SmartWoodID

Smart classification of Congolese timbers: deep learning techniques for enforcing forest conservation

Contract - B2 / 202 / P2 / SmartWoodID

SUMMARY

Context

A substantial part of the timber trade is still illegal, and illegal logging is the most profitable biodiversity crime. UN Environment estimates that illegal logging and the associated timber trade counts up to US$50 to $152 billion per year. Illegal logging involves a high risk of irreversible damage to ecosystems associated with the exploitation of highly sought after, sometimes protected, species. Timber regulations are already active (CITES, FLEGT, EUTR), but implementation and enforcement are a challenge. Currently, Belgium has the negative connotation of being the 'hub of illegal timber trade'. 27.5% of the total EU28 imports of primary tropical timber products are imported via Belgium (mainly via the port of Antwerp). Wood identification is a key process in the enforcement that needs to check whether the shipment corresponds with the products mentioned on the accompanying documents. For this reason, there is a growing demand for timber identification tools that can be applied by law enforcement officers.

The Tervuren xylarium is the Belgian governmental collection of wood samples. It is an internationally renowned part of the federal scientific heritage, housed by the Royal Museum for Central Africa and comprises reference material of 13 000 different botanical species. One of the growing actual functions of the collection is supporting forensic research through verification of a species' identity. The most common technique of timber identification is a wood anatomical assessment. Machine Learning methods are likely to be able to assist the wood identification process for non-specialists. Wood species have indeed characteristic features at different microscopical magnifications. However, some of those features are highly variable, which hampers the development of classical dichotomy identification keys that can be used by non-specialists. Moreover, many features seen on wood surfaces are to be understood as artifacts (fissures, traces of mechanical damage, fungi and insect attacks) and are not always easy to distinguish from diagnostic characteristics for the untrained eye. The Tervuren xylarium offers the most complete assemblage of reference material for the development of new wood identification approaches.

The project aimed at automating part of the wood identification process by applying artificial intelligence techniques for the analysis of wood anatomical images of timber species of the Democratic Republic of the Congo. The tree flora of Central Africa comprises 3013 species, 27 of these belong to the class 1 commercial timber species of the DRC and are actually intensively logged and traded, 20 to class 2 (have potentially a big commercial value), 44 to class 3 (are considered to be promoted) and 879 to class 4 (commercial value is not yet known). The project used xylarium samples of all the species of the four classes and took advantage of the power of modern deep learning approaches. The project relied on expert wood anatomical descriptions which served as annotated training data to develop the software. The project was unique because of the large number of African species, the application of deep learning and a database of standardized descriptions that are available online. In a first work package expert annotations of microscopic and mesoscopic images of transverse surfaces of 1000 Congolese wood species were made. In Work package 2 we developed an image processing pipeline for semi-automated annotation of microscopic and mesoscopic wood sections. In Work package 3 we focused on the production of a user-friendly interface.

Objectives

The SmartwoodID project aimed at improving the process of analysis of timber species from the Democratic Republic of the Congo (DRC), a large species-rich country where identification is a daunting task, leading to regular law breaches. The general ambition was to explore the possibilities for automating identification of timber species by applying machine learning, using large datasets of macroscopic features and high-resolution optical scans of endgrain surfaces.

The project had three research objectives:

- **Research objective 1:** Constructing a database of flat-bed scans from end-grain surfaces and macroscopic features of timber species from the Democratic Republic of the Congo.
- **Research objective 2:** Development of illustrated classification keys for wood identification of Congo.
- **Research objective 3:** Developing of an illustrated classification key assisted by machine learning.

Conclusions

The SmartWoodID image database offers new opportunities for developing identification systems based on recognition of diagnostic wood anatomical features. This database is unique since it covers a large number of African tree species and lower taxa of which the macroscopic structure is visualized and described. The Tervuren Wood Collection provides this thanks to its heritage of collecting reliable reference material over the span of more than a century. A total of 56% of all DRC tree species and lower taxa are currently available within the Tervuren Wood Collection. The first version of the SmartWoodID image database that is presented here consists of a set of 954 timber species and lower taxa present in the DRC forests. The database focuses on the macroscopic anatomical features that can be encountered on a high-resolution scan of end-grain wood surface. The database accounts for irregularities and natural variability, using multiple specimens with large end-grain surfaces. This makes it a robust reference database for research on wood in general and will allow the development of tools for aiding in law enforcement to combat illegal logging.

The study into identification keys (WP2) highlights the inherent limitations of the 31 expert-defined, accessible macroscopic cross-sectional features for taxonomic identification across a diverse range of timber species, such as those found in the Congo Basin. While classification accuracy improves at higher taxonomic ranks, genus- and family-level predictions remain limited due to overlapping anatomical features among taxa. Nevertheless, macroscopic cross-sectional features retain diagnostic value when applied within narrower taxonomic scopes. The successful discrimination of *Pterocarpus* species—once considered indistinguishable without laboratory-based methods—demonstrates that readily observable anatomical features can enable species-level identification in the field. Further improvements in diagnostic accuracy can be achieved by incorporating high-resolution, large-area imaging and multi-specimen datasets. These approaches more effectively capture intra-specific anatomical variability than conventional single-sample methods and enable the extraction of quantitative anatomical information at a finer scale. Integrating such enhancements offers promising pathways to increase both taxonomic resolution and classification reliability. However, practical constraints in field environments—such as time pressures and limited equipment—necessitate alternative strategies for reliable in situ wood identification. Continued progress will depend on advancing CV–based identification systems, particularly CNNs, which can directly process macroscopic images to deliver accurate and rapid classification. These models offer strong potential for scalable, efficient, and field-ready timber verification applications.

The results on WP3 show that CNN, such as the applied Xception architecture, can successfully extract features for classifying image patches of sanded cross-sectional images to classify different timbers at the genus level, and distinguishing between anomaly-free and anomalous wood. The performance on the test data varied for individual genera, with some benefiting from training on anomaly-free images, while for other genera, like *Cynometra*, higher recall was observed for the model trained on anomalous images. Grad-CAM analysis revealed the model's preference for regions on patches showing unobscured wood anatomical tissue, underscoring the importance of clear wood anatomy in training CNNs for wood identification. This could enable CNNs to capture diagnostic patterns more effectively, which in turn would lead to better discrimination between timbers, even when applied to anomalous specimens in the field. The inclusion of anomalous patches had a limited impact, but subtly enhanced performance on anomalous patches. The findings therefore suggest that CNNs (like Xception) demonstrate the highest proficiency in timber classification when trained on anomaly-free images, making this approach highly effective for developing CV-based wood identification models for deployment in the field. This demonstrates the potential of deep learning for automated timber genus identification. The results highlight that while all classification models capture similar underlying patterns, CNNs outperform identification keys on the 31 macroscopic cross-sectional IAWA features due to their ability to extract discriminative image features without relying on predefined descriptors. Performance trends across different taxonomic scopes emphasize the importance of training data diversity, as models trained on broader datasets exhibit greater generalization capabilities compared to those trained exclusively on commercial timbers.

Beyond standard multiclass classification CNN, object re-identification CNN approaches provide valuable alternatives, particularly in forensic contexts where identification may be less critical than ruling out certain timbers. The binary verification approach demonstrates strong performance in this regard, though effectiveness is constrained by ranking limitations and computational demands. Embedding-based re-identification, while computationally efficient, underperforms in this study, suggesting that improved mining strategies and loss functions could enhance its reliability. Additionally, object re-identification produces information on similarity to specific reference specimens, rather than producing a direct prediction of classes (e.g. genera in this study), providing valuable information for forensic researchers.

The results on an integrated key using CV and the 31 expert-defined features demonstrate that while integrating expert-defined macroscopic anatomical features can yield moderate improvements in CNN-based genus predictions, these benefits are highly dependent on both the specific genera and the depth of re-ranking applied. The overall results affirm that CNN models alone already encode substantial taxonomic information, likely due to their training on challenging diagnostic comparisons that extend beyond traditional anatomical descriptors. Crucially, re-ranking within the top two to five CNN predictions offers the most consistent performance gains across accuracy, precision, and recall—especially at the top three threshold. Beyond this range, performance diminishes due to misclassifications introduced by overemphasizing weak or misleading anatomical features. This finding underscores the limited but strategic utility of macroscopic cross-sectional wood anatomy for refining identifications. From a field application perspective, particularly in the context of frontline timber verification, the implications are twofold. First, the CNN model offers a rapid and accessible method for genus-level identification that already performs well in most cases. Second, refinement methods such as re-ranking must be applied selectively, as indiscriminate use—especially on protected taxa—can reduce recall, increasing the risk of overlooking high-priority timbers such as *Khaya*.

Future research should address the limitations of embedding-based models using the current state-of-the-art to offer a powerful approach for automated wood identification. These models should be explored in hybrid studies that integrate multiple diagnostic data modalities—macroscopic images, microscopic wood anatomy, and chemical fingerprinting—to improve classification accuracy and enable identification at sharper taxonomic resolution. As global efforts to combat illegal logging and enforce sustainable trade regulations intensify, advancing AI-driven timber identification will be essential for strengthening forensic capabilities and ensuring responsible resource management.

Keywords

Wood Anatomy
Wood Identification
Macroscopic wood anatomical assessment
Cross-section
Computer Vision-based wood identification
Object re-identification
Gradient-weighted class activation mapping