LETTER

Fresh Tea Sprouts Segmentation via Capsule Network

Chunhua QIAN[†,††], ***Student Member***, **Xiaoyan QIN**[†††], **Hequn QIANG**[††a)], **Changyou QIN**[††], ***and*** **Minyang LI**[†], ***Nonmembers***

SUMMARY The segmentation performance of fresh tea sprouts is inadequate due to the uncontrollable posture. A novel method for Fresh Tea Sprouts Segmentation based on Capsule Network (FTS-SegCaps) is proposed in this paper. The spatial relationship between local parts and whole tea sprout is retained and effectively utilized by a deep encoder-decoder capsule network, which can reduce the effect of tea sprouts with uncontrollable posture. Meanwhile, a patch-based local dynamic routing algorithm is also proposed to solve the parameter explosion problem. The experimental results indicate that the segmented tea sprouts via FTS-SegCaps are almost coincident with the ground truth, and also show that the proposed method has a better performance than the state-of-the-art methods.

key words: *tea sprouts segmentation, patch-based local dynamic routing, CapsNet, FTS-SegCaps*

1. Introduction

Fresh tea sprouts segmentation is an important technology of automatic tea picking systems based on machine vision. The segmentation results affect the automatic picking accuracy directly. Moreover, the area and contour information of the tea sprouts is mainly obtained from the segmentation results. In the last few years, many machine vision based fresh tea sprouts detection or segmentation methods are proposed. Zhang et al. [1] introduced a famous tea sprout identification and segmentation method based on improved watershed algorithm. Gui et al. [2] introduced a fresh tea sprouts detection method via Yolov5, the bottleneck attention module is added to the backbone network to suppress invalid information, but the detection accuracy is not high enough to achieve the requirements of automatic tea picking systems. A deep convolutional Encoder-Decoder Network based fresh tea sprouts segmentation method is proposed in [3], the SegNet is used as the backbone of the network to gain the segmentation results.

Most Convolutional Neural Networks (CNNs) based fresh tea sprouts segmentation methods have a better performance than the traditional methods. CNNs currently utilize pooling to route the information between layers, such as max-pooling and average pooling. However, pooling is a naive way of routing as it discards information about the precise location and pose of the entity within the region which can be valuable for the classification purpose [4]. Recently, a new structure called Capsule Network (CapsNet) is proposed in [5] to overcome these shortcomings. The main idea is to record the spatial relationship between local parts and whole object by encoding operation. The routing between capsules enables the network to recognize objects from different viewpoints not seen in the original CNNs. Therefore, we rethink the main processing steps of fresh tea sprouts segmentation task, find that the spatial relationship between tea sprouts can be utilized to improve the segmentation performance. Motivated by [6], [7], A novel method for Fresh Tea Sprouts Segmentation based on Capsule Network (FTS-SegCaps) is proposed in this paper. In order to solve the parameter explosion problem and further improve the segmentation performance, a patch-based local dynamic routing algorithm is also developed in the FTS-SegCaps. From the comparisons of different segmentation method, the segmented fresh tea sprouts via our method are more coincident with the ground truth. The experimental results show that the proposed method has a better performance than the state-of-the-art methods.

2. Methods

2.1 The Proposed FTS-SegCaps

The network structure of the proposed FTS-SegCaps is shown in Fig. 1, the input is a lager image (e.g. 360×480 pixels) containing fresh tea sprouts, and the output is the corresponding segmentation result. The network backbone is followed by U-Net and CapsNet, meanwhile, the skip connections concatenating together capsule types from earlier layer with the same spatial dimensions are also included. The FTS-SegCaps mainly contains four types of capsules: *PrimaryCaps*, *SecondaryCaps*, *TertiaryCaps*, *SegCaps*. We utilize a 2D convolution to convert pixel intensities to local features, each pixel is transformed into a single capsule which contains a 16-dimensional vector, and in this way we obtain the *PrimaryCaps* with a grid of 360×480 capsules. Then the *SecondaryCaps* and *TertiaryCaps* are designed to obtain high level features and the spatial relationship between local parts and the whole tea sprout. In the processing of routing calculation between capsule layers, we proposed

Manuscript received September 12, 2023.
Manuscript revised December 11, 2023.
Manuscript publicized January 17, 2024.
[†]The authors are with Nanjing Forestry University, Nanjing, China.
[††]The authors are with Suzhou Polytechnic Institute of Agriculture, Suzhou, China.
[†††]The author is with Global Institute of Software Technology, Suzhou, China.
a) E-mail: hqqiang@szai.edu.cn
DOI: 10.1587/transinf.2023EDL8062

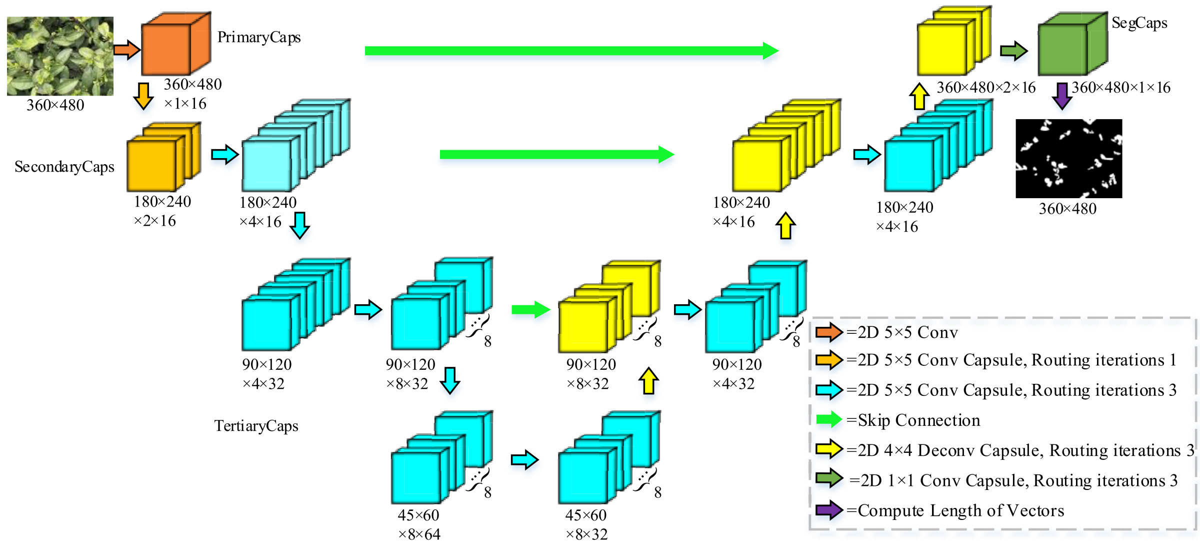

Fig. 1 The network structure of the proposed FTS-SegCaps. The U-Net is utilized as the backbone, and the skip connections concatenating together capsule types from earlier layer with the same spatial dimensions are also included.

a patch-based local dynamic routing algorithm to solve the memory burden and parameter explosion problem [8], the details are shown in Sect. 2.2. The child capsules are only routed to parent capsules within a defined path-based local window (the size is $h^{(l)} \times w^{(l)}$, l is the layer number), and the transformation matrices are shared for each member of the grid within the same capsule type. In Fig. 1, the routing iterations used for *PrimaryCaps* is 1, and for other routing layers are set to 3. In the proposed FTS-SegCaps, we also add the skip connection to overcome the gradient vanishing and over-fitting problem.

In order to obtain the fresh tea sprouts segmentation results, we design the decoder part following the baseline of U-Net. A 2D deconvolutional operation is utilized to reconstruct the same grid of the input image, and a 2D convolution is also employed to obtain the *SegCaps*. Then we compute the vectors length of each capsule in the *SegCaps* layer to obtain the final segmentation results. The main processing of deconvolutional operation is to upsample the child capsules to the corresponding parent capsules. A learned transformation matrix from layer l to $l+1$ is defined as $M_{t_j^{(l+1)}}$, where $t_j^{(l+1)}$ is a parent capsule type, and $U_{x,y|t_i^{(l)}}$ is the sub-grid of child capsules outputs at position (x, y) in layer l. Then, we reshape the child capsule outputs following the fractional striding formulation used in [9], and calculate the prediction vectors by,

$$\hat{\mathbf{u}}_{x,y|t_i^{(l)}} = M_{t_j^{(l+1)}} \cdot U_{x,y|t_i^{(l)}}, \qquad \forall t_j^{(l)} \in T^{(l)}, \tag{1}$$

where $T^{(l)} = \{t_1^{(l)}, t_2^{(l)}, \cdots t_m^{(l)} | m \in \mathbb{N}\}$, is a set of capsule types. Then, the $\hat{\mathbf{u}}_{x,y|t_i^{(l)}}$ is utilized as the input of the patch-based local dynamic routing algorithm to reconstruct the parent capsule.

Algorithm 1: Patch-Based Local Dynamic Routing

Set d, l, x, y, $b_{t_i^{(l)}|x,y} \leftarrow 0$

for d iterations **do**

for all child capsules types $t_i^{(l)}$ at (x, y) :

$$\mathbf{r}_{t_i^{(l)}} \leftarrow softmax(\mathbf{b}_{t_i^{(l)}})$$

for all parent capsule types $t_j^{(l+1)}$ at (x, y) :

$$\mathbf{p}_{x,y} \leftarrow \sum_n r_{t_i^{(l)}|x,y} \hat{\mathbf{u}}_{x,y|t_i^{(l)}}, \quad n = h^{(l)} \times w^{(l)}$$

for all parent capsule types $t_j^{(l+1)}$ at (x, y) :

$$\mathbf{v}_{x,y} = squash(\mathbf{p}_{x,y})$$

for all child capsules types $t_i^{(l)}$:

$$b_{t_i^{(l)}|x,y} \leftarrow mean_i(b_{t_i^{(l)}|x,y} + \hat{\mathbf{u}}_{x,y|t_i^{(l)}} \cdot \mathbf{v}_{x,y})$$

return $\mathbf{v}_{x,y}$

2.2 The Proposed Patch-Based Local Dynamic Routing Algorithm

The routing algorithm used in the original CapsNet is fully-connected, the parameter grows rapidly as the image gets larger. Therefore, a patch-based local dynamic routing algorithm is proposed to overcome these problems. In convolutional capsules, the parent capsule at position (x, y) in layer $l+1$ is defined as $\mathbf{p}_{x,y}$, which receives a set of prediction vectors $\hat{\mathbf{u}}_{x,y|t_i^{(l)}}$ calculated by an $h^{(l)} \times w^{(l)}$ grid of $z^{(l)}$ dimensional child capsules in layer l. The routing coefficients $r_{t_i^{(l)}|x,y}$ used to obtain the $\mathbf{p}_{x,y}$ are calculated by a Softmax operation,

$$r_{t_i^{(l)}|x,y} = \frac{\exp(b_{t_i^{(l)}|x,y})}{\sum_{t_j^{(l+1)}} \exp(b_{t_i^{(l)}|t_j^{(l+1)}})}, \tag{2}$$

where $b_{t_i^{(l)}|x,y}$ is the log prior probabilities to predict $\hat{\mathbf{u}}_{x,y|t_i^{(l)}}$ routed to $\mathbf{p}_{x,y}$, and the value is updated in Algorithm 1. Then, the output $\mathbf{v}_{x,y}$ is obtained via a squashing function[],

$$\mathbf{v}_{x,y} = \frac{\|\mathbf{p}_{x,y}\|^2}{1 + \|\mathbf{p}_{x,y}\|^2} \frac{\mathbf{p}_{x,y}}{\|\mathbf{p}_{x,y}\|}. \tag{3}$$

The main steps of the proposed patch-based local dynamic routing algorithm are summarized in Algorithm 1. d is the routing iterations used in the routing algorithm. The transformation martices $M_{t_j^{(l+1)}}$ are shared for each member of the grid within the same capsule type. Therefore, we utilize two kinds of routings in Fig. 1 according to the different capsule types.

3. Experiment and Evaluation

3.1 Dataset and Implement Details

The dataset used in the experiment contains 5180 RGB images at 360×480 resolution. Each image contains Longjing tea sprouts obtained at 45-degree angle shooting or overshoot. 3108 images are utilized as the training data, and the rest is test data. The ground-truth is obtained by manual segmentation with labels as 1, 2 for background and tea sprouts respectively. All input images are processed with the local contrast normalization operation.

The experiments are run with 2 NVIDIA Tesla P40 GPU, the TensorFlow and Keras are utilized to implement the model. To remain consistent, we train all models from scratch on the dataset. We utilize the weighted-BCE loss for the output in the FTS-SegCaps, set the learning rate to 0.0001, the routing iterations $d = 1$ for *PrimaryCaps*, $d = 3$ for other routing layers, and make $h^{(l)} = 9$, $w^{(l)} = 12$. In the experiment, the Overlap Error (OE) and Relative Difference Error (RDE) are employed to evaluate the segmentation results, the definitions of OE and RDE are in [3]. The Number of Training Parameters (NTP) and FPS are also used to further evaluate the proposed algorithm.

3.2 Segmentation Results and Discussion

In order to evaluate the proposed FTS-SegCaps, we implemented U-Net [10], SegNet [11], TS-SegNet [3] and SegCaps [6] to compare the performance on fresh tea sprouts

Table 1 Fresh tea sprouts testing set accuracy with different methods.

Method	OE			RDE			NTP	FPS
	Min	Max	Mean	Min	Max	Mean		
U-Net[10]	0.17	0.34	0.26	0.08	0.29	0.19	5.04×10^5	32
SegNet[11]	0.15	0.32	0.24	0.06	0.26	0.16	5.04×10^5	25
TS-SegNet[3]	0.08	0.31	0.20	0.01	0.19	0.10	5.21×10^5	24
SegCaps[6]	0.07	0.28	0.18	0.01	0.17	0.09	2.23×10^7	18
FTS-SegCaps	0.06	0.23	**0.15**	0.01	0.13	**0.07**	9.75×10^6	21

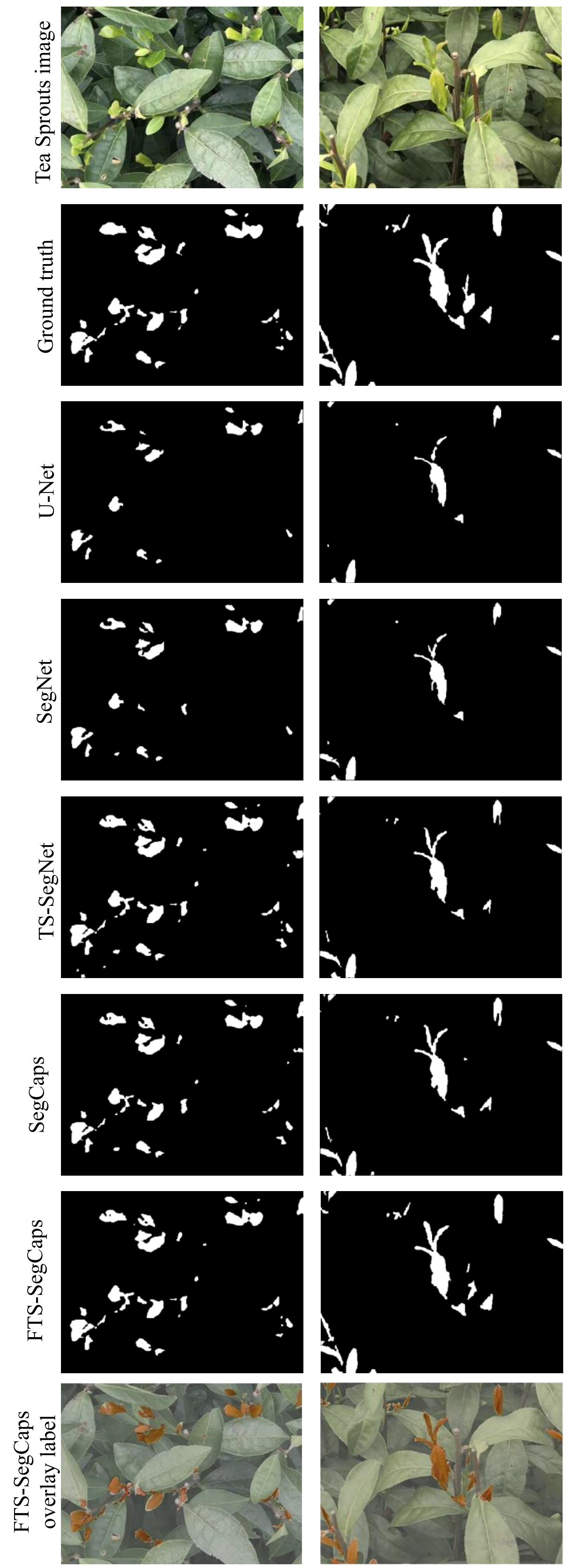

Fig. 2 Results on fresh tea sprouts segmentation with different methods.

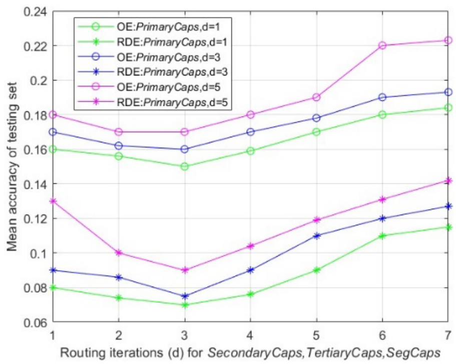

Fig. 3 FTS-SegCaps mean accuracy of testing set with different routing iterations.

segmentation. Table 1 shows the segmentation results of testing set with different methods. The comparison results indicate that the proposed FTS-SegCaps has a better performance on fresh tea sprouts segmentation. Compared to TS-SegNet, the mean of OE and RDE obtained by our method has dropped by 16.7% and 22.2% respectively. Moreover, the NTP in our method is decreased by 56.3% and the FPS increased by 16.7% compared to the SegCaps. The parameter explosion problem is basically solved with the contribution of the patch-based local dynamic routing algorithm.

Figure 2 shows some segmentation results of testing set with different methods. It can be observed that the tea sprouts region obtained by our method is closer to the ground-truth. Moreover, the rotated tea sprouts also can be segmented accurately with the contribution of the proposed CapsNet structure and dynamic routing. The spatial relationship between local parts and whole tea sprout is effectively utilized in our method, which leads to a better performance than the state-of-the-art methods.

In the experimental part, we also compared different routing iterations d to further verify the proposed FTS-SegCaps. Figure 3 shows the mean accuracy of fresh tea sprouts testing set with different routing iterations. The d used for *PrimaryCaps* is set to 1, 3, 5 respectively, and for other routing layers are set from 1 to 7 respectively. From the comparison of different routing iterations, we observed that too small or large routing iterations may lead to local optima or overfitting problem. In Fig. 3, the $d = 1$ for *PrimaryCaps* and $d = 3$ for other routing layers has the best performance on fresh tea sprouts segmentation, which is similar to the original CapsNet.

4. Conclusion

In this paper, a novel FTS-SegCaps method for fresh tea sprouts segmentation is introduced. The spatial relationship between local parts and whole fresh tea sprout is retained and effectively utilized by the CapsNet structure. A patch-based local dynamic routing algorithm is also proposed to solve the parameter explosion problem. With these strategies, the fresh tea sprouts with different postures can be segmented accurately via the proposed method. The experimental results indicate that the proposed FTS-SegCaps has a better performance than the state-of-the-art methods. Future work includes eliminating the effects of ambient lighting or damaged old leaves, further improving the segmentation performance, and collecting more data.

Acknowledgments

This research was supported by Jiangsu Modern Agricultural Industrial Technology System Project (JATS-2022-386), Jiangsu Vocational College Specialty Leading Person Training Project (2021GRFX052), Suzhou Polytechnic Institute of Agriculture Doctoral Promotion Program Research Fund(BS2022-15) and Young and Middle-aged Academic Leaders of Jiangsu Qinglan Program.

References

[1] L. Zhang, L. Zou, C. Wu, J. Jia, and J. Chen, "Method of famous tea sprout identification and segmentation based on improved watershed algorithm," Computers and Electronics in Agriculture, vol.184, 106108, May 2021.
[2] Z. Gui, J. Chen, Y. Li, Z. Chen, C. Wu, and C. Dong, "A lightweight tea bud detection model based on Yolov5," Computers and Electronics in Agriculture, vol.205, 107636, Jan. 2023.
[3] C. Qian, M. Li, and Y. Ren, "Tea sprouts segmentation via improved deep convolutional encoder-decoder network," IEICE Trans. Inf. & Syst., vol.E103-D, no.2, pp.476–479, Feb. 2019.
[4] A. Mobiny, H. Lu, H.V. Nguyen, B. Roysam, and N. Varadarajan, "Automated classification of apoptosis in phase contrast microscopy using capsule network," IEEE Trans. Med. Imag., vol.39, no.1, pp.1–10, Jan. 2019.
[5] S. Sabour, N. Frosst, and G.E. Hinton, "Dynamic routing between capsules," Advances in Neural Information Processing Systems (NIPS 2017), Jan. 2018.
[6] R. LaLonde, Z. Xu, I. Irmakci, S. Jain, and U. Bagci, "Capsules for biomedical image segmentation," Medical image analysis, vol.68, 101889, Feb. 2021.
[7] A. Deleruyelle, C. Versari, and J. Klein, "Self-mentoring: A new deep learning pipeline to train a self-supervised U-net for few-shot learning of bio-artificial capsule segmentation," Computers in Biology and Medicine, vol.152, 106454, Jan. 2023.
[8] A. Mobiny, H. Lu, H.V. Nguyen, B. Roysam, and N. Varadarajan, "Automated classification of apoptosis in phase contrast microscopy using capsule network," IEEE Trans. Med. Imag., vol.39, no.1, pp.1–10, Jan. 2020.
[9] Long J, Shelhamer E, Darrell T. "Fully convolutional networks for semantic segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol.39, no.4, pp.640–651, April 2017.
[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] 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.