ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 37,No. 5,Sep. 2022,pp. 1145－1156
DOI:10. 16337/j. 1004-9037. 2022. 05. 019

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

面向学位服照片生成的虚拟试衣方法

盛培卓[1]，李婷玉[1]，李天宝[2]，宋　丹[2]，刘安安[2]

(1. 天津大学智能与计算学部,天津 300354; 2. 天津大学电气自动化与信息工程学院,天津 300072)

摘　要: 为了解决现有虚拟试衣方法不能适用于学位服的问题,提出一种面向学位服照片生成的虚拟试衣方法。该方法首先对由服装变形模块和虚拟试穿模块构成的基于图像的虚拟试衣网络进行训练,将人像与学位服图像通过训练后的网络生成试衣结果。随后,将生成的学位服试衣结果通过背景融合模块与特定背景进行合成。实验过程中,本文构建了一个新的学位服与长裙的数据集。从实验结果来看,本文提出的算法能够在很大程度上减少原人像中衣服对学位服试穿的影响,能够较好地完成学位服的试穿工作并生成较为理想的试穿结果。

关键词: 虚拟试衣;学位服;人物特征保持;图像生成;背景融合

中图分类号: TP391　　**文献标志码:** A

Virtual Try-on Network for Graduation Photo Generation

SHENG Peizhuo[1], LI Tingyu[1], LI Tianbao[2], SONG Dan[2], LIU An'an[2]

(1. College of Intelligence and Computing, Tianjin University, Tianjin 300354, China; 2. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: In order to solve the problem that the existing virtual fitting methods cannot be applied to academic uniforms, a virtual try-on method oriented to the generation of academic uniforms is proposed. The method first trains the image-based virtual try-on network composed of the clothing deformation module and the virtual try-on module, and then generates try-on results through the trained network of the portrait and the academic dress image. Then, the generated academic dress try-on results are synthesized with the specific background through the background fusion module. During the experiment, this paper constructs a new dataset of academic dress and long skirt. From the experimental results, the algorithm proposed in this paper can greatly reduce the impact of the clothes in the original portrait on the academic dress try-on, and can better complete the academic dress try-on work and generate more ideal fitting results.
Key words: virtual try-on; academic dress; character retention; image generation; background fusion

引　　言

面向学位服照片生成的虚拟试衣方法旨在根据用户给定的人像以及所选的学位服种类生成对应穿着学位服的人像图片。该方法可以使学生足不出户拍摄学位毕业照,减少疫情期间学生的大规模聚

基金项目: 国家自然科学基金(61902277,61772359);天津市新一代人工智能重大专项(19ZXZNGX00110,18ZXZNGX00150);中国博士后科学基金(2020M680884);天津大学大学生创新创业训练计划(202010056483)。
收稿日期: 2021-05-23;**修订日期:** 2021-11-04

集,也可以帮助不在校的学生远程拍摄学位毕业照,并且可以使毕业照更加多样化。

目前最主流的虚拟试衣方法主要分为两种:一种是基于物理仿真的三维虚拟试衣技术;另一种是基于图像生成的二维虚拟试衣技术。早期研究中,与虚拟试衣相关的工作主要采用三维测量和构建三维服装模型的方法。Guan等[1]于2012年提出了针对所有人试穿(Dressing any person, DRAPE)的服装模型,该模型可以将因身体形状引起的衣服变形和因姿势变化引起的衣服变形进行"分解",这种"分解"提供了一个近似的物理服装变形,大大简化了服装的合成。随后,Hahn等[2]于2014年提出了一种基于自适应基的子空间服装仿真技术,可以只用几个基础向量再现出多种多样并具有细节的褶皱模式,更加方便地模拟了服装的变形。但是,上述两种方法的计算量非常大,因此很难得到广泛的应用。随后,随着神经网络与深度学习的发展,学者们开始利用生成对抗网络(Generative adversarial networks, GAN)来解决虚拟试穿问题。Jetchev等[3]于2017年提出了条件模拟生成对抗网络(Conditional analogy GAN, CAGAN),这是一种基于U-Net的生成对抗网络[4],这种网络无法处理较大的空间变形,因此利用这种方法无法产生逼真的结果。近年来,基于图像生成的二维虚拟试衣技术得到了众多研究者的关注。2018年,Han等[5]提出了基于图像的虚拟试穿网络(Image-based virtual try-on network, VITON),该方法首先使用形状上下文匹配算法[6]对衣服进行变形处理,然后使用U-Net生成器学习图像合成,将扭曲后的衣服与人体图像进行合成。2018年,Wang等[7]提出了面向特征保留的基于图像的虚拟试穿网络(Characteristic-preserving virtual try-on network, CP-VTON)。与VITON[5]相比,CP-VTON[7]引入了几何匹配模块[8],使得生成的试衣结果与VITON[5]相比更加自然,纹理更加清晰。Song等[9]提出的面向形体保持的基于图像的虚拟试穿网络(Shape-preserving image-based virtual try-on network, SP-VITON)针对形体提取问题在此基础上做了改进,通过引入密集人体姿态估计(Dense human pose estimation, DensePose)模型[10]透过原始图像中的服装来提取出较为真实的人体形体,很大程度上消除了原始服装对试衣结果产生的影响,进一步优化了虚拟试衣结果。

但是,上述方法都不太适合直接进行学位服的虚拟试衣工作,因为在之前提出的虚拟试衣技术中,大多是针对短袖、衬衫等服装,尚未有工作针对学位服这类服装的试穿进行研究。在试穿短袖这类服装时,试穿结果仅会受到人体上半身的影响。而在学位服的试穿中,不仅是人体上半身,下半身的形体及姿势也会对试穿结果造成影响。除此之外,学位服作为宽松型的裙装,在试衣时要求将人的上半身与下半身作为一个整体,因此上半身与下半身之间的姿势动作等均会相互影响,例如当人张开双臂时学位服的下摆也会随之变得相对宽大。同时,人体的腰部、胯部等作为人体上半身与下半身的连接处,其特征也会对试衣结果造成一定影响。但是,学位服宽松的特性也要求其在试穿时并不需要紧紧贴合人体的轮廓,这就使得不同体型人的试衣结果差距不会太大,并且一些不明显的人体姿势及形体特征并不会对试衣结果产生影响。另一方面,上述方法所采用的虚拟试衣数据集与学位服相差甚远,不能用于训练学位服的虚拟试衣网络。

为了解决上述问题,本文提出面向学位服照片生成的虚拟试衣方法,并取得较为理想的效果。本文的主要工作和创新点如下:

(1)分析了学位服等长裙类衣物数据,根据学位服的连体性等特点以及人物穿着学位服后的特点,有效提取了穿着学位服人物的人体特征,并设计了由服装变形模块和虚拟试穿模块构成的虚拟试衣网络,进行学位服虚拟试衣任务。

(2)在虚拟试衣的任务基础上,增加了背景融合模块,有助于生成更真实的毕业场景照片。

(3)首次收集并提出了学位服虚拟试衣数据集,该数据集弥补了虚拟试衣研究中长裙类数据集的空白,有助于推动相关工作的研究和发展。

1 基于图像生成的虚拟试衣方法

近年来,随着信息技术的发展和计算机运算能力的提高,深度学习方法和深度学习模型在各个领域取得了越来越好的应用效果[11]。将深度学习的方法应用于虚拟试衣任务,通过训练模型生成对应的虚拟试衣人物图像,可以实现基于图像生成的虚拟试衣方法。2017年,Zhu等[12]提出了一种利用生成对抗网络在穿着者身上生成指定服饰物品的方法,将复杂的图像生成过程分为两个阶段来确保生成对抗网络结构的整体连贯性:在第1阶段中通过有效的空间约束条件来指导生成合理的语义分割结果[13];在第2阶段中,使用具有成分映射层的生成模型来渲染最终图像,该网络的提出也为采用深度学习方法进行虚拟试衣任务提供了很好的思路。随后,VITON[5]和CP-VTON[7]网络的提出使得以特定的人物图像和衣物图像作为输入合成对应的虚拟试穿图像成为可能。VITON网络[5]采用了与服装无关的人物特征表示,其中人物特征包括人物姿势、人物身材以及人物身份信息(面部和头发)。通过这种与服装无关的人物特征表示可以最大化地摆脱原始衣物对人像信息的干扰,保留更多的任务形象信息,从而利用这些信息与给定衣物图像进行合成,最终生成更逼真的虚拟试衣图像。CP-VTON网络[7]在VITON网络[5]的基础上增加了几何匹配模块[8],通过对几何匹配模块进行训练使网络学习薄板样条变换,从而使衣服能够更好地与人体部位进行对齐,并且在试衣结果中更自然地保留衣服上原本的图案与纹理。但是,VITON[5]与CP-VTON[7]网络在提取原始图像中人体形体时仍然存在一定的缺陷,这是由于它们在对人物形体进行提取时只是对人体解析结果的身体分割部分进行下采样操作,即将穿着原始衣服的人体下采样到低分辨率来表示人体的形体,这样生成的形体提取结果会在很大程度上受到原始衣服的影响。当原始图像中的衣服较为宽松肥大时,可能提取到的身体形体也随之变得肥胖。针对形体提取的问题,SP-VITON网络[9]对数据集进行处理,通过对数据集中图像大小进行调整来增强人物身材变化。在此基础上,通过用基于稠密的人类姿势估计方法的DensePose模型[10]将每个像素映射到一个密集的姿势点集合,以此消除衣物的干扰,进而提取衣服遮挡下的人物身材信息,从而能够更准确地刻画人体信息。SP-VITON网络[9]对于人体体型提取方法的改进使得虚拟试穿技术可以应用于更多的服装类型。2019年,Dong等[14]提出了面向多姿态引导的虚拟试穿网络(Multi-pose guided virtual try-on network, MG-VTON),将虚拟试衣扩展到任意姿势下,该方法可以在不同姿势下将衣服转移到人像上。MG-VTON网络[14]分为3个阶段:首先合成所需的目标图像的人体解析图,以匹配所需的姿势和衣服形状;随后采用深度变形的生成对抗网络将衣服扭曲并与人体解析图进行合成;最后使用多姿势合成掩膜的渲染精化衣服上的纹理细节并去除伪影,使得合成效果更加逼真。

上述方法都是针对目标衣物进行相应的变形并采用一定方法将变形后的衣物与人体相拟合达到试衣效果。除此之外,Wu等[15]提出了一种新颖的M2E(Model to everyone)虚拟试穿网络,不需要单独的目标衣服图像,可以将模型图像中的衣服转移到待试穿人像上。此网络中包含3个关键模块,即姿势对齐网络(Pyramid attention network, PAN)、纹理细化网络(Texture refinement network, TRN)以及空间变换网络(Spatial transformer network, FTN)。同时,文献[15]以自监督的方式训练了算法框架,逐步将模型图像中的衣服转移到了人像中,解决了收集训练集困难的问题。Yoo等[16]于2016年提出了一种图像生成模型,该模型利用像素级别的域区分转移方法可以从人物图像中解析服装图像,在采用生成对抗网络中的判别器的同时还引入了一种新颖的域判别器用来使生成的图像与输入图像相关。另一种以人为中心的图像生成方法是姿势引导的人物图像生成,其可以基于单个姿势合成新的人物图像。典型方法是Neverova等在2018年提出的DensePose转换[17]和Dong等同年提出的软门控变形生成对抗网络(Soft-gated warping-GAN, SGWG)[18]。除了对于单个图像的虚拟试衣技术外,还有一些工作致力于实现基于视频的虚拟试穿。Dong等[19]于2019年提出了流指导变形的生成对抗网络

(Flow-navigated warping GAN, FW-GAN),该系统可以实现将衣服转移到人身上并且生成以任意姿势为条件的视觉逼真视频。

2 本文方法

2.1 方法框架

受第1节相关工作的启发,根据学位服的特点,本文提出一种面向学位服照片生成的虚拟试衣方法,如图1所示。本文方法以原始输入图像I_i和目标学位服图像c作为输入,目标为生成学位服人像照片$\hat{I}$。$\hat{I}$中除衣服以外的其他特征(如人物姿势、人物身材以及人物身份信息)都与原始输入图像I_i中一致,并且$\hat{I}$中的人物所穿衣服替换为c中图像,并且原始图像I_i中的衣服信息不对$\hat{I}$中的衣服信息造成任何干扰。

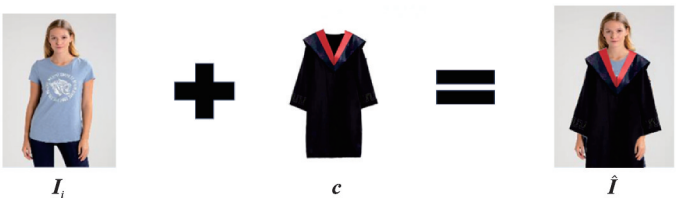

图1 面向学位服照片生成的虚拟试衣方法

Fig.1 Virtual try-on method for generation of graduation photo

在进行训练时,采用构建训练三元组(I_i, c, I_t)对网络进行训练可以达到较好的效果,其中I_t为I_i中人物穿着学位服c的图像,即目标生成图像$\hat{I}$的真实图像。但是,由于训练任务要求图像I_i与I_t中除人物穿着的衣服不同以外,图片中的其他信息,例如人体姿势与人物特征等均需相同,这就使得获取这样一对训练三元组几乎不可能。而VITON网络[4]可以解决此问题,即通过与服装无关的人体特征表示来保持用户的身体姿势、身材以及身份信息。

图2展示了面向学位服照片生成的虚拟试衣方法的整体框架。首先获取输入图像I_i的与服装无关

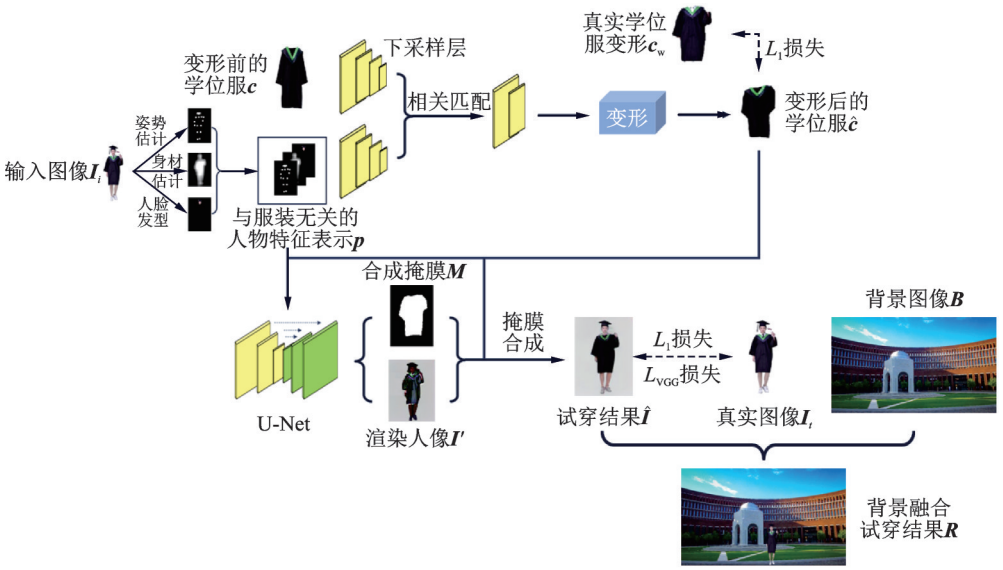

图2 面向学位服照片生成的虚拟试衣方法的整体框架

Fig.2 Overall framework of virtual try-on network for graduation photo generation

的人物特征表示$\boldsymbol{p}$,其中包括姿势估计(即姿势热图获取)、身材估计(即人物形状获取)及人脸与发型部分提取(即身份信息获取)。在训练阶段,以与服装无关的人物特征表示$\boldsymbol{p}$和变形前的学位服$\boldsymbol{c}$作为输入,将其通过服装变形模块后生成变形后的学位服$\hat{\boldsymbol{c}}$。随后,将$\hat{\boldsymbol{c}}$与$\boldsymbol{p}$作为虚拟试穿模块的输入,生成最终的虚拟试衣结果$\hat{\boldsymbol{I}}$。合成虚拟试衣结果$\hat{\boldsymbol{I}}$后,将其通过背景融合模块,即可生成相应背景之下的虚拟试衣结果$\boldsymbol{R}$。

2.2 与服装无关的人物特征表示

实现虚拟试衣任务目标必不可少的一个环节是提取原始图像$\boldsymbol{I}_i$中的人物特征。借助人物特征表示中的一些人体特征,才能使目标衣物准确地覆盖在人体相应部位的同时保持原图像$\boldsymbol{I}_i$中其他人体特征。由于在进行学位服虚拟试穿时相应的信息也需要得到保留,而VITON网络[5]中获取人物特征表示的方法对于穿着学位服的人像同样适用,因此本文采用VITON网络[5]中获取人物特征表示的方法提取本文的人物特征。同时,学位服与其他服装相比相对宽大,在试穿时人体体型的变化对试穿结果的影响不大,因此没有采用SP-VITON网络[9]中利用DensePose模型[10]获取人体身材的方法。本文获得3个不同方面人物特征表示的方法。

2.2.1 姿势热图

人体姿势是人物特征表示中的一个重要组成部分,也是决定着衣服变形状态与变形程度的一个重要因素。Cao等[20]在2017年提出了一种先进的姿势估计模型部分关联域(Part affinity fields, PAFs),利用这个模型可以有效地检测出人物图像的2D姿势并获取关键的姿势点。本文基于PAFs[20]估计人体姿势,将用户图像$\boldsymbol{I}_i$作为输入来获取原始图像中人体的18个姿势关键点的坐标。除获取这些关键点的二维坐标外,还需要获取其对应部位的语义位置属性(即每个关键点所以对应身体的部位,例如头部、颈部、肘部及膝部等),因此需要对这18通道的姿势关键点进行单独存储。本文用“1”填充每个关键点周围11×11部分的邻域,用“0”填充其余部分,这样就形成了用于表示人体姿势的18通道姿势热图。

2.2.2 人体身材

本文使用Liang等[21]于2018年提出的人体解析模型联合身体解析与姿态估计网络(Joint body parsing & pose estimation network, JPPNet)获得用户图像$\boldsymbol{I}_i$的人体分割图。随后对分割图下采样,生成模糊的二进制掩膜的一通道特征图,使得该特征图能够大致覆盖人体的不同部位,以此作为对人体身材的估计。

2.2.3 身份信息

身份信息指人的面部及头发部位等信息,也是在虚拟试衣任务中需要保留下来的关键信息,往往需要通过对人体进行解析后获取。本文同样采用JPPNet模型[21]生成人体分割图。在人体分割图中,不同的区域代表不同的语义信息,即人体的不同部位。在虚拟试衣任务中,通过对应的语义信息,将其中的面部和头发部分提取为三通道RGB图像,作为辨识人物身份的标志。将上述3方面的人体特征进行拼接即获得了文中用到的人物特征表示,如图3所示。该特征表示中包含虚拟试衣任务中所需的人物信息,作为后文虚拟试穿网络的输入可以使其达到更好的训练效果。

2.3 虚拟试穿网络

本文的虚拟试穿网络由服装变形模块与虚拟试穿模块组成。

2.3.1 服装变形模块

服装变形模块以与服装无关的人物特征表示$\boldsymbol{p}$以及变形前的学位服图像$\boldsymbol{c}$作为输入,输出变形后的学位服图像$\hat{\boldsymbol{c}}$。若以D表示服装变形模块,其功能即可表示为$\hat{\boldsymbol{c}}=D(\boldsymbol{p},\boldsymbol{c})$。在服装变形模块中,首先将

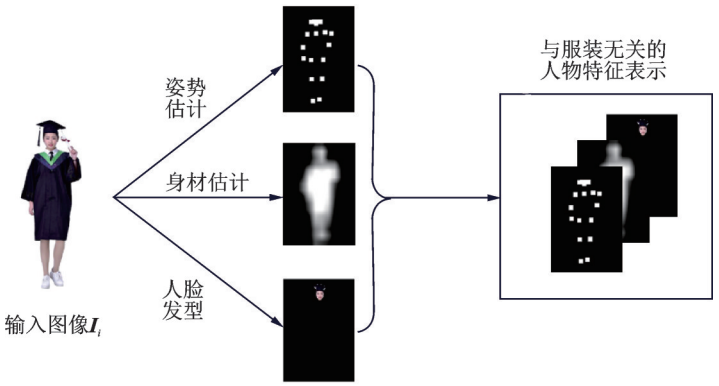

图3 与服装无关的人体特征表示

Fig.3 Clothing-free person representation

人物特征表示$\boldsymbol{p}$与学位服图像$\boldsymbol{c}$分别通过2个用于提取高级特征的网络,将提取到的2个高级特征通过1个相关层组合成为1个张量。随后,将该张量通过一个回归网络,计算空间变换参数θ。最后,根据人物特征表示$\boldsymbol{p}$,利用一个参数为θ的具有形状上下文匹配[22]的薄板样条(Thin plate spline, TPS)变换模块,将输入的学位服图像$\boldsymbol{c}$输出为变形后的学位服图像$\hat{\boldsymbol{c}}$。若用T表示该TPS模块,则该功能可对应表示为$\hat{\boldsymbol{c}}=T_\theta(\boldsymbol{c})$。服装变形模块的损失函数$L_D(\theta)$由变形学位服$\hat{\boldsymbol{c}}$与原始图像$\boldsymbol{I}_t$中的人像所穿衣服$\boldsymbol{c}_{\mathrm{w}}$之间的$L_1$损失组成,有

$$L_D(\theta)=\left\|\hat{\boldsymbol{c}}-\boldsymbol{c}_{\mathrm{w}}\right\|_1=\left\|T_\theta(\boldsymbol{c})-\boldsymbol{c}_{\mathrm{w}}\right\|_1 \tag{1}$$

在进行训练时,服装变形模块通过最小化损失函数$L_D(\theta)$来学习如何将目标衣物转移到目标人物的对应部位并对其逐渐优化,从而生成更理想的输出。

2.3.2 虚拟试穿模块

虚拟试穿模块旨在将变形后的学位服$\hat{\boldsymbol{c}}$与目标人像进行融合,生成最终的试穿结果$\hat{\boldsymbol{I}}$。在虚拟试穿模块T中,首先以与服装无关的人物特征表示$\boldsymbol{p}$与在服装变形模块中生成的变形后的学位服$\hat{\boldsymbol{c}}$作为输入,依次经过下采样层与上采样层构成的U-Net,在生成渲染人像$\boldsymbol{I}'$的同时生成学位服合成图的掩膜$\boldsymbol{M}$,随后将$\boldsymbol{M}$与$\boldsymbol{I}'$通过掩膜合成生成最终的试穿结果$\hat{\boldsymbol{I}}$,表达式为

$$\hat{\boldsymbol{I}}=\boldsymbol{M}\odot\hat{\boldsymbol{c}}+(\boldsymbol{1}-\boldsymbol{M})\odot\boldsymbol{I}' \tag{2}$$

虚拟试穿模块的损失函数L_T由L_1损失和VGG(Visual geometry group)感知损失[23]构成,表达式为

$$L_T=\lambda_{\mathrm{warp}}\left\|\hat{\boldsymbol{I}}-\boldsymbol{I}_t\right\|_1+\lambda_{\mathrm{VGG}}\lambda_i\left\|\phi_i(\hat{\boldsymbol{I}})-\phi_i(\boldsymbol{I}_t)\right\|_1+\lambda_M\left\|\boldsymbol{1}-\boldsymbol{M}\right\|_1 \tag{3}$$

式中:λ_{warp}表示L_1范数的权重;λ_{VGG}表示VGG感知损失范数的权重。式中的第1项代表试穿结果$\hat{\boldsymbol{I}}$与真实图像$\boldsymbol{I}_t$之间的L_1损失;第2项代表试穿结果$\hat{\boldsymbol{I}}$与真实图像$\boldsymbol{I}_t$之间的感知损失。通过L_1范数和VGG感知损失范数可以使得生成的结果$\hat{\boldsymbol{I}}$利用更多变形后的学位服中的信息并且使其更加平滑,从而使得生成的结果能够保留更多衣服上的细节纹理,并且使试穿结果看起来更加逼真自然。

2.4 背景融合模块

经过上述虚拟试穿网络得到试穿的人像结果后,本文将其通过一个背景融合模块使其能够自由地

与所需的背景进行拼接。如图4所示,将生成的学位服试穿结果$\hat{I}$与背景图片B作为输入,目标是生成一个图像R,R中除被$\hat{I}$替换的像素部分将背景遮挡外,其余部分均保留原有背景图片B中的特征,且$\hat{I}$经背景融合后人物特征(如人物姿态、衣服特征等)保持不变。

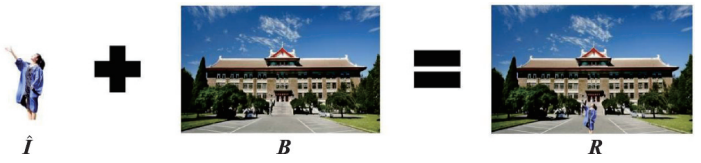

图4 背景融合目标

Fig.4 Goal of background fusion

如图5所示,在本文的背景融合模块中,首先对人物图像进行二值化处理得到二值化图像D。随后,通过图像边界两侧灰度级的突变获取图像的边界连续像素序列,进而得到人物的边缘轮廓序列并对其最外层进行绘制,从而获得以人物最外层边缘为界的掩膜图像m。由于使用仅通过提取边缘后的掩膜图像m不易与背景图像自然融合,因此本文使用图像膨胀技术对掩膜图像m边缘白色像素进行处理。通过膨胀去除任务边缘干扰色素得到更完美的掩膜图像$\hat{m}$,使人物边界向外部扩张,与背景融合更加自然。在原背景图像上确定人物位置,根据得到的掩膜图像$\hat{m}$对原背景图像B进行相应的像素替换,即可得到背景融合后的虚拟试穿结果R。

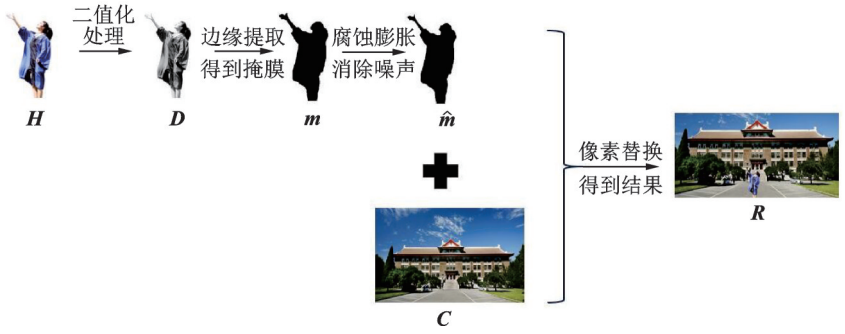

图5 背景融合整体框架

Fig.5 Background fusion overall framework

3 实验分析

3.1 数据集

本文构建了一个新的学位服照片数据集(https://github.com/jr011/-dataset),数据集信息如表1所示。首先,在淘宝、京东等网站上爬取了模特穿着学位服的图片,经过人工质量筛选后剩余1 176张人物图像I_t。由于学位服之间样式类似,最主要的区别为领子颜色区别,在此实验中本文共收集了5种领子颜色的学位服,分别为红色、黄色、粉色、绿色及银白色。因此,根据学位服图片c中领子颜色为其做对应,得到1 176组对应的训练集。由于数据量过少,可能导致网络学习效果不佳,本

表1 数据集信息

Table 1 Dataset information

数据集信息	量值
训练对总数量/组	16 225
学位服训练对数量/组	1 176
长裙训练对数量/组	15 049
分辨率/(像素×像素)	443×641

文尝试利用长裙数据集与学位服数据集进行合并对网络进行训练。因此，本文又在www.zalando.de网站上爬取了15 049组训练对，其中每组训练对包括裙子图像$\boldsymbol{c}$和人物图像$\boldsymbol{I}_i$，其中人物图像中的模特穿着对应裙子。由于网络上收集到的图片分辨率不同，且背景颜色非常丰富，会对后期的训练结果产生较大的影响，因此需要对训练集进行预处理。首先将其分辨率统一为443像素×641像素，随后对其背景颜色进行处理，将其背景统一处理成白色。

3.2 实验细节

本文在训练中使用Adam优化器[24]，其参数为：$\beta_1=0.5$，$\beta_2=0.999$，$\lambda_{\text{warp}}=\lambda_{\text{VGG}}=\lambda_M=1$。训练服装变形模块50 000步，训练虚拟试穿模块50 000步。在前50 000步的训练中将学习率设为0.000 1，对于后50 000步将学习率线性衰减为0。原始输入图像的分辨率为443像素×641像素，然后在输入过程中调整为256×192的固定大小，最终输出图像也具有相同的分辨率。对实验图片的预处理和背景合成部分在CPU i7-8550U处理器上运行，获取与服装无关的人体特征表示、网络训练及测试部分均在GPU GTX 1080 Ti上进行，Tensorflow版本为1.15，Python版本为3.7。

在第1阶段服装变形模块中，特征提取网络包含4个二步下采样卷积层以及2个一步卷积层，其滤波器的数量分别为64、128、256、512、512、512。而回归网络包含2个二步卷积层、2个一步卷积层以及1个全连接输出层，其滤波器数量分别为512、256、128、64。第2阶段的虚拟试穿模块中，在U-Net中包含6个二步下采样卷积层和6个上采样卷积层，下采样卷积层的滤波器数量分别为64、128、256、512、512、512，上采样卷积层滤波器数量分别为512、512、256、128、64、4。在每个卷积层之后是实例归一化层[25]以及斜率为0.2的Leaky ReLU[26]。

3.3 实验结果

图6展示了对服装变形模块进行训练的结果，将真实衣服变形和通过TPS生成的变形结果进行比较，可以看出除衣服边缘变形有一定差距以及手持物遮挡外，生成的变形结果比较准确。

图7展示了对虚拟试穿模块进行训练的结果，利用服装变形模块生成的变形后的学位服以及变形掩膜图像，虚拟试穿模块也可以较为准确地将变形后的学位服覆盖在人像的对应位置。

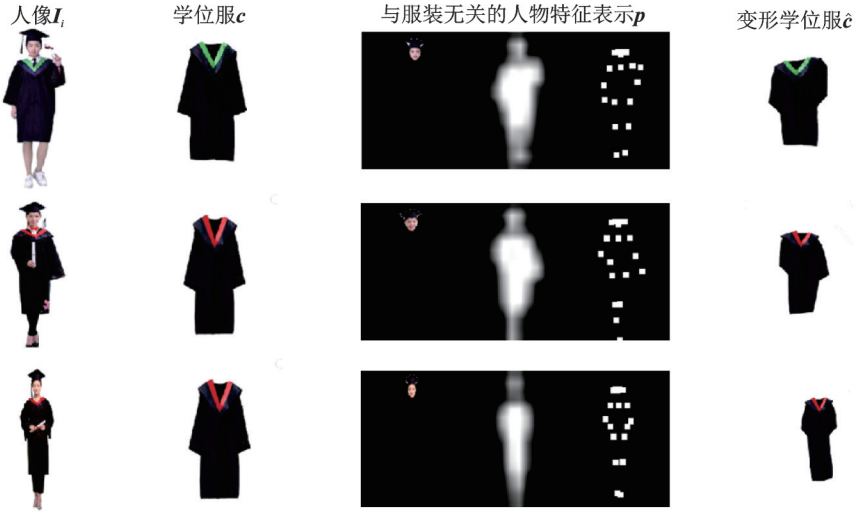

图6 服装变形模块训练结果

Fig.6 Training results of clothing deformation module

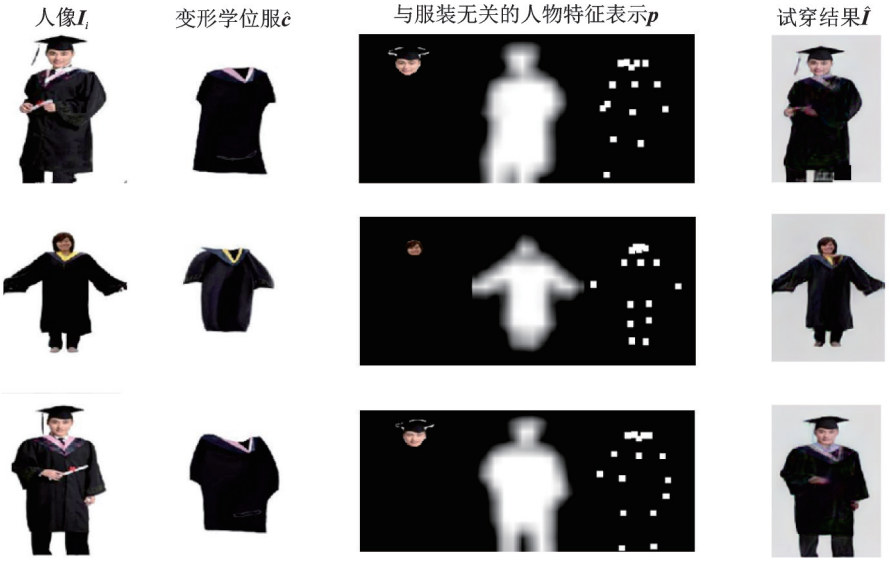

图7 虚拟试穿模块训练结果

Fig.7 Training results of virtual try-on module

图8展示了对虚拟试衣网络进行测试的结果。利用本文的方法,在测试结果的图像生成中,原有图像的人体服装对试衣结果几乎没有影响。模型可以使学位服较好地根据人像身体特征进行变形,将其覆盖在人体的对应区域,并且将人物特征及学位服上的褶皱较好地保留,从而产生较为逼真的试衣结果。同时利用本模型可以在不同的人像上进行虚拟试穿生成不同的结果,生成的虚拟试衣结果具有多样性,证明本文方法具有良好的鲁棒性。

图8 虚拟试衣结果

Fig.8 Virtual try-on results

将输出的虚拟试衣合成人像通过背景合成网络与背景进行合成,结果如图9所示。实验结果表明生成的图像有较强的真实性。

3.4 试衣结果定量分析

除可视化结果外,本文还对生成的虚拟试衣结果进行了定量分析。在虚拟试衣工作中,定量评价

图9 背景融合结果

Fig.9 Background fusion results

指标主要有起始分数(Inception score, IS)[27]、峰值信噪比(Peak signal-to-noise ratio, PSNR)和结构相似性(Structural similarity, SSIM)[28]等。本文参考VITON、CP-VTON中对结果的定量分析方法,同样采用IS指标[25]定量评估本文模型的图像生成质量。IS指标的计算采用Google的图片分类网络Inception Net V3模型,主要考虑生成图像的清晰度及生成图像的多样性两个方面。模型将生成器生成的图像数据作为图片分类网络的输入,并输出1 000维向量,其中不同维的向量代表该图像属于不同类别的概率。在衡量清晰度时,图像越清晰,网络对于图像的类别判断越准确,因此图像属于某一类的概率会越大。在衡量多样性时,当生成器生成的图像数据足够多时,其多样性也越强,图像数据在1 000类中的类别分布越平均。综合来看,生成的图像越清晰且生成图像的多样性越强时,IS值越高。表2展示了本文方法与VITON、CP-VTON及SP-VTON方法在原数据集进行试衣生成结果的IS值。通过对比可以看出,本文方法与VITON、CP-VTON及SP-VTON方法的IS值差异不大,证明本文方法在清晰度和多样性方面可以达到较好的标准,验证了本文方法的图像生成质量。

表2 本文方法与其他方法IS定量比较

Table 2 IS value comparison between the proposed method and other methods

方法	IS值
VITON	2.516
CP-VTON	2.748
SP-VITON	2.656
本文方法	2.810
真实数据	3.765

3.5 试衣失败情况分析

图10展示了本文虚拟试衣网络生成学位服照片的2种失败情况,其原因分别是:(1)训练集中人物的腿部几乎都呈竖直站立的状态,因而对于一些较为复杂罕见的人物姿势,网络无法准确地将服装覆盖到人体的对应部位;(2)人物原始服装颜色较深且形状较为复杂,在进行试穿时很难消除原始服装对试穿结果带来的影响。针对这一点,可以通过引入DensePose模块[10]提取衣服下的人体身材消除原始衣物的影响,对结果进行改进。除此之外,在进行试穿时手部和脚部等处的信息可能会被模糊或改变。通过对VITON[5]及CP-CTON[7]网络的实验结果进行观察,发现其也有相同的现象,即在虚拟试衣图像生成时会生成原本不存在的伪影,导致其结

图10 失败情况

Fig.10 Failure cases

果中原本裤子等非试衣区域的颜色发生改变。这是由于在生成虚拟试衣结果时,网络忽略了人体解析、服装和姿势之间的相互作用。针对这一点,为了使原有图像中的更多细节被更好地保留,可以利用保留头部的方法保留手部和脚部信息,即根据对应的语义信息直接将其从原图像中提取保留以对试衣结果进行优化。

4 结束语

本文提出了一种面向学位服照片生成的虚拟试衣方法,用于根据用户所提供的人物图像以及选择的目标学位服来合成虚拟试穿图像,并将虚拟试穿图像与特定的背景进行融合,生成逼真的毕业场景照片。文中针对学位服的特性,设计了面向学位服等长裙类衣物的虚拟试衣方法,并且构建了一个学位服虚拟试衣数据集。实验结果表明,本文方法能够在完成虚拟试穿任务时较好地保留原始图片中人物的姿势、形体特征及身份特征,从而将目标学位服很好地与人物图像相融合,并且保留衣服上的细节纹理。同时也能在保持人物特征及服装的情况下将生成的虚拟试衣图像与背景自然融合,有助于毕业生在线进行毕业照生成。

参考文献:

[1] GUAN Peng, REISS L, HIRSHBERG D A, et al. DRAPE: Dressing any person[J]. ACM Transaction of Graph, 2012, 31(4): 1-10.

[2] HAHN F, THOMASZEWSKI B, COROS S, et al. Gross: Subspace clothing simulation using adaptive bases[J]. ACM Transaction of Graph, 2014, 33(4): 1-9.

[3] JETCHEV N, BERGMANN U. The conditional analogy GAN: Swapping fashion articles on people images[C]//Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. Los Alamitos, CA, USA: IEEE, 2017: 2287-2292.

[4] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[5] HAN Xintong, WU Zuxuan, WU Zhe, et al. Viton: An image-based virtual try-on network[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Computer Society, 2018: 7543-7552.

[6] BELONGIE S, MALIK J, PUZICHA J. Shape matching and object recognition using shape contexts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 509-522.

[7] WANG Bochao, ZHENG Huabin, LIANG Xiaodan, et al. Toward characteristic-preserving image-based virtual try-on network[C]//Proceedings of 2018 European Conference on Computer Vision. Munich, Germany: Springer, 2018: 589-604.

[8] ROCCO I, ARANDJELOVIC R, SIVIC J. Convolutional neural network architecture for geometric matching[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA, USA: IEEE, 2017: 6148-6157.

[9] SONG Dan, LI Tianbao, MAO Zhendong, et al. SP-VITON: Shape-preserving image-based virtual try-on network[J]. Multimedia Tools and Applications, 2020, 79(45): 33757-33769.

[10] GULER R A, NEVEROVA N, KOKKINOS I. Densepose: Dense human pose estimation in the wild[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Computer Society, 2018: 7297-7306.

[11] HAO Tong, YU Ailing, PENG Wei, et al. Cross domain mitotic cell recognition[J]. Neurocomputing, 2016, 195: 6-12.

[12] ZHU Shizhan, URTASUN R, FIDLER S, et al. Be your own prada: Fashion synthesis with structural coherence[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Los Alamitos, CA, USA: IEEE, 2017: 1680-1688.

[13] LONG J, SHELHAMER E, DARRRLL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE Computer Society, 2015: 3431-3440.

[14] DONG Haoye, LIANG Xiaodan, SHEN Xiaohui, et al. Towards multi-pose guided virtual try-on network[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Los Alamitos, CA, USA: IEEE Computer Society, 2019: 9026-9035.

[15] WU Zhonghua, LIN Guosheng, TAO Qingyi, et al. M2e-try on net: Fashion from model to everyone[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York, the United States: Association for Computing Machinery, 2019: 293-301.

[16] YOO D, KIM N, PARK S, et al. Pixel-level domain transfer[C]//Proceedings of 2016 European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016: 517-532.

[17] NEVEROVA N, GULER R A, KOKKINOS I. Dense pose transfer[C]//Proceedings of 2018 European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 123-138.

[18] DONG Haoye, LIANG Xiaodan, GONG Ke, et al. Soft-gated warping-GAN for pose-guided person image synthesis[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc, 2018: 472-482.

[19] DONG Haoye, LIANG Xiaodan, SHEN Xiaohui, et al. FW-GAN: Flow-navigated warping GAN for video virtual try-on [C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Los Alamitos, CA, USA: IEEE Computer Society, 2019: 1161-1170.

[20] CAO Zhe, SIMON T, WEI Shien, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA, USA: IEEE, 2017: 7291-7299.

[21] LIANG Xiaodan, GONG Ke, SHEN Xiaohui, et al. Look into person: Joint body parsing & pose estimation network and a new benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(4): 871-885.

[22] BELONGIE S, MALIK J, PUZICHA J. Shape matching and object recognition using shape contexts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 509-522.

[23] JOHNSON J, ALAHI A, LI Feifei. Perceptual losses for real-time style transfer and super-resolution[C]//Proceedings of 2016 European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016: 694-711.

[24] DIEDERIK P, KINGM A, JIMMY B. Adam: A method for stochastic optimization[C]//Proceedings of 2015 International Conference on Learning Representations. San Diego, CA, USA: OpenReview.net, 2015: 13.

[25] ULYANOV D, VEDALDI A, LEMPITSKY V. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA, USA: IEEE, 2017: 6924-6932.

[26] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models[C]//Proceedings of 2013 International Conference on Machine Learning. Atlanta, GA, USA: JMLR.org, 2013: 3.

[27] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training gans[J]. Advances in Neural Information Processing Systems, 2016. DOI: 10.48550/arXiv.1606.03498.

[28] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

作者简介:

盛培卓(2000-),女,硕士研究生,研究方向:计算机视觉、多视图学习,E-mail: peizhuosheng@tju.edu.cn。

李婷玉(1998-),女,本科,研究方向:虚拟试衣,E-mail: 2878755056@qq.com。

李天宝(1996-),男,硕士研究生,研究方向:虚拟试衣、计算机视觉,E-mail: li-tianbao@tju.edu.cn。

宋丹(1992-),**通信作者**,女,博士,讲师,硕士生导师,研究方向:计算机图形学、虚拟试衣,E-mail: dan.song@tju.edu.cn。

刘安安(1982-),男,博士,教授,博士生导师,研究方向:计算机视觉、机器学习、三维模型检索,E-mail: liuanan@tju.edu.cn。

(编辑:张黄群)