Gradient Accumulation Effects on Critical Batch Size in 1B-Parameter Model Pre-Training

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: What is the impact of gradient accumulation on critical batch size thresholds for 1B-parameter models during pre-training on diagram-based coding tasks. Training large-scale models under given resources requires careful design of parallelism strategies. In particular, the efficiency notion of critical batch size (CBS), concerning the compromise between time and compute, marks the threshold beyond which greater data parallelism. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 1.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: How Does Critical Batch Size Scale in Pre-training?. Research question: What is the impact of gradient accumulation on critical batch size thresholds for 1B-parameter models during pre-training on diagram-based coding tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 1.5/10.

3 Results

8 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 1.5/10.

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/2412.01505v1
- http://arxiv.org/abs/2410.21676v4
- http://arxiv.org/abs/2507.07101v4