INT4 Quantization and Robustness of Multimodal Models on VQA-v2 Under Noise Conditions

Assignee Research

June 5, 2026

Abstract

This report synthesises findings from 2 peer-reviewed papers addressing the following research question: How does INT4 quantization affect the robustness of multimodal models on the VQA-v2 dataset under varying noise conditions compared to FP16 precision. 12 claims were extracted from source literature; 10 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey. Research question: How does INT4 quantization affect the robustness of multimodal models on the VQA-v2 dataset under varying noise conditions compared to FP16 precision?.

2 Methodology

Systematic literature search across multiple databases yielded 2 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.7/10.

3 Results

2 papers retrieved. 12 claims extracted; 10 independently verified. Quality review score: 7.7/10.

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
Large models have enabled remarkable achievements across various tasks.	✓	0.17
Large models often consist of billions of parameters.	×	0.12
Large models require vast amounts of computational resources for execution.	✓	0.24
The expansive scale and computational demands of large models pose considerable challenges when customizing them for par	✓	0.33
Parameter Efficient Fine-Tuning (PEFT) provides a practical solution by efficiently adjusting the large models over the	✓	0.40
PEFT refers to the process of adjusting the parameters of a pre-trained large model to adapt it to a specific task or do	✓	0.40
PEFT is particularly important when dealing with large-scale language models with high parameter counts.	✓	0.28
Fine-tuning large models from scratch can be computationally expensive and resource-intensive.	✓	0.27
Fine-tuning large models poses considerable challenges in the supporting system platform design.	✓	0.25
This survey presents comprehensive studies of various PEFT algorithms, examining their performance and computational ove	✓	0.26
This survey provides an overview of applications developed using different PEFT algorithms.	✓	0.22
This survey discusses common techniques employed in PEFT.	×	0.12

References

- https://doi.org/10.48550/arxiv.2403.14608
- https://openalex.org/W7127203421