Multimodal Data Augmentation Effects on Zero-Shot TabMWP Accuracy

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does the integration of multimodal data augmentation during pretraining affect zero-shot accuracy on TabMWP compared to text-only pretraining. 6 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: No "Zero-Shot" Without Exponential Data: Pretraining Concept Frequency Determines Multimodal Model Performance. Research question: How does the integration of multimodal data augmentation during pretraining affect zero-shot accuracy on TabMWP compared to text-only pretraining?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.2/10.

3 Results

14 papers retrieved. 6 claims extracted; 1 independently verified. Quality review score: 5.2/10.

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 concept acquisition process for classification includes concepts like 'dog', 'man', 'hat', 'grass', 'goldfish', 'sti	×	0.03
The concept frequency estimation for datasets CC-3M, CC-12M, YFCC-15M, and LAION-400M is provided for both classificatio	✓	0.23
Testing with synthetic pretraining concept distributions shows different performance metrics for classification and retr	×	0.11
The slope values for classification and retrieval tasks are 1.58 and 0.60, respectively.	×	0.02
The methodology includes references to specific models and datasets such as v0.1, Deepfloyd IF, and Semdedup.	×	0.03
The paper cites various studies and technical reports, including those by Achiam et al. (2023), Akyrek et al. (2022), a	×	0.04

References

- http://arxiv.org/abs/2407.17856v4
- http://arxiv.org/abs/2404.04125v3
- http://arxiv.org/abs/2501.13372v1