Scaling Training Data Enhances LAP Generalization to Unseen Robot Morphologies

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: How does the scaling of LAP's training dataset size affect its generalization capabilities to unseen robot morphologies, as measured by success rate metrics on RoboStack tasks. 11 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: LAP: Language-Action Pre-Training Enables Zero-shot Cross-Embodiment Transfer. Research question: How does the scaling of LAP's training dataset size affect its generalization capabilities to unseen robot morphologies, as measured by success rate metrics on RoboStack tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.2/10.

3 Results

16 papers retrieved. 11 claims extracted; 2 independently verified. Quality review score: 5.2/10.

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
LAP-3B achieves performance comparable to the π0.5-DROID on the seen embodiment.	×	0.05
LAP-3B attains over 50% average zero-shot success across three previously unseen embodiments and six real-world manipula	✓	0.24
LAP-3B delivers approximately a 2$\times$ improvement over the strongest baselines in zero-shot cross-embodiment generalization	✓	0.15
All open-sourced VLAs collapse to zero success rate in zero-shot cross-embodiment generalization performance.	×	0.13
LAP-3B is evaluated across four robot embodiments, ten real-world manipulation tasks, and the LIBERO simulation benchmar	×	0.10
LAP-3B adopts a Mixture-of-Transformers architecture combining a LAP-trained VLM backbone with a lightweight flow-matchi	×	0.04
LAP-3B differs from π0.5 only in the action representation used to supervise the VLM (language-actions versus FAST token	×	0.07
The VLM backbone in LAP-3B is optimized to predict structured language-actions using the cross-entropy loss.	×	0.06
The action expert in LAP-3B predicts continuous action chunks at:t+H via a flow-matching objective.	×	0.06
The overall training objective for LAP-3B is L = LFM + λ LCE.	×	0.06
The VLM and action expert in LAP-3B communicate solely through cross-attention.	×	0.06

References

- http://arxiv.org/abs/2505.05753v2
- http://arxiv.org/abs/1903.05635v2
- http://arxiv.org/abs/2602.10556v2