

基于多损失融合与谱归一化的图像超分辨率方法

许宁宁, 郑 凯

(华东师范大学 计算机科学与软件工程学院计算中心, 上海 200062)

摘要: 图像超分辨率重建研究存在结果客观衡量指标不断变优, 但是视觉感知质量依旧平滑的问题。同时, 应用生成对抗网络的超分辨率模型中的鉴别器 (discriminator) 设计存在一个普遍的问题, 即训练不稳定问题。针对以上问题作出两点改进, 一是提出多损失融合的方法, 寻求一种在 PSNR 指标与感知质量之间的平衡。通过将均方误差损失、感知损失、风格损失与对抗损失进行融合的方法, 在提高 PSNR 值的同时, 改善图像视觉质量。二是在基于生成对抗网络的超分辨率模型的鉴别器设计中引入谱归一化 (spectral normalization), 以实现更稳定有效的训练。结果显示, 改进后的方法得到了更高的 PSNR 指标与更逼真的视觉感知质量, 并进一步表明感知质量对于超分辨率重建的重要性。

关键词: 多损失融合; 谱归一化; 图像超分辨率

中图分类号: TP391 doi: 10.19734/j.issn.1001-3695.2019.02.0095

Multi-loss ensemble and spectral normalization for image super-resolution

Xu Ningning, Zheng Kai

(College of Computer Science & Software Engineering, East China Normal University, Shanghai 200062, China)

Abstract: Recently, the objective measurement index of image super-resolution has been improved continuously, but the quality of visual perception is still smooth. And there is a general problem with the discriminator design in the application of the super-resolution model, which is the instability of its training. Two improvements are made to the above problems. One was proposed a method of multi-loss ensemble, seeking a balance between PSNR indicators and perceived quality. By blending the mean square error loss, perceptual loss, style loss and adversarial loss, the PSNR value is improved while the visual quality is improved. The second is to apply Spectral Normalization in the discriminator design of the GAN-based super-resolution model to achieve more stable and effective training. The results show that the improved method yields a higher PSNR indicator and a more realistic visual perception quality, and further demonstrates the importance of perceived quality for super-resolution reconstruction.

Key words: multi-loss ensemble; spectral normalization; image super-resolution

0 引言

单幅图像超分辨率 (single image super-resolution, SISR) 的目标是将低分辨率 (low-resolution, LR) 的图像重建为高分辨率 (high-resolution, HR) 图像。近期, 深度学习在计算机视觉领域的进展^[1-6]已经广泛影响到超分辨率领域^[7-14], 相较于其他方法, 基于学习的方法取得了快速的进展。通常做法是使用深度神经网络对输入的图像进行抽象特征表达, 以全监督的方式进行训练, 通过基于数据的学习得到从 LR 到 HR 的非线性映射, 最后完成高分辨率的重建。

峰值信噪比 (peak signal-to-noise ratio, PSNR) 一直以来被用以衡量结果图像质量优劣, 结构相似性 (structural similarity, SSIM) 同样是一种比较常用的用于衡量两幅图像相似性的客观指标, 但是两者不可避免的都存在局限性, 在图像超分辨率领域, 这种局限性就体现在客观指标高的生成图像在视觉质量上不一定是最好的, 如图 1 所示, 重建的高指标图像可能存在视觉感知上过于平滑, 不具逼真纹理等缺点。

基于特征空间而非像素空间的感知相似性, 由 Dosovitskiy 等人^[15]提出并应用于图像生成领域, 之后 Gatys

等人^[16]对其进行改进并应用在图像风格迁移领域中, 用于分别衡量生成图像与原始图像的风格特征图与内容特征图之间的特征距离, 而非逐像素距离, 实验取得了开创性的进展。Johnson 等人^[9]在此基础上提出感知损失 (perceptual loss) 并在图像超分辨率领域得到应用, 在随后的基于感知损失的模型通常不再使用基于均方误差 (mean squared error, MSE) 的像素空间损失, 转而使用基于 VGG 模型^[17]的图像内容感知损失, 在特征空间优化感知损失函数, 用以增强重建图像的真实感, 但是重建结果与真实图像之间仍存在巨大的差距。

生成对抗网络 (generative adversarial networks, GAN)^[18]在图像生成工作中具有显著的优点, 能够生成具有逼真纹理的图像, 被引入到图像超分辨率研究中同样有助于得到更自然的图像重建结果。Ledig 等人^[19]提出的 SRGAN 模型成功将 GAN 模型应用在超分辨率实践中, 并在 GAN 架构中通过感知损失进行优化, 显著提高了重建图像的视觉质量。但是基于 GAN 框架的方法同时也增加了模型训练不稳定的问题。针对以上问题, 本文提出以下两点改进:

1) 多损失融合

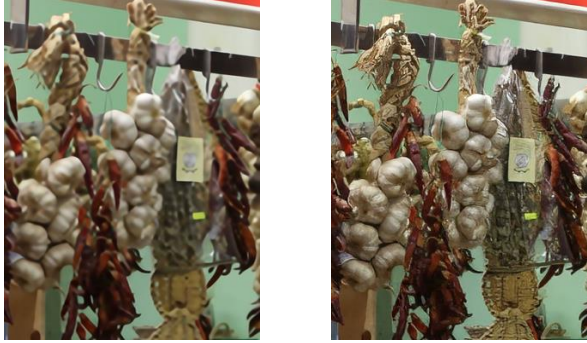
趋向 PSNR 的方法使用基于均方误差的像素损失, 而评价指标 PSNR 的计算原理又是基于逐像素均方误差的, 因而

收稿日期: 2019-02-10; 修回日期: 2019-03-27

作者简介: 许宁宁 (1994-), 女, 浙江台州, 硕士研究生, 主要研究方向为计算机视觉 (51174507023@stu.ecnu.edu.cn); 郑凯 (1968-), 男, 浙江宁波, 副教授, 硕士, 博士, 主要研究方向为计算机网络、云计算。

不断训练后的模型能够得到较高的 PSNR 指标, 但是在视觉感知却因缺少细节纹理而显得过于平滑;

趋向感知损失的方法能够使图像在视觉质量上得到改进, 但是客观指标的评价值又较低。所以本文对两种不同侧重的损失进行加权融合, 丰富损失函数的内涵, 经过训练后的模型能够在提升客观评价指标的情况下, 拥有更好的视觉感知质量。



(a)PSNR=26.761/SSIM=0.848 (b)PSNR=24.341/SSIM=0.776

图 1 高指标低视觉质量与低指标高视觉质量对比图

Fig. 1 High index low visual quality and low index high visual quality comparison chart

2) 谱归一化

基于 GAN 框架的模型, 存在训练过程不稳定的问题, 参数矩阵的谱范数被证明将会影响模型的泛化能力^[20], Miyato 等人^[21]提出的谱归一化 (spectral normalization) 的作用是限制鉴别器的 Lipschitz 范数, 防止参数矩阵在训练过程中对一个特定的方向敏感, 所以本文对原有 SRGAN 模型的鉴别器进行谱归一化处理, 相较于处理之前, 能够得到质量更优的生成结果。

1 相关理论

单幅图像超分辨率研究是图像处理领域, 也是计算机视觉领域的一个基础问题, 正在受到学界的广泛关注。近期深度学习的发展使得能够学习从 LR 图像到 HR 图像的更强大的映射功能。超分辨率卷积神经网络 SRCNN^[22]作为一个成功应用于图像超分辨率的深度模型, 也是首个应用卷积神经网络(convolutional neural networks, CNN)的端到端超分辨率方法, 在图像重建质量上明显优于先前的模型, 开启图像超分辨率的深度学习方法。

1.1 基于卷积神经网络的图像超分辨率

Dong 等人^[22]首次提出了应用卷积神经网络的图像超分辨率方法(SRCNN), 使得速度与重建结果相较传统方法都得到提升。随后, 随着网络结构的加深, 实验取得了越来越好的结果, Kim 等人^[14]在 VDSR 方法中通过创建浅层与深层网络之间的跳跃连接, 进行残差学习并增大学习率, 并表明更深的网络可以实现更好的准确性。Lim 等人^[24]在 EDSR 方法中引入残差网络结构并去除原有结构中 BN 层, 在标准数据集上取得更好结果。Shi 等人^[25]提出了一种子像素卷积层, 将从 LR 空间到 HR 空间的特征映射进行聚合, 使得可以直接将 LR 图像传输到网络中, 而不需要对 LR 图像进行双三次上采样, 从而减少了存储器使用和处理时间。这些方法在 PSNR 指标上取得了很高的评价。然而, 它们仍然无法恢复高频信息, 因为它们依赖于最小化 MSE 损失, 这导致了重建图像为均值最优而显得过于平滑, 视觉感知质量亟待提高。

1.2 感知损失

基于感知空间而非像素空间的感知驱动方法被证明可以

改善重建图像的视觉感知结果^[15], Johnson 等人^[9]提出基于感知损失函数的图像风格迁移和图像超分辨率的方法。感知损失衡量从图像识别网络 (例如 VGG 网络^[17]) 提取的两个图像的特征图之间的特征差异。实验结果表明^[9, 26], 最小化感知损失会导致较低的 PSNR, 但可以得到了更好的图像细节和边缘。但是, 这些方法无法完全恢复输入图像中丢失的高频细节。

1.3 生成对抗网络

生成对抗网络在最近被应用于 SISR, 以在超分辨率图像中合成感知上更逼真的高频细节。Ledig 等人^[19]提出的 SRGAN 方法是提升视觉质量的具有里程碑意义的结果之一, 在 GAN 架构中使用残差块并通过感知损失进行优化, 显著提高了重建图像的视觉质量。Sajjadi 等人^[26]扩展了 Ledig 等人的工作, 引入了一种受风格损失启发的纹理匹配损失^[16], 以促进重建的超分辨率图像具有与真实 HR 图像相同的纹理样式。虽然这些方法无法完全恢复输入图像中丢失的高频细节, 但它们却合成了高频细节, 因此结果看起来更具逼真的细节。但是, 这些方法倾向于产生任意高频伪像。此外, 这些基于 GAN 的 SR 方法采用感知损失, 最小化 VGG 特征的特征图 MSE。类似于像素上的 MSE, 简单地最小化 VGG 特征的特征图 MSE 将不足以完全表示特征图的实际特征。

2 多损失融合与谱归一化方法

本文在 SRGAN 模型基础上对损失函数进行改进, 进行多损失融合, 并对原有生成对抗网络中的鉴别器设计进行改进, 增加谱归一化处理, 通过稳定网络训练得到更优重建结果, 改进的作用过程如图 2 所示。改进的目标在于生成拥有较高 PSNR 指标的图像的同时, 改善超分辨率重建后图像的视觉感知质量。

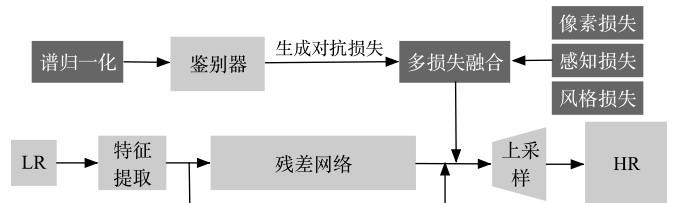


图 2 改进的作用过程

Fig. 2 improvement process

2.1 重建过程

对于给出的单幅低分辨率图像 x 和高分辨率图像 y , 超分辨的目标在于重建一个高分辨率的图像 \hat{y} , 使得其与原始图像 y 的差异最小。

模型中生成网络首先利用一个 3×3 的卷积神经网络从低分辨率图像 x 中提取特征图, 此过程可表示为

$$F_0 = f(x) \quad (1)$$

其中 f 表示特征提取函数, F_0 表示提取的特征, 同时作为下一阶段的输入。网络下一阶段是由多个残差块组成的链式模型。此过程可以表示为

$$F_k = R_k(F_{k-1}) \quad k=1,2,\dots,n \quad (2)$$

其中 R_k 表示第 k 个残差处理, F_{k-1} 和 F_k 分别表示第 k 个残差块的输入和输出。网络的最后一个阶段是将全局残差学习的结果与双三次插值的结果相加, 最终通过上采样得到超分辨率重建结果 \hat{y} , 此过程可表示为

$$\hat{y} = S(F_n + B(x)) \quad (3)$$

$$\hat{y} = G_\theta(x) \quad (4)$$

其中 B 表示对低分辨率图像的双三次插值处理, S 表示亚像素卷积 (sub-pixel convolution) ^[25], G 表示由 θ 参数化的生成

网络。

2.2 多损失融合

(x_i, y_i) 表示训练数据集中的训练对, 为了估计 \hat{y} , 需要通过最小化损失函数 \mathcal{L} 在训练过程中不断优化网络的参数 θ , 即:

$$\hat{\theta} = \arg \min_{\theta} \sum_i \mathcal{L}(\hat{y}_i, y_i) \quad (5)$$

评价指标峰值信噪比基于 MSE, 优化 MSE 为最小, 将使最终结果返回众多可能解的均值, 使得重建图像缺少自然纹理且因过于平滑而显得缺少真实感。

在本文中, 像素空间损失、感知损失、风格损失和对抗损失被用来提高重建图像的细节, 改善由传统像素空间均方误差损失造成的图像过于平滑问题。

2.2.1 像素空间损失

基于像素空间的均方误差损失是在图像超分辨率领域被最为广泛使用的损失函数, 最小化此损失函数有助于得到最优的 PSNR 指标。

$$\mathcal{L}_{\text{pixel}} = \sum_i \hat{y}_i - y_i^2 \quad (6)$$

2.2.2 感知损失

感知损失衡量的是图像在特征表示中的相似性, 感知损失相较逐像素损失在衡量图像相似性上具有更好的视觉感知结果。与文献[19]中相似, 本文使用预训练的 VGG-19 网络第五层池化层前的第四层卷积层作为特征映射。

$$\mathcal{L}_{\text{perceptual}} = \sum_i \phi(\hat{y}_i) - \phi(y_i)_2^2 \quad (7)$$

其中 ϕ 表示图像在 VGG-19 网络中的特征映射。

2.2.3 风格损失

Gatys 等人^[16]在图像风格迁移领域对提取的高层特征求解 Gram 矩阵来表示图像的风格特征, 风格特征同时刻画了图像的纹理特征, 所以最小化风格损失有助于增强重建图像与原始高分辨率图像的纹理相似性:

$$\mathcal{L}_{\text{style}} = \sum_i \text{Gram}(\phi(\hat{y}_i)) - \text{Gram}(\phi(y_i))_2^2 \quad (8)$$

其中 $\text{Gram}(\phi)$ 表示特征映射 ϕ 的 Gram 矩阵。

2.2.4 对抗损失

生成网络的目标在于生成逼真的图像使鉴别网络无法在重建图像与真实图像间作出区分。加入对抗损失有助于得到更加逼真的超分辨率重建结果, 对抗损失来源于鉴别网络对生成样本真实程度的判别:

$$\mathcal{L}_D = \sum_i \lg(1 - D_i(G_o(x_i))) \quad (9)$$

其中 $D_i(G_o(x_i))$ 表示重建结果为原始高分辨率图像的概率,

D 表示由 λ 参数化的鉴别网络。

2.2.5 总损失

趋向 PSNR 的损失函数能够得到较高的结果指标, 但是会导致视觉模糊与平滑, 趋向视觉感知的损失函数能够提高感知质量, 但是评价指标较低, 且缺少生成约束。因此为了平衡 PSNR 指标与视觉感知质量, 总损失融合了以上像素空间损失、感知损失、风格损失和对抗损失:

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{pixel}} + \beta \mathcal{L}_{\text{perceptual}} + \gamma \mathcal{L}_{\text{style}} + \delta \mathcal{L}_D \quad (10)$$

其中 $\alpha, \beta, \gamma, \delta$ 为平衡总损失的多损失融合系数, 以此来达到归一化的目的, 通过这些系数使得不同损失函数对总损失的影响权重接近。

2.3 谱归一化

生成对抗网络被应用于图像超分辨率研究之后, 进一步改进了重建图像的质量, 但是同时也引入了生成对抗模型普

遍存在的训练不稳定问题。许多研究表明 GAN 框架的这种不稳定源自网络中鉴别网络的性能控制问题, 在超分辨率应用中, 此问题的极端举例就是鉴别网络相较于生成网络过早进入完美状态, 总是能够完美鉴别超分辨率重建结果与原始高分辨率图像, 导致鉴别网络无法向生成网络提供梯度信息, 生成网络的训练也就停止了。Miyato 等人^[21]提出的谱归一化通过对鉴别网络中的参数进行归一化处理, 使得映射函数的梯度满足 Lipschitz 约束, 以稳定鉴别网络的训练。

本文将谱归一化应用到超分辨率模型的鉴别网络中, 其中鉴别器的网络层可表示为

$$x_L = a_L(W_L x_{L-1} + b_L) \quad (11)$$

其中 x_{L-1} 与 x_L 分别表示网络 L 层的输入与输出, W_L 表示参数矩阵, a_L 表示激活函数, b_L 表示偏差。

鉴别网络可用函数 D 表示为

$$D(x, \eta) = a_L(W_L(a_{L-1}(W_{L-1}(\dots a_1(W_1 x + b_1) \dots)))) \quad (12)$$

其中 η 表示各层网络参数 $\{W_1, \dots, W_L\}$ 。

鉴别网络的改进思想是限制参数 η 使得函数 D 满足 Lipschitz 约束。满足 Lipschitz 约束, 可以降低模型对扰动的敏感, 增强模型对输入扰动的稳定性, 提高模型的泛化能力, Lipschitz 约束的含义可表示为

$$\frac{D(x_1, \eta) - D(x_2, \eta)}{x_1 - x_2} \leq C \quad (13)$$

其中 C 表示一个常数。

Arjovsky 等人^[26]提出的 Wasserstein GAN 的限制 Lipschitz 约束的方法是简单的最大值约束, 即通过消减过大参数以保证矩阵中元素不超过给定值, 但是这同时也会破坏参数之间的数值比例关系, 所以本文使用另一种方法——谱归一化进行改进。

鉴别器中的多层神经网络可以通过多个单层神经网络进行递归分析, 对于网络中每一层的参数矩阵, 谱范数表示矩阵的最大奇异值, 可表示为

$$\sigma(W) := \sqrt{\max\{\lambda_i(W^H W)\}} \quad (14)$$

其中 W^H 表示 W 的转置共轭矩阵, λ 表示矩阵的特征值。

谱归一化即将参数矩阵除以它的谱矩阵, 以到达在不破坏矩阵内部参数关系的情况下, 满足 Lipschitz 约束, 可表示为

$$\bar{W}_{\text{SN}}(W) := \frac{W}{\sigma(W)} \quad (15)$$

当鉴别网络中的所有参数矩阵都经过谱归一化处理后, 相较于常规的网络, 类似于引入一个新的正则化项, 防止在训练过程中对特定方向敏感, 有助于生成网络达到更好的训练效果, 增加模型的泛化能力, 从而得到更加逼真的超分辨率重建结果。

3 实验

3.1 数据集

本文使用 DIV2K 和 Flickr2K 数据集进行训练, DIV2K 数据集包含 800 张 2K 高分辨率图像, 100 张 2K 验证图像和 100 张 2K 测试图像。Flickr2K 数据集包含 2650 张 2K 高分辨率图像, 并通过随机裁剪, 翻转和转置进行数据增强。对应低分辨率图像通过对高分辨率图像进行缩放因子为 4 的双三次插值降采样获取。对于测试, 本文使用标准数据集为 Urban100, DIV2K-valid 和 Manga109, 这些数据集中, DIV2K-valid 由自然场景图像组成, Urban100 包含具有不同频段细节的城市场景图像, Manga109 由日本漫画图像组成。

3.2 模型训练

用于训练的 LR-HR 图像对为由原始高分辨率图像经随机裁剪后大小为 128×128 的子图像和对应 32×32 的降采样子图像组成的图像对, 实验采用的降采样因子为 4, 并且缩放 LR 输入图像的范围到 $[0,1]$, HR 图像到 $[-1,1]$ 。多损失融合系数设置为 $\alpha=1e-2$, $\beta=1$, $\gamma=1e-3$, $\delta=5e-3$ 。本文使用 Adam 优化算法, 指数衰减率 $\beta_1=0.9$, $\beta_2=0.999$, 稳定常数 $\epsilon=1e-8$ 。训练时设置总迭代次数为 $5e5$, 批大小为 16, 初始学习率为 $1e-4$, 当分别经过 $5e4$, $1e5$, $2e5$, $3e5$ 次迭代后将学习率减半。实验环境为 NVIDIA Tesla P100 GPU, 完成 DIV2K 和 Flickr2K 数据集 50 万次迭代训练的完整模型的训练时长为 11 天, 完成 DIV2K 数据集 30 万次迭代训练的性能验证模型的训练时长为 5-6 天不等, 包括性能验证实验的 5 个模型在内, 本文共训练了 6 个不同模型。

3.3 结果

3.3.1 多损失融合性能验证

为了验证多损失融合的性能, 本文在相同条件下训练了应用不同损失函数的多个模型, 将多损失融合模型与表 1 中的不同组合损失模型进行比较, 使用 DIV2K 中包含 100 张 2K 图像的 DIV2K-valid 集测试, 其中性能验证模型设计如表 1 所示。

表 1 多损失融合性能验证模型设计

Table 1 Model design for multi-loss ensemble performance verification

模型	损失函数	描述
Net-PID	$\mathcal{L}_{Pixel} + \mathcal{L}_D$	像素损失+对抗损失
Net-PRD	$\mathcal{L}_{Perceptual} + \mathcal{L}_D$	感知损失+对抗损失
Net-SD	$\mathcal{L}_{Style} + \mathcal{L}_D$	风格损失+对抗损失
Net-E	$\mathcal{L}_{Pixel} + \mathcal{L}_{Perceptual} + \mathcal{L}_{Style} + \mathcal{L}_D$	多损失融合

实验设置 300000 次迭代, 每经过 5000 次训练后保存并输出网络模型, 并对模型在 DIV2K-valid 数据集中进行 PSNR 指标评价。结果显示使用多损失融合的模型相较于使用其他损失函数的模型, 在实验迭代训练过程中, 拥有最高的 PSNR 评价指标, 其中比较结果如图 3 所示。

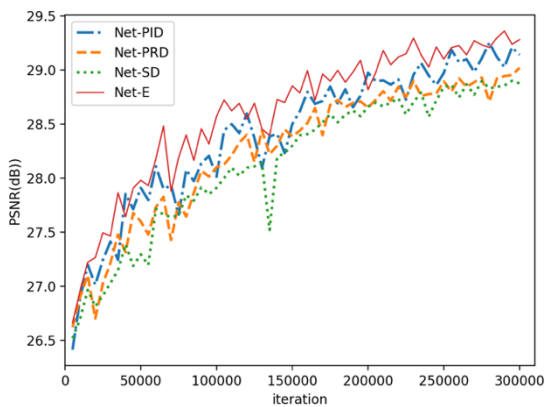


图 3 多损失融合性能验证 PSNR 指标比较结果

Fig. 3 Comparison of PSNR indicators for multi-loss ensemble performance verification

同时, 包含感知损失的模型拥有更好的视觉感知质量, 四种使用不同损失函数的模型都有较高的感知质量, 但是细节中仍有较大不同。Net-PID 模型的重建结果因缺少足够的高频信息, 图像缺少纹理细节。Net-PRD 模型生成结果存在较多伪影。Net-SD 模型应用风格损失, 无法生成更多高频信息, 和许多风格迁移图像一样, 导致重建图像过于平滑。Net-E 模型相较于其他模型, 拥有更加逼真的视觉重建结果, 改善

了过于平滑的现象, 细节中包含更多的高频信息, 其中比较结果如图 4 所示。

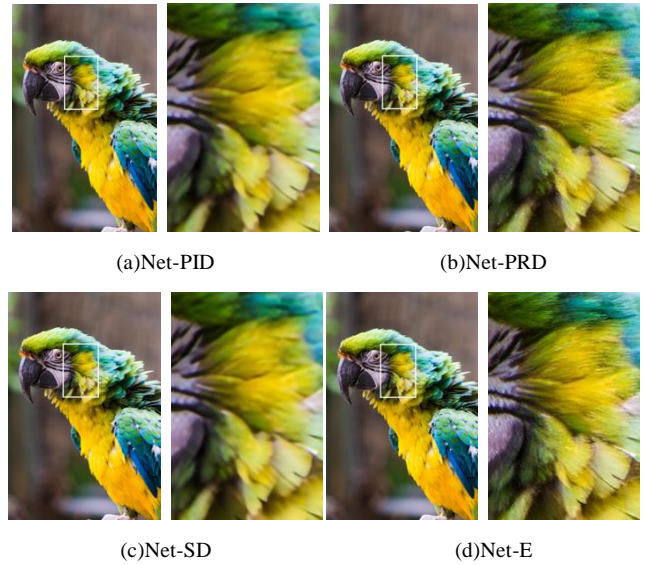


图 4 多损失融合性能验证视觉质量比较

Fig. 4 Comparison of visual quality of multi-loss ensemble performance verification

3.3.2 谱归一化性能验证

为了验证谱归一化的性能, 本文训练了经过谱归一化处理与未经过处理的两个模型进行比较, 验证模型设计如表 2 所示。

表 2 谱归一化性能验证模型设计

Table 2 Model design for spectral normalization performance verification

模型	描述
Net-SN	鉴别器应用谱归一化
Net-NoSN	鉴别器未应用谱归一化

实验设置 300000 次迭代, 每经过 5000 次训练保存并输出模型, 并对模型在 DIV2K-valid 数据集中进行 PSNR 指标检测, 结果显示鉴别器经过谱归一化处理后的模型, 在 300000 次迭代训练过程中, 在 DIV2K-valid 数据集中的 PSNR 评价指标要明显高于未经过谱归一化的模型, 其中比较结果如图 5 所示。

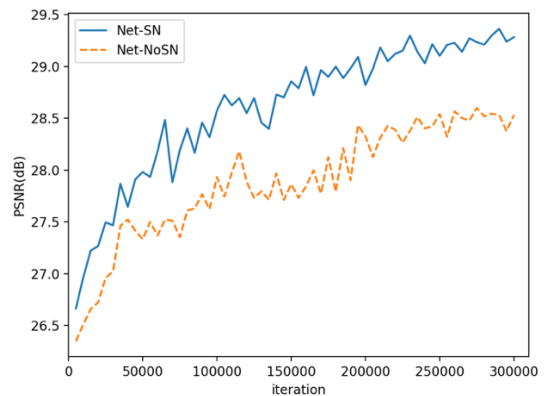


图 5 谱归一化性能验证 PSNR 指标结果比较

Fig. 5 Comparison of psnr indicator results for spectrum normalization performance verification

应用谱归一化的 Net-SN 模型, 通过谱范数对鉴别器的参数矩阵进行归一化, 使其满足 Lipschitz 约束, 改进了 GAN 的训练稳定性, 提高网络的性能。结果显示 Net-SN 模型的重建图像具有更清晰逼真的细节纹理, 其中比较结果如图 6

所示。



(a)Net-SN



(b)Net-NoSN

图 6 谱归一化性能验证视觉质量比较

Fig. 6 Comparison of visual quality of spectral normalization performance verification

3.3.3 与其他模型比较

将本文的方法 ELSN-SRGAN 分别与 bicubic, SRCNN^[22], VDSR^[14]和 SRGAN^[19]进行比较, 其中, VDSR 为趋向 PSNR 指标的方法, SRGAN 为趋向视觉感知质量的方法。在公开数据集 Urban100、DIV2K-valid 和 Manga109

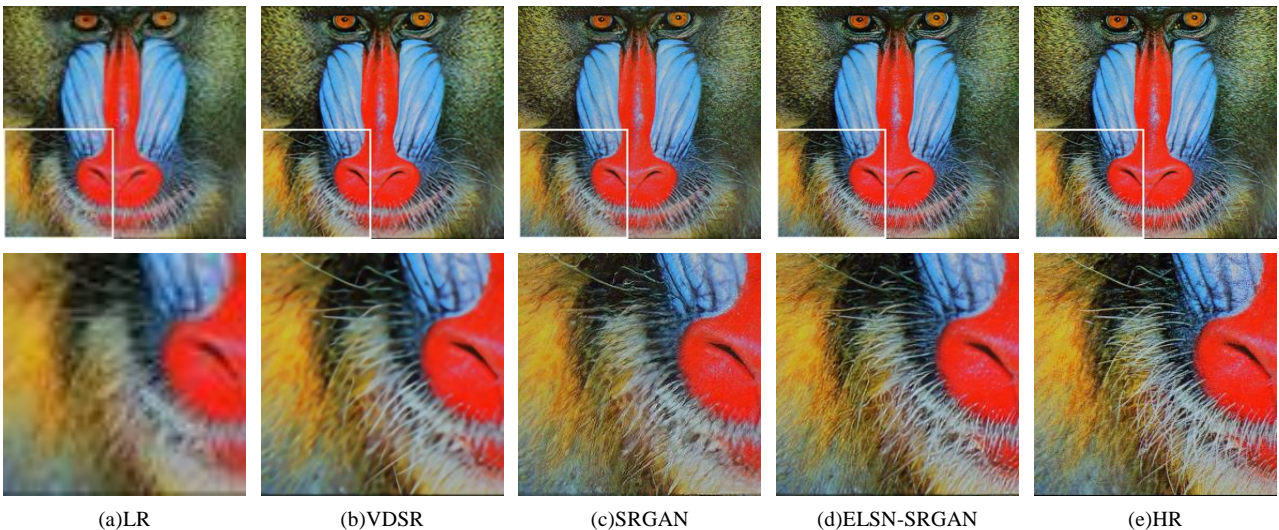
中做×4 倍超分辨率重建。实验结果显示本文的方法相较于未经过本文方法改进的 SRGAN 拥有更高的 PSNR/SSIM 指标, 且数值接近甚至超过了趋向 PSNR 的方法 VDSR, 同时, 本文的视觉感知质量远高于以上方法, 详细比较结果如表 3 与图 7 所示。

表 3 不同模型在不同数据集测试对比结果

Table 3 Different models test comparison results in different data sets

数据集	方法	PSNR	SSIM
Urban100	bicubic	23.14	0.6577
	SRCNN	24.52	0.7221
	VDSR	25.18	0.7524
	SRGAN	24.42	0.7325
	ELSN-SRGAN	25.58	0.7689
Manga109	bicubic	24.92	0.7866
	SRCNN	27.66	0.8555
	VDSR	28.82	0.8861
	SRGAN	28.09	0.8610
	ELSN-SRGAN	29.68	0.8842
DIV2K-valid	bicubic	26.66	0.8521
	SRCNN	27.78	0.8753
	VDSR	28.17	0.8841
	SRGAN	28.13	0.7744
	ELSN-SRGAN	29.17	0.7991

结果显示在在公开数据集 Urban100、DIV2K-valid 和 Manga109 中本文方法 ELSN-SRGAN 拥有最高的 PSNR 指标, 且 SSIM 指标接近于趋向指标的模型 VDSR, 且本文方法的图像重建结果比趋向视觉质量的模型 SGRAN 拥有更高的指标数值和更好的视觉感知结果, 在图像细节上具有更加逼真的纹理。



(a)LR

(b)VDSR

(c)SRGAN

(d)ELSN-SRGAN

(e)HR

图 7 不同模型视觉结果对比

Fig. 7 Comparison of visual results of different models

4 结束语

本文提出了一种改进的单幅图像超分辨率方法, 通过多损失融合和谱归一化提高了重建图像的 PSNR/SSIM 指标, 并且提高了视觉感知质量, 生成的图像达到了与 HR 图像接近的逼真结果, 做到了同时在 PSNR 评价指标与视觉感知两个方面改进超分辨率重建结果。

参考文献:

[1] Gao Huang, Zhuang Liu, Van Der Maaten L, et al. Densely connected

convolutional networks [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 2261-2269.

[2] Bulat A, Tzimiropoulos G. Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 109-117.

[3] Ilg E, Mayer N, Saikia T, et al. FlowNet 2. 0: Evolution of optical flow estimation with deep networks [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 1647-1655.

- [4] Bulat A, Tzimiropoulos G. Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 109-117.
- [5] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [6] Shrivastava A, Pfister T, Tuzel O, *et al.* Learning from simulated and unsupervised images through adversarial training [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . IEEE, 2017: 2242-2251.
- [7] Wang Y, Shen J, Zhang J. Deep Bi-Dense Networks for Image Super-Resolution [C]// 2018 Digital Image Computing: Techniques and Applications (DICTA) . IEEE, 2018: 1-8.
- [8] Tai Ying, Yang Jian, Liu Xiaoming. Image super-resolution via deep recursive residual network [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . IEEE, 2017: 2790-2798.
- [9] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution [C]// European Conference on Computer Vision. Springer, Cham, 2016: 694-711.
- [10] Dong Chao, Loy C, He Kaiming, *et al.* Image super-resolution using deep convolutional networks [J]. IEEE trans on pattern analysis and machine intelligence, 2016, 38 (2): 295-307.
- [11] Haris M, Widyanto M R, Nobuhara H. Inception learning super-resolution [J]. Applied optics, 2017, 56 (22): 6043-6048.
- [12] Kappeler A, Yoo S, Dai Q, *et al.* Video super-resolution with convolutional neural networks [J]. IEEE Trans on Computational Imaging, 2016, 2 (2): 109-122.
- [13] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1646-1654.
- [14] Dosovitskiy A, Brox T. Generating images with perceptual similarity metrics based on deep networks [C]// Advances in Neural Information Processing Systems. 2016: 658-666.
- [15] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2414-2423.
- [16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409.1556, 2014.
- [17] Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets [C]// Advances in neural information processing systems. 2014: 2672-2680.
- [18] 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) . IEEE, 2017: 105-114.
- [19] Yoshida Y, Miyato T. Spectral Norm Regularization for Improving the Generalizability of Deep Learning [J]. arXiv preprint arXiv: 1705.10941, 2017.
- [20] Miyato T, Kataoka T, Koyama M, *et al.* Spectral normalization for generative adversarial networks [J]. arXiv preprint arXiv: 1802.05957, 2018.
- [21] Dong Chao, Loy C, He Kaiming, *et al.* Image super-resolution using deep convolutional networks [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2016, 38 (2): 295-307.
- [22] Dong Chao, Loy C C, Tang Xiaoou. Accelerating the super-resolution convolutional neural network [C]// European Conference on Computer Vision. Springer, Cham, 2016: 391-407.
- [23] Lim B, Son S, Kim H, *et al.* Enhanced deep residual networks for single image super-resolution [C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops. 2017, 1 (2): 4.
- [24] 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. 2016: 1874-1883.
- [25] Sajjadi M S M, Schölkopf B, Hirsch M. Enhancenet: Single image super-resolution through automated texture synthesis [C]// Computer Vision (ICCV) , 2017 IEEE International Conference on. IEEE, 2017: 4501-4510.
- [26] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks [C]// International Conference on Machine Learning. 2017: 214-223.