Contrastive Fine-Tuning of Vision-Language Models for Zero-Shot Remote Sensing Classification

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the effect of fine-tuning pre-trained VLMs using different contrastive learning objectives on their zero-shot classification accuracy across remote sensing datasets like NWPU-RESISC45 and AID. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Vision-Language Modeling Meets Remote Sensing: Models, Datasets and Perspectives. Research question: What is the effect of fine-tuning pre-trained VLMs using different contrastive learning objectives on their zero-shot classification accuracy across remote sensing datasets like NWPU-RESISC45 and AID?.

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 5.3/10.

3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 5.3/10.

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.

References

- http://arxiv.org/abs/2306.11029v4
- http://arxiv.org/abs/2505.14361v1
- http://arxiv.org/abs/2508.11919v3