Performance Gaps Between 7B and 13B Vision-Language Models in Cross-Domain Object Grounding

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: Does the performance gap between 7B and 13B VLAs in object grounding persist when evaluated on cross-domain vision-language benchmarks such as LVIS or COCO-Text. We introduce InternVL 2.5, an advanced multimodal large language model (MLLM) series that builds upon InternVL 2.0, maintaining its core model architecture while introducing significant enhancements in training and testing strategies as well as data quality. In this work, we. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Expanding Performance Boundaries of Open-Source Multimodal Models with Model, Data, and Test-Time Scaling. Research question: Does the performance gap between 7B and 13B VLAs in object grounding persist when evaluated on cross-domain vision-language benchmarks such as LVIS or COCO-Text?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.8/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

References

- https://doi.org/10.48550/arxiv.2412.05271
- https://doi.org/10.48550/arxiv.2402.05935
- https://doi.org/10.48550/arxiv.2307.13721