Scaling One-to-Many Defense Strategies in Multimodal Models Without Retrieval Performance Loss

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: Can the one-to-many relationship-based defense strategy scale effectively to larger multimodal models (e.g., CLIP, BLIP) without degrading performance on standard retrieval metrics (R@1, R@5, R@10). 10 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Towards Better Instruction Following Retrieval Models. Research question: Can the one-to-many relationship-based defense strategy scale effectively to larger multimodal models (e.g., CLIP, BLIP) without degrading performance on standard retrieval metrics (R@1, R@5, R@10) on the SBU and MSCOCO benchmarks?.

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 4.2/10.

3 Results

16 papers retrieved. 10 claims extracted; 0 independently verified. Quality review score: 4.2/10.

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
The proposed marginal negative sampling strategy reduces computational complexity from combinatorial to linear, specific	×	0.04
Objective I (Univariate Conditional Modeling) models three conditional distributions: P(P|I, Q), P(I|P, Q), and P(IQ|P)	×	0.04
The multivariate objective (Eq. 7) formulates a more challenging ranking-based contrastive task by introducing a larger	×	0.03
The multivariate formulation exhibits greater robustness to competition-related issues compared to the univariate object	×	0.02
Table 2 presents experimental results comparing base models and their variants trained with InF-Embed on multiple instru	×	0.13
The FollowIR-7B model was trained on 50.7K instructions and 5.9M queries.	×	0.04
The Promptriever model was trained on 9.9K instructions, 9.9K queries, and 16.1K passages.	×	0.04
Adding InF-Embed to e5-base-v2 resulted in a performance change of -6.7 on one of the reported metrics.	×	0.04
Adding InF-Embed to e5-large-v2 improved a specific metric from 23.8 to 24.3.	×	0.05
Adding InF-Embed to ModernBERT-base resulted in a performance decrease of 5.8 on one of the reported metrics.	×	0.05

References

- http://arxiv.org/abs/2505.21439v1

- http://arxiv.org/abs/2405.18770v6
- http://arxiv.org/abs/2403.09513v1