Irwin Mango Grade Classification Competition

SHUN-TA WANG

National Taiwan Normal University. Department of Computer Science and Information Engineering

Compiled June 28, 2020

Irwin Mango is an important export agricultural product of Taiwan. Facing the price-cutting competition of other countries, post-harvest processing technology is a job that can enhance the value of products through technology. In the past, after harvesting, Irwin mangoes were manually screened into three grades of A, B, and C, namely for export, domestic sales, and processing. However, in recent years, the shortage of manpower due to rural population loss, and the process of sieving mangoes has also been shortened due to the shortening of the fresh-keeping period, resulting in an error of about 10% in the screening results. Therefore, this study proposed a modified form of stacking ensemble of convolutional neural network which includes AlexNet, ResNet18, VGGNet16, HRNet, and MobileNetV2 to achieve fast and accurate automated mango grades classification.

https://github.com/serendipity109

1. INTRODUCTION

Irwin Mango, one of the most important agricultural products exported from Taiwan, has continued to grow in recent years. Not only has it jumped into one of the three major export high-fresh fruit products, but it has also expanded its export countries to Japan, China, the United States and Hong Kong.

Although, with the cooperation of local governments in various countries, Taiwan's mango has increased its popularity and expanded its market share compared to the past, it still encounters price-cutting competition from other mango exporting countries (Philippines, Thailand). Therefore, many varieties have been improved and harvested. Processing technology and brand marketing and other work to enhance the value of products still need to be promoted by technological assistance.

One of the most urgent improvements is post-harvest processing technology. After harvesting, Irwin mangoes are screened into three grades of A, B and C according to quality, namely for export, domestic sales and processing. However, Irwin mango relies on manual screening. In addition to the shortage of human resources in the rural population, the screening process is also very short due to the shortening of the fresh-keeping period, which leads to an error of about 10% in the screening process. Estimated by the amount of export, 16 million NTD may be lost every year.

In the AI CUP 2020 Irwin Mango Grade Classification Competition, participants need to predict the grade of the input mango images. The strategy of the authors' team is fusion net on either feature-level or decision-level concatenate by five pre-trained models: ResNeXt (ResNeXt-50 32×4d)[1], AlexNet[2], VGG16[3], DenseNet (Densenet-121)[4] and ShuffleNet (ShuffleNetV2 with 1.0x output channels)[5]. This paper proposed a modified form of stacking ensemble of convolutional neural network which includes AlexNet[2], ResNet18[6], VGG16[3], HRNet[7], and MobileNetV2[8], with some specific image pre-processing by YOLOv4[9], data augmentations by RandAugment[10], loss function by FocalLoss[11].

2. RELATED WORK

Image pre-processing YOLO[9] is one of the best-known one-stage object detectors. One-stage detectors divide the image into a grid and predict bounding boxes and their classes inside them. In our study, we use pre-trained YOLOv4 to detect the bounding boxes of all the mangoes in each image. After detecting bounding boxes, we limit the long side to width and short side to height by rotating images.

Data augmentation RandAugment[10] is an effective and convenient way for augmentation combination, which is useful in image classification tasks. For each augmenting image, RandAugment only needs two hyper-parameters N (number of augmentation methods) and M (magnitude for all augmentation methods) to decide the augmentation method.

Feature extraction In our study, we use five different CNN(Convolutional Neural Network) models to extract a variety of different features, which are AlexNet[2], ResNet18[6], VGG16[3], HRNet[7], and MobileNetV2[8]. Among them, AlexNet, MobileNetV2 and ResNet18 have relatively simple architecture, which is more suitable for dealing with simple classification tasks. VGG16 has relatively deeper number of layers, making it easier to extract deeper features. HRNet puts more emphasis on high-resolution feature maps.

3. METHOD

A. Image pre-processing

Figure 1 shows an example of mango pre-processing. In the beginning, we crop the image by the bounding box detected from YOLOv4 [9], and then we limit the long side to width and short side to height by image rotation.

B. Image augmentation

For image augmentation, first we resize the images to different sizes according to different models. Figure shows in 2 . The sizes are determined by the average size of testing data, or half of it. In order to reduce the loss of information caused by interpolating. After that we use RandAugment [10] method, given hyper-parameters $N = 2$ (two augmentation method for each image) and $M = 14$ (magnitude = 14) to achieve data augmentation.

C. Stacking ensemble

After training each model, we do ensemble method for the output of each model, i.e. we do linear combination for each probability predicted by each model, training their weights and bias.

D. Loss function

We use FocalLoss[11] as our loss function, setting two hyper-parameters γ to 3 and α to [800/243, 800/293, 800/263] respectively. γ means the degree of ignoring easy samples, α means the weight of each class, which we usually use the reciprocal of their proportion of each class.

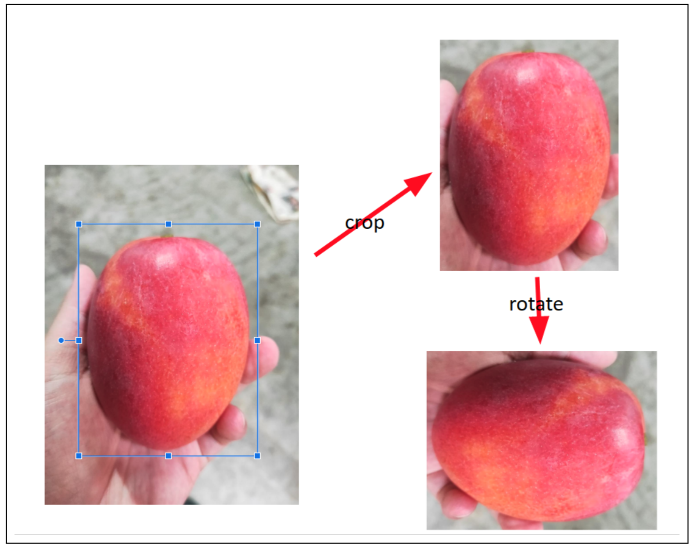

Fig. 1. Image pre-processing.

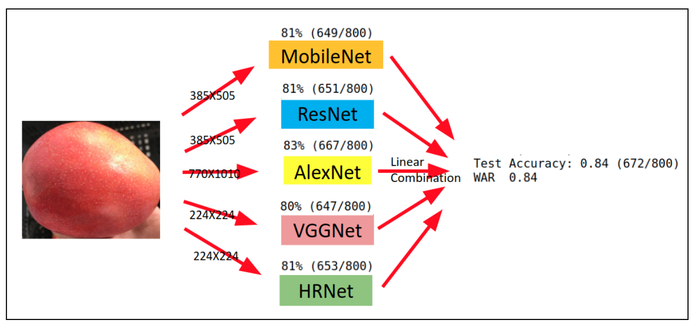

Fig. 2. Training diagram.

4. EXPERIMENTS

Our evaluation table 1 shows the comparison between our method and the baseline proposed by the authors' team. Our method outperforms the baseline by up to 13.75% in accuracy. In terms of each model, the outcome shows that simple model is more suitable for this task.

5. CONCLUSION

Irwin Mango is one of the most important agricultural products exported from Taiwan. Image classification by CNN method is going to replace the way

Table 1. Evaluation table

model	Accuracy	Proportion
$A+M+R+H+V$	0.84	$(672/800)$
$AlexNet$	0.83	$(667/800)$
$MobileNet$	0.81	$(649/800)$
$ResNet$	0.81	$(651/800)$
$HRNet$	0.81	$(653/800)$
$VGGNet$	0.80	$(647/800)$
$baseline$	0.7025	

of screening process, reducing the error rate. In this work we showed that an stacking ensemble method, with a state-of-the-art RandAugment method to do augmentation, improving the accuracy.

REFERENCES

1. S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* (2017), pp. 5987–5995.
2. A. Krizhevsky, "One weird trick for parallelizing convolutional neural networks," (2014).
3. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," (2014).
4. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2017 IEEE Conf. on Comput. Vis. Pattern Recognit. (CVPR) (2017).
5. N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," Lect. Notes Comput. Sci. p. 122–138 (2018).
6. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conf. on Comput. Vis. Pattern Recognit. (CVPR) (2016).
7. J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, and et al., "Deep high-resolution representation learning for visual recognition," IEEE Transactions on Pattern Analysis Mach. Intell. p. 1–1 (2020).
8. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2018 IEEE/CVF Conf. on Comput. Vis. Pattern Recognit. (2018).
9. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," (2020).
10. E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," (2019).
11. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," 2017 IEEE Int. Conf. on Comput. Vis. (ICCV) (2017).