

文章编号: 2095-2163(2022)12-0202-06

中图分类号: TP391.41

文献标志码: A

# 融合衣物特征的人体姿态迁移方法

李和彬, 丁纪峰

(大连民族大学 信息与通信工程学院, 辽宁 大连 116605)

**摘要:** 人体姿态迁移任务即将输入图片中的人物姿势转变为目标姿势,同时保证图片中衣物等外貌特征不变。本文提出了一种新型的生成式对抗网络用来生成人体姿态迁移后的图片,对抗网络的生成器使用注意力机制和风格迁移方法交替对人的外貌信息和姿势信息进行更新。此外,本文较以往的研究还增加了人物衣物的平铺图片作为网络输入,以此来补充人体姿态迁移过程中可能丢失的外貌信息,网络能够使生成图片中人物所穿衣物更加接近真实样式。通过大量实验证明,本文提出的网络在量化分数上获得了更高得分,图片的肉眼观感也更真实自然。

**关键词:** 姿态迁移; 注意力机制; 生成式对抗网络

## Design of human posture migration network based on clothing features classification

LI Hebin, DING Jifeng

(School of Information and Communication Engineering, Dalian Minzu University, Dalian Liaoning 116605, China)

**[Abstract]** The human pose transfer task is to change the human body in the input picture into the target posture, while ensuring that the appearance features such as clothes remain unchanged. In this paper, a new generative adversarial network is proposed to generate pictures of human posture after migration. The network generator uses attention mechanism and Style transfer block to update the appearance information and posture information alternately. In addition, compared with the previous research, this paper also adds the tile pictures of people's clothes as the network input to supplement the appearance information that may be lost during the migration of human posture. The network can more realistically and naturally restore the clothing style worn by people. Through a large number of experiments, it is proved that the network proposed in this paper shows a higher score in the quantitative score, and the visual perception of the picture is more realistic and natural.

**[Key words]** pose transfer; attention mechanism; Generative Adversarial Network

## 0 引言

人体姿态迁移任务近年来在计算机视觉领域引起了广泛的研究与关注。姿态迁移最初由 Ma 等人<sup>[1]</sup>首次提出,是指将输入图片中的人物姿势转变为目标姿势,同时保持人物的外貌不变。现已应用在众多场景中,如摄影编辑<sup>[2]</sup>和行人重识别<sup>[3-4]</sup>等任务。迄今为止该项技术已取得长足进步,如 Zhu 等人<sup>[4]</sup>提出了一种注意力转移模块应用于人体迁移任务,所研发网络能够较为准确地生成指定姿势的人物图片。

目前,虽然基于 GAN 网络的姿态迁移模型陆续涌现,并都能够较为准确地迁移人体姿势,但这些模型往往会出现人物外貌还原得不够真实的情况,尤其是在人物衣物的还原上,很多模型只能还原出大面积色块,而无法还原出衣物的细节纹理。究其原因,

往往是由于在姿态迁移任务中,同一个人不同姿势所展示的外貌信息有很大区别。因为角度不适合或衣物被遮挡的原因,仅靠一张输入图片很难完全包含人物衣物的所有细节。这种信息丢失会给神经网络的学习过程造成一定困难,使得即便如神经网络一般具有强大的学习能力也很难做到将衣物细节全部还原到位。

综上所述,本文提出了一种新型生成式对抗网络。该网络与传统网络的不同之处在于,传统网络中使用 2 张图片,即人物图片和姿态图片作为输入,研究中选用网络是将人物衣物的平铺图片作为网络的第三种输入,以此来补充姿态迁移过程中可能丢失的衣物信息。网络整体分为人物图片生成器、人物外貌鉴别器和人物姿势鉴别器。生成器具体结构如图 1 所示。图 1 中,分支一主要使用注意力机制,利用分支二提供的姿势信息对自身的

**作者简介:** 李和彬(1998-),男,硕士研究生,主要研究方向:人工智能、计算机视觉;丁纪峰(1978-),男,副教授,硕士生导师,主要研究方向:智能信息处理、计算机视觉。

**通讯作者:** 丁纪峰 Email: 535377131@qq.com

收稿日期: 2022-08-06

外貌信息进行更新,分支二则使用衣物特征融合模块,利用衣物的平铺图片信息以及分支一的外貌信息对自身的姿势信息进行更新。上下两条支路相互不停地指导对方更新信息,最后对分支一的外貌信息进行反卷积就得到了目标姿势的人物图片。

由于实验所用数据集除了人物图片外,还需要人物所穿衣物的平铺图片,现今网络中并无满足此要求的公开数据集,因此实验用数据集为本文在研

究过程中自行采集完成。本文通过该数据集在之后的实验中证明了网络的有效性。本文的贡献可总结为如下 2 个方面:

(1) 提出了一种新的思路用以补充姿态迁移任务中输入图片可能出现的外貌信息丢失。

(2) 提出了一种新型 GAN 网络,能够在人体姿态迁移任务中更真实地还原人物所穿衣物,并在实验中证明了该网络的有效性。

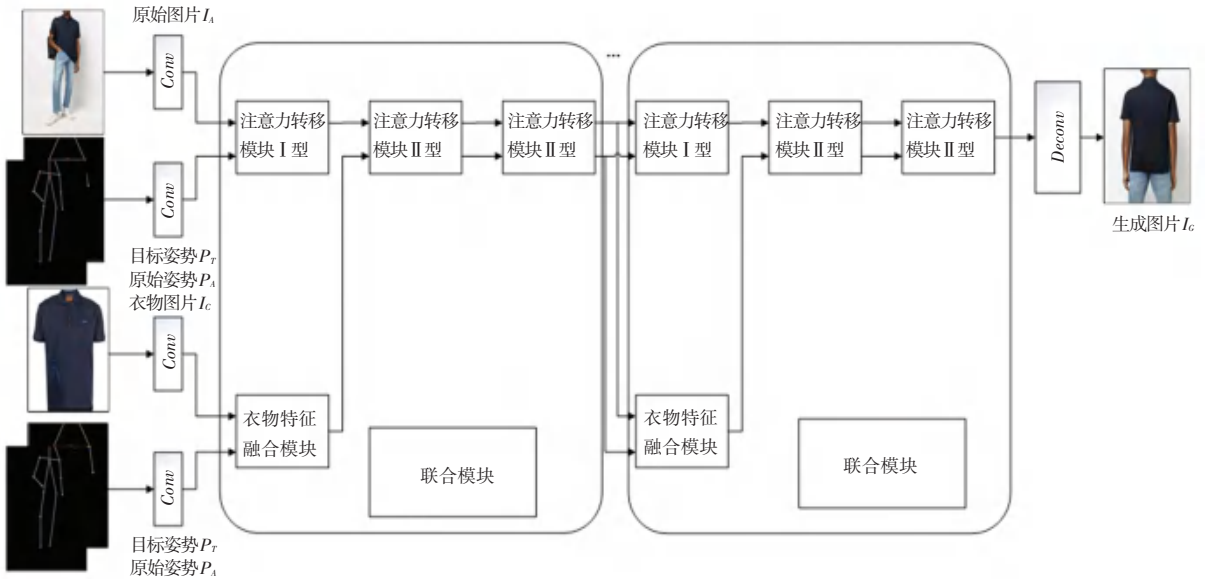


图 1 生成器结构

Fig. 1 The structure of the generator

## 1 相关工作

### 1.1 生成式对抗网络

生成式对抗网络 GAN 在现阶段的图像生成任务中表现出优异的性能<sup>[5-7]</sup>。GAN 网络由生成器和鉴别器组成<sup>[8]</sup>,其中生成器主要是产生目标图片并通过鉴别器的检测,鉴别器则着重于在鉴别的过程中判断出输入图片为真实图片、还是生成图片。生成器和鉴别器不断相互对抗,从而不断提升生成器的性能,最终生成理想图片。但是,原始 GAN 的缺陷也很明显,因其生成的图片有着较强的随机性,无法控制生成指定的图片,基于此,由 Mirza 等人<sup>[9]</sup>提出了条件 GAN、即 CGAN。CGAN 通过将属性信息输入到生成器与鉴别器中,能够显著增强生成网络的目的性<sup>[10]</sup>。以上 2 种 GAN 均为监督学习,而由 Zhu 等人<sup>[11]</sup>提出的 cycleGAN 是对数据集要求极低的一种无监督学习,cycleGAN 能够在无需成对图片的情况下完成图像域与图像域之间相互转化。能够由图像生成图像的网络还有由 Isola 等人<sup>[12]</sup>提出的

pix2pix 网络,其生成器网络为 U-Net<sup>[13]</sup>结构,能够保存不同分辨率下的像素级信息,适合为图片上色、提高清晰度等工作。

### 1.2 人物图像的生成

生成人物图像的 GAN 网络则不宜直接使用传统 GAN 结构,否则容易出现不可控制的人物形变。针对这个缺陷,现在生成人物图像的 GAN 网络往往会加入姿态信息作为限制。Ma 等人<sup>[1]</sup>首次提出了这个设想,研究构建的 PG2 网络首先使用了人体姿态信息用以引导人物图像生成。但由于姿态迁移代表着人物的大范围形变,若要直接得到理想图片,对网络的学习能力将会有很高的要求,随即又进一步提出了将任务分为两阶段的想法。Balakrishnan 等人<sup>[14]</sup>构建了一种 GAN 网络,能够将人物图像生成任务分为前景生成和背景生成两个阶段,最终再将前景和背景融合成目标图片。Ma 等人<sup>[1]</sup>提出的网络则将迁移任务分为粗图像生成阶段和精细图像生成阶段。其中,网络的第一阶段只生成姿势正确、但较为模糊的粗图像,网络的第二阶段则对粗图像进

行精细化处理。

除了网络方面的改进,学者们在人体姿态信息的采集上也取得了重要突破。姿态迁移中的骨骼关键点检测如今普遍使用的是 OpenPose<sup>[15]</sup>,此外还有性能更优的 DensePose<sup>[16]</sup>和 3D Pose<sup>[17]</sup>,由此检测得到的姿势信息中还包含了深度信息,而在训练后生成的图像质量也往往更佳。但是获取 DensePose 等信息的昂贵成本降低了实际上的泛用性,现今姿态迁移任务中的骨骼关键点检测仍然是以 OpenPose 方法为主流。

## 2 融合衣物特征的人体姿态迁移方法的设计

本文设计的生成器网络结构见图 1。网络至关重要的 2 部分为注意力转移模块和衣物特征融合模块。其中,注意力转移模块使用了 Zhu 等人<sup>[4]</sup>提出的注意力机制,利用衣物特征融合模块提供的人物姿态信息进行人物外貌信息的更新。衣物特征融合模块使用了风格迁移的思想<sup>[18]</sup>,使用衣物图片信息

和人物外貌信息对人体姿态信息进行更新。两者交替进行,每 3 个注意力转移模块和 1 个衣物特征融合模块组成 1 个联合模块,本文所用模型总共使用了 3 个联合模块。

网络输入信息包含提供人物外貌信息的原始图片  $I_A$ ,与图片  $I_A$  对应的原始姿态信息  $P_A$ 。图片  $I_T$  为真实目标图片, $P_T$  为与  $I_T$  对应的目标姿势。 $I_C$  为图片  $I_A$  中人物所穿衣物的平铺图片。 $I_A$  输入整体改观信息, $I_C$  补充衣物外观信息,网络在保持外观不变的同时完成姿势由  $P_A$  到  $P_T$  的转换,最终生成目标图片  $I_C$ 。

### 2.1 注意力转移模块

参考文献[4]的思想,本文设计的注意力转移模块如图 2 所示。注意力转移模块的输入为外貌信息  $I_{t-1}$  和姿势信息  $P_{t-1}$ 。最初姿势信息  $P_{t-1}$  是由图 1 中的原始姿势  $P_A$  和目标姿势  $P_T$  在深度轴上进行堆叠后,再经过卷积操作得到的。

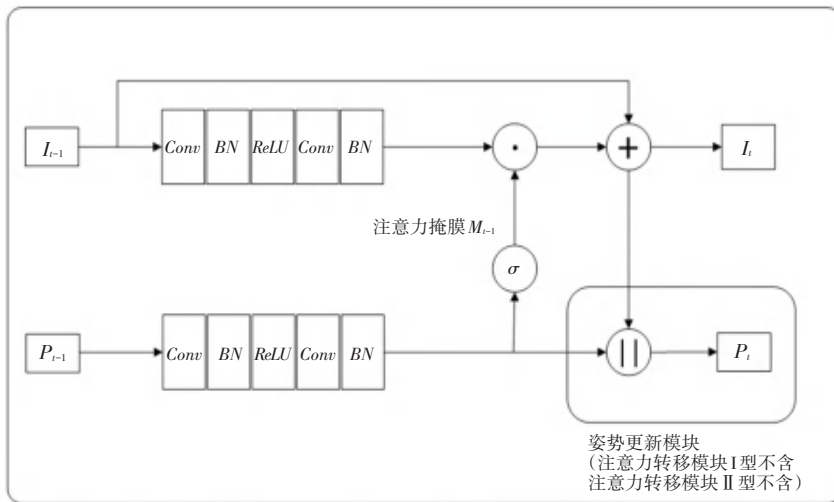


图 2 注意力转移模块结构

Fig. 2 Attention transfer block structure

在本模块中,外貌信息需要接受姿势信息的引导才能得到更新后的数据。而在人体姿态迁移任务中,图片中包含人物的区域显然更为重要,所以通过利用由姿势信息形成的注意力区域可以更好地完成外貌信息的卷积变换过程。

具体来说,输入的姿势信息  $P_{t-1}$  首先经过 2 次卷积操作,接着进行 sigmoid 操作得到注意力掩膜  $M_{t-1}$ 。 $M_{t-1}$  中的每个数值为 0 ~ 1 之间的数字,代表着这一位置上姿势信息的重要程度。然后要使用  $M_{t-1}$  让外貌信息在更新过程中能够关注更重要的信息。其数学公式可以表示为:

$$M_{t-1} = \sigma(\text{conv}(P_{t-1})) \quad (1)$$

得到注意力掩膜后,将经过 2 次卷积操作后的外貌信息  $I_{t-1}$  与注意力掩膜  $M_{t-1}$  进行元素相乘,这样一来外貌信息中包含人物信息的部分得到保留,与此同时则会抑制那些不含人物信息的区域。最后将元素相乘的积与输入的  $I_{t-1}$  进行残差连接,就得到了更新后的外貌信息  $I_t$ 。残差连接是为了防止训练过程中的网络退化现象。此处推导得到的数学公式为:

$$I_t = \text{conv}(I_{t-1}) \cdot M_{t-1} + I_{t-1} \quad (2)$$

以模块输出信息的数量分类,注意力转移模块有 2 种类型,详见图 2。注意力转移模块 I 型不包

含姿势更新模块, 模块输出只有更新后的外貌信息  $I_t$ 。注意力转移模块 II 型则包含姿势更新模块, 模块输出为更新后的外貌信息  $I_t$  和更新后的姿势信息  $P_t$ 。姿势更新模块将外貌和姿势信息进行深度堆叠, 在 II 型中,  $P_t$  可由如下公式进行描述:

$$P_t = conv(P_{t-1}) \parallel I_t \quad (3)$$

## 2.2 衣物特征融合模块

衣物特征融合模块负责对人体姿势信息进行更新, 需要外貌信息进行指导, 网络具体结构如图 3 所示。图 3 中,  $I_{t-1}$  为输入外貌信息,  $P_{t-1}$  为输入姿势信息。参见前文图 1, 衣物特征融合模块在网络中获取的最初的外貌信息为平铺衣物图片经过卷积操作后的数据。同时, 最初获取的姿势信息为  $P_A$  和  $P_T$  进行深度堆叠后再进行卷积操作得到的数据。

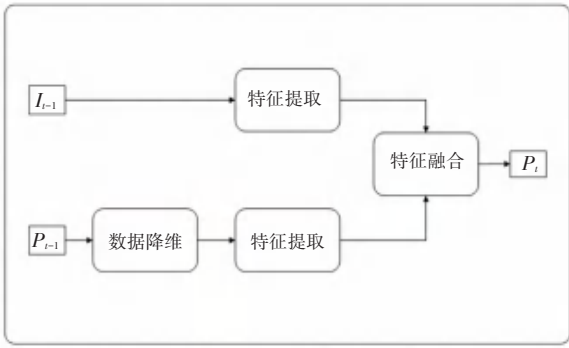


图 3 衣物特征融合模块结构

Fig. 3 Clothing feature integration block structure

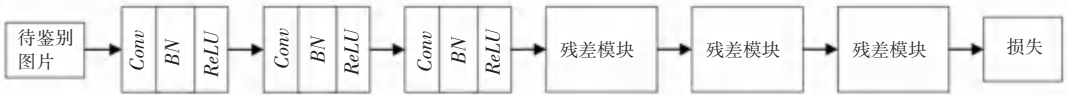


图 4 鉴别器结构

Fig. 4 The structure of the discriminator

## 2.4 损失函数设置

网络的损失函数包含 3 部分, 分别为:  $L_1$  损失、感知损失和对抗损失。对此拟做探讨分述如下。

(1)  $L_1$  损失。是真实目标图片  $I_T$  和生成图片  $I_G$  之间的逐个元素之差, 数学定义公式具体如下:

$$L_1(I_G, I_T) = \frac{1}{m} \sum_{i=0}^m (I_G^i - I_T^i) \quad (5)$$

其中,  $m$  表示图片数据的总元素个数、即图片的  $C$ 、 $H$  和  $W$  之积,  $I_G^i$  和  $I_T^i$  表示生成图片和真实图片的  $i$  个元素。

(2) 感知损失。考虑到只有  $L_1$  损失容易造成图片的失真<sup>[19]</sup>, 本文还一并使用了感知损失。在风格迁移和图像生成任务中经常会用到感知损失, 可以

首先, 模块中输入的姿势信息  $P_{t-1}$  进行数据降维操作, 这是因为  $P_{t-1}$  在网络前方的注意力转移模块中进行过深度堆叠, 维度变为外貌信息的 2 倍, 所以要先进行降维操作。随后, 将同维度的姿势信息和外貌信息用同一个特征提取网络进行特征提取, 将提取到的特征进行特征融合操作。具体做法是: 用姿态特征减去姿态特征的平均值, 再除以姿态特征的标准差。对此可用标准化后的值乘外貌特征的标准差, 再加上外貌特征的均值。计算方法如式 (4) 所示:

$$P_t = \sigma(F_I) \frac{F_P - \mu(F_P)}{\sigma(F_P)} + \mu(F_I) \quad (4)$$

其中,  $F_I$  是从外貌信息提取出的特征;  $F_P$  是姿势信息提取出的特征;  $\mu$  为取均值操作;  $\sigma$  为取标准差操作。

## 2.3 鉴别器网络

鉴别器分为姿势鉴别器  $D_P$  和外貌鉴别器  $D_A$ 。 $D_P$  用于判断生成图片  $I_G$  与目标姿势  $P_T$  的姿势一致性。 $D_A$  用于判断生成图片  $I_G$  与外貌图片  $I_A$  中的人物是否为同一个人。2 种鉴别器的结构相同, 这里给出的鉴别器设计结构如图 4 所示。图 4 中, 主要用到了残差网络。鉴别器网络的训练目标是将真实数据判别为真, 将生成数据判别为假, 如此则有助于生成数据向真实数据的分布靠近。

使生成图片更加平滑自然。本文使用预训练好的 VGG 网络进行真实目标图片  $I_T$  和生成图片  $I_G$  的特征提取, 再对提取出的特征计算感知损失, 数学定义公式具体如下:

$$L_P(I_G, I_T) = \frac{1}{m} \sum_{i=0}^m (\phi(I_G^i) - \phi(I_T^i)) \quad (6)$$

其中,  $\phi$  表示特征提取操作。

(3) 对抗损失。来自于姿势鉴别器  $D_P$  和外貌鉴别器  $D_A$  对真实图片或生成图片鉴别过程中产生的损失。数学定义公式具体如下:

$$L_{GAN} = E_{real \sim P_{data}(real)} [\log(D_P(real)) \cdot \log(D_A(real))] + E_{fake \sim P_g(fake)} [\log(1 - D_P(fake)) \cdot \log(1 - D_A(fake))] \quad (7)$$



其中, *real* 表示真实数据, *fake* 表示生成器生成数据。

综上分析后可得, 网络总损失可用如下公式计算求得:

$$L_{all} = \lambda_{L1} L_1 + \lambda_P L_P + \lambda_{GAN} L_{GAN} \quad (8)$$

其中,  $\lambda_{L1}$ 、 $\lambda_P$ 、 $\lambda_{GAN}$  表示 3 种损失的权重。

## 2.5 训练细节

研究中根据 GAN 网络的训练流程, 交替训练生成器 *G* 和 2 个鉴别器, 即  $D_P$  及  $D_A$ 。将  $(I_G, P_T)$  和  $(I_T, P_T)$  送入姿势鉴别器  $D_P$  中进行判别, 以保证姿势一致性。将  $(I_G, I_A)$  和  $(I_T, I_A)$  送入外貌鉴别器  $D_A$  中进行判别, 以保证外貌一致性。

训练过程使用了 Adam 优化器进行了 500 次训练,  $m$  优化器的  $\beta_1$  设置为 0.5,  $\beta_2$  设置为 0.999。学习率为动态学习率, 前 300 次训练学习率固定为 0.000 5, 后 200 次训练中学习率逐渐下降为 0。

## 3 实验过程

### 3.1 数据获取

实验用数据集的每组数据中需要同一人物的不同姿势和人物所穿衣物的平铺图片, 网络上并无合适的公开数据集, 所以本文数据集来源于自行制作。本文的数据集由 farfetch 服装网站进行采集 (<https://www.farfetch.cn/uk/shopping/men/items.aspx>), 该网站售卖的每件衣服包含了同一模特不同姿势的图片及平铺衣物图片。经过数据清洗后, 本文使用 40 000 组数据进行训练, 每组数据包含了同一模特的 2 张不同姿势的图片, 以及模特所穿衣物的 1 张平铺图片。

### 3.2 模型的定性比较和定量比较

本实验选用 SSIM<sup>[20-21]</sup> 和 IS<sup>[22]</sup> 两种指标分析模型性能, 将本文设计的模型和 Pose-Transfer 模型<sup>[4]</sup> 对测试集中的 5 000 组数据进行了对比。SSIM 指标用以评价生成图片和真实图片的结构相似度, 指标越接近 1、相似度越高。Inception Score、即 IS 指标用以评价生成式对抗网络的图片质量和多样性, 指标越高、说明生成式对抗网络的效果越好。

模型输出图片如图 5 所示。由图 5 可以看出, 本文设计的模型更好地保证了人体结构的合理性, 且在衣物还原部分的能力要优于 Pose-Transfer 模型, 从前 2 组实验结果来看, 本文所设计的模型还原的衣服质感更加偏向于真实衣物。从后 2 组实验结果看, 本文所用模型还原的衣服上面的花纹与真实衣物更加接近。

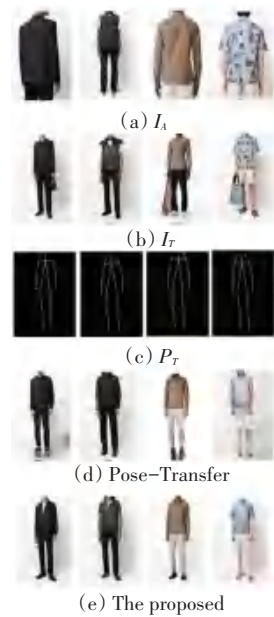


图 5 模型效果对比

Fig. 5 Model effect comparison

2 种模型的 SSIM 和 IS 评分见表 1。表 1 中, SSIM 表示本文所用模型生成的  $I_G$  与目标图片  $I_T$  的相似度更高。Inception Score 分数表示本文生成图片的质量与丰富性更高。

表 1 模型定量比较

Tab. 1 Quantitative comparison of models

使用模型	SSIM	Inception Score
本文设计模型	0.813 7	3.367 0
Pose-Transfer 模型	0.783 5	3.018 1

对于这 2 种方法, 从指标评分的定量比较和肉眼观感的定性比较来看, 本文设计的模型在测试集中的表现要优于 Pose-Transfer 方法, 尤其在衣物的重建方面, 本文设计的模型有着更真实的还原效果。本文取得的研究成果证明在网络中融入衣物特征能使姿态迁移任务的效果得到提升。

### 3.3 消融实验

为了寻找每个联合模块中注意力转移模块和衣物特征融合模块的最佳数量之比, 本文在固定生成器网络共使用 3 个联合模块, 每个联合模块包含 3 个注意力转移模块的条件下, 调整衣物特征融合模块的数量进行实验。实验记录了衣物特征融合模块和注意力转移模块数量之比为 1:1、1:2 和 1:3 情况下的数据指标, 结果如表 2。

由实验数据可看出, 当注意力转移模块和衣物特征融合模块的数量达到 1:1 时, 网络性能出现大幅度下降, 这是因为过多的衣物特征融合模块使网络的复杂度过高, 训练过程中出现了 loss 为 Nan 的

现象。而在此时就需要提前终止网络训练,所以没能达到其他情况的训练效果。而指标显示衣物特征融合模块和注意力转移模块的数量之比为 1:3 时网络效果最佳。

表 2 不同模块比例的性能比较

Tab. 2 Performance comparison of different proportions

衣物特征融合模块: 注意力转移模块	SSIM	Inception Score
1:3	0.813 7	3.367 0
1:2	0.764 9	3.062 1
1:1	0.426 1	2.945 2

## 4 结束语

本文为姿态迁移任务设计了一种新型网络结构,利用注意力机制和风格迁移方法将衣物平铺图片的信息融入到人物生成网络中。网络中的外貌信息与姿势信息相互指导对方更新,使外貌信息在渐进式的更新过程中最终转化为理想图片。与以往的研究相比,本文设计的网络结构能够让最终生成图片中人物所穿的衣物更接近真实图片。研究中已经通过实验证明了本网络生成的图片在清晰度与还原度上的优越性,同时本文取得的成果也说明衣物特征的输入能够提升人体姿态迁移的任务性能。

## 参考文献

[1] MA Liqian, JIA Xu, SUN Qianru, et al. Pose guided person image generation [C]// Advances in Neural Information Processing Systems. Long Beach; NIPS Foundation, 2017, 30: 406-416.

[2] WALKER J, MARINO K, GUPTA A, et al. The pose knows; Video forecasting by generating pose futures[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy; IEEE, 2017: 3332-3341.

[3] QIAN Xuelin, FU Yanwei, WANG Wenxuan, et al. Pose-normalized image generation for person re-identification [C]// Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany; dblp, 2018: 650-667.

[4] ZHU Zhen, HUANG Tengting, SHI Baoguang, et al. Progressive pose attention transfer for person image generation [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA; IEEE, 2019: 2342-2351.

[5] CHEN Ying, XIA Shixiong, ZHAO Jiaqi, et al. Person image synthesis through siamese generative adversarial network [J]. Neurocomputing, 2020, 417: 490-500.

[6] HSU C C, LIN C W, SU Wengtai, et al. Sigan: Siamese generative adversarial network for identity-preserving face hallucination[J]. IEEE Transactions on Image Processing, 2019, 28(12): 6225-6236.

[7] XU Xiaogang, CHEN Yingcong, JIA Jiaya. View independent

generative adversarial network for novel view synthesis [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. South Korea; IEEE, 2019: 7791-7800.

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

[9] MIRZA M, OSINDERO S. Conditional generative adversarial nets [J]. arXiv preprint arXiv:1411.1784, 2014.

[10] ANTIPOV G, BACCOUCHE M, DUGELAY J L. Face aging with conditional generative adversarial networks [C]//2017 IEEE International Conference on Image Processing (ICIP). Beijing, China; IEEE, 2017: 2089-2093.

[11] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Proceedings of the IEEE International Conference on Computer Vision. Italy; IEEE, 2017: 2223-2232.

[12] ISOLA P, ZHU Junyan, ZHOU Tinghui, et al. Image-to-image translation with conditional adversarial networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach; IEEE, 2017: 1125-1134.

[13] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]// International Conference on Medical Image Computing and Computer-assisted Intervention. Cham; Springer, 2015: 234-241.

[14] BALAKRISHNAN G, ZHAO A, DALCA A V, et al. Synthesizing images of humans in unseen poses [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City; IEEE, 2018: 8340-8348.

[15] LI Yining, HUANG Chen, LOY C C. Dense intrinsic appearance flow for human pose transfer [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach; IEEE, 2019: 3693-3702.

[16] GÜLER R A, NEVEROVA N, KOKKINOS I. Densepose: Dense human pose estimation in the wild [C]//Proceedings of the IEEE International Conference on Computer Vision. Italy; IEEE, 2017: 1501-1510.

[17] HUANG Xun, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization [C]//Proceedings of the IEEE International Conference on Computer Vision. Italy; IEEE, 2017: 1501-1510.

[18] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[19] SIAROHIN A, SANGINETO E, LATHUILIERE S, et al. Deformable gans for pose-based human image generation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City; IEEE, 2018: 7297-7306.

[20] HORE A, ZIOU D. Image quality metrics: PSNR vs. SSIM [C]//2010 20th International Conference on Pattern Recognition. Istanbul, Turkey; IEEE, 2010: 2366-2369.

[21] 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.

[22] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training gans [C]// Advances in Neural Information Processing Systems. Barcelona, Spain; NIPS Foundation, 2016, 29: 2234-2242.