Visual detection of eggs based on deep learning for egg picking robot

Zhihong Xu*, Mingyang Fan and Yapeng Zhang

College of ISE, Northeastern University, Shenyang, Liaoning, China

*Corresponding author. Email: 1624920617@qq.com

Abstract. The chicken farm is a typical labor-intensive production environment, and the mechanization of egg picking work is one of the development directions of the chicken industry. This article uses a camera as a sensor for visual detection. Given the limited computing resources of the robot, we improve the feature extraction part of Mask R-CNN network to reduce the memory loss of parameters and speed up the detection process. The experimental results show that compared with the classic method and the Mask R-CNN basic algorithm, the method in this paper has a higher recognition rate that can better support egg picking robots in egg recognition and pose estimation.

1. Introduction

The poultry and egg industry occupies a significant position in the national economy and is one of the pillar industries of agriculture in the world. Small bionic egg picking systems have a bright future in small and medium-sized farms.

For a long time, the problem of target detection has an essential position in the field of computer vision. Fu et al. [1] proposed a target matching and positioning algorithm that comprehensively applied SIFT features, Mahalanobis distance and radiation transformation methods, and applied it to the robot's grasping and positioning. This method uses stereo matching information and has relative low computational efficiency. Zhang et al. [2] combined gPb algorithm and Otsu to extract the saliency contour of the contour. Peng et al. [3] combined visual saliency with a target recognition method based on dynamic template matching, and proposed a target recognition algorithm for mobile robots in unknown environments. Krizhevsky et al. [4] used deep convolutional neural networks (Deep CNN) for target detection. Since then, researchers have respectively proposed algorithms such as R-CNN, Fast R-CNN, Mask R-CNN, YOLO series and SSD. Zhanget al. [5] applied soft-NMS to improve the R-CNN algorithm, and realized the detection of three types of ships: cargo ships, cruise ships, and yachts. Zhang Qi et al. [6] proposed an improved Fast-RCNN vehicle target detection method, which uses binocular vision to measure the distance of the vehicle, and solves the problem of low efficiency, poor accuracy and single detection algorithm caused by different characteristics of vehicles in different spatial scales and difficult to accurately obtain vehicle distance information. Ou et al. [7] used an improved deep learning algorithm Mask R-CNN to identify and segment the target on the RGB image, and used the Kinect depth sensor model to convert the two-dimensional image coordinates into three-dimensional space coordinates to perform the target object Three-dimensional modeling achieves the purpose of spatial positioning.
Taking into account that the processor onboard the robot is generally an embedded system and the computing resources are limited, this paper improves the feature extraction part on the basis of the Mask R-CNN algorithm to achieve a good egg detection effect.

2. Visual detection algorithm of eggs

2.1. Egg visual detection algorithm based on Mask R-CNN

The Mask R-CNN network can complete object detection and pose estimation at the same time, and has the advantages of small calculation amount, fast speed and high accuracy. It is suitable for real-time eggs detection. The algorithm flow is shown in Fig. 1.

![Mask R-CNN algorithm flowchart.](image)

Mask R-CNN [8] is developed on the basis of Faster R-CNN [9], adding RoIAlign and Fully Convolutional Network (FCN [10]). The output mask depends on the class label predicted by the classification branch, which separates the classification and mask generation.

2.2. Improvement of mask R-CNN

The general Mask R-CNN uses ResNet101 as a feature extraction network. Considering that egg detection is similar to a binary classification problem, the network can be improved and simplified to reduce the memory loss of parameters and speed up the detection speed. The main part of the typical ResNet101 network is shown in Fig. 2(a), in which the block1 in stage2 to stage5 is to reduce the length and width of each channel and increase the number of channels, and the block2 in stage4 is to deepen the number of network layers so as to extract more features. The simplified ResNet network backbone is shown in Fig. 2(b). The stage2 to stage5 of block1 in ResNet101 is replaced with a convolutional layer to achieve the same function, and the 22-layer block2 in stage4 is reduced to 12 layers.

An example of egg detection in the improved Mask R-CNN network is shown in Fig. 3. We took 1200 pictures of eggs varied in numbers in the diverse background environments to make sure the eggs appear in different poses. Each picture contains 1 to 6 eggs. The size of the pictures exceeds 3MB. The pictures are uniformly resized to 1024×1024. 1,000 pictures are used to train the neural network, and 200 pictures are used as the test set.
2.3. Transformation and template matching based on distance detection
The algorithm flow of egg detection based on distance transformation and template matching is shown in Fig. 4. For the better observation, pictures have been enhanced.
3. Experiment and analysis

3.1. Experimental environment configuration
The training of the neural network is conducted on the computer using GPU. The GPU model is NVIDIA GeForce GTX1050, and the computer system is Microsoft Windows 10, the Keras version is 2.1.5, and the deep learning framework Tensorflow 1.13.2 is used to build the required network. Numpy version is 1.17.4 and the programming language is python3.6. The mask used in the experiment was made by Labelme, as shown in Fig. 3.

3.2. Analysis of experiment results
The data set is used in the detection method based on template matching, the trained Mask R-CNN network and the improved Mask R-CNN network. We obtained the egg recognition accuracy, false recognition rate, and average time consumption of the three methods respectively. The experiment data results are shown in Table 1.

|                     | Accuracy  | False Detection Rate | Detection Speed |
|---------------------|-----------|----------------------|-----------------|
| Template Matching   | 78.17%    | 3.33%                | 4.20fps         |
| Mask R-CNN          | 92.51%    | 2.15%                | 0.51fps         |
| Improved Mask R-CNN | 94.18%    | 1.82%                | 0.76fps         |

After 200 times experimental tests, we averaged the results. Under the experimental conditions, the template matching algorithm has a higher detection speed advantage. The improved Mask R-CNN can accurately segment the entire egg from the environment. In a complex environment, the improved Mask R-CNN network has a better effect on egg recognition and pose estimation. We compared the detection results partially shown in Fig. 5.

4. Conclusion
This paper proposes an improved Mask R-CNN network for the goal of egg detection, and uses the improved Mask R-CNN, general Mask R-CNN and template matching methods to conduct experiments. Experimental results show that although the template matching algorithm has greater advantages in detection speed, the false detection rate of egg detection is relatively high. In contrast,
Mask R-CNN can achieve higher accuracy. Compared with the general Mask R-CNN network, the improved Mask R-CNN network has a greater improvement in accuracy and a faster detection speed, which can complete accurate recognition of eggs in complex environments.

References

[1] FU W P, QIN CH, LIU J, et al. Matching and Location of Image Object Based on SIFT Algorithm[J]. Chinese Journal of Scientific Instrument, 2011, 32 (1) : 163-169.

[2] Zhang Guimei, Zhang Song, Chu Jun. A new target detection method based on local contour features[J]. Acta Automatica Sinica, 2014, 40(10): 2346-2355.

[3] Peng Yuqing, Li Mu, Gao Qingqing, Zhang Yuanyuan. Mobile robot target recognition based on dynamic template matching[J]. Journal of Transducer Technology, 2016, 29(01): 58-63.

[4] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks[J]. Proceedings of the Neural Information Processing Systems, Lake Tahoe, USA, 2012:1097-1105.

[5] Zhang Xin, Guo Fuliang, Liang Yingjie, Chen Xiuliang. Detection and recognition of marine vessels based on R-CNN algorithm[J]. Application Research of Computers, 2020, 37(S1): 314-315+319.

[6] Zhang Qi, Hu Guangdi, Li Yusheng, Zhao Xin. Improved Fast-RCNN based binocular vision vehicle detection method[J]. Applied Optics, 2018, 39(06): 832-838.

[7] Ou Pan, Lu Kui, Zhang Zheng, Liu Zeyang. Target recognition and spatial positioning based on Mask RCNN[J]. Computer Measurement and Control, 2019, 27(06):172-176.

[8] He K, Gkioxari G, Piotr Dollár, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017.

[9] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[10] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651.

[11] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.