

# Temporal Robustness of CLIP-Based and Diffusion-Based Video Encoders in Zero-Shot Retrieval

Assignee Research

June 8, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the difference in robustness to temporal perturbations between CLIP-based and diffusion-based video encoders when evaluated on the ActivityNet zero-shot retrieval task. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Toward a Holistic Evaluation of Robustness in CLIP Models. Research question: What is the difference in robustness to temporal perturbations between CLIP-based and diffusion-based video encoders when evaluated on the ActivityNet zero-shot retrieval task?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.3/10.

## 3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.3/10.

## 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/2410.01534v2>
- <http://arxiv.org/abs/2506.01061v3>
- <http://arxiv.org/abs/2503.11495v1>