Deep Learning Model for Real-Time Multi-Class Detection on Food Ingredients Using Yolov4 Algorithm

Syafli Shovin[1], Tristyanti Yusnitasari[2], Teddy Oswari[3], Reni Diah Kusumawati[4], Nurasiah[5]
{syaflishovin@gmail.com[1] tyusnita@staff.gunadarma.ac.id[2], toswari@staff.gunadarma.ac.id[3]
reni_dk@staff.gunadarma.ac.id[4], nurasiah@staff.gunadarma.ac.id[5]}

Faculty of Computer Science, Gunadarma University, Jakarta, Indonesia[1,2,5]
Faculty of Economic, Gunadarma University, Jakarta, Indonesia[3,4]

Abstract. The position of food ingredients that may be piled and the similarity in terms of the shape, color, and texture of food ingredients become challenges to build a deep learning model that can perform optimal detection task on food ingredients. Therefore, the food ingredients will be trained and detected using the YOLOv4 algorithm because of its good performance. The data augmentation techniques from YOLOv4 algorithm also applied in this research to improve the variation and amounts of the collected dataset. To make the training more efficient, this research utilizes transfer learning method to adopt knowledge from YOLOv4 pre-trained model. The approach used in this research successfully creates a deep learning model for real-time multi-class detection on food ingredients with a reasonably good performance. The model shows performance with mAP@0.50 value of 84.90% and an average IoU of 72.77%.

Keywords: Deep Learning; Transfer Learning; YOLOv4; Food Ingredients Detection.

1 Introduction

Deep learning is widely applied for various purposes, such as face recognition, object detection, to autonomous vehicles. Various deep learning models have been widely developed in various fields, especially for object detection. One of the popular fields developed with deep learning that aims to perform object detection tasks is its application in the field of food. Deep learning is applied in the food field because of its advantages compared to conventional techniques [7]. The application of deep learning models to perform object detection and predict the name of food ingredients has various purposes, as one of them is conducting food quality screening [6].

Implementing deep learning to perform food detection tasks is a particular challenge. The challenge in detecting food objects depends on several factors, such as the possibility of the position of piled food ingredients, the similarity of food ingredient's shapes, colors, and textures become a challenge in the application of food ingredients detection. It takes an efficient algorithm to produce a deep learning model that can optimally extract the features from food ingredient images to detect food ingredients.

This research will utilize transfer learning techniques to produce deep learning models that can detect food ingredients with good performance. This research uses the architecture of the YOLOv4 model as the pre-trained model. YOLOv4 ability in real-time multi-class object detection is essential with the purpose of this research, which is to do real-time multi-class

BIC 2021, October 11-12, Padang, West Sumatera, Indonesia

DOI 10.4108/eai.11-10-2021.2319569

detection on food ingredients. Another advantage YOLOv4 has is that training on the YOLOv4 model can use a single GPU such as 1080 Ti or 2080 Ti, which means the model is more light-weighted. With these advantages, YOLOv4 can produce models with better speed and accuracy performance than other state-of-the-art detector models [1].

2 Materials and Methods

2.1 Image Dataset

The number of food ingredients classes used in this research includes 11 classes including fruits, vegetables, and animal protein. The 11 classes are banana, cabbage, carrot, cucumber, potato, shrimp, tomato, egg, lemon, broccoli, and orange. The data collected is in the form of an images file with .jpg extension. Datasets are taken from several open-source sites such as OpenImages, Google Image, and iStock. The data retrieved will be sorted back to ensure that the data used follows the research objective. The total number of datasets used is 2750, with each class having 250 images. Using Python-based program, the dataset will then be divided into two parts including training dataset and validation dataset. Training dataset used to train machine learning models that comprise 90% of the total dataset in this research. The rest 10% dataset used for the Validation Dataset to prevent overfitting.

2.2 Labeling and Annotating Dataset

Dataset that has been collected from various sources will be sorted to ensure that the data used is aligned with the purpose of this research. The data that has been sorted will then be given a bounding box & annotation label on the corresponding object. Creating bounding box & labeling annotation is done by using a Python-based open-source program called OpenLabeling. Then, the annotation results will be stored in the form of a file with a .txt extension that uses a file name corresponding to the file name of the photo dataset. The annotations result from the tool follow the YOLOv4 format, so that later the dataset can be trained with YOLOv4.

Fig. 1. Food ingredients was labeled accordingly with their class name

2.3 Data Augmentation

The augmentation data used in this research consists of several techniques adopted from YOLOv4 algorithm, such as photometric distortions and geometric distortions. The pre-trained model used in this research also adopted several other data augmentation techniques that aimed to simulate the presence of objects in image data. Some of the data augmentation techniques applied are CutMix, Mosaic Data Augmentation and Self-Adversarial Training (SAT).

The label for this data is adjusted following the mixture of the data [11]. While Mosaic Data Augmentation is a technique in which four images are combined into a single unit. It aims to train models to recognize the condition of objects beyond existing datasets. SAT is used to add noise when the model training process is carried out to produce a good performance against noise and adversarial attacks [1].

2.4 YOLOv4 Model Architecture

YOLOv4 is currently the fastest and most accurate models for real-time object detection. YOLOv4 adopts the form of a one-stage detector. The architectural selection used by YOLOv4 appears in Fig. 2.

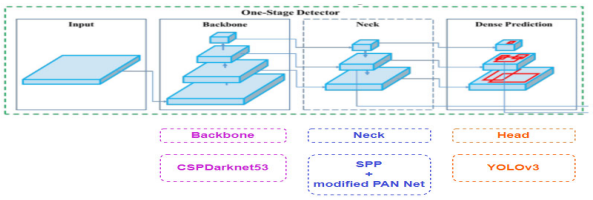

Fig. 2. YOLOv4 model architecture

The YOLOv4 model architecture and its weight will be obtained from Alex Bochkovskiy's GitHub repository using the Darknet repository. The Darknet neural network will be built by utilizing the Graphics Processing Unit (GPU) and the wights from YOLOv4 pre-trained model will be adopted to be used for the purpose of this research.

2.5 Fine Tuning Process

In this part, the model will be fine-tuned by tweaking the parameters to produce a deep learning model that can follow the purpose of this research. The adjusted parameter is being done to set the parameters of data augmentation, batches of the training, input size, learning rate, the filter in convolutional layer and the object class parameter in each YOLO layer. To fine-tune model from YOLOv4 pre-trained models, some parameters were set in Table 1.

Table. 1. Network parameters

Model	Input Size	Batch	Subdivisions	Learning Rate	Momentum	Decay	Filter on each Conv. layer before YOLO layer	Max. Batches
YOLOv4	416 x 416	64	16	0.001	0.949	0.0005	48	22000

2.6 Training Model

Once the model architecture and parameters are adjusted, the next stage is to train the model by utilizing transfer learning method using the knowledge of the pre-trained YOLOv4 model. The training is being done based on deep learning framework from Darknet platform using Google Colaboratory. Google Colaboratory provides a free Graphical Processing Unit (GPU) access, such as NVIDIA Tesla K80.

2.7 Model Evaluation

The evaluation process aims to compare the performance of several weights that have been obtained. After comparing, the weights with the best performance will be chosen. The evaluation process is done to avoid the possibility of using models that produce overfitting performance. The evaluation indicators that are being used is the Mean Average Precision (mAP) and Intersection Over Union (IoU) values.

Intersection Over Union value is ranged between 0 and 1, indicating how true the prediction bounding box (BB) is detected with the ground truth bounding box (BB). In the form of formulas and visualizations, IoU can be shown in Fig. 3.

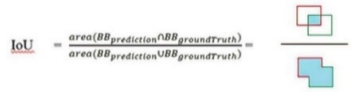

Fig. 3. IoU formulas and visualizations

Where the red box shows the predicted bounding box and the green box represents the ground truth bounding box. The greater the value of the IoU means the less distance there is between the predicted bounding box and the ground truth bounding box [4].

Mean Average Precision or mAP is a metric that is often used to measure the performance of models in object detection. The mAP calculation involves the IoU value to determine whether a prediction result is worth testing as true positive or not. As the name suggests, mAP is the mean average precision number of all classes with corresponding IoU thresholds [10]. mAP is calculated from the Average Precision (AP) that involves the Precision and Recall value. Where the Precision and Recall are defined in equation (1) and (2) respectively,

$$\text{Precision} = \frac{\text{TP}}{\text{FP+TP}} \ . \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN+TP}} \ . \tag{2}$$

From Precision and Recall values, the AP values will be obtained with formula as defined in (3) and can get used in the calculation of mAP in equation (4),

$$AP_k = \int_0^1 P_k(R_k) dR_k \ . \tag{3}$$

$$\text{mAP} = \frac{1}{k} \sum_{i=1}^{k} AP_i \ . \tag{4}$$

Where k represented the value assigned for each food ingredients class in this research, which shown in Table 2.

Table. 2. k value of each food ingredients classes

Class Name	Banana	Cabbage	Carrot	Cucumber	Potato	Shrimp	Tomato	Egg	Lemon	Broccoli	Orange
K value	0	1	2	3	4	5	6	7	8	9	10

3 Results and Discussion

This research successfully creates a deep learning-based model which perform the multi-class detection on food ingredients in real-time using the YOLOv4 algorithm. The model has a reasonably good performance with mAP@0.50 value of 84.90% and IoU value of 72.77%. The data used for the training process is data images of 11 food ingredients classes, including banana, cabbage, carrot, cucumber, potato, shrimp, tomato, egg, lemon, broccoli, and orange.

The stages carried out during the model development process in this research are collecting dataset, annotating the dataset, pre-processing the existing dataset, building darknet, fine-tuning the parameters to create new model architecture, conducting model training, evaluating models, and finally test the model with data outside the training and validation set. The testing process

using primary and secondary data sources. The primary data is from images taken with a smartphone camera and secondary data is from internet sources.

The model is also being tested with live webcam to see its ability to perform real-time detection as shown in Fig. 4.

Fig. 4. Examples of food ingredients detected by the model using YOLOv4 algorithm from live webcam

4 Related Work

Related works has been done for multi-class detection on food ingredients. Previous researchers have a research of Multi-Source Data Fusion Using Deep Learning for Smart Refrigerator which was discussed integrating deep learning models with cameras in a refrigerator to detect the name of the foods, such as fruits and vegetables in the refrigerator. They used the multi-fusion technique, which incorporated weights from several deep learning models, such as SSDs (ResNet), SSDs (VGG16), and SSDs (VGG19) [12]. Another research has comparing YOLOv4 and SSD performance in object detection with the same task, and the result was YOLOv4 model architecture produces better performance than SSDs with the same detection task [9]. The YOLO algorithm also used to detect orange fruits in orange orchards. The research compared different versions of YOLO architecture, namely YOLOv2, YOLOv3, and YOLOv4. As a result, YOLOv4 has the best performance by getting a mAP value of 90.8%. The enormous performance produced by the YOLOv4 algorithm from previous research is the basis for selecting the YOLOv4 algorithm in this research [8].

Various deep learning approaches has been used to classify fruits, namely Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). Integrating various approaches resulted in a new model with better performance than using each approach separately [3]. Another research has been done to produce a deep learning model to classify food ingredients, such as fruits and vegetables. From all previous research mentioned, the research only intended to detect fruits and vegetables as food ingredients [2], [5]. From the mentioned previous research, no one has conducted research to create a deep learning-based model to detect food ingredients, including the fruits, vegetables, and animal protein types that utilize transfer learning techniques using YOLOv4 model architecture.

5 Conclusions and Future Work

This research successfully showed that using YOLOv4 algorithm can produce new deep learning model that can do real-time multi-class detection on food ingredients. The results of this research showed that the model managed to get a mAP@0.50 score of 84.90% and average IoU score of 72.77%. Using the metrics mAP@0.50 means that the model will calculate the true positive value if the IoU threshold value of the bounding box area is at least 0.5. If the IoU value obtained from detection is below the threshold, it cannot be calculated as a true positive. The application of mAP and IoU as evaluation metrics helps to avoid ambiguity in seeing the performance results of object detection models.

For further research, variations of existing food classes with datasets that still adjust to the conditions to be tested can be added. In addition, the research results can be further developed to be integrated with various devices such as smartphones or even used in the agriculture world. According to many needs, further development can also be implemented in smart refrigerators, such as making it easier for users to find food recipes, get nutritional information from food, and others.

References

[1] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. (2020)

[2] Gao, X., Tao, Y., Ding, X., & Hou, R.: Research on food recognition of smart refrigerator based on SSD target detection algorithm. ACM International Conference Proceeding Series, 303–308. (2019)

[3] Gill, H. S., & Khehra, B. S.: An integrated approach using CNN-RNN-LSTM for classification of fruit images. Materials Today: Proceedings. (2021)

[4] Hofesmann E.: IoU a better detection evaluation metric | by Eric Hofesmann | Towards Data Science. (2020)

[5] Li, S., Lü, J., & Ni, S.: Integrated convolutional neural network and its application in fruits and vegetables recognition of intelligent refrigerator. Shuju Caiji Yu Chuli/Journal of Data Acquisition and Processing, 31(1), 205–212 (2016)

[6] Meenu, M., Kurade, C., Neelapu, B. C., Kalra, S., Ramaswamy, H. S., & Yu, YA.: Concise review on food quality assessment using digital image processing. Trends in Food Science and Technology, 118, 106–124. (2021)

[7] Minz, P. S., Sawhney, I. K., & Saini, C. S.: Algorithm for processing high definition images for food colourimetry. Measurement: Journal of the International Measurement Confederation, 158. (2020)

[8] Mirhaji, H., Soleymani, M., Asakereh, A., & Abdanan Mehdizadeh, S.: Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. Computers and Electronics in Agriculture, 191, 106533. (2021)

[9] Morera, Á., Sánchez, Á., Moreno, A. B., Sappa, Á. D., & Vélez, J. F.: Ssd vs. Yolo for detection of outdoor urban advertising panels under multiple variabilities. Sensors (Switzerland), 20(16), 1–23. (2020)

[10] Solawetz J.: What is Mean Average Precision (mAP) in Object Detection? (2020)

[11] Yun, S., Han, D., Chun, S., Oh, S. J., Choe, J., & Yoo, Y.: CutMix: Regularization strategy to train strong classifiers with localizable features. Proceedings of the IEEE International Conference on Computer Vision, 2019-October, 6022–6031. (2019)

[12] Zhang, W., Zhang, Y., Zhai, J., Zhao, D., Xu, L., Zhou, J., Li, Z., & Yang, S.: Multi-source data fusion using deep learning for smart refrigerators. Computers in Industry, 95, 15–21. (2018)