Interactive Deep Learning for Sorting Plant Images by Visual Phenotypes

Huimin Han[a], Ritvik Prabhu[a], Timothy Smith[a], Kshitiz Dhakal[b], Xing Wei[b], Song Li[b], and Chris North[a]

[a]Department of Computer Science, Virginia Tech
[b]School of Plant and Environmental Sciences, Virginia Tech

ABSTRACT

This paper proposes an interactive system called Andromeda that enables users to interact with machine learning models by sorting images in a reduced dimension plot. In our system, a dimension reduction algorithm projects the images into a 2D space representing similarities between the images based on visual features extracted by a deep neural network. With Andromeda, users can alter the projection by dragging a subset of the images into groups according to their domain expertise. The underlying machine learning model learns the new projection by optimizing a weighted distance function in the feature space, and the model re-projects the images accordingly. The users can explore multiple custom projections, and can export a model for future classification tasks. Our approach incorporates user preferences into machine learning model construction and allows reuse of pre-trained image processing models to accomplish new tasks based on user inputs. Using edamame pod images as an example, we transferred a maturity based model into a model that can classify number of seeds per pod to demonstrate the utility of our system.

Keywords: Deep learning feature visualization, visual back propagation, pre-trained deep learning model, dimension reduction, interactive visual analytics, computer vision

1. INTRODUCTION

In recent years, computer vision and artificial intelligence (AI) have played crucial roles in automating image-based decision processes in agriculture research and production. Many AI models, some implemented as phone apps, have been developed to diagnose plant diseases,[1] determine plant species,[2,3] and assess plant product quality.[4] However most published models follow a simple workflow of label-train-test-release cycle. Following this developmental workflow, a research group will collect samples, label samples, and use these labeled samples to train supervised machine learning models for either object detection or classification tasks. The goal of our project is to incorporate user perceptions in the model development through interactive machine learning.

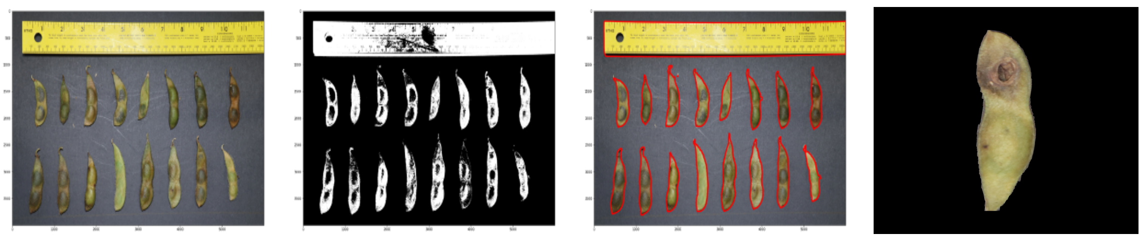
To address this need of incorporating user observations and perceptions into the machine learning processes, we have developed an interactive machine learning platform called Andromedawhere the end user can provide feedback to the machine learning models through regrouping the classified images. There are many potential usage cases for our interactive machine learning system. For example, our original machine learning model was trained to classify edamame pods based on their maturity stages include “diseased”, “ready to harvest” or “too late to harvest”.[4] However, while sorting the images for thousands of edamame pods, we also found many pods have different number of seeds per pod, which is important for consumer perception of the pod quality. Because the original model was only trained on the pod maturity and disease status, we could not reuse the model to predict number of seeds per pod. With an interactive system, we will be able to regroup the images to introduce new categories and reuse the trained model parameters for additional classification tasks. In the context of application of deep learning neural networks, this is particularly useful because a majority of parameters in the lower levels of the neural networks do not need to be retrained. Many of these parameters do carry useful information related to the shape and color of the pods. In this manuscript, using edamame as our model system, we demonstrate the function and utility of our interactive machine learning system.

Further author information: (Send correspondence to Chris North and Song Li)
Chris North: E-mail: north@vt.edu, Telephone: +1 540 231 2458
Song Li: E-mail: songli@vt.edu, Telephone: +1 540 231 2756

2. MATERIALS AND METHODS

2.1 Dataset and Preprocessing

Images used in this paper were collected by the Li Lab of Applied Machine Learning in Genomics and Phenomics at Virginia Tech.[5] This dataset comprises ready-to-harvest, late-to-harvest, and diseased pod images (100 images with 10-20 pods in each image). Figure 1 shows the sample raw data and image pre-processing results. We used an improved vegetation index, Excess Green minus Excess Red (ExG- ExR),[6] to identify pods for our data sets. ExR was subtracted from ExG with a zero threshold to create the ExG-ExR binary image. After computing a binary image from vegetation indices, we applied several morphological transformations.[7,8] We used dilation to increase the object area and closing and opening, which cleaned background noise by imputing missing pixel values. Finally, after vegetation indices and morphological transformations, we obtained a binary image mask with pods as white and background as black. Pods were detected by finding the contours of these masks.



(a) Sample raw data (b) Binary image (c) Object detection (d) Segemented disease instance

Figure 1. Sample raw data and preprocessing results

2.2 Feature Extraction and Visual Back Propagation

First, we use a deep neural network (DNN) to convert the images into meaningful quantitative representations.

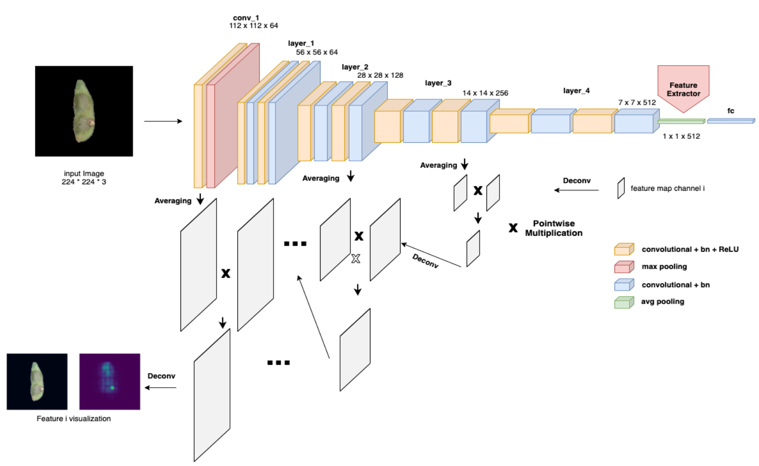


Figure 2. Resnet-18 Feature Extractor and Visual Back Propagation

- **CNN Representations**: Deep neural networks are a powerful tool to learn representations of data with multiple levels of abstraction. In particular, convolutional neural network (CNN) models are widely used in computer vision.[9,10] In this paper, we use the pre-trained ResNet-18 model from ImageNet,[11] which is one of the widely used CNN models to extract features from images. We use the last convolutional layer to extract 512 features from each image.[12]

- **Visual Back Propagation** To visualize each feature extracted from the ResNet-18 model, we utilize a modified version of the visual back propagation[13] method to visualize sets of pixels of the input image that contribute most to each feature. Starting from the 512 feature map from the last convolutional layer, we back propagate each feature and average the feature maps after each ReLU layer. The averaged feature map of the last convolutional layer is scaled-up via deconvolution and multiplied by the averaged feature map from the previous layer. The resulting intermediate mask is again scaled-up and multiplied. This process is repeated until we reach the input layer.

2.3 Dimension Reduction and Visual Analytics

To visualize similarities between images, we use a dimension reduction (DR) algorithm to project the 512-dimensional data into a 2-dimensional plot. To allow users to drag images and form new projections , we use an interactive framework called Andromeda.[14] A weighted Multi-Dimensional Scaling (MDS) algorithm with a weighted distance metric enables both forward and inverse projection. MDS projects the high-dimensional data to a 2D scatter plot.[15] A weighted distance function with user-specified weights on each dimension enables alternative projections that emphasize different dimensions. An inverse dimension reduction algorithm learns distance function weights for user-constructed layouts of the data points. Figure 3 shows the system design.

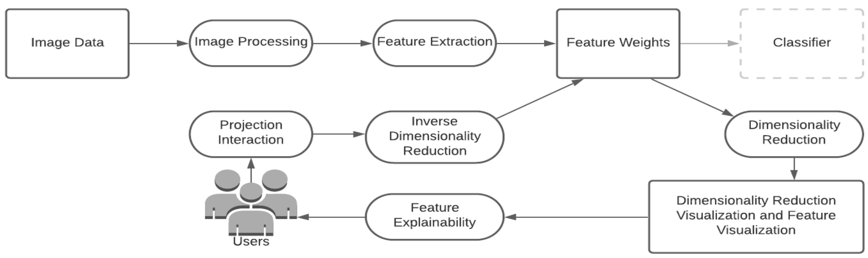


Figure 3. System Design

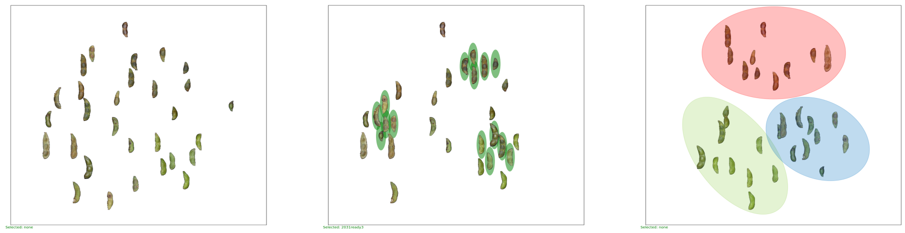
3. EXPERIMENTS AND RESULTS

These case studies validate the hypothesis that features extracted by deep neural networks can capture the abstractions specified by the users when dragging images in the plot. We also use visual back propagation method to uncover the learning process of the DNN model.

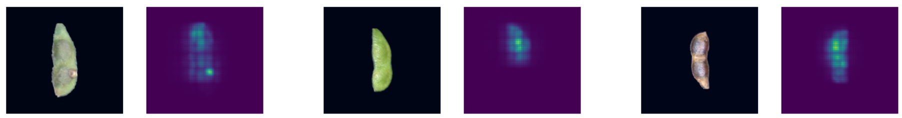
3.1 Classification of Pods Based on Maturity Stage

The maturity stage of each pod as either diseased, late-to-harvest, or ready-to-harvest is a feature easily determined by trained observers. The image data for 30 randomly chosen edamame pods are displayed on the 2D projection as shown in figure 4 (a). The initial weights for each feature are equal. In figure 4 (b) we interactively drag 15 pods highlighted in green in order to group them into 3 clusters according to desired phenotype categories. We assumed that, through this interaction, the underlying model would learn new weights for the feature spaces that satisfy the projection that we defined based on the stages of the pods. Figure 4 (c) shows the updated projection, which produced three main clusters of pods according to their maturity stage. The red cluster shows the pods that are too late to harvest, the blue cluster signifies the pods that are diseased and the green cluster signifies the pods that are ready to harvest. This indicates that the stage of each pod was effectively captured by the features extracted by the DNN model and successfully learned by the Andromeda model.

Furthermore, we can look at visualizations of the highly weighted features extracted by the DNN model on specific pods in order to understand what the most important visual features in our re-grouping represent to the model. In figure 4 (d) we see that one of the more important features used to determine a diseased pod highlights a salient discolored spot. Similarly, in figure 4 (e,f), areas of each pod correlating to important features are highlighted. This provides us with insight into which parts of the pod are important to the DNN for discerning a diseased, late-to-harvest, or ready-to-harvest product.



(a) Initial Layout (b) Regrouping (c) Updated Layout



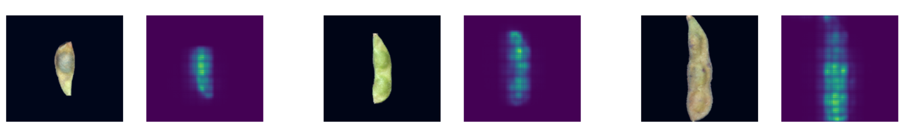
(d) Diseased pod (e) Ready-to-harvest pod (f) Late-to-harvest pod

Figure 4. Sorting by Phenotype: Disease, Ready-to-harvest, Late

3.2 Re-classification of Pods Based on Seed Numbers

The number of seeds per pod is an important feature that can potentially affect consumer acceptance of the product. However, the data were not originally collected to determine the number of seeds per pod. The number of seeds is a novel feature that could be observed directly by the end users but is not initially detected by the default projection. The chosen image data for 30 edamame pods are displayed on the 2D plot as Figure 5 (a) shows with equal initial weights applied for each feature. We interactively drag 15 pods highlighted in green to group them into 3 clusters as shown in 5 (b). We assume that by dragging a subset of the image data, the underlying model will learn the weights for the feature spaces that satisfy the projection we defined based on the number of seeds. Figure 5 (c) shows the updated projection. We find that "number of seeds" is well captured by the features extracted by the DNN model and learned by the Andromeda projection.

(a) Initial Layout (b) Regrouping (c) Updated Layout



(d) One Seed (e) Two Seeds (f) Three Seeds

Figure 5. Sorting by Phenotype: Number of Seeds

To explain the feature space with updated weights, we select the feature with highest weight as an example. In Figure 5 (d,e,f), the most relevant DNN feature mainly captures the overall shape of the pod to differentiate pods with different numbers of seeds.

These results indicate that Andromeda with features extracted by deep neural networks can indeed enable interactive sorting of pod images according to various human guided visual phenotypes. The resulting models could then be used as classifiers for larger collections of images. We plan to extend these methods to more complex input scenarios, such as images of pods on live plants captured in the field with mobile phones.

REFERENCES

[1] Mohanty, S. P., Hughes, D. P., and Salathé, M., "Using deep learning for image-based plant disease detection," *Front. Plant Sci.* **7**, 1419 (Sept. 2016).

[2] Barré, P., Stöver, B. C., Müller, K. F., and Steinhage, V., "LeafNet: A computer vision system for automatic plant species identification," *Ecol. Inform.* **40**, 50–56 (July 2017).

[3] Tan, J. W., Chang, S.-W., Abdul-Kareem, S., Yap, H. J., and Yong, K.-T., "Deep learning for plant species classification using leaf vein morphometric," *IEEE/ACM Trans. Comput. Biol. Bioinform.* **17**, 82–90 (Jan. 2020).

[4] Yu, D., Lord, N., Polk, J., Dhakal, K., Li, S., Yin, Y., Duncan, S. E., Wang, H., Zhang, B., and Huang, H., "Physical and chemical properties of edamame during bean development and application of spectroscopy-based machine learning methods to predict optimal harvest time," *Food Chem.* **368**, 130799 (Jan. 2022).

[5] Li, S., "Li lab of applied machine learning in genomics and phenomics." https://lilabatvt.github.io/.

[6] Meyer, G. and Camargo Neto, J., "Verification of color vegetation indices for automated crop imaging applications," *Computers and Electronics in Agriculture* **63**, 282–293 (10 2008).

[7] Raid, A., Khedr, W., El-dosuky, M., and Aoud, M., "Image restoration based on morphological operations," *International Journal of Computer Science, Engineering and Information Technology* **4**, 9–21 (07 2014).

[8] Gil, J. and Kimmel, R., "Efficient dilation, erosion, opening and closing algorithms," in [*ISMM*], (2000).

[9] Denker, J. S., Gardner, W. R., Graf, H. P., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D., Baird, H. S., and Guyon, I., "Neural network recognizer for hand-written zip code digits," in [*Proceedings of the 1st International Conference on Neural Information Processing Systems*], *NIPS'88*, 323–331, MIT Press, Cambridge, MA, USA (1988).

[10] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D., "Backpropagation applied to handwritten zip code recognition," *Neural Computation* **1**(4), 541–551 (1989).

[11] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," *CoRR* **abs/1512.03385** (2015).

[12] Bian, Y., Wenskovitch, J., and North, C., "Deepva: Bridging cognition and computation through semantic interaction and deep learning," in [*Proceedings of the IEEE VIS Workshop MLUI 2019: Machine Learning from User Interactions for Visualization and Analytics*], (2019).

[13] Bojarski, M., Choromanska, A., Choromanski, K., Firner, B., Ackel, L. J., Muller, U., Yeres, P., and Zieba, K., "Visualbackprop: Efficient visualization of cnns for autonomous driving," in [*2018 IEEE International Conference on Robotics and Automation (ICRA)*], 4701–4708 (2018).

[14] Zeitz, J., Dowling, M., Wenskovitch, J., Crandell, I., Wang, M., House, L., Leman, S., and North, C., "Observation-level and parametric interaction for high-dimensional data analysis," *ACM Trans. Interact. Intell. Syst.* **8** (June 2018).

[15] Cox, M. A. A. and Cox, T. F., [*Multidimensional Scaling*], 315–347, Springer Berlin Heidelberg, Berlin, Heidelberg (2008).