Gradient-Based Sparsification Trade-offs in Contrastive Learning Robustness and Efficiency

Assignee Research

June 2, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: What is the trade-off between inference efficiency (e.g., throughput in tokens/sec) and model robustness (e.g., accuracy on noisy CIFAR-10-C) when applying gradient-based sparsification to. Abstract Realizing increasingly complex artificial intelligence (AI) functionalities directly on edge devices calls for unprecedented energy efficiency of edge hardware. Compute-in-memory (CIM) based on resistive random-access memory (RRAM) 1 promises to meet such demand by. 6 claims were extracted from source literature; 6 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A compute-in-memory chip based on resistive random-access memory. Research question: What is the trade-off between inference efficiency (e.g., throughput in tokens/sec) and model robustness (e.g., accuracy on noisy CIFAR-10-C) when applying gradient-based sparsification to contrastive learning models?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.2/10.

3 Results

15 papers retrieved. 6 claims extracted; 6 independently verified. Quality review score: 7.2/10.

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
Compute-in-memory (CIM) based on resistive random-access memory (RRAM) promises to meet the demand for unprecedented ene	✓	0.54
Recent studies have demonstrated in-memory matrix-vector multiplication on fully integrated RRAM-CIM hardware.	✓	0.32
It remains a goal for a RRAM-CIM chip to simultaneously deliver high energy efficiency, versatility to support diverse m	✓	0.41
Efficiency, versatility and accuracy are all indispensable for broad adoption of the technology.	✓	0.22
The inter-related trade-offs among efficiency, versatility and accuracy cannot be addressed by isolated improvements on	✓	0.29
NeuRRAM is a RRAM-based CIM chip that simultaneously delivers versatility in reconfiguring CIM cores for diverse model a	✓	0.53

References

- https://doi.org/10.1007/s10994-019-05855-6
- https://doi.org/10.1038/s41586-022-04992-8
- https://doi.org/10.1186/s40537-021-00492-0