Qeios

Peer Review

Review of: "Grounding-IQA: Multimodal Language Grounding Model for Image Quality Assessment"

Junyong You[1]

1. Norwegian Research Centre, Bergen, Norway

The paper introduces a new paradigm termed "Grounding-IQA," which aims to enhance Image Quality Assessment (IQA) methods by integrating multimodal referring and grounding capabilities. Traditional IQA approaches focus largely on holistic score predictions, and while newer learning-based and Multimodal Large Language Models (MLLM)-based methods have improved perception, they often fail to provide fine-grained, spatially-aware quality evaluations. Grounding-IQA addresses this limitation by requiring models not only to assess overall image quality but also to pinpoint critical regions or objects that influence the quality. Two subtasks are defined within this paradigm: GIQA-DES (providing detailed quality descriptions along with bounding boxes for affected regions) and GIQA-VQA (answering detailed, spatially-referenced questions about local image quality). To support these tasks, the authors construct a large dataset named GIQA-160K through an automated annotation pipeline. Experimental results demonstrate that fine-tuning various MLLM architectures on GIQA-160K improves their grounding and quality assessment capabilities. However, there are still some concerns:

1. Representation of the paper: The paper is not always explained well. Some parts are difficult to follow. There are also too many abbreviations.

2. Complexity of the Pipeline: While the automated annotation pipeline is clearly described, it is quite intricate. There may be concerns about the accuracy and consistency of the pipeline—especially in challenging real-world scenarios—though the authors do attempt to mitigate this via multiple refinement steps. A more thorough quantitative analysis of the pipeline’s annotation quality (e.g., comparing automatically generated boxes with a human-produced gold standard at scale) would strengthen the claims.

doi.org/10.32388/ACBTY8

3. Generality and Scope: The paper relies on MLLMs and IQA datasets that already have rich descriptions. One might wonder how well the proposed approach generalizes to scenarios where comprehensive textual descriptions of image quality are not readily available. The authors could discuss how the paradigm might be adapted to less richly annotated domains or non-photorealistic imagery.

4. Qualitative Analysis: Although some qualitative results are shown, more thorough analysis (e.g., systematic examples of correct and incorrect grounding, or showing cases where the model fails due to subtle quality issues) would provide deeper insight. Detailed error analyses could help readers understand current limitations and future directions.

Declarations

Potential competing interests: No potential competing interests to declare.