

1 Background

Sawmill industry makes use of modern sensor systems and new computational methods for implementing digitalisation. These technological advancements enable developing new machine vision based measurements for quality control systems. The capabilities of the systems can be extended by using measured information from the raw material and the sawmill processes. The obtained quality information can be used in the control to optimise the sawing processes. Data-driven identification of wood species from sawn timber would be beneficial for guaranteeing that different species are not mixed in the process.

2 Problem Statement

The main objectives of the assignment are as follows:

- Study and analyse either macroscopic or microscopic image data of different wood species. The data to be used is described below.
- Train a convolutional neural network (CNN), classify the images to the predetermined classes (wood species) and quantitatively evaluate the model performance.

While studying the data and implementing the classifier, the following viewpoints should be considered and also documented in the report:

- Is there a need for data preprocessing? If yes, what kind of approach should be used?
- How to select the architecture for the CNN and utilise the relatively large number of images in such a way that the required computing effort is reasonable? Should transfer learning and a pre-trained model be used?
- How should the images be used for training, validation and testing?
- In which form should the colour image data be used as the input for the model? For example, depending on the data set, the image resolution is either 3264 x 2448 or 1024 x 768 pixels.
- How to affect the efficiency of the training process and convergence to an acceptable classification performance?
- What is or which are the relevant metrics for evaluating the classification performance? What is the baseline performance of a random classifier with the selected data set?

3 Data

For appropriate focusing of the task, select one data set from the following list. Focus on using the selected data set for solving the assignment.

3.1 Data Set 1: Macroscopic images of wood species

The **Forest Species Database – Macroscopic** (FSD-M) contains 2,942 macroscopic images of 41 different wood species. The colour images have been captured with a digital camera in macro mode and the spatial resolution of the images is 3264 x 2448 pixels. The total size of the data set is approx. 3.5 GB. The data set description including information about the encoding of the class labels into the filenames is available as a scientific article [1].

3.2 Data Set 2: Microscopic images of wood species

The **Forest Species Database – Microscopic** contains 2,240 microscopic images of 112 different wood species. The images have been captured through microscope optics and the spatial resolution of the images is 1024 x 768 pixels. The total size of the data set is approx. 8.5 GB. The data set description is also available as a scientific article [2].

4 General instructions

The general instructions are given as a separate document. They include the requirements, deadline and information on the grading.

If there are any problems with the assignment description and/or data, contact the person supervising the practical assignment. This should be done before inventing your own interpretations or making (too) radical assumptions.

References

- [1] Pedro L. Paula Filho, Luiz S. Oliveira, Silvana Nisgoski, and Alceu S. Britto. Forest species recognition using macroscopic images. *Machine Vision and Applications*, 25(4):1019–1031, May 2014.
- [2] J. Martins, L. S. Oliveira, S. Nisgoski, and R. Sabourin. A database for automatic classification of forest species. *Machine Vision and Applications*, 24(3):567–578, April 2013.