Video-JEPA Auxiliary Objectives and Cross-Domain Transfer Performance on HMDB-51

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: What is the impact of auxiliary objective variants in Video-JEPA on cross-domain transfer performance from mixed pretraining (UCF-101 + Something-Something V2 + ImageNet-100) to the HMDB-51 benchmark. 14 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: SceneMotifCoder: Example-driven Visual Program Learning for Generating 3D Object Arrangements. Research question: What is the impact of auxiliary objective variants in Video-JEPA on cross-domain transfer performance from mixed pretraining (UCF-101 + Something-Something V2 + ImageNet-100) to the HMDB-51 benchmark?.

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 3.7/10.

3 Results

4 papers retrieved. 14 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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
GPTEval3D's pairwise comparisons against manual verification and user study results show an agreement of 88.5% and 85.2%	×	0.05
SMC better conforms to the input text description in terms of the number of objects and the way they are arranged compar	×	0.08
MVDream sometimes omits an entire object category and disregards conditions on the number of objects.	×	0.03
GraphDreamer performs worse than MVDream as it almost always disregards the specified object numbers and layout.	×	0.05
GraphDreamer's outputs suffer from the Janus problem, generating objects in irregular shapes and blending objects togeth	×	0.03
MVDream also suffers from the Janus problem but to a lesser extent.	×	0.00
Using only 1 to 3 examples, SMC excels at generating results that respect the input text description.	×	0.11
Most generated arrangements by SMC have the correct number of objects, and the objects are arranged according to the des	×	0.07
SMC successfully generalizes the underlying motif even when the example it learned from is significantly different from	×	0.07
LayoutPrompter is similar to the ablated SMC without meta-program as it also uses in-context learning with a few example	×	0.07
Table 2 shows that LayoutPrompter and the SMC ablations generate arrangements with lower text-asset alignment, plausibil	×	0.08
The learning phase of SMC involves extracting visual programs from a few (1-3) examples.	×	0.10
The inference phase of SMC involves retrieving the appropriate meta-program and using an LLM to determine appropriate ca	×	0.05
SMC uses a simple Domain Specific Language (DSL) based on Python for its visual programs.	×	0.07

References

- http://arxiv.org/abs/2503.14756v3
- http://arxiv.org/abs/2604.18946v1
- http://arxiv.org/abs/2408.02211v2