

# ImageNet-100 and Video Dataset Mixing Effects on JEPA Action Recognition Performance

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What is the impact of mixing ImageNet-100 with video datasets during pretraining on the downstream action recognition accuracy of JEPA models compared to single-dataset regimes on UCF-101. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Factorized Latent Dynamics for Video JEPA: An Empirical Study of Auxiliary Objectives. Research question: What is the impact of mixing ImageNet-100 with video datasets during pretraining on the downstream action recognition accuracy of JEPA models compared to single-dataset regimes on UCF-101?.

## 2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.3/10.

## 3 Results

11 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 4.3/10.

## 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/2605.17165v1>
- <http://arxiv.org/abs/2207.01398v2>
- <http://arxiv.org/abs/2605.12241v1>