**Augmentation strategies for Plant Disease Classification**

BY

CHIA WAN JUN

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION TECHNOLOGY (HONOURS) COMMUNICATIONS AND NETWORKING

Faculty of Information and Communication Technology

(Kampar Campus)

FEBRUARY 2025

# COPYRIGHT STATEMENT



This Final Year Project report is submitted in partial fulfillment of the requirements for the degree of Bachelor of Information Technology (Honours) Communications and Networking at Universiti Tunku Abdul Rahman (UTAR). This Final Year Project report represents the work of the author, except where due acknowledgment has been made in the text. No part of this Final Year Project report may be reproduced, stored, or transmitted in any form or by any means, whether electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the author or UTAR, in accordance with UTAR's Intellectual Property Policy.

# ACKNOWLEDGEMENTS

I would like to express thanks and appreciation to my supervisor, Dr. Tan Hung Khoon and my moderator, Dr. Abdulrahman Aminu Ghali who have given me a golden opportunity to involve on the deep learning field study. Besides that, they have given me a lot of guidance in order to complete this project. When I was facing problems in this project, the advice from them always assists me in overcoming the problems. Again, a million thanks to my supervisor and moderator.

# ABSTRACT

This study explores the effectiveness of various data augmentation strategies for enhancing plant disease classification using the LeafGAN model. We propose a novel approach that integrates leaf region and disease symptom masking to improve the quality of synthetic images and, consequently, the performance of plant disease models. Three different configurations of the LeafGAN model were tested, with each model applying distinct masking techniques: **LeafGAN with LFLSeg** uses basic LeafGAN outputs, **SingleMask-LeafGAN** applies leaf region masking to isolate the leaf from the background, and **DualMask-LeafGAN** combines both leaf region and disease symptom masking for enhanced disease simulation. The models were evaluated based on their ability to generate realistic disease progression and recovery images, which were then used for data augmentation. Results show that **DualMask-LeafGAN**, incorporating both masking strategies, produced the most realistic and high-fidelity images, leading to superior augmentation quality. These findings highlight the potential of advanced data augmentation strategies in improving plant disease simulation, emphasizing the importance of targeted feature masking in enhancing the generalization and robustness of disease classification models in agricultural applications.

Area of Study: Deep Learning for Image Segmentation and Classification, Computer Vision in Agriculture

Keywords: Data Augmentation, Deep Learning, LeafGAN, Image Masking, Plant Disease Detection

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| *GAN* | Generative Adversarial Networks |
| *LFLSeg* | Label-free Leaf Segmentation |
| *LPIPS* | Learned Perceptual Image Patch Similarity |
| *FID* | Fréchet Inception Distance |
| *YOLO* | You Only Look Once |
| *IoU* | Intersection Over Union |
| *mAP* | Mean Average Precision |
| *SM-LeafGAN* | SingleMask-LeafGAN |
| *DM-LeafGAN* | DualMask-LeafGAN |

# Chapter 1

# Introduction


Figure 1.1 :Examples of plant diseases from the PlantVillage dataset [1].

Plant disease classification plays a vital role in securing agricultural productivity and food supply. Traditionally, disease identification has relied on manual observation, which is time-consuming, labor-intensive, and requires expert knowledge making it impractical for large-scale farming. These diseases, often caused by pathogens such as viruses, bacteria, and fungi, typically exhibit visible symptoms like yellowing, spots, or galls on plant leaves. However, relying solely on visual inspection is not only inefficient but also prone to human error, especially when symptoms are subtle or resemble multiple diseases. In response to these limitations, automated classification using artificial intelligence (AI) has emerged as a faster, more scalable, and more accurate alternative [2].

Recent advances in deep learning have significantly enhanced image-based plant disease detection. Pre-trained models utilizing techniques such as Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Diffusion models have achieved promising performance by increasing both speed and accuracy. However, the success of these models heavily depends on the quality and diversity of training datasets.

## 1.1 Problem Statement and Motivation

Acquiring high-quality plant disease datasets is hindered by environmental challenges, such as variable lighting, leaf overlap, and background noise, which complicate accurate data collection. These factors often lead to inconsistent image quality, making it difficult to capture clear and reliable visual symptoms of diseases. Additionally, seasonal variations and geographic differences further limit the ability to gather standardized and comprehensive datasets.

A major challenge in plant disease classification is the lack of diverse and balanced datasets. Many publicly available datasets contain an uneven distribution of disease classes, with some conditions significantly underrepresented. Training models on such imbalanced data introduce bias, often resulting in overfitting and poor generalization to real-world scenarios.

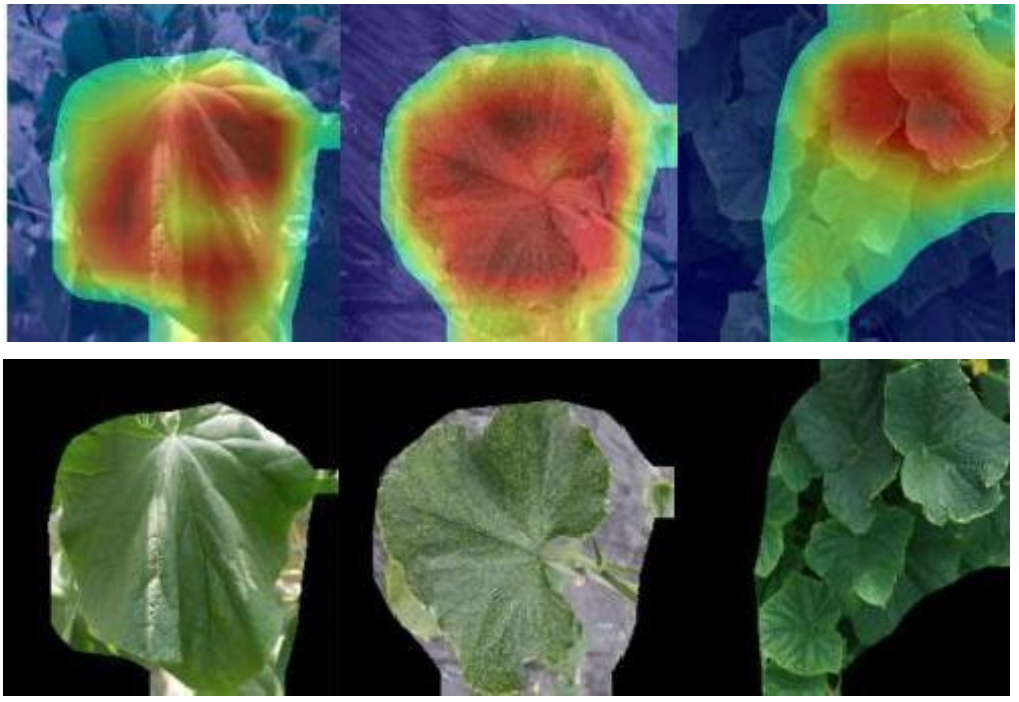


Figure 1.2: Examples of leaf segmented results from LeafGAN [3].

While basic augmentation techniques like rotation, flipping, and color variation increase dataset size, they often fail to replicate the intricate and diverse manifestations of plant diseases in real-world conditions. To overcome these limitations, advanced GAN-based methods, such as CycleGAN [4] and LeafGAN [3], have been developed for unpaired image-to-image translation to generate synthetic diseased leaf images. LeafGAN, for instance, employs Grad-CAM to generate a heatmap for rough leaf region localization and employs CycleGAN to train two models: one to superimpose disease symptoms on healthy leaves and another to remove disease symptoms from diseased leaves. However, the quality of the transformed images is heavily dependent on the accuracy of the segmentation algorithm, as Grad-CAM struggles to isolate leaf regions in images with multiple or overlapping leaves. This often results in the inclusion of background elements in the heatmap, leading to low-quality augmentations that compromise the realism of the generated images.

This research is motivated by the need to generate higher-quality synthetic training data through improved segmentation and masking strategies. Specifically, this study replaces the segmentation module in LeafGAN with a YOLOv5-based segmentation model to improve leaf

localization and masking precision. This is expected to produce more realistic augmented images that better support disease classification tasks.

## 1.2 Objectives

**Primary Objective:**

To enhance the diversity of existing plant disease datasets by applying advanced data augmentation techniques based on disease translation methods.

**Specific Objectives:**

(a) To study the impact of different leaf segmentation strategies on the quality of augmented disease images generated by LeafGAN. The current label-free segmentation method, LFLSeg, has shown limitations and produced suboptimal results.

(b) To propose a novel plant disease augmentation method incorporating alternative segmentation strategies aimed at improving the realism and reliability of synthetic images.

## 1.3 Contributions

This project's contributions include:

- This work demonstrates that segmentation quality is a crucial factor in enhancing the realism of GAN-generated synthetic data for plant disease augmentation. Experiments confirm that employing the YOLOv5 model for leaf localization significantly improves segmentation accuracy, resulting in higher-quality augmentations and generating promising results compared to traditional Grad-CAM-based approaches.

- A novel dual masking strategy is introduced to optimize disease augmentation. For healthy-to-diseased transformations, leaf segmentation is utilized to enable LeafGAN to apply disease symptoms accurately across any part of the leaf. Conversely, for diseased-to-healthy transformations, disease-region detection is employed to ensure precise targeting of affected areas, improving the model's ability to transform the correct regions. Experiments validate that this dual masking approach enhances the fidelity of bidirectional transformations.

## 1.4 Report Organization

This report is organised into 6 chapters: Chapter 1 Introduction, Chapter 2 Literature Review, Chapter 3 Proposed Method, Chapter 4 Experiment and Chapter 5 Conclusion. The first chapter is the introduction of this project which includes problem statement, project background and motivation, project scope, project objectives, project contribution, highlights of project achievements, and report organisation. The second chapter is the literature review carried out on several existing disease translation models using different methods and the strengths and weaknesses of each product. The third chapter discusses the overall system design of this project. The fourth chapter concerns the details on how to implement the design of the system and reports the results obtained from the experiments. The fifth chapter summarizes the key findings of the project, highlights the main contributions, and suggests directions for future research to further improve data augmentation for plant disease classification.

# Chapter 2
# Literature Review

In this section, we will explore GAN-based and diffusion-based models in the augmentation of plant disease datasets.

## 2.1 Previous Works on GANs-based model

Generative Adversarial Networks (GAN) is a minimax two-player game where both players learn from each other and attempt to outperform the other [5]. The model includes two main structures: a generator and a discriminator. The generator produces synthetic images that closely resemble real ones to challenge the discriminator, whose task is to differentiate between synthetic and real images. Besides, GAN models have been widely used in various areas such as the medical field [6, 7], gaming industry [8, 9] and art sector especially in style transfer [10, 11, 12]. The following research aims to extend the application of GAN models to the agriculture sector.

### 2.1.1 Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

CycleGAN [4] uses unpaired datasets for generating synthetic images, which has revolutionized the data augmentation method. Typically, the usage of paired datasets assists in generating the output image by learning the mapping between the paired datasets. For instance, health leaf images and disease leaf images are using simultaneously to guide the model training. However, this approach faces a crucial bottleneck due to the scarcity and difficulty in collecting paired data samples. CycleGAN introduces the unpaired dataset to address this problem. The main concept is to learn the special features from one image and apply them to other images without any paired training data.

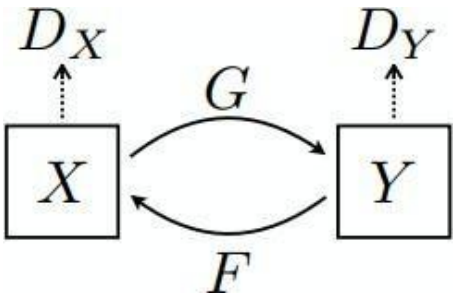



Figure 2.1: Overview of CycleGAN's architecture.

The process of the model is depicted in Figure 2.1. There are two different domain datasets, X (e.g., healthy plant dataset) and Y (e.g., diseased plant dataset). Additionally, generator G aids in translating the dataset from domain X to domain Y, while the reverse process is conducted using generator F. Two discriminators, Dx and Dy, distinguish between real images and synthetic images.

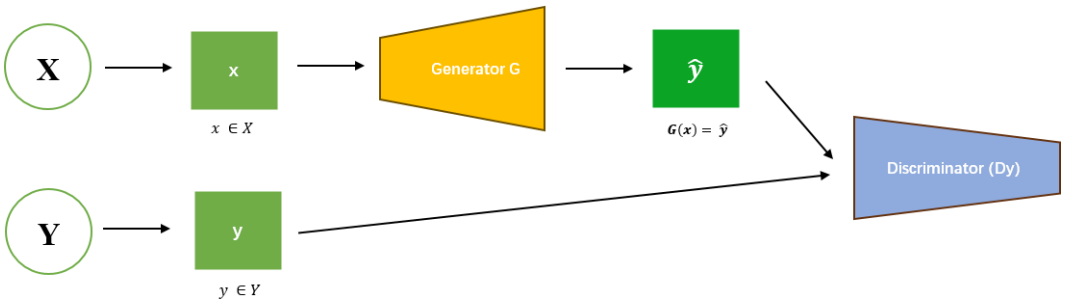



Figure 2.2: Overview of the translation Framework.

For instance, a healthy plant image $x$ is translated to a disease plant image though the generator G, i.e., $y \rightarrow G(x) = y\hat{}$. Then, Discriminator Dy identifies the synthetic images $y\hat{}$ among the original disease images $Y$.

**Loss function**

The final objective function includes the adversarial loss function for both the generator G and F, as well as a novel loss function called cycle consistency loss.

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &+ \lambda \mathcal{L}_{\text{cyc}}(G, F),\end{aligned} \tag{1}$$

The ***adversarial loss function*** (LGAN) encourages the discriminators to correctly classify synthetic images, maintaining the overall performance of the model. The generator G minimizes the difference between the real and fake images and generates compelling synthetic images, while discriminator Dy maximizes its ability to identify the results correctly. The formula for the adversarial loss is expressed as follows:

$$\begin{aligned}\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) =& \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] \\ &+ \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))]\end{aligned} \tag{2}$$

Utilizing unpaired datasets leads to a lack of guidance in image generation. Therefore, CycleGAN introduces a new loss function called ***cycle consistency loss*** to ensure successful translation though the process.

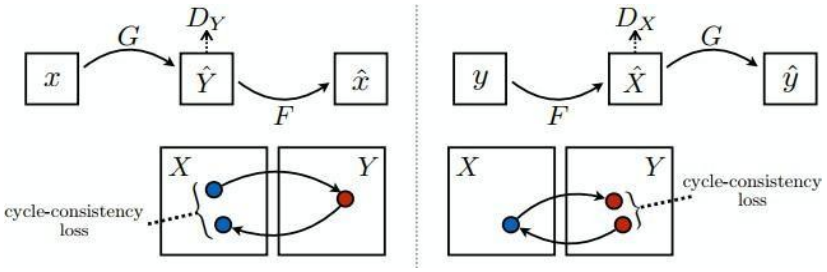



Figure 2.3: Cycle-consistency loss.

This model employs bidirectional translation concepts, meaning that if a horse image translates to the zebra, it can also translate back to the horse image. Similarly, the original image $x$ will be translate to $G(x) = y\hat{}$ then transfer back to the original domain $F(G(x)) \approx x$, as shown in Figure 2.3. Further explanation of the cycle consistency loss is to compare the distribution distance between the reconstructed image $x$ and original image $x$ to guarantee the similarity of them.

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]. \quad (3)$$

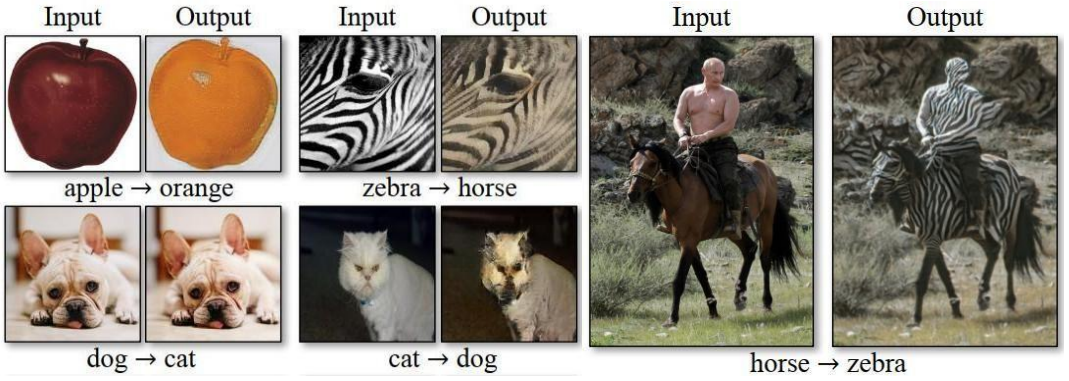



Figure 2.4: The failure cases of translating the images using CycleGAN.

Although the proposed model addresses the problem of insufficient datasets, it also has some limitations. For example, while CycleGAN is successful in translating colour and texture images, it struggles with geometric changes and various transformations, as depicted in Figure 2.4. Additionally, CycleGAN translates the entire image, which can lead to the generation of undesirable datasets. For instance, when translating a horse into a zebra, the human riding on it may also be affected due to the lack of masking or attention mechanism.

### 2.1.2 LeafGAN: An Effective Data Augmentation Method for Practical Plant Disease Diagnosis

Previous paper, CycleGAN, faces an issue in generating high-quality synthetic images and tends to transform the entire image into another image. Therefore, LeafGAN [3] is proposed to improve the existing problem found in CycleGAN. LeafGAN inherits the concept of unpaired datasets from CycleGAN and utilizes data augmentation methods to expand the dataset for diagnosing plant diseases.

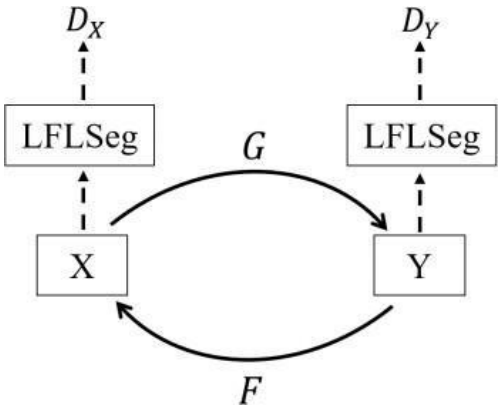



Figure 2.5: The architecture of LeafGAN.

Similar to the architecture of CycleGAN, which uses a GAN-based model and bidirectional generation in healthy plant image and arbitrary disease plant image. Moreover, this model integrates a label-free leaf segmentation module (LFLSeg) to emphasize the important regions of the dataset as shown in Figure 2.5.

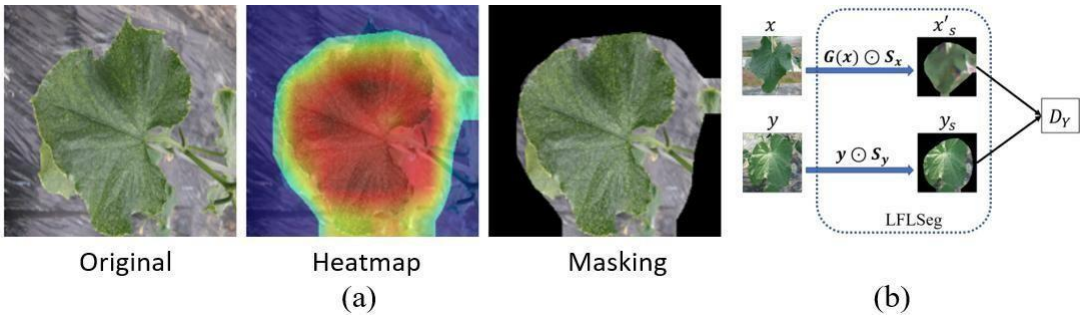



Figure 2.6: Result and segmentation process of LFLSeg model.

LFLSeg is employed to mask undesired regions, such as the background, with a focus on translating the symptoms of diseases in images. It uses Gram-CAM, which generates a heatmap to visualize the important regions on the image through specific colors, as shown in Figure 2.6 (a). Compared to traditional masking methods, it more accurately highlights the target.

The segmentation process is depicted in Figure 2.6 (b). A synthetic image is generated and after masking, it is added to the dataset pool to mislead the discriminator. Initially, a healthy image, denoted as $x$, is transformed into a synthetic disease image though the generator G which stimulates the symptoms. This transformation is represented as $G(x) = x'$. Following masking in LFLSeg, it is denoted as $G(x) \odot S_x = x'_s$ In contrast, another dataset $y$ exclusively masks the background of images, identified as $y \odot S_y = y_s$. Then, the discriminator $D_Y$ is tasked with identifying synthetic images.

**Loss function**

The final objective function includes mainly three different type of loss function: Adversarial Loss, Cycle Consistency Loss and Background Similarity Loss and defined as follow:

$$\mathcal{L}(G, F, D_X, D_X) = \mathcal{L}_{\text{adv}}(G, D_Y) + \mathcal{L}_{\text{adv}}(F, D_X) + \lambda[\mathcal{L}_{\text{cyc}}(G, F) + \mathcal{L}_{\text{bs}}(G, F)], \quad (4)$$

For both the ***adversarial loss*** and ***cycle consistency loss***, we find similarities with the CycleGAN loss function. Therefore, our focus in this paper is on introducing background ***similarity loss (Lbs).*** The core concept is to maintain the integrity of the background during transformation. This means that the synthetic background images should closely resemble their original counterparts. This is achieved by calculating the distance between the generated background images and the original ones, allowing us to identify any differences between them.

$$\mathcal{L}_{\text{bs}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[|(1 - S_x) \odot (G(x) - x)|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[|(1 - S_y) \odot (F(y) - y)|_1]. \quad (5)$$

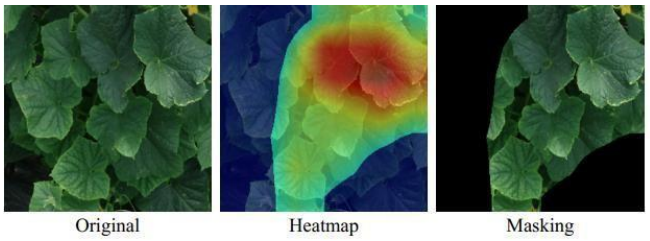



Figure 2.7: The failure result in segmentation.

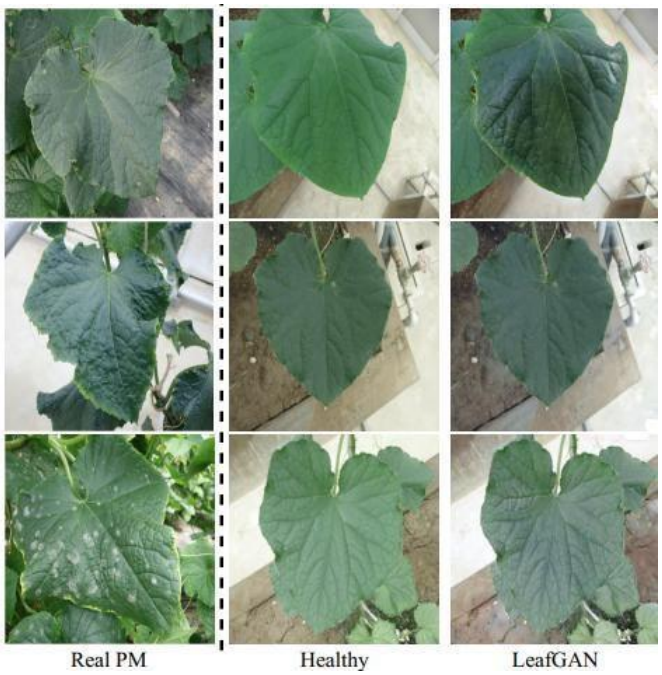



Figure 2.8: Failure result in translating the powdery mildew disease (PM).

The LFLSeg model effectively identifies most of the background and the target region but fails to classify multiple and overlapping leaves, as depicted in the first row of Figure 2.7. Additionally, the LeafGAN model encounters difficulties when dealing with powdery mildew (PM) disease, which exhibits complex characteristics across different disease stages. Consequently, it generates images with altered colours and insufficient representation of PM symptoms. In Figure 2.8, the first and second rows illustrate symptoms of powdery mildew in the first and middle stages, appearing as a darker blue colour. However, the later stage of PM disease (last row) is characterized by white spots. While this stage is typical, the problem arises from the limited examples in the training dataset.

## 2.2 Previous Works on diffusion-based model

The diffusion model comprises two main processes: the forward process, involving the addition of noise, and the backward process, also known as denoising, which removes the noise until the image becomes clear. This approach generates higher-quality and more diverse images compared to GANs models [13].

### 2.2.1 Denoising Diffusion Probabilistic Models

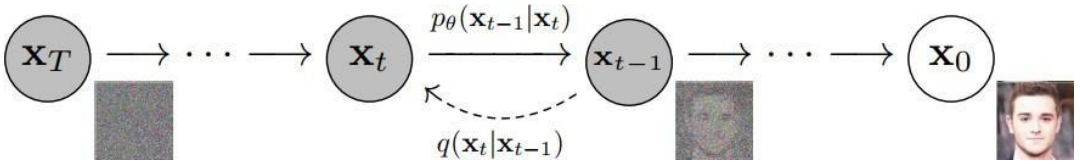



Figure 2.9: The process of Denoising Diffusion Probabilistic Models.

Denoising Diffusion Probabilistic Models (DDPM) [14] is a type of diffusion model, specifically a likelihood generative model, designed to produce images that belong to a similar distribution. In the forward process, Gaussian noise is gradually added to an input $x_0$ until it becomes a fully noised image $x_T$. This process is expressed as follows:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \tag{6}$$

Here, $x_t$ is the image at time step $t$, added with Gaussian noise following the variance $\beta_t$ at specific timestep t and scaling $x_{t-1}$ by $\sqrt{1-\beta_t}$. In other word, given an $x_{t-1}$ one can predict $x_t$.

Using the reparameterization trick, sampling from a standard Gaussian distribution and then scaling and shifting this sample using the model's learnable parameters, this equation can also be written as:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}) \tag{7}$$

This means $x_t$ is the combination of the $x_{t-1}$ with the noise $\epsilon$. This formula also predicts that a similar situation exists where $x_{t-1}$ consists of $x_{t-2}$ with noise $\epsilon$. Hence, it shows that an image at arbitrary timestep $x_t$ is represented as $x_0$ with noise $\epsilon$.

The backward process of the diffusion model tries to reconstruct the fully noised image $x_T$ into a clear noise image $x_0$. This process predicts the means and covariance matrix of Gaussian distribution as expressed in this equation:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)) \tag{8}$$

This model uses ***variational lower bound loss*** and is simplified as follows:

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2\right] \quad (9)$$

This equation aims to measure the difference between the noise added to the image $\epsilon$ and the noise generated by the diffusion process $q(x_t,\ t)$ at arbitrary time step t. Using this equation, it guides the diffusion process to generate positive images.

Similar to most diffusion models, a significant limitation of this model is the lengthy training time, as it necessitates 1000 timesteps to generate high-quality images. Moreover, while this model shows promise in improving image generation quality, it does not currently surpass existing models.

### 2.2.2 RePaint: Inpainting using Denoising Diffusion Probabilistic Models


Figure 2.10: Failure case of semantic meaningful generation.

Image inpainting is a technique that aids in recovering missing areas in an image. However, the existing approaches are limited to training on certain shapes of masks, and the generated images sometimes lack semantic meaningful changes, resulting in undesirable outcomes. For example, if we attempt to change the nose of a dog, it may translate into a different type of nose, such as that of a horse or another animal, as shown in Figure 2.10. Hence, RePaint, a model using the DDPM as the backbone and resample method is proposed [15]. It aims to generate different kinds of mask on training and enhance the accuracy of semantic generation ability.

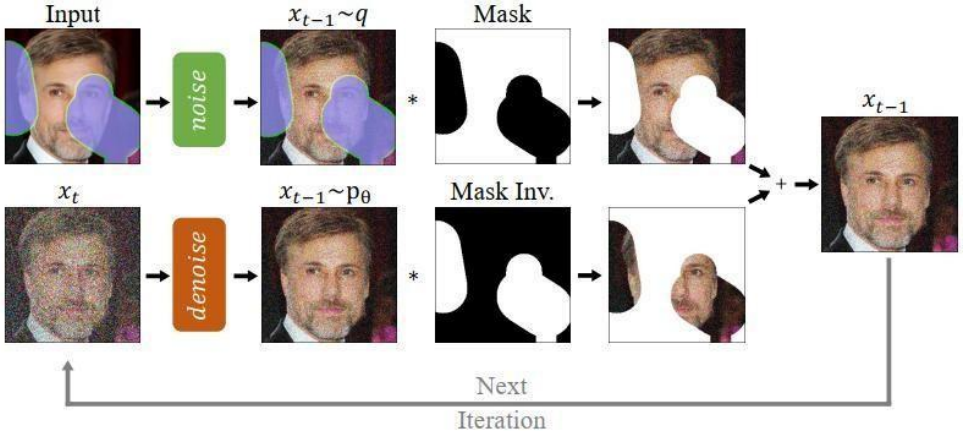



Figure 2.11: Overview of RePaint's architecture.

Figure 2.11 depicts the architecture of RePaint, which shares a similar process with DDPM in both the forward and backward processes. It identifies the image at arbitrary timesteps and combines it with a mask before proceeding to another iteration. In the forward process, the input image gradually accumulates Gaussian noise, using Equation (10) from the DDPM paper to predict intermediate image $x_t$ and $x_{t-1}$ at any timestep:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}) \tag{10}$$

During the forward process, an image $x_{t-1}$ gradually accumulates Gaussian noise, denoted as $x_{t-1}^{know}$, and the noise addition process is expressed as follows:

$$x_{t-1}^{\text{known}} \sim \mathcal{N}(\sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}) \tag{11}$$

The reverse process utilizes the image $x_t$ to denoise the image using a noise prediction equation inherited from DDPM and represents it as $x_{t-1}^{unknow}$. The denoising process is expressed as follow:

$$x_{t-1}^{\text{unknown}} \sim \mathcal{N}(\mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{12}$$

The $x_{t-1}^{know}$ is masked with the desirable area $m$ and $x_{t-1}^{unknow}$ is marked with an inverted mask $(1 - m)$. These two images are combined and indicated as follows:

$$x_{t-1} = m \odot x_{t-1}^{\text{known}} + (1 - m) \odot x_{t-1}^{\text{unknown}} \tag{13}$$

After combining the two images, this image serves as the input image during the next iteration until it outputs a fully denoised image.


| | T | r | LPIPS | T | r | LPIPS | T | r | LPIPS | T | r | LPIPS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Slowing down | 250 | 1 | 0.168 | 500 | 1 | 0.167 | 750 | 1 | 0.179 | 1000 | 1 | 0.161 |
| Resampling | 250 | 1 | 0.168 | 250 | 2 | 0.148 | 250 | 3 | 0.142 | 250 | 4 | 0.134 |

Figure 2.12: Result of using slow-down and resampling methods.

The known pixels are generated without considering the filling part, resulting in disharmony in the generated image. Additionally, the change in the image is hindered by the decrease in variance $\beta_t$ at each denoising step. The paper compares the slowing-down and resampling methods to solve this problem. The slowing-down approach involves using more time steps to generate one image, while the resampling method attempts to regenerate the image multiple times. Moreover, Learned Perceptual Image Patch Similarity (LPIPS) is used to evaluate the distance between the input and output images, where a higher number indicates more difference. The results show that despite the higher computational cost, resampling yields significantly better results compared to the slowing-down approach as shown in Figure 2.12.

Moreover, this model has the potential to further enhance the generative process time because the processing time, which necessitates at least 250 timesteps, exceeds that of the GAN model.

However, several models focus on addressing this issue, as evidenced in studies [16, 17]. Additionally, the disparity between the ground truth image and the generated image presents challenges in evaluation due to the enhancement of arbitrary masking during training, resulting in various output images. For example, the current evaluation metric, LPIPS, compares the diversity of the input and output images. Therefore, the utilization of Fréchet inception distance (FID) provides an alternative for analysing the quality and diversity of the images. However, FID typically requires evaluating images exceeding 1000, rendering this evaluation method infeasible for the current DDPM model.

### 2.2.3 Effective Data Augmentation with Diffusion Models

The current data augmentation techniques, such as rotation or flip, are limited in their ability to transform images with diverse and intricate features. For example, when applying these techniques to a car image, they often only alter the color or produce similar images from the dataset, failing to transform aspects like texture, brand, or unique structural elements. Additionally, pretrained models trained on large datasets like ImageNet can suffer from data leakage issues, as they tend to generate images that closely resemble the benchmarking dataset. Therefore, this paper proposes a text-to-image diffusion model (DA-Fusion) to increase diverse datasets using a more comprehensive data augmentation strategy [18].

To prevent the leakage of internet data, this model applies two approaches: a model-centric approach, which fine-tunes the weights of the model to remove class information, and a data-centric approach, which changes the class names to remove class knowledge.

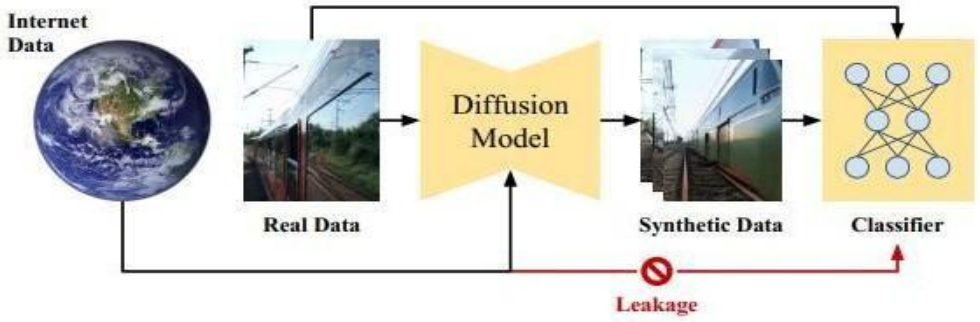



Figure 2.13: Leakage of Internet Data.

The model-centric method modifies the Stable diffusion model to eliminate concepts from the original dataset. Additionally, it fine-tunes the UNet, which is the backbone of the Stable diffusion model, to remove the effect of the prompt (class name). Consequently, it adjusts the loss function, as outlined in the Erased Stable Diffusion (ESD) [19], to assist in removing the concept of the original class name.

$$\min_{\theta} \mathbb{E}\left[\left\|\epsilon_{\theta}(x_t, t, \text{"class name"}) - \epsilon_{\theta^*}(x_t, t) + \eta(\epsilon_{\theta^*}(x_t, t, \text{"class name"}) - \epsilon_{\theta^*}(x_t, t))\right\|^2\right] \quad (14)$$

It aims to minimize the difference between the adjustment effect of the image given a specific class name compared to without modification. This helps the generation of the image to be less likely transformed into the specific image.

For the data-centric method to prevent dataset leakage, it changes the prompts of the original class name to a new prompt without any clues of the class name. By removing the class name from the prompt, the model is forced to rely on other contextual cues or general knowledge

rather than specific class information.

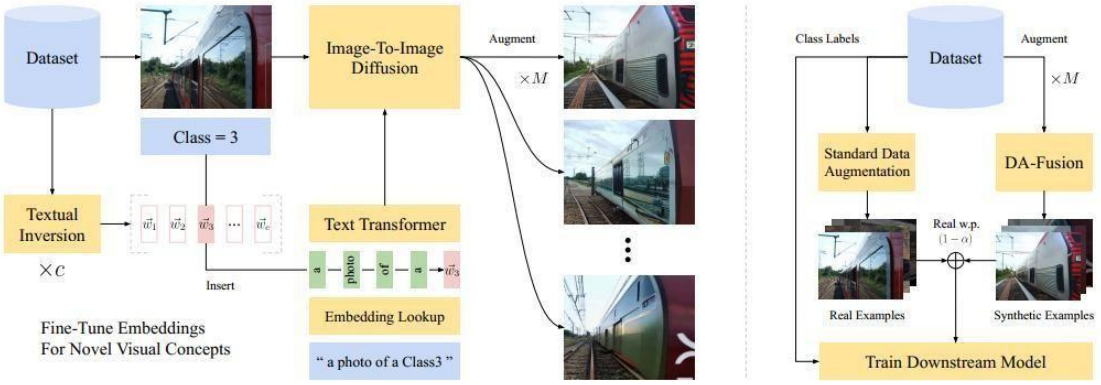



Figure 2.14: Process of data augmentation using DA-Fusion

Generally, the training process combines the dataset of the images and class labels to generate M versions of the augmented images. It applies the word-embedding method and uses the pretrained DDPM model throughout the entire process. Additionally, it trains the downstream model using the synthetic image and original images.

To guide the model in learning the new concept to generate diverse images, this model embeds a text encoder to learn the new words. It uses a word embedding vector $w\rightarrow\rightarrow\rightarrow_\iota$ to represent each class using a small number of labelled images. The loss function optimizes the learning process as shown below:

$$\min_{\vec{w}_0,\vec{w}_1,\ldots,\vec{w}_c} \mathbb{E}\left[\|\epsilon - \epsilon_\theta(\sqrt{\tilde{\alpha}_t}x_0 + \sqrt{1-\tilde{\alpha}_t}\epsilon, t, \text{"a photo of a } \vec{w}_i\text{"})\|^2\right] \tag{15}$$

It aids in generating a more accurate image by enhancing understanding through word embedding for a certain class. Moreover, this model employs an approach that splices real images as guidance for training the model. During the reverse process, it introduces another real image to guide the denoising process until it becomes clear.

$$x_{\lfloor St_0 \rfloor} = \sqrt{\tilde{\alpha}_{\lfloor St_0 \rfloor}}x_0^{\text{ref}} + \sqrt{1-\tilde{\alpha}_{\lfloor St_0 \rfloor}}\epsilon \tag{16}$$

At arbitrary timestep $St_0$ during the reverse process, a real image $x_0^{ref}$ is added as a reference during the generation and the denoising of Gaussian noise $\epsilon\sim N$ (0,1) continues until the timesteps are completed.

This model not only achieves success in few-shot classification but also in real-world weed recognition. However, it faces challenges in controlling the modification of the image. For instance, it cannot explicitly control the transformation of the image, such as changing the breed of a cat. Hence, prompt-based image editing has been suggested to address this limitation,

enabling localized changes without supervision. Moreover, the paper has mentioned that maintaining temporal consistency is crucial, as it assists in transforming realistic images across various background environments.

### 2.2.4 Patch Diffusion: Faster and More Data-Efficient Training of Diffusion Models

The diffusion model consumes a lot of time during training and is slow in generating an image due to the forward and backward processes. Additionally, it also requires a large dataset to train the model to achieve good results. This paper proposes Patch Diffusion, patch-wise diffusion training to address these two problems [20]. The core concept of this model focuses on smaller patches of the whole image, which decreases processing time and burden instead of adding noise throughout the entire image. Unlike conventional methods focusing on image distribution learning, Patch Diffusion adopts score-based generative modelling, prioritizing the assessment of scores, whether good or bad, over image generation distribution.

This model comprises three key components. Firstly, it randomly partitions the image into patches, incorporating patch location and size as adjustable parameters. Secondly, it integrates a pixel coordinate system to depict patch locations accurately. Lastly, it streamlines the reverse step, eliminating the need for separate patch sampling and merging, thereby enhancing efficiency and simplicity.

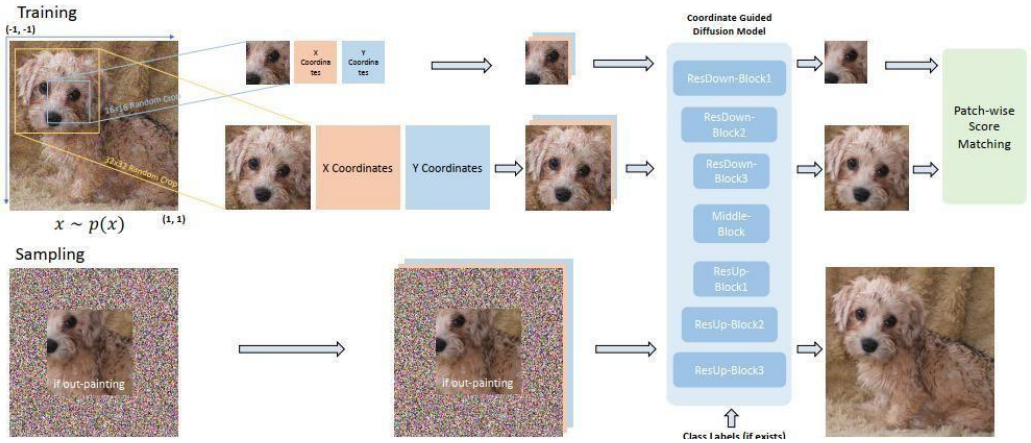



Figure 2.15: Architecture of Patch Diffusion.

During the training phase, the original image is cropped into larger and smaller sizes. Both images are indicated with $x_{i,j,s}$ where $(i,\ j)$ represents the left-upper coordinate of the patches and $s$ represents the size. The usage of different-sized patches helps capture the cross-dependency of the entire image. Additionally, the model employs an out-painting strategy to sample a new image, involving generating content beyond the boundaries of the input image. This is crucial as the full-size image provides the global score to aid in generating a high-quality image. A coordinate-guided diffusion model is utilized to incorporate spatial information through pixel coordinates, enhancing the generation process. This model concatenates the patches obtained from the previous step to provide guidance, serving as coordinate conditions during each reverse iteration.

The patch size selection is divided into two parts: stochastic and progressive. The stochastic method involves randomly cropping the image to create mini batches for training the model. Conversely, the progressive method increases the patch size from small to large during each iteration, as shown below:

$$s \sim p_s := \begin{cases} p & \text{when } s = R, \\ \frac{3}{5}(1-p) & \text{when } s = R//2, \\ \frac{2}{5}(1-p) & \text{when } s = R//4. \end{cases} \tag{17}$$

The resolution of the original image is denoted as $R$, and the ratio of the iteration is $p$. There are three situations: firstly, when the patch size is one-fourth of the original image (denoted as $R//4$), it selects the probability $\frac{2}{5}(1-p)$ of the iteration. Secondly, when the patch size is half of the original image, the probability $\frac{3}{5}(1-p)$ of the remaining iteration is selected. Lastly, it uses the original image throughout the total iteration. $p = 0.5$ is found to show that the scheduling trades off the efficiency of training time and the quality of the generated image. Furthermore, the model learns the score function on random-size patches rather than the entire image. The denoising score function is expressed as:

$$\mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}),\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\sigma_t^2\boldsymbol{I}),(i,j,s)\sim\mathcal{U}}\left|\left|D_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}}_{i,j,s};\sigma_t,i,j,s)-\boldsymbol{x}_{i,j,s}\right|\right|_2^2, \tag{18}$$

It aims to minimize Euclidean distance between the denoised patches $D_\theta(\tilde{x}_{i,j,s};\sigma_t,i,j,s)$ with the original patch $x_{i,j,s}$.

Patch diffusion addresses the issue of lengthy training times and achieves a 2x faster training rate compared to other models. Moreover, it demonstrates robust performance across datasets of various sizes, although there is still potential for enhancement in certain aspects of picture accuracy. Furthermore, there is discussion about enhancing the coordinate system by leveraging advanced positional embeddings, such as 5D coordinates encompassing spatial and viewing directions [21].

## 2.3 Limitation of Previous Studies

Previous studies have not extensively utilized diffusion models for augmenting plant datasets. While some research has mentioned generating new datasets in the field of plant datasets, there has been little focus on creating variable field conditions for crop datasets, leaving a gap in the literature. Furthermore, research on data augmentation strategies for plant disease detection is limited and does not comprehensively address different scenarios. For instance, LeafGAN has failed to analyse datasets with multiple layers of leaves, restricting its applicability to broader scenarios.

## 2.4 Proposed Solutions

This project aims to propose a model that employs the GAN model to augment new datasets capable of handling various conditions and backgrounds to address the dataset insufficiency problem. Moreover, this project aims to augment the dataset through two different methods. Firstly, by generating a variety of new healthy and diseased plant images. Secondly, by translating between healthy plant images and diseased plant images, and vice versa. Two different condition datasets have been utilized to highlight the complexity of background datasets, resulting in a more compelling and realistic dataset. Additionally, the transfer learning technique using the Yolov5 model is incorporated to focus on the translating region to aid in generating positive images.

# Chapter 3

# Dual Masking Strategy for LeafGAN

## 3.1 Segmentation Strategies in LeafGAN

The original LeafGAN framework suffers from two key limitations in its segmentation strategy. First, segmentation is only indirectly involved in the generation process. It primarily assists the discriminator by focusing attention on the leaf region during adversarial training, without directly guiding the generator. Consequently, disease translation is learned in a weakly supervised manner requiring the model to implicitly localize both the leaf and disease regions which limits its ability to synthesize accurate and realistic disease patterns. Second, the LFLSeg module used for segmentation often produces low-quality masks. These label-free masks tend to include background pixels or miss parts of the leaf, providing weak supervision that degrades the realism of generated images.

To address these issues, we propose two key enhancements. First, we replace LFLSeg with the **YOLOv5 model** for leaf localization, achieving more precise segmentation that isolates the leaf region and minimizes background interference. Second, we introduce a **dual masking strategy** that generates separate masks for both the leaf and disease regions. This additional spatial information allows the discriminator to focus more accurately on disease-affected areas, enhancing its ability to distinguish real images among synthetic images. As a result of the stronger adversarial feedback, the generator also improves, producing more realistic and disease-accurate outputs.

### 3. 2 DualMask-LeafGAN

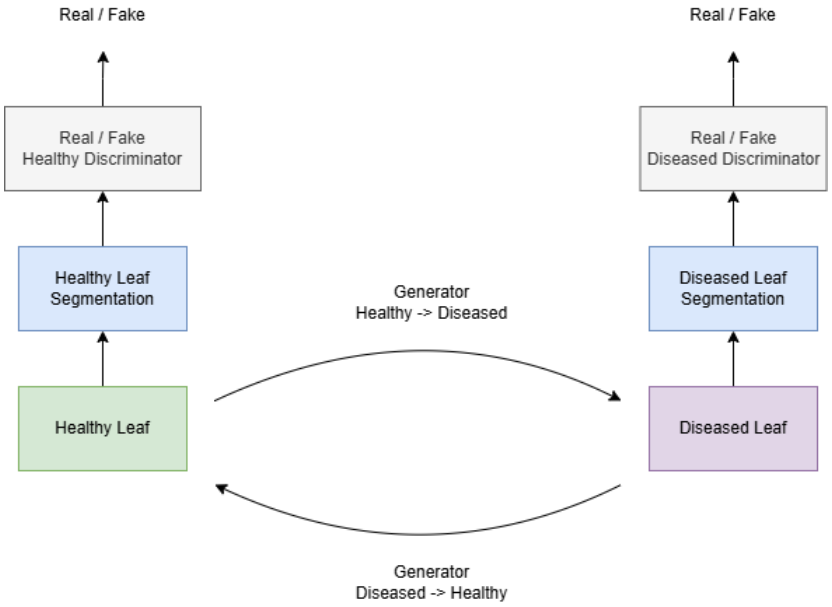



Figure 3.1: Overview of proposed DualMask-LeafGAN.

Figure 3.1 shows the architecture of DualMask-LeafGAN (DM-LeafGAN). DM-Leaf-GAN is an improved image translation framework aimed at generating more accurate and realistic representations of healthy and diseased leaf images. The model operates bidirectionally with two mapping functions, $G: H \rightarrow D$ and $F: D \rightarrow H$, where $H$ and $D$ denote the healthy and diseased leaf domains. Unlike the original LeafGAN, which relies on a single label-free segmentation mask (LFLSeg) for the entire leaf region, this approach employs two distinct segmentation modules for healthy and diseased leaves. These modules generate separate masks for both the leaf and the disease areas using YOLOv5, which provides enhanced precision in localization. When transforming a healthy leaf image $h \in H$ to a diseased version, the generator G creates a synthetic image h', and the corresponding masks $S_h$ and $S_d$ are applied to isolate important regions. The masked images $h'_s$ and real diseased $d_s$ are then fed into the diseased discriminator $D_D$, focusing its attention exclusively on the relevant leaf and disease areas for real or fake discrimination. The reverse transformation from diseased to healthy images follows a similar process with generator F and discriminator $D_H$. By incorporating dual masking and adversarial training, this method allows the discriminators to better identify subtle differences in the disease patterns, encouraging the generators to produce sharper and more faithful image translations. This strategy addresses the shortcomings of the previous single-mask method, improving disease localization and overall visual fidelity.

### 3.3 Loss Function for Dual-Mask LeafGAN

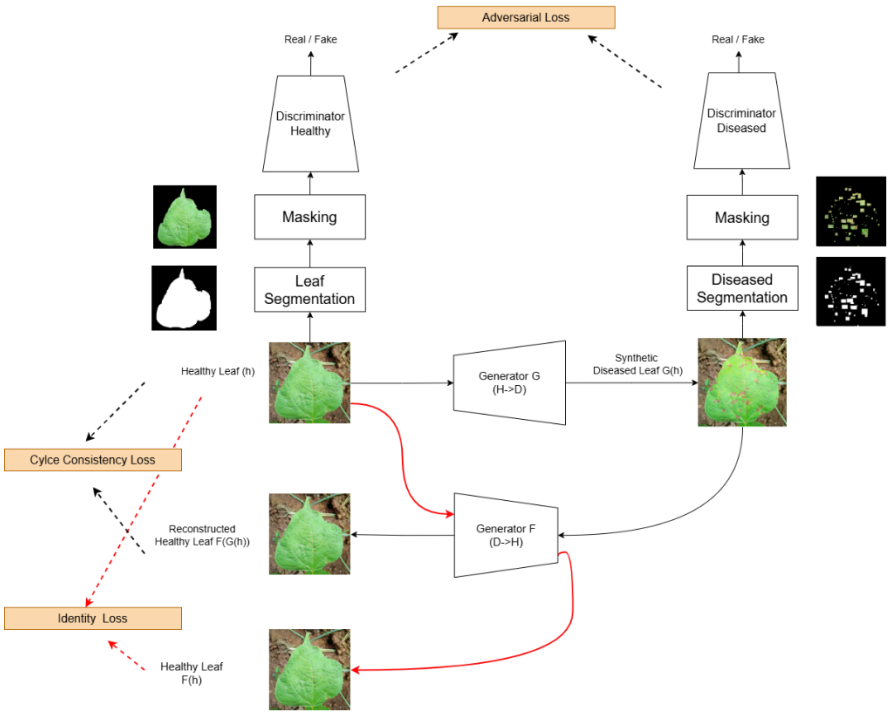



Figure 3.2: Loss functions used in DM-LeafGAN.

Figure 3.2 shows the healthy-to-diseased translation process and the corresponding loss functions. The training of Dual-Mask LeafGAN is guided by a composite loss function, adapted from the original LeafGAN, with additional considerations for dual-mask segmentation. As illustrated in Figure 3.2, the loss function comprises three components: adversarial loss, cycle-consistency loss, and identity loss. Each component plays a specific role in improving the quality and realism of the generated images.

#### A. Adversarial Loss

Adversarial loss encourages the generator to produce synthetic images that are indistinguishable from real images in their respective domains. In Dual-Mask LeafGAN, two discriminators are used one for healthy leaves ($D_H$) and another for diseased leaves $D_D$. The role of each discriminator is to differentiate between real and generated (fake) images within its domain.

As shown in Figure 3.2, generator G learns to transform a healthy leaf image $h \epsilon H$ into a diseased image $G(h) \in D$, which is then evaluated by $D_D$. Simultaneously, the reverse generator F transforms a diseased leaf $d \in D$ into a healthy version $F(d) \in H$, which is assessed by $D_H$. Both discriminators are guided using leaf-specific segmentation masks, enhancing their ability to focus on leaf regions during discrimination. The adversarial loss for

generator G and discriminator $D_D$ is defined as:

$$\mathcal{L}_{adv}(G, D_D) = E_{d\sim D}[\quad (D_D(d_s) - 1)^2] + E_{h\sim H}[(D_D(h'_s))^2]. \tag{19}$$

Similarly, the adversarial loss for F and $D_H$ is defined accordingly.

**B. Cycle-Consistency Loss**

Due to the unpaired nature of the dataset, there is no direct ground truth for the translated images. To address this, a cycle-consistency loss is introduced, ensuring that an image translated to the other domain and back again should resemble the original input. This loss enforces structure and semantic consistency across domains.

In practice, if a healthy leaf $h$ is transformed to diseased $G(h)$ and then reconstructed back to healthy via $F(G(h))$, the result should match the original $h$. The same applies in reverse for diseased to healthy transformations. The cycle-consistency loss is expressed as:

$$\mathcal{L}_{cyc}\mathrm{D}(G, F) = E_{h\sim H}\big[\big|F\big(G(h)\big) - h\big|\big] + E_{d\sim D}\big[\big|G\big(F(d)\big) - d\big|\big]. \tag{20}$$

This loss penalizes discrepancies between original and reconstructed images, thus promoting consistency.

**C. Identity Loss**

Identity loss is introduced to preserve color distribution and low-level features, especially when the input image already belongs to the target domain. It ensures that the generator does not unnecessarily alter images that require no transformation.

For example, feeding a diseased leaf image $d$ into generator G, which is designed for healthy-to-diseased translation, should ideally return the same image without changes. This constraint is critical to maintain visual realism and structural integrity.

The identity loss is defined as:

$$\mathcal{L}_{id}(G, F) = E_{d\sim D}[|G(d) - d|] + E_{h\sim H}[|F(h) - h|]. \tag{21}$$

**Overall Objective**

The total loss combines all three components with weighting parameters $\lambda_{cyc}$ and $\lambda_{id}$:

$$\mathcal{L}_{total} = \mathcal{L}_{adv} + \lambda_{cyc} \cdot \mathcal{L}_{cyc} + \lambda_{id} \cdot \mathcal{L}_{id} \tag{22}$$

By jointly optimizing this objective, the Dual-Mask LeafGAN achieves more precise, stable, and visually accurate domain translations, especially when guided by dual-mask segmentation under adversarial training.

# CHAPTER 4

# Experiment

## 4.1 Datasets

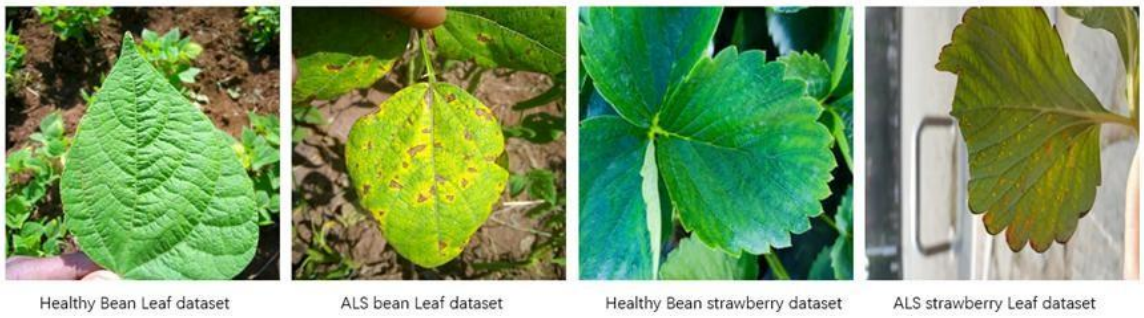



Figure 4.1: The sample of dataset images illustrating the types of leaf and diseases.

This study utilizes images of bean and strawberry leaves under two conditions: healthy and infected with Angular Leaf Spot. Except for the healthy strawberry dataset, which was collected from Roboflow, all other images were obtained from Kaggle. The bean dataset was originally collected in bean fields by the Makerere AI Lab in collaboration with the National Crops Resources Research Institute (NaCRRI), Uganda's official body responsible for agricultural research. These images reflect real-world variability in field conditions, such as lighting, leaf positioning, and background complexity. The healthy strawberry images from Roboflow provide clear and high-quality examples of non-infected leaves, useful for establishing a baseline during training. The total number of images used for each category is summarized in Table 1.

Table 1 The details of datasets

| | Healthy | Angular Leaf Spot (ALS) |
|---|---|---|
| Bean | 319 | 297 |
| Strawberry | 300 | 257 |

## 4. 2 Training Details

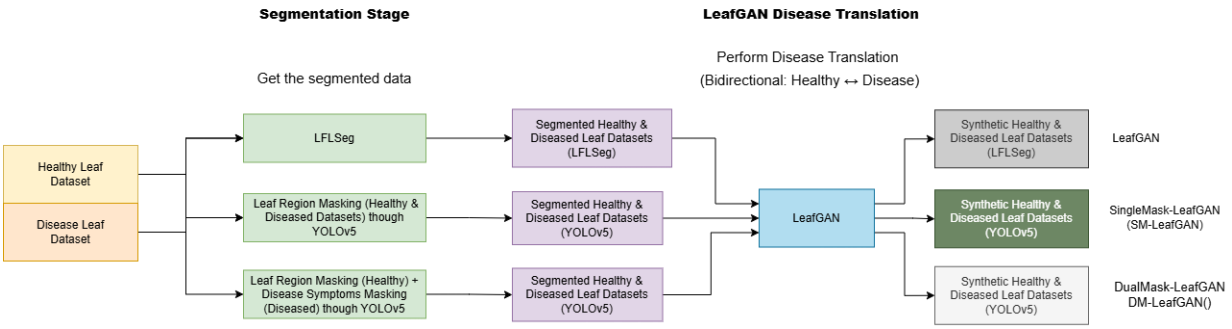



Figure 4.2: Overall research workflow.

The research begins with collecting healthy and diseased leaf images, which are then subjected to segmentation. Unlike the original LFLSeg method that masks only the leaf region, our approach uses YOLOv5 to perform dual masking—separating both the leaf region and disease symptoms for diseased leaves. This results in more precise segmented datasets for both healthy and diseased domains. These segmented datasets are then used to train the Dual-Mask Leaf-GAN, which performs bidirectional image translation. Finally, the trained model generates high-quality synthetic healthy and diseased leaf images that better simulate real-world variations.

### 4.2.1 Segmentation Strategies

To support the preprocessing phase of LeafGAN and enhance the quality of style translation, different segmentation strategies were explored to guide the model's attention toward relevant regions of interest. Three main approaches were evaluated: LFLSeg, leaf region masking (SM-LeafGAN), and combined leaf and disease symptom masking (DM-LeafGAN). Each strategy differs in how it defines and isolates informative areas for the generator to focus on during training.

(1) LFLSeg

LFLSeg represents the original segmentation strategy used in LeafGAN. It utilizes a pretrained module that generates heatmaps indicating areas of high relevance. However, this method lacks spatial precision and does not produce actual segmentation masks. As an attention-based approach, it offers only general guidance to the model without explicitly isolating leaf regions or disease symptoms, which can limit translation accuracy.

(2) Leaf Region Masking

This masking strategy was adopted in the SingleMask-LeafGAN (SM-LeafGAN) model to guide the generator with spatially constrained inputs. The strategy involves explicit segmentation of the entire leaf area to remove background distractions. The leaf regions were first manually annotated using Roboflow, and these annotated masks were used to train a YOLOv5x-seg model. The model was trained with an input size of 320, batch size of 32, and 150 epochs on Google Colab. Once trained, it could automatically generate binary masks that isolate leaf regions in both healthy and diseased images. These masks were applied during training to ensure that the generator focused only on the leaf areas, thereby improving the consistency and accuracy of the style translation.

(3) Leaf Region Masking + Disease Symptom Masking

DM-LeafGAN extends the previous approach by adding segmentation for visible disease symptoms. Both the leaf regions and the disease-affected areas (Angular Leaf Spot) were manually annotated using Roboflow. The annotated disease symptom data, in the form of bounding boxes, was used to train a standard YOLOv5 object detection model. The output bounding boxes were then converted into binary masks that specifically highlight symptomatic regions. Combined with the leaf masks from SM-LeafGAN, this dual-masking strategy provides detailed spatial guidance, enabling the model to distinguish between healthy and diseased regions. This enhances the disease-to-healthy image translation by helping the generator target specific areas for correction.

### 4.2.2 Disease Translation Models

To investigate the impact of segmentation-based masking on translation performance, three configurations of the LeafGAN model were evaluated. Each model was trained on the same dataset under consistent conditions to ensure fair comparison, with the objective of bidirectional image translation between healthy and diseased leaf domains.

- **LeafGAN**: This baseline configuration adopts the original LeafGAN architecture, using LFLSeg for preprocessing. No explicit region masking is applied images are translated in full, including background content. The model was trained for 150 epochs with a learning rate of 0.0005 and an identity loss weight ($\lambda_{id} = 0.6$).
- **SM-LeafGAN**: In this configuration, binary masks generated by the SM-LeafGAN model are applied to both healthy and diseased images. By masking out the background,

the model learns to focus exclusively on the leaf regions. Training was conducted for 100 epochs with a learning rate of 0.0007.

- **DM-LeafGAN**: This model introduces a dual-masking strategy. Leaf region masks are applied to healthy images, while symptom-specific masks—generated by the disease detection model—are used for diseased images. This selective masking directs the generator to concentrate on symptomatic regions during the disease-to-healthy transformation, potentially leading to more accurate correction and artifact reduction. Training settings matched those of SM-LeafGAN(100 epochs, learning rate of 0.0007).

Overall, these three configurations facilitate an in-depth analysis of how different segmentation strategies affect translation quality, particularly in reducing visual artifacts and improving symptom localization accuracy in synthetic images.

### 4.3 Result

#### 4.3.1 Leaf and disease segmentation result

To evaluate the performance of the proposed segmentation method, we compared our custom pipeline with the original LFLSeg module used in LeafGAN. Two segmentation tasks were considered: leaf region masking for healthy leaves and disease lesion masking for infected leaves. The final mAP0.5 scores achieved were 82.30% and 73.28%, respectively.

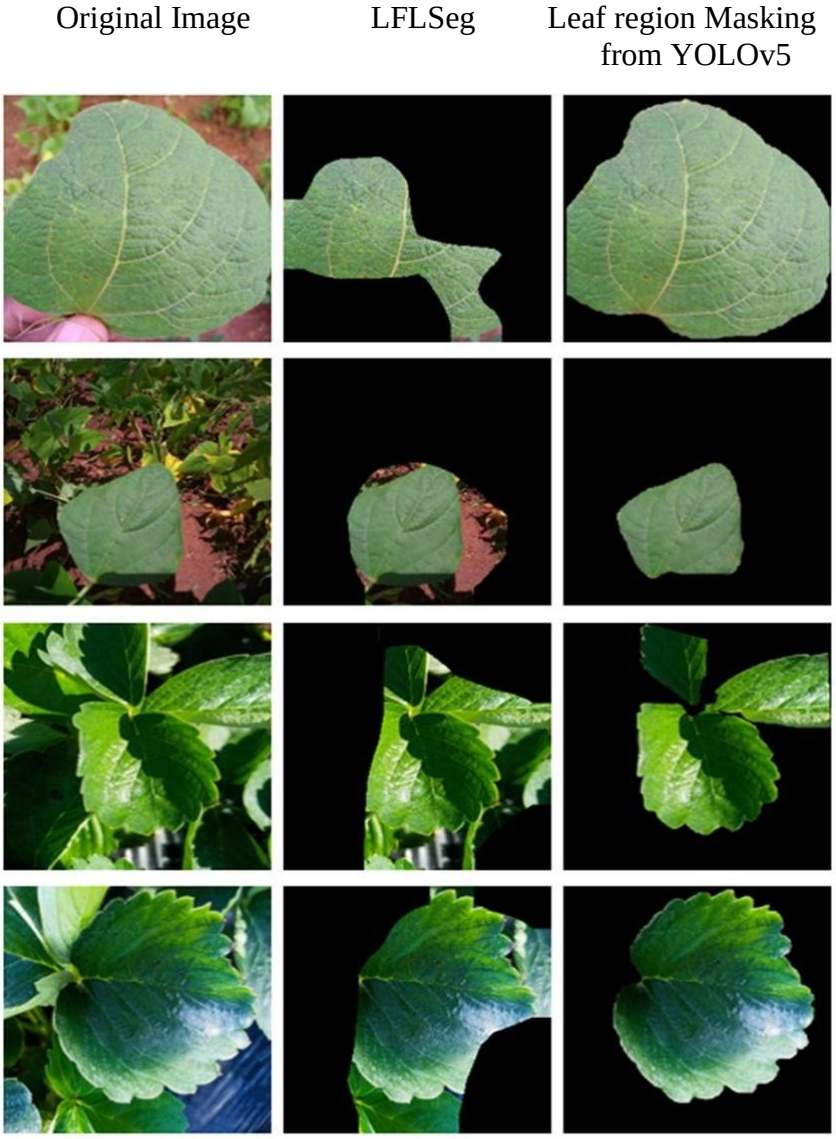



Figure 4.3: Comparison of Healthy Leaf Segmentation using different segmentation models.

Our approach leverages YOLOv5 to detect leaf regions, followed by binary masking to isolate the leaf from the background. It addresses a key limitation of LFLSeg, which employs Grad-CAM to highlight leaf regions, specifically its inability to accurately segment images containing multiple leaves. In contrast, our method reliably detects and isolates each leaf

instance. Compared to LFLSeg, our method also demonstrated improved spatial accuracy, particularly in cases with complex leaf contours and challenging lighting conditions, as illustrated in Figure 4.3.

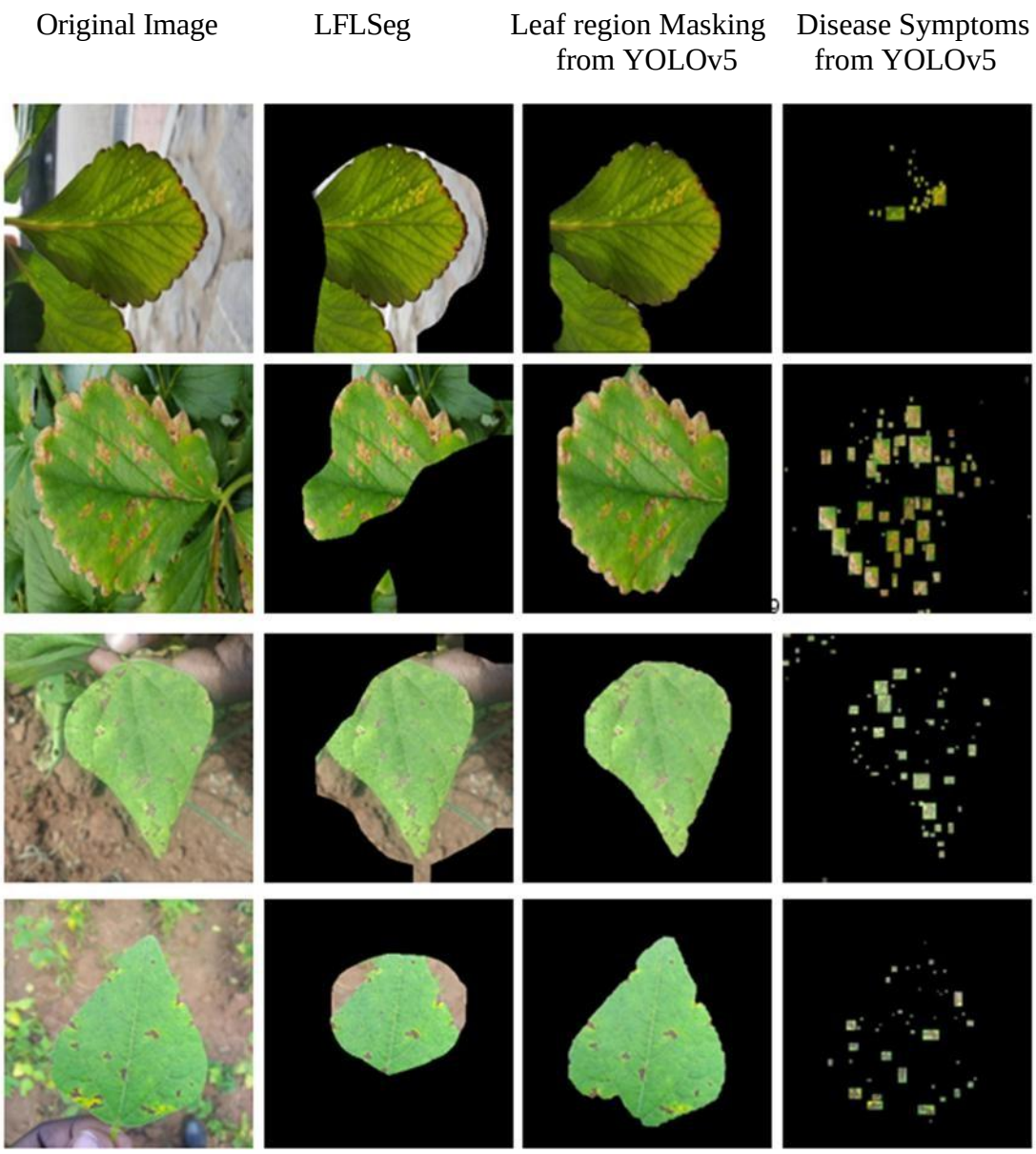



Figure 4.4: Comparison of Disease Leaf Segmentation using different segmentation models.

For disease lesion extraction, we proposed a novel strategy that incorporates lesion masks to guide LeafGAN more effectively. This enables more accurate translation of healthy patterns and realistic removal of disease features, as shown in Figure 4.4.

### 4.3.2 Disease Translation Results


Figure 4.5: Healthy bean leaves translated to the disease bean leaves result using different types of models.

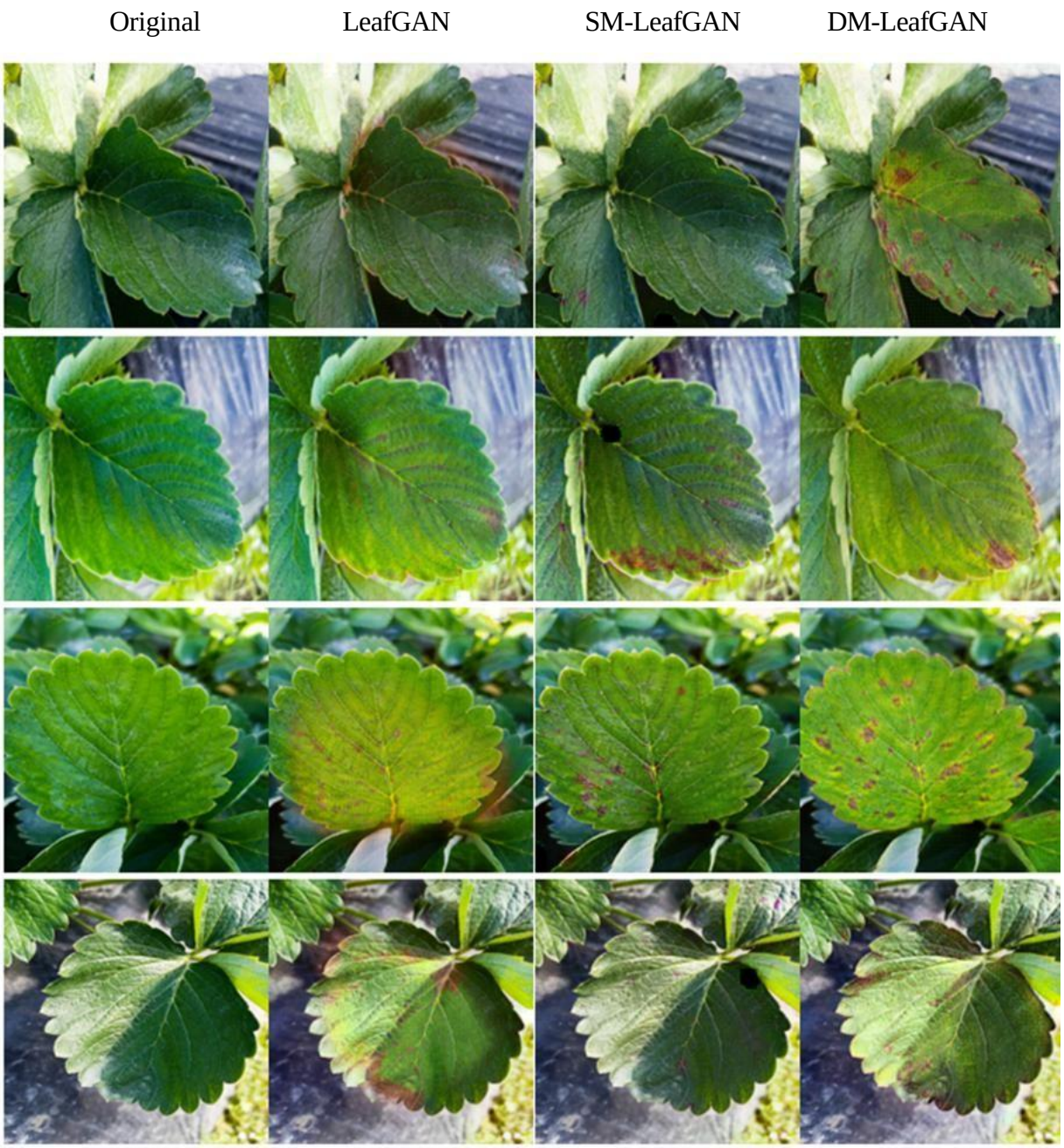



Figure 4.6: Healthy strawberry leaves translated to the disease strawberry leaves result using different types of models.

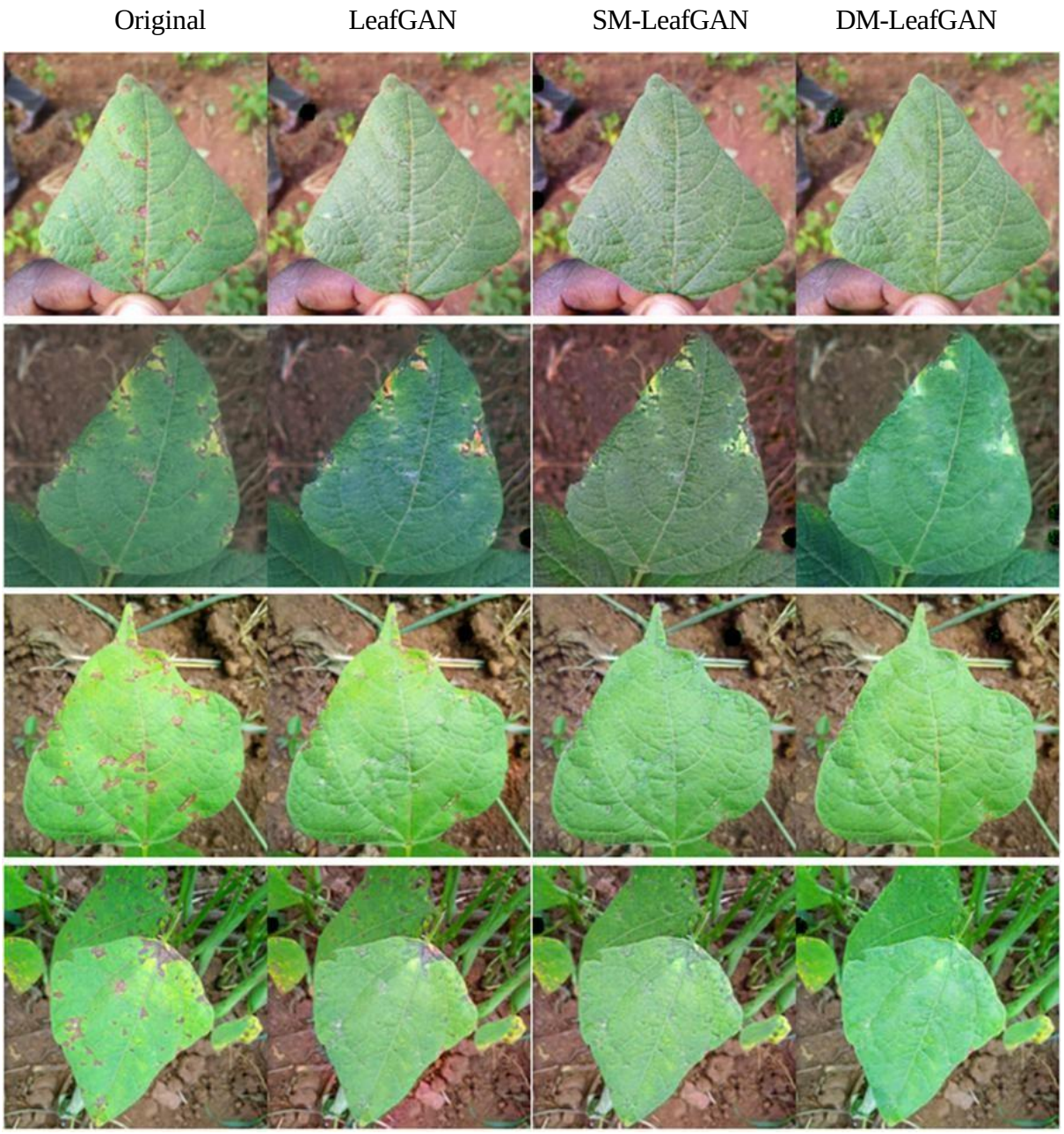



Figure 4.7: Diseased bean leaves translated to the healthy bean leaves result using different types of models.

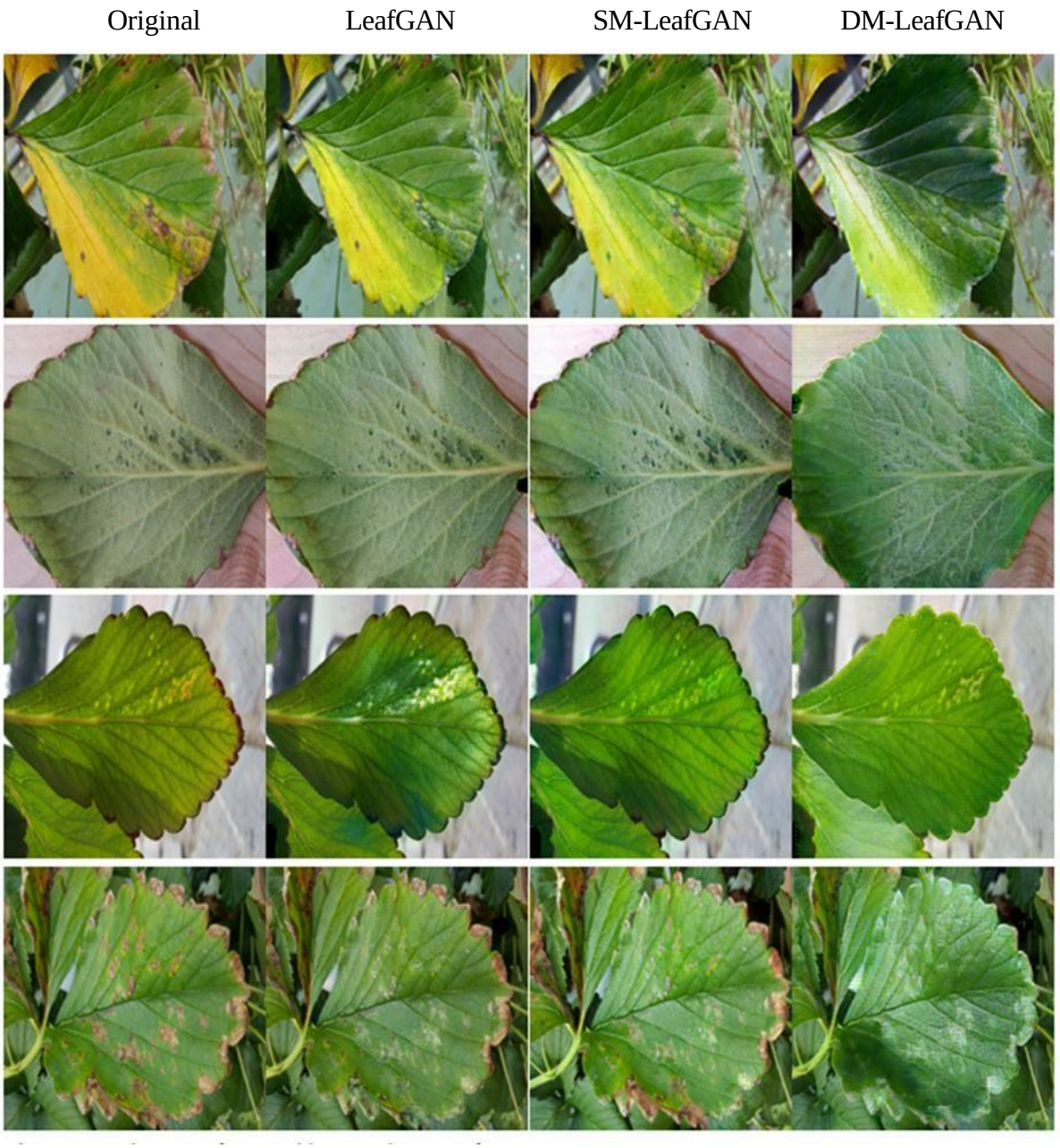



Figure 4.8: Diseased strawberry leaves translated to the healthy strawberry leaves result using different types of models.

This section evaluates the performance of three disease translation models: LeafGAN with LFLSeg, SingleMask-LeafGAN, and DualMask-LeafGAN, in simulating disease progression and recovery on bean and strawberry leaves. Each model was assessed based on its ability to transform leaf images from healthy to diseased states and vice versa.

LeafGAN with LFLSeg initiates the transformation by introducing early disease symptoms in healthy leaves, such as mild discoloration and sparse spotting. In diseased-to-healthy translations, it slightly reduces visible disease signs. However, this model often retains residual

features from the original state, indicating limited transformation depth.

SingleMask-LeafGAN improves upon the original LeafGAN by generating more distinct disease symptoms characterized by increased spotting and discoloration, including pinkish-purple lesions in healthy-to-diseased translations. It also achieves better symptom reduction in diseased leaves, producing more balanced intermediate results. Nonetheless, some artifacts from the original state remain.

DualMask-LeafGAN demonstrates the most effective and realistic transformations among the three. In the healthy-to-diseased translation, it simulates advanced disease symptoms with extensive discoloration and detailed spotting, closely mimicking naturally infected leaves. In the reverse direction, DualMask-LeafGAN successfully restores leaves to near-healthy appearances, effectively minimizing disease traces.

In summary, DualMask-LeafGAN delivers the highest fidelity in both transformation directions, followed by SingleMask-LeafGAN, which shows moderate improvements over the original LeafGAN with LFLSeg. While the baseline model provides basic functionality, its results lack completeness and realism. These findings highlight DualMask-LeafGAN as the most robust model for simulating plant disease dynamics, with promising applications in agricultural research and plant health monitoring. Further enhancements to the first two models could improve their translation accuracy and utility.

### 4.4 Failure Cases

Although the proposed models particularly DualMask-LeafGAN demonstrated strong performance in simulating disease progression and recovery, several failure cases reveal notable limitations that warrant further investigation.

Figure 4.9: Failure cases generated by SingleMask-LeafGAN.

As depicted in Figure 4.9, the first row of images reveals a poorly trained model for translating healthy to diseased leaves, marked by the appearance of gray artifacts. This deficiency is likely attributable to suboptimal hyperparameters, which hinder the model's ability to learn effectively, resulting in underwhelming performance as also observed in Figures 3 and 4 of the first row. Conversely, the second row showcases outcomes from the disease-to-healthy translation task, where the model achieves partial success in reducing disease symptoms but struggles to fully eliminate leaf spot areas, leaving residual green spots. Moreover, images 3 and 4 in the second row exhibit subtle visual distortions in the generated outputs, underscoring the model's persistent challenges in producing accurate translations.

Figure 4.10: Failure cases generated by DualMask-LeafGAN.

In the first row, which illustrates the translation from healthy to diseased, the sequence starts with a vibrant green leaf but quickly degrades as pinkish spots appear, followed by significant gray cutouts or missing sections in later images. These gaps indicate the model's difficulty in preserving the leaf's structural integrity, likely due to overfitting or poor boundary handling. As a result, the outcomes feature disrupted vein designs and grey artifacts in the leaves. In the second row, depicting disease-to-healthy translation, the initial leaf shows scattered disease spots. Although some symptom reduction is achieved, the restoration is hindered by an overly saturated green tone and irregular textures in subsequent images. This excessive coloration, along with subtle visual distortions, reveals the model's ongoing struggle to accurately normalize color and texture, ultimately failing to produce convincingly healthy leaves.

These failure cases highlight the importance of enhancing both the model's structural awareness and its ability to generate biologically coherent transformations. Future improvements could include integrating advanced loss functions, such as perceptual loss, to better preserve texture and structural consistency. Additionally, further training with fine-tuning and hyperparameter optimization may help address residual artifacts and improve generalization to complex leaf patterns.

# CHAPTER 5

# Conclusion

This study evaluated three LeafGAN models for plant disease translation, focusing on how different masking strategies influence the visual realism of generated images. Notably, our proposed DualMask-LeafGAN, which integrates both leaf region and disease symptom masking using YOLOv5, significantly outperformed the original LeafGAN with LFLSeg. The superior performance of this YOLO-based approach highlights that segmentation quality is a crucial factor in enhancing the realism of GAN-generated synthetic data.

SingleMask-LeafGAN, which utilized only leaf region masking, generated cleaner images by eliminating background interference and focusing on leaf areas. While it offered improved image quality compared to LeafGAN with LFLSeg, its outputs were less realistic than those produced by DualMask-LeafGAN. Although SingleMask-LeafGAN achieved moderate improvements, it occasionally retained artifacts from the original images, limiting its overall visual effectiveness.

DualMask-LeafGAN consistently delivered the most visually convincing results. It accurately simulated advanced disease symptoms and restored diseased leaves to near-healthy appearances with minimal residual artifacts. Its outputs closely resembled real-world leaf conditions, making it the most robust model for capturing the visual dynamics of plant disease.

The findings underscore the strong potential of DualMask-LeafGAN in simulating plant disease dynamics, particularly in terms of visual fidelity. Although some failure cases were observed, such as grey artifacts and occasional visual distortions, these issues were minor relative to its overall performance. Addressing these artifacts, for instance through the integration of perceptual loss, could further enhance output quality.

In conclusion, DualMask-LeafGAN emerged as the most effective model for realistic plant disease translation, offering high-fidelity image generation in both directions. Its potential applications include agricultural research, synthetic dataset generation, and plant health monitoring. Future work may focus on refining LeafGAN with LFLSeg and SingleMask-LeafGAN to narrow the performance gap and further enhancing DualMask-LeafGAN’s visual consistency.

## REFERENCES

[1] S. P. Mohanty, "PlantVillage-Dataset," GitHub, Mar. 14, 2023. https://github.com/spMohanty/PlantVillage-Dataset

[2] Committee on Science Breakthroughs 2030: A Strategy for Food and Agricultural Research et al., Science Breakthroughs to Advance Food and Agricultural Research by 2030. Washington, D.C.: National Academies Press, 2019. Available: https://nap.nationalacademies.org/catalog/25059/science-breakthroughs-to-advance-food-and-agricultural-research-by-2030

[3] Q. H. Cap, H. Uga, S. Kagiwada, and H. Iyatomi, "LeafGAN: An Effective Data Augmentation Method for Practical Plant Disease Diagnosis," IEEE Transactions on Automation Science and Engineering, pp. 1–10, 2020, doi: https://doi.org/10.1109/tase.2020.3041499.

[4] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," arXiv.org, 2017. https://arxiv.org/abs/1703.10593

[5] I. J. Goodfellow et al., "Generative Adversarial Networks," arXiv.org, Jun. 10, 2014. https://arxiv.org/abs/1406.2661

[6] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," Neurocomputing, vol. 321, pp. 321–331, Dec. 2018, doi: https://doi.org/10.1016/j.neucom.2018.09.013.

[7] T. Iqbal and H. Ali, "Generative Adversarial Network for Medical Images (MI-GAN)," Journal of Medical Systems, vol. 42, no. 11, Oct. 2018, doi: https://doi.org/10.1007/s10916-018-1072-9.

[8] F. Schubert, M. Awiszus, and B. Rosenhahn, "TOAD-GAN: a Flexible Framework for Few-Shot Level Generation in Token-Based Games," IEEE Transactions on Games, pp. 1–1, 2021, doi: https://doi.org/10.1109/tg.2021.3069833.

[9] F. Schubert, M. Awiszus, and B. Rosenhahn, "TOAD-GAN: a Flexible Framework for Few-Shot Level Generation in Token-Based Games," IEEE Transactions on Games, pp. 1–1, 2021, doi: https://doi.org/10.1109/tg.2021.3069833.

[10] Z. Zheng, J. Liu, and N. Zheng, "P$^{2}$-GAN: Efficient Stroke Style Transfer using Single Style Image," IEEE Transactions on Multimedia, pp. 1–13, 2022, doi: https://doi.org/10.1109/tmm.2022.3203220.

[11] A. H. Bermano et al., "State‐of‐the‐Art in the Architecture, Methods and Applications of StyleGAN," Computer Graphics Forum, vol. 41, no. 2, pp. 591–611, May 2022, doi: https://doi.org/10.1111/cgf.14503.

[12] E. Richardson et al., "Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, doi: https://doi.org/10.1109/CVPR46437.2021.00232.

[13] P. Dhariwal and A. Nichol, “Diffusion Models Beat GANs on Image Synthesis,” arXiv:2105.05233 [cs, stat], Jun. 2021, Available: https://arxiv.org/abs/2105.05233.

[14] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” arXiv:2006.11239 [cs, stat], Dec. 2020, Available: https://arxiv.org/abs/2006.11239

[15] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, “RePaint: Inpainting using Denoising Diffusion Probabilistic Models,” arXiv:2201.09865 [cs], Feb. 2022, Available: https://arxiv.org/abs/2201.09865 [18] E. Luhman and T. Luhman, “Denoising Synthesis: A module for fast image synthesis using denoising-based models,” Software Impacts, p. 100076, May 2021, doi: https://doi.org/10.1016/j.simpa.2021.100076.

[16] E. Luhman and T. Luhman, “Knowledge Distillation in Iterative Generative Models for Improved Sampling Speed,” arXiv:2101.02388 [cs], Jan. 2021, Available: https://arxiv.org/abs/2101.02388

[17] S. Du, K. Hao, H. Zhang, X. Tang, and B. Wei, “Patch Elastic Deformation: An Effective Data Augmentation Method,” Nov. 2022, doi: https://doi.org/10.1109/cac57257.2022.10055297.

[18] B. Trabucco, K. Doherty, M. Gurinas, and R. Salakhutdinov, “Effective Data Augmentation With Diffusion Models,” arXiv:2302.07944 [cs], Feb. 2023, Available: https://arxiv.org/abs/2302.07944

[19] R. Gandikota, J. Materzynska, J. Fiotto-Kaufman, and D. Bau, “Erasing Concepts from Diffusion Models,” arXiv.org, Jun. 20, 2023. https://arxiv.org/abs/2303.07345

[20] Z. Wang et al., “Patch Diffusion: Faster and More Data-Efficient Training of Diffusion Models,” arXiv.org, Oct. 18, 2023. https://arxiv.org/abs/2304.12526.

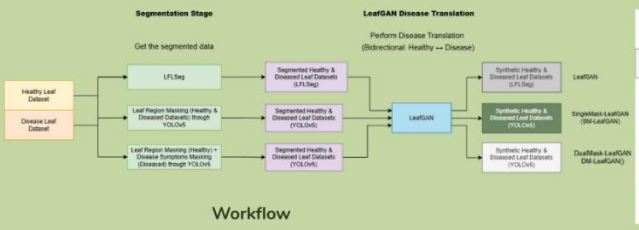
[21] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, Ravi Ramamoorthi, and R. Ng, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” arXiv (Cornell University), Mar. 2020, doi: https://doi.org/10.48550/arxiv.2003.08934.

**POSTER**

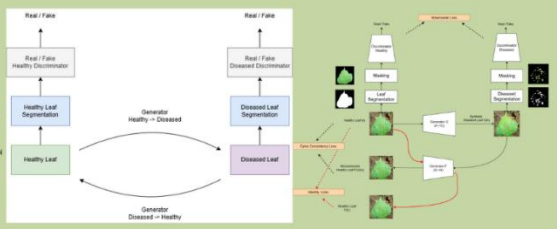
# Augmentation strategies for Plant Disease Classification

Student: Chia Wan Jun

Supervisor: Ts Dr Tan Hung Khoon




Workflow




LeafGAN

## Introduction

Manual plant disease detection is slow, error-prone, and not scalable. AI-powered image classification using deep learning (e.g., CNNs, GANs) has improved detection accuracy. However, performance is limited by imbalanced and low-quality datasets.

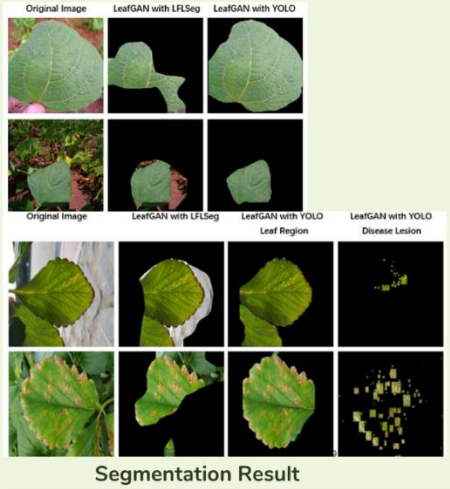
## Contribution

1. demonstrates that segmentation quality is a crucial factor in enhancing the realism of GAN-generated synthetic data for plant disease augmentation.
2. A novel dual masking strategy with leaf region masking and disease symptoms masking is introduced to optimize diseaseaugmentation.

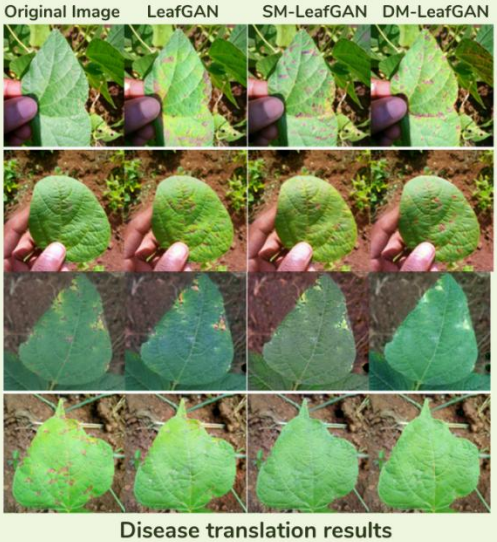
## Objective

- enhance the diversity of existing plant disease datasets by applying advanced data augmentation techniques based on disease translation methods.
- study the impact of different leaf segmentation strategies on the quality of augmented disease images
- propose a novel plant disease augmentation method incorporating alternative segmentation strategies aimed at improving the realism and reliability of synthetic images.

## Results




Segmentation Result




Disease translation results

## Conclusion

This study compared three LeafGAN models with different masking strategies for plant disease image augmentation. SingleMask-LeafGAN, using only leaf region masking, achieved the best classification accuracy but lacked visual realism. Model 3, combining leaf and disease symptom masking, produced the most realistic images but suffered from gray artifacts that slightly reduced accuracy. Future work will focus on integrating perceptual loss into DualMask-LeafGAN to enhance both visual quality and classification performance.