



(12) 发明专利申请

(10) 申请公布号 CN 116563669 A

(43) 申请公布日 2023. 08. 08

(21) 申请号 202310539771.7

(22) 申请日 2023.05.12

(71) 申请人 北京百度网讯科技有限公司  
地址 100085 北京市海淀区上地十街10号  
百度大厦2层

(72) 发明人 崔东林 李滨伯

(74) 专利代理机构 上海弼兴律师事务所 31283  
专利代理师 罗朗

(51) Int. Cl.

- G06V 10/774 (2022.01)
- G06V 10/764 (2022.01)
- G06V 20/40 (2022.01)
- G06N 3/088 (2023.01)
- G06V 10/74 (2022.01)
- G06V 10/82 (2022.01)

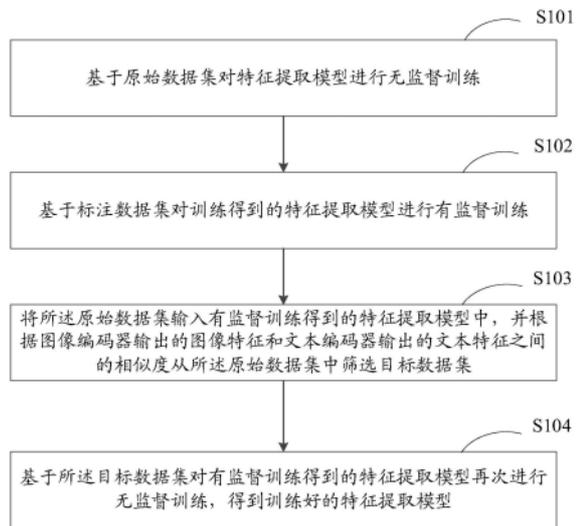
权利要求书3页 说明书9页 附图5页

(54) 发明名称

模型的训练方法、视频分类方法、装置及设备

(57) 摘要

本公开提供了一种模型的训练方法、视频分类方法、装置及设备,涉及人工智能技术领域,尤其涉及深度学习、视频处理等技术领域。具体实现方案为:基于原始数据集对特征提取模型进行无监督训练;其中,所述特征提取模型包括图像编码器和文本编码器;基于标注数据集对训练得到的特征提取模型进行有监督训练;将所述原始数据集输入有监督训练得到的特征提取模型中,并根据图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度从所述原始数据集中筛选目标数据集;基于所述目标数据集对有监督训练得到的特征提取模型再次进行无监督训练,得到训练好的特征提取模型。



1. 一种特征提取模型的训练方法,包括:

基于原始数据集对特征提取模型进行无监督训练;其中,所述原始数据集包括若干第一样本视频帧以及所述第一样本视频帧对应的文本说明,所述特征提取模型包括图像编码器和文本编码器,所述图像编码器用于对所述第一样本视频帧进行编码,得到对应的图像特征,所述文本编码器用于对所述文本说明进行编码,得到对应的文本特征;

基于标注数据集对训练得到的特征提取模型进行有监督训练;其中,所述标注数据集包括第二样本视频帧、针对所述第二样本视频帧标注的图像特征、所述第二样本视频帧对应的文本说明以及针对所述文本说明标注的文本特征;

将所述原始数据集输入有监督训练得到的特征提取模型中,并根据图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度从所述原始数据集中筛选目标数据集;

基于所述目标数据集对有监督训练得到的特征提取模型再次进行无监督训练,得到训练好的特征提取模型。

2. 根据权利要求1所述的训练方法,其中,所述根据图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度从所述原始数据集中筛选目标数据集,包括:

若图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度大于预设值,则将所述原始数据集中与所述图像特征对应的第一样本视频帧以及与所述文本特征对应的文本说明确定为目标数据集。

3. 根据权利要求1所述的训练方法,其中,所述根据图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度从所述原始数据集中筛选目标数据集,包括:

对于图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度,按照从高到低进行排序;

选取排序靠前的预设数量的相似度对应的图像特征和文本特征;

将所述原始数据集中与所述图像特征对应的第一样本视频帧以及与所述文本特征对应的文本说明确定为目标数据集。

4. 根据权利要求1-3中任一项所述的训练方法,其中,所述文本说明为标题。

5. 一种视频分类模型的训练方法,包括:

获取训练样本;其中,所述训练样本包括样本视频和所述样本视频的分类标签;

将所述样本视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述样本视频的分类结果;其中,所述视频分类模型包括依次连接的图像编码器和全连接层,所述图像编码器为利用权利要求1-4中任一项所述的训练方法训练得到的特征提取模型中的图像编码器;

根据所述分类标签和所述分类结果计算损失,并根据所述损失调整所述视频分类模型中全连接层的参数,直至满足收敛条件;

输出训练好的视频分类模型。

6. 根据权利要求5所述的训练方法,其中,所述视频分类模型还包括卷积层,设于所述图像编码器和所述全连接层之间;

所述根据所述损失调整所述视频分类模型中全连接层的参数,包括:根据所述损失调整所述视频分类模型中卷积层和全连接层的参数。

7. 一种视频分类方法,包括:

获取待分类视频;

将所述待分类视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述待分类视频的分类结果;

其中,所述视频分类模型根据权利要求5或6所述的训练方法训练得到。

8. 一种特征提取模型的训练装置,包括:

第一训练模块,用于基于原始数据集对特征提取模型进行无监督训练;其中,所述原始数据集包括若干第一样本视频帧以及所述第一样本视频帧对应的文本说明,所述特征提取模型包括图像编码器和文本编码器,所述图像编码器用于对所述第一样本视频帧进行编码,得到对应的图像特征,所述文本编码器用于对所述文本说明进行编码,得到对应的文本特征;

第二训练模块,用于基于标注数据集对训练得到的特征提取模型进行有监督训练;其中,所述标注数据集包括第二样本视频帧、针对所述第二样本视频帧标注的图像特征、所述第二样本视频帧对应的文本说明以及针对所述文本说明标注的文本特征;

数据筛选模块,用于将所述原始数据集输入有监督训练得到的特征提取模型中,并根据图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度从所述原始数据集中筛选目标数据集;

第三训练模块,用于基于所述目标数据集对有监督训练得到的特征提取模型再次进行无监督训练,得到训练好的特征提取模型。

9. 根据权利要求8所述的训练装置,其中,所述数据筛选模块具体用于在图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度大于预设值的情况下,将所述原始数据集中与所述图像特征对应的第一样本视频帧以及与所述文本特征对应的文本说明确定为目标数据集。

10. 根据权利要求8所述的训练装置,其中,所述数据筛选模块具体用于对于图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度,按照从高到低进行排序,并选取排序靠前的预设数量的相似度对应的图像特征和文本特征,以及将所述原始数据集中与所述图像特征对应的第一样本视频帧以及与所述文本特征对应的文本说明确定为目标数据集。

11. 根据权利要求8-10中任一项所述的训练装置,其中,所述文本说明为标题。

12. 一种视频分类模型的训练装置,包括:

第一获取模块,用于获取训练样本;其中,所述训练样本包括样本视频和所述样本视频的分类标签;

第一预测模块,用于将所述样本视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述样本视频的分类结果;其中,所述视频分类模型包括依次连接的图像编码器和全连接层,所述图像编码器为利用权利要求8-11中任一项所述的训练装置得到的特征提取模型中的图像编码器;

参数处理模块,用于根据所述分类标签和所述分类结果计算损失,并根据所述损失调整所述视频分类模型中全连接层的参数,直至满足收敛条件;

模型输出模块,用于输出训练好的视频分类模型。

13. 根据权利要求12所述的训练装置,其中,所述视频分类模型还包括卷积层,设于所述图像编码器和所述全连接层之间;

所述参数处理模块具体用于根据所述损失调整所述视频分类模型中卷积层和全连接层的参数。

14. 一种视频分类装置,包括:

第二获取模块,用于获取待分类视频;

第二预测模块,用于将所述待分类视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述待分类视频的分类结果;

其中,所述视频分类模型根据权利要求12或13所述的训练装置训练得到。

15. 一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-4中任一项所述的特征提取模型的训练方法、权利要求5或6所述的视频分类模型的训练方法或者权利要求7所述的视频分类方法。

16. 一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-4中任一项所述的特征提取模型的训练方法、权利要求5或6所述的视频分类模型的训练方法或者权利要求7所述的视频分类方法。

17. 一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1-4中任一项所述的特征提取模型的训练方法、权利要求5或6所述的视频分类模型的训练方法或者权利要求7所述的视频分类方法。

## 模型的训练方法、视频分类方法、装置及设备

### 技术领域

[0001] 本公开涉及人工智能技术领域,尤其涉及深度学习、视频处理等技术领域。

### 背景技术

[0002] 随着人工智能技术的发展,深度学习模型得到了广泛的应用,例如可以将深度学习模型应用于各种视频处理场景中。在训练深度学习模型时,往往需要用到大量的样本数据,样本数据的数据量越大,深度学习模型的学习效果越好。但是,如果样本数据中的样本噪声较大,会严重影响深度学习模型的学习效果。

### 发明内容

[0003] 本公开提供了一种特征提取模型的训练方法及装置、一种视频分类模型的训练方法及装置、一种视频分类方法及装置、电子设备、存储介质以及计算机程序产品。

[0004] 根据本公开的一方面,提供了一种特征提取模型的训练方法,包括:基于原始数据集对特征提取模型进行无监督训练;其中,所述原始数据集包括若干第一样本视频帧以及所述第一样本视频帧对应的文本说明,所述特征提取模型包括图像编码器和文本编码器,所述图像编码器用于对所述第一样本视频帧进行编码,得到对应的图像特征,所述文本编码器用于对所述文本说明进行编码,得到对应的文本特征;基于标注数据集对训练得到的特征提取模型进行有监督训练;其中,所述标注数据集包括第二样本视频帧、针对所述第二样本视频帧标注的图像特征、所述第二样本视频帧对应的文本说明以及针对所述文本说明标注的文本特征;将所述原始数据集输入有监督训练得到的特征提取模型中,并根据图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度从所述原始数据集中筛选目标数据集;基于所述目标数据集对有监督训练得到的特征提取模型再次进行无监督训练,得到训练好的特征提取模型。

[0005] 根据本公开的另一方面,提供了一种视频分类模型的训练方法,包括:获取训练样本;其中,所述训练样本包括样本视频和所述样本视频的分类标签;将所述样本视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述样本视频的分类结果;其中,所述视频分类模型包括依次连接的图像编码器和全连接层,所述图像编码器为利用上述特征提取模型的训练方法训练得到的特征提取模型中的图像编码器;根据所述分类标签和所述分类结果计算损失,并根据所述损失调整所述视频分类模型中全连接层的参数,直至满足收敛条件;输出训练好的视频分类模型。

[0006] 根据本公开的另一方面,提供了一种视频分类方法,包括:获取待分类视频;将所述待分类视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述待分类视频的分类结果;其中,所述视频分类模型根据视频分类模型的训练方法训练得到。

[0007] 根据本公开的另一方面,提供了一种特征提取模型的训练装置,包括:第一训练模块,用于基于原始数据集对特征提取模型进行无监督训练;其中,所述原始数据集包括若干第一样本视频帧以及所述第一样本视频帧对应的文本说明,所述特征提取模型包括图像编

码器和文本编码器,所述图像编码器用于对所述第一样本视频帧进行编码,得到对应的图像特征,所述文本编码器用于对所述文本说明进行编码,得到对应的文本特征;第二训练模块,用于基于标注数据集对训练得到的特征提取模型进行有监督训练;其中,所述标注数据集包括第二样本视频帧、针对所述第二样本视频帧标注的图像特征、所述第二样本视频帧对应的文本说明以及针对所述文本说明标注的文本特征;数据筛选模块,用于将所述原始数据集输入有监督训练得到的特征提取模型中,并根据图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度从所述原始数据集中筛选目标数据集;第三训练模块,用于基于所述目标数据集对有监督训练得到的特征提取模型再次进行无监督训练,得到训练好的特征提取模型。

[0008] 根据本公开的另一方面,提供了一种视频分类模型的训练装置,包括:第一获取模块,用于获取训练样本;其中,所述训练样本包括样本视频和所述样本视频的分类标签;第一预测模块,用于将所述样本视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述样本视频的分类结果;其中,所述视频分类模型包括依次连接的图像编码器和全连接层,所述图像编码器为利用上述特征提取模型的训练装置得到的特征提取模型中的图像编码器;参数处理模块,用于根据所述分类标签和所述分类结果计算损失,并根据所述损失调整所述视频分类模型中全连接层的参数,直至满足收敛条件;模型输出模块,用于输出训练好的视频分类模型。

[0009] 根据本公开的另一方面,提供了一种视频分类装置,包括:第二获取模块,用于获取待分类视频;第二预测模块,用于将所述待分类视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述待分类视频的分类结果;其中,所述视频分类模型根据上述视频分类模型的训练装置训练得到。

[0010] 根据本公开的另一方面,提供了一种电子设备,包括:至少一个处理器;以及与所述至少一个处理器通信连接的存储器;其中,所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行上述特征提取模型的训练方法、上述视频分类模型的训练方法或者上述视频分类方法。

[0011] 根据本公开的另一方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行上述特征提取模型的训练方法、上述视频分类模型的训练方法或者上述视频分类方法。

[0012] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现上述特征提取模型的训练方法、上述视频分类模型的训练方法或者上述视频分类方法。

[0013] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

## 附图说明

[0014] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0015] 图1是根据本公开一实施例的特征提取模型的训练方法的流程图;

[0016] 图2是根据本公开一实施例的特征提取模型的训练方法的示意图;

[0017] 图3是根据本公开另一实施例的特征提取模型的训练方法的示意图;

- [0018] 图4是根据本公开一实施例的特征提取模型的训练装置示意图；
- [0019] 图5是根据本公开一实施例的视频分类模型的训练方法的流程图；
- [0020] 图6是根据本公开一实施例提供的视频分类模型的结构示意图；
- [0021] 图7是根据本公开一实施例的视频分类模型的训练装置示意图；
- [0022] 图8是根据本公开一实施例的视频分类方法的流程图；
- [0023] 图9是根据本公开一实施例的视频分类装置示意图；
- [0024] 图10是用来实现本公开实施例的方法的电子设备的框图。

### 具体实施方式

[0025] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0026] 需要说明的是,本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0027] 根据本公开实施例,提供了一种特征提取模型的训练方法的实施例,需要说明的是,在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行,并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0028] 图1是根据本公开实施例的特征提取模型的训练方法的流程图,如图1所示,该训练方法包括以下步骤S101至步骤S104:

[0029] 步骤S101、基于原始数据集对特征提取模型进行无监督训练。其中,所述原始数据集包括若干第一样本视频帧以及所述第一样本视频帧对应的文本说明,所述特征提取模型包括图像编码器和文本编码器,所述图像编码器用于对所述第一样本视频帧进行编码,得到对应的图像特征,所述文本编码器用于对所述文本说明进行编码,得到对应的文本特征。

[0030] 在具体实施中,所述第一样本视频帧对应的文本说明可以为针对所述第一样本视频帧所在视频的内容的说明文字,例如可以为该视频的标题。

[0031] 其中,可以通过对样本视频进行截取,从截取得到的视频帧序列中选取多个第一样本视频帧。假设原始数据集包括N个第一样本视频帧和对应的N个文本说明,其中,N为大于1的整数。在对特征提取模型进行无监督训练的过程中,将第一样本视频帧输入特征提取模型的图像编码器中,得到图像特征,以及将第一样本视频帧对应的文本说明输入特征提取模型的文本编码器中,得到文本特征,以最大化N个配对的图像特征和文本特征之间的相似度,同时最小化 $N^2-N$ 个不配对的图像特征和文本特征之间的相似度为目标训练特征提取模型。

[0032] 在实际应用中,所述图像特征也可以称为图像向量,所述文本特征也可以称为文本向量。在如图2所示的例子中,原始数据集包括3个第一样本视频帧和对应的3个文本说明,将3个第一样本视频帧输入图像编码器分别得到图像向量F1、F2和F3,以及将3个文本说明输入文本编码器分别得到文本向量A1、A2和A3,其中,图像向量F1和文本向量A1配对,图像向量F2和文本向量A2配对,图像向量F3和文本向量A3配对,通过计算交叉熵损失,以最大化F1和A1、F2和A2以及F3和A3之间的相似度,同时最小化A2和F1、A3和F1、A1和F2、A3和F2、A1和F3以及A2和F3的相似度为目标训练包括图像编码器和文本编码器的特征提取模型。

[0033] 步骤S102、基于标注数据集对训练得到的特征提取模型进行有监督训练。其中,所述标注数据集包括第二样本视频帧、针对所述第二样本视频帧标注的图像特征、所述第二样本视频帧对应的文本说明以及针对所述文本说明标注的文本特征。

[0034] 需要说明的是,原始数据集的数据量通常远大于标注数据集的数据量。本实施例中原始数据集的数据量可以达到千万级。

[0035] 在步骤S102中,将标注数据集输入步骤S101训练得到的特征提取模型中,具体地,将第二样本视频帧输入特征提取模型中的图像编码器,得到预测的图像特征,以及将第二样本视频帧对应的文本说明输入特征提取模型中的文本编码器,得到预测的文本特征,根据预测的图像特征和标注的图像特征以及预测的文本特征和标注的文本特征计算损失,并根据计算的损失调整特征提取模型的参数,直至满足收敛条件。

[0036] 步骤S103、将所述原始数据集输入有监督训练得到的特征提取模型中,并根据图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度从所述原始数据集中筛选目标数据集。

[0037] 其中,上述相似度可以为余弦相似度,还可以为欧几里得距离、曼哈顿距离等。

[0038] 将原始数据集输入步骤S102训练得到的特征提取模型,输出的图像特征和文本特征之间的相似度越高,对应样本数据的噪声越小,相似度越低,对应样本数据的噪声越大,通过相似度对原始数据集进行筛选可以有效缓解大规模原始数据集中的样本噪声,从而得到样本噪声更小的目标数据集。

[0039] 步骤S104、基于所述目标数据集对有监督训练得到的特征提取模型再次进行无监督训练,得到训练好的特征提取模型。

[0040] 基于样本噪声更小的目标数据集对步骤S102训练得到的特征提取模型进行再次训练,可以有效提升特征提取模型的学习效果。

[0041] 其中,上述步骤S101、步骤S102和步骤S104中停止训练的条件可以根据实际情况进行设置,例如每次迭代的损失下降不明显,或者迭代的损失达到预设损失等。

[0042] 如图3所示,本公开实施例提供的特征提取模型的训练方法,首先利用原始数据集优化特征提取模型,此优化过程为无监督训练,其次利用标注数据集优化特征提取模型,此优化过程为有监督训练,然后利用有监督训练得到的特征提取模型对原始数据集进行筛选,得到目标数据集,最后利用目标数据集对特征提取模型进行优化,此优化过程为无监督训练。

[0043] 需要说明的是,为了进一步提高特征提取模型的学习效果,还可以对目标数据集进行进一步筛选,也即利用标注数据集再次优化步骤S104训练得到的特征提取模型,利用优化后的特征提取模型对目标数据集进行筛选,得到样本噪声更小的数据集,直至数据集

的样本噪声符合要求,最后利用符合要求的数据集对优化的特征提取模型进行最终训练,得到训练好的特征提取模型。

[0044] 在可选的一种实施例中,上述步骤S103具体包括:若图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度大于预设值,则将所述原始数据集中与所述图像特征对应的第一样本视频帧以及与所述文本特征对应的文本说明确定为目标数据集。其中,所述预设值可以根据实际情况进行设置,例如可以设置为60%。

[0045] 本实施例中,根据相似度是否大于预设值从原始数据集中筛选目标数据集,因此,筛选得到的目标数据集的样本噪声与预设值直接相关,其中,预设值越大,目标数据集的样本噪声越小,预设值越小,目标数据集的样本噪声越大。

[0046] 在可选的另一种实施例中,上述步骤S103具体包括以下步骤S103a至步骤S103c:

[0047] 步骤S103a、对于图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度,按照从高到低进行排序。

[0048] 步骤S103b、选取排序靠前的预设数量的相似度对应的图像特征和文本特征。其中,所述预设数量可以根据实际情况进行设置。

[0049] 步骤S103c、将所述原始数据集中与所述图像特征对应的第一样本视频帧以及与所述文本特征对应的文本说明确定为目标数据集。

[0050] 本实施例中,根据排序靠前的预设数量的相似度从原始数据集中筛选目标数据集,因此,筛选得到的目标数据集的样本噪声与预设数量相对相关,其中,预设数量越多,目标数据集的样本噪声相对越大,预设数量越小,目标数据集的样本噪声相对越小。

[0051] 根据本公开实施例,还提供了一种特征提取模型的训练装置实施例,其中,图4是根据本公开实施例的特征提取模型的训练装置示意图,该训练装置包括第一训练模块401、第二训练模块402、数据筛选模块403以及第三训练模块404。第一训练模块401用于基于原始数据集对特征提取模型进行无监督训练;其中,所述原始数据集包括若干第一样本视频帧以及所述第一样本视频帧对应的文本说明,所述特征提取模型包括图像编码器和文本编码器,所述图像编码器用于对所述第一样本视频帧进行编码,得到对应的图像特征,所述文本编码器用于对所述文本说明进行编码,得到对应的文本特征。第二训练模块402用于基于标注数据集对训练得到的特征提取模型进行有监督训练;其中,所述标注数据集包括第二样本视频帧、针对所述第二样本视频帧标注的图像特征、所述第二样本视频帧对应的文本说明以及针对所述文本说明标注的文本特征。数据筛选模块403用于将所述原始数据集输入有监督训练得到的特征提取模型中,并根据图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度从所述原始数据集中筛选目标数据集。第三训练模块404用于基于所述目标数据集对有监督训练得到的特征提取模型再次进行无监督训练,得到训练好的特征提取模型。

[0052] 需要说明的是,上述第一训练模块401、第二训练模块402、数据筛选模块403以及第三训练模块404对应于上述实施例中的步骤S101至步骤S104,四个模块与对应的步骤所实现的示例和应用场景相同,但不限于上述实施例所公开的内容。

[0053] 在可选的一种实施例中,所述数据筛选模块具体用于在图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度大于预设值的情况下,将所述原始数据集中与所述图像特征对应的第一样本视频帧以及与所述文本特征对应的文本说明确定为目标

数据集。

[0054] 在可选的另一种实施例中,所述数据筛选模块具体用于对于图像编码器输出的图像特征和文本编码器输出的文本特征之间的相似度,按照从高到低进行排序,并选取排序靠前的预设数量的相似度对应的图像特征和文本特征,以及将所述原始数据集中与所述图像特征对应的第一样本视频帧以及与所述文本特征对应的文本说明确定为目标数据集。

[0055] 根据本公开实施例,还提供了一种视频分类模型的训练方法的实施例,需要说明的是,在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行,并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0056] 图5是根据本公开实施例的视频分类模型的训练方法的流程图,如图5所示,该训练方法包括以下步骤S501至步骤S504:

[0057] 步骤S501、获取训练样本。其中,所述训练样本包括样本视频和所述样本视频的分类标签。

[0058] 步骤S502、将所述样本视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述样本视频的分类结果。其中,所述视频分类模型包括依次连接的图像编码器和全连接层,所述图像编码器为利用上述特征提取模型的训练方法训练得到的特征提取模型中的图像编码器。

[0059] 其中,利用上述特征提取模型的训练方法训练得到的特征提取模型构建视频分类模型,具体地,利用特征提取模型中的图像编码器构建视频分类模型,在图像编码器的后面连接全连接层,从而得到视频分类模型。

[0060] 步骤S503、根据所述分类标签和所述分类结果计算损失,并根据所述损失调整所述视频分类模型中全连接层的参数,直至满足收敛条件。

[0061] 步骤S504、输出训练好的视频分类模型。

[0062] 本公开实施例中,在视频分类模型的训练过程中,仅调整全连接层的参数,不调整图像编码器的参数。由于在上述特征提取模型的训练过程中,图像编码器已经学习了大量视频帧的图像特征,因此,在视频分类模型的训练过程中直接利用图像编码器的参数,不仅可以很好地理解视频帧的图像特征,还可以有效降低视频分类模型的训练成本。

[0063] 为了提高视频分类模型的学习效果,在可选的一种实施例中,所述视频分类模型还包括卷积层,如图6所示,设于所述图像编码器和所述全连接层之间。本实施例中,上述步骤S503具体包括:根据所述损失调整所述视频分类模型中卷积层和全连接层的参数。

[0064] 在具体实施中,所述卷积层的数量可以为一个,也可以为多个。在视频分类模型的训练过程中,仅调整卷积层和全连接层的参数,不调整图像编码器的参数。

[0065] 根据本公开实施例,还提供了一种视频分类模型的训练装置实施例,其中,图7是根据本公开实施例的视频分类模型的训练装置示意图,该训练装置包括第一获取模块701、第一预测模块702、参数处理模块703以及模型输出模块704。第一获取模块701用于获取训练样本;其中,所述训练样本包括样本视频和所述样本视频的分类标签。第一预测模块702用于将所述样本视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述样本视频的分类结果;其中,所述视频分类模型包括依次连接的图像编码器和全连接层,所述图像编码器为利用上述特征提取模型的训练装置得到的特征提取模型中的图像编码器。参

数处理模块703用于根据所述分类标签和所述分类结果计算损失,并根据所述损失调整所述视频分类模型中全连接层的参数,直至满足收敛条件。模型输出模块704用于输出训练好的视频分类模型。

[0066] 需要说明的是,第一获取模块701、第一预测模块702、参数处理模块703以及模型输出模块704对应于上述实施例中的步骤S501至步骤S504,四个模块与对应的步骤所实现的示例和应用场景相同,但不限于上述实施例所公开的内容。

[0067] 在可选的一种实施例中,所述视频分类模型还包括卷积层,设于所述图像编码器和所述全连接层之间。本实施例中,所述参数处理模块具体用于根据所述损失调整所述视频分类模型中卷积层和全连接层的参数。

[0068] 根据本公开实施例,还提供了一种视频分类方法的实施例,需要说明的是,在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行,并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0069] 图8是根据本公开实施例的视频分类方法的流程图,如图8所示,该训练方法包括以下步骤S801至步骤S802:

[0070] 步骤S801、获取待分类视频。

[0071] 步骤S802、将所述待分类视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述待分类视频的分类结果。其中,所述视频分类模型根据上述视频分类模型的训练方法训练得到。

[0072] 在具体实施中,可以从待分类视频中截取若干视频帧,将截取的视频帧输入视频分类模型中进行分类预测,从而得到待分类视频的分类结果。其中,所述视频分类模型可以对待分类视频进行不同的分类,例如对待分类视频的内容进行分类,分类结果可以包括美食、影视、旅游等,还可以对待分类视频的质量进行分类,分类结果可以包括高质量、中质量、低质量等。

[0073] 本公开实施例中,由于上述视频分类模型的训练方法的训练效果得到了提升,因此训练得到的视频分类模型的学习效果也得到了提升,即利用视频分类模型进行视频分类的分类结果更加准确。

[0074] 根据本公开实施例,还提供了一种视频分类装置实施例,其中,图9是根据本公开实施例的视频分类装置示意图,该视频分类装置包括第二获取模块901和第二预测模块902。第二获取模块901用于获取待分类视频。第二预测模块902用于将所述待分类视频中的至少一个视频帧输入视频分类模型中进行分类预测,得到所述待分类视频的分类结果;其中,所述视频分类模型根据上述视频分类模型的训练装置训练得到。

[0075] 需要说明的是,第二获取模块901和第二预测模块902对应于上述实施例中的步骤S801至步骤S802,两个模块与对应的步骤所实现的示例和应用场景相同,但不限于上述实施例所公开的内容。

[0076] 本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。

[0077] 根据本公开实施例,还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0078] 图10示出了可以用来实施本公开的实施例的示例电子设备1000的示意性框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0079] 如图10所示,设备1000包括计算单元1001,其可以根据存储在只读存储器 (ROM) 1002中的计算机程序或者从存储单元1008加载到随机访问存储器 (RAM) 1003中的计算机程序,来执行各种适当的动作和处理。在RAM 1003中,还可存储设备1000操作所需的各种程序和数据。计算单元1001、ROM 1002以及RAM 1003通过总线1004彼此相连。输入/输出 (I/O) 接口1005也连接至总线1004。

[0080] 设备1000中的多个部件连接至I/O接口1005,包括:输入单元1006,例如键盘、鼠标等;输出单元1007,例如各种类型的显示器、扬声器等;存储单元1008,例如磁盘、光盘等;以及通信单元1009,例如网卡、调制解调器、无线通信收发机等。通信单元1009允许设备1000通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0081] 计算单元1001可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元1001的一些示例包括但不限于中央处理单元 (CPU)、图形处理单元 (GPU)、各种专用的人工智能 (AI) 计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器 (DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元1001执行上文所描述的各个方法和处理,例如特征提取模型的训练方法、视频分类模型的训练方法或者视频分类方法。例如,在一些实施例中,上述特征提取模型的训练方法、视频分类模型的训练方法或者视频分类方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元1008。在一些实施例中,计算机程序的部分或者全部可以经由ROM 1002和/或通信单元1009而被载入和/或安装到设备1000上。当计算机程序加载到RAM 1003并由计算单元1001执行时,可以执行上文描述的特征提取模型的训练方法、视频分类模型的训练方法或者视频分类方法的一个或多个步骤。备选地,在其他实施例中,计算单元1001可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行上述特征提取模型的训练方法、视频分类模型的训练方法或者视频分类方法。

[0082] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列 (FPGA)、专用集成电路 (ASIC)、专用标准产品 (ASSP)、芯片上系统的系统 (SOC)、复杂可编程逻辑设备 (CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0083] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的

功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0084] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0085] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0086] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)和互联网。

[0087] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务器可以是云服务器,也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0088] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0089] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

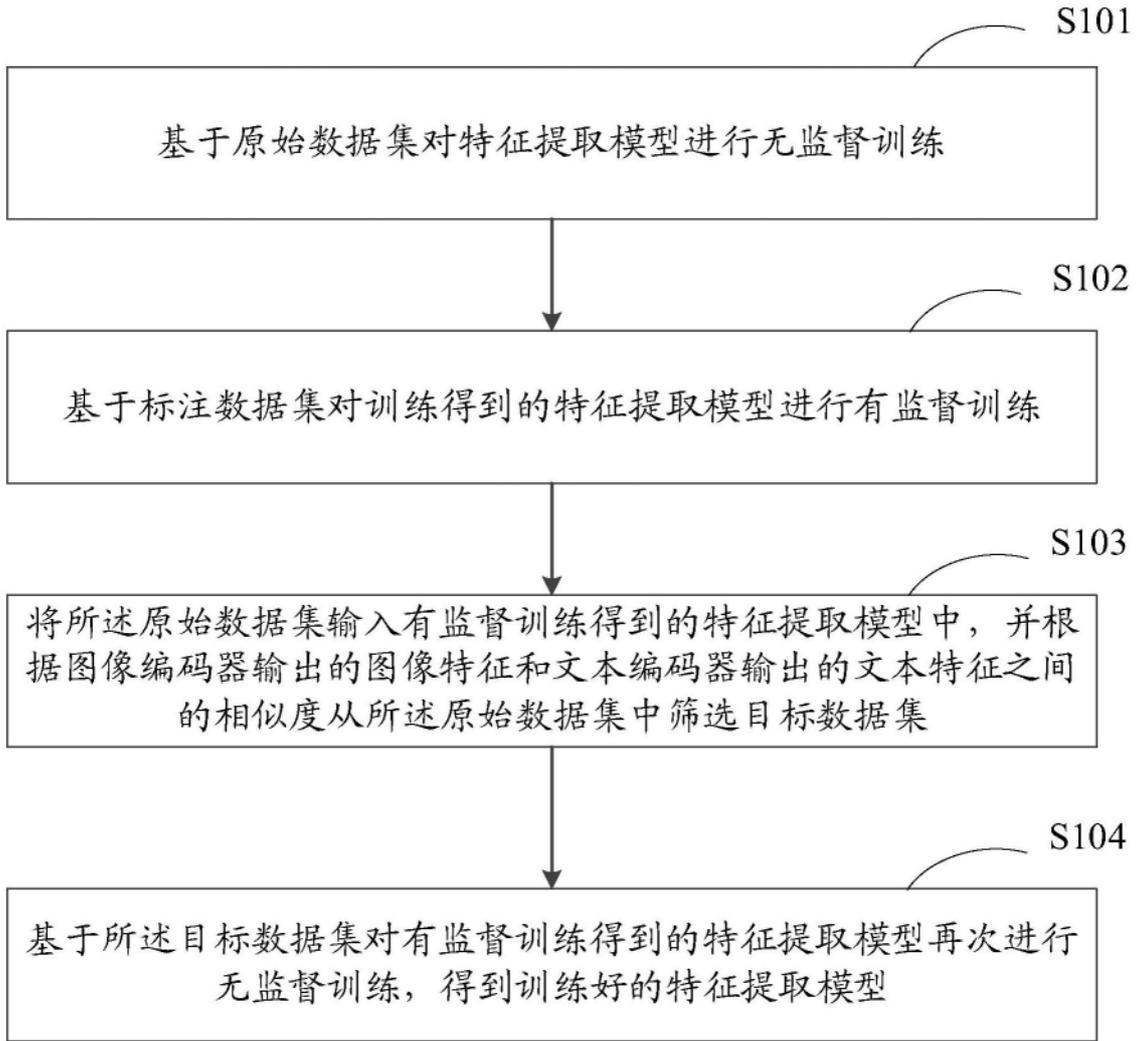


图1

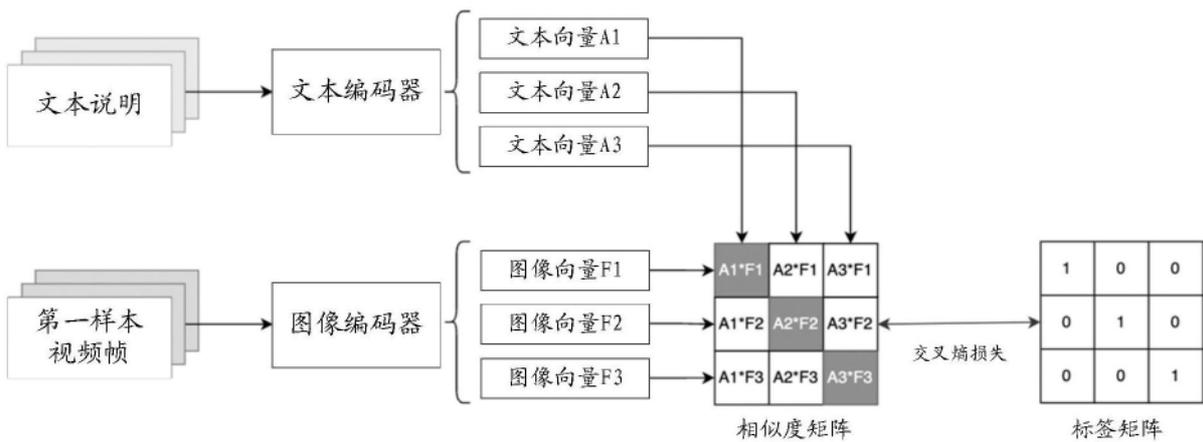


图2

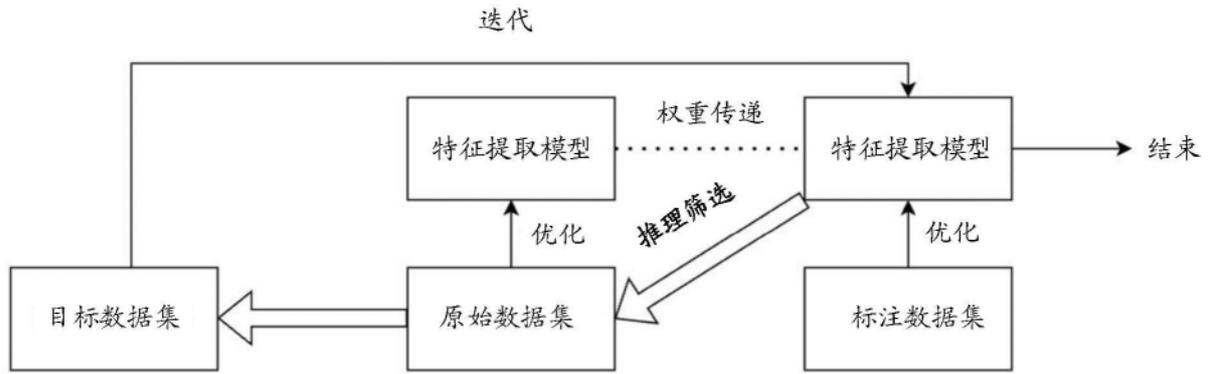


图3

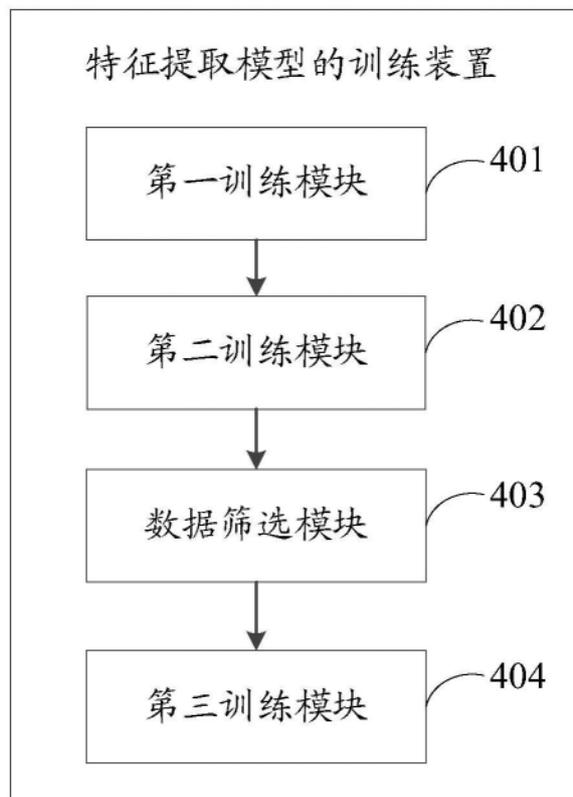


图4

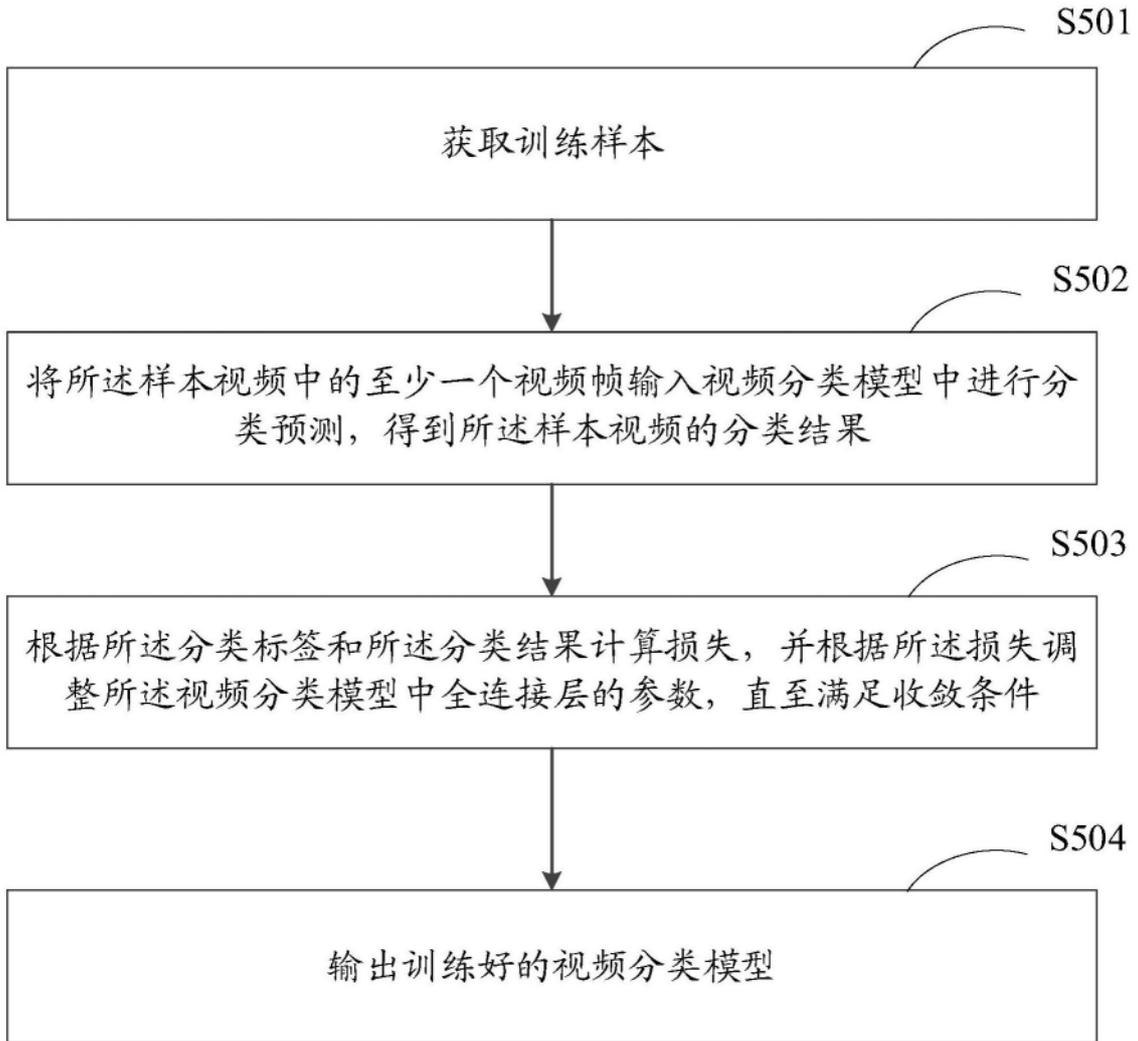


图5

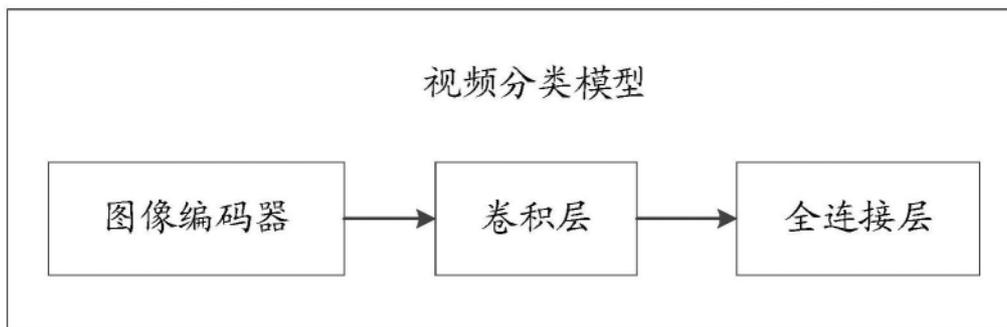


图6

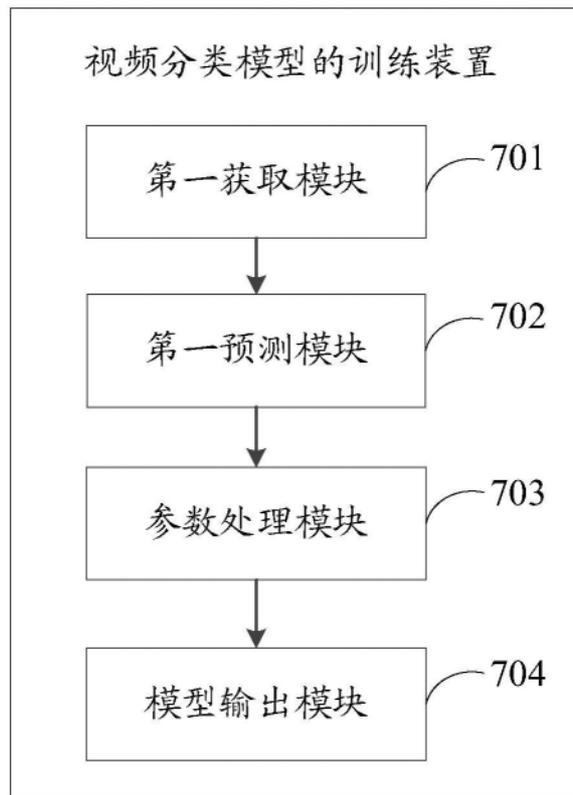


图7

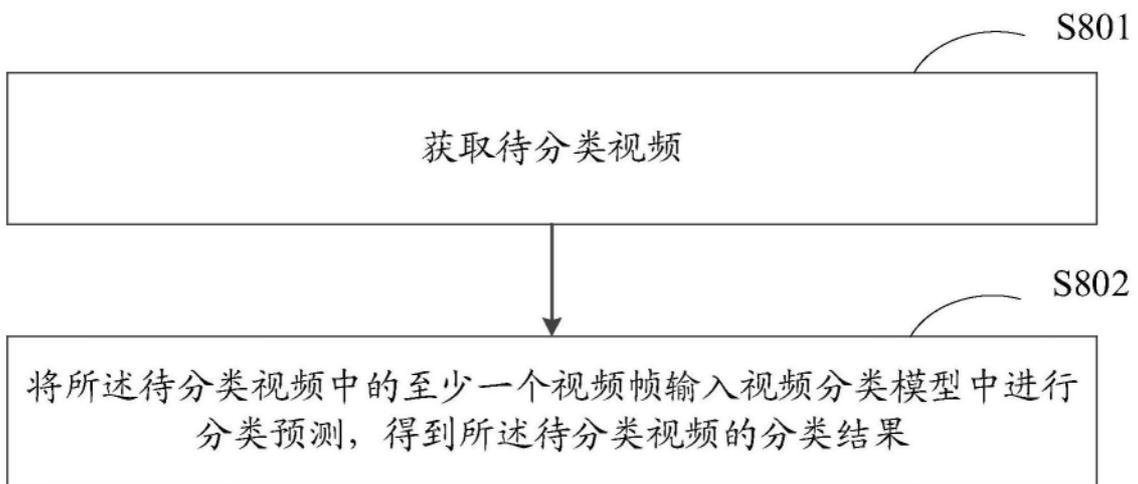


图8

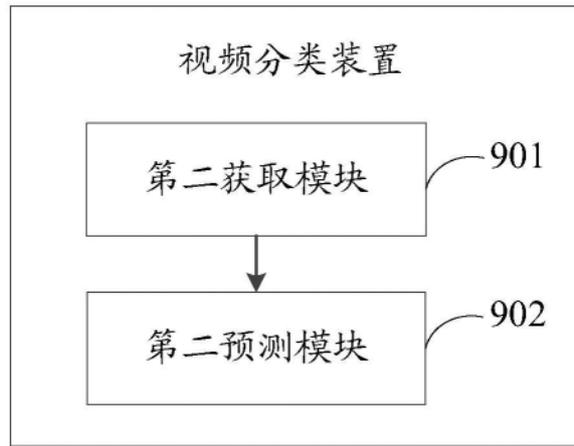


图9

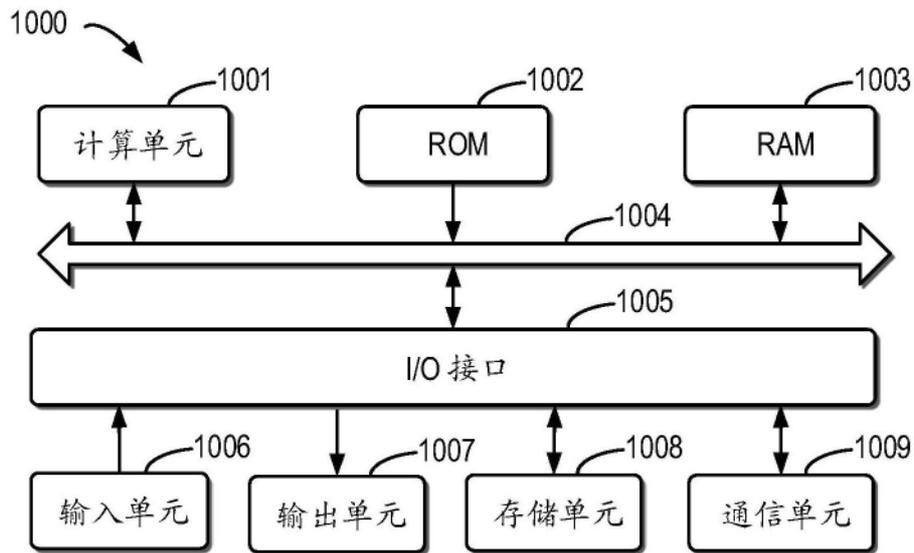


图10