Hierarchical multi-label Classification of Plant Images using Convolutional Neural Network

Mads Dyrmann, Søren Skovsen, and Rasmus Nyholm Jørgensen
Aarhus University
madsdyrmann@eng.au.dk

Effective weed control using herbicides relies on knowledge about which weeds are present in a field. This knowledge can be obtained automatically by analyzing images collected in the field. Image processing has previously been proposed as a way to solve the task of classifying weeds, [3, 6, 5, 1]. But handling weeds that cannot be be fully annotated remain an unsolved task, since supervised machine-learning methods require annotated data. These weeds are typically small and have not fully developed their visual characteristics. Such samples have previously been left out whereby the classifier performance does not reflect the performance in real-world applications.

However, it might be possible to describe plats at a higher taxonomic level and thereby train them supervised. E.g. "grass" rather than black-grass or "dicot" rather than "knotweed". With that in mind, we need a hierarchiucal multi-label classifier that can be trained on weeds that are only partly annotated. Moreover, a hierarchical classifier is desirable, as when classifying hard samples we would rather have a correct classification at a higher hierarchical level than a wrong classification at the species level. Here we present a convolutional neural network with a hierarchical output structure. This structure makes it possible to train the network using plant images that are not fully annotated and which we would normally have missed if we would only classify at the species level. Likewise, the classifier provides hierarchical outputs for each sample, which allows one to weight a detailed classification at the species level against a higher accuracy accuracy at the family-level or monocot/dicot-level.

Data material The data material consists of RGB images of weeds at early growth stages, collected in 225 Danish fields through five growing seasons using cellphone cameras, consumer cameras or machine-vision cameras. All images are taken vertically towards the ground with a ground sampling distance of 3 to 8 px/mm, which ensures that weeds smaller than a centimetre can still be seen. All plants are at early growth stages, typically between BBCH 11 and BBCH 19 [4]. Eight random samples are shown in Figure 2. A total of 90 697 plant images are used for training and 22 564 images are used for testing.

Although the goal is to classify the species of weeds, there are many weed samples for which it is not possible to annotate the species precisely. To facilitate a hierarchical classification of weeds, the weeds in this study are annotated at one of three levels: monocot/dicot-level, which is above the family-level, which is above the species-level (Figure 1). Thereby a plant will have one to three annotations: One if annotated at the monocot/dicot-level, two if annotated at family-level, and three if annotated at a species level.

The weeds belong to 38 species, which belong to 18 families, which are either monocots or dicots. The data set is likely to contain even more species and families, but since they are not annotated, the total number is unknown.

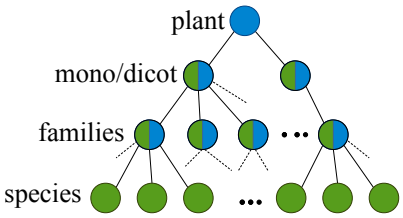

Figure 1: Hierarchical structure of labels. Blue nodes indicate decision nodes/loss nodes, while green nodes are output nodes.

	monocot/dicot	family	species
# of classes	2	18	38
# of fields	225	190	180
# of samples	90 697	43 293	38 423

Table 1: Summary of hierarchical labelling of training data.

Hierarchical classification The neural network is made up of two parts: A backbone network, which is used as a feature extractor, and a hierarchical classifier. The backbone consists of a standard ResNet50 network [2]. The classifier part is made dynamically from the training data; one output layer for each split in the hierarchy resulting in a total of 21 classifier layers – all connected to the backbone to enable rich feature maps. The hierarchical structure is ignored during training to avoid errors at higher hierarchical levels to influence the lower levels. The family level and species level can thereby be trained even though the monocot/dicot-level has not yet settled and therefore outputs the wrong predictions. All samples have one to three annotations, and are therefore only members of up to three of the 21 classifier layers. Therefore, those classifier layers (those coloured blue in Figure 1) to which these annotations belong are activated for a specific sample. All others are deactivated. The structure of the network and the loss aggregation renders possible training with mini-batches of mixed degree of annotation.

Evaluation Since not all paths in the hierarchy have the same numbers of decisions, the results will be skewed towards the branches with the fewest possibilities. Therefore, a path in the tree is chosen based on the the maximum output score of a given classifier layer.

Results and discussion The overall accuracies are found in Table 2, which shows that the accuracy is highest at the monocot/dicot discrimination level, while it is lowest at the species discrimination level.

When comparing the accuracies for plants that are fully annotated with those that are only annotated at the monocot/dicot-level or family level, there is a tendency that plants that can be annotated at species level are easier predictable at all hierarchi-

(a) 40% scale (b) 500% scale (c) 20% scale (d) 300% scale (e) 100% scale (f) 100% scale

Figure 2: Randomly selected samples. Please notice the variation in both scaling and occlusion. The annotations are as followed: a) FUMOF, b–d) PPDDD, e) STEME, f) VIOAR.

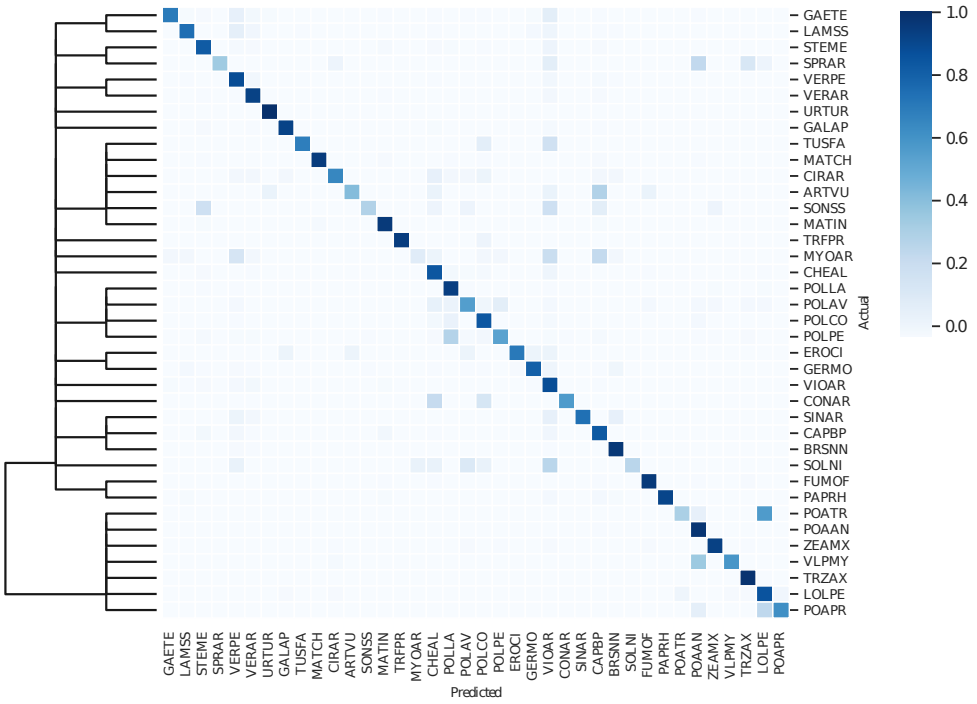

Figure 3: Confusion matrix for images annotated at the species-level. Cell colours illustrate the accuracy, while numbers indicate the number of samples. The tree at the axis indicate the hierarchical relationship

	monocot/dicot	family	species
Accuracy	0.956	0.894	0.877
Accuracy (annotated at species level)	0.992	0.899	0.877
Single classifier accuracy	0.956	0.902	0.971

Table 2: The accuracies at the three hierarchical levels. The accuracy at mono/dicot-level increases slightly when only considering fully-annotated samples, compared to all samples. When classifiers are evaluated independently, assuming the levels above are classified correctly, the accuracies increase to more than 90% for all hierarchical levels.

cal levels than plants, which are only partly annotated.

This is believed to be state-of-the-art when considering that the accuracy is comparable with the study by Dyrmann[1], while we have more than twice the number of samples and weed species.

The confusion matrices at species levels is shown in Figure 3. Interestingly, there is very little confusion within each family. Only in the knotweed-family (1POLF) and between grasses (1GRAF) there seem to be little confusion. Grasses are, however, rarely confused with any dicots at the species level, which is clear from Figure 3 and Table 2.

Conclusion A mix of partially annotated samples can be used simultaneously for training the classifier, which becomes able to classify weeds at three hierarchical levels; monocot/dicot, family, and species. Given the hierarchical structure, it is possible to utilize images that are only partly annotated. Moreover, the hierarchical structure renders possible classification at a higher hierarchical level, with increased accuracy as a result. The training data contains 18 annotated weed families and 38 weed species, which is believed to be state-of-art in terms of the number of weed species in an automated classification. The ability to distinguish weeds with this accuracy provides the basis for big potential savings regarding herbicide consumption in agriculture.

This renders possible training a classifier using all images of weeds, and not only the ones that can be annotated as a specific species. The method is trained and evaluated on more than 100 000 images that span 18 families and 38 species with accuracies ranging from 87% at the species level, up to 95.6% when discriminating monocots from dicots at the highest hierarchical level.

Acknowledgements The work was fonded by the Innovation-Fund, Denmark.

References

[1] M. Dyrmann. *Automatic Detection and Classification of Weed Seedlings under Natural Light Conditions*. PhD thesis, University of Southern Denmark, 2017. 1, 2

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv:1512.03385 [cs]*, Dec 2015. arXiv: 1512.03385. 1

[3] W. A. Kazmi. *Computer Vision based Weed Detection and 3D Plant Imaging*. PhD thesis, Aalborg University, 2014. 1

[4] U. Meier. *Growth stages of mono-and dicotyledonous plants*. Blackwell Wissenschafts-Verlag, 2 edition, 2001. 1

[5] M. Pflanz, H. Nordmeyer, and M. Schirrmann. Weed mapping with uas imagery and a bag of visual words based image classifier. *Remote Sensing*, 10(10):1530, Sep 2018. 1

[6] M. Weis, T. Rumpf, R. Gerhards, and L. Plümer. Comparison of different classification algorithms for weed detection from images based on shape parameters. 69:12, 2009. 1