

Semi-supervised dry herbage mass estimation using automatic data and synthetic images

Paul Albert^{1,3,5}, Mohamed Saadeldin^{2,3,5}, Badri Narayanan^{2,3,5}, Brian Mac Namee^{2,3,5},
Deirdre Hennessy^{4,5}, Aisling O'Connor^{4,5}, Noel O'Connor^{1,3,5}, Kevin McGuinness^{1,3,5}

¹School of Electronic Engineering, Dublin City University

²School of Computer Science, University College Dublin

³Insight Centre for Data Analytics

⁴Teagasc, ⁵VistaMilk

paul.albert@insight-centre.org

Abstract

Monitoring species-specific dry herbage biomass is an important aspect of pasture-based milk production systems. Being aware of the herbage biomass in the field enables farmers to manage surpluses and deficits in herbage supply, as well as using targeted nitrogen fertilization when necessary. Deep learning for computer vision is a powerful tool in this context as it can accurately estimate the dry biomass of a herbage parcel using images of the grass canopy taken using a portable device. However, the performance of deep learning comes at the cost of an extensive, and in this case destructive, data gathering process. Since accurate species-specific biomass estimation is labor intensive and destructive for the herbage parcel, we propose in this paper to study low supervision approaches to dry biomass estimation using computer vision. Our contributions include: a synthetic data generation algorithm to generate data for a herbage height aware semantic segmentation task, an automatic process to label data using semantic segmentation maps, and a robust regression network trained to predict dry biomass using approximate biomass labels and a small trusted dataset with gold standard labels. We design our approach on a herbage mass estimation dataset collected in Ireland and also report state-of-the-art results on the publicly released Grass-Clover biomass estimation dataset from Denmark. Our code is available at <https://git.io/J0L2a>.

1. Introduction

Local monitoring of the biomass composition of grassland has great potential to improve the reasonable use of fertilizers on dairy farms. Nitrogen over-fertilization has detrimental effects on the environment such as the pollution

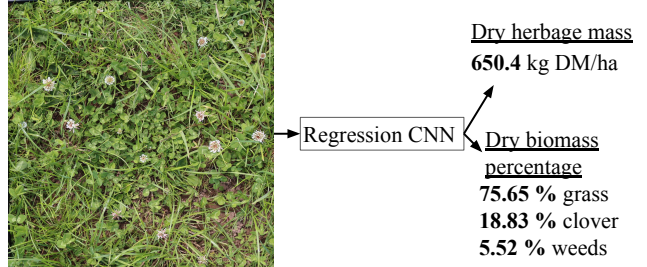


Figure 1: Overview of the dry herbage mass prediction task

of underground water or nearby rivers and a reduction in crop yield [3, 25, 39]. Clover is an important ally to reduce the need for nitrogen fertilization as it influences the impact of the fertilization process [52, 41]. Mapping the density of the clover content in grassland enables a targeted fertilization (as opposed to a uniform fertilization), which allows farmers to anticipate the amount of nitrogen required, and to limit over-fertilization. Balanced amounts of clover also have an important role to play in the final dry feed for the cow, as sufficient amounts of clover in the cow feed increases food intake and augments dairy production [38].

Species phenotyping proposes a direct application of computer vision where a canopy view of the objects is passed to an algorithm tasked with a computer vision problem. Some examples of these tasks include semantic segmentation [51, 18, 36], object counting [27, 6], classification [14, 17, 37], object detection [24, 47] and regression [38, 40]. The principal limitation when applying deep learning approaches to species phenotyping remains the large amount of annotated data required. Lower supervision alternatives using semi-supervised or unsupervised approaches can lower the annotation burden and enable a stronger con-

vergence than using a small number of annotated images alone. In the case of grass/clover biomass estimation this is even more important, as the annotation process is destructive. To accurately measure biomass the region of interest has to be cut, separated, and weighed in a laboratory whereas the collection of un-annotated images is fast and simple.

In this paper, we use a large collection of unlabeled images together with a small annotated subset to improve the accuracy of a dry herbage mass predicting convolutional neural network (CNN, see Figure 1). We first learn a weakly-supervised semantic segmentation network on synthetic images to estimate the species density in the herbage. We then use the segmentation masks to generate automatic biomass labels for the unlabeled images using a simple regression algorithm. Finally, we train a convolutional neural network on a mix of the automatically labeled data and a small number of manually labeled examples to improve the regression accuracy over training on the small number of manually labeled examples alone. We construct our algorithm on an Irish dry herbage mass dataset [19] and validate our results on a publicly available dry biomass dataset [51] collected in Denmark.

Our contributions are:

1. A herbage height aware, weakly supervised, semantic segmentation algorithm trained on synthetic images that is used to automatically label data;
2. An algorithm leveraging automatically labeled images to improve grass/clover/weed dry herbage mass estimation;
3. A detailed study of the importance of the low supervision elements for the final accuracy of our algorithm, and a comparison against the state-of-the-art on a publicly available dataset.

2. Related work

2.1. Image analysis for plant phenotyping

Plant phenotyping and dry matter prediction are excellent domains for the application for image analysis approaches since they enable insight to be extracted from the environment in a non-destructive manner. Existing works explore a variety of computer vision applications for plant phenotyping and in this section we review some of the most relevant to our work. Weed detection aims at localizing unwanted weeds to ultimately remove them by hand or using a robot. Common approaches include employing color filtering, edge detection, and area classification [44, 43, 54]; utilising color features used to train random forest algorithms and support vector machines [21, 22], or using neural networks used to semantically segment images [29]. Fruit or vegetable detection and counting reduces human labor by enabling automatic

fruit treatment or collection on the farm. Examples include tomato segmentation and counting using a convolutional neural network [1], large scale fruit detection in trees [47], or real-time detection using a lightweight neural network [9].

Some approaches use UAV imaging as opposed to ground-level image capture, introducing a fast solution to mapping weeds in a field [15, 22]. As well as using RGB images alone, additional sensors can be added to reduce the difficulty of the phenotyping task [35] such as radar or lidar [30].

2.2. Species biomass estimation from canopy view images

Biomass estimation from canopy view images aims at providing solutions for targeted fertilization in fields. This opens the way for automated fertilization, reducing costs for the farmer and reducing water pollution due to over-fertilization [3, 52]. The heavy occlusions present in canopy images (see Figure 1) poses significant challenges as the biomass estimate should account for elements hidden from the canopy view.

Himstedt *et al.* [20] study the biomass of clover in a legume-grass mixture and demonstrate a good capacity to detect clover from the legumes using morphological filtering and color segmentation to detect the clover. The authors were then able to accurately predict the clover biomass in a controlled environment under the assumption that the total biomass is known. Mortensen *et al.* [38] propose segment the grass clover mixture using color filtering and edge detection before employing a linear regressor to learn the mapping between coverage area of each species and dry biomass content. The authors were then able to directly predict the dry biomass of each element from the image alone. The cow feed scenario presents the added constraint of estimating dry biomass from an image of the fresh pasture.

Skovesen *et al.* [50] propose an improvement over previous work by using a neural network to segment images, and then fitting a linear regressor to the detected species percentages to predict the biomass percentages. To train the neural network, a synthetic dataset is generated using sample crops of relevant species pasted on a soil background. This allows the authors to generate an infinite amount of training images with ground truth from a similar visual domain for their segmentation algorithm. Based on this work, the Grass-Clover dataset challenge [51] asks entrants to improve the author’s baseline using the synthetic images together with a large collection of unlabeled real images and a small set of manually labeled real images.

2.3. Semantic segmentation on synthetic images and domain adaptation

Semantic segmentation aims at predicting the object that each pixel in an image belongs to [16]. The human annotation required for semantic segmentation tasks is exten-

sive, often requiring several hours per image [33]. This makes training strategies using fewer human annotated images attractive. Synthetic images promise to solve part of the problem by providing an unlimited amount of perfectly segmented training images. Popular synthetic datasets for semantic segmentation include The Grand Theft Auto V (GTA V) [45] or SYNTHIA [46] datasets that create synthetic images of cities using graphics engines.

Although the large quantity of labeled data allows a semantic segmentation neural network to converge on a synthetic dataset, the results need to generalize to real world data. Domain adaptation aims at learning domain agnostic features that can generalize from synthetic data to the real world. Domain adaptation strategies can be applied at different stages in a network: input adaptation, feature adaptation, or output adaptation. Input adaptation strategies aim at transforming synthetic images to look more realistic by applying a realistic style on synthetic images, often using a Generative Adversarial Network [60, 49, 48, 12].

Feature adaptation approaches aim to discover domain invariant (or aligned) features. Chen *et al.* [11] propose to use a maximum square loss to enforce a linear gradient increase between easier and harder classes. Luo *et al.* [34] use a significance aware adversarial information bottleneck; Chen *et al.* [13] propose a knowledge distillation approach by matching network activations to a network pretrained on ImageNet.

Output adaptation techniques constrain the network prediction directly to enforce better generalization. This can be achieved using adversarial approaches where the predictions made on synthetic and real data should be indistinguishable to a discriminator network [8], or by enforcing low entropy (more confident) predictions [56]. Batch normalization fine tuning on real data where the batch normalization parameters are tuned on the real images before evaluation has also been shown to be a simple but effective domain transfer strategy [32]. For a more detailed study of domain adaptation for semantic segmentation, we refer the reader to the domain review of Toldo *et al.* [55]

2.4. Semi-supervised learning and label noise

Training computer vision algorithms with limited supervision aims at learning representative features for a downstream task with little to no supervision. In the scope of this paper we train models using a small annotated subset together with a large amount of un-annotated images, which we refer to as a semi-supervised learning scenario. We additionally introduce label noise literature references, which tackles the scenario of approximately labeled data. This is related to the automatic labels we use in this paper.

Semi-supervised learning aims at learning robust features to solve a task using limited annotations. Annotations

are necessary in supervised learning to compute the weights of a neural network using gradient descent on a loss computed using the ground truth annotations. In the case of semi-supervised learning, only part of the dataset has been annotated by humans and the rest is unlabeled images. Iteratively approximating labels for the unlabeled data is a tedious task as the errors made by the network will be amplified (confirmation bias [5]). State-of-the-art semi-supervised learning uses consistency regularization mechanisms where labels are guessed using multiple views of a sample (different data augmentations) [7], sometimes coupled with pseudo-labeling [53].

Label noise proposes robust algorithms to mitigate approximate labeling. Approximate labelling can occur when a dataset is created from web queries [31] or when labels are inferred using label propagation [2]. Solutions for training a neural network on label noise datasets include lowering the contribution of noisy labels in the training loss [23], correcting the label using the network prediction [4], meta-learning inspired corrections [57], monitoring feature space consistency [42], or robust data augmentation [59].

3. Biomass prediction in grass-clover pastures

This section introduces the semi-supervised learning problem of dry biomass estimation of grass-clover pastures, the datasets used, the synthetic image generation process, the automatic labelling pipeline, and our automatic label robust biomass regression algorithm.

3.1. Semi-supervised biomass estimation in grass-clover pastures

We consider here a semi-supervised regression problem with $X_L = \{x_i\}_{i=1}^L$ labeled canopy images of grass and clover, and their corresponding label assignment $Y = \{y_i\}_{i=1}^L$, $Y \in \mathbb{R}^C$ where C is the number of species to predict. The small labeled set is complemented by a large set of unlabeled images $X_U = \{x_i\}_{i=1}^D$ with no corresponding labels and $|X_U| \gg |X_L|$. We note the complete dataset used to train the network $X = X_L \cup X_U$. This paper aims to solve the dry biomass prediction problem from images using a convolutional neural network $\Phi : X \rightarrow Y$ using unlabeled images to improve the regression accuracy.

3.2. Grass clover dry biomass datasets

We consider two different dry biomass prediction datasets, both centered around grass and clover biomass prediction. The first dataset we will refer to as the Irish dataset [19] consists of 528 images labeled with: total dry herbage mass (kg DM/ha), dry grass biomass percentage (%), dry clover biomass percentage (%), and dry weed biomass percentage (%). We study here the low supervision version of the dataset

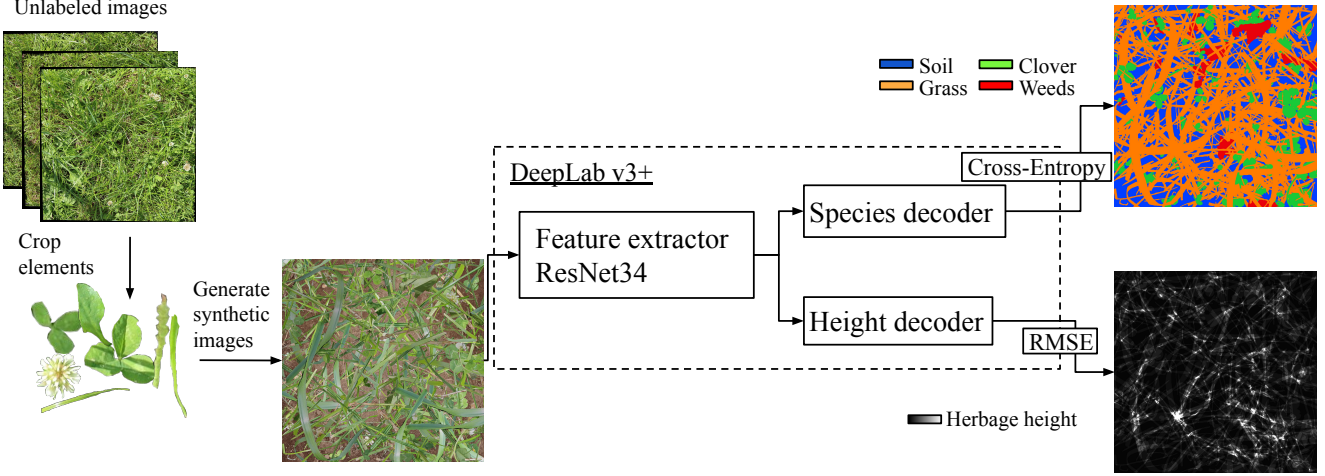


Figure 2: Herbage height aware semantic segmentation on synthetic images

which includes 156 fully annotated images (52 for training and 104 for validation) and an additional 594 unlabeled images. The images were collected using a high resolution Canon camera in Ireland in the summer of 2020.

The second dataset is the GrassClover dataset [51] which contains 152 images labeled with: dry grass biomass percentage (%), dry white clover biomass percentage (%), dry red clover biomass percentage (%), and dry weed biomass percentage (%). Contrary to the Irish dataset, the GrassClover dataset distinguishes between red and white clover species but does not target the direct estimation of the dry herbage biomass (kg DM/ha). The fully annotated images are completed with 31 600 unlabeled images without corresponding ground truth. The dataset was collected in Denmark in 2017 and 2018.

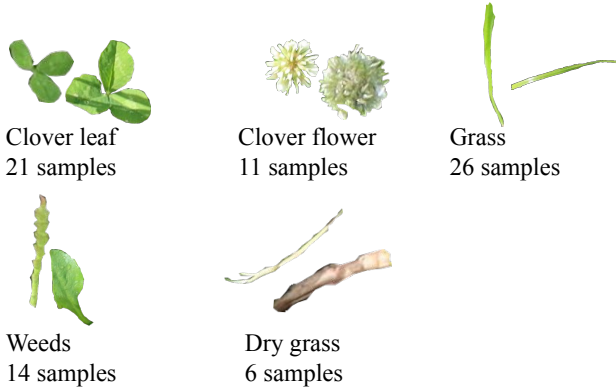


Figure 3: Cropped out samples for every species

3.3. Herbage height aware semantic segmentation on synthetic images

The task we aim to solve in this section is to first predict a semantic segmentation of the herbage into grass, clover (possibly red-white), and weeds; and second, a herbage height map. Since human annotation of ground truth for semantic segmentation can take up to several hours per image [33] and since a pixel specific herbage height is difficult to estimate in practice, we propose (similar to [51]) to train our semantic segmentation network Ψ on a synthetically generated dataset \tilde{X} . We generate the synthetic semantic segmentation images together with their 100% pixel-accurate synthetic segmentation ground truth using manually cropped out elements from the unlabeled images. In accordance to the low supervision scope of this paper, we only crop out 78 samples (see Figure 3) and collect 8 bare soil images to paste elements onto. The bare soil images are collected at the same site and using the same equipment as Hennessey et al. [19] during the Summer of 2021.

To produce images similar to the real images we aim to make predictions for, we respect the species ratio in images by enforcing the probability of a species to be pasted according to the observed average dry biomass distribution in the training dataset: 90% grass, 7% clover, 3% weeds. We draw the probability of each species to be pasted from a 3 component Dirichlet distribution with parameters (9, 2, 1) for (grass, clover, weeds). Once the species has been decided, we randomly draw a sample for this category and apply a series of transformation to increase the diversity of the synthetic images. The transformations include: (uniform) random rotation ($\pm 180^\circ$), random Gaussian blur (radius $\in [0, 5]$), random brightness change [0.6, 1], and random resizing (50 – 150%). Finally, we select a random center location to paste the sample on the background images

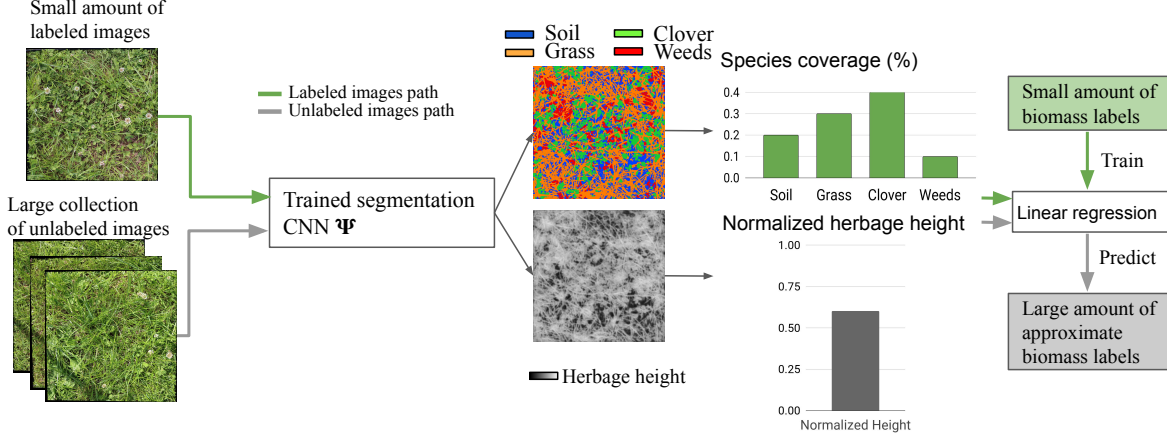


Figure 4: Automatic labeling from semantic segmentation

as well as a mask of the sample’s label on the ground truth map. We additionally approximate the herbage height in the synthetic images as the sum of the total number of successive elements pasted on a pixel. In the rest of the paper, this approximation made on synthetic images will be referred to as herbage height. For example, if three samples have been pasted at the same pixel (clover on top of grass on top of clover), we define the un-normalized herbage height as 3 for the given pixel. Once the synthetic dataset has been fully generated, we compute the 75th percentile of the herbage height for every pixel in all generated images (allowing us to filter outliers) and use this value to clip overly high herbage height numbers and produce a normalized herbage height between 0 and 1 for every pixel in every synthetic image. The normalized herbage height becomes the ground truth target for the segmentation network. Additionally, we found that the quality of the segmentation learnt by Ψ is best when the number of elements to paste is in $[400, 800]$ per image (randomly varied across images); beyond this the synthetic images become overly cluttered. Images are generated at a 2000×2000 resolution. The RGB images are stored in the JPEG format, the grayscale ground truth maps are stored as PNG images, and the herbage height matrix is stored as a compressed numpy array. Figure 4 illustrates the automatic labeling pipeline.

3.4. Generating synthetic images suitable for herbage mass estimation

To concurrently solve the tasks of semantically segmenting the herbage images and estimating the herbage height for every pixel in the images, we propose a herbage height aware semantic segmentation network Ψ consisting of a single feature extractor coupled with two decoder branches (see Figure 2). We concurrently train the species segmentation

branch using a pixel-level cross-entropy loss:

$$l_{\text{species}} = - \sum_{i=1}^C \tilde{y}_i \log(s_i),$$

where $S = s_{i=1}^C$ is the softmaxed prediction of the network and $\tilde{Y} = \tilde{y}_{i=1}^C$ are the synthetic segmentation labels. The herbage height branch is trained using a root mean square error (RMSE) loss:

$$l_{\text{height}} = \sqrt{\frac{1}{P} \sum_{p=1}^P (\tilde{h} - h)^2},$$

where P is the total amount of pixels in the images, h is the ground truth synthetic height label, and \tilde{h} is the network prediction (sigmoid). The total training loss of the segmentation network Ψ is $l = l_{\text{species}} + l_{\text{height}}$.

3.5. Automatic label prediction from species density estimations

The herbage height aware semantic segmentation network Ψ allows us to reduce the complexity of the biomass prediction problem by simplifying the input domain from high resolution real RGB images to the surface area occupied by each species in the canopy as well as an estimated herbage height map. From there, we compute the relative area occupied by each species in the canopy (in %) and the predicted herbage height over each image and train a simple ridge regression algorithm using the small number of labels, Y , to predict approximate labels for X_U . This intermediate task allows us to generate accurate automatic labels for X_U even if the number of images in X_L is very limited.

Table 1: Importance of data augmentation and batch normalization tuning when training on synthetic images.

	HRMSE					HRAE	RMSE			
	Total	Grass	Clover	Weeds	Avg.		Grass	Clover	Weeds	Avg.
Simple DA	357.35	328.66	55.74	26.75	137.05	35.26	8.11	6.87	3.22	6.07
+ ColorJitter	319.92	289.32	60.81	31.40	127.18	35.46	8.63	7.68	3.55	6.62
+ BN tuning	284.60	258.34	51.92	27.05	112.44	31.79	6.49	4.94	3.24	4.89

3.6. Regression on automatic labels with a trusted subset

Although the biomass information can be directly predicted using the automatic annotation process (as done in Skovsen et al. [51]), we propose to attempt to decrease the regression error further by solving the regression problem directly from the RGB images using a convolutional neural network, Φ , and both human-labeled and automatically labeled image datasets: X_L coupled with ground truth labels Y (the trusted set) and X_U coupled with approximate labels \tilde{Y} (the automatically labeled set). Φ is trained to predict the biomass composition (%) and the dry herbage mass (kg DM/ha) from RGB images alone; the automatic images are only used in Ψ to help predict the automatic labels \tilde{Y} for unlabeled images in X_U . To ensure that Φ will not over-fit to incorrect approximate labels, we use three mechanisms. First, we over-sample the trusted data to ensure that a fixed percentage will always be presented to the network in every mini-batch ($\frac{3}{4}$ approximate labels, $\frac{1}{4}$ trusted labels). Second, we use a label perturbation strategy where we randomly perturb the automatic labels to avoid over-fitting incorrect targets, and to avoid penalizing the network for making a prediction slightly different than the incorrect prediction. In practice, we randomly perturb the label by \pm two times the observed RMSE of the automatic labels on the validation set. Finally, we find that adding vertical flipping and randomly grayscaling to the input images to be interesting augmentations that preserve the full herbage information of the image and help further decrease validation error.

4. Experiments

4.1. Training details

We use two different neural networks to solve two distinct tasks. For the semantic segmentation network Ψ , we use a state-of-the-art architecture: DeepLabV3+ [10] where we duplicate the decoder to create the herbage height branch. Ψ is trained on 800 synthetic images and uses 200 synthetic images for validation. We use a ResNet34 [26] as the feature extractor, initialized on ImageNet [28], and with an output stride of 16 for both training and testing. We use the “poly” lr schedule [10] starting at 0.007, a batch size of 4, and train for 60 epochs. For the base data augmentation we resize images

to 1024 on the short size, randomly crop a 1024×1024 square, randomly flip horizontally, and normalize the images.

For the regression network Φ , we use a ResNet18 network [58] pretrained on ImageNet to solve the regression problem from RGB images directly. We train for 100 epochs, starting with a learning rate of 0.03 dividing it by 2 at epochs 50 and 80. We use the same base data augmentation as for Ψ but with a resolution lowered to 512×512 . For the strong(er) data augmentation, we add random vertical flipping and random grayscaling ($p = 0.2$). We train with a batch size of 12.

We use the Irish dataset [19] in its low supervision configuration (52 images are used for training, 104 for validation and 372 for testing) for our exploratory studies, and generate 1000 synthetic images to train Ψ according to the process described in Section 3.4. We validate our results on the GrassClover dataset [51] and use the full 152 fully annotated biomass images, dividing them into 100 for training and 52 for validation; we use the 174 images withheld for the Co-dalab¹ for testing. We make use of 800 randomly selected synthetic images out of the 8000 generated by the authors for Ψ , keeping 200 extra images for validation. We do not train the herbage height branch on the GrassClover dataset.

To evaluate the performance of the algorithms, we report the RMSE when predicting the dry biomass species percentage for both the Irish and GrassClover datasets. For the Irish dataset, we additionally report the RMSE of the global herbage mass prediction (HRMSE, kg DM/ha), the herbage relative absolute error $l_{\text{relative}} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i}$ (HRAE, in %) and the HRMSE specific to each species (kg DM/ha).

4.2. Semantic segmentation on synthetic images

To encourage Ψ to learn robust features that will generalize to unseen real images, we augment the synthetic images using color jittering and Gaussian blur. Furthermore, once the network has converged on the synthetic dataset and before predicting on the real images, we perform batch normalization tuning which is a common domain adaptation strategy [32] on the real images. An ablation study on the importance of the data augmentation and batch normalization tuning is given in Table 1, where we use the best performing regression algorithm from 4.3.

¹<https://competitions.codalab.org/competitions/21122>

Table 2: Ablation study for predicting approximate labels. We report the biomass prediction errors on a heldout validation set. **HL**: hard labels, **SL**: soft labels, **H**: herbage height

	HRMSE					HRAE	RMSE			
	Total	Grass	Clover	Weeds	Avg.		Grass	Clover	Weeds	Avg.
HL	351.54	332.88	51.34	28.29	137.50	41.61	6.82	6.20	3.25	5.42
SL	310.68	279.98	57.48	28.15	121.87	34.18	7.61	5.20	3.24	5.35
HL + SL	315.20	288.52	53.37	28.11	123.33	34.33	6.49	4.91	3.23	4.88
HL + SL + H	284.60	258.34	51.92	27.05	112.44	31.79	6.49	4.94	3.24	4.89

Table 3: Ablation study on training with approximate labels. We report results on the validation set using the linear regression baseline **LR** or training on the trusted data only **T**, the automatic data only **A**, or combinations of both **T+A**

	HRMSE					HRAE	RMSE			
	Total	Grass	Clover	Weeds	Avg.		Grass	Clover	Weeds	Avg.
LR	284.60	258.34	51.92	27.05	112.44	31.79	6.49	4.94	3.24	4.89
T	249.48	253.63	45.62	32.67	110.64	21.67	6.28	5.07	3.94	5.10
A	258.00	239.81	46.51	27.74	104.69	23.48	5.72	5.20	3.29	4.74
T + A	245.04	233.34	34.94	26.32	98.20	21.60	4.70	4.45	3.17	4.11
+ random GS	234.25	217.55	37.57	27.72	94.28	21.55	4.66	4.47	3.27	4.13
+ trusted oversampling	232.08	220.09	35.93	26.34	94.12	21.36	4.33	4.17	3.15	3.88
+ random perturbation	229.93	216.23	35.79	26.05	92.69	19.96	4.22	4.21	3.10	3.84

4.3. Regression from species coverage

We compare different sets of simple features to extract from the segmentation masks as well as the importance of the herbage height prediction when estimating the dry herbage mass. For features directly related to the dry biomass percentages, we compare averaging the most confident prediction for every pixel only (hard label, HL), averaging the full softmax prediction at each pixel (soft label, SL), or using the two sets of features jointly (HL+SL). In the regression model each feature is the average of the observations over the whole image: 4 features (soil %, grass %, clover %, weeds %) for HL or SL (8 for HL+SL), and 1 feature for the herbage height.

We fit a least squares L_2 regularized (ridge) regression algorithm to all features with a regularization factor of 1, and train on the small subset of annotated images before evaluating on the validation set (Table 2). First, we report the RMSE error of the total herbage mass error (kg DM/ha), as well as the detailed grass/clover/weed herbage mass estimation (kg DM/ha). Second, we report the relative RMSE for the total herbage mass (%) and the RMSE for the relative dry biomass estimation (%) for the grass/clover/weeds. We notice that using SL is better than HL when predicting the herbage mass, demonstrating the interest of capturing the full softmax information over the max prediction only. The precision of HL is still beneficial as we observe a good improvement in terms of dry biomass percentage RMSE

when the two sets of features are coupled. When adding the information about the herbage height, a decrease in HRMSE error is observed, validating the importance of the herbage height module in the segmentation architecture.

4.4. Biomass prediction using automatic labels and a trusted subset

We use the automatic labels to enhance the generalization of the regression CNN Φ in order to improve over the linear regression from the predictions of Ψ , especially in terms of herbage mass prediction. Table 3 reports the ablation study showing how the additional mechanisms we introduce allow us to be robust to the approximate automatic regression labels. The reported metrics are described in Section 4.1. We also compare the performance of the regression network against the linear regression from the prediction of Ψ .

4.5. Comparison against other works on the Grass-Clover dataset

We compare the improvements of our approach on the publicly released GrassClover dataset [51]. The target metrics for this dataset are limited to the dry biomass percentages, for which we report RMSE errors. Table 4 reports the performance of our algorithm with and without automatic labels on the test set available on the CodaLab challenge ² and compares against the best available results. We report

²<https://competitions.codalab.org/competitions/21122>

Table 4: Results on the GrassClover test set (RMSE).

	Grass	Clover			Weeds	Avg.
		Total	White	Red		
Skovsen <i>et al.</i> [51]	9.05	9.91	9.51	6.68	6.50	8.33
Naranayan <i>et al.</i> [40]	8.64	8.73	8.16	10.11	6.95	8.52
Trusted data	10.28	10.32	9.24	9.54	7.37	9.35
+ Automatic data	8.78	8.35	7.72	7.35	7.17	7.87

a lower RMSE on average than the methods we compare against and show that our algorithm is capable of using unlabeled images to reduce the biomass estimation error for every species over training on the small trusted subset alone.

5. Conclusion

This paper proposes to improve upon existing low supervision baselines in dry grass clover biomass prediction by making use of unlabeled images. To do so, we first train a herbage height aware semantic segmentation network on synthetic images that we use to generate automatic labels for the unlabeled data using a small set of labeled images. We then train a regression CNN on RGB images directly using the automatic labels to improve the accuracy over using the trusted data alone. We demonstrate the importance of our herbage height aware segmentation network when predicting dry herbage masses from canopy view images as well as the noise robust mechanisms we use to train on automatically labeled data. We improve over our baseline on the Irish dry herbage biomass dataset and set a new state-of-the-art performance level on the publicly available GrassClover dataset.

Acknowledgments

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/15/SIRG/3283 and SFI/12/RC/2289_P2.

References

- [1] Manya Afonso, Hubert Fonteijn, Felipe Schadeck Fiorentin, Dick Lensink, Marcel Mooij, Nanne Faber, Gerrit Polder, and Ron Wehrens. Tomato Fruit Detection and Counting in Greenhouses Using Deep Learning. *Frontiers in plant science*, 2020.
- [2] P. Albert, D. Ortego, E. Arazo, N.E. O’Connor, and K. McGuinness. ReLaB: Reliable Label Bootstrapping for Semi-Supervised Learning. In *International Joint Conference on Neural Networks (IJCNN)*, 2021.
- [3] Francisco Albornoz. Crop responses to nitrogen overfertilization: A review. *Scientia horticulturae*, 2016.
- [4] E. Arazo, D. Ortego, P. Albert, N. O’Connor, and K. McGuinness. Unsupervised Label Noise Modeling and Loss Correction. In *International Conference on Machine Learning (ICML)*, 2019.
- [5] E. Arazo, D. Ortego, P. Albert, N.E. O’Connor, and K. McGuinness. Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. In *International Joint Conference on Neural Networks (IJCNN)*, 2020.
- [6] Tewodros W Ayalew, Jordan R Ubbens, and Ian Stavness. Unsupervised Domain Adaptation for Plant Organ Counting. In *European Conference on Computer Vision*, 2020.
- [7] D. Berthelot, N. Carlini, E. Cubuk, A. Kurakin, K. Sohn, H. Zhang, and C. Raffel. ReMixMatch: Semi-Supervised Learning with Distribution Matching and Augmentation Anchoring. In *International Conference on Learning Representations (ICLR)*, 2020.
- [8] Matteo Biasetton, Umberto Michieli, Gianluca Agresti, and Pietro Zanuttigh. Unsupervised domain adaptation for semantic segmentation of urban scenes. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019.
- [9] Kushtrim Bresilla, Giulio Demetrio Perulli, Alexandra Boini, Brunella Morandi, Luca Corelli Grappadelli, and Luigi Manfrini. Single-shot convolution neural networks for real-time fruit detection within the tree. *Frontiers in plant science*, 2019.
- [10] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *European Conference on Computer Vision (ECCV)*, 2018.
- [11] Minghao Chen, Hongyang Xue, and Deng Cai. Domain adaptation for semantic segmentation with maximum squares loss. In *IEEE International Conference on Computer Vision (ICCV)*, 2019.
- [12] Yuhua Chen, Wen Li, Xiaoran Chen, and Luc Van Gool. Learning semantic segmentation from synthetic data: A geometrically guided input-output adaptation approach. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [13] Yuhua Chen, Wen Li, and Luc Van Gool. Road: Reality oriented adaptation for semantic segmentation of urban scenes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [14] Etienne David, Simon Madec, Pouria Sadeghi-Tehran, Helge Aasen, Bangyou Zheng, Shouyang Liu, Norbert Kirchessner,

- Goro Ishikawa, Koichi Nagasawa, and Minhajul A Badhon. Global Wheat Head Detection (GWHD) dataset: a large and diverse dataset of high-resolution RGB-labelled images to develop and benchmark wheat head detection methods. *Plant Phenomics*, 2020.
- [15] Aaron Etienne and Dharmendra Saraswat. Machine learning approaches to automate weed detection by UAV based sensors. In *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping*, 2019.
- [16] Mark Everingham and John Winn. The pascal visual object classes challenge 2012 (voc2012) development kit. *Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep.*, 2011.
- [17] Thomas Mosgaard Giselsson, Rasmus Nyholm Jørgensen, Peter Kryger Jensen, Mads Dyrmann, and Henrik Skov Midtby. A public image database for benchmark of plant seedling classification algorithms. *arXiv:1711.05458*, 2017.
- [18] Sebastian Haug and Jörn Ostermann. A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. In *European Conference on Computer Vision*, 2014.
- [19] D Hennessy, M Saad, B Mac Namee, N.E. O’Connor, K. McGuinness, P. Albert, B. Narayanan, and A. O’Connor. Using image analysis and machine learning to estimate sward clover content. In *European Grassland Federation Symposium*, 2021.
- [20] M Himstedt, T Fricke, and M Wachendorf. The benefit of color information in digital image analysis for the estimation of legume contribution in legume–grass mixtures. *Crop Science*, 2012.
- [21] N Islam, MM Rashid, S Wibowo, S Wasimi, A Morshed, C Xu, and S Moore. Machine learning based approach for Weed Detection in Chilli field using RGB images. In *International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2020.
- [22] Nahina Islam, Md Mamunur Rashid, Santoso Wibowo, Cheng-Yuan Xu, Ahsan Morshed, Saleh A Wasimi, Steven Moore, and Sk Mostafizur Rahman. Early Weed Detection Using Image Processing and Machine Learning Techniques in an Australian Chilli Farm. *Agriculture*, 2021.
- [23] Lu Jiang, Di Huang, Mason Liu, and Weilong Yang. Beyond Synthetic Noise: Deep Learning on Controlled Noisy Labels. In *International Conference on Machine Learning (ICML)*, 2020.
- [24] Yu Jiang, Changying Li, Andrew H Paterson, and Jon S Robertson. DeepSeedling: deep convolutional network and Kalman filter for plant seedling detection and counting in the field. *Plant methods*, 2019.
- [25] Xiaotang Ju, Xuejun Liu, Fusuo Zhang, and Marco Roelcke. Nitrogen fertilization, soil nitrate accumulation, and policy recommendations in several agricultural regions of China. *AMBIO: a Journal of the Human Environment*, 2004.
- [26] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [27] Saeed Khaki, Nima Safaei, Hieu Pham, and Lizhi Wang. Wheatnet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. *arXiv:2103.09408*, 2021.
- [28] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems (NeurIPS)*, 2012.
- [29] Petre Lameski, Eftim Zdravevski, Vladimir Trajkovik, and Andrea Kulakov. Weed detection dataset with RGB images taken under variable light conditions. In *International Conference on ICT Innovations*, 2017.
- [30] Michael A Lefsky, Warren B Cohen, Geoffrey G Parker, and David J Harding. Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience*, 2002.
- [31] W. Li, L. Wang, W. Li, E. Agustsson, and L. Van Gool. Web-Vision Database: Visual Learning and Understanding from Web Data. *arXiv: 1708.02862*, 2017.
- [32] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. In *International Conference on Learning Representations Workshop (ICLRW)*, 2016.
- [33] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision (ECCV)*, 2014.
- [34] Yawei Luo, Ping Liu, Tao Guan, Junqing Yu, and Yi Yang. Significance-aware information bottleneck for domain adaptive semantic segmentation. In *IEEE International Conference on Computer Vision (ICCV)*, 2019.
- [35] Cheryl L McCarthy, Nigel H Hancock, and Steven R Raine. Applied machine vision of plants: a review with implications for field deployment in automated farming operations. *Intelligent Service Robotics*, 2010.
- [36] Andres Milioto, Philipp Lottes, and Cyrill Stachniss. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. In *IEEE international conference on robotics and automation (ICRA)*, 2018.
- [37] Massimo Minervini, Andreas Fischbach, Hanno Scharf, and Sotirios A Tsafaris. Finely-grained annotated data20sets for image-based plant phenotyping. *Pattern recognition letters*, 2016.
- [38] Anders K Mortensen, Henrik Karstoft, Karen Sjøgaard, René Gislum, and Rasmus N Jørgensen. Preliminary results of clover and grass coverage and total dry matter estimation in clover-grass crops using image analysis. *Journal of Imaging*, 2017.
- [39] F Nájera, Y Tapia, C Baginsky, V Figueroa, R Cabeza, and O Salazar. Evaluation of soil fertility and fertilisation practices for irrigated maize (*Zea mays* L.) under Mediterranean conditions in central Chile. *Journal of soil science and plant nutrition*, 2015.
- [40] Badri Narayanan, Mohamed Saadeldin, Paul Albert, Kevin McGuinness, and Brian Mac Namee. Extracting pasture phenotype and biomass percentages using weakly supervised

- multi-target deep learning on a small dataset. In *Irish Machine Vision and Image Processing conference*, 2020.
- [41] Daniel Nyfeler, Olivier Huguenin-Elie, Matthias Suter, Emmanuel Frossard, John Connolly, and Andreas Lüscher. Strong mixture effects among four species in fertilized agricultural grassland led to persistent and consistent transgressive overyielding. *Journal of Applied Ecology*, 2009.
 - [42] D. Ortego, E. Arazo, P. Albert, N. O'Connor, and K. McGuinness. Towards Robust Learning with Different Label Noise Distributions. In *International Conference on Pattern Recognition (ICPR)*, 2020.
 - [43] Ajinkya Paikekari, Vrushali Ghule, Rani Meshram, and VB Raskar. Weed detection using image processing. *International Research Journal of Engineering and Technology (IRJET)*, 2016.
 - [44] Camilo Andra Pulido-Rojas, Manuel Alejandro Molina-Villa, and Leonardo Enrique Solaque-Guzmán. Machine vision system for weed detection using image filtering in vegetables crops. *Revista Facultad de Ingeniería Universidad de Antioquia*, 2016.
 - [45] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *European conference on computer vision*, 2016.
 - [46] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
 - [47] Inkyu Sa, Zongyuan Ge, Feras Dayoub, Ben Upcroft, Tristan Perez, and Chris McCool. Deepfruits: A fruit detection system using deep neural networks. *Sensors*, 2016.
 - [48] Swami Sankaranarayanan, Yogesh Balaji, Arpit Jain, Ser Nam Lim, and Rama Chellappa. Unsupervised domain adaptation for semantic segmentation with gans. *arXiv: 1711.06969*, 2017.
 - [49] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. In *IEEE conference on computer vision and pattern recognition (CVPR)*, 2017.
 - [50] Søren Skovsen, Mads Dyrmann, Jørgen Eriksen, René Gislum, Henrik Karstoft, and Rasmus Nyholm 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. In *International Conference on Precision Agriculture*, 2018.
 - [51] Søren Skovsen, Mads Dyrmann, Anders K Mortensen, Morten S Laursen, René Gislum, Jørgen Eriksen, Sadaf Farkhani, Henrik Karstoft, and Rasmus N Jørgensen. The GrassClover image dataset for semantic and hierarchical species understanding in agriculture. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
 - [52] Karen Sjøgaard. Nitrogen fertilization of grass/clover swards under cutting or grazing by dairy cows. *Acta Agriculturae Scandinavica Section B—Soil and Plant Science*, 2009.
 - [53] K. Sohn, D. Berthelot, C.-L. L, Z. Zhang, N. Carlini, E. Cubuk, A Kurakin, H. Zhang, and C. Raffel. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *arXiv: 2001.07685*, 2020.
 - [54] Jing-Lei Tang, Xiao-Qian Chen, Rong-Hui Miao, and Dong Wang. Weed detection using image processing under different illumination for site-specific areas spraying. *Computers and Electronics in Agriculture*, 2016.
 - [55] Marco Toldo, Andrea Maracani, Umberto Michieli, and Pietro Zanuttigh. Unsupervised domain adaptation in semantic segmentation: a review. *Technologies*, 2020.
 - [56] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
 - [57] N. Vyas, S. Saxena, and T. Voice. Learning Soft Labels via Meta Learning. *arXiv: 2009.09496*, 2020.
 - [58] S. Zagoruyko and N. Komodakis. Wide residual networks. *arXiv: 1605.07146*, 2016.
 - [59] H. Zhang, M. Cisse, Y.N. Dauphin, and D. Lopez-Paz. mixup: Beyond Empirical Risk Minimization. In *International Conference on Learning Representations (ICLR)*, 2018.
 - [60] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE international conference on computer vision (ICCV)*, 2017.