

# Metric Learning for Large Scale Agricultural Phenotyping

Zeyu Zhang<sup>a</sup>, Abby Stylianou<sup>b</sup>, and Robert Pless<sup>a</sup>

<sup>a</sup>George Washington University, Washington, D.C., United States

<sup>b</sup>Saint Louis University, St. Louis, Missouri, United States

## ABSTRACT

We explore a metric-learning approach to create representations of sorghum images grown in a field setting. We train a convolutional neural network to embed images so that images from the same variety have similar features and images of different varieties have different features. We show that these features are good at discriminating unseen cultivars, can be used to predict standard phenotypes (height and leaf length and width), and can be used to predict presence or absence of genetic mutations. We evaluate these results using TERRA-REF data from field-scale trials of hundreds of varieties of sorghum. This demonstrates an end-to-end solution for creating useful image phenotypes in realistic field conditions.

**Keywords:** image phenotyping, representation learning, sorghum

## 1. INTRODUCTION

High throughput phenotyping in agriculture seeks to automate the measurement of properties of plants as they grow. For example, in field grown sorghum, measurements like plant-height, canopy cover, leaf length and leaf width, are visual observable phenotypes that may be measurable automatically. Characterizing these phenotypes over time in large scale field trials opens up opportunities for improving plant-breeding and better understanding of the genotype-phenotype relationships. Some analytics pipelines have been built on platforms like PlantCV<sup>1</sup> but classical approaches often struggle in field conditions.

Recently, convolutional neural networks (CNNs) provide an approach that can use the availability of large training data set to better accommodate real world conditions. CNNs have been the backbone of approaches to extract traditional phenotypes such as leaf counting and segmentation,<sup>2-4</sup> panicle counting,<sup>5</sup> and prediction of end of season biomass.<sup>6</sup> Most related is recent work on Latent Space Phenotyping,<sup>7</sup> which trains a CNN/LSTM combination on time-sequences of imagery labeled by different conditions (like drought stress or nitrogen deficiency). Using simulated data and datasets captured in controlled environments, they show that on unseen cultivars, they can predict the presence of these conditions quite well.

In our approach, we demonstrate that the embedding can be trained using only labels of the cultivar variety. Even with these simple labels, the features that are learned to distinguish a large number of varieties are also sufficient to predict a wide variety of phenotypic features and genetic variations in realistic field conditions.

## 2. APPROACH

### 2.1 Dataset

To test our method, We create a curated RGB-image dataset based on TERRA.<sup>8,9</sup> Our dataset contains about 58,000 images from 350 cultivars (varieties) of the sorghum Biomass Association Panel<sup>10</sup> of sorghum grown under the TERRA-REF gantry in 2017. Each cultivar was grown in two spatially separated plots, so there are 700 total plots. They are captured on 105 days over the course of the 140 day growing season (some days did not have data captured).

Raw images from the field RGB sensor are  $3296 \times 2016$  pixels. We then split the dataset into training/testing set where 315 cultivars (630 plots) were randomly chosen as training data, and 35 varieties (70 plots) were held aside for test data. There is no overlap of either plots or cultivars between the training and the test sets.

## 2.2 Network Architecture and Training

We follow the Proxy Loss metric learning<sup>11</sup> approach, implemented on a ResNet-50. We use 25 images in each batch, where each image is labeled by plot number. The data augmentation strategy is to randomly crop a  $1000 \times 1000$  sub-image and resize it to  $448 \times 448$  (with appropriate blurring and sub-sampling to create anti-aliased images). The  $448 \times 448$  image is randomly flipped horizontally and vertically. At test time, the same pre-processing is done except the cropping is always the center  $1000 \times 1000$  image region and there is no flipping.

We use the SGD<sup>12</sup> optimizer with an initial learning rate of 0.01 and a momentum term of 0.9. We train for 40 epochs, at which point the training loss has converged. We follow the loss function in,<sup>11</sup> where each class has a ‘desired’ location and neural network learns a function such that points from that class are close to the desired location. As our embedding, we follow the suggestion by Nam et al<sup>13</sup> to use the activation of the penultimate CNN layer as embedding features (as opposed to the final output) because the penultimate features typically generalize better to unseen classes. This creates 2048-dimensional features for each image.

## 2.3 Prediction Tasks

In this section, we illustrate how we use the image embedding for the three different tasks of cultivar retrieval, phenotype prediction, and genetic marker classification.

### 2.3.1 Cultivar Retrieval

We first consider the task of identifying the cultivar observed in a test image. We perform this task using the common gallery style image retrieval-based approach from metric learning – we split the test set into query set and gallery set where every test cultivars present in both sets, then extract features from our metric learning model for all images in the two sets. For any single image in the query set, we compute the distance to the features from all of the images in the gallery set, and sort by their similarity. We then infer the cultivar class from the class of the  $k$ -Nearest Neighbors. Since all of the cultivars in the test set are unseen during the training stage. This task presents the generalizability of our model to unseen classes, showing whether the neural network learns to extract features for cultivar differentiation outside of the training set.

### 2.3.2 Phenotype Prediction

In the task of phenotype prediction, we attempt to predict different ground truth phenotypes using our learned feature representations, with the idea being that a good representation that can differentiate between different varieties of sorghum presumably learned something about the most important visual phenotypes that define those varieties. In order to perform this prediction, we use our trained model to create feature representations for images in both training and test sets. We then create a per-plot, per-day feature,  $v$ , by max pooling all of the feature representations from a single plot on a single day. For each of the test per-plot, per-day features, we then perform weighted  $k$ -NN and support vector machine regression to infer the phenotype of interest.

### 2.3.3 Genetic Marker Classification

The final task that we perform is to predict genetic markers for images in the test set. There are five genetic marker families whose genotype  $\times$  phenotype relationship is well-documented in the sorghum genomics literature: leaf wax group;<sup>14</sup> the dw group;<sup>15,16</sup> the dry stalk (d) locus group;<sup>17</sup> the ma group;<sup>18,19</sup> and the tan group.<sup>20</sup> Each sorghum variety in our dataset is considered to be reference (if the genetic marker is in its default configuration) or alternate (if it is mutated in one or more of the known locations in this marker family). Sorghum is a diploid species, meaning it has two copies of each of its 10 chromosomes; of a variety is reference in one copy and alternate in the other copy we drop that variety in this analysis.

For genetic marker classification, we use the similar approach described in section 2.3.2 to predict the genetic marker. Since this is a classification task, in the  $k$ -NN model, after the nearest neighbors acquired, the genetic marker of neighbors vote to get the prediction (reference vs. alternate) of query plots; for SVM model, we use the RBF kernel SVM classifier for the prediction.

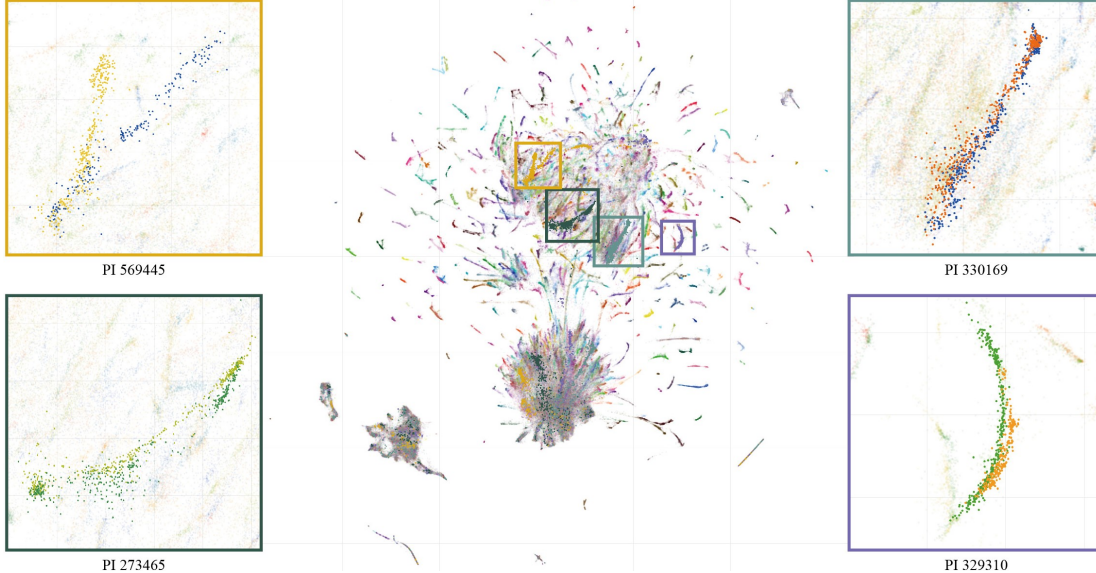


Figure 1. UMAP color coded by cultivars and examples of same cultivar color coded by different plot

### 3. RESULTS

#### 3.1 Embedding Space Visualization

In Figure 1, we show the UMAP visualization<sup>21</sup> of the features from all images in the dataset. The UMAP visualization projects the 2048-dimensional features into 2D, such that points that are neighbors in high dimensions remain neighbors in low dimensions. In the figure, we highlight four example cultivars in detail. In the zoomed view, they are colored by plot. Because the metric learning embedding function was only given *plot* labels during training – the fact that both plots from the same cultivar are mixed together indicates that the features encode features that are heritable and descriptive of sometimes subtle distinctions between similar cultivars.

#### 3.2 Cultivar Retrieval

For this task, we use the precision@ $k$  metric to evaluate performance. For each query image, there are  $k$  images retrieved. Every neighbor with same cultivar gives a score of 1, and 0 otherwise. Then this score is summed and divided by  $k$ . When using the GAP feature for this task, the precision@1 is 0.644 (Because there are 35 cultivars compared to chance performance of and precision@5 is 0.744. This shows the model can extract useful features to describe and identity unseen cultivars.

#### 3.3 Phenotype Prediction

In this task, we use the  $r^2$ -value as an evaluation metric. We consider three phenotypes that are available in the TERRA-REF dataset – canopy height, leaf-length, and leaf-width. The scores are shown in Figure 2a with both  $k$ -NN and SVM model.

It is instructive to look at the specific predictions and ground-truth value both at the scale of the whole season and for measurements during a day. We show this in Figure 2b. In these figures, the  $x$ -axis is the predicted result and the  $y$ -axis is the ground truth. The points are colored by date. In Figure 2b1, we enlarge the points from day 45 and 103 to emphasize the performance on single days. These plots all show the predictions from data driven features capture the overall trend of the phenotypes through the season well, even if the trend in a single day is somewhat weaker.

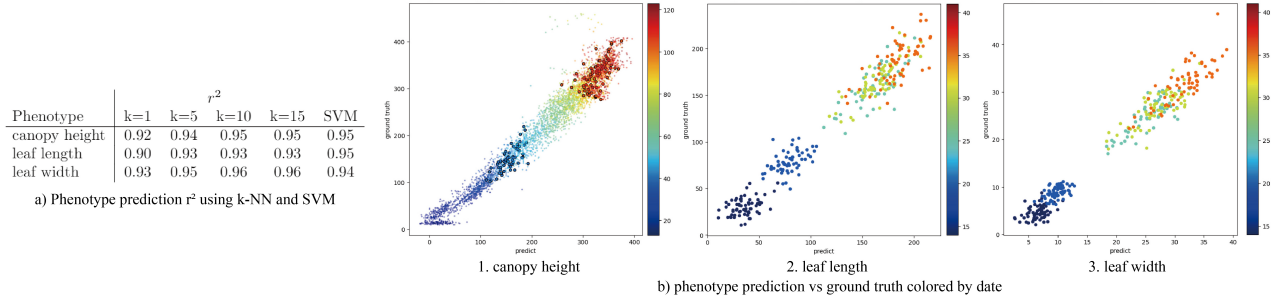


Figure 2. The phenotype prediction  $r^2$  results for canopy height, leaf length and width are shown on the left, for both  $k$ -NN and SVM regression model. On the right, we show the comparison of model prediction and the ground truth for these three phenotypes. Points are colored by date. Specially in figure b1, the results on day 45 and 103 are highlighted.

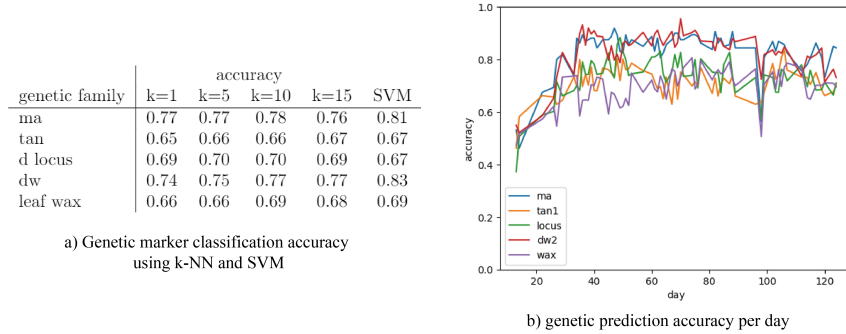


Figure 3. Genetic marker prediction results are shown on the left. Per day genetic marker prediction accuracy plot is shown on the right. It shows the accuracy goes up as the plant growing.

### 3.4 Genetic Marker Classification

For this task, we create a balanced test set for each genetic marker by selecting an equal number of reference and alternate cultivars from the test set for each marker to ensure that random chance performance on this task is 50%. The genetic marker classification results are shown in Figure 3a. We compare the result from  $k$ -NN model with different values of  $k$  and the SVM model. The results show that both classification approach work, but the SVM has consistently improved performance. This shows the embedding can capture the genotype  $\times$  phenotype relationship for different genetic markers using the visual features extracted from images.

In Figure 3b, we further illustrate the genetic marker prediction accuracy on each day. The result shows features extracted from image in early season have a harder time predicting the genetic maker. Accuracy increases over the course of the season, as the plants grow and the phenotypes controlled by each of the genetic marker families are better presented.

## 4. CONCLUSION

In large scale field trials, it is often the case that some labels come for free, such as the date of images, and the spatial plot labels. In this paper we explore the hypothesis that learning features tied to crop variety creates a rich descriptive space, and we demonstrate that this space can be used to extract common plant phenotypes and known genetic markers. For large, field scale plant trials (in contrast to greenhouse, or smaller scale controlled experiments), there may often be large amounts of visual data with limited labels, and exists already for many types of plants including wheat,<sup>22</sup> sunflower,<sup>23</sup> and lettuce.<sup>24</sup> Future research directions may include more exploratory work to understand and visualize what other known phenotypes of the plant are encoded within the features that are automatically learned, and potentially what novel phenotypes are available in the data that are not things that have been explicitly measured in the past.

## REFERENCES

- [1] Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., Doust, A. N., Feldman, M. J., Gilbert, K. B., Hodge, J. G., et al., “Plantcv v2: Image analysis software for high-throughput plant phenotyping,” *PeerJ* **5**, e4088 (2017).
- [2] Ayalew, T. W., Ubbens, J. R., and Stavness, I., “Unsupervised domain adaptation for plant organ counting,” in [*European conference on computer vision*], 330–346, Springer (2020).
- [3] Chen, Y., Baireddy, S., Cai, E., Yang, C., and Delp, E. J., “Leaf segmentation by functional modeling,” in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*], 0–0 (2019).
- [4] Giuffrida, M. V., Dobrescu, A., Doerner, P., and Tsiftaris, S. A., “Leaf counting without annotations using adversarial unsupervised domain adaptation,” in [*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*], 2590–2599, IEEE (2019).
- [5] Ubbens, J. R., Ayalew, T. W., Shirliffe, S., Josuttes, A., Pozniak, C., and Stavness, I., “Autocount: Unsupervised segmentation and counting of organs in field images,” in [*European Conference on Computer Vision*], 391–399, Springer (2020).
- [6] Masjedi, A., Carpenter, N. R., Crawford, M. M., and Tuinstra, M. R., “Prediction of sorghum biomass using uav time series data and recurrent neural networks,” in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*], 0–0 (2019).
- [7] Ubbens, J., Cieslak, M., Prusinkiewicz, P., Parkin, I., Ebersbach, J., and Stavness, I., “Latent space phenotyping: Automatic image-based phenotyping for treatment studies,” *Plant Phenomics* **2020**, 1–13 (01 2020).
- [8] Burnette, M., Kooper, R., Maloney, J. D., Rohde, G. S., Terstriep, J. A., Willis, C., Fahlgren, N., Mockler, T., Newcomb, M., Sagan, V., Andrade-Sanchez, P., Shakoor, N., Sidike, P., Ward, R., and LeBauer, D., “TERRA-REF data processing infrastructure,” in [*Proceedings of the Practice and Experience on Advanced Research Computing, PEARC 2018, Pittsburgh, PA, USA, July 22-26, 2018*], Sanielevici, S., ed., 27:1–27:7, ACM (2018).
- [9] LeBauer, D., Burnette, M. A., Demieville, J., Fahlgren, N., French, A. N., Garnett, R., Hu, Z., Huynh, K., Kooper, R., Li, Z., Maimaitijiang, M., Mao, J., Mockler, T. C., Morris, G., Newcomb, M., Ottman, M. J., Ozersky, P., Paheding, S., Pauli, D., Pless, R., Qin, W., Riemer, K., Rohde, G. S., Rooney, W. L., Sagan, V., Shakoor, N., Stylianou, A., Thorp, K., Ward, R., White, J. W., Willis, C., and Zender, C. S., “TERRA-REF, An Open Reference Data Set From High Resolution Genomics, Phenomics, and Imaging Sensors..” <https://datadryad.org/stash/dataset/doi:10.5061/dryad.4b8gtht99> (2020).
- [10] Brenton, Z. W., Cooper, E. A., Myers, M. T., Boyles, R. E., Shakoor, N., Zielinski, K. J., Rauh, B. L., Bridges, W. C., Morris, G. P., and Kresovich, S., “A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy,” *Genetics* **204**(1), 21–33 (2016).
- [11] Movshovitz-Attias, Y., Toshev, A., Leung, T. K., Ioffe, S., and Singh, S., “No fuss distance metric learning using proxies,” in [*Proc. International Conference on Computer Vision (ICCV)*], (Oct 2017).
- [12] Sutskever, I., Martens, J., Dahl, G., and Hinton, G., “On the importance of initialization and momentum in deep learning,” in [*International conference on machine learning*], 1139–1147, PMLR (2013).
- [13] Vo, N. and Hays, J., “Generalization in metric learning: Should the embedding layer be embedding layer?,” in [*2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*], 589–598, IEEE (2019).
- [14] Uttam, A., Madgula, P., Rao, Y., Tonapi, V., and Madhusudhana, R., “Molecular mapping and candidate gene analysis of a new epicuticular wax locus in sorghum (*sorghum bicolor* l. moench),” *Theoretical and Applied Genetics* **130** (10 2017).
- [15] Yamaguchi, M., Fujimoto, H., Hirano, K., Araki-Nakamura, S., Ohmae-Shinohara, K., Fujii, A., Tsunashima, M., Song, X., Ito, Y., Nagae, R., Wu, J., Mizuno, H., Yonemaru, J.-I., Matsumoto, T., Kitano, H., Matsuoka, M., Kasuga, S., and Sazuka, T., “Sorghum dw1, an agronomically important gene for lodging resistance, encodes a novel protein involved in cell proliferation,” *Scientific Reports* **6**, 28366 (06 2016).
- [16] Hilley, J. L., Weers, B. D., Truong, S. K., McCormick, R. F., Mattison, A. J., McKinley, B. A., Morishige, D. T., and Mullet, J. E., “Sorghum dw2 encodes a protein kinase regulator of stem internode length,” *Scientific Reports* **7** (7 2017).

- [17] Xia, J., Zhao, Y., Burks, P., Pauly, M., and Brown, P. J., “A sorghum nac gene is associated with variation in biomass properties and yield potential,” *Plant Direct* **2**(7), e00070 (2018).
- [18] Cuevas, H. E., Zhou, C., Tang, H., Khadke, P. P., Das, S., Lin, Y.-R., Ge, Z., Clemente, T., Upadhyaya, H. D., Hash, C. T., and Paterson, A. H., “The Evolution of Photoperiod-Insensitive Flowering in Sorghum, A Genomic Model for Panicoid Grasses,” *Molecular Biology and Evolution* **33**, 2417–2428 (06 2016).
- [19] Murphy, R. L., Morishige, D. T., Brady, J. A., Rooney, W. L., Yang, S., Klein, P. E., and Mullet, J. E., “Ghd7 (ma6) represses sorghum flowering in long days: Ghd7 alleles enhance biomass accumulation and grain production,” *The Plant Genome* **7**(2), plantgenome2013.11.0040 (2014).
- [20] Wu, Y., Li, X., Xiang, W., Zhu, C., Lin, Z., Wu, Y., Li, J., Pandravada, S., Ridder, D. D., Bai, G., Wang, M. L., Trick, H. N., Bean, S. R., Tuinstra, M. R., Tesso, T. T., and Yu, J., “Presence of tannins in sorghum grains is conditioned by different natural alleles of tannin1,” *Proceedings of the National Academy of Sciences* **109**(26), 10281–10286 (2012).
- [21] McInnes, L., Healy, J., and Melville, J., “Umap: Uniform manifold approximation and projection for dimension reduction,” *arXiv preprint arXiv:1802.03426* (2018).
- [22] Haghighattalab, A., Pérez, L. G., Mondal, S., Singh, D., Schinstock, D., Rutkoski, J., Ortiz-Monasterio, I., Singh, R. P., Goodin, D., and Poland, J., “Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries,” *Plant Methods* **12**(1), 1–15 (2016).
- [23] Gosseau, F., Blanchet, N., Varès, D., Burger, P., Campergue, D., Colombet, C., Gody, L., Liévin, J.-F., Mangin, B., Tison, G., et al., “Heliaphen, an outdoor high-throughput phenotyping platform for genetic studies and crop modeling,” *Frontiers in plant science* **9**, 1908 (2019).
- [24] Du, J., Lu, X., Fan, J., Qin, Y., Yang, X., and Guo, X., “Image-based high-throughput detection and phenotype evaluation method for multiple lettuce varieties,” *Frontiers in Plant Science* **11** (2020).