

基于相对生成对抗网络的低清小脸幻构

邵文泽*^{①②} 张苗苗^① 李海波^①

^①(南京邮电大学通信与信息工程学院 南京 210003)

^②(南京理工大学高维信息智能感知与系统教育部重点实验室 南京 210094)

摘要: 针对当前代表性低清小脸幻构方法存在的视觉真实感弱、网络结构复杂等问题,提出了一种基于相对生成对抗网络的低清小脸幻构方法(tfh-RGAN)。该文方法的网络架构包括幻构生成器和判别器两个部分,通过像素损失函数和相对生成对抗损失函数的联合最小化,实现生成器和判别器的交替迭代训练。其中,幻构生成器结合了残差块、稠密块以及深度可分离卷积算子,保证幻构效果和网络深度的同时降低生成器的参数量;判别器采用图像分类问题中的全卷积网络,通过先后去除批归一化层、添加全连接层,充分挖掘相对生成对抗网络在低清小脸幻构问题上的能力极限。实验结果表明,在不额外显式引入任何人脸结构先验的条件下,该文方法能够以更简练的网络架构输出清晰度更高、真实感更强的幻构人脸。从定量角度看,该文方法的峰值信噪比相比较之前的若干代表性方法可提高0.25~1.51 dB。

关键词: 图像处理; 超分辨率; 人脸幻构; 深度学习; 生成对抗网络

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2021)09-2577-09

DOI: 10.11999/JEIT200362

Tiny Face Hallucination via Relativistic Adversarial Learning

SHAO Wenzhe^{①②} ZHANG Miaomiao^① LI Haibo^①

^①(College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

^②(Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: Considering that previous tiny face hallucination methods either produced visually less pleasant faces or required architecturally more complex networks, this paper advocates a new deep model for tiny face hallucination by borrowing the idea of Relativistic Generative Adversarial Network (tfh-RGAN). Specifically, a hallucination generator and a relativistic discriminator are jointly learned in an alternately iterative training fashion by minimizing the combined pixel loss and relativistic generative adversarial loss. As for the generator, it is mainly structured as concatenation of a few basic modules followed by three $2\times$ up-sampling layers, and each basic module is formulated by coupling the residual blocks, dense blocks, and depthwise separable convolution operators. As such, the generator can be made lightweight while with a considerable depth so as to achieve high quality face hallucination. As for the discriminator, it makes use of VGG128 while removing all its batch normalization layers and embedding a fully connected layer additionally so as to fulfill the capacity limit of relativistic adversarial learning. Experimental results reveal that, the proposed method, though simpler in the network architecture without a need of explicitly imposing any face structural prior, is able to produce better hallucination faces with higher definition and stronger reality. In terms of the quantitative assessment, the peak signal-to-noise ratio of the proposed method can be improved up to 0.25~1.51 dB compared against several previous approaches.

Key words: Image processing; Super-resolution; Face hallucination; Deep learning; Generative adversarial networks

收稿日期: 2020-05-08; 改回日期: 2020-10-18; 网络出版: 2021-08-11

*通信作者: 邵文泽 shaowenze@njupt.edu.cn

基金项目: 国家自然科学基金(61771250, 61972213, 11901299), 中央高校基本科研业务费专项资金(30918014108)

Foundation Items: The Natural National Science Foundation of China (61771250, 61972213, 11901299), The Fundamental Research Funds for the Central Universities (30918014108)

1 引言

图像超分辨率(Super-Resolution, SR)是图像处理领域的一个经典问题。该技术旨在从输入的低分辨率(Low-Resolution, LR)图像中恢复不存在的细节以得到高分辨率(High-Resolution, HR)图像^[1,2]。尤其是对于室外大视域、不受限监控场景,摄像头采集到的远距离人脸影像的辨识度往往较低。如何从长、宽只有十几像素的超低清人脸影像自动重建高清人脸图片,成为国内外亟待解决的一项共同技术难题。

近年来,神经网络缘其强大的学习能力在众多的计算机视觉任务中达到了SoTA (State-of-The-Art)性能,受到了越来越广泛的关注。当前,不管是一般自然图像,还是本文关注的人脸影像,基于深度神经网络的SR方法无疑也取得了令人瞩目的成就。Dong等人^[1]提出的SR深度模型是第1个将卷积神经网络(Convolutional Neural Network, CNN)应用在图像SR领域的模型,记为SRCNN。虽然采用均方误差作为损失函数可以获得较高的峰值信噪比(Peak Signal-to-Noise Ratio, PSNR),但是由此引发的一个问题是,相较真实HR图像而言,SR结果过于平滑、视觉真实感较弱。Goodfellow等人^[3]提出的生成对抗网络(Generative Adversarial Networks, GAN)为图像SR开拓了一个新思路。GAN采用对抗学习的方式交替迭代训练生成器和判别器;生成器以生成能够成功欺骗判别器的图片为目标,判别器以最大程度区分生成器生成的图片与真实图片为目标。Yu等人^[4]提出的人脸SR判别生成网络(URDGN)是首次将GAN应用于人脸SR的深度模型。与原始GAN^[3]类似,URDGN通过生成器和判别器之间的对抗训练来增强SR人脸的真实感。然而,由于GAN本身存在的诸如训练不稳定、梯度消失、模式崩溃等内生缺陷,URDGN生成的幻构人脸往往存在显著可见的局部结构变形等视觉伪效应。

针对GAN的问题,文献^[5]从网络架构、训练方式、损失函数等不同角度提高GAN的训练稳定性,以期克服训练过程的梯度消失和模式崩溃。随着GAN的不断发展,人脸图像SR的质量也在逐步提升。Shao等人^[6]通过深入探索Wasserstein GAN(WGAN)^[7]在人脸图像SR上的潜能,在 16×16 像素到 128×128 像素的低清小脸幻构方法上取得了初步突破。为了充分挖掘小脸图像中的面部轮廓和语义信息,搭建网络时生成器引入了残差块和跳跃连接,并结合梯度惩罚(Gradient Penalty, GP)以及均方误差损失项对网络进行迭代训练。该方法不仅

有效提升了人脸图像SR的真实感,而且对于姿势、表情、光照和遮挡等变化具有较强的适应性。然而,相较原始GAN, WGAN-GP的计算成本大大增加了。同时,上述幻构方法^[6]的网络特征图包含大量通道数,进一步增加了该方法在实际训练和部署时的计算与存储负担。

此外,利用额外人脸结构先验辅助SR网络重建HR人脸成为一种新趋势。例如,Chen等人^[8]提出了一种嵌入面部结构先验的人脸图像SR网络(FSRNet),由粗SR网络、SR编码器、面部结构估计网络和SR解码器4个部分组成。为了生成逼真度高的幻构人脸,作者在FSRNet基础上进一步结合原始GAN^[3]提出FSRGAN。结果表明,基于均方误差损失的FSRNet重建的人脸图像在局部结构上得到大幅度提升,但是FSRGAN在改善SR结果过于平滑的问题时会产生显著可见的虚假纹理等视觉伪效应。

尽管原始GAN^[3]之后的大量变种模型在小脸幻构问题上都具有潜在应用价值的可能,但是Lucic等人^[5]的实证性研究表明,大部分模型未能真正实现令人信服的、优于原始GAN的训练稳定性和泛化性。对此,Jolicoeur-Martineau^[9]基于对原始GAN以及WGAN的深入反思,从训练机理的角度提出了一种能够真正显著提升原始GAN且包容WGAN的相对判别器。众所周知,原始GAN的判别器用于估计输入的真、假样本是真的概率,而生成器训练对应的是真、假样本空间之间的JSD (Jensen-Shannon Divergence)最小化问题。JSD取最大值时,判别器估计真、假样本是真的概率分别为1和0;JSD取最小值时,判别器估计真、假样本是真的概率均为0.5。由此可知,伴随着生成器的迭代训练,不仅期望判别器逐步提高生成的假样本是真的概率,实际上同时期望判别器逐步降低原始的真样本是真的概率,这与训练mini-batch中真、假样本各占一半的先验知识是一致的。虽然原始GAN的判别器易收敛到判定真样本是真的概率为1,但是由于梯度的消失,真样本实际上不再参与判别器的后续迭代。因此,原始GAN的训练只是逐步提高判别器判定假样本为真的概率直至1,而真样本判定为真的概率保持不变。另一方面,如果分别对照原始GAN与WGAN的生成器和判别器损失函数的梯度表达,有趣的发现是:与原始GAN基于JSD不同,WGAN基于IPM (Integral Probability Metrics),而由其决定的训练机理实际内含了判别器判定真样本为真的概率取低这个重要要素。为此,Jolicoeur-Martineau^[9]认为判别器度量的不

应是输入样本绝对为真的概率，而应该是相对为真的概率。基于该思想设计的相对判别器不仅实现了相较于原始GAN更好的训练稳定性和更高的生成质量，而且引入GP后相较WGAN-GP能够以更少的训练样本和时间生成质量更高的数据。

针对当前代表性低清小脸幻构方法存在的视觉真实感弱^[4,8]、网络结构复杂^[6,8]等问题，而深层原因往往归结于支撑这些方法的生成对抗模型本身^[3,7,9]，为此本文提出一种基于相对生成对抗网络的低清小脸幻构方法(tiny face hallucination via Relativistic GAN, tfh-RGAN)。具体地，提出方法的网络架构包括幻构生成器和判别器两个部分，通过像素损失函数和相对生成对抗损失函数的联合最小化，实现生成器和判别器的交替迭代训练。其中，幻构生成器结合了残差块(Residual Block)^[10]、稠密块(Dense Block)^[11]以及深度可分离卷积算子^[12]，保证幻构效果和网络深度的同时降低生成器的参数量；判别器采用了图像分类问题中的全卷积网络，通过先后去除批归一化层(Batch Normalization, BN)，添加全连接层(Fully Connected, FC)，充分挖掘相对生成对抗网络在小脸幻构问题上的能力极限。实验结果表明，在不额外显式引入任何人脸结构先验的条件下，本文方法能够以更简练的网络架构输出清晰度更高、真实感更强的幻构人脸。从定量角度看，本文方法的PSNR相较之前的若干代表方法可提高0.25~1.51 dB。

2 提出方法

图1所示为本文tfh-RGAN的网络架构示意图，上方为生成器，下方为判别器。生成器使用的卷积核尺寸均为3×3像素，而判别器交替使用了卷积核尺寸为3×3和4×4的卷积层，所有卷积层后都紧跟着一个LReLU激活函数。由图可见，tfh-RGAN将尺寸为16×16的小脸图像送入到生成器网络，小脸

图像先经过一个卷积层提取人脸的低层特征，然后经过一系列基本块逐步提取高层特征，跳跃连接结构将提取的低层特征与高层特征拼接融合后送入3个上采样层，每一个上采样层放大2倍，得到空间尺寸为128×128的多通道特征图，特征图经过两个卷积层进而得到3通道的幻构人脸图像。在这之后，幻构人脸图像与高清人脸图像一起作为判别器的输入共同经过一系列卷积层与全连接层，以利用提取的深度特征计算相对损失函数。通过联合最小化由相对生成对抗损失和像素损失构成的能量函数，最终实现生成器和判别器的交替迭代训练。

2.1 生成器

由图1可知，生成器是由若干基本块叠加而成的深度卷积网络，这些基本块主要负责提取高层特征。尽管本文的焦点问题不在于如何巧妙地设计生成器，但是生成器的网络结构对于充分发挥相对生成对抗网络在小脸幻构上的潜在优势还是极为关键的。不失一般性，本文采用Wang等人^[13]提出的RRDB (Residual-in-Residual Dense Block)作为生成器的基本块候选方案。为了实验结果的叙述方便，本文将嵌入RRDB基本块的小脸幻构生成器模型记为tfh-Generator_{RRDB}。为了在保持网络容量和网络深度的同时进一步降低网络参数量，本文将RRDB与MobileNets^[12]的深度可分离卷积算子进行结合，将轻量级的RRDB作为生成器的最终基本块，记为L-RRDB (Lightweight RRDB)，结构示意图如图2所示。其中， β 表示残差缩放参数，在各级残差结构引入 β 有助于抑制深层网络的训练不稳定性。在本文实验中，缩放参数 β 设置为0.2。

由图2可见，L-RRDB将Dense Block的标准卷积分解成了一个1×1卷积和一个深度卷积，1×1卷积作为输入特征图与深度卷积之间的线性连接，先降低输入特征图的通道数再将非线性激活变换的特

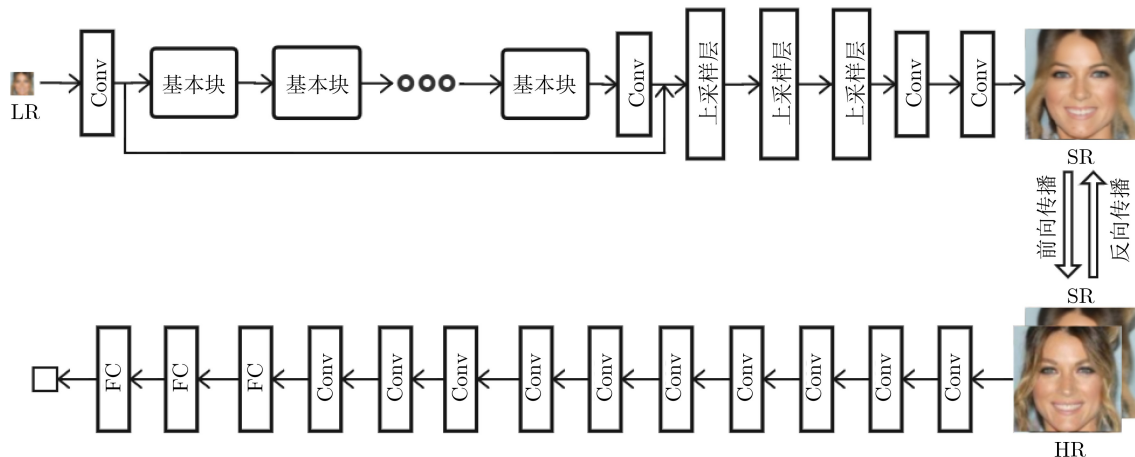


图1 tfh-RGAN的网络架构示意图

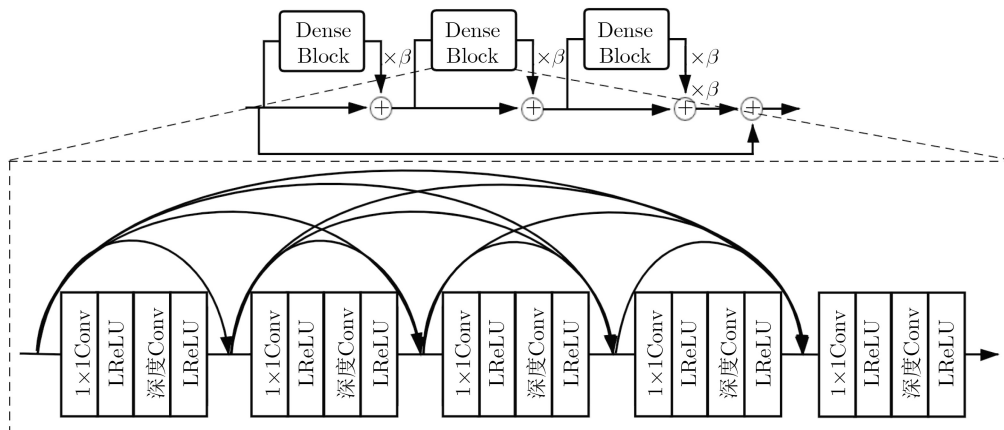


图2 L-RRDB结构示意图

征图传递给深度卷积，深度卷积执行滤波操作提取人脸特征。与原始MobileNets不同，由于L-RRDB的Dense Block输入通道数等于输出通道数，换言之其中卷积的输入通道数大于输出通道数，本文选择将 1×1 卷积放置于深度卷积之前。最终，本文将23个L-RRDB轻量级基本块嵌入如图1所示的生成器。与 $\text{tfh-Generator}_{\text{RRDB}}$ 相类似，本文将嵌入L-RRDB基本块的小脸幻构生成器模型记为 $\text{tfh-Generator}_{\text{L-RRDB}}$ 。注意的是，与稠密网络的Dense Block^[11]不同，RRDB和L-RRDB的稠密块均不包含BN层，仅由卷积和LReLU激活单元组成。原因在于，虽然BN能够加速训练速度，但是可能对生成判别机制带来伪影从而影响视觉效果，这在成像反问题中是一个普遍性实验发现^[6]。要指出的是，本文的模型架构具有较为良好的扩展性，当将L-RRDB升级替换为任意性能更优的基本块时，有望以更小的计算代价、更少的存储负担进一步提升人脸幻构效果。

2.2 判别器

本文采用Simonyan等人^[14]提出的大规模图像识别网络VGG128作为判别器的候选架构。VGG128的组成部分主要包括卷积层、BN层和FC层，卷积层负责逐步提取输入图像的语义特征，BN层加速训练速度，FC层输出整合深度特征以计算相对损失函数。为了充分挖掘相对生成对抗网络在小脸幻构上的能力极限，本文对原始VGG128做了精细化调整。相较于原始VGG128， tfh-RGAN 判别器有两处调整：第一，鉴于去除生成器的BN层可以改善生成器的幻构效果，本文借鉴该技巧，通过去除判别器的BN层以期提升图像真假鉴别性、改善视觉真实感；第二，出于提高判别器估计准确性的目的，同时为了避免特征图由于通道数下降幅度过大而导致细节信息的过度丢失，本文在VGG128的最后添加了一个全连接层。为了叙述方便，本文将生

成器嵌入L-RRDB而判别器利用原始VGG128的 tfh-RGAN 模型，记为 $\text{tfh-RGAN}_{\text{VGG128}}$ 。

2.3 损失函数

给定2.1节的生成器和2.2节的判别器， tfh-RGAN 将联合最小化由像素损失和相对生成对抗损失构成的能量函数，实现生成器和判别器的交替迭代训练。注意到，为了增强图像细节，很多成像反问题如超分辨率的深度网络往往是联合生成对抗损失、特征感知损失以及像素损失进行优化训练的。诚然，更多损失函数的结合有助于改善网络训练，但是同时也会带来额外的训练复杂度。例如，很多方法会分成两步训练：先用像素损失函数预训练网络，然后将生成对抗损失、特征感知损失以及像素损失3项结合去训练网络。此外，更多损失函数的结合涉及更复杂的超参数调节。再者，虽然基于VGG^[14]的感知损失函数能够增强图像细节，但是人脸与自然图像不同，人脸影像是一个含有较多平坦区域的特殊流形。为此，本着不额外增加训练复杂度这个前提，本文期望仅依靠相对生成对抗损失去提升幻构人脸的视觉真实感，以最大可能地探知Jolicoeur-Martineau^[9]的相对判别这一新思想的建模潜能。

当给定mini-batch的真、假样本对 $\{(x_{\text{HR}}, x_{\text{SR}})\}$ 时，原始GAN^[3]的判别器利用提取的深度特征直接返回样本为真的概率 $D(x_{\text{HR}}) = \sigma(C(x_{\text{HR}}))$ ， $D(x_{\text{SR}}) = \sigma(C(x_{\text{SR}}))$ 。其中， $C(\cdot)$ 表示刻画样本真假度的度量函数， $\sigma(\cdot)$ 为sigmoid激活函数， $x_{\text{SR}} = G(x_{\text{LR}})$ ， G 为生成器。相对判别器的不同之处在于，利用提取的深度特征度量了真、假样本相对为真的概率。具体地，式(1)定义了真样本 x_{HR} 相对mini-batch中的假样本 $\{x_{\text{SR}}\}$ 为真的概率：

$$D_{r-f}(x_{\text{HR}}, x_{\text{SR}}) = \sigma(C(x_{\text{HR}}) - E_{\text{SR}}[C(x_{\text{SR}})]) \quad (1)$$

式(2)定义了假样本 x_{SR} 相对mini-batch中的真样本 $\{x_{\text{HR}}\}$ 为真的概率：

$$D_{f-r}(x_{SR}, x_{HR}) = \sigma(C(x_{SR}) - E_{HR}[C(x_{HR})]) \quad (2)$$

其中, E_{HR} , E_{SR} 分别代表定义在 mini-batch 中真、假样本上的期望算子。根据式(1)、式(2), tfh-RGAN 的相对判别损失函数 L_D^R 为

$$L_D^R = -E_{HR}[\lg(D_{r-f}(x_{HR}, x_{SR}))] - E_{SR}[\lg(1 - D_{f-r}(x_{SR}, x_{HR}))] \quad (3)$$

同理, tfh-RGAN 的相对生成损失函数 L_G^R 为

$$L_G^R = -E_{HR}[\lg(1 - D_{r-f}(x_{HR}, x_{SR}))] - E_{SR}[\lg(D_{f-r}(x_{SR}, x_{HR}))] \quad (4)$$

对于像素损失, 本文采用类似均方误差的平均绝对误差作为正则项, 以保证幻构人脸与原始人脸在内容上的相似性, 具体为

$$L_{MAE} = E_{LR,HR}[\|G(x_{LR}) - x_{HR}\|_1] \quad (5)$$

其中, $E_{LR,HR}$ 代表定义在 mini-batch 中低清、高清样本上的期望算子。

至此, 结合式(1)~式(5), 给出本文的总损失函数如下为 $L = \alpha L_D^R + \alpha L_G^R + \gamma L_{MAE}$ 。其中, 第1项为最终的判别器损失函数, 后两项为最终的生成器损失函数, α, γ 为正值调节参数, 分别取为 4.5×10^{-5} 和 0.55。

3 实验结果与分析

本文采用名人属性数据集(Celebrity Face Attributes, CelebA)^[15]训练 tfh-RGAN。CelebA 包含 20 多万张尺寸为 128×128 的人脸图像, 是大约 1 万个名人的集合, 在姿态、表情、光照等不同维度上存在着广泛差异。本文实验只选择 CelebA 的前 2 万张人脸图片用于训练, 2 万张之后的 100 张图片用于验证, 最后的 260 张图片用于测试。为了训练 tfh-RGAN, 如上选择的 CelebA 图片作为原始 HR 图像, 8 倍双三次插值下采样后得到尺寸为 16×16 的 LR 图像。网络训练时, tfh-RGAN 将 HR 与 LR 图像分别输入判别器和生成器, 而测试时只将 LR 图像送入生成器得到幻构的 HR 人脸图像。具体地, 本文使用 $\beta_1 = 0.9, \beta_2 = 0.99$ 的 Adam^[16] 优化器训练网络。其中, 学习率最初设置为 1×10^{-4} , 迭代次数每增加 100000 次, 学习率减少为原来的 1/2, 总迭代次数为 500000。此外, mini-batch 的 LR 和 HR 图像对数目取为 16。本文所有实验均在配置了 NVIDIA GPU (1080 Ti) 的 pyTorch 框架下实现。

3.1 消融实验

本节分别从定性和定量两个角度, 验证最终模型 tfh-RGAN 在构造生成器和判别器过程中引入的不同模型要素的必要性和有效性。具体而言, 下文

重点讨论 tfh-RGAN 与前文 3 个不同版本退化模型的性能差异, 主要包括: 嵌入 RRDB 和 L-RRDB 的小脸幻构生成器模型 tfh-Generator_{RRDB}, tfh-Generator_{L-RRDB} 以及利用原始判别器 VGG128 的小脸幻构相对生成对抗模型 tfh-RGAN_{VGG128}。

3.1.1 定性分析

图3给出了 tfh-RGAN 与 3 个退化版本在 7 张人脸图像上的幻构结果。首先, tfh-Generator_{RRDB} 与 tfh-Generator_{L-RRDB} 相比较而言, 在姿态、表情以及遮挡等多变化情况下, 不论是整体轮廓还是局部结构, 两者所得的幻构结果没有太大差异。因此, tfh-Generator_{L-RRDB} 可实现与 tfh-Generator_{RRDB} 几乎相同的幻构视觉效果。在弱光照条件下 tfh-Generator_{L-RRDB} 的局部区域恢复能力甚至还略优于 tfh-Generator_{RRDB}。例如, 根据第 5 列幻构人脸, tfh-Generator_{L-RRDB} 恢复的嘴形比 tfh-Generator_{RRDB} 相对清晰些。然而, 非常明显的是, tfh-Generator_{RRDB} 与 tfh-Generator_{L-RRDB} 的幻构人脸都过于平滑、视觉真实感较弱。

然而, 遗憾的是, 相较于 tfh-Generator_{L-RRDB} 而言, 引入相对生成对抗学习机制的 tfh-RGAN_{VGG128} 未能有效增强幻构人脸图片的真实感。事实上, 图3显示 tfh-RGAN_{VGG128} 提升更多的是整体轮廓的清晰度, 仍然无法较好地恢复牙齿、胡须等脸部细节信息。整体对比 tfh-RGAN 与 tfh-RGAN_{VGG128} 的视觉效果不难看出, 前者的 SR 结果更加逼近原始 HR 人脸图片、真实感更突出。因此, tfh-RGAN 的恢复能力, 尤其是对纹理等细节的恢复能力更强。例如, 根据第 1, 3, 7 列的幻构人脸, tfh-RGAN 可以重建出更清晰的牙齿缝隙, 面部细小纹路更加明显, 胡须、眉毛也更具真实感。另外, tfh-RGAN 重建的眼睛结构更加具体明确, 尤其是瞳孔部分, 而 tfh-RGAN_{VGG128} 生成的瞳孔只有黑色。因此, tfh-RGAN 与 tfh-RGAN_{VGG128} 的视觉对比充分验证了精细化调整 VGG128 的必要性和有效性。

3.1.2 定量分析

目前存在多种图像重建质量评估方法, PSNR 和结构相似性指数 (Structural SIMilarity index, SSIM) 是最为常用的两种图像重建质量评估方法, 并且值越大表示重建结果越好。基于 PSNR 和 SSIM, 表1为 tfh-RGAN 消融实验的定量分析结果。由表1, tfh-Generator_{L-RRDB} 仅用 tfh-Generator_{RRDB} 1/2 不到的参数量, 却达到了与之几乎等同的 PSNR 和 SSIM。该结果与 3.1.1 小节的定性分析基本一致。因此, 使用深度可分离卷积不仅能够有效地降低网络的参数量, 而且能够很好地保持深度网络的

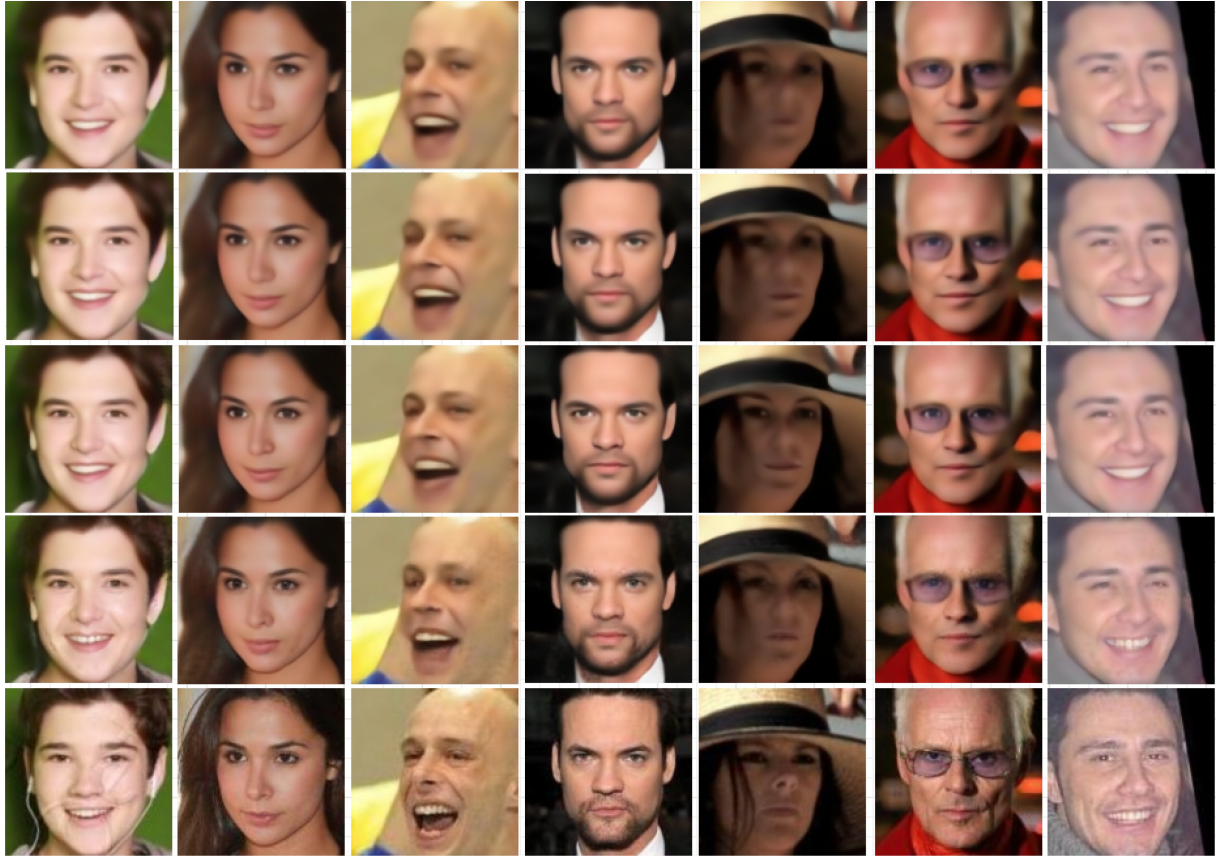


图3 tfh-RGAN消融实验的定性分析。1~4行分别对应tfh-Generator_{RRDB}, tfh-Generator_{L-RRDB}, tfh-RGAN_{VGG128}以及tfh-RGAN的幻构人脸; 第5行为原始HR人脸

表1 tfh-RGAN消融实验的定量分析

模型	PSNR	SSIM	参数量
tfh-Generator _{RRDB}	25.06	0.7313	16734915
tfh-Generator _{L-RRDB}	25.00	0.7295	7138947
tfh-RGAN _{VGG128}	24.89	0.7299	7138947
tfh-RGAN	24.73	0.7172	7138947

性能。同时,注意到, tfh-RGAN_{VGG128}的PSNR相较于tfh-Generator_{L-RRDB}有所下降,尽管tfh-RGAN_{VGG128}相对改善了幻构人脸的整体清晰度。这是由于tfh-Generator_{L-RRDB}的网络训练完全是基于像素损失的,与PSNR的计算模式一致。tfh-RGAN与tfh-RGAN_{VGG128}相比,PSNR和SSIM均有所下降,但下降幅度较小;而由3.1.1小节可知,tfh-RGAN的视觉真实感改善幅度较大。因此,结合定量与定性分析结果,基于相对生成对抗网络的低清小脸幻构方法总体达到了预期效果。

3.2 方法比较

为了验证本文提出方法的有效性和优越性,本小节将tfh-RGAN与当前5种代表性方法进行定性和定量比较,包括URDGN^[4],LCGE^[17],CNN-MNCE^[18],FSRNet^[8],FSRGAN^[8]。按照惯例,所

有人脸幻构方法均使用与tfh-RGAN完全相同的人脸数据集进行训练,以便进行公平的比较。

3.2.1 定性比较

图4所示为tfh-RGAN与代表性方法^[4,8,17,18]在10张人脸图像上的幻构结果。为了从不同角度清晰比较上述方法的幻构效果,选择呈现的10张不同性别的人脸图像具有明显的表情、姿态、光照以及遮挡变化。对于CNN-MNCE^[18]和LCGE^[17],二者均为两步渐进式SR方法。CNN-MNCE^[18]先在基于正则先验的SR框架中插入预训练的CNN图像去噪器实现人脸图像粗估计,然后利用残差补偿方法增强面部细节。LCGE^[17]先将LR人脸划分为5个LR面部分量,在此基础上通过5个并行的CNN模型分别实现不同HR面部分量的重建增强,最后将面部分量组装恢复成HR人脸。由图4仔细比较两者的幻构人脸,可以发现CNN-MNCE在恢复细节清晰度上比LCGE略有优势。但是,不管是LCGE还是CNN-MNCE,幻构人脸图像真实感均较弱,不仅面部光滑,而且当人脸有很大姿态或表情变化时,五官结构产生明显变形。

综合来看,基于生成对抗学习的SR方法尤其是FSRGAN和tfh-RGAN实现了更加真实的人脸幻

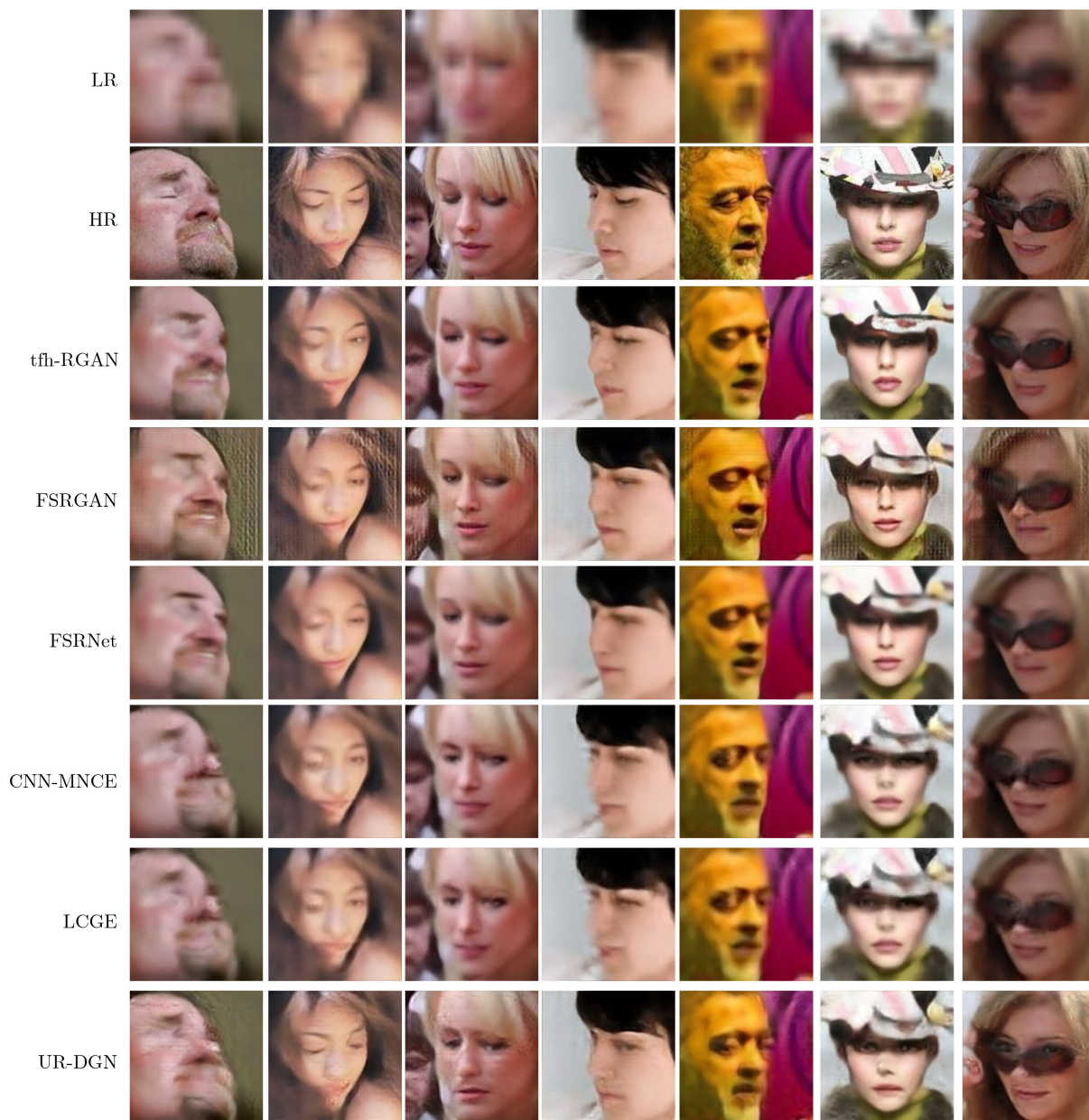


图4 tfh-RGAN与代表性方法的定性比较

构。但是，由于FSRGAN是在FSRNet基础上引入原始GAN训练而来的，伴随着视觉效果改善，原始GAN的内生性问题同时带来了显著可见的虚假纹理等视觉伪效应。不过注意到，相较LCGE, CNN-MNCE, FSRNet由于引入了人脸先验知识的确实实现了更清晰、更准确的局部结构重建效果，尽管依然缺失有助于提升视觉真实感的高频细节。而对于URDGN^[4]，虽然同样能够生成比LCGE和CNN-MNCE整体清晰度更好的人脸，然而在SR结果中不仅可以看到五官变形，而且与FSRGAN类似，同样存在虚假纹理等伪细节。基于上述实验结果比较性观察，本文提出的tfh-RGAN不仅重建出了没有伪效应、整体更清晰的SR人脸图像，而

且在局部区域中恢复的五官结构更为准确、高频细节更加丰富，在所有比较方法中幻构效果是最好的。为了清楚起见，图5展示了3张图片的局部放大区域，以充分验证tfh-RGAN相较其他方法^[4,8,17,18]的视觉效果优势。

3.2.2 定量比较

表2给出了tfh-RGAN与其他方法^[4,17,18,8]在260张图片上的平均PSNR和SSIM。出乎意料的是，尽管tfh-RGAN基于生成对抗学习机制，但依然取得了最优的PSNR (24.73 dB)和SSIM (0.7172)。而URDGN尤其是FSRGAN虽然整体上提升了图像的视觉感知清晰度，相较FSRNet^[8]和CNN-MNCE^[18]两个度量指标的取值却明显下降。

图5 tfh-RGAN与代表性方法^[4,17,18,8]在局部放大区域的定性比较

表2 tfh-RGAN与当前方法的定量比较

模型	URDGN ^[4]	LCGE ^[17]	CNN-MNCE ^[18]	FSRNet ^[8]	FSRGAN ^[8]	tfh-RGAN
PSNR (dB)	23.55	23.55	24.34	24.48	23.22	24.73
SSIM	0.6696	0.6673	0.6883	0.7133	0.6498	0.7172

注意到, 虽然URDGN引入了虚假噪声和可能的五官变形, 但是其PSNR指标明显优于FSRGAN, LCGE。一方面表明, FSRGAN的确引入了过多显著可见的虚假纹理等伪效应; 另一方面, 提升幻构人脸图像的视觉真实感对于图像质量评价的确也是必要的。事实上, 比较方法中只有CNN-MNCE, FSRNet和tfh-RGAN 3种方法的PSNR超过了24.00 dB, 且只有后两者的SSIM结果超过了0.7000。但是, 由上文可知, CNN-MNCE是两阶段逐步实现的方法, 而FSRNet额外引入了人脸结构先验知识。因此, tfh-RGAN是以更简练的网络架构、更明确的模型机理方式实现了清晰度更高、真实感更强的人脸幻构, 且总体上取得了定量评价和定性评价的统一。

4 结论

本文提出一种基于相对生成对抗网络的低清小脸幻构方法tfh-RGAN。利用嵌入轻量级RRDB基本块的幻构生成器和精细调参的VGG128判别器, tfh-RGAN通过最小化像素损失和相对生成对抗损失组成的能量函数, 实现了生成器和判别器的交替迭代训练。实验结果表明, tfh-RGAN能够以更简练的网络架构输出清晰度更高、真实感更强的幻构人脸, 实现了定量评价和定性评价的统一。很大程度上, tfh-RGAN成功挖掘了相对生成对抗学习在低清小脸幻构问题上的能力极限。

参考文献

- [1] DONG Chao, LOY C C, HE Kaiming, *et al.* Learning a deep convolutional network for image super-resolution[C]. Proceedings of the 13th European Conference on Computer Vision, Zurich, 2014: 184–199. doi: [10.1007/978-3-319-10593-2_13](https://doi.org/10.1007/978-3-319-10593-2_13).
- [2] 赵小强, 宋昭漾. 多级跳线连接的深度残差网络超分辨率重建[J]. 电子与信息学报, 2019, 41(10): 2501–2508. doi: [10.11999/JEIT190036](https://doi.org/10.11999/JEIT190036).
ZHAO Xiaoqiang and SONG Zhaoyang. Super-resolution reconstruction of deep residual network with multi-level skip connections[J]. *Journal of Electronics & Information Technology*, 2019, 41(10): 2501–2508. doi: [10.11999/JEIT190036](https://doi.org/10.11999/JEIT190036).
- [3] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, *et al.* Generative adversarial nets[C]. Proceedings of the 27th International Conference on Neural Information Processing Systems, Cambridge, UK, 2014: 2672–2680.
- [4] YU Xin and PORIKLI F. Ultra-resolving face images by discriminative generative networks[C]. Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 318–333. doi: [10.1007/978-3-319-46454-1_20](https://doi.org/10.1007/978-3-319-46454-1_20).
- [5] LUCIC M, KURACH K, MICHALSKI M, *et al.* Are GANs created equal? a large-scale study[EB/OL]. <https://arxiv.org/abs/1711.10337>, 2018.
- [6] SHAO Wenze, XU Jingjing, CHEN Long, *et al.* On potentials of regularized Wasserstein generative adversarial

- networks for realistic hallucination of tiny faces[J]. *Neurocomputing*, 2019, 364: 1–15. doi: [10.1016/j.neucom.2019.07.046](https://doi.org/10.1016/j.neucom.2019.07.046).
- [7] GULRAJANI I, AHMED F, ARJOVSKY M, *et al.* Improved training of Wasserstein GANs[EB/OL]. <https://arxiv.org/abs/1704.00028>, 2017.
- [8] CHEN Yu, TAI Ying, LIU Xiaoming, *et al.* FSRNet: End-to-end learning face super-resolution with facial priors[C]. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 2492–2501. doi: [10.1109/CVPR.2018.00264](https://doi.org/10.1109/CVPR.2018.00264).
- [9] JOLICOEUR-MARTINEAU A. The relativistic discriminator: A key element missing from standard GAN[EB/OL]. <https://arxiv.org/abs/1807.00734>, 2018.
- [10] HE Kaiming, ZHANG Xiangyu, and REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [11] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, *et al.* Densely connected convolutional networks[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2261–2269. doi: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [12] HOWARD A G, ZHU Menglong, CHEN Bo, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. <http://arxiv.org/abs/1704.04861>, 2017.
- [13] WANG Xintao, YU Ke, WU Shixiang, *et al.* ESRGAN: Enhanced super-resolution generative adversarial networks[C]. Proceedings of the 2018 European Conference on Computer Vision, Munich, Germany, 2018: 63–79. doi: [10.1007/978-3-030-11021-5_5](https://doi.org/10.1007/978-3-030-11021-5_5).
- [14] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. Proceedings of the 3rd International Conference on Learning Representations, San Diego, USA, 2015: 1–14.
- [15] LIU Ziwei, LUO Ping, WANG Xiaogang, *et al.* Deep learning face attributes in the wild[C]. Proceedings of 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 3730–3738. doi: [10.1109/ICCV.2015.425](https://doi.org/10.1109/ICCV.2015.425).
- [16] KINGMA D P and BA J. Adam: A method for stochastic optimization[EB/OL]. <https://arxiv.org/abs/1412.6980>, 2017.
- [17] SONG Yibing, ZHANG Jiawei, HE Shengfeng, *et al.* Learning to hallucinate face images via component generation and enhancement[C]. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017: 4537–4543.
- [18] JIANG Junjun, HU Yi, HU Jinhui, *et al.* Deep CNN denoiser and multi-layer neighbor component embedding for face hallucination[C]. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018: 771–778.
- 邵文泽：男，1981年生，博士，副教授，研究方向为变分方法、计算统计、表示学习及其成像与视觉应用。
- 张苗苗：女，1993年生，硕士生，研究方向为深度学习与人脸图像超分辨。
- 李海波：男，1965年生，博士，教授，研究方向为下一代智能视觉传感器网络和社交信号处理。

责任编辑：陈倩