

基于深度学习的卡口过车人脸超分辨率重建

万学俊¹, 谢敏怡², 赵周洲¹, 唐 轶³, 蒋 作³

(1. 玉溪市公安局 科技信息化支队, 云南 玉溪 653100; 2. 云南民族大学 电气信息工程学院;
3. 云南民族大学 数学与计算机科学学院, 云南 昆明 650500)

摘要: 针对交通卡口处获取的人脸图像分辨率低进而影响人脸识别的问题, 提出一种基于改进ESRGAN的人脸图像超分辨率重建算法。该算法在RRDB中进一步使用密集连接, 通过促进特征的传递更高效地利用不同层的特征, 加强了深层信息提取模块对人脸特征的提取; 在RDB中引入坐标注意力机制, 并将位置信息嵌入到通道注意力中以加强各通道之间的信息流通, 从而提高了网络对人脸特征的选择能力。利用改进后的算法在公开数据集和自制数据集上进行4倍和8倍的超分辨率重建测试, 并与主流超分辨率算法进行对比, 结果显示改进后的算法获得的人脸图像更加清晰, 能够提供更多的人脸细节特征。在客观评价指标上, 所提算法在4倍和8倍的超分辨率重建后, PSNR和SSIM相较于原算法都有所提升, 且在主流算法中均达到最大值。此外, 该算法在视觉效果上也有着较好表现, 能够更好地表达人脸特征。

关键词: 图像增强; 人脸超分辨率; ESRGAN; RRDB; 密集连接; CA

DOI: 10.11907/rjdk.241512

开放科学(资源服务)标识码(OSID):



中图分类号: TP391.41

文献标识码: A

文章编号: 1672-7800(2025)009-0213-08

Super-Resolution Reconstruction of Faces from Checkpoint Passing Vehicles Based on Deep Learning

WAN Xuejun¹, XIE Minyi², ZHAO Zhouzhou¹, TANG Yi³, JIANG Zuo³

(1. Science & Technology Informatization Squad, Yuxi Public Security Bureau, Yuxi 653100, China;
2. School of Electrical & Information Engineering, Yunnan Minzu University;
3. School of Mathematics & Computer Science, Yunnan Minzu University, Kunming 650500, China)

Abstract: To address the issue of low resolution of face images captured at traffic checkpoints that affects face recognition, a face image super-resolution reconstruction algorithm based on improved ESRGAN is proposed. This algorithm further employs dense connections in the RRDB to more efficiently utilize features from different layers by promoting feature transmission, thus enhancing the ability of the deep information extraction module to extract facial features. Additionally, a coordinate attention mechanism is introduced into the RDB, which strengthens the flow of information between channels by embedding positional information into channel attention, thereby improving the network's ability to select facial features. The improved algorithm is tested for 4x and 8x super-resolution reconstruction on public datasets and self-made datasets, and compared with mainstream super-resolution algorithms. The results show that the face images obtained by the improved algorithm are clearer and can provide more detailed facial features. In terms of objective evaluation metrics, the proposed algorithm achieves higher PSNR and SSIM values after 4x and 8x super-resolution reconstruction compared to the original algorithm, reaching the maximum among mainstream algorithms. Furthermore, the algorithm also performs well visually, better expressing facial features.

Key Words: image enhancement; face super-resolution; ESRGAN; RRDB; dense connections; CA

收稿日期: 2024-07-31

扫描二维码阅读全文:



作者简介: 万学俊(1984-), 男, 玉溪市公安局科技信息化支队民警, 研究方向为图像超分辨重建、大数据、人工智能; 谢敏怡(2000-), 男, 云南民族大学电气信息工程学院硕士研究生, 研究方向为图像超分辨重建; 赵周洲(1987-), 男, 玉溪市公安局科技信息化支队副科长、警务技术三级主管, 研究方向为视频监控系统建设、视频结构化、人像识别算法; 唐轶(1979-), 男, 博士, 云南民族大学数学与计算机科学学院教授、硕士生导师, 研究方向为深度学习; 蒋作(1979-), 男, 博士, 云南民族大学数学与计算机科学学院副教授、硕士生导师, 研究方向为深度学习、软件工程。本文通讯作者: 蒋作。

0 引言

在现代智能交通系统中,交通卡口的人脸图像采集是关键的一环,因为其采集到的人脸图像质量将直接影响到后续的人脸识别。在常见的人脸识别门禁中获取的人脸图像往往都是高清的,而交通卡口所拍摄的人脸大多是在非限制条件且被拍者无感的情况下获取^[1]。

受不同因素影响,卡口过车采集到的人脸图像具有亮度低、雨雾天气造成模糊和分辨率低导致模糊等特点。首先,在夜晚,由于光照强度显著降低,采集到的人脸图像亮度不足,直接影响了人脸识别效果;其次,天气波动如大雨和大雾天气会导致空气中悬浮颗粒增多,进而使拍摄到的人脸图像变得模糊不清,增加了人脸识别难度;再次,一些早期安装的监控设备由于技术限制,其本身性能欠佳,所拍摄的视频分辨率不高,意味着当使用人脸检测算法从这些低分辨率视频中截取人脸图像时,所得图像的分辨率也会相应较低,人脸细节信息可能会因此变得模糊,削弱了人脸识别准确性。特别是当车辆距离监控设备较远时,截取到的人脸在画面中所占比例更小,图像分辨率进一步降低,加剧了人脸细节特征的丢失,对后续人脸识别过程产生了更为不利的影响。

本文针对卡口过车人脸图像分辨率低导致模糊进而影响人脸识别这一情况,将图像超分辨率(Super-Resolution, SR)重建即超分引入到人脸图像增强中。通过人脸超分辨率(Face Super-Resolution, FSR)将低分辨率(Low Resolution, LR)人脸图像恢复为高分辨率(High Resolution, HR)人脸图像,经过处理后的人脸图像可以实现人脸细节增强,将人脸图像中的关键信息有效地展现出来,从而提高后续人脸识别算法准确率,为卡口过车全量信息提供技术支持。

近年来,深度学习的迅猛发展为人脸图像超分辨率提供了新的解决思路。学者们提出了多种基于深度学习的图像超分辨率方法,这些方法可分为有监督的和无监督的。当前,基于深度学习的超分辨率方法大多是有监督的,即训练时使用真实的HR图像和相对应的LR图像。根据网络类型的不同,有监督的超分方法又可分为基于卷积神经网络(Convolutional Neural Network, CNN)的方法和基于生成对抗网络(Generative Adversarial Network, GAN)的方法。

Dong等^[2]较早提出超分辨率卷积神经网络(Super-Resolution Convolutional Neural Network, SRCNN),该方法是根据传统基于稀疏编码的超分方法得来,并首次将CNN应用于超分领域。SRCNN首先通过双3次插值(Bicubic Interpolation)将低分辨率图像尺寸放大至与高分辨率图像一样,然后再用3个卷积层对图像分别进行提取和表达、非线性映射和重建操作。Zhang等^[3]提出残差密集网络

(Residual Dense Network, RDN),它是基于残差密集连接,通过融合多个残差密集模块有效地提取特征信息。Zhu等^[4]提出用于医学磁共振图像超分辨率的残差密集网络,其将浅层卷积特征输入到残差密集模块中以获取特征,且其中的每一层都直接与前一层连接,然后使用子像素卷积层进行上采样和超分辨率重建。大多数超分方法对每个缩放因子分别训练一个模型,从而导致计算效率低下。为了解决该问题,Hu等^[5]提出任意尺度的超分辨率网络Meta-SR,该方法能通过单一模型实现低分辨率图像进行任意缩放因子的超分。Liu等^[6]提出将基于深度学习和随机森林的网络用于人脸超分辨率,首先对低分辨率人脸图像进行上采样得到中间结果,然后从中间结果中提取人脸结构,最后利用先验信息和中间结果恢复人脸图像。

在超分过程中,当缩放因子较大时,生成的高分辨率图像会失去纹理细节,图像重建效果较差。随着GAN的强大生成优势在图像增强领域大放异彩,研究人员开始探索基于GAN的超分方法。首次将GAN应用于超分并取得显著效果的是Ledig等^[7]提出的超分辨率生成对抗网络(Super-Resolution Generative Adversarial Network, SRGAN),其首先对低分辨率图像进行处理并输送到GAN的生成器中进行训练,从而生成高分辨率图像,再将其输入判别器,通过原始高清图片辨别生成图片的真伪。SRGAN能够明显提升图像质量,拥有较好的视觉感知效果。Wang等^[8]在SRGAN的基础上对其进行改进,主要从网络结构、对抗损失和感知损失这3方面加以改进,提出了增强型超分辨率生成对抗网络(Enhanced Super-Resolution Generative Adversarial Network, ESRGAN),在真实度与纹理细节上,ESRGAN比SRGAN拥有更好的视觉效果。Lim等^[9]提出一种增强的深度超分辨网络(Enhanced Deep Super-Resolution Network, EDSR),EDSR通过去除批量归一化层(Batch Normalization, BN)、加深网络宽度和深度以提高图像超分辨重建效果。Ma等^[10]提出了一种深度迭代合并的网络DICGAN,其设计了一个两分支的网络结构,一个分支用于人脸恢复,另一个分支用于人脸特征点计算。同时,为了更有效地融合人脸图像的特征点,DICGAN还采用了注意力融合模块,而不是简单的拼接。

一些学者开始关注无监督的超分方法,以构建符合实际应用的超分方法,即直接对实际应用中的低分辨率图像进行超分辨率重建,而无需依赖额外的低分辨率和高分辨率图像对。Shocher等^[11]根据单个图像内部信息在不同尺度会重复出现的现象,提出一种基于无监督的零次超分方法(Zero-Shot Super-Resolution, ZSSR)。这是一种全卷积神经网络,所采用的损失函数是L1函数。ZSSR利用单个图像内部信息训练一个图像特定的CNN,在构建训练集时仅利用测试图像,使得模型可以适应各种图像。

但是,使用上述方法对交通卡口过车人脸图像超分辨率重建时仍存在问题:SRCNN网络结构较简单,在捕

获和重建复杂的面部细节时效果不理想;SRGAN在生成具有真实感纹理的人脸图像方面表现优异,但生成的图像存在的噪声和伪影仍需改进;由于EDSR是一个较为庞大的模型,因而在实际应用中会受到计算资源和时间的限制。

ESRGAN算法得到的人脸高分辨率图像画质更佳,且具有更自然和真实的纹理,但仍存在模糊、细节处理不足等问题。为进一步提升人脸超分辨率的效果,本文采用改进后的ESRGAN对卡口过车人脸图像进行超分辨率重建,并对不同放大倍数下的重建效果进行分析。与当前主流超分辨率方法相比,改进后的ESRGAN能够更好地恢复出较清晰的人脸图像,在评价指标上达到了最优。

1 相关模型

SRGAN在超分辨率上具有开创性,其能够在单个图

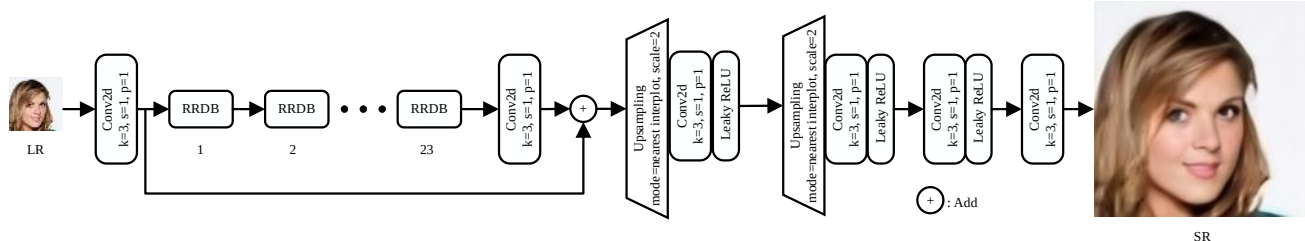


Fig. 1 ESRGAN generative network structure ($\times 4$)

图1 ESRGAN生成网络结构($\times 4$)

为了提高SRGAN生成高分辨率图像的质量,ESRGAN主要对SRGAN生成器的结构进行了两处改进。

第一处改进是移除了所有的BN层,如图2所示,其中(a)图是SRGAN的原始残差块(Residual Block, RB), (b)图是ESRGAN中移除了BN层的残差块。

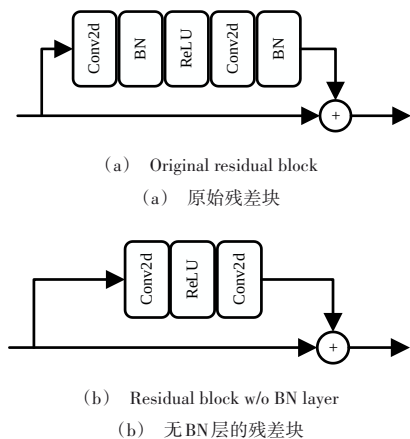


Fig. 2 Removing BN layer

图2 移除BN层

移除BN层能减少以峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)为导向的任务内存使用量和计算复杂度,从而提高性能^[12]。当训练和测试数据集的统计数据有

像超分辨率过程中生成逼真的纹理,但是超分得来的图像通常伴随着令人不快的伪影。为了进一步提升视觉质量,ESRGAN对SRGAN的网络结构、对抗损失和感知损失这3个关键组成部分进行了改进,从而提高了超分的整体感知质量。

1.1 生成网络

如图1所示(彩图扫OSID码可见,下同),ESRGAN仍然采用了SRResNet的基本架构^[7]。LR图像首先经过卷积层提取低层特征,然后输入到23个残差中的残差密集块(Residual-in-Residual Dense Block, RRDB)中逐层提取特征,最后通过多级上采样生成所需尺度的高分辨率人像,即放大倍数为 $\times 2$ 则使用1个上采样模块, $\times 4$ 则为2个上采样模块,依此类推。由此可见,在LR特征空间中即可完成大部分计算。由于LR图像对应的特征图尺寸相对较小,因而计算数据量也较小,从而可以加快超分速度。

较大差异时,BN层会引起伪影,限制了模型泛化能力,因而移除BN层可减少伪影,并提高模型泛化能力。

第二处改进是将Basic Block从SRGAN的原始残差块替换为RRDB。更多的层和连接总是能够提升性能,基于此RRDB,采用比SRGAN中原始残差块更深、更复杂的结构^[13]。如图3所示,RRDB由3个残差密集块(Residual Dense Block, RDB)构成,RDB通过结合多层残差网络和密集连接以提高特征提取能力。将通过3个残差密集块的特征图先乘以 β 倍的残差缩放后再输入到主残差网络中,其中残差缩放参数 $\beta=0.2$ 是一个介于0~1区间的超参数,用于缩小残差、防止不稳定以更好地促进训练。该结构具有更深、更复杂的网络,增加了网络容量,从而提高了网络性能。

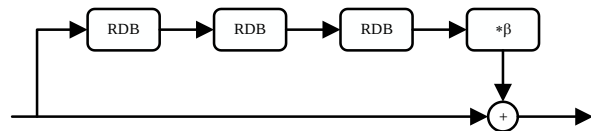


Fig. 3 RRDB network structure

图3 RRDB网络结构

其中,RDB是由去除BN层的残差块用密集连接的方式组成,具体结构如图4所示,每个密集块都含有5个 3×3 的卷积层,采用的激活函数是Leaky ReLU,这有助于提升网络稳定性和收敛速度。

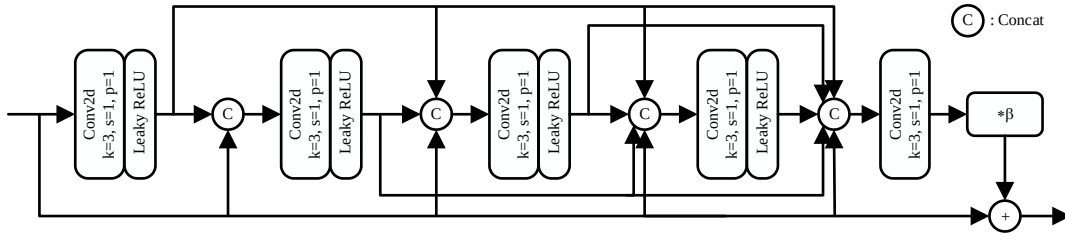


Fig. 4 RDB network structure

图4 RDB网络结构

1.2 生成网络改进

ESRGAN算法在较大放大倍数下重建后的图像仍然存在人脸细节模糊、显示不全等问题。为提高较大放大倍数下人脸图像生成质量,进一步增强其细节纹理,对ESRGAN的生成网络结构进行改进^[14-15]。

1.2.1 ERRDB

密集连接能够促进特征传递,更高效地利用不同层的特征进行特征融合,故在RRDB中引入密集连接以构建新的网络(Enhanced Residual-in-Residual Dense Block, ERRDB)。对每个残差密集块之前的所有输入进行拼接,并将拼接后的特征输入到对应的残差密集块中。引入密集连接的方法使得网络中的每一层都能够直接利用损失函数的梯度信息及输入信息,该策略能减少梯度消失问题,有助于训练出更深层次的网络结构^[16]。此外,密集连接中每一层的输出特征都能直接成为后续所有层的输入,这种设计不仅确保了信息在传播过程中的完整性,防止了信息丢失,还提升了特征利用率,加强了RRDB模块对人脸特征的提取。

改进后的ERRDB如图5所示。设其输入特征为 $X_{in}(h, w, c)$,其中 h 、 w 和 c 分别表示特征图的高、宽和通道数,则其特征提取过程可表示为:

$$X_1(h, w, c) = \text{RDB}_1(X_{in}) \quad (1)$$

$$X_2(h, w, 2c) = \text{RDB}_2[C_a(X_{in}, X_1)] \quad (2)$$

$$X_3(h, w, 4c) = \text{RDB}_3[C_a(X_{in}, X_1, X_2)] \quad (3)$$

其中, C_a 表示在通道维数上进行特征拼接操作,然后使用 3×3 的卷积核和LReLU激活函数实现深度信息融合,同时完成特征降维输出,最后将其与残差缩放参数 β 相乘后再与输入特征图 X_{in} 按元素相加,即可得到ERRDB模块最终的输出特征图 X_{out} 。

$$X_{out}(h, w, c) = \text{LReLU}[\text{Conv}_{3 \times 3}(X_3)] \times \beta + X_{in} \quad (4)$$

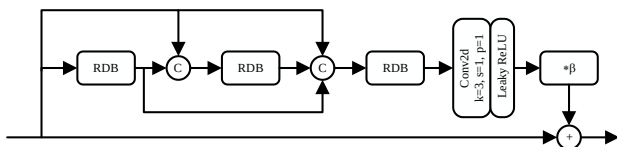


Fig. 5 ERRDB network structure

图5 ERRDB网络结构

1.2.2 CARDB

在生成网络的特征提取模块中引入注意力机制可进一步提升超分效果^[16-18],例如性能优越而参数量较少的坐标注意力机制(Coordinate Attention, CA)^[19]。因此,本文在特征提取模块RDB的最后一个卷积层后加入CA以构成坐标注意力残差密集块(Coordinate Attention Residual Dense Block, CARDB)。CA能高效地捕捉通道信息和位置信息的相关性,也能获取各通道之间的关系,从而加强生成网络对人脸特征的选择能力,提升算法对人脸超分辨率重建的效果。

CA的网络结构如图6所示。CA将通道注意力先分解为两个并行的一维特征编码过程,然后沿着这两个方向即垂直方向和水平方向对特征进行聚合,从而生成两个方向感知的特征图,再对这两个特征图进行编码以捕获输入特征图的长距离依赖关系和位置信息,最后通过乘法将这两个注意力图都应用于输入特征图,以强调注意区域的表示,这样便可通过CA在超分任务中生成有清晰细节纹理的人脸重建图像。

从图6可知,输入特征 X 的第 c 个通道通过空间范围分别为 $(H, 1)$ 和 $(1, W)$ 的2个平均池化层进行编码,则编码后此通道在高度为 h 处的元素和宽度为 w 的元素分别表示为:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (5)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (6)$$

池化后对第二个特征图的维度重新排列,再将2个聚合特征图在高度维上拼接起来,输入到卷积核为 1×1 的卷积模块中,得到:

$$f = \delta(\text{BN}(\text{Conv}_{1 \times 1}(C_a(z^h, z^w)))) \quad (7)$$

其中, f 为在垂直方向和水平方向上进行空间信息编码得来的中间特征图, C_a 是在空间维上进行拼接操作, δ 为非线性激活函数h-swish,如下:

$$h\text{-swish}(x) = x \times \text{ReLU6}(x + 3)/6 \quad (8)$$

再沿着第2维的空间维度将中间特征图分割为2个特征张量 f^h 和 f^w ,经过卷积层和激活函数得到与输入特征同样的通道数,即:

$$g^h = \delta(\text{Conv}_{1 \times 1}(f^h)) \quad (9)$$

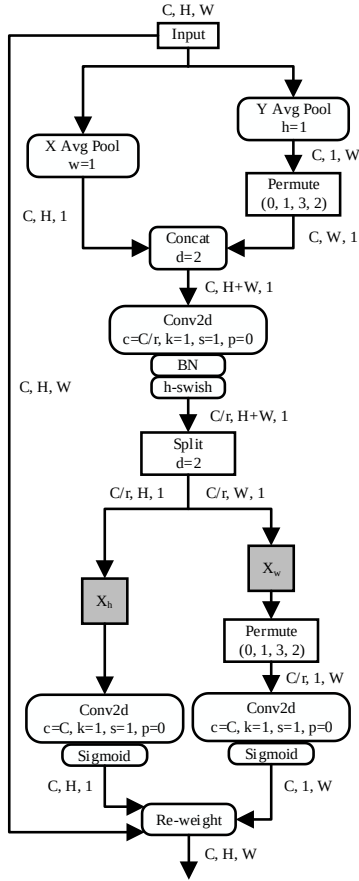


Fig. 6 CA network structure

图6 CA网络结构

$$g^w = \delta(\text{Conv}_{1 \times 1}(f^w)) \quad (10)$$

其中, δ 表示 Sigmoid 激活函数。将这 2 个输出分别扩展并用作注意力权重, 将其与输入特征图 X 相乘得到输出 Y 。

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (11)$$

在较大放大倍数如 $\times 4$ 和 $\times 8$ 的人脸图像超分辨率重建中, 由于输入的低分辨率图像尺寸较小, 计算数据量也较小, 因此改进后的 ESRGAN 生成网络即将图 1 中的 23 个 RRDB 替换为网络更深更复杂的 ERRDB, 并在 RDB 中引入坐标注意力机制 CA, 构成 CARDDB 替换原算法的 RDB 以增强网络的特征选择能力。而较小放大倍数 $\times 2$ 下, ESRGAN 超分辨率重建后得到的人脸图像相对而言较清晰, 能够较好地展示人脸特征并用于人脸识别, 因此不对 $\times 2$ 倍的生成网络进行改动。

1.3 鉴别网络

ESRGAN 根据相对鉴别网络替换了 SRGAN 的标准鉴别器^[20]。SRGAN 的标准鉴别器是估计一张输入图像 x 是真实自然的概率, 旨在鉴别输入图像是真实还是伪造的。而 ESRGAN 的相对鉴别器则是预测一张真实图像 x_r 相较于一张假的图像 x_f 更真实的概率, 它不仅仅将单个图像分类为真实或伪造, 而且考虑真实图像与伪造图像之间的相

对真实性。这种方法通过考虑生成图像相对于真实图像的相对真实性, 为生成器提供了更有意义的反馈。

在 SRGAN 中, 标准鉴别器可以表示为:

$$D(x) = \sigma(C(x)) \quad (12)$$

其中, $C(x)$ 是未转换的鉴别器输出, σ 是 Sigmoid 函数, 则相对鉴别器为:

$$D_{Ra}(x_r, x_f) = \sigma(C(x_r) - E_{x_f}[C(x_f)]) \quad (13)$$

其中, $E_{x_f}[\cdot]$ 表示对小批量中所有伪造数据取平均值, x_r 代表真实的高分辨率图像, $x_f = G(x_i)$ 是生成器 G 生成的图像, 而 x_i 则表示输入的低分辨率图像。

因此, 相对鉴别器损失函数可定义为:

$$L_D^{Ra} = -E_{x_r}[\log(D_{Ra}(x_r, x_f))] - E_{x_f}[\log(1 - D_{Ra}(x_f, x_r))] \quad (14)$$

相对应的生成器对抗损失为:

$$L_G^{Ra} = -E_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - E_{x_f}[\log(D_{Ra}(x_f, x_r))] \quad (15)$$

由式 (15) 可观察到, 生成器的对抗损失包含了 x_r 和 x_f 两部分, 因此 ESRGAN 的生成器在对抗训练中既可利用从生成图像中获得的梯度, 也可利用从真实图片中获得的梯度, 而在 SRGAN 中只有生成部分的数据起作用。

1.4 损失函数

基于更接近感知相似性的理论^[21-22], Johnson 等^[23]提出了感知损失, 并在 SRGAN 中得以扩展。感知损失先前是在预先训练的深度网络的激活层上定义, 最小化了两个激活特征之间的距离。这会引起两大不足: ①激活后的特征将变得非常稀疏, 尤其是在非常深的网络之后, 这导致了较弱的监督, 从而降低了性能; ②使用激活后的特征还会造成重建图像与真实图像的亮度不一致。鉴于此, ESRGAN 提出了新的感知损失, 即与之相反使用激活层之前的特征, 从而解决这两个问题。

因此, 生成器的总损失为:

$$L_G = L_{Percep} + \lambda L_G^{Ra} + \eta L_1 \quad (16)$$

其中, L_{Percep} 为所优化的感知损失, L_G^{Ra} 为生成器对抗损失, 内容损失是:

$$L_1 = E_{x_i} \|G(x_i) - y\|_1 \quad (17)$$

其评估了生成图像 $G(x_i)$ 与真实图像 y 之间的 1 范数距离, λ 和 η 是用于平衡不同损失项的参数。

2 实验与结果分析

2.1 实验环境及参数设置

实验所采用的操作系统是 Windows 11, 处理器为 13th Gen Intel (R) Core (TM) i7-13700H 2.40 GHz, 内存为 16 GB, 在深度学习框架 Pytorch 2.0.0+cu118 下使用显卡 NVIDIA GeForce RTX 4060 Laptop GPU 8GB 进行模型训练和测试。

批量大小设置为 8, 所使用的优化器是 Adam, 其中一

阶动量和二阶动量分别为 0.9 和 0.99, 初始学习率为 2×10^{-4} , 总迭代次数是 40 000, 每迭代 1 000 次, 便对验证集进行评估并保存训练权重。

2.2 数据集

实验所采用的数据集是基于公开的大规模人脸数据集 CelebA^[24] 和自制的交通卡口过车所拍摄的人脸数据集。CelebA 包含 202 599 张高清人脸居中图像, 且分辨率都为 178×218 , 从中选取 2 000 张图像并裁去图像边缘的背景区域, 将其重塑为 168×168 的图像以构成数据集 MiniCelebA, 然后按照 8:1:1 的比例将其划分为训练集、验证集和测试集。自制的交通卡口人脸数据集 YuxiFaces 包含 150 张真实的交通卡口过车人脸图像, 也将其作为模型的另一个测试集。

由于公开数据集 CelebA 中的图像经过裁剪背景区域后仅包含人的头像部分, 而卡口人脸图像则是由摄像头拍摄到的照片经过人脸检测算法截取出来的人脸图像部分, 两者具有相似性即都是仅包含人脸图像而无其他背景信息, 故可以用公开的人脸图像数据集训练好模型再用于卡口处人脸图像超分辨率重建, 而且实际卡口处监控拍摄到的人脸图像分辨率较低, 不适合用于需要大量高清图像的有监督的深度学习模型 ESRGAN 的训练, 故考虑使用公开数据集训练好模型后再将其应用于实际。

为获得低分辨率的人脸图像, 使用 OpenCV 库中的 `resize` 函数, 选择双 3 次插值法对 HR 图像分别进行 2 倍、4 倍和 8 倍的下采样处理, 以获得对应的 LR 图像并构成低分辨率和高分辨率图像对用于训练。

2.3 评价指标

采用结构相似度 (Structural Similarity Index Measure, SSIM) 和 PSNR 这 2 个常用的客观评价指标以衡量超分方法的性能。PSNR 和 SSIM 的值越大, 则表示超分辨率重建图像的质量越好, 即此种超分方法的性能越高。

SSIM 通过比较 2 幅图像的亮度、对比度和结构信息以评估它们之间的相似程度, 其取值范围为 0~1。SSIM 越接近 1, 表明 2 幅图像的内容和结构越相似, 即原始图像和重建图像越接近。给定图像 x 和 y , 则 SSIM 计算公式为^[25]:

$$SSIM = \frac{(2u_x u_y + c_1)(2\sigma_{xy} + c_2)}{(u_x^2 + u_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (18)$$

其中, u_x 和 u_y 分别表示图像 x 和 y 的平均值, 代表了亮度信息; σ_x^2 和 σ_y^2 为各自图像的方差, 代表了对比度信息; σ_{xy} 为 x 和 y 的协方差, 代表结构信息; c_1 和 c_2 是用于保持数值稳定性的常数, 其中 $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, L 为像素值的动态范围, $k_1 = 0.01$, $k_2 = 0.03$ 。

PSNR 通过比较信号的最大功率与信号噪声功率以衡量图像的重建质量, 它是基于均方误差 (Mean Square Error, MSE) 而计算, 单位是分贝, 较高的 PSNR 意味着超分辨率重建图像质量较高。其计算公式为^[25]:

$$PSNR = 10 \lg \frac{255^2}{\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2} \quad (19)$$

式中, m 和 n 分别为图像的行数和列数; $I(i, j)$ 是原始图像第 i 行、第 j 列的像素值; $K(i, j)$ 则是重建图像相对应的第 i 行、第 j 列的像素值。

在超分辨率重建的速度方面, 采用单张图片超分耗时 Times 来衡量模型对低分辨率图像进行超分辨率重建时平均所需要的时间, Times 越低表明模型处理单张低分辨率图像时所需时间更短, 即超分效率更高。

2.4 实验结果

本实验的目的是对卡口过车低分辨率人脸图像分别进行 2 倍、4 倍和 8 倍的超分辨率重建。为验证改进的 ESRGAN 算法的有效性, 在较大放大倍数 $\times 4$ 和 $\times 8$ 下将其与主流超分辨率算法 SRCNN、SRGAN、EDSR、ZSSR 和 ESRGAN 进行比较, 性能评价指标由 SSIM、PSNR 和 Times 给出。

2.4.1 对比实验

使用 MiniCelebA Test 和 YuxiFaces 这 2 个数据集作为测试集, 以验证改进的 ESRGAN 算法的有效性和泛化性。对这 2 个测试数据集进行 2 倍、4 倍和 8 倍的超分辨率重建后, 各种图像超分辨率重建算法的客观评价结果如表 1—表 3 所示。

Table 1 Comparison of 2x super-resolution reconstruction results of mainstream algorithms

Method	MiniCelebA Test		YuxiFaces	
	PSNR/dB	SSIM	PSNR/dB	SSIM
SRGAN	29.06	0.878 8	35.30	0.936 5
SRCNN	30.32	0.906 3	36.70	0.953 9
ZSSR	31.31	0.925 2	37.32	0.960 6
EDSR	34.28	0.947 0	40.38	0.976 1
ESRGAN	34.41	0.949 4	40.48	0.977 8

Table 2 Comparison of 4x super-resolution reconstruction results of different algorithms

Method	MiniCelebA Test		YuxiFaces	
	PSNR/dB	SSIM	PSNR/dB	SSIM
SRGAN	23.81	0.729 7	28.66	0.814 9
SRCNN	24.48	0.758 3	29.78	0.850 6
ZSSR	24.70	0.783 0	29.89	0.861 7
EDSR	27.92	0.844 1	32.61	0.904 1
ESRGAN	28.04	0.850 4	32.80	0.907 2
Ours	28.15	0.852 9	32.91	0.908 4

从表 1 可知, 原始的 ESRGAN 算法进行 2 倍的人脸超分辨率重建时效果优于其他主流算法。从较高的 PSNR 和 SSIM 值可以看出, 图像重建结果较好, 可应用于实际。

从表 2 和表 3 可知, 随着放大倍数的增加, 各算法重建图像的 PSNR 和 SSIM 都逐渐减小, 表明超分算法的性能随着放大倍数的增大而逐渐降低。在 $\times 4$ 和 $\times 8$ 的放大倍数下, 改进后的 ESRGAN 算法重建图像的 PSNR 和 SSIM 均优

Table 3 Comparison of 8x super-resolution reconstruction results of different algorithms

表 3 各算法 8 倍超分辨率重建结果比较

Method	MiniCelebA Test		YuxiFaces	
	PSNR/dB	SSIM	PSNR/dB	SSIM
SRGAN	19.49	0.557 5	23.81	0.658 1
SRCNN	20.22	0.583 0	24.92	0.706 0
ZSSR	20.57	0.627 8	25.20	0.728 0
EDSR	23.03	0.692 5	27.06	0.777 1
ESRGAN	23.12	0.701 7	27.08	0.781 8
Ours	23.29	0.704 5	27.33	0.783 9

Table 4 Comparison of Times results of ESRGAN algorithm before and after improvement

表 4 改进前后 ESRGAN 算法 Times 结果比较

Method	倍率	Times/s	
		MiniCelebA Test	YuxiFaces
ESRGAN	×4	0.106	0.092
Ours		0.170	0.168
ESRGAN	×8	0.088	0.084
Ours		0.153	0.150

于其他算法。与 ESRGAN 算法相比,本文算法进行 4 倍超分辨率重建时,在 2 个数据集上的 PSNR 平均提高了 0.11 dB,SSIM 平均提高了 0.001 9;在 ×8 倍率下,PSNR 平均提高了 0.21 dB,SSIM 平均提高了 0.002 5。由此可见,本文算法在进行较大倍率超分辨率重建时更有效。

表 4 展示了改进前后 ESRGAN 算法在 2 个测试集上的单张图片超分耗时 Times。可以看出,随着倍率的增大,Times 逐渐减小,这是因为用于 8 倍超分辨率重建的 LR 图像相对而言分辨率更低,从而使得计算特征图也较小,故超分时间更短。比较同一模型在同一倍率下的 Times 可以发现,其所需时间相差无几。综合 2 个数据集看,在 ×4 倍率下,改进后的 ESRGAN 算法对单张图片的超分耗时约为 0.169 s,比基线模型 ESRGAN 算法的 0.099 s 多出了 0.07 s;在 ×8 倍率下,改进后的算法多出了 0.066 s 左右。可以得知,本文算法以略微增加的时间成本取得了更好的人脸超分辨率图像重建效果,并且依旧满足实际卡口图片超分任务中的实时性要求。

图 7 展示了主流超分辨率算法在 MiniCelebA Test 上进行 2 倍图像超分辨率重建的直观比较结果。可以看出,ESRGAN 算法重建的人脸图像相较于其他算法更清晰,复原效果也最接近于原始的 HR 图片,能较好地满足实际应用需求。

图 8 展示了各算法在 MiniCelebA Test 上放大倍数为 4 的人脸超分辨率重建结果比较。通过总体比较可知,本文算法对人脸图像重建效果更好,而其他方法虽然均可提升图像质量,但是或多或少还存在着图像较模糊、有噪声等问题。改进前的 ESRGAN 算法在对图 8 中第一个人的头像进行超分时可知,其对左眼处理效果不佳。HR 图片中人物左眼的左下角部分有一些眼白,改进前的算法对图像超分后这部分全是瞳孔,改进后的算法则复原出了这部分



Fig. 7 Comparison of 2x reconstruction results of mainstream algorithms on MiniCelebA Test

图 7 在 MiniCelebA Test 上主流算法 2 倍重建结果比较

的人眼特征。由此可知,改进后的 ESRGAN 算法通过在 RRDB 中进一步引入密集连接,加强了特征传播过程中对信息利用。同时,引入注意力机制,关注重要的人脸特征,复原了更多人脸细节,得到的图像更加清晰。

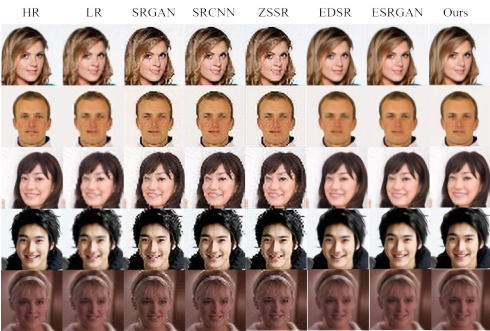


Fig. 8 Comparison of 4x reconstruction results of various algorithms on MiniCelebA Test

图 8 在 MiniCelebA Test 上各算法 4 倍重建结果比较

2.4.2 消融实验

本文验证了在 ESRGAN 算法中采用 ERRDB 模块和 CARDDB 模块的有效性,表 5 展示了在 8 倍超分辨率重建时 4 组消融实验的 PSNR 和 SSIM 值。其中,DC 代表引入密集连接的 ERRDB 模块,CA 表示加入坐标注意力机制的 CARDDB 模块,√和×分别表示使用与未使用该模块。

Table 5 Ablation experiment

表 5 消融实验

DC	CA	MiniCelebA Test		YuxiFaces	
		PSNR/dB	SSIM	PSNR/dB	SSIM
×	×	23.12	0.701 7	27.08	0.781 8
√	×	23.20	0.703 1	27.21	0.782 0
×	√	23.23	0.703 6	27.26	0.783 5
√	√	23.29	0.704 5	27.33	0.783 9

可以看出,同时使用这两个模块时,评价指标相对较高,在单独使用时它们对 ESRGAN 算法的超分性能均有一定提升,但坐标注意力机制的引入对算法的改进更大,表明 CA 通过捕获长距离依赖关系并保留精确的位置信息使得模型可以更好地捕获人脸细节特征。

3 结语

为改善卡口处获取的人脸图像分辨率低进而影响人脸识别等问题,本文基于改进的ESRGAN算法对人脸图像进行超分辨率重建。首先,在信息提取模块RRDB中引入密集连接,这一设计促进了特征在网络各层之间的传递,通过这种策略使得网络能够更有效地利用来自不同层级的特征信息,提高了其从低分辨率图像中学习高频细节的能力,加强了对人脸特征的提取;其次,将坐标注意力机制引入到RDB中,通过捕获位置信息和通道方面的相关性,加强了网络对人脸特征的选择能力。将改进后的ESRGAN算法与其他典型的超分辨率算法进行对比,实验结果表明,所提算法得到的重建图像在人脸细节上更清晰,有更好的视觉效果,在公开人脸数据集和自制人脸数据集上的评价指标PSNR和SSIM都达到了最优值,验证了所提方法的有效性,为提高交通卡口过车人脸识别准确率提供了一定的技术支持。

但同时,较低分辨率人脸图像经过超分后其评价指标还是较低,得到的人脸图像细节特征依旧不够明显,对人脸识别造成困难。因此,下一步工作是进一步提高较低分辨率图像的超分效果并减少算法处理图片的时间。

参考文献:

- SONG Z Q. Research on facial image enhancement based on attention mechanism [D]. Beijing: People's Public Security University of China, 2023.
宋治强. 融合注意力机制的人脸图像增强研究[D]. 北京: 中国人民公安大学, 2023.
- DONG C, LOY C C, HE K M, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307.
- ZHANG Y L, TIAN Y P, KONG Y, et al. Residual dense network for image restoration [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2472-2481.
- ZHU D M, QIU D F. Residual dense network for medical magnetic resonance images super-resolution [J]. Computer Methods and Programs in Biomedicine, 2021, 209(1): 106330-106330.
- HU X C, MU H Y, ZHANG X Y, et al. Meta-SR: a magnification-arbitrary network for super-resolution [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1575-1584.
- LIU Z S, SIU W C, CHAN Y L. Features guided face super-resolution via hybrid model of deep learning and random forests [J]. IEEE Transactions on Image Processing, 2021, 30(3): 4157-4170.
- LEDIG C, THEIS L, HUSZAR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4681-4690.
- WANG X T, YU K, WU S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks [C]//Proceedings of the European Conference on Computer Vision Workshops, 2018: 63-79.
- LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 136-144.
- MA C, JIANG Z Y, RAO Y M, et al. Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5569-5578.
- SHOCHER A, COHEN N, IRANI M. "Zero-shot" super-resolution using deep internal learning [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3118-3126.
- NAH S, KIM T H, LEE K M. Deep multi-scale convolutional neural network for dynamic scene deblurring [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3883-3891.
- ZHANG Y L, LI K P, LI K, et al. Image super-resolution using very deep residual channel attention networks [C]//Proceedings of the European Conference on Computer Vision, 2018: 286-301.
- SHANG T Z, DAI Q J, ZHU S C, et al. Perceptual extreme super-resolution network with receptive field block [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 440-441.
- RAKOTONIRINA N C, RASOANAIVO A. ESRGAN+: further improving enhanced super-resolution generative adversarial network [C]//IEEE International Conference on Acoustics, Speech and Signal Processing, 2020: 3637-3641.
- XIN Y X, ZHU F T, SHI P F, et al. Super-resolution reconstruction algorithm of images based on improved enhanced super-resolution generative adversarial network [J]. Laser and Optoelectronics Progress, 2022, 59(4): 381-391.
辛元雪, 朱凤婷, 史朋飞, 等. 基于改进增强型超分辨率生成对抗网络的图像超分辨率重建算法[J]. 激光与光电子学进展, 2022, 59(4): 381-391.
- HE Z M, HUANG Z C. Image super-resolution reconstruction algorithm based on coordinated attention generative adversarial network [J]. Microelectronics and Computer, 2023, 40(12): 35-44.
贺智明, 黄志成. 基于坐标注意力生成对抗网络的图像超分辨率重建[J]. 微电子学与计算机, 2023, 40(12): 35-44.
- YANG X Y, DENG M L, GAO H, et al. A super resolution reconstruction algorithm for face image based on coordinate attention mechanism [J]. Electronic Design Engineering, 2022, 30(19): 153-158.
杨晓雅, 邓森磊, 高辉, 等. 基于坐标注意力机制的人脸图像超分辨率重建算法[J]. 电子设计工程, 2022, 30(19): 153-158.
- HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- JOLICOEUR-MARTINEAU A. The relativistic discriminator: a key element missing from standard GAN [C]//Vancouver: International Conference on Learning Representations, 2018.
- BRUNA J, SPRECHMANN P, LECUN Y. Super-resolution with deep convolutional sufficient statistics [C]//Sanjuan: International Conference on Learning Representations, 2016.
- GATYS L A, ECKER A S, BETHGE M. Texture synthesis using convolutional neural networks [C]//Twenty-ninth Annual Conference on Neural Information Processing Systems, 2016: 262-270.
- JOHNSON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution [C]//Proceedings of the European Conference on Computer Vision Workshops, 2016: 694-711.
- LIU Z W, LUO P, WANG X G, et al. Deep learning face attributes in the wild [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 3730-3738.
- WANG C, LIU K, DU L. Evaluation and analysis of full reference image quality indicators [J]. Modern Electronics Technique, 2023, 46(21): 39-43.
王成, 刘坤, 杜砾. 全参考图像质量指标评价分析[J]. 现代电子技术, 2023, 46(21): 39-43.

(责任编辑: 孙 娟)