

CLAM vs. SimCLR Policy Performance on BridgeData V2 Under Visual Noise

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 3 peer-reviewed papers addressing the following research question: How does the performance of CLAM-trained policies on the BridgeData V2 benchmark compare to those trained with self-supervised contrastive learning methods like SimCLR when evaluated under high. 16 claims were extracted from source literature; 7 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Vision-Language-Action Models for Robotics: A Review Towards Real-World Applications. Research question: How does the performance of CLAM-trained policies on the BridgeData V2 benchmark compare to those trained with self-supervised contrastive learning methods like SimCLR when evaluated under high levels of visual noise?.

2 Methodology

Systematic literature search across multiple databases yielded 3 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.4/10.

3 Results

3 papers retrieved. 16 claims extracted; 7 independently verified. Quality review score: 6.4/10.

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-Action (VLA) models unify vision, language, and action data at scale.	✓	0.26
Vision, language, and action data have traditionally been studied separately.	✓	0.24
VLA models aim to learn policies that generalise across diverse tasks, objects, embodiments, and environments.	✓	0.26
VLA models are expected to enable robots to solve novel downstream tasks with minimal or no additional task-specific data	✓	0.29
Previous surveys on this topic focus narrowly on action representations or high-level model architectures.	✓	0.22
This work offers a comprehensive, full-stack review integrating both software and hardware components of VLA systems.	✓	0.27
The paper provides a systematic review of VLA strategy and architectural transition.	✓	0.19
The paper reviews VLA architectures and building blocks.	×	0.12
The paper reviews modality-specific processing techniques for VLAs.	×	0.14
The paper reviews learning paradigms for VLAs.	×	0.09
The paper reviews commonly used robot platforms for VLA deployment.	×	0.14
The paper reviews data collection strategies for VLAs.	×	0.12
The paper reviews publicly available datasets for VLAs.	×	0.11
The paper reviews data augmentation methods for VLAs.	×	0.12
The paper reviews evaluation benchmarks for VLAs.	×	0.09
All references in the paper are categorized by training.	×	0.09

References

- <https://doi.org/10.1109/access.2025.3609980>

- <https://doi.org/10.48550/arxiv.2312.08782>
- <https://doi.org/10.3390/ai6070158>