

# Secondary Publication



Chizhova, Maria; Pan, Junquan; Luhmann, Thomas; u. a.

## Towards automatic defects analyses for 3D structural monitoring of historic timber

Date of secondary publication: 06.06.2024

Version of Record (Published Version), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-956752

### Primary publication

Chizhova, Maria; Pan, Junquan; Luhmann, Thomas; u. a. (2024): „Towards automatic defects analyses for 3D structural monitoring of historic timber“. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Jg. XLVIII-2/W4-2024, pp. 103-110, Katlenburg-Lindau: Copernicus GmbH, doi: 10.5194/isprs-archives-xlviii-2-w4-2024-103-2024.

### Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>

## TOWARDS AUTOMATIC DEFECTS ANALYSES FOR 3D STRUCTURAL MONITORING OF HISTORIC TIMBER

M. Chizhova<sup>1</sup>, J. Pan<sup>1</sup>, T. Luhmann<sup>2</sup>, A. Karami<sup>3</sup>, F. Menna<sup>4</sup>, F. Remondino<sup>4</sup>, M. Hess<sup>1</sup>, Th. Eißing<sup>1</sup>

<sup>1</sup> University of Bamberg, Digital Technologies in Heritage Conservation, Centre of Heritage Conservation Studies and Technologies, Am Zwinger 4-6, 96047 Bamberg, Germany, (maria.chizhova, junquan.pan, mona.hess, thomas.eissing)@uni-bamberg.de

<sup>2</sup> IAPG, University of Applied Sciences Oldenburg, 26121 Ofener Str. 16, Oldenburg, Germany, luhmann@jade-hs.de

<sup>3</sup> CYCLOMEDIA, Van Voordenpark 1B, 5301 KP Zaltbommel, The Netherlands, akarami@cyclomedia.com

<sup>4</sup> 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy - Email: (fmenna, remondino)@fbk.eu

### Commission II

**KEY WORDS:** historic wooden timber, feature extraction, machine learning, 3D scanning, structure from motion, multi view photometric stereo, reflectance transformation imaging

#### ABSTRACT:

Stability of historic wooden constructions is changing with time and should be inspected appropriately for risk assessment and prevention. The stability or strength values of built-in historic timber are difficult or even impossible to be derived without invasive investigation, but this is particularly problematic for the monitoring of heritage objects. Luckily there are some visible timber surface features, like knots and cracks, which can act as individual evidence to estimate the wood strength as well as to adjust its grade class indicator. In the final project, we aim to compare different approaches for 3D digital documentation of historic wood timbers and focus on automatic knot detection using AI techniques. A first feasibility study reported here provides a scientific baseline for the development of an automated method to analyse historic timber stability using 3D surveying and recognised surface features. First results about texture and resolution properties are discussed here.

## 1. INTRODUCTION

### 1.1 Heritage timber monitoring during climate change

Climate change processes increasingly have massive impacts on available resources on earth. This means that the preservation of tangible cultural heritage is also facing new challenges. In this context, issues of energy balance and building efficiency are particularly relevant. Supporting the adaptive re-use of historic buildings in particular is one of the most important actions in the European Green Deal (European Green Deal, 2019).

Wood has always been a sustainable, resource- and energy-efficient material and is an organic and biological material, which is particularly affected by climate change impacts (e.g. pathogen spread, cracking). It is therefore important, to observe and understand the properties, condition and change of this material, in terms of heritage conservation as well as for future exploitation of wood as a building material.

### 1.2 Problem definition

The optimal monitoring of heritage timber constructions and buildings represents a complex system based on their documentation, digitisation, data modelling and analysis.

Timber stability is changing with time and should be inspected appropriately. Several problems have to be observed in this case. On the one hand, the stability or strength values of built-in historic timber are difficult or even impossible to be derived without invasive investigation (e.g. drilling), but this is particularly problematic for the monitoring of heritage objects. On the other hand, there are some visible surface features, like:

- knots (with radius as critical parameter for stability estimation),
- cracks (especially their position and passage relative to annual rings),
- fungal and mistletoe infection.

These can act as individual evidence to estimate the wood strength as well as to adjust its grade class indicator.

Since invasive methods are rather often inappropriate in heritage conservation, this project focuses on strength estimation based on visually detectable surface defects.

Automated optical inspection exists currently only in timber production, where new timber parts are recorded by a monitoring system and are additionally tested using invasive methods. For existing (built-in) timber structures, inspection is mainly carried out visually by assessors, which causes an increased amount of effort and time and does not include a truly objective analysis and documentation.

From the viewpoint of heritage conservation and dendrochronology, there are still numerous open questions and unsolved problems, amongst others:

- To date, there is a lack of standardised procedures for describing wood surfaces, e.g. through mathematical parameterisation of the 3D surfaces.
- There is a lack of systematic measurement series on different types of wood and surface types.
- The surface texture of wood (reflection properties, roughness) is excellently suited for optical 3D methods (photogrammetry, laser scanning, structured light projection recording); investigations on specific questions such as resolution, accuracy, illumination, colour calibration or realisation of reference systems are still pending.
- The potential of multi-dimensional optical and geometric feature extraction for automated evaluation of wood surfaces (e.g. by machine learning) has extensive perspectives to optimise monitoring of wood.

In general, 3D geometric information is still not sufficiently combined with existing dendrochronological methods and therefore this question offers a high research and exploitation potential.

### 1.3 State of the art

The strength estimation of built-in timber structures is a challenging task, especially for the investigation of cultural heritage objects, as invasive methods may only be used to a very limited extent in contrast to new timber. A successful investigation regarding the strength estimation of historic timber structures using visual sorting criteria was carried out in (Ebner, 2018). The visually detectable characteristics like knots and cracks have been defined as individual evidence for the estimation of strength value.

Automatic monitoring has existed in timber production for several decades. The pieces of wood are observed by a stationary optical measuring system that identifies and sorts defective wood. The focus here is on the detection of gross defects and visible surface features that are relevant for estimating their structural quality. The condition analysis of built-in timber is mainly carried out visually or manually, which is time-consuming and not objective enough. Conventional optical measuring methods can be used for automatic detection of visual feature surfaces.

Significant developments in imaging technologies over the last two decades have made substantial contributions to monitoring process. Various application examples from the cultural heritage and industrial sectors can be found in (Luhmann et al., 2023; Remondino & Stylianidis, 2016). Interesting in this context is a combination of photogrammetry with different illumination sources (Reflectance Transformation Imaging, Photometric Stereo; Karami et al., 2022a,b), which enables highly detailed object visualization with interactive virtual shading/illumination. This is particularly relevant for the investigation of fine surface defects (such as scratches or cracks).

Regarding automatic feature detection, various AI and machine learning algorithms are optimised for individual task modules. According to recent literature, methods exist that are specifically optimised for the automatic detection of

- branches (Ding et al. 2020, Yang et al. 2021),
- tree rings (overview of existing methods in Divya & Kaur, 2020),
- cracks (Liu et al., 2020).

Convolutional Neural Networks (CNN) (O'Shea & Nash, 2015) have emerged as a foundational technique in modern deep learning methodologies for image recognition tasks.

Their capability extends beyond recognizing static 2D images to processing dynamic 3D structures. Notably, many state-of-the-art models, including the real-time computer vision system 'You Only Look Once' (YOLO), are fundamentally grounded on CNN architectures. These models underscore the importance and versatility of CNNs in contemporary image analysis.

Several studies have employed the YOLO model for 2D images detection of features from wood surfaces. Examples include the detection of timber cracks (Liu et al., 2020) and wood knots (Han et al., 2023). The latest version of YOLO is YOLOv8, a continuation by Ultralytics, the developers of the prior YOLOv5 model (YOLOv8, 2023). This newest version exhibits enhanced capabilities in real-time segmentation, detection, and classification in comparison to its predecessors.

In general, there is no strategy or automated procedure that combines the individual detection methods with knowledge of structural and load-bearing capacity to enable reliable statements about the quality of existing and, in particular, old historical timber structures.

### 1.4 Motivation and goal

3D recording of existing buildings using optical measurement methods is an established process today. 3D object digitisation

results in spatial 2D and 3D data with local X,Y,Z coordinates, intensity and colour values as well as textures. This type of optical surface recording enables an as-built and true-to-deformation geometric modelling of a free-form object with integration of additional information (semantics). These factors can be documented using optical measurement methods and automatically analysed with AI/machine learning. In combination, this analysis allows objective and comprehensible conclusions about timber stability and grading class.

It is therefore important both, in terms of historic preservation to analyse and classify such surface defects, and for the future use of wood as a building material, to monitor and understand surface characteristics, stability and material changes.

In this ongoing project, we are working on a feasibility study for timber surfaces documentation with various 3D imaging technologies and previous results of knots detection from different sources (2D and 3D). Various imaging parameters like geometric precision, resolution, colour, texture, as well as advantages or disadvantages, like applicability at difficult recording conditions (e.g. roof structure) or detailed recording will be discussed. The feasibility study provides a scientific basis to develop a method for 3D structural monitoring and automatic stability analysis of built-in historic wooden timbers.

## 2. METHOD

In this feasibility study, we have applied different 3D imaging techniques for recording of historic timbers. In particular, we will focus on the use of AI techniques for inspection and diagnostic analysis using this data. In combination with CNN, diverse optical measuring methodologies offer invaluable insights beyond 2D data analysis. In our case, 2D images coming from digital cameras and derived from 3D recording techniques (orthophotos from point clouds, image patches from camera integrated in laserscanner) will be analysed for knotholes using the YOLOv8n model.

### 2.1 3D digitisation

The project implementation begins with the 3D digitisation of wood samples that show different machining marks and surface characteristics. Two recording scenarios were used for our investigations. In the first scenario, relatively small wood samples (between 15 cm and 1 m) have been recorded in the laboratory using different imaging techniques (Fig. 1):

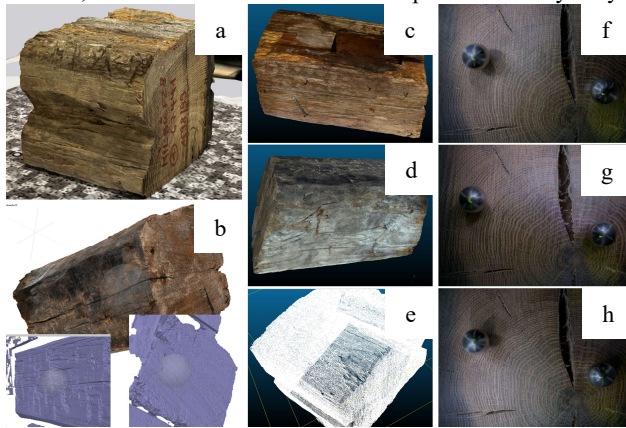
- Photogrammetric recording;
- Structured light scanning;
- Laser light section techniques;
- RTI (Reflectance Transformation Imaging);
- Multi-View Photometric Stereo (MVPS).

The pieces of wood could be captured from all sides as desired, and the recording settings (lighting conditions, object position) were the same for all objects and could be adjusted at any time.

In the second scenario, digitisation was carried out under real conditions, whereby an existing truss has been scanned. Due to accessibility problems on-site, our investigations were reduced to photogrammetry and structured light scanning, using hand-held scanners. Additionally, the whole roof construction has been scanned using terrestrial laser scanning (Faro Focus S 350).

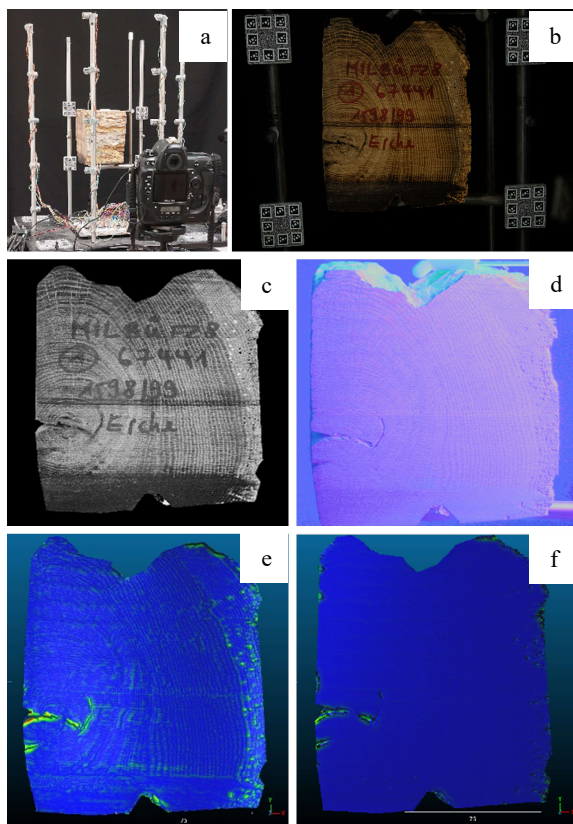
Photogrammetric imaging with the Nikon camera (50 mm lens) resulted in highly detailed textured 3D models with high detail reproduction in mm range. Nevertheless, fine geometries in sub-mm range (scratches, annual rings) were not captured geometrically, even though they are faithfully mapped and visually recognisable in the model texture. A comparable result demonstrates capture with 3D Scan App by iPhone 14 Pro. The same problem occurred with active scanning: fine features were

visually recognisable in the textures (excluding light sectioning), but not resolved geometrically. It also should be taken into account, that some timber surfaces can be painted or dusty/dirty.



**Figure 1.** Wood 3D recording results: a) 3D model from iPhone 14Pro, b) Macro-lens camera Nikon (textured & geometry only), c) Creaform structured light scanner, d) Hexagon measuring arm, e) Laser light section scanner SmartScan; f-h) wood surface under different illumination direction (RTI).

Alternatively, multi-view photometric stereo techniques, which can be used to produce albedo, normal map and higher spatial resolution depth maps have proved to be useful for 2D and 3D defect analysis (Karami et al., 2022a) (Fig. 2). Moreover, grazing angle images from the multiple lighting setup can be automatically chosen to enhance the detection of specific features according to their spatial orientation (Karami et al., 2022b).



**Figure 2.** Multi-view photometric stereo datasets and outputs: a) Multi-View Photometric Stereo system, b) combined image obtained as median of the 20 input images with different illumination, c) albedo, d) RTI normal map and roughness analysis from

3D point cloud obtained from e) integrate normals and f) photogrammetry respectively, which shows the benefits of photometric stereo in reproducing higher frequencies.

## 2.2 Methodological baseline

Considering the subtle changes on wood surfaces, aging marks, historical significance, and environmental impacts, it becomes statistically challenging to apply traditional machine learning techniques to the wood knots detection mission. Traditional machine learning methods predominantly rely on predetermined mathematical algorithms and manually curated feature sets for pattern identification. While such methods may be appropriate for discerning uniform patterns, exemplified by tree rings due to their inherent consistency, deep learning's inherent flexibility and robustness offers a better-suited avenue for tasks necessitating the discernment of multifaceted and heterogeneous patterns within datasets.

CNN-based frameworks, such as PointNet (Charles et al., 2017), pave the way for innovative approaches to 3D point cloud model analysis.

The workflow of deep learning for CNN model such as YOLO model on historical wood surface follows the general method of deep learning, which can be simplified as the following steps:

- **Data pre-processing**

In order to meet the requirements for training neural networks, it is essential to subject the collected data to rigorous pre-processing steps. This entails for example data cleaning to remove any anomalies or outliers, or normalization to ensure data consistency.

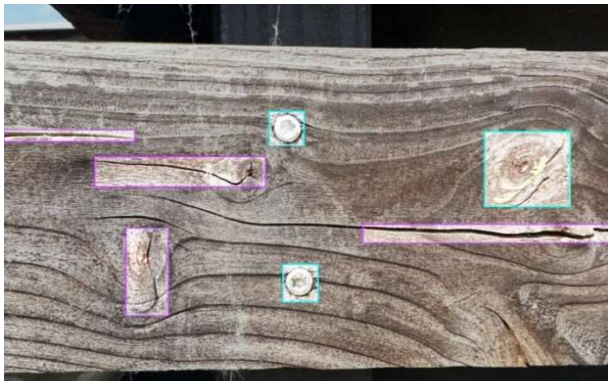
- **Model training**

Model training is the fundamental process in deep learning. It involves adjusting the weights of different parameters within a model using a training dataset. The primary goal of this process is to minimize the discrepancy between the model's predictions and the actual results (Fig. 3). This discrepancy is quantified using a loss function to evaluate the performance of the model. Iterative optimization techniques can also be used to adjust the model architecture to achieve superior prediction accuracy.

- **Model validation and testing**

Once the model has been trained, validation is required to ensure that the model has good generalization performance and does not overfit the training data. The validation dataset is used to evaluate the model's performance after each iteration during the training process and metrics such as accuracy, precision, recall and F1 score would be used to stage the training results. The F1 Score, also known as the F1 measure or F-score, is a statistical measure of classification model precision. It takes into account the precision and recognition possibility of a classification model, as well as correlated average of both. The F1 Score is particularly useful in cases, where there is an imbalance between positive and negative samples. Any discrepancies or problems identified at this stage may lead to further model optimization. Once the training process finished, the model will be applied to the test dataset. As the model has not seen this data during the training phase, this ensures an unbiased assessment of its performance in a real-world environment.

After obtaining a more reliable model, it should be combined with a real-life scenario, e.g. by adding it to a wood monitoring system, to verify its stability in practice and to detect problems that may arise in practice, for example due to the ambient light being too dark and thus not working correctly. This process requires repeated attempts to document the ability of the model applied in different scenarios, to repeat the previously mentioned process of training the model, and testing it in practice to achieve a more satisfactory result.



**Figure 3.** Labelled wood surface with wood crack (violet), wood knot (green) and wood nail (blue).

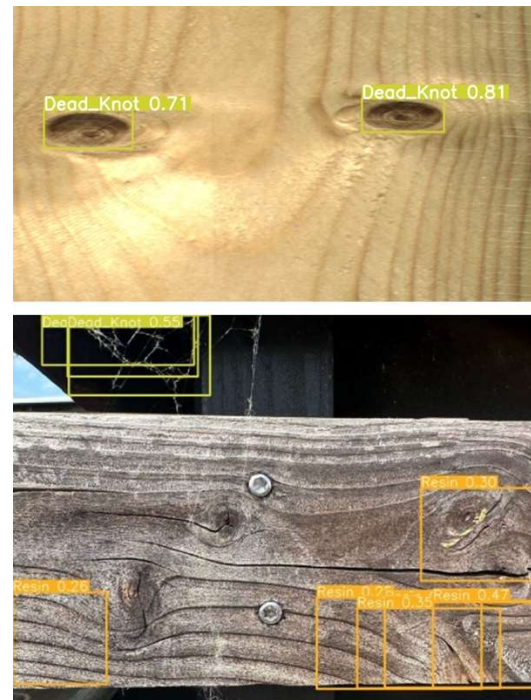
### 2.3 First results with YOLOv8n

For our initial experiment in wood knot detection, we selected the YOLOv8n model, which is the fastest model among all the pre-trained YOLOv8 model for image detection on the COCO dataset. The COCO (Common Objects in Context) dataset, created by the Microsoft Research team, is a widely used dataset for computer vision research, especially in the field of object detection, segmentation and image annotation. It contains a large number of images from diverse scenes and contexts, covering common objects and scenes in daily life. Our goal was to quickly assess the feasibility of the method. We chose the subset of the large-scale image dataset of wood surface defects (Kodytek et al., 2022), which was modified by Nouman Ahsan (Large Scale Image Dataset of Wood Surface Defects, 2023) This subset contains 4,000 labelled images of eight different categories of wood surface defects, such as live knots, dead knots, and knots with cracks. Taking advantage of the pre-labelled images in the dataset, we divided it into three parts: 81.25% for training, 12.5% for validation, and 6.25% for testing. The labelled test set allows us to evaluate the detection capability of the final trained model, while the validation set helps to assess the performance of the model after each training iteration.

We followed the standard YOLO training configurations recommended for initial experiments. The model was programmed to train for 30 epochs, with a batch size of 8 per epoch. In machine learning terminology, a 'batch' refers to the subset of samples used for each training iteration, known as an 'epoch', which represents a full training cycle through the dataset. For our experiment, this meant that the model processed approximately 12,188 batches from the 3,250 training samples.

Analysis of the training and test results showed that the trained model performed well in identifying both live and dead knots. However, it was less accurate at recognizing images of historic timber structures (Fig. 4), a limitation that was anticipated before the start of our study. This is likely because the model was only trained on a generic dataset without further specialised tuning, resulting in limited generalization.

In our future research, we plan to conduct a more in-depth investigation in line with our research objectives, which will include collecting data on various historical wood surface damages, testing and validating a range of deep learning models, and making appropriate modifications to the neural network to account for the data characteristics. This may include incorporating various filters or an attention mechanism to improve the model's ability to meet the detection performance requirements.



**Figure 4.** Top: Detection results using image from the dataset. Bottom: Detection results using image captured from real scene.

## 3. THE DOMINICAN CHURCH IN BAMBERG

### 3.1 The Dominican Church case study

In our initial research, we observed limitations in the available open-source datasets, particularly regarding the variability of historic wood surfaces and real-life scenarios. To conduct a robust preliminary study, we chose the Dominican Church in Bamberg as our first research object to obtain a set of results from real-life scenarios using tests from various models and strategies (Fig.5). The Dominican Church, constructed by the Dominican Order before 1400, is located in the historic city centre of Bamberg, and is no longer used as church since 1803. Our research is specifically targeting the historic wood roof structures of the church. The roof components, which are notable for their presence of original historic surfaces, have undergone a careful preservation of most of the original structure and sensitively applied structural improvements in a comprehensive conservation efforts (Eißing & Kraus, 2017). It therefore provides a unique opportunity to study and develop advanced restoration and conservation methods. Future work will aim to contribute to the conservation of this significant architectural heritage with observations on the roof structure.



**Figure 5.** Left: original image from Dominican Church roof. Right: cropped images (each in 640 x 640).

### 3.2 The general processes

As stated above, the aim of our project is to investigate a wood knot recognition approach using 2D and 3D data from recorded timbers. For photogrammetric analysis we used Nikon D3400 (APS-C) and Sony Alpha 4a (full frame) cameras to capture the entire structure with a focus on the wood knots, especially the dead wood knots on the historic wood surfaces. A collection of over a thousand images at various resolutions, including large scene shots with randomly distributed wooden knots and detailed close-ups of specific groups of knots, has been recorded. In addition, we have used a Faro Focus S 350 laser scanner to capture point clouds of the roof structure. From this data, orthophotos are created subsequently. The complete set of images provides the data base for AI based feature recognition.

The 3D pointclouds (from SfM and TLS) are further used to build a 3D BIM model in which the detected features will be geo-referenced.

For this preliminary study, identifying historic wood knots in 2D images proved to be more efficient and yielded results in line with expectations. Our immediate goal is to meticulously pre-process these 2D images to match the criteria of our standard dataset. This involves a detailed examination of several parameters.

### 3.3 Pre-processing of captured data

We began by selecting images that had the desired characteristics of wood knots, such as distinct features, optimal brightness, and contrast, among others. After curating an ideal set of images, we standardised the sizes to 640 x 640 pixels to ensure uniformity and compatibility with our analysis tools, tailored to the requirements of the YOLO model. Due to the uneven distribution of features, we applied a graded luminance filter to the cropped images to ensure that the dataset included a wide range of feature images under stable lighting conditions. Following these steps, we compiled a refined image set to serve as the basis for our custom dataset, which currently only includes image data from the wooden structures of the Dominican Church roof.

As a continuation of our pre-processing work, we carried out extensive labelling operations on selected images. The labelling classification consisted of three main categories: dead knots, live knots and wooden dowels. Due to the historical nature of the target building, dead knots are the most common feature among all three classes. However, the characteristics of wooden dowels can easily be confused with dead knots in timber frame analysis, making identification of the two a challenge. Therefore, we created the dowel category specifically to provide a more nuanced distinction between the two features (Fig.6). Our categorization effort not only improves the quality of the dataset, but also is critical for the accuracy of future automated feature recognition using deep learning models like YOLO. With such categorization, we expect the model to be able identifying accurately each type of wooden knot, providing more accurate data to support the conservation of historic buildings for further project phases.



Figure 6. Dead wood knot (left) and wood dowels (right).

### 3.4 Training process and experiments

As we had already used YOLOv8n in our first quick attempt, we developed a series of training experiments using both YOLOv8n and YOLOv8m and different epochs and additional data augmentation to achieve a better validation consistency within the dataset we had established from the Dominican Church. YOLOv8m (Medium) has a larger structure and offers greater accuracy than YOLOv8n (Nano). However, it requires more graphics processing power and takes longer to train on the same dataset. The complete original dataset with around 750 cropped images has been divided randomly into training set (80%), validation set (15%) and test set (5%). The test set will be used to verify the performance of the final model. For the following analysis of experiments, we will focus on the training and validation sets used during the training process.

In our first experiment with YOLOv8n, conducted over 10 epochs, the model achieved a peak precision of 0.62 and the mean Average Precision 50-90 (mAP 50-90) of 0.17 on the validation dataset, which is dominated by dead wood knots. mAP, which stands for mean Average Precision, is a performance metric that reflects the accuracy of object detection models across various classes and instances within a dataset. A high mAP value generally indicates that the model reliably detects most objects and accurately predicts their locations and classes. The term "mAP 50-90" refers to the mean of the Average Precision (AP) scores calculated at different IoU (Intersection over Union) thresholds, starting from 0.5 to 0.9, in increments of 0.05.

In particular, the precision for detecting dead knots on the wood surface reached 0.74. Subsequently, using YOLOv8m with identical parameters for a comparable number of epochs resulted in an improved maximum precision of 0.67 and increased mAP 50-90 of 0.2 on the validation datasets, with the precision specifically for dead knots increasing to 0.79. Despite this improvement, the loss plots show a more stable performance for YOLOv8n compared to YOLOv8m, albeit with lower overall precision values. This discrepancy suggests that YOLOv8m may require a longer training period to fully converge and realise its optimal performance.

Therefore, we continued with further controlled experiments by increasing the training iteration of the YOLOv8m to observe the statistical change on various parameters such as loss on both training and validation sets or mAP values.

To compare the training process using 10 epochs, we extended the time periods to 30 and 50 epochs (Fig.7).

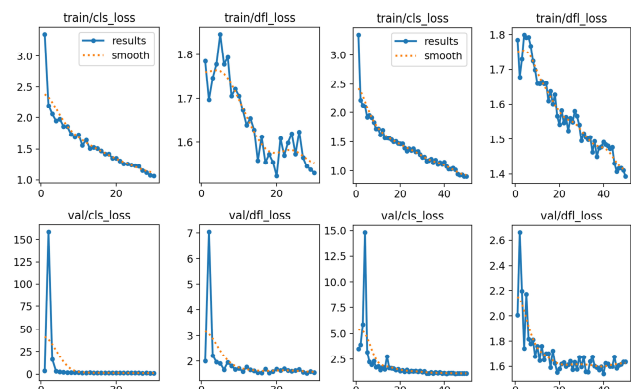


Figure 7. Results of YOLOv8m in 30 epochs (left) and 50 epochs (right).

The following diagrams show the change of CLS loss (classification loss, using the values of cross-entropy loss) and DFL loss (distribution focal loss) on both the training set and validation set (Fig.8).

Another loss, we also considered, is called BOX loss (bounding box loss). However, since we focus on the general performances within various iterations during the experiments, we may discuss this kind of loss in later research. From the experimental results, we can observe that the performance of the YOLOv8m model tends to stabilise as the number of training cycles increases, with a gradual decrease in the various loss metrics (including CLS loss and DFL loss). Although some fluctuations in performance still occur at higher training cycles, i.e., 50 epochs, overall, the model trained for longer cycles shows better performance in terms of the absolute value of loss metrics. This suggests that the model converges with increasing training time and that long-term training can still lead to improved performance even with short-term fluctuations.

In the meanwhile, the experimental results also show that the YOLOv8m model exhibits a steady decrease in classification loss and distribution focal loss as the number of training cycles increases. However, the recognition accuracy of the model still needs to be improved for class-specific distinctions, such as dead knot versus wood dowel, or under poor lighting conditions. The dataset we currently use is relatively small, containing only about 750 cropped and filtered raw images, which is often considered insufficient in deep learning and target detection tasks. In view of this, we plan to extend the dataset through data augmentation techniques to increase the variety and robustness of model training. Data augmentation will include image rotation, scaling, cropping, colour transformation and other operations to generate more varied images to simulate different recording conditions and background noise, with the expectation, that the model will maintain high accuracy in recognition across a wider range of scenarios. To address the limitations of our dataset's size and to enhance the robustness of our YOLOv8m model, we have implemented a series of data augmentation techniques, effectively tripling the size of our dataset.

The augmentations applied are as follows:

- **Shear transformations**

Images have been sheared horizontally by up to  $\pm 19$  degrees and vertically by up to  $\pm 14$  degrees, introducing slanting effects that simulate different perspectives and angles of viewing.

- **Brightness adjustments**

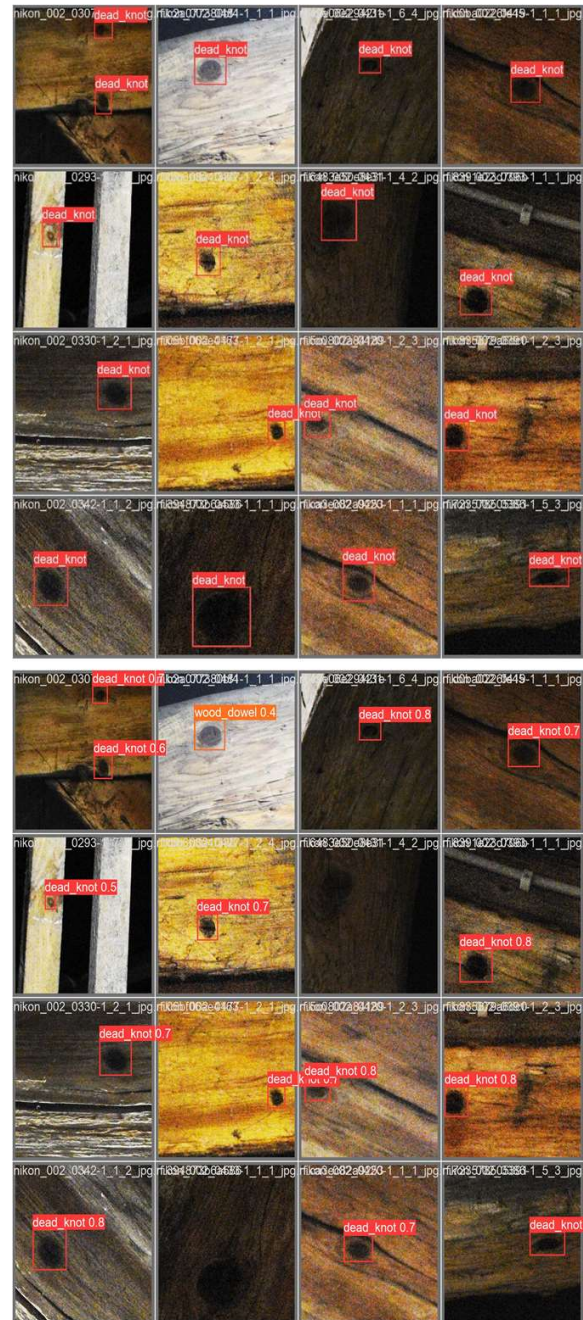
The brightness of images has been varied between -52% and +52%, which prepares the model to recognize objects under varying lighting conditions, from dimly lit to brightly illuminated environments.

- **Bounding box rotations**

The images along with their bounding boxes have been rotated 90 degrees in all possible orientations- clockwise, counter-clockwise, and upside down. This enhances the model's ability to detect objects regardless of their orientation in the image plane. Those augmentations were selected to reflect the potential variability in real-world scenarios and to ensure that the model is not only trained on a higher volume of data but also on data that challenges its recognition capabilities across a range of conditions.

Utilizing the augmented dataset, we re-distributed the data into training, validation, and test sets at random, as we did with the original dataset, and conducted comparative trainings for 30 and 50 epoch iterations. Post-augmentation, we observed several notable enhancements in the model's training performance: the overall loss declined more rapidly and with greater stability compared to the previous non-augmented results (Fig.9)

Furthermore, both the classification loss (CLS loss) and directional focal loss (DFL loss) decreased, maintaining a consistent downward trend across training and validation sets (Fig.10).

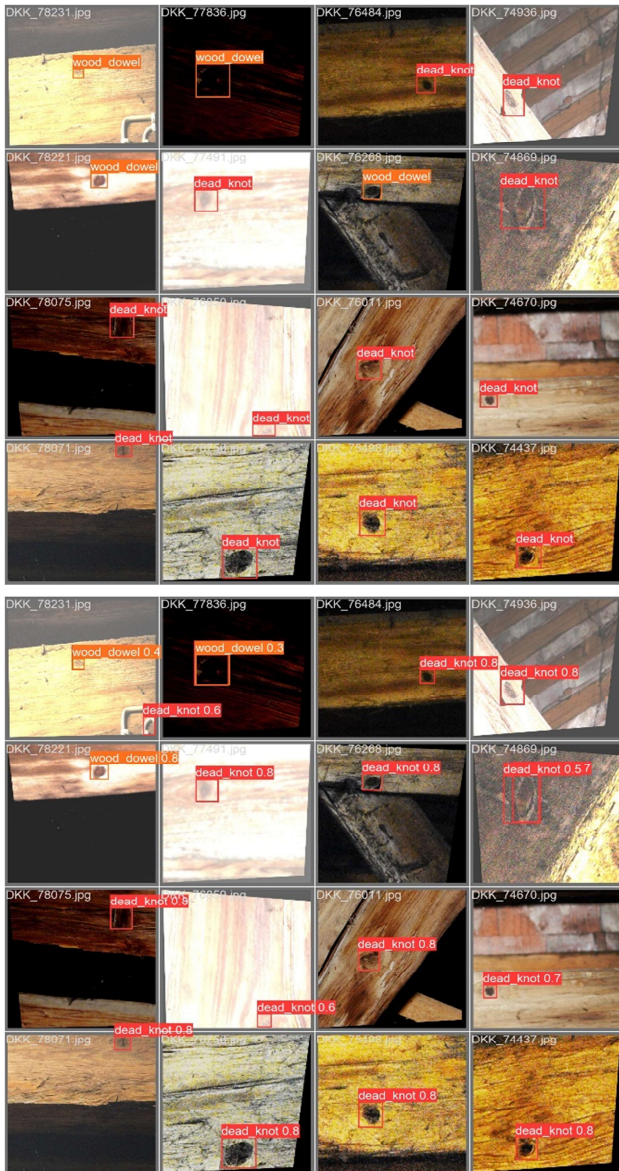


**Figure 8.** Results of YOLOv8m (50 epochs) on validation set; top: labels on validation set; bottom: predictions on validation.

train/cls\_loss      train/df\_loss      train/cls\_loss      train/df\_loss

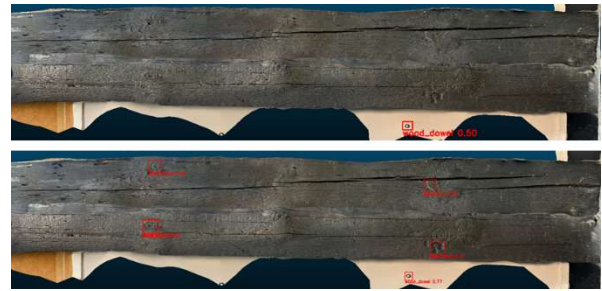
0      20      0      20      0      20      0      20      40

**Figure 9.** Results of YOLOv8m with data augmentation in 30 epochs (left) and 50 epochs (right).



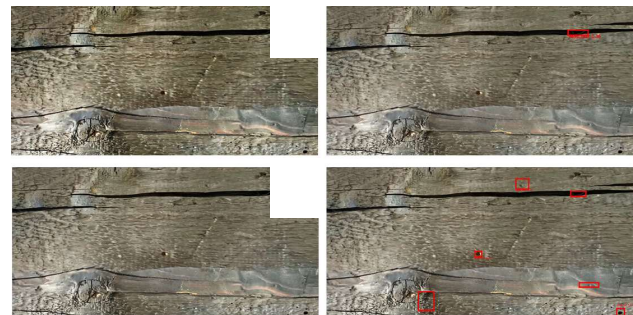
**Figure 10.** Results of YOLOv8m with data augmentation (50 epochs) on validation set; top: labels on validation set; bottom: predictions on validation set.

This contrasts with earlier results where loss reductions were beginning to plateau, and where DFL loss indicated a slight tendency towards overfitting beyond 45 epochs. These improvements suggest that the model's generalization ability has been strengthened, likely due to it learning more generalised features from the augmented training set. We also anticipate better detection performance in specialised scenarios, such as in challenging lighting conditions. The same workflow was applied for 2D images derived from 3D point clouds using structured light scanning and a photogrammetric approach. The orthogonal projection of textured models has been generated in CloudCompare software by different field of views from 5 to 30 degrees (whole timber) and two zoom levels: zoom level 1 means rendered image with 1647x801 px size and 3294x1602 px for zoom level 2. Figure 11 demonstrates detection results for 30° field of view and both zoom levels.



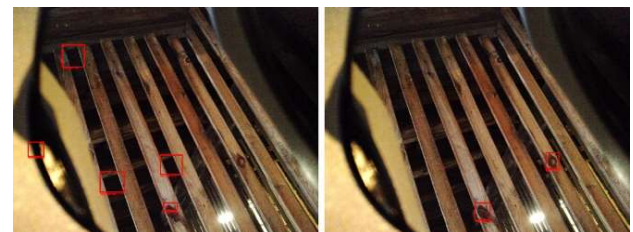
**Figure 11.** Detection results on rendered images with 30° field of view: top - 1647x801 px, bottom - 3294x1602 px.

In comparison to rendered images with 30° field of view, the images with 5° field of view have demonstrated a similar detection rate. In both cases a zoom level had a pivotal significance (Fig. 12).



**Figure 12.** Rendered images with 5° field of view from orthogonal projections of 3D surfaces (left) and detection results (right): a) 1647x801 px, b) 3294x1602 px.

Finally, 2D images from digital camera integrated in Faro Focus S 350 have been cropped in 640x640 as well as 960x960 size and processed in the same manner (Fig. 13).



**Figure 13.** Detection results from Faro Focus S 350 integrated camera cropped 640x640 (left) and 960x960 size (right).

### 3.5 Limitations

We have described a series of experiments conducted on data gathered from the Dominican Church, focusing on the effects of different models, training iterations, and data augmentation techniques on the recognition of historical wood features, primarily utilizing 2D images. During this research, we have identified several areas that warrant further attention:

- **Original data utilization**

Up until now, our dataset has been limited to two-dimensional photographic representations of the historical timber roof structure of the Dominican Church. The incorporation of three-dimensional point cloud data remains an untapped resource that we plan to use as next steps. Future studies will incorporate 3D data, aiming to develop a hybrid multi-source detection approach that enhances accuracy.

- **Feature class diversity**

Our labelling efforts predominantly focused on dead knots, with a minority of wood dowels that pose a classification challenge due to their similarity. An immediate research objective is to diversify the feature labels, ensuring a balanced representation within the dataset for accurate multi-class recognition.

- **Historical wood architecture in real scenarios**

The ultimate test for our system will be its performance in authentic settings. Factors such as lighting conditions, specific objectives, and feature complexity must be accounted for. To bolster the model's generalization capabilities, we may need to encompass a broader spectrum of real-world conditions or utilize generative AI algorithms to enrich scene representation in our dataset.

- **YOLOv8 Model Architecture**

While the YOLOv8 model stands as a leading real-time recognition framework, its suitability for our diverse targets and complex scenarios remains to be fully evaluated. We are contemplating augmentations to the existing YOLO architecture, potentially incorporating attention mechanisms or bespoke feature extractors, to optimize the extraction of distinct characteristics on historical wood surfaces.

#### 4. CONCLUSIONS

Based on the analysis of ongoing case studies and related experiments, several key conclusions have been drawn. Firstly, the use of pre-trained models such as YOLO for the identification of historic wood surface knots has demonstrated considerable accuracy after training on different datasets.

However, our experimental analysis of the Dominican Church roof structures shows that despite the use of identical datasets, different strategies and methodologies significantly affect model performance. Therefore, to construct an effective and universal detection system, it is imperative to consider not only data diversity, but also the integrated selection and combination of strategies, aiming to optimize the balance of model performance in different application scenarios.

Secondly, while our primary focus is on the identification of wood knots, other factors in real-world scenarios, such as wood cracks, fungal attack and weathering traces also significantly affect the physical properties of wood. Therefore, in addition to concentrating on knot identification, the potential impact of these factors on detection performance needs to be considered. In addition, the collection of a wider range of data under extreme conditions is crucial to improve the generalization capabilities of the model.

Finally, as an organic material, wood surface has identifiable biological characteristics, our research should not be limited to the dataset and model performance in deep learning alone. Future studies should delve deeper into the biological aspects of wood, such as the formation of knots or the biological mechanisms behind other damages, as this knowledge will provide significant auxiliary information for the recognition capabilities of deep learning models.

#### REFERENCES

Charles, R. Q., Su, H., Kaichun, M., Guibas, L. J., 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Proc. CVPR*, pp. 77-85.

Ding, F., Zhuang, Z., Liu, Y., Jiang, D., Yan, X., Wang, Z., 2020. Detecting Defects on Solid Wood Panels Based on an Improved SSD Algorithm. *Sensors*, 20(18):5315.

Divya, K., Kaur, S., 2021. Dendrochronology with Deep Learning. *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, London, United Kingdom, pp. 333-336, doi: 10.1109/ICIEM51511.2021.9445305

Ebner, F., 2018. Festigkeitsuntersuchungen an verbauten Konstruktionshölzern. Unpub. Master thesis in the master program Heritage Conservation.

Eißing, T., Kraus, S., 2017. Vom Pirckheimer Stübchen zum Koberger Schrankstübchen: Neue Erkenntnisse zu einer außergewöhnlichen Holzkonstruktion des späten 15. Jahrhunderts. In *Wartburg-Stiftung (Ed.), WARTBURG-JAHRBUCH 2016: Herausgegeben von der Wartburg-Stiftung in Zusammenarbeit mit dem Wissenschaftlichen Beirat* (pp. 131-167). Schnell und Steiner Regensburg.

European Green Deal, 2019. [https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal\\_en](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en)

Han, S., Jiang, X., & Wu, Z., 2023. An Improved YOLOv5 Algorithm for Wood Defect Detection Based on Attention. *IEEE Access*, 11, 71800–71810.

Karami, A., Menna, F., Remondino, F., 2022a. Combining Photogrammetry and Photometric Stereo to Achieve Precise and Complete 3D Reconstruction. *Sensors*, 22(21), p.8172.

Karami, A., Menna, F., Remondino, F., Varshosaz, M., 2022b. Exploiting light directionality for image-based 3d reconstruction of non-collaborative surfaces. *The Photogrammetric Record*, 37(177), pp.111-138.

Kodytek, P., Bodzas, A., & Bilik, P., 2022. A large-scale image dataset of wood surface defects for automated vision-based quality control processes. *F1000Research*, 10, 581.

Large Scale Image Dataset of Wood Surface Defects, 2023. Retrieved 14 January 2024, from <https://www.kaggle.com/datasets/nomihsa965/large-scale-image-dataset-of-wood-surface-defects>

Liu, Y., Hou, M., Li, A., Dong, Y., Xie, L., Ji, Y., 2020. Automatic detection of timber-cracks in wooden architectural heritage using yolov3 algorithm. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020, 1471-1476.

Luhmann, T., Robson, S., Kyle, S., Boehm, J., 2023. *Close-Range Photogrammetry and 3D Imaging*. de Gruyter, Berlin.

O'Shea, K., Nash, R., 2015. An Introduction to Convolutional Neural Networks. *arXiv:1511.08458*.

Remondino, F., Stylianidis, E., 2016. *3D Recording, Documentation and Management of Cultural Heritage*. Whittles Publishing Verlag, 978-1-84995-168-5

Yang, Y., Wang, H., Jiang, D., Hu, Z., 2021. Surface Detection of Solid Wood Defects Based on SSD Improved with ResNet. *Forests*. 2021; 12(10):1419. <https://doi.org/10.3390/f12101419>

YOLOv8: A New State-of-the-Art Computer Vision Model. (2023). Retrieved 1 Jan 2024, <https://yolov8.com/>