

Earth Syst. Sci. Data Discuss., referee comment RC3  
<https://doi.org/10.5194/essd-2022-394-RC3>, 2023  
© Author(s) 2023. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on **essd-2022-394**

Anonymous Referee #3

---

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-RC3>, 2023

---

The paper presents a dataset for the detection of fire and smoke in RGB images from different sources, including satellite data.

### **I'll start with the things I like:**

I appreciate the fact that the dataset is open and accessible in an ML-ready format. I appreciate the effort to generate annotations and control their quality and I like the related diagrams. I like the fact that you merge different types of related datasets. The paper is generally well-written and easy to follow.

### **Major Comments**

My main concern is that the dataset is not well motivated, and neither the related work nor the experiments are yet sufficient to back up the motivation. The paper is imbalanced with more emphasis and content given on choices that are trivial (though important), such as the annotation and quality control and less emphasis on the essential, i.e. the motivation and the evaluation. If this aims to stand as a dataset with impact, the evaluation section needs to be reworked. I expect the authors to show the performance on each type of dataset and show the benefit of adding the different types of datasets. The fact that FASDD has a better performance than FASDD\_CV and FASDD\_RS is not enough by itself, without enough context about the evaluation setup that makes it a fair comparison.

### **Main Questions**

- Why is this specific dataset needed? Why would anyone use this dataset instead of a domain-specific? The answer shouldn't be that fires are increasing in frequency or intensity and it should probably come from the evaluation. Here is a sample question that you may answer: Does training on fires from graphics help in detecting real fires from drones? There are many more like that.
- Is the use of these specific satellites motivated for fire detection? These satellites have low temporal resolution and may not really be useful for fire monitoring.
- Are the optical bands of the satellite suitable for fire detection? Is there any reason why anyone would use them for that?
- I don't see why having a standard image size is a bad thing and I am not sure about the large image sizes in a machine learning dataset. A 1000x1000 image size is impractical.
- The addition of images with objects confused with smoke/fire is interesting. How did you choose the categories of such objects?
- Is the training split that you propose part of the dataset? How did you choose the split? Is it a random one? If so, I really think you should define it more carefully.
- I would like to see an analysis of the performance per domain (drones, RS, graphics, etc) and generally a more thorough evaluation.
- The experimental setup is not very well described. What preprocessing do you do before feeding the samples to the models? How do you handle the different image sizes?
- The training of the models is not detailed enough and as the authors seem to conclude, it does not make for a fair comparison " the parameter configuration and training strategy of the two models are only as consistent as possible but not entirely consistent, which may lead to a loss of comparability to some extent". It's ok to not train the perfect models, they also seem to be working quite well. The problem is that the comparison does not make much sense. And, importantly, the comparison criteria do not motivate for the use of this specific dataset.

## **Suggestions for improvement**

It would be helpful to have the dataset characteristics of section 5 in one table, rather than in text.

It would help me as a reader to see an obvious connection of Figure 1 and 3 in the text. Maybe use bold or italics in the text when a step of the process appears.

The references for fire/smoke detection with remote sensing products are insufficient.

I am not sure why the authors insist on using the term modalities as way to differentiate between rgb images captured from different sources. Different modalities usually mean different types of sensing. If you always have optical images, I think it is confusing, if not invalid, to call them different modalities. In the end you only have image data, no sound, speech, text, etc.

## Minor comments

What are the 100,000 levels?

Line 32: "lead to human respiratory and cardiovascular diseases, and even endanger human life". This phrase suggests that human respiratory and cardiovascular diseases do not endanger human life. Reframe it please.

Line 33: "According to the 2022 report from International Association of Fire and Rescue Services, the frequency of global fire events has shown an increasing trend in the last decade". I read the report and did not see a clear mention of that trend. See Jones, Matthew W., et al. "Global and Regional Trends and Drivers of Fire Under Climate Change." *Reviews of Geophysics*, vol. 60, no. 3, 2022, p. e2020RG000726. Wiley Online Library, <https://doi.org/10.1029/2020RG000726>.

Line 143: What does it mean for a dataset to be "generalized by deep learning researchers"?

I am not sure why you need sections 2.2 and 2.3. As a reader, I am more interested in your choices, and motivation. Maybe you can think of a more logical structure for the text.