

全参考图像质量评价回顾与展望

高敏娟¹, 党宏社¹, 魏立力², 刘国军², 张选德³

(1. 陕西科技大学电气与控制工程学院, 陕西西安 710021; 2. 宁夏大学数学统计学院, 宁夏银川 750021;
3. 陕西科技大学电子信息与人工智能学院, 陕西西安 710021)

摘要: 全参考图像质量评价(Full Reference Image Quality Assessment, FR-IQA)是 IQA 领域广为研究的类型之一. 本文回顾了 FR-IQA 的发展历程, 对 FR-IQA 应用现状和通用 FR-IQA 问题的构建进行综述, 以及对 FR-IQA 算法进行总结和梳理. 并在此基础上, 重点分析了现有研究中存在的问题, 包括问题构建的合理性、建模的全面性问题、知识驱动与数据驱动结合的问题等. 基于对主观评价过程的深入分析, 结合现有研究存在的问题, 探讨了主观评分采用模糊建模和知识与数据联合驱动构建算法两个可能的研究方向, 以期对后续的研究者提供参考.

关键词: 图像质量评价; FR-IQA; 主观评分; 模糊建模; 知识与数据联合驱动

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2021)11-2261-12

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20200780

Review and Prospect of Full Reference Image Quality Assessment

GAO Min-juan¹, DANG Hong-she¹, WEI Li-li², LIU Guo-jun², ZHANG Xuan-de³

(1. School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China;

2. School of Mathematics and Statistics, Ningxia University, Yinchuan, Ningxia 750021, China;

3. School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China)

Abstract: Full reference image quality assessment (FR-IQA) is one of the types widely studied in the field of IQA. This paper reviews the development of FR-IQA, summarizes the application status of FR-IQA and the construction of general FR-IQA problems, and summarizes and combs FR-IQA algorithms. And on this basis, it focuses on the problems existing in the existing research, including the rationality of the problem construction, the comprehensive problem of modeling, and the problem of the combination of knowledge-driven and data-driven. Finally, based on the in-depth analysis of the subjective evaluation process and the existing problems in the existing research, the two possible research directions of subjective scoring using fuzzy modeling and knowledge-data-driven construction algorithm are discussed, in order to provide reference for subsequent researchers.

Key words: image quality assessment; FR-IQA; subjective rating; fuzzy modeling; knowledge-data joint driven

1 引言

图像是表达信息的重要载体之一. 随着信息技术的发展, 图像处理技术已被广泛应用于军事、航空航天、电子商务、医疗卫生、社会管理、工农业自动化等多个领域. 基于图像的处理过程通常涉及图像的采集、压缩、存储和传输等, 在这些过程中会不可避免地引入图像失真. 如: 图像采集受硬件条件和环境因素的限制可能会引入噪声; 图像压缩过程中, 有损压缩可能会带来图像的模糊和振铃现象. 这些图像失真直接影响着信息表达的充分性和准确性, 导致图像视觉质量的降低.

如何量化图像质量, 以便对图像质量进行控制或增强, 催生了对图像质量评价问题的研究.

人是图像的最终感知者, 图像质量评价最直接可靠的方法为人的主观评价. 但主观评价易受到观测者经验、情绪、年龄、审美情趣等个体因素的影响, 为消除这种影响通常邀请大量的观测者对同一幅图像质量进行打分, 以平均意见分值(Mean Opinion Score, MOS)作为图像的主观评分. 主观评价方法虽然被认为是图像质量评价的“金标准”, 但是该方法成本高, 耗时长, 且不能应用于实际系统中, 因此有必要研究客观图像质

量评价方法. 客观图像质量评价设计计算的模型自动地评价图像质量, 以尽可能达到与人的主观评价相一致的效果. 基于应用范围的不同, 客观图像质量评价可分为通用型和面向特定应用型. 通用型不预设特定的失真类型, 其评价模型试图胜任多种不同应用; 特定型则针对具体应用场景. 根据评价过程有无真实图像作为参考, 客观图像质量评价可分为全参考评价、部分参考评价和无参考评价^[1]. 全参考评价指在参考图像全部信息可获得的情况下, 评价失真图像的质量. 这里的“参考图像”可被认为是无失真图像或质量完美的图像.

在图像质量评价几十年的发展历程中, 研究者已做了大量的研究工作. 因此, 对这些研究工作进行适时的梳理和总结将非常必要. 本文首先回顾FR-IQA的发展历程, 对FR-IQA的应用现状和通用FR-IQA问题的构建进行综述, 以及对FR-IQA算法进行总结和梳理; 其次, 在回顾FR-IQA的基础上, 重点分析现有研究中存在的问题, 包括问题构建的合理性、建模的全面性问题、知识驱动与数据驱动结合的问题; 最后, 基于对主观评价过程的深入分析和现有研究存在的问题, 探讨主观评分采用模糊建模, 以及知识与数据联合驱动构建算法两个可能的研究方向.

2 全参考图像质量评价回顾

2.1 全参考图像质量评价的应用现状及挑战

大多数图像处理算法和设备都在致力于保持或提高供人类视觉欣赏的数字图像的外观质量, 因而开发与人类视觉感知一致的自动质量评估算法有着重要的应用价值^[2]. 在此, 对全参考评价应用的几个方面进行梳理.

(1) 用于评价和指导图像压缩编码.

在图像处理系统中, 图像压缩以有损压缩为主, 以尽可能低的压缩码率同时保持尽可能高的图像质量. 由于有损压缩不能完全恢复原始数据, 因而易使图像质量下降, 要想在图像质量和压缩码率之间做出均衡, 图像质量评价是做好这一均衡的前提. 峰值信噪比 (Peak Signal Noise Ratio, PSNR) 是目前编码标准 H. 264 和最新 H. 265 中广泛采用的客观评价方法. 将影响主观感知质量的视觉效应引入编码过程, 利用图像质量评价指标作为图像失真度量, 可使得编码器在人类主观感知质量准则下获得最大的编码性能^[3].

(2) 用于图像处理算法的测试基准.

图像质量指标可用作比较不同图像处理算法的评价准则, 目前涉及图像恢复算法、图像清晰化处理算法、图像融合算法等. 图像去噪算法^[4]中采用均方误差 (Mean Squared Error, MSE) 测量去噪后图像与干净图

像之间的偏差程度, 作为验证去噪算法性能的准则之一; 图像增强算法^[5]中采用结构相似性 (Structural Similarity, SSIM)^[6]和平均绝对误差 (Mean Absolute Error, MAE) 等指标对增强后的图像质量进行定量评价, 客观分析不同增强算法的鲁棒性和普适性; 图像融合算法^[7]采用 PSNR 衡量图像在融合过程中的失真程度, 采用 SSIM 衡量融合图像与源图像的结构相似度.

(3) 用作图像处理任务的优化目标.

将图像质量测度用作设计和优化新图像处理算法的目标是目前研究较少但非常有前景的一种应用^[8]. 图像处理方法通常以使 MSE 最小化为目的进行参数的调整. 早期研究中采用 SSIM 进行感知优化, 应用于图像恢复、无线视频流和图像合成中. 文献^[9]中采用 MAE (1-范数) 进行优化, 应用于图像恢复任务中, 性能优于 MSE (2-范数); 多尺度结构相似性指标 (Multiscale SSIM, MS-SSIM)^[10]被用于指导基于神经网络的图像超分辨率和压缩设计; 基于归一化拉普拉斯金字塔指标^[11]被用于优化图像渲染算法^[12]; 文献^[8]中采用不同全参考评价指标作为损失函数训练低层次视觉任务的深度神经网络.

(4) 用于特定场景的质量评价.

图像质量评价技术可推广到具体的应用场景. 在医学领域, 医学图像已成为临床医学诊断的重要依据, 依靠经验丰富的医生进行检测图像的方法效率低且成本高, 将图像质量评价用以自动医学图像质量检测是一种重要方法, 可节省医生的时间和避免患者的重新扫描^[13]. 在遥感领域, 遥感图像的应用发展对图像质量的要求越来越高, 客观图像质量评价可为遥感图像质量等级的判断提供技术支持^[14]. 在基于射线的检测技术中, 图像质量评价方法可为获取高质量的射线图像提供方法, 在后续缺陷检测和结构分析工作中具有重要意义^[15].

目前, 图像质量评价的研究及应用仍然是图像处理领域一个具有挑战性的问题. 首先, 图像质量评价任务本质上是对 HVS 的模拟, 这是整个人工智能的一个挑战. 对于人类观测者来说, 快速评估两幅图像之间的感知相似度几乎是毫不费力的, 但其潜在的过程却被认为是相当复杂的. 人究竟是如何认识外部世界并做出判断和预测的, 至今无法完全解读其机理. 而图像质量评价任务恰是模拟 HVS 这样一个复杂的系统, 实质上属于“黑箱系统”的模拟问题. 其次, 图像质量评价问题本身的构建是一个挑战. 图像质量评价任务是对图像感知质量做出判定, 属于判别问题, 但又与经典的判别任务不同. 经典的判别任务标签是确定的 (如人脸识别、指纹识别), 而图像质量评价任务的标签是主观质量, 人对图像质量的评判具有主观性和不确定性, 主观

质量应该如何刻画本身是值得研究的问题。再次,图像质量评价任务不仅与图像的退化类型有关,也与图像的内容有关。图像质量评价算法的设计既要用到图像低层视觉特征,也要用到图像高层视觉特征。如何有效提取特征,构建符合人眼视觉的评价指标是具有挑战性的问题。最后,现阶段大多数的评价算法都是基于各种 IQA 数据库和人工失真的图像,一方面,基于数据库的评价体系能否契合人类视觉值得进一步探讨,另一方面,实际场景中的图像失真更为复杂,对真实场景图像的评价仍有难度。图像质量评价是任务驱动的,在不同的应用场景下对图像质量的要求不同。例如,在消费电子里可应用通用质量评价,而在医学图像评价里则需根据医学标准界定何为高质量图像,这些都给质量评价问题带了挑战。

2.2 通用全参考图像质量评价问题的构建

全参考评价问题的研究一方面需要大量的参考图像和失真图像,另一方面需要主观评分来衡量客观评价模型的性能。为此,许多研究人员开发了包含大量图像和人工评分值的数据集,并制定了一系列的性能评估指标。

2.2.1 数据集

研究人员已建立了超过 20 个 IQA 数据集,在此介绍代表性的 LIVE^[16]、CSIQ^[17]、TID2008^[18]、TID2013^[19]、IVC^[20]、A57^[21]、Waterloo Exploration Database^[22]、BAPPS^[23]和 KADID-10k^[24]。

LIVE 数据集由美国德州大学奥斯丁分校图像与视频工程实验室开发,于 2006 年发布。LIVE 中包含 29 幅参考图像,5 种失真类型,4~5 种失真等级,共 779 幅失真图像。数据集中提供的图像主观评分 DMOS (Differential Mean Opinion Score) 值由 161 位观测者给出的约 25000 个数据统计得到,DMOS 取值范围为 [0, 100], 值越大表明图像质量越差。

CSIQ 数据集由美国俄克拉荷马州立大学电气与计算机工程学院开发,于 2010 年发布。CSIQ 中包含 30 幅参考图像,6 种失真类型,4~5 种失真等级,共 866 幅失真图像。数据集中提供的图像主观评分 DMOS 值由 25 位观测者给出的约 5000 个数据统计得到,DMOS 取值范围为 [0, 1]。

TID 数据集由乌克兰航空航天大学开发。TID2008 于 2009 年发布,包含 25 幅参考图像,17 种失真类型,4 种失真等级,共 1700 幅失真图像。数据集中提供的图像主观评分 MOS 值由 838 位观测者给出的 256428 个数据统计得到,MOS 取值范围为 [0, 9], 值越大表明图像质量越好。TID2013 是在 TID2008 基础之上的扩展版本,于 2015 年发布,包含 25 幅参考图像,24 种失真类型,5 种失真等级,共 3000 幅失真图像。数据集中提供

的图像主观评分 MOS 值由 971 位观测者给出的 524340 个数据统计得到,MOS 取值范围为 [0, 9]。

IVC 数据集由法国南特中央理工大学开发,于 2005 年发布。IVC 中包含 10 幅参考图像,4 种失真类型,共 185 幅失真图像。数据集中提供的图像主观评分 MOS 值由 15 位观测者给出的数据统计得到,MOS 取值范围为 [0, 5]。

A57 数据集由美国康奈尔大学开发,于 2007 年发布。A57 中包含 3 幅参考图像,6 种失真类型,共 54 幅失真图像。数据集中提供的图像主观评分 MOS 值由 7 位观测者给出的数据统计得到,MOS 取值范围为 [0, 1]。

Waterloo Exploration Database 数据集由加拿大滑铁卢大学电子与计算机工程学院开发,于 2017 年发布。数据集包含 4744 幅参考图像,4 种失真类型,5 种失真等级,共 94880 幅失真图像。

BAPPS 数据集由美国加州大学伯克利分校和 Adobe 公司共同开发,于 2018 年发布。BAPPS 中包含 187.7k 参考图像,425 种失真类型,共 375.4k 幅失真图像。主观评分方法采用 2AFC (Two Alternative Force Choice) 和 JND (Just Noticeable Difference)。

KADID-10k 数据库由德国康斯坦茨大学计算机与信息科学系开发,于 2019 年发布。该数据库中包含 81 幅参考图像,每幅参考图像对应 25 种失真类型,每种失真类型有 5 种失真等级,共 10125 幅失真图像。数据集中提供的主观评分 MOS 值由 2209 位观测者在众包平台上给出的数据统计得到,MOS 取值范围为 [1, 5]。

2.2.2 评估指标

IQA 算法性能评估的常用方法为计算主客观评分之间的相关性。广泛采用的评价指标有斯皮尔曼等级相关系数 SROCC、肯德尔等级相关系数 KROCC、皮尔逊线性相关系数 PLCC 和均方根误差 RMSE。一个好的算法应具有较高的 SROCC、KROCC、PLCC 值和较低的 RMSE 值。

记 $\{x_1, x_2, \dots, x_n\}$ 为测试图像的主观评分值, $\{y_1, y_2, \dots, y_n\}$ 为测试图像某种算法的客观评分值, n 为测试图像总数。SROCC 和 KROCC 用来计算数据之间排序的相关性,衡量算法预测的单调性。SROCC 定义为

$$\text{SROCC} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (1)$$

其中, d_i 为第 i 个测试图像主观评分 x_i 与客观评分 y_i 分别在各自数据序列中排序位置之差。

KROCC 定义为

$$\text{KROCC} = \frac{n_c - n_d}{0.5n(n-1)} \quad (2)$$

其中, n_c 为数据集中一致对的个数; n_d 为数据集中不一致对的个数.

PLCC 衡量算法预测的准确性, RMSE 衡量主客观评分之间的绝对误差. 计算时需要先进行回归分析, 建立主客观评分之间的非线性映射, 将 y_i 映射到 q_i .

$$q_i = \beta_1 \left[\frac{1}{2} - \frac{1}{1 + e^{\beta_2(y_i - \beta_3)}} \right] + \beta_4 y_i + \beta_5 \quad (3)$$

其中, $\beta_i (i=1, 2, \dots, 5)$ 为可调节的参数.

PLCC 定义为

$$PLCC = \frac{\sum_{i=1}^n (x_i - \bar{x})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (q_i - \bar{q})^2}} \quad (4)$$

其中, \bar{x} 和 \bar{q} 分别为 $\{x_1, x_2, \dots, x_n\}$ 和 $\{q_1, q_2, \dots, q_n\}$ 的均值.

RMSE 定义为

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - q_i)^2} \quad (5)$$

不断有新的评价方法被提出, 用以进行算法性能评估. 文献[22]提出三个评价准则来衡量算法的相对性能, 这三个准则从不同方面测试算法的鲁棒性和泛化能力, 分别为参考图像与失真图像区分性测试、排序一致性测试和两两偏好一致性测试. 文献[25]提出最大差异化竞争方法(Maximum Differentiation, MAD), 该方法首先由计算机合成一对图像, 使一个模型的响应最大化/最小化, 同时保持另一个模型不变. 然后重复这个过程, 但是两个模型的作用相反, 最后对这些生成的图像进行主观测试来评估两种模型的相对性能. 文献[26]在MAD方法的基础上提出gMAD(Group MAD), 将对两种模型的性能评估推广到多个模型, 通过模型间的进攻-防守游戏来评估算法优劣, gMAD方法在少量样本情况下算法性能比较具有优势.

2.3 全参考图像质量评价算法综述

传统的FR-IQA研究重点在于设计能够准确模拟HVS图像质量感知过程的特征, 而特征的设计依赖于对HVS的先验知识或假设. 目前, 大部分算法采用两阶段框架, 如图1所示. 第一阶段为局部结构相似性度量, 即计算图像局部质量; 第二阶段为池化, 将局部结构相似性综合为图像质量指标.

2004年, Wang等人^[6]假设HVS高度自适应于提取视觉场景中的结构信息, 提出结构相似性(SSIM)算法,

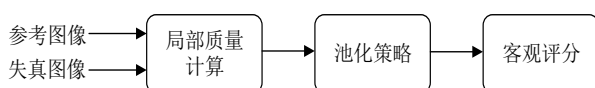


图1 FR-IQA算法两阶段框架

被认为是IQA领域具有里程碑意义的算法. 随后基于对SSIM的改进, Wang等人^[10]构建了多尺度结构相似性算法. 2005年, Sheikh等人^[27]提出信息保真度准则, 通过度量感知的参考图像和感知的失真图像之间信息共享的多少来给出图像质量评分. 随后, 又提出视觉信息保真度^[28], 该方法在小波域计算参考图像和失真图像经HVS感知通道后的信息量之比用来评价图像质量. 2006年, Sheikh等人^[29]提出针对全参考图像质量算法的统计评估方法. Shnayderman等人^[30]提出一种灰度图像质量度量方法, 该方法基于奇异值分解, 可有效测量相同失真类型下不同失真等级的图像质量以及不同失真类型的图像质量. Wang等人^[31]针对质量评估中第二阶段的三种空间域池化策略(Minkowski池化、局部质量加权池化、信息内容加权池化)进行了研究. Chen等人^[32]提出一种基于梯度的结构相似性算法, 算法性能优于SSIM尤其是对于模糊类型的失真. 2007年, Chandler等人^[33]提出基于小波的自然图像视觉信噪比算法, 该算法基于心理物理学对HVS的发现, 同时考虑初级和中级特征来量化失真图像的视觉保真度. 2009年, Sampat等人^[34]提出复小波结构相似性指标, 该算法的关键思想是图像失真会导致局部小波系数相位的变化, 通过在复小波域上提取特征来度量图像质量. Moorthy和Bovik^[35]提出用于质量评估的视觉重要性池化策略, 基于视觉注视和基于质量的加权方法可提高单尺度和多尺度SSIM的性能. 2010年, Chandler等人^[36]假定HVS使用多种策略来确定图像质量, 提出一种最明显失真的全参考评估方法. Zhang等人^[37]假定HVS根据低层特征感知图像质量, 以图像的一阶和二阶Riesz变换系数作为图像特征, 提出一种基于Riesz变换的特征相似性指标.

2011年, Wang等人^[38]在池化阶段采用信息量加权的策略, 提出IW-PSNR和IW-SSIM算法. Zhang等人^[39]以梯度和相位一致性为特征计算局部结构失真, 以相位一致性为显著性度量, 提出特征相似性算法(FSIM). 2012年, Liu等人^[40]提出梯度相似性的评价算法. Zhu和Wang^[41]提出一种基于多尺度视觉梯度相似度的全参考评估模型. Capodiferro等人^[42]提出一种基于两个不同指标的二维组合进行全参考评估的方法. 2013年, Xue等人^[43]采用梯度算子提取图像特征, 以标准差作为池化策略, 提出梯度模值相似度离差(GMSD)算法. Zhang等人^[44]假定图像中被感知到的边界是HVS语义认知的基础, 提出边界强度相似性算法(ESSIM). Wu等人^[45]基于内部生成机制启发, 将输入场景分解为具有视觉内容的预测部分和剩余内容的无序部分进行质量评估. Chang等人^[46]受启发于稀疏编码可以很好地描述初级视觉皮层中简单细胞的感受野, 提

出一种稀疏特征保真度评估指标. 2014年, Zhang等人^[47]提出视觉显著引导的评价算法(VSI). Preiss等人^[48]采用优化方法改善彩色图像差异指标的预测性能,特别是对于色域映射失真. 2015年, Pei等人^[49]提出基于高斯差分的全参考评价算法, Lee等人^[50]提出基于方向统计的彩色图像评价算法.

2016年, Nafchi等人^[51]提出了平均偏差相似性评价模型, 该模型利用了梯度相似度、色度相似度和偏差池化策略. Bau等人^[52]提出了结构对比度质量评价指标, 能够很好地表征具有结构失真的图像局部和全部视觉质量感知. Reisenhofer等人^[53]提出基于哈尔小波的评价算法. Li等人^[54]提出一种基于稀疏表示的具有自适应字典的图像质量评价算法, 该算法通过稀疏表示来捕获失真图像结构的变化来进行质量预测. 2017年, Ding等人^[55]综合梯度模、各向异性性和局部方向性构建特征, 提出各向异性的评价算法. Gu等人^[56]提出一种基于HVS的感知图像质量评价指标, 该指标采用多尺度的卷积运算、梯度模和颜色相似性, 并基于感知进行池化. 2018年, Sun等人^[57]通过度量参考图像和失真图像之间的超像素亮度相似性、超像素色度相似性和像素点梯度模相似性, 构造了超像素相似性算法. Ahar等人^[58]假设视觉质量评估任务与人类大脑稀疏编码的策略相关, 提出稀疏重要性等级测度. Claudio等人^[59]提出一种全参考评价方法, 该方法将两种独立的度量方法结合起来测量细节丢失和虚假细节对图像质量的影响. Jia等人^[60]提出基于对比度和视觉显著性相似性的评价算法. 2019年, Temel等人^[61]分析参考图像和失真图像之间误差信号的统计信息进行质量评估, 提出一种基于多尺度和多通道误差表示的频谱理解评估算法. 2020年, Kim等人^[62]考虑对失真的感知非线性地取决于失真类型, 提出一种动态感受野生成的全参考图像质量评估.

随着数据时代的到来, 近些年来数据驱动的方法开始兴起. 2012年, Charrier等人^[63]提出一种基于机器学习的图像质量度量方法, 该方法利用多支持向量机(SVM)对图像质量进行分类, 以特定SVM回归得到质量评分. Narwaria等人^[64]利用机器学习对基于奇异值分解的视觉特征进行质量评估, 以奇异值和向量构成视觉质量评估的所选特征, 将机器学习用于特征池化处理. 2013年, Liu等人^[65]提出采用多方法融合进行质量评估, 评分设置为来自多种方法评分的非线性组合, 并通过训练过程获得适当的权重. Gastaldo等人^[66]探讨了使用机器学习进行质量评估的优缺点, 表明在视觉质量评估中使用机器学习具有前景. 2015, Guha等人^[67]采用稀疏建模的方法学习图像的固有结构, 根据学习到的基本矢量比较参考图像和失真图像的结构来

估算失真图像的感知质量. 2016, Liang等人^[68]提出一种双路径深度卷积神经网络, 将相似场景但不一定对齐的图像作为参考进行质量评估. 2017年, Liu等人^[69]提出了一种基于并行增强思想的FR-IQA集成方法. Gao等人^[70]探讨了深度神经网络架构如何影响图像质量评估的性能, 提出了一种深度相似性的全参考评价框架, 该框架首先测量参考图像和失真图像特征(DNN模型产生的特征)之间的局部相似性, 然后聚合局部质量为整体质量得分. Wang等人^[71]提出的全参考IQA度量采用非负矩阵分解测量图像的退化, 以及应用极限学习机技术进行池化. Kim等人^[72]提出了一种基于卷积神经网络的深度图像质量评估模型, 区别于采用心理视觉研究的HVS特性, 该模型从IQA数据库的数据分布中学习HVS的行为, 而无需任何HVS的先验知识. 2018年, Bosse等人^[73]提出基于神经网络的评价方法, 网络采用端到端训练, 包括10个卷积层和5个池化层用于特征提取, 2个全连接层用于回归, 既可用于FR-IQA任务, 也可用于NR-IQA任务. Ma等人^[74]提出基于卷积神经网络(CNN)的几何变换不变性的图像质量评估, 阐述了图像预配准虽然能够提高传统知识驱动FR-IQA模型的质量预测性能, 但是因其计算量大而在实际应用中受限, 同时, 使用基于CNN的方法构建端到端IQA模型, 用于不需要配准的直接图像质量预测, 简单有效地解决了几何变换不变IQA问题. Zhang等人^[23]以深度特征作为感知度量提出LPIPS(Learned Perceptual Image Patch Similarity)评价指标, 证明了在监督、自监督, 无监督模型上得到的深度特征在模拟低层次感知相似性上都具有良好的表现. Prashnani等人^[75]提出基于图像对偏好的感知图像误差评估方法, 通过图像对学习(Pairwise-Learning)框架训练深度学习模型来预测对两幅失真图像的偏好. 2020年, Ding等人^[76]提出结构相似和纹理相似性统一(Deep Image Structure and Texture Similarity, DISTs)的评价指标, DISTs基于预训练的VGG网络, 将图像非线性变换为多尺度过完备表示, 然后通过距离度量来进行质量评估.

综上所述, 传统知识驱动的方法和近年来数据驱动的方法在FR-IQA研究上都取得了一定的进展. 传统方法利用先验知识, 手工提取特征来构建评价模型, 该方法本质上是将一个函数拟合到人的主观评判. 由于HVS的复杂性, 该方法精确建模HVS是非常困难的. 数据驱动的深度学习方法较之传统方法, 最大的特点在于无需利用先验知识, 可自动地学习特征, 直接从标注数据中学习评价模型. 该方法一方面“黑箱化”过程解释性差, 另一方面需要大量的训练数据, 目前将图像采样为块来进行数据增广的方法还存在问题.

3 现有研究存在的问题

图像质量评价研究虽取得了一定的进展,但仍存在许多需进一步讨论的问题.一般性地,FR-IQA中局部特征相似性计算方法是否能够模拟人对两幅图像的比较过程?FR-IQA中将局部质量综合为整体质量评分的池化方法,SSIM、ESSIM等算法采用算术平均,GMSD采用标准差,IW-SSIM、FSIM、VSI等采用不同视觉显著性度量进行加权,究竟如何有效地池化,还有待深入研究图像局部质量与整体质量之间的关系.主客观一致性度量指标中,如SROCC仅体现了排序的一致性,任一分量的轻微扰动将都会影响到评价结果.

关于现有图像质量评价研究中存在的问题,本文将重点从IQA问题构建的合理性、建模的全面性问题、知识驱动和数据驱动结合的问题三方面进行阐述.

3.1 问题构建的合理性

IQA研究的任务是对图像的视觉质量级别做出判定,属于判别问题,但又与经典的判别问题大不相同.经典的判别问题,如人脸识别和指纹识别,判别的目标是客观存在、确定无疑的.在人脸识别中,一幅人脸图像到底对应哪个人的脸,是确定无疑的;在指纹识别中,一幅指纹图像到底对应哪个人的指纹,同样是确定无疑的.但IQA判别的目标是图像的主观质量,而主观质量到底应该如何对其进行描述本身是值得研究的问题.

事实上,主观图像质量评价具有不确定性,这种不确定性包括随机性和模糊性.

(1) 随机性

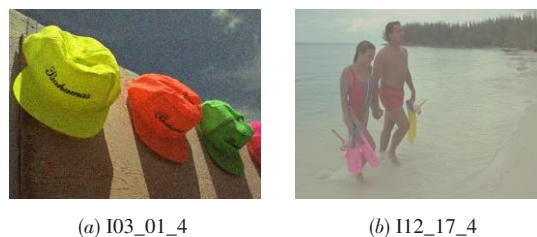
随机性指受观测者年龄、性别、知识结构、审美情趣等个体因素影响,不同观测者对同一幅图像可能会给予不同的评价;另外,同一个观测者在不同时间对同一幅图像也可能给予不同的评价.所幸的是,这种随机性背后呈现一定的统计规律性^[18,19].主观评分试验表明,对同一幅图像搜集的多个主观评分相对集中分布,且方差较小.因此取样本均值作为最终的主观评价,称为MOS值.目前图像质量评价领域通用的数据集LIVE、TID2013、TID2008、CSIQ、IVC和A57中都采用这种方式来获得MOS值.这种处理方式一定程度上消除了由随机性带来的不确定性.

(2) 模糊性

模糊性指“对于自己的主观感受,人类习惯于模糊的语义思维而不是准确的数量思维”.人类对图像质量进行评价时,并不会在实数轴上标定一个准确的实数作为评分,而只能给出类似“非常好、好、一般、较差、差”这样的语义评价,而且此时的“好”与彼时的“好”意义未必一样.这种语义变量之间的边界是模糊的,存在

“过渡区域”.

目前IQA研究采用的测试数据集在构建过程中,均邀请多个观测者对图像质量进行打分,然后将一幅图像获得的多个主观评分值“处理”成一个实数(MOS值或DMOS值).这种处理方式虽在一定程度上消除了主观的随机性,但也存在两方面的不足.一方面,未能充分挖掘一幅图像获得的多个评分值的信息,如TID2008中每幅图像获得了838个主观评分值,但仅取均值作为MOS值.838个主观评分值的具体分布包含丰富的信息,而均值无法完全体现这些信息.另一方面,最终采用实数(MOS值)来描述主观评分不能有效刻画主观判断的模糊性特点.如图2所示,TID2008中I03_01_4的MOS值为4.2414,而I12_17_4的MOS值为4.4848.“人类是否能够准确感知I03_01_4的感知质量是4.2414?”“能否准确感知两幅图像的质量只差0.2434?”“人类能否准确感知I03_01_4质量为4.2414,而I12_17_4的质量为4.4848?”显然,这种描述不能体现主观评价的模糊性特点,难以令人信服.



(a) I03_01_4 (b) I12_17_4

图2 TID2008中两幅失真图像

此外,目前广泛采用的主客观一致性度量在数值上缺乏鲁棒性.由SROCC的定义可以看出,SROCC反映了主客观评分之间“序”的一致性程度.客观评分序列 $\{y_1, y_2, \dots, y_n\}$ 中任何一个分量 y_i 的轻微扰动都可能改变序,从而对最终计算的SROCC值造成一定影响.因此,SROCC在数值上缺乏鲁棒性.

3.2 建模的全面性问题

目前已发表的大部分有关FR-IQA的研究都基于对HVS感知特点的认识来构造图像质量指标.如文献[44,55]利用了HVS的各向异性特点,文献[10]利用了多尺度特点.问题在于HVS至少同时具备这些特点,而不是在特定条件下具备其中某一个特点.因此,从全面模拟HVS的角度,图像质量评价研究应当同时利用目前所知的关于HVS的所有特点.

系统分析的公理化建模方法中,对于一个系统 $g = T(\mathcal{J})$ 假设已知关于系统的若干性质 $P_1, P_2, P_3, \dots, P_N$,且这些性质之间两两互不包含.那么已知的性质越多(内涵越大),则系统越确定(外延越小).当已知的性质足

够多时(N 充分大),则系统 $g=T(f)$ 完全由这些性质唯一确定.由这些性质出发,可推导出系统的具体形式(比如可能是某个常微分或偏微分方程).

IQA 研究的核心是对 HVS 的认识和建模,目前 IQA 方法大多仅考虑了 HVS 的某一特点, GMSD 算法提取图像梯度特征建模, ESSIM 算法提取图像边界强度特征建模.从以上系统分析的公理化建模方法中得到启示,为了尽可能全面准确地对 HVS 进行建模,须利用目前所知 HVS 的所有性质.

3.3 知识驱动和数据驱动结合的问题

现有的 FR-IQA 方法,大体可以分为知识驱动的方法和数据驱动的方法.知识驱动的方法(如文献[6,32~40])基于对 HVS 感知特点的认识或假定,启发式地构造图像质量指标,属于“自上而下”的方法;数据驱动的方法(以深度学习方法为代表^[73-75])基于标注数据“黑箱化”地学习图像质量模型,属于“自下而上”的方法,如图 3 所示.

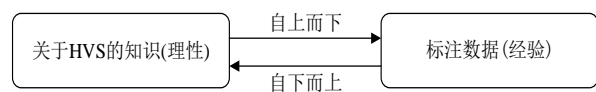


图 3 IQA 研究中可用的两类信息

知识驱动的方法利用了人类关于 HVS 的知识,但未能充分利用标注数据;而数据驱动的方法利用了标注数据,但未能充分利用人类关于 HVS 的知识.可以看出,两种方法分别利用了 IQA 研究中不同类的信息,二者各具优点.从尽可能利用所有可用信息的角度,如何将两种方法结合起来以充分发挥各自的优点还未得到进一步的研究.如:HVS 感知特点的先验知识可否迁移至深度网络中以增强其可解释性;深度网络的特征表示方法可否为构造更有效的先验模型提供启示等.如何充分利用 IQA 研究中可用的两类信息来建模,还有必要探究知识与数据驱动相结合的新方法.

4 全参考图像质量评价研究展望

图像质量评价研究的目的在于模拟人对图像质量的主观评价过程,构建同主观评价尽可能一致的客观图像质量指标,因此对主观评价过程的认识至关重要.本文通过对主观评价过程的深入分析形成了以下认识:主观图像质量评价是感知与理解基础上对图像视觉质量的判定,是高级的、涉及图像语义理解的过程;评价过程可划分为感知、理解和视觉质量评分三个阶段,其中感知与理解阶段具有非线性、非局部、各向异性、多尺度和稀疏性特点,而评分阶段具有模糊性特点.在此,简要解释 HVS 这几个特点.

(1)非线性特点. HVS 对图像的亮度、色度、对比

度、锐度等客观刺激的响应是非线性的.

(2)非局部特点.这一特点同语义对象在自然图像中的呈现方法有关.当人们观测一幅图像时,会自然地感知到其中的语义对象,而不是对局部结构进行逐一考察.

(3)各向异性特点.各向异性的概念源自热物理学,在均匀介质中某一点 P_0 处放置一个热源,则以 P_0 为中心的各个方向上热传导性质完全相同,热扩散具有各项同性特点.反之,在非均匀介质中某一点 P_0 处放置一个热源,则以 P_0 为中心的各个方向上的热传导性质并不相同,热扩散具有各向异性特点.通常,自然图像在空域某点为中心的不同方向上具有不同的正则性、曲率等性质,即自然图像具有局部各向异性结构,而 HVS 能够感知这种各向异性结构. HVS 具备各向异性结构感知能力是其具备语义理解能力的基础.

(4)多尺度特点. HVS 具有自适应多尺度聚焦能力,观测一幅图像时, HVS 自适应调节,对感兴趣区域使用小尺度,从而感知到更多细节;而对于其他区域,使用较大尺度,忽略掉细节,只感知到大概轮廓.

(5)稀疏性特点. HVS 对于图像的感知和理解过程是一个稀疏表示的过程.稀疏表示通常包含两个步骤:字典学习和稀疏编码.字典学习从样本中提取共性的结构,获得表示数据的基本原子;稀疏编码从字典中选取与样本相关性最强的少数原子来表示数据. HVS 具备类似的特点:首先,人类具有抽象概括能力,所谓抽象概括即提取样本的共性,这类似于字典学习;其次,当人类观测到一个对象时,会自然地所有记忆单元中触发与该对象相关性最强的记忆单元,这类似于稀疏编码.

(6)模糊性特点.关于自己的主观体验,人类更习惯于模糊的语义思维而不是准确的数量思维.比如:人们习惯于“我很开心”“我感觉天气很热”“我特别想念你”这样的表述,而不习惯于“我有 0.81 的开心”“我感觉到 0.87 的热度”“我对你有 0.93 的想念”这样的表述.人们难以对自己的主观感受进行准确量化,只能用语义表达模糊的范围.人们对于图像质量的主观评分也是如此.

值得注意的是,这里非线性、非局部、各向异性、多尺度和稀疏性是感知和理解阶段体现出的特点,而模糊性是感知和理解基础之上做出主观判决时体现出的特点.其中非局部、各向异性、多尺度和稀疏性特点是紧密地耦合在一起的.

基于以上对主观评价过程的认识,即关于 HVS 定性的知识,结合本文第 3 节中对 IQA 现有研究存在问题的分析,在此梳理 IQA 未来可研究的两个方向——主观评分模糊建模、知识与数据联合驱动的图像质量

评价。

4.1 主观评分的模糊建模

基于3.1节分析,IQA问题构建中存在两个问题:实数描述图像质量主观评分不能体现人的模糊语义思维;主客观一致性度量缺乏鲁棒性。那么应该如何对主观判断的模糊性进行描述呢?采用“模糊数”描述主观评分是否可行?

模糊数学建立在模糊集合论的基础之上。1965年扎德(Zadeh)对经典的集合论进行了推广,提出了模糊集合论的公理化体系。经典集合论中,集合及其运算同特征函数及其运算一一对应;模糊集合论中,模糊集及其运算同隶属度函数及其运算一一对应。隶属度函数“软化”了集合的边界,突破了经典集合论中元素与集合确定的二元关系(要么属于,要么不属于),刻画了元素隶属于集合的程度。模糊集合论基础之上可建立模糊测度,而模糊测度基础之上可建立模糊积分。工程应用中,模糊数学体现出以下两大优势:一方面,能很好地模拟人类的智能,对边界不清的主观判断进行更加准确的建模;另一方面,模糊计算在数值上具有较好的鲁棒性。

基于以上分析,IQA可研究的方向之一为,基于模糊建模,对图像质量主观评分进行更加令人信服的描述,以及构建鲁棒性更好的主客观一致性度量,从而建立新的IQA算法测试基准以改进IQA问题构建的合理性。具体研究问题有数据集的构建、主观评分试验设计、模糊数类型的选择、主观评分的模糊数刻画、模糊数序列与实数序列之间或两个模糊数序列之间的一致性度量等。

4.2 知识与数据联合驱动下的图像质量评价

由3.2节中建模全面性分析结合本节中对主观评价的认识可知,要实现HVS的全面建模,要同时利用HVS的六个特点来构造IQA指标,但采用传统方法很难在一个数学模型中同时体现这六个特点。本文发现,深度学习的网络结构具备与HVS完全对应的非线性、非局部、各向异性、多尺度、稀疏性等特点。

(1)非线性特点。深度学习(或者更一般的神经网络)最显著的特点就是非线性——这也是为何要引入神经网络的原因。大家共识的是无论有多深的网络,如果只是线性变换,那么它的表示最终都可以归纳到一个一层网络,它的表达能力显然是有限的。

(2)非局部特点。在深度学习里,局部性和非局部性往往同时使用,例如最典型的代表——卷积神经网络(CNN),其底部的层都使用支撑集比较小(例如 5×5 , 15×15 , 33×33)的滤波器对输入图像进行卷积(卷积核在图像不同区域共享参数),这就对应了CNN的局部性,此特点符合人眼的视野范围限制(感受野)——对

视觉系统的低层机能,即人眼观察一个图像,首先像一个滤波器对图像抽取低层的边缘、纹理等特征,并且这个滤波器在图像不同的区域工作原理是相同的;同时,随着层数增加,图像的分辨率不断降低(使用了pooling降低分辨率,保障网络规模不至于增长太快),对应滤波器覆盖范围不断变大——从局部到全局/非局部扩展,并且CNN网络的最后几层为全连接层,因而保障了最终的神经网络具有全局性/非局部性,这个时候的视觉系统已经从低层的局部非语义信息采集逐渐上升到可以获得图像的语义信息了,而这正好是后面分类器所需要的。

(3)多尺度特点。从CNN网络对图像维度和滤波器数目改变以及网络对图像不同level语义信息的表达能力来看(如人脸识别的应用),深度学习网络具有多尺度特点,正是这种多尺度特点完成了对图像不同层次语义的抽取。

(4)各向异性特点。各向异性表现在深度学习网络的参数(包括CNN网络中滤波器系数)都是自适应训练的,没有对称性等条件约束,大量样本数据的各项异性特点导致最终的神经网络模型也是各向异性的。

(5)稀疏性特点。为了稳定性,通常会采用dropout、舍弃微小的参数trick等方法,这些方面使得网络具有稀疏性的特点。此外,深度学习稀疏性还体现在对于神经网络的压缩——压缩网络参数,有利于减小计算量,降低过拟合。例如:Han等人^[77]提出的PQH(Pruning, Quantization and Huffman)算法,在经典的LeNet, AlexNet和VGG网络上压缩率为40~50倍而无明显的性能损失;Sun等人^[78]提出稀疏化深度卷积网络,在人脸识别网络DeepID获得10倍左右的压缩率且性能相当。

(6)模糊性特点。深度网络包含回归(regression)步骤,而这一过程可以“模糊化”。以IQA为例,当MOS值为模糊数时,则需要使用模糊回归。此时网络的训练需要借助模糊优化,计算复杂度有一定提高,但模型的鲁棒性也可能随之改善。也就是说,深度网络结构提供了模糊化“接口”。

值得注意的是,这里非线性、非局部、各向异性、多尺度和稀疏性是特征学习(feature learning)阶段体现出的特点,对应人类的感知和理解。模糊性是基于特征进行分类、回归时体现出的特点,对应人类的主观判决。同样,深度网络结构的非局部、各向异性、多尺度和稀疏性特点是紧密地耦合在一起的。

基于以上对深度网络特点与HVS感知特点之间关系的分析,IQA可研究的方向之一为,将先验知识迁移至深度网络中,利用关于HVS六个特点的先验知识(模型)来设计深度网络结构,借助已标注数据来学习图像质量模型。一方面该研究,较之已有的基于HVS部分

特点的方法,可实现对HVS更加全面准确的建模.另一方面,所研究的图像质量模型有以下特点:①既利用了关于人类视觉认知的先验知识(模型),又利用标注数据信息,实现知识与数据联合驱动;②基于知识构建网络结构,使得网络结构具有明确的物理意义,不再是“黑箱化”的;③基于先验知识和先验模型构建的网络由于具备某些先验结构,因此使其参数减少、自由度下降,同时网络的表示能力也会下降,但是针对特定应用,赋予网络先验结构一定会影响其应用中的性能,而且会减少对训练样本的需求.

5 总结

本文的工作主要包含三个方面:第一,针对全参考图像质量评价的研究,从问题构建和评价算法等方面进行了回顾;第二,在回顾的基础上,分析了现有研究中存在的问题,重点阐述问题构建的合理性、建模的全面性问题、知识驱动与数据驱动结合的问题;第三,结合存在的问题和主观评价过程的深入分析,探讨了主观评分采用模糊建模、知识与数据联合驱动构建算法两个可能的研究方向,以期对同行研究者提供参考.

参考文献

- [1] Chandler D M. Seven challenges in image quality assessment: Past, present, and future research[J]. *ISRN Signal Processing*, 2013, 2013: 1 – 53.
- [2] Mohammadi P, Ebrahimi-Moghadam A, Shirani S. Subjective and objective quality assessment of image: A survey [J]. *Majlesi Journal of Electrical Engineering*, 2014, 9(1): 419 – 423.
- [3] Mannos J, Sakrison D. The effects of a visual fidelity criterion of the encoding of images[J]. *IEEE Transactions on Information Theory*, 1974, 20(4): 525 – 536.
- [4] Buades A, Coll B, Morel J M. A review of image denoising algorithms, with a new one[J]. *Multiscale Modeling & Simulation*, 2005, 4(2): 490 – 530.
- [5] 李亮亮. 基于非下采样剪切波变换的图像增强算法研究[D]. 长春: 吉林大学, 2019.
Li L L. The research of image enhancement algorithm based on nonsubsampling shearlet transform[D]. Changchun, China: Jilin University, 2019. (in Chinese)
- [6] 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.
- [7] 沈瑜, 陈小朋, 刘成, 等. 基于混合模型驱动的红外与可见光图像融合[J]. *控制与决策*, 2021, 36(9): 2143 – 2151.
- [8] Shen Y, Chen X P, Liu C, et al. Infrared and visible image fusion based on hybrid model driving[J]. *Control and Decision*, 2021, 36(9): 2143 – 2151. (in Chinese)
- [9] Ding K Y, Ma K D, Wang S Q, et al. Comparison of full-reference image quality models for optimization of image processing systems[J]. *International Journal of Computer Vision*, 2021, 129(4): 1258 – 1281.
- [10] Zhao H, Gallo O, Frosio I, et al. Loss functions for image restoration with neural networks[J]. *IEEE Transactions on Computational Imaging*, 2017, 3(1): 47 – 57.
- [11] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[A]. *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*[C]. Pacific Grove, CA, USA: IEEE, 2003. 1398 – 1402.
- [12] Laparra V, Ballé J, Berardino A, et al. Perceptual image quality assessment using a normalized Laplacian pyramid [J]. *Electronic Imaging*, 2016, 2016(16): 1 – 6.
- [13] Laparra V, Berardino A, Ballé J, et al. Perceptually optimized image rendering[J]. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 2017, 34(9): 1511 – 1525.
- [14] Ma J J, Nakarmi U, Kin C Y S, et al. Diagnostic image quality assessment and classification in medical imaging: Opportunities and challenges[A]. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*[C]. Iowa City, IA, USA: IEEE, 2020. 337 – 340.
- [15] 郭从洲, 李可, 李贺, 等. 遥感图像质量等级分类的深度卷积神经网络方法[J/OL]. *武汉大学学报(信息科学版)*, 2020, DOI:10.13203/j.whugis20200292.
GUO C Z, LI K, LI H, et al. A deep convolutional neural network method for remote sensing image quality classification [J/OL]. *Geomatics and Information Science of Wuhan University*, 2020, DOI: 10.13203/j.whugis20200292. (in Chinese)
- [16] 李毅红, 韩焱, 潘晋孝, 等. 基于递变能量线性约束的X射线图像质量评价方法[J]. *电子学报*, 2017, 45(3): 669 – 673.
Li Y H, Han Y, Pan J X, et al. X-ray image quality evaluation based on linear constraint with variable energy[J]. *Acta Electronica Sinica*, 2017, 45(3): 669 – 673. (in Chinese)
- [17] Sheikh H R, Wang Z, Bovik A C, et al. Image and video quality assessment research at LIVE[EB/OL]. <http://live.ece.utexas.edu/research/quality>, 2020.
- [18] Larson E C, Chandler D M. Categorical subjective image

- quality database[EB/OL].<http://vision.okstate.edu/csiq>,2020.
- [18] Ponomarenko N, Egiazarian K. Tampere image database TID2008[EB/OL]. <http://www.ponomarenko.info/tid2008>, 2020.
- [19] Ponomarenko N, Jin L N, Ieremeiev O, et al. Image database TID2013: Peculiarities, results and perspectives[J]. *Signal Processing: Image Communication*, 2015, 30: 57 – 77.
- [20] Ninassi A, Callet P L, Atrousseau F. Subjective quality assessment IVC database 2005[EB/OL]. <http://www2.irc-cyn.ecnantes.fr/ivcdb>, 2020.
- [21] Chandler D M, Hemami S S. A57 Database 2007[EB/OL]. <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>. A57, 2020.
- [22] Ma K D, Duanmu Z F, Wu Q B, et al. Waterloo exploration database: New challenges for image quality assessment models[J]. *IEEE Transactions on Image Processing*, 2017, 26(2): 1004 – 1016.
- [23] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[A]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City, UT, USA: IEEE, 2018. 586 – 595.
- [24] Lin H H, Hosu V, Saupé D. KADID-10k: A large-scale artificially distorted IQA database[A]. 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)[C]. Berlin, Germany: IEEE, 2019. 1 – 3.
- [25] Wang Z, Simoncelli E P. Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities[J]. *Journal of Vision*, 2008, 8(12): 1 – 8.
- [26] Ma K D, Duanmu Z F, Wang Z, et al. Group maximum differentiation competition: Model comparison with few samples[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(4): 851 – 864.
- [27] Sheikh H R, Bovik A C, de Veciana G. An information fidelity criterion for image quality assessment using natural scene statistics[J]. *IEEE Transactions on Image Processing*, 2005, 14(12): 2117 – 2128.
- [28] Sheikh H R, Bovik A C. Image information and visual quality[J]. *IEEE Transactions on Image Processing*, 2006, 15(2): 430 – 444.
- [29] 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.
- [30] Shnayderman A, Gusev A, Eskicioglu A M. An SVD-based grayscale image quality measure for local and global assessment[J]. *IEEE Transactions on Image Processing*, 2006, 15(2): 422 – 429.
- [31] Wang Z, Shang X L. Spatial pooling strategies for perceptual image quality assessment[A]. 2006 International Conference on Image Processing[C]. Atlanta, GA, USA: IEEE, 2006. 2945 – 2948.
- [32] Chen G H, Yang C L, Xie S L. Gradient-based structural similarity for image quality assessment[A]. 2006 International Conference on Image Processing[C]. Atlanta, GA, USA: IEEE, 2006. 2929 – 2932.
- [33] Chandler D M, Hemami S S. VSNR: A wavelet-based visual signal-to-noise ratio for natural images[J]. *IEEE Transactions on Image Processing*, 2007, 16(9): 2284 – 2298.
- [34] Sampat M P, Wang Z, Gupta S, et al. Complex wavelet structural similarity: A new image similarity index[J]. *IEEE Transactions on Image Processing*, 2009, 18(11): 2385 – 2401.
- [35] Moorthy A K, Bovik A C. Visual importance pooling for image quality assessment[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2009, 3(2): 193 – 201.
- [36] 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.
- [37] Zhang L, Zhang L, Mou X Q. RFSIM: A feature based image quality assessment metric using Riesz transforms [A]. 2010 IEEE International Conference on Image Processing[C]. Hong Kong, China: IEEE, 2010. 321 – 324.
- [38] Wang Z, Li Q. Information content weighting for perceptual image quality assessment[J]. *IEEE Transactions on Image Processing*, 2011, 20(5): 1185 – 1198.
- [39] 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.
- [40] Liu A M, Lin W S, Narwaria M. Image quality assessment based on gradient similarity[J]. *IEEE Transactions on Image Processing*, 2012, 21(4): 1500 – 1512.
- [41] Zhu J Y, Wang N C. Image quality assessment by visual gradient similarity[J]. *IEEE Transactions on Image Processing*, 2012, 21(3): 919 – 933.
- [42] Capodiferro L, Jacovitti G, Di Claudio E D. Two-dimensional approach to full-reference image quality assessment based on positional structural information[J]. *IEEE*

- Transactions on Image Processing, 2012, 21(2): 505 – 516.
- [43] 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.
- [44] Zhang X D, Feng X C, Wang W W, et al. Edge strength similarity for image quality assessment[J]. IEEE Signal Processing Letters, 2013, 20(4): 319 – 322.
- [45] Wu J J, Lin W S, Shi G M, et al. Perceptual quality metric with internal generative mechanism[J]. IEEE Transactions on Image Processing, 2013, 22(1): 43 – 54.
- [46] Chang H W, Yang H, Gan Y, et al. Sparse feature fidelity for perceptual image quality assessment[J]. IEEE Transactions on Image Processing, 2013, 22(10): 4007 – 4018.
- [47] Zhang L, Shen Y, Li H Y. VSI: A visual saliency-induced index for perceptual image quality assessment[J]. IEEE Transactions on Image Processing, 2014, 23(10): 4270 – 4281.
- [48] 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.
- [49] Pei S C, Chen L H. Image quality assessment using human visual DOG model fused with random forest[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3282 – 3292.
- [50] Lee D, Plataniotis K N. Towards a full-reference quality assessment for color images using directional statistics[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3950 – 3965.
- [51] Nafchi H Z, Shahkolaei A, Hedjam R, et al. Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator[J]. IEEE Access, 2016, 4: 5579 – 5590.
- [52] 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.
- [53] Reisenhofer R, Bosse S, Kutyniok G, et al. A Haar wavelet-based perceptual similarity index for image quality assessment[J]. Signal Processing: Image Communication, 2018, 61: 33 – 43.
- [54] Li L D, Cai H, Zhang Y B, et al. Sparse representation-based image quality index with adaptive sub-dictionaries [J]. IEEE Transactions on Image Processing, 2016, 25(8): 3775 – 3786.
- [55] Ding L, Huang H, Zang Y. Image quality assessment using directional anisotropy structure measurement[J]. IEEE Transactions on Image Processing, 2017, 26(4): 1799 – 1809.
- [56] Gu K, Li L D, Lu H, et al. A fast reliable image quality predictor by fusing micro- and macro-structures[J]. IEEE Transactions on Industrial Electronics, 2017, 64(5): 3903 – 3912.
- [57] 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.
- [58] Ahar A, Barri A, Schelkens P. From sparse coding significance to perceptual quality: A new approach for image quality assessment[J]. IEEE Transactions on Image Processing, 2018, 27(2): 879 – 893.
- [59] Di Claudio E D, Jacovitti G. A detail-based method for linear full reference image quality prediction[J]. IEEE Transactions on Image Processing, 2018, 27(1): 179 – 193.
- [60] Jia H Z, Zhang L, Wang T H. Contrast and visual saliency similarity-induced index for assessing image quality [J]. IEEE Access, 2018, 6: 65885 – 65893.
- [61] Temel D, AlRegib G. Perceptual image quality assessment through spectral analysis of error representations[J]. Signal Processing: Image Communication, 2019, 70: 37 – 46.
- [62] Kim W, Nguyen A D, Lee S, et al. Dynamic receptive field generation for full-reference image quality assessment[J]. IEEE Transactions on Image Processing, 2020, 29: 4219 – 4231.
- [63] Ling W Y, Hu Y. Machine learning to design full-reference image quality assessment algorithm[J]. TELKOMNIKA Indonesian Journal of Electrical Engineering, 2013, 11(6): 3439 – 3444.
- [64] Narwaria M, Lin W S. 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.
- [65] Liu T J, Lin W S, Kuo C C J. Image quality assessment using multi-method fusion[J]. IEEE Transactions on Image Processing, 2013, 22(5): 1793 – 1807.
- [66] Gastaldo P, Zunino R, Redi J. Supporting visual quality assessment with machine learning[J]. EURASIP Journal on Image and Video Processing, 2013, 2013(1): 1 – 15.
- [67] Guha T, Nezhadarya E, Ward R K. Learning sparse models for image quality assessment[A]. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing

- (ICASSP)[C]. Florence, Italy: IEEE, 2014. 151 – 155.
- [68] Liang Y D, Wang J J, Wan X Y, et al. Image quality assessment using similar scene as reference[A]. Computer Vision – ECCV 2016[C]. Cham, GER: Springer International Publishing, 2016. 3 – 18.
- [69] Liu T J, Liu K H, Lin J Y, et al. A ParaBoost method to image quality assessment[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(1): 107 – 121.
- [70] Gao F, Wang Y, Li P P, et al. DeepSim: Deep similarity for image quality assessment[J]. Neurocomputing, 2017, 257: 104 – 114.
- [71] Wang S G, Deng C W, Lin W S, et al. NMF-based image quality assessment using extreme learning machine[J]. IEEE Transactions on Cybernetics, 2017, 47(1): 232 – 243.
- [72] Kim J, Lee S. Deep learning of human visual sensitivity in image quality assessment framework[A]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Florence, Italy: IEEE, 2017. 1969 – 1977.
- [73] Bosse S, Maniry D, Müller K R, et al. Deep neural networks for no-reference and full-reference image quality assessment[J]. IEEE Transactions on Image Processing, 2017, 27(1): 206 – 219.
- [74] Ma K D, Duanmu Z F, Wang Z. Geometric transformation invariant image quality assessment using convolutional neural networks[A]. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. Calgary, AB, Canada: IEEE, 2018. 6732 – 6736.
- [75] Prashnani E, Cai H, Mostofi Y, et al. PieAPP: Perceptual image-error assessment through pairwise preference[A]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City, UT, USA: IEEE, 2018. 1808 – 1817.
- [76] Ding K Y, Ma K D, Wang S Q, et al. Image quality assessment: Unifying structure and texture similarity[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 58(10), PP(99): 1.
- [77] Han S, Mao H, Dally W. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding[EB/OL].<https://arxiv.org/abs/1510.00149v4>, 2020.
- [78] Sun Y, Wang X G, Tang X O. Sparsifying neural network

connections for face recognition[A]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Las Vegas, NV, USA: IEEE, 2016. 4856 – 4864.

作者简介



高敏娟 女,1984年生,山西忻州人.博士研究生,主要研究方向为图像处理、图像质量评价.
E-mail:gaominjuan1984@163.com



党宏社 男,1962年生,陕西武功人.教授、博士生导师,主要研究方向为工业过程与优化、计算机控制和图像处理.
E-mail:danghs@sust.edu.cn



魏立力 男,1965年生,甘肃天水人.教授,主要研究方向为应用统计与数据分析.
E-mail:liliwei@nxu.edu.cn



刘国军 男,1978年生,宁夏吴忠人.教授,主要研究方向为小波分析、偏微分方程图像处理和图像质量评价.
E-mail:liugj@nxu.edu.cn



张选德(通讯作者) 男,1979年生,宁夏固原人.教授、博士生导师,主要研究方向为图像恢复、图像质量评价、稀疏表示和低秩逼近理论.
E-mail:zhangxuande@sust.edu.cn