SOVEREIGN: How do different architectural approaches for combining visual and linguistic features in multimodal models af

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 29, 2026

Abstract

A convolutional neural network (CNN) is one of the most significant networks in the deep learning field. Since CNN made impressive achievements in many areas, including but not limited to computer vision and natural language processing, it attracted much attention from both industry and academia in the past few years. The existing reviews mainly focus on CNN's applications in different scenarios without considering CNN from a general perspective, and some novel ideas proposed recently are not covered. In this review, we aim to provide some novel ideas and prospects in this fast-growing field.

1 Introduction

Analysis of: A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. Research goal: How do different architectural approaches for combining visual and linguistic features in multimodal models affect inference efficiency on Visual Genome tasks, measured by throughput (samples/second) and latency across GPU hardware platforms while maintaining competitive accuracy on scene understanding benchmarks?.

2 Methodology

Multi-query arXiv search (4 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

10 papers retrieved. 0 claims extracted, 0 verified. Tribunal: 6.5/10 $\rightarrow$ APPROVE (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/10.1109/tnnls.2021.3084827
- https://doi.org/10.1186/s40537-021-00444-8
- https://doi.org/10.1109/access.2021.3140175