Deep Learning-Based Segmentation of Early-Stage In-Bag Rice Root for Its Architecture Analysis

1 **Kai Zhu[1], Liang Gong[1*], Chenghui Lin[1], Tao Wang[1], Ke Lin[1], Da Jing Gu, Chengliang Liu[1*],**
2 **Third Author[1,2]**

3 [1]School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

4 [2]ParisTech Elite Institute of Technology, Shanghai Jiao Tong University, Shanghai 200240, China

5 *** Correspondence:**
6 Chengliang Liu   Liang Gong
7 profchlliu@163.com   gongliang_mi@sjtu.edu.cn

10 **Abstract**

11 The root architecture parameters are important to the study of plant growth state and the
12 segmentation of plant roots is the key to the measurement of these parameters. Most existing methods
13 use the threshold calculated by different algorithms to segment the roots in a grayscale image, which
14 requires a low noise background. We designed an automatic equipment to record the roots images of
15 rice seedlings planted in transparent bags. Those root images contain strong noise and it makes
16 existing methods invalid in our circumstances. In order to solve the segmentation problem of rice
17 roots under strong noise, we proposed a convolutional neural network based on U-Net and SE-
18 ResNet. The root images were preprocessed and cropped into small patches to fit CNN input
19 requirements. Experiments have showed that our method performs effectively in pixel-level
20 segmentation of rice seedling roots that contain tiny lateral roots. Our method achieves a pixel
21 accuracy of 99.5% and intersection over union (IoU) of 87.4%. This method provides a new
22 approach to automatic and fast pixel-level root segmentation, which is of great importance for the
23 attribute analysis of root morphology.

24 **1   Introduction**

25 The root system is the key organ of a plant to extract nutrients and water from the soil (Aroca, Porcel,
26 & Ruiz-Lozano, 2012). It contains the growth information and reveals the health state of the plant.
27 The research of plant root systems is very significant in biological field (Zobel, Kinraide, Baligar, &
28 Soil, 2007), as it is an important way to improve grain production. Root morphological attributes are
29 useful while evaluating the tolerance of plants to the fluctuation of growth conditions (Sekulska-
30 Nalewajko & Goclawski, 2009). Determining the parameters of root morphological characteristics
31 and analyzing root architecture are the crucial steps of root morphological attributes analysis.

32 However, the root features are hard to extract manually. Fortunately, some auto-analysis methods of
33 digital images have made this procedure more efficient (Andrés et al., 2008). The root system
34 segmentation is the foundation of most root morphological attributes analysis. Generally, the
35 segmentation of root objects means recognizing the root from the background in images (Strack,

36 2001). A Large number of segmentation methods have been proposed. (Y. Chen & Zhou, 2010) used
37 the Otsu method to segment the root images after median filtering, which is an image segmentation
38 algorithm based on dynamic threshold. (Gocławski, Sekulska-Nalewajko, Gajewska, & Wielanek,
39 2009) proposed a new method based on the color features of the wheat seedlings' roots. The image
40 was transformed into HCI from RGB color space firstly. Then threshold of component images were
41 selected. The transformation of color space is quite useful. (Sekulska-Nalewajko & Goclawski, 2009)
42 implemented the segmentation after the image had been transformed into HSI color space.

43 Although various methods have been used for root segmentation, almost all of them are limited to a
44 certain kind of root images under specific circumstances (Gocławski et al., 2009). Also, those
45 methods preferred simplex roots and oversimplified background to ensure the high contrast between
46 the root and background, which is the critical factor for the segmentation methods based on
47 threshold. Generally, if the root maintains its primitive condition without being washed or stained,
48 the collected images are very likely to have a low contrast so that the methods based on threshold
49 will be invalid. Meanwhile, the rice root system is composed of several tissues and there are different
50 root types, which form a complex structure (Rebouillat et al., 2009). That increases the segmentation
51 difficulty of the existing methods. The root images used in this paper are from the rice seedlings that
52 were planted in the transparent bags. Those images contain the reflective light and water drops,
53 which interfere the segmentation. As deep learning has achieved unparalleled results in computer
54 vision tasks, particularly the convolutional neural networks (CNN) inspired by the organization of the
55 visual field (Hubel & Wiesel, 1968), the accuracy of object detection and semantic segmentation has
56 been remarkably improved. Apparently, CNN can be used successfully in the life sciences (Zhou &
57 Troyanskaya, 2015), though the status might be more complex. In fact, the deep learning has already
58 been used for plant root analysis. (Pound et al., 2017) proposed a CNN model to classify the sliding
59 window over the root images. The wheat root tips are identified and localized after the whole root
60 image is scanned by the sliding window. (Ke et al.) used a CNN model for cucumber powdery
61 mildew recognition. (Yao, Zhang, & Xu, 2016) applied a CNN network for retinal blood vessel
62 segmentation with 23*23 image patches extracted from blood vessel images.

63 To solve the difficulties in root segmentation brought by the complex structure of rice root and the
64 strong noise in the images, a deep learning model of semantic segmentation is proposed. The
65 proposed CNN model uses the processed root images as the input and the mask images as the output.
66 The results show that our method performed quite well on rice root images with strong noise.

67 **2 MATERIALS AND MOTHODS**

68 **2.1 Data Collection**

69 In order to acquire the root image while keeping the roots' primitive shape, the rice seedlings should
70 not be moved during the processes. Some experiments use the MRI or X-ray to obtain roots under the
71 soil, although the equipment is quite expensive. Another solution is planting the rice seedlings in the
72 transparent medium to facilitate root image acquisition. But there has been little research about that,
73 the result of (Tracy, Black, Roberts, & Mooney, 2010) shows that the properties of medium had a
74 significant effect on root weight and root configuration. (Iyer-Pascuzzi et al., 2010) grew the rice
75 seedlings in the Gelzan CM agar to investigate the root traits which can distinguish rice genotypes.
76 This method is quite outstanding for obtaining root images except the demand for the expensive
77 equipment. However, the quantity and diversity of the image samples are the factors influencing the
78 capability of a CNN model (Barbedo, 2018). In this paper, we use transparent bags filled with
79 nutrient solution to cultivate the rice seedlings and the experiments have shown that the root can

80   cling to the capillary paper in these transparent bags. Figure 1 shows the captured root image of the
81   early-stage in-bag rice seedlings with the roots longer than 10 cm.

82   A multi-angle camera layout is adopted to improve the efficiency of image capturing and a light
83   source is installed on the side along with a bottom light to minimize the reflection. To avoid the
84   interference of ambient light, the imaging equipment is mounted in a darkroom. As shown in Figure
85   2, four cameras are installed in different angles to capture four root images each time. The shell of
86   the darkroom is made of aluminum-profile frame wrapped with metal skin to cut off the ambient
87   light. There is a conveyor belt and a window for transporting rice seedlings into this device, and the
88   window will be closed when the cameras start working. This procedure is controlled by a PLC that
89   receives the signals generated by the sensors mounted on the equipment. We used the CCD industrial
90   cameras with a resolution of 4608×3288 pixels, which will ensure the detection of the lateral roots.

91   **2.2   Image Process**

92   As shown in Figure 1, the original images contain part of the background and it is uncorrelated data.
93   The rice roots stretch like the thin strip, which is similar to the periphery of the background in
94   morphology. This will be difficult for the network to achieve precise segmentation if the original
95   images are directly used to train the CNN. We trimmed the useless part of pictures off and separated
96   the root system of different seedlings as shown in Figure 3(a). The mask images of the root system
97   were made manually with the help of Photoshop (Adobe Inc., USA). Usually, the size of a root image
98   with single seedling is larger than 500×2000 pixels, which is too large to be used as the input data of
99   a CNN. However, the down-sampling process brings the loss of information of small scale.
100   Particularly, in this experiment, the lateral root is so tiny that it just occupies about 3 pixels in width.
101   The lateral roots would be lost after the down-sampling. To solve this problem, we used a sliding
102   window with the size of 100×100 pixels to scan root images and each patch of scanning results
103   together with its mask image is considered as a sample for the CNN training.

104   Figure 3(b) is the mask of a root image and it is obvious that the white pixels standing for roots just
105   take up a very small proportion in the image. Generally, the severe class imbalance will cause several
106   problems (Lin et al., 2017). For example, the training will be inefficient. The model will degenerate
107   because it tends to predict pixels to be negative as most pixels of the training images are negative. To
108   avoid this condition, we set a threshold α and mark the proportion of white pixel in the sliding
109   window as γ. The stride of the sliding window is set as half of the window size if γ is bigger than α,
110   otherwise the stride is set as the size of the window. Specially, if $\gamma = 0$, the stride is set bigger than
111   the size of sliding window. These tricks can help to reduce the imbalance of pixel class by increasing
112   the proportion of root pixels. Figure 3(c) shows the stride of sliding window with different γ.

113   **2.3   CNN Network Design**

114   Many CNNs have achieved pixel-level segmentation to deal with the segmentation task for detailed
115   patterns. The skip-architecture was proposed to accomplish detailed segmentation (Long, Shelhamer,
116   & Darrell, 2015). It concatenated high-level features in encoding layers and appearance features in
117   decoding layers, which is proved to be effective on natural images (Hao, Yang, Liu, Mo, & Guo,
118   2017). The U-Net is a kind of CNN using the skip-architecture and has achieved outstanding
119   performance on biomedical image segmentation of different applications (Ronneberger, Fischer, &
120   Brox, 2015). After the down-sampling and up-sampling process, output images have the same size
121   with input images and the roots are segmented. These architectures help the U-Net combine the
122   information of different scales and enhance the precision of edges. As the characteristic of root
123   images are quite similar to the biomedical images, we chose the U-Net as the backbone of our CNN

124   model. In addition, the U-Net performs well on small dataset, which is useful as the obtention and
125   annotation of root images are difficult.

126   To improve the performance of our model, we embedded the ResNet module and squeeze-and-
127   excitation (SE) block into our model. The ResNet module is proposed to train a deeper neural
128   network (He, Zhang, Ren, & Sun, 2016), as the trouble of gradients vanishing/exploding is a big
129   obstacle when training a deeper neural networks. The SE-block is trained to recalibrate the response
130   of different channel-wise features and find the interdependencies between channels (Hu, Shen,
131   Albanie, Sun, & Wu, 2017). More detailed, the SE-block combined with the ResNet module called
132   SE-ResNet module, was used to replace the convolution layer in U-Net. Figure 4 shows the structure
133   of ResNet module and SE-ResNet module. The actual structure of SE-ResNet module used in this
134   paper is on the right side of Figure 4. The residual module is consisted of one ReLU layer and two
135   convolution layers. It has been demonstrated that SE-ResNet can significantly improve the
136   performance of CNN with minimal extra computational cost. Meanwhile, the increase of weights can
137   be ignored when compared to the total weights of the CNN. That means the embedding of SE-block
138   will not cause the problem of overfitting.

139   The structure of our CNN model called SE-ResUnet is instructed in Figure 5. The architecture of our
140   model includes a encoding path and a decoding path. Each colored rectangle stands for a module of
141   neural network and has been explained in the figure. The pair of number under the rectangle indicates
142   the size of output data in this layer and the number upon the rectangle indicates the channel quantity
143   of this layer. The encoding path contains 5 neural network blocks. Except the input layer and the
144   max-pooling layer, each block consists of a convolutional layer with a filter size of 3×3 pixels and
145   two SE-ResNet modules mentioned before. Correspondingly, the decoding path has a
146   deconvolutional layer concatenated with the copy of previous layer output, a convolutional layer with
147   a filter size of 3×3 pixels and two SE-ResNet modules. The patches were upsampled to 128×128
148   before training. During the encoding, the number of feature map increases from 1 to 256 and the size
149   of image decreases from 128×128 to 8×8 pixels. During the decoding, the number of feature map
150   decreases from 256 to 1 and the size of image increases from 8×8 to 128×128 pixels. To avoid
151   overfitting, we add a dropout layer after each max-pooling layer and deconvolutional layer
152   (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). In summary, our model is quite
153   different with the original U-Net in content, although they are similar in form. The difference
154   guarantees a good result of root segmentation.

155   The output layer of our model is a convolutional layer with a filter size of 1×1 and one channel. The
156   output layer is activated by the sigmoid function. The value of each pixel represents the probability.
157   Normally, the pixel is marked as root if its value is bigger than 0.5. Particularly, we use a threshold to
158   determine whether the pixel belongs root or not. The threshold is chosen to help the metric function
159   perform better on the validation data. The metric function is used to evaluate the result of
160   segmentation.

161   **2.4   Training and Optimization**

162   As illustrated above, we used the sliding window to obtain more than 7000 patches from 30 root
163   system images with large resolution as the dataset for training and validation. To improve the
164   generalization ability of the model, we applied some tricks of data augmentation such as flip.
165   Meanwhile, the annotation images were processed in the same way. In order to ensure the diversity
166   of test data split from the dataset, we divided the images into 10 grades standing for the proportion of
167   root region. In each grade, 10% of the patches were split as the validation set for parameter tuning

168 and the left was split as train set. Another 4 root system images were cut into patches as test set
169 without being balanced or augmented for evaluate the final model performance.

170 Here we used the binary cross-entropy loss(cost) function (H. Chen, Li, Chen, & Tang, 2016) as
171 shown in Equation 1. y_i stands for the ground truth of the pixel and p_i denotes the prediction value
172 calculated by the model. The optimization algorithm we chose to minimize the loss function is the
173 adaptive moment estimator (Adam) (Kingma & Ba, 2014). In Adam, the first and second moments of
174 gradients are used for updating and correcting the current learning rate (Hao et al., 2017). During the
175 training process, the parameters of Adam optimizer were set with initial learning rate = 0.01 and the
176 maximum number of epochs = 75. The other parameters were kept as their default value. The
177 learning rate would decay with the factor of 0.5 if there is no improvement in more than 5 epochs and
178 the minimal learning rate was set as 0. The batch size was 32 and the input images were converted to
179 grayscale with value between 0 and 1.

180
$$\boldsymbol{Loss = \sum_{i=1}^{m} -(y_i\, log(p_i) + (1 - y_i)\, log(1 - p_i))} \tag{1}$$

181 Our model was implemented with Keras, which is a high-level neural network deep learning API
182 written in Python with TensorFlow, CNTK, or Theano backend. We trained our model on a GPU
183 server with the Ubuntu 16.04 operating system and it costs 4 hours to finish the training with one
184 NVIDIA GeForce GTX 1080 GPU.

2.5   Performance Evaluation

186 Here we use the intersection over union (Long et al., 2015) and dice accuracy (Milletari, Navab, &
187 Ahmadi, 2016) as the metrics. The pixel accuracy is also considered. The intersection over
188 union(IoU) can evaluate the model performance effectively and eliminate the interference of
189 overmuch negative sample. The dice accuracy(DSC) indicates the overlap measurement between the
190 ground truth root regions and the segmentation results (prediction) of our model.

191 Those metrics are calculated by Equation 2, 3,4, where the *TP*, *FP* and *FN* denote the true
192 positive(the area which is both predicted and annotated as root area), false positive(the area which is
193 predicted as root area but annotated as background) and false negative(the area which is predicted as
194 background but annotated as root area) measurements. We extracted 10% of the total dataset to
195 calculate the metrics mentioned above.

196
$$\boldsymbol{Acc = \frac{1}{m}\sum_{i=1}^{m} f_i}, \quad \boldsymbol{f_i = \begin{cases} 1 & y_i = p_i \\ 0 & y_i \neq p_i \end{cases}} \tag{2}$$

197
$$\boldsymbol{IoU = \frac{TP}{FP+TP+FN}} \tag{3}$$

198
$$\boldsymbol{DSC = \frac{2TP}{FP+2TP+FN}} \tag{4}$$

3   Result

200 The proposed CNN was employed on our root dataset and the result shows that our model has
201 achieved excellent performance on root segmentation task. We achieve 99.5% on pixel accuracy and

202 87.4% on IoU. Those metrics demonstrated that our model had accomplished a favorable
203 performance in root segmentation task for different evaluation criteria. Our CNN model was
204 developed with Keras using a Tensorflow backend. We trained 75 epochs on one NVIDIA GeForce
205 GTX 1080 GPU.

206 Figure 6 shows the loss value and IoU accuracy of the proposed CNN model used on the root
207 segmentation task during the training process. The loss reduced and tended to be gentle on training
208 dataset, while the IoU increased and tended to be gentle. Due to the large proportion of backgrounds,
209 the accuracy is extremely high during the convergence process. As shown in Figure 6, the loss and
210 IoU tendency of validation dataset is the same with train set, indicating the model was well trained
211 and had no overfitting.

212 We used the trained model to implement the segmentation on a typical root system image and the
213 result is shown in Figure 7. The root system was segmented accurately and both the taproot and
214 lateral root were separated with the background correctly as shown in Figure 7,8. In Figure 7(a), the
215 noise similar to lateral root was removed, demonstrating that our proposed method was robust. As the
216 CNN requires images with size of 128 pixels, we trimmed the whole test root image into patches to
217 fit the CNN input. Each patch was segmented and saved in order, so that they can be stitched into a
218 complete root system image.

219 **4   Discussion**

220 **4.1   Comparison**

221 To illustrate the performance of the proposed model, we reproduced the experiment using the method
222 proposed by (Y. Chen & Zhou, 2010) as a comparison, where the Otsu method was used to segment
223 the root after median filtering. Meanwhile, it is necessary to compare the proposed method with U-
224 Net and investigate the effect of incorporating SE-ResNet.

225 Figure 8 shows the convergence process of loss, IoU and accuracy for U-Net. Compared with the
226 convergence process of our model in Figure 6, the IoU convergence speed is slow than proposed
227 model and final IoU value is also smaller. The experiment result strongly proved the effectiveness of
228 our model.

229 Table 1 shows the values of pixel accuracy, DSC and IoU for the previously mentioned methods,
230 when they are applied to the test set. Obviously, the result of deep learning methods are much better
231 than Otsu method. It is worth noting that metrics of deep learning methods on test set are better than
232 it on either train set or validation set. This phenomenon is understandable as the test set was obtained
233 without being balanced. our method excels than U-net at all three metrics, which shows that our
234 method is more effective.

235 As shown in Figure 9(b), the taproot is segmented correctly and part of lateral roots is distinguished.
236 The continuity of the final image is not influenced by using sub-images as the input of our CNN
237 model, since the sub-images and origin images are similar in structural features. However, due to the
238 low contrast of lateral root and background, most pixels of lateral roots are classed as background. As
239 shown in Figure 9(c)(d), the deep learning method performs much better than the Otsu method. Most
240 pixels of lateral roots are classed correctly. Meticulously, our method performs better in details,
241 especially the end of lateral roots.

242 **4.2   Limitations**

243 Although many methods have been proposed to solve the problem of root segmentation, they are
244 usually implemented in specific circumstances, where simple root and low noise background are
245 demanded. These methods usually segment the root by determining a threshold, which decides the
246 result of segmentation. Meanwhile, some processes have to be implemented to enhance the contrast.
247 However, the materials used here contain much noise and the gray value distribution of lateral roots
248 overlaps with the background, making those methods invalid. Based on the CNN's strong ability of
249 feature extraction, our method can segment the roots successfully.

250 Unfortunately, there are some limitations in our method. Firstly, as shown in Figure 10, there are
251 some mistakes in image annotation. The rice roots are so complex that we can not annotate every
252 lateral root correctly and the slender morphological character of roots makes it hard to divide the
253 boundary of root from background. Generally, the quality of image annotation determine the
254 performance of CNN. As shown in Figure 10, not all lateral roots are segmented. These problems are
255 caused by the errors of image annotation. Figure 11 shows the worst result, the lateral roots are too
256 many to be segmented correctly.

257 To decrease the difficulty of making image annotation, image enhancement might be a choice to help
258 improve segmenting the root correctly, for which it would increase the contrast between the root and
259 background. (Koller, Gerig, Szekely, & Dettwiler, 1995) proposed a multiscale based image
260 enhancement method to segment vascular image. Equally, applying these methods to the root images
261 may help improve the performance of CNN model.

262 ### 4.3 Conclusion

263 In this paper, we proposed a deep learning-based method to solve the problem of root segmentation.
264 An automatic equipment was designed to acquire the images of rice roots cultivated in transparent
265 bags. After the preprocessing, the root images were segmented by our CNN model.

266 Our SE-ResUnet model can segment the root images effectively. We achieved 96.2% on pixel
267 accuracy and 87.4% on IoU. It is of great improvement when compared to the Otsu method. Our
268 method achieved excellent performance on root segmentation, especially on the lateral root
269 segmentation and noise elimination. The experiments show that our method can be used to realize the
270 automatic root segmentation and it is important for the study of phenotypic parameters.

271 **Conflict of Interest**

274 **Author Contributions**

275 L.G., K.Z. conducted mathematical modeling and article writing. D.G. and T.W. helped building the
276 data collection equipment. K.L. and C.L. conducted the experimental verification.

277 **Funding**

280 **Acknowledgments**

The useful help and advice given by Linlizi Wu in sample collection and equipment design are gratefully acknowledged.

Reference

Andrés, C. K., Nagel, K. A., Klaus, P., Ulrich, S., Achim, W., & Hanno, S. J. N. P. (2008). Spatio-temporal quantification of differential growth processes in root growth zones based on a novel combination of image sequence processing and refined concepts describing curvature production. *177*(3), 811-821.

Aroca, R., Porcel, R., & Ruiz-Lozano, J. M. (2012). Regulation of root water uptake under abiotic stress conditions. *J Exp Bot, 63*(1), 43-57. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21914658. doi:10.1093/jxb/err266

Barbedo, J. G. A. J. B. E. (2018). Factors influencing the use of deep learning for plant disease recognition. *172*, 84-91.

Chen, H., Li, Y., Chen, C. L., & Tang, X. (2016). *Learning Deep Representation for Imbalanced Classification.* Paper presented at the Computer Vision & Pattern Recognition.

Chen, Y., & Zhou, X. (2010). *Plant root image processing and analysis based on 2D scanner.* Paper presented at the IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications.

Gocławski, J., Sekulska-Nalewajko, J., Gajewska, E., & Wielanek, M. (2009). An automatic segmentation method for scanned images of wheat root systems with dark discolourations. *International Journal of Applied Mathematics and Computer Science, 19*(4), 679-689. doi:10.2478/v10006-009-0055-x

Hao, D., Yang, G., Liu, F., Mo, Y., & Guo, Y. (2017). Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks.

He, K. M., Zhang, X. Y., Ren, S. Q., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)*, 770-778. Retrieved from <Go to ISI>://WOS:000400012300083. doi:10.1109/Cvpr.2016.90

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2017). Squeeze-and-Excitation Networks.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol, 195*(1), 215-243. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4966457.

Iyer-Pascuzzi, A. S., Symonova, O., Mileyko, Y., Hao, Y., Belcher, H., Harer, J., . . . Benfey, P. N. (2010). Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. *Plant Physiol, 152*(3), 1148-1157. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20107024. doi:10.1104/pp.109.150748

Ke, Lin, Liang, Gong, Yixiang, Huang, . . . Pan. Deep Learning-Based Segmentation and Quantification of Cucumber Powdery Mildew Using Convolutional Neural Network. *Frontiers in Plant Science*.

Kingma, D. P., & Ba, J. J. C. S. (2014). Adam: A Method for Stochastic Optimization.

Koller, T. M., Gerig, G., Szekely, G., & Dettwiler, D. (1995). *Multiscale detection of curvilinear structures in 2-D and 3-D imagedata.* Paper presented at the International Conference on Computer Vision.

Lin, T. Y., Goyal, P., Girshick, R., He, K., Dollar, P. J. I. T. o. P. A., & Intelligence, M. (2017). Focal Loss for Dense Object Detection. *PP*(99), 2999-3007.

Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Milletari, F., Navab, N., & Ahmadi, S. A. (2016). *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation.* Paper presented at the Fourth International Conference on 3d Vision.

Pound, M. P., Atkinson, J. A., Townsend, A. J., Wilson, M. H., Griffiths, M., Jackson, A. S., . . . French, A. P. (2017). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *Gigascience, 6*(10), 1-10. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29020747. doi:10.1093/gigascience/gix083

Rebouillat, J., Dievart, A., Verdeil, J. L., Escoute, J., Giese, G., Breitler, J. C., . . . Périn, C. J. R. (2009). Molecular Genetics of Rice Root Development. *2*(1), 15-34.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation.* Paper presented at the International Conference on Medical Image Computing & Computer-assisted Intervention.

Sekulska-Nalewajko, J., & Goclawski, J. (2009). *Segmentation and geometry identification of white roots in two-dimensional scanner images.* Paper presented at the International Conference on Perspective Technologies and Methods in Mems Design, 2009. Memstech.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. J. J. o. M. L. R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *15*(1), 1929-1958.

Strack, D. J. P. (2001). Root Methods: A Handbook : Edited By A. L. Smit, A. G. Bengough, C. Engels, M. van Noordwijk, S. Pellerin and S. C. van de Geijn, Springer Press, Berlin Heidelberg, 2000, 587 pp. 108 figs., DM 198. ISBN 3-540-66728-8. *57*(1), 143-143.

Tracy, S., Black, C., Roberts, J., & Mooney, S. (2010). *Visualising the effect of compaction on root architecture in soil using X-ray Computed Tomography.* Paper presented at the 19th World Congress of Soil Science, Soil Solutions for a Changing World.

337 Yao, Z., Zhang, Z., & Xu, L. Q. (2016). *Convolutional Neural Network for Retinal Blood Vessel Segmentation.* Paper
338   presented at the International Symposium on Computational Intelligence & Design.
339 Zhou, J., & Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model.
340   *Nat Methods, 12*(10), 931-934. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26301843.
341   doi:10.1038/nmeth.3547
342 Zobel, R. W., Kinraide, T. B., Baligar, V. C. J. P., & Soil. (2007). Fine root diameters can change in response to changes in
343   nutrient concentrations. *297*(1/2), 243-254.

344

345

346

Table 1. Comparison between the different methods

Root	Method	Pixel accuracy	DSC	IoU
	Otsu	0.973	0.655	0.487
1	U-net	0.995	0.916	0.845
	SE-ResUnet	0.996	0.936	0.880
	Otsu	0.971	0.744	0.582
2	U-net	0.994	0.915	0.844
	SE-ResUnet	0.996	0.937	0.882
	Otsu	0.972	0.689	0.525
3	U-net	0.995	0.916	0.846
	SE-ResUnet	0.996	0.936	0.880
	Otsu	0.950	0.640	0.47
4	U-net	0.988	0.898	0.815
	SE-ResUnet	0.991	0.921	0.854

347

348
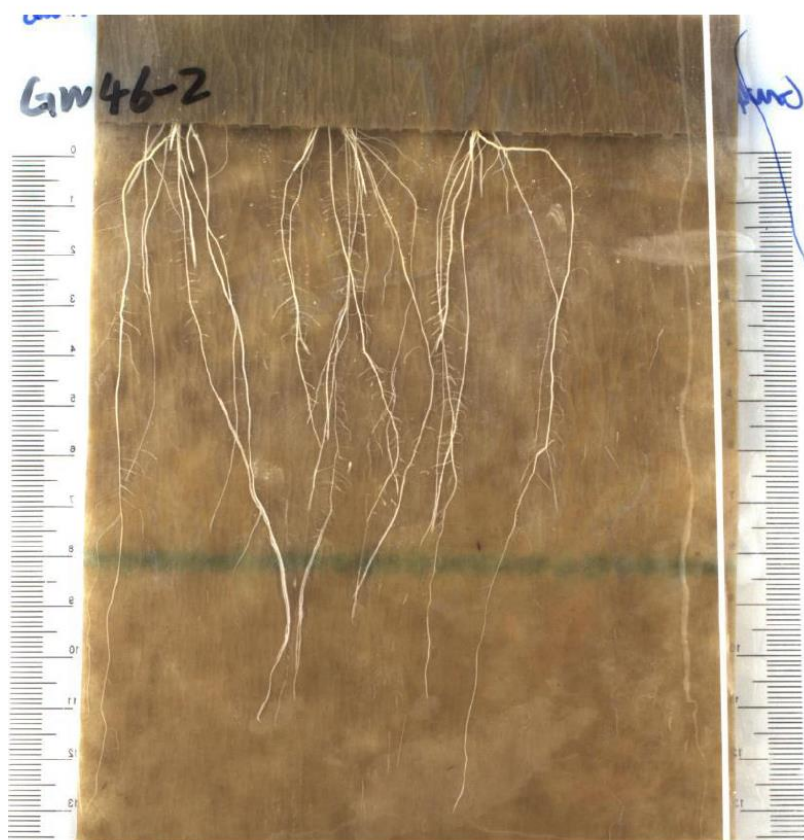
(a) Cultivation of the rice seedlings by transparent bag

(b)The water drops in the bag

(c) The reflective light

349

350   **Figure 1. Captured root system of rice seedlings. The root is in its primitive shape and it clings to the capillary paper**

351
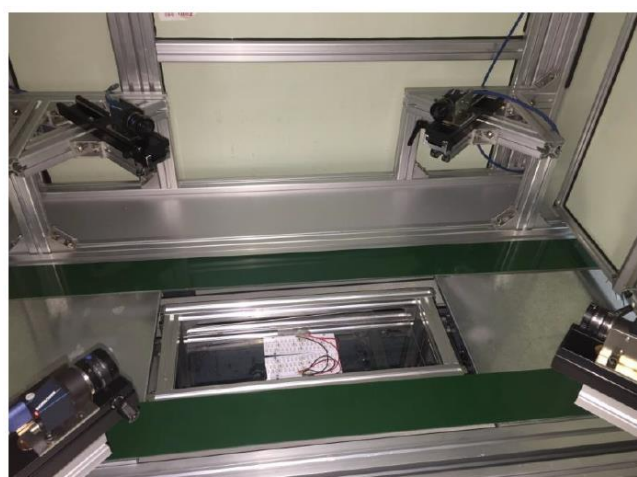
(a)Automatic imaging darkroom
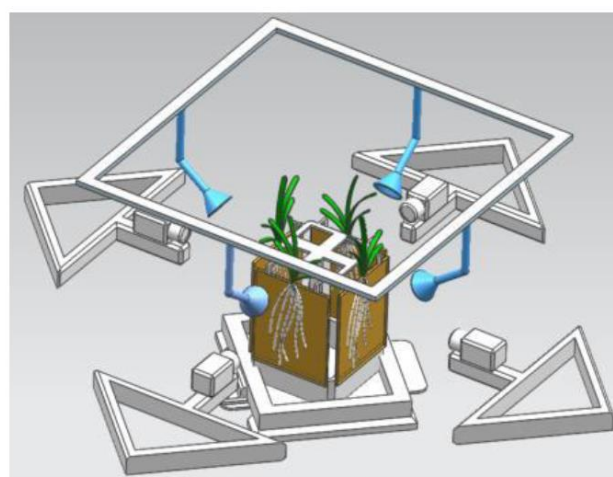
(b) Image acquisition

352

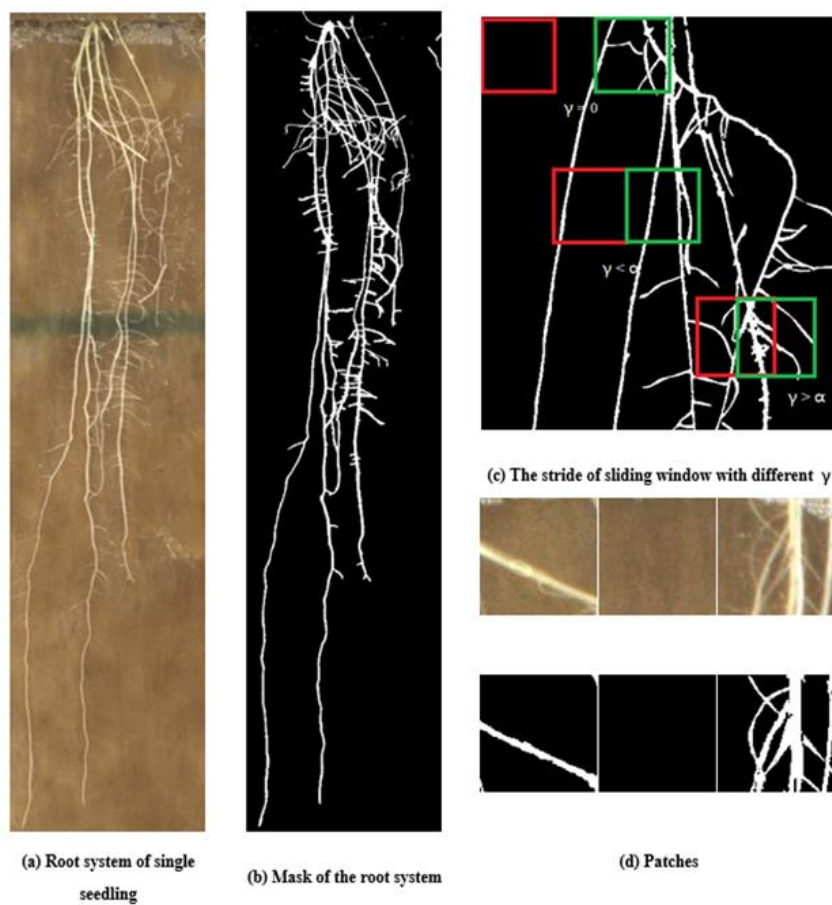353   **Figure 2. The equipment for collect root image**

354

(a) Root system of single seedling

(b) Mask of the root system

(c) The stride of sliding window with different γ

(d) Patches

355
356　　　**Figure 3. Process the root images for CNN training**

357
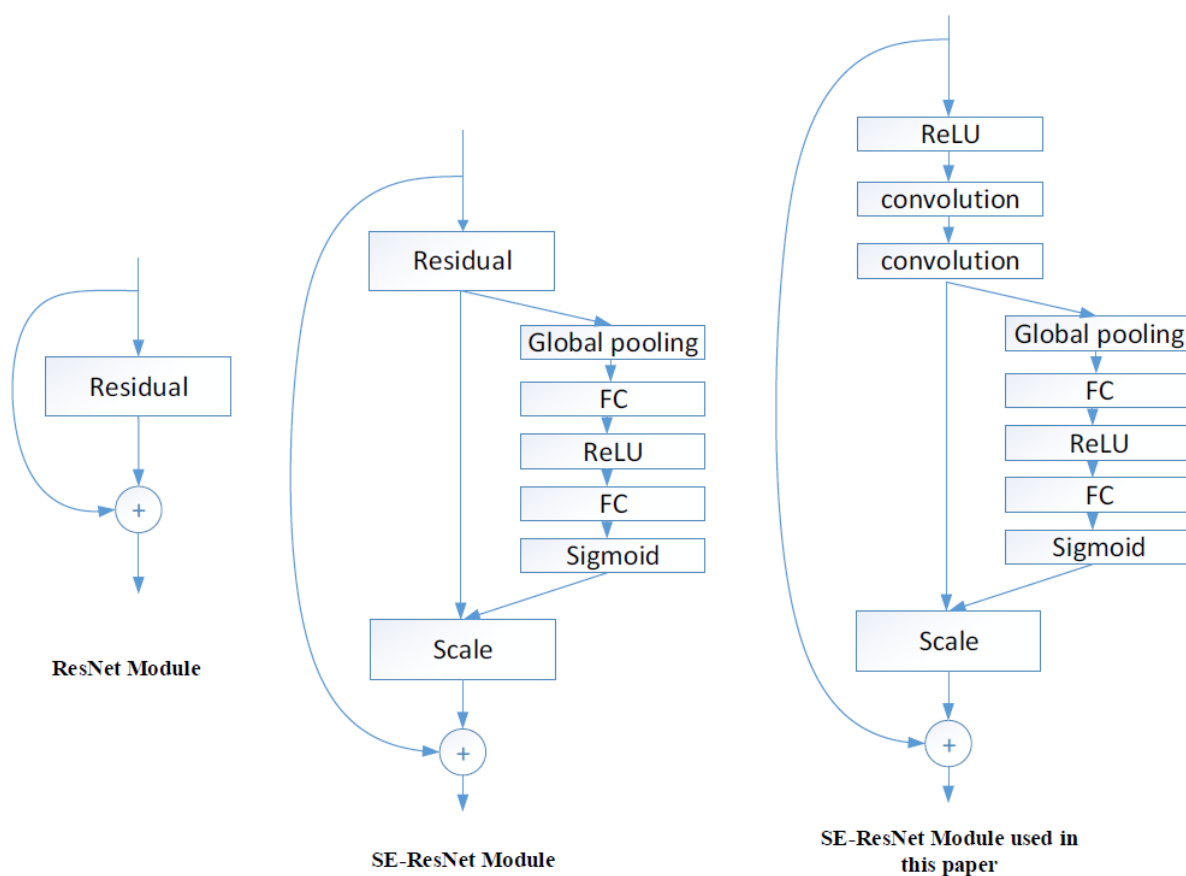
358   **Figure 4. The residual module and SE-ResNet module. The residual is scaled by the small branch net. FC means the full connection layer**

359

360
361
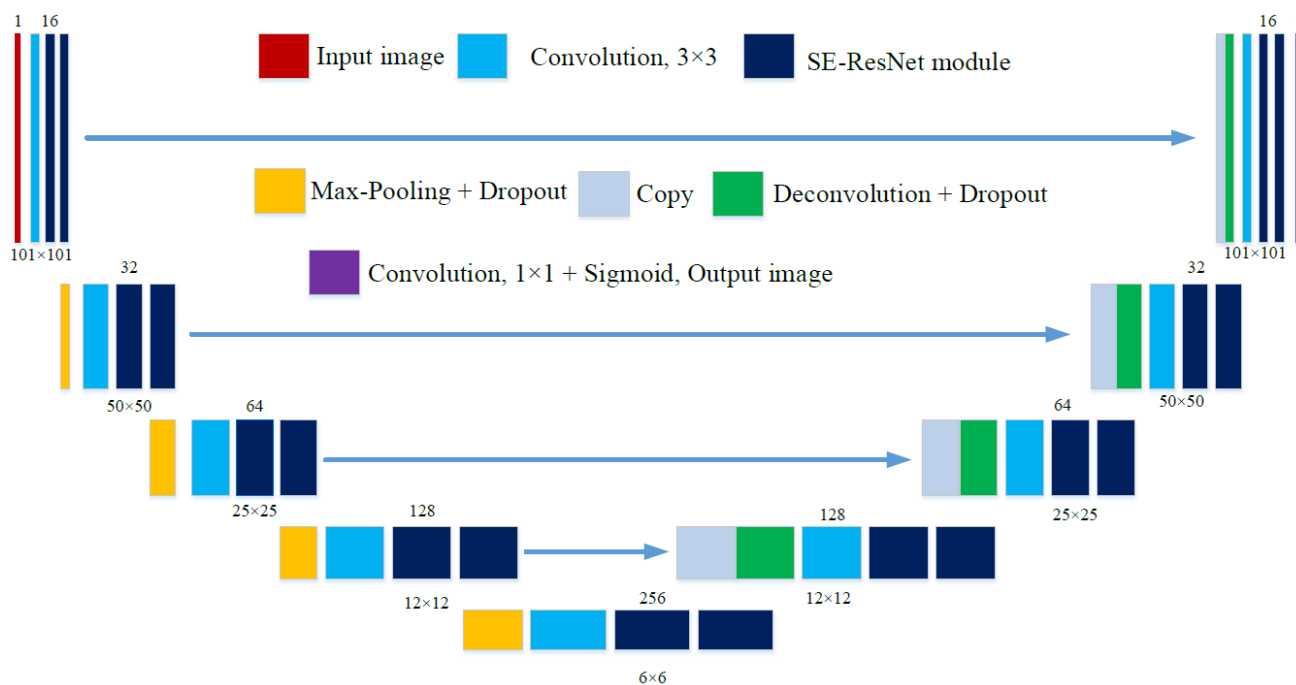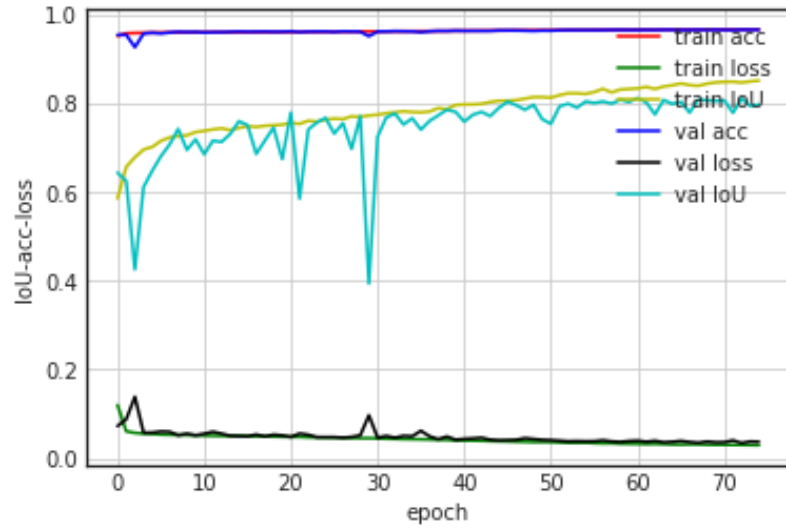
362   **Figure 5. The structure of our SE-ResUnet**
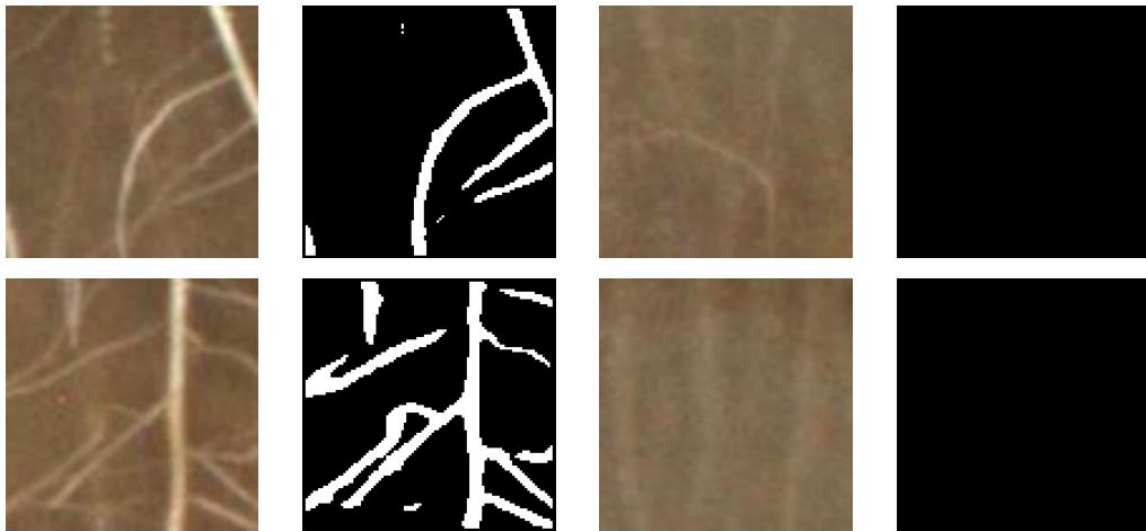
363

364
365

Figure 6. The convergence process on train set and validation set

366

367

(a) The lateral root was well segmented   (b) The root-like noise was romoved

368
369

370
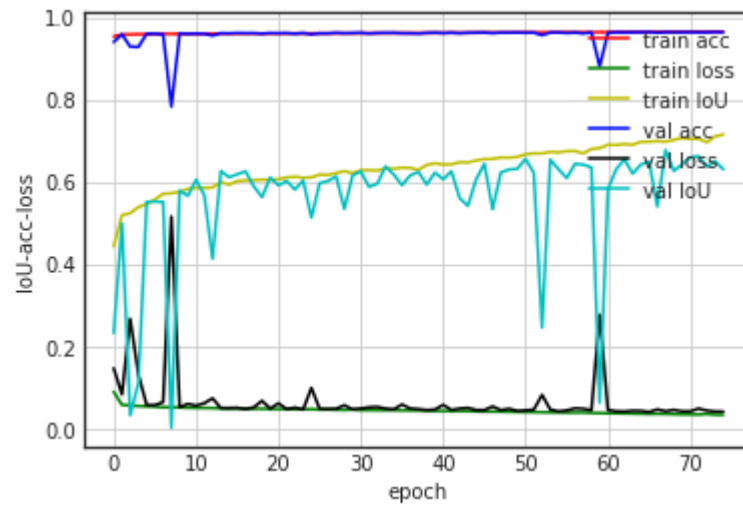
Figure 7. The segmentation experiment of proposed model

371
372

373   Figure 8. The convergence process of U-net

374

375

(a) Original root system image   (b) Segmentation result of   (b) Segmentation result of U-   (d) Segmentation result of our

                                          Otsu   net   model
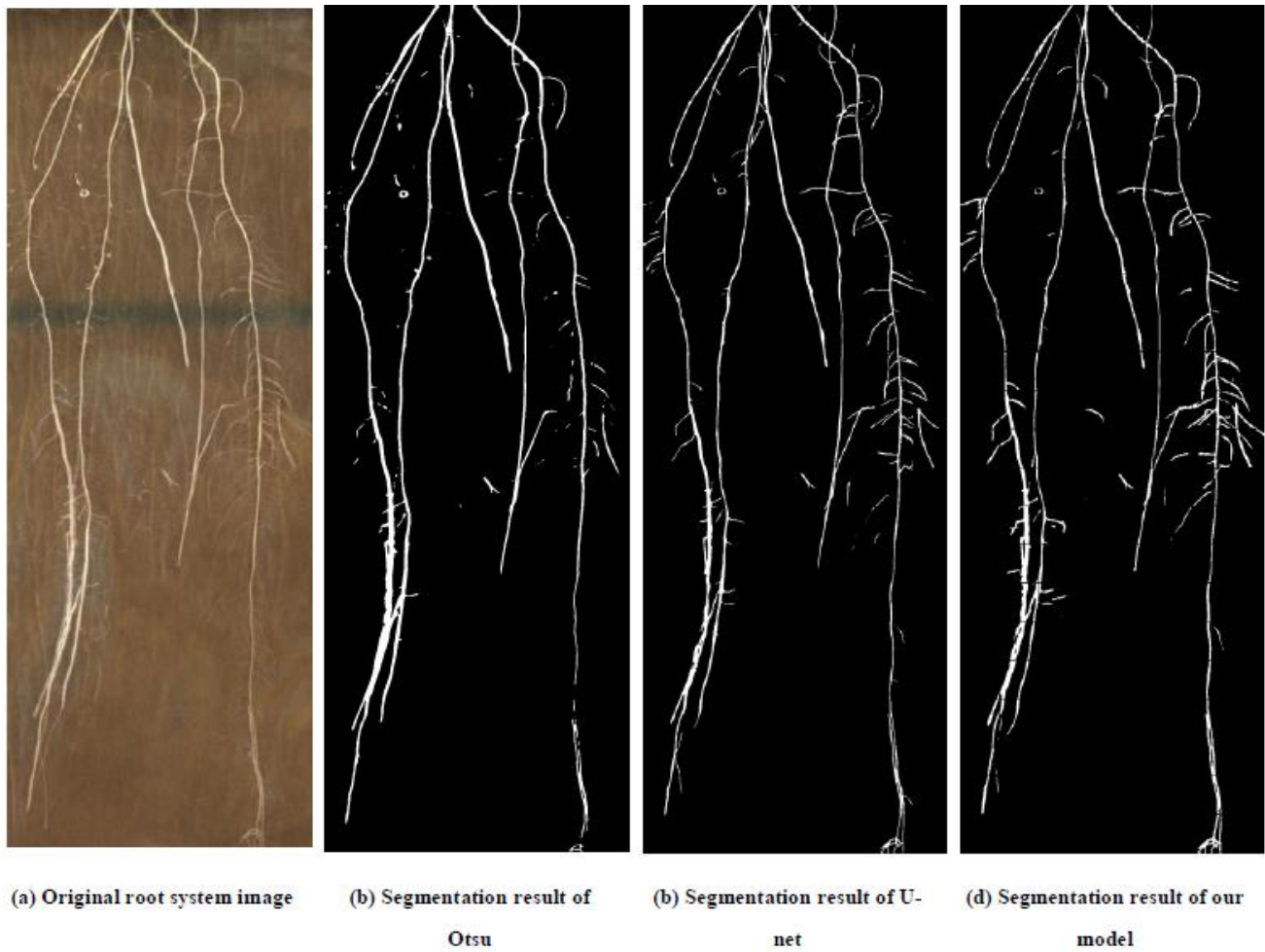
376
377   Figure 9. The segmentation comparison between different methods

378

(a) Some root was not segmented

(b) The noise was segmented as root

(c) The lateral root is too tiny to segment

(d) The junction of roots was not well segmented
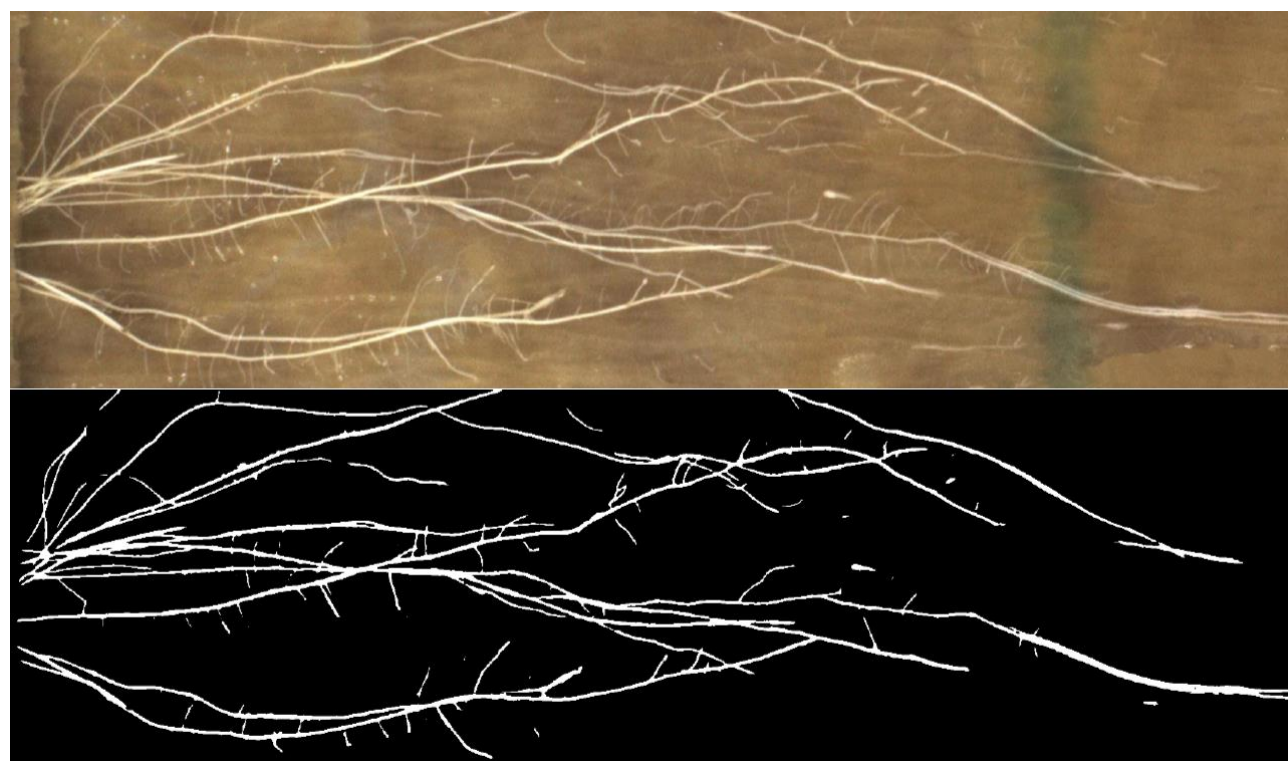
379
380
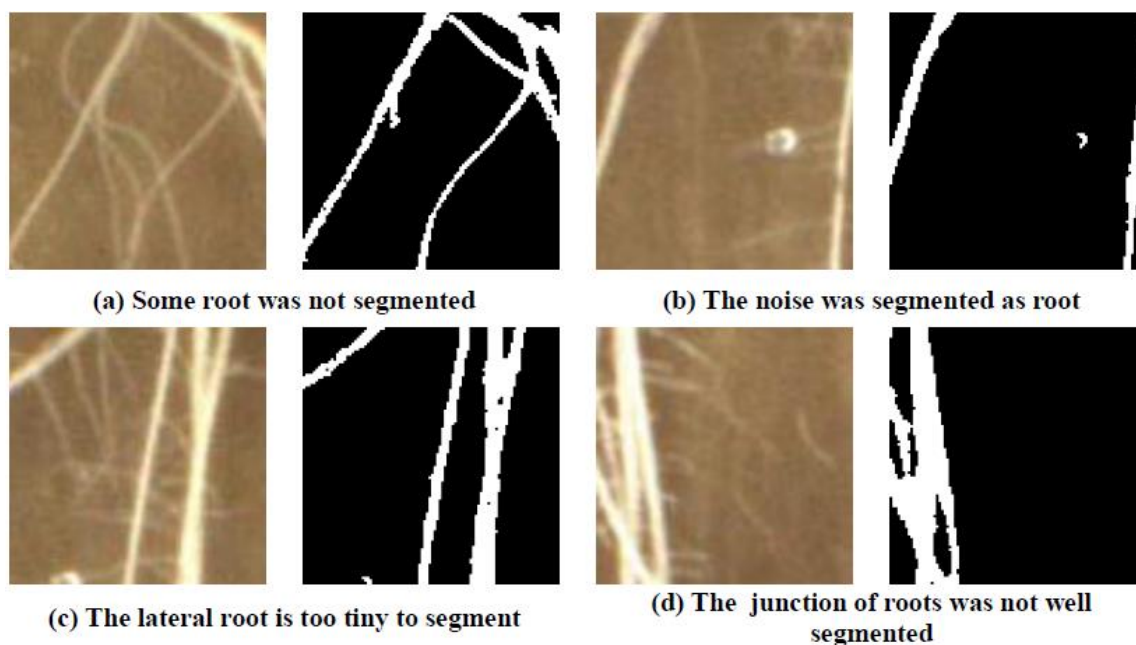
Figure 10.The defects of annotation image(mask)

381

382

383
384

Figure 11. The worst result

385