



Universidade Federal de Alagoas - UFAL
Instituto de Computação - IC
Curso de Ciência da Computação



Prof. Tiago Vieira

YCbCr Leaf Disease Segmentation

Eduardo Antônio de Lucena Lisboa

Maceió, 14 de fevereiro de 2022

Contents

1	Introduction	3
2	Objective	3
2.1	Related Works	3
2.2	Original Paper	3
2.3	Proposed Improvement	3
3	Materials and Methods	4
3.1	The BRACOL Dataset	4
3.2	Machine Learning Model	4
3.3	Code	5
4	Results	5
5	Conclusion	6
	References	7

1 Introduction

A common problem in Brazilian crops, as well in crops all around the world, is diseases and pests. All kinds of crops can suffer from such issues and preventing them is a highly sought subject.

Usually it can be hard to diagnose these diseases and pests in advance due to similarities between diseases or the size of the insect. Thus, identifying these diseases and treating them as fast as possible is fundamental to avoid partial or total crop losses.

One such plant is the Brazilian Arabica Coffee, which is known worldwide and one of Brazil's most important agricultural commodity, can be affected by diseases and pests and this can have enormous social and economical implications, due to the many families directly and indirectly involved in coffee plantations.

Being able to determine which part of the plant has a disease or a pest and how much of it has been affected can help to identify where to focus the treatment, if it should be a local treatment of a single or a few trees or if it should be of the whole plantation.

2 Objective

2.1 Related Works

In the literature, different color spaces have been used for various tasks [1][2] one of these color spaces is the YCbCr color space.

Initially used only in television because it uses less bandwidth to be transmitted and can be perceived just as well as a standard RGB image but with much less color information, has been shown to have excellent potential of increasing the accuracy of image segmentation, being it with faces [3], satellite images [4] and plants, being the latter used for maturity evaluation [5] or disease detection [6][7][8].

For this segmentation achieve its full potential, it must be paired with other methods and algorithms, and all these methods rely on a basic strategy that is the use of textural features [9].

Another important strategy commonly used is using the *k-means* clustering algorithm, which has been shown several times to be an efficient algorithm in segmentation problems [10] including plant diseases [11].

2.2 Original Paper

The original paper proposed a method of segmentation of leaf images using the *YCbCr* color space combined with the *K-Means* clustering algorithm to identify the diseased part and highlight it for easier identification.

The problem with this method is that if an image of a healthy leaf goes through this processing, it yields a wrong segmentation which can be confusing and an unnecessary expense of processing power, specially if a large amount of images are processed, as it may be the case in large coffee crops. Example images can be seen on Fig 1.

2.3 Proposed Improvement

To enhance the results obtained by the original paper, I intend to not only trying to better segment the image but also calculate the affected area size and the percentage of the leaf compromised, as well as creating two plots, one with the segmentation process, the segmented area

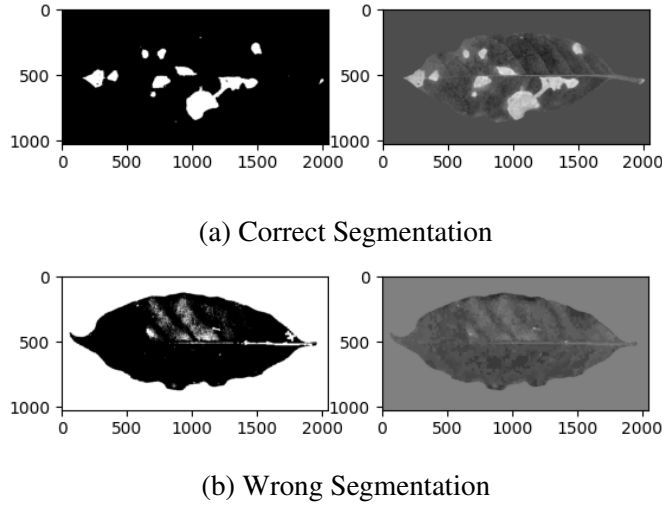


Figure 1: Segmentation Examples of Original Paper

and the original input image for a clearer understanding and the second showing the contours of the whole leaf and all located diseased areas, save the extracted information in a file for future analysis and implement a Jupyter notebook that can be used by anyone to classify their own coffee leaves images.

3 Materials and Methods

3.1 The BRACOL Dataset

The segmentation was done on the BRACOL Dataset [12], which is in the public domain, consisting of 1747 images of size 2048 x 1024 pixels. But due to the corruption of the compressed archive, only 1216 images could be used.

The images in BRACOL are labeled with healthy or diseased and the severity of each disease present, being these diseases *Rust*, *Miner*, *Cercospora* and *Phoma*, and the severity is divided in 5 different levels, being 0 a healthy leaf and 4 a very compromised leaf. All images were captured in similar conditions of lighting and a single leaf per image. The leaves were collected at different times of the year.

3.2 Machine Learning Model

The model for the images classification used was the same as used in the Deep Learning discipline with the model consisting of a *data augmentation* layer followed by a *normalization* layer, then two *2D convolutional* layers and a *2D max pooling* layer. After this come three more *2D convolutional* layers interspersed with three *2D max pooling* layers. Then we have a *dropout* layer, a *flattening* layer and three *dense* layers, the first two with *ReLU* activation and the last with *Softmax* activation.

The models were trained using 80% of the total amount of images in the dataset, being this percentage further divided in 80% for training and 20% for validation.

The hyperparameters *batch size* and *epochs* were fixed, having the values 32 and 50 respectively, and the *learning rate* varying, starting from 0,001. The training's accuracy and loss were

saved and plotted to assess the best learning rate for the final model, with the best value being found to be 0,0016.

The model fitting also used two *callback* functions, *EarlyStopping* and *ModelCheckpoint*. The first stops the model training if an arbitrary number of epochs pass without an arbitrary metric improving, in our case, the number of epochs (or *patience*) was 20 and the metric monitored was *accuracy*, and the second saves a version of the model every time an arbitrary metric improves, in our case, *accuracy*.

3.3 Code

The original paper's code was rewritten to clarify some steps and merge with the machine learning model and new processing. It was written in Python programming language version 3.9.10, and also tested in version 3.8.7 working without any problems.

The models were made using the TensorFlow library [13], version 2.8, and Keras [14] library, included in the TensorFlow library, and the plots were made with the Matplotlib library [15], version 3.5.1.

4 Results

The new model trained with the found learning rate value turned out to have worse results than the previously made model so the new ones were discarded and the original model was used instead.

The processing of the images yielded great results, achieving what was proposed. Although the segmentation itself was not improved, the classification of the images with the machine learning model to minimize the number of healthy images being processed, saving computational power for the diseased ones, allied with the new data extracted have made possible an improved visualization as well as a more in depth look on the leaves diseases severeness.

The results can be seen in Figs 2, showing the original images processed for the following examples, 3 and 4, where 3a, 3b and 3c as well as 4a, 4b and 4c show the new form of processing visualization of diseased images, Fig 5 showing how the data analyzed is presented and Fig 6 comparing the original paper's results with the new, where 6a is the only output of the original paper and 6b, 6c and 6d are the outputs of this paper.

The online version proposed was successfully implemented and can be found [here](#).

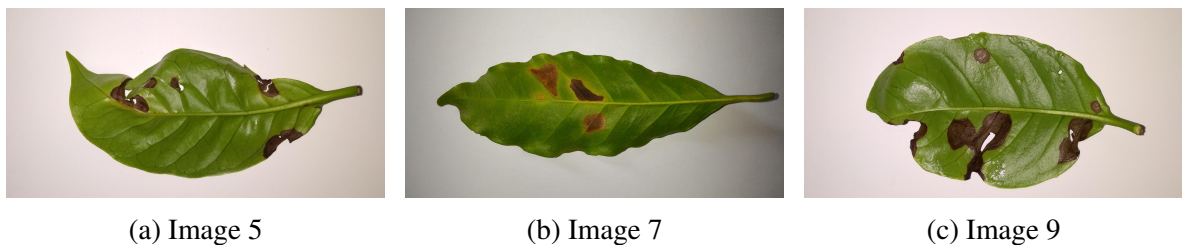
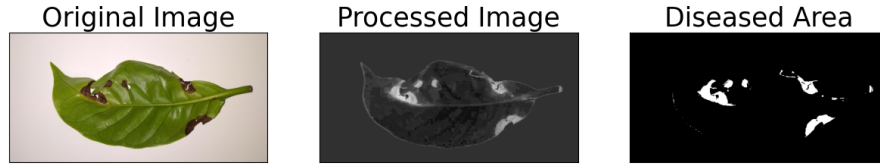
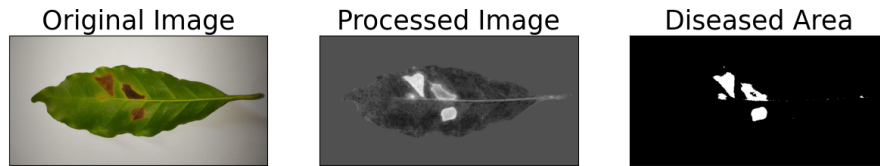


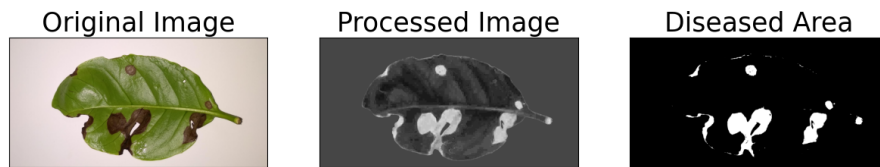
Figure 2: Original Images



(a) Image 5



(b) Image 7



(c) Image 9

Figure 3: Comparison Plots

5 Conclusion

The new data extracted from the images, along with the new form of visualization, can definitely be implemented in any crop around the world, providing an easy way to identify diseased areas with great accuracy and also showing important data about the diseases severeness, which can be used to put preventive measures in place while these diseases are not spread across the whole crop.

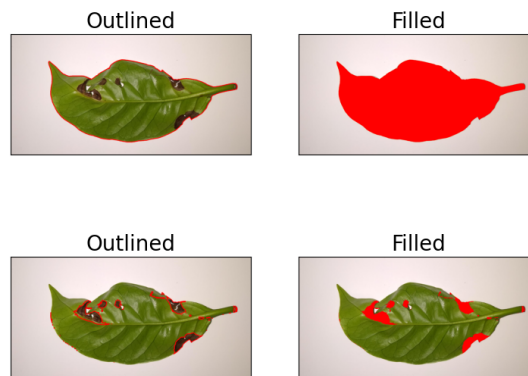
To further improve this paper, a possibility is to expand the machine learning model to distinguish between diseases or even expand the whole project to include not only Arabica Coffee leaves, but other plants as well. It is also possible to implement this method, expanded or not, in a mobile app for it to be readily accessible by anyone with a smartphone.

References

- [1] Y. Feng, G. Ren, K. He, Y. Liu, and L. Li, "Rgb color channel variation based segmentation of crop leaf lesion," in *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, 2015, pp. 592–596.
- [2] T. S. Bodhe and P. Mukherji, "Selection of color space for image segmentation in pest detection," in *2013 International Conference on Advances in Technology and Engineering (ICATE)*, 2013, pp. 1–7.
- [3] K. H. B. Ghazali, J. Ma, R. Xiao, and S. A. Iubis, "An innovative face detection based on YCgCr color space," *Physics Procedia*, vol. 25, pp. 2116–2124, 2012. [Online]. Available: <https://doi.org/10.1016/j.phpro.2012.03.358>
- [4] P. Ganesan, V. Rajini, B. S. Sathish, V. Kalist, and S. K. K. Basha, "Satellite image segmentation based on YCbCr color space," *Indian Journal of Science and Technology*, vol. 8, no. 1, p. 35, Jan. 2015. [Online]. Available: <https://doi.org/10.17485/ijst/2015/v8i1/51281>
- [5] R. Dhanesha, C. L. S. Naika, and Y. Kantharaj, "Segmentation of arecanut bunches using YCgCr color model," in *2019 1st International Conference on Advances in Information Technology (ICAIT)*. IEEE, Jul. 2019. [Online]. Available: <https://doi.org/10.1109/icaity47043.2019.8987431>
- [6] V. Singh and A. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41–49, Mar. 2017. [Online]. Available: <https://doi.org/10.1016/j.inpa.2016.10.005>
- [7] E. Prasetyo, R. D. Adityo, N. Suciati, and C. Fatichah, "Mango leaf image segmentation on hsv and ycbcr color spaces using otsu thresholding," in *2017 3rd International Conference on Science and Technology - Computer (ICST)*, 2017, pp. 99–103.
- [8] L. Wang, T. Yang, and Y. Tian, "Crop disease leaf image segmentation method based on color features," in *CCTA*, vol. 258, 08 2007.
- [9] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973. [Online]. Available: <https://doi.org/10.1109/tsmc.1973.4309314>
- [10] J. Baghel and P. K. Jain, "K-means segmentation method for automatic leaf disease detection," 2016.
- [11] T. G. Devi, P. Neelamegam, and A. Srinivasan, "Plant leaf disease detection using k means segmentation," 2018.
- [12] R. Krohling, J. Esgario, and J. A. Ventura, "Bracol - a brazilian arabica coffee leaf images dataset to identification and quantification of coffee diseases and pests," <https://www.doi.org/10.17632/yy2k5y8mxg.1>, 2019.
- [13] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga,

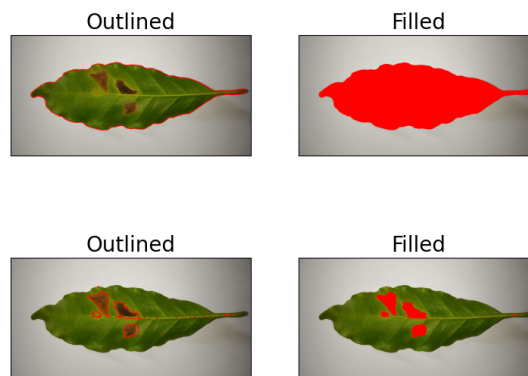
- S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>
- [14] F. Chollet *et al.* (2015) Keras. [Online]. Available: <https://github.com/fchollet/keras>
- [15] J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

Contour Areas



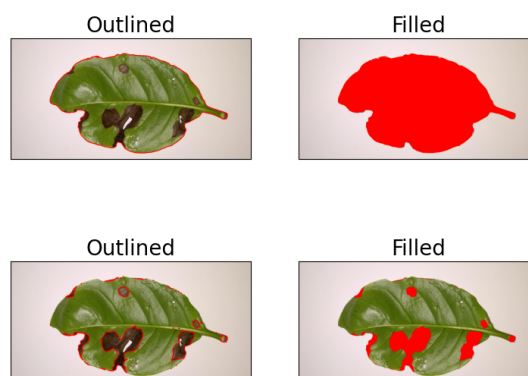
(a) Image 5

Contour Areas



(b) Image 7

Contour Areas



(c) Image 9

Figure 4: Contours Plots

```

IMAGE 5
Image shape:      1024 x 2048 px
Approximate leaf shape: 781 x 1603 px
Total leaf area:  679400 px^2
Diseased area:    55076 px^2
Diseased percentage: 8.11%

```

(a) Image 5

```

IMAGE 7
Image shape:      1024 x 2048 px
Approximate leaf shape: 576 x 1827 px
Total leaf area:  602802 px^2
Diseased area:    47979 px^2
Diseased percentage: 7.96%

```

(b) Image 7

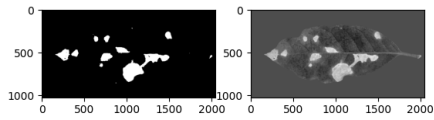
```

IMAGE 9
Image shape:      1024 x 2048 px
Approximate leaf shape: 846 x 1509 px
Total leaf area:  776972 px^2
Diseased area:    111477 px^2
Diseased percentage: 14.35%

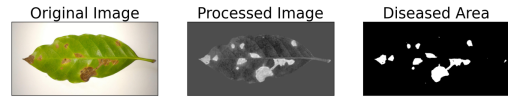
```

(c) Image 9

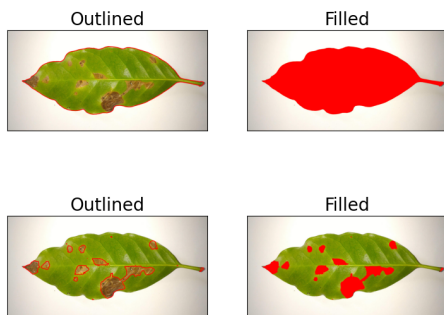
Figure 5: Informations



(a) Original Paper's Results
Contour Areas



(b) New Results - Comparison



(c) New Results - Contours

```

IMAGE 3
Image shape:      1024 x 2048 px
Approximate leaf shape: 693 x 1853 px
Total leaf area:  745865 px^2
Diseased area:    109107 px^2
Diseased percentage: 14.63%

```

(d) New Results - Info

Figure 6: Comparison - Original and New