Implementing and Analysing Instance Segmentation in Plant Disease Detection

Abstract:
Over seventy percent of the population in the Indian villages relies on farming, more than eighty percent of whom are either small on marginal in their scope (FAO, 2020). Thus, it is paramount for the farmers to know if their crop is disease infested or not. The methods used thus far require a good experience in the field and are not completely reliable. In this project, we delegate this task to deep learning models by giving them a better input for easier classification, thus making the process automatic with more reliable results. This project aims to provide one solution to this problem by segmenting images before giving them as input to the classifier. And it was observed that providing convolutional neural networks with segmented images did improve the performance of the classification model.
Keywords- Image segmentation, detectron2, plant disease detection.

1.Introduction:

Since India is an agrarian economy it is heavily reliant on agricultural production. As a result, the detection of diseases in plants is of paramount importance to the farmers and the nation as a whole. In the traditional approach for plant disease detection, researchers used naked eye inspection to identify and diagnose plant diseases. To accomplish this, a large team of specialists is required, as well as continual inspection of the plant, which is extremely expensive when working with big farmland. Simultaneously, farmers in India lack proper facilities and even the expertise to meet experts. As a result, consulting consultants is both expensive and time consuming. Thus the usage of an automated disease identification technique is advantageous for detecting a plant disease in its early stages. Making this process automatic has been researched on for quite a few years. Many models have been developed with varying accuracies all the way from machine learning (Jian, et al., 2010, Li,et al., 2011) to the more recent deep learning models (Buyssens, et al., 2012, El-Sayed, et al., 2013) . One approach towards making this process simpler is to make the dataset much easier for the classifier to comprehend, this can be achieved by the process of segmentation. Splitting of an image into various sections is called image segmentation (Al-Amri, et al., 2010, Singh, et al., 2010) these sections are made in order to identify different objects that the model finds. This way, the useful part of the image can be separated from the background which would be extremely helpful for the database since it would make the classification process of the model in which the data would be imputed much easier. This process of image segmentation is not new to the field of plant disease detection and has been implemented in various models with different techniques. In this article, we will concentrate on instance segmentation (Silberman, et al., 2014, Zhang, et al., 2015) . The premise for the selection of this method was to see how well it would perform when with real life images which have information in the background that is irrelevant to the image which makes the classification process difficult. Semantic segmentation (Noh, et al., 2015, Long, et al., 2015) is the predecessor to instance segmentation, which classifies each pixel as belonging to a certain label but cannot do so for separate instances of the same entity. For example, if there are two cars in an image, semantic segmentation will assign the exact same label to every pixel of both the cars. Instance segmentation on the other hand overcomes this and can label uniquely for every occurrence of a particular object. This makes instance segmentation work well with images with multiple objects in it.

This paper consists of 4 parts, methodology, dataset, result and conclusion, and practical implications. The architecture that is implemented in this study is a Mask R-CNN derived model which is created by facebook-AI called Detectron-2, which is fed input images from a manually created dataset, the methodology and dataset sections expatiate the above. The resulting segmented images after testing the model and the difference in classification accuracy of the obtained segmented images against non segmented images is mentioned in the results and conclusion section. Finally, practical implications state how this model might be helpful for other image detection models and its future applications.

2.Literature Review:

A survey of the various classification techniques that have been implemented in the plant disease detection was done by Ghaiwat et al.. In the earlier models which used SVM for the classification process had difficulty to determine the parameters if the data used for the training was not linearly separable (Ghaiwat, et al., 2014).

A vision-oriented algorithm that is used in conjunction with green pixel masking and the colour co-occurrence process (Dhaygude, et al., 2013). A defined threshold level is used in the removal of green pixels while masking is applied and then the image gets segmented. Krizhevsky, et al. trained a Deep Convolutional Neural Network on ImageNet Dataset to obtain better results than previous machine learning models.

Arti N. Rathod et al. did a survey in which the otsu method (Rathod, et al., 2013) and the otsu thresholding method (Badnakhe, et al., 2012) in tandem with k means clustering, was used for segmentation.

Semantic segmentation gained popularity in various image recognition softwares thus it was tested in the plant disease detection field as well to measure its performance.
Wang, et al. suggested an updated FCNs model based on FCNs for crop disease leaf image segmentation. It is primarily composed of coding and decoding networks. The coding network is an upgraded variant of the long-established VGG-16 network. Training of the network was done by using sundry images of maize leaf lesions. The model then learns the features of the lesion spot of the maize leaf images and thus the segmented output is formed (Wang, et al., 2019).

Arsenovic, et al. implemented a PlantDiseaseNet network which is a two-novel architecture consisting of PDNet-1 and PDNet-2 trained concurrently. PDNet-1 is used for detecting plant leaves by type, while PDNet-2 is in favor of classifying certain plant leaves. The backbone architecture of the PDNet-1 was a pre-trained AlexNet. Because of its architectural design, the learned model obtained an accuracy of 93.67 percent (Arsenovic, et al., 2019) .

Chronologically speaking R-CNN was the first instance segmentation technique that had significant improvements in its results over its predecessors. Shi, et al. expatiated on the various applications of neural networks like Convolutional Neural Networks and Recurrent Neural Networks due to their potential for future implementations. After CNNs gained popularity for performing well for image processing using instance segmentation, Girshick, et al, proposed the R-CNN model by using selective search technique in region proposals for manually generated features and AlexNet CNN is used for classification of images.
Fast R-CNN (Girshick, et al., 2014) was implemented to overcome the limitations of R-CNN like taking a lot of time for training.It uses an additional Region of Interest (ROI) pooling layer to R-CNN. Zagoruyko, et al modified Fast R-CNN to Multi-Path Network By including skip connections and a foveal structure and an integral loss function (Zagoruyko, et al., 2015). Skip connections are used for accessing multiple layers and Foveal structure is used for exploring objects at various resolutions. An integral loss function is used for adjustment of the network which yields localisation.

Faster R-CNN was an improved version of R-CNN which reduced processing time even more by the addition of region proposal networks. Ren, et at., proposed an additional network called Region Proposal Network (RPN) to Fast R-CNN to get Faster-RCNN model (Ren, et al., 2017). RPN uses classifiers, regressors and anchors to generate proposals.
Ghoury, et al. identified diseased spots in real life images of grapes.It was carried out using images from the internet and the PlantVillage dataset. Single Shot Detector (SSD) MobileNet v1 and Faster R-CNN Inception v2 were the two pre-trained deep learning models that were used. The Faster-R-CNN Inception v2 model correctly categorised nearly 95.57 percent of all testing images, with a classification accuracy ranging from 78 percent to 99 percent (Ghoury, et al., 2019) .

Along with Faster-RCNN ,an object mask prediction branch is added to build the Mask-RCNN model (He, et al., 2018).This model is one of the best performed models and sets as a benchmark model for various image processing fields. Another benefit of Mask R-CNN is its ease of generalisation to other similar functionalities.

Based on annotated photos of wheat spikes and FHB disease areas, two Mask-RCNN training models were developed (Su, et al., 2020). Furthermore, Mask-RCNN generated satisfactory results for detecting wheat FHB disease, with average P, R, and F1-score scores of 72.10 percent, 76.16 percent, and 74.04 percent, respectively. Ultimately, 911 diseased spikes were identified from 922 tests, yielding a detection rate of 98.81 percent.

Ganesh, et al. introduced an avant-garde instance segmentation architecture for orange detection and its segmentation. Augmentation of Mask R-CNN coupled with HSV input results, a multi-modal deep learning approach is proposed (Ganesh, et al., 2019) . The established framework's output is validated using images captured in orange groves under natural lighting conditions.Using RGB+HSV images , the average F1 score is equivalent to 0.89.

3.Methodology

In this decade of various advancements in the field of deep learning, the methods which were consecutively introduced, show promise towards making the process of learning features from a given input much more reliable. Out of several applications of deep learning models object detection is quite popular. It is defined as determining the position of objects in an image as well as the group to which each object belongs. As a result, standard object detection model pipelines can be classified into three stages: informative region selection, feature extraction, and classification.Generic object detection is a type of object detection which is mostly used in deep learning models. The aim of generic object detection is to locate and identify actual objects in a single image and mark them with rectangular BBs to indicate their certainty of presence. The architectures of generic object detection approaches are mostly divided into two groups. The first method is standard object detection,which involves generating of various region proposals and thereafter labelling each proposal into distinct object groups. The other considers object detection to be a regression or classification task, employing a centralised structure to specifically attain final results (categories and position).

One notable deep learning technique is instance segmentation which is used in this study. Instance segmentation became the successor of semantic segmentation, by overcoming the weakness of semantic segmentation of not being able to give unique labels to two similar objects in an image. It is able to assign a label to every pixel of an image consequently being able to segment it.

After instance segmentation gained popularity due to its application potential, there have been multiple models dedicated to improving it. Fast R-CNN, and Faster R-CNN were the notable beginning two models that had significant improvement in performance. Faster R-CNN introduced the concept of RPN(Region proposal network). Here, the features of the image that were calculated by the initial convolutional neural network of the Fast R-CNN, were reused for finding the region proposals instead of running them multiple times, i.e, instead of running a separate selective search algorithm. Which made the process a lot more efficient, consequently reducing the time taken for the processing of the images which was a major drawback of Fast R-CNN.

The Mask R-CNN is one of the most efficient instance segmentation architectures, it is an evolution of the RCNN, Fast RCNN, and Faster RCNN methods. This system works in two stages: (i) the creation of region proposals, and (ii) the classification of each input image is processed by a convolutional network, also known as the backbone structure.

The backbone can differ depending on the optimal tradeoff between efficiency, training speed, and computational power limitations.produced proposal. The Mask-RCNN design is made up of two pathways: bottom-up and top-down. The bottom-up segment is in control of convolutions and feature map generation, and the most common structure is ResNets or ResNeXts. The more layers there are, the longer it takes to learn, but the performance is better, especially in complex object detection.In this paper we use ResNet50 FPN backbone.The outputs of Backbone network are feature maps that are used to create anchor boxes in the Region Proposal Network (RPN). Each area with a high likelihood produces anchor boxes with varying ratios and sizes. (The Region of Interest (ROI) is passed through ROI align, a quantization-free bilinear interpolation that retains spatial detail. These fixed dimension ROIs go through three parallel processes: (1) object class and probability; (2) bounding box; and (3) segmentation mask.

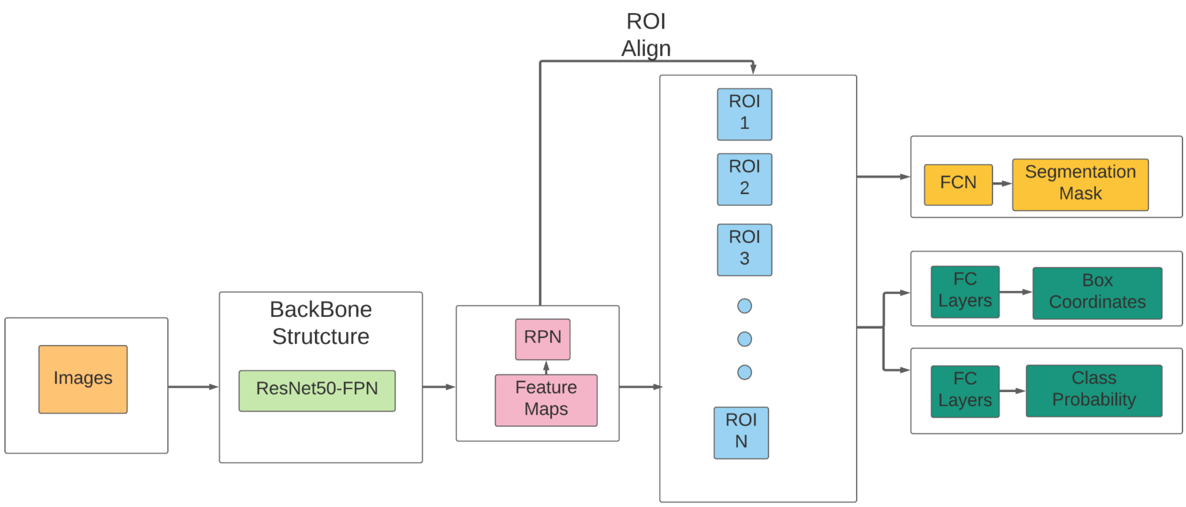

Figure. 1 Architecture of Mask R-CNN Model

The object detection algorithms that are implemented in detectron 2 are as follows:

3.1 Detectron 2

Detectron 2 was created by Facebook AI Research (FAIR) to provide an open source platform for further growth of the already blooming automated image detection research. The object detection algorithms that are implemented in detectron 2 are as follows:

- RPN
- Tensor mask

- Fast R-CNN
- DensePose
- PointRend
- Mask R-CNN
- Faster R-CNN
- and more...

The implemented detectron2 model contains ResNet-50 with Feature Pyramid Network (FPN). The model is iterated 720 times on the dataset for better accuracy with a learning rate of 0.0025.

3.2 Procedure

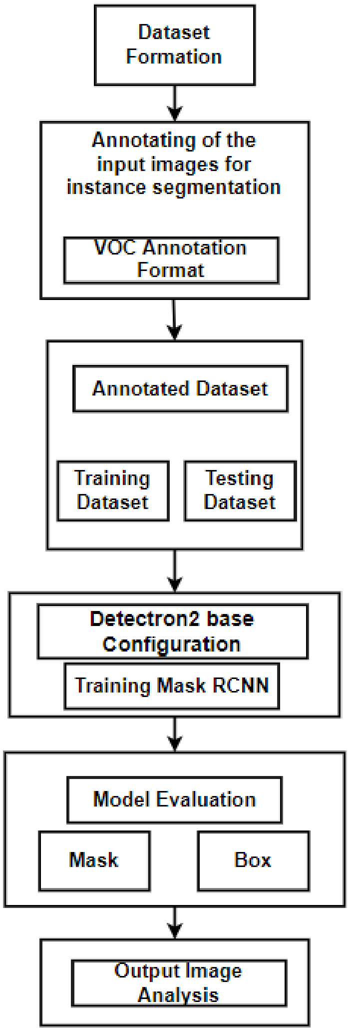

Figure.2 Working of the model

The flow chart of the working of the model presented in this study is shown in fig.2. The dataset collection is the first step, where annotation of the dataset according to the VOC format was done. This was followed by splitting the dataset into two parts, the training and testing sets. Next, the images were fed to the detectron-2 which is a derivation of the Mask R-CNN model, after the training of the model it learns how to segment images based on the training images given to it, consecutively the model is tested by using the testing dataset.

In our previous study, implementation of the ResNet-18 model was done which was used here to test the difference in result of segmented and non-segmented images during testing, and the analysis is done on the basis of the results.

4. Experimental Analysis

4.1. System Configuration

All of the experiments were implemented on a Tesla K80 runtime. The runtime was equipped with 2496 CUDA cores, 12GB GDDR5 VRAM.

4.2 Dataset

Annotations are the boundary of the object needed to identify the instance individually in an image. During the process of creation of the dataset used in this study, 70 images were extracted from an apple disease dataset from Plant Village(PV) dataset, which were then manually annotated. For creating the annotations, the VGG Image Annotator(VIA) tool is used.

The proposed model of this study is trained on 60 manually annotated images along with 10 images that are used to validate the model. Examples of the leaf images before and after annotation are shown in fig.3.

Figure.3 Above left images are before annotation and above right images are after annotation

5. Results and Conclusion

After training the Detectron-2 model with 60 annotated images, 10 images were used in the testing phase, some of the segmented images given as output can be seen in fig.4. It can be observed that there are a few minor inconsistencies in the segmentation process which would definitely improve with more training of the model.

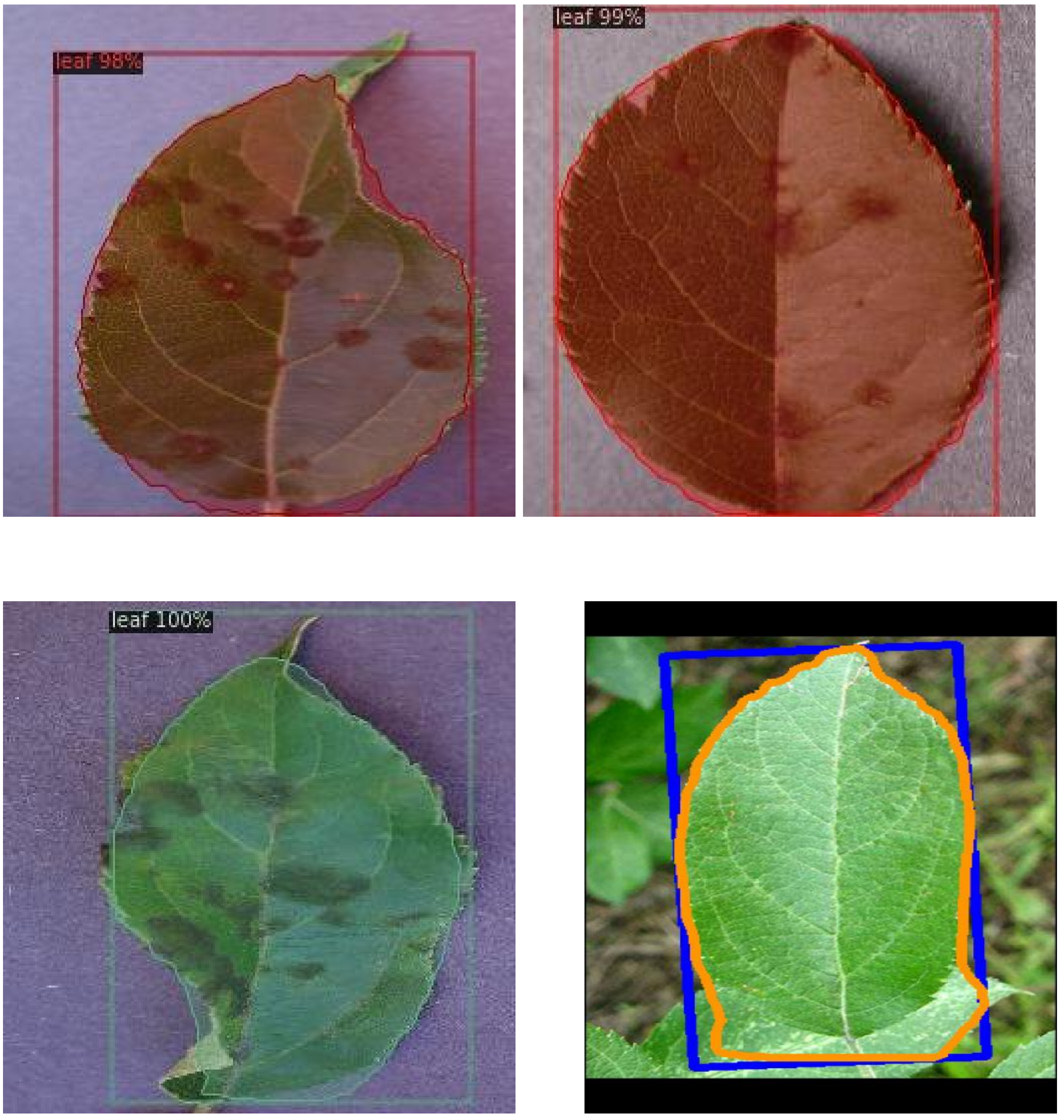

Figure.4 Segmented output images of the model

Also to check how these segmented images fare against non segmented images in the classification phase,the comparison of the accuracies of the ResNet model from our previous study when tested with images from a non segmented dataset that we got from Plant Village and the segmented images that we obtained from the detectron-2 model is done. Since the testing dataset for the segmented images is less (due to lack of publicly available annotated disease detection dataset) therefore the accuracy may vary if the dataset is bigger but would still perform better than a non segmented dataset during classification.

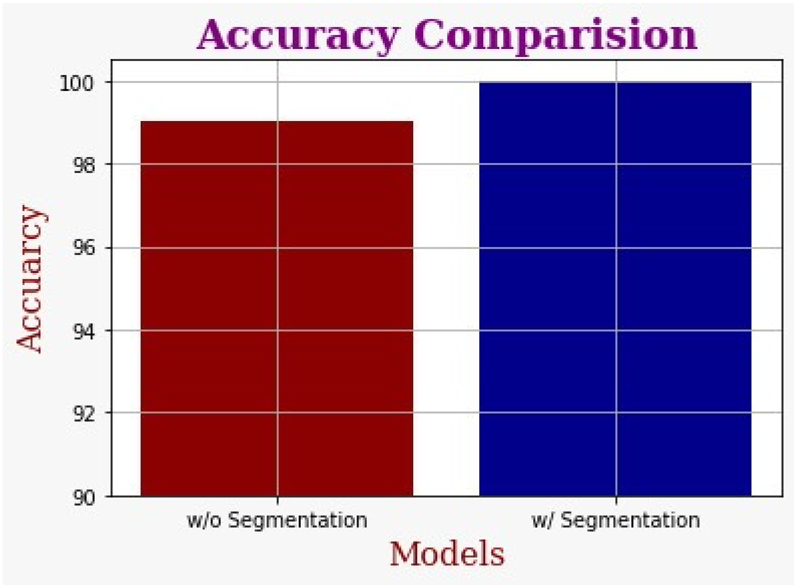

Figure.5 Accuracy Comparison between with segmentation and without segmentation models

From the confusion matrix of both the inputs we can see that the classification was amiss in one disease in the non segmented images (fig.6) and on the other hand it is perfect on the segmented images (fig.7). The ResNet network could fit the segmented images better than the non segmented images. The classification accuracy of the non segmented images was 98.9% whereas it was 100% for the segmented images.

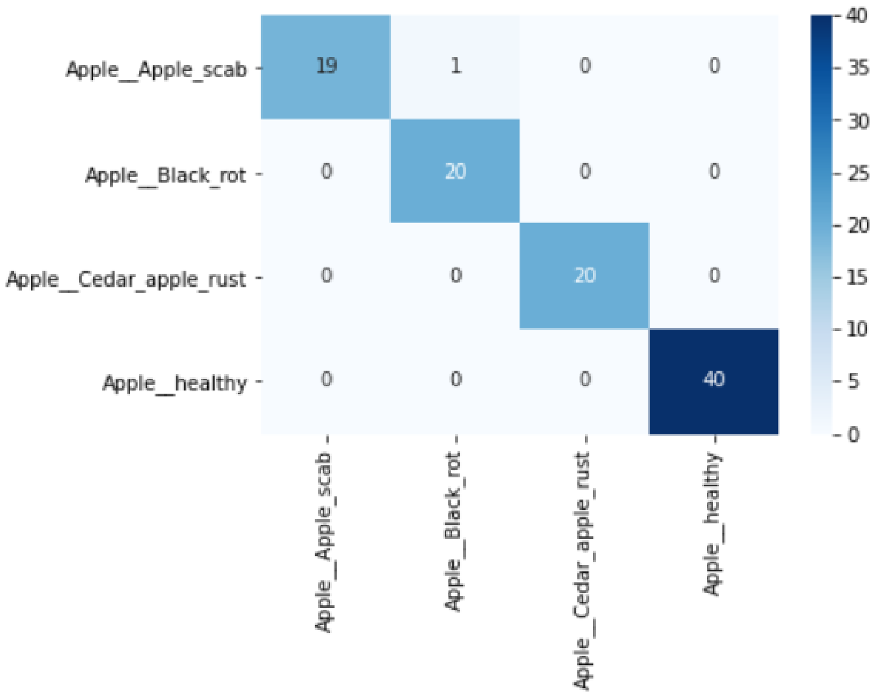

Figure.6 Confusion Matrix for Non-segmented images model

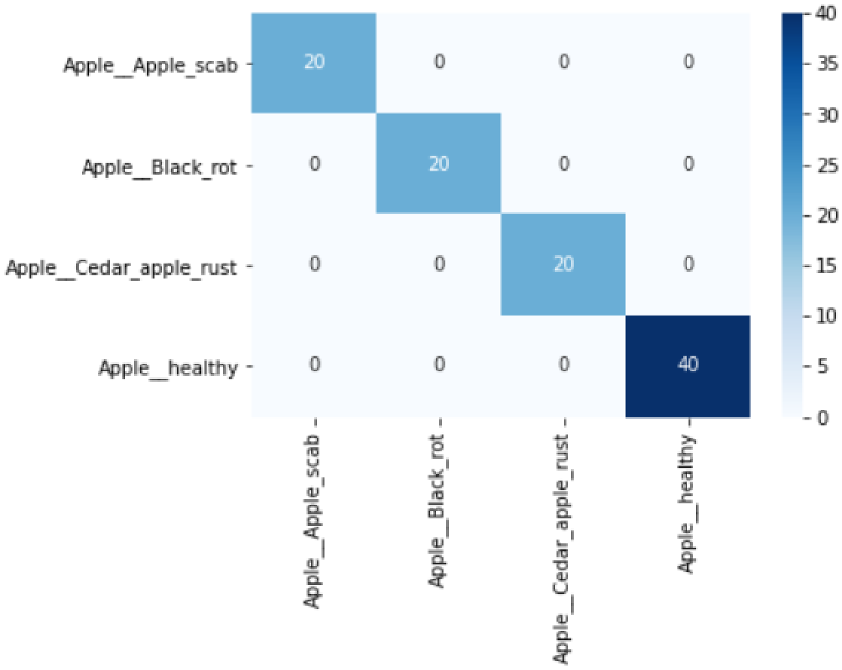

Figure.7 Confusion Matrix for Segmented images model

6.Practical Implications

The segmentation tool can be used for the training on any dataset but it would need to be annotated(with any format). Many image detection models fail when they are given real life images to classify and segmentation definitely helps in reducing the error margin by a huge margin. So this model can improve a dataset and thereby increase the classification accuracy of any model.

7.References

Al-Amri, S. S., & Kalyankar, N. V. (2010). Image segmentation by using threshold techniques. arXiv preprint arXiv:1005.4020.

Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., & Stefanovic, D. (2019). Solving current limitations of deep learning based approaches for plant disease detection. Symmetry, 11(7), 939.

Badnakhe, M. R., & Deshmukh, P. R. (2012). Infected leaf analysis and comparison by Otsu threshold and k-means clustering. International Journal of Advanced Research in Computer Science and Software Engineering, 2(3), 449-452.

Buyssens, P., Elmoataz, A., & Lézoray, O. (2012, November). Multiscale convolutional neural networks for vision–based classification of cells. In Asian Conference on Computer Vision (pp. 342-352). Springer, Berlin, Heidelberg.

Dhaygude, S. B., & Kumbhar, N. P. (2013). Agricultural plant leaf disease detection using image processing. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2(1), 599-602.

El-Sayed, M. A., Estaitia, Y. A., & Khafagy, M. A. (2013). Automated edge detection using convolutional neural network. Int. J. Adv. Comput. Sci. Appl.(IJACSA), 4(10).

Ganesh, P., Volle, K., Burks, T. F., & Mehta, S. S. (2019). Deep orange: Mask R-CNN based orange detection and segmentation. IFAC-PapersOnLine, 52(30), 70-75.

Ghaiwat, S. N., & Arora, P. (2014). Detection and classification of plant leaf diseases using image processing techniques: a review. International Journal of Recent Advances in Engineering & Technology, 2(3), 1-7.

Ghoury, S., Sungur, C., & Durdu, A. (2019). Real-Time Diseases Detection of Grape and Grape Leaves using Faster R-CNN and SSD MobileNet Architectures. In International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES 2019).

Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

Jian, Z., & Wei, Z. (2010, March). Support vector machine for recognition of cucumber leaf diseases. In 2010 2nd international conference on advanced computer control (Vol. 5, pp. 264-266). IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

Li, G., Ma, Z., & Wang, H. (2011, October). Image recognition of grape downy mildew and grape powdery mildew based on support vector machine. In International Conference on Computer and Computing Technologies in Agriculture (pp. 151-162). Springer, Berlin, Heidelberg.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE international conference on computer vision (pp. 1520-1528).

Rathod, A. N., Tanawal, B., & Shah, V. (2013). Image processing techniques for detection of leaf disease. International Journal of Advanced Research in Computer Science and Software Engineering, 3(11).

Ren, S., He, K., Girshick, R., Zhang, X., & Sun, J. (2016). Object detection networks on convolutional feature maps. IEEE transactions on pattern analysis and machine intelligence, 39(7), 1476-1481.

Shi, Z., & He, L. (2010, April). Application of neural networks in medical image processing. In Proceedings of the second international symposium on networking and network security (pp. 2-4).

Silberman, N., Sontag, D., & Fergus, R. (2014, September). Instance segmentation of indoor scenes using a coverage loss. In European Conference on Computer Vision (pp. 616-631). Springer, Cham.

Singh, K. K., & Singh, A. (2010). A study of image segmentation algorithms for different types of images. International Journal of Computer Science Issues (IJCSI), 7(5), 414.

Su, W. H., Zhang, J., Yang, C., Page, R., Szinyei, T., Hirsch, C. D., & Steffenson, B. J. (2020). Evaluation of mask RCNN for learning to detect fusarium head blight in wheat images. In 2020 ASABE Annual International Virtual Meeting (p. 1). American Society of Agricultural and Biological Engineers.

Wang, X. F., Wang, Z., & Zhang, S. W. (2019, August). Segmenting crop disease leaf image by modified fully-convolutional networks. In International Conference on Intelligent Computing (pp. 646-652). Springer, Cham.

Zagoruyko, S., Lerer, A., Lin, T. Y., Pinheiro, P. O., Gross, S., Chintala, S., & Dollár, P. (2016). A multipath network for object detection. arXiv preprint arXiv:1604.02135.

Zhang, Z., Schwing, A. G., Fidler, S., & Urtasun, R. (2015). Monocular object instance segmentation and depth ordering with cnns. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2614-2622).