

# Global-Local Contrastive Consistency Learning in Zero-Shot Text-Video Retrieval on MSVD

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does global-local contrastive consistency learning impact zero-shot text-video retrieval accuracy on the MSVD dataset compared to attention-based alignment methods. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Text-Video Retrieval With Global-Local Contrastive Consistency Learning. Research question: How does global-local contrastive consistency learning impact zero-shot text-video retrieval accuracy on the MSVD dataset compared to attention-based alignment methods?.

## 2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.0/10.

## 3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.0/10.

## 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/2405.12710v3>
- <http://arxiv.org/abs/2008.02531v2>
- <http://arxiv.org/abs/2605.17959v1>