电子设备实现本公开实施例所提供的方法。

[0168] 在该计算机程序被处理器701执行时,执行本公开实施例的系统/装置中限定的上述功能。根据本公开的实施例,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。

[0169] 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709被下载和安装,和/或从可拆卸介质711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。

[0170] 根据本公开的实施例,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公开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)。

[0171] 附图中的流程图和框图,图示了按照本公开各种实施例的系统、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。本领域技术人员可以理解,本公开的各个实施例和/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公开中。特别地,在不脱离本公开精神和教导的情况下,本公开的各个实施例和/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/或结合均落入本公开的范围。

[0172] 以上对本公开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,而并非为了限制本公开的范围。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。本公开的范围由所附权利要求及其等同物限定。不脱离本公开的范围,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公开的范围之内。

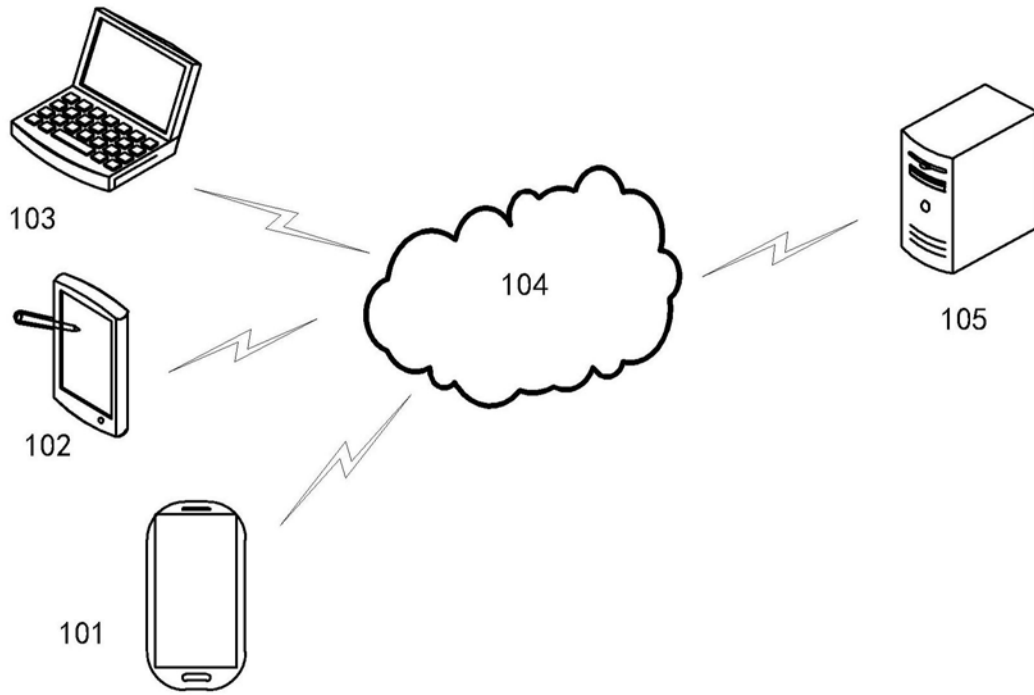


图1

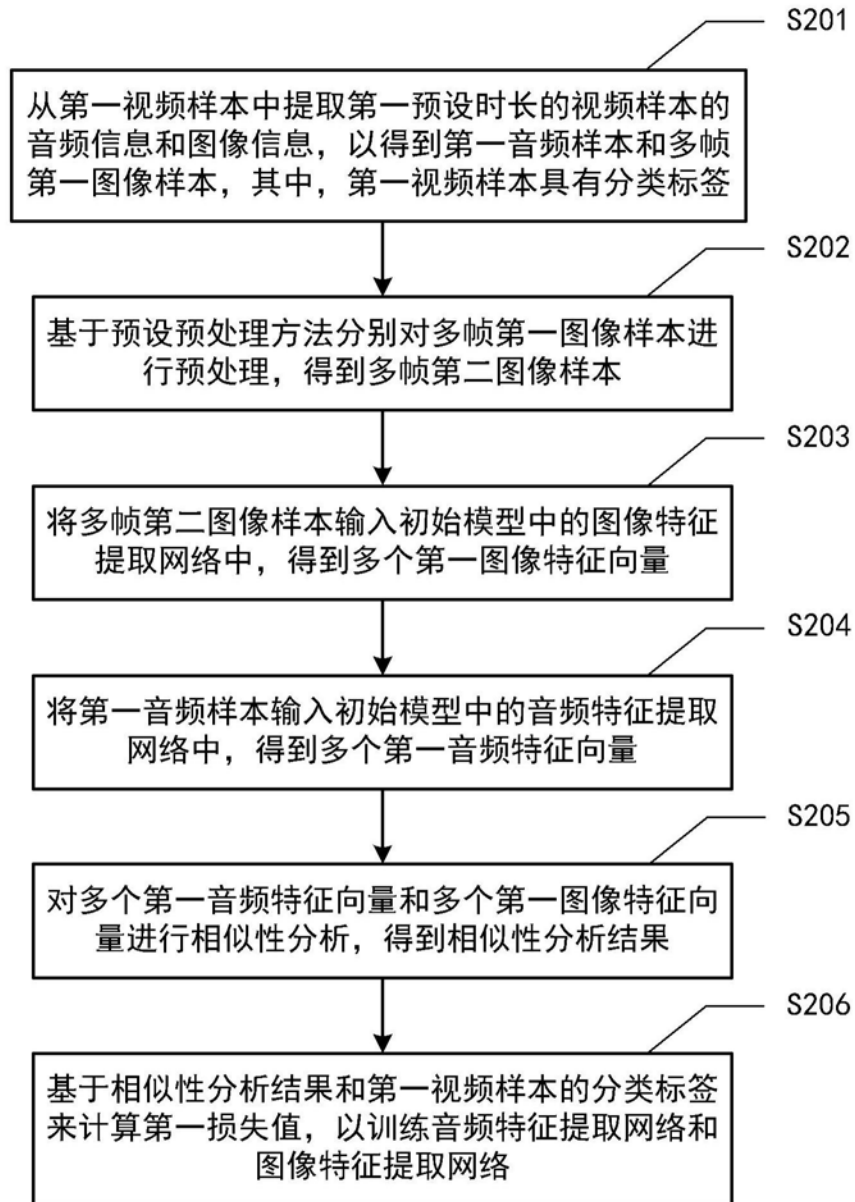


图2

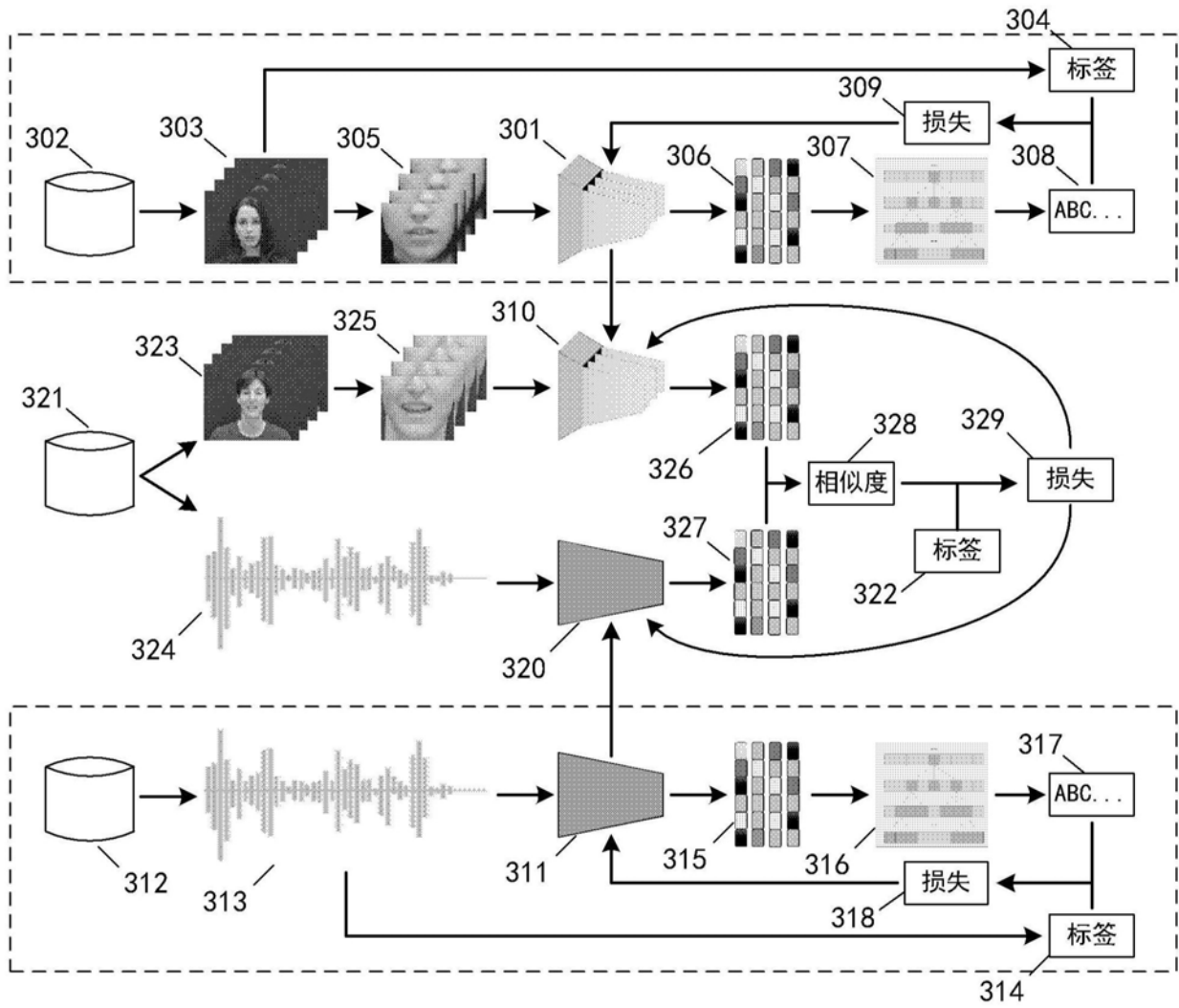


图3

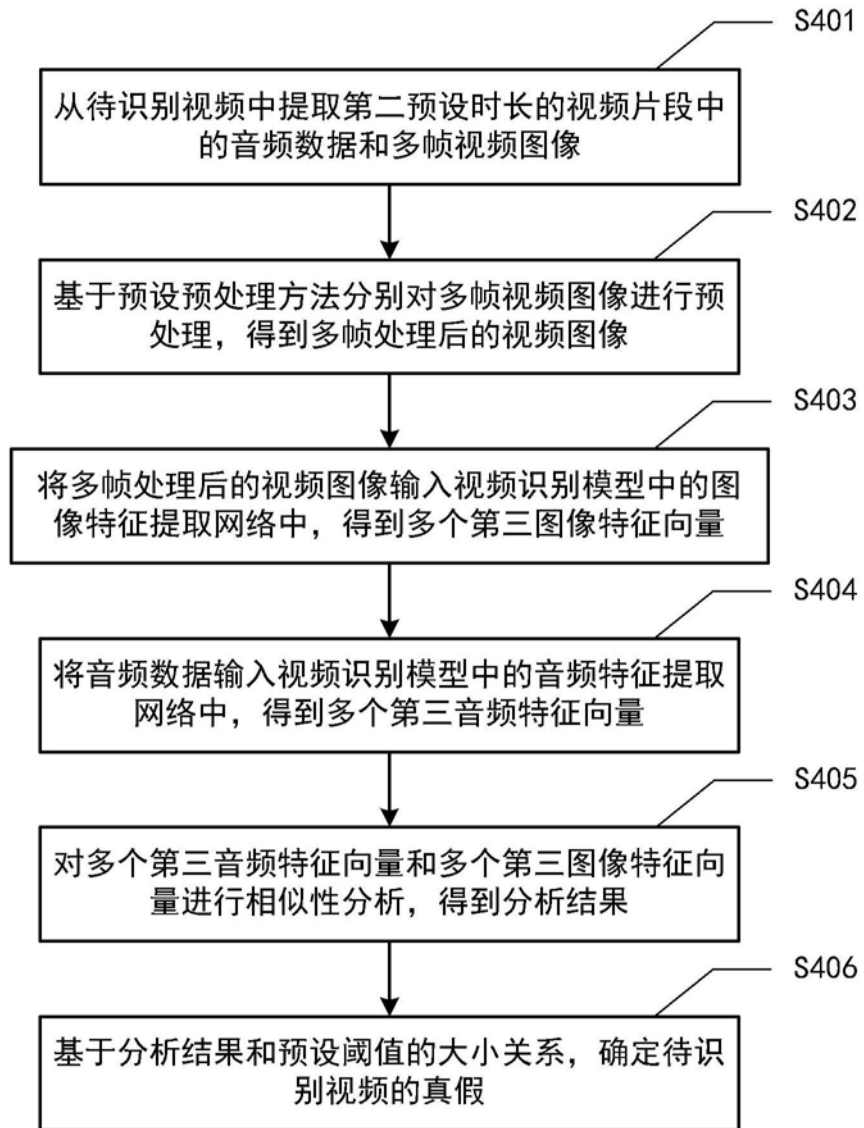


图4

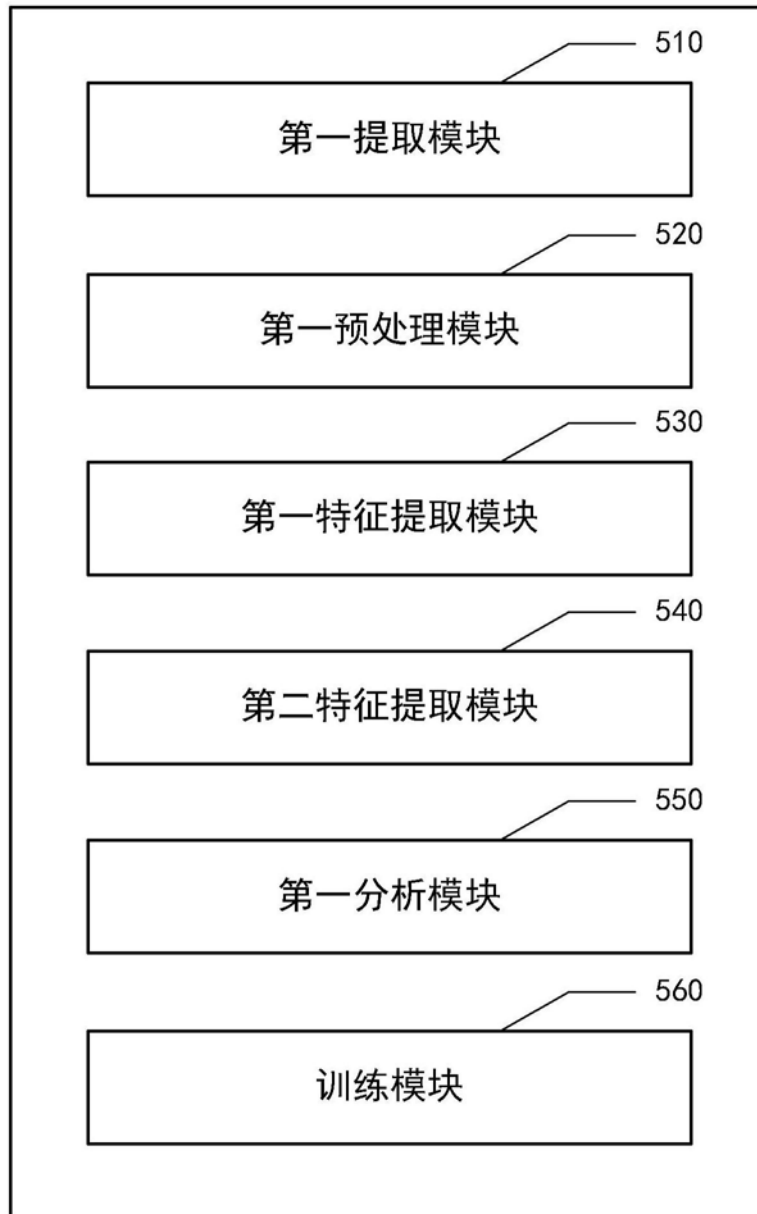


图5

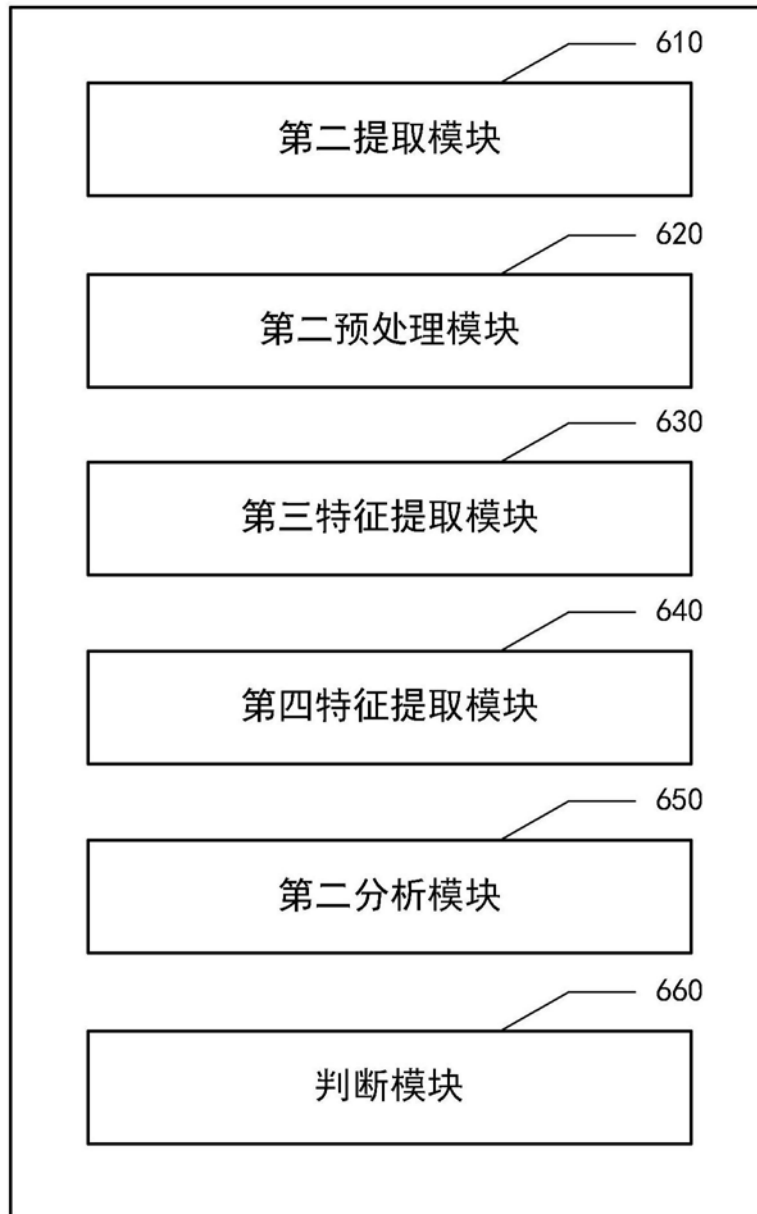


图6



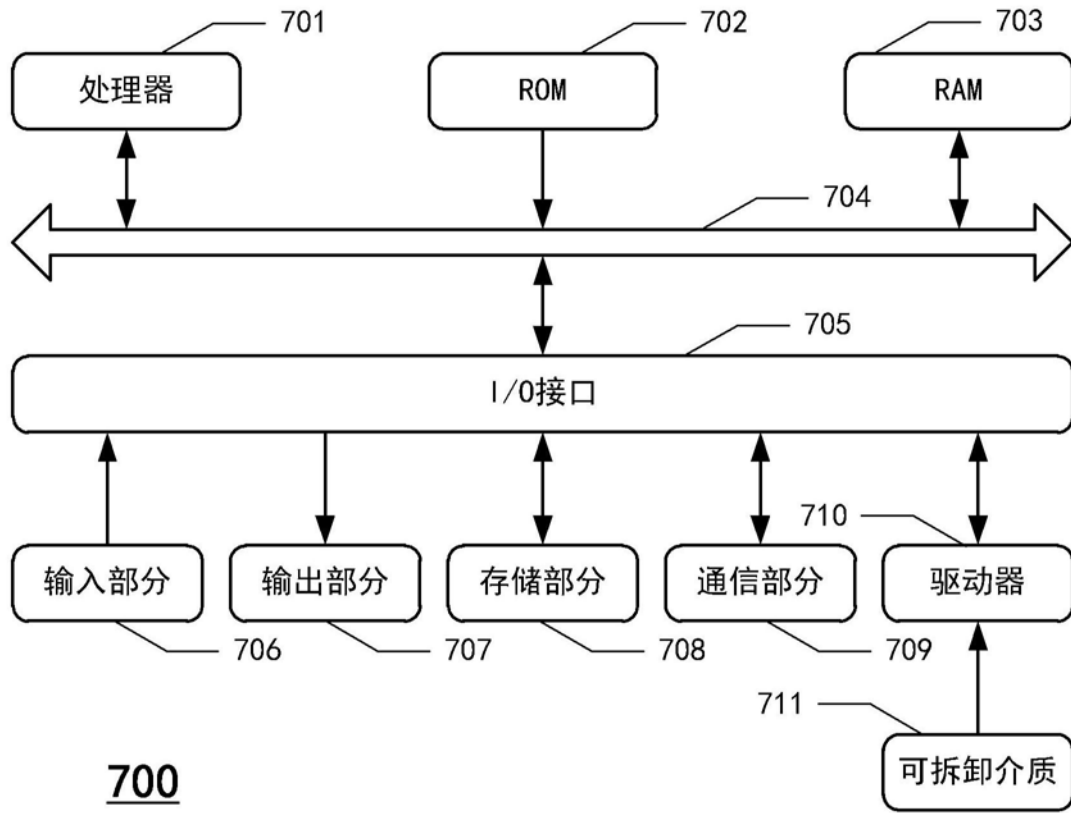


图7