

文章编号: 2095-2163(2022)01-0155-06

中图分类号: TP391

文献标志码: A

改进的生成对抗网络的超分辨率方法研究

朱煜君¹, 迟杏^{1,2}, 任志强¹, 张欣¹

(1 贵州大学 大数据与信息工程学院, 贵阳 550025; 2 贵州装备职业技术学院 电气自动化系, 贵阳 551499)

摘要: 本文针对基于深度学习的超分辨率重建网络 SRGAN 中存在的训练不稳定, 重建图像信息缺失等问题进行了分析研究, 做出相应的改进, 提出了一种改良的生成器残差块结构, 提高了重建效率, 增强了重建图像的质量; 引入 Wasserstein 距离代替 JS 散度来描述数据分布之间的差异, 解决了生成对抗网络训练不稳定的问题, 有效避免了梯度消失现象; 在内容损失函数方面, 选用更适合超分辨率重建任务的 L1 损失函数取代 MSE 函数, 均化了对误差的惩罚, 使细节更加丰富。在 Set5、Set14 与 BSD100 数据集上, 经过与原网络的对比实验结果证明, 平均 PSNR 最高峰值信噪比提高 1.8 dB, SSIM 结构相似度提升 6%, 证明了改进方法的有效性。

关键词: 超分辨率; 深度学习; 生成对抗网络; Wasserstein 距离

Research on improved super-resolution method of generative adversarial network

ZHU Yujun¹, CHI Xing^{1,2}, REN Zhiqiang¹, ZHANG Xin¹

(1 School of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China;

2 Department of Electrical Automation, Guizhou Equipment Vocational and Technical College, Guiyang 551499, China)

[Abstract] In view of the instability of training in the learning-based super-resolution reconstruction network SRGAN, the lack of information in the reconstructed image, etc. were analyzed and researched, and corresponding improvements were made. An improved generator residual block structure was proposed to speed up the reconstruction. Efficiency, enhances the quality of the reconstructed image; introduces Wasserstein distance instead of JS divergence to describe the difference between data distributions, solves the problem of instability in GAN network training, and effectively avoids the phenomenon of gradient disappearance; it is more suitable for content loss function selection. The L1 loss function of the super-resolution reconstruction task replaces the MSE function, which equalizes the error penalty and makes the details richer. The experimental results show that in the comparison with the original network on the Set5, Set14 and BSD100 data sets, the average PSNR highest peak signal-to-noise ratio is increased by 1.8dB, and the SSIM structure similarity is increased by 6%, which proves the effectiveness of the improved method.

[Key words] super resolution; deep learning; generative adversarial network; Wasserstein distance

0 引言

分辨率是评价图像清晰度的重要指标之一, 又可以分为显示分辨率与图像分辨率两种, 本文研究的内容聚焦于图像分辨率。拥有高分辨率的图像能够给人们展示出更多的信息, 使人能更快、更精准的定位所需要的部分, 图像的超分辨率重建技术应运而生, 即由单帧或序列的低分辨率 (Low Resolution, LR) 图像生成视觉效果更好的高分辨率图像 (High Resolution, HR), 在天文, 医学影像等领域都有着广泛的应用^[1]。图像的超分辨率技术经历了几个重要的发展阶段: 基于插值重建的方法, 该方法简单、易操作, 但重建后的图像较为模糊, 且存在振铃效

应, 重建效果较差^[2-3]; 基于深度学习的方法属于传统的机器学习领域, 通过神经网络强大的特征提取和数据表示能力学习高分辨率图像与低分辨率图像之间的普遍差异与联系, 构建出统计学模型, 这种方法生成的高分辨图片的效果更好, 方法更具有优势^[4]。2014年, Dong 等提出基于卷积神经网络的图像超分辨率模型 SRCNN (Super-Resolution Convolutional Neural Network), 通过在卷积层上进行特征映射得到超分辨率图像, 但因其网络深度不够、特征感受野小等问题, 当放大倍数过大时, 造成的生成图像过于平滑^[5]; 同年十月, 由 Goodfellow 等提出的生成对抗网络 (Generative Adversarial Network, GAN), 给基于学习的图像超分辨率技术提供了一

基金项目: 国家自然科学基金(61865002, 62065002); 贵州大学“双一流”研究重大项目(GDSYL2018001)。

作者简介: 朱煜君(1996-), 男, 硕士研究生, 主要研究方向: 深度学习、图像处理。

通讯作者: 张欣 Email: zyj_0702@qq.com

收稿日期: 2021-09-17

种全新的思路^[6];随后, Ledig 等将超分辨率重建技术与 GAN 相结合,提出了基于生成对抗网络的图像超分辨率 SRGAN (Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial),首次提出了内容损失加对抗损失的组合,提升了生成图像的真实感和感知相似性,具有较强的重建效率以及更好的重建效果^[7]。但 SRGAN 同时也遗留着生成对抗网络训练难收敛,网络易崩溃等问题,本文针对此问题,分析原因后做出相应优化,稳定了训练过程,加快了收敛;针对生成器输出图像信息不稳定的情况,改进了生成器网络结构,提高了网络的特征表达能力;同时引入更适合超分辨率重建的损失函数,提升了重建后图像的视觉效果。实验证明,对比于原网络,相应评价指标明显增高。

1 生成对抗网络

1.1 GAN 的模型理论

GAN 的核心思想来源于博弈论中的纳什均衡^[8-9]，“博弈”双方是生成对抗网络的基本组成部分:生成器 (Generator, G) 与判别器 (Discriminator, D)。“博弈”的内容是生成器通过学习输入的随机分布信息,进而学会伪造数据,再判别器对真实数据和生成的伪造数据进行判别。生成器会不断学习模仿真实数据的分布,而判别器就是不断的学习如何识别真假数据。双方不断的反馈、学习,优化过程就类似寻找二者之间的纳什均衡。

对于生成器而言,伪造出的数据越接近真实数据,表示生成器训练效果越好,当双方“博弈”达到动态均衡时取得最优解,此时的判别器已经无法辨别真实图像与生成器重建出的图像。

1.2 SRGAN 的模型

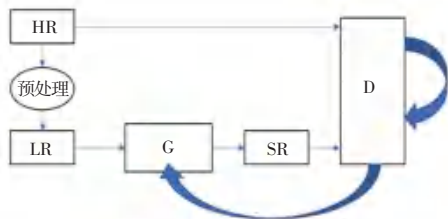
SRGAN 借助 GAN 的思想,把 GAN 模型引入图像的超分辨率领域,将原高分辨图像通过下采样预处理后得到的低分辨率图像送入生成器,生成器会模仿并伪造出超分辨率图像,判别器为了区分原高分辨率图像与伪造的超分辨率图像不断的学习优化,根据判别器的反馈结果不断更新,当判别器无法正确判别时,生成器训练完成。具体 SRGAN 流程如图 1 所示。

2 相应的改进

2.1 改进的网络结构

BN 层能一定程度上加速网络收敛,防止过拟合的出现,被广泛应用于图像分割、检测等领域,但由

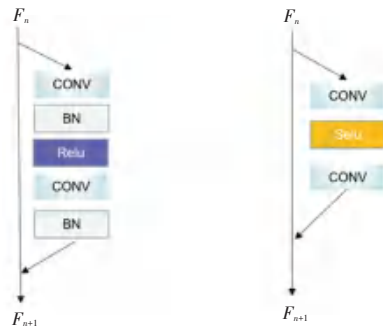
于 BN 层做标准化 (normalization) 的目标是浅层特征上的每一位像素点,对于超分辨率重建任务而言,堆积过多的 BN 层则会导致网络提取特征图像时原色彩分布被归一化,混淆了原始图像的对比度,不利于超分辨率图像的重建。通过进行对比实验,将生成器网络残差块中的规范化层 BN (Batch Normalization) 层删除,将原残差块中激活函数换成自适应能力更好的缩放指数型线性单元 *Selu* 函数^[10],残差块对比如图 2 所示。



HR: 原高分辨率图像; LR: 预处理后的低分辨率图像; G: 生成器; SR: 超分辨率图像; D: 判别器

图 1 SRGAN 网络流程

Fig. 1 SRGAN network flow



(a) 原残差块结构图 (b) 改进后残差块结构图

图 2 残差块对比图

Fig. 2 Comparison of residual blocks

通过后续实验证明,使用 *Selu* 函数的实际效果更为理想。16 个改进后的残差块组成的深层残差网络模块利用其强大的特征提取能力,提取的特征信息通过由 Sub-pixel Layer (亚像素卷积层) 为主构造的上采样块将其恢复成放大倍数 4 倍的超分辨率图像,为了得到更好的拟合效果,将原输出激活函数 *Sigmoid* 替换成双曲正切曲线 *Tan h* 函数,结构如图 3 所示。

2.2 改进的损失函数

原始的 GAN 是用并不连续的 *JS* 散度去描述真实分布与生成分布之间的距离^[11],其目标函数如公式 (1):

$$V(G, D) = -2 \lg 2 + 2JS(P_{data} \parallel P_g) \quad (1)$$

其中, P_{data} 与 P_g 分别为真实图像分布与伪造图像分布,通过分析 GAN 网络训练不稳定的因素,得

知当 P_{data} 与 P_g 的支撑集是高维空间中的低维流形 (manifold) 时, 其重叠部分为 0 的概率为 1; 当这两者的分布并无重叠部分或重叠部分可忽略, 则由公式 (1) 可知, 此时的 JS 散度就恒为定值 $-2\lg 2$, 从梯度下降的角度来分析, 此时梯度为 0, 出现梯度消失现象, 生成器无法从判别器的反馈中学习。针对这个问题, 本文使用 Wasserstein 距离来代替 JS 散度,

用 W 距离也称 *Earth - Mover (EM)* 距离来描述分布差异^[12]。 W 距离的定义如式 (2):

$$W(P_r, P_g) = \inf_{r \sim \prod(P_r, P_g)} E_{(x,y) \sim r} [\|x - y\|] \quad (2)$$

其中, P_r 表示真实数据分布; P_g 表示生成器生成的数据分布; $\prod(P_r, P_g)$ 是 P_r 和 P_g 所有可能组合起来的联合分布。

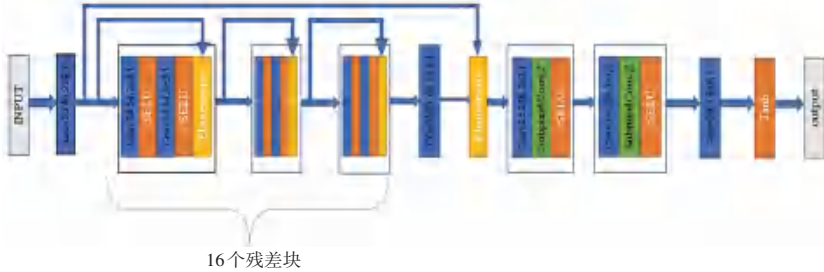


图 3 改进后的生成器网络结构图

Fig. 3 The improved generator network structure diagram

穷举出所有的可能, 从中找出最短的距离, 此时 GAN 的优化目标就变成找出真实图像的分布与生成器生成图像分布的 W 距离, 并不断根据反馈优化这个目标, 近似的由分类问题变成了回归问题。

网络损失函数 L^{SR} 由内容损失 L_X 和对抗损失 L_{Gen} 两部分组成, 如公式 (3):

$$L^{SR} = L_X + 10^{-3} L_{GEN} \quad (3)$$

虽然 MSE 损失函数具有较好收敛的特性, 但因其强大的约束力在图像的超分辨重建任务中对不同误差像素点的不均衡惩罚, 即增大了对高误差像素点的惩罚, 又对误差小的部分缺少相应的约束, 使得生成的图像过于平滑, 缺少细节。所以, 本文换成对高频信息更加敏锐 $L1$ 损失函数代替 MSE 损失函数。具体如公式 (4) 所示。

$$l_{L1} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} |I_{(x,y)}^{HR} - G(I_{(x,y)}^{LR})| \quad (4)$$

得到原高分辨图像与生成器生成的图像之间逐个像素之间的 $L1$ 距离, 这里的 r 为放大倍数, W 、 H 分别为对应图像的长和宽。

3 实验结果分析

实验的硬件环境为 CPU: Inter Xeon® W-2235 处理器, GPU: RTX 2080Ti, 运行内存为 32 G。用 Python3.6 作为本次实验的编程语言, 以 TensorFlow 框架搭建环境。本次实验使用的是超分辨率领域经典的 DIV2K 数据集, 其中包含 1 000 张高质量图像 (2K 分辨率), 将其中 800 张用来训练, 100 张验证图像, 最后 100 张测试图像, 再使用 Set5, Set14 以及

BSD100 数据集作为测试集来验证网络模型的性能。为证实本次实验的结果, 通过与传统方法中较为经典的双三插值 (Bicubic) 法, 基于学习的 SRCNN、VDSR 网络以及原始的 SRGAN 做比较与分析, 本次实验是在重建放大倍数为 4 倍的条件下进行, 在训练模型之前, 初始化生成器, 并优先对生成器迭代 100 轮, 将开始的学习率初始化为 0.000 1, 衰减率为 0.1, *Batch size* 设置为 16, 一共迭代 30 000 次, 使用交替训练的方法更新生成器与判别器。

3.1 评价标准

本文所采用客观评价标准与主观评价标准。客观评价标准为峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 与结构相似度 (Structural Similarity, SSIM); 主观评价使用平均主观意见分 (Mean opinion score, MOS)。

峰值信噪比 $PSNR$ 是使用最为普遍的一种图像客观评价指标, 其原本是表示信号可能的最大功率和噪声功率的比值, 但在用作图像的评价指标时是基于待做比较的图像之间对应像素点的差值来进行比较的。 $PSNR$ 的基本单位是 dB, $PSNR$ 越大则表示重建的图像越接近原图, 即失真越小, $PSNR$ 定义式如公式 (5):

$$PSNR = 10 \lg \frac{\alpha_{\max}^2}{\sigma_{MSE}^2} \quad (5)$$

其中, X_{\max} 为原图像像素的最大峰值, MSE 表示原图像与待比较图像之间的均方误差值。

结构相似度 $SSIM$ 是从亮度、对比度和结构 3 个层面来衡量图像的相似性。对应公式 (6) ~ (8)。

$$l(X, Y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (6)$$

$$c(X, Y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (7)$$

$$s(X, Y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (8)$$

其中, μ_x 与 μ_y 表示比较相似度的两张图像 X, Y 的均值; σ_x 与 σ_y 分别表示 X, Y 的方差; σ_{xy} 表示 X, Y 的协方差; C_1, C_2, C_3 均为常数。SSIM 的定义为式 (9) 为:

$$SSIM(X, Y) = l(X, Y) * c(X, Y) * s(X, Y) \quad (9)$$

SSIM 定义在闭区间 0~1 内, 在这个区间内的值越大, 表示图像失真越小。

3.2 客观结果分析

将实验结果分别在 Set5, Set14 数据集和 BSD100 数据集上进行验证, 计算通过不同超分辨率

方法重建出的超分辨率图像与真实高分辨率图像的 PSNR 和 SSIM 的平均值。通过对比发现, 经过双三插值法重建后的图像相比与其他几种方法, 无论是客观数据分析还是主观感受都较差, 图像模糊严重, 伪影较深, 而运用深度学习的另外几种方法则体现出明显的优越性。本文所使用的方法在 Set5, Set14 和 BSD100 验证集上的表现都强于同环境下的其他方法, 其他各种方法在不同数据集上的表现各有优劣, 其生成的图片在高频域即细节部分表现并不理想, 同时生成的图像在 PSNR 与 SSIM 指标上明显比本文的方法差。综合 Set5, Set14 和 BSD100 上的表现, 本文所使用的方法较原 SRGAN 网络在 PSNR 上提升 1.8 dB, 在 SSIM 上提升 6%, 也证明了本文所做改进的有效性, 客观评价指数见表 1, 各超分辨率方法重建效果图如图 4~6 所示。

表 1 客观评价指数对比表

Tab. 1 Comparison table of objective evaluation index

Validation	Evaluation	Bicubic	SRCNN	SRGAN	VDSR	Our
Set5	PSNR	26.34	29.17	26.93	29.37	29.59
	SSIM	0.77	0.83	0.79	0.83	0.85
Set14	PSNR	24.34	25.62	24.57	25.94	26.13
	SSIM	0.67	0.72	0.70	0.73	0.74
BSD100	PSNR	24.57	25.41	24.69	25.88	25.69
	SSIM	0.65	0.70	0.68	0.73	0.72

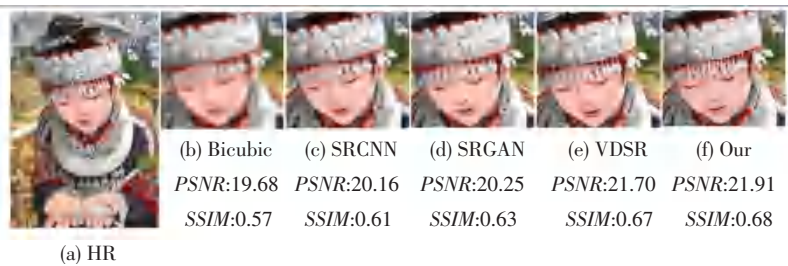


图 4 不同超分辨率方法在测试图“comic”中的局部对比图

Fig. 4 The partial comparison diagram of different super-resolution methods in the test chart "comic"

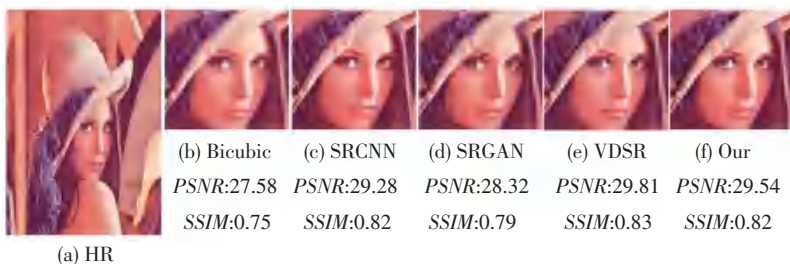


图 5 不同超分辨率方法在测试图“lenna”中的局部对比图

Fig. 5 The partial comparison diagram of different super-resolution methods in the test chart "lenna"

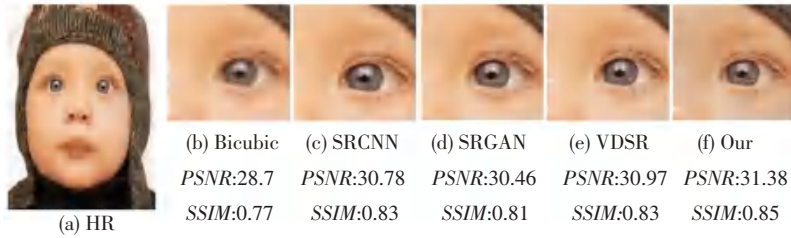


图 6 不同超分辨率方法在测试图“baby”中的局部对比图

Fig. 6 Partial comparison of different super-resolution methods in the test chart "baby"

3.3 主观结果分析

主观评价标准采用平均主观意见分 (Mean opinion score, MOS), 即随机找 30 个无视觉障碍的正常人, 让他们根据自己的主观视觉感受对使用不同超分辨率方法的测试图像进行评分, 共有 5 个等级, 从 1~5 分主观视觉效果依次递增, 统计后计算出平均分, 见表 2。

表 2 主观评价对比表

Tab. 2 Comparison table of subjective evaluation

Test	Subjective evaluation	Different super-resolution methods				
		Bicubic	SRCNN	SRGAN	VDSR	Our
Set5		1.83	3.56	3.20	3.53	3.62
Set14	MOS	1.62	3.26	3.14	3.57	3.78
BSD100		1.80	2.74	2.96	3.14	3.28

从主观评价来分析, 经过双三插值法重建出的图像可以看到明显的方块模糊, 在不同的测试集上 SRCNN 与 VDSR 的表现各有优劣, 但本文所使用的方法在主观真实感上都更强。经过多个对比, 不论是从客观评价标准 PSNR 峰值信噪比与 SSIM 结构相似度上来说, 还是从重建后的图主观意见得分 (MOS) 来分析, 本文所使用的方法都被证实是最优的。

3.4 对比实验分析

为探究在生成器网络中 BN 层以及相关激活函数对超分辨率重建效果的影响, 本文做出对比实验, 验证结果, 对比试验结果见表 3。

表 3 关于生成器改进的对比实验

Tab. 3 Comparative experiment on generator improvement

Type	Data Set		
	Set5	Set14	BSD100
	(PSNR/SSIM)	(PSNR/SSIM)	(PSNR/SSIM)
SRGAN	26.93/0.79	24.57/0.70	24.69/0.68
SRGAN-BN	27.41/0.81	25.37/0.71	25.23/0.68
SRGAN-BN+Selu	27.70/0.81	25.49/0.72	25.47/0.69

4 结束语

通过研究 SRGAN 网络中存在诸多的问题, 本文针对 GAN 网络收敛困难, 模型易崩溃的问题, 引入 Wasserstein 距离来代替 JS 散度, 优化对抗损失; 通过实验证明了去除生成器残差块中繁杂的 BN 层, 引入 SELU 激活函数, 提升了重建图像的精度, 稳定了网络收敛; 引入 L1 损失函数代替原 MSE 损失, 增加了生成图像的主观视觉效果, 增加了图像高频信息的重建。通过对比实验, 证明了本文所使用的方法是优于原网络的。接下来作者将继续实验探究基于 GAN 网络的图像超分辨重建的方法, 从分析判别器网络结构, 弥补重建过程中损失的高频信息入手, 通过实验找出一种更具有效率的超分辨重建的方法。

参考文献

- [1] 刘村, 李元祥, 周拥军, 等. 基于卷积神经网络的视频图像超分辨率重建方法[J]. 计算机应用研究, 2019, 36(4): 1-8.
- [2] BATZ M, EICHENSEER A, SEILER J, et al. Hybrid super-resolution combining example-based single-image and interpolation-based multi-image reconstruction approaches [C]// 2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015: 58-62.
- [3] 李斌, 马璐. 密集连接的生成对抗网络图像超分辨率重建[J]. 激光与光电子学进展, 2020, 57(22): 138-146.
- [4] 孙旭, 李晓光, 李嘉锋, 等. 基于深度学习的图像超分辨率复原研究进展[J]. 自动化学报, 2017, 43(5): 697-709.
- [5] DONG C, LOY C C, HE K, et al. Image Super-Resolution Using Deep Convolutional Networks[J]. IEEE Trans Pattern Anal Mach Intell, 2016, 38(2): 295-307.
- [6] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Networks [J]. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.
- [7] LEDIG C, THEIS L, F HUSZAR, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network [C]//Proceeding of the IEEE conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2017: 105-114.
- [8] NASH J F. Non-Cooperative Games[J]. Annals of Mathematics, 1951, 54: 286-295.
- [9] NASH J F. Equilibrium Points in N-Person Games [J]. Proceedings of the National Academy of Sciences, 1950, 36(1): 48-49.