https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

Automatic detection of instream large wood in videos using deep learning

Janbert Aarnink[1], Tom Beucler[1,2], Marceline Vuaridel[1], and Virginia Ruiz-Villanueva[1,3]

[1]Université de Lausanne, Faculty of Geosciences and the Environment (FGSE), Institute Earth Surface Dynamics (IDYST), Quartier UNIL-Mouline - Bâtiment Géopolis, 1015 Lausanne, Switzerland
[2]Université de Lausanne, Expertise Center for Climate Extremes, 1015 Lausanne, Switzerland
[3]University of Bern, Institute of Geography, Hallerstrasse 12, 3012, Bern, Switzerland

Correspondence: Janbert Aarnink (janbert.aarnink@unil.ch)

Abstract. Instream large wood (i.e., downed trees, branches and roots larger than 1m in length and 10cm diameter) has essential geopmorphological and ecological functions supporting the health of river ecosystems. Still, even though its transport during floods may pose a risk, it is rarely observed and, therefore, poorly understood. This paper presents a novel approach to detect pieces of instream wood from video. The approach uses a Convolutional Neural Network to detect wood automatically. We sampled data to represent different wood transport conditions, combining 20 datasets to yield thousands of instream wood images. We designed multiple scenarios using different data subsets with and without data augmentation and analyzed the contribution of each one to the effectiveness of the model using k-fold cross-validation. The mean average precision of the model varies between 35 and 93 percent, and is highly influenced by the quality of the data which it detects. When the image resolution is low, the identified components in the labeled pieces, rather than exhibiting distinct characteristics such as bark or branches, appear more akin to amorphous masses or 'blobs'. We found that the model detects wood with a mean average precision of 67 percent when using a 418 pixels input image resolution. Also, improvements of up to 23 percent could be achieved in some instances and increasing the input resolution raised the weighted mean average precision to 74 percent. We show that the detection performance on a specific dataset is not solely determined by the complexity of the network or the training data. Therefore, the findings of this paper can be used when designing a custom wood detection network. With the growing availability of flood-related videos featuring wood uploaded to the internet, this methodology facilitates the quantification of wood transport across a wide variety of data sources.

1 Introduction

Instream large wood includes downed trees, root wads, trunks, and branches of at least 10 centimetres in diameter and 1 meter in length (Platts et al., 1987) and is typically recruited from forested areas within the river catchment by landslides, debris flows and bank erosion. As it distributes along the river banks, wood plays a crucial role by trapping sediment, creating pools, and generating spatially varying flow patterns (Keller et al., 1995; Andreoli et al., 2007; Wohl et al., 2018). Therefore, instream wood is a crucial driver of the rivers' form and functioning and positively influences the diversity of the river ecosystem (Wohl et al., 2017). Whilst beneficial for biodiversity, wood can also be a hazard. Given that wood is typically transported

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

during high flow conditions, floods often significantly influence the redistribution of instream wood. During floods, wood may accumulate at bridges or narrow river sections, blocking the channel and causing larger inundations (Lucía et al., 2015). Additionally, the pressure from wood can become large enough to damage or even wipe out complete bridges (Diehl, 1997; Lyn et al., 2003). Costly wood removal efforts have long been the default mitigation strategy (Wohl, 2014), without considering the ecomorphological impact (Lassettre and Kondolf, 2012; Collins et al., 2012). However, these preventive efforts can even be counterproductive. Logjams upstream from infrastructure can trap wood transported during high-flow events, preventing it from accumulating at critical infrastructures downstream. A more complex river system resulting from instream wood can also dissipate more flood energy when compared to a channelized river (Curran and Wohl, 2003; Hassan et al., 2005). By building infrastructure, channelling, and removing wood from rivers, human influence has impacted wood regimes and the river ecosystem. To assess the health of a river, it is crucial to get a better understanding of instream large wood dynamics by assessing the quantity and transport. The amount of observations of instream wood is, however, scarce. To improve our knowledge of wood dynamics in various rivers, we can analyze monitoring systems or crowdsourced videos of floods to detect and quantify the flow of instream wood. While existing detection algorithms have proven effective at locations to which they are calibrated, they lack widespread applicability, limiting mitigation strategies.

Wood transport data is scarce. Although rarely monitored, estimating the quantity of wood in river systems and its temporal variation has gained traction over the last years. Different techniques can help assess a river's wood regime in terms of transport, such as Radio Frequency Identification (RFID), high-resolution aerial surveys and video monitoring (MacVicar et al., 2009). With RFID tags, individual pieces of wood are given a unique identity, and their movement can be registered and tracked. Aerial data can detect stored wood and wood jams (Haschenburger and Rice, 2004; Lassettre et al., 2008; Sanhueza et al., 2018). However, the best methods to quantify wood transport are video-based because such methods provide high temporal and spatial resolution (Ghaffarian et al., 2020). Using computer vision software combined with stationary cameras for detecting wood transport has provided a first insight into river wood dynamics (Lemaire et al., 2015; Zhang et al., 2021). The approach uses spatial and temporal pixel-level analyses to, for example, detect features like colours, edges and moving objects. Its first feature is a mask to identify potential floating objects that differ in colour from the water's surface. Combining these features eventually determines the likelyhood of wood being present. Even though the utility of the software was proven, and it is used to extract wood from videos at multiple sites (Zhang et al., 2021), it still requires manual, site-specific tuning. It is purposefully designed for a specific site to increase performance. When creating a method for a specific location, it performs well on data with which it is designed but becomes too specific and complex to generalize over a wide variety of datasets. Furthermore, the current method requires the camera to be angled in a fixed position to extract the wood-detection features, which decreases flexibility. Even when tuned to a specific location, its performance depends on seasonal and weather conditions (Ghaffarian et al., 2021).

Developments in mobile technology have enabled millions of people to use high-quality video cameras. During extreme weather events, videos of floods are often posted online, which can be an exciting source for wood transport analyzes. Citizen science projects like the Argentinian Storm Chasers project have shown the use of home videos to analyze hydraulic conditions during a storm (Le Coz et al., 2016). Similarly, crowdsourced videos can analyze wood regime characteristics (Ruiz-Villanueva

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

et al., 2019). However, quantifying wood transport from videos recorded from non-fixed standpoints presents a challenge, as existing tracking methods fail to analyze crowdsourced footage due to their dependency on a stationary camera angle, necessitating a more flexible wood detection solution.

After first being proposed in the 1940s (Prieto et al., 2016), the neural network has made a recent come-back that went hand in hand with developing high-performance parallel computing systems (Zhao et al., 2019). The neural network has recently been found to be an effective method for object detection (Lecun et al., 2015), and was used for instream plastic detection (van Lieshout et al., 2020). In our study, we propose using a Convolutional Neural Network (CNN) to provide a flexible first step to analyze wood transport in videos from various sources and circumvent the limitations of the current state-of-the-art wood detection algorithm. The CNN has multiple convolutional layers in which it analyzes video frames. Convolutions are used to extract hierarchical features from images to make predictions. Features like edges, corners and textures are combined to determine the class of an object. The algorithm can learn which features an object needs to have for it to be classified as a piece of instream wood. These detection features do not have to be individually hard-coded and are created by training the network with class examples. Depending on the architecture of the CNN, it can therefore be a couple of orders of magnitudes more complex and theoretically more effective at detecting wood. The training of a CNN requires a large amount of data, ideally from vastly different sources in different weather and flow conditions (Bengio et al., 2013). Popular CNN architectures that have been developed for object detection include the Single Shot MultiBox Detector (SSD), the faster R-CNN, the CenterNet, Retina-Net and You Only Look Once (YOLO) (Liu et al., 2016; Ren et al., 2017; Duan et al., 2019; Lin et al., 2020; Bochkovskiy et al., 2020).

2 Methods

For the main part of this study, we have chosen to train the fourth generation of the You Only Look Once (YOLO) network. Among other object detection networks, YOLO stands out as it offers both speed and accuracy as it directly predicts bounding boxes and class probabilities automatically and thereby outperforms its competitors (Bochkovskiy et al., 2020). Training a Convolutional Neural Network (CNN) that can detect wood in various conditions requires multiple steps. First, instream wood data is acquired and labelled. Subsequently, the dataset is trimmed and augmented to create a database of varying images containing instream wood. Once these steps are complete, a large part of the database is used to train the model, whilst a small part is used to validate the training performance. In this context, “database” refers to all the data used to create the CNN, whilst a “dataset” is a subset of the database consisting of all the data recorded by the same device at a specific location and date. Figure 1 gives an overview of the data collection and processing. It shows that we assess the performance of 14 different augmentation and sampling strategies from the datasets.

2.1 Data

For this study, we used five low-cost cameras, including three Android-phones and two Raspberry Pi camera modules. The cameras are installed at different locations, on different days and times of the day, with different orientations. We manually

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

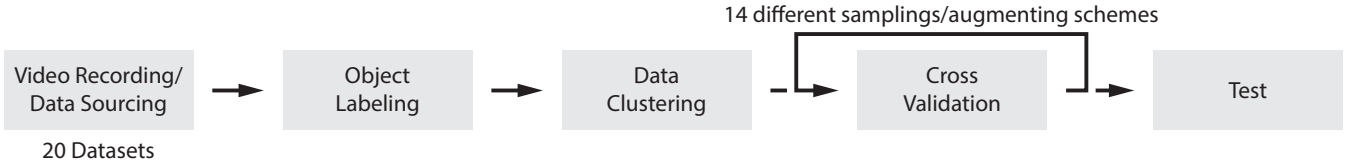


Figure 1. Overview of data collection and processing. Of the 20 labelled datasets obtained, 6 representative datasets are chosen for cross-validation (see Figure 4).

added wood to the river upstream from the cameras and let them record the wood passing by. We also used data from two locations in France that have been actively monitored for the last years at the Allier and the Ain Rivers (Zhang et al., 2021), and added carefully selected images of floating wood from online sources (purchased from istockphoto.com and dreamstime.com) to represent a small but diverse floating wood dataset. The final database consists of 15,228 images separated into 20 different datasets. For example, figure 2 shows 9 of the 20 different datasets.

2.1.1 Labeling

Labels are created on the data that is acquired to indicate where an instream piece of wood is located within each frame. Bounding boxes represent the four coordinates of a box's corners that fit around the piece of wood (see figure 8 for examples of bounding boxes). Initially, the labelling is manually done using labelling software called LabelImg (Viso.ai, 2022). To speed up the process, we devised a pseudo-labeling method. Only 10 per cent of the images in each dataset (1922 in total) were labelled by hand. A CNN (CenterNet, (Duan et al., 2019)) was trained to purposefully overfit that specific dataset by using only images from that particular dataset. With the CNN, the labels for the other 90 per cent of images are created. After that, the rest of the labels were checked manually. In 11 out of 15 cases, this method worked well. However, for four datasets, the CenterNet's performance was not sufficient to aid in the labelling of the other 90 per cent of images, as the mean Average Precision (mAP, see section 2.2.3) was below 20 per cent. Therefore, in these four cases, only the hand-labeled 10 per cent of images were used.

2.1.2 Data analysis

After labelling, we obtained 15,228 fully labelled video screenshots with bounding boxes (see figure 2) around the instream wood. When training a CNN, the goal is to have a lot of diverse data. Due to the natural way wood drifts and gets stuck, some videos contain minutes of the same piece recorded at the exact location. Therefore the data is trimmed. If, for subsequent frames, the labels were almost identical (with all bounding boxes within a certain percentage of each other), only 1 of the frames was kept in the database.

From the database, all labels were cropped out of their corresponding images and resized to 80 by 80 greyscale images, creating images of all 33,173 pieces of wood in the database without their surrounding image. Subsequently, the images were normalised and centred to eliminate circumstantial and camera-specific white-balance differences. This means that the average pixel intensity of each picture was set to 128 and the maximum or minimum pixel value to 255 or 0, respectively (see Figure 3).

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

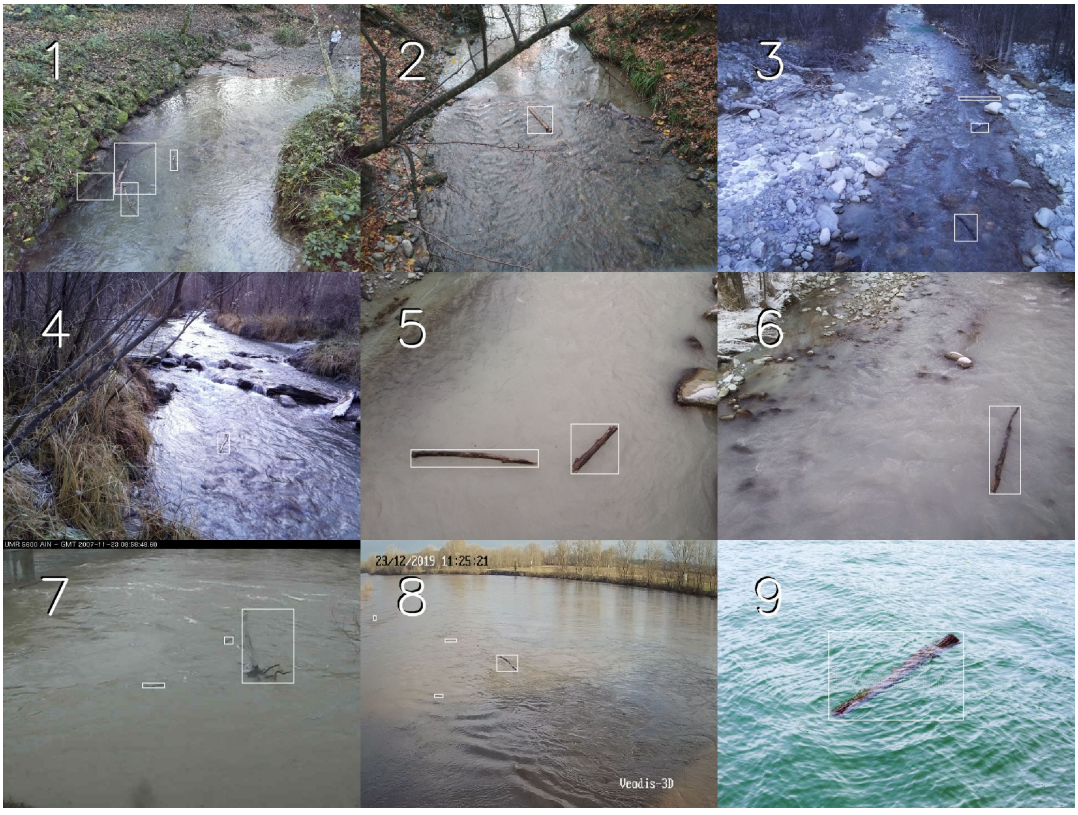


Figure 2. Data acquired at 9 out of 20 locations, with bounding boxes around instream wood. Images 1-2: La Chamberonne at UNIL campus, Switzerland (46.52373°N, 6.57729°E; 46.52296°N, 6.57577°E). Image 3: La Borgne d'Arolla, Switzerland (46.04814°N, 7.48884°E). Image 4: Dixence, Val d'Herens, Switzerland (46.17966°N, 7.4187°E). Images 5-6: La Borgne, Val d'Herens, Switzerland (46.1612°N, 7.44079°E; 46.10975°N, 7.49428°E). Image 7: Ain River, France. Image 8: Allier River, France. Image 9 is included to enhance data diversity.

We performed a Principal Component Analysis (PCA) on the set using the "clustimage" Python package (Taskesen, 2021). PCA is an unsupervised dimensionality reduction technique and, in our case, clusters of similar images in a predetermined number of clusters. The silhouette score evaluates the similarity of each of the 6400 pixels (arranged as 80×80 pixels) within a cluster compared to their dissimilarity across different clusters, effectively gauging the compactness and separation of the clusters.

After the PCA, we performed a t-Distributed Stochastic Neighbor Embedding (t-SNE) to visually review the diversity of cropped-out pieces of labeled wood. The stochastic nature of the t-SNE method means that each run might yield different results, so we used it for visualization purposes only in our study. Furthermore, we compared the relative size of the bounding boxes from dataset to dataset to understand the difference in the data.

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

Table 1. Data acquisition statistics.

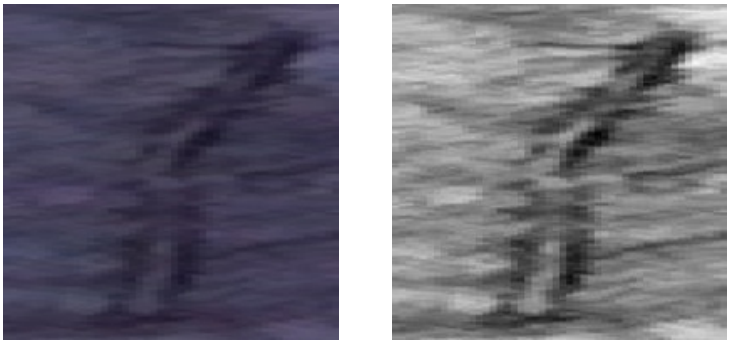
Dataset number	Amount of labeled images	Device	Resolution	Location	Number in example dataset (figure 2)
1	1.429	Raspberry Pi Camera	1920x1440	La Sorge, loc 1	1
2	601	Raspberry Pi Camera	1920x1440	La Sorge, loc 1	
3	1.076	Samsung Galaxy A4	3264x2448	La Sorge, loc 1	
4	478	Xiaomi Redmi 4X	4160x3120	La Sorge, loc 1	
5	344	Xiaomi Redmi 4X	4160x3120	La Sorge, loc 2	2
6	2.478	Raspberry Pi Camera	1920x1440	La Sorge, loc 2	
7	2.146	Raspberry Pi Camera	1920x1440	La Sorge, loc 2	
8	191	Samsung Galaxy A4	3264x2448	La Sorge, loc 2	
9	18	Xiaomi Redmi 2	3328x2496	Borgne d'Arolla	
10	138	Raspberry Pi Camera	1920x1440	Borgne d'Arolla	
11	1.046	Raspberry Pi Camera	1920x1440	Borgne d'Arolla	3
12	1.034	Raspberry Pi Camera	1920x1440	Dixence	4
13	157	Raspberry Pi Camera	1920x1440	La Borgne	5
14	2.340	Raspberry Pi Camera	1920x1440	La Borgne	6
15	1.236	Samsung Galaxy A4	3264x2448	La Borgne	
16	116	unknown	640x480	Ain	7
17	81	unknown	640x480	Ain	
18	176	unknown	1920x654	Allier	
19	134	unknown	1920x1080	Allier	8
20	9	differing	differing	differing	9
Average	678		2247x1673		
Total	15,228				

2.1.3 Data clustering: six representative datasets

In machine learning, the data must be split up into training, validation, and test data for evaluation purposes. The training data is used to optimize the neural network to detect the training examples, the validation data is used to optimize the neural network's hyperparameters (e.g., its architecture), and an unseen test dataset is used to analyze the network's performance (Xu and Goodacre, 2018). Even though the amount of labelled bounding boxes was large, the difference in the data sources was small as the 20 datasets had common locations and camera angles. Therefore, the traditional method of splitting the data into unique train and validation datasets was prone to high variance. To overcome this issue, cross-validation can be used where the same validation technique is used multiple times with different training and validation data.

As the database was separated into 20 different datasets (each with a different camera or location), in an ideal evaluation approach, a leave-one-out cross-validation using 20 different validation datasets would suggest how well different sampling

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024



(a) (b)

Figure 3. Original cutout (a) and greyscaled, normalized and centered cutout (b) (2020-11-29 Raspberry Pi 4 image 7411 label 2).

and augmentation scenarios would increase the model's effectiveness under different circumstances. However, as training 20 models 14 times uses too many resources, the 20 datasets were reduced to 6 representative validation datasets. We chose these six datasets not to be taken on the same day or at the exact location. For each of the six training instances, one of the six representative datasets was left out and used as a validation dataset (see figure 4). We repeated this process for 14 different training scenarios as described in Section 2.2.2.

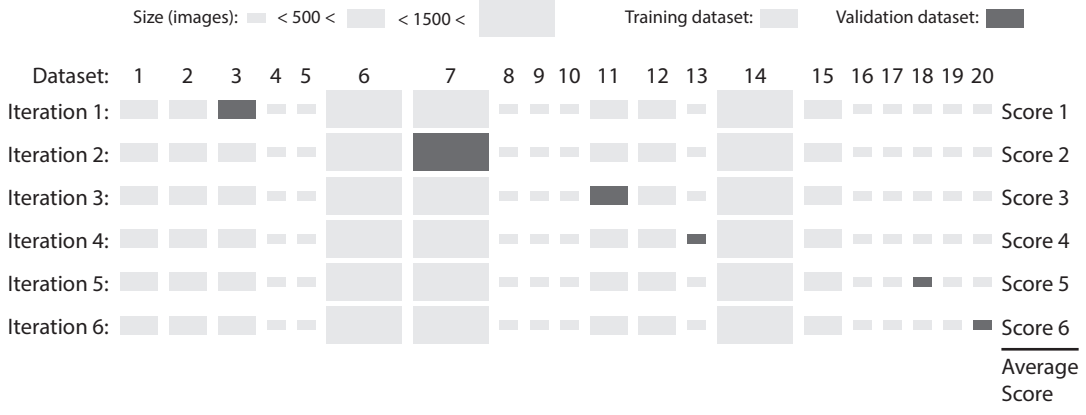


Figure 4. Cross validation scheme for one training scenario. This figure shows the distribution between the datasets used for training and those used for validation. The Y-axis contains all available datasets, and the X-axis contains the different training efforts, in which the dark grey dataset represents the validation data whilst the other 19 were used for training the model. Eventually, the six scores are averaged into the validation score. The size of the rectangle represents the size of the dataset.

Eventually, the scenario that yields the best performance was tested with an independent test dataset. The test dataset was gathered at the river Inn at the moment of an experimental flood at its tributary, the Spöl River. This location has been actively

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

studied for wood transport since 2018 and yields a valuable test dataset because an algorithm like ours could reduce human labor significantly.

2.2 Machine learning methods

2.2.1 Sensitivity test

The 20 training datasets have an average size of 761 images and a median size of 411, with some datasets containing almost 2,500 images. To reduce computational demands, we conducted tests to evaluate how the number of images sampled per dataset affects training performance, aiming to minimize computations for later scenarios. We conducted two tests with a fourfold difference in the total number of training images compared. In the first test, 2,000 images were augmented and sampled per dataset. Since the datasets did not all contain this exact number of images, they were randomly up- or down-sampled to reach this figure. We applied the same approach for the second test but with 500 images per dataset. Thus, a total training sample size of 38,000 images for the first test (2,000 images per dataset across 19 training datasets) was compared to 9,500 for the second test (500 images per 19 datasets). Following this comparison, we determined the optimal sample size for further exploration.

2.2.2 Thirteen training scenarios

As CNNs have not yet been trained to detect instream large wood, thorough testing of different training strategies is crucial. To enhance model performance, the database size can be expanded by adding more labelled images or through synthetic augmentation of existing data. Additionally, employing various sampling strategies may improve the algorithm's detection performance. The data used for wood detection can also vary significantly regarding camera angle, pixel size, and proximity to the stream (see Table 1). To determine the most effective sampling and augmentation strategies for different data types, 14 models were trained and evaluated against a baseline model. The baseline model was trained using only the labelled images without any modifications. The other 13 scenarios are detailed below:

1. **Trimmed, testing sensitivity to handling stationary frames**: When labels are similar in at least 3 subsequent frames, the images are deleted from the database. Detections were considered similar when all bounding boxes were within 4 per cent of their subsequent x and y location in the frame and 30 per cent of their width and height. In this scenario, 13,375 images from the total database are kept.

2. **Sampled V1, testing sensitivity to dataset size ("min500 max1200")**: As small datasets are undersampled compared to large datasets, in this scenario we sampled a minimum of 500 images per dataset and a maximum of 1,200 images per dataset. If the dataset was smaller than 500, we oversampled images randomly and added the duplicates to the dataset until we reached 500. We did not use all the data if the dataset was larger than 1,200. These numbers were chosen because when applying this sampling method to the 20 datasets, the total amount of images was approximately the same as when using all original data per dataset.

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

3. **Sampled V2, testing sensitivity to dataset size ("750")**: In this sampling scenario, to sample equally from every dataset, 750 images from each were used.

4. **Sampled V3, testing sensitivity to dataset size ("min500")**: To not delete data, in this scenario, only the small datasets were randomly oversampled to have a size of at least 500 images. As we kept all data in the other datasets, the total amount of training images was larger when compared to the baseline.

5. **Augmented V1, testing sensitivity to data augmentation ("mirrored rotated all")**: To increase the diversity of the data, the images were all used, and duplicates were mirrored and/or rotated. The rotation was kept between -15 and 15 degrees; in practice, the river is almost always at the bottom of the frame. This dataset contained twice the amount of images as the baseline.

6. **Augmented V2, testing sensitivity to data augmentation, ("mirrored rotated random")**: To increase the diversity of the data, the images were randomly selected to be mirrored and/or rotated. The rotation was kept between -15 and 15 degrees.

7. **Augmented V3, testing sensitivity to data augmentation ("only mirrored")**: The images were only randomly mirrored to disentangle the mirroring and rotation effect.

8. **Augmented V4, testing sensitivity to data augmentation ("only rotated")**: The images were only randomly rotated between -15 and 15 degrees, to disentangle the mirroring and rotation effect.

9. **Added V1, testing the sensitivity to data quality, added high definition non-floating wood**: In an attempt to increase the model understanding of non-living wood, we added photos of instream wood laying in the mostly dry riverbed from an earlier survey to the database. A total of 167 photos containing at least 1 example were added. The added data had pixel dimensions 4608 by 3456 and was of higher quality than the other 20 datasets (see Table 1). Here, the influence of bbox size and data quality were tested.

10. **Added V2, testing the sensitivity to data quality, added 12 datasets**: From videos found online, a subset of the frames were labelled and added to the training database. A total of 12 datasets ranging from 8 to 499 images per dataset were added from locations in North America, New Zealand, and Switzerland. The total amount of images added in this scenario was 1206, with an average pixel resolution of 1650x1133. Also, as taken mainly from the internet, the data was compressed. Therefore, the added data quality was worse than the original 20 datasets.

11. **Removed, testing the sensitivity to data quality, removed worst performing datasets**: As lower quality data can weaken the models' understanding of wood, in this scenario, the quality of the data can be analyzed by the effectiveness of the model trained in the base scenario to detect samples. In this scenario, the two datasets at which the base model performed the worst were removed from the data to see whether the detections on the other datasets got better.

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

12. **Merged, testing the sensitivity to adding a time component, merged three images into 1**: Because often the distinction between a piece of instream wood or flow features like eddies and wave ripples was not clear from a single image, in this scenario we merged three images into one image after converting them to greyscale. Therefore, instead of regular Red, Green and Blue bands, the model was trained on greyscale images at T-1, T and T+1, with T the timestep at which we were detecting. This was hypothesized to aid the detection as waves and eddies change during a short timestep whilst the wood holds its shape.

13. **Double Resolution, testing the sensitivity to increase the input image size to the model, from 416 to 832**: A CNN is trained based on a specific pre-defined image resolution. As standard, images are resized to a 416x416 image resolution before they are used in training and validation. Decreasing the images could result in data loss, especially in cases where the relative size of the wood pieces were small. Therefore, we tested the model's sensitivity to the input image size in this scenario by performing the same tests with double the resolution (832x832).

2.2.3 Model evaluation

Generally, a commonly used metric for object detection tasks to evaluate performance is mean Average Precision (mAP), which combines three different measures: precision, recall and intersection over union (IoU) (Zheng et al., 2020). Recall is the percentage of wood pieces detected by the algorithm out of all the logs that pass by. Precision indicates whether the piece the CNN detected were indeed instream wood. The object detection algorithm outputs either no bounding boxes or (multiple) bounding boxes for each image. Each bounding box indicates the outer limits of the object and has a confidence percentage corresponding to how certain the model is in its detection. More bounding boxes are classified as a detection when lowering the confidence threshold. Hence, the recall increases, and the precision decreases. The changes in precision and recall based on the threshold can be displayed in a precision-recall curve. The surface under the curve can be translated into a single average precision (AP) value for a specific IoU. However, this value does not compare different IoU thresholds. The IoU compares the label with detected bounding boxes by dividing their overlap by their combined total surface. For each IoU value, a different precision-recall curve can be created, resulting in different APs. When all different APs based on different thresholds and IoUs were combined into a single value, we got the mAP, which ranged between 0 and 100 %. With an upper limit of 100% mAP, the model would have labelled every instance of instream large wood exactly as the humans that labelled the training data. However, as human labelling is imperfect, the mean Average Precision was still not an objectively perfect performance index.

Different applications of object detection call for different thresholds in recall, precision and IoU, depending on the consequences. Depending on the large wood regime of a specific river, more emphasis can be laid on either the recall or the precision. When the amount of wood passing is very low, e.g., one piece of wood a month, increasing recall can ensure that a piece is not missed. However, when using a too high sensitivity, the model could wrongly detect wood in each frame, forcing the user to look at every image and delete all the false detections.

For this study, the same experiments were performed using 2000, 1000 and 500 augmented images for each of the 20 datasets initially. The performances were compared to test the model's sensitivity to the amount of images used. Then, for all of the

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

14 training scenarios we trained a model and validated it in the six validation datasets. To mitigate the variance in training results we performed the test three times and used the two best mAPs of every training instance to get robust results. After that, on a small subset of training scenarios, a newer YoloV7 model was trained to compare the results of different models on the same data. A final model was trained after determining which training strategies work best for which data types. This model was tested on a test dataset that had never been used in any of the analysis and training efforts. In this way, the test dataset represented a case in which an unrelated wood monitoring study would use the model for out-of-the-box detection.

Finally, neural networks for object detection are often considered black boxes, which decreases their trustworthiness. To increase transparency, algorithms were developed to reverse the detection process and find the input pixels from the image weighted the highest when the process decided whether or not to detect an object. For the YoloV4 algorithm, a Python package called "Yolime" was used (Sejr et al., 2021) for this purpose. Different instream wood samples from the database were hand-picked, and an algorithm was run to determine which pixels weighed the heaviest.

3 Results and discussion

Our research led to the development of a comprehensive database containing labelled videos of instream wood. The labels cover a broad spectrum of sizes, particularly when comparing different datasets. They ranged from clearly identifiable downed trees with distinct features such as bark, branches, and brown colour to bounding boxes resembling small blobs that spanned only a few pixels in both dimensions. Also, distinguishing water waves and eddies from pieces of instream wood can sometimes be challenging during labelling. Frequently, skipping through the frames provided a more precise understanding, as a wood piece would remain stationary in the frame while a wave would have dissipated. This observation suggested incorporating a temporal dimension could significantly enhance the detection process.

3.1 Training data: diverse but still clustered

Based on the analyses performed, the data utilized in this research appeared to be diverse. The PCA (Principal Component Analysis) yielded a low silhouette score, and a visual inspection of the t-SNE (t-Distributed Stochastic Neighbor Embedding) plot revealed only small clusters of similar data (see Supplementary Information), suggesting the presence of duplicates within the data. However, the analysis also uncovered similarities within each dataset. A comparison of the average bounding box sizes across datasets revealed distinct differences. Figure 5 illustrates the relative size of the bounding box compared to the overall image size, highlighting significant discrepancies in the sizes of labelled pieces of wood across different datasets. To adjust for the exponential distribution of calculated surface areas, we applied a square root transformation to the bounding box area for better visualization. The graph indicates that datasets 12, 18, and 19 were either of lower quality (due to lower pixel resolution) or were positioned further from the stream. This suggested that the model may learn to recognize instream wood as large pieces within the camera frame, with distinct bark and other features easily associated with wood in great detail in some datasets, while in others, it might identify it as merely a few pixels clustered together, resembling a blob.

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

Figure 5. The relative size of the wood pieces compared to the image size per dataset. The relative size is represented by the square root of the surface of the bounding box sizes divided by the square root of the total image size. To make the interpretation of the figure easier, we take the square root of the relative size. The datasets from table 1 are indicated by the number. The 'A' indicates datasets that are added in scenario 12. The 'D' indicates the dataset added in scenario 11.

3.2 Training results: fatabase vonfiguration matters most

First, analyses showed that the model's performance did not increase by oversampling data from 500 (9,500) to 2000 (38,000) images per dataset. The best mAP was similar for both tests. Figure 6 shows the difference between a training instance with 2,000 augmented images per dataset on the left and 500 augmented images per dataset on the right. It shows that increasing the data for a training instance slows down the model convergence without increasing eventual mAP. Testing the 14 different scenarios is hereafter done without the need for oversampling.

To test the model's sensitivity to stationary frames, dataset size, data augmentation and data quality, 13 different testing scenarios were performed. Table 2 shows the training results of different scenarios. When using the total amount of 15,228 labelled images, the average performance in mAP was 63.42%, which was similar to the results that van Lieshout et al. (2020) found when creating a plastic detection algorithm. The table shows the average mAP of three training sessions started from the same pre-trained weights for the six representative datasets. The two best performing versions of the model were considered in each training session. Therefore, the displayed mAP in the first column is the average of six mAPs. The following columns show the difference in performance (mAP) of the base scenario compared to the scenarios based on the six best performances. At the bottom of the table, the average of the six mAP and the weighted average are shown. As there was a large variety in the sizes of the datasets, to not overestimate the importance of small datasets, the weighted average compensated for the relative size of the datasets.

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

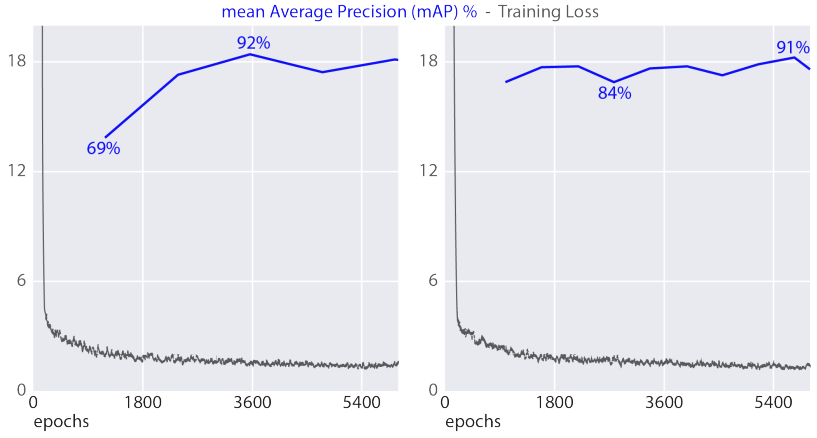


Figure 6. Training performance of dataset 13 (see Table 1) when using 2,000 images (left) and 500 images (right) per dataset (respectively 38,000 and 9,500 training images used). Blue: mean Average Precision, Grey: Complete Intersection-Over-Union Training Loss (Zheng et al., 2020)

Table 2. Neural Network performance for the base scenario in mean Average Precision (mAP, see section 2.2.3), and relative change in mAP when using the 13 different training scenarios. mAP changes of more than 3 per cent positive or negative are indicated in blue or red.

Scenario (dataset number)	Base **Jpgs** (mAP)	1 **Trimmed**	2 **Sampled V1**	3 **Sampled V2**	4 **Sampled V3**	5 **Augmented V1**	6 **Augmented V2**	7 **Augmented V3**	8 **Augmented V4**	9 **Added V1**	10 **Added V2**	11 **Removed**	12 **Merged**	13 **Double Res**
Sorge Samsung Galaxy A4 (3)	80.33	0	-3.33	-2.67	-1.33	-1.5	-2.17	-1.83	-2.5	-1.17	-1.83	-1.33	-18.83	3
Sorge Pi 4 (7)	71.17	-0.17	-1.67	-3.67	-0.83	-2.17	-3.83	0.17	-1.67	-0.17	-1.17	-1	-3.17	5.17
Borgne d'Arolla Pi 4 (11)	43.67	5	14.67	13.67	15.33	4.83	0	0.5	-2.5	-1.17	-1	2.5	6.17	17.17
Dixence Pi 4 (13)	92.83	0.5	-0.17	-1.67	0.67	1.17	2.33	-0.33	2.17	0.17	-0.67	0.17	1.17	-2.17
Allier (18)	35.33	-0.5	-3.83	-3.67	-3.67	-3.5	-1.17	0.33	-1.83	-0.33	3	-19	-1.5	14
Random Images (20)	57.17	-6.83	2	3.33	-4.67	-11.33	-9.83	-1.17	-16.5	22.83	-13.83	11.33	-*	-22.33
Average	63.42	-0.33	1.28	0.89	0.92	-2.08	-2.45	-0.39	-3.81	3.36	-2.58	-1.22	-*	2.47
Weighted Average	66.7	0.96	1.40	0.29	2.42	-0.50	-2.38	-.21	-1.94	-0.55	-1.14	-0.91	-*	7.27

* dataset 20 does not contain a timeseries and is therefore not possible to merge

Dataset 20 was small but had large variability incorporated in the images as they were taken from different sources. Therefore, the result from this dataset can be used to analyse the model's capacity to generalise the wood concept. On the other hand, the weighted average was a better performance metric when the wood detection we were interested in resembled the variety of the training data, which was, in this case, acquired mainly by cameras attached to a bridge.

285 Table 2 shows the increase in model performance when changing the sampling strategy. When adding data from videos of floods containing instream wood found on YouTube and Twitter (scenario 10), the average mAP for all datasets went down by 2.58%. This decrease in performance can arguably be attributed to the low quality of the data being confusing to the model. This observation was also strengthened by the vast decrease in performance (-13.82%) of this scenario when validating the

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

high-definition random wood images dataset. Also, the model performed better when tested on the Allier River dataset, which contains largely smaller (lower quality) samples of wood (see Figure 5). Instead of adding lower quality data, we added high-definition data of non-floating wood to the training data (Scenario 9), the large increase in performance when validating dataset 20 was explained by the algorithm generalizing the concept of wood. This increased the average performance but decreased the weighted average performance as the overall average label sizes of dataset 20 were relatively small.

With a significant difference between the best and the worst performing dataset, it can be argued that the algorithm and data quality is the limiting factor of the approach. This was reinforced by qualitative analysis of the two worst-performing datasets. As this data might confuse the model, another experiment was performed where the datasets with lower performance than 30 % mAP were excluded from the training data (Scenario 11). The results (table 2) showed a weighted average decrease of 0.91%. This decrease was primarily linked to the worse performance of the model on the Allier River dataset (18). The decrease was large because the excluded dataset was taken from the same data source on a different day. Therefore, it can be argued that the model was still shown to overfit the training data even with the precautionary measures. Reasoning the other way around, adding data (176 images in this case) from the same source but on a different day can increase model mAP by 19%. This shows a more practical implication for researchers. When starting a new monitoring project, is it good practice to label a training dataset and add it to the larger database. In this way, one can train a site-specific wood detection algorithm which is shown to be tens of percent better than the model out of the box. The above findings showed that although the validation data used to calculate the mAP was taken from a different data source than the training data, there was still similarity. Data from the same camera on different days or the exact location and date taken with different cameras were not completely different.

The results of scenario 12, where we merged three frames into one to integrate a time component, yielded exciting insights. For datasets where the model already demonstrated robust performance, the accuracy experienced a noticeable decline. On the other hand, on datasets where the initial model struggled, a remarkable improvement of nearly 10 per cent was observed. This suggests that incorporating temporal information might be particularly beneficial when distinguishing between subtle features, such as pices of wood and waves, proves challenging in a single frame. Further investigation into the nuanced impact of this temporal integration is needed to understand the specific scenarios where this approach is advantageous. These findings underscore the potential of leveraging temporal information to improve river wood detection. Lastly, scenario 13, where we doubled the image size after rescaling, shows increasing performance on 3 datasets specifically (7, 11 and 18). These are cases where the relative size of the wood examples is low (see image 5) and can therefore be missed when decreasing the image size too significantly before running detections.

The field of machine learning-based object detection moves fast. New versions of the existing state-of-the-art model are released every year. Therefore, we compared the performance of the 4th version of You-Only-Look-Once model to the 7th (Wang et al., 2022) version. When comparing the training results on the same data using the base scenario, the results are shown in table 3. Even though the model became more efficient and smaller (43% smaller from V4 to V7) and the input image size was larger (640x640 for V7 and 416x416 for V4), the performance did not drastically increase in our case. The average mAP went down by 4 per cent, whilst the weighted average went up by 2.5 per cent, mainly because the model performed better on the largest dataset. This indicates that the newer model can be trained to perform better for specific datasets. Still, when

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

Table 3. Comparison with YoloV7

	mAP @ 0.5		
Dataset No	**YoloV4**	**YoloV7**	**Difference**
1	80.33	77.18	-3.15
7	71.17	78.53	7.36
11	43.67	43.92	0.25
13	92.83	90.29	-2.54
18	35.33	21.07	-15.26
20	57.17	44.82	-12.35
Average	63.42	59.30	-4.12
Weighted Average	66.41	69.01	2.60

using a model without finetuning it to a particular study site, the 4th version of the YOLO model performed better. According to Wang et al. (2022), with more conventional ML benchmarks, the performance of YoloV7 did increase. This can indicate that the training data's quality and/or diversity is still the training process's bottleneck in our setup.

These results show that even though the model is effective in detecting wood in rivers, the mean average precision is still dependent on the composition of the data used for training and is, therefore, still overfitting in some instances.

To evaluate the model's performance, a novel test dataset was introduced, captured during a flood event in the River Inn. Notably, this dataset had never been used by the model during its training phase, and no adjustments were made to hyperparameters based on this new data. The mean average precision (mAP) from the flood dataset was 61 per cent when using the base scenario. Increasing the input resolution of the model to 832x832, as per scenario 13, did not increase the performance (60.5% mAP). It is essential to highlight that this accuracy was achieved despite the flood event's challenging conditions and the imagery's relatively low-quality nature. Images with dimensions of 1280x720 pixels were captured using a mobile phone in timelapse mode. Furthermore, it was found that the model is better at detecting larger pieces of wood than smaller pieces. The ability of the model to identify larger wood elements is essential for its practical applicability. Large wood components often constitute a substantial proportion of total wood transport within rivers. Hence, combining our deep learning model's proficiency in detecting wood and georectification techniques facilitates the quantification of wood transport in river systems. The results suggest that the model can be used to estimate and monitor wood transport dynamics in rivers, providing valuable insights into the ecological and geomorphic processes associated with fluvial environments.

3.3 Understanding model predictions: wood features, surrounding water, and object size

The effectiveness of CNN has been displayed in various fields (Kaur and Singh, 2023). However, due to the model being considered a black box, its trustworthiness can suffer. A CNN is believed to make detections by understanding the concept of wood and use wood features like bark, branches and colour to infer wood. However, it might take shortcuts and use different

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

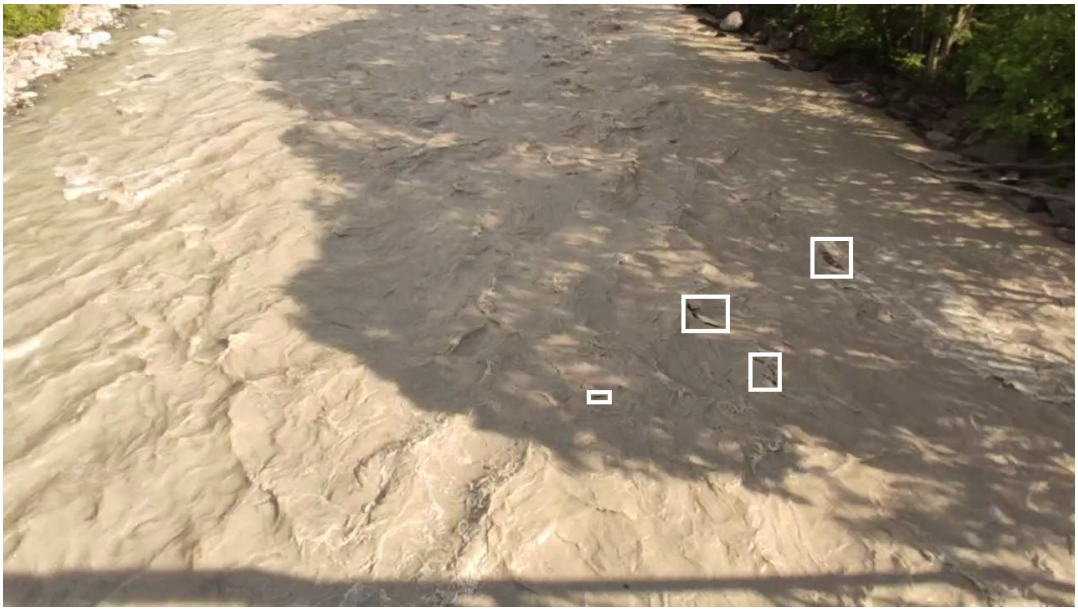


Figure 7. Example frame of the test dataset.

characteristics of the training data to determine whether an object is wood. If this is the case, the model would struggle to perform on datasets that do not contain those characteristics. Therefore, it is essential to understand the model predictions.

We used a picture of instream wood found online to analyze which pixels in a figure were weighed the heaviest by the model to determine whether an object was a piece of instream wood. Figure 8 shows the inference of the image when using the base scenario as described in section 2.2.2 as compared to scenario 9. It also indicates the pixels the neural network uses to detect wood in the image. Remarkably, not just the pixels representing wood were indicated as applicable to detect instream wood. The training data contained almost exclusively pieces of floating wood, and pieces on the bank that were not floating were not indicated. Therefore, the network seems to require the indication of water-containing waves next to the piece of wood to detect instream wood. Also, in the base scenario, most of the training data contained small pieces of wood with a small relative bounding box size. Therefore, in the left image, the confidence of the model in the detected piece being wood is low, as the training data did not contain a lot of similar high-definition images. In scenario 9, however, high-definition images of non-floating wood were added to the training database, and therefore, the inference yields different results. This image resembles the added images; consequently, the piece was indicated as wood with a higher certainty. Also, interestingly, the model seems to include the wood (bark) texture for its detection. These findings underscore the hypothesis in section 3.2 that there was a delicate balance between wood detection and small-object (blob) detection, primarily driven by the average size of samples in the training data.

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

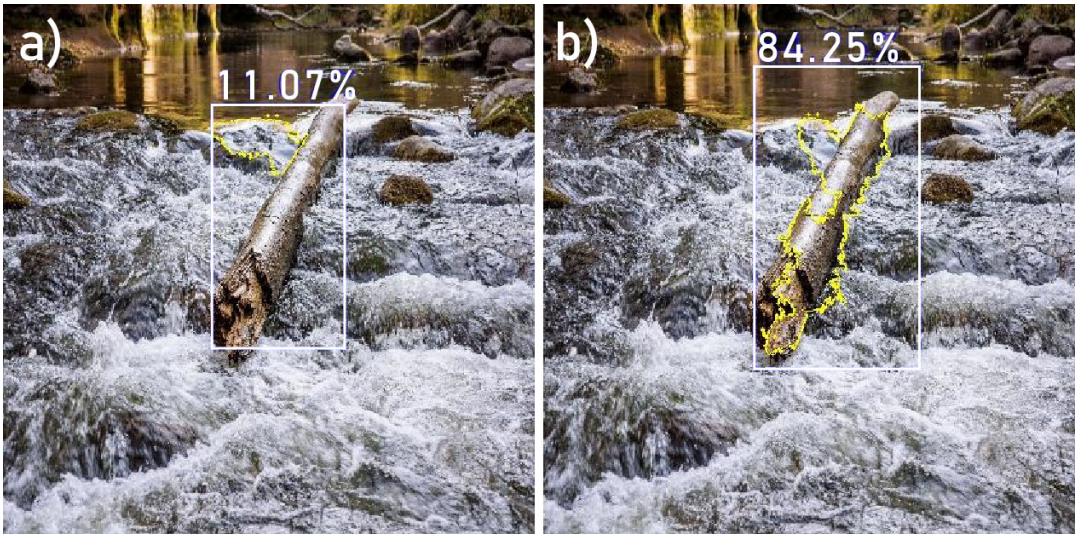


Figure 8. Wood detections according to a) Baseline scenario and b) Scenario 11: 'Added V1'. In yellow, we indicate the heaviest weighting pixels according to the neural networks.

4 Conclusions

We trained a Convolutional Neural Network to detect instream wood with a weighted average performance of 67 per cent mean Average Precision (mAP). On the best occasion, the model had a mAP of 93 per cent on one specific dataset. It has to be noted that the performance was sensitive to the quality of the images in the training data, as concluded by a wide range of results. On an unseen test dataset, its performance of 61 per cent mAP was in line with the results from the sensitivity tests. Efforts to improve the model's performance were, in some cases, successful. Depending on the data that was used for training, we enhanced up to a 23 per cent increase in mAP. Changing the sampling strategy by adding or removing training data yielded considerable differences in average performance. Also, although increasing in image input resolution increases the processing time and makes the method more costly, in some instances it did result in an almost 20 percent increase in mAP. Even though it was attempted to create a training database with various examples, the training results still indicate the model overfitting the training data.

This study shows that the model can generalise the concept of wood, but primarily when training data consists of high-definition photos of labeled wood. It still struggles to scale from high-definition images of wood to low-definition small samples. In the database, large examples of wood were notably different from examples that were only a couple of pixels in size and could, therefore, not display any characteristics of wood. A distinction between wood detection and blob detection can therefore be made. When creating a custom wood detection network, it is crucial to know the data type that must be analyzed and use the training datasets that resemble it. In the research process, a labeled training database of over 15.000 images was created. The training data is hosted publicly and can be used for future object detection refinements. Also, as the data is separated based on location and date, a customized model can be trained using the data that most closely represents the data of the

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

person interested. For a new wood detection study, custom-labeled data can be added to the training database to increase the performance even more. Adding only 176 labeled images of the same monitoring station but on a different day increased the model's performance by 19 percent.

The method described in this paper cannot be used in real-time. In future efforts, smaller versions of the tried models could be developed to run on in-field or mobile devices. Furthermore, a Tiny version of the YOLO model that scales down the model to be run locally is available. Lastly, as wave ripples often resemble pieces of wood and, in certain instances merging three subsequent frames seems to improve results already, future efforts of improving the model could benefit from including the time component of a video into the detection algorithm.

Code availability. https://github.com/janbertoo/Instream_Wood_Detection

Data availability. The data to which we have the rights is available at: 10.5281/zenodo.10822254 .

Author contributions. Study conception and design: JA, VRV
Data collection: JA, MV
Methodology design: JA, TB, VRV
Analysis and interpretation of results: JA, TB, VRV
Manuscript preparation: JA, VRV, TB.
All authors reviewed and approved the final version.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work has been supported by the Swiss National Science Foundation project PCEFP2186963 and the University of Lausanne.

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

References

Andreoli, A., Comiti, F., and Lenzi, M. A.: Characteristics, distribution and geomorphic role of large woody debris in a mountain stream of the Chilean Andes, Earth Surface Processes and Landforms, 32, 1675–1692, https://doi.org/10.1002/esp.1593, 2007.

Bengio, Y., Courville, A., and Vincent, P.: Representation learning: A review and new perspectives, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35, 1798–1828, https://doi.org/10.1109/TPAMI.2013.50, 2013.

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M.: YOLOv4: Optimal Speed and Accuracy of Object Detection, https://doi.org/10.48550/ARXIV.2004.10934, 2020.

Collins, B. D., Montgomery, D. R., Fetherston, K. L., and Abbe, T. B.: The floodplain large-wood cycle hypothesis: A mechanism for the physical and biotic structuring of temperate forested alluvial valleys in the North Pacific coastal ecoregion, Geomorphology, 139-140, 460–470, https://doi.org/10.1016/j.geomorph.2011.11.011, 2012.

Curran, J. H. and Wohl, E.: Large woody debris and flow resistance in step-pool channels, Cascade Range, Washington, Geomorphology, 51, 141–157, https://doi.org/10.1016/S0169-555X(02)00333-1, 2003.

Diehl, T.: Potential Drift Accumulation at Bridges, 1997.

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., and Tian, Q.: CenterNet: Keypoint triplets for object detection, Proceedings of the IEEE International Conference on Computer Vision, 2019-Octob, 6568–6577, https://doi.org/10.1109/ICCV.2019.00667, 2019.

Ghaffarian, H., Piégay, H., Lopez, D., Rivière, N., MacVicar, B., Antonio, A., and Mignot, E.: Video-monitoring of wood discharge: first inter-basin comparison and recommendations to install video cameras, Earth Surface Processes and Landforms, 45, 2219–2234, https://doi.org/10.1002/esp.4875, 2020.

Ghaffarian, H., Lemaire, P., Zhi, Z., Tougne, L., MacVicar, B., and Piégay, H.: Automated quantification of floating wood pieces in rivers from video monitoring: a new software tool and validation, Earth Surface Dynamics, 9, 519–537, https://doi.org/10.5194/esurf-2020-96, 2021.

Haschenburger, J. K. and Rice, S. P.: Changes in woody debris and bed material texture in a gravel-bed channel, Geomorphology, 60, 241–267, https://doi.org/10.1016/j.geomorph.2003.08.003, 2004.

Hassan, M. A., Hogan, D. L., Bird, S. A., May, C. L., Gomi, T., and Campbell, D.: Spatial and temporal dynamics of wood in headwater streams of the pacific northwest, Journal of the American Water Resources Association, 41, 899–919, https://doi.org/10.1111/j.1752-1688.2005.tb04469.x, 2005.

Kaur, R. and Singh, S.: A comprehensive review of object detection with deep learning, Digital Signal Processing, 132, 2023.

Keller, E. A., MacDonald, A., Tally, T., and Merrit, N. J.: Effects of large organic debris on channel morphology and sediment storage in selected tributaries of Redwood Creek, northwestern California, US Geological Survey Professional Paper, 1454, 1–29, 1995.

Lassettre, N. S. and Kondolf, G. M.: Large woody debris in urban stream channels: Redefining the problem, River Research and Applications, 28, 1477–1487, https://doi.org/10.1002/rra.1538, 2012.

Lassettre, N. S., Piegay, H., Dufour, S., and Rollet, A.: Decadal changes in distribution and frequency of wood in a free meandering river, the Ain River, France, Earth Surface Processes and Landforms, 33, 1098–1112, https://doi.org/10.1002/esp.1605, 2008.

Le Coz, J., Patalano, A., Collins, D., Guillén, N. F., García, C. M., Smart, G. M., Bind, J., Chiaverini, A., Le Boursicaud, R., Dramais, G., and Braud, I.: Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand, Journal of Hydrology, 541, 766–777, https://doi.org/10.1016/j.jhydrol.2016.07.036, 2016.

Lecun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, https://doi.org/10.1038/nature14539, 2015.

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

Lemaire, P., Piegay, H., MacVicar, B., Vaudor, L., Mouquet-Noppe, C., and Tougne, L.: An automatic video monitoring system for the visual quantification of driftwood in large rivers, III Wood in World Rivers, pp. 134–136, 2015.

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollar, P.: Focal Loss for Dense Object Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 42, 318–327, https://doi.org/10.1109/TPAMI.2018.2858826, 2020.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C.: SSD: Single shot multibox detector, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9905 LNCS, 21–37, https://doi.org/10.1007/978-3-319-46448-0_2, 2016.

Lucía, A., Comiti, F., Borga, M., Cavalli, M., and Marchi, L.: Dynamics of large wood during a flash flood in two mountain catchments, Natural Hazards and Earth System Sciences, 15, 1741–1755, https://doi.org/10.5194/nhess-15-1741-2015, 2015.

Lyn, D., Cooper, T., and Yi, Y.-K.: Debris Accumulation at Bridge Crossings: Laboratory and Field Studies, Publication FHWA/IN/JTRP-2003/10. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, https://doi.org/10.5703/1288284313171, 2003.

MacVicar, B., Piegay, H., Henderson, A., Comiti, F., Oberlin, C., and Pecorari, E.: Quantifying the temporal dynamics of wood in large rivers: field trials of wood surveying, dating, tracking, and monitoring techniques, Earth Surface Processes and Landforms, 34, 2031–2046, https://doi.org/10.1002/esp.1888, 2009.

Platts, W. S., Armour, C., Booth, G. B., Bryant, M., Bufford, J. L., Cuplin, P., Jensen, S., Lienkaemper, G. W., Wayne Minshall, G., Monsen, S. B., Nelson, R. L., Sedell, J. R., and Tuhy, J. S.: Methods for evaluating riparian habitats with applications to management., General Technical Report - US Department of Agriculture, Forest Service, 1987.

Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J., and Rojas, I.: Neural networks: An overview of early research, current frameworks and new challenges, Neurocomputing, 214, 242–268, 2016.

Ren, S., He, K., Girshick, R., and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1137–1149, https://doi.org/10.1109/TPAMI.2016.2577031, 2017.

Ruiz-Villanueva, V., Mazzorana, B., Bladé, E., Bürkli, L., Iribarren-Anacona, P., Mao, L., Nakamura, F., Ravazzolo, D., Rickenmann, D., Sanz-Ramos, M., Stoffel, M., and Wohl, E.: Characterization of wood-laden flows in rivers, Earth Surface Processes and Landforms, 44, 1694–1709, https://doi.org/https://doi.org/10.1002/esp.4603, 2019.

Sanhueza, D., Iroumé, A., Ulloa, H., Picco, L., and Ruiz-Villanueva, V.: Measurement and quantification of fluvial wood deposits using UAVs and structure from motion in the Blanco River (Chile), in: Proc. ofthe 5th IAHREurope Congress —NewChallenges inHydraulic ResearchandEngineering, edited by Aronne Armanini and Nucci, E., pp. 561–562, https://doi.org/10.3850/978-981-11-2731-1_216-cd, 2018.

Sejr, J. H., Schneider-Kamp, P., and Ayoub, N.: Surrogate Object Detection Explainer (SODEx) with YOLOv4 and LIME, Machine Learning and Knowledge Extraction, 3, 662–671, https://doi.org/10.3390/make3030033, 2021.

Taskesen, E.: Python package clustimage is for unsupervised clustering of images., https://erdogant.github.io/clustimage, 2021.

van Lieshout, C., van Oeveren, K., van Emmerik, T., and Postma, E.: Automated River Plastic Monitoring Using Deep Learning and Cameras, Earth and Space Science, 7, https://doi.org/10.1029/2019EA000960, 2020.

Viso.ai: Viso Suite: The One No Code Computer Vision Platform, 2022.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022.

https://doi.org/10.5194/egusphere-2024-792
Preprint. Discussion started: 16 April 2024

Wohl, E.: A legacy of absence: Wood removal in US rivers, Progress in Physical Geography, 38, 637–663, https://doi.org/10.1177/0309133314548091, 2014.

Wohl, E., Lininger, K. B., Fox, M., Baillie, B. R., and Erskine, W. D.: Instream large wood loads across bioclimatic regions, Forest Ecology and Management, 404, 370–380, https://doi.org/10.1016/j.foreco.2017.09.013, 2017.

Wohl, E., Scott, D. N., and Lininger, K. B.: Spatial Distribution of Channel and Floodplain Large Wood in Forested River Corridors of the Northern Rockies, Water Resources Research, 54, 7879–7892, https://doi.org/10.1029/2018WR022750, 2018.

Xu, Y. and Goodacre, R.: On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning, Journal of Analysis and Testing, 2, 249–262, 2018.

Zhang, Z., Ghaffarian, H., Macvicar, B., Vaudor, L., Antonio, A., Michel, K., and Piégay, H.: Video monitoring of in-channel wood: From flux characterization and prediction to recommendations to equip stations, Earth Surface Processes and Landforms, 46, 822–836, https://doi.org/10.1002/esp.5068, 2021.

Zhao, Z. Q., Zheng, P., Xu, S. T., and Wu, X.: Object Detection with Deep Learning: A Review, IEEE Transactions on Neural Networks and Learning Systems, 30, 3212–3232, https://doi.org/10.1109/TNNLS.2018.2876865, 2019.

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., and Ren, D.: Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression, AAAI, 34, 2020.