

# Diffusion Step Count Effects on Vision-Language Model Performance in Remote Sensing Detection

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the impact of varying diffusion process step counts in synthetic data generation on the downstream performance of vision-language models in remote sensing object detection tasks. 15 claims were extracted from source literature; 5 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: AeroGen: Enhancing Remote Sensing Object Detection with Diffusion-Driven Data Generation. Research question: What is the impact of varying diffusion process step counts in synthetic data generation on the downstream performance of vision-language models in remote sensing object detection tasks?.

## 2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.7/10.

## 3 Results

10 papers retrieved. 15 claims extracted; 5 independently verified. Quality review score: 5.7/10.

## 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 synthetic data produced by AeroGen are of high quality and diversity.	✓	0.22
The synthetic RSIOD data can significantly improve the detection performance of existing RSIOD models, i.e., the mAP met	✓	0.38
AeroGen is capable of generating high-quality images from text conditions and achieving significant progress in layout c	×	0.08
The application of generative models in remote sensing images is usually combined with specific tasks, such as change de	×	0.11
Utilizing generative diffusion models to fit the distribution of existing datasets and generate new samples to enhance t	×	0.09
AeroGen is the first generative model to support layout conditional control for both horizontal and rotated bounding box	×	0.11
The proposed AeroGen model allows for the specification of layout prior conditions with horizontal and rotated bounding	✓	0.17
The end-to-end data augmentation framework proposed by AeroGen eliminates the need for additional instance-pasting proce	×	0.12
The diversity-conditioned generator and generation quality evaluation in AeroGen enhance the diversity and quality of th	✓	0.18
The novel filtering mechanism in the data augmentation pipeline of AeroGen selects high-quality synthetic training image	✓	0.16
AeroGen achieves an FID of 38.57, CAS of 76.84, and YOLO Score of 29.8/54.2/31.6 on the DIOR dataset with HBB modality.	×	0.02
AeroGen achieves an FID of 35.07, CAS of 74.13, and YOLO Score of 29.6/57.6/32.0 on the DIOR-R dataset with OBB modality	×	0.02
AeroGen achieves an FID of 45.86, CAS of 51.19, and YOLO Score of 27.1/51.0/27.6 on the HRSC dataset with OBB modality.	×	0.01
AeroGen achieves an mAP of 41.69 and mAP50 of 64.12 with the strategy involving Synthesis, Filter, Augment, Semantic, an	×	0.02
AeroGen achieves an FID of 38.57, CAS of 76.84, and YOLO Score of 29.8/54.2/31.64with LMA and DCA enabled.	×	0.01

## References

- <http://arxiv.org/abs/2411.15497v3>
- <http://arxiv.org/abs/2505.14361v1>
- <http://arxiv.org/abs/2504.09480v1>