SRGAN(Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network)

在以往的超分辨率重建任务中，大家都使用最小化超分辨后的生成图像与真实高分辨率图像之间的MSE和最大化信噪比的方式来做评估，MSE的公式为：

The pixel-wise **MSE loss** is calculated as:

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

但是，使用这种这种最小化逐像素的相似性差异来进行优化的方法存在局限性，我们在最小化MSE的时候，也同时平均掉了高频的结构细节，使得图像的重建结果过于平滑（磨皮磨过头了）。而人的视觉感官对于图像的高频信息比较敏感，高频信息的损失会使得图像的视觉效果变差。同样， 更高的信噪比也不能说明重建得到的超分辨图像就一定是更好的结果。

作者还给出一张对比图来说明使用以往的传统最小化MSE评估结果的差距：

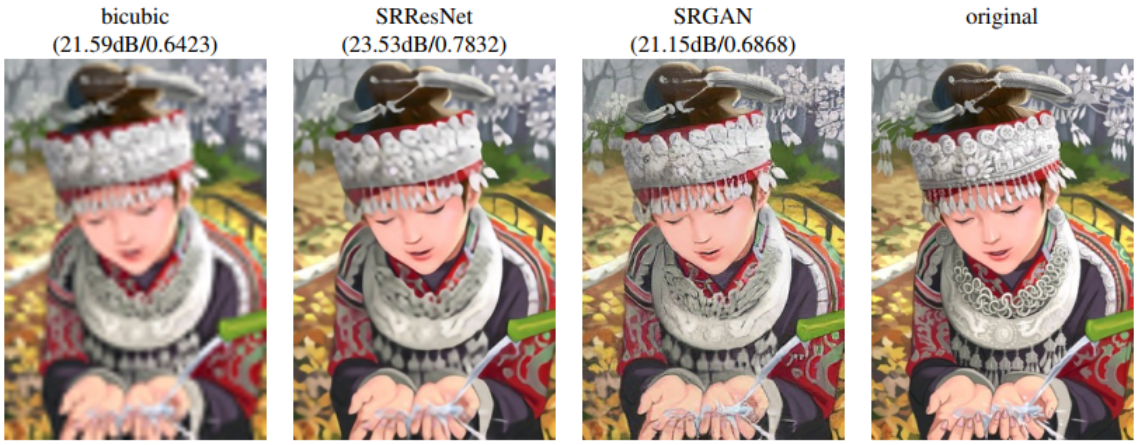

Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

其中上图中的左二是使用ResNet以MSE作为评估结果得到的，左三是使用GAN以作者自己提出的感知损失函数作为评估结果得到的。虽然这并严格满足控制变量的对比实验原则，但是也可以意会一下使用MSE过度磨皮的结果。

于是乎，作者自己提出了一种新的感知评估函数，用VGG based内容损失函数来代替MSE损失，同时结合了当下流行的GAN网络，来提高图像重建结果与原图之间的逼真度：

to perceptually relevant characteristics. We formulate the perceptual loss as the weighted sum of a content loss (l_{X}^{SR}) and an adversarial loss component as:

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

In the following we describe possible choices for the content loss l_{X}^{SR} and the adversarial loss l_{Gen}^{SR}.

这个损失函数主要由内容损失函数和对抗损失函数组成。其中内容损失函数如下：

VGG19 network, which we consider given. We then define the VGG loss as the euclidean distance between the feature representations of a reconstructed image $G_{\theta_G}(I^{LR})$ and the reference image I^{HR}:

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

Here $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the respective feature maps within the VGG network.

对抗损失函数如下：

fool the discriminator network. The generative loss l_{Gen}^{SR} is defined based on the probabilities of the discriminator $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ over all training samples as:

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

Here, $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ is the probability that the reconstructed image $G_{\theta_G}(I^{LR})$ is a natural HR image. For better gradient behavior we minimize $-\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$ instead of $\log[1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$ [22].

作者除了在提出了一个使用VGG based的感知函数以为，还使用了当下流行的GAN模型，下面是作者基于ResNet搭建的GAN网络：

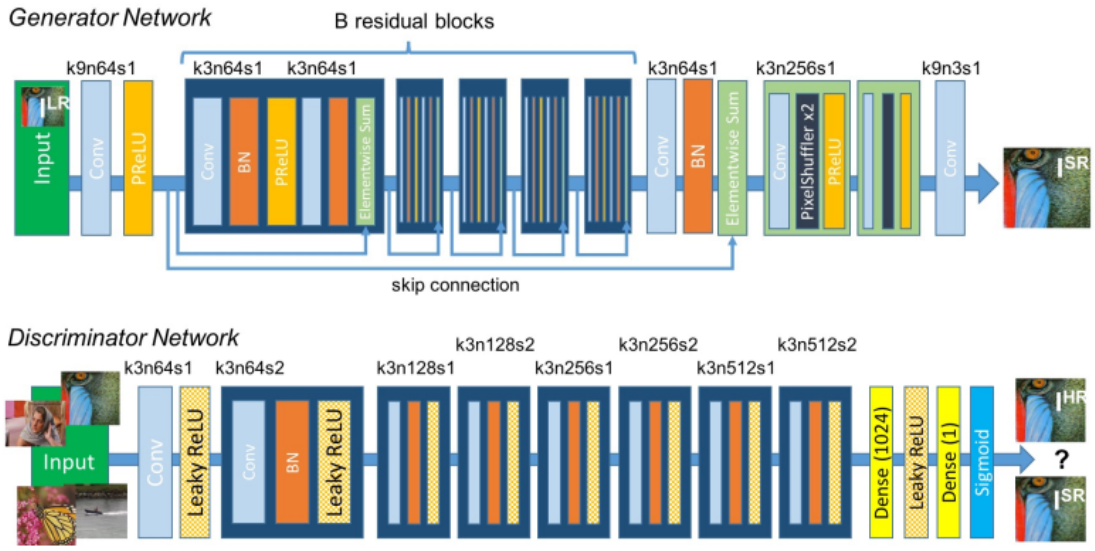

Figure 4: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

在生成网络上，作者使用了B个同样结构布局的残差块，然后使用了64个3*3的卷积核做卷积，再通过一个batch normalization层，用Parametric ReLU作为激活函数。在此之后，作者又加了两个同样的之前训练好的sub-pixel卷积层。

在辨别网络上，作者使用了α=0.2的Leaky ReLU，并且没有使用pooling。然后用了8个一样的3*3卷积核的VGG网络层，每一层使图像特征数量扩大一倍。最后使用了两个全连接层，一个最终的Sigmoid激活函数作为输出。

在完成了这些工作以后，作者还对评估函数所用的VGG网络的深度进行了探讨：

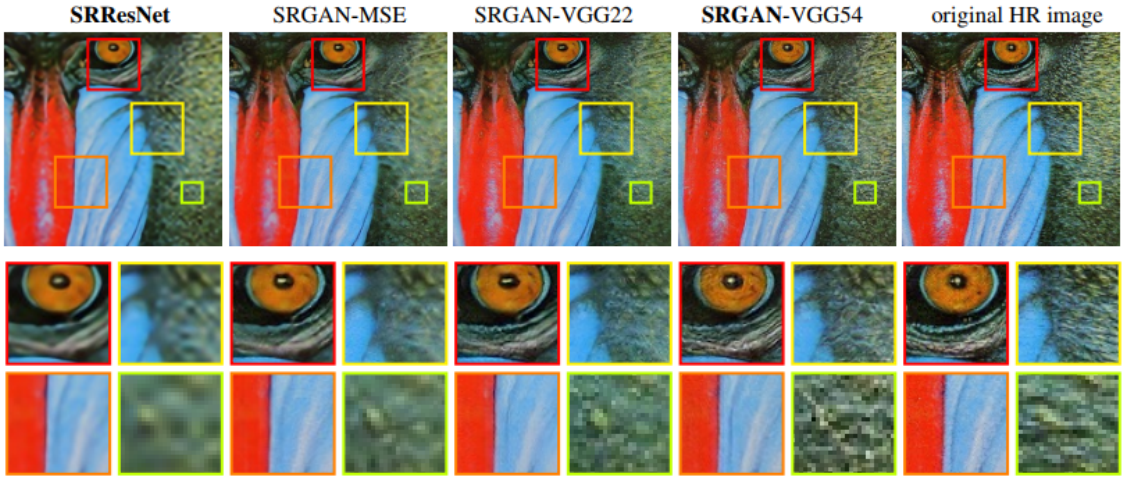

Figure 6: **SRResNet** (left: a,b), SRGAN-MSE (middle left: c,d), SRGAN-VGG2.2 (middle: e,f) and **SRGAN**-VGG54 (middle right: g,h) reconstruction results and corresponding reference HR image (right: i,j). [4× upscaling]

Table 2: Comparison of NN, bicubic, SRCNN [9], SelfExSR [31], DRCN [34], ESPCN [48], **SRResNet**, **SRGAN**-VGG54 and the original HR on benchmark data. Highest measures (PSNR [dB], SSIM, MOS) in bold. [4× upscaling]

作者认为，网络越深其超分辨重建的效果应该就越好。在他做的该组对比实验中，作者认为使用4个卷积层和5个pooling层得到的效果最佳。

当然了对于GAN里面的生成网络，作者认为使用ResNet的网络层数越深得到的效果也应该更好，但是这会带来更长的训练时间和更多的复杂的难以训练的高频参数。

另外，作者还对未来的工作进行了展望，他认为人的感官结果不能单纯的使用信噪比和单纯的数学公式来表示，他认为未来应该专注于在提出更好的感官评估函数。