This paper introduces a multi-resolution PV dataset that composes of various samples collected from satellite and aerial images. Such a dataset is of great interest to users because it helps to develop deep learning algorithms for automatic PV information extraction. The paper is well organized and the dataset is described in a clear fashion. Therefore, I would like to suggest for publication in the Earth System Science Data after a minor revision.

Response: Thank you for your encouraging comments.

Comments:

1) For the segmentation experiments (Section 4.1), the samples are divided into 80% training set and 20% testing set. Is the division performed randomly for PV08/03/01 or separately for each subcategory? The authors should make it clear because the two approaches lead to different model performance and segmentation results.

Response: Thank you for your advice. Actually, the division is performed for each subcategory of PV08/03/01. For example, the training (testing) set for the experiment on PV08 contains 80% (20%) samples from PV08_Rooftop and 80% (20%) samples from PV08_Ground. We have made it clear in the revised manuscript as "The experiments were conducted on PV08, PV03 and PV01 dataset, respectively. For each sub-category (e.g., PV08_Rooftop, PV PV08_Ground), all samples were separated into 80% training set (from which 20% samples were used for validation) and 20% testing set."

2) For the cross application (Section 4.2), I wonder whether authors divide training samples and testing samples in the same way as Section 4.1? The authors state that "fine-tuning means that the model was first pre-trained on PV03 (PV08) samples, then fine-tuned (fine-tuning process lasted 10 epochs) using PV08 (PV03) samples, and finally applied to PV08 (PV03) samples." Are the samples used for fine tuning the same as those used for direct training? If yes, how will the model perform if only using a small portion samples for fine-tuning.

Response: For the cross application, we divide training set and testing set in the same way as Section 4.1. The samples used for fine tuning are not the same as those used for direct training. Only a small portion (20%) samples from the training set of the target PV dataset are selected for fine-tuning. We have made it clear in the revised manuscript as "The

training set account for 80% of the whole dataset and the testing set is the remaining 20%. Only 20% samples from the training set of the target PV dataset are randomly selected for fine-tuning."