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

Søren Skovsen[1], Mads Dyrmann[1], Jørgen Eriksen[2], René Gislum[2], Henrik Karstoft[1], Rasmus N. Jørgensen[1]

[1]Department of Engineering, Aarhus University, Denmark.
[2]Department of Agroecology, Aarhus University, Denmark

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *Targeted fertilization of grass clover leys shows high financial and environmental potentials leading to higher yields of increased quality, while reducing nitrate leaching. To realize the gains, an accurate fertilization map is required, which is closely related to the local composition of plant species in the biomass. In our setup, we utilize a top-down canopy view of the grass clover ley to estimate the composition of the vegetation, and predict the composition of the dry matter of the forage. Using a deep learning approach, the canopy image is automatically pixel wise segmented and classified into white clover, red clover, grass and weeds. While robust grass and clover segmentation has proven to be a difficult task to automate, red and white clover discrimination in images is challenging, even for human experts, due to many visual similarities between the two clover species. Using high-resolution color images with a ground sampling distance of 4 to 6 pixels per mm and data simulation of hierarchical labels, a cascaded convolutional neural network was trained for segmentation and classification. Clover, grass and weeds was automatically segmented and classified with a pixel wise accuracy of 87.3 percent, while red clovers and white clovers could be distinguished automatically with 89.6 percent accuracy. Utilizing the image analysis on 179 images of mixed crop plots of ryegrass, white clover and red clover, demonstrated a linear correlation between the detected clover and clover species fractions in the canopy, and the corresponding compositions in harvested dry matter.*

Keywords. *Precision Agriculture, Targeted Fertilization, Grass-Clover, Deep Learning, Semantic Segmentation, Convolutional Neural Network, Data Simulation, Plant Classification.*

The authors are solely responsible for the content of this paper, which is not a refereed publication. Suggested citation: Cook, S., Lacoste, M., Evans, F., Ridout, M., Gibberd, M. & Oberthür, T. (2018). An On-Farm Experimental philosophy for farmer-centric digital innovation. In Proceedings of the 14th International Conference on Precision Agriculture. Montreal, CA: International Society of Precision Agriculture.

Introduction

Clover and grass are often grown in mixtures, as clover-grass leys with different species increase the yield stability (Eriksen et al. 2014; Frankow-Lindberg et al. 2009) and herbage quality (Phelan et al. 2015) compared to fertilized grass-only leys.

This is due to niche complementarity (Nyfeler et al. 2009) and the greater protein content of the clover (Egaard et al. 2009; Suter et al. 2014). Thus, properly managed clover-grass mixtures produce a greater yield compared to pure stands of grass and clover (Kirwan et al. 2007; Nyfeler et al. 2011).

Assessment of the grass to clover ratio is essential for optimal fertilization, as it has significant economic potential for the dairy industry. Grass-clover combines high productivity and low environmental impact if the nitrogen (N) supply is adjusted according to the clover content. Manure and fertilizers are used largely irrespective of clover content due to poor ability to assess clover content and lack of knowledge on the correlation between fertilizer (N) and clover content during the season.

It has been demonstrated that due to complementarity, inclusion of red clover in the traditional widely used ryegrass-white clover mixtures has advantages in terms of yield and increased clover content at a given fertilizer rate (Eriksen et al. 2014). However, the persistence of red clover is poorer than that of white clover, and thus monitoring of the species identity of the clovers in the three-species mixture is of huge importance.

Researches have previously attempted to estimate the forage composition of grass-clover using image analysis (Bonesmo et al. 2004; Himstedt et al. 2009; Himstedt et al. 2010; Himstedt et al. 2012; McRoberts et al. 2016; Mortensen et al. 2017) with limited robustness for autonomous mapping in agriculture. Using deep learning approaches for pixel wise vegetation classification in agriculture shows great promises (Mortensen et al. 2016, Dyrmann et al. 2016). Skovsen et al. (2017) demonstrated the use of convolutional neural networks (CNN), trained on simulated data, to perform robust segmentation of grass-clover canopies into grass, clovers, and weeds.

Continuing the work of Skovsen et al. (2017), we propose to increase the gained knowledge from the canopy analysis by distinguishing the clovers species at pixel level. Using canopy images of mixed crops of ryegrass, white clover and red clover, an additional CNN is trained to distinguish the two clover species. Combining the two CNNs in a cascading structure, we propose an improved pixel wise segmentation into grass, white clover, red clover and weeds.

Material and Methods

The proposed dry matter prediction relies on camera-based segmentation of field canopies. To evaluate each step in the process, pairs of grass-clover canopy images and corresponding dry matter analysis must be collected. While the segmentation task is evaluated only on manually pixel wise labeled test images, the dry matter prediction couples automated image segmentations to corresponding dry matter samples.

Collected Material

179 grass-clover field trial samples were photographed, harvested and separated into fractions of ryegrass, white clover, red clover and weeds, dried and weighed. The samples were collected in 2017 during the first three of the four seasonal cuts at an experimental site of 60 grass-clover plots at Aarhus University, Foulum, Denmark. The plots ranged from 1 to 4 years of age. Each sample measured 0.5 x 0.5 m, with yields ranging from 0.5 to 7.5 1000 kg ha^{-1} and clover fractions from 0 to 0.83 kg kg^{-1} dm, corresponding to clover fractions ranging from 0 to 83 percent. The plots were photographed top-down into the canopy using a Nikon D810A (Tokyo, Japan) camera with two external LED flashes. Achieving a ground sampling distance of 4 to 6 pixels per millimeter, the acquired sample images range from 2000 to 2500 pixels. Detailed information of the field trial samples and acquisition is published by Skovsen et al (2017).

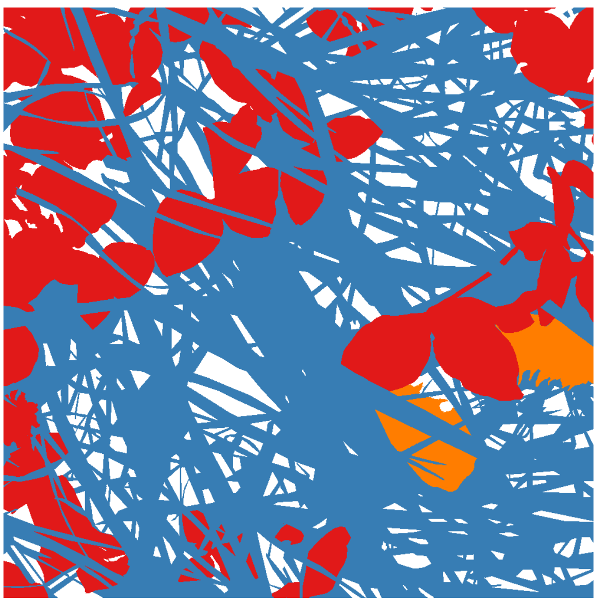
Image Labeling

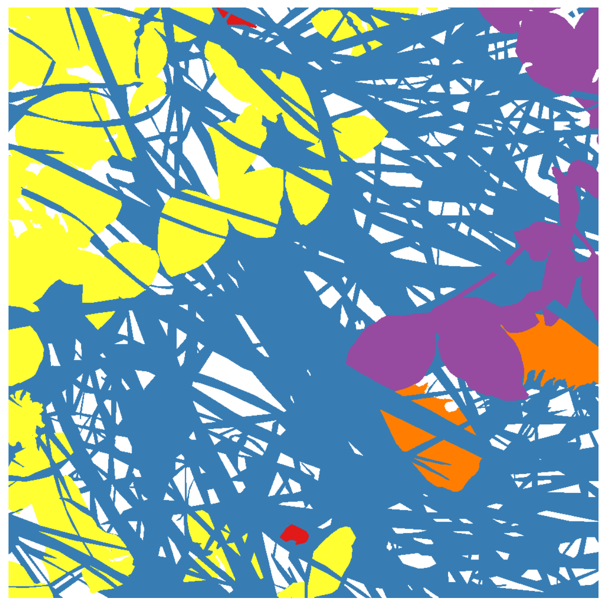
Labeled images are necessary to evaluate the performance of the image segmentation on ground truth data. A previous study in camera based semantic segmentation of grass-clover leys has hand-annotated 10 images into grass, clover, weeds and unknown pixels (Skovsen et al. 2017). The annotated images were crops of 1 megapixels of 10 canopy images with a dry matter clover fraction ranging from 5 to 100 percent. To support the increased classification into clover species, each annotated clover was individually classified into either white clover or red clover when distinguishable. This classification was performed by Søren Skovsen and relied on few specific traits:

- Red clover:
    - Erect stem structure, when visible.
    - Visible hair on the leaves and stems.
    - No toothed leaf margin
- White clover:
    - Toothed leaf margin
    - No hairs

When a single clover species was present in the dry matter analysis, all clovers in the corresponding image patch were reclassified accordingly. Fig. 1 shows a sample from the labeled test set, corresponding to 40 percent dry matter clover fraction. The majority of the clovers have been re-classified into species, based on the visual clues in the image.

(a) Image of a field trial sample **(b) 1000 x 1000 pixel image patch of a) to be labeled**





(c) Original label image (Skovsen et al. 2017) **(d) Extended hiearchical label image**

Fig. 1 Illustration of image labels. Colors are mapped as follows: red = clover, blue = grass, orange = weeds, yellow = white clover, purple = red clover, white = unknown

Image Simulation

Due to the high level of trainable parameters, convolutional neural networks require hundreds or thousands of labeled images for training (Oquab et al. 2014). Given the 179 high-resolution grass-clover images, this can be sufficient for training a pixel wise classification CNN, due to the high complexity of the corresponding label image. However, as previously shown (Skovsen et al. 2017), labeling the collected 179 images requires more than 3000 hours of manual labor.

Recent success in simulating vegetation images for CNN training show great potential in both maize (Dyrmann et al. 2016) and grass-clover (Skovsen et al. 2017) segmentation tasks. Extending the work of (Skovsen et al. 2017), the grass-clover image simulation program is expanded to provide hierarchical label images. The images are simulated using few segmented samples of clovers, grasses, and weeds, which are placed on top of an image of soil with random augmentations. By annotating the plant samples more accurately, the information can be passed through the simulation process to provide a more accurate label image. Following this approach, the segmented clover samples are additionally classified as either red clover or white clover before simulation, corresponding to the two clover species present in the collected images. Clover samples, with no visible distinguishable traits, retained the previous general classification, to preserve the natural clover variability in the simulated images. Consequently, the simulated images consist of six distinct classes: Soil, grass, weeds, clover, red clover and white clover, where the clover species are subsets of the general clover class. A sample of each vegetation class is illustrated in Fig. 2.

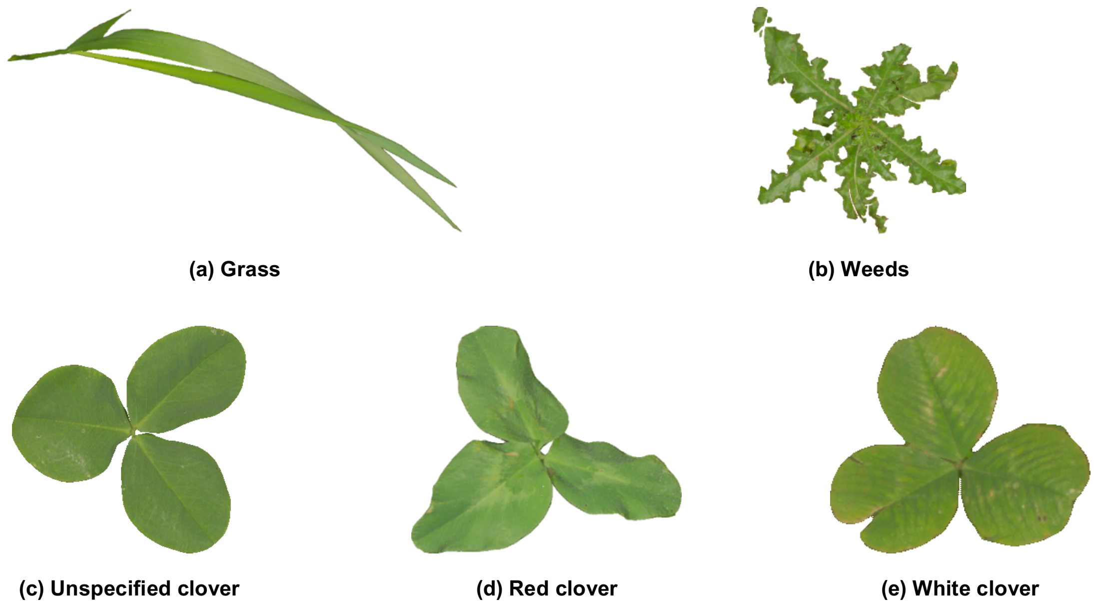
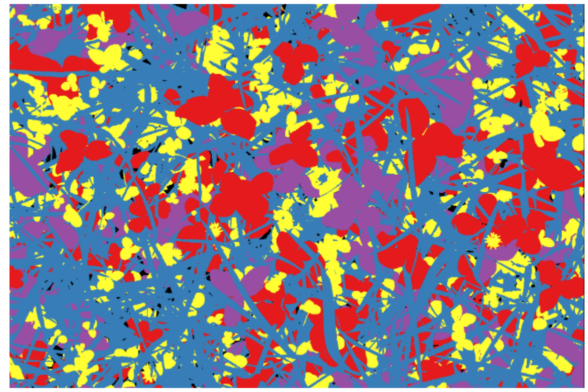


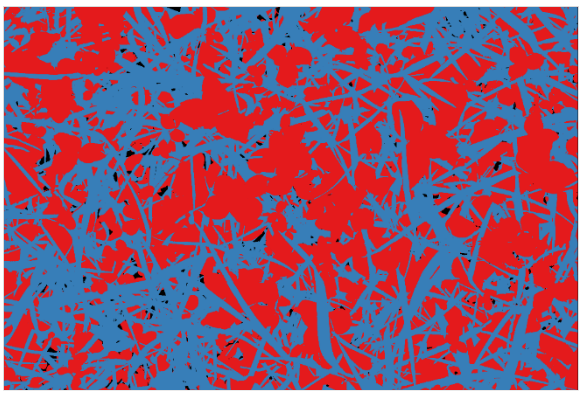
Fig. 2 Illustration of the hand-segmented sample classes used for simulation

To simulate images, a total of 73 white clovers, 33 red clovers, 43 unspecified clovers, 55 grasses, and 15 dandelions, 4 shepherds purse, and 5 thistles, the last three being categorized as weeds, were used. Varying the leaf area index, the clover fraction, the species of clover, the occurrence of weeds, the weed species, and the occurrence of clover flowers, 1720 high-resolution images were simulated along with the corresponding hierarchical labels for training. The images were then scaled to 75 percent size to expand the training data, and increase the invariance to the scale differences in the collected field image samples. With an estimate of 4 minutes per hand-segmented plant sample, we generated 36.3 gigapixels worth of training images, with hierarchical labels, using less than 20 hours of manual labor. One of the simulated training images is presented in Fig. 3 along with the corresponding label image, denoting the classification of each pixel.

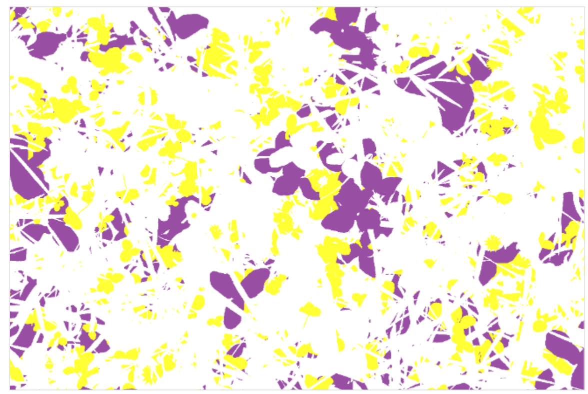
(a) Simulated training image



(b) Simulated hierarchical label for a)



(c) Mapping of b) into label image for upper CNN training



(d) Mapping of b) into label image for lower CNN training

Fig. 3 Illustration of a simulated image with corresponding labels. Colors are mapped as follows: red = clover, blue = grass, yellow = white clover, purple = red clover, black = soil, white = no label

Convolutional Neural Network

Training a CNN on a hierarchy of overlapping labels is a complex task. Instead, we propose to split the task into a cascade of CNNs, where each fork in the hierarchy is learned by a corresponding CNN instance. As the label hierarchy of the training images consists of two forks, a cascade of two CNNs is proposed. Fig. 4 shows the applied CNN architecture, FCN-8s by Long et al. (2015) and an example of corresponding in- and outputs.

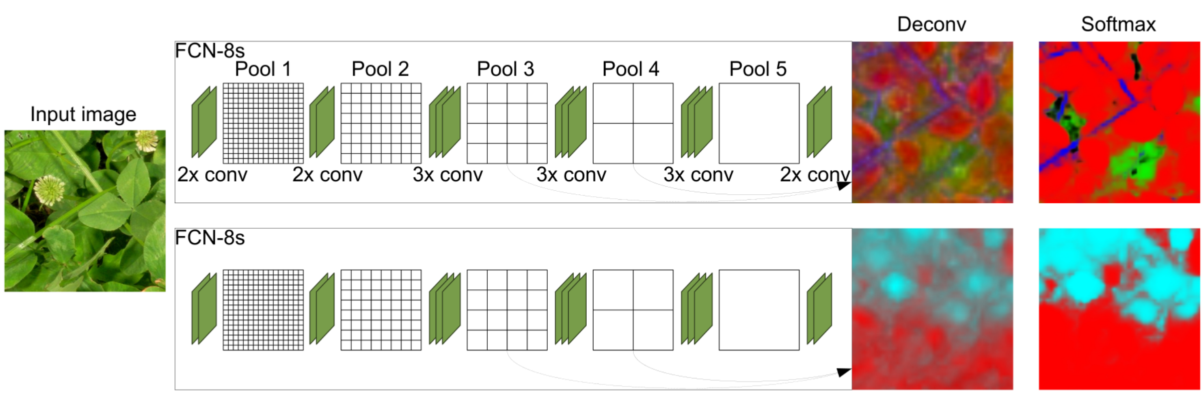


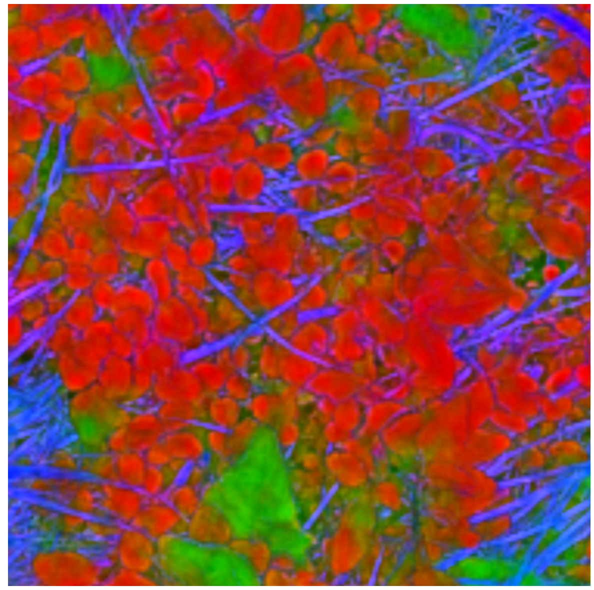
Fig. 4 Illustration of the two convolutional neural networks (CNN). The upper CNN is trained for grass (blue), weeds (green) and clover (red) segmentation. The lower CNN is trained for red clover (red) and white clover (cyan) distinction. When provided the same input image, the outputs are profoundly different

Following the approach of Long et al. (2015), both CNNs were initialized with pretrained weights from a VGG16 CNN (Simonyan and Zisserman 2014), trained for image classification on ImageNet. After converting the CNNs to fully convolutional networks, they were trained purely on the simulated grass-clover images. Both instances were trained for 10 epochs using the AdaGrad optimizer with a learning rate of 10^{-3} and weight decay of 10^{-4}. Each weight update was based on a mini-batch of four images, randomly cropped to 1200 x 1200 pixels to avoid overfitting. To increase the invariance of the segmentation, all images were augmented by randomly varying the brightness, saturation, hue and contrast. To cancel out learning biases from the simulated images, the score values from the trained CNNs are scaled by a class-specific constant accordingly.

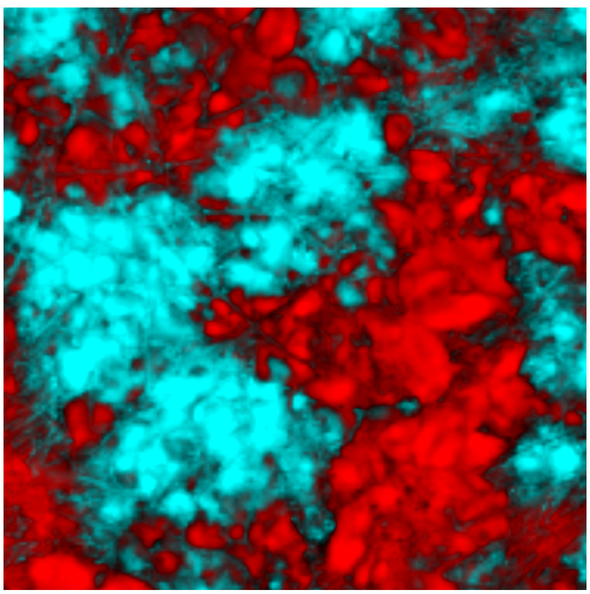
The principal difference between the two trained CNNs is located in the label image. Besides the different class labels, denoted by colors, the first set is densely covering all pixels, where the latter is sparse, as illustrated in Fig. 3 c and d. As the pixel level cross-entropy loss is relying on existing labels, the loss has been slightly modified. Instead of averaging the loss over all pixels in the mini-batch, it is averaged over all labeled pixels, and scaled corresponding to the label coverage in the mini-batch. The main consequence of the label sparsity is the transition from a segmenting CNN to a discriminating CNN. As the lower CNN is never taught to discriminate between grass, weeds and clover, it will provide a red clover and white clover evaluation on all pixels, though reliability only holds for clover pixels.

Fig. 5 illustrates the flow from input image to a cascaded output. Initially, the two CNNs process the same input image, and output a score map. The score map specifies the likelihood of each of the trained classes, illustrated by mapping the scores directly to RGB-channels as an image. Brighter colors indicate a higher likelihood of the given class. After the upper CNN has located clovers in the image, each clover detection is pixel wise substituted with the clover species likelihood of the lower CNN, as illustrated in Fig. 5 d.

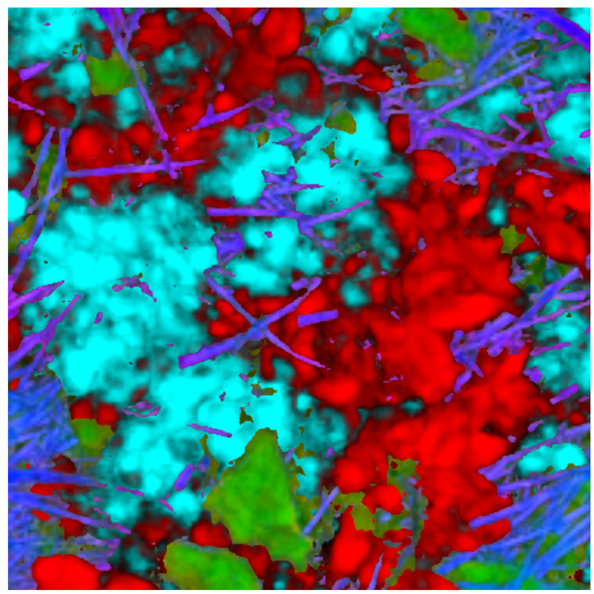
(a) Input image



(b) Score map of the upper CNN. (red = clover, blue = grass, green = weeds)



(c) Score map of lower CNN. (red = red clover, cyan = white clover)



(d) Cascaded score map. Clover detection in b) is replaced by clover species distinction in c). (blue = grass, green = weeds, red = red clover, cyan = white clover)

Fig. 5 Illustration of the flow from input image to a combined likelihood of each vegetation class

Results and Discussion

As the paper proposes a two-step approach for determining the dry matter composition, the presented results were separated similarly. First, the ability to semantically segment 10 test images was evaluated, followed by evaluation of the dry matter composition prediction on the collected samples.

Semantic Segmentation

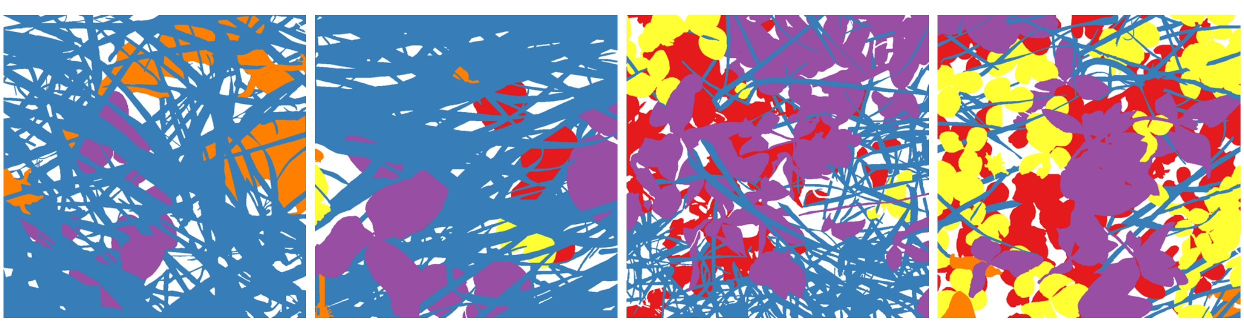
To evaluate the proposed method on the labeled test image patches, the uncropped canopy image was inputted to the CNNs to make use of the receptive field of 404 x 404 pixels of the CNN. Applying a standard arg-max segmentation scheme, each pixel is classified with the most probable class label. As previous approaches (Bonesmo et al. 2004, Skovsen et al. 2017) cannot distinguish clover species, our proposed method is compared both on the individual CNN performance, and on the cascaded CNN approach. When evaluating the individual CNN, each CNN is only evaluated on the isolated task. Consequently, the upper CNN is evaluated on discriminating grass, clover and weeds, and the lower CNN is evaluated on discriminating white and red clover, with no regards to actual clover segmentation performance. When evaluating the cascade of CNNs, an error in the upper CNN, will propagate into the lower CNN.

Table 1 Comparison of the approach by Bonesmo et al. (2014), Skovsen et al. (2017) and the proposed method of this paper. The compared evaluation measurements come from Skovsen et al. (2017). IoU: intersection over union. F.w: frequency weighted

	Grass, clover, weeds			White clover, red clover		
Method	**Pixel Accuracy**	**Mean IoU**	**F.w. IoU**	**Pixel Accuracy**	**Mean IoU**	**F.w. IoU**
(Bonesmo et al. 2004)	65.5	34.3	47.9	-	-	-
(Skovsen et al. 2017)	83.4	65.5	71.7	-	-	-
Individual CNN	87.3	69.9	79.4	89.6	73.4	81.1
Cascaded CNN	87.3	69.9	79.4	83.2	67.7	77.4

The results in Table 1 demonstrate the advantage of CNNs for grass-clover segmentation. While the improvement in grass, clover and weed segmentation since Skovsen et al. (2017) is substantial, these are mainly achieved tuning data augmentation and data simulation processes. The pixel wise accuracy of the red clover and white clover discrimination of 89.6 percent is noteworthy. To the best knowledge of the authors, this task has not been performed automatically before. When applying the cascaded structure, to locate the clovers, and to distinguish the species, the pixel wise accuracy drops 6 percent points due to missed clovers.

Samples of image patches in test set



Ground truth

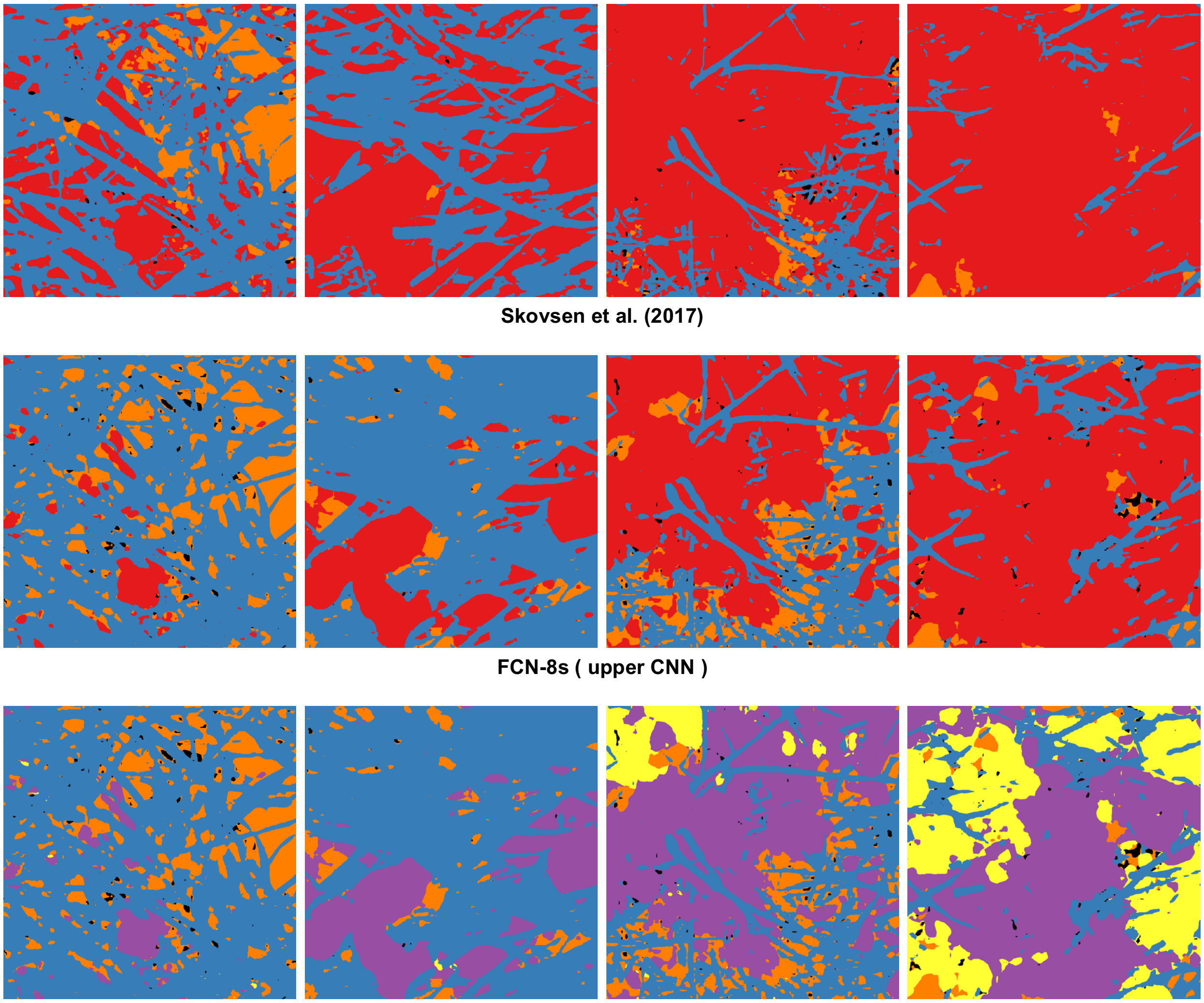


Fig. 6 Visual comparison of recent approaches for grass-clover segmentation. Columns from left to right correspond to 5%, 20%, 50% and 60% clover fraction in the harvested dry matter. Colors are mapped as follows: red = clover, blue = grass, orange = weeds, yellow = white clover, purple = red clover, black = soil, white = unknown class

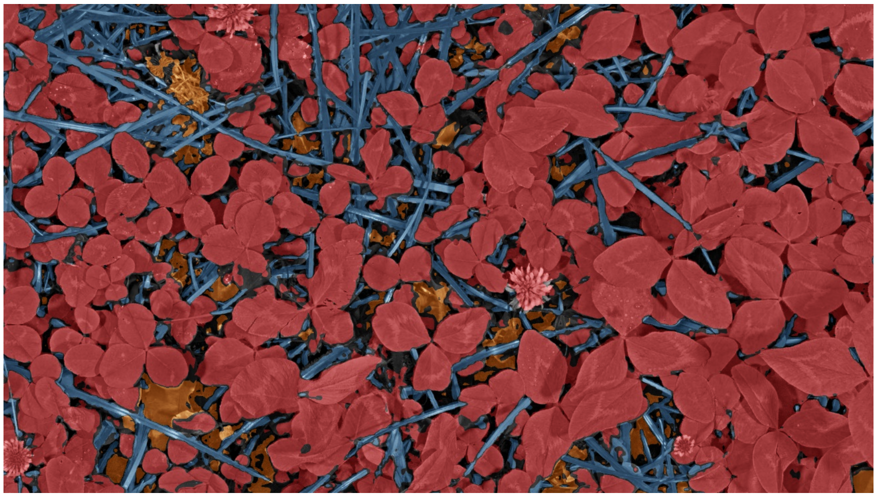
Fig. 6 shows a visual comparison between ground truth label and automatic segmentations, for four randomly selected images in the test set. The bias towards clover in the prior work of Skovsen et al. (2017) is notably decreased, which is believed to be the main reason for the increased accuracy. The spots of white clovers in column three and four are clearly identified by the cascaded CNN, while the annotated red clovers are generally identified correctly as well.

Dry Matter Prediction

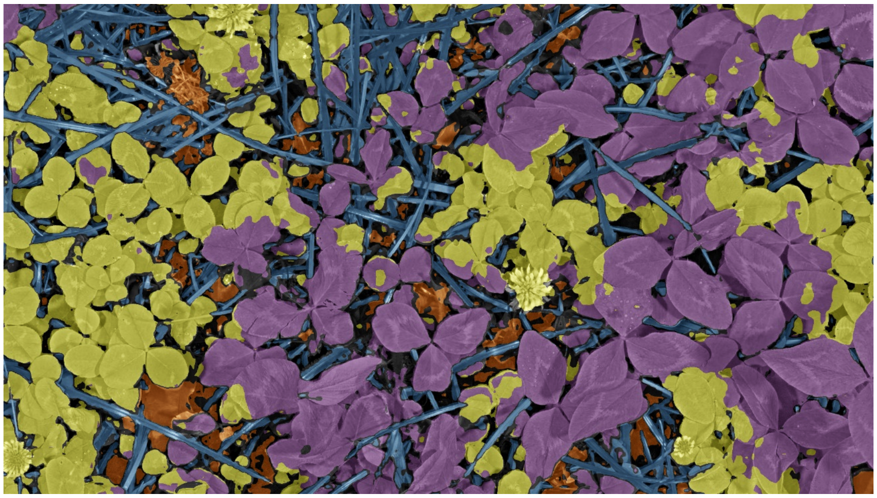
To evaluate the dry matter prediction, all 179 grass-clover image samples are passed through the cascaded CNNs, to automatically segment the images. To increase the accuracy of the semantic segmentation while ignoring unrecognized objects, the segmentation criteria is changed from the most likely class in any pixel, to a threshold on the softmax outputs, controlling the confidence of the classification. The class-specific thresholds are rooted in biases in the simulated training data and the dissimilarities between simulated and real data, and are determined based on visual inspection. The thresholds correspond to the certainty of the CNN that the given pixel belongs to the class. The selected certainty-thresholds were for weeds > 49%, for clover > 76%, for grass > 54%, for white clover > 39%, and for red clover > 59%.

An example of the final segmentation is shown in Fig. 7. This thresholded segmentation provides a more accurate segmentation than demonstrated in Table 1, and segments the grasses, weeds and clovers very accurately. Minor artifacts are visible in the clover species distinction, where parts single clover samples are classified as both species. Extending the approach to include instance segmentation, which can separate each plant sample in the image, is believed to minimize this effect.

(a) Input image



(b) Automated segmentation into grass, clover, and weeds



(c) Automated segmentation into grass, white clover, red clover, and weeds

Fig. 7 Illustration of the thresholded segmentation of a grass-clover image into categorical plant classes, and clover species. Colors are mapped as follows: blue = grass, orange = weeds, yellow = white clover, purple = red clover

As previously demonstrated by Skovsen et al. (2017), the clover fraction in the dry matter, can be correlated with the CNN-detected clover fraction of the detected vegetation. The same holds true for the upper CNN of the proposed method, which is invariable to both yield and time of season, as shown in Fig. 8.

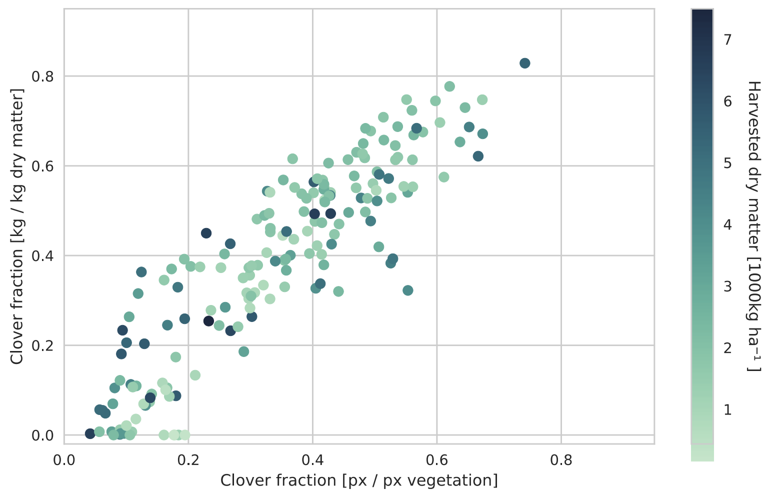
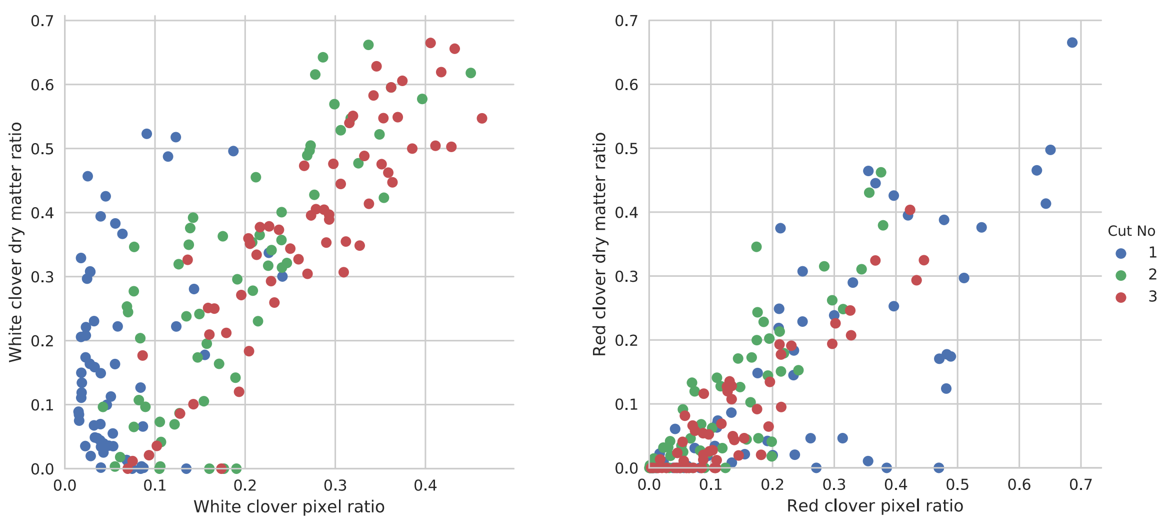


Fig. 8 Comparison of the visual clover clover ratio, determined by the CNN, and the clover ratio in the harvested dry matter

Using a comparable approach, the individual clover species fraction of the dry matter is presented with the individual clover species fraction of the vegetation pixels in Fig. 9. While the relationship between the autonomous canopy analysis and the dry matter composition is difficult to relate in the first seasonal cut, there is a clear relation in the following cuts, demonstrating the ability to discriminate the two clover species in real images.



(a) White clover comparison **(b) Red clover comparison**

Fig. 9 Comparison between the visual clover species in the canopy, determined by the CNNs, and the corresponding clover species in the harvested dry matter

As the yield and vegetation height was significantly higher in the first cut of the season, than the following (Skovsen et al. 2017), it is reasonable to believe that the white clovers were heavily occluded by the red clovers and ryegrass. This is supported by the observation that the amount of white clover was generally higher, than what was visible in the canopy. As the red clover dry matter fraction was often overestimated based on the canopy view, this supports the hypothesis of occlusion.

Conclusion

It has been shown that images with a hierarchy of class labels can be acquired efficiently using data simulation. Training a cascaded convolutional neural network on 3440 simulated images allowed for automatic pixel wise segmentation of images into clovers, grasses and weeds with an accuracy of 87.3 percent. By utilizing the hierarchical information in the training image labels, it was additionally possible to automatically distinguish between red clovers and white clovers with an accuracy of 89.6 percent. Employing the image analysis on 179 canopy images collected in grass-clover plots demonstrated clear correlations for predicting the clover fraction in the harvested dry matter. In the second and the third cut of the season, where the white clover is typically visible in the canopy, it was also possible to determine the clover species of the dry matter.

Acknowledgements

The work was founded by Green Development and Demonstration Programme (GUDP) under the Danish Ministry for Food, Agriculture and Fisheries, and Innovation Fund Denmark.

References

Bonesmo, H.; Kaspersen, K.; Bakken, A.K. Evaluating an image analysis system for mapping white clover pastures. Acta Agric. Scand. Sect. B—Soil Plant Sci. 2004, 54, 76–82.

Dyrmann, M.; Mortensen, A.K.; Midtiby, H.S.; Jørgensen, R.N. Pixel-wise classification of weeds and crops in images by using a fully convolutional neural network. In Proceedings of the International Conference on Agricultural Engineering, Aarhus, Denmark, 26–29 June 2016.

Egaard, K.S. Nitrogen fertilization of grass/clover swards under cutting or grazing by dairy cows. Acta Agric. Scand. Sect. B—Soil Plant Sci. 2009, 59, 139–150.

Eriksen, J.; Askegaard, M.; Søegaard, K. Complementary effects of red clover inclusion in ryegrass-white clover swards for grazing and cutting. Grass Forage Sci. 2014, 69, 241–250.

Frankow-Lindberg, B.E.; Halling, M.; Höglind, M.; Forkman, J. Yield and stability of yield of singleand multi-clover grass-clover swards in two contrasting temperate environments. Grass Forage Sci. 2009, 64, 236–245.

Himstedt, M.; Fricke, T.; Wachendorf, M. Determining the contribution of legumes in legume-grass mixtures using digital image analysis. Crop Sci. 2009, 49, 1910–1916.

Himstedt, M.; Fricke, T.; Wachendorf, M. The relationship between coverage and dry matter contribution of forage legumes in binary legume-grass mixtures. Crop Sci. 2010, 50, 2186–2193.

Himstedt, M.; Fricke, T.; Wachendorf, M. The benefit of color information in digital image analysis for the estimation of legume contribution in legume-grass mixtures. Crop Sci. 2012, 52, 943–950.

Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

McRoberts, K.C.; Benson, B.M.; Mudrak, E.L.; Parsons, D.; Cherney, D.J. Application of local binary patterns in digital images to estimate botanical composition in mixed alfalfa-grass fields. Comput. Electron. Agric. 2016, 123, 95–103.

Mortensen, A.K.; Dyrmann, M.; Karstoft, H.; Jørgensen, R.N.; Gislum, R. Semantic segmentation of mixed crops using deep convolutional neural network. In Proceedings of the International Conference on Agricultural Engineering, Aarhus, Denmark, 26–29 June 2016.

Mortensen, A.K.; Karstoft, H.; Søegaard, K.; Gislum, R.; Jørgensen, R.N. Preliminary Results of Clover and Grass Coverage and Total Dry Matter Estimation in Clover-Grass Crops Using Image Analysis. Journal of Imaging, 3(12), 59.

Nyfeler, D.; Huguenin-Elie, O.; Suter, M.; Frossard, E.; Connolly, J.; Lüscher, A. Strong mixture effects among four species in fertilized agricultural grassland led to persistent and consistent transgressive overyielding. J. Appl. Ecol. 2009, 46, 683–691.

Oquab M., Bottou L., Laptev I., Sivic J. Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks; Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition; Columbus, OH, USA. 23–28 June 2014; pp. 1717–1724.

Phelan, P.; Moloney, A.P.; McGeough, E.J.; Humphreys, J.; Bertilsson, J.; O'Riordan, E.G.; O'Kiely, P. Forage Legumes for Grazing and Conserving in Ruminant Production Systems. Crit. Rev. Plant Sci. 2015, 34, 281–326.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Skovsen, S., Dyrmann, M., Mortensen, A. K., Steen, K. A., Green, O., Eriksen, J., ...; Estimation of the Botanical Composition of Clover-Grass Leys from RGB Images Using Data Simulation and Fully Convolutional Neural Networks. Sensors, 2017, 17(12), 2930.

Suter, M.; Connolly, J.; Finn, J.A.; Loges, R.; Kirwan, L.; Sebastià, M.T.; Lüscher, A. Nitrogen yield advantage from grass-legume mixtures is robust over a wide range of legume proportions and environmental conditions. Glob. Chang. Biol. 2015, 21, 2424–2438.