

LETTER

Fresh Tea Shoot Maturity Estimation via Multispectral Imaging and Deep Label Distribution Learning

Bin CHEN^{†a)}, Member and JiLi YAN[†], Nonmember

SUMMARY Fresh Tea Shoot Maturity Estimation (FTSME) is the basement of automatic tea picking technique, determines whether the shoot can be picked. Unfortunately, the ambiguous information among single labels and uncontrollable imaging condition lead to a low FTSME accuracy. A novel Fresh Tea Shoot Maturity Estimating method via multispectral imaging and Deep Label Distribution Learning (FTSME-DLDDL) is proposed to overcome these issues. The input is 25-band images, and the output is the corresponding tea shoot maturity label distribution. We utilize the multiple VGG-16 and auto-encoding network to obtain the multispectral features, and learn the label distribution by minimizing the Kullback-Leibler divergence using deep convolutional neural networks. The experimental results show that the proposed method has a better performance on FTSME than the state-of-the-art methods.

key words: fresh tea shoot maturity estimation, deep label distribution learning, Kullback-Leibler loss function, label ambiguity

1. Introduction

About one in ten teas per year are not picked up in time due to lack of tea plucking workers. Automatic tea picking by machine is an effective way to solve the problems [1]. Fresh tea shoot maturity estimation, determined whether the shoot is picked, is a core step in automatic tea picking [2], [3]. Many researches about tea estimation have been reported in recent years. Liang et al. [4] achieved the estimation of tea quality by infusion colour difference analysis, the method was used in dried tea leaves. Hazarika et al. [5] introduced a fresh tea leave polyphenol estimating method using near infrared reflectance (NIR) spectroscopy and regression model. Kamagata et al. [6] estimated the tea growth stage using airborne hyperspectral data, the multiple regression analysis was included. A maturity level estimation method of leaves based on morphological feature was introduced in [7], the maturity level is divided into premature, mature and over-mature. Nandi et al. [8] proposed a mango fruit sorting and grading method by maturity level and size, four maturity levels were used. Most existed maturity estimation methods are based on single-label recognition (SLR) or multi-label recognition (MLR). For instance, the fresh tea shoot maturity is usually divided into immature, mature 1 to 5 and overmature. Actually it is attributed to the SLR problem, but are there any ambiguity information between mature 1 and 2? The answer is yes, the wrong estimation often occurs

when the two maturities are too close, which leads to a low estimating accuracy.

In order to overcome these issues, a novel Fresh Tea Shoot Maturity Estimating method via multispectral imaging and Deep Label Distribution Learning (FTSME-DLDDL) is proposed in this paper. The convolution neural network (CNN) based methods have achieved the inspiring performance on various tasks such as image classification [9], [10], segmentation [11], and object detection [12]. The features automatically extracted from the CNN are robust and have natural advantages. The deep label distribution learning is an improvement and extension of label distribution learning [13]–[15], it is developed to achieve the label distribution learning on CNN. In the proposed FTSME-DLDDL, we convert the label of each multispectral image into a discrete label distribution, then achieve the fresh tea shoot maturity estimating by multiple convolution neural networks and deep label distribution learning. An auto-encoding [16] network and Kullback-Leibler (KL) loss function are also employed to fuse the extracted multispectral features and achieve the deep label distribution learning separately. The experimental results show that the novel method leads to good estimating results, meanwhile, the proposed method can be used in many other shoot maturity estimating tasks.

2. Methods

2.1 Deep Label Distribution Learning

Deep label distribution learning is a general machine learning paradigm, and the label distribution represents the degree to which each label describes the instance. Let X be an instance, \mathbf{y} is the corresponding label distribution, $L = \{l_1, l_2, \dots, l_C\}$ is the label set, and $x = \phi(X; \theta)$ is the activation of the last layer in a CNN. The probability distribution of these activations can be calculated by:

$$\hat{y}_j = \frac{\exp(x_j)}{\sum_t \exp(x_t)} \quad (1)$$

then, the similar distribution $\hat{\mathbf{y}}$ is obtained by θ . In order to solve the $\hat{\mathbf{y}}$, a Kullback-Leibler (KL) divergence based measurement are employed, the best θ^* is determined:

$$\theta^* = \arg \min_{\theta} \sum_k y_k \ln \frac{y_k}{\hat{y}_k} \quad (2)$$

the loss function can be defined as:

Manuscript received March 11, 2020.

Manuscript revised May 13, 2020.

Manuscript publicized June 1, 2020.

[†]The authors are with College of Mathematics Physics and Information Engineering, Jiaxing University, Jiaxing 314000, China.

a) E-mail: chenbin@zjxu.edu.cn

DOI: 10.1587/transinf.2020EDL8038

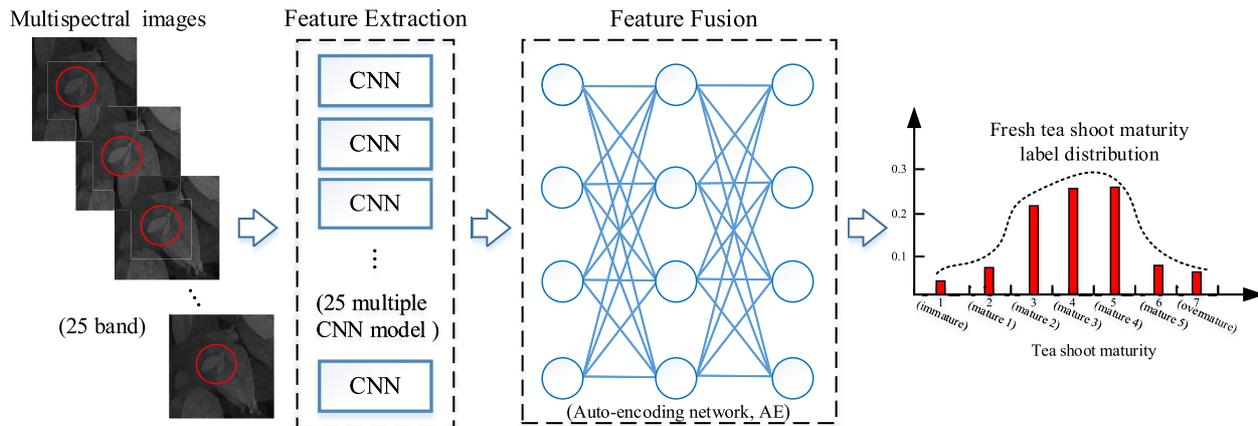


Fig. 1 Flow chart of the proposed FTSME-DLDDL method. The red circle area on multispectral image is the ROI gained from the fresh tea shoot segmentation method in [17], the multiple CNN model is based on VGG-16 [18], and the feature fusion structure is achieved by four auto-encoding layers.

$$T = - \sum_k y_k \ln \hat{y}_k \quad (3)$$

where, the Stochastic Gradient Descent (SGD) is utilized to minimize Eq. (3). The derivative of T with respect to θ can be calculated by:

$$\frac{\partial T}{\partial \theta} = (\hat{\mathbf{y}} - \mathbf{y}) \frac{\partial x}{\partial \theta} \quad (4)$$

the out of the DLDDL is l_{i^*} , where i^* is

$$i^* = \arg \max_i \hat{y}_i \quad (5)$$

The label distribution \mathbf{y} is usually hard to obtain in many datasets, fortunately, there are several label enhancement methods to fix that problem. Fuzzy set based label enhancement methods [19], [20] use the idea of fuzzy clustering, kernel membership and fuzzy operation to mining the related information between labels, then convert the logical label into label distribution. Another type of label enhancement methods are based on the graph or manifold [21] model, the topology between the instances is also included in these methods. Hence, we utilized the label enhancement to obtain the label distribution of fresh tea shoot maturity.

2.2 The Proposed FTSME-DLDDL Method

The flow chart of the proposed FTSME-DLDDL method is shown in Fig. 1, the input is 25-band images, and the output is the corresponding tea shoot maturity label distribution. A sample of estimated tea shoot maturity label distribution $\hat{\mathbf{y}}$ is shown in Fig. 1, it can be observed that the label ambiguity information is efficiently utilized via the discrete label distribution, which helps prevent the network from overfitting. The multiple CNN model is based on VGG-16 (remove the last layer and modify the last full connected layer), each band image has the corresponding feature extraction structure. In order to obtain the tea shoot maturity label distribution, a hierarchical auto-encoding (AE) subnetwork with

four layers is employed to fuse these multispectral features and further reduce the dimension. The size of each auto-encoding layer is 6400, 3200, 2000, 800 separately. The ReLU and dropout are utilized in the AE fusion subnetwork, the rate of dropout is 0.5. Then, we modified the output layer, and replaced the softmax loss function with the KL loss function.

During the training, the initialization of the deep network is performed randomly via a Gaussian distribution with zero mean and 0.02 standard deviation. The parameter used in ReLU is initialized to 0.2, the rate of dropout used in VGG-16 and AE fusion subnetwork is 0.5, the weight decay is set to 0.001, and the Stochastic Gradient Descent (SGD) optimization is also included.

3. Experiment and Evaluation

The experiments are run with 4 NVIDIA Tesla P40 20-GB GPU, the multiple CNN model is trained on TensorFlow, and other steps are implemented using MATLAB. The multispectral images are obtained by the hyperspectral camera XIMEA MQ022HG, the value of 25 bands are: 665.9 nm, 680.5 nm, 705.5 nm, 720.1 nm, 732.0 nm, 746.3 nm, 759.0 nm, 772.2 nm, 783.5 nm, 796.4 nm, 815.5 nm, 827.1 nm, 837.4 nm, 848.5 nm, 858.5 nm, 869.3 nm, 877.9 nm, 886.7 nm, 902.1 nm, 910.2 nm, 919.8 nm, 926.0 nm, 934.8 nm, 939.8 nm, 944.5 nm. In the experiment, we compare the FTSME accuracy with VGG-16+AE, ZF-Net+AE, DLDDL+ZF-Net+AE methods and false color image based methods (ZF-Net, VGG-16, DLDDL+ZF-Net, and DLDDL+VGG-16). The RGB channels of the false color image are composed of 944.5 nm, 886.7 nm, 665.9 nm separately. The hyperspectral camera and part of multispectral images are shown in Fig. 2. It can be observed that the tea shoot spectra and aging leaves have close spectral characteristics, which leads to the ambiguity information in single-label FTSME.

The spatial resolution of raw image is 409×216 , we resize the image to 224×224 , and ensure each image only has

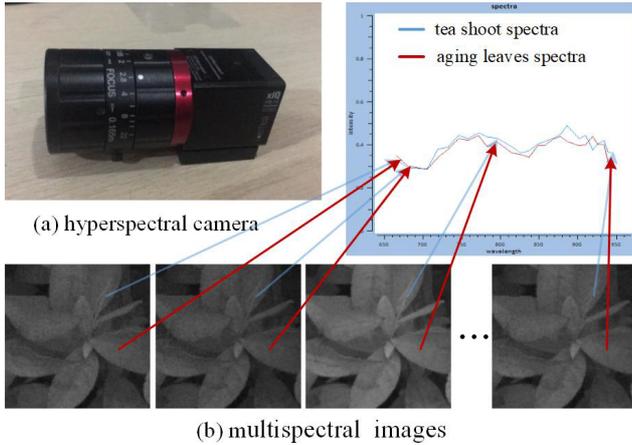


Fig. 2 The hyperspectral camera and part of multispectral images, the tea shoot spectra is shown in blue line, and the aging leaves spectra is shown in red line.

Table 1 Comparisons of different methods for FTSME on testing dataset.

Method	MAE
ZF-Net+AE	0.78 ± 0.05
VGG-16+AE	0.73 ± 0.04
DLDL+ZF-Net+AE	0.72 ± 0.03
The proposed FTSME-DLDDL	0.68 ± 0.02
ZF-Net (false color image based)	1.15 ± 0.07
VGG-16 (false color image based)	0.91 ± 0.06
DLDL+VGG-16 (false color image based)	0.87 ± 0.04
DLDL+ZF-Net (false color image based)	0.90 ± 0.04

one fresh tea shoot. The ROI of fresh tea shoot is based on the segmentation method in [17]. The dataset contains 5620 multispectral images, which were acquired in Hangzhou, China, from March 20th to April 4th, 2019. Each multispectral image has the corresponding tea shoot maturity label distribution obtained via the label enhancement method in [15]. As the fresh tea shoot maturity is usually divided into immature, mature 1 to 5 and overmature, the label distribution is $\mathbf{y} = (y_1, y_2, \dots, y_C)$, and we have $y_s \in [0, 1]$ and $\sum_{s=1}^C y_s = 1$. We also retain the single label l_n , for instance, $l_n = 1$ means the single label of the testing image is immature, and $l_n = 3$ means the single label is mature 2. The Mean Absolute Error (MAE) is employed to evaluate the performance of FTSME:

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^N |\hat{l}_n - l_n| \quad (6)$$

where the l_n is the ground-truth tea shoot maturity of n -th testing image, \hat{l}_n is the corresponding estimated value. Testing images that satisfy $|\hat{l}_n - l_n| \leq g$, $g \in \{1, 2, 3\}$, which can ensure the same maximum error for each category.

We use 80% of multispectral images for training and 20% for evaluation, the 10-fold-cross-validation is also used in the experiment. The last layer of ZF-Net used in ZF-Net+AE, DLDL+ZF-Net+AE methods is modified similarly with the proposed FTSME-DLDDL, and the VGG-16+AE method also has the similar processing. Table 1

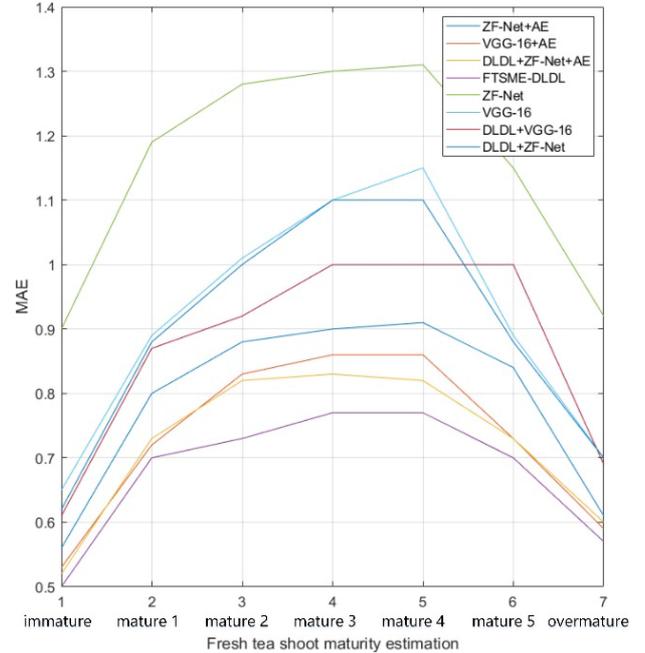


Fig. 3 The comparison of fresh tea shoot maturity MAE curves on testing dataset.

shows the different results of FTSME on testing dataset, the \hat{l}_n of DLDL based methods are obtained by the maximum operation. In Table 1, it can be observed that the multispectral images based methods (with AE layer) have better performance than the false color image based methods. The main reason is more tea shoot spectral features are obtained by the feature extraction and fusion layers in multispectral images based methods. Then, the fresh tea shoot maturity is accurately estimated by these extracted fusion features and CNN based structures. Comparing lines 1, 2 and lines 3, 4 in Table 1, we observe that the DLDL based methods have a significant improvement on MAE as the contribution of ambiguity information utilization via the label distribution learning. We also find the VGG-16 based methods perform better on FTSME than the ZF-Net based methods. The MAE obtained by our method is 0.68, it is much smaller than many other methods, which leads to stable estimating results.

The comparison of fresh tea shoot maturity MAE curves (consists of 7 points corresponding to immature, mature 1 to 5 and overmature) on testing dataset is shown in Fig. 3. It can be observed that the MAE curve obtained by the proposed FTSME-DLDDL has the best performance, and the estimation results are more close to the ground-truth. Comparing with our method, the DLDL+ZF-Net+AE method has the similar performance on points 1 and 7 in MAE curve, but the FTSME accuracy is much lower and unstable on points 2 to 5. From the comparison of each curve in Fig. 3, the FTSME accuracy of VGG-16 based methods is overall higher than the methods based on ZF-Net, and the DLDL based methods improve the accuracy by about 10%.

In Fig. 3, we also observe when the FTSME is imma-

ture (point 1 on MAE curve) or overmature (point 7 on MAE curve), the MAE value is smaller than the FTSME belonging to mature. The main reason is that the immature or overmature tea shoot has the highly distinguishable features, and the maturity is easily estimated. However, the features obtained from the mature tea shoot are usually indistinguishable, and the corresponding single tea shoot maturity label contains much ambiguity information. Therefore, the multi-spectral images and deep label distribution learning are employed in our method to overcome these issues.

4. Discussion and Conclusion

In this paper, we propose the FTSME-DLDDL, a multi-spectral imaging and deep label distribution learning based method to improve the estimating accuracy of fresh tea shoot maturity. The 25-band images are used as the input, and the output is the corresponding tea shoot maturity label distribution. The multiple VGG-16 and AE layers based network is developed to achieve the feature extraction and estimation, the Kullback-Leibler loss function is also employed to achieve the label distribution learning. The experimental results indicate that the proposed method leads to good performance on FTSME. The testing dataset MAE of our method is 0.68, which has a better results than the state-of-the-art methods.

FTSME is still a challenging task as the label ambiguity and the uncontrollable imaging condition, the estimating accuracy needs to be further improved. Future work includes the following aspects. More multispectral images should be collected. The feature extraction and fusion layer could be improved by other deep learning based networks. Furthermore, the proposed method can be utilized for many other shoot maturity estimating tasks.

Acknowledgments

This research was supported by Science and technology planning project of Jiaying, China, Grant number 2018AY11001; Zhejiang Provincial Natural Science Foundation, Grant number LQ18F020007, LY18F020021 and LQ20F020026.

References

- [1] L. Luo, X. Zou, J. Xiong, et al., "Automatic positioning for picking point of grape picking robot in natural environment," *Trans. Chinese Society of Agricultural Engineering*, vol.31, pp.14–21, 2015.
- [2] G. Ya, "Research on the application of automation software control system in tea garden mechanical picking," *International Conference on Applications and Techniques in Cyber Security and Intelligence*, pp.1830–1836, 2019.
- [3] W. Pei and X. Wang, "The two-dimension coordinates extraction of tea shoots picking based on image information," *Acta Agriculturae Zhejiangensis*, vol.28, pp.522–527, 2016.

- [4] Y. Liang, J. Lu, L. Zhang, S. Wu, and Y. Wu, "Estimation of tea quality by infusion colour difference analysis," *Journal of the Science of Food and Agriculture*, vol.85, no.2, pp.286–292, 2005.
- [5] A.K. Hazarika, S. Chanda, S. Sabhapondit, S. Sanyal, P. Tamuly, S. Tasrin, D. Sing, B. Tudu, and R. Bandyopadhyay, "Quality assessment of fresh tea leaves by estimating total polyphenols using near infrared spectroscopy," *Journal of Food Science and Technology*, vol.55, no.12, pp.4867–4876, 2018.
- [6] N. Kamagata, Y. Imai, Y. Akamatsu, et al., "Estimation of the tea growth stage using airborne hyperspectral data," *The 4th International Conference on O-CHA (Tea) Culture and Science*, pp.1–2, 2010.
- [7] P. Sirisomboon, M. Tanaka, T. Kojima, and P. Williams, "Nondestructive estimation of maturity and textural properties on tomato 'Momotaro' by near infrared spectroscopy," *Journal of Food Engineering*, vol.112, no.3, pp.218–226, 2012.
- [8] C.S. Nandi, B. Tudu, and C. Koley, "Computer vision based mango fruit grading system," *International Conference on Innovative Engineering Technologies (ICIET 2014)*, pp.28–29, 2014.
- [9] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geos. Remote Sens.*, vol.56, no.11, pp.6712–6722, 2018.
- [10] D. Guan, X. Tang, L. Wang, and J. Zhang, "Extreme learning machine with superpixel-guided composite kernels for SAR image classification," *IEICE Trans. Inf. & Syst.*, vol.E101-D, no.6, pp.1703–1706, June 2018.
- [11] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar, "Panoptic segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.9396–9405, 2019.
- [12] W. Zhou, C. Wu, Y. Gao, and X. Yu, "Automatic optic disc boundary extraction based on saliency object detection and modified local intensity clustering model in retinal images," *IEICE Trans. Fundamentals*, vol.E100-A no.9, pp.2069–2072, Sept. 2017.
- [13] X. Geng, "Label distribution learning," *IEEE Trans. Knowl. Data Eng.*, vol.28, no.7, pp.1734–1748, 2016.
- [14] B.-B. Gao, C. Xing, C.-W. Xie, J. Wu, and X. Geng, "Deep label distribution learning with label ambiguity," *IEEE Trans. Image Process.*, vol.26, no.6, pp.2825–2838, 2017.
- [15] M. Ling and X. Geng, "Soft video parsing by label distribution learning," *Frontiers of Computer Science*, vol.13, no.2, pp.302–317, 2019.
- [16] D.P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [17] 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. 2020.
- [18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [19] J.-Y. Jiang, S.-C. Tsai, and S.-J. Lee, "FSKNN: Multi-label text categorization based on fuzzy similarity and k nearest neighbors," *Expert Systems with Applications*, vol.39, no.3, pp.2813–2821, 2012.
- [20] X. Lin and X.-W. Chen, "Mr. KNN: Soft relevance for multi-label classification," *The 19th ACM International Conference on Information and Knowledge Management*, pp.349–358, ACM, 2010.
- [21] Y.-K. Li, M.-L. Zhang, and X. Geng, "Leveraging implicit relative labeling-importance information for effective multi-label learning," *2015 IEEE International Conference on Data Mining*, pp.251–260, IEEE, 2015.