

# Adversarial Robustness of ALIGN-Based Models on RSICD Under Varying Synthetic-to-Real Training Data Ratios

Assignee Research

June 12, 2026

## Abstract

Remote sensing images contain a wealth of Earth-observation information. Efficient extraction and application of hidden knowledge from these images will greatly promote the development of resource and environment monitoring, urban planning and other related fields. Remote sensing image caption (RSIC) involves obtaining textual descriptions from remote sensing images through accurately capturing and describing the semantic-level relationships between objects and attributes in the images. However, there is currently no comprehensive review summarizing the progress in RSIC based on deep learning.

## 1 Introduction

This paper examines: A Review of Deep Learning-Based Remote Sensing Image Caption: Methods, Models, Comparisons and Future Directions. Research question: What is the impact of varying the ratio of synthetic to real training data on the adversarial robustness of ALIGN-based models evaluated on RSICD?.

## 2 Methodology

Systematic literature search across multiple databases yielded 3 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.2/10.

## 3 Results

3 papers retrieved. 12 claims extracted; 10 independently verified. Quality review score: 7.2/10.

## 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 |
|--|----------|------------|
| Remote sensing images contain Earth-observation information.   | ✓        | 0.22       |
| Efficient extraction of hidden knowledge from remote sensing images promotes the development of resource and environment | ✓        | 0.23       |
| Efficient extraction of hidden knowledge from remote sensing images promotes the development of urban planning.          | ✓        | 0.20       |
| Remote sensing image caption (RSIC) involves obtaining textual descriptions from remote sensing images.                  | ✓        | 0.34       |
| RSIC requires accurately capturing and describing semantic-level relationships between objects and attributes in images. | ✓        | 0.23       |
| There is currently no comprehensive review summarizing the progress in RSIC based on deep learning prior to this paper.  | ✓        | 0.30       |
| The review covers six key aspects: encoder-decoder framework, attention mechanism, reinforcement learning, learning with | ✓        | 0.35       |
| The paper provides an explanation of datasets and evaluation metrics for RSIC.   | ✓        | 0.16       |
| The paper compares and analyzes the results of the latest RSIC models.   | ×        | 0.09       |
| The paper analyzes the pros and cons of different deep learning methods for RSIC.  | ✓        | 0.22       |
| The paper suggests future directions for RSIC.   | ×        | 0.15       |
| The primary objective of the review is to offer researchers a more profound understanding of RSIC.                       | ✓        | 0.24       |

## References

- <https://openalex.org/W7161452288>
- <https://doi.org/10.20944/preprints202505.1993.v1>
- <https://doi.org/10.3390/rs16214113>