

FaceInpainter:对异质领域的高保真人脸适应性

李佳, 李朝阳, 曹杰, 宋星光, 何然

中国科学院模式识别国家实验室

中科院脑科学与智能技术卓越中心

中国科学院大学人工智能学院, 北京, 中国

{jia.li, jie.cao}@cripac.ia.ac.cn, rhe@nlpr.ia.ac.cn,

zhaoyang0427@gmail.com, Wilson.song@hotmail.com



图1：通过FaceInpainter对野生人脸图像进行换脸的结果。我们的身份引导的脸部绘画框架可以在不同的领域产生高保真的结果，即照片或非写实风格，如油画、三维卡通、铅笔画和夸张画。

摘要

†通讯作者

在这项工作中，我们提出了一个名为FaceInpainter的新颖的两阶段框架，以实现在异质环境下的可控身份引导的脸部绘画（IGFI）。具体来说，通过明确地分解目标脸的前景和背景，第一阶段的重点是通过风格化的脸部绘画网络（SFI-Net）对固定的背景进行自适应的脸部拟合，其中包括目标的三维先验和纹理代码以及源脸的身份因子。处理源的新身份和目标的原始背景之间的不一致是很有挑战性的，涉及到融合边界上的脸部形状和外观。第二阶段由联合完善网络（JR-Net）组成，以完善交换后的面部。它利用AdaIN考虑身份和多尺度纹理代码，对来自SFI-Net的有面部遮挡的解码脸部进行特征转换。我们采用上下文损失来隐含地保留属性，使脸部变形和减少纹理失真。实验结果表明，我们的方法能够提供高质量的身份适应性，以适应异质性的工作，表现出具有竞争力的性能。

1. 简介

身份互换是一种技术，它能够根据源脸操纵脸部外观，同时保留目标脸的属性。DeepFakes[1]、FaceSwap[2]、FaceShifter[18]和SimSwap[5]是比较突出的方法。这些方法能够毫不费力地创建假脸，而可控性能和视觉真实性仍需改进。此外，人脸操作提供了更具挑战性和多样性的样本，以促进人脸取证[31]。

脸部互换任务主要从两个方面进行评估，即源脸和目标脸的身份和属性保真度。FaceShifter[18]采用了SPADE[28]和AdaIN[14]机制在解码器中整合了空间属性和身份风格，而源的一些重新冗余的身份特征，例如头发，使得在一些具有挑战性的情况下难以保留属性，如图2所示。SimSwap[5]使用基于鉴别器最后几层的弱特征匹配损失，以避免身份修改过程中的属性失真。

tion。尽管这种隐性约束效果很好，但它类似于[38, 49, 24]中常用的纹理匹配的感知损失，更适合于成对的数据学习。因此，在一些具有挑战性的情况下，SimSwap生成的人脸预先服务于大量的属性，使其看起来更像目标脸。在非逼真的

在这个领域，我们以FaceSwap[2]为例，这种基于3D的方法的主要问题包括低分辨率和面部周围明显的修改伪影。

在这项工作中，我们首先研究了异质性的身份互换问题。在一些工作中已经研究了异质脸的问题[42, 10, 11, 41]。在异质领域中，特别是在极端的姿势、表情和光照条件下，用一个新的身份完美地适应目标脸是一个挑战。我们提出了一个有效的身份引导的脸部绘画（IGFI）框架，以提高换脸模型的可控性和通用性。在IGFI过程中，目标脸的面部属性应该被保留下来，包括头部姿势、表情、照明、面部轮廓、发型和其他背景内容。此外，我们试图自适应地提高低分辨率场景的视觉质量。

在我们的工作中，身份特征是通过预先训练的最先进的人脸识别模型[6]从源脸部提取的。一般来说，身份特征包含源的高级语义表示，用于控制身份修改。至于目标的属性保存，3D拟合模型[12]被用来提取目标的3DMM参数，如表情和姿势，这对基于固定背景进行IGFI至关重要。为了保留纹理，一个预训练的脸部解析模型[19]被用来提取前景内容（脸部和颈部），并将其输入预训练的VGG网络[33]以获得纹理样式。之后，来自源脸和目标脸的形状和纹理方面的多个因素被重新组合成一个风格代码，以实现特定场景下高效和可控的IGFI，基于风格化的脸部绘画网络（SFI-Net）。特别是，为了保留高保真的非脸部区域（如背景、头发、衣服等），我们将它们直接添加到SFI-Net的最后一层。

为了削弱目标脸型和新身份的脸部绘画结果的不一致性，我们在第一阶段掩盖了对脸型表示至关重要的脸部和脖子区域。然而，对于一些复杂的场景，分割后的背景和生成的前景不能很好地结合。因此，在第二阶段，我们提出了一个联合细化网（JR-Net），通过采用自适应实例归一化（AdaIN）[14]对身份和多尺度纹理代码进行联合细化，以实现身份、属性和边界融合。通过超分辨率和闭塞感知的



图2：仍有一些具有挑战性的情况，例如，保留过多的源身份（FaceShifter[18]），预先提供更多的属性（SimSwap[5]），以及受到低分辨率和边界伪影的影响（FaceSwap[2]）。我们的模型在照相学（第1行）和非照相学（第2行）领域实现了更可控的人脸互换。

模块，JR-Net进一步优化了SFI-

Net的结果的视觉和闭塞感知。对于目标脸和变形脸之间的非对齐特征匹配，我们采用了上下文约束[23]，可以隐含地保持属性的一致性。如图1所示，我们的方法可以生成高质量的身份引导的交换脸，适应于各种异质领域，甚至是卡通和夸张的风格。

我们的贡献有三个方面。

- 我们首先研究了异质身份互换任务，并提出了一种新的解决方案来处理身份引导下的人脸涂抹（IGFI）问题。通过考虑高保真的身份和属性，我们的方法FaceInpainter比以前的脸部绘画方法取得了更可控和更高质量的结果。
- 我们提出了一个有效的IGFI框架。在第一阶段，我们引入了一个造型脸部绘画网络（SFI-Net），将身份和属性代码映射到被交换的脸部。第二阶段包含一个联合再细化网络（JR-Net），细化属性和身份细节，生成具有视觉自然融合边界的闭塞感知和高分辨率的交换脸。
- 由于实现了上下文属性和身份的高保真，我们在视觉上和数量上都实现了异质领域的良好概括。

2. 相关工作

2.1. 脸部彩绘

脸部彩绘主要应用于脸部完善[22, 45, 43, 34, 44]或组件编辑[7]。通过纳入VAE管道，Zheng等人[46]利用了隐藏部分图像的潜伏空间的平滑先验，从

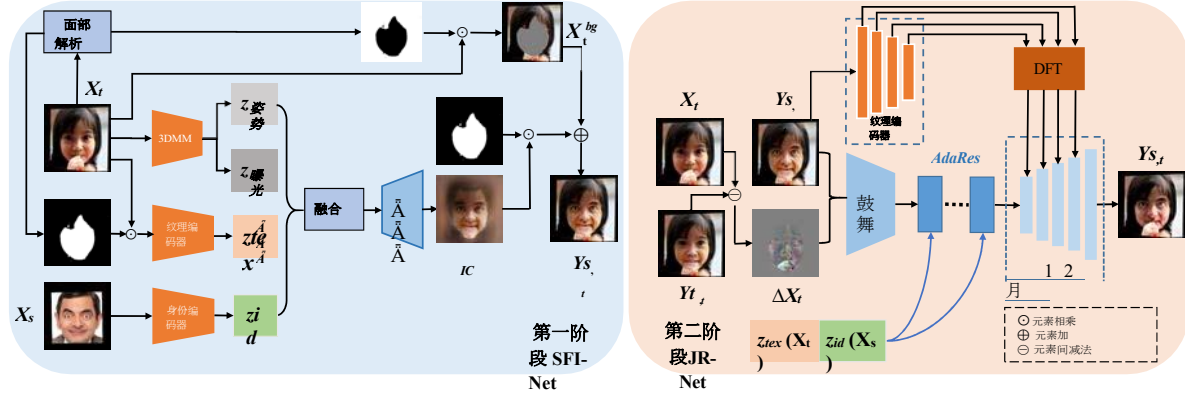


图3：FaceInpainter的框架。第一阶段（SFI-Net）从目标的属性代码中学习映射。

X_t 和源 X_s 的身份代码到基于 StyleGAN G_s 的绘画结果 I_c ，然后与固定的背景 X^{bg} ，得到交换后的脸部 $\hat{Y}_{t,s}$ 。第二阶段（JR-Net）是为细化SFI-Net的粗略结果而设计的。 z_{id} (X_s) 和 z_{tex} (X_t) 被融合，以控制AdaIN残余块中的非正常化参数，这有利于身份和属性的细化。 ΔX_t 和 $\hat{Y}_{t,t}$ 之间的残差图，作为闭塞表示。此外，字典特征转移（DFT）模块被用来提高面部分辨率。

其中生成管道推断出不同的绘画结果。为了产生一个可控的完整的脸部，Song等人。

[34]提出了一种几何感知的方法，在面部地标热图和解析图的约束下对面部进行涂抹和编辑。Deng等人[7]提出了r-FACE模型，用于控制基于参考图像的人脸组件的结构。

2.2. 身份互换

身份互换的目的是交换人脸特征而不对人脸的其他属性造成干扰[27, 18, 5]。FSGAN[27]结合了关于递归重演、脸部解析、内画和混合的级联模块，以实现高质量的交换再结果。FaceShifter[18]采用了一个两阶段的框架，将身份和属性嵌入与注意力图谱相适应，并以后发式错误确认的方式解决遮挡问题。SimSwap[5]利用ID注入模块，自适应地将身份嵌入到目标特征中，并解决弱特征匹配损失。

2.3. 基于计算机图形的方法

来自三维可变形人脸模型（3DMM）[48, 9, 12]或非线性3DMM[36, 35]的推理已被纳入各种人脸合成任务[47, 8]。FaceSwap[2]应用三维拟合和渲染，以获得对齐的面，其中

然后与目标进行混合。Hang等人[47]提出了一个旋转和渲染框架，利用三维人脸模型和高分辨率的GAN，分别用于创建配对数据和人脸内画。DiscoFaceGAN[8]解除了身份、表情、姿势的潜在表示。

和光照来精确控制具有不同属性的人脸的生成。

3. 办法

IGFI的任务应该是生成具有身份修改和属性保留的前景内容，以及与背景的自然融合。如图3所示，给定一个目标脸 X_t

，我们主要考虑三个重要因素：表情、姿势和纹理。我们利用先进的人脸识别网络[6]、3DMM[12]和纹理编码器[33]来获得解构的风格代码，这些代码被融合为一个变量来控制风格化人脸合成网络。作为一个内画任务，前景的遮挡已经被重新移动，新的人脸之间存在不一致的情况和原始目标背景。因此，SFI-Net生成的人脸 $\hat{Y}_{t,t}$ 中的人脸形状和属性的细节将由JR-Net来完善，描述如下。

3.1. 风格化的脸部彩绘网络

我们结合3DMM模型[3]来学习从源脸和目标脸的多个代码到IGFI结果的映射。作为一个参数化的模型，3DMM被用来拟合单一图像的形状 S 和纹理 T 。具体来说， S 被表示为。

$$S = S + B_{id} \alpha_{id} + \text{曝光 } \alpha_{ex}, \quad (1)$$

B 光 p

其中 S 是平均三维形状， B_{id} 是三维形状基数， α_{id} 是相应的系数向量。表达方式也是如此。单一图像的三维拟合被定义为

$$F(S) = f * P_r * R(p, y, r) * S + t_{2d}. \quad (2)$$

其中 f 是比例系数， P_r 是正字法投影矩阵， $R(p, y, r)$ 是由俯仰、偏航和滚动表示构建的旋转矩阵， t_{2d} 是平移矢量。面部姿态定义为 $z_{pos} = [f, R(p, y, r), t_{2d}]$ 。我们把 $z_{exp} = \alpha_{exp}$ 表示为表达式。

从目标脸部识别代码。一般来说，身份和前

对于非线性3DMM[36,

35], 用一个形状向量代表身份和表情因素。代替 α_{id} ，我们使用预先训练好的最先进的人脸识别模型[6]来获得身份潜伏代码 z_{id} 。

给定一个目标图像 X_t ，用现成的人脸分析器对 K 通道热图 M 进行响应，我们将其分开。

成背景 X_t^{bg} （如头发、衣服和其他非人区）和前景 X_t^{fg} （如脸部、颈部）。在

在我们的SFI-Net训练中， X_t 被送入一个纹理编码器。 E_{vgg} ，即预训练的VGG19网络[33]，以获得

目标脸部的纹理代码 z_{tex} ，其中包含皮肤颜色和光照信息。通过 $C = [z_{id}(X_s), z_{exp}(X_t), z_{pos}(X_t), z_{tex}(X_t^{fg})]$ ，我们通过由一个全连接层组成的融合模块得到IGF I的风格代码 z_c ，然后将 z_c

注入StyleGAN[17]，方法是

实施多尺度的AdaIN。

身份识别网络的编码特征

[6]一般包含发型，它可能对目标背景的保存有影响，同时修改由 $z_{id}(X_s)$ 引导的目标脸。为了保持高精细度的场景， X_t 中的背景区域直接与StyleGAN I_c 的输出融合，通过

$$Y^{\wedge}s_t = X_t \odot M_{bg} + I_c \odot (1 - M_{bg}), \quad (3)$$

其中， \odot 指的是元素相乘，而 M_{bg} 是指表示目标背景的掩码。

我们在 $Y^{\wedge}s_t$ 上使用多个代码的一致性损失，用于风格的人脸合成，具体如下。首先，我们通过以下方法计算 $Y^{\wedge}s_t$ 和 X_s 的身份特征之间的差距。

$$L_{id} = 1 - (z_{id}(Y^{\wedge}s_t), z_{id}(X_s)). \quad (4)$$

其中 $(-, -)$ 表示余弦相似性。

表情和姿势代码应该保持 $Y^{\wedge}s_t$ 和 X_t 的一致性，用以下方式表示

$$L_{exp} = |z_{exp}(X_t) - z_{exp}(X_t)|_1, \quad (5)$$

$$L_{pos} = |z_{pos}(X_t) - z_{pos}(X_t)|_1. \quad (6)$$

让 L_{GAN} 是辨别生成的人脸 $Y^{\wedge}s_t$ 和真实人脸 X_s 的对抗性损失，通过

我们利用一种半监督学习策略，即

重建损失用于惩罚在相同身份情况下 $Y^{\wedge}s_t$ 和 X_t 之间的像素级差异。它被表述为

$$L_{recon} = \frac{1}{2} \|Y^{\wedge}s_t - X_t\|_2^2 \quad \text{如果 } X_s = X_t \quad (8)$$

否则 0

FaceShifter[18]设计了自适应注意力规范化（AAD）层来整合属性和身份嵌入，其中注意力掩码的学习不受属性和身份保留损失的限制。请注意，由人脸识别模型[6]编码的 $z_{id}(X_s)$ 通常包含属于 X_t^{bg} 的额外样式，例如发型，这可能导致不可控制的应用。

图2行1)。必须采取强有力的纹理保存功能，以减轻多余的身份信息对交换后的脸部的影响。对此

最后，我们利用上下文损失[23]来衡量 $Y^{\wedge}s_t$ 和 X_t 之间的特征相似性。这种损失允许脸部形状变形并减少脸部变形后的纹理失真。

调换。它被表述为

$$L_{cx} = -\log(CX(F)_{vgg}^l(Y^{\wedge}s_t),_{vgg}(X_t)). \quad (9)$$

其中 l 表示预训练的VGG19网络的 $relu\{3, 2, 4\}$ 层。

此外，我们结合路径长度正则化[17]来优化SFI-Net的训练。它被表述为

$$L_{ppl} = E_{w, y \sim N(0, I)} (\frac{1}{J_w} \frac{\partial Y^{\wedge}s_t}{\partial w} - a)^2. \quad (10)$$

其中 $J_w = \frac{\partial Y^{\wedge}s_t}{\partial w}$ ， $w \sim f(C)$ 表示中间的融合风格代码 z 的潜伏代码通过映射网络 c ， y 是具有正态分布的随机图像， a 被设定为感知路径长度的指数移动平均值。SFI-Net的总损失如下

$$L_{SFI-Net} = L_{GAN} + \lambda_{id} L_{id} + \lambda_{exp} L_{exp} + \lambda_{pos} L_{pos} + \lambda_{rec} L_{rec} + \lambda_{CX} L_{CX} + \lambda_{ppl} L_{ppl}. \quad (11)$$

3.2. 联合完善网络

SFI-Net可以生成具有 X_s 的身份风格和 X_t 的其他身份无关的风格（如表情、姿势和纹理）的人脸。此外，SFI-Net的结果预为大部分的背景内容提供服务，除了脸部表情。由于解析模型的错误，出现了闭塞现象。其他问题主要是一些低分辨率的情况，以及混合 X_t^{bg} 和 I_t^{fg} 时的边界伪影。

$$\mathbf{L}_{GAN}(G,D)=\mathbf{E}_X[\log D(X_s)]+\mathbf{E}_Y[\log(1-D(Y^s_{s,t}))]. \quad (7)$$

我们提出了联合细化网络（JR-Net）来细化SFI-Net的面部绘画结果，如图所示

图3的第二阶段。为了使生成的图像具有闭塞感知能力，我们将 X 的残差图 t ，并 $Y^{\wedge}t_t$ 作为闭塞表示，并将其送入JR-Net。

$$\Delta X_t = X_t - Y^{\wedge}t_t. \quad (12)$$

JR-Net的属性嵌入来自于 X_t 的高水平纹理特征。具体来说，我们使用预先训练的VGG提取conv4 1层的纹理特征，并对其进行分析。使用自适应平均法，保持512个通道的风格代码。龄汇集。它与 X_s 的身份嵌入相融合，然后注入AdaIN [14]的残余块AdaRes中。

适应性转移 $\{Y^{\wedge}s_t, \Delta X_t\}$ 的特征样式。它是为---拟定为：

$$AdaIN(h^i, \gamma^i, \beta^i) = \{id, at\} \odot \frac{h^i - \mu^i}{\sigma^i} + \beta^i_{\{id, at\}}, \quad (13)$$

其中， $h^i \in \mathbb{R}^{C^i \times H_i \times W_i}$ 是 $\{Y^{\wedge}, \Delta X\}$ 的嵌入。 s, t

μ^i 和 σ^i 是 h^i 的均值和标准差，它们被用来执行实例诺言。恶化。设 $z^i_{tex} \in \mathbb{R}^{C_{tex} \times 1}$ 和 $z^i_{id} \in \mathbb{R}^{C_{id} \times 1}$ 是属性和身份嵌入在AdaIN的形式中 $\gamma^i_{\{id, at\}}$ 和 $\beta^i_{\{id, at\}} \in \mathbb{R}^{C^i \times H_i \times W_i}$ ，得到的是 $\{z^i_{tex}(X_t), z^i_{id}(X_s)\}$ 使用几个全连接层。此外，我们采用了超分辨率的细化方法基于高质量参考图像的多尺度成分字典的方案[20]。具体来说，在二维特征转移（DFT）模块中，通过AdaIN对离线生成的字典进行调制，基于纹理特征从 $\{Y^{\wedge}s_t, \Delta X_t\}$ 中提取相应成分（左眼、右眼、鼻子、嘴巴）的图谱。然后，匹配的复原特征被用于在去中心化过程中进行特征调制。通过相应的空间特征变换（SFT）层[40]，使编码器Dec。

在第二阶段，我们通过以下方式获得精炼的面孔

$$Y_{s,t} = JR-Net(\Delta X_t, Y^{\wedge}s_t, z_{tex}(X_t), z_{id}(X_s)). \quad (14)$$

为了更好地保留属性，我们利用上下文损失来衡量 $Y_{s,t}$ 和 X 的特征相似性。 t

$$LCX = -\log(CX(F^l_{vgg}(Y_{s,t}), F^l_{vgg}(X_t))). \quad (15)$$

为了提高身份精炼的性能，我们通过以下方式施加身份一致性损失

$$\hat{\gamma} = 1 - (z_{id}(Y_{s,t}), z_{id}(X_s)). \quad (16)$$

以类似的方式，我们将 $Y_{s,t}$ 和 X_t

我们进一步增加了感知损失，以改善高分辨率的交换脸的修复。

$$L_{vgg} = \frac{1}{N} \sum_{i=1}^N \|F^{i,vgg}_{vgg}(Y^{\wedge}_{s,t}) - F^{i,vgg}_{vgg}(X_t)\|_2 \quad \text{如果 } X=X_t \text{ 否则 } 0. \quad (18)$$

其中 $F^{i,vgg}_{vgg}$ 表示VGG19的第 i 个卷积层模型。我们设定 N 等于4。

JR-Net中的对抗性损失通过以下方式确保图像质量

$$L_{GAN}(G, D) = E_X [\log D(X_s)] + E_Y [\log(1-D(Y_{s,t}))]. \quad (19)$$

最后，JR-Net的总损失被表述为：

$$L_{JR-Net} = L_{GAN} + \lambda_{id} \hat{\gamma} + \lambda_{rec} + \lambda_{v} L_{vgg} + \lambda_{LCX} LCX. \quad (20)$$

4. 实验

实施细节。SFI-Net和JR-Net都是用CelebA-HQ[15]、FFHQ[16]和VG-HQ训练。GFace[29]数据集。我们对准并裁剪了源码和焦油。使用由[4]提取的五点地标来获得人脸。我们训练具有256×256分辨率的SFI-Net，最终的结果是：1.

JR-Net的稀释率为512×512。更多的结果可以在补充材料中找到。我们使用3DDFA2[12]获取3DMM参数。 $z_{exp}(X_t)$ 和 $z_{pos}(X_t)$ 的尺寸分别为10和12。 $z_{id}(X_s)$ 、 $z_{tex}(X_t)$ 和 $z_{tex}(X_t)$ 的尺寸为512。SFI-Net中融合模块的输出二值为1024。在公式中11和20， $\lambda_{id} = \lambda_{id} = 20$ ，其他权重设定为到10。

在JR-Net中，我们根据 X_t 的面部地标对 $Y^{\wedge}t_t$ 进行翘曲，使 ΔX_t 更侧重于闭塞。我们总共使用了8个AdaIN残差块。块的数量

在JR-Net中，下采样和上采样块分别被设置为3和4。至于数据增强，我们采用了多种随机操作，例如裁剪、翻转、旋转、模糊、亮度、饱和度和对比度的变化以及颜色抖动。

4.1. 与其他方法的比较

与以前的内画任务比较。一般的内画任务侧重于完成隐藏区域。对于脸部绘画任务，以前的方法[43,44,46]不能胜任产生身份引导的完成的任务。

之间的重建损失结合起来，当 X_s 和 X_t 是相同的身份时。

$$L_{\text{召回}} = \frac{1}{2} \|Y - X\|_F^2 \quad \text{如果 } X = X_s + X_t, \quad (17)$$

如图4所示，这些图像在面部区域造成了一些假象。我们的画像结果在高保真属性和身份因素方面有更好的表现。Forensics++[3]是一个大规模的人脸取证数据集，该数据集。对FaceForensics++的定性比较。脸部-由5,000个视频片段组成。我们比较了FaceInpainter

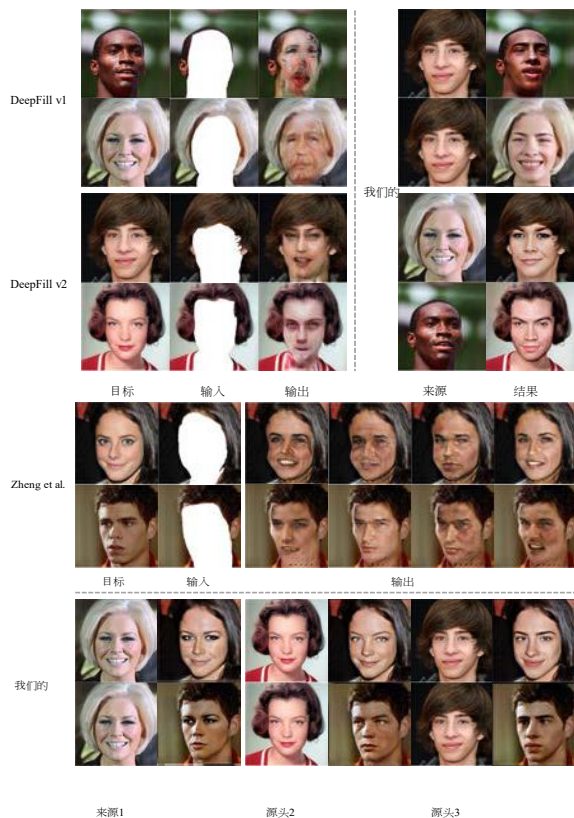


图4：与DeepFill v1[43]、DeepFill v2[44]和Zheng等人[46]的比较结果。我们的方法可以产生高质量的身份指导下的脸部绘画结果。



图5：与FaceSwap[2]、DeepFakes[1]和FaceShifter[18]在FaceForensics++[31]上的比较。我们的结果更好地实现了可控的属性保存和有效的身份修改。

与DeepFakes[1]、FaceSwap[2]和FaceShifter[18]在该基准上的表现相同。如图5所示，FaceShifter集中了



图6：与FSGAN[27]和FaceShifter[18]的更多比较结果。我们的结果在身份和属性方面都具有竞争性的视觉效果。



图7：与SimSwap[5]的更多比较结果。我们的方法可以生成交换后的人脸，脸颊和额头区域的伪影较少，同时正确跟踪目标的注视方向和表情。

在更多的身份修改上，源风格在交换后的脸上是不可控制的（例如第1行的眉毛，第2行的眼睛，第3行的眼镜，以及第4和第5行的头发）。在DeepFakes和FaceSwap的结果中也显示了类似的假象。我们的结果在目标脸的新身份和原始属性风格之间取得了更好的平衡，并且拥有高质量的面部区域的细节。

与FaceShifter的比较。如图6所示，FaceShifter能够生成高度可识别的换脸结果，但在保持属性细节方面不够好，例如第3行的凝视方向和皮肤颜色。

在第1行和第2行中，[18]提出的对齐属性一致性损失考虑了低层次的特征，并在一定程度上将目标的脸型风格带到生成的脸部（第4列）。我们的互换方法是进行高层次的

采用 L_{cx} 的特征匹配，结果保留了更高保真度的目标属性以及源脸的形状。

具有高分辨率的外观。与SimSwap的比较：SimSwap利用弱的feature匹配损失[5]来有效地提高保留面部属性的能力，也就是说，属性特征extractor放弃了鉴别器的前几层。SimSwap有能力修改身份并保留属性，但在脸部边界周围有明显的伪影（例如在额头和脸颊）。我们的结果是正确的

方法	身份证检索 ↑	姿势↓	表达式↓	SMD2↑	总的来说 ↑
DeepFakes [1]	81.96	4.14	0.187	8.835	86.47
换脸 [2]	54.19	2.51	0.148	9.573	61.11
FaceShifter [18]	97.38	2.96	0.136	8.483	102.77
SimSwap [5]	92.83	1.53	-	-	-
含有 $z_{exp}(X_t)$	97.23	2.10	0.121	8.499	103.51
w/o $z_{exp}(X_t)$	98.43	1.93	0.161	8.660	105.00
w/o $z_{pos}(X_t)$	97.93	4.97	0.140	8.428	101.25
w/o LCX	98.13	2.31	0.135	8.602	104.29
搭配 C	0.10	29.57	0.273	9.195	-20.55
自力更生协会-网络	97.93	1.68	0.117	8.407	104.54
不含DFT	98.43	2.18	0.139	8.773	104.88
w/o ΔX_t	98.30	2.13	0.134	8.654	104.69
不包括AdaRes	97.68	1.85	0.123	8.791	104.50
脸部画师	97.63	2.21	0.141	11.708	106.99

表1: 与DeepFakes的定量比较 [1],

FaceSwap[2]、FaceShifter[18]和SimSwap[5]在FaceForensics++[31]上。作为综合比较,我们通过减去其他人的ID检索和灰度差异模数之和(SMD2)[21]得分来获得总分。

正式代码	应该需要图像序列来进行训练	公开交换的结果		
		FaceForensics++	IIIT-CFW	中大素描
深度造假			×	×
换脸	×	×	×	×
脸部移位	×	×	×	×
模拟交换	×	×	×	×

表2: 我们在表3中只与FaceSwap比较的原因。

方法	身份证检索↑	姿势↓	Exp × 10 ↓	FID × 0.1 ↓	SMD2 × 0.1 ↑	总的来说↑
换脸IIIT-CFW	59.00	3.66	2.93	7.236	1.705	46.88
FaceInpainter IIIT-CFW	89.13	1.90	3.01	3.646	2.802	83.38
换脸中大	86.10	1.15	2.13	2.407	1.581	81.99
中大的FaceInpainter	88.36	0.99	2.23	2.148	2.843	85.84

表3:对IIIT-CFW[25]和CUHK[39]数据集的定量评估。



图8: 在CelebA-HQ数据集上的身份正面化和轮廓的更多定性结果[15]。我们的方法可以在给定目标脸的情况下,为不同身份的人生成正面和极端姿势的互换脸。

追踪凝视方向和表情,同时在没有视觉伪影的情况下传递新的身份,如图7所示。

额外的定性结果。如图8所示,IGFI的所有结果都是通过训练有素的FaceInpainter产生的,从各种身份(第1行)到相关的身份(第2行)。

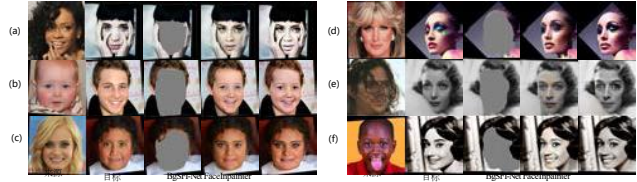


图9: JR-

Net处理了一些换脸合成的问题,例如融合边界、遮挡、脸部形状、皮肤颜色和低分辨率。

响应的正面视图(第2行),以及具有极端姿势的侧面视图(第3行)。根据目标属性,FaceInpainter可以生成具有相应身份的照片般真实的面孔。

定量比较。我们遵循[18, 5]中采用的实验协议,对身份转移和姿势保存进行定量评估。具体来说,我们采用CosFace[37, 26]来计算身份检索得分。与其他算法相比,我们的方法取得了更好的分数,如表1所示。我们使用[32]提取姿态特征。SimSwap的姿态得分较低,基于被调换的脸和目标脸之间的L-

2距离,但在ID检索上的表现相对较差,这说明了这表明它保留了过度的目标属性。至于表情估计,我们使用其他3DMM模型[9]来提取系数。我们的性能与FaceShifter相当,尽管只使用表情代码。

IGFI框架。在表1中, w/o DFT意味着w/ ΔX_t 和w/AdaRes, w/o ΔX_t 意味着w/AdaRes和w/o DFT, w/o AdaRes意味着w/ ΔX_t 和w/o DFT, FaceInpainter意味着w/ ΔX_t 、w/AdaRes、w/DFT。

考虑到交换后的人脸没有配对的高分辨率人脸作为参考图像来操作客观指标, Δ /PSNR和SSIM, 我们实现了基于灰度差异的评估[21]。通过DFT模块,我们的结果在图像清晰度方面取得了更好的成绩。

IIIT-CFW[25]被用于卡通脸部分类和Photo2Cartoon任务。CUHK[39]用于素描合成和面部素描识别。这两个数据集没有图像序列或视频。与上述任务不同的是,我们首先研究异质身份互换问题。对于IIIT-CFW和CUHK人脸素描,我们分别收集了N=1000和606张图像进行定量实验(表3)。第一张目标脸与第二张目标脸互换。

第N-i张脸的身份。Frechet Distance (FID) [13]被用来衡量质量和多样性。的调换的面孔。

人类评价。在对IGFI任务进行简单介绍后,10名用户被指示观察源、目标和交换的脸。参与者需要根据自己的经验来打分。分数被记录为



图10：基于FaceInpainter的各种野生脸部图像的IGFI结果。脸型是目标脸的一个重要属性，特别是在涉及到夸张的绘画和三维卡通时（第4和第6栏）。虽然我们实现了对异质领域的脸部适应，但独特的风格仍被保持。

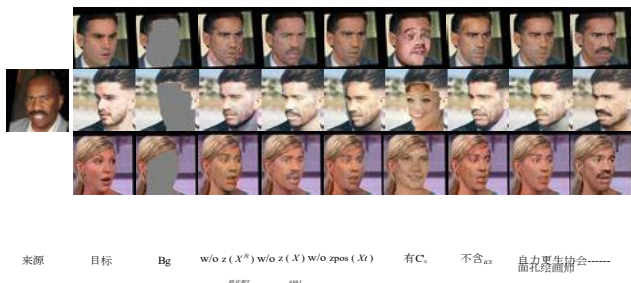


图11：SFI-Net的消融研究。

方法	身份	属性	边界	堵塞	现实主义
DeepFakes [1]	43.96	34.14	22.57	44.11	20.56
换脸 [2]	31.64	41.22	50.55	42.25	47.89
FaceShifter [18]	93.64	81.22	84.34	94.11	85.66
自力更生协会-网络	90.83	71.53	55.78	40.22	50.58
脸部画师	95.83	89.53	88.34	93.67	92.50

表4：用户研究结果。

1-100分为五个方面：(a)身份感知，(b)属性保存，(c)边界自然度，(d)闭塞感知，以及(e)整体真实性，分数越高意味着质量越高。有1000张来自FaceForensics++[31]的观察图像。我们收集了10000个人类的决定，平均分数显示在表4中。

4.2. 对FaceInpainter的分析

风格代码的必要性。 如图11所示，我们展示了不同风格代码设置的相应结果。在没有前景纹理代码的IGFI情况下，完成的面部区域存在纹理失真（第4栏）。如果没有表情代码，凝视方向或嘴部运动就不能很好地拟合（第5栏）。至于姿势，在脸部轮廓涂抹的情况下，当放弃姿势代码时，在合成的脸部中可以发现内部脸部和背景之间的明显差距（第6栏）。此外，受[30]的启发，我们使用一个可学习的编码器来利用多级属性代码 C_{att} ，而不是在SFI-Net中使用三维先验。 C_{att} 被认为包含了所有在

在SFI-Net的C语言

中，贡品风格代码，甚至更多的细节。

X_t 。那么融合后的风格代码 $C = \{C_{att}(X_t), Z_{id}(X_s)\}$

。被送到StyleGAN[17]进行风格化脸部绘画。如何-如图7所示，我们发现这些结果不能正确地适应身份、表情、质地和姿态。SFI-

Net的研究结果表明，所有的风格代码都是必要的。为IGFI框架（col 9）。

有效的语境损失：如图11所示，语境视觉损失有助于有效地保存焦油的纹理（第9和10栏）。如果舍弃 L_{cx}

，结果会表现出很差的性能，有很多纹理失真（第8栏）。完善JR-Net。

在第一阶段，面部遮挡已经被去除（图9（a），（d）和（e）），并且在这些图像之间存在着形状和外观的不一致。

因此，在SFI-

Net中生成的人脸中有关身份和属性的细节需要通过JR-Net的方式来完善。

5. 总结

我们提出了FaceInpainter来实现有效的身份互换，实现高保真的身份和属性。与以往的身份互换方法不同，FaceInpainter采用了一种内绘的方式进行互换，这样可以更好地保持各种异质场景的形成。SFI-Net使用三维先验来使渲染过程中目标属性的表达更加准确。此外，上下文结构和纹理细节在JR-Net中得到了很好的完善。广泛的实验表明，我们的方法可以处理高质量的、感知遮挡的身份适应异质领域。

6. 鸣谢

这项工作得到了北京市自然科学基金（批准号JQ18017）、中科院青年创新促进会（批准号Y201929）和国家自然科学基金（批准号U20A20223）的部分资助。

参考文献

- [1] Deepfakes github. <https://github.com/deepfakes/faceswap>.
- [2] Faceswap. <https://github.com/ondyari/FaceForensics/tree/master/dataset/FaceSwapKowalski>, Accessed:2019年9月30日。
- [3] Volker Blanz 和 Thomas Vetter. 一个用于合成3D脸部的可变形模型。在 *第26届计算机图形和交互技术年度会议上*, 第187-194页, 1999年。
- [4] 陈东, 任少卿, 魏一晨, 曹旭东, 和孙健. 联合级联人脸检测和对齐。在 *欧洲计算机视觉会议上*, 第109-122页。Springer, 2014.
- [5] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. Simswap: 一个高效的高保真人脸互换框架。在 *第28届ACM国际多媒体会议论文集中*, 第2003-2011页, 2020年。
- [6] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: 用于深度人脸识别的附加角差损失。在 *IEEE 计算机视觉和模式识别会议上*, 第4690-4699页, 2019年。
- [7] 邓启耀, 曹杰, 刘云帆, 柴振华, 李琦, 孙振安. *arXiv 预印本 arXiv:2006.02051*, 2020年, 参考文献引导的脸部组件编辑。
- [8] 邓宇, 杨蛟龙, 陈东, 文芳, 和童欣. 通过三维模仿-对比学习生成解耦和可控的人脸图像。在 *IEEE/CVF 计算机视觉和模式识别会议 (CVPR)* 上, 2020年6月。
- [9] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. 用弱监督学习进行精确的三维人脸重建。从单一图像到图像集。在 *IEEE 计算机视觉和模式识别研讨会论文集中*, 第0-0页, 2019年。
- [10] 傅朝友, 吴翔, 胡一波, 黄怀波, 和何然. 双变异生成的低拍异质人脸识别。在 *NeurIPS*, 2019.
- [11] 傅朝友, 吴翔, 胡一波, 黄怀波, 何然. Dvg-face: 用于异质人脸识别的双变量生成。 *TPAMI*, 2021.
- [12] Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei, and Stan Z Li. 实现快速、准确和稳定的三维密集人脸对齐。 *arXiv 预印本 arXiv:2009.09960*, 2020.
- [13] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 由两个时间尺度的更新规则训练的GAN收敛到局部纳什平衡。在 *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 6626-6637, 2017.
- [14] Xun Huang and Serge Belongie. 带有自适应实例规范化的实时任意风格转移。在 *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501-1510, 2017.
- [15] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. *arXiv 预印本 arXiv:1710.10196*, 2017.
- [16] Tero Karras, Samuli Laine, and Timo Aila. 基于风格的生成器架构, 用于生成式对抗网络。在 *IEEE 计算机视觉和模式识别会议论文集中*, 第4401-4410页, 2019年。
- [17] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 分析和提高stylegan的图像质量。在 *IEEE/CVF 计算机视觉和模式识别会议论文集中*, 第8110-8119页, 2020.
- [18] 李灵芝, 鲍建民, 杨浩, 陈东, 和文芳. 推进高保真身份互换的赝品检测。在 *IEEE/CVF 计算机视觉和模式识别会议 (CVPR)* 论文集, 2020年6月。
- [19] 李培培, 刘英禄, 史海林, 吴翔, 胡一波, 何然, 孙振南. 用于数据有限的人脸解析的双结构分解变体生成。In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 556-564, 2020.
- [20] 李晓明, 陈超峰, 周尚辰, 林贤辉, 左旺蒙, 张磊. 通过深度多尺度分量字典进行盲目的面部修复。In *ECCV*, 2020.
- [21] YF Li, NN Chen, and JC Zhang. 快速和高灵敏度的聚焦评估功能。 *计算机应用研究*, 27(4):1534-1536, 2010.
- [22] Y. Li, S. Liu, J. Yang, and M. Yang. 生成式人脸合成。In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5892-5900, 2017.
- [23] Roey Mechrez, Itamar Talmi, and Lihi Zelnik-Manor. 用非对齐数据进行图像转换的上下文损失。在 *欧洲计算机视觉会议 (ECCV)* 论文集中, 第768-783页, 2018.
- [24] 门一芳, 毛一鸣, 蒋玉宁, 马维英, 和连周辉. 使用属性分解的GAN的可控人物图像合成。在 *IEEE/CVF 计算机视觉和模式识别会议论文集中*, 第5084-5093页, 2020.
- [25] Ashutosh Mishra, Shyam Nandan Rai, Anand Mishra, and CV Jawahar. Iit-cfw: 一个野外卡通脸谱的基准数据库。In *European Conference on Computer Vision*, pages 35-47. Springer, 2016.
- [26] MuggleWang. Cosfacepytorch. https://github.com/MuggleWang/CosFace_pytorch, Accessed: 2018.
- [27] Yuval Nirkin, Yosi Keller, and Tal Hassner. Fsgan: 主题不可知的脸部交换和重演。在 *IEEE 计算机视觉国际会议论文集中*, 第7184-7193页, 2019年。
- [28] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 具有空间适应性的Normalization的语义图像合成。在 *IEEE 计算机视觉和模式识别会议上*, 第2337-2346页, 2019年。

- [29] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. 深度人脸识别。In *British Machine Vision Conference*, 2015.
- [30] Stanislav Pidhorskyi, Donald A Adjeroh, and Gianfranco Doretto. 对抗性潜伏自动编码器。在 *IEEE 计算机学会计算机视觉和模式识别会议 (CVPR) 论文集*, 2020年。
- [31] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics++: 学习检测被操纵的面部图像。在 *IEEE 国际计算机视觉会议论文集中*, 第1-11页, 2019年。
- [32] Nataniel Ruiz, Eunji Chong, and James M. Rehg. 没有关键点细粒度头部姿势估计。In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.
- [33] Karen Simonyan and Andrew Zisserman. 用于大规模图像识别的超深度卷积网络。In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [34] 宋林森, 曹杰, 宋凌霄, 胡一波, 和何然. 几何感知的人脸完成和编辑。在 *AAAI 人工智能会议论文集中*, 第33卷, 第2506-2513页, 2019年。
- [35] Luan Tran, Feng Liu, and Xiaoming Liu. 实现高保真的非线性三维人脸可变形模型。In *Proceeding of IEEE Computer Vision and Pattern Recognition*, Long Beach, CA, June 2019.
- [36] Luan Tran and Xiaoming Liu. 非线性3D人脸可变形模型。在 *IEEE 计算机视觉和模式识别 (CVPR)*, 犹他州盐湖城, 2018年6月。
- [37] 王浩, 王一彤, 周正, 季星, 龚迪红, 周景超, 李志峰, 刘伟. Cosface: 用于深度人脸识别的大边际余弦损失。In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265-5274, 2018.
- [38] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 高分辨率图像同步和语义操作的条件甘斯。在 *IEEE 计算机视觉和模式识别会议论文集中*, 第8798-8807页, 2018年。
- [39] 王小刚和唐晓鸥. 脸部照片-草图的协同和识别。 *IEEE transactions on pattern analysis and machine intelligence*, 31(11):1955-1967, 2008.
- [40] 王新涛, 余珂, 董超, 和陈变劳. 通过深度空间特征变换恢复图像超分辨率中的真实纹理。In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [41] Xiang Wu, Huaibo Huang, Vishal M Patel, Ran He, and Zhenan Sun. 用于异质人脸识别的分解变异表示。在 *AAAI 人工智能会议论文集*, 第33卷, 第9005-9012页, 2019年。
- [42] Aijing Yu, Haoxue Wu, Huaibo Huang, Zhen Lei, and Ran He. Lamp-hq: 一个大规模的多姿多彩的高品质数据库和nir-vis人脸识别的基准。 *International Journal of Computer Vision*, pages 1-17, 2021.
- [43] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. 带有上下文关注的生成性图像绘画。In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5505-5514, 2018.
- [44] 余佳辉, 林哲, 杨继美, 沈晓辉, 卢新, 和 Thomas S Huang. 用门控卷积进行自由形式的图像绘画。在 *IEEE 国际计算机视觉会议上*, 第4471-4480页, 2019年。
- [45] S. Zhang, R. He, Z. Sun, and T. Tan. Demeshnet: 盲目的脸部画像, 用于深度网状脸部验证。 *IEEE Transactions on Information Forensics and Security*, 13 (3) : 637-647, 2018.
- [46] C. Zheng, T. Cham, and J. Cai. Pluralistic image completion. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1438-1447, 2019.
- [47] Hang Zhou, Jihao Liu, Ziwei Liu, Yu Liu, and Xiaogang Wang. Rotate-and-render: 来自单视角图像的无监督逼真人脸旋转。在 *IEEE /CVF 计算机视觉和模式识别会议 (CVPR) 论文集*, 2020年6月。
- [48] 朱翔宇, 刘晓明, 雷震, 和 Stan Z Li. 全姿态范围内的人脸对齐。一个三维整体解决方案。 *IEEE transactions on pattern analysis and machine intelligence*, 41(1):78-92, 2017.
- [49] Zhen Zhu, Tengting Huang, Baoguang Shi, Miao Yu, Bofei Wang, and Xiang Bai. 用于生成人物图像的渐进式姿势注意力转移。In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2347-2356, 2019.