Robustness of Foundation Video Models Fine-Tuned on Synthetic Veo and Traditional Simulations

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: How does the robustness of foundation video models fine-tuned on synthetic Veo-generated environments compare to those fine-tuned on traditional simulation environments when evaluated on the. 14 claims were extracted from source literature; 8 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Rescaling Egocentric Vision: Collection, Pipeline and Challenges for EPIC-KITCHENS-100. Research question: How does the robustness of foundation video models fine-tuned on synthetic Veo-generated environments compare to those fine-tuned on traditional simulation environments when evaluated on the Charades-Ego benchmark under distribution shifts?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.7/10.

3 Results

9 papers retrieved. 14 claims extracted; 8 independently verified. Quality review score: 6.7/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
EPIC-KITCHENS-100 is the largest dataset in egocentric vision.	✓	0.26
EPIC-KITCHENS-100 contains 100 hours of video footage.	×	0.11
EPIC-KITCHENS-100 contains 20 million frames.	×	0.07
EPIC-KITCHENS-100 contains 90,000 annotated actions.	×	0.09
EPIC-KITCHENS-100 consists of 700 variable-length videos.	✓	0.17
EPIC-KITCHENS-100 captures activities in 45 distinct environments.	×	0.11
The data in EPIC-KITCHENS-100 was captured using head-mounted cameras.	✓	0.18
EPIC-KITCHENS-100 uses a novel annotation pipeline compared to the 2018 version by Damen et al.	×	0.13
The EPIC-KITCHENS-100 annotation pipeline yields 54% more actions per minute than the previous version.	✓	0.18
The EPIC-KITCHENS-100 annotation pipeline yields 128% more action segments than the previous version.	✓	0.18
EPIC-KITCHENS-100 includes footage collected in 2018 and new footage collected two years later.	✓	0.21
EPIC-KITCHENS-100 supports the challenge of evaluating whether models trained on 2018 data generalize to footage collect	✓	0.19
EPIC-KITCHENS-100 is aligned with six specific challenges: action recognition (full and weak supervision), action detect	✓	0.36
The paper provides baselines and evaluation metrics for each of the six defined challenges.	×	0.10

References

- https://doi.org/10.1561/2300000059
- https://doi.org/10.1007/s11263-021-01531-2

- https://doi.org/10.1109/access.2021.3140175