

图像质量评价概览

Dezeming Family

2022 年 3 月 23 日

目录

一 IQA 基本介绍	1
二 FR-IQA	1
2.1 MSE	1
2.2 SSIM	1
2.3 MS-SSIM	2
2.4 PSNR	2
2.5 LPIPS	2
2.6 VSI	2
2.7 FSIM	3
2.8 GMSD	3
2.9 IFC	3
三 RR-IQA	3
四 NR-IQA	3
4.1 GBVS	3
参考文献	3

一 IQA 基本介绍

图像质量评价 (Image quality assessment, IQA) 分为三种模式: 全参考 (full reference, FR)、部分参考 (reduced-reference, RR), 无参考 (no-reference, NR)。

实现一种新的图像质量评价体系, 需要在公开的 IQA 数据集上做大量的实验 [1] (比如 TID2008 数据集、TID2013 数据集)。IQA 包含了人类观察员对图像质量的主观打分, 需要计算失真图像与主观评分之间的相关性。当我们希望研究一种 IQA 算法时, 需要通过该算法计算出图像的 IQA 值以后, 评估算法的准确性, 评估方法有多种。

二 FR-IQA

- mean-squared error (MSE) 即均方误差。
- Structural similarity index metrics (SSIM [2]) 2004 年。
- Multiscale SSIM (MS-SSIM [4]) 2003 年。
- Peak signal to noise ratio (PSNR) 即峰值信号比。
- Learned perceptual image patch similarities (LPIPS [6]) 2018 年。
- Visual saliency-induced index (VSI [10]) 2014 年。
- Feature similarity index metrics (FSIM [8]) 2011 年。
- Gradient magnitude similarity deviation (GMSD [9]) 2014 年。
- Information fidelity criterion (IFC [5]) 2006 年。

2.1 MSE

MSE(Mean-Squared Error) 顾名思义, 就是求误差的平方, 然后再取均值。该方法又叫 MSD(Mean Squared Difference), 这是有效的图像误差计算方式。该方法如果不取平均, 就叫 SSD(Sum of Squared Difference)。

然后再介绍几种类似的方法。MAD(Mean Absolute Difference), 平均绝对差值 (即计算插值的绝对值, 再对每个像素求和取平均)。该方法如果不取平均, 只是求和, 就叫 SAD(Sum of Absolute Difference)。

2.2 SSIM

SSIM 假设人们观察图像时会提取其中的结构化信息。SSIM 评价体系下, 不但包含了传统的亮度和对比度的信息, 还包括了结构相似性信息。

使用均值来估计亮度 l 相似度 (因为均值可以反映像素值大小, 可以近似为亮度)。使用标准差来估计对比度 c 相似度 (因为标准差可以去除亮度的影响)。使用协方差来估计结构 s 相似度 (结构相似性应该与两幅图像的归一化值有关 $(\frac{A-\mu_A}{\sigma_A}, \frac{B-\mu_B}{\sigma_B})$, 其中, A, B 表示图像 A, B)。均值和方差的计算如下 (N 表示图像像素数):

$$\mu_A = \frac{1}{N} \sum_{i=1}^N A_i$$
$$\sigma_A = \left(\frac{1}{N-1} \sum_{i=1}^N (A_i - \mu_A)^2 \right)^{1/2}$$

之后, 用这三个相似度的指数乘积来定义总的相似度 [2]。

论文 [2] 中说明了 SSIM 应该作用于局部, 而不是全局, 因为全局的特征分布是非常不均匀的, 很难用简单的系数去评估全局的相似度。因此, 将整个图像分为一个一个的块来分别求 SSIM, 然后将所有的

块去平均。但是这样有一个问题，就是出现“块效应”（就是说，以前每个块的分析（采用 8X8 的窗口）都是单独进行的，块与块之间没有过渡，导致计算结果不理想）。因此采用一个 11X11 的高斯窗口（高斯权重设为 w_i ， $\sum w_i = 1$ ）来加权（ N 表示一个块内的像素数）：

$$\begin{aligned}\mu_A &= \sum_{i=1}^N w_i A_i \\ \sigma_A &= \left(\sum_{i=1}^N w_i (A_i - \mu_A)^2 \right)^{1/2} \\ \sigma_{AB} &= \sum_{i=1}^N w_i (A_i - \mu_A)(B_i - \mu_B)\end{aligned}\tag{二.1}$$

这里只介绍了 [2] 的具体思想和作用，关于 SSIM 的具体定义可在 [2] 中找到。

2 3 MS-SSIM

简单来说，就是通过不同的尺度来进行 SSIM 评价（或者说加上了尺度信息）。实际操作中采样高斯卷积的方式。这种评价方法比 SSIM 会更接近人类观察员的主观打分。

2 4 PSNR

峰值信噪比。其实是 MSE 的一种变体：

$$PSNR = 10 \times \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right)\tag{二.2}$$

PSNR 属于客观误差测量方法，PSNR 值越高则图像质量越高。但是正如均方误差，可能较高的 PSNR 视觉效果并不如较低的 PSNR，因为人眼对误差的敏感度并不是一个绝对值，与颜色的亮度、饱和度等方面都有关系。

2 5 LPIPS

对于人类来说，快速评估两个图像之间的感知相似度 (Perceptual Similarity) 是很容易的，但这个过程其实非常复杂。

LPIPS 是一个在视觉相似性判断上更接近人类感知的算法，本质上是深度学习方法，将深度特征图 (feature map) 用于衡量图像相似性的工具，由于深度特征图来源于卷积网络，所以更接近人的视觉判断。该网络不一定要在分类任务上进行训练、一些无监督模型也是可以的。根据论文 [6] 中描述，深度特征图的特征效果比低层次特征 (low-level metric, 例如方差、均值) 更好。

源码可见 [7]，该源码中包含了权重，可以直接进行测试。

2 6 VSI

VIS[10] 首先根据 GBVS[11]（一种基于图的视觉显著性特征检测提取的方法）计算特征图 VS(Visual Saliency)。

为了衡量对比度失真 (Contrast Change)，计算原图的图像梯度模（比如利用 Sobel 算子、Scharr 算子等，计算横向纵向的梯度，然后求模），得到 GM (gradient modulus)。

为了衡量色彩饱和度失真 (Change of Color Saturation)，首先将 RGB 图像进行空间变换，变换为 LMN 通道，之后只使用 L 通道计算梯度，得到色度特征。

之后，将三个特征进行组合，然后来计算两幅图像之间的 VSI。

2.7 FSIM

FSIM[8] 是 SSIM 的变种，本质上也是在考虑使用什么特征。

心理学和脑科学的研究发现不同频率的傅里叶波，当具有相同的相位时，一般就是视觉上可辨认的重要特征——这就意味着从某些一致性的相位上能够提取出特征信息来。

相位一致性 (phase congruency) 的提取方法可见 [12]。

同时还考虑了梯度幅值，采用 Schar 算子提取梯度 GM (gradient modulus) 即可。

之后，FSIM 会将两种特征进行组合，得到计算值。

2.8 GMSD

GMSD[9] 采用梯度信息进行相似度度量。该论文作者认为梯度是计算效率高且具有足够可信度的特征。

使用 prewitt 算子获取梯度模，然后计算出该像素的 GMS，之后可以每个像素加起来取平均，得到整张图的梯度模，叫做 GSM (即 GSM-mean)。

但是整张图的梯度模不足以直接作为评判依据，例如对于模糊而言，对边界的影响远大于对平坦地区的影响，因此，使用梯度模的方差效果会更好。

2.9 IFC

IFC [5] 即信息保真度准则，衡量失真图像与原图像的共同信息量。

通过小波变换，将图像分解为多个系数，根据系数进行评判。信息保真度来自于信息论的方法。

三 RR-IQA

四 NR-IQA

- Graph-based visual saliency (GBVS [11]) 2007 年。

4.1 GBVS

参考文献

- [1] <https://zhuanlan.zhihu.com/p/210887209>
- [2] Wang Z , Bovik A C , Sheikh H R , et al. Image Quality Assessment: From Error Visibility to Structural Similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4):600-612.
- [3] <https://blog.csdn.net/wjpwjpwj0831/article/details/120401764>
- [4] Zhou W , Simoncelli E P , Bovik A C . Multiscale structural similarity for image quality assessment[C]// Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on. 2003.
- [5] Sheikh H R , Member, IEEE, et al. An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics[J]. IEEE Transactions on Image Processing, 2006, 14(12):2117-2128.
- [6] Zhang R , Isola P , Efros A A , et al. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric[J]. IEEE, 2018.
- [7] <https://github.com/richzhang/PerceptualSimilarity>

- [8] Zhang L , Zhang L , Mou X , et al. FSIM: A Feature Similarity Index for Image Quality Assessment[J]. IEEE Transactions on Image Processing, 2011, 20(8):2378-2386.
- [9] Xue W , Zhang L , Mou X , et al. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index[J]. IEEE Transactions on Image Processing, 2014, 23(2):684-695.
- [10] Zhang L , Shen Y , Li H . VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment[J]. IEEE Transactions on Image Processing, 2014, 23(10):4270-4281.
- [11] Harel J , Koch C , Perona P . Graph-Based Visual Saliency[J]. Proc of Neural Information Processing Systems, 2007, 19:545-552.
- [12] [1] Image Features from Phase Congruency[J]. Videre A Journal of Computer Vision Research, 1999.