

视频生成方法、装置、电子设备及可读存储介质

相关申请的交叉引用

本申请要求于 2021 年 12 月 31 日递交、申请号为 202111672927.6、
名称为“视频生成方法、装置、电子设备及可读存储介质”的中国专利申
5 请的优先权，其全部内容通过引用并入本文。

技术领域

本公开涉及视频处理技术领域，尤其涉及一种视频生成方法、装置、
电子设备及可读存储介质。

背景技术

10 将用户在过去一定时间段内已发布的视频进行聚合生成一段视频，
用户观看生成视频的过程中可以唤醒一些深刻的记忆，因此，以回顾为
主题的视频深受用户的喜爱。

发明内容

本公开提供了一种视频生成方法、装置、电子设备及可读存储介质。
15 第一方面，本公开提供了一种视频生成方法，包括：

从原始视频素材集合中获取多个视频素材；所述原始视频素材集
合包括用户相关的视频；

获取用作背景音乐的目标音频素材；

20 针对各所述视频素材，对所述视频素材的各视频帧分别进行图像
特征提取，根据各所述视频素材的各所述视频帧分别对应的图像特征
信息进行分段处理，获得所述视频素材对应的目标视频片段；

将各所述视频素材分别对应的目标视频片段以及所述目标音频素
材进行聚合，生成目标视频；所述目标视频包括多个视频片段，所述多
个视频片段是分别基于多个目标视频片段得到的，所述目标视频中多
25 个所述视频片段按照发布时间顺序播放，且所述多个视频片段的时长

与所述目标音频素材中相应乐句的时长匹配。

作为一种可能的实施方式，所述获取用作背景音乐的目标音频素材，包括：

- 根据预设音频素材集合中各音频素材的音频特征、音频素材的节拍信息以及音频素材中特定音频片段的各乐句的时长，从所述预设音频素材集合中确定目标音频素材，其中，所述目标音频素材中的特定音频片段为所述目标视频的背景音乐。

- 作为一种可能的实施方式，所述根据预设音频素材集合中各音频素材的音频特征、音频素材的节拍信息以及音频素材中特定音频片段的各乐句的时长，从所述预设音频素材集合中确定目标音频素材，包括：

根据预设音乐特征集合对预设音乐素材集合包括的多个音乐素材进行排除，获得第一候选音频素材集合；

根据预设音频节拍，对所述第一候选音频素材集合包括的各音频素材进行排除，获得第二候选音频素材集合；

- 根据所述第二候选音频素材集合中，特定音频片段包括的各乐句时长满足预设时长条件的音频素材，确定所述目标音频素材。

作为一种可能的实施方式，若所述第二候选音频素材集合中不存在特定音频片段中乐句时长满足预设时长条件的音频素材，则所述方法还包括：

- 根据所述用户的偏好与预先指定的音频素材集合中各音频素材分别对应的音频特征进行匹配，若匹配成功，则根据匹配成功的音频素材确定所述目标音频素材。

作为一种可能的实施方式，所述方法还包括：

- 针对各所述视频素材，根据所述视频素材的各所述视频帧分别对应的图像特征信息进行加权计算，获得各所述视频帧分别对应的评价结果；

根据各所述视频帧分别对应的评价结果，从所述视频素材的各所述视频帧中抽取目标视频帧，以获得视频帧集合，所述视频帧集合包括

从所述多个视频素材中抽取的多个目标视频帧，所述视频帧集合用于生成所述目标视频的片头和/或片尾；

所述将各所述视频素材分别对应的目标视频片段以及所述目标音频素材进行聚合，生成目标视频，包括：

- 5 将各所述视频素材分别对应的目标视频片段、所述视频帧集合以及所述目标音频素材进行聚合，生成所述目标视频，所述目标视频包括根据所述视频帧集合生成的片头和/或片尾。

作为一种可能的实施方式，所述根据各所述视频素材的视频帧分别对应的图像特征信息进行分段处理，获得目标视频片段，包括：

- 10 根据所述视频素材的各视频帧分别对应的图像特征信息、所述目标音频片段中相应乐句的时长以及所述视频素材中原始音频的语句切分结果，对所述视频素材进行分段处理，获得所述目标视频片段。

作为一种可能的实施方式，所述目标音频片段包括相对应的原始音频中的一个或者多个完整语句。

- 15 第二方面，本公开提供了一种视频生成装置，包括：

视频处理模块，用于从原始视频素材集合中获取多个视频素材；所述原始视频素材集合包括用户相关的视频；

音频处理模块，用于获取用作背景音乐的目标音频素材；

- 20 所述视频处理模块，还用于针对各所述视频素材，对所述视频素材的各视频帧分别进行图像特征提取，根据各所述视频素材的各所述视频帧分别对应的图像特征信息进行分段处理，获得所述视频素材对应的目标视频片段；

- 25 视频聚合模块，用于将多个所述目标视频片段以及所述目标音频素材进行聚合，生成目标视频；所述目标视频包括多个视频片段，所述多个视频片段是分别基于多个目标视频片段得到的，所述目标视频中多个所述视频片段按照发布时间顺序播放，且所述多个视频片段的时长与所述目标音频素材中相应乐句的时长匹配。

第三方面，本公开提供一种电子设备，包括：存储器和处理器；

所述存储器被配置为存储计算机程序指令；

所述处理器被配置为执行所述计算机程序指令，使得所述电子设备实现第一方面任一项所述的视频生成方法。

第四方面，本公开提供一种可读存储介质，包括：计算机程序指令；
5 所述计算机程序指令被电子设备的至少一个处理器执行时，使得所述电子设备实现如第一方面任一项所述的视频生成方法。

第五方面，本公开提供一种计算机程序产品，当所述计算机程序产品被计算机执行时，使得所述计算机实现如第一方面任一项所述的视频生成方法。

10 第六方面，本公开提供一种计算机程序，当所述计算机程序被计算机执行时，使得所述计算机实现如第一方面任一项所述的视频生成方法。

附图说明

此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本公开的实施例，并与说明书一起用于解释本公开的原理。
15

为了更清楚地说明本公开实施例或现有技术中的技术方案，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地介绍，显而易见地，对于本领域普通技术人员而言，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

- 20 图 1 为本公开实施例提供的视频生成方法的应用场景图；
图 2 为本公开一实施例提供的视频生成方法的流程示意图；
图 3 为本公开另一实施例提供的视频生成方法的流程示意图；
图 4 为本公开另一实施例提供的视频生成方法的流程示意图；
图 5 为本公开另一实施例提供的视频生成方法的流程示意图；
25 图 6 为本公开一实施例提供的视频生成装置的结构示意图；
图 7 为本公开一实施例提供的电子设备的结构示意图。

具体实施方式

为了能够更清楚地理解本公开的上述目的、特征和优点，下面将对本公开的方案进行进一步描述。需要说明的是，在不冲突的情况下，本公开的实施例及实施例中的特征可以相互组合。

在下面的描述中阐述了很多具体细节以便于充分理解本公开，但本公开还可以采用其他不同于在此描述的方式来实施；显然，说明书中的实施例只是本公开的一部分实施例，而不是全部的实施例。

目前，通常会利用视频分析算法从用户在一定时间段内发布的视频中选择一部分视频片段，再将选择的这些视频片段进行聚合，从而生成回顾为主题的视频。然而，上述过程局限于如何选择视频片段，导致生成的视频整体不够生动，带给用户的感受无法达到预期，因此，如何生成更加生动精彩的回顾为主题的视频时亟待解决的问题。

本公开提供一种视频生成方法、装置、电子设备及可读存储介质，其中，该方法通过从用户在预设时间段内发布的视频素材中选择多个视频素材，并从选择的多个视频素材中提取一些目标视频片段，将抽取的目标视频片段与确定目标音频片段进行聚合，进而生成预设时间段对应的目标视频，使得用户能够通过目标视频回顾预设时间段内的深刻记忆；且目标视频中各视频片段的时长与所采用的目标音频片段中的乐句时长相匹配，使得目标视频的内容节奏与音频节奏相匹配，能够带给用户独特的体验感受。

图 1 为本公开提供的视频生成方法的场景示意图。图 1 所示的场景 100 包括：终端设备 101 和服务端设备 102，终端设备 101 上安装有客户端，客户端可以通过终端设备 101 与服务端设备 102 可以进行通信。

其中，客户端可通过终端设备 101 向用户展示获取目标视频的入口，基于用户针对该入口的触发操作生成操作指令，之后，客户端可以根据操作指令生成视频获取请求，客户端通过终端设备 101 向服务端设备发送视频获取请求，服务端设备 102 可以响应客户端通过终端设备 101 发送的视频获取请求，向客户端下发预设时间段对应的目标视

频。客户端接收到目标视频后，可加载目标视频，并通过终端设备 101 显示视频编辑页面，在视频编辑页面中播放展示目标视频，且用户可以通过视频编辑页面对视频进行一些调整，例如，更换目标视频的配乐等等。

- 5 服务端设备 102 可从用户相关的多个视频中选择多个视频素材，并从选择的多个视频素材中提取一些目标视频片段，且确定用作背景音乐的目标音频素材，再将抽取的目标视频片段与目标音频素材进行聚合，进而生成目标视频。由于服务端设备 102 可以预先离线生成目标视频，因此，服务端设备 102 接收到客户端发送的视频获取请求，即可快速响应视频获取请求，保证用户体验。

- 需要说明的是，除上述图 1 所示场景外，本公开提供的视频生成方法也可以终端设备本地执行，例如，客户端可以利用终端设备的资源预先从用户相关的视频中选择多个视频素材，并从选择的多个视频素材中提取一些目标视频片段，将抽取的目标视频片段与确定的目标音频片段均缓存在终端设备本地；当客户端接收到用户输入的视频获取请求时，将多个目标视频片段与目标音频素材进行聚合，生成目标视频。采用这样的方式，客户端可以利用终端设备在空闲时间段的资源进行视频素材以及音频素材的处理。其中，目标音频素材可以是终端设备预先从服务端设备获取的，也可以是在接收到用户输入的视频获取请求时，从服务端设备获取的。

当然，还可以采用其他方式实现本公开提供的视频生成方法，此处不一一赘述。

- 示例性地，本公开提供的视频生成方法可以由本公开提供的视频生成装置执行，该视频生成装置可以通过任意的软件和/或硬件的方式实现。示例性地，视频生成装置可以为手机、平板电脑、笔记本电脑、掌上电脑、车载终端、可穿戴设备、超级移动个人计算机（ultra-mobile personal computer, UMPC）、上网本或者个人数字助理（personal digital assistant, PDA）、个人计算机（personal computer, PC）、服务器、云

服务器、服务器集群等，本公开实施例对于视频生成装置的具体类型不作具体限定。

在下述实施例中，以电子设备执行视频生成方法为例进行详细介绍。

- 5 图 2 为本公开一实施例提供的视频生成方法的流程图。参照图 2 所示，本实施例的方法包括：

S201、从原始视频素材集合中获取多个视频素材。

- 其中，所述原始视频素材集合可以包括用户相关的视频，从原始视频素材集合中选择的多个视频素材用于生成用户对应的目标视频。其中，用户相关的视频可以包括但不限于：用户发布的视频、用户编辑但未发布的视频（如存储在草稿箱中的视频）、用户导出的视频等等；或者，用户相关的视频还可以与时间相关，即原始视频素材集合包括用户相关的且满足特定时间条件的视频，如，原始视频素材集合可以包括用户在过去一年中发布的视频、过去一年中拍摄的视频、过去一年中剪辑
- 10 的视频等等。
- 15

下述实施例中，以原始视频素材集合包括用于在预设时间段内发布的视频为例进行举例说明。

- 假设电子设备为服务端设备，服务端设备可以从用于存储用户发布的视频素材的数据库中获取用户在预设时间段内发布的视频素材；
- 20 假设电子设备为终端设备，通常客户端可以将用户已发布的视频素材、草稿箱中未发布的视频素材、用户下载的视频素材等缓存在终端设备本地的存储空间，因此，电子设备还可以从终端设备本地的存储空间获取用户在预设时间段内发布的视频素材。

- 本公开对于预设时间段的具体时长不做限定，例如，预设时间段可以
- 25 以为 1 年、半年、1 个季度等等。

其中，电子设备所确定的多个视频素材为用户在预设时长内发布的多个视频素材中的符合特定条件的视频素材，所选择的视频素材包括该用户在预设时间段内较具有意义的回忆的可能性较大。其中，用于

选择视频素材的特定条件可与视频素材在第一维度的信息相关，第一维度可以包括时长、播放、点赞、评论、收藏、转发、下载、分享等等。

作为一种可能的实施方式，电子设备可先分析用户在预设时间段内发布的所有视频素材的时长，将整个视频素材的时长大于第一预设时长的视频素材确定为候选视频素材，获得候选视频素材集合；再针对候选视频素材集合中每个候选视频素材分别在播放、点赞、评论、收藏、转发、下载、分享等一个或者多个第一维度的信息进行加权计算，得到每个候选视频素材分别对应的综合评分，再基于各候选视频素材对应的综合评分进行排序，确定综合评分满足预设条件的多个候选视频素材为用于生成预设时间段对应的目标视频的视频素材。

示例性地，针对预设时间段内的已发布的各视频素材，获取视频素材的点赞量 $X1$ 、评论量 $X2$ 、收藏量 $X3$ 以及转发量 $X4$ ，其中，点赞维度对应的权重为 $s1$ 、评论维度对应的权重为 $s2$ 、收藏维度对应的权重为 $s3$ 、转发维度对应的权重为 $s4$ ；基于每个维度对应的信息以及权重进行加权计算，获得视频素材对应的加权计算结果，上述加权计算通过公式可以表示为： $P=s1*X1+s2*X2+s3*X3+s4*X4$ ，其中， P 表示视频素材对应的加权计算结果。

其中，加权计算结果的数值越高，表示该已发布视频素材满足用于选择视频素材的特定条件的可能性越高；加权计算结果的数值越低，表示该已发布视频满足用于选择视频素材的特定条件的可能性越低。

需要说明的是，在进行加权计算之前，基于第一预设时长对用户预设时间段内已发布的视频素材进行筛选的目的是为了保证能够获得内容性较强的视频素材，视频素材时长太短，视频内容相对较少，可能无法分割出有效的用于生成目标视频的视频片段，也无法与配乐的乐句时长相匹配。

第一预设时长可以根据目标视频片段对应的最小时长确定。例如，目标视频片段最小时长为 $2s$ ，则第一预设时长可以大于 $2s$ ，例如，第一预设时长为 $3s$ 、 $4s$ 等等。

需要说明的是，本公开对于用于生成预设时间段对应的目标视频的视频素材的数量不做限定。例如，用于生成预设时间段对应的目标视频的视频素材的数量可以为一固定的数值，或者，用于生成预设时间段对应的目标视频的视频素材的数量可以预设数值范围内的任一数值，

5 可根据用户在预设时间段内发布的视频素材总量而定。示例性地，可以在电子设备设定用于生成预设时间段对应的目标视频的视频素材的数量为 5 个、10 个等等，或者，也可以设定用于生成预设时间段对应的目标视频的视频素材为[3, 10]中的任意数量。

此外，电子设备从原始视频素材集合中选择多个视频素材时，还可以先将用户转发的他人的视频、用户设置为私密状态的视频、用户发布之后删除的视频、以及广告推销等类型的视频进行排除，之后，再基于视频素材的时长以及视频素材在第一维度的信息进行筛选，以保护用户的隐私以及保证生成预设时间段对应的目标视频的内容是与用户在预设时间段内的回忆紧密相关的。

15 S202、获取用作背景音乐的目标音频素材。

在一些情况下，目标音频素材的时长、目标音频素材在时长以外的其他维度的音频特征等等满足本申请中作为目标视频的背景音乐的要求，则可以不需要对目标音频素材进行特殊的处理；在另一些情况下，目标音频素材的时长、目标音频素材中一些音频片段的音频特征等不

20 满足本申请中作为目标视频的背景音乐的要求，则可以从目标音频素材中提取满足条件的目标音频片段作为目标视频的背景音乐。

若以目标音频素材中的目标音频片段作为目标视频的背景音乐时，目标音频片段可以为电子设备通过综合分析预设音频素材集合中各音频素材在第二维度的音频特征，进而确定的目标音频素材中一特定音频片段。

25

预设音频素材集合可以包括用户收藏的音频素材、用户发布视频所使用的音频素材以及音频库中的音频素材等等。本公开对于预设音频素材集合所包括的音频素材的数量、音频素材的时长、音频素材的存

储格式等等参数均不作限定。

第二维度可以包括节奏、音频风格、音频情绪、音频场景、乐句的时长以及使用音频发布视频的次数等等一个或者多个维度。当然，第三维度还可以包括音频素材在其他维度的特征，本公开对于音频特征具体包括哪些维度的特征不做限定。

目标音频素材以及目标音频片段也可以是电子设备按照其他方式确定的音频。下文中，将在图 3 所示实施例详细介绍如何确定目标音频素材以及目标音频片段。

S203、针对各所述视频素材，对所述视频素材的各视频帧分别进行图像特征提取，根据各所述视频素材的各所述视频帧分别对应的图像特征信息进行分段处理，获得所述视频素材对应的目标视频片段。

本步骤中涉及的多个视频素材为步骤 S201 中从原始视频素材集合中获取的多个视频素材。

电子设备可采用预先训练好的视频处理模型，对选择的多个视频素材的每一个视频帧在第三维度进行特征提取，获得视频处理模型输出的针对视频素材中各视频帧的图像特征信息；基于视频素材中各视频帧对应的图像特征信息，获得视频帧对应评价结果；基于各视频帧分别对应的评价结果、视频素材的原始音频的语句切分结果以及目标音频片段中相应乐句的时长，从而确定所要切分的目标视频片段在视频素材中的位置，再进行分割，从而获得目标视频片段。

其中，上述第三维度可以包括图像风格、图像场景、图像主题、图像情绪、图像人物关系、图像显著性特征等等中的一个或者多个维度。此外，本公开对于视频处理模型的类型不做限定。例如，视频处理模型可以为神经网络模型、卷积模型等等。

其中，基于各视频帧分别对应的评价结果、视频素材的原始音频的语句切分结果以及目标音频片段中相应乐句的时长，获取目标视频片段，可以但不限于采用下述任一实施方式：

一种可能的实施方式，假设评价结果为数值时，电子设备可以通过

分析评价结果的数值较高的视频帧相对集中的视频帧范围，该视频帧范围可以设置等于或者超过预先设定的所生成的目标视频中每个视频片段的最大时长；之后，再基于针对视频素材的原始音频进行语句切分结果，在前述确定的视频帧范围内进行截取，获得目标视频片段，使得

5 目标视频片段对应的原始音频片段中包括的是一个或者多个完整语句。

另一种可能的实施方式，假设评价结果为数值时，电子设备还可以根据综合结果的数值最高的视频帧所处的位置，在评价结果的数值最高的视频帧所处位置前后进行搜索，获得目标视频片段，使得目标视频片段对应的原始音频片段中包括的是一个或者多个完整语句。

10 需要说明的是，采用上述任一方式确定的目标视频片段的时长可以小于目标音频片段中相对应的乐句的时长。示例性地，假设当前针对视频素材 2 执行上述第一种实施方式，若根据视频帧对应的评价结果确定相对评价结果的数值相对较高的视频帧范围为视频素材的第 15 帧至第 26 帧，以每秒 3 帧的速率播放，播放时长为 4s，视频素材 2 对应的

15 原始音频中，第 15 帧至第 26 帧所对应的原始音频片段包括 2 个完整语句，假设第 1 个语句和第 2 个语句时长均为 2s；而视频素材 2 对应的播放时长为目标音频片段中第 2 个乐句的时长为 3s，以每秒 3 帧的速率最多可以播放 9 帧，因此，需要从第 15 帧至第 26 帧中提取连续的 9 个及以下数量的视频帧，则可以根据两个语句分别对应的多个

20 视频帧的评价结果的数值高低的整体分布或者评价结果的数值最高的视频帧，确定其中一个语句对应的多个视频帧为需要提取的目标音频片段包括的视频帧。

示例性地，假设当前针对视频素材 3 执行上述第二种实施方式，若根据视频帧对应的评价结果确定相对评价结果的数值最高的视频帧为第 20 帧，视频素材 3 对应的原始音频中，第 20 帧所对应的语句时

25 长为 4s，前一个语句为 2s，后一个语句为 2s；而视频素材 3 对应的播放时长为目标音频片段中第 3 个乐句的时长为 8s，因此，可第 20 帧分别向前后搜索，确定第 20 帧对应的原始音频中的语句、前一个语句以

及后一个语句，确定其中一个语句对应的多个视频帧为需要提取的目标音频片段包括的视频帧。

电子设备通过对上述每个视频素材执行该过程，从而可以从选择的视频素材中提取相对应的目标视频片段。

- 5 需要说明的是，电子设备还可以通过其他方式从视频素材中提取目标视频片段。

S204、将多个目标视频片段以及目标音频素材进行聚合，生成目标视频；目标视频包括多个视频片段，所述多个视频片段是分别基于多个目标视频片段得到的，目标视频中多个所述视频片段按照发布时间顺
10 序播放，且所述多个视频片段的时长与目标音频素材中相应乐句的时长匹配。

多个目标视频片段即包括各视频素材对应的目标视频片段。

- 作为一种可能的实施方式，电子设备可以将多个目标视频片段按照发布时间顺序，填充至剪辑模板中，再与目标音频素材（或者目标音频素材的目标音频片段）进行聚合，从而生成目标视频。其中，剪辑模板用于指示一种或多种剪辑类型对应的剪辑操作方式，剪辑类型例如
15 包括转场、滤镜、特效等一种或多种。

- 由于在聚合的过程中，可能存在目标视频片段的时长大于或者小于相对应的乐句的时长，则可以将目标视频片段播放速度变快或者变慢，使得调整播放速度之后的目标视频片段与相对应的乐句时长匹配，
20 则目标视频包括的视频片段即上述调整了播放速度之后的目标视频片段。

- 此外，还可以在目标视频中，每个目标视频片段的每一视频帧中显示该目标视频片段的相关信息，例如，目标视频片段的发布时间、发布
25 地点、用户发布目标视频片段时的标题、是否为合拍类型视频等等信息。目标视频片段的相关信息可以显示在视频帧的任意位置，例如，左上角、右上角、左下角、右下角等位置，可以尽量不遮挡目标视频帧的画面，以保证用户的视觉体验。

本实施例提供的方法，通过从用户相关的视频中选择多个视频素材，并从选择的多个视频素材中提取一些目标视频片段，将抽取的目标视频片段与确定目标音频素材进行聚合，进而生成目标视频，用户能够通过目标视频回顾过去的深刻记忆；且目标视频中各视频片段的时长与所采用的目标音频素材中的乐句时长相匹配，使得目标视频的内容节奏与音频节奏相匹配，能够带给用户独特的体验感受。

图 3 为本公开另一实施例提供的视频生成方法的流程图。图 3 所示实施例主要用于介绍确定目标音频素材以及目标音频片的实施方式。参照图 3 所示，本实施例的方法包括：

10 S301、基于预设音频特征集合，对预设音频素材集合包括的多个音频素材进行排除，获得第一候选音频集合。

步骤 S301 相当于对音频在音频风格标签、音频情绪标签、音频语言标签、音频场景标签等维度进行筛选。

其中，预设音频素材集合可以包括音频库中的多个音频素材、用户收藏的音频素材、用户已发布视频所使用的音频素材等等。本公开对于确定预设音频素材集合的实现方式不做限定。电子设备可以获取预设音频素材集合中各音频素材的标识、音频素材的音频特征信息、音频素材的歌词信息等等，其中，音频素材的标识可以唯一标识音频素材，方便后续根据确定的目标音频素材的标识在预设音频素材集合中进行查找，音频素材的标识例如可以音频素材 ID、数字编号等等。

20 预设音频特征集合包括的音频特征为不推荐作为目标视频的配乐的音频素材在不同维度的音频特征。例如，预设音频特征集合可以包括音频风格特征、音频情绪特征、音频语言特征、音频场景特征等一个或者多个维度的音频特征。当然，预设音频特征集合还可以包括其他维度的音频特征，本公开对此不做限定。示例性地，预设音频特征集合包括：紧张、生气、疲惫等音频情绪特征。

25 电子设备根据预设音频特征集合对预设音频素材集合中的音频素材进行排除的过程中，可以将各音频素材对应的音乐特征与预设音频

特征集合中的各音频特征进行匹配, 若匹配成功, 则确定该音频素材为非候选音频; 若匹配不成功, 则确定该音频素材为候选音频素材, 从而获得第一候选音频素材集合。

第一候选音频素材集合可以包括多个候选音频素材。

- 5 S302、根据预设音频节拍, 对所述第一候选音频素材集合包括的各音频素材进行排除, 获得第二候选音频素材集合。

步骤 S302 相当于对音频素材在音频节奏维度进行筛选。

- 其中, 音频节拍可以通过每分钟的节拍数表示, 即 BPM (Beat Per Minute), BPM 可以用来衡量音频或者声音的节奏快慢, BPM 的值越高则表示节奏越快。
- 10

- 电子设备可以获取第一候选音频素材集合中各音频素材对应的 BPM, 再将音频素材对应的 BPM 与预设音频节拍进行对比, 满足预设音频节拍的音频素材为第二候选音频素材集合中的音频素材, 不满足预设音频节拍的音频素材为排除掉的音频素材, 排除的音频素材即无可能成为目标视频的配乐。
- 15

预设音频节拍可以为 BPM 范围, 例如, $BPM=[80, 130]$ 。

需要说明的是, 上述 S301 和 S302 的执行顺序不分先后, 也可以先执行 S302, 再执行 S301。

- 接下来, 电子设备可以根据第二候选音频素材集合中, 特定音频片段包括的各乐句时长满足预设时长条件的音频素材, 确定所述目标音频素材, 确定所述目标音频素材。示例性地, 可以包括步骤 S303 以至步骤 S305 实现。
- 20

S303、对第二候选音频素材集合中的音频素材分别进行音频识别, 获得第二候选音频素材集合中的音频素材分别对应的特定音频片段。

- 25 S304、针对第二候选音频集合中的各音频素材, 对所述音频素材的特定音频片段进行乐句切分, 获得所述特定音频片段包括的各乐句; 并根据所述目标音频片段包括的各乐句的时长是否满足预设时长条件, 确定所述音频素材是否为第三候选音频素材集合中的音频素材。

步骤 S303 以及 S304 相当于对音频素材在乐句的时长维度进行筛选，且是基于音频素材的特定音频片段所包括的乐句的时长进行筛选。

其中，特定音频片段可以理解为音频素材中的高潮音频片段。本公开对于从音频素材中获取相对应的特定音频片段的实现方式不做限定。

5 例如，特定音频片段可以通过音频处理模型对音频素材的特定属性进行分析，且可以结合用于生成预设时间段对应的目标视频的视频素材的数量，从目标音频中切分获得的音频片段。本公开对于音频处理模型的类型不做限定。例如，音频处理模型可以为神经网络模型、卷积模型等等。

10 针对第二候选音频素材集合中的各音频素材分别进行音频处理，获得第二候选音频素材集合中各音频素材分别对应的特定音频片段之后，可采用语音活动检测(Voice Activity Detection, VAD)技术对特定音频片段的歌词进行分析，确定特定音频片段中每一个乐句的歌词的起始位置和结束位置，即确定每一乐句对应的歌词的边界，将每一乐句的
15 歌词的起始位置和结束位置映射至特定音频片段中，从而获得每一乐句的时长。

若目标音频片段的各乐句的时长均满足第二预设条件（例如为预先设定的时长要求），则确定该特定音频片段所属的音频素材为第三候选音频素材集合中的音频素材；若特定音频片段的一个或多个乐句的
20 时长不满足第二预设条件，则确定该特定音频片段所属的音频素材为排除掉的音频素材。

需要说明的是，第二预设条件（例如为预先设定的时长要求）可根据目标视频中允许的单个视频片段的最小时长以及最大时长确定。例如，目标视频中，每个视频片段最短为 2s，最长为 5s，则第二预设条件
25 可以为大于或等于 3s，且小于或等于 6s，通过保留 1s 的时间余量，可用于转场，也可以避免后续在对视频片段进行聚合时，出现错误。

一种可能的情况，第二候选音频素材集合中一个或者多个音频素材符合条件，因此，第三候选音频素材集合包括一个或者多个音频素材，

则可以从第三候选音频素材集合包括的音频素材中确定目标音频素材，因此，该情况下，电子设备可以执行 S305。

另一种可能的情况，第二候选音频素材集合中的所有音频素材均不符合第二预设条件，因此，第三候选音频素材集合无音频素材，则无法从第三候选音频素材集合中确定目标音频素材，因此，该情况下，电子设备可以执行 S306。

S305、从第三候选音频素材集合包括的音频素材中，确定所述目标音频素材，目标音频片段为所述目标音频素材对应的特定音频片段。

一种可能的实施方式，电子设备可以对第三候选音频素材集合中的音频素材基于使用音频素材发布视频的总量、音频素材的收藏量、目标用户使用音频素材发布视频的数量等等信息进行加权计算，基于加权计算结果从高到低的顺序对第三候选音频素材集合中的各音频素材进行排序，基于排序结果，选择排在第一位的音频素材为目标音频素材。

另一种可能的实施方式，电子设备可以从第三候选音频素材集合中随机选择一个音频素材作为目标音频素材。

当然，电子设备还可以基于其他方式从第三候选音频素材集合中选择目标音频素材，本公开对于如何从第三候选音频素材集合中选择目标音频素材的实现方式不做限定。

需要说明的是，基于使用音频发布视频的总量、音频的收藏量、目标用户使用音频发布视频的数量等等信息对多个候选音频进行排序，也可以在步骤 S301 和 S302 之后、步骤 S303 之前执行，则执行 S303 至 S305 的过程中，可按照选择出的各候选音频的先后顺序进行。

S306、将用户的偏好与预先指定的音频素材集合中各音频素材对应的音频特征进行匹配，若匹配成功，则根据匹配成功的音频素材确定目标音频素材。

若用户的偏好与预先指定的音频素材集合中一个或多个音频素材的音频特征匹配成功，则可以从匹配成功的一个或多个音频素材中确

定一个作为目标音频素材。

例如, 可以按照基于使用音频素材发布视频的总量、音频素材的收藏量、目标用户使用音频素材发布视频的数量等信息进行加权计算, 基于加权计算结果从高到低的顺序对匹配成功的音频素材进行排序, 选择排在第一位的音频素材作为目标音频素材; 或者, 也可以随机从匹配成功的音频素材中选择一个作为目标音频素材。

若用户的偏好与在预预设音频的音频特征不匹配, 则可以从预设音频素材集合或者预先指定的音频素材集合中随机选择一个作为目标音频素材。

10 通过本实施例提供的方法选择目标音频素材, 使得选择的目标音频素材与用户在预设时间段内发布的多个视频素材中提取出的多个目标视频片段匹配, 通过选择的目标音频素材作为目标视频的背景音乐, 能够使得目标视频尽可能满足用户预期, 提升用户体验。

视频的片头和片尾是视频重要的部分, 能够提升视频的精彩程度, 因此, 通过图 4 所示实施例详细介绍如何生成目标视频的片头、片尾。

图 4 为本公开另一实施例提供的视频生成方法的流程图。参照图 4 所示, 本实施例提供的方法, 包括:

S401、从原始视频素材集合中获取多个视频素材。

其中, 所述原始视频素材集合包括用户在预设时间段内发布的视频素材, 从原始视频素材集合中选择的多个视频素材用于生成所述预设时间段对应的目标视频。

S402、获取用作背景音乐的目标音频素材。

本实施例中步骤 S401、S402 分别与图 2 所示实施例中步骤 S201、S202 类似, 可参照图 2 所示实施例的详细描述, 简明起见, 此处不再赘述。

需要说明的是, 由于目标视频包含片头和片尾, 因此, 在确定目标音频素材时, 不仅需要考虑到目标视频片段的时长, 还需要考虑到片头以及片尾需要分别对应一个乐句的时长, 以保证目标视频整体结构一致, 即

保证目标视频中每个视频片段与一完整乐句的时长相匹配。

S403、针对各视频素材,对视频素材的各视频帧分别进行图像特征提取,根据视频素材的各视频帧分别对应的图像特征信息,获得视频素材对应的目标视频片段以及目标视频帧,以获得多个目标视频片段和
5 视频帧集合。

视频帧集合包括从各所述视频素材中抽取的目标视频帧,所述视频帧集合包括的多个目标视频帧用于生成目标视频的片头和/或片尾。

其中,对多个视频素材分别进行视频帧图像特征提取,并根据各视频帧对应的图像特征信息进行分段处理,获取目标视频片段的实现方
10 式可参照图2所示实施例的详细描述,简明起见,此处不再赘述。

其中,视频帧集合包括的目标视频帧为用于生成目标视频的片头和/或片尾的图像素材。此处主要介绍如何获得视频帧集合包括的各目标视频帧。

作为一种可能的实施方式,目标视频帧可以是从小视频素材对应的
15 目标视频片段中抽取的一视频帧,且目标视频帧的图像特征信息符合条件。

示例性地,结合前文所述,在确定目标视频片段时,基于视频素材中各视频帧的图像特征信息,可以获得视频素材的各视频帧对应的评价结果,假设评价结果为数值时,则电子设备可以根据目标视频片段中
20 评价结果的数值最高的视频帧作为用于生成片头和/或片尾的图像素材。或者,电子设备也可以目标视频片段中评价结果的数值由高到低的顺序中,从靠前的多个视频帧中选择一个视频帧作为用于生成片头和/或片尾的图像素材。

通过针对每个视频素材执行上述过程,则可以从每个视频素材中
25 获得符合条件视频帧作为生成目标视频的片头和/或片尾的图像素材,即获得视频帧集合。

S404、根据视频帧集合生成片头和/或片尾,并将片头和/或片尾、以及多个目标视频片段与目标音频素材进行聚合,生成目标视频。

目标视频包括片头、分别基于多个目标视频片段得到的多个视频片段以及片尾，目标视频中多个所述视频片段按照发布时间顺序播放，且所述多个视频片段的时长与目标音频素材中相应乐句的时长匹配。

假设目标视频的背景音乐为目标音频素材中的目标音频片段，则

5 片头的时长与目标音频片段的第一个乐句的时长匹配，片尾的时长与目标音频片段的最后一个乐句的时长匹配，且各目标视频片段的时长与相应乐句的时长匹配，第 1 个目标视频片段的时长与目标音频片段中第 2 个乐句的时长匹配，第 2 个目标视频片段的时长与目标音频片段中第 3 个乐句的时长匹配，第 3 个目标音频片段的时长与目标音频

10 片段中第 4 个乐句的时长匹配，以此类推。

作为一种可能的实施方式，假设需要生成片头和片尾，电子设备可通过不同的剪辑模板分别对视频帧集合包括的多个目标视频帧进行剪辑，获得片头和片尾。剪辑模板用于指示将视频帧集合包括的目标视频帧剪辑成视频片段的剪辑方式，例如，剪辑方式包括但不限于转场方式、

15 特效、滤镜等等一种或多种。本公开对于剪辑模板不做限定。

由于片头的时长需要与目标音频片段的第一个乐句的时长匹配，片尾的时长需要与目标音频片段的最后一个乐句的时长匹配，因此，可根据目标音频片段的第一个乐句的时长选择用于剪辑片头的剪辑模板，根据目标音频片段的最后一个乐句的时长选择用于剪辑片尾的剪辑模

20 板。当然，在选择剪辑模板时，还可以考虑其他因素，例如，基于剪辑模板所生成的视频片段的风格、剪辑模板对于图像素材的要求等等。

需要说明的是，在聚合的过程中，可能存在目标视频片段的时长大于或者小于相对应的乐句的时长，则可以将目标视频片段播放速度变快或者变慢，使得调整播放速度之后的目标视频片段与相对应的乐句

25 时长匹配。

本实施例通过从多个目标视频片段中分别抽取视频帧生成目标视频的片头和片尾，使得目标视频更加精彩，使得目标视频能够使得目标视频尽可能满足用户预期，提升用户体验，且有利于提高用户发布该视

频的兴趣。

在一个具体的实施例中, 假设, 视频生成方法由服务端设备执行, 服务端设备具体包括: 业务服务端、音频算法服务端、图像算法服务端以及用于执行聚合过程的云服务。且服务端设备可以预先离线生成预设时间段 (如过去一年) 对应的目标视频, 当业务服务端接收到客户端发送的视频获取请求时, 向客户端发送预设时间段 (如过去一年) 对应的目标视频。

下面将通过图 5 所示实施例详细介绍, 服务端如何预先生成目标视频。图 5 所示实施例示出了上述场景中视频生成方法的具体流程。

10 参照图 5 所示, 本实施例包括:

S501、业务服务端向数据库发视频素材获取请求。

其中, 视频素材获取请求用于请求目标用户在预设时间段发布的视频。视频素材获取请求可以包括目标用户的标识以及用于指示请求获取目标用户在预设时间段发布的视频素材的指示信息。

15 S502、数据库向业务服务端发送目标用户对应的原始视频素材集合。

S503、业务服务端从目标用户对应的原始视频素材集合中, 确定多个视频素材。

20 S504、业务服务端基于预设音频特征集合以及预设音频节拍, 对预设音频素材集合进行排除, 获得第二候选音频素材集合。

S505、业务服务端将第二候选音频素材集合包括的各音频素材的标识发送给音频算法服务端。

25 S506、音频算法服务端根据接收到的音频素材的标识下载音频素材, 并对音频素材进行特定音频片段识别、乐句切分以及筛选, 确定目标音频素材。

S507、音频算法服务端向业务服务端发送目标音频素材的标识以及音频时间戳信息。

其中, 音频时间戳信息包括目标音频素材中的目标音频片段中各

乐句的时间戳信息，通过音频时间戳信息能够确定各目标音频片段中各乐句的起始时间、结束时间，进而可以获得各乐句的时长信息。

S508、业务服务端将多个视频素材的标识、目标音频素材的标识以及音频时间戳信息发送至图像算法服务端。

5 S509、图像算法服务端根据多个视频素材的标识、目标音频素材的标识，获取多个视频素材以及目标音频素材。

S510、图像算法服务端根据多个视频素材、目标音频素材以及音频时间戳信息，获得多个目标视频片段分别的时间戳信息以及多个目标视频片段与目标音频片段之间的视频聚合逻辑。

10 S511、图像算法服务端将多个目标视频片段分别的时间戳信息以及多个目标视频片段与目标音频片段之间的视频聚合逻辑发送至业务服务端。

S512、业务服务端将多个视频素材的标识、多个目标视频片段分别的时间戳信息、目标音频素材的标识、目标音频片段的时间戳信息以及
15 多个目标视频片段与目标音频片段之间的视频聚合逻辑发送给云服务。

S513、云服务根据多个视频素材的标识、多个目标视频片段分别的时间戳信息、目标音频素材的标识、目标音频片段的时间戳信息以及多个目标视频片段与目标音频片段之间的视频聚合逻辑进行视频聚合，生成目标视频。

20 生成的目标视频可以存储在云服务。

S514、客户端接收到用户发送的视频获取请求。

S515、客户端向业务服务端发送视频获取请求。

S516、业务服务端向云服务发送视频获取请求。

S517、云服务向业务服务端发送目标视频。

25 S518、业务服务端向客户端发送目标视频。

S519、客户端加载目标视频至视频编辑页面，进行播放。

之后，客户端可以基于用户的操作对目标视频进行剪辑并发布。

在图 5 所示实施例 5 中，音频算法服务端以及图像算法服务端的实

施方式可参照前述方法实施例的详细描述, 简明起见, 本实施例中不作赘述。

本实施例提供的方法, 通过从用户在预设时间段内发布的视频素材中选择多个视频素材, 并从选择的多个视频素材中提取一些目标视频片段, 将抽取的目标视频片段与确定目标音频片段进行聚合, 进而生成预设时间段对应的目标视频, 使得用户能够通过目标视频回顾预设时间段内的深刻记忆; 且目标视频中各视频片段的时长与所采用的目标音频片段中的乐句时长相匹配, 使得目标视频的内容节奏与音频节奏相匹配, 能够带给用户独特的体验感受。且该方案由客户端预先离线完成, 服务端设备接收到客户端发送的视频获取请求, 即可快速响应视频获取请求, 保证用户体验。

示例性地, 本公开还提供一种视频生成装置。

图6为本公开一实施例提供的视频生成装置的结构示意图。其中, 本实施例提供的视频生成装置可以为视频生成系统。参照图6所示, 视频生成装置600可以包括: 视频处理模块601、音频处理模块602以及视频聚合模块603。

其中, 视频处理模块601, 用于从原始视频素材集合中获取多个视频素材; 所述原始视频素材集合包括用户相关的视频。

音频处理模块602, 用于获取用作背景音乐的目标音频素材。

视频处理模块601, 还用于对所述视频素材的各视频帧分别进行图像特征提取, 根据各所述视频素材的各所述视频帧分别对应的图像特征信息进行分段处理, 获得所述视频素材对应的目标视频片段。

聚合模块603, 用于将各所述视频素材分别对应的目标视频片段以及所述目标音频素材进行聚合, 生成目标视频; 所述目标视频包括多个视频片段, 所述多个视频片段是分别基于多个目标视频片段得到的, 所述目标视频中多个所述视频片段按照发布时间顺序播放, 且所述多个视频片段的时长与所述目标音频素材中相应乐句的时长匹配。

本实施例提供的视频生成装置可以用于执行前述任一方法实施例

所示的技术方案，其实现原理以及技术效果类似，简明起见，可参照前述方法实施例的详细描述。

作为一种可能的实施方式，音频处理模块 602，具体用于根据预设音频素材集合中各音频素材的音频特征、音频素材的节拍信息以及音频素材中特定音频片段的各乐句的时长，从所述预设音频素材集合中确定目标音频素材，其中，所述目标音频素材中的特定音频片段为所述目标视频的背景音乐。

作为一种可能的实施方式，音频处理模块 602，具体用于根据预设音乐特征集合对预设音乐素材集合包括的多个音乐素材进行排除，获得第一候选音频素材集合；根据预设音频节拍，对所述第一候选音频素材集合包括的各音频素材进行排除，获得第二候选音频素材集合；根据所述第二候选音频素材集合中，特定音频片段包括的各乐句时长满足预设时长条件的音频素材，确定所述目标音频素材。

作为一种可能的实施方式，若所述第二候选音频素材集合中不存在特定音频片段中乐句时长满足预设时长条件的音频素材，则音频处理模块 602，还用于根据所述用户的偏好与预先指定的音频素材集合中各音频素材分别对应的音频特征进行匹配，若匹配成功，则根据匹配成功的音频素材确定目标音频素材。

作为一种可能的实施方式，视频处理模块 601，还用于针对各所述视频素材，根据所述视频素材的各所述视频帧分别对应的图像特征信息进行加权计算，获得各所述视频帧分别对应的评价结果；

根据各所述视频帧分别对应的评价结果，从所述视频素材的各所述视频帧中抽取目标视频帧，以获得视频帧集合，所述视频帧集合包括从所述多个视频素材中抽取的多个目标视频帧，所述视频帧集合用于生成所述目标视频的片头和/或片尾。

相应地，所述聚合模块 603，具体用于将各所述视频素材分别对应的目标视频片段、所述视频帧集合以及所述目标音频片段进行聚合，生成所述目标视频，所述目标视频包括根据所述视频帧集合生成的片头

和/或片尾。

作为一种可能的实施方式，视频处理模块 601，具体用于根据所述视频素材的各视频帧分别对应的图像特征信息、所述目标音频片段中相应乐句的时长以及所述视频素材中原始音频的语句切分结果，对所
5 述视频素材进行分段处理，获得所述目标视频片段。

作为一种可能的实施方式，所述目标音频片段包括相对应的原始音频中的一个或者多个完整语句。

作为一种可能的实施方式，视频生成装置 600 还可以包括：存储模块（图 6 中未示出），存储模块用于存储生成的目标视频。

10 作为一种可能的实施方式，若视频生成装置 600 为服务端设备时，视频生成装置还可以包括：通信模块（图 6 中未示出），通信模块用于接收客户端发送的视频获取请求，以及响应视频获取请求向客户端发送相应的预设时间段对应的目标视频。

需要说明的是，在图 6 所示装置实施例中未详细披露的细节可参
15 照前述方法实施例的描述，简明起见，装置实施例中未一一介绍。

示例性地，本公开还提供一种电子设备。

图 7 为本公开一实施例提供的电子设备的结构示意图。参照图 7 所示，本实施例提供的电子设备 700 包括：存储器 701 和处理器 702。

其中，存储器 701 可以是独立的物理单元，与处理器 702 可以通
20 过总线 703 连接。存储器 701、处理器 702 也可以集成在一起，通过硬件实现等。

存储器 701 用于存储程序指令，处理器 702 调用该程序指令，执行以上任一方法实施例的技术方案。

可选地，当上述实施例的方法中的部分或全部通过软件实现时，上
25 述电子设备 700 也可以只包括处理器 702。用于存储程序的存储器 701 位于电子设备 700 之外，处理器 702 通过电路/电线与存储器连接，用于读取并执行存储器中存储的程序。

处理器 702 可以是中央处理器（central processing unit，CPU），网

络处理器 (network processor, NP) 或者 CPU 和 NP 的组合。

处理器 702 还可以进一步包括硬件芯片。上述硬件芯片可以是专用集成电路 (application-specific integrated circuit, ASIC), 可编程逻辑器件 (programmable logic device, PLD) 或其组合。上述 PLD 可以是
5 复杂可编程逻辑器件 (complex programmable logic device, CPLD), 现场可编程逻辑门阵列 (field-programmable gate array, FPGA), 通用阵列逻辑 (generic array logic, GAL) 或其任意组合。

存储器 701 可以包括易失性存储器 (volatile memory), 例如随机存取存储器 (random-access memory, RAM); 存储器也可以包括非易
10 失性存储器 (non-volatile memory), 例如快闪存储器 (flash memory), 硬盘 (hard disk drive, HDD) 或固态硬盘 (solid-state drive, SSD); 存储器还可以包括上述种类的存储器的组合。

本公开还提供一种可读存储介质, 包括: 计算机程序指令; 计算机程序指令被电子设备的至少一个处理器执行时, 实现上述任一方法实
15 施例所示的视频生成方法。

本公开还提供一种计算机程序产品, 所述计算机程序产品被计算机执行时, 使得所述计算机实现上述任一方法实施例所示的视频生成方法。

需要说明的是, 在本文中, 诸如“第一”和“第二”等之类的关系术语
20 仅仅用来将一个实体或者操作与另一个实体或操作区分开来, 而并不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且, 术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含, 从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素, 而且还包括没有明确列出的其他要素, 或者是还包括为这
25 种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下, 由乐句“包括一个……”限定的要素, 并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

以上所述仅是本公开的具体实施方式, 使本领域技术人员能够理

解或实现本公开。对这些实施例的多种修改对本领域的技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本公开的精神或范围的情况下，在其它实施例中实现。因此，本公开将不会被限制于本文所述的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

5

权 利 要 求

1、一种视频生成方法，包括：

从原始视频素材集合中获取多个视频素材；所述原始视频素材集合包括用户相关的视频；

获取用作背景音乐的目标音频素材；

针对各所述视频素材，对所述视频素材的各视频帧分别进行图像特征提取，根据各所述视频素材的各所述视频帧分别对应的图像特征信息进行分段处理，获得所述视频素材对应的目标视频片段；

将各所述视频素材分别对应的目标视频片段以及所述目标音频素材进行聚合，生成目标视频；所述目标视频包括多个视频片段，所述多个视频片段是分别基于多个目标视频片段得到的，所述目标视频中多个所述视频片段按照发布时间顺序播放，且所述多个视频片段的时长与所述目标音频素材中相应乐句的时长匹配。

2、根据权利要求1所述的方法，其中，所述获取用作背景音乐的目标音频素材，包括：

根据预设音频素材集合中各音频素材的音频特征、音频素材的节拍信息以及音频素材中特定音频片段的各乐句的时长，从所述预设音频素材集合中确定目标音频素材，其中，所述目标音频素材中的特定音频片段为所述目标视频的背景音乐。

3、根据权利要求2所述的方法，其中，所述根据预设音频素材集合中各音频素材的音频特征、音频素材的节拍信息以及音频素材中特定音频片段的各乐句的时长，从所述预设音频素材集合中确定目标音频素材，包括：

根据预设音乐特征集合对预设音乐素材集合包括的多个音乐素材进行排除，获得第一候选音频素材集合；

根据预设音频节拍，对所述第一候选音频素材集合包括的各音频素材进行排除，获得第二候选音频素材集合；

根据所述第二候选音频素材集合中，特定音频片段包括的各乐句时长满足预设时长条件的音频素材，确定所述目标音频素材。

4、根据权利要求3所述的方法，其中，若所述第二候选音频素材集合中不存在特定音频片段中乐句时长满足预设时长条件的音频素材，则所述方法还包括：

根据所述用户的偏好与预先指定的音频素材集合中各音频素材分别对应的音频特征进行匹配，若匹配成功，则根据匹配成功的音频素材确定所述目标音频素材。

5、根据权利要求1所述的方法，还包括：

针对各所述视频素材，根据所述视频素材的各所述视频帧分别对应的图像特征信息进行加权计算，获得各所述视频帧分别对应的评价结果；

根据各所述视频帧分别对应的评价结果，从所述视频素材的各所述视频帧中抽取目标视频帧，以获得视频帧集合，所述视频帧集合包括从所述多个视频素材中抽取的多个目标视频帧，所述视频帧集合用于生成所述目标视频的片头和/或片尾；

所述将各所述视频素材分别对应的目标视频片段以及所述目标音频素材进行聚合，生成目标视频，包括：

将各所述视频素材分别对应的目标视频片段、所述视频帧集合以及所述目标音频素材进行聚合，生成所述目标视频，所述目标视频包括根据所述视频帧集合生成的片头和/或片尾。

6、根据权利要求1至5任一项所述的方法，其中，所述根据各所述视频素材的视频帧分别对应的图像特征信息进行分段处理，获得目标视频片段，包括：

根据所述视频素材的各视频帧分别对应的图像特征信息、所述目标音频片段中相应乐句的时长以及所述视频素材中原始音频的语句切分结果，对所述视频素材进行分段处理，获得所述目标视频片段。

7、根据权利要求6所述的方法，其中，所述目标音频片段包括相

对应的原始音频中的一个或者多个完整语句。

8、一种视频生成装置，包括：

视频处理模块，用于从原始视频素材集合中获取多个视频素材；所述原始视频素材集合包括用户相关的视频；

音频处理模块，用于获取用作背景音乐的目标音频素材；

所述视频处理模块，还用于针对各所述视频素材，对所述视频素材的各视频帧分别进行图像特征提取，根据各所述视频素材的各所述视频帧分别对应的图像特征信息进行分段处理，获得所述视频素材对应的目标视频片段；

视频聚合模块，用于将各所述视频素材分别对应的目标视频片段以及所述目标音频素材进行聚合，生成目标视频；所述目标视频包括多个视频片段，所述多个视频片段是分别基于多个目标视频片段得到的，所述目标视频中多个所述视频片段按照发布时间顺序播放，且所述多个视频片段的时长与所述目标音频素材中相应乐句的时长匹配。

9、一种电子设备，包括：存储器和处理器；

所述存储器被配置为存储计算机程序指令；

所述处理器被配置为执行所述计算机程序指令，使得所述电子设备实现如权利要求 1 至 7 任一项所述的视频生成方法。

10、一种可读存储介质，包括：计算机程序指令；

电子设备的至少一个处理器执行所述计算机程序指令，使得所述电子设备实现如权利要求 1 至 7 任一项所述的视频生成方法。

11、一种计算机程序产品，当所述计算机程序产品被计算机执行时，使得所述计算机实现如权利要求 1 至 7 任一项所述的视频生成方法。

12、一种计算机程序，当所述计算机程序被计算机执行时，使得所述计算机实现如权利要求 1 至 7 任一项所述的视频生成方法。