

Comments to the Authors

This work constructs a 100000-level FASDD based on multi-source heterogeneous flame and smoke images, which provides a challenging benchmark to drive the continuous evolution of fire detection models. However, there are still some issues that need to be addressed to demonstrate the reliability of the dataset.

1. In terms of sample annotation, what are the differences between satellite images with different spatial resolutions in the remote sensing field and RGB images in the CV field? Can different annotation formats be integrated or unified into one annotation format for the widespread use of FASDD?
2. How to consider the difference in spatial resolution between Sentinel-2 L1C and Landsat-8 TOA? How to solve the detection accuracy of targets with different sizes? What are the differences in detection accuracy between large and small targets during sample annotation, model training, and inference stages?
3. How to consider the difference in radiometric resolution between RGB and satellite images? How robust is the detection algorithm for data with different radiation resolutions? Please confirm through ablation experiments.
4. The FASDD holds rich variations in image size, resolution, illumination, scenario, image range, viewing angle, platform, and data source. How to consider data at different scales for various deep learning models? Please choose the latest model in the CV field to verify the reliability of FASDD and the superiority of Transformer-based models, such as DETR, etc.
5. What may be the reason for the poor validation performance of the transformer-based model compared to YOLOv5x?
6. The annotation and partition ratio of samples directly affect the accuracy of deep learning models. Please supplement ablation experiments to demonstrate the reliability and universality of FASDD datasets under different training sample ratios.
7. In the case of small samples, the Transformer-based model has poor convergence. The linear expansion of the transformer leads to a sharp increase in parameter size and insufficient local feature extraction. How to solve the problem of insufficient local feature extraction in transformer and poor target detection performance in FASDD?
8. Is the low accuracy of the Transformer-based model caused by overfitting? Please provide the

training accuracy and validation accuracy curve of the Transformer-based model.

9. Please increase the types and quantity of deep learning algorithms (each model type contains at least two algorithms) to fully validate the universality and reliability of the dataset.