

Article

Not peer-reviewed version

Experimental Design of *C. oleifera* diseases and pest Segmentation Based on CDM-DeeplabV3+ with A Residual Attention ASPP and Dual Attention Encoder

[Lin Li](#)*, Ruitian Guo, [Yongzhong Zhang](#)

Posted Date: 7 October 2024

doi: 10.20944/preprints202410.0402.v1

Keywords: Camellia oleifera diseases; Artificial intelligence; Semantic Segmentation; Experimental teaching; Science and education integration



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Experimental Design of *C. oleifera* Diseases and Pest Segmentation Based on CDM-DeeplabV3+ with A Residual Attention ASPP and Dual Attention Encoder

Lin Li *, Ruitian Guo and Yongzhong Zhang

School of Electronic Information and Physics, Central South University of Forestry and Technology, Changsha, 410004, Hunan, China; 20211609@csuft.edu.cn (R.G.); t20010595@csuft.edu.cn (Y.Z.)

* Correspondence: t20060540@csuft.edu.cn(L.L.);

Abstract: In order to deepen students' understanding of the deep learning network development and application process, this paper designed a deep learning-based teaching experiment simulation method for *C. oleifera* diseases and pest segmentation. The experimental process includes training data collection and preprocessing, network model training, result verification and analysis, etc., throughout the whole process of deep learning-based image segmentation task development. To this end, CDM-DeeplabV3+ network was proposed to segment *C. oleifera* leaf diseases fast and accurately, which was composed of a Residual Attention ASPP(CBAM-ASPP) and Dual Attention Encoder (DAM-Encoder). CBAM-ASPP obtained multiscale information by extending to five dilated convolutions and focused on the edge features using Convolutional Block Attention Module. The DAM-Encoder structure, which was connected with CBAM-ASPP in parallel, highlighted the tiny features by combining global and local features. Then, the *C. oleifera* diseases and pest dataset was built, and the model was compared with the traditional segmentation models DeeplabV3+, UNet, HrNetV2 and PSPNet. The experimental results show that the mIoU of CDM-DeeplabV3+ reaches 85.63%, which is 5.82% higher than that of the original model and the model parameters are reduced by nearly five times, proving the effectiveness and feasibility of segmenting pests and diseases on *C. oleifera*. The design of the experimental teaching content is inspired by the scientific research experiments of artificial intelligence, following the frontier of the discipline and easy to realize, reflecting the teaching concept of the integration of science and education.

Keywords: *Camellia oleifera* diseases; artificial intelligence; semantic segmentation; experimental teaching; science and education integration

1. Introduction

Camellia oleifera Abel, a species within the *Dungarunga* group or shrub category, holds a place in the esteemed *Camellia* family, scientifically known as *Theaceae*. It has high values in economic, edible, environmental and medicinal field [1]. In 2019, there were 4.4 million hectares of land used to grow *C.oleifera* in China, in which production exceeded 2.4 million tons [2]. *C.oleifera* cultivation is still facing great hazards from diseases and insect pests, some of which usually lead to 20%-40% reduction in fruit percentage and 10%--30% loss of seeds. It may cause the branches to wilt, even the entire plant to die in severe cases. Meanwhile, serious economic losses will occur, as disease spread rapidly and are difficult in control. At present, the control measures for diseases and insect pests of *C. oleifera* predominantly adopted are chemical interventions, and complemented by tea plantation sanitation. However, diseases are characterized by a lengthy incubation period and rapid dissemination, which is a challenge to determine the optimal timing for conventional control

measures. Conrad et al. [3] emphasize that manual detection of diseases is costly and inefficient. There is an pressing need for a fast, and accurate detection and segmentation method. However, conventional image processing techniques struggle to swiftly and accurately identify multiple diseases in intricate environments [4].

With the modernization, and intelligence of agriculture, accurately identifying disease categories and taking corresponding measures are significant. In agriculture, the rapid advancement and adoption of machine learning have proven to be a powerful and efficient solution [5]. Conventional mechanical learning methods, K-mean clustering, Markov Random Fields, Random Forests, and Support Vector Machines had been applied in the field of image segmentation for many years [6]. However, they exhibited inefficiencies in computation, limited scalability, and poor generalization performance, especially for handling complex data and tasks. Thanks to the groundbreaking advancements in deep learning, remarkable strides have been achieved in the field of plant disease segmentation [7,8]. Deep learning models focused on the efficient processing and full utilizing for the contextual information of images, through automatic feature extraction. Convolutional Neural Networks (CNNs), as a important component of deep learning, have emerged as the preferred network architecture for numerous model. Lu et al. [9] reviewed the application of CNN on diseases and insect pests of plant leaf. The convolutional layers were used to effectively capture local information in CNN. By stacking multiple layers of convolutional and pooling layers, CNN can learned hierarchical features, then efficiently comprehended both global and local contexts within the image. Consequently, the realm of image segmentation had witnessed the emergence of numerous innovative CNN-based architectures. Ramcharan et al. [10] applied transfer learning to train neural networks for the recognition about diseases of cassava, a significant economic crop. Lu et al. [11] introduced an automatic diagnosis system utilizing a supervised deep learning framework for identifying wheat disease. Moreover, it was packed into a real-time mobile application. Liu et al. [12] introduced an innovative approach for grape leaf disease classification, which relied on an improved CNN. In the network, the Inception structure and dense connectivity strategy were introduced to boost the extraction of multidimensional features. For detecting five common leaf diseases of apple trees, the Inception and Rainbow were combined in a new deep learning network by Jiang et al. [13]. Wang et al. [14] developed a new model that combined U-Net and DeeplabV3+, specifically tailored for cucumber foliar disease segmentation in challenging background. Deng et al. [15] proposed a cross-layer attentional mechanism combined with a multiscale convolutional module, which could greatly solved the problem of fuzzy edges and tiny disease in tomato foliar diseases.

Based on the above studies, we can infer that the segmentation of diseases and insect pests on important economic crops contributed to the development of agriculture. *C. oleifera* is also considered an economically significant crop, but the current relevant research mainly focused on pathological aspect. Meanwhile, DeeplabV3+ was a leading model of semantic segmentation model, which had the following advantages by comparing with other models such as UNet, UNet++, PSPNet, HrNetV2, et[16]:

- The Encoder-Decoder architecture enhanced feature extraction by utilizing deep networks within the encoder, while the decoder improved the precision of segmentation outcomes through a process of upsampling.
- The use of enhanced dilated convolutions was strategically employed to effectively gather information from various spatial scales, thereby significantly expanding the receptive field of convolutional layers. It allowed the model to capture a broader context.
- The ASPP module enhanced segmentation accuracy through the multiple dilated convolutions of different rates.

Due to the lack of public dataset, we built own dataset of *C. oleifera* diseases and insect pests. The dataset was comprised of 1264 images and 5 different diseases, which all were shot in their natural environments. Then, we utilized the DeeplabV3+ model directly on self-build dataset, but encountered the following challenges:

- The boundary of diseased leaves between the background was fuzzy, as shown in Figure 1. (A). This presented a challenge for the conventional DeeplabV3+ network. It struggled to effectively differentiate between background and disease features, resulting in a notable reduction in feature extraction accuracy.

- The gridding of dilated convolutions increased the loss of tiny disease information. The average pooling layer also lost edge and texture information. This presented a difficulty for the model in accurately segmenting these small diseased areas, as shown in Figure 1B.

- DeeplabV3+ adopted Xception as its backbone, which possessed a large number of parameters. The backbone produced complex computations and long segmentation duration. Consequently, it was unfit to be applied in practical crop production.

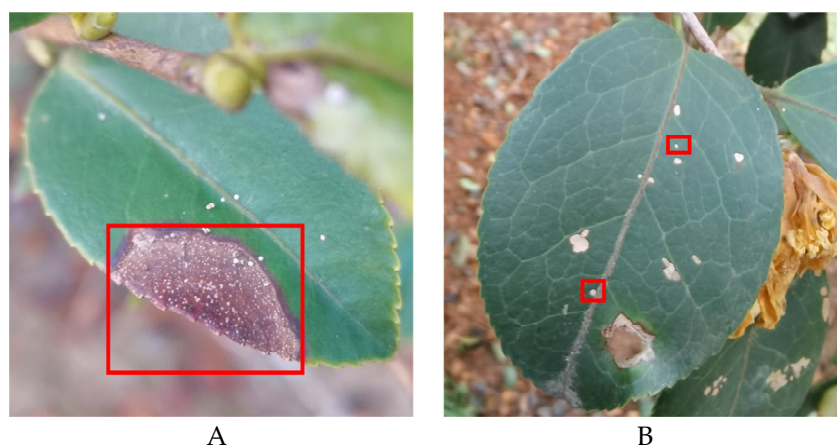


Figure 1. Figure A shows that the edges of the annotated portions are blurred with the background due to the similar background color. Figure B shows that the disease of the annotated portions is too tiny.

To solving the problem of fuzzy boundary between features and background, Khadidos et al. [17] introduced an approach that relied on level set evolution. The method assigned weights to energy terms based on their significance in boundary delineation, which served the dual purpose of improving both the precision of feature extraction and the accuracy of boundary segmentation. Xia et al. [18] integrated the Reverse Edge Attention Module (REAM) into their encoder architecture of model, strategically placing it between consecutive layers. We proposed a CBAM-ASPP module to solve not only the similarity of color between diseases and background, but also the presence of blurry boundaries. To improve the ability of capture features, we expanded convolution layers by assigning various dilation rate which captured features from different scales. Additionally, we removed the average pooling layer to better retain edge and texture information. By connecting Convolutional Block Attention Module (CBAM) with dilated convolutions, the model was directed to concentrate on salient features and distinguish the edges of diseased leaves more effectively.

The feature extraction of tiny disease also became a challenge. Gao et al. [19] introduced the H-SimAM Attention Mechanism which aimed to concentrate on potential symptoms within the complex background. Xiao et al. [20] developed the Context Enhancement Module and Feature Purification Module to enhance the detection accuracy, particularly for smaller targets. Due to the distinctive encoder-decoder architecture of DeeplabV3+, we have introduced the DANet attention mechanism to create an enhanced encoder. By focusing attention on feature maps via spatial and channel dimensions, the DANet mitigated impact of complex background and challenges of tiny object shapes.

The large model parameters and complex computations was also a serious problem. Liang et al. [21] discussed that lightweight design based on repetitive feature maps mainly concentrated on lightweighting the cheap operations. However, these methods often neglected generalization capabilities of network and huge computational demands in convolutional layers. To address this, the authors implemented a group convolution technique aimed at decreasing the computational

intensity of floating-point operations. MobileNetV2 was adopted as the backbone due to its suitability for mobile and embedded devices. The key role of MobileNetV2 contributed to the design of inverted residual blocks, which were integrated into depthwise separable convolutions and linear bottleneck layers. These innovations enabled the model to effectively decrease parameters without decreasing accuracy of segmentation detection. As a result, MobileNetV2 provided more practical values for mobile devices in agriculture.

The key contributions were encapsulated as follows:

1. We designed an experimental simulation method for insect pest and disease segmentation based on CDM-DeeplabV3+ network and employed a self-build dataset containing 1265 images that had five of diseases and insect pests of *C. oleifera*. All the data analysis in this research were concluded from this proprietary dataset.

2. A Residual Attention ASPP (CBAM-ASPP) was proposed, which extended dilated convolution amount to five and integrates CBAM modules to form a new Attention Convolution Unit. This unit significantly enhanced the extraction of key semantic information across varying scales. Additionally, the module removed the use of the average pooling layer and redirected attention towards edge and texture features, thereby reinforcing the performance of network to accurately segment edge features of leaf diseases.

3. A Dual-Attention Encoder Module (DAM- Encoder) was applied to improve feature extraction capabilities along both spatial and channel dimensions. Within the encoder, we integrated the CBAM-ASPP and DANet modules in a parallel configuration. It captured minutely detailed features from diverse perspectives, effectively addressing the problem of losing tiny features.

4. We changed the backbone to MobilenetV2. Unlike Xception, MobilenetV2 introduced the inverted residual block, which substantially decreased the number of parameters maintaining the precision of segmentation detection. The experimental results show that CDM-DeeplabV3+ achieved an mIoU of 85.63%, which had an increase of 5.82% to original model, and model parameters reduction by almost 5 times compared to original model.

2. Materials and Methods

DeeplabV3+ is an outstanding neural network proposed for image segmentation based on Encoder-decoder architecture [22,23]. The Encoder occupies an important position in feature extraction, while the Decoder refines these features to restore resolution and achieve detailed segmentation. Within the Encoder, the integration of the Xception and ASPP facilitates improved multi-scale feature fusion [24].

2.1. CDM-DeeplabV3+ Network

A simulation experimental method was proposed which conducted using CDM-DeeplabV3+ network for insect pest and disease segmentation. We introduce a CDM-DeeplabV3+ model aim to overcome three primary difficulties encountered in the segmentation: the problem of indistinct boundaries between diseases and the background, the propensity of models to loss tiny pest characteristics, and the computational challenges are posed by large parameter.

Figure 2A depicts the architectural design of the CDM-DeeplabV3+ model, which employs a traditional encoding and decoding framework. The encoding incorporates the MobilenetV2 as its backbone and integrates the residual attention space pyramid structure (CBAM-ASPP). To further enhance the model, we integrate the Dual Attention Network (DANet) within the architecture. Detailed explanations of each component follow in the subsequent sections.

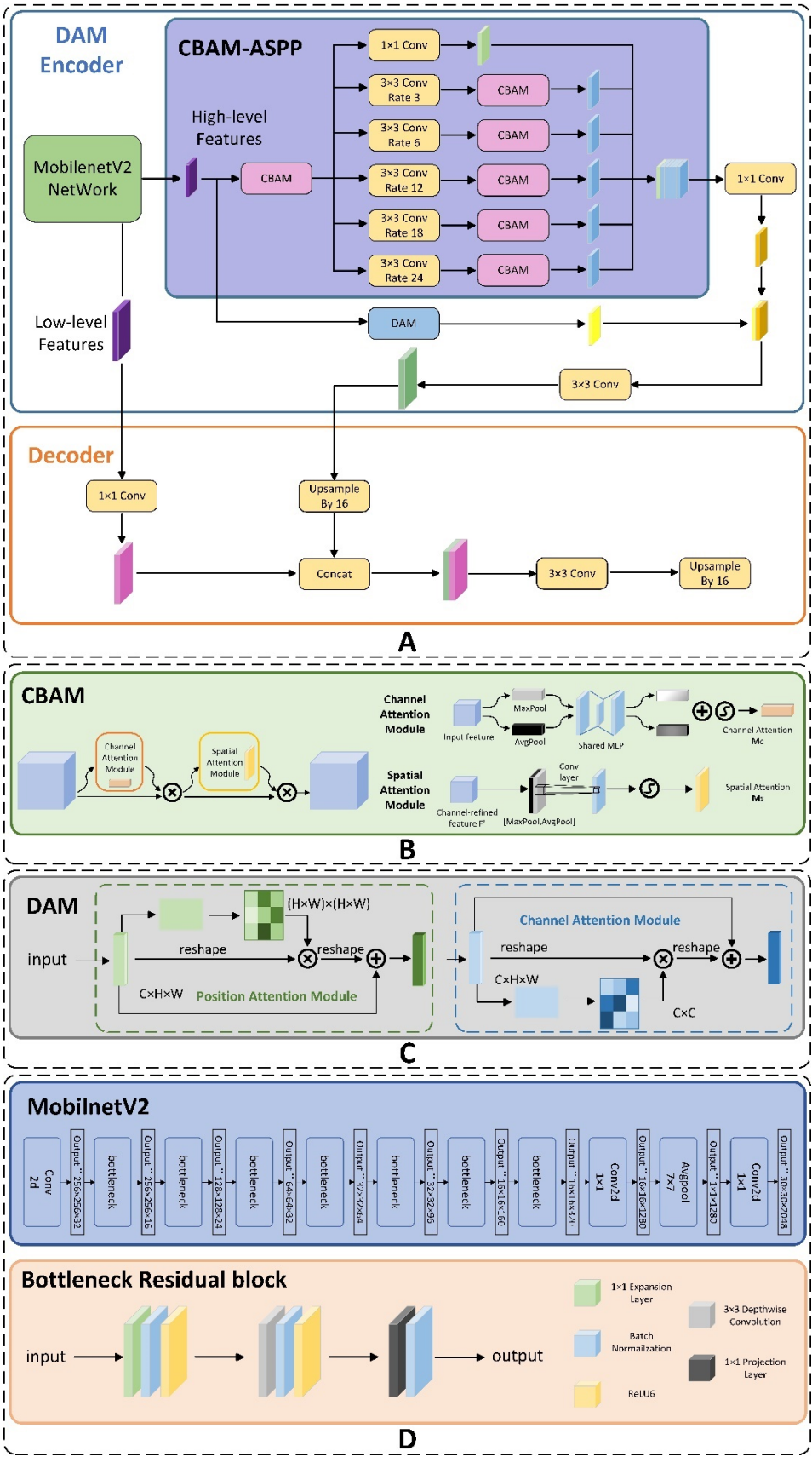


Figure 2. (A) CDM-DeeplabV3+ overall framework diagram. (B) CBAM structure schematic. (C) DANet structure schematic. (D) MobilenetV2 structure schematic.

2.2. CBAM-ASPP

Atrous Spatial Pyramid Pooling (ASPP), is an advanced approach that builds upon the foundations of Spatial Pyramid Pooling. Unlike SPP, ASPP is particularly effective at extracting semantic information at varying scales, thereby enhancing the ability of model to understand and represent features within image. ASPP substitutes conventional convolutions with dilated convolutions, which serve to enlarge the receptive field and bolster the capacity to extract semantic information [25]. ASPP employs multiple dilated convolution layers in parallel, each with a unique dilation rate. This design allows the network to simultaneously perceive features at various scales, thereby improving its ability to process information across multiple scales. The ASPP module integrates a 1x1 convolution, 3x3 dilated convolutional layers with dilation rates of 18, 12, and 6, accompanied by an average pooling layer, all designed for simultaneous feature refinement through multi-scale receptive fields [26]. The outcomes of these individual processes are then combined to produce a multi-scale fused feature map.

The CBAM is a straightforward and potent attention mechanism designed specifically for using in CNN [27]. It comprises two distinct attention components: channel attention and spatial attention. Figure 2B presents a schematic representation of the CBAM. This module utilizes its two attention dimensions to adaptively refine features, thereby enhancing the performance of model on feature extraction. Firstly, the feature graph F is inputted by the channel attention module, which generates a new feature map. This output is then combined with the original F through multiplication, forming an updated map labeled $F1$, as described in Equation (1). Subsequently, $F1$ passes through the spatial attention module to derive a further feature map. Ultimately, the computed result is multiplied by $F1$, giving rise to the final feature map $F2$, in accordance with the mathematical formulation outlined in Equation (2).

$$F1 = Mc(F) \otimes F \quad (1)$$

$$F2 = Ms(F1) \otimes F1 \quad (2)$$

The module for channel attention computes a channel attention map by assessing the interdependencies among channels. This process entails spatial operations through max and average pooling on the input maps, followed by utilization of a Multi-Layer Perceptron (MLP) to compute attention weights. Then, these weights are utilized to the input features by multiplication, thereby generating the channel attention feature maps, as per the mathematical formulation detailed in Equation (3).

$$Mc(F) = \delta \left(MLP(Avgpool(F)) + MLP(Maxpool(F)) \right) \quad (3)$$

The spatial attention mechanism exploits the correlations between various spatial positions to compute a 2D spatial attention map. The process begins by utilizing maximum and average pooling to input features across channel dimensions. The outputs of these pooling operations are then fused to form a comprehensive feature descriptor. The descriptor is fed into a convolutional layer, generating a spatial attention map. This attention map is multiplied with the original input feature map to get the final output, as described in mathematical formulation in Equation (4).

$$Ms(F) = \delta(f^{3 \times 3}([Avgpool(F); Maxpool(F)])) \quad (4)$$

After calculating both channel and spatial attentions, resulting maps are combined multiplicatively to generate the final adjusted feature map.

The classical ASPP structure had been the subject of extensive research. Qiu et al. [28] introduced a spatial pyramid pooling network, and designed to emphasize significant features for effective target detection. In disease and pest identification tasks, the ASPP method utilized diverse dilation rates to gathering information across multiple scales. Nonetheless, the feature extraction might be adversely affected by the gridding of dilated convolutions. Furthermore, the presence of parallel branches in the ASPP can led to the redundant extraction of similar features, which may complicated the process

of recognizing diseases [29]. This study introduced a CBAM-ASPP structure to address the gridding issue associated with dilated convolutions. The CBAM-ASPP structure effectively solved the difficulties in segmentation caused by differences in size, occlusions, overlaps, and alterations in illumination conditions for *C. oleifera* diseases and insect pests. Consequently, it improved segmentation results when dealing with intricate and varied tasks.

The ASPP module feeds these up-sampled features directly into dilated convolution layers. But the dilated convolution struggles to accurately capture the edge information of diseases and insect pests on *C. oleifera*. In CBAM-ASPP, the up-sampled original features are first processed by the CBAM. This module is designed to focus on both the spatial and channel aspects of the significant features, thereby improving the capability of model to accurately extract edges. In order to preserving more detailed edge and texture features related to *C. oleifera* diseases and to minimizing the possibility of overfitting the model, we remove the average pooling layer from the architecture [30]. To extract multiscale information from the diseases in *C. oleifera* and to facilitate their fusion, we extend the dilated convolution to incorporate five dilation rates: 3, 6, 12, 18, and 24, all while maintaining a convolution kernel size of 3x3. Different expansion rates can simulate different scaled receptive fields for capturing features at different scales. Dilated convolution, although adopt at capturing multi-scale features, may suffer from a limitation known as gridding artifacts. This occurs because the convolution kernel, when dilated, skips over certain points in the feature map, resulting in a sparse sampling pattern that can compromise the integrity of the learned representation. Moreover, employing excessively high dilation rates can lead to irregular feature extraction. To address this, we have integrated a 3x3 dilated convolution with a CBAM to create an Attention Convolution Module, and we have combined a 1x1 convolution with a CBAM to construct a Residual Module. By integrating convolution with the CBAM, we are able to effectively filter out extraneous features, thus facilitating the acquisition of consistent and significant features. The stability of network and robustness are improved by adding five attentional convolution modules and one residual module separately. This approach also helps maintain the integrity of spatial information within the network [31]. Ultimately, the five feature maps are combined using a cascading process, after which the output is fed into a 1x1 convolution module to extract the last high-level features.

2.3. DAM-Encoder

Given that some *C. oleifera* diseases present with small symptoms and are influenced by intricate environmental factors, it is essential to improve the detection of these minute features. To address this, the Dual Attention Network (DANet) is integrated into the Encoder of DeeplabV3+. DANet facilitates the dynamic fusion of local details with their corresponding global context, thereby enhancing the ability of network to discern these tiny features.

DANet is composed of two distinct architectural components: the Location Attention Module and the Channel Attention Module, each designed to emphasize spatial and channel relationships, respectively. These modules are engineered to capture long-range dependencies across both spatial locations and feature channels, respectively. Figure 2C presents a visual representation of the design architecture of the DANet.

The Positional Attention Module begins by accepting the input local features and processing them through a convolutional layer to reduce their dimensionality. This step results in the creation of three disparate feature maps, labeled as B, C, and D. Subsequently, the feature maps are transformed into an M by N matrix format, where N is calculated as the product of the original feature height (H) and width (W) of maps. Upon reshaping the feature matrix, it is subjected to matrix multiplication, followed by a softmax operation, which yields the spatial attention map, labeled as S, as per the mathematical formulation in Equation (5). This attention map is then applied to the original features and scaled by a factor of α before being subjected to an element-wise summation to produce the ultimate output, as shown in Equation (6).

$$S_{ji} = \frac{\exp(B_i \cdot C_j)}{\sum_{i=1}^N \exp(B_i \cdot C_j)} \quad (5)$$

$$E_j = \alpha \sum_{i=1}^N (S_{ji} D_i) + A_j \quad (6)$$

In the mathematical expressions, S_{ji} signifies the impact of the i -th position on the j -th position. B_i refers to the i -th feature at map B. C_j represents the j -th feature in map C. E_j denotes the j -th feature in map E. D_i corresponds to the i -th feature in map D, while A_j corresponds to the j -th position of the remote feature mapping.

The Channel Attention Module calculates map directly using the original features. Analogous to the spatial attention process, the local features are reconfigured into a shape of $M \times N$, as shown in Equation (7). The key distinction here is that the computation of the attention map involves direct operations on the original map without need to produce separate feature maps B, C, and D, as shown in Equation (8).

$$X_{ji} = \frac{\exp(A_i \cdot A_j)}{\sum_{i=1}^C \exp(A_i \cdot A_j)} \quad (7)$$

$$E_j = \beta \sum_{i=1}^N (X_{ji} A_i) + A_j \quad (9)$$

DANet optimizes the utilization of contextual information by transforming the feature outputs from its dual attention modules. Next, an element-wise addition is performed to merge the features, followed by the application of a convolutional layer to produce the final prediction map.

Integrating DANet into the Encoder enhances the clarity and effectiveness of presenting information, thereby mitigating misclassifications due to environmental influences. The proposed architecture greatly improves semantic coherence and facilitates superior segmentation of the minuscule features associated with diseases and pests.

Zhu et al. [32] discovered that the connection between DANet and the ASPP module yields different outcomes. Consequently, as depicted in Figure 3, this research proposes two distinct architectures for combining DANet with ASPP.

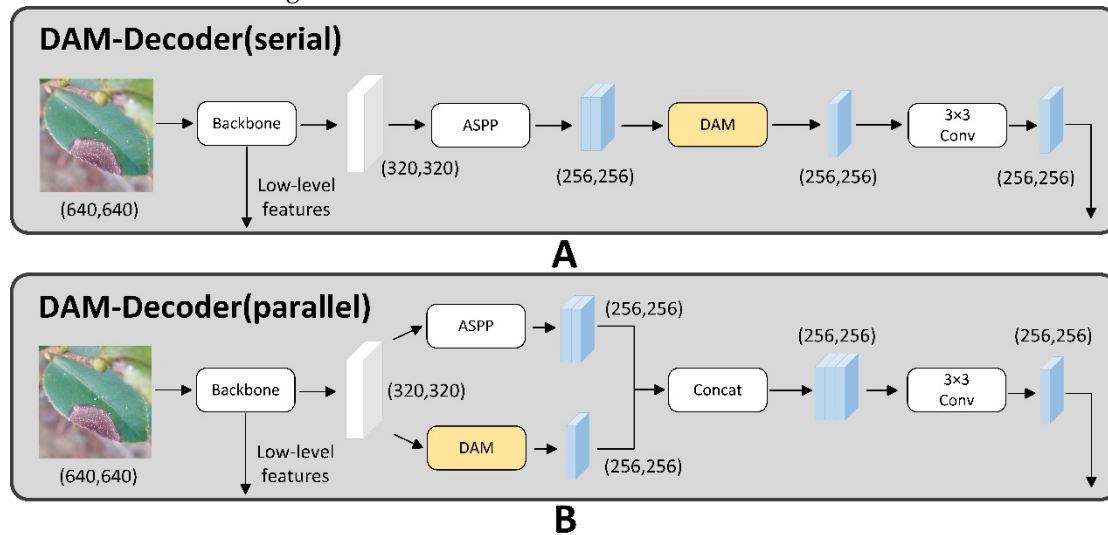


Figure 3. Schematic diagram of two diverse connections about ASPP and DAM.

In the first proposed structure, it is a sequential arrangement. The features extracted from the backbone are first inputted the ASPP module. Following this, the ASPP outputs are processed via the DANet module. Ultimately, the enhanced features derived from the DANet model are fused with the low-level features that extracted from the backbone.

In the second structure, DANet and ASPP are connected in parallel. This means that the features extracted from the backbone are inputted both modules concurrently. The resulting feature maps from each module are then subjected to further processing by a convolution module. Subsequently,

the processed features are fused with the low-level features originating from the backbone, enhancing the overall feature representation.

Experimental comparisons reveal that the second structural marginally surpass the first one in terms of overall performance. As a result, the second structure is chosen for our investigation.

2.5. MobilenetV2

The DeeplabV3+ model typically employs the Xception backbone, which is highly regarded for its effectiveness in semantic segmentation tasks. But segmentation of *C. oleifera* diseases and insect pests demand small models with few parameters, fast computation, and ease of deployment on mobile devices. The Xception model is distinguished by its arrangement of 36 convolutional layers, organized into 14 modular sections, each consisting of multiple convolutions. It also boasts a substantial number of parameters in its fully connected layer [33]. This complexity translates into a model that requires extensive training time, making it less ideal for use in segmenting diseases and insect pests in *C. oleifera*. Our approach adopts MobileNetV2 as the backbone. Unlike Xception, MobileNetV2 employs Depthwise Separable Convolutions (DSC) to construct inverted residual blocks and substitutes the traditional Rectified Linear Unit (ReLU) with a linear bottleneck layer. This design, characterized by its efficiency and reduced computational demands, is particularly suited for tasks focused on disease and insect pest segmentation within the context of *C. oleifera*. Figure 2D shows the schematic diagram of MobileNetV2.

Convolutional play a key role in many efficient neural network designs. They process the input data by scanning across the receptive field to selectively identify and extract the most significant features, Enhancing the network's capacity to discern crucial information within the regions of important. Conventional convolutional methods perform operations across both spatial and depth dimensions, which results in a substantial increase of parameters and computational load. Different from standard convolution, DSC employs a strategy where the full convolution splits into a pointwise convolution and a depthwise convolution. In a convolutional layer that includes K filters, each of dimension $F \times F$, operating on an input layer with D channels. Equation (9) presents the total parameter count.

$$K \times F \times F \times D \quad (9)$$

In the second step of DSC, we set the 1×1 convolution to $D \times K$. Under the same conditions, the total parameters as shown in Equation (10).

$$D \times F \times F + D \times K \quad (10)$$

The parameter count is decreased from $K \times F^2 \times D$ to $K \times F^2 + K \times D$ by employing DSC. This method reduces the computational cost by segregating spatial and depth-wise operations, allowing each input channel to be processed separately. Consequently, DSC exhibits a significant decrease in computational demands compared to conventional convolution, as supported by the research conducted by Dang et al. [34].

The residual block, introduced by He et al. [35], serves as the central component of the ResNet framework. Its main purpose is to mitigate the problems of gradient vanishing or explosion that often arise as the depth of a neural network grows. Gradient vanishing or exploding can cause weights to become extremely small or large, which may lead to instability in the loss function and weights of model. This instability can impede the model from effectively learning the data distribution, ultimately leading to poor performance. MobileNetV2 introduces inverted residual blocks to minimize the quantity of parameters without compromising network performance [36]. Figure 4 shows the schematic diagram of residual block and inverted residual block. These blocks first use a 1×1 convolution to increase the dimensionality of the input features. Then employs a depthwise separable convolution to capture spatial features within these enhanced dimensions. Subsequently, a second 1×1 convolution is utilized to reduce the feature dimensions, thereby completing the process of dimensionality reduction. The module also utilizes skip connections to connect inputs directly to

the outputs. The inverted residual block not only streamlines the transformation process but also conserves computational power and minimizes the number of parameters required.

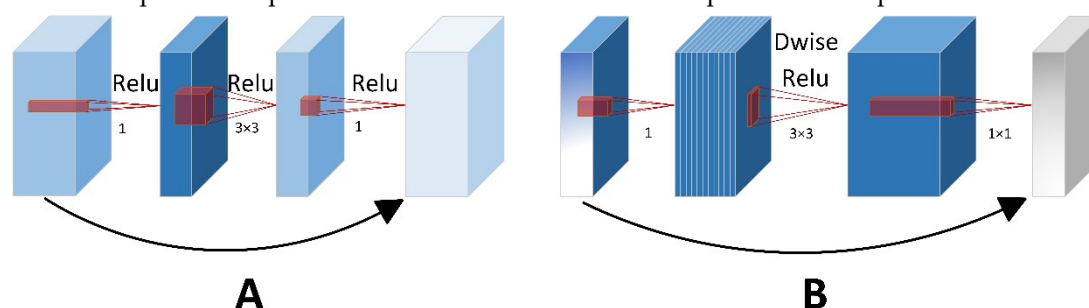


Figure 4. (A) schematic diagram of residual block structure. (B) schematic diagram of inverted residual structure.

ReLU activation function, a common choice in neural networks, serves to a certain extent alleviate the issue of overfitting by introducing non-linearity. MobileNetV2 introduces inverted residual blocks, which employs DSC to effectively extract spatial features in higher dimensions while producing output features of lower dimensionality. But the ReLU activation function tends to eliminate low-dimensional feature information. However, the linear bottleneck layer in neural networks is designed to capture low-dimensional spatial information without incurring significant information loss due to nonlinear transformations.

3. Results

The experimental framework for insect pest and disease segmentation on *C. oleifera* leaf using deep learning models is shown in Figure 5. Firstly, Images of *C. oleifera* with diseases were collected from different regions, including Jiading Town in Xinfeng County, Jiangxi Province; Yizhang County in Chenzhou City, Hunan Province; You County, Hunan Province; and Xingning City, Guangdong Province. Secondly, we performed image preprocessing on the collected photographs encompassing five categories of diseases and insect pests and annotated the augmented images. Then, the CDM-DeeplabV3+ model is trained using the images that have been processed and annotated for labels. Finally, the model is deployed to identify and segment the leaves symptom, generating the final segmentation outcomes to compare with the true result. Subsequently, the results are organized into four distinct segments to: (a) establish the experimental setup and parameter configurations. (b) parameter definition. (c) module performance assessment. (d) comparison of CDM-DeeplabV3+ model with other models.

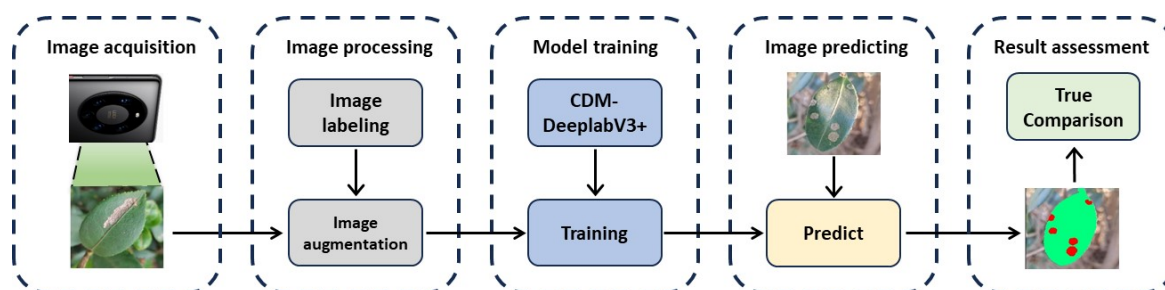


Figure 5. The process diagram of experimental simulation for insect pest and disease segmentation based on CDM-DeeplabV3+ network.

3.1. Image Acquisition and Dataset Preparation

Considering the absence of public dataset pertaining to the leaf diseases and insect pests of *C. oleifera*, we built own dataset. The images were shot in two separate periods: one from April to June, and the other from September to November of 2023. To ensure a more diversity and comprehensive

dataset, images were randomly collected across three different weather conditions: sunshine, overcast skies, and rainfall. The images were taken from various angles and under different light conditions using Huawei Mate40 camera. They were set to an aspect ratio of 1:1 and a resolution of 3072x3072 pixels. We collected a total of 3,172 high-quality images of *C. oleifera* diseases, which were rigorously vetted by experts and sorted into five categories: Anthracnose, Algae Leaf Spot Disease, Soft Rot Disease, Tea White Blister, and Worm Holes.

Data preprocessing aimed to highlight critical features, eased the learning process for models, reduced overfitting to particular datasets, and improved the ability of model to generalize new data. Given the high resolution of the *C. oleifera* images, traditional denoising techniques like Block-matching, Markov random field approaches, Contemporary wavelet and Three-dimensional (3D) filtering may inadvertently removed crucial details and texture information from the images. Consequently, we opted not to apply denoising procedures in our dataset. Our study addressed the challenges posed by the intricate backgrounds and inconsistent lighting of *C. oleifera* leaves by applying a range of image pre-processing techniques. These techniques encompassed mirroring, perspective correction, adjustments to contrast, and alterations to grayscale, all aimed at enhancing the complexity of the images for further analysis. We systematically applied one or more pre-processing methods to all leaves until the necessary criteria for training were satisfied. Following these pre-processing steps, the spatial distribution of leaf diseases was significantly improved. To expedite the training and reduce model size, the images were resized from 3072x3072 pixels to 640x640 pixels without compromising image quality. We utilized Labelme software to mark the locations of *C. oleifera* diseases and created corresponding mask maps. Figure 6 illustrated various disease labels, with each annotation of images saved in JSON format. A total of 1,265 photographs were selected following pre-processing to comprise the final dataset. Table 1 presented the number of each *C. oleifera* disease categories along with their respective percentages. From the total, 80% (1012 photographs) were randomly chosen for training purposes, while the remaining 20% (253 photographs) were set aside for validation purposes. The ratio of training to validation throughout the training will remain the same.

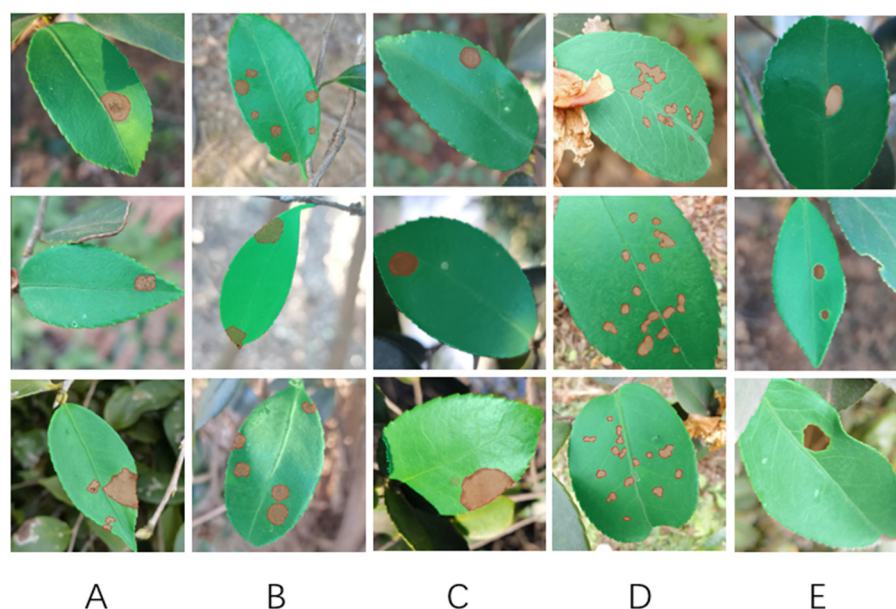


Figure 6. (A) Schematic of Anthracnose labeling. (B) Schematic of Algae leaf Spot Disease labeling. (C) Schematic of Soft Rot disease labeling. (D) Schematic of Tea white blister labeling. (E) Schematic of Worm Holes labeling.

Table 1. Number and proportion of C. oleifera diseases images.

Species	Anthracnose	Algae leaf Spot Disease	Soft Rot disease	Tea white blister	Worm Holes
typical example					
quantities	307	231	212	287	228
percentage	24.26%	18.26%	16.75%	22.38%	18.02%

3.2. Experimental Setup

To minimize hardware and network parameter influences on model training, all experiments were carried out under consistent hardware and software configurations. The specific hardware specifications used for training and prediction are presented in Table 2. Furthermore, to preserve consistency and to prevent any external factors from affecting the experimental outcomes, the parameters for each network are standardized across all tests. As shown in Table 2, this is the hyperparameter determined by us after a series of experiments.

Table 2. Parameters of environment.

Environment	Device	Parameter
Hardware environment	CPU	AMD Ryzen 7 5800H with Radeon Graphics
	GPU	NVIDIA GeForce RTX 3050 Ti Laptop GPU
	RAM	16GB
	Video memory	16GB
Software environment	OS	Ubuntu 20.04
	CUDA Toolkit	V11.1
	CUDNN	V8.0.4
	Python	3.8
	Torch	1.2.0
	torchvision	0.4

Table 3. Setting of experiment.

Hyperparameters	Parameters
Size of input images	640×640
Batch	8
Initial learning rete	0.0001
Optimizer	Adam
Momentum	0.9

3.3. Evaluation Indicators

The performance of model is assessed through a comprehensive set of five indicators: Mean Intersection over Union (mIoU), Recall, Precision, Parameter and Frames Per Second (FPS), providing a holistic understanding of its accuracy, efficiency, and resource consumption.

We first introduce the definitions of the four different types of regions: True Positive (TP): Represents the real diseased regions that have been identified as diseased. True Negative (TN): Represents the real healthy regions that have been identified as healthy. False Positives (FP): Represents the real healthy regions that have beenidentified as diseased. False Negative (FN): Represents the real diseased regions that have been identified as healthy. The definitions of these four regions in the segmentation results are shown in Table 4.

Table 4. Definitions of the four indicators of segmentation results.

Predicted	Regional	Definitions
Positive	Positive	TP
Positive	Negative	FN
Negative	Positive	FP
Negative	Negative	TN

The Intersection over Union (IoU) serves as an indicator to quantify the degree of overlap between the predicted segmented areas and actual segmented areas of *Camellia oleifera* diseases and insect pests. The mIoU, which are gained by averaging IoU across all diseases categories, is principal evaluation criterion in this study. It effectively assesses the performance of our model in semantic segmentation tasks, as shown in Equation (11).

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \quad (11)$$

Precision is defined as the proportion of correctly segmented *Camellia oleifera* leaf images relative to those that are incorrectly segmented within the dataset, as shown in Equation (12).

$$Precision(P) = \frac{TP}{TP + FP} \quad (12)$$

Recall, expressed as the ratio in Equation (13), refers to the proportion of actual *C. oleifera* leaf images that are correctly identified as needing segmentation out of all the instances that truly require segmentation.

$$Recall(R) = \frac{TP}{TP + FN} \quad (13)$$

FPS, or Frames Per Second, refers to the rate at which the model detects images. 'T' denotes the duration needed to process a single batch of images. During testing, we tested 100 times and determined the average FPS by calculating the mean of these 100 measurements to ensure accuracy, as shown in Equation (14).

$$FPS = \frac{1}{T} \quad (14)$$

3.4. Performance Evaluation of Each Module

The CBAM-ASPP module combines the CBAM with a dilated convolution, eliminating an average pooling layer. This structure facilitates the integration of multi-scale feature data, therefore significantly improving edge detection clarity. The DAM- Encoder integrates a DANet and CBAM-ASPP in parallel. This innovation allows the model to distinguish more small details of pests and diseases. Substituting the traditional backbone with MobileNetV2, which take the advantage of DSC, inverted residual blocks, and linear bottleneck layers, markedly decreases the quantity of parameters and accelerates computational efficiency.

3.4.1. Effectiveness of CBAM-ASPP

To overcome the issue of uneven sampling caused by dilated convolution, which can hinder the effective processing of edge information, we have incorporated dilated convolution with an attention module to create an attentional convolution module. The module addresses the challenge by selectively emphasizing important areas of the input. This approach not only enlarges the receptive field but also strengthens the extraction of edge features. Therefore, we perform experimental comparisons among five distinct attentional mechanisms: Convolutional Block Attention Module (CBAM), Dual Attention Network (DANet) [37], Efficient Channel Attention (ECA) [38], Shuffle Attention (SA) [39], and Spatial Group-wise Enhance (SGE) [40]. These mechanisms are evaluated for

their effectiveness when are incorporated into the ASPP module. The experimental results are presented systematically in Table 5 for clear interpretation and analysis.

Table 5. Comparison of Attention Mechanisms.

Method	mIoU(%)	Precision(%)
Without Attention	79.81	88.51
CBAM Attention	82.2	91.63
DANet Attention	81.05	92.05
ECA Attention	81.19	92.01
Shuffle Attention	79.63	89.95
SGE Attention	80.99	90.81

The results indicate that incorporating attention mechanisms greatly improves the mIoU of the model. DANet stands out due to its exceptional increase in precision, primarily credited to its capability in effectively capturing long-range dependencies across spatial and channel dimensions. This enhancement facilitates to overcome the limitations of dilated convolution. CBAM achieves the highest mIoU and demonstrates a substantial improvement in Precision. For maximizing segmentation accuracy, we opt to integrate CBAM into the ASPP module.

3.4.2. Effectiveness of DAM- Encoder

In the previous research, DANet has been incorporated into the DeeplabV3+ framework to improve the segmentation of small disease by integrating local features with their corresponding global features. In our study, we examine two distinct methods for integrating DANet to the DeeplabV3+. The one is a parallel arrangement of DANet and ASPP, the other is a serial configuration of DANet and ASPP. The corresponding findings are illustrated in Table 6. In the serial configuration, DANet processes features that have already been processed by ASPP. It might result in the loss of crucial details related to small-scale features. Conversely, when ASPP and DANet are connected in parallel, DANet processes high-level features which directly generated by backbone, allowing for the simultaneous extraction of disease features at different scales with ASPP. This parallel setup overlaps the outputs of DANet and ASPP, thereby enhancing the overall segmentation performance. Table 6 indicates that the parallel DAM- Encoder structure yields an mIoU of 81.96%, which is an improvement over the previous mIoU of 79.81%. Furthermore, the parallel DAM- Encoder experiences only a marginal reduction of 0.55% in Precision relative to the serial structure. Given these results, we decide to adopt the DAM- Encoder (parallel) configuration for the task of identifying and detecting minuscule pests and diseases on Camellia oleifera leaves.

Table 6. Comparison of DAM- Encoder structures in different connections.

Method	mIoU	Precision	Parameter	Recall
Without DAM	79.81	88.51	209.71	88.36
DAM- Encoder (serial)	79.2	92.66	256.82	84.46
DAM- Encoder (parallel)	81.96	92.01	283.27	87.99

3.4.3. Effectiveness of MobilenetV2

To optimize the efficiency of model by decreasing parameter count and computational load without compromising segmentation performance, we substitute the Xception backbone with the MobileNetV2. The results of this substitution are presented in Table 7. The modified parameters are diminished by roughly 8 times compared to the DeeplabV3+, and there is also a nearly 5-fold decrease compared to CDM-DeeplabV3+. Moreover, the adoption of MobilenetV2 as the backbone not only reduces the parameters and computational requirements but also maintains the integrity of vital features, leading to a modest increase in mIoU. According to the FPS measurements, the detection speed has improved. the system can process approximately twice as many images per second, from

around 10 to 20. The integration of the innovative CBAM-ASPP and DAM-Encoder designs does not result in a observably rise in the quantity of parameters, thus preserving the detection speed. The results clearly show that MobilenetV2 can markedly decrease the quantity of network parameters and simultaneously boost processing efficiency without sacrificing the performance of network. Hence, MobilenetV2 is chosen as the backbone for this study.

Table 7. Comparison of performance indicators base on different backbone.

Method	mIoU	Precision	Parameter	FPS
DeeplabV3+ (Xception)	79.81	88.51	209.71	10.49
DeeplabV3+ (MobilenetV2)	81.13	91.32	22.44	23.54
CDM-DeeplabV3+ (Xception)	83.12	91.29	295.48	10.41
CDM-DeeplabV3+ (MobilenetV2)	85.63	93.95	90.53	20.85

3.4.4. Ablation Experiments

To evaluate the efficacy of our CDM-DeeplabV3+ model designed specifically for disease segmentation in *C. oleifera* leaves, we conducted a series of ablation studies. Each structure, including CBAM-ASPP, DAM-Encoder, and MobilenetV2, are added incrementally for comparative analysis. Our analysis primarily focuses on the change in mIoU as a metric to evaluate the performance of each module and its promotion to the model. Having already discussed above the improvements in network performance achieved by adding single modules, we will delve into the experimental outcomes to demonstrate the impact of integrating multiple modules together. These experimental data are presented in Figure 7. The fused application of CBAM-ASPP and DAM-Encoder outperforming the single structure by 2.81% and 3.05% in mIoU, respectively. The integration of these two modules significantly enhances feature extraction, resolving the challenges of ambiguous boundaries and the detection of subtle characteristics. Consequently, this approach markedly bolsters the capabilities of network by various dimensions. Although integrating MobilenetV2 with the aforementioned modules does not lead to a substantial increase in mIoU, it notably diminishes the number of parameters by nearly 5 times. This reduction substantially enhances the efficiency of model, laying a solid foundation for real-time detection applications. The above conclusion of analysis are proved that introduction of our proposed modules have led to a considerable enhancement in the segmentation accuracy of diseased and insect pest *C. oleifera* leaves.

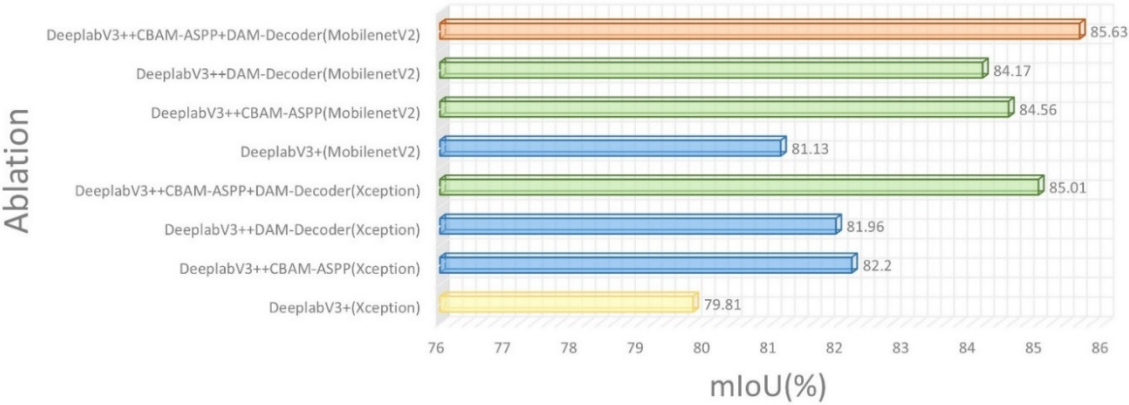


Figure 7. Ablation experiments with CDM-DeeplabV3+.

3.4.5. Edge and Tiny Problem Solving and Demonstration

In the introduction of this paper, we outline the challenges with the original DeeplabV3+ model, particularly highlighting the issues of edge blurring and the loss of small diseases. Figure 8 shows three distinct image types: the original image, its segmented image, and the composite image. In this study, as shown in Figure 8, we carry out a comparative examination of the segmentation outcomes

between the conventional DeeplabV3+ model and our innovative CDM-DeeplabV3+ design. Figure 8(a) reveals that the original DeeplabV3+ struggles to accurately outline the diseased areas. Conversely, CDM-DeeplabV3+ (as shown in Figure 8(A)) is proved has superior performance by delineating the jagged edges of diseases on *C. oleifera* leaves with greater precision. Additionally, Figure 8(c) and Figure 8(C) further highlight its effectiveness of edge segmentation by our proposed model. Regarding the segmentation of tiny diseases features, it is evident from a comparison of Figures 8(b) and 8(B) that our proposed network exceeds in detecting and segmenting these fine elements. The tea white blister, a common diseases of *C. oleifera* leaves, presents on leaves just like as numerous small, white, dot-like lesions. Figure 8(C) clearly shows that our model successfully segments all the area of tea white blister, effectively overcoming the difficulty of segmenting such tiny diseases. Figure 8(d) and Figure 8(D) depict situations where a *C. oleifera* leaf is simultaneously impacted by various diseases and insect pests. The CDM-DeeplabV3+ model still efficiently and accurately segments these complex overlapping issues.

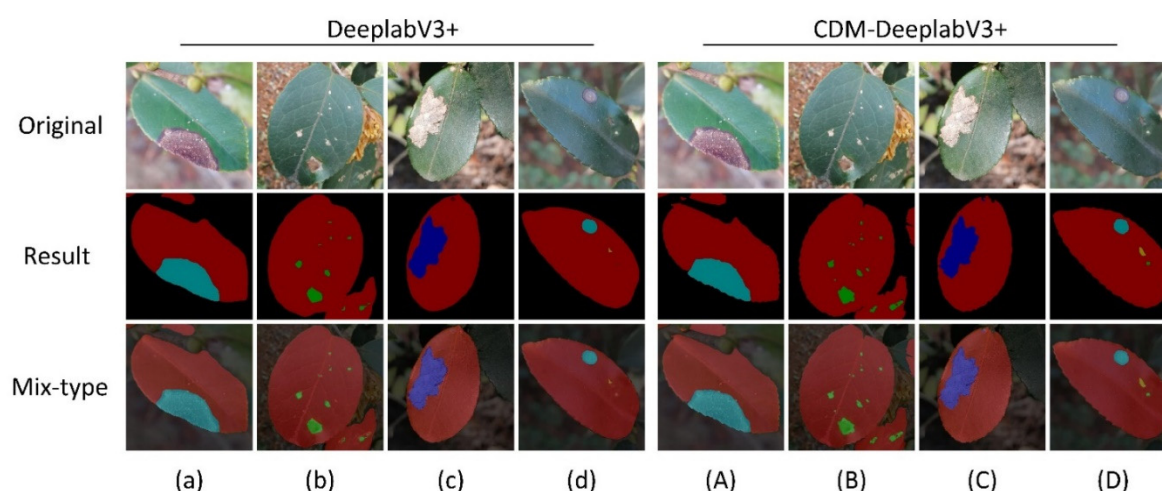


Figure 8. This figure shows that comparison of DeeplabV3+ and CDM-DeeplabV3+ about fuzz boundary and tiny disease segmentation.

3.5. Comparison of Other Models with the CDM-DeeplabV3+ Model

For this experiment, the performance of model is evaluated using the previously established metrics for the CDM-DeeplabV3+ architecture. Its performance was then compared to that of several advanced models, including UNet [41], UNet++ [42], PSPNet [43], and HrNetV2 [44]. Table 8 illustrates the results of each performance of model. PSPNet does not show significant advantages over the other models. Although it also uses a pyramid pooling structure, its receptive field is less extensive than the dilated convolutions employed in our approach, resulting in a lower mIoU. Moreover, the other evaluation metrics do not indicate outstanding performance. HrNetV2 (85.32% in mIoU, 85.58% in Precision) and UNet (78.70% in mIoU, 84.38% in Precision) exhibit better performances, compared with PSPNet (76.55% in mIoU, 80.37% in Precision). This indicates that HrNetV2 and UNet are equally effective in segmenting diseases and insect pests. We attribute the superior performance of HrNetV2 to continuous information exchange via iterative multiscale fusion. This process entails the integration of information from parallel representations at different resolutions, which enriches the data fusion and consequently improves segmentation accuracy. While HrNetV2 demonstrates strong segmentation capabilities, it comes with a parameter count that is roughly double that of our proposed network. In the context of agricultural disease and pest detection, the small model parameters are key. Therefore, our model can be applied in actual agriculture.

Table 8. A comparative evaluation of the core performance metrics among various models.

Model	mIoU	Precision	Parameter	Recall	FPS
DeeplabV3+	79.81	88.51	209.71	88.36	10.49
UNet-Vgg	78.70	88.33	94.9	86.82	6.73
UNet-Resnet	84.38	92.63	167	90.07	10.71
PSPNet-MobilenetV2	76.55	88.58	99.3	83.72	41.91
PSPNet-Resnet50	80.37	91.63	178.53	86.19	15.98
HrNetV2-W18	85.31	93.42	37.54	90.69	11.36
HrNetV2-W32	85.58	93.66	113.57	90.87	10.84
CDM-DeeplabV3+	85.63	93.95	90.53	90.53	20.85

In order to comparing the segmentation outcomes with different models, we put typical images into each models. This approach offers a visual means of understanding the disparities in segmentation performance among the models. Figure 5 presents the original images and the segmented results.

As depicted in Figure 9A, when leaf edges are indistinct and specific diseases closely resembled the background, some models, such as DeeplabV3+ and PSPNet, show overfitting. Even leaves were clearly distinguishable from the background, as shown in Figure 9C, the models also had difficulty precisely segmenting the edges of the diseased areas, resulting in inaccurate segmentation masks. The segmentation outcomes can effectively be discriminated between *Camellia oleifera* diseases and the background by our proposed models. In the case of severe disease, the leaves were eroded so that they blend with the dark background. But our model still could accurately segment serrated edges of the diseased leaves. Within the diseased leaves, it also could be depicted a jagged edge which more fit to diseases area, instead of an overfitting curve. The CBAM-ASPP module is effective in capturing multi-scale features. It retains vital edge information by prioritizing features located at the edges of diseases. The segmentation results of the CDM-DeeplabV3+ are corroborated more accurately, comparing with other networks.

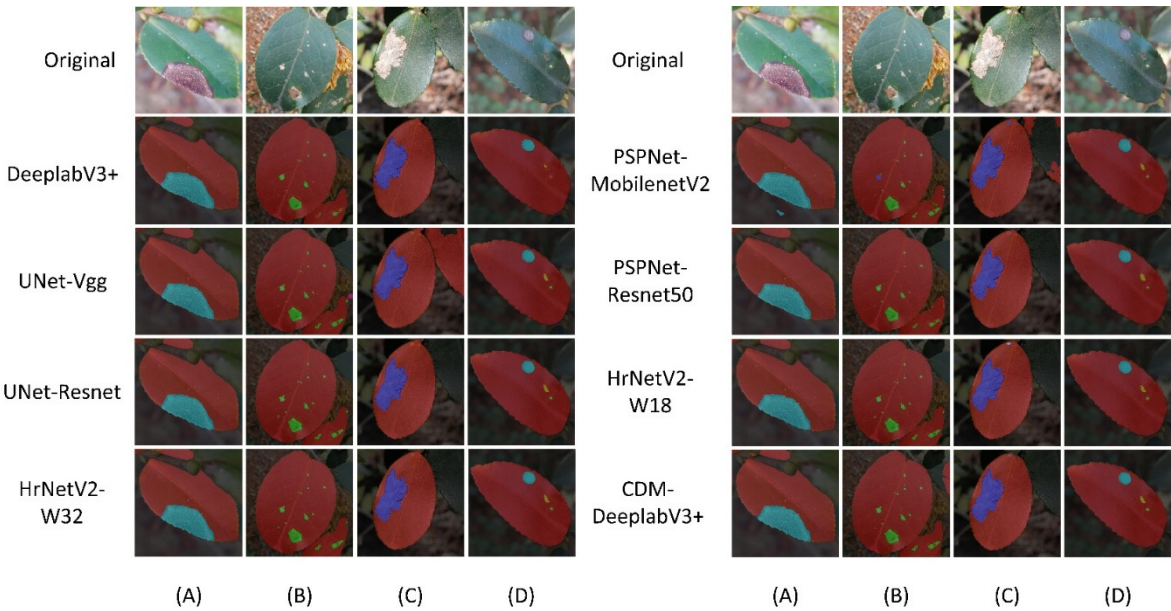


Figure 9. Comparison of segmentation effect of different models.

Tea white blister is primarily characterized by small white spots, as shown in Figure 9B. It is a challenge for original model, leading to the potential loss of information. CDM-DeeplabV3+ surpasses networks such as PSPNet in efficiently segmenting minuscule features. Additionally, it maintains robust recognition of Tea white blister even under varying illumination conditions. The capability of DAM-Encoder to integrate semantic information from both dimensions ensure robust

extraction of small details and improve semantic coherence. The improved performance contributes to our proposed architecture.

The loss function curves for various models are illustrated in Figure 10. Initially, the loss values among the models vary significantly. As increasing training epochs, these values drop quickly and tend to stabilize across most subsequent epochs. Notably, Figure 10 reveals that the proposed CDM-DeeplabV3+ model maintains consistently low loss, suggesting the robust generalization performance of network.

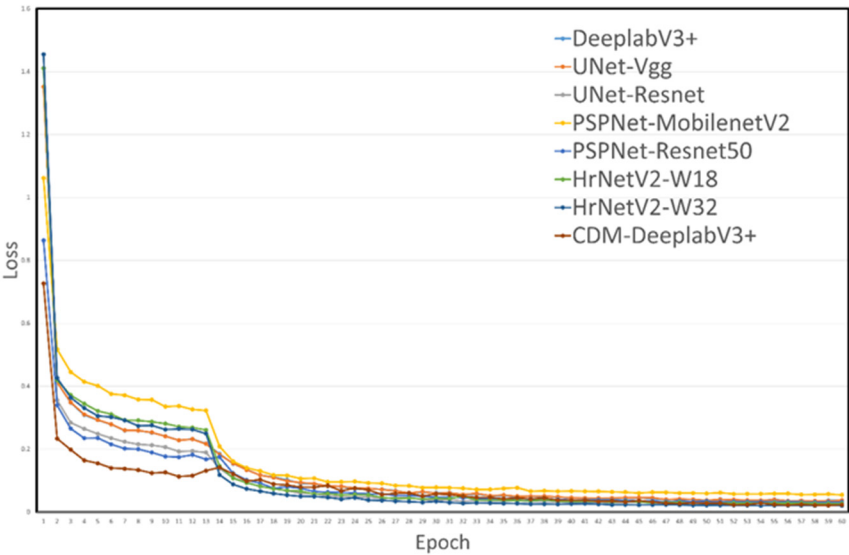


Figure 10. Comparison of loss function base on CDM-DeeplabV3+ and other models.

Authors should discuss the results and how they can be interpreted from the perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

5. Conclusions

To deepen students’ understanding of the process of developing and applying deep learning networks, we proposed the CDM-DeeplabV3+ network and designed a set of experimental simulation system for segmentation of insect pests and diseases on *C. oleifera* leaf based on this network. We enhanced edge extraction by replacing the ASPP module with the CBAM-ASPP module. Additionally, we adopted a parallel connection of the CBAM-ASPP and DANet modules to form the DAM-Decoder. This integration enhanced semantic consistency, resulting in clearer delineation of tiny features. We also substituted the Xception backbone with MobilenetV2 which significantly decreased the parameters of model. All experimental data were obtained by our self-build datasets, which were collected in real environment. Our model achieved an mIoU of 85.63% and a Precision of 93.95%. Meanwhile, the parameters of model had a factor of 5 reduction compared to original DeeplabV3+ model.

This experimental design is based on the results accumulated on scientific research experiments, the experimental content involves a number of professional courses such as image processing, pattern recognition, computer vision, etc., and the experimental process covers the whole process of the actual deep learning vision task.

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

Funding: This work was supported by the National Natural Science Foundation in China (Grant No.61902436) and Hunan Provincial Research Project on Teaching Reform in Colleges and Universities (Grant No. HNJG-20230458), Hunan Provincial Research Project on Teaching Reform for Degree and Graduate Students (Grant No. 2023JGYB57)

Data Availability Statement: Requests to access the datasets should be sent via email to guo1750524639@163.com.

Acknowledgments: We express our gratitude to the editor and reviewers for providing valuable suggestions aimed at enhancing the caliber of this paper. We sincerely acknowledge Bin Xie for their insightful guidance and unwavering support, which were instrumental in shaping the direction of this project. We also extend our thanks to Ruifeng Liu for their expertise in validation, which contributed significantly to the robustness and accuracy of our results. Additionally, we are grateful to Ziyang Shi for their meticulous efforts in project administration and data curation, ensuring the efficient management and execution of the research. We would also like to express our gratitude to Guoxiong Zhou for their invaluable assistance in validation, investigation, and securing funding.

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

References

1. Robards, K.; Prenzler, P.; Ryan, D.; Zhong, H. Camellia Oil and Tea Oil. In *Gourmet and health-promoting specialty oils*; Elsevier, 2009; pp. 313–343.
2. Chaydarreh, K.C.; Lin, X.; Guan, L.; Yun, H.; Gu, J.; Hu, C. Utilization of Tea Oil Camellia (*Camellia Oleifera* Abel.) Shells as Alternative Raw Materials for Manufacturing Particleboard. *Industrial Crops and Products* **2021**, *161*, 113221, doi:10.1016/j.indcrop.2020.113221.
3. Conrad, A.O.; Li, W.; Lee, D.-Y.; Wang, G.-L.; Rodriguez-Saona, L.; Bonello, P. Machine Learning-Based Presymptomatic Detection of Rice Sheath Blight Using Spectral Profiles. *Plant Phenomics* **2020**, doi:10.34133/2020/8954085.
4. Tian, H.; Wang, T.; Liu, Y.; Qiao, X.; Li, Y. Computer Vision Technology in Agricultural Automation—A Review. *Information Processing in Agriculture* **2020**, *7*, 1–19, doi:10.1016/j.inpa.2019.09.006.
5. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674, doi:10.3390/s18082674.
6. Likas, A.; Vlassis, N.; Verbeek, J.J. The Global K-Means Clustering Algorithm. *Pattern recognition* **2003**, *36*, 451–461, doi:10.1016/S0031-3203(02)00060-2.
7. Li, S.Z. Markov Random Field Models in Computer Vision. In *Proceedings of the Computer Vision—ECCV'94: Third European Conference on Computer Vision Stockholm, Sweden, May 2–6 1994 Proceedings, Volume II 3*; Springer, 1994; pp. 361–370.
8. Wang, G.; Sun, Y.; Wang, J. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Computational intelligence and neuroscience* **2017**, *2017*, doi:10.1155/2017/2917536.
9. Singh, A.K.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A. Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends in plant science* **2018**, *23*, 883–898, doi:10.1016/j.tplants.2018.07.004.
10. Lu, J.; Tan, L.; Jiang, H. Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification. *Agriculture* **2021**, *11*, 707, doi:10.3390/agriculture11080707.
11. Ramcharan, A.; Baranowski, K.; Ahmed, B. Deep Learning for Image-Based Cassava Disease Detection. *Frontiers in plant science* **2017**, *8*, 293051, doi:10.3389/fpls.2017.01852.
12. Jiang, P.; Chen, Y.; Liu, B.; He, D.; Liang, C. Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 59069–59080, doi:10.1109/ACCESS.2019.2914929.
13. Wang, C.; Du, P.; Wu, H.; Li, J.; Zhao, C.; Zhu, H. A Cucumber Leaf Disease Severity Classification Method Based on the Fusion of DeepLabV3+ and U-Net. *Computers and Electronics in Agriculture* **2021**, *189*, 106373, doi:10.1016/j.compag.2021.106373.
14. Deng, Y.; Xi, H.; Zhou, G.; Chen, A.; Wang, Y.; Li, L.; Hu, Y. An Effective Image-Based Tomato Leaf Disease Segmentation Method Using MC-UNet. *Plant Phenomics* **2023**, *5*, 0049, doi:10.34133/plantphenomics.0049.
15. 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 Proceedings of the European conference on computer vision (ECCV)*; 2018; pp. 801–818.
16. Khadidos, A.; Sanchez, V.; Li, C.-T. Weighted Level Set Evolution Based on Local Edge Features for Medical Image Segmentation. *IEEE Transactions on Image Processing* **2017**, *26*, 1979–1991, doi:10.1109/TIP.2017.2666042.
17. Xia, L.; Zhang, H.; Wu, Y.; Song, R.; Ma, Y.; Mou, L.; Liu, J.; Xie, Y.; Ma, M.; Zhao, Y. 3D Vessel-like Structure Segmentation in Medical Images by an Edge-Reinforced Network. *Medical Image Analysis* **2022**, *82*, 102581, doi:10.1016/j.media.2022.102581.

18. Gao, X.; Tang, Z.; Deng, Y.; Hu, S.; Zhao, H.; Zhou, G. HSSNet: A End-to-End Network for Detecting Tiny Targets of Apple Leaf Diseases in Complex Backgrounds. *Plants* **2023**, *12*, 2806, doi:10.3390/plants12152806.
19. Liang, Y.; Li, M.; Jiang, C.; Liu, G. CEModule: A Computation Efficient Module for Lightweight Convolutional Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* **2021**, *34*, 6069–6080, doi:10.1109/TNNLS.2021.3133127.
20. Chen, C.; Wang, C.; Liu, B.; He, C.; Cong, L.; Wan, S. Edge Intelligence Empowered Vehicle Detection and Image Segmentation for Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems* **2023**.
21. Ahmed, I.; Ahmad, M.; Khan, F.A.; Asif, M. Comparison of Deep-Learning-Based Segmentation Models: Using Top View Person Images. *IEEE Access* **2020**, *8*, 136361–136373, doi:10.1109/ACCESS.2020.3011406.
22. Zhang, D.; Ding, Y.; Chen, P.; Zhang, X.; Pan, Z.; Liang, D. Automatic Extraction of Wheat Lodging Area Based on Transfer Learning Method and Deeplabv3+ Network. *Computers and electronics in agriculture* **2020**, *179*, 105845, doi:10.1016/j.compag.2020.105845.
23. Song, Y.; Hu, J.; Wu, Q.; Xu, F.; Nie, S.; Zhao, Y.; Bai, S.; Yi, Z. Automatic Delineation of the Clinical Target Volume and Organs at Risk by Deep Learning for Rectal Cancer Postoperative Radiotherapy. *Radiotherapy and Oncology* **2020**, *145*, 186–192, doi:10.1016/j.radonc.2020.01.020.
24. 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 transactions on pattern analysis and machine intelligence* **2017**, *40*, 834–848, doi:10.1109/TPAMI.2017.2699184.
25. Song, X.; Fang, X.; Meng, X.; Fang, X.; Lv, M.; Zhuo, Y. Real-Time Semantic Segmentation Network with an Enhanced Backbone Based on Atrous Spatial Pyramid Pooling Module. *Engineering Applications of Artificial Intelligence* **2024**, *133*, 107988, doi:10.1016/j.engappai.2024.107988.
26. Lian, X.; Pang, Y.; Han, J.; Pan, J. Cascaded Hierarchical Atrous Spatial Pyramid Pooling Module for Semantic Segmentation. *Pattern Recognition* **2021**, *110*, 107622, doi:10.1016/j.patcog.2020.107622.
27. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional Block Attention Module. In Proceedings of the Proceedings of the European conference on computer vision (ECCV); 2018; pp. 3–19.
28. Qiu, Y.; Liu, Y.; Chen, Y.; Zhang, J.; Zhu, J.; Xu, J. A2sppnet: Attentive Atrous Spatial Pyramid Pooling Network for Salient Object Detection. *IEEE Transactions on Multimedia* **2022**.
29. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding Convolution for Semantic Segmentation. In Proceedings of the 2018 IEEE winter conference on applications of computer vision (WACV); Ieee, 2018; pp. 1451–1460.
30. Wang, Z.; Ji, S. Smoothed Dilated Convolutions for Improved Dense Prediction. In Proceedings of the Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2018; pp. 2486–2495.
31. Yu, D.; Wang, H.; Chen, P.; Wei, Z. Mixed Pooling for Convolutional Neural Networks. In Proceedings of the Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24–26, 2014, Proceedings 9; Springer, 2014; pp. 364–375.
32. Zhu, L.; Deng, Z.; Hu, X.; Fu, C.-W.; Xu, X.; Qin, J.; Heng, P.-A. Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection. In Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV); 2018; pp. 121–136.
33. Zhu, R.; Guo, D.; Wong, M.S.; Qian, Z.; Chen, M.; Yang, B.; Chen, B.; Zhang, H.; You, L.; Heo, J.; et al. Deep Solar PV Refiner: A Detail-Oriented Deep Learning Network for Refined Segmentation of Photovoltaic Areas from Satellite Imagery. *International Journal of Applied Earth Observation and Geoinformation* **2023**, *116*, 103134, doi:10.1016/j.jag.2022.103134.
34. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition; 2017; pp. 1251–1258.
35. Dang, L.; Pang, P.; Lee, J. Depth-Wise Separable Convolution Neural Network with Residual Connection for Hyperspectral Image Classification. *Remote Sensing* **2020**, *12*, 3408, doi:10.3390/rs12203408.
36. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition; 2016; pp. 770–778.
37. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual Attention Network for Scene Segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019; pp. 3146–3154.
38. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020; pp. 11534–11542.
39. Zhang, Q.-L.; Yang, Y.-B. Sa-Net: Shuffle Attention for Deep Convolutional Neural Networks. In Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); IEEE, 2021; pp. 2235–2239.
40. Li, X.; Hu, X.; Yang, J. Spatial Group-Wise Enhance: Improving Semantic Feature Learning in Convolutional Networks. *arXiv preprint arXiv:1905.09646* **2019**.

41. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv preprint arXiv:1804.03999* **2018**.
42. Peng, D.; Zhang, Y.; Guan, H. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. *Remote Sensing* **2019**, *11*, 1382, doi:10.3390/rs11111382.
43. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition; 2017; pp. 2881–2890.
44. Tang, B.; Liu, Z.; Tan, Y.; He, Q. HRTransNet: HRFormer-Driven Two-Modality Salient Object Detection. *IEEE Transactions on Circuits and Systems for Video Technology* **2022**, *33*, 728–742, doi:10.1109/TCSVT.2022.3202563.

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.