Segmentation and recognition of high throughput soybean pods with a supervised edge attention network and synthetic dataset

Si Yang, Lihua Zheng, Minjuan Wang*
CIEE, CAU
yangsi, zhenglh, minjuan@cau.edu.cn

Frédéric Boudon
CIRAD, UMR AGAP
frederic.boudon@cirad.fr

Tingting Wu, Shi Sun
ICS, CAAS
wutingting, sunshi@caas.cn

Peng He
Alibaba group
hepeng.he@antgroup.com

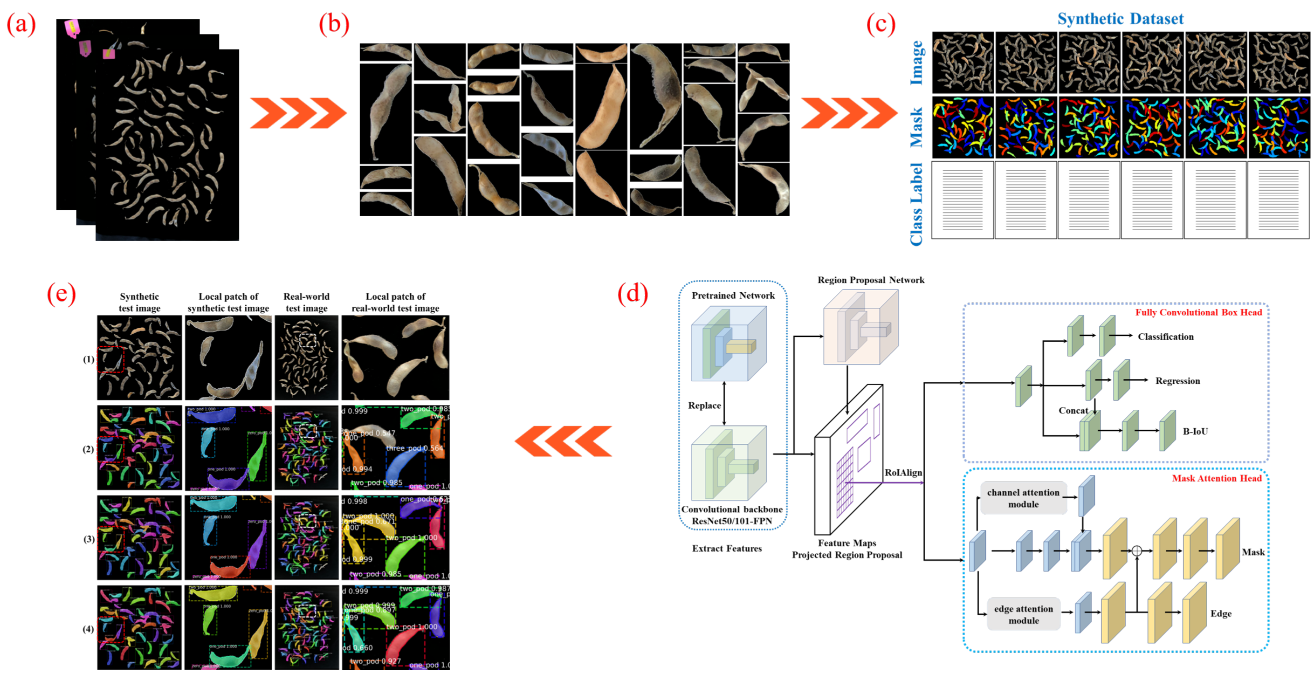

Figure 1: Overview of our method. (a) Real-world high throughput soybean pods dataset. (b) Single soybean pod image pool. (c) Examples of high throughput fine-labeled multi-class soybean pods dataset. (d) The improved SEANet with transfer learning, which is retrained on our synthetic dataset with the pretrained COCO weights. (e) Visualized results of the retrained model raw output (24 epoch, ResNet101-FPN backbone): (1) Example of synthetic test image and real-world test image, (2) Mask R-CNN, (3) SEANet, (4) Our method which improved the loss function of classification based on SEANet.

Over the past few years, the use of deep learning in crop phenotyping has grown exponentially [1, 7, 10]. However, deep learning requires a lot of labeled data to train an accurate algorithm with strong generalization ability [2, 5, 6, 16]. A synthetic images method is capable of generating almost unlimited amounts of labeled data with various representation under many conditions, however, it is difficult to achieve by applying image augmentation techniques on images captured from the real world [15]. Learning from synthetic images is one way to reduce the cost of manual annotation. Danielczuk et al. [4] trained an instance segmentation network on their generated category-agnostic depth image dataset. Yang et al.[17] mainly focus on one-category instance segmentation dataset generation. The studies of Toda et al. [15] demonstrate that synthetic images dataset, is sufficient for training an instance segmentation model to segment real-world high throughput barley seeds images.

To facilitate high throughput soybean pods precise phenotype data extraction, we modified SEANet by introducing the focal loss function [11] to better detect actual boundary of pods and to improve the accurate seed-per-pod estimation. Because of the large amount of labeled image data required to train a deep learning based instance segmentation model, we propose to train such model with synthetic high throughput soybean pods images prepared using our novel synthetic image approach. This approach can synthesize numerous multi-class soybean pods images and labeled ground truth images pair synchronously. The procedure to generate high throughput fine-labeled multi-class soybean pods synthetic images dataset is shown in Figure 2. Some examples of synthetic fine-labeled multi-class soybean pods dataset are shown in Figure 1(c).

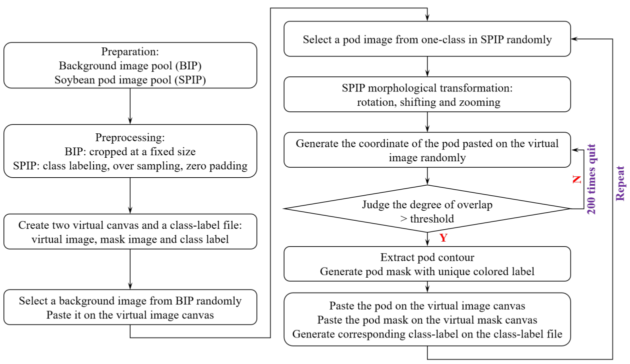

Figure 2: The procedure of high throughput fine-labeled multi-class soybean pods synthetic image pair (raw soybean pods image, labeled mask image) and relevant yaml file (class label file) generation method.

Supervised edge attention network (SEANet), which extend from Mask R-CNN with a fully convolutional box head and a mask head combined with a supervised edge attention module, was proposed by Chen et al.[3], as shown in Figure 1(d). Mask R-CNN [8] consists of a Faster R-CNN [14] object detection algorithm as well as a fully convolution network (FCN) semantic segmentation algorithm [13]. It can be trained by massive manual-labeled images dataset to segment specific categories of object. Usually, the cross entropy and mean square error loss function are commonly used in classification loss functions. In order to increase the accuracy of seed-per-pod estimation, we improved the classification loss function by introducing focal loss function [11] which reshape the loss function to down-weight easy examples and focus on training hard negatives. An evaluation was conducted on two feature extraction architectures (ResNet50/101-FPN [9] backbone). For the purpose of enhancing the diversity of dataset, left-right, up-down, rotation, brightness and Gaussian blur image augmentation methods were employed herein. Transfer learning is used by fine-tuning the pre-trained model weights based on the MS-COCO dataset[12].

The visualized results and the quantitative evaluation metrics of the instance segmentation with improved SEANet model are illustrated herein. The output of the improved SEANet model is a set of class names, bounding boxes coordinates and masks of soybean pods regions. Example of visualized results of synthetic test image and real-world test image are shown in Figure 1(e), showing that high throughput soybean pods are accurately located, segmented and classified by the improved SEANet model purely trained on our synthetic images dataset despite of their size, shape, color, orientation, location, occlusion and so on. And it showed the great performance of our method on pods instance segmentation, keeping effective boundary of pod and high seed-per-pod estimation regardless of the pods vary in shape. The experiment results also demonstrated that the instance segmentation model purely trained by our synthetic images dataset can achieve a promising performance. Table 1 and Table 2 summarized the quantitation evaluation metrics of the model retrained on our synthetic images dataset 24 epochs. The pretrained weight is COCO weights, and the evaluation backbone layers are ResNet 50-FPN and Resnet 101-FPN.

In this paper, segmentation and recognition of high throughput soybean pods based on a supervised edge attention network and synthetic dataset was constructed. The specific work is summarized as follow: (1) A novel method was proposed for automatically and rapidly synthesizing high throughput fine-labeled multi-class soybean pods image dataset. And a hybrid sim/real dataset was constituted for training and estimating our high throughput soybean pods instance segmentation and seed-per-pod estimation model. (2) The modified SEANet model purely trained on our synthetic images dataset can achieve a promising result for high throughput soybean pods instance segmentation and seed-per-pod estimation.

References

[1] Akshay L Chandra, Sai Vikas Desai, Wei Guo, and Vineeth N Balasubramanian. Computer vision with deep learning for plant phenotyping in agriculture: A survey. *arXiv preprint arXiv:2006.11391*, 2020. 1

[2] Ayan Chaudhury, Frédéric Boudon, and Christophe Godin. 3d plant phenotyping: All you need is labelled point cloud data. In *European Conference on Computer Vision*, pages 244–260. Springer, 2020. 1

[3] Xier Chen, Yanchao Lian, Licheng Jiao, Haoran Wang, YanJie Gao, and Shi Lingling. Supervised edge attention network for accurate image instance segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*, pages 617–631. Springer, 2020. 2

[4] Michael Danielczuk, Matthew Matl, Saurabh Gupta, Andrew Li, Andrew Lee, Jeffrey Mahler, and Ken Goldberg. Segmenting unknown 3d objects from real depth images using mask r-cnn trained on synthetic data. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 7283–7290. IEEE, 2019. 1

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1

[6] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016. 1

[7] Reem Ibrahim Hasan, Suhaila Mohd Yusuf, and Laith Alzubaidi. Review of the state of the art of deep learning for plant diseases: a broad analysis and discussion. *Plants*, 9(10):1302, 2020. 1

[8] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 2

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[10] Yu Jiang and Changying Li. Convolutional neural networks for image-based high-throughput plant phenotyping: a review. *Plant Phenomics*, 2020, 2020. 1

[11] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 1, 2

[12] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 2

[13] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 2

[14] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015. 2

[15] Yosuke Toda, Fumio Okura, Jun Ito, Satoshi Okada, Toshinori Kinoshita, Hiroyuki Tsuji, and Daisuke Saisho. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. *Communications biology*, 3(1):1–12, 2020. 1

[16] Si Yang, Lihua Zheng, Wanlin Gao, Bingbing Wang, Xia Hao, Jiaqi Mi, and Minjuan Wang. An efficient processing approach for colored point cloud-based high-throughput seedling phenotyping. *Remote Sensing*, 12(10):1540, 2020. 1

[17] Si Yang, Lihua Zheng, Peng He, Tingting Wu, Shi Sun, and Minjuan Wang. High-throughput soybean seeds phenotyping with convolutional neural networks and transfer learning. *Plant Methods*, 17(1):1–17, 2021. 1

Table 1: The quantitation evaluation metrics of different methods retrained on our synthetic images dataset with ResNet50-FPN backbone layer.

Method	Mask R-CNN		SEANet		Our method	
Test dataset	Synthetic	Real-world	Synthetic	Real-world	Synthetic	Real-world
Recall@[.5:.95]	0.796	0.435	0.888	0.461	0.889	**0.477**
mAP_{50}	0.992	0.603	0.998	0.640	0.998	**0.662**
mAP_{75}	0.992	0.401	0.994	0.434	0.996	**0.462**
mAP@[.5:.95]	0.863	0.342	0.873	0.360	0.875	**0.375**

Table 2: The quantitation evaluation metrics of different methods retrained on our synthetic images dataset with ResNet101-FPN backbone layer

Method	Mask R-CNN		SEANet		Our method	
Test dataset	Synthetic	Real-world	Synthetic	Real-world	Synthetic	Real-world
Recall@[.5:.95]	0.884	0.437	0.888	0.463	0.888	**0.492**
mAP_{50}	0.996	0.610	0.996	0.640	0.994	**0.675**
mAP_{75}	0.992	0.408	0.994	0.448	0.994	**0.479**
mAP@[.5:.95]	0.865	0.345	0.873	0.365	0.873	**0.389**