

Introduction

Good description of Cross-modality Data Augmentation for Aerial Object Detection with Representation Learning. With an introduction of an innovative solution to improve, object classification by means of Data Augmentation. They describe the usage of 2 recordings (RGB and IR)

is the research design appropriate?

The paper shows a very good and clear problem description, namely the need for effective data augmentation. I like the novel approach to combine data augmentation in both feature space and data space. In addition, the established methods (CutOut, MixUp, Mosaic) which outperform existing results.

However, there are some points where I see a need for improvement:

- only one source of data is used "DroneVehicle"
- For generalisation the ODD may be too small. No tests done with adverse weather conditions or any other situations with higher noise level
- there is a risk of overfitting: large-scale mosaic method accelerates convergence but also expanding the feature space. Further analysis of overfitting in different scenarios would be beneficial.

You need to analyze the real-world applicability.

Are the methods adequately described?

I see a clear description of the methods; the approach is well divided into distinct modules and I can see the novelty in the paper. So the techniques which are used are well applied – technically clear. But there is room for improvement, namely implementation details should be more explicit to enhance reproducibility. This may be a consequence of the research design. Broader ODD

Are the results clearly presented?

The data are well presented by comparing them with most of the baseline methods and state of the art data augmentation techniques (CutOut, MixUp, Mosaic). Some are missing: cross modality of feature space augmentation techniques.

In addition some tables (e.g. 4) show a lack of sufficient textual explanation. Improvement would improve readability.

Are the conclusions supported by the results?

The results support the conclusions, particularly in terms of method effectiveness and the complementarity of the approaches.

But as earlier said I see the need for additional dataset (broader ODD).

- address scalability and limitations
- include statistical rigour
- provide more qualitative evidence for improvements