

矩阵在图像压缩与评价中的应用与分析

SC18023140 赵杭天

一、图像质量评估方法 (Image Quality Assessment, IQA)

进行图像压缩之前，需要对图像压缩建立一套评价标准。而图像质量是非常主观的，所以，需要有一套方法，以人类的主观感受出发，科学合理的对图像进行评估。因此，本文从图像质量评估方法开始讨论。

根据对原图的参考程度，图像质量评估可分为全参考 (Full-Reference,FR)，部分参考 (Reduced-Reference,RR) 和无参考 (No-Reference,NR) 三种类型。考虑到普适性，本报告使用全参考与无参考中几种经典的评估方法。

1. 全参考 (Full-Reference, FR)

1.1 平均平方误差(Mean Squared Error, MSE)

$$D = \frac{1}{W * H} \sum_{x=1}^W \sum_{y=1}^H |S(x,y) - T(x,y)|$$

$S(x,y)$ 表示原始图像在坐标 (x,y) 的像素值， $T(x,y)$ 为待评价图像在坐标 (x,y) 的像素值。 W 代表图像宽度， H 代表图像高度。 D 代表待评估图像与原始图像质量区别 (或称失真度)，越小越好。

1.2 多层次结构相似性(Multi-Scale-Structural Similarity Index, MS-SSIM)

均值:

$$u_x = \frac{1}{R \times C} \sum_{i=1}^R \sum_{j=1}^C X(i,j)$$
$$u_y = \frac{1}{R \times C} \sum_{i=1}^R \sum_{j=1}^C Y(i,j)$$

其中 R 、 C 规定了每个局部区域的行列坐标范围，即 Rows、Columns;

方差:

$$\sigma_x^2 = \frac{1}{R \times C - 1} \sum_{i=1}^R \sum_{j=1}^C (X(i,j) - u_x)^2$$
$$\sigma_y^2 = \frac{1}{R \times C - 1} \sum_{i=1}^R \sum_{j=1}^C (Y(i,j) - u_y)^2$$

协方差:

$$\sigma_{XY} = \frac{1}{R \times C - 1} \sum_{i=1}^R \sum_{j=1}^C (X(i,j) - u_X)(Y(i,j) - u_Y)$$

中间方程组:

$$L(X, Y) = \frac{2u_X u_Y + C_1}{u_X^2 + u_Y^2 + C_1}$$

$$C(X, Y) = \frac{2\sigma_X \sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2}$$

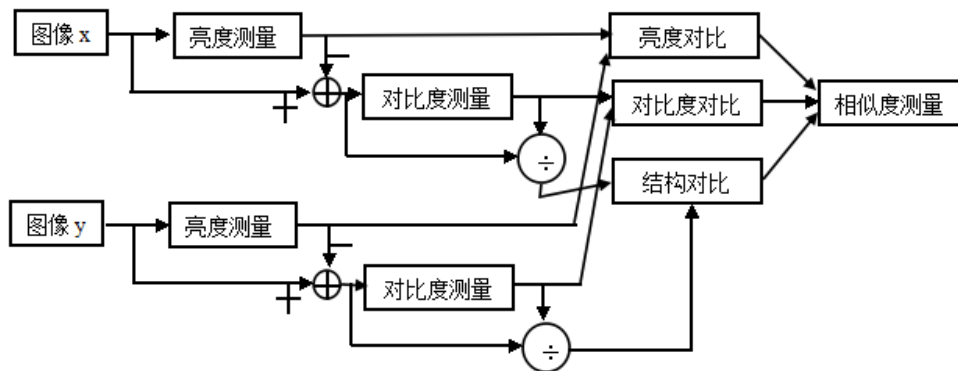
$$S(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X \sigma_Y + C_3}$$

其中 L(X,Y)是亮度对比因子, C(X,Y)是对比度因子, S(X,Y)是结构对比因子。

宽高以 2M-1 为因子进行缩小。当 M=1 时, 表示原始图像大小; 当 M=2 时, 表示原始图像缩小一半, 以此类推。

$$SSIM(X, Y) = [L_M(X, Y)]^{\alpha_M} \prod_{j=1}^M [C_j(x, y)]^{\beta_j} [S_j(x, y)]^{\gamma_j}$$

在文献[4]中, 实验得出 $\beta_1=\gamma_1=0.0448$, $\beta_2=\gamma_2=0.2856$, $\beta_3=\gamma_3=0.3001$, $\beta_4=\gamma_4=0.2363$, $\alpha_5=\beta_5=\gamma_5=0.1333$ 。这个评估算法相比之前的, 更贴近主观质量评价方法的结果。



SSIM 测量系统

1.3 梯度幅度相似性偏差(Gradient Magnitude Similarity Deviation, GMSD)

使用 Prewitt (或 Sobel、Scharr) 算子计算图像梯度:

$$h_x = \begin{bmatrix} \frac{1}{3} & 0 & -\frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \end{bmatrix}, \quad h_y = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

计算参考图像与目标图像的梯度幅值 m_r 和 m_d :

$$m_r(i) = \sqrt{(r \otimes h_x)^2(i) + (r \otimes h_y)^2(i)}$$

$$m_d(i) = \sqrt{(d \otimes h_x)^2(i) + (d \otimes h_y)^2(i)}$$

其中， \otimes 是卷积运算， r 是参考图像， d 目标图像。

梯度幅值相似性（GMS）如下：

$$GMS(i) = \frac{2m_r(i)m_d(i) + c}{m_r^2(i) + m_d^2(i) + c}$$

c 是一个常数，作用是为了防止分母为0。

标准差池化：

$$GMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (GMS(i) - GMSM)^2}$$

2. 无参考（No-Reference, NR）

2.1 基于自然场景统计的无参考图像质量评价方法（Blind/Referenceless Image

Spatial Quality Evaluator, BRISQUE)

局部归一化：

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C}$$

其中 μ 是局部均值， σ 是局部方差。

利用广义高斯分布拟合 MSCN 获得特征：

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right)$$

$$\beta = \alpha \sqrt{\frac{\Gamma(\frac{1}{\alpha})}{\Gamma(\frac{1}{3\alpha})}}$$

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt \quad a > 0$$

利用非对称广义高斯分布拟合 MSCN 相邻系数内积：

$$H(i, j) = \hat{I}(i, j) * \hat{I}(i, j + 1)$$

$$V(i, j) = \hat{I}(i, j) * \hat{I}(i + 1, j)$$

$$D1(i, j) = \hat{I}(i, j) * \hat{I}(i + 1, j + 1)$$

$$D2(i, j) = \hat{I}(i, j) * \hat{I}(i + 1, j - 1)$$

零均值非对称广义高斯分布(AGGD)的定义如下：

$$f(x; v, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{v}{(\beta_l + \beta_r)\Gamma(\frac{1}{v})} \exp\left(-\left(\frac{-x}{\beta_l}\right)^v\right) & x < 0 \\ \frac{v}{(\beta_l + \beta_r)\Gamma(\frac{1}{v})} \exp\left(-\left(\frac{-x}{\beta_r}\right)^v\right) & x \geq 0 \end{cases}$$

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma(\frac{1}{p})}{\Gamma(\frac{3}{p})}}$$

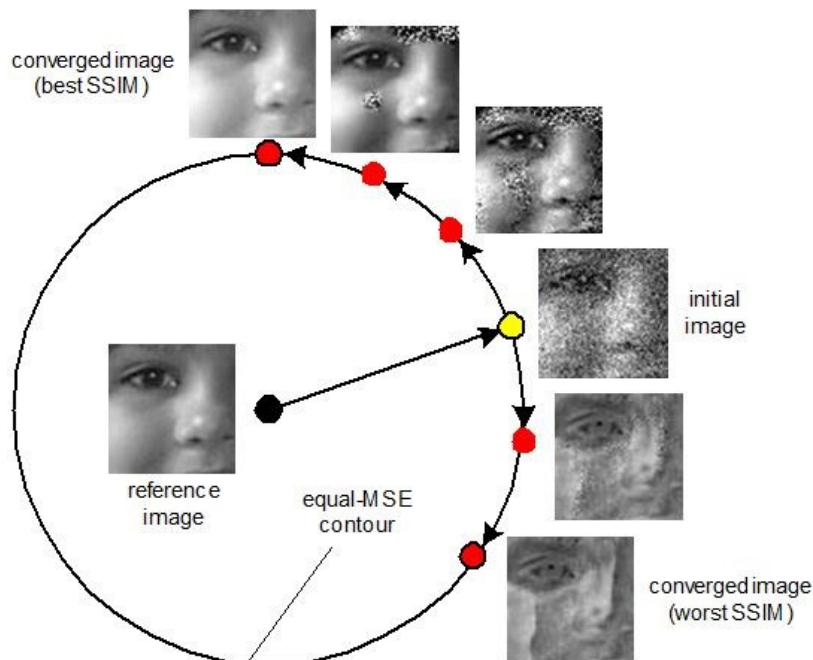
$$\beta_r = \sigma_r \sqrt{\frac{\Gamma(\frac{1}{p})}{\Gamma(\frac{3}{p})}}$$

$$\eta = (\beta_r - \beta_l) \frac{\Gamma(\frac{2}{p})}{\Gamma(\frac{1}{p})}$$

3. 讨论

1. Mean Square Error(MSE) / Peak Signal to Noise Ratio(PSNR)是图像/视频处理领域目前应用最广的性能量化指标, 简单易用速度快, 在目前应用最广泛的视频编码标准 H.264/AVC 和最新的 H.265/HEVC 中, PSNR 依然是最主要的客观评价方法。

但是, MSE/PSNR 与主观感知的一致性很低, 如下图中的情况, 圆内是原始图像, 圆外 6 幅加入不同失真, 明显主观质量相差很大, 但它们的 MSE 完全一样, 即 PSNR 也完全一样。失真图像按照 SSIM 排列, SSIM 的表现要好得多。



2. SSIM (Structural SIMilarity) 认为人类视觉感知能高度自适应提取场景中的结构信息, 主要有三大创新: 1) 虽然都是以逐点方式对比, PSNR 只对比当前点的像素值, 而 SSIM 以每个像素点为中心的小块(11 × 1)为单位, 对比小图像块反映的结构信息; 2) PSNR 只能对比像素值, 而 SSIM 对比图像块的三种统计特征, 即亮度(均值)、对比度(方差)和结构(协方差)信息; 3) 对比方式的差异, PSNR 采用差的平方, 而 SSIM 则为如下形式: $q = \frac{2ab+C}{a^2+b^2+C}$

3. 2011 年的 FSIM (Feature SIMilarity) 强调人类视觉系统理解图像主要根据图像低级特征, 以图像特征代替 SSIM 中的统计特征, 选择了相位一致性和梯度(3 × 3)幅值两种特征建立

FSIM, 又加入颜色特征建立 FSIMc, 并用相位一致性信息做加权平均, 评价准确性尤其在 TID2008 库上有大幅提升, 但速度比较慢。

12 年 TIP 的 GSM(Gradient SiMilarity), **强调梯度能传达重要的视觉信息**, 对场景理解至关重要, 梯度(5 × 5)特征和像素值结合就能达到不错的效果, 虽然性能比 FSIM 差一点, 但算法计算速度快很多。

14 年 TIP 的 GMSD (Gradient Magnitude Similarity Deviation) **强调图像梯度对图像失真敏感, 失真图像中不同局部结构的质量下降程度不同**, 可以只用梯度作为特征, pooling 采用标准差代替以前的均值。GMSD 比强调图像低级特征的 FSIM、强调图像梯度信息的 GSM 相比, 具有更好的时间效率并达到更大的人类主观一致性。至今仍在多个 IQA 数据库达到 state-of-the-art 性能。

5. 然场景统计(NSS)经常被应用于图像的质量评价。人们发现, 用高质量设备采集的自然图像(自然场景)有着一定的统计特征(如服从一些类高斯分布), 而图像的失真会使这些统计特征发生改变, 因此利用图像的一些统计特征作为图像特征, 可以完成图像的无参考评价。12 年的 BRISQUE 是一个经典的利用 NSS 进行 NR-IQA 的模型。观察到自然图像局部归一化亮度系数(MSCN), 强烈趋向于单位正态高斯特性, 因此作者假设失真会改变 MSCN 的分布状况, 并基于此提取特征。

无参考评估常用在缺失原图的图像压缩中, 例如图传系统的接收端进行图像质量监控。

二、图像压缩的矩阵方法

1. K-means

K-Means 算法的思想很简单, 对于给定的样本集, 按照样本之间的距离大小, 将样本集划分为 K 个簇。让簇内的点尽量紧密的连在一起, 而让簇间的距离尽量的大。

如果用数据表达式表示, 假设簇划分为(C_1, C_2, \dots, C_k), 则我们的目标是最小化平方误差 E:

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu^i\|_2^2$$

其中 μ_i 是簇 C_i 的均值向量, 有时也称为质心, 表达式为:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$$

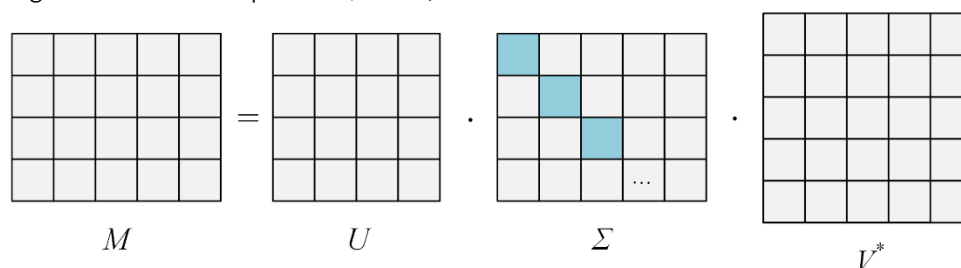
对于计算机来说, 每种颜色都会有一个对应 RGB 值, 比如黑色是[0,0,0], 白色是 [255,255,255], 所以 RGB 模式下, 最多可以区分 16581375(255 的三次方)种颜色。

另外我们知道, 一张图片的大小与分辨率正相关, 但其实也与图片颜色的复杂度是正相关的, 相同分辨率的情况下, 一张纯色图片是比一张五彩斑斓的图片要小的。

例如, 一张分辨率为 640*480 的图片, 其实就是由 307200 个 RGB 值组成。所以我们要做的就是对于这 307200 个 RGB 值聚类成 K 个簇, 然后使用每个簇内的质心点来替换簇内所有的 RGB 值, 这样在不改变分辨率的情况下使用的颜色减少了, 图片大小也就会减小了。

2. SVD

对于图像，可以描述为一个矩阵（张量） M ，由矩阵分析知识可知，可以对 M 进行奇异值分解（singular value decomposition, SVD）：



用数学公式表示：

$$M = [u_1, u_2, \dots, u_m] \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}$$
$$D = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r \end{bmatrix}$$

将矩阵 M 展开，可得：

$$M = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T$$

在上述奇异值分解 $M = U\Sigma V^*$ 中，

V 的列（rows）组成一套对 M 的正交“输入”或“分析”的基向量。这些向量是 M^*M 的特征向量。

U 的列（rows）组成一套对 M 的正交“输出”的基向量。这些向量是 MM^* 的特征向量。

Σ 对角线上的元素是奇异值，可视为是在输入与输出间进行的标量的“膨胀控制”。这些是 MM^* 及 M^*M 的特征值的非负平方根，并与 U 和 V 的行向量相对应。

如果我们将展开的矩阵 M 看作为一个图像的矩阵，则展开式中的每一个分量按 σ_i 的大小排序， σ_i 越大，说明越重要（K-L 变换证明了，在只能取有限个分量时，这样可保证均方误差最小）。而后面的权重很小，可以舍去。如果只取前面 k 项，则数据量为 $(m+n+1)k \ll mn$ 。这些保留的数据代表了原始图像中最主要、最具代表性的信息。这也就是图像压缩的基本原理。

三、实验

实验图片为一张色彩丰富的人像摄像图：



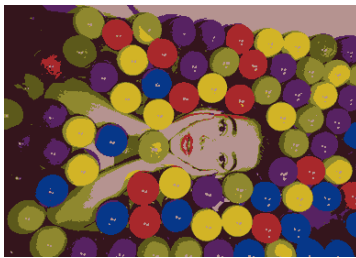
尺寸为 413x275 像素，PNG 格式，24 位深度；为合理度量压缩比（体积比），压缩存储格式也设为 PNG。

$$\text{体积比} = \frac{\text{压缩图像比特大小}}{\text{原始图像比特大小}} = 1 - \text{压缩比}$$

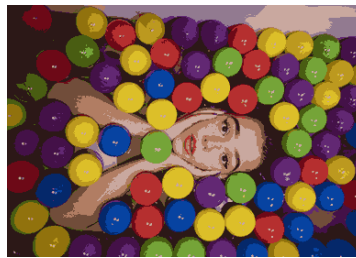
实验基于 Python3 语言完成，过程中使用到 PIL、OpenCV 图像开源库，numpy 数学函数库，sklearn 机器学习库，numba 加速运算库以及一些 Python 标准库。

1. K-means

使用不同的簇（clusters），对图像进行压缩：



n=8



n=16



n=24



n=32



n=64

表 1 不同簇的 K-means 压缩结果

	n=8	n=16	n=24	n=32	n=64
MSE	481.5031	235.2086	151.3340	113.0667	60.5863
MS-SSIM	0.8390	0.9095	0.9376	0.9538	0.9760
GMSD	0.0392	0.0281	0.0179	0.0129	0.0034
BRISQUE	93.2349	84.6054	61.6590	50.0125	27.5767
体积比	0.1227	0.1856	0.2551	0.3000	0.4280

2. SVD

使用不同的 k (主成分占比), 对图像进行压缩:

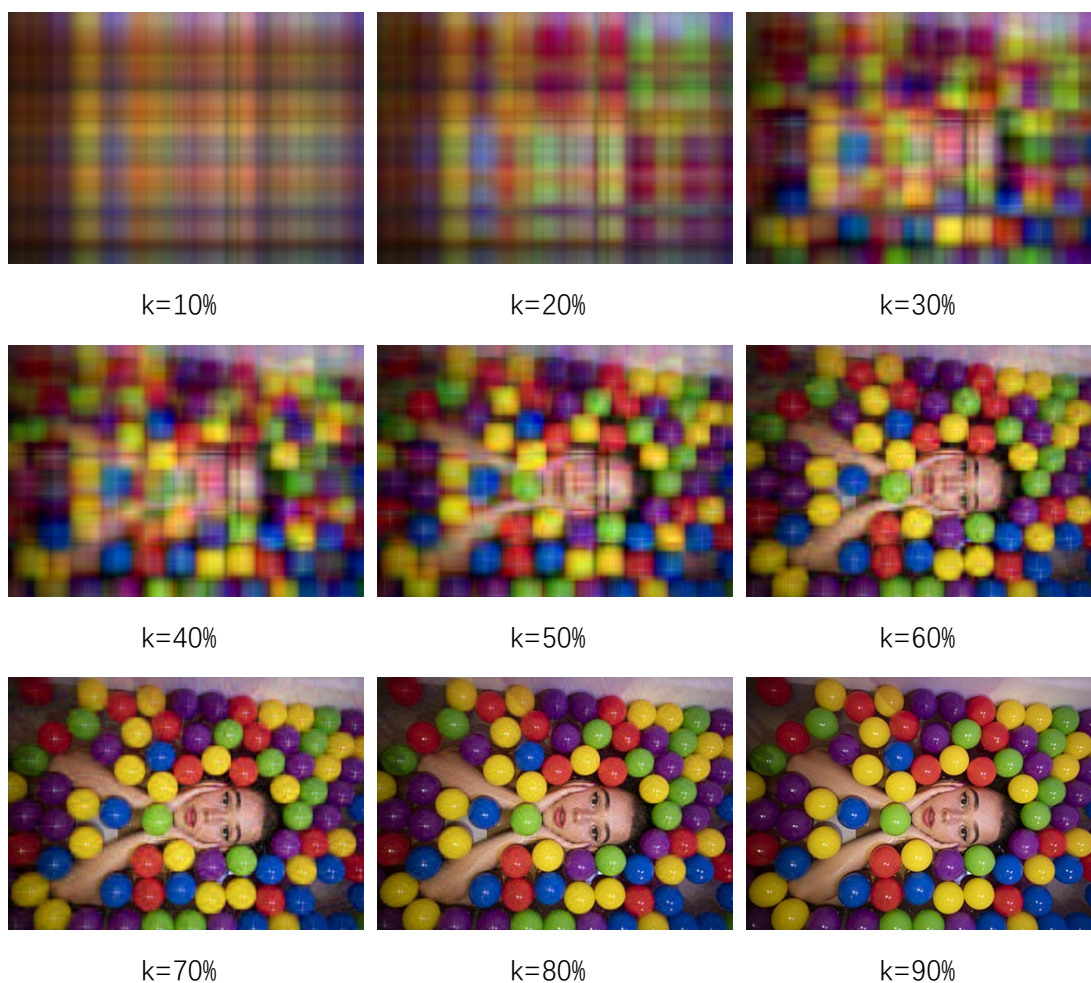


表 2 不同 k 的 SVD 压缩结果

	$k=0.1$	$k=0.2$	$k=0.3$	$k=0.4$	$k=0.5$
MSE	2626.5766	2315.0750	1465.4628	951.7588	515.1741
MS-SSIM	0.3301	0.4192	0.5634	0.6974	0.8360
GMSD	0.0758	0.0607	0.0446	0.0350	0.0241
BRISQUE	50.9457	48.8872	44.8513	42.9839	40.9860
体积比	0.5376	0.5806	0.6667	0.7258	0.7957
	$k=0.6$	$k=0.7$	$k=0.8$	$k=0.9$	
MSE	270.5524	136.7818	61.1140	17.1313	
MS-SSIM	0.9106	0.9570	0.9815	0.9947	
GMSD	0.0137	0.0050	0.0007	3.2296	
BRISQUE	35.2729	26.6558	13.8320	4.1117	
体积比	0.8763	0.9677	1.0483	1.0860	

四、分析

1. K-means 更容易达到更高的压缩比;
2. SVD 在主成分占比超过 0.8 时, 出现了体积大于原图的现象, 这主要是因为进行 PNG 保存时 (无损压缩) 出现了信息冗余。
3. K-means 的压缩损失主要体现在色彩丢失, 对图像纹理、轮廓几乎没有影响;
4. SVD 的压缩损失主要体系在边缘的变形、模糊方面, 色彩丢失不明显;
5. K-means 与 SVD 两者在达到 50%左右压缩比时 (K-means $n=64$; SVD $k=0.1$), 在目前最符合人眼主观判决的 GMSD 图像质量评价指标下, K-means 达到最好成绩。同时, K-means 对像素点 L2 距离的修改也更小 (MSE 更小)。
6. 从 K-L 变换推导到 PCA 和 SVD, 可以知道, PCA/SVD 取从大到小排序的前 σ_i 分量对应的特征向量作为主要信息, 其原理是在 K-L 变换中取 MSE 作为损失函数, 并使用 Lagrangian 乘子法构造代价函数, 并令其倒数为 0 得到的结果。而应用在图像压缩领域, “损失函数”可以是比较主观的东西, 至少 MSE 与人眼主观一致性比较差, 针对 IQA 开发出了如 GMSD 等评估标准, 用这些较好的 IQA 方法作为损失函数, 寻找极值, 可找到更合理的压缩方法。

【参考文献】

- [1] Image quality. Wikipedia. https://en.wikipedia.org/wiki/Image_quality
- [2] Sheikh, H.R.; Bovik A.C., Information Theoretic Approaches to Image Quality Assessment. In: Bovik, A.C. Handbook of Image and Video Processing. Elsevier, 2005.
- [3] 全参考图像质量评价方法整理与实用性探讨, 知乎, <https://zhuanlan.zhihu.com/p/24804170>
- [4] Zhou Wang, Eero P. Simoncelli and Alan C. Bovik “MULTI-SCALE STRUCTURAL SIMILARITY FOR IMAGE QUALITY ASSESSMENT”
- [5] 张贤达. 矩阵分析与应用 第二版[C]. 北京: 清华大学出版社, 2013