PLANT LEAF SEGMENTATION AND PHENOTYPIC ANALYSIS BASED ON FULLY CONVOLUTIONAL NEURAL NETWORK

Liying Cao[1], Hongda Li[1], Helong Yu[1,2*], Guifen Chen[3*], Heshu Wang[1]

[1] College of Information & Technology, Jilin Agricultural University, Jilin, China.
[2] Jilin precision agriculture and big data engineering research center, Changchun 130118, China.
[3] Changchun Humanities and Sciences College, Changchun 130117, China.
[*] Correspondence: yuhelong@jlau.edu.cn and chenguifen@jlau.edu.cn.

HIGHLIGHTS
- Modify the U-Net segmentation network to reduce the loss of segmentation accuracy.
- Reducing the number of layers U-net network, modifying the loss function, and the increase in the output layer dropout.
- It can be well extracted after splitting blade morphological model and color feature.

ABSTRACT. *From the perspective of computer vision, the shortcut to extract phenotypic information from a single crop in the field is image segmentation. Plant segmentation is affected by the background environment and illumination. Using deep learning technology to combine depth maps with multi-view images can achieve high-throughput image processing. This article proposes an improved U-Net segmentation network, based on small sample data enhancement, and reconstructs the U-Net model by optimizing the model framework, activation function and loss function. It is used to realize automatic segmentation of plant leaf images and extract relevant feature parameters. Experimental results show that the improved model can provide reliable segmentation results under different leaf sizes, different lighting conditions, different backgrounds, and different plant leaves. The pixel-by-pixel segmentation accuracy reaches 0.94. Compared with traditional methods, this network achieves robust and high-throughput image segmentation. This method is expected to provide key technical support and practical tools for top-view image processing, Unmanned Aerial Vehicle phenotype extraction, and phenotype field platforms.*

Keywords. Deep learning, Full convolution neural network, Image segmentation, Phenotype analysis, U-Net.

Plants are essential material resources for human survival (Hao et al., 2017). Therefore, researchers in agricultural-related industries are investing in the sustainable development of agriculture. In recent years, plant phenotypic analysis has solved many related problems based on the image analysis technology (Kim et al., 2017). The methods were used in plant growth monitoring, morphological analysis, and phenotype extraction by analyzing the images of different organs of plants. Further, the characteristics of the plant were analyzed, and the yield of the plant was predicted (Thomas et al., 2018). Various image processing and computer vision technologies were studied to achieve the non-destructive plant phenotype extraction for different experimental designs and specific image acquisition scenarios (Mubin et al., 2019).

Many studies have been conducted in plant image analysis (Bock et al., 2010; Zhang et al., 2019), including plant disease identification, leaf detection, leaf technology, leaf segmentation, and growth monitoring (Zhou et al., 2019). Most of the relevant data sources were collected in a controlled laboratory environment, greenhouse, or field. The coverage of wheat and other fine-leaved crops were extracted based on the improved k-means image segmentation (Wu et al., 2019). Although the higher accuracy (>90%) was achieved, such a traditional image processing algorithm requires a time-consuming tuning process. Praveen Kumar and Domnic (2018) proposed the image enhancement-based leaf segmentation method for *Arabidopsis* plants in Computer Vision Problems in Plant Phenotyping (CVPPP) dataset. The accuracy reached 95.4%, and it can be used for the segmentation of other round leaf plants. However, this method cannot be applied to the mass segmentation of extensive data, requiring human interaction as an input. Qian et al. (2015) carried out the accurate segmentation and chlorophyll diagnosis of corn multispectral image. The accuracy of the proposed local threshold processing-based segmentation was 95.59%. However, the extracted spectral feature parameter

modeling R2 was only about 0.596, which indicates that the accuracy of parameter extraction is not ideal. Chen et al. (2012) proposed a disease recognition method based on fuzzy clustering and vector machine segmentation of corn disease leaf images based on machine vision technology. The recognition rate was more than 95%, but the generalization ability was limited. Furthermore, many links involved human interactions, leading to inefficient disease recognition. Goclawski et al. (2012) segmented the dyed and scanned image of cucurbit leaves through a neural network and analyzed the color characteristics of non-biological stress. However, in their model, only the application of the back-propagation network was considered, and did not compare the solution proposed by them with other neural network models, so their network may not be the optimal architecture. Ma et al. (2019) adopted c (FCN) for the segmentation of seedlings and weeds in the rice seedling field image. The proposed SEGNET (segmentation network) model effectively classified the pixels of rice seedlings, background, and weeds in rice field images, achieving an average accuracy: 92.7%. The comparison of existing feature extraction methods is shown in table 1.

According to the analysis of the related works, the accuracy of leaf segmentation depends on the quality of the dataset, including the complexity of background, light, and other environmental factors. Besides, they were mostly based on supervised learning or traditional hand-crafted methods, which provide low robustness and accuracy. In order to overcome these shortcomings, this article proposes to adopt the U-Net for plant image segmentation, taking the following advantages: 1) the high-throughput segmentation and batch segmentation of plant leaf images, 2) high-precision leaf segmentation effect in different light intensity and different backgrounds.

U-Net convolutional neural network segmentation is based on a small sample training network. The model is modified in terms of the number of convolutional layers, activation function, and loss function. The improved model reduces the number of training weights, and the training speed is high. The result is high in accuracy. The feature extraction of the blade is from Semantic segmentation is performed in four aspects, and accuracy comparison is performed.

Table 1. Information on existing feature extraction methods.

	Existing Methods	Shortcoming	Accuracy
Leaf detection	Image based leaf segmentation and counting in rosette plants	Not suitable for big data	95.4%
Leaf segmentation	Based on the improved k-means image segmentation	Time consuming	90%
	Accurate segmentation and chlorophyll diagnosis of corn multispectral image	Low extraction accuracy	95.59%
	Clustering and vector machine image segmentation	Low efficiency	95%
	FCN,Segnet	Time consuming	92.7%

METHODS AND MATERIALS

DATASET

In this study, CVPPP dataset is used (Kumar and Domnic, 2020), which consists of four categories of plant images: A1, A2, A3, and A4. The CVPPP dataset was constructed for the challenge of leaf segmentation. Examples of images in the CVPPP dataset are shown in figure 1. The images are taken by Canon AD1000 power shot camera (Tokyo, Japan) with 7 mp. The images are composed of the top images of the plant placed in the flowerpot and tray. A1 and A2 subsets are *Arabidopsis* plants, acquired every 6 h in the daytime for 3 weeks and 20 min in the daytime for seven weeks, respectively. A1 has a complex light intensity and background environment, while A2 has a simple background environment but highly-varied plants. A3 subset is composed of tobacco plant images. A4 subset is composed of fir and maple. The images include diverse characteristics of plants, including different kinds of plants, different growth stages of plants, different numbers of plants. Also, the resolution, scene complexity, and light intensity are varied. Accordingly, the CVPPP is a challenging plant-segmentation dataset.

This article uses three sets of image data sets: two sets of *Arabidopsis* plants (A1, A2) and one set of tobacco plants (A3). The overall composition of the data set is shown in table 2.

THE PROPOSED SEGMENTATION NETWORK

This article proposes to adopt the U-Net to achieve the pixel-wise segmentation of plant images, which is learned in the end-to-end framework. The U-Net proposed in (Ronneberger et al., 2015) is improved and optimized in this article to achieve better segmentation and reduce training costs. In the proposed model, the modified data reading mode is employed, and the activation and loss functions are optimized in an adjusted framework. The model attempts to train a small number of samples, so the dropout layer is appropriately added to prevent network overfitting. In terms of network framework, this research model reduces the number of network layers, mainly because the training sample size is not set much, and the quality of image segmentation is guaranteed under the resolution of 256×256. The network parameters constructed were reduced to a certain extent, mainly in order to relieve the operating pressure of hardware equipment and reduce the training time while ensuring the segmentation effect.

The structure of the proposed network is depicted in figure 2. The first three layers on the left side of the network are the typical structure of encoding convolution blocks, which are composed of two 3×3 convolution layers (Conv) and one max-pooling layer. High-dimensional features are extracted through the encoder. The spatial dimension is gradually reduced by the max-pooling layers, while the number of feature channels is gradually increased as 32, 64, and 128. The 4th block, depicted at the bottom of the network, comprises two 3×3 convolution layers with the number of feature channels 256. The decoder part (5th-7th blocks) is composed of deconvolution blocks. The 5th and 6th deconvolution blocks are composed of two convolution layers and one following up-sampling layer. The last 7th block is
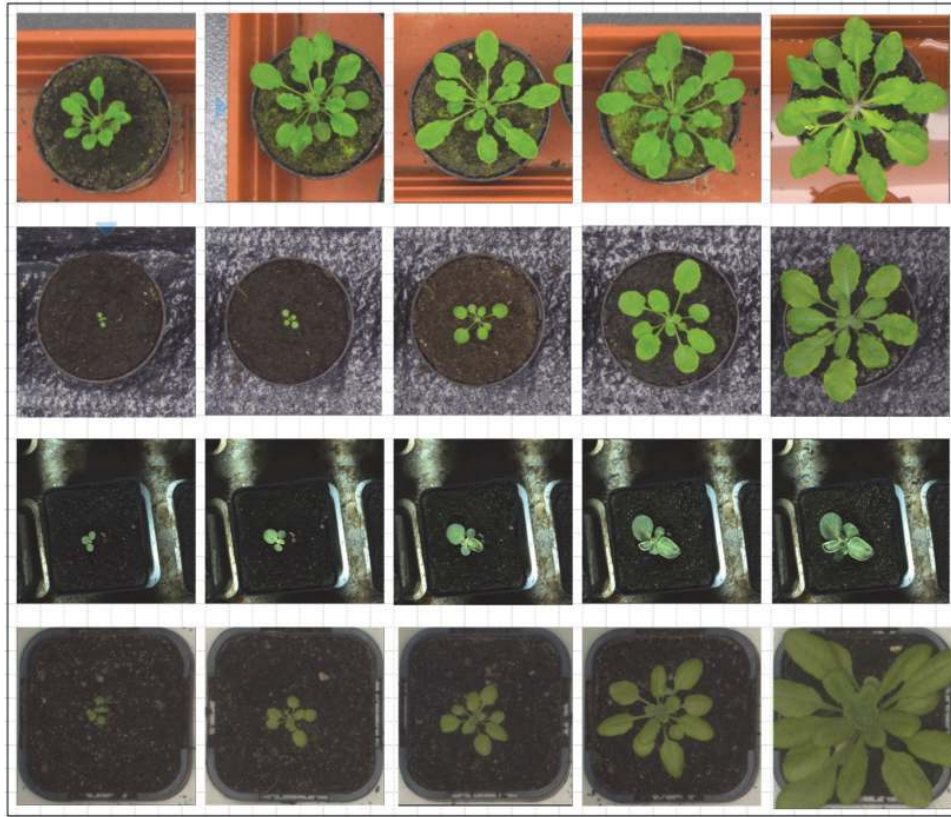
Figure 1. Examples of CVPPP dataset.

Table 2. The composition of the CVPPP dataset.

Dataset	Subjects	Wild-types	Mutants	Period	Total Images	Annotated Plants
A1	19	Yes	No	3	150	161
A2	24	Yes	Yes	7	1951	40
A3	20	-	-	2.5	34560	26

composed of two 3×3 convolution layers and one 1×1 convolution layer.

Further, the features from the previous blocks in the encoder are concatenated with the up-sampled features. Through these skip-connections, local features extracted in
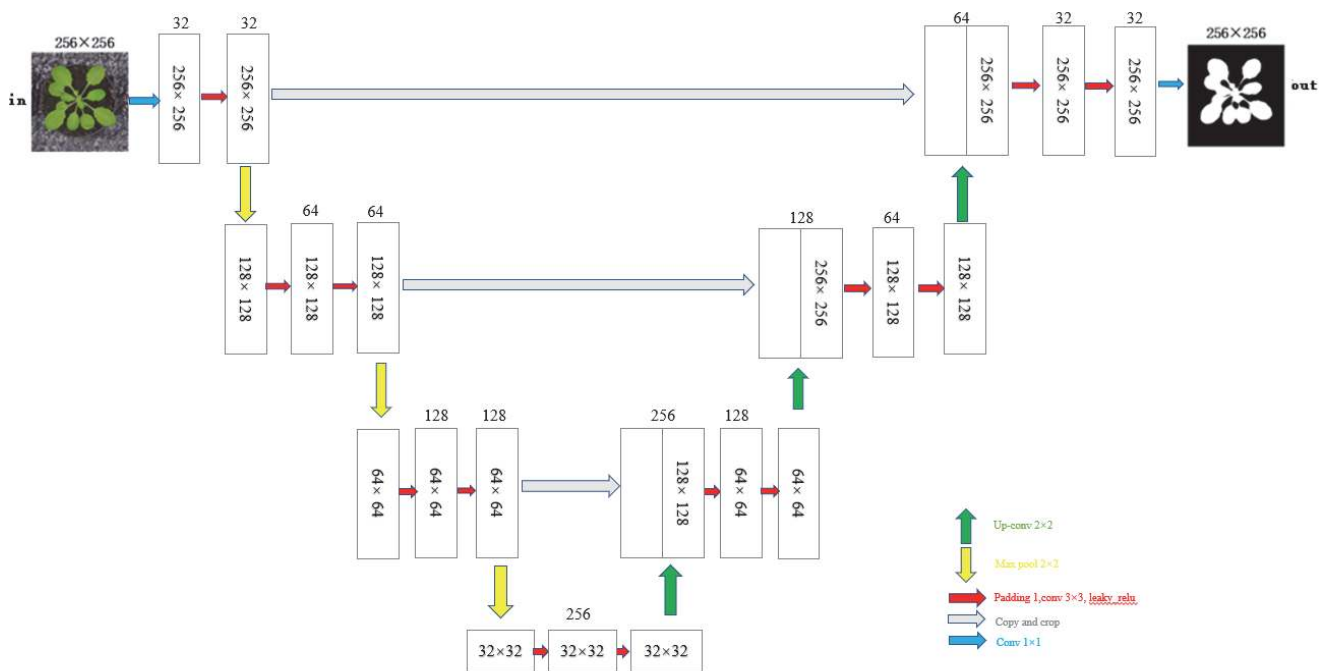

Figure 2. The overall structure of the proposed network model.

the encoder and the abstracted features from the previous block are fused to obtain better features while retaining the details. The last 1×1 convolution layer maps each component eigenvector to the required classification.

The batch normalization (BN) layer is also adapted to normalize the features of each layer. It makes the feature distribution of each layer more uniform, improving the convergence speed and the fault tolerance ability of the model. The parameters of the network are reduced to a certain extent, mainly in order to alleviate the pressure of hardware operation so that the training time is reduced while ensuring the segmentation accuracy(shown in table 2). Further, the dropout is adopted to appropriately prevent the network from overfitting, often induced by small training data. The network model is trained with an optimizer (SGD) in random gradient descent with momentum parameters. The learning rate is changed by the decay rate, which auxiliary optimizes the model.

ACTIVATION AND LOSS FUNCTIONS
Activation Function
A linear model is limited to model such a complex segmentation problem. Accordingly, the non-linear activation function was introduced to add non-linearity factors in deep networks. The widely used activation functions include Sigmoid, ReLu, Leaky-ReLu, and Tanh, which are plotted in figure 3.

The ReLu is a kind of linear activation function, mathematically expressed as follows:

$$f(x) = max(0, x) \qquad (1)$$

As the activation function of a neural network, it defines the non-linear output result after the linear transformation $w^{T}x + b$. As shown in figure 3, the output values of the ReLu are the same as the input value for $x > 0$; otherwise, ReLu deactivates the neurons. This situation has little effect on the forward propagation process in the traditional convolution

neural network model. However, in the back-propagation, the gradient will become 0 and result in the deactivation of neurons, which is called the dying ReLu problem (Lin and Shen, 2018; Tekchandani et al., 2020).

To overcome the dying ReLu problem, Leaky-ReLu was proposed where the values are not zero for negative inputs. Tanh function also has the same property. However, Leaky-ReLu has a simple linear relationship in the negative output part than the Tanh function, which is conducive to the rapid convergence of the model. Leaky-ReLu is defined as:

$$f(x) = \begin{cases} x & if\ x \geq 0 \\ \theta x & if\ x < 0 \end{cases} \qquad (2)$$

When the input value x is negative, its gradient is constant $\theta \in (0, 1)$, which alleviates the neuron deactivation problem in back-propagation.

In the proposed network, the Sigmoid activation function is used for the final output layer to activate the binary classification. Sigmoid can map a real number to an interval of (0,1), which is very suitable for binary classification, especially when the feature is complex. The Sigmoid activation function is not adopted for other layers because it is known to have the gradient vanishing problem in deeper layers during the back-propagation. When Sigmoid is close to the saturation area, the gradient closes to zero, and thus the propagated gradients are disappeared. Sigmoid activation is defined as:

$$S(x) = \frac{1}{1+e^{-x}} \qquad (3)$$

The models were compared using the above activation function training. The relationship between the loss and the number of iterations in the training process using different activation functions is shown in the figure 4.

The loss value of using the ReLU and Leaky ReLU activation function model is lower than the Tanh and Sigmoid functions in the early stage, indicating that the use of ReLU
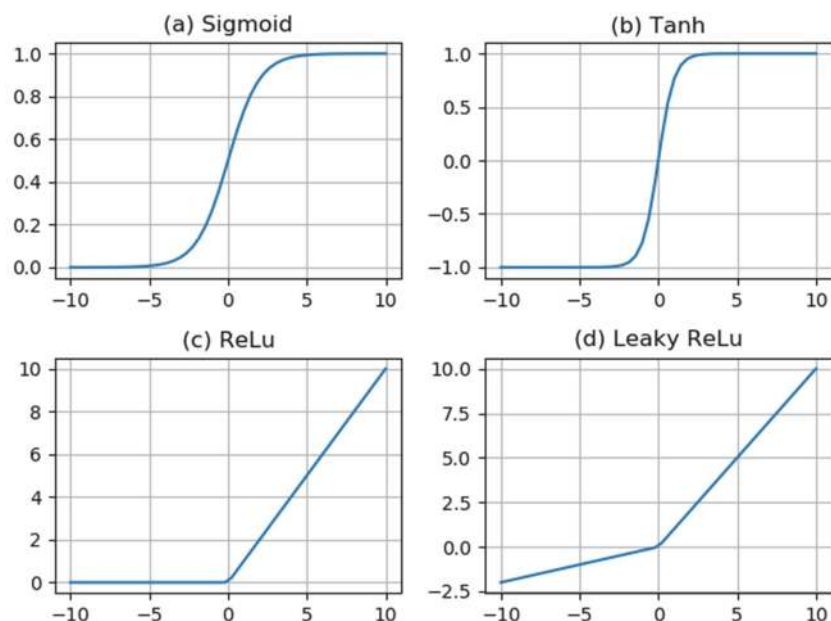

Figure 3. The widely used activation functions: (a) Sigmoid, (b) Tanh, (c) ReLu, and (d) Leaky-ReLu.
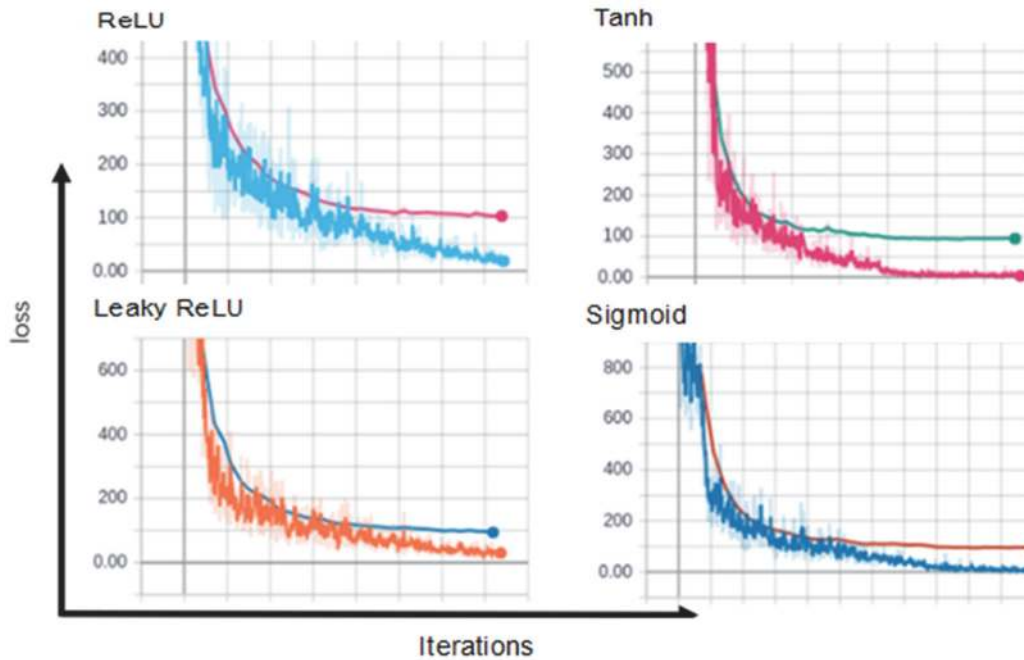
Figure 4. Relationship between loss and number of iterations of different activation functions during training.

and Leaky ReLU as the activation function is conducive to the rapid convergence of the model, and the specific comparison of the ReLU and Leaky ReLU functions has been carried out above. After the narrative, in the model training process, Leaky ReLU is selected as the activation function of the first few layers. For the binary classification task of the last layer, the Sigmoid function is still selected as the activation function of the output layer, and the loss function is used for classification output.

Loss Function

U-Net employed the cross-entropy loss function for multi-classification, defined as follows:

$$E = x \in \Omega w(x) log\left(p_{l(x)^{(x)}} \right) \tag{4}$$

where x and $w(x)$ represent the pixels and the weights. $log\left(p_{l(x)^{(x)}} \right)$ indicates the logarithm of probability p obtained by cross-entropy calculation of pixel x position in a category, and the subscript $l(x)$ represents the category.

The task of plant segmentation in *Arabidopsis* is formulated as a pixel-wise binary classification, and thus it does not require a complex multi-class cross-entropy loss. Accordingly, the modified cross-entropy function, the so-called binary cross-entropy function, is used as a loss function. However, in order to match the output of the last sigmoid function, the binary cross-entropy function is chosen as the loss function, and its mathematical expression is as follows.

$$L = -\left[ylogy^{'} + (1-y)log(1-y') \right] \tag{5}$$

where y is the ground-truth value, y' is the estimated value. y = 1 indicates that it is positive. y = 0 indicates that it is negative.

$$L = -logy' \tag{6}$$

$$L = -\log(1-y') \tag{7}$$

It can be seen from the equation that when the estimated value is close to 0, the output of the loss function is small, and the output value of the loss function is large when it is close to 1. It is very suitable for the two-class output of the last layer of the network.

Model accuracy and loss are shown in the figure 5.

Evaluation Metric for Segmentation Accuracy

The Intersection over Union (IoU) is used as the evaluation metric for the segmentation results, which is the ratio of the intersection and union of the predicted value and the real value of the image. Recall rate (R) represents the measurement of coverage, which measures that multiple leaf pixels are divided into white pixels.

The pixels that belong to the *Arabidopsis* leaf are labeled as 1, and all the other background pixels are labeled as 0. According to the estimated values, the pixels can be divided into positive (estimated value: 1) and negative (estimated value: 0) samples. Also, if the estimated values are matched to the ground-truth values, the pixels are called true. Otherwise, they are called false. Accordingly, all the estimated pixels are determined as one of the following four: true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN). TP represents that the estimated and ground-truth values are both 1, while TN represents that both are 0. FP represents that the estimated value is 1, but it is not matched as the ground-truth value, while FN represents that
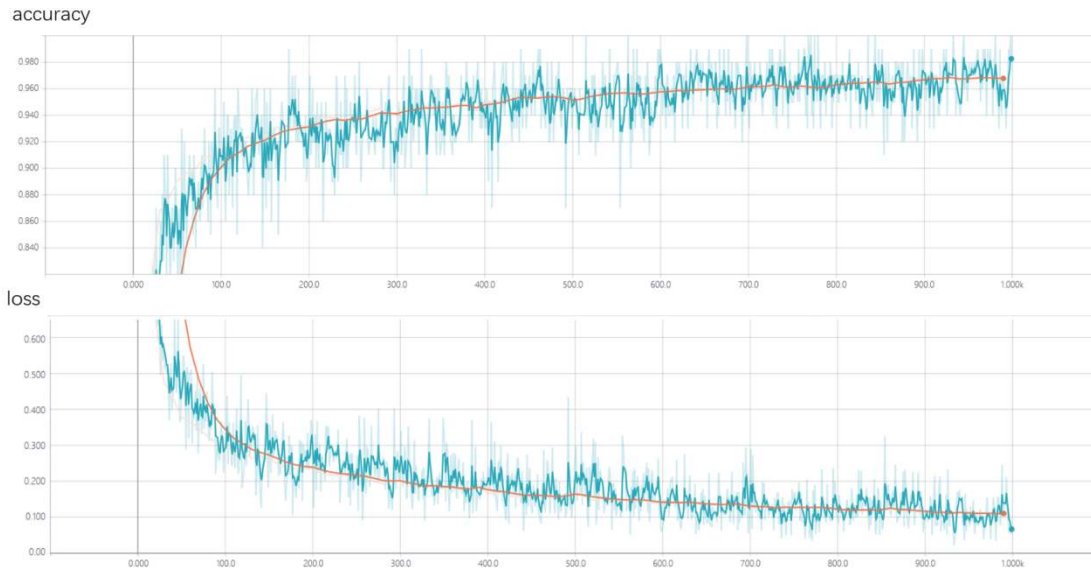
Figure 5. Model accuracy and loss during training.

the estimated value is 0, but it is not matched as the ground-truth value. The model realizes the binary segmentation for the canopy image of a single plant. Thus, TN does not affect the segmentation accuracy evaluation and is not included in the IoU evaluation. To ensure the robustness of the evaluation parameters, we randomly selected the mean values of 30 groups of data TP, FP, and FN as the evaluation parameters in the corresponding period. The recall rate R and IoU are defined as:

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$IoU = \frac{TP}{TP + FP} \tag{9}$$

In addition, the correlation coefficient of the segmented image is compared with the measured value according to the pixel precision.

Morphological Feature Extraction

The shape feature extraction from the segmented *Arabidopsis* plants is implemented using the OpenCV library. The description of morphological features can be divided into two categories: 1) contour-based shape description and 2) region-based shape description. The contour-based method describes the boundary contour of the target area. In contrast, the region-based method describes the shape of the target area through the geometric moment, area, and eccentricity. The minimum circumscribed circle, leaf center point, plant boundary frame, and plant contour of each plant after image segmentation are computed firstly, as shown in figure 6. Then, based on the basic features, the plant coverage, boundary frame area, circumscribed circle radius, leaf number, and other digital morphological features are extracted. The specific calculation equation is:

1. Circumscribed circle radius (r): On the contour of the plant, the longest distance between any two pixels:

$$r = \frac{max\left[dis\left(C_i, C_j \right) \right]}{2} \tag{10}$$

where C_i and C_j represent the two farthest pixels on the contour of the plant.

2. Bounding box area (S): Multiplication of the length and width in the plant bounding box as:

$$S = L \times H \tag{11}$$

where L is the length in the x-direction of the minimum bounding box, and H is the length in the y-direction. Plant bounding box refers to n rectangles with the smallest area, including the target plant area.

3. Coverage (C): Since the segmented image is a binary image, the pixel method is selected for coverage calculation, that is, the total number of pixels occupied by the target area is calculated:

$$C = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1} f\left(x, y \right) \tag{12}$$

In the binary graph, 1 and 0 are often used to represent the target and background, respectively. The calculation
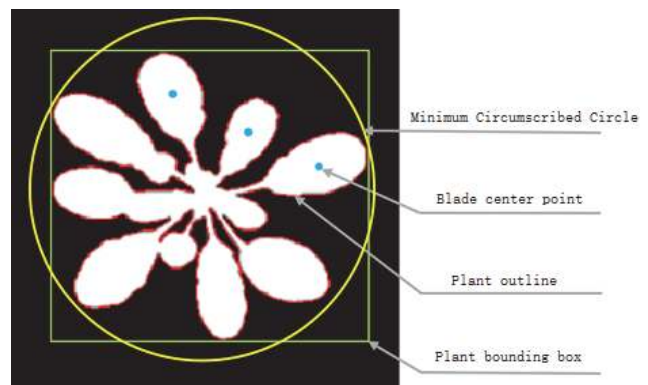

Figure 6. Extraction of phenotype parameters in segmentation results.

formula of the pixel method is to count the number of pixels with f(x, y) = 1. The area of each pixel can be calculated according to the length and width of the image. The coverage of plants in the image can be obtained by multiplying the total number of pixels in the region of interest after segmentation.

Color Feature Extraction Accuracy and Growth Analysis

Color cognition is a part of the basic research of the visual system, so color characteristics are an important parameter in the phenotype of plant leaves. The pixel value in the image consists of three channels: R, G, and B, which are arranged in a grid. In the process of data set collection, many factors will affect the quality of the captured image. Among them, the light intensity is more obvious, and different light intensity will change the value of the pixels in the image. In the original image, the light intensity of the V3 period and the V6 period are different, and the difference in the histogram of the green color channel in the image is also obvious, as shown in figure 7.

The segmented image is registered as a MASK with the original image, and the color feature RGB parameters of the MASK area are extracted and converted into HSV color space parameters. This approach is mainly because the HSV model in computer vision is based on the chroma, saturation, brightness, and other complex signals of the object to distinguish colors, which is similar to the principle of human eyes perceiving colors, and the HSV model can reduce the light intensity change to the color Distinguish the impact. The equation for converting specific RGB values into HSV space colors is:

$$H \begin{cases} 0° & , \Delta = 0 \\ 60° \times \left(\dfrac{\acute{G} - \acute{B}}{\Delta} + 0 \right) & , \acute{C}_{max} = R \\ 60° \times \left(\dfrac{\acute{B} - \acute{R}}{\Delta} + 0 \right) & , \acute{C}_{max} = G \\ 60° \times \left(\dfrac{\acute{R} - \acute{G}}{\Delta} + 0 \right) & , \acute{C}_{max} = B \end{cases} \tag{13}$$
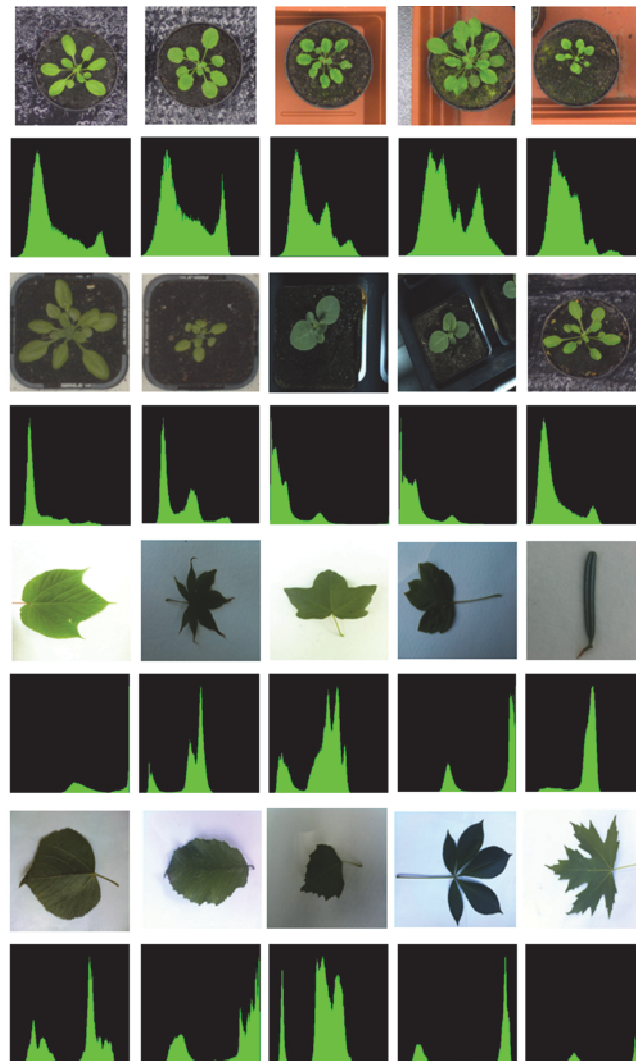

Figure 7. Color histogram of green color channel of some images in the testing set.

$$S = \begin{cases} 0 & , C_{max} = 0 \\ \dfrac{\Delta}{C_{max}} & , C_{max} \neq 0 \end{cases} \quad (14)$$

$$V = C_{max} \quad (15)$$

$$\acute{R} = R/255 \quad , \quad \acute{G} = G/255 \quad , \quad \acute{B} = B/255 \quad ,$$

$$C_{max} = max\left(\acute{R}, \acute{G}, \acute{B}\right) \quad , \quad C_{min} = min\left(\acute{R}, \acute{G}, \acute{B}\right) \quad ,$$

$$\Delta = C_{max} - C_{min}\circ$$

The extraction results of some leaf color parameters in the final test set image are shown in table 3.

Based on the results of the segmentation of plant image feature extraction method, the coverage, circumference, minimum circumscribed circle radius and bounding box area and other shape parameters were extracted in the test set composed of the CVPPP data set, and carried out in the test set composed of the Leafsnap data set Extraction of shape features such as leaf area, leaf length, leaf width, circumscribed circle radius, aspect ratio, shape factor, and density. In addition, this chapter also conducts parameter extraction based on color features in the test set, and finally focuses on the extracted data Related phenotypic analysis was carried out.

EXPERIMENTAL RESULTS AND ANALYSIS
Model Segmentation Effect

In this section, different activation functions are evaluated in the pixel-wise segmentation problem. The segmentation results are compared with the same network structure except for the activation function. The models are implemented in Tensorflow. The training time is about 45 minutes on a PC with two Intel Xeon (R) Gold 6148 CPUs (Santa Clara, Calif.), 256 GB RAM, and NVIDIA Quadro P6000 GPU (Santa Clara, Calif.). The segmentation results for different activation functions: ReLu, Sigmoid, Tanh, Leaky-ReLu are shown in figure 8. The images in the test set include various leaves with different leaf sizes, different light conditions, and different varieties of plant leaves.

Table 3. Test set image color feature parameter extraction results.

Color Feature	Dataset	Mean	Variance	Peak
R	A1	66.76	2969.66	-0.86
	A2	86.39	371.35	-0.48
	A3	73.13	3212.44	3.03
G	A1	78.22	2206.15	3.21
	A2	53.64	2469.14	8.63
	A3	107.23	1512.98	2.78
B	A1	95.42	4062.76	9.51
	A2	64.41	2880.03	4.05
	A3	69.03	305.62	1.15
H	A1	12.28	819.66	7.53
	A2	4.39	892.08	5.47
	A3	100.25	4001.71	7.45
S	A1	117.48	2576.66	-1.96
	A2	96.31	1032.23	7.12
	A3	56.45	3379.64	3.08
V	A1	77.51	512.34	2.76
	A2	18.92	1184.36	3.07
	A3	113.47	2365.98	4.46

In the research of computer image processing, image features, as the most basic attributes or characteristics to distinguish regions of interest in images, are a key part of related research. It is mainly composed of natural features that can be directly recognized by the human eye in an image and related features that are artificially defined by measuring or extracting the image after corresponding preprocessing. There are four basic methods for image feature extraction in the category of artificial features: shape, space, color, and texture.

According to the research object of this article, color feature and shape feature are selected for related phenotype extraction research. The CVPPP data set mainly extracts relevant features for the whole plant, while the Leafsnap data set extracts relevant features for plant leaves. The final extracted features will be subjected to analysis studies such as related growth analysis or comparison of differences between varieties.

The quantitative comparisons for different activation functions in terms of r and MIoU are summarized in table 4. As shown in figure 4, the model with the Leaky-ReLu provides the outperforming results over the other models. The recall rates R of the model are 0.94 and 0.96 for A1 and A2 datasets, respectively. The MIoU values are 0.93 and 0.95 for A1 and A2 datasets, respectively. The segmentation accuracy for A1 dataset is higher than that of A2 by 2%, which may be because the image quality of the A1 dataset is better than that of other datasets. Obvious features are conducive to model recognition. Nevertheless, the proposed model provides high accuracy of the segmentation for both A1 and A2 datasets regardless of the morphological structure and light conditions. For a single image, the segmentation time of the model is the same, about 1.2 s.

Shape Feature Extraction Accuracy and Growth Analysis

In the test set composed of the CVPPP data set, the average values of the above-mentioned phenotypic parameter extraction results are shown in table 5.

The performances of phenotypic parameter extraction are evaluated on 30 images selected in the test set. First, the images were segmented using the proposed model. Then, phenotypic parameters were extracted, including coverage, circumscribed circle radius, bounding box area, and perimeter. The extracted features were compared with the ones that were extracted from the Manual in segmentation data. The comparison results are shown in figure 9.

The vegetation coverage refers to the variation ratio of the vertical projection area of all plant canopy, branches, and leaves for crops, shrubs, trees, and weeds, in the ground of their growth-area (Zhang et al., 2017; Niu Yaxiao et al., 2018). In data processing, the dataset has been cropped to ensure the size of the background area is consistent. Therefore, the percentage of the leaf pixels in the entire segmented image is computed as the ratio of vertical projection area to the total background area. By multiplying the above proportion with the total area, the result covers different varieties in the same area. The plant growth is analyzed by taking the coverage as an example. The analyzed results are shown in figure 10. Note that the reason for not restoring to the real

Figure 8. Segmentation results for different activation functions. (a) the original image in the test set; (b) the ground-truth; (c)-(f): the segmentation results of the models using (c) ReLu; (d) Sigmoid; (E) Tanh; (f) Leaky-ReLu.

area is that the data set Beijing size is the same. The analysis based on the number of pixels does not affect the coverage analysis results.

In figure 10, (a) is the analysis result of coverage change of 10 different *Arabidopsis* plants (m1-m10) in A2 data set and (b) is the analysis result of coverage change of 5 different tobacco plants (m1-m5) in A3 data set. In the segmentation results of this research model, the coverage value in figure 9 is extracted and analyzed in pixels. The reason for not restoring to the real area is that the data set Background size is the same, and the analysis based on the number of pixels does not affect the coverage analysis results.

DISCUSSION

The extracted morphological features from the estimated segmentation are compared with ones from the ground-truth segmentation. The comparison shows that the four morphological feature parameters R2 are all higher than 0.96, and the NRMSE is less than 10%. It proves that the segmentation accuracy of the model is high enough to be used for further phenotypic extraction. Compared with the models in Wu

Table 5. Eigenvalues of plant shape parameters.

Datasets	Plant Area (cm²)	Plant Circumference (cm)	Minimum Circumscribed Circle Radius (cm)	Plant Bounding Box Area (cm²)
A1	165.132	124.725	11.282	274.514
A2	102.154	88.129	7.175	192.475
A3	121.734	116.524	9.248	202.756

et al. (2019) and Zhang et al. (2019), the Plant U-Net model not only provides higher segmentation accuracy than its improved K-means algorithm, but also has some advantages in coverage extraction. Also, compared with the model in Praveen Kumar and Domnic (2018), the proposed model can realize and batch image segmentation while avoiding a lot of human resources.

The proposed model is more robust and suitable for top view image segmentation of various crops over the methods. Moreover, the accuracy of the optimized U-Net model is higher than that of the FCN model in Zhe et al. (2018). As shown in figure 4, the proposed network can segment the leaves of *Arabidopsis* plants under different light conditions, different sizes of plants, and different types of plants. The

Table 4. Quantitative comparisons of segmentation accuracy for different activation functions.

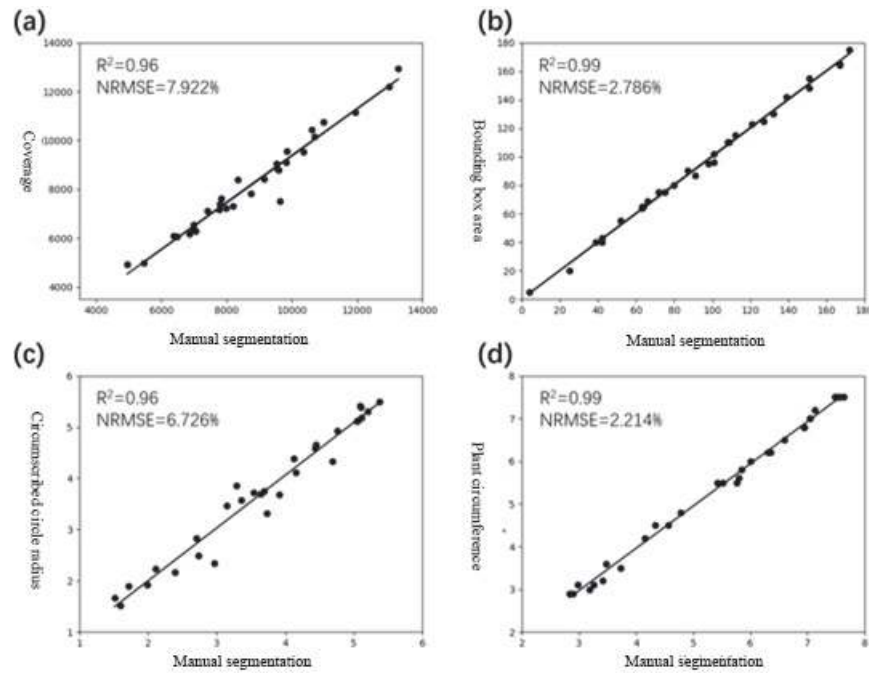| Dataset	ReLu		Sigmoid		Tanh		Leaky-ReLu	
	R	MIoU	R	MIoU	R	MIoU	R	MIoU
A1	0.91	0.90	0.85	0.84	0.82	0.81	0.94	0.93
A2	0.92	0.92	0.85	0.86	0.83	0.83	0.96	0.95
Average	0.915	0.91	0.85	0.85	0.825	0.82	0.95	0.94

Figure 9. Correlation comparison of phenotypic parameters extracted from the estimated segmentation and ground-truth: (a) the comparison of coverage extraction result, (b) the comparison of the extracted boundary box area, (c) the comparison of the extracted circumscribed circle radius, (d) the comparison of the extracted plant perimeter.

segmentation results for the A3 dataset also confirm the robustness of the proposed network for different lighting conditions, leaf size, and varieties of plants.

The numerical analysis of coverage extraction, where most of the segmentation values of the model are distributed in the upper half of the ground-truth segmentation value. It shows that the coverage values of the estimated segmentation are generally smaller than the ground-truth segmentation. This is mainly due to the following two reasons. First, there is FP judgment in the model, and the pixel points at the edge of the blade are not correctly determined (Wang et al., 2016). Second, some of the blades in the image are purely stuck with soil particles, overexposure, and shadow, affecting the recognition of the model. Also, there are a few cases in which the results of the estimated segmentation are worse than that of ground-truth segmentation, mostly because there are some weeds near the plant, the color of which is similar to the leaves so that the model cannot be recognized.

CONCLUSION

In this study, we propose a new U-Net model based on adjusting the model structure, activation function, and loss function of the U-Net segmentation network. The selected 120 images from multiple sub-datasets in the CVPPP dataset are used as training data to learn plant leaf segmentation. The remaining images are used as the test set in evaluating the segmentation performance of the proposed model. The experimental results show that the proposed segmentation network has strong edge detection ability in the A1 and A2 images. Also, the proposed model can accurately segment plant leaves while significantly reducing the workload that is usually required in the traditional hand-crafted methods. The morphological parameters extracted from the segmentation results are highly correlated with the measured values, which shows that the segmentation results of the model meet
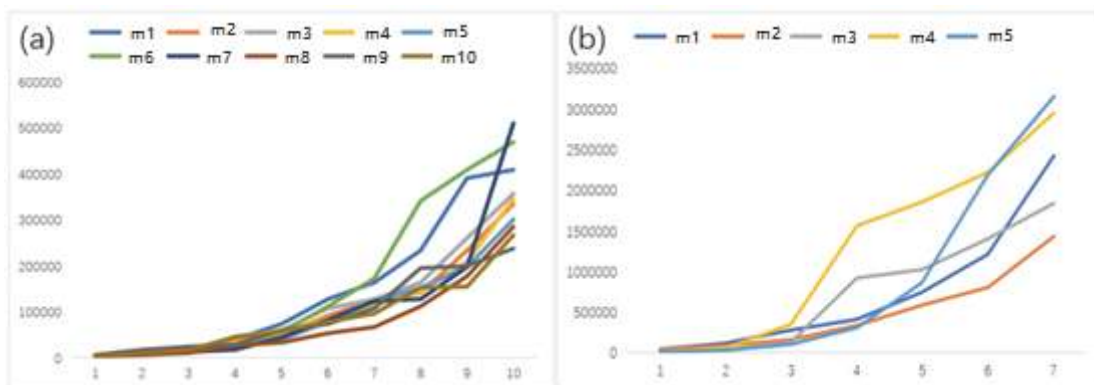

Figure 10. The analysis of plant coverage for different datasets: (a) Analysis for 10 different *Arabidopsis* plants in the A2 dataset, (b) analysis for 5 different tobacco plants in the A3 dataset.

the needs of phenotype extraction in terms of edge detection and segmentation accuracy. The proposed model can contribute to the segmentation of the top view image, the extraction of phenotype, and the evaluation of quantitative phenotype of greenhouse crops. It can be used as a valuable tool for breeding research.

In the numerical analysis of coverage extraction, the coverage value segmented by the model is generally smaller than the segmentation result. There are two main reasons for this situation. One is that there is an FP judgment in the model, and the pixels on the edge of the leaf are not correctly judged; the other is that some of the leaves in the image are purely soiled, overexposed, and shadows. And so on, affecting the recognition of the model. The current research results of this research are still limited to the analysis of the collected images, and real-time image analysis cannot be achieved. Take the phenotyping platform as an example. Generally speaking, the acquired plant image data has certain real-time analysis requirements. This research can achieve high-throughput segmentation of data concentration. The next step is to use the domestic 5G development wave and combine with the cloud data analysis platform to carry out the model to achieve real-time analysis in the true sense.

CONFLICTS OF INTEREST
The authors declare no competing interests.

DATA ACCESSIBILITY
The raw datasets and code are available in the Supporting Information.

AUTHOR CONTRIBUTIONS
Conceptualization, L. Cao and H. Yu; methodology, L. Cao and G. Chen; software, H. Li. and H. Wang; validation, L. Cao., H. Yu and H. Li.; writing—original draft preparation, L. Cao.; writing—review and editing, L. Cao and H. Yu.

REFERENCE
Bock, C. H., Poole, G. H., Parker, P. E., & Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Rev. Plant Sci., 29*(2), 59-107. https://doi.org/10.1080/07352681003617285

Chen, G., Jing, H., Li, D., & Si, X. (2012). Image recognition of maize diseases based on fuzzy clustering and support vector machine algorithm. *Sens. Lett., 10*(1-2), 433-438. https://doi.org/10.1166/sl.2012.1882

Goclawski, J., Sekulska-Nalewajko, J., & Kuzniak, E. (2012). Neural network segmentation of images from stained cucurbits leaves with colour symptoms of biotic and abiotic stresses. *Int. J. Appl. Math. Comput. Sci., 22*(3), 669-684. https://doi.org/10.2478/v10006-012-0050-5

Hao, F., Xuan, L., & Mu, C. (2017). Research on key technologies of intelligent agriculture based on agricultural big data. *Proc. Int. Conf. on Smart City & Systems Engineering.* https://doi.org/10.1109/ICSCSE.2016.0161

Kim, S. L., Solehati, N., Choi, I. C., Kim, K. H., & Kwon, T. R. (2017). Data management for plant phenomics. *J. Plant Biol., 60*(4), 285-297. https://doi.org/10.1007/s12374-017-0027-x

Kumar, J. P., & Domnic, S. (2020). Rosette plant segmentation with leaf count using orthogonal transform and deep convolutional neural network. *Machine Vision Appl., 31*(1), 6. https://doi.org/10.1007/s00138-019-01056-2

Lin, G., & Shen, W. (2018). Research on convolutional neural network based on improved Relu piecewise activation function. *Procedia Comput. Sci., 131*, 977-984. https://doi.org/10.1016/j.procs.2018.04.239

Ma, X., Deng, X., Qi, L., Jiang, Y., Li, H., Wang, Y., . . . Zhang, J. (2019). Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. *PLoS One, 14*(4). https://doi.org/10.1371/journal.pone.0215676

Mubin, N. A., Nadarajoo, E., Shafri, H. Z., & Hamedianfar, A. (2019). Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. *Int. J. Remote Sens., 40*(19), 7500-7515. https://doi.org/10.1080/01431161.2019.1569282

Niu Yaxiao, Z. L., & Han, W. (2018). Study on cotton coverage extraction method based on Lab color space. *J. Agric. Mach., 49*(10), 240-249.

Praveen Kumar, J., & Domnic, S. (2018). Image based leaf segmentation and counting in rosette plants. *Information Process. Agric., 6*(2), 233-246. https://doi.org/10.1016/j.inpa.2018.09.005

Qian, W., Sun, H., Li, M.-z., Song, Y.-y., & Zhang, Y.-e. (2015). Research on maize multispectral image accurate segmentation and chlorophyll index estimation. *Spectroscopy Spectral Analysis, 35*(1), 178-183.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Tekchandani, H., Verma, S., & Londhe, N. D. (2020). Mediastinal lymph node malignancy detection in computed tomography images using fully convolutional network. *Biocybernetics Biomed. Eng., 40*(1), 187-199. https://doi.org/10.1016/j.bbe.2019.05.002

Thomas, M., Dayananda, S., Nidamanuri, R. R., Nautiyal, S., Hanumaiah, N., Buerkert, A., & Wachendorf, M. (2018). Estimation of Vegetable Crop Parameter by Multi-temporal UAV-Borne Images. *Remote Sensing, 10*(5), 805. https://doi.org/10.3390/rs10050805

Wang, J. L., Yu, H., Zhao, S. S., Zheng, H. X., He, C., Cui, X. Y., . . . Wang, J. L. (2016). A new multi-scale analytic algorithm for edge extraction of strawberry leaf images in natural light. *Int. J. Agric. Biol. Eng., 9*(1), 99-108.

Wu, H., Cui, K., Zhang, X., Xue, X., Zheng, W., & Wang, Y. (2019). Fine leaf crop coverage extraction based on improved k-means image segmentation algorithm. *J. Agric. Mach., 50*(1), 42-50.

Zhang, J., Zhang, X., Meng, J., Hu, S., & Wu, L. (2017). Crop coverage extraction algorithm based on convolutional neural network with prior threshold optimization . *Signal Process., 33*(9), 1230-1238.

Zhang, P., Cui, M., Xie, Q., & Jiang, Y. (2019). Research on plant image recognition method based on deep convolution generation antagonism network. *J. Southwest Minzu University(Natural Sci. Ed.), 45*(2), 185-191.

Zhe, L., Zhang, X. L., Song, Y. Q., Zhu, Y., & Yuan, D. Q. (2018). Liver segmentation with improved U-Net and Morphsnakes algorithm. *J. Image Graphics, 23*(8), 1254-1262.

Zhou, C., Ye, H., Xu, Z., Hu, J., Shi, X., Hua, S., . . . Yang, G. (2019). Estimating maize-leaf coverage in field conditions by applying a machine learning algorithm to UAV remote sensing images. *Appl. Sci., 9*(11), 2389. https://doi.org/10.3390/app9112389