

文章编号: 2095-2163(2023)10-0147-04

中图分类号: TP391.4

文献标志码: A

# 集成学习算法在图像质量评估中的应用

牟卿志, 李玉婷, 孙宗升, 周 荃

(潍坊职业学院 机电工程学院, 山东 潍坊 262737)

**摘要:** 本文基于方法融合感知的思想, 探究了集成学习算法在图像质量评估中的应用, 对利用支持向量机及神经网络搭建集成学习框架进行图像质量评估的性能表现进行了分析。为保证集成学习后评估算法的通用性, 使用保真度、结构相似性、颜色质量与数据驱动学习等多种估计器指标进行集成, 使用 LIVE、多重失真 LIVE 及 TID 2013 3 种数据库进行验证。实验结果表明, 集成学习方法通常可以提高图像质量评估能力, 基于深度学习网络的评估能力增强优于基于支持向量机的增强, 当存在两种以上附加方法进行融合感知时, 二者较原有最佳性能方法均可取得较为有效地提升。

**关键词:** 图像质量评估; 多方法融合; 集成学习

## Application of ensemble learning algorithm in image quality assessment

MOU Qingzhi, LI Yuting, SUN Zongsheng, ZHOU Quan

(School of Mechanical and Electrical Engineering, Weifang Vocational College, Weifang Shandong 262737, China)

**[Abstract]** Based on the idea of method fusion perception, this paper explores the application of ensemble learning algorithm in image quality assessment, and analyzes the performance of an ensemble learning framework for image quality assessment combining support vector machine and neural network. In order to ensure the universality of the post evaluation algorithm of ensemble learning, this paper uses a variety of estimator indicators such as fidelity, structural similarity, color quality and data-driven learning. In the experiment, the LIVE, the multiply distorted LIVE, and the TID 2013 databases are used for verification. The experimental results show that ensemble learning methods can generally improve the image quality evaluation ability. The enhancement of evaluation ability based on deep learning network is better than that based on support vector machine. When there are more than two additional methods for fusion perception, both of them can achieve better effect than the original best performance method.

**[Key words]** image quality assessment; multi-method fusion; ensemble learning

## 0 引言

由于拍摄、存储和显示技术的进步, 图像信息交互已经成为在线媒体和社交网络的主流, 每天平均有数十亿张图片在网上共享。为了能够在有限的带宽内共享所有图像, 需要设计紧凑的压缩算法, 使图片不仅满足带宽要求, 而且能够保持适当的感知质量。然而, 海量图像数据使得人工主观评估的方法不再可行。因此, 迫切需要可靠的图像质量评估方法, 来提供用户体验基准, 以应对媒体与社交网络信息流传递的需求。

图像质量评估通常用于处理一幅或多幅图像, 按规律映射到评估器进行运算, 进而完成分数评估

工作<sup>[1]</sup>。按照映射时是否需要参考图像来看, 可分为全参考图像质量评价 (FR-IQA)、半参考图像质量评价 (RR-IQA) 和无参考图像质量评价 (NR-IQA) 3 类。其中, 无参考图像质量评价因其广泛的应用场景与较大的评价难度成为当今的研究热点<sup>[2]</sup>。在无参考图像质量评价中, 按映射规律, 评估器可分为基于保真度<sup>[3-4]</sup>、结构相似性<sup>[5-6]</sup>、颜色质量<sup>[7]</sup>与数据驱动学习<sup>[8]</sup>等多种评价指标。分析评估算法, 可知现有估计器都基于视觉系统不同方面的特征, 即使某种方法在特定感知策略上不如其他方法直接相关, 但其仍然可以包含其他方法无法提供的额外信息。因此, 融合多种方法提高整体评估能力的策略被提出, 并已成为图像质量评估工作

**作者简介:** 牟卿志(1996-), 男, 硕士, 助教, 主要研究方向: 计算机视觉、深度学习; 李玉婷(1991-), 女, 硕士, 助教, 主要研究方向: 图像质量分析与评估; 孙宗升(1990-), 男, 硕士, 工程师, 主要研究方向: 目标检测与分类; 周 荃(1981-), 女, 硕士, 副教授, 主要研究方向: 图像信息处理。

**通讯作者:** 牟卿志 Email: 920128084@qq.com

收稿日期: 2022-10-29

中的主要发展方向。随着融合与增强方法的改进,多方法融合可以产生更全面的质量估计器。目前常用多方法融合策略有基于线性组合<sup>[9]</sup>、回归<sup>[10]</sup>、集成学习策略<sup>[11]</sup>等的方法,其中基于集成学习策略的方法被证明表现优于常规融合方法。

在对算法评估能力探究,尤其是将多方法融合策略视为一个集成学习框架进行分析方面,文献[2]在 LIVE 和 TID 08 数据集对 5 种非特定失真的算法性能,结合算法运行时间情况评选出最佳方法。文献[12]则针对模糊失真图像的特性,按人工模糊与自然模糊进行了评估。本研究发现,框架不应局限于某些特定的质量指标、失真类型或学习方法。为了研究集成学习质量估计器的通用性,本文选用基于 4 种不同评价指标中的主流图像质量评估算法。在学习算法中,使用支持向量机与通用神经网络执行增强。所选用方法在 3 种最新且全面的数据库中进行多项训练和测试,最后使用基于准确性和线性进行度量,通过将融合方法与现有方法的评估能力进行比较,证明并探讨了不同融合方法在不同数据库上的评估能力。

## 1 实验设定

### 1.1 图像质量评估器选用

#### 1.1.1 基于保真度

保真度属性直白量化了评估图像相对于参考图像的变化,是图像和视频编码标准中用于衡量失真率的首选指标。峰值信噪比(PSNR)是其最为常用的像素级保真度量方法。进一步扩展指标,添加均值漂移灵敏度,定义为 PSNR-HMA(简称 HMA),这两个指标均用于失真率衡量操作。

#### 1.1.2 基于结构相似性

结构相似性是一种衡量两幅图像相似度的指标。文献[3]中提出的基于参考图像和扭曲图像在空间域中亮度、对比度和结构方面比较的完全参考度量方案-结构相似性(Structural Similarity, 简称为 SSIM),有效衡量了空间域中两幅图像结构的接近程度,是目前主流方法。

#### 1.1.3 基于颜色

在亮度稳定情况下,颜色成为人类视觉系统(HVS)的首要感知评价信息。在图像质量评估中,文献[7]所提出的 PerSIM 算法通过计算颜色-对立空间(LAB 空间)中色度通道的像素级保真度来引入颜色信息,通过视网膜细胞的对比敏感度带通特征,计算两幅图像相似性。

#### 1.1.4 基于数据驱动学习

数据驱动的质量估计器多采用无监督方式对使用的图像进行训练。图像通过多种预处理获得更具描述性的结构和颜色表示后,被送入线性解码器以获得稀疏表示,进而对稀疏表示进行比较,从而获得客观分数。数据驱动的质量估计器 UNIQUE 被证明是具备优秀表现的算法<sup>[8]</sup>。

### 1.2 集成学习算法

本文通过支持向量机和通用神经网络两种主流方法探究提升效果。在通用神经网络架构中,调整的唯一参数是单个隐藏层中的神经元数量,其被设置为实验中使用的质量估计器的总数。均方误差设置为代价函数;Levenberg-Marquardt 方法设置为训练函数。所采用支持向量机的策略中包括序列最小优化算法(Sequential Minimal Optimization, SMO)和线性核。

### 1.3 数据集

在估计器评估能力比较中,引入 LIVE<sup>[13]</sup>、多重失真 LIVE(MULTI)<sup>[14]</sup>和 TID 2013(TID13)数据库<sup>[15]</sup>。数据库中提供了参考图像及附带评价的失真图像,通过不同的失真图像及相关评价,可以有效对质量估计器的评估能力进行客观验证。

以 TID2013 数据集为例,其提供的参考图像及相关失真图像如图 1 所示。



(a) 参考图片

(b) 失真图片

图 1 TID2013 数据集中提供的参考图片与失真图片

Fig. 1 Reference and distorted images provided in the TID2013 dataset

### 1.4 数据分区和实验次数

实验中,质量估计器的评估能力通过  $k$  折交叉验证进行测量。其中,  $k$  设置为 4。在每次迭代中,每个数据库中总图像的 25% 被选为测试集,用于测量现有质量估计器的评估能力。每种方法都经过 10 次运行后取均值作为最终结果。

### 1.5 评价指标

本文采用精度和线性度两种评价指标。均方根误差(RSME)用于测量精度,而泊尔森相关系数(PLCC)用于测量预测线性度。相关公式如下:

$$RSME = \sqrt{\frac{\sum_{s=1}^N (x_s - y_s)^2}{N}} \quad (1)$$

$$PLCC = \frac{\sum_{s=1}^N (x_s - \mu_x)(y_s - \mu_y)}{\sqrt{\sum_{s=1}^N (x_s - \mu_x)^2} \cdot \sqrt{\sum_{s=1}^N (y_s - \mu_y)^2}} \quad (2)$$

其中,  $s$  为索引图像序号;  $x_s$  是评价分数;  $y_s$  是参考分数;  $\mu$  是图像均值;  $N$  是图像总数。

表 1 现有 IQA 方法的评估能力

Tab. 1 Evaluation capability of existing IQA methods

	均方根误差 ( $RMSE$ )					泊尔森相关系数 ( $PLCC$ )				
	PSNR	HMA	SSIM	PerSIM	UNIQUE	PSNR	HMA	SSIM	PerSIM	UNIQUE
LIVE	14.63	12.14	11.5	9.37	7.88	0.811	0.831	0.844	0.895	0.903
MULTI	12.85	9.14	10.81	8.93	7.94	0.795	0.862	0.847	0.883	0.886
TID13	1.35	1.12	1.26	0.85	0.88	0.807	0.824	0.828	0.861	0.857

## 2.2 集成学习后图像质量估计器评估能力变化

本节探究了在集成学习算法的弱势方法中添加新方法所导致的相对评估能力变化,以及现有方法与集成学习后的图像质量评估方法的比较。在集成学习时,每个类别均由表现最差的方法作为基础,加入所有方法列表中未使用过的最佳方法进行集成学习(该方法此时状态将变更为已使用),进行估计器能力评估,进而依次循环,继续进行方法列表下一个未使用过最佳方法的集成,直至所有方法全部集成为止。初始最弱势方法为表 1 中所列的 PSNR 方法,其  $RSME$  值为 14.63,  $PLCC$  值为 0.811。

实验结果表明,随着集成学习方法数量的增加,均方根误差 ( $RSME$ ) 呈现递减趋势,而泊尔森相关系数 ( $PLCC$ ) 增加,图像质量评估的准确性不断提升,线性度则不断优化。仅在少数例外情况下,泊尔森相关系数 ( $PLCC$ ) 在集成学习方法的数量方面出现线性度反向变化的行为。实验结果在 LIVE、MULTI 及 TID13 数据集的整体变化基本趋同。在 LIVE 数据集运行的标准误差线柱状图如图 2、图 3 所示。其中,主轴代表相应指标均值,误差线代表标准差。

在准确性及线性度分析方面,以 LIVE 数据集为例可以发现,集成学习方法通常可以提高图像质量评估能力。

集成单种方法时,基于神经网络集成学习的方式,使估计器评估能力在准确度与线性度方面分别

## 2 实验结果与分析

### 2.1 现有质量估计器的评估能力

现有质量估计器的评估能力见表 1。就准确性(均方根误差  $RMSE$ ) 和线性度(泊尔森相关系数  $PLCC$ ) 而言,最佳执行方法是 LIVE、MULTI 数据库中的 UNIQUE 和 TID13 数据库中的 PerSIM 方法。就整体评估能力来看,UNIQUE 估计器的评估误差较小,是评估能力最强的。

提升 25.52% 与 4.81%; 而使用支持向量机集成学习的质量估计器的评估能力提升不足,准确度与线性度提升能力仅为 7.17% 与 0.42%。由此可见,神经网络集成学习的方式对集成方法带来的误差减小较为敏感,可显著在原方法上完成优势信息补充。

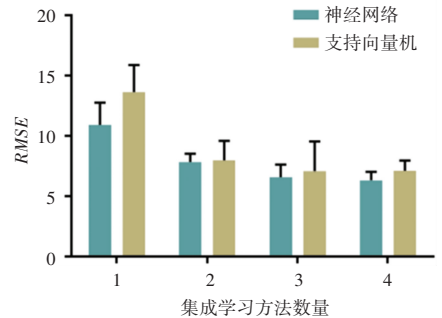


图 2 集成学习后估计器的均方根误差

Fig. 2  $RSME$  of estimator after ensemble learning

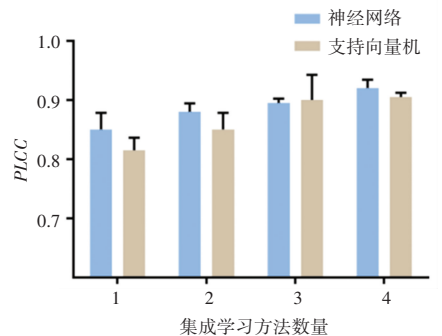


图 3 集成学习后估计器的泊尔森相关系数

Fig. 3  $PLCC$  of estimator after ensemble learning

集成两种及以上方法时,神经网络与支持向量机集成学习策略均取得显著的评估能力优化,在LIVE、MULTI、TID13数据集运行结果见表2。

因全部集成学习完成后的方案将所有图像质量评估算法都已融合在内,因此在进行评价时,选择将集成学习算法与原来的单一最佳方法进行对比,以验证多方法集成增强的有效性。在3种数据库权重

一致的情况下,神经网络集成算法较原最优方法的准确度与线性度分别提升16.68%与2.52%,而支持向量机的准确度与线性度提升9.01%与1.10%,两者均取得了较为有效地提升。考虑深度学习网络集成学习参数设置的便捷性,且其评估能力增强亦优于基于支持向量机的增强,因此该方法具有更强的适用性与推广性。

表2 现有与集成学习后的图像质量评估方法比较

Tab. 2 Comparison of existing and ensemble learning image quality assessment methods

	均方根误差 (RMSE)			泊尔森相关系数 (PLCC)		
	现有最佳方法	神经网络增强方法	支持向量机增强方法	现有最佳方法	神经网络增强方法	支持向量机增强方法
LIVE	7.88	6.31	7.15	0.903	0.924	0.905
MULTI	7.94	6.87	7.16	0.886	0.907	0.891
TID13	0.85	0.71	0.85	0.861	0.886	0.883

### 3 结束语

本文分析了基于多方法融合集成学习框架对图像质量提升的影响。实验结果表明,基于集成学习方法的评估能力优于现有的最佳方法,可以证明集成学习通常会提高图像质量评估算法的评估能力,增强级别取决于集成学习策略的类型。此外,使用两种或两种以上的附加方法增强评估能力最差的质量估计器,可以在所有场景中获得统计显著的改进,而不依赖于特定集成学习策略。

### 参考文献

[1] 张淑芳,张聪,张涛,等. 通用型无参考图像质量评价算法综述[J]. 计算机工程与应用,2015,51(19): 13-23,151.

[2] 王志明. 无参考图像质量评价综述[J]. 自动化学报,2015,41(6):1062-1079.

[3] EGIAZARIAN K, ASTOLA J, PONOMARENKO N, et al. New full-reference quality metrics based on HVS[C]//Proceedings of the second international workshop on video processing and quality metrics. 2006: 4.

[4] 岳广辉. 无参考图像质量评价方法研究[D]. 天津:天津大学,2019.

[5] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]//The Thirity-Seventh Asilomar Conference on Signals, Systems and Computers, 2003. IEEE, 2003: 1398-1402.

[6] 邵宇,孙富春,李洪波. 基于视觉特性的无参考型遥感图像质量评价方法[J]. 清华大学学报(自然科学版),2013,53(4):550-555.

[7] TEMEL D, ALREGIB G. PerSIM: Multi-resolution image quality assessment in the perceptually uniform color domain[C]//2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015: 1682-1686.

[8] TEMEL D, PRABHUSHANKAR M, ALREGIB G. UNIQUE: Unsupervised image quality estimation[J]. IEEE signal processing letters, 2016, 23(10): 1414-1418.

[9] 徐冬. 基于多元线性回归和稀疏化表示的高光谱遥感图像去噪[D]. 长沙:国防科学技术大学,2013.

[10] LIU T J, LIN W, KUO C C J. A multi-metric fusion approach to visual quality assessment[C]//2011 Third International Workshop on Quality of Multimedia Experience. IEEE, 2011: 72-77.

[11] 徐冬宇,厉小润,赵辽英,等. 基于多模型融合的高光谱图像质量评价[J]. 激光与光电子学进展,2019,56(2):92-101.

[12] 陈健,李诗云,林丽,等. 模糊失真图像无参考质量评价综述[J]. 自动化学报,2022,48(3):689-711.

[13] SHEIKH H. LIVE image quality assessment database release 2[J]. <http://live.ece.utexas.edu/research/quality>, 2005.

[14] JAYARAMAN D, MITTAL A, MOORTHY A K, et al. Objective quality assessment of multiply distorted images[C]//2012 Conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR). IEEE, 2012: 1693-1697.

[15] PONOMARENKO N, JIN L, IEREMEIEV O, et al. Image database TID2013: Peculiarities, results and perspectives[J]. Signal processing: Image communication, 2015, 30: 57-77.