(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115131698 A
(43)申请公布日 2022.09.30

(21)申请号 202210578444.8

(22)申请日 2022.05.25

(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 胡益珲 岑杰鹏 杨伟东 祁雷 马锴 陈宇

(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 贾允

(51)Int.Cl.
G06V 20/40(2022.01)
G06V 10/80(2022.01)
G06V 10/82(2022.01)
G06V 10/774(2022.01)
G06V 10/764(2022.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)
G06N 20/00(2019.01)
G06K 9/62(2022.01)
G06F 16/65(2019.01)

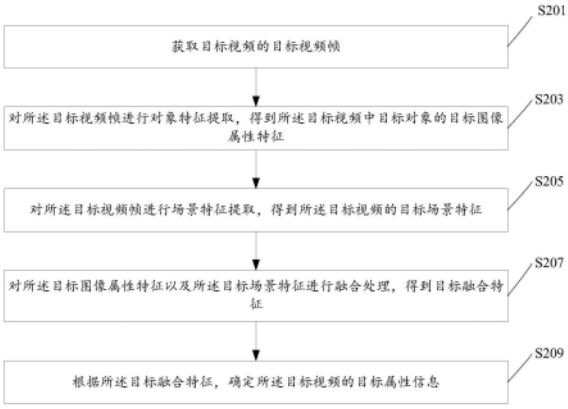
权利要求书3页 说明书16页 附图8页

(54)发明名称

视频属性确定方法、装置、设备及存储介质

(57)摘要

本申请公开了一种视频属性确定方法、装置、设备及存储介质，可以应用于云技术、人工智能、智慧交通、车联网等各种场景，所述方法包括：获取目标视频的目标视频帧；对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征；对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征；对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征；根据所述目标融合特征，确定所述目标视频的目标属性信息。本申请提高了确定的属性信息的准确率。



CN 115131698 A

1.一种视频属性确定方法,其特征在于,所述方法包括:

获取目标视频的目标视频帧;

对所述目标视频帧进行对象特征提取,得到所述目标视频中目标对象的目标图像属性特征;

对所述目标视频帧进行场景特征提取,得到所述目标视频的目标场景特征;

对所述目标图像属性特征以及所述目标场景特征进行融合处理,得到目标融合特征;

根据所述目标融合特征,确定所述目标视频的目标属性信息。

2.根据权利要求1所述的方法,其特征在于,所述对所述目标视频帧进行对象特征提取,得到所述目标视频中目标对象的目标图像属性特征,包括:

确定所述目标视频中至少两个目标对象各自的结构完整度和清晰度;

将结构完整度大于第一阈值且清晰度大于第二阈值的对象,确定为第一目标对象,将所述至少两个目标对象中除所述第一目标对象之外的对象确定为第二目标对象;

对所述第一目标对象进行对象特征提取,得到第一目标图像属性特征;

对所述第二目标对象进行对象特征提取,得到第二目标图像属性特征;

将所述第一目标图像属性特征以及所述第二目标图像属性特征作为所述目标图像属性特征。

3.根据权利要求2所述的方法,其特征在于,所述对所述第二目标对象进行对象特征提取,得到第二目标图像属性特征,包括:

对所述第二目标对象进行自重构特征提取,得到目标自重构特征;

对所述第二目标对象进行属性描述特征提取,得到目标描述特征;

将所述目标自重构特征以及所述目标描述特征,作为所述第二目标图像属性特征。

4.根据权利要求3所述的方法,其特征在于,所述对所述第二目标对象进行自重构特征提取,得到目标自重构特征,包括:

基于自重构特征提取模型,对所述第二目标对象进行自重构特征提取,得到所述目标自重构特征。

5.根据权利要求4所述的方法,其特征在于,所述自重构特征提取模型的训练方法包括:

将样本视频帧划分成至少两个网格图像;

对至少一个网格图像进行图像处理,得到处理后视频帧;所述图像处理包括网格图像的位置更换、网格图像中部分图像的遮挡处理中的至少一种;

基于所述处理后视频帧,对第一预设模型进行自重构特征提取训练,得到自重构特征;

基于所述自重构特征,对第二预设模型进行图像重构训练,得到重构视频帧;

在训练过程中,不断调整第一预设模型的第一模型参数以及第二预设模型的第二模型参数,直至所述第二预设模型输出的重构视频帧与所述样本视频帧相匹配;

将当前第一模型参数所对应的第一预设模型,作为所述自重构特征提取模型;所述当前第一模型参数为所述第一预设模型,在所述第二预设模型输出的重构视频帧与所述样本视频帧相匹配时的模型参数。

6.根据权利要求1所述的方法,其特征在于,所述方法还包括:

获取所述目标视频对应的目标音频以及目标文本;

对所述目标音频进行对象特征提取，得到所述目标对象的目标音频属性特征；

对所述目标文本进行对象特征提取，得到所述目标对象的目标文本属性特征。

7.根据权利要求6所述的方法，其特征在于，所述对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征，包括：

对所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征进行融合处理，得到所述目标融合特征。

8.根据权利要求7所述的方法，其特征在于，所述对所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征进行融合处理，得到所述目标融合特征，包括：

基于所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征，确定至少两个待融合特征；

对所述至少两个待融合特征进行融合，得到所述目标融合特征。

9.根据权利要求1所述的方法，其特征在于，所述对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征，包括：

对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标原料特征、目标命名特征以及目标类别特征；

所述对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征，包括：

对所述目标原料特征、所述目标命名特征、所述目标类别特征以及所述目标场景特征进行融合处理，得到所述目标融合特征；

所述根据所述目标融合特征，确定所述目标视频的目标属性信息，包括：

根据所述目标融合特征，确定所述目标对象的目标原料信息、目标命名信息以及目标类别信息。

10.根据权利要求1所述的方法，其特征在于，所述对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征，包括：

基于所述目标视频帧，确定所述目标对象的至少两个目标关联对象；

以所述至少两个目标关联对象各自对应的关联对象特征作为节点，构建有向无环图；所述有向无环图中的边表征所述边对应的两个关联对象特征之间的相似度；

基于所述有向无环图进行场景特征提取，得到所述目标场景特征。

11.一种视频属性确定装置，其特征在于，所述装置包括：

目标视频帧获取模块，用于获取目标视频的目标视频帧；

目标图像属性特征确定模块，用于对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征；

目标场景特征确定模块，用于对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征；

目标融合特征确定模块，用于对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征；

目标属性信息确定模块，用于根据所述目标融合特征，确定所述目标视频的目标属性信息。

12.一种视频属性确定设备,其特征在于,所述设备包括:处理器和存储器,所述存储器中存储有至少一条指令或至少一段程序,所述至少一条指令或至少一段程序由处理器加载并执行以实现如权利要求1-10任一所述的视频属性确定方法。

13.一种计算机存储介质,其特征在于,所述计算机存储介质存储有至少一条指令或至少一段程序,所述至少一条指令或至少一段程序由处理器加载并执行以实现如权利要求1-10任一所述的视频属性确定方法。

14.一种计算机程序产品,包括计算机指令,其特征在于,所述计算机指令被处理器执行时实现权利要求1-10任一所述的视频属性确定方法。

视频属性确定方法、装置、设备及存储介质

技术领域

[0001] 本申请涉及互联网技术领域，尤其涉及一种视频属性确定方法、装置、设备及存储介质。

背景技术

[0002] 相关技术中，直接使用在公开数据集上预训练好的图像模型提取特征，能够提取出较适合业务数据的粗粒度分类任务相关的视觉特征。但是缺少对多模态信息的综合建模利用，通常单依靠视觉信息来识别，缺失有辨识力的特征提取器，难区分细粒度的业务数据标签；且缺失对场景上下文的信息提取；因此，难以保证准确率和召回率。

发明内容

[0003] 本申请提供了一种视频属性确定方法、装置、设备及存储介质，可以提高确定的属性信息的准确率。

[0004] 一方面，本申请提供了一种视频属性确定方法，所述方法包括：

[0005] 获取目标视频的目标视频帧；

[0006] 对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征；

[0007] 对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征；

[0008] 对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征；

[0009] 根据所述目标融合特征，确定所述目标视频的目标属性信息。

[0010] 另一方面提供了一种视频属性确定装置，所述装置包括：

[0011] 目标视频帧获取模块，用于获取目标视频的目标视频帧；

[0012] 目标图像属性特征确定模块，用于对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征；

[0013] 目标场景特征确定模块，用于对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征；

[0014] 目标融合特征确定模块，用于对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征；

[0015] 目标属性信息确定模块，用于根据所述目标融合特征，确定所述目标视频的目标属性信息。

[0016] 另一方面提供了一种视频属性确定设备，所述设备包括处理器和存储器，所述存储器中存储有至少一条指令或至少一段程序，所述至少一条指令或所述至少一段程序由所述处理器加载并执行以实现如上所述的视频属性确定方法。

[0017] 另一方面提供了一种计算机存储介质，所述计算机存储介质存储有至少一条指令或至少一段程序，所述至少一条指令或至少一段程序由处理器加载并执行以实现如上所述

的视频属性确定方法。

[0018] 另一方面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可读存储介质中。计算机设备的处理器从计算机可读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以实现如上所述的视频属性确定方法。

[0019] 本申请提供的视频属性确定方法、装置、设备及存储介质，具有如下技术效果：

[0020] 本申请获取目标视频的目标视频帧；对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征；对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征；对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征；根据所述目标融合特征，确定所述目标视频的目标属性信息。本申请在确定视频属性信息过程中，融入了视频的场景特征，通过目标图像属性特征以及目标场景特征确定出融合特征，提高了确定的属性信息的准确率。

附图说明

[0021] 为了更清楚地说明本申请实施例或现有技术中的技术方案和优点，下面将对实施例或现有技术描述中所需要使用的附图作简单的介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，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它附图。

[0022] 图1是本申请实施例提供的一种视频属性确定系统的示意图；

[0023] 图2是本申请实施例提供的一种视频属性确定方法的流程示意图；

[0024] 图3是本申请实施例提供的一种对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征的方法的流程示意图；

[0025] 图4是本申请实施例提供的一种对所述第二目标对象进行对象特征提取，得到第二目标图像属性特征的方法的流程示意图；

[0026] 图5是本申请实施例提供的对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征的方法的流程示意图；

[0027] 图6是本申请实施例提供的一种基于目标视频帧构建的有向无环图的示意图；

[0028] 图7是本申请实施例提供的一种transformer模型的结构示意图；

[0029] 图8是本申请实施例提供的一种MLP网络的结构示意图；

[0030] 图9是本申请实施例提供的一种美食视频的属性信息确定方法的流程示意图；

[0031] 图10是本申请实施例提供的终端显示目标视频以及目标属性信息的页面；

[0032] 图11是本申请实施例提供的一种视频属性确定装置的结构示意图；

[0033] 图12是本申请实施例提供的一种服务器的结构示意图。

具体实施方式

[0034] 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，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0035] 人工智能(Artificial Intelligence,AI)是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,它企图了解智能的实质,并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。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。

[0036] 人工智能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、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0037] 计算机视觉技术(Computer Vision,CV)计算机视觉是一门研究如何使机器“看”的科学,更进一步的说,就是指用摄影机和电脑代替人眼对目标进行识别和测量等机器视觉,并进一步做图形处理,使电脑处理成为更适合人眼观察或传送给仪器检测的图像。作为一个科学学科,计算机视觉研究相关的理论和技术,试图建立能够从图像或者多维数据中获取信息的人工智能系统。计算机视觉技术通常包括图像处理、图像识别、图像语义理解、图像检索、OCR、视频处理、视频语义理解、视频内容/行为识别、三维物体重建、3D技术、虚拟现实、增强现实、同步定位与地图构建等技术,还包括常见的人脸识别、指纹识别等生物特征识别技术。

[0038] 机器学习(Machine Learning,ML)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,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0039] 本申请实施例提供的方案涉及人工智能的计算机视觉技术、机器学习等技术,具体通过如下实施例进行说明。

[0040] 需要说明的是,本申请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,以便这里描述的本申请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、产品或服务器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。

[0041] 请参阅图1,图1是本申请实施例提供的一种视频属性确定系统的示意图,如图1所示,该视频属性确定系统可以至少包括服务器01和客户端02。

[0042] 具体的,本申请实施例中,所述服务器01可以包括一个独立运行的服务器,或者分布式服务器,或者由多个服务器组成的服务器集群,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。服务器01可以包括有网络通信单元、处理器和存储器等等。具体的,所述服务器01可

以用于获取目标视频的目标视频帧;对所述目标视频帧进行对象特征提取,得到所述目标视频中目标对象的目标图像属性特征;对所述目标视频帧进行场景特征提取,得到所述目标视频的目标场景特征;对所述目标图像属性特征以及所述目标场景特征进行融合处理,得到目标融合特征;根据所述目标融合特征,确定所述目标视频的目标属性信息。

[0043] 具体的,本申请实施例中,所述客户端02可以包括智能手机、台式电脑、平板电脑、笔记本电脑、数字助理、智能可穿戴设备、智能音箱、车载终端、智能电视等类型的实体设备,也可以包括运行于实体设备中的软体,例如一些服务商提供给用户的网页页面,也可以为该些服务商提供给用户的应用。具体的,所述客户端02可以用于展示目标视频的目标属性信息。

[0044] 以下介绍本申请的一种视频属性确定方法,图2是本申请实施例提供的一种视频属性确定方法的流程示意图,本说明书提供了如实施例或流程图所述的方法操作步骤,但基于常规或者无创造性的劳动可以包括更多或者更少的操作步骤。实施例中列举的步骤顺序仅仅为众多步骤执行顺序中的一种方式,不代表唯一的执行顺序。在实际中的系统或服务器产品执行时,可以按照实施例或者附图所示的方法顺序执行或者并行执行(例如并行处理器或者多线程处理的环境)。具体的如图2所示,所述方法可以包括:

[0045] S201:获取目标视频的目标视频帧。

[0046] 在本申请实施例中,可以将提取目标视频中的视频帧,得到目标视频帧,目标视频帧可以为至少两个。目标视频可以为用于制作目标对象的视频,目标视频帧中包括目标对象,对于美食类视频,目标对象可以包括美食的原料、成品;对于手工作品类视频,目标对象可以包括手工作品的原料、成品。目标视频帧的提取方法包括均匀采帧和间隔采帧两种方式;其中,均匀采帧是指分段稀疏采样,从目标视频中均匀采样出N帧构成视频帧集合;间隔采帧是指按固定时间间隔采帧,比如1秒采一帧。目标视频帧可以为数字图像,也可以为频域图像。

[0047] S203:对所述目标视频帧进行对象特征提取,得到所述目标视频中目标对象的目标图像属性特征。

[0048] 在本申请实施例中,可以对目标视频帧进行对象特征提取,得到目标对象的目标图像属性特征。目标图像属性特征可以包括目标视频帧的RGB信息和时序信息;RGB色彩就是常说的光学三原色,R代表Red(红色),G代表Green(绿色),B代表Blue(蓝色)。自然界中肉眼所能看到的任何色彩都可以由这三种色彩混合叠加而成,因此也称为加色模式。

[0049] 在本申请实施例中,所述对所述目标视频帧进行对象特征提取,得到所述目标视频中目标对象的目标图像属性特征,包括:

[0050] 对所述目标视频帧进行对象特征提取,得到所述目标视频中的目标原料特征、目标命名特征以及目标类别特征。

[0051] 在本申请实施例中,对于制作目标对象的目标视频,目标图像属性特征可以包括目标原料特征、目标命名特征以及目标类别特征等。

[0052] 在本申请实施例中,如图3所示,所述对所述目标视频帧进行对象特征提取,得到所述目标视频中目标对象的目标图像属性特征,包括:

[0053] S2031:确定所述目标视频中至少两个目标对象各自的结构完整度和清晰度;

[0054] 在本申请实施例中,目标视频帧中可以包括多个目标对象,且每个目标对象在视

频帧中的完整度和清晰度均不相同。

[0055] S2033:将结构完整度大于第一阈值且清晰度大于第二阈值的对象,确定为第一目标对象,将所述至少两个目标对象中除所述第一目标对象之外的对象确定为第二目标对象;

[0056] 在本申请实施例中,可以根据各个目标对象在目标视频帧中的完整度和清晰度,对目标对象进行分类;并分别进行特征提取。对于第一目标对象,可以通过有监督训练模型提取其对象特征;对于第二目标对象,可以通过无监督训练模型提取其对象特征。

[0057] S2035:对所述第一目标对象进行对象特征提取,得到第一目标图像属性特征;

[0058] 在本申请实施例中,所述对所述第一目标对象进行对象特征提取,得到第一目标图像属性特征可以包括:

[0059] 基于有监督特征提取模型对所述第一目标对象进行对象特征提取,得到第一目标图像属性特征。

[0060] 在一些实施例中,所述有监督特征提取模型的训练方法包括:

[0061] 获取训练视频的训练视频帧;所述训练视频帧标注了训练对象的训练图像属性特征标签;

[0062] 在本申请实施例中,训练视频与目标视频可以为相同类型的视频,训练视频帧的提取方式与目标视频帧的提取方式相同。

[0063] 基于所述训练视频帧对预设机器学习模型进行图像属性特征提取训练,以调整所述预设机器学习模型的模型参数,至所述预设机器学习模型输出的训练图像属性特征标签与标注的训练图像属性特征标签相匹配;

[0064] 将输出的训练图像属性特征标签与标注的训练图像属性特征标签相匹配时的模型参数所对应的预设机器学习模型作为所述有监督特征提取模型。

[0065] 在本申请实施例中,预设机器学习模型可以为Video Swin Transformer,Video Swin Transformer的主干网络(backbone)的整体架构和Swin Transformer大同小异,多了一个时间维度,在做块分割(Patch Partition)的时候会有个时间维度的块大小(patch size)。Video Swin Transformer包括三个部分:video to token、model stages以及head。

[0066] 其中,Video to token:在image to token(用于将图像转化成令牌)中,是将4*4的图像块作为一组,而在Video to token(用于将视频转化成令牌)中,将2*4*4的视频块作为一组,而后再进行线性嵌入(embedding)以及位置嵌入(position embedding)。

[0067] Model stages:Model stages由多个重复的stage组成,每个Model stages包括Video Swin Transformer Block和Patch merging。

[0068] 1)Video Swin Transformer Block由可以分为两部分,Video W-MSA和Video SW-MSA。这里相当于将Swin Transformer Block计算由二维拓展到三维。

[0069] 2)Patch merging将相邻(2*2窗口内)令牌(token)特征合并,而后再利用线性层降维,相当于将token减少4倍,不过降维并没有保持维度不变,而是每次Patch merging之后特征维度仍然会增加2倍,和卷积神经网络(Convolutional Neural Network,CNN)中特征图减少,通道数增加很类似,这里每次进行patching merging,视频帧数是不变的。

[0070] head:在经过Model stages之后,得到了多帧数据的高维特征,用于视频分类的话需要进行简单的帧融合(average),可以使用头部(head)代码。

[0071] 在本申请实施例中，可以通过有监督特征提取模型快速、准确地提取目标视频帧中目标对象的目标图像属性特征。

[0072] S2037：对所述第二目标对象进行对象特征提取，得到第二目标图像属性特征；

[0073] 在本申请实施例中，如图4所示，所述对所述第二目标对象进行对象特征提取，得到第二目标图像属性特征，包括：

[0074] S20371：对所述第二目标对象进行自重构特征提取，得到目标自重构特征；

[0075] 在本申请实施例中，所述对所述第二目标对象进行自重构特征提取，得到目标自重构特征，包括：

[0076] 基于自重构特征提取模型，对所述第二目标对象进行自重构特征提取，得到所述目标自重构特征。

[0077] 在本申请实施例中，所述自重构特征提取模型的训练方法包括：

[0078] 将样本视频帧划分成至少两个网格图像；

[0079] 在本申请实施例中，样本视频帧基于样本视频确定，样本视频与训练视频可以相同或不同。

[0080] 对至少一个网格图像进行图像处理，得到处理后视频帧；所述图像处理包括网格图像的位置更换、网格图像中部分图像的遮挡处理中的至少一种；

[0081] 在本申请实施例中，可以对一个网络图像进行遮挡处理，同时对其中两个网格图像进行位置更换，从而得到处理后视频帧。

[0082] 在本申请实施例中，所述对至少一个网格图像进行图像处理，得到处理后视频帧，包括：

[0083] 在模型的第一次训练时，对至少一个网格图像进行第一次图像处理；所述第一次图像处理包括对第一数量的网格图像进行位置更换、对任一网格图像进行第一百分比的面积遮挡中的至少一种；

[0084] 在模型的第N次训练时，对至少一个网格图像进行第N图像处理；所述第N次图像处理包括对第N数量的网格图像进行位置更换、对任一网格图像进行第N百分比的面积遮挡中的至少一种；所述第N数量大于第N-1数量；所述第N百分比大于第N-1百分比；其中，N＝2，3，……，N为正整数。

[0085] 基于所述处理后视频帧，对第一预设模型进行自重构特征提取训练，得到自重构特征；

[0086] 在本申请实施例中，第一预设模型可以为编码器(encoder)模型。

[0087] 基于所述自重构特征，对第二预设模型进行图像重构训练，得到重构视频帧；

[0088] 在本申请实施例中，第二预设模型可以为解码器(decoder)模型。

[0089] 在训练过程中，不断调整第一预设模型的第一模型参数以及第二预设模型的第二模型参数，直至所述第二预设模型输出的重构视频帧与所述样本视频帧相匹配；

[0090] 将当前第一模型参数所对应的第一预设模型，作为所述自重构特征提取模型。当前第一模型参数为所述第一预设模型，在所述第二预设模型输出的重构视频帧与所述样本视频帧相匹配时的模型参数。

[0091] 在本申请实施例中，在模型训练过程中，可以将样本视频帧作为监督信号进行训练，并且对互换和遮挡的程度随着训练进程的推进逐渐增加。例如，在第一次训练时，对网

格图像的遮挡面积为10％，在第二次训练时，对所述网格图像的遮挡面积为20％；以此类推；网格图像的互换可以从两个网格图像互换增加至四个网格图像互换，从而提高模型的准确率。

[0092] 在本申请实施例中，可以通过自重构特征提取模型提取目标视频帧中的目标自重构特征，从而结合该特征确定视频属性信息，提高视频属性信息的准确率。

[0093] S20373：对所述第二目标对象进行属性描述特征提取，得到目标描述特征；

[0094] 在本申请实施例中，可以通过属性描述特征提取模型对所述第二目标对象进行属性描述特征提取，得到目标描述特征；属性描述特征可以用于描述第二目标对象的属性；其中属性描述特征提取模型可以基于图文匹配度训练的CLIP模型训练得到；CLIP(Contrastive Language-Image Pre-Training，对比语言－图像预训练)是在各种(图像，文本)对上训练的神经网络。可以用自然语言指示它来预测给定图像的最相关的文本片段，而无需直接针对任务进行优化。

[0095] S20375：将所述目标自重构特征以及所述目标描述特征，作为所述第二目标图像属性特征。

[0096] 在本申请实施例中，所述目标自重构特征以及所述目标描述特征可以为同一类型的特征或者不同类型的特征。

[0097] S2039：将所述第一目标图像属性特征以及所述第二目标图像属性特征作为所述目标图像属性特征。

[0098] 在本申请实施例中，可以将第一目标对象的第一目标图像属性特征以及第二目标对象的第二目标图像属性特征，作为所述目标图像属性特征。

[0099] S205：对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征。

[0100] 在本申请实施例中，场景特征为目标视频中场景对应的特征；在美食视频中，场景特征可以包括但不限于餐桌、锅、刀等对应的特征。

[0101] 在本申请实施例中，如图5所示，所述对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征，包括：

[0102] S2051：基于所述目标视频帧，确定所述目标对象的至少两个目标关联对象；

[0103] 在本申请实施例中，目标关联对象用于表征目标视频的场景，例如，在美食视频中，目标对象为美食的原材料、成品等；目标关联对象包括餐桌、锅、刀等。

[0104] S2053：以所述至少两个目标关联对象各自对应的关联对象特征作为节点，构建有向无环图；所述有向无环图中的边表征所述边对应的两个关联对象特征之间的相似度；

[0105] 在本申请实施例中，可以通过多个目标视频帧构建有向无环图；有向无环图中的节点表征关联对象特征，边表征所述边对应的两个关联对象特征之间的相似度。

[0106] S2055：基于所述有向无环图进行场景特征提取，得到所述目标场景特征。

[0107] 在本申请实施例中，可以通过有向无环图提取视频中的场景特征，具体的，可以通过图卷积神经网络(Graph Convolutional Networks，GCN)提取所述有向无环图中的场景特征，得到所述目标场景特征。

[0108] 在一个具体的实施例中，如图6所示，图6为一种基于目标视频帧构建的有向无环图的示意图；其中该图对应t＝k至t＝M之间的视频帧；通过GCN提取该有向无环图的场景特征，得到目标场景特征；最后，将目标场景特征与目标图像属性特征进行融合后，输入属性

信息预测模型,得到目标视频的目标属性信息。属性信息预测模型可以通过对多层感知机(MLP,Multilayer Perceptron)网络进行训练得到。

[0109] 在本申请实施例中,所述方法还包括:

[0110] 获取所述目标视频对应的目标音频以及目标文本;

[0111] 在本申请实施例中,可以对目标视频进行解析,提取其对应音频信号,得到目标音频;通过OCR(Optical Character Recognition,光学字符识别)提取目标视频对应的文本信息,得到目标文本。

[0112] 对所述目标音频进行对象特征提取,得到所述目标对象的目标音频属性特征;

[0113] 在本申请实施例中,可以通过音频属性特征提取模型,对所述目标音频进行对象特征提取,得到所述目标对象的目标音频属性特征;音频属性特征提取模型可以通过VGGish网络训练得到,最终得到固定维度的目标音频属性特征。其中,在大量的视频数据集上训练得到类VGG模型,该模型中生成128维的embedding。

[0114] 将基于tensorflow(一个基于数据流编程的符号数学系统)的VGG模型,称为VGGish。VGGish支持从音频波形中提取具有语义的128维embedding特征向量。“VGG”代表了牛津大学的Oxford Visual Geometry Group,该小组研究范围包括了机器学习到移动机器人。

[0115] VGG模型的特征如下:

[0116] (1)小卷积核(3*3卷积);

[0117] (2)小池化核(2*2的池化核);

[0118] (3)层数更深且特征图更宽。基于前两点外,由于卷积核专注于扩大通道数、池化专注于缩小宽和高,使得模型架构上更深更宽的同时,计算量的增加放缓;

[0119] (4)全连接转卷积。网络测试阶段将训练阶段的三个全连接替换为三个卷积,测试重用训练时的参数,使得测试得到的全卷积网络因为没有全连接的限制,因而可以接收任意宽或高为的输入。

[0120] 在本申请实施例中,通过音频属性特征提取模型,可以实现快速、准确地提取目标对象的目标音频属性特征。

[0121] 对所述目标文本进行对象特征提取,得到所述目标对象的目标文本属性特征。

[0122] 在本申请实施例中,对于视频中的文本信息,我们的数据主要来自标题和视频中OCR的识别结果,将上述两个文本使用BERT模型抽取相应的文本属性特征。BERT是2018年10月由Google AI研究院提出的一种预训练模型。BERT的全称是Bidirectional Encoder Representation from Transformers。可以通过BERT模型快速提取目标对象的目标文本属性特征。

[0123] S207:对所述目标图像属性特征以及所述目标场景特征进行融合处理,得到目标融合特征。

[0124] 在本申请实施例中,所述对所述目标图像属性特征以及所述目标场景特征进行融合处理,得到目标融合特征,包括:

[0125] 对所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征进行融合处理,得到所述目标融合特征。

[0126] 在本申请实施例中,所述对所述目标图像属性特征、所述目标场景特征、所述目标

音频属性特征以及目标文本属性特征进行融合处理，得到所述目标融合特征，包括：

[0127] 对所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征进行拼接，得到目标拼接特征；

[0128] 基于特征融合模型对所述目标拼接特征进行特征融合处理，得到所述目标融合特征。

[0129] 在本申请实施例中，所述特征融合模型的训练方法包括：

[0130] 获取标注了样本融合特征的样本拼接特征；

[0131] 在本申请实施例中，样本拼接特征的构建方法与目标拼接特征相同。

[0132] 基于所述样本拼接特征对预设多模态模型进行训练，以调整所述预设多模态模型的模型参数，至所述预设多模态模型输出的样本融合特征与标注的样本融合特征相匹配；

[0133] 在本申请实施例中，预设多模态模型可以为transformer(变换器网络)模型，transformer模型采用memorybank(信息存储)机制。

[0134] 将输出的样本融合特征与标注的样本融合特征相匹配时的预设多模态模型作为所述特征融合模型。

[0135] 在本申请实施例中，由于不同的特征是由不同监督信号、不同模型抽取得到的不同维度特征。他们的隐空间分布也不同，直接融合会导致特征表达的不一致，最终影响模型的识别性能。因此，可以通过transformer和memorybank机制来对齐不同模态的差异分布问题。首先，将多种特征顺序相接，并在连接处添加划分符号，输入到transformer模型中。如图7所示，图7为一种transformer模型的结构示意图，其中每个方块是拼接后的特征。transformer模型通过多头注意力机制(multi-head)对输入的拼接后多模态特征进行自注意力机制的学习，以及和memorybank中的特征进行相似度的计算，来调节不同模态的比重。最终将融合后的特征，存进memorybank中，并输出到下一层网络。transformer模型可以为缩放点积注意力机制(Scaled Dot-Product Attention)模型，该模型有两个序列X XX、YYY:序列X XX提供查询信息Q(query)，序列Y YY提供键、值信息K(key)、V(value)。Q就是词的查询向量，K是“被查”向量，V是内容向量。其中，Q是最适合查找目标的，K是最适合接收查找的，V就是内容，这三者不一定要一致，所以网络这么设置了三个向量，然后学习出最适合的Q，K，V，以此增强网络的能力。模型中包括矩阵乘法层(Mat Multiply，MatMul)、缩放层(scale)、遮罩层(mask)归一化指数函数层(Softmax)，其中，矩阵乘法层包括第一矩阵乘法层和第二矩阵乘法层；获取当前拼接特征的Q、K、V；将当前拼接特征的Q、K输入第一矩阵乘法层，将当前拼接特征的V输入第二矩阵乘法层；第一矩阵乘法层的输出端与缩放层的输入端连接，缩放层的输出端与遮罩层的输入端连接，遮罩层的输出端与归一化指数函数层的输入端连接，归一化指数函数层的输出端与第二矩阵乘法层的输入端连接；将第二矩阵乘法层的输出作为memorybank中已存储特征的Q，同时获取memorybank中已存储特征的K、V，分别输入第一矩阵乘法层、第二矩阵乘法层进行模型训练。

[0136] 在一些实施例中，还可以使用基于通道的注意力机制来计算各个特征在通道上的相似度，以此来进行特征融合。

[0137] 在本申请实施例中，所述对所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征进行融合处理，得到所述目标融合特征，包括：

[0138] 基于所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目

标文本属性特征，确定至少两种待融合特征；

[0139] 在本申请实施例中，可以通过梯度提升树(Gradient Boosting Decision Tree，GBDT)网络确定至少两种待融合特征；具体的，可以为同一训练视频的训练图像属性特征、所述训练场景特征、所述训练音频属性特征以及训练文本属性特征标注属性信息标签；并根据训练图像属性特征、所述训练场景特征、所述训练音频属性特征以及训练文本属性特征，对GBDT网络进行训练，在模型收敛时，得到训练图像属性特征、所述训练场景特征、所述训练音频属性特征以及训练文本属性特征各自对应的权重；并将权重大于预设权重阈值的训练特征对应的目标特征，确定为待融合特征。

[0140] 对所述至少两种待融合特征进行融合，得到所述目标融合特征。

[0141] 在本申请实施例中，可以选择目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征中的至少两种特征，进行融合处理，得到目标融合特征，从而提高确定的属性信息的准确率。

[0142] 在本申请实施例中，所述对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征，包括：

[0143] 对所述目标原料特征、所述目标命名特征、所述目标类别特征以及所述目标场景特征进行融合处理，得到所述目标融合特征。

[0144] S209：根据所述目标融合特征，确定所述目标视频的目标属性信息。

[0145] 在本申请实施例中，所述根据所述目标融合特征，确定所述目标视频的目标属性信息，包括：

[0146] 根据所述目标融合特征，确定所述目标对象的目标原料信息、目标命名信息以及目标类别信息。

[0147] 在本申请实施例中，目标图像属性特征可以包括目标原料特征、所述目标命名特征、所述目标类别特征，将这些特征与目标场景特征进行融合，可以得到目标融合特征。

[0148] 在本申请实施例中，可以将目标原料信息、目标命名信息以及目标类别信息作为三个预测任务，可以通过属性信息预测模型对所述目标融合特征进行属性信息预测处理，得到所述目标视频的目标属性信息；属性信息预测模型可以通过三层的MLP训练得到，如图8所示，图8为一种MLP网络的结构示意图，包括三个任务预测网络，每个任务预测网络对应一个损失值，其中每一层负责一个任务的预测，且每一个任务有单独的损失函数，最后的损失函数为各个任务损失值的加权和的平均值。

[0149] 在训练过程中，还可以结合速度梯度法(Fast Gradient Method，FGM)进行对抗训练；为训练时的梯度增加噪声，来增强模型的泛化性能。

[0150] 例如，在任务1中原始的损失函数为：

[0151] $L(\theta,x,y)=-\min_{\theta}\log p(y|x,\theta)$；其中，L代表损失函数，表征样本属性特征x对应的融合特征值与真实值y之间的差距大小；θ为参数。

[0152] 对梯度增加扰动$r_{adv}=\varepsilon\cdot\left(\frac{g}{\|g\|_2}\right)$，其中$g=\nabla_x(L(\theta,x,y))$，得到最终的损失函数为$L(\theta,x,y)+L_{adv}(\theta,x,y)$。

[0153] 在本申请实施例中，属性信息预测模型还可以通过其他多任务结构的模型训练得到，例如可以使用教师和学生网络的结构来进行多任务学习。针对不同维度的样本，网络学

习到的知识可以通过蒸馏的方式来进行传递。

[0154] 在一个具体的实施例中，分别采用实施例1、2确定视频的属性信息的预测准确率和召回率，结果如表1所示；实施例1：采用目标图像属性特征预测视频的属性信息；实施例2：对所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征进行融合处理，得到所述目标融合特征；根据目标融合特征预测视频的属性信息。

[0155] 表1

[0156]

	准确率	召回率
实施例1	79.6%	48.1%
实施例2	81.7%	68.0%

[0157] 在一个具体的实施例中，分别采用实施例3、4确定视频的属性信息的预测准确率和召回率，结果如表2所示；实施例3：根据有监督特征提取模型对所述目标视频帧进行对象特征提取，得到目标图像属性特征，进行视频的属性信息的预测；实施例4：引入CBAM、基于patch的随机Mask+Shuffle的重构任务，结合有监督特征提取模型得到的目标图像属性特征以及自重构特征，进行视频的属性信息的预测。

[0158] 表2

[0159]

	准确率	召回率
实施例3	55.5%	68.5%
实施例4	51.3%	66.2%

[0160] 在一个具体的实施例中，如图9所示，图9为一种美食视频的属性信息确定方法的流程示意图，所述方法包括：

[0161] S901：将视频解析为视频帧集合、音频信号以及文本信息；

[0162] S903：基于音频信息提取模型提取音频信号对应的音频特征；基于文本信息提取模型提取文本信息对应的文本特征；基于视频帧信息提取模型，提取视频帧集合对应的视频帧特征；

[0163] S905：基于图文匹配模型对视频帧集合进行属性描述特征提取，得到属性描述特征；基于特征重构模型对视频帧集合进行重构特征提取，得到自重构特征；

[0164] S907：根据变换器网络模型以及信息存储模型对所述音频特征、文本特征、视频帧特征、属性描述特征、自重构特征进行融合，得到目标融合特征；

[0165] S909：基于多任务学习模型对目标融合特征进行预测，得到视频的菜系标签、菜名标签以及材料标签。

[0166] 在本申请实施例中，所述方法还包括：

[0167] 向终端发送所述目标视频以及所述目标属性信息；以使所述终端展示所述目标视频以及所述目标属性信息。

[0168] 根据目标视频与目标属性信息的对应关系，构建视频属性映射关系。

[0169] 在一些实施例中，所述方法还包括：

[0170] 根据目标属性信息，确定关联属性信息；

[0171] 根据所述目标属性信息，确定目标视频的第一标签；

[0172] 根据所述关联属性信息，确定目标视频的第二标签。

[0173] 在本申请实施例中，关联属性信息可以为与目标属性信息的相似度大于预设值的

信息，第一标签可以为高置信度标签，第二标签可以为低置信度标签。

[0174] 在一些实施例中，所述方法还包括：

[0175] 根据所述目标属性信息，确定所述目标视频的类别信息；

[0176] 向终端发送所述目标视频、所述目标属性信息以及所述目标视频的类别信息。

[0177] 在本申请实施例中，如图10所示，图10为终端显示目标视频以及目标属性信息的页面；页面中展示了视频类别对应的一级分类以及二级分类标签；还展示了视频对应的属性标签，包括高置信度标签和低置信度标签。此外，还可以根据属性标签，构建标签树，并可以展示标签维度信息。

[0178] 在一些实施例中，所述方法还包括：

[0179] 接收终端响应于视频获取指令，发送的视频获取请求；所述视频获取请求携带待获取视频的关联信息；

[0180] 确定与所述待获取视频的关联信息匹配的目标属性信息，得到匹配属性信息；

[0181] 从所述视频属性映射关系中，查找与所述匹配属性信息对应的目标视频，作为所述待获取视频；

[0182] 向所述终端发送所述待获取视频。

[0183] 在本申请实施例中，如图10所示，用户在搜索框中输入“怎样炒土豆”，即可显示对应的视频，并可显示对应的视频，并显示视频对应的属性标签。

[0184] 在一些实施例中，所述方法还包括：

[0185] 确定所述目标视频对应的目标账户；

[0186] 根据所述目标属性信息，确定所述目标账户的账户属性信息。

[0187] 在一些实施例中，所述目标视频为至少两个，所述方法还包括：

[0188] 根据所述至少两个目标视频各自对应的目标属性信息，对所述至少两个目标视频进行分类。

[0189] 在一些实施例中，所述方法还包括：

[0190] 确定目标类别的目标视频集；

[0191] 获取所述目标视频集中各个目标视频的业务数据；

[0192] 在本申请实施例中，目标视频的业务数据可以包括但不限于目标视频的点击率、曝光量等数据。

[0193] 根据所述目标视频集中各个目标视频的业务数据，对所述目标视频集中目标视频进行排序；

[0194] 向终端发送所述目标视频集以及所述目标视频集中目标视频的排序结果；

[0195] 终端根据所述排序结果，显示所述目标视频集中的目标视频。

[0196] 在本申请实施例中，可以根据业务数据对多个目标视频进行排序，将业务指标较高的目标视频排在前面，将业务指标较低的目标视频排在后面；业务数据用于表征业务指标的高低，进一步反映出视频的质量，根据业务指标对视频进行排序，并按照排序来展示多个目标视频，可以提升用户体验，提升目标视频的点击率。

[0197] 由以上本申请实施例提供的技术方案可见，本申请实施例获取目标视频的目标视频帧；对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征；对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征；对所述

目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征；根据所述目标融合特征，确定所述目标视频的目标属性信息。本申请在确定视频属性信息过程中，融入了视频的场景特征，通过目标图像属性特征以及目标场景特征确定出融合特征，提高了确定的属性信息的准确率。

[0198] 本申请实施例还提供了一种视频属性确定装置，如图11所示，所述装置包括：

[0199] 目标视频帧获取模块1110，用于获取目标视频的目标视频帧；

[0200] 目标图像属性特征确定模块1120，用于对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征；

[0201] 目标场景特征确定模块1130，用于对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征；

[0202] 目标融合特征确定模块1140，用于对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征；

[0203] 目标属性信息确定模块1150，用于根据所述目标融合特征，确定所述目标视频的目标属性信息。

[0204] 在一些实施例中，所述目标图像属性特征确定模块可以包括：

[0205] 信息确定单元，用于确定所述目标视频中至少两个目标对象各自的结构完整度和清晰度；

[0206] 目标对象确定单元，用于将结构完整度大于第一阈值且清晰度大于第二阈值的对象，确定为第一目标对象，将所述至少两个目标对象中除所述第一目标对象之外的对象确定为第二目标对象；

[0207] 第一目标图像属性特征提取单元，用于对所述第一目标对象进行对象特征提取，得到第一目标图像属性特征；

[0208] 第二目标图像属性特征提取单元，用于对所述第二目标对象进行对象特征提取，得到第二目标图像属性特征；

[0209] 目标图像属性特征确定单元，用于将所述第一目标图像属性特征以及所述第二目标图像属性特征作为所述目标图像属性特征。

[0210] 在一些实施例中，所述第二目标图像属性特征提取单元包括：

[0211] 目标自重构特征确定子单元，用于对所述第二目标对象进行自重构特征提取，得到目标自重构特征；

[0212] 目标描述特征确定子单元，用于对所述第二目标对象进行属性描述特征提取，得到目标描述特征；

[0213] 第二目标图像属性特征确定子单元，用于将所述目标自重构特征以及所述目标描述特征，作为所述第二目标图像属性特征。

[0214] 在一些实施例中，所述目标自重构特征确定子单元可以包括：

[0215] 目标自重构特征提取子单元，用于基于自重构特征提取模型，对所述第二目标对象进行自重构特征提取，得到所述目标自重构特征。

[0216] 在一些实施例中，所述装置还可以包括：

[0217] 网格图像划分模块，用于将样本视频帧划分成至少两个网格图像；

[0218] 图像处理模块，用于对至少一个网格图像进行图像处理，得到处理后视频帧；所述

图像处理包括网格图像的位置更换、网格图像中部分图像的遮挡处理中的至少一种；

[0219] 自重构特征确定模块，用于基于所述处理后视频帧，对第一预设模型进行自重构特征提取训练，得到自重构特征；

[0220] 重构视频帧确定模块，用于基于所述自重构特征，对第二预设模型进行图像重构训练，得到重构视频帧；

[0221] 训练模块，用于在训练过程中，不断调整第一预设模型的第一模型参数以及第二预设模型的第二模型参数，直至所述第二预设模型输出的重构视频帧与所述样本视频帧相匹配；

[0222] 模型确定模块，用于将当前第一模型参数所对应的第一预设模型，作为所述自重构特征提取模型；所述当前第一模型参数为所述第一预设模型，在所述第二预设模型输出的重构视频帧与所述样本视频帧相匹配时的模型参数。

[0223] 在一些实施例中，所述装置还可以包括：

[0224] 文本获取模块，用于获取所述目标视频对应的目标音频以及目标文本；

[0225] 目标音频属性特征提取模块，用于对所述目标音频进行对象特征提取，得到所述目标对象的目标音频属性特征；

[0226] 目标文本属性特征提取模块，用于对所述目标文本进行对象特征提取，得到所述目标对象的目标文本属性特征。

[0227] 在一些实施例中，所述目标融合特征确定模块可以包括：

[0228] 特征融合单元，用于对所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征进行融合处理，得到所述目标融合特征。

[0229] 在一些实施例中，所述特征融合单元可以包括：

[0230] 待融合特征确定子单元，用于基于所述目标图像属性特征、所述目标场景特征、所述目标音频属性特征以及目标文本属性特征，确定至少两个待融合特征；

[0231] 特征融合子单元，用于对所述至少两个待融合特征进行融合，得到所述目标融合特征。

[0232] 在一些实施例中，所述目标图像属性特征确定模块可以包括：

[0233] 对象特征提取单元，用于对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标原料特征、目标命名特征以及目标类别特征；

[0234] 在一些实施例中，所述目标融合特征确定模块包括：

[0235] 目标融合特征确定单元，用于对所述目标原料特征、所述目标命名特征、所述目标类别特征以及所述目标场景特征进行融合处理，得到所述目标融合特征；

[0236] 所述目标属性信息确定模块包括：

[0237] 目标属性信息确定单元，用于根据所述目标融合特征，确定所述目标对象的目标原料信息、目标命名信息以及目标类别信息。

[0238] 在一些实施例中，所述目标场景特征确定模块可以包括：

[0239] 目标关联对象确定单元，用于基于所述目标视频帧，确定所述目标对象的至少两个目标关联对象；

[0240] 有向无环图构建单元，用于以所述至少两个目标关联对象各自对应的关联对象特征作为节点，构建有向无环图；所述有向无环图中的边表征所述边对应的两个关联对象特

征之间的相似度；

[0241] 目标场景特征确定单元，用于基于所述有向无环图进行场景特征提取，得到所述目标场景特征。

[0242] 所述的装置实施例中的装置与方法实施例基于同样地发明构思。

[0243] 本申请实施例提供了一种视频属性确定设备，该设备包括处理器和存储器，该存储器中存储有至少一条指令或至少一段程序，该至少一条指令或至少一段程序由该处理器加载并执行以实现如上述方法实施例所提供的视频属性确定方法。

[0244] 本申请的实施例还提供了一种计算机存储介质，所述存储介质可设置于终端之中以保存用于实现方法实施例中一种视频属性确定方法相关的至少一条指令或至少一段程序，该至少一条指令或至少一段程序由该处理器加载并执行以实现上述方法实施例提供的视频属性确定方法。

[0245] 本申请的实施例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可读存储介质中。计算机设备的处理器从计算机可读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以实现上述方法实施例提供的视频属性确定方法。

[0246] 可选地，在本申请实施例中，存储介质可以位于计算机网络的多个网络服务器中的至少一个网络服务器。可选地，在本实施例中，上述存储介质可以包括但不限于：U盘、只读存储器(ROM，Read-Only Memory)、随机存取存储器(RAM，Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0247] 本申请实施例所述存储器可用于存储软件程序以及模块，处理器通过运行存储在存储器的软件程序以及模块，从而执行各种功能应用以及数据处理。存储器可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、功能所需的应用程序等；存储数据区可存储根据所述设备的使用所创建的数据等。此外，存储器可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。相应地，存储器还可以包括存储器控制器，以提供处理器对存储器的访问。

[0248] 本申请实施例所提供的视频属性确定方法实施例可以在移动终端、计算机终端、服务器或者类似的运算装置中执行。以运行在服务器上为例，图12是本申请实施例提供的一种视频属性确定方法的服务器的硬件结构框图。如图12所示，该服务器1200可因配置或性能不同而产生比较大的差异，可以包括一个或一个以上中央处理器(Central Processing Units，CPU)1210(中央处理器1210可以包括但不限于微处理器MCU或可编程逻辑器件FPGA等的处理装置)、用于存储数据的存储器1230，一个或一个以上存储应用程序1223或数据1222的存储介质1220(例如一个或一个以上海量存储设备)。其中，存储器1230和存储介质1220可以是短暂存储或持久存储。存储在存储介质1220的程序可以包括一个或一个以上模块，每个模块可以包括对服务器中的一系列指令操作。更进一步地，中央处理器1210可以设置为与存储介质1220通信，在服务器1200上执行存储介质1220中的一系列指令操作。服务器1200还可以包括一个或一个以上电源1260，一个或一个以上有线或无线网络接口1250，一个或一个以上输入输出接口1240，和/或，一个或一个以上操作系统1221，例如Windows ServerTM，Mac OS XTM，UnixTM，LinuxTM，FreeBSDTM等等。

[0249] 输入输出接口1240可以用于经由一个网络接收或者发送数据。上述的网络具体实例可包括服务器1200的通信供应商提供的无线网络。在一个实例中，输入输出接口1240包括一个网络适配器(Network Interface Controller,NIC)，其可通过基站与其他网络设备相连从而可与互联网进行通讯。在一个实例中，输入输出接口1240可以为射频(Radio Frequency,RF)模块，其用于通过无线方式与互联网进行通讯。

[0250] 本领域普通技术人员可以理解，图12所示的结构仅为示意，其并不对上述电子装置的结构造成限定。例如，服务器1200还可包括比图12中所示更多或者更少的组件，或者具有与图12所示不同的配置。

[0251] 由上述本申请提供的视频属性确定方法、装置、设备或存储介质的实施例可见，本申请获取目标视频的目标视频帧；对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征；对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征；对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征；根据所述目标融合特征，确定所述目标视频的目标属性信息。本申请在确定视频属性信息过程中，融入了视频的场景特征，通过目标图像属性特征以及目标场景特征确定出融合特征，提高了确定的属性信息的准确率。

[0252] 需要说明的是：上述本申请实施例先后顺序仅仅为了描述，不代表实施例的优劣。且上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

[0253] 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于装置、设备、存储介质实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0254] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

[0255] 以上所述仅为本申请的较佳实施例，并不用以限制本申请，凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

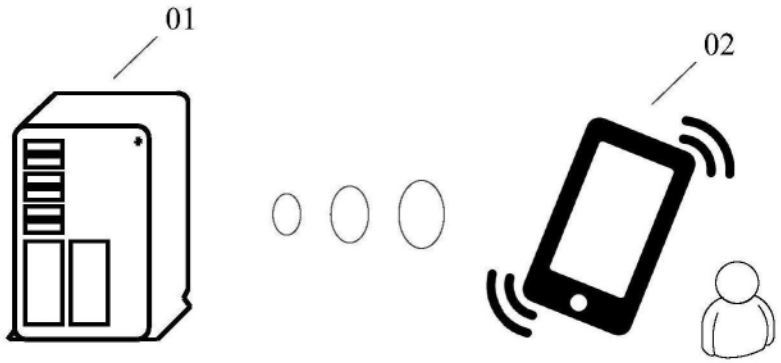

01
02

图1

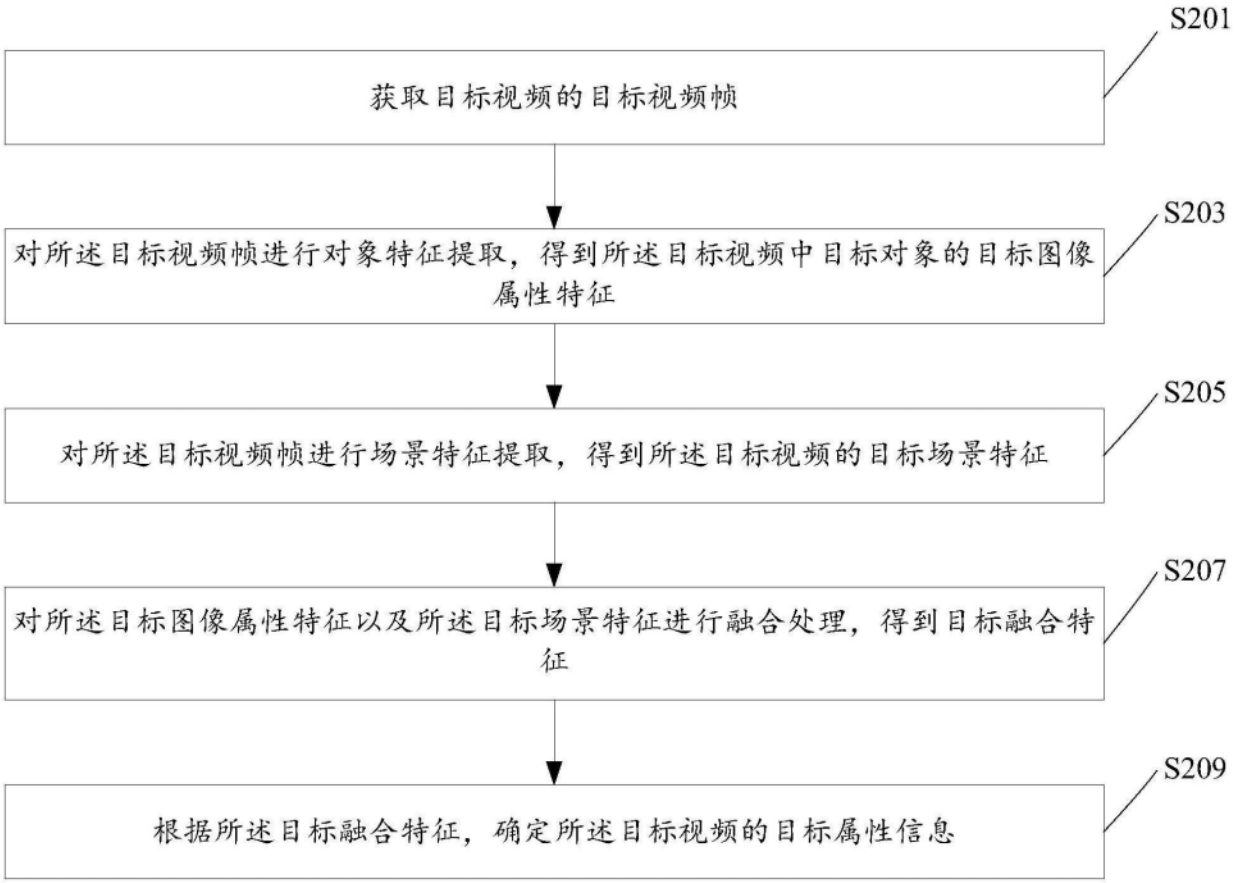

S201
获取目标视频的目标视频帧
S203
对所述目标视频帧进行对象特征提取，得到所述目标视频中目标对象的目标图像属性特征
S205
对所述目标视频帧进行场景特征提取，得到所述目标视频的目标场景特征
S207
对所述目标图像属性特征以及所述目标场景特征进行融合处理，得到目标融合特征
S209
根据所述目标融合特征，确定所述目标视频的目标属性信息

图2

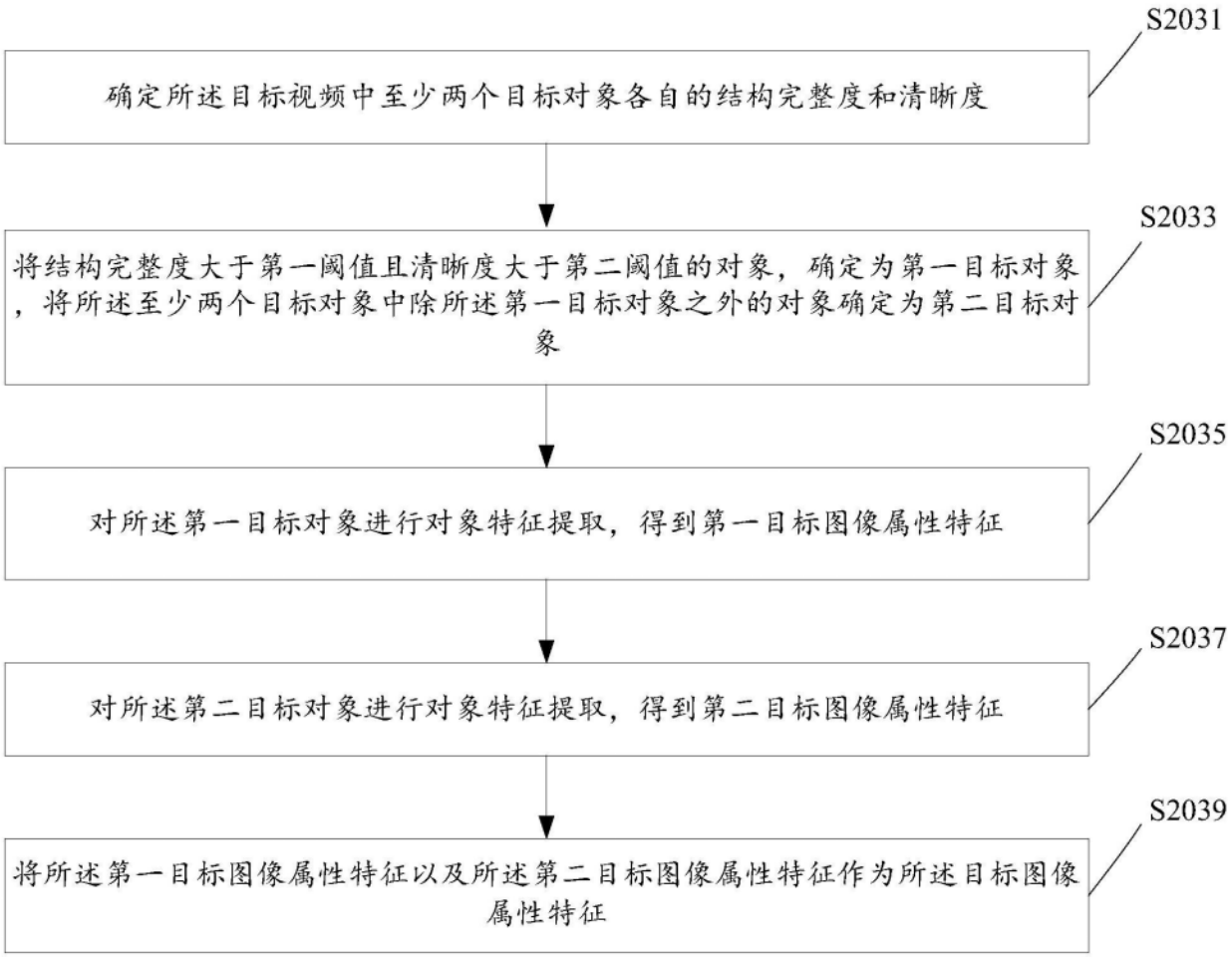

S2031
确定所述目标视频中至少两个目标对象各自的结构完整度和清晰度
S2033
将结构完整度大于第一阈值且清晰度大于第二阈值的对象，确定为第一目标对象，将所述至少两个目标对象中除所述第一目标对象之外的对象确定为第二目标对象
S2035
对所述第一目标对象进行对象特征提取，得到第一目标图像属性特征
S2037
对所述第二目标对象进行对象特征提取，得到第二目标图像属性特征
S2039
将所述第一目标图像属性特征以及所述第二目标图像属性特征作为所述目标图像属性特征

图3

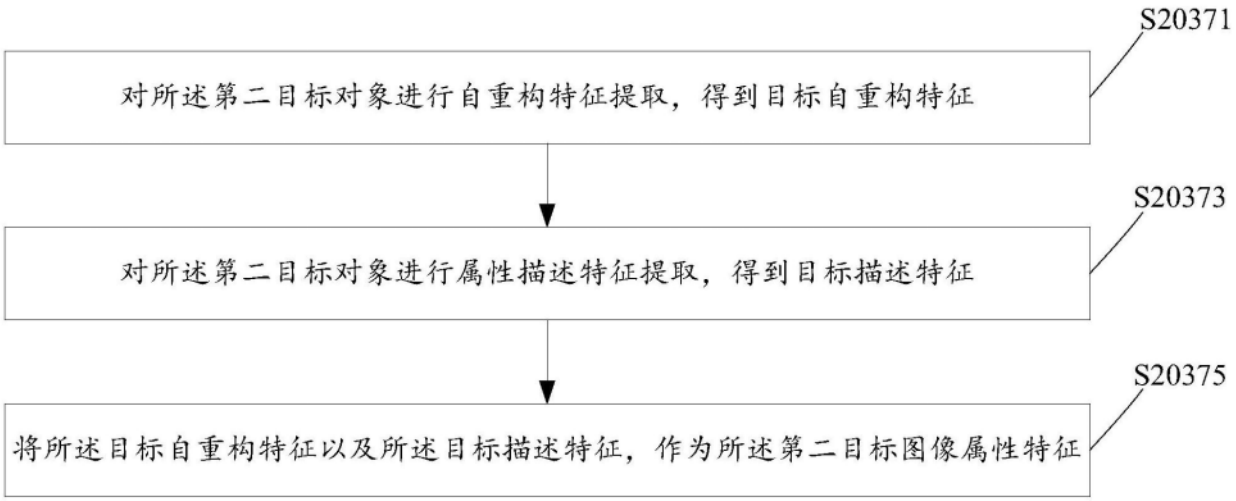

S20371
对所述第二目标对象进行自重构特征提取，得到目标自重构特征
S20373
对所述第二目标对象进行属性描述特征提取，得到目标描述特征
S20375
将所述目标自重构特征以及所述目标描述特征，作为所述第二目标图像属性特征

图4

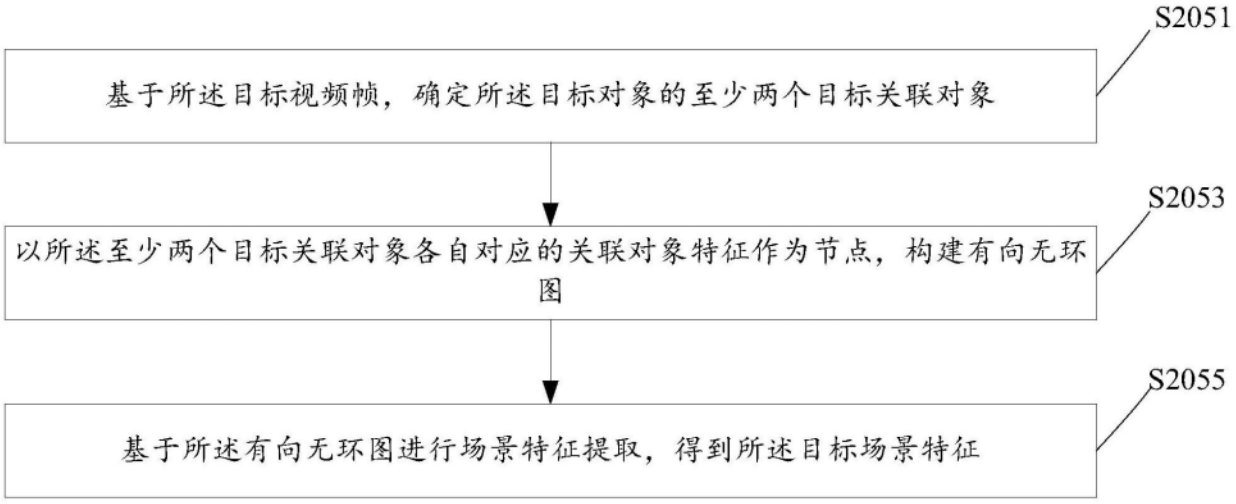

S2051
基于所述目标视频帧，确定所述目标对象的至少两个目标关联对象
S2053
以所述至少两个目标关联对象各自对应的关联对象特征作为节点，构建有向无环图
S2055
基于所述有向无环图进行场景特征提取，得到所述目标场景特征

图5

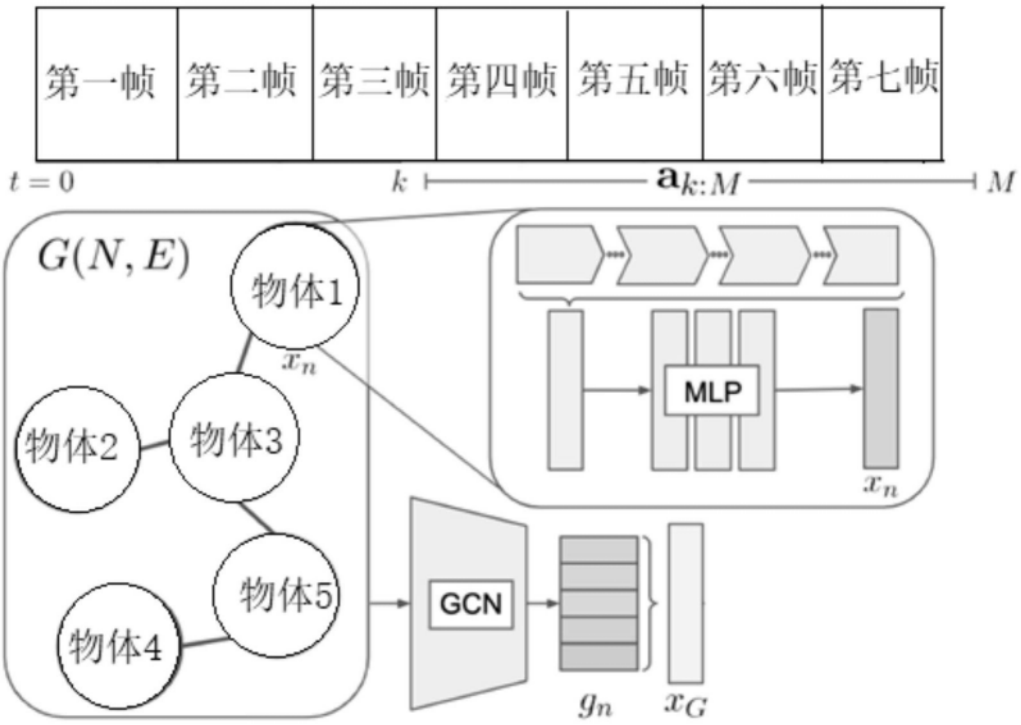

第一帧
第二帧
第三帧
第四帧
第五帧
第六帧
第七帧
$t=0$
k
$\mathbf{a}_{k:M}$
M
$G(N,E)$
物体1
物体2
物体3
物体4
物体5
x_n
MLP
x_n
GCN
g_n
x_G

图6

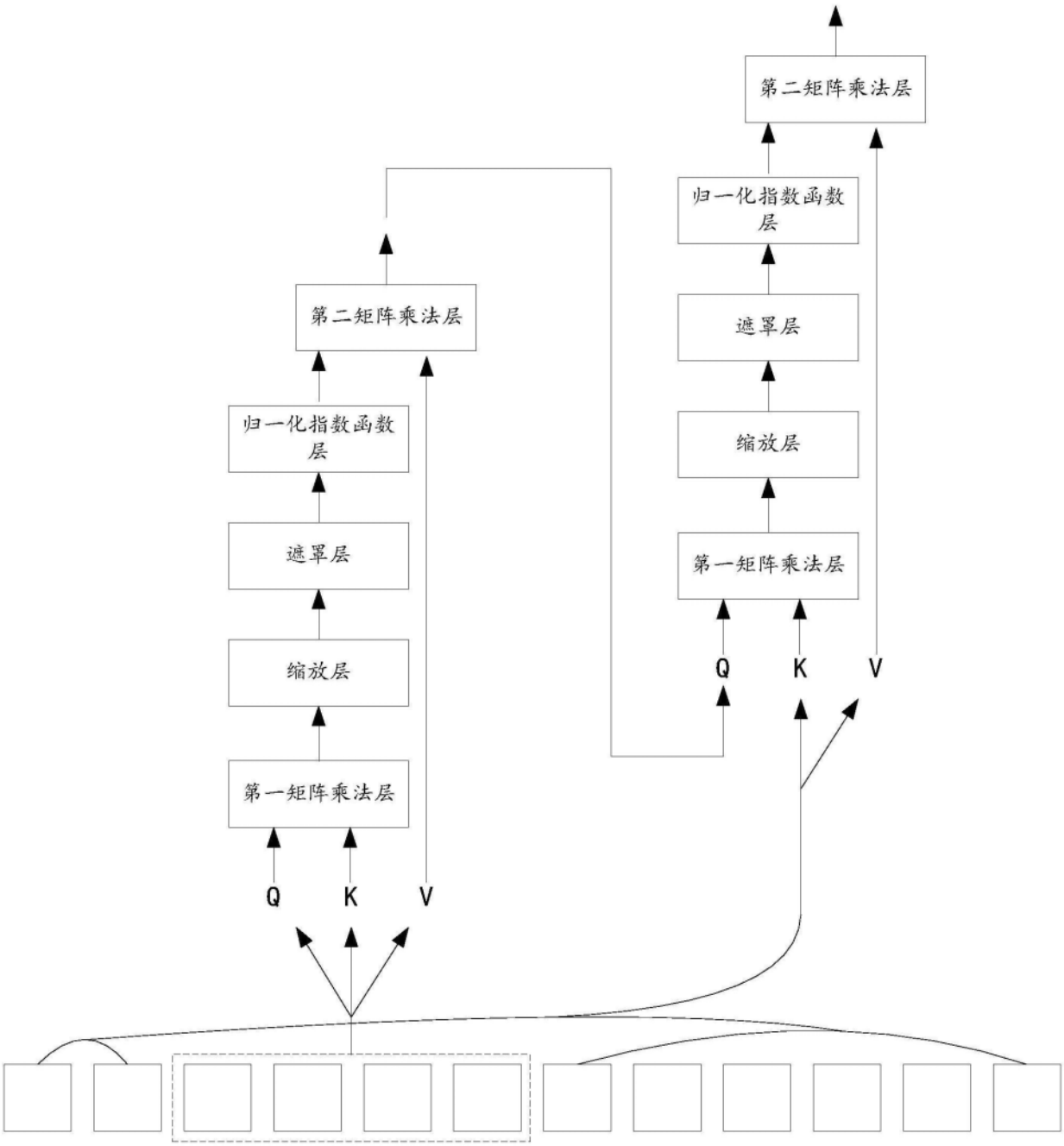

第二矩阵乘法层
归一化指数函数层
遮罩层
缩放层
第一矩阵乘法层
Q
K
V
第二矩阵乘法层
归一化指数函数层
遮罩层
缩放层
第一矩阵乘法层
Q
K
V

图7

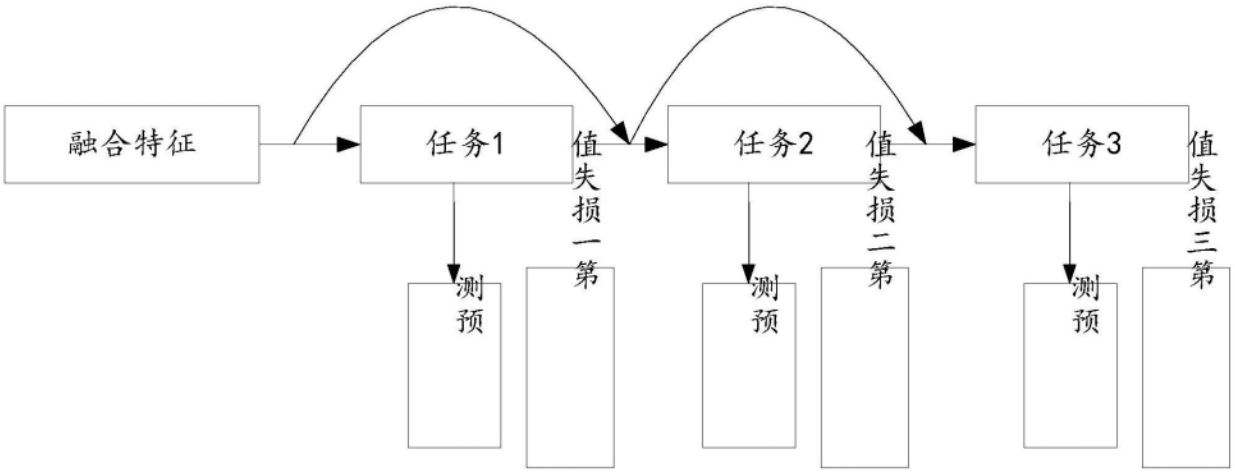

融合特征
任务1
任务2
任务3
值失损一第
值失损二第
值失损三第
测预
测预
测预

图8

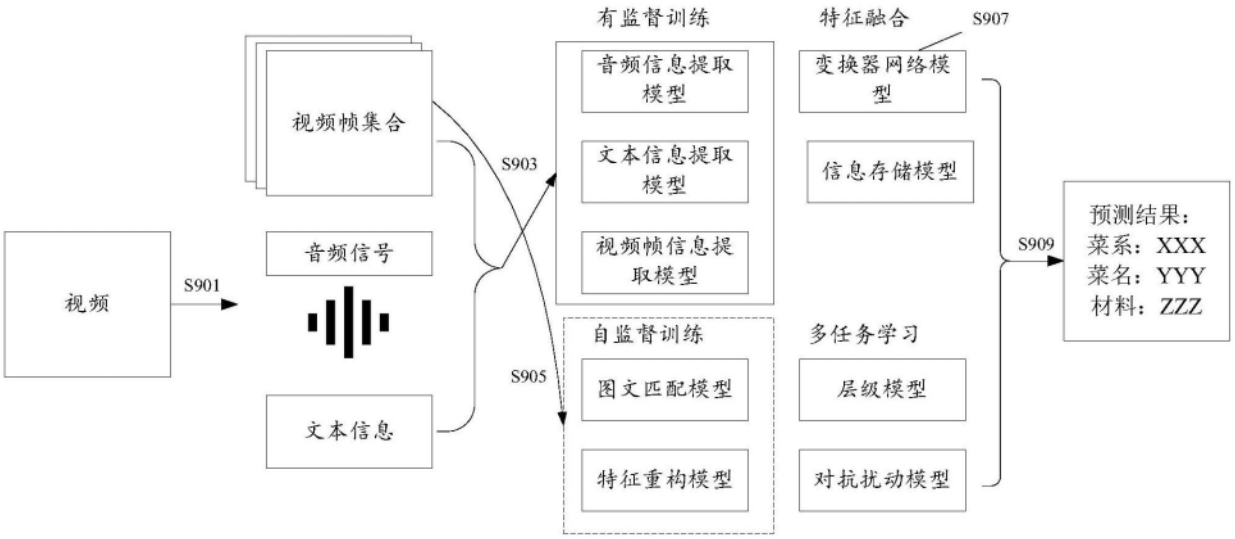

视频
S901
视频帧集合
音频信号
文本信息
S903
S905
有监督训练
音频信息提取模型
文本信息提取模型
视频帧信息提取模型
自监督训练
图文匹配模型
特征重构模型
特征融合
S907
变换器网络模型
信息存储模型
多任务学习
层级模型
对抗扰动模型
S909
预测结果：
菜系：XXX
菜名：YYY
材料：ZZZ

图9

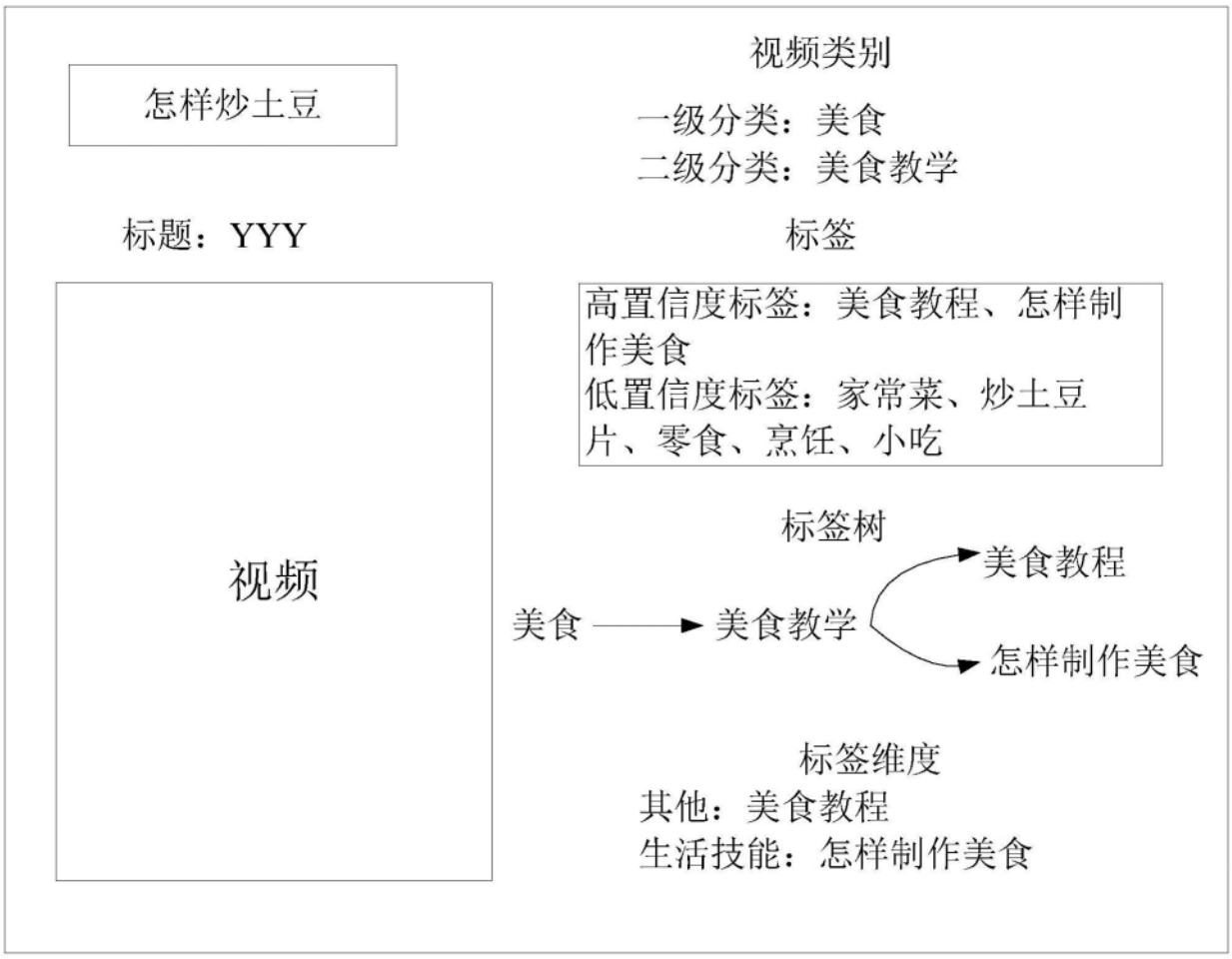

怎样炒土豆
视频类别
一级分类：美食
二级分类：美食教学
标题：YYY
标签
高置信度标签：美食教程、怎样制作美食
低置信度标签：家常菜、炒土豆片、零食、烹饪、小吃
视频
标签树
美食
美食教学
美食教程
怎样制作美食
标签维度
其他：美食教程
生活技能：怎样制作美食

图10

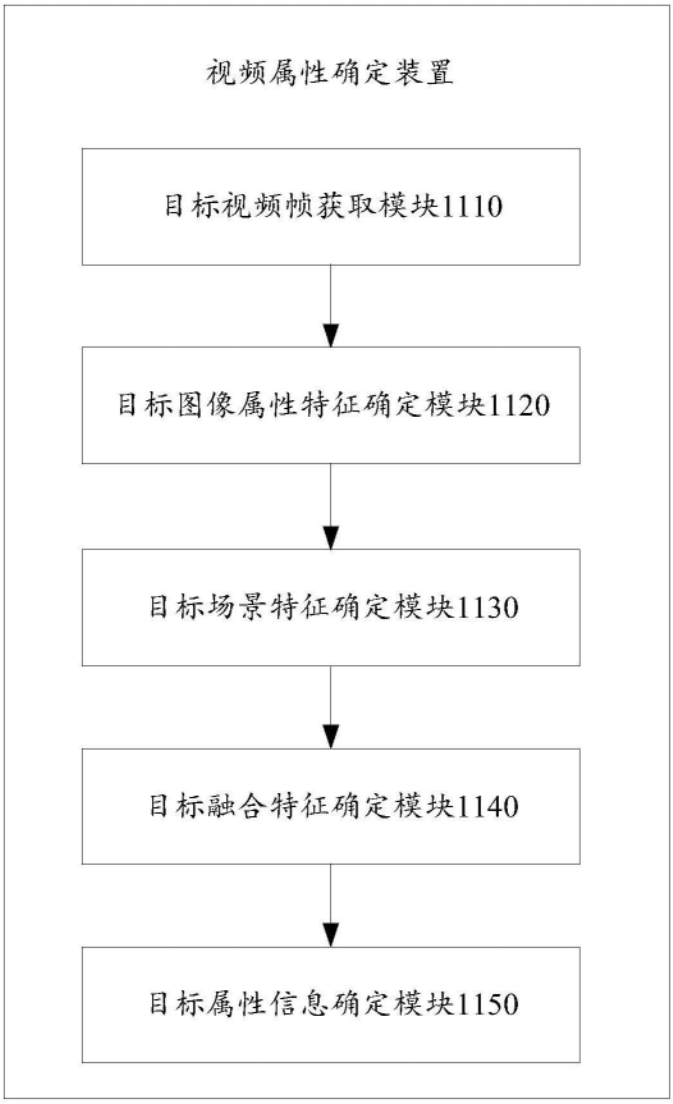

视频属性确定装置
目标视频帧获取模块1110
目标图像属性特征确定模块1120
目标场景特征确定模块1130
目标融合特征确定模块1140
目标属性信息确定模块1150

图11

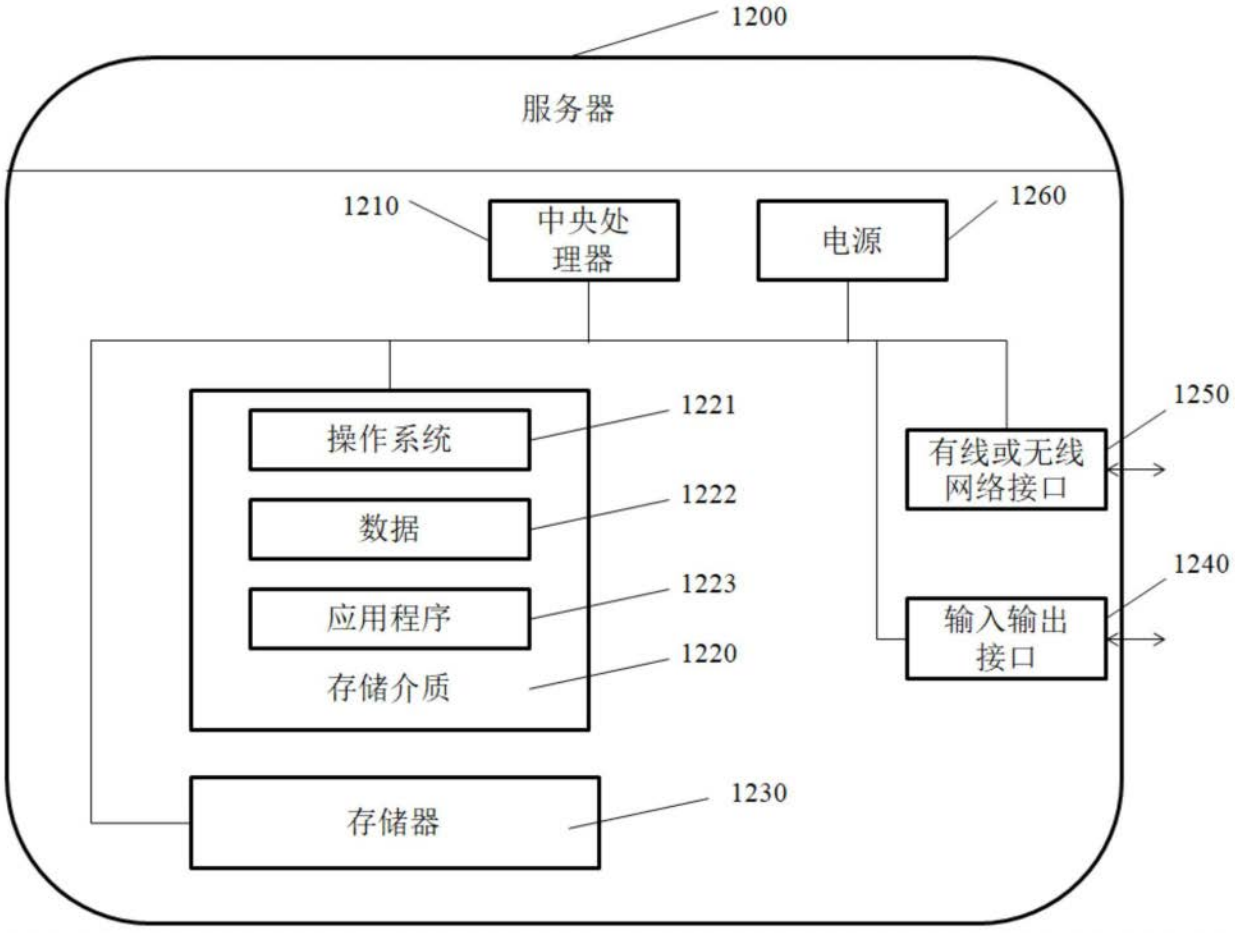

1200
服务器
1210
中央处理器
电源
1260
操作系统
1221
数据
1222
应用程序
1223
存储介质
1220
存储器
1230
有线或无线网络接口
1250
输入输出接口
1240

图12