Scaling Laws of 7B and 13B VLA Models in LongNav-R1 on R2R-CE

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 2 peer-reviewed papers addressing the following research question: Does the inference efficiency (latency/throughput) of 7B and 13B VLA models scale linearly with instruction complexity in LongNav-R1 on R2R-CE, and how does this correlate with their grounding and. The ability to autonomously navigate and explore complex 3D environments in a purposeful manner, while integrating visual perception with natural language interaction in a human-like way, represents a longstanding research objective in Artificial Intelligence (AI) and embodied. 7 claims were extracted from source literature; 7 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Large-Scale Model-Enhanced Vision-Language Navigation: Recent Advances, Practical Applications, and Future Challenges. Research question: Does the inference efficiency (latency/throughput) of 7B and 13B VLA models scale linearly with instruction complexity in LongNav-R1 on R2R-CE, and how does this correlate with their grounding and path completion performance?.

2 Methodology

Systematic literature search across multiple databases yielded 2 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.0/10.

3 Results

2 papers retrieved. 7 claims extracted; 7 independently verified. Quality review score: 8.0/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
Vision-Language Navigation (VLN) has evolved from geometry-driven to semantics-driven and, more recently, knowledge-driv	✓	0.30
With the introduction of Large Language Models (LLMs) and Vision-Language Models (VLMs), recent methods have achieved su	✓	0.39
Existing surveys primarily focus on traditional VLN settings and offer limited coverage of LLM-based VLN, particularly i	✓	0.35
The paper summarizes edge deployment and implementation requirements, datasets, and evaluation protocols.	✓	0.18
The paper analyzes the task evolution from path-following to goal-oriented and demand-driven navigation.	✓	0.22
Key challenges in VLN include reasoning complexity, spatial cognition, real-time efficiency, robustness, and Sim2Real ad	✓	0.25
Future research directions in VLN include knowledge-enhanced navigation, multimodal integration, and world-model-based f	✓	0.27

References

- https://doi.org/10.1007/s10791-026-09977-z
- https://doi.org/10.3390/s26072022