

(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
2024 年 12 月 19 日 (19.12.2024)



(10) 国际公布号
WO 2024/255652 A1

(51) 国际专利分类号:
H04N 21/85 (2011.01) **G06V 20/40** (2022.01)

(21) 国际申请号: PCT/CN2024/097380

(22) 国际申请日: 2024 年 6 月 4 日 (04.06.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202310701132.6 2023年6月13日 (13.06.2023) CN

(71) 申请人: 北京字跳网络技术有限公司
(**BEIJING ZITIAO NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市海淀区紫金数码园
4号楼2层0207, Beijing 100190 (CN)。

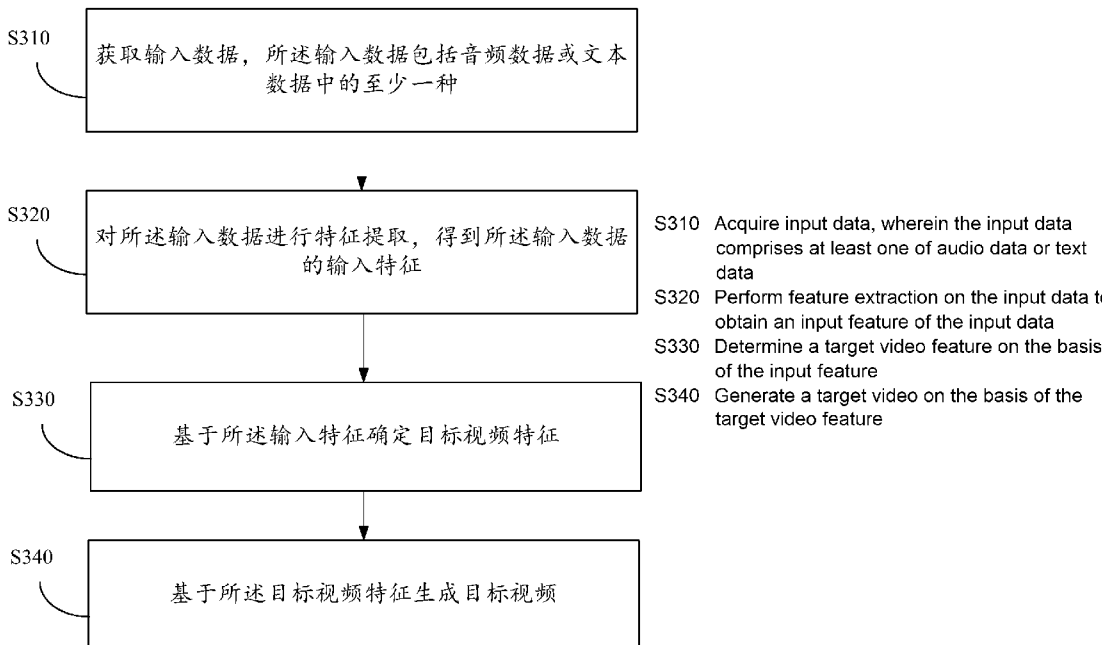
(72) 发明人: 张涛 (**ZHANG, Tao**); 中国北京市朝阳区七圣中街 12 号院融中心 B1 小邮局, Beijing 100028 (CN)。

(74) 代理人: 北京市金杜律师事务所 (**KING & WOOD MALLESONS**); 中国北京市朝阳区东三环中路 1 号环球金融中心办公楼东楼 20 层, Beijing 100020 (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, CV, 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, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) **Title:** VIDEO GENERATION METHOD AND APPARATUS, AND DEVICE, MEDIUM AND PROGRAM PRODUCT

(54) 发明名称: 视频生成方法、装置、设备、介质及程序产品



300

图 3

(57) **Abstract:** Provided in the present disclosure are a video generation method and apparatus, and a device, a storage medium and a program product. The method comprises: acquiring input data, wherein the input data comprises at least one of audio data or text data; performing feature extraction on the input data to obtain an input feature of the input data; determining a target video feature on the basis of the input feature; and generating a target video on the basis of the target video feature.

[见续页]

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, CV, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SC, 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, ME, 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))。

(57) 摘要: 本公开提供一种视频生成方法、装置、设备、存储介质及程序产品。该方法包括: 获取输入数据, 所述输入数据包括音频数据或文本数据中的至少一种; 对所述输入数据进行特征提取, 得到所述输入数据的输入特征; 基于所述输入特征确定目标视频特征; 基于所述目标视频特征生成目标视频。

视频生成方法、装置、设备、介质及程序产品

本申请要求于 2023 年 6 月 13 日提交中国国家知识产权局、申请号为 202310701132.6、发明名称为“视频生成方法、装置、设备、介质及程序产品”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。

技术领域

本公开涉及计算机技术领域，尤其涉及一种视频生成方法、装置、设备、介质及程序产品。

背景技术

基于多模态的视频生成技术（Multi-Modal-to-Video Generation）可以通过文字、语音等其他模态的数据来指导视频的生成。然而，现有的多模态视频生成技术中针对不同模态之间的匹配准确度不高，导致视频生成的准确性不高，降低了用户体验。

发明内容

本公开提出一种视频生成方法、装置、设备、存储介质及程序产品，以在一定程度上解决视频生成的准确性不高的技术问题。

本公开第一方面，提供了一种视频生成方法，包括：

获取输入数据，所述输入数据包括音频数据或文本数据中的至少一种；

对所述输入数据进行特征提取，得到所述输入数据的输入特征；

基于所述输入特征确定目标视频特征；

基于所述目标视频特征生成目标视频。

本公开第二方面，提供了一种视频生成装置，包括：

获取模块，用于获取输入数据，所述输入数据包括音频数据或文本数据；

提取模块，用于对所述输入数据进行特征提取，得到所述输入数据的输入特征；

匹配模块，用于基于所述输入特征确定目标视频特征；

生成模块，用于基于所述目标视频特征生成目标视频。

本公开第三方面，提供了一种电子设备，其特征在于，包括一个或者多个处理器、存储器；和一个或多个程序，其中所述一个或多个程序被存储在所述存储器中，并且被所述一个或多个处理器执行，所述程序包括用于执行根据第一方面或第二方面所述的方法的指令。

本公开第四方面，提供了一种包含计算机程序的非易失性计算机可读存储介质，当所述计算机程序被一个或多个处理器执行时，使得所述处理器执行第一方面或第二方面所述的方法。

本公开第五方面，提供了一种计算机程序产品，包括计算机程序指令，当所述计算机程序指令在计算机上运行时，使得计算机执行第一方面所述的方法。

从上面所述可以看出，本公开提供的一种视频生成方法、装置、设备、介质及程序产品，通过基于输入数据的输入特征确定对应的目标视频特征，从而生成目标视频。

附图说明

为了更清楚地说明本公开或相关技术中的技术方案，下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本公开的实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1为本公开实施例的视频生成架构的示意图。

图2为本公开实施例的示例性电子设备的硬件结构示意图。

图3为本公开实施例的视频生成方法的示意性流程图。

图4为本公开实施例的视频生成方法的示意性原理图。

图5为本公开实施例的视频生成装置的示意图。

具体实施方式

为使本公开的目的、技术方案和优点更加清楚明白，以下结合具体实施例，并参照附图，对本公开进一步详细说明。

需要说明的是，除非另外定义，本公开实施例使用的技术术语或者科学术语应当为本公开所属领域内具有一般技能的人士所理解的通常意义。本公开实施例中使用的“第一”、“第二”以及类似的词语并不表示任何顺序、数量或者重要性，而只是用来区分不同的组成部分。“包括”或者“包含”等类似的词语意指出现该词前面的元件或者物件涵盖出现在该词后面列举的元件或者物件及其等同，而不排除其他元件或者物件。“连接”或者“相连”等类似的词语并非限定于物理的或者机械的连接，而是可以包括电性的连接，不管是直接的还是间接的。“上”、“下”、“左”、“右”等仅用于表示相对位置关系，当被描述对象的绝对位置改变后，则该相对位置关系也可能相应地改变。

可以理解的是，在使用本公开各实施例公开的技术方案之前，均应当依据相关法律法规通过恰当的方式对本公开所涉及个人信息类型、使用范围、使用场景等告知用户并获得用户的授权。

例如，在响应于接收到用户的主动请求时，向用户发送提示信息，以明确地提示用户，其请求执行的操作将需要获取和使用到用户的个人信息。从而，使得用户可以根据提示信息来自主地选择是否向执行本公开技术方案的操作的电子设备、应用程序、服务器或存储介质等软件或硬件提供个人信息。

作为一种可选的但非限定性的实现方式，响应于接收到用户的主动请求，向用户发送提示信息的方式例如可以是弹窗的方式，弹窗中可以以文字的方式呈现提示信息。此外，弹窗中还可以承载供用户选择“同意”或者“不同意”向电子设备提供个人信息的选择控件。

可以理解的是，上述通知和获取用户授权过程仅是示意性的，不对本公开的实现方式构成限定，其它满足相关法律法规的方式也可应用于本公开的实现方式中。

图1示出了本公开实施例的视频生成架构的示意图。参考图1，该视频生成架构100可以包括服务器110、终端120以及提供通信链路的网络130。服务器110和终端120之间可通过有线或无线的网络130连接。其中，服务器110可以是独立的物理服务器，也可以是多个物理服务器构成的服务器集群或者分布式系统，还可以是提供云服务、云数据库、云计算、云函数、云存储、网络

服务、云通信、中间件服务、安全服务、CDN 等基础云计算服务的云服务器。

终端 120 可以是硬件或软件实现。例如，终端 120 为硬件实现时，可以是具有显示屏并且支持页面显示的各种电子设备，包括但不限于智能手机、平板电脑、电子书阅读器、膝上型便携计算机和台式计算机等等。终端 120 设备为软件实现时，可以安装在上述所列举的电子设备中；其可以实现成多个软件或软件模块(例如用来提供分布式服务的软件或软件模块)，也可以实现成单个软件或软件模块，在此不做具体限定。

需要说明的是，本申请实施例所提供的视频生成方法可以由终端 120 来执行，也可以由服务器 110 来执行。应了解，图 1 中的终端、网络和服务器的数目仅为示意，并不旨在对其进行限制。根据实现需要，可以具有任意数目的终端、网络和服务器的。

图 2 示出了本公开实施例所提供的示例性电子设备 200 的硬件结构示意图。如图 2 所示，电子设备 200 可以包括：处理器 202、存储器 204、网络模块 206、外围接口 208 和总线 210。其中，处理器 202、存储器 204、网络模块 206 和外围接口 208 通过总线 210 实现彼此之间在电子设备 200 的内部的通信连接。

处理器 202 可以是中央处理器 (Central Processing Unit, CPU)、视频生成器、神经网络处理器 (NPU)、微控制器 (MCU)、可编程逻辑器件、数字信号处理器 (DSP)、应用专用集成电路 (Application Specific Integrated Circuit, ASIC)、或者一个或多个集成电路。处理器 202 可以用于执行与本公开描述的技术相关的功能。在一些实施例中，处理器 202 还可以包括集成为单一逻辑组件的多个处理器。例如，如图 2 所示，处理器 202 可以包括多个处理器 202a、202b 和 202c。

存储器 204 可以配置为存储数据 (例如，指令、计算机代码等)。如图 2 所示，存储器 204 存储的数据可以包括程序指令 (例如，用于实现本公开实施例的视频生成方法的程序指令) 以及要处理的数据 (例如，存储器可以存储其他模块的配置文件等)。处理器 202 也可以访问存储器 204 存储的程序指令和数据，并且执行程序指令以对要处理的数据进行操作。存储器 204 可以包括易失性存储装置或非易失性存储装置。在一些实施例中，存储器 204 可以包括随机访问存储器 (RAM)、只读存储器 (ROM)、光盘、磁盘、硬盘、固态硬盘 (SSD)、闪存、存储棒等。

网络模块 206 可以配置为经由网络向电子设备 200 提供与其他外部设备的通信。该网络可以是能够传输和接收数据的任何有线或无线的网络。例如，该网络可以是有线网络、本地无线网络（例如，蓝牙、WiFi、近场通信（NFC）等）、蜂窝网络、因特网、或上述的组合。可以理解的是，网络的类型不限于上述具体示例。在一些实施例中，网络模块 306 可以包括任意数量的网络接口控制器（NIC）、射频模块、接收发器、调制解调器、路由器、网关、适配器、蜂窝网络芯片等的任意组合。

外围接口 208 可以配置为将电子设备 200 与一个或多个外围装置连接，以实现信息输入及输出。例如，外围装置可以包括键盘、鼠标、触摸板、触摸屏、麦克风、各类传感器等输入设备以及显示器、扬声器、振动器、指示灯等输出设备。

总线 210 可以被配置为在电子设备 200 的各个组件（例如处理器 202、存储器 204、网络模块 206 和外围接口 208）之间传输信息，诸如内部总线（例如，处理器-存储器总线）、外部总线（USB 端口、PCI-E 总线）等。

需要说明的是，尽管上述电子设备 200 的架构仅示出了处理器 202、存储器 204、网络模块 206、外围接口 208 和总线 210，但是在具体实施过程中，该电子设备 200 的架构还可以包括实现正常运行所必需的其他组件。此外，本领域的技术人员可以理解的是，上述电子设备 200 的架构中也可以仅包含实现本公开实施例方案所必需的组件，而不必包含图中所示的全部组件。

在现有的视频生成技术中，一般依赖跨模态匹配技术来解决不同模态数据之间的匹配问题，如视频、音频、文本、压缩编码等。然而，现有技术中针对不同模态之间的匹配准确度不高，导致基于用户的输入所生成的视频准确性不高，不能满足用户的需求。因此，如何提高视频生成的准确性成为了亟需解决的技术问题。

鉴于此，本公开实施例提供了一种视频生成方法、装置、设备、存储介质及程序产品。通过基于输入数据的输入特征确定对应的目标视频特征，从而生成目标视频。无论输入数据是何种模态均可以实现与视频特征的匹配，准确地生成目标视频，提高了视频生成的准确性。

具体地，用户可以输入一段文本数据，例如一段歌词，希望生成与该段歌词对应的视频，例如 MV。可以基于本公开实施例的视频生成方法，对文本数据进行特征提取得到文本特征。再基于该文本特征确定与该文本特征匹配的目

标视频特征。然后基于该目标视频特征生成对应的目标视频能够提高文本模态与视频模态的跨模态匹配的准确性，从而提高所生成的目标视频的准确度。

参见图 3，图 3 示出了根据本公开实施例的视频生成方法的示意性流程图。图 3 中，视频生成方法 300 可以进一步包括如下步骤。

在步骤 S310，获取输入数据，所述输入数据包括音频数据或文本数据中的至少一种。

其中，输入数据可以是用户基于交互界面通过用户操作直接输入或确定的数据。例如，用户可以直接输入文本数据，音频数据或视频数据，也可以选择文本数据，音频数据或视频数据，进而经由网络获取，如果是视频数据，获取视频数据中的音频数据或文本数据。在一些实施例中，输入数据可以来自于同一数据源，例如同一个音视频数据；也可以来自于不同的数据源。

在步骤 S320，对所述输入数据进行特征提取，得到所述输入数据的输入特征。

在一些实施例中，输入数据可以包括至少一个输入数据切片。其中，输入数据切片，即将用户输入的文本数据、音频数据或视频数据切分成片段。该片段可以包括具有单语义特性的单语义数据。一方面，输入数据切片可以将长输入数据切分为短输入数据，从而为并行化加速提供了基础。另一方面，输入数据切片将输入数据切分为具有单一语义的片段，避免进行特征匹配时同一段数据中包含多个语义的数据段所造成语义混淆。

在一些实施例中，对所述输入数据进行特征提取，得到所述输入数据的输入特征，包括：

对所述输入数据进行切分得到单语义数据；

基于所述单语义数据得到与输入数据对应的多个所述输入特征。

其中，输入数据通常包括多个语义的数据段，直接对多个语义的数据段进行特征提取可能导致语义混淆，降低最终生成视频的质量。例如，对于文本数据“text1, text2”，text1 和 text2 分别具有各自的语义，如果作为一个数据进行特征提取可能会造成语义的混乱。为了解决语义混乱的问题，可以将一个多语义数据拆分为多个单语义数据，例如将文本数据“text1, text2”拆分为两个单语义数据“text1”和“text2”分别进行特征提取得到 2 个输入特征 F_text1 和 F_text2。然后基于输入特征 F_text1 和 F_text2 来进行后续的特征匹配，则可以避免相互之间的语义交叉导致的语义混乱，能够有效的提升语义匹配的质量。

例如可以将“日照香炉生紫烟，遥看瀑布挂前川”拆分成“日照香炉生紫烟”和“遥看瀑布挂前川”。应了解，上述输入数据的切分仅为举例，并不旨在对输入特征的模态和切分得到的单语义数据的数量进行限制，输入数据可以包括文本数据、音频数据或视频数据中的至少一种，切分后可以包括更少或更多的单语义数据，在此不做限制。

在一些实施例中，基于所述单语义数据得到与输入数据对应的多个所述输入特征，进一步包括：

对所述单语义数据进行所述多模态特征提取处理，得到所述输入特征。

具体地，参见图 4，图 4 示出了根据本公开实施例的视频生成方法的示意性原理图。图 4 中，在对输入数据进行特征提取时，所使用的特征提取的处理方式与得到视频特征库时一致，例如可以使得对视频特征库中视频特征的特征提取网络对输入数据进行特征提取。特征提取将从单语义数据段中提取出一个特征向量，从而有效地降低数据的维度。

在步骤 S330，基于所述输入特征确定目标视频特征。

其中，输入特征可以是文本模态或音频模态的特征，可以基于该输入特征在视频特征库中进行特征匹配从中确定相匹配的目标视频特征，如图 4 所示。其中，视频特征库可以包括多个视频特征，视频特征可以具有对应的语义特性。可以将输入特征与视频特征的语义进行匹配，确定视频特征中的目标视频特征。进一步地，视频特征库还可以包括于视频特征对应的文本特征。

在一些实施例中，基于所述输入特征确定目标视频特征，包括：

计算所述输入特征与视频特征库中的视频特征之间的语义相关度；

将所述语义相关度最高的视频特征确定为所述目标视频特征。

具体地，可以基于输入特征与视频特征库中所有的视频特征计算余弦距离得到语义相关度，并选取最小余弦距离对应的视频特征作为匹配结果。

在一些实施例中，方法 300 还可以包括：

基于多个视频数据进行多模态特征提取处理，得到多个具有语义特性的所述视频特征。

其中，多个视频特征则形成了视频特征库。还可以基于主题、风格等属性对视频数据进行预分类，可以在后续的目标视频生成时为生成不同主题、风格的视频提供数据基础。针对每个视频数据均可以进行多模态特征提取，例如可以基于多模态特征提取网络对视频数据进行特征提取得到具有语义的视频特

征。因为后续需要与文本、音频等其他模态的数据进行检索匹配，所以不能像现有技术中所使用的单一模态的视频特征提取器得到单一模态的特征集合。多模态特征集合可以在后续的使用中不断更新，不需要在每次特征匹配时重新制作数据库。应了解，对输入数据进行特征提取的方式与得到多模态特征的特征提取方式一致。

在步骤 S340，基于所述目标视频特征生成目标视频。

其中，目标视频特征具有对应的视频片段，基于目标视频特征则可以生成对应的目标视频。

在一些实施例中，基于所述目标视频特征生成目标视频，包括：

基于所述目标视频特征确定对应的目标视频切片；

将所述目标视频切片的目标时间戳与所述输入数据的源时间戳对齐，得到对齐后的目标视频切片；

基于所述源时间戳将对齐后的所述目标视频切片进行拼接，得到所述目标视频。

其中，目标视频特征对应于每个输入特征，即对应于每个输入数据切片（即单语义数据）。在为每一个输入数据切片检索到对应的目标视频切片后，需要将这些目标视频切片组合成完整的视频。然而，输入数据与匹配到的目标视频切片的时间戳不总是对齐的，所以需要对齐二者的时间戳后进行目标视频切片的拼接，得到最终的目标视频。

在一些实施例中，当输入数据为文本数据时，可以基于预设播放速度设置所述文本数据的源时间戳。

其中，对于本身具有时间戳的输入数据，例如音频数据，则可以将其自身的时间戳作为源时间戳。对于本身不具有时间戳的输入数据，例如文本数据，则可以为其设置相应的时间戳。具体地，对于文本数据 text，可以基于预设播放速度 v 对文本数据进行播放，例如，文本数据 text 包括多个文本数据段 text1, text2, ……texti, ……，分别对应的数据长度为 $L1, L2, \dots, Li, \dots$ ，则文本数据段 text1 起始时间戳为 0，结束时间戳为 $L1/v$ ，文本数据段 text2 的起始时间戳为 $L1/v$ ，结束时间戳为 $(L1+L2)/v$ ，依此类推，可以对文本数据 text 设置时间戳。应了解，上述时间戳仅为示例，并不旨在对时间戳进行限制，设置的时间戳可以包括或不包括起始时间戳和/或结束时间戳，也可以包括其他位置的时间戳，在此不做限制。

在一些实施例中，将所述目标视频切片的目标时间戳与所述输入数据的源时间戳对齐，得到对齐后的目标视频切片，进一步包括：

基于所述目标时间戳得到所述目标视频切片的目标时长，以及基于所述源时间戳得到所述输入数据切片的源时长；

基于所述目标时长和所述源时长进行所述目标时间戳与所述源时间戳的对齐，得到对齐后的目标视频切片。

在一些实施例中，基于所述目标时长和所述源时长进行所述目标时间戳与所述源时间戳的对齐，得到对齐后的目标视频切片，包括：

将所述目标时长等于所述源时长的目标视频切片，直接作为对齐后的目标视频切片；

针对所述目标时长大于所述源时长的目标视频切片进行裁剪，得到对齐后的目标视频切片；

针对所述目标时长小于所述源时长的目标视频切片进行插帧延长，得到对齐后的目标视频切片。

具体地，可以逐个依次处理每个目标视频切片，来使目标视频切片与输入数据切片的时间戳对齐。以输入数据切片 S 的起始时间戳 s_1 、结束时间戳 s_2 和目标视频切片 T 的起始时间戳 t_1 、结束时间戳 t_2 为例：

若源时长 $s_2 - s_1 =$ 目标时长 $t_2 - t_1$ ，则将目标视频切片 T 作为匹配结果填入输入数据切片 S 对应的时间段 s_1s_2 。

若源时长 $s_2 - s_1 <$ 目标时长 $t_2 - t_1$ ，则对目标视频切片 T 进行裁剪，使得裁剪后的目标视频切片 T' 的时长 $t_2' - t_1' =$ 源时长 $s_2 - s_1$ ，并将裁剪后的目标视频切片 T' 作为匹配结果填入输入数据切片 S 对应的时间段 s_1s_2 。

若源时长 $s_2 - s_1 >$ 目标时长 $t_2 - t_1$ ，则对目标视频切片 T 进行插帧延长，使得插帧延长后的目标视频切片 T' 的时长 $t_2' - t_1' = s_2 - s_1$ ，并将插帧延长后的目标视频切片 T' 作为匹配结果填入输入数据切片 S 对应的时间段 s_1s_2 。

具体地，在完成所有目标视频切片的匹配和时间戳调整对齐后，就可以将所有目标视频切片按照输入数据的时间戳顺序排列好，并将这些目标视频切片渲染成最终的目标视频。

在一些实施例中，输入数据还可以包括用于指示目标视频的属性的指示信息。在一些实施例中，基于所述输入特征确定目标视频特征，包括：

基于所述输入特征与视频特征库中具有所述属性的视频特征进行匹配，确

定所述视频特征中的目标视频特征。

其中，属性可以包括风格、主题等，例如风格可以包括搞笑风格、古典风格等；主题可以包括自然主题、动物主题等。用户可以通过输入数据指示生成的目标视频的风格或主题等属性，例如用户输入一段风格为 F1 的视频数据 A，且指示信息包括目标视频的风格为 F2，则根据本公开实施例所生成的目标视频为将视频数据 A 的风格变为 F2。

需要说明的是，本公开实施例的方法可以由单个设备执行，例如一台计算机或服务器等。本实施例的方法也可以应用于分布式场景下，由多台设备相互配合来完成。在这种分布式场景的情况下，这多台设备中的一台设备可以只执行本公开实施例的方法中的某一个或多个步骤，这多台设备相互之间会进行交互以完成所述的方法。

需要说明的是，上述对本公开的一些实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于上述实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

基于同一技术构思，与上述任意实施例方法相对应的，本公开还提供了一种视频生成装置，参见图 5，所述视频生成装置包括：

获取模块，用于获取输入数据，所述输入数据包括音频数据或文本数据中的至少一种；

提取模块，用于对所述输入数据进行特征提取，得到所述输入数据的输入特征；

匹配模块，用于基于所述输入特征确定目标视频特征；

生成模块，用于基于所述目标视频特征生成目标视频。

为了描述的方便，描述以上装置时以功能分为各种模块分别描述。当然，在实施本公开时可以把各模块的功能在同一个或多个软件和/或硬件中实现。

上述实施例的装置用于实现前述任一实施例中相应的视频生成方法，并且具有相应的方法实施例的有益效果，在此不再赘述。

基于同一技术构思，与上述任意实施例方法相对应的，本公开还提供了一种非暂态计算机可读存储介质，所述非暂态计算机可读存储介质存储计算机指

令，所述计算机指令用于使所述计算机执行如上任一实施例所述的视频生成方法。

本实施例的计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。

上述实施例的存储介质存储的计算机指令用于使所述计算机执行如上任一实施例所述的视频生成方法，并且具有相应的方法实施例的有益效果，在此不再赘述。

所属领域的普通技术人员应当理解：以上任何实施例的讨论仅为示例性的，并非旨在暗示本公开的范围（包括权利要求）被限于这些例子；在本公开的思路下，以上实施例或者不同实施例中的技术特征之间也可以进行组合，步骤可以以任意顺序实现，并存在如上所述的本公开实施例的不同方面的许多其它变化，为了简明它们没有在细节中提供。

另外，为简化说明和讨论，并且为了不会使本公开实施例难以理解，在所提供的附图中可以示出或不示出与集成电路(IC)芯片和其它部件的公知的电源/接地连接。此外，可以以框图的形式示出装置，以便避免使本公开实施例难以理解，并且这也考虑了以下事实，即关于这些框图装置的实施方式的细节是高度取决于将要实施本公开实施例的平台的（即，这些细节应当完全处于本领域技术人员的理解范围内）。在阐述了具体细节（例如，电路）以描述本公开的示例性实施例的情况下，对本领域技术人员来说显而易见的是，可以在没有这些具体细节的情况下或者这些具体细节有变化的情况下实施本公开实施例。因此，这些描述应被认为是说明性的而不是限制性的。

尽管已经结合了本公开的具体实施例对本公开进行了描述，但是根据前面的描述，这些实施例的很多替换、修改和变型对本领域普通技术人员来说将是显而易见的。例如，其它存储器架构（例如，动态RAM(DRAM)）可以使用

所讨论的实施例。

本公开实施例旨在涵盖落入所附权利要求的宽泛范围之内的所有这样的替换、修改和变型。因此，凡在本公开实施例的精神和原则之内，所做的任何省略、修改、等同替换、改进等，均应包含在本公开的保护范围之内。

权 利 要 求 书

1、一种视频生成方法，包括：

获取输入数据，所述输入数据包括音频数据或文本数据中的至少一种；

对所述输入数据进行特征提取，得到所述输入数据的输入特征；

基于所述输入特征确定目标视频特征；

基于所述目标视频特征生成目标视频。

2. 根据权利要求1的方法，其中，对所述输入数据进行特征提取，得到所述输入数据的输入特征，包括：

对所述输入数据进行切分得到单语义数据；

基于所述单语义数据得到与输入数据对应的多个所述输入特征。

3. 根据权利要求1的方法，其中，所述基于所述输入特征确定目标视频特征，包括：

计算所述输入特征与视频特征库中的视频特征之间的语义相关度；

将所述语义相关度最高的视频特征确定为所述目标视频特征。

4、根据权利要求1的方法，其中，基于所述目标视频特征生成目标视频，包括：

基于所述目标视频特征确定对应的目标视频切片；

将所述目标视频切片的目标时间戳与所述输入数据的源时间戳对齐，得到对齐后的目标视频切片；

基于所述源时间戳将对齐后的所述目标视频切片进行拼接，得到所述目标视频。

5. 根据权利要求4的方法，其中，所述输入数据包括至少一个单语义数据；

将所述目标视频切片的目标时间戳与所述输入数据的源时间戳对齐，得到对齐后的目标视频切片，进一步包括：

基于所述目标时间戳得到所述目标视频切片的目标时长，以及基于所述源时间戳得到所述单语义数据的源时长；

基于所述目标时长和所述源时长进行所述目标时间戳与所述源时间戳的

对齐，得到对齐后的目标视频切片。

6. 根据权利要求 5 的方法，其中，基于所述目标时长和所述源时长进行所述目标时间戳与所述源时间戳的对齐，得到对齐后的目标视频切片，包括：

将所述目标时长等于所述源时长的目标视频切片，直接作为对齐后的目标视频切片；

针对所述目标时长大于所述源时长的目标视频切片进行裁剪，得到对齐后的目标视频切片；

针对所述目标时长小于所述源时长的目标视频切片进行插帧延长，得到对齐后的目标视频切片。

7. 根据权利要求 1 的方法，其中，所述输入数据还包括用于指示目标视频的属性的指示信息；

则基于所述输入特征确定目标视频特征，包括：

基于所述输入特征与视频特征库中具有所述属性的视频特征进行匹配，确定所述视频特征中的目标视频特征。

8. 根据权利要求 1 的方法，其中，所述输入数据为文本数据，所述方法还包括：基于预设播放速度设置所述文本数据的源时间戳。

9. 一种视频生成装置，包括：

获取模块，用于获取输入数据，所述输入数据包括音频数据或文本数据；

提取模块，用于对所述输入数据进行特征提取，得到所述输入数据的输入特征；

匹配模块，用于基于所述输入特征确定目标视频特征；

生成模块，用于基于所述目标视频特征生成目标视频。

10. 一种电子设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现如权利要求 1 至 8 任意一项所述的方法。

11. 一种非暂态计算机可读存储介质，所述非暂态计算机可读存储介质

存储计算机指令，所述计算机指令用于使计算机执行权利要求 1 至 8 任一所述方法。

12、一种计算机程序产品，包括计算机程序指令，当所述计算机程序指令在计算机上运行时，使得计算机执行权利要求 1 至 8 任一所述的方法。



100

图 1

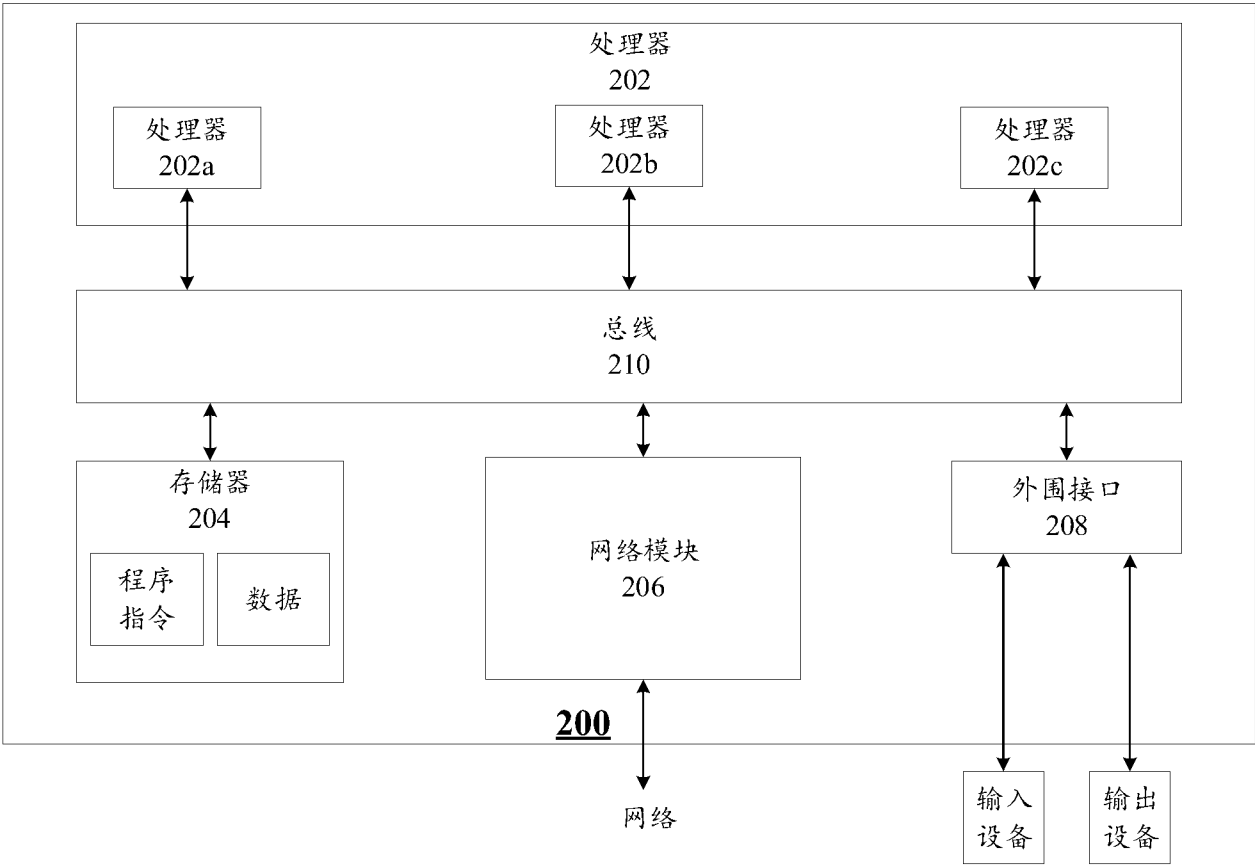


图 2

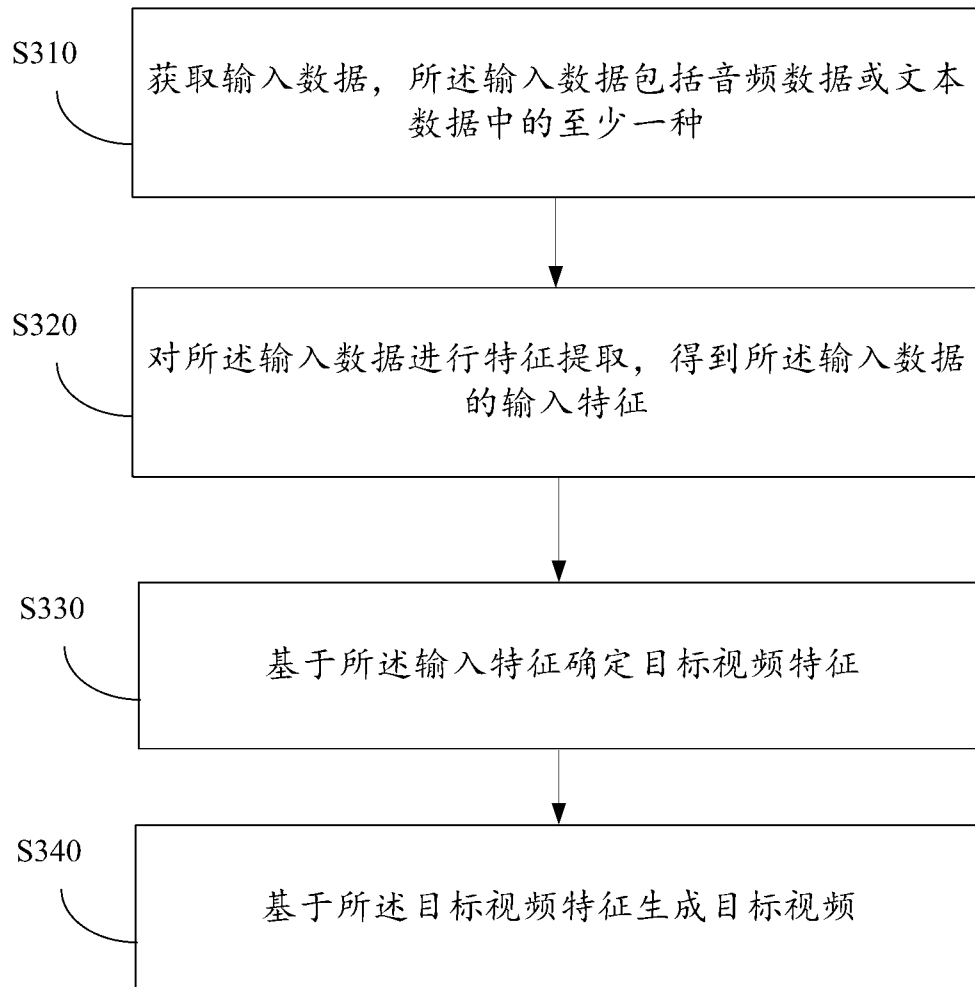
**300**

图 3

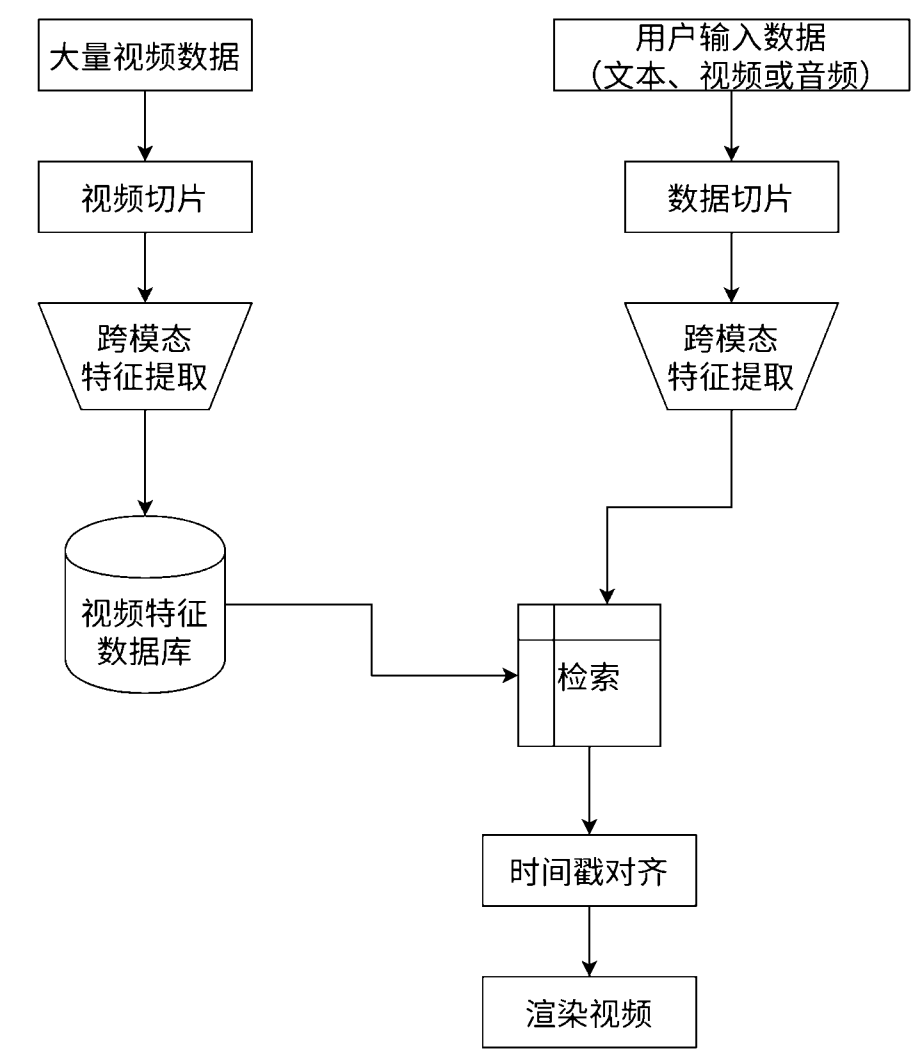


图 4

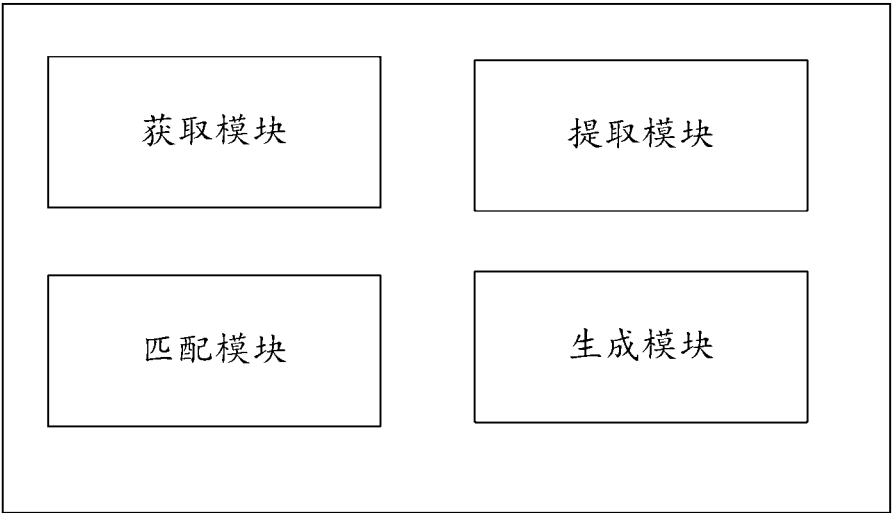


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/097380

A. CLASSIFICATION OF SUBJECT MATTER H04N 21/85(2011.01)i; G06V 20/40(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) IPC: H04N21/-, G06V20/- 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) CNABS, CNTXT, CNKI, VCN, CJFD: 视频, 帧, 片段, 生成, 产生, 构建, 合成, 制作, 生产, 创建, 特征, 特点, 关键字, 文本, 文字, 语音, 声音, 搜索, 检索, 查询, 查找, 搜寻, 寻找, 筛选, 检测, 定位, 匹配, 对比, 相似, 比对, 比较, 拼接, 合成, 合并, 连接, 融合, 拼合, 插接, 接合, 叠加, 整合, 拼缝, 时间戳, 时间, 时刻, 时钟, 对齐; EPTXT, USTXT, ENTXT, WOTXT, IEEE: video+, frame?, segment+, generate, generating, produce, producing, build+, synthesise, make, making, create, creating, feature?, keyword+, text?, voice?, sound?, search+, retrieve, find+, detect+, locate, match+, contrast+, resemble, compare, stitch+, synthesise, merge, join+, plug, overlay, integrate, timestamp, time?, moment?, clock?, align+;		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116634246 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 22 August 2023 (2023-08-22) description, paragraphs 0043-0098	1-12
X	CN 115967833 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 14 April 2023 (2023-04-14) description, paragraphs 0033-0123	1-12
X	CN 114242070 A (ALIBABA (CHINA) CO., LTD.) 25 March 2022 (2022-03-25) description, paragraphs 0026-0144	1-12
X	CN 115952317 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 11 April 2023 (2023-04-11) description, paragraphs 0036-0202	1-12
☒ 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 “D” document cited by the applicant in the international application “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 20 June 2024		Date of mailing of the international search report 21 June 2024
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 115237248 A (BEIJING YOUZHUJU NETWORK TECHNOLOGY CO., LTD.) 25 October 2022 (2022-10-25) entire document	1-12
A	CN 116012753 A (PING AN BANK CO., LTD.) 25 April 2023 (2023-04-25) entire document	1-12
A	US 2020201904 A1 (ADLAUNCH INTERNATIONAL INC.) 25 June 2020 (2020-06-25) entire document	1-12

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2024/097380

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116634246	A	22 August 2023	None			
CN	115967833	A	14 April 2023	None			
CN	114242070	A	25 March 2022	CN	114242070	B	24 March 2023
CN	115952317	A	11 April 2023	None			
CN	115237248	A	25 October 2022	None			
CN	116012753	A	25 April 2023	None			
US	2020201904	A1	25 June 2020	EP	3671487	A2	24 June 2020
				EP	3671487	A3	02 September 2020

A. 主题的分类

H04N 21/85(2011.01)i; G06V 20/40(2022.01)i

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. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: H04N21/-, G06V20/-

包含在检索领域中的除最低限度文献以外的检索文献

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))

CNABS, CNTXT, CNKI, VCN, CJFD: 视频, 帧, 片段, 生成, 产生, 构建, 合成, 制作, 生产, 创建, 特征, 特点, 关键字, 文本, 文字, 语音, 声音, 搜索, 检索, 查询, 查找, 搜寻, 寻找, 筛选, 检测, 定位, 匹配, 对比, 相似, 比对, 比较, 拼接, 合成, 合并, 连接, 融合, 拼合, 插接, 接合, 叠加, 整合, 拼缝, 时间戳, 时间, 时刻, 时钟, 对齐; EPTXT, USTXT, ENTXT, WOTXT, IEEE:video+, frame?, segment+, generate, generating, produce, producing, build+, synthesize, make, making, create, creating, feature?, keyword+, text?, voice?, sound?, search+, retrieve, find+, detect+, locate, match+, contrast+, resemble, compare, stitch+, synthesize, merge, join+, plug, overlay, integrate, timestamp, time?, moment?, clock?, align+;

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 116634246 A (北京字跳网络技术有限公司) 2023年8月22日 (2023 - 08 - 22) 说明书第0043-0098段	1-12
X	CN 115967833 A (北京字节跳动网络技术有限公司) 2023年4月14日 (2023 - 04 - 14) 说明书第0033-0123段	1-12
X	CN 114242070 A (阿里巴巴(中国)有限公司) 2022年3月25日 (2022 - 03 - 25) 说明书第0026-0144段	1-12
X	CN 115952317 A (北京字跳网络技术有限公司) 2023年4月11日 (2023 - 04 - 11) 说明书第0036-0202段	1-12
A	CN 115237248 A (北京有竹居网络技术有限公司) 2022年10月25日 (2022 - 10 - 25) 全文	1-12

☒ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

★ 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“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)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“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

“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

“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

“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

“&” 同族专利的文件

国际检索实际完成的日期

2024年6月20日

国际检索报告邮寄日期

2024年6月21日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

黄文波

电话号码 (+86) 028-62969345

PCT/ISA/210 表(第2页) (2022年7月)

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 116012753 A (平安银行股份有限公司) 2023年4月25日 (2023 - 04 - 25) 全文	1-12
A	US 2020201904 A1 (ADLAUNCH INT INC.) 2020年6月25日 (2020 - 06 - 25) 全文	1-12

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/097380

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	116634246	A	2023年8月22日	无			
CN	115967833	A	2023年4月14日	无			
CN	114242070	A	2022年3月25日	CN	114242070	B	2023年3月24日
CN	115952317	A	2023年4月11日	无			
CN	115237248	A	2022年10月25日	无			
CN	116012753	A	2023年4月25日	无			
US	2020201904	A1	2020年6月25日	EP	3671487	A2	2020年6月24日
				EP	3671487	A3	2020年9月2日