Plant disease identification using Vanillanet

Linsen Wei[1] , Zhuangda Wang[1] , Zuolin Tian[1] , Zihan Wang[2] , Hong Wu[2]

[1] School of Informatics, Xiamen University.
[2] Artificial Intelligence Research Institute, Xiamen University.
{23020241154451, 23020241154359, 23020241154439, 23020241154450, 36920241153257}@stu.xmu.edu.cn

Abstract

Plant disease identification is crucial for ensuring crop health and improving agricultural productivity. However, existing models often require substantial computational resources, limiting their applicability in resource-constrained environments. In this paper, we use VanillaNet-5, a lightweight and efficient deep learning model designed for plant disease classification. The model leverages a minimalist architecture, combining deep training and series activation to optimize both performance and computational efficiency. Experimental results on the PlantDoc dataset demonstrate that VanillaNet-5 achieves an accuracy of 54.237% and an F_1 score of 0.429, outperforming traditional models like MobileNetV3-L and ResNet50 in terms of both accuracy and latency. The proposed model offers a practical solution for real-time, large-scale plant disease monitoring, particularly in field conditions where computational resources are limited. This work contributes to the advancement of smart agriculture by providing an efficient and scalable framework for plant disease detection.

Introduction

Agriculture, as the foundation of human civilization, plays a pivotal role in ensuring food supply, fostering employment, and driving economic growth. As a cornerstone of the global economy, the health of crops directly impacts food security and agricultural productivity. However, the widespread occurrence of plant diseases and pests significantly hampers crop growth and yield, posing a severe threat to agricultural production. According to the Food and Agriculture Organization (FAO), up to 40% of global crop yields are lost annually due to pests and diseases. Consequently, the development of efficient and automated plant disease identification technologies is critical for enhancing crop protection, ensuring food security, and promoting sustainable agricultural practices (Xu et al. 2023).

Traditional methods for plant disease diagnosis heavily rely on human expertise and experience. While these methods are often accurate, they are time-consuming, costly, and impractical for large-scale, real-time agricultural monitoring and management. In recent years, advancements in artificial intelligence (AI), deep learning (DL), and computer vision (CV) have paved the way for automated approaches to detecting crop diseases through image analysis. Among these technologies, models such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) have demonstrated outstanding performance in computer vision tasks like image classification and object detection (Dosovitskiy et al. 2021). By training these models on large datasets of crop disease images, they can learn feature patterns and achieve high-precision classification of various disease types, thereby significantly improving the accuracy and efficiency of plant disease identification (Chang et al. 2024). Furthermore, research leveraging multimodal data, such as images, climate, and soil information, to enhance recognition performance is rapidly advancing. These technologies not only enable cost-effective identification but also integrate seamlessly with mobile devices and drone systems, greatly enhancing the efficiency of disease monitoring.

However, despite these advancements, several challenges remain. Leaf images captured in natural environments are often subject to various noise factors, and the deployment of deep learning models in real-world agricultural settings is still difficult (Mingyue Shao 2022). This paper aims to design and implement a lightweight and efficient model for leaf disease identification using VanillaNet (Chen et al. 2023). This research offers several key contributions:

- Developed a lightweight network architecture for crop disease recognition. This work addresses the limitations of data and computational resources commonly encountered in complex natural environments.

- The proposed approach promotes the automation and intelligence of leaf disease identification, contributing to more efficient and scalable agricultural solutions. Through these efforts, we aim to provide an effective and practical solution for smart agriculture, addressing the pressing challenges in disease management.

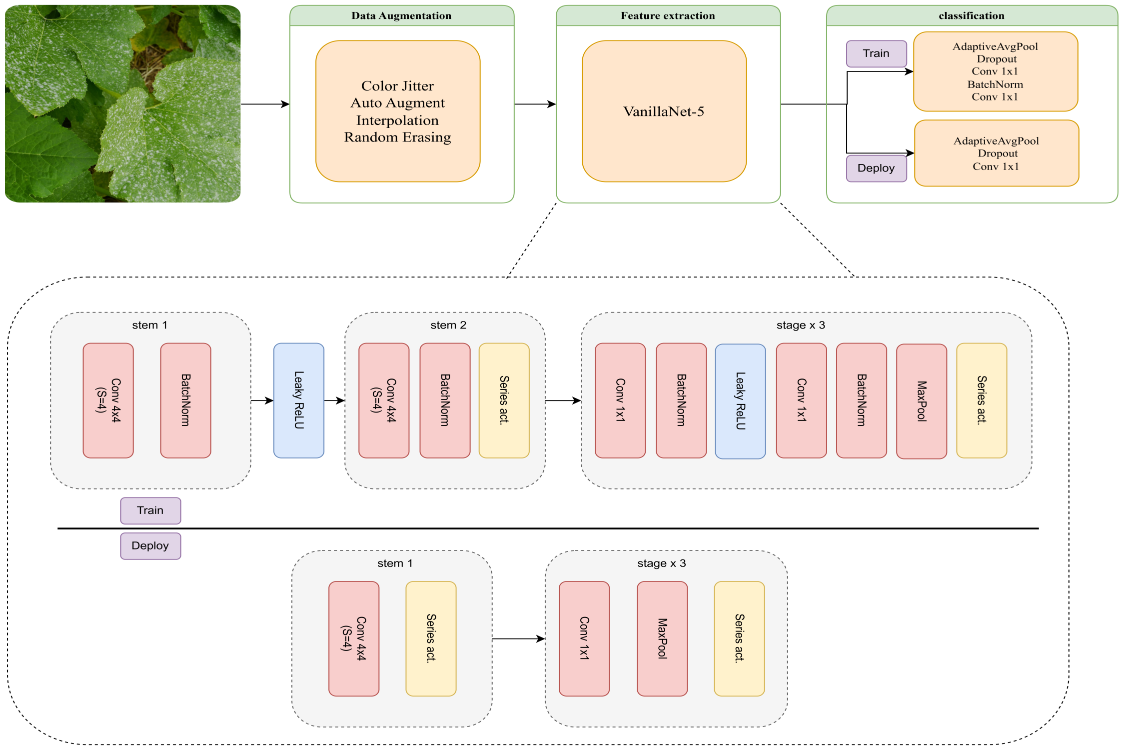


Figure 1: Vanillanet architecture for classification of leaf disease.

Related work

Crop disease recognition has evolved significantly over the years, transitioning from traditional machine learning approaches to advanced deep learning-based models.

Traditional Machine Learning Approaches: Traditional machine learning methods for crop disease identification often rely on handcrafted feature extraction combined with classifiers such as Support Vector Machines (SVMs), Random Forests (RFs), and K-Nearest Neighbors (KNNs). These techniques extract features like color, texture, and shape to classify and detect diseases. SVMs are particularly effective in handling small datasets and high-dimensional data, while RF, as an ensemble method, improves stability and accuracy by integrating multiple decision trees (Rodríguez-Lira et al. 2024). However, these methods heavily depend on the quality of the handcrafted features and the effectiveness of feature extraction. Additionally, the diversity of disease samples and environmental variations in natural settings limit the generalization performance of these models.

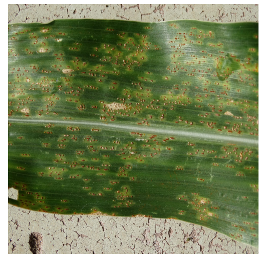
PlantDoc solutions: Recent studies have made significant progress in PlantDoc dataset detection and disease recognition. Cascade-DETR (Ye et al. 2023) improved detection accuracy by introducing Cascade Attention and IoU-aware Query Recalibration, enhancing localization performance in plant detection tasks. BIOCLIP (Stevens et al. 2024) utilized biological taxonomy information to learn finer-grained visual features, improving accuracy on the PlantDoc dataset. AgriDet (Pal and Kumar 2023) proposed a framework that combines image preprocessing, a multivariable grabbing algorithm, and an improved INC-VGGN neural network to effectively extract and classify plant diseases from complex backgrounds.

Deep Learning Approaches: Deep learning-based image recognition has revolutionized crop disease detection by automating feature extraction, significantly enhancing efficiency and accuracy (Wanjie Liang, 2023). Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) have been widely employed to process large datasets, achieving state-of-the-art performance in precision agriculture (Wanjie Liang 2023). These approaches enable timely interventions, effectively mitigating crop losses. For instance, GSMo-CNN adopts a multi-task prediction framework that leverages hierarchical classification to improve disease and species identification (J. P. S. Schuler 2022).

To address the limitations of deep learning in resource-constrained environments, lightweight models have gained significant attention. Wang Chunshan et al. proposed a multi-scale residual module-based model by adjusting the connections of residual layers and decomposing large convolution kernels, resulting in a lightweight and efficient disease recognition system (Chunshan Wang 2020). Saleem et al. demonstrated the superior performance of Xception, which achieved

(a) Apple rust leaf

(b) Bell_pepper leaf spot



(c) Corn rust leaf

(d) Potato leaf early blight

Figure 2: Four images of the Plantdoc.

a 99.81% accuracy on the PlantVillage dataset by integrating Adam optimization for enhanced performance (Saleem, Potgieter, and Arif 2020). De Ocamop and Dadios employed MobileNet for disease classification, showcasing its ability to handle small datasets effectively while maintaining high classification accuracy (Ocampo and Dadios 2018). MobileNet's compact architecture ensures suitability for scenarios with limited computational resources.

Despite these advancements, several challenges persist. First, the complexity of natural environments, including backgrounds with soil, weeds, and shadows, often obscures disease symptoms, making accurate identification difficult. Variations in lighting conditions and camera angles further complicate symptom recognition by altering leaf color and texture (Singh et al. 2020). Second, the diversity of plant diseases poses generalization challenges, as the same disease may present different symptoms across plant species and growth stages. Finally, deploying deep learning models in agricultural settings is constrained by the need for substantial computational power and large datasets, limiting access for smallholder farmers. These challenges underscore the need for lightweight, efficient models that perform reliably in diverse and resource-constrained environments.

Methods

This section presents a architecture for classification of leaf disease, illustrated in Figure 1. Initially, the input image undergoes a series of data augmentation processes to enhance the robustness of the model. These processes include techniques such as Color Jitter Auto Augment, which randomly adjusts the brightness, contrast, saturation, and hue of the images; interpolation, which is used to resize or transform images while maintaining their quality; and Random Erasing, which involves randomly erasing a portion of the image to simulate occlusions. Following these augmentations, the image is subjected to feature extraction using the VanillaNet-5 network from (Chen et al. 2023). The extracted features are then fed into a fully connected layer, which serves as the final step in the classification pipeline, producing the classification outcomes.

Data augmentation

In this work, we incorporate several data augmentation techniques to enhance the robustness and generalization ability of our plant recognition model. We apply color jittering to introduce random variations in brightness, contrast, saturation, and hue, allowing the model to become invariant to changes in lighting conditions. Additionally, AutoAugment is utilized, an automated augmentation policy search method that combines multiple transformations such as rotation, translation, and shearing to optimize performance. Random erasing is also employed, where a random rectangular region of the image is replaced with a constant value, encouraging the model to focus on diverse visual features rather than memorizing specific areas of the image. To handle images of varying resolutions, different interpolation methods are used during resizing, ensuring that the model is capable of dealing with multiple input image qualities. Finally, random resized cropping is applied to vary the scale and aspect ratio of the images, making the model robust to changes in object size and positioning. These augmentation strategies collectively increase the diversity of the training data, contributing to improved generalization and performance in real-world plant recognition tasks.

Network architecture

VanillaNet-5 is designed with a minimalist approach, emphasizing simplicity and efficiency. The network's architecture is composed of several key components: an initial stem layer that features a 4×4 convolution with a stride of 4 for preliminary feature extraction, followed by Batch Normalization and a LeakyReLU activation function. Subsequently, the stem 2 layer consists of a 4×4 convolution and a Series activation function, also followed by Batch Normalization. The main body of the model is structured into three repeating stages, each containing a 1×1 convolution, LeakyReLU activation, another 1×1 convolution, Batch Normalization, and a Series activation function, culminating in a MaxPool layer to reduce the spatial dimensions of the feature maps. This design aims to reduce model complexity while maintaining high performance.

Table 1: Comparison on PlantDoc. Latency is tested on Nvidia 3090-Ti GPU with batch size of 1.

Model	Params(M)	FLOPs(G)	Latency(ms)	$Acc(\%)$	F_1
MobileNetV3-L	4.24	0.22	1.77	42.797	0.382
Vgg19	139.68	19.63	4.11	33.475	0.216
Resnet50	23.56	4.10	2.01	52.542	0.492
Resnet101	42.56	7.82	3.68	52.119	0.478
Swin-B	86.77	15.46	5.04	34.322	0.188
PVTv2-B0	3.42	0.57	1.90	37.288	0.208
VanillaNet-5	17.34	8.46	1.02	54.237	0.429
VanillaNet-6	51.13	10.12	1.28	47.881	0.392
VanillaNet-8	60.19	13.51	1.61	37.712	0.312
VanillaNet-10	77.20	16.85	1.83	29.661	0.210

Table 2: Ablation study on VanillaNet-5

Model	Deep train.	Series act.	$Acc(\%)$	F_1
		✓	15.678	0.094
VanillaNet-5	✓		35.169	0.210
	✓	✓	54.237	0.429

Deep Train

VanillaNet-5 employs a training strategy that combines deep training with shallow inference, optimizing both performance and computational efficiency. During the training phase, the network incorporates two convolutional layers with an activation function LeakyReLU.

$$LeakyReLU(x) = \begin{cases} x & \text{if } x \geqslant 0 \\ \lambda \times x & else \end{cases} \quad (1)$$

To adaptively tune the model's learning dynamics, we define a parameter λ, representing the negative slope of the LeakyReLU activation, as a function of the current training epoch e and the total number of deep training epochs E. Specifically, $\lambda = \frac{e}{E}$, allowing λ to gradually increase as training progresses. This dynamic adjustment encourages a smoother learning curve by balancing the network's responsiveness to negative inputs throughout the training process.

At the end of the training phase, the two convolutional layers, along with their learned weights and biases, can be merged into a single convolutional layer. This merging process simplifies the network structure for inference, significantly reducing the computational overhead while maintaining the accuracy achieved during deep training. As a result, VanillaNet-5 achieves an efficient deployment model suitable for resource-constrained environments, without compromising on its capacity to learn complex data representations.

Series activation

Additionally, VanillaNet-5 introduces a Series activation function, which involves stacking multiple activation functions in parallel to significantly boost the network's non-linear expressive power. This approach allows the network to learn more complex representations during training. In the inference phase, the model simplifies the use of activation functions to facilitate faster and more efficient inference, making it suitable for deployment in resource-constrained environments. This design philosophy enables VanillaNet-5 to achieve a balance between performance and efficiency, demonstrating its potential for real-world applications.

In the context of neural networks, consider a single activation function applied to an input x, denoted by A(x), which may correspond to standard functions such as ReLU or Tanh. The concept of stacking instances of A(x) in parallel can be mathematically represented as:

$$A_s(x) = \sum_{i=1}^{n} a_i A(x + b_i) \quad (2)$$

Where n denotes the number of stacked activation functions, and a_i and $b_i$represent the scale and bias parameters of each activation, respectively. These parameters are introduced to prevent simple accumulation effects and ensure better control over the activation outputs.

Experiments

This section presents the experimental results of the proposed palnt disease identification model on the Plantdoc dataset. Ablation study is provided to investigate effectiveness of each component of the proposed model.

Experimental Setup and Evaluation Indicators

The experiments were conducted on an NVIDIA GeForce RTX 3090 Ti GPU with Ubuntu 20.04, using PyTorch 1.12. The system features a 12th i9-12900K CPU and 32GB of RAM, providing adequate resources for large datasets and complex models.

The experiments use the accuracy rate (Acc) as an evaluation index, which indicates the ratio of the number of correctly recognized samples to the total number of samples, and the specific formula is shown in Eq. (3). In addition, this paper uses the F_1 score as a complementary metric, which combines precision (P) and recall (R) into a single value, providing a balanced measure of a model's performance.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{6}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

Plantdoc dataset

The PlantDoc dataset is a highly valuable resource for visual plant disease detection, making it particularly suitable for real-world applications. With 2,598 images across 13 plant species and 17 disease classes, the dataset has been meticulously annotated, requiring around 300 hours of human effort. Unlike more controlled, laboratory-based datasets, PlantDoc emphasizes real-world conditions by providing images captured in diverse field environments. This characteristic offers a more accurate representation of plant disease in practical, outdoor settings, where variables such as lighting, weather, and background noise can influence the appearance of plant diseases. By using this dataset, we ensure that our model is better equipped to handle the challenges of real-world deployment, where conditions often vary significantly from those in controlled experiments. Figure 2 shows four images from the PlantDoc dataset, highlighting its diversity and relevance for field applications.

Comparison with other networks

This section compares the performance of different variants of VanillaNet with other CNNs and ViTs on the PlantDoc dataset, where none of the models used pre-trained weights. Among the VanillaNet variants, VanillaNet-5 performs the best, achieving 54.237% accuracy and an F_1 score of 0.429. It strikes a good balance between performance and computational efficiency, with 17.34 million parameters, 8.46 GFLOPs, and a latency of 1.02 ms, outperforming both MobileNetV3-L and ResNet50.

As the depth of VanillaNet increases, performance declines. VanillaNet-6 achieves 47.881% accuracy, VanillaNet-8 reaches 37.712%, and VanillaNet-10 performs the worst with 29.661% accuracy. These results suggest that deeper models may suffer from overfitting or diminishing returns on this dataset, and their performance might also be impacted by higher latencies.

In comparison with other CNNs and ViTs, VanillaNet-5 offers competitive performance with fewer parameters than ResNet50 (23.56M parameters) and other deeper models, while also maintaining a lower latency (1.02 ms vs. 2.01 ms for ResNet50). This makes VanillaNet-5 an efficient choice for plant recognition tasks, particularly when computational resources and inference time are constraints.

Ablation study

The ablation study evaluates the impact of deep training and series activation on VanillaNet-5's performance. As shown in Table 2, when only series activation was used, the model achieved poor performance with 15.678% accuracy and an F_1 score of 0.094. Using deep training alone improved the performance to 35.169% accuracy and an F_1 score of 0.210. However, the best results were obtained when both deep training and series activation were applied together, reaching 54.237% accuracy and an F_1 score of 0.429. These results confirm that combining both strategies significantly improves the model's performance, demonstrating their importance for effective plant recognition.

Conclusion

In this paper, we proposed VanillaNet-5, a lightweight and efficient model for plant disease identification, designed to operate effectively in resource-constrained environments. Experimental results on the PlantDoc dataset demonstrated that VanillaNet-5 achieved competitive performance, with an accuracy of 54.237% and an F_1 score of 0.429, outperforming other models in both accuracy and computational efficiency. The ablation study highlighted the importance of combining deep training and series activation in optimizing model performance. This work provides a practical solution for automated plant disease detection, offering a balance between performance and resource efficiency, and paving the way for future advancements in smart agriculture.

References

Chang, B.; Wang, Y.; Zhao, X.; Li, G.; and Yuan, P. 2024. A general-purpose edge-feature guidance module to enhance vision transformers for plant disease identification. *Expert Syst. Appl.*, 237(PC).

Chen, H.; Wang, Y.; Guo, J.; and Tao, D. 2023. VanillaNet: the Power of Minimalism in Deep Learning. arXiv:2305.12972.

Chunshan Wang, H. W. e. a., Ji Zhou. 2020. Improved Multi-scale ResNet for Vegetable Leaf Disease Identification. *Transactions of the Chinese Society of Agricultural Engineering*, 36(20): 209–217.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929.

J. P. S. Schuler, M. A.-N. e. a., S. Romani. 2022. Grouped Pointwise Convolutions Reduce Parameters in Convolutional Neural Networks. In *Mendel*, volume 28, 23–31.

Mingyue Shao, Q. F. X. C.-N. Z. W. Z., Jianhua Zhang. 2022. Research Progress on Deep Learning for Plant Leaf Disease Detection and Recognition. *Smart Agriculture*, (004-001).

Ocampo, A. L. P. d.; and Dadios, E. P. 2018. Mobile Platform Implementation of Lightweight Neural Network Model for Plant Disease Detection and Recognition. In *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*, 1–4.

Pal, A.; and Kumar, V. 2023. AgriDet: Plant Leaf Disease severity classification using agriculture detection framework. *Engineering Applications of Artificial Intelligence*, 119: 105754.

Rodríguez-Lira, D.-C.; Cordova-Esparza, D.-M.; Alvarez-Alvarado, J.; Terven, J.; Romero-González, J.-A.; and Rodriguez, J. 2024. Trends in Machine and Deep Learning Techniques for Plant Disease Identification: A Systematic Review. *Agriculture*.

Saleem, M.; Potgieter, J.; and Arif, K. 2020. Plant Disease Classification: A Comparative Evaluation of Convolutional Neural Networks and Deep Learning Optimizers. *Plants*, 9: 1319.

Singh, D.; Jain, N.; Jain, P.; Kayal, P.; Kumawat, S.; and Batra, N. 2020. PlantDoc: A Dataset for Visual Plant Disease Detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, CoDS COMAD 2020, 249–253. New York, NY, USA: Association for Computing Machinery. ISBN 9781450377386.

Stevens, S.; Wu, J.; Thompson, M. J.; Campolongo, E. G.; Song, C. H.; Carlyn, D. E.; Dong, L.; Dahdul, W. M.; Stewart, C.; Berger-Wolf, T.; Chao, W.-L.; and Su, Y. 2024. BioCLIP: A Vision Foundation Model for the Tree of Life. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 19412–19424.

Wanjie Liang, C. S. H. C.-W. Z., Jing Cao. 2023. Development of a Crop Disease Identification System Based on Deep Learning. *Journal of China Agricultural Mechanization*, 44(9): 169–175.

Xu, M.; Park, J. E.; Lee, J.; Yang, J.; and Yoon, S. 2023. Plant Disease Recognition Datasets in the Age of Deep Learning: Challenges and Opportunities. arXiv:2312.07905.

Ye, M.; Ke, L.; Li, S.; Tai, Y.-W.; Tang, C.-K.; Danelljan, M.; and Yu, F. 2023. Cascade-DETR: Delving into High-Quality Universal Object Detection. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 6681–6691.