Pretraining on growth stage from images facilitates Sorghum Biomass prediction

John Fozard
John Innes Centre
john.fozard@jic.ac.uk

1. Introduction

Sorghum is an ancient cereal grain used for both human consumption and as animal feed. Improving crop yields through breeding requires phenotyping systems that are able to assess the growth and yield of crops under varying environmental conditions and agronomic management strategies [1]. The TERRA-REF [5] plant phenotyping system has a gantry covering more than an acre of crops, and is able to image plants over a whole growing season using a number of different modalities.

For the Sorghum Biomass Prediction Kaggle challenge, a large training dataset of 277327 images of growing sorghum plants from this platform were provided. These were from 176 different species, planted in 349 plots. The goal of this challenge was to accurately predict the amount of plant material at the end of the growing season (end-of-season dry biomass, biomass), using visible light images. Additional data about the date of image capture, days after planting , amount of sunlight accumulated at the field site, cultivar id, and the height of the camera above the ground were supplied. Submissions were evaluated using predictions on a separate test set of 19442 images from 110 plots. Whilst most of the training data contained images of the same plot over an entire growing season (22 – 140 days after planting), the test data was restricted to subsets of data early (22 – 43 days after planting) mid (61–83 days) and late (96 days onwards) in the season. Submissions were scored according to the weighted root mean squared error in their predictions of the biomass for these plots. Errors on early and mid season data were weighted by factors 3- and 2-times larger than those on late-season data.

Eight teams competed in this competition, with the best scores for each team's submissions being listed on the leaderboards. During the competition, the public leaderboard showed scores on 25% of the test dataset. At the end of the competition, rankings on the private leaderboard were displayed, using the remaining 75% of the test data, scores on which ranged between 9098.9 and 10634.2. Here we describe the methods used for the submission with the first place score on the private leaderboard.

2. Methods

To reduce computational demands, all (RGB) images were initially reduced in size to a resolution of 224×224.

In an attempt to extract more informative features from the image data, we performed domain-adaptive pretraining on an alternative task. We fine-tuned a model (plant-stage) to predict the number of days after planting, which is supplied for each image within the training set. The additional features within the training dataset were not used as, apart from the cultivar, they strongly correlated with the number of days after planting. The training set was restricted to those plots with biomass lying between the 40th and 60th percentiles, assuming that these would adopt a typical growth profile.

The plant-stage model had a straightforward architecture. For each image, the 512 features from the average-pooling layer of a ResNet-18 [3] network, initialized with weights from pretraining on ImageNet [2], were input into a 2-layer perceptron, with two nodes in the hidden layer and one output node. An exponential linear unit (ELU) activation function was applied after the first layer. The training data, grouped by plot, was split 80:20 into training and validation sets. The Adam [4] optimizer with an initial learning rate of 10^{-5} was used to minimize the mean-squared prediction error. The model was trained for 20 epochs, and the final weights were obtained from the epoch with lowest validation loss. Model and training hyperparameters were not explored during the competition.

The biomass prediction (bp) model, used for the actual task of the competition, contained the same ResNet-18 backbone. The 512 features from the average pooling layer were concatenated with the standardized values of four of the additional features (date captured, days after planting, camera height and accumulated heat units) and input into a two-layer perceptron with 64 hidden units and one output node. Again, an ELU activation was applied after the first layer. In an attempt to quantify uncertainty in predictions, another version of the model (bp-uq) with a duplicate two-layer perceptron branch was trained to predict the logarithm of the variance of the predictions.

For biomass prediction, the training dataset was split into early (below 55 days), mid (55–89 days) and late stages (90 days or more), and separate models were trained independently on these three subsets. Five-fold cross-validation was performed for each stage, with the data again grouped by plot before being split into folds. Image data was augmented using random cropping and resizing to 224×224, random horizontal flips, and random changes in brightness, contrast and saturation, and Gaussian noise of amplitude 0.1 times the input was added to the additional features. Again, the Adam optimizer was used, with initial learning rate 10^{-5}, and the model for each fold was trained for 10 epochs. Mean-squared error was used as the loss for training the bp model. For the bp-uq model, the maximum likelihood loss [6] was used.

Final biomass predictions for each image were made by averaging the predictions from all five folds. For bp, predictions for each plot in the test set were made by averaging the predictions for all images from that plot. For bp-uq, the weighted average of predictions was used, with weights given by applying a softmax operator to the inverse of the predicted variances for all the images from the plot.

3. Results and discussion

Typical predictions from the plant-stage model (after epoch 13) are shown in Figure 1A. For a subset of low and high-yielding plots, there seemed to be differences between the actual and predicted days after planting that were most apparent during the mid to late-season (around 100 days). Linear regression indicated a negative correlation between the mean of the error in the prediction over the interval 100-120 days and the biomass for each plot (Figure 1B). Despite only explaining a small amount of the variation in the data ($R^2 = 0.072$), the 95% confidence interval of the slope (-476 ± 224) did not contain zero.

Both the bp-uq and bp models rapidly overfit the training data, with the epoch with the smallest loss on the validation set being between the first and the fifth for all folds. The mean best epoch was 1.87 for bp-uq and 1.93 for bp. Evaluated on the test data, the bp-uq model scored 9098.9 (first place) on the private leaderboard, and 7742.5 (between 4th and 5th) on the public leaderboard. Whilst not submitted as one of the final entries to the competition, the bp model scored 9090.9 on the private leaderboard, and 8031.4 on the public leaderboard, suggesting that the addition of the uncertainty quantification branch was of minimal benefit.

To assess the value of pre-training the Resnet-18 on the plant-stage task, models bp and bp-uq were trained using the original ImageNet Resnet-18 initial weights. These gave leaderboard scores of 9436.9 (private) and 7944.9 (public) for bp, and 9504.7 (private) and 8052.8 (public) for bp-uq. Whilst these scores are worse than with pre-training on the plant-stage task, further experiments are required to determine whether this difference is statistically significant.

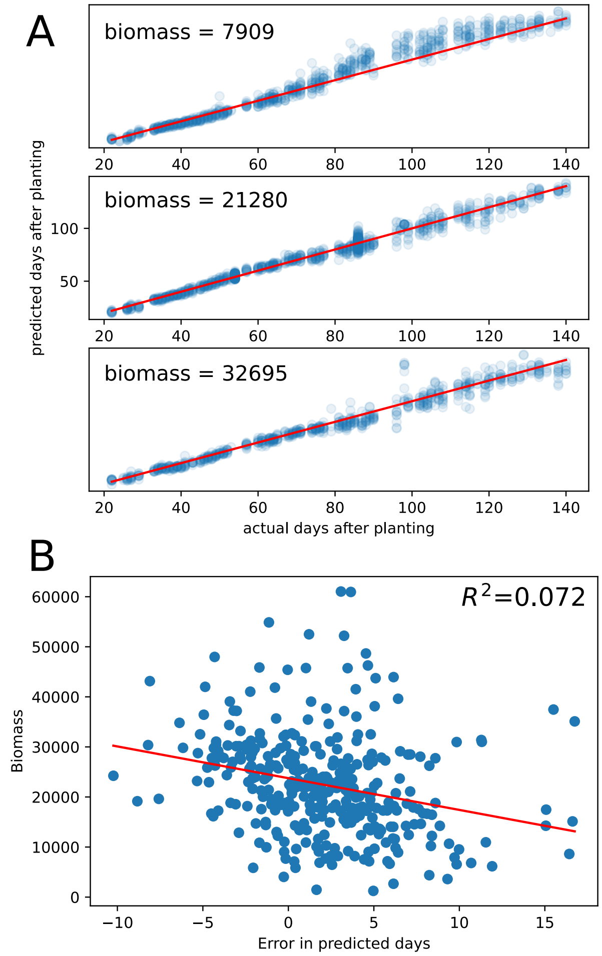

Figure 1. Results from plant-stage model. (A) Predicted vs actual days after planting for each image. Panels show results from three plots in the training set with different end-of-season dry biomass. Red line is the identity. (B) For each plot in the training set, end-of-season dry biomass against the mean error (predicted-actual) in the prediction of days after planting from the plant-stage model. Averages for each plot taken over images in the interval 100-120 (actual) days. Red line is the linear regression best-fit.

4. Conclusions

In this study, we presented a relatively straightforward model that was able to predict the end-of-season biomass. Domain-adaptive pre-training of the ResNet-18 backbone on the task of predicting the growth stage from images appeared to give models that made more accurate predictions than those using just initial weights from training on ImageNet.

It became clear that the problem was challenging, with training and validation losses indicating that models rapidly overfit the data. The substantial variation between scores on the public and private leaderboard suggests that further experiments are required to assess the effectiveness of this approach. It would be of particular interest to compare different pre-training tasks, such as predicting plant coverage or leaf area.

References

[1] Julia Bailey-Serres, Jane E Parker, Elizabeth A Ainsworth, Giles ED Oldroyd, and Julian I Schroeder. Genetic strategies for improving crop yields. *Nature*, 575:109–118, 2019. 1

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255, 2009. 1

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. 1

[4] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of 3rd International Conference on Learning Representations, ICLR*, 2015. 1

[5] David LeBauer, Maxwell A. Burnette, Jeffrey Demieville, Noah Fahlgren, Andrew N. French, Roman Garnett, Zhenbin Hu, Kimberly Huynh, Rob Kooper, Zongyang Li, Maitiniyazi Maimaitijiang, Jerome Mao, Todd C. Mockler, Geoffrey Morris, Maria Newcomb, Michael J Ottman, Philip Ozersky, Sidike Paheding, Duke Pauli, Robert Pless, Wei Qin, Kristina Riemer, Gareth Scott Rohde, William L. Rooney, Vasit Sagan, Nadia Shakoor, Abby Stylianou, Kelly Thorp, Richard Ward, Jeffrey W White, Craig Willis, and Charles S Zender. Terra-ref, an open reference data set from high resolution genomics, phenomics, and imaging sensors. dryad digital repository., 2020. 1

[6] David A. Nix and Andreas S. Weigend. Learning local error bars for nonlinear regression. In *Advances in Neural Information Processing Systems 7*, pages 489–496. 2