

Scaling Multimodal Models for Cross-Domain Language Understanding Beyond MMBench

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: Can scaling the size of multimodal models (e.g., LLaVA-v1.5-13B vs. LLaVA-v1.5-7B) improve performance on cross-domain language understanding tasks beyond what MMBench measures. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Creation-MMBench: Assessing Context-Aware Creative Intelligence in MLLM. Research question: Can scaling the size of multimodal models (e.g., LLaVA-v1.5-13B vs. LLaVA-v1.5-7B) improve performance on cross-domain language understanding tasks beyond what MMBench measures?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.8/10.

3 Results

11 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.8/10.

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

- <http://arxiv.org/abs/2503.14478v2>
- <http://arxiv.org/abs/2207.08179v1>
- <http://arxiv.org/abs/2408.03361v7>