# The GrassClover Image Dataset for Semantic and Hierarchical Species Understanding in Agriculture

Søren Skovsen[1*] Mads Dyrmann[1] Anders K. Mortensen[2] Morten S. Laursen[1]
René Gislum[2] Jørgen Eriksen[2] Sadaf Farkhani[1] Henrik Karstoft[1]
Rasmus N. Jørgensen[1]

[1]Dept. of Engineering, Aarhus University [2]Dept. of Agroecology, Aarhus University

## Abstract

*GrassClover is a diverse image and biomass dataset collected in an outdoor agricultural setting. The images contain dense populations of grass and clover mixtures with heavy occlusions and occurrences of weeds.*
*Fertilization and treatment of mixed crops depend on the local species composition. Therefore, the overall challenge is related to predicting the species composition in the canopy image and in the biomass. The dataset is collected with three different acquisition systems with ground sampling distances of 4–8 $\mathrm{px\,mm^{-1}}$. The observed mixed crops vary both in setting (field vs plot trial), seed compositions, yield, years since establishment and time of the season. Synthetic training images with pixel-wise hierarchical and instance labels are provided for supervised training. 31 600 unlabeled images are additionally provided for pre-training, semi-supervised training or unsupervised training. Furthermore, this paper provides challenges of semantic segmentation and prediction of the biomass compositions and a baseline model for this dataset.*

## 1. Introduction

Precision agriculture has the potential to revolutionize modern farming by tailoring treatments to match the variations and local properties of the fields. Mapping the local properties, however, relies on data acquisition and reliable analysis. Grass and clover are often grown together as a mixed crop to benefit from niche complements [10] and used as a feed crop in the dairy industry. Research has shown strong economic and environmental benefits by adjusting the amount of applied nitrogen fertilizer based on the local ratio between grass and clover. The GrassClover image dataset is designed to support advancements in this cross-domain of agriculture and computer vision.

*Corresponding author: ssk@eng.au.dk

Unlike traditional computer-vision datasets for semantic segmentation, as *e.g.* Pascal VOC [6], CityScapes [4], and COCO [14], all image classes and most object instances in GrassClover suffer from extreme occlusion. Entangled plants taking up the entire image suggests using a per-pixel classification of the canopy. In addition to image segmentation, the GrassClover dataset provides pairs of images and biomass compositions. The visual canopy composition of plant species can then be used to predict the labeled composition of the dense biomass.

By providing pixel perfect labeled synthetic images ready for training, researchers can compete in both image segmentation and biomass composition prediction. Hierarchically ordered class and instance labels support the development of more advanced methods, while a large number of unlabeled real images allows for self- and unsupervised approaches. The test set consists of a variety of species compositions, growth stages and weed infestations to represent real world variations in grass clover leys in Northern Europe.

Together with the GrassClover dataset, we present two challenges: 1) Pixel-wise classification of image canopies into grasses, red clovers, white clovers, weeds and soil. 2) Predict harvested biomass species compositions using canopy images. Other vision datasets in agriculture exist [3, 7, 9, 13], which target different challenges of detecting plant species and leaves from canopy images. However, to the best knowledge of the authors, no public agriculture datasets combine images of high Ground Sampling Distance (GSD) with both image segmentation and biomass composition prediction.

Bakken *et al.* have previously published an image dataset of grass and white clover swards for analyzing spatial and temporal interactions between the species [2]. 12 288 images were collected with uniform lighting and overlapping coverage. The GSD of 1.6 $\mathrm{px\,mm^{-1}}$ and lossy JPG compression supported the use of morphological operations [1] for pixel-wise classification. However, the absence of textural detail limits modern computer vision approaches. Since none of the images are labelled, they are not suited for supervised training or evaluation.

Skovsen *et al*. have demonstrated the use of simulated grass clover images for training semantic segmentation and predicting the clover percentage in biomass samples [12]. As proof of concept, [11] has extended this on a smaller dataset into hierarchical classification of clover into red and white clover.

| | Image | | | Biomass | |
|---|---|---|---|---|---|
| | Synthetic | Collected | | Basic | Adv. |
| Samples | 8000 | 31 600 | 15 | 435 | 272 |
| *Labels:* | | | | | |
| Class | ✓ | | ✓ | ✓ | ✓ |
| Sub-class | ✓ | | ✓ | | ✓ |
| Parts | ✓ | | | | |
| Instance | ✓ | | | | |

Table 1: Summary of the dataset. The data consists of synthetic images for supervised learning, a large number of unlabeled collected real images, 15 hand annotated images and two sets of images labeled with biomass compositions. Biomass class labels denotes grass, clover, weeds and soil. Sub-class labels add clover species of red clover and white clover.

## 2. Dataset

The GrassClover image dataset is centered around semantic segmentation of high resolution images of heavily occluded vegetation classes in grass clover fields. An overview of the available data is summarized in Table 1. The dataset provides 8000 synthetic high resolution images with pixel-perfect annotations, 15 pixel-wise annotated images of 1 Megapixel (Mpx) and 31 600 unlabeled images collected at five different sites with three acquisition platforms. Of these, 435 images are labeled with the biomass dry matter composition present in the image. The main image classes of the dataset are ryegrass (Lolium perenne), red clover (Trifolium pratense), white clover (Trifolium repens)[1], soil and weeds. In this context, weeds consist of all other species, the majority of which are dandelion, thistle, shepherd's purse, dock and couch grass. The GrassClover image dataset is made publicly available at https://vision.eng.au.dk/grass-clover-dataset.

[1]White and red clovers are visually slightly distinguishable in high resolution images: white clover leaves have toothed edges, while the surface of red clover leaves are covered with small hairs.

| | Nikon d810a | Sony a7 mk1 | IDS UI 3280CP |
|---|---|---|---|
| Resolution | 36 Mpx | 24 Mpx | 5 Mpx |
| 35 mm focal length equiv. | 35 mm | 35 mm | 63 mm |
| Lighting | 2× LED | Ring flash | Ring flash |
| Shutter speed | 1/1250 s | 1/160 s | 1/50 000 s |
| GSD [ $\mathrm{px\,mm^{-1}}$ ] | 4–6 | 4–6 | 5–8 |
| Sampling velocity | $0\,\mathrm{m\,s^{-1}}$ | $0\,\mathrm{m\,s^{-1}}$ | $5\,\mathrm{m\,s^{-1}}$ |

Table 2: Specification and comparison of the three image acquisition platforms.

### 2.1. Data specifications

The data is aimed at providing sufficiently high quality to avoid it being the limiting factor in later analyses. All collected images from the five test sites were captured with external lighting at minimum $4\,\mathrm{px\,mm^{-1}}$ and stored in lossless compressed raw files. The images are made available in the original format, 16 bit bayer-PNG or vendor-specific raw format depending on the imaging platform, and as JPG images developed from the raw formats.

### 2.2. Image acquisition platforms

The dataset has been collected with the three acquisition platforms based on the 36 Mpx Nikon d810a (Sigma 35 mm F1.4 DG HSM), 24 Mpx Sony a7 mk1 (Sony FE 35 mm F2.8 ZA) and 5 Mpx IDS UI-3280CP (Edmund Optics 16 mm F1.8 #86-571). The platforms are depicted in Figure 1a–1c. The cameras were set up to capture images with a canopy sampling distance of 4–8 pixel $\mathrm{mm^{-1}}$ depending on the height of the canopy and camera. A sample image from each platform, cropped to $400 \times 300$ pixels is presented in Figure 1d–1f. While the first two platforms were designed for still pictures in plot trials, the third was designed for large capacity image acquisition in fields. The number of collected images for each platform is shown in Figure 2. The camera systems are presented in detail in Table 2.

### 2.3. Image acquisition sites

The images have been collected in plot trials at two experimental sites and in the fields of three dairy farms in Denmark. All images collected with the IDS UI 3280CP camera originate from commercial grass clover fields. All biomass samples originate from outdoor plot trials located at Aarhus University, Foulum, Denmark, world geodetic system 1984 (WGS84: 56°29′43″N, 9°34′10″E) and at DLF Seed & Science grass clover breeding facility in Stevns, Denmark

(WGS84: 55°20′21″N, 12°23′08″E).

### 2.4. Synthetic training images

Based on previous work in the synthetic agricultural images [5, 12, 11], the synthetic images of grass clover mixtures were generated as:

1. Crop out several examples of every plant species and plant parts to be classified (Figure 3).
2. Randomly select a soil background image.
3. Loop until a preset leaf area index is reached:
    (a) Sample species example with a sampled class probability.
    (b) Randomly rotate and scale (70-120%) sample.
    (c) Add artificial shadow using a Gaussian filter on the sample mask.
    (d) Add sample with shadow to the background image.
    (e) Update class labels of the label image.
    (f) Update instance labels identifying each plant sample.

To allow for both between and within class weighing, and for hierarchical training, the synthetic images follow the hierarchical structure illustrated in Figure 4.

An example of a synthetic grass clover image is shown in Figure 5 with a visual representation of its class, sub-class and instance labels.

### 2.5. Unlabeled images

To supplement the synthesized grass clover images, 31 600 sampled images are made available. These can help minimizing the gap from synthesized images to real images using unsupervised or self-supervised approaches. As shown in Figure 2, the image collection started in 2017, then increased by a factor of 100 in 2018, due to the ATV-mounted IDS UI 3280CP.

### 2.6. Pixel-wise labeled images

Based on the observed image content across the collected images, six visual classes were defined: red clover, white clover, grass, weeds, soil and unknown. When the clover species could not be visually identified, they were annotated as clover. Since hand-labeling the heavily occluded images took 5-8 hours per Mpx, a subset of 15 image crops was annotated. 10 of the images were from plot trials of high biomass with varied clover content. 5 originated from sparser vegetation in October with varied clover content.

### 2.7. Biomass labeled images

Each biomass sample consists of: 1) A canopy image of a defined 0.5 m $\times$ 0.5 m of grass clover preceding the cut. 2) A composition of the harvested biomass with stems located in the square. An example is shown in Figure 6. After cutting the plants at a height of 5 cm, all plant samples were

(a) Nikon d810a. (b) Sony a7 mk1. (c) IDS UI 3280CP.

(d) Sample crop of (a). (e) Sample crop of (b). (f) Sample crop of (c).

Figure 1: The three image acquisition platforms. (a-c) illustrate the platforms in use. (d-f) show an equally sized image crop from the platforms above.

separated into ryegrass, clover and weeds. 272 of the samples had an extended sub-class separation of clovers into red clover and white clover. After drying the samples, each fraction was weighed individually to determine the dry matter yield and composition. 435 biomass samples were collected at two sites in Denmark as summarized in Figure 7. The acquisition spans the season of 2017, with additional samples in 2018. A histogram of the biomass distributions in the samples is shown in Figure 8.

### 2.8. Dataset splits

The data is separated into a training set and a test set. The training set consists of:

- 8 000 synthesized images with hierarchically ordered pixel-wise labels of classes, sub-classes and parts, and additional instance labels.
- 31 600 unlabeled images collected with three camera platforms.
- 152 randomly selected biomass labels and corresponding cropped images to learn a mapping from image content to biomass composition.

The test set consists of the 15 hand-labeled images and the 283 remaining biomass sample pairs. The labels, however, are kept for evaluation.

## 3. Hierarchical Semantic Segmentation

The task involves pixel-wise classification of grass clover images into five categories: grass, white clover, red clover, weeds and soil.

### 3.1. Tasks and metrics

Using a common approach for the entire test set, each pixel is to be classified into either of the five categories.

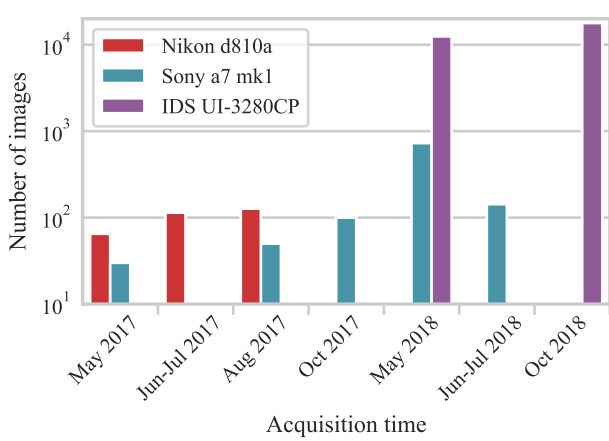



Figure 2: Number of collected images grouped by camera and date.

The class-wise Intersection over Union (IoU) metric of [8] is used to evaluate the performance of each class. The mean IoU is used to assess the overall segmentation performance.

$$\text{class } i \text{ IoU}: \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{1}$$

$$\text{mean IoU}: \frac{1}{N_{cl}} \frac{\sum_i n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{2}$$

where $n_{ij}$ is the number of pixels of class $i$ predicted as class $j$. $N_{cl}$ is the number of classes and $t_i = \sum_j n_{ij}$ is the total number of pixels of class i. Pixels labeled as unknown by the ground truth annotator will be disregarded when calculating the IoU. Clovers without species labels will be disregarded when calculating the clover species specific IoU.

### 3.2. Baseline

A hierarchical set of two FCN-8s [8] models was trained on 1720 synthetic images with sub-class labels to perform pixel-wise classification as described in [11]. The first FCN was trained to recognize grass, clover, weeds and soil. The second FCN was trained to differentiate between white clover and red clover, given a clover in the pixel. Predicted clover pixels by the first network were re-classified to the clover species predicted by the second network in the corresponding pixels. The classification output of each class was weighted after training to avoid training data biases. A sample classification output of the first FCN before softmax normalization is illustrated in Figure 9a. The final segmentation into grass, white clover, red clover, weeds and soil is shown in Figure 9b. The IoU performance of the baseline model is presented in Table 3.

| Intersection over Union | | | | | |
|---|---|---|---|---|---|
| Mean | Grass | White clover | Red clover | Weeds | Soil |
| 55.0% | 64.6% | 59.5% | 72.6% | 39.1% | 39.0% |

Table 3: Mean and per class Intersection over Union for semantic segmentation on the test set. The baseline result is provided by the two hierarchically trained FCN-8s models [11].

## 4. Biomass Composition Estimation

Targeted nitrogen fertilization of grass clover relies directly on the fraction of clover in the biomass. However, acquired image data in a grass clover field has the potential to provide much more information. The weed infestation is *e.g.* a useful metric in organic grass clover fields when planning crop rotations. The distribution of clover species alters the fertilization strategy, though less than the combined clover fraction itself. Occlusion in semantic segmentation increases the difficulty of classifying each pixel, yet every pixel to be classified is visible. Occlusion in biomass composition prediction, however, necessitates predicting the relative mass of each class, based only on the canopy view.

### 4.1. Tasks and metrics

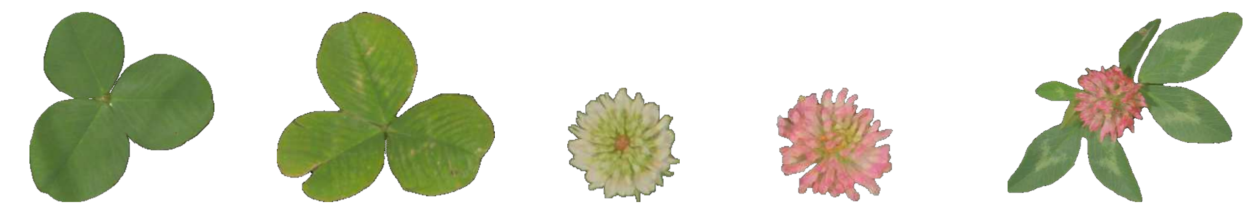
Predict the per sample distribution of the biomass classes: grass, clover, weeds, white clover and red clover. The evaluation of the prediction performance is based on the root mean square error (RMSE) and mean absolute error (MAE) of each biomass category prediction.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - t_n)^2} \tag{3}$$

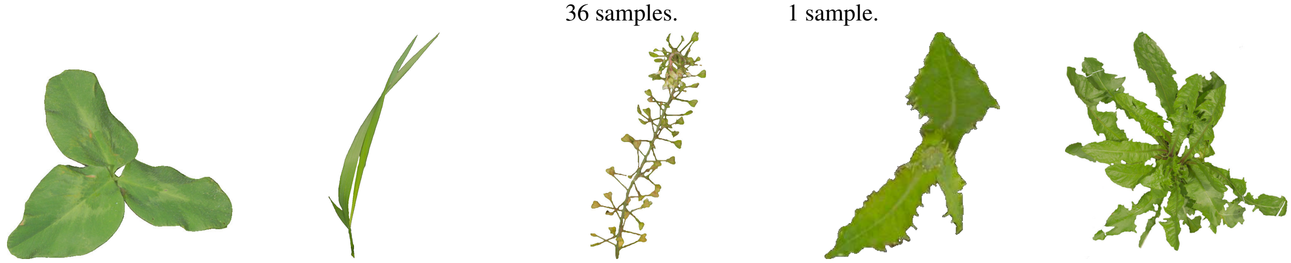
$$\mathrm{MAE} = \frac{1}{N}\sum_{n=1}^{N}|y_n - t_n| \tag{4}$$

where $N$ is the number of samples, $y$ is the predicted biomass fraction size and $t$ is the true biomass fraction size. All biomass samples will be used for evaluating the grass, clover and weeds fraction. Only biomass samples with advance labels can be used for evaluating the two clover species[2].

[2]The first seasonal cut of 2017 in Foulum is disregarded when predicting the clover species as the canopy height exceeded the height of white clovers, occluding them entirely.



(a) Unknown clover leaf. 43 samples.
(b) White clover leaf. 37 samples.
(c) White clover flower. 36 samples.
(d) Red clover flower. 1 sample.
(e) Red clover. 9 samples.



(f) Red clover leaf. 23 samples.
(g) Grass. 55 samples.
(h) Shepherd's purse. 4 samples.
(i) Thistle. 6 samples.
(j) Dandelion. 16 samples.

Figure 3: Illustration of the categories of plant classes used for generating synthetic images and the number of samples in each category. The plant cut-outs used for generating the synthetic images are released with the dataset.

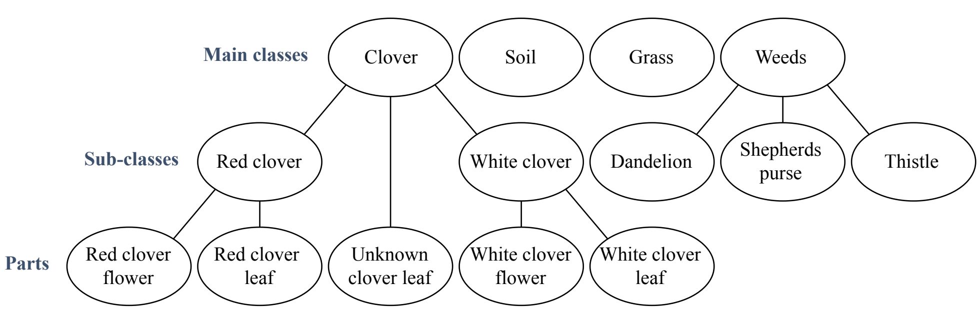
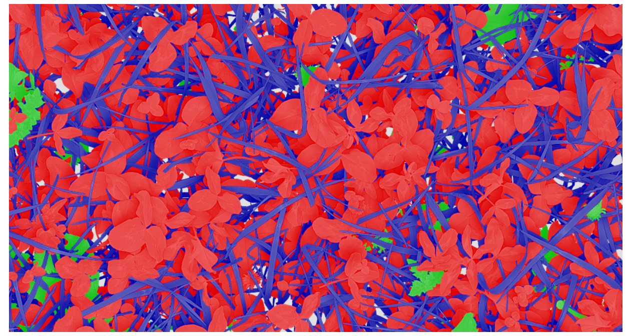
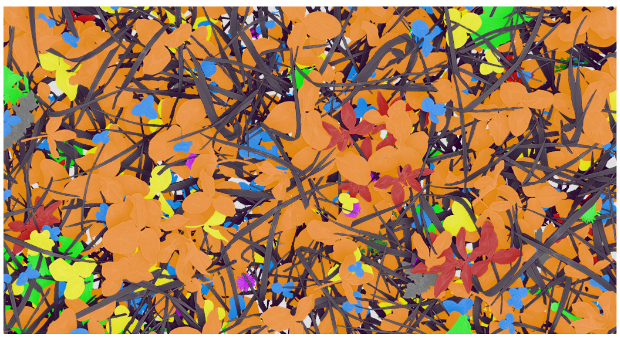



Figure 4: Hierarchical structure of the synthetic images. Every pixel in the synthetic image labels is labeled at the lowest hierarchical level of the corresponding plant cut-out.

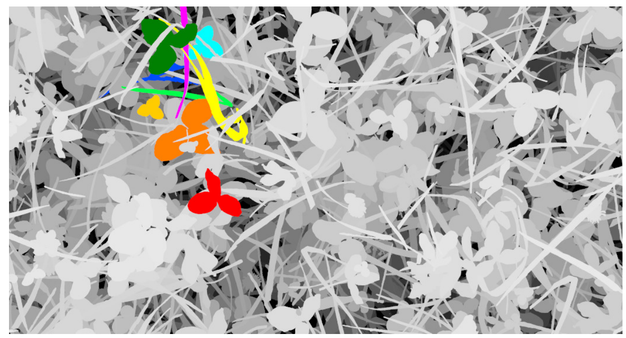
(a) Synthetic training image.



(b) Semantic segmentation label combined into main classes of (a). Red is clover, blue is grass, green is weeds, and light gray is soil.



(c) Semantic segmentation label for sub-classes of (a). Red is red clover, orange is red clover leaf, yellow is unknown clover leaf, blue is white clover leaf, purple is white clover flower, gray is thistle, dark gray is grass, light green is dandelion, dark green is shepherd's purse and light gray is soil.



(d) Instance segmentation label of individual plant samples of (a). Each plant takes a unique integer value. Nine plants have been highlighted with colors for visualization.

Figure 5: Example of a synthetic image and corresponding labels.

### 4.2. Baseline

Reusing the hierarchy of FCN-8s models from task one, each biomass-labeled image was segmented into vegetation classes, as illustrated in Figure 9b. The number of pixels classified into each category was summarized for each image. Based on the 261 training samples, a first order linear model was fitted for each category, to naïvely convert the visual fraction into the biomass fraction. This is demonstrated for white clover, red clover, grass and weeds in Figure 10. The baseline performance on the test set is presented in Table 4.

| | Grass | Clover | Weeds | White clover | Red clover |
|---|---|---|---|---|---|
| RMSE [%] | 9.05 | 9.91 | 6.50 | 9.51 | 6.68 |
| MAE [%] | 6.85 | 7.82 | 4.65 | 7.62 | 4.87 |

Table 4: Baseline class prediction errors on the biomass composition test set. The evaluation metrics are root mean square error and mean absolute error.

## 5. Discussion and Conclusion

In this work we present the GrassClover image dataset, specifically designed to support advancements in robust image analysis of heavily occluded mixed crops. 15 pixel-wise labeled images and a set of 435 images with biomass labels are used to evaluate the image analysis quality for real world applications. Using the synthetic hierarchical images, a baseline FCN-8s model is trained for semantic segmentation and made publicly available with the dataset. The baseline model provides coarse semantic segmentations reaching a mean IoU of 55.0%, and presents a naïve approach for estimating biomass compositions. Large sets of unlabeled images and instance labeled synthetic images are provided to motivate novel approaches to improve the current state of

(a) Original image with a defined square of 0.5 m $\times$ 0.5 m.

(b) Cropped image sample with a biomass label.

Figure 6: Example of a collected image and an image crop labeled with biomass content: 13.96 g grass, 35.67 g white clover, 5.40 g red clover, and 1.73 g weeds. The image is cropped without further processing to maintain image quality. The image crop is not necessarily square due to the pose of camera and frame.

the art in grass clover image analysis. Shortcomings of the baseline model lie mainly in the lack of recognizing weeds and soil across experimental sites. Possible improvements supported by the GrassClover dataset include the use of: 1) instance masks for improved boundary identification 2) hierarchically structured labels for training semantic segmentation 3) unlabeled collected images to minimize the feature space distance between real and artificially generated image classes 4) novel approaches for linking the canopy images to biomass compositions.

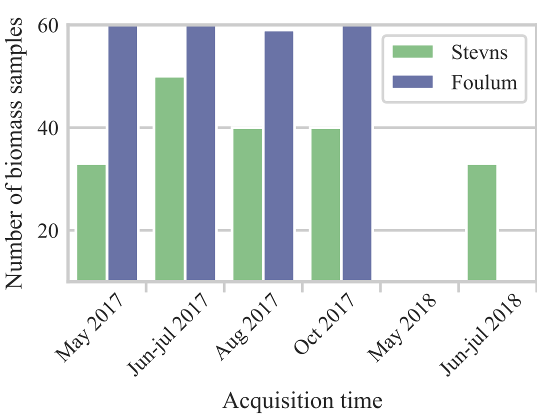



Figure 7: Biomass sample acquisition grouped by experimental site and time of season.

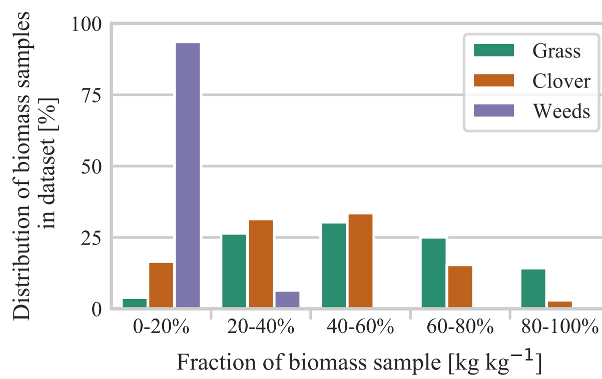


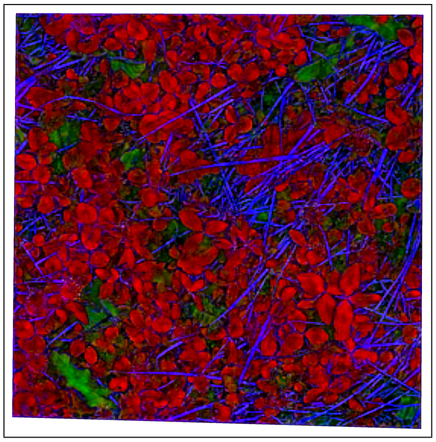

Figure 8: Distributions of the biomass groups in the dataset. *e.g.* 94% of the biomass samples in the dataset consists of less than 20% weeds.

## Acknowledgements

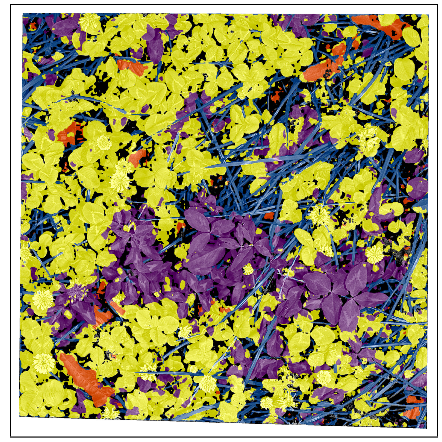
The work was funded by Green Development and Demonstration Programme (GUDP) under the Danish Ministry for Food, Agriculture and Fisheries, and Innovation Fund Denmark. The Nikon d810a acquisition platform was kindly provided by AgroTech, Teknologisk Institut. Access to plot trials at Stevns, Denmark was kindly provided by DLF Seed & Science.

## References

[1] H. B. *, K. Kaspersen, and A. K. Bakken. Evaluating an image analysis system for mapping white clover pastures. *Acta Agriculturae Scandinavica, Section B — Soil & Plant*

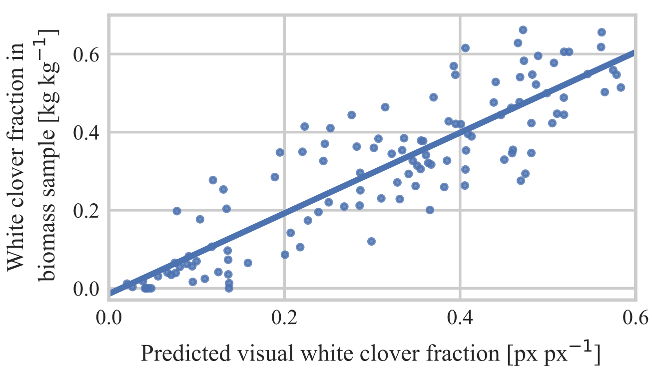


(a) Initial FCN classification output.

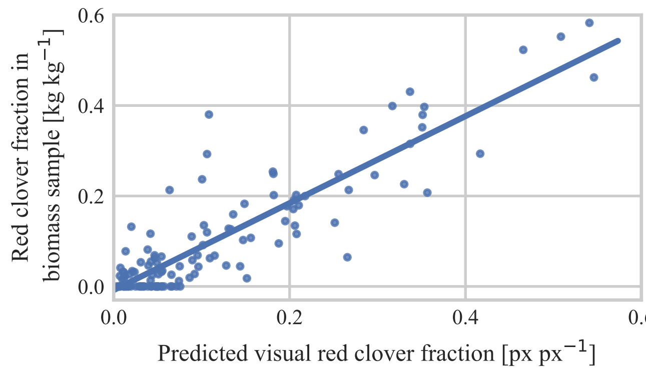


(b) Hierarchical semantic segmentation output.

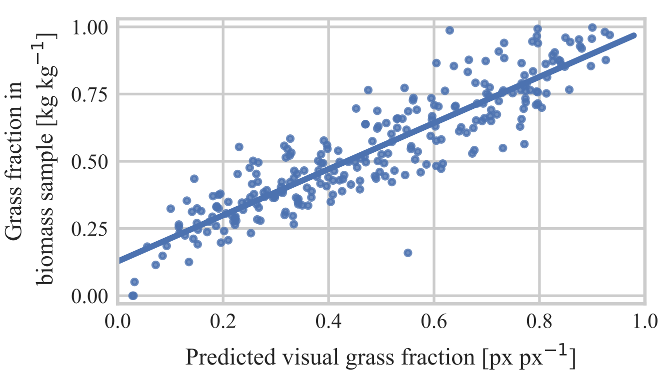
Figure 9: Baseline image analysis example of Figure 6b. a) shows the classification output before normalization. Increased intensity indicates a stronger confidence in the classification. Blue is grass, red is clover and weeds are green. b) Final hierarchical semantic segmentation using the baseline FCN models. Blue is grass, yellow is white clover, purple is red clover and orange is weeds.



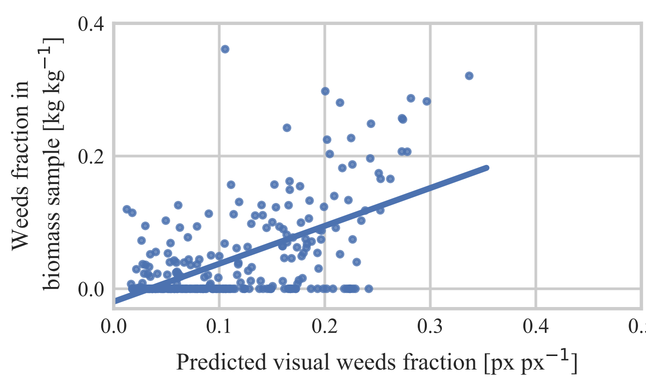

(a) White clover prediction.




(b) Red clover prediction.




(c) Grass prediction.




(d) Weeds prediction.

Figure 10: First order linear models on training samples for predicting the biomass compositions from predicted vegetation classes in the image.

*Science*, 54(2):76–82, 2004. 1

[2] A. K. Bakken, H. Bonesmo, and B. Pedersen. Spatial and temporal abundance of interacting populations of white clover and grass species as assessed by image analyses. *Dataset Papers in Science*, 2015, 2015. 1

[3] N. Chebrolu, P. Lottes, A. Schaefer, W. Winterhalter, W. Burgard, and C. Stachniss. Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. *The International Journal of Robotics Research*, 2017. 1

[4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[5] M. Dyrmann, A. Mortensen, H. Midtiby, and R. Nyholm Jørgensen. Pixel-wise classification of weeds and crops in images by using a fully convolutional neural network. 2016. Paper No. 608, CIGR 2016; Conference date: 26–29 June 2016. 3

[6] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, Jan 2015. 1

[7] T. M. Giselsson, M. Dyrmann, R. N. Jørgensen, P. K. Jensen, and H. S. Midtiby. A Public Image Database for Benchmark of Plant Seedling Classification Algorithms. *arXiv preprint*, 2017. 1

[8] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 4

[9] M. Minervini, A. Fischbach, H. Scharr, and S. A. Tsaftaris. Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters*, pages –, 2015. 1

[10] D. Nyfeler, O. Huguenin-Elie, M. Suter, E. Frossard, J. Connolly, and A. Lüscher. Strong mixture effects among four species in fertilized agricultural grassland led to persistent and consistent transgressive overyielding. *Journal of Applied Ecology*, 46(3):683–691, 2009. 1

[11] S. Skovsen, M. Dyrmann, J. Eriksen, R. Gislum, H. Karstoft, and R. N. Jørgensen. Predicting Dry Matter Composition of Grass Clover Leys Using Data Simulation and Camera-Based Segmentation of Field Canopies into White Clover, Red Clover, Grass and Weeds. *14th International Conference on Precision Agriculture*, 2018. 2, 3, 4

[12] S. Skovsen, M. Dyrmann, A. Mortensen, K. Steen, O. Green, J. Eriksen, R. Gislum, R. Nyholm Jørgensen, and H. Karstoft. Estimation of the botanical composition of clover-grass leys from rgb images using data simulation and fully convolutional neural networks. *Sensors*, 17(12), 12 2017. Special Issue: Sensors in Agriculture. 2, 3

[13] N. Teimouri, M. Dyrmann, P. Nielsen, S. Mathiassen, G. Somerville, and R. Jørgensen. Weed growth stage estimator using deep convolutional neural networks. *Sensors*, 18(5):1580, May 2018. 1

[14] T. yi Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, and C. L. Zitnick. Microsoft coco: Common objects in context, 2014. 1