

SOVEREIGN: How does the inference efficiency of VGG-style very deep convolutional networks compare to shallow baselines w

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 29, 2026

Abstract

Interest in Artificial Intelligence (AI) and its applications has seen unprecedented growth in the last few years. This success can be partly attributed to the advancements made in the sub-fields of AI such as machine learning, computer vision, and natural language processing. Much of the growth in these fields has been made possible with deep learning, a sub-area of machine learning that uses artificial neural networks. This has created significant interest in the integration of vision and language. In this survey, we focus on ten prominent tasks that integrate language and vision by discussi

1 Introduction

Analysis of: Trends in Integration of Vision and Language Research: A Survey of Tasks, Datasets, and Methods. Research goal: How does the inference efficiency of VGG-style very deep convolutional networks compare to shallow baselines when processing high-resolution video frames with corresponding subtitles?.

2 Methodology

Multi-query arXiv search (1 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score < 6.5.

3 Results

2 papers retrieved. 0 claims extracted, 0 verified. Tribunal: 3.8/10 → REJECT (revision_round=0). Policy: ESCALATE_TO_OWNER.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

References

- <https://doi.org/10.1613/jair.1.11688>
- <https://doi.org/10.48550/arxiv.2205.10839>