GRPO Optimization Enhances Zero-Shot Generalization in LLaVA-UHD over PPO

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: Does GRPO optimization improve the zero-shot generalization of LLaVA-UHD on out-of-distribution multimodal benchmarks compared to PPO. Large vision-language models have achieved outstanding performance, but their size and computational requirements make their deployment on resource-constrained devices and time-sensitive tasks impractical. Model distillation, the process of creating smaller, faster models that. 10 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Distilling Large Vision-Language Model with Out-of-Distribution Generalizability. Research question: Does GRPO optimization improve the zero-shot generalization of LLaVA-UHD on out-of-distribution multimodal benchmarks compared to PPO?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

3 Results

13 papers retrieved. 10 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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
Better relative and local visual space coherence with the teacher (indicated by better Mrel and Mneigh) can lead to enha	×	0.08
The inclusion of Lvlprox further strengthens teacher-student V-L alignments.	×	0.04
Improved alignments are effectively transferred to unseen concepts, demonstrating that the student has acquired better V	×	0.06
Language representations should capture precise, fine-grained, and meaningful semantic attributes for effective student	×	0.08
Enriching semantic details of label descriptions by prompting LLMs can enhance student's OOD generalization ability.	×	0.11
ChatGPT can generate informative, fine-grained, and meaningful descriptions for target classes while keeping sequence le	×	0.03
The student model S is vision-only, i.e., S = Simg.	×	0.08
Students are trained from scratch to avoid label contamination.	×	0.03
The student is evaluated on Xid, along with zero-shot and few-shot generalization on Xood.	×	0.09
Finetuning achieves much higher performance on Xood than training-free retrieval.	×	0.02

References

- http://arxiv.org/abs/2309.04041v2
- http://arxiv.org/abs/2307.03135v3
- http://arxiv.org/abs/2408.03361v7