Past Reviews

Reviewer #2 Review Attachment #1 (2026-06-12)

Lan Pham's Review for 260402G:

Appropriateness for JARS:
Good

Quality of writing:
Good

Clarity:
Satisfactory

Conciseness:
Satisfactory

References to literature:
Good

Relevance of figures:
Satisfactory

Originality:
Good

Significance of results:
Good

Technical accuracy of results:
Good

Rigor:
Good

Detail level of procedures:
Good

Substantiation of conclusions:
Satisfactory

Recommendation:
Can be made acceptable, dependent on revision

Exceptional Paper:
No

Reviewer Comments Required:
The primary objective of this manuscript is to conduct a systematic qualitative and quantitative evaluation of eight CAM-based XAI methods using YOLOv5 on the DIOR remote sensing dataset. The authors achieved this goal by introducing an adapted masking framework (M1, M2) and an entropy-based metric (M3)
ABSTRACT
The abstract lacks a clear take-home message. Since the study benchmarks eight CAM variants, the authors should explicitly state the main recommendation derived from the results and highlight its practical significance for remote sensing applications. The final part of the abstract should provide a concise, evidence-based conclusion that clarifies how the findings support more reliable and interpretable decision-making in safety-critical scenarios.
1. INTRODUCTION:
1. YOLOv5 is quite old now. The author should comment on why they didn't use newer versions like YOLOv8/v10.
2. In the Abstract (Lines 10-11), the authors explicitly state that the core focus and motivation of this study is to evaluate CAM-based XAI methods under "varying scene complexity and target density" (line 10) via systematic qualitative and quantitative (Line 11) analysis. However, in Section 1 (Introduction), there is a significant baseline literature gap. The authors completely fail to discuss, define, or contextualize what constitutes "scene complexity" or "target density" within the remote sensing domain
2. MATERIALS AND METHOD
The authors should introduce a new figure at section 2. For example "Figure X: Overview of the proposed CAM evaluation and benchmarking workflow". This diagram should visually map out the complete pipeline from the original input image to the final quantitative metric computation. This will significantly improve the readability, transparency, and overall structure of the methodology section.
Line 114-115: Figure 1 appears to contain a minor labeling inconsistency in the final output dimensions of the prediction heads. The authors are kindly encouraged to double-check and revise the dimension labels in Figure 1 to ensure strict alignment with the standard mathematical configuration of the underlying detector. The numbers 64, 128, and 256 seem to correspond to the feature map depths of the intermediate layers (such as within the Neck or Backbone components) rather than the final bounding box and class prediction outputs.
3. RESULT

Line 208 - 218: The authors should relocate the Dataset (3.1 section) description Section 2 "Materials and Methods". Additionally, the dataset should be described in more detail and link to abstract to achieve "varying scene complexity and target density".
Fig 6,6 have low resolution and small legend, axis size. The authors should replace them with high-resolution vector graphics and increase the font size.
Line 522-524: "The substantially higher runtime of Score-CAM...". The authors should correct their comments because Score-CAM takes 841.61 ms to run on the CPU, while other algorithms run on the GPU (4-6ms). This is unfair because the Score-CAM architecture is perfectly programmable on the GPU. The authors need to adjust their assessment or implement it on a GPU for an objective comparison.
Line 361-385. In Section 3.5.1 and Figure 14, the authors visually demonstrate that Eigen-CAM fails to capture multiple instances of the same object in a scene, whereas Score-CAM captures them broadly. Given that the DIOR dataset heavily features multi-instance scenes (e.g., parking lots full of vehicles, harbors full of ships), this is a fundamental flaw of Eigen-CAM. However, this critical drawback is completely buried in the global averages of Table 1. The authors should separate their quantitative evaluation (Table 1) into two sub-categories: "Single-instance images" vs. "Multi-instance images" to explicitly demonstrate this gap via metrics.
4. CONCLUSION
The authors should add a dedicated paragraph in conlusion section clearly discussing the "Limitations of the Study" before introducing the future research paths.

Comments to the Editor (Confidential):
I have used Chat GPT for grammatical and stylistic assistance