

面向真实场景的无参考图像质量评价研究综述

牛伟鹏, 董武*

北京印刷学院信息工程学院, 北京

收稿日期: 2026年9月20日; 录用日期: 2026年10月11日; 发布日期: 2026年10月26日

摘要

移动摄影、社交媒体与图像处理技术的发展, 使图像质量评价由受控合成失真研究逐步转向真实场景无参考质量评价。本文聚焦真实场景下的无参考图像质量评价(NR-IQA), 界定任务边界, 归纳复合失真、空间非均匀、质量感知对图像内容与语义的依赖、跨数据集域偏移及主观评分不确定性等挑战, 并比较代表性数据集与评测协议。按照技术路线演进, 系统梳理自然场景统计与浅层学习、卷积神经网络(CNN)与内容自适应建模、排序与自监督学习、跨域学习、Transformer及视觉语言模型等方法, 并从数据集内性能、跨数据集泛化、分数校准与不确定性估计、局部质量证据和语言解释等方面分析其优势、局限与适用条件。本文认为, 真实场景NR-IQA应由追求单一数据集内精度, 转向兼顾泛化性、校准性、可解释性与实际应用可行性。

关键词

无参考图像质量评价, 真实场景失真, Transformer, 视觉语言模型

No-Reference Image Quality Assessment for Real-World Scenarios: A Review

Weipeng Niu, Wu Dong*

School of Information Engineering, Beijing Institute of Graphic Communication, Beijing

Received: September 20, 2026; accepted: October 11, 2026; published: October 26, 2026

Abstract

The widespread adoption of mobile photography, the growth of social media, and advances in image processing technologies have shifted image quality assessment (IQA) research from the study of synthetic distortions under controlled conditions toward no-reference quality assessment in real-world

*通讯作者。

文章引用: 牛伟鹏, 董武. 面向真实场景的无参考图像质量评价研究综述[J]. 图像与信号处理, 2026, 15(4): 517-531.
DOI: 10.12677/jisp.2026.154046

scenarios. This review focuses on real-world no-reference image quality assessment (NR-IQA), delineates its task boundaries, summarizes key challenges, including composite distortions, spatial non-uniformity, the dependence of quality perception on image content and semantics, cross-dataset domain shifts, and uncertainty in subjective scores, and compares representative datasets and evaluation protocols. Following the evolution of technical approaches, we systematically review natural scene statistics and shallow learning, convolutional neural networks and content-adaptive modeling, ranking and self-supervised learning, cross-domain learning, Transformers, and vision-language models. We analyze their strengths, limitations, and applicability in terms of within-dataset performance and cross-dataset generalization, score calibration and uncertainty estimation, local quality evidence, and language-based explanations. We argue that NR-IQA in the real-world should move beyond the pursuit of accuracy on a single dataset and instead consider cross-dataset generalization, reliable score calibration, interpretability, and practical applicability.

Keywords

No-Reference Image Quality Assessment, Authentic Distortion Format, Transformer, Vision-Language Model

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

图像质量评价(Image Quality Assessment, IQA)旨在利用客观算法预测人类对图像视觉质量的主观感知,是成像、传输、压缩、增强和内容分发过程中的关键技术[1]。早期 IQA 研究主要面向受控环境下生成的合成失真, LIVE IQA 数据库为此类失真的感知建模与算法比较提供了标准实验基础[2]。随着移动摄影和网络内容的快速发展,图像内容与退化类型日益复杂, LIVE Challenge 数据库进一步将评价对象扩展至包含复合真实失真的自然图像[3]。在合成失真方面, TID2013 和 KADID-10k 又分别从失真类型与数据规模两个方面扩展了评价基准[4][5]。由于这些应用场景通常无法获得无失真参考图像,无参考图像质量评价(NR-IQA)逐渐成为研究重点,并在移动摄影、网络内容筛选、相机调优和自动增强等任务中具有直接应用价值。

从数据构成看,早期 IQA 数据集通常以少量高质量图像为参考,分别施加噪声、模糊和压缩等参数可控的合成失真。相比之下,真实图像却经历镜头与传感器、曝光与运动、ISP 多帧融合、降噪锐化、用户编辑及平台二次编码等连续链路。 BID 聚焦真实模糊[6], CID2013 揭示设备与成像管线差异[7], KonIQ-10k 覆盖大规模网络图像[8], SPAQ 面向智能手机摄影[9], FLIVE 则呈现更显著的内容长尾与局部退化[10]。因此,研究重心由“识别预设失真”转向“在未知内容与未知退化下估计主观质量”。

上述研究对象的变化也带来了相互耦合的建模困难:复合失真破坏单一退化先验;失真与处理在空间上非均匀,使整图平均掩盖关键区域;人脸、文字和主体等语义改变局部误差的感知权重;设备、分辨率、内容来源及评分协议又造成数据集间域偏移。众包评分还包含观察者差异、显示条件和量表使用偏差,超高分辨率图像则进一步放大局部细节与尺度变化带来的建模难度。因此,数据集内相关系数高并不等于开放环境中可靠,模型还应回答证据在哪里、尺度是否校准以及对新域判断有多大把握。

围绕 IQA 及 NR-IQA,国内已有多篇综述从不同角度总结相关研究。如表 1 所示,王志明从参考信息、特征与学习框架梳理 IQA [11];方玉明等归纳传统与深度无参考评价方法[12];曹玉东等侧重深度

IQA 的网络结构与训练策略[13]；鄢杰斌等按合成、真实和算法相关失真组织研究进展[14]。然而，真实拍摄复合失真通常仍是通用 IQA 体系中的一个分支，跨数据集协议、局部 - 语义证据、域适应和语言解释之间尚缺少贯通式讨论。基于上述不足，本文将真实场景 NR-IQA 作为独立问题域进行讨论，范围不包括视频质量评价和一般图像美学等其他方向。

本文的主要贡献归纳为以下三点：

- 1) 聚焦真实场景 NR-IQA，明确任务边界，并以复合失真、空间非均匀、内容相关和域偏移为核心困难。
- 2) 按质量证据演进梳理 NSS、CNN、排序与跨域学习、Transformer 和视觉语言模型，解释不同路线的能力来源。
- 3) 从跨数据集泛化、可信评价和语言解释角度比较方法，并讨论任务条件评价与可部署质量感知。

Table 1. Positioning of representative Chinese IQA reviews and this review
表 1. 代表性的中文 IQA 综述与本文定位比较

综述	组织主线	真实场景定位	相对本文的不足
王志明(2015)	参考信息、特征与学习框架	NR-IQA 中的失真与应用	未以复合真实失真和跨域泛化为主线[11]
方玉明等(2021)	监督/无监督、传统/深度方法	NR-IQA 的重要方向	未贯通真实数据集、跨域与语言解释[12]
曹玉东等(2021)	网络结构、 训练策略与数据集	深度 IQA 中的相关任务	侧重深度模型，真实场景问题域不独立[13]
鄢杰斌等(2022)	合成、真实及算法相关失真	三类失真之一	失真视角完整，部署与跨库讨论较少[14]
本文	真实场景方法演进与 关键能力比较	独立问题域	贯通基准、局部与语义证据、 跨域、可信和解释

2. 真实场景 NR-IQA 基础与评价基准

2.1. 问题定义与评价范式

主观评价通过单刺激、双刺激或成对比较采集观察者判断，并汇总为平均主观意见分数(MOS)或差分平均主观意见分数(DMOS)。为保证实验过程的规范性和结果的可靠性，国际电信联盟无线电通信部门 (ITU-R)发布了 BT.500 建议书，对电视图像质量主观评价中的观看条件、观察者筛选及数据处理等作出了基本规定[15]。与受控实验室评价不同，LIVE Challenge 采用在线众包方式，为真实场景图像的大规模主观评分提供了代表性范式。这些主观评分既描述了人类感知，也构成客观 IQA 训练与验证的目标尺度。与直接采集观察者判断不同，客观 IQA 则用算法逼近该感知尺度。按参考信息可用程度，客观评价分为全参考(FR)、半参考或减参考(RR)与无参考(NR)；FR 比较待测图与完整参考图，RR 只使用参考侧摘要，NR 方法只接收待测图并依赖统计、失真、语义或任务先验。

除参考信息的可用程度外，失真的生成方式也是区分实验室 IQA 与真实场景 IQA 的关键。实验室合成失真具有已知原图、类型和强度，适合机制验证。TID2013 强调多类型受控退化，KADID-10k 提供更大规模的合成组合，但二者仍难覆盖真实成像链中的混合、非线性和局部处理。本文所称“真实场景”是指由消费级或移动设备拍摄、网络传播与日常编辑自然产生，且通常缺少可配准无失真参考的图像。LIVE Challenge 代表真实复合退化，KonIQ-10k 强调开放网络内容，SPAQ 覆盖智能手机图像与属性评分，FLIVE 兼顾整图和局部质量。本文目标是预测整体感知质量及可选的属性、局部证据或解释；算法生成艺术图像、视频时域质量和一般美学偏好不作为正文主任务。

2.2. 真实场景失真的主要特征

真实退化首先表现为复合性与空间非均匀性：运动模糊可能只覆盖运动主体，HDR 融合会在轮廓附近产生重影，降噪和锐化又可能分别抹除暗部纹理并放大边缘光晕。图像块质量不能简单平均为整图 MOS，显著区域、小面积文字或人脸的失真常具有不成比例的影响。这要求模型同时学习局部质量、区域重要性及从局部到整体的非线性聚合。

内容与语义相关性使“低层异常”不能直接等价于“低质量”：浅景深、低纹理天空和电影色调可以是合法创作，面部失焦和文字模糊却往往严重损害可用性。与此同时，数据集的设备、场景、分辨率、裁剪方式、MOS 范围与标注人群不同，形成内容域、失真域和标签尺度的联合偏移；同一图像的主观分数也应被视为分布或区间，而非无噪声标量。

2.3. 真实场景图像质量数据集

真实数据集从专门采集的模糊与多设备图像，发展到大规模网络图片、智能手机照片以及整图 - 局部联合标注。BID 突出真实模糊，CID2013 分析不同设备的成像差异，LIVE Challenge 确立复合真实失真基准，KonIQ-10k 扩大生态有效样本规模，SPAQ 增加设备、场景和质量属性，FLIVE 则以近 4 万幅图像及局部标注支持从 patch 到整图的学习。图 1 展示了 LIVE Challenge、KonIQ-10k 和 SPAQ 中的代表性图像样例；表 2 则从图像数量、数据来源、分辨率、评分方式和数据特点等方面对主要数据集进行比较。



Figure 1. Representative image samples from Real-World IQA databases

图 1. 代表真实场景 IQA 数据集图像样例

这些数据集具有互补性而非可替代性：小型数据集适合历史对照和可解释分析，大型数据集更利于训练深度模型，SPAQ 与 FLIVE 则支持属性或局部证据。数据集构建还应保留评分分布与可靠性信息，并将分辨率和裁剪策略视为潜在的数据偏移来源。严谨实验应至少覆盖一个训练域和一个外部测试域，并公开去重、预处理、内容划分与 MOS 归一化方式，避免近重复内容或同一拍摄场景跨越训练与测试集合。

Table 2. Representative databases comparison for real-world image quality assessment

表 2. 代表性的真实场景图像质量数据集比较

数据集	图像数量	来源	分辨率	评分方式	特点
BID [6]	586	多相机自然拍摄	设备相关，约百万像素	实验室主观评分	突出真实模糊
CID2013 [7]	480	79 场景、8 类设备	随设备变化	实验室 MOS	分析相机管线差异
LIVE Challenge [3]	1162	移动设备真实图片	原生分辨率多样	大规模在线众包 MOS	经典复合失真基准
KonIQ-10k [8]	10,073	网络自然图片	统一到 1024 × 768	众包 MOS	规模大、生态有效
SPAQ [9]	11,125	66 部智能手机	原生高分辨率	总体及属性 MOS	含场景、EXIF 与属性
FLIVE [10]	39,810	开放网络图片	尺度与长宽比多样	整图和局部众包评分	支持局部到整体聚合

2.4. 评价指标与实验协议

设主观分数为 y_i , 模型输出为 q_i , 经非线性映射后的预测为 $\hat{y}_i$ 。SRCC 和 KRCC 衡量排序一致性, 对单调尺度变换较稳健; PLCC 与 RMSE 衡量校准后的线性相关和绝对误差。在无并列秩时, 常用定义如下, 其中 d_i 为两组秩之差, 和分别为一致对与不一致对数量。

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

$$\text{PLCC} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

$$\text{KRCC} = \frac{2(N_c - N_d)}{n(n-1)} \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

计算 PLCC 和 RMSE 前通常用五参数 Logistic 函数把原始预测映射到 MOS 尺度, 以分离排序能力与输出量纲; LIVE Challenge 的客观模型比较采用了非线性映射后相关性评价。映射只能在训练或验证数据上拟合; 若使用测试标签估计参数, 会造成信息泄漏。常见形式为:

$$f(q) = \beta_1 \left[\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(q - \beta_3))} \right] + \beta_4 q + \beta_5 \quad (5)$$

除指标的计算方式外, 数据划分、预处理与测试协议同样会显著影响评价结果。数据集内测试应说明随机划分比例、是否按参考内容或拍摄场景独立划分、输入分辨率、裁剪/多 patch 策略及预训练数据。LIVE IQA 按参考内容组织受控失真样本, TID2013 保留参考图像与失真等级的对应关系, KADID-10k 也需要避免同一参考内容跨越训练和测试侧。对 KonIQ-10k 应说明官方或随机划分, SPAQ 还应关注设备与场景独立性, FLIVE 则需说明整图和局部标注的使用方式。随机图像划分可能产生乐观估计, 因此固定划分和内容独立划分更适合比较模型能力。

跨数据集测试应冻结训练完成的特征与回归器, 明确训练库 - 测试库方向, 并避免用目标库标签调参。DGQA 通过失真引导的源域选择改善跨境评价[16]; 测试时适应方法则允许在无目标标签条件下更新部分参数[17]。二者的目标数据可见性不同, 必须分开报告。由于 NR-IQA 结果受随机划分影响较大, 建议重复实验并报告中位数或均值、标准差和最差分位, 而非只给出单次最优相关系数。

在明确任务边界、数据集特征和评价协议后, 下文进一步梳理真实场景 NR-IQA 的主要技术路线及其质量表征能力。

3. 真实场景 NR-IQA 方法演进

3.1. 基于自然场景统计与浅层学习的方法

自然场景统计(Natural Scene Statistics, NSS)假设高质量自然图像在适当归一化域具有稳定分布, 失真会改变系数的形状、相关性或尺度, 其基本流程见图 2 所示。DIIVINE 在小波子带上建模[18], BLINDS-II 在 DCT 域提取统计[19], BRISQUE 则使用空间域均值, 除对比度归一化系数及其邻域乘积, 并由 SVR 回归质量[20]。这类方法把“偏离自然性”转化为可解释、低成本的质量证据。

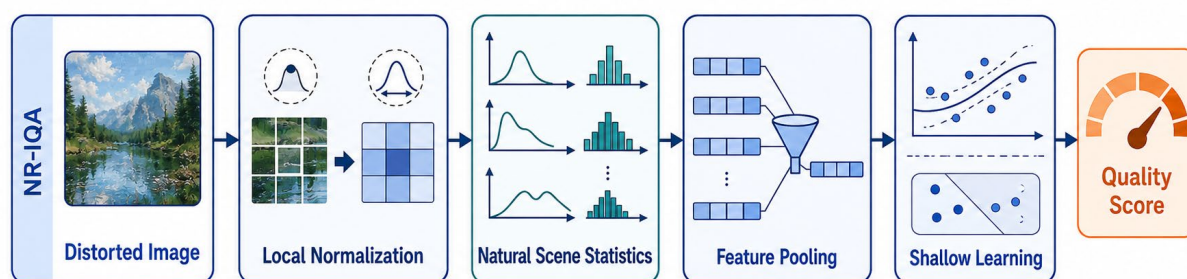


Figure 2. A general NR-IQA framework based on natural scene statistics

图 2. 基于自然场景统计的 NR-IQA 通用框架

NIQE 不依赖主观 MOS, 而是比较待测图像与高质量自然图像统计模型的距离[21]。IL-NIQE 进一步融合梯度、颜色和多域特征, 以扩大失真覆盖[22]。CORNIA 通过无监督字典学习从局部块中获得码本表示, 再以池化和回归完成评价[23]。NSS 与浅层学习数据需求低、部署轻量, 并可分解到亮度、对比度、梯度和颜色等指标。

其局限在于统计异常与感知失真并非一一对应: 浅景深、低纹理或风格化颜色可能偏离自然模型却不降低质量, 局部人脸或文字受损又可能被全局统计稀释。BRISQUE 的预定义局部统计难以覆盖复杂 ISP 链路, IL-NIQE 虽扩展多域特征, 仍缺少内容重要性建模。因此, 这类方法更适合作为轻量基线、先验或与深层特征互补, 而非开放真实场景的唯一评价器。

3.2. 基于 CNN 与内容自适应建模的方法

CNN 把手工统计扩展为可学习的局部质量表征, 图 3 概括了该类方法的典型框架。CNN-IQA 把图像划分为 patch 并回归局部分数[24]。DIQaM/WaDIQaM 进一步学习局部质量及其权重, 使整图预测从等权平均转向显著性或可靠性加权[25]。主线由“整幅图像单次回归”推进到“定位局部退化 - 估计局部质量 - 自适应聚合”, 但将整图 MOS 复制给所有 patch 会引入弱标签噪声。

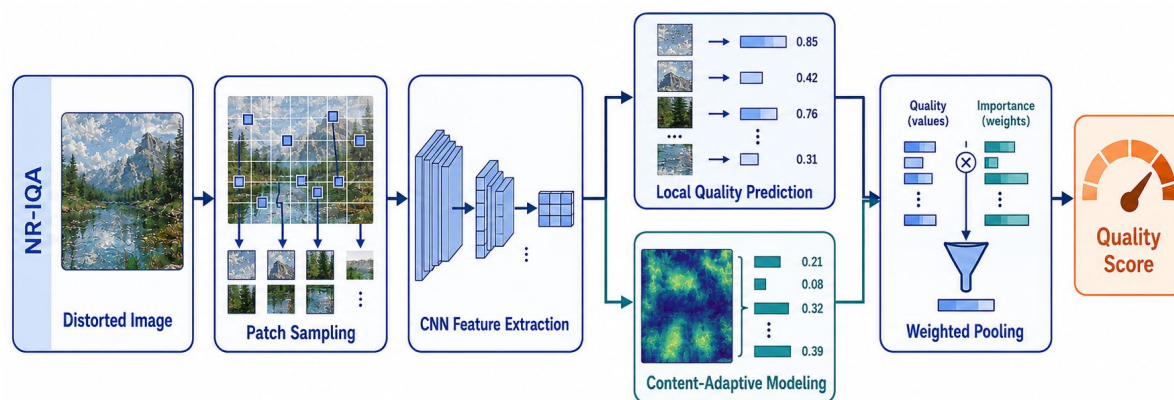


Figure 3. Local prediction and content-adaptive aggregation in CNN-based NR-IQA

图 3. CNN-NR-IQA 的局部预测与内容自适应聚合

MEON 以失真识别辅助质量预测[26]。DB-CNN 通过双线性池化融合面向合成失真的低层表示和面向真实图像的高层特征[27]。HyperIQA 则由超网络根据内容生成质量预测器参数, 使回归规则随图像语义与场景变化[28]。这些设计表明, 内容不是需要消除的干扰, 而是决定失真权重与聚合方式的条件。

PaQ-2-PiQ 利用整图和局部评分学习从 patch 到 picture 的映射, 适合生成质量热图和区域筛选。NIMA

预测评分分布而非单个 MOS, 为观察者分歧提供了更自然的表示[29]。CNN 擅长纹理、边缘和局部伪影, 计算成本也易于通过轻量骨干控制, 但受感受野和训练库偏差影响, 可能过度依赖场景语义或裁剪策略, 对远距离区域关系和跨数据集尺度变化仍不稳定。

3.3. 排序、自监督与跨域泛化方法

绝对 MOS 标注成本较高, 且不同数据集的评分量纲难以统一。为缓解这些问题, 研究者分别从排序监督、自监督表征和跨域学习三个方向展开探索。图 4 概括了三类方法的基本思路, 但它们可以独立或组合使用, 并非固定的串联流程。排序监督约束样本间质量次序, 自监督学习构造质量敏感表示, 跨域方法则进一步抑制数据集特有内容与评分尺度的影响。RankIQA 从合成失真等级构造排序对[30]。MetaIQA 通过元学习快速适应新失真或新数据集[31]。UNIQUE 联合学习质量分布与不确定性, 使实验室和真实失真共享表示[32]。相对监督减弱了对绝对尺度的依赖, 但最终部署仍需将排序空间校准到可解释分数。

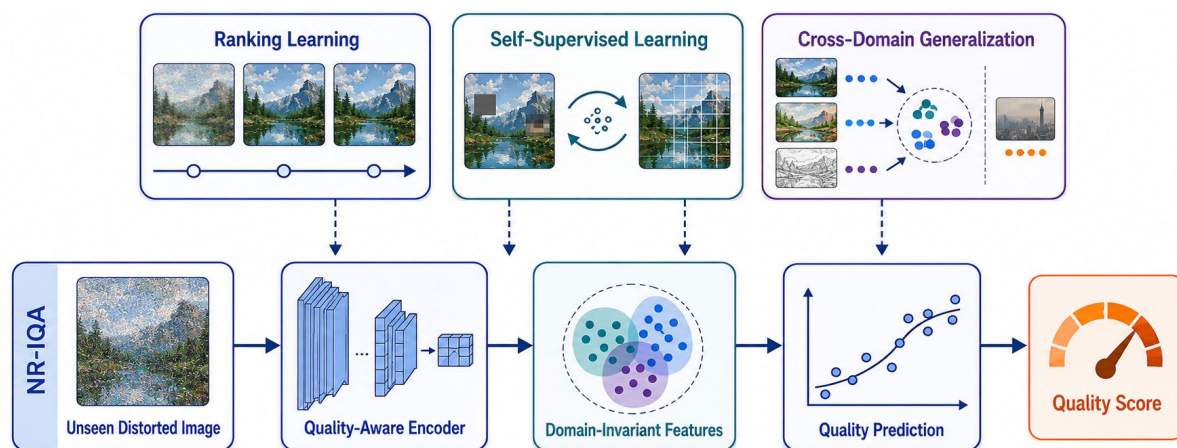


Figure 4. Ranking, self-supervised learning and cross-domain generalization for NR-IQA

图 4. NR-IQA 的排序、自监督学习与跨域泛化

自监督和对比学习利用无 MOS 图像构造质量相关预文本任务。CONTRIQUE 通过不同失真及增强视图学习对比表示[33]。Re-IQA 把内容与质量表征解耦[34], ARNIQA 学习失真流形, 以较少标注适配下游数据集[35]。ATTIQA 进一步借助视觉语言模型生成亮度、色彩、噪声和清晰度等属性伪标签进行大规模预训练[36]。关键不在于预训练数据越多越好, 而在于增强、伪标签和分类质量判断是否同向。

域适应与域泛化直接处理训练库和目标库的内容、失真及评分尺度差异。DGQA 通过失真引导的源域选择减少合成到真实迁移中的负迁移。测试时适应可在不访问目标标签的情况下用组对比和相对排序更新部分参数。QCN 以质量枢轴和几何排序空间增强跨库关系建模[37]。这些方法改善迁移, 但必须报告目标数据可见性、更新成本和失败域, 防止把目标库调参误称为泛化。

3.4. 基于 Transformer 与语义感知的方法

Transformer 把不同位置和尺度的图像块表示为 token, 并用自注意力建立长距离关系, 通用的模型框架如图 5 所示。MUSIQ 保留原生长宽比与尺度编码, 在多分辨率 token 之间联合建模[38]。TReS 结合相对排序与自一致性约束, 提高对裁剪和增强的稳定性[39]。TRIQ 则以 Transformer 编码全局质量关系[40]。相较固定感受野 CNN, 这类模型更容易判断分散失真与主体-背景关系。

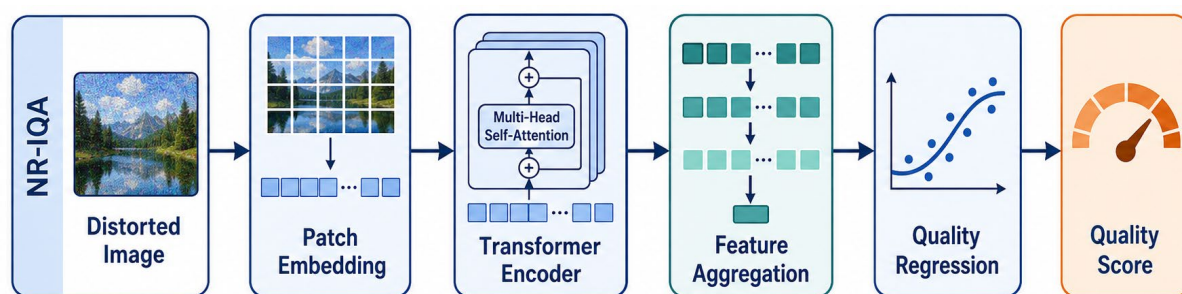


Figure 5. Token modeling and feature aggregation in Transformer-based NR-IQA

图 5. Transformer-NR-IQA 的 Token 建模与特征聚合

MANIQA 使用多维注意力强化通道和空间交互[41]。TOPIQ 从高层语义向下引导局部失真特征[42]。LoDa 在预训练 Transformer 旁注入轻量局部特征, 以较低适配成本补足质量敏感性[43]。方法演进由纯粹扩大全局感受野转向“语义决定关注区域、局部特征保留退化细节、跨尺度交互完成聚合”。

Transformer 的优势是全局内容建模和可变尺度交互, 弱点却恰与其高层预训练来源有关: 大 patch 和强语义不变性可能忽略微弱噪声、振铃、过锐化与局部重影。MUSIQ 以多尺度 token 缓解分辨率变化, LoDa 用局部特征注入补足细节感知。高分辨率全局注意力仍会带来显存和延迟压力, 因此部署时还需评估裁剪是否丢失关键退化。

3.5. 基于视觉语言模型的可解释评价方法

在 Transformer 强化全局语义建模的基础上, 视觉语言模型进一步将视觉质量证据与“好/差”“清晰/模糊”“自然/过度处理”等文本概念对齐, 其一般框架如图 6 所示。从模型形态看, CLIP-IQA [44]、LIQE [45]和 QualiCLI [46] P 等属于双编码器方法, 通常不包含显式的大语言模型; Q-Align 和 DepictQA 等则进一步引入语言生成或指令交互。无论采用何种路线, 文本解释都应能够回溯到局部视觉证据。文本先验扩展了类别与属性空间, 但提示词选择和语言偏见会改变输出尺度。

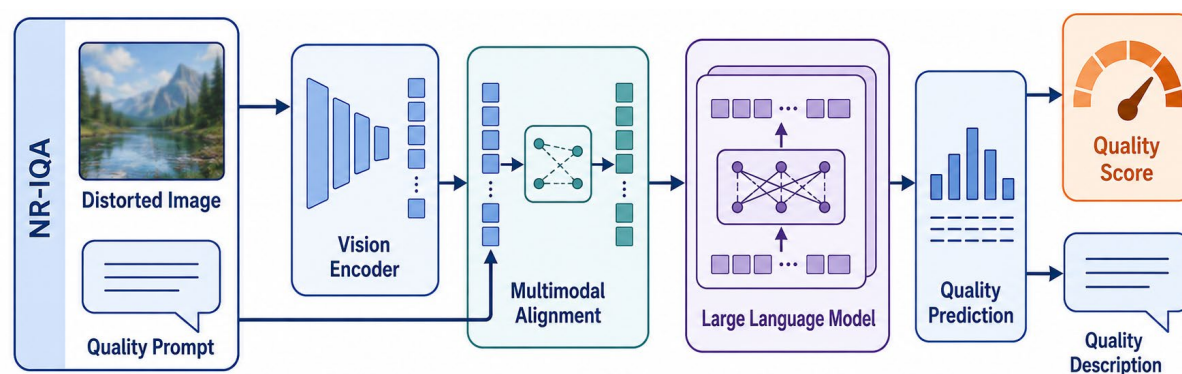


Figure 6. Multimodal alignment, quality scoring and description in VLM-based NR-IQA

图 6. 视觉语言 NR-IQA 的多模态对齐、质量评分与描述

Q-Bench 评估通用多模态模型的低层视觉能力[47], Q-Align 把连续 MOS 转换为文本定义的离散等级以训练多模态大模型[48]。DepictQA 将任务推进到图像质量比较、失真定位和自然语言描述[49]。PromptIQA 用少量图像-分数对作为上下文提示适应新的评分要求[50]。Align-IQA 则通过质量对齐强化感知判断[51]。NR-IQA 由此从输出单一分数, 发展为联合输出分数、属性、失真原因、比较结论与语言解释。

语言流畅不等于解释真实。LIQE 可能受提示词和图文先验影响而改变质量尺度，开放场景中的质量描述也可能与局部像素证据脱节。可信系统应把分数头、属性头、局部质量图和语言头对齐到共同视觉证据，利用反事实遮挡、区域一致性和人工核验检查解释；当模型无法定位或置信度低时，应显式表达不确定，而不是补写确定性描述。

总体而言，上述方法在质量证据、监督方式和输出形式上各有侧重，需要结合实验性能与应用条件进行横向比较。

4. 代表性方法比较与讨论

4.1. 不同技术路线比较

不同路线的差异本质上是质量证据、监督来源与聚合机制的差异。BRISQUE 体现可解释统计先验，HyperIQA 强化内容自适应的局部质量建模，QCN 重组跨域排序关系，MUSIQ 扩大局部 - 全局交互，Align-IQA 再把质量映射到开放文本概念。表 3 以能力等级作定性比较，“潜力强”不等于在所有数据集上已稳定实现。

Table 3. Comprehensive comparison of major technical routes for real-world NR-IQA
表 3. 真实场景 NR-IQA 主要技术路线综合比较

代表方法	质量特点	局部失真	内容语义	跨库泛化	文字解释	复杂度
NSS/浅层(BRISQUE [20]; NIQE [21])	自然统计、梯度、颜色	中	弱	弱 - 中	否	低
CNN/自适应 (DB-CNN [27]; HyperIQA [28])	局部纹理、多尺度与权重	强	中	中	否	中
排序/跨域(RankIQA [30]; QCN [37])	相对质量、增强与域关系	中 - 强	中	中 - 强	否	中
Transformer (MUSIQ [38]; TOPIQ [42])	多尺度 token 与全局关系	中 - 强	强	中	有限	高
VLM/MLLM (LIQE [45]; DepictQA [49])	图文等级、属性与推理	中	强	潜力强	是	高

4.2. 代表性方法性能比较

在表 3 定性比较不同技术路线的基础上，表 4 进一步汇总了 20 种代表性方法的数据集内性能，数值按各方法原论文或其公开比较表转录，用于观察技术路线而非建立统一排行榜。LoDa 的结果依赖局部适配器和预训练 Transformer 设置，PerceptCLIP 的结果则来自 CLIP 特征的轻量质量适配协议[52]。即使数据集相同，随机划分、场景去重、输入分辨率、patch 数量、预训练数据以及是否集成预测均可能不同；因此每个 SRCC/PLCC 对必须连同协议阅读，破折号表示相应来源未报告结果。

Table 4. Representative within-database performance (SRCC/PLCC)
表 4. 代表方法数据集内性能(SRCC/PLCC)

方法	LIVE Challenge	KonIQ-10k	SPAQ	FLIVE	主要特点
IL-NIQE [22]	0.508/0.508	0.523/0.537	0.713/0.712	0.294/0.332	多域 NSS
BRISQUE [20]	0.629/0.629	0.681/0.685	0.809/0.817	0.303/0.341	空间 NSS + SVR
WaDIQaM-NR [25]	0.682/0.671	0.804/0.807	—	0.455/0.467	局部加权
DB-CNN [27]	0.851/0.869	0.875/0.884	0.911/0.915	0.545/0.551	双线性 CNN
MetaIQA [31]	0.835/0.802	0.887/0.856	—	0.540/0.507	元学习

续表

HyperIQA [28]	0.855/0.871	0.908/0.921	0.916/0.919	0.535/0.623	内容自适应
KonCept512 [8]	—	0.921/0.937	—	—	大规模 CNN
MUSIQ [38]	—	0.916/0.928	0.917/0.921	0.646/0.739	多尺度 Transformer
TReS [39]	0.846/0.877	0.915/0.928	—	0.544/0.625	相对排序
DEIQT [53]	0.875/0.894	0.921/0.934	0.919/0.923	0.571/0.663	注意力面板
LIQE [45]	0.904/0.910	0.919/0.908	—	—	图文联合
Re-IQA [34]	0.840/0.854	0.914/0.923	0.918/0.925	0.645/0.733	质量/内容解耦
QPT-ResNet50 [54]	0.895/0.914	0.927/0.941	0.925/0.928	0.610/0.677	质量预训练
LoDa [43]	0.876/0.899	0.932/0.944	0.925/0.928	0.578/0.679	局部特征注入
CONTRIQUE [33]	0.845/0.857	0.894/0.906	0.914/0.919	0.580/0.641	对比预训练
MANIQA [41]	0.870/0.899	0.923/0.933	0.923/0.926	0.731/0.753	多维注意力
CLIP-IQA+ [44]	0.805/0.832	0.895/0.909	0.864/0.866	0.606/0.641	CLIP 提示
TOPIQ [42]	0.870/0.884	0.928/0.941	0.921/0.924	0.633/0.722	语义拓扑
YOTO-S [55]	0.869/0.903	0.926/0.938	—	0.570/0.652	FR/NR 统一
PerceptCLIP [52]	0.908/0.923	0.945/0.954	0.930/0.933	0.590/0.689	轻量 CLIP 适配

FLIVE 的相关系数整体低于 KonIQ-10k 和 SPAQ，反映其内容长尾、分辨率多样和真实退化组合更难。KonIQ-10k 具有更统一的图像尺度与筛选流程，SPAQ 集中于智能手机摄影，FLIVE 则覆盖更开放的网络内容和局部标注。这些差异说明数据集内数值不能直接代表所有场景。然而，数据集内性能可能掩盖模型对特定内容分布和 MOS 尺度的依赖。表 5 进一步给出了固定源模型且不使用目标域标签条件下的代表性跨数据集结果。在固定源模型、不使用目标标签的设置中，特征是否依赖数据集内容和 MOS 尺度往往比骨干网络大小更关键。

Table 5. Representative cross-database SRCC without target labels

表 5. 不使用目标标签的代表性跨数据集 SRCC

方法	LIVE Challenge → KonIQ-10k	KonIQ-10k → LIVE Challenge	特点
HyperIQA [28]	0.772	0.785	内容自适应仍受源域影响
CONTRIQUE [33]	0.676	0.731	预文本任务与目标分布存在差距
Re-IQA [34]	0.769	0.791	质量/内容解耦改善部分方向
QPT [54]	0.749	0.821	统一协议复现的质量预训练基线
QCN [37]	0.784	0.840	几何排序空间，两个方向均提升

4.3. 方法局限与应用选择

第一组矛盾是低层失真感知与高级语义建模：DB-CNN 利用卷积表征低层与高层质量特征，HyperIQA 使预测规则随内容变化，TOPIQ 再以高层语义引导局部失真建模。大型视觉语言模型虽能理解主体和场景，CLIP-IQA 的提示式评价仍可能因语义不变性掩盖细微退化。更稳妥的结构是保留高分辨率局部分支，以语义分支估计区域重要性，再由可审计聚合器输出整体与属性分数。

第二组矛盾是数据集内高性能与跨数据集泛化不足。QCN 用几何排序关系改善未知域评价，DGQA

通过源域选择减少负迁移, 测试时适应允许模型在无目标标签时更新。这些策略均可能引入新的数据可见性或稳定性条件。因此, 应用选择应优先看与部署域一致的外部测试, 而不是单一数据集最高 SRCC。

第三组矛盾是准确性、计算复杂度与可解释性。移动摄影和实时质量控制可选用 NIQE 等轻量统计基线, 或采用 HyperIQA 式内容自适应 CNN。大规模网络图像筛选适合离线预训练加轻量回归头, PaQ-2-PiQ 的局部评分还可支持区域筛选。跨设备服务应加入留域验证、校准和漂移监测; 面向审核或质量改进建议的场景, 可采用 Align-IQA 式语义对齐, 但必须把语言结论绑定到局部证据和置信度。上述比较表明, 真实场景 NR-IQA 的瓶颈已由单纯提高数据集内精度, 逐渐转向泛化、校准、标签可靠性与解释可信度等系统性问题。

5. 挑战与发展趋势

5.1. 跨数据集泛化与质量尺度校准

综合上述方法与实验比较, 真实场景 NR-IQA 的进一步发展主要受到跨数据集泛化与尺度校准、数据与标签质量、评价可信性与可解释性以及模型效率与实际部署四方面制约。跨数据集泛化的关键不在于继续提高某一源库上的平均相关性, 而在于模型能否在未知设备、未知内容和未知处理链上保持稳定排序。因此, 应把域泛化、测试时适应和持续更新视为三类具有不同目标数据访问权限与参数更新规则的应用设置: 域泛化不得接触目标数据; 测试时适应只允许使用无标签目标样本; 持续更新还必须监控旧域遗忘。基准应据此冻结数据访问边界, 并同时报告最差域性能、适应耗时和失败率。由此可见, 所谓通用质量模型首先应是协议可审计的模型, 而不是仅在混合数据集上取得更高平均分的模型。

质量尺度校准不是相关系数之后的附属步骤, 而是跨库部署成立的前提。不同数据集的 MOS 范围、锚点和观察者群体并不等价, 简单线性归一化只能统一数值区间, 不能统一感知含义。可行的基准应设置跨库共享锚图(由多个数据集共同评分的锚定图像)或成对比较样本, 由层次概率模型同时学习潜在质量尺度和数据集专属映射, 并为输出给出校准不确定性。据此, 跨数据集评价必须把“排序是否迁移”和“分数是否可解释”分开测量, 以校准误差和置信区间补充 SRCC 与 PLCC。

5.2. 数据规模与主观标签可靠性

真实数据建设的核心指标不应只是图像数量, 而应同时衡量设备、场景、光照、分辨率和处理链的覆盖度, 以及标签的可追溯性。数据集需要保留原始评分分布、样本置信区间、异常率和重复评分一致性, 并公开采集设备、场景类别和后处理信息。这样才能区分“图像本身存在感知争议”和“模型预测不稳定”。因此, 在相同标注预算下, 少量覆盖新域且具有重复评分的高信息样本, 比大量来源单一、只保留 MOS 的同质样本更有价值。

弱监督、主动学习和合成数据可以提高采集效率, 但三者应服务于同一个闭环: 先用不确定性和域差异选择最有信息的样本, 再以排序或属性标签降低人工判断难度, 最后用真实留出域检验合成退化是否有效。合成样本若不能复现真实成像链中的空间分布、ISP 痕迹和复合退化, 只会扩大训练规模而不会扩大有效覆盖。由此可见, 未来数据策略应从一次性建库转向“模型发现盲区 - 人工补充关键样本 - 外部域复核”的持续数据治理。

5.3. 可信与可解释质量评价

可信质量评价不应只输出单一质量分数, 还应说明判断依据及预测可靠程度。可解释性研究的重点应从展示注意力图或生成流畅描述, 转向验证解释是否真实反映模型的决策过程。语言解释需要与失真类型、空间位置、严重程度及受影响内容相对应, 并在相关视觉证据发生变化时与质量分数保持一致。

注意力机制只能提供相关性线索, 未来还应结合输入干预、反事实分析和概念归因等方法, 检验不同视觉因素对质量预测的实际影响。

不确定性估计应区分主观评分分歧带来的数据不确定性, 以及样本不足、域偏移和模型能力限制产生的模型不确定性。模型输出应由点预测逐步扩展为评分分布或预测区间, 并检验置信度与实际误差是否匹配。在实际系统中, 不确定性可用于识别低可信预测、触发人工复核、监测数据分布变化并辅助样本筛选。由此, 可信质量评价的关键不在于增加更多输出, 而在于保证质量分数、解释证据与预测置信度之间的一致性。

5.4. 模型效率与实际部署

真实场景 NR-IQA 需要处理高分辨率图像和局部细节, 模型效率与细粒度失真感知之间存在明显矛盾。结构层面可采用轻量化骨干、局部注意力和自适应区域选择减少冗余计算; 压缩层面可通过剪枝、量化和低秩分解降低模型大小与内存占用; 训练层面可利用知识蒸馏将复杂模型的质量排序、局部证据和语义信息传递给轻量模型。上述方法的目标不应只是减少参数量, 还应尽量保持跨数据集泛化和质量尺度稳定性。

实际部署评价应同时报告预测性能、模型大小、峰值内存、端到端延迟和设备能耗, 并明确输入分辨率及局部采样数量。面向资源受限设备, 可根据图像复杂度动态调整分辨率、采样区域和推理深度, 也可采用轻量模型筛选与复杂模型复核相结合的分级处理方式。未来研究需要由单纯追求数据集内精度, 转向准确性、实时性、稳定性与资源消耗之间的综合平衡。

6. 结论

真实场景 NR-IQA 面对复合、空间非均匀、内容相关且跨域的退化, 评价基准还受到数据集规模、MOS 尺度与主观不确定性的共同约束。仅用单次随机划分上的相关系数不能充分说明方法价值; 数据集内预测、跨数据集迁移、绝对尺度校准、局部证据和置信度应构成互补的评价体系。

在方法层面, 方法演进体现为质量证据持续扩展: NSS 提供可解释统计先验, CNN 强化局部与内容自适应, 排序、自监督和跨域学习缓解标签与迁移问题, Transformer 连接全局语义和局部退化, 视觉语言模型进一步输出属性、比较和语言解释。未来研究重点应由数据集内预测精度转向跨域泛化、可靠校准、证据一致的解释及资源受限场景下的高效部署。

参考文献

- [1] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, 600-612. <https://doi.org/10.1109/tip.2003.819861>
- [2] Sheikh, H.R., Sabir, M.F. and Bovik, A.C. (2006) A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Transactions on Image Processing*, **15**, 3440-3451. <https://doi.org/10.1109/tip.2006.881959>
- [3] Ghadiyaram, D. and Bovik, A.C. (2016) Massive Online Crowdsourced Study of Subjective and Objective Picture Quality. *IEEE Transactions on Image Processing*, **25**, 372-387. <https://doi.org/10.1109/tip.2015.2500021>
- [4] Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., et al. (2015) Image Database TID2013: Peculiarities, Results and Perspectives. *Signal Processing: Image Communication*, **30**, 57-77. <https://doi.org/10.1016/j.image.2014.10.009>
- [5] Lin, H., Hosu, V. and Saupe, D. (2019) KADID-10k: A Large-Scale Artificially Distorted IQA Database. 2019 *Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, Berlin, 5-7 June 2019, 1-3. <https://doi.org/10.1109/qomex.2019.8743252>
- [6] Ciancio, A., da Costa, A.L.N.T., da Silva, E.A.B., Said, A., Samadani, R. and Obrador, P. (2011) No-Reference Blur Assessment of Digital Pictures Based on Multifeature Classifiers. *IEEE Transactions on Image Processing*, **20**, 64-75.

- <https://doi.org/10.1109/tip.2010.2053549>
- [7] Virtanen, T., Nuutinen, M., Vaahteranoksa, M., Oittinen, P. and Hakkinen, J. (2015) CID2013: A Database for Evaluating No-Reference Image Quality Assessment Algorithms. *IEEE Transactions on Image Processing*, **24**, 390-402. <https://doi.org/10.1109/tip.2014.2378061>
 - [8] Hosu, V., Lin, H., Sziranyi, T. and Saupe, D. (2020) KonIQ-10k: An Ecologically Valid Database for Deep Learning of Blind Image Quality Assessment. *IEEE Transactions on Image Processing*, **29**, 4041-4056. <https://doi.org/10.1109/tip.2020.2967829>
 - [9] Fang, Y., Zhu, H., Zeng, Y., Ma, K. and Wang, Z. (2020) Perceptual Quality Assessment of Smartphone Photography. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 3677-3686. <https://doi.org/10.1109/cvpr42600.2020.00373>
 - [10] Ying, Z., Niu, H., Gupta, P., Mahajan, D., Ghadiyaram, D. and Bovik, A. (2020) From Patches to Pictures (PaQ-2-PiQ): Mapping the Perceptual Space of Picture Quality. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 3572-3582. <https://doi.org/10.1109/cvpr42600.2020.00363>
 - [11] 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.
 - [12] 方玉明, 睦相杰, 鄢杰斌, 刘学林, 黄丽萍. 无参考图像质量评价研究进展[J]. 中国图象图形学报, 2021, 26(2): 265-286.
 - [13] 曹玉东, 刘海燕, 贾旭, 李晓会. 基于深度学习的图像质量评价方法综述[J]. 计算机工程与应用, 2021, 57(23): 27-36.
 - [14] 鄢杰斌, 方玉明, 刘学林. 图像质量评价研究综述——从失真的角度[J]. 中国图象图形学报, 2022, 27(5): 1430-1466.
 - [15] International Telecommunication Union Radiocommunication Sector (2023) Methodologies for the Subjective Assessment of the Quality of Television Images. Recommendation ITU-R BT.500-15, Geneva. <https://www.itu.int/rec/R-REC-BT.500-15-202305-I>
 - [16] Li, A., Wu, J., Liu, Y. and Li, L. (2024) Bridging the Synthetic-to-Authentic Gap: Distortion-Guided Unsupervised Domain Adaptation for Blind Image Quality Assessment. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 28422-28431. <https://doi.org/10.1109/cvpr52733.2024.02685>
 - [17] Roy, S., Mitra, S., Biswas, S. and Soundararajan, R. (2023) Test Time Adaptation for Blind Image Quality Assessment. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 16696-16705. <https://doi.org/10.1109/iccv51070.2023.01535>
 - [18] Moorthy, A.K. and Bovik, A.C. (2011) Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality. *IEEE Transactions on Image Processing*, **20**, 3350-3364. <https://doi.org/10.1109/tip.2011.2147325>
 - [19] Saad, M.A., Bovik, A.C. and Charrier, C. (2012) Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain. *IEEE Transactions on Image Processing*, **21**, 3339-3352. <https://doi.org/10.1109/tip.2012.2191563>
 - [20] Mittal, A., Moorthy, A.K. and Bovik, A.C. (2012) No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing*, **21**, 4695-4708. <https://doi.org/10.1109/tip.2012.2214050>
 - [21] Mittal, A., Soundararajan, R. and Bovik, A.C. (2013) Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Processing Letters*, **20**, 209-212. <https://doi.org/10.1109/lsp.2012.2227726>
 - [22] Zhang, L., Zhang, L. and Bovik, A.C. (2015) A Feature-Enriched Completely Blind Image Quality Evaluator. *IEEE Transactions on Image Processing*, **24**, 2579-2591. <https://doi.org/10.1109/tip.2015.2426416>
 - [23] Ye, P., Kumar, J., Kang, L. and Doermann, D. (2012) Unsupervised Feature Learning Framework for No-Reference Image Quality Assessment. 2012 *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 16-21 June 2012, 1098-1105. <https://doi.org/10.1109/cvpr.2012.6247789>
 - [24] Kang, L., Ye, P., Li, Y. and Doermann, D. (2014) Convolutional Neural Networks for No-Reference Image Quality Assessment. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 1733-1740. <https://doi.org/10.1109/cvpr.2014.224>
 - [25] Bosse, S., Maniry, D., Muller, K., Wiegand, T. and Samek, W. (2018) Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Transactions on Image Processing*, **27**, 206-219. <https://doi.org/10.1109/tip.2017.2760518>
 - [26] Ma, K., Liu, W., Zhang, K., Duanmu, Z., Wang, Z. and Zuo, W. (2018) End-to-End Blind Image Quality Assessment Using Deep Neural Networks. *IEEE Transactions on Image Processing*, **27**, 1202-1213. <https://doi.org/10.1109/tip.2017.2774045>
 - [27] Zhang, W., Ma, K., Yan, J., Deng, D. and Wang, Z. (2020) Blind Image Quality Assessment Using a Deep Bilinear Convolutional Neural Network. *IEEE Transactions on Circuits and Systems for Video Technology*, **30**, 36-47.

- <https://doi.org/10.1109/tcsvt.2018.2886771>
- [28] Su, S.L., Yan, Q.S., Zhu, Y., Zhang, C., Ge, X., Sun, J., *et al.* (2020) Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Network. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 3664-3673. <https://doi.org/10.1109/cvpr42600.2020.00372>
 - [29] Talebi, H. and Milanfar, P. (2018) NIMA: Neural Image Assessment. *IEEE Transactions on Image Processing*, **27**, 3998-4011. <https://doi.org/10.1109/tip.2018.2831899>
 - [30] Liu, X., Van De Weijer, J. and Bagdanov, A.D. (2017) RankIQA: Learning from Rankings for No-Reference Image Quality Assessment. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 1040-1049. <https://doi.org/10.1109/iccv.2017.118>
 - [31] Zhu, H., Li, L., Wu, J., Dong, W. and Shi, G. (2020) MetaIQA: Deep Meta-Learning for No-Reference Image Quality Assessment. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 14131-14140. <https://doi.org/10.1109/cvpr42600.2020.01415>
 - [32] Zhang, W., Ma, K., Zhai, G. and Yang, X. (2021) Uncertainty-Aware Blind Image Quality Assessment in the Laboratory and Wild. *IEEE Transactions on Image Processing*, **30**, 3474-3486. <https://doi.org/10.1109/tip.2021.3061932>
 - [33] Madhusudana, P.C., Birkbeck, N., Wang, Y., Adsumilli, B. and Bovik, A.C. (2022) Image Quality Assessment Using Contrastive Learning. *IEEE Transactions on Image Processing*, **31**, 4149-4161. <https://doi.org/10.1109/tip.2022.3181496>
 - [34] Saha, A., Mishra, S. and Bovik, A.C. (2023) Re-IQA: Unsupervised Learning for Image Quality Assessment in the Wild. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 5846-5855. <https://doi.org/10.1109/cvpr52729.2023.00566>
 - [35] Agnolucci, L., Galteri, L., Bertini, M. and Del Bimbo, A. (2024) ARNIQA: Learning Distortion Manifold for Image Quality Assessment. 2024 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2024, 188-197. <https://doi.org/10.1109/wacv57701.2024.00026>
 - [36] Kwon, D., Kim, D., Ki, S., Jo, Y., Lee, H.E. and Kim, S.J. (2024) ATTIQA: Generalizable Image Quality Feature Extractor Using Attribute-Aware Pretraining. In: *Lecture Notes in Computer Science*, Springer, 284-300. https://doi.org/10.1007/978-981-96-0911-6_17
 - [37] Shin, N.H., Lee, S.H. and Kim, C.S. (2024) Blind Image Quality Assessment Based on Geometric Order Learning. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 12799-12808. <https://doi.org/10.1109/cvpr52733.2024.01216>
 - [38] Ke, J., Wang, Q., Wang, Y., Milanfar, P. and Yang, F. (2021) MUSIQ: Multi-Scale Image Quality Transformer. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 5128-5137. <https://doi.org/10.1109/iccv48922.2021.00510>
 - [39] Golestaneh, S.A., Dadsetan, S. and Kitani, K.M. (2022) No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency. 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2022, 3989-3999. <https://doi.org/10.1109/wacv51458.2022.00404>
 - [40] You, J. and Korhonen, J. (2021) Transformer for Image Quality Assessment. 2021 *IEEE International Conference on Image Processing (ICIP)*, Anchorage, 19-22 September 2021, 1389-1393. <https://doi.org/10.1109/icip42928.2021.9506075>
 - [41] Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., *et al.* (2022) MANIQA: Multi-Dimension Attention Network for No-Reference Image Quality Assessment. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, 19-20 June 2022, 1190-1199. <https://doi.org/10.1109/cvprw56347.2022.00126>
 - [42] Chen, C., Mo, J., Hou, J., Wu, H., Liao, L., Sun, W., *et al.* (2024) TOPIQ: A Top-Down Approach from Semantics to Distortions for Image Quality Assessment. *IEEE Transactions on Image Processing*, **33**, 2404-2418. <https://doi.org/10.1109/tip.2024.3378466>
 - [43] Xu, K., Liao, L., Xiao, J., Chen, C., Wu, H., Yan, Q., *et al.* (2024) Boosting Image Quality Assessment through Efficient Transformer Adaptation with Local Feature Enhancement. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 2622-2672. <https://doi.org/10.1109/cvpr52733.2024.00257>
 - [44] Wang, J., Chan, K.C.K. and Loy, C.C. (2023) Exploring CLIP for Assessing the Look and Feel of Images. *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**, 2555-2563. <https://doi.org/10.1609/aaai.v37i2.25353>
 - [45] Zhang, W., Zhai, G., Wei, Y., Yang, X. and Ma, K. (2023) Blind Image Quality Assessment via Vision-Language Correspondence: A Multitask Learning Perspective. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 14071-14081. <https://doi.org/10.1109/cvpr52729.2023.01352>
 - [46] Agnolucci, L., Galteri, L. and Bertini, M. (2024) Quality-Aware Image-Text Alignment for Real-World Image Quality Assessment. <https://arxiv.org/abs/2403.11176>

-
- [47] Zhang, Z., Wu, H., Zhang, E., Zhai, G. and Lin, W. (2024) Q-Bench⁺: A Benchmark for Multi-Modal Foundation Models on Low-Level Vision from Single Images to Pairs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **46**, 10404-10418. <https://doi.org/10.1109/tpami.2024.3445770>
 - [48] Wu, H., Zhang, Z., Zhang, W., *et al.* (2024) Q-Align: Teaching LMMs for Visual Scoring via Discrete Text-Defined Levels. *Proceedings of the 41st International Conference on Machine Learning*, Vienna, 21-27 July 2024, 54015-54029.
 - [49] You, Z., Li, Z., Gu, J., Yin, Z., Xue, T. and Dong, C. (2024) Depicting beyond Scores: Advancing Image Quality Assessment through Multi-Modal Language Models. In: *Lecture Notes in Computer Science*, Springer, 259-276. https://doi.org/10.1007/978-3-031-72970-6_15
 - [50] Chen, Z., Qin, H., Wang, J., Yuan, C., Li, B., Hu, W., *et al.* (2024) Promptiqa: Boosting the Performance and Generalization for No-Reference Image Quality Assessment via Prompts. In: *Lecture Notes in Computer Science*, Springer, 247-264. https://doi.org/10.1007/978-3-031-73232-4_14
 - [51] Yang, J., Fu, J., Zhang, Z., Liu, L., Li, Q., Zhang, W., *et al.* (2024) Align-IQA: Aligning Image Quality Assessment Models with Diverse Human Preferences via Customizable Guidance. *Proceedings of the 32nd ACM International Conference on Multimedia*, Melbourne, 28 October-1 November 2024, 10008-10017. <https://doi.org/10.1145/3664647.3681511>
 - [52] Zalcher, A., Wasserman, N., Beliy, R., Heinimann, O. and Irani, M. (2025) Don't Judge before You CLIP: A Unified Approach for Perceptual Tasks. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=uvQTYi6kbu>
 - [53] Qin, G., Hu, R., Liu, Y., Zheng, X., Liu, H., Li, X., *et al.* (2023) Data-Efficient Image Quality Assessment with Attention-Panel Decoder. *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**, 2091-2100. <https://doi.org/10.1609/aaai.v37i2.25302>
 - [54] Zhao, K., Yuan, K., Sun, M., Li, M. and Wen, X. (2023) Quality-Aware Pretrained Models for Blind Image Quality Assessment. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 22302-22313. <https://doi.org/10.1109/cvpr52729.2023.02136>
 - [55] Yun, Y.K. and Lin, W. (2026) You Only Train Once: A Unified Framework for Both Full-Reference and No-Reference Image Quality Assessment. *IEEE Transactions on Image Processing*, **35**, 4467-4479. <https://doi.org/10.1109/tip.2026.3661408>