Fruits and Vegetables Detection using the Improved YOLOv3

Changhua Xu[*]
Department of Engineering, Kyushu Institute of Technology
Sensui1-1, Tobata,804-8550, Kitakyushu, Fukuoka, Japan

Ziyue Liu
Department of Engineering, Kyushu Institute of Technology
Sensui1-1, Tobata,804-8550, Kitakyushu, Fukuoka, Japan

Joo Kooi Tan
Faculty of Engineering, Kyushu Institute of Technology
Sensui1-1, Tobata,804-8550, Kitakyushu, Fukuoka, Japan
E-mail: {xu-changhua, liu-ziyue, etheltan }@ss10.cntl.kyutech.ac.jp

Abstract

As the global aging intensifies, it is more convenient for a robot to go for buying things like fruits and vegetables instead of elderly, and it is more human-like to select items according to a user's personal preferences such as maturity of fruits, sweetness, etc. However, Fruits and vegetables are generally displayed in a disorderly manner. Therefore, detection and recognition of fruits and vegetables is a difficult task for a robot. This paper proposes an improved YOLOv3 and also pre-training the networks to detect fruits and vegetables, we then using Bilinear-CNN to classifyfruit's maturity. The effectiveness of the proposed method is shown by experiments.

Keywords: Deep learning, Neural Network, YOLOv3, detection, maturity classification.

1. Introduction

With the development of robotics and the intensification of global aging, it has become a trend to use a robot to help elderly persons in shopping task. While doing shopping, a robot which can choose items based on personal preferences is thought to be more human-like. Under this approach, the detection, recognition and classification abilities have become an important task in the whole shopping process.

About research on target detection algorithms, it began with some classic algorithms which are mainly divided into three stages: (1) Finding a candidate area on an image, (2) extracting features from the candidate area, and (3) using the features for training and then creating a classifier for recognition or classification. Zhang & Wu[1] investigated different multi-class Kernel SVMs with appearance descriptors for fruit classification. It used a combination of color, texture and shape features as descriptors to build model and got an accuracy of 88.2% on their dataset. Rocha *et al*[2] proposed a unified approach that can combine many features and classifiers to achieve a 97% accuracy of classification to the supermarket fruit data set. The conventional algorithms, which is simple though, costs a lot of time for data processing. And, due to artificial feature, the model has a poor generalization

With the development of computer equipment, the calculation of an algorithm employing a huge image database has greatly improved image detection technology not only in an accuracy but also in the terms of process speed. In this way, deep learning was proposed. Series of R-CNN algorithms[3-5] are classic methods for detection, and Faster-RCNN[5] is the Upgraded version of R-CNN algorithm. YOLO[6] was presented in 2016, it improved the speed of detection to 45 f/s at the expense of accuracy. The employed method of combining classification and localization tasks is believed to have a profound impact on solving object detection problems. Owing to the shortcomings of YOLO, the authors[6] have made a series of improvements to the algorithm and provided a 76.8% of mean average precision (mAP) in YOLOv3[7,8]. For fruit classification tasks based on deep learning, Hossain *et al*[9]. proposed two models in which a light CNN model and a VGG-16 fine-tuned model have achieved better performance than the previous studies [2,10].

In this paper, we create a dataset of fruits and vegetables and modify a new network based on YOLOv3 for detection. After that, we use Bilinear-CNN for classifying the maturity of fruits and vegetables, so that robot can choose those maturated fruit or vegetables according to a user's preference.

2. Methods

2.1. Creating *datasets*

Currently, a public training dataset of fruit and vegetable is not available. In this paper, we created two kinds of dataset, *i.e*., fruit and vegetable dataset and their maturity dataset(banana) for detection and classification. As for detection task, we collected 15 kinds of fruits and vegetables pictures from Internet and got totally 1840 pictures. As for classification task, we classified 300 banana pictures into 5 different maturity levels. Examples of 15 classes of fruits and vegetables pictures are shown in Fig. 1.

In addition, in order to optimize the generalization performance of detection process and an anti-interference ability in the real scene, we use gamma correction to change the brightness and saturation, and also add White Gaussian Noise to the pictures (dataset) for data augmentation.

2.2. An *improved network of feature extraction*

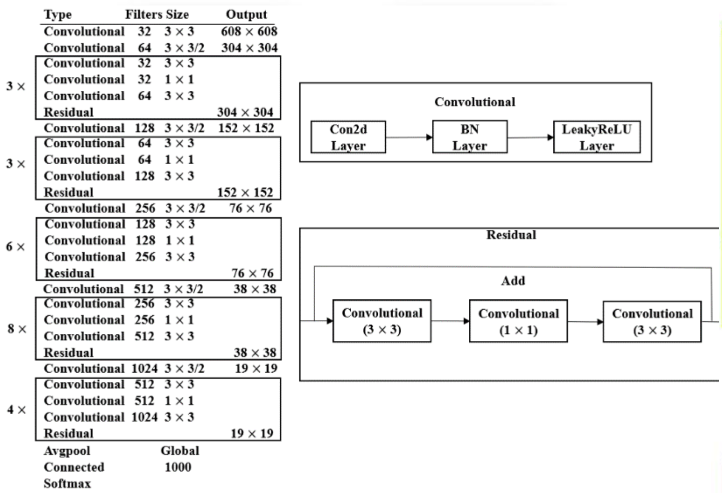
In YOLOv3, Darknet-53 is used for the feature extraction. There are a total of 5 blocks in darknet-53, and each block from top to bottom of the network has 1, 2, 8, 8, and 4 units. Since there are certain differences in the detail of texture within and between fruits or vegetables categories, so we need more shallow feature information to improve the deep network for detection. Instead of 1- 2-8-8-4

Fig. 1. Fifteen classes of fruits and vegetables image examples

unit's model in the original YOLOV3, the unit of each block of the proposed method is optimized to obtain a sequential unit number of 3-3-6-8-4-unit model. On the other hand, for each unit, we use two 3*3 convolutions and one 1*1 convolution to form a residual block. In this way, the proposed method not only increases the depth of the network layer to a certain extent, but also strengthens the expression of detailed features. In order to test the performance of the proposed unit model, we also modified the units to a 4-5-6-8-4-unit model for comparison. **Figure 2** shown the proposed improved network.

2.3. *Maturity classification*

We use Bilinear-CNN to classify maturity of fruit. In the classification process, the maturity feature is extracted using two VGG-16 networks. After detection, in order to eliminate the background effect and to improve the classification accuracy, we trim the targets from the detected images and drop them to the maturity networks for classification. The process is shown in the **Fig.3.**



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	608 × 608
	Convolutional	64	3 × 3/2	304 × 304
3 ×	Convolutional	32	3 × 3	
	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			304 × 304
	Convolutional	128	3 × 3/2	152 × 152
3 ×	Convolutional	64	3 × 3	
	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			152 × 152
	Convolutional	256	3 × 3/2	76 × 76
6 ×	Convolutional	128	3 × 3	
	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			76 × 76
	Convolutional	512	3 × 3/2	38 × 38
8 ×	Convolutional	256	3 × 3	
	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			38 × 38
	Convolutional	1024	3 × 3/2	19 × 19
4 ×	Convolutional	512	3 × 3	
	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			19 × 19
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig. 2. The improved network of darknet-53

3. Experiments

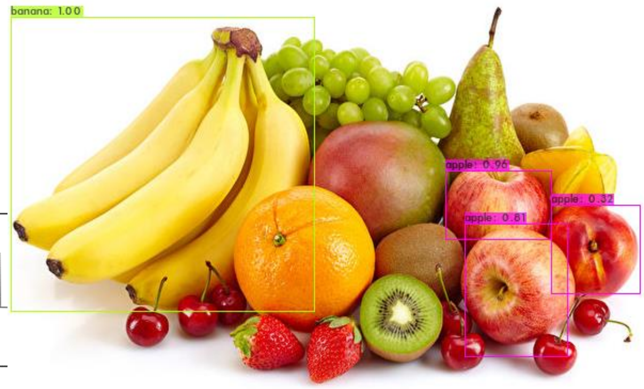
3.1. The performance of different units blocks model

In order to compare the performance of the proposed method with different kind of unit blocks, we used three kind of unit blocks network models in the experiment. Through comparative experiments, we evaluated the detection results of fruits and vegetables by the dataset mentioned in section 2.1. In the experiments, we used 1440 images for training and 400 images for test. For each class, we calculate the Average Precision (AP) to judge the accuracy of different unit model. The experimental results are shown in **Table 1.**

As shown in Table 1, the accuracy of the two proposed units has improved in many classes. especially among small targets such as garlic, potatoes, strawberries and cherry tomato. We also found that the result of 3-3-6-8-4 unit is superior then to 4-5-6-8-4 unit in the most classes.

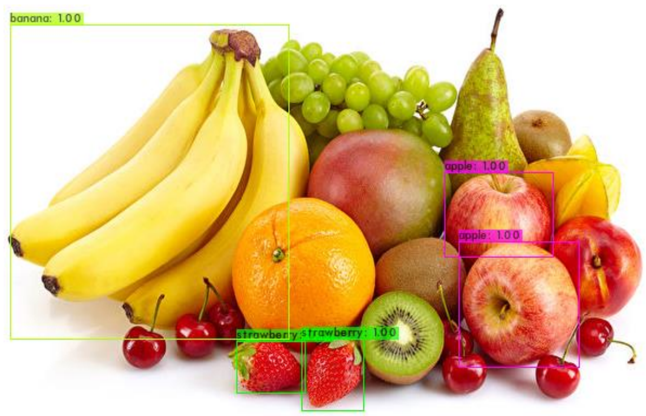
From the comparison of the results, we can get a more detailed effect difference, as shown in Fig.**4**, from which we can see that the network with a unit number of 4-5-6-8-4 eliminated false detections and was able to detect one strawberry. However, the unit number of 3-3-6-8-4 detected not only the banana and apple but also the two small strawberries correctly.

The mAP of the original 1-2-8-8-4unit model, 4-5-6-8-4-unit model come to be 79.04% and 81.41% respectively, the proposed 3-3-6-8-4-unit model gets a better result of 83.72%.



(a) Original 1-2-8-8-4

(b) 4-5-6-8-4



(c) 3-3-6-8-4

Fig. 4. The detection result of three kinds of unit model: (a)Original network, (b) network with a unit number of 4-5-6-8-4, (c) proposed network with a unit number of 3-3-6-8-4.

© The 2022 International Conference on Artificial Life and Robotics (ICAROB2022), January 20 to 23, 2022

Table 1. The comparison of the detection result on our data set

Number	Classes	YOLOV3 1-2-8-8-4 [%]	YOLOV3 4-5-6-8-4 [%]	YOLOV3 3-3-6-8-4 [%]
01	apple	72.3	84.2	85.8
02	banana	76.3	79.4	83.1
03	corn	84.0	86.1	89.3
04	dragon fruit	96.2	96.3	98.5
05	eggplant	91.7	74.7	76.7
06	garlic	77.0	86.1	87.3
07	honeydew	93.0	96.6	96.6
08	onion	65.7	68.8	71.5
09	peach	78.6	65.5	79.3
10	pear	69.7	66.8	67.6
11	bell pepper	77.6	81.3	80.7
12	potato	62.5	74.2	79.2
13	strawberry	73.6	89.5	87.2
14	cherry tomato	71.5	84.0	85.7
15	watermelon	95.8	87.8	87.3

3.2. Result of maturity classification

Table 2. The result of banana maturity classification

Maturity label	Number of pictures	Correct number of pictures	Accuracy [%]
0	47	47	100.0
1	33	28	84.8
2	40	25	62.5
3	30	22	73.3
4	55	53	96.4

In the maturity classification, we only consider "banana maturity". Totally 205 pictures of banana are used in the experiments. **Table 2** shown the result of maturity classification.

We trimmed the targets from the detected images to eliminate the influence of the background and achieved an accuracy of 85.37% with our test dataset. However, maturity label 2 did not get a good result. This is because some trimmed images have lower brightness, resulting in some misclassification images in the maturity label 2.

4. Conclusion

In this paper, we adopted an effectively improved YOLO algorithm to detect fruits and vegetables mainly by considering the balance of the units in network, and by adding another 3*3 convolution based on YOLOv3 to the residual block to improve the detection accuracy. As the result, we achieved the mAP of 83.72% of detection. For the classification task, we divided the bananas into five categories according to their maturity for training, and trimmed the detection result for classification, finally we obtained the average accuracy of 85.4%.

Elimination of the background noises from the detected images needs to be done. Practical use of the proposed method in more complicated environments also remains for future works.

References

1. Y. Zhang and L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine," Sensors, vol. 12, no. 9, pp. 12489- 12505, Sep. 2012.
2. A. Rocha, D.C. Hauagge, J. Wainer, S. Goldenstein, "Automatic fruit and vegetable classification from images," Computers and Electronics in Agriculture, vol. 70, no. 1, pp. 96-104, Jan. 2010.
3. Girshick R, Donahue J, Darrell T, et al: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2014)
4. Girshick R. Fast R-CNN[J]. Computer Science. (2015)
5. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis & Machine Intelligence. 39(6), 1137(2017)
6. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi: You Only Look Once: Unified, Real-Time Object Detection. 779-788(2016)
7. Joseph Redmon, Ali Farhadi: YOLO9000. Better, Faster, Stronger. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).6517-6525(2017).
8. Redmon J, Farhadi A: YOLOv3: An Incremental Improvement[J]. arXiv.org (2018)
9. M. Shamim Hossain, Muneer Al-Hammadi, and Ghulam Muhammad: Automatic Fruit Classification Using Deep Learning for Industrial Applications. IEEE Transactions on Industrial Informatics.15(2), pp. 1027 – 1034(2019).
10. S. R. Dubey and A. S. Jalal, "Robust approach for fruit and vegetable classification," Procedia Engineering, vol. 38, pp. 3449-3453, 2012.

Authors Introduction

Mr. Changhua Xu

He received his bachelor degree of Vehicle Engineering, China in 2018, now specializing in Control Engineering, Kyushu Institute of Technology, Japan. At present, his area of research is the development of human-machine collaboration system based on camera images.

Mr. Ziyue Liu

He received his B.E. degree in Mechanical Design, Manufacturing and Automation in 2014 from the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China. He obtained the M.E. degree in Department of Control Engineering in 2020 from the Graduate School of Engineering, Kyushu Institute of Technology in Japan. He is currently a Ph.D. candidate in the same University. His research interests are image processing, pattern recognition, and robotics.

Prof. Dr. Joo Kooi Tan

She is currently with Department of Mechanical and Control Engineering, Kyushu Institute of Technology, as Professor Her current research interests include ego-motion, three-dimensional shape and motion recovery, human detection and its motion analysis from videos. She was awarded SICE Kyushu Branch Young Author's in 1999, the AROB Young Author's Award in 2004, Young Author's Award from IPSJ of Kyushu Branch in 2004 and BMFSA Best Paper Award in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, The Institute of Electronics, Information and Communication Engineers, and The Biomedical Fuzzy Systems Association of Japan.