# HifiFace：三维形状和语义先验指导下的高保真人脸互换

王雨涵[1,2*], 陈旭[1,3*], 朱俊伟[1], 朱文清[1], 戴颖[1†] 王成杰[1], 李吉林[1], 吴永健[1],
黄飞跃[1]和纪蓉蓉[3,4†] 腾讯公司优图实验室
[2]浙江大学
[3]厦门大学信息学院人工智能系媒体分析与计算实验室
[4]厦门大学人工智能研究所
https://johann.wang/HifiFace



图1：由我们的HifiFace产生的脸部交换结果。目标图像中的脸被源图像中的脸所取代。

## 摘要

在这项工作中，我们提出了一种高保真的人脸互换方法，称为HifiFace，它可以很好地预处理源脸的脸部形状，并产生照片般逼真的结果。与其他现有的只使用人脸识别模型来保持身份相似性的换脸工作不同，我们提出了三维形状感知身份，通过三维模型和几何监督来控制脸部形状。
三维人脸重建方法。同时，我们在
引入语义面部融合模块，以选择
我们对编码器和解码器的组合进行了优化，并进行了自适应混合，这使得结果更加符合照片的真实性。在野外进行的大量人脸实验表明，我们的方法可以更好地保持身份，特别是在脸部形状上，并且可以比以前的最先进的方法产生更逼真的结果。

## 1 简介

脸部交换是一项从源脸部生成具有身份的图像，从目标图像生成具有属性（*如*姿势、表情、光照、背景*等*）的图像的任务（如图1所示），它在电影业[Alexander*等人*，2009]和电脑游戏中引起了很大的兴趣，具有很大的应用潜力。

为了产生高保真的人脸互换结果，有几个关键问题：（1）结果脸的身份，包括*脸部形状*应该与源脸接近。(2)结果应该是*照片般真实的*，忠实于目标脸的表情和姿势，并与目标图像的细节相一致，如照明、背景和遮挡。

为了保持生成的人脸的身份，以前的工作[Nirkin*等人*，2018；Nirkin*等人*，2019；Jiang*等人*，2020]通过3DMM拟合或人脸地标引导重现生成人脸内部区域，并将其混合到目标图像中，如图2（a）所示。这些方法在身份相似性方面很弱，因为3DMM不能模仿身份细节，而目标地标包含目标图像的身份。另外，混合阶段限制了脸部形状的变化。如图2(b)所示，[Liu *et al.* , 2019; Chen *et al.* , 2020]从人脸识别网络中获得支持。

*平等的贡献。于晗和徐晓明都是腾讯优图实验室的实习生时完成的工作。
†通讯作者。

工作，以提高身份相似度。然而，人脸识别网络更注重于纹理，对几何结构不敏感。因此，这些方法不能稳健地保留准确的脸部形状。

至于生成照片般真实的结果，[Nirkin *et al.*，2018; Nirkin *et al.*，2019]使用泊松混合来修复光照，但它往往会导致重影，并且不能处理复杂的外观条件。[Jiang等人，2020年；Zhu等人，2020年；Li等人，2019年]设计了一个额外的基于学习的阶段来优化照明或遮挡问题，但他们很挑剔，不能在一个模型中解决所有问题。

为了克服上述缺陷，我们提出了一个新颖而优雅的端到端学习框架，名为HifiFace，通过三维形状和语义先验来生成高保真的互换脸。具体来说，我们首先通过三维人脸重建模型对源脸和目标脸的系数进行回归，并将其作为形状信息重新组合。然后，我们将其与来自人脸识别网络的身份向量连接起来。我们明确地使用三维几何结构信息，并将重组的三维人脸模型与源的身份、塔拉的表情和目标的姿态作为辅助监督来执行精确的人脸形状转移。通过这种专门的设计，我们的框架可以实现更多的相似身份识别，特别是在脸型上。

此外，我们引入了一个语义面部融合（SFF）模块，使我们的结果更加符合照片的真实性。像照明和背景这样的属性需要空间信息，而高图像质量的结果需要详细的纹理信息。编码器中的低级特征包含空间和纹理信息，但也包含来自目标图像的丰富特征。因此，为了更好地保留属性而不损害身份，我们的SFF模块通过学习的自适应人脸掩码来整合低级别的编码器特征和解码器特征。最后，为了克服遮挡问题并实现完美的背景，我们也通过学习的人脸面具来混合输出到目标。与[Nirkin *et al.*，2019]使用目标图像的人脸掩码直接融合不同，HifiFace在扩张的人脸语义分割的指导下同时学习人脸掩码，这有助于模型更加关注面部区域并在边缘周围进行自适应融合。HifiFace在一个模型中处理图像质量、遮挡和照明问题，使结果更符合照片的真实性。广泛的实验表明，我们的结果超过了其他具有较大面部变化的野生人脸图像的先进技术（SOTA）。

我们的贡献可以概括为以下几点。

1. 我们提出了一个新颖而优雅的端到端学习框架，名为HifiFace，它可以很好地保留源脸的脸部形状，并产生高清晰度的脸部交换结果。
2. 我们提出了一种三维形状感知的身份提取器，它可以生成具有精确形状信息的身份向量，以帮助保留源脸的脸部形状。
3. 我们提出了一个语义面部融合模块，它可以解决遮挡和光照问题，并产生具有高图像质量的再结果。

## 2 相关工作

**基于三维的方法。**三维可变形模型（3DMM）将例子的形状和纹理转化为矢量空间表示[Blanz and Vetter, 1999]。[Thies等人，2016]通过对两张脸拟合一个三维可变形脸部模型，将表情从源脸转移到目标脸。[Nirkin等人，2018]通过3DMM转移表情和姿势，并训练一个人脸分割网络以保留目标脸部的遮挡。这些基于三维的方法遵循一个面向源的管道，如图2（a），它只通过三维拟合生成脸部区域，并通过目标脸部的面具将其混合到目标图像中。由于3DMM和渲染器不能模拟复杂的光照条件，它们受到不真实的纹理和光照的影响。此外，混合阶段限制了脸部形状。相比之下，我们的HifiFace通过3DMM的几何信息准确地保留了脸部形状，并通过对编码器和解码器特征的先导性重新组合实现了现实的纹理和属性。

**基于GAN的方法。**自[Goodfellow *et al.*，2014]提出以来，GAN在生成假图像方面表现出了巨大的能力。[Isola *et al.*，2017]提出了一个通用的图像到图像的翻译方法，这证明了条件GAN架构在交换脸部方面的潜力，尽管它需要配对数据。

基于GAN的人脸互换方法主要遵循面向源的管道或面向目标的管道。[Nirkin等人，2019年；Jiang等人，2020年]遵循图2(a)中面向源的管道，该管道使用人脸地标来组成人脸重现。但它可能带来较弱的身份相似性，而且混合阶段限制了脸部形状的变化。[Liu *et al.*, 2019; Chen *et al.*，2020; Li *et al.*，2019]遵循图2(b)中面向目标的管道，使用人脸识别网络提取身份，并使用解码器将编码器特征与身份融合，但他们不能稳健地预先提供准确的脸部形状，而且图像质量较差。相反，图2(c)中的HifiFace将人脸识别网络替换为3D形状感知的身份提取器，以更好地保持包括脸型在内的身份，并引入了一个SFF模式。为了进一步提高真实性，在译码器之后，还可以添加一个新的 "小程序"。

其中，FaceShifter [Li *et al.*, 2019]和Sim- Swap [Chen *et al.*，2020]遵循面向目标的管道，可以产生高保真的结果。FaceShifter [Li *et al.*，2019]利用了一个两阶段的框架，实现了最先进的身份识别性能。但是，尽管使用了一个额外的固定阶段，它还是不能完美地保留照明。然而，HifiFace可以在一个阶段很好地保留照明和身份。同时，HifiFace可以产生比FaceShifter更高质量的照片逼真的结果。[Chen *et al.*，2020]提出了弱特征匹配损失以更好地保留属性，但它损害了身份相似性。而HifiFace可以更好地保留属性，并且不损害身份。

## 3 办法

让$I_s$为源图像，$I_t$为目标图像，分别是。我们的目标是生成具有$I_s$的身份和$I_t$的属性的结果图像$I_r$。如图2（c）所示，我们的管道由四个部分组成：编码器部分、解码器

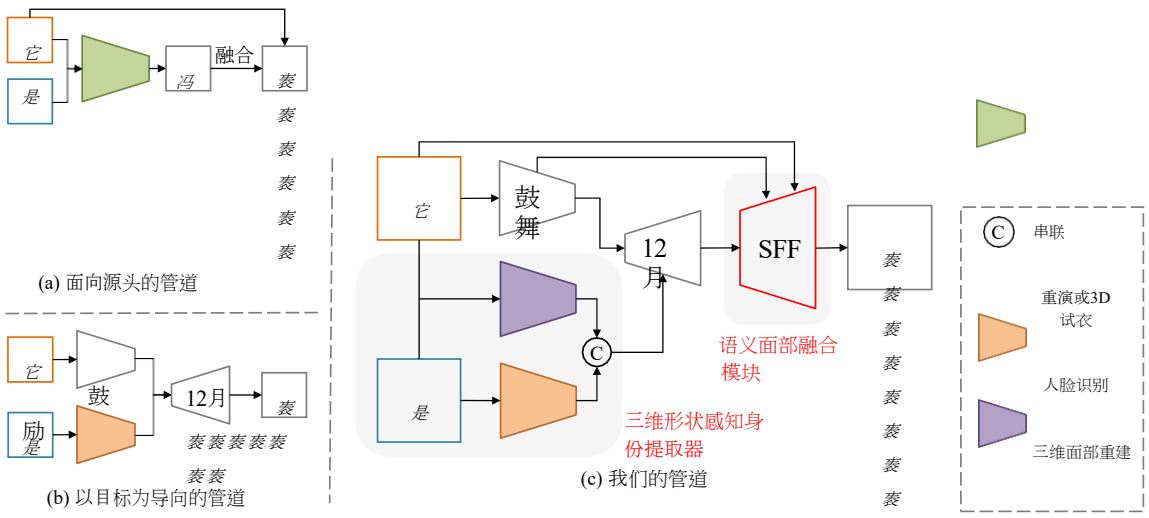


图2：以前的工作和我们的HifiFace的管道。(a) 面向源头的管道使用三维拟合或重演来生成内部人脸区域，并将其融合到目标图像中，其中$I_{Fr}$指的是结果中的人脸区域。(b) 面向目标的管道使用人脸识别网络来准确识别，并在解码器中结合编码器的特征和身份。(c) 我们的管道由四个部分组成：编码器部分、解码器部分、3D形状感知身份提取器和SFF模块。编码器从 它 那里提取特征，解码器将编码器的特征和三维形状感知的身份特征融合起来。最后，SFF模块帮助进一步提高图像质量。

部分，三维形状感知身份提取器（第3.1节），以及SFF模块（第3.2节）。首先，我们将$I_t$作为编码器的输入，并使用几个重块[He *et al.*, 2016]来获得属性特征。然后，我们使用三维形状感知身份提取器来获得三维形状感知身份。之后，我们在解码器中使用带有自适应实例归一化的重块[Karras *et al.*, 2019]来融合三维形状感知身份和属性特征。最后，我们使用SFF模块来获得更高的分辨率，使结果更符合照片的真实性。

### 3.1 三维形状感知身份提取器

大多数基于GAN的方法只使用人脸识别模型来获得人脸交换任务中的身份信息。

然而，人脸识别网络更注重纹理，对几何结构不敏感。为了获得更多
我们引入了3DMM，并使用预先训练好的最先进的3D人脸重建模型[Deng *et al.*, 2019]作为形状特征编码器，它代表了
脸部形状**S**由一个仿生模型决定。

$$\mathbf{S} = \mathbf{S}(\alpha, \beta) = \bar{\mathbf{S}} + \mathbf{B}_{id}\,\alpha + \mathbf{B}_{exp}\,\beta_{\circ} \tag{1}$$

其中，$\bar{\mathbf{S}}$是平均脸形；$\mathbf{B}_{id}$，$\mathbf{B}_{exp}$是身份和表情的PCA基数；$\alpha$和$\beta$是生成三维脸部的对应系数向量。

如图3(a)所示，我们将3DMM系数$\boldsymbol{c}_s$和$\boldsymbol{c}_t$，包含源脸和目标脸的身份、表情和位置，由3D人脸重构模型$\boldsymbol{F}_{3d}$。然后，我们通过$\boldsymbol{c}_{fuse}$生成一个新的3D人脸模型，包含源的身份、目标的表情和位置。
姿态。请注意，姿态系数并不决定脸部形状，但在计算损失时可能会影响二维地标的位置。我们不使用纹理和光照的共同效率，因为纹理重建仍然不令人满意。最后，我们将$\boldsymbol{c}_{fuse}$与预先训练好的最先进的人脸识别模型$\mathrm{F}_{id}$所提取的身份特征$\boldsymbol{v}_{id}$连接起来，得到最终的向量$\boldsymbol{v}_{sid}$，称为三维形状感知身份。因此，HifiFace实现了良好的身份信息，包括几何结构，这有助于保留源图像的脸

### 3.2 语义面部融合模块

**特征层面。**低级特征包含丰富的空间信息和纹理细节，这可能大大有助于产生更逼真的照片结果。在此，我们提出了SFF模块，不仅充分利用了低级编码器和解码器的特征，而且还克服了由于低级编码器特征中的目标身份信息而避免损害身份的矛盾。

如图3(b)所示，当解码器特征$\boldsymbol{z}_{dec}$的大小为目标的1/4时，我们首先预测一个面部面具$\boldsymbol{M}_{low}$。然后，我们用$\boldsymbol{M}_{low}$混合$\boldsymbol{z}_{dec}$，得到$\boldsymbol{z}_{fuse}$，表述为。

$$\boldsymbol{z}_{fuse} = \boldsymbol{M}_{low} \odot \boldsymbol{z}_{dec} + (1 - \boldsymbol{M}_{low}) \odot \sigma(\boldsymbol{z}_{enc})_{\circ} \tag{2}$$

其中$\boldsymbol{z}$指的是低级别的编码器特征，尺寸为1/4的，而$\sigma_{编码}$指的是一个重块。[He *et al.*, 2016]。部形状。

SFF的关键设计是调节恩人的注意力。
编码器和解码器，这有助于区分身份和贡献。具体来说，非面部区域的解码器特征可能会被插入的源身份信息所破坏，因此我们用干净的低级编码器特征来代替它，以避免潜在的伤害。而面部区域的解码器特征，包含了源脸部丰富的身份信息，不应受到目标的干扰，因此我们保留面部区域的解码器特征。

在特征级融合之后，我们生成$\boldsymbol{I}_{low}$，以计算辅助损失，从而更好地分解身份和属性。然后，我们使用一个4x升值模块$\boldsymbol{F}_{up}$，该模块包含几个重块，以更好地融合特征图。
基于$\boldsymbol{F}_{up}$，我们的HifiFace可以方便地生成更高的分辨率结果（*例如*，512 × 512）。

**图像层面。为了**解决闭塞问题并更好地保留背景，以前的工作[Nirkin等人，2019年；Natsume等人，2018年]直接使用目标脸部的面具。然而，它带来了伪影，因为脸部形状可能会改变。相反，我们使用SFF来学习一个视觉扩张的面具，并接受脸部形状的变化。具体来说，我们预测一个3通道的$\boldsymbol{I}_{out}$和1通道的$\boldsymbol{M}_r$，并混合$\boldsymbol{I}_{out}$到目标图像，由$\boldsymbol{M}_r$，表述为。

$$\boldsymbol{i}_r = \boldsymbol{m}_r \odot \boldsymbol{i}_{out} + (1 - \boldsymbol{m}_r) \odot \boldsymbol{i}_t 。 \quad (3)$$

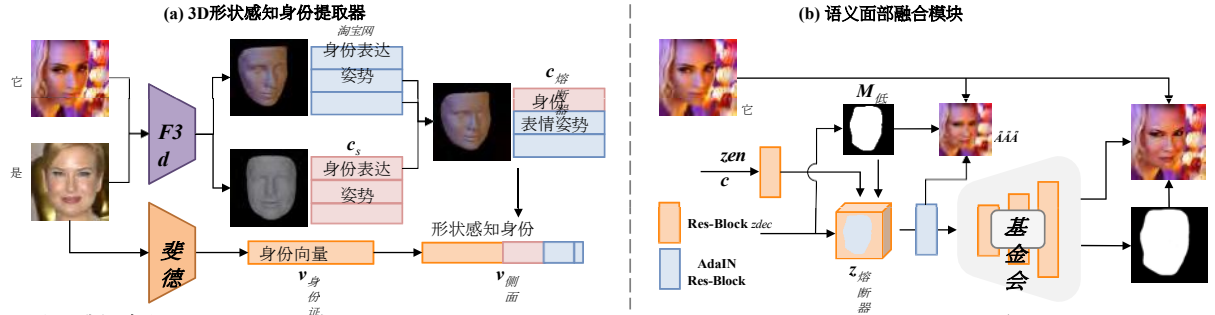


图3：3D形状感知身份提取器和SFF模块的细节。(a) 三维形状识别提取器使用$F_{3d}$（三维人脸重建网络）和$F_{id}$（人脸识别网络）来生成形状识别。(b) SFF模块通过$M_{low}$重新组合编码器和解码器的特征，并通过$M_r$进行最后的混合。$F_{up}$意味着上采样模块。

总之，在SFF模块的帮助下，HifiFace可以生成具有高图像质量的照片逼真的结果，并很好地保持照明和环境。请注意，尽管脸部形状发生了变化，这些能力仍然有效，因为面具已经被放大，我们的SFF受益于预测脸部轮廓的绘画。

### 3.3 损失功能

**三维形状识别（SID）损失。**SID损失包含形状损失和身份损失。我们使用二维地标关键点作为几何监督来约束人脸形状，这在三维人脸重建中被广泛使用[Deng *et al.*, 2019]。首先，我们使用网格渲染器，通过源图像身份和目标图像表达和姿势的系数来生成3D人脸模型。然后，我们通过回归3DMM系数，生成$I_r$和$I_{low}$的三维人脸模型。最后，我们将重建的人脸形状的三维面部标志顶点投影到图像获得地标$\{q^{fuse}\}$、$\{q^r\}$和$\{q^{low}\}$。

$$L_{形状} = \frac{1}{N}\sum_{n=1}^{N} ||q_n^{fuse} - q_n^r||_1 + ||q_n^{fuse} - q_n^{low}||_1 。 \quad (4)$$

此外，我们使用身份损失来保留源图像的身份。

$$L_{id} = (1 - cos(\upsilon_{id}(I_s), \upsilon_{id}(I_r))) + (1 - cos(\upsilon_{id}(I_s), \upsilon_{id}(I_{low}))) 。 \quad (5)$$

其中，$\upsilon_{id}$表示由$F_{id}$生成的身份向量，$cos(,)$表示两个向量的余弦相似度。最后，我们的SID损失被表述为。

$$L_{sid} = \lambda_{shape} L_{shape} + \lambda_{id} L_{id} 。 \quad (6)$$

其中$\lambda_{id} = 5$，$\lambda_{shape} = 0.5$。

**现实性损失。**真实性损失包含分割损失、重新构建损失、周期损失、感知损失和对抗性损失。具体来说，SFF模块中的$M_{low}$和$M_r$都是在SOTA人脸分割网络HRNet [Sun *et al.*, 2019]的指导下进行的。我们对目标图像的遮罩进行扩张，以消除脸部形状变化的限制，得到$M_{tar}$。分割损失表述为。

$$L_{seg} = ||R(M_{tar}) - M_{low}||_1 + ||M_{tar} - M_r||_1 , \quad (7)$$

其中$R(.)$表示调整大小的操作。

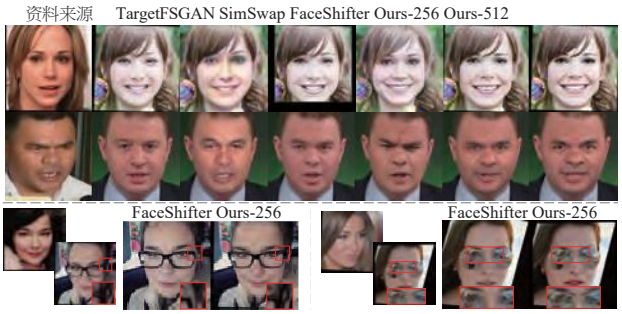


图4：与FSGAN、SimSwap和FaceShifter的比较。我们的结果能够很好地保留源脸的形状和目标属性，并且具有更高的图像质量，即使在处理遮挡的情况下。

如果$I_s$和$I_t$有相同的身份，预测的图像应该与$I_t$相同。所以我们使用重建损失来给出像素级的监督。

$$l_{rec} = ||i_r - i_t||_1 + ||i_{low} - r(i_t)|_1 。 \quad (8)$$

循环过程可以在换脸中进行任务也是如此。让$I_r$作为再目标图像，原始目标图像作为再源图像。在循环过程中，我们希望生成的结果具有再源图像的身份和再目标图像的属性，这意味着它应该与原始目标图像相同。循环损失是对像素监督的一种补充，可以帮助产生高保真的结果。

$$l_{cyc} = ||i_t - g(i_r, i_t)||_1 。 \quad (9)$$

其中$G$指的是HifiFace的整个发生器。

为了捕捉精细的细节，进一步提高真实性。我们遵循[Zhang *et al.*, 2018]中的Learned Perceptual Image Patch Similarity（LPIPS）损失和[Choi *et al.*, 2020]中的对抗性目标。因此，我们的真实性损失被表述为。

$$L_{real} = L_{adv} + \lambda_0 L_{seg} + \lambda_1 L_{rec} + \lambda_2 L_{cyc} + \lambda_3 L_{lpips}, \quad (10)$$

其中$\lambda_0 = 100$, $\lambda_1 = 20$, $\lambda_2 = 1$ 和 $\lambda_3 = 5$。

**整体损失。** 我们的全部损失总结如下。

$$L = L_{sid} + L_{real} 。 \quad (11)$$

| 方法 | ID↑ | 姿势↓ | 形状↓ | MAC ↓ | FPS↑ |
|---|---|---|---|---|---|
| 换脸 | 54.19 | 2.51 | 0.610 | - | - |
| 脸部移位 | 97.38 | 2.96 | 0.511 | 121.79 | 22.34 |
| 模拟换位 | 92.83 | **1.53** | 0.540 | **55.69** | **31.17** |
| 我们的-256 | **98.48** | 2.63 | **0.437** | 102.39 | 25.29 |

表1:FaceForensics++的定量实验。FPS是在GPU V100下测试的。


图5：（a）与AOT的比较。(b) 与DF的比较。

## 4 实验

**实施细节。**我们选择VGGFace2 [Cao *et al.* , 2018] 和Asian-Celeb [DeepGlint, 2020] 作为训练集。对于我们的分辨率为256 的模型（*即*, Ours-256），我们删除了尺寸小于256的图像，以提高图像质量。对于每张图片，我们用5个土地------来对齐脸部。标记并裁剪为256x256[Li *et al.* , 2019]，其中包含整个面部和一些背景区域。对于我们更精确的模型（*即*, Ours-512），我们采用了一个人像增强网络[Li *et al.* 响应地在$F_{up}$ 的SFF中添加另一个res-block，相比之下到Ors-256。具有相同身份的训练对的比例为50%。使用ADAM[Kingma and Ba, 2014]，β1 = 0；β2 = 0.99，学习率 = 0.0001。该模型以200K步进行训练，使用4个V100 GPU和32个批次大小。

### 4.1 定性比较

首先，我们将我们的方法与图4中的FSGAN [Nirkin *et al.* , 2019]、SimSwap [Chen et *al.* , 2020] 和FaceShifter [Li *et al.* , 2019]，图5中的AOT [Zhu *et al.* , 2020] 和Deeper-Forensics（DF）[Jiang *et al.*

如图4所示，FSGAN与目标脸部形状相同，它也不能很好地转移目标图像的照明。SimSwap不能很好地保留源图像的身份，特别是脸部形状，因为它使用了特征匹配损失，并且更注重属性。FaceShifter表现出很强的身份保留能力。

| 方法 | FF++ | | DFDC | |
|---|---|---|---|---|
| | AUC↓ | AP↓ | AUC↓ | AP↓ |
| 脸部换位 | 66.83 | 65.99 | 92.30 | 92.54 |
| 脸部换位 | 41.62 | 42.92 | 77.18 | 76.50 |
| 模拟交换 | 76.44 | 72.63 | 78.80 | 78.44 |
| 吾思-256 | **38.97** | **41.54** | **62.29** | **59.99** |

表2:FF++和DFDC的AUC和AP方面的结果。

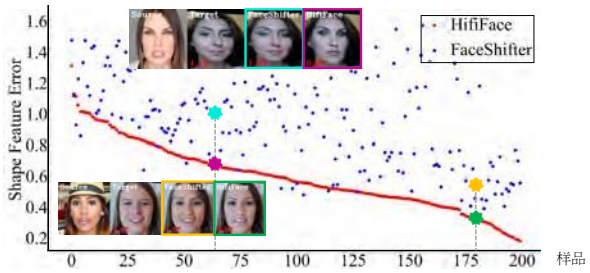


图6：在200 个形状差异较大的FF++对中，$I_r$和$I_s$的脸部形状误差。样本按HifiFace的形状误差进行排序。相同的列索引表示相同的源/目标对。

它有两个局限性：（1）属性恢复，而我们的HifiFace可以很好地保留所有的属性，如脸部颜色、表情和咬合。(2)
两个阶段的复杂框架工作，而HifiFace提出了一个更优雅的端到端框架，甚至有更好的恢复图像。如图5(a)所示，AOT是专门为克服照明问题而设计的，但在身份相似性和自由度方面比较弱。如图5(b)所示，DF减少了风格不匹配的坏情况，但在身份相似性方面也很弱。相比之下，我们的HifiFace不仅完美地保留了光照和脸部风格，而且很好地捕捉了源图像的脸部形状，生成了高质量的交换脸。更多的结果可以在我们的 补 充 材 料中找到。

### 4.2 量化比较

接下来，我们在FaceForensics (FF)++ [Rossler *et al.* , 2019]
数据集上对以下指标进行了定量比较。ID检索、姿势误差、脸型误差和脸部伪造检测算法的性能，再次证明我们的HifiFace的有效性。对于FaceSwap[FaceSwap,]和FaceShifter，我们从每个视频中均匀地抽出10帧，组成一个10K 的测试集。对于SimSwap和我们的HifiFace，我们用上述相同的源和目标对产生脸部交换的重新结果。

对于ID检索和姿势错误，我们遵循[Li *et al.* , 2019; Chen *et al.* , 2020]中的相同设置。如表1所示，HifiFace取得了最好的ID检索得分，并且在姿势保持方面与其他的人相比也是不错的。对于脸形误差，我们使用另一个三维脸部重建模型[Sanyal *et al*, 2019]对每个测试面的系数进行回归。误差是我们的HifiFace实现了最低的人脸形状误差，这是由交换的人脸和其源脸之间的身份系数的L2距离计算出来的。参数和速度的比较也显示在表1中，我们的HifiFace更快
与FaceShifter相比，它的生成质量更高。


图7：三维形状感知身份提取器的消融研究。

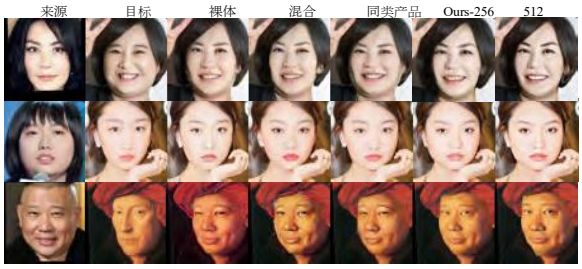


图8：SFF模块的消融研究。

为了进一步说明HifiFace在控制脸部形状方面的能力，我们在图6中可视化了HifiFace和FaceShifter [Li *et al.*，2019]
之间的样本形状差异。结果显示，当源和目标在脸部形状上有很大差异时，HifiFace明显优于Facehifter，95%的样本具有较小的形状误差。

此外，我们应用FF++ [Rossler *et al.*，2019]和DeepFake Detection Challenge (DFDC) [Dolhan- sky *et al.*，2019; selimsef, 2020]的模型来检验HifiFace的真实性。测试集包含了每种方法的10K互换脸和10K来自FF++的真实脸。如表2所示，HifiFace获得了最好的分数，表明更高的保真度有助于进一步提高人脸伪造检测。

### 4.3 对HifiFace的分析

**三维形状感知身份。**为了验证形状监督$L_{shape}$对脸部形状的有效性，我们训练另一个模型Ours-n3d，它取代了形状感知身份矢量与$\boldsymbol{F}$的法向同一矢量$_{id}$。 如图所示图7，Ours-n3d的结果几乎不能改变脸部形状或有明显的人工痕迹，而Ours-256的结果可以产生更多类似脸部形状的结果。

**语义面部融合。**为了验证SFF模块的必要性，我们与3个基线模型进行了比较：（1）"裸露"，删除了特征层面和图像层面的融合。(2) "混合"，去掉了特征级融合。(3) 'Concat'，用串联法取代特征级融合法。如图8所示，"Bare "不能很好地保留背景和遮挡，"Blend "缺乏可读性，而 "Concat在身份相似性方面很弱，这证明了SFF模式可以帮助保留属性并提高图像质量而不损害身份。

**脸部交换中的脸形保护。**脸部形状的保存对于脸部交换来说是相当困难的，这不是


图9：与直接使用蒙版混合的结果比较。Blend-T"、"Blend-DT "和 "Blend-R "分别是指通过目标图像的掩膜、目标图像的扩展掩膜和裸体结果的掩膜将裸体结果混合到目标图像上。

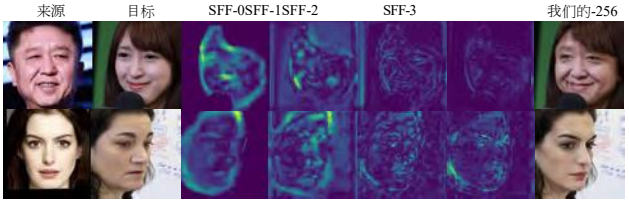


图10：SFF的差异特征图。

仅仅是因为难以获得形状信息，而且当脸部形状发生变化时，还面临着重新绘制的挑战。混合是一种保留闭塞和背景的有效方法，但当脸部形状发生变化时，它很难被应用。如图9所示，当源脸比目标脸更胖时（ 第1行），它可能会限制脸部的变化在Blend-T中的形状。如果我们使用Blend-DT或Blend-R，它不能遮挡处理得很好。当源脸比目标脸薄时（第2行），在Blend-T和Blend-DT中很容易带来脸部周围的伪影，在Blend-R中可能导致双重脸。相比之下，我们的HifiFace可以应用混合，没有上述问题，因为我们的SFF模块有能够对预测掩码的边缘进行涂抹。

为了进一步说明SFF如何解决这个问题，我们展示了SFF模型中每个阶段的差异特征图，命名为SFF-0~3，在$(I_s, I_t)$和$(I_t, I_t)$的输入之间。其中$(I_s, I_t)$获得了Ors-256，$(I_t, I_t)$实现了目标It-自己。在图10中，明亮的区域意味着脸部形状的地方变化或包含伪影。SFF模块重新组合了脸部区域和非脸部区域之间的特征，并更多地关注预测面具的轮廓，这为形状发生变化的区域的内画带来了很大好处。

## 5 结论

在这项工作中，我们提出了一种高保真的人脸交换方法，命名为HifiFace，它可以很好地保留源脸的脸部形状，并产生照片般真实的结果。我们提出了一个三维形状感知的身份提取器，以帮助预先提供包括脸部形状在内的身份。一个SFF模块被提出来，以实现特征级和图像级的更好结合，从而生成逼真的图像。广泛的实验表明，我们的方法在数量上和质量上都比以前的SOTA换脸方法产生更高的清晰度。最后但并非最不重要的是，HifiFace也可以作为一把锋利的矛，为人脸伪造检测领域的发展做出贡献。

## 参考文献


[Alexander *et al.* , 2009] Oleg Alexander, Mike Rogers, William Lambeth, Matt Chiang, and Paul Debevec.创造一个逼真的数字演员。数字艾米莉项目。In *2009 Conference for Visual Media Production*, pages 176-187.IEEE，2009年。

[Blanz and Vetter, 1999] Volker Blanz and 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。

[Cao *et al.* , 2018] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman.Vggface2:一个用于识别不同姿势和年龄的人脸的数据集。在*FG*，第67-74页。IEEE，2018。

[Chen *et al.* , 2020] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge.Simswap:一个高效的高帧率人脸互换框架。在*第28届ACM国际多媒体会议论文集中*，第2003-2011页，2020。

[Choi *et al.* , 2020] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha.Stargan v2:用于多领域的多样化图像合成。在*IEEE计算机视觉和模式识别会议论文集中*，第8188-8197页，2020年。

[DeepGlint, 2020] DeepGlint。http://trillionpairs.deepglint.com。*访问：2020-12-20*，2020。

[Deng *et al.* , 2019] 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会议上*，第0-0页，2019年。

[Dolhansky *et al.* , 2019] Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram, and Cristian Canton Ferrer.The deepfake detection challenge (dfdc) preview dataset. *arXiv preprint arXiv:1910.08854*, 2019.

[FaceSwap, ] FaceSwap. https://github.com/ondyari/faceforensics/tree/master/dataset/faceswapkowalski.*Accessed: 2020-12-20*.

[Goodfellow *et al.* , 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio.生成式对抗网。在*神经信息处理系统的进展中*，第2672-2680页，2014。

[He *et al.* , 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.用于图像识别的深度残差学习。In *Pro- ceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770-778, 2016.

[Huang *et al.* , 2020] Yuge Huang, Yuhan Wang, Ying Tai, Xiaom- ing Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang.Curricularface：深度人脸识别的自适应课程学习损失。在*IEEE计算机视觉和模式识别会议上*，第5901-5910页，2020年。

[Isola *et al.* , 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros.用条件对抗网络进行图像到图像的翻译。在*IEEE计算机视觉和模式识别会议上*，第1125-1134页，2017年。

[Jiang *et al.* , 2020] Liming Jiang, Ren Li, Wayne Wu, Chen Qian, and Chen Change Loy.Deeperforensics-1.0:一个用于真实世界人脸伪造检测的大规模数据集。在*IEEE计算机视觉和模式识别会议论文集中*，第2886-2895页。IEEE, 2020.

[Karras *et al.* , 2019] 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。在*IEEE计算机视觉和模式识别会议上*，第4401-4410页，2019年。

[Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba.Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Li *et al.* , 2019] Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen.Faceshifter:*arXiv preprint arXiv:1912.13457*, 2019.

[Li *et al.* , 2020] Xiaoming Li, Chaofeng Chen, Shangchen Zhou, Xianhui Lin, Wangmeng Zuo, and Lei Zhang.通过深度多尺度成分字典进行盲目的脸部修复。在*欧洲计算机视觉会议上*，第399-415页。Springer, 2020.

[Liu *et al.* , 2019] Jialun Liu, Wenhui Li, Hongbin Pei, Ying Wang, Feng Qu, You Qu, and Yuhao Chen.保存身份的生成对抗网络用于跨域人员再识别。*IEEE Access*, 7:114021-114032, 2019.

[Natsume *et al.* , 2018] Ryota Natsume, Tatsuya Yatagawa, and Shi- geo Morishima.Fsnet:基于图像的人脸互换的身份感知生成模型。In *Asian Conference on Computer Vision*, pages 117-132.Springer, 2018.

[Nirkin *et al.* , 2018] Yuval Nirkin, Iacopo Masi, Anh Tran Tuan, Tal Hassner, and Gerard Medioni.关于人脸分割、人脸互换和人脸感知。在*FG*，第98-105页。IEEE，2018年。

[Nirkin *et al.* , 2019] Yuval Nirkin, Yosi Keller, and Tal Hassner.Fs- gan:主体不可知的面部交换和重演。In *ICCV*, 2019.

[Rossler *et al.* , 2019] 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。

[Sanyal *et al.* , 2019] Soubhik Sanyal, Timo Bolkart, Haiwen Feng, and Michael J Black.在没有3D监督的情况下，学习从图像中回归3D脸部形状和前推力。在*IEEE计算机视觉和模式识别会议论文集中*，第7763-7772页，2019年。

[selimsef, 2020] selimsef. https://github.com/selimsef/dfdc_deepfake_challenge.*访问：2021-01-10*，2020。

[Sun *et al.* , 2019] Ke Sun, Yang Zhao, Borui Jiang, Tianheng Cheng, Bin Xiao, Dong Liu, Yadong Mu, Xinggang Wang, Wenyu Liu, and Jingdong Wang.*ArXiv preprint arXiv:1904.04514*, 2019.

[Thies *et al.* , 2016] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner.Face2face:实时人脸捕捉和RGB视频的重演。在*IEEE计算机视觉和特征识别会议的论文集中*，第2387-2395页，2016年。

[Zhang *et al.* , 2018] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang.深度特征作为感知指标的不合理的有效性。在*IEEE计算机视觉和模式识别会议论文集中*，第586-595页，2018。

[Zhu *et al.* , 2020] Hao Zhu, Chaoyou Fu, Qianyi Wu, Wayne Wu, Chen Qian, and Ran He.Aot:基于外观最优传输的身份互换的伪造检测。*神经信息处理系统进展》*，33，2020。

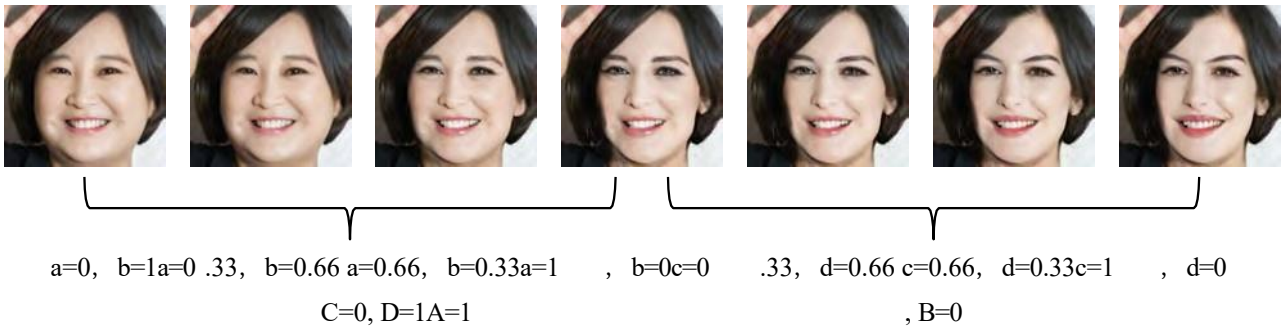


图11：不同成分的SID的内插结果。

## 网络结构

我们的HifiFace的详细结构在图12中给出。对于所有的剩余单元，我们使用Leaky ReLU（LReLU）作为激活函数。重采样是指平均池化或升值，用于改变特征图的大小。编码器中使用了实例归一化（IN）的Res-Blocks，而解码器中使用了自适应实例归一化（AdaIN）的Res-Blocks。

## 更多结果

为了分析三维人脸重建模型的形状信息和人脸识别模型的身份信息的具体影响，我们调整了SID的组成，以产生插值结果。它被表述为：。

$$\psi_{in} = a\psi_s + b\psi_t 。 \quad (12)$$

$$vin = c\upsilon_s + d\upsilon_t 。 \quad (13)$$

其中，$\psi_s$ ，$\psi_t$ 和$\psi_{in}$表示源、目标和内插图像的三维识别系数，$\upsilon_s$ ，$\upsilon_t$ 和$\upsilon_{in}$表示识别模型中源、目标和内插图像的识别向量。

正如我们在图11中看到的第1-4行，我们首先固定$c$=0和$d$=1，脸部形状仍然可以改变，但身份细节的湖泊。然后在第4-7行，我们固定$a$=1和$b$=0，身份变得更加相似。这些结果证明形状信息控制着形状和身份的基础，而身份矢量则有助于识别纹理。

最后，我们从互联网上下载了大量的野生人脸图像，并生成了更多的人脸交换结果，如图13和图14，以证明我们的方法具有强大的能力。更多的结果可以在HTTPS //johann.wang/HifiFace找到。

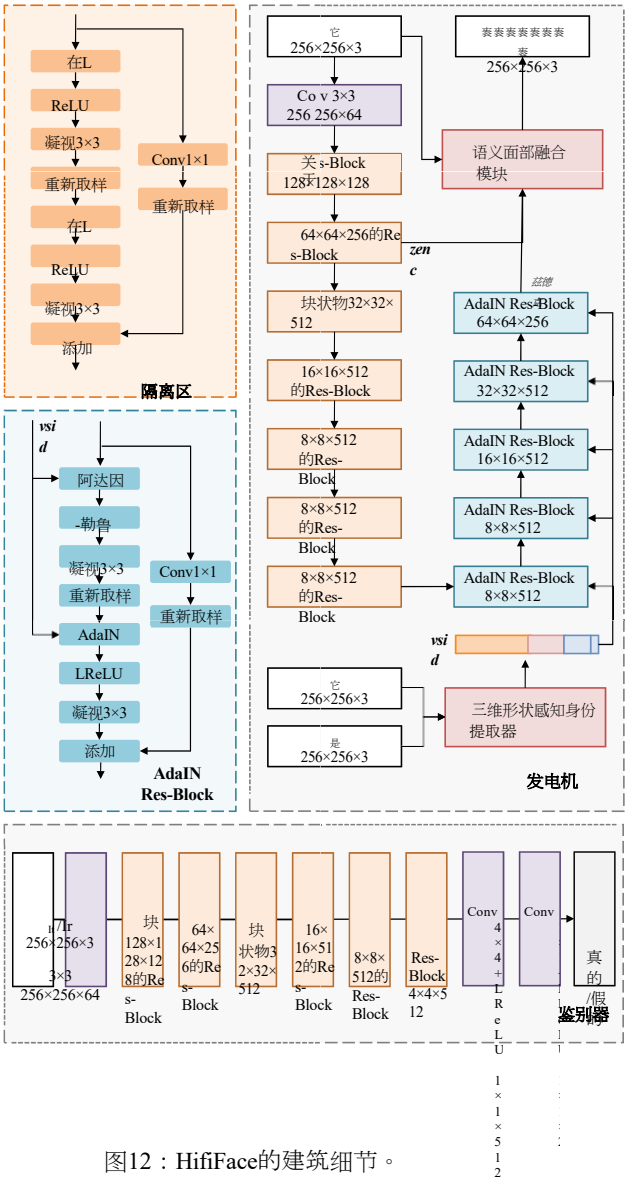


图12：HifiFace的建筑细节。

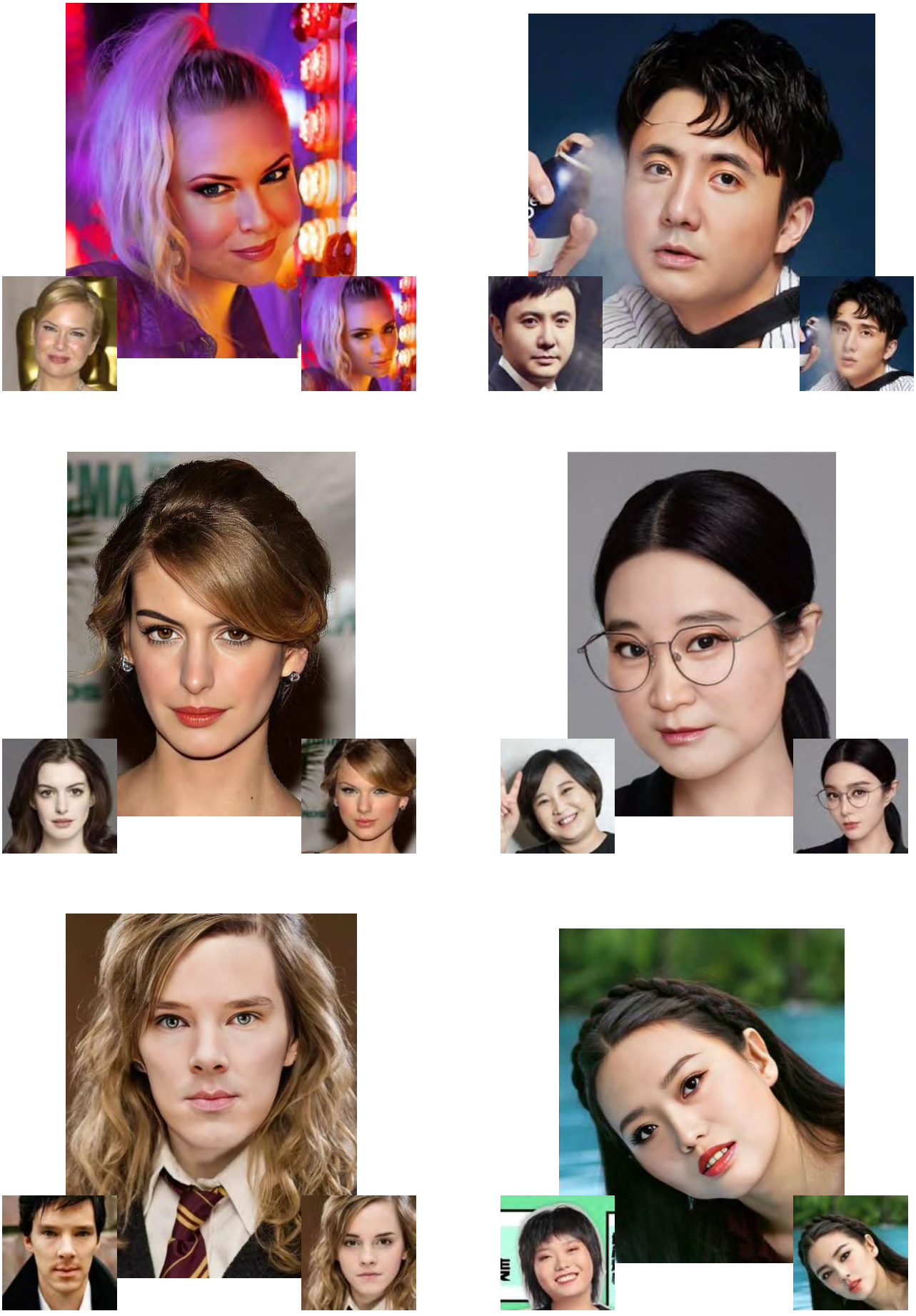

图13：更多关于高分辨率野生人脸的结果。目标图像中的脸被源图像中的脸所取代。

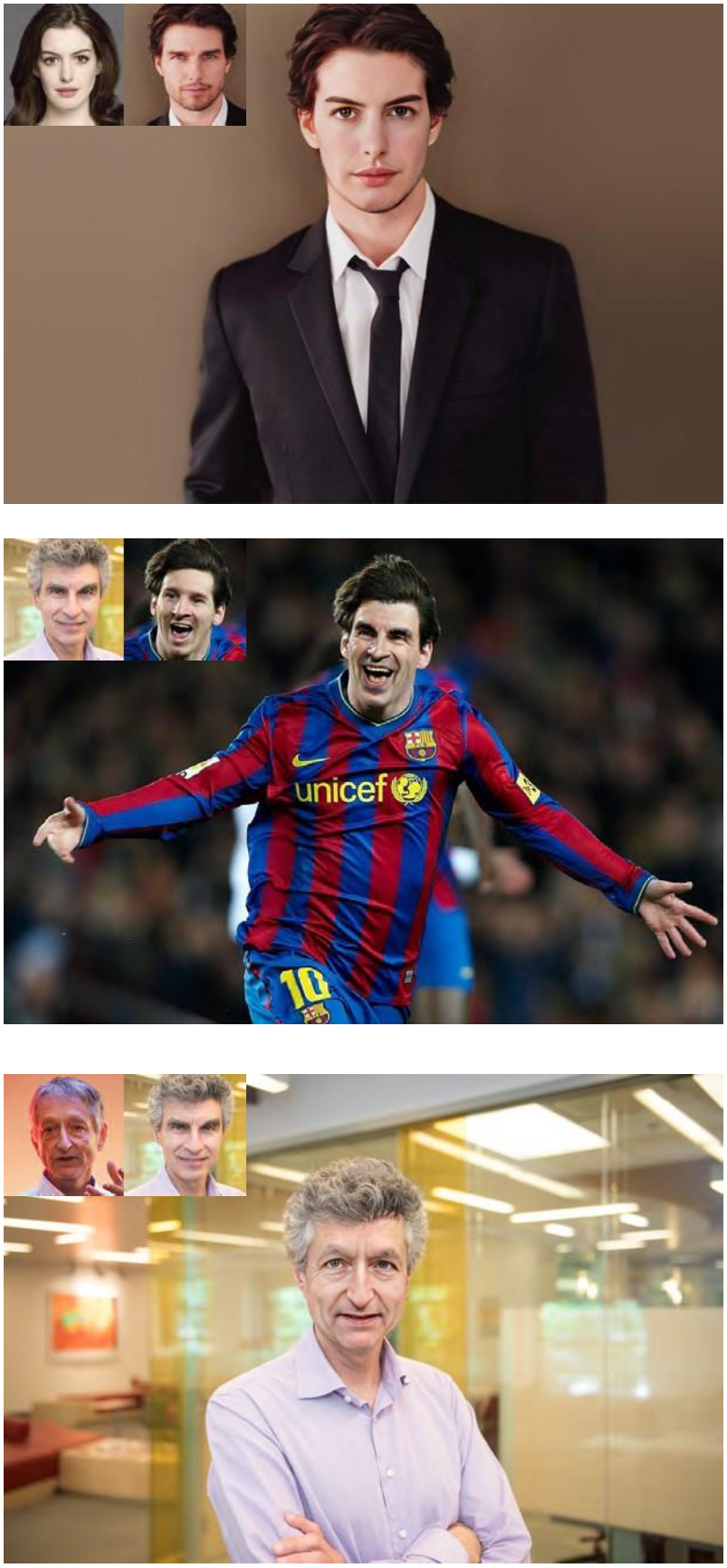


图14：高质量的真实世界照片上的一些结果。目标图像中的人脸被源图像中的人脸所取代。这表明我们的方法所生成的人脸可以非常自然地融入高分辨率拍摄的真实世界场景中。