2025 年 第 25 卷 第 6 期
2025, 25(6): 02442-11

科 学 技 术 与 工 程
Science Technology and Engineering

ISSN 1671—1815
CN 11—4688/T

DOI:10.12404/j.issn.1671-1815.2403052
引用格式:李培育, 张雅丽, 张奕博, 等. 基于自适应卷积与联合损失函数的人脸图像超分辨率重建[J]. 科学技术与工程, 2025, 25(6): 2442-2452.
Li Peiyu, Zhang Yali, Zhang Yibo, et al. Face image super-resolution reconstruction based on adaptive convolution and joint loss function [J]. Science Technology and Engineering, 2025, 25(6): 2442-2452.

基于自适应卷积与联合损失函数的人脸图像超分辨率重建

李培育, 张雅丽*, 张奕博, 赵益辰
(中国人民公安大学信息与网络安全学院, 北京 100038)

摘 要 针对当前人脸图像超分辨率重建算法模型卷积单一、感受野不足、单判别网络反馈信息不精确等问题,设计了一种基于自适应卷积与联合损失函数的算法。模型使用生成对抗网络架构,生成器方面,使用自适应卷积构造双路残差块并进一步组成高效的残差组,能自主学习在不同感受野下提取到的特征权重并补充单一支路遗漏的信息。判别器方面使用 Vgg 与 U-net 架构网络作为双判别网络,并使用双判别结果计算对抗损失,该损失与内容损失、感知损失组成联合损失函数。在 celeba 数据集上的实验表明,该算法与 RWSA 算法相比峰值信噪比(peak signal noise ratio, PSNR)值提高 1.166 dB,结构相似度(structure similarity, SSIM)值提高 0.037,学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)值优化 0.033,感知因子(perceptual index, PI)指标优化 0.119,与其他多种主流算法相比在图像细节清晰度方面具有优势。

关键词 超分辨率重建; 自适应卷积; 联合损失函数; 生成对抗网络; 卷积神经网络
中图法分类号 TP391; **文献标志码** A

Face Image Super-resolution Reconstruction Based on Adaptive Convolution and Joint Loss Function

LI Pei-yu, ZHANG Ya-li*, ZHANG Yi-bo, ZHAO Yi-chen
(College of Information and Cyber Security, People's Public Security University of China, Beijing 100038, China)

[**Abstract**] Content Aiming at the problems of single convolution model, insufficient Receptive field and inaccurate feedback information of single discriminant network in current face image super-resolution reconstruction algorithm, an algorithm based on adaptive convolution and joint Loss function was designed. A generation adversarial network architecture was used by the model. On the generator side, adaptive convolution was used to construct dual path residual blocks and further form efficient residual groups. It can independently learn feature weights extracted under different receptive fields and supplement missing information from a single branch. The subpixel convolution layers were used to complete quadruple reconstruction of face images. In terms of discriminators, Vgg and U-net architecture networks were used as dual discriminant networks, and dual discriminant results were used to calculate adversarial losses. The losses, content losses, and perceptual losses form a joint loss function. Experiments on the Celeba dataset show that compared with RWSA, this algorithm improves PSNR by 1.166 dB, SSIM by 0.037, LPIPS by 0.033, and PI by 0.119, compared with other mainstream algorithms, it has advantages in image detail clarity.

[**Keywords**] super-resolution reconstruction; adaptive convolution; joint loss function; generate countermeasure; convolutional neural network

近年来,各行业对于提升图像清晰度有很大需求,公安实战应用中希望将监控中得到的低分辨率人脸图像变得更具识别性,医学诊断等其他领域也希望使用的图像质量更好。图像超分辨率重建技术可以提高质量差的图像分辨率同时减少图像中的噪声。

目前,深度学习广泛用于图像超分辨率重建领域,Dong 等[1]使用卷积神经网络完成图像超分辨率

收稿日期: 2024-04-25; **修订日期**: 2024-12-13
基金项目: 中国人民公安大学安全防范工程双一流创新研究专项(2023SYL08)
第一作者: 李培育(1998—),男,汉族,河南信阳人,硕士研究生。研究方向:图像超分辨率重建、安全防范技术。E-mail:1448393699@qq.com。
* **通信作者**: 张雅丽(1977—),女,汉族,山西大同人,硕士,副教授。研究方向:安全防范技术与智能视频技术。E-mail:zhangyl_mail@163.com。

重建,并将重建部分的工作从网络前端调到网络尾端[2],有效降低了网络中间层的计算复杂度。Kim 等[3]加深了网络的结构,使用更多的卷积层来提取更丰富的特征,并使用了残差网络解决深层卷积带来的训练困难等问题。Tai 等[4]提出一种使用多记忆模块加强不同层次信息利用的算法。与之相似的是,Tong 等[5]提出的基于密集块的算法,将中间层的残差块所提取的特征都加权到后面残差块中,充分利用了低分辨率图像的信息,提高了模型精度。Zhang 等[6]在网络中加入通道注意力,使得网络在提取特征时对重要性不同的通道信息有所取舍。之后 Woo 等[7]补充了空间注意力机制,增加了网络对于空间信息的利用。通道和空间注意机制极大地改进了一般的重建方法,这激励了研究人员探索它们在人脸超分辨率重建中的应用,其中具有代表性的是引入了通道注意力的 E-SupResNet 模型[8]与引入空间注意力的 SPARNet 模型[9],使得网络在提取特征时对重要性不同的通道信息与空间信息有所取舍,进一步提高了重建算法的精度。为了解决普通卷积网络重建图像过于光滑的问题,Ledig 等[10]在重建网络之外加入了判别网络,在对抗学习中提高图像的感知质量,MLGE 模型[11]不仅设计判别器来区分人脸图像,而且应用人脸图像的边缘映射来重建人脸图像。Lim 等[12]对生成网络进行了改进,去除 BN 层来获得质量更高的图像。Wang 等[13]提出一种将密集网络与生成对抗网络相结合的算法,在生成器网络中使用密集模块,网络性能更好。Aakerberg 等[14]首先估计低分辨率人脸下采样的参数,如模糊核、噪声和压缩,然后生成具有估计参数的人脸图像对用于模型的训练。Deng 等[15]提出一种多尺度残差块改进的 SRGAN 模型,提取图像的全局和局部信息,提升模型效果。Tong 等[16]针对现有超分辨率模型提取惨层特征不足的问题,提出了一种多尺度特种融合方法,提升了图像重建质量。

以往基于生成对抗网络的算法仅以 Vgg(visual geometry group)网络作为判别器,在判别过程中会缺少对局部特征的注意,并且特定卷积神经网络模型中的普通卷积具有固定的感受野,在设计网络模型时只能通过多次实验来人为选取较为合适的卷积核大小,这不仅耗费大量的时间与计算机资源,而且人为选取很难穷尽所有卷积核大小之间的权重组合,这导致最终设计的模型往往不是理论上最优越的。

因此,现提出基于自适应卷积与联合损失函数的人脸图像超分辨率重建算法,构造生成对抗网络的架构,基于自适应卷积设计生成器,每个自适应卷积在提取特征时具有自适应的感受野,能够通过网络输入的特征来学习调节内部各卷积核的权重大小。使用残差学习策略将多路卷积单元构造成高效的特征提取残差块,并使用子像素卷积来重建得到高分辨率人脸图像。同时使用 Vgg 架构网络与 U-net 架构网络作为双判别器,能分别从整体与局部角度对人脸图像进行判别,将双判别器判别的结果作为判别损失,与像素损失、特征损失共同组成联合损失函数,共同约束生成器的重建输出。将本文提出的基于自适应卷积与联合损失函数的人脸超分辨率重建算法在 celeba 人脸数据集[17]上进行重建,以验证本文算法重建的人脸图像的质量和识别效果。

1　设计方法

1.1　网络整体框架

本文设计的人脸图像超分辨率重建算法网络结构主要由一个生成器 G 与 Vgg 架构判别器 D_{Vgg}、U-net 架构判别器 $D_{U\text{-}net}$ 组成。算法进行模型训练的过程如图 1 所示,其中低分辨率(low resolution, LR)图像由高分辨率(high resolution, HR)图像下采样得到,将 LR 图像输入 G 网络中,经过卷积神经网络重建过程生成新的超分辨率(super resolution, SR)图像,将 SR 图像与 HR 图像分别输入到两个判别器 D_{Vgg} 与 $D_{U\text{-}net}$ 中进行提取特征、真假判别,将双

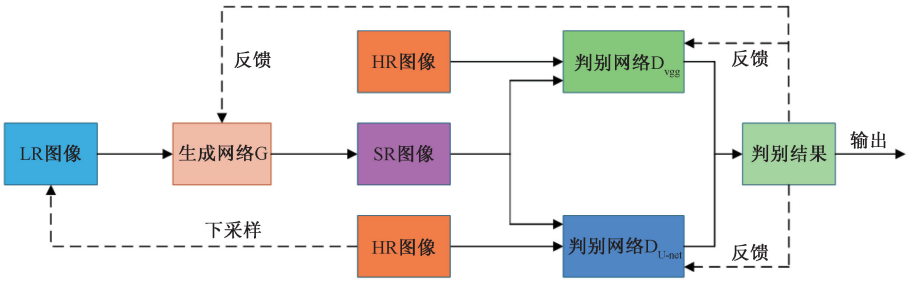


图 1　双判别器人脸超分辨率重建算法训练过程

Fig. 1　Training process of double discriminator face super-resolution reconstruction algorithm

判别器输出的判别结果共同组成损失函数约束生成与判别网络的优化训练。

1.2 生成器结构

本文网络模型的生成器总体结构如图 2 所示。其中浅层特征提取模块主要由一个卷积层与 Relu 激活函数[18]组成,将输入的低分辨率图像(low resolution,LR)提取特征得到浅层特征图,将浅层特征图作为残差组模块的输入,并计算得到深层特征,将浅层特征与深层特征进行特征融合作为上采样模块的输入,得到高分辨率图像,计算过程为

$$I^{HR} = f_{up}[f_l(I^{LR})] + f_h[f_l(I^{LR})] \tag{1}$$

式(1):$f_l(\cdot)$ 为浅层特征提取过程;$f_h(\cdot)$ 为残差组提取特征过程;$f_{up}(\cdot)$ 为上采样过程;I^{LR} 为输入的低分辨率图像;I^{HR} 为输出的高分辨率图像。

1.2.1 SK 自适应卷积

由于普通的卷积层受到固定感受野的限制,所提取到的特征会遗漏输入特征图的重要信息,本文构建生成器网络时使用 SK(selective Kernel)自适应卷积[19]作为部分的卷积层,SK 卷积结构如图 3 所示。

SK 卷积内部有多条分支,分别使用不同卷积来提取输入特征图信息,第一条分支卷积核大小为3 × 3,第 i 条分支卷积核大小为 $(1+2i)\times(1+2i)$,不同分支数量会对不同任务的模型产生不同影响,本文研究经过控制变量对比实验选取最优总分支数 $M=2$。

将特征图 $X \in \mathbf{R}^{C\times H\times W}$ 分别输入到两个分支中,第一条分支将 X 依次经过 3 ×3 卷积层得到提取特征 $U_1 \in \mathbf{R}^{C\times H\times W}$,第二条分支将 X 依次经过 5 ×5 卷积层得到提取特征 $U_2 \in \mathbf{R}^{C\times H\times W}$,其中 5 ×5 卷积是由扩张大小为 2、具有 3 ×3 核的空洞卷积[20]构成。将特征 U_1 与 U_2 经过元素求和得到融合特征 $U \in \mathbf{R}^{C\times H\times W}$,公式为

$$U = U_1 + U_2 \tag{2}$$

然后使用池化层处理特征 U,生成通道维度的统计值 $s \in \mathbf{R}^C$,特征 U 第 c 个通道的值在空间维度缩小后得到 s 的第 c 个元素,公式为

$$s_c = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} U_c(i,j) \tag{3}$$

式(3)中:H 和 W 为特征 U 每个通道上的高和宽。

进一步使用全连接层将特征 s 降低维度来提高效率,公式为

$$z = \delta[\beta(\boldsymbol{W}_S)] \tag{4}$$

式(4)中:δ 为 ReLU 激活函数;β 为批归一化;$\boldsymbol{W}_S$ 为转换矩阵。

使用 softmax 层处理特征 z 得到特征 U_1 与 U_2 的注意力向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$,a_c 与 b_c 为 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的第 c 个元素,公式为

$$a_c = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}}, \quad a_c + b_c = 1 \tag{5}$$

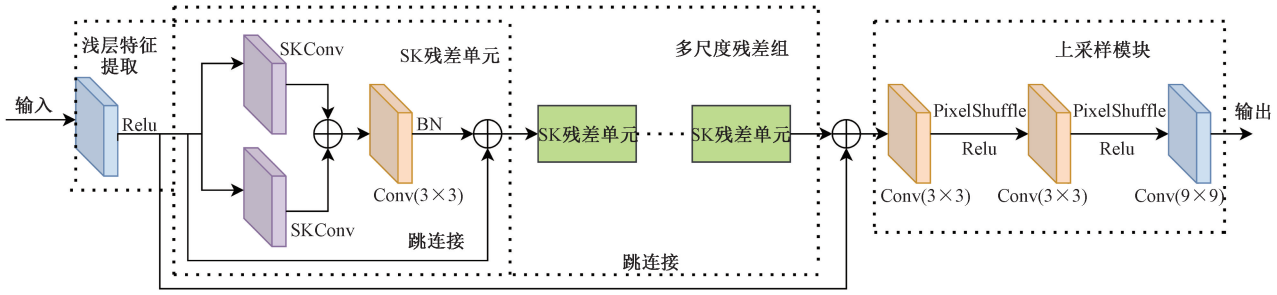
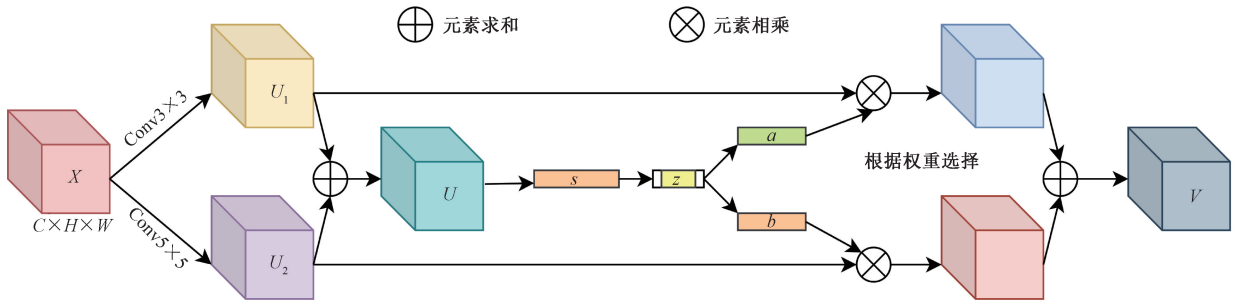


图 2 生成器网络结构

Fig. 2 Generator network structure



X 为输入的特征图;C、H、W 为输入图像的长、高、宽三维度数值;V 为最终计算得到的特征图

图 3 SK 卷积结构

Fig. 3 SK convolution structure

式(5)中:$\boldsymbol{A}$ 为转换矩阵;$\boldsymbol{A}$、$\boldsymbol{B} \in \mathbf{R}^{C\times 32}$;$A_c$ 为 $\boldsymbol{A}$ 的第 c 个元素。使用注意力向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 计算得到最终特征图 $V \in \mathbf{R}^{H\times W\times C}$,$V_c$ 为 V 的第 c 个元素,计算公式为

$$V_c = a_c U_1 + b_c U_2 \tag{6}$$

1.2.2　残差组模块

普通卷积网络是由多个卷积层进行简单叠加得到,每个中间层的特征信息经过尾端卷积层计算之后便不会保留,这使得每个卷积层都需要完整的提取前端输入的特征信息才能保证最终网络性能,而残差网络使用跳跃连接将每个卷积层的输入与输出相加,卷积层只需要学习不同于输入特征的新的信息,并且可以使深层网络结构中多余的冗余层实现输入与输出恒等映射,有效提高了卷积计算与梯度反馈的速度。

残差网络的特征很好地符合了人脸图像超分辨率重建的需求,作为网络输入的低分辨率人脸图像本身含有丰富的人脸特征信息,是重建高分辨率人脸图像的重要基础,若是将输入图像的特征经过每个卷积层完整地映射到网络的输出是十分低效的,因此本文研究使用全局残差与局部残差策略,让卷积只需要学习到输入图像本身特征之外的人脸信息,提高重建网络性能,本文设计的多尺度残差组总体结构如图 4 所示。

提升网络效能的方法除了增加网络深度,还可以增加网络的宽度,inception 模块[21]将不同的卷积层并行连接组合在一起,通过并行的不同卷积核提取特征信息,可以增加网络对不同尺度的适应性。为了降低网络计算复杂度,提高深层网络运行效率,在设计以上残差网络结构图中的残差单元时使用了一种与 inception 模块不同的多尺度残差策略,在传统的残差网络中,每个残差块仅通过简单的跳跃连接将输入直接加到输出上,这种设计虽然可以缓解梯度消失问题,但对于特征的多尺度变化处理不够灵活。而多尺度残差策略通过引入 SK 卷积,使得网络可以自适应地选择最合适的卷积核大小,从而更有效地捕捉到从细节到整体不同尺度的特征,这在一定程度上提高了人脸图像重建的精度和质量。所使用的 SK 残差单元结构如图 5 所示。

在使用自适应卷积从不同感受野提取特征信息的基础上,构造双路卷积结构,残差单元的每一条支路都使用 SK 自适应卷积作为卷积层,将输入结果进行逐元素叠加,连接一个核为 3×3 的卷积整合特征信息,通过跳连接与原始输入特征逐元素叠加,计算公式为

$$I_1 = f_{3\times 3}[f_{\mathrm{sk}_1}(I_x) + f_{\mathrm{sk}_2}(I_x)] \tag{7}$$

式(7)中:$f_{\mathrm{sk}}(\cdot)$ 为 SK 卷积过程;$f_{3\times 3}(\cdot)$ 为3×3 卷积过程;I_1 为深层特征提取模块输出;I_x 为输入。

将 16 个残差单元依次连接,将最后一个残差单元的输出通过跳连接与浅层特征提取模块的输出结果进行逐元素叠加,作为深层特征提取模块的输出。

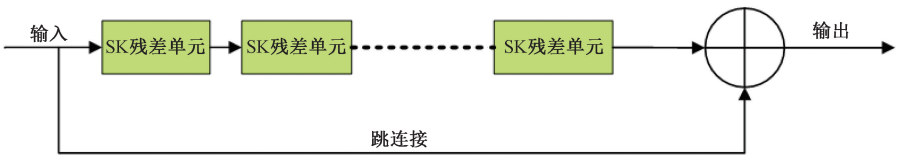


图 4　多尺度残差组总体结构

Fig. 4　Overall structure of multi-scale residual groups

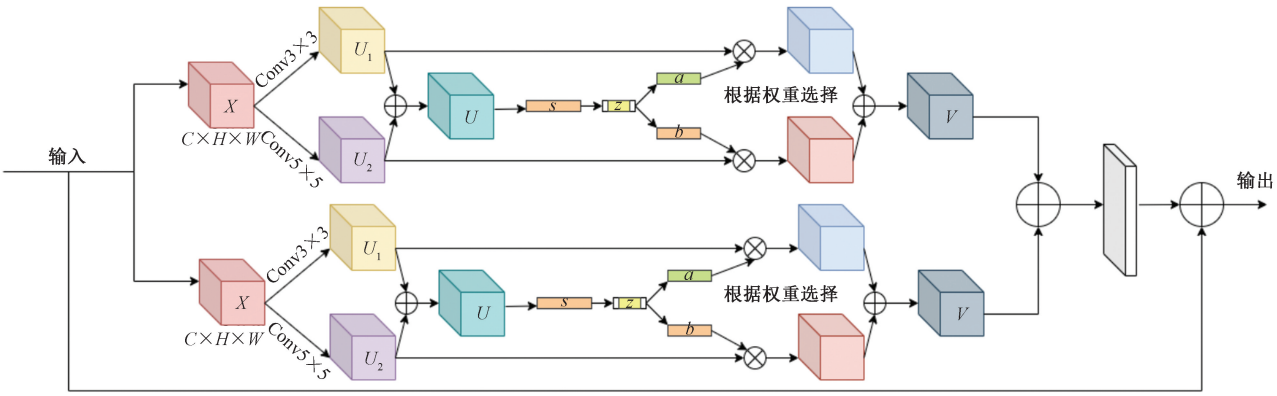


图 5　SK 残差单元

Fig. 5　SK residual unit

1.2.3 上采样模块

本文研究主要使用两个相邻的放大系数为2的子像素卷积层[22]作为上采样模块,实现对低分辨率人脸图像4倍的超分辨率重建。在第一个卷积层中,将残差组模块输出的特征 $E_1 \in \mathbf{R}^{C\times H\times W}$ 用3×3卷积进行通道扩充,得到特征图 $E_2 \in \mathbf{R}^{4C\times H\times W}$,使用PixelShuffle层将 E_2 重组得到尺寸放大2倍的特征图 $E_3 \in \mathbf{R}^{C\times H\times W}$,然后在第二个子像素卷积层中将 E_3 作为输入重复上述操作,得到尺寸放大4倍的特征图 $E_4 \in \mathbf{R}^{C\times H\times W}$,使用核为9×9的卷积层重建为高分辨率人脸图像,表示为

$$I_2 = f_{9\times 9}[\mu(f_{3\times 3}\{\mu[f_{3\times 3}(I_o)]\}] \tag{8}$$

式(8)中:$f_{3\times 3}(\cdot)$ 为3×3卷积计算过程;$\mu(\cdot)$ 代表PixelShuffle层重组过程;$f_{9\times 9}(\cdot)$ 为9×9卷积计算过程;I_2 为上采样模块输出;I_o 为上采样模块输入。

1.3 双判别器结构

SRGAN算法提出的Vgg网络使用深层卷积网络结构提取输入图像的特征信息,Real-Srgan[23]提出使用U-net结构判别器,U-net网络在图像分割领域具有较好的表现,下采样与上采样的结构以及特征融合的特性可以对每一个像素点进行判断类别,能更关注到人脸图像的局部细节信息。Vgg-19网络拥有19层的深层网络结构,能够捕捉图像的深层特征,可以更好地重建高频细节。U-net通过跳跃连接实现了不同层级特征的有效融合,这对于保持图像的空间信息十分有利。本文研究结合两者优点,使用Vgg架构网络 D_{Vgg} 与U-net架构网络 $D_{U\text{-}net}$ 作为模型的双判别器。结合VGG-19的深层特征提取能力和U-net的高效特征融合能力,实现优点互补,生成图像需要同时欺骗两个判别器,增加了生成模型的鲁棒性。但由于判别器的增加,所以训练复杂度也增加。D_{Vgg} 结构如图4所示,包含1个3×3卷积、6个连续的特征提取块,每个特征提取块包含卷积层、批归一化层与Leaky-Relu激活函数,使用全连接层与sigmold层输出判别结果。

$D_{U\text{-}net}$ 网络结构如图7所示,在下采样模块中使用3层卷积将初始层的特征下采样3次,每次对特征缩小尺寸、增加通道数,得到各低层特征。在上采样模块中进行相反的操作,增加尺寸、减少通道数,得到新的各层特征,将相同层次的特征相结合作为之后卷积层的输入,上采样层之后使用3个卷积层调整通道数输出判别结果。

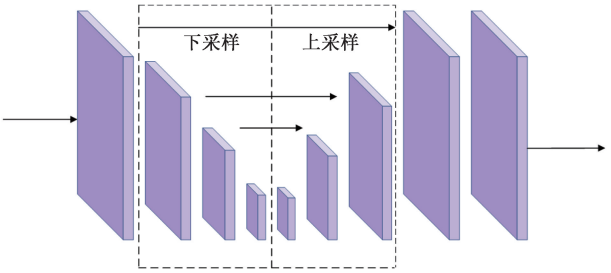


图7 判别网络 $D_{U\text{-}net}$ 结构

Fig. 7 The structure of discriminator network $D_{U\text{-}net}$

1.4 联合损失函数

联合损失函数包含两个方面,在模型训练的第一阶段,仅使用以均方误差(mean square error, MSE)为指导的损失函数,公式为

$$l_1 = l_{MSE} = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}[I_{x,y}^{HR} - G(I^{LR})_{x,y}]^2 \tag{9}$$

式(9)中:I^{LR} 为输入;l_{MSE} 为联合损失函数输出;$G(I^{LR})$ 表示生成图像;$I_{x,y}^{HR}$ 为输入高分辨率图像在(x, y)点的值;W、H 为输入图像的宽和高。

在模型训练的第二个阶段,额外融合以Vgg-19网络为基础的Vgg损失和以双判别网络为基础的对抗损失,总的损失函数公式为

$$l_2 = \alpha l_1 + \beta l_{Vgg} + \lambda_{D_1} l_{gen}^{D_1} + \lambda_{D_2} l_{gen}^{D_2} \tag{10}$$

式(10)中:l_1 为式(9)所代表的MSE损失;l_{Vgg} 为计算SR与HR图像深层特征图之间逐像素损失得到的感知损失;α 为MSE损失的加权系数;β 为Vgg损失的加权系数;λ_{D_1} 为判别网络 D_{Vgg} 对抗损失的加权系数;λ_{D_2} 为判别网络 $D_{U\text{-}net}$ 对抗损失的加权系数。

以往算法大多只使用Vgg-19网络单一层的输出作为图像深层特征,但是网络的中间层同样包含着重要的特征信息,本文研究融合了多层输出结果作为最终的图像深层特征并计算损失函数,表达

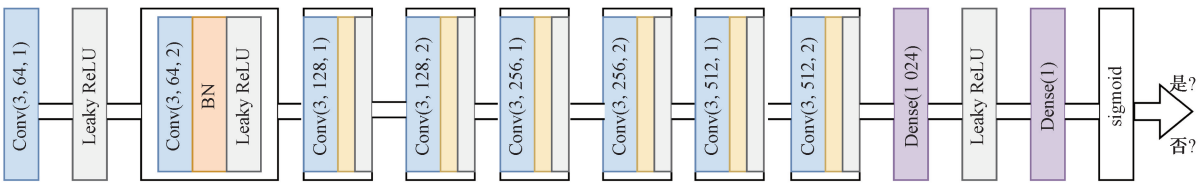


图6 判别网络 D_{Vgg} 结构

Fig. 6 The structure of discriminator network D_{Vgg}

式为

$$P = 0.1P_2 + 0.1P_7 + 0.1P_{16} + P_{19} \tag{11}$$

$$l_{\mathrm{Vgg}} = \frac{1}{W_i H_i}\sum_{x=1}^{W_i}\sum_{y=1}^{H_i}(P_{x,y}^{\mathrm{HR}} - P_{x,y}^{\mathrm{SR}})^2 \tag{12}$$

式中:P 为 Vgg-19 网络加权平均后的输出结果;P_i 为 Vgg-19 网络第 i 层的输出结果;W_i、H_i 为第 i 层特征图的宽和高;$P_{x,y}$ 为式(11)的最终深层特征输出结果在(x,y)点的值;$l_{\mathrm{gen}}^{D_1}$ 与 $l_{\mathrm{gen}}^{D_2}$ 表示判别网络 $\mathrm{D_{Vgg}}$ 与 $\mathrm{D_{U\text{-}net}}$ 计算的对抗损失,为了增强训练的稳定性,本文使用了相对平均判别(relative average discrimination, RaD[24]),公式为

$$D_{\mathrm{Ra}}(x_{\mathrm{r}},x_{\mathrm{f}}) = \sigma\{C(x_{\mathrm{r}}) - E_{x_{\mathrm{f}}}[C(x_{\mathrm{f}})]\} \tag{13}$$

$$l_{\mathrm{gen}}^{D_1} = -E_x\{\ln[1 - D_{\mathrm{Ra}}^1(x_{\mathrm{r}},x_{\mathrm{f}})]\} - E_x\{\ln[1 - D_{\mathrm{Ra}}^1(x_{\mathrm{f}},x_x)]\} \tag{14}$$

$$l_{\mathrm{gen}}^{D_2} = -E_x\{\ln[1 - D_{\mathrm{Ra}}^2(x_{\mathrm{r}},x_{\mathrm{f}})]\} - E_x[\ln(1 - D_{\mathrm{Ra}}^2(x_{\mathrm{f}},x_x))]\} \tag{15}$$

式中:$D_{\mathrm{Ra}}(x_{\mathrm{r}},x_{\mathrm{f}})$ 为数据 x_{r} 比数据 x_{f} 真实的概率;x_{r}、x_{f} 为 HR 图像数据与 SR 图像数据;σ 为 sigmoid 函数;$C(\cdot)$ 为判别器的原始输出;$E_x(\cdot)$ 表示求平均值。

为了平衡人脸图像真实感与失真之间的关系,本文算法与主流算法选取相同的损失加权系数,即 $\alpha = 1,\beta = 0.006,\lambda_{D_1} = 0.0015,\lambda_{D_2} = 0.0015$。

联合损失函数组合使得本文模型能够在保持像素级精度的同时,提升重建图像的感知质量和自然度。通过这种多方面的损失函数设计,模型不仅关注于像素级重建的准确性,还强调了重建图像的视觉效果,有效地平衡了重建过程中的细节保留和真实感提升。

然而,尽管该联合损失函数在多个方面考虑了重建质量,但对于特定的人脸图像超分辨率任务,可能还需要进一步的定制或优化。例如,考虑到人脸图像的特殊性,可能会加入特定的面部特征损失,如面部关键点或面部区域的特殊处理,以进一步提升面部图像的重建质量和自然度。

2　实验及结果分析

2.1　数据集与训练参数

本次实验采用 Pytorch 框架,在 Python3.7、CUDA 11.1、cuDNN 8.0.5 的虚拟环境中进行。本次实验的训练数据集使用 celeba 数据集的前 20 000 张人脸图像,测试数据集使用 celeba 数据集中后 500 张人脸图像。

训练时选用 Adam 优化器,重建倍数为 4 倍,设置每批次人脸图像数量为 32,为了提高网络训练效率,将原图随机裁出 96×96 的图像块作为 HR 人脸图像,并使用插值法将其 4 倍下采样作为生成网络输入的 LR 人脸图像。首先使用损失函数对生成网络进行学习率为 10^{-4}、周期为 100 批次的训练,再加入双判别器,使用 损失函数对生成网络以 10^{-4} 与 10^{-5} 的学习率分别进行 50 个周期的训练。

2.2　重建图像质量评价指标

本文研究共使用 4 种指标来评估重建人脸图像的质量。峰值信噪比 PSNR 从像素角度评价图像间的相似度,公式为

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I(i,j) - K(i,j)]^2 \tag{16}$$

$$\mathrm{PSNR} = 10\lg\frac{\mathrm{MAX}^2}{\mathrm{MSE}} \tag{17}$$

式中:MSE 是均方误差;I 和 K 分别为 HR 图像和 SR 图像在 (i,j) 点处的值;M 和 N 为图像长和宽;MAX=255,表示图像像素最大值。

结构相似度 SSIM 指标从结构角度衡量两个图像的相似性,计算公式为

$$S(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{18}$$

式(18)中:μ_x 和 μ_y 为两个图像的平均值;μ_x^2 和 μ_y^2 为两个图像的标准差;σ_{xy} 为两图像之间的协方差;C_1、C_2 为常数。

感知相似度 LPIPS 指标[25]使用特定神经网络,提取原始高分辨率人脸图像与生成图像的特征并计算两者的特征感知距离,计算公式为

$$d(x,x_0) = \sum_{l=1}^{19}\frac{1}{H_l W_l}\sum_{h=1,w=1}^{255} w_l \odot (\hat{y}_{hw}^l - \hat{y}_{0hw}^l)_2^2 \tag{19}$$

式(19)中:$\hat{y}^l,\hat{y}_0^l \in \mathbf{R}^{H_l\times W_l\times C_l}$ 表示在网络第 l 层提取的特征值;w 为宽度通道向量值;h 为高度通道向量值,最大值为 255;l 为输入的特征图层数,最大值为 19。

图像感知质量 PI 指标[26]用来衡量生成图像的主观感知质量,PI 值越低表示生成图像越符合人的主观感受,由量化指标 Ma[27]与 NIQE[28]计算得来,公式为

$$\mathrm{PI} = \frac{1}{2}[(10 - \mathrm{Ma}) + \mathrm{NIQE}] \tag{20}$$

2.3　结果分析

本节主要研究分析以上所提出的基于自适应卷积与联合损失函数的人脸图像超分辨率重建算法对低分辨率人脸重建的效果,使用从 celeba 数据集中选取的测试集进行部分对比实验,本文研究对生成对抗网络结构提出了 3 个方面的改进方法,将分别去除各改进模块的算法模型作为 Base 基础模

型,对使用改进模块前后的模型进行实验,分析各模块对于网络性能的不同影响,并将本文算法与主流算法进行重建结果比较分析。

首先研究自适应卷积对网络性能的影响,自适应卷积内部具有多个分支,当设置自适应卷积分支数为 1 时,就成为了普通的核为 3 × 3 的卷积单元[29],对自适应卷积的分支数进行对比实验,分析不同分支数的重建结果,以此确定在本算法中最优的分支数,重建图像指标评价结果如表 1 所示,其中 Base1 模型代表使用普通卷积的算法,Base1 + sk(2)模型表示使用分支数为 2 的自适应卷积,Base1 + sk(3)模型表示使用分支数为 3 的自适应卷积。

从表 1 中可以看出自适应卷积对算法结果的影响,使用含有两条卷积分支的自适应卷积相比使用普通卷积的算法重建结果在 4 个指标上均有优势,PSNR 值提高 1.455 dB,SSIM 值提高 0.029,LPIPS 值优化 0.048 1,PI 值优化 0.225 6,这说明相较于单核的卷积层,融合多核的卷积可以更有效地提取输入图像的特征信息,生成 SR 图像不仅在像素角度更接近原始 HR 图像,而且更符合 HR 图像的深层特征分布[30],能有效提高人眼感知质量。当分支数增加到 3 时,可以看到对于算法的作用很微弱,在 4 个指标上分别上下波动,说明在本算法中,融合核为 3 ×3 与 5 ×5 的卷积在感受野方面足够提取到输入图像的特征信息,额外融合核为 7 ×7 的卷积效果不明显[31]。本文研究对于生成对抗网络的训练采用了两阶段训练法,图 8 分别展示了在两个阶段训练过程中 PSNR 指标的变化折线图。

在仅训练生成器网络阶段,使用了自适应卷积的算法随着训练批次的增加,重建图像的 PSNR 指标上升速度明显快于普通卷积[32]的算法,这表明网络收敛更快,并且较早达到最优指标附近,在同时训练生成网络与判别网络阶段,使用自适应卷积的算法同样表现出收敛更快的性能,体现出自适应卷积能更高效的提取输入图像特征,不同感受野下的卷积层能学习到更多的人脸特征信息,使得指标折线图一直高于使用单一卷积的算法。由于自适应卷积分支数的继续增加并未明显提高网络性能,考虑到卷积复杂度对算法的负担,本文研究使用分支数为 2 的 SK 卷积作为最优模型的组成单元。依据车亚丽等[33]研究结果显示,自适应卷积在 Celeba 和 Helen 两种人脸数据集上都产生了较好的实验结果,对于无遮挡人脸数据集具有一定的泛化能力。

其次研究多尺度残差模块对算法结果的影响,将使用多尺度残差模块的算法与使用普通残差模块的算法进行对比,表 2 与图 9 分别表示两种算法重建的人脸图像指标评价结果与训练过程中 PSNR 指标的变化折线图,其中 Base2 表示未使用多尺度残差模块的算法,可以看到使用多尺度残差模块改进的算法重建结果 PSNR 值提高 0.627 dB,SSIM 值提高 0.023,LPIPS 值优化 0.026 2,PI 值没有明显变化,这说明在残差组模块中融合多路径的特征能够补充单一路径所忽略的特征信息,生成的 SR 图像细节上更清晰。多支路的主干网络同样能使网络较快学习到重要的人脸信息,质量评价指标变化速度更快。

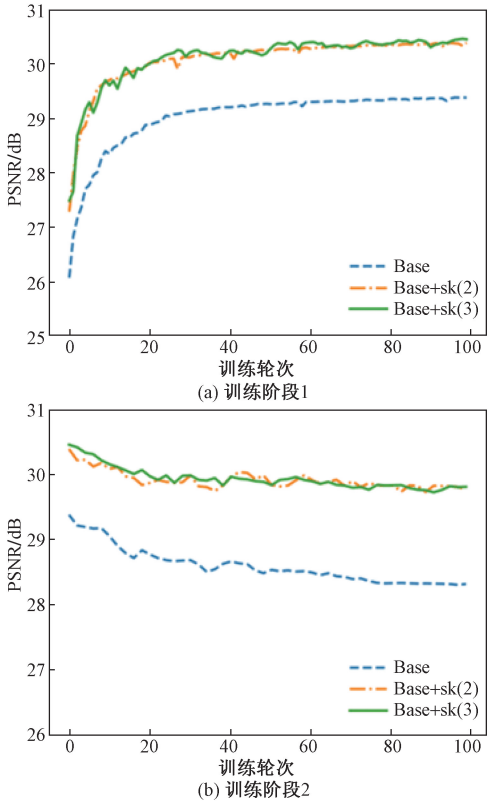


图 8 自适应卷积对 psnr 指标变化影响图

Fig. 8 Influence of adaptive convolution on psnr index change

表 1 自适应卷积对算法模型的影响

Table 1 Influence of adaptive convolution on algorithm model

模型	PSNR/dB ↑	SSIM ↑	LPIPS ↓	PI ↓
Base1	28.314	0.831	0.150 8	3.925 4
Base1 + sk(2)	29.769	0.860	0.102 7	3.699 8
Base1 + sk(3)	29.804	0.861	0.110 3	3.674 8

注:↑代表数值高为优;↓代表数值低为优。

表 2 多尺度残差模块对模型的影响

Table 2 Influence of multi-scale residual module on the model

模型	PSNR/dB ↑	SSIM ↑	LPIPS ↓	PI ↓
Base2	29.142	0.837	0.128 9	3.846 6
Base2 + sk multiscale residual	29.769	0.860	0.102 7	3.699 8

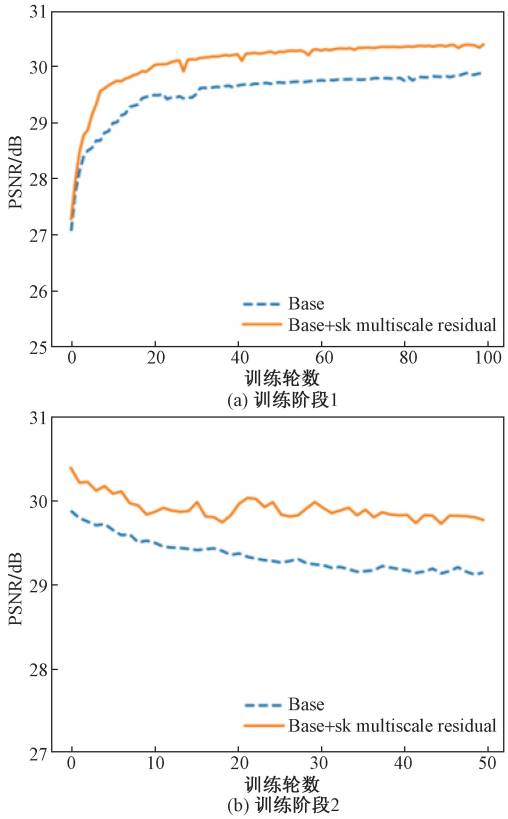


图 9　多尺度残差对 PSNR 指标变化影响图

Fig. 9　Impact of multi-scale residuals on the change of PSNR index

将使用双判别网络的算法与仅使用 Vgg 判别网络的算法进行对比研究,表 3 与图 10 分别表示两种算法重建的人脸图像指标评价结果与训练过程中 PSNR 指标的变化折线图,改进的算法相比单判别网络 PSNR 值没有明显变化,SSIM 值提高 0.011,PI 值优化 0.018 8,LPIPS 值优化 0.093 8,PSNR 的变化折线图并未始终与单判别的算法接近,这说明加入 U-net 网络的判别反馈之后,优化的损失函数主要针对对抗损失方面,主要能提高图像的感知质量。

为了验证本文算法重建人脸图像的优越性,与当前主流算法在测试集上进行对比实验。本文算法与 SRCNN、VDSR、EDSR、SRGAN、MLGE、RWSA 算法重建图像的质量指标评价结果如表 4 所示。其中 SRGAN、MLGE、RWSA 算法与本文算法一样采用生成对抗网络结构。本文算法运行时间优势不明显,但可以看到 EDSR 等缺少感知损失指导的算法在 PSNR 与 SSIM 指标上具有优势,本文算法因使用了生成对抗网络的架构与联合损失函数,在 LPIPS 和 PI 两个衡量图像感知质量的指标上均超过了其他算法,LPIPS 指标比第二名优化了 0.032,PI 指标比第二名优化了 0.119,相比于同样使用生成对抗网络结构的算法,本文算法具有更有效的特征提取模

表 3　双判别网络对模型的影响

Table 3　The influence of double discriminant network on the model

模型	PSNR/dB ↑	SSIM ↑	LPIPS ↓	PI ↓
Base3	29.651	0.849	0.121 5	3.993 6
Base3 + U-net	29.769	0.860	0.102 7	3.699 8

注:↑代表数值高为优;↓代表数值低为优。

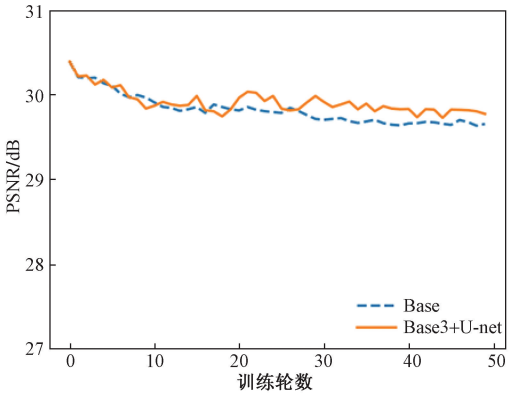


图 10　双判别网络对 PSNR 指标变化影响图

Fig. 10　Influence diagram of double discrimination network on PSNR index change

表 4　不同算法指标评价结果

Table 4　Index evaluation results of different algorithms

模型	PSNR/dB ↑	SSIM ↑	LPIPS ↓	PI ↓
Bicubic	27.977	0.811	0.330 0	7.619 5
SRCNN[1]	29.456	0.865	0.232 0	6.138 2
VDSR[3]	29.937	0.870	0.274 0	5.981 7
EDSR[12]	30.252	0.879	0.268 0	5.630 3
SRGAN[10]	27.968	0.814	0.161 0	4.031 3
MLGE[11]	28.381	0.815	0.136 0	4.019 0
RWSA[14]	28.603	0.823	0.135 0	3.818 0
本文模型	29.769	0.860	0.102 7	3.699 8

注:↑代表数值高为优;↓代表数值低为优。

块,在 PSNR 与 SSIM 指标上同样具有很大优势,证明了本文算法重建人脸图像的优越性。

为了更直观地看出算法重建的人脸图像的效果,上述算法各自重建的部分人脸图像整体对比图如图 11 所示,人脸图像细节对比图如图 12 所示。

从上述两张图中可以看出 bicubic 使用插值法图像质量最差,只能恢复脸部的大致轮廓,脸部细节模糊。SRCNN、VDSR、EDSR 随着卷积网络复杂度的提高所生成图像逐渐清晰,SRCNN 算法重建效果得到了增强,但由于网络层数较少,对特征信息提取有限,所以在细节信息上仍有损失。VDSR 对比 SRCNN 具有深度网络结构,并加入残差结构,但由于缺少联合损失约束,生成图像细节仍有提升的空间。SRGAN、VDSR、EDSR 三种算法生成效果得到增强,细节刻画提升较明显,但是由于部分纹理

图 11 不同模型重建人脸图像整体对比图

Fig. 11 Overall comparison of facial images reconstructed from different models

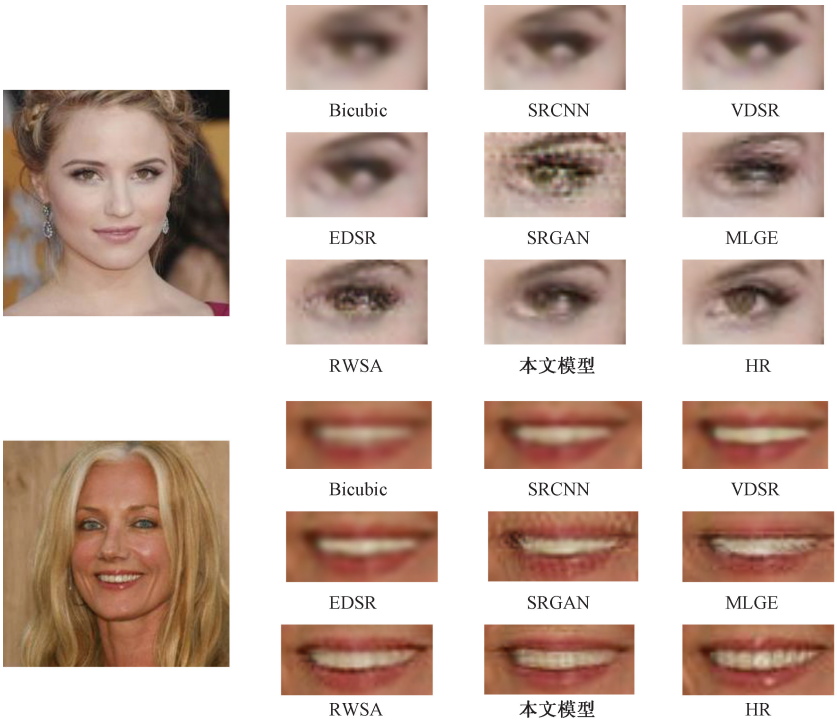


图 12 不同模型重建人脸图像细节对比图

Fig. 12 Comparison of facial image details reconstructed by different models

细节过于光滑,颜色对比度较低使得人脸图像增加了过多的失真,降低了人脸图像的真实感。相比较而言,本文算法重建的人脸图像具有较为清晰和真实的主观感受,使用 Vgg 架构网络作为判别器捕捉了人脸图像的全局特征,如整体结构、轮廓和主要组成部分的布局。这有助于确保重建的人脸图像在整体上与原始高分辨率图像保持一致性。U-net 架构网络作为另一个判别器,专注于图像的局部细节和纹理信息,如眼睛、鼻子和嘴巴的细节。这有助于提升图像细节的重建质量,使重建的人脸图像在眼睛等细节方面和 HR 图像更为相似。但是本文算法仅实现了倍数为 4 的图像重建工作,为了满足不同情况下的需求,进一步进行不同倍数的重建训练。

3 结论

(1) 算法采用的自适应卷积可以提高网络模型的精度,相较于单核的卷积层能更有效地提取输入图像的特征信息,生成的图像更符合真实图像的深层特征分布,能有效提高人眼感知质量,PSNR 值提高 1.455 dB, SSIM 值提高 0.029, LPIPS 值优化 0.048 1,PI 值优化 0.225 6。

(2)算法采用的多尺度残差结构能提高网络特征映射能力,融合多路径的特征能够补充单一路径

所忽略的特征信息,生成的图像细节上更清晰,PSNR 值提高 0.627 dB,SSIM 值提高 0.023,LPIPS 值优化 0.026 2。

(3)算法采用的双判别网络结构可以提高判别器判别效果,通过优化对抗损失可以提高图像的感知质量,SSIM 值提高 0.011,PI 值优化 0.018 8,LPIPS 值优化 0.093 8。

(4)提出了一种基于生成对抗网络的人脸超分辨率重建算法。一方面采用自适应卷积,根据输入特征调节内部卷积核的权重,提高生成图像的人眼感受质量;另一方面通过改进的多尺度残差结构,增加网络宽度,提高生成图像的清晰度。实验结果表明,本文方法对于人脸图像超分辨率重建上优于其他超分辨率算法,有效提高了现实环境中人脸超分辨率重建的质量。然而由于加入的双判别器网络结构增加了模型复杂度,所以该模型时间优势不明显,如何在保证人脸重建质量的同时使模型轻量化是之后尝试的改进方向。

参 考 文 献

[1] Dong C, Loy C, Tang X O. Accelerating the super-resolution convolutional neural network[C]//European Conference on Computer Vision. Cham: Springer, 2016: 391-407.

[2] Kim J, Lee J, Lee K. Deeply-recursive convolutional network for image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1637-1645.

[3] Kim J, Lee J, Lee K. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1646-1654.

[4] Tai Y, Yang J, Liu X M. Image super-resolution via deep recursive residual network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 3147-3155.

[5] Tong T, Li G, Liu X T, et al. Image super-resolution using dense skip connections[C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 4799-4807.

[6] 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 (ECCV). Berlin: Springer, 2018: 286-301.

[7] Woo S, Park J, Lee J, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 3-19.

[8] Chudasama V, Nighania K, Upla K, et al. E-comsupresnet: enhanced face super-resolution through compact network[J]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2021, 3(2): 166-179.

[9] Chen C F, Gong D H, Wang H, et al. Learning spatial attention for face super-resolution[J]. IEEE Transactions on Image Processing, 2020, 30: 1219-1231.

[10] Ledig C, Theis L, Huszár 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. Piscataway: IEEE, 2017: 4681-4690.

[11] Ko S, Dai B R. Multi-laplacian GAN with edge enhancement for face super resolution[C]// 25th International Conference on Pattern Recognition (ICPR). New York: IEEE, 2021: 3505-3512.

[12] 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. Piscataway: IEEE, 2017: 136-144.

[13] 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 (ECCV) Workshops. Berlin: Springer, 2018: 1-9.

[14] Aakerberg A, Nasrollahi K, Moeslund T B. Real-world super-resolution of face-images from surveillance cameras[J]. IET Image Processing, 2022, 16(2): 442-452.

[15] 邓酩, 柳庆龙, 侯立宪. 多尺度残差生成对抗网络的图像超分辨率重建[J]. 科学技术与工程, 2023, 23(31): 13472-13481.
Deng Ming, Liu Qinglong, Hou Lixian. Image super-resolution reconstruction of multi-scale residual generative adversarial networks[J]. Science Technology and Engineering, 2023, 23(31): 13472-13481.

[16] 仝卫国, 蔡猛, 庞雪纯, 等. 基于多尺度特征融合的超分辨率重建算法研究[J]. 科学技术与工程, 2022, 22(26): 11507-11514.
Tong Weiguo, Cai Meng, Pang Xuechun, et al. Super-resolution reconstruction algorithm based on multi-scale feature fusion[J]. Science Technology and Engineering, 2022, 22(26): 11507-11514.

[17] 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. Piscataway: IEEE, 2015: 3730-3738.

[18] Hahnloser R H R, Sarpeshkar R, Mahowald M A, et al. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit[J]. Nature, 2000, 405(6789): 947-951.

[19] Li X, Wang W H, Hu X L, et al. Selective Kernel networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 510-519.

[20] Yu F. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv: 1511.07122, 2015.

[21] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 1-9.

[22] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1874-1883.

[23] Wang X, Xie L, Dong C, et al. Real-esrgan: training real-world blind super-resolution with pure synthetic data[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 1905-1914.

[24] Jolicoeur-Martineau A. The relativistic discriminator: a key element missing from standard GAN[J]. arXiv preprint arXiv:

1807.00734, 2018.

[25] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 586-595.

[26] Blau Y, Mechrez R, Timofte R, et al. The 2018 PIRM challenge on perceptual image super-resolution[C]//Proceedings of the European Conference on Computer Vision (ECCV) Workshops. Piscataway: IEEE, 2018. DOI: 10.48550/arXiv. 1809.07517.

[27] Ma C, Yang C Y, Yang X K, et al. Learning a no-reference quality metric for single-image super-resolution[J]. Computer Vision and Image Understanding, 2017, 158: 1-16.

[28] Mittal A, Soundararajan R, Bovik A C. Making a "completely blind" image quality analyzer[J]. IEEE Signal Processing Letters, 2012, 20(3): 209-212.

[29] 章伟帆, 曾庆鹏. 多重放大的医学图像超分辨率重建[J]. 计算机工程与应用, 2022, 58(23): 230-237.
Zhang Weifan, Zeng Qingpeng. Multi-scale medical image super-resolution reconstruction[J]. Computer Engineering and Applications, 2022, 58(23): 230-237.

[30] 李靖宇, 程卫月, 李子翔, 等. 超分辨率重建的微小人脸识别算法[J]. 哈尔滨理工大学学报, 2022, 27(3): 52-58.
Li Jingyu, Cheng Weiyue, Li Zixiang, et al. Small face recognition algorithm based on super-resolution reconstruction[J]. Journal of Harbin University of Science and Technology, 2022, 27(3): 52-58.

[31] 张艳青, 马建红, 韩颖, 等. 真实场景下图像超分辨率重建研究综述[J]. 计算机工程与应用, 2023, 59(8): 28-40.
Zhang Yanqing, Ma Jianhong, Han Ying, et al. Review of research on real-world single image super-resolution reconstruction[J]. Computer Engineering and Applications, 2023, 59(8): 28-40.

[32] 吕佳, 许鹏程. 多尺度自适应上采样的图像超分辨率重建算法[J]. 计算机科学与探索, 2023, 17(4): 879-891.
Lyu Jia, Xu Pengcheng. Image super-resolution reconstruction algorithm based on multi-scale adaptive upsampling[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(4): 879-891.

[33] 车亚丽, 徐岩, 薛海丽, 等. 基于残差注意力机制的人脸图像超分辨率重建算法[J]. 测量科学与仪器, 2023, 13(9): 1-9.
Che Yali, Xu Yan, Xue Haili, et al. Face image super-resolution reconstruction algorithm based on residual attention mechanism[J]. Journal of Measurement Science and Instrumentation, 2023, 13(9): 1-9.