Intelligent Fruit Recognition System Using Deep Learning

Ha Huy Cuong Nguyen[1(✉)], Anh Tuan Luong[2], Trung Hai Trinh[3], Phan Hieu Ho[1], Phayung Meesad[4], and Thanh Thuy Nguyen[5]

[1] The University of Danang, Danang, Vietnam
nhhcuong@sdc.udn.vn, hophanhieu@ac.udn.vn
[2] Dong A University, Da Nang City, Vietnam
tuanla@donga.edu.vn
[3] Vietnam-Korea University of Information and Communication Technology, Danang, Vietnam
tthai@vku.udn.vn
[4] Faculty of Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
phayung.m@itd.kmutnb.ac.th
[5] Faculty of Computer Science, VNU University of Engineering and Technology, Hanoi, Vietnam
nguyenthanhthuy@vnu.edu.vn

Abstract. Industrial Revolution 4.0 has made us people more professional, automating all production stages from office work to project work on farms. In the precision agriculture, it is very urgent to bring new and effective solutions to using artificial intelligence for people to use and improve the manual steps gradually, and increase the automation feature. So, automatic fruit recognition technique is the latest trend and effective technique in precision agriculture. This paper proposes a technical solution for fruit classification using deep learning. Automatic fruit identification using computer vision is considered a challenging task. This is because there are similarities between fruits and changes in the external environment such as light affect the fruit recognition model. Most previously implemented techniques have some limitations since their testing and evaluation is done using a limited set of data sets. Some implementations, does not consider changes to the external environment for the image are considered in this implementation. In this paper, exploring part of the deep learning algorithms was achieved and discovered strengths and weaknesses for these algorithms. The knowledge was gained on deep learning and a model was built that could recognize fruits from images.

Keywords: Fruit recognition · Computer vision · Image processing · Image classification · Convolution Neural Network

1 Introduction

Artificial Intelligence (AI) is a system of machines or application software programs created by humans that can automatically perform intelligent behaviors

P. Meesad et al. (Eds.): IC2IT 2021, LNNS 251, pp. 13–22, 2021.
https://doi.org/10.1007/978-3-030-79757-7_2

such as the ability to communicate with humans, can learn and adapt intelligently to many situations. The narrow areas of AI include: Machine Learning; Deep Learning; Virtual Reality; Augmented Reality; Mixed Reality and Computer Vision. In the content of the topic, we mention the field of deep learning (Deep Learning). Deep Learning models, such as Convolutions Neural Networks (CNNs), have been successfully applied in image classification, text and speech recognition problems.

In recent years, we have witnessed many outstanding achievements in Computer Vision. Large image processing systems such as Facebook, Google or Amazon have included in their products intelligent functions such as facial recognition, self driving car development or automatic delivery.

Researchers in the country recently are focusing on the field of Machine Learning. Machine Learning is a method of analyzing data from which to automate the construction of analytical models. It can be said that this technology is very promising to bring optimal support for businesses with many real world applications, such as voice recognition and image recognition.

Advantages of deep learning models that automatically learn the characteristics of data to establish the new features and data classification. The problem of classification when meeting a large number of dimensions is often difficult. The classification model gives good results on the training set but low results on the test set. The common problem is that the data with the number of dimensions is too large up to thousands of dimensions and the data is separated in a large number of dimensions, so finding a good classification model is difficult because there are too many possibilities model selection.

In intelligent fruit recognition and precise agriculture, timely and accurate estimation of the maturity of the orchard will help improve and improve quality, ensure fruit yield and harvest planning. In addition, timely and accurate determination of the ripening period in the fruit maturity stages will allow to reduce farming costs, post-harvest storage costs and increase economic value, to meet the requirement market. These issues are of great concern in the direction of agricultural development to take advantage of high technology in the coming time in Vietnam.

The trained human investigators do the inspection for the quality of fruit by seeing and feeling. This method is significantly inconsistent, fickle and decisions may vary when considered by different investigators. In this type of environment, the analysis of fruit is done for several aspect criterion and it is a task that forms a sequence that includes the repetition of the actions or events. For this purpose, machine vision systems can be best suited for conventional analysis and quality assurance. In precise agriculture, computer vision systems and image processing are already a growing research area which is a significant analyzing technique.

To overcome all the problems mentioned above, the aim is to develop an automated system for recognition of particular fruits using machine learning techniques.

Image processing and computer vision techniques are used on a large scale today and are being applied in many domains or streams. Among these domains,

fruit recognition from image is one of them. Fruit recognition is a very challenging task as fruits may contain various similarities between them. Image recognition needs more computation power than most of the text base data classification used widely. However, the fruit recognition model can be beneficial for the people and can be used in a less expensive device.

Object detection is a specific application or process of detecting an object from the specified scene using a certain method or measure. Before the emergence of deep learning technology, the methods of finding the object from a scene were achieved by some prior knowledge and by establishing the mathematical models.

The work is organized as follows: Sect. 2 provides the Related Works; Sect. 3 presents Design and Implementation; Sect. 4 presents Performance Evaluation and Sect. 5 presents our conclusions and suggestions for future work finally.

The paper emphasizes on the use of image processing and computer vision technology that is used in the field of food industry and agriculture. Size, color, shape, texture and defect are the most important quality characteristics of fruits.

2 Related Works

Researcher Shadman [1] and colleagues proposed improved solution developed by Mango Yolo architecture (based on YOLO-v2, YOLO-v3, YOLO-v3 [2]) to develop show mangoes in the garden with an average accuracy of 0.983.

Dang et al. [3] proposed a system which specified the use of image processing and computer vision technology in the field of food industry and agriculture. Size, color, shape, texture and defect are the most important quality characteristics of agricultural products. The manual task of inspection of food done by the investigators is replaced by the computer vision systems which generated more authentic, equitable and non-destructive results. The quality inspection of computer vision included four main steps, namely, acquisition, segmentation, feature extraction and classification.

Li et al. [4] proposed a system that presents object detection based on deep learning for small samples. The proposed system contains three modules: syntactical samples generator, object detection network, and semantic relevant detection. The detection of indoor multiple small objects could be achieved with the use of this system. With considerations of subjective and objective approaches, the experimental results concluded that the proposed system produced better results than the other existing systems.

Tian et al. [5] proposed a system that uses the classical methodologies of object detection. The classic methodologies and the deep learning methodologies used in object detection are studied with respect to its relation and differences between them. Then on the basis study, the limitations and design of the model for object detection is proposed.

Avazjon et al. [6] proposed a system for recognizing fruits in treetops aimed for automatic fruit picking. For this purpose, an expert system must be equipped to an intelligent robotic fruit picker. The robotic fruit picker must analyze the information obtained by vision sensors. The vision sensors used are cameras and

laser scanners. The robot must detect physical objects in the observed scene and must decide whether the detected object is fruit or leaf or branch. In uncontrolled situations, such as overlapped fruits or shades, it must make correct decisions. For the detection of fruits on treetop, an approach based on detection of convex surfaces is used. It is assumed that most of the fruits have convex shape. Hence, it is used to distinguish fruits from leaves and branches of the tree.

Rajeshwari et al. [7] proposed a system for accurate apple recognition with robotic fruit picking. A machine learning method is applied to a detection system to improve the accuracy of fruit recognition and achieve a better performance. To help the robot to perform the operation, a three-dimensional (3D) data of a scene is captured directly and accurately by laser radar or depth camera.

Bargoti et al. [8] proposed a system of optimal structure that is applied by design pattern recognition and classification. For solving of pattern recognition and classification tasks, the main problem in artificial networks is the evaluation of the number of hidden neurons. Multi layer perception is the most useful artificial neural network for estimating the functional structure in classification.

Munera et al. [9] proposed a system for green fruit feature extraction. This is done with a combination of the techniques of color model conversion, thresholds, histogram equalization, spatial filtering with Laplace and Sober operators and Gaussian blur. Along with this system, in order to recognize and count the number of green fruits, an algorithm is built and tested.

3 Design and Implementation

The automatic fruit recognition system is developed to make use of computer vision and image processing techniques in the food industry. Segmentation is used to simplify image analysis and image features and descriptors are used for feature extraction. On the basis of color, shape, size and texture, the quality analysis of fruits are done. [2] The architecture of the system is depicted in Fig. 1. The user must give an input image to the system by specifying the path of the image. After that the user will get the output as the name of fruit along with the input image. Fruit prediction is done using the algorithm Deep Convolution Neural Network in Keras. There are two models as one for fresh fruit recognition and second for rotten fruit recognition.

3.1 Fresh Fruit Recognition

There are total 13 classes of fruits used for training and validation purposes. These 13 classes of fruits include apple, blueberry, lemon, mango, orange, pear, pineapple, pomegranate and walnut. The required data set of these fruits were taken from the Fruit-360 data set of Kaggle. The number of fruit images taken for the purpose of training and validation are represented in Fig. 2.

The first step will be to provide an image data set to the system. The data set comprises of 13 classes. The second step will comprise of storing the data set into its memory. This image data set will be then used for pre processing.

Table 1. Fruit images distributed used for training and validation purpose

Number	Name	Train imgs	Test imgs
0	Apple	6404	2134
1	Avocado	918	309
2	Banana	1430	484
3	Blueberry	462	154
4	Lemon	982	330
5	Mango	916	308
6	Orange	479	160
7	Pear	5037	1689
8	Pineapple	983	329
9	Pomegranate	492	164
10	Rambutan	492	164
11	Strawberry	1230	410
12	Walnut	735	249

Pre-processing of the image is done to remove the noise and get clear image. The third step will comprise of feature extraction from it. The features such as edge detection, texture and color are included. Edge detection can be done by detecting the edges of the image. Texture is used to detect the skin disease and defects present on the fruit. Color features are used to identify the color of the fruit including mean and variance of the RGB color model. Based on the features extracted, Convolution Neural Network (CNN) in Keras will be trained.

Due to the high level of performance given by deep learning, a subset of machine learning, it is becoming very popular across many types of data. CNN is built for classifying images which makes use of deep learning. The building of CNN is easier with the use of Keras library in Python. The model uses layers of CNN such as Conv2D and MaxPooling2D. Conv2D is also called as 2D Convolution Layer. This layer produces tensor of outputs by creating convolution kernel that is convoluted with the input layer. MaxPooling2D is used for max pooling operation for spatial data. Spatial data can be defined as representation of information about a physical object by numerical values. Selecting the maximum element from the region of the feature map covered by the filter is the operation performed by max pooling layer. To reduce the dimensions of the feature maps, pooling layers are used so. Thus, it can be concluded that use of pooling layers reduces the number of parameters to learn and the amount of computation performed in the network.

In compiling the model, a loss function, an optimizer and a list of met rices are used. The model includes categorical cross entropy as a loss function and admix as an optimizer [10]. These two were used because after trying with several other loss functions and optimized, amongst them, they gave the best results. A

summary of the model is generated using the function after carrying out the first 3 steps. Figure 3 shows the CNN architecture designed for fresh fruit recognition.

The fourth step will be predicting the image based on trained model. The test image will be used in this step. Users need to specify the path of the image that is where the image is stored in the computer system. The image path should be precise. If the image path is incorrect, then prediction will not be done, and error will be thrown. If the path is correct, then users will get the desired output. Trained model will then check for all the possibilities from the given image. The possibilities will include getting the model and weights.

For the purpose of graphical user interface, Flask was used. Flask is a web framework that provides tools, libraries and technologies that allow one to build a web application. This web application is designed and implemented with sample input and output as depicted in Fig. 3.

There are total 3 classes of fruits used for training and validation purposes. These 3 classes of fruits include apple, orange and banana. The required data set of these fruits was taken from the Fruit-360 data set of Kaggle. The number of fruit images taken for the purpose of training and validation are represented in Table 1. The procedure used is the same as used for fresh fruit recognition. In compiling the model, a loss function, an optimizer and a list of met rices are used. The model includes categorical cross entropy as a loss function and ada max as an optimizer. These two were used because after trying with several other loss functions and optimizers, amongst them, they gave the best results. A summary of the model is generated using the function after carrying out the first 3 steps.

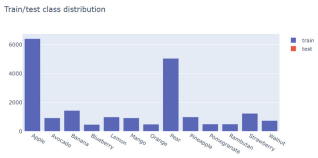

Fig. 1. Summary of training rotten fruit recognition model

No GUI for rotten fruit recognition was implemented because there were only 3 classes of fruits used. But console outputs were generated for this category of recognition. The sample inputs and outputs of the working of system is shown in Fig. 3.

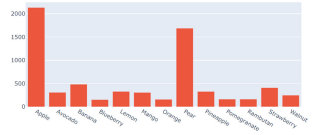

Fig. 2. There is a near identical class distribution in train and test sets

3.2 Improved ResNet50 Fruit Recognition Model

There are total 13 classes of fruits used for training and validation purposes. The required data set of these fruits was taken from the Fruit-360 data set of Kaggle. The number of fruit images taken for the purpose of training and validation are represented in Table 1. This is paper, approach ResNet50 Fruit Recognition Model Architecture from this model, we recommend three more layers: Dropout, Dense and Softmax.

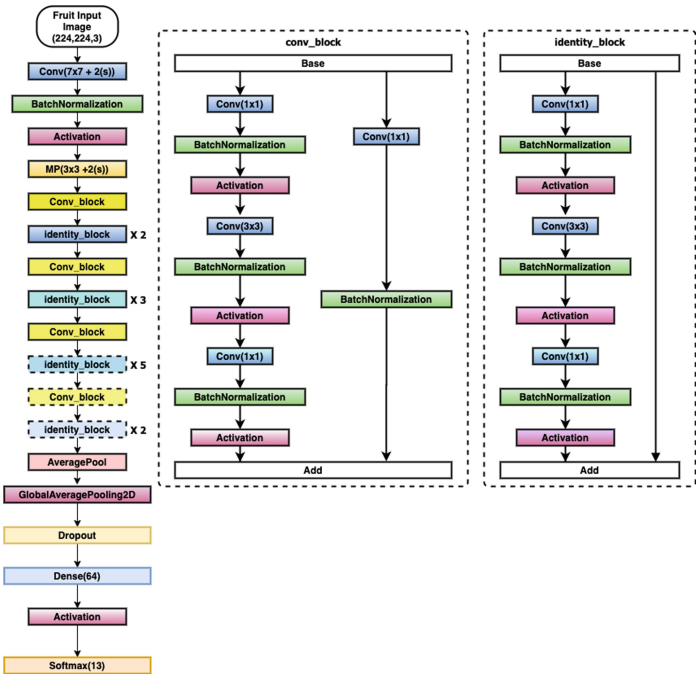

Fig. 3. ResNet50 fruit recognition model architecture

Figure 3. Example network architectures for ImageNet. Based on the above plain network, in this model insert 3 shortcut connections which turn the network into it counterpart residual version. The models are trained on the 5.30 milltion training images, and evaluated on the 50k validation images.

4 Performance Evaluation

The implemented models were checked for classification performance with different optimizers. This is done by representing the values for training accuracy, training loss, validation accuracy and validation loss, respectively. These variations are calculated for fresh fruit recognition with use of different optimizers as follows (Fig. 4).

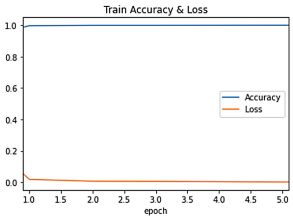

Fig. 4. ResNet50 fruit recognition train and validation

After comparing with optimizers, the best results were obtained from soft max optimizer than others. Hence use of soft max optimizer for fresh fruits recognition was done. Also, false predictions for some fruits were observed. They were as follows.

The image of Pomegranate was inputted as the test image as shown in Fig. 5, and the output was displayed as Apple with 98.88% of accuracy as shown in Fig. 6. Thus, prediction was false because the model considered the color feature of the input image and the prediction for the color of Apple was more in percentage as compared to Pomegranate.

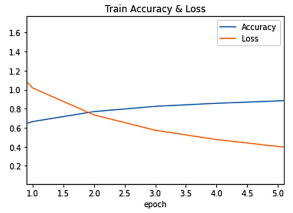

Fig. 5. Convolutional neural network fruit recognition model train

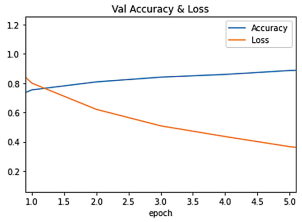

Fig. 6. Convolutional neural network fruit recognition model validation

In the paper, the refining and release techniques were applied to all layers of the input image data set trained from the improved R-CNN model. In order to maintain the calculation results through the trained steps, the set of typical

classes of the trained base model. Experiments conducted to extract some of the selected data a subset compared to the initially divided classes. In the experiment script in all the proposed method retains deep learning rate is 0.00005 for the initial class and 0.000055 for the final layer. After 20 min of training, the accuracy achieved is 99.33%.

5 Conclusion

Currently, in precision agriculture, it is very necessary to apply artificial intelligence solutions to create a correct agriculture from the selection of raw materials to the final stage of harvesting. So here is the article we emphasize on the use of computer vision and image processing technology used in the food industry and agriculture sectors. In addition to previously known identifiers: size, color, shape, texture and defects are the most important quality characteristics of the fruit. We fixed the problems encountered during the manual fruit inspection, with the computer vision systems in use. The implemented system provides authentic, fair and non-destructive results for the fruit grading task. In the article, we have conducted research on previously achieved deep learning algorithms from which we discover the strengths and weaknesses of these algorithms. Knowledge is acquired through deep learning and a built-in model is able to recognize fruit from images.

Acknowledgment. The authors wish to express their appreciation to the Ministry of Education and Training for supporting this research project as part of the Ministerial Program of Science and Technology CTB.2021.DNA. “Research on applying deep learning model to recognize ripe pineapple period in Quang Nam - Da Nang”.

References

1. Sakib, S., Ashrafi, Z., Siddique, M., Bakr, A.: Implementation of Fruits Recognition Classifier using Convolutional Neural Network Algorithm for Observation of Accuracies for Various Hidden Layers. arXiv e-Journal, pp. 1–4 (2019)
2. Alexey, A.B.: Yolo mark Apple detection during different growth stages in orchards using the improved YOLO-V3 model (2018)
3. Chung, D.T.P., Van Tai, D.: A fruits recognition system based on a modern deep learning technique. In: Journal of Physics: Conference Series, vol. 1327, pp. 1–5 (2019)
4. Hou, L., Wu, Q., Sun, Q., Yang, H., Li, P.: Fruit recognition based on convolution neural network. In: IEEE, 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), pp. 18–22 (2016)
5. Tian, Y., Yang, G., Wang, Z.: Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Comput. Electron. Agric. **157**, 417–426 (2019)
6. Marakhimov, A.R., Khudaybergenov, K.K.: Approach to the synthesis of neural network structure during classification. Int. J. Comput. 20–26 (2020)
7. Rajeshwari, P., Abhishek, P., Srikanth, P., Vinod, T.: Object detection: an overview. Int. J. Trend Sci. Res. Dev. (IJTSRD) **3**, 1663–1665 (2019)

8. Bargoti, S., Underwood, J.: Deep fruit detection in orchards. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, pp. 3626–3633 (2017). https://doi.org/10.1109/ICRA.2017.7989417
9. Munera, S., Amigo, J.M., Blasco, J., Cubero, S., Talens, P., Alexios, N.: Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging. J. Food Eng. **214**(3), 29–39 (2017)
10. Nguyen, H.H.C., Nguyen, D.H., Nguyen, V.L., Nguyen, T.T.: Smart solution to detect images in limited visibility conditions based convolutional neural networks. In: Advances in Computational Collective Intelligence. ICCCI 2020. Communications in Computer and Information Science, vol. 1287, pp. 641–650. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63119-2-52