© IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

Cross-Modality Wood Log Tracing*

1st Georg Wimmer
Department of Computer Sciences
University of Salzburg
Salzburg, Austria
Email: gwimmer@cs.sbg.ac.at

2nd Rudolf Schraml
Department of Computer Sciences
University of Salzburg
Salzburg, Austria
Email: rschraml@cs.sbg.ac.at

3rd Lukas Lamminger
Department of Computer Sciences
University of Salzburg
Salzburg, Austria
Email: lukas.lamminger@stud.sbg.ac.at

4th Alexander Petutschnigg
University of Applied Sciences Salzburg
Kuchl, Austria
Email: alexander.petutschnigg@fh-salzburg.ac.at

5th Andreas Uhl
Department of Computer Sciences
University of Salzburg
Salzburg, Austria
Email: uhl@cs.sbg.ac.at

***Abstract*—The proof of origin of logs is becoming increasingly important. In the context of Industry 4.0 and to prevent illegal logging there is an increased interest to track each individual log. In the near future more and more sawmills will be equipped with a computed tomography (CT) scanner. In order to establish wood log traceability from the forest to the sawmill this work investigates log recognition based on RGB log end images captured in the forest and CT log images captured in the sawmill. The advantage of that approach is that CT scanners are already applied in big saw mills to optimize the saw cut and so the logs only have to be recorded once more in the forest which saves time and cost. To bridge the domain shift between CT and RGB images, we apply widely known domain adaption approaches and present a novel filtering approach. Log recognition is done using a convolutional neural network (CNN) based method using the triplet loss for CNN training and a novel shape descriptor. The results (equal error rate of 13%) show that the recognition of logs using different imaging modalities (RGB and CT) is indeed feasible, despite the challenging experimental setup.**

***Index Terms*—log tracing, computerized tomography, deep learning, domain adaption**

I. Introduction

Wood log biometrics are a physically marking free approach to establish log traceability, that would be easy as well as cheap to implement. The motivation for wood log tracing is that on the one hand, illegal logging can be better combated and, on the other hand, the identification of each individual log forms a basis for steps towards forest-based industry 4.0. Other tracking technologies for wood logs (e.g. punching, coloring, barcoding log ends and Radio Frequency Identification (RFID)) require the physical marking of each tree which costs time and money. Wood log tracing based on digital log end images only requires to install one camera at the harvesting device (e.g. harvester) and another camera at the sawmill.

The basic assumption is that a single log can be identified based on two different cross-section (CS) images of one log end, the first being captured in the forest during harvesting, the second being captured at the sawmill to facilitate log traceability in the forest–sawmill supply chain. Previous works in this field investigated the general applicability of biometric log identification using digital log end images. These works only applied log tracking between images that were recorded at the same place, at the same time and with the same device, which is a much easier scenario as in a practical setup where logs are recorded in the forest and later once again at the sawmill. For a literature review we refer to [1] and to two findings presented in [2], [3].

About a decade ago, the first fast ($\approx$ 120 m/min) and robust industrial CT-Scanner was developed within the CT-Pro Project in 2010 [4], [5] and a few scanners are already installed in Germany, France and Chile, for example. CT-scanning in combination with an exact log positioning at the saw intake increases the value of each log up to about 20% [6], [7]. CT scanning will become state-of-the-art in the sawmill industry, resulting in corresponding data availability that can be potentially used for wood log tracing. This work elaborates the question if logs can be identified by means of an RGB CS-image captured at the forest site and an CT CS-image captured in the sawmill. Hence, there would be no need to capture RGB CS-images in the sawmill which further saves time and costs.

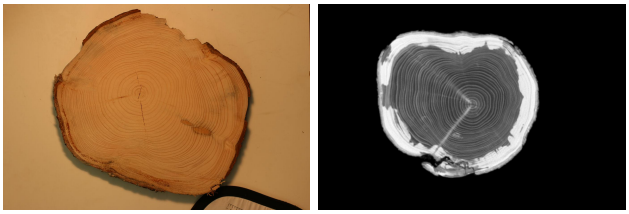
First, all the images are segmented to remove the background of the log images and then domain adaption is applied to bridge the domain shift between CT and RGB images. We apply well known CNN based domain adaption approaches as well as a novel hand-crafted filtering approach. Finally, log recognition is applied by CNNs that were trained with the triplet loss function and a novel shape feature descriptor.

II. The CT and RGB log image dataset

Our two employed datasets consist of images taken from the same 100 logs. The images were acquired from both ends of the logs and show the CS.

The first dataset shows images of discs that were cut off from both ends of the logs. The discs were sanded and recorded indoor using a Canon 70D camera. The camera has been fixed so the images were all taken under the same viewpoint and the same scale, but the discs were differently rotated. Because of the sanded surface, the CS-Images offer

This work is partially funded by the Austrian Science Fund (FWF) under Project Number I 3653.



(a) Sanded (b) CT

Fig. 1. Example images of the same log from both datasets.

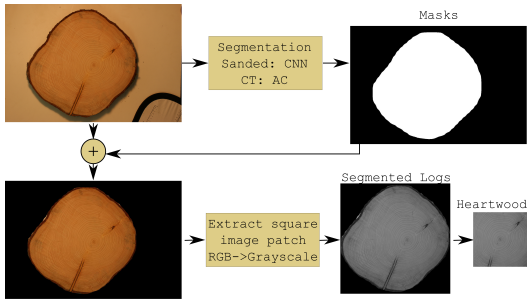


Fig. 2. Segmentation, square image patch extraction (Segmented Logs) and core image (Heartwood Logs) extraction of a log CS image

perfect visibility of the tree ring pattern (see Fig. 1.(a)). We further denote this dataset as Sanded. The Sanded dataset consists of 6-9 images per side of a log with 1321 images in total.

For the second dataset, the 100 logs with the removed discs from both ends of the logs are once again recorded using a CT scanner. About all 4.5 mm along the length of a log a CT image is taken from the CS. We only employ the first and the last 15 images, those that are most close to one of the ends of the log (in that way this dataset is acquired in a similar way as the Sanded dataset). Thus, the CT dataset consists of 3000 images. The CT images of a log all have the exactly same rotation, scale and perspective. Since the images were taken at slightly different longitudinal positions of the log, images of the same log can show different branches and hence there are clear differences even between the images of one side of a log. The CT images offer perfect visibility of the tree ring patter for the heartwood (the inner area of a CS), but hardly any visibility of the tree ring pattern for the sapwood (the outer area of a CS) as can be seen in Fig. 1.(b). So for the recognition of CT images based on RGB images, it may be beneficial to only use the center area of the CS since the outer area of a log in a CT image does not contain any useable information except of the log shape.

III. CS Segmentation

The CS-Images are segmented to remove the background of the log images and to get segmentation masks which can be used for log shape feature descriptors. For the images of the Sanded dataset, we apply the Mask R-CNN framework [8] to get a segmentation mask that separates the CS from the background. As net architecture we employ the ResNet-50 architecture using a model pretrained on the COCO dataset. The segmentation net is trained on the MVA log dataset, a log dataset of 2270 log CS images with manually segmented masks (see [9] for further information on the MVA log dataset). For further information on the CNN-based segmentation see [10]. For the CT images, the CNN-based segmentation does not work because of the lack of manually segmented CT log images. However, the CT images already have a mostly black background outside of the CS and so segmentation is a rather easy task and has been applied using the active contour method [11]. For both CT and RGB images, the background is set to black and the images are reduced to the smallest possible square shaped image section so that the CS is still completely included in the image together with a five pixel thick black border on each side of the image. We furtherly denote the log images with black background as 'Segmented Logs'. Next, we use an algorithm [12] that finds the middle point (pith) MP of a log based on the tree ring pattern to extract a smaller image patch that is centered at MP. If not mentioned otherwise, the side length of the square shaped extracted image patch is half the size of the side length of the Segmented Log image. In that way, most parts of the sapwood (which does not contain any information in CT images) are removed, while the images still contain the perfectly visible tree ring pattern of the heartwood. We further denote those images as 'Heartwood' images. For our experiments, the RGB images are transformed to grayscale to reduce the difference between CT and RGB images. A schematic representation of the segmentation of a log image and the extraction of the Segmented Logs and the Heartwood is presented in Fig. 2.

IV. Domain Adaption

There is a huge difference between the two imaging modalities (i.e. domains) of CT and common RGB recorded images that makes it difficult to directly compare them. To bridge an image domain shift of this extent, typically domain adaptation methods are applied in literature. Domain adaptation using image-to-image translation is independent of the final task to be performed on the images as only the image itself is converted. Image-to-image translation gained popularity during the last years generating highly attractive and realistic output [13], [14]. In this work we apply three different deep learning based domain adaption approaches:

1) Cycle-GAN (cG [13]): Generative adversarial networks (GANs) that make use of the so-called cycle-consistency loss.
2) pix2pix (p2p [14]): Applies conditional adversarial networks as a general-purpose solution to image-to-image translation.
3) Style transfer (ST [15]): Uses neural representations to separate and recombine content and style of arbitrary images.

For all 3 approaches, we used the original code of the authors provided on github. Domain adaption is applied in a 3-

fold cross validation setup. Fig. 3 illustrates a CT and RGB (Sanded) CS-Image of the same log and the domain adapted images using the three domain adaption approaches. As we can observe, the outputs of the domain adaption approaches do not really match the domain shift and include quite a lot of obvious faults (especially p2p). Because of the rather disappointing results of the employed domain adaption approaches we developed a novel filtering approach to bridge the gap between CT and RGB images. The aim of that approach is to display the tree ring pattern, the possibly most distinctive wood log feature, in the same way for both imaging modalities while simultaneously putting branches in the background. Branches complicate the wood log recognition when comparing CS-Images taken at different (longitudinal) positions of the log, since then different wood knots may be visible on the CS-Images.

We employ directional Gaussian 2D filters with zero mean. The size of the filters is 5×5. The standard deviations to build the filters in horizontal (x) and vertical (y) direction are $\sigma_x = 3$ and $\sigma_y = 1$. In that way the Gaussian filters have a clear direction and we can construct filters in 8 different directions (0°, 22.5°, 45°, ... 157.5°). The filters are shown in Fig. 4 (a). To highlight the tree ring pattern, the direction of the filters has to be the same as for the tree rings. In that way the filter responses are high at the position of a tree ring border. For this, we subdivide the log in 16 different sectors, where each sector covers the part of the log within a range of 22.5° using MP as center point. Then each sector is filtered separately with the filter that has the same direction as the respective sector (see Fig. 4(b)). Finally, the filter responses of the 16 different sectors are concatenated to form a filter response image $FR(I)$ of the log image I. Since the direction of wood knots is going from the center MP outwards (see Fig. 1), image regions showing wood knots generate small filter response values and therefore take a back seat in the filter response images. For RGB and CT images, the tree ring borders are shown in different ways (dark lines for RGB images and white lines for CT images). Hence, the filtering has to be applied slightly different for the two imaging modalities, so that the high filter responses marking a tree ring border occur at the same position for CT and RGB images. In case of the CT images, we use the original filter response $FR(I)$ and for the RGB images we use the inverse filter response $-FR(I)$. We furthermore denote the filter response images as 'Filtered'. Fig. 5 shows images of the same log but from different datasets and their filter responses. As we can see, the filtering does produce visually similar outcomes for CT and RGB images.

V. Wood Log Recognition using the CNN Triplet Loss and Shape Features

The big advantage of the triplet loss [16] compared to common loss functions (e.g. SoftMax loss) is that CNNs can be applied to subjects that have not been used during training, which is necessary for any biometric application. The triplet loss requires three input images at once (a so called triplet), where two images belong to the same class (the so called Anchor image and a sample image from the same class, further denoted as Positive) and the third image belongs to a different class (further denoted as Negative). The triplet loss learns the network to minimize the distance between the Anchor and the Positive and maximize the distance between the Anchor and the Negative. The triplet loss using the squared Euclidean distance is defined as follows:

$$L(A,P,N) = \max(||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + \alpha, 0), \quad (1)$$

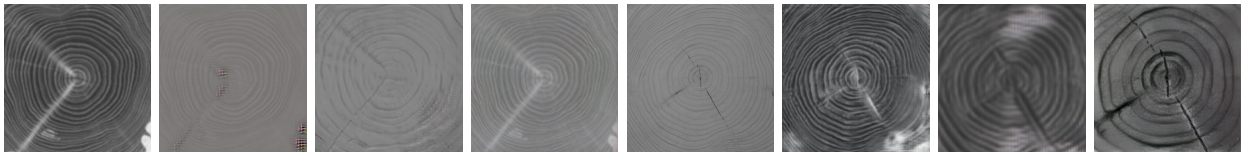
where A is the Anchor, P the Positive and N the Negative. α is a margin that is enforced between positive and negative pairs and is set to $\alpha = 1$. $f(I)$ is an embedding (the CNN output) of an input image I.

Fig. 6 schematically shows the training of a CNN using the triplet loss. To specifically train the net to recognize the log shown in a CT image based on a gallery of Sanded log images, the Anchor is set to be an CT image whereas the Positive and the Negative are set to be images of the Sanded database.

Summarized this means the CNN is trained to create an embedding $f(I)$, such that the Euclidean distances between embeddings of CS-Images from the same class (log) but different image modality is small, whereas the Euclidean distance between embeddings of any pairs of log images from different logs and image modalities is large. We employ hard triplet selection [16] (only those triplets are chosen for training that actively contribute to improving the model) and the Squeeze-Net (SqNet) architecture [17]. SqNet is a small neural networks that is specifically created to have few parameters and only small memory requirements.

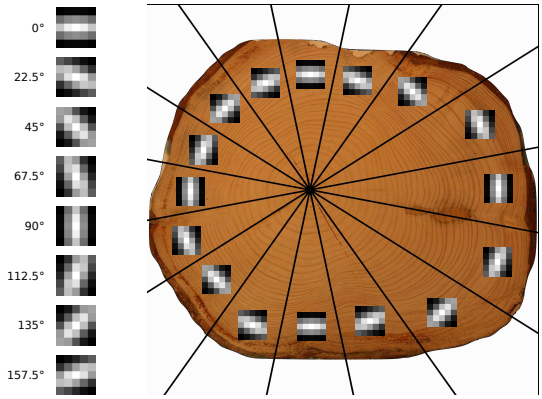
The size of the CNN's last layer convolutional filter is adapted so that a 64-dimensional output vector (embedding) is produced. To make the CNN more invariant to rotations and increase the amount of training data, we employ data augmentation for CNN training. The images are randomly rotated in the range of 0-360 ° and random shifts in horizontal and vertical directions are applied by first resizing the input images to a size of 234×234 and then extracting a patch of size 224×224 at a random position of the resized image (± 10 pixels in each direction). The CNNs are trained for 800 epochs.

Additional to the CNN method, we developed a novel method to describe the shape of the log CS. The advantage of shape features is that the shape of the log is the same for log images of different imaging modalities (aside of slight variations due to different longitudinal positions where the CS-Images were recorded). The shape features are extracted from the binary segmentation masks (see Sec. III). For our shape feature, further denoted as 'LogShape', we use the length of the mayor and minor axis of the ellipse that has the same normalized second central moment as the region of the log from the segmented log mask. As third feature of the LogShape descriptor, we employ the distance between the centroid C of the log and its MP. A schematic representation



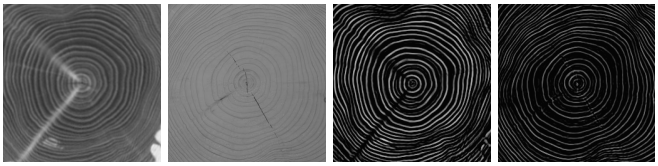
(a) CT (b) CT2RGB cG (c) CT2RGB p2p (d) CT2RGB ST (e) RGB (f) RGB2CT cG (g) RGB2CT p2p (h) RGB2CT ST

Fig. 3. Examplar outcomes of the domain adaption approaches



(a) Filters (b) Filters per sector

Fig. 4. Division of the log in 16 sectors and the associated filters for each sector



(a) CT original (b) Sanded orig. (c) CT FR (d) Sanded FR

Fig. 5. Filter response (FR) images of images from the same log but different datasets.

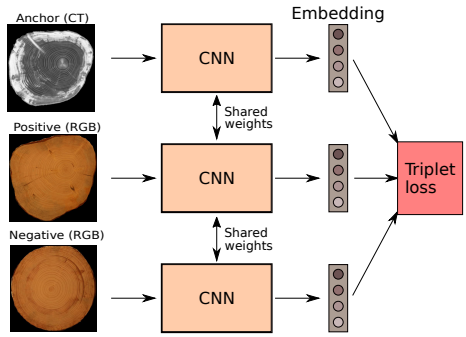


Fig. 6. CNN training using the triplet loss

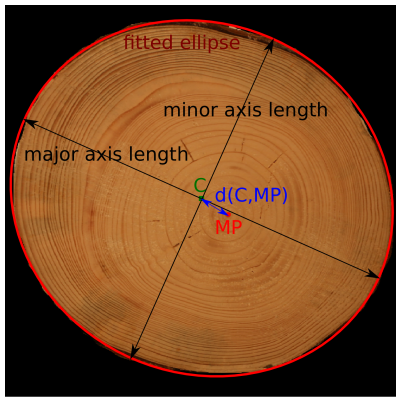


Fig. 7. The three features (minor and mayor axis length, distance $d(C, MP)$ between centroid C and middle point MP) extracted by the LogShape descriptor (best watched in color)

of the three features extacted by the LogShape descriptor is given in Figure 7.

Since the images of the CT and Sanded dataset have different resolutions, we normalize the shape feature vector f of a log l separately on the CT and Sanded dataset: $\|f(l)\| = (f(l) - \overline{f})/\sigma(f)$, where $\overline{f}$ is the mean and $\sigma(f)$ the standard deviation over all feature vectors of a dataset. In that way we balance the different resolutions and can directly compare shape features from both datasets.

In an additional experiment, the shape feature is combined with CNN features by concatenating the feature vectors of both descriptors where the shape features are multiplied by a factor of 5 (the shape descriptor only consists of 3 features per image whereas the CNN feature vector consists of 64 features per image). The shape feature descriptor is combined with the CNN descriptor applied to the Heartwood images (those images contain no information on the shape of a log) and we further denote this combination of features as 'CNN+LogShape'.

We also experimented with other shape features like Zernike moments, which were already utilized in a work about shape features for wood log recognition [18] and performed best in that study. However, Zernike moments do only work when applied to a single dataset but not for our cross-dataset setup recognizing CT images based on RGB images.

VI. Experimental Setup

For the training and evaluation of the CNNs, we employ a 2-fold cross validation. The CNN is trained two times, each time using one fold for training and the other one for evaluation. Each fold consists of 50 out of the 100 logs from both datasets, where all images of a log (no matter which log side or which

dataset) are in the same fold. The evaluation of the LogShape descriptor is also done separately on the same 2 folds (training is not needed for the LogShape descriptor) . Distances between feature vectors of log images are measured using the Euclidean distance d.

As performance measures we employ the equal error rate (EER). The similarity score between two log images is the inverse distance ($1/d$) between the feature vectors of the log images. We only employ the similarity scores between CT and RGB images for the EER computation, but not between images of the same imaging modality. In that way, a CT image of a log can only be identified by the similarity of its feature vector to the feature vectors of the RGB images. We have to consider that the two CNNs (one per fold) per experiment have a different mapping of the images to the CNN output feature space. Thus, CNN feature vectors of different folds cannot be compared in the evaluation and the EER has to be computed separately for each of the two folds. We report the mean EER over the two folds (also for the shape feature).

As already mentioned before, the datasets consist of images from both ends of the logs. Since the logs have a length of about 4 metres, there are no obvious visible similarities between the two ends of a log. To be able to employ the maximum number of images for CNN training, the two sides of a log are considered as different classes thus resulting in 200 classes in total. To avoid any bias by assigning different classes to the two sides of a log, we exclude those triplets during training where the Anchor and the Negative are from the same log but different sides. To avoid any bias in our evaluation, we do not use the similarity scores between images from different sides of the same log for EER computation (for CNN as well as LogShape).

VII. Results

First, we want to find out if it is more beneficial to apply the cross-domain CNN recognition to the Segmented Logs (whole logs) or to the Heartwood (only centers of the logs), without any domain adaptation. It turned out that using the Heartwood (EER=27.0%) performs better than using the Segmented Logs (EER=29.6%).

Next, we want to to find out if domain adaption is able to increase the performance of the CNNs. The domain adaption approaches were applied to Heartwood images, which provided better results in terms of the optical appearence of the domain transformed images as well as the EER rates. In Table I, we present the results for applying CNNs to the domain adapted images as well as the Filtered images (the filter responses (FR) of our proposed filtering approach). The domain adaption approaches either transfer the domain from CT to RGB (RGB in Table I) meaning that the CT images are transfared to RGB and the RGB images are kept the same, or the other way around where the RGB images are transfered to CT (CT in Table I).

As we can see in Table I, our proposed Filtered approach clearly achieves the best recognition rate (EER=20.4%). The domain adaption approaches mostly achieve better results than for employing the original Heartwood images (EER=27.0%), but the results are clearly worse than for using the hand-crafted Filtered approach.

Method	Cycle-GAN		pix2pix		Style tr.		Filtered
Domain	CT	RGB	CT	RGB	CT	RGB	FR
EER	28.0	23.1	25.8	25.3	33.3	23.8	20.4

TABLE I
CNN RESULTS FOR THE THREE DOMAIN ADAPTION APPROACHES AND THE FILTERING APPROACH (FILTERED)

The results using filtered images can be furthermore improved by reducing the Heartwood image patch that is extracted from the Segmented Log image. For a reduction of the side length of the Heartwood image patch down to to 40% of the side length from the Segmented Log image (instead of 50% as in Table I), we achieve a EER rate of 18.4%.

In Table II, we present the results of the LogShape descriptor as well as the combination of the LogShape descriptor with our best performing CNN feature (filtered Heartwood with 40% side length).

Method	CNN	Logshape	CNN+LogShape
EER	18.4	16.9	13.0

TABLE II
RECOGNITION PERFORMANCE (EER IN [%]) OF CNN, LOGSHAPE AND THE COMBINATION OF BOTH FEATURES

We can observe that the proposed hand-crafted shape feature performs better than our best performing CNN. But the clearly best results is achieved by the combination of the two methods with an EER of 13%.

VIII. Conclusion

In this work we investigated if log tracing is feasible in a scenario where RGB images of logs are recorded and the logs are later once again recorded at the sawmill using a CT scanner. We were able to show that the recognition of CT images based on a gallery of RGB images with high quality, well visible tree ring pattern and constant image recording conditions (Sanded dataset) is feasible, despite the huge difference in the imaging modalities of CT and RGB recorded log images. We further showed that hand-crafted approaches (the proposed filtering approach and the proposed shape descriptor) can outperform deep learning approaches in this scenario. The best result (EER =13%) was achieved using a combination of CNN features extracted from filtered images of the heartwood and shape features extracted from the segmentation masks of the logs. In future work, we aim to find out if the same log tracing scenario is also possible using RGB images directly recorded in the forest during harvesting the wood.

References

[1] R. Schraml, J. Charwat-Pessler, K. Entacher, A. Petutschnigg, and A. Uhl, "Roundwood tracking using log end biometrics," in *Proceedings of the Annual GIL Meeting (GIL'2016)*, ser. LNI. Gesellschaft für Informatik, 2016, pp. 189–192.

[2] R. Schraml, H. Hofbauer, A. Petutschnigg, and A. Uhl, "On rotational pre-alignment for tree log end identification using methods inspired by fingerprint and iris recognition," *Machine Vision and Applications*, vol. 27, no. 8, pp. 1289–1298, 2016.

[3] R. Schraml, K. Entacher, A. Petutschnigg, T. Young, and A. Uhl, "Matching score models for hyperspectral range analysis to improve wood log traceability by fingerprint methods," *Mathematics*, vol. 8, no. 7, p. 10, 2020. [Online]. Available: https://www.mdpi.com/2227-7390/8/7/1071

[4] F. Giudiceandrea, E. Ursella, and E. Vicario, "A high speed ct scanner for the sawmill industry," in *Proceedings of the 17th International Nondestructive Testing and Evaluation of Wood Symposium*, Sopron, HU, September 2011.

[5] ——, "From research to market: a high speed ct scanner for the sawmill industry," in *Proceedings XXXI Scuola Annuale di Bioingegneria*, Brixen, ITA, 2012.

[6] A. Berglund, O. Broman, A. Grönlund, and M. Fredriksson, "Improved log rotation using information from a computed tomography scanner," *Computers and Electronics in Agriculture*, vol. 90, pp. 152 – 158, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168169912002347

[7] S. Staengle, F. Bruechert, A. Heikkila, T. Usenius, A. Usenius, and U. Sauter, "Potentially increased sawmill yield from hardwoods using x-ray computed tomography for knot detection," *Annals of Forest Science*, vol. 72, no. 1, pp. 57–65, 2015. [Online]. Available: http://dx.doi.org/10.1007/s13595-014-0385-1

[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[9] R. Schraml, H. Hofbauer, A. Petutschnigg, and A. Uhl, "On rotational pre-alignment for tree log end identification using methods inspired by fingerprint and iris recognition," *Machine Vision and Applications*, vol. 27, no. 8, pp. 1289–1298, 2016.

[10] G. Wimmer, R. Schraml, H. Hofbauer, A. Petutschnigg, and A. Uhl, "Two-stage cnn-based wood log recognition," *CoRR*, 2021.

[11] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[12] R. Schraml and A. Uhl, "Pith estimation on rough log end images using local fourier spectrum analysis," in *Proceedings of the 14th Conference on Computer Graphics and Imaging (CGIM'13)*, Innsbruck, AUT, Feb. 2013.

[13] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976.

[14] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *CoRR*, vol. abs/1703.10593, 2017. [Online]. Available: http://arxiv.org/abs/1703.10593

[15] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *CoRR*, vol. abs/1508.06576, 2015. [Online]. Available: http://arxiv.org/abs/1508.06576

[16] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 815–823.

[17] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size," *CoRR*, vol. abs/1602.07360, 2016. [Online]. Available: http://arxiv.org/abs/1602.07360

[18] R. Schraml, A. Petutschnigg, and A. Uhl, "Validation and reliability of the discriminative power of geometric wood log end features," in *Proceedings of the IEEE International Conference on Image Processing (ICIP'15)*, Quebec, CAN, 2015.