



(12) 发明专利申请

(10) 申请公布号 CN 114513653 A

(43) 申请公布日 2022. 05. 17

(21) 申请号 202210073911.1

(22) 申请日 2022.01.21

(71) 申请人 腾讯科技(深圳)有限公司  
地址 518057 广东省深圳市南山区高新区  
科技中一路腾讯大厦35层

(72) 发明人 常勤伟 杨天舒

(74) 专利代理机构 广州三环专利商标代理有限公司 44202  
专利代理师 杜维

(51) Int. Cl.

H04N 17/00 (2006.01)

H04N 21/234 (2011.01)

H04N 21/44 (2011.01)

G06K 9/62 (2022.01)

G06V 10/80 (2022.01)

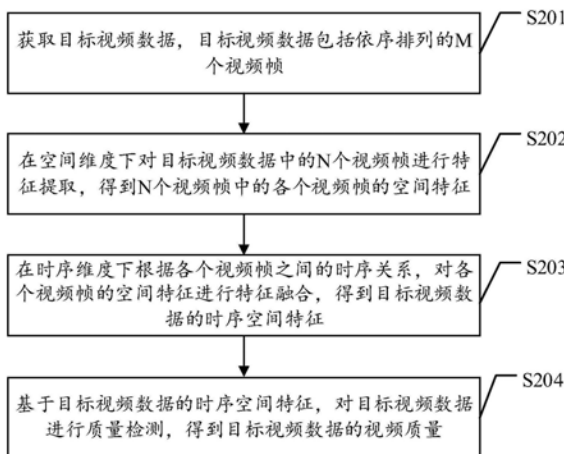
权利要求书2页 说明书16页 附图6页

(54) 发明名称

视频处理方法、装置、设备、计算机程序产品及存储介质

(57) 摘要

本申请实施例公开了一种视频处理方法、装置、设备、计算机程序产品及存储介质,其中视频处理方法包括:获取目标视频数据,所述目标视频数据包括依序排列的M个视频帧,M为正整数;在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视频帧中的各个视频帧的空间特征,N为正整数且N小于或等于M;在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的空间特征进行特征融合,得到所述目标视频数据的时序空间特征;基于所述目标视频数据的时序空间特征,对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量。本申请实施例可提升质量检测结果的准确性。



1. 一种视频处理方法,其特征在于,包括:

获取目标视频数据,所述目标视频数据包括依序排列的M个视频帧,M为正整数;

在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视频帧中的各个视频帧的空间特征,N为正整数且N小于或等于M;

在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的空间特征进行特征融合,得到所述目标视频数据的时序空间特征;

基于所述目标视频数据的时序空间特征,对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量。

2. 根据权利要求1所述的方法,其特征在于,任一视频帧的空间特征包括多个特征元素,所述在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的空间特征进行特征融合,得到所述目标视频数据的时序空间特征,包括:

根据所述各个视频帧的空间特征包括的特征元素的数量,分别确定每个视频帧对应的特征元素重组尺寸;

按照所述每个视频帧对应的特征元素重组尺寸,分别对所述每个视频帧的空间特征中的特征元素进行元素重组,得到所述每个视频帧的重组空间特征;

在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的重组空间特征进行特征融合,得到所述目标视频数据的时序空间特征。

3. 根据权利要求2所述的方法,其特征在于,所述在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的重组空间特征进行特征融合,得到所述目标视频数据的时序空间特征,包括:

按照预设的滑动步长采用滑动窗口在所述N个视频帧中进行滑动;

在每次滑动所述滑动窗口后,将所述N个视频帧中当前位于所述滑动窗口内的各个视频帧均作为目标视频帧;

在时序维度下根据各个目标视频帧之间的时序关系,对所述各个目标视频帧的重组空间特征进行三维卷积处理,得到一个子时序空间特征;

在所述滑动窗口结束滑动后,采用得到的各个子时序空间特征,构建所述目标视频数据的时序空间特征。

4. 根据权利要求2所述的方法,其特征在于,所述在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的重组空间特征进行特征融合,得到所述目标视频数据的时序空间特征,包括:

遍历所述各个视频帧,并将当前遍历的当前视频帧的重组空间特征确定为当前重组空间特征;

在时序维度下根据所述各个视频帧之间的时序关系,确定所述当前视频帧的相邻视频帧;

基于所述相邻视频帧的重组空间特征,对所述当前重构空间特征进行特征提取,得到所述当前视频帧对应的子时序空间特征;

在所述各个视频帧均被遍历后,采用所述各个视频帧对应的子时序空间特征,构建所述目标视频数据的时序空间特征。

5. 根据权利要求1-4中任一项所述的方法,其特征在于,所述目标视频数据的时序空间

特征包括多个子时序空间特征,所述基于所述目标视频数据的时序空间特征,对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量,包括:

对所述目标视频数据的时序空间特征中的各个子时序空间特征进行全局池化处理,基于全局池化处理结果确定所述目标视频数据的视频特征向量;

采用所述视频特征向量对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量。

6. 根据权利要求1-4中任一项所述的方法,其特征在于,所述在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视频帧中的各个视频帧的空间特征,包括:

针对所述目标视频数据中的N个视频帧中的任一视频帧,在空间维度下对所述任一视频帧进行特征提取,得到一个或多个目标特征图;

对各个目标特征图进行融合处理,得到所述任一视频帧的空间特征。

7. 根据权利要求6所述的方法,其特征在于,所述对各个目标特征图进行融合处理,得到所述任一视频帧的空间特征,包括:

对各个目标特征图进行全局池化处理,得到所述各个目标特征图的均值和标准差;

对所述各个目标特征图的均值和标准差进行拼接处理,得到所述任一视频帧的空间特征。

8. 一种视频处理装置,其特征在于,包括:

获取单元,用于获取目标视频数据,所述目标视频数据包括依序排列的M个视频帧,M为正整数;

处理单元,用于在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视频帧中的各个视频帧的空间特征,N为正整数且N小于或等于M;

所述处理单元,还用于在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的空间特征进行特征融合,得到所述目标视频数据的时序空间特征;

所述处理单元,还用于基于所述目标视频数据的时序空间特征,对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量。

9. 一种计算机设备,其特征在于,包括处理器和存储器,其中,所述存储器用于存储计算机程序,所述计算机程序被所述处理器执行时,实现如权利要求1-7中任一项所述的方法。

10. 一种计算机存储介质,其特征在于,所述计算机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1-7中任一项所述的方法。

11. 一种计算机程序产品,包括计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1-7中任一项所述的方法。

## 视频处理方法、装置、设备、计算机程序产品及存储介质

### 技术领域

[0001] 本申请涉及计算机技术领域,尤其涉及一种视频处理方法、装置、设备、计算机程序产品及存储介质。

### 背景技术

[0002] 目前,普遍存在对视频数据进行质量检测的需求,如对视频数据的清晰度进行评估;在此种情况下,通常按照一定规则对视频数据进行抽帧解码,得到若干视频帧,一个视频帧为一个图像数据,并分别获取各个视频帧的质量,从而将各个视频帧的质量之间的均值作为视频数据的视频质量。但在获取各个视频帧的过程中,容易出现拖影、运行模糊等情况,导致得到的视频质量的精确度较低。基于此,如何提升质量检测结果的准确性成为一个研究热点。

### 发明内容

[0003] 本发明实施例提供了一种视频处理方法、装置、设备、计算机程序产品及存储介质,可以更加精确地获取目标视频数据的视频质量,提升质量检测结果的准确性。

[0004] 一方面,本申请实施例提供了一种视频处理方法,所述方法包括:

[0005] 获取目标视频数据,所述目标视频数据包括依序排列的M个视频帧,M为正整数;

[0006] 在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视频帧中的各个视频帧的空间特征,N为正整数且N小于或等于M;

[0007] 在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的空间特征进行特征融合,得到所述目标视频数据的时序空间特征;

[0008] 基于所述目标视频数据的时序空间特征,对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量。

[0009] 另一方面,本申请实施例提供了一种视频处理装置,所述装置包括:

[0010] 获取单元,用于获取目标视频数据,所述目标视频数据包括依序排列的M个视频帧,M为正整数;

[0011] 处理单元,用于在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视频帧中的各个视频帧的空间特征,N为正整数且N小于或等于M;

[0012] 所述处理单元,还用于在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的空间特征进行特征融合,得到所述目标视频数据的时序空间特征;

[0013] 所述处理单元,还用于基于所述目标视频数据的时序空间特征,对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量。

[0014] 再一方面,本申请实施例提供了一种计算机程序产品,所述计算机程序被处理器执行时实现如下步骤:

[0015] 获取目标视频数据,所述目标视频数据包括依序排列的M个视频帧,M为正整数;

[0016] 在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视

帧中的各个视频帧的空间特征,  $N$  为正整数且  $N$  小于或等于  $M$ ;

[0017] 在时序维度下根据所述各个视频帧之间的时序关系, 对所述各个视频帧的空间特征进行特征融合, 得到所述目标视频数据的时序空间特征;

[0018] 基于所述目标视频数据的时序空间特征, 对所述目标视频数据进行质量检测, 得到所述目标视频数据的视频质量。

[0019] 再一方面, 本申请实施例提供了一种计算机设备, 所述计算机设备包括处理器、存储器, 其中, 所述存储器用于存储计算机程序, 所述计算机程序被所述处理器执行时实现如下步骤:

[0020] 获取目标视频数据, 所述目标视频数据包括依序排列的  $M$  个视频帧,  $M$  为正整数;

[0021] 在空间维度下对所述目标视频数据中的  $N$  个视频帧进行特征提取, 得到所述  $N$  个视频帧中的各个视频帧的空间特征,  $N$  为正整数且  $N$  小于或等于  $M$ ;

[0022] 在时序维度下根据所述各个视频帧之间的时序关系, 对所述各个视频帧的空间特征进行特征融合, 得到所述目标视频数据的时序空间特征;

[0023] 基于所述目标视频数据的时序空间特征, 对所述目标视频数据进行质量检测, 得到所述目标视频数据的视频质量。

[0024] 再一方面, 本申请实施例提供了一种计算机存储介质, 所述计算机存储介质存储有计算机程序, 所述计算机程序适于由处理器加载并执行如下步骤:

[0025] 获取目标视频数据, 所述目标视频数据包括依序排列的  $M$  个视频帧,  $M$  为正整数;

[0026] 在空间维度下对所述目标视频数据中的  $N$  个视频帧进行特征提取, 得到所述  $N$  个视频帧中的各个视频帧的空间特征,  $N$  为正整数且  $N$  小于或等于  $M$ ;

[0027] 在时序维度下根据所述各个视频帧之间的时序关系, 对所述各个视频帧的空间特征进行特征融合, 得到所述目标视频数据的时序空间特征;

[0028] 基于所述目标视频数据的时序空间特征, 对所述目标视频数据进行质量检测, 得到所述目标视频数据的视频质量。

[0029] 本申请实施例在获取目标视频数据后, 可先在空间维度下对目标视频数据中的  $N$  个视频帧进行特征提取, 得到  $N$  个视频帧中的各个视频帧的空间特征; 然后, 在时序维度下根据各个视频帧之间的时序关系, 对各个视频帧的空间特征进行特征融合, 得到目标视频数据的时序空间特征, 使得该时序空间特征不仅可指示各个视频帧的空间信息, 还可指示各个视频帧之间的时序信息, 从而使得在基于该时序空间特征对目标视频数据进行质量检测时, 可全局地利用目标视频数据的整体信息, 有效减少单个视频帧中因对象运动或场景变换而产生的拖影情况对质量检测的影响, 提升质量检测结果的准确性。

## 附图说明

[0030] 为了更清楚地说明本申请实施例技术方案, 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。

[0031] 图1是本申请实施例提供的一种视频处理方案的流程示意图;

[0032] 图2是本申请实施例提供的一种视频处理方法的流程示意图;

[0033] 图3是本申请实施例提供的一种卷积块的示意图;

- [0034] 图4是本申请实施例提供的另一种视频处理方法的流程示意图；
- [0035] 图5是本申请实施例提供的一种特征提取模块的结构示意图；
- [0036] 图6是本申请实施例提供的一种特征融合模块的结构示意图；
- [0037] 图7是本申请实施例提供的一种视频处理装置的结构示意图；
- [0038] 图8是本申请实施例提供的一种计算机设备的结构示意图。

### 具体实施方式

[0039] 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。

[0040] 随着互联网技术的不断发展,人工智能(Artificial Intelligence, AI)技术也随之得到更好的发展。所谓的人工智能技术是指利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术;其主要通过了解智能的实质,生产出一种新的能以人类智能相似的方式做出反应的智能机器,使得智能机器具有感知、推理与决策等多种功能。相应的, AI技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术(Computer Vision, CV)、语音处理技术、自然语言处理技术以及机器学习(Machine Learning, ML)/深度学习等几大方向。

[0041] 其中,机器学习是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机设备具有智能的根本途径;其专门研究计算机设备怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。而深度学习则是一种利用深层神经网络系统,进行机器学习的技术;机器学习/深度学习通常可包括人工神经网络、强化学习(Reinforcement Learning, RL)、有监督学习、无监督学习等多种技术;所谓的有监督学习是指采用类别已知(具有标注类别)的训练样本进行模型训练的处理方式,无监督学习是指采用类别未知(没有被标记)的训练样本进行模型训练的处理方式。

[0042] 基于AI技术中的机器学习/深度学习技术,本发明实施例提出了一种视频处理方案,以实现提升对目标视频数据进行质量检测的准确性,并获取精确度较高的视频质量。在具体实现中,该视频处理方案可由一个计算机设备执行,该计算机设备可以是终端或者服务器;其中,此处所提及的终端可以包括但不限于:智能手机、平板电脑、笔记本电脑、台式计算机、智能手表、智能电视、智能车载终端等;终端内可运行各式各样的客户端(application, APP),如视频播放客户端、社交客户端、浏览器客户端、信息流客户端、教育客户端,等等。此处所提及的服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN(Content Delivery Network, 内容分发网络)、以及大数据和人工智能平台等基础云计算服务的云服务器,等等。

[0043] 需要说明的是,在其他实施例中,该模型训练方案也可由服务器和终端共同执行;服务器和终端通过网络(如有线网络、无线网络等)连接,对此不作限定。并且,本发明实施例所提及的计算机设备可以位于区块链网络外,也可以位于区块链网络内,对此不作限定;所谓的区块链网络是一种由点对点网络(P2P网络)和区块链所构成的网络,而区块链则是指一种分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式,其本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块(或称为区块)。

[0044] 参见图1所示,本申请实施例所提出的视频处理方案的大致原理如下:

[0045] ①获取目标视频数据的视频链接。可选的,计算机设备在检测到用户的质量检测操作时,可将该质量检测操作所指示的链接作为目标视频数据的视频链接;也可将最近上传的视频数据的视频链接作为目标视频数据的视频链接,本申请对此不作限定。其中,最近上传的视频数据可以指的是上传时间与当前系统时间之间的差值小于或等于预设时长的视频数据;该预设时长可以是计算机设备设置的,也可以指用户按照意愿设置的,如1分钟(min)或30秒(s)等。或者,计算机设备还可将任一视频发布平台所需发布的视频数据的视频链接作为目标视频数据的视频链接,或者已经发布的视频数据的视频链接作为目标视频数据的视频链接。

[0046] ②采用目标视频数据的视频链接,下载目标视频数据;其中,目标视频数据包括M个视频帧,M为正整数。需要说明的是,在本申请的具体实施方式中,涉及用户执行的质量检测操作所指示的链接对应的视频数据等,当本申请实施例运用到具体产品或技术中时,与用户相关的数据(如用户输入的链接对应的视频数据、用户上传的视频数据等)的获取需要获得用户许可或者同意,且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。具体的,计算机设备在用户执行质量检测操作或上传视频数据时,可以显示提示界面或提示弹窗,该提示界面或提示弹窗用于提示用户当前正在获取目标视频数据,且该提示界面或提示弹窗中包括确认选项;若用户对上述提示界面或提示弹窗中的确认选项进行了选择操作,则计算机设备可开始执行获取目标视频数据的相关步骤,否则结束。

[0047] ③在空间维度下对目标视频数据中的N个视频帧进行特征提取,得到N个视频帧中各个视频帧的空间特征,N为正整数且N小于或等于M。

[0048] ④对各个视频帧的空间特征进行时序维度的特征融合,并基于特征融合结果获取目标视频数据的视频质量。可以理解的是,目标视频数据的视频质量可以是一个数值,也可以是一个评价指标,如优或良等,本申请对此不作限定。

[0049] ⑤输出结果。需要说明的是,若目标视频数据的视频质量为一个数值时,计算机设备可以直接输出目标视频数据的视频质量,也可以根据该视频质量确定目标视频数据的评价指标,并输出目标视频数据的评价指标。具体的,在根据该视频质量确定目标视频数据的评价指标时,若该视频质量大于第一预设评价阈值,则计算机设备可以确定目标视频数据的评价指标为优;若该视频质量小于或等于第一预设评价阈值且大于第二预设评价阈值,则计算机设备可以确定目标视频数据的评价指标为良;若该视频质量小于或等于第二预设评价阈值,则计算机设备可以确定目标视频数据的评价指标为差,等等;其中,第一预设评价阈值大于第二预设评价阈值。可选的,第一预设评价阈值和第二预设评价阈值可以是计

算机设备设置的,也可以是用户按照意愿设置的,本申请对此不作限定。

[0050] 值得注意的是,上述只是示例性的阐述了视频处理方案的大致原理,并不对此进行限定。例如,在上述流程中,计算机设备在通过步骤①获取目标视频数据的视频链接后,可通过步骤②下载目标视频数据;但在其他实施例中,计算机设备可以直接获取目标视频数据,并执行步骤③~⑤以输出相应的结果。又如,若目标视频数据是即将发布的视频数据,则计算机设备在通过步骤④确定目标视频数据的视频质量后,也可以不输出结果,而是直接根据目标视频数据的视频质量,判断是否需对该目标视频数据执行取消发布的操作。具体的,若目标视频的视频质量不满足发布要求,则可取消发布该目标视频数据。

[0051] 经实践表明,本申请实施例提出的视频处理方案可至少具有如下几点有益效果:

[0052] 第一点:可在时序维度下对各个视频帧的空间特征进行特征融合,得到目标视频数据的时序空间特征,并基于该时序空间特征实现对目标视频数据整体质量的自动化评分流程;第二点:利用目标视频数据的整体信息,即时序空间特征,对目标视频数据进行质量检测,有效减少单个视频帧中由于对象运动或场景变换造成的拖影的影响,从而提升质量检测的准确性,以获取精确度较高的视频质量;第三点:本申请实施例所提出的视频处理方案可支持同时对多个目标视频数据进行质量检测,这样可实现快速地确定各个目标视频数据的视频质量,效率高,极大的提高了单位时间内审核视频的数量;第四点:可提供一种评价标准,并按照统一的评价标准对各个目标视频数据进行质量检测,以实现对各个目标视频数据进行质量检测的公平性;第五点:可输出结果,以便于用户直观地掌握目标视频数据的质量检测结果。

[0053] 进一步的,本申请实施例可应用于云技术、人工智能、智慧交通、驾驶等各种应用场景,如质量审核场景、质量自检场景,等等。

[0054] 例如,当上述视频处理方案应用于质量审核场景时,计算机设备可以对用户上传至各视频平台的每个视频数据进行质量检测,以实现每个视频数据的质量审核。需要说明的是,用户自制视频数据在各视频平台爆发式增长,但视频数据的视频质量直接影响用户的观看体验,故计算机设备对用户上传的视频数据进行质量审核具有重大意义。进一步的,计算机设备在对用户上传的视频数据进行质量审核的过程中,可向相应视频平台的管理者反馈未通过质量检测的视频数据,也可直接对未通过质量检测的视频数据进行进一步的处理,如下架未通过质量检测的视频数据等,本申请对此不作限定。

[0055] 又如,当上述视频处理方案应用于质量自检场景时,计算机设备可以对用户执行的质量检测操作所指示的视频数据进行质量检测,并输出用户执行的质量检测操作所指示的视频数据的视频质量,以便于用户直观地掌握自制视频数据的视频质量。

[0056] 基于上述视频处理方案的相关描述,本申请实施例提出一种视频处理方法,该视频处理方法可以由上述所提及计算机设备执行,该计算机设备可以是终端或者服务器;或者,该视频处理方法可由终端和服务器共同执行。为便于阐述,后续均以计算机设备执行该视频处理方法为例进行说明;请参见图2,该视频处理方法可包括以下步骤S201-S204:

[0057] S201,获取目标视频数据,目标视频数据包括依序排列的M个视频帧。

[0058] 其中,M为正整数;可以理解的是,一个视频帧可以指的是一个图像数据,也就是说,一个视频帧可以指的是一张图片。

[0059] 可选的,计算机设备可以获取目标视频数据的视频链接,并通过该视频链接对目



标视频数据进行下载,也可以从用户提供的移动存储器中获取目标视频数据,该移动存储器可以是移动硬盘或者USB闪存盘(USB flash disk)等。计算机设备也可以从计算机设备自身的存储空间中获取目标视频数据,例如当目标视频数据存在于计算机设备的只读存储器(Read Only Memory,ROM)时,计算机设备可以从该只读存储器中获取目标视频数据。本申请对目标视频数据的具体获取方式不作限定。

[0060] S202,在空间维度下对目标视频数据中的N个视频帧进行特征提取,得到N个视频帧中的各个视频帧的空间特征;其中,N为正整数且N小于或等于M。

[0061] 在一种实施方式中,计算机设备对目标视频数据进行抽帧处理,得到目标视频数据中的N个视频帧;也就是说,计算机设备可以从目标视频数据中的M个视频帧中抽取N个视频帧,进而在空间维度下对目标视频数据中的N个视频帧进行特征提取,得到N个视频帧中的各个视频帧的空间特征,此情况下的N小于M。通过对目标视频数据的抽帧处理,可减少需进行特征提取的视频帧,进而有效减小特征提取所需的运算资源,提升特征提取的效率。

[0062] 另一种实施方式中,计算机设备可以将目标视频数据中的每个视频帧均作为待进行特征提取的视频帧;也就是说,计算机设备可以将目标视频数据中的所有视频帧作为N个视频帧,进而在空间维度下对目标视频数据中的N个视频帧进行特征提取,得到N个视频帧中的各个视频帧的空间特征,此情况下的N等于M。通过对每个视频帧的特征提取,可全面地获取目标视频数据中每个视频帧的空间特征,进而全面地获取到目标视频数据的空间信息,可对目标视频数据的视频质量进行全面评估,从而有效提升质量检测结果的准确性;并且,可避免单个视频帧中因对象运动或场景变换而产生的拖影、运动模糊等情况对质量检测的影响,从而提升质量检测结果的准确性。

[0063] 可以理解的是,各个视频帧的空间特征可以用于指示各个视频帧的语义特征,也就是说,计算机设备在空间维度下对N个视频帧中的各个视频帧进行特征提取,可得到目标视频数据的语义内容,该目标视频数据的语义内容可对最终评分结果(即目标视频数据的视频质量)产生巨大影响;换句话说,各个视频帧的空间特征可对目标视频数据的视频质量产生巨大影响。

[0064] 需要说明的是,计算机设备可以采用卷积神经网络(Convolutional Neural Networks,CNN)对各个视频帧进行特征提取;所谓的卷积神经网络是一类包含卷积计算且具有深度结构的前馈神经网络,是深度学习的代表算法之一。可选的,此处的卷积神经网络可以为ResNet-50(一种残差网络),即计算机设备可以采用ResNet-50来对各个视频帧进行特征提取,ResNet-50可包括多个网络层,每个网络层包括至少一个block(或称为卷积块),一个block包括3个卷积层(conv)。例如,如图3所示,ResNet-50中的一个block可包括:卷积核为 $1 \times 1$ 且通道数为64的第一个卷积层,卷积核为 $3 \times 3$ 且通道数为64的第二个卷积层,以及卷积核为 $1 \times 1$ 且通道数为256的第三个卷积层。

[0065] 进一步的,计算机设备可以在ImageNet(一种大型通用物体识别的开源数据集)上对ResNet-50进行预训练,并采用经过预训练的ResNet-50来对各个视频帧进行特征提取。在此种情况下,在ImageNet上预训练过的ResNet-50具有对不同语义信息的判别能力,从而能够更好地提取语义特征,也就是说,计算机设备可以采用在ImageNet上预训练过的ResNet-50,更好地在空间维度上对各个视频帧进行特征提取,从而得到更加准确的各个空间特征。

[0066] 可选的,计算机设备还可以使用其他不同的结构网络来对各个视频帧进行特征提取,如ResNet-18(一种残差网络)、ResNet-101(一种残差网络)等,本申请对此不作限定。

[0067] S203,在时序维度下根据各个视频帧之间的时序关系,对各个视频帧的空间特征进行特征融合,得到目标视频数据的时序空间特征。

[0068] 可以理解的是,目标视频数据中每个视频帧都在变化,在时序维度上,若参考观测帧前后的一系列视频帧,则很容易判断出目标视频数据的视频质量变化情况,也就是说,时序维度的评估可以是反映目标视频数据是否具有较高视频质量的客观手段,以便于后续更加客观全面地对目标视频数据进行质量检测。

[0069] 在此种情况下,计算机设备在得到各个视频帧的空间特征后,可以采用能较好地融合帧间信息的三维卷积(3D conv)神经网络来捕获每一个视频帧前后相邻的视频帧,也就是说,计算机设备可以在时序维度下根据各个视频帧之间的时序关系,采用3D卷积神经网络对各个视频帧的空间特征进行特征融合,得到目标视频数据的时序空间特征。可选的,计算机设备还可以采用其他时序相关的神经网络对各个视频帧的空间特征进行特征融合,如循环神经网络(Recurrent Neural Network,RNN)、长短期记忆网络(Long Short-Term Memory,LSTM)等,本申请对此不作限定。

[0070] 需要说明的是,3D卷积神经网络可对连续的多帧图像(即多个视频帧)进行卷积操作,通过堆叠多个连续的视频帧组成一个立方体,然后在立方体中运行3D卷积核;在这个结构中,卷积层中每一个特征图(Feature map)都会与上一层中多个邻近的连续帧相连,因此可以捕捉运动信息。所谓的特征图是卷积神经网络卷积层的输出结果,一般是图像在同一层次不同基上的描述,通常包含原始图像的特征。

[0071] S204,基于目标视频数据的时序空间特征,对目标视频数据进行质量检测,得到目标视频数据的视频质量。

[0072] 需要说明的是,目标视频数据的视频质量可以指的是该目标视频数据的清晰度,也可以指的是该目标视频数据的流畅度,等等。

[0073] 进一步的,计算机设备基于目标视频数据的时序空间特征,对目标视频数据进行质量检测,得到目标视频数据的视频质量的具体实施过程可以包括:对目标视频数据的时序空间特征进行全局池化(Global Pooling,GP)处理,基于全局池化处理结果确定目标视频数据的视频特征向量;采用视频特征向量对目标视频数据进行质量检测,得到目标视频数据的视频质量。

[0074] 本申请实施例在获取目标视频数据后,可先在空间维度下对目标视频数据中的N个视频帧进行特征提取,得到N个视频帧中的各个视频帧的空间特征;然后,在时序维度下根据各个视频帧之间的时序关系,对各个视频帧的空间特征进行特征融合,得到目标视频数据的时序空间特征,使得该时序空间特征不仅可指示各个视频帧的空间信息,还可指示各个视频帧之间的时序信息,从而使得在基于该时序空间特征对目标视频数据进行质量检测时,可全局地利用目标视频数据的整体信息,有效减少单个视频帧中因对象运动或场景变换而产生的拖影情况对质量检测的影响,提升质量检测结果的准确性。

[0075] 请参见图4,是本申请实施例提供的另一种视频处理方法的流程示意图。该视频处理方法可以由上述所提及的计算机设备执行,该计算机设备可以是终端或者服务器;或者,该视频处理方法可由终端和服务器共同执行。为便于阐述,后续均以计算机设备执行该视

频处理方法为例进行说明;请参见图4,该视频处理方法可包括以下步骤S401-S406:

[0076] S401,获取目标视频数据,目标视频数据包括依序排列的M个视频帧。

[0077] S402,在空间维度下对目标视频数据中的N个视频帧进行特征提取,得到N个视频帧中的各个视频帧的空间特征;其中,N为正整数且N小于或等于M。

[0078] 需要说明的是,任一视频帧的空间特征包括多个特征元素。

[0079] 具体的,计算机设备可以针对目标视频数据中的N个视频帧中的任一视频帧,在空间维度下对该任一视频帧进行特征提取,得到一个或多个目标特征图;并对各个目标特征图进行融合处理,得到该任一视频帧的空间特征。换句话说,计算机设备在空间维度下对该任一视频帧进行特征提取,可以得到该任一视频帧对应的一组目标特征图,该任一视频帧对应的一组目标特征图包括一个或多个目标特征图。

[0080] 可以理解的是,若计算机设备可以采用卷积神经网络对上述任一视频帧进行特征提取,那么上述任一视频帧对应的一组目标特征图包括的目标特征图的数量与相应的卷积核的数量相同。具体的,计算机设备在空间维度下对上述任一视频帧进行特征提取时,若计算机设备采用k个卷积核对该任一视频帧进行特征提取,则可以得到k个目标特征图,即该任一视频帧对应的一组目标特征图中包括的目标特征图的数量为k,k为正整数。

[0081] 进一步的,计算机设备对各个目标特征图进行融合处理,得到任一视频帧的空间特征的具体实施过程可以包括:对各个目标特征图进行全局池化处理,得到各个目标特征图的均值(mean)和标准差(Standard Deviation,std);并对各个目标特征图的均值和标准差进行拼接处理,得到任一视频帧的空间特征。在此种情况下,该任一视频帧的空间特征包括各个目标特征图的均值和标准差;相应的,该任一视频帧的空间特征中的特征元素的数量等于各个目标特征图的数量2倍。举例来说,假设计算机设备对该任一视频帧进行特征提取,得到k个目标特征图,则计算机设备对各个目标特征图进行全局池化处理,可以得到k个均值和k个标准差,那么计算机设备可以对此处的k个均值和k个标准差进行拼接处理,得到该任一视频帧的空间特征,且该任一视频帧的空间特征包括2k个特征元素。

[0082] 可选的,计算机设备在对各个目标特征图的均值和标准差进行拼接处理时,计算机设备可以按照目标特征图的顺序,依次对每个目标特征图的均值和标准差进行拼接处理,也可以先对每个目标特征图的均值进行拼接处理,再对每个目标特征图的标准差进行拼接处理,本申请对拼接处理的具体实施方式不作限定。

[0083] 需要说明的是,计算机设备对各个目标特征图进行全局池化处理可以指的是:对各个目标特征图进行全局平均池化处理;可选的,计算机设备也可以对各个目标特征图进行全局最大池化处理,得到各个目标特征图的最大值,并对各个目标特征图的最大值进行拼接处理,得到该任一视频帧的空间特征,等等。

[0084] 可以理解的是,计算机设备在空间维度下对目标视频数据中的各个视频帧进行特征提取,可以得到每个视频帧对应的至少一个目标特征图;并分别对每个视频帧对应的至少一个目标特征图进行融合处理,得到每个视频帧的空间特征。

[0085] 可选的,各个视频帧的空间特征是通过特征提取模块得到的。例如,如图5所示,计算机设备可通过特征提取模块对N个视频帧进行特征提取,得到N组目标特征图,一组目标特征图包括一个或多个目标特征图;进一步的,计算机设备可以分别对每个视频帧对应的各个目标特征图进行全局池化处理,得到每个视频帧对应的各个目标特征图的均值和标准

差,并对每个视频帧对应的各个目标特征图的均值和标准差进行拼接处理,得到每个视频帧的空间特征,即N个空间特征。需要说明的是,图5仅示例性的表示了通过特征提取模块对N个视频帧进行特征提取的具体过程,本申请对此不作限定;例如视频帧1对应的目标特征图的数量还可以为2或3等。

[0086] 又如,假设N的取值为720,则输入特征提取模块的总帧数为720帧,且假设视频画面长宽均为512,即每个视频帧的长宽均为512,那么具体可包括表1所示的特征提取模块网络结构:

[0087] 表1

流程	张量大小
输入目标视频数据	(720, 3, 512, 512)
经过 CNN (ResNet-50)	(720, 2048, 16, 16)
经过 GP	mean (720, 2048, 1, 1) std (720, 2048, 1, 1)
拼接处理	(720, 4096)

[0088] 参见上述表1所示,目标视频数据输入网络前的张量大小为(720,3,512,512),此处的张量大小用于指示目标视频数据对应的特征图的数量和尺寸;也就是说,在输入网络(即特征提取模块)前,可在目标视频数据中选取720个视频帧,每个视频帧包括3个输入通道,且每个通道对应的特征图的长宽均为512,即此时每个视频帧包括3个长宽均为512的特征图。在经过ResNet-50按帧提取空间特征后,得到大小为(720,2048,16,16)的张量;也就是说,计算机设备通过ResNet-50对每个视频帧进行特征提取后,可以得到每个视频帧对应的2048个目标特征图,且每个目标特征图的长宽均为16。接下来,再经过全局池化(GP),分别得到大小均为(720,2048,1,1)的mean张量和std张量;也就是说,计算机设备对每个视频帧的各个目标特征图进行全局池化处理后,可得到每个视频帧对应的2048个均值和2048个标准差,且每个均值的长宽均为1,每个标准差的长宽均为1。进一步的,计算机设备将二者组合并去掉冗余维度最终得到目标视频数据的空间特征为(720,4096)大小的张量;也就是说,计算机设备可对每个视频帧对应的2048个目标特征图的均值和标准差进行拼接处理,得到720个视频帧的空间特征,且每个视频帧的空间特征包括4096个特征元素。

[0089] S403,根据各个视频帧的空间特征包括的特征元素的数量,分别确定每个视频帧对应的特征元素重组尺寸。

[0090] 需要说明的是,任一视频帧对应的特征元素重组尺寸所指示的长宽之间的乘积等于该任一视频帧的空间特征包括的特征元素的数量。

[0091] 举例来说,假设任一视频帧的空间特征包括4096个特征元素,那么计算机设备可以确定该任一视频帧对应的特征元素重组尺寸所指示的长宽均为64,此时该任一视频帧对应的特征元素重组尺寸所指示的长宽之间的乘积为4096。

[0092] 可选的,任一视频帧对应的特征元素重组尺寸所指示的长宽可以相同,也可以不同,本申请对此不作限定。

[0094] S404,按照每个视频帧对应的特征元素重组尺寸,分别对每个视频帧的空间特征中的特征元素进行元素重组,得到每个视频帧的重组空间特征。

[0095] 可以理解的是,任一视频帧的重组空间特征包含的特征元素与该任一视频帧的空间特征中的特征元素相同。举例来说,假设任一视频帧的空间特征包括特征元素A、特征元素B、特征元素C以及特征元素D,那么该任一视频帧对应的特征元素重组尺寸所指示的长宽可以均为2,且该任一视频帧的重组空间特征包括特征元素A、特征元素B、特征元素C以及特征元素D。

[0096] 可选的,计算机设备可以通过reshape(一种元素重组函数)操作,分别对每个视频帧的空间特征进行元素重组,得到每个视频帧的重组空间特征。

[0097] S405,在时序维度下根据各个视频帧之间的时序关系,对各个视频帧的重组空间特征进行特征融合,得到目标视频数据的时序空间特征。

[0098] 在一种实施方式中,计算机设备在时序维度下根据各个视频帧之间的时序关系,对各个视频帧的重组空间特征进行特征融合,得到目标视频数据的时序空间特征的具体实施过程可以包括:按照预设的滑动步长采用滑动窗口在N个视频帧中进行滑动;在每次滑动上述滑动窗口后,将N个视频帧中当前位于滑动窗口内的各个视频帧均作为目标视频帧;在时序维度下根据各个目标视频帧之间的时序关系,对各个目标视频帧的重组空间特征进行三维卷积处理,得到一个子时序空间特征;在滑动窗口结束滑动后,采用得到的各个子时序空间特征,构建目标视频数据的时序空间特征。在此种情况下,计算机设备可以采用3D卷积神经网络对各个视频帧的重组空间特征进行特征融合。

[0099] 可选的,上述预设的滑动步长可以是计算机设备设置的,也可以是用户按照意愿设置的,如1或2等,本申请对此不作限定。相应的,上述滑动窗口的深度可以为3,也可以为5,本申请对此不作限定;具体的,若滑动窗口的深度为3,那么当前位于滑动窗口内的视频帧的数量为3,那么计算机设备可对3个目标视频帧的重组空间特征进行三维卷积处理。需要说明的是,此处的滑动步长和滑动窗口的深度均是针对时序维度所设置的,本申请对空间维度下的滑动步长和滑动窗口的长宽不作限定;也就是说,上述滑动窗口可以指的是三维滑动窗口,则计算机设备可以采用三维滑动窗口在N个视频帧的重组空间特征组成的立方体中进行滑动,并在时序维度下根据各个目标视频帧之间的时序关系,对所有视频帧的重组空间特征进行三维卷积处理,得到各个子时序空间特征。

[0100] 需要说明的是,若计算机设备采用单个3D卷积核对各个目标视频帧的重组空间特征进行三维卷积处理,则可得到包括单个特征图的一个子时序空间特征;若计算机设备采用多个3D卷积核对各个目标视频帧的重组空间特征进行三维卷积处理,则可得到包括多个特征图的一个子时序空间特征;其中,一个子时序空间特征包括的特征图的数量与3D卷积核的数量相同。换句话说,上述所提及的一个子时序空间特征可以包括每个卷积核对应的特征图,即可以包括上述各个目标视频帧在每个卷积核下的特征图,且一个子时序空间特征包括的特征图的数量等于3D卷积核的数量。例如假设计算机设备采用H个3D卷积核对上述各个目标视频帧的重组空间特征进行三维卷积处理,则可得到一个包括H个特征图的子时序空间特征,H为正整数。

[0101] 另一种实施方式中,计算机设备在时序维度下根据各个视频帧之间的时序关系,对各个视频帧的重组空间特征进行特征融合,得到目标视频数据的时序空间特征的具体实

施过程可以包括:遍历各个视频帧,并将当前遍历的当前视频帧的重组空间特征确定为当前重组空间特征;在时序维度下根据各个视频帧之间的时序关系,确定当前视频帧的相邻视频帧;基于相邻视频帧的重组空间特征,对当前重组空间特征进行特征提取,得到当前视频帧对应的子时序空间特征;在各个视频帧均被遍历后,采用各个视频帧对应的子时序空间特征,构建目标视频数据的时序空间特征。可选的,计算机设备可以基于相邻视频帧的重组空间特征,通过RNN或LSTM等神经网络对当前重组空间特征进行特征提取,得到当前视频帧对应的子时序空间特征,本申请对计算当前视频帧对应的子时序空间特征的具体实施方式不作限定。

[0102] 可选的,目标视频数据的时序空间特征是通过特征融合模块得到的。例如,如图6所示,当计算机设备对目标视频数据中的N个视频帧进行特征提取时,可以得到N个视频帧的空间特征;在此种情况下,计算机设备可以通过特征融合模块对每个视频帧的空间特征中的特征元素进行元素重组,得到N个重组空间特征,并对各个重组空间特征进行特征融合,得到目标视频数据的时序空间特征。可以理解的是,计算机设备通过特征提取模块得到的N个空间特征,可经过reshape操作(即元素重组操作)得到长宽相等的张量,即长宽相等的重组空间特征;进一步的,可采用3D卷积神经网络对各个视频帧的重组空间特征进行特征融合,得到目标视频数据的时序空间特征;其中,3D卷积神经网络可以捕获每一个视频帧前后相邻的视频帧,使特征融合模块能较好地融合帧间信息。

[0103] S406,基于目标视频数据的时序空间特征,对目标视频数据进行质量检测,得到目标视频数据的视频质量。

[0104] 具体的,目标视频数据的时序空间特征包括多个子时序空间特征,计算机设备基于目标视频数据的时序空间特征,对目标视频数据进行质量检测,得到目标视频数据的视频质量的具体实施过程可以包括:对目标视频数据的时序空间特征中的各个子时序空间特征进行全局池化处理,基于全局池化处理结果确定目标视频数据的视频特征向量;采用视频特征向量对目标视频数据进行质量检测,得到目标视频数据的视频质量。

[0105] 需要说明的是,若上述多个子时序空间特征是计算机设备通过3D卷积神经网络对目标视频数据的重组空间特征进行三维卷积处理所得的,则计算机设备在对各个子时序空间特征进行全局池化处理可以指的是:以单个卷积核为单位,分别对每个卷积核在各个子时序空间特征对应的特征图进行全局池化处理,得到每个卷积核对应的全局池化处理结果。换句话说,针对H个卷积核中的任一卷积核而言,可从每个子时序空间特征中提取该任一卷积核对应的特征图,从而采用提取出的多个特征图构成一组特征图。例如假设有2048个依序排列的卷积核,23个子时序空间特征;且按照卷积核的排序顺序,每个子时序空间特征包括2048个 $2 \times 2$ 的特征图,针对第1个卷积核而言,便可从每个子时序空间特征中均提取第1个 $2 \times 2$ 的特征图来构建得到一组特征图,该组特征图便可包含23个 $2 \times 2$ 的特征图。基于此原理便可得到H组特征图,进而可分别对每组特征图进行全局池化处理,得到H个全局池化处理结果;可见在此种情况下,计算机设备对目标视频数据的时序空间特征中的各个子时序空间特征进行全局池化处理,得到的全局池化处理结果的数量等于计算机设备使用的卷积核的数量。

[0106] 可选的,此处的全局池化处理可以指的是全局平均池化处理,则一个全局池化处理结果可以指的是一组特征图中所有特征元素之间的均值;此处的全局池化处理也可以指

的是全局最大池化处理,则一个全局池化处理结果可以指的是一组特征图包括的所有特征元素中的最大值,本申请对全局池化处理的具体实施方式不作限定。

[0107] 需要说明的是,计算机设备可以对每组特征图中各个特征图的长宽两个维度进行归一化处理,得到归一化处理后的各组特征图,并对归一化处理后的各组特征图进行全局池化处理,得到全局池化处理结果。

[0108] 进一步的,当计算机设备采用视频特征向量对目标视频数据进行质量检测,得到目标视频数据的视频质量时,计算机设备可以通过全连接层(Fully Connected Layers, FC)对视频特征向量进行计算,以实现目标视频数据进行质量检测的目的,从而得到目标视频数据的视频质量。具体的,计算机设备可以确定视频特征向量对应的权重向量,该权重向量中的权重元素与视频特征向量中的特征元素一一对应,也就是说,该权重向量可以用于指示视频特征向量中每个特征元素对应的权重值;在此种情况下,计算机设备可以基于该权重向量对视频特征向量中的各个特征元素进行加权求和,得到目标视频数据的视频质量。

[0109] 例如,如图6所示,在得到目标视频数据的时序空间特征后,可对目标视频数据的时序空间特征中的各个子时序空间特征进行全局池化处理,即对相应的各组特征图进行全局池化处理,得到视频特征向量,并通过全连接层对视频特征向量进行计算,得到目标视频数据的视频质量。

[0110] 又如,假设N的取值为720,且每个视频帧的空间特征包括4096个特征元素,则计算机设备可以向特征融合模块输入每个视频帧的空间特征,即输入大小为(720,4096)的空间特征,那么具体可包括表2所示的特征融合模块网络结构:

[0111] 表2

	流程	张量大小
	输入空间特征	(720, 4096)
[0112]	reshape	(1, 1, 720, 64, 64)
	3D conv	(1, 2048, 23, 2, 2)
	GP	(1, 2048, 1, 1)
[0113]	FC	(1, 1)

[0114] 参见上述表2所示,计算机设备可以向特征融合模块输入大小为(720,4096)空间特征,也就是说,计算机设备可以将720个视频帧的空间特征输入特征融合模块,每个空间特征包括4096个特征元素。然后,经过reshape操作(即元素重组操作)得到形如(1,1,720,64,64)的特征图,其中第一个数值1指的是1个批次,第二个数值1指的是输入通道数;也就是说,计算机设备可以对各个空间特征中的特征元素进行元素重组,得到720个长宽均为64的重组空间特征。接下来,再传入3D conv(即3D卷积),此处的3D conv的总体框架依旧采用ResNet-50结构,但ResNet-50原始的二维卷积均被替换为三维卷积,以此来进行特征融合;在此种情况下,经过3D conv后,得到大小为(1,2048,23,2,2)的张量;也就是说,计算机设备可以对720个重组空间特征进行特征融合,得到2048组特征图,即2048个3D卷积核在各个

子时序空间特征对应的特征图,每组特征图的深度为23,且长宽均为2,即每组特征图可以包括23个长宽均为2的特征图。进一步的,在经过全局池化(GP)归一化特征图的长宽两个维度,得到大小为(1,2048,1,1)的张量;也就是说,计算机设备可以对各组特征图中的各个特征图的长宽两个维度进行归一化处理,并对归一化处理后的各组特征图进行全局池化处理,得到2048组特征图的全局池化处理结果,即2048个3D卷积核对应的特征图的全局池化处理结果,且每个全局池化处理结果为一个数值,即每个全局池化处理结果的长宽均为1。在此种情况下,去掉冗余维度后经过全连接层(FC)得到大小为(1,1)的张量,也就是说,计算机设备可以通过全连接层对2048个全局池化处理结果确定的视频特征向量进行计算,得到目标视频数据的视频质量,且目标视频数据的视频质量为一个数值,即目标视频数据的视频质量的长宽均为1。

[0115] 本申请实施例可获取目标视频数据,该目标视频数据包括依序排列的M个视频帧;并先在空间维度下对目标视频数据中的N个视频帧进行特征提取,得到N个视频帧中的各个视频帧的空间特征;然后,可根据各个视频帧的空间特征包括的特征元素的数量,分别确定每个视频帧对应的特征元素重组尺寸,并按照每个视频帧对应的特征元素重组尺寸,分别对每个视频帧的空间特征中的特征元素进行元素重组,得到每个视频帧的重组空间特征;进一步的,在时序维度下根据各个视频帧之间的时序关系,对各个视频帧的重组空间特征进行特征融合,得到目标视频数据的时序空间特征,使得该时序空间特征不仅可指示各个视频帧的空间信息,还可指示各个视频帧之间的时序信息,可全面地利用目标视频数据的空间信息和时序信息,减少单个视频帧带来的拖影、运动模糊等问题对质量检测的影响,提升质量检测结果的准确性。

[0116] 基于上述视频处理方法的相关实施例的描述,本申请实施例还提出了一种视频处理装置,该视频处理装置可以是运行于计算机设备中的一个计算机程序(包括程序代码)。该视频处理装置可以执行图2或图4所示的视频处理方法;请参见图7,所述视频处理装置可以运行如下单元:

[0117] 获取单元701,用于获取目标视频数据,所述目标视频数据包括依序排列的M个视频帧,M为正整数;

[0118] 处理单元702,用于在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视频帧中的各个视频帧的空间特征,N为正整数且N小于或等于M;

[0119] 所述处理单元702,还用于在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的空间特征进行特征融合,得到所述目标视频数据的时序空间特征;

[0120] 所述处理单元702,还用于基于所述目标视频数据的时序空间特征,对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量。

[0121] 在一种实施方式中,任一视频帧的空间特征包括多个特征元素,处理单元702在所述在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的空间特征进行特征融合,得到所述目标视频数据的时序空间特征时,可具体用于:

[0122] 根据所述各个视频帧的空间特征包括的特征元素的数量,分别确定每个视频帧对应的特征元素重组尺寸;

[0123] 按照所述每个视频帧对应的特征元素重组尺寸,分别对所述每个视频帧的空间特征中的特征元素进行元素重组,得到所述每个视频帧的重组空间特征;



[0124] 在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的重组空间特征进行特征融合,得到所述目标视频数据的时序空间特征。

[0125] 另一种实施方式中,处理单元702在所述在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的重组空间特征进行特征融合,得到所述目标视频数据的时序空间特征时,可具体用于:

[0126] 按照预设的滑动步长采用滑动窗口在所述N个视频帧中进行滑动;

[0127] 在每次滑动所述滑动窗口后,将所述N个视频帧中当前位于所述滑动窗口内的各个视频帧均作为目标视频帧;

[0128] 在时序维度下根据各个目标视频帧之间的时序关系,对所述各个目标视频帧的重组空间特征进行三维卷积处理,得到一个子时序空间特征;

[0129] 在所述滑动窗口结束滑动后,采用得到的各个子时序空间特征,构建所述目标视频数据的时序空间特征。

[0130] 另一种实施方式中,处理单元702在所述在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的重组空间特征进行特征融合,得到所述目标视频数据的时序空间特征时,可具体用于:

[0131] 遍历所述各个视频帧,并将当前遍历的当前视频帧的重组空间特征确定为当前重组空间特征;

[0132] 在时序维度下根据所述各个视频帧之间的时序关系,确定所述当前视频帧的相邻视频帧;

[0133] 基于所述相邻视频帧的重组空间特征,对所述当前重构空间特征进行特征提取,得到所述当前视频帧对应的子时序空间特征;

[0134] 在所述各个视频帧均被遍历后,采用所述各个视频帧对应的子时序空间特征,构建所述目标视频数据的时序空间特征。

[0135] 另一种实施方式中,所述目标视频数据的时序空间特征包括多个子时序空间特征,处理单元702在基于所述目标视频数据的时序空间特征,对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量时,可具体用于:

[0136] 对所述目标视频数据的时序空间特征中的各个子时序空间特征进行全局池化处理,基于全局池化处理结果确定所述目标视频数据的视频特征向量;

[0137] 采用所述视频特征向量对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量。

[0138] 另一种实施方式中,处理单元702在在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视频帧中的各个视频帧的空间特征时,可具体用于:

[0139] 针对所述目标视频帧中的N个视频帧中的任一视频帧,在空间维度下对所述任一视频帧进行特征提取,得到一个或多个目标特征图;

[0140] 对各个目标特征图进行融合处理,得到所述任一视频帧的空间特征。

[0141] 另一种实施方式中,处理单元702在对各个目标特征图进行融合处理,得到所述任一视频帧的空间特征时,可具体用于:

[0142] 对各个目标特征图进行全局池化处理,得到所述各个目标特征图的均值和标准差;

[0143] 对所述各个目标特征图的均值和标准差进行拼接处理,得到所述任一视频帧的空间特征。

[0144] 根据本申请的一个实施例,图2或图4所示的方法所涉及各个步骤均可由图7所示的视频处理装置中的各个单元来执行的。例如,图2中所示的步骤S201可由图7中所示的获取单元701执行,步骤S202-S204均可由图7中所示的处理单元702执行。又如,图4中所示的步骤S401可由图7所示的获取单元701执行,步骤S402-S406均可由图7所示的处理单元702执行,等等。

[0145] 根据本申请的另一个实施例,图7所示的视频处理装置中的各个单元可以分别或全部合并为一个或若干个另外的单元来构成,或者其中的某个(些)单元还可以再拆分为功能上更小的多个单元来构成,这可以实现同样的操作,而不影响本申请的实施例的技术效果的实现。上述单元是基于逻辑功能划分的,在实际应用中,一个单元的功能也可以由多个单元来实现,或者多个单元的功能由一个单元实现。在本申请的其它实施例中,视频处理装置也可以包括其他单元,在实际应用中,这些功能也可以由其它单元协助实现,并且可以由多个单元协作实现。

[0146] 根据本申请的另一个实施例,可以通过在包括中央处理单元(CPU)、随机存取存储介质(RAM)、只读存储介质(ROM)等处理元件和存储元件的例如计算机的通用计算设备上运行能够执行如图2或图4中所示的相应方法所涉及各步骤的计算机程序(包括程序代码),来构造如图7中所示的视频处理装置,以及来实现本申请实施例的视频处理方法。所述计算机程序可以记载于例如计算机存储介质上,并通过计算机存储介质装载于上述计算设备中,并在其中运行。

[0147] 本申请实施例在获取目标视频数据后,可先在空间维度下对目标视频数据中的N个视频帧进行特征提取,得到N个视频帧中的各个视频帧的空间特征;然后,在时序维度下根据各个视频帧之间的时序关系,对各个视频帧的空间特征进行特征融合,得到目标视频数据的时序空间特征,使得该时序空间特征不仅可指示各个视频帧的空间信息,还可指示各个视频帧之间的时序信息,从而使得在基于该时序空间特征对目标视频数据进行质量检测时,可全局地利用目标视频数据的整体信息,有效减少单个视频帧中因对象运动或场景变换而产生的拖影情况对质量检测的影响,提升质量检测结果的准确性。

[0148] 基于上述方法实施例以及装置实施例的描述,本申请实施例还提供一种计算机设备。请参见图8,该计算机设备至少包括处理器801、输入接口802、输出接口803以及计算机存储介质804。其中,计算机设备内的处理器801、输入接口802、输出接口803以及计算机存储介质804可通过总线或其他方式连接。

[0149] 计算机存储介质804可以存储在计算机设备的存储器中,所述计算机存储介质804用于存储计算机程序,所述计算机程序包括程序指令,所述处理器801用于执行所述计算机存储介质804存储的程序指令。处理器801(或称CPU(Central Processing Unit,中央处理器))是计算机设备的计算核心以及控制核心,其适于实现一条或多条指令,具体适于加载并执行一条或多条指令从而实现相应方法流程或相应功能;在一个实施例中,本申请实施例所述的处理器801可以用于进行一系列的视频处理,具体包括:获取目标视频数据,所述目标视频数据包括依序排列的M个视频帧,M为正整数;在空间维度下对所述目标视频数据中的N个视频帧进行特征提取,得到所述N个视频帧中的各个视频帧的空间特征,N为正整数

且N小于或等于M;在时序维度下根据所述各个视频帧之间的时序关系,对所述各个视频帧的空间特征进行特征融合,得到所述目标视频数据的时序空间特征;基于所述目标视频数据的时序空间特征,对所述目标视频数据进行质量检测,得到所述目标视频数据的视频质量。

[0150] 本申请实施例还提供了一种计算机存储介质(Memory),所述计算机存储介质是计算机设备中的记忆设备,用于存放程序和数据。可以理解的是,此处的计算机存储介质既可以包括计算机设备中的内置存储介质,当然也可以包括计算机设备所支持的扩展存储介质。计算机存储介质提供存储空间,该存储空间存储了计算机设备的操作系统。并且,在该存储空间中还存放了适于被处理器加载并执行的一条或多条的指令,这些指令可以是一个或一个以上的计算机程序(包括程序代码)。需要说明的是,此处的计算机存储介质可以是高速RAM存储器,也可以是非不稳定的存储器(non-volatile memory),例如至少一个磁盘存储器;可选的,还可以是至少一个位于远离前述处理器的计算机存储介质。在一个实施例中,可由处理器加载并执行计算机存储介质中存放的一条或多条指令,以实现上述有关图2或图4所示的视频处理方法的实施例中的各个方法步骤。

[0151] 本申请实施例在获取目标视频数据后,可先在空间维度下对目标视频数据中的N个视频帧进行特征提取,得到N个视频帧中的各个视频帧的空间特征;然后,在时序维度下根据各个视频帧之间的时序关系,对各个视频帧的空间特征进行特征融合,得到目标视频数据的时序空间特征,使得该时序空间特征不仅可指示各个视频帧的空间信息,还可指示各个视频帧之间的时序信息,从而使得在基于该时序空间特征对目标视频数据进行质量检测时,可全局地利用目标视频数据的整体信息,有效减少单个视频帧中因对象运动或场景变换而产生的拖影情况对质量检测的影响,提升质量检测结果的准确性。

[0152] 需要说明的是,根据本申请的一个方面,还提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机存储介质中。计算机设备的处理器从计算机存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述图2或图4所示的视频处理方法实施例方面的各种可选方式中提供的方法。

[0153] 并且,应理解的是,以上所揭露的仅为本申请较佳实施例而已,当然不能以此来限定本申请之权利范围,因此依本申请权利要求所作的等同变化,仍属本申请所涵盖的范围。

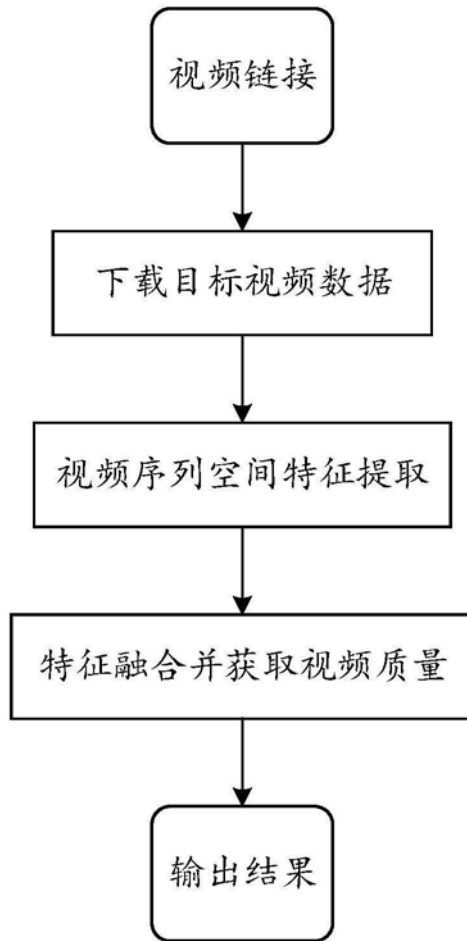


图1

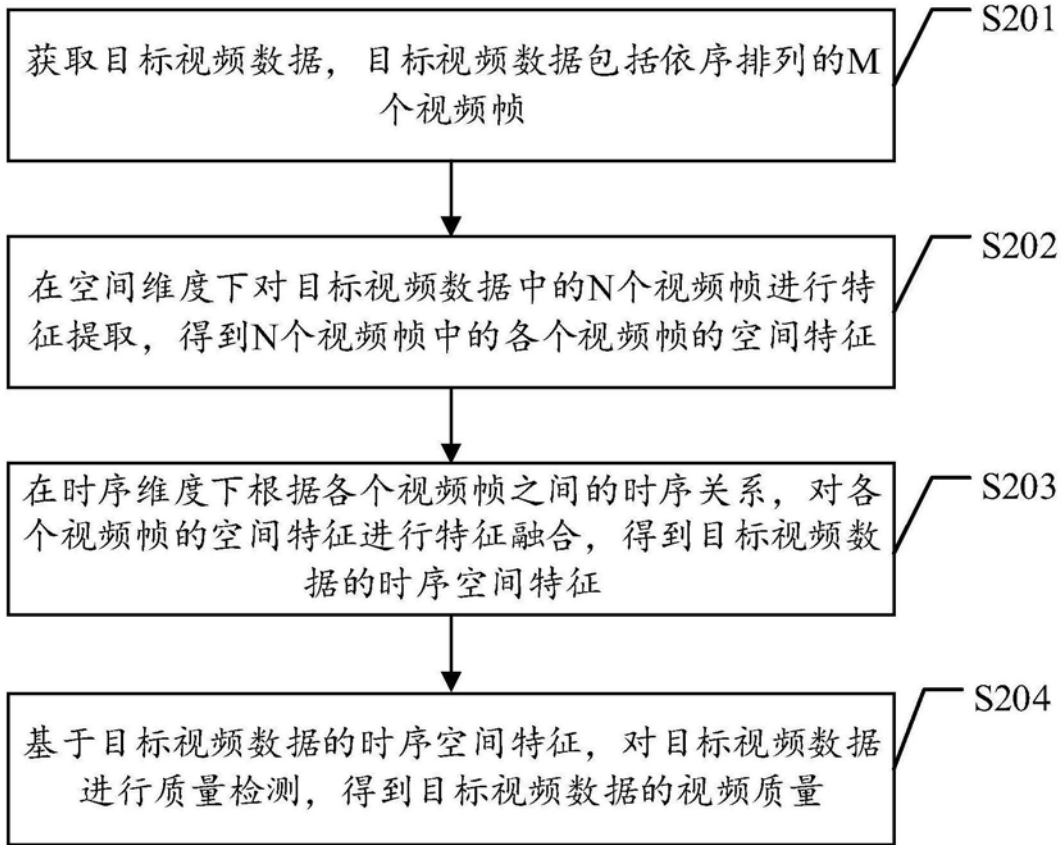


图2

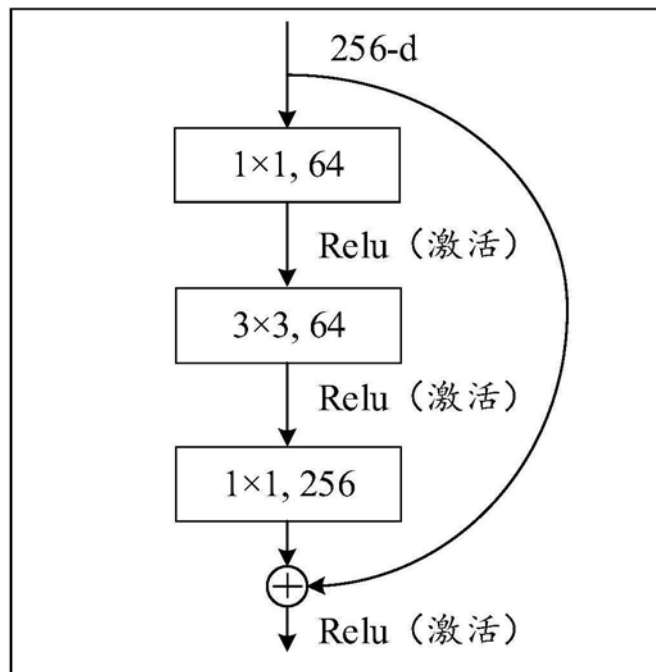


图3

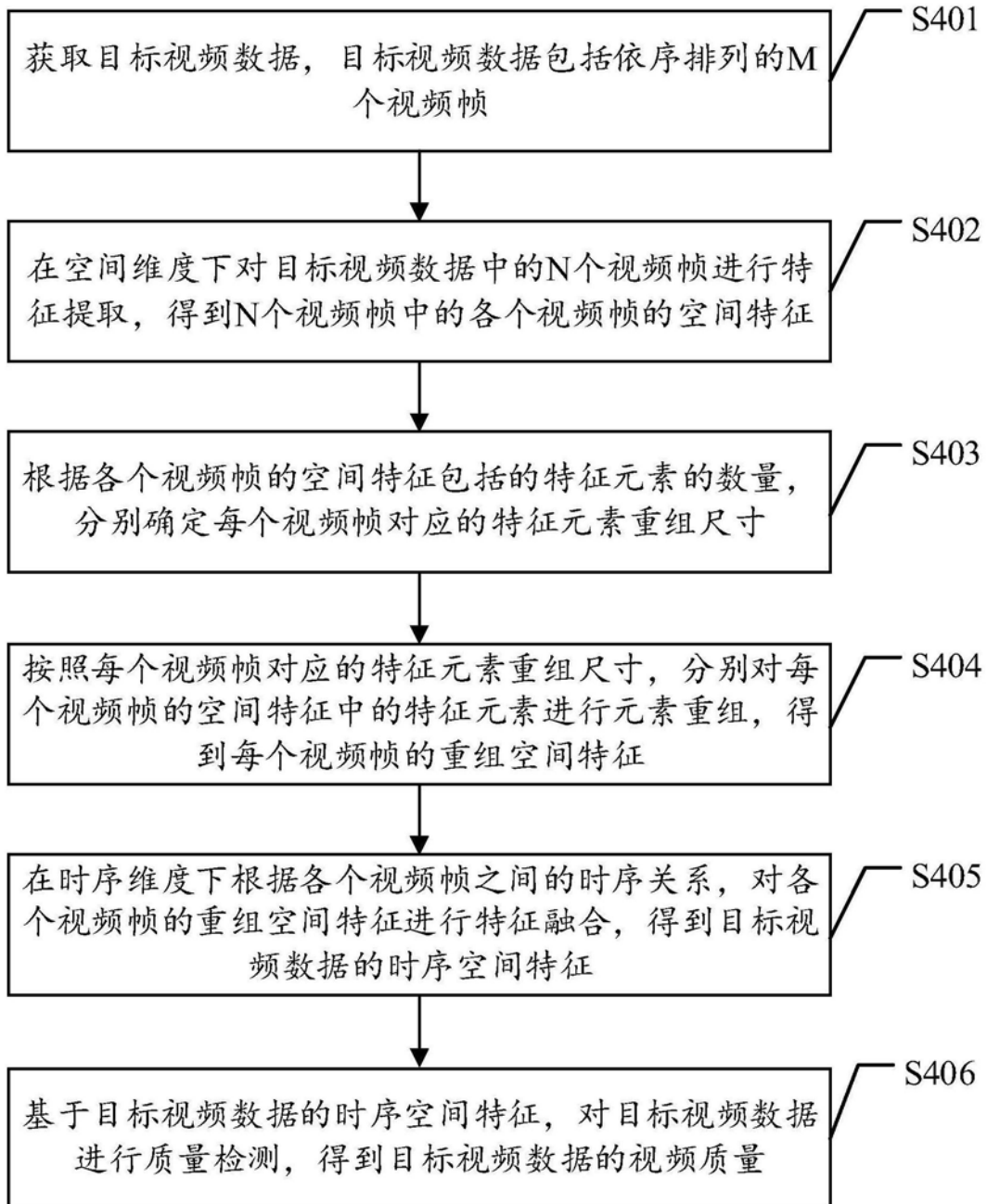


图4

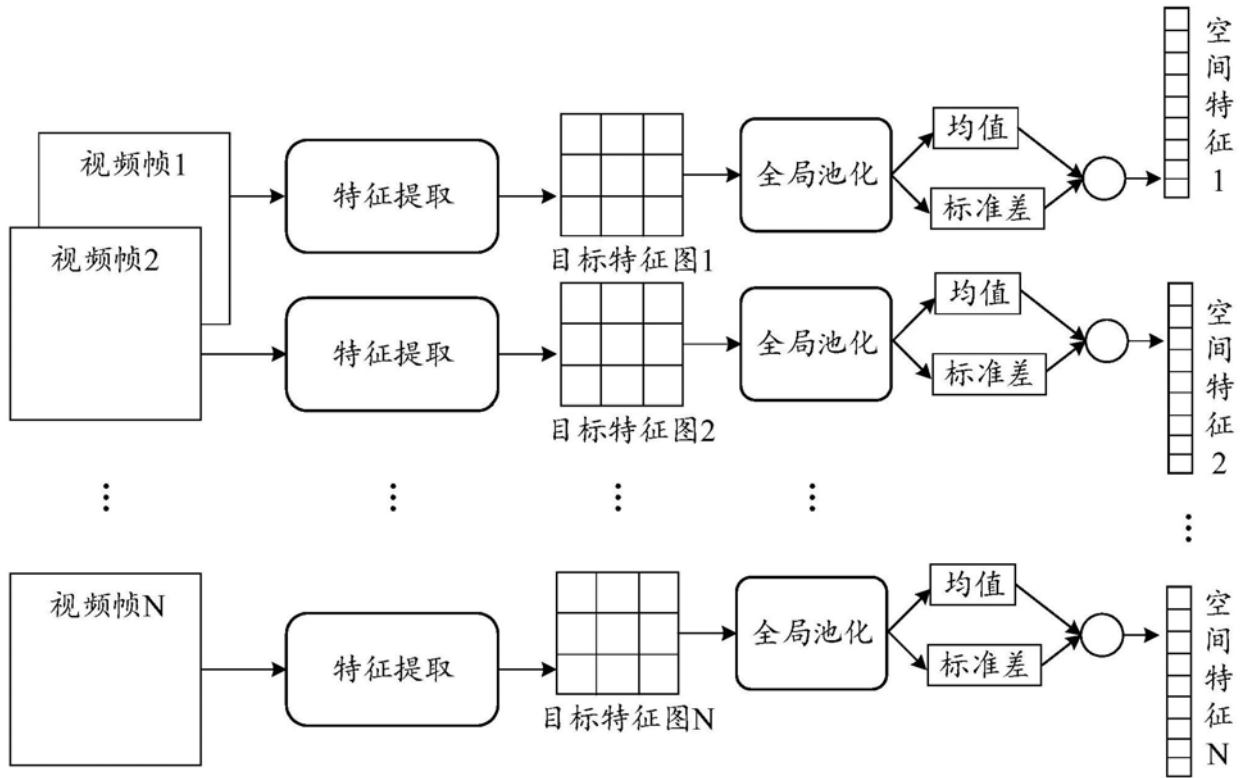


图5

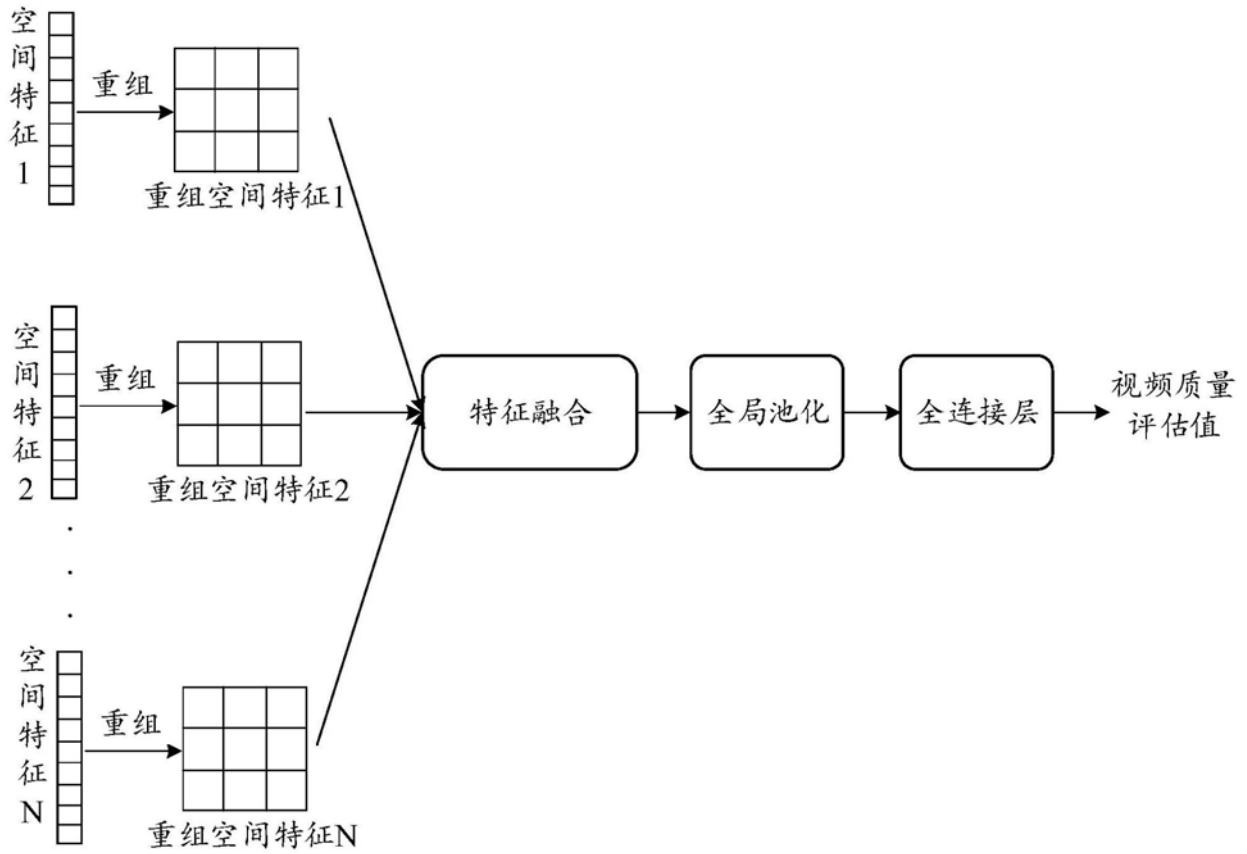


图6

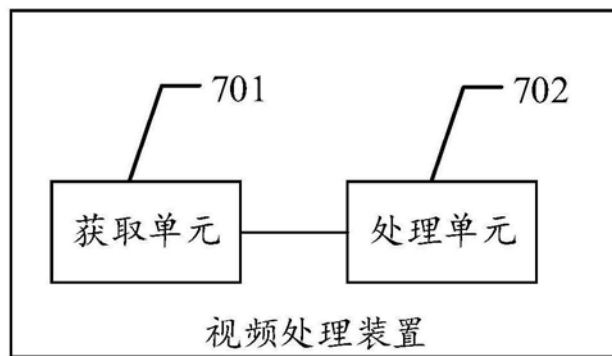


图7



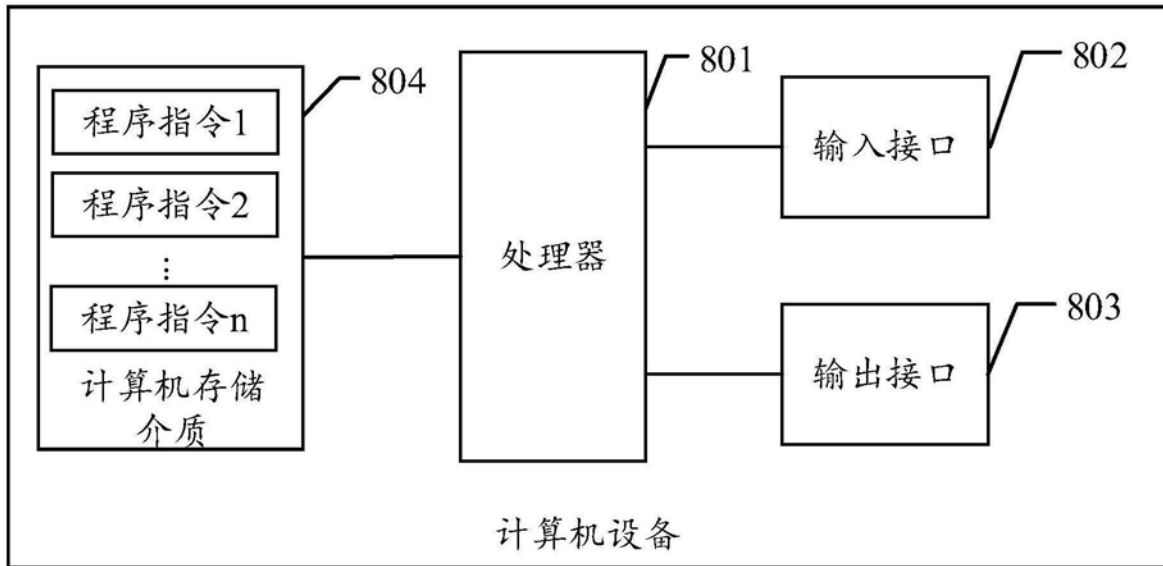


图8