SOVEREIGN: Introducing Routing Functions to Vision-Language Parameter-Efficient Fine-Tuning

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 27, 2026

Abstract

Mainstream parameter-efficient fine-tuning (PEFT) methods, such as LoRA or Adapter, project a model's hidden states to a lower dimension, allowing pre-trained models to adapt to new data through this low-rank bottleneck. However, PEFT tasks involving multiple modalities, like vision-language (VL) tasks, require not only adaptation to new data but also learning the relationship between different modalities. Targeting at VL PEFT tasks, we propose a family of operations, called routing functions, to enhance VL alignment in the low-rank bottlenecks. These feature routing functions adopt linear ope

1 Introduction

Analysis of: Introducing Routing Functions to Vision-Language Parameter-Efficient Fine-Tuning with Low-Rank Bottlenecks. Research goal: What is the impact of dynamic routing overhead on inference latency when adapting vision-language models to new domains, measured through wall-clock time comparisons on ImageNet-1K and COCO captioning benchmarks?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

11 papers retrieved. 3 claims extracted, 3 verified. Tribunal: 8.3/10 $\rightarrow$ APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
Routing functions adopt linear operations and do not introduce new trainable parameters.	✓	0.28
Routing functions significantly improve performance of the original PEFT methods, achieving over 20% improvement on VQAv	✓	0.37
When fine-tuning a pre-trained multimodal model such as CLIP-BART, smaller but consistent improvements are observed acro	✓	0.34

References

- http://arxiv.org/abs/2507.22398v3
- http://arxiv.org/abs/2403.09377v2
- http://arxiv.org/abs/2602.06370v1