



(12) 发明专利

(10) 授权公告号 CN 114663552 B

(45) 授权公告日 2022.08.16

(21) 申请号 202210573730.5

G06N 3/04 (2006.01)

(22) 申请日 2022.05.25

G06K 9/62 (2022.01)

(65) 同一申请的已公布的文献号
申请公布号 CN 114663552 A

(56) 对比文件

(43) 申请公布日 2022.06.24

CN 105354876 A, 2016.02.24

CN 104156966 A, 2014.11.19

(73) 专利权人 武汉纺织大学

CN 111768472 A, 2020.10.13

CN 112613439 A, 2021.04.06

地址 430073 湖北省武汉市洪山区纺织路1号

CN 112232914 A, 2021.01.15

CN 103578004 A, 2014.02.12

(72) 发明人 姜明华 史衍康 余锋 花爱玲
周昌龙 宋坤芳

CN 111787242 A, 2020.10.16

CN 112233222 A, 2021.01.15

(74) 专利代理机构 武汉科皓知识产权代理事务
所(特殊普通合伙) 42222

CN 113012303 A, 2021.06.22

CN 113781164 A, 2021.12.10

专利代理师 王琪

CN 113052980 A, 2021.06.29

CN 114419335 A, 2022.04.29

(51) Int. Cl.

US 2017024928 A1, 2017.01.26

G06T 11/60 (2006.01)

CN 103597519 A, 2014.02.19

G06V 40/10 (2022.01)

GB 201510752 D0, 2015.08.05

G06V 10/26 (2022.01)

US 2021133919 A1, 2021.05.06

(续)

G06V 10/74 (2022.01)

审查员 夏兵秀

G06V 10/764 (2022.01)

权利要求书4页 说明书9页 附图2页

G06V 10/82 (2022.01)

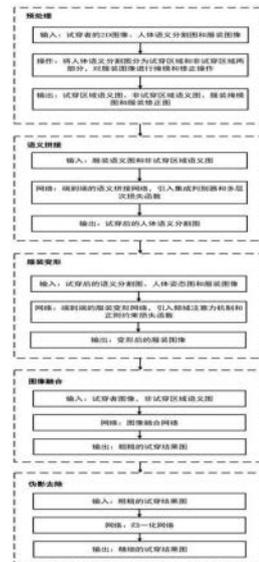
(54) 发明名称

更多的细节,生成高分辨率的试衣图片。

一种基于2D图像的虚拟试衣方法

(57) 摘要

本发明公开了一种基于2D图像的虚拟试衣方法,涉及虚拟试衣技术领域。本发明方法包括:输入虚拟试衣需要的图像数据,经过预处理操作生成试穿区域语义图、非试穿区域语义图和服装修正图;经过端到端的语义拼接网络,在引入集成判别器和多层次损失函数的约束下,生成试穿后的人体语义分割图;通过服装变形网络,在引入频域分类器和正则约束损失函数的约束下,预测出变形后的服装图像;将试穿者图像和变形后的服装图像拼接得到粗糙的试穿结果图;经过归一化网络计算并缩小伪影区域,最终得到精细的试穿结果图。本发明方法解决现有技术中服装与人体错位产生伪影的问题,使得生成的图片保持



CN 114663552 B

[接上页]

(56) 对比文件

JP 2020097803 A,2020.06.25	US 2020015575 A1,2020.01.16
KR 20140125507 A,2014.10.29	WO 2019193467 A1,2019.10.10
US 2021049811 A1,2021.02.18	KR 20210099353 A,2021.08.12
CN 104813340 A,2015.07.29	WO 2022002961 A1,2022.01.06
WO 2021008166 A1,2021.01.21	US 2019333267 A1,2019.10.31
CN 110852941 A,2020.02.28	WO 2016109884 A1,2016.07.14
EP 3091510 A1,2016.11.09	WO 2020131518 A1,2020.06.25
JP 2016038812 A,2016.03.22	CN 111709874 A,2020.09.25
CN 109740529 A,2019.05.10	CN 112330580 A,2021.02.05
	US 2017084035 A1,2017.03.23
	US 2020151807 A1,2020.05.14

1. 一种基于2D图像的虚拟试衣方法,其特征在于,包括如下步骤:

步骤(1),输入试穿者的2D图像、人体语义分割图和服装图像,经过预处理操作将人体语义分割图分为试穿区域和非试穿区域两部分,对服装图像进行掩模和修正处理,生成服装掩模图和服装修正图;

步骤(2),输入服装语义图和非试穿区域语义图,经过端到端的语义拼接网络,在引入集成判别器和多层次损失函数的约束下,生成试穿后的人体语义分割图;

步骤(3),根据试穿后的语义分割图、人体姿态图和服装图像,通过服装变形网络,在引入频域分类器和正则约束损失函数的约束下,预测出变形后的服装图像;

步骤(3)中服装变形网络是端到端的网络,输入是试穿后的语义分割图、服装图像和人体姿态图,在人体姿态图中若干个关键点位置的约束下,经过两条支路通过深度卷积网络分别提取试穿后的语义分割图和服装图像的关键点位置的像素信息,引入频域分类器,使用插值法预测变形后服装图像的像素信息,引入正则约束损失函数控制服装的变形程度,最后输出变形后的服装图像;

正则约束损失函数表示为:

$$\mathcal{L}_{nom}(W, M) = \sqrt{\alpha \sum_x \sum_y (W(x) - M(x))^2 + \beta \sum_x \sum_y (W(y) - M(y))^2}$$

上式中, $W(x)$, $M(x)$ 分别表示变形前后坐标点的x值,

$W(y)$, $M(y)$ 分别表示变形前后坐标点的y值,

$(W(x) - M(x))^2$ 表示x轴方向上坐标的变化量的平方,

$(W(y) - M(y))^2$ 表示y方向上坐标的变化量, α, β 为两个可学习的参数,用于控制服装在x轴和y轴方向的变形程度;

服装变形网络的损失函数表示为:

$$\mathcal{L}_c = \lambda_1 \mathcal{L}_{CGAN} + \lambda_2 \mathcal{L}_{FM} + \lambda_3 \mathcal{L}_{nom}$$

上式中, \mathcal{L}_{CGAN} 是条件对抗损失函数, \mathcal{L}_{FM} 是像素交叉熵损失函数, \mathcal{L}_{nom} 是正则约束损失函数, $\lambda_1, \lambda_2, \lambda_3$ 是系数,用于调节三个损失函数对总损失函数的影响;

步骤(3)中频域分类器指在频域空间,通过设置阈值来对特征进行划分,一方面更多地把关注点放在生成低频特征信息上,增强图像的内容信息,另一方面借助试穿后的语义分割图区别开噪点和高频信息,提取更多有用的高频信息,使用插值法增强图像高频区域的纹理精细度,从而提高生成图片的质量;频域分类器的具体流程包括:

首先,由卷积层提取服装图像的空间域特征图,使用离散傅里叶变换得到频域特征图,可表示为:

$$F_{i,j} = DFT(S_{i,j})$$

其中, $S_{i,j}$ 表示空间域特征图中第i行第j列的空间特征分量, $F_{i,j}$ 表示频域特征图中第i

行第*j*列的频率特征分量, $DFT()$ 是离散傅里叶变换函数;

然后,通过神经网络结构学习得到频域特征图中每个频率特征分量的重要性权值,将权值大于阈值的频率特征分量标识为低频信息特征分量,小于阈值的频率特征分量标识为高频信息特征分量,可表示为:

$$Freq_{i,j,z} = \begin{cases} Freq_{i,j,0}, & k_{i,j} \geq \lambda \\ Freq_{i,j,1}, & k_{i,j} < \lambda \end{cases}$$

其中, $Freq_{i,j,z}$ 表示第*i*行第*j*列类别为*z*的频率特征分量,当*z*=0时表示该频率特征分量为低频信息特征分量,当*z*=1时表示该频率特征分量为高频信息特征分量, $k_{i,j}$ 表示第*i*行第*j*列频率特征分量的重要性权值, λ 是阈值;

最后,通过试穿后的语义分割图中的位置信息进一步在高频信息特征通道中筛选出噪声信息,使用低通滤波器来过滤掉图像噪声,使用插值法填充图像内容和纹理信息;

步骤(4),根据试穿者图像,对非试穿区域语义图进行着色得到非试穿区域彩色图,再与变形后的服装图像进行拼接得到粗糙的试穿结果图;

步骤(5),将粗糙的试穿结果图和试穿后的语义分割图输入到归一化网络中,计算并缩小伪影区域,最终得到精细的试穿结果图。

2.如权利要求1所述的一种基于2D图像的虚拟试衣方法,其特征在于:步骤(1)中输入试穿者的2D图像、人体语义分割图和服装图像后,预处理过程包括两部分:

一部分,输入人体语义分割图,包含7种语义信息依次为:头发、面部、脖子、手部、下装、衣服和手臂的语义信息,并且各语义信息使用不同像素值进行标注,可表示为: $Y = \{a_1, a_2, a_3, \dots, a_7\}$,在预处理过程中根据各语义信息像素值的不同,将人体语义分割图划分为非试穿区域语义图: $Y_1 = \{a_1, a_2, a_3, a_4, a_5\}$ 和试穿区域语义图: $Y_2 = \{a_6, a_7\}$,其中,非试穿区域语义图包括:头发、面部、脖子、手部、下装的语义信息,试穿区域语义图包括:衣服和手臂的语义信息;

另一部分,输入服装图像,对服装图像进行掩膜处理,即将服装图像中的服装区域的像素值设置为0,背景的像素值设置为255,得到服装掩模图,根据试穿区域语义图中像素点的位置信息对服装图像进行仿射变换,使得服装图像的轮廓与试穿区域语义图的轮廓一致,从而得到服装修正图,为后面的服装变形做准备。

3.如权利要求1所述的一种基于2D图像的虚拟试衣方法,其特征在于:步骤(2)中语义拼接网络是端到端的网络,输入是服装语义图和非试穿区域语义图,先经过卷积层提取图像特征并进行下采样,再经过反卷积层将压缩后的图像上采样成原本的尺寸,得到粗糙的试穿后的人体语义分割图,最后将粗糙的试穿后的人体语义分割图输入到集成判别器进行相似度计算,经过多轮语义拼接后输出相似度最高的试穿后的人体语义分割图;集成判别器是由*n*个相同网络结构的深度卷积网络拼接而成,*n*的大小与输入图片的分辨率有关,深度卷积网络由池化层、卷积层、归一化层和激活层依次连接构成;集成判别器的输入是真实图片和预测图片,输出是两个图片的特征相似度,*n*个深度卷积网络分别在*n*个不同尺度上区分真实图像和预测图像,并且通过设置系数来控制每个深度卷积网络的预测结果对最终

结果的影响,从而鼓励语义拼接网络生成更多的细节;每个深度卷积网络对应一个特征匹配损失函数,合起来构成集成判别器的多层次损失函数,对预测结果进行细粒度优化,鼓励语义拼接网络生成更精细的细节;

多层次损失函数表示为:

$$\mathcal{L}_{mul}(D_1, D_2, \dots, D_n) = \gamma_1 \mathcal{L}_{FM}(G(1), D_1) + \gamma_2 \mathcal{L}_{FM}(G(2), D_2) + \dots + \gamma_n \mathcal{L}_{FM}(G(n), D_n)$$

上式中, $\mathcal{L}_{FM}(G(1), D_1)$ 表示深度卷积网络 D_1 对语义拼接网络生成的 $G(1)$ 图像的特征匹配损失函数, $\gamma_1, \gamma_2, \dots, \gamma_n$ 是系数,控制不同尺度的深度卷积网络损失对多层次损失函数的影响,即 $\gamma_1 + \gamma_2 + \dots + \gamma_n = 1$ 。

4. 如权利要求1所述的一种基于2D图像的虚拟试衣方法,其特征在于:步骤(5)中去除伪影的处理过程包括:首先,通过深度卷积网络提取粗糙的试穿结果图的图像特征,经过上采样得到粗糙试穿结果图的语义分割图,与试穿后的语义分割图求差值,该过程可以表示为:

$$Artifact = R - S$$

其中, $Artifact$ 表示伪影区域的像素信息, R 表示粗糙的试穿结果图, S 表示试穿后的语义分割图;

得到一张只有小部分像素值大于0的图像,其中像素值大于0的像素点即为伪影区域的像素点,并根据伪影区域的位置信息将该区域在粗糙的试穿结果图中标识出来;然后,经过归一化网络,学习到两个型变系数 α 和 β ,分别控制伪影区域在x轴和y轴方向的缩小程度,用于缩小伪影区域的大小;最后,经过上采样层,将调整后的特征图片上采样成原图大小,并使用插值法对损失的像素点值进行填充,生成更加精细的试穿结果图。

5. 如权利要求1所述的一种基于2D图像的虚拟试衣方法,其特征在于:步骤(5)中归一化网络的输入是粗糙的试穿结果图,输出是精细的试穿结果图,归一化网络包括获取伪影语义图和消除伪影信息两个阶段:

第一个获取伪影语义图阶段,包含两个子网络,设为网络1-1、1-2,其中网络1-1中包含 n 层卷积层,通过卷积层提取图像特征,以及 n 层反卷积层,通过反卷积将图像映射为原尺寸大小,卷积层的输出作为反卷积层的输入,卷积层与反卷积层之间通过残差结构连接,在最后一层反卷积后输出结果;其中网络1-2包含图像求差模块,输入两种相同尺寸的图像,输出两种图像求差后的结果图;其中 n 的取值与输入图片的尺寸有关;

第二个消除伪影信息阶段,包含一个网络2-1,其中网络2-1由两条并联支路构成,其中,第一条支路由一个卷积层,一个激活层、一个区域正则化层、一个下采样,一个上采样层、和一个相加层依次串联而成,第二条支路由单一的区域正则化层构成,在相加层后输出最后的结果;

其中,所述区域正则化层,其作用是对伪影区域进行标准化,首先获得伪影区域的空间位置,然后通过卷积网络计算型变系数 α 和 β ,引入平面变形损失函数来控制伪影区域的缩减程度,防止因过度缩减导致错误修改人体部位的像素点信息,以此来辅助缩小伪影区域;所述相加层,其作用是对两条支路的输出结果进行融合;

其中,平面变形损失函数表示为:

$$\mathcal{L}_{rul}(C_x, C_y) = \alpha \sum_{i=-1,1} \sum_x \sum_y |C_x(x+i, y) - C_x(x, y)| + \beta \sum_{i=-1,1} \sum_x \sum_y |C_y(x, y+j) - C_y(x, y)|$$

上式中, C_x, C_y 分别表示伪影区域特征点的 x, y 坐标,

$|C_x(x+i, y) - C_x(x, y)|$ 表示两个节点之间的欧氏距离, i, j 是形变量, α, β 是型变系数。

一种基于2D图像的虚拟试衣方法

技术领域

[0001] 本发明涉及虚拟试衣技术领域,并且更具体地,涉及一种基于2D图像的虚拟试衣方法。

背景技术

[0002] 近年来,随着数字技术的快速发展和人工智能技术的不断升级,为虚拟试衣技术的发展提供了重要动力和支撑。“宅经济”悄然兴起,日常购衣也纷纷转至线上。虚拟试衣的出现可以使卖方更客观地展示服饰优缺点,使交易双方可以更为直观地了解信息,促成交易,减少不必要的工作量,提高工作效率,满足用户需求。

[0003] 目前,国内外很多研究机构都在做相关研究,但是针对2D的合成图像仍然局限于低分辨率,这是影响在线消费者满意度的关键因素。随着图像分辨率的增加,变形后的服装与人体之间未对齐区域的伪影在最终结果中变得明显,最终结果中的身体部位和衣服的纹理清晰度方面表现不佳,严重影响了试衣效果。

[0004] 公开号为CN 112232914A的中国专利公开了“基于2D图像的四阶段虚拟试衣方法及装置”,通过预测试衣后的语义分割图和手臂图像,来辅助生成试穿合成图像,解决了现有技术中存在的吻合有伪影的问题,但该技术不能实现在高分辨2D图像的虚拟试衣,对高分辨率图像的纹理清晰度保持效果不佳。

发明内容

[0005] 针对现有技术的以上缺陷或改进需求,本发明提供了一种基于2D图像的虚拟试衣方法,其目的在于解决高分辨率试衣图像中出现伪影的问题,保持更多的纹理细节,提高试衣效果。

[0006] 为实现上述目的,按照本发明的一个方面,提供了一种基于2D图像的虚拟试衣方法,包括如下步骤:

[0007] (1)输入试穿者的2D图像、人体语义分割图和服装图像,经过预处理操作将人体语义分割图分为试穿语义区域和非试穿语义区域两部分,对服装图像进行掩模和修正处理,生成服装掩模图和服装修正图;

[0008] (2)输入服装语义图和非试穿区域语义图,经过端到端的语义拼接网络,在引入集成判别器和多层次损失函数的约束下,生成试穿后的人体语义分割图;

[0009] (3)根据试穿后的语义分割图、人体姿态图和服装图像,通过服装变形网络,在引入频域分类器和正则约束损失函数的约束下,预测出变形后的服装图像;

[0010] (4)根据试穿者图像,对非试穿区域语义图进行着色得到非试穿区域彩色图,再与变形后的服装图像进行拼接得到粗糙的试穿结果图;

[0011] (5)将粗糙的试穿结果图和试穿后的语义分割图输入到归一化网络中,计算并缩小伪影区域,最终得到精细的试穿结果图。

[0012] 优选的,所述的一种基于2D图像的虚拟试衣方法中步骤(1)的预处理过程包括两

部分：一部分，输入人体语义分割图，包含7种语义信息依次为：头发、面部、脖子、手部、下装、衣服和手臂的语义信息，并且各语义信息使用不同像素值进行标注可表示为： $Y = \{a_1, a_2, a_3, \dots, a_7\}$ ，在预处理过程中根据各语义信息像素值的不同，将人体语义分割图划分为非试穿区域语义图： $Y_1 = \{a_1, a_2, a_3, a_4, a_5\}$ 和试穿区域语义图： $Y_2 = \{a_6, a_7\}$ ，其中，非试穿区域语义图包括：头发、面部、脖子、手部、下装的语义信息，试穿区域语义图包括：衣服和手臂的语义信息；另一部分，输入服装图像，对服装图像进行掩膜处理，即将服装图像中的服装区域的像素值设置为0，背景的像素值设置为255，得到服装掩模图，根据试穿区域语义图中像素点的位置信息对服装图像进行仿射变换，使得服装图像的轮廓与试穿区域语义图的轮廓一致，从而得到服装修正图，为后面的服装变形做准备。

[0013] 优选的，所述的一种基于2D图像的虚拟试衣方法中语义拼接网络是端到端的网络，输入是服装语义图和非试穿区域语义图，先经过卷积层提取图像特征并进行下采样，再经过反卷积层将压缩后的图像扩展成原本的尺寸，得到粗糙的试穿后的人体语义分割图，最后将粗糙的试穿后的人体语义分割图输入到集成判别器进行相似度计算，经过多轮语义拼接后输出相似度最高的试穿后的人体语义分割图。在语义拼接过程中引入多层次损失函数，对预测结果进行细粒度优化，鼓励语义拼接网络生成更精细的细节。

[0014] 优选的，集成判别器是由n个相同网络结构的深度卷积网络拼接而成，n的大小与输入图片的分辨率有关，深度卷积网络由池化层、卷积层、归一化层和激活层依次连接构成。集成判别器的输入是真实图片和预测图片，输出是两个图片的特征相似度。n个深度卷积网络分别在n个不同尺度上区分真实图像和预测图像，并且通过设置系数来控制每个深度卷积网络的预测结果对最终结果的影响，从而鼓励语义拼接网络生成更多的细节。每个深度卷积网络对应一个特征匹配损失函数，合起来构成集成判别器的多层次损失函数。

[0015] 优选的，多层次损失函数表示为：

$$[0016] \quad \mathcal{L}_{mul}(D_1, D_2, \dots, D_n) = \gamma_1 \mathcal{L}_{FM}(G(1), D_1) + \gamma_2 \mathcal{L}_{FM}(G(2), D_2) + \dots + \gamma_n \mathcal{L}_{FM}(G(n), D_n)$$

[0017] 上式中， $\mathcal{L}_{FM}(G(1), D_1)$ 表示深度卷积网络 D_1 对语义拼接网络生成的 $G(1)$ 图像的特征匹配损失函数， $\gamma_1, \gamma_2, \dots, \gamma_n$ 是系数，控制不同尺度的深度卷积网络损失对多层次损失函数的影响，即 $\gamma_1 + \gamma_2 + \dots + \gamma_n = 1$ 。

[0018] 优选的，所述的一种基于2D图像的虚拟试衣方法中服装变形网络是端到端的网络，输入是试穿后的语义分割图、服装图像和人体姿态图，在人体姿态图中18个关键点位置的约束下，经过两条支路分别提取试穿后的语义分割图和服装图像的关键点位置的像素信息，引入频域分类器，使用插值法预测变形后服装图像的像素信息，引入正则约束损失函数控制服装的变形程度，最后输出变形后的服装图像。

[0019] 正则约束损失函数表示为：

$$[0020] \quad \mathcal{L}_{nom}(W, M) = \sqrt{\alpha \sum_x \sum_y (W(x) - M(x))^2 + \beta \sum_x \sum_y (W(y) - M(y))^2}$$

[0021] 上式中， $W(x), M(x)$ 分别表示变形前后坐标点的x值，

[0022] $W(y), M(y)$ 分别表示变形前后坐标点的 y 值,

[0023] $(W(x) - M(x))^2$ 表示 x 轴方向上坐标的变化量的平方,

[0024] $(W(y) - M(y))^2$ 表示 y 方向上坐标的变化量, α, β 为两个可学习的参数, 用于控制服装在 x 轴和 y 轴方向的变形程度;

[0025] 服装变形网络的损失函数表示为:

$$[0026] \quad \mathcal{L}_c = \lambda_1 \mathcal{L}_{CGAN} + \lambda_2 \mathcal{L}_{FM} + \lambda_3 \mathcal{L}_{nom}$$

[0027] 上式中, \mathcal{L}_{CGAN} 是条件对抗损失函数, \mathcal{L}_{FM} 是像素交叉熵损失函数, \mathcal{L}_{nom} 是正则约束损失函数, $\lambda_1, \lambda_2, \lambda_3$ 是系数, 用于调节三个损失函数对总损失函数的影响。

[0028] 优选的, 频域分类器指在频域空间, 通过设置阈值来对特征进行划分, 一方面更多地把关注点放在生成低频特征信息上, 增强图像的内容信息, 另一方面借助试穿后的语义分割图区别开噪点和高频信息, 提取更多有用的高频信息, 使用插值法增强图像高频区域的纹理精细度, 从而提高生成图片的质量。频域分类器的具体流程包括:

[0029] 首先, 由卷积层提取服装图像的空间域特征图, 使用离散傅里叶变换得到频域特征图, 可表示为:

$$[0030] \quad F_{i,j} = DFT(S_{i,j})$$

[0031] 其中, $S_{i,j}$ 表示空间域特征图中第 i 行第 j 列的空间特征分量, $F_{i,j}$ 表示频域特征图中第 i 行第 j 列的频率特征分量, $DFT()$ 是离散傅里叶变换函数;

[0032] 然后, 通过神经网络结构学习得到频域特征图中每个频率特征分量的重要性权值, 将权值大于阈值的频率特征分量标识为低频信息特征分量, 小于阈值的频率特征分量标识为高频信息特征分量, 可表示为:

$$[0033] \quad Freq_{i,j,z} = \begin{cases} Freq_{i,j,0}, & k_{i,j} \geq \lambda \\ Freq_{i,j,1}, & k_{i,j} < \lambda \end{cases}$$

[0034] 其中, $Freq_{i,j,z}$ 表示第 i 行第 j 列类别为 z 的频率特征分量, 当 $z=0$ 时表示该频率特征分量为低频信息特征分量, 当 $z=1$ 时表示该频率特征分量为高频信息特征分量, $k_{i,j}$ 表示第 i 行第 j 列频率特征分量的重要性权值, λ 是阈值;

[0035] 最后, 通过试穿后的语义分割图中的位置信息进一步在高频信息特征通道中筛选出噪声信息, 使用低通滤波器来过滤掉图像噪声, 使用插值法填充图像内容和纹理信息。

[0036] 可选的, 样条插值方法包括: 张力样条插值法、规则样条插值法和薄板样条插值法。

[0037] 优选的, 所述的一种基于 2D 图像的虚拟试衣方法中步骤 (4) 包括: 首先提取试穿者图像各像素点值, 并将该值重新赋给对应区域的非试穿区域语义图, 对非试穿区域语义图重新着色得到非试穿区域彩色图, 然后将非试穿区域彩色图与变形后的服装图像一起输入到语义拼接网络, 得到粗糙的试穿结果图。在语义拼接网络中, 先经过卷积层提取非试穿区

域彩色图与变形后的服装图像特征并进行下采样,再经过反卷积层将压缩后的图像扩展成原本的尺寸,得到粗糙的试穿后的人体语义分割图,最后将粗糙的试穿结果图输入到集成判别器进行相似度计算,经过多轮语义拼接后输出相似度最高的试穿结果图。在语义拼接过程中引入多层次损失函数,对预测结果进行细粒度优化,鼓励语义拼接网络生成更精细的细节。

[0038] 优选的,所述的一种基于2D图像的虚拟试衣方法中去除伪影的处理过程包括:首先,通过深度神经网络提取粗糙的试穿结果图的图像特征,经过上采样得到粗糙试穿结果图的语义分割图,与试穿后的语义分割图求差值,该可以表示为:

[0039] $Artifact=R-S$

[0040] 其中, $Artifact$ 表示伪影区域的像素信息, R 表示粗糙的试穿结果图, S 表示试穿后的语义分割图;

[0041] 可以得到一张只有小部分像素值大于0的图像,其中像素值大于0的像素点即为伪影区域的像素点,并根据伪影区域的位置信息将该区域在粗糙的试穿结果图中标识出来;然后,经过多层归一化网络,学习到两个型变系数 α 和 β ,分别控制伪影区域在x轴和y轴方向的缩小程度,用于缩小伪影区域的大小;最后,经过上采样层,将调整后的特征图片上采样成原图大小,并使用插值法对损失的像素点值进行填充,生成更加精细的试穿结果图。

[0042] 可选的,插值法包括:最近邻插值法、双线性插值法和双三次插值法。

[0043] 优选的,归一化网络的输入是粗糙的试穿结果图,输出是精细的试穿结果图,归一化网络包括获取伪影语义图和消除伪影信息两个阶段:

[0044] 第一个获取伪影语义图阶段,包含两个子网络,设为网络1-1、1-2,其中网络1-1中包含n层卷积层,通过卷积层提取图像特征,以及n层反卷积层,通过反卷积将图像映射为原尺寸大小,卷积层的输出作为反卷积层的输入,卷积层与反卷积层之间通过残差结构连接,在最后一层反卷积后输出结果;其中网络1-2包含图像求差模块,输入两种相同尺寸的图像,输出两种图像求差后的结果图。其中n的取值与输入图片的尺寸有关,通常当输入图片的尺寸为 1024×768 时, $n=3,4$ 的实验效果较好。

[0045] 第二个消除伪影信息阶段,包含一个网络2-1,其中网络2-1由两条并联支路构成,其中,第一条支路由一个卷积层,一个激活层、一个区域正则化层、一个下采样,一个上采样层、和一个相加层依次串联而成,第二条支路由单一的区域正则化层构成,在相加层后输出最后的结果。

[0046] 其中,所述区域正则化层,其作用是对伪影区域进行标准化,首先获得伪影区域的空间位置,然后通过卷积网络计算型变系数 α 和 β ,引入平面变形损失函数来控制伪影区域的缩减程度,防止因过度缩减导致错误修改人体部位的像素点信息,以此来辅助缩小伪影区域;所述相加层,其作用是对两条支路的输出结果进行融合。

[0047] 优选的,平面变形损失函数表示为:

[0048]
$$\mathcal{L}_{rnl}(C_x, C_y) = \alpha \sum_{i=-1,1} \sum_x \sum_y |C_x(x+i, y) - C_x(x, y)| + \beta \sum_{i=-1,1} \sum_x \sum_y |C_y(x, y+i) - C_y(x, y)|$$

[0049] 上式中, C_x, C_y 分别表示伪影区域特征点的 x, y 坐标,

$|C_x(x+i, y) - C_x(x, y)|$ 表示两个节点之间的欧氏距离, i, j 是形变量, α, β 是型变系数。

[0050] 总体而言, 通过本发明所构思的以上技术方案与现有技术相比, 能够取得下列有益效果:

[0051] (1) 通过预处理对输入图片进行预处理, 区分试穿区域和非试穿区域, 能够更好的保留非试穿区域的像素信息, 使试穿结果图像保留更多的细节; 对服装图像进行空间上的修正, 便于后续更好的对服装图片进行变形。

[0052] (2) 使用包含更多关键点信息的人体姿态图来辅助生成语义分割图像和服装变形图像, 不仅使得语义分割的精度更高, 而且增加了服装变形的鲁棒性, 引入正则约束损失函数和频域分类器, 解决了因异常扭曲变形导致服装纹理失真的问题。

[0053] (3) 使用归一化网络解决了在高分辨率情况下, 服装与人体不吻合而出现伪影的问题, 大大提高了高分辨率2D图像的虚拟试穿效果, 提高了试穿效果和用户体验。

附图说明

[0054] 图1为本发明实施例提供的一种基于2D图像的虚拟试衣方法的流程示意图;

[0055] 图2为本发明实施例提供的一种基于2D图像的虚拟试衣方法的归一化网络图。

具体实施方式

[0056] 为了使本发明的目的、技术方案及优点更加清楚明白, 以下结合附图及实施例, 对本发明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用以解释本发明, 并不用于限定本发明。此外, 下面所描述的本发明各个实施方式中所涉及到的技术特征只要彼此之间未构成冲突就可以相互组合。

[0057] 如图1所示, 是实施例提供的一种2D图像的虚拟试衣方法流程示意图; 实施例提供的一种2D图像的虚拟试衣方法包括如下步骤:

[0058] (1) 输入试穿者的2D图像、人体语义分割图和服装图像, 经过预处理操作将人体语义分割图分为试穿语义区域和非试穿语义区域两部分, 对服装图像进行掩模和修正处理, 生成服装掩模图和服装修正图;

[0059] 其中, 预处理过程包括两部分: 一部分, 输入人体语义分割图, 包含7种语义信息依次为: 头发、面部、脖子、手部、下装、衣服和手臂的语义信息, 并且各语义信息使用不同像素值进行标注可表示为: $Y = \{a_1, a_2, a_3, \dots, a_7\}$, 在预处理过程中根据各语义信息像素值的不同, 将人体语义分割图划分为非试穿区域语义图: $Y_1 = \{a_1, a_2, a_3, a_4, a_5\}$ 和试穿区域语义图: $Y_2 = \{a_6, a_7\}$, 其中, 非试穿区域语义图包括: 头发、面部、脖子、手部、下装的语义信息, 试穿区域语义图包括: 衣服和手臂的语义信息; 另一部分, 输入服装图像, 对服装图像进行掩模处理, 即将服装图像中的服装区域的像素值设置为0, 背景的像素值设置为255, 得到服装掩模图, 根据试穿区域语义图中像素点的位置信息对服装图像进行仿射变换, 使得服装图像的轮廓与试穿区域语义图的轮廓一致, 从而得到服装修正图, 为后面的服装变形做准备。

[0060] (2) 输入服装语义图和非试穿区域语义图, 经过端到端的语义拼接网络, 在引入集

成判别器和多层次损失函数的约束下,生成试穿后的人体语义分割图,其中服装语义图是提前准备好的,是网络的输入数据;

[0061] 其中,语义拼接网络是端到端的网络,输入是服装语义图和非试穿区域语义图,先经过卷积层提取图像特征并进行下采样,再经过反卷积层将压缩后的图像扩展成原本的尺寸,得到粗糙的试穿后的人体语义分割图,最后将粗糙的试穿后的人体语义分割图输入到集成判别器进行相似度计算,经过多轮语义拼接后输出相似度最高的试穿后的人体语义分割图。在语义拼接过程中引入多层次损失函数,对预测结果进行细粒度优化,鼓励语义拼接网络生成更精细的细节。

[0062] 其中,集成判别器是由n个相同网络结构的深度卷积网络拼接而成,n的大小与输入图片的分辨率有关,通常当输入图片的尺寸为 1024×768 时, $n=3, 4$ 的实验效果较好。深度卷积网络由池化层、卷积层、归一化层和激活层依次连接构成。集成判别器的输入是真实图片和预测图片,输出是两个图片的特征相似度。n个深度卷积网络分别在n个不同尺度上区分真实图像和预测图像,并且通过设置系数来控制每个深度卷积网络的预测结果对最终结果的影响,从而鼓励语义拼接网络生成更多的细节。每个深度卷积网络对应一个特征匹配损失函数,合起来构成集成判别器的多层次损失函数。

[0063] 其中,多层次损失函数表示为:

$$[0064] \quad \mathcal{L}_{mul}(D_1, D_2, \dots, D_n) = \gamma_1 \mathcal{L}_{FM}(G(1), D_1) + \gamma_2 \mathcal{L}_{FM}(G(2), D_2) + \dots + \gamma_n \mathcal{L}_{FM}(G(n), D_n)$$

[0065] 上式中, $\mathcal{L}_{FM}(G(1), D_1)$ 表示深度卷积网络 D_1 对语义拼接网络生成的 $G(1)$ 图像的特征匹配损失函数(该损失函数为现有的), $\gamma_1, \gamma_2, \dots, \gamma_n$ 是系数,控制不同尺度的深度卷积网络损失对多层次损失函数的影响,即 $\gamma_1 + \gamma_2 + \dots + \gamma_n = 1$ 。

[0066] (3) 根据试穿后的语义分割图、人体姿态图和服装图像,通过服装变形网络,在引入频域分类器和正则约束损失函数的约束下,预测出变形后的服装图像,其中人体姿态图是提前准备好的,是网络的输入数据;

[0067] 其中,服装变形网络是端到端的网络,输入是试穿后的语义分割图、服装图像和人体姿态图,在人体姿态图中18个关键点位置的约束下,经过两条支路分别提取试穿后的语义分割图和服装图像的关键点位置的像素信息,引入频域分类器,使用插值法预测变形后服装图像的像素信息,引入正则约束损失函数控制服装的变形程度,最后输出变形后的服装图像,其中18个关键点包括:鼻子、颈部、右肩、右肘、右手腕、左肩、左肘、左手腕、右臀部、右膝盖、右脚踝、左臀部、左膝盖、左脚踝、右眼、左眼、右耳、左耳。

[0068] 其中,频域分类器指在频域空间,通过设置阈值来对特征进行划分,一方面更多地把关注点放在生成低频特征信息上,增强图像的内容信息,另一方面借助试穿后的语义分割图区别开噪点和高频信息,提取更多有用的高频信息,使用插值法增强图像高频区域的纹理精细度,从而提高生成图片的质量。频域分类器的具体流程包括:

[0069] 首先,由卷积层提取服装图像的空间域特征图,使用离散傅里叶变换得到频域特征图,可表示为:

$$[0070] \quad F_{i,j} = DFT(S_{i,j})$$

[0071] 其中, $S_{i,j}$ 表示空间域特征图中第 i 行第 j 列的空间特征分量, $F_{i,j}$ 表示频域特征图中第 i 行第 j 列的频率特征分量, $DFT()$ 是离散傅里叶变换函数;

[0072] 然后, 通过神经网络结构学习得到频域特征图中每个频率特征分量的重要性权值, 将权值大于阈值的频率特征分量标识为低频信息特征分量, 小于阈值的频率特征分量标识为高频信息特征分量, 可表示为:

$$[0073] \quad Freq_{i,j,z} = \begin{cases} Freq_{i,j,0}, & k_{i,j} \geq \lambda \\ Freq_{i,j,1}, & k_{i,j} < \lambda \end{cases}$$

[0074] 其中, $Freq_{i,j,z}$ 表示第 i 行第 j 列类别为 z 的频率特征分量, 当 $z=0$ 时表示该频率特征分量为低频信息特征分量, 当 $z=1$ 时表示该频率特征分量为高频信息特征分量, $k_{i,j}$ 表示第 i 行第 j 列频率特征分量的重要性权值, λ 是阈值;

[0075] 最后, 通过试穿后的语义分割图中的位置信息进一步在高频信息特征通道中筛选出噪声信息, 使用低通滤波器来过滤掉图像噪声, 使用插值法填充图像内容和纹理信息。

[0076] 其中, 样条插值方法包括: 张力样条插值法、规则样条插值法和薄板样条插值法。

[0077] 其中, 正则约束损失函数表示为:

$$[0078] \quad \mathcal{L}_{nom}(W, M) = \sqrt{\alpha \sum_x \sum_y (W(x) - M(x))^2 + \beta \sum_x \sum_y (W(y) - M(y))^2}$$

[0079] 上式中, $W(x)$, $M(x)$ 分别表示变形前后坐标点的 x 值, $W(y)$, $M(y)$ 分别表示变形前后坐标点的 y 值, $(W(x) - M(x))^2$ 表示 x 轴方向上坐标的变化量的平方, $(W(y) - M(y))^2$ 表示 y 方向上坐标的变化量, α, β 为两个可学习的参数, 用于控制服装在 x 轴和 y 轴方向的变形程度;

[0080] 其中, 服装变形网络的损失函数表示为:

$$[0081] \quad \mathcal{L}_c = \lambda_1 \mathcal{L}_{CGAN} + \lambda_2 \mathcal{L}_{FM} + \lambda_3 \mathcal{L}_{nom}$$

[0082] 上式中, \mathcal{L}_{CGAN} 是条件对抗损失函数 (该损失函数为现有的), \mathcal{L}_{FM} 是像素交叉熵损失函数 (该损失函数为现有的), \mathcal{L}_{nom} 是正则约束损失函数, $\lambda_1, \lambda_2, \lambda_3$ 是系数, 用于调节三个损失函数对总损失函数的影响。

[0083] (4) 根据试穿者图像, 对非试穿区域语义图进行着色得到非试穿区域彩色图, 再与变形后的服装图像进行拼接得到粗糙的试穿结果图。

[0084] 首先提取试穿者图像各像素点值, 并将该值重新赋给对应区域的非试穿区域语义图, 对非试穿区域语义图重新着色得到非试穿区域彩色图, 然后将非试穿区域彩色图与变形后的服装图像一起输入到语义拼接网络, 得到粗糙的试穿结果图。在语义拼接网络中, 先经过卷积层提取非试穿区域彩色图与变形后的服装图像特征并进行下采样, 再经过反卷积

层将压缩后的图像扩展成原本的尺寸,得到粗糙的试穿后的人体语义分割图,最后将粗糙的试穿结果图输入到集成判别器进行相似度计算,经过多轮语义拼接后输出相似度最高的试穿结果图。在语义拼接过程中引入多层次损失函数,对预测结果进行细粒度优化,鼓励语义拼接网络生成更精细的细节。

[0085] (5)将粗糙的试穿结果图和试穿后的语义分割图输入到归一化网络中,计算并缩小伪影区域,最终得到精细的试穿结果图。

[0086] 其中,去除伪影的处理过程包括:首先,通过深度神经网络提取粗糙的试穿结果图的图像特征,经过上采样得到粗糙试穿结果图的语义分割图,与试穿后的语义分割图求差值,该过程可以表示为:

[0087] $Artifact=R-S$

[0088] 其中, $Artifact$ 表示伪影区域的像素信息, R 表示粗糙的试穿结果图, S 表示试穿后的语义分割图;

[0089] 可以得到一张只有小部分像素值大于0的图像,其中像素值大于0的像素点即为伪影区域的像素点,并根据伪影区域的位置信息将该区域在粗糙的试穿结果图中标识出来;然后,经过归一化网络,学习到两个型变系数 α 和 β ,分别控制伪影区域在x轴和y轴方向的缩小程度,用于缩小伪影区域的大小;最后,经过上采样层,将调整后的特征图片上采样成原图大小,并使用插值法对损失的像素点值进行填充,生成更加精细的试穿结果图。

[0090] 其中,插值法包括:最近邻插值法、双线性插值法和双三次插值法。

[0091] 其中,归一化网络的输入是粗糙的试穿结果图,输出是精细的试穿结果图,归一化网络包括获取伪影语义图和消除伪影信息两个阶段:

[0092] 第一个获取伪影语义图阶段,包含两个子网络,设为网络1-1、1-2,其中网络1-1中包含n层卷积层,通过卷积层提取图像特征,以及n层反卷积层,通过反卷积将图像映射为原尺寸大小,卷积层的输出作为反卷积层的输入,卷积层与反卷积层之间通过残差结构连接,在最后一层反卷积后输出结果;其中网络1-2包含图像求差模块,输入两种相同尺寸的图像,输出两种图像求差后的结果图。其中n的取值与输入图片的尺寸有关,通常当输入图片的尺寸为 1024×768 时, $n=3,4$ 的实验效果较好。

[0093] 第二个消除伪影信息阶段,包含一个网络2-1,其中网络2-1由两条并联支路构成,其中,第一条支路由一个卷积层,一个激活层、一个区域正则化层、一个下采样,一个上采样层、和一个相加层依次串联而成,第二条支路由单一的区域正则化层构成,在相加层后输出最后的结果。

[0094] 其中,所述区域正则化层,其作用是对伪影区域进行标准化,首先获得伪影区域的空间位置,然后通过卷积网络计算型变系数 α 和 β ,引入平面变形损失函数来控制伪影区域的缩减程度,防止因过度缩减导致错误修改人体部位的像素点信息,以此来辅助缩小伪影区域;所述相加层,其作用是对两条支路的输出结果进行融合。

[0095] 其中,平面变形损失函数表示为:

$$[0096] \quad \mathcal{L}_{rnl}(C_x, C_y) = \alpha \sum_{i=-1,1} \sum_x \sum_y |C_x(x+i, y) - C_x(x, y)| + \beta \sum_{i=-1,1} \sum_x \sum_y |C_y(x, y+j) - C_y(x, y)|$$

[0097] 上式中, C_x, C_y 分别表示伪影区域特征点的 x, y 坐标, $|C_x(x+i, y) - C_x(x, y)|$ 表示两个节点之间的欧氏距离, i, j 是形变量, α, β 是型变系数。

[0098] 如图2所示,是实施例提供的一种2D图像的虚拟试衣方法的归一化网络图;

[0099] 归一化网络是输入是粗糙的试穿结果图,输出是去除伪影后的精细的试穿结果图。归一化网络有两个阶段构成:第一个阶段用来获取伪影语义图,包含3层卷积层、3层反卷积层和求差模块;第二个阶段用来消除伪影信息,包含两个支路,其中一条支路包括卷积层、激活层、区域正则化层、下采样层和上采样层,另一条支路只包括区域正则化层,最后通过相加层将两条支路的结果进行融合。

[0100] 本发明不仅使得语义分割的精度更高,而且增加了服装变形的鲁棒性,使试穿结果图像保留更多的细节大大提高了高分辨率2D图像的虚拟试穿效果,提高了试穿效果和用户体验。

[0101] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。本申请实施例中的方案可以采用各种计算机语言实现,例如,面向对象的程序设计语言Java和直译式脚本语言JavaScript等。

[0102] 尽管已描述了本申请的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本申请范围的所有变更和修改。

[0103] 显然,本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样,倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内,则本申请也意图包含这些改动和变型在内。

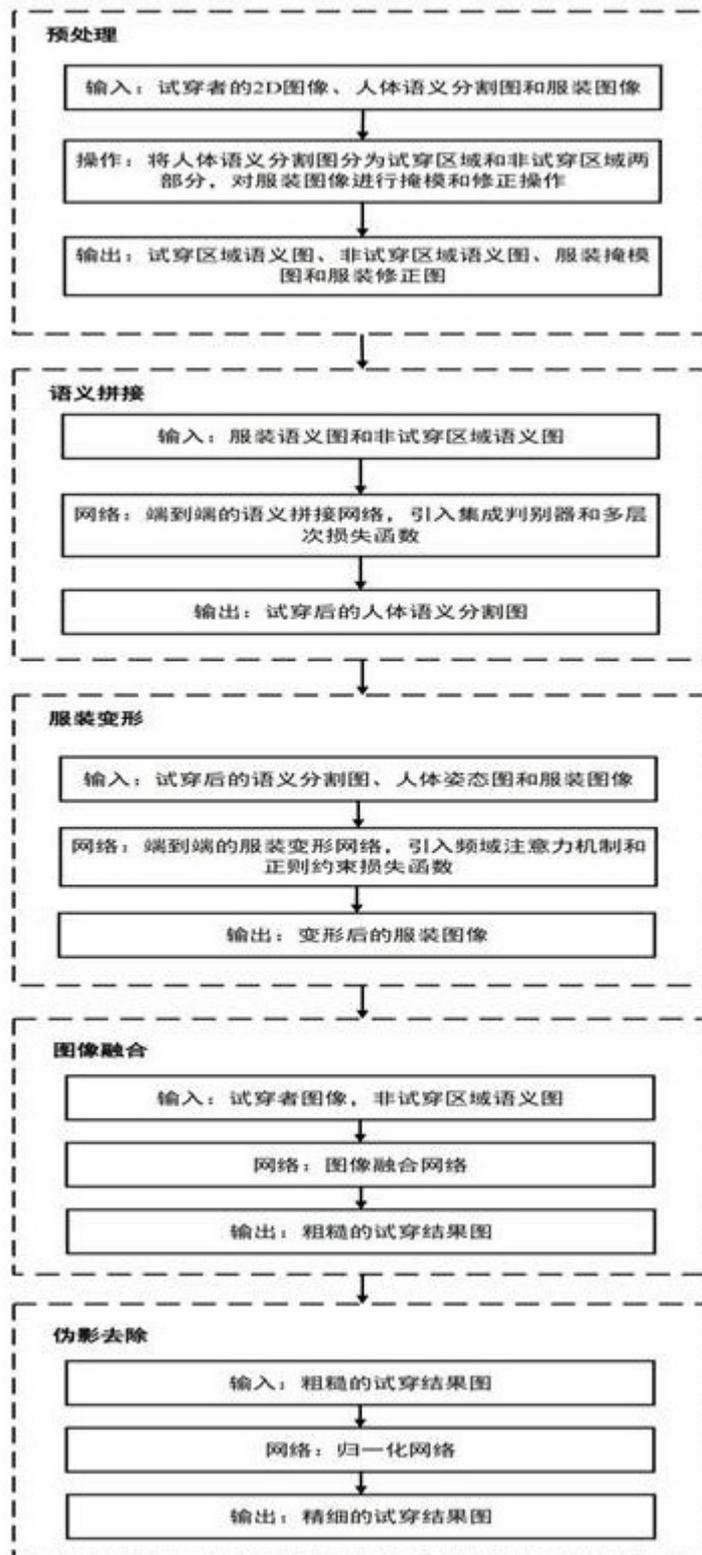


图1

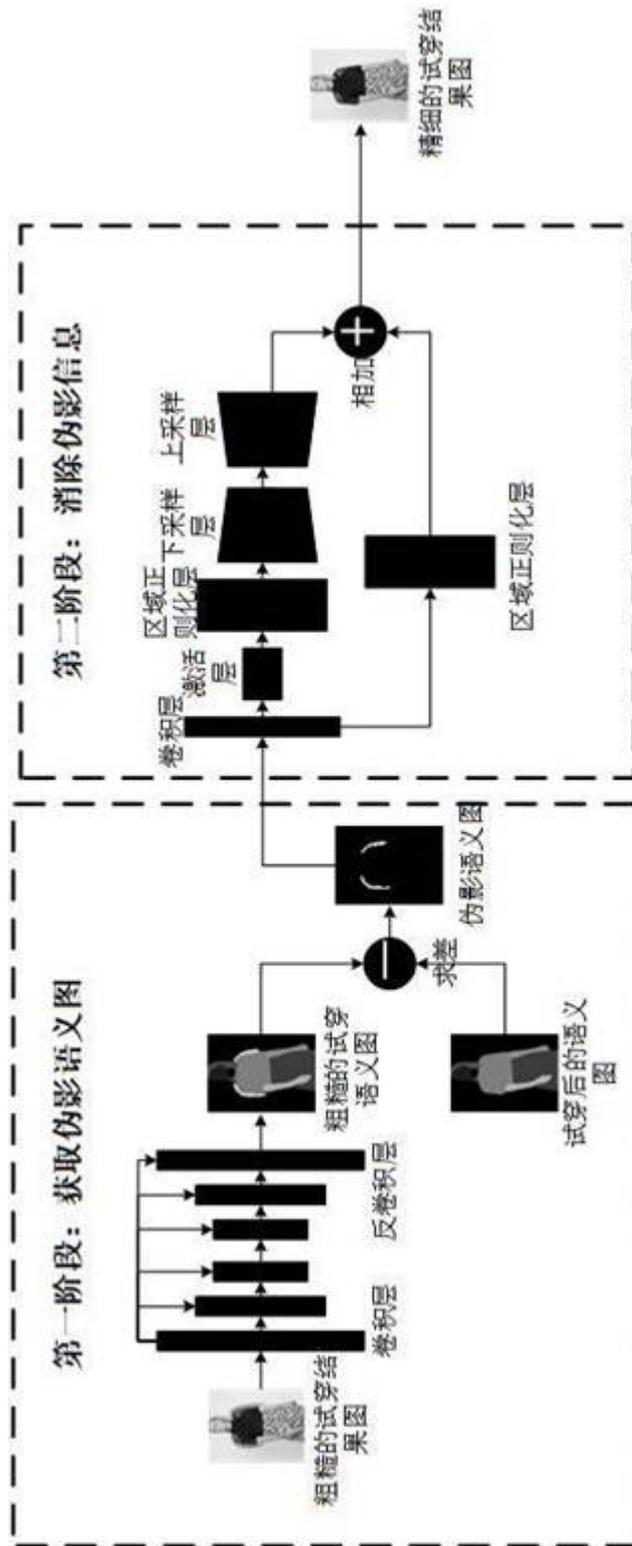


图2