

人工智能大作业——图像超分实验

1. 实验目标

输入大小为 $h \times w$ 的图像 X ，输出为一个 $sh \times sw$ 的图像 Y ， s 为放大倍数。

2. 数据集简介

本次实验采用的是 BSDS500 数据集，其中训练集包含 200 张图像，验证集包含 100 张图像，测试集包含 200 张图像。

数据集来源：

https://download.csdn.net/download/weixin_42028424/11045313

3. 数据预处理

数据预处理包含两个步骤：

(1) 将图片转换成 YCbCr 模式

由于 RGB 颜色模式色调、色度、饱和度三者混在一起难以分开，因此将其转换成 YcbCr 颜色模式，Y 是指亮度分量，Cb 表示 RGB 输入信号蓝色部分与 RGB 信号亮度值之间的差异，Cr 表示 RGB 输入信号红色部分与 RGB 信号亮度值之间的差异。

(2) 将图片裁剪成 300×300 的正方形

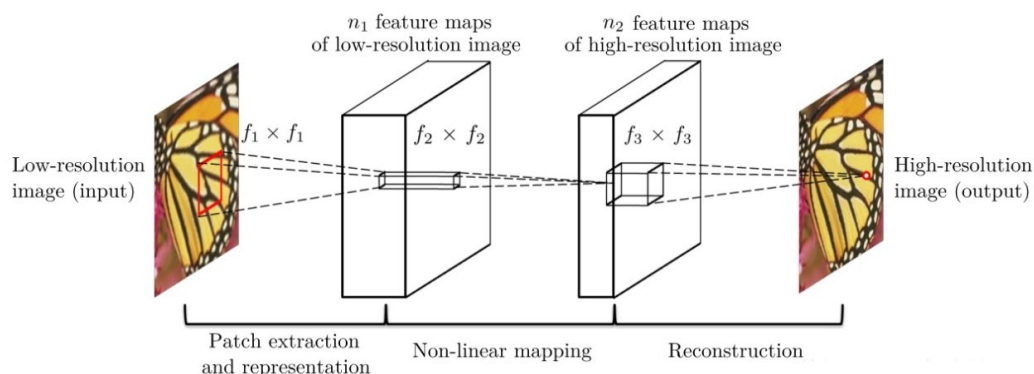
由于后面采用的神经网络输入图片要求长宽一致，而 BSDS500 数据集中的图片长宽并不一致，因此需要对其进行裁剪。这里采用的方式是先定位到每个图片中心，然后以图片中心为基准，向四个方向拓展 150 个像素，从而将图片裁剪成 300×300 的正方形

4. 网络结构

本次实验尝试了 SRCNN 和 FSRCNN 两个网络。

4.1 SRCNN

SRCNN 由 2014 年 Chao Dong 等人提出，是深度学习在图像超分领域的开篇之作。其网络结构如下图所示：



该网络对于一个低分辨率图像，先使用双三次插值将其放大到目标大小，再通过三层卷积网络做非线性映射，得到的结果作为高分辨率图像输出。

作者对于这三层卷积层的解释：

(1) 特征块提取和表示：此操作从低分辨率图像 Y 中提取重叠特征块，并将每个特征块表示为一个高维向量。这些向量包括一组特征图，其数量等于向量的维数。

(2) 非线性映射：该操作将每个高维向量非线性映射到另一个高维向量。每个映射向量在概念上都是高分辨率特征块的表示。这些向量同样包括另一组特征图。

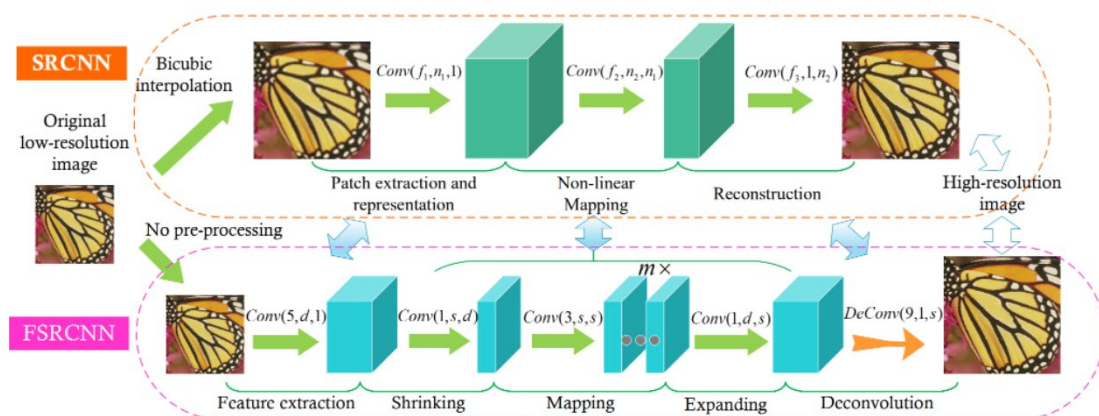
(3) 重建：该操作聚合上述高分辨率 patch-wise（介于像素级别和图像级别的区域）表示，生成最终的高分辨率图像。

各层结构：

- 输入：处理后的低分辨率图像
- 卷积层 1：采用 9×9 的卷积核
- 卷积层 2：采用 1×1 的卷积核
- 卷积层 3：采用 5×5 的卷积核
- 输出：高分辨率图像

4.2 FSRCNN

FSRCNN 由 2016 年 Chao Dong 等人提出，与 SRCNN 是相同作者。其网络结构如下图所示：



FSRCNN 在 SRCNN 基础上做了如下改变：

(1)FSRCNN 直接采用低分辨率的图像作为输入，不同于 SRCNN 需要先对低分辨率的图像进行双三次插值然后作为输入；

(2)FSRCNN 在网络的最后采用反卷积层实现上采样；

(3)FSRCNN 中没有非线性映射，相应地出现了收缩、映射和扩展；

(4)FSRCNN 选择更小尺寸的滤波器和更深的网络结构。

各层结构：

- 输入层：FSRCNN 不使用 bicubic 插值来对输入图像做上采样，它直接进入特征提取层
- 特征提取层：采用 $1 \times d \times (5 \times 5)$ 的卷积层提取
- 收缩层：采用 $d \times s \times (1 \times 1)$ 的卷积层去减少通道数，来减少模型复杂度
- 映射层：采用 $s \times s \times (3 \times 3)$ 卷积层去增加模型非线性度来实现低分辨率到高分辨率的映射
- 扩张层：该层和收缩层是对称的，采用 $s \times d \times (1 \times 1)$ 卷积层去增加重建的表现力
- 反卷积层： $s \times 1 \times (9 \times 9)$
- 输出层：输出高分辨率图像

5. 评估指标

本次实验尝试了 PSNR 和 SSIM 两个指标。

5.1 PSNR

PSNR(Peak Signal to Noise Ratio)为峰值信噪比，计算公式如下：

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - Y(i, j))^2$$

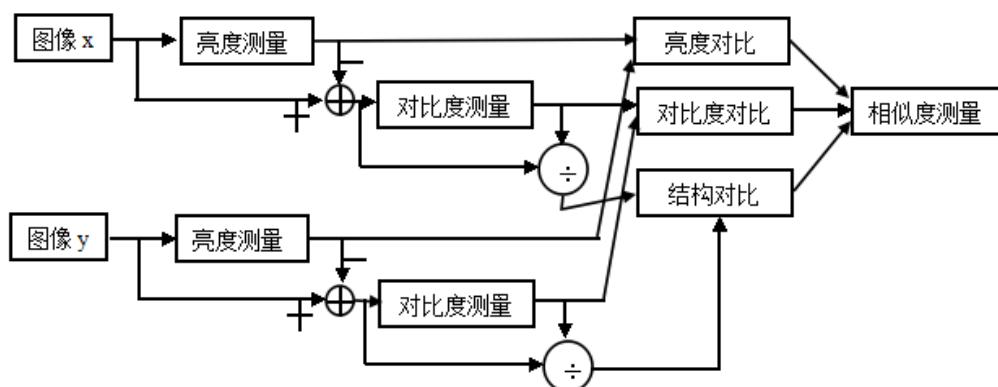
$$PSNR = 10 \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right)$$

其中，n 为每像素的比特数。

PSNR 的单位是 dB，数值越大表示失真越小，一般认为 PSNR 在 38 以上的时候，人眼就无法区分两幅图片了。

5.2 SSIM

SSIM(Structural Similarity)为结构相似性，由三个对比模块组成：亮度、对比度、结构。



(1) 亮度对比函数

图像的平均灰度计算公式：

$$\mu_X = \frac{1}{H \times M} \sum_{i=1}^H \sum_{j=1}^M X(i, j)$$

亮度对比函数计算公式：

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

(2) 对比度对比函数

图像的标准差计算公式：

$$\sigma_X = \left(\frac{1}{H \times M} \sum_{i=1}^H \sum_{j=1}^M (X(i, j) - \mu_X)^2 \right)^{\frac{1}{2}}$$

对比度对比函数计算公式：

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

(3) 结构对比函数

结构对比函数计算公式：

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

综合上述三个部分，得到 SSIM 计算公式：

$$SSIM(x, y) = f(l(x, y), c(x, y), s(x, y)) \\ = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma$$

其中， $\alpha, \beta, \gamma > 0$ ，用来调整这三个模块的重要性。

SSIM 函数的值域为[0, 1]，值越大说明图像失真越小，两幅图像越相似。

6. 模型训练/测试

设定 epoch 为 500 次，保存验证集上 PSNR 最高的模型。两个模型在测试集上的表现如下表所示：

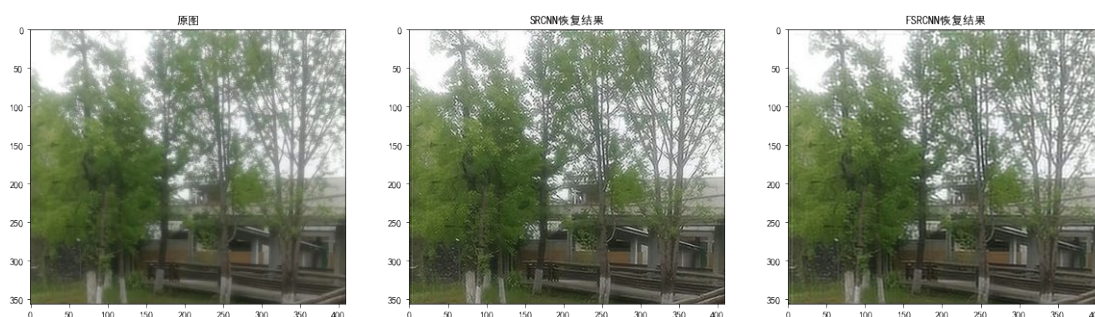
	SRCNN	FSRCNN
平均 PSNR (db)	28.57	26.98
平均 SSIM	0.92	0.93

从结果可以发现，FSRCNN 的 PSNR 比 SRCNN 低，但 FSRCNN 的 SSIM 比 SRCNN 高，说明 PSNR 和 SSIM 并不存在完全正相关的关系。

7. 实图测试

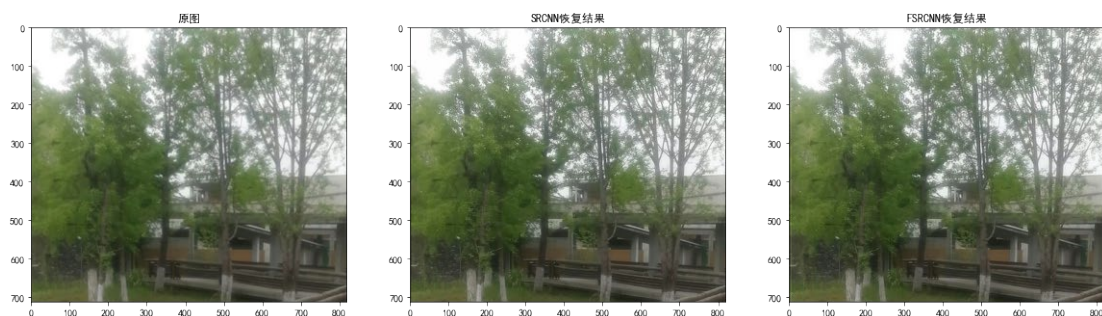
为了直观感受两个模型的效果，我用自己拍摄的图进行实图测试，效果如下：

s=1 (放大倍数=1)



当放大倍数=1 时，SRCNN 的超分结果比 FSRCNN 的超分效果要更好一些，这和两个模型平均 PSNR 的数值相吻合。

s=2 (放大倍数=2)



当放大倍数=2 时，SRCNN 的超分结果和 FSRCNN 的超分效果相差不大。