**计算机视觉实践-练习3**

许余 123106222809

1. **实验目标**

实现一种图像超分辨率方法在Set5数据集上的测试，得到超分辨率图像，测量其与原始真实图像之间的PSNR，SSIM指标值。对所选择方法的细节进行介绍，并试着讨论该方法可能存在的优缺点，以及可能改进的方向。

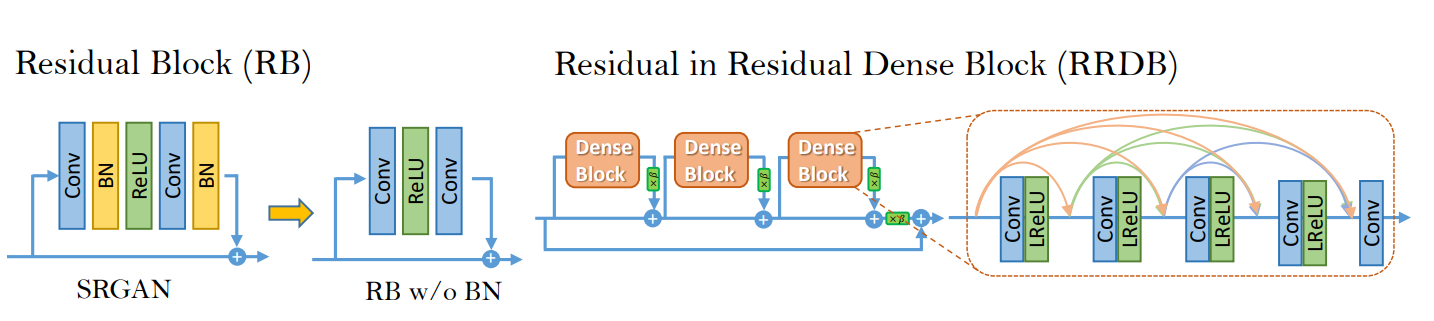
测试方式：先将图像用Bicubic插值进行下采样，再使用超分辨率算法进行处理，将得到的超分辨率图像与真实的原始图像进行对比。

1. **实验原理**
2. ESRGAN

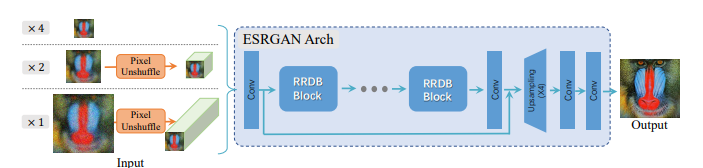
ESRGAN是ECCV2018的文章，该方法在PIRM218-SR比赛取得冠军。论文分析SRGAN能够生成更多的纹理细节，但它纹理往往不够自然，也常伴随着一些噪声。然后深入研究并改进了SRGAN的三个关键部分——网络结构、对抗损失函数和感知损失函数。ESRGAN针对SRGAN进行了以下改进。

1. 新的生成器结构

如图2.1，相对于SRGAN，去掉了所有的BN层，高模型的泛化能力，减少计算复杂度和内存占用。提出RRDB(Residual in Residual Dense Block)，这里的Dense指的不是全连接而是卷积层中有着密集的残差链接，这样做的好处是可以获得更深入、更复杂的结构，网络容量也变得更高。



**图2.1 RRDB结构**



**图2.2 ESRGAN生成器结构**

1. 网络插值

将残差信息乘以一个 0 到 1 之间的数（通过实验最终确定0.2），这样可以使训练更稳定，在保持纹理的同时的减少伪影。

1. 判别器

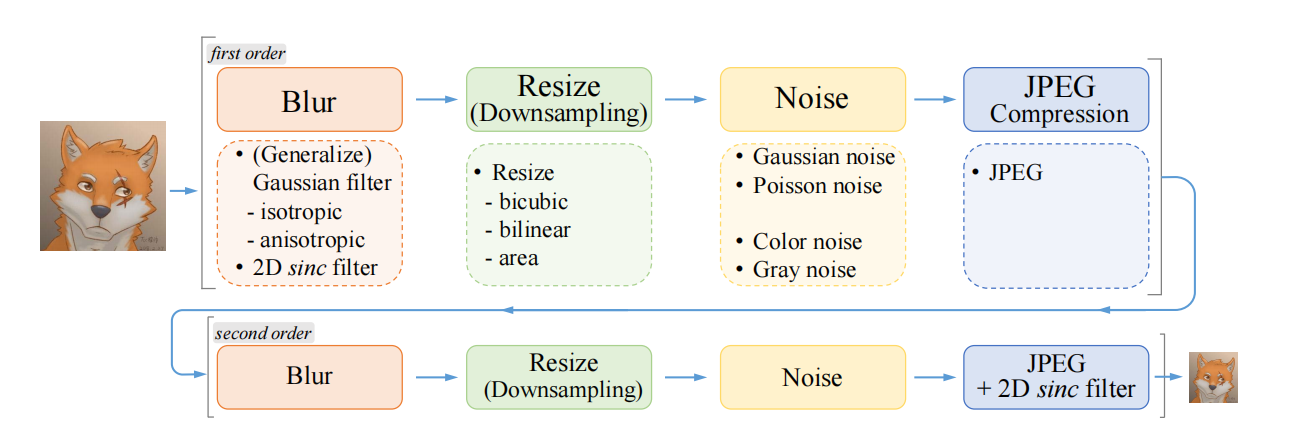
使用RaGAN改进了判别器，它学习判断“一幅图像是否比另一幅图像更真实”，而不是“一幅图像是真实的还是假的”。

1. Real-ESRGAN

Real-ESRGAN与 ESRGAN相比采用了同样的生成器架构,但是相对于ESRGAN在其它方面做了一些改进。

1. 数据高阶退化模型

高阶退化模型（High-order Degradation Model）是Real-ESRGAN最重要的创新点。经典的退化模型不能模拟一些复杂的退化问题，特别是未知的噪声和复杂的伪影，这是因为合成的低分辨率图像与现实的退化图像仍然有很大的差距。因此，Real-ESRGAN将经典的退化模型扩展到高阶过程，以模拟更实际的退化。所谓高阶退化模型通俗的说就是将经典退化算法排列组合，论文将退化算法分为Blur、Resize、Noise、JPEG Compression四类，如图2.3。



**图2.3 Real-ESRGAN的退化算法**

从论文代码中可以看出，整个退化模型循环两遍上面四种退化过程，每个过程随机选一种算法。

1. U-Net判别器

Real-ESRGAN中的鉴别器对复杂的训练输出需要更大的鉴别能力，它还需要为局部纹理产生精确的梯度反馈，而不是只区分全局样式。因此使用更加强大的U-Net作为判别器。输出每个像素的真实度值，并可以向生成器提供详细的每像素反馈，增强了图像对细节上的对抗学习。

1. 谱归一化

论文将ESRGAN中的VGG鉴别器改进为U-Net设计。U-Net结构和复杂的退化也增加了训练的不稳定性。因此，作者采用谱归一化（SN）来稳定训练，实现局部细节增强和伪影抑制的良好平衡。

1. **实验过程**
2. Set5图像下采样

使用Python的Pillow库，以下采样因子4对Set5数据集中的图像进行处理，采用Bicubic插值方法，然后保存用于后续的超分辨率操作，这一步的代码保存为sample.py，下采样结果图片放在Set5\_0.25文件夹。

1. 图像超分辨率

下载Real-ESRGAN的官方项目，使用官方提供的realesrgan-x4plus模型，对上一步下采样得到的图像进行超分辨率，使分辨率恢复到原来的大小，超分辨率的图像放在Set5\_0.25\_x4文件夹。

1. 评估

使用Python的skimage.metrics库，测量其与原始真实图像之间的峰值信噪比（PSNR）和结构相似性指标（SSIM），并对结果进行分析，这一步的代码保存在evaluate.py。

1. **结果分析**

**表4.1 图像超分辨率的评估结果**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| image | baby | bird | butterfly | head | woman |
| resolution | 512x512 | 288x288 | 256x256 | 280x280 | 228x344 |
| PSNR | 29.165 | 25.499 | 21.390 | 28.118 | 24.933 |
| SSIM | 0.820 | 0.767 | 0.759 | 0.666 | 0.808 |

PSNR是一个衡量图像质量的客观指标，值越高表示图像质量越好，失真越少。在这组数据中，baby图像的PSNR最高（29.165），表明其超分辨率后的图像质量最好，失真最少。butterfly的PSNR最低（21.390），表明在超分辨率处理过程中可能引入了较多的噪声或失真。

SSIM是一个衡量两幅图像相似度的指标，范围在-1到1之间，值越接近1表示两幅图像越相似。在超分辨率任务中，通常用来比较超分辨率后的图像与原始高分辨率图像之间的相似度。baby图像的SSIM值最高（0.820），意味着其超分辨率后的图像与原始高分辨率图像的结构最为相似。head图像的SSIM值相对较低（0.666），可能意味着在超分辨率处理过程中，图像的结构信息损失较多。

1. **改进方向**

由于Real-ESRGAN是在ESRGAN的基础上改进而来，所以可以按照前者的改进思路去做一些尝试。

众所周知，CNN 在长距离建模能力方面存在局限性，而 Transformer 随能全局建模但受到其二次计算复杂度的限制。因此，以 Mamba 为代表的状态空间模型（SSMs）已然成为一种有前景的方法。首先，它不仅擅长于建模长距离交互作用，同时还能保持线性计算复杂度。有学者提出VM-UNet模型，即在U-Net引入了视觉状态空间（VSS）作为捕捉广泛上下文信息的基础块，并构建了一个不对称的编码器-解码器结构。可以考虑使用VM-UNet作为判别器结构，但是VM-UNet的复杂结构进一步增大了训练的不稳定性，考虑采用谱归一化进行稳定训练。