

# MaxInfo Chunk-Based Key-Frame Selection vs. Uniform Sampling in Long-Form Video-Language Tasks

Assignee Research

June 6, 2026

## Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does the performance of MaxInfo’s Chunk-based key-frame selection compare to uniform sampling in downstream video-language tasks (e.g., QA, captioning) when evaluated on long-form video. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: T\*: Re-thinking Temporal Search for Long-Form Video Understanding. Research question: How does the performance of MaxInfo’s Chunk-based key-frame selection compare to uniform sampling in downstream video-language tasks (e.g., QA, captioning) when evaluated on long-form video benchmarks like ActivityNet or Charades-Ego?.

## 2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.5/10.

## 3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 7.5/10.

## 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/2410.09380v1>
- <http://arxiv.org/abs/2504.02259v3>
- <http://arxiv.org/abs/2507.00033v1>