Synthetic First-Person Video Fine-Tuning for Zero-Shot Action Recognition on AVA

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How does fine-tuning multimodal action recognition models on synthetic first-person video data affect zero-shot generalization performance on the AVA benchmark compared to real-world third-person. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: An Evaluation of Large Pre-Trained Models for Gesture Recognition using Synthetic Videos. Research question: How does fine-tuning multimodal action recognition models on synthetic first-person video data affect zero-shot generalization performance on the AVA benchmark compared to real-world third-person datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.2/10.

3 Results

4 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.2/10.

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/2506.20967v2
- http://arxiv.org/abs/2409.16382v1
- http://arxiv.org/abs/2410.02152v1