Towards Confirmable Automated Plant Cover Determination - Supplementary Material

Matthias Körschens[1], Paul Bodesheim[1], Christine Römermann[1,2,3], Solveig Franziska Bucher[1,3], Josephine Ulrich[1], and Joachim Denzler[1,2,3]

[1] Friedrich Schiller University Jena, Jena, Germany
[2] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[3] Michael Stifel Center Jena, Jena, Germany
{matthias.koerschens,paul.bodesheim,christine.roemermann, solveig.franziska.bucher,josephine.ulrich,joachim.denzler}@uni-jena.de

Baseline Comparison

In this supplementary material we provide additional experimental results with two other baselines to put the results of our network into context.

Setup

The first baseline we compare our network to is a constant predictor. This predictor uses the training set in each split and always predicts the average cover percentage in the training set during evaluation over the respective test set.

We also compare our network with a default U-Net as described in the original paper [1]. For a fair comparison, we also show the results of our network with the same training settings as the U-Net, which differ slightly from the results in the main paper.

Here, we use a batch size of 8 instead of 16, and also a learning rate of 0.001 instead of 0.01 as in the main paper. The reason for this is that the large batch size leads to a memory overflow in our 1-GPU setup for the U-Net and the higher learning rate leads to diverging weights of the U-Net. The rest of the parameters as well as the data augmentations are the same as in the main paper.

Experimental Results

Cover Predictions The numerical results of our experiments can be seen in Table 1. We can see that the errors of the constant predictor are much higher compared to those of our model. Hence, the constant predictor can be clearly outperformed because it only achieves a total MAE of 9.88% and MSAE of 0.84.

Using the U-Net as a feature extractor, we can see that the errors are comparable to the ones of our model. In total, our model achieves an MAE of 5.39% and MSAE of 0.51 here, as opposed to 5.54% MAE and 0.52 MSAE for the U-Net. We notice that our extractor network still slightly outperforms the U-Net

Table 1. The mean values and standard deviations of the absolute errors (MAE) and scaled absolute errors (MSAE) for a constant predictor (CP), our network and a U-Net as feature extractors.

Plants	Tri_pra	Pla_lan	Med_lup	Cen_jac	Ach_mil
MAE (CP)	26.44 (± 13.34)	8.55 (± 5.28)	9.62 (± 5.24)	8.99 (± 4.59)	3.37 (± 2.64)
MAE (U-Net)	10.55 (± 10.81)	5.73 (± 5.06)	8.03 (± 7.00)	6.05 (± 5.11)	2.20 (± 2.77)
MAE (Ours)	9.61 (± 10.22)	5.71 (± 5.37)	8.10 (± 7.34)	5.73 (± 4.99)	2.23 (± 2.69)
MSAE (CP)	0.79 (± 0.40)	0.74 (± 0.46)	0.82 (± 0.45)	0.67 (± 0.34)	0.99 (± 0.77)
MSAE (U-Net)	0.32 (± 0.32)	0.50 (± 0.44)	0.69 (± 0.60)	0.45 (± 0.38)	0.64 (± 0.81)
MSAE (Ours)	0.29 (± 0.31)	0.50 (± 0.47)	0.69 (± 0.63)	0.43 (± 0.37)	0.65 (± 0.79)
Plants	Lot_cor	Sco_aut	Grasses	Dead_Litter	
MAE (CP)	4.50 (± 3.55)	3.21 (± 3.30)	9.55 (± 11.53)	14.72 (± 11.41)	
MAE (U-Net)	3.37 (± 3.75)	2.37 (± 3.46)	4.81 (± 7.65)	6.75 (± 9.04)	
MAE (Ours)	3.37 (± 3.74)	2.32 (± 3.40)	4.27 (± 6.34)	7.21 (± 9.17)	
MSAE (CP)	0.70 (± 0.55)	0.85 (± 0.87)	0.93 (± 1.12)	1.04 (± 0.81)	
MSAE (U-Net)	0.53 (± 0.59)	0.63 (± 0.91)	0.47 (± 0.74)	0.48 (± 0.64)	
MSAE (Ours)	0.53 (± 0.58)	0.61 (± 0.90)	0.41 (± 0.62)	0.51 (± 0.65)	

with respect to both error measures. However, it should also be noted that the U-Net model has about 11 times more parameters than our model (34 million vs. 3 million) and in our experiments, the U-Net took about three to four times longer for training compared to our model.

Segmentation Results As the segmentation results with these settings using our model are mostly the same as with the settings in the main paper, we only analyze the segmentations of the U-Net here, which can be seen in Figure 1 and Figure 2.

In general, we notice that the segmentations have a similar quality to the ones that our network returns. However, due to the structure of the U-Net, some parts differ. As the U-Net returns segmentations of the same resolution as the input image, these segmentations also have the potential to be more exact than those of our network. This can be seen especially when focussing on the grass segmentations and *P. lanceolata* in Figure 1. However, the shortcuts in the U-Net, which skip a large part of the network and thus enable such an exact segmentation, can also be detrimental to the results, since only very small local features are included in the identification process. *T. pratense* in Figure 1 shows this effect quite well. We can see that most of these plants are identified as *C. jacea*, as the network appears to simply ignore the "bigger picture". While the segmentations in Figure 2 also show the same problems, they are not as prevalent as in Figure 1 because the plants are much smaller in this image.

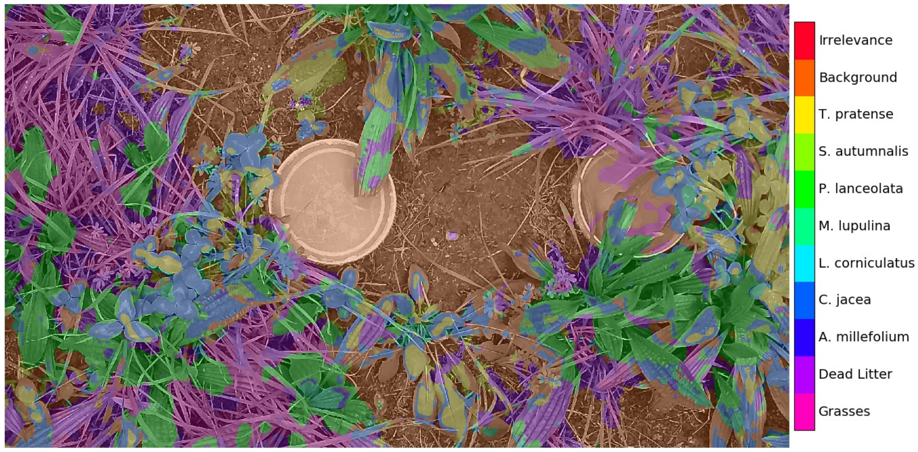

Fig. 1. The segmentations using a U-Net as feature extractor. We can see that the segmentations are more exact compared to the segmentations of our proposed network shown in the main paper (e.g. *P. lanceolata* and grasses). However, since often very local information is used, some segmentations are wrong (e.g. *T. pratense*).

Conclusion

We can conclude that our approach with the usage of our network and the U-Net both outperform the constant predictor baseline by a large margin and our network and the U-Net as feature extractor yield similar results, numerically and visually. It should, however, still be noted that despite similar results, our network has a much smaller number of parameters and much shorter training time. With regard to the segmentations both networks have some advantages and some disadvantages. It might be possible to combine their advantages in a novel network in a future. For now, both networks could be used for solving the task at hand.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

Fig. 2. The segmentations using a U-Net as feature extractor. These segmentations look similar to the ones that our proposed network generates. However, they also suffer from similar problems as discussed in the main text and in the caption of Figure 1.