Overview

This paper investigates a novel concept for monitoring wood in rivers, developing on existing algorithm development of CNNs for image recognition. There are numerous applications and possible impacts for this work, both for research and monitoring. The authors make a good case for the necessity of the research and offer a good grounding in some of the key concepts for a reader who is knew to machine learning methods, as well as those more familiar with them.

The writing needs to be improved throughout, as there are numerous occasions of informal writing which feel out of place, such as 'made a recent come-back'. Likewise, there are spelling mistakes and issues on consistency between American and English, which I am aware can be a challenge when writing in a non-native language. This can be helped by making sure both the spell-check and dictionary of all text in the document are set to one or the other. Moreover, some of the text struggles to convey the complexity of the methods in places, with repetition followed by missing detail.

The scenario design is clear, however how these scenarios fit in with some of the other analyses being undertaken is less apparent. There seems to be several sections which are additional scenarios/tests throughout the paper which are not clearly explained in the methods. Furthermore, there are numerous scenarios which are outlined in the methods which are not commented on in the results or discussion. These should either be discussed, or possibly removed (placed in supplementary) for the revision. Some of the additional analyses could then be incorporated as scenarios to make it easier to understand for the reader. Moreover, as the methods are quite complex, an improved schematic overview of the workflow would benefit the manuscript.

The results and discussion are currently presented as one. It would be best to separate them in this instance, with a discussion focussing on the reasons why some scenarios performed better, the limitations of the design, and the impact this may have for wood monitoring. This seems to be the largest element missing from this paper. Overall, it has great potential for helping to improve wood monitoring in rivers, but this is only briefly covered in the discussion, despite the overwhelming literature relating to the importance of wood in rivers, the hazards they present, and the different methods currently being used to monitor them. The results themselves are also not covered in full which is surprising as the order of magnitude of results is similar to those results that are covered.

The figures and tables throughout are of good standard, and with some adjustment would be suitable for publication. The only major differences would be the inclusion/adjustment of the methodology schematic as the visual elements would help the reader alongside the text, as well as the inclusion of a location map for the dataset origins in Figure 2.

Overall, the paper shows good promise and with some adjustments to the content, bolstering some of the justification, improving the writing standard, and focusing on the relevance of the work, would provide a useful addition to the field.

General Comments

Section	Comment
Introduction	The opening to the manuscript is very clear laying out both the justification for this research as well as introducing concepts around the importance of in-channel vegetation and the anthropogenic forces governing this area. Some aspects could do with a greater reference to existing literature, as well as making sure to give examples when the authors make certain statements regarding current methods and results. Take care also not to generalise fields of research too much, to state the wood transport data is scarce seems a stretch, as there are examples of research into this. Perhaps rephrasing these types of statements to be softer would be beneficial, even with the caveat of this being a growing area of interest. Some of the latter paragraphs don't tend to flow well between one-another, and potentially some better link statements and reordering would help this feel less 'chunked'. Some suggestions provided herein. The final paragraph is almost to detailed for the introduction, but not detailed enough for the methods. I would suggest removing part of this (line 67 onwards) into the methods section to help justify your choice of CNNs. The new last sentence would act as an aim of the paper, and as such, a few points to act as objectives could help round of the introduction nicely. This is currently not clearly stated and so would need to be included, even if not explicitly listed.
Methods	There seems to be a large omission of background into the different possible algorithms that could have been used (both CNNs and other) at the beginning of this section. With the addition and expansion on some of the text from the introduction, as well as more explanation into how they work and why they are suited for instream wood detection, this would be more suitable, especially as this paper feels like it is trying to reach both those already involved in machine learning methods, and those who want to detect wood better but do not have the experience. Likewise, the choice of YOLO as an algorithm is not fully justified in the text, and so could do with some more supportive reasoning here. Again, there is some use of informal phrases, such as 'the goal is to have a lot of diverse data', whereas, 'CNNs require large volumes of diverse data for effective training', would be better suited. Overall, there is a tendency to omit some key pieces of information from the methods as outlined in the specific comments below. These stop the reader from fully understanding the methods that have been undertaken, and without them would not be suitable for full replicability. The overall layout of the methods section is well though out with clear subsections to help guide the reader through what is clearly an extensive

	piece of analysis, but the detail makes it hard to truly understand what's going on in places. Although the splitting of the data into train, validation, and test, is somewhat stated at the beginning of the methods, I don't think it is overly clear that this is done on a dataset basis for each of the following scenarios. I think some clearer wording and phrasing here would help the reader to know exactly what is being used as training, validation, and testing data. Again, I am not sure if a better overall methodology schematic would be suitable, building on figure 1 primarily. The scenarios used for assessing model development are in themselves valid, but seem to place arbitrary values and not investigate the scale over which these operate. I.e. a minimum of 500, but does not overly look into the effects of changing this arbitrary value, which could have better answered a few scenarios in more detail. This is partly answered in the sensitivity analysis, but using two values does not really give a sense of how the ability of the network to learn with different numbers of training images develops. I am also not sure the sensitivity analysis is really that different from some of your scenarios, and as such maybe rolling them into one section is better here. When outlining your scenarios, you also sometimes state the number of images altered, and how this effected training dataset size compared to baseline, but not always, even for similar scenarios. Try to be consistent in the information provided. Overall, there is a lot of key information in the methods, and it is relatively well laid out. However, the discrepancies between information given, and a lack of lineage in terms of going from raw data to a final test of an optimised model could be improved. Certain sections seem out of place, with some underdiscussed. The methods themselves seem sound and novel given the prior work in this area and could make a valid contribution to the research in this field which is great to see.
Results/ Discussion	I think there is a need to separate out the results and discussion. Although in some cases a more fluid approach is justified, in this scenario I think it becomes a little unclear as to what you have found and what you are discussing. This links to your section headers. The methods has sensitivity analysis and the scenarios as separate, but your results combines these. I think it would be wise to align the headings so each one refers to each section of the methods, this will help the reader. From doing this, it should help to solve a follow up change, whereby your headings for each part of the results/discussion summarise your findings. I really like this idea so the reader has a clear description of what the next section is discussing, these could work nicely in a separate discussion section along with some other suggestions further in. I am unsure whether the PCA is adding much to the argument here, a lack of access to supplementary to check this means further comment cannot be made. However, figure 5 shows quite similar datasets on the whole with a few outliers, contrary to the narrative.

	Your results show the outputs of several augmentation scenarios and sampling scenarios (1-8), but these are barely covered in any results or discussion, with most of the writing focussing on scenarios 9-13, which begs the question of why these scenarios were done or even included to begin with? The values in table 2 suggest the changes in model outputs are not drastically different than those that are discussed either. The end of the results/discussion section feels a bit rushed at points and does not really go into some of the detail on interesting questions surrounding model selection and the test dataset. I am also unsure of the comparison to an extra downloaded image of wood in rivers, as it seems quite detached from the traditional imagery which is primarily captured in these scenarios shown in Figure 7.
Conclusion	Overall, the conclusion is well written and summarises the results nicely. I think the assumption that the model does not understand 'what wood is' until you add some extra photos is overly critical, as the model is understanding 'what wood is' when transported and observed from a bridge or other viewing point. Finally, the next steps could be clearly grouped into some neat bullet points to show where the research should be heading.

Specific Comments

Section	Line	Comment
Introduction	Lines 19-20	There are quite a few more important sources of large wood, such as windthrow and natural mortality, or influence from fauna.
	Line 25	I think localised rather larger is more appropriate here when talking about inundation.
	Line 26	Could likely do with some more up to date references here to show advances in this area. E.g. https://www.mdpi.com/2076-3417/13/18/10454#B65-applsci-13-10454 and https://www.mdpi.com/2077-1312/10/7/911
	Line 36	Although this is introductory, I think some reference to those existing algorithms, and why they are location specific would be beneficial, especially as one of the goals of the paper appears to be to reduce the site-specific nature of current wood detection algorithms.
	Line 40	Comma needed after high-resolution aerial surveys. This occurs in other places where you are listing.
	Line 41	Give some examples of how RFID is being used, such as https://onlinelibrary.wiley.com/doi/full/10.1002/esp.3463?casa_token=G-p1V7DmbDEAAAAA%3AcGP5b3hqKzIPyE8YEHHSPK78ppWrXGMST7iPG-JZUsnupImrtM2Vs6gkX-LIQYRsjPiCq-bqfgCXEA, https://onlinelibrary.wiley.com/doi/abs/10.1002/esp.1888?casa_token=w1NFWrbZJ7gAAAAA:OYvqxuFyU7vHaDZ2FQ3hHxIDP474jAXCKdoHvR_oKZbKkLphb0btepE7Yw0yjn9ZpJW3KPwQb6tyYQ, and the authors prior work using RFID to improve CNNs.
	Line 54	Introducing a sentence along the lines of 'they are also limited by their spatial locations, and rely on specific setups being installed prior to an event' This would lead nicely into the use of citizen science.
	Line 61	Similar to the above, another link sentence here would help the flow, think along the lines of 'Advances in machine learning methods may help to overcome this and allow for widespread wood detection'.
	Line 62	I am not sure this first sentence is needed, feels informal and unnecessary.
	Line 67	Starting 'The CNN has multiple…', move this to next paragraph in methods and needs expanding (see comments relating to this below). Then add in some objectives as to how you plan to run, test, and evaluate your algorithm development.

Methods	Line 78	The first sentence explaining your choice of YOLO algorithm should be proceeded by a small review (one paragraph) on how CNNs work, and why they may be more suitable for detection than other algorithms, building on the information from the introduction. Following this, I feel that just saying YOLO was chosen for speed and accuracy is limited, especially as this is tested on generic imagery in their paper not large wood. Is there any reason to suggest it would be better for large wood? If not, have other studies trying to detect wood compared between algorithms? Think this is quite a crucial area to justify.
	Line 80	With above changes, a new paragraph could likely be started with 'Training a CNN…'
	Line 86	Combine these two sentences for better flow.
	Line 89 -95	I think some more details about the quality of the cameras here would be useful, such as resolution. Do you have a record of how much wood was added to each stream? How long have the monitoring programmes been underway in France and is any of that manual input to the channel? Where are the online videos from and what helps to make the images and wood a more diverse setting? These are all questions that need addressing. Does each image represent a single piece of wood, or are there more pieces of wood in each image, relating to the 15,228 number here. I think you should refer to table 1 here, and also adjust figure 2 as outlined below in figures and tables section.
	Line 98	Figure 2 also shows bounding boxes, maybe reference this instead.
	Lines 99 – 105	This section could do with some additional clarity, especially as this is an additional CCN alogirthm being deployed I assume? Were you checking that the automated bounding boxes for these were detecting wood, as this is not clear which dataset you are referring to by saying the labels were checked manually. There is also no explanation of why this worked better for 11 of the 15 datasets, or what your tolerance for acceptable mAP was, especially as not good enough was below 20 percent. Moreover, when images were checked manually, were incorrectly labelled frames eliminated or adjusted, or left as incorrect?
	Line 112	Why 80*80, is this purely incidental that no wood was larger than this, I also assume this is in pixel size not other units?
	Line 116	There is no statement of why this PCA was undertaken, and it only becomes clearer when reading the results. You need to add some context as to why this is undertaken. Furthermore, if the results of the t-SNE test are stochastic, could you not

		run the test numerous times to assess the diversity, akin to monte-carlo scenarios?
	Line 125	Swap must for 'is typically', if smaller datasets don't allow it, there is not always a split in this fashion or a separate test dataset.
	Line 125 – 142	This section is trying to explain a somewhat complex training and validation procedure, whereby computational trade-offs mean omitting some of your data as validation. However, it feels as though how these 6 examples were selected is not overly clear, besides not being at the same place and time. It may have made sense to use dataset 14 also, purely as that would give you validation samples at a range of sizes. The section took a while to become clear as to what the process was, and that datasets weren't being dropped from training, just the number of validation sets dropped. Perhaps trying to simplify the wording in places and go through the order. For example, 6 validation cycles were run, for each one a single dataset was dropped for validating and the model trained on the remaining 19. These 6 were chosen to represent a range of conditions, and reduced computational overhead by not undertaking 20 validation cycles.
	Line 139	Where has this extra dataset come from, and why was it not introduced with the other datasets? Who has been studying this, a research group(s) or monitoring agencies? This is a useful case example that in essence the paper could have been framed around. I.e. instead of can we implement a cool algorithm, can we reduce human labour of monitoring wood?
	Lines 145 – 152	This seems to be an odd way to do your sensitivity test, as although you are trying to identify the effect of number of inputs on output quality, if these inputs are multiplied for smaller datasets, then they are not adding any extra information, only overtraining the model? Would it not have been better to undertake this at a smaller number of images to assess performance, on possibly a limited number of datasets. I.e. for 10 of your datasets over 300, or 8 over the 1000, sequentially go from including 100, 200, 300, etc, and then quantify at what point there is no improvement in the model? This in itself could be one of the scenarios.
	Line 154	Although it is clear there are no river based large wood studies to learn from, it seems that casting the net a little wider shows these studies have been used in similar ways on living trees and perhaps other wood related scenarios. E.g. https://ieeexplore.ieee.org/abstract/document/9643113?casa_token=Vm749u_aLtQAAAAA:IQ8hGqEscqOOTf4M5Co8uVAJ1qsiJtDGUoMrQDFj-oSM14tTKiVBbKzIUl1GOOTwZ5AGgRy_qw

	Line 158	Although in principle I can understand how all these parameters effect wood detection, but has any worked actually been conducted on this? If so, reference it.
	Lines 158 – 160	It says 14 were trained and compared to a baseline, but there are 13 outlined. Either say 13 models trained, or 14 including the baseline for which the other 13 are compared to.
	Lines 162-163	Why were the values of 4% and 30% chosen, is there a rationale for this? The logic behind this makes sense, but just need to clarify reasoning for thresholds, even if they were just decided as no previous study to base upon.
	Line 170	Can the dataset size be given for total number of images, i.e. how similar is 'approximately the same'. This also feeds into informal language comments.
	Line 171	Why such a high number, when lowering it slightly could reduce the need for oversampling from some datasets? This appears similar to the sensitivity analysis you performed.
	Lines 176 – 179	Can you specify the number that were rotated vs mirrored, as the and/or makes it unclear if this was randomly done and randomly distributed. Were any both mirrored and rotated? I agree, that only partial rotation is necessary, this seems like a sensible decision to have made.
	Line 180	Again, how many were altered, and what proportion were mirrored or rotated. I think this needs more detail so the user knows what was done. How much extra data did this result in?
	Line 188	I feel that the inclusion of the phrase 'non-living wood' implies you are adding living wood, as opposed to wood that is not floating. Could be removed.
	Line 189	Change example to wood sample.
	Lines 192 – 196	This is an interesting scenario, primarily as these are open source datasets, with lower quality, but greater geographical diversion. In essence, I am not sure this is just testing data quality. Again, with the addition of other datasets, I think they should be mentioned in the original introduction of data, and their locations (even if approximate) included on a figure map. They can be highlighted/commented that they are only used for testing or specific scenarios, but curious as to why they were not included from the outset?
	Line 199	Which datasets were removed?
	Lines 201 – 206	This is really well explained and justified here, so should therefore be a model for your other scenarios where the justification is weaker.

	Lines 207 – 211	This is an interesting scenario to assess, as many secondary data sources may be of lower quality. However, are the double-precision images used either a) the down sampled images at 416*416 resampled again to a higher resolution, or b) the original images resampled to 832*832? The text make it seems like you double the resolution of the down sampled image (scenario a), as opposed to changing the original resampling (scenario b). Make this clear either way.
	Lines 213 – 226	This is a really well written section on the statistics being used, what they mean, and how a reader should interpret them.
	Lines 232 – 239	How is this sensitivity test different to the one introduced earlier, and why has this one got more dataset sizes to test the sensitivity? This is also not referred to in the results as far as I can tell, so what is the purpose of this section? The variance method also adds some confusion, is each of these models run several times, and the best results taken? If so, why the best results, does that not overestimate model performance? This could be explained better. You then talk about comparing between models, which again is fine but is very brief as to why, needs more explanation. You then mention a final model, is this not just your optimal model from all your testing?
	Line 238	Which dataset is this, is it the same as the one introduced previously on the river Inn? Again, this needs to be stated.
	Lines 240 – 244	This seems like a really sensible addition and is good to see some unpicking of what is happening behind the scenes. Maybe a brief idea of how this works, and what you hope to find and why you picked certain images (one river, across rivers, different angles?) would be sensible? In this case you might hope to hypothesise why some images/datasets are less well classified? This would be nice to see expanded on in the discussion.
Results/ Discussion	Line 248	I am not sure 'blob' is appropriate here, and if they are so small how can you be certain these are pieces of wood? You mention wood remaining stationary, does that mean moving wood was not included in the study?
	Lines 254 – 264	Unfortunately, no supplementary could be found on the online interface for comparison. However, I do wonder if whether double panelling a figure to include one of these plots for clustering with figure 5 could help to show the variation. I would also argue that the relative sizes of the bounding boxes compared to images were not that different, with many similar distributions and a few outliers, primarily from external datasets which is to be expected.

		You also state that for 12, 18, and 19, the drop in relative size could be due to low camera resolution or distance from stream, but 12 is one of the model setup cameras so surely you know this, and could tell for the others by looking at the original images?
	Line 265	Assume this is meant to be Database Configuration…
	Lines 266-270	As this is both a results and discussion section currently, there is a lack of discussion here about why this may be, and that by oversampling images you may not see an improvement in model performance purely due to the model become more tuned to those specific examples.
	Line 270	This is very important, if you do not now oversample, in your scenarios where you mentioned oversampling smaller datasets, did you now not do this? This seems like quite a big change. If so, I think the sensitivity results need to come within the methods inclusive so that you do not explain changes in your methods during the results.
	Line 273	What were these results, and are they really comparable considering the differences in the object types?
	Line 274 – 276	This section is not overly clear, I think it needs better wording to explain what is being done here, especially regarding the multiple training rounds. This feeds back into above comments at the end of the methods.
	Lines 281 – 284	I can see what is trying to be said here, about training for specific or general wood detection, but feel it could have been said better. This is also the first mention of how cameras were mounted, perhaps this should be mentioned in the data section also.
	Lines 285 – 293	There is a focus here on the high-definition wood images in this analysis, and yet there are only 9 images in the dataset. As such, are larger changes in mAP not more likely due simply to the lower number of objects to compare against? This is somewhat shown by the weighted average, and so overstating the importance of a vast performance decrease or increase here may be unjustified. The narrative however, that good wood images lead to better training than poor wood images, is justified by the average and weighted average outputs.
	Lines 294 – 295	Has a significance test been undertaken here? Are these broadly speaking not the only two factors, apart from manual labelling to begin with for training. What are the worst performing models?
	Lines 296 – 297	This sounds like you have added in an extra scenario, rather than describing one of your scenarios.

		Change 'where the datasets with lower performance than 30% mAP were excluded' to 'where the datasets with a mAP of lower than 30% were excluded'.
	Lines 300 – 301	Which scenario is this, can't find a reference to 19% in the table that is positive? If this is just assuming the inverse, then the addition of these images back wouldn't be the same 19% as the base conditions would be a different value.
	Lines 301 – 306	This is a really important and useful point, and should be one of the key take home messages that adding to existing databases with some data from a site improves the algorithms performance. Check some wording here though, especially when speculating performance benefits.
	Lines 307 – 313	This is an interesting section about whether the time component is critical. However, I fell it is overplayed in its significance. Of the two worst performing datasets (11 and 18) only one shows an increase of 6%, the other a decrease. Therefore, to say improvements of nearly 10% are made is an exaggeration. Arguably, this is somewhat upstaged by the large decrease in one of the better performing datasets (3).
	Lines 313 – 316	Make this a separate paragraph as it feels separate from the temporal component. Compared to the emphasis placed on scenario 12, scenario 13 appears to show much greater performance gains, and the importance of image resolution in tracking wood. As this has implications for how wood should be monitored, both from a hardware and software perspective, it likely needs more attention and discussion around the trade-offs between image resolution, computational efficiency, and expected wood size. Line 315 references image 5, is this from figure 2 as these seem to be larger wood size, if not, please be clearer as to what this refers to.
	Lines 317 – Onwards	This almost feels like a different section or subsection, as it is a change from training and validating to assessing the model used. It seems as though this section itself however is limited in just comparing two models, moreover, these results have differences greater than many of the scenarios provided above, which indicates that model choice may be more important than datasets, something that is not discussed in great detail. As a result, the take home would switch from the importance of data, to the importance of model selection in getting the best outputs…
	Line 329	Perhaps, if a new subsection is introduced for the above, this should be moved prior to it.

	Line 334	Reference figure 7 here, as it is not referenced anywhere in the text
	Lines 335 – 337	You identify that the model is better at identifying large wood, and then state how large wood components compromise the greatest proportion of transport, but this needs to be referenced to support this. Furthermore, small wood components also play a role in increasing the total volume of log jams etc and so important to monitor. Commenting on how this is missed in the dataset is probably needed. If possible, it would be great to look at those that are missed and estimate the size of these to identify a limit of detection. However, that may be beyond the scope of this investigation and potential for future research.
	Line 338	Have these images been georectified in the processing? If so this needs to be explained for reproducibility. Moreover, if they have then they could be used to identify the limits of detection for wood as per above?
	Line 342	Give examples here please, and comment on how they may differ or align to wood detection (e.g. shape and background).
	Line 342 – 346	I think this needs to be reworded, at times this sounds speculative and also non-scientific. The theory of not being able to detect outside of the training sample is sound, just the transmission of this information is not clear enough.
	Line 347	Where was this from and why not use one of your current data? Again, this points to questions going back to your initial data introduction, and consistently adding new bits of information.
	Lines 350 – 360	Does this not come back to simple survey and image design. If most of your images are from roads and bridges overlooking rivers, and you provide an image much closer to the channel, it will struggle, until as you say you include images of large bits of wood close up. Therefore, to use the word remarkably again seems a little overstated.
	Lines 357-358	Can you expand on how you know it is using the wood texture, is this hypothesised from the location of the pixels used, or can this be proven?
Conclusion	Lines 363 – 371	This is a nice start to your conclusion, summarising your results well to give an overview of the paper. However, there is no comment on how increasing data sizes or changing their angles/mirroring had no effect.
	Lines 372 – 382	I feel that to say your model struggles on the definition of wood, unless its given high-quality images of wood not in rivers, is overly harsh on your model. The purpose of this paper and method is to detect wood in rivers, likely from monitoring stations above the rivers surface (on bridges etc). So the

		model works if it detects these well, and shouldn't necessarily be able to detect wood such as in Figure 8. Therefore, the model CAN generalise the concept of wood 'in rivers', which is the main purpose is it not? I think the word blob should be removed throughout, perhaps in this instance they are best referred to as fragments or segments, i.e. not all the wood is on show? Make sure this distinction is first explained when replacing the initial occurrence of the word 'blob'. This may be clarified by an earlier point, is this 19% increase simply the opposite of the 19% reduction when the Allier dataset (18) is removed? If so. This is not 19% (e.g. 20% decrease from 100 is 80, a 20% increase from 80 is not 100). If this is a separate analysis, make sure this is clear during the methods and results. It could even be viewed as an additional scenario (e.g. adding same site from different date).
	Lines 383 – 387	This could likely be grouped into areas of future research. 1) real time monitoring 2) algorithm development and miniaturisation 3) temporal imagery for object detection. These could also form some structure for a separated discussion, allowing room to discuss the impacts of the research.

Figures

Figure	Comment
Figure 1	This figure could benefit from labelling the boxes with the sections of the method that they refer to. This will allow readers to quickly understand which bit of the process they are referring to. Make sure the naming matches to, it will help the reader. This could also be improved by creating this as an overall schematic of the methods, which would better describe the whole process as mentioned prior.
Figure 2	It is great to see some visual examples of what these images look like, and how they differ, especially in regard to the additional imagery. However, I think it would be good to possibly remove one or two images, and add an inset location map showing where in the world these were taken from, rather than coordinates in the caption. This would give a better idea to the readers of where your data is coming from. You could colour or size location dots based on the number of images from a location as well.
Table 1	Could this table also have a column or some stars which denote the datasets used in validation, these are mentioned later on but will help the reader when scanning back and forth. Consider making either camera lowercase, or the unknown and differing upper case.
Figure 3	Why is this figure not further up in the manuscript? It is referenced first several pages earlier and causes confusion in the current section. Appreciate this may just be a current formatting error for the preprint.
Figure 4	No changes required for this figure, it is clearly laid out, shows the size of datasets, and helps to explain what is happening in terms of the number of training vs validation datasets.
Figure 5	Again, another clear figure which adds to the manuscript and is broadly easy to interpret. The inclusion of a double headed arrow along the x axis, pointing to larger wood and smaller wood may help with interpretation, so readers know if the value is indicating a lot of the image is the woods bounding box, or little.
Figure 6	This figure is good, however it could do with stretching along the x axis, as this will help to show the variation in IOU training loss which show subtle differences.
Table 2	The table layout is fine, but the text is a little hard to read in places. For those reading in non-colour or with colour-confusion, perhaps as well as colours a marker could be used to quickly attribute greater than 3% increases or decreases.

Figure 7	A useful figure, make sure it is referenced in the text. Are these bounding boxes ones predicted by the model or drawn manually for users. It could be better to include boxes created by the model as well to show the types of wood it is missing (perhaps detected and missed wood as two separate colours?).
Figure 8	Are the bounding boxes in this figure manually drawn? If so, they should probably better align with the extent of the wood. Likewise, as the percentage is referring to overlap in bounding box size, perhaps indicating the bounding box of the detected wood would help to illustrate these differences? Otherwise, this is a very helpful and useful figure.