(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111047548 A
(43)申请公布日 2020.04.21

(21)申请号 202010168492.0

(22)申请日 2020.03.12

(71)申请人 腾讯科技(深圳)有限公司
地址 518000 广东省深圳市南山区高新区科技中一路腾讯大厦35层

(72)发明人 吴现 李琛 戴宇荣

(74)专利代理机构 广州华进联合专利商标代理有限公司 44224
代理人 毛丹

(51)Int.Cl.
G06T 5/50(2006.01)
G06T 7/10(2017.01)
G06N 3/04(2006.01)

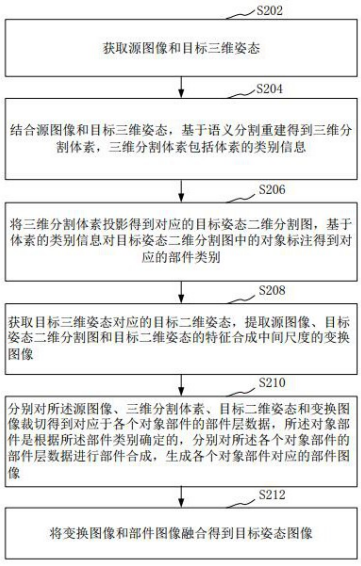
权利要求书3页 说明书17页 附图9页

(54)发明名称

姿态变换数据处理方法、装置、计算机设备和存储介质

(57)摘要

本申请涉及一种姿态变换数据处理方法、装置、计算机设备和存储介质，涉及人工智能的图像处理技术，包括：获取源图像和目标三维姿态，基于语义分割重建得到包括体素的类别信息的三维分割体素，将三维分割体素投影得到对应的目标姿态二维分割图，基于类别信息对目标姿态二维分割图中的对象标注得到部件类别；获取目标三维姿态对应的目标二维姿态，提取源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像；分别对源图像、三维分割体素、目标二维姿态和变换图像裁切得到各个对象部件的部件层数据，分别对各个对象部件的部件层数据进行部件合成，生成部件图像；将变换图像和部件图像融合得到目标姿态图像，提高姿态变换图像质量。



CN 111047548 A

1.一种姿态变换数据处理方法,所述方法包括:

获取源图像和目标三维姿态;

结合所述源图像和目标三维姿态,基于语义分割重建得到三维分割体素,所述三维分割体素包括体素的类别信息;

将所述三维分割体素投影得到对应的目标姿态二维分割图,基于所述体素的类别信息对所述目标姿态二维分割图中的对象标注得到对应的部件类别;

获取所述目标三维姿态对应的目标二维姿态,提取所述源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像;

分别对所述源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据,所述对象部件是根据所述部件类别确定的;

分别对所述各个对象部件的部件层数据进行部件合成,生成各个对象部件对应的部件图像;

将所述变换图像和所述部件图像融合得到目标姿态图像。

2.根据权利要求1所述的方法,其特征在于,所述结合所述源图像和目标三维姿态,基于语义分割重建得到三维分割体素包括:

将所述源图像和目标三维姿态输入三维体素网络;

所述三维体素网络对所述源图像和目标三维姿态进行编码,得到编码结果,对所述编码结果提取特征并解码输出携带体素的类别信息的三维分割体素。

3.根据权利要求2所述的方法,其特征在于,所述三维体素网络对所述源图像和目标三维姿态进行编码,得到编码结果,对所述编码结果提取特征并解码输出携带体素的类别信息的三维分割体素包括:

对所述源图像进行编码并提取特征得到第一特征;

对所述目标三维姿态进行编码并提取特征得到第二特征;

将所述第一特征和第二特征合并得到合并特征,将所述合并特征输入堆叠沙漏网络,通过所述堆叠沙漏网络解码得到所述三维分割体素。

4.根据权利要求1所述的方法,其特征在于,所述提取所述源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像包括:

将所述源图像、目标姿态二维分割图和目标二维姿态合并形成输入矩阵;

将所述输入矩阵输入至中间尺度生成网络;

所述中间尺度生成网络依次通过下采样层、残差块层和上采样层对所述输入矩阵进行特征提取得到所述中间尺度的变换图像。

5.根据权利要求4所述的方法,其特征在于,所述中间尺度生成网络的训练包括以下步骤:

获取第一训练样本;所述第一训练样本包括源训练图像、目标姿态训练二维分割图、目标训练二维姿态和对应的标签变换图像;

将所述源训练图像、目标姿态训练二维分割图、目标训练二维姿态输入中间尺度生成网络,经过所述中间尺度生成网络所包括的各层依次处理后,输出对应的训练变换图像;

根据所述训练变换图像和所述标签变换图像的差异反向传播调整所述中间尺度生成网络的网络参数,以得到已训练的中间尺度生成网络。

6.根据权利要求5所述的方法，其特征在于，所述根据所述训练变换图像和所述标签变换图像的差异反向传播调整所述中间尺度生成网络的网络参数，以得到已训练的中间尺度生成网络包括：

通过预训练感知网络分别提取标签变换图像和训练变换图像的特征得到特征图，计算特征图之间的距离得到感知损失；

根据所述标签变换图像、源训练图像、训练变换图像对中间尺度生成网络和判别网络进行对抗学习，得到对抗损失；

通过判别网络中多个不同尺度的卷积层分别计算标签变换图像和训练变换图像在多个不同尺度的特征距离，统计所述多个不同尺度的特征距离得到特征匹配损失；

根据所述感知损失、对抗损失、特征匹配损失确定目标损失，根据所述目标损失反向传播调整所述中间尺度生成网络的网络参数，以得到已训练的中间尺度生成网络。

7.根据权利要求1所述的方法，其特征在于，所述分别对所述源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据包括：

获取所述源图像对应的源二维分割图，根据所述源二维分割图对所述源图像进行分割得到对应于各个对象部件的部件层数据；

获取各个对象部件对应的裁切信息，基于各个对象部件的中心位置，对所述三维分割体素、目标二维姿态、变换图像分别裁切得到与对应的裁切信息匹配的部件层数据。

8.根据权利要求1所述的方法，其特征在于，所述分别对所述各个对象部件的部件层数据进行部件合成，生成各个对象部件对应的部件图像包括：

获取所述各个对象部件对应的层级生成网络；

将部件层数据输入匹配的对象部件的层级生成网络；

各个对象部件的层级生成网络分别输出与对象部件匹配的部件图像。

9.根据权利要求8所述的方法，其特征在于，所述层级生成网络的训练包括以下步骤：

获取第二训练样本；所述第二训练样本包括源训练部件图像、训练部件三维分割体素、目标训练部件二维姿态、部件变换图像和对应的标签部件图像，所述第二训练样本中的各个样本对应当前对象部件；

将所述源训练部件图像、训练部件三维分割体素、目标训练部件二维姿态、部件变换图像输入所述当前对象部件对应的层级生成网络，经过所述层级生成网络所包括的各层依次处理后，输出对应的训练部件图像；

根据所述训练部件图像和所述标签部件图像的差异反向传播调整所述层级生成网络的网络参数，以得到已训练的当前对象部件对应的层级生成网络。

10.根据权利要求9所述的方法，其特征在于，所述根据所述训练部件图像和所述标签部件图像的差异反向传播调整所述层级生成网络的网络参数，以得到已训练的当前对象部件对应的层级生成网络包括：

根据所述标签部件图像、源训练部件图像、训练部件图像对当前对象部件对应的层级生成网络和判别网络进行对抗学习，得到部件对抗损失，所述标签部件图像是训练集中未被遮挡的与当前对象部件对应的图像；

通过预训练感知网络分别提取标签部件图像和训练部件图像的特征得到特征图，计算特征图之间的距离得到部件感知损失；

根据所述部件对抗损失、部件感知损失确定目标部件损失；

根据所述目标部件损失反向传播调整所述当前对象部件对应的层级生成网络的网络参数，以得到已训练的层级生成网络。

11.根据权利要求1至10中任一项所述的方法，其特征在于，将所述变换图像和所述部件图像融合得到目标姿态图像包括：

将所述部件图像根据二维分割图融合到所述变换图像的对应对象部件区域得到初始全局姿态图像；

将所述源图像、目标姿态二维分割图、目标二维姿态和初始全局姿态图像合并输入至精细尺度生成网络，输出得到目标姿态残差图像；

将所述目标姿态残差图像与所述初始全局姿态图像融合得到所述目标姿态图像。

12.一种姿态变换数据处理装置，所述装置包括：

获取模块，用于获取源图像和目标三维姿态；

三维分割体素模块，用于结合所述源图像和目标三维姿态，基于语义分割重建得到三维分割体素，所述三维分割体素包括体素的类别信息；

投影模块，用于将所述三维分割体素投影得到对应的目标姿态二维分割图，基于所述体素的类别信息对所述目标姿态二维分割图中的对象标注得到对应的部件类别；

中间尺度模块，用于获取所述目标三维姿态对应的目标二维姿态，提取所述源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像；

部件图像生成模块，用于分别对所述源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据，所述对象部件是根据所述部件类别确定的，分别对所述各个对象部件的部件层数据进行部件合成，生成各个对象部件对应的部件图像；

融合模块，用于将所述变换图像和所述部件图像融合得到目标姿态图像。

13.根据权利要求12所述的装置，其特征在于，所述三维分割体素模块还用于将所述源图像和目标三维姿态输入三维体素网络，所述三维体素网络对所述源图像和目标三维姿态进行编码，得到编码结果，对所述编码结果提取特征并解码输出携带体素的类别信息的三维分割体素。

14.一种计算机设备，其特征在于，包括存储器和处理器，所述存储器中存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行权利要求1至11中任一项权利要求所述方法的步骤。

15.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行权利要求1至11中任一项权利要求所述方法的步骤。

姿态变换数据处理方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及图像处理领域，特别是涉及姿态变换数据处理方法、装置、计算机设备和存储介质。

背景技术

[0002] 姿态变换是新兴的图像合成研究技术，目的是将给定的对象图像转换到目标姿态，同时维持对象外观的一致性。

[0003] 相关技术中，可以采用深度学习技术解决姿态变换问题，通过空间变换模块将给定图像特征形变至目标姿态图像，可是当视觉特征因为姿态间隔太大而改变时，将难以得到令人满意的结果，也无法解决遮挡问题，导致姿态变换图像的质量低。

发明内容

[0004] 基于此，有必要针对上述技术问题，提供一种姿态变换数据处理方法、装置、计算机设备和存储介质，基于语义的对象部件合成高质量的局部部件图像，并且使用带语义分割的三维体素表示来消除二维表示的歧义问题，能够提高目标姿态图像的质量。

[0005] 一种姿态变换数据处理方法，所述方法包括：获取源图像和目标三维姿态；

结合所述源图像和目标三维姿态，基于语义分割重建得到三维分割体素，所述三维分割体素包括体素的类别信息；将所述三维分割体素投影得到对应的目标姿态二维分割图，基于所述体素的类别信息对所述目标姿态二维分割图中的对象标注得到对应的部件类别；获取所述目标三维姿态对应的目标二维姿态，提取所述源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像；分别对所述源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据，所述对象部件是根据所述部件类别确定的；分别对所述各个对象部件的部件层数据进行部件合成，生成各个对象部件对应的部件图像；将所述变换图像和所述部件图像融合得到目标姿态图像。

[0006] 一种姿态变换数据处理装置，所述装置包括：获取模块，用于获取源图像和目标三维姿态；三维分割体素模块，用于结合所述源图像和目标三维姿态，基于语义分割重建得到三维分割体素，所述三维分割体素包括体素的类别信息；投影模块，用于将所述三维分割体素投影得到对应的目标姿态二维分割图，基于所述体素的类别信息对所述目标姿态二维分割图中的对象标注得到对应的部件类别；中间尺度模块，用于获取所述目标三维姿态对应的目标二维姿态，提取所述源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像；部件图像生成模块，用于分别对所述源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据，所述对象部件是根据所述部件类别确定的，分别对所述各个对象部件的部件层数据进行部件合成，生成各个对象部件对应的部件图像；融合模块，用于将所述变换图像和所述部件图像融合得到目标姿态图像。

[0007] 一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行上述姿态变换数据处理方法的步骤。

[0008] 一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行上述姿态变换数据处理方法的步骤。

[0009] 上述姿态变换数据处理方法、装置、计算机设备和存储介质，采用带分割的三维体素表示来消除二维表示所带来的歧义问题，包含了源图像的人体形状和目标姿态信息，不仅能表示姿态变换后的对象三维坐标，还能识别出各个部件类别，从而使得后续高质量的局部部件合成得以实现，分别独立合成完整的各个对象部件，从而解决各部件间的遮挡问题并保证局部合成结果的高质量，通过生成变换图像，从粗到细的框架让最终结果达到高清分辨率，融合得到的目标姿态图像具有非常高的鲁棒性和准确度。

附图说明

[0010] 图1为一些实施例中提供的姿态变换数据处理方法的应用环境图。

[0011] 图2为一些实施例中姿态变换数据处理方法的流程图。

[0012] 图3为一些实施例中中间尺度生成网络的训练的流程图。

[0013] 图4为一些实施例中层级生成网络的训练的流程图。

[0014] 图5为一些实施例中人体姿态变换网络的示意图。

[0015] 图6为一些实施例中人脸层级表示的详细结构示意图。

[0016] 图7为一些实施例中部件图像合成结果的示意图。

[0017] 图8为一些实施例中在视频动作序列上的部分结果示意图。

[0018] 图9为一些实施例中人体姿态变换结果对比的示意图。

[0019] 图10为一些实施例中人体姿态变换结果对比的示意图。

[0020] 图11为一些实施例中人体姿态变换结果对比的示意图。

[0021] 图12为一些实施例中模型生成图像剥离实验的评测结果。

[0022] 图13为一些实施例中姿态变换数据处理装置的结构图。

[0023] 图14为一些实施例中计算机设备的内部结构框图。

具体实施方式

[0024] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0025] 可以理解，本申请所使用的术语“第一”、“第二”等可在本文中用于描述各种元件，但除非特别说明，这些元件不受这些术语限制。这些术语仅用于将第一个元件与另一个元件区分。

[0026] 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。

[0027] 人工智能技术是一门综合学科，涉及领域广泛，既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、

大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术包括计算机视觉技术以及机器学习/深度学习等方向。

[0028] 计算机视觉技术(Computer Vision, CV)计算机视觉是一门研究如何使机器“看”的科学,更进一步的说,就是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等机器视觉,并进一步做图形处理,使电脑处理成为更适合人眼观察或传送给仪器检测的图像。作为一个科学学科,计算机视觉研究相关的理论和技术,试图建立能够从图像或者多维数据中获取信息的人工智能系统。

[0029] 机器学习(Machine Learning, ML)是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0030] 本申请的方案涉及基于人工智能的图像处理技术。具体通过如下实施例进行说明。

[0031] 图1为一些实施例中提供的姿态变换数据处理的应用环境图。如图1所示,在该应用环境中,包括终端110以及服务器120。终端110可以将源图像和目标三维姿态发送到服务器120中,服务器120可以结合源图像和目标三维姿态,基于语义分割重建得到包括体素的类别信息的三维分割体素;将三维分割体素投影得到对应的目标姿态二维分割图,基于体素的类别信息对目标姿态二维分割图中的对象标注得到对应的部件类别。获取目标三维姿态对应的目标二维姿态,提取所述源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像;分别对源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据,对象部件是根据部件类别确定的;分别对各个对象部件的部件层数据进行部件合成,生成各个对象部件对应的部件图像;将变换图像和部件图像融合得到目标姿态图像。服务器120将目标姿态图像返回给终端110,终端110可以显示目标姿态图像。

[0032] 在一些实施例中,也可以在终端110根据源图像和目标三维姿态,通过上述实施例的步骤得到并显示目标姿态图像。终端110中可以安装有用于姿态变换的应用程序,该应用程序上的部署包括三维分割体素模块、中间尺度模块和部件图像生成模块。

[0033] 服务器120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,可以是提供云服务器、云数据库、云存储和CDN等基础云计算服务的云服务器。终端110可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表等,但并不局限于此。终端110以及服务器120可以通过网络等通讯连接方式进行连接,本申请在此不做限制。

[0034] 如图2所示,在一些实施例中,提出了一种姿态变换数据处理方法,本实施例主要以该方法应用于上述图1中的服务器120或终端110来举例说明。具体可以包括以下步骤:

步骤S202,获取源图像和目标三维姿态。

[0035] 具体地,源图像表示姿态变换前的图像,目标三维姿态将目标姿态采用三维表示。目标三维姿态用于描述与源图像匹配的对象变换后的目标姿态。目标三维姿态可以为一个

或多个，表示需要将源图像中的对象变换为一个或多个姿态。源图像中的对象可以为人体、植物、动物等存在姿态的人或物体。源图像也可以为一个或多个，表示需要将一个或多个源图像中的对象转换姿态。一个源图像中可以包括一个或多个待转换姿态的对象。

[0036] 在一些实施例中，目标三维姿态包括三维人体关节点的位置，使用热力图表示，可以通过三维姿态估计技术得到。

[0037] 在一些实施例中，目标三维姿态包括多个，生成目标三维姿态对应的目标姿态图像，将各个目标姿态图像组合生成目标姿态视频。如可以将同一个人像生成包括多个不同姿态变换的视频。

[0038] 步骤S204，结合源图像和目标三维姿态，基于语义分割重建得到三维分割体素，三维分割体素包括体素的类别信息。

[0039] 具体地，语义分割是计算机视觉中的基本任务，在语义分割中需要将视觉输入分为不同的语义可解释类别，即分类类别在真实世界中是有意义的。语义分割可以使用卷积神经网络为每个体素分配类别标签，卷积层可以有效地捕捉图像或目标三维姿态中的局部特征，并以层级的方式将许多这样的模块嵌套在一起，通过一系列卷积捕捉图像或目标三维姿态的复杂特征。可通过卷积神经网络基于语义分割提取源图像和目标三维姿态的特征，从而重建得到包含了体素的类别信息的三维分割体素，通过标注分割均匀的网格拓扑得到各个体素的类别信息。三维分割体素是在三维空间中标注出每个体素的类别，其中包含了源图像的对象形状和目标姿态信息，不仅能表示姿态变换后的人体三维坐标，并且能识别出各个对象的部件类别。对象的部件类别根据对象的特征进行划分，如当对象为人体时，部件类别可以包括脸，左手，右手，左腿，右腿，躯干，背景。

[0040] 在一个实施例中，通过已训练的三维分割体素，将源图像和目标三维姿态输入三维分割体素得到输出的三维分割体素。其中，三维体素网络的网络结构可自义，可以是深度神经网络。

[0041] 步骤S206，将三维分割体素投影得到对应的目标姿态二维分割图，基于体素的类别信息对所述目标姿态二维分割图中的对象标注得到对应的部件类别。

[0042] 具体地，将三维分割体素投影到图像的水平平面，即xy平面上，得到二维分割图，由于三维分割体素是基于目标三维姿态得到的，与目标姿态对应，所以得到的是对应的目标姿态二维分割图。将相同类别信息的体素对应的区域合并得到二维分割图上对象的各个部件类别对应的区域，从而对每个区域标注不同的部件类别，基于该二维分割图，可以将人体姿态变换问题视为图像到图像的转换问题。

[0043] 步骤S208，获取目标三维姿态对应的目标二维姿态，提取源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像。

[0044] 其中，目标二维姿态将目标姿态采用二维表示，目标二维姿态用于描述与源图像匹配的对象变换后的目标姿态，可以用二维的热力图表示。当目标三维姿态有多个时，分别对应相同数量的目标二维姿态，目标二维姿态与目标三维姿态表示的是同一个姿态，可以通过二维姿态估计算法生成。中间尺度的变换图像是指分辨率小于目标图像分辨率，但姿态为转换后姿态的粗略图像，通过从粗到细的框架，先生成中间尺度的变换图像，再结合后续的部件图件让最终结果达到目标图像的高清分辨率。

[0045] 具体地，合成中间尺度的变换图像时引入了目标二维姿态来增强目标的空间注意

力。提取源图像、目标姿态二维分割图和目标二维姿态的特征，可以通过深度神经网络学习输入的特征，将源图像的纹理转变为与目标姿态二维分割图和目标二维姿态匹配的目标姿态纹理，从而合成中间尺度的变换图像。

[0046] 在一个实施例中，通过已训练的中间尺度生成网络，将源图像、目标姿态二维分割图和目标二维姿态输入中间尺度生成网络得到输出的中间尺度的变换图像。其中，中间尺度生成网络的网络结构可自义，可以是深度神经网络。

[0047] 步骤S210，分别对源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据，对象部件是根据部件类别确定的。分别对各个对象部件的部件层数据进行部件合成，生成各个对象部件对应的部件图像。

[0048] 其中，部件层数据是输入的各个图像中对应对象部件的数据，对象部件是基于语义分割得到的部件。根据部件类别确定对象部件，可以从部件类别中过滤与对象本身无关的类别，得到对象部件，如过滤掉背景类别。如部件类别包括的类别有脸，左手，右手，左腿，右腿，躯干，背景，过滤掉背景这个与对象本体无关的部件类别，得到对象部件包括脸，左手，右手，左腿，右腿，躯干。可以理解，由于各个对象部件占整个对象的比例不同，裁切得到的各个部件层数据的图片分辨率也可不同。在一个实施例中，依据每个对象部件在整个对象中的相对长度，设定部件层数据的图像分辨率。

[0049] 具体地，分别对各个对象部件的部件层数据进行部件合成，合成一个无视遮挡的完整目标部件图像。部件图像可以是目标部件和部件变换图像之间的残差，从而在下一步中可通过残差与变换图像进行融合得到目标姿态图像。合成各个对象部件对应的完整的对象部件，避免了目标姿态下的遮挡现象。由于生成了完整的部件，当遮挡发生时，之后的全局融合步骤可以得到更加正确真实的结果图像。

[0050] 在一个实施例中，通过已训练的与各个对象部件对应的层级生成网络，将部件层数据输入与对象部件匹配的层级生成网络，得到输出的部件图像。其中，与各个对象部件对应的层级生成网络的网络结构可自义，各个对象部件的层级生成网络独立训练，可以是深度神经网络。

[0051] 步骤S212，将变换图像和部件图像融合得到目标姿态图像。

[0052] 具体地，融合的方式可自定义，可先将变换图像放大变换至与目标姿态图像匹配的分辨率，再将各个部件图像分别叠加至与放大后的变换图像中的匹配的对象部件区域，得到目标姿态图像。因为每个部件图像是独立生成的，所以可能存在它们之间全局一致性无法得到保障的问题，在一些实施例中，对目标姿态图像进一步进行后处理得到精准的目标姿态图像。如将各个对象部件区域的分界处进行滤波处理，或通过精细尺度生成网络得到修复图像，将修复图像与目标姿态图像融合得到精准的目标姿态图像。

[0053] 上述姿态变换数据处理方法，提出了一个新的深度学习框架，采用带分割的三维体素表示来消除二维表示所带来的歧义问题，包含了源图像的人体形状和目标姿态信息，不仅能表示姿态变换后的对象三维坐标，还能识别出各个部件类别，从而使得后续高质量的局部部件合成得以实现，分别独立合成完整的各个对象部件，从而解决各部件间的遮挡问题并保证局部合成结果的高质量，通过生成变换图像，从粗到细的框架让最终结果达到高清分辨率，融合得到的目标姿态图像具有非常高的鲁棒性和准确度。

[0054] 在一些实施例中，步骤S204包括：将源图像和目标三维姿态输入三维体素网络，三

维体素网络对源图像和目标三维姿态进行编码,得到编码结果,对编码结果提取特征并解码输出携带体素的类别信息的三维分割体素。

[0055] 具体地,三维体素网络可以是深度神经网络,包括卷积层,池化层、连接层等网络结构,通过相互连接的各个网络层对源图像和目标三维姿态进行编码,得到编码结果,对编码结果提取特征并解码输出携带体素的类别信息的三维分割体素。三维体素网络的输入是源图像和目标三维姿态,输出是携带体素的类别信息的三维分割体素。在进行模型训练时,可以采用有监督的训练方法训练三维体素网络,在一个实施例中,训练数据来自于数据集中标注好的三维人体模型,训练数据包括源训练图像、目标训练三维姿态,标注好的三维人体体素。在一个实施例中,将源训练图像、目标训练三维姿态输入三维体素网络,经过三维体素网络所包括的各层依次处理后,输出对应的训练三维分割体素。根据训练三维分割体素和标注好的三维体素的差异反向传播调整三维体素网络的网络参数,以得到已训练的中间尺度生成网络。

[0056] 在一些实施例中,使用交叉熵损失函数来描述训练三维分割体素和标注好的三维体素的差异以反向传播调整三维体素网络的网络参数,从而得到已训练的三维体素网络。其中交叉熵损失函数的定义如下:

$$Lvol = -\sum_{i}^{H}\sum_{j}^{W}\sum_{k}^{D}\sum_{c}^{N}\hat{V}(i,j,k,c)\log(\,fs(V((i,j,k,c))$$

$Lvol$:是交叉熵损失,H,W,D分别是高度,宽度,深度,N是部件类别数,i,j,k,c分别表示对应的变量,$\hat{V}$是三维体素的正确标注,V是网络输出的三维体素的标注,f_S是softmax函数。

[0057] 本实施例中,通过已训练的三维体素网络直接得到携带体素的类别信息的三维分割体素,使得三维分割体素的重建通过三维体素网络来完成,高效准确。

[0058] 在一些实施例中,三维体素网络对源图像和目标三维姿态进行编码,得到编码结果,对编码结果提取特征并解码输出携带体素的类别信息的三维分割体素包括:对源图像进行编码并提取特征得到第一特征,对目标三维姿态进行编码并提取特征得到第二特征,将第一特征和第二特征合并得到合并特征,将合并特征输入堆叠沙漏网络,通过堆叠沙漏网络解码得到三维分割体素。

[0059] 具体地,经过三维体素网络中数个卷积层和池化层分别对源图像和目标三维姿态进行编码并提取特征,编码和提取特征的方式可以一致或不同。将第一特征和第二特征组合生成一个矩阵输入堆叠沙漏网络。在一个实施例中,堆叠沙漏网络包括两个,第一个沙漏网络生成初始特征,第两个沙漏网络让初始特征更精细,从而解码得到三维分割体素。

[0060] 本申请实施例中,通过采用堆叠沙漏网络作为三维体素网络的架构,分别对源图像和目标三维姿态进行编码并提取特征,然后合并提取的特征并输入到堆叠沙漏网络中并解码出三维分割体素,可以提高三维分割体素生成的准确度。

[0061] 在一些实施例中,步骤S208包括:将源图像、目标姿态二维分割图和目标二维姿态合并形成输入矩阵,将输入矩阵输入至中间尺度生成网络,中间尺度生成网络依次通过下采样层、残差块层和上采样层对输入矩阵进行特征提取得到中间尺度的变换图像。

[0062] 具体地,中间尺度生成网络用于依次通过下采样层、残差块层和上采样层对输入矩阵进行特征提取,从而输出低分辨率的粗糙结果图像。下采样层用于降低特征尺寸,使网

络聚焦于全局语义信息，形成对输入图像的整体的理解。残差块层用于使得计算结果为残差，减少计算量。上采样层是结合下采样各层的信息和输入信息来还原细节信息并逐步还原图像精度。在一个实施例中，通过有监督的训练方法来训练中间尺度生成网络，其中损失函数可根据需要自定义，在一个实施例中通过感知损失、对抗损失、特征匹配损失中的至少一种损失反向传播调整中间尺度生成网络的网络参数，以得到已训练的中间尺度生成网络。其中感知损失和特征匹配损失是为了使输出图像和标签图像更接近，对抗损失的作用是为了保证结果图像保持与源图像的对象特征一致性。

[0063] 本实施例中，通过已训练的中间尺度生成网络直接得到中间尺度的变换图像，使得变换图像的合成通过中间尺度生成网络来完成，高效准确。

[0064] 在一些实施例中，中间尺度生成网络的训练包括以下步骤：获取第一训练样本，第一训练样本包括源训练图像、目标姿态训练二维分割图、目标训练二维姿态和对应的标签变换图像。将源训练图像、目标姿态训练二维分割图、目标训练二维姿态输入中间尺度生成网络，经过中间尺度生成网络所包括的各层依次处理后，输出对应的训练变换图像；根据训练变换图像和标签变换图像的差异反向传播调整中间尺度生成网络的网络参数，以得到已训练的中间尺度生成网络。

[0065] 其中，第一训练样本包括源训练图像、目标姿态训练二维分割图、目标训练二维姿态和对应的标签变换图像，训练变换图像就是通过中间尺度生成网络对输入源训练图像、目标姿态训练二维分割图、目标训练二维姿态提取特征进行预测得到的图像。标签变换图像是真实的中间尺度的图像。

[0066] 具体地，在获取到第一训练样本后，将源训练图像、目标姿态训练二维分割图、目标训练二维姿态输入中间尺度生成网络，经过中间尺度生成网络所包括的各层依次处理后，输出训练变换图像，根据训练变换图像与标签变换图像的差异构建损失函数，再按照最小化该损失函数的方向反向传播，调整中间尺度生成网络的网络参数并继续训练，直至满足训练结束条件或所有的训练样本训练完毕。

[0067] 在一个实施例中，如图3所示，根据训练变换图像和所述标签变换图像的差异反向传播调整中间尺度生成网络的网络参数，以得到已训练的中间尺度生成网络包括：

步骤S302，通过预训练感知网络分别提取标签变换图像和训练变换图像的特征得到特征图，计算特征图之间的距离得到感知损失。

[0068] 具体地，预训练感知网络是指已训练的可以感知图像质量的网络，如可以是VGG-19预训练网络，通过感知损失使得变换图像结果能更接近正确标注图像。在一个实施例中，感知损失函数的表达式如下：

$$\tilde{L}_{per}=\sum_{i=1}^{L}\| \Phi_i(\hat{I}_t)-\Phi_i(\tilde{I}_t) \|_i$$

其中Φ_l指的是预训练网络的第l层特征，$\hat{I}_t$指的是标签变换图像，是真实的目标图像，$\tilde{I}_t$指的是合成的训练变换图像，L表示计算感知损失使用预训练网络的总层数。

[0069] 步骤S304，根据标签变换图像、源训练图像、训练变换图像对中间尺度生成网络和判别网络进行对抗学习，得到对抗损失。

[0070] 具体地，通过对抗损失使得变换图像结果能更接近真实图像。对抗学习是通过让

两个机器学习模型相互博弈的方式进行学习，得到期望的机器学习模型。将中间尺度生成网络与判别网络进行对抗学习，中间尺度生成网络的目标是根据输入得到所期望的输出。判别网络的目标是将生成网络的输出从真实的图像中尽可能分辨出来。判别网络的输入包括中间尺度生成网络的输出以及真实的图像。两个网络相互对抗学习、不断调整参数，最终目的是中间尺度生成网络要尽可能地欺骗判别网络，使判别网络无法判断生成网络的输出结果是否真实。

[0071] 其中，判别网络模型中参数的调整方向是朝着使判别网络模型的损失值变小的方向进行调整，使得判别网络模型的判别能力变强，而中间尺度生成网络中参数的调整方向是朝着判别网络模型的损失值变大的方向进行调整，使得判别网络模型不容易将中间尺度生成网络的输出从真实的图像中尽可能分辨出来。在进行对抗学习时，可以对模型参数进行多次调整。

[0072] 在一个实施例中，对抗损失函数的表达式如下：

$$\tilde{L}adv = \log(D(\hat{I}_t, I_s)) + \log(1 - D(\tilde{I}_t, I_s))$$

其中D_i是判别网络第i层特征，$\hat{I}_t$指的是标签变换图像，$\tilde{I}_t$指的是合成的训练变换图像，I_s指的是源训练图像。

[0073] 步骤S306，通过判别网络中多个不同尺度的卷积层分别计算标签变换图像和训练变换图像在多个不同尺度的特征距离，统计多个不同尺度的特征距离得到特征匹配损失。

[0074] 具体地，特征匹配损失用于使得变换图像结果能更接近正确标注图像，通过多个不同尺度的标签变换图像和训练变换图像之间的特征距离，使得提取的特征更能准确的合成高质量的变换图像。

[0075] 在一个实施例中，特征匹配损失函数的表达式如下：

$$\tilde{L}_{FM} = \sum_{i=1}^{T} \left\| D_i(\hat{I}_t, I_s) - D_i(\tilde{I}_t, I_s) \right\|_1$$

D_i是判别网络的第i层特征，T是总层数，$\hat{I}_t$指的是标签变换图像，$\tilde{I}_t$指的是合成的训练变换图像，I_s指的是源训练图像。

[0076] 步骤S308，根据感知损失、对抗损失、特征匹配损失确定目标损失，根据目标损失反向传播调整中间尺度生成网络的网络参数，以得到已训练的中间尺度生成网络。

[0077] 具体地，可根据公式$\tilde{L} = \tilde{L}_{adv} + \lambda_{FM}\tilde{L}_{FM} + \lambda_{per}\tilde{L}_{per}$计算得到目标损失，其中$\tilde{L}_{adv}$表示对抗损失，$\tilde{L}_{FM}$表示特征匹配损失，$\tilde{L}_{per}$表示感知损失，$\lambda_{FM}$表示特征匹配损失的权重，$\lambda_{per}$表示感知损失的权重。最小化该目标损失的方向反向传播，调整中间尺度生成网络的网络参数并继续训练，直至满足训练结束条件或所有的训练样本训练完毕。在一个实施例中，λ_{FM}与λ_{per}分别为1。

[0078] 本申请实施例中，通过将感知损失、对抗损失、特征匹配损失加权得到目标损失，将不同类型的损失结合起来训练中间尺度生成网络，提高了变换图像结果的质量。

[0079] 在一些实施例中，步骤S210包括：获取源图像对应的源二维分割图，根据源二维分割图对源图像进行分割得到对应于各个对象部件的部件层数据，获取各个对象部件对应

的裁切信息，基于各个对象部件的中心位置，对三维分割体素、目标二维姿态、变换图像分别裁切得到与对应的裁切信息匹配的部件层数据。

[0080] 具体地，源二维分割图用于描述源图像的对象形状和初始姿态信息，不仅能表示初始姿态的人体二维坐标，并且包括对对象标注的不同的部件类别。为了建立源图像和目标图像间的对象层对应关系，源二维分割图包括与三维分割体素一致的部件类别。

[0081] 在一些实施例中，通过部件分割网络得到源图像对应的源二维分割图，部件分割网络用于确定输入图像对应的二维分割图，为了平衡效率和性能，可以使用U-Net架构作为部件分割网络。部件分割网络可以通过有监督的训练方法训练得到，训练样本数据包括：源训练图像和标签二维分割图，其中标签二维分割图可以通过将源图像对应的三维分割体素投影到xy平面，得到二维对象部件分割的正确标注，可以使用交叉熵损失函数来训练部件分割网络。源图像对应于各个对象部件的部件层数据可以通过源图像和源二维分割图逐像素相乘得到。

[0082] 在一些实施例中，使用交叉熵损失函数来描述训练二维分割图和标注好的标签二维分割图的差异以反向传播调整部件分割网络的网络参数，从而得到已训练的部件分割网络。其中交叉熵损失函数的定义如下：

$$\mathrm{L}par = -\sum_{i}^{H}\sum_{j}^{W}\sum_{c}^{N}\hat{S}_s(i,j,c)\log(\,fs(S_s((i,j,c))$$

$\mathrm{L}par$ 是交叉熵损失，H，W分别是高度，宽度，N是部件类别数，i，j，c分别表示对应的变量，$\hat{S}_s$是标签二维分割图，S_s是网络输出的训练二维分割图，$\mathrm{f_S}$是softmax函数。

[0083] 通过$I_s^m = I_s \otimes S_s^m$得到源图像对应于各个对象部件的部件层数据，$I_s^m$表示源图像对应部件类别为m的部件层数据，其中$\mathrm{I_S}$指的是源图像，m表示部件类别，$S_s^m$表示源二维分割图中部件类别为m的区域。

[0084] 裁切信息是指与目标部件层对应的图像分辨率信息、裁切中心位置等。其中裁切中心位置分别在各个待裁切图像对应的对象部件区域的中心。根据裁切信息得到三维分割体素、目标二维姿态、变换图像分别对应的裁切中心位置，以裁切中心位置为中心根据匹配的图像分辨率信息，进行裁切得到对应于各个对象部件的部件层数据。

[0085] 本实施例中，通过源二维分割图对源图像进行高效的分割，通过部件类别分别对其他图像进行裁切，准确快速地得到各个部件层数据。

[0086] 在一些实施例中，步骤S210包括：获取各个对象部件对应的层级生成网络；将部件层数据输入匹配的对象部件的层级生成网络；各个对象部件的层级生成网络分别输出与对象部件匹配的部件图像。

[0087] 其中，层级生成网络用于将被裁切的源图像、三维分割体素、目标二维姿态、变换图像中对应于对象部件的部件层数据合并输入，经过层级生成网络中的各个网络层处理，合成一个无视遮挡的完整的部件图像。层级生成网络的输出可以是目标部件和变换图像部件间的残差，从而在下一步中可通过层级生成网络输出的残差与变换图像进行融合得到目标姿态图像。层级生成网络输出的是合成完整的对象部件，避免了目标姿态下的遮挡现象。由于生成了完整的部件，当遮挡发生时，之后的全局融合步骤可以得到更加正确真实的结果图像。

[0088] 各个对象部件的层级生成网络独立训练得到，使得各个对象部件图像可以独立的合成从而无遮挡。

[0089] 在一些实施例中，对于存在对称特征的对象部件共用一个层级生成网络，如左右手臂共用一个手臂层级生成网络，左右腿共用一个腿部层级生成网络。可以理解，当存在对称特征的对象部件共用一个层级生成网络时，共用的层级生成网络的输入包括两个对称特征的对象部件对应的图像。如利用人体对称性，手臂层级生成网络和腿部层级生成网络的输入还包括另一侧的部件图像，这一额外的部件图像当原部件被遮挡时，可以提供更多的外观信息。

[0090] 在一个实施例中，通过有监督的训练方法来训练层级生成网络，其中损失函数可根据需要自定义，在一个实施例中通过感知损失和对抗损失中的至少一种损失反向传播调整层级生成网络的网络参数，以得到已训练的中间尺度生成网络。其中感知损失是为了使输出图像和标签图像更接近，对抗损失用于评判某一对象部件是否完整。感知损失可以只度量可见区域，从而消除被遮挡的未知区域的影响。

[0091] 本实施例中，通过已训练的层级生成网络直接得到与对象部件匹配的部件图像，使得各个部件图像独立无遮挡来完成，高效准确。层级生成网络用于更精确地处理重要的对象部件。既要保留源姿势中的纹理细节，又要合成目标姿势中的缺失区域。

[0092] 在一些实施例中，层级生成网络的训练包括以下步骤：获取第二训练样本；所述第二训练样本包括源训练部件图像、训练部件三维分割体素、目标训练部件二维姿态、部件变换图像和对应的标签部件图像，第二训练样本中的各个样本对应当前对象部件。将源训练部件图像、训练部件三维分割体素、目标训练部件二维姿态、部件变换图像输入当前对象部件对应的层级生成网络，经过层级生成网络所包括的各层依次处理后，输出对应的训练部件图像；根据训练部件图像和所述标签部件图像的差异反向传播调整层级生成网络的网络参数，以得到已训练的当前对象部件对应的层级生成网络。

[0093] 其中，第二训练样本包括源训练部件图像、训练部件三维分割体素、目标训练部件二维姿态、部件变换图像和对应的标签部件图像，这些都对应同一个对象部件。训练部件图像就是通过层级生成网络对输入数据提取特征进行预测得到的图像。标签部件图像是与当前对象部件对应的无遮挡的真实图像。

[0094] 具体地，在获取到第二训练样本后，将源训练部件图像、训练部件三维分割体素、目标训练部件二维姿态、部件变换图像输入层级生成网络，经过层级生成网络所包括的各层依次处理后，输出训练部件图像，根据训练部件图像与标签部件图像的差异构建损失函数，再按照最小化该损失函数的方向反向传播，调整层级生成网络的网络参数并继续训练，直至满足训练结束条件或所有的训练样本训练完毕。

[0095] 在一些实施例中，如图4所示，根据训练部件图像和所述标签部件图像的差异反向传播调整所述层级生成网络的网络参数，以得到已训练的当前对象部件对应的层级生成网络包括：

步骤S402，根据标签部件图像、源训练部件图像、训练部件图像对当前对象部件对应的层级生成网络和判别网络进行对抗学习，得到部件对抗损失，标签部件图像是训练集中未被遮挡的与当前对象部件对应的图像。

[0096] 具体地，通过部件对抗损失来评判对象部件是否完整，在一个实施例中，部件对抗

损失的函数表达式如下：

$$\mathrm{L}^{m}adv = \log(D(I_r^m, I_s^m)) + \log(1 - D(I_t^m, I_s^m))$$

其中，m指的是当前对象部件，I_s^m指的是源图像中的当前对象部件层，I_t^m指的是训练层级生成网络时输出的当前对象部件对应的训练部件图像，I_r^m指的是训练集中随机筛选的未被遮挡的当前对象部件层。

[0097] 步骤S404，通过预训练感知网络分别提取标签部件图像和训练部件图像的特征得到特征图，计算特征图之间的距离得到部件感知损失。

[0098] 具体地，通过部件感知损失使得合成结果能更接近正确标注图像，部件感知损失只度量可见区域，从而消除被遮挡的未知区域的影响。

[0099] 在一个实施例中，部件感知损失通过以下公式计算得到：

$$L_{per}^m = L_{per} \otimes S_t^m$$

其中，L_{per}表示整个图像的感知损失，m表示部件类别，S_t^m表示目标姿态二维分割图中部件类别为m的区域。L_{per}^m表示部件类别为m的部件感知损失。

[0100] 步骤S406，根据部件对抗损失、部件感知损失确定目标部件损失，根据目标部件损失反向传播调整当前对象部件对应的层级生成网络的网络参数，以得到已训练的层级生成网络。

[0101] 具体地，可根据公式$\mathrm{L}^m = \mathrm{L}_{adv}^m + \lambda_{per}^m L_{per}^m$计算得到目标部件损失，其中$\mathrm{L}_{adv}^m$表示部件对抗损失，$L_{per}^m$表示部件感知损失，$\lambda_{per}^m$表示部件感知损失的权重。最小化该目标损失的方向反向传播，调整层级生成网络的网络参数并继续训练，直至满足训练结束条件或所有的训练样本训练完毕。在一个实施例中，λ_{per}^m为1。

[0102] 本申请实施例中，通过将部件对抗损失、部件感知损失加权得到目标损失，将不同类型的损失结合起来训练层级生成网络，提高了部件图像结果的质量。

[0103] 在一些实施例中，步骤S212包括：将部件图像根据二维分割图融合到变换图像的对应对象部件区域得到初始全局姿态图像；将源图像、目标姿态二维分割图、目标二维姿态和初始全局姿态图像合并输入至精细尺度生成网络，输出得到目标姿态残差图像；将目标姿态残差图像与初始全局姿态图像融合得到目标姿态图像。

[0104] 具体地，初始全局姿态图像与目标姿态图像具有相同的分辨率，先将变换图像变换至目标姿态图像对应的分辨率大小，再将部件图像根据部件所处的位置分别叠加至变换后的变换图像的对应区域得到初始全局姿态图像。在一个实施例中，通过以下公式得到初始全局姿态图像：

$$\dot{\mathrm{I}}_t = \tilde{\mathrm{I}}_t + \sum_{m}^{M} \tilde{\mathrm{I}}_t^m \otimes S_t^m$$

其中，$\tilde{\mathrm{I}}_t$表示变换图像，S_t^m表示目标姿态二维分割图中部件类别为m的区域，$\dot{\mathrm{I}}_t$表示初始全局姿态图像，$\tilde{\mathrm{I}}_t^m$表示部件m的部件图像，M表示对象部件的总数目。

[0105] 精细尺度生成网络用于改善初始结果，并确保合成高质量全局一致的姿态变换结果。精细尺度生成网络的输入是源图像、目标姿态二维分割图、目标二维姿态和初始全局姿

态图像的合并，输出是目标姿态图像和初始全局姿态图像间的残差。在一个实施例中，通过以下公式得到目标姿态图像：

$$I_t = \hat{I}_t + \dot{I}_t$$

其中，I_t表示目标姿态图像，$\hat{I}_t$表示精细尺度生成网络的输出，$\dot{I}_t$表示初始全局姿态图像。

[0106] 在一个实施例中，通过有监督的训练方法训练精细尺度生成网络，精细尺度生成网络训练的目标函数，目标损失可以是对抗损失、感知损失和特征匹配损失的加权总和。在一个实施例中，可根据公式$L = L_{adv} + \lambda_{FM} L_{FM} + \lambda_{per} L_{per}$计算得到精细尺度生成网络的目标损失，其中$L_{adv}$表示对抗损失，$L_{FM}$表示特征匹配损失，$L_{per}$表示感知损失，$\lambda_{FM}$表示特征匹配损失的权重，$\lambda_{per}$表示感知损失的权重。最小化该目标损失的方向反向传播，调整精细尺度生成网络的网络参数并继续训练，直至满足训练结束条件或所有的训练样本训练完毕。在一个实施例中，λ_{FM}与λ_{per}分别为1。

[0107] 本申请实施例中，通过从粗到细的框架，通过精细尺度生成网络维持精细的局部纹理细节，合成更高分辨率的结果图像。

[0108] 以下以应用于人体姿态变换，如图5所示，展示了本申请实施例提出的层级端到端人体姿态变换网络，对本申请实施例提供的姿态变换数据处理方法进行说明，包括以下步骤：

1、获取源图像和目标三维姿态，源图像是1024*1024的包括人体的图像，目标三维姿态是256*256*64的三维姿态。

[0109] 2、在输入三维体素网络时，可以将源图像进行下采样得到256*256的图像，然后将下采样后的源图像和目标三维姿态输入至三维体素网络，可以理解的是，也可以将原分辨率的源图像输入三维体素网络，由三维体素网络进行下采样得到256*256的图像。三维体素网络输出256*256*64的带分割的三维体素，包括体素的类别信息，分为7个部件类别，分别为脸，左手，右手，左腿，右腿，躯干，背景。

[0110] 3、将三维体素投影到图像的xy平面上得到目标姿态二维分割图，目标姿态二维分割图标注了目标图像的人体部件类别，其中人体部件类别包括脸，左手，右手，左腿，右腿，躯干。

[0111] 具体地，基于该二维分割图，可以将人体姿态变换问题视为图像到图像转换问题。

[0112] 4、获取目标三维姿态对应的目标二维姿态，为256*256的二维图像，将256*256的源图像、256*256的目标二维姿态、256*256的二维分割图输入至中间尺度生成网络，输出对应于目标三维姿态的512*512分辨率的变换图像。

[0113] 具体地，引入了目标二维姿态来增强目标的空间注意力。

[0114] 5、确定源图像、三维分割体素、目标二维姿态、变换图像对应于各个对象部件的部件层数据。

[0115] 具体地，一些身体部件，例如人脸，在变换过程中可视度改变很大，并且包含重要信息纹理。为了更好地解决这一问题，额外地利用人体语义表示来指示身体各个语义部件的独立合成。通过三个基于部件的层级生成网络用于更加精确地合成重要的人体组件，包括脸，手臂，腿。这些层级生成网络可以合成高质量的结果，不仅保持了源图像中的纹理细

节,同时生成了目标姿态下被遮挡的区域。由于使用了三维分割体素表示,尽管这些重要部件的可见性会在目标姿态下完全改变,依然可以正确解决遮挡问题,因为对应的三维体素是完整的。

[0116] 为了建立源图像和目标图像间的身体层对应关系,和三维体素表示相同,我们将源图像分割为7个部件类别。通过部件分割网络得到源图像对应的二维部件分割图,源图像的各个部件层可以通过源图像和二维部件分割图逐像素相乘得到。

[0117] 根据中心的位置,这些身体部件在对应的三维分割体素、目标二维姿态、变换图像中被裁切,由于最终结果的分辨率是1024*1024,依据每个部件层在整个身体中的相对长度,人脸、手臂、腿部的分辨率被分别设定为128*128、256*256、512*512。

[0118] 6、将各个对象部件的部件层数据分别输入至与对象部件匹配的层级生成网络,得到各个对象部件对应的部件图像。

[0119] 将被裁切的结果和源图像的部件层合并输入到对应的层级生成网络中,合成一个无视遮挡的完整目标部件图像。由于人身体的对称性,让左右的同类部件共用同一个生成器,所以总共有三个层级生成网络来合成一张人脸、两条手臂和两条腿。并且,为了更好地利用人体对称性,手臂生成器和腿部生成器的输入还包括另一侧的初始部件图像,当原部件被遮挡时这一额外的部件图像可以提供更多的外观信息,层级生成网络的输出结果是目标部件和变换图像部件间的残差。如图6所示,展示了人脸层级表示的详细结构,将源图像中的人脸部分、三维分割体素中的人脸部分、目标二维姿态中的人脸部分、变换图像中的人脸部分输入人脸层级生成网络,可以得到128*128的人脸部件图像。如图7从左到右依次展示了真实目标图像、真实目标部件图像、没有层级表示的合成结果和本方案层级生成网络的合成结果,可见,本方案的结果质量更好,是完整的部件图像。

[0120] 7、将变换图像和部件图像融合得到目标姿态图像。

[0121] 具体地,将512*512的变换图像上采样至1024*1024的分辨率,然后将128*128的人脸部件图像、256*256的左、右手臂部件图像、512*512的左右腿部部件图像根据二维分割图加到上采样后的变换图像的对应位置上,得到初始全局姿态图像。

[0122] 将源图像、目标姿态二维分割图、目标二维姿态和初始全局姿态图像合并输入至精细尺度生成网络,输出得到目标姿态残差图像。可以理解的是,输入前可调整各个图像的大小,使各个输入图像分辨率保持一致。

[0123] 本申请实施例提供的姿态变换数据处理方法,经过实验,在两个人体数据集上都取得了很好的人体姿态变换结果,定量和定性的比较结果都超越了当前最优的三个技术方案,归功于提出的层级姿态变换合成框架,三维分割体素表示和基于部件的层级表示,可以更好地处理遮挡问题,维持精细的局部纹理细节,合成更高分辨率的结果图像。同时,具备很好的鲁棒性和准确度,也可以应用于视频动作序列的合成任务,如图8所示,本申请方案在视频动作序列上的部分结果。

[0124] 如图9所示,是本申请方案在Human3.6M数据集里的部分人体姿态变换结果,及与其他三个技术方案(DSC,LW-GAN,Pix2pixHD)的比较。

[0125] 如图10所示,是本申请技术方案在自己构建的运动视频数据集里的部分人体姿态变换结果,及与其他三个技术方案(DSC,LW-GAN,Pix2pixHD)的比较。

[0126] 把本申请生成的人体姿态变换结果和其他三种当前最好方法进行了比较。三种方

法包括Pix2pixHD, DSC, LW-GAN, 它们在数据集上被重新训练。使用结构相似性(SSIM)和图像块感知相似性(LPIPS)这两个量化指标来评估各个方法合成图像的优劣。比较结果如图11是在两个数据集中的评测结果(箭头表示数值更优的方向),可以看到本申请方法明显优于其他三种方法。

[0127] 为了评测本申请中各个部分的重要性,构建了基础模型并将其余部分依次逐个添加,从而进行剥离实验。基础模型Baseline的输入是源图像和目标二维姿态,包含中间尺度生成网络和精细尺度生成网络。在此基础上添加三维体素网络,从而将投影得到的目标姿态二维分割图作为生成网络的输入,这一模型被标记为Baseline+V。再添加部件生成网络,但是只使用传统的对抗损失函数训练部件生成网络,从而可能合成不完整的目标部件图像,这一模型被标记为Baseline+V+PL-。因此,本申请完整方法的模型被标记为Baseline+V+PL。如图12展示了各个模型生成图像的评测结果,可以看到LPIPS指标随着各个部件的添加逐渐变好。

[0128] 本申请实施例提供的姿态变换数据处理方法场景上可以用于视频制作,也可以用于美颜相机等,可以联合部署于智能摄像头、后台服务器的CPU(central processing unit,中央处理器)或GPU(Graphics Processing Unit,图形处理器)上,也可以用于手机等移动设备的CPU上。

[0129] 如图13所示,在一些实施例中,提供了一种姿态变换数据处理装置,该姿态变换数据处理装置可以集成于上述的服务器120或者终端110中,具体可以包括:

获取模块502,用于获取源图像和目标三维姿态。

[0130] 三维分割体素模块504,用于结合源图像和目标三维姿态,基于语义分割重建得到三维分割体素,所述三维分割体素包括体素的类别信息。

[0131] 投影模块506,用于将三维分割体素投影得到对应的目标姿态二维分割图,基于所述体素的类别信息对目标姿态二维分割图中的对象标注得到对应的部件类别。

[0132] 中间尺度模块508,用于获取目标三维姿态对应的目标二维姿态,提取所述源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像。

[0133] 部件图像生成模块510,用于分别对源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据,对象部件是根据部件类别确定的,分别对各个对象部件的部件层数据进行部件合成,生成各个对象部件对应的部件图像。

[0134] 融合模块512,用于将变换图像和部件图像融合得到目标姿态图像。

[0135] 在一些实施例中,三维分割体素模块504还用于将源图像和目标三维姿态输入三维体素网络,三维体素网络对源图像和目标三维姿态进行编码,得到编码结果,对编码结果提取特征并解码输出携带体素的类别信息的三维分割体素。

[0136] 在一些实施例中,三维分割体素模块504还用于对源图像进行编码并提取特征得到第一特征;对目标三维姿态进行编码并提取特征得到第二特征;将第一特征和第二特征合并得到合并特征,将合并特征输入堆叠沙漏网络,通过堆叠沙漏网络解码得到三维分割体素。

[0137] 在一些实施例中,中间尺度模块508还用于将源图像、目标姿态二维分割图和目标二维姿态合并形成输入矩阵;将输入矩阵输入至中间尺度生成网络;中间尺度生成网络依次通过下采样层、残差块层和上采样层对输入矩阵进行特征提取得到中间尺度的变换图

像。

[0138] 在一些实施例中，装置还包括：

中间尺度生成网络训练模块，用于获取第一训练样本；第一训练样本包括源训练图像、目标姿态训练二维分割图、目标训练二维姿态和对应的标签变换图像；将源训练图像、目标姿态训练二维分割图、目标训练二维姿态输入中间尺度生成网络，经过中间尺度生成网络所包括的各层依次处理后，输出对应的训练变换图像；根据训练变换图像和所述标签变换图像的差异反向传播调整中间尺度生成网络的网络参数，以得到已训练的中间尺度生成网络。

[0139] 在一些实施例中，中间尺度生成网络训练模块，还用于通过预训练感知网络分别提取标签变换图像和训练变换图像的特征得到特征图，计算特征图之间的距离得到感知损失；根据标签变换图像、源训练图像、训练变换图像对中间尺度生成网络和判别网络进行对抗学习，得到对抗损失；通过判别网络中多个不同尺度的卷积层分别计算标签变换图像和训练变换图像在多个不同尺度的特征距离，统计多个不同尺度的特征距离得到特征匹配损失；根据感知损失、对抗损失、特征匹配损失确定目标损失，根据目标损失反向传播调整中间尺度生成网络的网络参数，以得到已训练的中间尺度生成网络。

[0140] 在一些实施例中，部件图像生成模块510还用于获取源图像对应的源二维分割图，根据源二维分割图对所述源图像进行分割得到对应于各个对象部件的部件层数据；获取各个对象部件对应的裁切信息，基于各个对象部件的中心位置，对三维分割体素、目标二维姿态、变换图像分别裁切得到与对应的裁切信息匹配的部件层数据。

[0141] 在一些实施例中，部件图像生成模块510还用于获取各个对象部件对应的层级生成网络；将部件层数据输入匹配的对象部件的层级生成网络；各个对象部件的层级生成网络分别输出与对象部件匹配的部件图像。

[0142] 在一些实施例中，装置还包括：

层级生成网络训练模块，用于获取第二训练样本；第二训练样本包括源训练部件图像、训练部件三维分割体素、目标训练部件二维姿态、部件变换图像和对应的标签部件图像，第二训练样本中的各个样本对应当前对象部件；将源训练部件图像、训练部件三维分割体素、目标训练部件二维姿态、部件变换图像输入所述当前对象部件对应的层级生成网络，经过层级生成网络所包括的各层依次处理后，输出对应的训练部件图像；根据训练部件图像和所述标签部件图像的差异反向传播调整层级生成网络的网络参数，以得到已训练的当前对象部件对应的层级生成网络。

[0143] 在一些实施例中，层级生成网络训练模块还用于根据标签部件图像、源训练部件图像、训练部件图像对当前对象部件对应的层级生成网络和判别网络进行对抗学习，得到部件对抗损失，标签部件图像是训练集中未被遮挡的与当前对象部件对应的图像；通过预训练感知网络分别提取标签部件图像和训练部件图像的特征得到特征图，计算特征图之间的距离得到部件感知损失；根据部件对抗损失、部件感知损失确定目标部件损失；根据目标部件损失反向传播调整所述当前对象部件对应的层级生成网络的网络参数，以得到已训练的层级生成网络。

[0144] 在一些实施例中，融合模块512还用于将部件图像根据二维分割图融合到所述变换图像的对应对象部件区域得到初始全局姿态图像；将源图像、目标姿态二维分割图、目标

二维姿态和初始全局姿态图像合并输入至精细尺度生成网络，输出得到目标姿态残差图像；将目标姿态残差图像与初始全局姿态图像融合得到目标姿态图像。

[0145] 图14示出了一些实施例中计算机设备的内部结构图。该计算机设备具体可以是图1中的终端110。如图14所示，该计算机设备包括通过系统总线连接的处理器、存储器、网络接口、输入装置和显示屏。其中，存储器包括非易失性存储介质和内存储器。该计算机设备的非易失性存储介质存储有操作系统，还可存储有计算机程序，该计算机程序被处理器执行时，可使得处理器实现姿态变换数据处理方法。该内存储器中也可储存有计算机程序，该计算机程序被处理器执行时，可使得处理器执行姿态变换数据处理方法。计算机设备的显示屏可以是液晶显示屏或者电子墨水显示屏，计算机设备的输入装置可以是显示屏上覆盖的触摸层，也可以是计算机设备外壳上设置的按键、轨迹球或触控板，还可以是外接的键盘、触控板或鼠标等。

[0146] 该计算机设备具体还可以是图1中的服务器120，可以包括比图中所示不同的部件。

[0147] 本领域技术人员可以理解，图14中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0148] 在一些实施例中，本申请提供的姿态变换数据处理装置可以实现为一种计算机程序的形式，计算机程序可在如图14所示的计算机设备上运行。计算机设备的存储器中可存储组成该姿态变换数据处理装置的各个程序模块，比如，图13所示的获取模块502、三维分割体素模块504、投影模块506、中间尺度模块508、部件图像生成模块510和融合模块512。各个程序模块构成的计算机程序使得处理器执行本说明书中描述的本申请各个实施例的姿态变换数据处理装置中的步骤。

[0149] 例如，图14所示的计算机设备可以通过如图13所示的姿态变换数据处理装置中的获取模块502获取源图像和目标三维姿态，通过三维分割体素模块504结合源图像和目标三维姿态，基于语义分割重建得到三维分割体素，三维分割体素包括体素的类别信息，通过投影模块506将三维分割体素投影得到对应的目标姿态二维分割图，基于体素的类别信息对目标姿态二维分割图中的对象标注得到对应的部件类别。通过中间尺度模块508获取目标三维姿态对应的目标二维姿态，提取源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像。通过部件图像生成模块510分别对源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据，对象部件是根据部件类别确定的，分别对各个对象部件的部件层数据进行部件合成，生成各个对象部件对应的部件图像。通过融合模块512将变换图像和部件图像融合得到目标姿态图像。

[0150] 在一些实施例中，提供了一种计算机设备，包括存储器和处理器，存储器存储有计算机程序，计算机程序被处理器执行时，使得处理器执行上述姿态变换数据处理方法的步骤。此处姿态变换数据处理方法的步骤可以是上述各个实施例的姿态变换数据处理方法中的步骤。

[0151] 在一些实施例中，提供了一种计算机可读存储介质，存储有计算机程序，计算机程序被处理器执行时，使得处理器执行上述姿态变换数据处理方法的步骤。此处姿态变换数据处理方法的步骤可以是上述各个实施例的姿态变换数据处理方法中的步骤。

[0152] 应该理解的是，虽然本申请各实施例的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，各实施例中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些子步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

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，所述的计算机程序可存储于一非易失性计算机可读取存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory，ROM)、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器(Random Access Memory，RAM)或外部高速缓冲存储器。作为说明而非局限，RAM可以是多种形式，比如静态随机存取存储器(Static Random Access Memory，SRAM)或动态随机存取存储器(Dynamic Random Access Memory，DRAM)等。

[0153] 以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

[0154] 以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

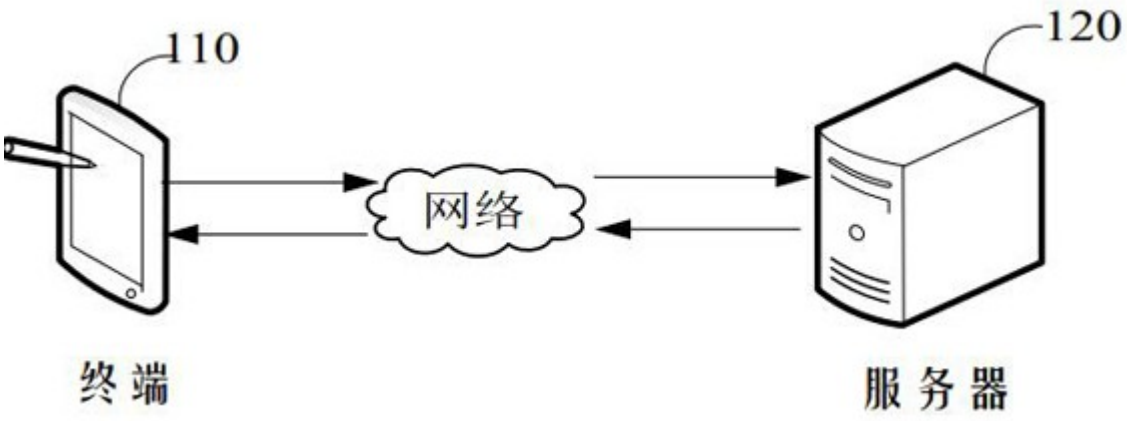


图1

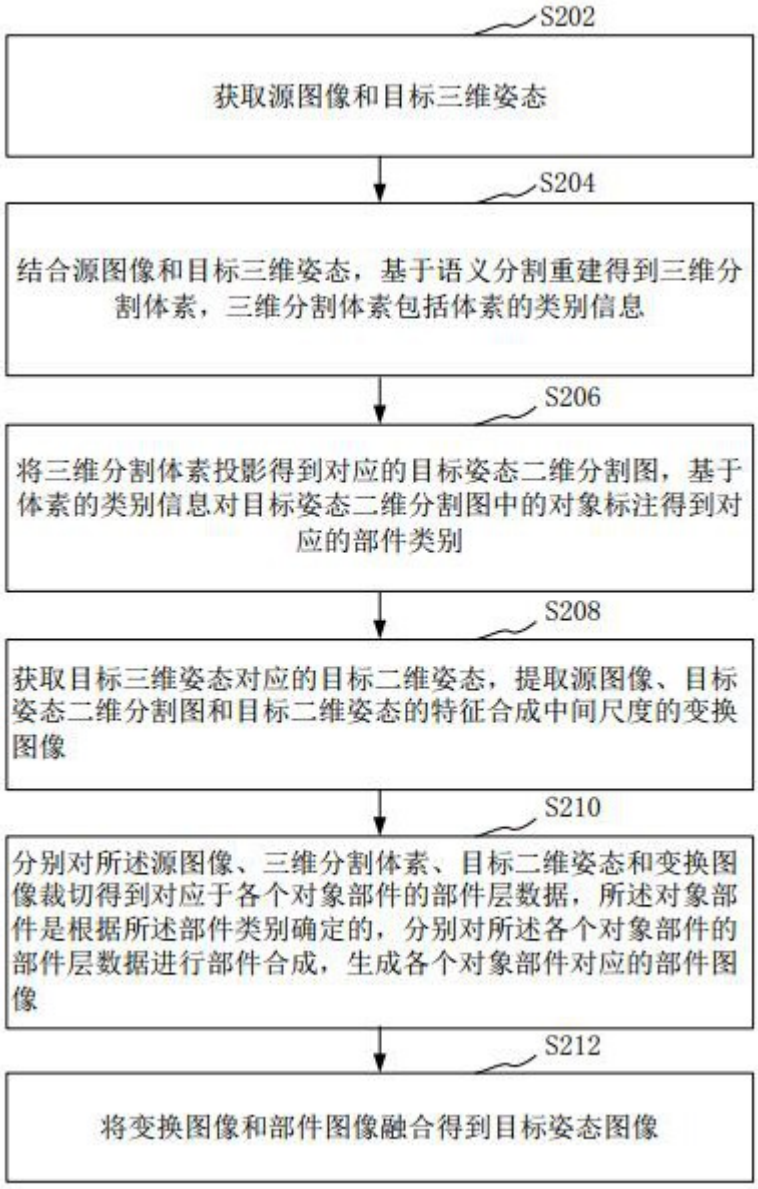

S202
获取源图像和目标三维姿态
S204
结合源图像和目标三维姿态，基于语义分割重建得到三维分割体素，三维分割体素包括体素的类别信息
S206
将三维分割体素投影得到对应的目标姿态二维分割图，基于体素的类别信息对目标姿态二维分割图中的对象标注得到对应的部件类别
S208
获取目标三维姿态对应的目标二维姿态，提取源图像、目标姿态二维分割图和目标二维姿态的特征合成中间尺度的变换图像
S210
分别对所述源图像、三维分割体素、目标二维姿态和变换图像裁切得到对应于各个对象部件的部件层数据，所述对象部件是根据所述部件类别确定的，分别对所述各个对象部件的部件层数据进行部件合成，生成各个对象部件对应的部件图像
S212
将变换图像和部件图像融合得到目标姿态图像

图2

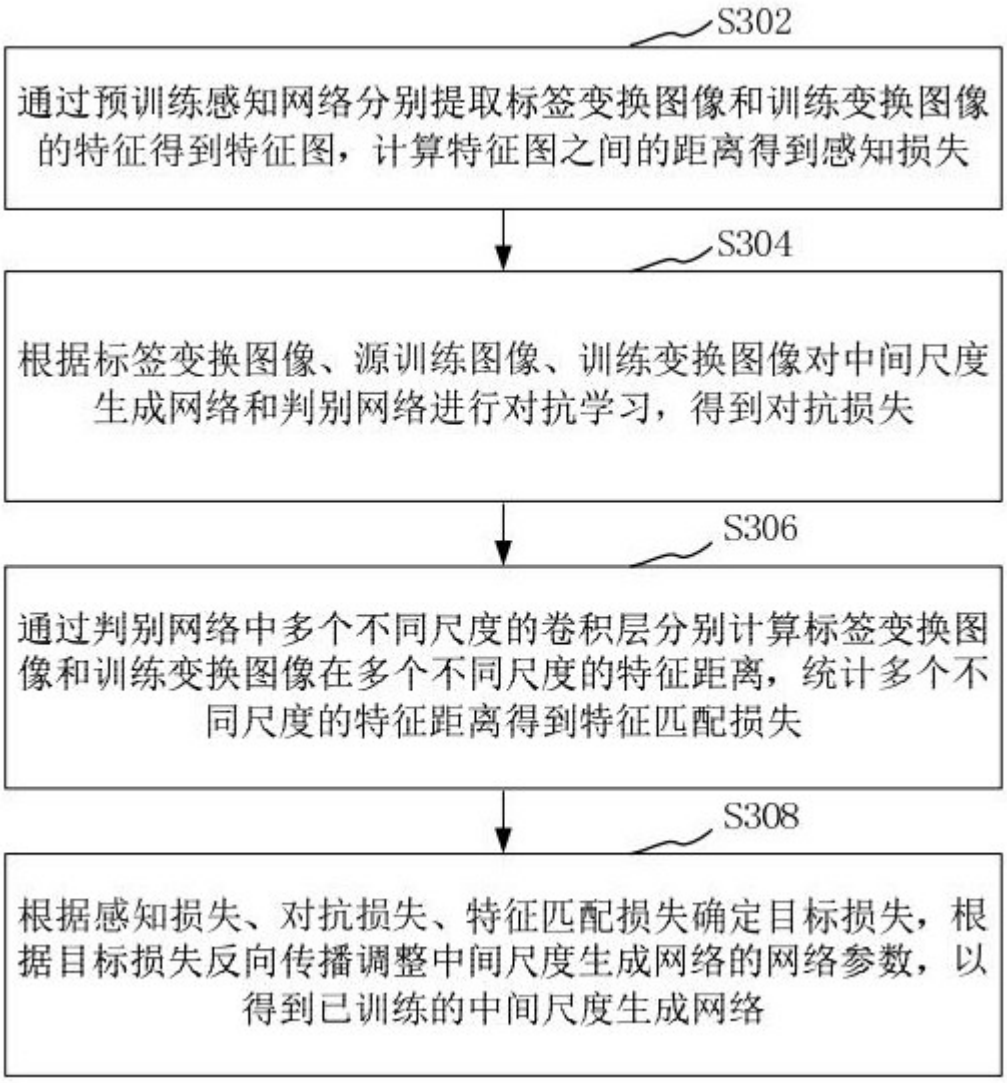

S302
通过预训练感知网络分别提取标签变换图像和训练变换图像的特征得到特征图，计算特征图之间的距离得到感知损失
S304
根据标签变换图像、源训练图像、训练变换图像对中间尺度生成网络和判别网络进行对抗学习，得到对抗损失
S306
通过判别网络中多个不同尺度的卷积层分别计算标签变换图像和训练变换图像在多个不同尺度的特征距离，统计多个不同尺度的特征距离得到特征匹配损失
S308
根据感知损失、对抗损失、特征匹配损失确定目标损失，根据目标损失反向传播调整中间尺度生成网络的网络参数，以得到已训练的中间尺度生成网络

图3

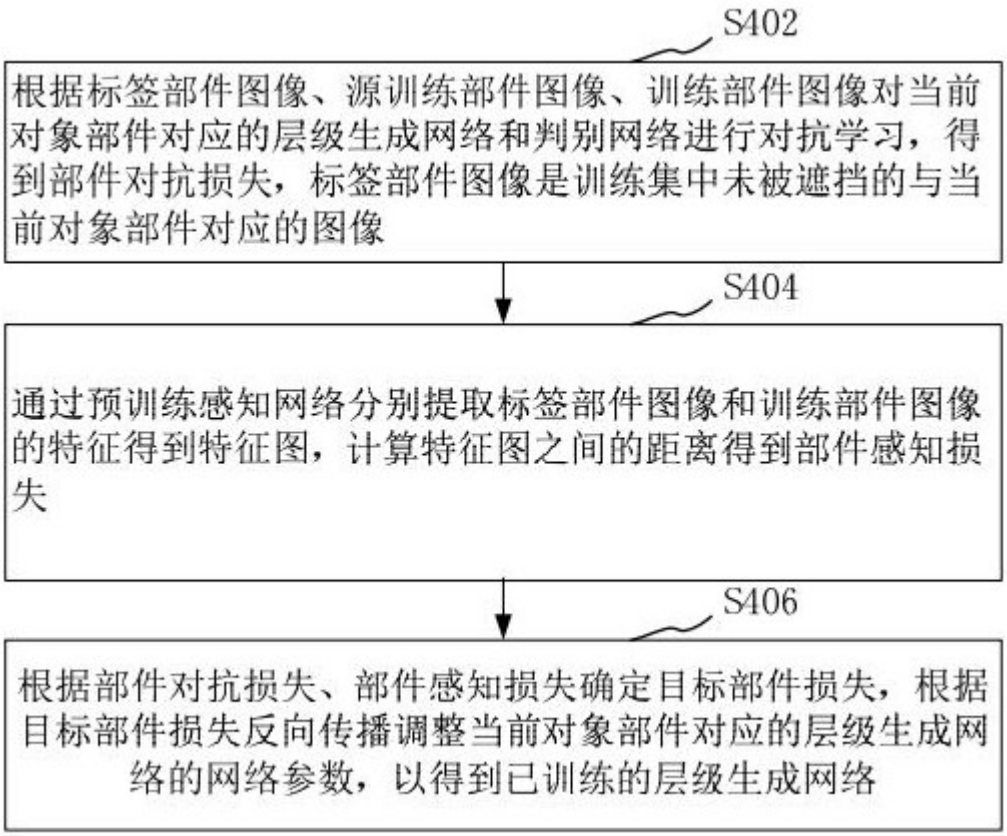

S402
根据标签部件图像、源训练部件图像、训练部件图像对当前对象部件对应的层级生成网络和判别网络进行对抗学习，得到部件对抗损失，标签部件图像是训练集中未被遮挡的与当前对象部件对应的图像
S404
通过预训练感知网络分别提取标签部件图像和训练部件图像的特征得到特征图，计算特征图之间的距离得到部件感知损失
S406
根据部件对抗损失、部件感知损失确定目标部件损失，根据目标部件损失反向传播调整当前对象部件对应的层级生成网络的网络参数，以得到已训练的层级生成网络

图4

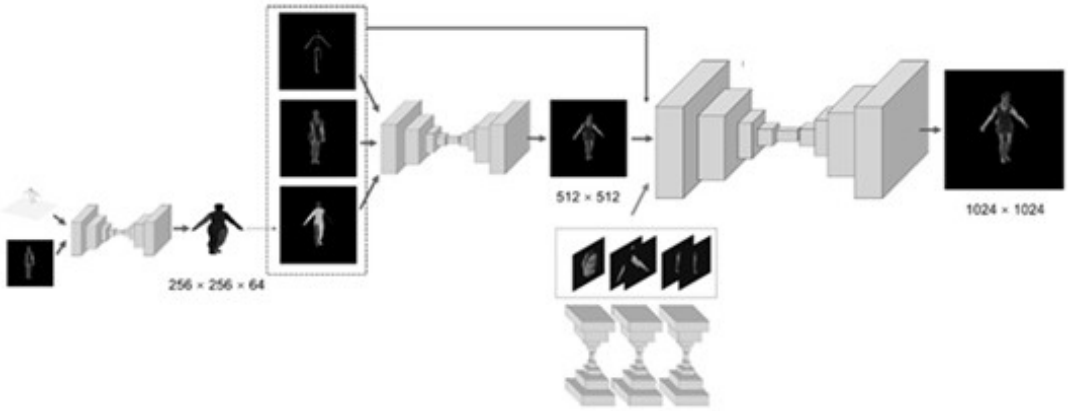

256 × 256 × 64
512 × 512
1024 × 1024

图5

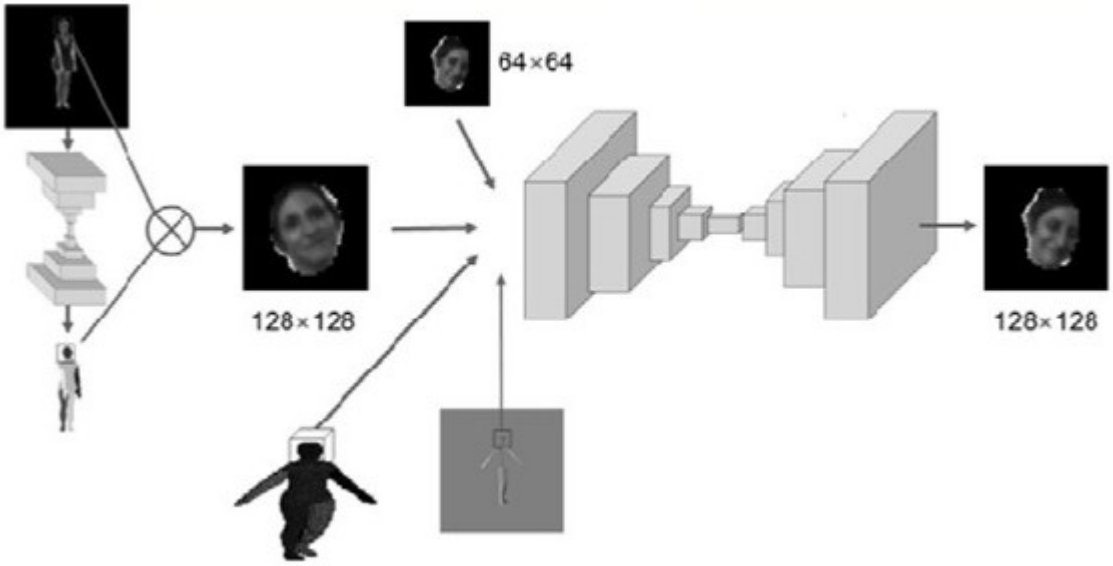

64×64
128×128
128×128

图6

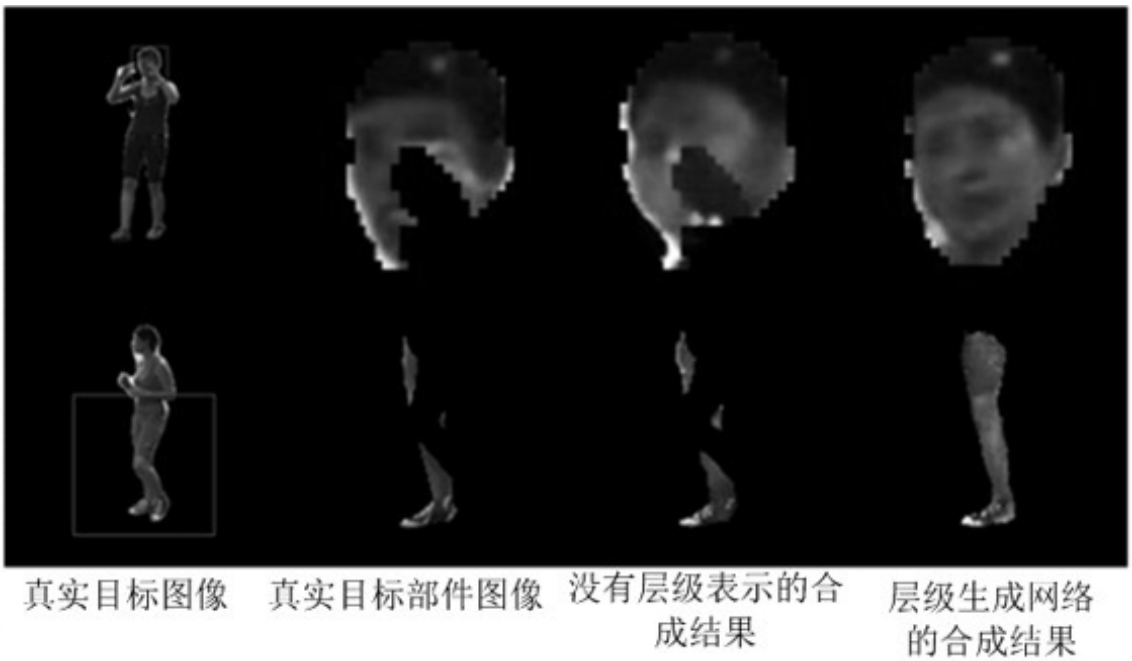

真实目标图像
真实目标部件图像
没有层级表示的合成结果
层级生成网络的合成结果

图7

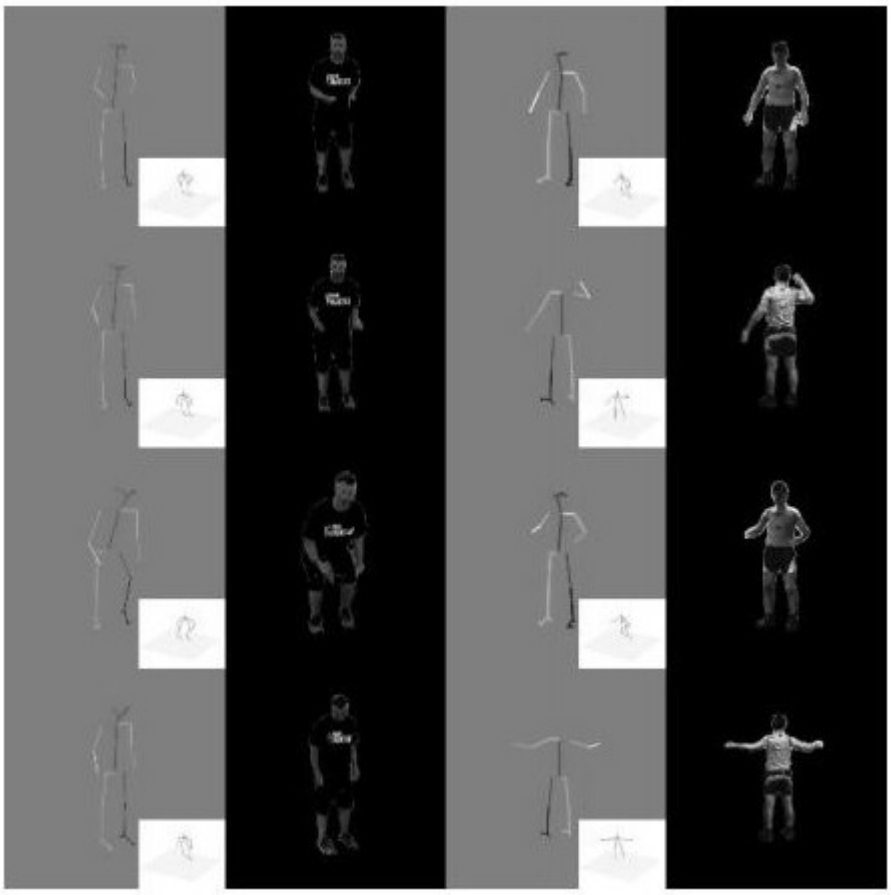


图8

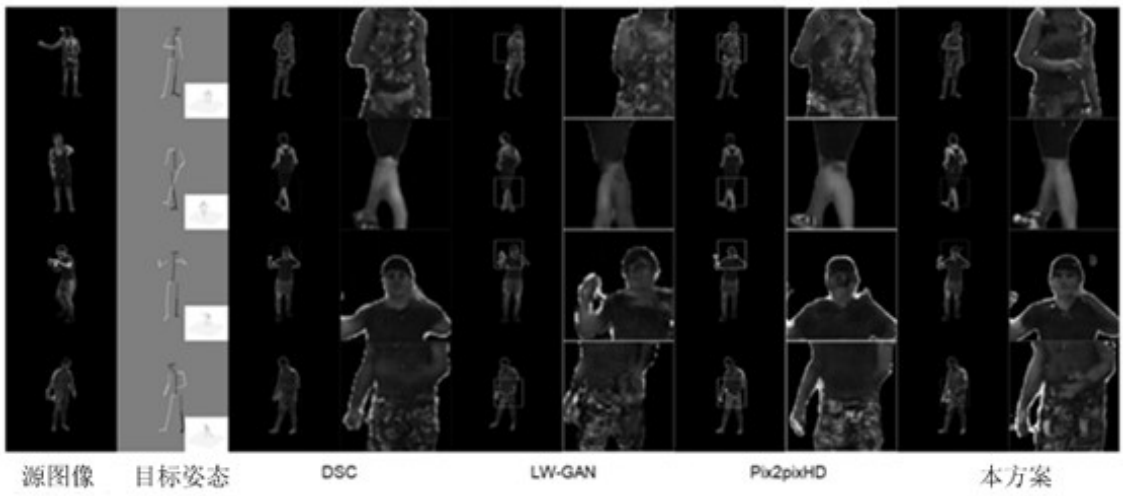


图9

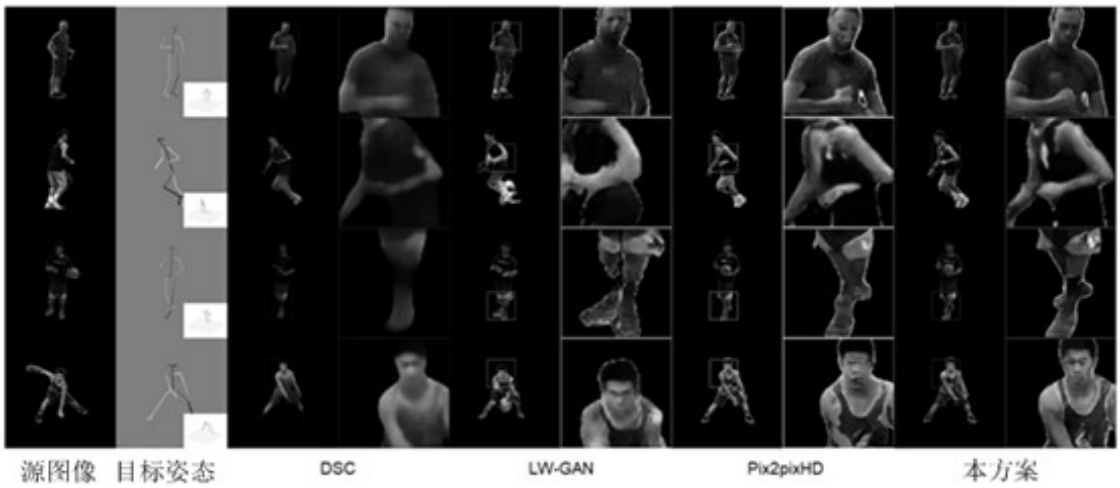


图10

Method	SSIM (↑)[42]	LPIPS (↓) [47]
Pix2pixHD [41]	0.9457	0.0510
DSC [37]	0.9174	0.0652
LWGAN [26]	0.8477	0.1599
Our method	**0.9469**	**0.0473**

Method	SSIM (↑)[42]	LPIPS (↓) [47]
Pix2pixHD [41]	0.9244	0.0778
DSC [37]	0.8819	0.1807
LWGAN [26]	0.8234	0.1992
Our method	**0.9276**	**0.0765**

图11

Method	SSIM (↑)[42]	LPIPS (↓)[47]
Baseline	0.9370	0.0525
$Baseline + V$	**0.9478**	0.0478
$Baseline + V + PL^{-}$	0.9457	0.0488
$Baseline + V + PL$	0.9469	**0.0473**

图12

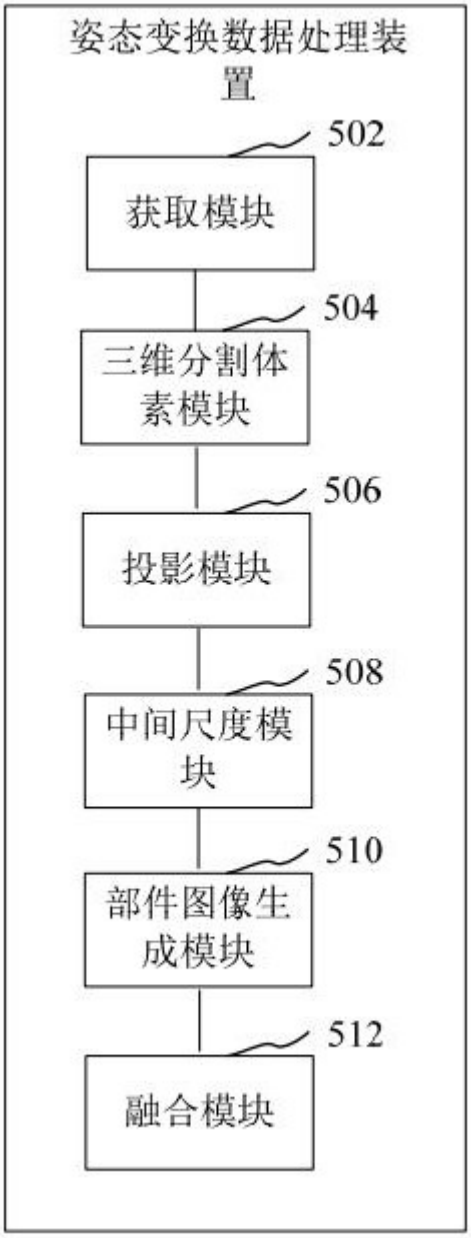

姿态变换数据处理装置
502
获取模块
504
三维分割体素模块
506
投影模块
508
中间尺度模块
510
部件图像生成模块
512
融合模块

图13

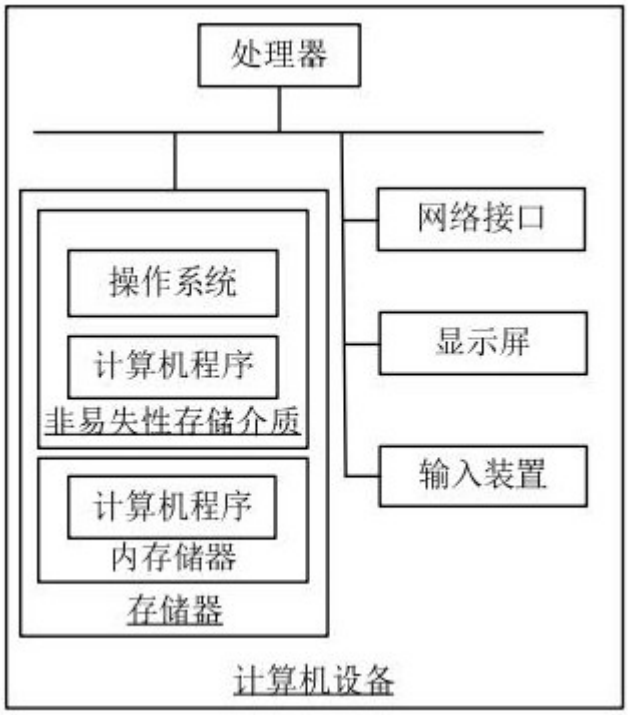


图14