Benchmark archaeology: investigate Docvqa score discrepancy for Qwen2.5 — reported 14.1%–94.3% (spread 80.3pp) across 3

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: Benchmark archaeology: investigate Docvqa score discrepancy for Qwen2.5 — reported 14.1%–94.3% (spread 80.3pp) across 3 papers. Sources: 'VisionSelector: End-to-End Learnable Vis' (94.3%);. 10 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Zero-shot and Few-shot Learning with Knowledge Graphs: A Comprehensive Survey. Research question: Benchmark archaeology: investigate Docvqa score discrepancy for Qwen2.5 — reported 14.1%–94.3% (spread 80.3pp) across 3 papers. Sources: 'VisionSelector: End-to-End Learnable Vis' (94.3%); 'VisionSelector: End-to-End Learnable Vis' (94.3%); 'DocHop-QA: Towards Multi-Hop Reasoning o' (14.1%). Identify evaluation protocol differences (few-shot, prompting, preprocessing)..

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 3.8/10.

3 Results

4 papers retrieved. 10 claims extracted; 0 independently verified. Quality review score: 3.8/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
The survey covers 50 papers on KG-aware ZSL and 46 papers on KG-aware FSL.	×	0.12
The survey paper [25] includes 19 papers about KG-aware ZSL and 21 papers about KG-aware FSL.	×	0.11
The benchmarking paper [26] reviews around 10 ZSL methods that mainly utilize class attribute and text information as th	×	0.06
The survey paper [6] does not cover state-of-the-art ZSL methods proposed in recent 3 years nor KG-aware ZSL methods.	×	0.07
The survey is suitable for AI researchers, especially those who are to enter the domain of ML with sample shortage, thos	×	0.07
The survey provides a more fine-grained method categorization and additional technical analysis on KGs and their constru	×	0.10
The survey paper [8] briefly categorizes different external knowledge used in ZSL with incomplete reviews on KG-aware ZS	×	0.11
The survey paper [7] systematically reviews the ZSL methods by 2019.	×	0.05
The survey paper [15] systematically reviews the FSL methods by 2020.	×	0.06
The survey paper [6] focuses on their evaluation and result comparison on image classification task.	×	0.07

References

- http://arxiv.org/abs/2112.10006v6
- http://arxiv.org/abs/2311.14544v1
- http://arxiv.org/abs/2308.10783v2