DeepWheat: Estimating Phenotypic Traits From Images of Crops Using Deep Learning

Shubhra Aich\textsuperscript{1}, Imran Ahmed\textsuperscript{1}, Ilya Obsyannikov\textsuperscript{1}, Ian Stavness\textsuperscript{*1}, Anique Josuttes\textsuperscript{2}, Keegan Strueby\textsuperscript{2}, Hema Sudhakar Duddu\textsuperscript{2}, Curtis Pozniak\textsuperscript{2,3}, and Steve Shirtliffe\textsuperscript{2}

\textsuperscript{1}Department of Computer Science, University of Saskatchewan, Canada
\textsuperscript{2}Department of Plant Sciences, \textsuperscript{3}Crop Development Centre, University of Saskatchewan, Canada

Abstract

In this paper, we investigate the problem of estimating the phenotypic traits of plants from color images and elevation maps of field plots. We focus on emergence and biomass traits — two important indicators of crop growth and health. We employ a state-of-the-art deconvolutional network for segmentation and convolutional architectures, with residual learning in the final stages, for trait estimation. Our intention was to design estimation architectures that behave like high dimensional nonlinear regression models. To the best of our knowledge, this is the first work on emergence counting and biomass estimation based on deep learning. Evaluation was performed on two different species of wheat, grown in field plots for an experimental plant breeding study. Our framework achieves satisfactory performance with mean and standard deviation of absolute difference of 1.20 and 1.53 counts for emergence and 1.45 and 2.05 for biomass estimation. Our results for counting wheat plants from field images is comparable to the accuracy reported for the similar, but arguably less difficult, task of counting leaves from pictures of rosettes grown in pots. Our results for biomass estimation improve upon all previously proposed approaches in the literature.

1. Introduction

Measuring the phenotypic traits of crops, which are the differences in plant characteristics caused by the interaction of the plant’s genetics and the environment, is important in plant breeding research as it allows the breeders to select crop varieties with desirable physical characteristics, such as high yield, resistance to stress, and ability to be easily harvested. Traditionally, phenotypic measurements are made manually in the field, which is both labor intensive and potentially inaccurate due to substantial sub-sampling involved. To overcome these drawbacks, image based automated phenotypic traits estimation is emerging as a vibrant subfield of application oriented computer vision research with the goal of capturing more accurate information at a large scale for better crop production.

In many crops, including wheat, emergence (the density of plants within the field) and biomass (the total mass of each plant) are important phenotypes. Emergence is important because a vigorous and uniform crop stand is needed to compete for moisture, nutrients and sunlight. Plants that emerge late will have a lower yield than the early emerging ones due to the increase in competition for sunlight and essential nutrients [19]. Determining biomass in different crop varieties is important because it is correlated with yield [30] and photosynthetic activity, and is an indicator of overall plant health [11]. These phenotypes are also labour intensive and destructive to measure manually: emergence typically requires physically touching plants in the field to determine which leaves belong to which plant, and biomass measurements are made by cutting out plants from the field and measuring their mass. Furthermore, these phenotypes are traditionally measured on only a sub-sample of the experimental area, which can result in sampling error. The combination of high importance and high measurement difficulty makes these phenotypes good candidates for image-based phenotyping in any crop breeding programs.

Despite the significance of emergence and biomass in crop breeding, little computer vision research has been done on automated estimation of these traits from images. A few studies have looked at plant density estimation in maize [29, 28, 27] and wheat [20, 15] from RGB images. All of these previous methods employ a traditional image processing pipeline that requires hand-tuned parameters tailored to the specific crop of interest. In the wheat studies,
the plant counting algorithm depends on the accurate segmentation of leaves, followed by extracting regional properties of the leaves as features, and then training a simple artificial neural network (ANN) [20] or a support vector machine (SVM) [15]. In both papers, the initial segmentation of the plant foreground from the soil background is accomplished with simple naive approaches: Otsu thresholding on the “b” channel of Lab image or a predefined RGB transformation channel \((2G - 2B - 2.4R)\). However, simple threshold-based segmentations are not robust to changes in lighting conditions and variations in plant/soil color found in different field environments. Indeed, we found that these segmentation approaches gave very poor results for the images used in our study and were therefore not useful benchmarks for comparison.

A number of previous studies have attempted to estimate biomass, but most have done so from field-based measurements and are therefore not applicable to image datasets. A few studies have used aerial images as a basis for biomass estimation. In [24], naive linear regression models are computed from plant height and plant coverage in aerial images. In [23], different linear and nonlinear combinations of height measured with an ultrasonic sensor, leaf area index measured with a plant canopy sensor, and vegetation index from canopy reflectance obtained using a portable spectrometer are used as the predictors and biomass is used as the response of the multiple linear regression model. The product of leaf area index and dry matter content per leaf area is regarded as the estimation of above ground biomass (AGB) in [22]. The authors also provide a comparison against the models developed using exponential regression, partial least square regression and simple artificial neural networks. In [18], AGB was estimated from height information obtained from the Digital Terrain Model (DTM) derived from LiDAR data. For each plot, simple statistical measures of height, such as mean, quadratic mean, standard deviation, skewness, kurtosis, and percentile of height along with height bins at fixed intervals, are used as the predictors for regression modeling. A similar approach is taken in [17] with additional vegetation indices extracted from hyperspectral data. In terms of the list of predictor variables, the approach in [9] can be considered an extended version of the other two [17, 18] with height information plus the vegetation indices based on both hyperspectral and unmanned aerial vehicle (UAV) images. We have implemented these previous methods as a benchmark against which we evaluate our approach.

In this paper, we propose completely data-driven frameworks for emergence counting and biomass estimation. We incorporate state-of-the-art deep learning concepts and architectures to build the individual modules within these frameworks. The name “DeepWheat” refers to our overall system because of this first employment of deep learning in this domain and since we have used the image dataset of two species of the crop wheat for the evaluation. Although we evaluate our approach on wheat, our design allows the frameworks to be generalized over other types of crops with minimal additional manual intervention. We also propose novel data augmentation methods in order to improve training from relatively small sets of aerial images. To the best of our knowledge, this is the first work on image-based phenotypic trait estimation of crops with deep learning.

The rest of the paper is organized as follows. Next section describes both our emergence count and biomass estimation frameworks in detail. Section 3 describes the experimental results along with the evaluation process and datasets used. Finally, we finish the paper with conclusive remarks and possible directions for future work.

2. Our Approach

In this section, we describe the design of both emergence count and biomass estimation frameworks in detail. Although the final approximation models for both tasks are quite similar, the overall workflows are different.

2.1. Emergence Count

Figure 1 depicts the overall computational procedure for counting crop emergence. First, we loosely segment the plant regions from the RGB plot images through the segmentation module described later. Next, we extract all the segmented patches from the whole image, as indicated by the red rectangles in Figure 1 and input each patch image to the counting module to get the individual emergence counts for each patch. Finally, we sum up all the predicted counts for a single plot image to get the overall prediction for emergence count for that particular plot. In this framework, both the segmentation and the counting modules comprise deep architectures which we describe below.

2.1.1 Segmentation

The reason behind segmenting the plots into smaller patches is twofold. First, due to the very high resolution of plot images (\(\sim 2500 \times 7500\)), it is not feasible to do the emergence counting task on the whole image at once from a computational standpoint. Instead, either sequential or parallel counting over the disjoint plant regions appears to be the practically feasible alternative. Second, data-driven approaches like deep learning to develop the counting model requires a large number of training samples, whereas we have only a few high-resolution plot images available for that purpose. Therefore, we require a way to generate a moderate number of training samples for the counting model from a small number of whole plot images. Segregating the plant regions roughly from the soil or background...
Figure 1: Workflow of the emergence count procedure. First, we loosely segment the plant regions from RGB plot images with the segmentation module. Next, small patches containing plants are extracted via connected component analysis. Then, the counting module provides us individual counts for all the patches which are finally summed up to get the overall emergence count for a single plot.

Figure 2: Manual ground-truth generation for relaxed segmentation of plants. The contours shown in red indicate the manual drawing of contours around plant regions. Later, these contours are filled with simple morphological hole-filling operation to create the binary segmentation mask.

followed by the generation of non-overlapping patches provides us with more than a hundred of subsamples from each plot image for further training of the counting model.

From the design perspective, we relax the output of the segmentation module from exact segmentation to a soft or relaxed segmentation for several reasons. First, generating the exact ground-truth manually for images like the ones shown in Figure 2 is a more tedious and time-consuming process than defining loose or relaxed contours around plants. Moreover, for deep networks, learning to count from the subsamples with a somewhat reduced background around the plant regions is not much different compared to that of exactly segmented plants. This is because the background is uniform in this case and so, it is unlikely that the model would behave like a decision or regression tree exploiting any of the distinctive features of the background region. In addition, precise localization of many leaf regions is extremely difficult even by careful visual inspection. Hence, there is a high probability that going for the precise segmentation, the deep segmentation architecture might miss very thin or hard-to-detect regions of the plants which would ultimately deteriorate the counting performance since the model responsible for counting would assume the segregated leaves as different instances rather than a single one. Based on these considerations, soft segmentation is preferred over precise segmentation for our counting framework.

To do soft segmentation with deep learning, some sort of encoder-decoder architecture is needed, where the encoder extracts features necessary for that particular segmentation task from the input in the front-end of the network and the decoder part in the back-end predicts the segmentation mask leveraging those features [26, 8, 21, 12]. For our task, we chose the SegNet architecture [8, 7] (Figure 3) for sev-
Figure 3: SegNet architecture [8]. \( C \equiv \) Convolution, \( N \equiv \) Batch Normalization [14], \( R \equiv \) Rectified Linear Unit (ReLU), \( D \equiv \) Deconvolution or Convolution-transpose, \( P \equiv \) Max-Pooling, and \( U \equiv \) Max-Unpooling. The argument list for all the blocks is \(#\text{InputMaps}, \#\text{OutputMaps}, \text{FilterSize}, \text{Stride}, \text{Padding}\). All the pooling and unpooling operations are \(2 \times 2\) with stride of 2.

Several reasons. First, the soft segmentation problem we are dealing with is easier than the exact segmentation ones in terms of relaxation from exact to soft. Second, this problem is simpler than general multi-class semantic segmentation both in terms of the cardinality of the output category and the nature of the domain since the diversity of the pixel intensities in a single image is highly restrained compared to that of natural images. Furthermore, our concern is not to get an overall-high precision segmentation mask, rather we are conservative only about not missing any plant regions in the image and so, a small amount of under-segmentation is not an issue for the counting model afterward.

Therefore, an easier to train segmentation architecture with a reasonable number of parameters seems to be the right choice for our task and SegNet fulfills our criteria almost perfectly in this regard. However, unlike other encoder-decoder variants of deep models like deconvolutional network [21] which compress the size of the feature maps into a vector with fully connected (FC) layers at the backend of the encoder subnetwork to account for global context, the design of SegNet eliminates these FC layers with the sufficiency assumption of semi-global context instead of purely global one. We are not concerned about the justification of this strong assumption in a generalized setting in this paper. Rather, from the visual exploration of the plot images as shown in Figure 2, we think that this semi-global assumption holds for our domain also. In addition, SegNet has a substantial computational advantage with about 90% reduction in the number of training parameters [8] due to the elimination of FC layers.

2.1.2 Counting by Regression

Figure 4 illustrates the network used for counting the emergence. To design this network, initially, we take inspiration from the leaf counting architecture [7]. However, we find very slow convergence with that model. This might be because, in most images used for the leaf counting problem in [7], the leaves are clearly separable from each other with very little occlusion, whereas almost all the patches extracted after segmentation, contain severe occlusion. Moreover, our objective is to count the number of plants, which is more difficult than leaf counting. One obvious way to address this issue is to increase both the inherent nonlinearity and the capacity of the model by making the model deeper. However, simply adding more layers did not work in our case. Rather, to make the model converge at a reasonable speed, we incorporated the concept of residual learning with additional layers [13], where the feature maps in the subsequent layers are learned as the residue with respect to the previous feature maps (Figure 4). Since information changes gradually between consecutive layers, it is easy to learn the changes only with respect to the already learned maps than to do it from scratch. Consequently, the residual networks are shown to be converging faster than their plain counterparts. Moreover, instead of simply adding plain layers, the incorporation of residual blocks adds flexibility to the model in terms of depth, since bypassing layers through identity mapping is easier with the shortcut connection present in the residual block [13]. The residual block in our paper is simplified in the sense that the number of in-
2.2. Biomass Estimation

For biomass estimation, we have both 5 channel orthomosaics (Blue, Green, Red, NIR, RedEdge) and digital elevation maps (DEM). Sample RGB images are shown in Figure 5. The pixel values of the DEM files indicate the elevation of the objects from the ground, which is plant leaves in our case. Note that, the RGB images of the plots available for emergence counts in the previous section and biomass estimation here are from different sources. The RGB images for biomass estimation are lower resolution (≈ 120 × 480) than those used for emergence counting (see Section 3.1).

Above ground biomass refers to the weight of all plant material above the ground. We expect that there is a relationship between biomass and height or elevation values of the DEM images, but this relationship is difficult to observe from simple biomass versus elevation graphs. However, representing values from each plot as a different dimension in $\mathbb{R}^n$ space, we have found small angles (30° - 32°) in our dataset between the normalized elevation vector and the biomass vector. This suggests a nonlinear relationship between these two quantities and we take this as motivation for further computational analysis.

Now, to apply any data-hungry models like deep learning to estimate biomass from these images, one of the main obstacles is the extremely low number of available samples (≈ 150) for training and testing. One of the obvious ways to overcome this drawback is to figure out a suitable data-augmentation strategy. In this paper, we have devised a novel, simple and effective randomized data augmentation scheme that can be utilized to generate a sufficiently large number of augmented samples from each image. The idea is based on swapping similar superpixels in the image randomly. We call this approach the randomized minimal region swapping (RMRS) algorithm. The steps of the RMRS algorithm are as follows:

1. Get the list of $K$ superpixels from RGB to grayscale image and sort by their mean values.

2. Generate a randomized list of length $N$ of the number of random swaps needed to generate the pool of $N$ augmented samples from a single image. The random integer values are in the range $[low, \lfloor K/2 \rfloor]$, where $low$ is the predefined threshold for the minimum number of swaps needed to create an augmented sample.

3. For each number $r$ in the list generated in step 2, generate a randomized list of length $r$ of either even or odd superpixel indices in the range $[1, \lfloor K/2 \rfloor]$ and swap minimal rectangular regions between those even(odd) superpixels and their consecutive odd(even) counterparts in the sorted list. Even-odd consideration is necessary to avoid un-augmentation by repeated swaps.
Figure 6: Sample RGB plot images (left) with corresponding DEMs (right). Range of the DEM values is converted to grayscale for the purpose of visualization.

Figure 7: Normalized summation of the elevation for the samples augmented from a single image. The first point represents the elevation of the original sample and the rest (499) are the augmented ones. The range of normalized elevation is in the range $[\sim 0.99, 1.0]$ indicating that the total elevation for all the samples are similar to the original.

In our implementation, we use SLIC [6] as the superpixel algorithm. Figure 6 shows sample augmentation results for a single image along with the original one. As can be seen, it is impossible to identify the augmented samples as the artificial ones by looking only at RGB images, even though the corresponding DEMs appear to be highly discretized. Hence, as part of a further exploratory analysis, we plot the normalized summation of all pixel values or elevations of each DEM file for all the augmented samples along with the original one. Figure 7 shows this normalized elevation plot for a single image and its augmented samples. As you can see, the normalized elevation varies in the range $[\sim 0.99, 1.0]$, which means that although the augmented DEM files look different and discretized, the contents of the DEM pixels remain nearly constant after being augmented by the RMRS algorithm.

In addition to increasing the number of training samples, augmenting data this way has another advantage as a byproduct. We hypothesize that the spatial relationships among the pixels in DEM images have little to do with the prediction of biomass since plants can be found in almost any region in the plot images. Therefore, the counting model should learn to map the pixel values from DEM images into the real valued space of biomass in an almost spatially invariant manner. For data augmentation by RMRS algorithm, new samples are just different permutations of the original one. From the practical standpoint, the interpretation might be that to generate an augmented sample, we swap the plants with similar color information within the plot. Thus, by learning to predict from this augmented dataset, the model may intrinsically learn a spatially invariant mapping from color and elevation to biomass.

Finally, we use a similar network architecture for biomass estimation (Figure 8). The only difference between this model and the emergence count one is that the parameters and the placement of the computational blocks or layers are slightly modified to fit the model into this problem.
### 3. Experiments

This section contains the experimental details for our work. First, we describe the datasets used for both tasks. Next, training procedure and implementational details of the networks are provided. Finally, the evaluation metrics are described and the evaluation results are reported in comparison to previous work.

#### 3.1. Datasets

The dataset used for emergence count consists of 274 wheat (Triticum durum) plots of 1.5m × 3.7m area. High-resolution aerial images (∼ 2500 × 7500 pixels per plot) were captured for each plot by walking through the field with a GoPro Hero 5 camera [3] mounted on a monopod with a gimbal for stabilization. Covering plots with this device has the advantage of getting very high-resolution images appropriate for detailed computational analysis compared to other remote sensing technologies.

For biomass estimation, we used aerial drone images for 48 wheat (Triticum aestivum) plots for two dates: June 27 and July 20, 2016. The UAV images have been captured using a MicaSense RedEdge camera [4] on a DraganFly Commander drone [2]. The RedEdge camera includes five different sensors, one for each band: Blue (∼ 465 – 485nm), Green (∼ 550 – 570nm), Red (∼ 658 – 678nm), NIR (∼ 820 – 860nm), and RedEdge (∼ 707 – 727nm). Output from these sensors were post-processed using the Agisoft Photoscan [1] to generate the orthomosaic image and the digital elevation map. For each of these dates, manual ground truth measurement of biomass have also been conducted. For manual counting, plants were cut randomly from the plots at ground-level using sickles, dried, and then weights of those plants were noted. The dataset is randomly split into two equal subsets for training and testing.

#### 3.2. Training and Implementation

We used Torch [10] as the deep learning framework. To train the segmentation network, we generated 0.25M subsamples of size 224 × 224 from 10 high-resolution plot images. The network was trained for 30 epochs over this augmented dataset. SGD-momentum was used as the optimizer with a fixed learning rate, momentum, and weight decay of 0.01, 0.9, and 0.0001 respectively, over the training period.

Both the emergence count and biomass estimation networks were trained with similar parameter settings. Adam optimizer was used with learning rate and weight decay both set to 0.0001. However, for emergence network training, we slowed down the training rate later based on our observation on the training statistics. Training for the emergence network was conducted for 100 epochs, whereas the biomass estimation network was trained with different combinations of input channels for 50 epochs with the same initial parameter settings.

Note that the emergence count network was trained on 7855 patches extracted from 37 images and their slightly augmented versions. On the other hand, the biomass network was trained with about 0.15M augmented training samples generated by the RMRS algorithm.

#### 3.3. Evaluation

Here, we provide three evaluations of our approach. First, we assess the performance for the relaxed binary segmentations generated by our segmentation network. Next, both emergence count and biomass estimation networks are evaluated based on the metrics listed in Equation 1 below. Among these metrics, we take MAD and SDAD from the leaf counting benchmark [7]. The others are simply different forms of these measures.

| Precision | Recall | Accuracy |
|-----------|--------|----------|
| 85.59     | 83.76  | 93.76    |

Table 1: Binary segmentation results

**Emergence evaluation:** Precision, recall, and accuracy are measured to evaluate the segmentation network (Table 1). Results for precision (∼ 86%) and recall (∼ 84%) are somewhat low because the ground truth segmentations are not precise, but loosely defined contours covering all the plant regions in the images. To justify our outputs, we have visually checked almost all the test segmentation results and find very few plant regions undetected by the network.

\[
\begin{align*}
N & = \text{Number of samples} \\
%\text{Underestimate}(U) & = \frac{\sum |a_i - t_i| (a_i < t_i)}{\sum t_i} \\
%\text{Overestimate}(O) & = \frac{\sum |a_i - t_i| (a_i > t_i)}{\sum t_i} \\
%\text{Difference}(D) & = U + O \\
\text{Mean Absolute Difference (MAD)} & = \frac{\sum |a_i - t_i|}{N} \\
\text{Std Absolute Difference (SDAD)} & = \sqrt{\frac{\sum (|a_i - t_i| - \text{MAD})^2}{N - 1}}
\end{align*}
\]

(1)

| Problem     | MAD  | SDAD | %U  | %O  | %D  |
|-------------|------|------|-----|-----|-----|
| Emergence   | 1.20 | 1.53 | 19.82 | 8.54 | 28.36 |
| Leaf Count[7] | 1.62 | 2.30 | - | - | - |

Table 2: Evaluation metrics for the emergence count model

Table 2 lists the evaluation metrics for our emergence counting network. As stated in the introduction, we did not find appropriate literature to benchmark our approach. The
closest approach is the one used for Arabidopsis and Tobacco leaf counting problem [7]. Sample images for this problem are also shown in Figure 9. Although the performance of our approach with %D of 28% and MAD and SDAD of 1.20 and 1.53 may seem somewhat poor, in comparison to the statistics of a rosette leaf counting problem, it is evident that our results are comparable to one of the best leaf counting system currently available, and the present problem is harder due to outdoor images and wheat plants with thin, overlapping leaves.

Figure 9: Downscaled sample images from the A1, A2, A3, and A4 training sets of CVPPP-2017 dataset [5].

As can be seen, the model trained with only H(DEM) as input gives %D of \( \sim 26\% \), which is \( \sim 4\% \) and \( \sim 2\% \) lower than the model trained with RGBH and all the channels. At this point, it is unclear whether the deep learning model takes care of any of the RGB texture in the biomass image. Intuitively, although color information or greenness of the RGB image might be important, the texture information is not that significant for biomass estimation. However, there is a high variance in the color information under different weather conditions. For instance, if the weather is overcast, crops will appear dark-green, for sunny weather, it will be yellowish-green, and so on. Another critical issue is that after augmenting data using RMRS algorithm, albeit the very local texture property and the total energy of the images are more or less preserved, semi-local texture property is destroyed. We are not sure whether this lack of semi-local texture causes the network trained with RGBE input to perform poorer than the one with only DEM input. This issue can only be explored further if sufficient raw training samples are available in future.

On the other hand, the fact that the model works better when two extra non-visible wavelengths, such as, \( NIR \) and \( RedEdge \), are provided along with \( RGB \), is consistent with the plant science literature [25] where vegetation indices extracted from hyperspectral and visible wavelength data are used as strong indicators of photosynthetic measurements of plants. However, the utility of hyperspectral data for biomass estimation is still an open question.

| Method        | MAD  | SDAD | %U   | %O   | %D   |
|---------------|------|------|------|------|------|
| \( H_1 + MARS \) [18] | 1.66 | 2.03 | 23.00 | 6.61 | 29.61 |
| \( H_2 + PLS \) [17]  | 3.86 | 2.72 | 17.34 | 51.58 | 68.92 |
| \( H_2 + MARS \) [18, 17] | 1.74 | 2.07 | 5.08  | 25.88 | 30.96 |
| \( OHS + MLR \) [9]    | 1.67 | 1.63 | 21.46 | 8.21  | 29.67 |
| Ours (H)            | 1.45 | 2.05 | 18.46 | 7.42  | 25.88 |

Table 4: Comparison of biomass results to other methods

In Table 4, we provide comparison against the recent literature. We implemented the methods described in [18, 17, 9] on our data for comparison. These three papers reported the effect of different feature combinations from the set of simple statistical features based on height and different vegetation indices as the predictor variables for their regression models. In this table, we enlist the best combination of features found based on the performance on our dataset. \( H_1, H_2, \) and \( H_3 \) indicates slightly different variations statistical height features and \( OHS \) stands for the combination of \( H_3 \) and Optimized Soil-Adjusted Vegetation Index (OSAVI). Also, \( MARS \) (Multivariate Adaptive Regression Splines), \( PLS \) (Partial Least Squares), and \( MLR \) (Multivariate Linear Regression) are different linear and nonlinear regression algorithms. As can be seen, even with such tiny amount of original training data, the best performance of our deep model (trained with \( DEM(H) \)) is \( \sim 4\% \) better than the recent nonlinear regression model for biomass.

**Acknowledgment**

This work was funded by a Canada First Research Excellence Fund grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

**4. Conclusion and Future Work**

In this paper, we have developed three different deep learning models; one for segmenting plant regions from the background and the other two for emergence count and biomass estimation from aerial images. Our results show better biomass estimation accuracy than previous methods and similar accuracy for outdoor emergence counting as previous studies of indoor leaf counting. Although we have only evaluated our model on particular species of wheat, we expect that our design methodology allows for general-
ization of these models to other types of crops with minimal changes. As future work, we plan to evaluate our networks with other crops that have different plant morphologies, such as pulses and oilseeds. We also plan to further investigate the use of digital elevation maps together with non-visible wavelengths of light as input for biomass estimation and the requirement of more accurate segmentation for the emergence counting pipeline.

References

[1] Agisoft. http://www.agisoft.com/.

[2] DraganFly. http://www.draganfly.com/.

[3] GoPro. https://gopro.com/.

[4] MicaSense-RedEdge Camera. https://www.micasense.com/rededge/.

[5] Leaf Counting Challenge. https://www.plant-phenotyping.org/CVPPP2017-challenge, 2017.

[6] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. S¨usstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11):2274–2282, Nov 2012.

[7] S. Aich and I. Stavness. Leaf counting with deep convolutional and deconvolutional networks. CoRR, abs/1708.07570, 2017.

[8] V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet: A deep convolutional encoder-decoder architecture for scene segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP(99):1–1, 2017.

[9] J. Bendig, K. Yu, H. Aasen, A. Bolten, S. Bennertz, J. Broscheit, M. L. Gnyp, and G. Bareth. Combining uav-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. International Journal of Applied Earth Observation and Geoinformation, 39:79 – 87, 2015.

[10] R. Collobert, K. Kavukcuoglu, and C. Farabet. Torch7: A matlab-like environment for machine learning. In BigLearn, NIPS Workshop, 2011.

[11] J. Dai, B. Bean, B. Brown, W. Bruening, J. Edwards, M. Flowers, R. Karow, C. Lee, G. Morgan, M. Ottmann, J. Ransom, and J. Wiersma. Harvest index and straw yield of five classes of wheat. Biomass and Bioenergy, 85:223 – 227, 2016.

[12] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. G. Rodriguez. A review on deep learning techniques applied to semantic segmentation. CoRR, abs/1704.06857, 2017.

[13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, June 2016.

[14] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In F. Bach and D. Blei, editors, Proceedings of the 32nd International Conference on Machine Learning, volume 37 of Proceedings of Machine Learning Research, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR.

[15] X. Jin, S. Liu, F. Baret, M. Hemerl, and A. Comar. Estimates of plant density of wheat crops at emergence from very low altitude uav imagery. Remote Sensing of Environment, 198:105 – 114, 2017.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 25, pages 1097–1105. Curran Associates, Inc., 2012.

[17] G. V. Laurin, Q. Chen, J. A. Lindsell, D. A. Coomes, F. D. Frate, L. Guerriero, F. Pirotti, and R. Valentini. Above ground biomass estimation in an african tropical forest with lidar and hyperspectral data. ISPRS Journal of Photogrammetry and Remote Sensing, 89:49 – 58, 2014.

[18] G. V. Laurin, N. Puletti, Q. Chen, P. Corona, D. Papale, and R. Valentini. Above ground biomass and tree species richness estimation with airborne lidar in tropical ghana forests. International Journal of Applied Earth Observation and Geoinformation, 52:371 – 379, 2016.

[19] K. Lawles, W. Raun, K. Desta, and K. Freeman. Effect of delayed emergence on corn grain yields. Journal of Plant Nutrition, 35(3):480–496, 2012.

[20] S. Liu, F. Baret, B. Andrieu, P. Burger, and M. Hemmerl. Estimation of wheat plant density at early stages using high resolution imagery. Frontiers in Plant Science, 8:739, 2017.

[21] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In IEEE International Conference on Computer Vision (ICCV), pages 1520–1528, Dec 2015.

[22] X. Quan, B. He, M. Yebra, C. Yin, Z. Liao, X. Zhang, and X. Li. A radiative transfer model-based method for the estimation of grassland aboveground biomass. International Journal of Applied Earth Observation and Geoinformation, 54:159 – 168, 2017.

[23] B. Reddersen, T. Fricke, and M. Wachendorf. A multisensor approach for predicting biomass of extensively managed grassland. Computers and Electronics in Agriculture, 109:247 – 260, 2014.

[24] M. Schirrmann, A. Hamdor, A. Garz, A. Ustuyuzhanin, and K.-H. Dammer. Estimating wheat biomass by combining image clustering with crop height. Computers and Electronics in Agriculture, 121:374 – 384, 2016.

[25] P. J. SELLERS. Canopy reflectance, photosynthesis and transpiration. International Journal of Remote Sensing, 6(8):1335–1372, 1985.

[26] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4):640–651, April 2017.

[27] D. Shrestha and B. Steward. Shape and size analysis of corn plant canopies for plant population and spacing sensing. Applied Eng. in Agric., 21(2):295–303, 2005.

[28] D. Shrestha, B. Steward, and S. Birrell. Video processing for early stage maize plant detection. Biosystems engineering, 89(2):119–129, 2004.
[29] D. S. Shrestha and B. L. Steward. Automatic corn plant population measurement using machine vision, paper number 011067. In 2001 ASAE Annual Meeting. American Society of Agricultural and Biological Engineers, 2001.

[30] J. M. Soriano, M. Malosetti, M. Rosell, M. E. Sorrells, and C. Royo. Dissecting the old mediterranean durum wheat genetic architecture for phenology, biomass and yield formation by association mapping and qtl meta-analysis. *PLOS ONE*, 12(5):1–19, 05 2017.