Mixed-Dataset Pretraining with Factorized Objectives in Video-JEPA Transfer Learning

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the impact of mixed-dataset pretraining with factorized objectives on the transfer learning performance of Video-JEPA models when fine-tuned on low-data regimes of the AVA dataset. 8 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Fine-T2I: An Open, Large-Scale, and Diverse Dataset for High-Quality T2I Fine-Tuning. Research question: What is the impact of mixed-dataset pretraining with factorized objectives on the transfer learning performance of Video-JEPA models when fine-tuned on low-data regimes of the AVA dataset?.

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 4.5/10.

3 Results

10 papers retrieved. 8 claims extracted; 1 independently verified. Quality review score: 4.5/10.

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
Fine-tuning LlamaGen on Fine-T2I achieves an 80.7% win rate for visual quality compared to a counterpart without further	×	0.13
Fine-tuning LlamaGen on Fine-T2I achieves a 65.3% win rate for text-image alignment.	✓	0.19
Fine-T2I boosts the visual quality the most, possibly due to the extremely high aesthetic quality of the dataset.	×	0.10
Fine-tuning on Fine-T2I leads to clear preference gains on both LlamaGen and SD-XL models.	×	0.12
Models fine-tuned on Fine-T2I generate much better results than those fine-tuned on T2I-2M and BLIP3o-60k datasets on bo	×	0.12
GenEval benchmark results improve for both LlamaGen and SD-XL after fine-tuning on Fine-T2I.	×	0.10
Fine-T2I dataset contains extremely high aesthetic quality images.	×	0.14
Fine-T2I enhances controllability on standard automatic protocols as indicated by GenEval improvements.	×	0.04

References

- http://arxiv.org/abs/2605.17165v1
- http://arxiv.org/abs/2311.16109v1
- http://arxiv.org/abs/2602.09439v1