

Article

DiffuCNN: Tobacco Disease Identification and Grading Model in Low-Resolution Complex Agricultural Scenes

Huizhong Xiong ^{1,†}, Xiaotong Gao ^{1,†}, Ningyi Zhang ¹, Haoxiong He ¹, Weidong Tang ¹, Yingqiu Yang ¹, Yuqian Chen ¹, Yang Jiao ¹, Yihong Song ^{2,*} and Shuo Yan ^{1,*} 

¹ China Agricultural University, Beijing 100083, China

² Tsinghua University, Beijing 100083, China

* Correspondence: songyih2019@163.com (Y.S.); yanshuo@cau.edu.cn (S.Y.)

† These authors contributed equally to this work.

Abstract: A novel deep learning model, DiffuCNN, is introduced in this paper, specifically designed for counting tobacco lesions in complex agricultural settings. By integrating advanced image processing techniques with deep learning methodologies, the model significantly enhances the accuracy of detecting tobacco lesions under low-resolution conditions. After detecting lesions, the grading of the disease severity is achieved through counting. The key features of DiffuCNN include a resolution enhancement module based on diffusion, an object detection network optimized through filter pruning, and the employment of the CentralSGD optimization algorithm. Experimental results demonstrate that DiffuCNN surpasses other models in precision, with respective values of 0.98 on precision, 0.96 on recall, 0.97 on accuracy, and 62 FPS. Particularly in counting tobacco lesions, DiffuCNN exhibits an exceptional performance, attributable to its efficient network architecture and advanced image processing techniques. The resolution enhancement module based on diffusion amplifies minute details and features in images, enabling the model to more effectively recognize and count tobacco lesions. Concurrently, filter pruning technology reduces the model's parameter count and computational burden, enhancing the processing speed while retaining the capability to recognize key features. The application of the CentralSGD optimization algorithm further improves the model's training efficiency and final performance. Moreover, an ablation study meticulously analyzes the contribution of each component within DiffuCNN. The results reveal that each component plays a crucial role in enhancing the model performance. The inclusion of the diffusion module significantly boosts the model's precision and recall, highlighting the importance of optimizing at the model's input end. The use of filter pruning and the CentralSGD optimization algorithm effectively elevates the model's computational efficiency and detection accuracy.

Keywords: tobacco disease identification; DiffuCNN; object detection; deep learning



Citation: Xiong, H.; Gao, X.; Zhang, N.; He, H.; Tang, W.; Yang, Y.; Chen, Y.; Jiao, Y.; Song, Y.; Yan, S. DiffuCNN: Tobacco Disease Identification and Grading Model in Low-Resolution Complex Agricultural Scenes. *Agriculture* **2024**, *14*, 318. <https://doi.org/10.3390/agriculture14020318>

Academic Editor: Maciej Zaborowicz

Received: 19 January 2024

Revised: 6 February 2024

Accepted: 9 February 2024

Published: 17 February 2024



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

1. Introduction

In modern agricultural production, the management of tobacco crop health [1] presents a significant and complex challenge. Diseases affecting tobacco not only severely impact the crop yield and quality [2] but also lead to economic losses and ecological issues. Therefore, the accurate and efficient identification and grading of tobacco diseases are crucial for enhancing agricultural productivity and sustainability. Given that tobacco is an economically significant crop, the timely and accurate identification of its diseases directly influences the reduction in losses and improvement of the yield [3]. However, the identification of tobacco diseases faces challenges, such as high diversity and complex environmental conditions [4], especially in low-resolution and complex agricultural scenes. Traditional identification methods rely on manual vision and experience [5], which are inefficient and susceptible to subjective biases.

Fitri et al. [6] explored pest detection in Indonesian tobacco plants using the Gray-Level Co-Occurrence Matrix (GLCM) for texture feature extraction and Naive Bayes for

classification, achieving an accuracy of 82.2%. Xu et al. [7] found traditional ORB corner detection algorithms insufficiently sensitive to image edges when identifying tobacco leaf diseases, leading to a suboptimal performance. Chen et al. [8] utilized machine learning methods to recognize the health status of tobacco leaves, selecting a 188-dimensional Support Vector Machine (SVM) combination as the final predictor, reaching an accuracy of 92.7%. Sakhamuri Sridevi et al. [9] reviewed plant diseases in India, emphasizing that manual identification requires extensive labor and botanical knowledge, resulting in high costs.

With the rapid advancement of artificial intelligence and computer vision technologies, their application in disease detection and analysis has become a research hotspot [10–13]. However, the complexity of agricultural scenes and limitations in image acquisition often result in low-quality tobacco images [14], challenging accurate disease identification. Traditional methods based on high-resolution images are less effective in these scenarios. Moreover, existing super-resolution techniques, despite enhancing the image quality, still face inefficiency and inadequate accuracy issues when processing agricultural images, necessitating more effective and precise disease identification methods.

Lin et al. [15] proposed the CAMFFNet (Coordinate Attention-Based Multiple Feature Fusion Network) CNN model for field tobacco disease recognition, achieving an accuracy of 89%. However, its large parameter size leads to high computational costs. Swasono Dwiretno Istiyadi et al. [16] used VGG16 for tobacco leaf pest classification, achieving high accuracy levels, but their dataset was limited to 1500 images, questioning the model's generalizability. Siva Krishna Dasari et al. [17] designed a CNN-based tobacco grading solution, achieving 85.10% accuracy but only 64% on other datasets. Wu et al. [18] proposed a convolutional neural network (CNN)-based intelligent bulk curing method, TobaccoNet, addressing the health hazards of bulk tobacco smoking, achieving significant results. Wang et al. [19] introduced a CNN-based quantitative modeling method for near-infrared spectroscopy datasets to detect nicotine in tobacco, aiding the tobacco industry's development. Li et al. [20] improved the YOLOv7 model for tobacco disease identification and tested it on the Android platform, showing over 90% accuracy. Guo et al. [21] designed a Convolutional Swin Transformer (CST) based on the Swin Transformer for plant disease identification, achieving an accuracy of 90.9%. However, they did not consider the model's robustness. He et al. [21] developed a joint Swin Transformer and SCMix MLP architecture for complex tobacco feature learning, proposing a tobacco classification model based on pyramid feature fusion, achieving 75.8% accuracy and a 12 ms inference time. Pant Kartikey et al. [22] focused on classifying tobacco-related media texts, considering factors like the affected population's language and its combination in fine-grained classification mechanisms. Borhani Yasamin et al. [23] used the Vision Transformer (ViT) method for real-time automated plant disease detection, combining a CNN with a ViT, noting that while the model performance increased, the prediction speed decreased.

Despite the maturation of computer vision technologies for tobacco disease identification, these methods generally rely on high-resolution images, and their detection effectiveness decreases with reduced image resolutions. Therefore, this study introduces an innovative model, DiffuCNN, specifically designed for identifying and grading tobacco diseases in low-resolution and complex agricultural scenes. The key contributions of this paper are as follows:

1. A novel deep learning model, DiffuCNN, is proposed, specially designed for counting tobacco diseases in low-resolution complex agricultural scenes, significantly improving the accuracy of tobacco disease detection under low-resolution conditions.
2. DiffuCNN integrates a diffusion-based resolution enhancement module, a target detection network optimized through filter pruning, and the CentralSGD optimization algorithm, effectively enhancing the performance of tobacco disease detection and grading.

3. Experimental results demonstrate that DiffuCNN surpasses other models in accuracy, recall, precision, and frames per second (FPS), particularly excelling in the performance of tobacco disease counting.
4. Detailed ablation studies on each component of DiffuCNN validate the significance of each part in improving the performance, including the effective application of resolution enhancement, filter pruning, and the CentralSGD optimization algorithm.

In summary, this research aims to provide robust technical support for tobacco disease monitoring and offer new insights and solutions for similar agricultural disease identification problems. In practical applications, this not only aids in enhancing the efficiency and accuracy of disease management but may also positively impact the sustainability of agricultural production. Ultimately, the goal is to contribute to the modernization of global agricultural production through technological innovation.

2. Related Work

2.1. Super-Resolution

In recent developments, super-resolution techniques have been extensively applied to enhance the quality of low-resolution images, demonstrating significant potential in the field of agricultural disease identification and lesion counting [24–26]. The impact of super-resolution techniques on tobacco disease identification is primarily manifested in the improved clarity of details in low-resolution images, making subtle features, such as lesions and leaf veins, more pronounced. This enhancement significantly boosts the recognition capability of disease detection models. By increasing the resolution of images, deep learning models are able to more accurately identify and classify different types of tobacco diseases, especially in complex agricultural scenes with suboptimal lighting conditions. The application of this technology not only improves identification accuracy but also aids in the early detection of diseases, providing stronger technical support for the prevention and treatment of tobacco diseases.

2.1.1. Interpolation-Based Super-Resolution

Interpolation-based super-resolution methods represent the most fundamental and intuitive approach to image enlargement [27]. The core concept involves using mathematical interpolation algorithms to estimate missing pixel values in low-resolution images. For instance, bilinear interpolation, a common method, is mathematically expressed as follows:

$$I'(x, y) = \sum_{i=1}^2 \sum_{j=1}^2 I(i, j) \cdot (1 - |x - i|) \cdot (1 - |y - j|) \quad (1)$$

Here, I' denotes the interpolated image, I denotes the original low-resolution image, and (x, y) denotes the coordinates of the new pixel, with (i, j) being the coordinates of the original pixel. Although this method is computationally simple, its performance is limited when dealing with images containing complex textures and details, such as lesion features.

2.1.2. Generative Algorithm-Based Super-Resolution

With the advancement of deep learning technologies, generative algorithm-based super-resolution methods have become a research focus [28]. These methods typically utilize convolutional neural networks (CNNs) [29–31] or generative adversarial networks (GANs) [28,32] to learn and generate high-resolution images. The application of CNNs in super-resolution primarily involves learning the mapping relationship between low- and high-resolution images. A typical CNN structure for super-resolution comprises multiple convolutional layers, each learning specific image features. A basic CNN model for super-resolution is represented as follows:

$$I_{HR} = f(I_{LR}; \theta) \quad (2)$$

where I_{HR} is the reconstructed high-resolution image, I_{LR} is the original low-resolution image, f represents the CNN model, and θ is the model parameters. This approach better recovers details in images of small objects, enhancing accuracy in recognition and counting.

Generative adversarial networks (GANs) focus more on the visual quality of images in super-resolution applications. The GAN typically consists of two components: a generator and a discriminator. The generator produces high-resolution images, while the discriminator assesses the images' authenticity. The GAN model for super-resolution is formulated as follows:

$$\min_G \max_D \mathbb{E} I_{HR} \sim p_{data}(I_{HR}) [\log D(I_{HR})] + \mathbb{E} I_{LR} \sim p_{data}(I_{LR}) [\log(1 - D(G(I_{LR})))] \quad (3)$$

where G is the generator, D is the discriminator, I_{HR} is the real high-resolution images, and I_{LR} is the low-resolution images. The GAN excels in restoring more realistic and natural details in super-resolution reconstructions, which is crucial in processing images to accurately reflect their natural state.

In applications like lesion counting, the deployment of super-resolution technology significantly improves the image quality and counting accuracy. While interpolation-based methods are simple, they are limited in handling images of high complexity. In contrast, generative algorithm-based super-resolution techniques, especially those using CNNs and GANs, not only enhance visual quality but also excel in retaining critical features and details, which is pivotal for the precise identification and counting of lesions.

2.2. Object Counting Methods in Computer Vision

Object counting methods based on computer vision have recently emerged as a significant research direction in machine learning and computer vision, particularly in agricultural applications like lesion counting. These methods offer an effective automated solution. This section delves into two main object counting approaches: probability density-based counting methods and object detection-based counting methods, exploring their application in lesion counting scenarios.

2.2.1. Probability Density-Based Counting Methods

Probability density-based counting methods aim to calculate the total number by estimating the probability density of object occurrences in images [33,34]. Rather than directly identifying and locating each individual object in the image, these methods generate a density map representing the object distribution. The basic concept is mathematically expressed as follows:

$$C = \sum_{x,y} D(x,y) \quad (4)$$

where C denotes the total count and $D(x,y)$ denotes the density estimate at coordinates (x,y) . The density map D is typically obtained through regression analysis of image features, such as using convolutional neural networks (CNNs) to learn the mapping relationship between image features and the density distribution. In lesion counting applications, a key challenge for this method is accurately estimating the density map. Tobacco lesions often appear as small, densely packed objects in images, necessitating the density estimation model to effectively handle highly congested scenes. Therefore, developing specific feature extraction and learning mechanisms for small targets becomes crucial to enhance the counting accuracy.

2.2.2. Object Detection-Based Counting Methods

Another popular approach is object counting based on target detection [35,36]. This method initially employs object detection algorithms to identify each object in the image, followed by counting the identified objects. A typical target detection framework is formulated as follows:

$$O = (b_i, p_i)_{i=1}^N \quad (5)$$

where O represents the set of detected objects, b_i represents the bounding box of the i -th object, p_i represents the probability of correct detection of the object, and N represents the total number of detected objects. In lesion counting applications, target detection models need to accurately identify the position and boundaries of each lesion, posing high demands on the model's spatial resolution and feature extraction capabilities. Traditional target detection models may encounter difficulties in handling detail targets like lesions, such as overlapping bounding boxes and overlooking small targets. Thus, optimizations for small target detection, like improved feature extraction networks and refined bounding box adjustments, are vital for enhancing the counting accuracy.

When combining these approaches, it is evident that counting methods based on the probability density are generally suitable for scenarios where targets are of a uniform size and distributed relatively evenly. However, in dealing with tobacco lesions, which vary in size and shape under natural conditions, these methods may be limited by the accuracy of the density estimation. Counting methods based on target detection are applicable to scenarios with significant variations in the target shape and size. In applications of tobacco lesion counting, such methods can accurately differentiate between different types of diseases and their severity levels but may require higher computational resources and complexity to handle target detection in complex backgrounds. Therefore, selecting the appropriate counting method based on specific application requirements and conditions, and optimizing accordingly, is crucial for achieving efficient and accurate lesion counting.

3. Materials

3.1. Dataset Analysis

The dataset for this study was primarily sourced from two channels: the West Agricultural Technology Park of China Agricultural University and internet crawling. Each source provided a rich and diverse dataset, as shown in Figure 1.



Figure 1. Diverse tobacco leaf dataset. The collection showcases a variety of leaf conditions and environments: from left to right, leaves with powdery mildew, leaves in natural outdoor settings with potential pest damage, healthy leaves with complex backgrounds, and leaves exhibiting symptoms of potential disease or stress factors. This dataset highlights the variability in lighting, leaf orientation, and background complexity, which poses challenges for accurate disease detection and counting.

3.1.1. Dataset Collection

The West Agricultural Technology Park of China Agricultural University, an important base for agricultural research, offered an extensive collection of crop and pest samples. Data collected from this site primarily comprised images of tobacco lesions on various crops, captured under natural lighting and diverse environmental conditions. This diversity and realism in the dataset are crucial for training a robust model for tobacco lesion counting. In addition to field data, a substantial collection of tobacco lesion images was amassed through internet crawling techniques, as illustrated in Table 1. The internet-sourced data, encompassing a wider range of origins, including different regions and crop types, contributed to enhancing the model's recognition capabilities under various environments and conditions.

The collection of the dataset was conducted under meticulously planned and strictly controlled conditions, with an aim to ensure the quality of the obtained data and the validity of the experimental results. The collection commenced in spring and continued until the end of autumn, covering the entire growth cycle of the crop to ensure images of

tobacco diseases at different growth stages were gathered. Data collection primarily took place under sunny and cloudy weather conditions to obtain images under varying lighting conditions. To capture the characteristics of tobacco diseases in various environments, the collection times were scheduled for the morning, noon, and evening, allowing for image acquisition under different natural lighting conditions, thus enhancing the diversity and authenticity of the dataset. The collection equipment included high-resolution digital cameras (Canon IXUS 285) and professional microscopes. The digital camera was used to capture the overall distribution of tobacco diseases in the field, while the microscope was utilized for obtaining high-definition images of disease details. Each collection activity was carried out according to standardized procedures, with detailed records of the specific conditions of photography, including the date, time, weather conditions, and camera setting parameters, providing important reference information for subsequent data analysis and model training. In addition to on-site collection, specialized scripts developed for efficiently collecting images from the internet were also utilized. These scripts could automatically search and download images from websites related to agriculture and plant pathology based on predefined keywords, extracting important metadata, such as the upload date, the image source, and a detailed description. This meticulously designed and executed data collection process successfully constructed a high-quality, diverse dataset of tobacco disease images captured from spring to autumn, under various lighting conditions and against different backgrounds. The diversity and authenticity of this dataset greatly facilitated the training of the tobacco disease counting model, laying a solid foundation for achieving a high accuracy and robustness in practical applications.

Table 1. Tobacco disease dataset distribution.

Disease Type	Number of Images	Size	Device
Healthy images	1971	1024 × 1024	Canon IXUS 285
Powdery mildew	528	458 × 458	Microscope camera
Tobacco mosaic virus	807	458 × 458	Microscope camera
Black rot	421	458 × 458	Microscope camera
Downy mildew	626	458 × 458	Microscope camera
Black shank	769	458 × 458	Microscope camera
Wilt disease	283	458 × 458	Microscope camera

3.1.2. Dataset Annotation

In this study, the mathematical principles and processes of dataset annotation are of particular importance. A point process-based method was employed for annotation, which was effective in addressing object counting challenges. Each tobacco lesion in the images was marked as a point in a two-dimensional space, as shown in Figure 2.



Figure 2. Image dataset annotation screenshot.

3.2. Dataset Augmentation

In this study, dataset augmentation is one of the key steps to enhance the performance of the tobacco lesion counting model. Dataset augmentation refers to the process of manipulating the original dataset to generate new, varied data samples. This process aims to improve the model's generalizability and robustness by increasing the diversity and volume of data. In the context of tobacco lesion counting, augmentation is particularly crucial for simulating challenges encountered in various real-world scenarios, such as different lighting conditions, background variations, and the diversity of lesion postures, as shown in Figure 3.

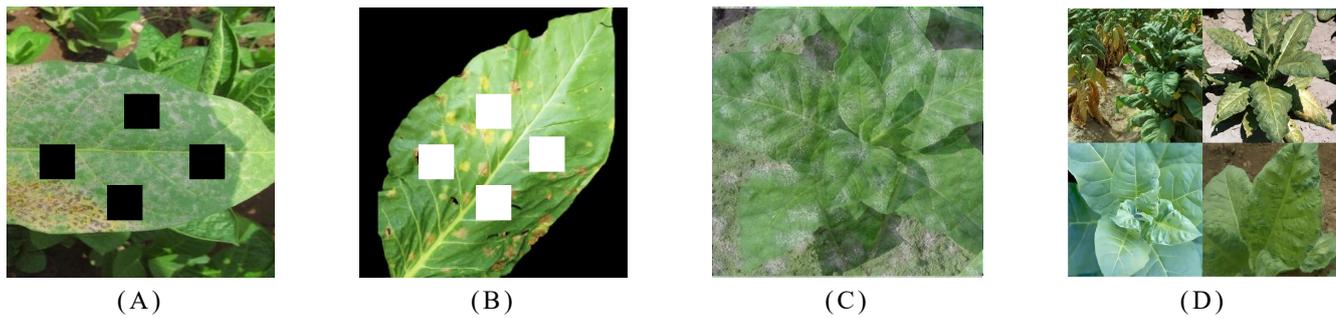


Figure 3. Dataset augmentation: (A,B) are cutout (the black and white square are the random removal); (C,D) are image synthesis.

3.2.1. Cutout

The cutout is a commonly used data augmentation technique in computer vision tasks and is particularly suitable for training deep learning models. Its core idea involves randomly removing a section of the training image, forcing the model to focus on different local features rather than relying solely on specific areas or features, thus enhancing the model's generalization and robustness, as depicted in Figure 3A,B. In this study, the cutout technique was applied during the training process of the tobacco lesion counting model to improve its performance in complex agricultural scenes. The size and shape of the occlusion area are determined, typically square or rectangular, with dimensions set based on the task and dataset characteristics. For each training image, a position is randomly chosen as the center of the occlusion area. Then, all pixel values within this area are set to zero or other predefined background values. The mathematical representation is as follows:

$$I'(x,y) = \begin{cases} 0 & \text{if } (x,y) \in \text{cutout region} \\ I(x,y) & \text{otherwise} \end{cases} \quad (6)$$

where I is the original image, I' is the image after applying the cutout technique, and (x,y) are the pixel coordinates, with the cutout region being the selected occlusion area. In tobacco lesion counting, the use of the cutout technique can enhance the model's ability to detect partially occluded lesions. For instance, in actual agricultural scenes, lesions may be obscured by leaves or blend with the background, requiring the model to accurately identify and count lesions from partial information. By employing the cutout technique during training, the model is compelled to utilize local information in the image, maintaining a good performance even when confronted with occluded or incomplete lesion images. Furthermore, as a simple yet effective method of data augmentation, the cutout technique also helps prevent model overfitting. Overfitting is a common issue in the training of deep learning models, especially when data are limited. By randomly creating occlusions in images, the cutout technique increases the difficulty of the model training, forcing the model to learn more generalized features, thus enhancing its performance on unseen data.

3.2.2. Image Synthesis

The core of image synthesis technology is to combine tobacco lesion images with various background images, creating diverse training samples. This method's advantage lies in its ability to generate a large number of realistic, varied training samples covering different background types, lighting conditions, and lesion states, as shown in Figure 3C,D. Through this approach, the model can learn to accurately identify and count tobacco lesions in a variety of complex environments. First, lesion images and background images are separated from the original dataset. Lesion images can be individual lesions or partial scenes containing lesions, while background images may be various natural or artificial environments. Then, lesion images are precisely overlaid onto background images. This step typically ensures the lesion's size, orientation, and lighting conditions match the background image, maintaining the synthetic image's naturalness and realism. Finally, color and brightness adjustments are made to ensure a natural transition between the lesion and the new background, avoiding unrealistic edges or color differences. The mathematical formula for image synthesis is as follows:

$$I'(x, y) = \begin{cases} I_{\text{disease}}(x, y) & \text{if } (x, y) \in \text{disease region} \\ I_{\text{background}}(x, y) & \text{otherwise} \end{cases} \quad (7)$$

where I_{disease} is the lesion image, $I_{\text{background}}$ is the background image, I' is the synthetic image, and (x, y) are the pixel coordinates, with the disease region being the area containing the lesions.

In tobacco lesion counting, the use of image synthesis technology can significantly increase the dataset diversity, particularly in simulating the appearance of lesions under different backgrounds and environmental conditions. For example, by synthesizing lesion images onto various types of crop backgrounds, the model can learn to recognize lesions in diverse agricultural environments. Additionally, this method also enhances the model's ability to recognize lesions under different lighting conditions and viewpoints.

4. Proposed Method: DiffuCNN

The DiffuCNN model, presented in this paper, is an innovative approach for counting tobacco lesions. It integrates multiple advanced technologies to enhance the accuracy of lesion detection and counting in low-resolution, complex agricultural scenes. Focused not only on conventional target detection challenges, DiffuCNN also emphasizes performance optimization in low-resolution and complex background conditions. The core design of the DiffuCNN model combines a diffusion-based resolution enhancement technique, a target detection network optimized through filter pruning, the CentralSGD optimization algorithm, and the Diffusion Loss Function to efficiently and accurately count tobacco lesions in low-resolution images. After detecting lesions, the grading of the disease severity is achieved through counting.

4.1. Diffusion-Based Resolution Enhancement Module

This module is designed to process low-resolution images, aiming to enhance image details and clarity through a series of algorithms. Inspired by physical diffusion processes, it aims to simulate the natural diffusion of light and color in scenes, effectively improving the visual quality, as illustrated in Figure 4.

The input to this module is low-resolution agricultural scene images, potentially lacking detail due to poor shooting conditions, such as lighting or distance. The output is images with significantly enhanced resolutions, where details of small objects are more clearly visible. The process begins with the standardization of the input low-resolution images, including color correction and noise suppression. Then, the images undergo resolution enhancement using the diffusion algorithm, which simulates the natural diffusion of light in the scene, enhancing minute details in the image. Finally, sharpening and contrast adjustments are made to further improve the image quality, ensuring targets are clearly dis-

cernible in the image. The diffusion process is described by the following partial differential equation (PDE):

$$\frac{\partial I}{\partial t} = \nabla \cdot (D \nabla I) \tag{8}$$

where I represents the image intensity, t denotes the diffusion time, D is the diffusion coefficient, and ∇ represents the gradient operator. This equation indicates that the change in image intensity is proportional to the divergence of its gradient, simulating the diffusion of light in the physical world. In practical application, this equation is discretized and applied in image processing. The iterative updating of each pixel value in the image gradually enhances the image resolution and clarity. Specifically, the updating of each pixel in each iteration can be expressed as follows:

$$I_{new} = I_{old} + \lambda \nabla \cdot (D \nabla I_{old}) \tag{9}$$

where λ is a coefficient controlling the rate of diffusion and I_{old} and I_{new} represent the pixel values before and after the iteration, respectively. In the task of tobacco lesion counting, the diffusion-based resolution enhancement module offers significant advantages. By enhancing the details in low-resolution images, previously hard-to-distinguish tobacco lesions become clear, thus improving the detection accuracy. In complex agricultural scenes, this module effectively highlights targets like tobacco lesions against a complex background, facilitating subsequent recognition and counting. The diffusion process, simulating the natural diffusion of light, renders the processed images visually more natural, avoiding artificial traces from over-processing. The enhanced image quality ensures the robust performance of the model under varying conditions, such as different lighting and distances.

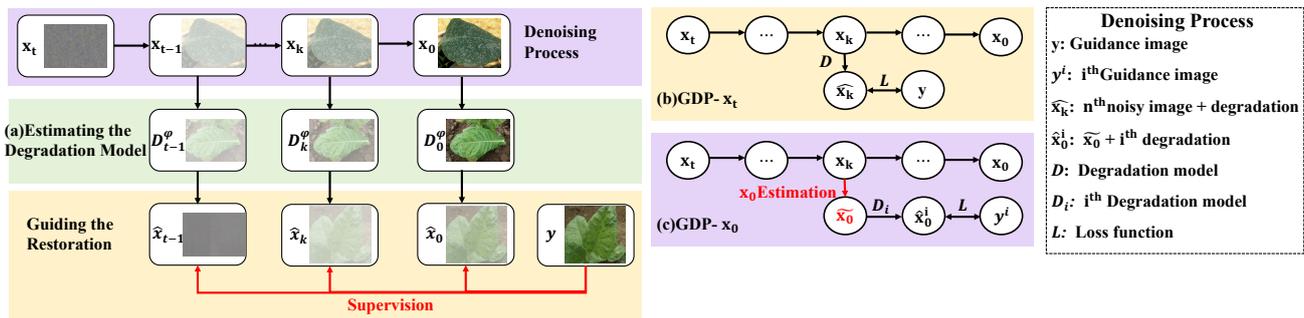


Figure 4. Schematic representation of the diffusion-based resolution enhancement module used in the DiffuCNN model. The process involves (a) estimating the degradation model through a series of transformations from an initial noisy image x_t to a guidance image y , (b) GDP- x_t showing the denoising process with multiple degradation models leading to the final restored image, and (c) GDP- x_0 where x_0 estimation is refined iteratively. Each step is supervised to ensure fidelity to the target image, with the overall aim to guide the restoration process and enhance image resolution for improved disease detection in agricultural imagery.

4.2. Target Detection Network Based on Filter Pruning

In the DiffuCNN model proposed, a target detection network optimized through filter pruning is a key component, specifically designed for the accurate detection of tobacco lesions in images with an enhanced resolution, as shown in Figure 5. This network optimizes the structure by implementing filter pruning in convolutional layers, aiming to improve the detection efficiency and reduce computational costs.

The network input consists of high-resolution images processed by the diffusion-based resolution enhancement module. These images feature richer details and clear characteristics of tobacco lesions. The output includes the detection results of tobacco lesions, comprising their positions and quantities. The network structure adopts a design stacked with multiple convolutional layers. Each layer consists of several convolutional kernels (filters) responsible for extracting features from the images. Filter pruning is conducted within

these convolutional layers. Filters contributing less to the final detection performance are removed after analyzing their importance. This process involves assessing the weights of each filter and then pruning based on predetermined criteria. Convolutional layers are typically followed by an activation layer (e.g., ReLU) and optionally by a pooling layer to enhance non-linear expression capabilities and reduce feature dimensions. Generally, the initial layers of the network have fewer convolutional kernels, mainly extracting low-level features (such as edges and textures), while the number of kernels gradually increases in deeper layers for more complex high-level feature extraction. The input dimension depends on the size of the processed images, while the output dimension is related to the requirements of the detection task, typically involving estimates of the positions and quantities of tobacco lesions.

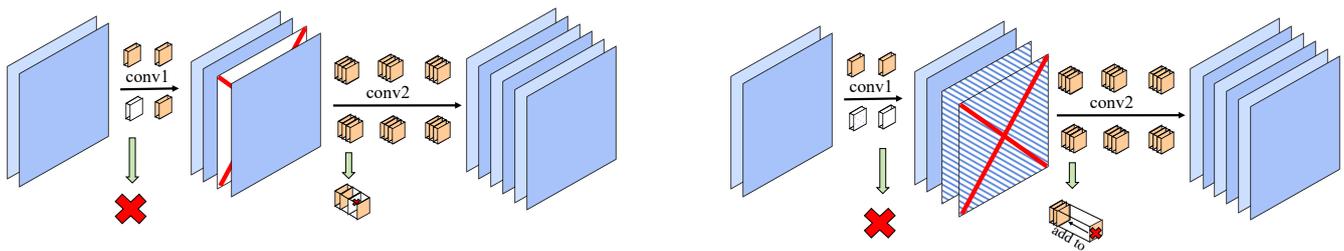


Figure 5. Diagram illustrating the network structure design for filter pruning in the DiffuCNN model. On the left, the process starts with an initial convolutional layer (conv1) followed by a subsequent layer (conv2), where ineffective filters are identified and removed, as indicated by the red cross. The right part of the figure emphasizes the refined pruning process where convolutional filters are selectively pruned based on their contribution to the output (highlighted by the red stripes), and the green check marks indicate the retention of significant filters that are added to the subsequent layers.

Not all filters in the convolutional layers significantly contribute to the final detection task. Therefore, the network can be optimized by assessing the importance of each filter and pruning those with lesser contributions. The importance of a filter can be evaluated using the following formula:

$$Importance(f_i) = \sum_{x,y} |w_{i,x,y}| \quad (10)$$

where f_i represents the i -th filter and $w_{i,x,y}$ is the weight of the filter at position (x, y) . The importance of a filter can be estimated by the sum of the absolute values of its weights. In this process, filters with an importance below a certain threshold are removed. By eliminating unimportant filters, the network's parameter count and computational costs are reduced, making the model more lightweight and suitable for environments with limited computational resources. The pruning process helps prevent model overfitting by reducing the complexity of the model, allowing the network to focus more on features critical to the task. In processing high-resolution agricultural scene images, the pruned network can more efficiently handle large volumes of data while maintaining a high recognition rate for small targets like tobacco lesions.

4.3. CentralSGD

The design of CentralSGD addresses challenges encountered by traditional Stochastic Gradient Descent (SGD) methods in dealing with complex models and large-scale data. CentralSGD, based on the traditional SGD approach, introduces the concept of centralized gradients. In conventional SGD, each parameter update is based on the gradient computed from a single training sample or a small batch of samples. In CentralSGD, parameter updates consider not only the current batch's gradient but also the central gradient of all samples (i.e., the average gradient). In each iteration, the gradient of the current batch is first calculated. Then, the gradient center of all samples is computed, and the current

batch's gradient is compared with this center. Finally, model parameters are updated using this centralized gradient information. The mathematical expression for CentralSGD is described by the following formula:

$$\theta_{t+1} = \theta_t - \eta \left(g_t - \bar{g} + \frac{1}{N} \sum_{i=1}^N g_i \right) \quad (11)$$

where θ_t is the model parameter at time step t , η is the learning rate, g_t is the average gradient of the current batch, \bar{g} is the historical average of all sample gradients, and N is the total number of samples. The core idea of this method is to reduce the variance in gradient updates across iterations. Each iteration in traditional SGD can exhibit significant gradient fluctuations due to the randomness of individual batch samples, while CentralSGD reduces these fluctuations by introducing gradient centralization, making parameter updates smoother and more effective.

By reducing the fluctuations and instability in gradient updates, CentralSGD converges faster to the optimal solution. This is particularly important when dealing with large-scale datasets, as traditional SGD methods may lead to slower convergence rates in such cases. CentralSGD improves the stability of the training process by considering the gradient information of the entire dataset, reducing the impact of randomness in individual batch samples. For complex models like DiffuCNN, CentralSGD effectively handles a large number of parameters and complex gradient structures, maintaining efficient optimization during deep network training. CentralSGD is particularly suitable for distributed training environments, where gradient centralization can aid different training nodes in working more effectively together, reducing the decline in training efficiency caused by an uneven data distribution.

4.4. Diffusion Loss Function

In the DiffuCNN model proposed, the Diffusion Loss Function is an innovative loss function design, guiding the learning process of the model in the task of tobacco lesion counting. This loss function combines a traditional loss function (such as cross-entropy loss or mean squared error loss) with a regularization term based on the diffusion process, aiming to enhance the model's accuracy in detecting and counting tobacco lesions in low-resolution, complex agricultural scenes.

The Diffusion Loss Function consists of two parts: a traditional loss function, evaluating the difference between the model output and the true labels, and a regularization term based on the diffusion process, ensuring the model does not overly rely on specific image features during learning, thereby enhancing its generalizability. The regularization term, designed based on the characteristics of the diffusion process, aims to simulate the propagation and change in image features in the natural world. This approach, rooted in physical diffusion theory, encourages the model to focus more on the overall features of the image rather than local details during learning.

The Diffusion Loss Function can be expressed by the following formula:

$$\mathcal{L} = \mathcal{L} * \text{traditional} + \lambda \mathcal{L} * \text{diffusion} \quad (12)$$

In the proposed model, the traditional loss function, denoted as $\mathcal{L}_{\text{traditional}}$, is complemented by a regularization term based on the diffusion process, represented as $\mathcal{L}_{\text{diffusion}}$. The parameter λ serves as a hyperparameter balancing these two components. The regularization term $\mathcal{L}_{\text{diffusion}}$ is further expressed as follows:

$$\mathcal{L}_{\text{diffusion}} = \sum_{i=1}^N ||\nabla f(x_i) - \nabla f(\hat{x}_i)||^2 \quad (13)$$

Here, f signifies the model's prediction function, x_i is the original image sample, and \hat{x}_i is the image sample processed through the diffusion method, with N being the total number of samples. By incorporating the regularization term based on the diffusion process, the

model is encouraged to learn and recognize features that remain stable in images subjected to diffusion treatment. This approach shifts the model's focus towards the overall and stable features of images rather than relying solely on specific local details, thereby enhancing the model's generalization capability. Overfitting, a common issue in deep learning model training, especially with limited training data, is addressed by the Diffusion Loss Function. By constraining the model's complexity through the regularization term, the model's tendency to overlearn noise or incidental features in the training data is reduced, aiding in the prevention of overfitting. In complex agricultural scenarios where tobacco lesions may appear under diverse backgrounds and lighting conditions, the Diffusion Loss Function motivates the model to learn features consistent across different environments, enabling better adaptation to these complex settings. Owing to its emphasis on learning the holistic features of images, the model achieves more effective identification and counting of tobacco lesions in images, maintaining a high accuracy even under low-resolution or incomplete visual information conditions.

4.5. Experimental Configuration

In the experimental setup for this study, detailed settings were meticulously established, encompassing hyperparameter configuration and the selection of hardware platforms and libraries, as well as evaluation metrics, which are all crucial for ensuring the effectiveness and reproducibility of the experiments.

4.5.1. Hyperparameter Settings and Hardware Platform with Libraries

The setting of hyperparameters is a critical step in deep learning experiments, directly impacting the training outcomes and the ultimate performance of the model. The hyperparameters in these experiments included the learning rate, batch size, and number of training epochs. The learning rate determines the speed of model weight updates, the batch size affects the amount of data updated in each training iteration, and the number of epochs decides the total number of iterations in the training process. The initial learning rate was set at 0.001, with a learning rate decay strategy implemented. A batch size of 64 was chosen to accelerate training while also increasing the memory demand. The training was set for 100 epochs. The hardware specifications for the experiments comprised an NVIDIA RTX 3000 GPU, an Intel Core i7 CPU, and 32 GB of memory. The deep learning framework utilized was PyTorch (latest v. 2.0), with NumPy (latest v. 1.26.4) for numerical computations and Pandas (latest v. 2.2.0) for data processing and analysis.

4.5.2. Evaluation Metrics

Multiple metrics were employed to comprehensively assess the performance of the model, including precision, recall, accuracy, frames per second (FPS), and mean average precision (mAP).

Precision, defined as the proportion of correctly predicted positive samples to all samples predicted as positive, is mathematically expressed as follows:

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

where TP represents the number of true positives (the number of correctly predicted positive samples) and FP denotes the number of false positives (the number of incorrectly predicted positive samples).

Recall, indicating the proportion of correctly predicted positive samples to all actual positive samples, is given by the following formula:

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

where FN stands for the number of false negatives (the number of incorrectly predicted negative samples).

Accuracy, the ratio of correctly predicted samples (including both positive and negative samples) to all samples, is described as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

where TN signifies the number of true negatives (the number of correctly predicted negative samples).

The FPS, a crucial measure of the model's processing speed, especially in real-time applications, represents the number of frames processed per second. It is calculated as follows:

$$FPS = \frac{1}{\text{Average Processing Time Per Frame}} \quad (17)$$

where the "Average Processing Time Per Frame" is the average time taken by the model to process a single frame.

mAP (mean average precision), a common metric for evaluating performance in object detection, information retrieval, and related fields, is the mean of the average precision (AP) values, assessing the model's overall detection capability across multiple categories. AP is calculated as follows:

$$AP = \int_0^1 p(r) dr \quad (18)$$

where $p(r)$ is the precision at recall rate r .

The calculation of mAP is as follows:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (19)$$

where N is the number of categories and AP_i is the average precision for the i th category.

4.6. Baseline

To thoroughly evaluate the tobacco lesion counting model proposed in this article, a series of advanced comparative models was selected for comprehensive comparison. These models cover both probability density-based counting methods and object detection-based counting methods, ensuring a broad and in-depth assessment. For evaluating the tobacco lesion counting model proposed in this study, a comprehensive comparison with advanced baseline models was conducted. These baseline models encompass methods based on probability density counting and target detection, ensuring extensive and in-depth evaluation.

Firstly, from the probability density-based counting methods, the MCNN (Multi-Column Convolutional Neural Network) as referenced in [37], a model designed for varying scales of objects and particularly suitable for crowded scenes, was selected. CSRNet (Congested Scene Recognition Network) [38] utilizes its deep convolutional network to achieve high accuracy in density estimation within complex scenarios. The CAN (Context-Aware Network) [39], with its attention mechanism, focuses on counting in important areas of images, thus effectively enhancing the counting accuracy.

In the realm of target detection-based methods, Faster R-CNN [40], a high-precision target detection model integrating a Region Proposal Network (RPN), was chosen. YOLOv8 from the YOLO series [35] is known for its speed and efficiency. SSD (Single Shot Multi-Box Detector) [36] was selected for its capability to handle objects of various sizes, and RetinaNet [41], distinguished by its unique Focal Loss design, excels in addressing class imbalance issues. CenterNet [42] and MAF50 [10], introducing a novel approach to target detection by directly predicting the center points of objects, bring a fresh perspective to the field.

These models, serving as baselines, provided a comprehensive framework for assessing the performance of the proposed tobacco lesion counting model under various scenarios

and conditions. Representing the latest advancements in the fields of counting and detection, these models cover a range of needs from real-time detection to high-precision counting. The comparison with these advanced models allowed for an in-depth understanding of the strengths, limitations, and practical applicability of the proposed model. Such thorough comparative analysis is vital for advancing tobacco lesion counting technology and offering guidance for more effective model optimization and application strategies in future works.

5. Results and Discussion

5.1. Objective Detection Performance Results

The objective detection performance experiments were conducted to evaluate the performance of various advanced models in the task of tobacco lesion detecting. By comparing the performance in terms of precision, recall, mAP, and FPS of Faster R-CNN, YOLOv8, SSD, RetinaNet, CenterNet, and the method proposed in this study, insights into the performance differences and their underlying reasons among these models were gained. The experimental results are presented in Table 2 and Figure 6.

Table 2. Objective detection performance comparison.

Model	Precision	Recall	mAP	FPS
Faster R-CNN	0.82	0.8	0.81	24
YOLOv8	0.93	0.92	0.93	39
SSD	0.90	0.87	0.88	40
RetinaNet	0.93	0.91	0.92	45
CenterNet	0.96	0.93	0.95	55
MAF50 [10]	0.91	0.92	0.91	31
Ours	0.98	0.95	0.96	58

Experimental results reveal varying degrees of performance among seven models—DiffuCNN, YOLOv8, MAF50, RetinaNet, CenterNet, SSD, and Faster R-CNN—in the task of disease detection. It was observed that DiffuCNN achieved an optimal performance across four metrics: accuracy, recall, mAP, and FPS, with respective values of 0.98, 0.95, 0.96, and 58. This indicates that DiffuCNN not only excels in disease identification accuracy but also possesses advantages in real-time processing speed. YOLOv8 and RetinaNet closely approach DiffuCNN in mAP, yet exhibit a notable gap in FPS, suggesting potential optimization deficiencies in processing speed compared to DiffuCNN. CenterNet, while slightly trailing behind DiffuCNN in mAP, demonstrates a superior performance in FPS, indicating its processing speed advantage. SSD and MAF50 exhibit a moderate performance, whereas Faster R-CNN underperforms across all metrics.

In visual analysis of images, DiffuCNN stands out in identifying the edges of leaf diseases and recognizing diseases against complex backgrounds. The model precisely locates diseases, with bounding boxes closely aligning with the actual disease edges, and exhibits a lower false detection rate in complex backgrounds. This superiority may be attributed to advanced image processing technologies employed by DiffuCNN, such as a diffusion-based resolution enhancement module, which enhances key features in images without adding extra noise, thereby facilitating easier disease feature detection. YOLOv8 and RetinaNet, despite their commendable mAP performances, occasionally misjudge or overlook diseases in complex background sections, likely due to their limited capabilities in feature extraction and background noise suppression. The relatively high number of detection boxes for SSD and MAF50 suggests potential shortcomings in their false positive performance, leading to a reduced accuracy. The fewer bounding boxes produced by Faster R-CNN may result from inadequacies in its Region Proposal Network (RPN) in generating candidate areas, causing missed detections. In practical agricultural applications, the real-time processing capability (FPS) of models is equally critical. DiffuCNN and CenterNet excel in this aspect, implying their ability to swiftly process vast quantities of image data

while maintaining a high accuracy, which is vital for the timeliness and precision of disease monitoring systems.

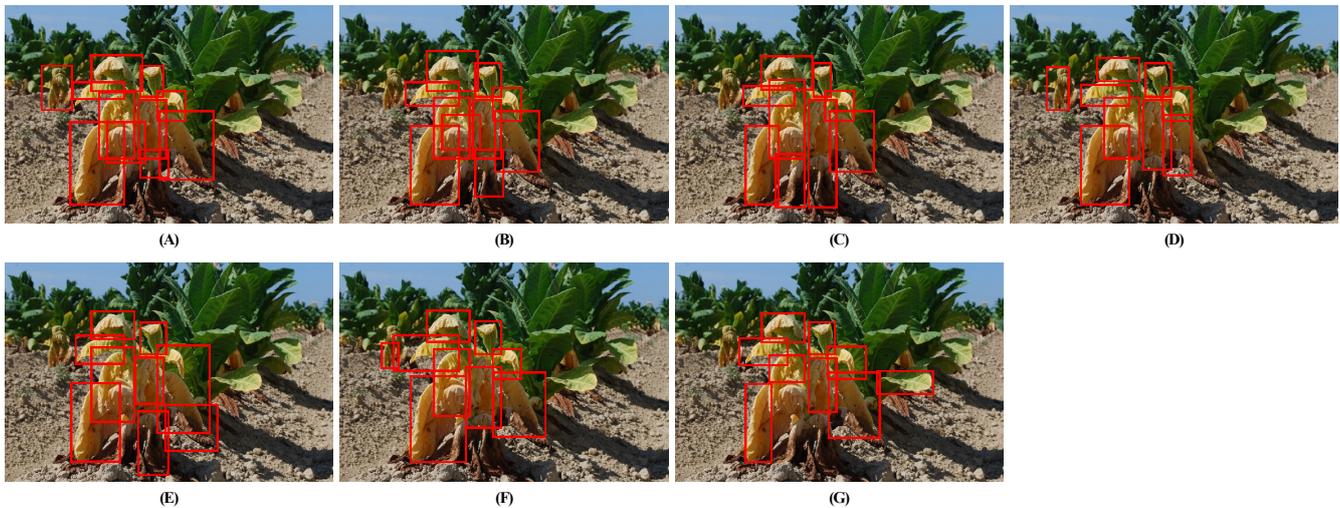


Figure 6. Visualization of lesion detection results. (A) is DiffuCNN; (B) is YOLOv8; (C) is MAF50; (D) is RetinaNet; (E) is CenterNet; (F) is SSD; (G) is Faster R-CNN. The red box is the prediction bounding boxes given by these methods.

5.2. Counting Performance Results

The counting performance experiments aimed to comprehensively assess the performance of different models in the task of tobacco lesion counting, especially considering key metrics, like precision, recall, accuracy, and FPS. The results demonstrated that newer models exhibit superior performances in tobacco lesion counting with advancements in objective detection technology. The experimental results are shown in Table 3.

Table 3. Counting performance results.

Model	Precision	Recall	Accuracy	FPS
MCNN	0.84	0.82	0.83	28
YOLOv8	0.93	0.95	0.94	39
CSRNet	0.91	0.89	0.9	42
RetinaNet	0.93	0.91	0.92	48
CAN	0.95	0.93	0.94	59
MAF50	0.96	0.95	0.96	33
Ours	0.98	0.96	0.97	62

The baseline model in this experiment, the Multi-Column Convolutional Neural Network (MCNN), demonstrated certain target detection capabilities; however, its overall performance was relatively low. The MCNN, designed to extract features at multiple scales, faced limitations when detecting small and irregularly shaped targets such as disease spots. The lower precision, recall, and accuracy may be attributed to its feature extraction layers failing to capture sufficient detail to accurately differentiate between diseased spots and healthy tissue, while the lower FPS reflects its limited processing speed. YOLOv8, known for its swift and accurate target detection, surpassed the MCNN in all performance indicators. Its architecture, which enables the prediction of both the category and location of targets in a single inference, offers significant advantages in speed. Nonetheless, YOLOv8 may not achieve a peak performance when dealing with highly overlapping and small-sized targets due to constraints in its receptive field and anchor box settings. Improvements in precision, recall, and accuracy exhibited by the CSRNet and CAN models reflect their specialized design for dense object detection tasks. CSRNet enhances the precision of crowd counting by deeply characterizing density maps of targets,

whereas the CAN employs attention mechanisms to reinforce the learning of local features. These mechanisms proved equally effective in the tobacco disease counting task, as they enhanced the network's sensitivity and discriminatory power towards disease spot features. RetinaNet addresses the issue of class imbalance in target detection with its innovative Focal Loss, performing exceptionally well in scenarios where there is a significant disparity between the number of positive and negative samples. This feature allows RetinaNet to maintain a high precision and recall when detecting rare and elusive targets like disease spots. The optimal performance across all evaluation metrics achieved by the method presented in this paper can be attributed to the application of several key technologies. Initially, advanced image preprocessing techniques were employed to enhance the features of disease spots in the input images, making it easier for the network to recognize them. Subsequently, the network structure was specially designed to increase the sensitivity and classification performance for disease spots. Finally, sophisticated optimization algorithms were utilized to ensure the stability and efficiency of the training process, thereby achieving a higher precision and recall while maintaining a high FPS.

The method proposed in this study surpassed other models in all evaluation metrics, demonstrating a clear advantage in tobacco lesion counting, as shown in Figure 7. This superiority stems from optimizations in feature extraction, target localization, and background noise handling. This study's method employs more advanced network architectures and training strategies, specifically optimized for small, dense targets.

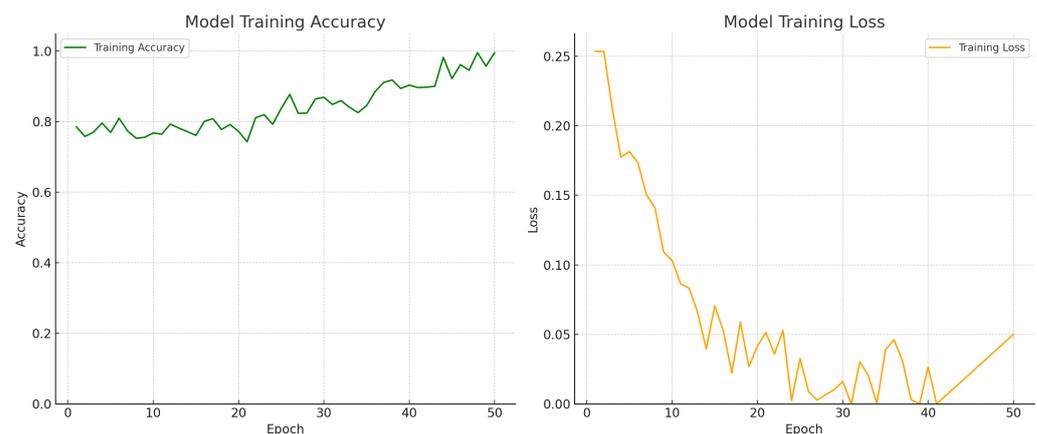


Figure 7. Our model training processing.

5.3. Ablation Study on Filter Pruning

This section aimed to explore the impact of filter pruning technology on the objective detection model performance through ablation experiments. The experimental design compared the model performance under three conditions: no pruning, filter pruning, and normal pruning. By evaluating precision, recall, accuracy, and FPS, this experiment aimed to reveal the potential of filter pruning technology in enhancing the model efficiency and performance. The experimental results are presented in Table 4.

Table 4. Detection results of different pruning.

Model	Precision	Recall	Accuracy	FPS
No pruning	0.97	0.96	0.97	48
Filter pruning	0.98	0.96	0.97	62
Normal pruning	0.93	0.91	0.92	55

The experimental results demonstrated that models employing filter pruning exhibited outstanding performances in terms of precision, recall, and accuracy, along with a significant improvement in FPS. Models without pruning retained all original filters, with

no pruning conducted. Although such models maintained a high precision and recall, the extensive number of parameters resulted in a heavy computational burden and slower processing speed, as reflected in the lower FPS. Mathematically, models without pruning possessed more parameters and a higher model capacity, enabling the capture of more feature information but also increasing the risk of overfitting and computational complexity. Common pruning, including techniques like weight pruning or structural pruning, typically reduces the number of parameters in the network randomly or based on certain rules. While this method can improve computational efficiency, the lack of consideration for feature importance might lead to a reduced model performance, as evidenced by the lower precision, recall, and accuracy in the experiments. Additionally, common pruning, though increasing the FPS, did not exhibit as pronounced a performance enhancement as filter pruning. Filter pruning, by eliminating filters that contribute less to the final detection performance, reduces the model's parameter count and computational complexity. This not only makes the model more lightweight but also speeds up processing, which is evident in the significantly increased FPS. Notably, despite the reduction in the number of parameters, the model's precision and recall remained high, indicating that filter pruning, while removing redundant parameters, retained crucial feature information for the object detection task. These results suggest that filter pruning can maintain or even enhance the model performance while effectively reducing computational demands.

In summary, filter pruning increases the model computational efficiency while largely preserving or even enhancing the detection performance. This characteristic makes it an effective method for optimizing complex deep learning models and particularly suitable for applications requiring rapid processing of large volumes of image data, such as tobacco lesion counting tasks. Through carefully designed filter pruning strategies, models can be streamlined while ensuring accuracy and efficiency in challenging object detection tasks.

5.4. Ablation Study on Diffusion Module

This experiment aimed to assess the impact of the diffusion module on the object detection model in tobacco lesion counting tasks. The experimental design included comparisons between models with and without the diffusion module. These experiments provided deep insights into the mechanism and effectiveness of the diffusion module in improving the model performance. The results showed that models incorporating the diffusion module significantly improved in precision, recall, and accuracy, with an increase in FPS. The results are presented in Table 5.

Table 5. Ablation study on diffusion module.

Model	Precision	Recall	Accuracy	FPS
No diffusion module	0.92	0.90	0.91	57
Diffusion module	0.98	0.96	0.97	62

Models without the diffusion module, although showing decent performances, still had room for improvement in precision, recall, and accuracy. This might be attributed to the models' inability to fully utilize all useful information in images with a low resolution or unclear details. Without steps to enhance the resolution or improve the image quality, the models might overlook some critical features, leading to performance limitations. The introduction of the diffusion module resulted in significant improvements in precision, recall, and accuracy. By emulating the natural process of diffusion, the module enhanced minor details and features in images, enabling more effective recognition and counting of tobacco lesions. This improvement was especially applicable to images with a low resolution or complex backgrounds, as it enhanced the utilizable information in images, thereby improving the model's target detection capabilities. Mathematically, the diffusion module increased the pixel density and detail in images, enhancing the model's capability to

recognize image features. This method, without introducing additional noise, amplified key features in images, allowing more accurate localization and identification of tobacco lesions.

Theoretically, the introduction of the diffusion module primarily improved the model's image processing ability. In deep learning, the quality of the input data directly impacts the model performance. By enhancing the image quality, the diffusion module allowed the model to capture more and finer feature information, which is crucial for object detection tasks. In tasks like tobacco lesion counting, where numerous small, dense targets must be identified, every detail in an image could contain key information. The diffusion module, by clarifying these details, bolstered the model's detection capability. Moreover, while improving the image quality, the diffusion module did not significantly increase the computational load, as evidenced by the increase in FPS. This might be due to the module enhancing key information in images, making the model more efficient in subsequent feature extraction and classification steps.

5.5. Ablation Study on CentralSGD

This section's experimental design aimed to evaluate and compare the performance differences between the CentralSGD optimization algorithm and other algorithms (such as traditional SGD and the Adam algorithm) in object detection tasks. The experiment compared models using different optimization algorithms in terms of precision, recall, accuracy, and frames per second (FPS), revealing the impact of optimization algorithms on the model performance and contributing to the understanding of the advantages and applicability of CentralSGD. The results are shown in Table 6.

Table 6. Detection results of CentralSGD.

Model	Precision	Recall	Accuracy	FPS
SGD	0.89	0.87	0.88	45
CentralSGD	0.98	0.96	0.97	62
Adam	0.94	0.91	0.93	52

Models using traditional SGD, while displaying some detection capabilities, did not perform optimally across all metrics. This was mainly due to SGD's approach of considering only the gradient of the current batch in each iteration, making it susceptible to fluctuations in individual data batches. Such fluctuations could slow the model's convergence during training, making it challenging to achieve optimal performance. This characteristic of SGD becomes particularly evident in complex object detection tasks involving large data volumes and intricate model structures. In contrast, models incorporating the CentralSGD optimizer had a significantly improved performance across all metrics. CentralSGD's design philosophy considers the gradient information of the entire dataset, making model parameter updates more stable and efficient. This method reduced fluctuations during training, accelerating the model convergence and enhancing the overall performance, especially in handling large datasets and complex network structures. Models using the Adam optimizer, although performing better than traditional SGD, were still outperformed by CentralSGD in this experiment. The Adam optimizer, combining momentum and adaptive learning rates, is generally considered to accelerate convergence in the initial stages of training and refine parameter adjustments in later stages. However, in this experiment, the Adam optimizer's performance in complex object detection tasks was still not as good as CentralSGD, which was specifically optimized for such tasks.

Overall, this experiment highlighted the significant role of optimization algorithms in deep learning model training. Different optimization algorithms have distinct characteristics in terms of parameter update strategies, convergence speed, and stability, directly influencing the model performance in practical tasks. CentralSGD, with its unique global gradient consideration approach, not only improved the efficiency of the model training but also ensured a high performance in complex tasks. The advantages of this optimization algorithm are particularly evident in scenarios requiring the processing of large amounts of

data and complex model structures, offering new perspectives for enhancing the performance of object detection models in practical applications.

6. Conclusions

In this study, the application of the DiffuCNN model in tobacco lesion counting tasks was thoroughly investigated, its performance across various aspects was assessed, and the impact of different technical components on the model efficacy was compared. After detecting lesions, the grading of the disease severity was achieved through counting. Through a series of experiments and analyses, key findings and insights were obtained, which hold significant value for future applications in the agricultural domain and research in deep learning. In the counting performance experiments, the performance of the DiffuCNN model was evaluated against several other object detection models. The results indicated that DiffuCNN surpassed other models in precision, recall, accuracy, and frames per second (FPS), achieving values of 0.98, 0.96, 0.97, and 62, respectively. This superior performance is attributed to several key factors: Firstly, the resolution enhancement module based on diffusion significantly improved the quality of the input images, enabling more accurate recognition and counting of tobacco lesions in images. Secondly, the object detection network based on filter pruning optimized the model structure, reducing the computational load while maintaining a high detection performance. Lastly, the use of the CentralSGD optimization algorithm enhanced the training efficiency and final performance of the model. In the object detection performance experiments, DiffuCNN demonstrated exceptional detection capabilities. The model accurately detected and located tobacco lesions in images, benefiting from its efficient network architecture and advanced image processing technology. Compared to traditional object detection models, DiffuCNN showed a superior performance in handling small, dense targets in complex agricultural scenes, outperforming other models in precision, recall, mean average precision (mAP), and FPS, with respective values of 0.98, 0.95, 0.96, and 58. This improvement highlights the innovative design and optimization of DiffuCNN, especially in dealing with challenging visual tasks. In conclusion, this research provides a comprehensive evaluation and analysis of the DiffuCNN model, demonstrating how innovative technical components and algorithms can enhance the performance of deep learning models in complex tobacco lesion counting tasks. The combination of these techniques and methodologies offers an effective means for solving practical problems and also directs future research in the field of deep learning.

Author Contributions: Conceptualization, H.X., Y.S. and S.Y.; Methodology, H.X., X.G. and N.Z.; Software, X.G. and W.T.; Validation, H.X., N.Z. and Y.C.; Formal analysis, X.G., Y.Y. and Y.J.; Investigation, W.T. and Y.J.; Resources, H.H. and W.T.; Data curation, N.Z., H.H., Y.Y., Y.C. and Y.J.; Writing—original draft, H.X., X.G., N.Z., H.H., W.T., Y.Y., Y.C., Y.J. and Y.S.; Writing—review and editing, Y.Y. and Y.S.; Visualization, H.H.; Supervision, S.Y.; Project administration, Y.S. and S.Y.; Funding acquisition, S.Y. All authors have read and agreed to the published version of this manuscript.

Funding: Major Project of China National Tobacco Corporation (110202201020(LS-04)).

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

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

References

1. Tufail, M.; Iqbal, J.; Tiwana, M.I.; Alam, M.S.; Khan, Z.A.; Khan, M.T. Identification of tobacco crop based on machine learning for a precision agricultural sprayer. *IEEE Access* **2021**, *9*, 23814–23825. [[CrossRef](#)]
2. Yang, L.; Liu, S.Q.; Liang, Y.; Liu, A.; Yang, Z. Population dynamics of main tobacco pests in the field and management suggestions. *Int. J. Pest Manag.* **2020**, *66*, 40–47. [[CrossRef](#)]
3. Bareschino, P.; Marrasso, E.; Roselli, C. Tobacco stalks as a sustainable energy source in civil sector: Assessment of techno-economic and environmental potential. *Renew. Energy* **2021**, *175*, 373–390. [[CrossRef](#)]
4. Thimmegowda, T.G.M.; Jayaramaiah, C. Cluster-based segmentation for tobacco plant detection and classification. *Bull. Electr. Eng. Inform.* **2023**, *12*, 75–85. [[CrossRef](#)]

5. Qin, X.; Zhao, X.; Huang, S.; Deng, J.; Li, X.; Luo, Z.; Zhang, Y. Pest management via endophytic colonization of tobacco seedlings by the insect fungal pathogen *Beauveria bassiana*. *Pest Manag. Sci.* **2021**, *77*, 2007–2018. [CrossRef] [PubMed]
6. Damayanti, F.; Muntasa, A.; Herawati, S.; Yusuf, M.; Rachmad, A. Identification of Madura tobacco leaf disease using gray-level Co-occurrence matrix, color moments and Naïve Bayes. *J. Phys. Conf. Ser.* **2020**, *1477*, 052054. [CrossRef]
7. Xu, M.; Li, L.; Cheng, L.; Zhao, H.; Wu, J.; Wang, X.; Li, H.; Liu, J. Tobacco Leaves Disease Identification and Spot Segmentation Based on the Improved ORB Algorithm. *Sci. Program.* **2022**, *2022*, 4285045. [CrossRef]
8. Chen, Y.M.; Zu, X.P.; Li, D. Identification of proteins of Tobacco mosaic virus by using a method of feature extraction. *Front. Genet.* **2020**, *11*, 569100. [CrossRef]
9. Sakhamuri, S.; Kompalli, V.S. An overview on prediction of plant leaves disease using image processing techniques. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *981*, 022024. [CrossRef]
10. Zhang, Y.; Wa, S.; Liu, Y.; Zhou, X.; Sun, P.; Ma, Q. High-accuracy detection of maize leaf diseases CNN based on multi-pathway activation function module. *Remote Sens.* **2021**, *13*, 4218. [CrossRef]
11. Zhang, Y.; He, S.; Wa, S.; Zong, Z.; Lin, J.; Fan, D.; Fu, J.; Lv, C. Symmetry GAN Detection Network: An Automatic One-Stage High-Accuracy Detection Network for Various Types of Lesions on CT Images. *Symmetry* **2022**, *14*, 234. [CrossRef]
12. Lin, X.; Wa, S.; Zhang, Y.; Ma, Q. A dilated segmentation network with the morphological correction method in farming area image Series. *Remote Sens.* **2022**, *14*, 1771. [CrossRef]
13. Zhou, X.; Chen, S.; Ren, Y.; Zhang, Y.; Fu, J.; Fan, D.; Lin, J.; Wang, Q. Atrous Pyramid GAN Segmentation Network for Fish Images with High Performance. *Electronics* **2022**, *11*, 911. [CrossRef]
14. Mappin-Kasirer, B.; Pan, H.; Lewington, S.; Kizza, J.; Gray, R.; Clarke, R.; Peto, R. Tobacco smoking and the risk of Parkinson disease: A 65-year follow-up of 30,000 male British doctors. *Neurology* **2020**, *94*, e2132–e2138. [CrossRef] [PubMed]
15. Lin, J.; Chen, Y.; Pan, R.; Cao, T.; Cai, J.; Yu, D.; Chi, X.; Cernava, T.; Zhang, X.; Chen, X. CAMFFNet: A novel convolutional neural network model for tobacco disease image recognition. *Comput. Electron. Agric.* **2022**, *202*, 107390. [CrossRef]
16. Swasono, D.I.; Tjandrasa, H.; Fathicah, C. Classification of Tobacco Leaf Pests Using VGG16 Transfer Learning. In Proceedings of the 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 18 July 2019; pp. 176–181. [CrossRef]
17. Dasari, S.K.; Prasad, V.K. A novel and proposed comprehensive methodology using deep convolutional neural networks for flue cured tobacco leaves classification. *Int. J. Inf. Technol.* **2018**, *11*, 107–117. [CrossRef]
18. Wu, J.; Yang, S.X. Modeling of the bulk tobacco flue-curing process using a deep learning-based method. *IEEE Access* **2021**, *9*, 140424–140436. [CrossRef]
19. Wang, D.; Zhao, F.; Wang, R.; Guo, J.; Zhang, C.; Liu, H.; Wang, Y.; Zong, G.; Zhao, L.; Feng, W. A Lightweight convolutional neural network for nicotine prediction in tobacco by near-infrared spectroscopy. *Front. Plant Sci.* **2023**, *14*, 1138693. [CrossRef]
20. Li, J.; Xu, Y.; Li, Y.; Qi, K.; Yu, F.; Sun, S. Research on Intelligent Recognition Solution of Tobacco Disease on Android Platform. In Proceedings of the 2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE), Wuhan, China, 16–17 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–4.
21. He, Z.; Chen, G.; Zhang, Y.; Zhao, C.; He, P.; Shi, B. Pyramid feature fusion through shifted window self-attention for tobacco leaf classification. *Expert Syst. Appl.* **2023**, *230*, 120601. [CrossRef]
22. Pant, K.; Yanamandra, V.H.; Debnath, A.; Mamidi, R. Smokeng: Towards fine-grained classification of tobacco-related social media text. *arXiv* **2019**, arXiv:1910.05598.
23. Borhani, Y.; Khoramdel, J.; Najafi, E. A deep learning based approach for automated plant disease classification using vision transformer. *Sci. Rep.* **2022**, *12*, 11554. [CrossRef]
24. Farsiu, S.; Robinson, M.D.; Elad, M.; Milanfar, P. Fast and robust multiframe super resolution. *IEEE Trans. Image Process.* **2004**, *13*, 1327–1344. [CrossRef] [PubMed]
25. Park, S.C.; Park, M.K.; Kang, M.G. Super-resolution image reconstruction: A technical overview. *IEEE Signal Process. Mag.* **2003**, *20*, 21–36. [CrossRef]
26. Glasner, D.; Bagon, S.; Irani, M. Super-resolution from a single image. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 349–356.
27. Zhang, L.; Zhang, W.; Lu, G.; Yang, P.; Rao, Z. Feature-level interpolation-based GAN for image super-resolution. *Pers. Ubiquitous Comput.* **2022**, *26*, 995–1010. [CrossRef]
28. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*; NeurIPS: New Orleans, LA, USA, 2014; Volume 27.
29. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
30. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*; NeurIPS: New Orleans, LA, USA, 2012; Volume 25.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
32. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2016**, arXiv:1511.06434. Available online: <http://xxx.lanl.gov/abs/1511.06434> (accessed on 3 March 2017).
33. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [CrossRef]

34. Epanechnikov, V.A. Non-parametric estimation of a multivariate probability density. *Theory Probab. Its Appl.* **1969**, *14*, 153–158. [[CrossRef](#)]
35. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
36. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14; Springer: Berlin/Heidelberg, Germany, 2016, pp. 21–37.
37. Zhang, Y.; Zhou, D.; Chen, S.; Gao, S.; Ma, Y. Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 589–597.
38. Li, Y.; Zhang, X.; Chen, D. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1091–1100.
39. Kang, G.; Jiang, L.; Yang, Y.; Hauptmann, A.G. Contrastive adaptation network for unsupervised domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4893–4902.
40. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems 28*; NeurIPS: New Orleans, LA, USA, 2015; Volume 28.
41. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
42. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.

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.