Gradient Checkpointing and Memory-Efficient Methods in Vision GNN Training Trade-offs

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does gradient checkpointing impact the accuracy-throughput tradeoff when training Vision GNNs on multimodal datasets compared to traditional memory-efficient methods like gradient accumulation. Graph Neural Networks (GNNs) have emerged as an efficient alternative to convolutional approaches for vision tasks such as image classification, leveraging patch-based representations instead of raw pixels. These methods construct graphs where image patches serve as nodes, and. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Explaining Vision GNNs: A Semantic and Visual Analysis of Graph-based Image Classification. Research question: How does gradient checkpointing impact the accuracy-throughput tradeoff when training Vision GNNs on multimodal datasets compared to traditional memory-efficient methods like gradient accumulation?.

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 4.0/10.

3 Results

14 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 4.0/10.

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 ViG-Small variant of the Vision GNN achieves a classification accuracy of 68.6% on the ImageNet validation set in th	×	0.05
In the early layers (1-4) of the ViG-Small model, visual similarity (Sl_vis) is greater than 0.6 and embedding similarit	×	0.04
In the final layers of the ViG-Small model, spatial distances (Dl) increase to approximately 8.8 and visual similarity (	×	0.04
Embedding similarity (Sl_emb) increases significantly in the final two layers of the ViG-Small model, indicating the eme	×	0.03
The graph modularity scores (Ql) for adversarial images in the ImageNet-a dataset are significantly lower (starting at 0	×	0.04
The top-1 classification accuracy of the ViG-Small model drops from 68.6% on the ImageNet validation set to 2.7% on the	×	0.02
The correct class probability (pl) falls from 0.486 in the final layers on the ImageNet validation set to 0.018 on the I	×	0.02
The ViG-Small model's graph structure evolves through intermediate layers, showing a trade-off between local and global	×	0.05
The ViG-Small model's dynamic graph construction provides an interpretable view of how the model processes information,	×	0.06
The ViG-Small model's graph structure mirrors the hierarchical feature learning observed in traditional CNNs, with a tra	×	0.07
The ViG-Small model's graph structure adapts to the input content, providing a self-documenting receptive field.	×	0.04

References

- http://arxiv.org/abs/2406.00552v4
- http://arxiv.org/abs/2504.19682v1
- http://arxiv.org/abs/2604.00086v1