Plant Phenomics
A SCIENCE PARTNER JOURNAL

SPJ
SCIENCE PARTNER JOURNALS

# Manuscript Template

## FRONT MATTER

### Title

Improved Transformer for Time Series Senescence Root Recognition

**Authors**

Hui Tang[1]†,   Xue Cheng[1]†, Qiushi Yu[1], JiaXi Zhang[1] Nan Wang[1]*, Liantao Liu[2]*

**Affiliations**

[1]College of Mechanical and Electrical Engineering, Hebei Agricultural University, 10 071000, Baoding, China
[2]College of Foreign Languages, Hebei Agricultural University, 071000, Baoding, China
†These authors contributed equally to this work.
*Address correspondence to: liultday@126.com (L.L.); cmwn@163.com (N.W.)

### Abstract

The root is an important organ for plants to obtain nutrients and water, and its phenotypic characteristics are closely related to its functions. Deep learning-based high-throughput in-situ root senescence feature extraction has not yet been published. In light of this, this paper suggests a technique based on the TransFormer neural network for retrieving cotton's in-situ root senescence properties. High-resolution in-situ root pictures with various levels of senescence are the main subject of the investigation. By comparing the semantic segmentation of the root system by general Convolutional Neural Networks and TransFormer Neural Networks, SegFormer-UN (large) achieves the optimal evaluation metrics with mIoU, mRecall, mPrecision and mF1 metrics values of 81.52%, 86.87%, 90.98% and 88.81% respectively. The segmentation results indicate more accurate predictions at the connections of root systems in the segmented images. In contrast to two algorithms for cotton root senescence extraction based on deep learning and image processing, the in-situ root senescence recognition algorithm using the SegFormer-UN model has a parameter count of 5.81M and operates at a fast speed, approximately four minutes per image. It can accurately identify senescence roots in the image. We propose that the SegFormer-UN model can rapidly and non-destructive identify senescence root in in-situ root images, providing important methodological support for efficient crop senescence research.

## MAIN TEXT

### 1. Introduction

The root is composed of fine root and their root hairs, which are the key tissues that absorb soil nutrients and water, directly affecting the functions of the root [1]. The root is in direct contact with the soil, and because of this, its features adapt to environmental changes before those of above-ground plants, giving a clear indication of how crops are growing [2-3]. Root senescence is an important root trait that significantly affects the process of aboveground senescence [4]. And exploring the law of root senescence is an important aspect of revealing the senescence of aboveground plants.

The basis for root senescence research is the acquisition of dynamic in situ root pictures. Finding and identifying roots is challenging due to the soil's shade and lack of transparency. Conventional manual excavation techniques including digging, soil coring, and soil column methods [5–7] cannot be used to research the senescence law of roots. The in situ cultivation method cultivates plants while allowing for the observation of root characteristics in an undisturbed environment [8]. In recent years, the commonly used in-situ cultivation methods include aeroponics, hydroponics, gel culture, and germination paper culture [9–12]. These techniques have some usefulness for researching in-situ roots, but using soil as a culture medium is more challenging, and the quick cultivation cycle makes it impossible to examine the complete plant life cycle. The minirhizotron method is an old in-situ root research technique that uses dirt as a culture medium, however, it has drawbacks such as high costs, small in-situ root pictures, and inefficient collection [13-14]. In contrast, methods such as X-ray Computed Tomography (XCT) [15], magnetic resonance imsenescence (MRI) [16], ground penetrating radar (GPR) [17], and electrical capacitance (EC) [18] have also been reported, but they are plagued by low imsenescence quality [19]. In contrast to earlier methods, the digital imsenescence device method [20,21] uses a high-resolution image acquisition device that has been widely used for root acquisition, and offers the advantages of being less expensive, easier to install, and demonstrating imsenescence. To get in-situ root images of crops throughout their whole growth period, which can be used for studies on root senescence, this study makes use of the RhizoPot, which the team developed in the early stages [22,23].

Root identification is an approach to root phenotypic research, and traditional recognition methods include manual delineation and semi-automatic interactive recognition. The manual representation suffers from issues including low efficiency, a heavy workload, and a high rate of result error [24]. The researcher's visual observations serve as the foundation for the root semi-automatic interactive software, and the auxiliary software identifies the roots, which can increase job productivity and lower labor expenses. However, relying on the arbitrary skill and practical knowledge of manual root separation makes it challenging to achieve high-throughput in-situ root image recognition [25,26].

A deep learning-based semantic segmentation approach speeds root identification research [27]. And by continually extracting the features from the image's region of interest, this technique achieves pixel-by-pixel categorization of the image. Since it was the first semantic segmentation model, FCN has produced positive results [28]. Based on this approach, Shoaib Kamal et al. accomplished weed-crop segmentation [29]. The processing results for FCN are not fine enough, and the relationship between the pixels is not taken into account. To achieve end-to-end training with superior training results than FCN, SegNet utilizes a symmetric encoder-decoder structure based on FCN and adds nonlinear upsampling [30]. SegRoot, a high-throughput root analysis tool created by Tao Wang and colleagues that distinguishes between the soil and the root system, is based on the SegNet model [31]. Originally developed as a segmentation model in the same year, UNet was first applied to the recognition of medical images. With skip connection structures added between levels, UNet is a U-shaped structure based on an encoder-decoder that integrates low-level characteristics and high-level semantic information to increase segmentation accuracy and make it more appropriate for small datasets [32]. As vascular tissue and the crop root system architecture (RSA) are comparable, the model was applied to root identification. For instance, better UNet models were used to build root detection tools like ChronoRoot, RootPainter, RootDetector, etc [33–35]. Pyramid pooling, on which PSPNet is built, aggregates the context data of many receptive fields and enhances the ability to gather global information [36]. Based on enhanced PSPNet, Rui Zhang and colleagues implemented agricultural regional segmentation [37]. To reduce model parameters and increase accuracy, Google's DeepLabV3plus model leverages Depthwise separable convolution and includes an encoder and decoder structure based on V3 [38]. In an earlier study conducted by our team, the enhancement of upsampling based on the V3plus network was documented [39,40]. Convolutional neural networks are the foundation for root feature extraction currently, but deeper networks, more input parameters, and more computation are needed to optimize and improve semantic segmentation models [41,42].

Li et al., through improvements to UNet, achieved pixel accuracy of 0.9917, Intersection Over Union of 0.9548, and F1-score of 95.10 in the task of peanut root recognition [43]. Lu et al. achieved pixel accuracy (PA) of 97.7 in segmenting chili pepper roots, with an average F1-score over 90 [44]. Shen et al., by enhancing deeplabV3plus for cotton root segmentation, reported F1-score, recall, and precision values of 0.9773, 0.9847, and 0.9702, respectively. Compared with the above papers, this study compares the two types of root segmentation models, and the data used are from real root images with the same configuration environment, and the experimental results obtained were relatively accurate.

The machine translation issue was addressed in the initial proposal for the attention mechanism [45]. To process NLP, the Google team created a TransFormer neural network that is entirely based on the attention mechanism block and uses the multi-head self-attention mechanism in place of convolution [46]. The long-distance dependence issue is resolved by the TransFormer neural model, which is capable of performing parallel computations. Since the structure of the NLP input data is different from the structure of the picture data, it adopts a sequence structure. It has been published on how to apply TransFormer neural networks to the field of picture recognition [47]. This methodology achieves picture categorization by converting the image's pixel and location data into a sequence via a patch operation. Hamoud Alshammari used this model to implement the Olive Disease Classification issue for plant disease detection [48]. Based on this model, Jiqing Chen et al. finished classifying maize seeds [49]. The Vision TransFormer neural model can process images, but it has drawbacks, including a lot of parameters and calculations and a finite number of model stacks. Swin TransFormer (Swin) uses the window self-attention module and constructs a hierarchical structure to tackle the challenging problem of self-attention computing. Image data can now be transmitted between windows thanks to the Shifted Windows Multi-Head Self-Attention (SW-MSA) structure [50], broadening the range of applications for TransFormer, including the detection of plant diseases, the identification of weeds, and the detection and identification of animals [51–54].

The visual attention mechanism model has a stronger ability to capture global receptive fields, extract more contextual information, and have good modal fusion ability, which has been widely used in agriculture[55]. Morphological information on color differences within plant tissues can be extracted based on RGB images [56], such as the analysis of sorghum senescence based on the degree of greenness in leaf tissues in the context of nitrogen and water availability [57], and the study of chlorophyll degradation during the senescence process of wheat and madder [58]. Typically, changes in root color indicate the degree of root senescence, and when the root turns black, it is thought to be dead [59]. There aren't many reports on deep learning-based root senescence recognition. To accurately identify in-situ roots, the goal of this paper is to create a root segmentation model based on a visual attention mechanism. To assess this model’s processing effectiveness, we compare it to two widely used models: the convolutional model and the TransFormer model. The tracking and extraction of time-series root image characteristics are also lacking in root phenotypic investigations, and the present dynamic root identification is manually extracted by software during the root growth phase [60], which has issues with high cost and poor accuracy. In this study, full-fertility root pictures were precisely identified using this model by applying varying levels of nitrogen stress, which offers fresh perspectives for the investigation of crop time-series root dynamics properties.

## 2. Materials and Methods

The Crop Growth Regulation Laboratory at Hebei Agricultural University (Baoding City, Hebei Province, 38.85°N, 115.30°E) has an artificial climate chamber where this experiment was conducted. Figure 1 depicts the proposed method’s flow. To begin with, the root system time-series photos are obtained, and the dataset is created and tagged with two data kinds. Data type 1 is root binary annotation and data type 2 is root senescence annotation. Subsequently, the improved model SegFormer-UN proposed in this paper is implemented for root segmentation training based on data type 1, and finally the senescent root recognition training is implemented based on data type 2.

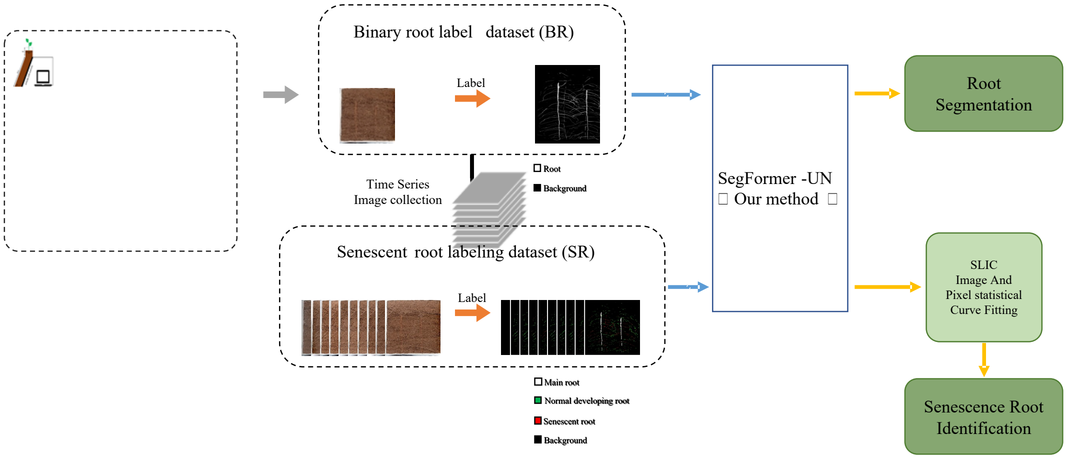

Fig1. The Overall Process of Dataset Acquisition and Model Training

### 2.1 Data acquisition

In-situ root image acquisition was acquired using a high-throughput in-situ root research device (Rhizopot) (Figure 2A), which consists of a root’s growth device, and a flatbed scanner (Epson Perfection Version 39, Suwa, Japan) and a laptop. The root growth device is a trapezoid-like device composed of acrylic plates and attached to a scanner. The front of the device is 20 cm wide and 30 cm high, the side is 8 cm wide and the top protrusion is 6 cm high. The growth device and scanner are at an angle of 60° to the ground and covered with a black straight plate to prevent light exposure. The flatbed scanner was connected to a PC. And multiple scanner devices could be connected on the PC side through Epson’s dedicated API (Figure 2B), and the final output image of the scanning device was in JPEG format.

As the object of study, the cotton plants were cultivated in two levels of nitrogen concentration soil: low nitrogen soil environment 0g N/kg-1; and Normal soil environment (urea N fertilizer) 138 mg N/kg-1 (Zhu et al., 2022). In addition, 138 mg N/kg-1 soil (superphosphate), and 138 mg N/kg-1 soil (potassium chloride) were applied to all soils. The greenhouse was maintained under the following environmental conditions:14/10 h (day/night), 28/25°C (day/night), daytime light intensity 600μmol·m-2·s-1, and relative humidity 45-50%. Other cultivation conditions are as shown in Zhu et al., paper [61].

Image acquisition involved continuous capturing over a period of 100 days from the beginning of the crop growth stage. The pixel size of the image acquisition was 1200 dpi (317.5 mm), with a resolution of 10200×14039 and a RGB image depth of 24 bits. The images were formatted as JPEG. There are 8 sets of root images, each consisting of over 100 images. Due to the unreliable factors during the scanning process, images containing noise and unclear content were excluded, limiting each set to 100 images. In order to complete the training of the model, a hundred images were randomly selected from the 8 sets for annotation. The selected images were divided into training sets, validation sets and test sets with proportions of 70%, 20%, and 10%, respectively. The remaining images were used to assess the actual performance of the model. Adobe Photoshop CC 2020 (Adobe Inc., San Jose, CA, United States) was used for image annotation. The annotation process involved opening the images with Adobe Photoshop, creating a new layer, using the lasso tool to select the roots for annotation, and filling the selected roots with white using the paint bucket tool. This process was repeated until all root annotations were completed, and the soil background was filled with black. The annotation of two types of data is illustrated Figure 2C. Four randomly selected localizations are shown for example diagrams of binary and senescence labeling.

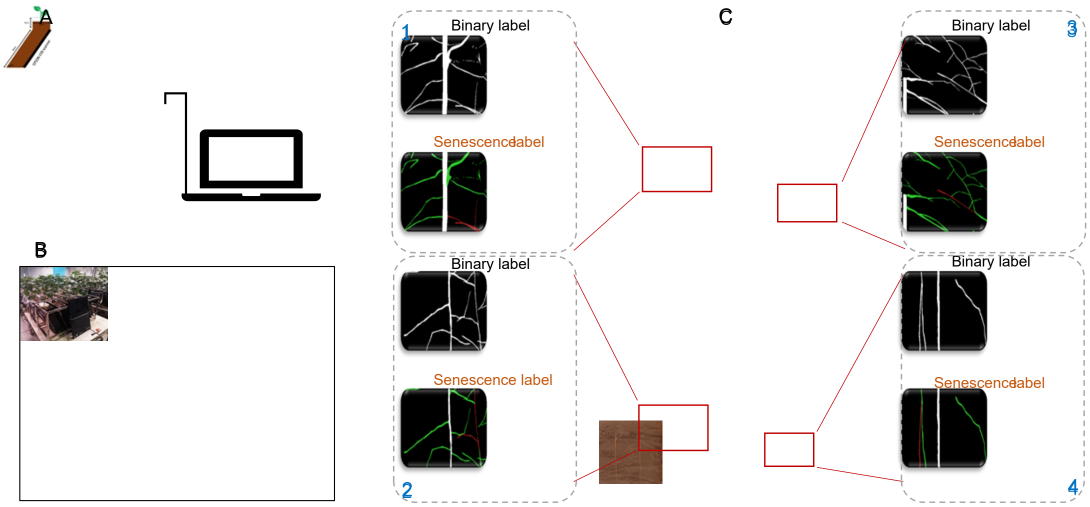

Fig2. Image acquisition device and annotation method, (a) Image acquisition device, (b) Cluster acquisition, (c) Two annotation methods

**Root Binary Annotation Dataset (RB):** To finish training the model, the necessary dataset must be labeled. The root is symbolized in the labeled images by the color white, which has the color value [255,255,255]. The soil background is represented in black with a color value of [0, 0, 0]. The image is in PNG format.  The actual annotated images, as shown in the binary label, took approximately 3 hours for annotation per image. Due to the high resolution of the images in the training set, the original images are cropped into smaller images using windows of size 768×768 dpi. In cases where the original image size is less than 768×768, black pixels are used to fill the gaps. The cropped real images and masks can fit completely after cropping. After cropping 100 large images, a total of 18883 smaller images are obtained. These are divided into three datasets using a ratio of 70%, 20%, and 10%, resulting in a training set of 13216 images, a validation set of 3778 images, and a test set of 1889 images. During model training, to ensure compatibility with the input requirements of the model, the images are resized to 512×512 dpi during the image reading stage.

**Root Senescence Annotation Dataset (RS):** In order to complete the labeling of the root senescence and reduce the labeling time, professional agronomy experts classify actual root systems based on binary images. The labeling categories are divided into four categories: soil as the background is represented in black with the color value of [0,0,0]; normally growing root systems are represented in green with the color value of [0,255, 0]; senescent roots are represented in red with the color value of [255,0,0]; and the main roots are represented in white with the color value of [255,255,255]. The annotation of aging roots requires approximately 1 hour per image. Similarly, to manage the large image size, the entire image is cropped into smaller images with a size of 768×768 dpi. After cropping, the training set consists of 2088 images, the validation set has 300 images, and the test set has 600 images. Additionally, the images are resized to 512×512 dpi during the reading stage.

## 2.2 Model structure

This paper utilizes a U-shaped encoder and decoder structure and is based on the SegFormer model [62]. The encoder structure starts by extracting features through four TransFormer blocks, halving the size of the feature map after each extraction, and saving the output feature map of each block for feature fusion in the decoder. Both the encoder and decoder have four layers and are symmetrical. The decoding process entails stitching the TransFormer block's output feature map, fusing it using Convolution, and then up-sampling to finish downscaling and dimensionalizing the image. The pixel-by-pixel classification of the image is finally finished by the segmentation head, and Figure 3 depicts the model's overall structure. This model employs two distinct backbones (Small and Large), whose depth varies, and whose model layers are deeper for Large than for Small.

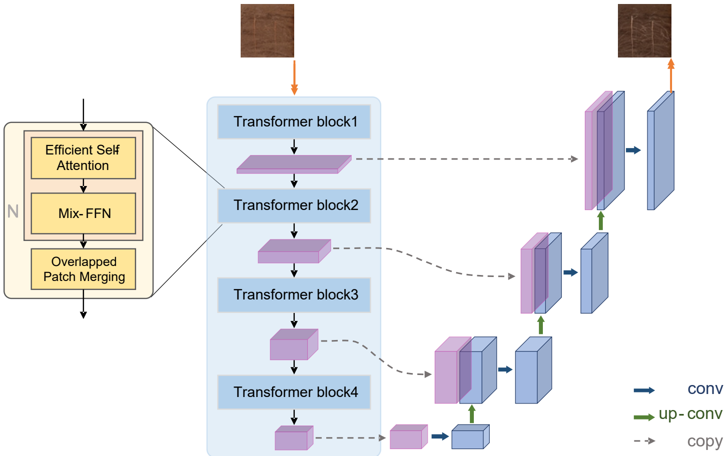

Fig3. The overall structure diagram of the model

### 2.2.1 The TransFormer block of the encoder

In this study, the Overlapped Patch Merging method is used to process the VIT's Patch Merging method, which can guarantee the local continuity of these patches at each stage of feature extraction. The patch interacts during feature extraction by varying patch size (K), stride between two adjacent patches (S), and padding size (P). Setting K=7, S=4, P=3 for the initial picture input and K=3, S=2, P=1 for the subsequent phases will complete this operation. The above operation enables to obtain a convolution-like multi-scale output map in different feature processing stages after a given input image, assuming that the pixel size of the input image is $H \times W \times 3$ and the size of the feature map at different stages is as shown in Equation 1:

$$\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i \quad i \in \{1,2,3,4\} \#(1)$$

$C$ in the formula denotes the image dimension of each stage.

Efficient Self-Attention: The original single attention (SA) module is three times the sequence and input into SA (Q, K, V) for calculation, $Q = ZW_Q$ $\mathrm{K} = ZW_K$ $\mathrm{V} = ZW_V$, $W_Q$ $W_K$ $W_V \in \mathbb{R}^{C \times d}$, Where Q/K/V represents the weight matrix; L represents sequence length, $C$ represents hidden channel size, and $d$ represents output channel size. The final SA calculation is shown in Equation 2:

$$SA = softmax\left(\frac{ZW_Q(ZW_K)^T}{\sqrt{d_K}}\right)(ZW_V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)(V)\#(2)$$

$d_K$ is the dimension of K, which is to prevent the inner product from being too large.

To improve the calculation efficiency, the calculation is simplified by reducing the K series, such as Equations 3 and 4:

$$\hat{K} = Reshape\left(\frac{L}{R}, d \cdot R\right)(K)\#(3)$$

$$K = Linear(d \cdot R, d)(\hat{K})\#(4)$$

$R$ represents the reduction ratio, $\hat{K}$ is to reshape the dimension to $\frac{L}{R} \times (d \cdot R)$, and then through $Linear$, $\hat{K}$ is reduced from $d \cdot R$ dimension to $d$ dimension, and finally $K \in \mathbb{R}^{\frac{L}{R} \times C}$.

Mix-FFN: Reduces the nonlinearity of the function by mixing the feed-forward network (FFN) after the Efficient Self-Attention calculation. The calculation formula is such as 5:

$$x_{out} = MLP(GELU(Dconv_{3 \times 3}(MLP(x_{in})))) + x_{in}\#(5)$$

$x_{in}$ is the input feature map, $Dconv_{3 \times 3}$ is the depthwise convolutions with convolution kernel 3, $GELU$ is the activation function, and $MLP$ is the multi-layer perceptron.

### 2.2.2 Convolution of decoder

Convolution is mainly used to complete the segmentation operation in the decoder. The decoder used in this paper is similar to UNet, which recovers through convolution operations at different scales. Each layer of the decoder is mainly composed of upsampling and convolutional layers for feature map fusion and image size restoration. Finally, the segmentation head is used to achieve pixel by pixel classification to predict the actual root system.

### 2.3 Extraction method of senescence root

Tasks involving senescence root extraction can be completed using the SegFormer-UN (small) suggested in this page. The senescence time series root dataset (RS) is used to train the model, which converges after 100 generations of training and infers the properties of the senescence root time series. This paper uses the superpixel approach (SLIC) for error correction because the shooting environment has an impact on the results of recognition. Senescence images are segmented into pixel blocks using the SLIC method, resulting in blocks of varying forms. To consistently cover non root pixels and protect the root block segmentation outcomes, a black and white prediction image is utilized. Each pixel block's color weight value is calculated, and the color with a higher weight is used to replace the original color block. Calculate the senescence root and normal root pixels of each image once the temporal dataset has been repaired. Then, fit curves to the two-pixel categories, and contrast the results with the actual values. The algorithm is as follows.

**Algorithm 1:** Extraction Method and Analysis of Senescence Root

```
Input: Time series images, $A_i$; The Image array of Binary Root, $A_b$; The image list, $L_i$
Output: The Senescence repair image, $A_s$; Pixel statistics table
Stage I: Senescence Root Extraction
$A_s$ = SegFormer-UN ($A_i$)
Return $A_s$
Stage II: Senescence Root Image Correction
Region size =60 // Superpixel size
Ruler = 20 // Superpixel Smoothness
Iterate = 50 // Iterations
$A_{slic}$ = Superpixel SLIC ($A_s$, Region size, Ruler, Iterate) // Superpixel SLIC Algorithm
$A_l$, class = Image And ($A_{slic}$, $A_b$) // Image And Algorithm
for $i$ = 0 →class do
  if $A_{sr}[A_l == i]$ then
    $A_{sr}[A_l == i]$=0
  End if
end for
Return $A_{sr}$
Stage III: Pixel value statistical and analysis
Pixel value=255
for $i$ = 0 →$L_i$ do
  for $j$ = 0 → Pixel value do
    $D_i[i]$ $[j]$ = Len ($A_s[A_s == j]$) // Calculate the total number of pixels
  end for
end for
$C$ = Polyfit ($D_i$) // Polynomial Curve Fitting
Return $D_i$, $C$
```

Among them, $A_s$ represents the senescence root matrix, $A_s$ represents the SLIC segmentation matrix, class represents the SLIC segmentation category, $A_l$ represents the image and algorithm output results, $A_{sr}$ represents the correction matrix, $D_i$ pixel statistical results, $C$ represents the fitting curve results.

## 2.4 Experimental setup

To compare various semantic segmentation algorithms, a total of seven models are selected based on the usability and reproducibility of the code. The four convolutional neural networks are UNet, SegNet, PSPNet, and DeepLabV3plus, TransFormer neural networks are TransUNet, SwinUNet, and SETR [63-65].

The training and inference of all the above models are based on the Ubuntu 22.10 Operating System, and the configuration of the system is as follows: the processor is Intel i5-12400F (2.5GHz) with 32G RAM, the graphics card model is NVIDIA RTX 3060 with 12GB graphics memory, and a memory bit width of 192 bits. The deep learning framework used for training is Pytorch 11.6.

All models are trained in the same training environment with the same model hyperparameter values to ensure fairness during training. The parameters betas are set to 0.9 and 0.999, the initial learning rate is set to 0.0001, and the learning rate decay strategy uses the cosine annealing learning rate decay strategy in all of the models. They also all adopt the Adaptive Moment Estimation with decoupled weight decay (adamW) optimizer (CosineAnnealingLR). Moreover, the model training spans 100 generations. The appropriate data enhancement is done during the training phase to guarantee the training effect and boost the robustness and generalizability of the model. The common techniques for data augmentation include flipping, cropping, and altering image properties (brightness, contrast, saturation, and hue).

## 2.5 Evaluation of the model

To validate various model performances, the model needs to be evaluated with the help of evaluation metrics. In this paper, the root system and soil background are equivalent to the pixel-by-pixel classification of images, so it is necessary to use the confusion matrix to count the classification results and actual values, to further obtain a variety of classification basis. In this paper, four evaluation metrics are Precision (Formula 6), Recall (Formula 7), and IoU (Formula 8). and F1 (Formula 9), which are calculated as shown below:

$$Precision = \frac{TP}{TP + FP} \times 100\%\#(6)$$

$$Recall = \frac{TP}{TP + FN} \times 100\%\#(7)$$

$$IoU = \frac{TP}{TP + FN + FP} \times 100\%\#(8)$$

$$F1 = 2 \times Precision \times \frac{Recall}{Precision + Recall} \times 100\%\#(9)$$

In the formula, TP stands for the number of root pixels that were accurately predicted to be roots, FP for the number of background pixels that were accurately predicted to be roots, FN for the number of background pixels that were correctly predicted to be roots, and TN for the number of background pixels that were accurately predicted to be backgrounds. IoU can assess how closely classification results for pixels match actual values. F1 is the harmonic average of Precision and Recall, which are used to confirm that the rate of pixel classification is correct. Each of the four evaluation criteria has a value range of 0 to 1.

## 3. **Results**

### 3.1 Segmentation results

This paper compares a total of eight models, including PSPNet, SegNet, UNet, DeeplabV3plus, TransUNet, SwinUNet, SETR, and the method proposed in this paper. Each model is trained using the same configuration for 100 epochs, and all model losses reach convergence, and the model training effect is tested based on the test set to obtain the evaluation indexes. Considering the indicators and segmentation results comprehensively, the optimal model is selected. The numerical comparison results of the model indicators are shown in Table 1. The SegFormer-UN (Small) proposed in this paper uses a lightweight backbone with relatively small FLOPs and Params. The mIoU and mRecall indicators are also higher than other comparative models at 81.06% and 86.29%, respectively. Since mPrecision and mRecall are a pair of contradictory values, mF1 is used to measure the quality of both. The mF1 value of SegFormer-UN (Small) is 88.47%. The best evaluation indicators are obtained when a deeper backbone model SegFormer-UN (Large) is used, with the largest values of mIoU, mRecall, and mF1, which are 81.52%, 86.87%, and 88.81%, respectively. The deeper backbone increases FLOPs and Params, but still lower than that of the other TransFormer neural networks.

Table1-Evaluation Metrics of the model

| Method | Flop(M) | Params(M) | Root IoU (%) | Root Recall (%) | Root Precision (%) | F1(%) |
|---|---|---|---|---|---|---|
| UNent | 786318.75 | 34.53 | 80.87% | 85.46% | 91.75% | 88.33% |
| DeepLabV3plus | 95122.49 | 22.44 | 78.03% | 82.07% | 91.49% | 86.14% |
| PSPnet | 28367.25 | 1.51 | 76.43% | 79.61% | **92.49%** | 84.83% |
| SegNet | 937999.47 | 39.79 | 79.79% | 85.27% | 90.09% | 87.52% |
| SETR | 275440.73 | 85.88 | 75.86% | 81.07% | 88.53% | 84.37% |
| TransUNet | 351034.49 | 91.52 | 79.53% | 84.33% | 90.92% | 87.31% |
| SwinUNet | 136384.82 | 41.34 | 79.99% | 85.07% | 90.73% | 87.67% |
| SegFormer-UN(Small) | 15528.10 | 5.81 | 81.06% | 86.29% | 90.96% | 88.47% |
| SegFormer-UN(Large) | 60972.07 | 23.11 | **81.52%** | **86.87%** | 90.98% | **88.81%** |

The root segmentation results are shown in the Figure 4. Among the aforementioned models, except for SETR, all segmentation models exhibit excellent results in root segmentation. The segmentation of root edges is smooth, and overall recognition is clear, although there are some differences in certain details. However, the SETR model shows more spikiness in the segmentation of root edges, and there are instances of interrupted root recognition.

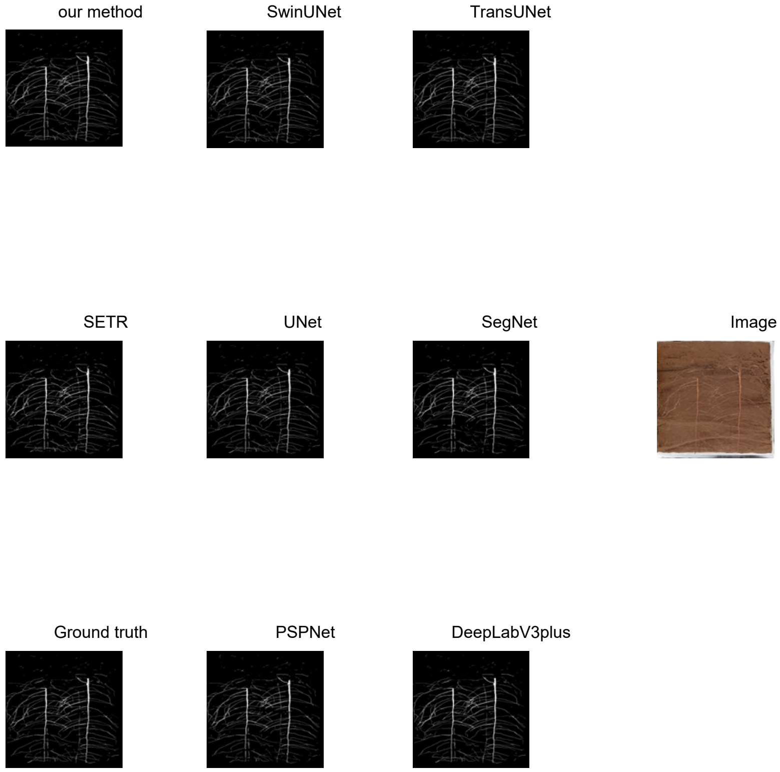

Fig4. Root Segmentation Image

Our improved method handles the boundaries of root systems more effectively (as indicated by the blue box in Figure 5). SegFormer-UN can recognize and connect the low-contrast roots, and recognize more and more complete roots, while other models exhibit discontinuities in root recognition, and some models are unable to identify certain roots. However, when dealing with the issue of significant soil particle occlusion, the results of all the aforementioned models are not ideal (as highlighted by the red box in Figure 4). Due to severe soil particle occlusion, some roots are in a blurred and indistinct state, making it challenging for all segmentation models to accurately identify the roots. However, the model proposed in this study, along with SwinUNet and TransUNet, performs relatively better compared to other models in segmenting obscured roots.

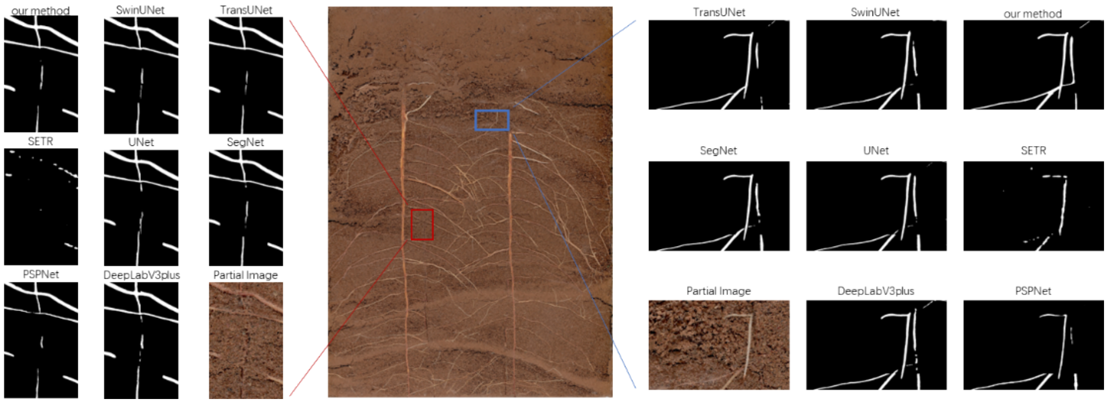

Fig5. Root Segmentation Results

In order to further validate the effectiveness of the improved model, we conduct ablation experiments on a series of methods adopted in this paper, and the comparison results are shown in Table 2. The experimental methods include: improvement of two up-sampling methods sub-pixel (SegFormer-Sub-pixel) and deconvolution (SegFormer-Trans-Conv), UNet decoder (SegFormer-UN), Depthwise separable convolution+UNet decoder (SegFormer-DU), and DeeplabV3plus decoder (SegFormer-DP). The results show that the model structure using the improved up-sampling method alone has decreased compared to the original model evaluation indicators. Among them, the mIoU, mRecall, mPrecision, and mF1 of SegFormer-Trans-Conv have decreased by 79.09%, 84.17%, 90.34%, and 86.98%, respectively; the mIoU, mRecall, mPrecision, and mF1 of SegFormer-Sub-pixel decreased to 78.99%, 84.02%, 90.37%, and 86.90%, respectively. The performance of the model varies after changing the decoder. The SegFormer-DP indicator value decreases, and the required Params also increases. The SegFormer-UN and SegFormer-DU indicators have both improved and reduced model FLOPs and Params. Among them, SegFormer-UN has the greatest improvement, with four indicators of 81.52%, 86.87%, 90.98%, and 88.81% and the FLOPs and Params of model have been slightly reduced. As we continue to increase the depth of the model backbone, the improvement effect becomes more obvious. The four indicator values increase by 1.34%, 1.49%, 0.34%, and 0.99%, respectively, compared to the original model, but these will be accompanied by a doubling of FLOPs and Params.

Table2-Evaluation Metrics of the model

| Method | Flop(M) | Params(M) | Root IoU (%) | Root Recall (%) | Root Precision (%) | F1(%) |
|---|---|---|---|---|---|---|
| SegFormer(Small) | 125096.17 | 6.08 | 80.18% | 85.38% | 90.64% | 87.82% |
| SegFormer-Trans-Conv (Small) | 135087.00 | 6.28 | 79.09% | 84.17% | 90.34% | 86.98% |
| SegFormer-Sub-pixel (Small) | 125096.17 | 6.08 | 78.99% | 84.02% | 90.37% | 86.90% |
| SegFormer -DP (Small) | 38425.59 | 7.81 | 79.54% | 84.57% | 90.61% | 87.33% |
| SegFormer -DU (Small) | 10218.50 | 3.93 | 80.92% | 86.34% | 90.66% | 88.37% |
| SegFormer -UN(Small) | 15528.10 | 5.81 | 81.06% | 86.29% | 90.96% | 88.47% |
| SegFormer -UN(Large) | 60972.07 | 23.11 | **81.52%** | **86.87%** | **90.98%** | **88.81%** |

## 3.2 Extraction of time-series root senescent features

Senescence root color weights can be calculated automatically based on annotated images using the SegFormer-UN model, which leads to better root extraction performance (Processing time and image recognition) than image processing. Based on deep learning extraction methods, GPU can be used to accelerate inference operations and complete a 512× 512 dpi image inference takes about 1 second, and completing a complete image takes about 4 minutes. It is possible to notice that root senescence is increasing by using an improved model to extract the sequence senescence root dataset with ten-day intervals, as illustrated in Figure 6A. Similar to this, although deep learning can swiftly extract roots, there still exist senescent and normal root systems in the same root system in a mixed state. The mixed colors will entirely be replaced by colors with large weight based on the superpixel correction results displayed in Figure 6B. Using superpixel correction and conducting manual recognition for comparison, it was found that the accuracy of aging root system identification relatively improved after correction. Small blocks of misclassified roots are corrected without affecting the overall root system classification. Therefore, from a pixel perspective, the corrected identification results for aging root systems are more accurate. The training and inference configurations are the same as those for root segmentation.

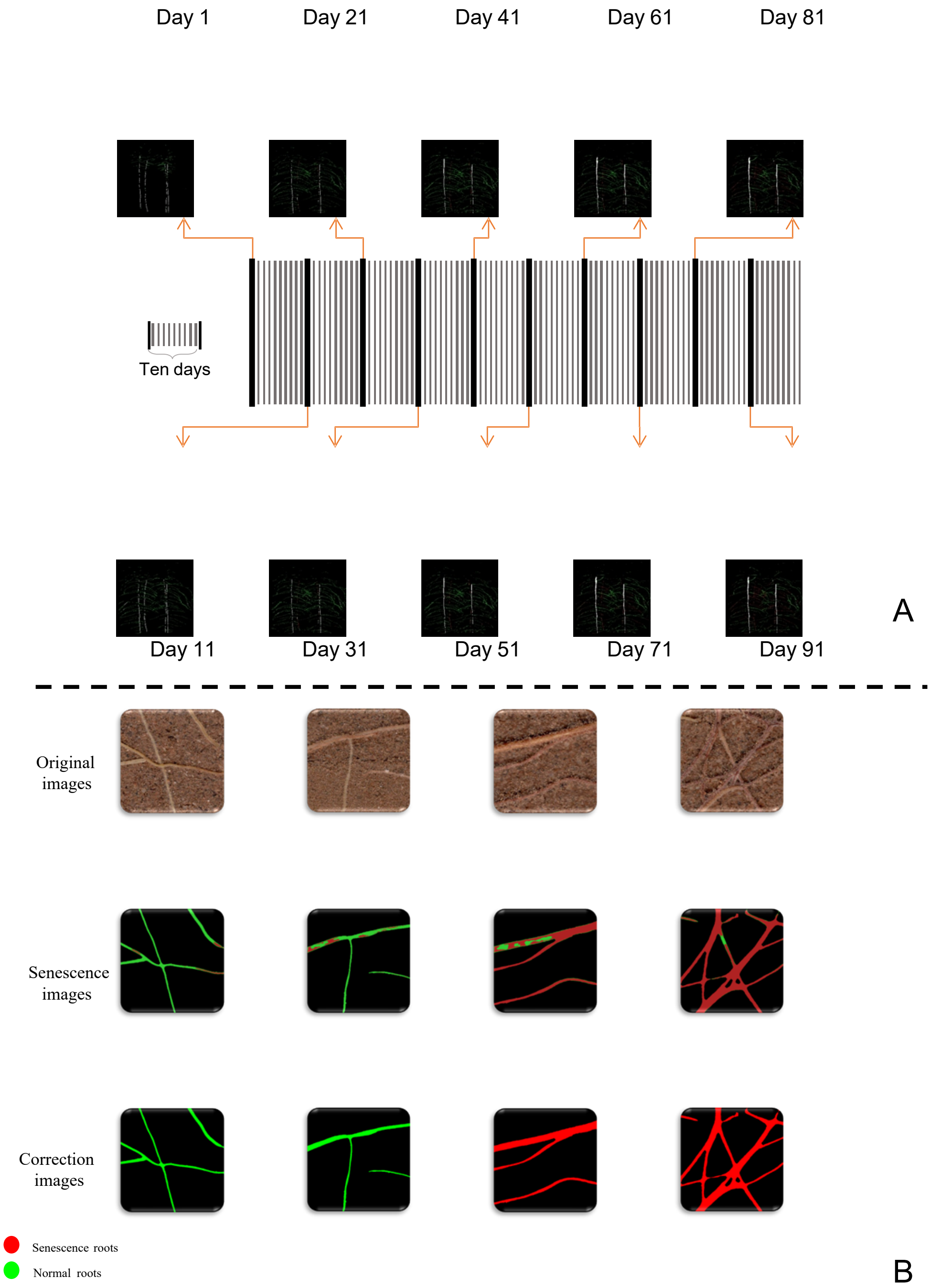

Fig6. (a)Time series senescence root results; (b) Senescence correction results

### 3.3 Exploration of Time Series Root Senescence Law

Due to the similarity of adjacent images in temporal data, dimensionality reduction and clustering are used to find the maximum difference between images and to find the proportion of normal and senescence root pixels under temporal and statistical nodes

(Figure 7A and Figure 7B). The clustering results prove that the temporal dataset is classified into ten-time intervals with different time intervals according to the growth order. The proportion of senescence root in the initial growth stage is zero, and the proportion of senescence root increases with time. However, compared to normal root, the proportion of senescence root is not significant, and the trend of senescence root in the later stage of root growth is increasing. The trend of senescence root changes over time is consistent with the observation results of the naked eye. For the normal root, the percentage of the normal root system increases due to the early stage of being in the growth phase with vigorous root growth, and the total root grows slowly in the middle and late stages, and the true ratio of the normal root decreases accordingly after the senescent root system increased. This paper performs cubic polynomial fitting on the proportion of normal and senescence pixels. The fitting results are shown in the figure, and the two curves have good fitting effects. The R-Squared ($R^2$) of the senescence fitting curve is 0.98, and the Mean Squared Error (MSE) is 0.00025. The $R^2$ of the normal fitting curve is 0.94, and the MSE is 0.00113.

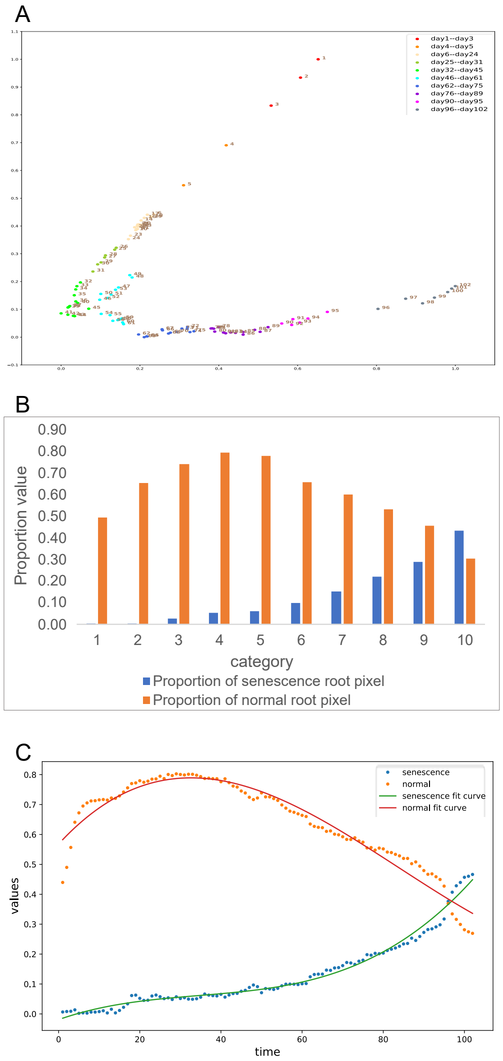

Fig7. (a) Time series senescence cluster analysis results; (b) Pixel ratio histogram of normal and senescence root at time nodes; (c) Pixel fitting curves for normal and senescence root

4. **Discussion**

### 4.1 Data annotation logic

We employ two forms of labeling rather than unique senescent labeling to label the data due to labeling expense and training issues. First of all, when compared to a senescent label, semantic segmentation reduces labeling time by two hours, and integrating two datasets can cut total labeling time. There isn't a strict restriction for the number of senescent labels in the dataset because the data we gather consists of serialized images with a high degree of similarity. Second, we merely employ the black box model of deep learning in our senescent extraction to simply understand the color difference of the root system.

### 4.2 Model analysis

The findings of the model comparison show that the TransFormer neural network's index evaluation is superior to that of the convolutional neural network. This is a result of the self-attention mechanism's capacity for direct global relationship modeling, which increases the image's receptive field and gathers additional contextual data. However, in terms of processing and parameters, a TransFormer neural network frequently exceeds a convolutional neural network. The SegFormer model reported in this research improves the self-attention mechanism, producing the smallest SegFormer(Small) model with fewer parameter values and better performance than traditional convolutional neural networks. However, the details on the root of the TransFormer neural network are more detailed, which is also a result of the self-attention mechanism's capacity to process global information. The processing results of the TransFormer neural network and the convolutional neural network in terms of actual image segmentation do not differ significantly.

Although the output results of the model are not significantly different from the rest of the convolution and TransFormer, the detailed recovery of the feature map is still insufficient. The original model decoding method is the output feature map of the four TransFormer blocks, using a straightforward and lightweight MLP plus sampling method for feature stitching fusion output results.

The initial approach in this paper was improved using two different upsampling methods (subpixel, deconvolution), in light of earlier studies by this experimental group, and it is evident from the experimental results that the above two methods are ineffective for model enhancement. This is due to the fact that adjusting the upsampling alone is insufficient to restore the target, there is an unavoidable loss of details in the preceding procedures, and the contextual information from the backbone extraction is not effectively used.

As target information and dimensions are restored, this paper adopts a tweak in the decoder structure to minimize detail loss. This paper uses the DeeplabV3plus decoder structure initially but does a bad job of utilizing scale feature maps, which leads to subpar processing results for root. While gradual convolution and feature stitching improve semantic segmentation, we adopted the UNet decoder structure for feature restoration. By using a skip connection to mix data from various trunk stages throughout the upsampling recovery process, low-level detail data and high-level semantic data may be recovered and combined more effectively, improving segmentation accuracy. After improvement, we discovered that the model’s computational complexity was lowered by an eighth compared to the original decoding structure, and the number of parameters decreased by around 0.27M, shortening training time. This is owing to the original model’s excessive usage of hidden layers and the MLP structure, which also causes the loss of spatial information in the feature map, resulting in a huge number of parameters.

To further simplify the model, the use of Depthwise separable convolution instead of convolution further reduces the number of parameters and computational complexity. Depthwise separable convolution first uses deep convolution to perform spatial convolution on each channel of the feature map, and then uses point-by-point convolution to fuse the channel information, which simplifies the convolution operation through step-by-step operations. Although the number of parameters is reduced compared to the convolutional method, the evaluation indicators are not as good as those of the convolutional indicators.

The SegFormer -UN training results achieved were the best of all models, and in addition to making improvements to the decoder, we also attempted to enhance the depth of the backbone, which entails an increase in computation and parameter count. Although the model’s parameter count has increased considerably compared to SegFormer (Small), SwinUNet, and TransUNet, and there are certain advantages, the computation increase is not considerable.

### 4.3 Senescent root extraction

#### 4.3.1 Traditional methods

Before classification, the taproot must be removed because its existence will prevent correctly extracting senescent roots from the root segmentation image after it has been obtained [66]. To build the full root, first, remove the majority of the vertical taproot from the predicted image using the opening operation, then calculate the root edge using the image gradient, enlarge it using blur and threshold, and finally combine it with the opening operation results. Lastly, lateral root extraction is achieved by images and operations. The categorization root pixel block is then intentionally picked to determine the mean and variance of the pixels (one normal root and two senescent roots). The senescent root pixels are filtered by mean and variance to retain the normal root pixels, and the retained results are filtered and sharpened. The senescent root pixel is then removed once again by means and variance, and both the final filtered image and the original image are pixel enhanced. If the calculated value is greater than 0, the original image pixel block is assigned to the filtered image until the sliding window is finished. Pixel enhancement involves calculating the two-norm difference between the two image pixel blocks in accordance with the sliding window with the initial pixel core size of 2 and the maximum pixel size of 4. [67]. Figure 8 depicts the image processing procedure.

Mean and variance of RGB channel pixels value are easily determined by artificially creating three separate pixel blocks, and the outcome of removing the senescent root is acquired after pixel-by-pixel processing. The lower, darker root system has been eliminated, leaving only the top, light-colored live root visible in the image. While some pixel values around the classification threshold are deleted from the local result map, the upper portion of the normal root of the complete root system is mistakenly left behind.

Senescent root and normal root are now separated on the same root, however, this is an atypical occurrence. The computation time for extraction is too long compared to deep learning because this method cannot be computed using GPU; it takes roughly 31 minutes to calculate one image, and this time rises when dealing with dense distributions of roots.

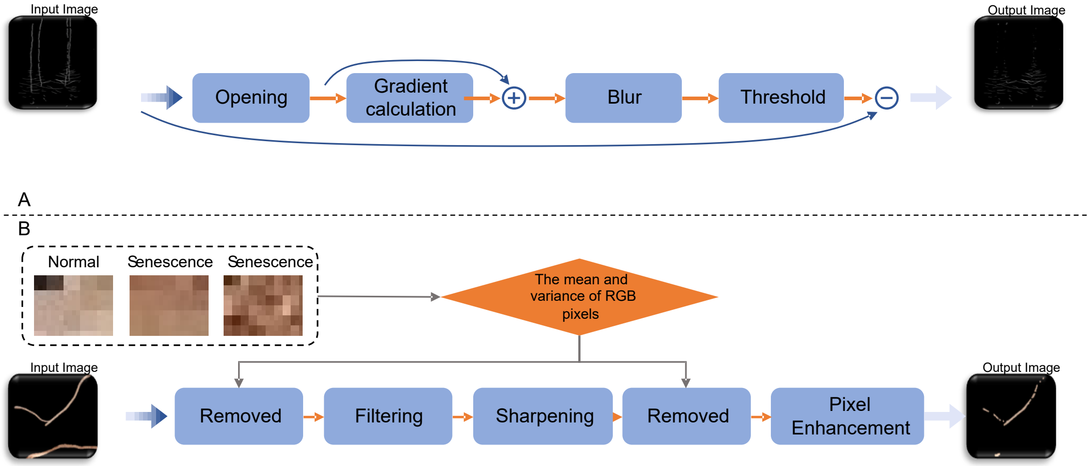

Fig8. (a) Flowchart for removing the taproot; (b) Flowchart for extracting senescence root

### 4.3.2 SegFormer-UN

At present, color classification based on deep learning is widely used in different fields and has high practicability. For instance, Ziyuan Yang et al. classified feces using deep learning to extract shallow information in the medical industry [68]. Liu et al. successfully identified the apple root system in agriculture using various seedling root colors [69]. As a result, the strategy described in this work, which is based on deep learning, is effective. The modified model for root senescent extraction has a high extraction effectiveness in the inference stage, according to experiments, and it extracts data quickly. This is due to the trained model's ability to automatically determine the senescent root's classification threshold and utilize GPU acceleration to speed up processing. The use of the SegFormer (Small) model in the inference stage is characterized by short processing time and high extraction efficiency. Considering the application of this method to mobile devices, a relatively smaller model is chosen. Compared to image processing, the overall training samples are small and the cost is relatively low, although the model costs extra during training. Zhu et al. showed that manually tracking and extracting the senescence root from time series information takes a lot of time [70]. This study improves efficiency by achieving one-stop, quick identification and analysis with deep learning and image processing.

More image processing is required in this piece since deep learning approaches have an issue with intermittent error recognition. Then, we classify senescence photos into pixel blocks, which are further separated into blocks with various forms, using the super-pixel approach. Then, we cover the image with binary prediction, keeping only the necessary root, and recolor the processing output to mostly remove false positives. Although the aforementioned method can fix a tiny percentage of wrong colors, it cannot fix vast mistake recognition areas. There are many errors in root senescence identification, including subjective annotation errors and the impact of model performance. However, compared to purely manual recognition, the efficiency improvement brought about by the model is beyond what can be achieved manually. Moreover, varied shooting environments might result in varying exposure, brightness, and contrast amongst temporal data photos, which can result in inaccurate root recognition as a result of the aforementioned factors.

## 5. Conclusion

This paper uses TransFormer neural networks to recognize roots and extract senescence roots from temporal pictures. In comparison to TransFormer and generic convolutional neural networks, the SegFormer-UN model performs better and can accurately partition cotton roots. The best root mIoU, mRecall, and mF1 index values were obtained by the model, which were 81.52%, 86.87%,and 88.81%, respectively. It is more accurate to use the real root segmentation image connection. The updated picture based on SegFormer-UN was proven to be more precise and effective in root extraction than conventional image processing techniques after two senescence root extraction approaches were verified. But, we are unable to receive full acknowledgment. We will eventually broaden the variety of root senescence samples and enhance network performance. The senescence laws of cotton root can be investigated using the senescence root extraction technique created by our research center.

## Acknowledgments

**Author contributions:** L.L. and N.W. initiated and designed the research. X.C., H.T., and J.Z. performed the experiments and collected the data. H.T. wrote the code and tested the methods., Q.Y. and H.T. analyzed the data and wrote the manuscript. R.X. translated and polished the manuscript. All authors contributed to the article and approved the submitted version.

**Funding:** This study was supported by grants from the National Natural Science Foundation of China (No. 32272220 and 32172120), and the Top-notch Talent Plan Program of the Education Department of Hebei Province (BJ2021058), and Central Guiding Local Science and Technology Development Fund Project (236Z7402G).

**Competing interests:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Data Availability:** The dataset could be given upon reasonable request from the corresponding author Nan Wang.

## Supplementary Materials

Relevant supporting materials have been uploaded to the submission system.

## References

[1] Li, A.; Zhu, L.; Xu, W.; Liu, L.; Teng, G. Recent Advances in Methods for *in Situ* Root Phenotyping. *PeerJ* **2022**, *10*, e13638, doi:10.7717/peerj.13638.

[2] Dong, H.; Niu, Y.; Li, W.; Zhang, D. Effects of Cotton Rootstock on Endogenous Cytokinins and Abscisic Acid in Xylem Sap and Leaves in Relation to Leaf Senescence. *Journal of Experimental Botany* **2008**, *59*, 1295–1304, doi:10.1093/jxb/ern035.

[3] Kunkle, J.M.; Walters, M.B.; Kobe, R.K. Senescence-Related Changes in Nitrogen in Fine Roots: Mass Loss Affects Estimation. *Tree Physiology* **2009**, *29*, 715–723, doi:10.1093/treephys/tpp004.

[4] Chen, Y.; Dong, H. Mechanisms and Regulation of Senescence and Maturity Performance in Cotton. *Field Crops Research* **2016**, *189*, 1–9, doi:10.1016/j.fcr.2016.02.003.

[5] LIU Feng-shan; ZHOU Zhi-bin; HU Shun-jun; DU Hai-yan; CHEN Xiu-long Influence of different soil coring methods on estimation of root distribution characteristics. *Acta Prataculturae Sinica* **2012**, *21*, 294–299.

[6] FEI Luyang Comparative Analysis of Double Ring Method and Single Ring Soil Column Method in Measuring Soil Infiltration on Loess Surface. *Soil and Water Conservation in China* **2020**, 47-50+5, doi:10.14123/j.cnki.swcc.2020.0193.

[7] YANG Yichen; YANG Xiwen; WU Yin; HUANG Yuan; XU Lili; FU Jinzhou; GUO Fangfang; ZHOU Sumei; HE Dexian Discussion on Improving Precision of Wheat Root Research by Cube Sampling Method. *Journal of Henan Agricultural Sciences* **2021**, *50*, 36–46, doi:10.15933/j.cnki.1004-3268.2021.11.005.

[8] LI Long; LI ChaoNan; MAO XinGuo; WANG JingYi; JING RuiLian Advances and Perspectives of Approaches to Phenotyping Crop Root System. *Scientia Agricultura Sinica* **2022**, *55*, 425–437.

[9] Jie, H.; Austin, P.T.; Kong, L.S. Effects of Elevated Root Zone CO2 and Air Temperature on Photosynthetic Gas Exchange, Nitrate Uptake, and Total Reduced Nitrogen Content in Aeroponically Grown Lettuce Plants. *Journal of Experimental Botany* **2010**, *61*, 3959–3969, doi:10.1093/jxb/erq207.

[10] Marié, C.L.; Kirchgessner, N.; Flütsch, P.; Pfeifer, J.; Hund, A. RADIX: Rhizoslide Platform Allowing High Throughput Digital Image Analysis of Root System Expansion. *Plant Methods* **2016**, *12*, doi:10.1186/s13007-016-0140-8.

[11] Piñeros, M.A.; Larson, B.G.; Shaff, J.E.; Schneider, D.J.; Falcão, A.X.; Yuan, L.; Clark, R.T.; Craft, E.J.; Davis, T.W.; Pradier, P.-L.; et al. Evolving Technologies for Growing, Imaging and Analyzing 3D Root System Architecture of Crop Plants: Digital Phenotyping of Root System Architecture. *J. Integr. Plant Biol.* **2016**, *58*, 230–241, doi:10.1111/jipb.12456.

[12] Wu, J.; Wu, Q.; Pagès, L.; Yuan, Y.; Zhang, X.; Du, M.; Tian, X.; Li, Z. RhizoChamber-Monitor: A Robotic Platform and Software Enabling Characterization of Root Growth. *Plant Methods* **2018**, *14*, doi:10.1186/s13007-018-0316-5.

[13] Parker, C.J.; Carr, M.K.V.; Jarvis, N.J.; Puplampu, B.O.; Lee, V.H. An Evaluation of the Minirhizotron Technique for Estimating Root Distribution in Potatoes. *J. Agric. Sci.* **1991**, *116*, 341–350, doi:10.1017/S0021859600078151.

[14] Taylor, B.N.; Beidler, K.V.; Strand, A.E.; Pritchard, S.G. Improved Scaling of Minirhizotron Data Using an Empirically-Derived Depth of Field and Correcting for the Underestimation of Root Diameters. *Plant Soil* **2014**, *374*, 941–948, doi:10.1007/s11104-013-1930-7.

[15] Scotson, C.P.; Duncan, S.J.; Williams, K.A.; Ruiz, S.A.; Roose, T. X-ray Computed Tomography Imaging of Solute Movement through Ridged and Flat Plant Systems. *Eur J Soil Sci* **2021**, *72*, 198–214, doi:10.1111/ejss.12985.

[16] Bagnall, G.C.; Koonjoo, N.; Altobelli, S.A.; Conradi, M.S.; Fukushima, E.; Kuethe, D.O.; Mullet, J.E.; Neely, H.; Rooney, W.L.; Stupic, K.F.; et al. Low-Field Magnetic Resonance Imaging of Roots in Intact Clayey and Silty Soils. *Geoderma* **2020**, *370*, 114356, doi:10.1016/j.geoderma.2020.114356.

[17] Zhang; Derival; Albrecht; Ampatzidis Evaluation of a Ground Penetrating Radar to Map the Root Architecture of HLB-Infected Citrus Trees. *Agronomy* **2019**, *9*, 354, doi:10.3390/agronomy9070354.

[18] Schierholt, A.; Tietz, T.; Bienert, G.P.; Gertz, A.; Miersch, S.; Becker, H.C. Root System Size Response of Bzh Semi-Dwarf Oilseed Rape Hybrids to Different Nitrogen Levels in the Field. *Annals of Botany* **2019**, *124*, 891–901, doi:10.1093/aob/mcy197.

[19] Li, A.; Zhu, L.; Xu, W.; Liu, L.; Teng, G. Recent Advances in Methods for *in Situ* Root Phenotyping. *PeerJ* **2022**, *10*, e13638, doi:10.7717/peerj.13638.

[20] Mohamed, A.; Monnier, Y.; Mao, Z.; Lobet, G.; Maeght, J.-L.; Ramel, M.; Stokes, A. An Evaluation of Inexpensive Methods for Root Image Acquisition When Using Rhizotrons. *Plant Methods* **2017**, *13*, 11, doi:10.1186/s13007-017-0160-z.

[21] Nahar, K.; Pan, W.L. High Resolution in Situ Rhizosphere Imaging of Root Growth Dynamics in Oilseed Castor Plant (Ricinus Communis L.) Using Digital Scanners. *Model. Earth Syst. Environ.* **2019**, *5*, 781–792, doi:10.1007/s40808-018-0564-4.

[22] Xiao, S.; Liu, L.; Zhang, Y.; Sun, H.; Zhang, K.; Bai, Z.; Dong, H.; Li, C. Fine Root and Root Hair Morphology of Cotton under Drought Stress Revealed with RhizoPot. *J Agro Crop Sci* **2020**, *206*, 679–693, doi:10.1111/jac.12429.

[23] Zhao, H.; Wang, N.; Sun, H.; Zhu, L.; Zhang, K.; Zhang, Y.; Zhu, J.; Li, A.; Bai, Z.; Liu, X.; et al. RhizoPot Platform: A High-Throughput in Situ Root Phenotyping Platform with Integrated Hardware and Software. *Front. Plant Sci.* **2022**, *13*, 1004904, doi:10.3389/fpls.2022.1004904.

[24] Le Bot, J.; Serra, V.; Fabre, J.; Draye, X.; Adamowicz, S.; Pagès, L. DART: A Software to Analyse Root System Architecture and Development from Captured Images. *Plant Soil* **2010**, *326*, 261–273, doi:10.1007/s11104-009-0005-2.

[25] Betegón-Putze, I.; González, A.; Sevillano, X.; Blasco-Escámez, D.; Caño-Delgado, A.I. My ROOT: A Method and Software for the Semiautomatic Measurement of Primary Root Length in Arabidopsis Seedlings. *Plant J* **2019**, tpj.14297, doi:10.1111/tpj.14297.

[26] Pound, M.P.; French, A.P.; Atkinson, J.A.; Wells, D.M.; Bennett, M.J.; Pridmore, T. RootNav: Navigating Images of Complex Root Architectures. *Plant Physiology* **2013**, *162*, 1802–1814, doi:10.1104/pp.113.221531.

[27] WANG Zhen; YAO Lingjie; ZHONG Fangqiang Application of Image Semantic Segmentation in Smart Agriculture. *Information & Computer* **2022**, *34*, 32-34+49.

[28] Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation 2015.

[29] Kamal, S.; Shende, V.G.; Swaroopa, K.; Bindhu Madhavi, P.; Akram, P.S.; Pant, K.; Patil, S.D.; Sahile, K. FCN Network-Based Weed and Crop Segmentation for IoT-Aided Agriculture Applications. *Wireless Communications and Mobile Computing* **2022**, *2022*, 1–10, doi:10.1155/2022/2770706.

[30] Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation 2016.

[31] Wang, T.; Rostamza, M.; Song, Z.; Wang, L.; McNickle, G.; Iyer-Pascuzzi, A.S.; Qiu, Z.; Jin, J. SegRoot: A High Throughput Segmentation Method for Root Image Analysis. *Computers and Electronics in Agriculture* **2019**, *162*, 845–854, doi:10.1016/j.compag.2019.05.017.

[32] Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation 2015.

[33] Gaggion, N.; Ariel, F.; Daric, V.; Lambert, É.; Legendre, S.; Roulé, T.; Camoirano, A.; Milone, D.H.; Crespi, M.; Blein, T.; et al. *ChronoRoot: High-Throughput Phenotyping by Deep Segmentation Networks Reveals Novel Temporal Parameters of Plant Root System Architecture*; Plant Biology, 2020;

[34] Smith, A.G.; Han, E.; Petersen, J.; Olsen, N.A.F.; Giese, C.; Athmann, M.; Dresbøll, D.B.; Thorup-Kristensen, K. *RootPainter: Deep Learning Segmentation of Biological Images with Corrective Annotation*; Plant Biology, 2020;

[35] Peters, B.; Blume-Werry, G.; Gillert, A.; Schwieger, S.; von Lukas, U.F.; Kreyling, J. As Good as Human Experts in Detecting Plant Roots in Minirhizotron Images but Efficient and Reproducible: The Convolutional Neural Network "RootDetector." *Sci Rep* **2023**, *13*, 1399, doi:10.1038/s41598-023-28400-x.

[36] Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network 2017.

[37] Zhang, R.; Chen, J.; Feng, L.; Li, S.; Yang, W.; Guo, D. A Refined Pyramid Scene Parsing Network for Polarimetric SAR Image Semantic Segmentation in Agricultural Areas. *IEEE Geosci. Remote Sensing Lett.* **2022**, *19*, 1–5, doi:10.1109/LGRS.2021.3086117.

[38] Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation 2018.

[39] Shen, C.; Liu, L.; Zhu, L.; Kang, J.; Wang, N.; Shao, L. High-Throughput in Situ Root Image Segmentation Based on the Improved DeepLabv3+ Method. *Front. Plant Sci.* **2020**, *11*, 576791, doi:10.3389/fpls.2020.576791.

[40] Kang, J.; Liu, L.; Zhang, F.; Shen, C.; Wang, N.; Shao, L. Semantic Segmentation Model of Cotton Roots In-Situ Image Based on Attention Mechanism. *Computers and Electronics in Agriculture* **2021**, *189*, 106370, doi:10.1016/j.compag.2021.106370.

[41] Zhong, G.; Ling, X.; Wang, L. From Shallow Feature Learning to Deep Learning: Benefits from the Width and Depth of Deep Architectures. WIREs Data Min & Knowl 2019, 9, e1255, doi:10.1002/widm.1255.

[42] Salas, J.; De Barros Vidal, F.; Martinez-Trinidad, F. Deep Learning: Current State. IEEE Latin Am. Trans. 2019, 17, 1925–1945, doi:10.1109/TLA.2019.9011537.

[43] Li, Y.; Huang, Y.; Wang, M.; Zhao, Y. An Improved U-Net-Based in Situ Root System Phenotype Segmentation Method for Plants. Front. Plant Sci. 2023, 14, 1115713, doi:10.3389/fpls.2023.1115713.

[44] Lu, W.; Wang, X.; Jia, W. Root Hair Image Processing Based on Deep Learning and Prior Knowledge. Computers and Electronics in Agriculture 2022, 202, 107397, doi:10.1016/j.compag.2022.107397.

[45] Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation 2014.

[46] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need 2017.

[47] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale 2021.

[48] Alshammari, H.; Gasmi, K.; Ben Ltaifa, I.; Krichen, M.; Ben Ammar, L.; Mahmood, M.A. Olive Disease Classification Based on Vision Transformer and CNN Models. *Computational Intelligence and Neuroscience* **2022**, *2022*, 1–10, doi:10.1155/2022/3998193.

[49] Chen, J.; Luo, T.; Wu, J.; Wang, Z.; Zhang, H. A Vision Transformer Network SEEDVIT for Classification of Maize Seeds. *J Food Process Engineering* **2022**, *45*, doi:10.1111/jfpe.13998.

[50] Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows 2021.

[51] Lu, J.; Wang, W.; Zhao, K.; Wang, H. Recognition and Segmentation of Individual Pigs Based on Swin Transformer. *Animal Genetics* **2022**, *53*, 794–802, doi:10.1111/age.13259.

[52] Meng, X.; Yang, Y.; Wang, L.; Wang, T.; Li, R.; Zhang, C. Class-Guided Swin Transformer for Semantic Segmentation of Remote Sensing Imagery. *IEEE Geosci. Remote Sensing Lett.* **2022**, *19*, 1–5,

doi:10.1109/LGRS.2022.3215200.

[53] Agilandeeswari, L.; Meena, S.D. SWIN Transformer Based Contrastive Self-Supervised Learning for Animal Detection and Classification. *Multimed Tools Appl* **2023**, *82*, 10445–10470, doi:10.1007/s11042-022-13629-x.

[54] Wang, Y.; Zhang, S.; Dai, B.; Yang, S.; Song, H. Fine-Grained Weed Recognition Using Swin Transformer and Two-Stage Transfer Learning. *Front. Plant Sci.* **2023**, *14*, 1134932, doi:10.3389/fpls.2023.1134932.

[55] Wang, Z.; Zhang, Z.; Lu, Y.; Luo, R.; Niu, Y.; Yang, X.; Jing, S.; Ruan, C.; Zheng, Y.; Jia, W. SE-COTR: A Novel Fruit Segmentation Model for Green Apples Application in Complex Orchard. Plant Phenomics 2022, 2022, 0005, doi:10.34133/plantphenomics.0005.

[56] Großkinsky, D.K.; Syaifullah, S.J.; Roitsch, T. Integration of Multi-Omics Techniques and Physiological Phenotyping within a Holistic Phenomics Approach to Study Senescence in Model and Crop Plants. *Journal of Experimental Botany* **2018**, *69*, 825–844, doi:10.1093/jxb/erx333.

[57] Neilson, E.H.; Edwards, A.M.; Blomstedt, C.K.; Berger, B.; Møller, B.L.; Gleadow, R.M. Utilization of a High-Throughput Shoot Imaging System to Examine the Dynamic Phenotypic Responses of a C4 Cereal Crop Plant to Nitrogen and Water Deficiency over Time. *Journal of Experimental Botany* **2015**, *66*, 1817–1832, doi:10.1093/jxb/eru526.

[58] Cai, J.; Okamoto, M.; Atieno, J.; Sutton, T.; Li, Y.; Miklavcic, S.J. Quantifying the Onset and Progression of Plant Senescence by Color Image Analysis for High Throughput Applications. *PLoS ONE* **2016**, *11*, e0157102, doi:10.1371/journal.pone.0157102.

[59] Hendrick, R.L.; Pregitzer, K.S. The Demography of Fine Roots in a Northern Hardwood Forest. *Ecology* **1992**, *73*, 1094–1104, doi:10.2307/1940183.

[60] Zhu, L.; Liu, L.; Sun, H.; Zhang, Y.; Liu, X.; Wang, N.; Chen, J.; Zhang, K.; Bai, Z.; Wang, G.; et al. The Responses of Lateral Roots and Root Hairs to Nitrogen Stress in Cotton Based on Daily Root Measurements. *J Agronomy Crop Science* **2022**, *208*, 89–105, doi:10.1111/jac.12525.

[61] Zhu, L.; Liu, L.; Sun, H.; Zhang, K.; Zhang, Y.; Li, A.; Bai, Z.; Wang, G.; Liu, X.; Dong, H.; et al. Low Nitrogen Supply Inhibits Root Growth but Prolongs Lateral Root Lifespan in Cotton. *Industrial Crops and Products* **2022**, *189*, 115733, doi:10.1016/j.indcrop.2022.115733.

[62] Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers 2021.

[63] Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation 2021.

[64] Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation 2021.

[65] Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.S.; et al. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers 2021.

[66] RUAN Zihang; HUANG Yong; WANG Meng; SHI Qiang; ZHANG Jinling Highlight removal method of tomato surface based on image processing. *China Cucurbits and Vegetables* **2023**, *36*, 64–70, doi:10.16861/j.cnki.zggc.2023.0066.

[67] Burgos-Artizzu, X.P.; Ribeiro, A.; Tellaeche, A.; Pajares, G.; Fernández-Quintanilla, C. Analysis of Natural Images Processing for the Extraction of Agricultural Elements. *Image and Vision Computing* **2010**, *28*, 138–149, doi:10.1016/j.imavis.2009.05.009.

[68] Yang, Z.; Leng, L.; Kim, B.-G. StoolNet for Color Classification of Stool Medical Images. *Electronics* **2019**, *8*, 1464, doi:10.3390/electronics8121464.

[69] Liu, Y.; Qian, J.; Yang, X.; Di, B.; Zhou, J. Study on Measurement Method for Apple Root Morphological Parameters Based on Labview. *Plant Methods* **2019**, *15*, 149, doi:10.1186/s13007-019-0535-4.

[70] Zhu, L.; Liu, L.; Sun, H.; Zhang, Y.; Zhu, J.; Zhang, K.; Li, A.; Bai, Z.; Wang, G.; Li, C. Physiological and Comparative Transcriptomic Analysis Provide Insight Into Cotton (Gossypium Hirsutum L.) Root Senescence in Response. *Front. Plant Sci.* **2021**, *12*, 748715, doi:10.3389/fpls.2021.748715.