Overview

Thanks to the authors for taking on-board the suggestions of the prior reviews and clearly outlining the changes they have made.

The modifications to the paper have helped to improve its overall quality and robustness. Special attention has been paid to make the methodology clearer to the reader and as such it is easier to follow and understand what has been undertaken. Likewise, inclusion of more detail about how changes to data were undertaken and the number of images and samples has helped to make it clearer the volumes of data used overall and in the different scenarios.

There are a few outstanding changes which are advised below, and with this the work is of publishable quality and will be of benefit for researchers in this field in the future.

General Comments

Section	Comment
Introduction	The introduction is improved, both by clearly stating what the paper is aiming to do and giving a broader overview of the current literature on the topic. Likewise, the shift in introducing CNNs to the method section helps to create a better break point in the front end of the paper.
Methods	The choice of the YOLO algorithm as a starting point here is much better justified, alongside providing context to other algorithms that could be used and why they were not. This is a considerable improvement over the prior version of the paper. The sources of data are also much better explained, and the variation in capturing technique and methods (natural vs seeded). This is also echoed in the data description for table 1. Likewise, the PCA purpose is much better explained and clearer for the reader which is excellent to see. Although section 2.2.1 may be better using a range of values, it is now better justified as to finding whether less data would decrease performance or not as opposed to finding an optimal size. However I think as you have some different sized scenarios, and the 2000 sensitivity example will be using multiples of the same image, I feel this section could just be removed with no impact on the paper. The inclusion of all the datasets in the scenarios really helps to see how much data is being used in each. Changes to the following sections also help to distinguish between what the scenarios are doing and any following tests that were undertaken.
Results	The section on training data clustering is much improved, with the changes made helping to inform the reader of the variation and/or similarities in your data. Again, for the sensitivity, I think unless you could rerun to look at say 250, 100 images as well this isn't rally adding much to the narrative or useful to

	the reader as it hasn't identified a threshold at which time and performance separate/converge.

I feel that with the separation of the results and discussion, although this has helped distinguish between what you have found and why that may be, it would make more sense still for the test on the river Inn to be placed in the results. This should be relatively straightforward with a split halfway through that paragraph in the discussion. I see less of an issue of the YOLO algorithm results being in the discussion, as this is more or a discussion of the algorithm selection than an aim of the paper to find results of.

Overall though the separation of the two sections has been done well and is easy to read and follow. |
| Discussion | Clarity around the discussion relating to adding similar images was welcomed, especially from the same river at a different time point. There were some clarifications in calculations here too. The addition of discussions around earlier scenario tests was also welcomed as this was missing prior and provides useful guidance for database configuration. However, there still seems to be no reference to the augmented scenarios around rotation and mirroring (apart from a line in the conclusions). These appear to show changes in line with other scenarios but are not overly discussed. Adding a sentence or two discussing these after the sampling discussion would be sensible, even if to mention the impact not being as much as expected.

There may need to be some reference to the visual similarity of images in for the river Inn experiment, was this likely to be a good match to the training images for example?

Overall though the discussion is well framed in which each component being discussed is countered with a point about how this may impact future practitioners in the field which is useful for the paper to have both applicability and need.

I wonder if for using the Yolime on images, whether the same could easily be replicated for a standard image from one of your bridge datasets perhaps? You note the importance of surrounding water to detect wood, perhaps this would also be shown in your other images? This could then be combined with your current figure 8 to have 4 images, two algorithsm at two locations?

It was good to see the addition of low-cost camera limitations, but perhaps adding a comment weighting up the benefits vs drawbacks of low-cost cameras for science vs monitoring could be included. I.e. are they good enough for monitoring and alerting of large wood flows, but not good enough for quantifying and understanding wood transport dynamics? |
| Conclusion | The conclusion has been reworded in to succinctly summarise the substantial amount of work that has gone into this paper. This is effective at displaying the key messages to the reader. |

Specific Comments

Section	Line	Comment
Introduction	83	Start a new paragraph with 'in this study', helps to really showcase what the aim is rather than being lost in the last introduction paragraph.
Methods	137	Can you add 'both using manual and automated approaches' to the end of the first sentence (or similar), just so that readers are aware from the outset that there is a mix and that's why the pseudo-labelling is introduced.
	153	Was this a specified percentage or did it vary, needed to be included to decide what counts as overlapping?
	189	Although this section is now much clearer with the introduction of an extra validation dataset. I still believe including this in the data acquisition section would be best to avoid confusion over bringing extra datasets in. Alternatively, adding a statement in the data paragraph that additional data is used for validation or in some scenario cases would also be sufficient.
	214	Add in 'for' between points and mean.
	269	The sentence here relating to the figure does not make sense. A) I think you mean figure 5, and B), change an 11 and A 12 to 'The 11 and the 12 descriptors are the self-gathered extra datasets'.
	274	Why was dataset 18 not removed as well if it is shown to be one of the worse?
	286	Just to clarify, as I don't think this is particularly explicit still. Are these images resized to 832x832 from the original image, or by 'doubling' the resolution of the sampled 416x416 images. Mainly as doubling the resampled images implies something about the image size (in bytes), whereas resampling to 832 tells you something about image detail retained.
	316	New line at 'After that' and perhaps change this to subsequently or additionally instead.

Results	352	Insert 'test' or 'database configuration' before scenarios to distinguish from baseline scenario.
Conclusion	507	Change can not to cannot, although correct 'can not' over emphasises the negative, whereas in reality there is potential for real-time use.
	509	Merge the two sentences into one here, there seems to be no benefit of separating them out.

Figures

Figure	Comment
Figure 2	Could this include a reference to the source? Whether its primary collection, secondary (i.e. the Allier and Ain), and presumably purchased imagery for number 9? Either in the figure or the caption.
Table 1	The inclusion of the number of unique labels is great for working out which sites may be contributing more to influence the model performance.