Computer Science and Application 计算机科学与应用, 2023, 13(2), 191-203
Published Online February 2023 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2023.132020

Hans汉斯

人脸图像性别转移鲁棒模型研究

卢　维[1,2]，何　强[1,2]

[1]北京建筑大学，理学院，北京
[2]北京建筑大学，大数据建模理论与技术研究所，北京

收稿日期：2023年1月6日；录用日期：2023年2月3日；发布日期：2023年2月13日

摘　要

人脸图像性别转移属于图像风格迁移问题的特例，运用一般的生成对抗网络模型往往不能对人脸部分进行高质量迁移，且无关背景域常常出现扭曲模糊现象，人脸肤色也不能保持原颜色。针对上述问题，本文在基于改进MUNIT的人脸图像性别转换模型的基础上，提出具有鲁棒性质的人脸图像性别转移模型。首先对输入模型的人脸图像进行人脸解析(Face Parsing)，准确将图像中的人脸部分输入到模型中进行训练学习，以解决图像中无关背景域对模型训练的影响；其次构造新的损失函数，将模型生成前后的人脸部分做基于颜色的直方图匹配(Histogram Matching)，从而将人脸性别转移前后的肤色保持一致；最后对公开人脸数据集CeleBA进行属性筛选，以减少人脸遮挡，眼镜等影响模型训练的不利因素，从而提高生成图像的质量。实验结果表明，与其他经典算法相比，本文所提方法可以有效保留图像背景区域以及人脸肤色，并生成效果更好的人脸性别转移图像。

关键词

生成对抗网络，人脸解析，直方图匹配，无监督样式迁移，人脸性别转换

A Robust Model of Gender Transfer in Facial Images

Wei Lu[1,2], Qiang He[1,2]

[1]School of Science, Beijing University of Civil Engineering and Architecture, Beijing
[2]Institute of Big Data Modelling and Technology, Beijing University of Civil Engineering and Architecture, Beijing

Received: Jan. 6th, 2023; accepted: Feb. 3rd, 2023; published: Feb. 13th, 2023

Abstract

Gender transfer of face image is a special case of image style transfer problem. The use of the gen-

文章引用: 卢维, 何强. 人脸图像性别转移鲁棒模型研究[J]. 计算机科学与应用, 2023, 13(2): 191-203.
DOI: 10.12677/csa.2023.132020

eral generative adversity-network model often cannot transfer the face part of high quality, and the irrelevant background domain often appears distorted and fuzzy, and the face skin color can not maintain the original appearance. To solve these problems, this paper proposes a robust face image gender transfer model based on MUNIT's improved face image gender transfer model. Firstly, Face Parsing was performed on the face image input to the model, and the face part of the image was accurately input to the model for training and learning, so as to solve the influence of the irrelevant background domain on the model training. Secondly, a new loss function was constructed to perform color based on Histogram Matching on faces before and after the generation of the model, so as to ensure the consistency of skin color before and after face gender transfer. Finally, attribute screening was carried out on the public face data set CeleBA to reduce the adverse factors affecting model training such as face occlusion and glasses, so as to improve the quality of the generated images. The experimental results show that, compared with other classical algorithms, the proposed method can effectively preserve the background area of the image and human skin color, and generate better facial gender transfer images.

Keywords

Generate Adversarial Networks, Face Parsing, Histogram Matching, Unsupervised Style Transfer, Face Gender Transfer

1. 引言

近年来，随着人工智能技术的不断发展，人们的生活方式发生了极大改变，各种关于人脸的深度学习技术运用到生活的方方面面，如人脸识别[1] [2] [3]，人脸美化[4] [5] [6]以及人脸属性编辑[7] [8] [9]等等。当下各类短视频软件是比较火爆的娱乐方式之一，其内部强大的美颜效果以及各种滤镜都受到用户的追捧，其中关于人脸图像性别转移的滤镜一经发布都能引起社交网络传播热潮，在国内外都受到用户的广泛关注。

人脸图像性别转移就是将男(女)性的人脸在保持原本身份的前提下转换成为女(男)性的人脸，如下图1所示，一般通过生成对抗网络(Generative Adversarial Networks, GAN) [10]实现。人脸图像性别转移任务可以视为风格迁移问题的一种特殊情况。Zhu 等人[11]引入循环一致性损失，将图像从源域转移到目标域，可以在不配对的数据集中间进行图像风格转换，但该方法生成的结果会产生粗糙的纹理，并且内容被过度保留，生成的图像质量不够精细。在 CycleGAN 模型问世后，有许多相关研究人员基于 GAN 网络对风格迁移做了进一步的开发应用。Chen 等人[12]在 2018 年提出对抗性门控网络(Gated GAN)，通过门控转换器使输入图像完成不同风格的迁移工作。Sanakoyeu 等人[13]通过对抗判别器对输入图像进行编码解码，风格化图像集合整体迁移，并利用编码器完成重建损失。陈等人[14]提出将每个风格集中到 StyleBank 部分，在转换新风格图像时只需要重新训练 StyleBank 就能完成风格迁移。Liu 等人[15]通过感知损失以及保留图像对象，提高训练效率，并提出新的目标函数，增加了输出图像的风格多样性。Ma 等人[16]将图像分为内容特征与风格特征，且两者可以完全分离，利用对偶一致性损失来实现语义相关的风格迁移。Huang 等[17]在 2018 年提出多模态无监督图像转换网络(MUNIT)，它将图像的隐藏编码进一步细化为图像内容编码和图像风格编码，通过改变编码的方式来完成图像的风格交换，但对于特定人脸图像性别转

换问题，其图像生成结果仍存在人脸图像模糊，背景图像扭曲，面部身份保留效果不好等缺点。

针对上述问题，本文在基于改进的 MUNIT 人脸图像性别转换模型的基础上，提出一种具有鲁棒性质的人脸图像性别转移模型，并通过实验验证了其有效性。

本文的主要贡献如下：

1) 将输入模型的人脸图像进行 Face Parsing [18]操作，将具体的人脸部分准确输入到模型中进行训练学习，可以有效避免无关背景域对于模型训练学习的影响，同时完好地保留了图像背景部分。

2) 设计人脸肤色损失函数，将模型结果生成前后的人脸部分做基于颜色的 Histogram Matching [19]，以此可以保留人脸原本的肤色，从而提高了性别转换的图像效果。

3) 在训练策略上，对 CelebA 数据集的人脸图像根据属性进行简单的筛选，减少遮挡，眼镜等影响模型生成结果的因素，从而提高图像的生成质量。

本文将具有鲁棒性质的人脸图像性别转移模型在数据集 CelebA 上进行实验，通过主观视觉评价，以及基于内容准确率和结构相似度的客观评价指标，表明了所提方法的先进性。

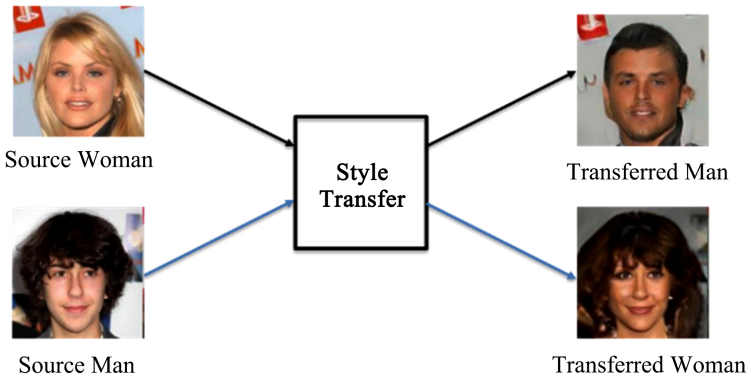


Figure 1. Face image gender transfer
图 1. 人脸图像性别转移

2. 相关知识

2.1. MUNIT 模型

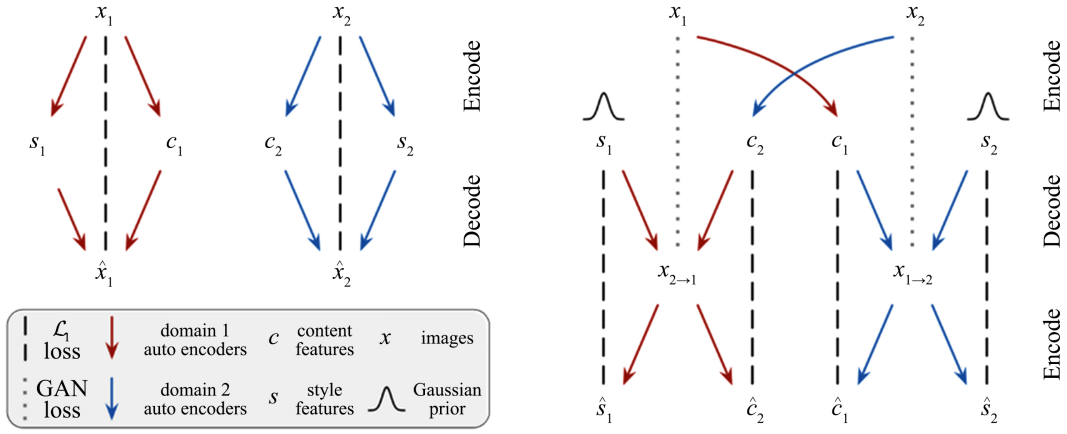
MUNIT 模型脱胎于 Liu 等人提出的非监督图像翻译模型(UNIT) [20]，作者认为风格编码 s 与内容编码 c 为相互独立的图像信息空间。在不同的域之间，内容编码空间是共享的。内容空间中包含一些图像内物体像素级属性，例如边缘信息、相对位置、朝向等信息，而风格编码则蕴含一些风格特征信息例如颜色、纹理等等。假设两个不同的域 $X1$ 和 $X2$ 的风格编码空间分别为 $s1$ 和 $s2$，图像共享的内容编码空间为 $c1$，图像的风格迁移过程如下式所示：

$$P(c1,s2)=G2\left(P(c1),P(s2)\right) \tag{1}$$

其中，$G2$ 为图像风格空间 $s2$ 的风格迁移生成器。编码器通过参数学习分别将风格编码空间 $s2$ 和内容编码空间 $c1$ 从不同的图像域 $X2$ 和 $X1$ 中提取出来。作者假设前两者的分布相互独立，解码器就可以通过参数学习和损失函数的指导，学习到风格分布 $P(s2)$ 和内容分布 $P(c1)$ 的联合分布 $P(c1,s2)$。而学习到的联合分布就是将风格 $s2$ 融合到内容 $c1$ 的风格迁移图像结果。MUNIT 方法可以通过改变不同的风格编码进行多次风格迁移。

MUNIT 网络的风格解码过程如图 2 所示。网络结构与 CycleGAN [10]的循环对称结构类似。图像经

过解码器 E 后生成对应的内容编码 c 和风格编码 s。将不同风格图像的风格编码 s 交换之后，利用生成器 G 还原成图像，完成一次单向的风格迁移过程。通过两个相同且对称的风格迁移过程，风格图像 $x1$ 和 $x2$ 分别变为 $x1 \to 2$ 和 $x2 \to 1$。如图 2 中(a)过程所示，图像需要经过重建损失，即将图像 x 通过编码器生成其对应风格内容编码后再次重新组合，确保生成器的图像生成能力准确，避免出现模式崩塌的情况。



(a) Within-domain reconstruction　　(b) Cross-domain translation

Figure 2. MUNIT style encoding and decoding process (from literature [17])
图 2. MUNIT 风格编解码过程(摘自文献[17])

2.2. MUNIT 模型损失函数

MUNIT 模型的整体网络结构如图 3 所示，我们集合编码器，解码器，鉴别器的 loss，当作最后优化的目标，其为对抗性损失和双向重建损失项的加权和，如下：

$$\begin{aligned}&\min_{E_1,E_2,G_1,G_2}\min_{D_1,D_2} L\left(E_1,E_2,G_1,G_2,D_1,D_2\right)\\&= L_{GAN}^{x1}+L_{GAN}^{x2}+\lambda_x\left(L_{recon}^{x1}+L_{recon}^{x2}\right)+\lambda_c\left(L_{recon}^{c1}+L_{recon}^{c2}\right)+\lambda_s\left(L_{recon}^{s1}+L_{recon}^{s2}\right)\end{aligned} \tag{2}$$

这里的 $\lambda_x,\lambda_c,\lambda_s$ 是控制每项 loss 的权重参数。(2)式中前两项为对抗损失，使用 GANs 来匹配翻译后图像的分布和目标数据的分布。

$$L_{GAN}^{x2}=E_{c1\sim p(c1),s2\sim p(s2)}\left[\log\left(1-D_2\left(G_2\left(c1,s2\right)\right)\right)\right]+E_{x2\sim p(x2)}\left[\log D_2\left(x2\right)\right] \tag{3}$$

这里 D_2 是鉴别生成的图像是否符合域 X_2 的分布，鉴别器 D_1 以及 L_{GAN}^{x1} 有类似定义。(2)式中第三项为图像的重建损失，给定一个从数据分布中采样的图像，我们能够在编码和解码后重建它。

$$L_{recon}^{x1}=E_{x1\sim p(x1)}\left[\left\|G_1\left(E_1^c\left(x_1\right),E_1^s\left(x1\right)\right)-x1\right\|_1\right] \tag{4}$$

(2)式中第四、五项为图像的内容风格损失，给出一个来自于 latent distribution 的 latent code (style 或者 content)，我们能够在编码和解码后重构它。

$$L_{recon}^{c1}=E_{c1\sim p(c1),s2\sim q(s2)}\left[\left\|E_2^c\left(G_2\left(c1,s2\right)\right)-c1\right\|_1\right] \tag{5}$$

$$L_{recon}^{s2}=E_{c1\sim p(c1),s2\sim q(s2)}\left[\left\|E_2^s\left(G_2\left(c1,s2\right)\right)-s2\right\|_1\right] \tag{6}$$

这里的 $q\left(s2\right)$ 表示先验分布 $N\left(0,1\right)$，$p\left(c1\right)$ 是由 $c1=E_1^c\left(x1\right)$ 和 $x1\sim p\left(x1\right)$ 给出。

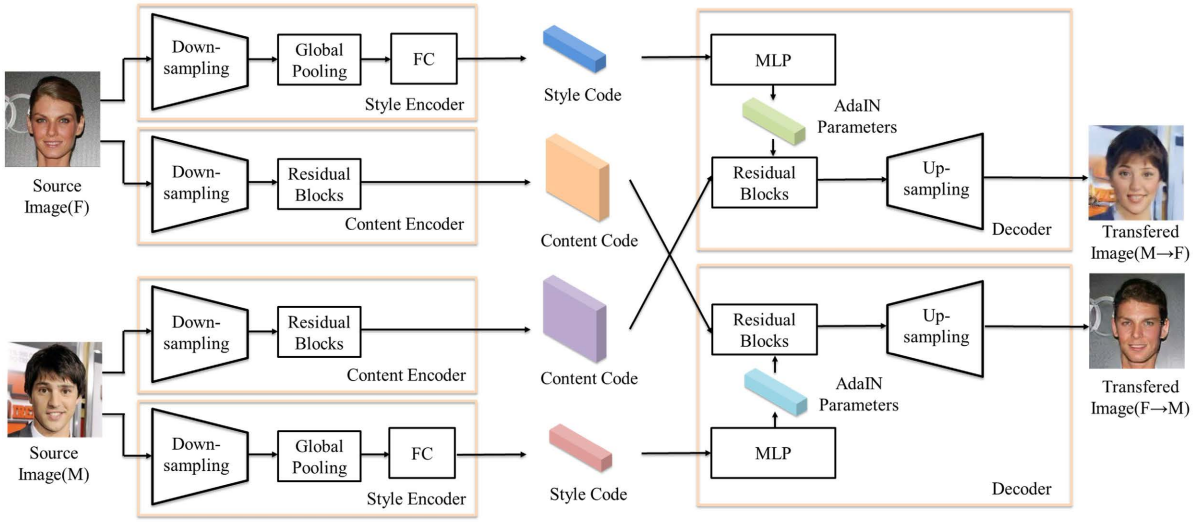


Figure 3. MUNIT model overall network structure
图 3. MUNIT 模型整体网络结构

2.3. 人脸图像风格迁移

关于人脸图像的风格迁移任务最大的难点就是难以得到配对的数据集，因此需要无监督的风格迁移模型来进行训练，一般采用循环生成对抗网络 CycleGAN 或者无监督样式迁移网络 MUNIT 完成迁移任务。Kim 等[21]提出 UGATIT 模型，将辅助分类器得到的特征图输入到注意力模块，以便于更好区分源域和目标域，使模型迁移效果更加优秀，但容易改变图像无关背景。石达等[22]提出来基于改进 CycleGAN 的人脸性别伪造图像生成模型，通过在循环生成对抗网络 CycleGAN 的生成器后加入混合注意力和自适应残差块，结合相对损失函数得到了不错的人脸图像性别转换效果，但仍无法解决无关背景域的影响。Liu 等[23]在多模态无监督图像翻译网络(MUNIT)的基础上引入新的人脸性别概率性掩膜，促进实现性别转移和身份保留的目标，同时通过人脸稀疏特征学习到关于人脸性别的决定性因素，最终获得了较好的性别转换效果，但对于人脸面部颜色，细节的部分仍有改进的空间。由于人脸图像性别转移没有配对数据集的特殊性，本文将基于无监督样式迁移 MUNIT 模型的基础上进行研究。

3. 人脸图像性别转移鲁棒模型

3.1. 总体网络结构

本文方法脱胎于 MUNIT 模型，并在基于改进的 MUNIT 人脸图像性别转换模型的基础上再次做出优化。改进的 MUNIT 模型在生成器部分加入了混合注意力机制 CBAM [24]以及动态实例归一化 DIN [25]，对模型结果背景扭曲现象有了一定的缓解，但是还未完全解决此现象，同时未考虑到人脸图像性别转换前后肤色随机变化的问题，为此我们提出具有鲁棒性质的人脸图像性别转移模型。

完整的改进 MUNIT 网络模型如图 4 所示，网络首先将输入模型的人脸图像进行人脸解析，得到人脸解析图以及存放解析信息的特征图，然后将图像中的人脸部分包括面部五官以及脖子部分输入到模型当中，通过内容编码器和风格编码器分别提取并交换图像的内容特征和风格特征，最终完成人脸图像性别转换过程。其中，风格编码器由下采样部分(Down-sampling)、全局池化层(Global Pooling)和全连接层(Fully Connected, FC)组成；内容编码器由下采样部分，动态实例归一化(DIN)残差模块(Residual Blocks)和混合注意力模块组成。而解码器则是由多层感知机(Multilayer Perceptron)、残差模块和上采样部分(Up-Sampling)构成。

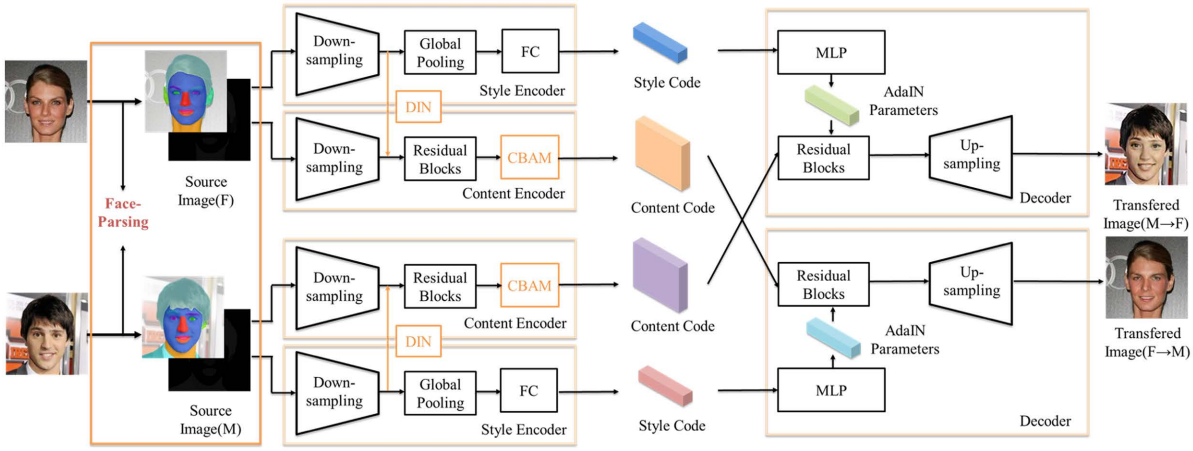


Figure 4. Network structure of robust model of gender transfer in face image
图 4. 人脸图像性别转移鲁棒模型网络结构图

3.2. 人脸解析

人脸解析，是将人的头部包含人脸五官构成进行分解，得到头发、面部皮肤、眼睛、眉毛、鼻子、嘴、耳朵等区域，在深度学习当中可以作为分类任务实现。图像分类任务在深度学习领域的研究已经比较成熟，Yu 等人[26]在 2018 年提出双分支语义分割网络 BiSeNet，采用小步长的 Spatial Path 以及快速下采样的 Context Path，同时引入新的特征融合模块 Feature Fusion Module 来对特征进行合理的融合，最终得到更高速率以及更高准确率的语义分割结果，我们将采用训练好的 BiSeNet 模型完成本文的人脸解析任务。

原图　　人脸解析图　　空间位置图

Figure 5. Face parsing by BiSeNet
图 5. 基于 BiSeNet 的人脸解析图

如上图 5 所示，第一列是输入模型的人脸原图，第二列是经过解析的人脸各部分分布图，第三列是记录各部分标签的空间位置图。通过人脸解析图我们可以将图像中的人脸部分进行精准选择，将人脸部分准确输入到风格迁移模型中进行训练学习，从而避免了无关背景图像对训练的影响，同时可以完美保留图像的背景域。从三组人脸图像解析图可以看出第一行无遮挡的正脸图像分解的效果最为准确，第二三行因为有眼镜，鲜花的遮挡导致部分人脸分析出现错漏，因此输入到改进 MUNIT 模型的人脸数据集最好为正脸无遮挡的人脸图像，才能有最好的人脸图像性别转换效果。

3.3. 人脸肤色损失函数

无监督样式迁移模型 MUNIT 可以对图像做多风格的样式迁移，但对于人脸图像性别迁移任务，该方法无法使模型生成前后的人脸肤色保持一致，因此我们设计了新的损失函数以解决这个问题。若要使模型生成前后的人脸肤色保持一致，我们采用直方图匹配的方法。

直方图匹配，又称直方图规定化，即变换原图的直方图为规定的某种形式的直方图，从而使两幅图像具有类似的色调和反差。我们对模型生成前后的人脸图像做人脸解析，将除去眼睛，眉毛，嘴巴以外的人脸部分提取出来，做基于颜色的直方图匹配，得到具有相同颜色分布的人脸，从而保留了原人脸图像的肤色。

$$L_{face} = \left\| I_s^B - HM\left(I_s^B \circ M_{face}^1, I_r \circ M_{face}^2\right) \right\|_2 \tag{7}$$

$$M^1 = FP\left(I_s^B\right) \tag{8}$$

$$M^2 = FP\left(I_r\right) \tag{9}$$

其中，公式(7)表示人脸肤色损失函数，I_s^B 表示经过迁移的人脸图像，$HM\left(A,B\right)$ 表示对 y 域 A 与域 B 进行直方图匹配，$I_s^B \circ M_{face}$ 表示原图像与解析得到的人脸 mask 相乘提取图像中的人脸区域。公式(8)，(9)表示对 I_s^B 迁移图像，I_r 原图像进行 FP (Face Parsing)人脸解析。

最后我们将人脸肤色损失函数加入到模型的总损失函数当中，如下公式(10)所示。

$$\begin{aligned} &\min_{E_1,E_2,G_1,G_2} \min_{D_1,D_2} L\left(E_1,E_2,G_1,G_2,D_1,D_2\right) \\ &= L_{GAN}^{x1} + L_{GAN}^{x2} + \lambda_x\left(L_{recon}^{x1} + L_{recon}^{x2}\right) + \lambda_c\left(L_{recon}^{c1} + L_{recon}^{c2}\right) \\ &\quad + \lambda_s\left(L_{recon}^{s1} + L_{recon}^{s2}\right) + \lambda_f \cdot L_{face} \end{aligned} \tag{10}$$

4. 实验与分析

4.1. 数据集

本文在综合考虑后，选用公开数据集 CelebFaces Attributes Dataset (CelebA)。CeleBA 数据集是一个大规模的人脸属性数据集，包括 10,177 个身份，202,599 张人脸图像，且每张照片都有特征标注信息，包含性别以及各种人脸特征等 40 多项信息。将 CeleBA 数据集的训练集输入模型进行训练，验证集和测试集输入模型进行测试。为减少无关背景因素对于图像生成结果的影响，我们对数据集的标注信息进行预处理，选取年轻人并对图像做合适的裁剪，将图片大小调整为 256*256。最后男性人脸训练集和测试集数量分别是 46,372 和 4564，女性人脸训练集和测试集的数量分别是 90,016 和 10,014。

4.2. 实验细节

本文实验的服务器配置如表 1 所示，模型训练的部分参数设置如表 2 所示。

Table 1. Experimental server configuration
表 1. 实验服务器配置

操作系统	ubuntu18.04	CPU	15 核 AMD EPYC 7543 32-Core Processor
内存	80G	GPU	RTX3090
显存	24G	Python	3.6.13
Pytorch	1.10.2	Cuda	11.3

Table 2. Some parameters of the experiment
表 2. 实验的部分参数设置

max_iter	1,000,000	batchsize	1
l_r	0.0001	λ_x	10
λ_c	1	λ_s	1
λ_f	1	gamma	1

将预先处理好的 CeleBA 数据集输入到模型进行训练。在实验中，模型训练次数 max_iter 为 1,000,000，batchsize 设为 1，初始学习率 l_r 设置为 0.0001，将式(4)中的图像重建损失权重 λ_x 设置为 10，图像风格及内容的重建损失权重 λ_c，λ_s 均设置为 1，人脸面部 mask 的直方图匹配损失权重 λ_f 设置为 1，每次学习率衰减的大小 gamma 设置为 0.5。在模型的训练过程中，使用 Adam [27]优化器对梯度下降进行优化。

4.3. 评价指标

图像风格迁移结果主观性非常大，因计算机很难对转移前后的图像风格变化给出定性的评价结果。因此，本文将结合主观视觉评价与客观指标评价对模型结果进行解析。主观视觉评价将本文模型生成结果与同等条件下其他模型生成结果随机采样，依靠不同用户的评价选出人脸性别转换效果最优的模型。客观评价指标结合内容准确率和结构相似度进行综合评判。

1) 内容准确率。内容准确率即模型生成的伪造数据通过判别器的概率，也就代表了模型生成结果的有效性。本文使用 InceptionV3 网络[28]作为分类模型。将分类模型在 CeleBA 数据集中进行预训练得到基准的内容准确率，然后将本文模型生成的伪造图像输入到预训练后的分类模型中，如果伪造的图像足够真实可以通过分类模型，将其计入正确样本，最后将正确样本与输入样本数相除即可得到最后的内容准确率，准确率越高代表模型生成效果越好。

2) 结构相似度。本文基于 FID (Fréchet Inception Distance)指标来计算男女面部特征之间的相似度。FID 代表了真实人脸图像与模型伪造的人脸图像的特征向量之间距离的一种度量。这种视觉特征是使用 Inception v3 图像分类模型提取特征并计算得到的。FID 在最佳情况下的得分为 0.0，表示两组图像相同。分数越低代表两组图像越相似，或者说二者的统计量越相似。FID 计算式如式(11)所示：

$$\mathrm{FID}=\left\|\mu_1-\mu_2\right\|_2^2+T_r\left(\sum_1+\sum_2-2\left(\sum_1\sum_2\right)^{\frac{1}{2}}\right) \tag{11}$$

其中，μ_1 和 $\sum_1$ 为输入的人脸数据集的均值和协方差矩阵，μ_2 和 $\sum_2$ 为模型生成数据集的均值和协方差矩阵，T_r 表示矩阵对角线上元素的总和。

4.4. 效果评估

4.4.1. 主观视觉评价

本文将预处理过的 CeleBA 数据集输入到改进的 MUNIT 模型，原始 MUNIT 模型以及 CycleGAN 模型中进行训练和测试，横向对比每种方法的生成结果。本文所做实验均采用经过 1,000,000 次迭代的生成模型，且同一种实验采用相同的测试数据，只保留生成方法和训练数据的不同。实验结果如图 6、图 7 所示。

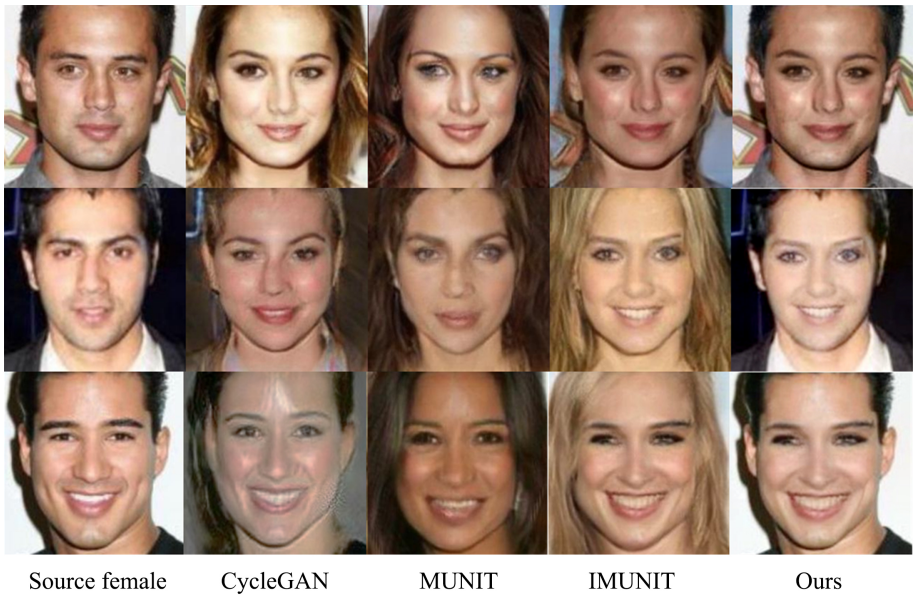


Figure 6. Male to female
图 6. 男性转为女性

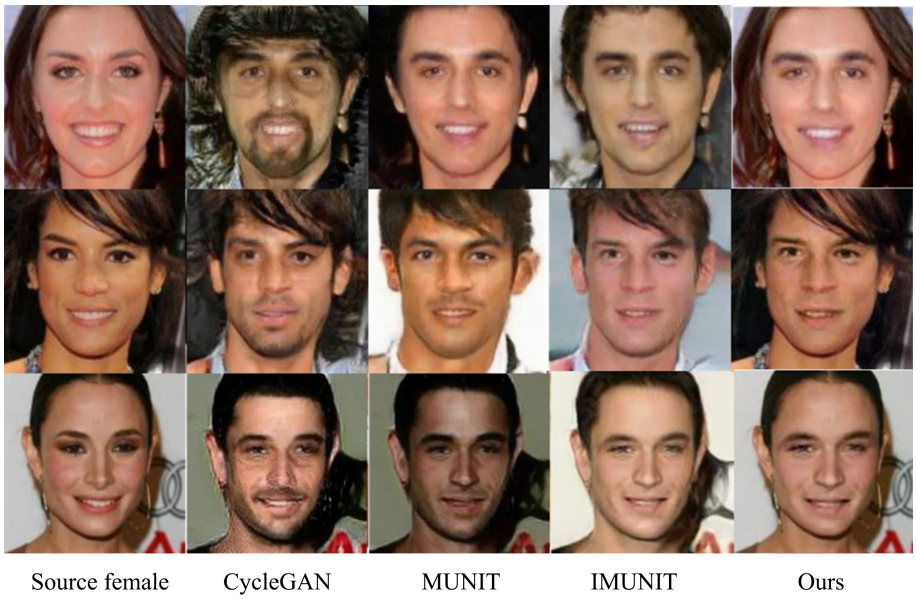


Figure 7. Female to male
图 7. 女性转为男性

图 6，图 7 从左到右每列分别为原图像，CycleGAN 生成的性转图像，原 MUNIT 生成的性转图像，加入 CBAM 和 DIN 的 IMUNIT 模型以及本文模型生成的性转图像。从图 6 男性到女性的性别转换结果可以看出，本文方法在能够完好的保留原图像的背景区域以及模型生成前后的人脸肤色，并且生成更加优秀的性转图像效果。显然，图 6，图 7 中不难看出本文方法优秀的背景域及人脸肤色的保留效果，相比于其他经典风格迁移方法，本文所提方法有显著优势；在人脸部分，本文结果与 IMUNIT 模型的结果基本保持一致，但从整体上看，本文方法的生成效果更为优秀。

我们随机选取 10 张人脸图像，男女各 5 张，输入到 CycleGAN，MUNIT，IMUNIT 和本文方法生成的结果组合成问卷，交由 259 名用户进行评选，选取性别转换后效果最好的图像(模型)。所得结果如图 8 所示，显然，本文所提方法在人脸图像性别转换上表现的最好。

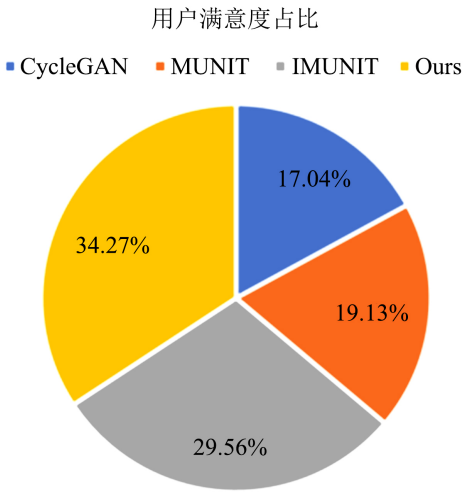


Figure 8. User satisfaction survey pie chart
图 8. 用户满意度调查饼状图

4.4.2. 客观指标评价

1) 消融实验

本文在 IMUNIT 的基础上逐步增加 Face Parsing 和人脸肤色损失函数 L-Face，下面将分别计算在不同改进策略下的内容准率和 FID 得分。

如表 3 所列，添加 Face Parsing 操作后生成模型对人脸部分进行准确迁移，在 CeleBA 数据集上，伪造女性和伪造男性的内容准确率分别提高了 0.011 和 0.026；继续添加损失函数 L-face，使模型保持人脸肤色一致，内容准确率再提高了 0.021 和 0.031。模型中添加 Face Parsing 后，在 CeleBA 数据集上，伪造女性和伪造男性的 FID 得分分别降低了 7.86 和 4.77；继续增加损失函数 L-Face 后，FID 再降低了 4.14 和 3.33。从表 3 和表 4 可以看出，本文在 IMUNIT 模型上进行的改进是行之有效的。

Table 3. Content accuracy under different conditions on the CeleBA dataset
表 3. CeleBA 数据集上不同条件下的内容准确率

	IMUNIT	IMUNIT + FP	IMUNIT + FP + L-Face
男转女	0.935	0.946	0.967
女转男	0.583	0.609	0.640

Table 4. FID scores under different conditions on the CeleBA dataset
表 4. CeleBA 数据集上不同条件下的 FID 得分

	IMUNIT	IMUNIT + FP	IMUNIT + FP + L-Face
男转女	63.59	55.73	51.59
女转男	37.45	32.68	29.35

2) 与其他方法对比

本文方法与其他方法的内容准确率和 FID 得分的对比结果如表 5、表 6 所列。本文方法在男女性别转换的实验中内容准确率相较于其他方法都更加优秀，说明基于本文方法生成的人脸图像更加真实。基于本文模型的男转女 FID 得分低于 IMUNIT 模型以及原始的 MUNIT 模型，高于 CycleGAN 模型，说明本文方法在身份保留方面还有进步的空间，需要继续改进；在男转女的 FID 的得分结果在几种方法中最低，说明本文方法具有更好的模型性能，使模型的人脸生成结果更真实，效果更好。

Table 5. Content accuracy of each model on CeleBA dataset
表 5. CeleBA 数据集上各模型的内容准确率

	CycleGAN	MUNIT	IMUNIT	Ours
男转女	0.876	0.886	0.935	0.958
女转男	0.379	0.573	0.583	0.690

Table 6. FID scores of each model on the CeleBA dataset
表 6. CeleBA 数据集上各模型的 FID 得分

	CycleGAN	MUNIT	IMUNIT	Ours
男转女	37.45	86.34	63.59	52.47
女转男	43.47	45.56	37.45	33.20

5. 结束语

本文在无监督样式迁移 MUNIT 的基础上完成人脸图像性别迁移任务，为解决模型性转结果图像背景区域扭曲模糊以及人脸肤色随机改变的缺点，本文提出具有鲁棒性质的人脸图像性别转移模型。首先对输入模型的人脸图像进行 Face Parsing，直接提取人脸部分进行训练学习，可以完好保留图像背景域，同时减少了无关区域对模型训练的影响，从而提高了图像生成质量；构建新的人脸肤色损失函数，将模型生成前后的人脸部分进行 Histogram Matching，以此可以保持性转前后的人脸肤色保持一致。通过最后的实验结果可得，本文所述方法可以更好地保留图像背景区域，解决了模型生成前后肤色不一致的问题，产生了更高质量的性别转换图像。同时，我们可以看到本文方法仅仅关注于人脸部分区域的性别转换，没有照顾到头发等其他关于性别的显著特征，未来可以考虑与人脸 3D 结合，将 3D 头发与 2D 人脸进行组合，以生成更加信服的人脸性别转换结果。

基金项目

北京市教育委员会科学研究计划项目资助(KM202110016001, KM202210016002)。北京建筑大学科学

研究基金(KYJJ2017017, PG2021094)；住房和城乡建设部科学技术计划北京建筑大学北京未来城市设计高精尖创新中心开放课题(NO. UDC2019033324, UDC201703332)；北京建筑大学课程建设重点培育项目“高等数学 ZDXX202008”。

参考文献

[1] Schroff, F., Kalenichenko, D. and Philbin, J. (2015) FaceNet: A Unified Embedding for Face Recognition and Clustering. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 815-823. https://doi.org/10.1109/CVPR.2015.7298682

[2] Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I. and Zafeiriou, S. (2022) ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 5962-5979. https://doi.org/10.1109/TPAMI.2021.3087709

[3] Meng, Q., Zhao, S., Huang, Z. and Zhou, F. (2021) MagFace: A Universal Representation for Face Recognition and Quality Assessment. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 14220-14229. https://doi.org/10.1109/CVPR46437.2021.01400

[4] Deng, H., Han, C., Cai, H., Han, G. and He, S. (2021) Spatially-Invariant Style-Codes Controlled Makeup Transfer. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 6545-6553. https://doi.org/10.1109/CVPR46437.2021.00648

[5] Sun, Z., Chen, Y. and Xiong, S. (2021) SSAT: A Symmetric Semantic-Aware Transformer Network for Makeup Transfer and Removal. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 2325-2334. https://doi.org/10.1609/aaai.v36i2.20131

[6] Nguyen, T., Tran, A. and Hoai, M. (2021) Lipstick Ain’t Enough: Beyond Color Matching for In-the-Wild Makeup Transfer. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), 13300-13309. https://doi.org/10.1109/CVPR46437.2021.01310

[7] Sun, J., Wang, X., Zhang, Y., Li, X., Zhang, Q., Liu, Y. and Wang, J. (2021) FENeRF: Face Editing in Neural Radiance Fields. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 7662-7672. https://doi.org/10.1109/CVPR52688.2022.00752

[8] Otberdout, N., Ferrari, C., Daoudi, M., Berretti, S. and Bimbo, A. (2021) Sparse to Dense Dynamic 3D Facial Expression Generation. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 20353-20362. https://doi.org/10.1109/CVPR52688.2022.01974

[9] Xu, Y., Yin, Y., Jiang, L., Wu, Q., Zheng, C., Loy, C.C., Dai, B. and Wu, W. (2022) TransEditor: Transformer-Based Dual-Space GAN for Highly Controllable Facial Editing. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 19-20 June 2022, 7673-7682. https://doi.org/10.1109/CVPR52688.2022.00753

[10] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C. and Bengio, Y. (2014) Generative Adversarial Nets. *Proceedings of the NIPS* 2014 *Workshop on High-Energy Physics and Machine Learning*, Montreal, 13 December 2014, 2672-2680.

[11] Rai, H. and Shukla, N. (2018) Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *IEEE International Conference on Computer Vision*, *ICCV* 2017, Venice, 22-29 October 2017, 2223-2232.

[12] Chen, X., Xu, C., Yang, X., *et al*. (2018) Gated-GAN: Adversarial Gated Networks for Multi-Collection Style Transfer. *IEEE Transactions on Image Processing*, **28**, 546-560. https://doi.org/10.1109/TIP.2018.2869695

[13] Sanakoyeu, A., Kotovenko, D., Lang, S., *et al*. (2018) A Style-Aware Content Loss for Real-Time HD Style Transfer. *European Conference on Computer Vision*, Munich, 8-14 September 2018, 698-714. https://doi.org/10.1007/978-3-030-01237-3_43

[14] Chen, D., Yuan, L., Liao, J., *et al*. (2017) Stylebank: An Explicit Representation for Neural Image Style Transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 1897-1906. https://doi.org/10.1109/CVPR.2017.296

[15] Liu, H., Michelini, P.N. and Dan, Z. (2018) Artsy-GAN: A Style Transfer System with Improved Quality, Diversity and Performance. 2018 *24th International Conference on Pattern Recognition* (*ICPR*), Beijing, 20-24 August 2018, 79-84. https://doi.org/10.1109/ICPR.2018.8546172

[16] Ma, Z., Li, J., Wang, N., *et al*. (2020) Semantic-Related Image Style Transfer with Dual-Consistency Loss. *Neurocomputing*, **406**, 135-149. https://doi.org/10.1016/j.neucom.2020.04.027

[17] Huang, X., Liu, M., Belongie, S.J. and Kautz, J. (2018) Multimodal Unsupervised Image-to-Image Translation. 15*th European Conference on Computer Vision*, Munich, 8-14 September 2018, 179-196.

https://doi.org/10.1007/978-3-030-01219-9_11

[18] Fan, M., Lai, S., Huang, J., Wei, X., Chai, Z., Luo, J. and Wei, X. (2021) Rethinking BiSeNet for Real-Time Semantic Segmentation. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 19-25 June 2021, 9711-9720. https://doi.org/10.1109/CVPR46437.2021.00959

[19] Wilmot, P., Risser, E. and Barnes, C. (2017) Stable and Controllable Neural Texture Synthesis and Style Transfer Using Histogram Losses.

[20] Liu, M., Breuel, T.M. and Kautz, J. (2017) Unsupervised Image-to-Image Translation Networks.

[21] Kim, J., Kim, M., Kang, H. and Lee, K. (2020) U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation.

[22] 石达, 芦天亮, 杜彦辉, 等. 基于改进 CycleGAN 的人脸性别伪造图像生成模型[J]. 计算机科学, 2022, 49(2): 31-39.

[23] Liu, X., Wang, R., Peng, H., *et al*. (2021) Sparse Feature Representation Learning for Deep Face Gender Transfer. *IEEE International Conference on Computer Vision*, Montreal, 11-17 October 2021, 4070-4080. https://doi.org/10.1109/ICCVW54120.2021.00454

[24] Woo, S., Park, J., Lee, J. and Kweon, I. (2018) CBAM: Convolutional Block Attention Module. 15*th European Conference on Computer Vision*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[25] Jing, Y., Liu, X., Ding, Y., Wang, X., Ding, E., Song, M. and Wen, S. (2020) Dynamic Instance Normalization for Arbitrary Style Transfer. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 4369-4376. https://doi.org/10.1609/aaai.v34i04.5862

[26] Yu, C.Q., Wang, J.B., Peng, C., Gao, C.X., Yu, G. and Sang, N. (2018) BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. *European Conference on Computer Vision*, Munich, 8-14 September 2018, 334-349. https://doi.org/10.1007/978-3-030-01261-8_20

[27] Kingma, D. and Ba, J. (2014) Adam: A Method for Stochastic Optimization. Computer Science.

[28] Szegedy, C., Vanhoucke, V., Ioffe, S., *et al*. (2016) Rethinking the Inception Architecture for Computer Vision. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 2818-2826. https://doi.org/10.1109/CVPR.2016.308