



南开大学
Nankai University

南 开 大 学

计 算 机 学 院

图像超分辨率研讨报告

崔江浩、丁屹、卢麒萱

年级：2020 级

专业：计算机科学与技术

2022 年 5 月 29 日

摘要

关键字：图像超分辨率、生成对抗网络、感知损失函数

目录

一、 摘要	1
二、 介绍	1
(一) 卷积神经网络的设计	2
(二) 损失函数	2
(三) 贡献	2
三、 方法	2
(一) 对抗网络架构	3
(二) 感知损失函数	3
1. 内容损失	3
2. 对抗损失	4
四、 实验总结	4
五、 结论	4

一、 摘要

从低分辨率图像还原高分辨率图像的困难任务被称为超分辨率，其受到了计算机视觉研究领域的广泛关注，具有广泛的研究范围。

尽管使用更快，更深层次卷积神经网络在单图像超分辨率上的准确性和速度取得了突破，但是一个核心问题在很大程度上仍然没有解决：当我们在进行大比例上采样时，如何恢复更精细的纹理细节？这种基于优化的超分辨率方法主要由目标函数的选择驱动。最近的工作主要集中于最小化重建错误平均值平方。这样带来的结果虽然有高峰值信噪比，但通常缺乏高频细节，在感官上令人不满意：它们没有高分辨率所对应的高保真度。

在本文中，我们提出了 SRGAN，图像超分辨率（SR）的生成对抗网络（GAN）。据我们所知，这是第一个能够推断出真实自然的 4 倍超分辨率图像的框架。为了实现这一点，我们提出了一个由对抗损失和内容损失组成的感知损失函数。对抗性损失器使用一个经过训练的分类器网络来区分超分辨率图像和原始真实图像，以将结果导向自然图像处理器。此外，我们使用由感知相似性而不是像素相似性激活的内容损失器。

在公共测试集数据中，我们的深层残差网络能够从重度下采样的图像中修复真实材质。在 MOS 测试中，使用 SRGAN 获得的高分辨率图像比任何其他图像超分辨率方法都要接近原始图像。

二、 介绍

在本文中，我们提出了一个超分辨率生成对抗网络 SRGAN，其中我们采用了一个具有跳跃连接和偏离 MSE 的深度残差网络 ResNet 作为唯一优化目标。与之前的工作不同的是，我们使用 VGG 网络的高级特征图，结合了一个识别器来定义一种新的感知损失，该识别器激励了难以从高分辨率图像 HR 中分离出来的特征。

这里我们将重点讨论单图像超分辨率 SISR，不再进一步讨论从多幅图像恢复 HR 图像的方法。

相关工作：

- 基于预测的过滤方法，速度快但是会生成过于光滑的纹理
- 建立低分辨率和高分辨率图像信息之间的复杂映射，通常依赖于训练数据
- 将基于梯度轮廓先验的边缘重定向 SR 算法和基于学习的优点相结合，在避免边缘伪影的同时重建真实的纹理细节
- 邻域嵌入方法通过在低维簇中寻找相似的 LR 训练快，并结合它们对应的 HR 块进行重建
- 基于卷积神经网络 CNN 的 SR 算法表现出了优异的性能

算法表现出了优异的性能。Wang et al 基于学习的迭代收缩和阈值算法 LISTA 在前馈网络结构中编码了一个稀疏表示。Dong 等人使用双三次插值对输入图像进行上采样，并对三层深度全卷积网络进行端到端训练，以实现最先进的 SR 性能。随后，研究表明，让网络直接学习缩放滤波器可以进一步提高精度和速度。Kim 等人利用他们的深度递归卷积网络 DRCN，提出了一种高性能的架构，允许长范围像素依赖，同时保持模型参数的数量很小。与我们的论文特别相关的是 Johnson 等人和 Bruna 等人的工作，他们依靠更接近感知相似性的损失函数恢复在视觉上更有说服力的 HR 图像。

(一) 卷积神经网络的设计

研究表明,更深层次的网络架构可能很难训练,但有可能大幅提高网络的准确性,因为它们允许非常复杂的建模映射。为了有效地训练这些更深层次的网络结构,批处理归一化经常被用来抵消内部协变最移位。更深层次的网络架构也被证明可以提高 SISR 的性能。另一个简化深度 cnn 训练的强大设计选择是最近引入的残差块和跳跃连接的概念。跳跃连接减轻了网络体系结构建模身份映射的负担,这在本质上是微不足道的,然而,用卷积核表示可能不是微不足道的。在 SISR 的背景下,学习缩放滤波器在精度和速度方面也是有益的。这是对 Dong 等人的改进,在将图像输入 CNN 之前,使用双三次插值来扩大 LR 的观察值。

(二) 损失函数

像素级损失函数(如 MSE)处理恢复丢失的高频细节(如纹理)具有固有的不确定性:最小化 MSE 以寻找合理解,这些解通常过于光滑,因此具有较差的观感。使用在神经网络特征空间计欧式距离的损失函数,并结合对抗性训练。结果表明,能生成视觉上更优越的图像,并可用于解决解码非线性特征表示的病态问题。与这项工作类似,使用从预先训练的 VGG 网络提取的特征,而不是简单的像素级误差评估,能获得在观感上超分辨率和风格都更令人信服的结果。

(三) 贡献

在这篇文章中,我们描述了第一个非常深入的 ResNet 架构。我们的主要贡献是:

- 我们通过 PSNR 和结构相似度 SSIM 来测最高缩放因子 (4X) 的图像 SR, 并采用针对 MSE 优化的 16 块深度 ResNet (SRResNet)
- 我们提出了 SRGAN, 这是一个基于 GAN 的网络, 针对新的感知损失进行了优化。在这里, 我们将基于 MSE 的内容损失替换为在 VGG 网络的特征映射上计算的损失, 它对像素空间的变化更不稳定
- 我们通过对来自三个公共基准数据集的图像进行广泛的平均意见评分 MOS 测试确认, SRGAN 在很大程度上是一种新的技术, 用于建立具有高缩放因子 (4X) 的逼真 SR 图像

三、 方法

在 SISR 中, 我们从低分辨率的输入图像 I^{LR} 生成超分辨率的图像 I^{SR} , I^{LR} 是高分辨率图像 I^{HR} 的低分辨率版本, 高分辨率图像只在训练期间可用。在训练期间, I^{LR} 为使用比例因子为 r 进行的对 I^{HR} 高斯滤波下采样操作获得的图像。对于一个有着 C 个颜色通道的图像, 我们使用大小为 $W \times H \times C$ 的实张量描述 I^{LR} , 使用大小为 $rW \times rH \times C$ 的实张量描述 I^{HR} 和 I^{SR}

我们的最终目标是训练一个生成函数 G , 它可以对 LR 输入图像预测对应的 HR 图像。对此, 我们训练了一个参数化 θ_G 的前馈 CNN G_{θ_G} 生成网络, $\theta_G = \{W_{1:L}; b_{1:L}\}$ 表示 L 层网络的权重和偏差, 是通过 SR 的特定损失函数 l^{SR} 进行优化得到。对于训练图片 $I_n^{HR}, n = 1, \dots, N$ 及相应的 $I_n^{LR}, n = 1, \dots, N$, 我们求解:

$$\hat{\theta}_G = \arg \min_{\theta_G} \frac{1}{N} \sum_{n=1}^N l^{SR}(G_{\theta_G}(I_n^{LR}), I_n^{HR}) \quad (1)$$

我们将专门设计几个相异模型的损失函数的加权组合作为新的感知损失模型 l^{SR} 。

(一) 对抗网络架构

我们进一步定义了一个分类器网络 D_{θ_D} ，使用 G_{θ_G} 以一种交替的方式来优化它，来解决对抗的 min-max 问题：

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{train}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})} [\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))] \quad (2)$$

这个公式的一般想法是，它允许训练生成模型 G ，目的可能是欺骗可微分类器 D ，该分类器被训练来区分超分辨率图像和真实图像。通过这种方法，我们的生成器可以学习创建与真实图像高度相似的解，因此很难通过 D 进行分类。生成网络 G 的核心使用两个卷积层其中包含 3×3 内核和 64 个特征映射，然后使用 batch 标准化层和 ParametricReLU 作为激活函数。通过两个经过训练的亚像素卷积层来提高输入图像的分辨率。

为了从生成的 SR 样本中区分真实的 HR 图像，我们训练了一个分类器网络。LeakyReLU 激活 ($\alpha = 0.2$)，避免整个网络的最大池化。训练判别器网络来求解方程 2 中的最大化问题。它包含 8 个卷积层，每层 3×3 个卷积核，与 VGG 网络一样，从 64 个核增加到 512 个核，增加了 2 倍。当特征数增加一倍时，采用跨步卷积来降低图像分辨率。在生成的 512 个特征图之后，再经过两个密集层和一个最终的 sigmoid 激活函数来获得样本分类的概率。

(二) 感知损失函数

我们感知损失函数的定义 l^{SR} 对生成网络的性能至关重要。而 l^{SR} 通常基于 MSE 建模，我们改进了 Johnson 和 Bruna 等人的成果，并设计了一个损失函数，根据感知相关特征评估解决方案。我们将感知损失表示为内容损失 l_X^{SR} 和对抗损失如下：

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}} \quad (3)$$

perceptual loss (for VGG based content losses)

1. 内容损失

像素级 MSE 损失函数的计算方法：

$$l_{MSE}^{SR} = \frac{1}{r^2 W H} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2 \quad (4)$$

这是图像 SR 最广泛使用的优化目标，许多最先进的方法都依赖于此。然而，在获得特别高的 PSNR 的同时，MSE 优化问题的解决方案往往缺乏高频率内容，导致过于光滑的纹理。

我们不依赖像素级损失，而是基于 Gatys、Bruna 和 Johnson 等人的思想，并使用更接近感知相似性的损失函数。我们基于 Simonyan 和 Zisserman 描述的预训练 19 层 VGG 网络的 ReLU 激活层来定义 VGG 损失。用 $\phi_{i,j}$ 我们表示 VGG19 网络中第 i 个 maxpooling 层之前的第 j 个卷积（激活后）得到的特征映射，然后我们将 VGG 损失定义为由构图像 G 的特征 $G_{\theta_G}(I^{LR})$ 和原始图像 I^{LR} 表示之间的欧氏距离

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2 \quad (5)$$

这里 $W_{i,j}$ 和 $H_{i,j}$ 描述了 VGG 网络中各自特征图的维度。

2. 对抗损失

除了到目前为止所描述的内容损失, 我们还将 GAN 的生成成分添加到感知损失中。这鼓励我们的网络倾向于通过尝试得到欺骗分类器的解决方案。生成损失 l_{Gen}^{SR} 定义基于分类器对于所有训练样本的概率 $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ 为:

$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (6)$$

这里的 $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ 表示重构图像 $G_{\theta_G}(I^{LR})$ 是自然 HR 图像的概率。为了更好的梯度行为, 我们选择最小化 $-\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$ 而不是 $\log[1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$

四、 实验总结

我们通过 MOS 测试证实了 SRGAN 优越的感知性能。我们进一步表明, 标准的量化措施, 如 PSNR 和 SSIM, 无法捕获和准确评估与人类视觉系统相关的图像质量。这项工作的重点是超分辨率图像的感知质量, 而不是计算效率。与 Shi 等人的模型相比, 该模型没有对视频 SR 进行实时优化。然而, 对网络架构的初步实验表明, 浅层网络有可能在质量表现上略有下降的情况下提供非常有效的替代方案。与 Dong 等人的研究相比, 我们发现更深层次的网络架构是有益的。我们推测 ResNet 设计对更深层次网络的性能有很大的影响。

五、 结论

我们描述了一个深度残差网络 SRResNet, 当使用广泛使用的 PSNR 公共数据集时, 它达到了一个新的里程碑。我们也强调了 PSNR 的一些局限性, 提出了 SPGAN 算法, 该算法通过训练 GAN 来增强内容损失函数的对抗性损失, 通过大量的 MOS 测试, 我们已经证实, 在较大的缩放因子 (4X) 下, SRGAN 重建比使用最先进的其他参考方法得到的重建图像在很大程度上更加逼真。