Classification of coffee beans using Deep Learning

Igor G. Lube[1], Gustavo M. Almeida[1]

[1]*Instituto Federal do Espírito Santo – Campus Serra*
Rodovia ES-010 - Km 6,5 – Manguinhos, 29173-087, Espírito Santo, Brazil
igorgl_@hotmail.com
gmaia@ifes.edu.br

Abstract. The global agribusiness of coffee includes, annually, resources that reach 91 billion dollars and involves half a billion people. The coffee market is characterized by a set of activities of enormous complexity, dynamism, and a growing level of demand from consumers regarding the quality of the drink. This imposes high quality control on producing, consuming, and exporting countries. Currently, the definition of the quality and therefore the value of coffee is based on manual classification, that is, a person plays the role of a trained (certified) classifier to qualify coffee samples. Thus, the current classification process suffers from the subjectivity of the classifiers and a great difficulty in standardizing the process due to possible inconsistencies in the process. Given this scenario, the present work proposes to develop a system for classifying coffee samples considering shape and imperfections. The classification process will be done by using computer vision through Deep Learning and regional convolutional networks (R-CNN) where the intrinsic defects present in the sample will be identified. Among the benefits of automating the coffee classification process, the following stand out: Cost reduction, agility and standardization of the classification.

Keywords: Classification of coffee beans. Deep Learning. Mask R-CNN

1 Intrododuction

The importance of coffee in the world economy is well known, being the second most consumed beverage in the world according to the International Coffee Organization [1] which also reveals Brazil as the first place in coffee production and export worldwide, which makes this product one of the most important sources of income in the country and in several municipalities, generating more than 8 million jobs nationwide [2]. One of the main difficulties for producers is the evaluation of the quality of their crops, which must be quick and reliable to keep up with the volatility of the prices of these commodities. The quality of the drink depends, among other factors, on the quality of the grains used and the amount of imperfections and impurities mixed with the good grains. Defects found in coffee samples are accounted to define the type of coffee. These defects are counted manually by trained evaluators. The subjectivity of the process is great because each evaluator may have a different classification from the same sample and the quality of the evaluation also suffers negative effects during the work shift for reasons inherent to human physiology and environmental conditions such as fatigue, stress, poor lighting etc. The speed and reliability are extremely dependent on the professional, which makes it difficult to standardize this stage of the production chain since each human being has its particularities.

In this behalf, the present work proposes a method for classifying coffee beans using computer vision to increase speed and standardization of the process.

2 Related works

Pizzaia et al. [3] propose a multilayer Perceptron neural network (MLP) to classify coffee beans through the shape, size and color of the samples. The images used contained good and defective coffee beans previously classified by specialists. The grains were segmented by Otsu binarization and the areas, rounding and the average value of the RGB layers of each grain were calculated. These characteristics were fed into the MLP network that classifies the grains as good or defective. The network obtained an average of 94% accuracy.

In the work of Vasconcellos et al. [4] a system using K-Means clustering is proposed to classify four types of coffee beans, namely the good, broca damage, in parchment and black beans. The samples were treated with the White-Patch luminosity standardization algorithm. The area, rounding and average of the RGB channels of each grain were used in the classifier, obtaining a 90.74% hit rate.

This article proposes a method for classifying coffee beans using deep learning and Regional Convolutional Neural Networks (R-CNN). Section 3 presents the theoretical basis used to execute the method. Section 4 presents the system and procedures used. In section 5, the discussions and results are shown.

3 Theoretical foundation

This section describes the Mask R-CNN framework used to implement this method.

Defined by He et al. [5] as a simple and flexible framework for instance segmentation, Mask R-CNN is based on a framework proposed by Ren et al. [6] called Faster R-CNN, a system for object detection and semantic segmentation, which is fast, robust and intuitive.

Semantic segmentation is the task of classifying the pixels of an object belonging to a class and object detection aims to classify the objects and localize them using a bounding box. Instance segmentation is the combination of these two tasks from the classical computer vision.

Mask R-CNN adds a new mask prediction branch that works in parallel with the existing Faster R-CNN bounding box classification and regression branch. This new branch is a small FCN (Full Connection Network) applied pixel by pixel in each region of interest (RoI) and in addition to this FCN there is a layer called RoIAlign with the function of improving the precision of the created masks and decoupling the branch from the masks class prediction, guaranteeing an improvement of 10% to 50% in the accuracy of the masks.

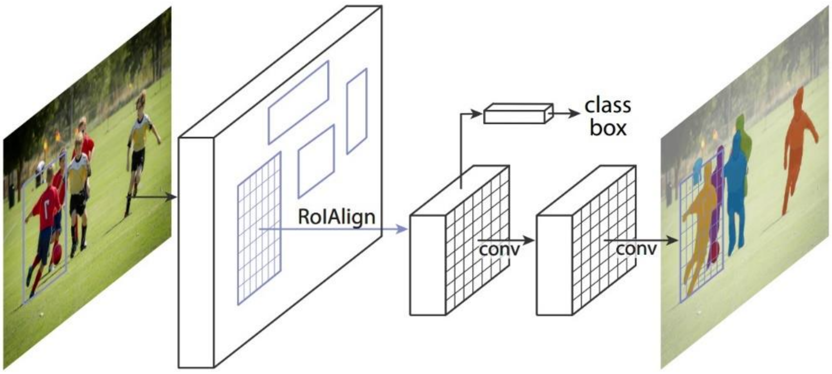

Figure 1. Framework of Mask R-CNN

4 Method

The database of coffee beans was assembled from samples of beans obtained by the partnership between the Serra and Alegre campuses of IFES (Federal Institute of Science and Technology of Espírito Santo). The samples contain good grains and defective grains (broca, parchment, black and bad) separated and classified in a traditional way by specialists.

4.1 Image dataset

The treatment of the images was by segmentation in which an image with several grains was subdivided into smaller images with only one grain. Because of the low number of samples, it was necessary to use the Data

Augmentation process that applies rotations, mirroring, among other strategies to increase the database and, consequently, the accuracy of the network.

Figur2 2. Grains before (left) and after (right) segmentation

The grain classes have characteristics that differentiate them by attributes such as color, appearance and uniformity as shown in Table 1.

Table 1. Characteristics of coffee beans

Class	Characteristics
Good	Color and appearance uniformity and free from significant imperfections
Bad	Broken, non-uniform, small grains and various imperfections
Broca	dark canals due to the presence of the coffee berry borer in the planting
Black	Blackish color due to high level of undesirable fermentation
Parchment	Yellowish parchment surrounding the grain that has not been completely removed

The grains with defects of the type parchment, black and broca have their characteristics well defined, however in the class of bad grains several defects were grouped as: broken grains, misshapen, green etc. Due to the lack of sufficient samples of these defects, it was necessary to leave them in the same class.

The grain image bank has the following classes and quantities according to the Table. 2:

Table 2. Grain dataset quantities by class

Class	Training	Validation
Good	200	8
Bad	200	5
Broca	96	5
Black	66	11
Parchment	44	4
Total	606	33

4.2 Image annotations

For the supervised training of the network to take place, it is necessary that the images contain the annotations of the location of the grains and their respective classes, which are the Regions of Interest (RoI) and the Labels. For this, the VGG Image Annotator software (VIA tool) was used.

4.3 Training

The algorithm used to implement the Mask R-CNN was done in Python with a Resnet50 backbone. The algorithm is an adaptation of the one proposed by the original article by He et al [5]. and the main changes were in the re-scaling of images, bounding box and in the learning rate.

The algorithm of this work received the images with annotations for training, however the database is too small to train a network from scratch. To solve this problem, the technique known as fine tuning was used, which is a type of transfer learning where characteristics learned by network trained in a larger database is trasfered to a smaller one.

The proposed network was trained by transferring learning from a pre-trained network based on the Microsoft COCO database: Common Objects on Context [7]. With fine tuning it is possible to have a network with good results even with a reasonably small database. Table 3 shows the hyperparameters used for training the network and their respective values.

Tabela 3. Parameters used in network training

Parameters	value	Description
Number of epochs	45	Number of training epochs
Steps per epoch	100	Number of samples per epoch
Images per GPU	2	Number of images processed per GPU
Learning rate	0.001	Network learning rate
Minimum confidence	0.6	Ignores detections with less than 60% confidence

5 Results and discussions

The network was trained with 606 images of grains segmented for 45 epochs using the COCO Dataset for transferring learning. As a result, a network was obtained with good performance in classifying and building masks for the proposed classes.

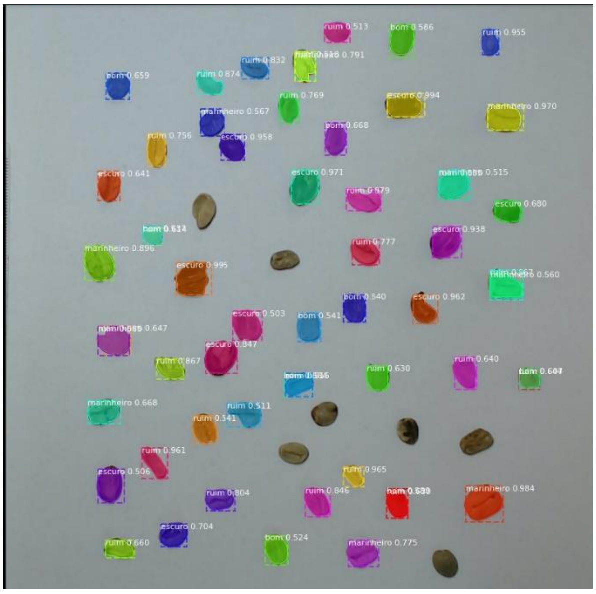

Figure 3. Masks obtained for all grain classes

The table 4 with the confusion matrix of the classification results is shown below.

Tabela 4. Confusion matrix

		Predicted class				
		Good	Bad	Broca	Black	Parchment
Real class	Good	7	0	0	1	0
	Bad	1	2	0	2	0
	Broca	0	0	5	0	0
	Black	0	0	0	11	0
	Parchment	0	1	0	0	3

Table 5 shows the performance metrics of the classifier.

Table 5. Classifier performance

Class	Accuracy	Precision	Recall	F-score
Good	93,94%	87,5%	87,5%	87,5%
Bad	87,88%	66,67%	40%	50%
Broca	100%	100%	100%	100%
Black	90,91%	78,57%	100%	88%
Parchment	96,97%	100%	75%	85,7%
Average	93,94%	86,55%	80,50%	82,24%

The results obtained show that the network obtained a good efficiency to correctly classify all classes of grains except for bad grains. The probable explanation of the poor performance for this class is the great variety of characteristics that define it, being possible a bad grain to be confused with other types as they have some attribute in common.

Compared to previous studies [4], this network obtained a satisfactory result related to brocated grains, which are difficult to differentiate from good grains, as they are different only by the small dark holes resulting from the attack of the coffee borer (Hypothenemus hampei). This greater efficiency is mainly due to the network's ability to identify the recurrent pattern of holes in the inner part of the grains, which is not feasible if the network only has data on the edge pixels and the average value of the RGB channels.

However, the proposed network is not able to correctly classify some grains due to the reliability being below the adjusted minimum, as it is possible to observe in some grains without masks in figure 3. This performance can be improved if the network is trained in a larger database and more complete, mainly for bad grains, which could become several new classes with well-defined defects.

6 Conclusion

One of the great advantages of using convolutional neural networks, and more specifically Mask R-CNN, is to skip steps for preparing the input data. Because of the great robustness of this type of network, it is not necessary to adjust the images, whether in scale, rotation, lighting, etc., as the network can learn the attributes of the data with spatial invariability, making CNN in general very efficient for the computational vision task.

The work proposed in this article shows that it is perfectly feasible to classify coffee beans using only the learning of network characteristics, without the need for extra parameters. However, the lack of a large and diversified database reduces the consistency of the results obtained, with the acquisition of this image bank being the next step towards better network performance.

The reason for the low number of samples was the difficulty in obtaining images of grains and types of imperfections that was properly classified by specialists. Several cooperatives were contacted, but there was no positive feedback in a timely manner.

The network's response speed provides a great time advantage over manual classification, showing that this method, once improved, can greatly increase the productivity and standardization of the coffee bean classification process.

References

[1] International Cofee Organization, year = 2019, url = http://www.ico.org/prices/new-consumption-table.pdf, urldate = 2020-08-10.
[2] Ministry of Agriculture, Livestock and Supply, Brazil, year = 2019, url = https://www.gov.br/agricultura/pt-br/assuntos/politica-agricola/cafe/cafeicultura-brasileira, urldate = 2020-08-05.
[3] J. P. L. Pizzaia, I. R. Salcides, G. M. Almeida, R. Contarato, R. Almeida. *Arabica coffee samples classification using a Multilayer Perceptron neural network*. In 13th IEEE International Conference on Industry Applications (INDUSCON), pp. 80-84, 2018.
[4] B. M. Vasconcellos, G. M. Almeida, R. B. Contarato, R. Almeida. *Classificação de amostras de café Usando visão Computacional.* In Simpósio Brasileiro de Automação Inteligente, pp. 2505-2508, 2019.
[5] K. He, G. Gkioxari, P. Dollar and R. Girshick. *Mask r-cnn*. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
[6] S. Ren, K. He, R. Girshick and J. Sun. *Faster r-cnn: Towards real-time object detection with region proposal networks*. In *Advances in neural information processing systems*, pp. 91–99, 2015.
[7] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollar and C. L. Zitnick. *Microsoft coco: Common objects in context. In European conference on computer vision*, pp. 740–755, 2014.