Vol.25 No.12 Dec. 2022

文章编号: 2096-1472(2022)-12-26-04

DOI:10.19644/j.cnki.issn2096-1472.2022.012.006

基于真实图像退化的人脸超分辨率重建算法

李洁沁,谢丁峰

(湖南工业职业技术学院信息工程学院, 湖南 长沙 410000) ☑Lijieqin4568@163.com; coolboyxie@163.com



摘 要:针对大多数人脸超分算法的预设退化模型与真实图像的退化方式差距大,导致人脸重建的效果不理想的问题,提出了一种针对真实图像退化的人脸超分辨率重建算法。首先设计了一种混合退化模型,通过对运动模糊、高斯噪声等多种退化形式进行合成用于模拟真实图像退化空间,生成接近现实场景的低分辨率图像。然后采用基于小波域的超分辨重建网络预测得到高分辨率图像的小波系数,并经过小波逆变换得到超分辨率图像。在高清人脸数据集(FFHQ)和真实人脸数据集(RealSR)上的实验结果表明,研究提出的算法不仅能有效提升重建效果,而且适用于真实场景下的人脸超分辨率重建。

关键词:退化模型,人脸图像,超分辨率重建,小波变换,卷积神经网络中图分类号:TP391 文献标识码:A



LI Jieqin, XIE Dingfeng

(School of Information Engineering, Hunan Industry Polytechnic, Changsha 410000, China)

⊠Lijieqin4568@163.com; coolboyxie@163.com

Abstract: A large gap between the preset degradation model of most face hyper-resolution algorithms and the degradation of real images leads to the poor effect of face reconstruction. In view of this problem, this paper proposes a face super-resolution reconstruction algorithm for real image degradation. Firstly, a hybrid degradation model is designed, which combines motion blur, Gaussian noise and other degradation forms to simulate the real image degradation space and generate low-resolution images close to the real scene. Then, the wavelet coefficients of the high-resolution image are predicted by the super-resolution reconstruction network based on wavelet domain, and the super-resolution image is obtained by inverse wavelet transform. Experimental results on FFHQ (Flickr-Faces-HQ) and RealSR datasets show that the proposed algorithm not only effectively improves the reconstruction effect, but also is suitable for face super-resolution reconstruction in real scenes.

Keywords: degradation model; face image; super-resolution reconstruction; wavelet transform; convolution neural network

1 引言(Introduction)

人脸超分辨率(Face Super-Resolution, Face SR),也称为面部重建,是指将低分辨率(Low-Resolution, LR)人脸图像重建成相应的高分辨率(High-Resolution, HR)人脸图像^[1]。与自然图像不同,人脸图像具有独特而复杂的结构,并且面部的纹理、五官存在差别,直接将超分辨率重建技术

应用到人脸图像中无法得获得理想的效果。目前,大部分的算法通常采用插值降采样^[2-4]得到模型训练所需的低分辨率图像,但其退化模型与真实图像的退化方式差异较大,限制了模型的性能,使得人脸图像重建后的效果不理想。文献[5]采用增强生成对抗网络的判别学习能力实现人脸重建,使人脸细节重建效果得到较大的提升,但重建后的图像容易出现虚

假信息。文献[6]通过相机采集真实HR、LR图像对,使用拉普拉斯金字塔预测网络进行超分辨率重建,但该算法需要先对LR图像配准且采集数据非常烦琐。

针对以上问题,本文提出一种针对真实图像退化的人脸超分辨率重建算法。如图1所示,首先设计了一种新的混合退化模型,模拟真实图像的多种退化方式从而生成LR图像,同时对原始HR图像进行小波分解,得到多个高频分量和一个低频分量。然后将得到的LR图像与小波分解后的低频分量作为输入送入基于小波域的超分辨率网络中,经过特征提取、小波预测,最后通过小波逆变换得到SR图像。

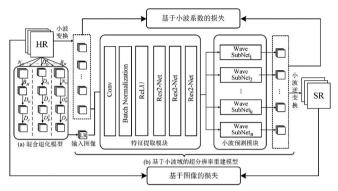


图1 针对真实图像退化的人脸超分辨率重建模型示意图 Fig.1 The diagram of real image degradation—based face super—resolution reconstruction model

2 图像退化(Image degradation)

一般情况下,图像超分辨率重建在生成LR图像过程中往往采用的是单一的退化模型^[7-9],但在实际生活中。图像退化包含多种因素,单一的退化模型不足以反映复杂情况。因此本文提出一种混合退化模型,如图1中(a)所示,通过对真实场景下常见多种退化操作进行随机组合模拟现实场景下图像退化的随机性和复杂性,使得到的低分辨率图像的真实性更高,从而提高图像超分辨率重建的效果。

2.1 运动模糊

运动模糊是相机在曝光时被拍摄物体和相机间发生相对移动,使得不同位置的像素发生叠加,从而导致成像质量下降^[10],是最常见的一种图像退化方式。现实场景中,相机的抖动、物体的非静止状态都会造成图像出现不同程度的模糊情况,模糊空间定义如下^[11]:

$$B_M = K \otimes S + N \tag{1}$$

式(1)中, B_M 和S分别为模糊图像和清晰图像,K为模糊核矩阵, \otimes 代表卷积操作,N为加性噪声,通过调整模糊核的大小,可以得到不同模糊程度的图像。

传统图像退化通常对HR图像进行模糊再下采样,但在现实场景下,真实LR图像可能是模糊的,因此本文对HR和LR两种图像都进行模糊操作,模糊核大小从 $\{5\times5,7\times7,9\times9\}$ 均匀采样,并对模糊核从 $\{0,\pi\}$ 均匀旋转角度,通过两者结合扩展模糊的退化空间。

2.2 噪声

图像在获取和传输过程中会受到各种噪声的干扰,通常是由不良照明或高温引起的传感器噪声。因此,除了广泛采用的高斯噪声,本文还引入了JPEG压缩噪声。

高斯噪声采用具有协方差矩阵 Σ 的三维零均值高斯噪声模型 $(N(0,\Sigma))$ 。当 $\Sigma=\sigma^2I(I)$ 为单位矩阵)时,该模型变成与信道无关的加性高斯白噪声;而当 $\Sigma=\sigma^2I(I)$ 为元素都为1的 3×3 矩阵)时,则变成灰度加性高斯噪声。通过设置概率 $P=\{0.4,0.2,0.2\}$,选择三种不同状态下的高斯噪声。

JPEG数字压缩是目前广泛采用的图像压缩标准,质量因子决定了压缩程度,范围为[0,100],质量因子越小,压缩程度越高,容易出现方块效应^[12]。因此,将压缩因子设置为[30,95],并从中均匀选择。

2.3 下采样

传统的下采样方法中,邻插值简单且容易实现,但生成的图像带有锯齿边缘和马赛克现象。为了使下采样得到的图像更平滑,本文选用双线性插值,同时使用双三次插值操作保留一部分图像的高频部分。此外,采用一种Down-up下采样方式,即先用比例因子s/a对图像进行下采样,然后用比例因子a进行上采样的下采样方法,计算方式如下:

$$D_{\text{down-up}}^{s} = D_{\text{down}}^{s/a} D_{\text{up}}^{a} \tag{2}$$

其中。 D_{total}^{a} 代表比例因子为s/a的下采样, D_{up}^{a} 代表比例因子为a一采样,两种采样的插值方法从双线性插值和双三次插值中随机选择,a 从[1/2,s] 中选择。

考虑到现实场景下真实的低分辨率图像包含复杂的情况,它可能是高分辨图像在不同顺序下的噪声、模糊、下采样版本,也可以是包含其中一种或者几种。为了模拟真实图像的退化空间,本文设计了一种新的混合退化模型,由现实场景中常见的各项退化操作组成退化序列:

$$\left\{B_{M}, N_{G}, D_{JPEG}, D_{B}^{2}, D_{D}^{3}, D_{\text{down-up}}^{s/a}\right\} \tag{3}$$

其中, B_M 、 N_G 、 D_{JPEG} 、 D_B^2 、 D_D^3 、 $D_{down-up}^{s/a}$ 分别代表运动模糊、高斯噪声及JPEG数字压缩、与双线性与双三次插值、Down-up下采样。

对退化序列中的退化操作进行随机选择和组合,任意一种退化模型或者其他传统退化模型是本文退化模型特例。此外,噪声特性随着模糊核尺寸的变化而发生改变,从而扩展退化空间,对各种退化操作中的参数进行均匀采样实现了数据扩充。图2给出了退化模型示意图,通过调整不同的退化操作和退化参数将HR图像退化成LR图像。

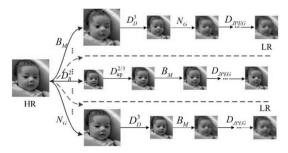


图2 退化模型示意图

Fig.2 The diagram of the degradation model

3 基于小波域的超分辨率重建模型(Superresolution reconstruction model based on wavelet domain)

基于小波域的超分辨率网络结构如图1中(b)所示,首先对原始HR图像进行小波变换,将得到的低频小波图像与退化后的LR图像结合,输入到特征提取网络。然后通过小波预测网络预测得到与特征提取网络输入同样数量的小波系数,最后由小波逆变换得到最终的SR图像。

3.1 小波分解

由于图像的高频小波系数会随其模糊程度的增加而逐渐消失,因此获得清晰的SR图像的关键在于还原低频信息的同时,也能够还原高频小波系数^[13]。为了重建图像的高频信息,本文将小波变换和卷积神经网络结合,从而获取更好的SR图像。采用Haar^[14]变换对图像进行处理,通过高通滤波和低通滤波对图像进行迭代分解,计算不同分辨率下的小波系数。二维Haar变换可以看成分别对图像的所有行和所有列进行一维Haar分解,得到四个细节系数(A、H、V、D),分别代表低频信息、水平高频信息、垂直高频信息及对角高频系数。人脸图像经过二维离散小波变换及小波系数经过二维离散逆变换后的结果如图3所示。

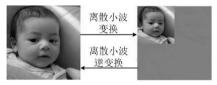


图3二维离散小波变换结果

Fig.3 The results of 2D discrete wavelet transform

3.2 特征提取网络

特征提取网络将 $3\times h\times w$ 大小的低分辨率图像作为输入,首先通过一个卷积层提取特征,卷积核大义为 3×3 , Stride为1,Pad为1。然后输入到三个堆叠的Res2—Net^[15]模块中,Res2—Net结构在减少网络参数的同时,能够很好地提取图像的多尺度特征,提升模型的性能。通过特征提取网络最终得到一个大小为 $N_e\times h\times w$ 的特征,其中 N_e 为最后一层卷积的通道数,本文设置为1,024,h为输入图像高度值,w为输入图像宽度值。

3.3 小波预测网络

小波预测网中包含 N_w 个独立的小波预测子模块,每个子模块中卷积核大小为 3×3 , Stride 为 1, Pad 为 1, 接收特征提取网络输出的特征作为输入,生成与LR大小相同的 N_w 组小波系数。此外,为了使网络具有可扩展性,每个子模块高度独立,不与其他子模块进行信息传递。预测网络中子模块的数目不同,对应实现不同超分尺度的放大效果 (N_w =4n, n为小波包分解的级别,超分因子 r=2n)。

由于小波分解系数的过程可以通过二维离散小波逆变换反向推导,因此对小波预测网络得到的 N_w 组小波系数进行二维离散小波逆变换,最终生成与原始HR图像大小 $[3\times(r\times h)\times(r\times w)]$ 相等的重建图像。

3.4 损失函数

人脸超分辨率重建任务中,通常采用像素的均方误差 (Mean Square Error, MSE)作为损失函数。但是,仅仅将 MSE损失最小化很难获取高频纹理信息,最终导致重建后 的图像过于平滑。由于纹理细节能够通过高频小波系数进行 描述,因此本文将超分辨率问题从原始图像空间转化到小波域,采用基于小波的损失改善纹理重建的效果,并与整个图像的MSE损失函数结合,共同约束网络,从而使网络重建后 的图像更清晰。

整幅图像的MSE损失如下:

$$I_{\text{full-image}} = \left\| \hat{Y} - Y \right\|_{F}^{2} \tag{4}$$

式(4)中, Ŷ和Y分别代表预测的HR图像和真实HR图像。

基于小波的损失函数如下:

$$l_{\text{wavelet}} = \left\| \hat{C} - C \right\|_{E}^{2} \tag{5}$$

式(5)中, \hat{C} =(\hat{c}_1 , \hat{c}_2 ,···, \hat{c}_{N_w})和C=(c_1 , c_2 ,···, c_{N_w})分别代表预测小波系数和真实HR图像产生的小波系数。

此外,为防止高频小波系数收敛为零,引入纹理损失函数,其定义如下:

$$l_{\text{texture}} = \sum_{i=1}^{N_w} \max(\alpha \|c_i\|_F^2 + \varepsilon - \|\hat{c}_i\|_F^2, 0)$$
 (6)

式(6)中, λ ,为小波系数的数量,r为超分辨率比例因子,对于n级小波分解, $\lambda = 2n$, $N_w = 4n$ 。 γ_i 为权重参数, $\alpha 与 \varepsilon$ 为松弛值,防止小波系数收敛为0,总的目标函数如下:

$$l_{\text{total}} = \omega_1 l_{\text{full-image}} + \omega_2 l_{\text{wavelet}} + \omega_3 l_{\text{texture}}$$
 (7)

式(7)中, ω_1 、 ω_2 和 ω_3 为权重参数。

▲实验(Experiment)

4 实验设置

本文使用Pytorch实现改进后的小波超分辨率网络,初始学习率设置为0.002, 迭代5,000次,每1,000次学习率减半。使用Adam优化器实现网络参数优化。Epoch和Batch size分别设置为400和64。此外,实验的硬件配置如下:英特尔核心I5-9400F CPU和32 GB RAM, 搭载4块NVIDIA GeForce RTX 2060 GPU。

4.2 实验数据

数据集包含两个部分,训练集从CelebA $^{[16]}$ 中选取,共25万张图像,将其统一为128×128的HR图像,然后通过退化模型退化后得到大小为32×32的LR图像,作为小波网络的输入。测试集为高清人脸数据集(Flickr-Faces-HQ,FFHQ) $^{[17]}$ 和真实人脸数据集(Real-world Super-Resolution,RealSR) $^{[6]}$ 数据集,共24,000 张。

4.3 实验结果

本文使用2—8 倍超分因子对LR图像进行重建,并以峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和人脸特征相似性(Identity Similarity, IS)作为评价指标在FFHQ和RealSR数据集上评估本文提出算法的性能。

为了评估退化模型对人脸重建效果的影响,本文通过三种退化模型对原始HR图像进行退化操作,退化结果如图4(a) 所示,其中Type1为双三次下采样,Type2为高斯噪声+双线性插值+JPEG数字压缩,Type3为本文退化模型。采用基于

小波域的超分辨率网络对三种退化模型的退化图像进行8 倍超分,结果如图4(b)所示,通过对比发现,本文所用退化模型在视觉上能取得更好的重建效果,说明改进退化模型能够提高人脸重建表现。

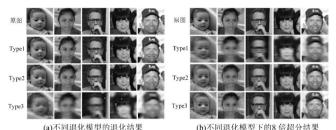


图4 不同退化模型的退化结果和重建效果

different degradation models

Fig.4 Degradation results and reconstruction effects of

图5展示了本文提出的算法和双三次插值算法(Bicubic)¹¹⁸,级联递归卷积神经网络模型(PFSR)¹¹⁹,身份判别卷积神经网络模型(SICNN)^[20],基于小波域的卷积神经网络模型(Wavelet-SRNet)^[21]的8倍超分重建结果,表1中对比了这几种算法在两个测试集上的重建表现。实验结果表明本文提出的算法在不同的超分因子下都取得了最好的表现:(1)当超分因子为2时,各类算法都取得了较好的重建效果,本文提出的算法在两个数据集上获得了最优的重建表现;(2)随着超分因子的增加,Bicubic、PFSR、SICNN的重建表现显著降低,Wavelet-SRNet和本文算法仍然能够保持良好的重建效果,然而在真实低分辨率数据集RealSR上本文的算法重建效果明显优于Wavelet-SRNet。

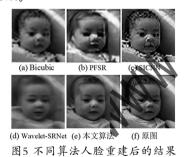


Fig.5 Results of face reconstruction with different algorithms 表 1 不同算法在FFHQ、RealSR数据集上的超分指标对比 Tab.1 Comparison of super-resolution indicators of different algorithms on FFHQ and RealSR datasets

数据集	超分	Bicubic		PFSR		SICNN		Wavelet-SRNet		本文算法	
	因子	PNSR	IS	PNSR	IS	PNSR	IS	PNSR	IS	PNSR	IS
FFHQ	2×	24.86	0.8392	22.57	0.8004	26.91	0.8522	30.06	0.9097	32.67	0.9301
	4×	22.47	0.8096	20.11	0.7665	23.75	0.8306	26.47	0.8814	27.48	0.9188
	8×	20.30	0.7983	17.32	0.7151	20.64	0.8092	20.91	0.8725	23.86	0.9010
RealSR	2×	23.38	0.7845	20.03	0.7587	25.33	0.8296	27.84	0.8827	29.49	0.9065
	4×	22.09	0.7216	18.26	0.6809	23.17	0.8103	25.63	0.8301	28.31	0.8796
	8×	20.27	0.6010	17.52	0.5722	20.89	0.7798	23.59	0.7824	25.64	0.8531

5 结论(Conclusion)

本文提出了一种针对真实图像退化的人脸超分辨率重建算法,该算法设计了一种混合退化模型来模拟真实图像退化空间,用于生成接近现实场景的低分辨率图像,并通过对退化操作中的参数进行均匀采样实现数据扩充。此外,采用基于小波域的超分辨率网络,通过结合图像空域特征和频域特征,并引入Res2-Net模块增强模型多尺度特征提取能力,从而提升人脸图像重建的效果。实验结果表明,与其他算法相比,本文算法在不同超分因子下都取得了最好的重建效果,尤其是在真实低分辨率数据集RealSR上仍然能够取得良好的重建效果,表明本文提出的算法适用于现实场景下的人脸超分辨率重建。

参考文献(References)

- [1] 苏衡,周杰,张志浩.超分辨率图像重建方法综述[J].自动化学报,2013,39(08):1202-1213.
- [2] LI X, ORCHARD M T. New edge-directed interpolation[J]. IEEE Transactions on Image Processing, 2001, 10(10): 1521-1527.
- [3] 孙毅堂,宋慧慧,张开华,等.基于极深卷积神经网络的人脸超分辨和重建算法[]].计算机应用,2018,38(04):1141-1145.
- [4] 徐文博,孙广玲,陆小锋.预训练网络引导的人脸图像超分辨率重建[J].工业控制计算机,2020,33(06):36-38.
- [5] Y.O. X, PORIKLI F. Ultra—resolving face images by discriminative generative networks[C]// European Conference on Computer Vision. Amsterdam: Springer, 2016:318–333.
- [6] CAI J, ZENG H, YONG H, et al. Toward real-world single image super-resolution: A new benchmark and a new model[C]// IEEE International Conference on Computer Vision. Seoul: IEEE, 2019:3086–3095.
- [7] 徐川,端木春江.基于卡通纹理分解和稀疏表示的图像超分辨率重建[]].软件工程,2016,19(5):15-17.
- [8] HARIS M, SHAKHNAROVICH G, UKITA N. Deep back-projection networks for super-resolution[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018:1664–1673.
- [9] ZHANG T, SONG S, JIA Z, et al. Object motion deblurring in single image under static background[J]. IEEE Access, 2020, 8:218069–218080.
- [10] 唐述,万盛道,杨书丽,等.空间尺度信息的运动模糊核估计方 法[[].软件学报,2019,30(12):3876-3891.
- [11] 方睿,周愉,刘鹏,等.基于多尺度循环残差神经网络的图像去运动模糊[]].计算机工程与设计,2022,43(03):786-793.
- [12] KUMAR A, KANSAL A, SINGH K. Anti-forensic approach for JPEG compressed images with enhanced image quality and forensic undetectability[J]. Multimedia Tools and Applications: An International Journal, 2020, 79(3):8061–8084.