摘要

随着数字成像技术的飞速革新与视觉媒体应用的爆炸式增长，图像质量评价（Image Quality Assessment, IQA）已发展成为计算机视觉与多媒体处理领域中的一项核心基础技术。其在图像增强、压缩、传输、重建乃至人工智能生成内容（AIGC）等众多任务中，对算法优化与系统性能提升具有不可或缺的指导作用。然而，图像质量本质上是人类视觉系统的一种主观感知体验，这使得构建大规模、高一致性的主观评价数据集面临周期长、成本高、标注一致性难保证等严峻挑战。现有数据集的规模因而受到严重限制，并直接导致当前基于深度学习的 IQA 模型普遍陷入过拟合的困境，表现出对训练集外失真类型、复杂场景内容以及跨数据集分布的泛化能力显著不足。尤其是在面对未知失真（如新兴 AI 生成模型特有的伪影）以及真实开放环境下的复杂视觉内容时，模型性能往往急剧衰退。因此，如何突破数据标注的瓶颈，构建具有强泛化能力的 IQA 模型，已成为推动该技术在实际场景中落地应用的关键科学问题。

根据模型在预测时所能获取的参考图像信息量，图像质量评价方法可系统性地划分为全参考图像质量评价（FR-IQA）、半参考图像质量评价（RR-IQA）与无参考图像质量评价（NR-IQA，亦称盲图像质量评价 BIQA）三类。其中，无参考图像质量评价方法不依赖于任何形式的原始参考图像信息，更符合监控、移动摄影、社交媒体内容审核等实际应用场景的苛刻条件，因而展现出极其广阔的应用前景。鉴于此，本文聚焦于提升无参考图像质量评价模型的泛化能力这一核心目标，分别从数据利用、范式优化、与模态融合三个具有内在联系的创新维度展开系统性研究，提出了一套层次分明、互为补充的具有高泛化性能的 NR-IQA 训练新框架。

在数据利用层面，为破解标注数据稀缺的根本性难题，本文提出了一种基于置信度量化伪标签学习的半监督无参考图像质量评价框架（CPL-IQA）。该框架将熵最小化与图结构标签传播思想相结合，为大量无标注样本生成具有可靠性度量的伪质量标签。通过设计模型训练与标签优化交替进行的迭代机制，该框架能够充分协同利用有限的标注数据与海量的无标注数据，实现端到端的半监督训练。这不仅从源头上缓解了模型对有限标注数据的过拟合，更通过引入无标注数据中蕴含的丰富视觉结构信息，显著增强了模型对未见过失真与内容的泛化能力。

在范式优化层面，本文首次从因果推断的视角，深刻揭示了制约 NR-IQA 模型泛化能力的本质原因。通过构建 NR-IQA 任务的结构因果图，清晰指出了图像内容与失真类型作为混淆变量，在失真图像与主观评分之间引入了虚假关联。为切断这些混淆路径，本文提出了一种基于因果学习的无参考图像质量评价训练范式（Causal-IQA），将模型的学习目标从拟合图像与质量分数间的相关关系从根本上转变为探究其内在的因

果关系。该方法创新性地设计了反事实数据划分机制，并依据后门准则对模型进行因果干预，从而有效剥离了图像内容与失真类型对质量评分的混淆效应，引导模型学习本质、鲁棒且独立于混杂因子的质量特征，最终实现泛化性能的质的飞跃。

在模态融合层面，为进一步挖掘先验知识以增强模型的质量感知能力，本文探索了视觉-语言模型在 NR-IQA 任务中的巨大潜力，提出了一种基于自适应提示学习的多模态无参考图像质量评价方法（MMP-IQA）。该方法以在大规模图文对上预训练的 CLIP 模型为基石，设计了条件跨模态融合模块，通过计算视觉质量原型并与文本描述进行交互，实现了视觉与语言模态在深度语义层面的协同感知。针对传统离散提示工程耗时费力且泛化性弱的缺陷，该方法引入了连续可学习的质量提示向量，使模型能够端到端地自适应捕获与图像质量高度相关的上下文语义。此外，本文还构建了一个集相对质量排序、线性关系度量与数据集特定绝对质量评估于一体的三重评估器框架，并提出了双自适应权重调整策略，确保了模型在混合多数据集训练中的稳定性与高效性。

本文所提出的三种方法在多个权威公开 IQA 数据集上进行了全面而严格的实验验证，测试场景覆盖了真实失真、合成失真以及前沿的人工智能生成（AIGC）图像等多种复杂情况。大量的实验结果表明，无论是在同类数据集内的性能表现，还是在极具挑战性的跨数据集、跨失真类型的泛化测试中，本文所提出的三种方法均超越了现有主流及前沿方法，在 SRCC 与 PLCC 等关键评价指标上取得了领先优势，充分验证了其在泛化能力、鲁棒性及实际应用价值上的优异表现。

总的来说，本文围绕“如何构建具有强泛化能力的无参考图像质量评价模型”这一核心科学问题，系统性地从数据、优化和模态三个维度展开研究。大量的实验和理论结果不仅验证了所提出方法的有效性，更为解决无参考图像质量评价中的泛化难题提供了新的理论框架和技术路径，具有重要的理论意义和实际应用价值。

关键词：无参考图像质量评价，泛化能力，半监督学习，因果关系，多模态训练

Research on the Generalization of No-Reference Image Quality Assessment Models

Yan Zhong (Applied Mathmatics Major)
Supervised by Prof. Ming Jiang and A.P. Tingting Jiang

ABSTRACT

With the rapid innovation of digital imaging technology and the explosive growth of visual media applications, Image Quality Assessment (IQA) has developed into a core foundational technology in the fields of computer vision and multimedia processing. It plays an indispensable guiding role in algorithm optimization and system performance improvement for numerous tasks, such as image enhancement, compression, transmission, reconstruction, and even Artificial Intelligence Generated Content (AIGC). However, image quality is inherently a subjective perceptual experience of the human visual system, which poses severe challenges to constructing large-scale, highly consistent subjective evaluation datasets—including long cycles, high costs, and difficulty in ensuring annotation consistency. Consequently, the scale of existing datasets is severely limited, directly leading to the prevalent overfitting dilemma of current deep learning-based IQA models. These models exhibit significantly insufficient generalization capabilities across unseen distortion types, complex scene contents, and cross-dataset distributions. Particularly when facing unknown distortions (such as artifacts unique to emerging AI-generated models) and complex visual content in real open environments, model performance often degrades sharply. Therefore, breaking through the bottleneck of data annotation and constructing IQA models with strong generalization capabilities has become a key scientific issue driving the practical application of this technology.

Based on the amount of reference image information available to models during prediction, IQA methods can be systematically classified into three categories: Full-Reference Image Quality Assessment (FR-IQA), Reduced-Reference Image Quality Assessment (RR-IQA), and No-Reference Image Quality Assessment (NR-IQA, also known as Blind Image Quality Assessment, BIQA). Among them, NR-IQA methods do not rely on any form of original reference image information, making them more compatible with the stringent requirements of practical application scenarios such as surveillance, mobile photography, and social media content moderation, thus demonstrating extremely broad application prospects. In view of this, this thesis

focuses on the core goal of improving the generalization ability of NR-IQA models, conducting systematic research from three inherently interconnected innovative dimensions—data utilization, paradigm optimization, and modal fusion—and proposes a hierarchical, complementary new training framework for NR-IQA with high generalization performance.

At the data utilization level, to address the fundamental problem of scarce annotated data, this thesis proposes a semi-supervised NR-IQA framework based on confidence-quantified pseudo-label learning. Integrating entropy minimization with graph-structured label propagation, this framework generates pseudo-quality labels with reliability metrics for a large number of unannotated samples. By designing an iterative mechanism where model training and label optimization alternate, the framework can fully leverage both limited annotated data and massive unannotated data synergistically to achieve end-to-end semi-supervised training. This not only alleviates model overfitting to limited annotated data at the source but also significantly enhances the model's generalization ability to unseen distortions and contents by introducing rich visual structure information contained in unannotated data.

At the paradigm optimization level, this thesis reveals the inherent reasons restricting the generalization ability of NR-IQA models from the perspective of causal inference for the first time. By constructing a structural causal graph for the NR-IQA task, it clearly identifies that image content and distortion type, as confounding variables, introduce spurious correlations between distorted images and subjective scores. To block these confounding paths, this thesis proposes a causal learning-based training paradigm for NR-IQA, fundamentally transforming the model's learning objective from fitting the correlational relationship between images and quality scores to exploring their inherent causal relationship. This method innovatively designs a counterfactual data partitioning mechanism and performs causal intervention on the model in accordance with the backdoor criterion, thereby effectively decoupling the confounding effects of image content and distortion type on quality scores. It guides the model to learn essential and robust quality features independent of these confounders, ultimately achieving a qualitative leap in generalization performance.

At the modal fusion level, to further mine prior knowledge and enhance the model's quality perception ability, this thesis explores the great potential of vision-language models in NR-IQA tasks and proposes a multimodal NR-IQA method based on adaptive prompt learning. Built on the CLIP model pre-trained on large-scale image-text pairs, this method designs a conditional cross-modal fusion module. By computing visual quality prototypes and interacting with text descriptions, it achieves collaborative perception of visual and linguistic modalities at the deep semantic level. Addressing the drawbacks of traditional discrete prompt engineering

—being time-consuming, labor-intensive, and having weak generalization—this method introduces continuous learnable quality prompt vectors, enabling the model to adaptively capture contextual semantics highly relevant to image quality in an end-to-end manner. Additionally, this thesis constructs a triple evaluator framework integrating relative quality ranking, linear relationship measurement, and dataset-specific absolute quality assessment, and proposes a dual adaptive weight adjustment strategy to ensure the model' s stability and efficiency in mixed multi-dataset training.

The three proposed methods in this thesis have undergone comprehensive and rigorous experimental validation on multiple authoritative public IQA datasets, covering various complex scenarios such as real distortions, synthetic distortions, and cutting-edge AIGC-generated images. Extensive experimental results demonstrate that both in performance within the same dataset and in highly challenging cross-dataset and cross-distortion generalization tests, the three proposed methods outperform existing mainstream and state-of-the-art approaches, achieving leading advantages in key evaluation metrics such as SRCC and PLCC. This fully verifies their excellent performance in generalization ability, robustness, and practical application value.

In summary, this thesis systematically conducts research from three dimensions—data, optimization, and modality—centered on the core scientific question of "how to construct NR-IQA models with strong generalization capabilities." A large number of experimental and theoretical results not only verify the effectiveness of the proposed methods but also provide new theoretical frameworks and technical paths for solving the generalization problem in NR-IQA, holding significant theoretical significance and practical application value.

KEY WORDS: No-Reference Image Quality Assessment, Generalization Capability, Semi-Supervised Learning, Causality, Multi-Modal Training