

Comment on essd-2022-75

Robbe Neyns (Referee)

Referee comment on "UGS-1m: fine-grained urban green space mapping of 31 major cities in China based on the deep learning framework" by Qian Shi et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-75-RC2>, 2022

The paper makes a clear contribution to the field of urban green space mapping by providing a new annotated dataset of urban green spaces, a new deep learning framework and a detailed green space map of 34 cities in China. However, I still have several question/comments that need to be addressed before considering publication.

General remarks

- One of the key contributions of the paper is the use of adversarial training is the use of adversarial training to increase the generalisation capacity of the deep learning model. Yet the actual added value of the adversarial part of the model is only addressed to a limited extent. If I understand correctly (this should also be clarified in the paper), the other semantic segmentation models were not finetuned using adversarial training. This means that one could deduce the effectiveness of the adversarial training approach from the difference in accuracy with the other models. But this will always leave the question: is the improvement related to the model structure of the generator or due to the inclusion of a discriminator? In short, I would have liked to see the difference in performance with/without the discriminator.
- Related to the previous remark I would have liked a description of the spatial differences in accuracy in the discussion section. Do some cities show a lower accuracy? Could we explain this by the type of green space that is dominant? Difference in phenological phase when the imagery was taken?...
- The discussion of the type of errors in the resulting green space map is very limited. Do you mainly notice problems at the edges of green spaces? For high/low or dense/sparse vegetation?
- Are there any remarks in relation to the dataset? For example, in figure 12 image 3 (starting from above) I noticed that the residential area has been indicated as non-vegetated while there is clearly vegetation next to the buildings.

Specific remarks

- Which value did you use for N in section 3.1.2?
- Figure 8: I believe the final 8x8x1 image is flattened before it is fed through the softmax function?
- How was the parameter optimization performed?
 - A batch size of 8 image pairs is rather small, was this decided through parameter optimization experiments or because of memory limitations?
 - The same learning rate, batch size and number of epochs was used for all models? Was this done to facilitate the experiments?