

Interactive Deep Learning for Exploratory Sorting of 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

^aDepartment of Computer Science, Virginia Tech

^bSchool of Plant and Environmental Sciences, Virginia Tech

ABSTRACT

This paper proposes an interactive system called Andromeda¹ that enables users to interact with machine learning models to allow for exploratory sorting of images through a cognitive approach that uses 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 to learn about the visual support for different groupings based on explainable-AI feedback. Our approach incorporates user preferences into machine learning model construction and allows transfer learning from pre-trained image processing models to accomplish new tasks based on user inputs. Using edamame pod images as an example, we interactively re-project the images into different groupings based on maturity and disease, and identify important visual features from the pixels highlighted by the model.

Keywords: Deep learning feature visualization, dimension reduction, interactive visual analytics

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 have been developed to diagnose plant diseases,² determine plant species,^{3,4} and assess plant product quality.⁵ Most published models follow a simple workflow consisting of a label-train-test-release cycle.

However, researchers sometimes need a more exploratory approach in which they interactively explore many alternative labelings based on their domain expertise. Thus, the goal of this paper is to create an algorithm that aids users in performing exploratory sorting on the fly, as they might sort physical seed pods on a tabletop, augmented with interactive machine learning that incorporates user perception feedback.

To dynamically incorporate user’s exploratory feedback into the machine learning processes, we developed an interactive machine learning platform within a computational notebook (Jupyter) called Andromeda. Users can provide feedback to the machine learning models through regrouping the projected images. First, we exploit transfer learning⁶ to extract image features using a popular pre-trained image-processing neural network. Second, the user is provided with a 2-D interactive projection where they can visualize similarities and re-group the images. Third, a machine-learning algorithm for inverse projection learns the revised similarities between the user-specified groups and renders a new projection of the images. It uses the interactive feedback from the user to learn a novel distance model that weights the visual features extracted by the network. Fourth, the explainable-AI highlights the pixels in the images corresponding to the up-weighted image features, thus providing visual justification for the user’s grouping. In this manuscript, using images of edamame pods as our model system, we demonstrate the functionality and utility of our interactive machine learning approach.

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.⁷ 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),⁸ 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.^{9,10} 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.

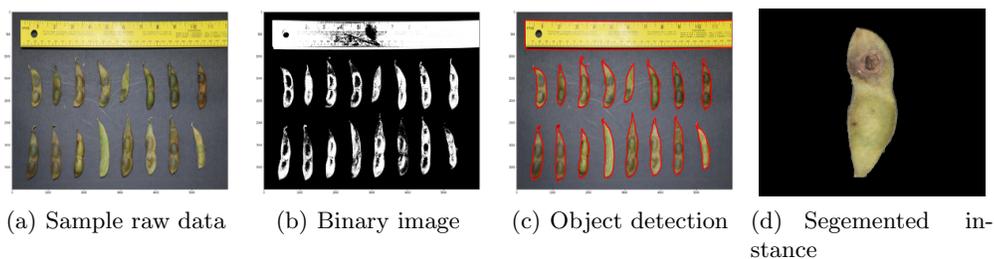


Figure 1. Sample raw data and preprocessing results for a diseased pod

2.2 Feature Extraction and Visual Back Propagation

We use a convolutional neural network (CNN) to convert images into meaningful quantitative representations.

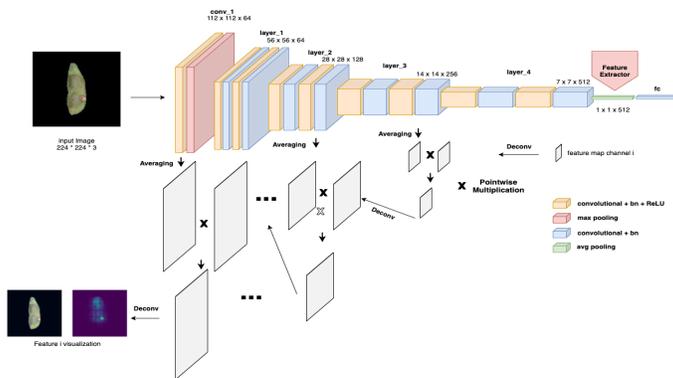


Figure 2. Resnet-18 Image Feature Extractor and Visual Back Propagation for Explainable-AI

- CNN Representations:** Convolutional neural networks are a powerful tool to learn representations of image data with multiple levels of abstraction. In particular, convolutional neural network (CNN) models are widely used in computer vision.^{11,12} Although any pre-trained CNN model could be applied, in this paper we use the pre-trained ResNet-18 model from ImageNet.¹³ ResNet was introduced in 2015 and won several competitions in computer vision¹⁴, and 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.¹⁵
- Visual Back Propagation** To visualize each feature extracted from the ResNet-18 model, we utilize a modified version of the visual back propagation¹⁶ 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 image layer. We initiate visual back propagation using feature weights from the interactive dimension reduction model, thus enabling explainability of the projection.

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.¹ A weighted Multi-Dimensional Scaling (MDS) algorithm with a weighted distance metric enables both forward and inverse projection. Although any dimension reduction algorithm could be applied (such as PCA or t-sne), MDS closely matches the user’s cognitive task of sorting images by mapping high-dimensional image similarities to 2D distances. Proximal images in the projection are similar in the weighted high-dimensional image feature space¹. MDS also easily adapts to different distance functions for experimentation. While MDS projects the high-dimensional data to a 2D scatter plot,¹⁷ a weighted distance function with user-specified weights on each dimension enables alternative projections that emphasize different dimensions. After interactive sorting, an inverse dimension reduction algorithm learns distance function weights for the user-modified projections of the images. Figure 3 shows the system design.

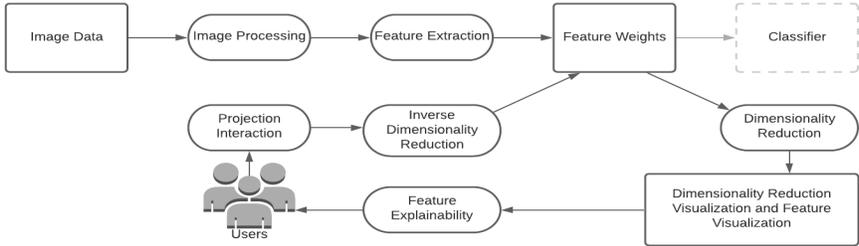


Figure 3. Andromeda System Design

3. INTERACTIVE CASE STUDY

These case studies validate the hypothesis that our interactive system with features extracted by the deep neural network can dynamically capture the novel abstractions interactively specified by the users when dragging images in the plot, and that the visual back propagation method provides explainability to uncover the important visual features identified through the interactive learning process that support the user’s abstractions.

3.1 Pods Based on Maturity Stage

The maturity stage of each pod as either diseased, late-to-harvest, or ready-to-harvest is a phenotype that can be determined by trained observers. Here we test if a user can sort the images according to these phenotypes with the help of the machine learning. The image data for 30 randomly chosen edamame pods are displayed on the 2D projection as shown in figure 4 (a); note that the color coding was added to show that the default plot does not capture the desired phenotypes as clear clusters. The initial weights for each feature are equal. In figure 4 (b) the user interactively drags 15 pods highlighted in green in order to group them into 3 clusters according to desired phenotype categories. 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 are diseased and the green cluster are ready to harvest. This indicates that the desired phenotypes of each pod were effectively captured by the weighted features and successfully learned by Andromeda.

Furthermore, the explainable-AI visualizations of specific pods depict the most important visual features in our re-grouping as learned by the interactive model. In figure 4 (d) we see that one of the more important visual features learned to determine disease phenotype is 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 for visually discerning a diseased, late-to-harvest, or ready-to-harvest product.

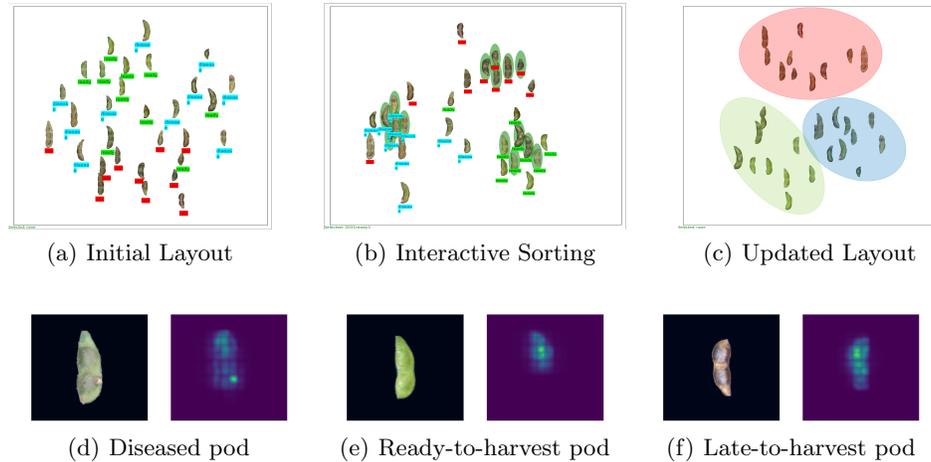


Figure 4. Interactive Sorting by Phenotype: Diseased, Ready-to-harvest, Late-to-harvest

3.2 Pods Based on Seed Numbers

The number of seeds per pod is an important phenotype that potentially affects consumer acceptance of the product. However, the images were not originally collected to determine the number of seeds. The number of seeds is a novel visual feature that can be observed directly by the end users but is not initially clustered in the default projection. Images of 30 edamame pods are displayed in the 2D plot in Figure 5 (a) with equal weights applied for each feature. The user interactively drags 15 pods highlighted in green to group them into 3 clusters according to number of seeds, as shown in 5 (b). Figure 5 (c) shows the updated projection with clear clustering. We find that the “number of seeds” phenotype is well captured by the weighted features learned by Andromeda.

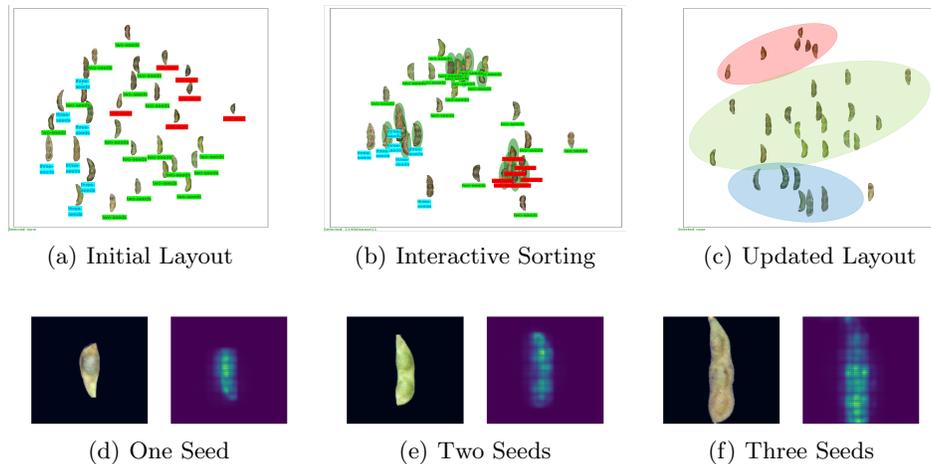


Figure 5. Interactive Sorting by Phenotype: Number of Seeds

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

These case study results indicate that Andromeda with features extracted by CNN deep neural networks can indeed enable interactive sorting of pod images according to various human-guided visual phenotypes. In future work, the resulting weighted feature 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] 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).
- [2] 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).
- [3] 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).
- [4] 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).
- [5] 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).
- [6] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q., “A comprehensive survey on transfer learning,” *Proceedings of the IEEE* **109**(1), 43–76 (2021).
- [7] Li, S., “Li lab of applied machine learning in genomics and phenomics.” <https://lilabatvt.github.io/>.
- [8] 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).
- [9] 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).
- [10] Gil, J. and Kimmel, R., “Efficient dilation, erosion, opening and closing algorithms,” in [*ISMM*], (2000).
- [11] 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).
- [12] 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).
- [13] He, K., Zhang, X., Ren, S., and Sun, J., “Deep residual learning for image recognition,” *CoRR* **abs/1512.03385** (2015).
- [14] Schaetti, N., “Character-based convolutional neural network and resnet18 for twitter author profiling : Notebook for pan at clef 2018,” (09 2018).
- [15] 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).
- [16] 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).
- [17] Cox, M. A. A. and Cox, T. F., [*Multidimensional Scaling*], 315–347, Springer Berlin Heidelberg, Berlin, Heidelberg (2008).