ISSN 2616-5775 Vol. 6, Issue 12: 70-78, DOI: 10.25236/AJCIS.2023.061208

Research on Grape Leaf Disease and Pest Detection Method Based on Style Transfer Assistance

Qi Li, Xiaohui Pang*

Shaanxi University of Science and Technology, Xi'an, China
sustpxh@163.com
**Corresponding author*

Abstract: *Grape cultivation is affected by a variety of diseases and pests. Addressing the issues of insufficient labeled samples in existing grape leaf disease and pest identification models and limited generalization capability of the networks, this paper proposes a grape leaf disease and pest detection method assisted by image style transfer. During the preprocessing stage, gamma correction is used to reduce the impact of lighting variations in the real environment on grape leaf image detection. In the data augmentation phase, an improved cycle generative adversarial network (CycleGAN) is employed, adding a category loss to incorporate category labels during the generator's training process. Category label information for grape diseases and pests, along with images created by the generator, are input into the discriminator. This converts samples of original healthy grape leaves from various growth stages into samples that represent affected regions, enriching the data augmentation for image samples input into the detection network and increasing the variety of the samples. In the grape leaf image detection process, a visual attention mechanism module is introduced into the Backbone structure of YOLOv8 to optimize the original feature extraction network. This module assigns and dynamically adjusts different attention weights to the diseased and pest-infested areas, thus augmenting the salience of grape leaf disease features against the complete background of the leaf. Experimental results show that the improved detection algorithm achieves a 4% increase in accuracy and a 7.1% increase in recall rate on the test set for grape leaf diseases and pests compared to the baseline algorithms, providing a foundation for the selection of pesticide usage in the development of smart agriculture.*

Keywords: *Grape Leaf Pests and Diseases, Light Correction, Style Transfer, Attention Mechanism, Yolov8 Algorithm*

1. Introduction

Currently, grapes have become an important crop among agricultural products and are widely planted in many countries and regions around the world. In China, the planting area and yield of grapes are both quite considerable. According to the 2023 Global Wine Report data released by the International Organization of Vine and Wine (OIV), as of 2022, there is a global grape planting area of 7.3 million hectares. During the cultivation of grapes, grape diseases and pests are one of the significant challenges faced, and the prevention and control of grape diseases and pests have a significant impact on grape yield and quality.

In recent years, with the continuous development of deep learning technology, automated detection methods for pests and diseases based on computer vision have been widely studied and applied. Xin He et al. adopted the Multi-Scale ResNet method to identify grape leaf diseases, employing a novel approach to explore the fine-grained recognition problem of crop diseases. They used MaskR-CNN to obtain leaf information and applied multi-scale convolution to enhance the identification rate, ultimately achieving an accuracy of 90.83%[1]. Wenjuan Guo et al. have employed the Squeeze-and-Excitation Networks, Efficient Channel Attention, and Convolutional Block Attention Module to introduce attention mechanisms into the Fast Region-based Convolutional Neural Network, YOLOx, and Single Shot Multibox Detector, enhancing vital features and weakening irrelevant ones to maintain the model's real-time performance and improve the model's detection accuracy[2]. Ji M et al. proposed a United Model based on a comprehensive method that integrates multiple convolutional neural networks architectures, which by combining multiple CNNs, is capable of extracting complementary discriminative features for the recognition of four types of grape leaf diseases, achieving an accuracy rate of 99.17%[3]. However, the dataset used in the experiment has a uniform background, which means that this method is not suitable

ISSN 2616-5775 Vol. 6, Issue 12: 70-78, DOI: 10.25236/AJCIS.2023.061208

for the detection of diseases and pests with complex backgrounds. Although progress has been made in the detection of grape diseases and pests using deep learning, there are still challenges and issues to contend with. Due to the diversity of grape diseases and pests, current studies often involve images with single-background and individual diseased leaves, which to some extent limits the network's generalization ability. Different grape varieties may exhibit different disease and pest characteristics due to varying growth environments and stages, leading to insufficient labeled samples. Thus, optimizing network structures and datasets to achieve high-quality, accurate detection of uneven grape leaf disease samples poses a more challenging task for the current technology.

To address the issues mentioned above, this study takes three common grape leaf diseases – black rot, black measles, and leaf blight – as the research subjects. Building on the standard YOLOv8 algorithm, it proposes a grape leaf disease and pest detection method aided by image style transfer to enhance the model's generalizability under complex backgrounds. The method first employs gamma correction for light preprocessing before feeding images into the detection network, reducing the impact of lighting variations in actual grape plantation environments on the detection of grape leaf images. Addressing the complexity of actual grape plantation scenes and the shortage of training samples for diseases and pests in different varieties at different growth stages, an improved CycleGAN is utilized. By adding a category loss and incorporating category labels during the training process of the generator, the model feeds both generated images and grape disease and pest category label information into the discriminator. It transforms original healthy grape leaf samples at various growth stages into samples of affected areas, augmenting the input data for the detection network and increasing sample diversity. To improve the accuracy of identifying and distinguishing between affected and healthy regions on grape leaves, a visual attention mechanism module is introduced into the backbone structure of YOLOv8. This enhancement optimizes the existing feature extraction network by assigning and dynamically adjusting different attention weights to pest and disease areas, thereby increasing the prominence of grape leaf disease features against the complete leaf backgrounds.

2. Correlation Theory

2.1 Overview of YOLOv8

The YOLO (You Only Look Once) object detection algorithm is an end-to-end, single-stage detection algorithm predicated on the division of an image into multiple grids, with predictions for object detection being made in each individual grid. In this paper, the latest YOLOv8 model from Ultralytics, which was released in January 2023, is employed. It is built upon the successful foundation of previous YOLO versions and introduces new features and improvements to further enhance performance and flexibility. YOLOv8 introduces a state-of-the-art (SOTA) architecture, offering object detection networks with resolutions of P5640 and P61280, in addition to a YOLACT-based instance segmentation model. In comparison to YOLOv5, the YOLOv8 algorithm substitutes the C3 structure in the backbone and neck with a more gradient-rich C2f structure, and it tailors the channel count to different scale models instead of applying a uniform parameter set across all models. This considerably enhances the algorithm's efficiency. In the head portion, YOLOv8 has adopted the contemporary decoupled head structure, segregating the classification and detection heads and transitioning from Anchor-Based to Anchor-Free methodologies. With respect to loss computation, YOLOv8 employs the Task Aligned Assigner strategy for the assignment of positive samples and integrates the Distribution Focal Loss. The training data augmentation incorporates the operation of disabling the Mosaic augmentation in the last 10 epochs, borrowed from YOLOX, resulting in an effective improvement in accuracy.

2.2 Overview of Image Style Transfer

Image Style Transfer, a subdivision within the realm of computer vision, represents an image processing methodology whereby the semantic content of an image is rendered through divergent styles. This technique allows for the transposition of one image's style onto another, thereby engendering images adorned with variegated artistic styles, coloration methods, or textural attributes [4-5]. Early methods of image style transfer primarily relied on optimization techniques, such as minimizing a loss function to match the content and style features of an image. Classic methods for image style transfer include "Neural Style Transfer" proposed by Gatys et al., which utilizes convolutional neural networks to extract image features and achieves image style transfer by minimizing both content loss and style loss[6-7]. With the emergence of Generative Adversarial Networks (GAN), image style transfer has been significantly improved. CycleGAN, proposed by ZhuJY et al. as a classic style transfer algorithm, uses two

ISSN 2616-5775 Vol. 6, Issue 12: 70-78, DOI: 10.25236/AJCIS.2023.061208

unidirectional GANs to form a cyclic GAN, which retains key information in the images and solves the problem of requiring paired training data in image style transfer[8]. It has strong application value in fields such as film and television production and style design. The architecture of the CycleGAN algorithm is illustrated in Figure 1.

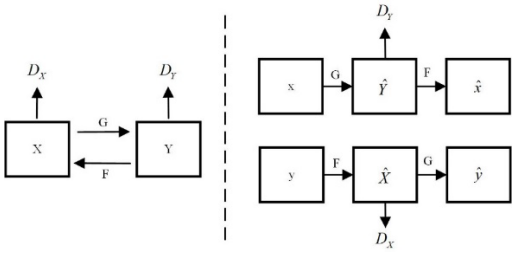


Figure 1: CycleGAN algorithm structure

3. Grape Pest and Disease Detection Algorithm Based on Style Transfer Assistance

The focus of this study revolves around the disease and pest infestation of grape leaves throughout their entire lifecycle. The analysis primarily examines three prevalent grape leaf diseases: black rot, black measles, and leaf blight. The datasets for each of these grape leaf diseases, as well as for healthy leaves, are depicted in Figure 2.

（a）Black Rot　（b） Black Measles　（c） Leaf Blight　（d）Healthy Leaf

Figure 2: Data set of each grape leaf disease and healthy leaf

3.1 Experimental Data Acquisition

This study is applied to intelligent pesticide spraying vehicles in vineyards, and the dataset for detecting targets consists of two parts. The first part includes 1,000 images each of grape black rot, black measles, and leaf blight publicly available from Baidu Paddle Paddle; the second part comprises 3,000 images of healthy grape leaves collected from the field at various growth stages. After data augmentation, a total of 2,000 images each for black rot, black measles, and leaf blight, as well as 2,000 images of healthy grape leaves, were input into the target detection network. These images were divided into a training set of 5,600 images, a test set of 1,600 images, and a validation set of 800 images, based on a ratio of 7:2:1. The image size was scaled to 256×256, and the Imglabel software was used to annotate the regions with grape leaf diseases in the images.

3.2 Adaptive Light and Dark Light Image Preprocessing Based on γ Correction

During the detection of grape leaf images, the complex background of the vineyard environment and the uneven lighting conditions in the captured grape leaf images result in problems such as variations in brightness, contrast, and color in the grape leaf images that need to be detected. This leads to an increased rate of false detection of grape leaf diseases and pests by the model. To better extract image features and address this challenge, γ (gamma) correction is used to preprocess the input images for lighting conditions.

Gamma correction is a common image processing technique that modifies an image's brightness and contrast by adjusting its gamma value, making the image clearer and more vivid. The higher the gamma value, the lower the brightness and the higher the contrast of the image. Before applying gamma correction, image pixel intensities must be scaled to the range of [0, 1.0] to ensure the corrected image results are accurate and stable. Gamma correction is also known as power-law transformation and can be utilized in various image processing scenarios, such as image enhancement, color correction, and medical image processing, among others. In gamma correction, the transformation of pixel values follows a formula expressed as Formula 1.

ISSN 2616-5775 Vol. 6, Issue 12: 70-78, DOI: 10.25236/AJCIS.2023.061208

$$V_{out} = V_{in}^{\gamma} \tag{1}$$

In the formula, V_{in} is the pixel value of the input image, and V_{out} is the pixel value of the output image. γ is the correction parameter, which is usually between 0.1 to 1.0.

The larger the value of gamma, the more the details in the dark areas of the image will be enhanced, and the smaller the value of gamma, the more the details in the bright areas of the image will be enhanced. As can be observed from the comparison in Figure 3, gamma correction can normalize grape leaf images with varying brightness to a balanced state, thus enhancing the accuracy of the models in detecting diseases and pests in grape leaves.

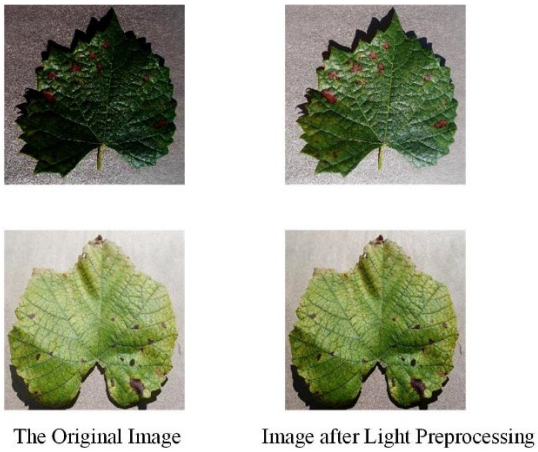


Figure 3: Grape leaf images before and after light pretreatment

3.3 Style Transfer Based on the Improved Cyclic Generative Adversarial Network (CycleGAN)

The number of image samples of grape leaves with diseases and pests is limited. Before inputting the dataset into the detection network, the images are subjected to transformations such as translation, rotation, scaling, and flipping to generate multiple images with different perspectives and variations, thereby increasing the diversity of the samples. This improves the training effect of the target detection algorithm and enhances the robustness of the model. However, traditional data augmentation methods have the drawback of reducing data correlation, which may make it difficult for models to generalize in real-world scenarios. Inspired by image style transfer techniques, this paper adopts CycleGAN to perform data augmentation on the image samples input into the detection network. CycleGAN mainly transfers color and texture changes, but due to the complex background variation in photos taken by cameras in actual grape plantations, variations in grape leaf sizes, random occlusions, and other issues, it can easily confuse background with target features. Therefore, this paper improves the original loss functions by adding a category loss on top of the existing adversarial loss and cycle consistency loss. During the training process of the generators, the semantic meaning of the task is emphasized by introducing category labels into the discriminator along with the generated images, enhancing the precision of the transformation from healthy leaves to diseased leaves.

The style transfer model in this paper consists of two generators (G, F) and two adversarial discriminators (DX, DY). The generator that converts images of healthy grape leaves to images of grape leaves with diseases and pests is called G, and the one that converts images from diseased to healthy leaves is called F. The discriminator that determines whether an image is of a healthy grape leaf is called DX, and the one that determines if an image is of a grape leaf with diseases and pests is called DY.

The generators G and F consist of an encoder and decoder. The encoder uses three convolutional layers to extract different features of grape leaves, followed by six residual blocks that further extract image information while preserving the input data features. The decoder then performs upsampling using two transposed convolutions, followed by a convolutional layer, and the resulting image matrix is activated by the Tanh function to produce the final output image. Instance normalization is employed during the training of the generators to normalize the features and weights extracted from each image, ensuring that the generators produce more stable feature representations for each instance. The generative network of the style transfer model is shown in Figure 4. The adversarial discriminators DX and DY use a PatchGAN structure to assess the difference between the generated images and the real images of the target domain.

ISSN 2616-5775 Vol. 6, Issue 12: 70-78, DOI: 10.25236/AJCIS.2023.061208

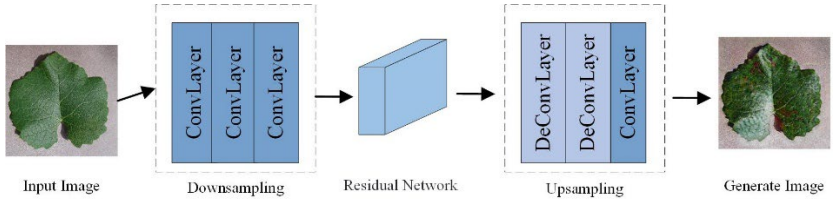


Figure 4: Generation network of style migration model

The original loss function of CycleGAN is composed of an adversarial loss and a cycle consistency loss. The adversarial loss is a binary classification loss based on discriminators, which primarily controls the generator to produce images that are more realistic to the target domain. The cycle consistency loss uses symmetrical cyclic constraints to maintain the consistency between the original image and the regenerated (cyclic) image, guiding the generator to produce images as close as possible to the input image. For grape leaves, thoroughly learning the detail features of the diseased areas and the background characteristics is key to generating high-quality images. To better learn local features of the input image, a category loss is introduced during the training of generators G and F, which brings in category labels. This allows generated images to be trained under the target domain's supervisory signals and submits them to discriminators DX and DY along with the generated images. When training discriminators, the real images are trained under the supervisory signal of the source domain, and the cross-entropy loss function is used to calculate the difference between the category predictions of the generated images and the real category of the target domain images. The generator minimizes the cross-entropy loss to improve the categorical accuracy of the images. The category loss is shown in Equation 2.

$$L_{\text{class}} = -\sum_{c=1}^{C} y_c \log(p_c) \tag{2}$$

Herein, C is the number of categories of grape leaf diseases and pests, y_c represents the value of the c-th category, and p_c represents the probability of the c-th category as predicted by the model.

Therefore, the total loss function of the network is represented by Equation 3.

$$LOSS=L_{GAN}(G,D_Y,X,Y)+L_{GAN}(F,D_X,Y,X)+\lambda_1 L_{cycle}(G,F)+\lambda_2 L_{class} \tag{3}$$

λ_1 and λ_2 are non-negative hyperparameters that are used to adjust the different impacts of loss on the overall effect.

After training, this paper transformed 3000 images of healthy grape leaves into 1000 images each with features indicative of black rot, black measles, and leaf blight. The grape leaf images augmented with disease and pest features through image style transfer are presented in Figure 5.

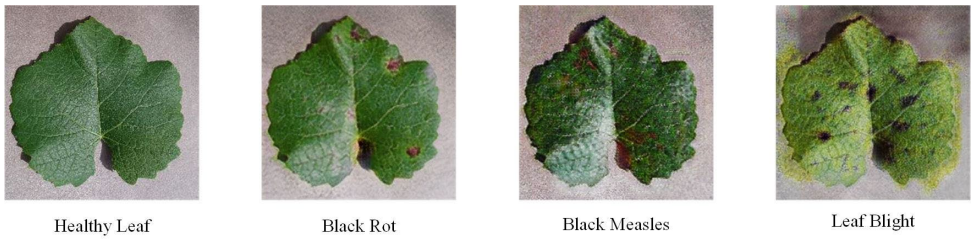


Figure 5: Image style transfer expanded grape leaf pests and diseases image

3.4 Visual Attention Module is Introduced in YOLOv8 Backbone Network

In actual grape vineyards, distinguishing the texture and color information of disease spots on diseased grape leaves is challenging due to the complex and irregular backgrounds of the leaves. Some diseased grape leaves may only have a few spots, making direct detection quite difficult.To enhance the detection capability of disease spots on diseased grape leaves, a visual attention mechanism is introduced into the backbone structure of the YOLOv8 model, assigning and dynamically adjusting different attention weights to the pest-affected areas, saving computational resources, enhancing the algorithm's feature extraction of disease targets, and strengthening the model's learning capacity.

In the YOLOv8 model, the convolutional layers in the backbone structure are responsible for extracting low-level and mid-level features from the image. By adding a Spatial Attention module, which

ISSN 2616-5775 Vol. 6, Issue 12: 70-78, DOI: 10.25236/AJCIS.2023.061208

uses positional weights to produce a spatial attention feature map, the model can focus on regions of the leaf that may have pests or diseases, emphasizing features related to those regions. As the depth of the network increases, the feature extraction capability of the backbone network becomes insufficient, leading to an inability to effectively integrate high-quality contextual information, thus reducing the model's detection precision. The Non-local module adjusts the importance of different positions by computing the similarity between various positions in the input image. This helps the model to capture global contextual information in the image by modeling long-range dependencies between pixels and locating abnormal areas on the leaf, thereby enhancing the accuracy of pest and disease detection. The improved YOLOv8 algorithm framework is depicted in Figure 6.

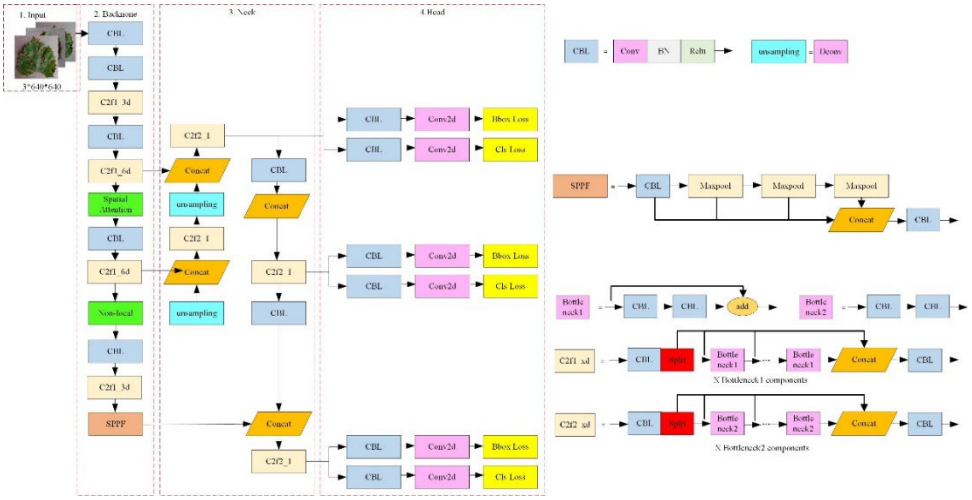


Figure 6: Improved yolov8 algorithm framework

The Spatial Attention module is a mechanism for attention that is used for computer vision and image processing tasks. It adjusts the importance of each pixel in an input image by learning positional weights [9]. By performing element-wise multiplication of the positional weights and the original feature map, the module can focus its attention on key areas of the image, particularly those areas that are relevant to the task. In order to calculate the attention weights, the input feature map undergoes a convolutional operation that transforms it into a lower-dimensional tensor, reducing the number of channels to one to obtain positional weights. Subsequently, a convolutional layer further transforms the converted feature map to obtain the corresponding weights for each position, as shown in Equation 4, where X is the transformed feature map, and f represents the convolutional operation. To ensure that the weight values fall within a reasonable range, a softmax operation is performed on the weights to ensure that each position's weight is between 0 and 1 and that the sum is 1. The normalization process is seen in Equation 5, where W is the weight vector. Finally, the normalized weights are element-wise multiplied with the original feature map to produce the weighted feature map. The feature weighting process is shown in Equation 6, where X is the original feature map, and Y is the feature map after weighting.

$$W = f(X) \tag{4}$$

$$soft\max(W) = \frac{\exp(W)}{\sum_i \exp(W_i)} \tag{5}$$

$$Y = X \odot soft\max(W) \tag{6}$$

Non-local is a module or method capable of capturing long-range dependencies between pixels in an image by directly capturing remote dependencies through computing interactions between any two positions, constructing a convolutional kernel with the same size as the feature map to capture the global interaction and contextual information between pixels, achieving non-local interactions among pixels. This global dependency is beneficial for handling inter-object relationships, motion representation across large spans, etc.[10]. Its representation is shown in Equation 7.

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \tag{7}$$

In this context, x_i represents a position in the input feature map, where i denotes the output position, such as spatial, temporal, or spatio-temporal index. Its response should be computed by

ISSN 2616-5775 Vol. 6, Issue 12: 70-78, DOI: 10.25236/AJCIS.2023.061208

enumerating over j. The functions f, g are used to compute similarity and fusion respectively. The term $C(x)$ is a normalization factor used to scale the similarity coefficient. The final output y is obtained after normalizing with the response factor $C(x)$.

4. Experimental Results and Analysis

4.1 Experimental Environment

The algorithm presented in this paper is implemented within the deep learning framework, Pytorch, utilizing hardware comprising a single NVIDIA GeForce RTX 2080Ti GPU with 16GB of memory, and running on the Ubuntu18.04 operating system. The algorithm development is based on the Python language. The network model parameters used in the experiments in this paper are as follows: Batch_size is set to 64, the weight decay factor is 0.0001, the number of training iterations is set to 300, and the initial learning rate is 0.001.

4.2 Evaluation Index

The evaluation metrics adopted in the experiments of this paper for the detection of grape leaf diseases and pests include detection accuracy (the proportion of samples that are correctly predicted as positive out of the total samples predicted as positive, Precision, P), mean Average Precision (mAP), and recall rate (the proportion of the true positive samples correctly predicted out of the original samples, Recall, R).

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{9}$$

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \tag{10}$$

Where TP (True Positives) represents the number of images that have disease features and the disease type is correctly identified, FP (False Positives) represents the number of images that have disease features but the disease type is misidentified, and FN (False Negatives) represents the number of images that actually have disease features but are not recognized as having a disease.

4.3 Analysis of Experimental Results

To facilitate the analysis of the improvement effects, the paper designs ablation studies to verify the effectiveness of each enhancement, with the results shown in Table 1.

Table 1: Ablation study

Serial number	Image ray preprocessing	Style transfer model	Visual attention module	mAP/%
1	×	×	×	89.3
2	√	×	×	89.8
3	×	√	×	91.6
4	×	×	√	91.4
5	√	√	×	92.3
6	√	×	√	91.7
7	×	√	√	95.6
8	√	√	√	96.4

Using the grape leaf diseases and pests dataset employed in this paper, the algorithm presented herein was compared with mainstream detection network models such as Faster R-CNN and YOLOv5, with the results displayed in Figure 7.

To verify that preprocessing the lighting of grape leaf images can improve the detection rate of diseased grape leaves, this paper selected grape leaf images taken under the poor lighting conditions of a cloudy day. The comparison results of detection before and after light preprocessing are shown in

ISSN 2616-5775 Vol. 6, Issue 12: 70-78, DOI: 10.25236/AJCIS.2023.061208

Figure 8, with an average increase of about 3% in the detection confidence for grape diseases.

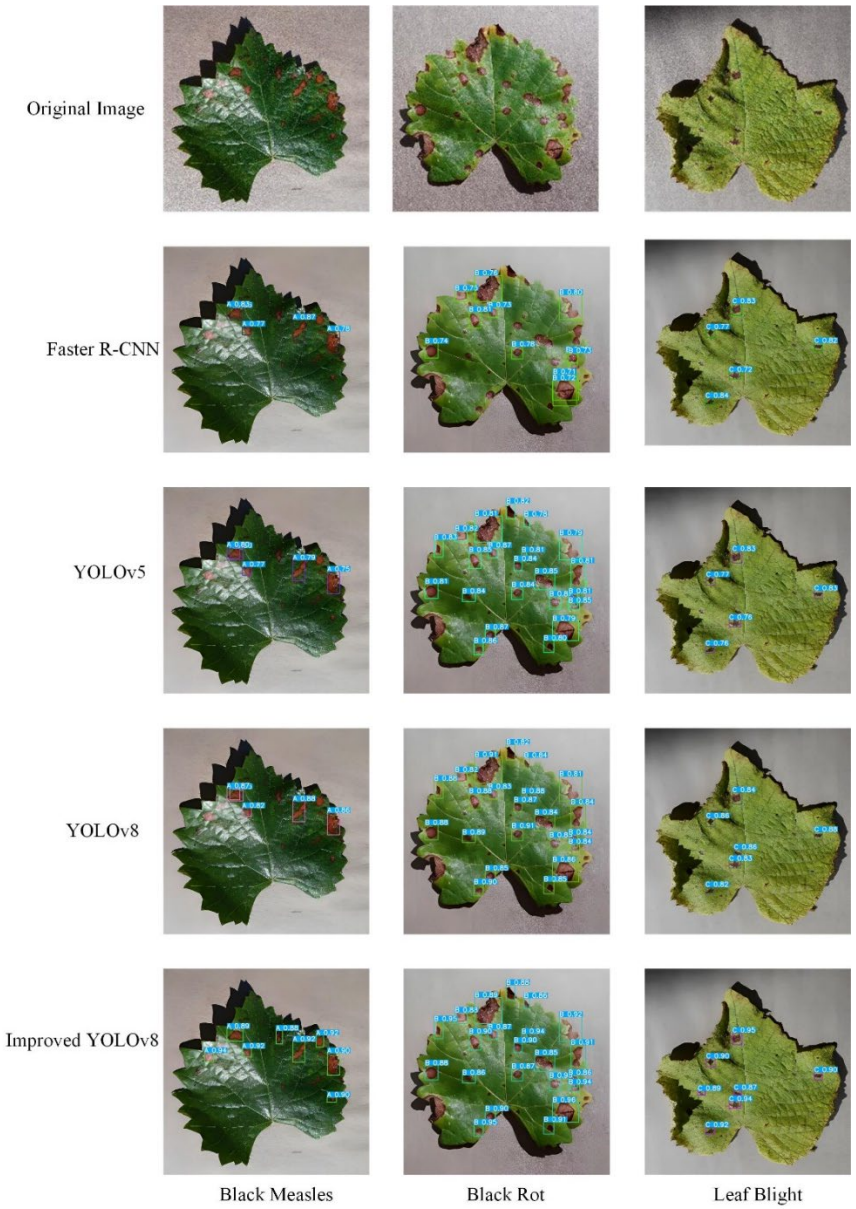


Figure 7: Detection results of grape pests and diseases by different algorithms

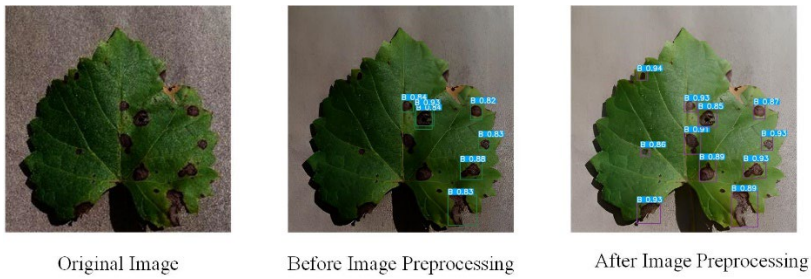
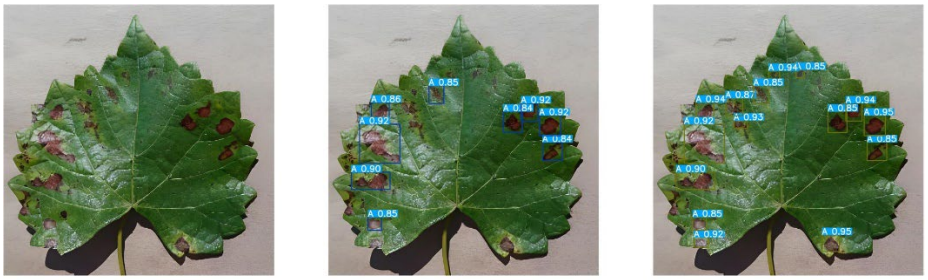


Figure 8: Detection results of grape leaf pests and diseases before and after image preprocessing

In order to prove that using style transfer to augment dataset samples can improve the detection rate of diseased grape leaves, this paper compared it with models created using traditional data augmentation techniques for images, such as translation, rotation, scaling, and flipping. The comparison results are demonstrated in Figure 9, with an average increase of approximately 4% in the detection confidence for grape diseases.

ISSN 2616-5775 Vol. 6, Issue 12: 70-78, DOI: 10.25236/AJCIS.2023.061208



Original Image | Traditional Data Enhancement Methods | Improved Data Enhancement Methods

Figure 9: Detection results of grape leaf pests and diseases by different data enhancement methods

5. Conclusion

This paper investigates the detection of three common diseases and pests affecting grape leaves, including black rot, black measles, and leaf blight. By improving the cycle generative adversarial network and combining publicly available images of black rot, black measles, and leaf blight from Baidu PaddlePaddle with images of healthy grape leaves collected in the field from various growth stages, a comprehensive grape disease and pest target dataset was constructed. With modifications to the YOLOv8 algorithm, a visual attention mechanism was introduced in the Backbone structure, giving and dynamically adjusting different attention weights to diseased and pest-ridden regions, thereby enhancing the feature extraction capabilities of the algorithm for disease targets. The experimental results show that compared to the YOLOv8 algorithm, the improved network model has better adaptability under different conditions of grape leaf diseases, achieving an average precision of 96.4% for grape leaf disease and pest detection and an average detection speed of 30 FPS. This meets the real-time detection requirements of intelligent pesticide sprayers in vineyards. Future work can focus on the grading research of the severity of grape leaf diseases and pests, providing an optimal solution for the selection of pesticide dosage for smart pesticide sprayers.

References

[1] He, X., Li, S., & Liu, B. Grape Leaf Disease Recognition Based on Multi-Scale Residual Neural Network [J]. Computer Engineering, 2021, 47(5), 285-291, 300.

[2] Guo, W., Feng, Q., Li, X., Yang, S., & Yang, J. Grape leaf disease detection based on attention mechanisms [J]. International Journal of Agricultural and Biological Engineering, 2022, 15(5):8.

[3] Ji, M., Zhang, L., & Wu, Q. Automatic grape leaf diseases identification via United Model based on multiple convolutional neural networks [J].Information Processing in Agriculture, 2020, 7(3):418-426.

[4] Y. Liu, F. E. T. Munsayac, N. T. Bugtai and R. G. Baldovino. Image Style Transfer with Feature Extraction Algorithm using Deep Learning[C]. 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM),2021, pp. 1-5.

[5] Zhang, Y., Tian, Y., & Hou, J.Csast: content self-supervised and style contrastive learning for arbitrary style transfer. Neural Networks: The Official Journal of the International Neural Network Society, 2023, 164: 146-155.

[6] Gatys L A, Ecker A S, Bethge M.A neural algorithm of artistic style [J]. Journal of Vision, 2015.

[7] Gatys, L. A, Ecker, A. S., & Bethge, M. Image Style Transfer Using Convolutional Neural Networks[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016,2414-2423.

[8] Zhu J Y, Park T, Isola P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[C]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242-2251.

[9] Woo, S., Park, J., Lee, J., & Kweon, I. (2018). CBAM: Convolutional Block Attention Module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018, 3-19.

[10] Wang, X., Girshick, R., Gupta, A., & He, K. Non-local neural networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 7794-7803.