An Annotated Dataset for Apple Leaf Disease

Alakananda Mitra[1], Saraju P. Mohanty[2], and Elias Kougianos[3]

[1]*Nebraska Water Center, University of Nebraska-Lincoln, Lincoln, NE, 68588, USA.*
[2]*Department of Computer Science, University of North Texas, Denton, TX, 76707, USA.*
[3]*Department of Electrical Engineering, University of North Texas, Denton, TX, 76707, USA.*

Abstract

Image annotation is foundational for building accurate and effective AI models, especially in computer vision applications. However, annotating large datasets is time-consuming and resource-intensive. This study annotated 850 apple leaves from a publicly available dataset for three diseases. These annotated images can be used for image segmentation-based applications. We also tested the annotated dataset's accuracy by using ~~them~~ in a Mask Region-based Convolutional Neural Network (Mask-RCNN).

Keywords— Annotated Data, Precision Agriculture, Image Segmentation.

1 Introduction

Plant disease is an abiotic stress that negatively impacts crop quality and decreases crop yield. Early disease detection is crucial for disease management and prevention. Artificial intelligence (AI) shows great promise for automated disease detection with high accuracy and in no time by classification-, object detection-, and image segmentation-based methods. However, image annotation is crucial in AI, especially in computer vision. Image annotation labels the images used to train the machine learning (ML) models. However, this process is tedious and time-consuming. ~~here~~, we annotated apple leaves for disease detection.

2 Dataset Annotation and Pre-processing

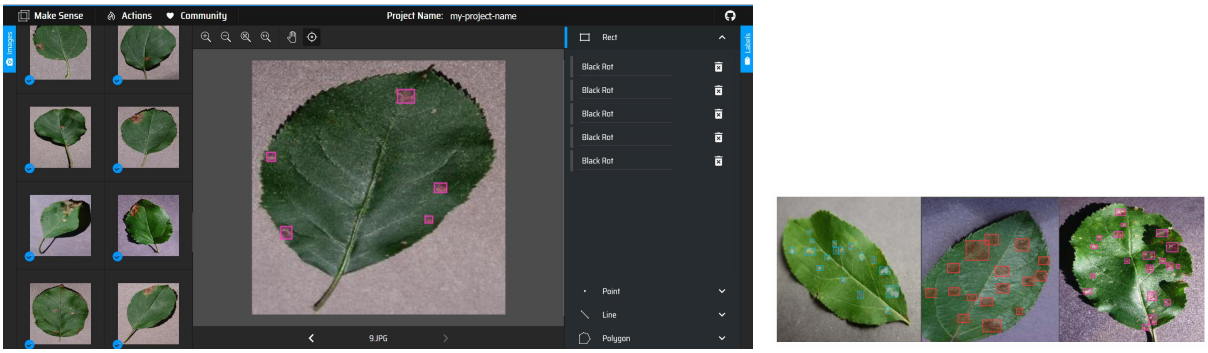
~~The dataset~~ annotation is the most arduous and time-consuming job in deep learning (DL) research, where images are used as input for the DL model. In this research, we ~~described~~ a dataset that we annotated. The original dataset is ~~publicly~~ available *PlantVillage* dataset [1]. The dataset has 3171 apple leaf images of 4 classes: Black Rot, Apple Scab, Cedar Apple Rust, and healthy. We annotated 850 randomly selected infected leaves.

~~Table.1~~ describes the annotated dataset details. A web-based open-source annotation tool [2] was used to annotate the data. ~~*RectLabel*~~ tool was used, and the two diagonally opposite corners' coordinates were saved in ~~*.xml*~~ file. Fig. 1(a) shows the annotation method, and some sample annotated leaves are shown in Fig. 1(b). The annotated data was randomly divided into *train* and *valid* folders at 70 : 30 ratio. We selected an additional 175 leaf images (no annotation) from the dataset for testing purposes.

To test the annotations' correctness, we used a Mask Region-based Convolutional Neural Network (Mask-RCNN) model. We also used image augmentation, such as horizontal and vertical flips, affine rotation, affine scaling, and edge detection.

Table 1: Annotated Dataset Details

Types	Symptoms	Number of Images		
		Total	Train	Validation
Black Rot	Initially small, purple-black lesions on leaf skin; then spherical spots + yellow-brown center and finally brown-purple rim.	300	210	90
Cedar Apple Rust	Initially, orange-red dots grow to orange-yellow patches with red edges; later, bright yellow dots.	250	175	75
Apple Scab	Initially yellow-green radial or circular patches; then brown to black with defined edges; serious disease when curled leaves; patches grow bigger by connecting.	300	210	90



(a) (b)

Figure 1: a) Annotation Tool. b) Sample of Annotated Images. [3]

3 Annotated Dataset Verification

To evaluate the accuracy of the annotated data, we used a pre-trained Mask-RCNN (on ImageNet) to detect apple leaf disease through instance segmentation. A learning rate schedule was used in the training. 83.8% *mean average precision* (mAP) was obtained with a backbone network *ResNet101* and *Feature Pyramid Network*. The detailed work was published in [3].

4 Conclusion

This abstract aims to help ~~the~~ researchers with the annotated dataset for apple leaf diseases for any image-segmentation-based work. This dataset can also be used to develop an automated annotation tool for plant disease detection.

Acknowledgement

This dataset was annotated for the study published in [3].

References

[1] Geetharamani G. and Arun Pandian J. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers and Electrical Engineering*, 76:323–338, 2019.

[2] Make Sense. https://www.makesense.ai/. ~~Accessd~~ on 13 June 2022.

[3] Alakananda Mitra, Saraju P. Mohanty, and Elias Kougianos. A smart agriculture framework to automatically track the spread of plant diseases using mask region-based convolutional neural network. In Luis M. Camarinha-Matos, Luis Ribeiro, and Leon Strous, editors, *Internet of Things. IoT through a Multi-disciplinary Perspective*, pages 68–85, Cham, 2022. Springer International Publishing.