

EfficientViT Integration in PaLI Balances Image-Text Retrieval Accuracy and Model Efficiency

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does the integration of EfficientViT into PaLI affect the trade-off between image-text retrieval accuracy and model size, as measured by retrieval metrics (e.g., R@1, R@5) on COCO and Flickr30k. 8 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: FiCo-ITR: bridging fine-grained and coarse-grained image-text retrieval for comparative performance analysis. Research question: How does the integration of EfficientViT into PaLI affect the trade-off between image-text retrieval accuracy and model size, as measured by retrieval metrics (e.g., R@1, R@5) on COCO and Flickr30k benchmarks while keeping the text-only performance stable on LAVIS?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.2/10.

3 Results

12 papers retrieved. 8 claims extracted; 2 independently verified. Quality review score: 5.2/10.

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
FiCo-ITR library quantifies computational costs of large-scale ITR, revealing trade-offs between retrieval performance a	✓	0.16
FG methods aim to map visual and textual information to a joint space where relevant samples are aligned at the instance	×	0.05
CG search employs bitwise hash codes to find retrieval samples that are more broadly relevant to the query instead of ex	×	0.03
The broader search criteria of CG search allows for more efficient computational costs compared to FG search.	×	0.05
Direct empirical comparative evaluations of recent representative FG and CG ITR models are lacking in the literature.	✓	0.20
Traditional ITR benchmark datasets such as Flickr30K and MS-COCO are small compared to real-world applications.	×	0.08
FiCo-ITR library is implemented and distributed to bridge methodological differences between CG and FG ITR subfields.	×	0.11
FiCo-ITR library is available in the Python Package Index (PyPI) and GitHub.	×	0.06

References

- <http://arxiv.org/abs/2407.20114v3>
- <http://arxiv.org/abs/2506.16691v1>
- <http://arxiv.org/abs/2109.10739v1>