Automated Corn Ear Height Prediction Using Video-Based Deep Learning

Johnson Wong
School of Medicine
Indiana University
Indianapolis, USA
johnwong@iu.edu

Hao Sha, Mohammad Al Hasan, George Mohler
Computer and Information Science
Indiana University - Purdue University
Indianapolis, USA
{haosha,alhasan,gmohler}@iupui.edu

Steve Becker and Curtis Wiltse
Beck's Superior Hybrids
Atlanta, IN, USA
{steve.becker,cwiltse}@beckshybrids.com

***Abstract*—In corn breeding, hand-measurement of ear height is a labor-intensive process, thus limiting scalability. Here we show that it is feasible to automate estimation of the average ear height of a row of corn in experimental fields used for corn breeding. For this purpose we use point pattern analysis on predicted shank-node locations extracted from video captured on uncalibrated cameras moving through a plot at a fixed height from the ground (4 feet and 2 feet). First, a convolutional neural network-based object detection system (YOLOv3) was trained to detect the ear-stalk connection point and applied to the collected videos. Detected ear position and time information from each frame were super-imposed into a point pattern and point-features were then extracted. Using ridge regression to predict the average ear height per plot, we achieved 0.772 concordance, 2.989 inches root mean squared error, and 2.263 inches mean absolute error compared with hand-measured average ear height. Feature weight importance suggests that one camera may be sufficient for prediction without significant decrease in accuracy. This deep learning system can be utilized by mounting cameras onto the plot combine harvester to collect the necessary videos during harvest and could be expanded to quantify other phenotype measurements of interest that are labor-intensive to collect.**

***Index Terms*—precision agriculture, corn ear height, convolution neural network, yolo**

I. Introduction

Plant breeding [1] as a scientific discipline involves the utilization of a plant species' genetic diversity to develop new plant varieties for farmers, ranchers, foresters and gardeners. Any given plant breeding program will generate hundreds to tens of thousands of potential new plants or varieties each year. In Figure 1, a satellite image of an experimental grid of corn hybrids at Beck's is shown. Each grid cell contains a 2 row plot of corn corresponding to a particular strain and Beck's Hybrids runs trials in over 50,000 such plots across the U.S. each season.

Until recently, these new plants and/or varieties have been scored or measured visually for important traits such as plant architecture, disease/insect/abiotic stress tolerance, flower color, fruit shape, etc. Advancements in sensing technology and artificial intelligence have created opportunities to automate much of trait scoring or measurement. In corn breeding, hand-measurement of ear height is a labor-intensive process, thus limiting scalability. Corn height is an important trait, as corn stalks are more likely to bend and/or break if the corn ear is too high off the ground.

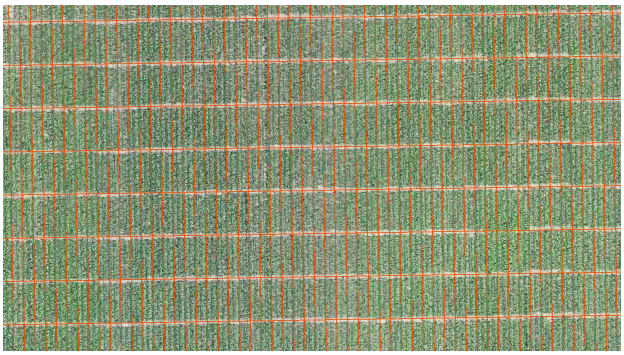

Fig. 1. Satellite image of experimental grid of corn hybrids.

In this paper we outline a prototype system for automating the measurement of corn ear height in experimental corn breeding yield trials. An overview of the system is provided in Figure 2. First, each two row plot corresponds to a different corn hybrid strain being tested. Video is then taken at a fixed height through each experimental plot. A "you only look once" (yolo) convolution neural network [2] is used to detect the node where the corn ear meets the stalk (shank node). A point pattern is then constructed by super-imposing the detected bounding box centers. Analysis of the point pattern is then employed to estimate the average ear height of experimental plot. This data can then be combined with genetic information for each strain to select new hybrids to test the following season.

The outline of the paper is as follows. In Sub-section I-A we provide an overview of related work on computer vision with applications to corn growth and precision agriculture. In Section II we provide details on our general methodology for corn ear detection from video and point pattern analysis to estimate average ear height per experimental plot. In Section III we provide accuracy results for our methodology and discuss the important features used in making predictions. Finally, in Section IV we discuss directions for future work.

A. Related work

Our work presented here is part of a growing body of literature on computer vision and deep learning methods for precision agriculture. Corn plant disease detection using

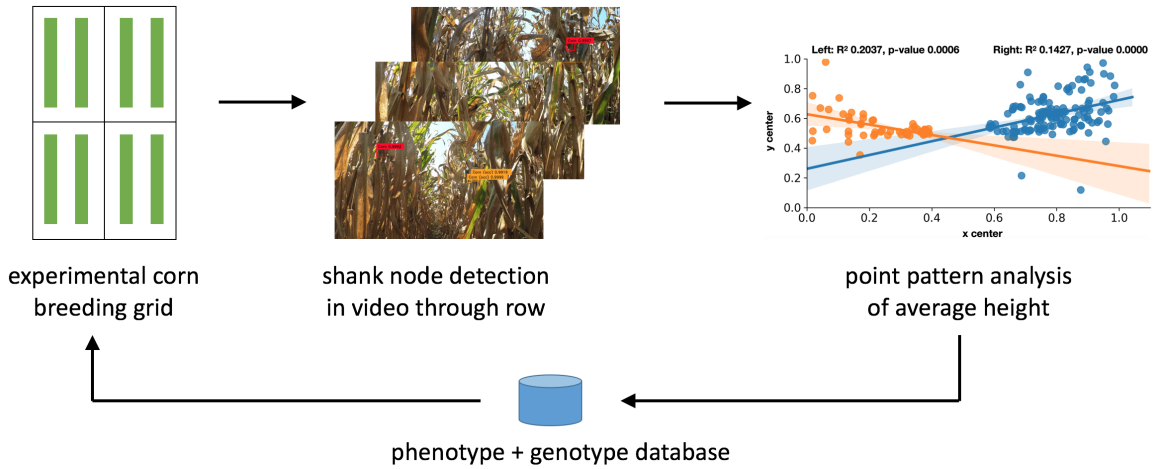

Fig. 2. (Left) Each 2 row plot corresponds to a different corn hybrid strain. (Center) Video through one experimental plot along with yolo-based detection of node where corn ear meets the stalk. (Right) Point pattern analysis of detected bounding box centers to estimate average ear height of experimental plot. (Bottom) Average ear height (phenotype) is combined with strain genotype and analyzed to select strains to test next season.

convolution neural networks is recently considered in [3] and in [4] broken corn detection is developed using yolo. In [5], moldy corn grain detection is considered and in [6] the authors develop methods for corn seeding monitoring. Methods have been developed for the detection of other plants, from strawberries [7], apples [8] and tomatoes [9] to weeds [10].

In other recent work, video from UAVs and satellite imagery are used to measure corn plant height [11] and crop yield [12], however the resolution in these works is much coarser than what is needed for corn breeding. Other research has focused on measuring corn height in indoor laboratory settings [13]. To our knowledge, this is the first work to develop an automated system for measuring ear height within the context of corn breeding at scale within a 2 inch margin of error.

II. Methods

A. Shank-node detection

As discussed in the previous section, corn ear height is an important characteristic to measure in corn breeding. While one approach would be to detect the corn ear in each frame of the video, this leads to errors as the orientation of the ear may not be fixed (some ears are upright, whereas others may face down). We therefore detect the node where the shank meets the stalk (see Figure 3). For this purpose we use a yolo-v3 convolution neural network [14]. Unlike CNN classification networks, yolo solves a regression problem for predicting both class probabilities (binary class indicating shank-node) and bounding box locations.

The yolo-v3 architecture includes 106 convolution layers. The layers predominantly utilize 3x3 filters and the network includes residual skip connections to avoid the vanishing gradient issue in very deep networks. The network performs object detection at three scales by down-sampling the input image by a factor of 8, 16, and 32. We also tested the use of binary classification CNN based models applied to sub-images to detect shank-nodes, however yolo-v3 provided superior accuracy with reduced computational cost.

We train yolo-v3 on 413 annotated images (245 $4'$-camera and 168 $2'$-camera). The training converges after 6,000 iterations, rendering $mAP@0.5 \sim 77.34\%$. Using a confidence threshold at 0.25, we obtained a training precision ~ 0.78, recall ~ 0.75, and F1-score ~ 0.77. For detection, we adopted a different confidence threshold at 0.1, for capturing more objects.

B. Average ear height estimation

Mapping each bounding box and confidence directly to the video's labeled height is one possible approach to estimating average ear height, but careful sample weighting to normalize for differing counts of bounding boxes per video would be needed. Instead, we aimed to characterize each video as a whole. Each bounding box was assigned a weight equal to the confidence (linear mapping to 0-1 weight) times the hann window, $w(t) = .5(1 - \cos(2\pi t))$, function based on the relative video frame index t ($0 \leq t \leq 1$, to decrease contribution of frames from beginning and end of video).

Next we super-impose the bounding-box centers outputted by yolo across the video and analyze the resulting point pattern (shown in Figure 5). Given each experimental plot consists of 2 rows and the camera passed through the middle, each side of the super-imposed 1920x1080 pixel frames produces a cluster around a line positioned at the shank-nodes. We derive two types of features:

1) The x, y centers of each bounding box (pixel coordinates) were calculated and grouped by video and left/right side of video (center determined by $x = 1920/2$). The x coordinate of points (horizontal image pixel index) from the right side was mirrored around the center to standardize points to the "left" side coordinate system. Each group of points were fit with linear regression. Slope, intercept, and covariances were taken as features.

Fig. 3. Yolo corn shank node detection. Video available at https://www.youtube.com/watch?v=6xZgesxGD9k

2) The x, y pixel coordinate bounding box centers were converted to polar coordinates using (960,270) and (960,540) as origins for camera heights 2 ft and 4 ft, respectively. These points were chosen by eye based on the center point distribution and could represent the horizon vanishing point. The x coordinate was again mirrored from the right side. The center of mass (using confidence weight) of the polar angle was calculated for the left and right sides, and the weights were also binned every 15 degrees using a histogram over the polar angle. After removing bins with 0 count across all videos, this gave 52 features.

We then use supervised learning based on these features to map the above statistics of the point pattern to the average ear height of each experimental plot (currently measured by hand at Beck's Hybrids across over 50,000 experimental plots each season. We fit a number of different sklearn [15] models including linear, lasso and ridge regression, KNN, XGboost [16] and random forest.

III. Results

In Table I we compare the root mean square deviation (RMSD), mean absolute error (MAE) and correlation between model predicted height and human labeled height on a held-out validation set. The best performing model is ridge regression with a mean absolute error of 2.263 inches.

In Figure 4 we plot the range of errors for each of the competing models. We find that 50% of predictions generally fall within 2.5 inches of absolute error and 90% fall within 5 inches. We note that the human labels are generated by taking several sampled height measurements in each plot rather than measuring each plant individually, so part of the error may be due to label noise.

In Table II we display the features used in the supervised models along with their correlations with average ear height. Here we observe that the 4ft camera features performed better than the 2ft camera, which may be due to the stage of growth (near harvest) where the height of the camera was closer to the height of the ears of corn. For detection earlier in the season the 2ft camera may yield better results. Based on random forest importance ranking, ultimately covariance and polar histogram features were removed, and using slope, intercept, and angle information resulted in the best model performance.

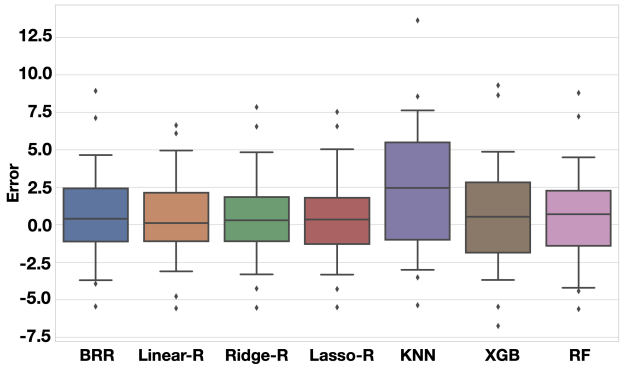

Fig. 4. Box and whisker plot of mean absolute error for competing models predicting average ear height per video.

Method	RMSD	MAE	CORR
Bayesian Ridge Regression	3.115	2.320	0.760
Linear Regression	3.064	2.394	0.762
Ridge Regression	**2.989**	**2.263**	**0.772**
Lasso Regression	3.007	2.286	0.769
KNN	4.960	3.921	0.368
XGBoost	3.671	2.782	0.622
Random Forests	3.271	2.540	0.743

TABLE I
MODEL ACCURACY FOR HEIGHT (INCHES) PREDICTION ON VALIDATION SET.

IV. Future Work

Our results indicate that automated corn ear height measurement is possible for large scale corn breeding within an error of 2 inches per experimental plot. Measurements can be taken using a camera fixed to the harvester or sprayer, or potentially via unmanned vehicles or drones. The framework used here may be applied to measuring other important traits such as yield per plot, percentage of broken stalks or ears on the ground, root lodging, stalk lodging, stay-green and intactness. Other sensors, such as lidar, may also be used in place of the video camera and may provide more accurate estimates in some situations. The methodology we outlined here is also

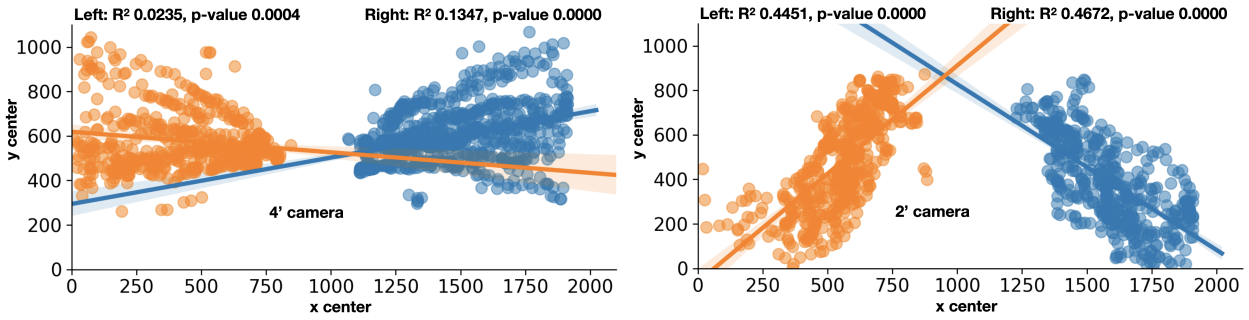

Fig. 5. Point pattern captured by $4'$ camera (left) and $2'$ camera (right), in the same lane. The markers are placed at the bounding box centers. Orange : points on the left half; Blue: points on the right half.

Feature	Kendall tau	Pearson r
4ft left intercept	-0.580	-0.721
4ft right polar angle	0.543	0.698
4ft right intercept	-0.518	-0.627
4ft left polar angle	0.512	0.683
4ft left slope	0.442	0.536
2ft right polar angle	0.439	0.625
2ft left polar angle	0.406	0.581
2ft right intercept	-0.369	-0.545
4ft right slope	0.304	0.431
2ft left intercept	-0.218	-0.384
2ft left slope	-0.072	0.000
2ft right slope	0.072	0.075

TABLE II
POINT PATTERN FEATURES AND CORRELATION WITH AVERAGE EAR HEIGHT.

not restricted to corn breeding and can be used for measuring important traits in other precision agriculture applications.

Further improvements to the model are also possible. Combining corresponding features from the left and mirrored right side per frame did not improve performance. It is possible that there was an overall tilt to the camera and left/right polar angles could be used to rotate the reference frame to enforce symmetry. Training to detect horizon or vanishing point could help with this as well, and be used to model a dynamic origin point for the polar coordinate method. Simple video stabilization was implemented but did not yield accuracy improvements. Additionally, if object tracking can be performed on bounding box positions over time, these paths could also be used to address one or both of the above issues.

Modeling on bounding box data directly was done with hard cutoffs for confidence and frame number. Since the model is essentially a series of two regressions, adapting a similar weighting scheme and normalizing by the bounding box count per video might yield improved results.

REFERENCES

[1] A. R. Hallauer, W. A. Russell, and K. Lamkey, "Corn breeding," *Corn and corn improvement*, vol. 18, pp. 463–564, 1988.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[3] S. Mishra, R. Sachan, and D. Rajpal, "Deep convolutional neural network based detection system for real-time corn plant disease recognition," *Procedia Computer Science*, vol. 167, pp. 2003–2010, 2020.

[4] Z. Liu and S. Wang, "Broken corn detection based on an adjusted yolo with focal loss," *IEEE Access*, vol. 7, pp. 68281–68289, 2019.

[5] W. Wang, X. Bai, X. Chu, H. Jiang, B. Jia, Y. Yang, and D. Kimuli, "Detection of moldy corn kernels using image processing technique based on the android operating system," in *2017 ASABE Annual International Meeting*, p. 1, American Society of Agricultural and Biological Engineers, 2017.

[6] B. Li, X. Liu, T. Zhang, F. Tian, Y. Tan, and S. Li, "Corn seeding monitoring system based on image processing," in *2018 ASABE Annual International Meeting*, p. 1, American Society of Agricultural and Biological Engineers, 2018.

[7] N. Lamb and M. C. Chuah, "A strawberry detection system using convolutional neural networks," in *2018 IEEE International Conference on Big Data (Big Data)*, pp. 2515–2520, IEEE, 2018.

[8] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved yolo-v3 model," *Computers and electronics in agriculture*, vol. 157, pp. 417–426, 2019.

[9] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "Yolo-tomato: A robust algorithm for tomato detection based on yolov3," *Sensors*, vol. 20, no. 7, p. 2145, 2020.

[10] M. H. Asad and A. Bais, "Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network," *Information Processing in Agriculture*, 2019.

[11] F. Furukawa, K. Maruyama, Y. K. Saito, and M. Kaneko, "Corn height estimation using uav for yield prediction and crop monitoring," in *Unmanned Aerial Vehicle: Applications in Agriculture and Environment*, pp. 51–69, Springer, 2020.

[12] R. Fieuzal, C. M. Sicre, and F. Baup, "Estimation of corn yield using multi-temporal optical and radar satellite data and artificial neural networks," *International journal of applied earth observation and geoinformation*, vol. 57, pp. 14–23, 2017.

[13] D. Shrestha, B. Steward, T. Kaspar, P. Robert, *et al.*, "Determination of early stage corn plant height using stereo vision," in *6th International Conference on Precision Agriculture and Other Precision Resources Management.*, pp. 1382–1394, 2002.

[14] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[16] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.