

基于改进超分辨率生成对抗网络的图像重建算法

查体博, 罗林, 杨凯*, 张渝, 李金龙

西南交通大学物理科学与技术学院, 四川 成都 610031

摘要 针对现有基于像素损失的超分辨率图像重建算法对纹理等高频细节的重建效果差问题, 提出了一种基于改进超分辨率生成对抗网络(SRGAN)的图像重建算法。首先, 去除了生成器中的批归一化层, 并结合多级残差网络和密集连接, 用残差套残差密集块提高了网络提取特征的能力。然后, 结合均方误差与感知损失作为指导生成器训练的损失函数, 既保留了图像的高频细节, 又避免了伪影的出现。最后, 去除了判别器的最后一层 Sigmoid 层, 以更好地收敛训练过程, 并用相对损失函数指导判别器的训练。在 COCO 数据集上的实验结果表明, 相比原始 SRGAN 算法, 本算法在 Set5 数据集上的峰值信噪比(PSNR)、结构相似性(SSIM)分别提高了 0.86 dB、0.0123; 在 Set14 数据集上的在 PSNR、SSIM 分别提高了 0.69 dB、0.0090, 且本算法的平均意见指数和视觉效果远优于其他算法。

关键词 图像处理; 超分辨率; 深度学习; 残差网络; 生成对抗网络

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.0810005

Image Reconstruction Algorithm Based on Improved Super-Resolution Generative Adversarial Network

Zha Tibo, Luo Lin, Yang Kai*, Zhang Yu, Li Jinlong

School of Physical Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan 610031, China

Abstract Aiming at the problem that the existing pixel loss-based super-resolution image reconstruction algorithms have poor reconstruction effect on high-frequency details, such as textures, an image reconstruction algorithm based on an improved super-resolution generative adversarial network (SRGAN) is proposed in this paper. First, remove the batch normalization layers in the generator, combine the multi-level residual network and dense connections, and use the residual-in-residual dense blocks to improve the network's ability for feature extraction. Then, the mean square error and perceptual loss are combined as the loss function to guide the generator training, which preserves the image's high-frequency details and avoids the artifacts' appearance. Finally, the last Sigmoid layer of the discriminator is removed to better converge the training process, and the relativistic loss function is used to guide the discriminator training. The experimental results on the COCO dataset show that compared with the original SRGAN algorithm, the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of the algorithm in the Set5 data set are increased by 0.86 dB and 0.0123, respectively, in the Set14 data set, the PSNR and SSIM of the algorithm are improved by 0.69 dB and 0.0090, respectively. The mean opinion index and visual effect of the algorithm are far better than other algorithms.

Key words image processing; super-resolution; deep learning; residual network; generative adversarial network

OCIS codes 100.3020; 100.6640; 110.3010

收稿日期: 2020-07-28; 修回日期: 2020-08-13; 录用日期: 2020-09-10

基金项目: 国家自然科学基金(61771409)

* E-mail: yangkai_swjtu@163.com

1 引言

单张图像超分辨率(SISR)可从退化的低分辨率(LR)图像中重建出令人满意的高分辨率(HR)图像^[1]。SISR可应用于各种计算机视觉任务,如监视成像、医学成像和图像生成。图像的超分辨率(SR)是一个不适定的逆过程,对输入的LR图像可输出多个解,因此,人们提出了多种基于插值、基于重构和基于学习的SR算法。Dong等^[2]首先在图像SR中引入了三层卷积神经网络(CNN),提出了基于CNN的图像超分辨率(SRCNN)算法。之后,Dong等^[3]训练了基于快速CNN的图像超分辨率(FSRCNN)算法,运用反卷积层、小尺寸卷积核以及更多的映射层,解决了SRCNN算法训练时间长、重建耗时的问题。Kim等^[4]借鉴深度学习中的残差学习^[5]思想提出了超深度卷积网络的精确图像超分辨率(VDSR)算法,该算法用很深的网络结构,进一步提升了图像的重建质量。Kim等^[6]提出了基于深度循环卷积神经网络(DRCN)的图像SR算法,进一步提升了图像重建的效果。Shi等^[7]通过改进反卷积的方式,提出了一种基于高效亚像素CNN的实时单图像和视频超分辨率(ESPCN)算法。Caballero等^[8]在ESPCN算法的基础上提出了基于时空网络和运动补偿的实时视频超分辨率(VESPCN)算法。胡诗语等^[9]提出了基于密集连接与激励模块的图像SR网络,将局部特征和整体特征通过密集连接和激励监督后输出至重建网络。目前,基于CNN构建的SR重建算法得到了快速发展,新的网络结构和训练策略不断提升了SR重建图像的质量,尤其是图像的峰值信噪比(PSNR),但这些算法均倾向于在没有足够高频细节的情况下输出过于平滑的结果,且基于PSNR的度量结果与人类的主观评价不一致^[10-11]。

Goodfellow等^[12]提出的生成对抗网络(GAN)是近年来新的研究方向,Ledig等^[13]提出了一种超分辨率GAN(SRGAN)算法,该算法主要由生成器(Generator)和判别器(Discriminator)组成。生成

器的作用是对图像进行重建,判别器的作用是区分真实图像与重建图像,两者的目标相互对抗,通过迭代训练达到平衡。此外,SRGAN算法用感知损失^[14]使网络生成的图像更接近真实图像。相比利用均方误差(MSE)损失的算法,SRGAN算法可明显提高重建图像的整体视觉效果。但对纹理细节的重构中存在模糊和伪影问题,且在损失函数的设置上还有较大提升空间。

针对上述问题,本文提出了一种基于深度GAN的SR重建算法。首先,用残差套残差密集块(RRDB)^[15]提高生成器的特征提取能力,用密集连接使生成器结构的网络容量更高。然后,去除生成器的归一化(BN)层,提升网络性能表现的同时减少伪影现象;且去除BN层有助于提高算法的鲁棒性、降低计算复杂度和内存使用。训练时发现,不包含MSE的损失函数指导的生成器网络能更好地重建图像的细节部分,但也会产生很多伪影。因此,将MSE与感知损失作为指导生成器训练的损失函数。最后,去掉判别器模型中的Sigmoid激活函数,引入Wasserstein距离,以更好地衡量重建图像和真实图像的差异,提升判别器模型训练的收敛速度;并用相对损失(Relativistic loss)^[16]函数训练SRGAN的判别器,以充分训练生成器。

2 基本原理

2.1 生成器网络模型

现有以提高PSNR为导向的SR重建算法倾向输出过于平滑的结果,没有足够的高频细节,且PSNR度量的结果与人们的主观评价并不一致^[13]。为了提高SR重建图像的视觉效果,用感知损失和MSE损失作为损失函数;为使生成图像更接近真实图像,进一步结合图像先验知识改善恢复的纹理细节。与基于PSNR的算法相比,SRGAN算法重建图像的整体视觉效果更好。但SRGAN算法重建的图像与原始图像仍然存在明显差距。为了进一步提高SRGAN算法恢复图像的质量,首先,保留SRGAN算法中生成器的整体架构,如图1所示,其

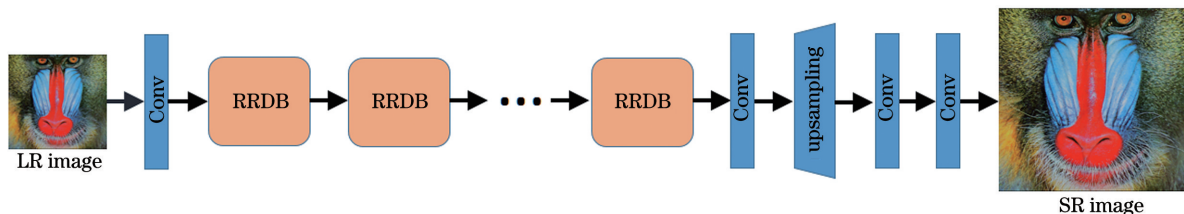


图1 生成器的结构

Fig. 1 Structure of the generator

中, Conv 为 CNN 中的卷积层。然后, 对生成器的基本模块 (Basic block) 进行改进, 具体步骤如下。

1) 删除所有 BN^[17] 层, 如图 2 所示。其中, RB 为残差块, ReLU 为修正线性单元。

2) 用 RRDB 替换原始基本模块, 如图 3 所示。其中, LReLU 为带泄露的修正线性单元, 在输入值为负时, 其梯度为一个常数。

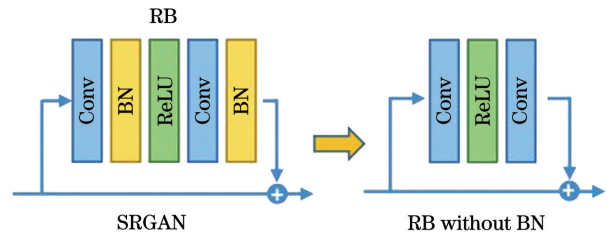


图 2 去掉 BN 层的网络结构

Fig. 2 Network structure after removing the BN layer

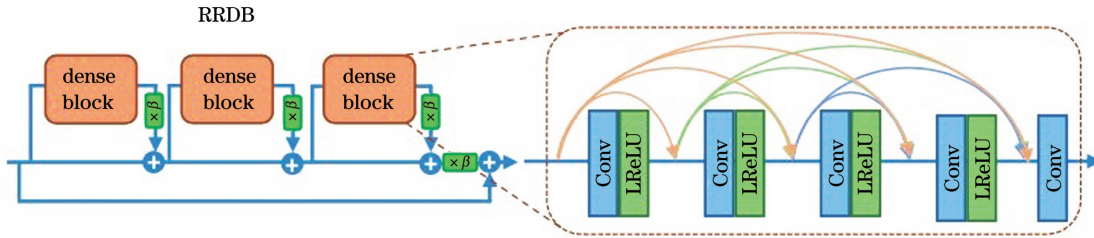


图 3 RRDB 的结构

Fig. 3 Structure of the RRDB

去除 BN 层已经被证明在处理以 PSNR 为导向^[15]的任务时, 可提高网络的性能并降低计算复杂度。基于增强深度残差网络的多尺度单张图像超分辨率 (MDSR) 算法^[15] 和增强的 SRGAN (ESRGAN)^[18] 模型均通过去除 BN 层优化网络结构。BN 的基本原理是在训练过程中, 用一批数据的均值和方差对模型不同层的输入数据进行归一化处理。但当模型训练集和测试集数据的数学统计特征差异较大时, 会引入影响视觉效果的伪影, 限制了 SR 模型生成图像的能力。当网络非常深且在 GAN 框架下进行训练时, 去除 BN 层有助于提高网络的鲁棒性、降低计算复杂度和内存使用。为了获得稳定的训练和一致的表现, 在改进的生成器模型中, 去除了 BN 层, 保留 SRGAN 的高层架构设计。已有研究表明, 更多的层和连接可以提升重建图像的视觉效果^[15], 因此, 采用比原始 SRGAN 中 RB 更深、更复杂的 RRDB。将 RRDB 和主路径中使用的密集块作为残差结构的基本单元, 使改进生成器结构的网络容量更高。除了改进网络结构外, 在每个密集块后都添加了残差缩放层, 即在残差添加到主路径之前乘以常数 β (0 和 1 之间), 以缩小残差, 降低训练过程中不稳定的概率。

2.2 生成器损失函数

传统的 SR 算法过分强调 MSE 损失, 其损失函数可表示为

$$l_{MSE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^r \sum_{y=1}^H [I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y}]^2, \quad (1)$$

式中, I^{LR} 为 HR 图像 I^{HR} 对应的 LR 图像, W 、 H 分别为 I^{LR} 的宽和高, r 为 I^{HR} 到 I^{LR} 的下采样倍数, G_{θ_G} 为由参数 θ_G 构建的前馈神经网络 (生成器网络), $G_{\theta_G}(I^{LR})$ 为重建图像。

SRGAN 算法中的感知损失函数可表示为

$$l^{SR} = l_{VGG}^{SR} + 10^{-3} l_{Gen}^{SR}, \quad (2)$$

式中, l_{VGG}^{SR} 为内容损失, $10^{-3} l_{Gen}^{SR}$ 为对抗性损失成分的加权和。其中, 内容损失可表示为

$$l_{VGG,i,j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x'=1}^{W_{i,j}} \sum_{y'=1}^{H_{i,j}} \left\{ \phi(I_{x',y'}^{HR}) - \phi_{i,j} [G_{\theta_G}(I^{LR})]_{x',y'} \right\}^2, \quad (3)$$

式中, $W_{i,j}$ 和 $H_{i,j}$ 为 VGG (Visual geometry group) 网络中某层特征图的宽和高, $\phi_{i,j}$ 为 VGG 网络内第 i 个最大池化层前通过第 j 个卷积 (激活之后) 获得的特征映射, $\phi(I_{x',y'}^{HR})$ 为真实图像 I^{HR} 经过 VGG 网络后提取的某层特征图中像素 (x', y') 的像素值。因此, SRGAN 算法的内容损失可以理解的真实图像和生成图像分别经过 VGG 网络后特征图的欧氏距离。对抗性损失可表示为

$$l_{Gen}^{SR} = \sum_{n=1}^N -\log \{ D_{\theta_D} [G_{\theta_G}(I^{LR})] \}, \quad (4)$$

式中, $D_{\theta_D} [G_{\theta_G}(I^{LR})]$ 为重建图像 $G_{\theta_G}(I^{LR})$ 是原始 I^{HR} 的概率, D_{θ_D} 为由参数 θ_D 构建的判别器网络, N 为一次批训练的样本数量。在训练过程中发现, 不包含 MSE 的损失函数指导的生成器网络虽然能很好地重建图像的细节部分, 但会产生影响图像质量的伪影。因此, 将 MSE 与感知损失结合

起来作为指导生成器训练的损失函数,可表示为

$$l^{SR} = l_{VGG}^{SR} + \lambda_1 l_{Gen}^{SR} + \lambda_2 l_{MSE}^{SR}, \quad (5)$$

式中, λ_1 和 λ_2 为不同损失对应的权重。

2.3 判别器网络模型

将 WassersteinGAN(WGAN)^[19] 引入图像 SR 重建算法中,以衡量重建图像和真实图像之间的差异。在原始 SRGAN 算法中,判别器最后一层为 Sigmoid 激活函数,可看成一个二元分类器,输出是输入图像为真的置信概率。本模型的判别器包括

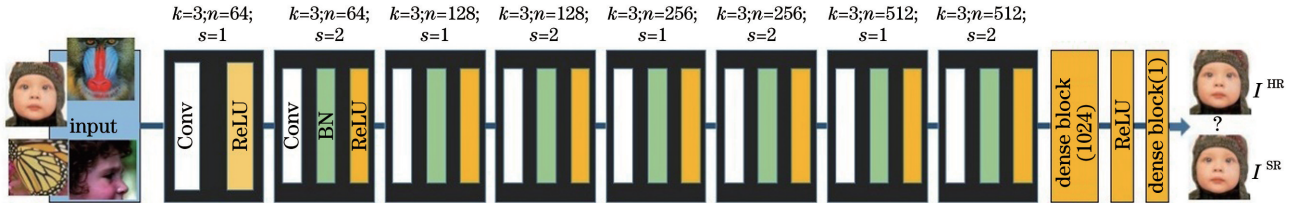


图 4 判别器的结构

Fig. 4 Structure of the discriminator

2.4 判别器损失函数

在原始 SRGAN 算法中,指导判别器训练的损失函数如(2)式所示。经过充分训练,判别器能将更多的真实图像判断为真,将生成图像判断为假。通过训练生成器,使生成的图像骗过判别器的识别,从而提高被判断为真实图像的概率。原始 GAN 忽视了数据集的先验知识,即 GAN 判别器判别的图像中,真实图像和生成图像的概率均为 0.5。生成器和判别器均经过很好的训练,且生成器生成的图像足够接近真实图像。针对真实数据集和生成数据集,判别器的判别概率符合

$$C(x_f) \approx C(x_r), \quad (6)$$

式中, x_f 为生成数据集, x_r 为真实数据集, $C(x_f)$ 和 $C(x_r)$ 为判别器判断生成数据集和真实数据集为真的概率。

原始 GAN 中,假设判别器不知道输入为一半真实数据和一半生成数据,最终训练的结果是判别所有图像全部为真,即 $D_{\theta_D} [G_{\theta_G} (I^{LR})] \approx 1$ 。在训练过程中,判别器判别真实图像为真的概率始终为 1,判断生成图像为真的概率不断上升。但实际中,判别器可以被继续训练,如果不对图像进行单独判断,直接对比生成图像和真实图像,有可能发生 $C(x_f) > C(x_r)$ 的情况,其训练过程^[16] 如图 5 所示。SRGAN 中的标准判别器可表示为

$$\begin{cases} D(x_r) = X_{\text{Sigmoid}} [C(x_r)] \rightarrow 1 \\ D(x_f) = X_{\text{Sigmoid}} [C(x_f)] \rightarrow 0 \end{cases}, \quad (7)$$

式中, $D(\cdot)$ 为标准判别器, X_{Sigmoid} 为 Sigmoid 激活

8 个结构相似的模块(由 Conv 层、ReLU 和 BN 层组成),对于每个图像,鉴别器输出的结果为重建图像和 HR 图像的相似性,判别器的结构如图 4 所示。其中, $k \times k$ 为卷积核的尺寸, $n \times n$ 为特征图的尺寸, s 为步长。相比 GAN, WGAN 的改进主要包括: 1) 不再使用 log 函数作为非线性手段; 2) 不再使用 Sigmoid 激活函数; 3) 每次重新计算梯度时,对超出范围的梯度进行截断; 4) 优化方法不再使用 Adam 方法。

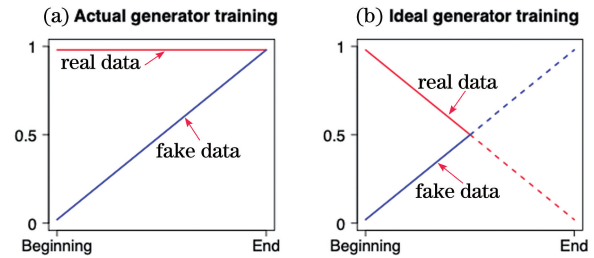


图 5 训练过程示意图。(a)实际的训练曲线;(b)理想的训练曲线^[16]

Fig. 5 Schematic diagram of the training process.

(a) Actual training curve; (b) ideal training curve^[16]

函数。可以发现,标准判别器仅判断单张图像是否为真,真实图像输出值趋近于 1,生成图像输出值趋近于 0。本算法使用的判别器可表示为

$$\begin{cases} \bar{D}(x_r, x_f) = X_{\text{Sigmoid}} \{C(x_r) - E[C(x_f)]\} \rightarrow 1 \\ \bar{D}(x_f, x_r) = X_{\text{Sigmoid}} \{C(x_f) - E[C(x_r)]\} \rightarrow 0 \end{cases}, \quad (8)$$

式中, $\bar{D}(\cdot)$ 为相对损失判别器, $E(\cdot)$ 为对所有数据取平均值的操作。与标准判别器相比,相对损失^[18] 判别器会对比真实图像和生成图像,并输出真实图像比生成图像更接近真实图像的概率,从而提升判别器的准确率。具体判断方法可表示为

$$\begin{aligned} P(x_r = X_{\text{real}}) &= E_{x_f \sim Q} [P(x_r \geq x_f)] = \\ &= E_{x_f \sim Q} \{X_{\text{Sigmoid}} [C(x_r) - C(x_f)]\} = \\ &= E_{x_f \sim Q} [D(x_r, x_f)], \end{aligned} \quad (9)$$

$$\begin{aligned} P(x_f = X_{\text{real}}) &= E_{x_r \sim P} [P(x_f \geq x_r)] = \\ &= E_{x_r \sim P} \{X_{\text{Sigmoid}} [C(x_f) - C(x_r)]\} = \\ &= E_{x_r \sim P} [D(x_f, x_r)], \end{aligned} \quad (10)$$

式中, X_{real} 为真实图像, $P(x_r = X_{\text{real}})$ 为相对损失判别器判断真实数据为真的概率, $P(x_f = X_{\text{real}})$ 为相对损失判别器判断生成数据为真的概率, $x_r \sim P$ 为在真实数据 P 上的随机分布, $x_f \sim Q$ 为在生成数据 Q 上的随机分布, $x_r \geq x_f$ 表示真实图像比生成图像更真实, $x_f \geq x_r$ 表示生成图像比真实图像更真实。可以看出, $D(x_f, x_r) = X_{\text{Sigmoid}} [C(x_f) - C(x_r)]$, 则判别器损失函数可表示为

$$L_D = -E_{x_r \sim P} \left\{ \log \left\{ E_{x_f \sim Q} [D(x_r, x_f)] \right\} \right\} - E_{x_f \sim Q} \left\{ \log \left\{ 1 - E_{x_r \sim P} [D(x_f, x_r)] \right\} \right\}. \quad (11)$$

可以看出, 上述方法需要寻找所有可能的 x_r 与 x_f 组合, 如果一个批训练中包含 h 张图像, 需计算 h^2 次, 计算量较大且耗费计算机的内存。因此先分别对真实图像批训练集和生成图像批训练集求平均, 再进行相应的计算, 此时判别器的相对平均损失函数 $L_{\bar{D}}$ 可表示为

$$L_{\bar{D}} = -E_{x_r \sim P} \left\{ \log [\bar{D}(x_r)] \right\} - E_{x_f \sim Q} \left\{ \log [1 - \bar{D}(x_f)] \right\}, \quad (12)$$

$$\bar{D}(x) = \begin{cases} X_{\text{Sigmoid}} [C(x) - E_{x_f \sim Q} C(x_f)], & x = X_{\text{real}} \\ X_{\text{Sigmoid}} [C(x) - E_{x_r \sim P} C(x_r)], & x = X_{\text{fake}} \end{cases}, \quad (13)$$

式中, X_{fake} 为生成图像, $\bar{D}(x_r)$ 、 $\bar{D}(x_f)$ 分别为真实图像批训练集和生成图像批训练集求平均后的值, 将 $\bar{D}(x)$ 代入(12)式, 可得到指导判别器反向传播的损失函数 $L_{\bar{D}}$

$$L_{\bar{D}} = -E_{x_r \sim P} \left\{ \log \left\{ X_{\text{Sigmoid}} [C(x) - E_{x_f \sim Q} C(x_f)] \right\} \right\} - E_{x_f \sim Q} \left\{ \log \left\{ 1 - X_{\text{Sigmoid}} [C(x) - E_{x_r \sim P} C(x_r)] \right\} \right\}. \quad (14)$$

3 实验结果与分析

3.1 实验环境

3.1.1 实验平台

实验使用的硬件平台: 系统为 Windows10, CPU 为 i5-7500, 电脑内存为 12 GB。软件平台: 数据预处理部分用 Matlab 软件, 模型搭建与训练用 PyCharm 编译的程序, 深度神经网络库为 Pytorch, 该深度神经网络库集成度高, 代码可读性强, 如图 6 所示。

3.1.2 数据集

实验使用的数据集为 SR 训练中常用的 COCO 数据集^[20], 该数据集是由 Microsoft Company 提供

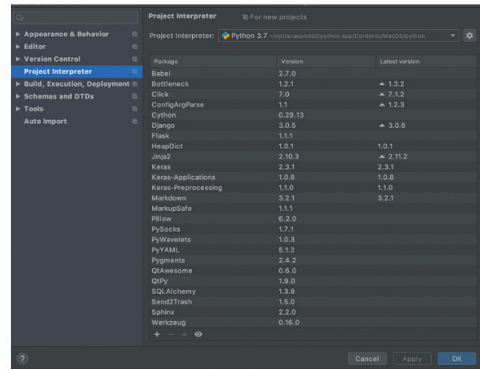


图 6 网络的训练环境

Fig. 6 Training environment of the network

的图像识别数据集, 共包含 91 个类别, 120000 张图像。SRGAN 算法使用的训练集包含 30 万张图像, 因此还需对 COCO 数据集进行数据增强。实验用 Set5、Set14、BSD100 以及铁轨图像 4 个数据集, 其中, Set5 数据集仅包含 5 张图像, 且细节较少; Set14 数据集包含 14 张图像, 人物照片居多, 图像的复杂程度及细节处于中等; BSD100 数据集包含 100 张图像, 图像细节丰富, 重建的挑战性较大; 铁轨图像数据集更注重模型的实际重建效果, 均源于实际生活中, 重建难度较大。

3.1.3 数据增强

为了在一定程度上避免过拟合情况, 深度学习模型的训练往往需要大量的图像。在不同的应用背景下, 一般通过对样本图像的变换增强数据集, 如尺度变换、平移变换、图像翻转、对比度变换。

3.1.4 训练参数设置

1) 实验进行的是倍数为 4 的超分辨率重建。
2) 将尺寸为 128 pixel × 128 pixel 的 HR 图像进行随机裁剪, 并通过双三次采样获得对应的 LR 图像。

3) 批训练(Batch size)中每批包含 16 张图像。

4) 训练过程包括两个阶段: 第一阶段仅使用 MSE 损失函数进行预训练, 初始学习率为 2×10^{-4} , 进行 500 次批训练, 并用第一阶段得到的权重参数对第二阶段要训练的生成器进行初始化; 第二阶段用(5)式作为损失函数指导整个模型的训练, 其中, $\lambda_1 = 5 \times 10^{-3}$, $\lambda_2 = 10^{-2}$, 初始学习率为 10^{-4} , 每经过 50000 个批训练, 学习率降为原来的一半。

5) 训练用 Adam 算法进行优化, 生成器和判别器的参数 β 分别为 0.9 和 0.999。

3.1.5 平均意见指数

主观评价方法是人们通过自己的视觉感受对不

同图像进行评价,将主观评价的最终得分称为平均意见指数(MOI)。当 2 张图像的视觉效果截然不同时,人们对图像进行比较客观的评价;但当 2 张图像的差别较小时,人们的评价结果是随机或掺杂大量个人喜好的。对于评价指标,余伟等^[21]给出了人们评价 1 张图像需要参考的评价标准,如表 1 所示。

表 1 图像质量的评价标准

Table 1 Evaluation standard of the image quality

Evaluation standard	Score
No change in image quality	5
Slight change in image quality can be seen	4
Slightly hinder viewing	3
Hinder viewing	2
Seriously obstructing viewing	1

表 1 中的方法将每张 SR 算法重建的图像和 HR 图像都成对出现一次,如重建 O 张图像,对比 h 种重建算法,完整的评价需进行 $h \times O$ 次判断,导致评价者的判断能力出现疲劳,从而降低评价的科学性和准确性。因此,采用相对主观评价法,每次待评价图像包括 LR 图像和不同 SR 算法重建的图像,评价者根据图像视觉质量对不同算法进行对比。MOI 测试中,选用双盲测试方法。双盲测试指测试者和受试者均不了解某种待测试物品所属的类别,测试者整理和分析实验数据时也不知道实验数据所属的具体类别,从而有效消除受试者的个人喜好及测试者的心理暗示因素对实验结果的影响。因此,在 SR 重建评价任务中,开发了一个用于 MOI 评价系统的图形用户界面(GUI)。该系统的评价过程中,不存在实验人员,且图像顺序被随机打乱,只有计算机知道图像和与之对应的 SR 算法,双盲测试系统的界面如图 7 所示。



图 7 MOI 测试系统的界面

Fig. 7 Interface of the MOI test system

MOI 测试系统的具体测试方法:将软件以及图像数据集打包发送给 10 位受试人员,让受试人员进行打分测试。测试完成后,受试人员将显示的测试

结果发送给测试者,然后计算最终的 MOI。表 2 为图像质量主观评价的打分标准,得到的 MOI 为

$$X_{MOI} = \frac{\sum_{k=1}^z n_k \times O_k}{\sum_{k=1}^z n_k}, \quad (15)$$

式中, z 为评价人数,每张图像的每一种得分记为 O_k , n_k 为每种得分的评价人数。

表 2 图像质量主观评价打分表

Table 2 Scoring table for subjective evaluation of image quality

Relative measurement scale	Score
Quality is the worst in this picture group	1
Quality is below average in this picture group	2
Quality is on average in this picture group	3
Quality is above average in this picture group	4
Quality is the best in this picture group	5

3.2 仿真实验

3.2.1 原始 SRGAN 与本算法的评价指标

将本算法与原始 SRGAN 算法进行了对比实验,并在 Set5 数据集上进行测试,结果如表 3、图 8、图 9 所示。可以发现,相比原始 SRGAN 算法,本算法的 PSNR 提高了 0.86 dB,SSIM 提高了 0.0123,迭代 60 个周期后开始趋于稳定,PSNR 和 SSIM 接近最终训练结果,分别为 29.60 dB 和 0.8558。

表 3 不同算法在 Set5 数据集上的测试结果

Table 3 Test results of different algorithms on the Set5 data set

Algorithm	SRGAN ^[13]	Ours	Difference
PSNR/dB	28.74	29.60	↑ 0.86
SSIM	0.8435	0.8558	↑ 0.0123

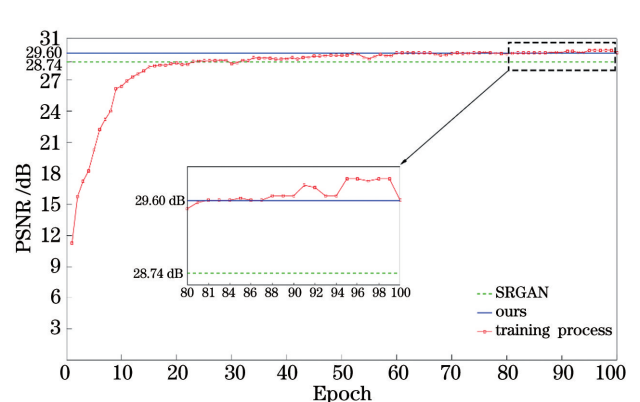


图 8 不同算法在 Set5 测试集的 PSNR

Fig. 8 PSNR of different algorithms in the Set5 test set

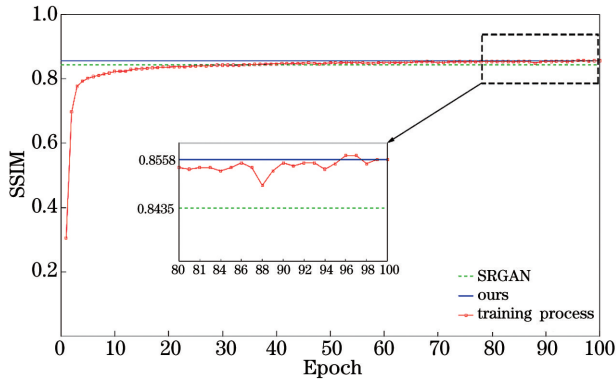


图 9 不同算法在 Set5 测试集上的 SSIM

Fig. 9 SSIM of different algorithms on the Set5 test set

本算法和原始 SRGAN 算法在 Set14 数据集上的实验结果如表 4、图 10、图 11 所示。可以发现,相比原始 SRGAN 算法,本算法的 PSNR 提高了 0.69 dB, SSIM 提高了 0.0090,迭代 60 个周期后开始趋于稳定,PSNR 和 SSIM 接近最终训练结果,分别为 26.44 dB 以及 0.746。

表 4 不同种算法在 Set14 数据集上的测试结果

Table 4 Test results of different algorithms on the Set14 data set

Algorithm	SRGAN ^[13]	Ours	Difference
PSNR/dB	25.75	26.44	↑ 0.69
SSIM	0.7370	0.7460	↑ 0.0090

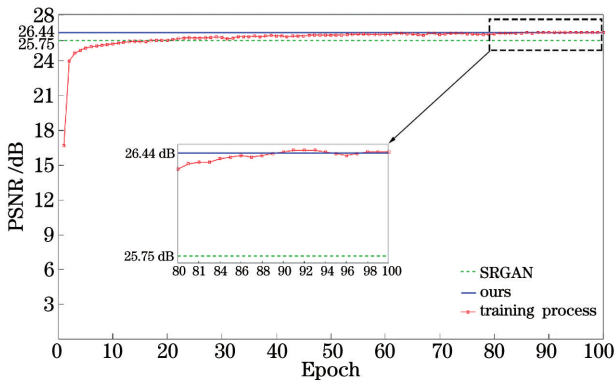


图 10 不同算法在 Set14 测试集上的 PSNR

Fig. 10 PSNR of different algorithms on the Set14 test set

表 6 BN 层对算法性能的影响

Table 6 Influence of BN layer on algorithm performance

Data set	SRGAN(with BN)			Ours(with BN)			SRGAN(without BN)			Ours(without BN)		
	PSNR /dB	SSIM	Time /s	PSNR /dB	SSIM	Time /s	PSNR /dB	SSIM	Time /s	PSNR /dB	SSIM	Time /s
Set5	28.69	0.8415	0.21	29.02	0.8489	0.20	29.25	0.8483	0.20	29.60	0.8547	0.18
Set14	24.96	0.7187	0.43	26.10	0.7233	0.39	25.67	0.7203	0.42	26.41	0.7398	0.38
BSD100	24.01	0.6488	0.45	25.15	0.6511	0.43	25.10	0.6503	0.44	25.52	0.6548	0.41

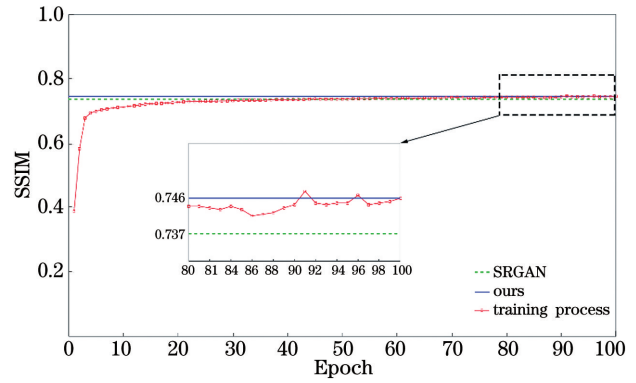


图 11 不同算法在 Set14 测试集的 SSIM

Fig. 11 SSIM of different algorithms in the Set14 test set

表 5 为不同算法在 BSD100 测试集上得到的结果,可以发现,相比原始 SRGAN 算法,本算法的 PSNR 提高了 0.90 dB,SSIM 提高了 0.0047。由于 BSD100 数据集包含 100 张图像,每个训练周期都对 BSD100 数据集进行测试会花费大量的时间,因此未绘制 BSD100 训练过程中的 PSNR、SSIM 曲线。

表 5 不同算法在 BSD100 数据集上的测试结果

Table 5 Test results of different algorithms on the BSD100 data set

Algorithm	SRGAN ^[13]	Ours	Difference
PSNR/dB	24.65	25.55	↑ 0.90
SSIM	0.6502	0.6549	↑ 0.0047

为了验证 BN 层对算法性能的影响,用 SRGAN 算法和本算法设置了一个迭代 300 次的实验,分别测试了有无 BN 层时算法的性能,测得的平均 PSNR、SSIM、重建时长如表 6 所示。可以发现,去掉 BN 层的算法重建图像的平均 PSNR 和 SSIM 均较高,且重建时间较短,这表明去除 BN 层可加快图像重建的速度、提高重建图像的质量。

3.2.2 不同 SR 算法的评价指标

将训练好的算法与目前主流的算法进行了对比,包括插值算法、SRCNN、VDSR、SRResNet,不同算法在 Set5、Set14 以及 BSD100 测试集中的测试

结果如表 7 所示。可以发现,相比目前流行的 SR 算法,本算法的 PSNR 较低,原因是本算法不是以 PSNR、SSIM 为导向进行训练。SRGAN 算法及本算法主要强调图像的真实感与细节恢复能力,且 PSNR、SSIM 不能代表图像的视觉效果。

SRResNet 算法在所有数据集上的 PSNR 和 SSIM 均最大,但 SRResNet 仅利用 SRGAN 的生成器进行训练,也是以 MSE 作为损失函数指导生成器的训练,其重建图像的视觉效果远不如 SRGAN 算法。

表 7 不同算法在三个数据集下的性能

Table 7 Performance of different algorithms under three data sets

Dataset	Algorithm	Bicubic	SRCNN ^[2]	VDSR ^[5]	SRResNet ^[13]	Ours
Set5	PSNR /dB	28.43	30.14	31.35	31.92	29.60
	SSIM	0.8211	0.8647	0.8838	0.8998	0.8558
	MOI	1.44	2.4	3.18	3.28	4.66
Set14	PSNR /dB	25.99	27.18	28.01	28.39	26.44
	SSIM	0.7486	0.7861	0.7674	0.8116	0.746
	MOI	1.42	2.43	3.18	3.59	4.38
BSD100	PSNR /dB	25.96	26.9	27.29	27.52	25.55
	SSIM	0.6675	0.7101	0.7251	0.7603	0.6549
	MOI	1.36	2.45	3.2	3.57	4.42

3.2.3 本算法与 SRGAN 算法的重建效果

图 12 为本算法与 SRGAN 算法的重建效果,可以发现,相比原始 SRGAN 算法,本算法能恢复出纹理细节清晰的图像。对于 Set14 数据集中狒狒的面部,原始 SRGAN 算法重建的图像中出现了很多裂纹,而本算法重建的猩猩面部更平滑;对于 BSD100

数据集中的草地,原始 SRGAN 算法出现了很多伪影,而本算法能提供完整的细节信息;对于草地上鸟的羽毛,尽管细节部分已经相当模糊,但本算法重建的图像细节更丰富,细腻性更好、MOI 更高。

3.2.4 多种算法的重建效果

为进一步测试本算法的重建性能,选取来自 Set5、Set14 的 3 张图像,对比了本算法和其他算法的重建效果,重建后的图像及其 MOI 值如图 13 所示。从放大的图像局部细节可以发现,双三次插值法的重建效果最差,细节模糊不清。几种基于像素损失的算法虽然恢复了图像的局部细节,但重建图像整体过于平滑和模糊,难以分辨帽子的编织纹理和饰片的边缘等高频信息;本算法重建的图像能更好地恢复细节纹理,如毛线的纤维层次、蝴蝶翅膀的锯齿纹理、女孩帽子饰片的局部层次。尽管输入网络的 LR 图像已经非常模糊,但本算法仍然能很好地恢复图像细节,有效避免了其他算法重建图像中的过度平滑问题;且本算法重建图像的 MOI 值最高,这表明本算法的主观视觉效果更好。

3.2.5 铁轨图像集

为进一步验证本算法在铁路场景下应用的前景,对比测试了多种算法对于铁路领域灰度图像的重建效果,结果如图 14 所示。可以看出,铁路灰度图像重建难度更大,一方面明暗对比强烈,人眼不易分辨局部图像信息;另一方面,缺乏色度信息,对图

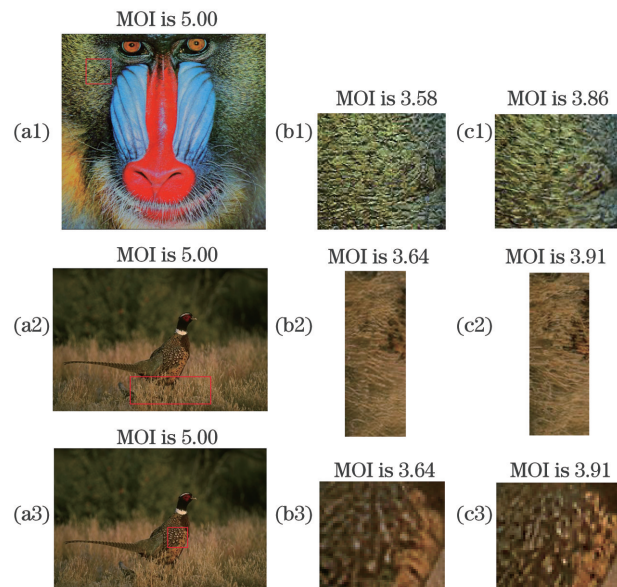


图 12 2 种算法的重建效果。(a)原始图像;(b)SRGAN 算法;(c)本算法

Fig. 12 Reconstruction effects of two algorithms.
(a) Original image; (b) SRGAN algorithm;
(c) our algorithm

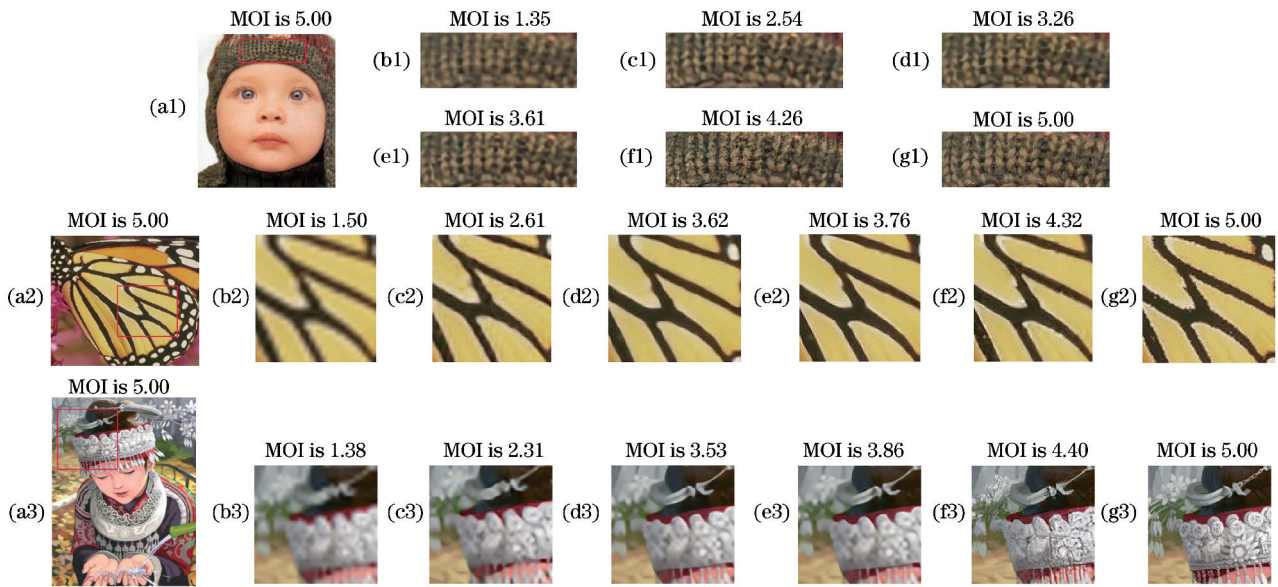


图 13 5 种算法的重建结构。(a)整体原始图像;(b)双三次插值算法;(c)SRCNN 算法;(d)VDSR 算法;(e)SRRResNet 算法;(f)本算法;(g)局部原始图像

Fig. 13 Reconstruction results of 5 different algorithms. (a) Overall original image; (b) bicubic interpolation algorithm; (c) SRCNN algorithm; (d) VDSR algorithm; (e) SRRResNet algorithm; (f) our algorithm; (g) partial original image

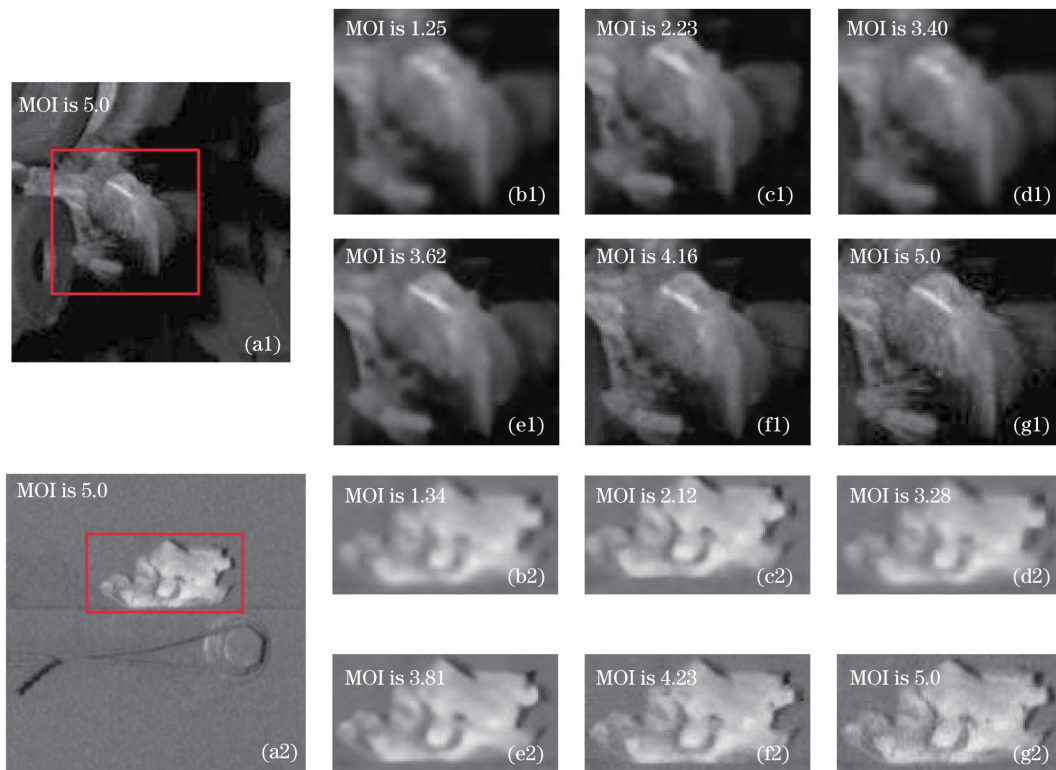


图 14 5 种算法重建的铁轨图像。(a)整体原始图像;(b)双三次插值算法;(c)SRCNN 算法;(d)VDSR 算法;(e)SRRResNet 算法;(f)本算法;(g)局部原始图像

Fig. 14 Railroad track image reconstructed by 5 different algorithms. (a) Overall original image; (b) bicubic interpolation algorithm; (c) SRCNN algorithm; (d) VDSR algorithm; (e) SRRResNet algorithm; (f) our algorithm; (g) partial original image

像信息的提取更加依赖图像的边缘和纹理细节。相比其他算法,本算法重建图像的纹理清晰、边缘明显,没有平滑的痕迹,精确地还原了原始图像的细节,能用于提高铁路部件缺陷检测的精度。

用不同算法对多个数据集中的图像进行 4 倍超分辨率重建,并对其重建效果进行评价。实验结果表明,本算法重建的图像具有更高的清晰度、更精确的纹理细节,获得的视觉效果也最佳。

4 结 论

本算法基于 SRGAN 的主体架构,通过引入更精密的 RRDB、去除 BN 层改善生成器的性能。然后在感知损失和对抗损失基础上添加 MSE 损失,融入 WGAN 思想,并去除 Sigmoid 激活函数。最后采用相对损失函数指导判别器训练,避免追求像素尺度上的一致,引导生成器产生更自然、真实的特征。在多个自然图像数据集和铁路数据集上的重建结果表明,相比 SRGAN 算法,本算法可以减少伪影,提升图像细节,缩短重建时间;且在恢复纹理等细节信息上的效果更佳,避免了过度平滑,视觉感知效果也更好。之后还需对 GAN 进行研究,以提高深度 CNN 在图像 SR 任务中的表现,进一步提升网络对图像特征的提取能力。尽管本算法采用了 MOI 主观评价指标,但由于人们品味与喜好的不同,主观评价值的变化很大,且 PSNR、SSIM 不能很好地表示图像真实的视觉效果,因此对 SR 重建效果的评价需要一个更普适和尽可能接近人眼视觉效果

参 考 文 献

- [1] Xie C, Zeng W L, Lu X B. Fast single-image super-resolution via deep network with component learning [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 29(12): 3473-3486.
- [2] 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*, 2016, 38(2): 295-307.
- [3] Dong C, Loy C C, Tang X. 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.
- [4] 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.
- [5] 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.
- [6] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution [EB/OL]. [2020-07-04]. <https://arxiv.org/abs/1511.04491>.
- [7] 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.
- [8] Caballero J, Ledig C, Aitken A, et al. Real-time video super-resolution with spatio-temporal networks and motion compensation [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2848-2857.
- [9] Hu S Y, Wang G D, Zhao Y, et al. Image super-resolution network based on dense connection and squeeze module [J]. *Laser & Optoelectronics Progress*, 2019, 56(20): 201005.
胡诗语, 王国栋, 赵毅, 等. 基于密集连接与激励模块的图像超分辨率网络 [J]. *激光与光电子学进展*, 2019, 56(20): 201005.
- [10] Yang C Y, Ma C, Yang M H. Single-image super-resolution: a benchmark [M] // Fleet D, Pajdla T, Schiele B, et al. *Computer Vision-ECCV 2014. Lecture Notes in Computer Science*. Cham: Springer, 2014, 8692: 372-386.
- [11] Zhang S F, Zhang C, Zhang T, et al. Review on universal no-reference image quality assessment algorithm [J]. *Computer Engineering and Applications*, 2015, 51(19): 13-23, 151.
张淑芳, 张聪, 张涛, 等. 通用型无参考图像质量评价算法综述 [J]. *计算机工程与应用*, 2015, 51(19): 13-23, 151.
- [12] Goodfellow I J, Pouget A J, Mirza M, et al. Generative adversarial nets [EB/OL]. [2020-07-02]. <https://arxiv.org/abs/1511.04491>.
- [13] 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.
- [14] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution [M] // Leibe B, Matas J, Sebe N, et al. *Computer Vision-*

- ECCV 2016. Lecture Notes in Computer Science. Cham: Springer, 2016, 9906: 694-711.
- [15] 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.
- [16] Jolicoeur-Martineau A. The relativistic discriminator: a key element missing from standard GAN[EB/OL]. [2020-07-02]. <https://arxiv.org/abs/1807.00734>.
- [17] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. [2020-07-04]. <https://arxiv.org/abs/1502.03167>.
- [18] Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks[M]//Leal-Taixé L, Roth S, et al. Computer Vision-ECCV 2018. Lecture Notes in Computer Science. Cham: Springer, 2019, 11133: 63-79.
- [19] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs [EB/OL]. [2020-07-05]. <http://arxiv.org/abs/1704.00028>.
- [20] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context [M]//Fleet D, Pajdla T, Schiele B, et al. Computer Vision-ECCV 2014. Lecture Notes in Computer Science. Cham: Springer, 2014, 8693: 740-755.
- [21] Yu W, Xu J J, Liu Y Y, et al. No-reference quality evaluation for gamut mapping images based on natural scene statistics[J]. Laser & Optoelectronics Progress, 2020, 57(14): 141006.
余伟, 徐晶晶, 刘玉英, 等. 基于自然场景统计的色域映射图像无参考质量评价[J]. 激光与光电子学进展, 2020, 57(14): 141006.