



GLAD: Advanced Attention Mechanism-Based Model for Grape Leaf Disease Detection

Venkata Nagaraju Thatha¹, Polukonda Mary Kamala Kumari², Uddagiri Sirisha³, Valisetty Venkata Ram Manoj⁴, Surapaneni Phani Praveen^{3*}

¹ Department of Information Technology, MLR Institute of Technology, Hyderabad 500049, India

² Department of CSE, Lakireddy Balireddy College of Engineering, Mylavaram 521230, Andhra Pradesh, India

³ Department of CSE, P V P Siddhartha Institute of Technology, Vijayawada 520007, Andhra Pradesh, India

⁴ Department of CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada 520008, Andhra Pradesh, India

Corresponding Author Email: phani.0713@gmail.com

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290230>

ABSTRACT

Received: 7 September 2023

Revised: 14 January 2024

Accepted: 18 February 2024

Available online: 25 April 2024

Keywords:

Yolov5-You only look once version5, convolutional neural network, shuffle attention, convolutional block attention module, efficient channel attention, multi-channel attention, agricultural disease detection

Diseases affecting grape leaves can have a wide variety of symptoms and a complicated history in vineyards, making detection and diagnosis a formidable issue. The complexity of these problems is frequently too much for existing detection algorithms to handle. Hence, a new method called GLAD (Grape Leaf Disease Detection) was developed. GLAD makes use of the PLANT-VILLAGE dataset, which has been hand-picked to detect grape diseases. We added the self-attention mechanism to make it more effective, and it now can collect data on grape leaf illnesses all over the world better. Adaptively spatial feature fusion (ASFF) technology and BiFPN feature fusion network provide more robust models and improve grape leaf disease fusion by reducing complex background interference. The Shuffle Attention approach is also used to make identifying diseases in grape leaves easier. The dataset is enriched using data augmentation methodologies and transfer learning to identify diseases affecting grape leaves. As part of this process, the model's parameters are adjusted using data from other plant disease datasets. Despite several obstacles, the experimental findings show that the suggested model is intelligent enough to identify grape leaf disorders. Its real-time target detection capabilities are on full display when it outperforms state-of-the-art methods. A powerful and effective tool for the agricultural sector, GLAD is a major step forward in solving the problems associated with grape leaf disease identification.

1. INTRODUCTION

Grapes are an economically valuable and widely consumed fruit. Grape leaves are highly vulnerable to infections from bacteria, viruses, and fungi. The quality and quantity of grapes can be diminished if these diseases spread throughout the plant. Diseases on grape leaves used to be identified manually using standard phytopathology techniques. It's a lot of work and takes a long time, though. As more and more acreage is devoted to grape cultivation, the accuracy of manual identification methods declines. For the sake of grape production's continued growth, automatic identification of leaf diseases is essential. Automated identification techniques are more efficient, precise, and trustworthy than their manual counterparts. Also, unlike manual methods, they may be utilized to keep tabs on broad stretches of land.

Maintaining healthy crops, decreasing crop losses, and using as little pesticides as possible are all crucial in agriculture [1]. The key to successful disease prevention and early diagnosis is vigilance. Powdery mildew, brown blotch, and anthracnose are just a few of the diseases that can affect grape vines. The quantity and quality of fruit produced may be drastically affected by these illnesses. Detecting grape diseases via tried-and-true approaches like producers'

anecdotes or experts' advice can be time-consuming, inefficient, and behind the times. Grape leaf photos are utilized to detect, identify, and direct disease prevention efforts [2] to overcome these constraints. The leaves of sick grapes tend to be discolored and distorted.

Classifying and identifying crop diseases using traditional methods [3-5] requires laborious manual feature selection that is highly context dependent. However, deep learning's introduction of convolutional neural networks (CNNs) has radically altered the process of identifying agricultural diseases. A lot of work has been done to categorize agricultural illnesses using various methods [6-12]. By extracting high-dimensional features from photos of the sick leaves, CNNs can accurately diagnose crop leaf illnesses even in the presence of complicated backdrops.

Face recognition [13], navigation systems [14], obstacle identification [15], pedestrian detection [16], abnormal activity recognition [16], monitoring physical activity [17, 18], fruit detection [19], and weed detection [20] are just a few examples of the many uses for CNNs. In order to help farmers reduce crop failures and maximize yields, these technologies can be utilized to perform automatic identification of crop leaf diseases. Scientists from China and other nations have been investigating object detection algorithms to create crop disease

detection models. They have tried out various models using tomato disease datasets, including Faster R-CNN and Single Shot Multibox Detector. The most encouraging results in disease identification have been achieved by combining faster R-CNN with VGG16. Researchers have successfully used Faster R-CNN on time-series images of grape leaves for dynamic identification of grape leaf diseases. Using an improved Faster R-CNN model, the authors [20] were able to get good disease detection results in bitter melon leaves. The authors [21] also used an internal dataset to train the SSD (Single Shot Multibox Detector) model, which led to an 83.90% overall accuracy in detecting agricultural illnesses.

Several methods, such as an improved model based on MobileNetV2 and YOLOv3 [22], have been proposed by researchers to identify agricultural illnesses. This model's many strengths include its small memory footprint, high detection accuracy, and lightning-fast identification times. However, CNNs' limited perception is a significant limitation. Because of convolutional computation, CNNs often only extract characteristics from nearby regions of an image [23].

We propose GLAD, a deep learning model for grape leaf disease detection, to overcome this shortcoming. With GLAD, CNN-based object detection is improved over YOLOv5 [24]. GLAD can gain a comprehensive overview of the image by including the self-attention mechanism. This allows GLAD to ignore irrelevant noise and zero in on relevant regions, leading to more precise detections. The GLAD model is an advancement on the YOLOv5 deep learning model for detecting diseases in grape leaves. It facilitates the model's acquisition of more contextual information by creating a global perception eld [25]. Several novel elements are included in the GLAD model. They are

- As a result of BIFPN [26], grape leaf diseases are enhanced on multiscale features, and features become more prominent.
- The Shuffle Attention mechanism [27] enables the GLAD model to place greater emphasis on grape leaf diseases during the detection process.
- The ASFF [28] detection head automatically filters out irrelevant information and suppresses interference from complex backgrounds when detecting grape leaf diseases.
- Through transfer learning, the GLAD model can be accelerated in situations with limited sample sizes with an increase in accuracy and robustness.

2. RELEVANT LITERATURE

Presented Table 1 encapsulates various research studies have explored innovative grape leaf disease detection approaches, employing diverse techniques such as convolutional neural networks (CNNs), transfer learning, and attention mechanisms. Noteworthy models include DICNN, integrating dense connectivity for feature reuse; GLD-DTL, utilizing MobileNetV3 for accurate recognition; and GeT, a Ghost-convolution enlightened Transformer. These models showcase high accuracies ranging from 86.65% to 99.93%. Some studies focus on integrating attention mechanisms like SE, ECA, and CBAM into popular architectures like Faster R-CNN and YOLOx to enhance real-time detection precision. Additionally, techniques such as LoRa combined with CNN and GANs in conjunction with Faster R-CNN contribute to the diversity of methods explored for grape leaf disease identification. These studies contribute to advancing the eld by

addressing challenges related to disease presentation variance and complex background interference.

Our research aims to address the challenges general target detection models face in accurately identifying grape leaf disease targets. The main objectives of this research include:

- To overcome these issues encountered in the intelligent recognition of grape leaf diseases, we have developed an enhanced model called GLAD.
- This model incorporates several improvements that enhance the fusion of grape leaf disease features across different scales, prioritize the grape leaf disease areas, and improve detection accuracy for minor target grape leaf diseases.
- Additionally, it leverages global information to infer grape leaf diseases better and exhibits enhanced resilience against complex background interference during detection. Using various technologies, we have significantly improved the accuracy of intelligent grape leaf disease detection while achieving real-time detection speeds.
- Implementing our proposed model can enable prompt and precise detection of grape leaf diseases, replacing inefficient manual inspections. This facilitates the implementation of targeted measures for disease control, ultimately enhancing production efficiency and grape quality.

3. MATERIALS AND METHODS

3.1 Image acquisition

Detection and classification of grape diseases are made easier using the plant village dataset. Four thousand six hundred and twenty-two images show grape leaves with many symptoms, including black rot, esca measles, leaf spot, and healthy leaves. The dataset is imbalanced, with Black Rot being the largest category (1,180 images). Esca measles is the most common classification, followed by Leaf spot. The dataset also contains 423 images of healthy grape leaves for reference. The varying quantities of images indicate the prevalence of certain grape leaf diseases. Esca measles accounts for approximately 34% of the dataset, while Black Rot and Leaf spot represent 29% and 26%, respectively. Healthy images make up about 10.4% of the collection. The plant village dataset provides researchers with a valuable resource to train and evaluate grape disease detection models. It can be used to develop more accurate algorithms for automated diagnosis and classification of grape leaf conditions.



Figure 1. Sample images from dataset

Table 1. Prior works on grape leaf disease detection

Reference	Key Concept	Pre-Trained Models Used	Attention Mechanism Used
[29]	We are using transfer learning and convolutional neural networks to identify grape diseases.	AlexNet, GoogleNet and ResNet-18.	-
[30]	They are utilizing four cascaded Inception structures to extract multidimensional features.	Trained from Scratch (DICNN-dense Inception convolutional neural network)	-
[31]	The system compared AlexNet, GoogLeNet, and ResNet-18 to select the best pre-trained network for integrating with RCNN for multiple object detection in images.	AlexNet, GoogLeNet, and ResNet-18	-
[32]	A grape leaf spot identification method using ne-grained GANs was proposed to augment local spot area images before training deep learning models for better generalization.	Faster R-CNN +GAN	-
[33]	To train a DCNN model to detect and categorize grape illnesses using RGB images of the leaves.	Xception and Inception V3 training with ResNet50, VGG16, and VGG19	-
[34]	Using ResNet50 and ResNet101 to directly combine the Fc1000 deep feature. ResNet50 and ResNet10.	ResNet50 and ResNet10	-
[35]	The proposal for GLD-DTL, a technique for disease recognition in grape leaves using DTL, was made by introducing the rst framework of MobileNetV3. This method could accurately identify six prevalent grape leaf illnesses by mechanically extracting discriminative features from photos of sick grape leaves.	MobileNetV3 model	-
[36]	Based on the detection features of the four models, an MDF decision-making approach was developed and an MDF model was acquired.	ShuffleNet V2	-
[37]	In order to create the Ghost-convolution enlightened Transformer (GeT) model, we suggested merging Ghost-convolution with the Transformer encoder. In order to train its Transformer layers, the networks take feature maps generated by compact Ghost blocks.	Ghost-convolution and Transformer encoder	-
[38]	A grape leaf black rot detection system using super-resolution picture augmentation and convolutional neural networks.	YOLOv3-SPP network	No
[39]	A new network called CASM-AMFMNet (coordinated attention shuffle mechanism-asymmetric multiscale fusion module net) was used to detect diseases in grape leaves.	CoatNet	-
[40]	To improve real-time detection precision, strengthen essential features, and weaken unrelated features, this study introduced squeeze-and-excitation networks, efficient channel attention (ECA), and convolutional block attention modules into Faster Region-based Convolutional Neural Networks (R-CNN), YOLOx, and single shot multi-box detectors (SSD).	YOLOx,SSD,F-RCNN +Attention mechanisms	Yes
[41]	Residual blocks, RFFBs, and attention modules for convolutional blocks comprise the bulk of GrapeNet.	Layers of convolution, residual blocks, residual feature fusion blocks (RFFBs), and convolutional block attention modules (CBAMs)	Yes
[42]	This paper showcases the endings of a computer vision system that combines LoRa with Deep Learning. The system can efficiently identify illnesses in grape leaves using low-resolution photos.	Long Range (LoRa), CNN	-
[43]	Using the Inception-v1 module, Inception-ResNet-v2 module, and SE-blocks, a deep-learning-based Faster DR-IACNN model with better feature extraction is provided for grape leaf disease detection.	Faster R-CNN ,Resnet	Yes

Figure 1 shows a sample of the dataset. The images show different grape leaf diseases, including black-rot, esca measles, leaf-spot & healthy leaves.

3.2 Image pre-processing and augmentation

Table 2. Insights into the dataset

No of Images with Augmentation	No of Images Without Augmentation	Class
3000	423	healthy
3000	1383	esca
3000	1076	measles
3000	1180	leaf-spot
12000	4062	black-rot
		Total

Before training the model, data augmentation is performed

to improve the data and ensure consistency within the sample space. The following data augmentations are applied (Table 2):

- **Image rotation:** Images are randomly rotated by angles ranging from 90 to 270 degrees to capture different shooting perspectives.
- **Colour dithering:** The "chroma," "saturation," & "contrast" of the images are randomly adjusted to simulate various lighting conditions. This results in images with different color characteristics.
- **Sharpen processing:** This technique enhances the edge definition of the images, producing variations in the level of image sharpness.

These data augmentation strategies aim to diversify the dataset, enabling the model to learn from a wider range of visual variations. This improves the model's ability to generalize to different scenarios.

3.3 Proposed framework

The GLAD model is an improved version of the YOLOv5 model specifically designed for grape leaf disease detection. The GLAD model incorporates several new features, including:

- The Transformer module, which expands the model's global receptive field and allows it to gather more contextual information.
- The BIPFN feature fusion network improves multiscale feature fusion efficiency.
- Detecting minor target grape leaf diseases is improved through the Shuffle Attention (SA) mechanism.
- With the ASFF detection head, you can detect grape leaf diseases by automatically filtering out irrelevant information.

The GLAD model can now detect grape leaf diseases more effectively. An average accuracy of 92.3% was attained by the GLAD model when tested on the PLANT-VILLAGE dataset. In comparison, the YOLOv5 model only managed an average accuracy of 88.2%; this is a giant leap forward.

An innovative and encouraging method for detecting

diseases in grape leaves is the GLAD model. This method outperforms the current ones in accuracy and robustness, and it has multiple applications for detecting illnesses in grape leaves.

3.3.1 YOLOv5 framework

In this study, we used YOLOv5 as the basis for finding diseases on grape leaves. The newest member of the YOLO family is, YOLOv5, is a one-stage method for finding objects. We also added the SPPF (spatial pyramid pooling fast) tool to improve the effectiveness of training and make it easier to learn. The SPPF module not only speeds up the training process, but it also cuts down on unnecessary gradient information, which helps people learn better. YOLOv5 uses a network called PAFPN (Path Aggregation Feature Pyramid Network) to combine features. This network has three layers of recognition at different levels. This multiscale feature fusion module lets the model learn from features at different scales, which makes object recognition more accurate. One of the best things about YOLOv5 is that it is small and light. This makes it easy to set up and allows real-time tracking on IoT (Internet of Things) devices with limited resources.

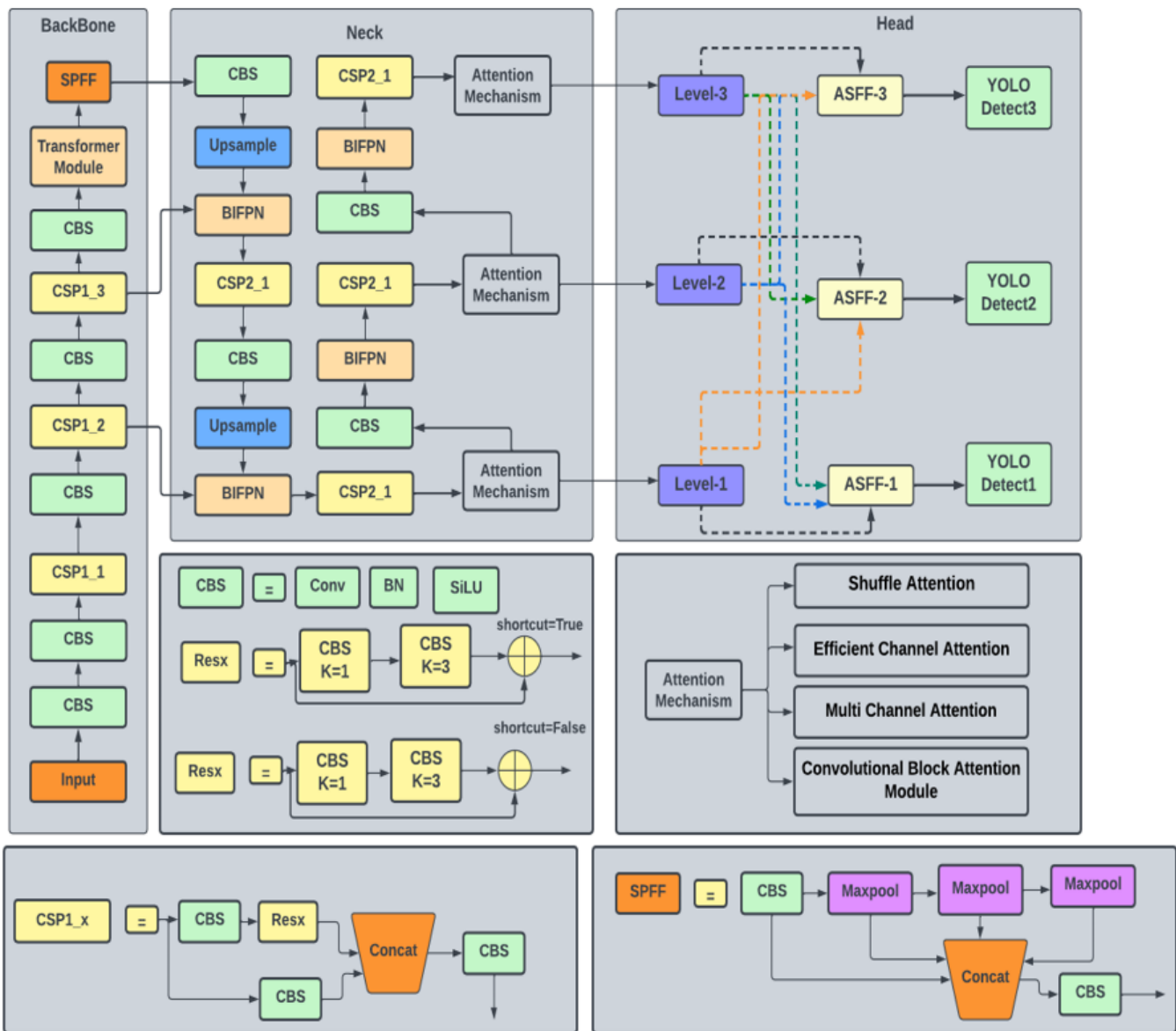


Figure 2. An improved YOLOv5 GLAD model

Taking all of these things into account, the researchers chose YOLOv5 as the basic framework for spotting diseases on grape leaves. They then made some new changes and developed the GLAD model, shown in Figure 2.

3.3.2 Self-attention system of transformer

Images of grape leaves can show how diseases are spread in different ways. Some diseases, like leaf blight, only affect small parts of a leaf and stay in one place. Detection depends more on local knowledge and high-level features in these situations. Other diseases, like black rot, are all over the leaf and can only be found by looking at the whole plant. It is important to get global semantic knowledge to improve the network's ability to figure out where it is. Most of YOLOv5's basic structure is based on convolutional neural networks (CNNs). CNN can only get information about long-distance interactions if they focus on how images next to each other are related. In addition, they need to make long-range semantic links better, and that's a problem for networks that want to ignore noise all over the feature map to focus on specific areas of interest. There is a strong correlation between the actual perception elds of extracted features and their theoretical counterparts, as they are significantly smaller in reality. This shows that convolutional operations alone aren't enough to determine how local picture features depend on things far away.

To get around this problem, the researchers developed GLAD, a new model that includes the self-attention process. Figure 3 shows the self-attention mechanism, which lets the model take in global semantic knowledge and improve its ability to and its way around.

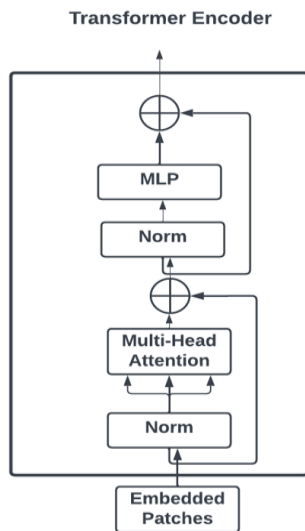


Figure 3. Self attention system for GLAD model

Self-attention processes have been suggested to get around the fact that CNNs are inherently local. The Transformer is a self-attention device that has been shown to work well in tasks like processing natural language and computer vision. The Transformer works by paying attention to different parts of the input series and learning how they are related. This means that the Transformer can pick up long-distance relationships that would be hard for CNNs to pick up.

Adopting visual Transformer networks shows how important it is to set up remote dependencies if you want to make big changes to how computer vision jobs are done. In short, adding self-attention mechanisms, especially the

Transformer, gets around the fact that CNNs can't capture long-range dependencies and lets the network use global semantic information successfully. This makes it easier for the network to nd and pinpoint tea diseases. The attention can be calculated as:

$$S_a = attention(Q_v, K_v, V_v) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where, S_a is the self-attention mechanism, Q_v is the query vector, K_v is the key vector, V_v is the value vector.

3.3.3 Enhanced object detection through multiscale feature fusion with BIFPN

Within the same disease category, our collection of grape leaf diseases revealed a broad variation of sizes and forms. This necessitates incorporating traits from different scales. To make YOLOv5's multiscale feature fusion network more effective, we swapped out the original PAFPN for BiFPN. Multiscale feature fusion is a module that is part of the BiFPN plugin. This updated PAFPN for YOLOv5 incorporates residual connections into the actual network topology and is based on EfficientDet. Nodes that contribute less to feature fusion are removed using cross-connections in PAFPN, while input and output nodes of the same scale are connected using jump connections. In contrast to PAFPN, BiFPN distributes feature information uniformly across multiple scales using weights.

Our GLAD model efficiently fuses multiscale aspects of grape leaf diseases by adhering to the concepts of BiFPN. As a result, the GLAD model's capacity for representing features is improved while the number of parameters is decreased. Figure 4 depicts the internal organisation of BIFPN.

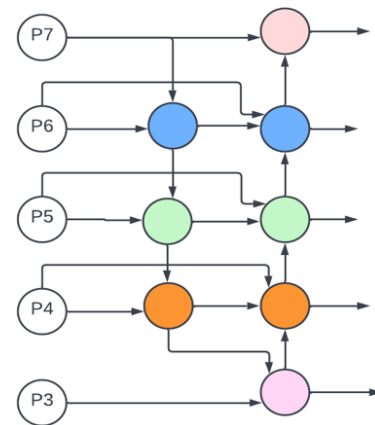


Figure 4. Structure of BIFPN

3.3.4 Attention mechanisms

Shuffle Attention: Detecting small target grape leaf diseases can be improved using the Shuffle Attention (SA) mechanism. SA combines spatial attention and channel attention mechanisms within blocks, which allows for their effective integration. Spatial attention captures the spatial dependence between pixels, while channel attention captures the dependence between channels. Using both types of attention mechanisms simultaneously can yield better results but requires increased computational resources. SA addresses this issue by incorporating the Channel Shuffle operation, which allows for efficient computation of both spatial and channel attention.

First, SA incorporates global information and generates a channel feature using Global Average Pooling as $s \in \mathbb{R}^{C/2G \times 1 \times 1}$. Z_{k1} . It is incorporated in Eq. (2).

$$S = F_{gp}(Z_{k1}) = \frac{1}{A \times B} = \sum_{m=1}^A \sum_{n=1}^B (Z_{k1})(m, n) \quad (2)$$

An adaptive selection module based on this channel feature is then applied, along with an activation function based on sigmoid geometry, in order to produce a precise feature that provides guidance for adaptive selection. It is incorporated in Eq. (3).

$$Z'_{k1} = \sigma(F_c(s)). Z_{k1} = \sigma(V_1 s + b_1) Z_{k1} \quad (3)$$

The SA mechanism then uses group norm (GN) to normalize the spatial feature map Z_{k2} . This normalized feature map is then passed through $F_c(\cdot)$ to enhance its representation. Spatial attention can be calculated using Eq. (4) as the final output.

$$Z'_{k2} = \sigma(V_2 \cdot GN(Z_{k2}) + b_2) \cdot Z_{k2} \quad (4)$$

The SA mechanism can be used to improve the detection of small target grape leaf diseases by selectively focusing on the characteristics of these diseases. This is done using spatial and channel attention mechanisms to capture the relevant information from the input image. The SA mechanism is a promising new approach for improving the accuracy of object detection algorithms.

The initial convolutional neural network (CNN) design of YOLOv5 included shuffle attention. Shuffle attention improves the model's capacity to convey the semantics of grape leaf disease traits by efficiently merging channel and spatial attention processes. One benefit of the suggested GLAD model is its ability to be partitioned and parallelized, which allows for the extraction of attention regions and the focus on the unique features of grape leaf diseases.

Convolutional Block Attention Module: Convolutional neural networks (CNNs) can be more precise with the help of an attention mechanism called the Convolutional Block Attention Module (CBAM). CBAM analyzes the output of a CNN, and the model's ability to classify images is enhanced by emphasizing only the most relevant elements. By zeroing in on the most pertinent aspects of an input image, CBAM can improve the detection of small-target grape leaf diseases. A combination of spatial and channel attention techniques is used to pick out the crucial details in the input image and bring them to the fore.

CBAM's spatial attention mechanism is useful for pinpointing key regions of interest within an input image. To create an attention map, we combine the results from the CNN's convolutional layers and apply a sigmoid activation function to the combined data. By highlighting the most crucial regions of the input image, the attention map directs the CNN's processing power where it is most needed.

CBAM's channel attention mechanism aids in selecting the most relevant channels from the given input image. A sigmoid activation function is applied to the combined output of the CNN's convolutional layers across the channel dimension to create an attention map. By highlighting the most crucial channels in the input image, the attention map directs the CNN's processing power where it is most needed.

CBAM can aid in the accuracy of CNNs by selecting and focusing on the most significant elements in the input image

through a combination of spatial and channel attention techniques. This is especially useful for jobs that require discriminating between items or classes, such as object detection. By including it in the model's feature extraction layers, YOLOv5 now benefits from the CBAM attention module.

Efficient Channel Attention: ECA (Efficient Channel Attention) is a channel gating method that uses a feature map to learn the relevance of each channel. ECA is a superior attention mechanism because it uses a more effective method than SE-Net's global average pooling layer. This can be useful for reducing the computational complexity of the model without sacrificing accuracy, which is especially relevant for disease detection in grape leaves. First, a weight is determined for each channel in the feature map, which is used by the channel gating mechanism in ECA. The feature map is gated using these weights so only channels with large values can proceed to the next layer. By zeroing in on the most crucial channels in the feature map, ECA is able to boost the precision of the model. SE-Net's global average pooling layer assigns a weight to each channel in the feature map. However, unlike ECA's channel gating technique, the global intermediate pooling layer averages all the pixel values in each channel.

ECA's potential to reduce the model's computational complexity makes it a promising tool for disease detection in grape leaves. As disease identification in grape leaves is a computationally intensive task, this is significant as it can lead to better model speed and performance.

Multi-Channel Attention (MCA): It combines a single-stage attention mechanism with multiple attention heads to learn the significance of distinct channels on a feature map. As a result, MCA can pick up on subtler interdependencies between the channels, which can prove helpful in situations like disease detection in grape leaves. For MCA to function, the feature map must first be segmented into separate channels. Then, individual "attention heads" deal with each channel individually. Each channel's attention head figures out how significant its channel is by gauging it against the others. With several attention heads, MCA can learn intricate interdependencies between channels. Complex dependency learning might be helpful for problems like disease detection in grape leaves, where distinguishing between similar diseases can be difficult. For instance, the symptoms of two distinct grape leaf diseases may be identical, yet their effects on the leaves may differ. MCA can discern these subtle distinctions by learning the interdependencies between the channels in the feature map.

3.3.5 Adaptive spatial feature fusion

Grape leaf disease detection is difficult because of the diverse context and changing disease severity. To solve these problems, we suggest a new detection method called adaptive spatial feature fusion (ASFF). Grape leaf disease diagnosis is greatly aided by ASFF's ability to automatically alter out the noise and minimize interference from complicated backdrops. It also makes it easier to combine data from multiple levels of illness analysis.

We swapped over YOLOv5's old detecting head for a new ASFF one. Adjusting the fusion ratio between feature layers is a key function of the ASFF detection head. During multiscale feature fusion, ASFF decreases inference overhead by altering spatially connecting information and improves the invariance of the feature ratio. As a result, these targets for grape leaf diseases can be detected more easily. Figure 5 depicts the

internal organization of ASFF.

The YOLOv5 network's "neck" produces results labelled "Level1, Level 2, and Level 3." For illustration, think about ASFF-3. After fusion, ASFF-3's input is derived from a weighted total of the inputs at Levels 1, 2, and 3. Multiplication and adding of masses characterize this fusion process, as seen in Eq. (5).

$$b'_{ij} = \alpha_{ij} * a_{ij}^{1 \rightarrow 1} + \beta_{ij} * a_{ij}^{2 \rightarrow 1} + \chi_{ij} * a_{ij}^{3 \rightarrow 1} \tag{5}$$

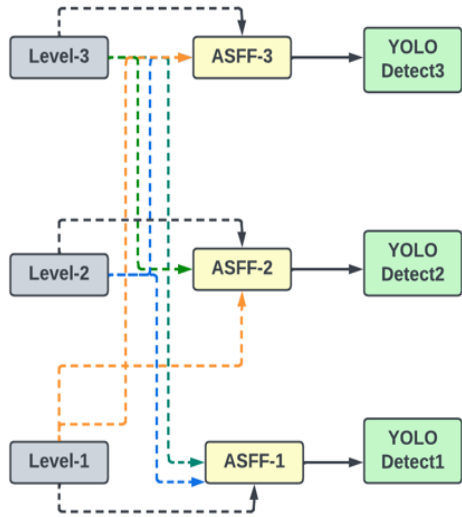


Figure 5. Structure of ASFF

4. RESULTS & DISCUSSIONS

The experiment was conducted on a computer system running the Windows operating system. The programming language used for implementation was Python, with GPU acceleration through the CUDA framework. Both the training and testing environments were identical. For more information about the experimental environment, please refer to Table 3.

Table 3. Training details

Training Environment	Details
Programming language	Python
Pytorch version	1.8.2
GPU	RTX 5060
CPU	R7 5800
Operating system	Microsoft Windows 8

4.1 Evaluation metrics

Grape leaf disease detection models are assessed on precision, recall, and mAP.

- True Positives (TP): Accurate grape leaf disease identification.
- False Positives (FP): Grape leaf diseases are misidentified as healthy leaves.

Number of grape leaf diseases missed (false negatives). Precision is the ratio of genuine positives (correctly identified grape leaf diseases) to the model's anticipated true positives. It measures how successfully the model eliminates false positives. Recall, also known as sensitivity or true positive rate, is the percentage of true positives discovered. This shows that the model can describe all grape leaf diseases—popular performance metric Mean Average Precision averages

detection accuracy from numerous types. The model's utility is assessed by accuracy and recall.

4.2 Performance evaluation and comparative analysis

SGD was used to train the GLAD model. A 0.001 beginning learning rate (lr) was modified using linear scaling. Transfer-learned weights employed in early training underwent 300 iterations. Figure 6 shows the results of numerous attentional visualization methods.

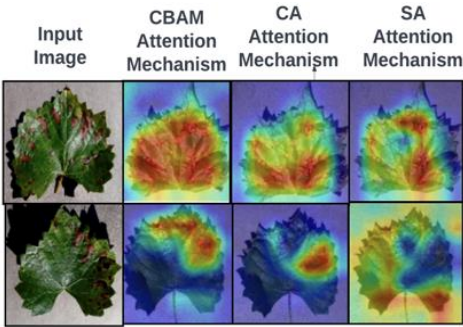


Figure 6. Presentation of results utilizing various attention mechanisms

Table 4. Analysis of YOLO-V5 models using various attention strategies for disease detection in grapes

Model	Prec	Rec	FPS	mAP@0.5
Yolov5+Transformer	83.90	82.99	54	81.67
Yolov5+Transformer+BIFPN	85.34	83.98	55	83.32
Yolov5+Transformer+BIFPN+ASFF+SA	85.67	84.56	54	84.67
Yolov5+Transformer+BIFPN+ASFF+ECA	85.88	83.45	53	84.89
Yolov5+Transformer+BIFPN+ASFF+MCA	85.99	83.78	54	85.78
Yolov5+Transformer+BIFPN+ASFF+CBAM	86.67	84.97	53	87.86

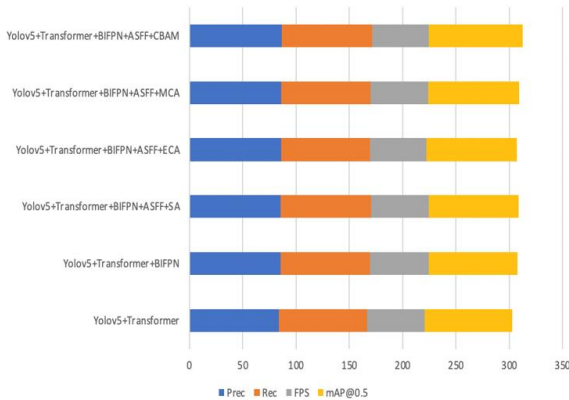


Figure 7. Analysis of YOLO-V5 models using various attention strategies for disease detection in grapes

Table 4 and Figure 7 compare YOLOv5 model setups with feature improvement methods. Precision, recall, FPS, and mean Average Precision at 0.5 IoU are assessed. All configurations have high precision, recall, and mAP scores, indicating good object detection. Compared to the basic YOLOv5+Transformer setup, feature enhancement techniques like BiFPN (Bilateral Feature Pyramid Network),

ASFF (Adaptive Spatial Feature Fusion), SA (Squeeze-and-Attention), ECA (Efficient Channel Attention), MCA (Multi-Context Attention), and CBAM improve model performance. Incorporating more enhancement techniques leads to higher mAP@0.5 scores, a crucial measure for evaluating the model's accuracy in detecting objects. However, it is worth noting that including these enhancement techniques also results in a slight decrease in the frames per second (FPS) metric. This reduction in FPS is anticipated as the additional computations required for feature enhancements demand more processing time.

5. CONCLUSION

The paper presents the GLAD model as a real-time system for detecting grape leaf illnesses. It was designed to overcome the difficulties caused by the wide variety of grape leaf diseases and complicated background interferences. Notably, the GLAD model greatly improves global perception during disease diagnosis by integrating Transformer's self-attention mechanism with multiscale feature fusion. The experimental results highlight the GLAD model's exceptional performance, with an outstanding average accuracy of 87.86% and an efficient detection speed of 53FPS. Incorporating Transformer's self attention mechanism into the GLAD model enhances its accuracy by allowing it to capture complicated patterns and nuanced information across different illness presentations. Multiscale feature fusion is crucial when navigating and understanding complex background interferences since it consolidates information from various sizes. These technologies work together to make the GLAD model a powerful tool for detecting diseases in grape leaves.

Ultimately, the GLAD model's ability to achieve faster and more accurate results is mainly due to the deliberate integration of Transformer's self attention and multiscale feature fusion. To improve detection accuracy even further, future efforts will center on synergistic model integration, which combines insights from several models and tackles any problems with false and missing detections.

REFERENCES

- [1] Food and Agriculture Organization of the United Nations. Plant diseases and pests. Available: <http://www.fao.org/emergencies/emergency-types/plant-pests-and-diseases/en/>.
- [2] Carlson, G.A. (1970). A decision theoretic approach to crop disease prediction and control. *American Journal of Agricultural Economics*, 52(2): 216-223. <https://doi.org/10.2307/1237492>
- [3] Sirisha, U., Chandana, B.S. (2023). Utilizing a hybrid model for human injury severity analysis in traffic accidents. *Traitement du Signal*, 40(5): 2233-2242. <https://doi.org/10.18280/ts.400540>
- [4] Sirisha, U., Chandana, B.S., Harikiran, J. (2023). NAM-YOLOV7: An improved YOLOv7 based on attention model for animal death detection. *Traitement du Signal*, 40(2): 783-789. <https://doi.org/10.18280/ts.400239>
- [5] Srinivasu, P.N., Sirisha, U., Sandeep, K., Praveen, S.P., Maguluri, L.P., Bikku, T. (2024). An interpretable approach with explainable AI for heart stroke prediction. *Diagnostics*, 14(2): 128. <https://doi.org/10.3390/diagnostics14020128>
- [6] Ghaffari, R., Laothawornkitkul, J., Iliescu, D., Hines, E., Leeson, M., Napier, R., Taylor, J.E. (2012). Plant pest and disease diagnosis using electronic nose and support vector machine approach. *Journal of plant diseases and protection*, 119: 200-207. <https://doi.org/10.1007/BF03356442>
- [7] Zhang, D., Chen, G., Zhang, H., Jin, N., Gu, C., Weng, S., Chen, Y. (2020). Integration of spectroscopy and image for identifying fusarium damage in wheat kernels. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 236: 118344. <https://doi.org/10.1016/j.saa.2020.118344>
- [8] Wang, Z., Wang, K., Pan, S., Han, Y. (2018). Segmentation of crop disease images with an improved K-means clustering algorithm. *Applied engineering in agriculture*, 34(2): 277-289.
- [9] Kamath, R., Balachandra, M., Prabhu, S. (2020). Crop and weed discrimination using Laws' texture masks. *International Journal of Agricultural and Biological Engineering*, 13(1): 191-197. <https://doi.org/10.25165/j.ijabe.20201301.4920>
- [10] Bi, C., Chen, G. (2011). Bayesian networks modeling for crop diseases. In *Computer and Computing Technologies in Agriculture IV: 4th IFIP TC 12 Conference, CCTA 2010, Nanchang, China*, pp. 312-320. https://doi.org/10.1007/978-3-642-18333-1_37
- [11] Ferentinos, K.P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145: 311-318. <https://doi.org/10.1016/j.compag.2018.01.009>
- [12] Schroff, F., Kalenichenko, D., Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 815-823. <https://doi.org/10.1109/CVPR.2015.7298682>
- [13] Yang, W., Fan, S., Xu, S., King, P., Kang, B., Kim, E. (2019). Autonomous underwater vehicle navigation using sonar image matching based on convolutional neural network. *IFAC-PapersOnLine*, 52(21): 156-162. <https://doi.org/10.1016/j.ifacol.2019.12.300>
- [14] Sivaraman, S., Trivedi, M.M. (2014). Active learning for on-road vehicle detection: A comparative study. *Machine Vision and Applications*, 25: 599-611. <https://doi.org/10.1007/s00138-011-0388-y>
- [15] Zhang, L., Lin, L., Liang, X., He, K. (2016). Is faster R-CNN doing well for pedestrian detection? In *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands*, pp. 443-457. https://doi.org/10.1007/978-3-319-46475-6_28
- [16] Sirisha, U., Chandana, B.S. (2023). Gitaar-git based abnormal activity recognition on ucf crime dataset. In *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, pp. 1585-1590. <https://doi.org/10.1109/ICSSIT55814.2023.10061116>
- [17] Praveen, S.P., Suntharam, V.S., Ravi, S., Harita, U., Thatha, V.N., Swapna, D. (2023). A novel dual confusion and diffusion approach for grey image encryption using multiple chaotic maps. *International Journal of Advanced Computer Science and Applications*, 14(8): 971-984. <https://doi.org/10.14569/IJACSA.2023.01408106>
- [18] Sivaraman, S., Trivedi, M.M. (2014). Active learning for

- on-road vehicle detection: A comparative study. *Machine Vision and Applications*, 25: 599-611. <https://doi.org/10.1007/s00138-011-0388-y>
- [19] Zhang, L., Lin, L., Liang, X., He, K. (2016). Is faster R-CNN doing well for pedestrian detection? In *Computer Vision—ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, pp. 443-457. https://doi.org/10.1007/978-3-319-46475-6_28
- [20] Sirisha, U., Chandana, B.S. (2023). Privacy preserving image encryption with optimal deep transfer learning based accident severity classification model. *Sensors*, 23(1): 519. <https://doi.org/10.3390/s23010519>
- [21] Praveen, S.P., Nakka, R., Chokka, A., Thatha, V.N., Vellela, S.S., Sirisha, U. (2023). A novel classification approach for grape leaf disease detection based on different attention deep learning techniques. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(6): 1199-1209.
- [22] Polly, R., Devi, E.A. (2022). A deep learning-based study of crop diseases recognition and classification. In *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, pp. 296-301. <https://doi.org/10.1109/ICAIS53314.2022.9742950>
- [23] Pan, X., Ge, C., Lu, R., Song, S., Chen, G., Huang, Z., Huang, G. (2022). On the integration of self-attention and convolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, pp. 815-825. <https://doi.org/10.1109/CVPR52688.2022.00089>
- [24] Ultralytics-Yolov5. Available online: <https://github.com/ultralytics/yolov5>
- [25] Tian, Y., Wang, Y.T., Wang, J.G., Wang, X., Wang, F.Y. (2022). Key problems and progress of vision transformers: The state of the art and prospects. *Acta Automatica Sinica*, 48(4): 957-979.
- [26] Tan, M., Pang, R., Le, Q.V. (2020). EfficientDet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 10781-10790. <https://doi.org/10.1109/CVPR42600.2020.01079>
- [27] Madhuri, A., Jyothi, V.E., Praveen, S.P., Altaee, M., Abdullah, I.N. (2023). Granulation-based data fusion approach for a critical thinking worldview information processing. *Journal of Intelligent Systems and Internet of Things*, 9(1): 49-68. <https://doi.org/10.54216/JISIoT.090104>
- [28] Liu, S., Huang, D., Wang, Y. (2019). Learning spatial fusion for single-shot object detection. *arXiv preprint arXiv:1911.09516*. <https://doi.org/10.48550/arXiv.1911.09516>
- [29] Alkan, A., Abdullah, M.U., Abdullah, H.O., Assaf, M., Zhou, H. (2021). A smart agricultural application: Automated detection of diseases in vine leaves using hybrid deep learning. *Turkish Journal of Agriculture and Forestry*, 45(6): 717-729. <https://doi.org/10.3906/tar-2007-105>
- [30] Liu, B., Ding, Z. (2020). Grape leaf disease identification using improved deep convolutional neural networks. *Frontiers in Plant Science*, 11: 542365. <https://doi.org/10.3389/fpls.2020.01082>
- [31] Lauguico, S., Concepcion, R., Tobias, R.R., Bandala, A., Vicerra, R.R., Dadios, E. (2020). Grape leaf multi-disease detection with confidence value using transfer learning integrated to regions with convolutional neural networks. In *2020 IEEE region 10 conference (TENCON)*, Osaka, Japan, pp. 767-772. <https://doi.org/10.1109/TENCON50793.2020.9293866>
- [32] Zhou, C., Zhang, Z., Zhou, S., Xing, J., Wu, Q., Song, J. (2021). Grape leaf spot identification under limited samples by fine grained-GAN. *IEEE Access*, 9: 100480-100489. <https://doi.org/10.1109/ACCESS.2021.3097050>
- [33] Math, R.M., Dharwadkar, N.V. (2022). Early detection and identification of grape diseases using convolutional neural networks. *Journal of Plant Diseases and Protection*, 129(3): 521-532. <https://doi.org/10.1007/s41348-022-00589-5>
- [34] Peng, Y., Zhao, S., Liu, J. (2021). Fused-deep-features based grape leaf disease diagnosis. *Agronomy*, 11(11), 2234. <https://doi.org/10.3390/agronomy11112234>
- [35] Yin, X., Li, W., Li, Z., Yi, L. (2022). Recognition of grape leaf diseases using MobileNetV3 and deep transfer learning. *International Journal of Agricultural and Biological Engineering*, 15(3): 184-194. <https://doi.org/10.25165/j.ijabe.20221503.7062>
- [36] Yang, R., Lu, X., Huang, J., Zhou, J., Jiao, J., Liu, Y., Gu, P. (2021). A multi-source data fusion decision-making method for disease and pest detection of grape foliage based on ShuffleNet V2. *Remote Sensing*, 13(24): 5102. <https://doi.org/10.3390/rs13245102>
- [37] Lu, X., Yang, R., Zhou, J., Jiao, J., Liu, F., Liu, Y., Gu, P. (2022). A hybrid model of ghost-convolution enlightened transformer for effective diagnosis of grape leaf disease and pest. *Journal of King Saud University-Computer and Information Sciences*, 34(5): 1755-1767. <https://doi.org/10.1016/j.jksuci.2022.03.006>
- [38] Zhu, J., Cheng, M., Wang, Q., Yuan, H., Cai, Z. (2021). Grape leaf black rot detection based on super-resolution image enhancement and deep learning. *Frontiers in plant science*, 12: 695749. <https://doi.org/10.3389/fpls.2021.695749>
- [39] Suo, J., Zhou, G., Cai, W. (2022). CASM-AMFMNet: A network based on coordinate attention shuffle mechanism and asymmetric multi-scale fusion module for classification of grape leaf diseases. *Frontiers in Plant Science*, 13: 846767. <https://doi.org/10.3389/fpls.2022.846767>
- [40] Guo, W., Feng, Q., Li, X., Yang, S., Yang, J. (2022). Grape leaf disease detection based on attention mechanisms. *International Journal of Agricultural and Biological Engineering*, 15(5): 205-212. <https://doi.org/10.25165/j.ijabe.20221505.7548>
- [41] Lin, J., Chen, X., Pan, R., Cao, T., Cai, J., Chen, Y., Zhang, X. (2022). Grapenet: A lightweight convolutional neural network model for identification of grape leaf diseases. *Agriculture*, 12(6): 887. <https://doi.org/10.3390/agriculture12060887>
- [42] Zinonos, Z., Gkelios, S., Khalifeh, A.F., Hadjimitsis, D.G., Boutalis, Y.S., Chatzichristofis, S.A. (2021). Grape leaf diseases identification system using convolutional neural networks and Lora technology. *IEEE Access*, 10: 122-133. <https://doi.org/10.1109/ACCESS.2021.3138050>
- [43] Xie, X., Ma, Y., Liu, B. (2020). A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks. *Frontiers in Plant Science*, 11: 529357. <https://doi.org/10.3389/fpls.2020.00751>