

基于极限学习机的底层特征全参考彩色图像 质量评价方法

马月梅,付 浩,刘国军,杨 玲,魏立力

(宁夏大学数学统计学院,宁夏 银川 750021)

[摘要] 作为图像质量的监测和评价工具,图像质量评价(image quality assessment, IQA)在各种图像处理系统中发挥着重要的作用,理想的 IQA 方法应该与人类视觉系统(human visual system, HVS)相一致. 目前 HVS 对图像的理解主要是依据图像的底层特征,本文提出了一种新的全参考(full reference, FR)彩色图像 IQA 方法. 首先,提取了结构对比度指标(structural contrast index, SCI)、梯度、局部二值模式(local binary pattern, LBP)和色度四类底层特征图,用于刻画图像的不同特征属性;其次,利用不同的特征池化策略对每类特征分别处理,将其组成一组相似特征向量作为图像质量的检测器并采用极限学习机(extreme learning machine, ELM)建立回归模型,得到客观的质量分数;最后,与目前流行的 8 种 FR IQA 方法在 5 个标准 IQA 数据库上进行数值实验. 结果表明,该方法整体性能优于其他方法,能够有效地提高大多数失真类型的预测精度.

[关键词] 彩色图像质量评价,底层特征,局部二值模式,梯度,结构对比度指标,极限学习机

[中图分类号] TP391.41; TP181 [文献标志码] A [文章编号] 1001-4616(2022)04-0091-11

Full Reference Color Image Quality Assessment Method via Low-level Features Combination with Extreme Learning Machine

Ma Yuemei, Fu Hao, Liu Guojun, Yang Ling, Wei Lili

(School of Mathematics and Statistics, Ningxia University, Yinchuan 750021, China)

Abstract: As an image quality monitoring and evaluation tool, image quality assessment(IQA) plays an important role in various image processing systems. The ideal IQA method should be consistent with human visual system(HVS). Suppose HVS understanding an image mainly according to its low-level features, a novel Full Reference(FR) color IQA method. Firstly, four different types of low-level feature maps are extracted, namely structural contrast index(SCI), gradient, local binary pattern(LBP), and chroma, which are used to characterize different feature attributes of the image. Secondly, different feature pooling strategies are employed to process each type of features respectively, and a set of similar feature vectors are formed as the detector of image quality. Then, extreme learning machine(ELM) is used to establish regression model and map the feature vectors into an objective quality score. Finally, extensive experiments performed on five benchmark IQA databases and compared with eight state-of-the-art FR IQA metrics. The results demonstrate that the overall performance of proposed method is better than other methods, and can effectively improve the accuracy of IQA index on most of distortions.

Key words: color image quality assessment, low-level features, local binary pattern, gradient, structural contrast index, extreme learning machine

随着多媒体时代的到来,以及各种新的手持设备和智能手机的逐渐普及,许多图像处理和计算机视觉领域的研究对设计客观图像质量评价(image quality assessment, IQA)算法的需求不断增加. 需要参考图像和失真图像的 IQA 方法称为全参考型(full-reference, FR),由于易于处理、计算简单、性能优越,FR 一直是

收稿日期:2022-03-28.

基金项目:国家自然科学基金项目(62061040)、宁夏区重点研发计划项目(2019BEG03056)、宁夏自然科学基金项目(2021AAC03039).

通讯作者:刘国军,博士,教授,博士生导师,研究方向:图像处理的小波和偏微分方程,图像质量评价,机器学习方法的研究. E-mail:liugj@nxu.edu.cn

IQA 研究领域的基础。

图像的失真能够影响人类的主观感知,通常主观感知的定义有两个必要过程:识别和解释。在识别过程中,人类通过感知图像的结构信息进行识别,如结构相似度(structural similarity, SSIM)^[1]等。由于图像信息通常具有冗余性,目前人类视觉系统(human visual system, HVS)主要基于底层特征加强对于图像的理解,例如边、角和线等。换言之,底层特征的显著性为HVS识别场景内容传递了关键信息,通过比较参考图像和失真图像之间的底层特征集,可以设计出一个良好的客观IQA指标。在解释过程中,图像的内容认为是可以理解的,其语义信息能够表达图像的内容及其属性。文献[2]提出了一种基于语义信息的IQA框架,认为在人类感知过程中,失真会降低视觉内容从而影响语义获取和识别,同时人对图像的主观感知是高级的、语义的过程。然而,对失真图像提取语义信息比较困难,往往需要通过构造底层特征到高级语义之间的映射来实现语义信息的提取,从而提取有效且具有代表性的底层特征显得十分必要。

FR IQA模型的基本框架分为两步:特征提取和池化。在池化过程中,存在两种不同的策略,即经典的IQA方法以及基于学习的IQA方法。经典的FR IQA主要是利用图像本身的特征来进行质量评价,如PSNR^[3]、SSIM^[1]、FSIM^[4]、VIS^[5]等都属于经典的方法,它们主要是采用算术平均或加权策略进行池化^[6]。基于学习的IQA主要是利用机器学习技术来模拟不同失真效果和图像质量之间的关系,通过使用人类主观质量分数对参数进行优化训练,使得基于学习的IQA模型所获得的评价性能更符合人类的感知。

目前,卷积神经网络(convolutional neural networks, CNN)和支持向量回归(support vector regression, SVR)是研究人员采用较多的两种学习算法。文献[7]提出使用奇异值分解(singular value decomposition, SVD)中的奇异向量作为特征来量化图像中的主要结构信息,并利用加权平均池化方法进行融合,最后采用SVR进行图像质量预测。文献[8]提出的多方法融合(multi-method fusion, MMF) IQA,利用机器学习将多个现有的FR IQA方法结合起来,并从候选方法集中选择一个子集,实验证明当子集包含三种质量评估方法时,获得了较好的性能。文献[9]提出了一种基于CNN的FR IQA模型,即深度图像质量评估(deep image quality assessment, Deep-QA),它在没有任何先验知识的情况下学习HVS的视觉敏感特性。Deep-QA通过使用深度CNN,每个像素的视觉权重通过使用失真图像、其客观性误差图以及真实性主观评分来获得,同时利用端对端的优化训练,实现与人类主观评分相一致的评价得分。文献[10]提出了一个基于深度相似性(deep similarity, DeepSim)的FR IQA框架,该方法利用ConvNet模型,将失真图像和参考图像分别进行特征提取,通过计算不同特征图之间的相似性,最后将相似性结果池化得到全局质量分数。

然而,在使用这两种学习方法时可能会遇到一些问题,如过度拟合、人工干预试验、耗时和局部优化等^[11]。Huang等^[12]后续提出了极限学习机(extreme learning machine, ELM)学习方法,它是由单隐层前馈神经网络(single-hidden layer feed forward neural networks, SLFN)构成。ELM方法在执行过程中不需要调整网络的输入权值以及隐层的偏置,而且只需要设置网络的隐层节点个数就可以产生唯一的最优解。此外,ELM已被证明具有较高的学习准确性,已被广泛地应用于人脸识别、图像分割、人类行为识别和IQA^[13]等,其效率优于神经网络(neural networks, NN)和SVR。

通过上述分析,研究提出了基于极限学习机的底层特征(low level features based extreme learning machine, LLF-ELM)彩色FR IQA算法。本文的主要贡献如下:

(1)分别提取参考图像和失真图像四种不同的底层特征图,其中结构对比度指标(structural contrast index, SCI)图能够很好地刻画图像纹理的感知复杂度;梯度图可以有效反映图像局部对比度损失;加权局部二值(local binary pattern, LBP)直方图保留了更多的图像局部纹理结构信息,以及色度图能够较好地表达颜色信息在HVS视觉感知过程中的重要性;

(2)考虑到不同特征具有不同的属性,在图像质量评价时,使用同一种特征相似性度量方法将无法有效地刻画不同特征,本研究采用了不同的底层特征图相似特征指标,包括SCI图之间的相关系数,梯度图之间、色度图之间的偏差系数,以及加权LBP直方图之间的距离系数;

(3)采用ELM回归模型,学习特征向量与主观得分之间复杂的非线性关系,将特征向量映射为最终的图像质量分数;

(4)在TID2013, TID2008, CSIQ, IVC, MICT 5个标准数据库上与目前流行的FR IQA方法进行对比实验,结果表明提出的彩色IQA方法与主观评价具有较好的一致性。

1 基本原理

1.1 底层特征

1.1.1 色度

由于底层特征通常是从图像的亮度通道计算而来,它们并不能很好地处理色彩的变换,同时人眼对彩色图像与灰度图像感知能力也具有较大的差异^[14],从颜色空间和色度学等角度出发,融合图像的色度信息建立彩色 IQA 评价模型. 首先,给定一幅 RGB 空间的彩色图像,将其转换到 LMN 空间^[15],转换公式为:

$$\begin{bmatrix} L \\ M \\ N \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (1)$$

式中, L 为亮度通道, M 和 N 两个色度通道将作为刻画由颜色失真引起的质量退化的特征.

1.1.2 结构对比度指标

对比度掩蔽(contrast masking, CM)是指一幅原始图像中由于另一种成分的出现而引起的图像中某些部分的可见度有所降低,如图 1 所示.



图 1 对比度掩模效应例子

Fig. 1 Contrast mask effect example

图 1 是 CM 效应的一个例子,即一幅原始图像注入值为 50 的伪加性高斯白噪声生成的失真图像,其中红色和蓝色框分别表示均匀和复杂的纹理局部图像区域. 观察可知,HVS 很容易在均匀区域(如图中的红框)中感知到伪加性高斯白噪声失真,而在复杂纹理图案的图像区域(蓝框)中几乎感知不到. 这意味着 HVS 会根据图像的纹理特征感知不同的视觉质量与失真.

考虑到图像纹理复杂度是建模 CM 效应的重要条件,本文采用 SCI 作为纹理复杂度度量,其旨在同时考虑纹理图像的结构度和对比度强度. 将亮度图像 L 进行离散余弦变换(discrete cosine transform, DCT),选取一幅大小为 $n \times n$ 的 DCT 块,计算块系数的结构度(KT)和对比度强度(CI)的比率,则纹理图像的感知复杂度 SCI(τ)定义如下:

$$\tau = \frac{CI^\alpha}{KT^\beta}, \quad (2)$$

式中, α, β 为模型参数;式(2)中的对比度强度 CI 定义为 $CI = m_0/n^2$,其中 n 是 DCT 块的高度. 结构度 KT 反映了纹理图案的随机性,它是归一化 DCT 块的交流(alternating current, AC)系数值的峰度,定义为:

$$KT = \frac{m_4}{m_2^2}, \quad (3)$$

式中, m_k 是归一化 AC 系数(p)的 k 阶矩

$$m_k = \sum_{\omega \in B, \omega \neq 0} \omega^k p(\omega), \quad (4)$$

式中, ω 是第 (u, v) 个 DCT 系数的空间频率值, u 为水平方向频率, v 为垂直方向频率, $\omega = \delta \cdot \sqrt{u^2 + v^2}$, δ 为非负常数, $p(\omega)$ 是 ω 处 DCT 归一化系数的值,定义为:

$$p(\omega) = \frac{(\varepsilon + |c(\omega)|^\lambda)}{\sum_{\omega \in B, \omega \neq 0} (\varepsilon + |c(\omega)|^\lambda)}, \quad (5)$$

式中, $c(\omega)$ 是 ω 的 DCT 系数; ε 是一个正常数避免分母取到 0, λ 为调整参数.

在本文中, 使用逆 SCI($\tau^* = \tau^{-1}$) 进行 FR IQA 建模, 使得纹理复杂度越低的图像纹理区域具有更高的 τ^* 值. 逆 SCI(τ^*) 的模型参数根据经验设置为 $\varepsilon = 0.25$, $\lambda = \beta = \alpha = 1$. 为了便于计算, 对 τ^* 进行缩放, 公式简化为:

$$\tau^{-1}(\text{SCI}) = \frac{\sum_{(u,v) \in B} \{(u^2 + v^2)^2 \cdot (\varepsilon + |c(u,v)|)\}}{\sum_{(u,v) \in B} \{(u^2 + v^2) \cdot (\varepsilon + |c(u,v)|)\}^2} = \frac{m_4}{m_2^2}, \quad (6)$$

1.1.3 梯度

对于数字图像, 梯度幅值被定义为沿两个正交方向的图像方向梯度的均方根. 为了简化计算并引入一些噪声不敏感性, 利用 Prewitt 滤波器计算梯度, 对参考图像 $R(i,j)$ 和失真图像 $D(i,j)$ 分别计算梯度图, 结果如下:

$$\begin{aligned} G_R(i,j) &= \sqrt{(h_x * R)^2(i,j) + (h_y * R)^2(i,j)}, \\ G_D(i,j) &= \sqrt{(h_x * D)^2(i,j) + (h_y * D)^2(i,j)}. \end{aligned} \quad (7)$$

式中, h_x, h_y 分别是 Prewitt 滤波器的水平和垂直方向导数; $G_R(i,j), G_D(i,j)$ 分别表示在 (i,j) 处的参考图像和失真图像梯度幅值.

1.1.4 旋转不变一致 LBP

Ojala 等^[16-18] 提出旋转不变一致 LBP, 其原理是根据图像中的像素对其邻域的相对值进行编码, 并获取编码的统计信息, 从而识别出相应的纹理特征, 公式如下:

$$\text{LBP}_{P,r}^{\text{riu2}} = \begin{cases} \sum_{i=0}^{P-1} S(g_i - g_c), & U(\text{LBP}_{P,r}) \leq 2 \\ P + 1, & \text{其他} \end{cases} \quad (8)$$

$$U(\text{LBP}_{P,r}) = |S(g_{P-1} - g_c) - S(g_0 - g_c)| + \sum_{i=1}^{P-1} |S(g_i - g_c) - S(g_{i-1} - g_c)| \quad (9)$$

$$S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (10)$$

式中, LBP 上标“riu2”表示值小于等于 2 的旋转不变“均匀”模式. U 值定义为该模式中空间转换的次数(按位 0/1 变化); g_c, g_i 分别表示中心像素点和邻域像素点; P 是邻域的像素点总个数; r 是周围邻域像素点个数.

对参考图像和失真图像分别进行 LBP 运算后, 可分别得到 $R_LBP_{P,r}^{\text{riu2}}$ 图和 $D_LBP_{P,r}^{\text{riu2}}$ 图, 其图中每个位置的像素值表示该位置处的可见程度, 而具有高可见度的图像区域对后续的感知处理具有重要的作用.

1.2 构造相似特征向量

图像的失真会导致提取的特征发生变化, 因此, 本小节通过计算原始图像和失真图像的特征相似性来度量图像的失真程度. 对于不同的特征, 运用了不同相似性计算方法, 将参考图像 R 和失真图像 D 之间的相似性测量分为四个部分: SCI、梯度、色度和 LBP.

首先, R_{SCI} 和 D_{SCI} 之间的 SCI 相似性用相关系数 CC 表示, 计算如下:

$$S_1 = \text{CC}(R_{\text{SCI}}, D_{\text{SCI}}) = \frac{\sum_{i=1}^M \sum_{j=1}^N (R_{\text{SCI}}(i,j) - \bar{R}_{\text{SCI}}(i,j)) (D_{\text{SCI}}(i,j) - \bar{D}_{\text{SCI}}(i,j))}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N (R_{\text{SCI}}(i,j) - \bar{R}_{\text{SCI}}(i,j))^2} \sqrt{\sum_{i=1}^M \sum_{j=1}^N (D_{\text{SCI}}(i,j) - \bar{D}_{\text{SCI}}(i,j))^2}}, \quad (11)$$

其次, 梯度图 G_R, G_D 的相似性采用相似性标准偏差, 定义如下:

$$G(i,j) = \frac{2G_R(i,j)G_D(i,j) + T_1}{G_R^2(i,j) + G_D^2(i,j) + T_1}, \quad (12)$$

$$S_2 = \text{SD}_G = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (G(i,j) - \bar{G}(i,j))^2}{M \times N}}. \quad (13)$$

式中, T_1 是增加结果稳定性的正常数; $M \times N$ 是图像中的像素总数. 同样, M_R, N_R 和 M_D, N_D 分别为参考图像和失真图像的 M 和 N 色度分量, 相似性计算如下:

$$C(i, j) = \frac{2(M_R(i, j)M_D(i, j) + N_R(i, j)N_D(i, j)) + T_2}{M_R^2(i, j) + M_D^2(i, j) + N_R^2(i, j) + N_D^2(i, j) + T_2}, \quad (14)$$

$$S_3 = SD_C = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (C(i, j) - \bar{C}(i, j))^2}{M \times N}}. \quad (15)$$

式中, T_2 是一个正常数.

最后, 将参考图像和失真图像的 LBP 图表示成加权 LBP 直方图 H_R 和 H_D 的形式, 使其刻画纹理图像的全局结构特征, 公式如下:

$$H_R(k) = \sum_{i=1}^m \sum_{j=1}^n |R_L(i, j)| f(R_LBP_{p,r}^{riu2}(i, j), k), \quad (16)$$

$$H_D(k) = \sum_{i=1}^m \sum_{j=1}^n |D_L(i, j)| f(D_LBP_{p,r}^{riu2}(i, j), k).$$

$$f(x, y) = \begin{cases} 1, & x=y \\ 0, & \text{其他} \end{cases} \quad (17)$$

式中, $k \in [0, K]$, $K=9$ 为旋转不变 LBP 模式的最大值, $m \times n$ 表示图像大小; R_L 和 D_L 分别表示参考图像和失真图像的亮度分量图. 最后再通过 l_1 范数来计算 LBP 加权直方图相似度, 公式如下:

$$S_4 = \|H_R - H_D\|_1. \quad (18)$$

2 回归模型

给定一幅彩色图像, 通过计算局部相似性可得到一个 4 维的向量 $S = (S_1, S_2, S_3, S_4) = (CC_{sci}, SD_G, SD_C, HD_{LBP})$, 最终目标是找到特征向量 S 的一个函数 f , 映射得到图像质量的最终得分 Q , 即

$$Q = f(S), \quad (19)$$

式中, Q 可以归一化为 $[0, 1]$.

为估计 f , 本文采用 ELM^[12] 来学习特征向量 S 和主观质量分数之间的关系, 网络模型如图 2 所示.

ELM 目标是为了寻找一个 $f(S_i)$ 函数, 使其与主观质量得分之间的误差最小.

$$f(S_i) = \sum_{j=1}^l \beta_j g(W_j \cdot S_i + b_j), \quad i=1, \dots, n \quad (20)$$

式中, S_i 是第 i 幅图像所对应的一组特征向量, n 为样本数, W_j 和 b_j 分别是输入层到隐层的权重向量和偏置向量, β_j 为输出权重向量, g 是有界分段连续的激活函数, 采用 Sigmoid 函数.

为了更进一步分析, 可以将式 (20) 写成矩阵形式

$$Y_H \beta = Y, \quad (21)$$

式中, Y_H 为 ELM 网络的隐层输出矩阵, 可以表示为

$$Y_H = \begin{pmatrix} g(w_1 \cdot S_1 + b_1) & \cdots & g(w_l \cdot S_1 + b_l) \\ \cdots & \cdots & \cdots \\ g(w_1 \cdot S_n + b_1) & \cdots & g(w_l \cdot S_n + b_l) \end{pmatrix}_{n \times l}, \quad (22)$$

$$\beta = \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_l \end{pmatrix}_{l \times 1}, \quad Y = \begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix}_{n \times 1}$$

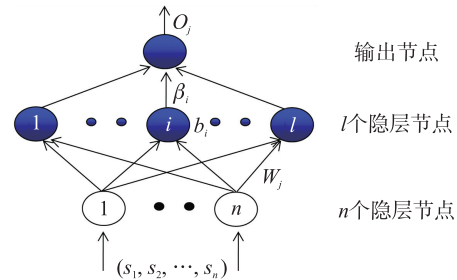


图 2 极限学习机网络结构

Fig. 2 Extreme learning machine network structure

式中, l 是 SLFN 中隐层节点个数, n 为输入特征数量.

与传统的学习算法不同^[19], ELM 采用最小权重范数和最小二乘法代替标准优化方法. 输出权重向量 β 可估计为

$$\beta = YY_H^* \quad (23)$$

式中, Y_H^* 是矩阵 Y_H 的 Moore-Penrose 广义逆, Bartlett^[20] 表明标准权重越小, 前馈神经网络的泛化性能越好. ELM 通过采用 Moore-Penrose 广义伪逆, 使得 β 在所有优化解中具有最小的权重范数. 对于上述的问题的求解, 正交投影法^[21]可以有效地用于计算 Moore-Penrose 逆, 也可根据岭回归理论^[22]对输出权重向量 β 进行估计. LLF-ELM 算法流程图如图 3 所示.

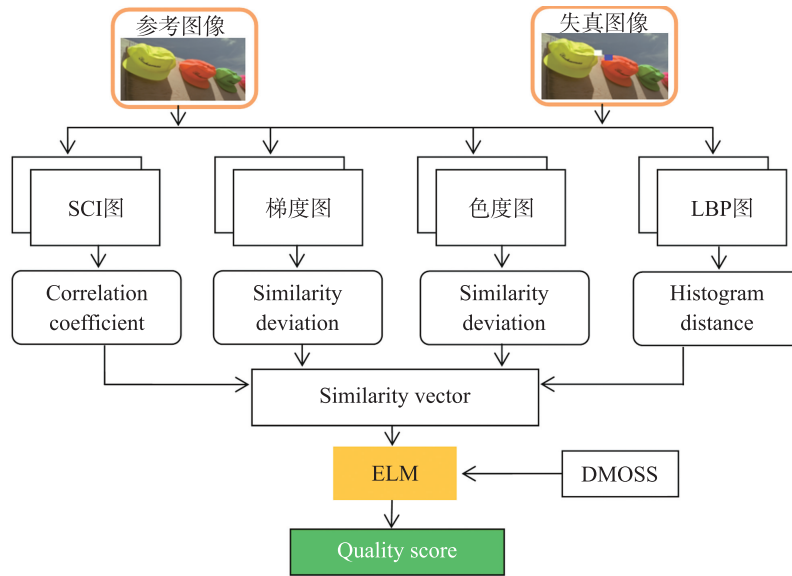


图 3 LLF-ELM 算法流程图

Fig. 3 Flowchart of low level features-extreme learning machine metric

3 结果与讨论

3.1 实验数据与实验设置

在实验中, 采用单隐层前馈神经网络 ELM 用于建立非线性回归模型, 用少量的隐层节点来训练大量样本, 可以节省大量训练时间同时, 能够达到较好的评价性能. 在 TID2013^[23], TID2008^[24], CSIQ^[25], IVC^[26], MICT^[27] 5 个不同的 IQA 数据库上进行了比较实验. 对于 5 个数据库, 使用相同数量的隐层节点 300 个. 实验部分与目前流行的 6 种基于非学习池化策略的 FR IQA 方法: GWL-SSIM^[28], SSIM^[1], FSIM^[4], PSNR^[3], VSI^[5] 和 SC-QI^[16], 2 种基于学习池化策略的 FR IQA 方法 SVD-SVR^[7], Deep-QA^[9] 进行比较.

为了说明算法的有效性, 即客观评价与 HVS 的一致性程度, 通常采用将客观评价值与主观评价值 (mean opinion score, MOS) 进行分析. 采用 4 个标准评估 IQA 指标来刻画: Spearman 等级相关系数 (spearman rank-order correlation coefficient, SROCC), Kendall 等级相关系数 (kendall rank-order coefficient, KRCC), Pearson 线性相关系数 (pearson linear correlation coefficient, PLCC) 和均方根误差 (root mean squared error, RMSE). 前两个相关系数可以测量图像质量评价算法的单调性, 其数值越大, 说明相关性越好. 计算 PLCC 和 RMSE 这两个指标需要进行回归分析, 提供一个客观得分与 MOS 之间的一个非线性映射. 对于非线性回归, 使用下面的映射函数^[29].

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x - \beta_3)}} \right) + \beta_4 x + \beta_5, \quad (24)$$

式中, $\beta_i, i=1, 2, \dots, 5$ 是相应选择的参数. 一个好的客观评价算法相应的有较高的 SROCC, KROCC, PLCC 同时有较低的 RMSE.

3.2 实验结果与分析

3.2.1 各个算法在数据库上的比较

首先,将 LLF-ELM 算法与对比算法在 5 个标准数据库中进行了性能评价. 在每个数据库上,随机选择 80% 的失真图像用于训练,20% 的失真图像用于测试. 取训练—测试组的 1 000 次迭代均值结果作为最终评价,如表 1 所示. 将 5 个标准数据库上 2 个最佳评价算法的 SROCC,PLCC 值(KROCC 和 RMSE 具有相同的评价效果)用黑体标注. 由表中的数据可以看出本文方法在各个数据库中的指标均取得了前两名的评价效果,在 5 个标准数据库中的 SROCC 值分别为 0.902 7、0.921 2、0.970 1、0.943 1 和 0.936 7,其中在 TID2013 中排名第二,其他四个数据库中排名第一. 同时本文方法 SROCC 和 PLCC 值在 TID2013 和 CSIQ 上,分别比基于 SVR 和 CNN 的 3 种 FR IQA 方法约高 0.02 和 0.01,与目前流行的 NR IQA 方法相比,本文方法获得了更高的预测精度.

表 1 不同评价算法在 5 个数据库中的实验结果(每个数据库上 2 个最佳评价算法的相关系数值用黑体标注)

数据库	评价指标	GWL-SSIM[28]	SSIM[1]	FSIM[4]	PSNR[3]	VSI[5]	SC-QI[16]	LLF-ELM
TID2013	SROCC	0.795 7	0.741 7	0.801 5	0.639 6	0.896 5	0.905 2	0.902 7
	KROCC	0.627 4	0.558 8	0.628 9	0.469 8	0.718 3	0.735 2	0.732 1
	PLCC	0.852 9	0.789 5	0.858 9	0.450 4	0.900 0	0.907 1	0.916 9
	RMSE	0.647 2	0.760 8	0.634 9	1.137 2	0.540 4	0.521 9	0.494 9
TID2008	SROCC	0.874 7	0.774 9	0.880 5	0.524 5	0.897 9	0.905 1	0.921 2
	KROCC	0.695 4	0.576 8	0.694 6	0.469 6	0.712 3	0.724 9	0.758 7
	PLCC	0.869 6	0.773 2	0.873 8	0.530 9	0.876 2	0.889 9	0.922 4
	RMSE	0.662 7	0.851 1	0.652 5	1.137 2	0.646 6	0.612 0	0.519 5
CSIQ	SROCC	0.961 7	0.875 6	0.924 2	0.805 7	0.942 3	0.943 4	0.970 1
	KROCC	0.821 0	0.690 7	0.756 7	0.608 0	0.785 7	0.787 0	0.847 0
	PLCC	0.944 1	0.861 3	0.912 0	0.800 1	0.927 9	0.926 8	0.969 5
	RMSE	0.086 6	0.133 4	0.107 7	0.157 5	0.097 9	0.098 6	0.065 9
IVC	SROCC	0.920 9	0.901 8	0.929 2	0.688 5	0.899 3	0.886 0	0.931 4
	KROCC	0.755 1	0.722 3	0.756 4	0.522 0	0.721 7	0.708 1	0.782 8
	PLCC	0.928 5	0.911 9	0.937 6	0.719 9	0.841 7	0.761 8	0.949 5
	RMSE	0.449 6	0.499 9	0.423 6	13.36 8	0.499 9	0.545 6	0.370 7
MICT	SROCC	0.924 0	0.879 4	0.905 9	0.613 0	0.865 9	0.817 9	0.938 1
	KROCC	0.756 7	0.693 9	0.730 2	0.444 7	0.673 7	0.620 1	0.795 9
	PLCC	0.929 9	0.888 7	0.907 8	0.642 6	0.736 2	0.618 2	0.950 0
	RMSE	0.475 9	0.573 8	0.524 8	0.958 8	0.617 5	0.697 3	0.388 8
Weight Average	SROCC	0.850 2	0.779 7	0.849 1	0.631 6	0.902 8	0.907 7	0.919 7
	KROCC	0.682 9	0.592 2	0.673 3	0.490 9	0.725 3	0.735 7	0.759 9
	PLCC	0.875 6	0.802 0	0.874 8	0.538 6	0.890 8	0.892 3	0.928 1
	RMSE	0.558 6	0.681 5	0.553 1	1.371 1	0.507 1	0.491 6	0.432 4

除此之外,还绘制了本文方法在 TID2013,TID2008,CSIQ,MICT,IVC 5 个数据库上预测得分与主观意见得分 MOS 的散点图,如图 4 所示. 从图中可知,该算法的散点图均匀地分布在曲线的两侧,这说明其客观评价得分与主观得分有较高的一致性.

3.2.2 对不同失真类型的评价效果比较

在 TID2008 数据库上测试了不同方法在 17 种失真类型的预测精度,其中每种失真类型的 2 个最佳算法的 SROCC 值用黑色粗体显示. 与目前流行的 6 种基于非学习和 2 种基于学习(浅层学习和深度学习各 1 种)的 FR IQA 方法进行比较,由表 2 可以看出,LLF-ELM 方法最优值与次优值约十余次,由此表明该方法在大多数失真类型上与部分 FR IQA 方法比较具有相同的竞争精度.

在表 2 中,对于模式噪声(Pattern-noise)和掩模噪声(masked noise)失真类型,LLF-ELM 方法的 SROCC 指标分别达到了 0.854 1 和 0.882 5,远优于其他评价方法;对于彩色高斯加性噪声(AWGN-color)失真类型,LLF-ELM 方法的 SROCC 值达到了 0.93,比 Deep-QA 和 VSI 指标分别约高 0.07、0.02;同时对于其他类型的噪声失真,也获得了很好的评价性能. 其主要原因是:1)融合的底层特征对质量预测较有

效. SCI 特征可以有效地削弱噪声对识别图像内容的细节信息造成的影响,LLF-ELM 方法对各种噪声失真均获得了较理想的评价结果;2)针对不同特征的属性采用不同的相似性度量池化是合理的,有助于提高 IQA 性能.从整体上看,LLF-ELM 方法组合的四种底层特征具有互补性,使得对不同的失真类型获得了优异的评价效果.当然,对于高斯模糊(Gaussian Blur,GB),JPEG-2000 压缩(JPEG-2000 Compression),JPEG 传输误差(JPEGTransmission Errors),JPEG-2000 传输误差(JPEG-2000 Transmission Errors),平均偏移(Mean Shift)这五种失真类型的评价效果相对最优的评价方法仍有很大的提升空间.

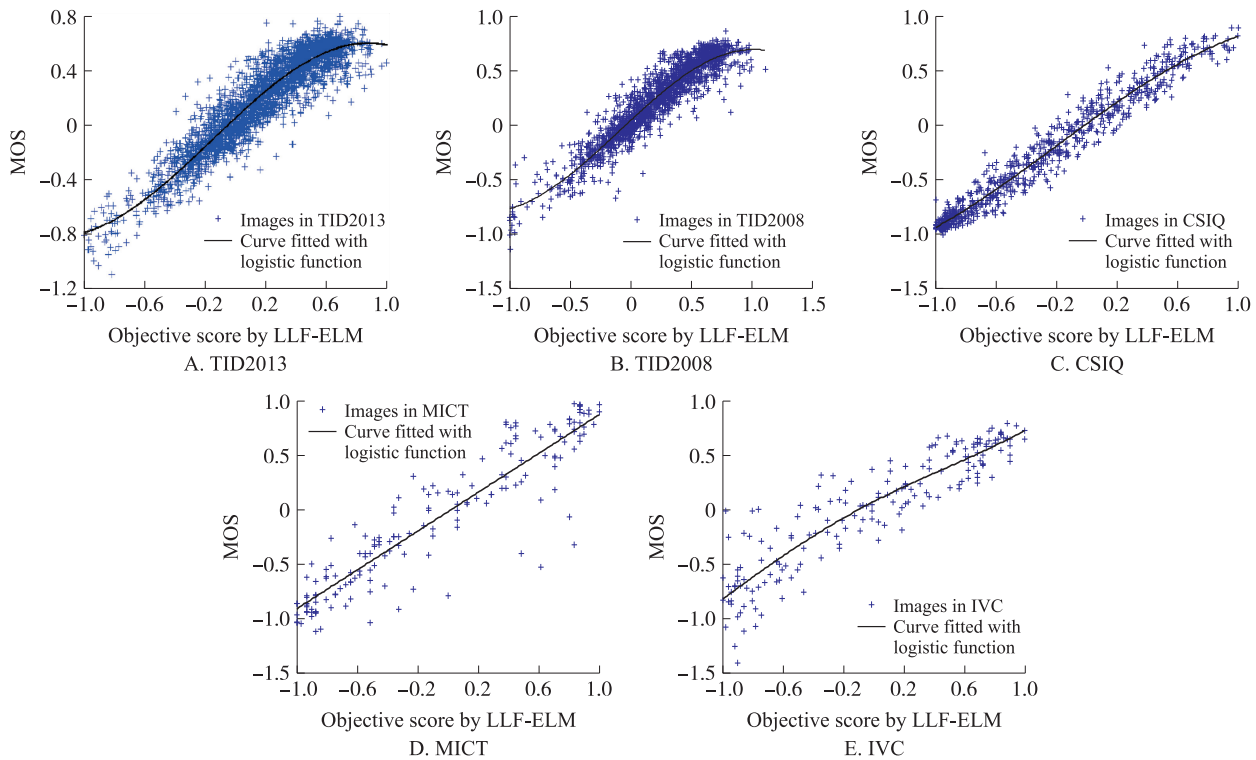


图 4 本文算法在 5 个数据库中预测得分与主观意见得分 MOS 的散点图

Fig. 4 The scatter plot of the prediction score and subjective opinion score MOS in the five databases

表 2 各个算法对 TID2008 标准数据库 17 种失真类型的实验结果比较

Table 2 Comparisons of experimental results of 17 distortion types in TID2008 standard database by each algorithm

distortion types	GWL-SSIM [28]	SSIM [1]	FSIM [4]	PSNR [3]	VSI [5]	SC-QI [16]	SVD-SVR [7]	Deep-QA [9]	LLF-ELM
awgn	0.8666	0.810 7	0.856 6	0.911 4	0.922 9	0.939 3	0.760 0	0.980 2	0.947 4
awgn-color	0.831 1	0.802 9	0.852 7	0.906 8	0.911 8	0.907 4	0.720 3	0.863 1	0.931 0
spatial corr-noise	0.891 2	0.814 4	0.848 3	0.922 9	0.929 6	0.941 2	0.787 5	0.970 4	0.944 4
masked noise	0.765 7	0.779 5	0.802 1	0.848 7	0.773 4	0.834 5	0.636 3	0.795 2	0.882 5
high-fre-noise	0.906 8	0.872 9	0.909 3	0.932 3	0.925 3	0.919 6	0.863 8	0.974 4	0.940 8
impulse noise	0.711 0	0.673 2	0.745 2	0.917 7	0.829 8	0.860 5	0.663 0	0.725 3	0.938 7
quanti-noise	0.854 8	0.853 1	0.856 4	0.869 9	0.873 1	0.902 7	0.813 0	0.901 1	0.907 1
Gaussian blur	0.945 6	0.954 4	0.947 2	0.868 2	0.952 9	0.962 1	0.812 0	0.950 4	0.931 0
denoising	0.973 4	0.953 0	0.960 3	0.938 1	0.969 3	0.958 8	0.889 3	0.929 3	0.971 4
jpg-comp	0.937 4	0.925 2	0.927 9	0.901 1	0.961 6	0.951 3	0.885 5	0.940 0	0.951 9
jpe2k-comp	0.981 8	0.962 5	0.977 3	0.830 0	0.984 8	0.963 7	0.902 7	0.958 3	0.979 0
jpg-trans-error	0.865 0	0.867 8	0.870 8	0.766 5	0.916 0	0.886 4	0.834 7	0.880 0	0.878 0
jpg2k-trans-error	0.898 6	0.857 7	0.854 4	0.776 5	0.894 2	0.899 6	0.792 8	0.928 2	0.891 7
pattern-noise	0.788 3	0.710 7	0.749 1	0.593 1	0.769 9	0.725 6	0.660 0	0.760 0	0.854 1
block-distortion	0.834 6	0.846 2	0.849 2	0.585 2	0.629 5	0.862 2	0.801 3	0.517 4	0.905 3
mean shift	0.649 9	0.723 1	0.672 0	0.697 4	0.671 4	0.704 7	0.515 2	0.652 1	0.669 6
contrast	0.662 4	0.524 6	0.648 1	0.612 6	0.655 7	0.616 9	0.436 0	0.838 2	0.718 8

3.2.3 交叉验证

此外,为了测试本文方法的泛化性和鲁棒性,进行了交叉数据库评估. 实验中,选择 TID2013(3 000 幅图像)和 CSIQ(866 幅图像)作为训练数据,与两种基于学习的方法(即 SVD-SVR^[7]和 FR-DCNN)进行了比较,其 SROCC 值如表 3 所示. 其中较为遗憾的是,FR-DCNN(基于深度学习)方法只有 TID2013 和 CSIQ 分别作为测试集和训练集的 SROCC 指标值,而在训练集数量(3 000 幅图像)相同情况下,本文方法 SROCC 值达到 0.946 0,比其高了约 0.01;同时随着训练集数量(866 幅图像)的减少,本文方法性能并没有明显的下降,由此说明本文方法有较好的普适性.

表 3 交叉数据库验证的 SROCC 值(对于训练和测试集中出现的相同图像在表中由“-”表示;“-”表示缺失值)

Table 3 SROCC values for cross-database validation (for the same image that appears in the training and test sets, “-” is indicated in the table, and “--” denotes missing values)

数据库	评价方法	TID2013	TID2008	CSIQ	MICT	IVC
TID2013	SVD-SVR[7]	-	0.883 2	0.893 1	0.855 8	0.875 5
	FR-DCNN[33]	-	--	0.931 0	--	--
	LLF-ELM	-	0.910 0	0.946 0	0.882 1	0.904 8
CSIQ	SVD-SVR[7]	0.745 8	0.755 0	-	0.832 7	0.882 8
	LLF-ELM	0.866 1	0.896 1	-	0.914 3	0.901 7

3.2.4 不同底层特征和池化策略的比较

LLF-ELM 方法分别提取了色度、SCI、梯度和 LBP 4 个特征,将 4 个特征分别记为 f_1 、 f_2 、 f_3 、 f_4 . 在使用 ELM 方法特征进行融合时,由于每种特征对最终预测的质量得分贡献不同,为了度量每个特征对最终质量评价结果的影响,对 4 个特征分别进行了感知特征分析来衡量重要性. 表 4 给出了在数据库中四组特征利用 ELM 融合得到的 SROCC 和 PLCC 值. 由图可以看出,SROCC 和 PLCC 这两个相关系数都是在同时使用四种特征建立回归模型的情形下,达到了最高值. 所以,本文提取的四种特征对最终的预测结果都是有用的,具有互补的效果.

表 4 不同特征组合对模型预测精度的影响比较

Table 4 Comparison of the influence of different feature combinations on model prediction accuracy

数据库	评价指标	f_1	f_2	f_3	f_4	$f_1+f_2+f_3+f_4$
TID2013	SROCC	0.661 3	0.629 1	0.856 8	0.437 7	0.902 7
	PLCC	0.642 1	0.638 2	0.887 0	0.602 6	0.916 9
TID2008	SROCC	0.595 2	0.655 4	0.882 8	0.527 2	0.921 2
	PLCC	0.594 1	0.654 2	0.906 2	0.657 0	0.922 4
CSIQ	SROCC	0.755 2	0.707 0	0.966 4	0.696 7	0.970 1
	PLCC	0.746 0	0.734 5	0.960 3	0.724 4	0.969 5

为了进一步说明 4 种特征对于质量评价的有效性和基于 ELM 池化技术的优点,进行了以下分析. 将提出的融合 4 种底层特征的算法与基于传统的 SSIM^[8]和基于 SVD^[7]算法进行了比较;对于池化策略,所提出的 ELM 与 SVR 进行比较,如表 5 所示. 首先在使用相同池化技术进行融合的前提下,比较不同特征的评价性能,结果表明所提出的特征在 4 个数据库中的 SROCC 值均表现最佳. 其次,在使用相同特征方法的前提下,验证不同池化策略的性能,可以发现基于 ELM 的方法在数据库中的预测精度都较优于基于 SVR 的方法. 因此,将基于 4 种不同类型的底层特征与基于 ELM 的池化策略相结合,与当前流行的算法相比,具有较强的竞争力.

表 5 对三种不同的底层特征描述符和两种池化方法在四个数据库上进行 SROCC 比较

Table 5 Compares SROCC on four databases for three different low-level feature descriptors and two pooling methods

数据集/模型	TID2008	CSIQ	MICT	IVC
SVD-SVR[7]	0.788 9	0.887 5	0.866 8	0.882 9
MMF-SVR[8]	0.917 6	0.957 5	0.921 3	0.925 3
LLF-SVR	0.920 4	0.968 9	0.926 2	0.928 0
SVD-ELM	0.792 8	0.892 3	0.870 4	0.888 5
MMF-ELM	0.919 0	0.969 6	0.927 6	0.927 3
LLF-ELM	0.921 2	0.970 1	0.936 7	0.931 4

4 结论

本文提出了一种基于ELM的底层特征全参考彩色IQA方法(LLF-ELM)。该方法由4个特征模块组成,分别为SCI、梯度、LBP以及色度图,针对不同的特征图采用不同的池化策略,将其组合成一组特征向量,最后采用ELM训练图像质量特征向量到主观得分之间的映射。在5个公开数据集上测试了LLF-ELM方法,并与其他8种具有代表性的IQA方法进行了比较,5个数据集上的SROCC值都超过了0.9,在数据集TID2013中的SROCC值达到了0.9027,结果表明LLF-ELM方法在预测精度方面拥有良好的表现。与目前流行的基于非学习和学习的FR IQA方法相比,本文方法在各种标准数据库和不同失真类型上预测的准确性和鲁棒性方面获得了最优的评价性能。

[参考文献]

- [1] 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.
- [2] JI W, WU J, SHI G, et al. Blind image quality assessment with semantic information[J]. Journal of visual communication and image representation, 2019, 58: 195–204.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 刘国军, 高丽霞, 陈丽奇. 广义平均的全参考型图像质量评价池化策略[J]. 光学精密工程, 2017, 25(3): 742–748.
- [7] NARWARIA M, LIN W. SVD-based quality metric for image and video using machine learning[J]. IEEE transactions on systems, man and cybernetics, part B(cybernetics), 2012, 42(2): 347–364.
- [8] LIU T J, LIN W, KUO C C J. Image quality assessment using multi-method fusion[J]. IEEE transactions on image processing, 2013, 22(5): 1793–1807.
- [9] KIM J, LEE S. Deep learning of human visual sensitivity in image quality assessment framework[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). USA: IEEE, 2017.
- [10] GAO F, WANG Y, LI P, et al. DeepSim: Deep similarity for image quality assessment[J]. Neurocomputing, 2017, 257(27): 104–114.
- [11] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1–3): 489–501.
- [12] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE transactions on systems, man and cybernetics, part B(cybernetics), 2012, 42(2): 513–529.
- [13] SURESH S, BABU R V, KIM H J. No-reference image quality assessment using modified extreme learning machine classifier[J]. Applied soft computing journal, 2009, 9(2): 541–552.
- [14] 岳靖, 刘国军, 付浩. 四元数谱余量彩色图像质量评价[J]. 激光与光电子学进展, 2019, 56(3): 31009–1–8.
- [15] GEUSEBROEK J M, BOOMGAARD R V D, SMEULDERS A W M, et al. Color invariance[J]. IEEE transactions on pattern analysis and machine intelligence, 2001, 23(12): 1338–1350.
- [16] BAE S H, KIM M. A novel image quality assessment with globally and locally consistent visual quality perception[J]. IEEE transactions on image processing, 2016, 25(5): 2392–2406.
- [17] OJALA T, PIETIKAINEN M, MAENPAA T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE transactions on pattern analysis and machine intelligence, 2002, 24(7): 971–987.
- [18] 刘丽, 谢毓湘, 魏迎梅, 等. 局部二进制模式方法综述[J]. 中国图象图形学报, 2018, 19(12): 1696–1720.
- [19] LIU T J, LIN W, KUO C C J. Image quality assessment using multi-method fusion[J]. IEEE Transactions on image processing, 2013, 22(5): 1793–1807.
- [20] BARTLETT P L. The sample complexity of pattern classification with neural networks: The size of the weights is more

- important than the size of the network[J]. IEEE transactions on information theory,2002,44(2):525-536.
- [21] RAO C R. Generalized inverse of matrices and its applications[M]. New York,USA:Wiley. 1971.
- [22] HOERL A E, KENNARD R W. Ridge regression: biased estimation for nonorthogonal problems[J]. Technometrics, 1970, 12(1):55-67.
- [23] PONOMARENKO N, JIN L, LEREMEIEV O, et al. Image database TID2013: peculiarities, results and perspectives[J]. Signal processing: image communication, 2015, 30:55-77.
- [24] PONOMARENKO N, LUKIN V, ZELENSKY A, et al. TID 2008—a database for evaluation of full-reference visual quality assessment metrics[J]. Advances of modern radio electronics, 2009, 10:30-45.
- [25] LARSON C, CHANDLER D M. Categorical image quality(CSIQ) database 2009[OL]. <http://vision.okstate.edu/csiq>.
- [26] NINASSI A, CALLET P L, AUTRUSSEAU F. Subjective quality assessment IVC database 2005[OL]. <http://www2.irccyn.ecnantes.fr/ivcdb>.
- [27] HORITA Y, SHIBATA K, KAWAYOKE Y, et al. MICT: Image quality evaluation database 2000[OL]. <http://mict.eng.utoyama.Ac.jp/mictdb.html>.
- [28] 高丽霞, 刘国军, 张选德. 快速提升的结构相似度 IQA 算法[J]. 计算机辅助设计与图形学学报, 2017, 29(07):1314-1323.
- [29] KANG L, YE P, LI Y, et al. Convolutional neural networks for no-reference image quality assessment[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Beijing:IEEE Computer Society, 2014:1733-1740.

[责任编辑:黄 敏]