

LETTER

Tea Sprouts Segmentation via Improved Deep Convolutional Encoder-Decoder Network

Chunhua QIAN^{†,††a)}, Student Member, Mingyang LI[†], and Yi REN^{††}, Nonmembers

SUMMARY Tea sprouts segmentation via machine vision is the core technology of tea automatic picking. A novel method for Tea Sprouts Segmentation based on improved deep convolutional encoder-decoder Network (TS-SegNet) is proposed in this paper. In order to increase the segmentation accuracy and stability, the improvement is carried out by a contrastive-center loss function and skip connections. Therefore, the intra-class compactness and inter-class separability are comprehensively utilized, and the TS-SegNet can obtain more discriminative tea sprouts features. The experimental results indicate that the proposed method leads to good segmentation results, and the segmented tea sprouts are almost coincident with the ground truth.

key words: tea sprouts segmentation, TS-SegNet, skip connections, contrastive-center loss function

1. Introduction

Tea sprouts segmentation is the core technology of automatic tea picking systems based on machine vision [1]. The tea sprouts region in the image can be automatically gained by the segmentation method. Much research has focused on tea sprouts segmentation or recognition, a color and shape features based segmentation method [2] was utilized to recognize the tea sprouts. Tang et al. [3] introduced a tea sprouts recognition method based on multiple threshold segmentation, but the threshold value calculated by the image histogram suffers the ambient lighting issue. Shao et al. [4] proposed an improved k-means algorithm to recognize the tea sprouts. An image histogram equalization and HSI color model were also employed in [4]. Other research has focused on tea category or tea disease recognition. Wu et al. [5] introduced a tea category identification method based on optimal wavelet entropy and weighted k-Nearest neighbors. But the proposed method [5] only classifies three kinds of tea, there are many kinds of tea available nowadays. Zhang et al. [6] proposed a tea disease recognition method via tea leaf color and texture features, and the Back Propagation (BP) neural network is utilized to recognize the different tea diseases.

Most existed tea sprouts segmentation methods are not deep learning based, the color or shape features are designed by artificial experience, which lack robustness and

represent-tation ability. The segmentation results gained by the existed methods are not high enough to achieve the tea automatic picking. For these issues, a novel method for Tea Sprouts Segmentation based on deep convolutional encoder-decoder Network (TS-SegNet) is proposed in this paper. The network structure of TS-SegNet is shown in Fig. 1, it is an end-to-end method, the input is a tea image, and the output is the segmentation results of tea sprouts. The proposed TS-SegNet is based on the deep learning segmentation architectures SegNet [7]. The SegNet is a well-known architecture in computer vision for semantic segmentation [8], [9], but has not been used much so far for the tea sprouts segmentation task. SegNet utilizes the pooling indices to achieve the upsampling operation, but it does not capture multiscale information. So the skip connections used in U-Net [10] and contrastive-center loss function are employed in the proposed TS-SegNet to capture both coarse level and fine level information at the deconvolutional layers. The experimental results show that the novel method leads to good results, meanwhile, the segmentation of tea sprouts can be widely used in the tea automatic picking.

2. Tea Sprouts Segmentation Methods

2.1 Contrastive-Center Loss Function

The tea sprouts features extracted from the deep neural network are separable but not that discriminative enough, and the inter-class separability is not considered. So, a contrastive-center loss function is employed to overcome these issues. The intra-class compactness and inter-class separability is achieved in Eq. (1):

$$L_c = \frac{1}{2} \sum_{i=1}^m \frac{\|x_i - c_{y_i}\|_2^2}{\left(\sum_{j=1, j \neq y_i}^k \|x_i - c_j\|_2^2\right) + \gamma} \quad (1)$$

where L_c is the contrastive-center loss, m is the training sample number in a patch, x_i is the i th training sample. c_{y_i} is the y_i th class center of extracted features, γ is a constant preventing the denominator equal to 0. In our experiments, we set $\gamma = 1$ by default.

The updates of network parameters and class center simultaneously in Eq. (2), Eq. (3):

$$\frac{\partial L_c}{\partial x_i} = \frac{x_i - c_{y_i}}{\left(\sum_{j=1, j \neq y_i}^k \|x_i - c_j\|_2^2\right) + \gamma}$$

Manuscript received August 8, 2019.

Manuscript revised September 11, 2019.

Manuscript publicized November 6, 2019.

[†]The authors are with Nanjing Forestry University, Nanjing, China.

^{††}The authors are with Suzhou Polytechnic Institute of Agriculture, Suzhou, China.

a) E-mail: borisqiang@sina.com

DOI: 10.1587/transinf.2019EDL8147

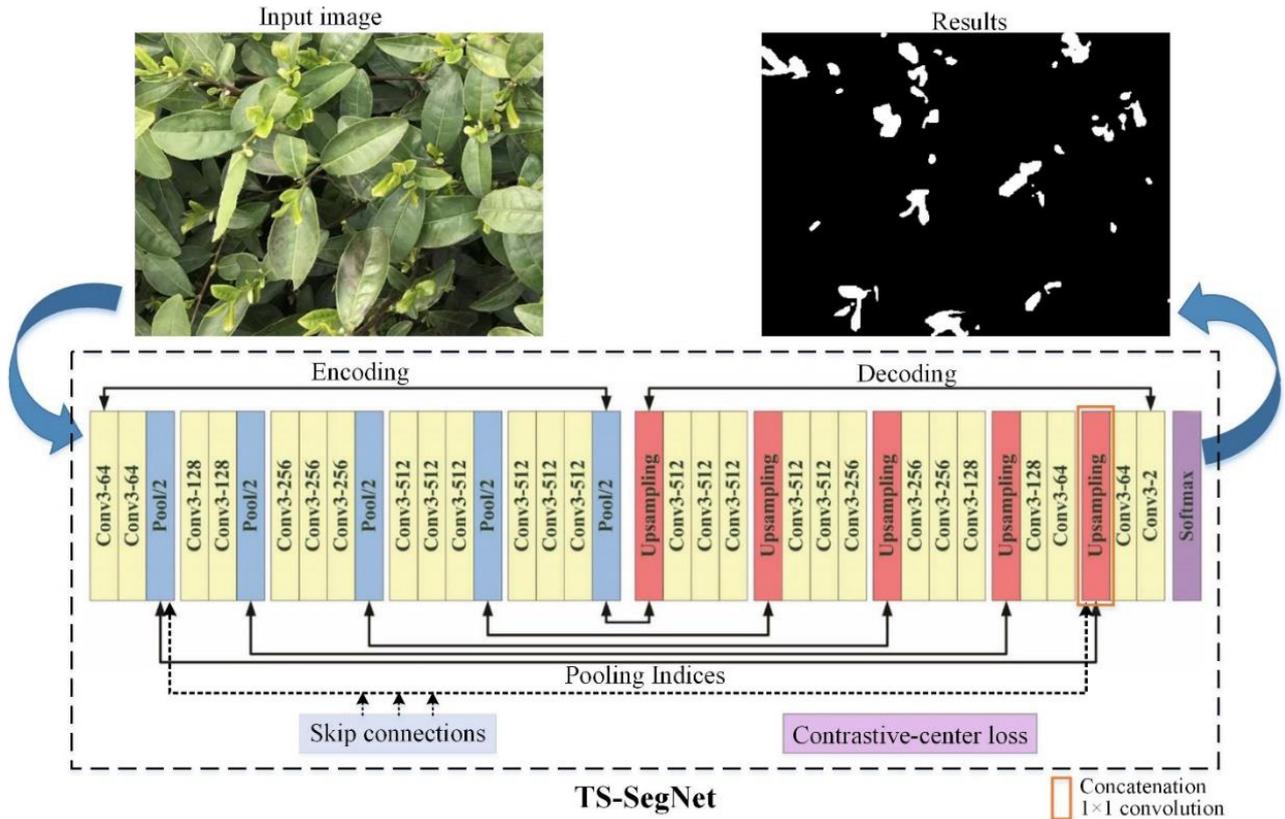


Fig. 1 The network structure of the proposed TS-SegNet

$$\frac{\partial L_c}{\partial c_n} = \sum_{i=1}^m \begin{cases} \frac{\|x_i - c_{y_i}\|_2^2 \sum_{j=1, j \neq y_i}^k (x_i - c_j)}{\left[\left(\sum_{j=1, j \neq y_i}^k \|x_i - c_j\|_2^2 \right) + \gamma \right]^2} & \text{if } y_i = n \\ \frac{c_{y_i} - x_i}{\left(\sum_{j=1, j \neq y_i}^k \|x_i - c_j\|_2^2 \right) + \gamma} & \text{if } y_i \neq n \\ \frac{(x_i - c_n) \|x_i - c_{y_i}\|_2^2}{\left[\left(\sum_{j=1, j \neq y_i}^k \|x_i - c_j\|_2^2 \right) + \gamma \right]^2} & \text{if } y_i \neq n \end{cases} \quad (2)$$

$$(3)$$

where $n = 1, 2, \dots, m$ is the serial number of current class center. Both the distances of training samples to their corresponding class centers and the sum distances of training samples to their non-corresponding class centers are optimized by the contrastive-center loss function.

The convolutional layers in the SegNet also suffer the problem of ignoring inter-class separability during the training process. In order to overcome these issues, the proposed tea sprouts segmentation network is improved by the contrastive-center loss function. Therefore, the TS-SegNet can obtain more discriminative tea sprouts features with the contribution of Eq. (1), and the proposed method has a good performance on the accuracy of tea sprouts segmentation.

2.2 The Proposed TS-SegNet

The network structure of the proposed method is shown in Fig. 1, it is based on the SegNet. The input of the method is a tea sprouts image, and the output is the tea sprouts segmentation results. The TS-Segnet encoding part consists of 13 convolutional layers in the VGG-16 network [11], and the fully connected layers are discarded in order to retain high resolution feature maps. The TS-Segnet decoding part consists an inverse VGG-16 network for the up-sampling, each layer in encoding has the corresponding layer in decoding. Each convolutional layer uses a 3×3 kernel, the size of max-pool layers is 2×2 and the ReLU activation function is employed in the TS-Segnet. In the end, a softmax layer with 2 outputs is used to implement 2-label classification as background (1) and tea sprouts (2).

A U-Net [10] type skip connection is employed only at the last up-sampling layer as shown in Fig. 1 (Dotted arrow part) to incorporate feature maps with fine details. At this layer, a 1×1 convolution layer is utilized to enhance coarser and finer information for the tea sprouts segmentation, and also to reduce the number of parameters for the final convolutional layer. A contrastive-center loss function is also utilized to increase the intra-class compactness and inter-class separability of the extracted tea sprouts features. Therefore, the proposed TS-SegNet has a good performance on tea sprouts segmentation task with the improvement of

skip connection and contrastive-center loss function.

3. Experiment and Evaluation

3.1 Dataset and Implement Details

In the experiment, the dataset consists of 1296 training and 864 testing RGB images (Longjing tea, overshoot or 45-degree angle shooting) at 360×480 resolution. The ground truth is provided with manual segmentation with labels as 1 and 2 for background and tea sprouts respectively. All the input RGB images are performed the local contrast normalization operation [12].

We initialized the encoder and decoder weights by the technique described in [13], and also utilized the stochastic gradient descent with a fixed learning rate of 0.01 to train the proposed TS-SegNet. The training set is shuffled and each mini-batch consists of 6 images. We used the contrastive-center loss as the objective function in the network training part. In the last convolutional layer of decoding part, a 1×1 convolution operation was employed to receive more tea sprouts coarser and finer information via skip connection, the number of parameters in the layer is also reduced.

For each tea sprouts region, overlap error (OE) and relative difference error (RDE) were used as the evaluation metric [14]. OE and RDE are defined as:

$$OE = 1 - \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

$$RDE = \frac{abs(|A| - |B|)}{|B|} \quad (5)$$

A is the segmentation results to be evaluated, B denotes the reference manual delineation results, where $||$ calculates the area of a logical image input (pixel numbers).

3.2 Results and Discussion

To validate the efficiency of the proposed TS-SegNet, we implemented SegNet [7], U-Net [10], U-SegNet [12], Graph cuts [15] and Level set [16] to benchmark the performance on tea sprouts segmentation. The qualitative comparisons of TS-SegNet segmentation with different deep architectures can be seen in Fig. 2. The qualitative results show that the proposed TS-SegNet has a superior performance as compared to the other deep architectures. From the TS-SegNet overlay label (Fig. 2 last row), it can be observed that the tea sprouts segmentation result gained by our method is complete in shape, especially for the small tea sprouts. It also can be observed that the proposed TS-SegNet can reduce the impact of individual shape and color differences from tea sprouts, ambient illumination and shooting angle during image acquisition, which makes it suitable for the tea automatic picking.

Table 1 shows the tea sprouts testing set accuracy with different methods, the comparison results indicate that the proposed TS-SegNet has the minimum OE and RDE. The minimum value of RDE in the tea sprouts testing set is 0.01

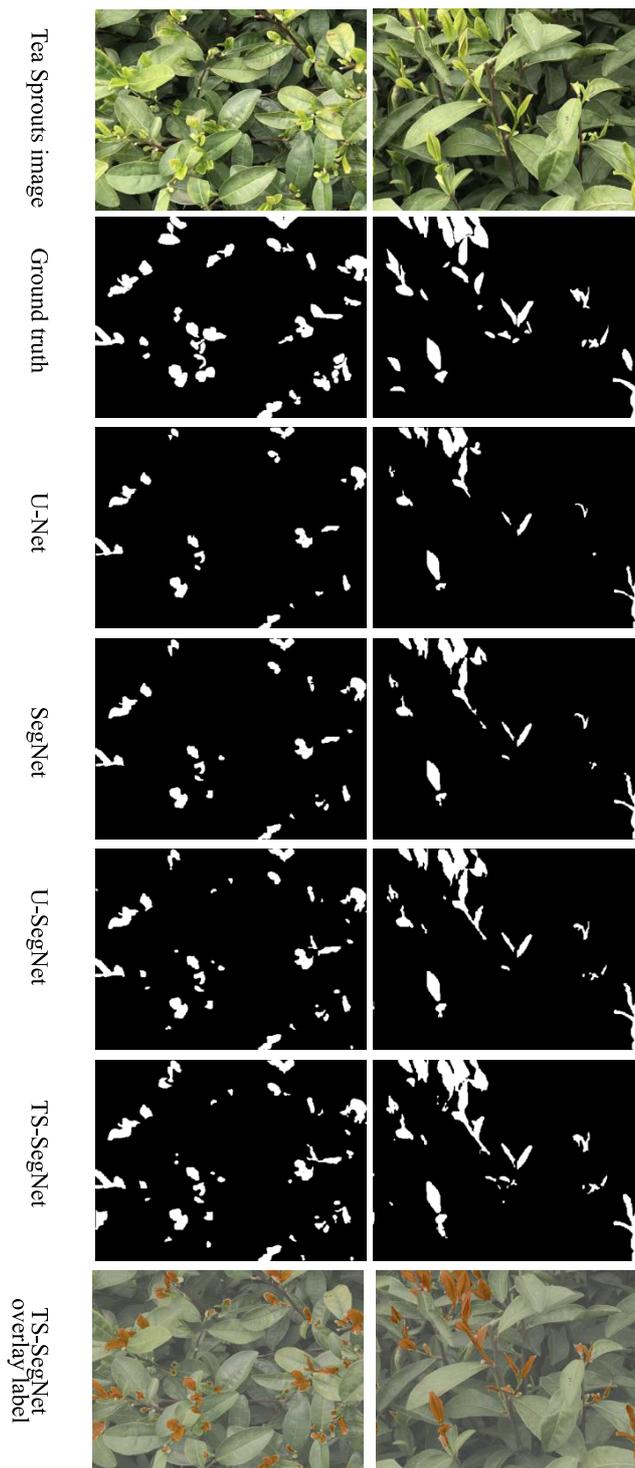


Fig. 2 Results on tea sprouts segmentation with different deep architectures

(gained by our method with a 45K iterations), and the corresponding segmentation result is almost coincident with the ground truth, the only difference is a small part of the segmented tea sprout stalk.

In the experimental part, we also compared different loss functions to verify the proposed method. Table 2 shows

Table 1 Tea sprouts testing set accuracy with different methods.

| Method | OE | | | RDE | | |
|----------------|------|------|-------------|------|------|-------------|
| | Min | Max | Mean | Min | Max | Mean |
| Graph cuts[15] | 0.17 | 0.45 | 0.36 | 0.09 | 0.39 | 0.28 |
| Level set[16] | 0.17 | 0.44 | 0.35 | 0.08 | 0.39 | 0.27 |
| U-Net[7] | 0.18 | 0.37 | 0.28 | 0.09 | 0.32 | 0.20 |
| SegNet[10] | 0.16 | 0.35 | 0.25 | 0.07 | 0.28 | 0.17 |
| U-SegNet[12] | 0.14 | 0.35 | 0.24 | 0.05 | 0.26 | 0.15 |
| TS-SegNet | 0.09 | 0.32 | 0.20 | 0.01 | 0.20 | 0.11 |

Table 2 TS-SegNet testing set mean accuracy with different loss functions.

| Loss function | OE | RDE |
|-------------------------|-------------|-------------|
| Softmax | 0.23 | 0.14 |
| Center loss | 0.21 | 0.13 |
| Contrastive-center loss | 0.20 | 0.11 |

the tea sprouts testing set accuracy with different loss functions, the network structure used is as shown in Fig. 1. The comparison results indicate that the contrastive-center loss function has a better performance than other loss functions. The main reason is the intra-class compactness and inter-class separability of the tea sprouts features are effectively measured via the contrastive-center loss function.

4. Conclusion

In this paper, a novel TS-SegNet method for tea sprouts segmentation is presented. With the contribution of contrastive-center loss function and the skip connections, the intra-class compactness and inter-class separability are comprehensively utilized, and more discriminative tea sprouts features are automatically extracted. Hence, the accuracy of tea sprouts segmentation can be efficiently increased through these gained discriminative features. The experimental results show that the proposed method leads to high accuracy, and has the minimum OE and RDE compared with other deep architectures. The segmented tea sprout regions are complete and almost coincident with the ground truth, which can be used in the tea automatic picking. Future work includes eliminating the effects of the tea sprouts stalks during the segmentation and further improving the segmentation effect.

Acknowledgments

This research was supported by Suzhou Science and Technology Bureau (SNG2017056, SNG2017058), Jiangsu Overseas Research & Training Program for University Prominent Young & Middle-aged Teachers and Presidents, Science and Technology Project (KYPY201702).

References

[1] H. Zhang, Y. Chen, W. Wang, et al., "Positioning method for tea picking using active computer vision," *Transactions of the Chinese Society for Agricultural Machinery*, vol.45, pp.61–78, Sept. 2014.

[2] F. Yang and Y. Qing, "Recognition of the tea sprout based on color and shape features," *Transactions of the Chinese Society for Agricultural Machinery*, vol.1, pp.119–122, Sept. 2009.

[3] X. Tang, X.M. Wu, F.G. Zhang, et al., "Contrastive research on tender tea recognition based on multiple threshold segmentation methods," *Agricultural Equipment & Technology*, vol.6, pp.10–15, Dec. 2013.

[4] P.D. Shao, M.H. Wu, X.W. Wang, J. Zhou, and S. Liu, "Research on the tea bud recognition based on improved k-means algorithm," *2018 2nd International Conference on Electronic Information Technology and Computer Engineering*, pp.1–5, Nov. 2018.

[5] X. Wu, J. Yang, and S. Wang, "Tea category identification based on optimal wavelet entropy and weighted k-Nearest Neighbors algorithm," *Multimedia Tools and Applications*, vol.77, pp.3745–3759, Feb. 2018.

[6] S.T. Zhang, Z.Y. Wang, X.G. Zhou, et al., "Recognition of tea disease spot based on hyperspectral image and genetic optimization neural network," *Transactions of the Chinese Society of Agricultural Engineering*, vol.22, pp.200–207, Nov. 2017.

[7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.39, no.12, pp.2481–2495, Jan. 2017.

[8] R. Wu, M. Feng, W. Guan, et al., "A mutual learning method for salient object detection with intertwined multi-supervision," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.8150–8159, June 2019.

[9] P. Zhang, W. Liu, H. Wang, Y. Lei, and H. Lu, "Deep gated attention networks for large-scale street-level scene segmentation," *Pattern Recognition*, vol.88, pp.702–714, April 2019.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, pp.234–241, Nov. 2015.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[12] K. Jarrett, K. Kavukcuoglu, and Y. LeCun, "What is the best multi-stage architecture for object recognition," *2009 IEEE 12th International Conference on Computer Vision*, pp.2146–2153, Sept. 2009.

[13] P. Kumar, P. Nagar, C. Arora, and A. Gupta, "U-Segnet: fully convolutional neural network based automated brain tissue segmentation tool," *2018 25th IEEE International Conference on Image Processing*, pp.3503–3507, Oct. 2018.

[14] B. Chen, Y. Chen, G. Yang, J. Meng, R. Zeng, and L. Luo, "Segmentation of liver tumor via nonlocal active contours," *2015 IEEE International Conference on Image Processing*, pp.3745–3748, Oct. 2015.

[15] S. Dai, K. Lu, J. Dong, Y. Zhang, and Y. Chen "A novel approach of lung segmentation on chest CT images using graph cuts," *Neurocomputing*, vol.168, pp.799–807, Nov. 2015.

[16] Y. Guo, A. Şengür, and J.-W. Tian, "A novel breast ultrasound image segmentation algorithm based on neutrosophic similarity score and level set," *Computer methods and programs in biomedicine*, vol.123, pp.43–53, Jan. 2016.