One-to-Many Image-Text Relationships Enhance Vision-Language Model Robustness Against Adversarial Attacks

Assignee Research

June 3, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What is the impact of leveraging one-to-many image-text relationships on the retrieval accuracy of vision-language models when evaluated against gradient-based multimodal adversarial attacks. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 0.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Multimodal Adversarial Defense for Vision-Language Models by Leveraging One-To-Many Relationships. Research question: What is the impact of leveraging one-to-many image-text relationships on the retrieval accuracy of vision-language models when evaluated against gradient-based multimodal adversarial attacks?.

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 0.2/10.

3 Results

11 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 0.2/10.

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.

References

- http://arxiv.org/abs/2403.10883v2
- http://arxiv.org/abs/2405.18770v6
- http://arxiv.org/abs/2103.15670v3