Earth Syst. Sci. Data Discuss., referee comment RC2
https://doi.org/10.5194/essd-2022-394-RC2, 2023

Comment on essd-2022-394

Anonymous Referee #2

Referee comment on "FASDD: An Open-access 100,000-level Flame and Smoke Detection Dataset for Deep Learning in Fire Detection" by Ming Wang et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-394-RC2, 2023

The authors propose a new dataset for fire detection made by combining different types of images. They call it the Flame and Smoke detection dataset (FASDD) and claim that it is the largest set of data on fires. It combines both images of flames and of smokes which are elements associated with the presence of an existing fire. After reviewing precedent datasets, the authors start describing the procedure followed to collect the images and perform the labelling, which is partly done in an automated way by relying on open-source tools that use pre-trained DL models. I recognise the merit of putting together different sources and datasets into a unique and curated catalogue which is perhaps among the largest in terms of size. However, before recommending publication, I would like the following points to be addressed:

- What are the classes for the classification task? (maybe fire/non-fire/smoke/non-smoke ?)
- Are there cases where both flames and smoke are present?
- What are the criteria followed to sample negative instances?
- Why the format 1000x1000 for the satellite images? I would say it is not so practical for training due to memory issues.
- What is the usefulness of having both photos (CV) and satellite images (RS) in the same dataset? In principle it could seem a rather arbitrary choice with little or no advantage for model training.

FInally, I think that more attention should be given to discuss possible applications of this dataset. It is mentioned that in general the FASDD should help advance research in fire detection, but I would like the autors to make an extra effort in envisioning concrete use cases or ways in which the FASDD could improve existing applications and areas of research.