FOOD DETECTION WITH IMAGE PROCESSING USING CONVOLUTIONAL NEURAL NETWORK (CNN) METHOD

Assyifa Ramdani[1]
School of Electrical Engineering
Telkom University
Bandung, Indonesia
[1]assyifarmdn@student.telkomuniversity.ac.id

Agus Virgono[2]
School of Electrical Engineering
Telkom University
Bandung, Indonesia
[2]avirgono@telkomuniversity.ac.id

Casi Setianingsih[3]
School of Electrical Engineering
Telkom University
Bandung, Indonesia
[3]setiacasie@telkomuniversity.ac.id

Abstract - Currently, the payment process at restaurants is still manual and inefficient because it uses a cash register. A cashier will check what food is ordered, then count it with the cash register. This is not efficient. So food detection devices and automatic food price estimates have the answer to these deficiencies. Food detection aims to facilitate payment at restaurants, and automatic food price estimation using the Convolutional Neural Network (CNN) classification method. The detection accuracy of 6 types of food using the CNN method was obtained 100% with 80% data partition training data and 20% test data with epoch 9000 and learning rate 0.0002, with a detection time of fewer than 10 seconds.

Keywords: food detection, computer vision, CNN, deep learning.

I. INTRODUCTION

The restaurant is one of the means to carry out the Food Service Industry or part of tourism accommodation that plays a role in meeting customer needs [1]. The restaurant is divided into several types, including the canteen. Canteen is a restaurant associated with offices, factories, and schools, a place where workers or students usually get lunch, take breaks with snacks or work hours interlude, study hours. Due to the accumulation of customer populations at certain hours, then food payments have to queue. The queue is a problem that is common in the community or the production process of goods and services [2]. The queue can occur because the level of service demand is higher than the facility's ability to provide services. In restaurants, queues are standard at lunch or dinner hours. Convolutional Neural Network (CNN) is a development from Artificial Neural Network (ANN) to classify the image, image segmentation, and object recognition with high accuracy and high performance [3]. Using the Convolutional Neural Network (CNN) classification, the food detection system is a representation of advanced stages of knowledge by explaining what food is in one frame of the picture taken. This system is applied to food analysis in restaurants, and automatic billing, where the cashier only makes bills.

One of the problems with food payment occurs at the Canteen of Engineering, Telkom University. Food payment has two stages: buying food at one of the shops and food is calculated using a manual calculator and then getting a receipt. The consumer goes to the cashier and pays, and then the receipt is stamped, then the consumer must come to the previous store to submit a stamped note, the payment stage is very impractical. This situation raises the need for applications that make it possible to detect food types and estimate the total price of food. Automatic detection of food images plays an important role. Focus on the restaurants; food recognition algorithms can enable price monitoring, making it possible to simplify the services offered by restaurants. Using the Convolutional Neural Network (CNN) classification, the food detection system is a representation of advanced stages of knowledge by explaining what food is in one frame of the picture taken. This system is applied to food analysis in restaurants, and automatic billing, where the cashier only makes a final bill.

Regarding food image detection for restaurants, Eduardo Aguilar et al. [4] describe food detection using public dataset UNIMIB2016, achieving about 90% accuracy. The convolutional neural network (CNN) showed significantly higher accuracy than did traditional support vector machine-based methods with handcrafted features that the convolution kernels show that color dominates the feature extraction process [5].

In this paper, we take the case of payment at the Canteen of Engineering, Telkom University. Payment using a food detection system so that we no longer use a receipt paper. Detection of food images using the CNN method. For the dataset, we built a food dataset according to the food that is often bought in the engineering canteen. Our contributions are: 1) we built a dataset for food detection by using food that is often bought, and 2) using the CNN method to detect a food image with 80% partition training data and 20% test data with epoch 9000 and learning rate 0.0002.

II. RELATED WORK

1. Estimating Food Calories for Multiple-Dish Food Photos

Food has a calorie; in a meal, the consumer usually does not know how many calories are in a food that is served. Food detection is needed. In a food photograph generally includes several types of food served, and in one picture can be detected what food is served. CNN is a food detection method with quite a reasonable accuracy. CNN can detect any food in an image that contains several types of dishes, CNN is also used for the classification of dishes or what food is served

978-1-7281-9336-6/20/$31.00 ©2020 IEEE

2. Deep Structured Convolutional Neural Network for Tomato Diseases Detection

Outbreaks of plant diseases can cause threats to food security. Technology is needed for machine learning and for recognizing objects. CNN is one technique or method for object detection and object identification in depth. CNN also has architectures such as AlexNet and VGGNet. The architecture is useful for obtaining information on various lengths to capture various correlations between pixels. Use of CNN architecture to further improve performance. This architecture is useful as an effort to modify the number of filters, add depth or width to the network. CNN's accuracy without adding CNN architecture reaches 89% accuracy. Adding VGGNet architecture, making detection accuracy reaches 95.24

3. Image Recognition with Deep Learning

Choosing foods is essential for specific purposes, such as a diet rating system, to avoid obesity, diabetes, and others. Food references can be found by way of food classification. Image classification is challenging because the food image dataset is not linear; for example, there is one food with another, which is not the same, and there is also one food with another food that has similarities of food types and shapes. Foods have many characteristics to be classified, and the classification method can use existing methods in computer vision. CNN is a neural network class that is very effective for the task of image classification and object detection. Food dataset classification using CNN obtained 92.86% accuracy.

III. RESEARCH METHOD

A. Image Processing

Image Processing is a method for converting input images into digital form images, carrying out some operations on them, obtaining enhanced images, or extracting some information so that it becomes output following the intended use [6]. Image processing is one of the signal processing with input in the form of images (image) and output images that have been transformed with specific techniques. Image processing has a variety of processes, such as sampling, quantization, and noise. Samplings are the process of determining the color of specific pixels in the image of a continuous image. Quantization is the process of associating average colors with certain color levels. Noise (noise) is an image or pixel that interferes with image quality.

B. Food

Food is processed material derived from animals or plants, which will be eaten by living things to get energy and nutrition after processing. In a food can have characters and classifications such as color [7], shape, texture, and others. The content of food is very diverse, such as calorie content [8]. Food classification is essential to avoid disease, live a healthy lifestyle, apply diet in diabetics [9], sufferers of food allergies, and help find food calories, nutritional value, and as a food reference [10]. Very diverse foods become a challenge in the classification of food. Food classification can be done using existing computer vision (Jahan, Kekha, & Quadri, 2018).

C. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a method in deep learning that is used for image classification, semantic segmentation, object detection, and feature extraction [11]. CNN is included in the type of deep neural network because of the high network depth and is widely applied to image data [12]. CNN is used for image classification or object detection [13]. Object Detection can be achieved because CNN utilizes the convolution process so that the computer gets information. Object detection can vary, exist for the classification of food objects such as fruit, for example, mangosteen [14], tomatoes and mushrooms, and disease objects. Convolutional Neural Network is divided into 2 main parts: the extraction layer feature and the fully connected layer. Feature extraction consists of a convolutional layer and a pooling layer. Layers on CNN, among others

- Convolution Layer
  This layer serves to find the texture/pattern in the image by conducting the convolution process between the filter and the image. In the convolution layer located on the first layer in the Inception architecture, a 5x5 dimension image is performed between the input image and the $3x3$ filter, with a 1-pixel stride or shift filter.

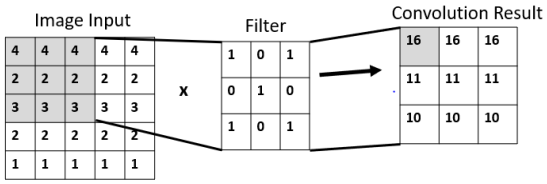

Figure 1. Convolution layer process

The input image is multiplied by the filter using equation 2.1. Then the right side displays the results of the convolution layer. Multiplication of input images with filters such as (4x1) + (4x0) + (4x1) + (2x0) + (2x1) + (2x0) + (3x1) + (3x0) + (3x1), the results obtained are 16. The results of this layer will go through a process in pooling layer.

- Pooling Layer
  The pooling layer does the process of dividing/partitioning several parts in a square shape. After that, the pooling layer will take the maximum value of the pixel part of the convolution layer results. This layer, will find the maximum value by using the max-pooling filter $2x2$ and stride 1 or shifting the filter by 1 pixel.

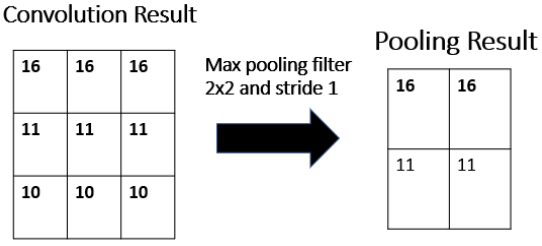

Figure 2. Pooling Layer Process

The results of the pooling layer obtained the most significant values. The results of this layer will proceed to the next process, the fully connected layer.

- Fully Connected Layer
  Last layer is fully connected layer. This layer functions to store the whole layer's results by connecting other neurons to one neuron and storing it in shape. At the fully connected layer, this is the arrangement the results of the previous layer. The result of pooling calculations are 16 and 11, and then 2 vectors are worth 16 and 11. The next stage is softmax; at this stage, the probability obtained from each vector will then be matched with the class that has been formed by the system. By using equation 2.2, the calculation can be seen as follows

$$\sigma(z_1) = \frac{e^{16}}{e^{16} + e^{11}}$$

$$\sigma(z_1) = 0.993307$$

$$\sigma(z_1) = \frac{e^{11}}{e^{16} + e^{11}}$$

$$\sigma(z_1) = 0.006692$$

  The values obtained are 0.993307 and 0.006692 from the vector.

In the food dataset, pictures are taken with the same distance and brightness to obtain the right image used for analysis because it has variability in proper image processing. The type of food used in this final project research is rice (id=1), fried chicken (id=2), spinach (id=3), fried chili sauce (id=4), tofu (id=5), and chicken braised in coconut milk (id=6). Taking pictures is done by placing 3 types of food on a plate with the same color.

Food images are taken with a webcam to produce a right image and a tripod.

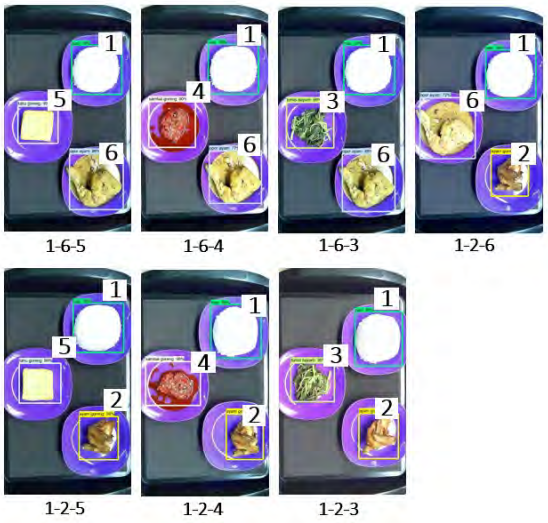

Figure 3. Classes of Food

IV. SYSTEM DESIGN AND OVERVIEW

A. System Overview

In this research, food detection will be applied using Convolutional Neural Network (CNN) base on desktop. The following is the design of food detection base on desktop :

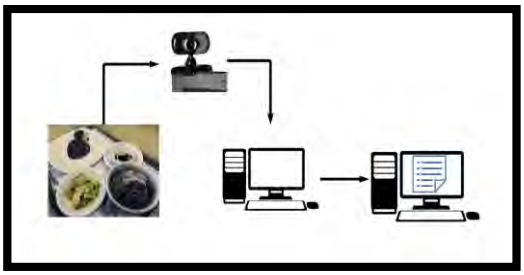

Figure 4. System Overview

In figure 4. Shows the process of food detection. For food information, we must have a dataset about the food, which is will recognition. The dataset will be on the process like training and testing. Food will be captured using a webcam, and the food image will be processed on the desktop. The food image will be recognized using the CNN method, and the data that has been processed will be displayed in a desktop application.

B. Data Collecting

Data collection or dataset is taken from the image of food to be processed; namely, there are 6 types of food for research. The food is available at the Canteen of Engineering, Telkom University. We were shooting in outdoor places so that the light is based on shooting from 1 pm to 5 pm. The shooting distance between the camera and food is 35 cm from food.

Table 1. Data Collecting

Distance / Time / Place	Food Class to-	Displacement Plate to-
13.00-17.00 / Canteen of Engineering Telkom University / 35cm Lux Detail : 1.00 p.m – 3.00 p.m (100 – 600) 3.00 p.m – 5.00 p.m (10 – 200)	Class I (1-2-3)	1-3-2
		3-1-2
		2-1-3
		1-2-3
		3-2-1
		2-3-1
	Class II (1-2-4)	1-4-2
		4-1-2
		2-1-4
		4-2-1
		2-4-1
		1-2-4
	Class III (1-2-5)	1-5-2
		1-2-5
		5-2-1
		5-1-2
		2-1-5
		2-5-1
	Class IV (1-2-6)	1-6-2
		1-2-6
		6-2-1
		6-1-2
		2-1-6
		2-6-1
	Class V (1-6-3)	1-3-6
		1-6-3
		3-6-1
		3-1-6
		6-1-3
		6-3-1
	Class VI	1-4-6

Distance / Time / Place	Food Class to-	Displacement Plate to-
	(1-6-4)	1-6-4
		4-6-1
		4-1-6
		6-1-4
		6-4-1
	Class VII (1-6-5)	1-6-5
		1-5-6
		6-5-1
		6-1-5
		5-1-6
		5-6-1

C. Data Retrieval

The test carried out in the detection of food images through the application, and the results of food image output have been identified. Food images that will be tested based on images of food that has been stored in the storage and taking pictures directly using a webcam camera. Food images will be tested based on each food's position on the tray; in one tray, there are 3 different types of food. to facilitate food testing, for example with the following code:

- Rice = 1
- Fried chicken = 2
- Spinach = 3
- Fried Chili Sauce= 4
- Tofu = 5
- Chicken Braised in Coconut Milk= 6

D. Data Partition Testing

Data partition functions to train the system to detect images. The test was carried out on 420 images and divided into partitions 90% of training data (378 images) 10% of test data (42 images), 80% of training data (336 images) 20% of test data (84 images), 70% of training data (294 images) 30% of test data (126 pictures), 60% of training data (252 images) 40% of test data (168 images), and 50% of training data (210 images) 50% of test data (210 images). Tests carried out at 9och epoch and learning rate 0.0002.

Table 2. Data Partition Testing

Testing to-	Testing Data (%)	Training Data (%)	Dimensions of Images	Size of Images
1	10	90	720x1280	<200Kb
2	20	80		
3	30	70		
4	40	60		
5	50	50		

V. IMPLEMENTATION AND TESTING

1. System Implementation

Figure 5. System Implementation

Figure 6. Detection Result

In figure 6. The shows detection result of food image that is Ayam Goreng is fried chicken (id=2), Nasi is rice (id=1), and Tahu Goreng is tofu (id=5)

The following is a display of desktop applications that can recognize the food images. Testing the accuracy of the image data system from the webcam is done to determine the performance of the CNN classification system that has been designed. This test was conducted with total data of 420 food images. The image used is the distance parameters of 35 cm, outdoor light intensity.

2. Testing Scenarios

A. Partition data

In this test, the dataset is divided into several data partitions. The partition has five data partitions.

Table 3. Total Image Partition data

Testing to-	Testing Data (%)	Training Data (%)	Total Image of Training Data	Total Image of Testing Data
1	10	90	378	42
2	20	80	336	84
3	30	70	294	126
4	40	60	252	168
5	50	50	210	210

Table 4. Partition data

Learning rate 0.0002, *Epoch* 9000	
Partition	Accuracy (%)
90% training data 10% testing data	69.01
80% training data 20% testing data	100
70% training data 30% testing data	69.43

978-1-7281-9336-6/20/$31.00 ©2020 IEEE

Learning rate 0.0002, *Epoch* 9000	
Partition	Accuracy (%)
60% training data 40% testing data	99.07
50% training data 50% testing data	86.10

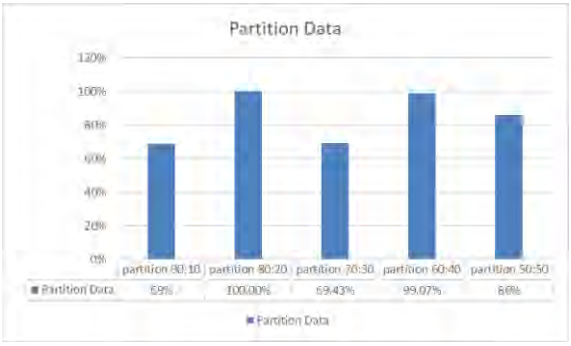

Figure 7. Partition data

B. Epoch and Learning Rate

For having the best accuracy of food detection. CNN must have to test the epoch and learning rate for the best system to recognize the food image.

Epoch is compiling the dataset all through one time forward and backward process through all the neural network nodes 1 time.

Table 5. Epoch

Learning rate 0.0002	
Epoch	Accuracy (%)
1000	71.29
3000	63.88
5000	87.56
7000	78.7
9000	100

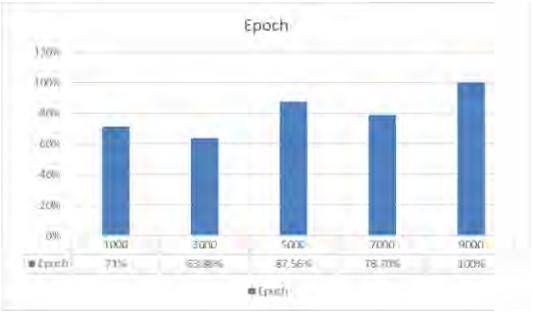

Figure 8. Epoch

The learning rate serves to get the highest accuracy. This is because the greater the value of the learning rate will reduce the value of error/loss and increase the accuracy and accuracy of detection

Table 6. Learning Rate

Epoch 9000	
Learning rate	Accuracy (%)
0.1	0
0.01	94.44
0.001	94.44
0.0001	83.33
0.0002	100

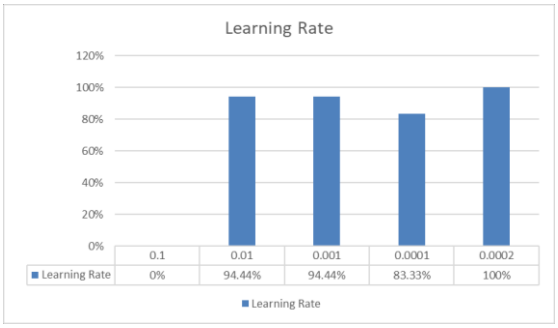

Figure 9. Learning Rate

3. Experiment Result

This study uses 84 images to be tested. There are 12 pictures in class I - VII food. Tests are also carried out with the different position of food in one image.

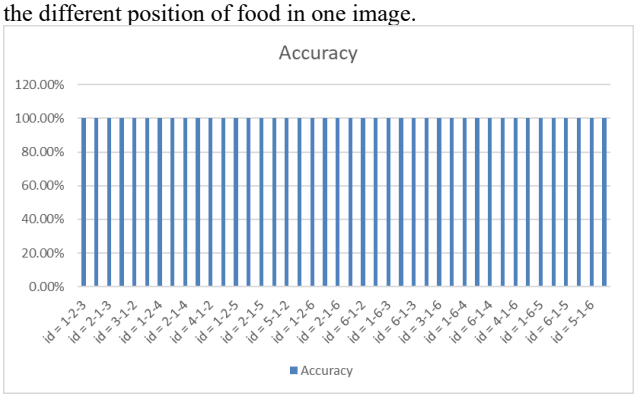

Figure 10. Accuracy

Table 7. Speed of Detection

No	Class	Average Detection Speed (seconds)
1	1-2-3	5.90
2	1-2-4	4.88
3	1-2-5	4.7
4	1-2-6	7.10
5	1-6-3	7.22
6	1-6-4	4.64
7	1-6-5	4.72

4. Confusion Matrix Result

The calculation of the confusion matrix below uses CNN with partition 80% data training and 20% data testing, with epoch 9000 and learning rate 0.0002. The way to calculate the confusion matrix:

978-1-7281-9336-6/20/$31.00 ©2020 IEEE

Table 8. Confusion Matrix Result

		Predicted Class						
		I	II	III	IV	V	VI	VII
Actual Class	I	12	0	0	0	0	0	0
	II	0	12	0	0	0	0	0
	III	1	0	12	0	0	0	0
	IV	0	0	0	12	0	0	0
	V	0	0	0	0	12	0	0
	VI	0	0	0	0	0	12	0
	VII	0	0	0	0	0	0	12

1. Accuracy

Overall accuracy results are based on the following equation:

$$\text{Accuracy} = \frac{TP_I+TP_{II}+TP_{III}+TP_{IV}+TP_V}{N} \quad (1)$$

TP_i is a number of true positives images identified of Class-i, and N is the total of data image in the dataset. The experiment achieves 100% accuracy in this testing data.

2. Precision

Precision result of each class is based on the following equation:

$$\text{Prec}_i = \frac{TP_i}{TP_i+\sum FP_i} \quad (2)$$

With Prec_i is the precision of Class-i, TP_i is a number of true positives images identified of Class-i, and FP_i is a number of false positives images identified of Class-i. The experiments of each class achieve the precision following accuracy:

a. Class I is 100%
b. Class II is 100%
c. Class III is 100%
d. Class IV is 100%
e. Class V is 100%
f. Class VI is 100%
g. Class VII is 100%

VI. CONCLUSION

The food detection system can classify 6 types of food using the Convolutional Neural Network (CNN) algorithm. The best results are obtained on the 80% partition of training data and 20% of test data, getting 100% accuracy and an average speed of fewer than 10 seconds. The food detection system can classify 6 types of food using the Convolutional Neural Network (CNN) algorithm well on existing image data or stored in storage with 80% data partition for training data and 20% for test data and get an accuracy of 100% and can detect food with an average speed of fewer than 10 seconds. Have an epoch or step of 9383 steps with a loss of 0.05 and a learning rate of 0.0002.

VII. REFERENCES

[1] A. G. Subakti, "Analisis Kualitas Pelayanan di Restoran Saung Mirah Bogor," *Jurnal Binus University,* vol. 5, pp. 49-56, 2014.

[2] W. M. W. Handoko and A. R. Widjojo, "Analisis Tingkat Pelayanan Optimal pda Rumah Makan Mie Ayam Mas Yudi Jl. Sagan Kidul No.20 Yogyakarta," *Jurnal Universitas Atma Jaya Yogyakarta,* vol. 25, pp. 73-89, 2013.

[3] V. Ristiawanto, B. Irawan, C. Setianingsih, "Wood Classification with Transfer Learning Method and Bottleneck Features" *2019 International Conference on Information and Communications Technology (ICOIACT),* 2019.

[4] A. Eduardo, R. Beatriz, B. Marc, and R. Petia, "Grab, Pay and Eat: Sematic Food Detection for Smart Restaurants" *IEEE Transactions On Multimedia*, Vol. 1, No. 1, 2020.

[5] K. Hakuto, A. Kiyoharu, O. Makoto, "Food Detection and Recognition Using Convolutional Neural Network" *ACM Multimedia*, 2004.

[6] Kavita, Saroha, Ritika, R. Bala, and M. Siwach, "Review Paper on Overview of Image Processing and Image Segmentation," *International Jurnal of Research in Computer Application and Robotics,* vol. 1, pp. 1-13, 2013.

[7] S. M. A. Mohammed A. Subhi, "A Deep Convolutional Neural Network for Food Detection and Recognition," *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES),* 2018.

[8] T. Ege and K. Yanai, "Estimating Food Calories for Multiple-Dish Food Photos," *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR),* 2017.

[9] B. N. K. S. Md Tohidul Islam, T. Jabid and S. Rahman, "Image Recognition with Deep Learning," *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS),* 2018.

[10] M. T. Turan and E. Erzin, "Detection of Food Intake Events From Throat Microphone Recordings Using Convolutional Neural Networks," *2018 IEEE International Conference on Multimedia & Expo Workshop (ICMEW),* 2018.

[11] J. Wu, "Introduction to Convolutional Neural Networks," *National Key Lab for Novel Software Technology,* 2017.

[12] I. W. S. E.P, A. Y. Wijaya and R. Soelaiman, "Klasifikasi Citra Menggunakan Convolutional Neural Networks (CNN) pada Caltech 101," *Jurnal Teknik ITS,* vol. 5, 2016.

[13] F. H. B. E. Z. Haiqiang, "Combining Convolutional and Recurrent Neural Networks for Human Skin Detection," *IEEE Signal Processing,* vol. 24, 2017.

[14] L. M. Azizah, S. F. Umayah, S. Riyadi, C. Damarjati, and N. A. Utama, "Deep Learning Implementation Using Convolutional Neural Network in Mangosteen Surface Defect Detection," *2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE),* 2017.