VLA-Adapter Robustness Across Model Sizes in Cross-Domain Robotics Tasks

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does VLA-Adapter's robustness to novel robot setups (e.g., unseen grippers or environments) compare across different model sizes (e.g., 1B vs. 7B parameters) on the RoboBench suite, evaluated. 11 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: VLA-Adapter: An Effective Paradigm for Tiny-Scale Vision-Language-Action Model. Research question: How does VLA-Adapter's robustness to novel robot setups (e.g., unseen grippers or environments) compare across different model sizes (e.g., 1B vs. 7B parameters) on the RoboBench suite, evaluated using cross-domain adaptation metrics like domain shift sensitivity?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

14 papers retrieved. 11 claims extracted; 2 independently verified. Quality review score: 4.5/10.

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
VLA-Adapter achieves a success rate of 95.0% on LIBERO-Long with a B1 backbone, which is a 9.2% improvement over OpenVLA	×	0.06
VLA-Adapter achieves a success rate of 95.2% on LIBERO-Long with a B2 backbone, which is a 7.7% improvement over OpenVLA	×	0.06
VLA-Adapter achieves a success rate of 95.4% on LIBERO-Long with a B3 backbone, which is a 0.9% improvement over OpenVLA	×	0.06
VLA-Adapter uses 24.7GB of VRAM for training with an 8 batch size, which is 1/14$\times$ of OpenVLA-OFT.	×	0.04
VLA-Adapter achieves a throughput of 219.2Hz with an 8-dim chunk, which is 3$\times$ of OpenVLA-OFT.	×	0.05
VLA-Adapter maintains a performance of 97.3% on LIBERO, which is comparable to OpenVLA-OFT's 97.1%.	×	0.04
VLA-Adapter is effective when the backbone is frozen, with only the ActionQuery and Policy trained from scratch.	×	0.08
VLA-Adapter is compared with OpenVLA-OFT and SmolVLA in Table 3.	×	0.04
VLA-Adapter is designed to bridge the gap between vision-language representations and actions more effectively.	✓	0.16
Current VLA models typically require large-scale embodied data for pre-training Multimodal Large Language Models (MLLMs)	✓	0.17
VLA models face bottlenecks including reliance on large-scale VLMs, slow fine-tuning speed, high GPU memory consumption,	×	0.10

References

- http://arxiv.org/abs/2509.09372v2
- http://arxiv.org/abs/2508.13073v2
- http://arxiv.org/abs/2602.03973v1