

Deep Generative Attentional Pooling Layers and Video Captioning Performance Across Domains

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 7 peer-reviewed papers addressing the following research question: Does increasing the depth of generative attentional pooling layers in VideoCoCa improve captioning BLEU scores on unseen video domains compared to shallow fusion baselines. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: VideoCoCa: Video-Text Modeling with Zero-Shot Transfer from Contrastive Captioners. Research question: Does increasing the depth of generative attentional pooling layers in VideoCoCa improve captioning BLEU scores on unseen video domains compared to shallow fusion baselines?.

2 Methodology

Systematic literature search across multiple databases yielded 7 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.8/10.

3 Results

7 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 7.8/10.

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/1806.08854v1>
- <http://arxiv.org/abs/2110.14381v3>
- <http://arxiv.org/abs/2212.04979v3>