
Cassava Leaf Disease Classification using Separable Convolutions UNet

Patike Kiran Rao^{1,2}, R Sandeep Kumar^{1,2}, Dr K Sreenivasulu²

1 Faculty of Engineering and Technology, Research Scholar @ M S Ramaiah University of Applied Science, Bangalore, Karnataka, India.

2 Department of Computer Science & Engineering, Assistant Professor @ G PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY, Kurnool, AP, India

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

Abstract: In this work, we develop and trained deep learning models for the segmenting and classification of cassava leaf disease as Blight or Mosaic. As the second-largest provider of carbohydrates in Africa, cassava is a key food security crop grown by smallholder farmers because it can withstand harsh conditions. At least 80% of household farms in Sub-Saharan Africa grow this starchy root, but viral diseases are major sources of poor yields. Our emphasis here was on two major cassava diseases that occur in Nigeria which are the Cassava Mosaic Disease (CMD) and the Cassava Bacterial Blight disease (CBBB). A total of 46 models were trained in five categories from over 21397 cassava leaf images was collected at different times of day containing leaves at different levels of symptom manifestation. One model diagnosed the healthy leaf and the other model detected the diseases that are present on the leaf when diagnosed as an unhealthy leaf and two most accurate models were exported. A 5-fold cross-validation was used to test the Separable Convolutions UNet model developed for health diagnosis and the Separable Convolutions UNet model developed for disease detection which yielded accuracies of 83.9% and 61.6% respectively.

1. Introduction:

Cassava is one of the most important staple food crops in Africa. Three continents, Africa, Asia and Latin America produce large amounts of cassava roots. Over 500 million people in the tropical world particularly Africa depend on cassava as one of their major staple foods. In Asia and Latin America, productions are largely used as raw materials for industries, as animal feed or for export markets. In Africa, the bulk of production is depended on as food by humans. Cassava is one of the most significant food security crops in some parts of sub-saharan Africa and has proved to be the most dependable crop in a number of countries as the last line of defence against famine. In the last decade, cassava is cultivated not just for human consumption in sub-saharan Africa but also to provide raw materials for emerging industries that depend on products from the roots, particularly starch.

The ability of cassava to thrive or do well on poor soils gives it an advantage over yam and the other root and tubers, grains or legumes in Africa. For many years to come cassava will continue to be an important source of carbohydrate to millions of people, particularly the rural and urban poor in Africa. Cassava, therefore, has to be managed more effectively than it is currently experiencing to increase its yield per unit area to ensure that more than enough roots in particular are produced to satisfy domestic, industrial and food security requirements at all times.

Among the factors that affect cassava production, diseases and pests still remain the major constraints that can bring Africa's cassava production to a halt. The recent East African Cassava Mosaic pandemic and the food shortages that resulted from it adds value to the above statement. African cassava mosaic disease is still widespread and causes severe yield losses in production systems that depend on susceptible cultivars. Cassava bacterial blight, anthracnose, bud necrosis, leaf spots and root rot diseases affect yields of cassava in almost all producing countries in Africa. Information on yield losses due to diseases are often based on estimates but observations indicate that losses are significant in most of the cassava growing areas of Ghana.

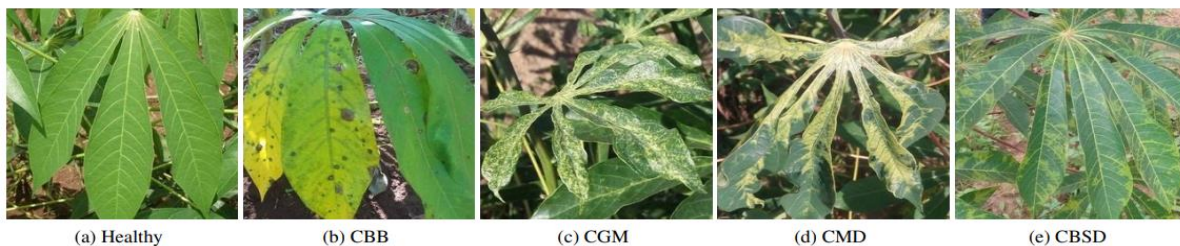
What is the Importance of Cassava Diseases?

- Diseases cause low yields of edible roots.
- Low yields due to diseases affect incomes of farmers.
- Food security is reduced by diseases.
- Severe outbreaks of diseases such as cassava bacterial blight can result in famine (in whole communities or countries).
- Cassava diseases that affect stems can lead to loss or shortages in the supply of planting materials.
- Loss of leaves through diseases can affect the availability of leafy vegetables.

- Loss of leaves and poor yield of storage roots can affect livestock production in communities that use cassava as animal feed.

Unfortunately, however, cassava farmers in most producing countries in Africa do very little or nothing to control diseases and pests of the crop. An impression that exists among cassava farmers (particularly in Ghana) is that cassava cuttings will give some root yields no matter where they are planted even if no attention is paid to the plants. Also, it is common to meet farmers who regard symptoms of certain diseases of cassava as signs of plant maturity. These incorrect impressions need to be corrected to increase yields of cassava. Farmers must be made aware of diseases and their importance and why it is necessary that diseases must be controlled.

This text is, therefore, written to introduce cassava as a crop with diseases that need to be controlled to increase yields to meet demands for consumption, food security and raw material requirements of industry. Symptoms of diseases have been shown or described to make disease identification an easy exercise. Actions that can be undertaken to control specific diseases have been described. Agriculture extension agents, farmers and students of agriculture will find this small text a very useful guide to controlling diseases of cassava.



2. Related Work:

In this section, we describe in detail the major previous works that motivated our work.

2.1. Separable Convolutions

Separable Convolutions were initially introduced by [1] and then later implemented by [1]. Separable convolution is a form of factorization which factorizes a standard convolution into a depth wise convolution and a point wise convolution (1×1 convolution). A standard convolution layer works by applying a convolution kernel to all of the channels of the input image and takes a weighted sum of the input pixels covered by the kernel sliding across all input channels of the image. This means that for a standard convolution, no matter how many input channels are available, the output channel is one. However, in depth wise separable convolutions, features are only learned from the input channels so the output layer has the same number of channels as the input.

This is known as depth wise convolution followed by a point wise (1×1) convolution layer which computes the weighted sum of all output channels into a single output (Figures 1 and 2).

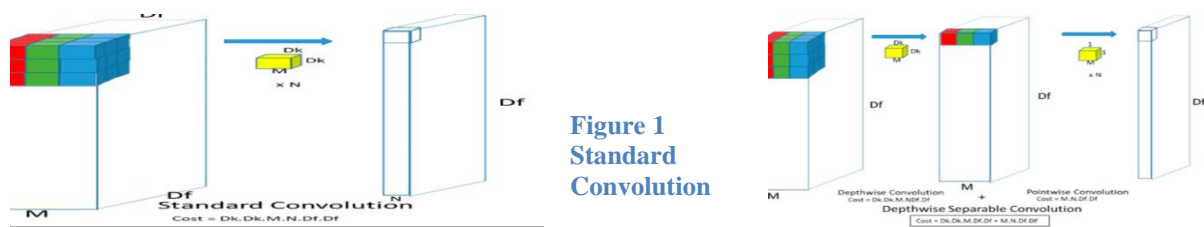


Figure 2 Separable Convolution

2.2. Fully Convolution Networks(FCNS)

The most fundamental idea behind FCNs [10] is that they are only made up of locally connected layers (convolution, pooling, and up sampling) without fully connected or dense layers. This tends to reduce the time required for computation and the number of parameters. It also means that an FCN will work regardless of the input image size. FCNs are typically made up of:

- Down sampling/Contraction/Encoding Path: On this path, the model extracts and interprets the contextual information on the input image.
- Up sampling/Expanding/Decoding Path: The specific localization or construction of segmentation maps from the extracted context in the encoding path.
- Skip Connections/Bottlenecks: Combines information from encoding and decoding paths by summing feature maps

2.3. U-Net

The U-Net architecture is designed as an improvement of the FCN architecture specifically for the segmentation of medical images. The major difference between U-Net and FCN is U-Net is symmetrical and the bottleneck layers that combine information from the encoding and decoding paths do so by concatenating the feature maps whereas they are summed in the FCN architecture. The encoding path of U-Net is made of four blocks each containing two 3x3 unpadded convolutions with a ReLU activation layer and a 2x2 max-pooling layer. The number of feature channels is also doubled after each down sampling step but the size of feature maps is reduced due to max-pooling. The decoding path contains 2x2 up sampling with 3x3 standard convolutions. Each convolution is followed by a concatenation of features from corresponding layers in the encoding path. This helps to transfer the localization information that is learned during down sampling from the encoding to the decoding path.

3. Materials and Methods:

In this section, we outline the proposed technique, describe the Separable-UNet architecture and experiments conducted.

3.1. Setup

The training was based on the Keras with a Tensorflow backend as the deep learning framework on a work station enabled with an NVidia Tesla K40c GPU (12GB memory) and Intel ® Xeon (R) CPU E5-2603 V4 @ 1.70 GHz with 12CPUs. CuDNN 7.0 library was used with the benchmark function enabled to ensure that the fastest algorithms are used.

3.2. Datasets

The datasets used to evaluate the performance of Separable-UNet are the Kaggle challenge dataset for the segmentation of neuronal structures in electron microscopic (EM) stacks [40,41] and the MSD challenge brain tumor segmentation dataset [42].

3.3. Kaggle Challenge Dataset

The cassava leaf images were taken with a commonly available Sony Cybershot 20.2-megapixel digital camera in experimental fields belonging to the International Institute of Tropical Agriculture (IITA), outside of Bagamoyo, Tanzania. The entire cassava leaf roughly centered in the frame was photographed to build the first dataset (Figure S6). Over a four-week period, 11,670 images were taken. Images of cassava diseases were taken using several cassava genotypes and stages of maturity (as described in Table S1) in order to provide the full range of symptoms for each given disease to the deep learning model. Each of the diseases or types of pest damage was distinctive and the variation of symptom expression between varieties was minor in comparison to the contrasts between diseases.

3.4. Data Pre-Processing

The resolution of the images of the Kaggle challenge EM stacks is originally 512×512 but was resized to 256×256 due to computational limitations. Data augmentation techniques were used due to the small number of available training images. A smaller number of images might lead to a concept known as over fitting where a trained model performs very well on training data but performs poorly on new test data. These augmentation techniques included horizontal flip, zoom range, height and width shift range. The number of images of the EM stacks dataset after augmentation increased to 120. The resolution of images in the BRATs dataset (240×240) was also reduced to 144×144 . Center cropping and normalization of data to ensure 0 mean and unit variance was also employed and the original 3D slices converted to 2D slices for training and testing of Separable-UNet. In all, there are 25347 casava leaf image samples.

3.5. Separable U-Net

Separable U-Net follows a similar architecture as U-Net with a few a few modifications. Except for the first convolution layer which has a standard convolution, all other convolution layers are made of depth wise separable convolution layers. Separable U-Net architecture with convolution encoder and decoder using separable convolution based U-Net architecture. It can construct the model thinner and less computationally expensive.

The original depthwise separable convolution is the depthwise convolution followed by a pointwise convolution.

1. Depthwise convolution is the channel-wise $n \times n$ spatial convolution. Suppose in the figure above, we have 5 channels, then we will have $5 \times n \times n$ spatial convolution.
2. Pointwise convolution actually is the 1×1 convolution to change the dimension.

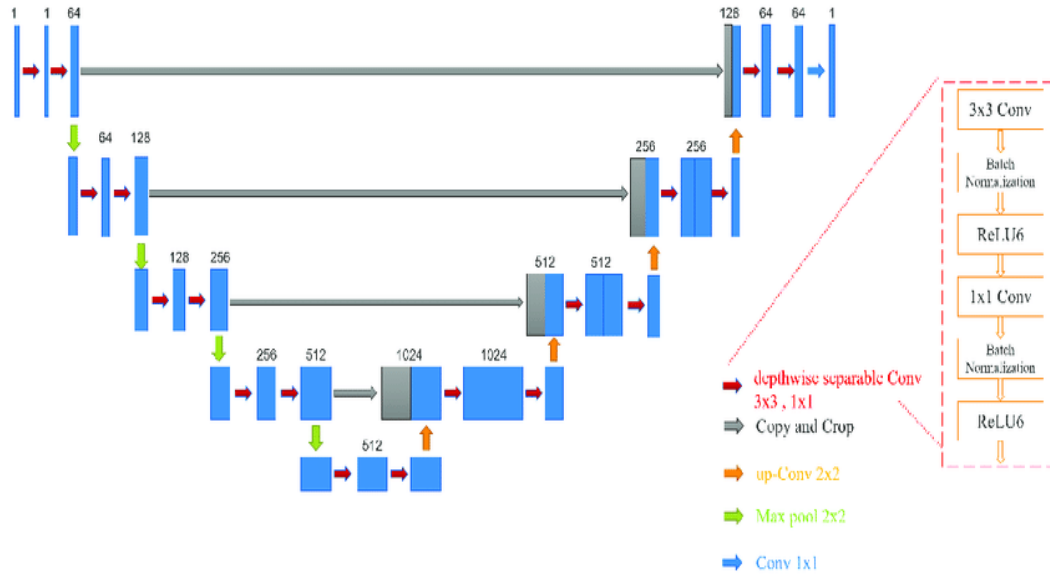


Figure 3 Separable U-Net Architecture

4. Results

Experiments measuring the computational requirements of Separable U-Net, its inference speed, and segmentation performance on the mentioned datasets conducted in this session.

4.1. Ablation Study

An extensive ablation study is performed to evaluate the performance of the proposed model and to support the final design decisions made in this study. Two different modifications are made to the architecture design and they include:

- U-Net (GN =32) – Original U-Net architecture with GN only with 32 groups.
- U-Net with Separable layers – U-Net architecture with depth wise separable layers replacing standard convolution layers and batch normalization layers only.

The performance of these modifications is reported alongside the original U-Net architecture, the proposed Separable Convolution with CReLU.

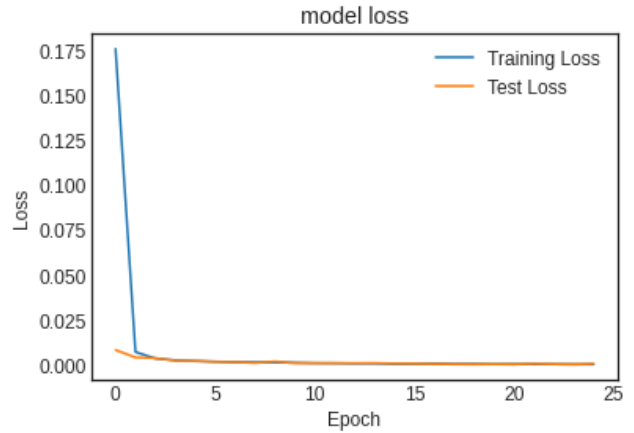
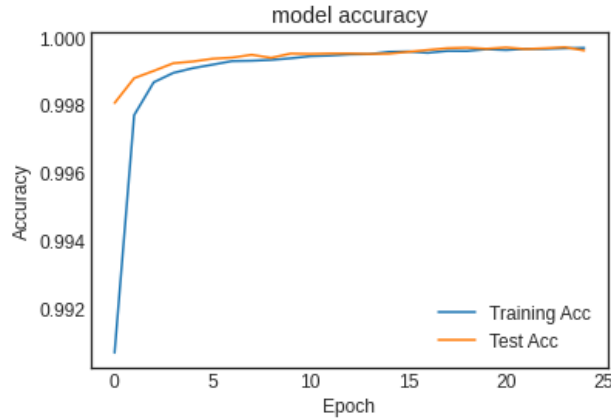
4.2. Results of Cassava Kaggle Dataset

Separable UNet is seen to achieve comparable performance in terms of accuracy, mean IOU and Dice co-efficient, while being more computationally efficient than the original U-Net model.

U-Net (GN = 32)

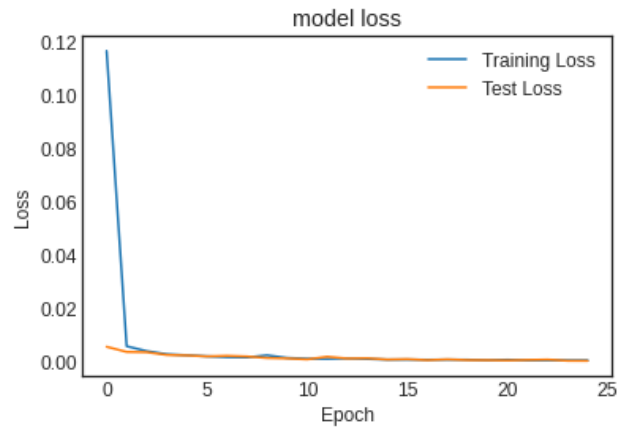
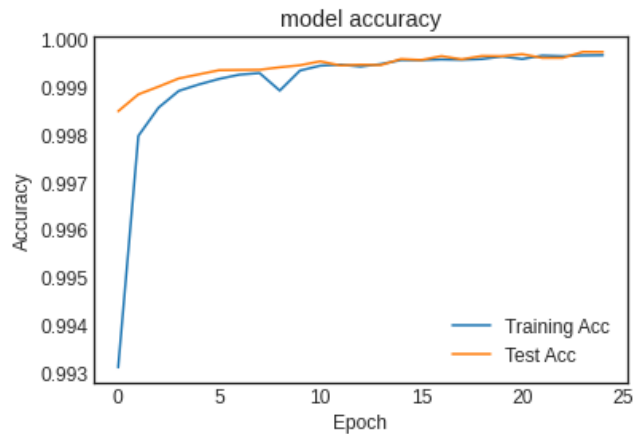
U-Net and SD-UNet are trained from scratch for only four epochs and their mean dice scores on validation data compared alongside their inference speed. The choice of a smaller number of epochs is due to the ability of the Adam optimization algorithm reaching a minimum quickly and each epoch runs 1000 iterations over the dataset. SD-UNet achieves a better loss and mean dice co-efficient compared to the U-Net model. Its inference speed is also faster on a single Tesla K40C gpu device. The training curve in Figure 10 also shows that WS with GN also significantly improves the training loss and obtains a smoother curve. Pixel wise, accuracy has been accepted as a general metric but is not necessarily the best form of performance evaluation mostly due to class imbalance. This means that accuracy could be very high or very low depending on the scale of pixel imbalance that exists in the dataset and, therefore, is not necessarily always correlated with the Dice co-efficient which measures the difference in the overlap between each pixel in an image and its prediction. The Dice co-efficient is not dependent on the balance of data and is more accurate compared to pixel accuracy. Sample tumor segmentation visualizations are shown in Figure 11 and it is interesting to note that while SD-UNet achieves comparable performance with U-Net on large tumor segmentations, it significantly outperforms U-Net on smaller tumor segmentations.

U-Net



Total params: 3,690,049
 Trainable params: 3,689,665
 Non-trainable params: 384

S U-Net



Total params: 3,561,281
 Trainable params: 3,560,897
 Non-trainable params: 384

5. Discussion and Conclusion

Early detection is necessary to help in preventing complications that may arise due to late detection of cassava leaf detections. However, with the increasing availability of large cassava leaf data, the workload on formers and other experts in the field has also increased. To help provide easier, accurate and timely detections, several deep learning methods have been proposed and most have chalked great successes in these tasks. The U-Net architecture is one such model that is widely accepted researchers for image segmentation tasks.

In recent times, mobile handheld devices have been enabled with processing functionalities that were only imaginable for large computers in the past. However, deep learning applications require even higher computations. This makes it very challenging to deploy deep learning applications on handheld or embedded devices. The U-Net architecture, for instance, requires over 2M FLOPs and over 370 megabytes (Mb) of storage space which are really high demands. Moreover, not much attention has been paid to applying deep learning methods on resource-constrained devices in areas of plant imaging.

In this study, separable convolution U-Net, has been presented for the segmentation of plant data on devices with limited computational budgets. The separable architecture makes use of depth wise separable convolutions. Our findings show that the proposed architecture is only 15.8 Mb in size which is 23 x smaller than the U-Net and requires 8x less computational complexity while maintaining decent accuracy results.

Furthermore, in the absence of experts of different unforeseen reason, being able to deploy separable convolution on a device such as a mobile phone could help anybody in obtaining segmentation results given the availability of images. Separable Unet robustness is also demonstrated during test results to perform significantly better than the original UNet architecture on smaller plant leaf datasets. Separable convolution will also be applied to different kinds of plant leaf disease data for further testing of its performance.

Acknowledgement

P Kiran Rao obtained the Degree of Master of Technology in CSE from JNTU Aanantapuramu. Presently he is pursuing a Doctoral degree in faculty of engineering from M S Ramaiah University of Applied Sciences, Bangalore. Moreover, he is working as an Assistant Professor in the CSE department for G Pullaiah College of Engineering & Technology, Kurnool. His research interests include cloud computing, deep learning, and medical image processing.

R Sandeep Kumar obtained the Degree of Master of Technology in CSE from JNTU Hyderabad. Presently he is pursuing a Doctoral degree in faculty of engineering from M S Ramaiah University of Applied Sciences, Bangalore. Moreover, he is working as an Assistant Professor in the CSE department for G Pullaiah College of Engineering & Technology, Kurnool. His research interests include cloud computing, deep learning, and medical image processing.

Dr K.Sreenivasulu received Ph. D degree in CSE from JNTU Kakinada ,india, in 2016. He obtained M.Tech in Computer Science from JNTU College of Engineering Anantapur in 2003 and BE in Computer Science and Engineering from Bangalore University in 1997.He is presently working as professor & Controller of Examinations ,G.Pullaiah College of Engineering and Technology Kurnool. He is having more than 22 years of experience in teaching. He has guided several Graduate & Post graduate students in their Academic projects. He is life member of IE(I), and MISTE. He has more than 12 research publications in proceedings of National, International Conferences, National and International Journals.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analysis, or interpretation of data; in writing of the manuscript, or in the decision to publish the results.

References:

1. Sifre, L. Rigid-Motion Scattering for Image Classification. Ph.D. Thesis, Ecole Polytechnique, Palaiseau, France, 2014. CMAP Rigid-Motion Scattering For Image Classification. Available online: https://www.di.ens.fr/data/publications/papers/phd_sifre.pdf (accessed on 17 February 2020).
2. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal (2018) Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 68 (6), pp. 394–424.
3. G. Scelo and T. L. Larose (2018) Epidemiology and risk factors for kidney cancer. *Journal of Clinical Oncology* 36 (36), pp. 3574–3581.
4. M. M. Nguyen, I. S. Gill, and L. M. Ellison (2006) The evolving presentation of renal carcinoma in the united states: trends from the surveillance, epidemiology, and end results program. *The Journal of urology* 176 (6), pp. 2397–2400.
5. M. Sun, F. Abdollah, M. Bianchi, Q. Trinh, C. Jeldres, R. Thuret, Z. Tian, S. F. Shariat, F. Montorsi, P. Perrotte, et al. (2012) Treatment management of small renal masses in the 21st century: a paradigm shift. *Annals of surgical oncology* 19 (7), pp. 2380–2387.
6. N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, et al. (2019) The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445.
7. P Kiran Rao, Dr Subarna Chatterjee et al.(2020) Diagnosis of Kidney Renal Cell Tumor through Clinical data mining and CT scan image processing: A Survey <https://doi.org/10.26452/ijrps.v11i1.1778>. pp. 13-24.