

图像生成方法、装置、计算机设备及计算机可读存储介质

本申请要求于 2021 年 01 月 28 日提交中国专利局，申请号为 202110118922.2，申请名称为“图像生成方法、装置、计算机设备及计算机可读存储介质”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

本申请涉及图像处理技术领域，特别涉及一种图像生成方法、装置、计算机设备、计算机可读存储介质及计算机程序产品。

背景技术

虚拟换装是指应用图像融合技术，将用户的人体图像和包含目标衣物的衣物图像进行融合，得到用户穿戴该目标衣物后的图像，从而在用户无需真正试穿目标衣物的情况下，就能够了解到目标衣物的穿戴效果。

目前，在虚拟换装过程中，通常应用图像融合模型，分别对人体图像和衣物图像进行特征提取，基于提取到的两个图像特征生成新的图像，即用户穿戴目标衣物的图像。但是，在上述过程中，由于图像融合模型所提取的是粗略的图像特征，在生成图像时容易导致新生成的图像缺失细节信息，进而导致图像生成效果失真，虚拟换装的效果较差。

发明内容

本申请的实施例提供了一种图像生成方法、装置、计算机设备、计算机可读存储介质及计算机程序产品。

一种图像生成方法，由计算机设备执行，该方法包括：

获取包含目标人体的第一人体图像以及包含目标衣物的第一衣物图像；

基于该第一人体图像中该目标人体的姿态，对该第一衣物图像进行变换，得到第二衣物图像，该第二衣物图像包含该目标衣物，该目标衣物的姿态与该目标人体的姿态相匹配；

分别对该第二衣物图像、该第一人体图像中裸露区域的图像以及该第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，该裸露区域是该第一人体图像中的该目标人体中未被衣物遮挡的区域；及

基于该衣物特征、该皮肤特征以及该人体特征，生成第二人体图像，该第二人体图像中该目标人体穿戴该目标衣物。

一种图像生成装置，该装置包括：

第一获取模块，用于获取包含目标人体的第一人体图像以及包含目标衣物的第一衣物图像；

衣物变换模块，用于基于该第一人体图像中该目标人体的姿态，对该第一衣物图像进行变换，得到第二衣物图像，该第二衣物图像包含该目标衣物，该目标衣物的姿态与该目标人体的姿态相匹配；

特征提取模块，用于分别对该第二衣物图像、该第一人体图像中裸露区域的图像以及该第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，该裸露区域是该第一人体图像中的该目标人体中未被衣物遮挡的区域；及

第一生成模块，用于基于该衣物特征、该皮肤特征以及该人体特征，生成第二人体图像，该第二人体图像中该目标人体穿戴该目标衣物。

一种计算机设备，包括存储器和一个或多个处理器，所述存储器存储有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行上述图像生成方法的步骤。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，其上存储有计算机可读指令，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理

器执行上述图像生成方法的步骤。

一种计算机程序产品或计算机程序，所述计算机程序产品或计算机程序包括计算机可读指令，所述计算机可读指令存储在计算机可读存储介质中，计算机设备的处理器从所述计算机可读存储介质读取所述计算机可读指令，所述处理器执行所述计算机可读指令，使得所述计算机设备执行上述图像生成方法的步骤。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征、目的和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例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 是本申请一个实施例提供的一种图像生成方法的实施环境示意图；
- 图 2 是本申请一个实施例提供的一种图像生成方法的流程图；
- 图 3 是本申请一个实施例提供的一种图像融合模型的示意图；
- 图 4 是本申请另一个实施例提供的一种图像生成方法的流程图；
- 图 5 是本申请一个实施例提供的一种图像融合模型的示意图；
- 图 6 是本申请一个实施例提供的一种空间变换网络的示意图；
- 图 7 是本申请一个实施例提供的一种特征拼接的示意图；
- 图 8 是本申请一个实施例提供的一种解码网络的示意图；
- 图 9 是本申请一个实施例提供的一种图像融合模型的训练方法流程图；
- 图 10 是本申请一个实施例提供的一种模型训练方法的示意图；
- 图 11 是本申请一个实施例提供的一种图像生成装置的结构示意图；
- 图 12 是本申请一个实施例提供的一种终端的结构示意图；
- 图 13 是本申请一个实施例提供的一种服务器的结构示意图。

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述，显然，所描述的实施例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。

本申请中术语“第一”“第二”等字样用于对作用和功能基本相同的相同项或相似项进行区分，应理解，“第一”、“第二”、“第 n”之间不具有逻辑或时序上的依赖关系，也不对数量和执行顺序进行限定。

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。本申请实施例涉及人工智能技术中的计算机视觉技术以及机器学习技术。

以下对本申请涉及的术语进行解释：

空间变换网络(Spatial Transformer Network, STN)：是一种不需要引入额外数据标签，便可以在网络中对数据，如特征图等，进行空间变换的操作网络。

薄板样条算法(Thin Plate Spline, TPS)：是一种插值算法，能够用于图像变形(image warping)等任务中，通过少量的控制点来驱动图像进行变化。

生成器(Generator)：是指基于深度神经网络搭建的用于生成图像的网络，通常的结构为 encoder-decoder 结构，即下采样-上采样结构。

人体语义信息：在本申请实施例中，人体语义信息指的是将人体中各个身体部位通过不同的像素值来表示的信息。

图 1 是本申请实施例提供的一种图像生成方法的实施环境示意图，参见图 1，该实施环境包括终端 110 和服务器 140。

其中，终端 110 安装和运行有支持图像生成的应用程序，例如，该应用程序是虚拟换装应用程序、电商类应用程序，该应用程序能够调用图像生成功能，对用户的人体图像和目标衣物的图像进行融合，生成新的人体图像，该新的人体图像能够呈现该用户穿着该目标衣物的效果。在一些实施例中，该终端 110 可以是智能手机、平板电脑、笔记本电脑、台式计算机、物联网设备和便携式可穿戴设备，物联网设备可为智能电视、智能音箱、智能空调、智能车载设备等；便携式可穿戴设备可为智能手表、智能手环、头戴设备等等，在一些实施例中，该终端 110 是智能家居，例如，该终端 110 是智能镜子、智能衣柜等，本申请实施例对该终端 110 的设备类型不做限定。示例性的，终端 110 是用户使用的终端，终端 110 中运行的应用程序内登录有用户账号。终端 110 可以泛指多个终端中的一个，本实施例仅以终端 110 来举例说明。

服务器 140 可以是一台服务器、多台服务器、云计算平台和虚拟化中心中的至少一种。服务器 140 用于为支持图像生成的应用程序提供后台服务。在一些实施例中，服务器 140 承担主要图像生成工作，终端 110 承担次要图像生成工作；或者，服务器 140 承担次要图像生成工作，终端 110 承担主要图像生成工作；或者，服务器 140 或终端 110 分别可以单独承担图像生成工作。

在一些实施例中，服务器 140 包括：接入服务器、图像生成服务器和数据库。接入服务器用于为终端 110 提供数据接入服务。图像生成服务器用于提供图像生成有关的后台服务，该图像生成服务器可以搭载图形处理器，支持图形处理器多线程并行计算。图像生成服务器可以是一台或多台。当图像生成服务器是多台时，存在至少两台图像生成服务器用于提供不同的服务，和/或，存在至少两台图像生成服务器用于提供相同的服务，比如以负载均衡方式提供同一种服务，本申请实施例对此不加以限定。图像生成服务器中可以设置有图像融合模型，在模型训练和应用过程中，该图像生成服务器可以搭载图像处理器 GPU（Graphics Processing Unit，图形处理器），并支持图像处理器并行运算。其中，上述服务器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、域名服务、安全服务、CDN（Content Delivery Network，内容分发网络）、以及大数据和人工智能平台等基础云计算服务的云服务器。

上述终端 110 与图像生成平台 140 可以通过有线或无线通信方式进行直接或间接地连接，本申请实施例对此不作限定。

在一些实施例中，上述实施环境可以部署在区块链系统中，示例性的，上述服务器为区块链上的节点设备，该服务器中搭载有图像生成模型，终端响应于虚拟换装事件，向区块链系统中的该服务器发送虚拟换装指令，其中，虚拟换装事件可以是用户通过与终端的交互触发的虚拟换装操作，也可以是终端通过检测设备在检测到符合触发条件时自动触发的事件。该虚拟换装指令携带有用户的私钥签名，服务器应用该用户的公钥对该虚拟换装指令进行验证，响应于验证通过，由服务器中的图像生成模型，基于该虚拟换装指令执行虚拟换装任务。需要说明的是，上述终端也可以是区块链系统中的节点设备，本申请实施例对此不作限定。

本申请实施例提供的图像生成方法，能够与多种应用场景相结合。例如，在电商类应用程序中，用户需要对应用程序中所展示的衣物进行试穿时，计算机设备可以调用该图像生成功能，基于用户的人体图像和待试穿衣物的衣物图像，生成新的人体图像，该新的人体图像能够呈现出用户穿戴该待试穿衣物的效果。当然，本申请实施例提供的图像生成方

法也可以应用于智能镜子、智能衣柜等智能家居上，本申请实施例对该图像生成方法的应用场景不作限定。

具体地，计算机设备基于目标人体的姿态，对原始的第一衣物图像进行变换，得到与目标人体的姿态相贴合、并且保留了大量细节信息的第二衣物图像，再基于不同特征提取任务，分别从不同维度对第二衣物图像和第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，也即是，得到三个不同维度的、更细粒度的、更精确的特征，从而基于这些特征所生成的第二人体图像能够包含更丰富的细节信息，使第二人体图像具有较高的真实度，确保虚拟换装效果良好。

图2是本申请实施例提供的一种图像生成方法的流程图。该方法可以应用于上述终端或者服务器，而终端和服务器均可以视为一种计算机设备，因此，在本申请实施例中，以计算机设备作为执行主体，对该图像生成方法进行介绍，参见图2，该实施例具体可以包括以下步骤：

201、计算机设备获取包含目标人体的第一人体图像以及包含目标衣物的第一衣物图像。

其中，该第一人体图像可以是目标人体的全身图像或者半身图像，第一衣物图像在人体上的穿着部位与目标人体的第一人体图像所展示的人体区域相匹配，例如，该第一人体图像是目标人体的上半身图像，该第一衣物图像为上半身衣物对应的图像，如上衣的图像；再例如，该第一人体图像是目标人体的下半身图像，该第一衣物图像为下半身衣物图像，如裤子的图像、半裙的图像等；再例如，该第一人体图像是目标人体的全身图像，该第一衣物图像为全身衣物图像，如连衣裙的图像等。该第一人体图像和第一衣物图像可以是存储在计算机设备中的图像，也可以是计算机设备从视频中截取的图像，还可以是具备图像采集功能的设备实时采集的图像，例如，计算机设备与相机建立连接，相机将实时拍摄的图像发送给该计算机设备，或者，该计算机设备配置有摄像头，通过该摄像头来实时采集图像，本申请实施例对具体采用哪种图像不作限定。

202、计算机设备基于该第一人体图像中该目标人体的姿态，对该第一衣物图像进行变换，得到第二衣物图像，该第二衣物图像包含该目标衣物，该目标衣物的姿态与该目标人体的姿态相匹配。

在本申请实施例中，计算机设备基于目标人体的姿态对包含目标衣物的第一衣物图像进行变换，使目标衣物的姿态与目标人体的姿态相贴合，得到第二衣物图像，该第二衣物图像充分保留目标衣物的细节信息，从而后续对第二衣物进行特征提取时，能够提取到丰富的细节特征。

203、计算机设备分别对该第二衣物图像、该第一人体图像中裸露区域的图像以及该第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，该裸露区域是该第一人体图像中的该目标人体中未被衣物遮挡的区域。

在一种可能实现方式中，该计算机设备从多个维度对该第二衣物图像和该第二人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征。示例性的，对特征提取任务进行任务拆解，拆解为三个维度的特征提取子任务，即计算机设备分别从衣物维度、皮肤维度以及整体人像维度进行特征提取，得到衣物特征、皮肤特征以及人体特征。在本申请实施例中，通过任务拆解的方式分别提取各个维度的特征，一方面，能够提取到更全面、更细粒度的特征，另一方面，各个特征提取子任务分开执行，也能够降低计算机设备在特征提取时的数据处理负担。

204、计算机设备基于该衣物特征、该皮肤特征以及该人体特征，生成第二人体图像，该第二人体图像中该目标人体穿戴该目标衣物。

在一种可能实现方式中，该计算机设备可以将多维度的特征，即衣物特征、皮肤特征以及人体特征进行拼接，得到拼接特征，拼接特征包含了衣物的细节信息、人体皮肤信息

以及人体全局细节信息，计算机设备可以对该拼接特征进行解码，生成该用户穿戴该目标衣物的人体图像，即第二人体图像，以达到虚拟换装的效果。

本申请实施例提供的技术方案，通过先基于目标人体的姿态，对原始的第一衣物图像进行变换，得到与目标人体的姿态相贴合，并且保留了大量细节信息的第二衣物图像，再基于不同特征提取任务，分别从不同维度对第二衣物图像和第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，也即是，得到三个不同维度的、更细粒度的、更精确的特征，从而基于这些特征所生成的第二人体图像能够包含更丰富的细节信息，使第二人体图像具有较高的真实度，确保虚拟换装效果良好。

上述实施例是对本申请实施方式的一个简要介绍，在一种可能实现方式中，上述第二衣物图像的获取步骤、特征提取步骤以及第二人体图像的生成步骤可以通过图像融合模型执行，该图像融合模型为已训练的模型，该图像融合模型可以是存储在计算机设备中的模型，也可以是网络中的模型。图3是本申请实施例提供的一种图像融合模型的示意图，参见图3，该图像融合模型包括衣物区域预测网络301、空间变换网络302、衣物编码网络303、皮肤编码网络304、人像编码网络305以及解码网络306。其中，该衣物区域预测网络301用于确定目标人体穿戴目标衣物后，该目标衣物的覆盖区域；空间变换网络302用于基于目标人体的姿态，对衣物图像进行仿射变换；衣物编码网络303、皮肤编码网络304以及人像编码网络305用于从不同维度提取人体图像和衣物图像的特征；解码网络306用于对提取到的多维度的特征进行解码，生成新的人体图像。需要说明的是，上述对图像融合模型的说明，仅是一种示例性说明，本申请实施例对该图像融合模型的结构不做限定，在本申请实施例中，仅以图3所示的图像融合模型进行图像生成的过程为例进行说明。图4是本申请实施例提供的一种图像生成方法的流程图，以下结合图3和图4，对图像生成方法进行说明，在一种可能实现方式中，该图像生成方法包括以下步骤。

401、计算机设备获取包含目标人体的第一人体图像以及包含目标衣物的第一衣物图像。

在一种可能实现方式中，计算机设备响应于虚拟换装指令，获取第一人体图像以及第一衣物图像。示例性的，该计算机设备是用户使用的终端，该计算机设备所运行的目标应用程序支持虚拟换装功能，例如，该目标应用程序是电商应用程序，计算机设备响应于用户在该电商应用程序中所触发的虚拟换装指令，获取第一人体图像和第一衣物图像。其中，该第一人体图像可以是预先存储在该计算机设备中的图像，也可以是计算机设备通过摄像头实时拍摄的图像；该第一衣物图像可以是预先存储在该计算机设备中的图像，也可以是电商应用程序中的图像，即用户在该电商应用程序中所选中的目标衣物所对应的图像。需要说明的是，本申请实施例对该第一人体图像和第一衣物图像得到获取方法不做限定。

在一种可能实现方式中，计算机设备获取到该第一人体图像和第一衣物图像后，可以对该第一人体图像和第一衣物图像进行预处理，示例性的，计算机设备对该第一人体图像和第一衣物图像进行的预处理可以是调整图像的尺寸、对图像进行图像增强处理中的至少一种，其中，图像增强处理可以是对图像进行降噪、增强对比度等处理中的至少一种，以抑制图像中的无用信息，改善图像的视觉效果，将图像调整为更适合机器分析处理的形式。

402、计算机设备基于该第一人体图像中目标人体的姿态以及该第一衣物图像，确定第一区域图像。

其中，该第一区域图像用于指示该目标人体穿戴该目标衣物时，该目标衣物的遮挡区域，示例性的，该第一区域图像可以为二值化的掩模图像，该目标衣物遮挡区域的像素点表示为第一数值，其他区域的像素点表示为第二数值。

在一种可能实现方式中，计算机设备先对该第一人体图像进行人体区域识别，将该第一人体图像中属于相同身体区域的像素点设置为相同像素值，得到人体语义图像，再基于该人体语义图像以及该第一衣物图像，确定该第一区域图像。其中，该人体语义图像用于

指示该目标人体的姿态, 计算机设备能够基于该人体语义图像对目标人体的各个身体区域进行区分。

在一种可能实现方式中, 计算机设备通过人体区域识别网络获取该人体语义图像。其中, 该人体区域识别网络是基于卷积神经网络所构建的, 包括至少一个用于提取人体语义特征的运算层, 例如, 该人体区域识别网络可以是 FCN (Fully Convolutional Networks, 全卷积网络), 本申请对该人体区域识别网络的结构不做限定。示例性的, 计算机设备将第一人体图像输入该人体区域识别网络, 由该人体区域识别网络中的至少一个卷积层对该第一人体图像进行特征提取, 通过反卷积对最后一个卷积层所输出的特征图进行上采样, 使该特征图恢复到与该第一人体图像相同的尺寸, 再基于上采样后的特征图进行像素点分类, 确定每个像素点所属的身体区域, 从而得到该人体语义图像。上述对人体语义图像获取方法的说明, 仅是一种可能实现方式的示例性说明, 本申请实施例对具体采用哪种方法获取人体语义图像不做限定。

需要说明的是, 在一种可能实现方式中, 上述人体区域识别网络可以包含在图像融合模型中, 图 5 是本申请实施例提供的一种图像融合模型的示意图, 如图 5 所示, 在图像融合模型 501 中的衣物区域预测网络 502 之前, 设置有该人体区域识别网络 503。在一种可能实现方式中, 该人体区域识别网络也可以不包括在图像融合模型中, 即计算机设备在获取到第一人体图像和第一衣物图像之后, 先基于该第一人体图像获取人体语义图像, 再将人体语义图像和该第一衣物图像输入图像融合模型, 由图像融合模型执行后续的图像生成步骤, 本申请实施例对此不做限定。

在一种可能实现方式中, 计算机设备获取到人体语义图像后, 将该人体语义图像以及该第一衣物图像输入图像融合模型的衣物区域预测网络 (Mask Prediction Network, MPN)。其中, 该衣物区域预测网络可以实现为 encoder-decoder (编码器-解码器) 结构, 编码器和解码器部分均包括多个卷积层。示例性的, 计算机设备通过该衣物区域预测网络, 提取该人体语义图像对应的人体姿态特征以及该第一衣物图像中该目标衣物的轮廓特征, 即通过该衣物区域预测网络中编码器部分的卷积层, 对人体语义图像和第一衣物图像进行下采样; 计算机设备基于该人体姿态特征以及该目标衣物的轮廓特征, 确定该第一区域图像, 即通过该衣物区域预测网络中解码器部分的卷积层, 对编码器所输出的特征进行上采样, 得到该第一区域图像。其中, 该第一区域图像中, 该目标人体穿戴该目标衣物时, 该目标衣物的遮挡区域的像素点为第一数值, 其他区域的像素点为第二数值。需要说明的是, 上述对第一区域图像获取方法的说明, 仅是一种可能实现方式的示例性说明, 本申请实施例对具体采用哪种方法获取该第一区域图像不做限定。

在一种可能实现方式中, 该衣物区域预测网络还可以输出第二区域图像, 该第二区域图像用于指示该第一区域图像用于指示该目标人体穿戴该目标衣物时, 该目标人体图像的裸露区域, 即裸露的皮肤区域, 例如, 手部区域等。如图 3 所示, 将人体语义图像 D 和第一衣物图像 C 输入衣物预测网络, 该衣物预测网络输出第一区域图像 M^{clothes} 和第二区域图像 M^{skin} 。在一种可能实现方式中, 该第二图像区域可以应用于后续的特征提取步骤, 使图像融合模型能够确定出裸露的皮肤区域的特征, 如皮肤区域的轮廓特征、位置特征等, 从而便于图像融合模型更准确的生成新的人体图像。

在本申请实施例中, 通过先获取第一区域图像, 即先确定目标衣物所呈现出的姿态, 再进行衣物仿射变换, 能够使得衣物图像与目标人体的姿态相贴合, 有效提高衣物仿射变换的准确度, 提高仿射变换所得到的衣物图像的真实度。

403、计算机设备基于该第一区域图像, 对该第一衣物图像进行变换, 得到第二衣物图像。

在本申请实施例中, 计算机设备可以通过图像融合模型中的空间变换网络对该第一衣物进行变换, 得到该第二衣物图像。图 6 是本申请实施例提供的一种空间变换网络的示意

图,如图6所示,该空间变换网络包括变换矩阵预测子网络601和衣物生成子网络602,示例性的,该变换矩阵预测子网络和衣物生成子网络均包括多个运算层,该运算层可以是卷积层、池化层等,本申请实施例对此不做限定,其中,该变换矩阵预测子网络用于确定第一衣物图像在仿射变换时所应用的仿射变换矩阵,该衣物生成子网络包括衣物细节特征,能够对衣物图像进一步修饰。

以下结合图6对该第二衣物图像的获取方法进行说明。在一种可能实现方式中,计算机设备将该第一区域图像以及该第一衣物图像输入图像融合模型的空间变换网络,通过该空间变换网络基于该第一衣物图像以及该第一掩模区域之间的映射关系,对该第一衣物图像进行变换,得到中间衣物图像,第一掩模区域用于指示第一衣物图像中的目标衣物所在的区域,计算机设备再通过该空间变换网络确定该中间衣物图像所对应的纹理特征,基于该纹理特征以及该中间衣物图像,生成该第二衣物图像。

在一些实施例中,计算机设备在将第一衣物图像输入空间变换网络之前,还可以获取该第一衣物图像对应的第一掩模区域,即该第一衣物图像对应的初始区域图像,该初始区域图像用于指示该第一衣物图像中该目标衣物的所在区域,该初始区域图像用于辅助确定该第一区域图像。

在一些实施例中,计算机设备基于第一掩模区域确定目标衣物的初始轮廓,并基于第一区域图像确定目标衣物在变换后对应的目标轮廓,基于初始轮廓以及目标轮廓,确定初始轮廓和目标轮廓之间的映射关系,基于映射关系,对第一衣物图像进行变换,得到中间衣物图像。

在一些实施例中,计算机设备基于该初始区域图像进行衣物轮廓识别处理,确定目标衣物的初始轮廓,基于该第一区域图形确定目标衣物在变换后对应的目标轮廓,通过将初始轮廓与目标轮廓进行轮廓比对,获取该初始轮廓和目标轮廓之间的映射关系,即目标衣物在变换前后的对应关系,该初始轮廓和目标轮廓之间的映射关系能够辅助计算机设备在对目标衣物进行变换的过程中,对目标衣物的形状进行调整。

如图6所示,计算机设备将第一衣物图像C、初始区域图像M和第一区域图像 $M^{clothes}$ 输入空间变换网络,由该空间变换网络中的变换矩阵预测子网络,预测该第一衣物图像对应的仿射变换矩阵T,再调用TPS(Thin Plate Splines,薄板样条函数)算法,基于该仿射变换矩阵T对该第一衣物图像进行变换,即基于该仿射变换矩阵T调整该第一衣物图像中各个像素点的位置,得到中间衣物图像 C^{raw} ,该中间衣物图像中,目标衣物的姿态与目标人体的姿态相贴合。

在其中一些实施例中,空间变换网络包括衣物生成子网络,计算机设备能够基于衣物生成子网络进一步对中间衣物图像进行修饰,添加细节信息,计算机设备基于所述衣物生成子网络,按照预先学习的衣物细节特征,对所述中间衣物图像进行特征提取,得到所述中间衣物图像对应的纹理特征。

在一个具体地应用中,如图6所示,衣物生成子网络输出包含纹理特征的纹理特征图像 M^{α} 和中间衣物图像 C^{raw} ,生成第二衣物图像 C^{wap} 。在一种可能实现方式中,衣物生成子网络所包括的衣物细节特征,可以是在模型训练阶段所学习到的,也即是,衣物生成子网络基于预先学习的衣物细节特征,对该中间衣物图像进行特征提取,得到该中间衣物图像对应的纹理特征;在一些实施例中,也可以将第一衣物图像输入该衣物生成子网络,由该衣物生成子网络对该第一衣物图像中目标衣物的衣物细节特征进行学习,基于学习到的衣物细节特征来预测该纹理特征,本申请实施例对此不作限定。需要说明的是,上述对获取第二衣物图像的方法的说明,仅是一种可能实现方式的示例性说明,本申请实施例对具体采用哪种方法获取第二衣物图像不作限定。在本申请实施例中,通过衣物生成子网络对中间衣物图像进行进一步修饰,能够最大限度保留该目标衣物的纹理,使生成的第二衣物图像的视觉效果更加自然。

需要说明的是，上述对衣物图像进行变换的方法的说明，仅是一种可能实现方式的示例性说明，在一些实施例中，也可以应用光流变换的方式代替仿射变换，以更好地保留衣物得到细节信息，本申请实施例对采用哪种方式对衣物图像进行变换不作限定。

需要说明的是，上述步骤 402 和步骤 403 是基于该第一人体图像中该目标人体的姿态，对该第一衣物图像进行变换，得到第二衣物图像，该第二衣物图像包含该目标衣物，该目标衣物的姿态与该目标人体的姿态相匹配的步骤。在本申请实施例中，通过先对衣物进行仿射变换，得到贴合人体姿态的衣物图像，在这一过程中，不仅能够准确捕捉到目标人体的姿态信息，也能够充分保留衣物的细节信息，例如，最大限度的保留衣物的纹理信息、领口、袖口等细节信息。

404、计算机设备对第二衣物图像进行特征提取，得到衣物特征。

在本申请实施例中，计算机设备可以通过图像融合模型的衣物编码网络对该第二衣物图像进行特征提取，得到衣物特征。在一种可能实现方式中，该衣物编码网络可以采用 ResUnet（一种图像生成网络）中编码网络的结构，该衣物编码网络包括多个级联的运算层，计算机设备通过该多个级联的运算层，逐层提取第二衣物图像的特征，得到各个运算层输出的编码向量，作为衣物子特征，各个衣物子特征包含了第二衣物图像中目标衣物的全局和局部特征信息，该计算机设备将各个衣物子特征，作为该目标衣物对应的衣物特征。示例性的，每个运算层输出的衣物子特征尺度不同，例如，位置越靠前的运算层所输出的衣物子特征的尺度越大，能够保留更多的衣物细节信息，位置靠后的运算层所输出的衣物子特征的尺度越小，能够体现衣物的全局特征。在一些实施例中，可以将上述衣物子特征均表示为向量的形式，以确保各衣物子特征的表达形式的一致性，实现特征的统一表达。在另一些实施例中，可以将尺度较大的衣物子特征表示为矩阵的形式，尺度较小的衣物子特征表示为向量的形式，能够更为精准地对不同尺度衣物子特征进行差异性地表达，提高特征表达的准确性。

在一种可能实现方式中，如图 3 所示，该计算机设备可以将该第二衣物图像与上述步骤 402 中所获取的第二区域图像均输入衣物编码网络，由衣物编码网络基于第一衣物图像和第二区域图像执行后续的衣物特征提取步骤。其中，第二区域图像能够准确的表示出目标人体穿戴该目标衣物后的裸露区域，在衣物特征提取过程中，应用该第二区域图像，能够辅助衣物编码网络确定目标衣物与目标人体的相对位置关系，即目标衣物对目标人体的遮挡情况，使该衣物编码网络提取到更准确的衣物特征。

需要说明的是，上述对衣物特征获取方法的说明，仅是一种可能实现方式的示例性说明，本申请实施例对具体采用哪种方法获取该衣物特征不作限定。

405、计算机设备对第一人体图像中裸露区域的图像进行特征提取，得到皮肤特征。

其中，该裸露区域是该第一人体图像中的该目标人体中未被衣物遮挡的区域。在一种可能实现方式中，计算机设备可以通过皮肤区域识别网络，对该第一人体图像中的裸露区域，即皮肤区域，进行识别，得到一个二值化图像，在该二值化图像中，属于皮肤区域的像素点表示为第三数值，属于其他区域的像素点表示为第四数值。需要说明的是，本申请实施例对该皮肤区域识别网络的结构，以及，皮肤区域识别的方法不作限定。

在一种可能实现方式中，该计算机设备可以基于该二值化图像，对该第一人体图像进行图像分割，得到该第一人体图像中裸露区域的图像，将该裸露区域的图像输入图像融合模型的皮肤编码网络，通过该皮肤编码网络对该第一人体图像中裸露区域的图像进行特征提取，得到该皮肤特征。在本申请实施例中，该皮肤编码网络的结构与上述衣物编码网络的结构相同，该皮肤特征的提取过程与上述步骤 404 中衣物特征的提取过程同理。该计算机设备获取该皮肤编码网络中各个运算层所输出的不同尺度的皮肤子特征，将该多个皮肤子特征作为该皮肤特征。

406、计算机设备对第一人体图像进行特征提取，得到人体特征。

在一种可能实现方式中, 计算机设备通过该图像融合模型的人像编码网络对该第一人体图像进行特征提取, 得到该人体特征。在本申请实施例中, 该人像编码网络的结构与上述衣物编码网络的结构相同, 该人体特征的提取过程与上述步骤 404 中衣物特征的提取过程同理。该计算机设备获取人像编码网络中各个运算层所输出的不同尺度的人体子特征, 将该多个人体子特征作为该人体特征。

需要说明的是, 上述步骤 404 至步骤 406 是分别对该第二衣物图像、该第一人体图像中裸露区域的图像以及该第一人体图像进行特征提取, 得到衣物特征、皮肤特征以及人体特征的步骤。在本申请实施例中, 以先获取衣物特征、再获取皮肤特征、人像特征的执行顺序进行描述, 在一些实施例中, 也可以同时获取该衣物特征、皮肤特征和人像特征, 本申请实施例对这三种特征的获取顺序, 即上述步骤 404、步骤 405 和步骤 406 的执行顺序不作限定。在本申请实施例中, 通过任务拆解的方式, 从多个维度分别进行特征提取, 一方面, 能够提取到更全面、更细粒度的特征, 另一方面, 三个特征提取任务分别执行, 也能够降低计算机设备在特征提取过程中的运算负担, 提高设备性能。

407、计算机设备基于该衣物特征、该皮肤特征以及该人体特征, 生成第二人体图像。

在一种可能实现方式中, 计算机设备对该衣物特征、该皮肤特征以及该人体特征进行拼接, 得到拼接特征。示例性的, 该衣物特征、该皮肤特征以及该人体特征均包括至少两个尺度的子特征, 计算机设备分别对该衣物特征、该皮肤特征以及该人体特征中相同尺度的子特征进行拼接, 得到至少两个拼接子特征, 再对该至少两个拼接子特征进行拼接, 得到该拼接特征。图 7 是本申请实施例提供的一种特征拼接的示意图, 参见图 7, 相同尺度的衣物子特征 701、皮肤子特征 702 和人像子特征 703 拼接在一起, 得到拼接子特征 704, 再对各个尺度的拼接子特征 704、705、706 进行拼接, 得到拼接特征 707。需要说明的是, 上述对特征拼接方法的说明, 仅是一种示例性说明, 该计算机设备也可以通过其他方法对各个维度的特征进行拼接、融合, 本申请实施例对此不做限定。

在本申请实施例中, 计算机设备通过图像融合模型中的解码网络, 基于该拼接特征进行解码, 生成该第二人体图像。其中, 该第二人体图像中该目标人体穿戴该目标衣物。在一种可能实现方式中, 该解码网络可以采用 ResUnet 中解码网络的结构, 该解码网络包括多个级联的运算层, 计算机设备通过该多个运算层对拼接特征进行上采样, 生成该第二人体图像。图 8 是本申请实施例提供的一种解码网络的示意图, 计算机设备可以将拼接特征中不同尺度的拼接子特征, 分别输入解码网络中对应的运算层, 示例性的, 将第一尺度的拼接子特征输入第一尺度的运算层 801, 将第二尺度的拼接子特征输入第二尺度的运算层 802。在本申请实施例中, 不同尺度的拼接子特征能够保留不同粒度的细节特征, 该解码网络通过对不同尺度的拼接子特征进行解码, 所生成的第二人体图像, 能够充分保留仿射变换后的衣物的细节信息、人体皮肤信息以及人体全局细节信息, 达到真实的虚拟换装效果。

本申请实施例提供的技术方案, 通过先基于目标人体的姿态, 对原始的第一衣物图像进行变换, 得到与目标人体的姿态相贴合且保留了大量细节信息的第二衣物图像, 再基于不同特征提取任务, 分别从不同维度对第二衣物图像和第一人体图像进行特征提取, 得到衣物特征、皮肤特征以及人体特征, 也即是, 得到三个不同维度的、更细粒度的、更精确的特征, 从而基于这些特征所生成的第二人体图像能够包含更丰富的细节信息, 使第二人体图像具有较高的真实度, 确保虚拟换装效果良好。并且, 在本申请实施例中, 通过任务拆解的方式, 能够有效降低计算机设备在特征提取、图像生成过程中的运算负担, 节省运算资源开支, 提高设备性能。应用本申请实施例提供的技术方案, 在不同应用场景中, 基于不同分辨率的图像, 均能生成较为逼真的虚拟换装人像, 本方案应用于商业场景中具有良好的鲁棒性。

上述实施例中的图像融合模型为计算机设备中存储的预先训练好的模型, 该图像融合

模型为该计算机设备训练的模型，或者为其他设备训练的模型。图 9 是本申请实施例提供的一种图像融合模型的训练方法流程图，参见图 9，在一种可能实现方式中，该图像融合模型的训练方法包括以下步骤：

901、计算机设备获取第一初始模型和第二初始模型。

其中，该第一初始模型和第二初始模型的结构相同，该第一初始模型和第二初始模型的结构如图 3 所示，均为基于 CNN（Convolutional Neural Networks，卷积神经网络）所构建的模型。

902、计算机设备获取包含第一人体的第一样本人体图像、包含第一衣物的第一样本衣物图像以及包含第二衣物的第二样本衣物图像。

其中，该第一样本人体图像中的第一人体穿戴该第一衣物，该第一衣物和第二衣物不同。在一种可能实现方式中，该第一样本人体图像携带有标注信息，该标注信息用于对第一样本人体的各个身体区域进行区分，以便于模型执行后续的图像生成步骤。当然，该第一样本人体图像、第一衣物图像和第二衣物图像中还可以标注有其他信息，本申请实施例对此不作限定。

在一种可能实现方式中，计算机设备获取到样本人体图像和样本衣物图像后，可以对图像进行预处理，例如，预处理可以是调整图像尺寸，也可以进行图像增强处理等。

903、计算机设备通过该第一初始模型对该第一样本人体图像以及该第二样本衣物图像进行图像融合，得到第二样本人体图像。

在本申请实施例中，计算机设备将该第一样本人体图像和第二样本衣物图像输入第一初始模型，由该第一初始模型基于该第一人体的特征和第二衣物的特征，生成第二样本人体图像。其中，该第二样本人体图像中，该第一人体穿戴该第二衣物。该计算机设备通过该第一初始模型生成第二样本人体图像的过程，与上述步骤 402 至步骤 407 的过程同理。

904、计算机设备通过该第二初始模型对该第二样本人体图像以及该第一样本衣物图像进行图像融合，得到目标样本人体图像。

在本申请实施例中，计算机设备将该第二样本人体图像和第一衣物图像输入第二初始模型，由该第二初始模型基于第一人体的特征和第一衣物的特征，生成目标样本人体图像。其中，该目标样本人体图像中，该第一人体穿戴该第一衣物。该计算机设备通过该第二初始模型生成目标样本人体图像的过程，与上述步骤 402 至步骤 407 的过程同理。

905、计算机设备基于该第一样本人体图像和该第二样本人体图像之间的误差、该第一样本人体图像和该目标样本人体图像之间的误差，对该第一初始模型进行训练。

在本申请实施例中，计算机设备可以获取多个维度的误差。在一种可能实现方式中，计算机设备基于该第一样本人体图像和该第二样本人体图像之间的像素值分布差异，确定第一误差。其中，该第一误差用于衡量生成图像的分布一致性（Adversarial Loss），也即是，衡量模型所生成的人体图像与输入的人体图像之间的像素值分布差异，以及，模型所生成的人体图像与输入的人体图像中裸露区域的图像之间的像素值分布差异。在一种可能实现方式中，计算机设备分别获取第一样本人体图像和第二样本人体图像的像素值分布直方图，获取两个像素值分布直方图之间的误差，作为该第一误差，当然，计算机设备也可以通过其他方式获取该第一误差，本申请实施例对此不作限定。在应用实施例中，通过获取第一误差，即约束像素值分布的一致性，能够帮助模型生成更真实的人体图像，以及，更好地还原人体图像中的皮肤。

在一种可能实现方式中，计算机设备将该第一样本人体图像所对应像素值矩阵和该第二样本人体图像所对应像素值矩阵之间的误差，确定为第二误差。其中，该第二误差用于衡量模型的输入图像和输出图像之间的像素值一致性（Preserving Loss），在本申请实施例中，该第二误差能够指示模型的输入人体图像的衣服区域和胳膊区域与输出人体图像中相应区域的一致性差异。在一种可能实现方式中，计算机设备可以应用 L1 损失函数，来获

取该第一样本人体图像和第二样本人体图像之间的第二误差，当然，该计算机设备也可以通过其他方法获取该第二误差，本申请实施例对此不作限定。本申请实施例中，通过获取第二误差，即对图像之间的像素值一致性进行约束，能够使模型充分保留人体的下肢、下装、面部、头发等部位的细节信息。

在一种可能实现方式中，计算机设备将该第一样本人体图像所对应像素值矩阵和该目标样本人体图像所对应像素值矩阵之间的误差，确定为第三误差。其中，该第三误差用于表示两个模型之间的循环一致性误差，该第三误差用于衡量第一初始模型所输入的第一样本人体图像和第二初始模型的输出目标样本人体图像之间的误差。在一种可能实现方式中，计算机设备可以应用 L1 损失函数，来获取第一样本人体图像和目标样本人体图像之间的第三误差，当然，该计算机设备也可以通过其他方法获取该第三误差，本申请实施例对此不作限定。在本申请实施例中，通过获取第三误差，即通过循环一致性约束的设计，能够使第二初始模型所输出的目标样本人体图像最大限度的逼近原始的第一样本人体图像。

在一种可能实现方式中，计算机设备将该第一样本人体图像中衣物区域的图像特征和目标样本人体图像中衣物区域的图像特征之间的误差，确定为第四误差。其中，该第四误差用于指示目标样本人体图像中的衣物与原始的第一样本人体图像中的衣物之间的差异。在本申请实施例中，计算机设备从该第一样本人体图像中，分割出衣物区域的图像，得到第一分割图像，从目标样本人体图像中，分割出衣物区域的图像，得到第二分割图像，将第一分割图像和第二分割图像分别输入 VGG (Visual Geometry Group Network, 视觉几何组网络) 进行特征提取，获取 VGG 网络中每个运算层所输出的图像特征，计算机设备获取第一分割图像和第二分割图像在相同运算层上的图像特征之间的误差，即，获取运算层上的图像特征之间的欧氏距离，计算机设备基于获取到的多个欧式距离确定该第四误差，例如，对该多个欧式距离取平均值，得到该第四误差。需要说明的是，上述对第四误差获取方法的说明，仅是一种示例性说明，本申请实施例对具体采用哪种方法获取第四误差不作限定。在本申请实施例中，通过获取第四误差，即对衣物特征的一致性进行约束，能够提高模型所生成的人体图像中衣物的真实度。

在本申请实施例中，计算机设备基于该第一误差、第二误差、第三误差以及该第四误差，对该第一初始模型的模型参数进行调整。示例性的，该计算机设备分别将该第一误差、第二误差、第三误差以及第四误差，反向传播至该第一初始模型，求解该第一初始模型中各个运算层的参数，以实现对该第一初始模型中各个运算层的参数的调整。需要说明的是，上述对第一初始模型的参数进行调整的方法的说明，仅是一种示例性说明，本申请实施例对采用哪种方法调整该第一初始模型的参数不作限定。

906、计算机设备基于该第二样本人体图像和该目标样本人体图像之间的误差、该第一样本人体图像和该目标样本人体图像之间的误差，对该第二初始模型进行训练。

在一种可能实现方式中，计算机设备基于该第二样本人体图像和该目标样本人体图像之间的像素值分布差异，确定第五误差；将该第二样本人体图像所对应像素值矩阵和该目标样本人体图像所对应像素值矩阵之间的误差，确定为第六误差；将该第一样本人体图像所对应像素值矩阵和该目标样本人体图像所对应像素值矩阵之间的误差，确定为第七误差；将该第一样本衣物图像的特征和第二样本衣物图像的特征之间的误差，确定为第八误差；基于该第五误差、第六误差、第七误差以及该第八误差，对该第二初始模型的模型参数进行调整。需要说明的是，该第五误差、第六误差、第七误差、第八误差的获取过程分别与上述第一误差、第二误差、第三误差、第四误差的获取过程同理。在本申请实施例中，计算机设备可以将该第五误差、第六误差、第七误差以及第八误差，反向传播至该第二初始模型，求解该第二初始模型中各个运算层的参数。需要说明的是，本申请实施例对采用哪种方法调整该第二初始模型的参数不作限定。

907、计算机设备响应于该第一初始模型和该第二初始模型均满足模型训练条件，获

取训练完成的该第二初始模型，作为图像融合模型。

其中，该模型训练条件由开发人员进行设置，本申请实施例对此不作限定。示例性的，该模型训练条件包括训练次数阈值，若模型训练次数达到该训练次数阈值，则获取训练完成的第二初始模型，作为该图像融合模型；若模型训练次数未达到该训练次数阈值，则继续获取下一组样本图像对该第一初始模型和第二初始模型进行训练。示例性的，该模型训练条件包括误差阈值，若模型输出结果对应的误差值均小于该误差阈值，且，模型输出结果对应的误差值均小于该误差阈值的次数达到目标次数，则确定该第一初始模型和该第二初始模型均满足模型训练条件，获取训练完成的第二初始模型，作为图像融合模型；否则，继续获取下一组样本图像进行模型训练。

在本申请实施例中，针对虚拟换装训练数据的不成对性，引入了循环一致性的自监督模型训练方法，图 10 是本申请实施例提供的一种模型训练方法的示意图，以下结合图 10 对上述模型训练过程进行说明，如图 10 所示，计算机设备将第一样本人体图像 I_1 和第二样本衣物图像 C_2 输入第一初始模型，由第一初始模型输出中间态的虚拟换装结果，即第二样本人体图像 I_2 ，计算机设备再将第二样本人体图像 I_2 和第一样本衣物图像 C_1 输入第二初始模型，获取第二初始模型输出的目标样本人体图像 $\bar{I}_1$ ，计算机设备通过控制 $\bar{I}_1 \approx I_1$ ，

即使目标样本人体图像与第一样本人体图像趋于一致，从而实现循环一致性的自监督模型训练。且，在本申请实施例中，在基于原始的人体图像和原始的衣物图像生成新的人体图像时，通过任务拆解，将虚拟换装任务拆解成衣物仿射变换、人体皮肤生成以及人体全局信息合成这三个任务，对各个子任务分别设置相应的编码网络，分别进行不同维度的特征提取，一方面，能够获取到更全面、更细粒度的特征，另一方面，三个任务分别单独执行，能够降低计算机设备在虚拟换装任务中的运算量负担，提升计算机设备在虚拟换装任务中的设备性能。在损失函数设置方面，本申请实施例中，提出了 4 种损失函数，获取四个维度的误差值，使训练出的图像融合模型能够自适应的保留相关部位的特征，例如人体下肢、面部、头发等部位得到特征，使图像融合模型能够自适应的生成被衣服掩盖的人体皮肤，例如在长袖上衣换成短袖上衣时，准确的构造出人体的胳膊区域的图像，使模型能够达到照片级别的虚拟换装效果。

上述所有可选技术方案，可以采用任意结合形成本申请的可选实施例。

图 11 是本申请实施例提供的一种图像生成装置的结构示意图，参见图 11，该装置包括：

第一获取模块 1101，用于获取包含目标人体的第一人体图像以及包含目标衣物的第一衣物图像；

衣物变换模块 1102，用于基于该第一人体图像中该目标人体的姿态，对该第一衣物图像进行变换，得到第二衣物图像，该第二衣物图像包含该目标衣物，该目标衣物的姿态与该目标人体的姿态相匹配；

特征提取模块 1103，用于分别对该第二衣物图像、该第一人体图像中裸露区域的图像以及该第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，该裸露区域是该第一人体图像中的该目标人体中未被衣物遮挡的区域；

第一生成模块 1104，用于基于该衣物特征、该皮肤特征以及该人体特征，生成第二人体图像，该第二人体图像中该目标人体穿戴该目标衣物。

在一种可能实现方式中，该衣物变换模块 1102 包括：

区域确定子模块，用于基于该第一人体图像中该目标人体的姿态以及该第一衣物图像，确定第一区域图像，该第一区域图像用于指示该目标人体穿戴该目标衣物时，该目标衣物的遮挡区域；

衣物变换子模块，用于基于该第一区域图像，对该第一衣物图像进行变换，得到该第

二衣物图像。

在一种可能实现方式中，该区域确定子模块包括：

区域识别单元，用于对该第一人体图像进行人体区域识别，将该第一人体图像中属于相同身体区域的像素点设置为相同像素值，得到人体语义图像，该人体语义图像用于指示该目标人体的姿态；

区域确定单元，用于基于该人体语义图像以及该第一衣物图像，确定该第一区域图像。

在一种可能实现方式中，该区域确定单元，用于将该人体语义图像以及该第一衣物图像输入图像融合模型的衣物区域预测网络；通过该衣物区域预测网络，提取该人体语义图像对应的人体姿态特征以及该第一衣物图像中该目标衣物的轮廓特征，基于该人体姿态特征以及该目标衣物的轮廓特征，确定该第一区域图像。

在一种可能实现方式中，该衣物变换子模块，还用于将该第一区域图像以及该第一衣物图像输入图像融合模型的空间变换网络，该空间变换网络用于对图像进行变换；通过该空间变换网络基于该第一衣物图像以及该第一掩模区域之间的映射关系，对该第一衣物图像进行变换，得到中间衣物图像；通过该空间变换网络确定该中间衣物图像所对应的纹理特征，基于该纹理特征以及该中间衣物图像，生成该第二衣物图像。

在一种可能实现方式中，该衣物变换子模块，还用于基于第一掩模区域确定目标衣物的初始轮廓，并基于第一区域图像确定目标衣物在变换后对应的目标轮廓；基于初始轮廓以及目标轮廓，确定初始轮廓和目标轮廓之间的映射关系；

基于映射关系，对第一衣物图像进行变换，得到中间衣物图像。

在一种可能实现方式中，空间变换网络包括衣物生成子网络；该衣物变换子模块，还用于基于衣物生成子网络，按照预先学习的衣物细节特征，对中间衣物图像进行特征提取，得到中间衣物图像对应的纹理特征。

在一种可能实现方式中，该特征提取模块 1103，还用于通过图像融合模型的衣物编码网络对该第二衣物图像进行特征提取，得到该衣物特征；通过该图像融合模型的皮肤编码网络对该第一人体图像中裸露区域的图进行特征提取，得到该皮肤特征；通过该图像融合模型的人像编码网络对该第一人体图像进行特征提取，得到该人体特征。

在一种可能实现方式中，该第一生成模块 1104 包括：

特征拼接子模块，用于对该衣物特征、该皮肤特征以及该人体特征进行拼接，得到拼接特征；

图像生成子模块，用于基于该拼接特征进行解码，生成该第二人体图像。

在一种可能实现方式中，该衣物特征、该皮肤特征以及该人体特征均包括至少两个尺度的子特征；

该特征拼接子模块，用于分别对该衣物特征、该皮肤特征以及该人体特征中相同尺度的子特征进行拼接，得到至少两个拼接子特征；对该至少两个拼接子特征进行拼接，得到该拼接特征。

在一种可能实现方式中，该第二衣物图像的获取步骤、该特征提取步骤以及该第二人体图像的生成步骤通过图像融合模型执行。

在一种可能实现方式中，该装置还包括：

第二获取模块，用于获取第一初始模型和第二初始模型，该第一初始模型和第二初始模型的结构相同；

第三获取模块，用于获取包含第一人体的第一样本人体图像、包含第一衣物的第一样本衣物图像以及包含第二衣物的第二样本衣物图像，该第一样本人体图像中该第一人体穿戴该第一衣物；

第二生成模块，用于通过该第一初始模型对该第一样本人体图像以及该第二样本衣物图像进行图像融合，生成第二样本人体图像，该第二样本人体图像中，该第一人体穿戴该

第二衣物；

第三生成模块，用于通过该第二初始模型对该第二样本人体图像以及该第一样本衣物图像进行图像融合，生成目标样本人体图像，该目标样本人体图像中，该第一人体穿戴该第一衣物；

第一训练模块，用于基于该第一样本人体图像和该第二样本人体图像之间的误差、该第一样本人体图像和该目标样本人体图像之间的误差，对该第一初始模型进行训练；

第二训练模块，用于基于该第二样本人体图像和该目标样本人体图像之间的误差、该第一样本人体图像和该目标样本人体图像之间的误差，对该第二初始模型进行训练；

第四获取模块，用于响应于该第一初始模型和该第二初始模型均满足模型训练条件，获取训练完成的该第二初始模型，作为图像融合模型。

在一种可能实现方式中，该第一训练模块，用于基于该第一样本人体图像和该第二样本人体图像之间的像素值分布差异，确定第一误差；将该第一样本人体图像所对应像素值矩阵和该第二样本人体图像所对应像素值矩阵之间的误差，确定为第二误差；将该第一样本人体图像所对应像素值矩阵和该目标样本人体图像所对应像素值矩阵之间的误差，确定为第三误差；将该第一样本人体图像中衣物区域的图像特征和目标样本人体图像中衣物区域的图像特征之间的误差，确定为第四误差；基于该第一误差、第二误差、第三误差以及该第四误差，对该第一初始模型的模型参数进行调整。

在一种可能实现方式中，该第一训练模块，还用于将第一误差第二误差、第三误差以及第四误差，反向传播至第一初始模型，求解第一初始模型中各个运算层的参数；基于求解结果，对第一初始模型中各个运算层的参数进行调整。

在一种可能实现方式中，该第二训练模块，用于基于该第二样本人体图像和该目标样本人体图像之间的像素值分布差异，确定第五误差；将该第二样本人体图像所对应像素值矩阵和该目标样本人体图像所对应像素值矩阵之间的误差，确定为第六误差；将该第一样本人体图像所对应像素值矩阵和该目标样本人体图像所对应像素值矩阵之间的误差，确定为第七误差；将该第一样本人体图像中衣物区域的图像特征和目标样本人体图像中衣物区域的图像特征之间的误差，确定为第八误差；基于该第五误差、第六误差、第七误差以及该第八误差，对该第二初始模型的模型参数进行调整。

本申请实施例提供的装置，通过先基于目标人体的姿态，对原始的第一衣物图像进行变换，得到与目标人体的姿态相贴合且保留了大量细节信息的第二衣物图像，再基于不同特征提取任务，分别从不同维度对第二衣物图像和第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，也即是，得到三个不同维度的、更细粒度的、更精确的特征，从而基于这些特征所生成的第二人体图像能够包含更丰富的细节信息，使第二人体图像具有较高的真实度，确保虚拟换装效果良好。

需要说明的是：上述实施例提供的图像生成装置在图像生成时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的图像生成装置与图像生成方法实施例属于同一构思，其具体实现过程详见方法实施例。

上述技术方案所提供的计算机设备可以实现为终端或服务器，例如，图 12 是本申请实施例提供的一种终端的结构示意图。该终端 1200 可以是：智能手机、平板电脑、MP3 播放器(Moving Picture Experts Group Audio Layer III, 动态影像专家压缩标准音频层面 3)、MP4 (Moving Picture Experts Group Audio Layer IV, 动态影像专家压缩标准音频层面 4) 播放器、笔记本电脑或台式电脑。终端 1200 还可能被称为用户设备、便携式终端、膝上型终端、台式终端等其他名称。

通常，终端 1200 包括有：一个或多个处理器 1201 和一个或多个存储器 1202。处理器

1201 可以包括一个或多个处理核心，比如 4 核心处理器、8 核心处理器等。处理器 1201 可以采用 DSP (Digital Signal Processing, 数字信号处理)、FPGA (Field-Programmable Gate Array, 现场可编程门阵列)、PLA (Programmable Logic Array, 可编程逻辑阵列) 中的至少一种硬件形式来实现。

存储器 1202 可以包括一个或多个计算机可读存储介质，该计算机可读存储介质可以是非暂态的。存储器 1202 还可包括高速随机存取存储器，以及非易失性存储器，比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中，存储器 1202 中的非暂态的计算机可读存储介质用于存储至少一条计算机可读指令，该至少一条计算机可读指令用于被处理器 1201 所执行以实现本申请中方法实施例提供的图像生成方法。

在一些实施例中，终端 1200 还可选包括有：外围设备接口 1203 和至少一个外围设备。处理器 1201、存储器 1202 和外围设备接口 1203 之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与外围设备接口 1203 相连。具体地，外围设备包括：射频电路 1204、显示屏 1205、摄像头组件 1206、音频电路 1207、定位组件 1208 和电源 1209 中的至少一种。

外围设备接口 1203 可被用于将 I/O (Input/Output, 输入/输出) 相关的至少一个外围设备连接到处理器 1201 和存储器 1202。在一些实施例中，处理器 1201、存储器 1202 和外围设备接口 1203 被集成在同一芯片或电路板上；在一些其他实施例中，处理器 1201、存储器 1202 和外围设备接口 1203 中的任意一个或两个可以在单独的芯片或电路板上实现，本实施例对此不加以限定。

射频电路 1204 用于接收和发射 RF (Radio Frequency, 射频) 信号，也称电磁信号。射频电路 1204 通过电磁信号与通信网络以及其他通信设备进行通信。显示屏 1205 用于显示 UI (User Interface, 用户界面)。该 UI 可以包括图形、文本、图标、视频及其它们的任意组合。摄像头组件 1206 用于采集图像或视频。音频电路 1207 可以包括麦克风和扬声器中的至少一种。定位组件 1208 用于定位终端 1200 的当前地理位置。电源 1209 用于为终端 1200 中的各个组件进行供电。

在一些实施例中，终端 1200 还包括有一个或多个传感器 1210。该一个或多个传感器 1210 包括但不限于：加速度传感器 1211、陀螺仪传感器 1212、压力传感器 1213、指纹传感器 1214、光学传感器 1215 以及接近传感器 1216。

本领域技术人员可以理解，图 12 中示出的结构并不构成对终端 1200 的限定，可以包括比图示更多或更少的组件，或者组合某些组件，或者采用不同的组件布置。

图 13 是本申请实施例提供的一种服务器的结构示意图，该服务器 1300 可因配置或性能不同而产生比较大的差异，可以包括一个或多个处理器 (Central Processing Units, CPU) 1301 和一个或多个的存储器 1302，其中，该一个或多个存储器 1302 中存储有至少一条计算机可读指令，该至少一条计算机可读指令由该一个或多个处理器 1301 加载并执行以实现上述各个方法实施例提供的方法。当然，该服务器 1300 还可以具有有线或无线网络接口、键盘以及输入输出接口等部件，以便进行输入输出，该服务器 1300 还可以包括其他用于实现设备功能的部件。

在示例性实施例中，还提供了一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，例如包括至少一条计算机可读指令的存储器，上述至少一条计算机可读指令可由处理器执行以完成上述实施例中的图像生成方法。例如，该计算机可读存储介质可以是只读存储器 (Read-Only Memory, ROM)、随机存取存储器 (Random Access Memory, RAM)、只读光盘 (Compact Disc Read-Only Memory, CD-ROM)、磁带、软盘和光数据存储设备等。

在示例性实施例中，还提供了一种计算机程序产品，该计算机程序产品包括至少一条计算机可读指令，该至少一条计算机可读指令存储在计算机可读存储介质中。计算机设备

的处理器从计算机可读存储介质读取该至少一条计算机可读指令，处理器执行该至少一条计算机可读指令，使得该计算机设备实现该图像生成方法所执行的操作。

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步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，该程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

上述仅为本申请的可选实施例，并不用以限制本申请，凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

权利要求书

1、一种图像生成方法，其特征在于，由计算机设备执行，所述方法包括：

获取包含目标人体的第一人体图像以及包含目标衣物的第一衣物图像；

基于所述第一人体图像中所述目标人体的姿态，对所述第一衣物图像进行变换，得到第二衣物图像，所述第二衣物图像包含所述目标衣物，所述目标衣物的姿态与所述目标人体的姿态相匹配；

分别对所述第二衣物图像、所述第一人体图像中裸露区域的图像以及所述第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，所述裸露区域是所述第一人体图像中的所述目标人体中未被衣物遮挡的区域；及

基于所述衣物特征、所述皮肤特征以及所述人体特征，生成第二人体图像，所述第二人体图像中所述目标人体穿戴所述目标衣物。

2、根据权利要求 1 所述的方法，其特征在于，所述基于所述第一人体图像中所述目标人体的姿态，对所述第一衣物图像进行变换，得到第二衣物图像，包括：

基于所述第一人体图像中所述目标人体的姿态以及所述第一衣物图像，确定第一区域图像，所述第一区域图像用于指示所述目标人体穿戴所述目标衣物时，所述目标衣物的遮挡区域；及

基于所述第一区域图像，对所述第一衣物图像进行变换，得到所述第二衣物图像。

3、根据权利要求 2 所述的方法，其特征在于，所述基于所述第一人体图像中所述目标人体的姿态以及所述第一衣物图像，确定第一区域图像，包括：

对所述第一人体图像进行人体区域识别，将所述第一人体图像中属于相同身体区域的像素点设置为相同像素值，得到人体语义图像，所述人体语义图像用于指示所述目标人体的姿态；及

基于所述人体语义图像以及所述第一衣物图像，确定所述第一区域图像。

4、根据权利要求 3 所述的方法，其特征在于，所述基于所述人体语义图像以及所述第一衣物图像，确定所述第一区域图像，包括：

将所述人体语义图像以及所述第一衣物图像输入图像融合模型的衣物区域预测网络；

通过所述衣物区域预测网络，提取所述人体语义图像对应的人体姿态特征以及所述第一衣物图像中所述目标衣物的轮廓特征；及

基于所述人体姿态特征以及所述目标衣物的轮廓特征，确定所述第一区域图像。

5、根据权利要求 2 所述的方法，其特征在于，所述基于所述第一区域图像，对所述第一衣物图像进行变换，得到所述第二衣物图像，包括：

将所述第一区域图像以及所述第一衣物图像输入图像融合模型的空间变换网络，所述空间变换网络用于对图像进行变换；

通过所述空间变换网络基于所述第一衣物图像以及第一掩模区域之间的映射关系，对所述第一衣物图像进行变换，得到中间衣物图像，所述第一掩模区域用于指示所述第一衣物图像中的目标衣物所在的区域；

通过所述空间变换网络，确定所述中间衣物图像所对应的纹理特征；及

基于所述纹理特征以及所述中间衣物图像，生成所述第二衣物图像。

6、根据权利要求 5 所述的方法，其特征在于，所述通过所述空间变换网络基于所述第一衣物图像以及第一掩模区域之间的映射关系，对所述第一衣物图像进行变换，得到中间衣物图像，包括：

基于所述第一掩模区域确定所述目标衣物的初始轮廓，并基于所述第一区域图像确定所述目标衣物在变换后对应的目标轮廓；

基于所述初始轮廓以及所述目标轮廓，确定所述初始轮廓和所述目标轮廓之间的映射关系；

基于所述映射关系，对所述第一衣物图像进行变换，得到中间衣物图像。

7、根据权利要求5所述的方法，其特征在于，所述空间变换网络包括衣物生成子网络；

所述通过所述空间变换网络，确定所述中间衣物图像所对应的纹理特征，包括：

基于所述衣物生成子网络，按照预先学习的衣物细节特征，对所述中间衣物图像进行特征提取，得到所述中间衣物图像对应的纹理特征。

8、根据权利要求1所述的方法，其特征在于，所述分别对所述第二衣物图像、所述第一人体图像中裸露区域的图像以及所述第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，包括：

通过图像融合模型的衣物编码网络对所述第二衣物图像进行特征提取，得到所述衣物特征；

通过所述图像融合模型的皮肤编码网络对所述第一人体图像中裸露区域的图进行特征提取，得到所述皮肤特征；及

通过所述图像融合模型的人像编码网络对所述第一人体图像进行特征提取，得到所述人体特征。

9、根据权利要求1所述的方法，其特征在于，所述基于所述衣物特征、所述皮肤特征以及所述人体特征，生成第二人体图像，包括：

对所述衣物特征、所述皮肤特征以及所述人体特征进行拼接，得到拼接特征；及

基于所述拼接特征进行解码，生成所述第二人体图像。

10、根据权利要求9所述的方法，其特征在于，所述衣物特征、所述皮肤特征以及所述人体特征均包括至少两个尺度的子特征；

所述对所述衣物特征、所述皮肤特征以及所述人体特征进行拼接，得到拼接特征，包括：

分别对所述衣物特征、所述皮肤特征以及所述人体特征中相同尺度的子特征进行拼接，得到至少两个拼接子特征；及

对所述至少两个拼接子特征进行拼接，得到所述拼接特征。

11、根据权利要求1所述的方法，其特征在于，所述第二衣物图像的获取步骤、所述特征提取步骤以及所述第二人体图像的生成步骤通过图像融合模型执行。

12、根据权利要求11所述的方法，其特征在于，所述方法还包括：

获取第一初始模型和第二初始模型，所述第一初始模型和第二初始模型的结构相同；

获取包含第一人体的第一样本人体图像、包含第一衣物的第一样本衣物图像以及包含第二衣物的第二样本衣物图像，所述第一样本人体图像中所述第一人体穿戴所述第一衣物；

通过所述第一初始模型对所述第一样本人体图像以及所述第二样本衣物图像进行图像融合，生成第二样本人体图像，所述第二样本人体图像中，所述第一人体穿戴所述第二衣物；

通过所述第二初始模型对所述第二样本人体图像以及所述第一样本衣物图像进行图像融合，生成目标样本人体图像，所述目标样本人体图像中，所述第一人体穿戴所述第一衣物；

基于所述第一样本人体图像和所述第二样本人体图像之间的误差、所述第一样本人体图像和所述目标样本人体图像之间的误差，对所述第一初始模型进行训练；

基于所述第二样本人体图像和所述目标样本人体图像之间的误差、所述第一样本人体图像和所述目标样本人体图像之间的误差，对所述第二初始模型进行训练；及

响应于所述第一初始模型和所述第二初始模型均满足模型训练条件的事件，获取训练完成的所述第二初始模型，作为图像融合模型。

13、根据权利要求12所述的方法，其特征在于，所述基于所述第一样本人体图像和

所述第二样本人体图像之间的误差、所述第一样本人体图像和所述目标样本人体图像之间的误差，对所述第一初始模型进行训练，包括：

基于所述第一样本人体图像和所述第二样本人体图像之间的像素值分布差异，确定第一误差；

将所述第一样本人体图像所对应像素值矩阵和所述第二样本人体图像所对应像素值矩阵之间的误差，确定为第二误差；

将所述第一样本人体图像所对应像素值矩阵和所述目标样本人体图像所对应像素值矩阵之间的误差，确定为第三误差；

将所述第一样本人体图像中衣物区域的图像特征和目标样本人体图像中衣物区域的图像特征之间的误差，确定为第四误差；及

基于所述第一误差、所述第二误差、所述第三误差以及所述第四误差，对所述第一初始模型的模型参数进行调整。

14、根据权利要求13所述的方法，其特征在于，所述基于所述第一误差、所述第二误差、所述第三误差以及所述第四误差，对所述第一初始模型的模型参数进行调整，包括：

将所述第一误差所述第二误差、所述第三误差以及所述第四误差，反向传播至所述第一初始模型，求解所述第一初始模型中各个运算层的参数；

基于求解结果，对所述第一初始模型中各个运算层的参数进行调整。

15、根据权利要求12所述的方法，其特征在于，所述基于所述第二样本人体图像和所述目标样本人体图像之间的误差、所述第一样本人体图像和所述目标样本人体图像之间的误差，对所述第二初始模型进行训练，包括：

基于所述第二样本人体图像和所述目标样本人体图像之间的像素值分布差异，确定第五误差；

将所述第二样本人体图像所对应像素值矩阵和所述目标样本人体图像所对应像素值矩阵之间的误差，确定为第六误差；

将所述第一样本人体图像所对应像素值矩阵和所述目标样本人体图像所对应像素值矩阵之间的误差，确定为第七误差；

将所述第一样本人体图像中衣物区域的图像特征和目标样本人体图像中衣物区域的图像特征之间的误差，确定为第八误差；及

基于所述第五误差、所述第六误差、所述第七误差以及所述第八误差，对所述第二初始模型的模型参数进行调整。

16、一种图像生成装置，其特征在于，所述装置包括：

第一获取模块，用于获取包含目标人体的第一人体图像以及包含目标衣物的第一衣物图像；

衣物变换模块，用于基于所述第一人体图像中所述目标人体的姿态，对所述第一衣物图像进行变换，得到第二衣物图像，所述第二衣物图像包含所述目标衣物，所述目标衣物的姿态与所述目标人体的姿态相匹配；

特征提取模块，用于分别对所述第二衣物图像、所述第一人体图像中裸露区域的图像以及所述第一人体图像进行特征提取，得到衣物特征、皮肤特征以及人体特征，所述裸露区域是所述第一人体图像中的所述目标人体中未被衣物遮挡的区域；及

第一生成模块，用于基于所述衣物特征、所述皮肤特征以及所述人体特征，生成第二人体图像，所述第二人体图像中所述目标人体穿戴所述目标衣物。

17、一种计算机设备，包括存储器和一个或多个处理器，所述存储器存储有计算机可读指令，其特征在于，所述一个或多个处理器执行所述计算机可读指令时实现权利要求1至15中任一项所述的方法的步骤。

18、一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，存储有计算

机可读指令，其特征在于，所述计算机可读指令被一个或多个处理器执行时实现权利要求 1 至 15 中任一项所述的方法的步骤。

19、一种计算机程序产品，包括计算机可读指令，其特征在于，所述计算机可读指令被一个或多个处理器执行时实现权利要求 1 至 15 中任一项所述的方法的步骤。