CLASSIFICATION OF TREE SPECIES ON THE BASIS OF TREE BARK TEXTURE

L. Ganschow[1], T. Thiele[1], N. Deckers[2], R. Reulke[2,*]

[1] VINS 3D GmbH, Wegedornstrasse 32, 12524 Berlin, Germany - (lene.ganschow, tom.thiele)@vins3d.de
[2] Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany - (niklas.deckers, ralf.reulke)@hu-berlin.de

KEY WORDS: Convolutional Neural Network, Image Segmentation, Forest Inventory, Pretraining and Finetuning, Integrated Positioning System, Plant Classification

ABSTRACT:

Forest inventory is an important topic in forestry and a digital solution which works on the basis of tree images is looked for. Implementing a system which automatically classifies tree species is the overall goal. In this paper the implementation of a convolutional neural net for solving this classification problem is executed and evaluated. The objective is creating a system which works well on unseen data and deriving guidelines and constraints to guarantee good accuracy results. Images including tree segmentation and the corresponding labels are provided as training data. The tree species classification takes the segmentation results of a stereo vision based image segmentation algorithm as input. The basic idea consists of cropping the tree images into quadratic boxes before feeding them into the neural net. First, each box is classified separately and then the results are evaluated to get a classification for the whole tree. Methods for result improvement include altering box size, using overlapping boxes, artificially enlarging the training set, pretraining and finetuning. Cropping a tree image into boxes of a specific size and accumulating the single results to get a classification of the whole tree leads to an accuracy of 96.7% provided that specific constraints like minimum box number and the projected size of the tree on image plane are considered. Finally, ways to further improve performance are pointed out.

1. INTRODUCTION

To classify objects on an image, models based on machine learning algorithms are created by learning features from training data in order to distinguish between different categories. Object detection can be seen as prior step which needs to be applied before classification as it comprises finding previously defined objects (for example trees) on an image. The task described in this paper is the classification of tree species on the basis of provided data.

In the field of forestry machine learning algorithms may be used for remote sensing, e.g. creating an overview of the tree population in forests by automatically determining parameters like tree height, diameter at breast height or species. The goal is to create a solution which is faster than manual classification.

The start-up VINS 3D GmbH used their own approach to segment trees on images and provided the data for this research.

Previously presented models either classify only on the basis of leaf images or use leaf images and add additional pictures of blossom or stem (Lee et al., 2015) (Lee et al., 2016). Two of the main advantages are that data acquisition is not seasonally limited and that data collection gets easier.

For solving the problem of tree species classification convolutional neural networks (CNNs) are chosen. It has been shown that CNNs perform better on image recognition tasks than any other supervised learning technique (Russakovsky et al., 2015) (Ciregan et al., 2012). Since 2012, all algorithms which got ranked on the top places in the well known image recognition challenge ILSVRC use CNNs. Well-known CNNs are for example AlexNet or VGGNet.

*Corresponding author

As it has already been shown that CNNs work well on other image classification tasks, they could constitute a good approach for classifying images of tree species. The task is comparable to previous plant classification studies. The data used in this presentation differs in terms of limited image resolution, the additional availability of depth information, the possibility to identify individual trees and thereby using several images of a tree captured from different viewing angles and the usage of images which show exclusively tree bark.

2. RELATED WORK

As network architectures are improved rapidly it is not possible to denounce one which always gives the best results (Goodfellow et al., 2016). Therefore, using one of the architectures which gave good results in a renowned image classification challenge (like ILSVRC) which has the goal of solving a similar problem has become common practice.

Looking at existing literature, color is often used as feature for classifying plants. This approach holds some disadvantages as colors can rapidly change due to weather conditions or camera equipment (Yalcin , Razavi, 2016). Therefore, color (alone) does not seem to be a sufficient characteristic for plant classification. However, some illumination changes might be balanced out with appropriate preprocessing steps (Yalcin , Razavi, 2016).

Taking leaves as single classification feature holds some difficulties as well. Deciduous plants could only be classified in summer. Furthermore, (Lee et al., 2015) showed that the shape of a leaf does not work well as feature to classify plants. In another study, (Lee et al., 2016) proposed to use a set of plant parts for classification. These included (amongst others) branch, flower, fruit and stem. Similarly, (Reyes et al., 2015) used a training set of images which depicted either the whole

plant or different parts. An average precision of 48.6% was achieved with this approach, while images of flower and leaf can be classified with a higher accuracy (approximately 65% and 58%) (Reyes et al., 2015). (Lee et al., 2015) confirm with their study that CNN-learned features result in better accuracy values than hand-selected features. Learning features automatically holds the advantage of being able to look at a large number of images in a relatively short time. This feature extraction method results in a smaller error (Reyes et al., 2015).

With pretraining the convolutional neural network (Yalcin , Razavi, 2016), (Lee et al., 2016), (Lee et al., 2015), (Reyes et al., 2015) and (Choi, 2015) were able to improve their classification results. (Lee et al., 2016) pretrained their CNN on the ImageNet dataset and determined a 9.5% higher accuracy score (reaching 71.2%). When the available training set is small, pretraining is often applied to take advantage of learning from a bigger dataset with the goal of learning more robust features (Lee et al., 2015). Futhermore, it has been shown that a CNN which was trained on one specific dataset can also achieve good results on another dataset (Lee et al., 2015).

(Lee et al., 2015) classified plants on the basis of leaf images and used a dataset collected at the Botanic Gardens in Kew, England, comprising 44 different classes for finetuning. Pretraining was performed on ImageNet dataset (Lee et al., 2015). (Choi, 2015) applied finetuning as well. He showed that pretraining on ImageNet and finetuning on the Life CLEF Plant Identification dataset is highly effective.

Using leaves entails the obvious disadvantage that it is impossible to classify deciduous trees in winter as well as trees without any leaves and thereby restricts the application. Furthermore, combining images of several parts of the plant for classification needs much more image material. For this reason, the approach of classifying trees only on the basis of tree bark images is evaluated in this paper. As grayscale images are used for this approach, the features learned by the neural network are not based on colors. Consequently, this approach is not affected by color changes resulting from illumination changes due to variable weather conditions or camera parameters.

It can be expected that these results can easily be improved by applying simple methods like increasing training time (by increasing the number of epochs), including more training data or using one of the CNN architectures employed in ILSVRC.

3. RESEARCH DESIGN

With the approach for segmenting trees introduced by (Thiele, 2015) the input data for this research was prepared. The training images are cropped into quadratic disjoint boxes as shown in figure 1. This ensures on the one hand that the shape of images fed into the network is consistent and does not depend on tree outlines. On the other hand, compared to a classification approach using single pixels as input, the neighborhood and thus potentially important features (lines, curves, general structure) are included and get analyzed. It is assumed that execution time is reduced by using boxes instead of pixels. A finer grained classification can later be obtained by applying overlapping boxes and averaging the results for each pixel.

Each box contains either only pixels that were classified as tree by the segmentation algorithm or only pixels that were

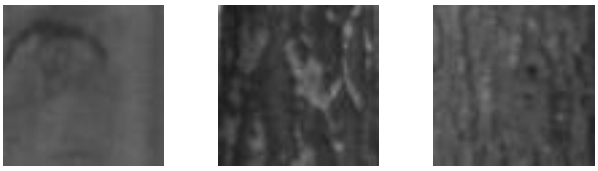


Figure 1. 60x60px boxes of tree bark of different tree species.

classified as background. All other boxes are excluded from the training set. The side length of a box is initially set to 60px, but other sizes are evaluated as well. Undersampling is used to create balanced classes. The prepared boxes are divided up into a training set (90%) and a test set (10%). As the images were taken in a forest, the background images do also show trees, only at a larger distance. This might be a problem when trying to classify trees. Initially, the boxes are not overlapping to ensure that the test data does not contain boxes which the algorithm already saw in the training set. To generalize further, one test setting comprises training on two datasets and testing on a third one. This ensures that no tree from the testing set has already been fed into the neural network. The output of the CNN can either be used with a threshold as a binary classifier or by combining the outputs of several neural networks to get more confidence in the predicted score.

Preparing the training data by manually classifying the tree images into tree species classes was done manually. This data was used for training as well as testing purposes. During the manual classification process some errors in the data, for example two trees which were mistakenly segmented as only one tree, could be eliminated.

The classification process is divided into three steps: data preparation, model training and the actual classification of tree images.

The data is prepared as follows: For each tree the according label (tree species) is retrieved and boxes of a fixed size (similar to the tree/background classification) which show exclusively tree bark are cropped out of the tree images.

The data has to be divided into training, validation and test set. Different net architectures can now be tested against each other. The goal is to create a model with a good performance and F1 score, also on unseen data. Optimization approaches like additional preprocessing, data inspection methods and testing different box sizes are therefore used.

The impact of three-dimensional depth information is investigated by scaling the tree image to a specific distance with respect to the given disparity maps. Each image has a corresponding disparity map which holds the disparity value for each pixel. Knowing the disparity values and camera parameters (baseline, focal length and size of the detector element), the distance between the stereo camera and the depicted object can be calculated.

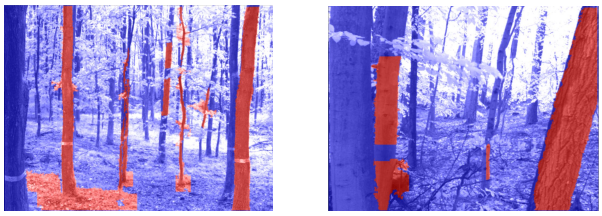
Additionally, finetuning and pretraining approaches are looked at with the goal of further model improvement.

In the end, it should be possible to classify single boxes that show a section of the tree bark by using tree images as input and outputting a classification based on the results of several tree boxes. The tree boxes are derived from several images showing the same tree from different perspectives.

4. DATA

Figure 2. Exemplary greyscale images

Greyscale images as exemplary depicted in figure 2 were used as input images. Figure 3 shows the results of the stereo vision based image segmentation algorithm used by VINS 3D GmbH. This approach produces a binary classification with red areas depicting trees and blue areas showing background. These examples point out some difficulties concerning the tree segmentation task. On image a) it can be seen that large parts of the ground are falsely classified as part of the tree on the left side. Another problem that arises when classifying with this approach is that trees which stand very close to each other are not always recognized as single trees. For example only one of the two trees on the right side of image a) is classified as tree. The trees on the left on images a) and b) do not get recognized. However, the algorithm recognizes also thin trees in the background and the majority of tree stems are segmented in the data sets.



(a) (b)

Figure 3. Results of stereo vision based image segmentation algorithm.

5. RESULTS AND DISCUSSION

The classification of tree species constitutes a complex problem. Therefore a deep CNN is needed. Using a CNN which has only a few layers results in an inability to learn complex features which are necessary to distinguish different tree species.

Changing the image box size results in an accuracy change. It can be observed that the accuracy (as well as precision, recall and F1 score) decreases with smaller box size. This can be explained as there are less structures recognizable in a 10x10px box compared to a 60x60px box.

Checking the quality of the training data is very important for sorting out falsely classified images or images that do not show the expected richness of detail, for example due to overexposure. By improving training data quality the neural net's performance can increase considerably.

The relatively low-resolution grayscale images demonstrate furthermore that it is not necessarily required to provide high quality images for a classification task. CNNs are also able to detect features on images with limited level of detail.

A CNN will not produce good classification results if the majority of the training data comes from only one class. Sampling methods work well to overcome this imbalance problem. In this case, undersampling is used, meaning boxes get randomly removed from a class until all classes contain as many boxes as the class with the former smallest box number.

One approach to overcome the problem of scarce training data is decreasing box size. Cropping the tree images into smaller boxes leads to a twentyfold increase in the number of image boxes and thereby improves the F1 score significantly (from 70% to 83%). The maximum F1 score is achieved with 40x40px boxes. It could be assumed that precision and recall would increase with smaller box size as the training set gets bigger. This is true for box side lengths of 60px, 50px and 40px. However, the F1 score decreases when the image boxes are smaller than 40x40px. Smaller images show less information, reducing the probability that features needed for classification are depicted in the image. In this case, the picture section is often not large enough to classify the image box correctly. Over 10 000 image boxes produced by 221 individual trees are used for training. Based on these results, a box size of 40x40px is henceforth used for upcoming test settings.

The number of available training and test boxes can be further increased by cropping overlapping boxes from the original image. This approach may lead to the problem of overfitting as it is no longer ensured that a part of a tree that is depicted in a test box has not yet been seen in a training box. However, training set size increases dramatically.

Combining the elimination of bad quality image boxes, the specification of 40x40px boxes and the overlapping of boxes, the F1 score increases. The new test setting comprises nine undersampled classes which need to get distinguished. This results in an F1 score of 87% (precision = 87%, recall = 86%). The metrics per tree species class are visualized in table 1.

Class	Precision	Recall	F1 Score	Support
BI	0.99	0.89	0.94	261
DGL	0.88	0.87	0.88	237
EI	0.78	0.92	0.84	198
FI	0.91	0.83	0.87	256
GKI	0.80	0.74	0.77	251
LAS	0.90	0.79	0.84	265
RBU	0.69	0.91	0.78	176
TAS	0.97	0.97	0.97	235
THJ	0.85	0.90	0.87	218
Average / Total	**0.87**	**0.86**	**0.87**	**2097**

Table 1. Precision, recall and F1 score for each tree species regarding the classification on single box level. 40x40px boxes and shift of 10px was applied. "Support" is the number of test image boxes that were classified as the according class.

To check if the learned model is able to generalize, training is performed on two datasets, while a third one provides the image boxes for testing. The results show that some overfitting has occured during training as the average F1 score decreased from 87% to 74%. Precision is now 18 percentage points lower (69%) and recall drops to 80% (from 84%). Overfitting is minimized by using drop out layers. It is nonetheless possible, that the CNN learned to distinguish between tree species on the basis of e.g. lighting conditions which might be different for the various datasets. As most datasets comprise mainly one tree species, a correlation between lighting and tree species cannot be ruled out. Therefore, using techniques to align the data

https://doi.org/10.5194/isprs-archives-XLII-2-W13-1855-2019

coming from the different datasets (e.g. with normalization) should be further looked into to improve results and minimize overfitting.

The usage of depth information (which is provided for every image) might further improve performance. All trees get scaled in such a way that they look like they have been captured from the same distance. Scaling the images could help balancing size differences in tree bark structures resulting from the different distances the photos have been taken at. The goal is to find a distance which is not too far away so that structures are still visible and the trees do not get too thin to crop boxes out of them. On the other hand, the distance should not be too short to avoid getting very blurry images as the scaling factor gets too large. As 40px was found to be the optimal box width, the maximal scaling distance is calculated for every tree (diameter) so that the tree is at least 40px wide on the resulting image. For evaluating the effects of scaling images before training, the CNN is trained on a "normal" (not scaled) subset of the training data and afterwards on the same subset containing the scaled image versions. Precision, recall and F1 score of the testing set without prior scaling are shown in table 2. The results of the CNN applied after scaling are presented in table 3. On average, the CNN performed better with images that were not scaled obtaining an F1 score of 79%, whereas the F1 score produced by the CNN based on scaled images is slightly lower with 76%.

Tree Species	Precision	Recall	F1 Score	Support
BI	0.89	1.00	0.94	25
DGL	0.64	0.75	0.69	24
EI	0.96	1.00	0.98	27
GKI	0.96	0.64	0.77	42
RBU	0.50	0.64	0.56	22
Average / Total	**0.79**	**0.81**	**0.79**	**140**

Table 2. Precision, recall, F1 score and support for CNN classification without image scaling.

Tree Species	Precision	Recall	F1 Score	Support
BI	0.94	0.95	0.95	316
DGL	0.81	0.76	0.78	341
EI	0.80	0.79	0.79	325
GKI	0.64	0.62	0.63	327
RBU	0.59	0.65	0.62	291
Average / Total	**0.76**	**0.75**	**0.76**	**1600**

Table 3. Precision, recall, F1 score and support for CNN classification with image scaling.

By scaling the images the training set could be enlarged by a factor of about ten. A possible reason for the general decrease of the F1 score is that many tree images had to be scaled to appear at a distance of 3.5m, although their original distance was much larger. As the image resolution of the camera is not very good, many image boxes will only show blurred artifacts which make it difficult to recognize certain structures. Without scaling, two trees of the same species might differ a lot due to different structure sizes, but at least the image features are as clear as they can be. With a higher resolution camera, it might be beneficial to try this approach again as much more training data can be generated. Although the overall F1 score decreased after scaling, it is higher for some of the classes. Birch, douglas fir and beech got a slightly higher F1 score after scaling. Most oak images were captured from the same distance (between 7 and 8m). As the tree bark structures are already of a similar size, normalization by scaling should not be necessary in this case. On the contrary, scaling the images would only lead to blurred images which contain less detail than before. Therefore, it is plausible that the F1 score of this class decreases after scaling. Consequently, the image distribution related to the distance should be examined to decide if scaling is beneficial. It has to be considered that the time for creating such a training set of scaled images is much larger than using the unscaled images. This increase of training time leads to the necessity of evaluating advantages and disadvantages of a scaling approach.

Using pretrained network parameters for initialization instead of random values is often used to improve performance. In this experiment, a CNN without pretraining and finetuning is used as baseline and achieves an average F1 score of 65% (60x60px boxes) and 75% (50x50px boxes). Pretraining on ImageNet and subsequently finetuning only the top layer does not increase the F1 score. No matter which box size is used, precision and recall are equal or less than the baseline values. However, fastening the parameters of all layers except the ones of the top layers causes an improvement. The average F1 score reaches 83% when using 60x60px boxes and 81% for 50x50px boxes. This result shows that a dataset which comprises many more and very different classes can still be used for pretraining for a much more specific classification task. As the complexity of learned features increases with the depth of a layer, it can be deduced that the top layers are sufficient to learn the specifics necessary for distinguishing between the given tree species. The preceding layers served as basis by learning to detect simpler features, e.g. edges, which are necessary both for the 1000 classes of the ImageNet dataset as well as for the nine or respectively eleven tree species classes. Another big difference between the two datasets is that the ImageNet dataset consists of colored images, while the tree dataset contains only grayscale images. Concluding, the color channel number of the dataset used for pretraining can be higher than the one of the finetuning dataset and still give good results.

Increasing the number of boxes per tree raises the probability of a correct classification of an individual tree but results in increased runtime. The resulting boxes are fed into the CNN outputting the prediction of a tree species for each box. The class which gets predicted by most boxes is taken as classification result for the whole tree. Results can be improved by applying the constraint that a minimum number of boxes need to predict a specific class to be counted as classification. This results in a possible recommendation of how many boxes a tree should at least consist of for a reliable classification. This recommendation depends on the data and affected by several factors, e.g. image resolution. Therefore, such guidelines are not generally valid for every dataset. Applying the constraint that at least ten boxes need to vote for one class in order to classify the whole tree increases the score to 88.8%. Concluding, the accuracy depends on the number of boxes created and thereby also on the tree distance. Consequently, all trees from this dataset that can be cut into more than 40 image boxes out of which at least half vote for the same class, will be correctly classified with a probability of 96.7%. In order to derive general guidelines, more data needs to be evaluated to get a more robust result and to be able to predict specific guideline values for diameter and maximum distance of a tree.

6. CONCLUSIONS

Although images of tree bark look very similar to an untrained human and resolution and quality of the given images are limited, good accuracy results can be obtained. An F1 score

of almost 97% could be achieved by dividing the tree image into boxes. Nonetheless, some important aspects have to be considered. Preprocessing the data, particularly checking for unusable samples and preparing a uniform and balanced dataset is essential for obtaining useful results. Pretraining the network on a large image set containing many training samples (e.g. ImageNet dataset) and afterwards finetuning the net on the original data can further improve the results.

Applying the trained CNN model to an image of a whole tree, the image box method was shown to be beneficial. Every box is fed into the CNN and produces a classification result. Classifying the tree as member of the class which most boxes predicted presents one possibility to combine all outputs. This approach was shown to be successful when paying attention to the use of a sufficient number of boxes. Further approaches for determining the class of a tree with respect to the box classification need to be researched. An alternative would be using the returned score (between 0 and 1) per class and accumulating this output over all boxes. More training data and prior knowledge which helps to narrow down possible tree species can further improve the accuracy results. However, constraints (e.g. minimal needed number of boxes per tree) need to be validated with a larger amount of training data in order to obtain generalized guidelines.

Another suggestion for improvement is about constantly giving feedback to the CNN in order to learn while new samples are fed into the network. This feedback system could be realized by a subsequent direct evaluation after the CNN has classified a tree.

Finally, an appropriate system consisting of a computing unit with the classification implementation and a camera needs to be developed for being able to classify tree species directly on the go. This system might predict a tree class for every tree which is captured with the camera and also count occurrences of different tree species in order to get a tree species distribution overview for the observed forest part.

Concluding, the experiments show that image classification models based on CNN are powerful and work well for classifying images even when the differences between species are difficult to recognize for a layman. Limited image resolution and no possibility to distinguish classes based on colors do not lead to bad classification accuracy. It will be interesting to determine to what extent the classification results can even be improved either by the use of higher quality image data or by applying one of the other improvement suggestions mentioned before.

REFERENCES

Choi, S., 2015. Plant identification with deep convolutional neural network: Snumedinfo at lifeclef plant identification task 2015. *CLEF*.

Ciregan, D., Meier, U., Schmidhuber, J., 2012. Multi-column deep neural networks for image classification. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3642–3649.

Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. The MIT Press.

Lee, S. H., Chan, C. S., Wilkin, P., Remagnino, P., 2015. Deep-plant: Plant identification with convolutional neural networks. *2015 IEEE International Conference on Image Processing (ICIP)*, 452–456.

Lee, S. H., Chang, Y. L., Chan, C. S., Remagnino, P., 2016. Plant identification system based on a convolutional neural network for the lifeclef 2016 plant classification task. *CLEF*.

Reyes, A. K., Caicedo, J. C., Camargo, J. E., 2015. Fine-tuning deep convolutional networks for plant recognition. *CLEF*.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115, 211-252.

Thiele, T., 2015. Automatic Tree Analysis by Means of Stereo Vision. Master's thesis, Warsaw University of Life Sciences, Eberswalde University of Sustainable Development, Warsaw (Poland), Eberswalde (Germany).

Yalcin, H., Razavi, S., 2016. Plant classification using convolutional neural networks. *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–5.