

Article

Determination Model of Epidermal Wettability for Apple Rootstock Cutting Based on the Improved U-Net

Xu Wang ¹, Lixing Liu ¹, Jinxuan Zou ¹, Hongjie Liu ^{1,2}, Jianping Li ^{1,2}, Pengfei Wang ^{1,2} and Xin Yang ^{1,2,*}

¹ College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding 071000, China; 20221090026@pgs.hebau.edu.cn (X.W.); 20211090024@pgs.hebau.edu.cn (L.L.); 20232090225@pgs.hebau.cn (J.Z.); liuhj@hebau.edu.cn (H.L.); ljpd527@hebau.edu.cn (J.L.); jdwpf@hebau.edu.cn (P.W.)

² Hebei Province Smart Agriculture Equipment Technology Innovation Center, Baoding 071001, China

* Correspondence: yangxin@hebau.edu.cn; Tel.: +86-159-3345-9619

Abstract: Keeping the epidermis of apple rootstock cuttings moist is important for maintaining physiological activities. It is necessary to monitor the epidermis moisture in real time during the growth process of apple rootstock cuttings. A machine vision-based discrimination model for the moisture degree of cuttings' epidermis was designed. This model optimizes the structure of the semantic segmentation model U-Net. The model takes the Saturation channel and Value channel information of the cutting images in the HSV color space as the characteristics of the cuttings' moisture, so that the model has good performance in the blue-purple supplementary light environment. The average accuracy of the improved model is 95.07% for dry and wet cuttings without supplementary light, and 84.83% with supplementary light. The humidification system implanted in the model can control the atomizer to complete the task of moisturizing the cuttings' epidermis. The average moisture retention rate of the humidification system for cuttings was 92.5%. Compared with the original model, the moisturizing effect of the humidification system increased by 26.87%. The experimental results show that the improved U-Net model has good generalization and high accuracy, which provides a method for the design of an accurate humidification system.

Keywords: epidermis moisture; semantic segmentation; U-Net; HSV; humidification system



Citation: Wang, X.; Liu, L.; Zou, J.; Liu, H.; Li, J.; Wang, P.; Yang, X. Determination Model of Epidermal Wettability for Apple Rootstock Cutting Based on the Improved U-Net. *Agriculture* **2024**, *14*, 2223.

<https://doi.org/10.3390/agriculture14122223>

Academic Editor: Margarita García-Vila

Received: 14 November 2024

Revised: 2 December 2024

Accepted: 2 December 2024

Published: 5 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The cutting propagation of apple rootstock can provide superior germplasm for molecular-assisted breeding with the advantage of a short cycle and the preservation of the mother plant's excellent traits. It is of great significance in the mining of important trait genes as well as for improving the efficiency of breeding programs [1,2]. The moisture level of the apple rootstock cutting epidermis is an external manifestation of its water content. Achieving precise monitoring of the cutting's epidermal moisture is an essential prerequisite for transpiration assessment, phenotypic trait analysis, and appropriate water supplementation [3]. The evaluation of the moisture level of apple rootstock cuttings' epidermis mainly relies on manual observation at present, which is a labor-intensive and inefficient method.

As a substitute for manual vision, machine vision can observe micrometer-scale targets without contact. The model trained by a neural network can capture required features rapidly from images which can accomplish the production task and improve production efficiency [4,5]. The RGB images are often regarded as input data for neural network training models [6], with the advantage of being convenient to collect and economically inexpensive. But they are prone to the influence of lighting [7]. During cutting trials in a plant incubator, blue-purple supplemental lighting is utilized for the fill light to promote photosynthesis in the cuttings' certain period. But the blue-purple light interferes with the expression of information in the RGB images. So it is necessary for the neural network

model to improve the resistance of light interference. The HSV images represent strong resistance to light noise in the target detection. Yao et al. [8] converted the input RGB data to HSV data, endowing the input data with the ability to resist light noise interference, which enhances the model's detection accuracy under complex lighting conditions. Li et al. [9] used the saturation and value information in HSV color space to quickly segment rice plant pixels and extract vegetation coverage in a complex field environment, reducing the impact of light changes on target detection to a certain extent. The above-mentioned method provides a solution approach for the difficulty in collecting target features under poor lighting conditions by selecting HSV color space. The classification of cutting epidermal wettability is a distinction of pixel-level optical characteristics, where the neural network model needs to effectively recognize and classify the pixel features of the images.

Image segmentation technology can assign semantic labels to every pixel in the image [10]. For pixel-level image classification and recognition, achieving end-to-end pixel-level segmentation provides a solution for the classification problem of cutting epidermal moisture [11]. With the rapid development of computer theory and hardware devices, image segmentation has become a research hotspot in the field of machine vision [12]. The PSP-Net network utilizes a featured pyramid structure, generating feature maps with different receptive fields through pooling layers of varying scales, and then aggregates these feature maps to obtain excellent global information perception capabilities [13]. Li et al. [14] improved the PSP-Net semantic segmentation model by integrating the tree skeletal information of the kiwifruit tree, which allows for efficient monitoring of the growth data of the kiwifruit tree canopy. The Deeplab network achieved accurate classification of every pixel in the image with the adoption of different techniques, involving multiple convolutional layers and dilated convolution techniques, which can perform well in dense dataset prediction tasks [15]. Cao et al. [16] put forward that the DeepLabV3+ neural network is improved by integrating channels and spatial attention mechanism module. This improvement can enhance the model's focus on particularly important features. The improved model achieved the segmentation of rice and weeds in complex backgrounds. U-Net is a convolutional neural network with a U-shaped architecture. This U-shaped neural network is often applied into image segmentation tasks in the medical field [17]. Liu et al. [18] trained a straw coverage detection model by way of U-Net. They employed the cross-entropy loss function as the loss function to improve the model's predictive performance. U-Net has shown excellent performance in various image segmentation tasks. Its effective feature extraction and fusion mechanisms enable it to perform well on small datasets [19]. Due to the scarcity of apple rootstock cutting epidermis image data, which constitutes a small dataset segmentation task, this study improved the U-Net to obtain the U-DSE-AG-Net model for the classification of cutting epidermis wetness. During the classification process of cutting epidermis wetness, it is challenging to capture the differences in pixel color channels due to varying degrees of wetness. Utilizing attention mechanisms to optimize the neural network structure can enhance the model's performance.

The attention mechanism is playing a crucial role in the field of deep learning [20,21]. The model is able to selectively pay attention to the key parts of input data by simulating the attention mechanism of the human visual system, which can process input data more efficiently [22,23]. The channel attention mechanism module focuses on the importance of different feature channels. By modeling the interdependency between channels, it automatically learns the significance of each feature channel and assigns different weight coefficients to each, thereby enhancing important features and suppressing unimportant ones [24]. Zhang et al. [25] introduced a channel attention mechanism that integrates low-resolution feature maps with strong semantic information from higher layers at the channel level, improving the detection effect of wheatear on the pyramid network model. Huo et al. [26] utilized the SE attention mechanism module to improve YOLOv5s' ability to identify obstacles, which can enhance the working efficiency of sugarcane harvesters.

To improve the classification performance of the model and reduce the impact of light changes on the outcomes, this study utilizes the HSV color space to obtain the

cutting images' information from the Saturation channel and the Value channel, which are more resistant to light noise interference. This can enhance the expressive effect of the moisture channel information in the cutting images and accomplish the preprocessing of input data. The U-DSE-AG-Net integrates the DSE module and the AG module into the skip connection layers of the U-Net. The DSE module was obtained by improving the SE attention mechanism module. In response to the requirements of the classification task of the epidermal wettability of apple cuttings, the loss function L_{ch} was designed, which can improve the segmentation accuracy and generalization ability of the neural network model. U-DSE-AG-Net is able to accurately segment cutting images in the condition of complex light noise interference and achieve the classification of cutting epidermal wettability.

2. Materials and Methods

2.1. Acquisition of Datasets

The apple rootstock cutting images data were obtained from the plant incubator in the Agricultural Intelligent Equipment and Information Laboratory at the Agricultural University of Hebei, Baoding, China, as shown in Figure 1. To ensure the correctness of the experiment conclusions, the designed dataset meets two requirements. It must include photos with different lighting conditions and epidermal wettability. RGB images of the cuttings were captured using a 1080P resolution camera with a shooting angle α of 15° , a vertical height h of 20 cm, and a horizontal distance d of 30 cm from the cell tray. A schematic diagram of the image data collection system is shown in Figure 2.



Figure 1. Apple rootstock cuttings in the plant incubator.

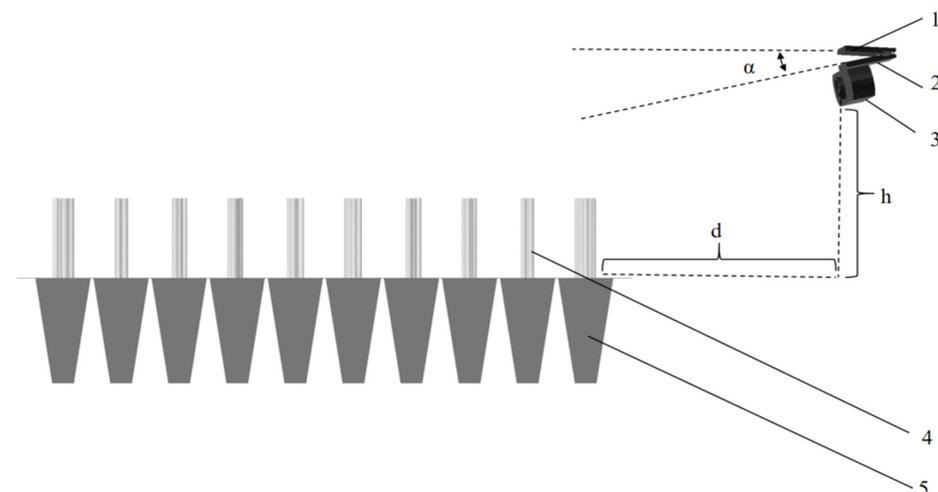


Figure 2. Schematic diagram of the cuttings image collection system 1: Camera bracket; 2: Camera bracket 2; 3: Camera; 4: Apple rootstock cuttings; 5: Cell tray.

During the cutting period of 0–20 days, apple rootstock cuttings require blue-purple light supplementary lighting to meet their growth and development needs. Due to the transpiration of the cuttings and the evaporation from the environment, the exchange

rate of moisture in the epidermal cells of the cuttings is relatively fast. It is necessary to monitor the moisture level of the cuttings' epidermis in real time during this stage and to maintain the cuttings' epidermis wetness to sustain their normal physiological activities. Images in this period were selected to create the dataset, with typical sample examples shown in Figure 3. Cuttings are divided into epidermis-dried cuttings and epidermis-wet cuttings. Epidermis-dried cuttings means that, after undergoing sufficient evaporation and the physiological activities of the cuttings, the cuttings' epidermis has no water film; epidermis-wet cuttings refer to how, after full atomization and humidification, the cuttings' epidermis adheres to a layer of water film. The thickness of the water film is about 0.01 mm. During the whole process of image collection, cutting images were divided by type of light source and epidermal wettability. Eventually, 1200 original RGB cutting images were obtained, including 300 images of epidermis-dried cuttings without supplementary lighting, 300 images of epidermis-moist cuttings without supplementary lighting, 300 images of epidermis-dried cuttings with supplementary lighting, and 300 images of epidermis-moist cuttings with supplementary lighting. The size of the cropped images was set to 512×512 pixels. When the cutting images were collected, the environment parameters were specified, such as air temperature at $24\text{ }^{\circ}\text{C}$, air humidity at 85%, substrate temperature at $26\text{ }^{\circ}\text{C}$, and substrate moisture at 60%.

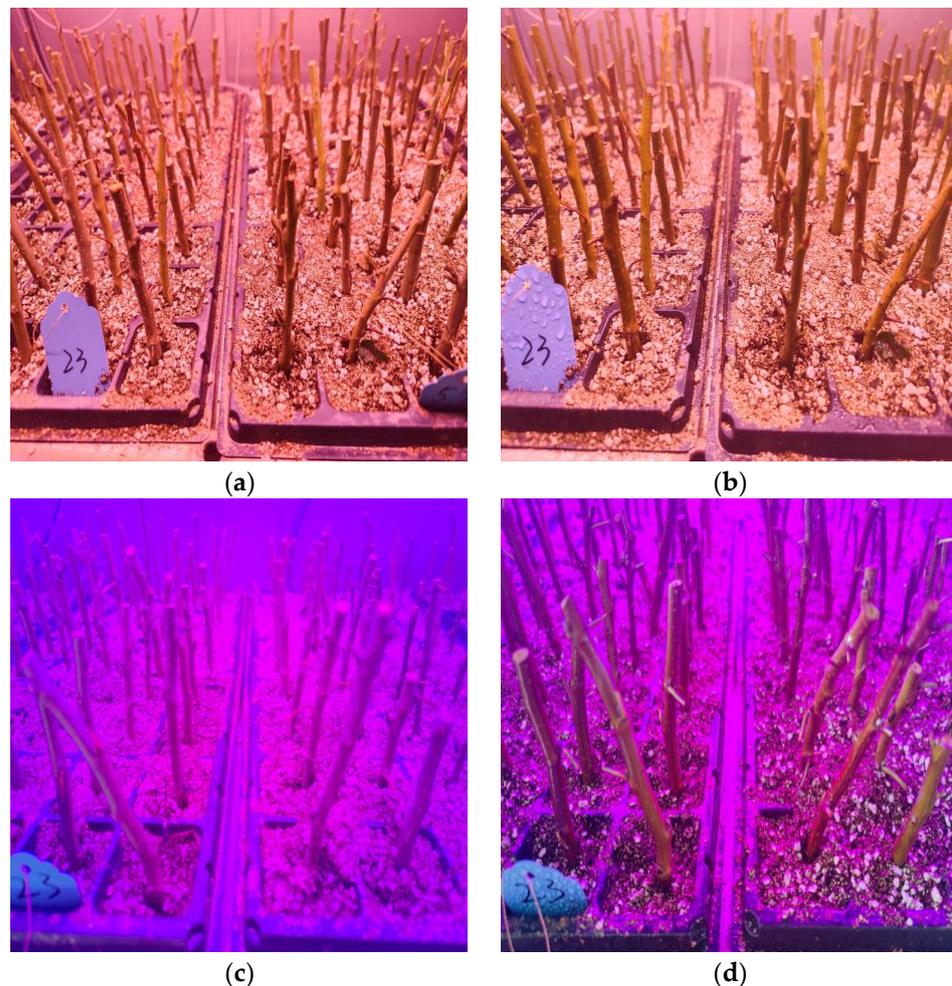


Figure 3. Images of apple rootstock cuttings. (a) Epidermis-dried cutting image without supplementary lighting; (b) Epidermis-wet cutting image without supplementary lighting; (c) Epidermis-dried cutting image with supplementary lighting; (d) Epidermis-wet cutting image with supplementary lighting.

2.2. Data Enhancement and Expansion

Dosovitskiy et al. [27] demonstrated the value of data augmentation in learning invariance for neural networks in feature learning. Data augmentation can improve the model's generalization and robustness, reduce overfitting, enhance the accuracy of model predictions, and save on data collection costs [28]. In order to achieve precise prediction of the moisture level of cutting epidermis by the model, this study increases the number of images in the existing dataset through random rotation geometric transformations. The rotation angle of the images is 180° . The initial 1200 photos were expanded to 2400 by using this method. Labelme was used to note the cutting images, which were divided into four main categories, including Wet cutting and Dry cutting in the environment without fill light and blue-purple fill light. The dataset was divided into training, testing, and validation sets in an 8:1:1 ratio to support model training.

2.3. Selection of Color Space

Selecting the appropriate color space is fundamental for neural networks to perform image segmentation for the image data input to the neural network [29]. Currently, common color spaces mainly include RGB, HSV, and so on [30,31]. The RGB image uses changes in the red (R), blue (B), and green (G) color channels and their superposition of one another to conduct feature description on the images [32]. According to Figure 3, the image is divided into four types, including the epidermis-dried image without supplementary lighting (a), the epidermis-wet image without supplementary lighting (b), the epidermis-dried image with supplementary lighting (c), and the epidermis-wet image with supplementary lighting (d). Visual processing was performed on the cutting images of the same genotype from Figure 3, resulting in Figure 4, which shows the histograms of the R, G, and B color channels of the cutting images. Each color channel is treated as an independent array, and the pixel counts for the corresponding channels are calculated to obtain (a), (b), (c), and (d). In Figure 4, comparisons between (a) and (c), and between (b) and (d), respectively, indicate that the number of pixels in the R channel of the cutting images increases dramatically after the addition of supplementary lighting, while the pixel counts in the G and B channels are approximately zero. This suggests that the supplementary lighting has a significant impact on the distribution of the RGB color channel histograms of the cutting images. From the comparisons between (a) and (b), and between (c) and (d), we found that the histograms of the corresponding color channels before and after humidification of the cutting show a small difference in area and an inconspicuous change in peak shape. Figure 4 illustrates that the RGB color channel information of the cutting epidermis differs slightly before and after humidification. Therefore, taking RGB images directly as the dataset for training the model is not conducive to classifying the surface moisture level of the cutting.

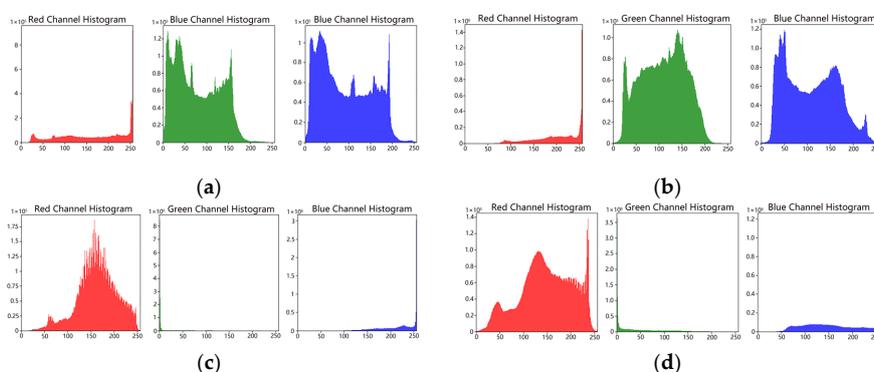


Figure 4. RGB channel histograms of the apple rootstock cutting images. (a) RGB channel histograms of the epidermis-dried cutting image without supplementary lighting; (b) RGB channel histograms of the epidermis-wet cutting image without supplementary lighting; (c) RGB channel histograms of the epidermis-dried cutting image with supplementary lighting; (d) RGB channel histograms of the epidermis-wet cutting image with supplementary lighting.

As depicted in Figure 5, the water film adhering to the cutting's epidermis absorbs and refracts a portion of the light in actual production. The water film alters the path of light transmission and reduces the reflectivity of the cutting's epidermis. This leads to a decrease in the value of the cutting images. The scattering effect of the water film on light makes the color components of the cutting's epidermis more salient and the saturation more diverse.

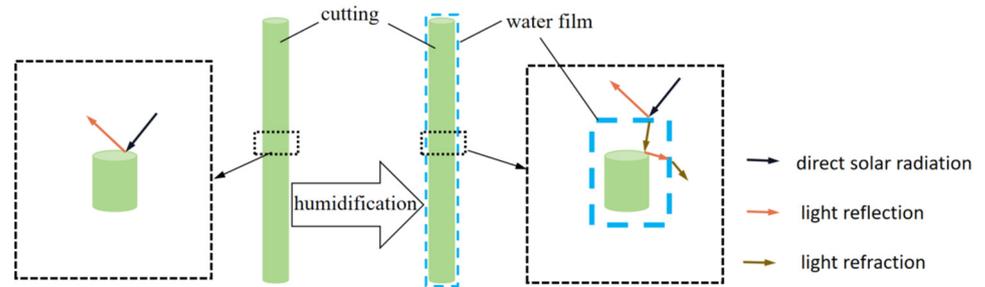


Figure 5. Schematic diagram of light transmission paths before and after humidification of the cuttings.

The HSV color space separates the image into three components: Hue, Saturation, and Value [33,34]. The conversion process of the RGB color space to the HSV color space is as follows: The process of normalizing the values of the three channels of an RGB image from the range [0, 255] to [0, 1] is given by Equation (1):

$$\begin{cases} R' = \frac{R}{255} \\ G' = \frac{G}{255} \\ B' = \frac{B}{255} \end{cases} \quad (1)$$

R' , G' , B' represent the normalized results of the R, G, and B color channels. Value is the maximum value among RGB, showing the value of color. Its calculation process is given by Equation (2):

$$V = \max(R', G', B') \quad (2)$$

The calculation of Saturation is given by Equation (3):

$$S = \begin{cases} \frac{V - \min(R', G', B')}{V} & V \neq 0 \\ 0 & V = 0 \end{cases} \quad (3)$$

Hue depends on the maximum Value channel of RGB. Suppose $M = \max(R', G', B')$, $m = \min(R', G', B')$, so the calculation of Hue is given by Equation (4):

$$H = \begin{cases} 60^\circ \times \frac{G' - B'}{M - m} & M = R' \\ 60^\circ \times (2 + \frac{B' - R'}{M - m}) & M = G' \\ 60^\circ \times (4 + \frac{R' - G'}{M - m}) & M = B' \\ 0 & M = m \end{cases} \quad (4)$$

Figure 6 represents the histograms of the H, S, and V channels of the cutting images, obtained (a), (b), (c), (d) after normalization treatment.

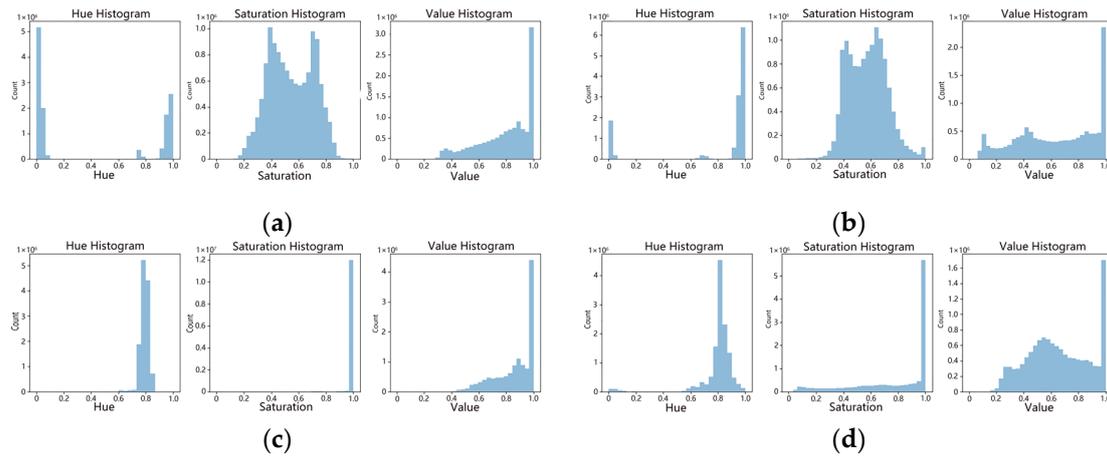


Figure 6. HSV channel histograms of the NO. 23 apple rootstock cutting images: (a) HSV channel histograms of dry epidermis without supplementary light cuttings images; (b) HSV channel histograms of wet epidermis without supplementary light cuttings images; (c) HSV channel histograms of dry epidermis with supplementary light cuttings images; (d) HSV channel histograms of wet epidermis with supplementary light cuttings images.

With the comparison of (a) and (b), (c) and (d) in Figure 6, the variation range of the tonal value of the images is small after humidifying the dried cuttings. This proved that the Hue channel had a poor effect on the wetness of the cuttings, which was not suitable for classifying the wetness of the cuttings. The information of the images in the HSV color space can be characterized by Equation (5).

$$\begin{cases} \varphi_i = \frac{1}{N} \sum_{j=1}^N P_{ij} \\ \psi_i = \left[\frac{1}{N} \sum_{j=1}^N (P_{ij} - \varphi_i)^2 \right]^{\frac{1}{2}} \end{cases} \quad (5)$$

P_{ij} is the i th color channel of the j th pixel of the image, and N is the number of pixels. φ_i reflects the average value of each color channel. ψ_i defines the dispersion degree of pixel values within the color channel. Equation (5) can reflect the information characteristics of the histogram of the image in the HSV color space. These characteristics are the main basis for the model to classify the moisture degree of cuttings' epidermis.

The pixel distribution range of the Saturation channel in the inserted spike image (b) increased from [0.04, 1.00] to [0.02, 1.00] compared to the image (a) in Figure 6. And the pixel distribution range of the Saturation channel in the inserted spike image (d) increased from [0.61, 1.0] to [0.05, 1.0] compared to the image (c) in Figure 6. The saturation value of cuttings with wet epidermis was higher than that of cuttings with dry epidermis, proving that the images of the cuttings with wet epidermis showed more saturation with more vivid and saturated color. The pixel distribution range of the Value channel in the inserted spike image (b) increased from [0.33, 1.00] to [0.08, 1.00] compared to the image (a) in Figure 6. And the pixel distribution range of the Value channel in the inserted spike image (d) increased from [0.41, 1.0] to [0.15, 1.0] compared to the image (c) in Figure 6. The Value of cuttings with moist epidermis is lower than that of cuttings with dry epidermis, indicating that cuttings with moist epidermis have lower Value and darker images. After the cuttings were humidified, the water film covered on the surface changed the light characteristics. The number of low Value pixels increased significantly, resulting in a decrease in the overall Value of the image, which was consistent with the actual performance of the cuttings before or after humidification. The number of pixels of HSV channels in cuttings images still conforms to this rule in the environment with supplementary light.

Therefore, converting the RGB image of cuttings into HSV color space is conducive to the understanding and recognition of the wetness of the epidermis of cuttings in the subsequent image segmentation process. The neural network can complete the wetness classification of cuttings by learning the information from the Saturation channel and Value channel of the images.

2.4. U-DSE-AG-Net Neural Network

2.4.1. Neural Network Backbone Structure and Improvement

U-DSE-AG-Net is based on the improvement of U-Net. U-Net is a kind of encoder–decoder segmentation network proposed by Ronneberger et al. [35], which preserves the low-level details of the encoder and the high-level semantic information of the decoder through hopping connection layer. As shown in Figure 7, the U-Net network architecture consists of a contraction path (left) and an expansion path (right). The shrink path is a typical convolutional network architecture, while the 3×3 convolutional kernel is repeatedly applied for two convolutional operations. After each convolution operation, the ReLU activation function is used for data operation, making data with nonlinear characteristics, and then 2×2 maximum pooling is carried out to realize a down-sampling operation with step length of 2. In each down-sampling step, the number of feature channels is doubled compared to the previous layer, the number of channels in f_1 feature layer is 64, and the number of channels in f_5' feature layer is 512 after three operations. When the expansion path returns from the f_5' layer to the f_1' layer, skip connection was used to connect corresponding contraction paths of the same layer. The feature layers of the two paths are stacked to achieve fusion. After the fusion, the feature layer is performed twice with the convolution kernel of 3×3 , and then the 2×2 convolution check data are used to carry out the “transposed convolution” operation layer by layer to realize the upward transmission of data. Finally, the feature layer with the same size as the input image is obtained.

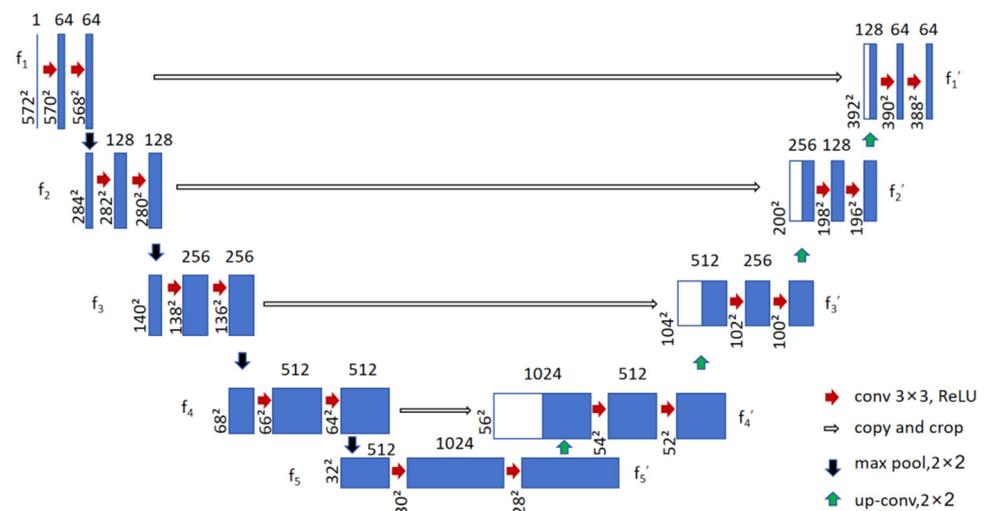


Figure 7. U-Net backbone architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is indicated at the top of the box. The white boxes represent the feature maps that have been copied. The arrows represent different operations. f_{1-5} : The number of layers 1–5 of the encoder contraction path; f_{1-5}' : Number of decoder extension path layers 1–5.

As shown in Figure 8, in order to improve the attention of U-Net to important features, the Double Squeeze and Excitation (DSE) networks and the Attention Gate (AG) were embedded in the skip connection layers between the encoder and the decoder of U-Net structure, which results in U-DSE-AG-Net. U-DSE-AG-Net architecture can effectively improve the performance of the model when performing the task of cutting epidermis wetness classification, benefiting from these two modules.

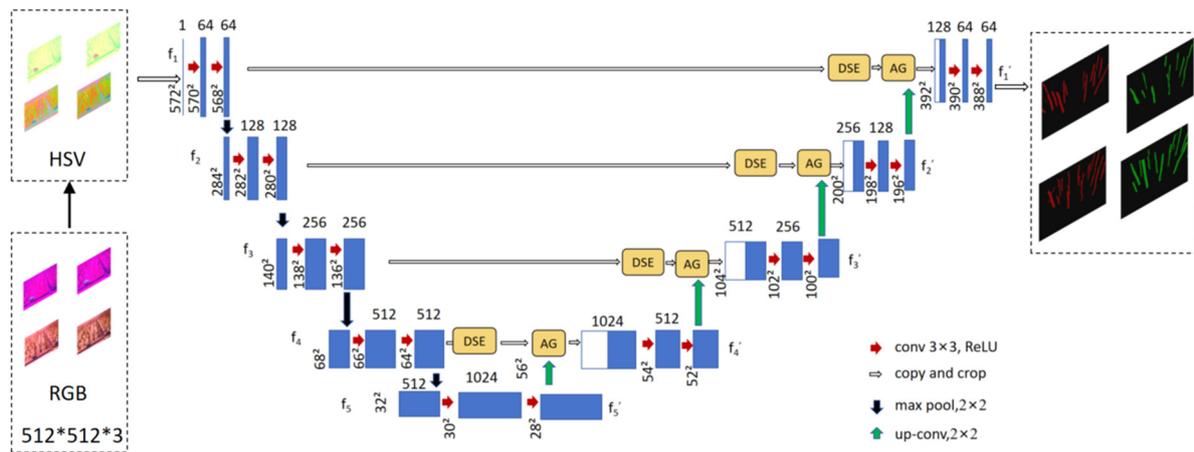


Figure 8. U-DSE-AG-Net network architecture.

The DSE attention mechanism module is based on the improvement of the SE attention mechanism module, which classifies tasks according to the wetness of cuttings’ epidermis. The SE module (Squeeze-and-Excitation networks) is the channel attention mechanism module [36]. By focusing on the characteristics of the saturation and channel value information of the cuttings, the neural network can learn better, so that the neural network can pay more attention to the channel information of the image. The Attention Gate (AG) module is inspired by the additive attention model [37], and the gating unit is introduced to realize dynamic selection and weighting of input features, and the attention coefficient determines the importance of input features, so that the model can focus on the information that is more critical to that task. The specific operation process of the model on the data is shown in Table 1. The arrow in Table 1 indicates the direction of data transmission.

In this study, the model is expected to pay more attention to the important features of epidermis wetness in the cuttings under the condition of blue-purple supplementary light. In the case of chaotic background together with blue and purple light as interference noise, more characteristic information of cuttings epidermis wetness should be obtained during model training. Non-target information such as the background information in the plant incubator should not play an important role in the epidermis wetness classification of cuttings. The classification task is mainly based on the feature information of the saturation and value of the cuttings’ images, so the model should pay attention to the channel information of the feature layer rather than the spatial information.

Compared with the Convolutional Block Attention Module (CBAM), the SE and ECA Efficient Channel Attention (ECA) modules only take the channel dimension into account without capturing features in the spatial dimension to save computational costs. Compared with the ECA module, the SE module is more suitable for the scene with a large number of channels. By learning the correlation between channels, the SE module can filter out channel-specific information and enhance the expression ability of the model. With enough resources, the SE module is a better choice [38].

In the process shown in Figure 9, the SE module network structure is divided into two steps, including Squeeze and Excitation. The SE attention mechanism module performs convolution in the input feature layer X to obtain the feature mapping U of dimension $H \times W \times C$. The feature mapping U is compressed by $F_{sq}(\cdot)$. When compressed, the spatial dimension of U is compressed to 1×1 , and the number of channels remains unchanged. Each channel obtains a path descriptor, and this descriptor function generates a feature vector containing global information about the channel. This $1 \times 1 \times C$ vector is the output of the compression operation, which will be the input data for the subsequent excitation operation. The excitation operation F_{ex} adopts a self-selection transfer mechanism, which takes $1 \times 1 \times C$ eigenvector as input and outputs the weights of each channel. F_{scale} is a cumulative operation. By multiplying channels by channels, the weights generated by

excitation operations are applied to the original feature graph to adjust the importance of different channels, and the generated outputs are used for subsequent layers of the network.

Table 1. Parameter calculation flowchart of U-DSE-AG-Net. ↓: Data transmission to the lower network; ↑: Data transmission to upper network; →: Data transmission to the right network.

| Encoder | Skip Connection Layer | Decoder |
|--|--|---|
| Input (512, 512, 3) ↓ | | |
| f_1 Conv2d filters = 64 (512, 512, 64) Conv2d filters = 64 (512, 512, 64) Maxpoolings = 2 (256, 256, 64) | → DSE (512, 512, 64) → AG (512, 512, 128) → ↑ UpSampling2D (512, 512, 128) | Conv2d filters = 64 (512, 512, 64) Conv2d filters = 64 (512, 512, 64) Concatenate (512, 512, 192) f_1' |
| ↓ | | |
| f_2 Conv2d filters = 128 (256, 256, 128) Conv2d filters = 128 (256, 256, 128) Maxpoolings = 2 (128, 128, 128) | → DSE (256, 256, 128) → AG (256, 256, 256) → → ↑ UpSampling2D (256, 256, 256) | UpSampling2D (512, 512, 128) Conv2d filters = 128 (256, 256, 128) Conv2d filters = 128 (256, 256, 128) Concatenate (256, 256, 384) f_2' |
| ↓ | | |
| f_3 Conv2d filters = 256 (128, 128, 256) Conv2d filters = 256 (128, 128, 256) Conv2d filters = 256 (128, 128, 256) Maxpoolings = 2 (64, 64, 256) | → DSE (128, 128, 256) → AG (128, 128, 512) → → ↑ UpSampling2D (128, 128, 512) | UpSampling2D (256, 256, 256) Conv2d filters = 256 (128, 128, 256) Conv2d filters = 256 (128, 128, 256) Concatenate (128, 128, 768) f_3' |
| ↓ | | |
| f_4 Conv2d filters = 512 (64, 64, 512) Conv2d filters = 512 (64, 64, 512) Conv2d filters = 512 (64, 64, 512) Maxpoolings = 2 (32, 32, 512) | → DSE (64, 64, 512) → AG (64, 64, 512) → ↑ UpSampling2D (64, 64, 512) | UpSampling2D (128, 128, 512) Conv2d filters = 512 (64, 64, 512) Conv2d filters = 512 (64, 64, 512) Concatenate (64, 64, 1024) f_4' |
| ↓ | | |
| f_5 Conv2d filters = 512 (32, 32, 512) Conv2d filters = 512 (32, 32, 512) Conv2d filters = 512 (32, 32, 512) | → | UpSampling2D (64, 64, 512) ↑ f_5' |

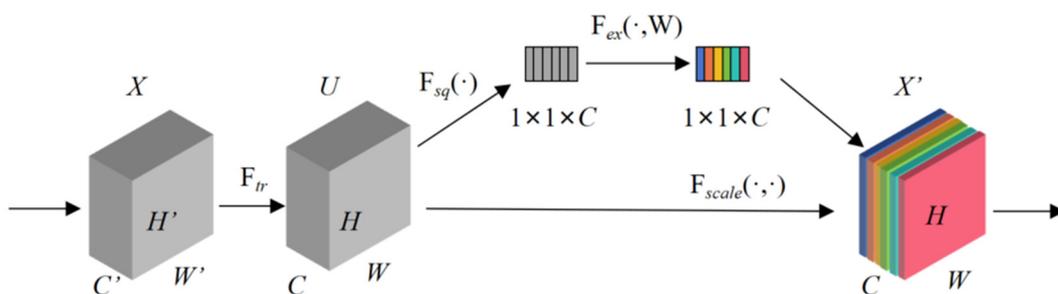


Figure 9. Squeeze and Excitation block.

(1) Squeeze

The input feature map is average pooled at the global level, and the spatial dimension (i.e., width and height) information is compressed into a channel descriptor at that level,

which is represented as a direct summation of the spatial dimensions. Equation (6) describes the extruded eigenvector on the c channel. Channels are average pooled at the global level and the entire spatial information input to this level is squeezed into a channel descriptor, which is expressed as a direct summation over and division by the spatial dimensions.

$$F_{sq}(u_c) = \frac{1}{H \times W} \sum_{a=1}^H \sum_{b=1}^W u_c(a, b) \tag{6}$$

F_{sq} represents the compression operation, $H \times W$ is the number of channels in the input feature map, and u_c represents the feature mapping of the input feature map in (a, b) space.

(2) Excitation

The excitation operation calculates the weight of the input feature map to make the network adaptively learn the activation difficulty of each channel, which highlights the useful features and improves the accuracy of the network. The process is shown in Equation (7).

$$s_c = \sigma(W_2\delta(W_1u_c)) \tag{7}$$

When the input vector u_c enters the excitation stage, it passes through two fully connected layers, where W_1 and W_2 are the weights of the first fully connected layer and the second fully connected layer, respectively. σ and δ , respectively, represent Sigmoid function and ReLU activation function. After the excitation operation, a vector s_c of $1 \times 1 \times C$ is obtained, where each element corresponds to the weights of the corresponding channels in the original feature graph, and these weights are applied to the feature graph U for F_{scale} operation, as shown in Equation (8).

(3) Scale Operation

$$X_c = F_{scale}(U_c, s_c) = U_c \cdot s_c \tag{8}$$

$F_{scale}(U_c, s_c)$ represents the multiplication of the vector s_c and the feature graph U_c . X_c represents the final output feature map of the SE attention mechanism module. Figure 10 shows the operation flow of image data in the SE network.

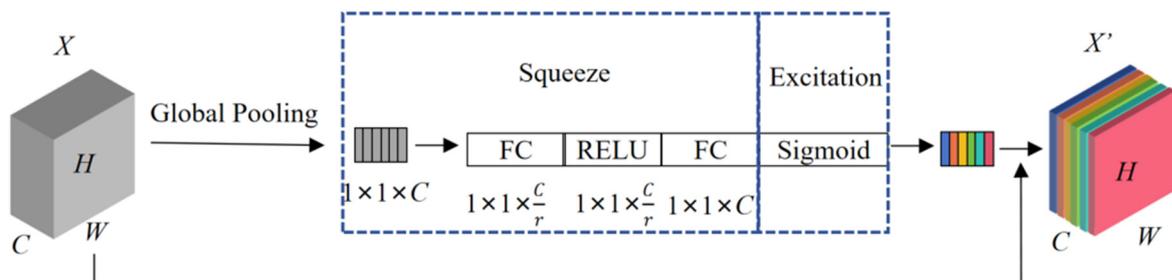


Figure 10. Schematic diagram of SE net structure.

In order to improve the attention mechanism module’s attention degree to channels, a Squeeze path F_{sq2} is added to the original SE module to obtain the DSE attention mechanism module (Double Squeeze and Excitation networks), as shown in Figure 11. The compression paths F_{sq1} and F_{sq2} globally average pool the feature layers U , calculate the feature layers u_1 and u_2 with the number of channels $C/2$, and sum u_1 and u_2 point to point to obtain the feature vector $sq_c = 1 \times 1 \times C$ before excitation. In the two compression processes, the channel information of feature mapping sq_c with key features is preserved, which improves the probability of key features appearing. The feature mapping sq_c performs excitation again and outputs the feature vector s_c .

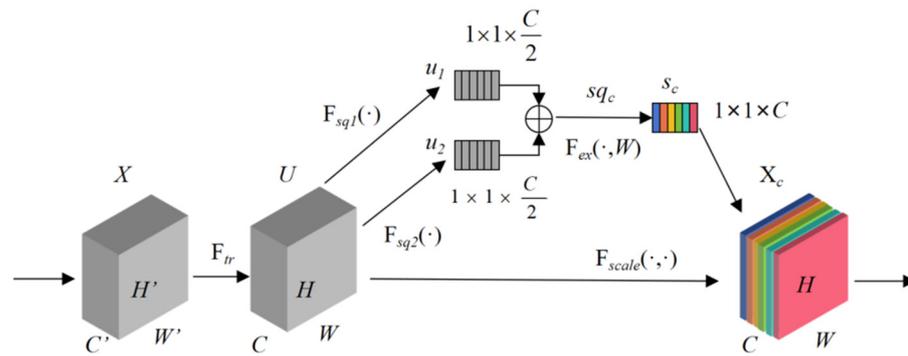


Figure 11. Double Squeeze and Excitation networks.

The DSE attention mechanism module is embedded in the skip connection layer after the encoder convolution, and while contributing weight to the feature channel of the input image, it learns the features of different categories in the image to improve the classification accuracy. Compared with the output layer embedded in the network, placing the DSE module in the skip connection layer will enrich the image channel features in the encoder unit, and the extracted high-level semantic information can be more effectively fused with the decoder information, so that the model can better obtain segmentation performance.

Attention gating mechanism module (Attention Gate, AG) is a deep learning-enhanced model. The network structure before and after improvement is shown in Figure 12, and the input information is X_c and X' .

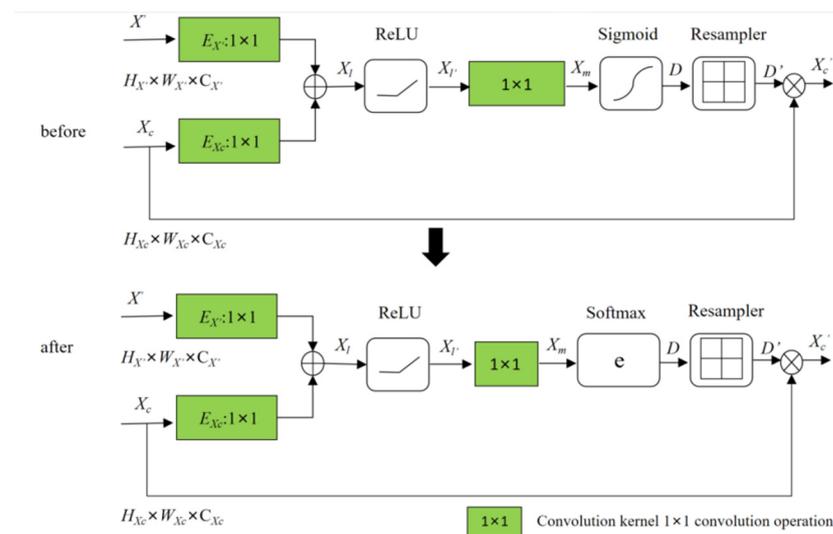


Figure 12. Attention Gate before and after module structure diagram.

The data dimension input by the AG module is determined by the output data of the DSE module of each skip connection layer. At the same time, it obtains parameters from the neural network UpSampling2D data of the deep layer. The dimensions of these two input data are consistent. The data dimension relationship can be reflected in Figure 8 and Table 1. $X_c(H_{Xc} \times W_{Xc} \times C_{Xc})$ is the data after the DSE attention mechanism module, including the information of the Saturation channel and Value channel of the feature map; X' comes from the data uploaded by the decoder one layer deeper than the DSE attention mechanism module. Because the depth is deeper than the feature layer X_c , it contains more semantic information and can more accurately guide the model to learn the key information in the feature layer X_c . The number of channels in the feature layer $C_{X'} = 2C_{Xc}$, so the feature layer X' is down-sampled, and the feature layer X' and X_c are, respectively, convolved by 1×1 convolution to ensure that the dimensions of the output feature layer $E_{X'}$ and

E_{X_c} are consistent. $E_{X'} \oplus E_{X_c}$ means that the two feature layers are added point by point to calculate the middle feature layer X_l . The ReLU function is used to carry out nonlinear operation on the feature layer X_l and output the feature layer X_l' , and the 1×1 convolution kernel is used to carry out convolution operation, and the output feature layer X_m realizes the fusion of the two feature layers. The AG attention gating mechanism module uses Sigmoid function to obtain the attention coefficient vector matrix D for the fused feature layer. The Resampler module restores the dimensions of the attention coefficient vector matrix D , and the restored vector matrix D' and the feature graph X_c weighted operation " \otimes " output the subsequent feature layer $X_{c'}$.

Since both ends of the Sigmoid function are close to saturation when the values are 0 or 1, the gradient in these regions is almost 0. This local gradient and the weighted operation result of the channel data corresponding to the feature graph X_c are 0, resulting in the loss of important features. Therefore, in order to improve the performance of the AG module, the Sigmoid function is replaced by Softmax function to calculate the feature attention coefficient. The Softmax function is defined as Equation (9):

$$S(x_{ij}) = \frac{e^{x_{ij}}}{\sum_{k=1}^n e^{x_k}} (k = 1, 2, \dots, n) \quad (9)$$

" e " is the base of the natural logarithm, x_{ij} is the value of the (i, j) element in the feature layer X_m , n is the total number of categories, and $S(x_i)$ is the feature attention coefficient of element x_i . The Softmax function uses an exponential function whose output D' is a non-singular matrix. The matrix D' element corresponds to the value of $[0, 1]$, and the weighted operation result of the channel data corresponding to the feature graph X_c is not 0, so as to avoid the loss of important features.

2.4.2. Loss Function Selection and Improvement

When training a neural network model, the loss function is a key indicator to evaluate the difference between the predicted value and the actual value of the model. The dice loss function is applicable to class-unbalanced segmentation tasks, while the cross-entropy loss function is applicable to class-balanced segmentation tasks.

In the testing process of the apple rootstock cuttings' epidermis wetness classification model, the pixels of cuttings and background images need to be divided on the Hue channel, with about a 1:9 ratio between cuttings and background pixels. This classification task is an imbalance segmentation task. When there is a significant class imbalance in the dataset, the dice loss function focuses directly on the overlap between the predicted results (cuttings vs. background) and the true labels, making the model more focused on the segmentation effect on the minority category (cuttings). In the task of segmentation of cuttings and background images, the dice loss function adopts continuous probability value, which is defined as Equation (10):

$$L_{Di}(p, t) = 1 - \frac{2 \sum p_{ij} t_{ij} + \varepsilon}{\sum p_{ij} + \sum t_{ij} + \varepsilon} \quad (10)$$

" p " represents the probability value predicted by the model, " t " represents the binary value (0 or 1) of the true label of the cuttings image, "0" represents the cuttings with epidermis wetness, and "1" represents the cuttings with dry epidermis. p_{ij} and t_{ij} represent the predicted and true value of the (i, j) pixel, ε is a small smoothing term that should be avoided with a denominator of zero. The dice loss function improves the segmentation accuracy of cuttings and background by maximizing the overlap between model prediction results and real labels.

After dividing the pixels of the cuttings, the model needs to distinguish the wetness degree of the cuttings' epidermis. The number of all kinds of cutting images should be the same, and the environmental parameters should be consistent when collecting images. And the category distribution of the images of cuttings with wet epidermis and dry epidermis

should be balanced. Therefore, each pixel of the two kinds of images should be carefully compared to realize the classification of the wetness degree of the cuttings. When the class distribution of the dataset is relatively balanced, the cross-entropy loss function will calculate the loss of each pixel independently, which is conducive to the model capturing the details of the image. And the cross-entropy loss function is suitable for the scene of fine segmentation boundary. In the binary classification problem of cuttings with wet and dry epidermis, the cross-entropy loss function is defined as Equation (11):

$$L_J(y, \hat{y}) = -\frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W [y_{ij} \log(\hat{y}_{ij}) + (1 - y_{ij}) \log(1 - \hat{y}_{ij})] \quad (11)$$

This function directly measures the difference between the probability distribution predicted by the model and the true label. HW is the total number of pixels in the cuttings image. \hat{y} is the prediction probability of the (i, j) pixel sample that the model considers it belongs to the class 1, while y is the true label of the (i, j) pixel of the cuttings image. The model predicts the pixels of the cuttings with wet epidermis and those with dry epidermis simultaneously. The sum of the predicted probabilities of the two categories is 1.

To improve the performance of the classification model, the dice loss function is designed as Equation (12), combining the advantages of the dice loss with the cross-entropy loss function.

$$L_{ch} = \alpha L(p, t) + \beta L(y, \hat{y}) + \frac{\lambda}{2} \sum_i \sum_j \omega_{ij}^2 + \eta \quad (12)$$

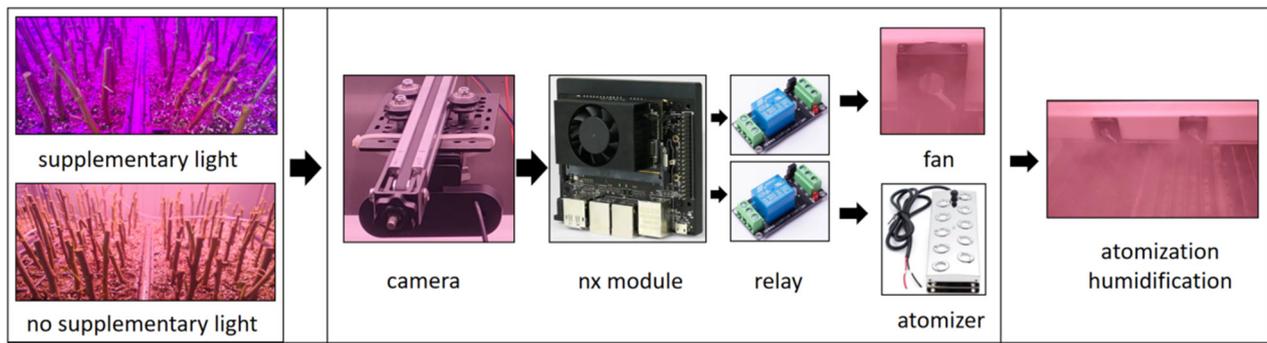
α, β represents the weight of dice loss (the segmentation loss of cuttings and background image) and cross-entropy loss function (cuttings represent wetness classification loss), as well as the segmentation of cuttings and background image, identifying the tonal channel information. The wetness classification of cuttings is based on the calculation of two channel data, saturation and value. Considering the reasonable distribution of computing power, the setting is the weight of loss function $\beta > \alpha$ and $\alpha + \beta = 1$. In order to prevent overfitting and improve the generalization ability of the model, the L^2 regularization term $\frac{\lambda}{2} \sum_i \sum_j \omega_{ij}^2$ should be introduced into the reconstruction loss function. ω_{ij} is the weight vector of the element of (i, j) , while λ is a hyperparameter that controls the regularization intensity.

In practical application, it should be ensured that the humidification system can humidify the cuttings' epidermis as much as possible, so that most of the cuttings' epidermis can remain wet. Therefore, when classifying the wetness of cuttings' epidermis, it is necessary to predict the wet cuttings image as dry, which is less harmful than the probability of predicting the dry cuttings image as wet. So the required model has a certain deviation between the predicted value and the actual value. It is necessary to increase the bias parameter η ($\eta > 0$) of the loss function L_{ch} to meet this demand, so as to increase the number of pixels predicted by the model with a pixel category of 1, improving the probability of the cuttings being predicted to be dry, as well as enhancing the practicability of the model.

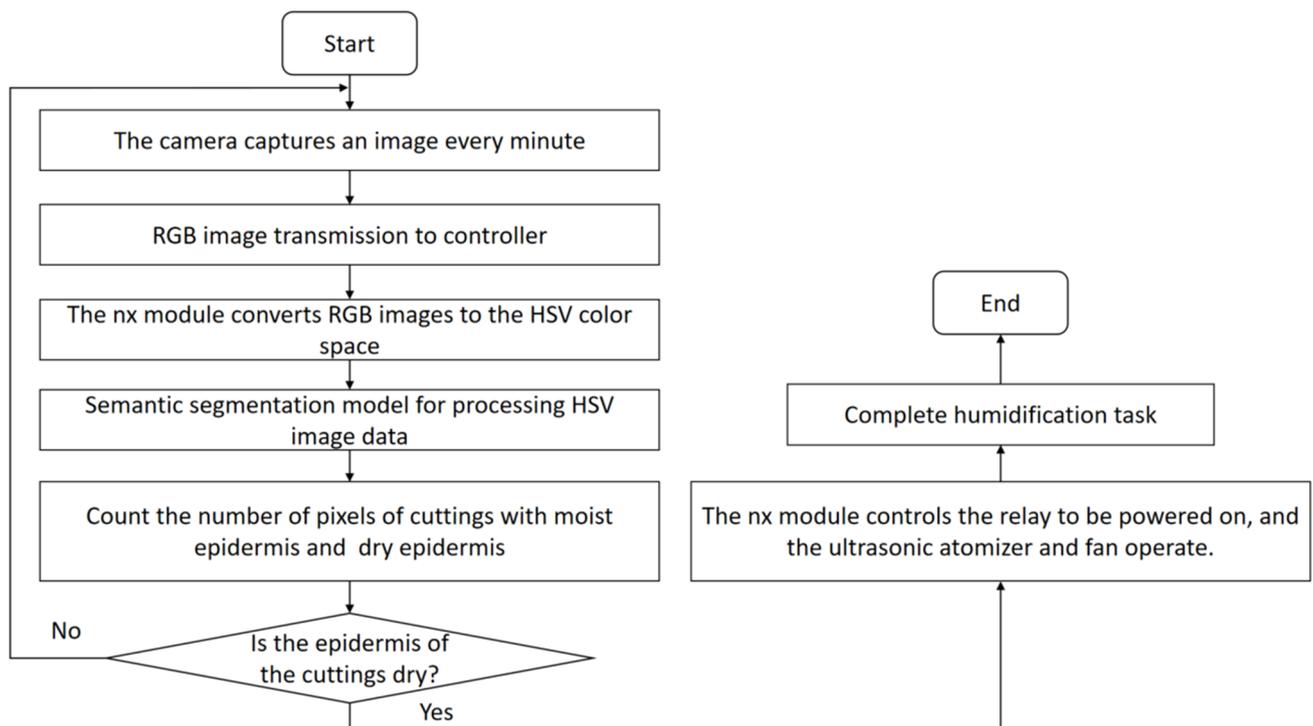
2.5. Surface Humidification System for Cuttings

2.5.1. Design for Humidification System

To achieve the goal of moisturizing the epidermis of the cuttings, a 0.01 mm thick water film is attached to the epidermis of the cuttings. As shown in Figure 13, this study designed an automatic humidification control system for the epidermis of the cuttings.



(a)



(b)

Figure 13. Design of humidification control system for cuttings. (a) Schematic diagram of humidification system operation; (b) Control flowchart of humidification system.

As shown in Figure 13a, the humidifying control system for inserting ears consists of a camera, module, relay, fan, and atomizer. Figure 13b shows the software design of the humidification control system for cuttings in this study. The camera collects images of the inserted cutting epidermis image every minute, and the image data are input into the neural network model of the module for prediction. The prediction results of the insertion epidermis image serve as the basis for the operation of the fan and atomizer. In the prediction results, the proportion of dry inserted epidermis image pixels to the total number of inserted cuttings pixels is greater than 50%. When the atomized and humidified insertion epidermis image pixels account for more than 90% of the total number of pixels, the fan and atomizer will stop running.

2.5.2. Test of Humidification System

This study aims to validate the performance of the neural network model by conducting experiments on the automatic humidification system of the cuttings' epidermis, as shown in Figure 14.

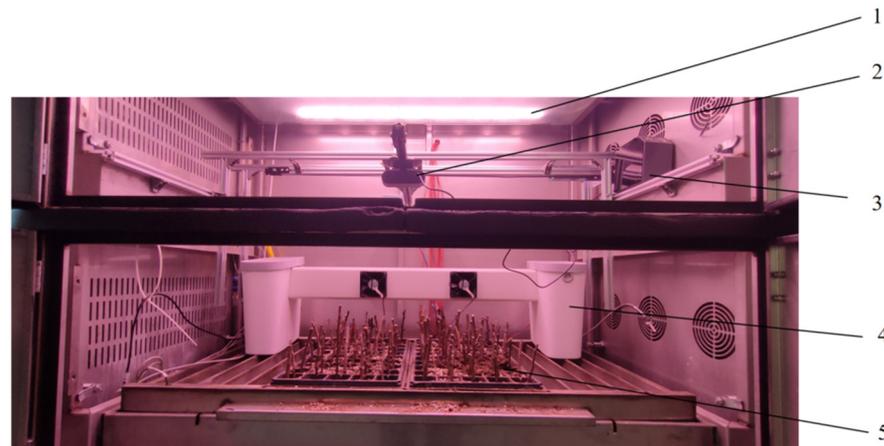


Figure 14. Insertion humidification system. 1. Fill light; 2. 1080P camera; 3. Controller; 4. Atomization humidification system; 5. Cuttings.

The rated power of the blue-purple light is 50 W, and the wavelength adjustment range of the light source spans from 400 nm to 475 nm. The power of the light source under the light environment of the non-blue-purple lamp is also set to 50 W. The relative humidity of the experimental environment is 95%. The temperature of the experimental environment is 22 °C. The observation angle set by the camera aligns with the image acquisition system depicted in Figure 13, ensuring that the number of cuttings observed within the camera's field of view meets the prediction requirements of the model. The controller consists of Nvidia Jetson Xavier nx modules, with 16 GB of RAM and the ability to run the neural network model. The fog droplet generation device is an ultrasonic atomizer that utilizes electronic high-frequency oscillation at a frequency of 1.7 MHz to generate water mist by breaking up the structure of liquid water molecules through high-frequency resonance of ceramic atomization plates. The negative pressure fan at the droplet outlet transports the water mist to the surface of the cuttings at a speed of 3.2 m/s, where the droplets deposit to form a water film for moisture retention. A total of 300 cuttings of apple rootstocks were subjected to hardwood cutting experiments to serve as test subjects for evaluating the efficiency of the humidification system.

3. Results

3.1. Configuration of Model Training Environment Parameters

Parameters for model U-DSE-AG-Net training workstation: Intel Xeon E5-2630×2 processor that was made by Intel which is located in Santa Clara, CA, USA, 64 GB memory, disk storage space 4 TB. The type of graphics card is RTX 3090 Ti, with 24 GB storage space. The workstation is made by Dell (Xiamen, China) company. The construction of the training environment is based on Pytorch, selecting Adam optimizer and setting the training cycle epoch 200 times. For each epoch, it will save the weight once, with a maximum learning rate of 0.1 as well as a minimum learning rate of 0.0001. $\alpha = 0.356$, $\beta = 0.644$, $\lambda = 0.5$, $\eta = 0.005$.

3.1.1. Evaluation Index and Result Based on Loss Value

In the process of model training, Loss and Val_Loss can be used as the basis for selecting the best model [39,40]. Loss and Val_Loss continue to decrease during the training process, while the performance of the model on the validation set is gradually improved. When Val_Loss shows an upward trend and Loss indicates a downward trend, the model begins to overfit. When both Loss and Val_Loss continue to decline, however, Val_Loss is always higher than Loss, and the model begins to underfit. In order to verify the performance of the U-DSE-AG-Net neural network model, it is necessary to compare it with the control group composed of the DeepLabV3+ neural network and the PSPNet neural network model. The model of the U-SE-Net neural network can be constructed by

embedding the SE module into the skip connection layer of the U-Net neural network, and then the ablation study can be carried out. All neural network models are trained on the same set of cuttings' image data. The two types of loss value change processes recording the training process of the neural network model are shown in Figure 15.

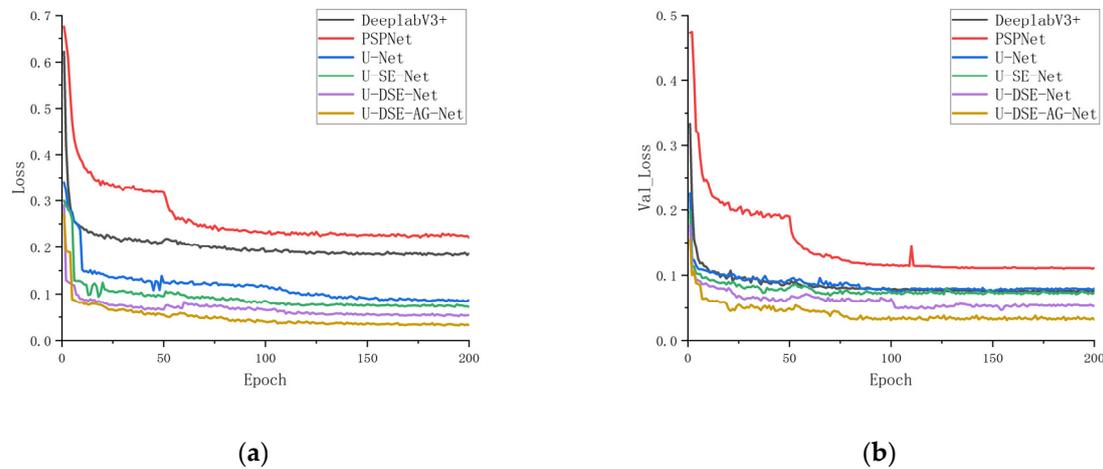


Figure 15. The neural networks are trained to recognize Loss and Val_Loss. (a): Loss. (b): Val_Loss.

In the comparison experiment, the decrease rate of loss value as well as the final value of the DeepLabV3+ and PSPNet neural networks in the control group is lower than that of U-Net, which proves that the skip connection layer of U-Net fuses deep-level features with shallow features. It improves the efficiency and accuracy of image segmentation. Both Loss and Val_loss of the four neural network models (U-Net, U-SE-Net, U-DSE-Net, and U-DSE-AG-Net) in the ablation group shows a downward trend while the performance of the models gradually improves. There is no significant increase in outliers during the decline in loss value. The regularization term of the loss function L_{ch} improves the generalization ability of the model, which avoids overfitting the model. When Epoch reaches 200, it is shown in Figure 15a. The Loss of the six neural network models is 0.221, 0.183, 0.088, 0.074, 0.053, and 0.033. As is shown in Figure 15b, the Val_Loss of the six neural network models is 0.076, 0.111, 0.090, 0.078, 0.058, and 0.037.

As indicated in Figure 15, in the ablation test, the decrease rate of the loss value of U-SE-Net is faster than that of U-Net, which proves that the SE module makes the neural network pay more attention to the information of Hue channel and Saturation channel of the cuttings' characteristic images, reducing the capture of spatial information and accelerating the gradient decline. U-DSE-Net's decrease rate in Loss value is slower than that of U-Net and U-SE-Net, but the final value of the two types of loss values is smaller than U-Net as well as U-SE-Net. As is indicated, under the effect of the new module DSE, the extended Squeeze phase increases the number of parameters of the neural network operation, resulting in an increase in the time of a single iteration of the model weight. Therefore, the decrease rate in Loss value is relatively low, while the model accuracy improves with the increase in channel information. The decrease rate in loss value of the neural network U-DSE-AG-Net is only faster than that of U-DSE-Net. Compared with all other neural networks, the final value of the two types of loss values is the smallest and the decline curve is stable, which proves that the generalization of the model is the best while the prediction result is more accurate. This shows that module DSE-AG combines the characteristic information of the deep layer of the network, allocating the weight of the characteristics transmitted by the skip connection layer according to the significance, as well as reducing the interference of background light noise.

3.1.2. Evaluation Indices and Results Based on Confusion Matrix

The confusion matrix represents the relationship between the predicted results of the model in the test dataset as well as the true label, which includes TP, TN, FP, and FN. Taking the wetness of cuttings in the study as an example, TP indicates that the actual category of cuttings is epidermal dryness, while the predicted category is also epidermal dryness. TN indicates that the actual category of cuttings is epidermal wet, and the predicted category is also epidermal wet. FP indicates that the actual category of cuttings is epidermal wet, while the predicted category is epidermal dryness. FN indicates that the actual category of cuttings is epidermal dryness, while the predicted category is epidermal wet. In binary classification, the confusion matrix is used to obtain the above four results (Figure 16).

| | | True Class | |
|-----------------|----------|------------|----------|
| | | Positive | Negative |
| Predicted Class | Positive | TP | FP |
| | Negative | FN | TN |

Figure 16. Confusion matrix diagram for the degree of epidermal wetness classification of cuttings.

The evaluation indices based on the confusion matrix include Precision, Recall, *F1*-Score, and Accuracy. As shown in Equation (13), precision represents the proportion of the TP in the sample for which the model predicts a positive example.

$$P = \frac{TP}{TP + FP} \times 100\% \quad (13)$$

As shown in Equation (14), recall represents the proportion of *TP* predicted by the model to all positive examples.

$$R = \frac{TP}{TP + FN} \times 100\% \quad (14)$$

F1-Score is the harmonic mean of the accuracy rate of model prediction and recall rate. The computation formula is Equation (15).

$$F1 = \frac{2PR}{P + R} \times 100\% \quad (15)$$

As is shown in Figure 17, in order to save the computing resource, each of the five Epochs is one group, which assesses the *F1*-Score of the neural network model once. The *F1*-Score of the six neural networks is on the rise. The *F1*-Score of the U-DSE-AG-Net neural network model is greater than that of other neural network models with the final value of 0.948. This proves that the accuracy rate *P* and recall rate *R* of the U-DSE-AG-Net neural network model are high. It will be more precise than other neural network models when identifying the significant sample.

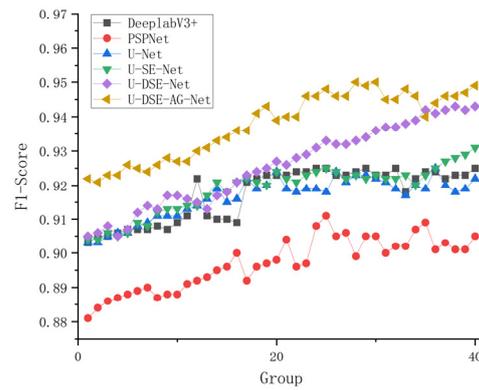


Figure 17. F1-Score of neural network models.

Accuracy represents the proportion of samples in which the model predicts the correct degree of epidermal wetness from the total sample. It can be seen in Equation (16).

$$A = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{16}$$

In order to verify the accuracy rate of each neural network model in classifying the degree of wetness of cuttings, a segmentation test should be conducted on the four types of images, as seen in Figure 3. In order to ensure the scientific prediction of the model, the cuttings' genotypes used in the experiment are different from those in Figure 3. In this study, the cutting images of three genotypes were selected for the experiment. Each genotype has 100 rootstock pictures. A total of 300 images were randomly selected from each type of images. Each input image is 512 × 512 pixels. Figure 18 shows the comparison between the segmentation effect of each neural network and manual annotation. The red split information represents dry epidermis pixels of the cuttings. The green split information represents wet epidermis pixels of the cuttings.

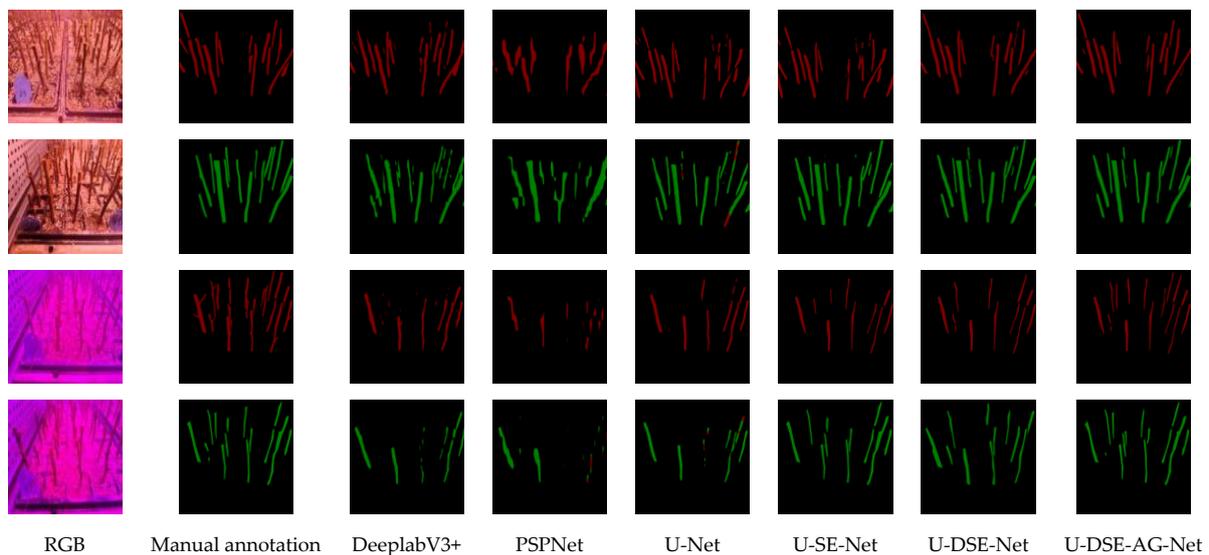


Figure 18. Segmentation results of different neural networks.

The accuracy rate of each neural network model for image segmentation results is shown in Table 2.

Table 2. Segmentation results of four kinds of cuttings by neural network models.

| Neural Network | Accuracy Rating (A)/% | | | | | | | | | | | |
|----------------|---------------------------|-------|-------|---------------|-------|-------|------------------------|-------|-------|---------------|-------|-------|
| | No Supplementary Lighting | | | | | | Supplementary Lighting | | | | | |
| | Wet Epidermis | | | Dry Epidermis | | | Wet Epidermis | | | Dry Epidermis | | |
| | N40 | N59 | G935 | N40 | N59 | G935 | N40 | N59 | G935 | N40 | N59 | G935 |
| DeepLabV3+ | 79.03 | 78.55 | 89.63 | 69.02 | 67.77 | 89.54 | 23.99 | 34.15 | 37.84 | 32.57 | 35.09 | 35.12 |
| PSPNet | 44.23 | 55.21 | 65.03 | 33.38 | 37.46 | 63.21 | 20.11 | 10.56 | 27.49 | 24.16 | 23.33 | 27.47 |
| U-Net | 87.22 | 85.69 | 90.04 | 85.23 | 86.52 | 90.42 | 31.28 | 29.55 | 42.75 | 35.26 | 45.07 | 49.06 |
| U-SE-Net | 90.89 | 90.11 | 92.15 | 89.33 | 88.29 | 92.64 | 69.24 | 79.02 | 78.58 | 71.59 | 65.22 | 78.84 |
| U-DSE-Net | 92.11 | 90.01 | 93.14 | 90.31 | 91.28 | 93.47 | 79.62 | 81.18 | 83.35 | 82.32 | 80.58 | 84.93 |
| U-DSE-AG-Net | 95.01 | 94.99 | 95.25 | 94.73 | 95.05 | 95.44 | 85.62 | 88.44 | 88.16 | 89.09 | 88.97 | 89.68 |

N40, N59, and G935 in Table 1 represent the varieties of cuttings. N40 refers to Xinjiang wild apple 40. N59 refers to Xinjiang wild apple 59. G935 is a rootstock variety with good stress resistance. The accuracy rate of rootstock cutting recognition of these three genotypes by the statistical model reflects the generalization of the model.

As shown in Figure 19, the average accuracy of the model segmentation in different scenes was compared in this study. Figure 19a shows the average segmentation accuracy of the models for the cutting images of three genotypes without a blue-purple supplementary light environment. Figure 19b shows the average segmentation accuracy of the models for the cutting images of three genotypes in the blue-purple supplemental light environment. The average recognition accuracy of the U-DSE-AG-Net model in identifying cuttings with wet epidermis and dry epidermis without blue-purple supplemental light was 95.07% and 95.07%, respectively. The average recognition accuracy of the U-DSE-AG-Net model in identifying cuttings in blue-purple supplemental light was 87.41% and 89.24%, respectively.

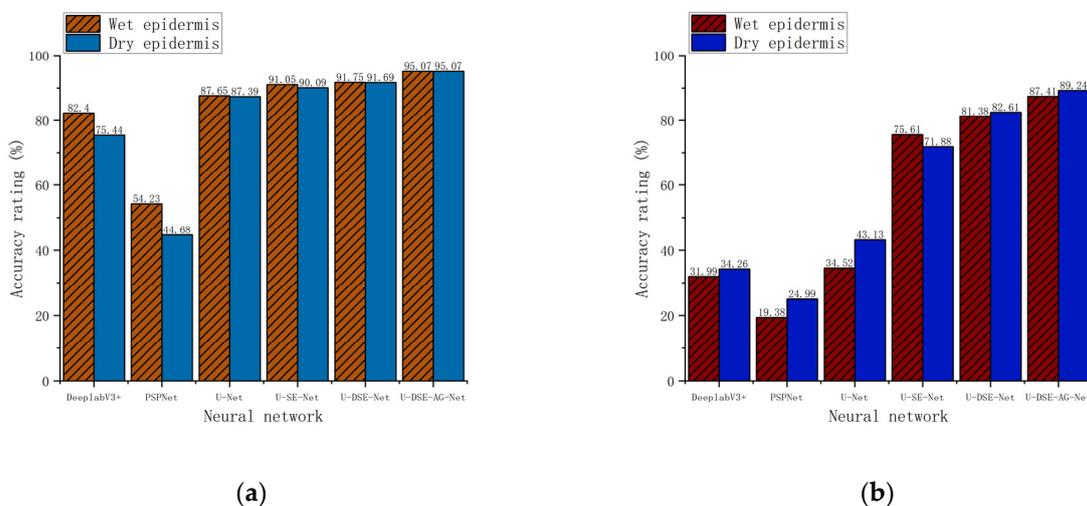


Figure 19. Average accuracy of model recognition in different scenes. (a) The average segmentation accuracy of the models for the cutting images of three genotypes without a blue-purple supplementary light environment; (b) The average segmentation accuracy of the models for the cutting images of three genotypes in the blue-purple supplemental light environment.

From Figure 18 and Table 2, the accuracy of all neural network models without supplementary lighting in classifying the degree of wetness of cuttings is higher than that in the case of supplementary lighting, which proves that fill light has exerted a bigger influence on the classification of model. When the U-DSE-AG-Net model has no supplementary lighting, the accuracy of classification of wetness of cuttings with wet epidermis and dry epidermis is 95.25% and 95.44%. When it has supplementary lighting, the accuracy of classification of wetness of cuttings with wet epidermis and dry epidermis is 88.16% and 89.68%. The U-DSE-AG-Net model, for the classification of four types

of cuttings images, is more accurate than other neural networks for the classification of cuttings' epidermis wetness, and the prediction effect is the best. When comparing the U-DSE-AG-Net model with the U-Net model, the classification accuracy of the cuttings with wet epidermis and dry epidermis increases by 5.21% and 5.02%, respectively, under the condition of no supplementary lighting, as well as increases by 45.41% and 40.62% with blue-purple supplementary lighting. The U-DSE-AG-Net model has a strong ability to resist light noise interference.

As seen in Tables 2 and 3, and Figure 19, the Loss function of the ablation experimental groups U-Net, U-SE-Net, U-DSE-Net, and U-DSE-AG-Net neural network is the function L_{ch} . Compared with the control group DeepLabV3+ and PSPNet neural network, on the condition of no supplementary lighting, the accuracy of classification of cuttings' epidermis wetness in the ablation experimental group is higher than that in the control group, which proves that the loss function L_{ch} combines the advantages of cross-entropy loss function and dice loss. It can also accurately segment the cuttings' image pixels from the background image pixels, which realizes the accurate classification of the wetness of cuttings' pixels. The accuracy of the model in predicting the dry epidermis of cuttings in the ablation test group is higher than that of the wet epidermal image, which proves that the bias parameter η in the loss function L_{ch} biases the prediction results more in favor of the dry epidermis. The predicted results are favorable for the moisturizing of cuttings.

Table 3. Performance parameter statistics of humidification system.

| Neural Network | Number of Images | Average Detection Time (ms) | Number of Converted Frames (Fps) | Moisture Retention Rate of Cuttings (%) | |
|----------------|------------------|-----------------------------|----------------------------------|---|------------------------|
| | | | | No Supplementary Lighting | Supplementary Lighting |
| DeeplabV3+ | 300 | 85.86 | 34.93 | 85.43 | 32.04 |
| PSPNet | 300 | 164.07 | 18.28 | 61.31 | 23.46 |
| U-Net | 300 | 78.99 | 37.98 | 90.22 | 41.04 |
| U-SE-Net | 300 | 75.53 | 40.80 | 90.62 | 73.53 |
| U-DSE-Net | 300 | 77.16 | 38.88 | 92.37 | 83.66 |
| U-DSE-AG-Net | 300 | 78.03 | 38.45 | 95.14 | 89.87 |

4. Discussion

1. In this study, the classification of epidermis wetness of the cuttings included wet and dry categories. The classification of the epidermis wetness of the cuttings needs to be further divided into more detailed grading according to the actual requirements of production. The factors affecting the epidermis wetness of the cuttings include not only the interaction with the external environment, but also the consumption and generation of water by the physiological activities of the cuttings themselves (transpiration consumes water, while photosynthesis produces water). Therefore, the epidermis wetness of the cuttings is a dynamic process, which requires not only the observation of the changes in the external environment, but also the physiological activities of the cuttings themselves. Spectroscopy technology can accurately reflect the content of chlorophyll as well as carotene in vegetation cells. It can also indirectly reflect the intensity of photosynthesis in epidermal cells of cuttings. Microscopy imaging technology is used to observe the degree of stomatal opening and closing of cuttings' epidermal cells, which could also be used to monitor the intensity of transpiration of cuttings' epidermal cells. Therefore, in order to establish a model of the change in the epidermis wetness of the cuttings, the image segmentation technology is combined with spectral technology as well as microscope imaging technology to fully explore the dynamic communication relationship between the cuttings and the external environment, constructing a dynamic classification model of the change in the wetness of the epidermis of the cuttings.
2. In the process of applying machine vision technology to the monitoring of the moisture of the epidermis of cuttings, the intensity and color of the fill light have a direct

impact on the imaging results. Therefore, when constructing the cuttings model of the epidermal wetness of cuttings, it is necessary to further explore the relationship between the spectral information, transmission mode, transmission path of external light, and the imaging of epidermal wetness of cuttings. The physical characteristics of the epidermis of cuttings of different genotypes are diverse, such as roughness, fluff, or not. It not only leads to the difference in the adhesion state of the water film, but also changes the transmission of the fill light line as well as influences the effect on the expression of the epidermis of the cuttings in the color space, so it is necessary to explore the influence mechanism of the physical properties of the epidermis of the cuttings and the influence of light source on the wetness of the epidermis of the cuttings. In order to further improve the understanding of the wetness information of cuttings by the neural network model, the advantages of multiple color spaces, such as RGB, HSV, LAB, HIS, and so on, can be combined in the future to obtain the features of the epidermal wetness of cutting images more efficiently.

3. With regard to classifying the epidermal wetness of cuttings, aiming at further improving the model transferability of the neural network model, it is necessary to change the supervised training method in this study to the unsupervised training method in the future, so that the neural network can independently divide the level of the epidermis wetness of the cuttings, improving the classification effect of the neural network on the epidermis wetness of the cuttings. In order to better embed the mobile terminal to provide data for the humidification system, the U-DSE-AG-Net model is needed for lightweight design in the future.

5. Conclusions

The neural network of image segmentation is used to design a classification model for the epidermal wetness of cuttings. This model is designed to detect accuracy and efficiency of the epidermal wettability during the growth process for apple rootstock cuttings, and to realize the moisture retention of cuttings.

1. This study converted RGB images of the cuttings into the HSV color space, achieving the effective expression of the wetness information of the cutting epidermis. The Hue channel and the Saturation channel information can be used as the basis for classifying the wetness of the epidermis of cuttings in the environment of supplemental lighting, improving the classification ability of the model in complex light environments.
2. The module DSE strengthens the model's ability to capture Hue channel and Saturation channel information based on the module SE. The module DSE is integrated with the improved module AG, assigning non-negative weights to important features, which reduces the error of prediction. The skip connection layers of the U-Net embedded module DSE and module AG result in U-DSE-AG-Net. This network structure can weaken the lighting noise interference in the skipping connection layer. The comparative test and ablation test show that the U-DSE-AG-Net neural network model has the best performance. Loss and Val_Loss are the smallest; that is, 0.033 and 0.037, respectively. The *F1*-Score is improved by 3.2% compared to U-Net. The accuracy of the model in predicting the wetness and dryness of the cuttings epidermis is increased by 45.41% and 40.62% in the supplementary blue-purple light environment. The model has a solid ability to resist light noise interference.
3. The experiment of identifying the moisture content of the epidermis of three kinds of cuttings with genotypes N40, N59, and G935 was carried out. The average accuracy of the model was 91.69%. The detection speed rate of the model was 38.45 fps. The average moisture retention rate of the humidification system for cuttings was 92.51%. The system can realize the real-time monitoring of the moisture content of the cuttings' epidermis and ensure consistent moisturizing. The model has good generalization and practicability. It qualifies as an economical, non-contact, as well as non-destructive monitoring method.

Author Contributions: Conceptualization, X.Y.; methodology, X.Y.; software, X.W. and L.L.; formal analysis, X.W. and L.L.; investigation, X.W. and L.L.; resources, X.Y.; writing—original draft preparation, X.W., L.L. and X.Y.; writing—review and editing, X.W., L.L., J.Z., H.L. and X.Y.; visualization, X.W., L.L., P.W. and J.L.; supervision, P.W. and H.L.; project administration X.Y. and J.L.; funding acquisition, X.Y., J.L. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the earmarked fund for CARS (CARS-27), and supported by the Earmarked Fund for Hebei Apple Innovation Team of Modern Agro-industry Technology Research System (HBCT2024150202), and supported by the Hebei Province Graduate Innovation Funding Project (CXZZBS2024084).

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Xu, L.; Huang, H. GeNetic and epigeNetic controls of plant regeneration. *Curr. Top. Dev. Biol.* **2014**, *108*, 1–33. [[CrossRef](#)] [[PubMed](#)]
- Liu, K.; Yang, A.; Yan, J.; Liang, Z.; Yuan, G.; Cong, P.; Zhang, L.; Han, X.; Zhang, C. MdAIL5 overexpression promotes apple adventitious shoot regeneration by regulating hormone signaling and activating the expression of shoot development-related genes. *Hortic. Res.* **2023**, *10*, uhad198. [[CrossRef](#)] [[PubMed](#)]
- Wang, X.; Liu, L.; Xie, J.; Wang, X.; Gu, H.; Li, J.; Liu, H.; Wang, P.; Yang, X. Research Status and Prospects on the Construction Methods of Temperature and Humidity Environmental Models in Arbor Tree Cuttage. *Agronomy* **2023**, *14*, 58. [[CrossRef](#)]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
- Camargo, A.; Smith, J.S. Image pattern classification for the identification of disease causing agents in plants. *Comput. Electron. Agric.* **2009**, *66*, 121–125. [[CrossRef](#)]
- Yuan, J.; Zhou, W.; Luo, T. DMFNet: Deep multi-modal fusion Network for RGB-D indoor scene segmentation. *IEEE Access* **2019**, *7*, 169350–169358. [[CrossRef](#)]
- Zhang, Y.; Yang, J.; Deng, H.; Zhou, Y.; Miao, Y. Semantic segmentation model for greenhouse tomato images using RGB. *Trans. Chin. Soc. Agric. Eng.* **2024**, *40*, 295–306. [[CrossRef](#)]
- Yao, Y.; Peng, Y.; Chen, Z.; He, W.; Wu, Q.; Huang, W.; Chen, W. An Improved YOLO Algorithm Supporting Anti-illumination Target Detection. *Automot. Eng.* **2023**, *45*, 777–785. [[CrossRef](#)]
- Feng, S.; Yang, X.; Li, G.; Zhao, D.; Yu, F.; Xu, T. Unsupervised extraction of rice coverage with incorporating CLAHE-SV enhanced Lab color features. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 195–206. [[CrossRef](#)]
- Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
- Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. BiseNet: Bilateral segmentation Network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 325–341.
- Li, L.; Hu, W.; Lu, J.; Zhang, C. Leaf vein segmentation with self-supervision. *Comput. Electron. Agric.* **2022**, *203*, 107352. [[CrossRef](#)]
- Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890. [[CrossRef](#)]
- Li, Z.; Yu, J.; Pan, S.; Jia, Z.; Niu, Z. Individual Tree Skeleton Extraction and Crown Prediction Method of Winter Kiwifruit Trees. *Smart Agric.* **2023**, *5*, 92–104. [[CrossRef](#)]
- Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
- Cao, Y.; Zhao, Y.; Yang, L.; Li, J.; Qin, L. Weed Identification Method in Rice Field Based on Improved DeepLabv3+. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 242–252.
- Li, C.; Tan, Y.; Chen, W.; Luo, X.; Gao, Y.; Jia, X.; Wang, Z. Attention unet++: A nested attention-aware u-net for liver ct image segmentation. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 345–349. [[CrossRef](#)]
- Liu, Y.; Zhou, X.; Wang, Y.; Yu, H.; Geng, C.; He, M. Straw coverage detection of conservation tillage farmland based on improved U-Net model. *Opt. Precis. Eng.* **2022**, *30*, 1101–1112. [[CrossRef](#)]
- Cai, W.; Wang, B.; Zeng, F. CUDU-Net: Collaborative up-sampling decoder U-Net for leaf vein segmentation. *Digit. Signal Process.* **2024**, *144*, 104287. [[CrossRef](#)]
- Chaudhari, S.; Mithal, V.; Polatkan, G.; Ramanath, R. An attentive survey of attention models. *ACM Trans. Intell. Syst. Technol. (TIST)* **2021**, *12*, 1–32. [[CrossRef](#)]

21. Zhu, D.; Wen, R.; Xiong, J. Lightweight corn silk detection network incorporating with coordinate attention mechanism. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 145–153. [[CrossRef](#)]
22. Han, X.; Zhao, C.; Wu, H.; Zhu, H.; Zhang, Y. Image classification method for tomato leaf deficient nutrient elements based on attention mechanism and multi-scale feature fusion. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 177–188. [[CrossRef](#)]
23. Wang, Z.; Ma, F.; Zhang, Y.; Zhang, F.; Ji, P.; Cao, M. Crop disease recognition using attention mechanism and multi-scale lightweight network. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 176–183. [[CrossRef](#)]
24. Su, B.; Shen, L.; Chen, S.; Mi, Z.; Song, Y.; Lu, N. Multi-features Identification of Grape Cultivars Based on Attention Mechanism. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 226–233+252. [[CrossRef](#)]
25. Zhang, Q.; Hu, S.; Shu, W.; Cheng, H. Wheat Spikes Detection Method Based on Pyramidal Network of Attention Mechanism. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 253–262. [[CrossRef](#)]
26. Huo, P.; Ma, S.; Su, C.; Ding, Z. Emergency obstacle avoidance system of sugarcane basecutter based on improved YOLOv5s. *Comput. Electron. Agric.* **2024**, *216*, 108468. [[CrossRef](#)]
27. Dosovitskiy, A.; Fischer, P.; Springenberg, J.T.; Riedmiller, M.; Brox, T. Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1734–1747. [[CrossRef](#)] [[PubMed](#)]
28. 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. *Comput. Electron. Agric.* **2021**, *189*, 106370. [[CrossRef](#)]
29. Ma, Y.; Bian, M.; Fan, Y.; Chen, Z.; Yang, G.; Feng, H. Estimation of Potassium Content of Potato Plants Based on UAV RGB Images. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 196–203+233. [[CrossRef](#)]
30. Zhang, P.; Chen, Z.; Ma, S.; Yin, D.; Jiang, H. Prediction of soybean yield by using RGB model with skew distribution pattern of canopy leaf color. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 120–126. [[CrossRef](#)]
31. Song, C.; Qu, X.; Hu, G.; Su, T. Crop Identification in Mature Stage with Remote Sensing Based on NDVI-NSSI Space and HSV Transformation. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 193–200. [[CrossRef](#)]
32. Xiong, X.; Yu, L.; Yang, W.; Liu, M.; Jiang, N.; Wu, D.; Chen, G.; Xiong, L.; Liu, Q. A high-throughput stereo-imaging system for quantifying rape leaf traits during the seedling stage. *Plant Methods* **2017**, *13*, 1–17. [[CrossRef](#)]
33. Li, T.; Sun, M.; Ding, X.; Li, Y.; Zhang, G.; Shi, G.; Li, W. Tomato recognition method at the ripening stage based on YOLO v4 and HSV. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 183–190. [[CrossRef](#)]
34. Liu, P.; Liu, L.; Wang, C.; Zhu, Y.; Wang, H.; Li, X. Determination Method of Field Wheat Flowering Period Based on Machine Vision. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 251–258. [[CrossRef](#)]
35. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015, Proceedings, Part III 18*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. [[CrossRef](#)]
36. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*; pp. 7132–7141. [[CrossRef](#)]
37. Bahdanau, D. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473. [[CrossRef](#)]
38. Zhang, Y.; Yi, P.; Zhou, D.; Yang, X.; Yang, D.; Zhang, Q.; Wei, X. CSANet: Channel and spatial mixed attention CNN for pedestrian detection. *IEEE Access* **2020**, *8*, 76243–76252. [[CrossRef](#)]
39. Jin, C.; Ben, X.; Chao, J. Apple inflorescence recognition of phenology stage in complex background based on improved YOLOv7. *Comput. Electron. Agric.* **2023**, *51*, 211–219. [[CrossRef](#)]
40. Li, S.; Zhang, S.; Xue, J.; Sun, H. Lightweight target detection for the field flat jujube based on improved YOLOv5. *Comput. Electron. Agric.* **2022**, *202*, 107391. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.