SED-SFT Mitigation of Mode Collapse in Multimodal Models on LVIS Benchmark

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: To what extent does SED-SFT mitigate mode collapse in multimodal models like Flamingo when fine-tuned on the LVIS object detection benchmark compared to CE loss. 12 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: SED-SFT: Selectively Encouraging Diversity in Supervised Fine-Tuning. Research question: To what extent does SED-SFT mitigate mode collapse in multimodal models like Flamingo when fine-tuned on the LVIS object detection benchmark compared to CE loss?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.4/10.

3 Results

11 papers retrieved. 12 claims extracted; 2 independently verified. Quality review score: 4.4/10.

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
SED-SFT significantly enhances generation diversity with a negligible computational overhead increase compared with CE l	✓	0.34
SED-SFT yields average improvements of 2.06 and 1.20 points in subsequent RL performance over standard CE-based baseline	✓	0.38
All experiments follow a standard SFT-then-RL pipeline to evaluate whether selectively encouraging diversity during SFT	×	0.14
We conduct experiments on two instruction-tuned backbones: Qwen2.5-Math-7B-Instruct and Llama-3.2-3B-Instruct.	×	0.11
For SFT, we sample 20,000 examples from the Micomind dataset.	×	0.04
For RL, we use the Math (Level 1) training split.	×	0.04
For SFT, we follow GEM's training setup, using a learning rate of 2 $\times$ 10-5 and DeepSpeed stage-2.	×	0.04
For RL, we apply GRPO implemented in the Verl framework with batch size 256; other hyperparameters use the default GRPO	×	0.03
We generate RL training samples by sampling each prompt 8 times with Qwen2.5-Math-7B-Instruct and filter out prompts whe	×	0.06
SED-SFT consistently outperforms the baselines.	×	0.10
The Cross-Entropy objective strongly drives the model policy $\pi$$\theta$ to quickly converge along the single correct path y, le	×	0.10
Updates should be strategically restricted to positions that potentially contain alternative reasoning branches to mitig	×	0.04

References

- http://arxiv.org/abs/2602.07464v1
- http://arxiv.org/abs/2504.09480v1
- http://arxiv.org/abs/2605.29400v1