Classifying maturity of cherry tomatoes using Deep Transfer Learning techniques

Danh Phuoc Huynh¹, My Van Vo¹, Nghin Van Dang¹, and Tri Quoc Truong¹
¹Faculty of Engineering, Van Lang University, 45 Nguyen Khac Nhu street, Co Giang ward, District 1, Ho Chi Minh city, Vietnam
tri.truong@vlu.edu.vn

Abstract. This research studies a method to classify the tomatoes’ maturity by using deep transfer learning techniques. We carry out sorting systems adopting three pre-trained convolutional neural networks of VGG16, VGG19, and ResNet101. The experimental results show that the VGG19 model obtains a high precision on both the train set and the test set.

1. Introduction
Tomatoes are one of the popular vegetable crops that are grown all over the world. They are healthy because of their good source of vitamin A and C. Along with the constant increase in demands for tomatoes, the consumer also pays more attention to their quality. Tomatoes maturity level plays an important role in product quality. However, the traditional classification of tomatoes is often based on the practical experience of farmers. It takes a lot of time and inaccuracy in harvesting the crop. There have been researches in the field of automatic classification of tomatoes over the last decade [1, 2, 3].

The rapid growth of recent techniques in image processing, computer vision and especially Machine Learning (ML) paved the way for high-quality agriculture. In [4], the authors implemented a fuzzy architecture based on RGB color to classify six categories of tomatoes. This approach obtained a good classification in RGB color space. The limitation of this paper was that the result is a suited prediction in the conditional environment with controlled lighting, fixed black background. In another study, by combining histogram feature extracting and Naïve Bayes Classifier, Kusuma et al. proposed a new tomato maturity classification [5]. The histogram feature acquired the pixel intensity values of the tomato’s image, the information of a relative, brightness, and contrast. Hence, the Naïve Bayes Classifier learned the extracted values obtained from the histogram feature to make a more accurate prediction. According to the results, the proposed method produced an accuracy rate of 76%. In fact, the main concern of the classification method was accuracy and computational cost.

Recently, Deep Learning (DL) has attracted intensive research interest in the classification of tomatoes. By using a lot of labelled images of tomatoes as the input, the DL model is trained to learn the tomato’s features. After a period of training, the “learned model” was then used to predict what type of tomatoes. However, the traditional DL techniques acquired big data to conduct a good classification. In recent years, many researchers have used transfer learning methods to overcome these problems [6, 7]. For instance, to detect the tomato crop disease, Ouhami et al. obtained deep learning models DensNet, 161 and 121 layers, and VGG16 [6]. The results show that transfer learning has a suitable architecture for the task of plant disease detection with less time-consuming and high accuracy.

In this study, we focused on the task of maturity classification of tomatoes. Based on the dataset of 1374 tomato images, we divided tomatoes into 3 categories: green, yellow, and red. By taking advantage
of the transfer learning method, we employed three Deep Learning models of VGG16, VGG19, and ResNet101 to figure out the best model with great accuracy.

2. Dataset description
In this study, we acquired the dataset, named Fruits-360, which contained 90483 images of 131 popular fruits and vegetables in [8]. For simplicity, we chose 1374 images of tomatoes and divided into 3 classes: green, yellow, and red. The examples of 3 types of tomatoes were illustrated in Figure 1. All images in this dataset extracted the tomato from the background. Tomatoes were then scaled down to 100 × 100 pixels.

![Figure 1. Three classes of tomato in scale of 100 × 100 pixels.](image)

3. Deep Transfer Learning Model
3.1 Convolutional Neural Networks
Convolutional Neural Networks (CNNs) are some types of Deep Learning (DL) models. The fundamental structure of DL model is illustrated in Figure 2. Here, the input the input $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$, remarked by blue nodes, is fed into the network through the Input layer. In image processing $\mathbf{x}$, expresses the pixel matrix of the image. In image processing, $\mathbf{x}$ expresses the pixel matrix of the image. Then, $\mathbf{x}$ is combined with weight matrix $\mathbf{W}$ to generate the input $\mathbf{W}$ of the Hidden layer which is remarked by yellow nodes. We note that a deep neural network consists of numerous hidden layers and numerous nodes in each layer. Hence, the output of the previous layer becomes the input of the next layer. At the final output layer remarked by red nodes yields:

$$\hat{y} = f(W; x)$$

where, $f$ is a function which describes the relationship between the input and the output. $\hat{y}$ is the predicted result. The algorithm will modify the weight matrix $W$ such that $\hat{y}$ is close to the true output $y$. In the problem of classification of tomatoes’ maturity, $y$ is considered as red, yellow or green fruit.

Similarly, CNNs focus on solving the tasks relating digital image processing, computer vision. Such roles of CNNs are remarked as object classification, object detection and recognition or image segmentation [9]. In case of object classification tasks, the typical CNNs can be divided into two blocks: block of image’s feature learning (called Block 1) and block of classification (called Block 2) as shown in Figure 3.

Block 1 consists of convolutional (abbr. *conv*) layer, max pooling layer and Rectified Linear Unit (ReLU) layer which is followed by a convolutional layer. The objective of the convolutional layer is to extract the feature of the image, such as edges, color or gradient orientation. Subsequently, the max pooling layer learns the dominant feature. Furthermore, the pooling layer is responsible for decreasing the size of convolved features expressed as a matrix. It makes the computational cost reduced, which is acquired to process the dataset.

Block 2 contains Flatten layer, Fully connected layer and Soft-max layer. Here, the image represented as a matrix is flattened into a column vector at Flatten layer. Then, the flattened output is fed into the Fully connected layer for classification purposes. At the rest of CNNs, by using the Softmax technique, the model is able to compute the highest score relating to the input image. We note that in our study, at the output we obtain three scores (between 0 and 1) that is predicted to be green, yellow, and red.
3.2 Pre-trained model CNNs

In order to reduce the demand on large dataset and computational cost of DL model, we obtained a transfer learning method which takes advantage of pre-trained model. This study takes pre-trained weights from the VGG16, VGG19, and ResNet101 [10, 11]. These models trained on big labelled images from scratch for similar tasks of fruit classification. Hence, using a pre-trained model gains much faster and more accurate compared with models trained on a large dataset.

VGG16 is a famous deep CNNs which was proposed by Simonyan K and Zisserman A in [10]. Number 16 refers to the fact that this CNN has 16 layers consisting of 13 conv layers followed by the maxpooling and 3 fully connected layer. The architecture of the VGG16 model is shown in Fig 4. Meanwhile, VGG19 is a version of the VGG family model. As illustrated in Figure. 5, similar to that of VGG16 model, it contains 16 layers of conv layers and 3 fully connected layers.
Figure 4. Architecture of VGG16 model which consists of 16 layers.

4. Results
The dataset is divided into two parts of 80% for training, named train test and 20% for testing, named test set. Figure 5 shows the accuracy rate (right) and the loss evaluation as training the model in case of VGG16 model. Consequently, Figure 6 and 7 respectively show the similar manner of VGG19 and ResNet101 models. We executed the training of all three models in 30 epochs. As shown in Table 1, we can observe that the VGG19 obtains excellent accuracy on both train set and test set.

Figure 5. Accuracy and loss evolution of VGG16 model over 30 training epochs.

Figure 6. Accuracy and loss evolution of VGG19 model over 30 training epochs.
5. **Conclusion**

This research was conducted to study tomatoes classification by using deep transfer learning techniques. We employed 3 pre-trained CNNs models of VGG16, VGG19, and ResNet101. The experimental results show that the VGG19 model obtained a high precision of 94.14% accuracy in evaluating the level of how the cherry tomatoes ripen. However, in this classification task, we only considered 3 classes of tomatoes, so that it is easy to obtain high accuracy. In future work, we will consider enlarging classes of data and train the above CNNs model from scratch for real tomatoes classification.
References

[1] Fuentes A, Yoon S, Kim S C and Park D S 2017 A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition Sensors 17(9)

[2] Rangarajan A K, Purushothaman R and Ramesh A 2018 Tomato crop disease classification using pre-trained deep learning algorithm Procedia computer science 133 1040.

[3] Zhang L, Jia J, Gui G, Hao X, Gao W and Wang M 2018 Deep learning based improved classification system for designing tomato harvesting robot IEEE Access 6 67940.

[4] Villaseñor-Aguilar M J, Botello-Álvarez J E, Pérez-Pinal F J, Cano-Lara M, León-Galván M F, Bravo-Sánchez M G and Barranco-Gutierrez 2019 Fuzzy classification of the maturity of the tomato using a vision system Journal of Sensors.

[5] Kusuma A, and Putra M D M 2018 Tomato maturity classification using naive bayes algorithm and histogram feature extraction Journal of Applied Intelligent System 3(1) 39.

[6] Ouhami M, Es-Saady Y, El Hajji M, Hafiane A, Canals R, and El Yassa M 2020 Deep transfer learning models for tomato disease detection In International Conference on Image and Signal Processing 65.

[7] Pattnaik G, Shrivastava V.K, and Parvathi K 2020 Transfer learning-based framework for classification of pest in tomato plants Applied Artificial Intelligence 34(13) 981.

[8] Oltean M and Muresan H Fruits 360 dataset on github URL: https://github.com/Horea94/Fruit-Images-Dataset

[9] Girshick R, Donahue J, Darrell T, and Malik J 2015 Region-based convolutional networks for accurate object detection and segmentation IEEE transactions on pattern analysis and machine intelligence 38(1) 142.

[10] Simonyan K, and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.VGG19.

[11] He K, Zhang X, Ren S, and Sun J 2016 Deep residual learning for image recognition In Proceedings of the IEEE conference on computer vision and pattern recognition 770.