

Multimodal Alignment Strategies for Cross-Lingual Retrieval in MSVD-Indonesian Adaptation

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 7 peer-reviewed papers addressing the following research question: How do different multimodal alignment strategies affect cross-lingual retrieval performance and robustness when adapting English pre-trained models to the MSVD-Indonesian benchmark. Multimodal learning on video and text has seen significant progress, particularly in tasks like text-to-video retrieval, video-to-text retrieval, and video captioning. However, most existing methods and datasets focus exclusively on English. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: MSVD-Indonesian: A Benchmark for Multimodal Video-Text Tasks in Indonesian. Research question: How do different multimodal alignment strategies affect cross-lingual retrieval performance and robustness when adapting English pre-trained models to the MSVD-Indonesian benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 7 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.2/10.

3 Results

7 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 5.2/10.

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/2506.15415v1>
- <http://arxiv.org/abs/2306.11341v2>
- <http://arxiv.org/abs/2508.11281v3>