HESS-2021-20

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

Thank you for considering our paper for revision. You will find below a point-by-point answer to the reviews. They correspond to slightly adapted versions of our answers made during the interactive process (https://doi.org/10.5194/hess-2021-20-AC1 and https://doi.org/10.5194/hess-2021-20-AC2). The small differences between the answers are related to minor changes that we deemed necessary when we applied our proposed modifications.

The legend of the marked-up version of our revised manuscript is as follows:

- Red: additions to the manuscript
- Blue: modifications to the manuscript
- Green: text & sections that were moved (if parts were moved and modified, blue has the priority)

Note that, unless explicitly stated otherwise, section numbers in our response reflect the new organization of the manuscript that is presented at the end of this document.

Response to Reviewer #1

Comment #1: Regarding the use of alternative ways to obtain topographic data

We agree that obtaining ground surveys for the field-of-view for each camera is a significant challenge. An alternative could be to use the camera images in conjunction with a high resolution digital surface model (DSM). An open access lidar DSM is available across the UK. Over other parts of the world, it might be necessary to use a DSM that is commercially available (for example the 12m WorldDEM), but the accuracy of the results using a low resolution DSM would need to be carefully evaluated.

We added the following sentence in the conclusion section:

"*Future work will focus on the merging of the water segmentation results with lidar digital surface model (DSM) data available at 1m resolution over the UK (Environment Agency, 2017). This would allow the water segmentation algorithms to provide a direct estimate of the water levels in the areas that are studied, without requiring any ground-surveys.*"

Comment #2: Regarding camera movement

From our inspection of the datasets, we found that the camera movement was negligible (a maximum of 2-3 pixels, even on objects far away from the camera). We were especially careful to choose landmark pixel locations that were not too close to the edge of the landmarked object in order to avoid confusions. We would expect typical intensity-based image registration algorithms to deteriorate the results due to changes in image illumination and movement within the image (of boats, wildlife and debris), although we have not tested this.

We added the following sentence at the end of Section 3.1.1 (Dataset presentation):

"*An inspection of the datasets and results showed that the impact of camera movement was negligible. Machine-Learning based landmark detection algorithms (e.g, Vandaele et al., 2018) could have been used otherwise, but they are unnecessary in the context of this study.*"

Comment #3: Regarding the use of area features instead of landmarks

The starting point of this study was to show how well we were able to automate the annotation process in comparison with the manual landmark annotation approach used in our former study (Vetra-Carvalho et al. 2020). We agree that the use of landmarked areas as opposed to landmarked pixels could strengthen the robustness of our results, but it would have required a new and more complex ground survey that was not in the scope of this study.

To address this comment, we have added the following sentence at the end of Section 3.1.1, after the modification of Comment #2:

"*Also note that this work focuses on a simple process relying on single pixel landmark locations annotated by Vetra-Carvalho et al., (2020b). The use of landmarked areas of multiple pixels sharing the same height could likely help to increase the detection performance and should be considered for an optimal use of this landmark-based approach.*"

Comment #4: Literature

We added citations to the literature mentioned by the reviewer at line 41 of the introduction section as follows:

"*There have been a number of citizen science projects that investigated the use of crowdsourced observations of river level (e.g. Royem et al., 2012; Lanfranchi et al., 2014; Etter et al., 2020; Lowry et al., 2019; Walker et al., 2019; Baruch, 2018).*"

In our introduction section, we added the following paragraph regarding the use of river cameras:

"*Several studies have already attempted to use videos and still camera images in order to observe flood events. Surface velocity fields can be computed using videos (e.g., Muste et al., 2008; Le Boursicaud et al.,2016; Creutin et al. 2003; Perks et al., 2020). Still images can be used to observe the water-levels, either manually (e.g., Royem et al, 2012; Schoener, 2018; Etter et al, 2020) or automatically, for example by considering image processing edge detection techniques (Eltner et al. 2018). Under the right conditions, these automated water-level estimation techniques can provide good* *accuracy with uncertainties of only a few mms (Gilmore et al., 2013; Eltner et al. 2018). However, the performance of these approaches lacks portability (Eltner et al.,2018.).* "

Response to Reviewer #2

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 revised the organization of our manuscript. The new *table of contents* is located at the end of this document. 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 minor changes out of this answer but they are given in colour in our marked-up version of the revised manuscript (see above for legend).

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, we made 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 analysed a set of TL approaches for improving the performance of deep water segmentation networks. This paper builds on the work of Vandaele et al (2020) and studies 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, this work carries out novel experiments realised 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.

We added a clarification in Section 2.3.3 to 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.

See our answer to Reviewer 1 Comment #1 regarding the alternative ways to obtain topographic data.

Comment #1c Changes to the title of our paper

Following the reviewer's comment, we made 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

We added 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"

The manuscript was checked and some sentences were rephrased to address this comment.

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

We introduced some material on transfer learning in Section 1. In order to fit the typical HESS organization, we merged former 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 were 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 added 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 Database, temporal distribution of available discharge data, https://www.bafg.de/SharedDocs/Bilder/Bilder_GRDC/grdcStations_tornadoChart.jpg, *Last accessed: 15 March 2021*

We rephrased the presentation of the Moy de Vitry et al (2019) paper (line 119) and moved it to new Section 2.3.2:

The experiments presented in this work use the SOFI index to track water-level changes. Moy de Vitry et al. (2019) introduced the SOFI index to extract flood level information from a deep semantic segmentation network trained from scratch on an image dataset annotated with water labels. The SOFI index is related to the percentage of pixels in the image that are estimated as water pixels by the network, as

$$SOFI = \frac{\#Pixels_{Flooded}}{\#Pixels_{Total}} \quad (1)$$

This non-dimensional index allows the authors to monitor the evolution of water-levels in their datasets, and can be computed on the entire water mask or only a sub-region.

Eq 1 corresponds to the mathematical expression for the SOFI index following its definition in Moy de Vitry et al. (2019).

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

We clarified 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 made the following changes:

(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 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 Section 2.2.3 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 removed the F^1, F^2, F^3 and F^4 scores (and their mentions) from the main paper.

Comment #8: LBWLE captures floods rather than droughts

See response to Comment #1b.

Comment #9: Purpose of setting-up two experiments

We made the following modifications in the introduction of our new Section 3 (Experiments):

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 the nearest available river gauge data for validation was used instead. For some of the cameras, the nearest gauge is several km away.

New Table of Contents: