Spatial-Temporal Corruption Robustness Benchmarks in Cross-Modal Generalization

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: To what extent do spatial-temporal corruption robustness benchmarks like Mini SSV2-C generalize to other multimodal domains (e.g., audio-visual or vision-language tasks), and how do models like. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Benchmarking Cross-Domain Audio-Visual Deception Detection. Research question: To what extent do spatial-temporal corruption robustness benchmarks like Mini SSV2-C generalize to other multimodal domains (e.g., audio-visual or vision-language tasks), and how do models like OpenPangu-7B-MLA perform in these cross-domain evaluations?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

10 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.8/10.

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/2506.00358v3
- http://arxiv.org/abs/2110.06513v2
- http://arxiv.org/abs/2405.06995v3