Comment on essd-2022-259
Anonymous Referee #2


This is a well written paper that develops a deep learning-based green roofs' mapping framework in incorporating very high-resolution remote sensing satellites. Also, this study used data augmentation to improve the accuracy of the mapping and provide a robust dataset of green roof locations and areas in London. Overall, it is helpful for others to study aerial remote sensing mapping. However, there are some problems, which must be solved before it is considered for publication. If the following problems are well-addressed, this reviewer believe that the essential contribution of this paper are important for designing sustainable buildings and studying urban microclimates.

General Comments:

In ABSTRACT: authors are suggested to refine the abstract, focusing on the novelty of the research rather than providing extensive background information. In addition, semantic segmentation models coupled with data augmentation strategies are important in this study. It is helpful to provide the validation accuracy accordingly in this section.

In INTRODUCTION: authors are suggested to start broad in the general background, then narrow in on the relevant topic that will be pursued in the paper. The introduction sections are to highlight the challenges currently faced by green roofs' mapping research. I suggest that the first three paragraphs be summarized in one paragraph. In addition, the detailed description of the mentioned algorithms (UNet) can be moved to the second section since we are not developing new models.

In DATA and METHODS: there are many datasets mentioned and may be clearer if summarized in a table. For example, recording information such as the coverage of the study area, the date of data acquisition and the spatial resolution of the imagery. We all know that convolutional neural networks are data-driven models. Well, how many positives and negatives are there before and after using data augmentation, and is there an improvement in model performance and by how much? Also, although deep learning or
semantic segmentation models are a black box, it is helpful to provide formulas. Besides, I have some doubts about the details of the algorithm framework. For example, the loss curve during optimization. Furthermore, the general area of the study area is larger than a 256*256 image patches. How did you deal with it when predicting?

In RESULTS: several tables record the model performance. Recommendations for the structure of research paper. In Tables 3 and 5, I found that the values of TP, TN, FP, and FN vary greatly. Also, despite the high accuracy of the models in Tables 4 and 6, the performance is weaker on the IOU metrics. Why is this happening?

The image segmentation algorithm used is supervised classification. As a result, the classification results are constrained by the labels, and the model’s generalization is limited. Even though we are able to assign labeling tasks to individuals, we cannot classify locations where the image is occluded. Is the current weakly supervised or unsupervised segmentation a significant advancement?