Earth Syst. Sci. Data Discuss., referee comment RC1
https://doi.org/10.5194/essd-2022-394-RC1, 2022

Comment on essd-2022-394

Anonymous Referee #1

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-RC1, 2022

Summary

Fire has wide impacts on Earth Systems and human society, and efficient fire detection could promote better understanding, modeling, and preventing fires. Wang et al., synthesized a comprehensive dataset (FASDD) which covers multiple scenarios (e.g., computer vision and remote sensing) across different spatial scales. The paper is generally well written, however, I have the following major concerns especially for the data generation, data validation, and the usefulness of the data. Meanwhile some expressions listed below are less rigorous. The current version was not acceptable.

Major comments

(1) Data generation: it’s well known that near infrared (NIR) and short-wave infrared (SWIR) are two commonly used bands for fire detection while the authors only used the visible bands via visual interpretation. If only based on true colors, it can hardly convince me about the generality of the dataset for large spatial scale fire detection (e.g., some land surface items could show similar colors with fires in remote sensing images and thus mislead machine leaning models). Also, the results in Table 2 showed the really low performance on detecting remote sensing fires even with advanced machine learning models. With such a lower accuracy (e.g., relative to MTBS fire products or MODIS fire products listed below), how can the data help improve fire detection.

Sparks, A. M., Boschetti, L., Smith, A. M., Tinkham, W. T., Lannom, K. O., & Newingham, B. A. (2014). An accuracy assessment of the MTBS burned area product for shrub–steppe fires in the northern Great Basin, United States. International Journal of Wildland Fire, 24(1), 70-78.

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., & Justice, C. O. (2018). The Collection 6 MODIS burned area mapping algorithm and product. Remote sensing of environment, 217, 72-85.

(2) Data generation: for active fire detection, middle infrared and thermal bands are also important but were ignored.

(3) Data generation: for the fire detection, the authors used the top-of-atmosphere reflectance instead of atmospheric-corrected land surface reflectance which should be a problem. For example, if the smoke is white or grey, how to classify smoke versus clouds only through visual interpretation of true-colors? Meanwhile, the CV fire images should be obtained on land surface with a much higher spatial resolution (can be with sub-meter resolution in Fig. 4). Can such kinds of CV image trained machine learning models be directly used to large-scale remote sensing data without atmospheric-correction and with different spatial resolutions?

(4) Data annotation: the "minimum bounding rectangle" was used to label the images. Commonly, fire detection is to classify whether each pixel is burned or not instead of a rectangle (e.g., the MODIS fire product in Giglio et al., (2018) and Giglio et al., (2016)). Meanwhile, the fire patch perimeter was always not rectangle (Laurent et al., 2018), therefore a bounding rectangle could contain both burned and unburned pixels, right?

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., & Justice, C. O. (2018). The Collection 6 MODIS burned area mapping algorithm and product. Remote sensing of environment, 217, 72-85.

Giglio, L., Schroeder, W., & Justice, C. O. (2016). The collection 6 MODIS active fire detection algorithm and fire products. Remote sensing of environment, 178, 31-41.

Laurent, P., Mouillot, F., Yue, C., Ciais, P., Moreno, M. V., & Nogueira, J. M. (2018). FRY, a global database of fire patch functional traits derived from space-borne burned area products. Scientific data, 5(1), 1-12.

(5) Data annotation: Line 205-207, such descriptions were less rigorous, are there any numbers or thresholds to quantify such descriptions?

(6) Data validation: the true-color based visual interpretation could also involve biases, therefore it's important to validate the generated data against other reliable fire dataset, such as the MTBS data

(7) Line 245-246, the dataset consists of 95,314 computer vision fire samples but only 5,773 remote sensing samples. Due to the data imbalance, the model performance (Table 2) on FASDD data therefore mainly depends on the model performance on FASDD_CV. For the limited number of remote sensing fire samples, most samples in each region were distributed within ten days of a specific year (Table 1). Can such limited number of wildfires reflect all the seasonal and interannual changes of environmental conditions and fire dynamics so that machine learning models could learn from enough data? To my knowledge, the fire occurrence conditions changed across seasons and therefore the fire detectability could also be affected.

(8) Evaluation in section 4: the evaluation mainly showed the extent to which machine learning models could detect fires of generated FASDD data, therefore whether FASDD is reliable remains unknown. The FASDD data needed to be validated against other reliable fire products.

(9) There are many existing remote sensing fire products (e.g., MTBS, MODIS fire products) and CV fire data sets. I understand that combining the two kinds of dataset was the main difference of FASDD relative to other datasets, however, the authors did not show why combining these two kinds of dataset is important? Can combining these two kinds of datasets improve fire detection relative to existing fire detection algorithms (e.g., the method for MODIS or MTBS fire product)? If not, why people use such complex dataset (with different spatial resolution and without atmospheric-correction)

Minor comments

(1) Line 210-211, how the detection model was trained on FASDD_CV data with different spatial resolution and employed to images with different spatial resolution?

(2) Line 278-279: the original images have different sizes, right? How to unify them to the same image size?

(3) Line 280, in addition to the epoch and batch size, what about other critical parameters like the learning rate? Also any strategy for overfitting?

(4) Evaluation metrics: instead of predicting boxes, the model performance on classifying pixels (burned or not) is also very important.

(5) Line288: “IoU” firstly mentioned but without its full name

(6) Line 307: “which demonstrates the difficulty of fire detection on Remote Sensing images”, such results may imply that critical information is missing for fire detection (see my major comment 1 and 2) and thus results in the low detectability.