

# 基于先验特征与谱归一化的人脸超分辨

万杰林 冷拓\* 倪超杰

(上海大学计算机工程与科学学院 上海 201900)

**摘要** 图像超分辨技术指在不丢失信息的情况下将低分辨率(LR)图像转换成高分辨率(HR)图像。该技术在人像上的实现有着广泛的应用场景如人脸识别、人脸对齐等,但传统的超分辨方法在人脸图像上恢复程度低,并且不稳定。对此,提出SN-FSRGAN模型。使用人脸先验特征指导超分辨率;引入谱归一化用于稳定基于GAN的超分辨率网络训练结果。通过在数据集Helen与CelebA上实验显示,所提出的方法在PSNR、SSIM与视觉感官上皆取得了对比ESRGAN、FSRGAN等模型而言较优的结果。

**关键词** 人脸超分辨 生成对抗网络 人脸先验特征 谱归一化

中图分类号 TP3

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.07.030

## FACE SUPER-RESOLUTION BASED ON PRIOR FEATURES AND SPECTRAL NORMALIZATION

Wan Jielin Leng Tuo\* Ni Chaojie

(College of Computer Engineering and Science, Shanghai University, Shanghai 201900, China)

**Abstract** The purpose of image super-resolution technology is to convert low-resolution (LR) images into high-resolution (HR) images without losing information. The realization of this technology on portraits has a wide range of application scenarios such as face recognition, face alignment, etc., but the traditional super-resolution method has a low degree of recovery on face images and is unstable. In this regard, we propose a SN-FSRGAN model. Face prior features were used to guide super-resolution; and spectral normalization was introduced to stabilize GAN-based super-resolution network training results. Experiments on the Helen and CelebA datasets show that the proposed method has achieved better results in terms of PSNR, SSIM and visual senses compared with models such as ESRGAN and FSRGAN.

**Keywords** Face super-resolution GAN Face prior features Spectral normalization

## 0 引言

单幅图片超分辨(Single Image Super-Resolution)的问题<sup>[8]</sup>旨在不丢失信息的情况下将一幅低分辨率(LR)图像恢复成高分辨率(HR)。从本质上来说,是在寻找一个较好的从低维到高维的映射。人脸超分辨作为超分辨问题中特殊的分支,所研究的目标是恢复低分辨下的人脸图像。相较前者,人脸超分辨中可以利用到的信息更多,因此人脸超分辨在极低分辨率( $16 \times 16$ 、 $32 \times 32$ )下也有实用价值。而在这种情况下

恢复的人脸往往跟实际的人脸存在一定偏差,因此人脸超分辨也称为人脸幻觉。目前,人脸超分辨常用于辅助其他人脸算法,如在公安刑事侦查案件中,可以通过监控中低分辨率的人脸进行超分辨,来得到高分辨清晰的人脸图像,便于来辅助对罪犯的追踪。

超分辨问题在20世纪60年代由Harris提出,之后提出的方法主要可以划分成三类:基于插值的方法、基于统计学的方法和基于生成对抗网络的方法。

基于插值的方法主要包括有最近邻插值、双线性插值以及双三次插值方法等。但这类方法没有考虑像素点之间的相互关系,最终恢复的HR图像会比较模

糊。目前,基于插值的方法主要作为神经网络中用来提升分辨率的一层。

基于统计学的方法,主要是指通过卷积神经网络<sup>[2]</sup>来表示一个低分辨率到高分辨率的映射。在 2012 年卷积神经网络被用于图像分类问题上以来,CNN 成了图像处理方面的主流。在超分辨问题上,Dong 等<sup>[4-5]</sup>证明 CNN 来代替之前的映射 patch,并提出了使用 3 层卷积的 SRCNN。Kim 等<sup>[13]</sup>提出了使用比 SRCNN 更深的网络 VDSR 来表示更复杂的映射。在 He 等<sup>[6]</sup>提出残差网络中,网络中的浅层特征跳跃连接到深层网络中,可以有效地避免梯度消失。而 Ledig 等<sup>[7]</sup>提出了将残差网络应用到超分辨问题上的 SRResNet。相比之前的工作,Ioffe 等<sup>[10]</sup>从网络结构出发,通过移除 Batch-Normalization 层来提升网络的灵活性,并提出了 EDSR<sup>[11]</sup>。基于统计学的方法需要描述恢复的图像和真实图像之间的距离来作为损失函数,而上述方法所使用的损失函数为均方误差 (Mean Square Error),所有像素点上的权重都一样,这使得网络输出的图像更加的平滑,相比于真实图像缺乏锐度。

基于生成对抗网络的超分辨方法,通过生成对抗网络中生成器 G 来完成图像从低分辨率到高分辨率的工作,以及判别器 D 预测输入图像来自真实样本的概率,并尽可能将生成的样本与真实样本区分开,形成一种对抗的形式,具体算法流程见图 1。Ledig 等<sup>[7]</sup>通过上述思想,提出了 SRGAN。相比于 SRCNN 等方法,SRGAN 的生成器中使用了 VGG 网络中高层的特征来描述恢复图像的感知损失<sup>[19-20]</sup>。而 ESRGAN<sup>[14]</sup>在 SRGAN 的基础上,结合 EDSR 中对 BN 层的讨论,对 BN 层进行了移除。此外,ESRGAN 中还提出一种相对判别器,来判断真实图像与生成图像的相对程度,而不是判断是否为真实图像的概率。

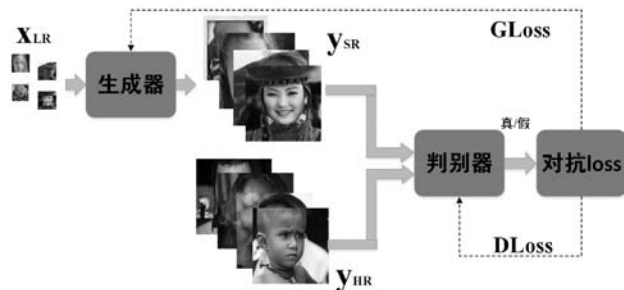


图 1 基于生成对抗网络超分辨算法流程

在 GAN 中,由于通过 JS 散度来描述两个概率分布的距离,导致网络难以训练,以及训练结果不稳定的问题。WGAN<sup>[21-22]</sup>中对这一问题进行了分析,提出用 Wasserstein 距离来替代之前使用的 JS 散度,并证明当

判别器满足 1-Lipschitz 条件,则可以准确地描述两个分布之间的距离。对于 1-Lipschitz 条件 WGAN 中没有给出一个较好的求解方法,SNGAN<sup>[23]</sup>中对卷积网络中的 1-Lipschitz 条件进行了分析,得出了每一层满足 1-Lipschitz 条件的卷积网络也满足 1-Lipschitz 条件的结论,并提出了一种谱归一化方法。

在人脸超分辨中,Yu 等<sup>[16]</sup>提出使用人脸自带的先验知识来辅佐超分辨的工作的网络 FSRNet,将人脸先验特征估计网络的结果添加到超分辨网络中,在融合了人脸特征之后,输出恢复的高分辨率图像。并融合 SRGAN 的思路,提出了 FSRGAN。

本文算法在 SRGAN 的基础上,结合 FSRNet 中利用人脸先验特征的方法,以及 SNGAN 中稳定 GAN 训练结果的方法,提出一种 SN-FSRGAN 模型。在生成器上,首先对 LR 图像进行一个粗糙恢复,然后进行人脸先验特征的估计,再将先验特征融入到网络中进行一个精细的恢复;在判别器上,结合谱归一化方法,来稳定 GAN 的训练结果。通过实验证明,SN-FSRGAN 相比于其他的方法,恢复的 SR 图像质量更好,模型更加稳定,在 PSNR 和 SSIM 上都要优于其他方法。

## 1 结合先验特征的超分辨生成器网络

人脸超分辨相比于一般场景的超分辨方法,可以使用人脸先验特征来辅佐超分辨工作。人脸先验特征主要包括人脸关键点和人脸解析图。人脸解析图包括皮肤、下颚、鼻子、嘴巴、上嘴唇、下嘴唇、左眼、右眼、左眉毛、右眉毛等组件,如图 2 所示。

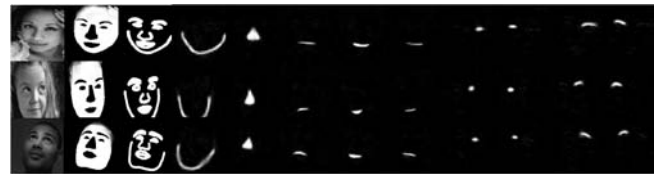


图 2 人脸解析图中所使用的人脸组件

### 1.1 网络结构

在生成器网络结构中,我们参考了 FSRNet 中对先验特征的利用的方法。为了获取先验特征,生成器网络中穿插人脸特征估计网络,整体网络结构如图 3 所示。超分辨生成器网络分为 4 部分,包括粗糙的超分辨网络、精细的编码网络、精细的解码网络和先验特征估计网络。在粗糙的超分辨网络进行一个粗糙的恢复,再由先验特征估计网络对人脸先验特征进行估计,最后将人脸特征融入超分辨网络中,进行精细的恢复。其中,模型所实现的放大倍率为 4 倍。

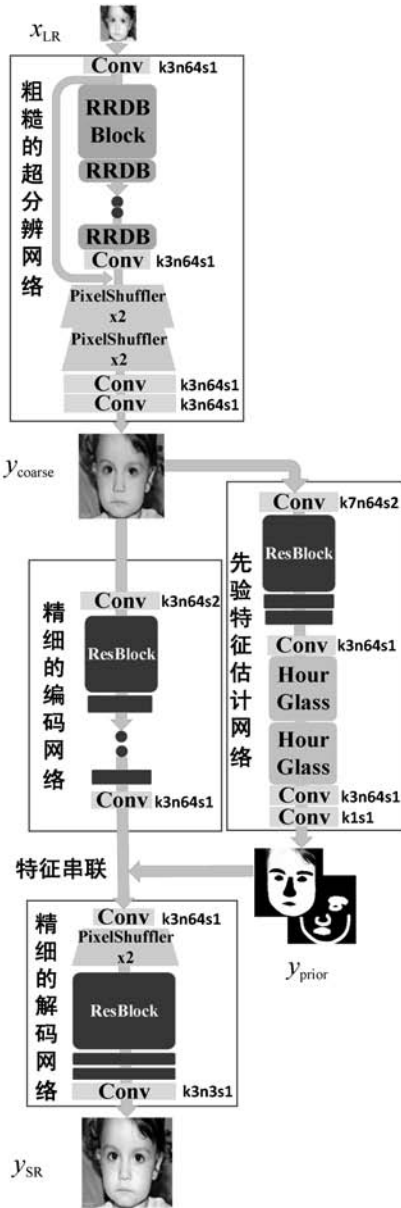


图3 超分辨率生成器网络结构

1) 粗糙的超分辨率网络。在提升超分辨率的过程中, 细节纹理往往是最容易丢失的, 而整体形状则容易被估计出来。通过粗糙的超分辨率网络来对尺度进行一个提升, 并使得整体形状得以保持, 从而估计到较为准确的先验特征。

粗糙的超分辨率网络在结构上, 参考了 ESRGAN<sup>[14]</sup> 中的结构 RRDB, 相比其他超分辨率模型中使用的基本块, RRDB 中不仅移除了 BN 层, 并且在网络中结合了 DenseBlock<sup>[12]</sup> 和 ResBlock<sup>[6]</sup> 的思想, 可以在一定程度上避免梯度消失和过拟合。在网络中, 先经过前面的特征融合之后, 再进行尺度提升, 这可以使得在同等计算资源下网络更深; 将浅层的特征跳跃连接到更深的网络中, 来加快训练速度, 同时还可以避免梯度消失。

$$y_{coarse} = F_{粗糙}(x_{LR}) \quad (1)$$

式中:  $x_{LR}$  是输入的低分辨率图像, 以及得到粗糙恢复

的图像  $y_{coarse}$ 。

在尺度提升的方法上, 模型中采用了亚像素卷积<sup>[26]</sup>。在亚像素卷积层中, 为了在  $n$  个通道上特征尺度提升  $r$  倍, 将通过将通道数量增加  $r \times r$  倍, 之后再每  $r \times r$  个特征进行排列来提升尺度。在 PyTorch 中, 可以通过组合 Conv 层与 PixelShuffle 层来实现:

```
import torch.nn as nn

self.pixel_shuffle = nn.Sequential(
    nn.Conv2d(64, 64 * 4, kernel_size = 3, stride = 1, padding = 1,
    bias = False),
    nn.ReLU(True),
    nn.PixelShuffle(2)
)
```

相比传统的插值方法, 亚像素卷积灵活性更高, 但需要消耗一定的计算资源。

2) 先验特征估计网络。人脸先验特征估计网络将从  $y_{coarse}$  中估计关键点与人脸映射特征, 在实验中, 我们需要使用已经标记好的样本来指导先验特征估计工作。在先验特征估计网络的表达式(2)中,  $y_{prior}$  为所估计的特征。

$$y_{prior} = F_{先验}(y_{coarse}) \quad (2)$$

在先验特征估计网络中使用网络结构中包括几个前导的残差块, 以及之后的两个沙漏 (HourGlass) 结构, 最后通过卷积核为  $1 \times 1$  的卷积层来得到不同的人脸特征图。

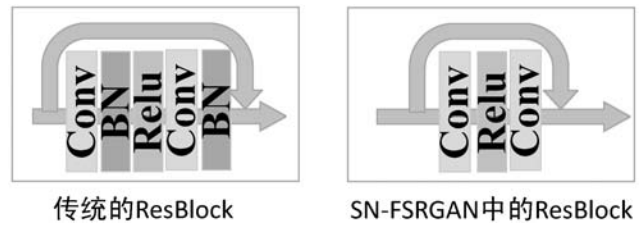


图4 ResBlock 结构

相较于之前的超分辨率算法中所使用的残差块 ResBlock, EDSR 中对于其中的 Batch-Normalization 层进行了详细的分析, 并进行了对 BN 层的移除。本文模型在 ResBlock 上, 同样也移除了 BN 层, 见图4。移除 BN 层带来的好处是可以减少计算成本, 通过扩充了网络的灵活性, 使得到的图像更加切合真实场景。

考虑到计算成本, 在前导的卷积层中, 使用了卷积核为  $7 \times 7$ 、步长为 2 的卷积层来进行下采样。其中, HourGlass 结构常用于估计人脸特征的工作中<sup>[30]</sup>, 在 HourGlass 中前半段进行下采样, 后半段进行上采样, 并且在对称的相同尺度特征之间进行跳跃连接, 可以最大程度地利用不同尺度下的特征。

3) 精细的编码和解码网络。精细的编码网络是为了提取粗糙的人脸图像中的特征信息, 见式(3)。

而精细的解码网络的目的是把人脸图像中的特征信息与先验特征进行融合,来得到更加真实的人脸图像,见式(4)。

$$y_f = F_{\text{编码}}(y_{\text{coarse}}) \quad (3)$$

$$y_{\text{SR}} = F_{\text{解码}}(y_{\text{prior}}, y_f) \quad (4)$$

式中: $y_f, y_{\text{SR}}$ 分别表示精细的编码网络得到的特征和恢复的高分辨率图像。

在精细的编码网络中,首先通过一个卷积核为  $3 \times 3$ ,步长为 2 的卷积层来下采样,之后再通过  $k$  个残差块来提取图像特征。精细的解码网络中,首先将估计的人脸特征与精细的编码网络所提取的信息,串联在一起,在 PyTorch 中可以通过 `cat` 函数将两个相同尺度的特征拼接在一起,来完成这部操作。在通过前导的卷积层之后,再通过亚像素卷积层提升特征的尺度。在通过  $n$  个残差块之后,输出恢复的高精度图像。在实验中,我们参考 Yu 等<sup>[16]</sup> 的设定,所使用的  $k = 13, n = 5$ 。

### 1.2 损失函数

在生成器网络中,损失函数由三部分组成,包括粗糙恢复的图像与真实图像、估计的先验特征与标记好的特征,以及恢复好的 SR 图像与真实图像 HR 之间的距离。

$$l_{\text{pixel}}^{\text{SR}} = \frac{1}{2N} \sum_{i=1}^N \{ |y_{\text{HR}} - y_{\text{SR}}| + |y_{\text{HR}} - y_{\text{coarse}}| + \lambda | \tilde{y}_{\text{prior}} - y_{\text{prior}} | \} \quad (5)$$

式中: $y_{\text{HR}}$ 表示高分辨的目标图像,  $\tilde{y}_{\text{prior}}$ 表示已经标记好的特征。在损失函数中,我们采用了 L1 loss,相比之前的 L2 loss,计算量更小,而且对一些错误的输入具有一定的鲁棒性。在实验中,所使用  $\lambda = 10^{-4}$ 。

## 2 结合谱归一化的判别器网络

在 SRGAN 中,沿用了 GAN 中所使用的判别器,通过判别器 D 来预测图像是否来自真实样本。但局限于 GAN 本身难以训练,难以收敛,在这一部分中我们将谱归一化方法引入到判别器中。

### 2.1 网络结构

关于 GAN 不稳定的问题,在 WGAN<sup>[21-22]</sup> 中已经展开了详细的讨论与分析,提出了只要判别器满足 1-Lipschitz 条件(6),那么就可以准确地描述生成数据的分布与真实数据分布的距离,从而指导生成器生成的数据更逼近真实样本。

$$\frac{\|f(x) - f(x')\|}{\|x - x'\|} \leq 1 \quad (6)$$

SNGAN<sup>[23]</sup> 中对于 1-Lipschitz 条件进行了分析,在卷

积神经网络中,只要每一层都满足 1-Lipschitz 条件,那么整个网络是满足 1-Lipschitz 条件的,而对卷积层,只要它的最大奇异值不超过 1,那么它就满足 1-Lipschitz 条件。SNGAN 中提出约束每一层卷积的奇异值,但由于求大矩阵的奇异值计算量非常大,于是提出一种迭代求奇异值的方法——谱归一化。我们将谱归一化应用到模型的判别器中,判别器中使用的网络结构如图。在 SN-FSRGAN 模型中的判别器参考了 VGG 网络,但我们的网络中使用卷积核为  $3 \times 3$ ,步长为 2 的卷积层替代了池化层,具体网络结构见图 5。

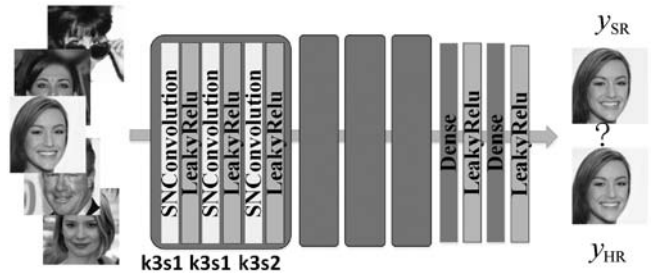


图 5 判别器网络结构

在谱归一化的实现中,可以通过 PyTorch 中关于谱归一化函数的实现,在 Conv 层外嵌套 SpectralNorm 方法,可以便捷地实现卷积层的谱归一化:

```
import torch.nn.utils.spectral_norm as SpectralNorm
SpectralNorm(nn.Conv2d(64, 128, kernel_size = 3, stride = 1, padding = 1))
```

### 2.2 损失函数

SRGAN 中的损失函数由两部分组成,内容损失以及对抗损失。

$$l^{\text{SR}} = l_X^{\text{SR}} + 10^{-1} l_{\text{Gen}}^{\text{SR}} \quad (7)$$

1) 内容损失。SRGAN 中通过 VGG 网络中高层的特征来描述图片的特征,并且使用越高层的特征,细节会恢复得越好。

$$l_{\text{VGG}/i,j}^{\text{SR}} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(y_{\text{HR}})_{x,y} - \phi_{i,j}(y_{\text{SR}})_{x,y})^2 \quad (8)$$

式中: $\phi_{i,j}$ 表示 VGG 网络中第  $i$  个池化层前第  $j$  个卷积层的特征,  $W_{i,j}, H_{i,j}$ 表示该层的行列数。在实验中,我们参照了 SRGAN 使用了第 5 个池化层前第 3 个卷积层的特征。与 ESRGAN 中提出的使用激活层之前的特征作为内容损失相比,我们在模型中采用的是激活层之后的特征。

结合上一节中所使用的损失函数,SN-FSRGAN 所使用的内容损失为:

$$l_X^{\text{SR}} = l_{\text{pixel}}^{\text{SR}} + \lambda_p l_{\text{VGG}/i,j}^{\text{SR}} \quad (9)$$

2) 对抗损失。为了实现 G 与 D 的对抗过程,来得到更加逼真的高分辨率人脸图像,SN-FSRGAN 中使用对抗损失函数为:

$$l_{\text{Gen}}^{\text{SR}} = E[\log(D(y_{\text{HR}}))] + E[\log(1 - D(G(x_{\text{LR}})))] \quad (10)$$

为了减少计算量,我们使用式(11)作为生成器所使用的对抗损失。

$$l_{\text{Gen}}^{\text{SR}} = E[-\log D(G(x_{\text{LR}}))] \quad (11)$$

### 3 实验及分析

#### 3.1 数据集及实验环境

本实验基于 Python 语言,所使用的深度学习框架为 PyTorch,采用 GPU 为 NVIDIA RTX 2060,显存大小为 6 GB。本文分别在 Helen 数据集<sup>[27]</sup>和 CelebA 数据集<sup>[28]</sup>上进行了相关实验。

Helen 数据集包括 2 330 幅人脸 RGB 图片,以及图片中人脸的 194 个关键点的具体坐标以及面部组件解析图。其中,不同的人脸图像,分辨率各不相同。我们使用了其中 1 800 幅人脸作为训练集,530 幅图片作为测试集。我们通过关键点位置对人脸图像进行一个粗糙的裁剪,重组得到  $128 \times 128$  分辨率的高分辨率图片。再通过双三次线性插值得到低分辨  $32 \times 32$  的图片,裁剪过程见图 6。此外,我们对每一幅图片进行 90、180、270 度的旋转,以及水平翻转,同时也对水平翻转后的图进行 90、180、270 度的翻转,将原来的数据集扩充了 7 倍。

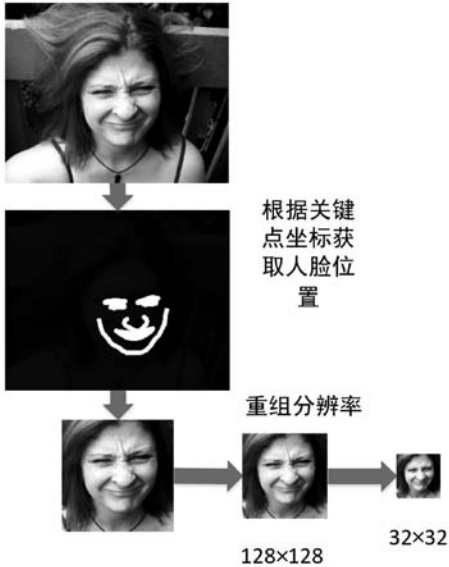


图 6 Helen 数据集裁剪过程

CelebA 数据集包括 30 000 幅人脸图像,分辨率为  $178 \times 218$ ,在实验中我们使用其中 25 000 幅作为训练集,5 000 幅作为测试集。相比 Helen 数据集,CelebA 数据集中不包含先验特征,我们采用了 Lee 等<sup>[29]</sup>中对 CelebA 预处理好的 CelebAMask-HQ 数据集,其中包括了已经裁剪好的 CelebA 人脸图片,以及相应的人脸解析。在对 CelebA 数据集的处理上我们也使用 Helen

数据集一样的翻转处理,但在 CelebA 数据集的实验中,我们仅使用了人脸解析图而没有使用关键点。

#### 3.2 评价标准

在 SISR 问题中,主要使用的评价标准为峰值信噪比<sup>[9]</sup>(PSNR)和结构相似性<sup>[25]</sup>(SSIM),本文也采用这两种标准来对图片进行评价。

1) PSNR。峰值信噪比实质上是以两幅图像的均方误差(式(12))来描述两幅图像的距离,峰值信噪比越大,那么两幅图像的均方误差也就越小,图像失真越小,如式(13)所示。

$$E_{\text{MS}} = \frac{1}{WH} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [y_{\text{HR}}(i,j) - y_{\text{SR}}(i,j)]^2 \quad (12)$$

$$R_{\text{PSN}} = 10 \times \log\left(\frac{(2^n - 1)^2}{M_{\text{MSE}}}\right) \quad (13)$$

往往使用 MSE 作为损失函数的方法,都有着较高的 PSNR 值。在 SRGAN 中指出 PSNR 值高的图片,并不一定有着较好的视觉感受,而选择使用 VGG 高层特征而不是 MSE。因此基于 GAN 的方法 PSNR 值虽然不高,但在视觉感知上却要优于其他的方法。

2) SSIM。结构相似性,从亮度、对比度、结构三个方面来比较两幅图像的相似性。

$$c(X,Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2}$$

$$l(X,Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1}$$

$$s(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}$$

$$S_{\text{SIM}} = l(X,Y) \cdot c(X,Y) \cdot s(X,Y) \quad (14)$$

式中: $\mu_X, \mu_Y$ 表示图像  $X, Y$  的均值; $\sigma_X, \sigma_Y$ 表示  $X, Y$  的方差; $\sigma_{XY}$ 表示  $X, Y$  的协方差; $C_1, C_2, C_3$ 为常数。SSIM 的取值范围为  $[0, 1]$ ,值越大表示图像失真程度越小。

#### 3.3 模型对比分析

将本文的模型与 SRGAN、ESRGAN、FSRGAN 进行比较,为了公平起见,我们在同一个数据集上进行训练。实验结果如表 1 - 表 2 所示。

表 1 在 Helen 数据集上的对比结果

评价指标	SRGAN	ESRGAN	FSRGAN	SN-FSRGAN
PSNR	23.75 3	23.16 6	24.95 4	26.71 1
SSIM	0.669 3	0.646 2	0.719 8	0.817 1

表 2 在 CelebA 数据集上的对比结果

评价指标	SRGAN	ESRGAN	FSRGAN	SN-FSRGAN
PSNR	22.643 2	23.754 6	26.127 4	27.815 9
SSIM	0.601 2	0.657 2	0.776 4	0.853 2

从表 1、表 2 可知,我们的模型相比于 SRResNet 和 ESRGAN,在数值指标有着显著提升。而相比于结合了人脸先验特征的 FSRGAN,SN-FSRGAN 在 PSNR 上要高出 1.6 dB,在 SSIM 上要高 0.075 左右。

我们在 Helen 数据集和 CelebA 数据集中分别选取了两幅图像进行展示,见图 7。在视觉上,SN-FSRGAN 相比 ESRGAN、SRResNet 等一些没有利用人脸先验特征的方法,人脸扭曲程度小,并有更好的视觉感受。相比于 FSRGAN、SN-FSRGAN 中没有穿插 BN 层,这使得我们的结果中几乎没有模糊阴影。此外,得益于带有谱归一化的判别器,训练结果会比 FSRGAN 更加的稳定,生成的图像会更加真实。

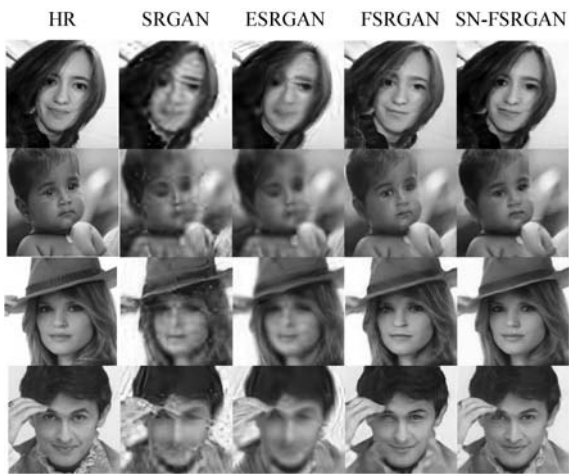


图 7 各算法重建结果对比

### 3.4 对比实验

1) 内容损失替换。相比于 ESRGAN 中采用激活层前的特征作为内容损失,我们保持了使用激活层之后的特征。我们在 Helen 数据集上进行了一组对比试验。在整体上保留了其他参数不变,仅修改了内容损失。相比于 SN-FSRGAN 中使用激活层前的特征作为内容损失,以激活层之后的特征作为内容损失所估计的特征会更加准确,见图 8。而精准的特征融入超分辨网络中可以得到更加细致的人脸,如图中眼角、牙齿等一些局部人脸特征会更加贴合真实样本。



图 8 使用激活层前的特征(上)和使用激活后的特征(下)

2) ResBlock 中 BN 层保留。相比于传统的 ResBlock,我们的 ResBlock 中移除了 Batch-Normalization 层,有

许多有利的证据证明了这一举措可以提高 PSNR,我们同样做了一组对照实验,来验证去除了 ResBlock 的好处。在 CelebA 数据集上进行了一组对比实验,与 SN-FSRGAN 相比,在 ResBlock 基本块中增加了 BN 层,见图 4。其余配置参数进行保留,结果见图 9。带有 BN 层的 ResBlock 不仅在肉眼感官上要明显与 HR 图像存在差异,而且在 PSNR/SSIM 上明显低于 SN-FSRGAN 的结果。

图像对比	PSNR/SSIM
HR	
带BN的 ResBlock	20.4506/0.6603
不带BN的 ResBlock	27.8159/0.8532

图 9 移除 BN 层对比结果

3) 仅使用人脸解析图带来的影响。由于在 CelebA 中缺乏关于人脸特征的标识,我们采用了 CelebAMask-HQ 数据集。在 CelebAMask-HQ 数据集中,没有人脸关键点的标识,这在一定程度上会影响实验的结果。对此,我们对 Helen 数据集增加了一组不使用人脸关键点特征,而仅使用人脸解析图映射的对比实验。由于两组结果在肉眼上,难以对比,我们仅展示两组在 PSNR 和 SSIM 的数值,见表 3。

表 3 Helen 数据集不使用人脸关键点的对比结果

评价指标	仅使用解析图	SN-FSRGAN
PSNR	26.405 3	26.711 0
SSIM	0.807 9	0.817 1

## 4 结 语

图像超分辨作为近些年机器学习领域的热点研究问题,而作为研究分支的人脸超分辨,更是受到研究者的青睐。目前,人脸超分辨常常作为其他人脸算法的前置算法,如人脸识别、人脸对齐。通过本算法,可以得到更加清晰的高分辨人脸,这可以提高其他检测工作的精度,例如公安机关刑事侦查中可通过摄像头中低分辨率的人脸图像就可以完成对罪犯的追踪。

本文基于 SRGAN,结合人脸特征与谱归一化方法,提出一种人脸超分辨模型 SN-FSRGAN,相比于其他方法,可以恢复更加清晰的人脸。在 Helen、CelebA 数据集上,SN-FSRGAN 比之前恢复最好的模型 FSRGAN 在 PSNR/SSIM 上分别提升了 1.6/0.075。但由于网络深度的原因,导致算法计算量太大,以及在网络中没有使用加快训练的方法(BN 层),使训练时间相比

其余方法都要长。接下来的研究目标可以研究如何加快模型训练的速度,以及进一步提升图像重建的效果。

## 参 考 文 献

- [ 1 ] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. *Neural Computation*,1989,1(4):541-551.
- [ 2 ] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks[C]//*Advances in Neural Information Processing Systems*,2012:5211-5221.
- [ 3 ] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//27th International Conference on Neural Information Processing Systems,2014:2672-2680.
- [ 4 ] Dong C, Loy C, He K M, et al. Learning a deep convolutional network for image super-resolution [ C ]//European Conference on Computer Vision,2014:184-199.
- [ 5 ] Dong C, Loy C, He K M, et al. Image super-resolution using deep convolutional networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,2016,38(2):295-307.
- [ 6 ] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[ C ]//IEEE Conference on Computer Vision and Pattern Recognition,2016:770-778.
- [ 7 ] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network [ C ]//IEEE Conference on Computer Vision and Pattern Recognition,2017:105-114.
- [ 8 ] Glasner D, Bagon S, Irani M. Super-resolution from a single image[ C ]//12th IEEE International Conference on Computer Vision,2009:349-356.
- [ 9 ] Yang C Y, Ma C, Yang M H. Single-image super-resolution: A benchmark[ C ]//European Conference on Computer Vision,2014:372-386.
- [ 10 ] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [ C ]//32nd International Conference on International Conference on Machine Learning,2015:448-456.
- [ 11 ] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[ EB ]. arXiv:1707.02921,2017.
- [ 12 ] Huang G, Liu Z, Maaten L V D, et al. Densely connected convolutional networks[ C ]//IEEE Conference on Computer Vision & Pattern Recognition,2017:2261-2269.
- [ 13 ] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[ C ]//IEEE Conference on Computer Vision & Pattern Recognition,2016:1646-1654.
- [ 14 ] Wang X T, Yu K, Wu S X, et al. ESRGAN: Enhanced super-resolution generative adversarial networks[ C ]//European Conference on Computer Vision,2018:63-79.
- [ 15 ] Zhu S Z, Liu S F, Loy C, et al. Deep cascaded bi-network for face hallucination[ C ]//European Conference on Computer Vision,2016:614-630.
- [ 16 ] Yu X, Fernando B, Ghanem B, et al. Face super-resolution guided by facial component heatmaps [ C ]//European Conference on Computer Vision,2018:219-235.
- [ 17 ] Chen Y, Tai Y, Liu X M, et al. FSRNET: End-to-end learning face super-resolution with facial priors [ C ]//IEEE Conference on Computer Vision and Pattern Recognition, 2018:2492-2501.
- [ 18 ] Bulat A, Yang J, Tzimiropoulos G. To learn image super-resolution, use a GAN to learn how to do image degradation first[ C ]//European Conference on Computer Vision,2018:187-202.
- [ 19 ] Johnson J, Alahi A, Li F. Perceptual losses for real-time style transfer and super-resolution [ C ]//European Conference on Computer Vision,2016:694-711.
- [ 20 ] Bruna J, Sprechmann P, LeCun Y. Super-resolution with deep convolutional sufficient statistics [ EB ]. arXiv:1511.05666v1,2015.
- [ 21 ] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[EB]. arXiv:1701.07875,2017.
- [ 22 ] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs[ C ]//31st International Conference on Neural Information Processing Systems,2017:5769-5779.
- [ 23 ] Miyato T, Kataoka T, Koyama M, et al. Spectral normalization for generative adversarial networks[ EB ]. arXiv:1802.05957,2018.
- [ 24 ] He K M, Zhang X Y, Ren S Q, et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification[ C ]//IEEE International Conference on Computer Vision,2015:1026-1034.
- [ 25 ] Wang Z, Bovik A C, Sheikh H R. Image quality assessment: From error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*,2004,13(4):600-612.
- [ 26 ] Shi W Z, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[ C ]//IEEE Conference on Computer Vision and Pattern Recognition,2016:1874-1883.
- [ 27 ] Le V, Brandt J, Lin Z, et al. Interactive facial feature localization[ C ]//European Conference on Computer Vision,2012:679-692.
- [ 28 ] Liu Z W, Luo P, Wang X G, et al. Deep learning face attributes in the wild[ EB ]. arXiv:1411.7766,2015.
- [ 29 ] Lee C H, Liu Z W, Wu L Y, et al. MaskGAN: Towards diverse and interactive facial image manipulation[ EB ]. arXiv:1907.11922v1,2019.
- [ 30 ] Newell A, Yang K Y, Deng J. Stacked hourglass networks for human pose estimation [ C ]//European Conference on Computer Vision,2016:483-499.