

基于颜色名称的彩色图像质量评价

马畅, 张选德

引用本文:

马畅, 张选德. 基于颜色名称的彩色图像质量评价[J]. 液晶与显示, 2022, 37(1): 56–65.

MA Chang, ZHANG Xuan-de. Color image quality assessment based on colornames[J]. *Chinese Journal of Liquid Crystals and Displays*, 2022, 37(1): 56–65.

在线阅读 View online: <https://doi.org/10.37188/CJLCD.2021-0189>

您可能感兴趣的其他文章

Articles you may be interested in

一种低光照度下的彩色数码设备色彩还原能力评价方法

Evaluation method of color restoration ability of color digital equipment under low illumination

液晶与显示. 2020, 35(5): 477–485 <https://doi.org/10.3788/YJYXS20203505.0477>

结合图像签名和最优传输的颜色传递方法

Color transfer method for joint image signature and optimal transmission

液晶与显示. 2021, 36(8): 1151–1165 <https://doi.org/10.37188/CJLCD.2020-0300>

基于稀疏表示的SOM多失真图像质量评价方法

Multi-distorted image quality assessment algorithm based on sparse representation and SOM

液晶与显示. 2018, 33(10): 877–883 <https://doi.org/10.3788/YJYXS20183310.0877>

本征图像分解的稀疏表示彩色图像去噪算法

Color image denoising algorithm based on intrinsic image decomposition and sparse representation

液晶与显示. 2019, 34(11): 1104–1114 <https://doi.org/10.3788/YJYXS20193411.1104>

基于图像分割和区域匹配的灰度图像彩色化算法

Gray image colorization algorithm based on image segmentation and region matching

液晶与显示. 2019, 34(6): 619–626 <https://doi.org/10.3788/YJYXS20193406.0619>

文章编号:1007-2780(2022)01-0056-10

基于颜色名称的彩色图像质量评价

马 畅, 张选德*

(陕西科技大学 电子信息与人工智能学院, 陕西 西安 710021)

摘要: 图像质量评价(Image Quality Assessment, IQA)是计算机视觉领域研究的基本问题之一。目前,绝大多数图像质量模型都是基于灰度图像构建的,而彩色图像质量评价至今依然是 IQA 领域的开放性问题。彩色图像质量评价研究的关键在于建立与人类色彩认知能力相一致的彩色信息的量化描述。本文基于颜色名称(Colornames, CN)构建彩色图像质量评价模型,将图像的每个像素值映射为 CN 概率向量,利用 Wasserstein 距离计算两幅图像的感知色差,以亮度和梯度特征作为补充,在池化阶段采用显著性加权得到客观图像质量评分。在公开测试数据集上的实验结果表明,提出的模型在 TID2008、TID2013 和最新的 KADID-10k 数据集上表现最佳,其 SROCC 值分别为 0.900 9, 0.890 1, 0.863 7。总体评价效果与目前最好的传统方法(非深度学习方法)相当;而对于颜色失真,则具有明显的优势。

关键词: 彩色图像质量评价;颜色名称;Wasserstein 距离;人类视觉系统

中图分类号: TP391 **文献标识码:** A **doi:** 10.37188/CJLCD.2021-0189

Color image quality assessment based on colornames

MA Chang, ZHANG Xuan-de*

(College of Electronic Information and Artificial Intelligence,
Shaanxi University of Science and Technology, Xi'an 710021, China)

Abstract: Image quality assessment (IQA) is one of the basic research issues in the field of computer vision. At present, most image quality models are constructed based on grayscale images, and color image quality assessment is still an open issue in the field of IQA. The key of color image quality assessment research is to construct a quantitative description of color information consistent with human color cognition. This paper constructs a color image quality assessment model based on colornames (CN). It maps each pixel value of the image to a CN probability vector, uses the Wasserstein distance to calculate the perceived color difference of two images, uses the lightness and gradient features, and uses the saliency weighting in the pooling stage to obtain the objective image quality scores. The experimental results on the public test databases show that the proposed model performs best on TID2008, TID2013 and the latest KADID-10k databases, with SROCC values of 0.900 9, 0.890 1 and 0.863 7, respectively. The overall assessment effect is comparable to the current best traditional method (Non-deep learning method). But for color distortion, it has obvious advantages.

Key words: color image quality assessment; colornames; Wasserstein distance; human visual system

收稿日期:2021-07-16;修订日期:2021-08-11.

基金项目:国家自然科学基金(No.61871260)

Supported by National Natural Science Foundation of China(No.61871260)

* 通信联系人, E-mail: zhangxuande@sust.edu.cn

1 引言

IQA 是图像处理领域的基本问题之一,且在图像压缩、视频通讯、图像恢复等众多问题中有着十分重要的应用。例如:图像压缩中需要在图像质量和压缩率之间作均衡,视频通讯中需要对图像质量进行实时监控,而图像恢复系统中也需要一个合适的图像质量指标对其性能进行评价。IQA 的研究目标在于利用数学模型来模拟人类视觉系统(Human Vision System, HVS)对图像质量的感知和评价过程,构建同主观评价尽可能一致的客观图像质量指标^[1]。根据参考图像的可用性,图像质量评价可以分为3种类型:全参考(Full-Reference, FR)、部分参考(Reduced-Reference, RR)和无参考(No-Reference, NR)^[2]。本文针对全参考图像质量评价进行研究,构建彩色图像质量评价模型。

当前大多数图像质量评价算法都是针对灰度图像设计的,如 SSIM (Structure SIMilarity)^[3], GMSD (Gradient Magnitude Similarity Deviation)^[4]等,而彩色图像质量评价方面的工作相对较少。Li Leida 等人提出了基于稀疏表示和重构残差的彩色图像质量评价(Sparse Representation and Reconstruction Residual, SRRR)算法^[5],使用自然彩色图像训练的过完备颜色字典表示参考图像和失真图像,构造两个特征图度量图像的结构和颜色失真,计算重构残差度量图像的对比度变化,还引入亮度相似性以得到彩色图像的最终质量得分。Jens Preiss 等人^[6]提出了一种使用彩色图像差异(Color Image Difference, CID)作为目标函数来优化色域映射的算法,在解决传统色域映射算法生成图像中包含各种视觉伪像问题的同时给出了改进的彩色图像差异(Improved Color Image Difference, ICID)指标,提高了对彩色图像质量的预测性能。Dogancan Temel 和 Ghassan AlRegib 提出基于多尺度和多通道误差表示的频谱理解的图像质量评价(Spectral Understanding of Multi-scale and Multi-channel Error Representations, SUMMER)算法^[7],该方法关注误差图像(参考图像与失真图像之差)的幅度谱,解决了灰度图像的光谱统计量忽略了的颜色信息以及 HVS 的选择性和层次性的问题。Sun

Wen 等人提出基于超像素的图像质量评价(SuperPixel-Based SIMilarity, SPSIM)算法^[8],基于感知上有意义的超像素图像块计算亮度、色度和梯度相似度,根据梯度区域一致性来进一步调整这3个特征,最后用纹理复杂度作为池化阶段的加权函数,得到了与主观评分较高的一致性。

对于彩色图像,我们不能将其简单地转到灰度域进行评价,而是要引入颜色特征来度量图像的色彩变化,以提高算法对彩色图像质量的评价性能。一种直观的彩色图像质量评价方法是在颜色通道中计算逐像素的保真度,如 PerSIM (Perceptual SIMilarity)^[9]和 FSIMc (Feature SIMilarity extend to Color)^[10],但是各个颜色通道之间的差异未必对应于颜色之间的感知差异。从人类感知的角度来看,通常颜色空间不是均质(Homogeneous)的度量空间^[11],因此彩色图像质量评价研究中不应对各个颜色通道作分离处理,而应着眼于整体感知的颜色并计算颜色差异。从根本上讲,彩色图像质量评价的关键在于建立与 HVS 色彩感知能力相一致的彩色描述与量化方法。但色彩的描述与量化是计算机视觉领域至今尚未完全解决的问题,这使得彩色图像质量评价成为了 IQA 领域的开放性问题。

CN^[12]是近年来颇为知名的颜色描述方法,这种方法用11维概率向量来描述颜色,这个向量的每个分量表示了色彩属于11个可被准确感知亦可用语义描述的颜色概率。CN 的独到之处在于使用了颜色的语义描述,这些语义描述间接地反映了人类的色彩感知能力。本文利用 CN 构建彩色图像质量模型,该模型将参考图像和失真图像的每个像素值映射为 CN 概率向量,并利用 Wasserstein 距离计算两个向量分布之间的差异来度量两幅图像间的感知色差;然后将参考图像与失真图像转换到各通道相互独立的对抗颜色空间(Opponent Color Space)^[13],并在其亮度通道中计算能够表征图像结构信息的梯度特征。由于 HVS 对亮度变化感知比对颜色更敏感,且人眼对颜色的感知与亮度关系密切,所以我们在模型中加入亮度特征作为补充。在池化阶段使用视觉显著性作为加权函数来获得图像质量得分。在几个公开数据集上的实验结果表明,提出的模型能够获得很好的评价效果。

2 颜色名称

人们通常使用颜色名称(CN)轻而易举地描述人眼所看到的世界。而在计算机视觉中,我们学习图像像素值与 CN 之间的映射关系,以将语义颜色标签分配给图像像素来描述图像的颜色信息。这里使用的 11 种基本颜色名称包括黑色、蓝色、棕色、灰色、绿色、橙色、粉红色、紫色、红色、白色和黄色^[12]。

为了获得图像像素值与 CN 之间更准确的映射关系,我们使用概率潜在语义分析(Probabilistic Latent Semantic Analysis, PLSA)模型从 Google 图像上搜索大量真实世界的图像构成的数据集中学习颜色名称。PLSA 是由 Hofmann^[14]提出的一种用于文档分析的生成模型。给定一组文档 $D = \{d_1, \dots, d_N\}$, 每个文档都用单词表 $W = \{w_1, \dots, w_M\}$ 描述,这些单词是由潜在主题 $Z = \{z_1, \dots, z_K\}$ 产生的。在 PLSA 模型中,文档 d 中单词 w 的条件概率由式(1)计算:

$$p(w | d) = \sum_{z \in Z} p(w | z) p(z | d), \quad (1)$$

其中: $p(z | d)$ 和 $p(w | z)$ 都是离散多项式分布,可以使用 EM 算法^[14]通过最大化对数似然函数 L 进行估算:

$$L = \sum_{d \in D} \sum_{w \in W} n(d, w) \log p(d, w), \quad (2)$$

其中 $p(d, w) = p(d) p(w | d)$, $n(d, w)$ 是文档 d 中出现单词 w 的次数。

在学习颜色名称的问题中,图像对应文档,像素值对应单词,颜色名称对应潜在主题。我们将图像中的像素值建模为由颜色名称生成的颜色值,PLSA 模型的目的是找到最能解释所观察数据(图像像素)的潜在主题(颜色名称)。该过程可以理解将为 $p(w | d)$ 分解为单词-主题分布 $p(w | z)$ 和文档-主题固定比例 $p(z | d)$,则在 $p(w | z)$ 中可以得到主题(颜色名称)在单词(像素值)上的分布,如图 1 所示。

由于标准的 PLSA 模型无法利用图像标签信息,因此研究提出 PLSA 模型的改进版本^[12]用于学习颜色名称。改进的 PLSA 模型使用图像标签来定义图像中颜色名称出现频率的先验分布,与图像标签相对应的颜色名称被认为比其他

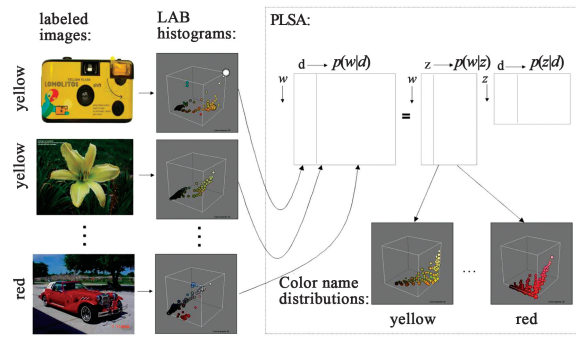


图 1 用于学习颜色名称的标准 PLSA 模型概述^[12]

Fig. 1 Overview of standard PLSA model for learning colornames

颜色具有更高频率的先验;同时向标准的 PLSA 模型中的对数似然函数添加一个正则化因子,以抑制 CN 概率分布中的次级模式,保证概率分布的单峰性。假设分布 $p(z | d)$ 是从参数为 α_{l_d} 的 Dirichlet 分布生成的,其中 l_d 是文档 d 的标签。向量 α_{l_d} 的长度为 K (主题数),其中对于 $z = l_d$ 有 $\alpha_{l_d}(z) = c \geq 1$, 否则 $\alpha_{l_d}(z) = 1$ 。通过改变 c 就可以控制图像标签 l_d 对分布 $p(z | d)$ 的影响。然后,将 $p(z | w)$ 分布与它的单峰形式 $\rho_z^{mz}(w)$ 之间的差作为正则化因子加入对数似然函数:

$$L = \sum_{d \in D} \sum_{w \in W} n(d, w) \log p(d, w) - \gamma \sum_{z \in Z} \sum_{w \in W} (p(z | w) - \rho_z^{mz}(w))^2. \quad (3)$$

在公式(2)中添加正则化因子会使得函数 $p(z | w)$ 接近 $\rho_z^{mz}(w)$, 由于 $\rho_z^{mz}(w)$ 是单峰的,因此这将抑制 $p(z | w)$ 中的次级模式。最后通过两个参数 c 和 γ 来调整两种改进的强度。

3 基于 CN 的彩色图像质量评价模型

3.1 感知色差

这里利用 Wasserstein 距离来计算参考图像与失真图像的 CN 概率向量之间的差异,度量两幅图像之间的感知色差。首先利用文献^[12]中提供的映射关系表将参考图像和对应的失真图像中的每个像素映射为 11 维 CN 概率向量,然后利用两个向量之间的距离来度量感知色差。如何度量两个概率向量(分布)之间的距离呢? 常用的方法有 KL 散度和 Wasserstein 距离,其中

Wasserstein 距离也被称为推土机距离 (Earth Mover Distance, EMD), 是将一个直方图转换为另一个直方图所必须付出的最小代价。Wasserstein 距离较之 KL 散度具有更好的数学性质, 即使两个分布的支撑集没有重叠或者重叠非常少, 仍然能反映两个分布的远近。因此选用 Wasserstein 距离来度量两个 CN 概率向量的差异:

$$\text{CND}_i = \text{WS}(Y(f_i^r), Y(f_i^d)), \quad (4)$$

其中, f^r 表示参考图像, f^d 表示失真图像, i 表示“图像块”索引, $Y()$ 表示图像像素值到 CN 概率向量的映射, $\text{WS}()$ 表示 Wasserstein 距离运算符, CND 表示颜色名称距离。图 2 是参考图像和失真图像以及它们的 CND 图的示例。



图 2 颜色名称距离图

Fig.2 Colornames distance map

Wasserstein 距离是两个概率分布或直方图之间在感知上有意义的统计指标^[15]。给定两个概率分布 P_1 和 P_2 , 则 P_1 和 P_2 之间的 Wasserstein 距离可定义为:

$$W(P_1, P_2) = \inf_{\gamma \in \Pi(P_1, P_2)} E_{(x,y) \sim \gamma} [\|x - y\|], \quad (5)$$

其中, $\inf$ 表示最大下界, $\Pi(P_1, P_2)$ 是 P_1 和 P_2 分布组合起来的所有可能的联合分布的集合, 对于每一个可能的联合分布 γ , 可以从中采样 $(x, y) \sim \gamma$ 得到一个样本 x 和 y , 并计算出这对样本的距离 $\|x - y\|$, 所以可以计算该联合分布 γ 下样本对距离的期望值 $E_{(x,y) \sim \gamma} [\|x - y\|]$ 。在所有可能的联合分布中能够对这个期望值取到的下界 $\inf_{\gamma \in \Pi(P_1, P_2)} E_{(x,y) \sim \gamma} [\|x - y\|]$ 就是 Wasserstein 距离。

3.2 梯度相似度

图像梯度是 IQA 问题中最常用的特征之一, 它对失真很敏感, 并且可以反映图像的对比度和结构信息。有多种不同的算子可用于计算图像梯

度, 这里使用 Scharr 算子^[16]。首先, 利用公式 (6) 将图像转换到对抗颜色空间 (Opponent Color Space) 中, 该空间亮度与颜色信息完全分开, 且各颜色通道相互独立^[13]。

$$\begin{bmatrix} L \\ M \\ N \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (6)$$

其中, L 表示亮度通道, M 和 N 表示颜色通道。上述转换中的权重针对 HVS 进行了优化^[17]。然后, 我们从 L 通道计算图像的水平和垂直梯度, 分别用 $G_x(i)$ 和 $G_y(i)$ 表示, 并计算其梯度幅度 $GM(i)$, 其中 i 代表第 i 个像素。

图像 $f(i)$ 的水平 and 垂直梯度 $G_x(i)$ 和 $G_y(i)$ 计算为:

$$G_x(i) = \frac{1}{16} \begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix} * f(i), \quad (7)$$

$$G_y(i) = \frac{1}{16} \begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & 3 \end{bmatrix} * f(i), \quad (8)$$

将 $f(i)$ 的梯度幅度定义为 $GM(i) = \sqrt{G_x^2(i) + G_y^2(i)}$ 。图 3 展示了参考图像和失真图像的梯度幅度 (Gradient Magnitude, GM) 图, 我们可以在它们的 GM 图之间观察到明显的差异, 这表明 GM 图在反映图像的局部结构信息以及对比度失真等方面具有良好的能力。

参考图像 f_r 和失真图像 f_d 之间逐像素的梯

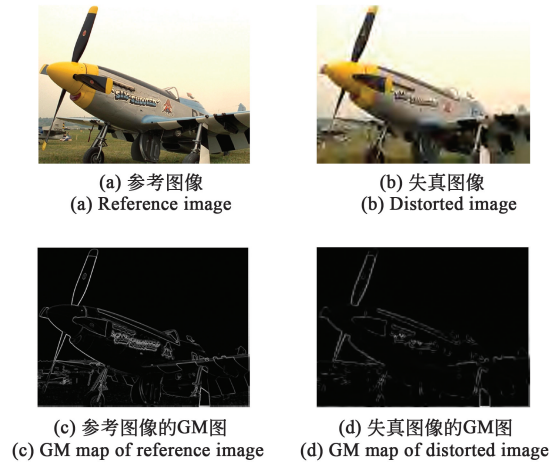


图 3 梯度幅度图

Fig.3 Gradient magnitude maps

度幅度的相似度定义为:

$$S_G(i) = \frac{2GM_r(i)GM_d(i) + GM_1}{GM_r^2(i) + GM_d^2(i) + C_1}, \quad (9)$$

其中 $GM_r(i)$ 和 $GM_d(i)$ 分别表示参考图像 f_r 和失真图像 f_d 中第 i 个像素的梯度幅度, C_1 是一个正常数, 以保持 $S_G(i)$ 的稳定性。

3.3 亮度相似度

与颜色信息相比, HVS 对图像的亮度变化更加敏感^[18]。亮度总是对图像的感知质量有很大影响, 因此在彩色图像质量评价的问题中, 仍然需要考虑图像的亮度特征。亮度相似度是基于每个图像块的平均值来计算的, 参考图像与失真图像的每一组图像块对的平均值构成一组平均值对。根据恰可察觉失真 (Just Noticeable Distortion, JND) 模型, 我们知道人类视觉系统无法察觉到处于一定阈值以下的图像内容的变化^[19]。因此, 我们不必使用所有平均值对来度量亮度失真, 因为

一些差异较小的平均值对不会影响人类对图像质量的感知, 甚至还会缓和人眼对亮度变化较大的图像区域质量的感知, 因此, 我们仅考虑亮度差异较大的平均值对, 然后计算两组选定平均值图像块对之间的相关性。

具体来说, 对于对抗颜色空间亮度通道中的每一个图像块, 首先计算它们的平均值, 并将其用于表示这个图像块的亮度级别。参考图像块 $\{f_i^r\}_{i=1}^K$ 和失真图像块 $\{f_i^d\}_{i=1}^K$ 之间的绝对亮度差用 $l = \{l_i\}_{i=1}^K$ 表示, 其中 K 是图像块的总数, $l_i = |\mu(f_i^r) - \mu(f_i^d)|$, $\mu()$ 是平均运算。然后确定亮度差足够大的图像块对:

$$(l^r, l^d) = \{(\mu(f_i^r), \mu(f_i^d))\}, \quad (10)$$

$$|\mu(f_i^r) - \mu(f_i^d)| \geq \text{median}(l),$$

其中 $\text{median}()$ 表示中位数计算。最后, 亮度相似度的得分计算如式 (11) 所示:

$$Q_L = \frac{\sum_{i=1}^T (l_i^r - \mu(l^r)) \cdot (l_i^d - \mu(l^d)) + C_2}{\sqrt{\sum_{i=1}^T (l_i^r - \mu(l^r))^2 \cdot \sum_{i=1}^T (l_i^d - \mu(l^d))^2 + C_2}}, \quad (11)$$

其中 T 表示亮度差足够大的图像块对的数量, $C_2 > 0$ 是避免除零的小常数, l_i^r 和 l_i^d 分别表示 l^r 和 l^d 的第 i 个元素。

3.4 基于 CN 的彩色图像质量评价模型

以基于 CN 定义的感知色差为基础, 以梯度幅值相似性和亮度相似性作为补充, 我们构建了一个全参考彩色图像质量评价模型, 简称为 CNCI (CN based Color image quality Index)。CNCI 的总体框架如图 4 所示。

参考图像 f_r 和失真图像 f_d 之间的相似度图 $S(i)$ 包括两个部分, 一个是颜色名称距离图 (CND), 另一个是梯度相似度图 (S_G)。我们用参考图像和失真图像之间的颜色名称距离图来表示两幅图像间的感知色差, 用梯度相似度图表示图像的结构变化, 然后将二者融合, 如公式 (12) 所示:

$$S(i) = (S_G(i))^{\alpha} \cdot (1 - \text{CND}(i))^{\beta}. \quad (12)$$

我们采用 SDSP 视觉显著模型^[20]来计算参考图像的视觉显著图 VS_r 和失真图像的视觉显著图 VS_d , 使用 $w(i) = \max(VS_r(i), VS_d(i))$ 来加权

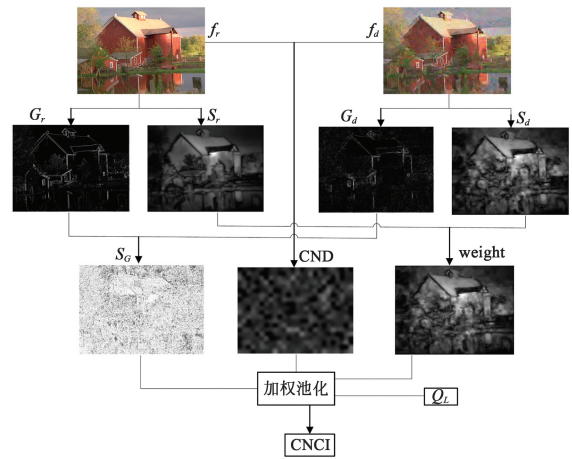


图 4 CNCI 模型总体框架图

Fig.4 CNCI model overall framework diagram

$S(i)$, 为图像的不同区域赋予不同的权重大小, 从而产生一个分数:

$$Q_S = \frac{\sum_{i=1}^W \sum_{j=1}^H S(i) \cdot w(i)}{\sum_{i=1}^W \sum_{j=1}^H w(i)}, \quad (13)$$

其中, W 和 H 表示相似度图的大小。考虑亮度对感知质量有很大影响, 因此我们将该模型的最

终质量得分 Q 定义为 Q_S 和 Q_L 的线性组合:

$$Q = a \cdot Q_S + b \cdot Q_L, \quad (14)$$

其中 a 和 b 是用于调整两个分量相对重要性的参数,满足 $a + b = 1$ 。

4 实验结果分析

4.1 数据集和评价指标

我们在 5 个数据集 TID2008^[21]、TID2013^[22]、CSIQ^[23]、LIVE^[24] 和 KADID-10k^[25] 上测试所提出的彩色图像质量评价模型的性能。这些数据集中包含参考图像、失真图像以及针对失真图像所收集的平均主观分数(Mean Opinion Scores, MOS)或差异平均主观分数(Differential Mean Opinion Scores, DMOS)。其中, TID2008 包含 1 700 张失真图像, 17 种失真类型和 4 个失真等级; TID2013 包含 3 000 张失真图像, 24 种失真类型和 5 个失真等级; CSIQ 包含 866 张失真图像, 5 种失真类型和 4~5 个失真等级; LIVE 包含 779 张失真图像, 5 种失真类型和 4~5 个失真等级; KADID-10k 包含 10 125 张失真图像, 25 种失真类型和 5 个失真等级。

图像质量评价模型的性能通常以主客观评分之间的一致性来衡量。常用的评价指标有 Spearman 秩相关系数(SROCC)、Kendall 秩相关系数(KROCC)、Pearson 线性相关系数(PLCC)和均方根误差(RMSE)。其中, SROCC 和 KROCC 用来衡量主客观评分之间的一致性, PLCC 和 RMSE 用来衡量模型预测的准确性。SROCC、KROCC、PLCC 越大或 RMSE 越小, 代表模型性能越好。在计算 PLCC 和 RMSE 前需要先进行回归分析, 建立主客观评分间的非线性映射。这里采用公式(15)中的 logistic 回归函数, 其中 Q 代表 IQA 方法计算得到的客观评分,

P 代表 Q 的回归值, $\beta_i \mid i=1, 2, 3, 4, 5$ 为要拟合的参数。

$$P = \beta_1 \left[\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (Q - \beta_3)}} \right] + \beta_4 Q + \beta_5. \quad (15)$$

4.2 实验参数

在本文方法中需要设置的参数有 C_1 和 C_2 、 α 和 β 、 a 和 b 。我们选取 TID2008 数据集中前 8 幅参考图像和对应的 544 幅失真图像作为测试子集, 在该子集上选取不同参数组进行数值实验, 通过最高的 SROCC 值来确定最佳参数组, 最终分别取 C_1 、 C_2 、 α 、 β 、 a 和 b 的值为 386, 0.001, 0.6, 0.02, 0.7, 0.3。

4.3 性能比较

实验中采用的对比算法包括 PSNR、SSIM^[3]、MS-SSIM^[26]、FSIM^[10]、FSIMc^[10]、GMSD^[4]、SRRR^[5]、RVSIM^[27] 和 SUMMER^[7]。这些算法均采用作者公布的代码及其参数设置。表 1 列出了本文算法与 9 个对比算法在 5 个测试数据集上的评价结果, 其中加粗显示了排名前两位的实验结果。从表中可以看出, CNCI 出现 12 次, SRRR 出现 10 次, FSIMc 出现 8 次, GMSD 出现 6 次, FSIM 出现 4 次。在 TID2008 数据集上, CNCI 的 SROCC 和 KROCC 值均为最高, PLCC 和 RMSE 值与 SRRR 相当, 并优于大多数算法, 这表明其可以获得与主观评价较高的一致性。在 TID2013 和 KADID-10k 数据集上, CNCI 的 4 个指标均为最高, 具有很好的质量评价能力, 这是因为 CNCI 考虑了颜色信息, 对于颜色失真类型的图像能够获得更好的评价效果。而 SRRR 和 SUMMER 算法中也引入了颜色信息, 但其总体性能并不突出, 只在 CSIQ 数据集上 SRRR 算法的性能略高于 CNCI。

表 1 不同 IQA 模型在 TID2008、TID2013、LIVE、CSIQ 和 KADID-10k 数据集上的实验结果比较

Tab.1 Comparison of experimental results of different IQA models on TID2008, TID2013, LIVE, CSIQ and KADID-10k databases

数据集	性能指标	PSNR	SSIM	MS-SSIM	FSIM	FSIMc	GMSD	SRRR	RVSIM	SUMMER	CNCI
TID2008	SROCC	0.524 5	0.774 9	0.852 8	0.880 5	0.884 0	0.890 6	0.888 5	0.737 5	0.828 9	0.900 9
	KROCC	0.369 6	0.576 8	0.654 3	0.694 6	0.699 1	0.709 0	0.702 4	0.562 8	0.644 2	0.718 4
	PLCC	0.530 9	0.773 2	0.842 0	0.873 8	0.876 2	0.878 6	0.882 5	0.795 4	0.820 9	0.881 1
	RMSE	1.137 2	0.851 1	0.729 2	0.652 5	0.646 8	0.640 8	0.631 2	0.813 3	0.766 4	0.634 7

续 表

数据集	性能指标	PSNR	SSIM	MS-SSIM	FSIM	FSIMc	GMSD	SRRR	RVSIM	SUMMER	CNCI
TID2013	SROCC	0.639 6	0.741 7	0.7859	0.801 5	0.851 0	0.803 8	0.875 8	0.675 7	0.855 6	0.890 1
	KROCC	0.469 8	0.558 8	0.604 7	0.628 9	0.666 5	0.633 4	0.691 4	0.514 6	0.666 8	0.713 5
	PLCC	0.701 7	0.789 5	0.832 9	0.858 9	0.876 9	0.855 1	0.892 0	0.782 5	0.861 3	0.895 0
	RMSE	0.883 2	0.760 8	0.686 1	0.634 9	0.595 9	0.642 8	0.560 4	0.771 9	0.630 2	0.553 0
LIVE	SROCC	0.875 5	0.947 9	0.944 3	0.963 4	0.964 5	0.960 3	0.958 8	0.960 0	0.932 7	0.950 5
	KROCC	0.686 4	0.796 3	0.792 1	0.833 7	0.836 3	0.826 9	0.821 3	0.820 3	0.768 0	0.802 9
	PLCC	0.872 1	0.944 9	0.942 8	0.959 7	0.961 3	0.942 1	0.957 6	0.957 0	0.927 9	0.946 9
	RMSE	13.368 0	8.945 5	9.095 0	7.678 0	7.529 6	9.159 6	7.824 2	7.927 4	9.739 8	8.682 6
CSIQ	SROCC	0.805 7	0.875 6	0.913 3	0.924 2	0.931 0	0.957 2	0.947 9	0.897 9	0.830 0	0.936 0
	KROCC	0.608 0	0.690 7	0.739 3	0.756 7	0.769 0	0.813 4	0.796 4	0.723 4	0.658 9	0.774 2
	PLCC	0.800 1	0.861 3	0.899 1	0.912 0	0.919 2	0.954 2	0.943 7	0.923 6	0.817 7	0.922 9
	RMSE	0.157 5	0.133 4	0.114 9	0.107 7	0.103 4	0.078 6	0.086 9	0.100 7	0.151 1	0.101 1
KADID-10k	SROCC	0.675 7	0.723 9	0.802 3	0.829 4	0.853 8	0.847 4	0.808 3	0.717 7	0.723 1	0.863 7
	KROCC	0.487 6	0.536 7	0.608 8	0.639 0	0.665 1	0.664 0	0.612 9	0.537 7	0.540 4	0.674 4
	PLCC	0.661 2	0.722 7	0.801 2	0.826 3	0.851 2	0.804 8	0.807 0	0.730 2	0.670 3	0.861 9
	RMSE	0.940 3	0.748 3	0.642 3	0.609 7	0.568 2	0.642 6	0.639 3	0.739 7	0.925 6	0.549 0

实验进一步验证了模型关于单一失真类型的评价效果。表 2 列出了本文算法与 9 个对比算法在 TID2013 数据集上对每一种失真类型关于 SROCC 指标的评价结果,并加粗显示了排名前两位的实验结果。从表中可以看出,CNCI 模型

可以在 TID2013 中大部分失真类型上产生较好的效果,且在 AGN、JPEG、JP2K、JGTE 和 J2TE 失真类型上获得最佳效果。就单一失真类型来看,CNCI 能够达到与 RMSD 和 SUMMER 相当的性能。

表 2 不同 IQA 模型在 TID2013 数据集上单一失真性能(SROCC)的比较

Tab. 2 Comparison of single distortion performance (SROCC) of different IQA models on the TID2013 databases

性能指标	PSNR	SSIM	MS-SSIM	FSIM	FSIMc	GMSD	SRRR	RVSIM	SUMMER	CNCI
AGN	0.933 7	0.867 1	0.864 6	0.897 3	0.910 1	0.946 1	0.909 8	0.865 2	0.938 0	0.942 7
ANC	0.856 8	0.772 6	0.773 0	0.820 8	0.853 7	0.868 9	0.885 9	0.814 5	0.888 5	0.870 6
SCN	0.924 1	0.851 5	0.854 4	0.875 0	0.890 0	0.934 8	0.893 5	0.840 5	0.939 1	0.934 4
MN	0.838 5	0.776 7	0.807 3	0.794 4	0.809 4	0.708 5	0.794 9	0.820 2	0.796 0	0.738 4
HFN	0.916 5	0.863 4	0.860 4	0.898 4	0.904 0	0.916 4	0.905 2	0.885 0	0.947 2	0.924 7
IN	0.926 3	0.750 3	0.762 9	0.807 2	0.825 1	0.763 3	0.829 6	0.881 1	0.915 9	0.864 5
QN	0.863 6	0.865 7	0.870 6	0.871 9	0.880 7	0.905 7	0.872 7	0.752 4	0.870 9	0.867 3
GB	0.912 3	0.966 8	0.967 3	0.955 1	0.955 1	0.911 4	0.951 2	0.975 5	0.965 0	0.961 3
DEN	0.945 7	0.925 4	0.926 8	0.930 2	0.933 0	0.952 5	0.946 1	0.894 9	0.958 9	0.946 4
JPEG	0.925 2	0.920 0	0.926 5	0.932 4	0.933 9	0.949 9	0.947 3	0.929 4	0.947 6	0.953 8
JP2K	0.893 4	0.946 8	0.950 4	0.957 7	0.958 9	0.965 6	0.967 7	0.952 1	0.966 5	0.968 8
JGTE	0.787 4	0.849 3	0.847 5	0.846 4	0.861 0	0.840 1	0.891 0	0.842 9	0.840 9	0.909 3
J2TE	0.858 1	0.882 8	0.888 9	0.891 3	0.891 9	0.913 5	0.906 3	0.885 5	0.912 2	0.921 1
NEPN	0.693 0	0.782 1	0.796 8	0.791 7	0.793 7	0.814 3	0.795 4	0.749 7	0.799 6	0.804 3
Block	0.130 0	0.572 0	0.480 1	0.548 9	0.553 2	0.663 0	0.258 5	0.603 8	0.376 7	0.437 2

续 表

性能指标	PSNR	SSIM	MS-SSIM	FSIM	FSIMc	GMSD	SRRR	RVSIM	SUMMER	CNCI
MS	0.764 7	0.775 2	0.790 6	0.753 1	0.748 7	0.735 6	0.650 0	0.602 8	0.660 9	0.728 7
CTC	0.460 8	0.377 5	0.463 4	0.468 6	0.467 9	0.325 3	0.434 3	0.091 6	0.905 4	0.473 9
CCS	0.689 2	0.414 1	0.409 9	0.274 8	0.835 9	0.190 7	0.686 6	0.437 8	0.665 4	0.816 1
MGN	0.908 5	0.780 3	0.778 6	0.846 9	0.856 9	0.888 0	0.862 4	0.824 7	0.918 9	0.904 2
CN	0.883 8	0.856 6	0.852 8	0.912 1	0.913 5	0.929 8	0.911 3	0.899 1	0.943 2	0.926 2
LCNI	0.938 6	0.905 7	0.906 8	0.946 6	0.948 5	0.963 1	0.931 5	0.915 5	0.969 7	0.954 7
ICQD	0.908 7	0.854 2	0.855 5	0.876 0	0.881 5	0.909 8	0.891 3	0.806 5	0.906 9	0.887 8
CHA	0.775 9	0.877 5	0.878 4	0.871 5	0.892 5	0.851 7	0.883 8	0.917 8	0.893 7	0.880 1
SSR	0.915 2	0.946 1	0.948 3	0.956 5	0.957 6	0.968 4	0.960 9	0.944 1	0.969 2	0.962 9

4.4 消融实验

为了验证颜色名称距离(CND)特征对彩色图像质量评价模型预测性能的提升作用,实验在TID2008数据集上对CND特征、梯度特征和亮度特征进行消融实验。表3列出了在TID2008数据集上,CNCI模型仅使用CND特征、梯度特征和亮度特征,3个特征之间两两组合以及同时使用3个特征时所能达到的SROCC值。可以看出,同时使用3个特征能够得到最高的SROCC

值。由于HVS对亮度感知比对颜色感知更加敏感,因此仅使用亮度通道提取的梯度特征和仅使用亮度特征的效果要稍好于仅使用CND特征。但是从表3中可以看出,引入CND特征后,彩色图像质量评价模型的评价性能有明显提升,这证明了颜色名称距离对两幅彩色图像之间的感知色差度量的有效性。

5 结 论

本文提出一个基于颜色名称的彩色图像质量评价模型(CNCI),该模型采用颜色名称距离度量图像的整体感知颜色差异,同时结合梯度幅度相似性度量图像的结构变化。在质量分数池化阶段,将视觉显著性用作加权函数来表示局部图像区域的重要性,并加入亮度相似性作为补充,进一步提高算法的性能。在5个公开测试数据集上对CNCI和其他最新或著名的9个算法进行了比较,实验结果表明,该模型能够获得与主观评价更好的一致性,并且在TID2008、TID2013和最新的KADID-10k数据集中获得最佳效果,其SROCC值分别为0.900 9,0.890 1,0.863 7。

表3 CNCI模型在TID2008数据集上的消融实验性能(SROCC)比较

Tab.3 Comparison of ablation experimental performance (SROCC) of CNCI model on TID2008 database

特征组合	SROCC
CND	0.584 0
梯度	0.747 6
亮度	0.661 7
CND+梯度	0.843 0
CND+亮度	0.786 6
梯度+亮度	0.862 2
CND+亮度+梯度	0.900 9

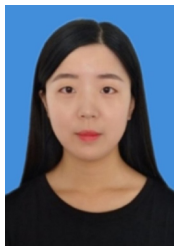
参 考 文 献:

- [1] 姚军财. 基于人眼对比度敏感视觉特性的图像质量评价方法[J]. 液晶与显示, 2011, 26(3): 390-396.
YAO J. Image quality assessment method based on contrast sensitivity characteristics of human vision system[J]. *Chinese Journal of Liquid Crystals and Displays*, 2011, 26(3): 390-396. (in Chinese)
- [2] WANG Z, BOVIK A C. *Modern Image Quality Assessment* [M]. San Rafael: Morgan & Claypool Publishers, 2006.
- [3] 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.

- [4] XUE W F, ZHANG L, MOU X Q, *et al.* Gradient magnitude similarity deviation: a highly efficient perceptual image quality index [J]. *IEEE Transactions on Image Processing*, 2014, 23(2):684-695.
- [5] LI L D, XIA W H, FANG Y M, *et al.* Color image quality assessment based on sparse representation and reconstruction residual [J]. *Journal of Visual Communication and Image Representation*, 2016, 38(7):550-560.
- [6] PREISS J, FERNANDES F, URBAN P. Color-image quality assessment: from prediction to optimization [J]. *IEEE transactions on image processing*, 2014, 23(3):1366-1378.
- [7] TEMEL D, ALREGIB G. Perceptual image quality assessment through spectral analysis of error representations [J]. *Signal Processing: Image Communication*, 2019, 70(2):37-46.
- [8] SUN W, LIAO Q M, XUE J H, *et al.* SPSIM: a superpixel-based similarity index for full-reference image quality assessment [J]. *IEEE Transactions on Image Processing*, 2018, 27(9):4232-4244.
- [9] TEMEL D, ALREGIB G. PerSIM: Multi-resolution image quality assessment in the perceptually uniform color domain [C]// 2015 *IEEE International Conference on Image Processing*, Quebec City, QC, Canada:ICIP, 2015: 1682-1686.
- [10] ZHANG L, ZHANG L, MOU X Q, *et al.* FSIM: a feature similarity index for image quality assessment[J]. *IEEE Transactions on Image Processing*, 2011, 20(8): 2378 - 2386.
- [11] KINSMAN T, FAIRCHILD M, PELZ J. Color is not a metric space implications for pattern recognition, machine learning, and computer vision [C]// 2012 *Western New York Image Processing Workshop*, Rochester, NY, USA:WNYIPW, 2012:37 - 40.
- [12] WEIJER J, SCHMID C, VERBEEK J, *et al.* Learning color names for real-world applications[J]. *IEEE Transactions on Image Processing*, 2009,18(7):1512-1523.
- [13] 吴金建. 基于人类视觉系统的图像信息感知和图像质量评价[D].西安:西安电子科技大学, 2014.
WU J J. Image information perception and image quality assessment based on human visual system[D]. Xi'an: Xidian University, 2014. (in Chinese)
- [14] HOFMANN T. Probabilistic latent semantic indexing [C]// *The 22nd annual international ACM SIGIR conference on Research and development*. Berkeley, California, USA: SIGIR, 1999: 50-57.
- [15] LEVINA E, BICKEL P. The earth mover's distance is the mallows distance: some insights from statistics[C]// *Proceedings Eighth IEEE International Conference on Computer Vision*, Vancouver, BC, Canada:ICCV, 2001: 251-256.
- [16] JAHNE B, HAUSSECKER H, GEISSLER P. *Handbook of computer vision and applications* [M]. San Diego: Academic press, 1999.
- [17] GEUSEBROEK J M, VANDEN B R, SMEULDERS A W M, *et al.* Color and scale: the spatial structure of color images [C]// *European Conference on Computer Vision*, Berlin, Germany:ECCV, 2000: 331-341.
- [18] VANDEN B L, CHRISTIAN J. *Vision models and applications to image and video processing* [M]. Berlin: Springer Science & Business Media, 2013.
- [19] WU J J, SHI G M, LIN W S. Survey of visual just noticeable difference estimation[J].*Frontiers of Computer Science*, 2019, 13(1): 4-15.
- [20] ZHANG L, GU Z Y, LI H Y. SDSP: a novel saliency detection method by combining simple priors[C]// 2013 *IEEE international conference on image processing*. Melbourne, VIC, Australia: ICIP, 2013: 171-175.
- [21] PONOMARENKO N, LUKIN V, ZELENSKY A, *et al.* TID2008-a database for evaluation of full-reference visual quality assessment metrics[J]. *Advances of Modern Radioelectronics*, 2009, 10(4): 30-45.
- [22] PONOMARENKO N, IEREMEIEV O, LUKIN V, *et al.* Color image database TID2013: Peculiarities and preliminary results[C]// *European workshop on visual information processing*. Paris, France: EUVIP, 2013: 106-111.
- [23] LARSON E C, CHANDLER D M. Most apparent distortion: full-reference image quality assessment and the role of strategy[J]. *Journal of electronic imaging*, 2010, 19(1): 011006.
- [24] SHEIKH H R, SABIR M F, BOVIK A C. A statistical evaluation of recent full reference image quality assessment algorithms[J].*IEEE Transactions on Image Processing*, 2006, 15(11): 3440-3451.

- [25] LIN H H, HOSU V, SAUPE D. KADID-10k: A large-scale artificially distorted IQA database [C] // 2019 *Eleventh International Conference on Quality of Multimedia Experience*. Berlin, Germany: QoMEX, 2019: 1-3.
- [26] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment [C]. *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*. Pacific Grove, CA, USA: ACSSC, 2003, 2:1398-1402.
- [27] YANG G Y, LI D S, LU F, *et al.* RVSIM: a feature similarity method for full-reference image quality assessment [J]. *EURASIP Journal on Image and Video Processing*, 2018(1): 1-15.

作者简介:



马 畅(1997—),女,黑龙江大庆人,硕士研究生,2019 年于陕西科技大学获得学士学位,主要从事图像处理及图像质量评价方面的研究。E-mail: machang971@163.com



张选德(1979—),男,宁夏固原人,博士,教授,2013 年于西安电子科技大学获得博士学位,主要从事图像恢复、图像质量评价、视觉跟踪、多光谱图像处理等方面的研究。E-mail: zhangxuande@sust.edu.cn