agronomy

Article

U-Net with Coordinate Attention and VGGNet: A Grape Image Segmentation Algorithm Based on Fusion Pyramid Pooling and the Dual-Attention Mechanism

Xiaomei Yi †, Yue Zhou †, Peng Wu *, Guoying Wang, Lufeng Mo, Musenge Chola, Xinyun Fu and Pengxiang Qian

College of Mathematics and Computer Science, Zhejiang A and F University, Hangzhou 311300, China; yxm@zafu.edu.cn (X.Y.); 2021611012016@stu.zafu.edu.cn (Y.Z.); molufeng@gmail.com (L.M.); musengechola1@gmail.com (M.C.); fxy111@stu.zafu.edu.cn (X.F.); 202224070120@stu.zafu.edu.cn (P.Q.)

* Correspondence: wp@zafu.edu.cn

† These authors contributed equally to this work.

Abstract: Currently, the classification of grapevine black rot disease relies on assessing the percentage of affected spots in the total area, with a primary focus on accurately segmenting these spots in images. Particularly challenging are cases in which lesion areas are small and boundaries are ill-defined, hampering precise segmentation. In our study, we introduce an enhanced U-Net network tailored for segmenting black rot spots on grape leaves. Leveraging VGG as the U-Net's backbone, we strategically position the atrous spatial pyramid pooling (ASPP) module at the base of the U-Net to serve as a link between the encoder and decoder. Additionally, channel and spatial dual-attention modules are integrated into the decoder, alongside a feature pyramid network aimed at fusing diverse levels of feature maps to enhance the segmentation of diseased regions. Our model outperforms traditional plant disease semantic segmentation approaches like DeeplabV3+, U-Net, and PSPNet, achieving impressive pixel accuracy (PA) and mean intersection over union (MIoU) scores of 94.33% and 91.09%, respectively. Demonstrating strong performance across various levels of spot segmentation, our method showcases its efficacy in enhancing the segmentation accuracy of black rot spots on grapevines.

Keywords: grape black rot; U-Net; ASPP; dual attention module; feature pyramid

Citation: Yi, X.; Zhou, Y.; Wu, P.; Wang, G.; Mo, L.; Chola, M.; Fu, X.; Qian, P. U-Net with Coordinate Attention and VGGNet: A Grape Image Segmentation Algorithm Based on Fusion Pyramid Pooling and the Dual-Attention Mechanism. *Agronomy* **2024**, *14*, 925. https://doi.org/10.3390/agronomy14050925

Academic Editor: Yanbo Huang

Received: 24 March 2024
Revised: 24 April 2024
Accepted: 25 April 2024
Published: 28 April 2024

1. Introduction

Grapes are one of the most important fruits in the world, and the healthy and stable development of their industry is of great significance to the national economic development and farmers' income increase. In the cultivation of grapes, the larger the planting area, the larger the scale of damage when disease occurs and the greater the economic loss caused. Among grape leaf diseases, black rot, brown spot, and verticillium are the most common, of which black rot is one of the most important grape diseases worldwide. Black rot is a fungal disease that causes yield loss in grapes, showing black spots on leaves and fruit, and it is prevalent in the wetter spring and early summer seasons and affects a wide range of areas. Therefore, the rapid and accurate identification of grape leaf diseases and implementation of preventive measures can greatly reduce the degree of its harm in favor of increasing grape production and income [1]. At present, grape diseases mainly rely on agricultural experts for on-site identification, and manual identification is subjective, time-consuming, and labor-intensive, so it is important to develop a fast, accurate, and intelligent grape disease identification system [2].

Computer vision is widely used in the field of agriculture, and with the development of image processing and computer technology, image segmentation methods have experienced three basic stages, as follows: classical segmentation methods, machine learning methods, and deep learning methods. These methods have been applied in agricultural disease detection.

Traditional image segmentation techniques, such as threshold segmentation, can distinguish lesions from the background by using color and texture properties. Then, each pixel in the image is compared; if its gray value is greater than the threshold, the pixel is classified into one category, and if its gray value is less than the threshold, it is classified into another category. Dutta et al. [3] proposed a method for the efficient real-time segmentation of diseased leaves on kohlrabi plots by adjusting VI and Otsu thresholds. ZixiLiu et al. [4] used the Otsu method, OpenCV morphological operation, and morphological transformation method to outline the outline of the object for corn gray spots, corn rust, large corn spots, and healthy corn leaves and used the outline to obtain the difference set between the corn leaves and the background to obtain a complete corn leaf image. Classical image segmentation methods require high image quality, and if the environmental conditions change during the image quality, the recognition results will be poor or even invalid. Therefore, the versatility and robustness of these methods cannot be satisfying, and the accuracy in practical applications cannot be guaranteed.

With the development of machine learning, many researchers have started to try to apply it to disease speckle segmentation to improve the accuracy and robustness of segmentation. Attiquekhan et al. [5] used a genetic algorithm (GA) to add a feature selection step that further speeds up the process of obtaining improved classification results using support vector machines. Ambarwari et al. [6] used Support Vector Machine (SVM) with RBF kernel for plant species recognition with 82.67% accuracy. S. Appeltans et al. [7] removed soil pixels from hyperspectral images through LDA classification and a custom noise filtering algorithm. Machine learning methods can yield satisfactory segmentation results using small sample sizes, but these methods require multiple steps of image preprocessing and are relatively complex to execute. In addition, machine learning-based segmentation methods are relatively weak in unstructured environments and require researchers to manually design feature extraction and classifiers, which makes the work more difficult.

With the improvement of computer hardware performance, deep learning has been rapidly developed. Currently, common deep learning algorithms include the full convolutional neural network algorithm FCN proposed by Long et al. [8] for the problem of extremely high memory cost and low computational efficiency of CNN. Zhang et al. [9] established a full convolutional network (FCN)-based segmentation model for wheat spikelets, which effectively achieves the segmentation of wheat spikelets in the field environment. Badrinarayanan et al. [10] proposed SegNet, which uses an inverse convolutional filter to replace the traditional up-sampling operation, eliminating the need to learn to increase the sampling rate. Zhao et al. [11] proposed PSPNet, global prior information that is effective in obtaining high-quality results in scene semantic analyses. DeepakKumar et al. [12] used pyramid scene parsing network (PSPNet) and fuzzy rule model to develop an innovative multilevel model (PSGIC) for estimating wheat leaf rust and its infection level. Chen et al. successively proposed Deeplab [13], DeeplabV2 [14], DeeplabV3 [15], and DeeplabV3+ [16], which can efficiently extract multi-scale image semantic information. Cai et al. [17,18] used a modified DeeplabV3+ to segment maple leaves and spots, and then assessed the extent of disease damage. Ronneberger et al. [19] proposed a new model called U-Net. The U-Net network improves the FCN network by combining encoding paths that capture contextual information and decoding paths used for precise positioning, which splices high-resolution features with decoder up-sampled output features by jumping structures. Yi, Liu, et al. [20,21] performed algorithm improvement based on U-Net for light bark birch and rice segmentation. Chen et al. [22], based on the U-Net network, proposed BLSNet to improve the accuracy of rice lesion segmentation through attention mechanism and multi-scale extraction. Aiming at the problems of low crop classification accuracy, insufficient plant disease feature extraction, and inaccurate disease edge segmentation in the traditional plant classification model, this paper proposes an improved U-Net-based plant disease segmentation method, i.e., CVU-Net. The experimental results show that CVU-Net can take into account various requirements, such as accuracy and average intersection-to-union

ratio, can segment small lesions well, and has good segmentation effects on the edges of lesions.

The contribution of this paper mainly includes the following three parts:

1. A grape black rot spot segmentation model CVU-Net is proposed to achieve the accurate segmentation of grape black rot spots.
2. A dual-attention mechanism is incorporated into the U-Net encoding network, enabling the model to better capture the edge, texture, and semantic information of the target, thereby producing more accurate segmentation results.
3. The use of multiple atrous convolutions in multi-scale ASPP for parallel sampling of the input image for feature extraction, enriching the semantic information by expanding the sensory field, and encoding the global context using image-level features can avoid the problem of segmentation error due to falling into the local features, thus improving the segmentation performance of the network.

Because some of the diseased spots on grape leaves are small, and the edges of the lesions are blurry, there is no way for traditional deep learning methods to identify them accurately, so to improve the accuracy of disease semantic segmentation, this paper proposes an improved U-Net network, called CVU-Net network. In this network, we use the VGG network as the backbone feature extraction network [23], add the attention mechanism module to the feature extraction network part, and introduce the ASPP module to increase the field of view of the filter. We compared the segmentation performance of traditional U-Net, DeeplabV3+, PSP-NET, and the CVU-Net network proposed in this paper on the grape disease dataset. The experimental results show that CVU-Net outperforms the other compared networks in terms of segmentation performance. The improved method in this paper significantly improves the segmentation capability of the network, which effectively improves the segmentation accuracy of disease images.

The remainder of this article is structured as follows. We start with a description of the materials and methods of the experiments in Section 2. In Section 3, experiment results with a detailed discussion about the experiment are given. In Section 4, we discuss the experiments and suggest directions for future work. Finally, the conclusions of the experiment are presented in Section 5.

2. Materials and Methods

2.1. Data Acquisition and Preprocessing

For this study, we utilized the publicly available Plant Village dataset, which comprises 54,309 RGB images showcasing symptoms of 26 common diseases found on the leaves of 14 different plant species. Among these, this paper specifically selected 262 images of grape leaves afflicted with black rot as our test subjects. All of these images are verified by researchers with expertise in grape diseases.

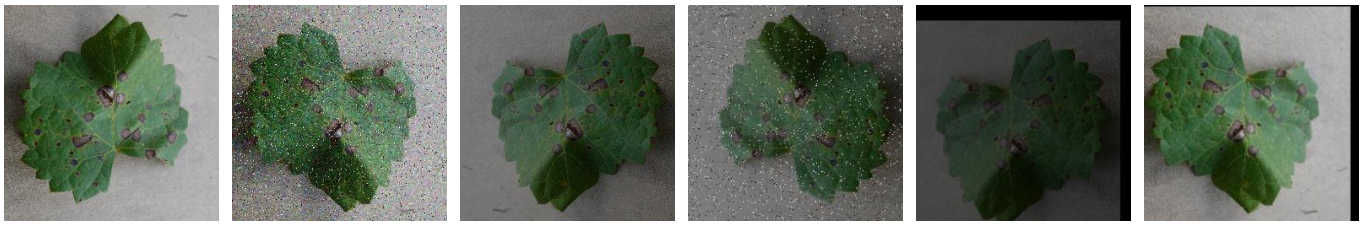
To label and segment the dataset, this paper employed the LabelMe 4.5.13. LabelMe is a visual annotation tool developed in Python 3.7.0 and designed using the Qt5 graphic library. It is used for labeling images for tasks such as semantic segmentation and target detection. Moreover, LabelMe supports label generation in VOC and COCO formats.

In this paper, LabelMe was used to delineate the shape and location of grape leaf spots by drawing closed regions through polygons. The labeled data were stored in JSON format, and the data labels were subsequently converted into binary PNG images using the json_to_dataset command. In these binary images, the black areas represent the background, while the red areas represent the leaf spots, as illustrated in Figure 1.

In the real environment, the collected image datasets may be affected by weather, light, dust, etc., so being more consistent with the real environment also further improves the robustness and generalization ability of the model. This article performs data enhancement through random transformation, adjusting image brightness and contrast, adding noise and translation. Subsequently, the images were resized to a resolution of 256 × 256 pixels. This process resulted in a total of 2096 experimental data images, thereby creating the grape leaf spot dataset, PD1. Figure 2 illustrates some of the augmentation outcomes.

(**a**) (**b**)

Figure 1. Image annotation status: (**a**) original image; (**b**) image marking results. Black represents the background, and red represents the lesions.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 2. Data augmentation: (**a**) original image; (**b**) flipped and added noise; (**c**) flipped and reduced brightness; (**d**) added noise and reduced brightness; (**e**) flipped and shifted and reduced brightness; (**f**) flipped and shifted.

Using the dataset construction methodology described above, the grape disease images were randomly divided, with 90% allocated to the training set and the remaining 10% designated as the test set. To account for the inherent randomness in this process, multiple tests were conducted to enhance accuracy.

2.2. Data Enhancement

(1) Random rotation transformation

To verify more possibilities, this article simulates multi-angle shooting datasets and uses rotation and mirror-flipping methods for data enhancement. Random rotation is calculated as follows:

Set the pixel coordinates of the image before rotation to (x, y), and its coordinates are expressed as follows:

$$\begin{cases} x = \gamma \cdot \cos(\alpha) \\ y = \gamma \cdot \sin(\alpha) \end{cases} \tag{1}$$

After rotating by angle β, the coordinates of the corresponding pixel point in the image are (x', y'), and the coordinates at this time are expressed as follows:

$$\begin{cases} x' = \gamma \cdot \cos(\alpha - \beta) \\ y' = \gamma \cdot \sin(\alpha - \beta) \end{cases} \tag{2}$$

The equivalent transformation is as follows:

$$\begin{cases} \begin{cases} x' = \gamma \cdot \cos(\alpha)\cos(\beta) + \gamma \cdot \sin(\alpha)\sin(\beta) \\ y' = \gamma \cdot \sin(\alpha)\cos(\beta) - \gamma \cdot \cos(\alpha)\sin(\beta) \end{cases} \end{cases} \tag{3}$$

Putting Formula (1) into Formula (3), we obtain the following:

$$\begin{cases} x' = x\cos(\beta) + y\sin(\beta) \\ y' = y\cos(\beta) - x\sin(\beta) \end{cases} \tag{4}$$

Mirror flipping includes vertical mirror flipping and horizontal mirroring flipping. The vertical mirror flip uses the horizontal midline as the axis and flips vertically, and the horizontal mirror flip uses the vertical midline as the axis and flips horizontally.

(2) Brightness and contrast adjustment

Due to the influence of weather and light, the clarity of the dataset will be affected when collecting the dataset. To better fit the situation of grape diseases in the natural environment, this article expands the dataset by adjusting the brightness and contrast to make the dataset model as close as possible to various situations encountered in the natural environment.

Adjust brightness as follows: Change the brightness of an image by directly adding, subtracting, multiplying, or dividing operations on each pixel value of the image. Let R represent the original RGB value, represent the adjusted RGB value, g is the adjustment factor, and the brightness adjustment formula is as shown in (5).

$$R = R'(1 + g) \tag{5}$$

Adjust contrast as follows: Fine and effective contrast adjustment can be achieved by training a neural network to learn a contrast transformation function. Assuming m to be the median of image brightness, the meanings of R, R′ and g are the same as above, and the specific calculation method is shown in (6).

$$R = m + (R' - m)(1 + g) \tag{6}$$

(3) Add noise as follows:

By adding noise to simulate the interference factors that would appear in the real world, the performance of the segmentation algorithm can be tested and evaluated by adding noise to the image, and the robustness of the algorithm can be further improved.

Gaussian noise is a kind of random noise that obeys Gaussian distribution (also called normal distribution). It is characterized by adding random disturbances in the form of a bell-shaped curve to the image. It has two parameters, mean and variance, where the mean reflects the symmetry The direction of the axis and the variance represent the width of the normal distribution curve, and its probability density function is shown in Equation (7). Among these, the random variable is x, the mathematical expectation is μ, and the variance is σ^2.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{7}$$

Salt-and-pepper noise, also known as impulse noise, is a type of noise commonly used in image processing. Its characteristic is that black pixels or white pixels will appear randomly in the image, or they may appear at the same time. The occurrence of salt-and-pepper noise can be introduced due to sensor failure, signal transmission errors, or other issues during image acquisition.

Salt-and-pepper noise usually causes obvious black and white spots in the image, seriously affecting the look and quality of the image. Gaussian noise will make the image as a whole feel blurry and distorted, reducing the clarity and contrast of the image. This paper uses a combination of Gaussian noise and salt-and-pepper noise to improve the robustness of the algorithm and greatly increase the generalization effect of the model.

2.3. Experiment Platform and Evaluation Metrics

The hardware and software configurations used for the experiments in this paper are shown in Table 1.

Table 1. Experimental hardware and software configuration.

Item	Detail
CPU	12thGenIntel(R) Core (TM)i5-12400@2.50 GHz
RAM	16 GB
Operating system	Windows11 64-bit
CUDA	CUDA 11.6
Python	Python3.7
Optimizer	Adam

In our experiments, this paper employed two evaluation metrics, pixel accuracy (PA) and mean intersection over union (MIoU), to assess the segmentation performance of grape disease images. The formula is as follows, where k denotes the total number of categories, P_{ij} denotes the number of pixels belonging to category i but predicted to belong to category j, P_{ii} denotes the number of pixels correctly predicted, and P_{ij} and P_{ji} denote false positive and false negative results, respectively.

(1) Pixel accuracy (PA)

PA represents the ratio of correctly predicted pixels to the total number of pixels. Its calculation formula is as follows:

$$PA = \frac{\sum_{i=0}^{k} P_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} P_{ij}} \tag{8}$$

(2) Mean intersection over union (MIoU)

MIoU is a widely used evaluation metric in experimental studies of semantic segmentation. It involves calculating the ratio of the intersection between the real and predicted sets to the union of the real and predicted sets for each category, and then calculating the average across all categories. The calculation formula is as follows:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{P_{ii}}{\sum_{i=0}^{k} P_{ij} + \sum_{i=0}^{k} P_{ji} - P_{ii}} \tag{9}$$

2.4. Network Architecture

U-Net is a neural network model that consists of an encoder–decoder architecture as shown in Figure 3. The encoder part uses the CNN architecture as a contraction path to extract image features and reduce resolution, and the contraction path has four sub-blocks, each of which consists of two consecutive 3 × 3 convolutions, the ReLU activation function, and the maximum pooling layer for down-sampling. Two 3 × 3 convolution operations can effectively reduce the neural network complexity and keep the original segmentation accuracy unchanged. In each down-sampling step, the number of feature channels is doubled. The decoder part consists of convolutional blocks containing up-sampling operations to form an extended path to repair the image detail information, locate the boundary of the segmented object, and gradually restore the spatial resolution of the feature map. In the expansion path, the sub-blocks contain two consecutive 3 × 3 convolutions, the ReLU activation function, and the up-sampling inverse convolution layer. Up-sampling expands the feature map to twice its original size and restores missing detail information. Splicing is a unique U-Net feature that clips the low-level detail features captured by the down-sampling process in the same layer and splices them into the high-level semantic features extracted using the up-sampling process. The final output segmentation result combines both the object category recognition basis provided by the low-resolution information

and the accurate positioning segmentation basis provided by the high-resolution features, which improves the problem of insufficient information in the up-sampling process and achieves accurate segmentation.

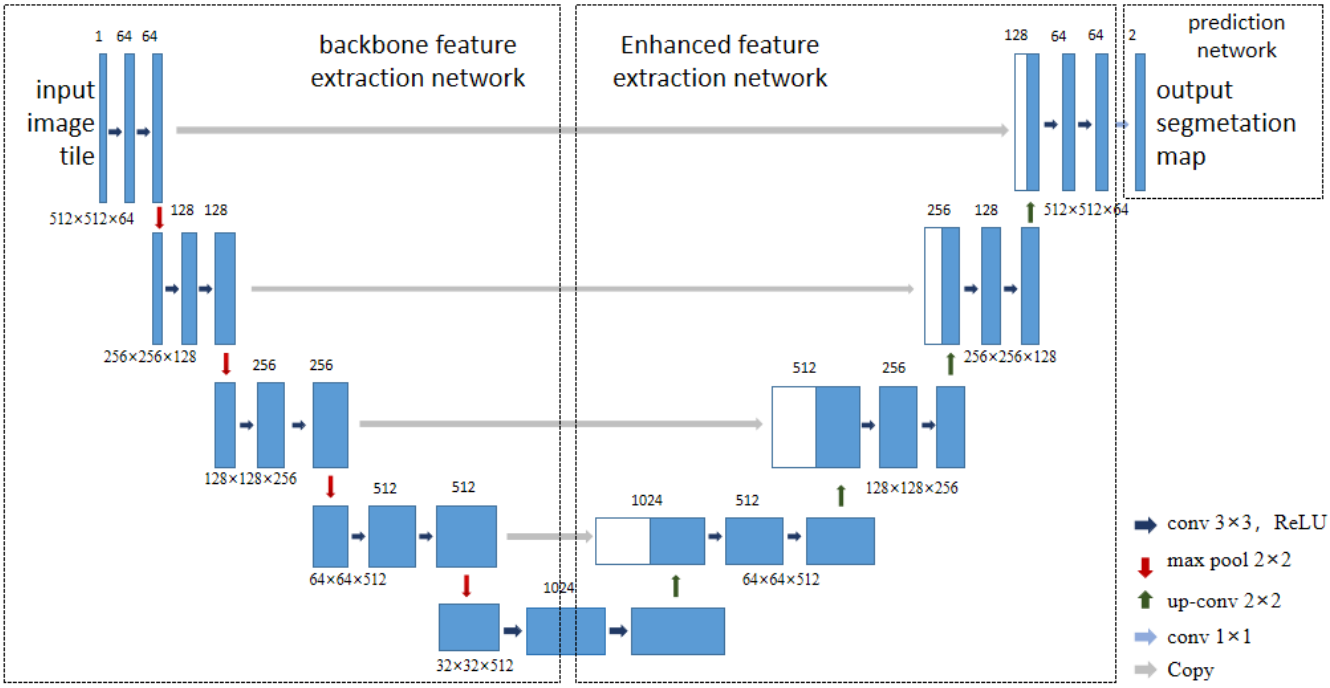


Figure 3. U-Net structure.

The U-Net model achieves excellent segmentation results on a variety of datasets, but the U-Net model also has some shortcomings. Firstly, the redundancy is too large, as each pixel point needs to take a patch, and then the similarity of the patches of two neighboring pixels is very high, which leads to a very large amount of redundancy, resulting in very slow network training. Secondly, high classification accuracy and localization accuracy cannot coexist; when the sensory field is chosen to be larger, the dimensionality reduction multiplier of the corresponding pooling layer behind it will increase, which will lead to lower localization accuracy, but if the sensory field is smaller, then the classification accuracy will be lower. Then, the shallow network information is directly input into the decoder part will cause the poor segmentation of lesion edges. To improve the segmentation performance of the model, and at the same time improve the abovementioned shortcomings, this paper makes the following improvements on the traditional U-Net model structure: (1) Use VGG to replace the U-Net feature extraction network as follows: based on the U-Net framework, the network used for the method feature extraction is replaced with VGG, which greatly improves the training accuracy of the network, and obtains a more accurate segmentation algorithm. (2) Add an ASPP module as follows: replace ordinary convolution with atrous convolution and add spatial pyramid pooling structure to effectively reduce the loss of local information and lack of correlation of long-distance information caused by the gridding effect, and different scale features can be obtained without using pooling layer. (3) Adding CA as follows: CA is added to the feature extraction module and ASPP module to reduce the loss of accuracy to train a more accurate segmentation method and improve the segmentation accuracy of the proposed model for grape disease images. The improved model structure is shown in Figure 4.

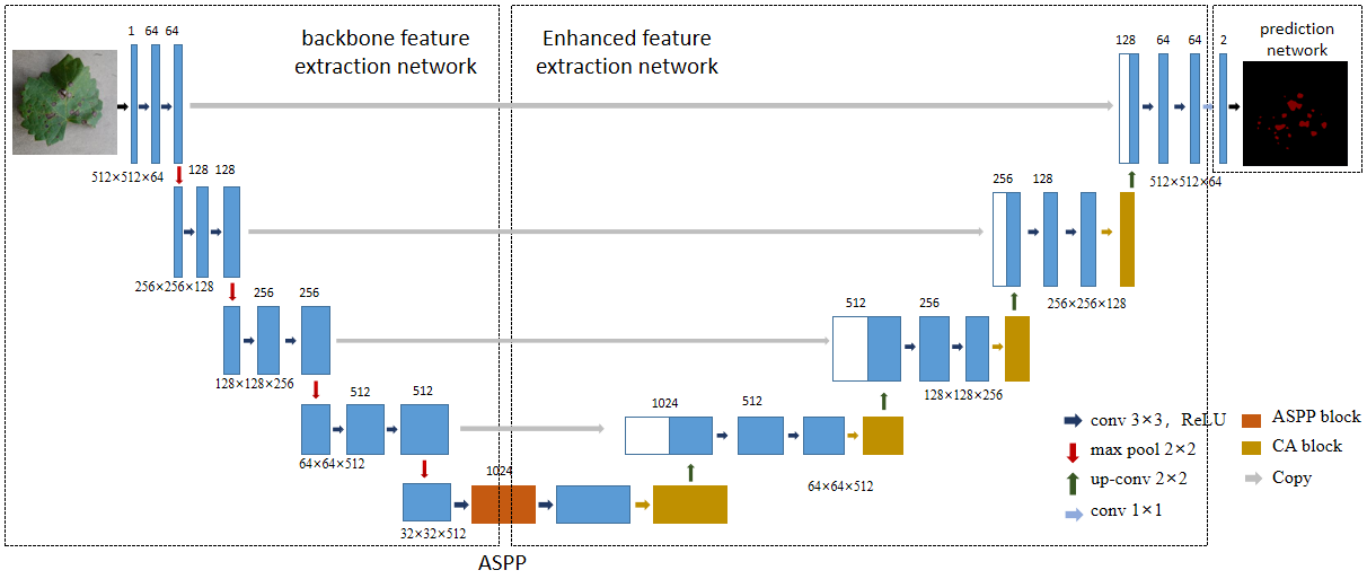
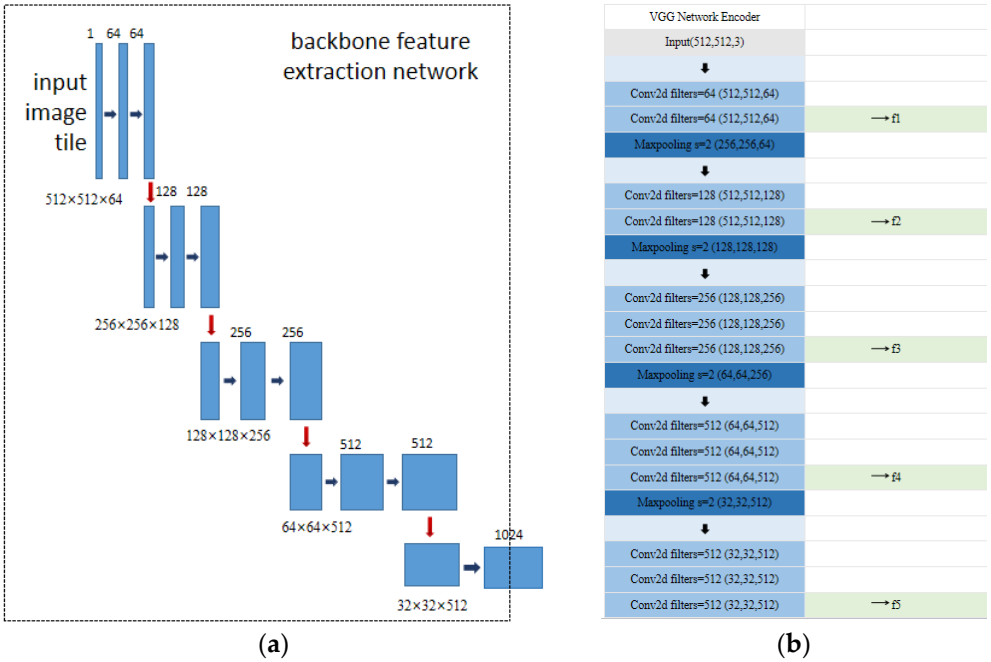


Figure 4. CVU-Net network structure. The orange color block represents the location in which the ASPP module is added, and the yellow color block represents the location in which the attention mechanism is added.

2.5. Optimizing the Feature Extraction Part

In this paper, U-Net is used as the basic framework for constructing the model, and the convolutional layer and the max pooling layer of the VGG16 network are used as the encoder of the U-Net network to improve the efficiency and accuracy of the image feature interpretation to improve the semantic segmentation accuracy of the U-Net and to reduce the influence of other factors on the interpretation accuracy of the model. The backbone feature extraction network part is shown in Figure 5.



(**a**) (**b**)

Figure 5. Backbone feature extraction network structure: (**a**) backbone network feature extraction model; (**b**) backbone feature extraction partial implementation approach.

The encoder can use VGG to obtain feature layer after feature layer for stacking for convolution and max pooling. Five initial effective feature layers can be obtained using the backbone feature extraction part for the next stacking and stitching.

2.6. Attention Mechanism

Attention mechanisms in deep learning select the information that is more critical to the task from a large amount of information, and combining attention mechanisms with fast convolution can better improve the performance of semantic segmentation tasks. To date, the most popular attention mechanism is still squeeze and excite (SE) attention [24]. SENet is designed to enable the network to perform dynamic channel feature recalibration to improve the network's representational capabilities, the structure of which is shown in Figure 6. From the structure, it can be seen that for an input X, it is convolved to obtain a feature map (U), for which an SE module can be attached to attach the channel attention; for U, the spatial information of each of its channels is first compressed to a single value, i.e., a vector of size $1 \times 1 \times C$ is obtained from the U of size $H \times W \times C$. Then, a set of FC layers are applied to the vector to perform a weighting adjustment to obtain a $1 \times 1 \times C$ channel attention vector; finally, the channel attention vector is weighted to U to form a weighted feature map. However, SENet only considers encoding inter-channel information and ignores the importance of positional information, which is crucial for capturing object structure in visual tasks.

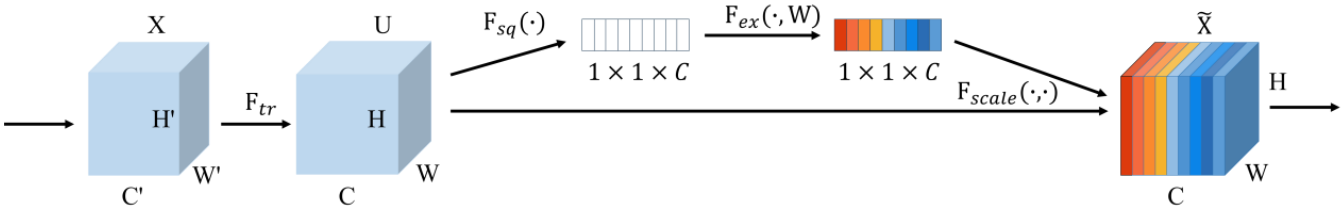


Figure 6. SENet structure.

Subsequent work, such as BAM [25] and CBAM [26], exploits positional information by reducing the number of channels followed by a large-size convolution, which is then used to compute spatial attention. However, convolution can only establish local relationships but cannot model the long-term dependencies necessary for visual tasks.

Coordinate attention (CA) [27], on the other hand, enables lightweight networks to pay attention to a large area while avoiding incurring large computational overheads by embedding location information into the channel attention. To mitigate the loss of location information caused by 2D global pooling, CA decomposes the channel attention into two parallel 1D feature encoding processes to efficiently integrate spatial coordinates to input information into the generated attention graph. Specifically, CA uses two 1D global pooling operations to aggregate input features along vertical and horizontal directions into two separate direction-aware feature maps, respectively. These two feature maps embedded with direction-specific information are then encoded into two separate attention maps, each capturing the long-range dependencies of the input feature maps along one spatial direction. Thus, location information can be provided in advance in the generated attention maps. The two attention maps are then applied to the input feature map using multiplication to emphasize the representation of interest. Because this attention operation distinguishes spatial directions (i.e., coordinates) and generates coordinate-aware feature maps, the proposed method is referred to as coordinate attention.

The CA module encodes channel relations and long-range dependencies through precise position information, similar to the SE module, which is also divided into two steps, coordinate information embedding and coordinate attention generation; its specific structure is shown in Figure 7.

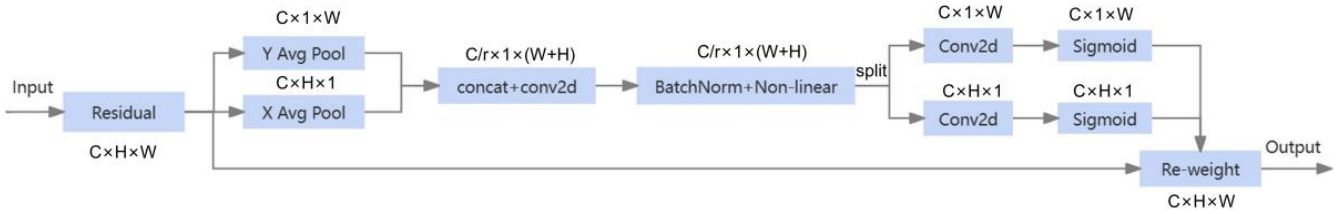


Figure 7. CA structure.

For an input X, each channel is initially encoded using a pooling kernel of size (H, 1) along the horizontal coordinate direction or a pooling kernel of size (1, W) along the vertical coordinate direction. This results in the expression of the output for the cth channel with a height of h as follows:

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} X_C(h, i) \tag{10}$$

Similarly, the output of the cth channel with width w is expressed as follows:

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \le i < H} X_C(j, w) \tag{11}$$

With the above transformation, the features are aggregated along two directions, resulting in a pair of direction-aware feature maps that are able to obtain a global sensory field and accurately encode positional information. They are then modified using the 1×1 convolutional transform function F_1 as follows:

$$f = \delta\left(F_1\left(\left[z^h, z^w\right]\right)\right) \tag{12}$$

where $[\cdot,\cdot]$ denotes the cascade operation along the spatial dimension, and δ is a nonlinear activation function, which is an intermediate feature map that encodes spatial information in the horizontal and vertical directions. This is the shrinkage rate used to control the size of the SE block. Then, f is decomposed into two independent tensors $f^h \in R^{C/r \times H}$ and $f^w \in R^{C/r \times W}$ along the spatial dimension, F_h and F_w are transformed using two additional 1×1 convolutions, and f^h and f^w are transformed into tensor inputs X with the same number of channels to obtain g^h and g^w, respectively, as follows:

$$g^h = \sigma\left[F_h\left(f^h\right)\right] \tag{13}$$

$$g^w = \sigma[F_w(f^w)] \tag{14}$$

where σ represents the sigmoid function. To reduce the computational overhead and model complexity, papers typically decrease the number of channels in f using an appropriate shrinkage rate r. Subsequently, g^h and g^w are expanded and employed as attention weights, respectively. Finally, the output Y of the CA module can be expressed as:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \tag{15}$$

2.7. FPN-Based Feature Fusion Branching

During the process of CNN learning image features, the image resolution gradually decreases due to deep convolutional operations. This can result in lower-resolution deep features at the output, leading to recognition errors for objects that occupy a relatively small percentage of pixels in the image. To enhance multi-scale detection accuracy, it is beneficial to combine features from different network layers during training.

Feature pyramid network (FPN) [28] is a method used for fusing feature maps from different layers to enhance the feature extraction process. Its specific structure is depicted

in Figure 8. FPN can fuse feature maps that capture different scales of information. As illustrated in the figure, FPN generates a new set of deep features by up-sampling the deep features twice, stacking them with the shallow features, and then convolving them to produce a new set of deep features. Feature fusion occurs sequentially, allowing the prediction network to incorporate five preliminary and effective feature maps generated by the VGG component of the U-Net backbone network. The fused feature map contains richer semantic and spatial information because it incorporates features from various levels. This enrichment contributes to the improved segmentation performance of the U-Net network.

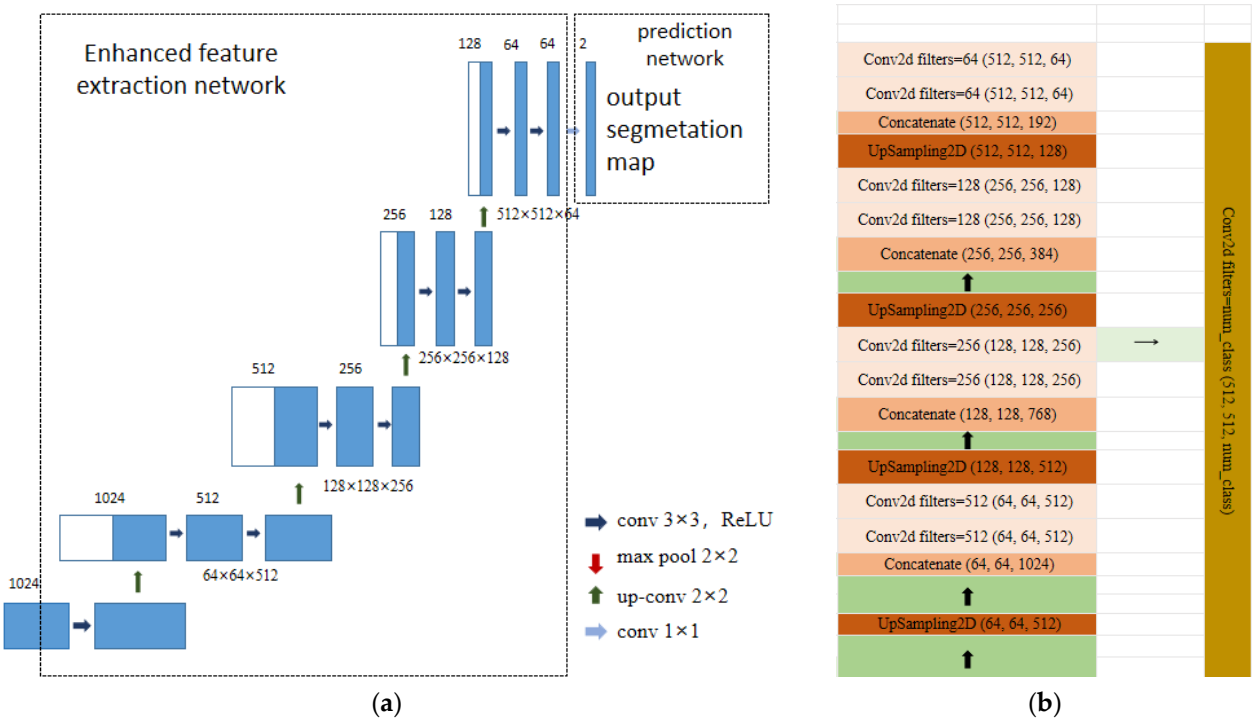


Figure 8. Enhancement of the structure of the part of the feature extraction network. (**a**) Enhanced feature extraction partial model; (**b**) enhancement of the feature extraction component implementation approach.

Feature layer after feature layer can be obtained using VGG for stacking for convolution and max pooling. Five initial valid feature layers can be obtained using the backbone feature extraction part for the next stacking and stitching.

2.8. Multi-Scale Feature Fusion for Hollow Space Pyramid Pooling ASPP

The pooling operation of the semantic segmentation network in the process of expanding the receptive field and aggregating contextual information makes it easy to lose position and dense semantic information, while atrous convolution reduces the dependence on parameters and calculation processes on the basis of ensuring the image resolution properties. It requires fewer parameters to achieve the expansion effect of the effective receptive field of the convolution kernel and effectively aggregate contextual information. Consider a 2D atrous convolution that applies atrous convolution on the input feature map x for each position i and filter w of the input feature map y, as follows:

$$y[i] = \sum_{k=1}^{K} x[i + r \cdot k] \cdot w[k] \tag{16}$$

where k denotes the convolution kernel size and r denotes the sampling rate. The above formula indicates that a new filter is obtained by inserting $r - 1$ zero values along each spatial dimension between two consecutive filter values. Then, the feature mapping x is convolved through this filter to obtain the final feature map. Consequently, atrous

convolution can control the sensory field of the filter and the compactness of the network output features by adjusting the sampling rate, all without increasing the number of parameters or computational effort.

Multi-scale fusion's atrous spatial pyramid pooling ASPP uses atrous convolution with multi-level atrous sampling rates to sample feature maps in parallel, allowing the ASPP module to learn image features from different receptive fields [29]. Because the dilated convolution with a large sampling rate will degenerate into a 1×1 convolution due to the inability of the image boundary response to capture long-range information, the image-level features obtained through global average pooling are integrated into the ASPP module, that is, the image-level features. The feature map outputs by the four convolution branches are input into a 1×1 convolution layer and then bilinearly up-sampled to a specific spatial dimension. The calculation process is as shown in the following formula:

$$Y = \text{Concat}(\text{image}(X), H_{1,1}(x), H_{6,3}(x), H_{12,3}(x), H_{18,3}(x)) \quad (17)$$

In the formula, $H_{r,n}(x)$ represents the atrous convolution with sampling rate r and convolution kernel size n×n on level features, image(x) represents using the global average pooling method to extract image-level features from the input x, and the ASPP structure is shown in Figure 9.

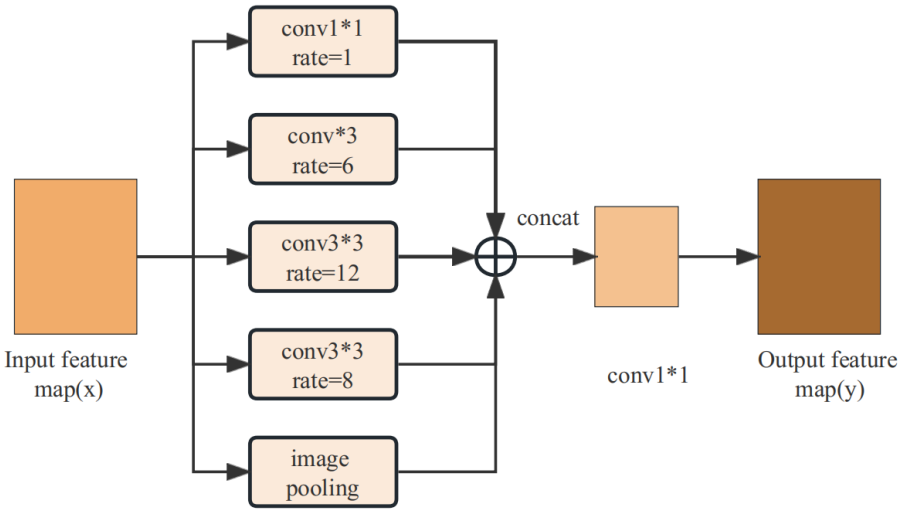


Figure 9. ASPP structure.

The ASPP structure expands the sensory field and enhances semantic information through parallel sampling using atrous convolution at multiple sampling rates. Additionally, image-level features effectively capture global contextual information and account for context relationships, thereby preventing segmentation errors arising from overreliance on local features and ultimately improving target segmentation accuracy. Therefore, before up-sampling, the feature map containing high-level semantic information is input to the ASPP module to obtain features of different scales, which helps to improve the network's lesion extraction performance.

3. Results

3.1. Determination of Training Parameters

Because too small and too large learning rates can lead to very slow model convergence and model non-convergence, it is necessary to determine an appropriate initial learning rate. This article designs and tests the accuracy of the U-Net model trained with four initial learning rates. The results are shown in Figure 10. It can be seen that when the learning rate is 0.0001, the epoch is 100, and the average intersection and merger ratio of the method on the PD1 dataset is 86.81%, which achieves good segmentation results. On this basis, based

on the empirical values of commonly used network training hyperparameters and repeated testing, starting network hyperparameters are provided for subsequent experiments on the CVU-Net model, as shown in Table 2.

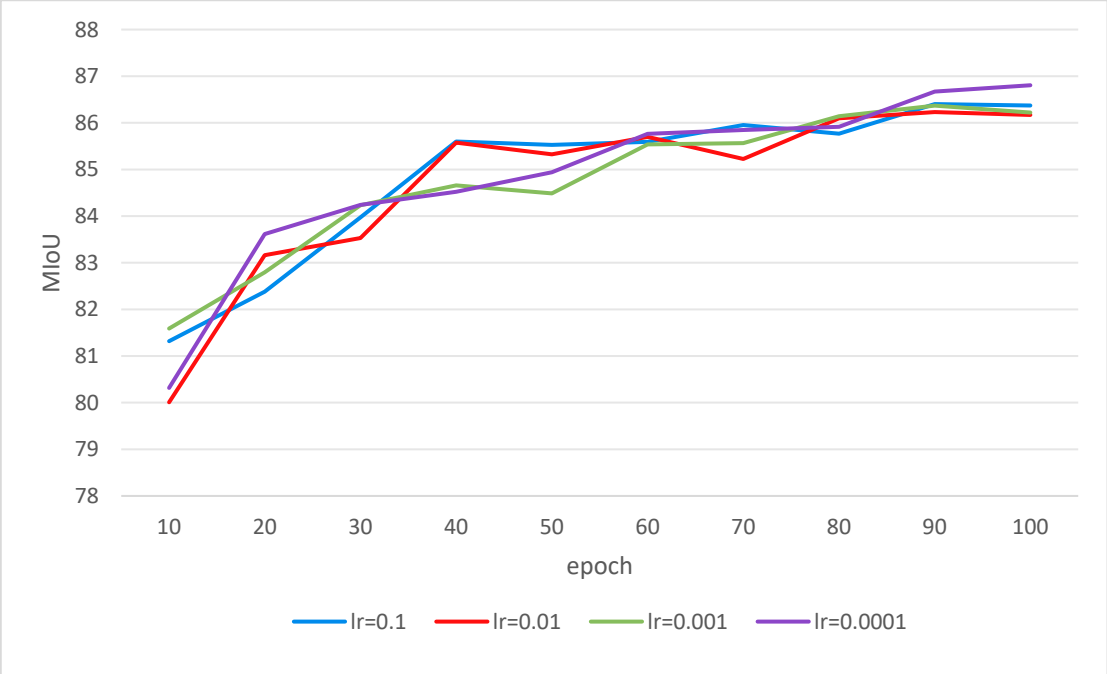


Figure 10. Model average intersection ratio versus learning rate and number of iterations.

Table 2. Training parameters.

Epoch	Batch Size	Lr	Input Shape
100	16	0.0001	512 × 512

3.2. Comparison of Different Attention Mechanisms

To verify the difference between using different attention mechanisms on the detection performance of the algorithm, while controlling other variables consistently, this experiment will add the following four attention mechanisms: SENet, CBAM, ECA, and CA to the original model for comparison and analysis. The original model is an improved U-Net model with added VGG and ASPP modules. Using MIoU and PA as indicators, segmentation experiments were conducted on the grape disease image test set with complex backgrounds. Table 3 shows the comparison results of different attention mechanisms. As can be seen from the table, the MIoU and PA indicators of the CA attention mechanism are the highest, reaching 91.09% and 94.33%, respectively. Therefore, this paper selects CA as the most appropriate attention mechanism based on its performance and uses the training set to evaluate the segmentation performance of the CVU-Net model.

Table 3. Comparative results of different attention mechanisms.

Method	MIoU (%)	PA (%)
original	90.06	92.98
SENet	90.43	93.46
CBAM	90.40	93.44
ECA	90.48	93.66
CA	91.09	94.33

3.3. Ablation Experiments

To assess the performance of the proposed CVU-Net method in the task of grape disease semantic segmentation, it was compared with traditional semantic segmentation methods such as FCN, PSPNet, U-Net, and DeeplabV3+. MIoU and PA were selected as metrics to evaluate the segmentation performance of each method.

To test the generalization ability of CVU-Net and verify its robustness, segmentation and comparison experiments were conducted on the constructed training and test sets. To confirm the effectiveness of the CVU-Net concept, which includes using a network that provides better segmentation than the original feature extraction network (VGG), incorporating an ASPP module into the jump connection section and adding CA to both the enhanced feature extraction module and the ASPP module, the following ablation experiments were performed on the test set.

(1) VU-Net: Based on the traditional U-Net architecture, the feature extraction network is replaced with the VGG network, which has a superior segmentation effect.
(2) AVU-Net: Building upon VU-Net, an ASPP module is integrated into the jump connection layer.
(3) CVU-Net1: Extending AVU-Net, CA is introduced into the enhanced feature extraction module.
(4) CVU-Net: Further enhancing AVU-Net, CA is integrated into both the enhanced feature extraction module and the ASPP module.

Table 4 presents the experimental results of different configurations on the PD1 dataset. It is evident that CVU-Net outperforms the other configurations, indicating that the addition of the CA module after the feature extraction module and ASPP module effectively enhances the model's segmentation capabilities.

Table 4. CVU-Net ablation experiment results.

Method	MIoU (%)	PA (%)
U-Net	86.16	90.66
scheme1	89.37	92.61
scheme2	90.06	92.98
scheme3	90.78	93.67
scheme4	91.09	94.33

3.4. Fivefold Cross Validation

To further compare the performance of different models or parameter settings, we found the best model or parameter configuration. Fivefold cross-validation experiments were performed for different parameter selections. We divided the dataset into five equally sized subsets. In each iteration of the fivefold cross-validation, four of the five subsets were used to train the model, while the remaining subsets were used to test its performance. This process was performed five times, ensuring that each subset was used once as a test set. We chose four parameter schemes, as shown in the Table 5. For the performance indicators MIoU and PA, the mean of five cross-validations was calculated, and the experimental results are shown in the Table 6.

Table 5. Specific parameter settings for different solutions.

Scheme	Bitch Size	Lr
scheme1	8	0.001
scheme2	8	0.0001
scheme3	16	0.001
scheme4	16	0.0001

Table 6. Fivefold cross-validation results.

Method	MIoU (%)	PA (%)
scheme1	90.88	92.61
scheme2	91.03	92.98
scheme3	91.07	93.67
scheme4	91.13	94.33

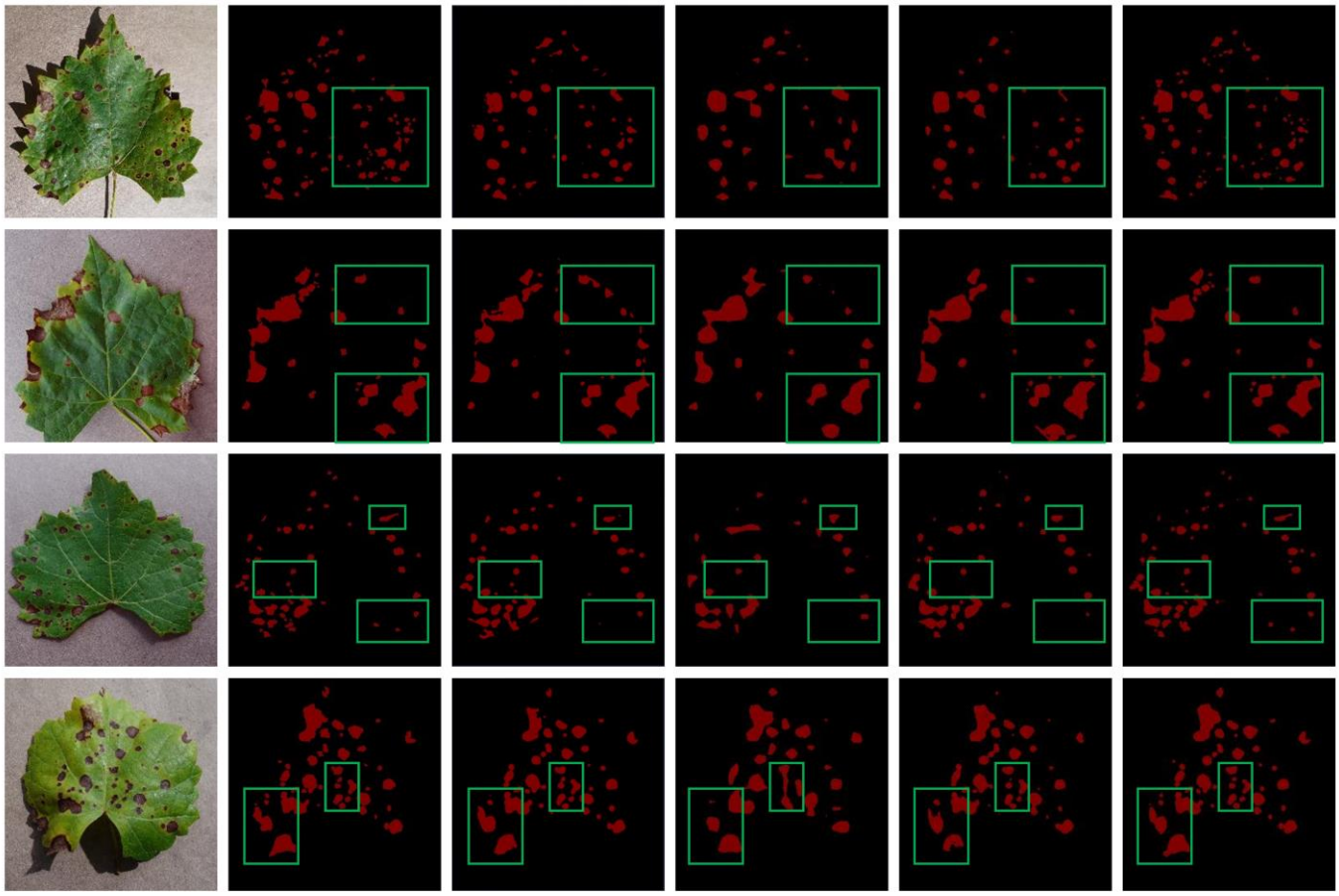
It can be seen from the experimental results that when the bitch size is 16 and the learning rate is 0.0001, the values of MIoU and PA are the highest, reaching 91.18% and 94.40%, respectively. After weighing the evaluation indicators of different schemes, we finally chose scheme4 to carry out the next experiment.

3.5. Performance Comparison of Different Segmentation Methods

This paper compared CVU-Net with traditional U-Net, PSPNet, and DeeplabV3+. The comparison results of different segmentation algorithms are presented in Table 7. As shown in the table, the improved method in this paper achieves a pixel accuracy (PA) of 94.33%, which is 3.67%, 3.57%, and 5.41% higher than that of the traditional U-Net algorithm, PSPNet algorithm, and DeeplabV3+ algorithm, respectively. Regarding the mean intersection over union (MIoU), the improved method in this paper attains a value of 91.13%. In terms of MIoU, it outperforms the traditional U-Net algorithm, PSPNet algorithm, and DeeplabV3+ algorithm by 4.97%, 5.51%, and 5.44%, respectively. These experimental results demonstrate that the incorporation of the depth attention mechanism in this paper's method enhances the model's feature extraction capability and significantly improves the accuracy of grape semantic segmentation. The visualization of the segmentation results is shown in Figure 11. In Figure 11, the first column represents the original grape leaf images, the second column depicts the manually labeled images, the third column displays the segmentation results from the DeeplabV3+ model, the fourth column shows the segmentation results from the PSPNet model, the fifth column exhibits the segmentation results from the U-Net model, and finally, the sixth column demonstrates the segmentation results from the CVU-Net model proposed in this paper. It can be seen from the visualization results that the U-Net model segmentation is more accurate, but small lesions will be missed and misidentified; the PSPNet model is not effective and will identify dense small lesions as one large lesion. Spot edge detection is not accurate enough; the DeeplabV3+ model will miss the detection of small lesions and produce unclear edge segmentation of large lesions. CVU-Net is more accurate in segmenting the edges of lesions and small lesions, which is basically consistent with the annotation situation, and can achieve very good accuracy results. The visualization results prove that adding the ASPP module can enhance the model's perception of the input image and capture a wider range of contextual information. Adding CA to the feature extraction module and ASPP module can help the model further learn the correlation between features and focus on important feature channels to more accurately segment the lesion area and lesion edge.

Table 7. Comparison results of different segmentation methods.

Method	MIoU (%)	PA (%)
U-Net	86.16	90.66
PSPNet	85.62	90.76
DeeplabV3+	85.69	88.92
CVU-Net	91.13	94.33



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 11. Segmentation effect of different algorithms: (**a**) original image; (**b**) ground truth; (**c**) U-Net; (**d**) PSPNet; (**e**) DeeplabV3+; (**f**) CVU-Net. The green boxes represent areas where there is a large difference between the different methods.

3.6. Disease Spot Grading and Comparison Experiments

Because there is no clear grading standard for the degree of grape leaf spots, to more accurately analyze the grading of the degree of grape leaf black rot spots, this paper takes the standard Grapevine Downy Mildew Disease classification method [30] developed by the People's Republic of China as a reference to develop a grading standard for grape black rot leaf spots. This paper is based on the principle of pixel point statistics. Using Python to achieve the statistics of the area of the disease spot, the leaves are divided into three levels, as follows: level 1, level 2, and level 3. the specific grading standards are shown in Table 8.

Table 8. Grading criteria for black rot spots on grapes.

Disease Spot Level	The Range of k	Quantity
Level 1	$0 \leq k \leq 5\%$	808
Level 2	$5\% \leq k \leq 25\%$	800
Level 3	$25\% \leq k \leq 50\%$	488

Where k is the proportion of the diseased area to the whole image, the principle calculation formula is as follows:

$$k = \frac{A_{spot}}{A_{image}} = \frac{\sum (x, y) \in R_{spot} n}{\sum (x, y) \in R_{image} n} \quad (18)$$

In the formula, A_{spot} is the area of the lesion area, A_{image} is the area of the whole image, R_{spot} indicates the lesion area, and R_{image} indicates the image area.

To measure the effectiveness of this model, based on the grading PD1, a comparison experiment was conducted using the traditional U-Net, VGG + U-Net, and ASPP + VGG + U-Net with the method of this paper, as shown in the following table, in which VU-Net denotes the model of the traditional U-Net introducing the VGG network, and AVU-Net denotes the model of the traditional U-Net introducing the VGG network and the model of ASPP module.

As can be seen from Table 9, the highest segmentation accuracy of all models for the level 3 category in the experiment may be due to the larger area of the level 3 leaf spot. A comparison of the segmentation accuracy of each model for the level 3 lesions is shown in Figure 12.

Table 9. Comparative accuracy experiments of different models.

Model	PA (%)		
	Level 1	Level 2	Level 3
U-Net	78.97	87.98	90.69
VU-Net	79.47	89.19	92.26
AVU-Net	80.87	90.3	93.53
CVU-Net	84.55	91.33	94.69

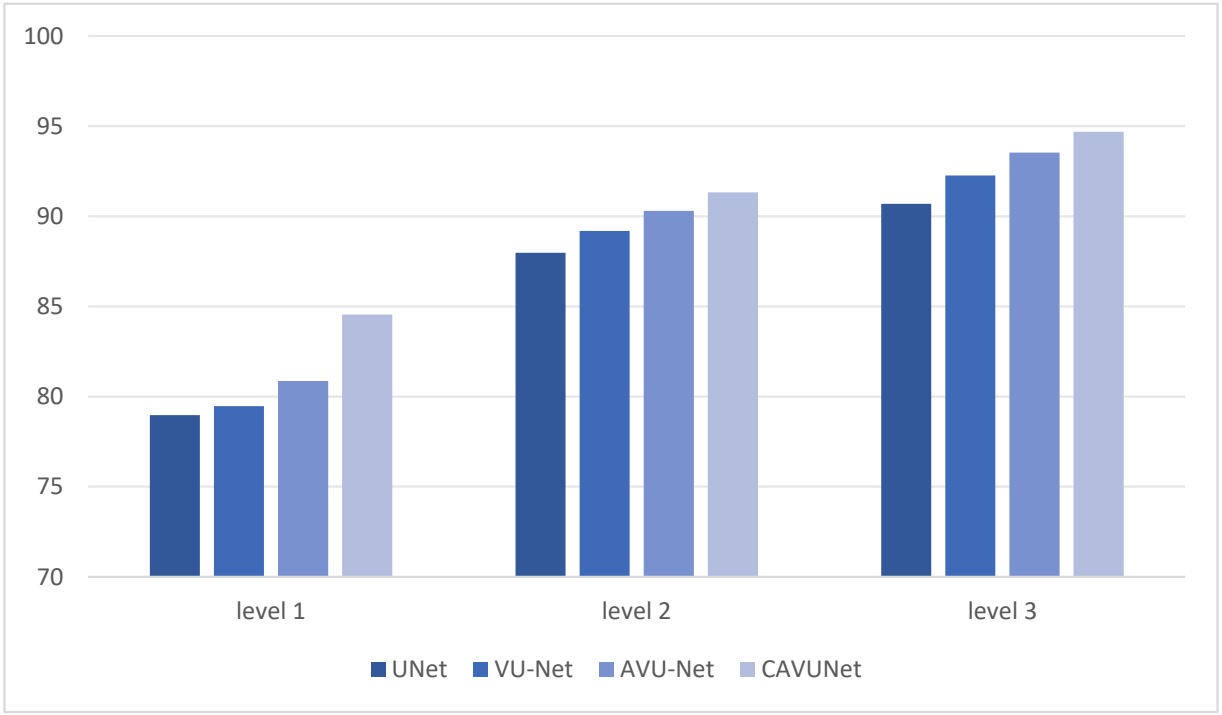


Figure 12. Comparison of segmentation accuracy of each model for graded lesions.

It can be clearly seen from Figure 12 and Table 9 that compared with the U-Net model, VU-Net model, and AVU-Net model, the segmentation accuracy of the level 3 category of this model has increased by 4.0%, 2.43%, and 1.16%, respectively. For the other two types of lesion levels, this model is improved compared to the U-Net model, VU-Net model, and AVU-Net model.

4. Discussion

The main work of this article includes the following four parts: first, improve the segmentation accuracy of the algorithm, improve the segmentation accuracy for low-level disease categories, and effectively improve the algorithm's segmentation accuracy for low-level disease categories; second, select further research and experiments will be carried out on grape leaves with different degrees of disease; the third is to study how to reduce the interference of uncertain factors, such as noise and shadows in the image, on the segmentation accuracy of the algorithm; and the fourth is to conduct further research on the unclear segmentation of lesion edges and the misdetection or missed detection of small lesions.

It can be seen from the experiments that the method CVU-Net proposed in this paper can extract the diseased areas in the images more effectively than methods such as U-Net, DeeplabV3+, and PSPNet. The PA of the whole grape disease image dataset reaches 94.33%, and MioU reaches 91.09%, which are 4.93% and 3.67% higher than the traditional U-Net network, respectively. The robustness of CVU-Net was fully verified by comparing it with the other three semantic segmentation methods on the grape disease test set. Although CVU-Net segmented the grape disease image more accurately than the other test methods, its segmentation of the occluded region was not accurate for the grape disease leaves that were occluded by leaves in more complex cases. Therefore, we recommend constructing a relevant dataset and conducting further experimental studies in the future to address this issue.

5. Conclusions

In response to the low accuracy of grape disease image segmentation, this paper proposes a segmentation method CVU-Net based on a deep learning network. Our method combines the U-Net model with the VGG network, significantly improving the training accuracy of the network and achieving more precise segmentation results.

We incorporate the ASPP module into the skip connection part, expanding the receptive field and aggregating context information to avoid the loss of position information and dense semantic information caused by pooling operations while reducing the dependence on parameters and calculation processes. It can help the model better capture the edge information of the image and retain the detailed features of the image, allowing the model to produce more refined and accurate segmentation results.

In this paper, we introduce CA into the feature extraction module and ASPP module, which can better restore the edge information of objects and further improve the feature extraction capabilities of the method, reducing missed objects. Experiments on PD1 show that our method can effectively extract the areas of grape leaf black rot disease spots and achieve more accurate and efficient segmentation of disease spots. However, the segmentation effect on other disease images of grape leaves is unknown. In the next step, we will pre-train the model on other grape disease image datasets to achieve the segmentation and recognition of different diseases in real environments.

Author Contributions: Supervision, methodology, X.Y.; software, writing—original draft, Y.Z.; conceptualization, investigation, P.W.; formal analysis, resources, G.W. and L.M.; writing—review and editing, M.C., X.F. and P.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Acknowledgments: We thank Yi Xiaomei, Wu Peng, and Wang Guoying for their help in providing road guidance and material collection for this study.

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

References

1. Yuan, H.; Zhu, J.; Wang, Q.; Cheng, M.; Cai, Z. An improved DeepLab v3+ deep learning network applied to the segmentation of grape leaf black rot spots. *Front. Plant Sci.* **2022**, *13*, 795410. [CrossRef] [PubMed]
2. Alajas, O.J.; Concepcion, R.; Dadios, E.; Sybingco, E.; Mendigoria, C.H.; Aquino, H. Prediction of Grape Leaf Black Rot Damaged Surface Percentage Using Hybrid Linear Discriminant Analysis and Decision Tree. In *2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 25–27 June 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
3. Dutta, K.; Talukdar, D.; Bora, S.S. Segmentation of unhealthy leaves in cruciferous crops for early disease detection using vegetative indices and Otsu thresholding of aerial images. *Measurement* **2022**, *189*, 110478. [CrossRef]
4. Liu, Z.; Du, Z.; Peng, Y.; Tong, M.; Liu, X.; Chen, W. Study on corn disease identification based on pca and svm. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; IEEE: Piscataway, NJ, USA, 2020; Volume 1, pp. 661–664.
5. Khan, M.A.; Lali MI, U.; Sharif, M.; Javed, K.; Aurangzeb, K.; Haider, S.I.; Altamrah, A.S.; Akram, T. An optimized method for segmentation and classification of apple diseases based on strong correlation and genetic algorithm based feature selection. *IEEE Access* **2019**, *7*, 46261–46277. [CrossRef]
6. Ambarwari, A.; Adrian, Q.J.; Herdiyeni, Y.; Hermadi, I. Plant species identification based on leaf venation features using SVM. *Telkomnika Telecommun. Comput. Electron.Control.* **2020**, *18*, 726–732. [CrossRef]
7. Appeltans, S.; Pieters, J.G.; Mouazen, A.M. Detection of leek white tip disease under field conditions using hyperspectral proximal sensing and supervised machine learning. *Comput. Electron. Agric.* **2021**, *190*, 106453. [CrossRef]
8. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
9. Zhang, D.; Wang, D.; Gu, C.; Jin, N.; Zhao, H.; Chen, G.; Liang, H.; Liang, D. Using neural network to identify the severity of wheat Fusarium head blight in the field environment. *Remote Sens.* **2019**, *11*, 2375. [CrossRef]
10. 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] [PubMed]
11. 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.
12. Kumar, D.; Kukreja, V. Application of PSPNET and fuzzy Logic for wheat leaf rust disease and its severity. In Proceedings of the 2022 International Conference on Data Analytics for Business and Industry (ICDABI), Virtual, 25–26 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 547–551.
13. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
14. Chen, L.C.; 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]
15. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
16. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
17. Cai, M.; Yi, X.; Wang, G.; Mo, L.; Wu, P.; Mwanza, C.; Kapula, K.E. Image segmentation method for sweetgum leaf spots based on an improved DeeplabV3+ network. *Forests* **2022**, *13*, 2095. [CrossRef]
18. Wu, P.; Cai, M.; Yi, X.; Wang, G.; Mo, L.; Chola, M.; Kapapa, C. Sweetgum Leaf Spot Image Segmentation and Grading Detection Based on an Improved DeeplabV3+ Network. *Forests* **2023**, *14*, 1547. [CrossRef]
19. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015, Part III 18*; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
20. Yi, X.; Wang, J.; Wu, P.; Wang, G.; Mo, L.; Lou, X.; Liang, H.; Huang, H.; Lin, E.; Maponde, B.T.; et al. AC-UNet: An improved UNet-based method for stem and leaf segmentation in Betula luminifera. *Front. Plant Sci.* **2023**, *14*, 1268098. [CrossRef] [PubMed]
21. Liu, S.; Huang, Z.; Xu, Z.; Zhao, F.; Xiong, D.; Peng, S.; Huang, J. High-throughput measurement method for rice seedling based on improved UNet model. *Comput. Electron. Agric.* **2024**, *219*, 108770. [CrossRef]
22. Chen, S.; Zhang, K.; Zhao, Y.; Sun, Y.; Ban, W.; Chen, Y.; Zhuang, H.; Zhang, X.; Liu, J.; Yang, T. An approach for rice bacterial leaf streak disease segmentation and disease severity estimation. *Agriculture* **2021**, *11*, 420. [CrossRef]
23. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
24. 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, 18–22 June 2018; pp. 7132–7141.
25. Park, J.; Woo, S.; Lee, J.Y.; Kweon, I.S. Bam: Bottleneck attention module. *arXiv* **2018**, arXiv:1807.06514.
26. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
27. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.

28. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
29. Habijan, M.; Galić, I.; Romić, K.; Leventić, H. AB-ResUNet+: Improving Multiple Cardiovascular Structure Segmentation from Computed Tomography Angiography Images. *Appl. Sci.* **2022**, *12*, 3024. [CrossRef]
30. Zhang, X.; Guo, J.; Wang, Y.; Wang, Q.; Li, Z.; Wang, X. Effects of Shelter Cultivation on the Growth and Disease Occurrence of Table Grape. *Acta Bot. Boreal.-Occident. Sin.* **2023**, *43*, 255–264.