PAPER • OPEN ACCESS

Colombian fruit and vegetables recognition using convolutional neural networks and transfer learning

To cite this article: O I Álvarez-Canchila *et al* 2020 *J. Phys.: Conf. Ser.* **1547** 012020

View the article online for updates and enhancements.

Colombian fruit and vegetables recognition using convolutional neural networks and transfer learning

O I Álvarez-Canchila[1], D E Arroyo-Pérez[1], A Patiño-Saucedo[2], H Rostro González[2], and A Patiño-Vanegas[1]

[1] Facultad de Ciencias Básicas, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia

[2] Departamento de Ingeniería Electrónica, Universidad de Guanajuato, Salamanca, México

E-mail: `imoscaralvarez@gmail.com, alberto.patino@ugto.mx`

Abstract. Automatic image recognition is a convenient option for labeling and categorizing fruits and vegetables in supermarkets. This paper proposes the design and implementation of an automatic classification system for Colombian fruits, by training a convolutional neural network. A database was created to train and test the system, which consisted of 4980 images, labeled in 22 classes, each corresponding to pictures of the same kind of fruit, trying to reproduce the variability of a real case scenario with occlusions, different positions, rotations, lightings, colors, etc., and the use of bags. On-training data augmentation was used to further increase the robustness of the model. Additionally, transfer learning was implemented by taking the parameters of a pretrained model used for fruit classification as the new initial parameters of the proposed convolutional network, achieving an increase of the classification accuracy compared with the same model when trained with random initial weights. The final classification accuracy of the network was 98.12% which matches the scores achieved on previous works that performed fruit classification on less challenging datasets. Considering top-3 classification we report an accuracy of 99.95%.

1. Introduction

Fast and reliable automatic classification of fruit and vegetable images remains an open challenge that if solved can greatly benefit the experience of customers in supermarkets and fruit stores. The difficulty of the problem arises from the high variability of the data: there are thousands of classes and subclasses of produce, with intra-class variations in size, quantity, state of maturation, the use of bags, etc. Not to mention the usual complexity of data recorded by cameras, where occlusions, pose, illumination and background changes are commonplace.

In the past, image classification tasks were tackled by obtaining hand-engineered features from explicit knowledge about the data, feeding these features to standard classifiers, with works such as Dubey, *et al.* [1], Rocha, *et al.* [2], Zawbaa, *et al.* [3] and Zhang, *et al.* [4] for the case of fruit and vegetable classification. Nowadays, the state of the art method to perform image classification is the convolutional neural network (CNN) [5], which allows the system to find the relevant features by itself. The success of CNNs is due to their ability to supervisedly learn invariate representations of images helped by spatial filters [6]. The parameters of the filter are updated by backpropagation, a generalization of the gradient descent algorithm for multiple sequentially connected layers. In the case of fruit classification, CNNs have been recently exploited, with works such as Patiño-Saucedo, *et al.* [7]. It is worth to mention that works on fruit classification literature are limited to locally gathered datasets (*e.g.* the database gathered in [2] is confined to fruits and vegetables commonly found in Brazil) and as far a we know a general fruit classification system has not yet been attempted.

In the present study, a system is developed for the recognition of fruits and vegetables produced in Colombia, on a data-set gathered for this purpose. The motivation is to provide supermarkets with an option to automatically classify the produce, since it is commonly carried out by cashier personnel, but there are cases in which fruits do not have an automatic detection mechanism such as bars codes, leading to the memorization of codes and delays in the attention to the customer. In order to achieve the automatic classification of fruits and vegetables we propose the training of a CNN, in which the system must show as a result the produce category delivered to the network regardless of its kind, color, quantity, texture, etc.

The data set for the development of this study consists of 4980 fruit photographs divided into 22 categories that were taken with a Samsung WB50F camera with a resolution of 4608 x 3456 pixels, 3 color channels and a depth of 24 bits. The system was designed by implementing a CNN of the AlexNet type (Krizhevsky, *et al.* [8]), the use of data augmentation to increase the robustness of the model [9], and transfer learning (TL) to make it learn from diverse kind of data [10]. A comparative analysis of the details of training with and without transfer learning is carried out.

2. Contribution

This work constitutes the first attempt to build an automatic fruit and vegetable classification system from images with products commonly found in Colombian markets. The data-set created for this work features the use of transparent bags for some of the categories. Similar looking fruits were collected such as celery, Figure 1(a), chives, Figure 1(b), yam, Figure 1(c), yucca, Figure 1(d), mandarin, Figure 1(e), naranjilla, Figure 1(f), pear, Figure 1(g), green apple, Figure 1(h). This and the use of bags such as in Figure 1(e) mandarin, Figure 1(f) naranjilla add complexity to the classification task. In Figure 2 we observe the performance of the system for four samples. Note that for some challenging fruit samples such as mandarin Figure 2(b), easily confused with naranjilla Figure 2(c), the classification confidence drops. But for most challenging fruits such as shown in Figure 2(a) and Figure 2(c), the system performs correctly.

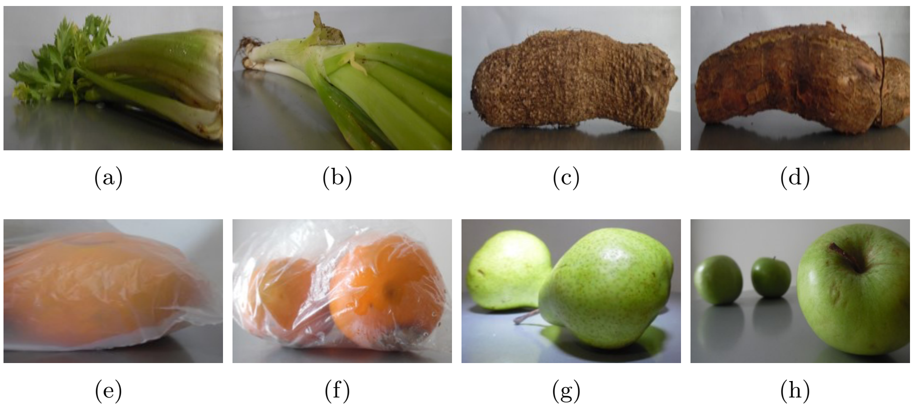

Figure 1. Samples of similar fruits and vegetables from the data-set: (a) celery, (b) chives, (c) yam, (d) yucca, (e) mandarin, (f) naranjilla, (g) pear, (h) green apple.

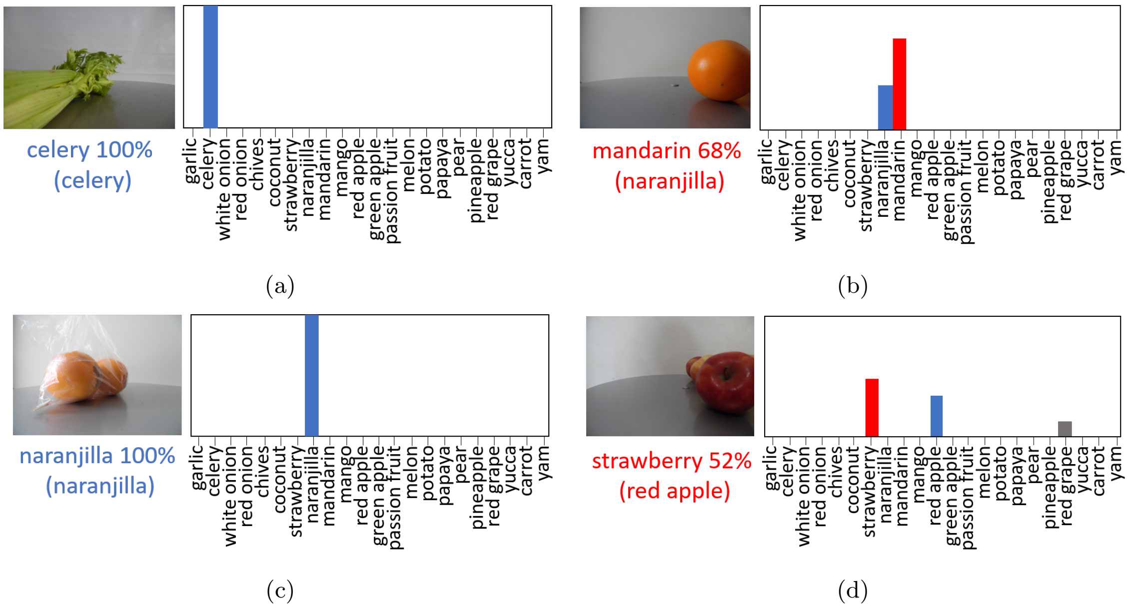

Figure 2. Image recognition of similar fruits and vegetables: (a) celery, (b) naranjilla, (c) naranjilla, (d) red apple.

3. Methodology

3.1. Dataset

For the elaboration of the data-set the photographs were taken with a digital camera, a tripod and a rotating platform. The data-set comprises 22 categories labeled as follows: garlic, celery, white onion, red onion, chives, coconut, strawberry, naranjilla, mandarin, mango, red apple, green apple, passion fruit, melon, yam, potato, papaya, pear, pineapple, red grape, yucca and carrot; totaling 4980 photographs with a resolution of 4608 x 3456 pixels. For training the network, a sub-sampling of a factor of 72 was performed on the entire data set producing images of 48 x 64 pixels.

For the gathering of the database three lighting conditions were taken into account: natural scarcely illuminated, natural fully illuminated, and white artificial light. The photographs represent high variability in position and number of fruits devising a real scenario.

3.2. Convolutional neural network training

An adaptation of the AlexNet network was implemented to train the dataset consisting of 6 convolutional layers, 3 densely connected layers, MaxPooling operators after the first, second and fifth convolutional layers and the Dropout regularization technique after the fourth convolutional layers, first and second densely connected layers. A scheme of the CNN can be seen in Figure 3.

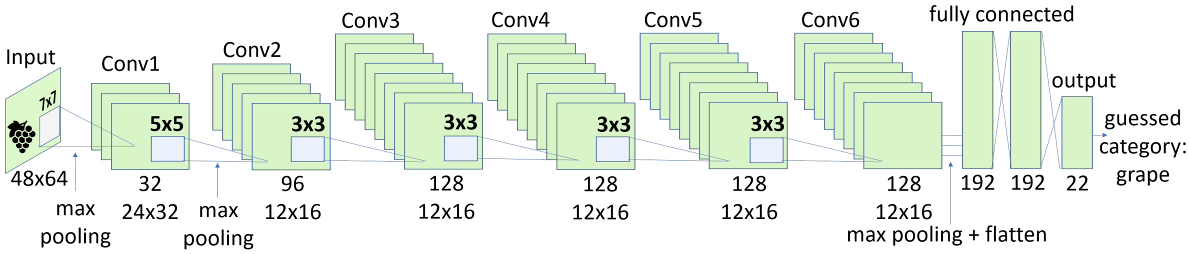

Figure 3. Proposed AlexNet inspired for fruit classification.

Journal of Physics: Conference Series **1547** (2020) 012020 doi:10.1088/1742-6596/1547/1/012020

First, a baseline model CNN BASE (no bags) is trained by considering only the images without bags, 2947 images, which constitute 59% of the full data-set. The samples without bags were split in 80% for training and 20% for testing. Data augmentation of the training samples was used, as shown in Figure 4(a) where the original picture is transformed with horizontal flip Figure 4(b), rotation Figure 4(c), vertical flip Figure 4(d) and scaling Figure 4(e). This is usual practice in order to increase the robustness of the model to data variability, so it can generalize better to unseen examples.

The second model CNN (full datatet) follows a similar procedure (same training/test split and data augmentation) but the whole data-set of 4980 images was used for training and testing.

The third model CNN+TL (full dataset) was also trained with the whole dataset but saught to increase the performance of the second model by transfering the weights or parameters of the first model. This is known as transfer learning and is done by making a loan of the weights of the first layers of a model trained with another data, maintaining random weights in the final layers of classification, with the objective of efficiently extracting the high-level characteristics of the images, which accelerates the learning process of the target network.

In all of the three models, training was performed in 300 epochs. The experiments were performed in a Colaboratory Jupyter notebook environment belonging to Google [11], executed in the cloud with Python and the Keras API, with TensorFlow [12] backend which uses a Tesla K80 GPU. In total, 3 experiments were performed: CNN BASE (no bags), CNN (full dataset) and CNN+TL (full dataset).

(a) (b) (c) (d) (e)

Figure 4. Data augmentation of a training example: (a) original, (b) horizontal flip, (c) rotation, (d) vertical flip, (e) zoom.

Table 1 shows the number of samples per class for training, totaling 2947 for the "no bags" model and 4980 for the model trained on the whole data-set.

Table 1. Size of the training set per class of bagless and full (complete) data-set.

Class	Without bags	Complete	Class	Without bags	Complete
Garlic	147	239	Green apple	96	202
Celery	143	254	Passion fruit	139	291
White onion	158	258	Melon	94	164
Red onion	151	241	Yam	165	259
Chives	123	213	Potato	174	260
Coconut	142	225	Papaya	153	236
Strawberry	115	171	Pear	156	283
Naranjilla	118	190	Pineapple	127	225
Mandarin	129	255	Red grape	100	174
Mango	150	227	Yucca	158	248
Red apple	67	111	Carrot	142	249

4. Results

Numerical results are given in Table 2 and shown in Figure 5. Figure 5 displays the results on each of the three proposed training methods. In the execution of the training it is evident that when transfer learning is applied, the accuracy of the classification of the network increases, from 97.70% to 98.12%. It is also noted that when the network is trained without bags, the classification is higher at 98.39% which is expected as the data is less complex.

In the three experiments, we selected the training images using sampling without replacement of the group of each image class. We repeat this procedure 10 times and we report the average classification accuracy (μ), average error ($\epsilon = 1 - \mu$). Additionally, we report the top-3 classification accuracy on Table 3. This metric takes a correct classification when the expected classification is among the three first most active neurons in the output layer. For the three models it achieves almost 100% accuracy.

Table 2. Classification accuracy on the test set for the three models.

CNN BASE (no bags)	CNN (full dataset)	CNN+TL (full dataset)
98.39%	97.70%	98.12%

Table 3. Top-3 classification accuracy for the three models.

CNN BASE (no bags)	CNN (full dataset)	CNN+TL (full dataset)
99.96%	99.82%	99.95%

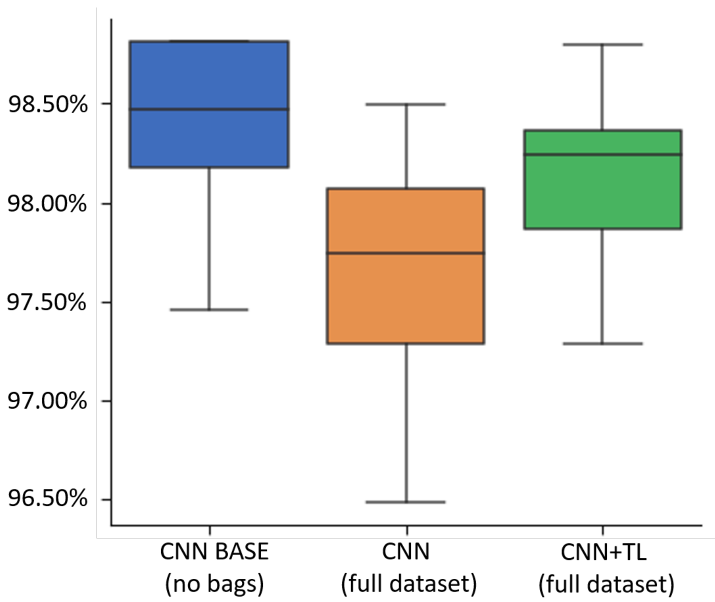

Figure 5. Boxplot of the classifiaction accuracies on the test set.

5. Conclusions

In this work, we implemented a fruit classification system for Colombian fruits and vegetables using a convolutional neural network. A data-set was created with the purpose of analyzing its performance. Unlike other fruit data-sets, this takes into account the use of fruits and vegetables inside bags as a natural source of variability, which represents an additional challenge to the system. Moreover, this is the first time that the use of transfer learning on fruit classification is analyzed, specifically the transfer of knowledge from a model trained with bagless fruit images to one which includes bags.

From the numerical results obtained, we intend to highlight the use of transfer learning in neural networks which is the re-utilization of weights in one or several layers of a pre-trained network to accelerate training and improve the performance of the model. The CNN model with transfer learning trained in this work achieves an accuracy matching state of the art for the Colombian fruits data-set.

References

[1] Dubey S R and Jalal A 2012 *Procedia Engineering* **38** 3449
[2] Rocha A, Hauagge D C, Wainer J and Goldenstein S 2010 *Computers and Electronics in Agriculture* **70(1)** 96
[3] Zawbaa H M, Abbass M, Hazman M and Hassenian A E 2014 Automatic fruit image recognition system based on shape and color features *International Conference on Advanced Machine Learning Technologies and Applications (AMLTA), Communications in Computer and Information Science* vol 488 (Egypt: Springer) pp 278–290
[4] Zhang Y, Wang S, Ji G and Phillips P 2014 *Journal of Food Engineering* **143** 167
[5] Lecun Y, Bottou L, Bengio Y and Haffner P 1998 *Proceedings of the IEEE* **86(11)** 2278
[6] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J *et al.* 2018 *Pattern Recognition* **77** 354
[7] Patino-Saucedo A, Rostro-Gonzalez H and Conradt J 2018 Tropical fruits classification using an alexnet-type convolutional neural network and image augmentation *International Conference on Neural Information Processing* (Cambodia: Springer) pp 371–379
[8] Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *Proceedings of the 25th International Conference on Neural Information Processing Systems* vol 1 (New York: Curran Associates Inc.) pp 1097–1105
[9] Taylor L and Nitschke G 2018 Improving deep learning using generic data augmentation *IEEE Symposium Series on Computational Intelligence (SSCI)* (Bangalore: IEEE)
[10] Yosinski J, Clune J, Bengio Y and Lipson H 2014 How transferable are features in deep neural networks? *Proceedings of the 27th International Conference on Neural Information Processing Systems* vol 2 (Cambridge: MIT Press) pp 3320–3328
[11] Bisong E 2019 Google colaboratory *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners* (Berkeley: Apress) pp 59–64
[12] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray D G, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y and Zheng X 2016 Tensorflow: A system for large-scale machine learning *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (Savanna: USENIX Association) pp 265–283