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

丁屹、卢麒萱、崔江浩

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、卷积神经网络、双三次插值法、基于稀疏编码的 SR 方法、深度网络、生成对抗网络、感知损失函数

目录

第一章　问题描述

图像超分辨率（超分辨率成像，Super-resolution imaging，缩写 SR，是一种提高影片分辨率的技术，通过硬件或软件的方法提高原有图像的分辨率，通过一系列低分辨率的图像来得到一幅高分辨率的图像过程就是超分辨率重建。图像超分辨率技术分为超分辨率复原和超分辨率重建。目前，图像超分辨率研究可分为三个主要范畴：基于插值、基于重建和基于学习的方法。

本文后将详细介绍图像超分辨率的三种深度学习方法：

1. SRCNN 方法
2. VDSR 方法
3. SRGAN 方法

第二章　SRCNN

一、　前言

作者提出了一种深度学习的方法用于解决超分辨率问题 (super-resolution, SR)，具体的是，使用 CNN 来拟合低分辨率图像和高分辨率图像的映射，这是一种端到端的方法；虽然传统的基于稀疏编码的 SR 方法（the sparse-coding-based method）也能被看作是 CNN 的；但是作者提出的方法，这个 CNN 是个轻量级模型，其能够在实际应用更快速的同时，保证图像重建达到当前先进水平 (state-of-art)。

解决 SR 问题大部分是基于例子的策略（example-based strategy），主要有两种思路：

1. 利用同一图像的内部相似性。

2. 从外部低分辨率和高分辨率示例图片对中学习映射函数。

基于稀疏编码的 SR 方法就是学习映射函数的一种代表方法，该方法包括以下几步：

1. 从输入图片中密集地进行采样形成大量的重叠的 patch，并且对这些 patch 进行预处理（例如减去均值、标准化）。

2. 使用一个低分辨率的字典（low-resolution dictionary）对 patch 进行编码。

3. 使用一个高分辨率的字典（high-resolution dictionary）对输出进行编码，用来构建高分辨率的 patch。

4. 将重叠的 patch 进行聚合产生最终的输出。

作者提出的 Super-Resolution Convolutional Networks（SRCN）具有以下几个优点：

1. 结构简单且准确率高。

2. 卷积核数和层数适中，在实际的及时应用中，即使使用 CPU 也能得到很快的相应速度。

二、　前人相关工作

对于大部分的 SR 算法都是关注灰度图片或单通道的图片；至于彩色图片，这些传统的算法都是先将彩色图片（RGB）转换到不同的色彩空间（例如 YCbCr、YUV），再在亮度通道上使用 SR 算法。

三、　超分辨率的卷积神经网络

(一)　构想

首先将输入图片使用双三次插值法（bicubic interpolation）放大到想要的大小，这是唯一需要做的预处理，并且把经过预处理后的图片记作 Y。我们的目标是学习出一个映射 F，使得 F(Y) 和真实高分辨率图片 X 尽可能的相似。如图2.1所示：

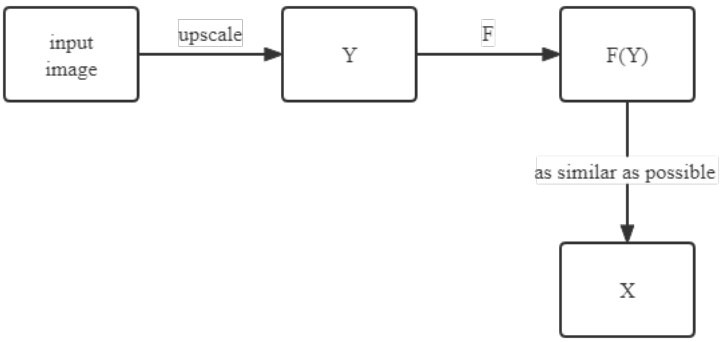


图 2.1: 流程图

我们想要学习的这个函数 F，其实可以分解成以下三部分操作：

1.Patch extraction and representation: 从 Y 中提取出 patch，然后将每个 patch 映射成一个高维向量，并且这些向量可以组成一个 featuremap，这个操作等同于使用带有偏置的卷积核在图片上进行卷积运算，我们可以写成：

$$F_1(Y) = max(0, W_1 * Y + B_1)$$

其中 W1、B1 分别表示为卷积核和偏置，* 表示卷积运算；W1 对应于 n1 个大小为 c×f1×f1 的卷积核 (c 表示输入图片的通道数)。

2.Non-linear mapping: 将上一步得到的高维向量经过非线性变换转换为另一个高维向量，这些高维向量也能得到一个 featuremap，这个操作对应于使用 n2 个大小为 1×1 的卷积核进行运算，当然卷积核的大小是可以改变的，使用更大的卷积核会有更好的泛化效果，我们可以写成：

$$F_2(Y) = max(0, W_2 * F_1(Y) + B_2)$$

其中 W2 对应于 n2 个大小为 n1×f2×f2 的卷积核。虽然增加更多的卷积层能增加非线性，但是会使得模型复杂度增加，从而需要更多的训练时间。

3.Reconstruction：将上一步生成的高维向量合成最终的高分辨率图片输出，这个操作可以看成在 featuremap 上使用预先定义的卷积核进行 averaging 的过程，这是一个线性过程：

$$F(Y) = W_3 * F_2(Y) + B_3$$

其中 W3 对应于 c 个大小为 n2×f3×f3 的卷积核。

以上的操作如图2.2所示。

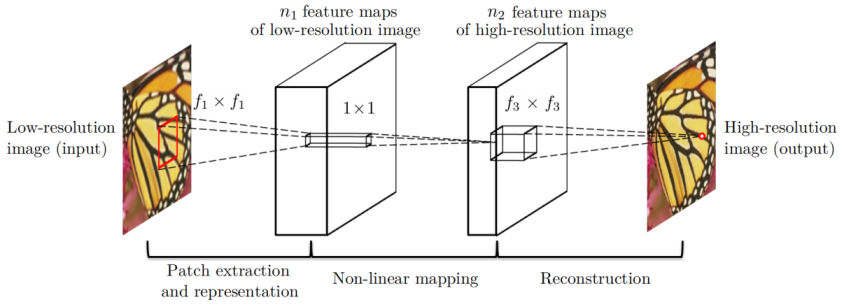


图 2.2: 学习模型

(二)　和基于稀疏编码的 SR 方法的关系

传统的基于稀疏编码的 SR 方法可以被看成一个卷积神经网络，如图2.3所示：

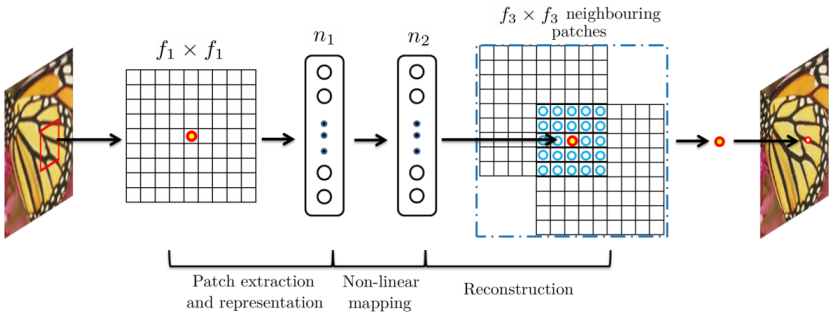


图 2.3: 卷积神经网络

在基于稀疏编码的 SR 方法模型中，我们可以把 f1×f1 低分辨率的 patch 看成是从输入图片提取出来的，然后稀疏编码求解程序将 patch 投影到一个字典上，如果这个字典的大小为 n1，那么相当于在输入图片上使用 n1 个大小 f1×f1 的卷积核进行卷积运算，如上图的左半部分所示。稀疏编码求解程序会迭代的处理 n1 维度的向量，从而得到一个 n2 维度的向量，一般来说 n1、n2 是相等的，这个时候稀疏编码求解程序起的作用就是一个大小为 1×1 的非线性映射运算符，如上图的中间部分所示。然后再对 n2 维度的向量投影到另一个字典空间目的是产生高分辨率的 patch，重叠部分的 patch 会进行平均操作，如上图的右半部分所示。

在 SRCNN 中，以上的低分辨率字典、高分辨率字典、非线性映射、减去均值和平均化的操作都包括在需要优化的卷积核里了，所以说该模型是一个包括所有操作的端到端的映射。而且之所以 SRCNN 的表现更好是因为 SRCNN 使用更多的像素信息用于图像重构。例如当我们设置 f1=9,f2=1,f3=5,n1=64,n2=32 时，高分辨率像素使用了 (9+5-1)^2=169 个像素的信息，而传统的方法只使用了 (5+5-1)^2=81 个像素的信息。

(三)　训练

该模型的参数有：W1,W2,W3,B1,B2,B3。给定一组高分辨率图片 Xi 以及对应低分辨率图片 Yi，我们使用 MSE 作为损失函数：

$$L(\varnothing) = \frac{1}{n}\sum_{i=1}^{n} \|F(Y_i;\varnothing) - X_i\|^2$$

其中 n 表示训练样本的数量，使用 MSE 作为损失函数会使得 PSNR 值变得很高，损失函数使用 SGD 进行优化，参数的更新过程如下所示：

$$\triangle_{i+1} = 0.9 \cdot \triangle_i - \mu \cdot \frac{\partial L}{\partial W_i^l}, W_{i+1}^l = W_i^l + \triangle_{i+1}$$

参数 Wi 使用均值为 0 标准差为 0.001 的高斯分布进行初始化，参数 Bi 初始化为 0，前两层的学习率为 10^-4，最后一层的学习率为 10^-5，而且作者发现在最后一层使用小的学习率对于 SRCNN 收敛很重要。为了避免在训练时边界影响，所有 CNN 都采用 no padding 的策略。

四、 实验结果

(一) 训练集大小对实验结果的影响

使用大型的训练集可能对 SRCNN 的表现有提升，但是训练集大小的影响不大，不像如分类问题那么明显。

(二) SR 中已学习的卷积核

第一层的 featuremap 中包括不同的结构，例如不同方向上的边缘；第二层主要是光照强度的不同。

(三) 模型大小和模型表现的折衷

卷积核的数量

通过增加卷积核的数量能得到更好的性能，但是卷积核越多，图像重构的时间越长。

卷积核大小

适当的增大卷积核大小能获得更多的结构信息，所以会得到更好的结果。增大第二层卷积核的大小能显著提升性能，这说明利用周围的信息在映射阶段是有益的；但是增大第二层卷积核大小会使得模型复杂度、以及降低了部署的时间。

层的数量

四层网络比三层网络的收敛速度慢，但是给定足够的时间，四层网络最后也会追上三层网络的。但是并不像图像分类问题上网络深度越深效果越好，SR 问题并不能看出网络深度对网络表现的影响，而且随着网络深度的增加可能还会导致性能的下降。

导致以上结论的原因可能是 SRCNN 没有池化层和全连接层，所以它对初始化参数和学习率很敏感。当模型层数增加时，我们很难保证找到一组合适的学习率使得其收敛，即使模型收敛了，也可能陷入一个不好的局部最小值，并且经过足够的训练时间，已学习的卷积核的多样性会变少。而且在图像分类领域，不适当的增加模型深度也会使得准确率的下降或退化。

五、 总结

作者提出了基于稀疏编码的卷积方法，并将其实现为一个深度神经网络 SRCNN，能够学习端到端的低分辨率图片到高分辨率图片的映射，而且这是个轻量级的结构，有着超越其他 state-of-art 方法的表现。

第三章　VDSR

一、　摘要

作者提出了一种高精度单幅图像超分辨率方法。该方法使用了一个受 VGG 网络启发的深度进化网络，用于图像网络分类。增加网络深度会显著提高准确性，最终模型使用 20 个重量层。通过在深层网络结构中多次级联小过滤器，以有效的方式利用大图像区域上的上下文信息。然而，对于非常深的网络，收敛速度成为训练中的一个关键问题。作者提议一个简单而有效的训练程序，只学习残差，并使用极高的学习率 (比 SRCNN 高 104 倍) 由可调梯度削波启用。

二、　介绍

单幅图像超分辨率 (SISR)：解决在给定低分辨率 (LR) 图像的情况下生成高分辨率 (HR) 图像的问题。

SISR 广泛用于计算机视觉应用，早期方法包括插值，如双三次插值和 Lanczos 重采样利用统计图像先验的更强大的方法或内部补丁复发。

目前，学习方法被广泛用于模拟从 LR 到 HR 面片的映射。邻居嵌入方法插值面片子空间。稀疏编码方法使用基于稀疏信号表示的学习紧凑字典。最近，兰登森林和卷积神经网络 (CNN) 在准确性方面也有了很大的提高。

SRCNN 成功地将深度学习技术引入到超分辨率问题中，但在三个方面存在局限性:

- 它依赖于小图像区域的上下文;
- 训练收敛太慢;
- 网络只对单一尺度起作用。

作者提出了一个基于非常深的卷积网络的高精度随机共振方法。如果使用小的学习率，非常深的网络收敛太慢。如果使用高学习率提高收敛率会导致梯度爆炸，于是用剩余学习和梯度削波来解决这个问题。此外，作者扩展了工作处理单一网络中的多尺度随机共振问题。

三、　相关工作

SRCNN 是一种具有代表性的基于深度学习的 SR 方法。因此，让我们用我们提出的方法进行分析和比较。

(一)　图像超分辨率卷积网络

SRCNN 模型由三层组成: 面片提取/表示、非线性映射和重建。

作者认为增加深度可以显著提高性能。使用了 20 个配重层 (每层 3×3), 深度关系网 (20vs3), 用于重建的信息更大 (41×41 对 13×13)。

Training SRCNN 直接建模高分辨率图像。SRCNN 的两个目的: 将输入传送到末端层和重构残差。由于网络直接对残差图像进行建模，可以以更高的精度实现更快的收敛。

Scale SRCNN 针对单个比例因子进行训练，仅与指定的比例一起工作。为了处理多尺度随机共振 (可能包括分数因子)，需要为每个感兴趣的尺度构建单独的单尺度随机共振系统。

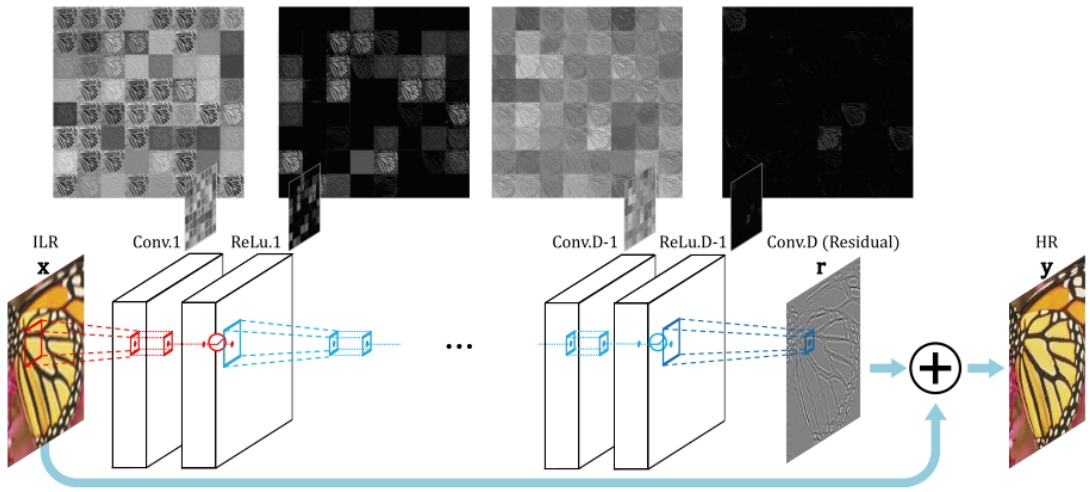
作者为了使网络符合各种场景，设计并训练了一个单一网络来有效地处理多尺度随机共振问题。

细节部分,输出图像与输入图像具有相同的大小,通过在训练期间每层填充零,而来自 SRCNN 的输出小于输入。简单地对所有层使用相同的学习速率，而 SRCNN 对不同层使用不同的学习速率，以实现稳定的收敛。

四、　推荐的方法

(一)　推荐的网络

超分辨率图像重建，使用非常深度卷积网络，配置如图所示。



基于 CNN 的方法可以从超分辨率方法中获益，因为超分辨率方法中经常使用对图像细节建模的方法。

使用非常深度网络来预测密集输出的一个问题：每次应用卷积运算时特征图的大小变小了。

如果所需的环绕区域非常大，这种方法就无效。裁剪后，最终图像太小，视觉效果不佳。为了解决这个问题，在卷积之前填充零以保持所有特征图的大小（包括输出）相同。

一旦图像细节被预测，它们被添加回输入 ILR 图像以给出最终图像 (HR)。

(二)　训练

描述最小化的目标，以便找到模型的最佳参数:

设 x 表示插值的低分辨率图像，y 表示高分辨率图像，给定训练数据集 $\{x^{(i)}, y^{(i)}\}$，目的是预测 $\hat{y} = f(x)$，其中 $\hat{y}$ 是目标 HR 图像的估计值。要使平方误差 $\frac{1}{2}||y - f(x)||^2$ 最小化。

Residual-Learning 利用剩余学习来解决渐晕/梯度爆炸问题问题。定义残差图像 $r = y - x$，要去预测残差图像，损失函数变为 $\frac{1}{2}||r - f(x)||^2$，$f(x)$ 为网络预测。

High Learning Rates for Very Deep Networks 在真实的时间限制下，深层模型可能存在无法收敛的问题。学习率高，以促进训练是常规策略。但是简单地将学习速率设置得很高也会导致梯度消失/爆炸。为此，在抑制爆炸梯度的同时，采用可调的梯度削波来最大限度地提高速度。

Adjustable Gradient Clipping 可调梯度裁剪的使用仅限于训练 CNN。其中一个常用的策略是对个体进行裁剪预定义范围 $[-\theta, \theta]$ 的梯度。随机梯度下降常用于训练，学习率乘以调整步长。如果使用高学习率，很可能将 调整为较小，以避免在高学习率状态下爆发梯度。但是，随着学习率被退火以变得更小，有效梯度 (梯度乘以学习率) 接近零，并且如果学习率以几何方式降低，则训练可能需要指数多次迭代才能收敛。为了达到最快的收敛速度，我们剪切梯度 $[-\frac{\theta}{\gamma}, \frac{\theta}{\gamma}]$，$\gamma$ 表示当前学习速率。

Multi-Scale 使用更多的参数来定义网络。考虑到经常使用分数比例因子，需要一种经济的方法来存储和检索网络。为此，训练了一个多尺度模型。使用这种方法，参数在所有预定义的比例因子之间共享。

残差和非残差网络的性能表 (PSNR) 如图：

Epoch	10	20	40	80
Residual	36.90	36.64	37.12	37.05
Non-Residual	27.42	19.59	31.38	35.66
Difference	9.48	17.05	5.74	1.39

(a) Initial learning rate 0.1

Epoch	10	20	40	80
Residual	36.74	36.87	36.91	36.93
Non-Residual	30.33	33.59	36.26	36.42
Difference	6.41	3.28	0.65	0.52

(b) Initial learning rate 0.01

Epoch	10	20	40	80
Residual	36.31	36.46	36.52	36.52
Non-Residual	33.97	35.08	36.11	36.11
Difference	2.35	1.38	0.42	0.40

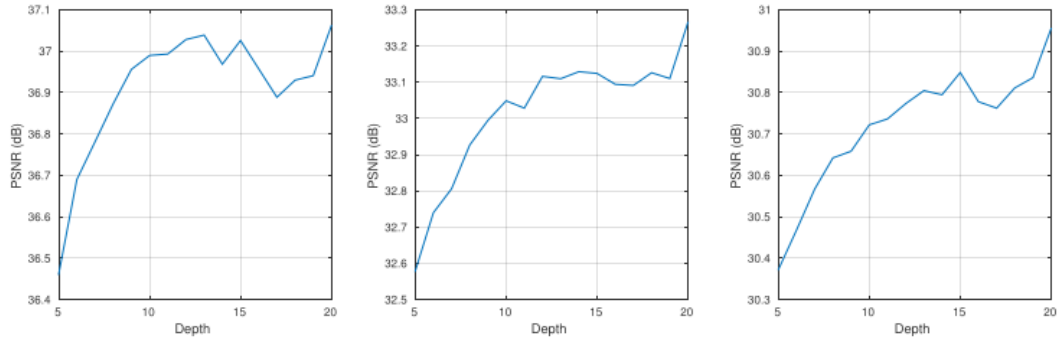
(c) Initial learning rate 0.001

五、 了解性质

讨论了作者提出的方法的三个性质：

- 学习深度越深越好；
- 该剩余学习网络可以比标准的 CNN 快得多且大大提高了性能；
- 该网络是适用于多种规模的单一模型。

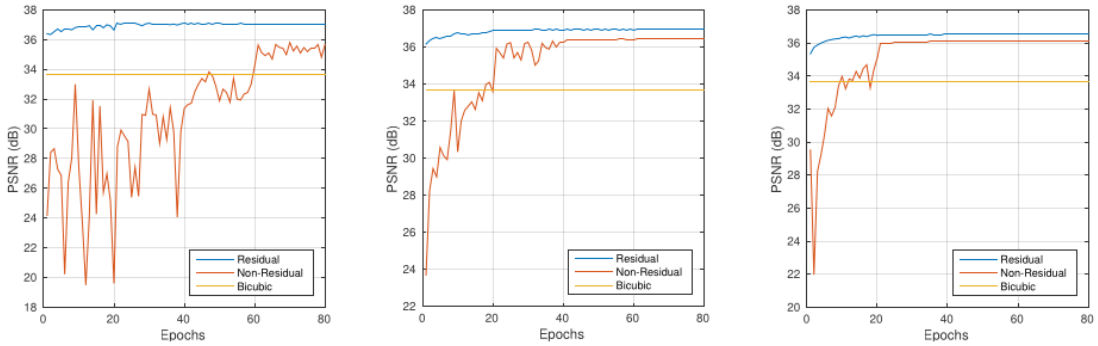
通过实验证明，非常深的网络可以显著提高随机共振的性能。训练和测试深度范围从 5 到 20 的网络 (仅计算权重层，不包括非线性层)，结果如图：



(a) Test Scale Factor 2　**(b)** Test Scale Factor 3　**(c)** Test Scale Factor 4

使用深度 10(权重层) 和比例因子 2。不同学习率下的性能曲线如图:

Figure 3: Depth vs Performance



(a) Initial learning rate 0.1　**(b)** Initial learning rate 0.01　**(c)** Initial learning rate 0.001

在收敛时，剩余网络表现出优越的性能。

六、　实验结果

描述用于训练和测试我们方法的数据集后，给出训练所需的参数，将该方法与几种最先进的 SISR 方法进行了比较。

(一)　用于训练和测试的数据集

作者采用了参考文献中使用的各种不同类型具有挑战性的数据集。

(二)　训练参数

使用深度为 20 的网络。训练使用大小为 64 的 batch size。动量和重量衰减参数设置为 0.9 和 0.0001。在 80 个时期内训练所有实验，9960 次迭代。学习率最初设置为 0.1，每 20 个时期减少 10 倍。总体而言，学习率降低了 3 倍，并且在 80 个周期后停止学习。

(三)　基准测试

遵循公开可用的黄等人的框架，该框架将双三次插值应用于图像的颜色分量，裁剪图像边界附近的像素。

(四)　与最先进方法的比较

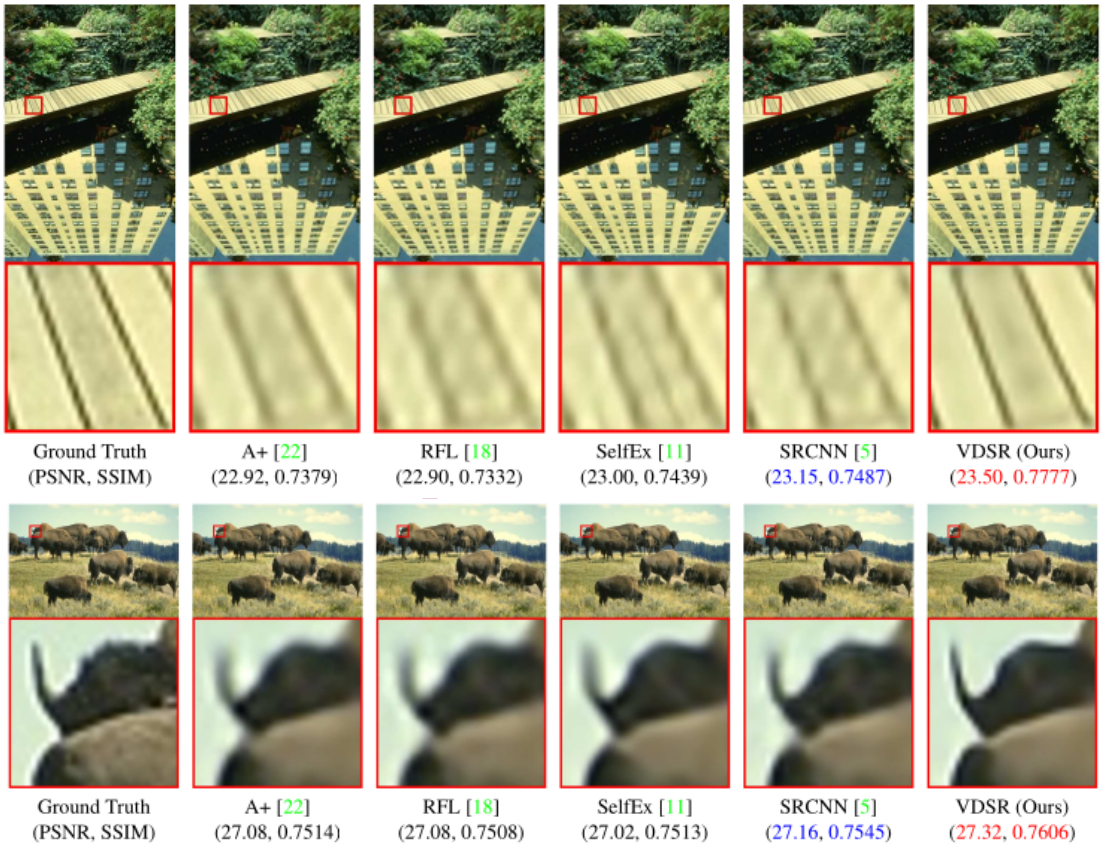
分为定量和定性的比较，与 A+, RFL, SelfEx 和 SRCNN 进行比较。

下表是定量比较的结果：

Dataset	Scale	Bicubic PSNR/SSIM/time	A+ [22] PSNR/SSIM/time	RFL [18] PSNR/SSIM/time	SelfEx [11] PSNR/SSIM/time	SRCNN [5] PSNR/SSIM/time	VDSR (Ours) PSNR/SSIM/time
Set5	×2	33.66/0.9299/0.00	36.54/0.9544/0.58	36.54/0.9537/0.63	36.49/0.9537/45.78	36.66/0.9542/2.19	37.53/0.9587/0.13
	×3	30.39/0.8682/0.00	32.58/0.9088/0.32	32.43/0.9057/0.49	32.58/0.9093/33.44	32.75/0.9090/2.23	33.66/0.9213/0.13
	×4	28.42/0.8104/0.00	30.28/0.8603/0.24	30.14/0.8548/0.38	30.31/0.8619/29.18	30.48/0.8628/2.19	31.35/0.8838/0.12
Set14	×2	30.24/0.8688/0.00	32.28/0.9056/0.86	32.26/0.9040/1.13	32.22/0.9034/105.00	32.42/0.9063/4.32	33.03/0.9124/0.25
	×3	27.55/0.7742/0.00	29.13/0.8188/0.56	29.05/0.8164/0.85	29.16/0.8196/74.69	29.28/0.8209/4.40	29.77/0.8314/0.26
	×4	26.00/0.7027/0.00	27.32/0.7491/0.38	27.24/0.7451/0.65	27.40/0.7518/65.08	27.49/0.7503/4.39	28.01/0.7674/0.25
B100	×2	29.56/0.8431/0.00	31.21/0.8863/0.59	31.16/0.8840/0.80	31.18/0.8855/60.09	31.36/0.8879/2.51	31.90/0.8960/0.16
	×3	27.21/0.7385/0.00	28.29/0.7835/0.33	28.22/0.7806/0.62	28.29/0.7840/40.01	28.41/0.7863/2.58	28.82/0.7976/0.21
	×4	25.96/0.6675/0.00	26.82/0.7087/0.26	26.75/0.7054/0.48	26.84/0.7106/35.87	26.90/0.7101/2.51	27.29/0.7251/0.21
Urban100	×2	26.88/0.8403/0.00	29.20/0.8938/2.96	29.11/0.8904/3.62	29.54/0.8967/663.98	29.50/0.8946/22.12	30.76/0.9140/0.98
	×3	24.46/0.7349/0.00	26.03/0.7973/1.67	25.86/0.7900/2.48	26.44/0.8088/473.60	26.24/0.7989/19.35	27.14/0.8279/1.08
	×4	23.14/0.6577/0.00	24.32/0.7183/1.21	24.19/0.7096/1.88	24.79/0.7374/394.40	24.52/0.7221/18.46	25.18/0.7524/1.06

可见该方法优于以前所有方法并且相对较快。

下面两图是与性能最好的方法比较。第一张图中，只有该方法完美地重构了中间的线。第二张图中，轮廓在该方法中是清晰和生动的，而在其他方法中它们是严重模糊或失真的。



（五）　结论

作者提出了一个使用非常深度的网络来超分辨率的方法。由于收敛速度慢，训练非常深的网络很困难，于是使用剩余学习和极高的学习率来快速优化非常深度的网络。收敛速度被最大化，并且使用梯度裁剪来保证训练的稳定性。已经证明了该方法在基准图像上比现有的方法有更大的优势。今后该方法可能会应用于其他图像恢复问题，如去噪和压缩伪像消除。

第四章　SRGAN

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，是通过对 SR 的特定损失函数 l^{SR} 进行优化得到。对于训练图片 $I_n^{HR}, n = 1, ..., N$ 及相应的 $I_n^{LR}, n = 1, ..., N$，我们求解：

$$\hat{\theta}_G = arg \ \min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{SR}(G_{\theta_G}(I_n^{LR}), I_n^{HR}) \tag{4.1}$$

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 $G\theta_G$ 以一种交替的方式来优化它，来解决对抗的 min-max 问题：

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{train}(I^{HR})}[logD_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))] \tag{4.2}$$

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 4.2 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：

$$l^{SR} = \underbrace{\underbrace{l_X^{SR}}_{content\ loss} + \underbrace{10^{-3} l_{Gen}^{SR}}_{adversarial\ loss}}_{perceptual\ loss\ (for\ VGG\ based\ content\ losses)} \tag{4.3}$$

内容损失

像素级 MSE 损失函数的计算方法：

$$l_{MSE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2 \tag{4.4}$$

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

$$l^{SR}_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2 \tag{4.5}$$

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

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^{SR}_{Gen} 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:

$$l^{SR}_{Gen} = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \tag{4.6}$$

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。