深度学习与计算机视觉课程项目

——超分辨率

一、选择了什么问题

当前，大多数超分辨率方法都依赖于低分辨率和高分辨率图像对，以全监督的方式训练网络。但是，这种图像对在实际应用中比较难以实现，目前常规方法是使用下采样方法来人工生成相应的低分辨率图像。但这种方法在处理过程中会引入伪影并去除一些自然的图像特性，这种缺陷使得利用这种人工处理图像训练的超分辨率网络很难推广到自然图像上。

基于这个问题，这次的挑战比赛设计了基于源域和目标域的图像超分任务，本次报告我们组选择了 Real World Super-Resolution Challenge: Track 1 Same Domain，这项超分任务是使模型能够基于源域的输入图像，生成最优质量的超分辨率图像。

二、这个问题的特点和数据是什么样的

本次比赛提供的数据是 DIV2K， Flickr2K 两个数据集，总共有 2650 张自然图像，其中分为 HR 高清图像训练集，大小为 2040*1404，然后还有缩小各种倍数的低分辨率验证集和测试集，本次实验使用的是源数据集中的图像。

三、采用了何种方法解答该问题

我们对比了 EDSR、SRGAN 和 ESRGAN 的结果。首先看一下 EDSR 和 SRGAN 的模型框架：

3.1 EDSR 模型框架

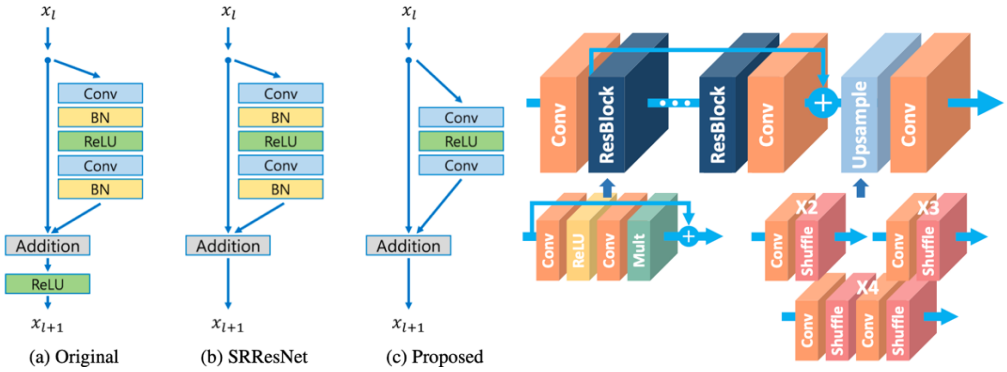


图 1 Residual blocks（左），EDSR 网络框架（右）

如图 1，EDSR 的残差块把 BN 层移除掉，而且和 SRResNet 相似，相加后不经过 relu 层。框架结构和 SRResNet 非常相似，但移除了 BN 层和大多数 ReLU 层，ReLU 层只在残差块里才有。最终的训练版本有 B=32 个残差块，F=256 个通道。并且在训练 x3，x4 模型时，采用 x2 的预训练参数。

3.2 SRGAN 模型框架

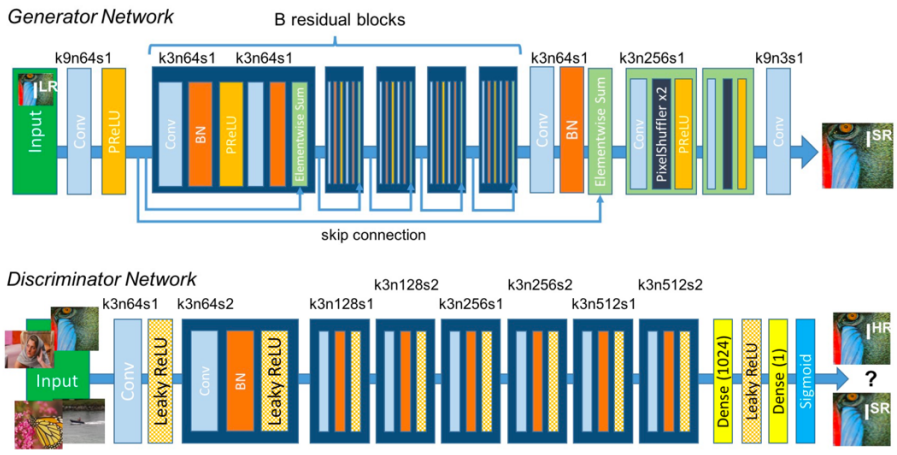


图 2 SRGAN

SRGAN 是在 SRResNet 的基础上加上一个鉴别器。GAN 的作用，是额外增加一个鉴别器网络和 2 个损失（G_loss 和 D_loss），用一种交替训练的方式训练两个网络。

生成网络中，如图 2（上），包含多个残差块，每个残差块中包含两个 3×3 的卷积层，卷积层后接 BN 层，PReLU 作为激活函数。最后接两个 2× 亚像素卷积层(sub-pixel convolution layers)用来增大特征尺寸。

在判别网络中，如图 2（下），包含 8 个卷积层，随着网络层数加深，特征个数不断增加，特征尺寸不断减小，选取的激活函数为 LeakyReLU，最后通过两个全连接层和最终的 sigmoid 激活函数得到预测为自然图像的概率。

EDSR 和 ESRGAN 都使用了已经训练好的模型来测试效果。用 SRGAN 训练了 2000 个 epoch。

四、最终的模型架构是什么

对比结果是 ESRGAN 的感知效果最好。所以使用了 ESRGAN 的模型。ESRGAN 是在 SRGAN 的基础上进行改进得到的模型：

4.1 生成网络的改进

4.1.1 对残差块（Residual Block）的改进

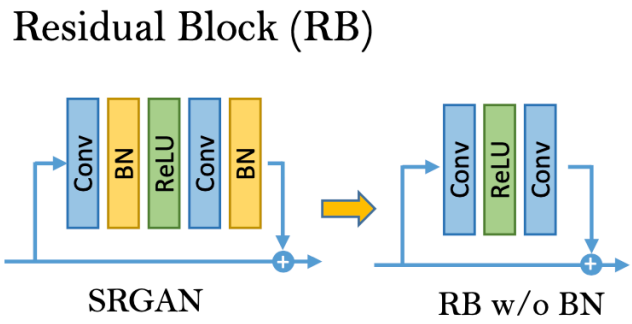


图 3 SRGAN（左）与 ESRGAN（右）的残差块比较

SRGAN 因为使用了 BN，所以会产生伪影。因此 ESRGAN 去除了 BN，对于不以 PSNR 为主要指标的任务（如超分辨率和去模糊）来说，去掉 BN 层可以提高视觉效果、减小计算复杂度和内存占用。

4.1.2 对残差块连接的改进

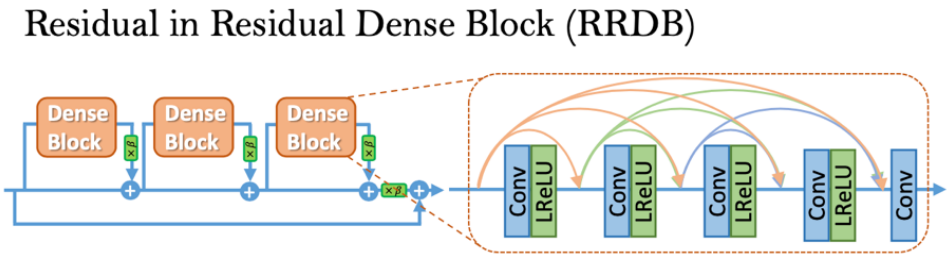


图 4 ESRGAN 残差块的连接

SRGAN 的残差块是顺序连接的，ESRGAN 把这些残差块用密集连接的方式连在一起。生成网络里的特征提取部分最终变成了图 4。

4.2 判别网络的改进

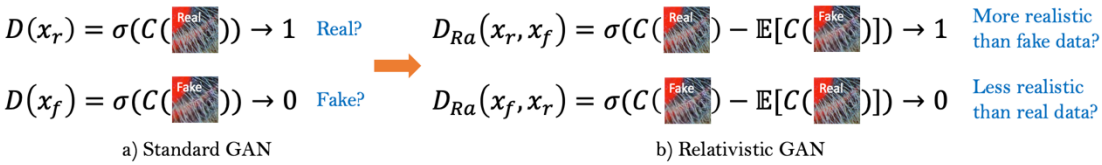


图 5 标准判别器和 Relativistic 判别器

ESRGAN 基于 Relativistic GAN 改进了判别网络，把标准的判别器换成 Relativistic average Discriminator(RaD)。SRGAN 中的判别器用于估计输入到判别器中的图像是真实且自然图像的概率，而 ESRGAN 的判别器则是尝试估计真实图像相对来说比 fake 图像更逼真的概率。

判别器的损失函数定义为：

$$L_D^{Ra} = -\mathbb{E}_{x_r}[\log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[\log(1 - D_{Ra}(x_f, x_r))]$$

对应的生成器的对抗损失函数为：

$$L_G^{Ra} = -\mathbb{E}_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[\log(D_{Ra}(x_f, x_r))]$$

4.3 感知域损失

感知损失是利用卷积神经网络提取出的特征，通过比较生成图片经过卷积神经网络后的特征和目标图片经过卷积神经网络后的特征的差别，使生成图片和目标图片在语义和风格上更相似。感知域损失定义在预训练的深度网络的激活层，这一层中两个激活后的特征的距离会被最小化。

与之相反，ESRGAN 提出了一个更有效的感知域损失，使用了激活前的特征，用一个训练好的 VGG19 来给出特征。这样会克服两个缺点：第一，激活后的特征是非常稀疏的，特别是在很深的网络中。这种稀疏的激活提供的监督效果是很弱的，会造成性能低下；第二，使用激活后的特征会导致重建图像与 ground-truth 的亮度不一致。

所以，最终的生成器损失函数为：

$$L_G = L_{\text{percep}} + \lambda L_G^{Ra} + \eta L_1$$

五、怎么训练模型以及模型在测试数据集上表现

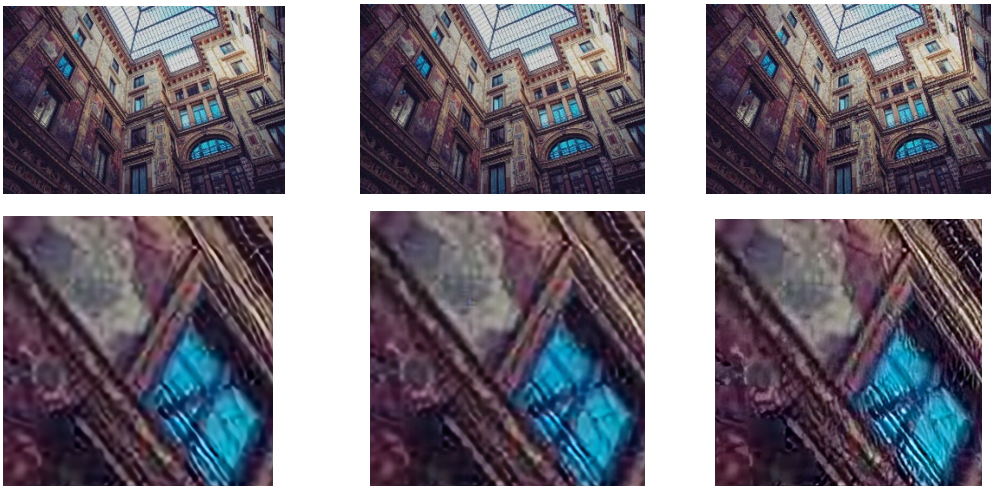
5.1 实验

本次实验使用 EDSR, ESRGAN, SRGAN 三种架构进行实验，使用的数据集：训练集 2650 张

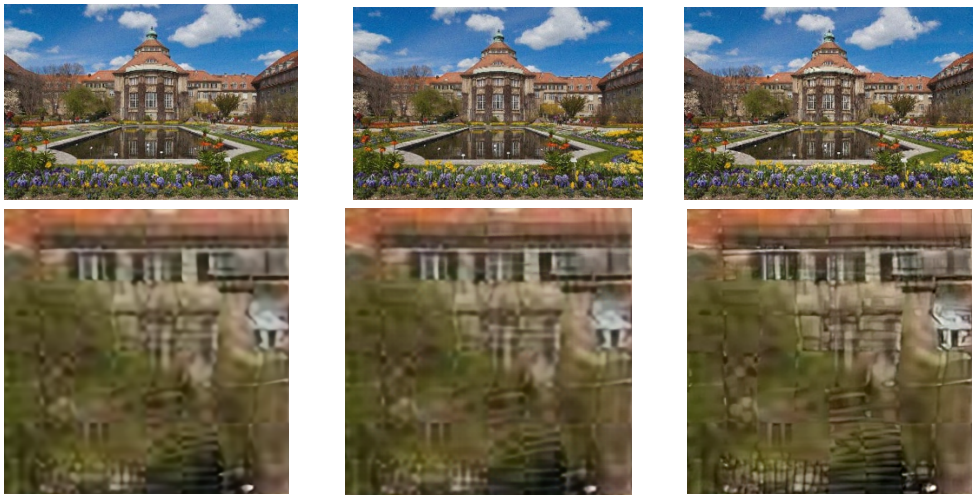
图片，验证集 100 张图片，实验环境为：linux 服务器，NVIDIA DGX-2:Tesla V100 32G*2。每个网络运行 2000 轮，多次调整参数，得到最终的 test_imagine 程序，下面为实验结果：

Esrgan_result　　edsr_img　　Srgan_result

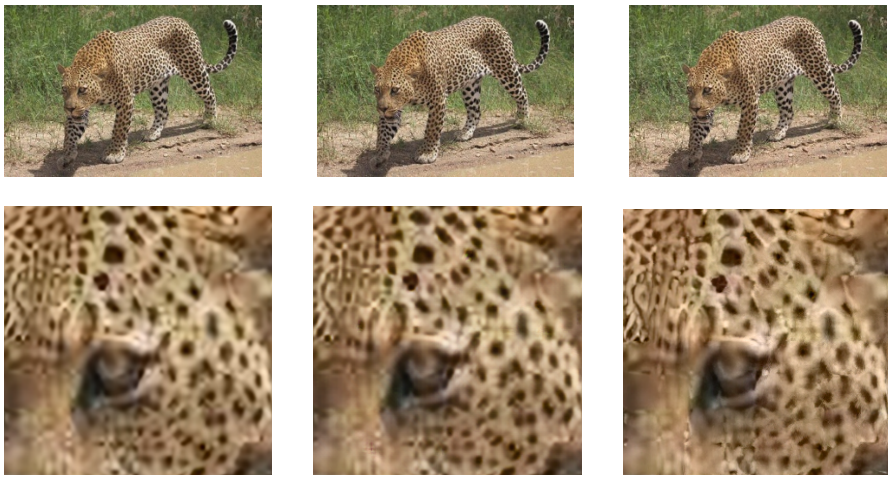
第一组



第二组



第三组



5.2 图像质量评价

图像质量评价可以分为主观评价方法和客观评价方法。

5.2.1 主观评价：

观察者对图像质量进行主观评分，一般采用平均主观得分(Mean opin-ion score, MOS)或平均主观得分差异(Direrential mean opinion score, DMOS)（即人眼对无失真图像和有失真图像评价得分的差异），本小组寻找第三方评分人员测得评分如下：

模型	EDSR	ESRGAN	SRGAN
MOS 得分	4.0	4.6	4.3

5.2.2 客观评价：

本次实验使用 psdr 和 ssim 进行评分：PSNR（Peak Signal to Noise Ratio），峰值信噪比，即峰值信号的能量与噪声的平均能量之比，通常表示的时候取 log 变成分贝（dB），由于 MSE 为真实图像与含噪图像之差的能量均值，而两者的差即为噪声，因此 PSNR 即峰值信号能量与 MSE 之比；SSIM（structural similarity）结构相似性，也是一种全参考的图像质量评价指标，它分别从亮度、对比度、结构三方面度量图像相似性。

SRGAN 的 PSNR 和 SSIM 值：

```
[converting LR images to SR images] PSNR: 24.7648 dB SSIM: 0.6971: 100%|████████| 100/100 [00:05<00:00,
17.59it/s]
```

5.2.3 讨论：

这些实验显示出优越的感知能力。拟议框架的性能完全基于视觉比较。标准量化指标，PSNR 和 SSIM 显然无法捕获和准确评估关于人类视觉系统的图像质量[47]。因为 PSNR 不足以量 SR 结果的感知质量，未来的工作将通过收集主观的平均意见得分（mos）等指标。焦点在这部作品中，超分辨率图像而不是计算效率。初步实验在网络架构上，建议较浅的网络有潜力在质量表现的小幅下降。更深的网络体系结构以及知觉损失将是未来工作的一部分。我们发现了更深层次的网络架构有益于结果的改善。

5.2.4 结论：

1）我们使用三种网络：EDSR,SRGAN,ESRGAN 在一个数据集上进行实验，其中 esrgan 的结果最好，mos 得分最高。

2）训练轮数太少，各个模型只训练了 2000 轮，导致结果并不是很出色，需要增加训练时间。