

Synthetic vs. Real Gesture Data Scalability in Video Encoder Edge Deployment

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the scalability of video encoders trained on synthetic vs. real gesture data affect their inference throughput (measured in FPS) and memory footprint when deployed on edge devices for. 10 claims were extracted from source literature; 1 was 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: Towards Synthetic Data Generation for Improved Pain Recognition in Videos under Patient Constraints. Research question: How does the scalability of video encoders trained on synthetic vs. real gesture data affect their inference throughput (measured in FPS) and memory footprint when deployed on edge devices for real-time applications?.

2 Methodology

Systematic literature search across multiple databases yielded 15 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

15 papers retrieved. 10 claims extracted; 1 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.

5 Extracted Claims

Claim	Verified	Confidence
The study generated a synthetic dataset containing 8,600 synthetic heads per perspective and texture.	×	0.08
The synthetic dataset generation pipeline captures 3D facial movements from a small set of participants and translates t	✓	0.17
The 'Mixed to Real' training configuration achieved an AUROC of 0.78.	×	0.04
The 'Synth to Real' training configuration achieved an AUROC of 0.581.	×	0.04
The 'Real to Real' training configuration achieved an AUROC of 0.741.	×	0.04
Training with 10 textures per patient resulted in an AUROC of 0.642.	×	0.03
Training with multiple views resulted in an AUROC of 0.653.	×	0.02
The experiments utilized a SlowFast-R50 network architecture.	×	0.03
The models were optimized for 100 epochs using SGD with a batch size of 64 and a learning rate of 0.01.	×	0.04
The study compares three training configurations: real data only, synthetic data exclusively, and a mix of both.	×	0.07

References

- <http://arxiv.org/abs/2409.16382v1>
- <http://arxiv.org/abs/2410.02152v1>
- <http://arxiv.org/abs/2108.08339v3>