CHERRY SEGMENTATION AND IDENTIFICATION BASED ON DeepLabV3 IN COMPLEX ORCHARD ENVIRONMENT
/
基于DeepLabV3 的复杂果园环境下樱桃分割与识别

Jinlong WU*[1,2], Ronghui MIAO[1,2]
[1] College of information science and engineering, Shanxi agricultural university, Taigu / China;
[2] College of agricultural engineering, Shanxi agricultural university, Taigu / China
Tel: +86 18503482797; E-mail: wujinlong8192@163.com
DOI: https://doi.org/10.35633/inmateh-72-61

Keywords: *DeepLabV3, cherry segmentation, complex orchard environment, residual network (ResNet), atrous convolution*

ABSTRACT
Aiming at the problems of less research on cherry segmentation and identification, with slow recognition speed and low classification accuracy in agricultural products, a method based on DeepLabV3 was proposed to realize the rapid segmentation and identification of cherry in complex orchard environment. Complex environment mainly includes front lighting, back lighting, cloudy and rainy days, single fruit, multi fruit, fruit overlap, and branch and leaf occlusion. This model proposed the Atrous Spatial Pyramid Pooling (ASPP) module to effectively extract multi-scale contextual information, and solved the problem of target segmentation at multiple scales. The obtained data was divided into training, validation and testing sets in 7:1:2 ratios, and the residual network 50 (ResNet50) was selected as backbone of the DeepLabV3. Experimental results show that the algorithm in this paper can segment cherry quickly and accurately, the mean intersection over union (MIoU) was 91.06%, the mean pixel accuracy (MPA) was 93.05%, and the kappa coefficient was 0.89, which was better than fully convolutional networks (FCN), SegNet, DeepLabV1 and DeepLabV2. It is demonstrated that this study can provide technical support for intelligent segmentation of agricultural products.

摘要
针对当前樱桃分割与识别研究较少，农产品分割与识别速度慢、分类精度低等问题，本文提出一种基于 DeepLabV3 模型的复杂果园环境下樱桃目标的快速分割与识别方法。DeepLabV3 模型提出的孔洞空间金字塔池化模块可有效地提取多尺度语境信息，解决多尺度下的目标分割难题。复杂果园环境主要包括顺光、逆光、阴雨天气、单果、多果、果实重叠和枝叶遮挡等情况。本研究选取 ResNet50 作为该模型的骨干网络，将获取的图像数据按照 7:1:2 的比例划分成训练集,验证集和测试集。实验结果表明，本文提出的方法可以对复杂果园背景下的樱桃进行快速准确分割，分割的 MIoU 值为 91.06%，MPA 值为 93.05%，kappa 系数为 0.89，均优于 FCN、SegNet、DeepLabV1 和 DeepLabV2 方法，该方法能够为农作物智能分割与识别提供技术支持。

INTRODUCTION
Cherry contains various essential amino acids for human body, it has high medicinal value and can enhance immunity. Hence, utilizing deep learning to realize the segmentation and identification for cherry is a pivotal step in the industrialization process of cherry *(Shuvo et al., 2021)*. In recent years, deep learning has been widely applied in computer vision task, focusing on solving the problems of segmentation *(Yang et al., 2019)*, detection *(Wang and He, 2021)*, recognition *(Liu et al., 2023)*, and target tracking *(Lu et al., 202*1). Among them, image segmentation is the process of understanding images at pixel levels and obtaining target category labels corresponding to each pixel, which has received widespread attention in the field of agriculture.

Recently, several researches on cherry detection and segmentation have been conducted. An image analysis algorithm for the classification of cherry in real time was developed, and histogram analysis was performed on the RGB and HSV colour spaces *(Reyes et al., 2021)*.

[1] *Jinlong Wu, Lect.; Ronghui Miao, Lect.*

[2] *Jinlong Wu, Ph.D.Stud.; Ronghui Miao, Ph.D.Stud.*

A cherry recognition method was proposed based on colour channel transform, in which the RGB colour channel was converted into the M channel in CMYK mode to realize threshold segmentation and edge detection, the experimental results showed that with high efficiency and stability, the algorithm can meet the requirements of cherry fruit recognition *(Yang et al., 2019).* An algorithm based on the improved convolutional neural network (CNN) was proposed to detect the appearance of cherry, and the results demonstrated that the accuracy was 99.4%, which was superior to other methods *(Mohammad et al., 2020).* An instance segmentation model was proposed for automated pruning decisions of sweet cherry, and the results indicated that instance segmentation was a promising approach to make automated pruning decisions in sweet cherry trees trained to the upright fruiting offshoots (UFO) architecture *(Daniel and Manoj, 2023)*. An automated method was proposed for detection of powdery mildew disease in cherry leaf images, the proposed method used an automated strategic removal of background from the image and then extracted the desired diseased portion, and a set of public arXiv e-prints data were used to test the proposed algorithm, the tested algorithm achieved accuracy of 99% *(Gupta et al., 2017)*.

The rapid and accurate segmentation of cherry in complex orchard environment are influenced by several factors: the fruits overlap with each other; the leaves would block the fruits; the lighting environment is complex; the cherry picking robot itself has limited computing resources; the low efficiency in running complex algorithms. These bring difficulties for the rapid and accurate detection of cherry. Manual picking is still the main method at present. Therefore, with scarce agricultural labour and increased picking costs, replacing manual picking with cherry picking robots has important practical significance and broad application prospects.

This paper focuses on cherry segmentation in complex orchard environment, ResNet backbone network combined with DeepLabV3 model was proposed to pre-process and segment the collected cherry images. The training and testing sets of cherry images were established to achieve segmentation and recognition, and the recognition results were analysed and evaluated.

MATERIALS AND METHODS

Data collection

The experimental cherry images were collected in Taigu orchard of Shanxi Province in May 2022. The manual data collection was carried out using an OPPO Reno5 camera model. The shooting distance between the device and the objects was 150 mm-250 mm. A total of 1251 images were collected, with an image resolution of 4032 pixels × 3024 pixels. Figure 1 shows the collected cherry images, which include complex environment such as front lighting, back lighting, cloudy and rainy days, single fruit, multi fruit, fruit overlap, and branch and leaf occlusion. The collected dataset was divided into training, validation and testing sets in 7:1:2 ratios. Seventy percent (876 images) and ten percent (125 images) were randomly selected as the training and validation sets for model training and parameter optimization. The remaining twenty percent (250 images) were used as the testing set to verify the performance of the model.

a. Back lighting+multi fruit *b. Front lighting+multi fruit* *c. Multi fruit+branch and leaf occlusion* *d. Single fruit+No occlusion* *e. Multi fruit+ rainy days*

Fig. 1 - Sample images of cherry in complex orchard environment

Dataset production

In order to ensure that the dataset can meet the PASCALVOC format, it is necessary to annotate the dataset. In this paper, the annotation tool Labelme was used to annotate the cherry images, where cherry was labelled in red, and the background was labelled in black. The annotation results were shown in figure 2.

a. Original image 1 *b. Annotation image 1* *c. Original image 2* *d. Annotation image 2* *e. Original image 3* *f. Annotation image 3*

Fig. 2 - Annotation results of partial cherry images

Cherry segmentation and identification model

This paper constructed a model for segmenting and identifying cherry based on the DeepLabV3 network. The backbone of DeepLabV3 network was ResNet. In order to reduce the parameters and accelerate the training speed of the network, the maximum pooling layer which continuously repeated in the network was removed, and the ASPP after the convolutional layer was added.

ResNet backbone

All neural networks were composed by a combination of convolutional and pooling layers before the ResNet was proposed. Generally, the more convolutional and pooling layers there are, the more comprehensive the image feature information can be obtained, and the better the learning effect of the model. However, in practical experiments, it was found that as the convolutional and pooling layers were stacked, the optimization effect actually deteriorated, and the accuracy of test and training data decreased *(Yu et al., 2022)*. This reason of the phenomenon is that the deepening of the network can lead to gradient disappearance, explosion, and degradation problems *(Zhang et al., 2022)*.

a. Gradient disappearance and explosion

Gradient disappearance: If the error gradient of each layer is less than 1, the deeper the network, the closer the gradient approaches 0 during backpropagation.

Gradient explosion: If the error gradient of each layer is greater than 1, the deeper the network, the larger the gradient value during backpropagation.

b. Degradation problem

As the number of layers increases, the prediction effect actually deteriorates. As shown in figure 3, it is evident that the training and testing performance of 56 layer network is worse than that of 20 layer network.

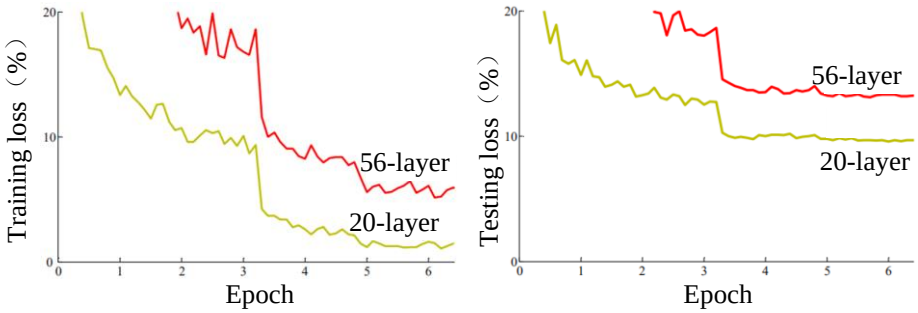


Fig. 3 - Degradation problem

There were already measures such as batch normalization (BN) to alleviate the problems of gradient explosion or disappearance. To solve the degradation problem in deep networks, certain layers of the neural network can be artificially made to skip the connections of neurons in the next layer, connect them in layers, and weaken the strong connections between each layer. In order to achieve good results and reduce degradation, *He et al. (2016)* proposed a new network structure-ResNet, this network won the first place in the classification and target detection tasks in the ImageNet contest, and the error rate was only 3.57%. Compared with the classic models such as AlexNet *(Han et al., 2017)*, VGGNet *(Noubigh et al., 2021)*, and GoogleNet, ResNet possesses prominent image classification performance, its structure can accelerate the training process of the neural networks, and improve the accuracy of the model in image classification.

ResNet is a deep learning network that implements feature extraction, which is a simplified framework for training deep learning models. ResNet has five depth structures (18, 34, 50, 101, and 152), it is composed of several small blocks, which can effectively solve the overfitting problem caused by too many layers. Each layer is composed of multiple blocks, and then form the entire network. The block contains a normal convolutional output layer, a dedicated input-output link branch. The block structure of the ResNet backbone network is shown as figure 4.

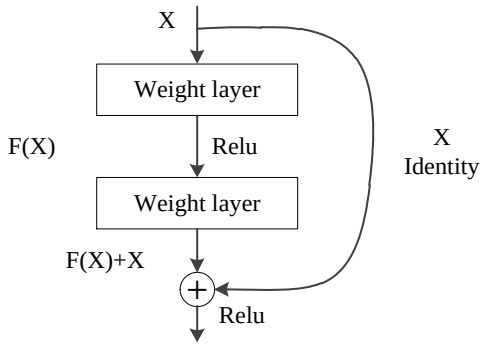


Fig. 4 – Block in ResNet

When the input is X, the learned features are denoted as $H(X)$, then the residual is $F(X)$, in which $H(X)$ is the observed value, and X is the estimated value, residual is the difference between the observed value and the estimated value. Because residual learning is easier than directly learning the original features, then the learning problem of the original features can be solved as $H(X)=F(X)+X$. When the residual is 0, the network only performs identity mapping, and the network performance will not decrease.

DeepLabV3 model

With both serial and parallel atrous convolution modules, DeepLabV3 model can achieve multi-scale object segmentation. It can extract multi-scale contextual information by the cascading and paralleling modules through various atrous rates of atrous convolution effectively *(Chen et al., 2017)*. This model proposes the ASPP module, which improves the ability of extracting global contextual information by the combination hierarchical image information. It can solve the problem of object segmentation at multiple scales, and the structure is shown in figure 5.

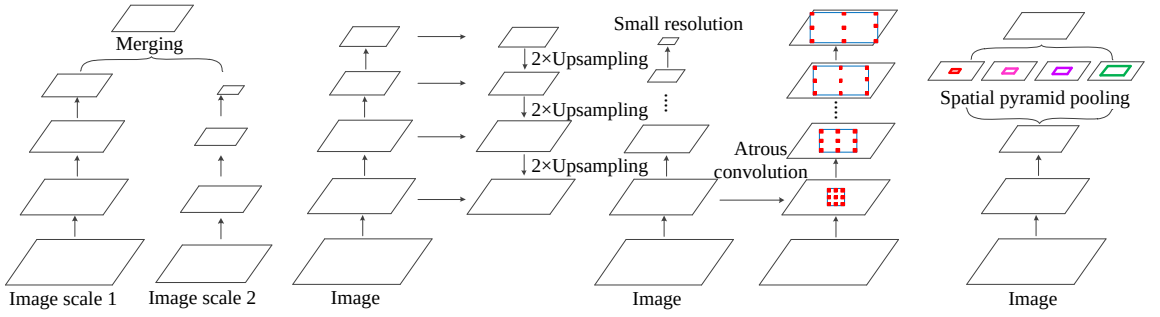


Fig. 5 – DeepLabV3 for capturing multi-scale context

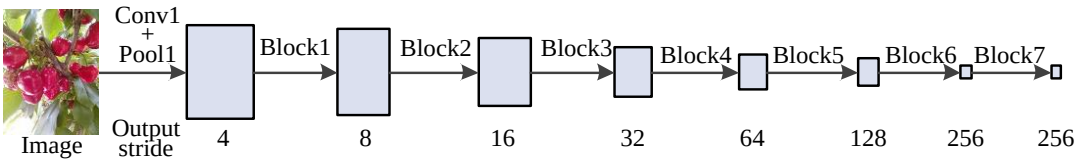
Atrous convolution: In order to solve the problem of the invariance of local image spatial information transformation that hinder dense prediction tasks, atrous convolution is used for pre-training. It mainly removes the last few layers of the dense feature network by down sampling and the corresponding filter kernels for up sampling, which is equivalent to inserting holes between different filter weights to extract more compact features. And this architecture can control the resolution of deep convolutional neural networks (DCNNs) in calculating feature responses without learning new additional parameters.

Atrous space pyramid pooling *(Celik and Talu, 2022)*: This module plays a role in convolutional features to obtain corresponding contextual information through arbitrary scales.

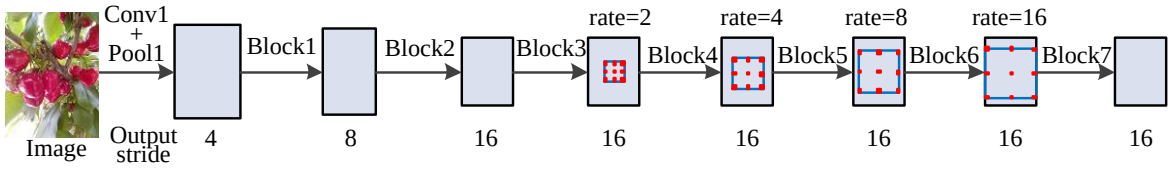
Context module: Typically cascaded behind the model to obtain long-range contextual information, this module exists in DeepLabV1 and conditional random field recurrent neural networks (CRF-RNN).

Encoder and Decoder: Encoders are typically applied to pre trained networks for image classification. Therefore, a decoder is usually used to restore image resolution after encoder. These modules often exist in FCN, SegNet, U-Net *(Park et al., 2021)*, and RefineNet *(Viswanathan and Chu, 2005)*.

DeepLabV3 uses ResNet as the backbone network, which makes it easier to obtain global information in deeper modules. By using the last few blocks to transform them into atrous convolutions, it can maintain the computational complexity and resolution, resulting in denser feature responses and avoiding the loss of detail content in the segmentation network. Specifically, the structures of Block1 to Block4 are directly copied from the original structure of ResNet, and then Block4 is replicated three times resulting in Block 5-7. The difference between them is the use of different expansion rates, as shown in figure 6 and 7.



a. cascading modules with convolution while without deep layers



b. When output stream=16, perform atrous convolution of rate>1 after Block3

Fig. 6 – Cascading modules with inverse convolution

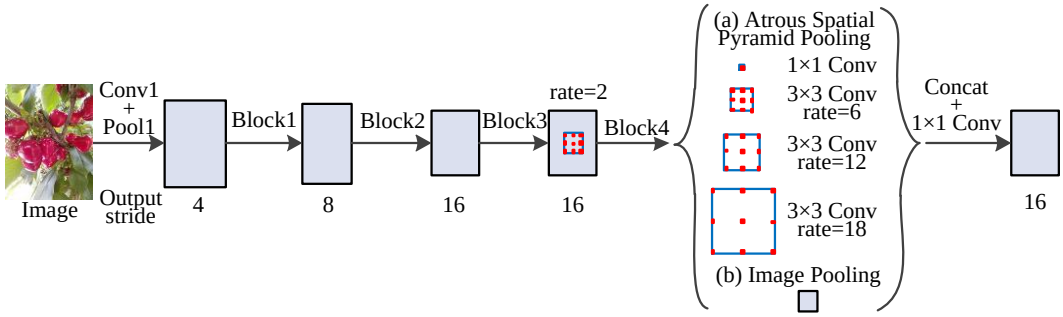


Fig. 7 – ASPP based on image level features

Model evaluation

In this paper, the confusion matrix, mean intersection over union, overall accuracy of classification, user accuracy, producer accuracy and kappa coefficient were used to evaluate the classification results. Confusion matrix is mainly used to compare the classification results with the actual measured values. Here, r is category, X_{ij} represents the percentage of category i judged as the category j by the classifier in the total number of category i; X_{ii} is the number of pixels in row i and column i in the confusion matrix (the number of correct classifications); X_{i+} and X_{+i} are the total number of pixels in row i and column i respectively; N is the total pixels.

MIoU: Inference calculation is performed separately on each category dataset. The calculated intersection of the predicted area and the actual area is divided by the union of the predicted area and the actual area, and the results obtained from all categories are averaged. The value is calculated as formulas (1) and (2).

$$IoU = \frac{TP}{TP + FP + FN} \tag{1}$$

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{Pii}{\sum_{j=0}^{K} Pij + \sum_{j=0}^{K} Pji - Pii} \tag{2}$$

MPA: Its meaning is to calculate the proportion of correctly classified pixels for each class separately and then accumulate to calculate the average, that is, the proportion of correctly predicted pixels in the category to the total pixels (sum of diagonal element values/sum of total element values). The higher the accuracy, the better the quality of the model.

MPA first calculates the pixel accuracy (*PA*) for each class, and then averages the *PA* for all classes, as shown in formulas (3) and (4):

$$PA = \frac{\sum_{i=0}^{K} Pii}{\sum_{i=0}^{K}\sum_{j=0}^{K} Pij} \tag{3}$$

$$MPA = \frac{1}{K+1}\sum^{K}\frac{P_{ii}}{\sum^{K} P_{ij}} \tag{4}$$

Assuming there are $K+1$ classes (from class 0 to class K, which contain an empty class or background), P_{ij} represents the number of pixels that originally belongs to class i but were predicted to be class j, P_{ii} (true positives) represents the true numbers, while P_{ij} and P_{ji} are interpreted as false positives and false negatives respectively.

Overall classification accuracy (*OA*) is equal to the sum of correctly classified pixels divided by the total pixels, as shown in formula (5):

$$OA = \frac{\sum_{i=1}^{r} X_{ii}}{\sum_{i=1}^{r}\sum_{j=1}^{r} X_{ij}} \tag{5}$$

User accuracy (*UA*) indicates the probability that a certain type of sample is correctly classified, as shown in formula (6):

$$UA = \frac{X_{ii}}{X_{i+}} \tag{6}$$

Producer accuracy (*PA*) represents the probability that a certain type of sample in the classification diagram is correctly classified, as shown in formula (7):

$$PA = \frac{X_{ii}}{X_{+i}} \tag{7}$$

Kappa coefficient can make full use of the information of confusion matrix. It can be used as a comprehensive index for the evaluation of classification accuracy. The relationship between classification quality and kappa statistics are: the range of kappa coefficient is 0.0-0.2, the classification quality is “Difference”; the range of kappa coefficient is 0.2-0.4, the classification quality is “Commonly”; the range of kappa coefficient is 0.4-0.6, the classification quality is “Good”; the range of kappa coefficient is 0.6-0.8, the classification quality is “Very Good”; the range of kappa coefficient is 0.8-1.0, the classification quality is “Excellent”. The calculation formula of kappa coefficient is (8):

$$K = \frac{N\sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r}(X_{i+}X_{+i})}{N^2 - \sum_{i=1}^{r}(X_{i+}X_{+i})} \tag{8}$$

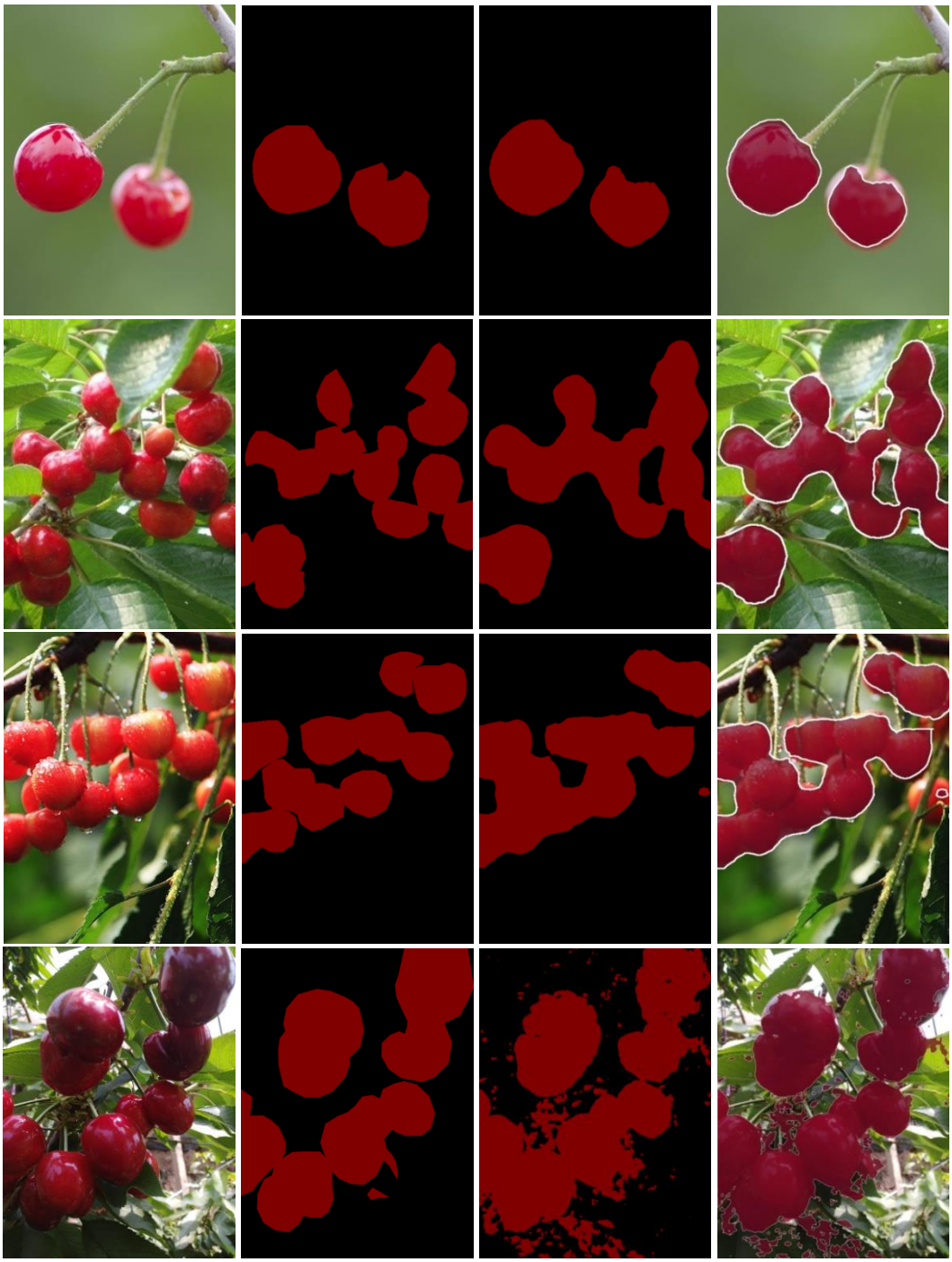
Experimental environment configuration and network parameter settings

The operating system is Windows 10, the CPU model is Intel (R) Core (TM) i7-12700F CPU@2.10GHz, the GPU model is NVIDIA GeForce RTX 3080, the running memory is 32GB, the hard drive is 1T. The programming language is Python 3.9, the deep learning framework is Python 1.13.0, and the GPU acceleration libraries are CUDA11.7 and CUDNN8.4.1.

To shorten the training time of the network, the freezing training method was adopted. When freezing the backbone network for training, the initial learning rate was 0.001, the batch processing amount was 16, and the momentum was 0.9. The network training learning rate after thawing was 0.0001, the batch size was 8, and the iteration epoch was 300.

RESULTS AND DISCUSSION

Figure 8 shows the segmentation and recognition results of cherry. In the segmentation results, 8(a) are the original images, 8(b) are the annotation images, 8(c) are the segmentation results in mask, and 8(d) are the segmentation results on the original images. In 8(b) and 8(c), red represents the cherry region, and black represents the background. The backbone network in the model is ResNet50. From figure 8(c) and 8(d), it can be seen that the DeepLabV3 model can effectively achieve segmentation of cherry in different orchard environment.



a. Original images b. Annotation images c. Segmentation results in mask d. Segmentation results on the original images

Fig. 8 – Segmentation and recognition results of cherry in DeepLabV3

Tables 1 and 2 show the confusion matrices corresponding to the classification prediction results of figure 8 (the first image), which expressed in numerical and percentage forms respectively. The former can clearly express the number of correctly and incorrectly classified pixels for different categories, while the latter can represent the accuracy of classification.

From tables 1 and 2, it can be seen that the total number of rows represents the total number of other pixels classified into that category, the total number of columns represents the true ground truth values for each category, and the diagonal elements represent the correct number of classification for each category. By calculating the sum of diagonal elements, the number of correctly classified objects can be calculated as 56397, with an overall classification accuracy of 92.53%. Table 2 shows the percentage of correct and incorrect classification for each category. It can be seen from the table that the classification accuracies of the cherry and background regions are relatively high, at 91.07% and 93.10% respectively.

Table 1

Confusion matrix of segmentation results (numerical)

Categories	Cherry	Background	Total of row
Cherry	15622	3023	18645
Background	1532	40775	42307
Total of Column	17154	43798	60952

Table 2

Confusion matrix of segmentation results (Percentage)

Categories	Cherry	Background	Total of row
Cherry	91.07	6.90	97.97
Background	8.93	93.10	102.03
Total of Column	100.00	100.00	200.00

The UA and PA of each category can also be calculated by the confusion matrix, table 3 shows the UA and PA corresponding to figure 8 (the first image). It can be seen that the UA and PA of background and cherry are both higher, which can reach over 83%.

Table 3

User accuracy and producer accuracy of classification prediction results for 9 feature vectors

Classification results	UA	UA(%)	PA	PA(%)
Cherry	15622/18645	83.79	15622/17154	91.07
Background	40775/42307	96.38	40775/43798	93.10

Identification results of different methods

Table 4 is the classification results compared with FCN, SegNet, DeepLabV1 and DeepLabV2 of the testing images. The MIoU and MPA of FCN are 78.81% and 76.54% respectively, which are the worst among all methods. Compared with traditional CNN, the input images in FCN can be of any size and the algorithm is more efficient. However, its segmentation results are not precise and not sensitive to the details of the image, at the same time, the relationship between pixels has not been fully considered, there is a lack of spatial consistency *(Wang et al., 2020)*.

The MIoU and MPA of SegNet are 89.28% and 88.45% respectively, which are better than FCN. The main reason is that SegNet utilizes the maximum pooling position index for upsampling, which not only reduces the computational complexity but also better preserves the edge information of the image *(Li et al., 2021)*.

With the upgrade of the version for DeepLab series methods, the segmentation results have also been correspondingly improved. On the basis of DeepLabV1, DeepLabV2 can achieve better performance by using multi-scale processing and ASPP. DeepLabV3 expands ASPP module to better solve segmentation problems at multiple scales *(Kwak and Sung, 2021)*. Among all the methods, the identification results of our method are the highest, the MIoU is 91.06%, the MPA is 93.05%, and kappa coefficient is 0.89, and the overall classification effect is “Good”.

Table 4

Classification results of different methods

Methods	MIoU (%)	MPA (%)	Average Kappa coefficient
FCN	78.81	76.54	0.73
SegNet	89.28	88.45	0.82
DeepLabV1	83.23	81.06	0.83
DeepLabV2	88.75	85.37	0.85
Our method	91.06	93.05	0.89

CONCLUSIONS

In this paper, DeepLabV3 was adopted to realize the rapid segmentation and identification of cherry in complex orchard environment. Compared with FCN, SegNet, DeepLabV1 and DeepLabV2, the identification results of our method are the highest, and the overall classification effect is “Good”. The algorithm in this paper can achieve cherry target segmentation in various complex backgrounds, such as front lighting, back lighting, cloudy and rainy days, single fruit, multi fruit, fruit overlap, and branch and leaf occlusion. It can provide theoretical basis for object detection, segmentation, and localization of fruits and vegetables in complex orchard environments. In image segmentation and recognition, future work can improve the network structure and introduce attention mechanisms to increase the acquisition of image feature information, thus, improving the recognition accuracy and extraction speed of the model.

The method in this paper mainly focuses on the application of intelligent cherry picking robots, and it has not been verified in embedded devices. Generally, deploying image segmentation models on embedded devices requires smaller model size and higher recognition speed. Therefore, further research is needed on the lightweight of the models while ensuring its accuracy and recognition speed, which can solve the problem of weak processing performance in embedded devices. The subsequent research on lightweight of the models will be implemented such as network pruning and efficient network structures.

ACKNOWLEDGEMENT

This research was funded by Shanxi Agricultural University Youth Science and Technology Innovation Fund, grant number 2019023; Shanxi Province Applied Basic Research Youth Project, grant number 202203021212428; Shanxi Province Applied Basic Research Youth Project, grant number 202203021212414.

REFERENCES

[1] Celik, G., & Talu, M. (2022). A new 3D MRI segmentation method based on Generative Adversarial Network and Atrous Convolution. *Biomedical Signal Processing and Control*, 71, 103155. https://doi.org/10.1016/j.bspc.2021.103155

[2] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE transactions on pattern analysis and machine intelligence, 40*(4), 834-848. https://doi.org/10.1109/tpami.2017.2699184

[3] Daniel, B., & Manoj, K. (2023). Automated pruning decisions in dormant sweet cherry canopies using instance segmentation. *Computers and Electronics in Agriculture*, 207, 107716. https://doi.org/10.2139/ssrn.4202299

[4] Gupta, V., Sengar, N., Dutta, M., Travieso, C. & Alonso, J. (2017). Automated segmentation of powdery mildew disease from cherry leaves using image processing. *International Conference and Workshop on Bioinspired Intelligence (IWOBI), 15*(10), 27-37. https://doi.org/10.1109/iwobi.2017.8006454

[5] Han, X., Zhong, Y., Cao, L., & Zhang, L. (2017). Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. *Remote Sensing, 9*(8), 848. https://doi.org/10.3390/rs9080848

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778. https://doi.org/10.1109/cvpr.2016.90

[7] Kwak, J., & Sung, Y. (2021). DeepLabV3-Refiner-Based Semantic Segmentation Model for Dense 3D Point Clouds. *Remote Sensing, 13*(8), 1565. https://doi.org/10.3390/rs13081565

[8] Li, Q., Wang, H., Li, B., Yanghua, T., & Li, J. (2021). IIE-SegNet: Deep semantic segmentation network with enhanced boundary based on image information entropy. *IEEE Access,* 9, 40612-40622. https://doi.org/10.1109/access.2021.3064346

[9] Liu, K., Lin, K., & Zhu, C. (2023). Research on Chinese traditional opera costume recognition based on improved YOLOv5. *Heritage Science*, *11*(1), 1727. https://doi.org/10.1186/s40494-023-00883-x

[10] Lu, P., Ding, Y., & Wang, C. (2021). Multi-small target detection and tracking based on improved YOLO and sift for drones. *International journal of innovative computing, information and control*, *17*(1), 205-224. http://www.ijicic.org/ijicic-170114.pdf

[11] Mohammad, M., Ahmad, J., Khalegh, J., & Zhang, Y. (2020). Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach. Postharvest Biology and Technology, 166, 111204. https://doi.org/10.1016/j.postharvbio.2020.111204

[12] Park, J., Choi, J., Seol, S., Byun, J., & Kim, Y. (2021). A method for adequate selection of training data sets to reconstruct seismic field data using a convolutional U-Net. *Geophysics*, 1-103. https://doi.org/10.1190/geo2019-0708.1

[13] Reyes, J. F., Contreras, E., Correa, C., & Melin, P. (2021). Image analysis of real-time classification of cherry fruit from colour features. *Journal of Agricultural Engineering, 52*(4). https://doi.org/10.4081/jae.2021.1160

[14] Shuvo, M., Ahommed, R., Reza, S., & Hashem, M. (2021). CNL-UNet: A novel lightweight deep learning architecture for multimodal biomedical image segmentation with false output suppression. *Biomedical Signal Processing and Control*, 70, 102959. https://doi.org/10.1016/j.bspc.2021.102959

[15] Viswanathan, N., & Chu, C. (2005). FastPlace: efficient analytical placement using cell shifting, iterative local refinement and a hybrid net model. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *24*(5), 722-733. https://doi.org/10.1145/981066.981072

[16] Wang, D., & He, D. (2021). Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering*, *210*(6), 271-281. https://doi.org/10.1016/j.biosystemseng.2021.08.015

[17] Wang, Z., Xie, L., & Qi, J. (2020). Dynamic pixel-wise weighting-based fully convolutional neural networks for left ventricle segmentation in short-axis MRI. *Magnetic resonance imaging: An International journal of basic research and clinical applications, 66*(1), 131-140. https://doi.org/10.1016/j.mri.2019.08.021

[18] Yang, W., Zhang, J., Xu, Z., & Hu, K. (2019). Real-time DeepLabv3+ for pedestrian segmentation. *Journal of Optical Technology c/c of Opticheskii Zhurna*l, *86*(9), 570. https://doi.org/10.1364/jot.86.000570

[19] Yang, R., Wu, M., Bao, Z., & Zhang, P. (2019). Cherry recognition based on colour channel transform. *The 2019 International Conference on Artificial Intelligence and Computer Science*, 292-296. https://doi.org/10.1145/3349341.3349419

[20] Yu, L., Zeng, Z., Liu, A., Xie, X., Wang, H., Xu, F., & Hong, W. (2022). A Lightweight Complex-Valued DeepLabv3+ for Semantic Segmentation of PolSAR Image. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15. https://doi.org/10.1109/jstars.2021.3140101

[21] Zhang, X., Bian, H., Cai, Y., Zhang, K., & Li, H. (2022). An improved tongue image segmentation algorithm based on Deeplabv3+ framework. *IET Image Processing*, 16. https://doi.org/10.1049/ipr2.12425