

舞蹈动作生成方法、计算机设备及存储介质

本申请要求于 2022 年 11 月 17 日提交中国专利局、申请号为 202211441749.0、申请名称为“舞蹈动作生成方法、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。

技术领域

本申请涉及人工智能领域，尤其涉及一种舞蹈动作生成方法、计算机设备及存储介质。

背景技术

舞蹈是由音频和舞蹈动作组成的一种可以传递情感的高级艺术，其中，音频和舞蹈动作如何匹配成为舞蹈编排的重点和难点，对于专业的舞蹈演员可以根据自己对音频情感的理解编排舞蹈动作，然而，这种人工编排舞蹈的方式依赖于专业的舞蹈演员，对于普通舞者或者用户无法自己完成舞蹈编排。因此，如何有效地实现对舞蹈的自动化编排非常重要。

发明内容

本申请实施例提供了一种舞蹈动作生成方法、计算机设备及存储介质，可以自动化地生成舞蹈动作，并提高了舞蹈动作的中提取多个音频片段；将所述多个音频片段输入预训练的编码质量。

第一方面，本申请实施例提供了一种舞蹈动作生成方法，包括：

获取待编舞音频，并从所述待编舞音频模型，得到所述多个音频片段中每个音频片段的第一动作特征，其中，所述编码模型是由样本音频和所述样本音频对应的样本舞蹈动作训练得到的；

根据所述每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与所述每个音频片段的第一动作特征相似的第二动作特征；

将所述每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征，并根据所述第三动作特征确定所述待编舞音频的舞蹈动作，其中，所述解码模型是由所述样本音频对应的样本动作特征以及所述样本音频对应的所述样本舞蹈动作训练得到的，所述样本音频对应的样本动作特征是利用所述预训练的编码模型对所述样本音频进行编码得到的，所述第三动作特征配置为指示所有音频片段的动作特征。

第二方面，本申请实施例提供了一种舞蹈动作生成装置，包括：

获取单元，配置为获取待编舞音频，并从所述待编舞音频中提取多个音频片段；

编码单元，配置为将所述多个音频片段输入预训练的编码模型，得到所述多个音频片段中每个音频片段的第一动作特征，其中，所述编码模型是由样本音频和所述样本音频对应的样本舞蹈动作训练得到的；

确定单元，配置为根据所述每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与所述每个音频片段的第一动作特征相似的第二动作特征；

解码单元，配置为将所述每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征，并根据所述第三动作特征确定所述待编舞音频的舞蹈动作，其中，所述解码模型是由所述样本音频对应的样本动作特征以及所述样本音频对应的所述样本舞蹈动作训练得到的，所述样本音频对应的样本动作特征是利用所述预训练的编码模型对所述样本音频

进行编码得到的，所述第三动作特征配置为指示所有音频片段的动作特征。

第三方面，本申请实施例提供了一种计算机设备，所述计算机设备包括：处理器和存储器，所述处理器配置为执行：

获取待编舞音频，并从所述待编舞音频中提取多个音频片段；

将所述多个音频片段输入预训练的编码模型，得到所述多个音频片段中每个音频片段的第一动作特征，其中，所述编码模型是由样本音频和所述样本音频对应的样本舞蹈动作训练得到的；

根据所述每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与所述每个音频片段的第一动作特征相似的第二动作特征；

将所述每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征，并根据所述第三动作特征确定所述待编舞音频的舞蹈动作，其中，所述解码模型是由所述样本音频对应的样本动作特征以及所述样本音频对应的所述样本舞蹈动作训练得到的，所述样本音频对应的样本动作特征是利用所述预训练的编码模型对所述样本音频进行编码得到的，所述第三动作特征配置为指示所有音频片段的动作特征。

第四方面，本申请实施例还提供了一种计算机可读存储介质，该计算机可读存储介质中存储有程序指令，该程序指令被执行时，配置为实现上述第一方面所述的方法。

本申请实施例可以获取待编舞音频，并从待编舞音频中提取多个音频片段；将多个音频片段输入预训练的编码模型，得到多个音频片段中每个音频片段的第一动作特征，其中，编码模型是由样本音频和样本音频对应的样本舞蹈动作训练得到的；根据每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与每个音频片段的第一动作特征相似的第二动作特征；将每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征，并根据第三动作特征确定待编舞音频的舞蹈动作，其中，解码模型是由样本音频对应的样本动作特征以及样本音频对应的样本舞蹈动作训练得到的，样本音频对应的样本动作特征是利用预训练的编码模型对样本音频进行编码得到的，第三动作特征配置为指示所有音频片段的动作特征。通过这种方式可以自动化地生成舞蹈动作，满足了用户对生成舞蹈动作的自动化、智能化需求，并提高了舞蹈动作的质量。

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，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，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1是本申请实施例提供的一种舞蹈动作生成方法的流程示意图；

图2是一种多帧舞蹈动作的示意图；

图3是一种人体关键点的示意图；

图4是本申请实施例提供的另一种舞蹈动作生成方法的流程示意图；

图5是本申请实施例提供的又一种舞蹈动作生成方法的流程示意图；

图6是本申请实施例提供的又一种舞蹈动作生成方法的流程示意图；

图7是本申请实施例提供的一种舞蹈动作生成装置的结构示意图；

图8是本申请实施例提供的一种计算机设备的结构示意图。

具体实施方式

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。

人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。

人工智能技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大图像处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

语音技术(Speech Technology)的关键技术有自动语音识别技术(ASR)和语音合成技术(TTS)以及声纹识别技术。让计算机能听、能看、能说、能感觉,是未来人机交互的发展方向,其中语音成为未来最被看好的人机交互方式之一。

自然语言处理(Nature Language processing, NLP)是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。因此,这一领域的研究将涉及自然语言,即人们日常使用的语言,所以它与语言学的研究有着密切的联系。自然语言处理技术通常包括文本处理、语义理解、机器翻译、机器人问答、知识图谱等技术。

机器学习(Machine Learning, ML)是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习/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。

基于上述人工智能技术中所提及的机器学习等技术,本申请提出了一种舞蹈动作生成方案,通过预训练的编码模型对待编舞音频进行编码得到待编舞音频的第一动作特征,并从通过模型学习得到的动作库中确定出与第一动作特征相似的第二个动作特征,进一步通过解码模型对第二个动作特征进行解码,生成该待编舞音频的舞蹈动作。通过这种方式,可以实现自动生成舞蹈动作,并提高了生成的舞蹈动作的质量。

本申请实施例提供的舞蹈动作生成方法可以应配置为一种舞蹈动作生成装置,该舞蹈动作生成装置可设置于计算机设备中,在某些实施例中,该计算机设备可以包括但不限于智能手机、平板电脑、笔记本电脑、台式电脑、车载智能终端、智能手表等智能终端设备。

在某些实施例中,本申请实施例提供的舞蹈动作生成方法可以应配置为舞蹈编排的场景:例如根据待编舞音频生成与该待编舞音频匹配的舞蹈动作等。当然,以上应用场景仅仅是示例说明,在其他实施例中,本申请实施例的舞蹈动作生成可以应用到任意与舞蹈动作生成相关联的场景中。

下面结合附图对本申请实施例提供的舞蹈动作生成方法进行示意性说明。

具体请参见图 1，图 1 是本申请实施例提供的一种舞蹈动作生成方法的流程示意图，本申请实施例的舞蹈动作生成方法可以由舞蹈动作生成装置执行，其中，舞蹈动作生成装置设置于终端或计算机设备中，其中，终端或计算机设备的具体解释如前。具体地，本申请实施例的方法包括如下步骤。

S101：获取待编舞音频，并从该待编舞音频中提取多个音频片段。

本申请实施例中，计算机设备在从该待编舞音频中提取多个音频片段时，可以获取待编舞音频的节拍信息，并根据节拍信息从待编舞音频中提取多个音频片段。

在一个实施例中，计算机设备在根据节拍信息从待编舞音频中提取多个音频片段时，可以根据指定节拍从待编舞音频中提取多个音频片段。例如，该指定节拍可以为 1/2 节拍，计算机设备可以提取待编舞音频中每个节拍的 1/2 节拍对应的多个音频片段。在其他实施例中，该指定节拍可以为其他节拍，本申请对此不做具体限定。

本申请通过从待编舞音频中提取多个音频片段，以为每个音频片段生成匹配的舞蹈动作，有助于提高为待编舞音频生成的舞蹈动作的质量。

S102：将多个音频片段输入预训练的编码模型，得到多个音频片段中每个音频片段的第一动作特征。

本申请实施例中，计算机设备可以将多个音频片段输入预训练的编码模型，得到多个音频片段中每个音频片段的第一动作特征，其中，所述编码模型是由样本音频和样本音频对应的样本舞蹈动作对初始的编码模型进行训练得到的。在某些实施例中，该第一动作特征的数据形式可以包括但不限于矩阵、多边形网格（Polygon Mesh Data, MMD）、三维通用模型格式（FilmBox, FBX）等数据形式。

在一个实施例中，计算机设备在将多个音频片段输入预训练的编码模型，得到多个音频片段中每个音频片段的第一动作特征时，可以将多个音频片段中的每个音频片段输入预训练的编码模型，得到与每个音频片段对应的动作特征向量，并根据每个音频片段对应的动作特征向量确定第一动作特征。

进一步地，当第一动作特征的数据形式为矩阵时，计算机设备在根据每个音频片段对应的动作特征向量确定第一动作特征时，可以将每个音频片段中每个动作对应的动作特征向量作为行向量，从而根据多个音频片段中每个音频片段的每个动作对应的动作特征向量组成第一动作特征，其中，一个音频片段可以包括一个或多个动作，每个动作对应一个动作特征向量。例如，第一动作特征可以为 $t \times 512$ 的矩阵， t 为音频片段的数量。

当第一动作特征的数据形式为矩阵时，矩阵的列数可以根据动作库中的动作的数量来确定，例如，动作库中的动作的数量为 64，则可以确定第一动作特征矩阵为 $t \times 64$ 的矩阵。

S103：根据每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与第一动作特征相似的第二个动作特征。

本申请实施例中，该动作库中是预先存储多种舞蹈类别的舞蹈动作的动作特征，该动作特征的数据形式包括但不限于矩阵。例如，该动作库可以为 $T \times 24 \times 3$ 的矩阵， T 配置为指示 T 帧舞蹈动作，24 配置为指示人体中该舞蹈动作的 24 个关键点，3 配置为指示每个舞蹈动作具有三个维度的坐标位置。如图 2 所示，图 2 是一种多帧舞蹈动作的示意图，如图 2 所示，每个人体动作对应为一帧舞蹈动作，图 2 包括多帧舞蹈动作。如图 3 所示，图 3 是一种人体关键点的示意图，图 3 中标注的 0-23 的数字配置为指示人体的多个关键点，其中，舞蹈动作是根据每个关键点的位置确定的，不同位置的各个关键点可以组成多个舞蹈动作。

在一个实施例中，计算机设备在根据第一动作特征，从动作库预先存储的多种舞蹈动作

的动作特征中确定与第一动作特征相似的第二个动作特征时，可以获取每个音频片段的第一动作特征中包括的多个第一动作特征向量，其中，每个第一动作特征向量配置为指示一个动作；并根据每个音频片段的每个第一动作特征向量从动作库预先存储的多种舞蹈动作的动作特征中确定与每个第一动作特征向量对应的第二动作特征向量；以及根据每个音频片段的每个第二动作特征向量确定每个音频片段的第二动作特征。

在一个实施例中，计算机设备在根据每个音频片段的每个第一动作特征向量，从动作库预先存储的多种舞蹈动作的动作特征中确定与每个第一动作特征向量对应的第二动作特征向量时，可以获取每个音频片段的每个第一动作特征向量与动作库中每个动作特征向量的距离，该动作库中的每个动作特征向量配置为指示预先存储的一个舞蹈动作；从动作库中获取与每个音频片段的每个第一动作特征向量距离最短的动作特征向量作为每个音频片段的第二动作特征向量。

在一个实施例中，计算机设备在获取每个音频片段的每个第一动作特征向量与动作库中每个动作特征向量的距离时，可以利用欧几里得算法计算每个音频片段的每个第一动作特征向量与动作库中每个动作特征向量的距离。

在一个实施例中，第一动作特征具体为第一动作特征矩阵，计算机设备可以针对每个音频片段，获取音频片段对应的第一动作特征矩阵中的各个行向量，得到音频片段对应的多个第一动作特征向量，其中，每个行向量配置为指示一个动作；针对每个音频片段的每个第一动作特征向量，从动作库预先存储的多种舞蹈动作的动作特征中确定与第一动作特征向量对应的第二动作特征向量，得到每个音频片段对应的多个第二动作特征向量；针对每个音频片段，将音频片段对应的多个第二动作特征向量进行组合，得到音频片段对应的第二动作特征矩阵，其中，第二动作特征矩阵配置为表示所述第二动作特征。

本申请通过从动作库中确定出与第一动作特征对应的第二动作特征，有助于后续更有效地根据解码模型对每个音频片段的第二动作特征进行解码，得到第三动作特征，进一步根据第三动作特征确定出待编舞音频的舞蹈动作。

S104: 将每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征，并根据第三动作特征确定待编舞音频的舞蹈动作。

本申请实施例中，计算机设备可以将每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征，并根据第三动作特征确定待编舞音频的舞蹈动作，其中，解码模型是由样本音频对应的样本动作特征以及样本音频对应的样本舞蹈动作对初始的解码模型进行训练得到的，样本音频对应的样本动作特征是利用预训练的编码模型对样本音频进行编码得到的，第三动作特征配置为指示所有音频片段的动作特征，其中，该第三动作特征包括所有音频片段中人体的关键点数量和每个关键点的位置。在某些实施例中，该第三动作特征包括但不限于第三动作特征矩阵，例如，该第三动作特征可以为 $T*24*3$ 的矩阵，其中， T 配置为指示音频片段的数量，24 表示人体的关键点，3 配置为指示每个关键点的三维坐标。

计算机设备在根据第三动作特征确定待编舞音频的舞蹈动作时，可以根据第三动作特征确定所有音频片段中人体的关键点数量和每个关键点的位置；根据所有音频片段中人体的关键点数量和每个关键点的位置确定待编舞音频的舞蹈动作。例如，假设第三动作特征为 $T*24*3$ 的矩阵，则计算机设备可以根据 $T*24*3$ 的矩阵确定所有音频片段中人体的关键点数量和每个关键点的位置，并根据所有音频片段中人体的关键点数量和每个关键点的位置确定所有音频片段中的人体动作，进一步将所有音频片段的人体动作确定为待编舞音频的舞蹈动作。

本申请实施例通过预训练的编码模型对待编舞音频的音频片段进行编码，得到对应的第一动作特征，并从动作库预先存储的多种舞蹈动作的动作特征中确定与第一动作特征对应的第二动作特征，有助于利用第二动作特征通过预训练的解码模型解码生成更准确、高质量的待编舞音频的舞蹈动作。

请参见图 4，图 4 是本申请实施例提供的另一种舞蹈动作生成方法的流程示意图，本申请实施例的舞蹈动作生成方法可以由舞蹈动作生成装置执行，其中，舞蹈动作生成装置设置于终端或计算机设备中，其中，终端或计算机设备的具体解释如前。具体地，本申请实施例主要是对编码模型的训练过程的说明，具体包括如下步骤。

S401：获取待编舞音频，并从该待编舞音频中提取多个音频片段。

S402：获取样本数据集，样本数据集包括多个样本舞曲数据，每个样本舞曲数据包括样本音频和样本舞蹈动作。

S403：根据每个样本舞曲数据的样本音频和样本舞蹈动作对初始的编码模型进行训练，得到预训练的编码模型。

在一个实施例中，计算机设备在根据每个样本舞曲数据的样本音频和样本舞蹈动作对初始的编码模型进行训练，得到预训练的编码模型时，可以从每个样本舞曲数据的样本舞蹈动作中提取样本动作特征；将样本动作特征和样本音频输入第一编码模型进行训练，得到预训练的编码模型。在某些实施例中，该样本动作特征包括但不限于矩阵的数据形式。

在一个实施例中，计算机设备在从每个样本舞曲数据的样本舞蹈动作中提取样本动作特征时，可以获取每个样本舞曲数据的每个样本舞蹈动作对应的人体的关键点数量和关键点位置，关键点位置包括每个关键点的坐标；并将每个样本舞蹈动作对应的人体的关键点数量和关键点位置输入初始的编码模型，提取得到样本动作特征。在某些实施例中，每个样本舞蹈动作的人体的关键点可以包括人体的 24 个关键点，在某些实施例中，每个样本舞蹈动作的关键点位置可以包括每个样本舞蹈动作的三维坐标数据。

在一个实施例中，计算机设备在将样本动作特征和样本音频输入初始的编码模型进行训练，得到预训练的编码模型时，可以将样本动作特征和样本音频输入初始的编码模型进行训练，得到第一样本动作特征，根据第一样本动作特征从动作库预先存储的多种舞蹈动作的动作特征确定与第一样本动作特征相似的第二样本动作特征；将第二样本动作特征输入初始的解码模型，得到第三样本动作特征；根据第三样本动作特征调整初始的解码模型的模型参数，并将第二样本动作特征输入调整后的解码模型进行训练，得到预训练的解码模型。在某些实施例中，第三样本动作特征配置为指示样本音频中人体的关键点数量以及每个关键点的位置。

在一个实施例中，计算机设备在根据第三样本动作特征调整初始的解码模型的模型参数，并将第二样本动作特征输入调整后的解码模型进行训练，得到预训练的解码模型时，可以根据第三样本动作特征确定样本音频的舞蹈动作；将确定的样本音频的舞蹈动作与样本音频的样本舞蹈动作进行对比，根据对比结果调整初始的解码模型的模型参数；将第二样本动作特征输入调整模型参数后的解码模型重新训练，以得到预训练的解码模型。

计算机设备在根据第三样本动作特征确定样本音频的舞蹈动作时，可以根据第三样本动作特征确定样本音频中人体的关键点数量和每个关键点的位置；根据样本音频中人体的关键点数量和每个关键点的位置确定样本音频的舞蹈动作。

进一步地，计算机设备在将确定的样本音频的舞蹈动作与样本音频的样本舞蹈动作进行对比时，可以将样本音频的舞蹈动作的动作特征矩阵与该样本舞蹈动作的动作特征进行对比。

进一步地，计算机设备在根据对比结果调整第一编码模型的模型参数时，可以计算样本

音频的舞蹈动作的动作特征中的各个向量与该样本舞蹈动作的动作特征的各个向量的向量距离，当样本音频的舞蹈动作的动作特征中的向量与该样本舞蹈动作的动作特征的向量的向量距离大于第一距离阈值时，根据样本音频的舞蹈动作的动作特征中的向量与该样本舞蹈动作的动作特征矩阵的向量的向量距离调整初始的编码模型的模型参数。

本申请通过初始的编码模型可以得到各个样本舞蹈动作与样本动作特征的映射关系，有助于根据样本动作特征和样本音频训练初始的编码模型，得到预训练的编码模型，并得到样本动作特征、样本音频、样本音频的舞蹈动作的动作特征矩阵三者之间的映射关系。

S404：将多个音频片段输入预训练的编码模型，得到多个音频片段中每个音频片段的第一动作特征。

S405：根据每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与每个音频片段的第一动作特征相似的第二动作特征。

S406：将每个音频片段的第二动作特征输入预训练的解码模型，得到第三动作特征，并根据第三动作特征确定待编舞音频的舞蹈动作。

本申请实施例通过初始的编码模型得到各个样本舞蹈动作与样本动作特征的映射关系，有助于根据样本动作特征和样本音频训练初始的编码模型，得到预训练的编码模型，并得到样本动作特征、样本音频、第一样本动作特征三者之间的映射关系，以实现在测试时通过预训练的编码模型生成待编舞音频的第一动作特征。

请参见图 5，图 5 是本申请实施例提供的又一种舞蹈动作生成方法的流程示意图，本申请实施例的舞蹈动作生成方法可以由舞蹈动作生成装置执行，其中，舞蹈动作生成装置设置于终端或计算机设备中，其中，终端或计算机设备的具体解释如前。具体地，本申请实施例主要是对解码模型的训练过程的说明，具体包括如下步骤。

S501：获取待编舞音频，并从该待编舞音频中提取多个音频片段。

S502：将多个音频片段输入预训练的编码模型，得到多个音频片段中每个音频片段的第一动作特征。

S503：根据每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与每个音频片段的第一动作特征相似的第二动作特征。

S504：将样本音频输入预训练的编码模型，得到与样本音频对应的第一样本动作特征。

S505：根据第一样本动作特征对预设的解码模型进行训练，得到预训练的解码模型。

在一个实施例中，计算机设备在根据第一样本动作特征对初始的解码模型进行训练，得到预训练的解码模型时，可以根据第一样本动作特征从动作库预先存储的多种舞蹈动作的动作特征确定与第一样本动作特征相似的第二样本动作特征；将第二样本动作特征输入初始的解码模型进行训练，得到预训练的解码模型。

在一个实施例中，计算机设备在将第二样本动作特征输入初始的解码模型进行训练，得到预训练的解码模型时，可以将第二样本动作特征输入初始的解码模型，得到第三样本动作特征；根据第三样本动作特征调整初始的解码模型的模型参数，并将第二样本动作特征输入调整后的解码模型进行训练，得到所述预训练的解码模型。

在一个种实施方式中，计算机设备可以根据第三样本动作特征确定样本音频的舞蹈动作；将确定的样本音频的舞蹈动作与样本音频的样本舞蹈动作进行对比，根据对比结果调整初始的解码模型的模型参数；将第二样本动作特征输入调整模型参数后的解码模型重新训练，得到预训练的解码模型。

计算机设备在根据第三样本动作特征确定样本音频的舞蹈动作时，可以根据第三样本动

作特征确定样本音频中人体的关键点数量和每个关键点的位置；根据样本音频中人体的关键点数量和每个关键点的位置确定样本音频的舞蹈动作。

在一个种实施方式中，计算机设备在根据第三样本动作特征调整模型参数时，可以根据第三样本动作特征和样本动作特征调整初始的解码模型的模型参数。

进一步地，计算机设备在根据第三样本动作特征和样本动作特征调整初始的解码模型的模型参数时，可以根据第三样本动作特征矩阵中的向量与样本动作特征矩阵中的向量之间的向量距离调整初始的解码模型的模型参数。

进一步地，当第三样本动作特征中的向量与样本动作特征中的向量之间的向量距离大于第二距离阈值时，计算机设备可以根据第三样本动作特征中的向量与样本动作特征中的向量之间的向量距离调整初始的解码模型的模型参数。

S506:将每个音频片段对应的第二动作特征输入预训练的解码模型,得到第三动作特征,并根据第三动作特征确定待编舞音频的舞蹈动作。

本申请实施例通过预训练的编码模型生成样本音频的第一样本动作特征,并根据第一样本动作特征对初始的解码模型进行训练,得到预训练的解码模型,有助于在测试时更准确地生成待编舞音频的第三动作特征,从而生成更准确,质量更高的舞蹈动作。

请参见图6,图6是本申请实施例提供的又一种舞蹈动作生成方法的流程示意图,通过获取待编舞音频61,并从待编舞音频中提取多个音频片段62,将多个音频片段输入预训练的编码模型,得到多个音频片段中每个音频片段的第一动作特征63,从动作库预先存储的多种舞蹈动作的动作特征中确定与每个音频片段的第一动作特征相似的第二动作特征,将每个音频片段的第二动作特征输入预训练的解码模型,得到所有音频片段的第三动作特征64,从而根据第三动作特征确定待编舞音频的舞蹈动作。

请参见图7,图7是本申请实施例提供的一种舞蹈动作生成装置的结构示意图。具体的,舞蹈动作生成装置设置于计算机设备中,装置包括:获取单元701、编码单元702、确定单元703、解码单元704;

获取单元701,配置为获取待编舞音频,并从所述待编舞音频中提取多个音频片段;

编码单元702,配置为将所述多个音频片段输入预训练的编码模型,得到所述多个音频片段中每个音频片段的第一动作特征,其中,所述编码模型是由样本音频和所述样本音频对应的样本舞蹈动作训练得到的;

确定单元703,配置为根据所述每个音频片段的第一动作特征从动作库预先存储的多种舞蹈动作的动作特征中确定与所述每个音频片段的第一动作特征相似的第二动作特征;

解码单元704,配置为将所述每个音频片段对应的第二动作特征输入预训练的解码模型,得到第三动作特征,并根据所述第三动作特征确定所述待编舞音频的舞蹈动作,其中,所述解码模型是由所述样本音频对应的样本动作特征以及所述样本音频对应的所述样本舞蹈动作训练得到的,所述样本音频对应的样本动作特征是利用所述预训练的编码模型对所述样本音频进行编码得到的,所述第三动作特征配置为指示所有音频片段的动作特征。

进一步地,获取单元701从所述待编舞音频中提取多个音频片段时,具体配置为:

获取所述待编舞音频的节拍信息;

根据所述节拍信息从所述待编舞音频中提取多个音频片段。

进一步地,所述第一动作特征具体为第一动作特征矩阵;确定单元703根据所述每个音频片段的第一动作特征,从动作库预先存储的多种舞蹈动作的动作特征中确定与所述每个音频片段的第一动作特征相似的第二动作特征时,具体配置为:

针对每个所述音频片段，获取所述音频片段对应的第一动作特征矩阵中的各个行向量，得到所述音频片段对应的多个第一动作特征向量，其中，每个所述行向量配置为指示一个动作；

针对所述每个音频片段的每个所述第一动作特征向量，从所述动作库预先存储的多种舞蹈动作的动作特征中确定与所述第一动作特征向量对应的第二动作特征向量，得到所述每个音频片段对应的多个第二动作特征向量；

针对所述每个音频片段，将所述音频片段对应的多个第二动作特征向量进行组合，得到所述音频片段对应的第二动作特征矩阵，其中，所述第二动作特征矩阵配置为表示所述第二动作特征。

进一步地，确定单元 703 从所述动作库预先存储的多种舞蹈动作的动作特征中确定与所述第一动作特征向量对应的第二动作特征向量时，具体配置为：

获取所述第一动作特征向量与所述动作库中每个动作特征向量的距离，所述动作库中的所述每个动作特征向量配置为指示预先存储的一个舞蹈动作；

从所述动作库中获取与所述第一动作特征向量距离最短的动作特征向量，作为所述第一动作特征向量对应的第二动作特征向量。

进一步地，编码单元 702 将所述多个音频片段输入预训练的编码模型，得到与所述多个音频片段中每个音频片段的第一动作特征矩阵之前，还配置为：

获取样本数据集，所述样本数据集包括多个样本舞曲数据，每个样本舞曲数据包括样本音频和样本舞蹈动作；

从所述每个样本舞曲数据的样本舞蹈动作中提取样本动作特征；

将所述样本动作特征和所述样本音频输入初始的编码模型进行训练，得到所述预训练的编码模型。

进一步地，编码单元 702 从所述每个样本舞曲数据的样本舞蹈动作中提取样本动作特征时，配置为：

获取每个所述样本舞蹈动作对应的人体的关键点数量和关键点位置；

将所述每个样本舞蹈动作对应的人体的关键点数量和关键点位置输入所述初始的编码模型，提取所述样本动作特征。

进一步地，解码单元 704 将所述每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征之前，还配置为：

将所述样本音频输入所述预训练的编码模型，得到所述样本音频的第一样本动作特征；并根据所述第一样本动作特征从动作库确定与所述第一样本动作特征相似的第二样本动作特征；

将所述第二样本动作特征输入初始的解码模型，得到第三样本动作特征；并根据所述第三样本动作特征确定所述样本音频的舞蹈动作；

将确定的所述样本音频的舞蹈动作与所述样本音频的样本舞蹈动作进行对比，根据对比结果调整所述初始的解码模型的模型参数，以得到所述预训练的解码模型。

进一步地，解码单元 704 根据所述第三动作特征确定所述待编舞音频的舞蹈动作时，具体配置为：

根据所述第三动作特征确定所述待编舞音频中人体的关键点数量和每个关键点的位置；

根据所述待编舞音频中人体的关键点数量和每个关键点的位置确定所述待编舞音频的舞蹈动作。

本申请实施例通过预训练的编码模型对待编舞音频的音频片段进行编码，得到每个音频片段的第一动作特征，并从动作库预先存储的多种舞蹈动作的动作特征中确定与第一动作特征相似的第二动作特征，有助于利用第二动作特征通过预训练的解码模型解码生成更准确、高质量的待编舞音频的舞蹈动作。

请参见图 8，图 8 是本申请实施例提供的一种计算机设备的结构示意图。具体的，所述计算机设备包括：存储器 801、处理器 802。

在一种实施例中，所述计算机设备还包括数据接口 803，所述数据接口 803，配置为传递计算机设备和其他设备之间的数据信息。

所述存储器 801 可以包括易失性存储器 (volatile memory)；存储器 801 也可以包括非易失性存储器 (non-volatile memory)；存储器 801 还可以包括上述种类的存储器的组合。所述处理器 802 可以是中央处理器 (central processing unit, CPU)。所述处理器 802 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) 或其任意组合。

所述存储器 801 配置为存储程序，所述处理器 802 可以调用存储器 801 中存储的程序，配置为执行如下步骤：

获取待编舞音频，并从所述待编舞音频中提取多个音频片段；

将所述多个音频片段输入预训练的编码模型，得到所述多个音频片段中每个音频片段的第一动作特征，其中，所述编码模型是由样本音频和所述样本音频对应的样本舞蹈动作训练得到的；

根据所述每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与所述每个音频片段的第一动作特征相似的第二动作特征；

将所述每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征，并根据所述第三动作特征确定所述待编舞音频的舞蹈动作，其中，所述解码模型是由所述样本音频对应的样本动作特征以及所述样本音频对应的所述样本舞蹈动作训练得到的，所述样本音频对应的样本动作特征是利用所述预训练的编码模型对所述样本音频进行编码得到的，所述第三动作特征配置为指示所有音频片段的动作特征。

进一步地，处理器 802 从所述待编舞音频中提取多个音频片段时，具体配置为：

获取所述待编舞音频的节拍信息；

根据所述节拍信息从所述待编舞音频中提取多个音频片段。

进一步地，所述第一动作特征具体为第一动作特征矩阵；处理器 802 根据所述每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与所述每个音频片段的第一动作特征相似的第二动作特征时，具体配置为：

针对每个所述音频片段，获取所述音频片段对应的第一动作特征矩阵中的各个行向量，得到所述音频片段对应的多个第一动作特征向量，其中，每个所述行向量配置为指示一个动作；

针对所述每个音频片段的每个所述第一动作特征向量，从所述动作库预先存储的多种舞蹈动作的动作特征中确定与所述第一动作特征向量对应的第二动作特征向量，得到所述每个音频片段对应的多个第二动作特征向量；

针对所述每个音频片段，将所述音频片段对应的多个第二动作特征向量进行组合，得到

所述音频片段对应的第二动作特征矩阵，其中，所述第二动作特征矩阵配置为表示所述第二动作特征。

进一步地，处理器 802 从所述动作库预先存储的多种舞蹈动作的动作特征中确定与所述第一动作特征向量对应的第二动作特征向量时，具体配置为：

获取所述第一动作特征向量与所述动作库中每个动作特征向量的距离，所述动作库中的所述每个动作特征向量配置为指示预先存储的一个舞蹈动作；

从所述动作库中获取与所述第一动作特征向量距离最短的动作特征向量，作为所述第一动作特征向量对应的第二动作特征向量。

进一步地，处理器 802 将所述多个音频片段输入预训练的编码模型，得到与所述多个音频片段中每个音频片段的第一动作特征矩阵之前，还配置为：

获取样本数据集，所述样本数据集包括多个样本舞曲数据，每个样本舞曲数据包括样本音频和样本舞蹈动作；

从所述每个样本舞曲数据的样本舞蹈动作中提取样本动作特征；

将所述样本动作特征和所述样本音频输入初始的编码模型进行训练，得到所述预训练的编码模型。

进一步地，处理器 802 从所述每个样本舞曲数据的样本舞蹈动作中提取样本动作特征时，配置为：

获取每个所述样本舞蹈动作对应的人体的关键点数量和关键点位置；

将所述每个样本舞蹈动作对应的人体的关键点数量和关键点位置输入所述初始的编码模型，提取所述样本动作特征。

进一步地，处理器 802 将所述每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征之前，还配置为：

将所述样本音频输入所述预训练的编码模型，得到所述样本音频的第一样本动作特征；并根据所述第一样本动作特征从动作库确定与所述第一样本动作特征相似的第二样本动作特征；

将所述第二样本动作特征输入初始的解码模型，得到第三样本动作特征；并根据所述第三样本动作特征确定所述样本音频的舞蹈动作；

将确定的所述样本音频的舞蹈动作与所述样本音频的样本舞蹈动作进行对比，根据对比结果调整所述初始的解码模型的模型参数，以得到所述预训练的解码模型。

进一步地，处理器 802 根据所述第三动作特征确定所述待编舞音频的舞蹈动作时，具体配置为：

根据所述第三动作特征确定所述待编舞音频中人体的关键点数量和每个关键点的位置；

根据所述待编舞音频中人体的关键点数量和每个关键点的位置确定所述待编舞音频的舞蹈动作。

本申请实施例通过预训练的编码模型对待编舞音频的音频片段进行编码，得到每个音频片段的第一动作特征，并从动作库预先存储的多种舞蹈动作的动作特征中确定与每个音频片段的第一动作特征相似的第二动作特征，有助于利用每个音频片段的第二动作特征通过预训练的解码模型解码生成更准确、高质量的待编舞音频的舞蹈动作。

本申请的实施例还提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现本申请图 1、图 4 或图 5 所对应实施例中描述的方法，也可实现图 7 所述本申请所对应实施例的装置，在此不再赘述。

所述计算机可读存储介质可以是前述任一实施例所述的设备的内部存储单元，例如设备的硬盘或内存。所述计算机可读存储介质也可以是所述设备的外部存储设备，例如所述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，所述计算机可读存储介质还可以既包括所述设备的内部存储单元也包括外部存储设备。所述计算机可读存储介质配置为存储所述计算机程序以及所述终端所需的其他程序和数据。所述计算机可读存储介质还可以配置为暂时地存储已经输出或者将要输出的数据。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一计算机可读存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，所述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory, ROM）或随机存储记忆体（Random Access Memory, RAM）等。

以上所揭露的仅为本申请的部分实施例而已，当然不能以此来限定本申请之权利范围，本领域普通技术人员可以理解实现上述实施例的全部或部分流程，并依本申请权利要求所作的等同变化，仍属于本发明所涵盖的范围。

权 利 要 求 书

1. 一种舞蹈动作生成方法，其中，包括：

获取待编舞音频，并从所述待编舞音频中提取多个音频片段；

将所述多个音频片段输入预训练的编码模型，得到所述多个音频片段中每个音频片段的第一动作特征，其中，所述编码模型是由样本音频和所述样本音频对应的样本舞蹈动作训练得到的；

根据所述每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与所述每个音频片段的第一动作特征相似的第二动作特征；

将所述每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征，并根据所述第三动作特征确定所述待编舞音频的舞蹈动作，其中，所述解码模型是由所述样本音频对应的样本动作特征以及所述样本音频对应的所述样本舞蹈动作训练得到的，所述样本音频对应的样本动作特征是利用所述预训练的编码模型对所述样本音频进行编码得到的，所述第三动作特征配置为指示所有音频片段的动作特征。

2. 根据权利要求 1 所述的方法，其中，所述从所述待编舞音频中提取多个音频片段，包括：

获取所述待编舞音频的节拍信息；

根据所述节拍信息从所述待编舞音频中提取多个音频片段。

3. 根据权利要求 1 所述的方法，其中，所述第一动作特征具体为第一动作特征矩阵；所述根据所述每个音频片段的第一动作特征，从动作库预先存储的多种舞蹈动作的动作特征中确定与所述每个音频片段的第一动作特征相似的第二动作特征，包括：

针对每个所述音频片段，获取所述音频片段对应的第一动作特征矩阵中的各个行向量，得到所述音频片段对应的多个第一动作特征向量，其中，每个所述行向量配置为指示一个动作；

针对所述每个音频片段的每个所述第一动作特征向量，从所述动作库预先存储的多种舞蹈动作的动作特征中确定与所述第一动作特征向量对应的第二动作特征向量，得到所述每个音频片段对应的多个第二动作特征向量；

针对所述每个音频片段，将所述音频片段对应的多个第二动作特征向量进行组合，得到所述音频片段对应的第二动作特征矩阵，其中，所述第二动作特征矩阵配置为表示所述第二动作特征。

4. 根据权利要求 3 所述的方法，其中，从所述动作库预先存储的多种舞蹈动作的动作特征中确定与所述第一动作特征向量对应的第二动作特征向量，包括：

获取所述第一动作特征向量与所述动作库中每个动作特征向量的距离，所述动作库中的所述每个动作特征向量配置为指示预先存储的一个舞蹈动作；

从所述动作库中获取与所述第一动作特征向量距离最短的动作特征向量，作为所述第一动作特征向量对应的第二动作特征向量。

5. 根据权利要求 1 所述的方法，其中，所述将所述多个音频片段输入预训练的编码模型，得到与所述多个音频片段中每个音频片段的第一动作特征矩阵之前，还包括：

获取样本数据集，所述样本数据集包括多个样本舞曲数据，每个样本舞曲数据包括样本音频和样本舞蹈动作；

从所述每个样本舞曲数据的样本舞蹈动作中提取样本动作特征；

将所述样本动作特征和所述样本音频输入初始的编码模型进行训练，得到所述预训练的编码

模型。

6.根据权利要求 5 所述的方法，其中，所述从所述每个样本舞曲数据的样本舞蹈动作中提取样本动作特征，包括：

获取每个所述样本舞蹈动作对应的人体的关键点数量和关键点位置；

将所述每个样本舞蹈动作对应的人体的关键点数量和关键点位置输入所述初始的编码模型，提取所述样本动作特征。

7.根据权利要求 1 所述的方法，其中，所述将所述每个音频片段对应的第二动作特征输入预训练的解码模型，得到第三动作特征之前，还包括：

将所述样本音频输入所述预训练的编码模型，得到所述样本音频的第一样本动作特征；并根据所述第一样本动作特征从动作库确定与所述第一样本动作特征相似的第二样本动作特征；将所述第二样本动作特征输入初始的解码模型，得到第三样本动作特征；并根据所述第三样本动作特征确定所述样本音频的舞蹈动作；

将确定的所述样本音频的舞蹈动作与所述样本音频的样本舞蹈动作进行对比，根据对比结果调整所述初始的解码模型的模型参数，以得到所述预训练的解码模型。

8.根据权利要求 1 所述的方法，其中，所述根据所述第三动作特征确定所述待编舞音频的舞蹈动作，包括：

根据所述第三动作特征确定所述待编舞音频中人体的关键点数量和每个关键点的位置；

根据所述待编舞音频中人体的关键点数量和每个关键点的位置确定所述待编舞音频的舞蹈动作。

9.一种计算机设备，其中，包括处理器、输入设备、输出设备和存储器，所述处理器、输入设备、输出设备和存储器相互连接，其中，所述存储器配置为存储计算机程序，所述计算机程序包括程序，所述处理器被配置配置为调用所述程序，执行如权利要求 1-8 任一项所述的方法。

10.一种计算机可读存储介质，其中，该计算机可读存储介质中存储有程序指令，该程序指令被执行时，配置为实现如权利要求 1-8 任一项所述的方法。