

Comments to the Author:

Thanks for submitting the interesting manuscript, and I really appreciate your attempt to use deep learning technique to solve the problems in the field of hydrology. Before your revision, I would like you to consider my following suggestions:

1. emphasize the scientific significance of your research.

The novelty of the paper hasn't been clearly proposed. The main contents of your work can be classified into two parts: firstly, water segmentation based on transfer learning approaches; secondly, the development of corresponding algorithms to monitor the water level with or without the help of landmarks. However, neither the scientific significance of the contents were fully reflected in the paper, especially in introduction and title.

If you want to address the significance of transfer learning, some extra experiments will be necessary. Not only the comparison between different transfer learning approaches should be carried on, but the advantages of transfer learning over directly training models should be clarified with data through series of experiments. I suggest that you design more detailed experiments to explore the relationships between the number of training data and the superiority of transfer learning approaches. This will also help to directly explain why you use transfer learning.

LBWLE is not important or novel as expected. It is highly influenced by the number and quality of the landmarks, so you should focus more on how to increase the accuracy of semantic segmentation model itself as suggested previously. Though the significance of the second point may be smaller, yet if you want to emphasize the second point, there should be more discussions on LBWLE algorithm. The advantage of LBWLE over previous methods (e.g. SOFI) and the necessities of the LBWLE were not fully discussed in the paper. Add some reference on the advantages and disadvantages of the previous methods to show the novelty of your method.

Accordingly, the title also has to be changed to put emphasis on transfer learning, and the introduction need to be restructured. The introduction is suggested to be organized to present more about the development of water semantic segmentation methods and the corresponding water level indexes rather than the history of water level monitoring approaches.

2. Suggestions on language and writing style.

As for language, avoid using too much first person expression like "we". As for writing style, the paper is organized more like a paper published in computer science journals, e.g., separating the introduction and background. adjust your paper structure referred to other papers in HESS if you still would like to publish the paper in a hydrological journal.

3. Check your reference.

Please make sure every paper you refer to can support your ideas. E.g. in line 26: “The network of river gauging stations is declining globally (Vorosmartyetal., 2001)”, the expression is doubted, if it is the idea in the paper? In line 119, the originality of Moy de Vitry etal.(2019) lied on two parts, the development of a water segmentation model and the proposal of SOFI, the expression in your paper (the biggest originality...) was not rigorous.

4. Give an explanation when computer science terms first appear, like “fine-tune”.

5. The two transfer learning approaches cannot be fully understand only from the present description, it is recommended to use figure to illustrate it. And comment on what’s the essential difference between the two transfer learning approaches.

6. The structure of the original semantic segmentation deep learning model would be altered when be transferred, e.g., adding an fully-connected layer to adjust the model task from multi-class to two-class. It is a necessary step but not mentioned in your paper.

7. A lot of metrics for evaluating the results has been listed in Table 3. Are they all useful for the evaluation? Only remain the metrics helpful for the following discussion and explain what the metrics represent before discussing the results.

8. If the LBWLE can only capture the flood rather than drought (Figure 8)? If so, this will limit the value of the algorithm in hydrological problems such as hydrological model calibration.

9. Illustrate what’s your purpose on setting up two experiments (two-week and year-long) ? Do they correspond to different application scenes respectively? What substantial conclusions can be reached through the comparison of two experiments, why not only use the year-long series?

In conclusion, though your research is really interesting, I recommend you to supplement the suggested experiments and reorganize the structure. I’m looking forward to your reply to my questions.