Classification and detection of chili and its flower using deep learning approach

W H M Saad¹,², S A A Karim³, M S J A. Razak⁴, S A Radzi⁵, Z M Yussof⁴

¹ Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer (FKEKK), Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia.
² Centre for Telecommunication Research & Innovation (CeTRI), Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia.
³ Department of Electrical Engineering, Politeknik Ibrahim Sultan, 81700 Pasir Gudang, Johor, Malaysia.
⁴ SolokFertigasi by Perwira MSJ Sdn Bhd, 75460 Duyong, Melaka, Malaysia.

wira_yugi@yahoo.com

Abstract. This study presents an application of using Deep Neural Network (DNN) based detector to detect chili and its flower in the chili plant image. Detecting chili on its plant is important for the development of a robotic vision to automatically picking the chili. Only one type of local chili variety is used in this study from the species of Capsicum frustecens. Five hundred of chili plant images were captured from multiple angles and each of images was marked and labelled for any present of chili and its flower. These images were divided into 70-30 per cent proportion for training with validation and testing purposes accordingly. This project uses Faster Regions with Convolutional Neural Networks (R-CNN) as a deep learning model for training that contains around 177 numbers of layers including input and output layers. The classifier network was trained to optimize all parameters involved in chili and its flower classification and detection. The classification and detection accuracy are measured on the tested images. The result shows very good accuracy in validation and testing for classification and detection especially with the image of chili plan is upright position.

1. Introduction

Detection and classification play an important role in image processing. Automatic recognition system has been applied in several applications to make things easier for human to complete their task. For this study, a primary concern is to detect and classify the different parts of the chili plant namely the chili fruit itself and its flower.

There are many types and variety of chili plant that has been commercialized for local and export market. Chilies come in many sizes, colors and shape. The one that we focused on this study is a specific small variety of a species called Capsicum frustecens. The local name in Malaysia for this type of chili is called ‘cili padi’, and the variety that we use for this study is called ‘Cili Padi Bara’. The image sample of the chili and its plant are shown in Figure 1.

There are a few challenges that need to be overcome in classifying and detecting the chili on the plant as the color of unripe chili look like the color of its leaf. Sometimes, the shape of the chili from a certain angle on a 2D image is somehow almost like the shape of the leaf too. By using a bare-naked eye, chili’s outer texture that is slightly shiny in comparison to its leaf make it easier to spot with a good lighting conditions.
and a close distance to the eye since the size of this type of chili is relatively small. Due to this knowledge, it is recommended that to have good lighting of images while collecting the image of the chili for training and validation. At the same time, capturing the image with high number of pixel or literally closer distance in between lens and the plant is recommended to achieve a good accuracy of detection and classification and at the same time it is suitable to be used on the field of chili plucking application as this activity is normally done in a sunny bright day.

A recent study shows that robot vision and artificial intelligence has been a good companion in automatic fruit detection and fruit picking automation [1]. Most of the study has been demonstrated with a relatively large round shape type of fruits such as tomatoes and apples [2]. Each type of fruits and its plant has its own characteristic and the challenge to recognize each type of fruits is different than the other especially by using a traditional computer vision method [5]. Traditionally, each of the feature for the object that we want to detect the need to be learned and understood. By using a deep learning approach, detecting an object like that is becoming simpler with better output accuracy. The drawback of deep learning approach is often caused by high computation cost, but with the current advancement of GPU and FPGA embedded system, it shortens the training time and making the deployment of the system in real-time to be possible.

There are a few deep learning architectures models that have been substantiated in classifying and detecting the fruits [1-4]. One of the good examples is Faster Regions with Convolutional Neural Networks (Faster R-CNN) that has been used in detecting several types of fruits including small fruits [3]. In this study, we are using Faster R-CNN deep learning architecture models for classifying and detecting both chili fruits and its flower on the chili plant image.

This paper consists of five sections including this introduction. Section 2 explains in detail the methodology and development of the network use and its usage in the classification and detection of chili and chili’s flower. After that, Section 3 deliberates on the testing result and discussion. Lastly, Section 4 concludes the study.

2. Methodology
The development of this system makes autonomous chili farming seems possible in the very near future. This study includes the labelling of a large number of chilies and their flowers. Five hundred images with resolution 1000x1500 pixels which are resized from original 4000x6000 pixels of the captured image are used in training. As mention previously, the DNN based object detector which is used in this project is Faster R-CNN and Matlab is used in training with the validation process, detection process and designing of GUI for deployment.

Next, when the input data to an algorithm is too large to be processed, and it is suspected to be notoriously redundant, the input data is transformed into a reduced representation set of features, also named features vector. Transforming the input data into the set of features is called feature extraction. The features used in this project are aspect ratio, height, width, elongation, skewness, compactness,
orientation, area, perimeter and extent. This section presents the technology inputs and processing steps needed in the chili plant identification and recognition. Each of these steps is described in Figure 2 and Table 1.

2.1. Classification process of chili and its flower
The deep learning model used in this study to detect chili and chili flower is Faster R-CNN with 177 layers, including input and output layers. The detection showed in this manuscript uses a prediction threshold of 0.65. Faster R-CNN uses search selective method to find the regions of interests and passes them to a ConvNet. It tries to find the areas that might be an object by combining similar feature into rectangular boxes.

![Flowchart](image)

**Figure 2.** Flowchart for overall process and development

2.2. Development of GUI for system deployment
Matlab App Designer enables the user to create professional apps without having to be a professional software developer. Figure 3 shows the GUI created using App Designer. The ‘Browse Image’ button locates the desire validation image to be viewed and detected. The ‘Predict’ button will start the inferencing on the selected image for the detection of chili and chili flower. The diagram section will show the image with detected chili and chili flower after the ‘Predict’ button was pushed. The bottom
part of the GUI will show the prediction time, and the number of chili and chili flower detected after the prediction process completed.

**Table 1.** A slightly more complex table with a narrow caption.

| Step 1:  | Image labelling.  |
|---------|-------------------|
| Step 2:  | Randomly split labelled image dataset into training and validation dataset (80-20 ratio). |
| Step 3:  | Neural network model training is done using Faster-RCNN and ResNet-50 as the feature extractor. |
| Step 4:  | Training is done using 1000x1500 pixels resolution instead of the original 4000x6000 pixels resolution due to GPU memory constraints. |
| Step 5:  | A few basic detection tests are done using the trained neural network model. |
| Step 6:  | Performance of the deep neural network model is evaluated using mean average precision (MAP). |
| Step 7:  | A graphical user interface (GUI) is built using Matlab App Designer. |
| Step 8:  | The deep neural network model is implemented into the GUI. |

**Figure 3.** Illustration of designed GUI on chili and chili flower detection.

The constructional code begins with loading the image and the preset for its function, as seen in Figure 4. Setting variable ‘doTraining’ to true will perform Faster R-CNN training and ‘doEval’ is used to start evaluation whereas ‘doSingleTest’ is used to perform chili detection for only one single image, it is used with conjunction with the GUI.

```matlab
data = load('AllImage\infoAll.mat');
imageDataset = data.combineData;
imageDataset = struct2table(imageDataset);
doTraining = false;
doEval = false;
doSingleTest = true;
```

**Figure 4.** Implementation code for detection in GUI
3. Result

The DNN based object detector using Faster R-CNN bring out the output of the detection with a certain threshold of confidence label. In this project, a threshold of 0.65 is used since that is the best number before the detection started to wrongly detect the chili due to the size is too small and the leaf of the plant might resemble the shape and texture of the chili inside of the used image. In other words, if an object is detected to have more than 65% resemble chili or chili flower, only the chili and chili flower will be labelled and shown on the resulting image. One of the few results of the chili and flower detection is as shown in Figure 5. The number labelled above each box shows their respective confidence level.

Result evaluations in Figure 6 show that the average precision (AP) of chili detection is 0.36, while the AP for chili flower detection is 0.50. Thus, the mean average precision (MAP) of the object detector developed is 43%. It is understandable why flower has higher AP compared to chili since chili flower has more distinctive features to leaf compared to chili. From observations, it can be deduced that the detector has a substantially acceptable level of accuracy. One of the main causes of low MAP is due to the mislabeling of objects in the training images. The training dataset consists of some chilies and chili’s flowers which are not labelled entirely. Since AP computations involve the overlapping of detection or also known as intersection over union (IoU) of object detection ground truth, the poor labelling of ground truth results in low MAP. One of the examples of poorly labelled ground truth image is as shown in Figure 7, where a lot of flowers and chilies are not labelled, and at some points, two chilis are labelled as one.

![Figure 5. Results of inferencing on an image containing chili and flower](image_url)

![Figure 6. MAP of chili and its flower detection tested with pre labelled image.](image_url)
Figure 7. Example of poorly manual label chili and its flower on the chili plant

When inferencing on the image from Figure 7, the results show even better detection than the manual label image. As shown in Figure 8, more chilies and chili’s flowers that did not been labelled or mislabeled can be detected by the developed system. However, when two chilis or more that is close to each other in the image, the detector still detects them as one chili due to the poor manual labelling in training dataset.

Figure 8. Results of inferencing on the same image in Figure7

One of the main features of the object detector utilized in the detection is the shape of the object. Due to that reason, when the input image is inverted, some the shape of the leaf resembling chili, which causes the object detector to label some leaves to be chili. The detection is more accurate when the input image is in an upright orientation.

4. Conclusion and future suggestion
The complete classification and detection system for chili and its flower detection has been demonstrated. The system shows an effective and reliable classification and detection of chili and its flower images captured by a camera. The developed system is capable of detecting even chili and its flower that is mislabeled when marking it manually for the dataset in training and validation. The detection shows a promising prediction when the confidence level is 65 per cent and higher.

There are a few things that we can do to improve the entire chili detection. One of the ways is to use more dataset and properly labelled the overlap chili as an individual object. Other than that, we can also develop a dataset using a raw data of dual-camera images as it can give information on the depth for better object shape of the feature vector.
Acknowledgements
Authors would like to thank the Machine Learning & Signal Processing (MLSP) research group under Center for Telecommunication Research and Innovation (CeTRI) and Rehabilitation Engineering & Assistive Technology (REAT) research group under Center of Robotics & Industrial Automation (CeRIA) of Universiti Teknikal Malaysia Melaka (UTeM), Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer (FKEKK), UTeM for supporting this work and providing the use of the existing facilities to complete this project. Authors would thank SolokFertigasi establishment under Perwira MSJ Enterprise management for their collaboration and willingness on sharing images of chilis captured from their chili field.

References
[1] W. C. Seng and S. H. Mirisae, “A New Method for Fruits Recognition System,” in Proceedings of the 2009 International Conference on Electrical Engineering and Informatics, ICEEI 2009, 2009.
[2] Ren, S., K. He, R. Gershick, and J. Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE Transactions of Pattern Analysis and Machine Intelligence. Vol. 39, Issue 6, June 2017, pp. 1137-1149.
[3] Girshick, R. "Fast R-CNN." Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile, Dec. 2015, pp. 1440-1448.
[4] Zitnick, C. L., and P. Dollar. "Edge Boxes: Locating Object Proposals from Edges." European Conference on Computer Vision. Zurich, Switzerland, Sept. 2014, pp. 391-405.
[5] Karis, M. S., Saad, W. H. M., Ali, N. M Khor, C. W., Razif, N. R. A., Selamat, A., Abidin, A. F. Z., Ahmad, M. K. A., “Fruit sorting based on machine vision technique”, Journal of Telecommunication, Electronic and Computer Engineering, 8(4), pp-31-35