Paper Review: 03

Title: Deep Learning-Based Intelligent Apple Variety Classification System and Model Interpretability Analysis.

Published date: 19th February, 2023

Link: https://www.mdpi.com/2304-8158/12/4/885

1. Introduction

There are over 7500 varieties of apples, and people spend a lot of time sorting, packing, and labeling apples before selling them. Several rapid and non-destructive techniques for differentiating apple varieties have emerged, including electronic noses, visible and near-infrared spectroscopy, and image processing-based methods.

Deep learning has been successfully applied to the automatic identification, classification, and detection of fruits and vegetables with strong capabilities in feature learning. Recently, CNNs have been used for apple recognition tasks, including quality assessment and bruise detection. A few studies have also reported on the use of CNNs in apple variety recognition. Apple quality detection and grading tasks involve fewer classes, while apple variety classification tasks are recommended to be performed with as many classes as possible to approximate real-world scenarios.

The related work mentioned above and most applications of CNN approaches in the food field are generally focused on the performance of different models, while fewer studies have involved and investigated the interpretability of models. However, model interpretability is highly correlated with its credibility. Based on the above, we employed two frameworks of CNNs with transfer learning to automatically classify 13 types of apples. The obtained results contribute to autonomous robotic fruit harvesting and post-harvest technology and further accelerate the development of agro-based industries. We used five CNNs from two different frameworks to classify 13 classes of apple. We used three visualization methods to reveal how the "black box" models make classification decisions.

A publicly available benchmark fruit dataset, Fruits-360, contains 90,483 images from 131 categories of fruits and vegetables, including 8538 images of 13 classes of apples from a wide range of varieties.

2.2. Training and Testing Datasets Set-Up

Thirteen classes of apple images from Fruits-360 were used to build two datasets, with training-to-testing ratios of 2.4:1.0 and 1.0:3.7, respectively.

2.3. Network Architectures

Series networks are neural networks with layers arranged one after the other, with only one input layer and one output layer. AlexNet and VGG-19 are representative series networks that have achieved good performance in image recognition and classification.

A directed acyclic graph network (DAG network) is another structure of neural network used for deep learning, and the residual network (ResNet) is a type of DAG network with residual connections that bypass the main network layers.

2.4. Transfer Learning

In general, CNNs perform best when trained on large datasets, but currently used CNNs are trained on small datasets and are therefore overfitted, making the results unscientific and unconvincing.

To overcome the above problems, transfer learning was used. This method uses the information collected from an established model to start over with a different problem, and can help reduce the dependence of deep networks on computer hardware and training time.

2.6. Metrics for Performance Evaluation of CNN-Based Models

A confusion matrix is a table layout used to describe and visualize the performance of a trained model on a testing set.

The accuracy, precision, recall, and F1-score of a model are calculated from the total correct positive cases and the total negative cases over the total number of cases. The macro-average values of precision, recall, and F1 are also calculated.

3.1. Performance of the Different Trained Models

The number of misclassified images increased noticeably for each trained model in Figure 3, resulting in a decrease in the overall classification accuracy of each trained model. All five models achieved high overall classification accuracies on dataset A, but the size of the testing set on dataset B significantly decreased the overall classification accuracy of all models, as well as other metrics. However, VGG-19 and the three ResNets still maintained good performance on dataset B.

We set up two datasets to evaluate the performance of CNN-based models, and found that the model size, overall classification accuracy, and training and testing times increased as the model depth increased. This is consistent with many recent studies that suggest that network depth is a key factor in leading the results.

The ultimate goal of a multi-class classification task is to achieve the most accurate recognition of a single image in the shortest possible time. The training times for the models were long, but still acceptable. VGG-19 achieved the highest accuracy and macro-F1 on both datasets, while ResNet-50 achieved the highest accuracy and macro-F1 on both datasets. However, VGG-19 took 4.7 times longer to recognize an image than ResNet-50, which reduces its efficiency in practice.

3.2. Model Interpretability Analysis 3.2.1. Feature Visualization

Traditional fixed-feature-based machine learning methods are not robust or suitable for complex tasks, because fruits have many inter-class and intra-class similarities and variations. However, CNNs are able to automatically learn and integrate features from training images and use them for classification tasks.

Different models interpret the same class of apple in different ways, and different models learn the true differences between classes of apples. For instance, AlexNet and VGG-19 generate different feature visualization images for each type of apple. The feature visualization images generated by the three ResNet-based models were different and abstract, which is difficult to interpret. This phenomenon is due to the different depths of the models, because deeper layers can extract more advanced and complex features than the relatively shallow model.

3.2.2. Strongest Activations

VGG-19 and ResNet-based models recognize apples based on their contours or shapes, whereas the strongest activations show that DAG networks and AlexNet-based models recognize apples based on their entire region.

LIME was applied to CNN-based models to improve the interpretability of the models.

Section 3.2 presents three visualization methods to explore the five CNN-based models' working mechanisms in this task. The feature visualization images show the different understanding of apple images by different trained models, while the strongest activations and LIME images show how and why different trained models make classification decisions.

4. Conclusions

Five CNNs from two different structures were used on two datasets to identify and classify 13 classes of apples. The results showed that the dataset configuration had a significant effect on the classification results, and that the model sizes, accuracies, and training and testing times increased as the model depth increased.

Our future work will focus on increasing the number of apple classes used for classification, developing a new method that can automatically perform hyperparameter optimization, and investigating the interpretability of CNN-based models.