# 基于生成对抗网络的单幅图像超分辨率论文解读与实验改进

吕帅

暨南大学

lvshuai@stu2022.jnu.edu.cn

# 摘要

本文首先对SRGAN[1]的论文进行解读与分析，接着通过Pytorch框架[2,3]复现了SRGAN模型。此外，考虑到动画和照片超分辨率之间的区别，重新设计了生成器的损失函数，取消了VGG19预训练的输入。针对GAN模型本身存在的训练困难问题，再SRGAN的基础上提出了SRWGAN[4]、SRWGAN-GP（GP指带有梯度惩罚的WGAN）模型。得出的结论是SRWGAN-GP在10k训练示例中优于SRGAN，并在附录中给出了详细的测试结果。

# 引言

给定一个低分辨率(Low Resolution, LR)图像，试图恢复或估计其高分辨率(High Resolution, HR)版本的图像被称为单图像超分辨率(Super Resolution, SR)。SR在图像处理领域非常有用，可用于提高视频和照片的质量，改善医学图像的清晰度，制作电影特效，甚至还可以用于提高卫星图像的分辨率，从而更好地监测地球的变化。SR通常通过使用深度学习网络来实现，首先，系统会训练一个深度学习模型，使其能够识别低分辨率图像中的细节。然后，当低分辨率图像被输入到系统中时，模型会根据其学到的知识生成更高分辨率的图像。值得一提的是，SR也可以与其他图像处理技术相结合，例如图像拼接技术，这可以让系统从多个角度拍摄的低分辨率图像中提取更多的信息，从而生成更加逼真的高分辨率图像。

# 论文解读

该部分是对Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network论文的分析与解读。

## 2.1 传统方法

传统放大图片都是采用插值的方法，最常用的有双线性插值（Bilinear）、双立方插值（bicubic）等，这些方法的速度非常快，而基于深度学习的超分辨率方案的开山之作叫做SRCNN。理解起来，传统插值方法可以看做把像素复制放大倍数后，用某种固定的卷积核去卷积；而基于卷积神经网络的超分辨率方法是去学习这个卷积核，根据构建出的超分图像与真实的高分辨率图像（Ground Truth）的差距去更新网络参数。

## 2.2 传统损失函数设计

基于CNN的损失函数一般都采用均方误差（Mean squared error, MSE）——即构建出来的超分图像与真实超分图像逐像素差的平方：

形状

中度可信度描述已自动生成

其中，是网络参数，是低分辨率图片，是重建出来的高分辨率图像，是真实的高分辨率图像，分别是图片数量、图片宽和高，都可以看成常数。这种方法存在一个问题：虽然最小化MSE能够使得峰值信噪比（Peak signal-to-noise ratio, PSNR）最大化，由于PSNR是基于像素级图像（pixel-wise image）的差异来定义的，

因此最高的PSNR不一定能反映人感官上最好的结果，过于平滑会使得图像内部物体的边缘看起来模糊，如下图所示：

日历

描述已自动生成

从数学角度分析，SR问题本身就是一个不适定问题（ill posed problem），因为对于一个给定的LR图像来说，存在多种对应的HR图像，使得评估一个SR方法变得困难和模糊，因此PSNR越高，效果不一定越好。其中，若一个数学物理定解问题的解存在、唯一并且稳定，则称该问题是适定的（well posed）。

## 2.3 作者工作

SRGAN论文作者的贡献可以总结为两点：

**第一、**首先把GAN引入到超分辨领域：与标准的GAN结构相比，SRGAN生成器的输入**不再是噪声**，而是低分辨率图像；而判别器结构跟普通的GAN没有什么区别，用于判别生成器生成的SR和真实的HR之间的区别。其模型结构图如下，生成器采用SRResNet架构，采取“卷积层+BN层+PReLU+残差块”的堆叠；判别器类似于VGG模型，即“VGG+LeakyReLU+Max-Pooling”。

图表, 图示, 条形图

描述已自动生成

**第二、**设计了一种全新的生成器的Loss——感知损失（perceptual loss），其由内容损失（content loss）和对抗损失（adversarial loss）通过加权求和方式取得：

文本

中度可信度描述已自动生成

**内容损失**X有两种选择，一种选择就是传统的MSE，另一种选择是作者提出的VGG loss，即分别得到真实图片（Ground Truth）与生成图片在VGG19下的某个层的特征图，然后在特征图下求MSE。总的来说，相较于MSE，VGG loss多了一个，如下式所示：

文本

描述已自动生成

**对抗损失**定义如下，其中表示判别器对于生成超分图片的输出，其值越大越好，因此越小越好。

文本

低可信度描述已自动生成

# 实验

在Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network论文中，作者没有给出官方代码，本次实验主要参考GitHub开源代码，并在其基础上改进。该部分内容包括数据集选取与预处理、数据集差异性分析、改进的模型SWGAN和SWGAN-GP、实验训练策略、实验结果。

## 3.1 数据集选取与预处理

项目数据集选自Danbooru2018[5]，该动漫数据集的像素值均为512x512，每张图像由RGB通道组成。我从14W张数据集中随机选取10400个动漫图像，其中10000张用于模型训练，200张用于验证，200张用于测试，其概览如下：

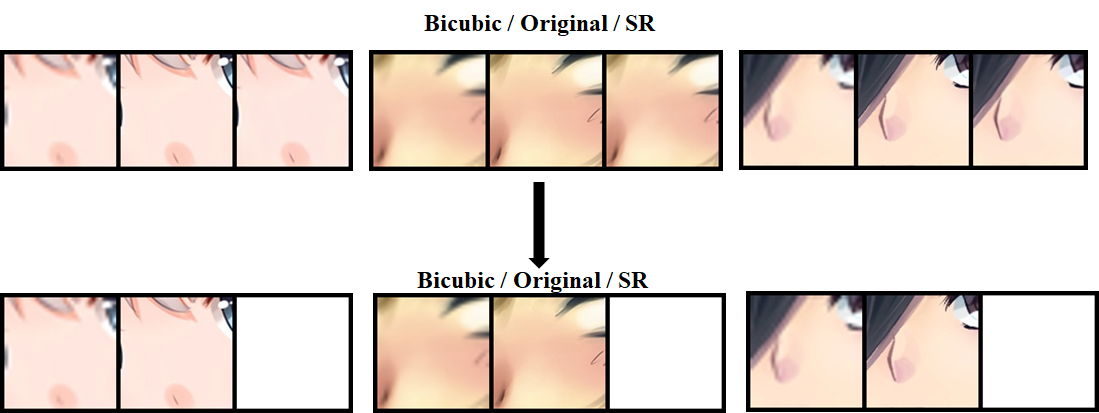


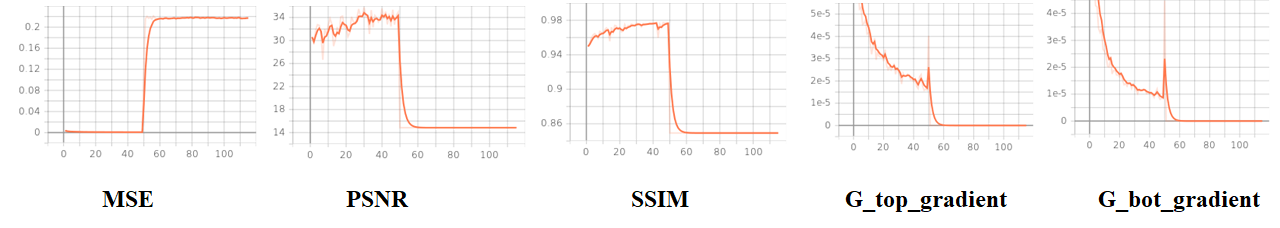
SRGAN网络的输入图像是数据集中的一张原图像（HR）和一个双三次插值下采样（LR）的图像对。由于GPU（RTX 3070）内存（8GB）限制，SRGAN输入的图像对规约为128 x 128 HR + 32 x 32 LR，SRWGAN-GP，输入的图像对规约为96 x 96 HR + 24 x 24 LR。

## 问题分析与解决

### 3.2.1问题1

**问题1：**如果直接使用SRGAN论文给出的模型，来训练动漫数据集，可能会崩溃，在训练达到大约50epoch时，生成器给出的SR直接会直接变为白色，MSE等相关指标也停止更新，如下图所示：





**原因分析1：**为了方便分析，我们把生成器的梯度打印出来，发现50epoch后为0，因此我们有理由怀疑是生成器的损失函数设计出现了问题。原始SRGAN的损失函数包括三个部分：MSE损失、VGG损失和对抗性损失。其中VGG损失基于预训练的19层VGG网络，而VGG19是从真实生活照片训练得到的模型，**会导致检索到的特征不适用于动漫**。具体地，动漫的像素和照片的像素分布是不同的。照片的颜色丰富，而且内容在像素之间变化很大（图像多为高频信息），而动漫图像通常为单一颜色，像素通常没有或只有轻微的变化（图像多为低频信息）。此外，动漫没有像VGG19这样的预训练模型可用。

**解决1**：放弃VGG损失，并提出了新的生成器损失函数为：

### 3.2.2问题2

**问题2：**GAN训练及其不稳定，包括MSE的振荡变化（以MSE为例）、突然地内存溢出，如下图所示：



**原因分析2**：GAN模型本身存在的训练崩溃问题。

**解决2**：参考论文Improved Training of Wasserstein GANs[6]，因为SRGAN先于该论文发表，因此我们把Wasserstein引入SRGAN进行尝试，来度量不同数据集之间的扰动差异，称该模型为SRWGAN。

### 3.2.3问题3

**问题3：训练SRWGAN时会出现条状、点状纹理。**



**原因分析3：引入噪声过大，模型对抗噪声的鲁棒性不好。**

**解决3**：调整对抗损失的比λ为1e-5、引入梯度惩罚（Gradient Penalty, GP）使得模型输入有微小变化的时候，网络权重不会产生太大的变化，即模型SRWGAN-GP。

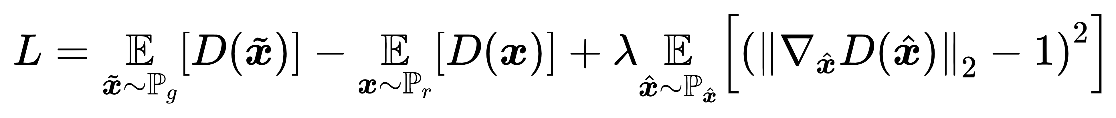
## 3.3 SRWGAN

SRWGAN而改进后相比原始SRGAN的算法实现代码只需要改动四点：

* 判别器最后一层去掉sigmoid。
* 生成器和判别器的loss不取log。
* 每次更新判别器的参数之后把它们的绝对值截断到不超过一个固定常数c，即将权重裁剪到某个连接范围[-c, c]。
* 不要用基于动量的优化算法（包括momentum和Adam），推荐RMSProp，SGD也行。

## 3.4 SRWGAN-GP

SRWGAN-GP使用的损失函数是Ishaan Gulrajani提出的带有梯度惩罚的L1对抗损失和鉴别损失，该惩罚项主要是为了让生成器判别器的梯度值尽可能地在1附近，否则就会惩罚。



## 3.5 训练过程与分析

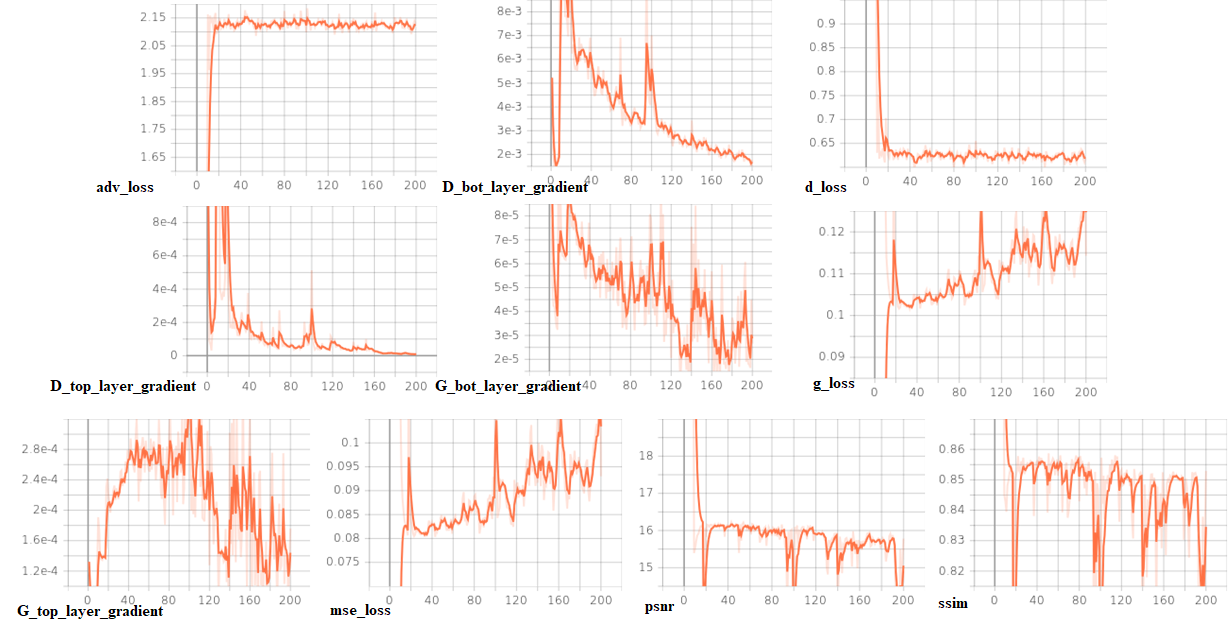
### 3.5.1超参数设置

SRGAN和SRWGAN的batch size设置为32，训练了200个epoch。其中SRGAN训练时长约为6.6h（一个epoch约为1min55s左右），SRWGAN训练时长约为3.8h（一个epoch约为1min10s左右）。

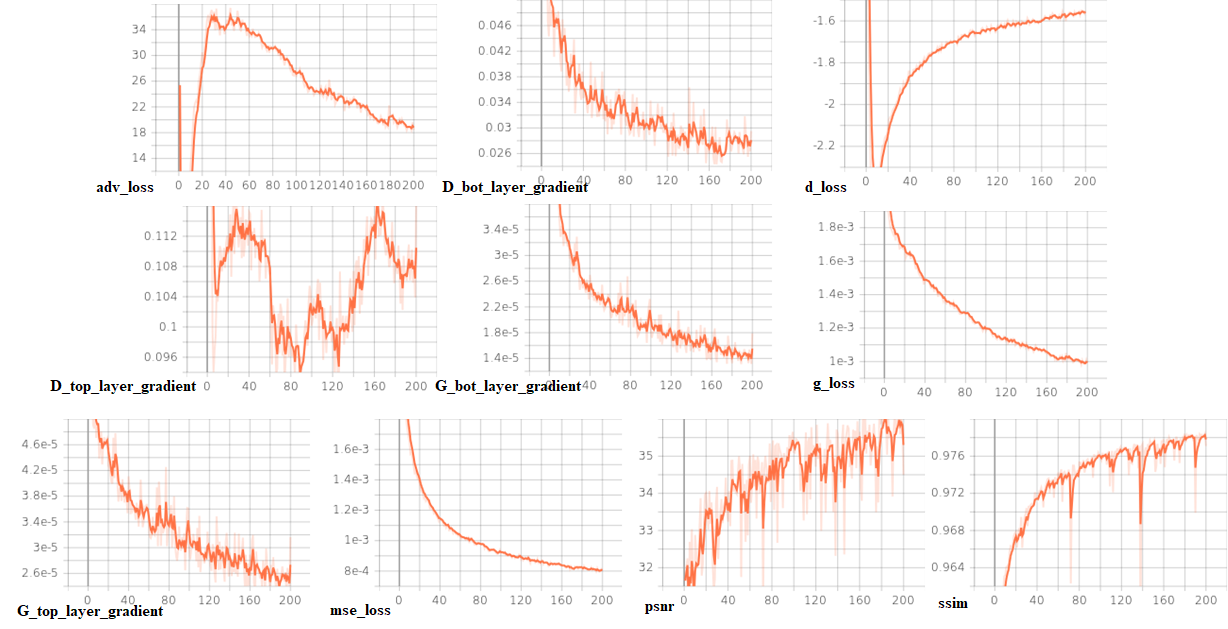
|  |  |  |
| --- | --- | --- |
|  | SRGAN | SRWGAN（GP） |
| bATCH SIZE（批大小） | 32 | 32 |
| EPOCH（轮次） | 200 | 200 |
| 对抗损失占比 | 1e-2 | 1e-5 |
| training time（训练时长） | 6.6h | 3.8h |

### 3.5.1可视化与结果分析

为了了解模型和训练是否工作良好，分别打印了SRGAN与SRWGAN-GP在生成器和判别器的第一层和最后一层权重的梯度的abs均值、生成器和判别器的损失值、MSE、PSNR、SSIM。SRGAN可视化如下：



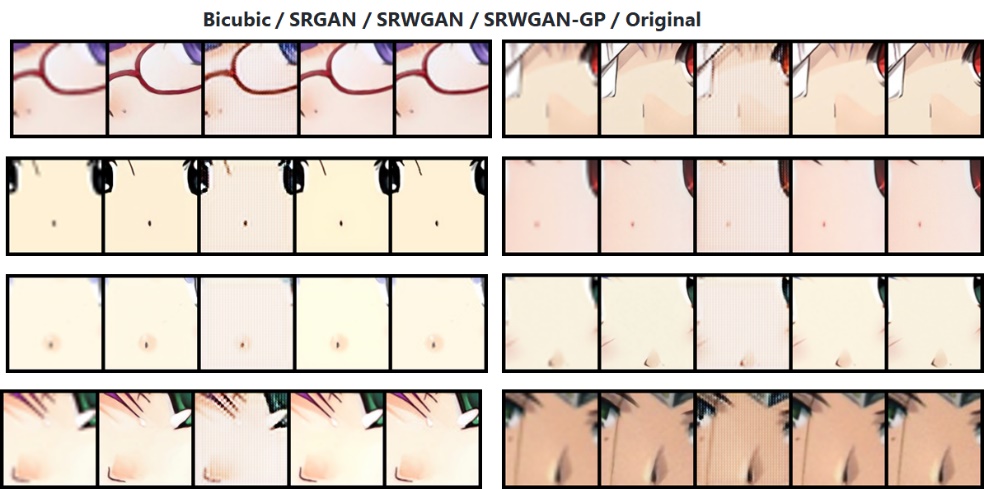
SRWGAN-GP可视化如下：



分析结论：SRWGAN-GP的MSE损失平稳下降、具有稳定的梯度、收敛速度更快。SRWGAN-GP的SSIM和PSNR等评估指标峰值分别约为0.977、38，SRGAN分别约为0.855、16。

# 测试结果

实验结果包括如下的1）LR的下采样的双立方体。2）基于修改损失函数的SRGAN的SR。3）SRWGAN的SR。4）SRWGAN-GP的SR。5）数据集的HR。



一般来说，SRWGAN会出现条纹，SRWGAN-GP和SRGAN几乎无法分辨，但SRWGAN-GP更加符合HR的颜色分布。在提交的“训练过程SR结果\vis\_test”文件夹中有更多的测试结果。

# 结论和未来工作

由于动漫和照片的分布是不同的，动漫超分辨率也是一个需要解决的问题，在使用SRGAN解决动漫超级分辨率的问题时，修改了损失函数、不同的超参数，并删除了预训练模型的输入。SRWGAN-GP解决方案在这个问题上总体上优于SRGAN，训练时间更短，结果更好。之后的工作中，我计划探索和应用更多的GAN模型来改善单一动漫图像的结果，同时利用RNN的优势来处理动漫视频。

# 贡献

此项目主要对SRGAN论文的核心进行了解读，并解决了动漫数据集SR出现的一系列问题，随后将WGAN引入到SRGAN中，解决了训练困难的问题，接着将GP引入SRWGAN中使得SR图像更加平滑，在zip中有详细的代码说明与相应的可视化结果与训练过程。

# 致谢

感谢林老师为我打开了计算机视觉的大门，使得我更关注传统CV与深度学习的结合，感谢林老师在汇报结束时为我提供了非常有益的建议。感谢我的母亲支持我购买价格昂贵的笔记本电脑，让我有环境进行该实验的训练与研究。

# 参考文献

[1] 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. 2017: 4681-4690.

[2] https://github.com/goldhuang/SRGAN-PyTorch

[3] <https://github.com/tensorlayer/srgan>

[4] Arjovsky, Martin; Chintala, Soumith; Bottou, Léon. Wasserstein Generative Adversarial Networks. International Conference on Machine Learning. PMLR: 214–223.

[5] https://hyper.ai/datasets/9000

[6] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein gans[J]. Advances in neural information processing systems, 2017, 30.

[7] https://github.com/JustinhoCHN/SRGAN\_Wasserstein

[8] <https://github.com/kolak33/SRWGANGP>