Longuetaud *et al. Annals of Forest Science* (2025) 82:5
https://doi.org/10.1186/s13595-025-01276-9

Annals of Forest Science

DATA PAPER Open Access

Traceability of oak (*Quercus petraea* (Matt.) Liebl. and *Quercus robur* L.) logs: the Biomtrace database

Fleur Longuetaud[1*], Frédéric Mothe[1], Dorian Martinetto[1,2], Phuc Ngo[2], Alexandre Piboule[3], Daniel Rittié[1], Frédéric Bordat[1], Philippe Jacquin[1], Isabelle Debled-Rennesson[2], Aymeric Albert[4] and Claudine Richter[4]

Abstract

Key message The Biomtrace database contains 33,390 RGB images of the butt end cross-section of 5135 French oak logs. Each log was photographed several times with different camera orientations during an initial shooting session. For more than half of the logs, additional photos were taken at least 3 weeks after the first photo session. Cross-sections were segmented on all the images using the PointRend convolutional neural network. Spatial calibration was carried out by a specific algorithm using a checkerboard pattern present in each image. This image database was created with the aim of developing algorithms for the biometric traceability of logs, involving artificial intelligence approaches that require large databases. But other applications are also possible, such as the automatic extraction of information on the size and quality of logs. The Biomtrace database is available at https://doi.org/10.57745/9DBCL4, and associated metadata are available at https://metadata-afs.nancy.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/feda0a0e-041a-4190-9a73-5159b10ff0f0.

Keywords Image analysis, RGB images, Oak, Segmentation, Biometric traceability, Tree log identification

1 Background

Today, establishing effective wood traceability from forest to industry has become a priority. Traceability should make it possible to combat illegal logging and illegal timber export, provide consumers with information about the local origin of products and optimise the flow of raw material in the supply chain, as well as processing, particularly when the wood is also characterised in terms of quality.

In France, there is a particular issue with the oak resource, which is subject to illegal export. Stricter controls are therefore needed for this resource. In addition, the new European regulation against deforestation and forest degradation (EUDR[1]) is due to come into force in 2025 and will require all wood-based products to be traceable. More generally, the issue of wood traceability has been the focus of increasing attention and investment around the world in recent years (e.g., Fig. 2 in Tonouéwa et al. (2024) and Fig. 2 in Elias (2024)). This is reflected, for example, in numerous recent publications reviewing

Handling editor: Irène Gabriel.

This article is part of the topical collection on ARBRE: Documenting rapid changes occurring in forest socio-ecosystems.

*Correspondence:
Fleur Longuetaud
fleur.longuetaud@inrae.fr
[1] Université de Lorraine, AgroParisTech, INRAE, Silva, 54000 Nancy, France
[2] Université de Lorraine, CNRS, LORIA, F-54000 Nancy, France
[3] Office National des Forêts, Pôle Recherche, Développement et Innovation, F-54600 Villers-lès-Nancy, France
[4] Office National des Forêts, F-94704 Maisons-Alfort, France

[1] https://green-business.ec.europa.eu/deforestation-regulation-implementation_en

existing methods (among others, Shirmohammadi, 2022; Kaulen et al., 2023; Elias, 2024).

Physical markers such as plastic tags or RFID chips, or the marking of logs with paint, are unreliable and easily falsified. It is therefore necessary to develop other methods that are technologically and economically realistic (Knowles et al., 2017; Kaulen et al., 2023).

One promising approach seems to be the biometric traceability of logs based on the analysis of photos of their cross-section (CS), i.e. traceability without labelling, based on the intrinsic characteristics of the objects studied and on the assumption that they are unique in their characteristics in the same way than fingerprints for humans. A recent literature review on traceability methods in the forestry sector concludes that biometric traceability approaches, without physical markers, are the most promising and that research should continue in this direction (Kaulen et al., 2023).

In this article, we present the Biomtrace database created to develop and test methods for the biometric traceability of logs. This database mainly contains RGB images of *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. untreated log ends (i.e., without any further preparation like sanding and polishing). A large proportion of the logs were photographed on two dates about a month apart to take account of changes in the appearance of CSs and their environment as a result of time spent in the forest or on a log yard.

The first study on this subject was carried out by Barrett (2008), whose aim was to link photos of stumps with photos of the corresponding CSs of logs in order to combat illegal logging. The method was based on pseudo-Zernike polynomial moments to provide an orientation-invariant comparison algorithm. The final matching was done manually.

More recently, approaches based on texture feature analysis for fingerprints or iris recognition have been applied by Schraml et al. (2014, 2015a, 2015, 2016). In these studies on the recognition of spruce logs, the texture linked to the presence of clearly visible annual growth rings is essential. In CSs of oak logs, however, the rings are not visible or are only very slightly visible, therefore other methods have to be developed. Schraml et al. (2016) highlighted the sensitivity of the algorithms to log rotation and compared three methods for rotational prealignment of logs to improve matching.

More recently, approaches based on convolutional neural networks (CNN) appeared and Wimmer et al. (2021) for spruce and Holmström et al. (2023) for pine showed their superiority compared with more traditional approaches based on image-processing techniques.

However, these approaches based on artificial intelligence (AI) require large image databases to train the algorithms. The Biomtrace database is much larger than any of the databases used in the publications cited above and concerns oak, a group of species that has not been dealt with to date in terms of traceability, with a cross-sectional appearance that is very different from that of softwood species, in particular with annual rings that are only slightly visible.

The database also contains manual reference measurements to validate the pre-processing algorithms used to segment the CSs and measure their size using a calibration based on a black and white checkerboard placed on the log when the image was captured.

The data was collected as part of a research project also named Biomtrace.

2 Methods

2.1 Image collection

A photo campaign, carried out between March 2023 and March 2024, produced 33,390 images of CSs of the butt end of oak logs from 5135 different logs collected in 62 sites (Fig. 1).

Most of the logs belonged to Office National des Forêts (ONF), which manages French public forests. The logs were grouped together on logging sites. From one site to another, the time elapsed since the trees were felled could vary. Logging sites were usually identified by an ONF identifier beginning with the letters *FB* followed by 9 digits (e.g., *FB220018133*). For 21 sites, the ONF identifier was unknown and was replaced by *XX* followed by a number. ONF identifiers could be used to make the link with log size information recorded in the field by ONF employees. It would then be possible to compare the results of the algorithms, in particular for the segmentation of log CSs in the images (Sect. 2.5.2) and the automatic measurement of the black and white checkerboard pattern (Sect. 2.4.2), with these manual measurements in the field in order to check their consistency. For each site, the file `Biomtrace/tables/database_description.txt` gives the name of the forest in which it is located and its GPS coordinates.

For 25 sites with a total of 2284 logs, it was only possible to take photos on one single date d_1. This number is calculated as the total number of logs (5135) minus the number of logs with at least two dates for taking photos (2851). The aim here was to build up a large database of images in which to search for matches when developing traceability algorithms. The larger the database, the more powerful the algorithms will need to be in order to identify a photo of a given log taken at an earlier date. For

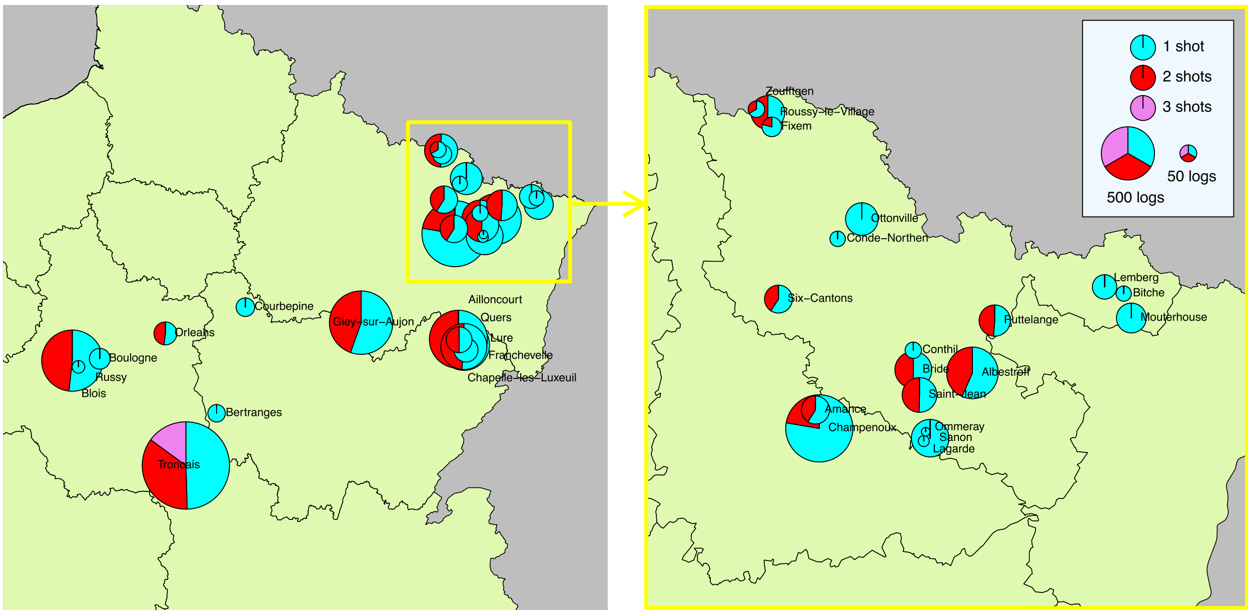

Fig. 1 Location of the logging sites. The surface area of the discs is proportional to the number of logs photographed

the other 37 sites representing 2851 logs, we were able to carry out at least two photo sessions on two dates (d_1 and d_2 or d_1, d_2 and d_3) several weeks apart, to take account of changes in the appearance of the CSs over time. The time between the first shot of a log and the second shot varied from 25 to 63 days. For five sites, all in Tronçais forest, a third session was held, 14 to 17 days after the second shot. The number of photos and logs at shot #1 (first shots), shot #2 (second shots) and shot #3 (third shots) are presented in Table 1. Figure 2 illustrates the changes in appearance of the CSs and their surroundings between successive shots. The difference between the dates d_1, d_2 and d_3 for visiting logging sites and the sessions for taking photos of a log (shot #1, shot #2 and shot #3) is that it was possible to visit a site on a date d_2 (already visited on a date d_1) but have new logs present that had not already been photographed on d_1. In this case, for such a log, the date on which the first photos were taken was d_2, but these photos belonged to the shot #1 dataset.

The 33390 photos are available on the data repository in Biomtrace/color_images (Longuetaud et al, 2024).

Table 1 Number of photos and logs in the respective photo campaigns (shot #1, shot #2 and shot #3)

	Shot #1	Shot #2	Shot #3	Total
Number of photos	21150	11404	836	33390
Number of logs	5135	2851	209	5135

2.2 Protocol for taking photos of a log

The photos were all taken with smartphones. The size and quality of the images depended on the smartphone used and the parameters set on it. The type of smartphone used depended on the operators in the field. Most of the images (90.1%) were 4032 × 3024 pixels. In the photo-taking protocol, a resolution greater than 12 million pixels was required to ensure minimum quality. This instruction was not respected for 1.6% of the images (2560 × 1440 for only 25 photos and 4080 × 2296 for 500 photos) which were nevertheless kept in the database.

Before the photos were taken, a black and white checkerboard pattern was positioned on the log (Fig. 3) so that the scale of the image could be recovered later (Sect. 2.4).

We asked the operators to frame the CSs as best they could. With a few exceptions, the centre of the image was located within the CS of interest.

When the photos were taken during a log measurement and identification operation by ONF-staff with the application of a numbered plastic tag, four photos were taken by turning the camera before affixing the tag as explained in Fig. 4, then a final photo was taken with the tag so that an identifier could be assigned to all photos of a given log in the database.

When the photos were taken after the log had been marked with a plastic tag, only four photos were taken (Fig. 4), all with the tag.

Images were saved in JPG format.

It should be noted that, depending on the camera, the saved image may include information about the

Fig. 2 Successive shots of five logs. Site identifier is shown on the bottom line

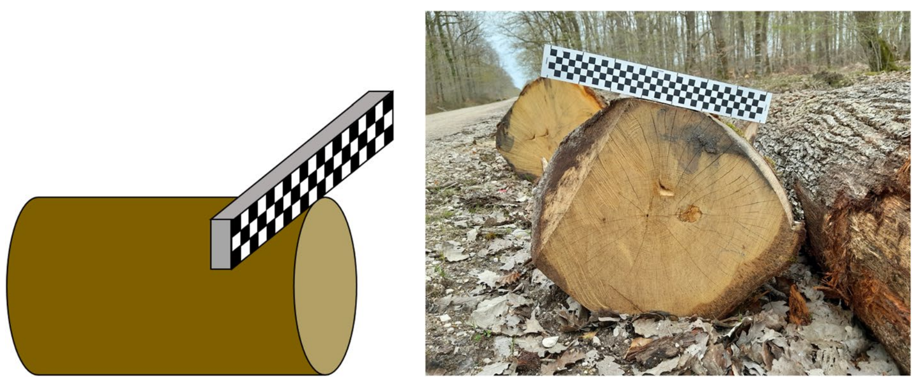

Fig. 3 Positioning of the black and white checkerboard pattern

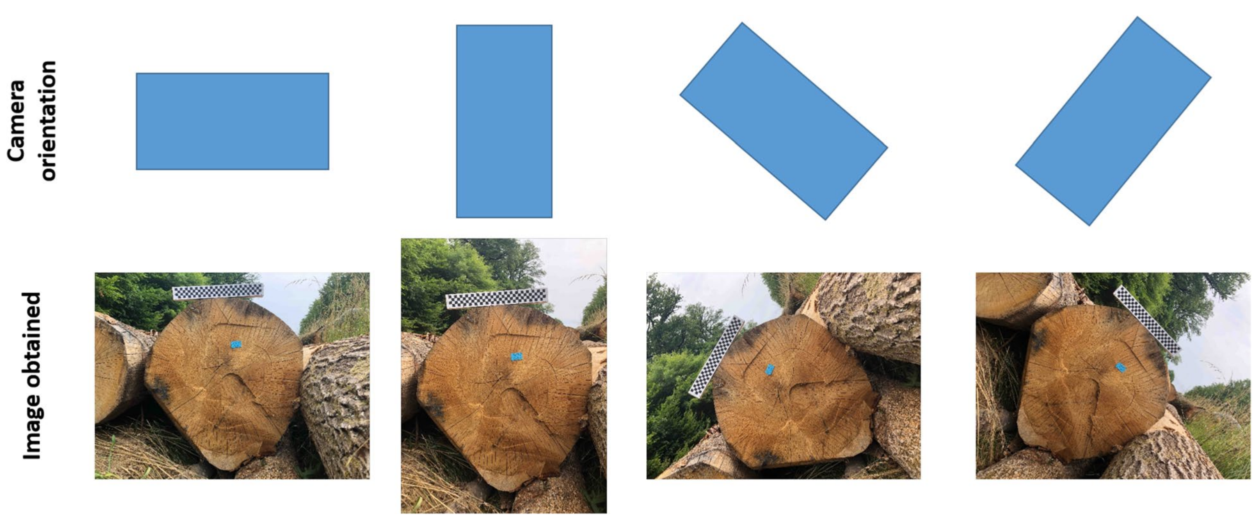

Fig. 4 Protocol of taking four photos per log by turning the camera. The order of the photos depends on the operator

rotation of an angle (90, 180 or 270 degrees) depending on the position of the camera when the photo was taken. This rotation information is recorded in the Exif metadata.

2.3 Checking and renaming images

For each logging site, the photos were checked by an operator to exclude some (e.g., blurred or cut images, log without tag) and to record the ONF log numbers written on the plastic tags.

The image files were then renamed using the log number prefixed by the site identifier and followed by the "photo sequence number" using an R script (R Core Team, 2023) developed to automate the process. The "photo sequence number" was a simple digit from 1 to 4 (or 5) for the first photo session of each log or *R1* to *R4* for photos from the second session, taken a few weeks later, or *RR1* to *RR4* for photos from the third session. For instance, *FB220038073_02256_2.jpg* refers to the second image at shot #1 of log *02256* taken at the logging site *FB220038073* and *FB220056383_07079_R4.jpg* refers to the fourth image of the second photo session (shot #2) of log *07079* at the logging site *FB220056383*.

2.4 Image scale: get the pixel size

The diameter of the logs is an important identification criterion. In order to retrieve pixel size information from the images, a black and white checkerboard pattern was positioned on the log to be photographed (except on one of the 33,390 images, namely *FB220022047_20137_5.jpg*). In this project, a grid of 35 × 4 black and white squares of each 1 × 1 cm was used. It was positioned so that its printed side (checkerboard) was aligned as far as possible with the plane of the CS to be photographed (Fig. 3).

It is likely that in the near future the use of a physical checkerboard will no longer be necessary and that all smartphones will be equipped with distance sensors that will enable the scale to be obtained automatically during acquisition.

2.4.1 Manual measurements

The image scale (i.e., pixel size in mm in the plane of the wood CS) was measured manually on a subset of 1129 images in order to build and validate software for automatic scale measurement on all images (Sect. 2.4.2). Pixel size was also measured manually on two additional images for which the algorithms had failed, bringing the total to 1131 images measured manually. The ImageJ software (Schneider et al., 2012) was used to calculate the surface area in pixels of the 35 × 4 squares of the checkerboard by drawing a quadrilateral passing through the four corners (ignoring distortions). When the checkerboard was not fully visible, the measurement was taken on a rectangle shorter in length but still four squares wide. The width of the pixels in mm was calculated as the square root of the measured area in mm^2 divided by the corresponding area in number of pixels. The pixel width was also estimated manually for image *FB220022047_20137_5.jpg* on which the checkerboard had been forgotten by comparing the number of pixels of the CS and the average surface area of the CS in mm^2 obtained from the other images of the same log for which the checkerboard was present. The results are provided in the Biomtrace/table/database_images.txt (Longuetaud et al, 2024) table (`pixelSize_manual_mm`).

2.4.2 Automatic measurements

Specific algorithms were developed to detect the checkerboard and calculate the size of the pixels on each image. The source code of the C++ implementations of two software packages (*detec_mire* and *pixel_grid*) using the OpenCV library (Bradski, 2000) is given in the `Biomtrace/tools` directory. The results are given in the table `Biomtrace/table/database_images.txt` providing the pixel size

Fig. 5 The two images (*FB230003243_04079_R1.jpg* and *XX20_02679_R3.jpg*) for which the algorithms failed to detect the black and white checkerboard pattern

Table 2 Comparison of the two algorithms for automatic measurement of the black and white checkerboard pattern. RMSE and bias (and their relative values) are calculated on 1129 manually measured images including the checkerboard; the number of fails are calculated on the 33,389 images including the checkerboard

Program	Pixel width (mm)		Number of pixels/cm		Total number of fails
	RMSE	Bias	RMSE	Bias	
detec_mire	0.0088 *(3.5%)*	0.0026 *(1.0%)*	0.6442 *(1.5%)*	− 0.4380 *(− 1.0%)*	2
pixel_grid	0.0044 *(1.7%)*	0.0008 *(0.3%)*	0.4741 *(1.1%)*	− 0.1201 *(− 0.3%)*	2

of each photo (`pixelSize_detecmire_mm` and `pixelSize_pixelgrid_mm`).

Both software failed to detect the checkerboard in just two images (Fig. 5). In the first image (on the left), the photo is very blurred. In the second image (on the right), the checkerboard is very small in the image with relatively few pixels for each square. As shown in Table 2, *pixel_grid* proved to offer the best accuracy with a root mean square error (RMSE) on pixel size of less than 5 microns. As it stands, we recommend using *pixel_grid* for automatic calibration.

2.5 Segmenting the cross-sections

A necessary pre-processing step before applying traceability algorithms is the segmentation of the wood cross-sections S in the images.

2.5.1 Manual segmentation

The CSs were manually segmented on a subset of 1311 images using the image edition software Gimp (https://www.gimp.org/). It is important to note that the images were opened in Gimp without applying any rotation given by the Exif orientation metadata (see Sect. 2.2). The underbark contour was drawn using the free select tool, then filled in with white. After reverting the selection, the outside of the section (including bark) was filled in black.

We preferred to draw the underbark contour of the CSs because the bark can be damaged during logging or during transport of the logs. In addition, for large logs the bark can be very irregular and therefore difficult to measure accurately. This should not be a problem for the traceability algorithms, which will be rather based on the singularities visible on the area of wood.

In addition, on the oak log CSs, notches related to the felling of the trees, oriented in a plane slightly different from that of the main part of the CSs, were sometimes visible at the periphery of the CSs. Depending on the angle from which they were photographed, their surface could be very different from one photo to another. We therefore had to decide whether or not to include them in the area of interest. Here, we chose to include them systematically, as shown in Fig. 6.

Fig. 6 Rule for taking account of harvesting-related facets when manually delimiting cross-sections. Manual segmentations of the cross-sections are highlighted in yellow

The binary masks for these 1311 images are saved in the `Biomtrace/segmentation_masks/manual/manual_1` directory.

After testing two CNNs (PointRend and YOLOv9) for the automatic segmentation of our 33390 images, we selected the results that diverged the most between the two CNNs (on a segmented surface criterion) to add some manual segmentation on those cases considered difficult. The manual segmentations of these 173 additional images can be found in the directory `Biomtrace/segmentation_masks/manual/manual_2`.

2.5.2 Automatic segmentation

The segmentation task was performed by the CNN PointRend (Kirillov et al., 2020). An illustration showing the comparative results of different CNNs on an example is given in Fig. 7. In this Figure, PointRend gives the best results. The PointRend implementation used is from the platform Detectron2 (https://ai.meta.com/tools/detectron2/). The network was pre-trained using the MS COCO dataset (Lin et al., 2014), then fine-tuned using the manual segmentations (Sect. 2.5.1).

Of the 1311 manual segmentations (Sect. 2.5.1), 116 were set aside for independent evaluation on a test set not

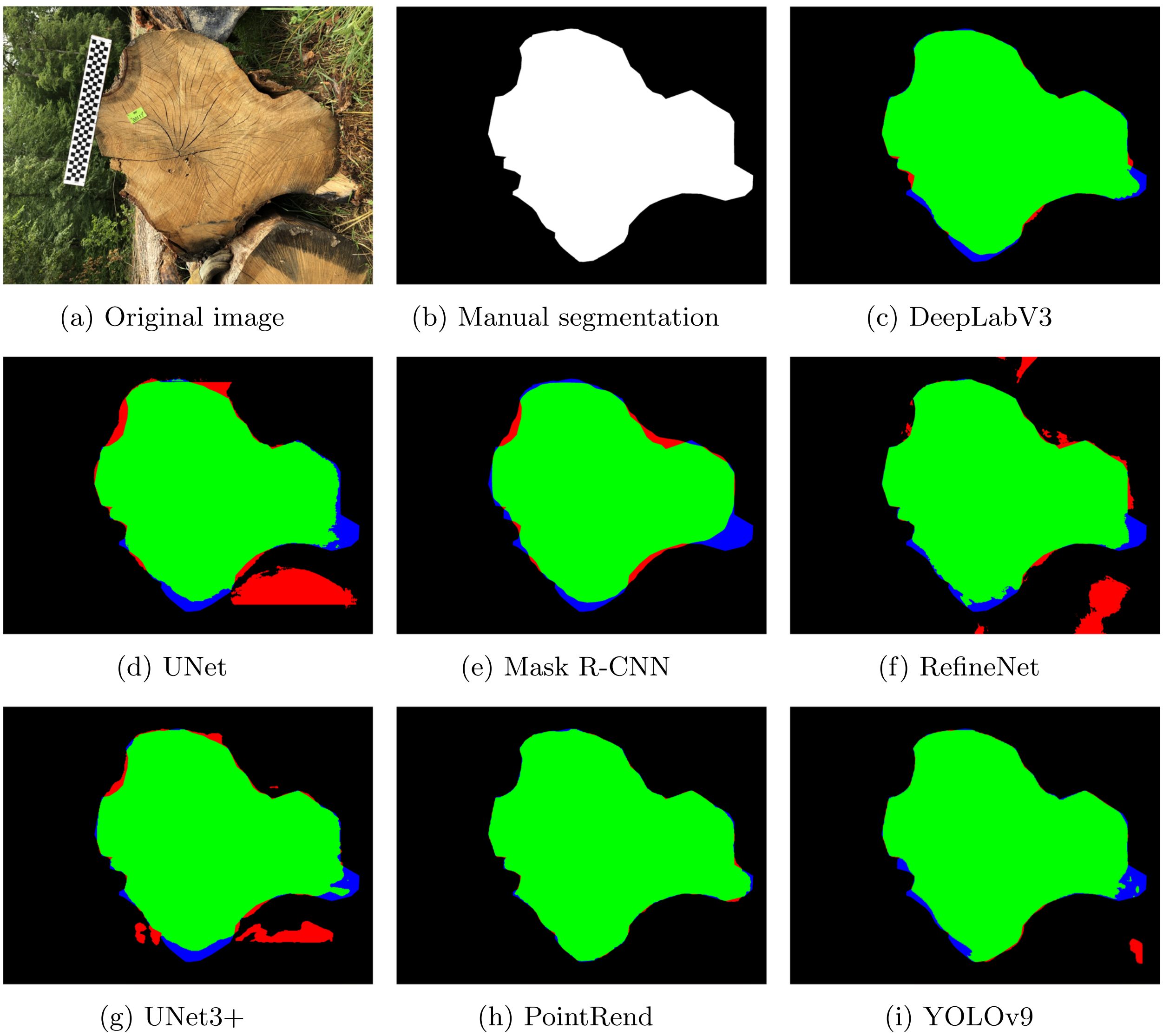

Fig. 7 Segmentation of the cross-section in image *FB210033929_00117_1.jpg* with different CNNs. In green: true positives (TP) which are pixels of the cross-section correctly classified; In black: true negatives (TN) which are the correctly classified background pixels; In red: false positives (FP) which are the background pixels attributed to the cross-section; In blue: False Negatives (FN) which are the cross-section pixels attributed to the background

used for training. The list of these 116 images is provided in the `Biomtrace/tables/segmentation_test_set.txt` table to enable subsequent comparisons on the same dataset. The accuracy (i.e., rate of correctly classified pixels) for the CNN PointRend was 99.1%.

CNN PointRend was then applied to the 33390 images in the database. The resulting masks are given in the `Biomtrace/segmentation_masks/automatic` directory in PNG format. Then, measurements of the size and shape of these masks were obtained automatically by using the *Set Measurements* and *Analyze Particles* functions of the ImageJ software in batch mode. The surface area of the segmented CSs, the mean diameter and several shape descriptors are given in the `Biomtrace/table/database_images.txt` table (*`segm_area_pixel2`* and following variables prefixed by *`segm_`*). The definitions and units of the different variables are given in the `Biomtrace/table/database_images_variables.txt` table.

In the absence of manual segmentation on the large majority of these images, several criteria were used to detect the most obvious errors, among them: The variation in the surface area of the CS measured in mm^2 between the different photos of the same log taken during the same session and the distance between the centre of the image and the centre of gravity of the segmented area. In five cases out of 33,390 images (*FB220034467_00573_3.png, FB220046062_00857_1.png, FB220046062_00857_5.png, FB220046062_00916_3.png* and *FB220062509_00829_2.png*), the wrong log was segmented (several CSs were visible on these images), and in this case the original image was modified (manual deletion of the wrong log) in order to obtain a correct segmentation mask for these five images. Although this happened relatively rarely and was not a problem for traceability, several other errors could occurred, such as holes in the masks, segmented CS halves in the case of a log split in two through the centre and pieces of a neighbouring CS taken in the segmented area. Some examples are given in Fig. 8. Automatic image segmentation could be further improved by using more recent deep learning methods such as the Vision Transformers (ViT) instead of CNNs, as was done in Nack et al. (2024), and by improving post-processing.

3 Access to the data and metadata description

The database (Longuetaud et al., 2024) is available at Recherche Data Gouv repository: https://doi.org/10.57745/9DBCL4. Associated metadata access is at https://metadata-afs.nancy.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/feda0a0e-041a-4190-9a73-5159b10ff0f0.

The arborescence of the Biomtrace directory was detailed in the previous section and is summarized in Fig. 9.

The `Biomtrace/tables` directory includes two main tables, both in the form of an array with variables in columns and a header line indicating the variable names. Each data table is accompanied with a variable table (with the same name followed by `_variable`) describing each variable.

- The site table `database_sites.txt` (62 lines × 24 variables) provides the name and location of each logging site. The dates of the photo sessions and the number of logs and images taken during each session are also recorded.

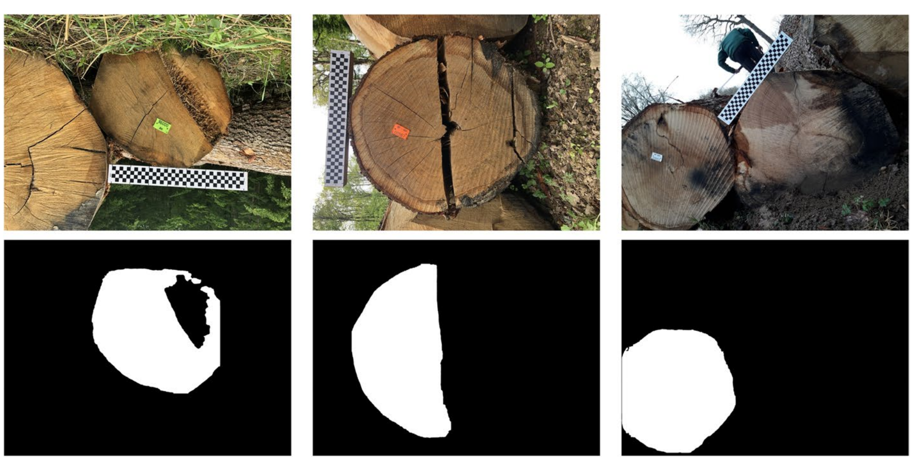

Fig. 8 Examples of automatic segmentations for which errors were detected: A hole in the area to be segmented (on the left; Image *FB210033929_00096_2.jpg*); Only one half of the cross-section segmented for a log split in two (in the centre; Image *FB220034467_00567_R1.jpg*); An adjacent cross-section segmented instead of the main cross-section (on the right; Image *FB220034467_00573_3.jpg*)

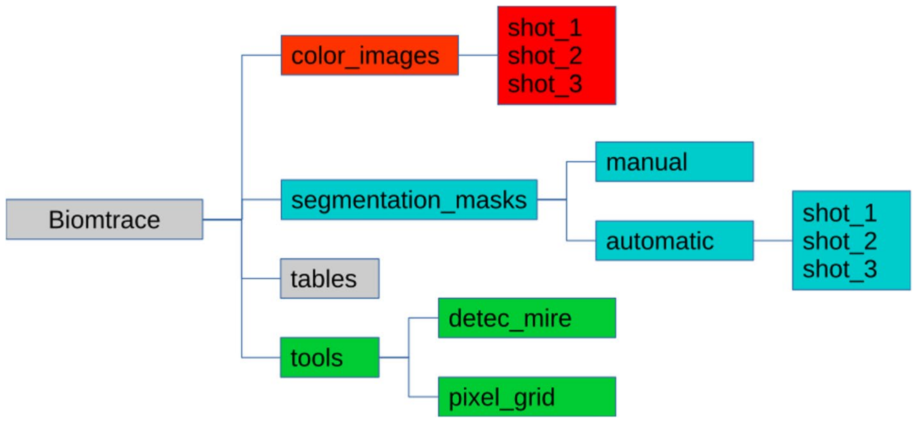

Fig. 9 Organisation of the data files

- The image table `database_images.txt` (33390 lines × 20 variables) provides the name of each image, the log name, the main data recorded in the Exif metadata (width and height in pixels, rotation, shooting date, camera model), the size of the pixels in the CS plane as calculated by two software programs (Sect. 2.4.2) or measured manually for some images (Sect. 2.4.1) and geometric measurements of the CS (size and shape) obtained using ImageJ software on automatically generated segmentation masks (Sect. 2.5.2).

The `Biomtrace/tables` directory also includes four other files, each containing a list of images. These are lists of images from the test sets that have been used to evaluate segmentation and traceability algorithms:

- Segmentation: The `segmentation_test_set.txt` file provides the name of the 116 images used in the segmentation algorithm test set (Sects. 2.5.2 and 5).
- Traceability: There are three files listing images of the test sets used in the work by Martinetto et al. (2025) in experiments A and C (Sect. 5). The `traceability_test_set_Experiment_A_and_C_d1.txt` file provides the name of the 4085 images taken at d_1 used in the test sets of experiments A and C (same data set for the two experiments concerning photos taken at d_1). The `traceability_test_set_Experiment_A_d2.txt` file provides the name of the 800 images taken at d_2 used in the test set of experiment A. And `traceability_test_set_Experiment_C_d2.txt` file provides the name of the 7868 images taken at d_2 used in the test set of experiment C.

4 Technical validation

The log numbers have been carefully identified on the images by an operator to enable the files to be renamed correctly and ensure that each number (site identifier and ONF tag number) was assigned to all the photos of the same log, and that each number corresponded to a single log.

The manual measurements (checkerboard and CS segmentation) were carefully carried out by experienced operators.

The automatic measurements (checkerboard and CS segmentation) were obtained by software tools specially designed for this task and validated using the manual measurements. For the automatic measurement of the checkerboard pattern, a comparison of the outputs of the two software packages was carried out in order to detect any measurement errors in one or the other. Additional manual measurements of the checkerboard were carried out for images for which the two algorithms diverged the most from each other, in order to further improve the algorithms.

Overall, the results were carefully examined to ensure that no significant errors had occurred. Applying traceability algorithms to this dataset also enabled errors to be corrected, such as the segmentation of the wrong log when there were several logs on the image and none properly centred, as required in the photo-taking protocol.

5 Reuse potential and limits

All data tables are in plain text format with tabulations as separator. They can be imported easily with R or with any spreadsheet software. RGB images are in standard JPG format and binary segmentation masks are in PNG format.

The Biomtrace database is suitable for the development of image processing algorithms for untreated CSs of oak logs.

One of the main objectives of setting up this database, which was also the aim of the Biomtrace project, was to develop methods for the biometric traceability of logs from the forest to the sawmill. This database is made available to the scientific community with the aim of improving existing biometric traceability methods.

Automatic assessment of log volume and/or quality on the basis of CS images would also be an interesting objective as mentioned in the literature review carried out by Sandvik et al. (2024). For example, features such as eccentricity, rot, cracks, sapwood and growth rings can be observed in the images.

Researchers may also be interested in the degradation and changes in appearance of the logs between dates d_1 and d_2.

We give below some examples of past uses of the Biomtrace database.

Based on these images, two algorithms, which are variants of each other, for automatically detecting and measuring the black and white checkerboard pattern were developed and are described in a Master's report (Serratore and Suck, 2023), in an internal report (Mothe et al., 2024) and summarised in this paper. The reports and source code are available on the data repository in the `Biomtrace/tools` directory. These algorithms make it possible to obtain the scale of the images and therefore to measure the diameter or surface area of the CSs which can be an important criterion for traceability or for optimising processing times by eliminating, for example, candidate logs which would not be compatible from the point of view of their dimensions.

An algorithm for segmenting oak CSs in the images has also been developed using the PointRend CNN. A study comparing CNNs and more recent approaches such as Vision Transformers is currently underway, which could lead to even better segmentation results. Segmentation of CSs is a necessary pre-processing step before applying a biometric traceability or quality assessment algorithm, and this step influences the final results.

Then, an initial traceability algorithm based on deep learning was developed from a sub-part of this database (the entire database was not yet available at the time of this work) (Martinetto et al., 2025). This algorithm consisted of identifying features in the images and then searching for a correspondence between an image taken at a date d_2 and an image of the same log taken at a date d_1, a few weeks before d_2, belonging to a large database of recorded images.

With regard to the possible limitations of this database, we could mention the presence of a numbered plastic label on most of the CSs. This label could affect the detection of certain features or, conversely, make it easier to recognise logs. The work by Martinetto et al. (2025) has shown that by removing these tags with an algorithm, the biometric recognition of the logs was not affected.

It was also possible that part of the CS may not be clearly visible due to the environment around the log, or the presence of earth or vegetation, depending on the season and the location.

In this database, between the two dates on which the photos were taken, the appearance of the CSs changed, but very rarely were the logs moved between the two sessions. To make the situation even more realistic from the point of view of traceability, more photos of logs that have been moved should be added, as this changes the perspective from which the photos were taken.

Acknowledgements
Thanks to Alexandre Serratore and Hugo Suck for programming a first version of the image calibration software. Thanks also to Adriel Rajaonarivo and Anaelle Rajaonarivo for manually segmenting the images.

Authors' contributions
FL initiated and drafted the funding applications (FBF, LUE, Grand Est Region, ADEME), managed the project from a scientific point of view, took part in data collection and analysis, supervised the students and wrote the manuscript; FM was involved in developing the software, collecting and analysing the data and writing the manuscript; DM carried out his doctoral thesis on the basis of this database and played a major role in developing the algorithms and analysing the data as well as writing the manuscript; PN was involved in writing the funding applications, developing the algorithms, analysing the data, supervising the students and writing the manuscript; AP was involved in collecting and analysing the data and supervising the students; DR was involved in collecting the data; FB was involved in collecting the data; PJ participated in the development of algorithms; ID was involved in supervising the students; AA initiated the Biomtrace project and the search for funding and took part in data collection and analysis; CR initiated the Biomtrace project and the search for funding and took part in data collection and analysis. The authors read and approved the final manuscript.

Funding
This work was carried out as part of the Biomtrace project funded by France Bois Forêt (grant numbers: 23/RD/1698 and 24/RD/1868).
This work was also supported by "Lorraine Université d'Excellence", part of the France 2030 Program, reference ANR-15-IDEX-04-LUE, and by joint funding for a doctoral thesis (2024–2027) from Région Grand Est (grant number: 24-DOC-112) and ADEME (grant number: 24ESD0544).
SILVA laboratory is supported by a grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE).

Data availability
The images and the database are available at Research Data Gouv repository: https://doi.org/10.57745/9DBCL4. The licence of the database is Etalab Open License 2.0 compatible CC-BY 2.0. The metadata for the database are availble at: https://metadata-afs.nancy.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/feda0a0e-041a-4190-9a73-5159b10ff0f0.

Code availability
The source code generated as part of this work is available in the database repository or in the publications cited in the article.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors gave their informed consent to this publication and its content.

Competing interests
The authors declare that they have no competing interests.

Received: 18 July 2024 Accepted: 15 January 2025
Published online: 17 February 2025

References
Barrett W (2008) Biometrics of cut tree faces. In: Advances in computer and information sciences and engineering. Springer Netherlands, Dordrecht, p 562–565. https://doi.org/10.1007/978-1-4020-8741-7_100
Bradski G (2000) The OpenCV Library. Dr Dobb's J Softw Tools 25(11):120–123
Elias M (2024) Timber traceability and sustainable transportation management: a review of technologies and procedures. Bulletin of the Transilvania University of Brasov Series II, Forestry Wood Industry Agricultural Food Engineering, pp 11–52
Holmström E, Raatevaara A, Pohjankukka J et al (2023) Tree log identification using convolutional neural networks. Smart Agric Technol 4:100201. https://doi.org/10.1016/j.atech.2023.100201
Kaulen A, Stopfer L, Lippert K et al (2023) Systematics of forestry technology for tracing the timber supply chain. Forests 14(9):1718. https://doi.org/10.3390/f14091718
Kirillov A, Wu Y, He K, Girshick R (2020) Pointrend: image segmentation as rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), p 9799–9808, June 2020. https://doi.org/10.48550/arXiv.1912.08193
Knowles C, Boston K, Berecibar E (2017) A new method for tagging and tracking logs. Int For Rev 19(3):294–305. https://doi.org/10.1505/146554817821865036
Lin TY, Maire M, Belongie S, et al (2014) Microsoft coco: common objects in context. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer, pp 740–755
Longuetaud F, Mothe F, Martinetto D, et al (2024) "Biomtrace". [Dataset]. V1. Recherche Data Gouv. https://doi.org/10.57745/9DBCL4
Martinetto D, Wimmer G, Ngo P et al (2025) A new approach to biometric wood log traceability combining traditional methods and deep learning. Smart Agric Technol 10:100686. https://doi.org/10.1016/j.atech.2024.100686
Mothe F, Jacquin P, Ngo P, et al (2024) Calibration spatiale d'une image à l'aide d'une mire à damier noir et blanc. Tech. rep., INRAE. https://doi.org/10.57745/II8ZY0
Nack F, Stemmer MR, Stivanello ME (2024) Comparison of modern deep neural networks architectures for cross-section segmentation in images of log ends. IEEE Lat Am Trans 22(4):286–293
R Core Team (2023) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/
Sandvik YJ, Futsæther CM, Liland KH et al (2024) A comparative literature review of machine learning and image processing techniques used for scaling and grading of wood logs. Forests 15(7):1243
Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9(7):671–675
Schraml R, Charwat-Pessler J, Petutschnigg A et al (2015) Towards the applicability of biometric wood log traceability using digital log end images. Comput Electron Agric 119:112–122. https://doi.org/10.1016/j.compag.2015.10.003
Schraml R, Hofbauer H, Petutschnigg A et al (2016) On rotational pre-alignment for tree log identification using methods inspired by fingerprint and iris recognition. Mach Vis Appl 27:1289–1298. https://doi.org/10.1007/s00138-016-0814-2
Schraml R, Charwat-Pessler J, Uhl A (2014) Temporal and longitudinal variances in wood log cross-section image analysis. In: 2014 IEEE International Conference on Image Processing (ICIP). IEEE ICIP, p 5706–5710. https://doi.org/10.1109/ICIP.2014.7026154
Schraml R, Hofbauer H, Petutschnigg A, Uhl A (2015b) Tree log identification based on digital cross-section images of log ends using fingerprint and iris recognition methods. In: Azzopardi G, Petkov N (eds) Computer Analysis of Images and Patterns. Springer International Publishing, Cham, p 752–765. https://doi.org/10.1007/978-3-319-23192-1_63
Serratore A, Suck H (2023) Calibration automatique basée sur la détection du damier. Master's thesis, M1 informatique, Université de Lorraine, Nancy. https://hal.univ-lorraine.fr/hal-03603108
Shirmohammadi M (2022) A review of traceability systems in the timber industry. IntechOpen
Tonouéwa JFMF, Biaou SSH, Assèdé ESP et al (2024) Timber traceability, determining effective methods to combat illegal logging in africa: a review. Trees For People 18:100709. https://doi.org/10.1016/j.tfp.2024.100709
Wimmer G, Schraml R, Hofbauer H, et al (2021) Two-stage CNN-based wood log recognition. In: Computational Science and Its Applications. ICCSA 2021. Lecture Notes in Computer Science, vol 12955, Springer, pp 115–125. https://doi.org/10.1007/978-3-030-87007-2_9

Publisher's Note