PAPER • OPEN ACCESS

Fruit Maturity Classification Using Convolutional Neural Networks Method

To cite this article: Selly Anatya *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **1007** 012149

View the article online for updates and enhancements.

Fruit Maturity Classification Using Convolutional Neural Networks Method

Selly Anatya [1], Viny Christanti Mawardi*[,2], Janson Hendryli [3]

[1,2,3]Informatics Engineering Department, Faculty of Information Technology,
    Universitas Tarumanagara

* viny@untar.ac.id

Abstract. Difficulty in finding information about levels maturity based on the type of fruit using data textual, make a search system using image as query needed. The concept of Content-Based Image Retrieval (CBIR) will search and display images again which are relevant based on the visual features query image. In this study an application was made to classify 5 classes of fruit, Star fruit, Mango, Melon, Banana and Tomato. Which in each class divided again into 52 sub-classes consisting of the type and level of fruit maturity with a total of 5030 training data images. The method used to classify and extract images features is Convolutional Neural Network (CNN). After the image is classified, the search process is carried out to determine the fruit that is similar to the classified image. The results of the classification accuracy of 1294 images are 61%. While the retrieval of 50 images has a precision value of 88.93%.

1. Introduction

Each fruit has unique characteristics from the shape, size, texture, skin color, aroma, taste and nutritional content. The ripening process that occurs in fruit correlates with physical characteristics namely, skin color, shape, size, and texture [1]. Measurement of the level of maturity can generally be done manually using the senses of human vision and special knowledge. But manual measurement requationuires knowledge and other weaknesses, that is, with the visual abilities of each person that is different will result in judgments to be subjective and inconsistent [2]. Example of fruit maturity that can be seen based on fruit visualization can be seen in Figure 1. The maturity level of fruit can be seen based on changes in color or changes in texture.
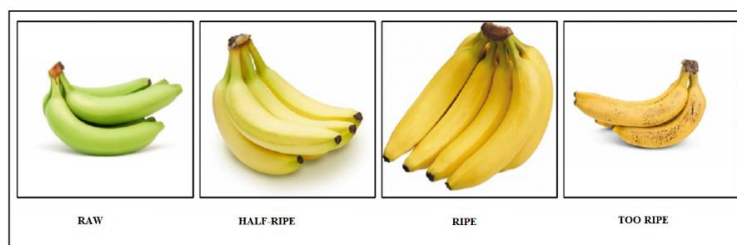

Fig 1. Banana Maturity Level

There are several studies conducted to detect the level of fruit maturity. Nina Sularida et al who succeeded in identifying the maturity level of bananas based on color features [3]. But this research is still limited to identifying one type of fruit and only based on one image feature. Fatma and Ahmed also

conducted research conducted using several methods to classify the four levels of maturity of bananas. One of the methods used in his research is Artificial Neural Network [4]. Zakhayu and Viny have also conducted previous studies to detect animal image classes with CNN and conduct a retrieval process for searching for images of similar animals. This study detected a number of classes and sub-classes of animals which were quite numerous and produced an accuracy of 89.6% [5].

There are many methods for doing the classification that have been developed at this time. One of the image recognition techniques that is growing rapidly now is deep learning. Convolutional Neural Network is one of the methods in deep learning that can be used to carry out the classification process. In 2012, CNN was successfully used to classify 1000 image classes in the ImageNet Large Scale Visual Recognition competition by Alex Kriszhevsky's team [6]. CNN method was chosen because of its ability to detect unique features in images automatically from the results of training on algorithmic neural networks [7].

Content-Based Image Retrieval is one kind of application of the vision of a computer, which is a technique searches image by doing a comparison between the image query and the image in the database that is taken by the information that is contained in the image query the [8]. Content-Based Image Retrieval can also be interpreted as a technique for searching an image that are related and have similar characteristics from a collection of images. The characteristics of this image can be a feature of color, shape, and texture. Sarita et al try using CNN to retrieve image with 3 steps, the extraction of features, calculating the similarity based on feature, the return of the image is relevant in semantics and the acquisition of the image [8].

In this research a system to detect fruit and its level of maturity will be built based on the image from the photo. The recognition process is used for more than one type of fruit with different levels of maturity. We use CNN to detect classes and fruit maturity which are then used in other image search processes that are similar to the images of the fruit. We try not only detect the class of fruit maturity but try to retrieve image that has similar maturity with the test image. The number of fruit classes used in the program is 5 classes consisting of Starfruit, Mango, Melon, Banana, and Tomato. Five classes are divided into 52 sub-classes consisting of 15 types of fruit with 3 or 4 levels of maturity.

2. Convolutional Neural Networks

CNN consists of three layers, namely the Convolution Layer, the Pooling Layer, and the Fully-Connected Layer. From the combination of these layers, an architectural model will be built that can be used to recognize objects and carry out the classification process. At present, several CNN architectures have been made such as LeNet-5 (1998), AlexNet (2012), ZFNet (2013), GoogleNet (2014), VGGNet (2014) and ResNet (2015) [9]. Architecture is the architecture of the best that is used in the race ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The architecture that is used for the CNN model is ResNet50 architecture. ResNet or Residual Neural Network created by Kaiming He, to race ILSVRC in the year 2015 [9].

The design is using ResNet50 because the layer in the network will have a shortcut connection. Shortcut connection is connecting layer directly to the 2 to 3 layers further that contains functions activation nonlinear (ReLU) and Batch Normalization. With things, this can avoid problems gradient and also the process of training. it will be easier. Because it can be used as many layers, so that accuracy is increased.

2.1. Forward Pass

Forward Pass is a process of bringing the input data to the inside of each neuron in the layer to the forward direction [10]. The process is to perform the calculation of the value of the neuron output of the input data. The results of the predicted output can be either classifier or regression.

The beginning in Forward Pass is the data going through the convolution layer. Convolution Layer will produce an activation map of the dot operations performed. After that, the activation is done with the ReLU function and pooling layer to reduce the size of the activation map. Then carried out repeatedly according to the amount of convolution layer.

The last stage in Forward Pass is happening in Fully-Connected Layer. In Fully-Connected Layer data are in vector form is processed with weight and bias values. The results of the process are, then calculated the value of SoftMax and its error values against the train data, to perform classification [5].

- *Convolution Layer*
  The convolution layer consists of a set of convolution kernels or filters. Each volume of the input will do the convolution with filters to extract features from images. The size of the filter is much smaller than the volume of the input.
  In the convolution process, the filter will be shifted to all parts of the input volume. Each shift is performed dot operations between values of the input volume and filter. Filter shift starts from the top right to bottom left, the number of shifts that are done depends on the amount of stride. To determine the size of the dimensions of the output of the results of convolution used equation (1) [11]:

$$Output\ Size = \frac{(W-N)}{S} + 1 \qquad (1)$$

Where W is size of input dimension *(WxW)*, N is size of filter dimension *(NxN), and S* is number of strides.

- *Pooling Layer*
  Pooling Layer is the layer that is used to reduce the dimensions of the map activation, this step is also called as down sampling [12]. This step is done to reduce complexity so that it can speed up computing because the parameters that need to be changed are increasingly reduced.

- *Fully-Connected Layer*
  In the Fully-Connected Layer, there are 3 types of layers, namely the input layer, hidden layer, and output layer. Fully-Connected neurons in the layer are mutually connected to one each other. Each connection has a weight [12]. Neurons in the input layer are vectors of the results of flatten that were carried out previously.
  To produce the output value of neuron, multiplication is performed between the value of the input neuron and weight, the multiplication results are then summed up and added with the bias value. The formula for calculating the output value can be written as equation (2) [12]:

$$y = b + \sum_i x_i \cdot w_i \qquad (2)$$

Where y is value of the output neuron, b is value of bias, x_i is value of the input neuron, and w_i is value of weight.

2.2. Backward Pass

Backward Pass or Backpropagation is the stage of the training network, it will be calculated the value of the error for updating the value of weight. Weight of this will affect the results of the prediction of network neural clone and can reduce the value of the error.

The process of Backward Pass is moving backward which starts from a layer of output to the layer input. Before neuron output will be calculated the value of the error value, the value of this and will be used as a reference in updating the value of weight and bias. Weight and bias values are updated with gradient descent. This process is done by calculating the partial derivative value using the Chain Rule method. The derivative value of the neuron before and after it is then multiplied to get the local gradient value, this value will then be used as a new weight value [5].

- *Categorical Cross Entropy Loss*
  Categorical Cross-Entropy Loss is a function to calculate the value of loss or error in the classification model. Value losses are calculated from the value of neuron output, which results in SoftMax and the value of the target class is the form of one-hot encoding. The Categorical Cross-Entropy Loss is equation (3) [12]:

$$\mathcal{L}(\hat{y}, t) = -\frac{1}{n}\sum_{i=1}^{n}[t_i \cdot \log(\hat{y}_i) + (1 - t_i) \cdot \log(1 - \hat{y}_i)] \quad (3)$$

Where $\mathcal{L}$ is loss function, n is amount of image data, i is index in image data and an index on vector of the target class, t_i is target class in the form of one-hot encoding, and $\hat{y}_i$ is predicted value of SoftMax.

- *Chain Rule*
  Chain Rule is a rule for completing derivative composition functions [12]. To update the weight value, it is necessary to calculate the gradient value of the loss to weight function by finding its partial derivative. So from that used Chain Rule to seek derivative part of the function of these. The Chain Rule formula for the derivative of the total loss to weight is written at equation (4):

$$\frac{\partial Loss}{\partial w_{i,j}} = (\hat{y} - t) \times x_i \quad (4)$$

Where $\hat{y}$ is predicted value of SoftMax, t is target class in the form of one-hot encoding, x_i is value of the input neuron, and $w_{i,j}$ is value of weight to update.

- *Stochastic Gradient Descent*
  Stochastic Gradient Descent is an algorithm for updating weight values [13]. In this algorithm, the old weight will be reduced by a portion of the loss value for each weight. The formula for updating weights can be in equation (5) [8]:

$$w'_{i,j} = w_{i,j} - \alpha \times \frac{\partial Loss}{\partial w_{i,j}} \quad (5)$$

Where $w'_{i,j}$ is new weight value, $w_{i,j}$ is initial weight value, α is learning rate, and $\frac{\partial Loss}{\partial w_{i,j}}$ is chain rule value of total loss to weight.

3. Experiment Result

Classification testing on CNN models that have been trained is done using a confusion matrix and classification report. While retrieval testing is done by using the model of CNN to extract features. In this study we used 2 architectural models as seen at table 1.

Table 1. The Experiment Model

Model	Input Shape	Learning Rate	Total Epochs
Model 1	(300, 300, 3)	0.00001	1000
Model 2	(250, 250, 3)	0.0001	100

From table 2 we can see, model 1 has a higher accuracy value and a lower loss value compared to model 2, namely an accuracy of 93.79% and a loss of 0.201704. When viewed from the architecture the input dimensions in Model 2 are smaller than Model 1, this can cause more information to be unsuccessfully extracted, making Model 2 harder to classify images than Model 1.

Table 2. Formatting sections, subsections and subsubsections.

Model	Total Epochs	Accuracy Value	Loss Value
Model 1	250	87.27%	0.383847
	500	89.39%	0.297934
	1000	93.79%	0.201704
Model 2	100	4.41%	3.8524

Besides, a large number of epochs can also affect the accuracy value and loss in the model in classifying. The more epochs, the more models will learn the characteristics of the image. Thus, the model can increasingly recognize the class of the image.

Then the model used for the program is Model 1 because it has a greater accuracy value and lower loss. In the model of the results of this training, the process will then be tested. Tests carried out by classifying validation data, amounting to 1294 images. After doing the classification, the performance of the model is assessed from the evaluation metrics namely confusion matrix.

3.1. Confusion Matrix Result

The confusion matrix shows the predicted results and the actual class of 1294 data validation classifications. The matrix table shows the prediction results of validation data is quite successful in classifying super classes, this can be seen from the color distribution in the diagonal confusion matrix table 3.

The most validation data successfully predicted correctly were in the "Tomat Beef Setengah Matang" class, which is 46 imagery data. This is due to the training data in the "Tomat Beef Setengah Matang" class, although it consists of a variety of perspectives and backgrounds, but the shape and color are almost the same for each image in the subclass, in addition to the total number of training images there are 150 images, which is the number most compared to the image of training in other Tomato subclasses.

Table 3. Confusion Matrix

	Starfruit	Mango	Melon	Banana	Tomato
Starfruit	277	4	0	5	5
Mango	10	266	12	0	19
Melon	0	0	124	0	0
Banana	1	0	0	260	0
Tomato	15	2	0	2	292

Validation data, which is most often misclassified, is in the "Tomat Cherry Mentah" subclass which is most often predicted as the Belimbing Wuluh Matang. This is due to the similarity of the color and shape features in the two images data. In the "Tomat Cherry Mentah" and "Belimbing Wuluh Matang" subclass, they are oval-shaped and green.

Image Retrieval test results on 50 validation image data based on Top 15 relevant images, obtained an average precision value of 86.65% and a recall value of 27.50%. This shows the CNN model is still good enough to be used as a method of feature extraction of fruit image features. We can see the image retrieval result at table 4.

Table 4. Example of Image Retrieval Result

The top 10 precision value generated from the feature extraction method with the CNN model, which is 100%. This can be due to the test data in the class whose image data has more uniform characteristics, in terms of color, shape or texture. The greatest recall value is owned by "Pisang Cavendish Sangat Matang", which is 53.33%. This is because the image data in the "Pisang Cavendish Sangat Matang" subclass has the shape and color features that are most easily distinguished from other images in its class, besides that, the "Pisang Cavendish Sangat Matang" only has a bit of relevant image data.

The smallest precision value generated from the feature extraction method with the CNN model belongs to the "Tomat Beef Setengah Matang" subclass, which is 53.17%. This insignificant retrieval result can be caused by image data in the class whose image data consists of very diverse perspectives and backgrounds, making it difficult to find relevant images. The smallest recall value is owned by "Pisang Cavendish Mentah", which is equationual to 13.99%. In the "Pisang Cavendish Mentah", the subclass has a small recall value because, the training image data in the class has only 18 images, while the average number of images in the other Banana subclass is 60 images, so the "Pisang Cavendish Mentah" subclass is difficult to recognize by the CNN model.

The results of this Image Retrieval test show images with a dark background will be significantly more retrieval results. This can be caused, because on a dark background the object will be more easily distinguished from the background, compared with a bright background. Also, the pixel value on a dark background is smaller, to minimize the computational process when extracting features.

4. Conclusions
The conclusions obtained after the testing of the application of maturity of fruit based on the type of fruit with a method of CNN is as follows:
1. Training that is done to the Model 1 and Model 2 states that the Model 1 is a model of pitch well which has a value of accuracy of 93.79 % and the value of loss amounted to 0.201704.
2. Model 1, which has been tested by using the validation data as many as 1964 images have a performance that is quite good, which is the value of accuracy of 61%. The model successfully carries out a classification of the class. It is showing a model of CNN suitable to perform the classification of the image that have characteristics typical of the striking.
3. The most difficult subclasses to predict are the "Pisang Ambon Mentah" and "Melon Rock Setengah Matang". "Pisang Ambon Mentah" often predictable as "Pisang Cavendish Mentah", regard it can be caused due to similarity traits in both subclass are of the features of color and also because of the amount of data to train the class "Pisang Ambon Mentah" only amounted to 18 images, while the number of average data on class Bananas as much as 60 images. While for "Melon Rock Setengah Matang" often predictable as "Melon Rock Mentah", regard it can be caused due to the similarity characteristic image of the subclass that of in terms of texture.
4. The test of Model 1 to perform the extraction features of the 50 validations data has a performance that is quite good, which is the value of the average precision of 88.93%. It is showing a model of CNN can do the extraction features on the image of the fruit, although the various background.

5. References
[1] Akbar, Yohanita Maulina, and Rudiati Evi Masithoh. "Aplikasi Analisis Multivariat Berdasarkan Warna Untuk Memprediksi Brix Dan PH Pada Buah-Buahan." PhD diss., [Yogyakarta]: Universitas Gadjah Mada, 2014.
[2] Chan, Andri, Paulus Liem, Ng Poi Wong, and Toni Gunawan. "Segmentasi buah menggunakan metode k-means clustering dan identifikasi kematangannya menggunakan metode perbandingan kadar warna." *Jurnal SIFO Mikroskil* 15, no. 2 (2014): 91-100.
[3] Sari, Jayanti Yusmah, and Ika Purwanti Ningrum Purnama. "Identifikasi Tingkat Kematangan Buah Pisang Menggunakan Metode Ektraksi Ciri Statistik Pada Warna Kulit Buah." *Ultimatics: Jurnal Teknik Informatika* 10, no. 2 (2018): 98-102.

[4] Mazen, Fatma MA, and Ahmed A. Nashat. "Ripeness classification of bananas using an artificial neural network." *Arabian Journal for Science and Engineering* 44, no. 8 (2019): 6901-6910.

[5] Rian, Zakhayu, Viny Christanti, and Janson Hendryli. "Content-Based Image Retrieval using Convolutional Neural Networks." In *2019 IEEE International Conference on Signals and Systems (ICSigSys)*, pp. 1-7. IEEE, 2019.

[6] Demush, Rostyslav. "A Brief History of Computer Vision (and Convolutional Neural Networks)." *Hackernoon, Veljača* (2019). Accessed from: https://hackernoon.com/a-briefhistory-of-computer-vision-and-convolutional-neural-networks-8fe8aacc79f3,   4 September 2019

[7] Dertat, Arden. "Applied deep learning-part 4: Convolutional neural networks." *Towards Data Science. November* 8 (2017): 2017.

[8] Saritha, R. Rani, Varghese Paul, and P. Ganesh Kumar. "Content based image retrieval using deep learning process." *Cluster Computing* 22, no. 2 (2019): 4187-4200.

[9] Das, S. "CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more…." *Medium* November 16 (2017). https://medium.com/@sidereal/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5.(accessed January 10, 2020).

[10] Rumelhart, David E., Bernard Widrow, and Michael A. Lehr. "The basic ideas in neural networks." *Communications of the ACM* 37, no. 3 (1994): 87-93.

[11] Deshpande, Adit. "A beginner's guide to understanding convolutional neural networks, part 2." *URL: https://adeshpande3. github. io/A-Beginner% 27s-Guide-To-Understanding-Convolutional-Neural-Networks-Part-2/(visited on 04/20/2018)* (2017).

[12] Saha, Sumit. "A comprehensive guide to convolutional neural networks—the ELI5 way." *Towards Data Science* 15 (2018).

[13] Rojas, Raul. "The backpropagation algorithm." In *Neural networks*, pp. 149-182. Springer, Berlin, Heidelberg, 1996.