Thanks for the revised manuscript, which contains a much improved version of your manuscript, and that addressed most of the major points raised previously.

Regarding my previous major points, I note the authors effort, and specifically:

1. I understand the authors added more information in the introduction about different CNNs and other deep learning techniques. I think what is still missing is in the discussion section a paragraph where the authors discuss what other neural networks could (or could have not) achieved, based on the existing literature. This will help to underpin the authors' method as the "way to go" for automatic detection.

2. It is now much clearer what the aim of this paper is from the onset.

3. The paper structure is now much more readable and I appreciate the authors' effort to achieve that.

4. Whilst the methodology is substantially improved, some aspects are still a bit unclear to me, specifically about where and how the cameras were placed. I think a schematic would really help (even as a subfigure to figure 2 or a standalone figure), as also suggested in my previous review. Also, for Figure 2 the map is very welcome; however, the locations are impossible to identify, I recommend in-set maps to show where bridges/capturing points are located.

Other minor line-by-line comments:

General: I note the effort done by the authors to use one type of spelling only. There are a few still to be solved (e.g., utilized, analyze, hypothesized, greyscale now turned into grayscale)

Line 83: "which we attribute to the lack of uniform data" is there a reference(s) that you can use to substantiate this?

I note the response from the authors about iPhones (and by extent to many other brands) which is perfectly fine. However, can they add the information provided in the text (i.e., that the phones used in the study were those only available to the authors, so there could be uncertainty about other brands and processing software)?

Line 92: CNN and YOLO have already been defined above, so no need to re-instate the acronym

Line 150: "a lot of", this is not formal scientific writing

Line 150-151: "Naturally, the wood floats and moves in a f low or is deposited or trapped by an obstacle (i.e., river bank, boulders, trees), some videos contain minutes" sentence needs rephrasing

The authors mention that they credited the images used for the dataset, but this is not reported in any of the figures or in the text, please correct.

Line 638: whilst I thoroughly appreciate the authors' encouragement for this reviewer to get a promotion to full professorship, I am still a "Dr"

References: please note the typo in "Panici (2021)" and in-text citation