Pruning Strategies in ARS: Accuracy-Throughput Trade-offs on GSM-8K and MATH

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How do alternative pruning strategies in ARS compare to the proposed method in terms of accuracy-throughput trade-offs on the GSM-8K and MATH benchmarks. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Exploring 3D Dataset Pruning. Research question: How do alternative pruning strategies in ARS compare to the proposed method in terms of accuracy-throughput trade-offs on the GSM-8K and MATH benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

3 Results

13 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
The training pipeline is conducted in three phases, all using a fixed random seed of 42.	×	0.04
The data augmentation protocols outlined in the original papers for each model are strictly followed.	×	0.02
Table 5 presents overall accuracy (OA) and mean accuracy (mAcc) on ShapeNet55, ModelNet40, and ScanObjectNN.	×	0.13
The Many Medium Few benchmark table shows class rank distribution with counts of 6000, 4000, and 2000.	×	0.03
The Uniform (80) benchmark table shows performance metrics for different coreset sizes.	×	0.04
The Figur benchmark table shows changes in overall accuracy (OA) and mean accuracy (mAcc) for different coreset sizes.	×	0.13
The code is available on Github.	×	0.09
The method unfolds in three parts: targeting prior mismatch bias, minimizing representation error, and introducing a ste	×	0.08
The theoretical analysis guides the method, which is shown in Fig 2.	×	0.03
The method decomposes the dilemma by decoupling structural likelihood from class priors during post-training.	×	0.03
Reweighting scales the posterior, implicitly forcing the model to fit a distorted prior.	×	0.02
The pruning weights w define the target distribution; thus, any slight misalignment in w shifts the global optimum.	×	0.05
log p(x | y) describes the semantic geometry of the data manifold, which is shared regardless of the evaluation.	×	0.03

References

- http://arxiv.org/abs/2603.00651v1
- http://arxiv.org/abs/2604.25926v1
- http://arxiv.org/abs/1807.10577v1