SR的一些评估方案

PSNR

PSNR全称为"Peak Signal-to-Noise Ratio"，中文意思即为峰值信噪比。是用来衡量图像质量的指标之一。其是基于MSE（均方误差）定义的，对于一个给定大小的m*n的原始图像I和对其添加噪声后的图像K，其MSE可以定义为：

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

MSE就是原图与对比图每个像素的差值的平方的平均数。

而PSNR可以定义为：

$$PSNR = 10 \cdot log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

其中 MAX_I 为图像的最大像素值（即如果是8位二进制像素，那么其值为255），PSNR的单位为dB。对于多图像通道的图片，会分别对三个通道求MSE再进行平均值计算，进而求PSNR。

很显然的，PSNR是衡量目标图片与原图的相似性的，一般来说：

- 高于40dB：说明图像质量极好
- 30—40dB：表明图像质量可以接受
- 20—30dB：表明图像质量差
- 低于20dB：图像质量不可接受

SSIM

SSIM（Structural Similarity），结构相似性，是一种衡量两幅图像相似度的指标。给定两张图像x与y，其结构相似性可以按照如下算法求出：

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

其中 μ 代表均值， σ 代表方差， σ_{xy} 是x和y的协方差。 $c_1 = (k_1L)^2, c_2 = (k_2L)^2$ 是用来维持稳定的常数，L是像素值的动态范围（如最常见的0-255，L=256）， $k_1 = 0.01, k_2 = 0.03$

结构相似性的范围为-1到1，当两张图完全一致时，SSIM的值为1

SSIM的计算公式似乎是由三个方面的相似度所相乘得到的：照明度(l)，对比度(c)，结构(s)（但似乎得不到上面的式子）

$$l(x,y)=\frac{2\mu_x\mu_y+c_1}{\mu_x^2+\mu_y^2+c_1}$$

$$c(x,y)=\frac{2\sigma_x\sigma_y+c_2}{\sigma_x^2+\sigma_y^2+c_2}$$

$$s(x,y)=\frac{\sigma_{xy}+c_3}{\sigma_x\sigma_y+c_3}$$

LPIPS

LPIPS，全称为Learned Perceptual Image Patch Similarity（感知图像相似度），是一种用于测量两幅图像之间的感知相似性的指标，与传统的PSNR和SSIM不同，其是通过深度学习方法得到的一个网络，用于模拟基于人类视觉感知的图像相似性。

torchvision.models中就有可以直接使用的模型。

FID

FID（Frechet Inception Distance）是一种用于评估生成模型性能的指标，特别是在生成对抗网络中（GANs）广泛使用。其用于测量生成图像和真实图像分布之间的差异。

FID的计算是基于两个图像分布之间的特征向量空间的Frenchet距离。也是需要经过Inception网络上进行前向传播来进行特征向量的提取。

CLIP-IQA

主要使用CLIP模型来对图像的质量进行评估，主要思想是构造了一对“good and bad”prompt，并计算图像和文本的相似度，主要用于给出和人类感知相近的分数。

输入应该是【正面描述】+【负面描述】+【需要评估的图片】，最后输出这张图片与描述的相似程度。

MUSIQ

MUSIQ是一种多尺度的图像评估的transformer架构，解决了CNN架构需要图像进行特定大小的裁剪，可以处理任意分辨率的图像。最后通过tranformer输出的结果作为图像质量的得分。