Spatio-Temporal Side Tuning Enhances Robustness in Multimodal Video-Language Models for Pedestrian Attribute Recognition

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 6 peer-reviewed papers addressing the following research question: How does spatio-temporal side tuning affect the robustness of multimodal video-language models compared to full fine-tuning when evaluated on the RICAP benchmark for pedestrian attribute recognition. 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: Spatio-Temporal Side Tuning Pre-trained Foundation Models for Video-based Pedestrian Attribute Recognition. Research question: How does spatio-temporal side tuning affect the robustness of multimodal video-language models compared to full fine-tuning when evaluated on the RICAP benchmark for pedestrian attribute recognition under motion blur conditions?.

2 Methodology

Systematic literature search across multiple databases yielded 6 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

6 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/2404.17929v1
- http://arxiv.org/abs/2505.11842v3
- http://arxiv.org/abs/2504.10018v2