HESS-2021-20
***Vandaele et al.* Deep learning for the estimation of water-levels using river cameras**

Response to Reviewer #2

We would like to thank the reviewer for their comments and suggestions. We respond to these comments below. For clarity, we have introduced some extra numbering (1a, 1b, 1c etc.) to address the separate points made in each comment.

As part of our response to the comments, we propose to revise the organization of our manuscript. The new proposed *table of contents* is located at the end of this document. If our manuscript is accepted for revision, in addition to the changes proposed below, we intend to update parts of the text to better fit this new organization (section introductions for example). for the sake of clarity, we preferred to keep these changes out of this answer.

Comment #1a: Emphasize the scientific significance of your research, including the transfer learning aspects

In order to highlight the novelty of our manuscript and answer the reviewer's comment, if our manuscript is accepted for revision, we would like to make the following changes:

(1) Modify the end of the introduction (starting line 58, ending line 69) to give a better introduction to transfer learning and our previous paper:

"Over the last decade, transfer learning (TL) techniques have become a common tool to try to overcome the lack of available data (Reyes et al, 2015; Sabbatelli et al, 2018). The aim of these techniques is to repurpose efficient machine learning models trained on large annotated datasets of images to new related tasks where the availability of annotated datasets is much more limited (see Section 2 for more details). Vandaele et al (2020) successfully analyzed a set of TL approaches for improving the performance of deep water segmentation networks. In this paper, we build on the work of Vandaele et al. (2020) and study the performance of these water segmentation networks for the automation of river-level estimation from river-camera images, in the context of flood-related studies. In particular, we carry out novel experiments realized with new river-camera datasets and metadata, that consider the use of several methods to extract quantitative water-level observations from the segmented river-camera images."

(2) Create a new Section 2 that would be our Methodology section and would encompass:
- the former Section 2 introducing semantic segmentation, deep learning and transfer learning
- the former Section 3 presenting the application of transfer learning
- the presentation of SOFI and the new LBWLE method moved from former sections 4.1.4 and 4.2.2

We think that this new organization will help the reader to understand our research more clearly.

Comment #1b Importance and novelty of LBWLE

Our goal with LBWLE was to propose a way to provide quantitative water-level observations in accepted units (m), using the landmark-annotated dataset at our disposal.

If our manuscript is accepted for revision, we would like to add a clarification in the new Section 2.3 of our reorganized paper where we would compare the two criteria (SOFI and LBWLE):

When compared to the SOFI index, water-level estimation using landmarks and LBWLE is at a disadvantage because of the necessary and time-consuming ground-survey of the location observed by the camera. Furthermore, landmarks can mostly only be used when the river is out-of-bank, so the approach is not likely to capture drought events. However, the main advantage of this approach compared to SOFI is that it allows estimation of quantitative river levels in accepted units of length (e.g., m). The SOFI index values are dimensionless percentages and to convert them to a height measurement an appropriate scaling must be obtained by calibration with independent data.

Please also see our answer to Reviewer 1's first comment regarding the alternative ways to obtain topographic data (https://doi.org/10.5194/hess-2021-20-AC1).

Comment #1c Changes to the title of our paper

Following the reviewer's comment, we would suggest the following title change to our paper:

Deep learning for automated river-level monitoring through river camera images: an approach based on water segmentation and transfer learning

Comment #1d Emphasis on water-level segmentation in Section 1- Introduction

If we are invited to revise the manuscript, we will add an additional material about transfer learning to the introduction (see response to Comment #1a). In our view it is necessary to also discuss other types of water level observation and hydrological uses of river cameras in the introduction (see response to Reviewer 1 Comment 4).

Comment #2a: Suggestions on Language and Writing style - use of first person "we"

If we are invited to revise the manuscript, it will be checked and some sentences will be rephrased to address this comment.

Comment #2b: Separate introduction and transfer learning background sections

If we are invited to revise the paper we will introduce some material on transfer learning in Section 1. In order to fit the typical HESS organization, we will merge Section 2 and 3 into a single Methodology section (see response to Comment #1a and the new table of contents presented at the end). Some of the ideas on transfer learning from former Section 2 will be rephrased and retained as part of a "Definitions" subsection of our new merged Methodology section. This is to avoid Section 1 becoming excessively long while explaining the important concepts used in our work so that they can be readily understood by HESS readers.

Comment #3: Check your references

Vörösmarty et al (2001) include a substantial section on the decline of river gauge data worldwide. We intend to add two more recent references that provide further evidence on this point (Mishra and Coulibaly, 2009; Global Runoff Data Center, 2016).

Mishra, A.K. and Coulibaly, P., 2009. Developments in hydrometric network design: A review. *Reviews of Geophysics*, *47*(2).

Global Runoff Data Center, 2016, Global Runoff Data Base, temporal distribution of available discharge data, https://www.bafg.de/SharedDocs/Bilder/Bilder_GRDC/grdcStations_tornadoChart.jpg, *Last accessed: 15 March 2021*

We would rephrase the presentation of the Moy de Vitry et al (2019) paper (line 119):

Moy de Vitry et al (2019) used a deep semantic segmentation network trained from scratch on an image dataset annotated with water labels. They post-processed the network output in order to produce a generic algorithm for flood level trend monitoring. The SOFI index that they introduced in their post-processing step is related to the percentage of pixels in the image that are estimated as water pixels by the network . This non-dimensional index allows the authors to monitor the evolution of water-levels in their datasets. In our experiments, we will use the SOFI index to track water-level changes (see Section 4).

We would also give the mathematical expression for the SOFI index in the new Section 2.3 of the reorganized manuscript following Moy de Vitry et al. (2019).

Comment #4: Give an explanation when computer science terms first appear, like "fine-tune"

We would like to clarify the explanation of fine-tuning that was given after its first appearance at line 142. Note that given this comment (as well as the restructuring of the manuscript proposed in our responses to comments #1a, #5 and #6), we propose to:

(1) Insert the explanation of fine-tuning in the new Section 2.2.3:

In Vandaele et al. (2020), the most successful approach considered for applying transfer learning to the semantic segmentation networks is fine-tuning: with fine-tuning, the filter weights obtained by training the network over the source problem are used as initial weights for training the network over the target problem.

(2) Avoid mentioning fine-tuning in the former Section 3.3 (reorganized section 2.2.2) and use a more generic term instead: the application of transfer learning to train the networks.

(3) Rename Section 3.4 (reorganized section 2.2.3): *Applying transfer learning to train the networks.*

Comment #5: Explanation of the two transfer learning approaches

We explained the two transfer learning approaches in former Section 3.4. We hope that these explanations will be more prominent and easier to follow in our reorganized manuscript, where they will appear in section 2.2.3.

Comment #6: Network structure change for binary segmentation

We talk about the change in dimension due to the switch to binary segmentation in our transfer learning approach in the former Section 3.4 of our paper. Similarly to Comment #5 and #4 , we hope that the reorganized version of the manuscript could help to avoid the confusion.

Comment #7: Table 3 metrics

These metrics are widely used for evaluating flood extent models (e.g. Stephens et al, 2014). However, we understand that including them all in the main paper may not be useful for all readers. We propose to remove the F^1, F^2, F^3 and F^4 scores from the main paper, but show these in Appendices or Supplementary Materials.

Comment #8: LBWLE captures floods rather than droughts

See response to Comment #1b.

Comment #9: Purpose of setting-up two experiments

If our manuscript is accepted for revision, we propose to make the following modifications in the introduction of Section 4:

Two experiments were carried out with this study.

The first experiment presented in Section 3.1 is designed to address the suitability of our approach for the automatic derivation of water level observations using river cameras images and landmarks from a ground survey. Landmarks and associated manually derived water-levels are available for a two-week flood event (Vetra-Carvalho et al, 2020). These data allow us to validate our LBWLE approach for water-level estimation in accepted units of length (m) with co-located water-levels estimated by a human observer.

With the second experiment presented in Section 3.2, our approach is applied to larger, one year, datasets of camera images that include a larger range of river flow rates and stages. This experiment allows us to better understand the suitability and robustness of the LBWLE and SOFI water-level measurements. However, manually derived co-located water levels are not available for this period, so we use the nearest available river gauge data for validation. For some of the cameras, the nearest gauge is several km away.

New Table of Contents:

1. Introduction *[former Section 1]*

2. Transfer learning for water segmentation and river-level estimation
   New section encompassing former sections 2 and 3, and addition of a new section concerning SOFI and LBWLE.
   2.1 Definitions *[former Section 2]*
      2.1.1 Water segmentation [*former Section 2.1*]
      2.1.2 Deep Learning for automated water segmentation *[former Section 2.2]*
      2.1.3 Transfer Learning *[former Section 2.3]*
   2.2 Transfer Learning for deep water semantic segmentation networks *[former section 3]*
      Former Section 3.1 is removed to avoid repetition of material
      2.2.1 Network architectures and source datasets *[former Section 3.2]*
      2.2.2 Target datasets for water semantic segmentation *[former Section 3.3]*
      2.2.3 Applying transfer learning to train the networks [*former Section 3.4]*
   2.3 River-level estimation using water segmentation [*New section to present SOFI (former part of Section 4.2.2 moved here) and LBWLE (former part of Section 4.1.4 moved here)]*

3. Experiments *[former Section 4]*
   3.1 Application on a practical case for flood observation *[former Section 4.1]*
      3.1.1 Datasets *[former Section 3.1.1]*
      3.1.2 Evaluation Protocol *[former Section 4.1.2]*
      3.1.3 Landmark classification results *[former Section 4.1.3]*
      3.1.4 Estimating the water-level using the landmark classification [*former Section 4.1.4]*
      LBWLE explanation is moved to 2.3
   3.2 Performance evaluation for year long water-level analysis *[former Section 4.2]*
      3.2.1 Year-long river-camera images datasets *[former Section 4.2.1]*
      3.2.2 Evaluation protocol *[former Section 4.2.2]*
      => SOFI discussion has been moved to Section 2.3
      3.2.3 Landmark-based water-level estimation analysis *[former Section 4.2.3]*
      3.2.4 Full image SOFI analysis *[former Section 4.2.4]*
      3.2.5 Windowed image SOFI index analysis *[former Section 4.2.5]*

4. Conclusion *[former Section 5]*