

## (12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织  
国际局

(43) 国际公布日  
2023 年 1 月 5 日 (05.01.2023)



(10) 国际公布号  
**WO 2023/273628 A1**

(51) 国际专利分类号:  
**G06K 9/00** (2022.01)

(21) 国际申请号: PCT/CN2022/092377

(22) 国际申请日: 2022 年 5 月 12 日 (12.05.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:  
202110731049.4 2021年6月30日 (30.06.2021) CN

(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

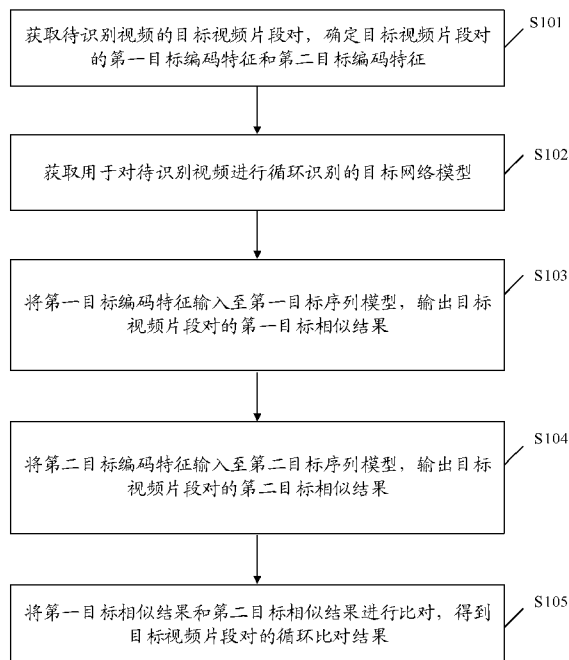
(72) 发明人: 郭卉 (GUO, Hui); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(74) 代理人: 深圳市深佳知识产权代理事务所(普通合伙) (SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦 B 座 18C2、18D、18E、18E2, Guangdong 518001 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN,

(54) **Title:** VIDEO LOOP RECOGNITION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种视频循环识别方法、装置、计算机设备及存储介质



S101 ACQUIRE A TARGET VIDEO CLIP PAIR IN A VIDEO TO UNDERGO RECOGNITION, AND DETERMINE A FIRST TARGET ENCODING FEATURE AND A SECOND TARGET ENCODING FEATURE OF THE TARGET VIDEO CLIP PAIR  
S102 ACQUIRE A TARGET NETWORK MODEL USED FOR PERFORMING LOOP RECOGNITION ON THE VIDEO TO UNDERGO RECOGNITION  
S103 INPUT THE FIRST TARGET ENCODING FEATURE INTO A FIRST TARGET SEQUENCE MODEL, AND OUTPUT A FIRST TARGET SIMILARITY RESULT FOR THE TARGET VIDEO CLIP PAIR  
S104 INPUT THE SECOND TARGET ENCODING FEATURE INTO A SECOND TARGET SEQUENCE MODEL ASSOCIATED, AND OUTPUT A SECOND TARGET SIMILARITY RESULT FOR THE TARGET VIDEO CLIP PAIR  
S105 COMPARE THE FIRST TARGET SIMILARITY RESULT WITH THE SECOND TARGET SIMILARITY RESULT, AND OBTAIN A CYCLIC COMPARISON RESULT FOR THE TARGET VIDEO CLIP PAIR

图 3

(57) **Abstract:** Embodiments of the present application disclose a video loop recognition method and apparatus, a computer device, and a storage medium. The method comprises: acquiring a target video clip pair in a video to undergo recognition, and determining a first target encoding feature and a second target encoding feature; acquiring a target network model; inputting the first target encoding feature into a first target sequence model associated with first modal information, and outputting a first target similarity result for the target video clip pair; inputting the second target encoding feature into a second target sequence model associated with second modal



MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE,  
PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE,  
SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,  
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区  
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,  
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,  
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,  
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,  
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,  
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,  
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

information, and outputting a second target similarity result for the target video clip pair; and comparing the first target similarity result with the second target similarity result, and obtaining a cyclic comparison result for the target video clip pair. By adopting the embodiments of the present application, the accuracy of video loop recognition can be improved.

(57) 摘要: 本申请实施例公开了一种视频循环识别方法、装置、计算机设备及存储介质, 包括: 获取待识别视频的目标视频片段对, 确定第一目标编码特征和第二目标编码特征; 获取目标网络模型; 将第一目标编码特征输入至与第一模态信息相关联的第一目标序列模型, 输出目标视频片段对的第一目标相似结果; 将第二目标编码特征输入至与第二模态信息相关联的第二目标序列模型, 输出目标视频片段对的第二目标相似结果; 将第一目标相似结果和第二目标相似结果进行比对, 得到目标视频片段对的循环比对结果。采用本申请实施例, 可以提高视频循环识别的准确率。

## 一种视频循环识别方法、装置、计算机设备及存储介质

本申请要求于2021年06月30日提交中国专利局、申请号为202110731049.4、申请名称为“一种视频循环识别方法、装置、计算机设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

### 5 技术领域

本申请涉及计算机技术领域，尤其涉及视频循环识别。

### 背景技术

10 轮播识别是指对不断重复的某一视频片段的视频循环识别，以提高视频质量。若将现有的图像识别技术应用于视频循环识别的过程，则具有图像识别功能的计算机设备在提取到某一待识别视频的每一视频帧的图像特征时，可以将每一视频帧的图像特征分别与后续视频帧的图像特征进行逐帧匹配，以根据统计到的重复帧的数量比例来判断重复的视频片段。

然而，一旦基于统计到的重复帧的数量比例来进行视频循环识别，这将导致该计算机设备会误将关系错乱的重复视频帧判决为循环视频帧，以至于不支持视频剪辑等应用场景。

### 15 发明内容

本申请实施例提供一种视频循环识别方法、装置、计算机设备及存储介质，可以提高视频循环识别的准确率。

本申请实施例一方面提供一种视频循环识别方法，包括：

20 获取待识别视频的目标视频片段对，确定目标视频片段对的第一目标编码特征和第二目标编码特征；第一目标编码特征所对应的第一模态信息不同于第二目标编码特征所对应的第二模态信息；

获取用于对待识别视频进行循环识别的目标网络模型；目标网络模型包括与第一模态信息相关联的第一目标序列模型和与第二模态信息相关联的第二目标序列模型；

25 将第一目标编码特征输入至第一目标序列模型，输出目标视频片段对的第一目标相似结果；

将第二目标编码特征输入至第二目标序列模型，输出目标视频片段对的第二目标相似结果；

将第一目标相似结果和第二目标相似结果进行比对，得到目标视频片段对的循环比对结果；循环比对结果用于指示待识别视频的视频类型。

30 本申请实施例一方面提供一种视频循环识别装置，包括：

目标编码特征获取模块，用于获取待识别视频的目标视频片段对，确定目标视频片段对的第一目标编码特征和第二目标编码特征；第一目标编码特征所对应的第一模态信息不同于第二目标编码特征所对应的第二模态信息；

35 目标网络模型获取模块，用于获取用于对待识别视频进行循环识别的目标网络模型；目标网络模型包括与第一模态信息相关联的第一目标序列模型和与第二模态信息相关联的第二目标序列模型；

第一目标相似结果确定模块，用于将第一目标编码特征输入至第一目标序列模型，输

—2—

出目标视频片段对的第一目标相似结果；

第二目标相似结果确定模块，用于将第二目标编码特征输入至第二目标序列模型，输出目标视频片段对的第二目标相似结果；

目标相似结果比对模块，用于将第一目标相似结果和第二目标相似结果进行比对，得到目标视频片段对的循环比对结果；循环比对结果用于指示待识别视频的视频类型。

本申请实施例一方面提供了一种计算机设备，包括：处理器和存储器；

处理器与存储器相连，其中，存储器用于存储计算机程序，计算机程序被处理器执行时，使得该计算机设备执行以上方面所述的方法。

本申请实施例一方面提供了一种计算机可读存储介质，计算机可读存储介质存储有计算机程序，该计算机程序适于由处理器加载并执行，以使得具有该处理器的计算机设备执行以上方面所述的方法。

本申请实施例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上方面所述的方法。

在本申请实施例中，该计算机设备在对待识别视频进行循环识别时，可以事先获取待识别视频的目标视频片段对，并确定目标视频片段对的第一目标编码特征以及第二目标编码特征。其中，这里的第一目标编码特征对应的第一模态信息(例如，视频模态信息)不同于第二目标编码特征对应的第二模态信息(例如，音频模态信息)。在获取到用于对待识别视频进行循环的目标网络模型时，该计算机设备可以充分利用该目标视频片段对的视频信息，即不仅可以通过目标网络模型中的第一目标序列模型，确定该目标视频片段对在第一模态信息下的第一目标相似结果，还可以通过目标网络模型中的第二目标序列模型，确定该目标视频片段对在第二模态信息下的第二目标相似结果。进一步地，该计算机设备可以将该第一目标相似结果和第二目标相似结果进行比对，以更加准确的得到该目标视频片段对的循环比对结果，进而提高了视频循环识别的准确率。

## 附图说明

图1是本申请实施例提供的一种网络架构的结构示意图；

图2是本申请实施例提供的一种基于多模态信息的视频对循环识别模型的模型架构图；

图3是本申请实施例提供的一种视频循环识别方法的流程示意图；

图4是本申请实施例提供的一种残差块的结构示意图；

图5是本申请实施例提供的一种目标序列模型的结构示意图；

图6是本申请实施例提供的一种深度神经网络的网络示意图；

图7是本申请实施例提供的一种视频循环识别方法的流程示意图；

图8是本申请实施例提供的一种数据交互的场景示意图；

图9是本申请实施例提供的一种视频循环识别装置的结构示意图；

图10是本申请实施例提供的一种计算机设备的示意图。

## 具体实施方式

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。

5 请参见图1，图1是本申请实施例提供的一种网络架构的结构示意图。如图1所示，该网络架构可以包括服务器10和用户终端集群。该用户终端集群可以包括一个或者多个用户终端。如图1所示，该用户终端集群具体可以包括用户终端100a、用户终端100b、用户终端100c、...、用户终端100n。如图1所示，用户终端100a、用户终端100b、用户终端100c、...、用户终端100n可以分别与上述服务器10进行网络连接，以便于每个用户终端可以通过该网络  
10 网络连接与服务器10进行数据交互。其中，这里的网络连接不限定连接方式，可以通过有线通信方式进行直接或间接地连接，也可以通过无线通信方式进行直接或间接地连接，还可以通过其他方式，本申请在此不做限制。

其中，该用户终端集群中的每个用户终端均可以包括：智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表、车载终端、智能电视等具有视频循环识别功能的智能终端。应当理解，如图1所示的用户终端集群中的每个用户终端均可以安装有目标应用(即应用客户端)，当该应用客户端运行于各用户终端中时，可以分别与上述图1所示的服务器10之间进行数据交互。其中，该应用客户端可以包括社交客户端、多媒体客户端(例如，  
15 视频客户端)、娱乐客户端(例如，游戏客户端)、教育客户端、直播客户端等应用客户端。其中，该应用客户端可以为独立的客户端，也可以为集成在某客户端(例如，社交客户端、教育客户端以及多媒体客户端等)中的嵌入式子客户端，在此不做限定。

如图1所示，本申请实施例中的服务器10可以为该应用客户端对应的服务器。该服务器10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，还可以是提供云计算服务的云服务器。其中，本申请实施例将不对用户终端和服务器的数量进行限制。

25 以便于理解，本申请实施例可以在图1所示的多个用户终端中选择一个用户终端作为目标用户终端。例如，本申请实施例可以将图1所示的用户终端100a作为目标用户终端，该目标用户终端中可以集成有目标应用(即应用客户端)。此时，该目标用户终端可以通过该应用客户端对应的业务数据平台与服务器10之间实现数据交互。其中，这里的目标应用可以运行有已经训练完成的目标网络模型，该目标网络模型是一种基于多模态信息的视频对循环识别模型，该目标网络模型可以为用于对待识别视频进行循环识别的神经网络模型，通过  
30 该目标网络模型的循环比对结果可以预测待识别视频的视频类型，即判断该待识别视频是否为循环视频，以便后续处理(例如，审核拦截等)。

应当理解，本申请实施例提出了一种基于目标网络模型的视频循环识别方法，该方法可以涉及人工智能领域中的机器学习方向。可以理解的是，所谓人工智能(Artificial Intelligence，简称AI)是利用数字计算机或者数字计算机控制的计算模拟、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  
35 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。

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、  
5 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、自动驾驶、智慧交通等几大方向。

其中,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  
10 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。

其中,由于属于循环视频类型的视频在视频播放过程中的视频质量较差,为了有效提  
15 高视频质量,因此在视频审核过程中,需要对视频进行循环识别。然而,相关技术中的视频循环识别方案在不同视频帧的比对过程中,仅关注参与比对的两个图像帧之间的图像特征是否匹配,而并未充分利用整个视频信息,从而降低了视频循环识别的准确率。

可以理解的是,视频信息是非常丰富的,可以包括视频模态信息、音频模态信息、台词文本模态信息、视频标题模态信息以及封面模态信息等等多种模态信息,每种模态信息  
20 所刻画的视频维度均不同。因此,在对某一视频进行循环识别时,本申请实施例提供了一种基于多模态信息的视频对循环识别模型(即目标网络模型)。

可以理解的是,为了充分利用视频的多模态信息,本申请实施例所涉及的目标网络模型可以包括多个与某种模态信息相关联的目标序列模型(即序列相似度学习模型),且一个目标序列模型用于确定目标视频片段对在对应模态信息下的目标相似结果。比如,该目标网络模型中可以包括与视频模态信息相关联的目标序列模型(例如,目标序列模型1)、与音频  
25 模态信息相关联的目标序列模型(例如,目标序列模型2)、与台词文本模态信息相关联的目标序列模型(例如,目标序列模型3)、与视频标题模态信息相关联的目标序列模型(例如,目标序列模型4)、与封面模态信息相关联的目标序列模型(例如,目标序列模型5)等。

比如,具备视频循环识别功能的计算机设备(例如,图1所示的用户终端100a或者图1所示的服务器10)可以运行有上述目标网络模型,该目标网络模型中的目标序列模型可以以2  
30 个为例,具体可以包括与第一模态信息(例如,视频模态信息)相关联的第一目标序列模型和与第二模态信息(例如,音频模态信息)相关联的第二目标序列模型。可以理解的是,该计算机设备在对待识别视频进行循环识别时,可以事先获取待识别视频的目标视频片段对,并确定目标视频片段对的第一目标编码特征以及第二目标编码特征。在获取到目标网络模型  
35 时,该计算机设备可以充分利用该目标视频片段对的视频信息,即不仅可以通过目标网络模型中的第一目标序列模型,确定该目标视频片段对在第一模态信息下的第一目标相似结果,还可以通过目标网络模型中的第二目标序列模型,确定该目标视频片段对在第二模态

—5—

信息下的第二目标相似结果。进一步地，该计算机设备可以将该第一目标相似结果和第二目标相似结果进行比对，以更加准确的得到该目标视频片段对的循环比对结果，进而提高了视频循环识别的准确率。

5 为了便于理解，进一步地，请参见图2，图2是本申请实施例提供的一种基于多模态信息的视频对循环识别模型的模型架构图。如图2所示，本申请实施例中的计算机设备可以为具备视频循环识别功能的计算机设备，该计算机设备可以为上述图1所示的用户终端集群中的任意一个用户终端，例如，用户终端100a，该计算机设备也可以为上述图1所示的服务器10，这里将不对计算机设备进行限定。

10 其中，本申请实施例中的计算机设备在获取到图2所示的视频2V时，可以将该视频2V作为待识别视频，以对该视频2V进行视频审核。该视频2V可以为与该计算机设备具有网络连接关系的某一用户终端所发送的视频。应当理解，该计算机设备在获取到该视频2V时，可以对该视频2V进行分段处理，以得到N个视频片段，进而可以根据这N个视频片段，确定视频2V的目标视频片段对。其中，这里的N为正整数。

15 比如，若图2所示的视频2V的视频时长为20秒，则该计算机设备可以选取一个切割间隔时长(例如，5秒)，将该视频2V均匀的切分为4个视频片段，如图2所示，这4个视频片段按时间轴的顺序排列，可以依次包括视频片段P<sub>1</sub>、视频片段P<sub>2</sub>、视频片段P<sub>3</sub>以及视频片段P<sub>4</sub>。进一步地，该计算机设备可以从这4个视频片段中随机选取两个视频片段，以组成该视频2V的目标视频片段对。其中，该视频2V的目标视频片段对具体可以包括：由视频片段P<sub>1</sub>和20 视频片段P<sub>2</sub>所组成的目标视频片段对(例如，视频片段对2a)、由视频片段P<sub>1</sub>和视频片段P<sub>3</sub>所组成的目标视频片段对(例如，视频片段对2b)、由视频片段P<sub>1</sub>和视频片段P<sub>4</sub>所组成的目标视频片段对(例如，视频片段对2c)、由视频片段P<sub>2</sub>和25 视频片段P<sub>3</sub>所组成的目标视频片段对(例如，视频片段对2d)、由视频片段P<sub>2</sub>和视频片段P<sub>4</sub>所组成的目标视频片段对(例如，视频片段对2e)、由视频片段P<sub>3</sub>和30 视频片段P<sub>4</sub>所组成的目标视频片段对(例如，视频片段对2f)。

25 为了提高视频循环识别的准确率，本申请实施例中的计算机设备需要充分利用目标视频片段对的多模态信息，因此，该计算机设备可以确定每个目标视频片段对的目标编码特征。其中，这里的目标编码特征可以包括每个目标视频片段对在第一模态信息下的编码特征(即第一目标编码特征)、以及在第二模态信息下的编码特征(即第二目标编码特征)。比如，35 这里的第一模态信息可以为视频模态信息，这里的第二模态信息可以为不同于第一模态信息的音频模态信息。如图2所示，视频片段对2a的第一目标编码特征可以为编码特征21a，且该视频片段对2a的第二目标编码特征可以为编码特征22a。视频片段对2b的第一目标编码特征可以为编码特征21b，且该视频片段对2b的第二目标编码特征可以为编码特征22b。以此类推，该视频片段对2f的第一目标编码特征可以为编码特征21f，且该视频片段对2f的第二目标编码特征可以为编码特征22f。

35 可以理解的是，该视频片段对2a的目标编码特征可以包括编码特征21a(例如，视频编码特征S<sub>1</sub>和视频编码特征S<sub>2</sub>)、以及编码特征22a(例如，音频编码特征Y<sub>1</sub>和音频编码特征Y<sub>2</sub>)。其中，视频编码特征S<sub>1</sub>是指视频片段对2a中的视频片段P<sub>1</sub>在第一模态信息下的编码特征，音频编码特征Y<sub>1</sub>是指视频片段对2a中的视频片段P<sub>1</sub>在第二模态信息下的编码特征；视频编

—6—

码特征 $S_2$ 是指视频片段对2a中的视频片段 $P_2$ 在第一模态信息下的编码特征，音频编码特征 $Y_2$ 是指视频片段对2a中的视频片段 $P_2$ 在第二模态信息下的编码特征。

进一步地，该计算机设备可以获取用于对待识别视频进行循环识别的目标网络模型(例如，图2所示的目标网络模型2000M)。其中，该目标网络模型2000M中可以包括与第一模态信息相关联的目标序列模型210M(即第一目标序列模型)、和与第二模态信息相关联的目标序列模型220M(即第二目标序列模型)。其中，本申请实施例可以将目标序列模型210M(即第一目标序列模型)中的序列表征学习层称之为第一序列表征学习层，且将第一目标序列模型中的相似度量层称之为第一相似度量层。本申请实施例可以将目标序列模型220M(即第二目标序列模型)中的序列表征学习层称之为第二序列表征学习层，且将第二目标序列模型中的相似度量层称之为第二相似度量层。

可以理解的是，该计算机设备可以将目标视频片段对的第一目标编码特征输入至目标序列模型210M，通过该第一序列表征学习层，对第一目标编码特征进行序列特征学习，且将序列特征学习到的第一目标学习特征输入至第一相似度量层，由该第一相似度量层输出目标视频片段对的第一目标相似结果(例如，图2所示的目标相似结果1G)。其中，目标相似结果1G中的数值“0”可以指示目标视频片段对在第一模态信息下不相似，目标相似结果1G中的数值“1”可以表示目标视频片段对在第一模态信息下相似。与此同时，该计算机设备可以将目标视频片段对的第二目标编码特征输入至目标序列模型220M，通过该第二序列表征学习层，对第二目标编码特征进行序列特征学习，且将序列特征学习到的第二目标学习特征输入至第二相似度量层，由该第二相似度量层输出目标视频片段对的第二目标相似结果(例如，图2所示的目标相似结果2G)。其中，目标相似结果2G中的数值“0”可以指示目标视频片段对在第二模态信息下不相似，目标相似结果2G中的数值“1”可以表示目标视频片段对在第二模态信息下相似。此时，该计算机设备可以将第一目标相似结果和第二目标相似结果进行比对，得到目标视频片段对的循环比对结果。

比如，针对视频片段对2a这一目标视频片段对而言，该计算机设备可以将该视频片段对2a的编码特征21a输入至图2所示的目标序列模型210M，通过该目标序列模型210M中的第一序列表征学习层和第一相似度量层，输出视频片段对2a的第一目标相似结果(例如，数值为“0”的相似结果)，这意味着该视频片段对2a在第一模态信息下不相似。与此同时，该计算机设备还可以将该视频片段对2a的编码特征22a输入至图2所示的目标序列模型220M，通过该目标序列模型220M中的第二序列表征学习层和第二相似度量层，可以输出视频片段对2a的第二目标相似结果(例如，数值为“1”的相似结果)，这意味着该视频片段对2a在第二模态信息下相似。此时，该计算机设备可以将视频片段对2a的第一目标相似结果和视频片段对2a的第二目标相似结果进行比对，从而可以得到视频片段对2a的循环比对结果(例如，数值为“0”的非循环视频结果)，这意味着该视频片段对2a中的两个视频片段并不重复。

又比如，针对视频片段对2d这一目标视频片段对而言，该计算机设备可以将该视频片段对2d的编码特征21d输入至图2所示的目标序列模型210M，通过该目标序列模型210M中的第一序列表征学习层和第一相似度量层，输出视频片段对2d的第一目标相似结果(例如，

-7-

数值为“1”的相似结果),这意味着该视频片段对2d在第一模态信息下相似。与此同时,该计算机设备还可以将该视频片段对2d的编码特征22d输入至图2所示的目标序列模型220M,通过该目标序列模型220M中的第二序列表征学习层和第二相似度度量层,可以输出视频片段对2d的第二目标相似结果(例如,数值为“1”的相似结果),这意味着该视频片段对2d在第二模态信息下相似。此时,该计算机设备可以将视频片段对2d的第一目标相似结果和

5 视频片段对2d的第二目标相似结果进行比对,从而可以得到视频片段对2d的循环比对结果(例如,数值为“1”的循环视频结果),这意味着该视频片段对2d中的两个视频片段互相重复。

其中,目标视频片段对的循环比对结果可以用于指示待识别视频的视频类型。可以理解的是,若图2所示的循环比对结果中不存在数值为“1”的循环视频结果,即全为数值为“0”的非循环视频结果,则该计算机设备可以将视频2V的视频类型确定为非循环视频类型。可选的,若图2所示的循环比对结果中存在数值为“1”的循环视频结果,则该计算机设备可以直接将视频2V的视频类型确定为循环视频类型。

由此可见,本申请实施例中的计算机设备在对视频2V进行循环识别时,可以充分利用视频2V的目标视频片段对的多模态信息,不仅需要通过目标网络模型2000M中的与第一模态信息相关联的目标序列模型210M,确定视频2V的目标视频片段对的第一相似结果,还需要通过目标网络模型2000M中的与第二模态信息相关联的目标序列模型220M,确定视频2V的目标视频片段对的第二相似结果,进而可以基于目标视频片段对的第一相似结果和第二相似结果,共同确定目标视频片段对的循环比对结果,以根据循环比对结果确定待识别视频的视频类型,进而提高了视频循环识别的准确率。

其中,具备视频循环识别功能的计算机设备通过一种基于多模态信息的视频对循环识别模型(即目标网络模型),确定目标视频片段对的循环比对结果的具体实现方式可以参见下述图3-图8所对应的实施例。

进一步地,请参见图3,图3是本申请实施例提供的一种视频循环识别方法的流程示意图。如图3所示,该方法可以由具备视频循环识别功能的计算机设备执行,该计算机设备可以为用户终端(例如,上述图1所示的用户终端集群中的任意一个用户终端,例如,具备模型应用功能的用户终端100a),也可以为服务器(例如,上述图1所示的服务器10),在此不做限定。为便于理解,本申请实施例以该方法由具备视频循环识别功能的服务器执行为例进行说明,该方法至少可以包括以下S101-S105:

30 S101,获取待识别视频的目标视频片段对,确定目标视频片段对的第一目标编码特征和第二目标编码特征。

其中,待识别视频为需要进行循环视频识别的视频,即需要识别其视频类型是否为循环视频类型。该待识别视频可以被划分为多个视频片段,例如,不同的视频片段涉及待识别视频的不同时间区间。基于多个视频片段可以组成至少一个视频片段对,一个视频片段对中包括多个视频片段中任意两个视频片段。本步骤中的目标视频片段对为该至少一个视频片段对中的一个。

具体地,该计算机设备在获取到待识别视频时,可以对待识别视频进行分段处理,以

得到N个视频片段。其中，这里的N为正整数。其中，这里的待识别视频可以为与计算机设备具有网络连接关系的用户终端所发送的，也可以为该计算机设备对应的用户通过该计算机设备的摄像组件所采集到的，还可以为该计算机设备对应的用户通过应用客户端所剪辑的，这里将不对待识别视频的获取方式进行限定。进一步地，该计算机设备可以从N个视频片段中获取视频片段 $P_i$ 和视频片段 $P_j$ ，且将视频片段 $P_i$ 和视频片段 $P_j$ 作为待识别视频的目标视频片段对；其中，这里的i和j均为小于或者等于N的正整数，且i不等于j。进一步地，该计算机设备可以对目标视频片段对中的每个视频片段分别进行第一特征提取，得到目标视频片段对的第一目标编码特征(即某一模态信息下的编码特征)。与此同时，该计算机设备还可以对目标视频片段对中的每个视频片段分别进行第二特征提取，得到目标视频片段对的第二目标编码特征(即另一模态信息下的编码特征)。

应当理解，该计算机设备通过目标网络模型对待识别视频进行循环识别的过程中可以包括视频准备过程(即分段处理和特征提取处理)。由于对待识别视频进行分段处理后所得到的视频片段的段长与后续输入至目标网络模型的输入特征相关，因此，该计算机设备需要根据实际应用情况并综合考虑目标网络模型的模型处理速率，合理选取用于对待识别视频进行分段处理的切割间隔时长。其中，可以理解的是，该计算机设备在对获取到的待识别视频进行分段处理时，可以先确定该待识别视频的视频时长，进而可以基于确定的视频时长，对获取到的待识别视频进行分段处理，以得到N个视频片段。

其中，该计算机设备可以获取视频时长对应的切割时长选择范围，进而可以从切割时长范围中，选取一个数值(stime)作为切割间隔时长，且根据切割间隔时长对待识别视频进行分段处理，从而得到N个视频片段。比如，若计算机设备获取到的待识别视频(例如，视频1)的视频时长为20秒，则该计算机设备可以获取该视频时长对应的切割时长选择范围(例如，[4,10])，进而可以在这一切割时长选择范围内选取一个数值(例如，5秒)作为切割间隔时长，进而可以根据切割间隔时长，对视频1进行分段处理，换言之，该计算机设备可以每隔5秒切分一段，从而得到4个具有5秒视频时长的视频片段。

可选的，该计算机设备还可以根据待识别视频的视频时长，先对待识别视频进行均匀分段处理，以将该待识别视频一分为二，从而得到两段具有相同时长的视频片段(例如，初始视频片段)，进而可以对每一个初始视频片段再进行均匀分段处理，直到分段处理后的初始视频片段的视频时长达到某一时长阈值(例如，10秒)时，结束对待识别视频的分段处理。比如，若该计算机设备获取到的待识别视频(例如，视频2)的视频时长为40秒，则该计算机设备可以先对该视频2进行分段处理，以得到两段具有20秒视频时长的初始视频片段(例如，初始视频片段a和初始视频片段b)，然后再对初始视频片段a进行分段处理，以得到两段具有10秒视频时长的视频片段(例如，视频片段 $P_1$ 和视频片段 $P_2$ )，且对初始视频片段b进行分段处理，以得到两段具有10秒视频时长的视频片段(例如，视频片段 $P_3$ 和视频片段 $P_4$ )。其中，该计算机设备对视频2进行分段处理后所得到的视频片段可以为4个具有10秒视频时长的视频片段。

进一步地，该计算机设备可以从N个视频片段中获取视频片段 $P_i$ 和视频片段 $P_j$ ，且将视频片段 $P_i$ 和视频片段 $P_j$ 作为待识别视频的目标视频片段对。应当理解，为了充分利用目标视

频片段对的多模态信息，该计算机设备不仅可以对目标视频片段对中的每个视频片段分别进行第一特征提取，得到目标视频片段对的第一目标编码特征，还可以对目标视频片段对中的每个视频片段分别进行第二特征提取，得到目标视频片段对的第二目标编码特征。

其中，可以理解的是，在第一模态信息为视频模态信息时，该计算机设备可以将目标视频片段对中的每个视频片段对应的视频帧分别作为待处理视频帧，且基于待处理视频帧的帧率确定抽帧参数。进一步地，该计算机设备可以基于抽帧参数对待处理视频帧进行抽帧处理，得到与待处理视频帧相关联的待编码视频帧。比如，当待处理视频帧为视频时长为6秒，且该待处理视频帧的帧率为25fps时，这意味着该待处理视频帧可以每秒传输25帧视频帧。此时，该计算机设备可以根据确定的抽帧参数(例如，每秒均匀抽取3帧视频帧)，对该待处理视频帧进行抽帧处理，以得到18帧视频帧，进而可以将这18帧视频帧作为与待处理视频帧相关联的待编码视频帧。

进一步地，该计算机设备可以获取与视频模态信息相关联的视频编码模型。其中，这里的视频编码模型可以是基于样本图像数据库所训练的，该样本图像数据库可以为一个用于视觉对象识别软件研究的大型可视化数据库。可选的，该视频编码模型还可以为开源的模型，例如，该视频编码模型可以为Imagenet预训练模型，该Imagenet预训练模型是一种基于大型通用物体识别开源数据集(例如，imagenet数据集)所训练的深度学习网络模型。可以理解的是，该计算机设备可以采用开源的Imagenet预训练模型中的残差网络(例如，ResNet-101)中的池化结果作为视频编码特征。

进一步地，请参见表1，表1是本申请实施例提供的一种残差网络结构示意图。其中，该残差网络结构示意图为ResNet-101的网络结构，该表1中可以包括卷积层(Layer name)、输出图像尺寸(Output size)以及每一个卷积层中的卷积信息。如表1所示，残差网络结构示意图中可以包括7个卷积网络层，具体可以包括卷积网络层1(例如，Conv1)、卷积网络层2(例如，Conv2\_x)、卷积网络层3(例如，Conv3\_x)、卷积网络层4(例如，Conv4\_x)、卷积网络层5(例如，Conv5\_x)、卷积网络层6(例如，Max pool，即池化层)以及卷积网络层7(例如，full connection，即全连接层)。其中，残差网络中的101层网络是指卷积网络层和全连接层的层数，而并未计算激活层或者池化层。

如表1所示，卷积网络层1中具有64个7x7的卷积，且步幅为2。而卷积网络层2、卷积网络层3、卷积网络层4以及卷积网络层5均为一种残差块(blocks)，例如，卷积网络层2中包括一个最大池化层(步幅为2的3x3的池化层)以及3个残差块，且每个残差块均包括3层，具体可以包括64个1x1的卷积、64个3x3的卷积、256个1x1的卷积。其中，全连接层中的N可以为学习的类别数量。具体请参见下述表1：

表1

| Layer name      | Output size | ResNet-101             |           |
|-----------------|-------------|------------------------|-----------|
| Conv1           | 300×500     | 7×7, 64, stride 2      |           |
| Conv2_x         | 150×250     | 3×3 max pool, stride 2 |           |
|                 |             | 1×1, 64                | ×3 blocks |
|                 |             | 3×3, 64                |           |
|                 |             | 1×1, 256               |           |
| Conv3_x         | 76×126      | 1×1, 128               | ×4 blocks |
|                 |             | 3×3, 128               |           |
|                 |             | 1×1, 512               |           |
| Conv4_x         | 38×63       | 1×1, 256               | ×3 blocks |
|                 |             | 3×3, 256               |           |
|                 |             | 1×1, 1024              |           |
| Conv5_x         | 19×32       | 1×1, 512               | ×3 blocks |
|                 |             | 3×3, 512               |           |
|                 |             | 1×1, 2048              |           |
| Max pool        |             | 1×2048                 |           |
| full connection |             | 1×N                    |           |

为便于理解，进一步地，请参见图4，图4是本申请实施例提供的一种残差块的结构示意图。如图4所示，残差块4k可以为上述表1中的卷积网络层2中的一个残差块。其中，该残差块4k中可以包括3个卷积层，具体可以包括卷积层41j(例如，64个1x1的卷积)、卷积层42j(例如，64个3x3的卷积)以及卷积层43j(例如，256个1x1的卷积)。

其中，计算机设备通过图4所示的残差块4k对输入特征进行卷积处理的过程中，既保持了精度又减少了计算量。如图4所示，计算机设备在将特征1(例如，一个具有256维度的输入特征)输入至该残差块4k中的卷积层41j时，可以用64个1x1卷积对该特征1进行降维处理，以得到特征2(例如，具有64维的降维特征)，以减少计算量。然后，该计算机设备可以将特征2输入至卷积层42j，进而可以用64个3x3的卷积对特征2进行卷积处理，得到图4所示的特征3(例如，具有64维的卷积特征)。最后，该计算机设备可以将特征3输入至卷积层43j，以通过256个1x1卷积进行特征还原，从而可以得到图4所示的特征4(例如，具有256维的还原特征)。其中，该特征4可以通过卷积层43j的输出特征以及特征1的残差所共同确定。图4中的线性整流函数(Rectified Linear Unit, relu)，又称修正线性单元，是一种人工神经网络中

常用的激活函数。

应当理解，本申请实施例中的计算机设备可以将待编码视频帧输入至上述视频编码模型，进而可以通过该视频编码模型，对待编码视频帧进行编码处理，以得到待编码视频帧对应的视频编码特征。进一步地，该计算机设备可以基于待编码视频帧对应的视频编码特征，得到目标视频片段对的第一目标编码特征。其中，第一目标编码特征包括视频片段 $P_i$ 对应的视频编码特征 $S_i$ ，以及视频片段 $P_j$ 对应的视频编码特征 $S_j$ 。

比如，该计算机设备可以将目标视频片段对中的视频片段 $P_i$ 对应的待编码视频帧输入至视频编码模型，由该视频编码模型输出视频片段 $P_i$ 对应的视频编码特征 $S_i$ 。又比如，该计算机设备可以将目标视频片段对中的视频片段 $P_j$ 对应的待编码视频帧输入至视频编码模型，由该视频编码模型输出视频片段 $P_j$ 对应的视频编码特征 $S_j$ 。此时，该计算机设备可以将视频编码特征 $S_i$ 和视频编码特征 $S_j$ 作为目标视频片段对的第一目标编码特征。

可见，当第一模态信息为视频模态信息时，可以基于待处理视频帧的帧率确定抽帧参数，从而准确的从待处理视频帧中抽取可表征目标视频片段的待编码视频帧，由此得到准确的第一目标编码特征。

其中，可以理解的是，在第二模态信息为音频模态信息时，该计算机设备可以将目标视频片段对中的每个视频片段对应的音频帧分别作为待处理音频帧，且对待处理音频帧进行音频准备处理，得到与待处理音频帧相关联的待编码音频帧。其中，该计算机设备在对待处理音频帧进行音频准备处理的过程中，可以对待处理音频帧进行重采样，进而可以对重采样后的待处理音频帧进行短时傅里叶变换，以得到待处理音频帧对应的频谱图。进一步地，该计算机设备可以将该频谱图映射到滤波器组中，以得到该待处理音频帧对应的初始声谱。该计算机设备可以通过下述公式(1)，确定初始声谱对应的稳定声谱，且根据组帧时长(例如，0.96秒)对该稳定声谱进行组帧处理，以得到待编码视频帧。

具体地，计算机设备确定稳定声谱的计算公式可以为下述公式(1)所示：

$$S = \log(\text{mel-spectrum} + 0.01) \quad (1)$$

其中，mel-spectrum是指初始声谱，0.01是指偏置参数，以避免初始声谱为0时需要取对数的情况。

比如，该计算机设备可以对待处理音频帧进行重采样，以得到16KHZ的单声道音频，进而可以通过使用25毫秒的Hann时窗，10毫秒的帧移，对重采样后的待处理音频帧进行短时傅里叶变换，以得到对应的频谱图。进一步地，该计算机设备可以将该频谱图映射到梅尔标度滤波器组(mel-scale filter banks, mel滤波器组)，以得到待处理音频帧对应的初始声谱(例如，梅尔声谱，mel声谱)。进一步地，该计算机设备可以通过上述公式(1)，确定mel声谱对应的稳定的mel声谱(即稳定声谱)。该计算机设备在得到稳定声谱之后，可以根据0.96秒的组帧时长，对稳定声谱进行组帧处理，且组帧处理过程中并没有帧的重叠，每一帧都包含64个mel频带，时长为10毫秒(即总共96帧)。

进一步地，该计算机设备可以获取与音频模态信息相关联的音频编码模型。其中，这里的音频编码模型可以是基于样本音频数据集所训练的，该样本音频数据集可以为携带有音频标记的视频音轨(例如，10秒时长的音频)组成的数据集。可选的，该音频编码模型还可

以为开源的模型，例如，该音频编码模型还可以为Vggish模型，该Vggish模型旨在为音频事件检测提取常见的大规模评估任务，且可以每秒提取1个128维的编码特征。其中，该Vggish模型的输出的音频编码特征的数据格式可以为[nums\_frames, 128]，其中，nums\_frames可以为待编码音频帧的帧长，nums\_frames=待处理音频帧的音频时长/0.96。

应当理解，本申请实施例中的计算机设备可以将待编码音频帧输入至上述音频编码模型，通过音频编码模型对待编码音频帧进行编码处理，得到待编码音频帧对应的音频编码特征。进一步地，该计算机设备可以基于待编码音频帧对应的音频编码特征，得到目标视频片段对的第二目标编码特征。其中，第二目标编码特征包括视频片段 $P_i$ 对应的音频编码特征 $Y_i$ ，以及视频片段 $P_j$ 对应的音频编码特征 $Y_j$ 。

比如，该计算机设备可以将目标视频片段对中的视频片段 $P_i$ 对应的待编码音频帧输入至音频编码模型，由该音频编码模型输出视频片段 $P_i$ 对应的音频编码特征 $Y_i$ 。又比如，该计算机设备可以将目标视频片段对中的视频片段 $P_j$ 对应的待编码音频帧输入至音频编码模型，由该音频编码模型输出视频片段 $P_j$ 对应的音频编码特征 $Y_j$ 。此时，该计算机设备可以将音频编码特征 $Y_i$ 和音频编码特征 $Y_j$ 作为目标视频片段对的第二目标编码特征。

可见，当第一模态信息为音频模态信息时，可以基于待处理音频帧进行音频准备处理，从而准确确定出可以表征目标视频片段的待编码音频帧，由此得到准确的第二目标编码特征。

其中，本申请实施例中的视频编码模型和音频编码模型可以独立于本申请实施例中的目标网络模型，还可以是该目标网络模型中的两个子模型，这里将不对其进行限定。

S102，获取用于对待识别视频进行循环识别的目标网络模型。

具体地，该计算机设备还可以获取用于对待识别视频进行循环识别的目标网络模型(例如，图2所示的目标网络模型2000M)。其中，该目标网络模型可以包括与第一模态信息相关联的第一目标序列模型(例如，图2所示的目标序列模型210M)、和与第二模态信息相关联的第二目标序列模型(例如，图2所示的目标序列模型220M)。

在一种可能的实现方式中，第一目标序列模型可以包括第一序列表征学习层和第一相似度度量层，第二目标序列模型包括第二序列表征学习层和第二相似度度量层。

例如，该第一目标序列模型可以为与视频模态信息相关联的序列相似度学习模型，该第一目标序列模型可以用于识别该目标视频片段对在视频模态信息下的目标相似结果，该第二目标序列模型可以为与音频模态信息相关联的序列相似度学习模型，该第二目标序列模型可以用于识别待目标视频片段对在音频模态信息下的目标相似结果。

S103，将第一目标编码特征输入至第一目标序列模型，输出目标视频片段对的第一目标相似结果。

该第一目标相似结果可以体现出基于第一模态信息，目标视频片段对中两个视频片段将的相似程度。

在一种可能的实现方式中，S103包括：将所述第一目标编码特征输入至所述第一目标序列模型，通过所述第一序列表征学习层对所述第一目标编码特征进行序列特征学习，将序列特征学习到的第一目标学习特征输入至所述第一相似度度量层，由所述第一相似度

量层输出所述目标视频片段对的第一目标相似结果。

其中，该目标视频片段对可以包括视频片段 $P_i$ 和视频片段 $P_j$ ； $i$ 和 $j$ 均为小于或者等于 $N$ 的正整数，且 $i$ 不等于 $j$ ，这里的 $N$ 可以为待识别视频中的视频片段的总数量。本申请实施例中的第一目标序列模型中的第一序列表征学习层可以包括与视频片段 $P_i$ 相关联的第一网络层和与视频片段 $P_j$ 相关联的第二网络层，且该第一网络层与该第二网络层具有相同的网络结构。具体地，该计算机设备可以将第一目标编码特征输入至第一目标序列模型。这里的第一目标编码特征可以包括视频编码特征 $S_i$ 和视频编码特征 $S_j$ ，视频编码特征 $S_i$ 可以为视频片段 $P_i$ 在第一模态信息下的编码特征；视频编码特征 $S_j$ 可以为视频片段 $P_j$ 在第一模态信息下的编码特征。进一步地，该计算机设备可以通过第一序列表征学习层中的第一网络层，对视频编码特征 $S_i$ 进行序列特征学习，得到视频编码特征 $S_i$ 对应的学习特征 $X_i$ 。同理，该计算机设备可以通过第一序列表征学习层中的第二网络层，对视频编码特征 $S_j$ 进行序列特征学习，得到视频编码特征 $S_j$ 对应的学习特征 $X_j$ 。进一步地，该计算机设备可以将学习特征 $X_i$ 和学习特征 $X_j$ 作为第一目标学习特征，且将第一目标学习特征输入至第一相似度量层，由第一相似度量层输出第一目标学习特征的相似度。此时，该计算机设备可以基于第一目标学习特征的相似度，确定目标视频片段对的第一目标相似结果。

可以理解的是，本申请实施例中的相似度量层(即 Similarity 层)均可以采用 cosine-similarity 这一计算方法，对输入至该相似度量层的两个特征计算相似度。具体计算公式可以参见下述公式(2)：

$$L(x,y)=x^T y \quad (2)$$

其中，这里的 $x$ 可以为输入至该相似度量层的一个输入特征， $y$ 为输入至该相似度量层的另一个输入特征。

比如，若该相似度量层为第一目标序列模型中的第一相似度量层，则这里的 $x$ 可以为该第一目标序列模型中的第一网络层所输出的学习特征 $X_i$ ，这里的 $y$ 可以为该第一目标序列模型中的第二网络层所输出的学习特征 $X_j$ 。

为便于理解，进一步地，请参见图5，图5是本申请实施例提供的一种目标序列模型的结构示意图。如图5所示，本申请实施例中的目标序列模型500M可以为与某一模态信息相关联的目标序列模型。例如，该目标序列模型500M可以为上述图2所示的目标网络模型2000M中的目标序列模型210M，即与第一模态信息(例如，视频模态信息)相关联的目标序列模型。

其中，本申请实施例中的待识别视频的目标视频片段对可以包括视频片段 $P_1$ 和视频片段 $P_2$ ，该视频片段 $P_1$ 在视频模态信息下的视频编码特征可以为图5所示的视频编码特征 $5S_1$ ，且该视频片段 $P_2$ 在视频模态信息下的视频编码特征可以为图5所示的视频编码特征 $5S_2$ 。

如图5所示，该目标序列模型500M中可以包括图5所示的序列表征学习层51L(即第一序列表征学习层)和相似度量层52L(例如，第一相似度量层)，且该序列表征学习层51L中可以包括网络层511L以及网络层512L。其中，网络层511L可以为与视频片段 $P_1$ 相关联的第一网络层，该网络层512L可以为与视频片段 $P_2$ 相关联的第二网络层，且该网络层511L与该网络层512L具有相同的网络结构。

-14-

可以理解的是,在本申请实施例中,计算机设备可以将视频编码特征 $5S_1$ 和视频编码特征 $5S_2$ 一并输入至图5所示的序列表征学习层51L,以得到目标视频片段对的第一目标学习特征。换言之,该计算机设备可以将视频编码特征 $5S_1$ 输入至序列表征学习层51L中的网络层511L,通过该网络层511L对视频编码特征 $5S_1$ 进行序列特征学习,得到该视频编码特征 $5S_1$ 对应的学习特征 $5X_1$ 。与此同时,该计算机设备还可以将视频编码特征 $5S_2$ 输入至序列表征学习层51L中的网络层512L,通过该网络层512L对视频编码特征 $5S_2$ 进行序列特征学习,得到该视频编码特征 $5S_2$ 对应的学习特征 $5X_2$ 。

为便于理解,进一步地,请参见表2,表2是本申请实施例提供的一种深度神经网络(Deep Neural Networks, 简称DNN)的网络结构示意图。其中,表2所示的深度神经网络可以用于对视频片段对应的模态编码特征进行序列特征学习。例如,表2所示的网络结构示意图可以为图5所示的网络层511L对应的网络结构示意图或者网络层512L对应的网络结构示意图。

其中,该表2中可以包括子网络层、子网络层中的每个卷积层(Layer name)、每个卷积层的输入/输出特征尺寸(Input/Output size)以及每个卷积层中的卷积核(Kernel)。如表2所示,本申请实施例中的第一序列表征学习层中的第一网络层可以包括第一子网络层(例如,子网络层 $L_1$ )、第二子网络层(例如,子网络层 $L_2$ )、第三子网络层(例如,子网络层 $L_3$ )以及第四子网络层(例如,子网络层 $L_4$ )。其中,表2所示的每个卷积层中的步幅(stride)和空洞卷积参数(dilate)均可以根据实际情况进行动态调整。具体请参见下述表2:

表2

| 子网络层       | Layer name | Input/Output size          | Kernel                       |           |
|------------|------------|----------------------------|------------------------------|-----------|
| 子网络层 $L_1$ | Conv1      | 18×2048, 18×1024           | 7×7, 1024, stride 2, dilate1 |           |
|            | Conv2      | 18×1024, 18×1024           | 7×7, 1024, stride 2, dilate1 |           |
|            | Conv3      | 18×2048, 9×1024            | 7×7, 1024, stride 2, dilate2 |           |
|            | Concate    | (18×2048, 9×1024), 27×1024 | -                            |           |
| 子网络层 $L_2$ | Conv4_x    | 27×1024, 14×1024           | 1×1, 1024, dilate2           | ×1 blocks |
|            |            |                            | 3×3, 1024                    |           |
|            |            |                            | 1×1, 1024                    |           |
| 子网络层 $L_3$ | Conv5_x    | 14×1024, 7×512             | 1×1, 512, dilate2            | ×1 blocks |
|            |            |                            | 3×3, 512                     |           |
|            |            |                            | 1×1, 512                     |           |
| 子网络层 $L_4$ | Conv6_x    | 7×512, 1×512               | 1×7, 512                     | ×1 blocks |

在通过第一序列表征学习中的第一网络层,对视频编码特征 $S_i$ 进行序列特征学习的具体实施方式中,该计算机设备可以通过第一网络层中的第一子网络层,对视频编码特征 $S_i$ 进行特征转换,从而可以得到视频编码特征 $S_i$ 对应的第一转换特征。其中,第一网络层中的第一子网络层可以包括第一卷积层(例如,表2所示的Conv1这一卷积层)、第二卷积层(例

如,表2所示的Conv2这一卷积层)以及空洞卷积层(例如,表2所示的Conv3这一卷积层)。比如,Conv3这一空洞卷积层中的步幅和空洞卷积参数均可以为2。该特征拼接层(例如,表2所示的Concate这一拼接层)用于将Conv1这一卷积层中的输出特征和Conv2这一卷积层的输出特征进行拼接处理,以得到用于输入至第二子网络层中的输入特征。

5 可以理解的是,在该计算机设备将视频编码特征 $S_i$ 输入至第一网络层中的第一子网络层(例如,表2所示的子网络层 $L_1$ )时,该计算机设备可以通过空洞卷积层,对视频编码特征 $S_i$ 进行卷积处理,以得到视频编码特征 $S_i$ 对应的第一卷积特征。与此同时,该计算机设备还可以通过第一卷积层,对视频编码特征 $S_i$ 进行卷积处理,以得到视频编码特征 $S_i$ 对应的第二卷积特征,且将该第二卷积特征输入至第二卷积层,通过第二卷积层对第二卷积特征进行  
10 卷积处理,以得到第三卷积特征。此时,该计算机设备可以将第一卷积特征和第三卷积特征进行拼接处理,从而得到视频编码特征 $S_i$ 对应的第一转换特征。

进一步地,该计算机设备可以先将第一转换特征输入至第二子网络层(例如,表2所示的子网络层 $L_2$ ),通过第二子网络层对第一转换特征进行特征转换,从而得到第一转换特征对应的第二转换特征。然后,该计算机设备可以将第二转换特征输入至第三子网络层,通过第三子网络层对第二转换特征进行特征转换,以得到第二转换特征对应的第三转换特征。  
15 最后,该计算机设备再将第三转换特征输入至第四子网络层,通过第四子网络层中的最大池化层,对第三转换特征进行最大池化处理,以得到视频编码特征 $S_i$ 对应的学习特征 $X_i$ 。

通过第一网络层包括的多个子网络层依次对视频编码特征进行特征转换,可以提高对视频编码特征的有效数据的提取精度,得到更为准确的学习特征。

20 便于理解,进一步地,请参见图6,图6是本申请实施例提供的一种深度神经网络的网络示意图。其中,该深度神经网络(例如,图6所示的网络层6L)可以用于对目标视频片段对中的某一视频片段对应的编码特征进行序列特征学习。该网络层6L可以为上述图5所示的网络层511L或者上述图5所示的网络层512L。

其中,该网络层6L可以包括子网络层 $6L_1$ (即第一子网络层)、子网络层 $6L_2$ (即第二子网络层)、子网络层 $6L_3$ (即第三子网络层)以及子网络层 $6L_4$ (即第四子网络层)。如图6所示,该子网络层 $6L_1$ 可以包括第一卷积层(例如,卷积层601j)、第二卷积层(例如,卷积层602j)以及空洞卷积层(例如,卷积层603j)。可以理解的是,空洞卷积层可以包括以下作用:(1)扩大感受野:在深度神经网络中为了增加感受野且降低计算量,往往需要进行降采样以增加感受野,但这样将会导致空间分辨率降低。为了能不丢失分辨率,且仍然扩大感受野,本申请  
25 实施例可以使用空洞卷积。(2)捕获多尺度上下文信息:空洞卷积层中的空洞卷积参数(dilation rate)可以根据实际情况设置,具体含义是指在卷积核中填充(dilation rate-1)个0,因此,当设置不同dilation rate时,感受野就会不一样,也即获取了多尺度信息。

其中,图6所示的卷积层603j可以为空洞卷积参数为2的卷积层,即空洞卷积层,这意味着该计算机设备可以在卷积核中填充1个0。如图6中的区域 $Q_1$ 所示,该计算机设备可以根据卷积层603j中的 $3 \times 3$ 的卷积核,会在图像中找到 $5 \times 5$ 的区域,并从中根据空洞卷积参数的数值选择对应位置进行相乘操作。其中,图6所示的卷积层601j和卷积层602j均可以为空洞卷积参数为1的卷积层,即常规卷积层。可以理解的是,当空洞卷积参数的为1时,这意  
35

意味着该计算机设备无法在卷积核中填充0,如图6中的区域 $Q_2$ 所示,该计算机设备可以根据卷积层601中的 $3 \times 3$ 的卷积核,会在图像中找到 $3 \times 3$ 的区域。

其中,根据上述表2所示,该计算机设备在对特征1(例如,具有 $18 \times 2048$ 维度的视频编码特征 $S_i$ )进行序列特征学习的具体实施方式中,可以将该特征1作为子网络层 $6L_1$ 的输入特征。此时,该计算机设备可以通过该子网络层 $6L_1$ 中的卷积层603j中的1024个 $7 \times 7$ 的卷积核,对该特征1进行卷积处理进行卷积处理,得到第一卷积特征(例如,具有 $9 \times 1024$ 维度的特征2)。与此同时,该计算机设备还可以通过该子网络层 $6L_1$ 中的卷积层601j中的1024个 $7 \times 7$ 的卷积核,对该特征1进行卷积处理,得到第二卷积特征(例如,具有 $18 \times 1024$ 维度的特征3)。进一步地,该计算机设备可以将该特征3作为卷积层602j的输入特征,通过卷积层602j中的1024个 $7 \times 7$ 的卷积核,对特征3进行卷积处理,得到第二卷积特征(例如,具有 $18 \times 1024$ 维度的特征4)。进一步地,该计算机设备可以将特征2和特征4进行拼接处理,得到子网络 $6L_1$ 的输出特征(即具有 $27 \times 1024$ 维度的第一转换特征)。

然后,该计算机设备可以将具有 $27 \times 1024$ 维度的第一转换特征输入至子网络 $6L_2$ ,通过该子网络 $6L_2$ 中的卷积层604j中的1024个 $1 \times 1$ 的卷积核、卷积层605j中的1024个 $3 \times 3$ 的卷积核、以及卷积层606j中的1024个 $1 \times 1$ 的卷积核,得到具有 $14 \times 1024$ 维度的第二转换特征。

进一步地,该计算机设备可以将具有 $14 \times 1024$ 维度的第二转换特征输入至图6所示的子网络 $6L_3$ ,通过该子网络 $6L_3$ 中的卷积层607j中的512个 $1 \times 1$ 的卷积核、卷积层608j中的512个 $3 \times 3$ 的卷积核、以及卷积层609j中的512个 $1 \times 1$ 的卷积核,得到具有 $7 \times 512$ 维度的第三转换特征。

最后,该计算机设备可以将具有 $7 \times 512$ 维度的第三转换特征输入至图6所示的子网络 $6L_4$ ,通过该子网络 $6L_4$ 中的卷积层610j(即最大池化层)中的512个 $1 \times 7$ 的卷积核,对第三转换特征进行最大池化处理,得到具有 $1 \times 512$ 维度的学习特征 $X_i$ 。

同理,该计算机设备可以将视频编码特征 $S_j$ 输入至与图6所示的网络层 $6L$ 具有相同网络结构的第二网络层中,可以得到视频编码特征 $S_j$ 对应的学习特征 $X_j$ ,且将学习特征 $X_i$ 和学习特征 $X_j$ 作为目标视频片段对的第一目标学习特征。

如图5所示,该计算机设备确定的第一目标学习特征可以为图5所示的学习特征 $5X_1$ 和学习特征 $5X_2$ ,此时,该计算机设备将学习特征 $5X_1$ 和学习特征 $5X_2$ 一并输入至图5所示的相似度量层52L,由该相似度量层52L输出学习特征 $5X_1$ 和学习特征 $5X_2$ 的相似度,进而可以基于相似度量层52L输出的相似度,确定目标视频片段对的第一目标相似结果。

其中,这里的相似度量层52L输出的相似度可以采用sigmoid这一激活函数映射到0至1之间。如图5所示,该计算机设备可以获取与目标序列模型500M相关联的相似阈值(例如,0.5),其中,这里的相似阈值可以根据实际情况进行动态调整。若学习特征 $5X_1$ 和学习特征 $5X_2$ 的相似度(例如,0.4)未达到相似阈值,则该计算机设备可以确定该目标视频片段对的第一目标相似结果指示视频片段 $P_1$ 和视频片段 $P_2$ 在视频模态信息下不相似。若学习特征 $5X_1$ 和学习特征 $5X_2$ 的相似度(例如,0.8)达到相似阈值,则该计算机设备可以该确定目标视频片段对的第一目标相似结果指示视频片段 $P_1$ 和视频片段 $P_2$ 在视频模态信息下相似。

由此,通过第一目标序列模型包括的第一网络层和第二网络层可以分别对目标视频片

段对中的两个视频片段进行序列特征学习，从而更为准确的识别出这两个视频片段间的特征相似性，提高了第一目标相似结果的准确度。

S104，将第二目标编码特征输入至第二目标序列模型，输出目标视频片段对的第二目标相似结果。

5 该第二目标相似结果可以体现出基于第二模态信息，目标视频片段对中两个视频片段将的相似程度。

在一种可能的实现方式中，S104包括：将第二目标编码特征输入至第二目标序列模型，通过第二序列表征学习层对第二目标编码特征进行序列特征学习，将序列特征学习到的第二目标学习特征输入至第二相似度度量层，由第二相似度度量层输出目标视频片段对的第二目标相似结果。

10 具体地，该计算机设备可以将第二目标编码特征输入至与第二模态信息相关联的第二目标序列模型。这里的第二目标编码特征可以包括音频编码特征 $Y_i$ 和音频编码特征 $Y_j$ ，音频编码特征 $Y_i$ 可以为视频片段 $P_i$ 在第二模态信息下的编码特征；音频编码特征 $Y_j$ 可以为视频片段 $P_j$ 在第二模态信息下的编码特征。进一步地，该计算机设备可以通过第二序列表征学习层中的与音频编码特征 $Y_i$ 相关联的网络层(即第三网络层)，对音频编码特征 $Y_i$ 进行序列特征学习，得到音频编码特征 $Y_i$ 对应的学习特征 $Z_i$ 。同理，该计算机设备可以通过第二序列表征学习层中的与音频编码特征 $Y_j$ 相关联网络层(即第四网络层)，对音频编码特征 $Y_j$ 进行序列特征学习，得到音频编码特征 $Y_j$ 对应的学习特征 $Z_j$ 。进一步地，该计算机设备可以将学习特征 $Z_i$ 和学习特征 $Z_j$ 作为第二目标学习特征，且将第二目标学习特征输入至第二相似度度量层，由第二相似度度量层输出第二目标学习特征的相似度。此时，该计算机设备可以基于第二目标学习特征的相似度，确定目标视频片段对的第二目标相似结果。

20 其中，该计算机设备确定目标视频片段的第二目标相似结果的具体实施方式可以参见上述S103中确定目标视频片段的第一目标相似结果的具体实施方式，这里将不再继续进行赘述。

25 S105，将第一目标相似结果和第二目标相似结果进行比对，得到目标视频片段对的循环比对结果。

具体地，该计算机设备可以将第一目标相似结果和第二目标相似结果进行比对。若第一目标相似结果指示目标视频片段对在第一模态信息下相似，且第二目标相似结果指示目标视频片段对在第二模态信息下相似，则该计算机设备可以得到目标视频片段对的循环视频结果。可选的，若第一目标相似结果指示目标视频片段对在第一模态信息下不相似，或者第二目标相似结果指示目标视频片段对在第二模态信息下不相似，则该计算机设备可以得到目标视频片段对的非循环视频结果。此时，该计算机设备可以将循环视频结果或者非循环视频结果作为目标视频片段对的循环比对结果。其中，目标视频片段对的循环比对结果可以用于指示待识别视频的视频类型。

35 可以理解的是，若目标视频片段对的第一目标相似结果与第二目标相似结果均为数值为“1”的相似结果，则该计算机设备可以确定该目标视频片段对的循环比对结果为数值为“1”的循环视频结果，这意味着目标视频片段对中的两个视频片段中的画面和音频均互相

重复。可选的,若目标视频片段对的第一目标相似结果或者第二目标相似结果为数值为“0”的相似结果,则该计算机设备可以确定该目标视频片段对的循环比对结果为数值为“0”的非循环视频结果,这意味着目标视频片段对中的两个视频片段不重复。

可见,通过充分利用该目标视频片段对在不同模态下的信息,并基于不同模态间相似  
5 度结果的一致性来确定目标视频片段对是否为循环视频结果,以更加准确的得到该目标视频片段对的循环比对结果,进而提高了视频循环识别的准确率。

在本申请实施例中,该计算机设备在对待识别视频进行循环识别时,可以事先获取待  
识别视频的目标视频片段对,并确定目标视频片段对的第一目标编码特征以及第二目标编  
码特征。其中,这里的第一目标编码特征对应的第一模态信息(例如,视频模态信息)不同于  
10 第二目标编码特征对应的第二模态信息(例如,音频模态信息)。在获取到用于对待识别视频  
进行循环的目标网络模型时,该计算机设备可以充分利用该目标视频片段对的视频信息,  
即不仅可以通过目标网络模型中的第一目标序列模型,确定该目标视频片段对在第一模态  
信息下的第一目标相似结果,还可以通过目标网络模型中的第二目标序列模型,确定该目  
标视频片段对在第二模态信息下的第二目标相似结果。进一步地,该计算机设备可以将该  
15 第一目标相似结果和第二目标相似结果进行比对,以更加准确的得到该目标视频片段对的  
循环比对结果,进而提高了视频循环识别的准确率。

进一步地,请参见图7,图7是本申请实施例提供的一种视频循环识别方法的流程示意  
图。该方法可以由具备视频循环识别功能的计算机设备执行。其中,该计算机设备可以为  
用户终端(例如,上述图1所示的用户终端集群中的任意一个用户终端,例如,用户终端100a),  
20 也可以为服务器(例如,上述图1所示的服务器10)。该方法还可以由具备模型应用功能的用  
户终端和具备模型训练功能的服务器共同执行。在此不做限定。该方法至少可以包括以下  
S201-S211:

S201,获取待识别视频的目标视频片段对,确定目标视频片段对的第一目标编码特征  
和第二目标编码特征。

25 S202,获取用于对待识别视频进行循环识别的目标网络模型。

S203,将第一目标编码特征输入至第一目标序列模型,通过第一序列表征学习层对第  
一目标编码特征进行序列特征学习,将序列特征学习到的第一目标学习特征输入至第一相  
似度度量层,由第一相似度度量层输出目标视频片段对的第一目标相似结果。

S204,将第二目标编码特征输入至第二目标序列模型,通过第二序列表征学习层对第  
30 二目标编码特征进行序列特征学习,将序列特征学习到的第二目标学习特征输入至第二相  
似度度量层,由第二相似度度量层输出目标视频片段对的第二目标相似结果。

S205,将第一目标相似结果和第二目标相似结果进行比对,得到目标视频片段对的循  
环比对结果。

其中,本申请实施例中的视频循环识别方法可以包括模型训练过程以及模型应用过程。  
35 可以理解的是,该S201-S205阐述了模型应用过程,该模型应用过程的具体实施方式可参见  
上述图3所对应实施例中对S101-S105的描述,这里将不再继续进行赘述。

若循环比对结果为循环视频结果,则该计算机设备可以确定待识别视频的视频类型为

循环视频类型，进一步地，该计算机设备还可以基于循环视频类型生成循环提示信息，且将循环提示信息返回至用户终端。其中，该用户终端可以为待识别视频的发送端。

为便于理解，进一步地，请参见图8，图8是本申请实施例提供的一种数据交互的场景示意图。如图8所示，本申请实施例中的服务器80B可以为具备视频循环识别功能的计算机设备，该服务器80B可以为上述图1所示的服务器10。本申请实施例中的用户终端80A可以为与服务器80B具有网络连接关系的用户终端，例如，该用户终端可以为上述图1所示的用户终端集群中的任意一个用户终端，例如，用户终端100b。

如图8所示，用户终端80A可以将待识别视频8V发送至服务器80B，以使该服务器80B获取训练好的目标网络模型，对待识别视频8V进行循环识别，以确定该待识别视频8V的视频类型。其中，如图8所示，该服务器80B接收到该待识别视频8V时，可以对该待识别8V进行分段处理，以获取该待识别视频8V的目标视频片段对。其中，该待识别视频8V的目标视频片段对可以包括图8所示的视频片段 $P_1$ 和视频片段 $P_2$ 。

进一步地，该服务器80B可以确定目标视频片段对的第一目标编码特征和第二目标编码特征。在获取到目标网络模型(例如，图8所示的网络模型8000M)时，该服务器80B可以将第一目标编码特征和第二目标编码特征一并输入至网络模型8000M，由该网络模型8000M对这一目标视频片段对进行循环识别，以得到该目标视频片段对的循环比对结果。

可以理解的是，若该循环比对结果为循环视频结果，则该服务器80B可以确定待识别视频8V的视频类型为循环视频类型，此时，该服务器80B可以基于确定的循环视频类型，生成用于提示该用户终端80A的循环提示信息(例如，图8所示的循环提示信息80F)，进一步地，该服务器80B可以将该循环提示信息80F返回至用户终端80A。

应当理解，该用户终端80A可以根据循环提示信息80F，确定是否需要服务器80B对待识别视频8V进行过滤处理。当用户终端80A对应的用户确定待识别视频8V中的重复片段需要进行过滤时，该用户可以执行用于指示服务器80B对该待识别视频8V进行过滤的触发操作，以使用户终端在响应该触发操作时，向服务器80B发送过滤请求。服务器80B在接收到该过滤请求时，可以对待识别视频8V进行过滤处理，进而可以将过滤处理后的待识别视频8V存储至服务器80B的数据库。

可以理解的是，在服务器80B为接入至区块链网络中的区块链节点时，该服务器80B还可以将该过滤处理后的待识别视频8V进行打包处理，生成包含过滤处理后的待识别视频8V的待验证区块，进而可以将该待验证区块广播至区块链网络中的共识节点，以使共识节点进行区块共识，且在接收到的区块共识节点指示共识成功时，服务器80B可以将该待验证区块写入至区块链网络中的区块链，即将过滤处理后的待识别视频8V写入区块链网络。其中，区块链是分布式数据存储、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。区块链(Block chain)，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，每一个数据块中包含了一批网络交易的信息，用于验证其信息的有效性(防伪)和生成下一个区块。

其中，该模型训练过程具体可以参见下述S206-S211的描述。

S206，获取用于训练初始网络模型的样本视频片段对以及样本视频片段对的样本标签。

具体地,该计算机设备可以从样本视频片段集合中随机获取一个视频片段(例如,视频片段A),以作为初始样本视频片段。进一步地,该计算机设备可以对该初始样本视频片段进行整体攻击(加高斯噪声、改变视频片段A的视频帧率),以得到该初始样本视频片段对应的目标样本视频片段。进一步地,该计算机设备可以为这一初始样本视频片段和这一目标样本视频片段所组成的视频片段对进行标注,以得到携带第一样本标签(例如,正样本标签)的视频片段对。其中,本申请实施例可以将携带第一样本标签的视频片段对称之为正样本视频片段对。此外,该计算机设备还可以从样本视频片段集中选择K个与初始样本视频片段不重复的视频片段,进而可以分别对选择的K个视频片段中的每个视频片段与该初始样本视频片段进行标注,以得到K个携带第二样本标签(例如,负样本标签)的视频片段对。其中,本申请实施例可以将携带第二样本标签的视频片段对称之为负样本视频片段对。进一步地,该计算机设备可以将正样本视频片段对和负样本视频片段对作为用于训练初始网络模型的样本视频片段对,且将第一样本标签和第二样本标签均称之为对应样本视频片段对的样本标签。

比如,该计算机设备在获取正样本视频片段对时,可以对视频片段A进行均匀抽帧处理,使得视频片段A的帧率由原来的25fps更新为12fps,进而可以将具有25fps的视频片段A和具有12fps的视频片段A作为需要标注互相重复的正样本视频片段对。

S207,分别获取样本视频片段对在第一模态信息下的第一样本编码特征和在第二模态信息下的第二样本编码特征。

具体地,该计算机设备可以对样本视频片段对中的每个视频片段分别进行第一特征提取,得到样本视频片段对的第一样本编码特征(即第一模态信息下的编码特征)。与此同时,该计算机设备还可以对样本视频片段对中的每个视频片段分别进行第二特征提取,得到样本视频片段对的第二样本编码特征(即第二模态信息下的编码特征)。

其中,样本视频片段对可以包括第一样本视频片段(例如,视频片段A)和第二样本视频片段(例如,视频片段B)。该样本视频片段对的第一样本编码特征可以包括视频片段A在第一模态信息下的视频编码特征(例如,视频编码特征 $S_A$ )、以及视频片段B在第一模态信息下的视频编码特征(例如,视频编码特征 $S_B$ )。该样本视频片段对的第二样本编码特征可以包括视频片段A在第二模态信息下的视频编码特征(例如,音频编码特征 $Y_A$ )、以及视频片段B在第二模态信息下的视频编码特征(例如,音频编码特征 $Y_B$ )。

S208,将第一样本编码特征输入至第一初始序列模型,通过第一初始序列模型中的第一序列表征学习层和第一相似度度量层,输出样本视频片段对的第一预测相似结果。

其中,第一初始序列模型中的第一序列表征学习层可以包括与视频片段A相关联的第一网络层和与视频片段B相关联的第二网络层,且该第一网络层与该第二网络层具有相同的网络结构。具体地,该计算机设备可以将第一样本编码特征输入至第一初始序列模型。进一步地,该计算机设备可以通过第一初始序列模型中的第一序列表征学习层中的第一网络层,对视频编码特征 $S_A$ 进行序列特征学习,得到视频编码特征 $S_A$ 对应的学习特征 $X_A$ 。同理,该计算机设备可以通过第一序列表征学习层中的第二网络层,对视频编码特征 $S_B$ 进行序列特征学习,得到视频编码特征 $S_B$ 对应的学习特征 $X_B$ 。进一步地,该计算机设备可以将学习

特征 $X_A$ 和学习特征 $X_B$ 作为第一样本学习特征,且将第一样本学习特征输入至第一相似度量层,由第一相似度量层输出第一样本学习特征的相似度。此时,该计算机设备可以基于第一样本学习特征的相似度,确定样本视频片段对的第一预测相似结果。

5 S209,将第二样本编码特征输入至第二初始序列模型,通过第二初始序列模型中的第二序列表征学习层和第二相似度量层,输出样本视频片段对的第二预测相似结果。

具体地,该计算机设备可以将第二样本编码特征输入至与第二模态信息相关联的第二初始序列模型。进一步地,该计算机设备可以通过第二序列表征学习层中的与音频编码特征 $Y_A$ 相关联的网络层(即第三网络层),对音频编码特征 $Y_A$ 进行序列特征学习,得到音频编码特征 $Y_A$ 对应的学习特征 $Z_A$ 。同理,该计算机设备可以通过第二序列表征学习层中的与音频编码特征 $Y_B$ 相关联网络层(即第四网络层),对音频编码特征 $Y_B$ 进行序列特征学习,得到音频编码特征 $Y_B$ 对应的学习特征 $Z_B$ 。进一步地,该计算机设备可以将学习特征 $Z_A$ 和学习特征 $Z_B$ 作为第二样本学习特征,且将第二样本学习特征输入至第二相似度量层,由第二相似度量层输出第二样本学习特征的相似度。此时,该计算机设备可以基于第二样本学习特征的相似度,确定样本视频片段对的第二预测相似结果。

15 其中,该S207S209的具体实施方式可参见上述图3所对应实施例中对S102-S104的描述,这里将不再继续进行赘述。

S210,将第一预测相似结果和第二预测相似结果进行比对,得到样本视频片段对的预测循环结果对应的预测标签。

具体地,该计算机设备可以将第一预测相似结果和第二预测相似结果进行比对。若第一预测相似结果指示样本视频片段对在第一模态信息下相似,且第二预测相似结果指示样本视频片段对在第二模态信息下相似,则该计算机设备可以得到样本视频片段对的循环视频结果。可选的,若第一预测相似结果指示样本视频片段对在第一模态信息下不相似,或者第二预测相似结果指示样本视频片段对在第二模态信息下不相似,则该计算机设备可以得到样本视频片段对的非循环视频结果。此时,该计算机设备可以将循环视频结果或者非循环视频结果作为样本视频片段对的预测循环结果,进而可以基于预测循环结果确定样本视频片段对对应的预测标签。

其中,这里的预测循环结果可以包括正样本视频片段对的第一预测循环结果和负样本视频片段对的第二预测循环结果;这里的预测标签可以包括第一预测循环结果对应的第一预测标签和第二预测循环结果对应的第二预测标签。

30 S211,基于预测标签和样本标签,对初始网络模型进行迭代训练,将迭代训练后的初始网络模型作为用于对待识别视频进行循环识别的目标网络模型。

具体地,该计算机设备可以基于样本标签所指示的正样本视频片段对和负样本视频片段对之间的样本比例,确定与初始网络模型的模型损失函数相关联的损失权重参数。进一步地,该计算机设备可以基于第一预测标签和第一样本标签,得到正样本视频片段对的正样本损失,且基于第二预测标签和第二样本标签,得到负样本视频片段对的负样本损失。此时,该计算机设备可以基于正样本损失、负样本损失以及损失权重参数,以得到模型损失函数对应的模型损失,进而可以基于模型损失,对初始网络模型进行迭代训练,得到模

- 22 -

型训练结果。进一步地,该计算机设备可以基于模型训练结果,确定用于对待识别视频进行循环识别的目标网络模型。

具体地,本申请实施例中的初始网络模型的模型损失函数可以参见下述公式(3)-(4):

$$E = -\sum_{x,y \in P_{\text{same}}} \ln(Pr(x,y)) - K \sum_{x,y \in P_{\text{diff}}} \ln(1 - Pr(x,y)) \quad (3)$$

$$Pr(x,y) = \frac{1}{1 + e^{-L(x,y)}} \quad (4)$$

其中,  $L(x,y)$ 是指由初始网络模型输出的预测相似结果,(例如,数值为“1”或者数值

为“0”的预测相似结果)。 $\sum_{x,y \in P_{\text{same}}} \ln(Pr(x,y))$  可以为计算机设备基于第一预测标签和第一

样本标签所得到的正样本视频片段对的正样本损失, $\sum_{x,y \in P_{\text{diff}}} \ln(1 - Pr(x,y))$  可以为该计算机设

备基于第二预测标签和第二样本标签所得到的负样本视频片段对的负样本损失。这里的K是指与初始网络模型的模型损失函数相关联的损失权重参数。比如,若样本标签所指示的正样本视频片段对和负样本视频片段对之间的样本比例为1:10,则这里的损失权重参数可以为0.1。通过采用正负样本视频片段对进行模型训练,可以避免训练过程中出现过拟合的问题,而且可以有助于提升目标网络模型的精度。而且,还可以通过控制正负样本视频片段对的比例,均衡正负样本视频片段对的比例,进一步提升训练效率,提升目标网络模型的精度。

应当理解,该计算机设备在基于模型损失对初始网络模型进行迭代训练的过程中,可以采用基于SGD(Stochastic Gradient Descent)的梯度下降法确定初始网络模型的模型参数w和偏置参数b,在每次迭代过程中,通过将确定到的模型损失(即预测结果误差)反向传播到初始网络模型,进而可以确定梯度并更新该初始网络模型的参数。

首先,该计算机设备可以初始网络模型的模型参数可以进行参数初始化,即所有卷积层可以采用方差为0.01,均值为0的高斯分布进行初始化。其次,该计算机设备可以设置学习参数,使得所有网络参数都需要学习。该计算机设备可以将初始网络模型的学习率设置为0.005。进一步地,该计算机设备在对初始网络模型进行迭代训练时,可以进行多轮迭代,即该计算机设备可以将样本视频片段对划分多个批次,并将每个批次所抓取的一定数量的样本视频片段对作为当前的训练样本。

具体过程可以为:该计算机设备可以把初始网络模型的所有参数都设为学习状态,对该初始网络模型进行迭代训练时,需要对当前批次所抓取的样本视频片段对的样本编码特征(例如,第一样本编码特征以及第二样本编码特征)进行前向计算,以得到预测循环结果,进而可以根据此样本视频片段对的预测循环结果计算损失函数,把模型损失值回传到初始网络模型中,进而可以通过采用SGD随机梯度下降法,后向计算并得到全部模型参数的更新值,以更新模型参数,从而实现一次权重优化。

其中,该计算机设备还可以获取与初始网络模型相关联的模型收敛条件,其中,该模型收敛条件可以为模型损失持续N轮(例如,10轮)未继续下降,即停止模型训练。可选的,

该模型收敛条件还可以为模型损失小于模型收敛条件中的损失阈值，即停止模型训练。这里将不对其进行限定。

可以理解的是，若模型训练结果指示迭代训练后的初始网络模型满足模型收敛条件，则将满足模型收敛条件的初始网络模型作为用于对待识别视频进行循环识别的目标网络模型。可选的，若模型训练结果指示迭代训练后的初始网络模型不满足模型收敛条件，则该计算机设备可以基于不满足模型收敛条件的模型损失函数，对初始网络模型的模型参数进行调整。进一步地，该计算机设备可以将调整模型参数后的初始网络模型作为过渡网络模型，对过渡网络模型进行迭代训练，直到迭代训练后的过渡网络模型满足模型收敛条件时，将满足模型收敛条件的过渡网络模型作为用于对待识别视频进行循环识别的目标网络模型。通过上述对模型训练可信度的有效评估，可以识别出不同迭代阶段的模型精度，从而提高模型训练效率。

由此可见，本申请实施例的计算机设备通过对样本视频片段对的样本编码特征(即第一样本编码特征和第二样本编码特征)，对初始网络模型进行迭代，以得到用于对待识别视频进行循环识别的目标网络模型，在出现大量待识别视频时，无需人工参与视频循环识别，可以直接通过该目标网络模型，快速确定待识别视频的视频类型，以判断该待识别视频的视频类型，进而可以提高视频循环识别的效率。此外，由于训练好的目标网络模型是基于不同模态信息的样本编码特征所训练的，因此，训练好的目标网络模型在进行循环识别时，能够充分利用待识别视频的多模态信息，提高循环识别的准确率。可以理解的是，由于目标网络模型可以基于样本视频片段对在多模态信息下的样本编码特征进行训练，进而可以快速支持更丰富的业务扩展，即根据业务需求，定制不同类型目标网络模型。例如，该计算机设备可以训练一个关注视频模态信息、音频模态信息以及台词文本模态信息的目标网络模型。又比如，该计算机设备还可以训练一个关注音频模态信息、视频标题模态信息以及封面模态信息的目标网络模型，这里将不对其进行限定。

进一步地，请参见图9，图9是本申请实施例提供的一种视频循环识别装置的结构示意图。如图9所示，该视频循环识别装置1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)，例如，该视频循环识别装置1为一个应用软件；该视频循环识别装置1可以用于执行本申请实施例提供的方法中的相应步骤。如图9所示，该视频循环识别装置1可以运行于具有视频循环识别功能的计算机设备，该计算机设备可以为上述图1所对应实施例中的服务器10，也可以为上述图1所对应实施例中的用户终端集群中的任意一个运行有目标网络模型的用户终端，例如，用户终端100a。该视频循环识别装置1可以包括：目标编码特征获取模块11，目标网络模型获取模块12，第一目标相似结果确定模块13，第二目标相似结果确定模块14，目标相似结果比对模块15，视频类型确定模块16，循环提示信息生成模块17，样本视频对获取模块18，样本编码特征获取模块19，第一预测相似结果确定模块20，第二预测相似结果确定模块21，预测相似结果比对模块22以及迭代训练模块23。

该目标编码特征获取模块11，用于获取待识别视频的目标视频片段对，确定目标视频片段对的第一目标编码特征和第二目标编码特征；第一目标编码特征所对应的第一模态信息不同于第二目标编码特征所对应的第二模态信息；

-24-

其中,该目标编码特征获取模块11包括:分段处理单元111,目标视频对确定单元112,第一特征提取单元113以及第二特征提取单元114。

该分段处理单元111,用于在获取到待识别视频时,确定待识别视频的视频时长,且基于视频时长对待识别视频进行分段处理,得到N个视频片段;N为正整数;

5 该目标视频对确定单元112,用于从N个视频片段中获取视频片段 $P_i$ 和视频片段 $P_j$ ,将视频片段 $P_i$ 和视频片段 $P_j$ 作为待识别视频的目标视频片段对; $i$ 和 $j$ 均为小于或者等于N的正整数,且 $i$ 不等于 $j$ ;

该第一特征提取单元113,用于对目标视频片段对中的每个视频片段分别进行第一特征提取,得到目标视频片段对的第一目标编码特征;

10 其中,第一模态信息为视频模态信息;

该第一特征提取单元113包括:待处理视频帧确定子单元1131,待编码视频帧确定子单元1132,视频编码特征确定子单元1133以及第一目标编码特征确定子单元1134。

该待处理视频帧确定子单元1131,用于将目标视频片段对中的每个视频片段对应的视频帧分别作为待处理视频帧,且基于待处理视频帧的帧率,确定抽帧参数;

15 该待编码视频帧确定子单元1132,用于基于抽帧参数对待处理视频帧进行抽帧处理,得到与待处理视频帧相关联的待编码视频帧;

该视频编码特征确定子单元1133,用于获取与视频模态信息相关联的视频编码模型,将待编码视频帧输入至视频编码模型,通过视频编码模型对待编码视频帧进行编码处理,得到待编码视频帧对应的视频编码特征;

20 该第一目标编码特征确定子单元1134,用于基于待编码视频帧对应的视频编码特征,得到目标视频片段对的第一目标编码特征;第一目标编码特征包括视频片段 $P_i$ 对应的视频编码特征 $S_i$ ,以及视频片段 $P_j$ 对应的视频编码特征 $S_j$ 。

其中,该待处理视频帧确定子单元1131,待编码视频帧确定子单元1132,视频编码特征确定子单元1133以及第一目标编码特征确定子单元1134的具体实现方式可以参见上述图  
25 3所对应实施例中对第一目标编码特征的描述,这里将不再继续进行赘述。

该第二特征提取单元114,用于对目标视频片段对中的每个视频片段分别进行第二特征提取,得到目标视频片段对的第二目标编码特征。

其中,第二模态信息为音频模态信息;

30 该第二特征提取单元114包括:待编码音频帧确定子单元1141,音频编码特征确定子单元1142以及第二目标编码特征确定子单元1143。

该待编码音频帧确定子单元1141,用于将目标视频片段对中的每个视频片段对应的音频帧分别作为待处理音频帧,且对待处理音频帧进行音频准备处理,得到与待处理音频帧相关联的待编码音频帧;

35 该音频编码特征确定子单元1142,用于获取与音频模态信息相关联的音频编码模型,将待编码音频帧输入至音频编码模型,通过音频编码模型对待编码音频帧进行编码处理,得到待编码音频帧对应的音频编码特征;

该第二目标编码特征确定子单元1143,用于基于待编码音频帧对应的音频编码特征,

得到目标视频片段对的第二目标编码特征；第二目标编码特征包括视频片段 $P_i$ 对应的音频编码特征 $Y_i$ ，以及视频片段 $P_j$ 对应的音频编码特征 $Y_j$ 。

其中，该待编码音频帧确定子单元1141，音频编码特征确定子单元1142以及第二目标编码特征确定子单元1143的具体实现方式可以参见上述图3所对应实施例中对第二目标编码特征的描述，这里将不再继续进行赘述。

其中，该分段处理单元111，目标视频对确定单元112，第一特征提取单元113以及第二特征提取单元114的具体实现方式可以参见上述图3所对应实施例中对S101的描述，这里将不再继续进行赘述。

该目标网络模型获取模块12，用于获取用于对待识别视频进行循环识别的目标网络模型；目标网络模型包括与第一模态信息相关联的第一目标序列模型和与第二模态信息相关联的第二目标序列模型；

在一种可能的实现方式中，第一目标序列模型包括第一序列表征学习层和第一相似度量层；第二目标序列模型包括第二序列表征学习层和第二相似度量层；

该第一目标相似结果确定模块13，用于将第一目标编码特征输入至第一目标序列模型，输出目标视频片段对的第一目标相似结果；

在一种可能的实现方式中，该第一目标相似结果确定模块13还用于将第一目标编码特征输入至第一目标序列模型，通过第一序列表征学习层对第一目标编码特征进行序列特征学习，将序列特征学习到的第一目标学习特征输入至第一相似度量层，由第一相似度量层输出目标视频片段对的第一目标相似结果；

其中，目标视频片段对包括视频片段 $P_i$ 和视频片段 $P_j$ ； $i$ 和 $j$ 均为小于或者等于 $N$ 的正整数，且 $i$ 不等于 $j$ ； $N$ 为待识别视频中的视频片段的总数量；第一序列表征学习层包括与视频片段 $P_i$ 相关联的第一网络层和与视频片段 $P_j$ 相关联的第二网络层，且第一网络层与第二网络层具有相同的网络结构；

该第一目标相似结果确定模块13包括：目标编码特征输入单元131，第一序列特征学习单元132，第二序列特征学习单元133，相似度输出单元134以及相似结果确定单元135。

该目标编码特征输入单元131，用于将第一目标编码特征输入至第一目标序列模型；第一目标编码特征包括视频编码特征 $S_i$ 和视频编码特征 $S_j$ ；视频编码特征 $S_i$ 为视频片段 $P_i$ 在第一模态信息下的编码特征；视频编码特征 $S_j$ 为视频片段 $P_j$ 在第一模态信息下的编码特征；

该第一序列特征学习单元132，用于通过第一序列表征学习层中的第一网络层，对视频编码特征 $S_i$ 进行序列特征学习，得到视频编码特征 $S_i$ 对应的学习特征 $X_i$ ；

其中，第一网络层包括第一子网络层、第二子网络层、第三子网络层以及第四子网络层；

该第一序列特征学习单元132包括：第一特征转换子单元1321，第二特征转换子单元1322，第三特征转换子单元1323以及最大池化处理子单元1324。

该第一特征转换子单元1321，用于通过第一网络层中的第一子网络层，对视频编码特征 $S_i$ 进行特征转换，得到视频编码特征 $S_i$ 对应的第一转换特征；

其中，第一子网络层包括第一卷积层、第二卷积层以及空洞卷积层；

该第一特征转换子单元1321还用于:

在将视频编码特征 $S_i$ 输入至第一网络层中的第一子网络层时,通过空洞卷积层对视频编码特征 $S_i$ 进行卷积处理,得到视频编码特征 $S_i$ 对应的第一卷积特征;

通过第一卷积层,对视频编码特征 $S_i$ 进行卷积处理,得到视频编码特征 $S_i$ 对应的第二卷积特征,且将第二卷积特征输入至第二卷积层,通过第二卷积层对第二卷积特征进行卷积处理,得到第三卷积特征;

将第一卷积特征和第三卷积特征进行拼接处理,得到视频编码特征 $S_i$ 对应的第一转换特征。

该第二特征转换子单元1322,用于将第一转换特征输入至第二子网络层,通过第二子网络层对第一转换特征进行特征转换,得到第一转换特征对应的第二转换特征;

该第三特征转换子单元1323,用于将第二转换特征输入至第三子网络层,通过第三子网络层对第二转换特征进行特征转换,得到第二转换特征对应的第三转换特征;

该最大池化处理子单元1324,用于将第三转换特征输入至第四子网络层,通过第四子网络层中的最大池化层,对第三转换特征进行最大池化处理,得到视频编码特征 $S_i$ 对应的学习特征 $X_i$ 。

其中,该第一特征转换子单元1321,第二特征转换子单元1322,第三特征转换子单元1323以及最大池化处理子单元1324的具体实现方式可以参见上述图6所对应实施例中对学习特征的描述,这里将不再继续进行赘述。

该第二序列特征学习单元133,用于通过第一序列表征学习层中的第二网络层,对视频编码特征 $S_j$ 进行序列特征学习,得到视频编码特征 $S_j$ 对应的学习特征 $X_j$ ;

该相似度输出单元134,用于将学习特征 $X_i$ 和学习特征 $X_j$ 作为第一目标学习特征,且将第一目标学习特征输入至第一相似度度量层,由第一相似度度量层输出第一目标学习特征的相似度;

该相似结果确定单元135,用于基于第一目标学习特征的相似度,确定目标视频片段对的第一目标相似结果。

其中,该目标编码特征输入单元131,第一序列特征学习单元132,第二序列特征学习单元133,相似度输出单元134以及相似结果确定单元135的具体实现方式可以参见上述图3所对应实施例中对S103的描述,这里将不再继续进行赘述。

该第二目标相似结果确定模块14,用于将第二目标编码特征输入至第二目标序列模型,输出目标视频片段对的第二目标相似结果;

在一种可能的实现方式中,该第二目标相似结果确定模块14还用于将第二目标编码特征输入至第二目标序列模型,通过第二序列表征学习层对第二目标编码特征进行序列特征学习,将序列特征学习到的第二目标学习特征输入至第二相似度度量层,由第二相似度度量层输出目标视频片段对的第二目标相似结果;

该目标相似结果比对模块15,用于将第一目标相似结果和第二目标相似结果进行比对,得到目标视频片段对的循环比对结果;循环比对结果用于指示待识别视频的视频类型。

其中,该目标相似结果比对模块15包括: 比对单元151,第一结果确定单元152,第二

结果确定单元153以及循环比对结果确定单元154。

该比对单元151, 用于将第一目标相似结果和第二目标相似结果进行比对;

该第一结果确定单元152, 用于若第一目标相似结果指示目标视频片段对在第一模态信息下相似, 且第二目标相似结果指示目标视频片段对在第二模态信息下相似, 则得到目标视频片段对的循环视频结果;

该第二结果确定单元153, 用于若第一目标相似结果指示目标视频片段对在第一模态信息下不相似, 或者第二目标相似结果指示目标视频片段对在第二模态信息下不相似, 则得到目标视频片段对的非循环视频结果;

该循环比对结果确定单元154, 用于将循环视频结果或者非循环视频结果作为目标视频片段对的循环比对结果。

其中, 该比对单元151, 第一结果确定单元152, 第二结果确定单元153以及循环比对结果确定单元154的具体实现方式可以参见上述图3所对应实施例中对S105的描述, 这里将不再继续进行赘述。

该视频类型确定模块16, 用于若循环比对结果为循环视频结果, 则确定待识别视频的视频类型为循环视频类型;

该循环提示信息生成模块17, 用于基于循环视频类型生成循环提示信息, 将循环提示信息返回至用户终端; 用户终端为待识别视频的发送端。

该样本视频对获取模块18, 用于获取用于训练初始网络模型的样本视频片段对以及样本视频片段对的样本标签; 初始网络模型包括第一初始序列模型和第二初始序列模型; 第一初始序列模型包括第一序列表征学习层和第一相似度度量层; 第二初始序列模型包括第二序列表征学习层和第二相似度度量层;

该样本编码特征获取模块19, 用于分别获取样本视频片段对在第一模态信息下的第一样本编码特征和在第二模态信息下的第二样本编码特征;

该第一预测相似结果确定模块20, 用于将第一样本编码特征输入至第一初始序列模型, 通过第一初始序列模型中的第一序列表征学习层和第一相似度度量层, 输出样本视频片段对的第一预测相似结果;

该第二预测相似结果确定模块21, 用于将第二样本编码特征输入至第二初始序列模型, 通过第二初始序列模型中的第二序列表征学习层和第二相似度度量层, 输出样本视频片段对的第二预测相似结果;

该预测相似结果比对模块22, 用于将第一预测相似结果和第二预测相似结果进行比对, 得到样本视频片段对的预测循环结果对应的预测标签;

该迭代训练模块23, 用于基于预测标签和样本标签, 对初始网络模型进行迭代训练, 将迭代训练后的初始网络模型作为用于对待识别视频进行循环识别的目标网络模型。

其中, 样本视频片段对包括正样本视频片段对和负样本视频片段对; 正样本视频片段对为携带第一样本标签的视频片段对; 负样本视频片段对为携带第二样本标签的视频片段对; 第一样本标签和第二样本标签均属于样本标签; 预测循环结果包括正样本视频片段对的第一预测循环结果和负样本视频片段对的第二预测循环结果; 预测标签包括第一预测循

环结果对应的第一预测标签和第二预测循环结果对应的第二预测标签;

该迭代训练模块23包括: 损失权重参数确定单元231, 样本损失确定单元232, 迭代训练单元233, 第一模型确定单元234, 参数调整单元235以及第二模型确定单元236。

该损失权重参数确定单元231, 用于基于样本标签所指示的正样本视频片段对和负样本视频片段对之间的样本比例, 确定与初始网络模型的模型损失函数相关联的损失权重参数;

该样本损失确定单元232, 用于基于第一预测标签和第一样本标签, 得到正样本视频片段对的正样本损失, 且基于第二预测标签和第二样本标签, 得到负样本视频片段对的负样本损失;

该迭代训练单元233, 用于基于正样本损失、负样本损失以及损失权重参数, 得到模型损失函数对应的模型损失, 且基于模型损失, 对初始网络模型进行迭代训练, 得到模型训练结果;

该第一模型确定单元234, 用于若模型训练结果指示迭代训练后的初始网络模型满足模型收敛条件, 则将满足模型收敛条件的初始网络模型作为用于对待识别视频进行循环识别的目标网络模型。

该参数调整单元235, 用于若模型训练结果指示迭代训练后的初始网络模型不满足模型收敛条件, 则基于不满足模型收敛条件的模型损失函数, 对初始网络模型的模型参数进行调整;

该第二模型确定单元236, 用于将调整模型参数后的初始网络模型作为过渡网络模型, 对过渡网络模型进行迭代训练, 直到迭代训练后的过渡网络模型满足模型收敛条件时, 将满足模型收敛条件的过渡网络模型作为用于对待识别视频进行循环识别的目标网络模型。

其中, 该损失权重参数确定单元231, 样本损失确定单元232, 迭代训练单元233, 第一模型确定单元234, 参数调整单元235以及第二模型确定单元236的具体实现方式可以参见上述图7所对应实施例中对S211的描述, 这里将不再继续进行赘述。

其中, 该目标编码特征获取模块11, 目标网络模型获取模块12, 第一目标相似结果确定模块13, 第二目标相似结果确定模块14, 目标相似结果比对模块15, 视频类型确定模块16, 循环提示信息生成模块17, 样本视频对获取模块18, 样本编码特征获取模块19, 第一预测相似结果确定模块20, 第二预测相似结果确定模块21, 预测相似结果比对模块22以及迭代训练模块23的具体实现方式可以参见上述图7所对应实施例中对S201-S211的描述, 这里将不再继续进行赘述。另外, 对采用相同方法的有益效果描述, 也不再赘述。

进一步地, 请参见图10, 图10是本申请实施例提供的一种计算机设备的示意图。如图10所示, 该计算机设备1000可以为具备视频循环识别功能的计算机设备, 该计算机设备1000可以包括: 至少一个处理器1001, 例如, CPU, 至少一个网络接口1004, 存储器1005, 至少一个通信总线1002。其中, 通信总线1002用于实现这些组件之间的连接通信。网络接口1004可选地可以包括标准的有线接口、无线接口(如WI-FI接口)。存储器1005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。存储器1005可选地还可以是至少一个位于远离前述处理器1001的存储装置。如图10所示, 作为一种计算机存储介质的存储器1005可以包括操作系统、网络通信模块、用户接口模块

以及设备控制应用程序。其中，在一些实施例中，该计算机设备还可以包括图10所示的用户接口1003，比如，若该计算机设备为图1所示的具有视频循环识别功能的用户终端(例如，用户终端100a)，则该计算机设备还可以包括该用户接口1003，其中，该用户接口1003可以包括显示屏(Display)、键盘(Keyboard)等。

5 在图10所示的计算机设备1000中，网络接口1004主要用于进行网络通信；而用户接口1003主要用于为用户提供输入的接口；而处理器1001可以用于调用存储器1005中存储的设备控制应用程序，以实现：

获取待识别视频的目标视频片段对，确定目标视频片段对的第一目标编码特征和第二目标编码特征；第一目标编码特征所对应的第一模态信息不同于第二目标编码特征所对应的第二模态信息；

10 获取用于对待识别视频进行循环识别的目标网络模型；目标网络模型包括与第一模态信息相关联的第一目标序列模型和与第二模态信息相关联的第二目标序列模型；

将第一目标编码特征输入至第一目标序列模型，输出目标视频片段对的第一目标相似结果；

15 将第二目标编码特征输入至第二目标序列模型，输出目标视频片段对的第二目标相似结果；

将第一目标相似结果和第二目标相似结果进行比对，得到目标视频片段对的循环比对结果；循环比对结果用于指示待识别视频的视频类型。

应当理解，本申请实施例中所描述的计算机设备1000可执行前文图3和图7所对应实施例中对该视频循环识别方法的描述，也可执行前文图9所对应实施例中对该视频循环识别装置1的描述，在此不再赘述。另外，对采用相同方法的有益效果描述，也不再进行赘述。

20 本申请实施例还提供一种计算机可读存储介质，该计算机可读存储介质存储有计算机程序，该计算机程序包括程序指令，该程序指令被处理器执行时实现图3和图7中各个步骤所提供的视频循环识别方法，具体可参见图3以及图7各个步骤所提供的实现方式，在此不再赘述。

25 计算机可读存储介质可以是前述任一实施例提供的数据传输装置或者计算机设备的内部存储单元，例如计算机设备的硬盘或内存。该计算机可读存储介质也可以是该计算机设备的外部存储设备，例如该计算机设备上配备的插接式硬盘，智能存储卡(smart media card, SMC)，安全数字(secure digital, SD)卡，闪存卡(flash card)等。进一步地，该计算机可读存储介质还可以既包括该计算机设备的内部存储单元也包括外部存储设备。该计算机可读存储介质用于存储该计算机程序以及该计算机设备所需的其他程序和数据。该计算机可读存储介质还可以用于暂时地存储已经输出或者将要输出的数据。

30 本申请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，使得该计算机设备可执行前文图3或者图7所对应实施例中对该视频循环识别方法的描述，在此不再赘述。另外，对采用相同方法的有益效果描述，也不再进行赘述。

35

本申请实施例的说明书和权利要求书及附图中的术语“第一”、“第二”等是用于区别不同对象，而非用于描述特定顺序。此外，术语“包括”以及它们任何变形，意图在于覆盖不排他的包括。例如包括了一系列步骤或单元的过程、方法、装置、产品或设备没有限定于已列出的步骤或模块，而是可选地还包括没有列出的步骤或模块，或可选地还包括对于

5 于这些过程、方法、装置、产品或设备固有的其他步骤单元。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业

10 技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

本申请实施例提供的方法及相关装置是参照本申请实施例提供的方法流程图和/或结构示意图来描述的，具体可由计算机程序指令实现方法流程图和/或结构示意图的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。这些计算机程序指令可提供到通用计算机、专用计算机、嵌入式处理机或其他可编程视频循环识别设备的处理器以

15 产生一个机器，使得通过计算机或其他可编程视频循环识别设备的处理器执行的指令产生用于实现在流程图中的一个流程或多个流程和/或结构示意图中的一个方框或多个方框中指定的功能的装置。这些计算机程序指令也可存储在能引导计算机或其他可编程视频循环识别设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图中的一个流程或多个流程和/或结构示意图中的一个方框或多个方框中指定的功能。这些计算机程序指令也可装载到计算机或其他可编程视频循环识别设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图中的一个流程或多个流程和/或结构示意图中的一个方框或多个方框中指定的功能的步骤。

20 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。

## 权 利 要 求

1、一种视频循环识别方法，所述方法由计算机设备执行，所述方法包括：

获取待识别视频的目标视频片段对，确定所述目标视频片段对的第一目标编码特征和第二目标编码特征；所述第一目标编码特征所对应的第一模态信息不同于所述第二目标编码特征所对应的第二模态信息；

获取用于对所述待识别视频进行循环识别的目标网络模型；所述目标网络模型包括与所述第一模态信息相关联的第一目标序列模型和与所述第二模态信息相关联的第二目标序列模型；

将所述第一目标编码特征输入至所述第一目标序列模型，输出所述目标视频片段对的第一目标相似结果；

将所述第二目标编码特征输入至所述第二目标序列模型，输出所述目标视频片段对的第二目标相似结果；

将所述第一目标相似结果和所述第二目标相似结果进行比对，得到所述目标视频片段对的循环比对结果；所述循环比对结果用于指示所述待识别视频的视频类型。

2、根据权利要求1所述的方法，所述第一目标序列模型包括第一序列表征学习层和第一相似度量层；所述第二目标序列模型包括第二序列表征学习层和第二相似度量层，所述将所述第一目标编码特征输入至所述第一目标序列模型，输出所述目标视频片段对的第一目标相似结果，包括：

将所述第一目标编码特征输入至所述第一目标序列模型，通过所述第一序列表征学习层对所述第一目标编码特征进行序列特征学习，将序列特征学习到的第一目标学习特征输入至所述第一相似度量层，由所述第一相似度量层输出所述目标视频片段对的第一目标相似结果；

所述将所述第二目标编码特征输入至所述第二目标序列模型，输出所述目标视频片段对的第二目标相似结果，包括：

将所述第二目标编码特征输入至所述第二目标序列模型，通过所述第二序列表征学习层对所述第二目标编码特征进行序列特征学习，将序列特征学习到的第二目标学习特征输入至所述第二相似度量层，由所述第二相似度量层输出所述目标视频片段对的第二目标相似结果；

3、根据权利要求1所述的方法，所述获取待识别视频的目标视频片段对，确定所述目标视频片段对的第一目标编码特征和第二目标编码特征，包括：

在获取到待识别视频时，确定所述待识别视频的视频时长，且基于所述视频时长对所述待识别视频进行分段处理，得到N个视频片段；所述N为正整数；

从所述N个视频片段中获取视频片段 $P_i$ 和视频片段 $P_j$ ，将所述视频片段 $P_i$ 和所述视频片段 $P_j$ 作为所述待识别视频的目标视频片段对；所述i和所述j均为小于或者等于所述N的正整数，且所述i不等于所述j；

对所述目标视频片段对中的每个视频片段分别进行第一特征提取，得到所述目标视频片段对的第一目标编码特征；

— 32 —

对所述目标视频片段对中的每个视频片段分别进行第二特征提取，得到所述目标视频片段对的第二目标编码特征。

4、根据权利要求3所述的方法，所述第一模态信息为视频模态信息；

5 所述对所述目标视频片段对中的每个视频片段分别进行第一特征提取，得到所述目标视频片段对的第一目标编码特征，包括：

将所述目标视频片段对中的每个视频片段对应的视频帧分别作为待处理视频帧，且基于所述待处理视频帧的帧率，确定抽帧参数；

基于所述抽帧参数对所述待处理视频帧进行抽帧处理，得到与所述待处理视频帧相关联的待编码视频帧；

10 获取与所述视频模态信息相关联的视频编码模型，将所述待编码视频帧输入至所述视频编码模型，通过所述视频编码模型对所述待编码视频帧进行编码处理，得到所述待编码视频帧对应的视频编码特征；

15 基于所述待编码视频帧对应的视频编码特征，得到所述目标视频片段对的第一目标编码特征；所述第一目标编码特征包括所述视频片段 $P_i$ 对应的视频编码特征 $S_i$ ，以及所述视频片段 $P_j$ 对应的视频编码特征 $S_j$ 。

5、根据权利要求3所述的方法，所述第二模态信息为音频模态信息；

所述对所述目标视频片段对中的每个视频片段分别进行第二特征提取，得到所述目标视频片段对的第二目标编码特征，包括：

20 将所述目标视频片段对中的每个视频片段对应的音频帧分别作为待处理音频帧，且对所述待处理音频帧进行音频准备处理，得到与所述待处理音频帧相关联的待编码音频帧；

获取与所述音频模态信息相关联的音频编码模型，将所述待编码音频帧输入至所述音频编码模型，通过所述音频编码模型对所述待编码音频帧进行编码处理，得到所述待编码音频帧对应的音频编码特征；

25 基于所述待编码音频帧对应的音频编码特征，得到所述目标视频片段对的第二目标编码特征；所述第二目标编码特征包括所述视频片段 $P_i$ 对应的音频编码特征 $Y_i$ ，以及所述视频片段 $P_j$ 对应的音频编码特征 $Y_j$ 。

30 6、根据权利要求2所述的方法，所述目标视频片段对包括视频片段 $P_i$ 和视频片段 $P_j$ ；所述 $i$ 和所述 $j$ 均为小于或者等于 $N$ 的正整数，且所述 $i$ 不等于所述 $j$ ；所述 $N$ 为所述待识别视频中的视频片段的总数量；所述第一序列表征学习层包括与所述视频片段 $P_i$ 相关联的第一网络层和与所述视频片段 $P_j$ 相关联的第二网络层，且所述第一网络层与所述第二网络层具有相同的网络结构；

35 所述将所述第一目标编码特征输入至所述第一目标序列模型，通过所述第一序列表征学习层对所述第一目标编码特征进行序列特征学习，将序列特征学习到的第一目标学习特征输入至所述第一相似度度量层，由所述第一相似度度量层输出所述目标视频片段对的第一目标相似结果，包括：

将所述第一目标编码特征输入至所述第一目标序列模型；所述第一目标编码特征包括视频编码特征 $S_i$ 和视频编码特征 $S_j$ ；所述视频编码特征 $S_i$ 为所述视频片段 $P_i$ 在所述第一模态

信息下的编码特征；所述视频编码特征 $S_j$ 为所述视频片段 $P_j$ 在所述第一模态信息下的编码特征；

通过所述第一序列表征学习层中的所述第一网络层，对所述视频编码特征 $S_i$ 进行序列特征学习，得到所述视频编码特征 $S_i$ 对应的学习特征 $X_i$ ；

5 通过所述第一序列表征学习层中的所述第二网络层，对所述视频编码特征 $S_j$ 进行序列特征学习，得到所述视频编码特征 $S_j$ 对应的学习特征 $X_j$ ；

将所述学习特征 $X_i$ 和所述学习特征 $X_j$ 作为第一目标学习特征，且将所述第一目标学习特征输入至所述第一相似度度量层，由所述第一相似度度量层输出所述第一目标学习特征的相似度；

10 基于所述第一目标学习特征的相似度，确定所述目标视频片段对的第一目标相似结果。

7、根据权利要求6所述的方法，所述第一网络层包括第一子网络层、第二子网络层、第三子网络层以及第四子网络层；

所述通过所述第一序列表征学习层中的所述第一网络层，对所述视频编码特征 $S_i$ 进行序列特征学习，得到所述视频编码特征 $S_i$ 对应的学习特征 $X_i$ ，包括：

15 通过所述第一网络层中的所述第一子网络层，对所述视频编码特征 $S_i$ 进行特征转换，得到所述视频编码特征 $S_i$ 对应的第一转换特征；

将所述第一转换特征输入至所述第二子网络层，通过所述第二子网络层对所述第一转换特征进行特征转换，得到所述第一转换特征对应的第二转换特征；

20 将所述第二转换特征输入至所述第三子网络层，通过所述第三子网络层对所述第二转换特征进行特征转换，得到所述第二转换特征对应的第三转换特征；

将所述第三转换特征输入至所述第四子网络层，通过所述第四子网络层中的最大池化层，对所述第三转换特征进行最大池化处理，得到所述视频编码特征 $S_i$ 对应的学习特征 $X_i$ 。

8、根据权利要求7所述的方法，所述第一子网络层包括第一卷积层、第二卷积层以及空洞卷积层；

25 所述通过所述第一网络层中的所述第一子网络层，对所述视频编码特征 $S_i$ 进行特征转换，得到所述视频编码特征 $S_i$ 对应的第一转换特征，包括：

在将所述视频编码特征 $S_i$ 输入至所述第一网络层中的所述第一子网络层时，通过所述空洞卷积层对所述视频编码特征 $S_i$ 进行卷积处理，得到所述视频编码特征 $S_i$ 对应的第一卷积特征；

30 通过所述第一卷积层，对所述视频编码特征 $S_i$ 进行卷积处理，得到所述视频编码特征 $S_i$ 对应的第二卷积特征，且将所述第二卷积特征输入至所述第二卷积层，通过所述第二卷积层对所述第二卷积特征进行卷积处理，得到第三卷积特征；

将所述第一卷积特征和所述第三卷积特征进行拼接处理，得到所述视频编码特征 $S_i$ 对应的第一转换特征。

35 9、根据权利要求1所述的方法，所述将所述第一目标相似结果和所述第二目标相似结果进行比对，得到所述目标视频片段对的循环比对结果，包括：

将所述第一目标相似结果和所述第二目标相似结果进行比对；

若所述第一目标相似结果指示所述目标视频片段对在所述第一模态信息下相似，且所述第二目标相似结果指示所述目标视频片段对在所述第二模态信息下相似，则得到所述目标视频片段对的循环视频结果；

若所述第一目标相似结果指示所述目标视频片段对在所述第一模态信息下不相似，或者所述第二目标相似结果指示所述目标视频片段对在所述第二模态信息下不相似，则得到所述目标视频片段对的非循环视频结果；

将所述循环视频结果或者所述非循环视频结果作为所述目标视频片段对的循环比对结果。

10、根据权利要求1所述的方法，所述方法还包括：

若所述循环比对结果为循环视频结果，则确定所述待识别视频的视频类型为循环视频类型；

基于所述循环视频类型生成循环提示信息，将所述循环提示信息返回至用户终端；所述用户终端为所述待识别视频的发送端。

11、根据权利要求2所述的方法，所述方法还包括：

获取用于训练初始网络模型的样本视频片段对以及所述样本视频片段对的样本标签；所述初始网络模型包括第一初始序列模型和第二初始序列模型；所述第一初始序列模型包括所述第一序列表征学习层和所述第一相似度度量层；所述第二初始序列模型包括所述第二序列表征学习层和所述第二相似度度量层；

分别获取所述样本视频片段对在所述第一模态信息下的第一样本编码特征和在所述第二模态信息下的第二样本编码特征；

将所述第一样本编码特征输入至所述第一初始序列模型，通过所述第一初始序列模型中的所述第一序列表征学习层和所述第一相似度度量层，输出所述样本视频片段对的第一预测相似结果；

将所述第二样本编码特征输入至所述第二初始序列模型，通过所述第二初始序列模型中的所述第二序列表征学习层和所述第二相似度度量层，输出所述样本视频片段对的第二预测相似结果；

将所述第一预测相似结果和所述第二预测相似结果进行比对，得到所述样本视频片段对的预测循环结果对应的预测标签；

基于所述预测标签和所述样本标签，对所述初始网络模型进行迭代训练，将迭代训练后的初始网络模型作为用于对所述待识别视频进行循环识别的所述目标网络模型。

12、根据权利要求11所述的方法，所述样本视频片段对包括正样本视频片段对和负样本视频片段对；所述正样本视频片段对为携带第一样本标签的视频片段对；所述负样本视频片段对为携带第二样本标签的视频片段对；所述第一样本标签和所述第二样本标签均属于所述样本标签；所述预测循环结果包括所述正样本视频片段对的第一预测循环结果和所述负样本视频片段对的第二预测循环结果；所述预测标签包括所述第一预测循环结果对应的第一预测标签和所述第二预测循环结果对应的第二预测标签；

所述基于所述预测标签和所述样本标签，对所述初始网络模型进行迭代训练，将迭代

训练后的初始网络模型作为用于对所述待识别视频进行循环识别的所述目标网络模型，包括：

基于所述样本标签所指示的所述正样本视频片段对和所述负样本视频片段对之间的样本比例，确定与所述初始网络模型的模型损失函数相关联的损失权重参数；

5 基于所述第一预测标签和所述第一样本标签，得到所述正样本视频片段对的正样本损失，且基于所述第二预测标签和所述第二样本标签，得到所述负样本视频片段对的负样本损失；

10 基于所述正样本损失、所述负样本损失以及所述损失权重参数，得到所述模型损失函数对应的模型损失，且基于所述模型损失，对所述初始网络模型进行迭代训练，得到模型训练结果；

若所述模型训练结果指示迭代训练后的初始网络模型满足模型收敛条件，则将满足所述模型收敛条件的初始网络模型作为用于对所述待识别视频进行循环识别的所述目标网络模型。

13、根据权利要求12所述的方法，所述方法还包括：

15 若所述模型训练结果指示迭代训练后的初始网络模型不满足模型收敛条件，则基于不满足所述模型收敛条件的所述模型损失函数，对所述初始网络模型的模型参数进行调整；

将调整模型参数后的初始网络模型作为过渡网络模型，对所述过渡网络模型进行迭代训练，直到迭代训练后的过渡网络模型满足所述模型收敛条件时，将满足所述模型收敛条件的过渡网络模型作为用于对所述待识别视频进行循环识别的所述目标网络模型。

20 14、一种视频循环识别装置，包括：

目标编码特征获取模块，用于获取待识别视频的目标视频片段对，确定所述目标视频片段对的第一目标编码特征和第二目标编码特征；所述第一目标编码特征所对应的第一模态信息不同于所述第二目标编码特征所对应的第二模态信息；

25 目标网络模型获取模块，用于获取用于对所述待识别视频进行循环识别的目标网络模型；所述目标网络模型包括与所述第一模态信息相关联的第一目标序列模型和与所述第二模态信息相关联的第二目标序列模型；

第一目标相似结果确定模块，用于将所述第一目标编码特征输入至所述第一目标序列模型，输出所述目标视频片段对的第一目标相似结果；

30 第二目标相似结果确定模块，用于将所述第二目标编码特征输入至所述第二目标序列模型，输出所述目标视频片段对的第二目标相似结果；

目标相似结果比对模块，用于将所述第一目标相似结果和所述第二目标相似结果进行比对，得到所述目标视频片段对的循环比对结果；所述循环比对结果用于指示所述待识别视频的视频类型。

15、一种计算机设备，包括：处理器和存储器；

35 所述处理器与存储器相连，其中，所述存储器用于存储计算机程序，所述处理器用于调用所述计算机程序，以使得所述计算机设备执行权利要求1-13任一项所述的方法。

—36—

16、一种计算机可读存储介质，所述计算机可读存储介质中存储有计算机程序，所述计算机程序适于由处理器加载并执行，以使得具有所述处理器的计算机设备执行权利要求1-13任一项所述的方法。

5 17、一种包括指令的计算机程序产品，当其在计算机上运行时，使得所述计算机执行权利要求1-13任意一项所述的方法。

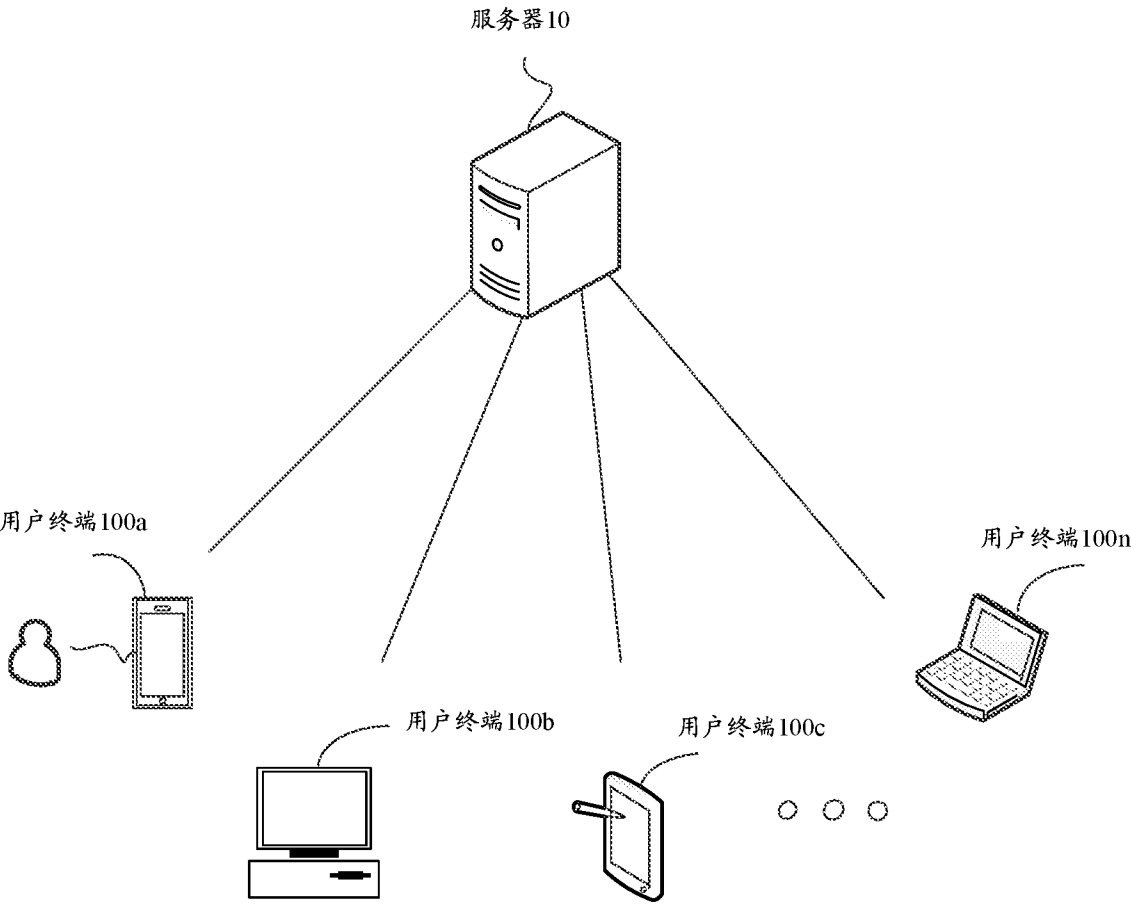


图 1

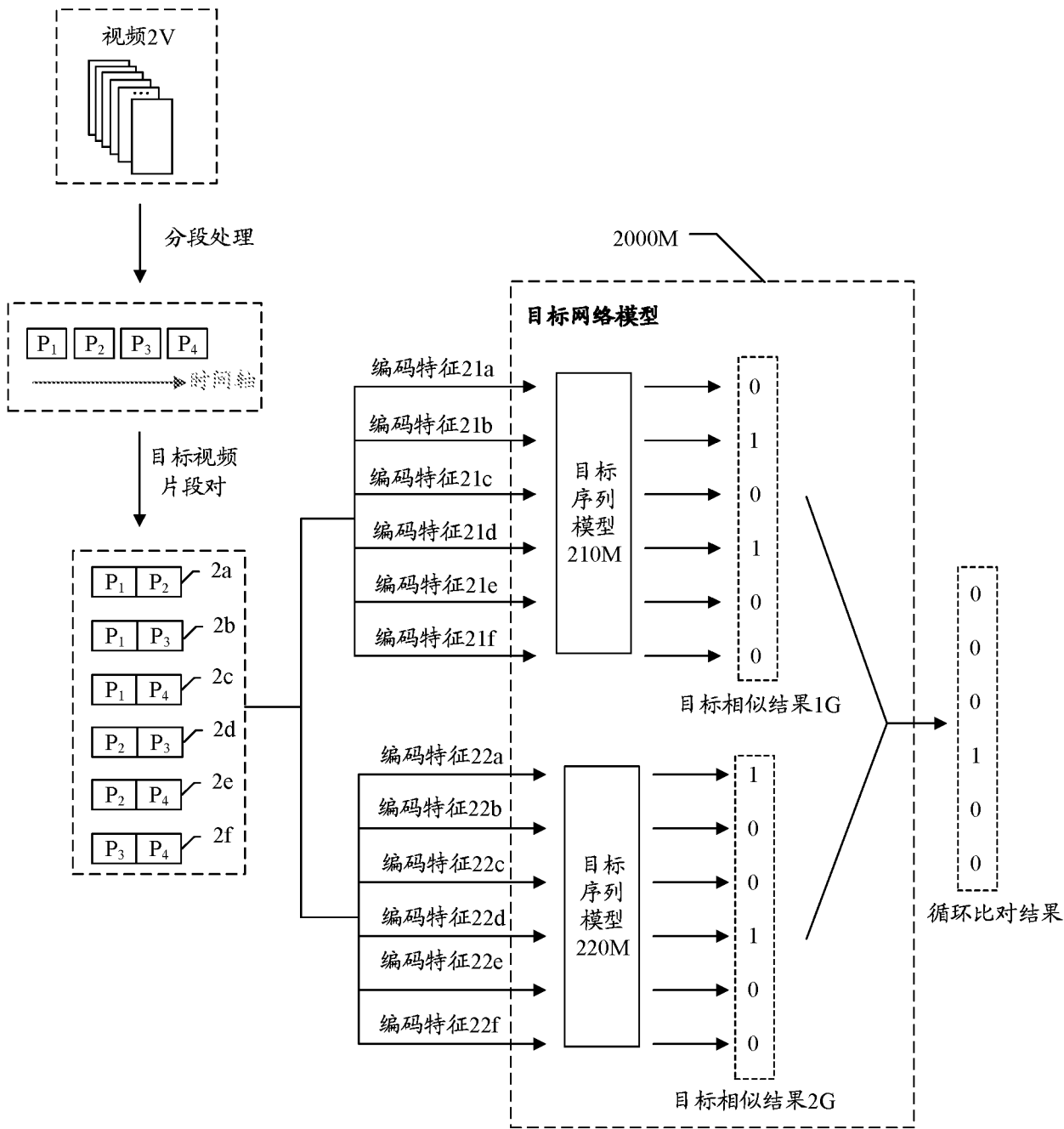


图 2

—3/10—

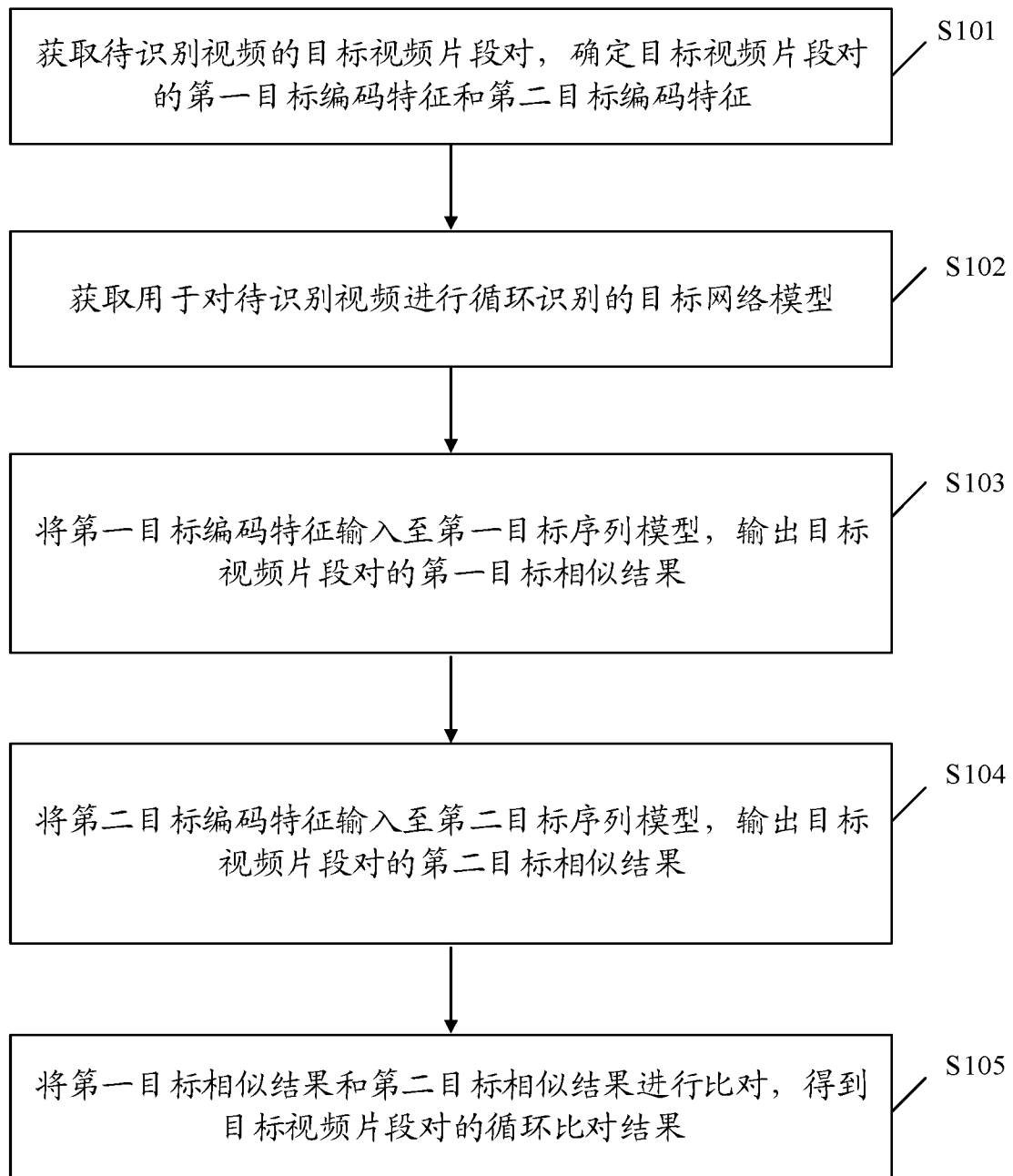


图 3

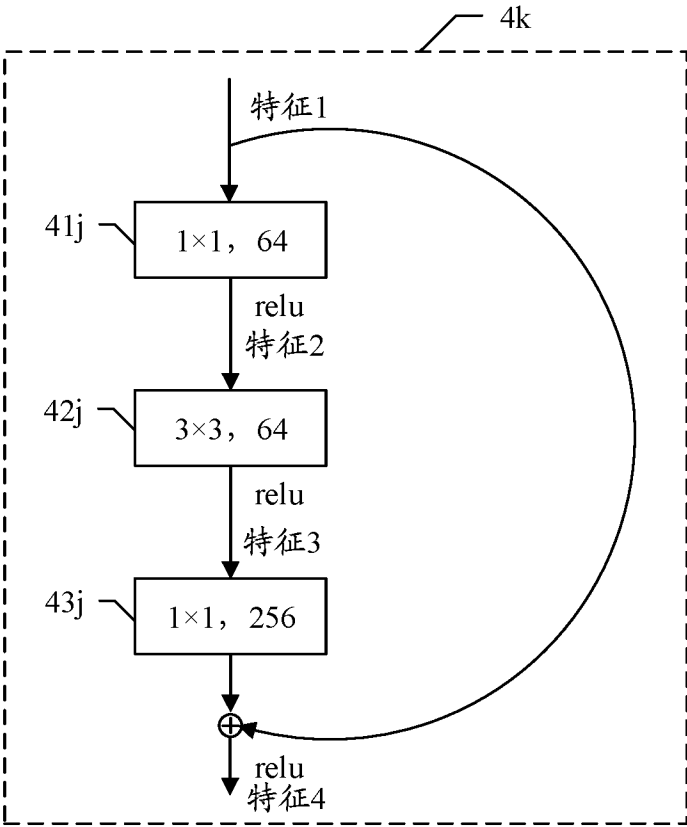


图 4

—5/10—

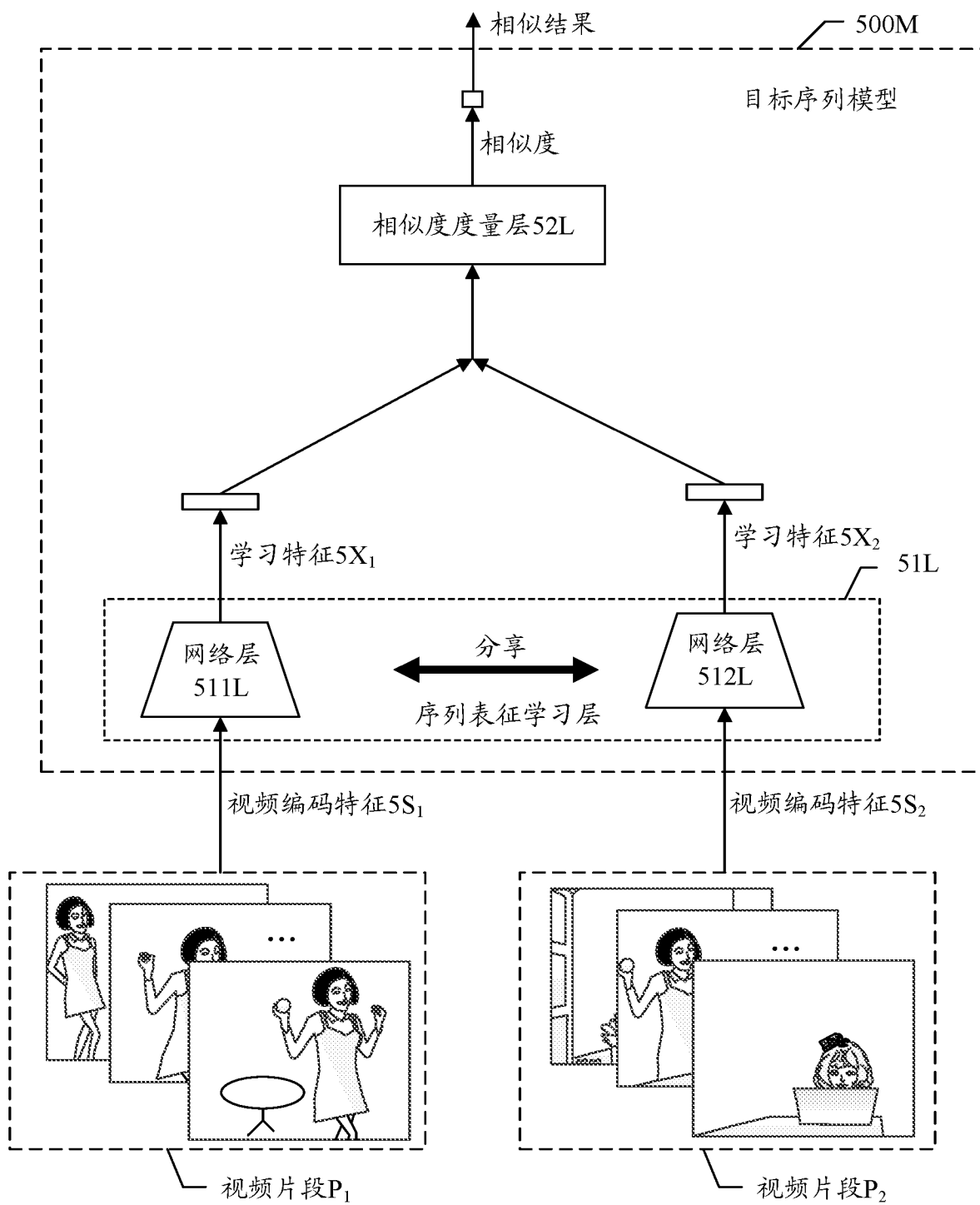


图 5

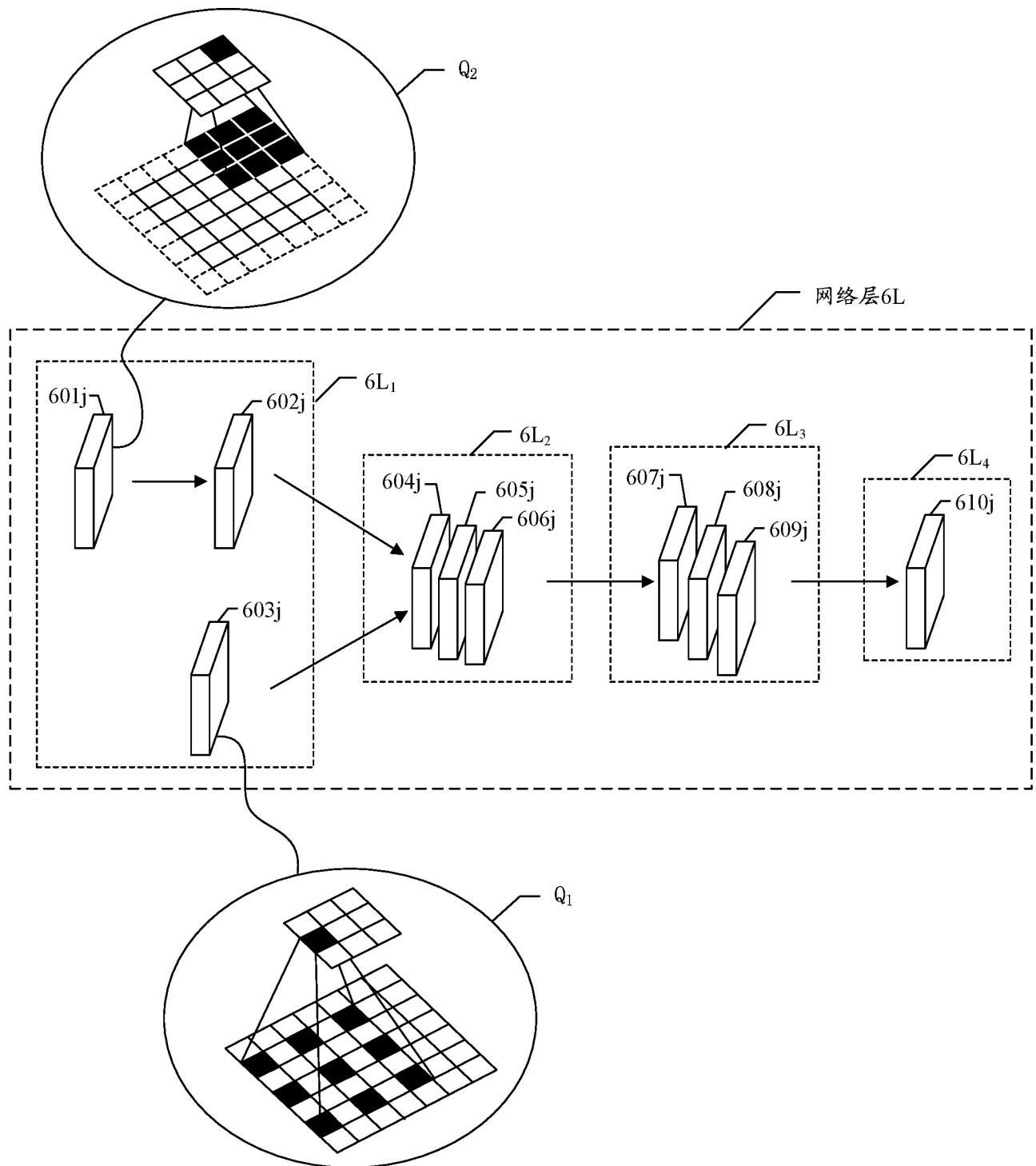


图 6

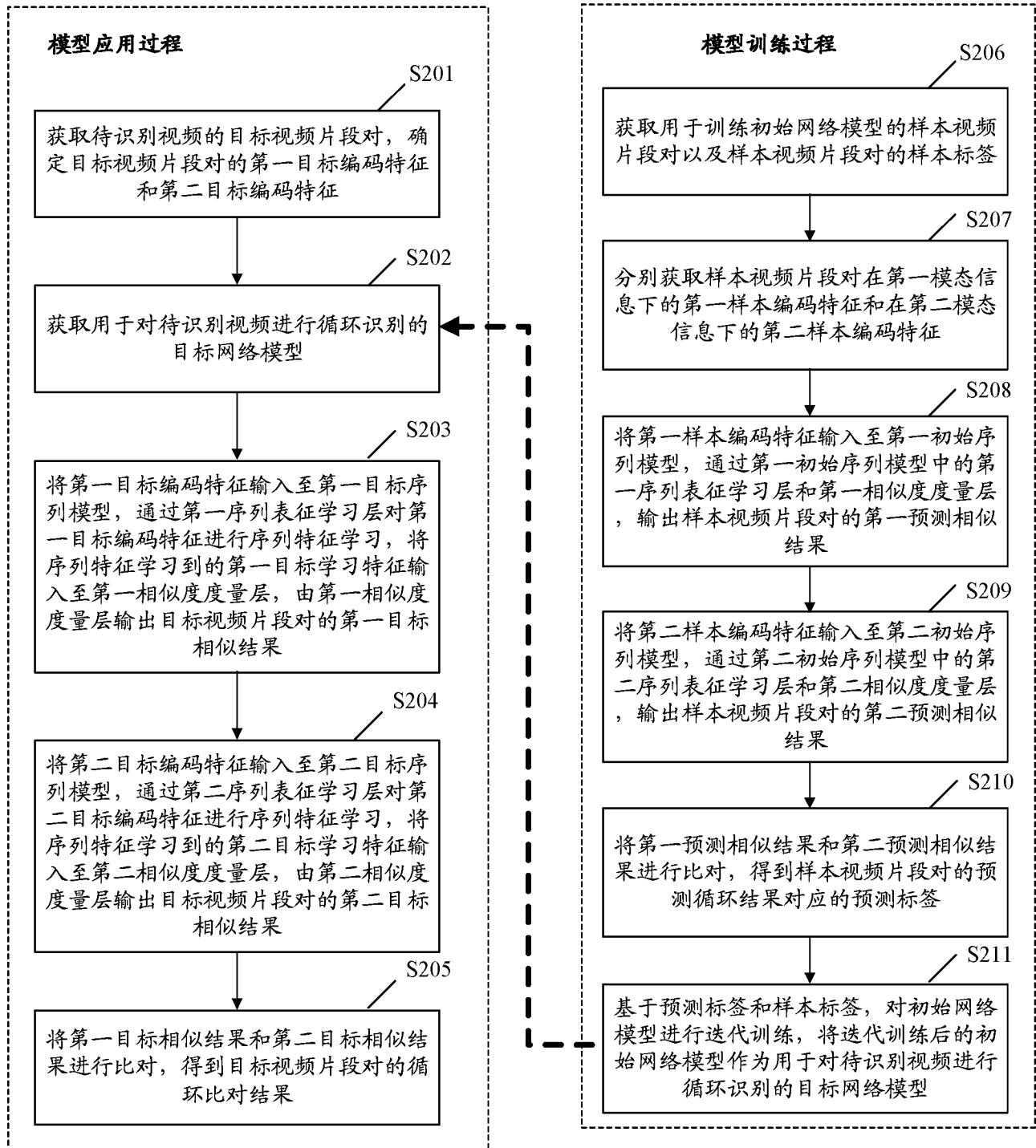


图 7

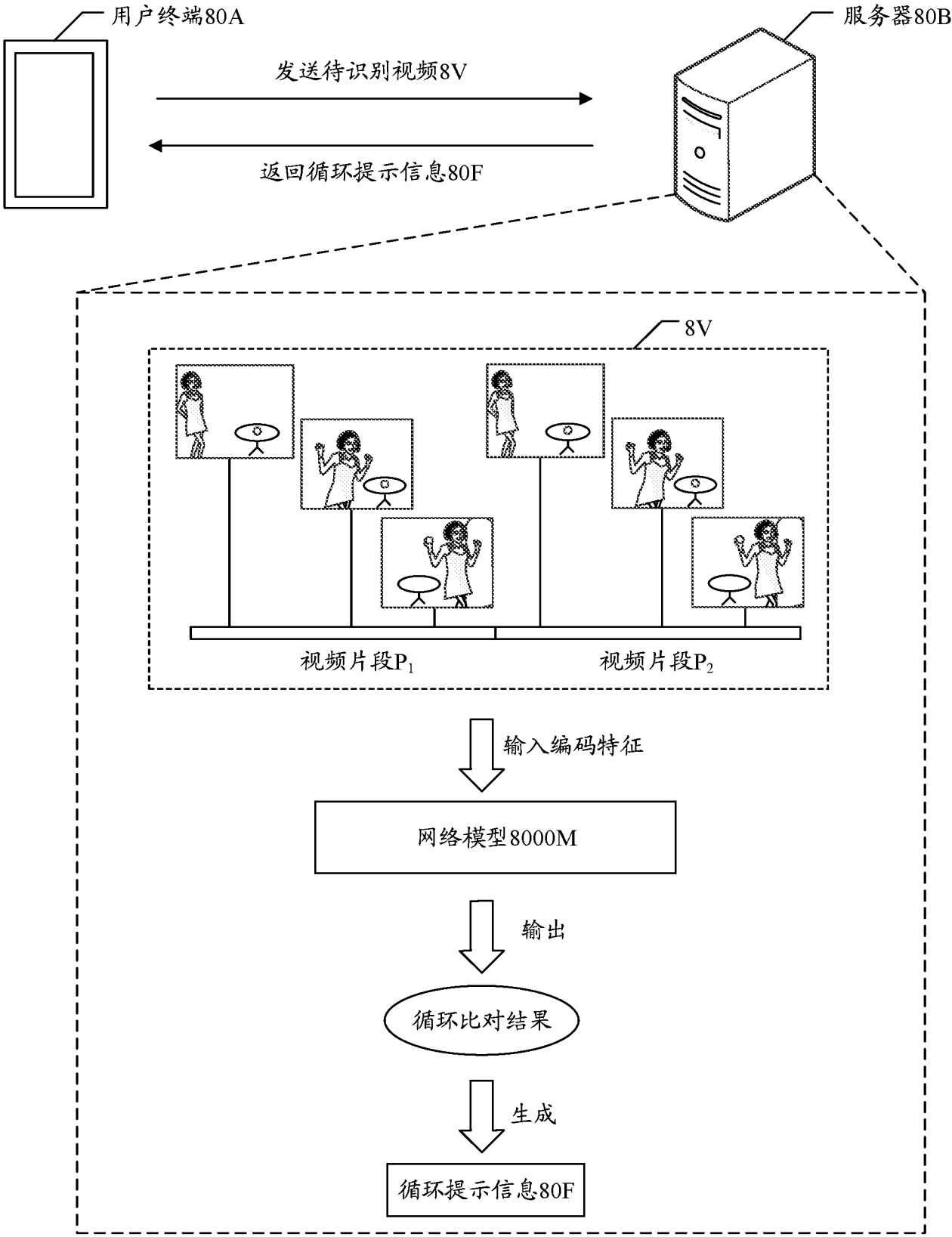


图 8

-9/10-

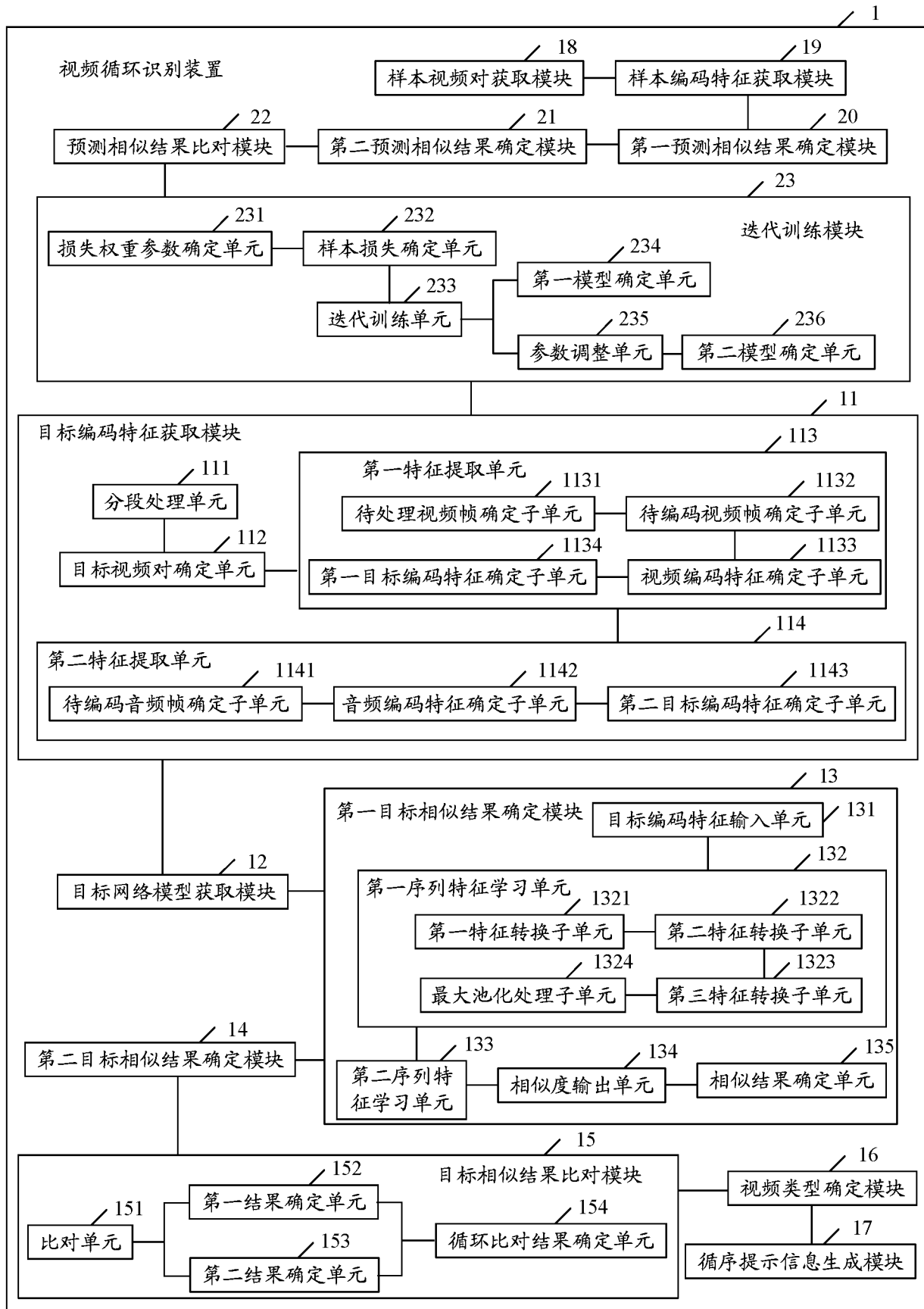


图9

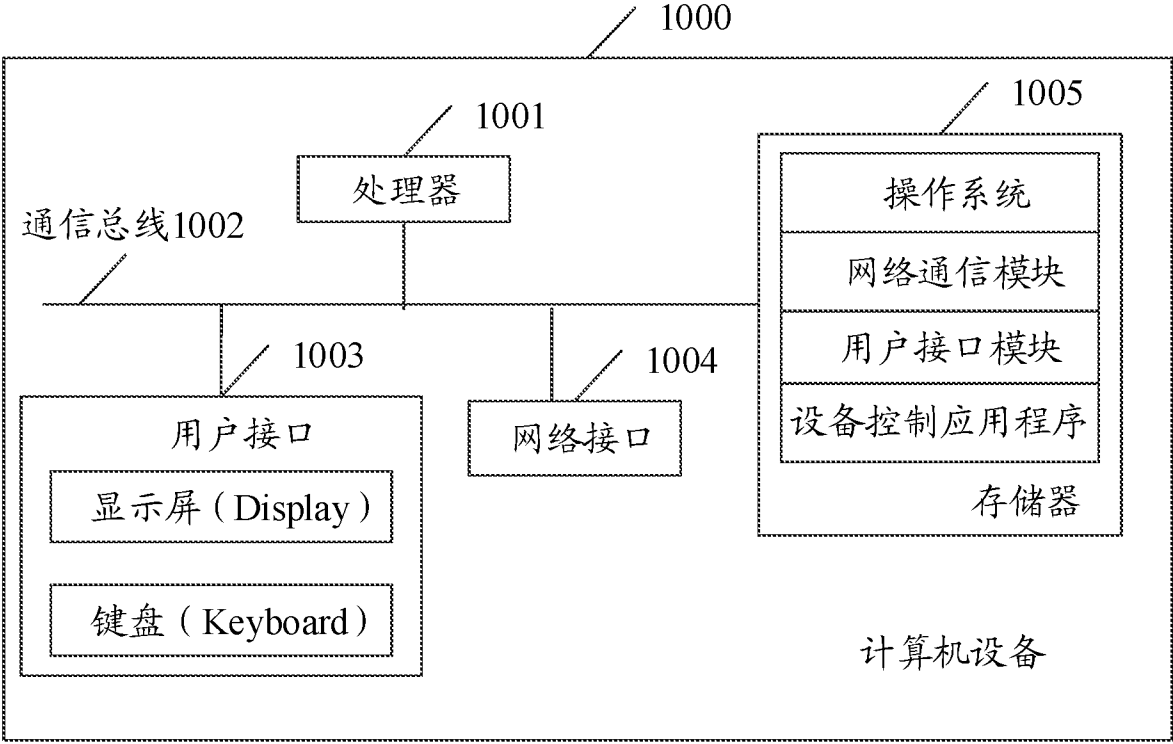


图 10

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/092377

**A. CLASSIFICATION OF SUBJECT MATTER**

G06K 9/00(2022.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI, CNABS, WPABS, CNTXT, ENTXT, ENTXTC, CJFD, CNKI: 视频, 循环, 识别, 片段, 模型, 编码, 相似度, 相似, 文本, 视觉, 图像, 音频, 模态, 比较, 匹配, 对比, 去重, 查重, video, loop, identification, identif+, segment, model, encoding, encode, similarity, similar, text, visual, image, audio, modality, comparison, match, contrast, deduplication, check

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                  | Relevant to claim No. |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| PX        | CN 113177538 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 27 July 2021 (2021-07-27)<br>claims 1-15, and description, paragraphs 24-168 | 1-17                  |
| A         | CN 109389096 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 26 February 2019 (2019-02-26)<br>entire document                    | 1-17                  |
| A         | CN 109391825 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 26 February 2019 (2019-02-26)<br>entire document                             | 1-17                  |
| A         | CN 101937506 A (FUDAN UNIVERSITY) 05 January 2011 (2011-01-05)<br>entire document                                                   | 1-17                  |
| A         | CN 109977262 A (MEGVII TECHNOLOGY LTD.) 05 July 2019 (2019-07-05)<br>entire document                                                | 1-17                  |
| A         | WO 2019241346 A1 (GOOGLE L.L.C.) 19 December 2019 (2019-12-19)<br>entire document                                                   | 1-17                  |



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance  
 “E” earlier application or patent but published on or after the international filing date  
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  
 “O” document referring to an oral disclosure, use, exhibition or other means  
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  
 “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone  
 “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art  
 “&” document member of the same patent family

Date of the actual completion of the international search

15 July 2022

Date of mailing of the international search report

27 July 2022

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/  
CN)  
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing  
100088, China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2022/092377**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) |            |    | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|------------|----|--------------------------------------|
| CN                                        | 113177538  | A  | 27 July 2021                         | CN                      | 113177538  | B  | 24 August 2021                       |
| CN                                        | 109389096  | A  | 26 February 2019                     | CN                      | 109389096  | B  | 02 October 2020                      |
| CN                                        | 109391825  | A  | 26 February 2019                     | US                      | 2019320002 | A1 | 17 October 2019                      |
|                                           |            |    |                                      | WO                      | 2019024919 | A1 | 07 February 2019                     |
|                                           |            |    |                                      | EP                      | 3664402    | A4 | 10 June 2020                         |
|                                           |            |    |                                      | EP                      | 3664402    | A1 | 10 June 2020                         |
| CN                                        | 101937506  | A  | 05 January 2011                      | CN                      | 101937506  | B  | 17 October 2012                      |
| CN                                        | 109977262  | A  | 05 July 2019                         | CN                      | 109977262  | B  | 16 November 2021                     |
| WO                                        | 2019241346 | A1 | 19 December 2019                     | EP                      | 3740935    | A1 | 25 November 2020                     |
|                                           |            |    |                                      | US                      | 2021089777 | A1 | 25 March 2021                        |
|                                           |            |    |                                      | CN                      | 111742345  | A  | 02 October 2020                      |
|                                           |            |    |                                      | EP                      | 3740935    | B1 | 22 December 2021                     |

|                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                  |                                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|-----------------------------------|
| <b>A. 主题的分类</b><br>G06K 9/00 (2022.01) i<br><br>按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类                                                                                                                                                                                                                                                                                                                                                 |                                                                                  |                                   |
| <b>B. 检索领域</b><br>检索的最低限度文献 (标明分类系统和分类号)<br>G06K<br><br>包含在检索领域中的除最低限度文献以外的检索文献<br><br>在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))<br>DWPI, CNABS, WPABS, CNTXT, ENTXT, ENTXTC, CJFD, CNKI: 视频、循环、识别、片段、模型、编码、相似度、相似、文本、视觉、图像、音频、模态、比较、匹配、对比、去重、查重、video, loop, identification, identif+, segment, model, encoding, encode, similarity, similar, text, visual, image, audio, modality, comparison, match, contrast, deduplication, check |                                                                                  |                                   |
| <b>C. 相关文件</b>                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                  |                                   |
| 类 型*                                                                                                                                                                                                                                                                                                                                                                                                                                | 引用文件, 必要时, 指明相关段落                                                                | 相关的权利要求                           |
| PX                                                                                                                                                                                                                                                                                                                                                                                                                                  | CN 113177538 A (腾讯科技深圳有限公司) 2021年7月27日 (2021 - 07 - 27)<br>权利要求1-15, 说明书第24-168段 | 1-17                              |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                   | CN 109389096 A (北京字节跳动网络技术有限公司) 2019年2月26日 (2019 - 02 - 26)<br>全文                | 1-17                              |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                   | CN 109391825 A (腾讯科技深圳有限公司) 2019年2月26日 (2019 - 02 - 26)<br>全文                    | 1-17                              |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                   | CN 101937506 A (复旦大学) 2011年1月5日 (2011 - 01 - 05)<br>全文                           | 1-17                              |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                   | CN 109977262 A (北京旷视科技有限公司) 2019年7月5日 (2019 - 07 - 05)<br>全文                     | 1-17                              |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                   | WO 2019241346 A1 (GOOGLE LLC) 2019年12月19日 (2019 - 12 - 19)<br>全文                 | 1-17                              |
| <input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。                                                                                                                                                                                                                                                                                                                                                |                                                                                  |                                   |
| * 引用文件的具体类型:<br>“A” 认为不特别相关的表示了现有技术一般状态的文件<br>“E” 在国际申请日的当天或之后公布的在先申请或专利<br>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)<br>“O” 涉及口头公开、使用、展览或其他方式公开的文件<br>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件<br>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件<br>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性<br>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性<br>“&” 同族专利的文件                                   |                                                                                  |                                   |
| 国际检索实际完成的日期<br>2022年7月15日                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                  | 国际检索报告邮寄日期<br>2022年7月27日          |
| ISA/CN的名称和邮寄地址<br>中国国家知识产权局 (ISA/CN)<br>中国北京市海淀区蓟门桥西土城路6号 100088<br>传真号 (86-10)62019451                                                                                                                                                                                                                                                                                                                                             |                                                                                  | 受权官员<br>田竞<br>电话号码 86-10-62411749 |

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2022/092377

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 |            |    | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|------------|----|----------------|
| CN          | 113177538  | A  | 2021年7月27日     | CN   | 113177538  | B  | 2021年8月24日     |
| CN          | 109389096  | A  | 2019年2月26日     | CN   | 109389096  | B  | 2020年10月2日     |
| CN          | 109391825  | A  | 2019年2月26日     | US   | 2019320002 | A1 | 2019年10月17日    |
|             |            |    |                | WO   | 2019024919 | A1 | 2019年2月7日      |
|             |            |    |                | EP   | 3664402    | A4 | 2020年6月10日     |
|             |            |    |                | EP   | 3664402    | A1 | 2020年6月10日     |
| CN          | 101937506  | A  | 2011年1月5日      | CN   | 101937506  | B  | 2012年10月17日    |
| CN          | 109977262  | A  | 2019年7月5日      | CN   | 109977262  | B  | 2021年11月16日    |
| WO          | 2019241346 | A1 | 2019年12月19日    | EP   | 3740935    | A1 | 2020年11月25日    |
|             |            |    |                | US   | 2021089777 | A1 | 2021年3月25日     |
|             |            |    |                | CN   | 111742345  | A  | 2020年10月2日     |
|             |            |    |                | EP   | 3740935    | B1 | 2021年12月22日    |