May 18th, 2024

Corn Condition Identification - Report

By Esmeralda Hernandez

I worked with a corn disease image dataset that contained images of corn leaves that came from corn crops that were healthy or were infected with one of the following diseases: blight, common rust, or gray leaf spot. The three corn diseases have different causes as well as approaches to healing them. As such, I used Roboflow as a tool to perform object detection to create a model that can identify whether a corn plant was healthy or infected with one of the three diseases mentioned above from an image of the corn's leaf.

I used a total of 280 images to train and test my model which included 65 images of healthy corn, 60 images of corn with gray leaf spot, 70 images of corn with common rust, and 85 images of corn with blight. In Roboflow, I made 284 annotations in total and used 70% of the images to train, 20% for validation, and the remaining 10% of the images for testing my MS COCO model with a YOLO-NAS architecture which is used for object segmentation, and context recognition.

My initial hypothesis was that given the smaller image dataset, I would get a model with a precision of 75%. I was pleasantly surprised however to discover that the model I had developed has a mean average precision (mAP) of 97%, precision of 85.1%, and recall of 93.8%. The high precision signifies that my model performed well in predicting the true identification of the state of the corn leaves in the images. The high recall I obtained complements my high precision as it shows that my model does well in correctly classifying images of corn to their actual condition.

While I have good precision and recall for my model, I must recognize that my initial observation on my smaller image dataset proved to have negative effects. It became evident while manually testing my model that it was overfitted and not general enough to make correct identification predictions on images that were unlike those that were entered for each given class in my dataset. I uploaded 18 images to test the model's performance and 7 of the images were identified incorrectly. Unsurprisingly, due to the fact it was the least represented disease in my dataset, out of the 5 images of gray leaf spot corn that I uploaded, only one was predicted correctly. While running these test cases though, it became apparent to me that the way that the leaves were captured in the images contributed greatly to the overfitting problem as I noticed there were patterns in the images used to train the data that most likely aided in the misclassification of diseases. In the images below there is one image for each of the four classes in my model, as you can visually notice, each one has images taken at different angles, backgrounds, and different sizes which takes up space in the image. The images that I got from my dataset for healthy leaves were all caught very close to the leaf as shown in Figure 2. As such, the other three images I ran through my model that were healthy and similarly photographed were all predicted accurately to be healthy. Yet, when I included an image, such as the first one in Figure 1, that was more zoomed out compared to the others and was healthy, the model incorrectly identified this as blight. This was not surprising as when observing the images of corn with blight, there were more images, compared to the other classes, that had similarly captured the corn leaf in the image as the first image in Figure 1.

In the future, I would improve this model by ensuring that there would be more diversity in the angles and views for each distinct corn class, whether healthy or infected, to aid in improving the model overfitting problem. Additionally, I would add a larger dataset of images when creating a model, especially for one that is meant to distinguish between the four classes of healthy, common rust, blight, and gray leaf spot.

Figure 1

Figure 2