Multi-resolution Outlier Pooling for Sorghum Classification

Chao Ren[1], Justin Dulay[1], Gregory Rolwes[1], Duke Pauli[2], Nadia Shakoor[3], and Abby Stylianou[1]

[1] Saint Louis University
[2] University of Arizona
[3] Donald Danforth Plant Science Center

Abstract

Automated high throughput plant phenotyping involves leveraging sensors, such as RGB, thermal and hyperspectral cameras (among others), to make large scale and rapid measurements of the physical properties of plants for the purpose of better understanding the difference between crops and facilitating rapid plant breeding programs. One of the most basic phenotyping tasks is to determine the cultivar, or species, in a particular sensor product. This simple phenotype can be used to detect errors in planting and to learn the most differentiating features between cultivars. It is also a challenging visual recognition task, as a large number of highly related crops are grown simultaneously, leading to a classification problem with low inter-class variance. In this paper, we introduce the Sorghum-100 dataset, a large dataset of RGB imagery of sorghum captured by a state-of-the-art gantry system, a multi-resolution network architecture that learns both global and fine-grained features on the crops, and a new global pooling strategy called Dynamic Outlier Pooling which outperforms standard global pooling strategies on this task.

1. Introduction

Sorghum is widely used as an agricultural feed substitute, a gluten-free ancient grain, a source of bio-fuel, and even as popcorn in some food communities. Demand for sorghum for a variety of purposes has risen with the need for better food and energy sources, motivating the need to rigorous plant breeding strategies to select for traits that are valued for each purpose (e.g., more grain for food uses or more biomass for bio-fuel production).

Automated high throughput plant phenotyping involves leveraging sensors, such as RGB, thermal and hyperspectral cameras (among others), to make large scale and rapid measurements of the physical properties of plants for the purpose of better understanding the difference between crops and facilitating rapid plant breeding programs. One of the most basic phenotyping tasks is to determine the cultivar (or

Figure 1: The TERRA-REF Field and Gantry-based Field Scanner in Maricopa, Arizona (top), with sorghum being grown in the field. Sorghum is a hugely important cereal crop, widely used as a source of grain, agricultural feed, and even bio-fuel. Over several seasons, hundreds of varieties of both bio-energy and grain sorghum were grown in the TERRA-REF field (middle and bottom), and were imaged daily for the purpose of high throughput phenotyping.

species) in a particular sensor product. In experiments with a large number of related cultivars being grown simultaneously, this is a challenging fine-grained visual categorization task due to the low inter-class variability.

Our contributions towards this high throughput phenotyping task are threefold:

- A large scale sorghum cultivar classification dataset consisting of tens of thousands of images captured from a gantry based RGB sensor;

- A multi-resolution network architecture that learns relevant features at different scales; and

- A new global pooling strategy called Outlier Pooling which significantly outperforms standard pooling approaches in this agricultural setting.

2. Background

2.1. TERRA-REF Field and Gantry-based Field Scanner

In 2016, the Advanced Research Project Agency–Energy (ARPA-E) funded the Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform, or TERRA-REF[2]. The TERRA-REF project stood up a state-of-the-art gantry based system for monitoring the full growth cycle of over an acre of crops with a cutting-edge suite of imaging sensors, including stereo-RGB, thermal, short- and long-wave hyperspectral, and laser 3D-scanner sensors. The goal of the TERRA-REF gantry was to perform in-field automated high throughput plant phenotyping, the process of making phenotypic measurements of the physical properties of plants at large scale and with high temporal resolution, for the purpose of better understanding the difference between crops and facilitating rapid plant breeding programs. Due to the technical demands of high-throughput phenotyping, it is most often performed in controlled environments (e.g., greenhouses with imaging platforms). Controlled environments play a very important role in understanding plant performance by providing management of the abiotic environment, and the ability to reproduce experimental conditions year-round, but plant performance in field settings, both in terms of growth and yield parameters, is strongly influenced by variability in weather, soil conditions and other environmental parameters that cannot be observed in the greenhouse. The TERRA-REF gantry system was designed to meet the sort of technical requirements for high throughput phenotyping in a field setting. The TERRA-REF field and gantry system are shown in Figure 1, and example data captured from its RGB, 3D-scanner and thermal cameras are shown in Figure 2.

Over the course of its first several years in operation, the TERRA-REF platform collected multiple petabytes of sen-
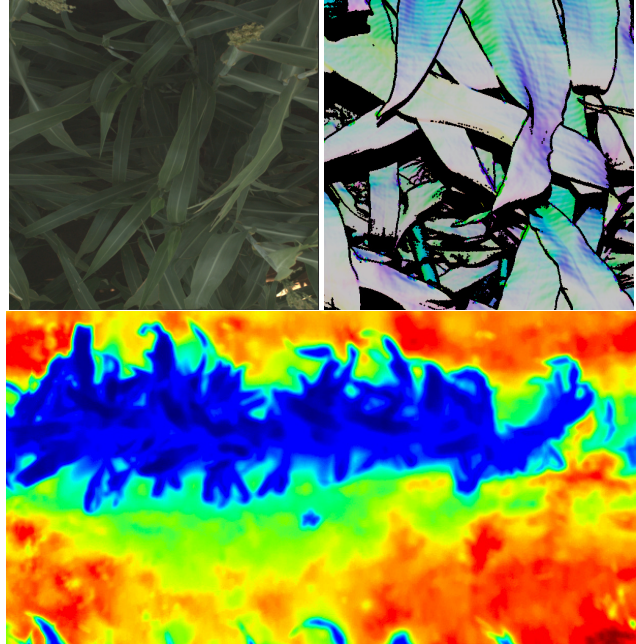

Figure 2: **Example data from the TERRA-REF gantry system.** (top-left) RGB data. (top-right) 3D-scanner data (false color, where color indicates the surface normal and value indicates depth from the scanner). (bottom) Thermal data. In this paper we focus on data from the RGB camera.

sor data capturing the full growing cycle of sorghum plants from the sorghum Bioenergy Association Panel [1], a set of 390 sorghum cultivars whose genomes have been fully sequenced and which show promise for bio-energy usage.

2.2. Phenotyping from Aerial Data

While in-field sorghum phenotypes have been determined using aerial RGB data from drones[12, 10, 3, 18, 4, 11], UAV datasets are limited in their temporal resolution due to the labor required in capturing the data, and their spatial resolution is limited both by the sensors that are able to be mounted on board a drone and the time constraints of the drone operator (flights from lower to the ground are higher resolution but take longer to complete). By comparison, the data from the TERRA-REF Gantry-based Field Scanner has both high spatial resolution (the gantry has extremely high quality sensors and the height of the gantry is placed to optimally image the plants thoughout the growing cycle), and temporal resolution (data is captured every day).

3. Dataset & Classification Task

In this paper, we describe the Sorghum-100 dataset, a curated subset of the RGB imagery captured during the TERRA-REF experiments, labeled by cultivar and day after planting. The dataset will be released publicly and there
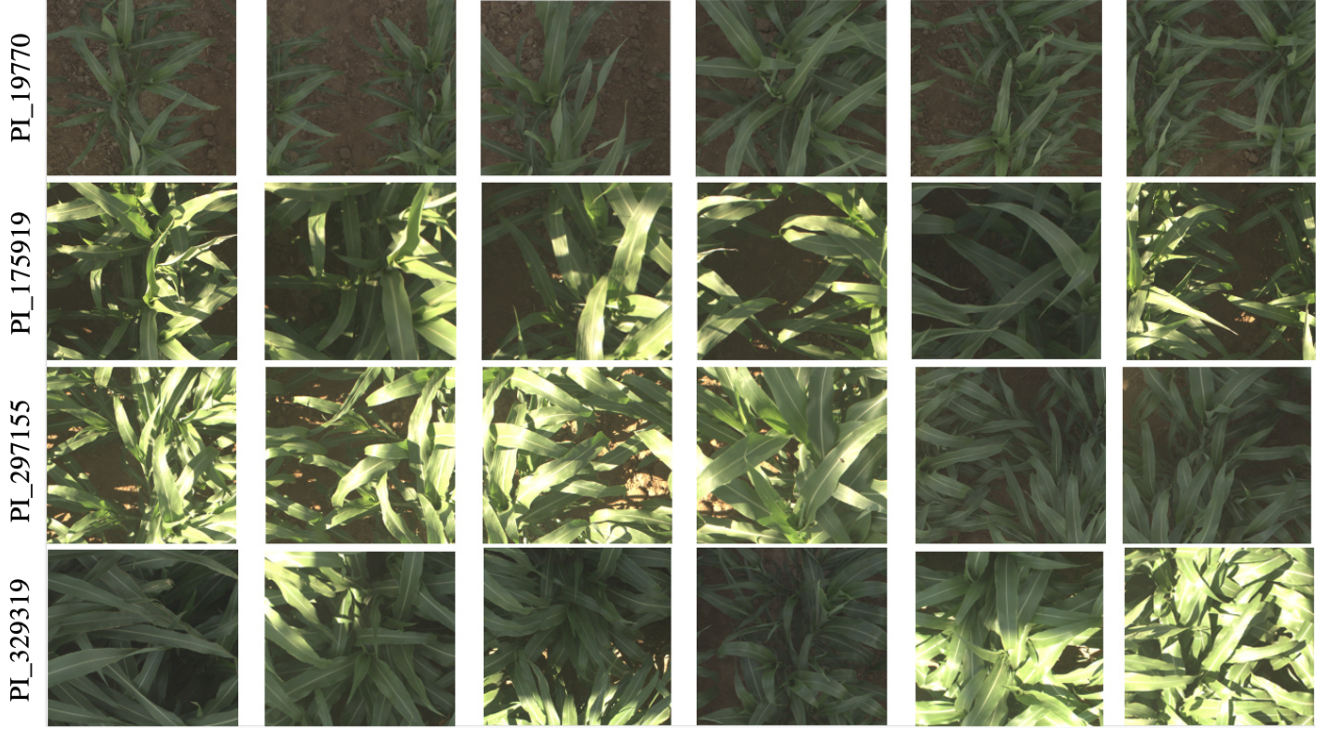
Figure 3: Multiple images from the dataset representing different cultivars. The rows represent different culitvars. The columns represent different captured dates respectively: June 1st, 3rd, 7th, 17th, 19th and 27th, 2017.

will be a corresponding Kaggle competition. This data could be used to develop and assess a variety of plant phenotyping models which seek to answer questions relating to the presence or absence of desirable traits (e.g., "does this plant exhibit signs of water stress?"). In this paper we focus on the question: "What cultivar is shown in this image?" Predicting the cultivar in an image is an especially good challenge problem for familiarizing the machine learning community with the TERRA-REF data. At first blush, the task of predicting the cultivar from an image of a plant may not seem to be the most biologically compelling question to answer – in the context of plant breeding, the cultivar, or parental lines are typically known. A high accuracy machine learning predictor of the species captured by the sensor data, however, can be used to determine where errors in the planting process may have occurred. For example, seed may be mislabeled prior to planting, or planters may get jammed, depositing seeds non-uniformly in a field [14]. Both types of errors are surprisingly common and can cause major problems when processing data from large-scale field experiments with hundreds of cultivars and complex field planting layouts.

The Sorghum-100 dataset consists of 48,106 images and 100 different sorghum cultivars grown in June of 2017 (the images come from the middle of the growing season when the plants were quite large but not yet lodging – or falling

over). In Figure 3, we show a sample of images from four different cultivars. Each row includes six images from different dates in June. This figure highlights the high interclass visual similarity between the different classes, as well as the high variety in the imaging conditions from one day to the next, or even over the course of a day.

The dataset is divided into a training dataset and a testing dataset. Each cultivar was grown in two separate plots in the TERRA-REF field as shown in Figure 4 (top) to account for extremely local field or soil conditions that might impact the growth of plants in one particular plot. We leverage this natural split in the data when dividing our dataset between train and test – images for a given cultivar in the training dataset come from one plot, while the test images from that same cultivar come from the other plot. This means that a model cannot achieve high performance by memorizing features that aren't meaningful phenotypes (e.g., by memorizing patterns observed in the dirt).

The original data from the TERRA-REF gantry is not divided by plot. In Figure 4 (bottom), we show an original RGB image captured by the TERRA-REF system, and one plot crop that was extracted from that image highlighted in blue. Plot crops were extracted based on the highly accurate metadata reported by the gantry regarding where the camera was positioned when the image was captured, and the known geographic boundaries of the plots on the ground.

Figure 4: **TERRA-REF Field Organization.** (top) Each experimental cultivar was planted in two different plots at distant locations in the field (borders were planted with well-known cultivars). In the top figure, each plot is labeled by its cultivar name, and plots from the same cultivar have matching colors. For each non-border cultivar, we include images from one of the plots in our training data, and images from the other plot in our test data, requiring models to generalize across field locations (as opposed to, for example, overfitting on unique ground features that are not relevant to the cultivar). (bottom) Original data from the sensor is pre-processed to crop regions that confidently only consist of plants from a single plot. The blue rectangle in the image above shows a the ground boundaries of a plot projected onto the image.

By defining the plot based on the ground coordinates, some parts of the plants can actually be cut off at the edges of the plot crop, due to the growth of the plants and the viewpoint of the sensor. We chose this behavior over the alternate of allowing a padding around each plot – the padding approach would perhaps avoid the issue of chopping off portions of a plant in the images, but would introduce the potentially more problematic issue of having a significant number of pixels from adjacent plots (which come from different cultivars) included in the image.

4. Global Pooling Strategies

A very common template for deep convolutional neural network architectures is to end with a final convolutional layer that is significantly smaller spatially than the input volume with a large number of channels, followed by a 'global pooling' operation, which converts each channel to a single number. This global pooled output can then be passed through one or more fully connected layers. This template is seen in a variety of popular models used in both classification and deep metric learning, including AlexNet [8], ResNet [5], ResNeXt [16], Inception [15] and DenseNet [7].

The two most common strategies for performing the global pooling that converts each channel in the final convolutional layer to a single number are Global Average Pooling, where the value in the pooled layer, n, per channel, j, is the average value of that channel in the preceding layer:

$$\text{layer}_n(j) = \frac{\sum \text{layer}_{n-1}(j)}{\text{width}_{n-1} \times \text{height}_{n-1}} \tag{1}$$

and Global Max Pooling, where the output at channel i is determined by the max value of channel j in the preceding layer:

$$\text{layer}_n(j) = \max\left(\text{layer}_{n-1}(j)\right) \tag{2}$$

Both approaches were introduced in [9]. While there are other global pooling strategies, such as global sum pooling (which takes the sum of all of the activations, and is proportional to global average pooling) and learned global pooling operators [17], global average pooling and global max pooling are by far the most commonly used.

Intuitively, we can reason that a Global Average Pooling strategy is well suited to problem domains where the relevant features are evenly distributed throughout an image, while Global Max Pooling is well suited to problem domains where the relevant features are sparse. In high throughput phenotyping (especially in this gantry based system), however, relevant features are often sparse, such as panicles or flowers that are only present in some images from some species from some parts of the season, but may have more than one instance where present (such as the pink

4

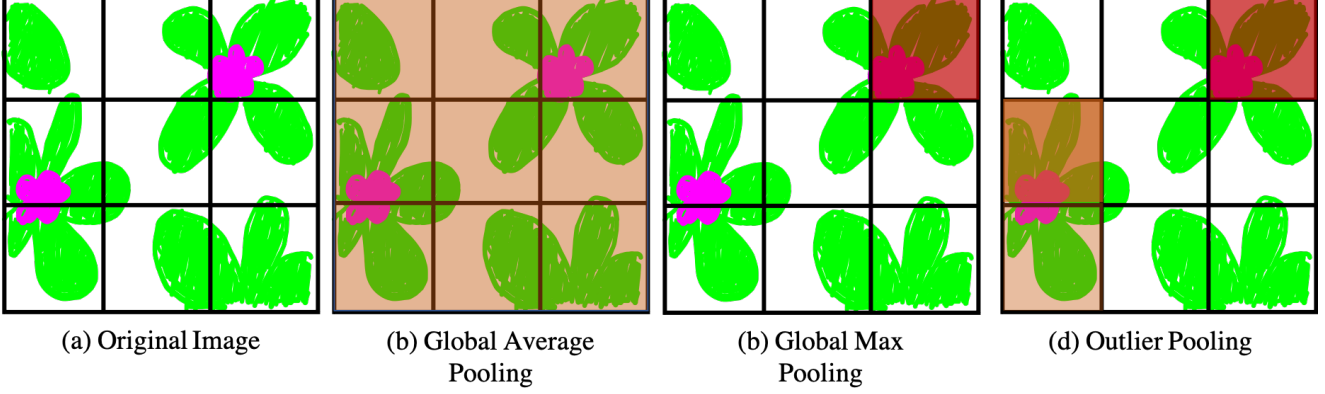| (a) Original Image | (b) Global Average Pooling | (b) Global Max Pooling | (d) Outlier Pooling |

Figure 5: Imagine a hypothetical learned high level filter that activates highly at locations with pink flowers, like those shown in (a). In this figure, we visualize the activation of such a filter when passed through a Global Average Pooling layer (b), a Global Max Pooling layer(c), and our proposed Outlier Pooling (d). Higher activations being passed through the global pooling operation are visualized with darker red colors, while lower values are visualized in lighter colors (locations which do not contribute any activation have no coloring). In Global Average Pooling, the filter averages together active and in-active regions, under-weighting the flowers in the image. Global Max Pooling only considers one of the flowers. Our proposed Outlier Pooling passes activations for regions near both flowers.

flowers in Figure 5). In these cases, global average pooling would under-weight the multiple instances of the interesting object, while global max pooling would only consider one of them.

4.1. Outlier Pooling

We propose a novel global pooling strategy that is well suited to these settings where relevant features may be interspersed throughout an image. The approach is called Outlier Pooling, where the output at channel j is determined by the average value of any locations in the preceding layer that exceed a threshold:

$$\text{layer}_n(j) = \frac{\mathbf{I} \sum \text{layer}_{n-1}(j)}{\sum \mathbf{I}} \quad (3)$$

where $\mathbf{I}$ is an indicator function for every spatial location, x, in channel j of layer_{n-1}:

$$\mathbf{I}_{x_j} := \begin{cases} 1 & \text{if } x_j >= t_j \\ 0 & \text{if } x_j < t_j \end{cases} \quad (4)$$

and t_j is the threshold for channel j:

$$t_j = \mu_{\text{layer}_{n-1}(j)} + \lambda \sigma_{\text{layer}_{n-1}(j)} \quad (5)$$

$\mu_{\text{layer}_{n-1}(j)}$ is the channel mean computed per image, $\sigma_{\text{layer}_{n-1}(j)}$ is the channel standard deviation computed per image, and λ is a parameter that is fine-tuned based on the dataset (we empirically find a value of 2 to perform best for the Sorghum-100 dataset).

4.2. Dynamic Outlier Pooling

Very early in the training process, the most relevant features have not yet been learned. When initializing features from a model pretrained on ImageNet, all of the images in the Sorghum-100 dataset produce extremely similar output features as there is very low variability in these images relative to the entire space of ImageNet images. This means that there are very few locations whose activations will vary significantly from the mean, leading to difficulty in training with Outlier Pooling early on.

To account for this challenge, we use a Dynamic Outlier Pooling approach, which acts more like average pooling early in the training process and becomes progressively more focused on outliers as the training progresses. In this Dynamic Outlier Pooling, the output for a channel j is:

$$\text{layer}_n(j) = \frac{w_1 \mathbf{I} \sum \text{layer}_{n-1}(j) + w_2 (1 - \mathbf{I}) \sum \text{layer}_{n-1}(j)}{\text{width}_{n-1} \times \text{height}_{n-1}} \quad (6)$$

where $\mathbf{I}$ is the indicator function from Equation 4, and w_1 (the relative weighting of the values above the threshold) and w_2 (the relative weighting of the values below the threshold) change over the course of training according to the following schedule:

$$w_1 = 1 + \frac{\text{current epoch}}{\text{total epoches}} \quad (7)$$

$$w_2 = 1 - \frac{\text{current epoch}}{\text{total epoches}} \quad (8)$$
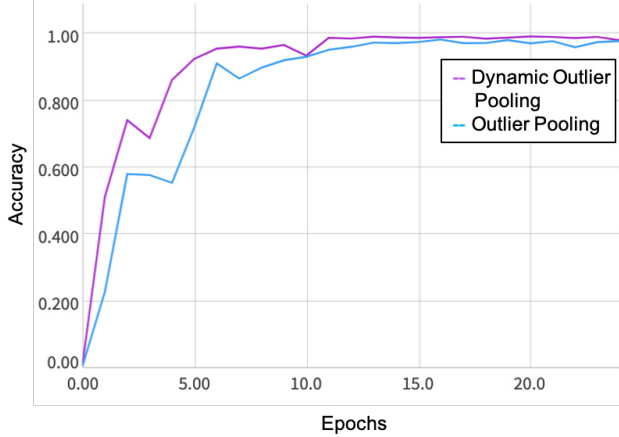
5

Figure 6: Using Dynamic Outlier Pooling, as described in Section 4.2 leads to faster convergence on the Sorghum-100 training set when compared with the non-Dynamic Outlier Pooling.

The impact of this dynamic weighting is that as training progresses, the features that do not exceed the threshold become less impactful on the global pooled feature in $layer_n$. A comparison of the convergence between Dynamic Outlier Pooling and the non-Dynamic version is shown in Figure 6, where we can see that the Dynamic version leads to faster convergence.

4.3. Multi-Resolution Crop Pooling

In addition to considering the appropriate pooling strategy for the gantry-based high throughput phenotyping domain, it is also important to consider the best visual scale of data to consider. The recent advances in commercial GPU technologies with large amounts of memory (e.g., 16GB Nvidia V100 GPUs, such as those used in the experiments in this paper) allow for training on relatively high resolution imagery – but there are still limits to the practical image resolution and size that can be used in training, and trade-offs between the size of the image and the number of images that can be fit in a batch. Moreover, it is desirable that the neural network 'see' both global scale features (e.g., "What does this whole plot of plants look like?") as well as local, high resolution features (e.g., "What does the structure of this flower look like?"). In order to do this while still having reasonable batch sizes and training, we propose a Multi-Resolution Crop Pooling approach, as seen in Figure 7.

This architecture has two parallel branches – one that is a low resolution 'global scale' branch, and one that is a high resolution 'local scale' branch. In the global scale branch, the input is a resized version of the original image (during training, the image is resized and then randomly cropped to be square; during testing, the image is resized and then center cropped).

In the local scale branch, there are four models with shared weights. A random crop at original resolution is fed into each model. We then perform global pooling across the final convolutional layer from all four crops. The intuition behind this is that, as an example, you may see a flower in only one of the high resolution crops, and we want the network to give 'full credit' for the presence of that flower, without down-weighting the features corresponding to that flower based on their absence in the other high resolution crops.

The global pooled features from these two branches are then concatenated and fed through a final fully connected layer.

This architecture is most similar to the feature pyramid architecture from [6], which trained two networks: one on crops from the ImageNet [13] dataset for 'object scale' features, and one on crops from the Places [19] dataset, for 'scene scale' features. At test time, they took different scale crops of the test image, ran each through whichever network the crop was suited for (smaller crops through the ImageNet trained network and larger crops through the Places trained network), computed the max feature per crop, and then concatened each of these max pooled features. Our approach differs in that we do not train our two networks on separate datasets, but rather different scales drawn from the same data, and most importantly in the cross-crop pooling that occurs before the features are concatenated with each other. We additionally utilize the proposed Dynamic Outlier pooling as the global pooling strategy both for the low resolution images in the global scale branch, and for the combination of the high resolution crops in the local scale branch.

5. Experiments

In the following experiments, we train separate networks comparing our Dynamic Outlier Pooling with Global Average Pooling and Global Max Pooling. We additionally do this comparison training a single low resolution image of the entire plot (only the global scale branch of our proposed network), four high resolution crops (only the high res branch of our proposed network), and finally our proposed multi-resolution model.

We initialize the trunk of each of our models with a ResNet-50 model [5] pre-trained on ImageNet. For experiments where only a single, low resolution image is used, during training we resize the original images to be 512 pixels on its shortest side, and then take a random 512×512 crop (at test time, we take a center crop). For experiments where high resolution crops are used, we first resize the image so its shortest side is 1024 pixels, and then take four random 512×512 crops (at test time, we take the four non-overlapping crops from a 1024×1024 center crop).

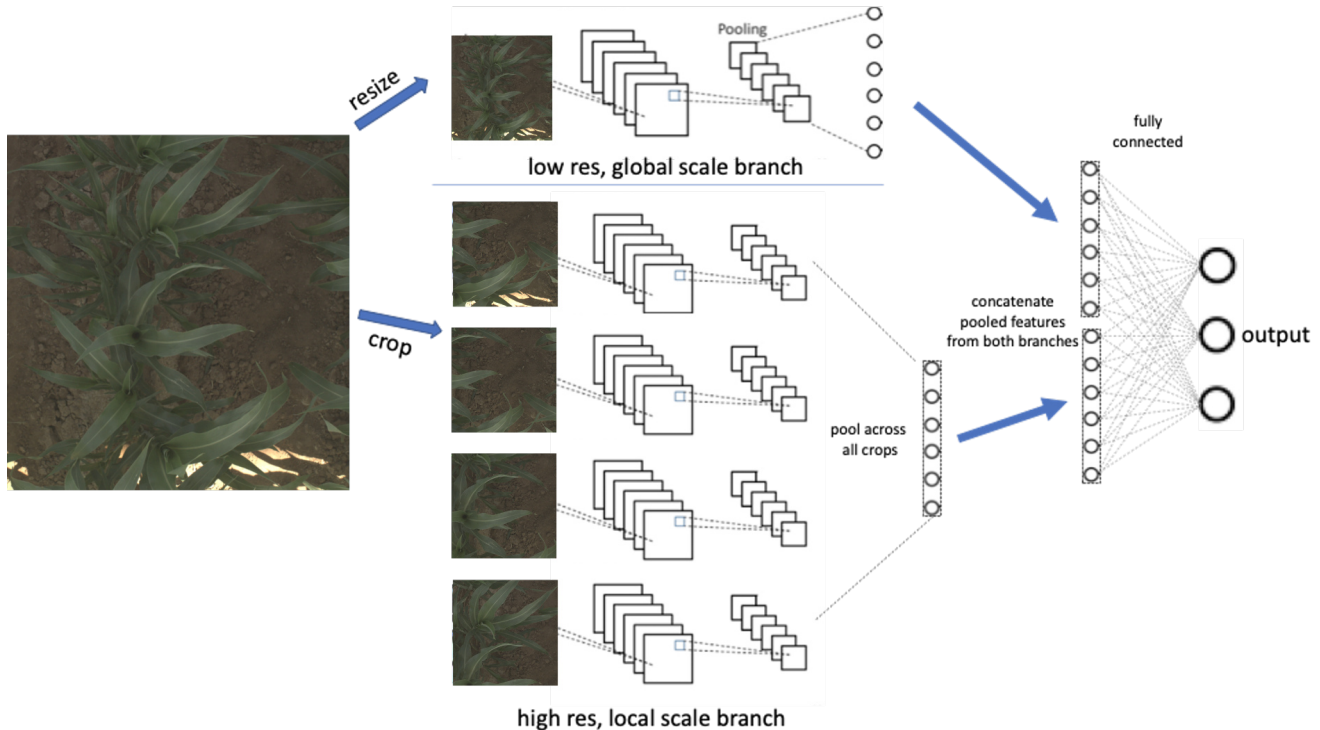We normalize by channel means and standard deviations

Figure 7: We propose a parallel network architecture that consists of both a low resolution, global scale branch, and a high resolution, local scale branch. While this diagram only shows a single convolutional layer followed by a global pooling operation, in reality we use a ResNet-50 trunk. In the high resolution branch, we take four random crops at original resolution, and perform global pooling across all four crops' final convolutional layers. The features from the two branches are then concatenated and fed through a final fully connected layer.

and perform random horizontal and vertical flips. For experiments using dynamic outlier pooling, we experimentally found a outlier pooling threshold of 2.0 yielded the best performance on this dataset.

In Table 1, we report the results of an ablation study of the different combinations of image resolutions, as well as different global pooling strategies. We tune the learning rate for each experiment, and report results after 20 epochs. The results show that using multiple high resolution crops, regardless of the global pooling strategy, outperforms using the more global low resolution crop of the whole plot, and that combining the high resolution and low resolution images outperforms training just the global or local scale networks individually. Additionally, we can see that our proposed dynamic outlier pooling approach significantly outperforms both global max pooling and global average pooling, regardless of the resolutions considered.

6. Conclusion

In this paper, we introduced the Sorghum-100 cultivar classification dataset which includes tens of thousands of images from 100 bio-energy lines of sorghum grown in the TERRA-REF field. This is the first gantry-based sorghum dataset, which has higher temporal and spatial resolution than similar UAV based datasets, making the data suitable for generating models for true high-throughput phenotyping in the field. While this paper presents the dataset in the context of cultivar classification, we hope to in the future release more data products and metadata, including data from additional sensors and both hand- and algorithm- generated phenotypes with the goal of supporting the development of machine learning models for more sophisticated high-throughput phenotyping tasks.

In addition, we presented a novel approach to training a deep convolutional neural network on the cultivar classification task, using a multi-resolution network architecture, and new global average pooling strategy called outlier pooling that is well suited to agricultural domains where the relevant features may be interspersed sparsely throughout the imagery. We performed an ablation highlighting the improvement in the classification task based on both the multi-resolution network architecture and the new outlier pooling approach, and show that the best performance is achieved when the two components are combined.

Pooling Crops (Resolution)	Average	Max	Dynamic Outlier
Whole Image (low res)	0.72137	0.64615	0.73829
Multi-crop (just high res)	0.74296	0.6518	0.76923
Multi-crop (high and low res)	0.76326	0.65486	**0.78788**

Table 1: In this table, we show the accuracy of different cropping and global pooling strategies on the Sorghum-100 test set. The performance of using high resolution crops is higher than using only one global, low resolution crop, but using both high and low resolution crops performs either individually. Additionally, using the Dynamic Outlier Pooling approach out-performs either Global Average Pooling or Global Max Pooling regardless of the resolution considered.

Possible modifications to the multi-resolution outlier pooling approach include having a larger discrepancy between the resolutions seen in the global and local scale branches, considering a feature pyramid approach that is more complex than simply two different resolutions, computing channel thresholds per batch rather than per image, and allowing varying threshold multipliers for different channels. While we look forward to exploring these questions in future work, the simple settings chosen in this work achieve significantly superior performance over baselines that consider only a single resolution or standard global pooling approaches.

References

[1] Zachary W Brenton, Elizabeth A Cooper, Mathew T Myers, Richard E Boyles, Nadia Shakoor, Kelsey J Zielinski, Bradley L Rauh, William C Bridges, Geoffrey P Morris, and Stephen Kresovich. A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy. *Genetics*, 204(1):21–33, 2016. 2

[2] Maxwell Burnette, Rob Kooper, J. D. Maloney, Gareth S. Rohde, Jeffrey A. Terstriep, Craig Willis, Noah Fahlgren, Todd Mockler, Maria Newcomb, Vasit Sagan, Pedro Andrade-Sanchez, Nadia Shakoor, Paheding Sidike, Rick Ward, and David LeBauer. TERRA-REF data processing infrastructure. In Sergiu Sanielevici, editor, *Proceedings of the Practice and Experience on Advanced Research Computing, PEARC 2018, Pittsburgh, PA, USA, July 22-26, 2018*, pages 27:1–27:7. ACM, 2018. 2

[3] Yuhao Chen, Javier Ribera, Christopher Boomsma, and Edward J Delp. Plant leaf segmentation for estimating phenotypic traits. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3884–3888. IEEE, 2017. 2

[4] Wei Guo, Bangyou Zheng, Andries B. Potgieter, Julien Diot, Kakeru Watanabe, Koji Noshita, David R. Jordan, Xuemin Wang, James Watson, Seishi Ninomiya, and Scott C. Chapman. Aerial imagery analysis – quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Frontiers in Plant Science*, 9:1544, 2018. 2

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 4, 6

[6] L. Herranz, S. Jiang, and X. Li. Scene recognition with cnns: Objects, scales and dataset bias. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 571–579, 2016. 6

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. 4

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 2012. 4

[9] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network, 2013. cite arxiv:1312.4400Comment: 10 pages, 4 figures, for iclr2014. 4

[10] Ali Masjedi, Jieqiong Zhao, Addie M Thompson, Kai-Wei Yang, John E Flatt, Melba M Crawford, David S Ebert, Mitchell R Tuinstra, Graeme Hammer, and Scott Chapman. Sorghum biomass prediction using uav-based remote sensing data and crop model simulation. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 7719–7722, 2018. 2

[11] Andries B. Potgieter, Barbara George-Jaeggli, Scott C. Chapman, Kenneth Laws, Luz A. Suárez Cadavid, Jemima Wixted, James Watson, Mark Eldridge, David R. Jordan, and Graeme L. Hammer. Multispectral imaging from an unmanned aerial vehicle enables the assessment of seasonal leaf area dynamics of sorghum breeding lines. *Frontiers in Plant Science*, 8:1532, 2017. 2

[12] Javier Ribera, Fangning He, Yuhao Chen, Ayman F Habib, and Edward J Delp. Estimating phenotypic traits from uav based rgb imagery. *arXiv preprint arXiv:1807.00498*, 2018. 2

[13] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 6

[14] Davinder Sharma, Jagadish Rane, Rajender Singh, Vijay Kumar Gupta, and Ratan Tiwari. Comparison of different planting methods to determine the precision of phenotyping wheat in field experiments. In *Advances in Plant & Microbial Biotechnology*, pages 77–83. Springer, 2019. 3

[15] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 4

[16] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5987–5995, 2017. 4

[17] Boxue Zhang, Qi Zhao, Wenquan Feng, and Shuchang Lyu. Alphamex: A smarter global pooling method for convolutional neural networks. *Neurocomputing*, 321:36–48, 2018. 4

[18] Z. Zhang, A. Masjedi, J. Zhao, and M. M. Crawford. Prediction of sorghum biomass based on image based features derived from time series of uav images. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 6154–6157, 2017. 2

[19] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning Deep Features for Scene Recognition using Places Database. In *Advances in Neural Information Processing Systems*, 2014. 6