

融合注意力的生成式对抗网络单图像超分辨率重建

彭晏飞^{**}, 张平甲^{*}, 高艺, 訾玲玲

辽宁工程技术大学电子与信息工程学院, 辽宁 葫芦岛 125105

摘要 基于深度学习的单图像超分辨率重建方法已经比较完善, 重建图像具有较高的客观评价价值或具有较好的视觉效果, 但是图像感知效果和客观评价价值不能均衡提升。针对这一问题, 提出一种融合注意力的生成式对抗网络单图像超分辨率重建方法。首先去掉残差网络中会破坏图像原本的对比度信息、影响图像生成质量的批归一层, 其次是构造注意力卷积神经网络残差块, 可有效地在特征映射中进行自适应特征细化, 改善重建结果在大尺度因子下缺乏高频信息和纹理细节的问题, 最后构造像素损失函数, 使用鲁棒性较好的 Charbonnier 损失函数替代均方差损失函数, 用总变差正则项平滑训练结果。实验结果表明, 在 4 倍放大因子下, 与其他方法在 Set5、Set14、Urban100、BSDS100 测试集上进行测试比较, 本文方法峰值信噪比平均值提升 2.88 dB, 结构相似性平均值提升 0.078。实验数据和效果图表明, 该方法主观上具有丰富的细节, 客观上具有较高的峰值信噪比值和结构相似性值, 实现了视觉效果和客观评价指标值的均衡提升。

关键词 图像处理; 超分辨率; 残差学习; 卷积神经网络; 注意力; 生成式对抗网络

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202158.2010012

Attention Fusion Generative Adversarial Network for Single-Image Super-Resolution Reconstruction

Peng Yanfei^{**}, Zhang Pingjia^{*}, Gao Yi, Zi Lingling

School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China

Abstract Deep learning-based single-image super-resolution reconstruction method has been relatively perfect. The reconstructed image has a high objective evaluation value or a good visual effect; however, the image perception effect and objective evaluation value cannot be improved in a balanced manner. To address this problem, this paper proposes a single-image super-resolution reconstruction method based on an attention fusion generative adversarial network. In the proposed method, first, the batch layer that destroys the original image contrast information and affects the quality of image generation in the residual network is removed. Then, the residual block of the attention convolutional neural network, which can effectively perform adaptive feature refinement in the feature map, is constructed. To improve the reconstruction results that lack high-frequency information and texture details under large-scale factors, a pixel-loss function is constructed to replace the mean squared error-loss function with a more robust Charbonnier loss function, and a total variation regular term is used to smooth the training results. The experimental results show that compared with other methods on the Set5, Set14, Urban100, and BSDS100 test sets under $4\times$ magnification factor, the average peak signal-to-noise ratio and average structure similarity increased by 2.88 dB and 0.078, respectively. The experimental data and renderings demonstrate that the proposed method is subjectively rich in details, objectively has a high peak signal-to-noise ratio and structural similarity value, and achieves a balanced improvement of visual effects and objective evaluation index values.

Key words image processing; super-resolution; residual learning; convolutional neural network; attention; generative adversarial network

OCIS codes 100.6640; 100.3010; 100.2000

收稿日期: 2020-11-25; 修回日期: 2020-12-21; 录用日期: 2021-01-06

基金项目: 国家自然科学基金(61702241, 61602226)、辽宁省教育厅高等学校基本科研项目(LJ2017FBL004)

通信作者: *1308192862@qq.com; **pengyf75@126.com

1 引言

近年来,在图像超分辨率领域,单图像超分辨率的关注度越来越高,如何重建出含有丰富细节的高分辨率图像并且有较高的客观判别值,成为图像处理中的研究热点。单幅图像超分辨率(SISR)是一个没有唯一解的病态求逆问题,通过对低分辨率(LR)图像提取关键信息用来生成相对应的高分辨率图像。超分辨率目前已成功应用于视频监控^[1]、医学图像处理^[2]、人脸识别^[3]、遥感影像^[4]等领域。

图像超分辨率是一个传统研究热点,传统研究方法常用基于插值^[5]、基于重建^[6]的方法,随着机器学习和深度学习的发展又产生了基于学习^[7]的方法,基于插值的重建方法通过使用相邻像素预测,得到的未知像素值插入到相应位置,重建速度较快,但缺点也很明显,放大倍数较大时,生成图像的高频细节缺失严重。基于重建的超分辨率方法通常需要多帧多图像,配合先验知识进行低分辨率图像到高分辨率图像的重建。该方法可通过先验信息对高频信息进行恢复,但对低分辨率图像要求较高,需要多张强相关图像。基于学习的方法通过对大量数据进行分析学习提取出关键参数,并在学习中不断优化,最终构建出从低分辨率到高分辨的映射模型,目前基于学习的方法主要有基于近邻嵌入、稀疏编码表示和基于深度学习,其中基于深度学习的卷积神经网络(CNN)算法重建出的图像质量较高,受应用场景限制较小,是目前最主流的超分辨率重建方法。

Dong 等^[8]首先采用卷积神经网络来实现超分辨率方法(SRCNN),该方法相比传统算法在精度有了很大的提升,但该网络中的网络层数较少,收敛太慢,使用卷积核较小,提取出的特征都是局部特征在纹理细节的恢复上难以奏效,导致重建效果较差。在 Dong 等的研究基础上 Shi 等^[9]通过增加亚像素卷积层,可高效、实时地将低分辨率图像转换成高分辨率图像。He 等^[10]提出残差网络(ResNet),解决了随着网络层数增加梯度消失的问题,通过简单的残差块堆叠,网络能达到很深的深度。Kim 等^[11]在 He 等的基础上提出了非常深的超分辨率网络(VDSR),使用了更加深层的 VGG 网络结构模型,达到了 20 层的网络深度,通过残差学习大幅提升了网络收敛速度。Lim 等^[12]提出增强深度超分辨率网络(EDSR),去除了残差块中的批归一化(BN)层,节约了约 40% 的内存使用量,可构建具有更好性能的更大模型。Lai 等^[13]提出了拉普拉斯金字塔超分

分辨率网络(LapSRN),用 Charbonnier 损失函数替换 L2 损失函数,对网络进行深度监督训练,并实现高质量的重构。Zhang 等^[14]提出一种残差通道注意力网络(RCAN),通过学习不同通道的重要性赋予其不同的权重,提出了 Residual in Residual(RIR)结构来构建深层的可训练网络,收敛效果远超只单纯堆叠残差块的神经网络。

受到生成对抗网(GAN)的启发,2017 年,Ledig 等^[15]提出了超分辨率生成对抗网络(SRGAN),首次在图像超分辨率领域中使用 GAN,通过完善一些细节上的真实感提升了图像的感知效果。2019 年, Wang 等^[16]提出增强型超分辨率生成对抗网络(ESRGAN),用 Dense Block 替换 Residual Block,并去掉 BN 层,虽然生成图像的峰值信噪比(PSNR)值不是很理想,但人眼感官效果大大提升。受其启发,本文在 SRGAN 的基础上,为了提高模型运算速度,去除了残差块中的 BN 层,其次在残差块中使用了卷积神经网络注意模块(CBAM)^[17]。使用通道注意力机制(CA),考虑特征通道之间的相互依赖性自适应地调整特征,提高网络的表达能力。并设计一种有关空间注意力机制(SA)的模型,可自动对图像中高频信息区域给予更多注意,增加网络中高频细节信息。在训练中,使用鲁棒性更好的 Charbonnier 损失函数替换掉原本的均方差(MSE)损失函数,并对生成的图像进行总变差(TV)正则化进行平滑处理,使图像的高频信息不会过于突兀。通过以上方法表明本文方法可均衡提升图像感知效果和客观评价价值。

2 本文方法

本文以 SRGAN 为基础,受注意力机制、Charbonnier 损失和单幅图像超分辨率的增强深度残差网络(EDSR)的启发,提出一种融入注意力机制的生成对抗网络实现图像超分辨率重建。所提方法分为三部分:1) 构造了注意力卷积神经网络残差块,能更好地自适应特征细化,增加图像的高频信息;2) 构造像素损失函数使用鲁棒性较好的 Charbonnier 损失函数,同时加入 TV 正则项用来稳定 GAN 的训练;3) 去除了会破坏了原始图像对比度信息、在回归任务并无太大作用的 BN 层,将 Relu 激活函数换成拟合效果更好的 PRelu 激活函数,合理地增加了网络深度,加入多个更新后的残差块以获得更多的参数,能得到更好的拟合效果。总体结构流程如图 1 和 2 所示。

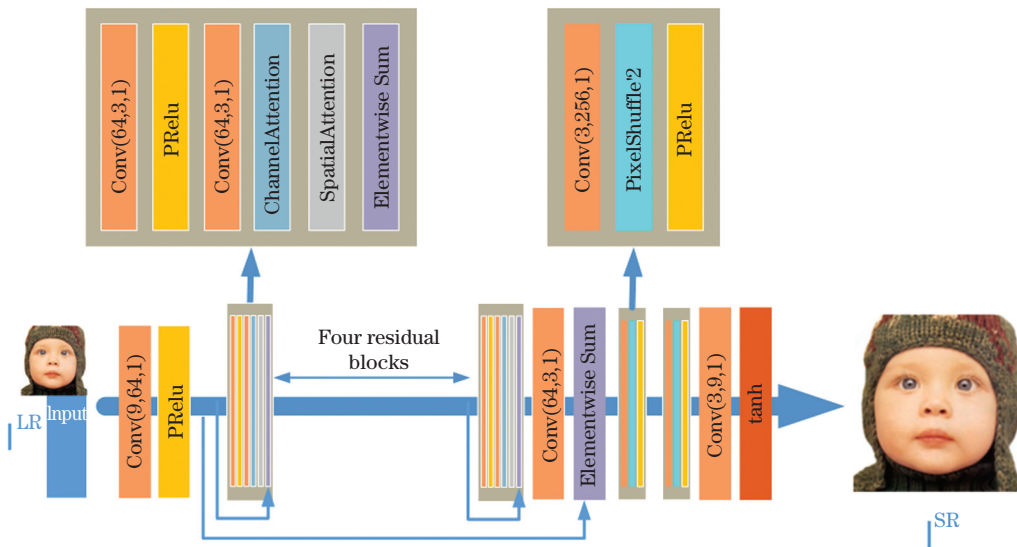


图 1 生成器网络结构

Fig. 1 Generator network structure

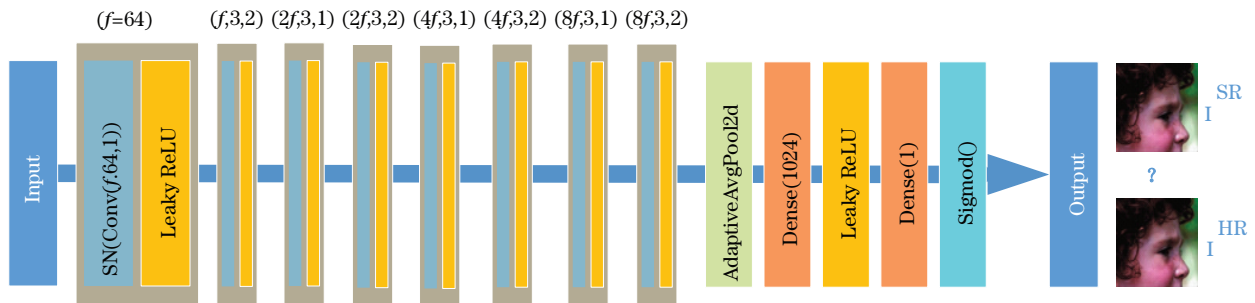


图 2 判别器网络结构

Fig. 2 Discriminator network structure

2.1 残差注意力块

注意力机制是模拟人类视觉特有的大脑信号处理机制设计的模型,通过快速扫描全局图像,获得需要重点关注的目标区域,而后对这一区域投入更多注意力资源,以获取更多需要关注目标的细节信息,而抑制其他无用或不重要的信息。残差注意力块的核心思想是在原本的残差块中加入注意力模块,用来学习图像中不同位置相应的空间注意力。为了在多个混合通道和空间信息中实现卷积运算提取信息特征,将注意力处理纳入到大规模分类任务中,能提

高 CNN 网络的性能。本文采用 CBAM 中的 SA 和 CA 来构建残差注意力块,该残差块通过强调沿着通道和空间轴这两个维度的有意义特征(图 3),使得每个分支可以分别学习到在通道和空间轴上哪些是重要信息以及它的所在位置,模块通过学习强调或抑制某些信息,增加网络中的信息流动性。Woo 等^[18]通过广泛的消融研究验证了注意力模块的有效性,可广泛应用于增强 CNN 的表达能。下面详细介绍通道注意力和空间注意力机制在模型中的作用。

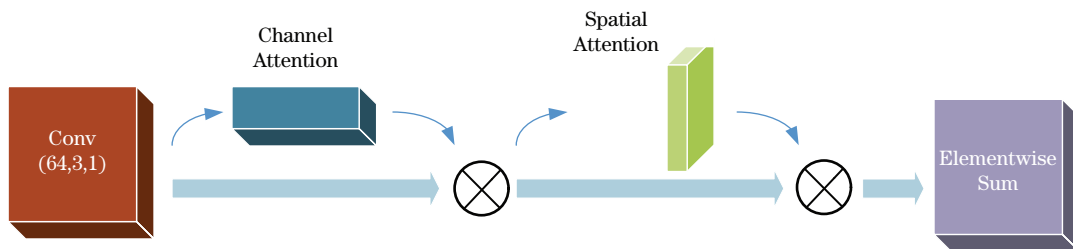


图 3 通道和空间注意力模块的使用

Fig. 3 Use of channel and spatial attention modules

2.1.1 通道注意力

低分辨率图像中一般只有很少的高频信息,大部分都是低频信息。在重建图像的过程中,平滑区域中多数都是低频信息比较容易重建,但高频信息的重要性往往在重建过程中被弱化。现在的深度学习学习得到的模型都有较高的客观评价,但是高频细节缺失,图像边缘锐化程度较低,使得人眼感官效果较差。在本文模型中通过压缩输入特征图的空间维数来计算通道注意力值,采用平均池化法进行空间信息的聚合。同时使用平均池化和最大池化的特性来提高网络的表示能力,有效地增强高频信息的拟合效果,生成细节更加逼真的重建图像。首先,使用平均池化层和最大池化层操作聚合特征映射的空间信息,生成平均集合特征 T_{avg}^c 和最大集合特征 T_{max}^c ,平均集合特征可对通道进行整体细节的提取,最大集合特征针对高频信息进行处理,可给予包含高频信息的通道更多的注意力。然后,这两个数据被转发到由一个隐藏层和多层感知器(MLP)组成的共享网络,生成通道注意映射 $W_c \in \mathbb{R}^{C \times 1 \times 1}$ 。根据上述描述,通道注意计算为

$$W_c(T) = \sigma \{ \text{MLP}[\text{AvgPool}(T)] + \text{MLP}[\text{MaxPool}(T)] \} = \sigma \{ \mathbf{N}_1 [\mathbf{N}_0(T_{\text{avg}}^c)] + \mathbf{N}_1 [\mathbf{N}_0(T_{\text{max}}^c)] \}, \quad (1)$$

式中: σ 为 sigmoid 函数; T 的大小为 $C \times H \times W$; \mathbf{N}_0 为 $C \times \frac{C}{r}$ 的矩阵; \mathbf{N}_1 为 $\frac{C}{r} \times C$ 的矩阵; r 为递减率; $W_c(T)$ 大小为 C ,即 T 的通道个数。

2.1.2 空间注意力

图像超分辨率重建的一个难点就是高频信息太少,并且深度网络难以定位其位置,不能有针对性地进行复原。在本文模型中利用特征间的空间关系生成空间注意图,低频区域残差值无限趋近于 0,高频区域的残差值要大于低频区域的残差值,对特征图沿着通道轴进行池化,能有效突出携带高频信息的区域。通过对通道轴的特征图信息进行平均池化和最大池化融合得到平均池化特征二维图 T_{avg}^s 和最大池化特征二维图 T_{max}^s 。将生成的两组二维图进行拼接,在通过标准 3×3 卷积层对他们之间和周围的特征信息进行融合,得到空间注意力图 A_{sa} ,再对其使用 sigmoid 函数激活便得到空间注意力值 $W_s(T)$,空间注意力的计算公式为

$$W_s(T) = \sigma \{ f^{3 \times 3} \{ [\text{AvgPool}(T); \text{MaxPool}(T)] \} \} = \sigma \{ f^{3 \times 3} [T_{\text{avg}}^s; T_{\text{max}}^s] \}, \quad (2)$$

式中: $f^{3 \times 3}$ 表示 3×3 的卷积。 T_{avg}^s 和 T_{max}^s 的计算公式为

$$T_{\text{avg}}^s(h, \omega) = \frac{\sum_{c=1}^C T_2^c(h, \omega)}{C}, \quad (3)$$

$$T_{\text{max}}^s(h, \omega) = \max_{c=\{1, \dots, C\}} T_2^c(h, \omega), \quad (4)$$

式中: (h, ω) 表示空间位置, $h \in \{1, \dots, H\}$, $\omega \in \{1, \dots, W\}$;第 c 个通道表示为 T_2^c ,其中 $c \in \{1, \dots, C\}$, C 为特征图的通道数。

2.2 损失函数构建

本文使用了内容损失和感知损失相结合的混合损失函数,同时,使用了特征图计算 Charbonnier 损失函数来生成像素损失函数,再加上 TV 损失来约束噪声,平滑生成图像。

本文使用空间注意力对图像高频信息进行重建,使用 Charbonnier 是为了保证图像低频信息能正确稳定的重建。MSE 损失函数在图像超分辨率重建中对于高频信息有较好的生成效果,同时能提高生成图像的 PSNR 值,但是也会在其他区域中生成许多影响观感的虚影和不规则的像素点。MSE 损失约束太强,对较大的损失惩罚力度较大,但是在较小的损失上无能为力,生成的图像有很低的损失值但过于平滑,同时它不适用于最后一层含 Sigmoid 激活函数的神经网络。L1 损失函数重建时间较长,但是它可对图像的高频信息恢复具有较强约束力,可有效地去除重建图像中的虚影和不规则的像素点。对比 L1 和 MSE 损失函数,本文使用鲁棒性更好的 Charbonnier 损失函数,兼顾了 L1 和 MSE 的优点,在花费更少的训练时间的同时,还能取得较高的 PSNR 值。Charbonnier 损失函数表示为

$$L_{\text{char}}(y, \hat{y}) = \epsilon E_{x, y \sim P_{\text{data}}(x, y)} \{ \sigma [y - G(x)] \}, \quad (5)$$

式中: β 是超参数; y 表示原始图像; $G(x)$ 为生成器生成的图像; $\sigma(x) = \sqrt{x^2 + \epsilon^2}$, $\epsilon = 10^{-7}$ 。

在图像的超分辨率重建过程中,图像中的噪声会对生成图像产生非常大的影响,加入一些正则项可以保持图像的光滑性,消除图像复原可能带来的伪影。于是本文加入了 TV 正则项,其中 δ 为超参数。

$$L_{\text{TV}} = \delta \| \hat{y} \|_{\text{TV}}. \quad (6)$$

文中使用的判别器网络损失函数共包括四个部分:Charbonnier 损失函数 L_{char} 、感知损失函数 L_{per} 、对抗损失函数 L_{adv} 、规则项 L_{TV} 。整体的损失函数沿用了 SRGAN 的损失超参数 $\alpha = 6 \times 10^{-3}$, $\beta =$

10^{-3} , Charbonnier 损失函数替换了原本的 MSE 损失函数, L_{TV} 是为了搭配感知损失函数, 处理重建过程中产生的伪影, 系数的选定如图 4 所示, 当 L_{TV} 的系数过大时, 与原图相比具有较大的色差, 在 $\epsilon =$

$2 \times e^{-8}$ 时和原图的色差最小, 在 $\epsilon = 2 \times e^{-9}$ 时, 图片中会有少许的斑点产生, 最终采用 $\epsilon = 2 \times e^{-8}$ 。

$$L = L_{char} + \alpha L_{per} + \beta L_{adv} + \epsilon L_{TV}, \quad (7)$$

式中: $\alpha = 6 \times 10^{-3}, \beta = 10^{-3}, \epsilon = 2 \times e^{-8}$ 。



图 4 ϵ 系数取不同值重建效果图

Fig. 4 Reconstruction effects with different values of ϵ coefficient

2.3 归一化处理

BN 层常用于图像分割、目标检测等领域, 对大部分任务都能加速收敛, 还可实现正则化, 有效防止图像过拟合, 对于一些不需要太细致的任务, 能找出关键特征, 就有较好的效果, 则 BN 就能发挥积极的作用, 超分辨率重建任务基本都是将小图像通过像素插值, 变成大图像, 并没有色彩、明暗度、对比度的变化, 如果在超分辨率重建任务中使用这种像素级别的任务, 图像的色彩分布都会被归一化, 一些图像特有的细节往往会被 BN 层给忽略, 不利于图像生成, BN 层的存在还会增加网络的复杂度, 去除 BN 层的同时再将 Relu 激活函数替换成能更好拟合、可保留更多的浅层特征信息、在深层特征更具区分度的 PRelu 激活函数。再通过通道和空间注意力模块, 为了保证训练更加稳定, 残差块在相加前, 经过卷积和注意力处理的一路乘以 0.105, 经测试该模型在计算效率上优于 SRGAN, 还能有更好的拟合效果, 最终的结构如图 5 所示。

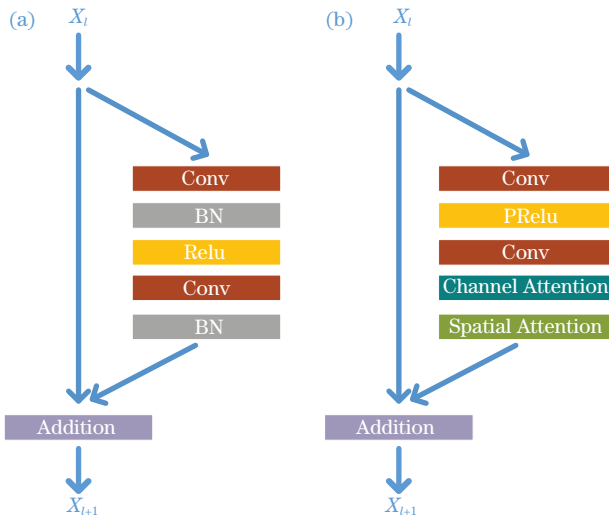


图 5 残差块比较。(a) SRGAN; (b) 本文模型

Fig. 5 Comparison of residual blocks.

(a)SRGAN; (b) proposed model

3 实验结果与分析

3.1 实验数据集

本文使用的训练数据集是 DIV2K^[19] 800 张高清图像数据集。验证集使用的是 DIV2K 中编号为 801~900 的 100 张图片, 为了与其他方法形成更好的客观对比, 以使用最广泛、认可度较高的 Set5、Set15、Urban100、BSDS100 四个数据集进行测试^[20-23]。实验时的 LR 图像是由原图经过图像双三次下采样 4 倍创建的。

依据计算机性能, 图像在输入前被统一裁剪成 88×88 大小的小图像, 训练 batch_size 设置为 8, num_work 设置为 4, epoch 设置为 200, 验证 batch_size 设置为 1, num_work 设置为 4。生成网络和判别网络的优化器都是 Adam 优化器。

3.2 运行环境

实验的硬件设备为 AMD Ryzen 7 4800H @ 2.9 GHz 处理器, 16 GB 运行内存, NVIDIA GeForce RTX2060 6 G 显卡。配置的虚拟环境为 Anaconda3, cuda_10.1, Pycharm (Python3.7), cudnn-10.1, 深度学习框架为 Pytorch, 操作系统是 Win10。

3.3 质量评价

本文使用图像超分辨率领域常用的 PSNR 和结构相似性 (SSIM), PSNR 是基于误差敏感的图像质量评价, 其值越大说明图像的失真越小, SSIM 是一种符合人类视觉的图像质量评价标准, 基于局部图案的亮度、对比度进行计算, 其值越接近 1, 说明生成图像的质量越好, 具体公式为

$$R_{PSNR} = 10 \times$$

$$\lg \frac{255^2 \times W \times H \times C}{\sum_{i=1}^W \sum_{j=1}^H \sum_{z=1}^C [\bar{x}(i, j) - x(i, j)]^2 + 1 \times 10^{-9}}, \quad (8)$$

式中: W 、 H 、 C 分别表示图片的宽度、长度、和通道数; x 表示 SR 图像; \bar{x} 表示原始图像。

$$R_{SSIM}(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)}, \quad (9)$$

式中: μ_X 、 μ_Y 分别表示图像 X 、 Y 的均值; σ_X 、 σ_Y 分别表示图像 X 、 Y 的方差; σ_{XY} 表示 X 图像和 Y 的协方差; C_1 、 C_2 为常数。

3.4 结果分析

实验在数据集 Set5、Set14、Urban100、BSDS100 上进行了测试,为了验证本文算法的有效性,分别与 Bicubic、ESPCN 等经典图像超分辨率重建算法进行比较,同时还与感知驱动的方法 SRGAN,以及在此基础上进行了改进的 ESRGAN 进行对比。

这些算法性能优越,具有代表性,且与本文算法具有较强的相关性。为了体现实验的公平性,所有实验的训练模型均在 DIV2K 的训练集上进行训练,训练中的 epoch 均为 200。因此,所测结果与原论文中的数据有所出入。训练过程中的生成器损失函数值和判别器损失函数值变化如图 6 和图 7 所示。

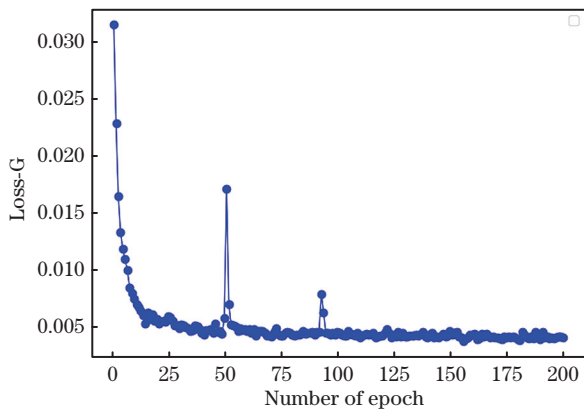


图 6 生成器函数损失值变化曲线

Fig. 6 Variation curve of generator function loss value

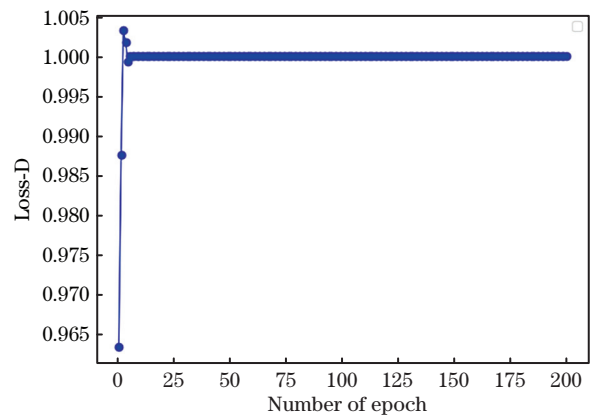


图 7 判别器函数损失值变化曲线

Fig. 7 Variation curve of discriminant function loss value

本文在 Set5 测试集中选出“baby”“butterfly”图像,在 Set14 测试集中选出“pepper”图像,在 BSDS100 测试集中选出“fish”图像,在 Urban100 测试集中选出“room”图像来进行可视化展示,使用多个数据集展示,是为了证明本文模型的通用性。如图 8~12 所示。通过重建的局部放大图可以看出, Bicubic 的重建图像高频信息丢失较多,图像严重失真。图像整体的视觉效果较为模糊,重建图像的质量最差。ESPCN 生成的图像细节都接近原图,但还是较模糊,尽管其 PSNR 值较高,但是边缘锯齿状较为明显,主观感觉较差。SRGAN 和 ESRGAN 都是基于感知损失建立的模型,整体视觉效果远超前两个方法,但 SRGAN 还是较为模糊,图像细节较少,ESRGAN 整体感观最好,但 PSNR 值不是最高,说明生成的细节部分是模型推理填充的,不是原图中真实的细节。本文方法生成的图像在视觉效果上与 ESRGAN 相近,部分图像细节更接近原图, PSNR 值和 SSIM 值均高于其他方法,再次证明本文方法是一种能均衡提升视觉效果和客观评价指标的良好方法。

表 1、表 2 分别表示各超分辨率重建方法在 PSNR 值和 SSIM 值。从表 1 和表 2 可以看出,本

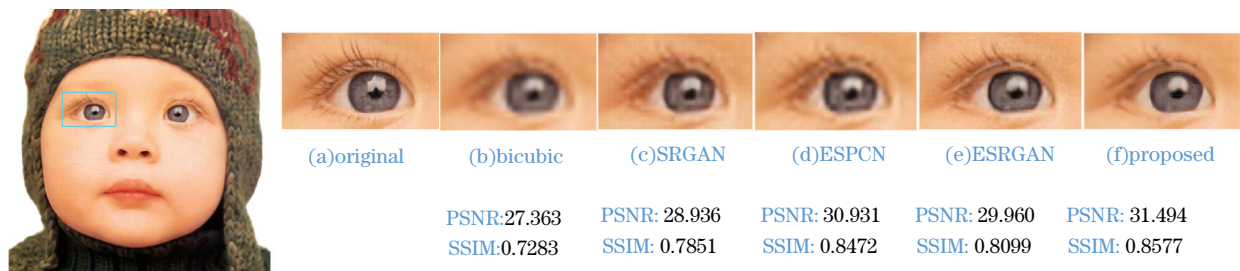


图 8 五种算法在 Set5 测试集中的“baby”重建效果局部放大对比图

Fig. 8 Partial enlarged comparison diagrams of the “baby” reconstruction effect of five algorithms in Set5 test set

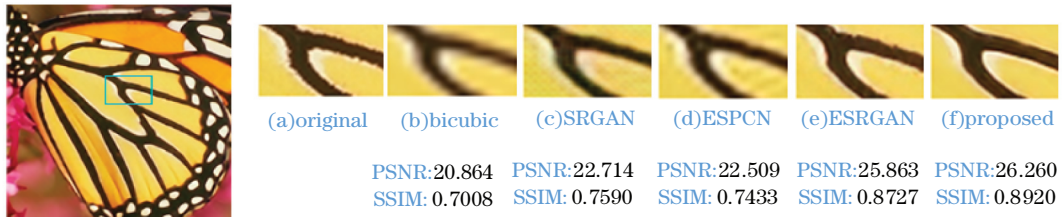


图 9 五种算法在 Set5 测试集中的“butterfly”重建效果局部放大对比图

Fig. 9 Partial enlarged comparison diagrams of the “butterfly” reconstruction effect of five algorithms in Set5 test set

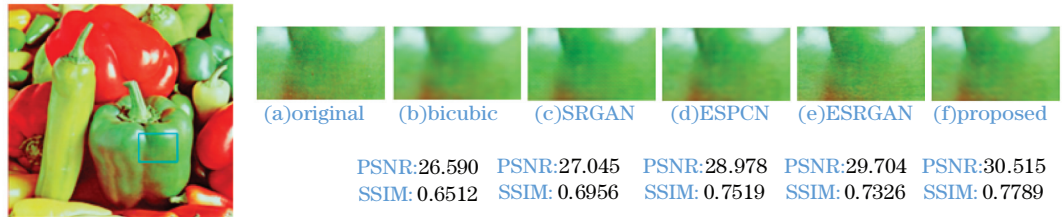


图 10 五种算法在 Set14 测试集中的“pepper”重建效果局部放大对比图

Fig. 10 Partial enlarged comparison diagrams of the “pepper” reconstruction effect of five algorithms in Set14 test set

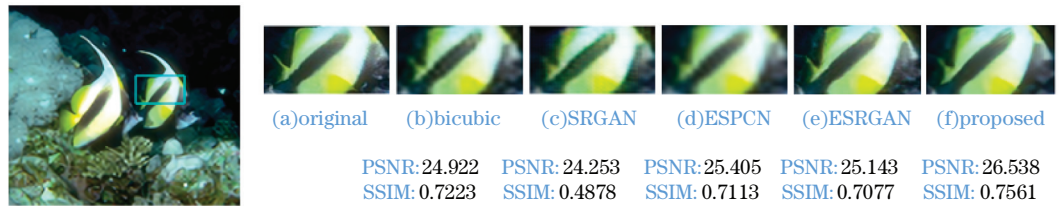


图 11 五种算法在 BSDS100 测试集中“fish”重建效果局部放大对比图

Fig. 11 Partial enlarged comparison diagrams of the “fish” reconstruction effect of five algorithms in BSDS100 test set

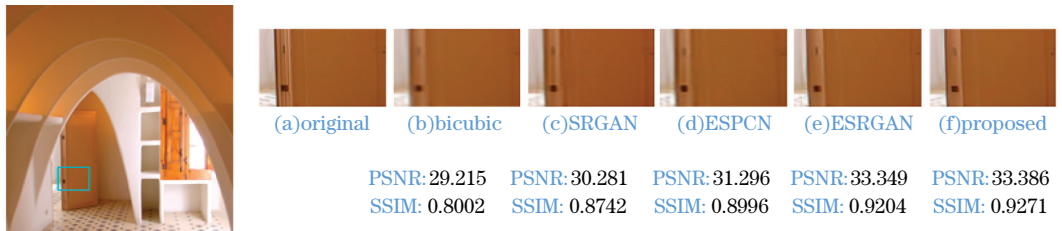


图 12 五种算法在 Urban100 测试集中“room”重建效果局部放大对比图

Fig. 12 Partial enlarged comparison diagrams of the “room” reconstruction effect of five algorithms in Urban100 test set

表 1 各超分辨率重建方法 PSNR 值对比

Table 1 Comparison of PSNR values of various super-resolution reconstruction methods

Test set	Scale	Bicubic	ESPCN	SRGAN	ESRGAN	Proposed
Set5	4×	26.692	27.594	26.628	28.543	29.510
Set14	4×	24.565	25.186	24.568	24.532	26.443
Urban100	4×	21.706	22.300	22.113	22.792	23.917
BSDS100	4×	24.641	25.043	24.472	25.322	25.884

表 2 各超分辨率重建方法 SSIM 值对比

Table 2 Comparison of SSIM values of various super-resolution reconstruction methods

Test set	Scale	Bicubic	ESPCN	SRGAN	ESRGAN	Proposed
Set5	4×	0.7730	0.7895	0.8023	0.8145	0.8517
Set14	4×	0.6732	0.6983	0.7019	0.6711	0.7377
Urban100	4×	0.6317	0.6595	0.6774	0.7050	0.7415
BSDS100	4×	0.6401	0.6702	0.6713	0.6514	0.7002

文算法在四倍放大重建尺度因子下,四个测试集的 PSNR 和 SSIM 指标值都比其他算法要高。相较于插值方法,提升尤为明显,这是因为深度学习具有更多可自学习的参数,可更好地重建学习模型,实现更好的表达能力。但是效果值相较于 SRGAN 提升最为明显,这是因为本文算法构建了更为合理的损失函数,同时去除了 BN 层,可有效地增加网络深度,生成拟合度更好的网络模型,使得训练的网络模型在图像视觉效果的提升和客观评价值的提升上找到一个平衡点。本文算法相较于 ESRGAN 的 PSNR 数值提升大多都在 1 dB,提升幅度明显。以 Set5 为例,本文算法相较于 Bicubic、ESPCN、SRGAN、ESRGAN,其 PSNR 值分别提高了 2.818 dB、1.916 dB、2.882 dB、0.967dB,SSIM 值分别提高 0.0787、0.0622、0.0494、0.0372,由此可以看出,本文方法在客观质量评价中,整体重建效果更好。

为了展示 Charbonnier 模块和 CA&SA 模块的性能,本文进行了消融实验,分别在 SRGAN 的基础

上搭配不同模块组成不同模型,用来分析不同模块对重建效果的影响。在放大尺度因子为 4 时,在 SRGAN 模型的基础上分别进行去除 BN 层的操作生成对比实验一,在 SRGAN 上加 CA&SA 模块生成对比实验二,在 SRGAN 上加 Charbonnier 模块生成对比实验三,以及本文模型即三种模块同时作用到 SRGAN 模型上时产生的效果,实验结果得到的 PSNR 值和 SSIM 值如表 3 和表 4 所示。可以看出,加上任意一个模块均比原模型取得了更好的重建效果值,在本文模型中同时使用三种模块效果提升最大。主观效果如图 13~15 所示,在 SRGAN 的基础上分别加入不同模块客观评价价值均有提升,但视觉效果的提升不是特别明显,仅略有提升,单一模块对整体模型的生成效果影响较小,然而三种模块同时作用在模型上视觉效果提升非常显著,由此可以看出,增加各个模块均有益于图像的重建,由消融实验可充分证明本文算法的可行性和有效性。

表 3 不同模块组合的模型在 4 个测试集上的 PSNR 值

Table 3 PSNR values of models with different module combinations on four test sets

Method	Set5	Set14	Urban100	BSDS100
SRGAN	26.628	24.568	22.113	24.472
SRGAN-BN	27.863	25.269	22.743	25.228
SRGAN+CA&SA	27.865	25.375	22.693	25.198
SRGAN+Charbonnier	28.153	25.724	22.919	25.404
SRGAN+Charbonnier+CA&SA-BN	29.510	26.443	23.917	25.884

表 4 不同模块组合的模型在 4 个测试集上的 SSIM 值

Table 4 SSIM values of models with different module combinations on four test sets

Method	Set5	Set14	Urban	BSDS100
SRGAN	0.8023	0.7019	0.6774	0.6713
SRGAN-BN	0.8058	0.7079	0.6820	0.6728
SRGAN+CA&SA	0.8057	0.7068	0.6792	0.6754
SRGAN+Charbonnier	0.8203	0.7180	0.6939	0.6822
SRGAN+Charbonnier+A&SA-BN	0.8517	0.7377	0.7415	0.7002

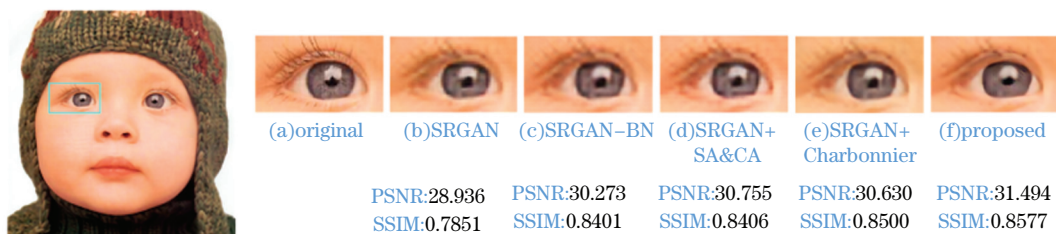


图 13 Set5 测试集中的“baby”在消融实验中重建效果局部放大对比图

Fig. 13 Partial enlarged comparison diagrams of the “baby” reconstruction effect in ablation experiment in Set5 test set

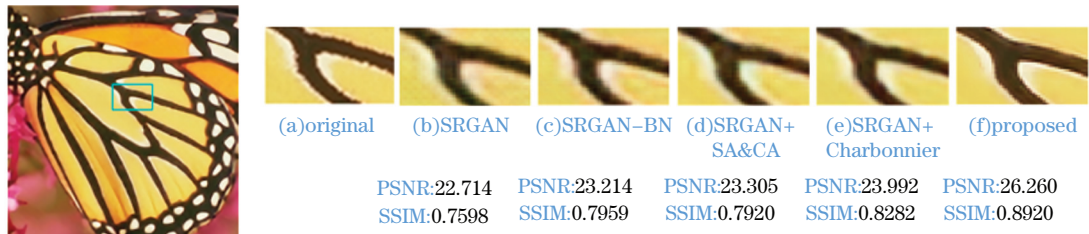


图 14 Set5 测试集中的“butterfly”在消融实验中重建效果局部放大对比图

Fig. 14 Partial enlarged comparison diagrams of the “butterfly” reconstruction effect in ablation experiment in Set5 test set

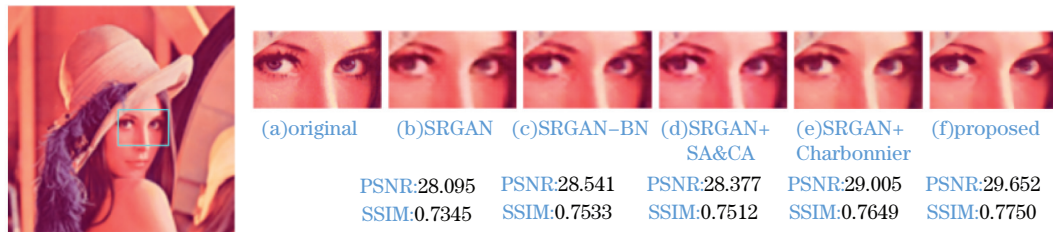


图 15 Set14 测试集中的“lenna”在消融实验中重建效果局部放大对比图

Fig. 15 Partial enlarged comparison diagrams of the “lenna” reconstruction effect in ablation experiment in Set14 test set

4 结 论

本文在 SRGAN 的基础上,提出一种融入注意力机制的生成对抗网络实现单图像超分辨率重建方法,首先去除了残差块中影响图像生成质量的 BN 层,提高了网络的训练速度,同时使用拟合效果更好的 PReLU 激活函数。构造注意力卷积神经网络残差块,增加了图像的高频信息,又引入了鲁棒性更好的 Charbonnier 损失函数替代了 MSE 损失函数,最后使用 TV 正则项平滑训练结果。经过前文的论述,相较于其他方法,本文测试图像的 PSNR 值和 SSIM 值都较高,同时也具有较好的视觉效果,证明了本文方法可实现图像质量和客观评价指标的均衡提升。在未来的研究中将会在拓宽网络深度和网络宽度中进一步改进网络,来构建更好的网络模型以进一步提高图像重建质量。

参 考 文 献

- [1] Zou W W W, Yuen P C. Very low resolution face recognition problem[J]. IEEE Transactions on Image Processing, 2012, 21(1): 327-340.
- [2] Liu K W, Ma Y, Xiong H X, et al. Medical-image super-resolution reconstruction method based on residual channel attention network [J]. Laser & Optoelectronics Progress, 2020, 57(2): 021014. 刘可文, 马圆, 熊红霞, 等. 基于残差通道注意力网络的医学图像超分辨率重建方法[J]. 激光与光电子学进展, 2020, 57(2): 021014.
- [3] Zhang H Y, Yang W M, Wang H S. 3D face recognition combining local keypoints with isogeodesic curves [J]. Laser & Optoelectronics Progress, 2020, 57(22): 221503. 张红颖, 杨维民, 王汇三. 结合局部关键点集与测地线的三维人脸识别[J]. 激光与光电子学进展, 2020, 57(22): 221503.
- [4] Chang J X, Wang S X, Yang Y W, et al. Hierarchical optimization method of building contour in high-resolution remote sensing images[J]. Chinese Journal of Lasers, 2020, 47(10): 1010002. 常京新, 王双喜, 杨元维, 等. 高分遥感影像建筑物轮廓的逐级优化方法[J]. 中国激光, 2020, 47(10): 1010002.
- [5] Zhu S Y, Zeng B, Liu G H, et al. Image interpolation based on non-local geometric similarities [C] // 2015 IEEE International Conference on Multimedia and Expo (ICME), June 29-July 3, 2015, Turin, Italy. New York: IEEE Press, 2015: 1-6.
- [6] Zhang K B, Gao X B, Tao D C, et al. Single image super-resolution with non-local means and steering kernel regression [J]. IEEE Transactions on Image Processing, 2012, 21(11): 4544-4556.
- [7] Peng Y F, Gao Y, Du T T, et al. Single image super-resolution reconstruction method for generative adversarial network [J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(9): 1612-1620. 彭晏飞, 高艺, 杜婷婷, 等. 生成对抗网络的单图像

- 超分辨率重建方法[J]. 计算机科学与探索, 2020, 14(9): 1612-1620.
- [8] Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9906: 391-407.
- [9] Shi W Z, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1874-1883.
- [10] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [11] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1646-1654.
- [12] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1132-1140.
- [13] Lai W S, Huang J B, Ahuja N, et al. Deep Laplacian pyramid networks for fast and accurate super-resolution[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5835-5843.
- [14] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 286-301.
- [15] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 105-114.
- [16] Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks [M] // Leal-Taixé L, Roth S. Computer vision-ECCV 2018 Workshops. Lecture notes in computer science. Cham: Springer, 2019, 11133: 63-79.
- [17] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [18] Woo S, Hwang S, Kweon I S. StairNet: top-down semantic aggregation for accurate one shot detection [C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Lake Tahoe, NV, USA. New York: IEEE Press, 2018: 1093-1102.
- [19] Timofte R, Agustsson E, van Gool L, et al. Ntire 2017 challenge on single image super-resolution: methods and results [C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops, July 21-26, 2017, Honolulu, HI, USA, 2017: 114-125.
- [20] Lei P C, Liu C, Tang J G, et al. Hierarchical feature fusion attention network for image super-resolution reconstruction[J]. Journal of Image and Graphics, 2020, 25(9): 1773-1786.
雷鹏程, 刘丛, 唐坚刚, 等. 分层特征融合注意力网络图像超分辨率重建[J]. 中国图象图形学报, 2020, 25(9): 1773-1786.
- [21] Yuan Y, Liu S Y, Zhang J W, et al. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 814-81409.
- [22] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 5197-5206.
- [23] Matsui Y, Ito K, Aramaki Y, et al. Sketch-based manga retrieval using manga109 dataset[J]. Multimedia Tools and Applications, 2017, 76(20): 21811-21838.