

Article

Data Augmentation Method for Plant Leaf Disease Recognition

Byeongjun Min ^{1,2,†}, Taehyun Kim ^{3,†} , Dongil Shin ¹  and Dongkyoo Shin ^{1,2,*} ¹ Department of Computer Engineering, Sejong University, Seoul 05006, Republic of Korea² Department of Convergence Engineering for Intelligent Drones, Sejong University, Seoul 05006, Republic of Korea³ Department of Agriculture Engineering, National Institute of Agricultural Sciences, Jeonju 63240, Republic of Korea

* Correspondence: shindk@sejong.ac.kr

† These authors contributed equally to this work.

Abstract: Recently, several plant pathogens have become more active due to temperature increases arising from climate change, which has caused damage to various crops. If climate change continues, it will likely be very difficult to maintain current crop production, and the problem of a shortage of expert manpower is also deepening. Fortunately, research on various early diagnosis systems based on deep learning is actively underway to solve these problems, but the problem of lack of diversity in some hard-to-collect disease samples remains. This imbalanced data increases the bias of machine learning models, causing overfitting problems. In this paper, we propose a data augmentation method based on an image-to-image translation model to solve the bias problem by supplementing these insufficient diseased leaf images. The proposed augmentation method performs translation between healthy and diseased leaf images and utilizes attention mechanisms to create images that reflect more evident disease textures. Through these improvements, we generated a more plausible diseased leaf image compared to existing methods and conducted an experiment to verify whether this data augmentation method could further improve the performance of a classification model for early diagnosis of plants. In the experiment, the PlantVillage dataset was used, and the extended dataset was built using the generated images and original images, and the performance of the classification models was evaluated through the test set.

Keywords: plant disease recognition; data augmentation; imbalanced dataset; convolutional attention



Citation: Min, B.; Kim, T.; Shin, D.; Shin, D. Data Augmentation Method for Plant Leaf Disease Recognition. *Appl. Sci.* **2023**, *13*, 1465. <https://doi.org/10.3390/app13031465>

Academic Editor: Zhengjun Qiu

Received: 1 December 2022

Revised: 17 January 2023

Accepted: 20 January 2023

Published: 22 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recently, rapid climate change has had a great impact on the natural ecosystem and human life, and specifically on crop productivity [1]. In particular, an increase in temperature generally shortens the life cycles of plant pathogens while increasing their density, which also increases the possibility of generating mutants with strong pathogenicity [2]. In addition, it causes problems such as crop diseases and pests, which were rare in the past, or the expansion of the areas vulnerable to plant diseases that occur only in a limited region. Currently, the decline in crop yields due to plant diseases is estimated at around 16% worldwide [3], but if climate change continues, it may become very difficult to maintain current crop productivity. Therefore, the problem of crop productivity reduction due to climate change may intensify in the future, and there is also a growing lack of professional manpower to help solve it.

Deep learning has made a breakthrough in the field of computer vision in recent years, and there has been much research on how to use it for the early diagnosis of plant diseases [4–9]. Deep learning is suitable for diagnosing visual disease symptoms in the leaves and stems of plants because it is possible to extract and learn high-level features from images. Using this deep learning method, Mohanty et al. [10] conducted a plant image disease study based on a deep convolutional neural network model. The authors

conducted an experiment to classify 26 disease images from 14 crop species using Google LeNet and AlexNet models and reported an average accuracy of 99.3%. Ferentinos et al. [11] also pointed out that models taken on datasets generated in controlled experimental environments achieved 99.5% accuracy when tested with models AlexNet, AlexNetOWTBn, GoogLeNet, OverFeat, and VGG, but did not perform the same in real-world environments. The accuracy dropped to 33% when using the learned model in the real world. Hassan et al. [12] conducted a study on image-based plant disease recognition using the Shallow VGG network. The Shallow VGG network is constructed using only a few layers from a pre-trained VGG network and serves to perform feature extraction from input data. The authors proposed to classify the features extracted through this through the Random Forest and Xgboost models. As a result of the experiment, it was reported that the two combined models proposed have better classification performance than the Vanilla VGG Network model. Lee et al. [13] performed image-based plant leaf disease research based on Recurrent Neural Network. The authors point out that CNN-based classification models do not clearly capture disease areas, and to solve this problem, they proposed a network combining Gated Recurrent Units (GRU) and Attention modules. They verified through experiments that the proposed method can capture clear disease regions unlike CNN and reported that classification performance was improved.

Although many vision-based plant disease recognition advanced studies using deep learning are in progress, some diseases are not common in nature, making it difficult to collect the same amount of data from healthy plant samples. Therefore, datasets for plant disease diagnosis collected from natural environments often lack samples of these rare disease classes, a problem known as imbalanced data in machine learning. When trained from such imbalanced data, most supervised learning models suffer from the overfitting problem, in which decision boundaries are biased toward major classes [14,15]. To solve this problem, previous work has adopted a manual augmentation method using professional manpower, but this is very inefficient and has high costs.

In this paper, we propose a data augmentation method based on image-to-image translation that can increase the sample diversity of diseased leaf datasets with insufficient numbers. The proposed augmentation method performs translation between healthy and diseased leaf images through the cycle-consistent generative adversarial networks (Cycle GAN) [16]. However, vanilla CycleGAN shows poor plant leaf image translation results, and we point out and improve on the following two problems: (1) poor reflection of the evident texture of the target disease, and (2) poor preservation of the shape of the input leaf image due to indiscriminate transformation in the background region. To solve this problem, we further utilize an Attention mechanism and a binary mask that explicitly indicates the location of the leaves. The attention mechanism is a method of selectively focusing on key information within given information and can greatly improve the expressive ability of the proposed plant leaf translation model. In deep learning, attention mechanisms originated in human perceptual systems as a method of selecting key information to focus on [17,18]. Many studies combining attention mechanisms have been conducted, and it has been demonstrated that the model can significantly improve the feature expression ability [17–21]. We used the Convolutional Block Attention Module (CBAM) [22] to construct a network that allows the generator to decide which features should be preferentially selected to deceive the discriminator. CBAM is a lightweight attention module that can learn attention maps corresponding to spaces and channels from a given feature map. We use CBAM to configure the generator network as a Pre-Activation Residual Attention Network and perform adaptive feature refinement during the training process. Also, we train a background loss function that regulates the generator to not perform unnecessary transformations on background regions. The background loss suppresses unnecessary transformations between the input data and the transformed image through a binary mask explicitly representing the background region of the input image. This assists the attention module in quickly focusing on the leaf itself, excluding the background where it cannot perform a transformation. Through these improvements, we generated a more plausi-

ble diseased leaf image compared to existing methods and conducted an experiment to verify whether this data augmentation method could further improve the performance of a classification model for early diagnosis of plants. We conducted experiments using apple, potato, and grape leaf data provided by the Plant Village [23] dataset and verified whether the proposed method can solve the problems of images produced by vanilla CycleGAN. We then compared the experimental results of various classification models to verify whether the extended dataset from the images generated by the proposed method can improve performance. Experimental results prove that the proposed augmentation method can solve the problem of lack of diversity in some disease samples that are difficult to collect by generating usefully translated images and contributing to improving classification performance.

The main contributions of this study are as follows:

- In this study, we focus on the overfitting problem caused by imbalanced data problems created by some disease images that are rarely observed in natural environments and propose a novel plant image augmentation method based on CycleGAN to solve this problem.
- We point out two issues that vanilla CycleGAN needs to address for image augmentation: (1) poor reflection of the evident texture of the target disease, and (2) poor preservation of the shape of the input leaf image due to indiscriminate transformation in the background region.
- To improve the poor image conversion results of CycleGAN, we utilized the attention mechanism and binary mask, and solved the bias problem of classification models by constructing an extended dataset.
- We have conducted extensive experiments to verify that the proposed method can improve the overfitting problem due to imbalanced dataset.

2. Related Studies

Recently, various data augmentation studies on the early diagnosis of plants using Generative Adversarial Networks (GAN) are being conducted. Most of them aim to correct problems with existing models to verify that more realistic images can be generated or to solve overfitting problems by replacing insufficient data.

Wu et al. [24] worked on recovering tomato disease data via DCGAN (Deep Convolutional GAN) [25]. Their contribution demonstrates that samples generated via DCGAN can improve the efficiency of classification models. However, it was confirmed that the generated samples are not clear and are not likely images in reality. Also, there is a problem of not being able to control the class of the image to be created.

Deng et al. [26] proposed an RHAC_GAN model that improves ACGAN (Auxiliary Classifier GAN) to solve the problem of tomato disease data augmentation. RHAC_GAN aims to solve the problem of lack of diversity among generated images by using hidden variables obtained from the discriminator along with inputs from the generator. Regarding the lack of diversity in the generated images, the authors explain that traditional ACGAN generators have difficulty learning different information within a class. In addition, a network was constructed by combining the Attention module and the Residual Block so that the generated images reflected obvious disease features. Plant leaf images generated by RHAC_GAN are sharper and more realistic compared to DCGAN and ACGAN.

Cap et al. [27] proposed a LeafGAN model that solved CycleGAN's problem of changing unwanted content such as background when transforming plant leaf images. Their contribution is to propose an LFLSeg module that can segment leaf regions relatively simply and use it to control the transformation of background regions. The author proposed a background loss function to control the transformation of the background area by the generator and used only the input from which the background was removed as the input of the discriminator so that the discriminator for the background area did not react. The experiment was conducted based on actual data collected in the natural environment, and

it was reported that classification performance was partially improved when learning was performed with a dataset augmented by the proposed model.

Xu et al. [28] proposed a Style-Consistent Image Translation (SCIT) model that aims at converting to the target domain while preserving information about the parts related to styles independent of labels in the translation process of plant images. To this end, we proposed a style loss calculated through a pre-trained vgg19 network. The authors verified through experiments that the amount of change to the target domain can be adjusted according to the change in the coefficient value of style loss.

3. Materials and Methods

3.1. Data Augmentation Method for Plant Leaf Disease Recognition

The plant data augmentation method proposed in this study was studied by borrowing the structure of CycleGAN, as shown in Figure 1, and the main suggestion is to improve the expression ability of the model using the Pre-Activation Residual Attention Block (PaRAB) and Binary Mask. These improvements allow the proposed model to generate visually plausible plant leaf data. The proposed method solves the bias problem of early diagnosis models of plants by contributing to increasing the diversity of samples by making diseased samples with insufficient numbers from normal plant leaves.

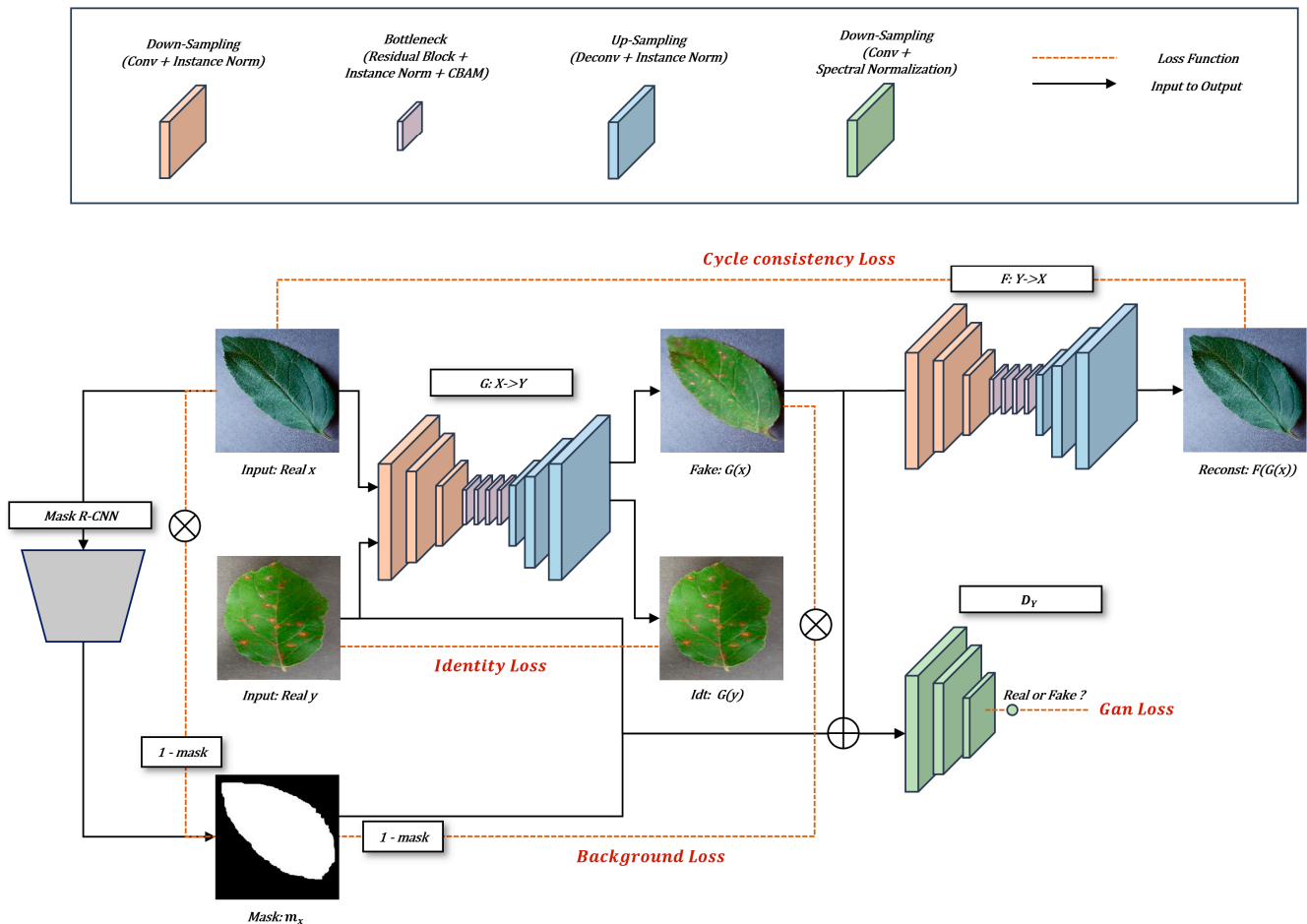


Figure 1. Data Augmentation Method for Plant Leaf Disease Recognition.

The entire training process of the proposed method and each loss function can be confirmed in Figure 1. The red line represents the elements needed to calculate each loss function, and the black line represents the elements used as inputs for each model. Basically, two generators $G: X \rightarrow Y$ and $F: Y \rightarrow X$ are utilized, where X and Y are the domains to be transformed, X means a healthy leaf class, and Y means a disease-affected leaf class.

to be transformed. Consequently, the two generators are converted to domains that are opposite to each other. In addition, the discriminators competing with the generators D_Y and D_X are required for training, where D_Y distinguishes the fake leaf image $G(x)$ from the real leaf image $y \in Y$, and D_X distinguishes the fake leaf image $F(y)$ from the real leaf image $x \in X$. G , D_Y , and F , D_X , which correspond to each other as a pair, proceed with adversarial learning and perform transformation between both domains.

In addition, we utilize the binary mask M_x and M_y required to regulate the generator from converting to unnecessary areas. The binary mask explicitly denotes leaf areas, M_x points to healthy leaf areas, and M_y points to disease-infected leaf areas. The leaf mask is obtained using a pre-trained Mask R-CNN [29] model, and for time efficiency, all mask data used for learning were extracted before training. the binary mask data is used by the generator to learn the background loss function and is also passed along as input to the discriminator as additional information about the leaf region. Since this method is used only as an input of a discriminator that is not actually used during the test process, it is reasonable because there is no need to have mask data when making a new sample by the generator during the test process. Additionally, giving additional information to discriminators is already well known to further enhance the model's expressiveness in previous studies such as CGAN and ACGAN. This allows the generator to create new samples that preserve the shape of the input leaves well and assists the attention module within PaRAB to quickly focus on the leaf region, excluding backgrounds where transformation cannot be performed.

3.2. Pre-Activation Residual Attention Block

In the initial training stage, the generator of CycleGAN attempts to transform the entire image region evenly and indiscriminately and learns to deceive the discriminator. This also includes background areas that are unnecessary for transformation. Moreover, CycleGAN lacks an attention mechanism, so the generator cannot discern the features that should be considered more important in the translation between the two domains. Due to these problems, the generator requires an exceptionally long learning time to accurately reflect the texture of the target domain.

To solve this problem, we adopt CBAM and propose a Pre-activation Residual Attention Block (PaRAB), which allows adaptive feature refinement for the given feature responses by explicitly modeling the interdependence including the channel and spatial. As shown in Figure 2, PaRAB is configured by deploying an attention module before combining the output of the pre-activation residual and the low-level features delivered via skip connection. We construct a generator network through PaRAB and train the learnable parameters of the attention module using an adversarial loss function. This allows the generator to selectively emphasize features that are advantageous for deceiving the discriminator while suppressing less important ones. As a result, the generator can create an image that clearly shows the texture of the target domain.

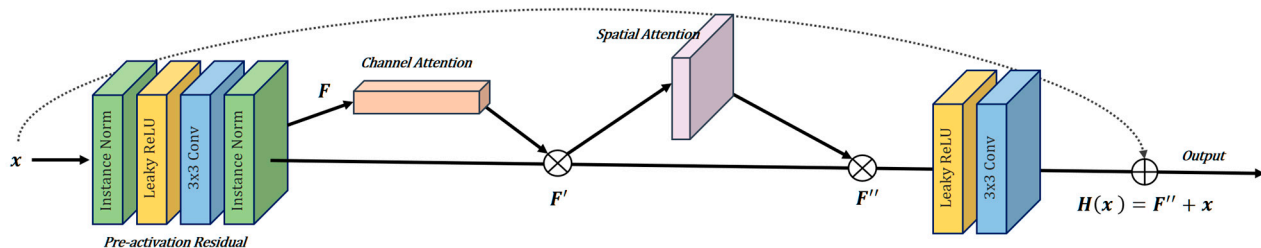


Figure 2. Pre-activation Residual Attention Block (PaRAB).

As shown in Figure 3, the two sub-modules within the PaRAB are calculated sequentially, and the attention process takes the feature map $F \in \mathbb{R}^{C \times H \times W}$ as an input to generate a channel attention map $M_c \in \mathbb{R}^{C \times 1 \times 1}$ and a spatial attention map $M_s \in \mathbb{R}^{1 \times H \times W}$, which

are used by multiplying them sequentially with each given input. The attention module is placed after the second instance normalization layer. These attention processes are as follows:

$$\begin{aligned} F' &= M_C(F) \otimes F \\ F'' &= M_S(F') \otimes F' \end{aligned} \quad (1)$$

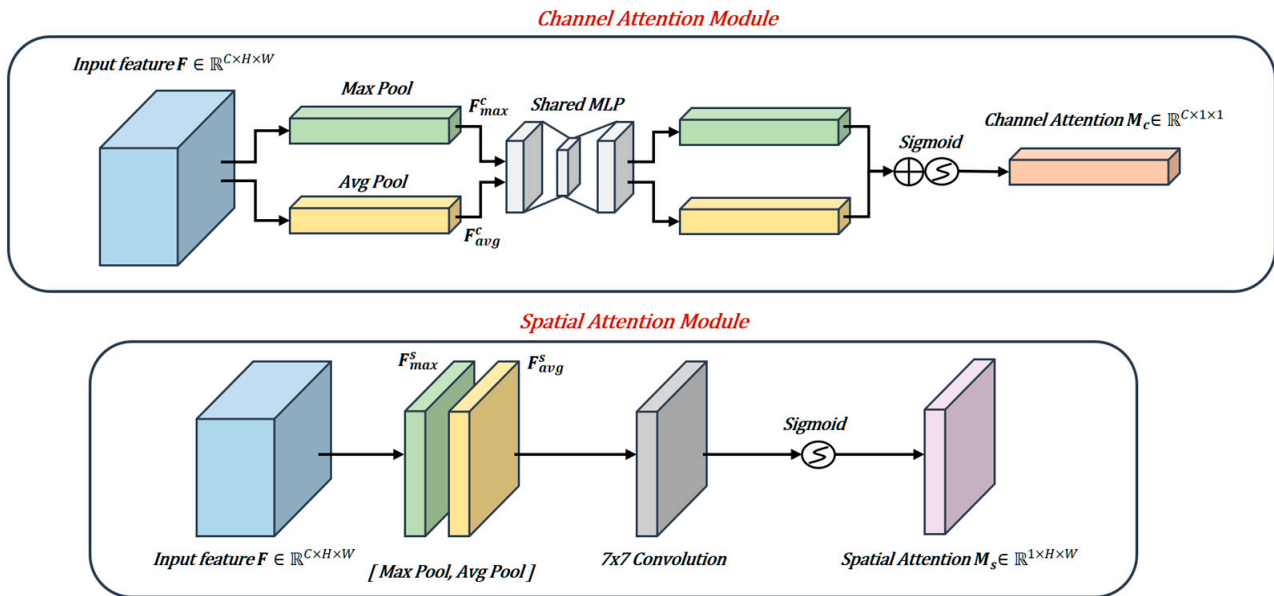


Figure 3. Channel-Spatial Attention module.

The channel attention module integrates the information of a given input feature F to generate an attention map M_c that models the interdependence between channels. These calculations are performed through average pooling and max pooling operations, and they generate channel descriptions $F_{avg}^c \in \mathbb{R}^{C \times 1 \times 1}$ and $F_{max}^c \in \mathbb{R}^{C \times 1 \times 1}$. The calculated two-channel descriptions pass through the shared network with a bottleneck compressed by a reduction ratio r , and then add up to finally obtain the channel attention map M_c through the sigmoid function. These channel attention processes are as follows:

$$M_c(F) = \sigma \left(\text{MLP}(P_{avg}(F)) + \text{MLP}(P_{max}(F)) \right) \quad (2)$$

The spatial attention module integrates the information of a given input feature F to generate an attention map M_s that models the interdependence between spatial. These calculations are performed through average pooling and max pooling operations, and they generate spatial descriptions $F_{avg}^s \in \mathbb{R}^{1 \times H \times W}$ and $F_{max}^s \in \mathbb{R}^{1 \times H \times W}$. Then, by integrating the two spatial descriptions, the model finally obtains the spatial attention map M_s using the sigmoid activation function after a 7×7 convolution. These spatial attention processes are as follows:

$$M_s(F) = \sigma \left(f^{7 \times 7}([P_{avg}(F); P_{max}(F)]) \right) \quad (3)$$

Since the two sub-modules proceed sequentially, the execution order can be changed arbitrarily. The authors report that, in general, using the channel attention module first showed better performance, and in this case, the input of the spatial attention module, used later, corresponds to $F' \in \mathbb{R}^{C \times H \times W}$.

3.3. Network Architecture

The generator network receives a real image as a single input, consisting of three down-sampling layers, an up-sampling layer, and six bottleneck layers. The down-sampling layer and the up-sampling layer used 3×3 convolutional filters, the stride value was 2 and the padding value was 1. The bottleneck layer was composed of the Pre-Activation Residual

Attention Block introduced in Section 3.2. The entire network referred to the structure of Pre-activation ResNet [30], and Instance normalization (IN) [31] and Leaky ReLU were applied to all hidden layers. Table 1 shows the shape of the input image and the detailed generator network configuration. Also, unlike using tanh activation in the generator's output layer in general, using linear activation helped to obtain clearer results.

Table 1. Related parameter and Output shape of Generator.

Layer	Activation	Normalization	Output Shape
Image	-	-	$(256 \times 256 \times 3)$
Conv 7×7	-	-	$(256 \times 256 \times 64)$
Down-sample	Leaky ReLU	Instance Normalization	$(128 \times 128 \times 128)$
Down-sample	Leaky ReLU	Instance Normalization	$(64 \times 64 \times 256)$
Down-sample	Leaky ReLU	Instance Normalization	$(32 \times 32 \times 512)$
PaRAB $\times 6$	Leaky ReLU	Instance Normalization	$(32 \times 32 \times 512)$
Up-sample	Leaky ReLU	Instance Normalization	$(64 \times 64 \times 256)$
Up-sample	Leaky ReLU	Instance Normalization	$(128 \times 128 \times 128)$
Up-sample	Leaky ReLU	Instance Normalization	$(256 \times 256 \times 64)$
Conv 1×1	Linear	-	$(256 \times 256 \times 3)$

Since the discriminator receives the results of the channel-wise concatenation operation from the input image and the corresponding mask, unlike the generator, the input channel is 6. We utilized Spectral Normalization techniques to stabilize discriminator training. This method, proposed in SNGAN [32], is a simple and efficient weight normalization method, which prevents the gradient of the discriminator from exploding by making the Lipschitz constant bounded. This is very important because generator network learning relies entirely on feedback from discriminators. Therefore, spectral normalization (SN) was applied to all layers, and the activation function was Leaky ReLU. The down-sampling layer used a 4×4 filter, the stride value was 2 and the padding value was 1. For other layers, use the stride value of 1. The overall structure of the model follows the configuration of the discriminator in PatchGAN [33]. We have confirmed that the best results are not to use feature normalization techniques such as Instance Normalization and Batch Normalization. Table 2 shows the shape of the input image used and the detailed discriminator network configuration.

Table 2. Related parameter and Output shape of Discriminator.

Layer	Activation	Normalization	Output Shape
Image \oplus Mask	-	-	$(256 \times 256 \times 6)$
Down-sample	Leaky ReLU	Spectral Normalization	$(128 \times 128 \times 64)$
Down-sample	Leaky ReLU	Spectral Normalization	$(64 \times 64 \times 128)$
Down-sample	Leaky ReLU	Spectral Normalization	$(32 \times 32 \times 256)$
Conv 4×4	Leaky ReLU	Spectral Normalization	$(31 \times 31 \times 512)$
Conv 1×1	Linear	Spectral Normalization	$(30 \times 30 \times 1)$

3.4. Loss Function

Image mapping functions G and F perform adversarial learning with discriminators D_X and D_Y . The adversarial loss function used is the least squares loss function, and the objective function is as follows:

$$\begin{aligned}\mathcal{L}_{adv}(G, D_Y) &= \mathbb{E}_{y \sim p_{data(y)}} \left[(D_Y(y \oplus m_y) - 1)^2 \right] + \mathbb{E}_{x \sim p_{data(x)}} \left[(D_Y(G(x) \oplus m_x))^2 \right] \\ \mathcal{L}_{adv}(F, D_X) &= \mathbb{E}_{y \sim p_{data(y)}} \left[(D_X(x \oplus m_x) - 1)^2 \right] + \mathbb{E}_{x \sim p_{data(x)}} \left[(D_X(F(y) \oplus m_y))^2 \right]\end{aligned}\quad (4)$$

where \oplus denotes a concatenation channel-wise operation that is calculated using a given real image and corresponding mask data. Since the test phase does not use the discriminator,

the mask can serve as an additional input. We confirmed that the discriminator learns more stably when provided with such additional information. In addition, the cycle-consistency loss and identity loss used for training are as follows:

$$\mathcal{L}_{cycle} = \mathbb{E}_{x \sim p_{data(x)}} [|F(G(x)) - x|_1] + \mathbb{E}_{y \sim p_{data(y)}} [|G(F(y)) - y|_1] \quad (5)$$

$$\mathcal{L}_{idt} = \mathbb{E}_{y \sim p_{data(y)}} [|G(y) - y|_1] + \mathbb{E}_{x \sim p_{data(x)}} [|F(x) - x|_1] \quad (6)$$

The purpose of the cycle-consistency loss is to limit transformations that do not reflect the characteristics of the input data. Therefore, it minimizes the difference when restoring fake images $G(x)$ and $F(y)$, generated from input images to the original domain. The identity loss helps preserve the color and tint of each domain. The background loss helps to prevent the transformation of the background regions between the mapped image and the input image, and is as follows:

$$\mathcal{L}_{bg} = \mathbb{E}_{x \sim p_{data(x)}} [|m_x^b \otimes (G(x) - x)|_1] + \mathbb{E}_{y \sim p_{data(y)}} [|m_y^b \otimes (F(y) - y)|_1] \quad (7)$$

where \otimes denotes an element-wise product operation calculated using a given real image and corresponding background mask data. Basically, the mask m_x represents the leaf region, so $m_x^b = 1 - m_x$ is calculated and used for the background loss function.

Finally, the loss function for training the proposed model consists of a combination of the four loss functions described above, where λ is a coefficient for adjusting the balance of each loss, and the final loss function is as follows:

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda_{cycle} \mathcal{L}_{cycle} + \lambda_{idt} \mathcal{L}_{idt} + \lambda_{bg} \mathcal{L}_{bg} \quad (8)$$

4. Results and Discussion

In this chapter, we conduct performance evaluation experiments to verify the validity of the proposed model. We verify the improved image quality with the proposed method, and then validate the classification performance with a dataset containing augmented plant leaf images.

4.1. Plant Village Dataset

The Plant Village dataset provides healthy and unhealthy leaf images for 14 crops, and there are various studies using it. An experiment was conducted using apple, potato, and grape leaf data among 14 crops, and the sample images are shown in Figures 4–6. As shown in Tables 3–5, the dataset configuration for the experiments of the proposed generative models basically extracted and used 500 data for each class. The minor class, which lacked the number of data, was randomly extracted whenever the learning data pair was configured. Since the proposed model performs only conversions for two domains, an experiment was conducted by constructing a pair of health and disease classes individually to learn a single model.



Figure 4. Sample image of disease-infected grapes (Black rot, Scab, Rust).



Figure 5. Sample image of disease-infected grapes (Black rot, Esca, Blight).



Figure 6. Sample image of disease-infected potatoes (Early blight, Late blight).

Table 3. Configuring the Apple dataset used to train the proposed model.

Class	Train	Support	Ratio
Healthy	500	1645	30.4%
Black rot	500	621	80.5%
Scab	500	630	79.4%
Rust	275	275	100%
Total	1775	3171	55.9%

Table 4. Configuring the Grape dataset used to train the proposed model.

Class	Train	Support	Ratio
Healthy	423	423	100%
Black rot	500	1180	42.4%
Esca	500	1383	36.2%
Blight	500	1076	46.5%
Total	1923	4062	47.3%

Table 5. Configuring the Potato dataset used to train the proposed model.

Class	Train	Support	Ratio
Healthy	152	152	100%
Early Blight	500	1000	50%
Late Blight	500	1000	50%
Total	1152	2152	53.5%

As shown in Tables 6–8, a fixed number of samples were used for testing from the original data for the dataset composition for the experiment of classification models. This is to construct a correct test set because there are very few classes. Therefore, in the case

of the apple and grape dataset, 200 samples were configured as a test set, and in the case of the potato dataset, 100 samples were configured as a test set. Afterward, the dataset composition for learning combines data not used for testing and augmented data and uses them for model learning. In the case of the apple and grape datasets, 500 samples were basically used for learning, and 400 samples were used for learning for the potato datasets.

Table 6. Configuring the Apple dataset used to train classification models.

Class	Train/Aug	Test	Support
Healthy	300/200	200	1645
Black rot	300/200	200	621
Scab	300/200	200	630
Rust	75/425	200	275
Total	2000	800	3171

Table 7. Configuring the Grape dataset used to train classification models.

Class	Train/Aug	Test	Support
Healthy	223/227	200	423
Black rot	300/200	200	1180
Esca	300/200	200	1383
Blight	300/200	200	1076
Total	2000	800	4062

Table 8. Configuring the Potato dataset used to train classification models.

Class	Train/Aug	Test	Support
Healthy	52/348	100	152
Early Blight	300/100	100	1000
Late Blight	300/100	100	1000
Total	1200	300	2152

4.2. Experiment Setup

The training parameters for learning the proposed model are shown in Table 9, and the results of the proposed model and the CycleGAN model are evaluated through FID (Frechet Inception Distance) [34]. In addition, through the image, it is verified whether the conversion problem for the background area and the clear disease texture are reflected.

Table 9. Parameters used in training.

Parameter	Value
Total Iteration	100,000
β_1, β_2 (Adam)	0.5/0.999
Learning Rate for G, D	0.0002
λ_{cycle}	10
λ_{idt}	5
λ_{bg}	10
r (Reduction ratio)	16

An experiment is conducted to verify whether the plant leaf images generated from the proposed method can further improve the performance of a classification model for early diagnosis of plants. The models used to evaluate classification performance are ResNet-18 [35], DenseNet-161 [36], MobileNet-v2 [37], and EfficientNet-b0 [38], and the experiment was conducted through fine-tuning from the pre-learned weights of the ImageNet [39] dataset. Precision, recall, and F1 Score were used to evaluate the performance of classification models.

4.3. Results

In the result of the apple leaf in Figure 7, the black rot disease image generated through CycleGAN does not preserve the shape of the original leaf image well by performing the translation on some background areas. On the other hand, the results generated through the proposed method reflect more distinct disease characteristics for both Rust and Black rot classes. In addition, the generated results well preserve the shape of the original real leaf image, so the validity of the proposed method can be confirmed. Even in the grape leaf results in Figure 7, the diseased leaf images generated through CycleGAN do not preserve the shape of the input image in common and reflect the poor texture. The cause of these problems is that the generator performs training to deceive the discriminator by evenly and indiscriminately trying to transform the entire training image area. In contrast, the translation results of the proposed method show that explicitly regulating transformation only in the leaf region requiring transformation can create visually valid plant leaf data. The result of potato leaf transformation in Figure 7 also shows that the transformation through the proposed method produces much more stable and realistic results. The result of potato leaf transformation in Figure 7 also shows that the transformation through the proposed method produces much more stable and realistic results. In addition, the generator can select the features necessary for conversion and unnecessary features well by partially performing conversion only on the necessary areas in the original image and preserving the color of the original image in some areas.

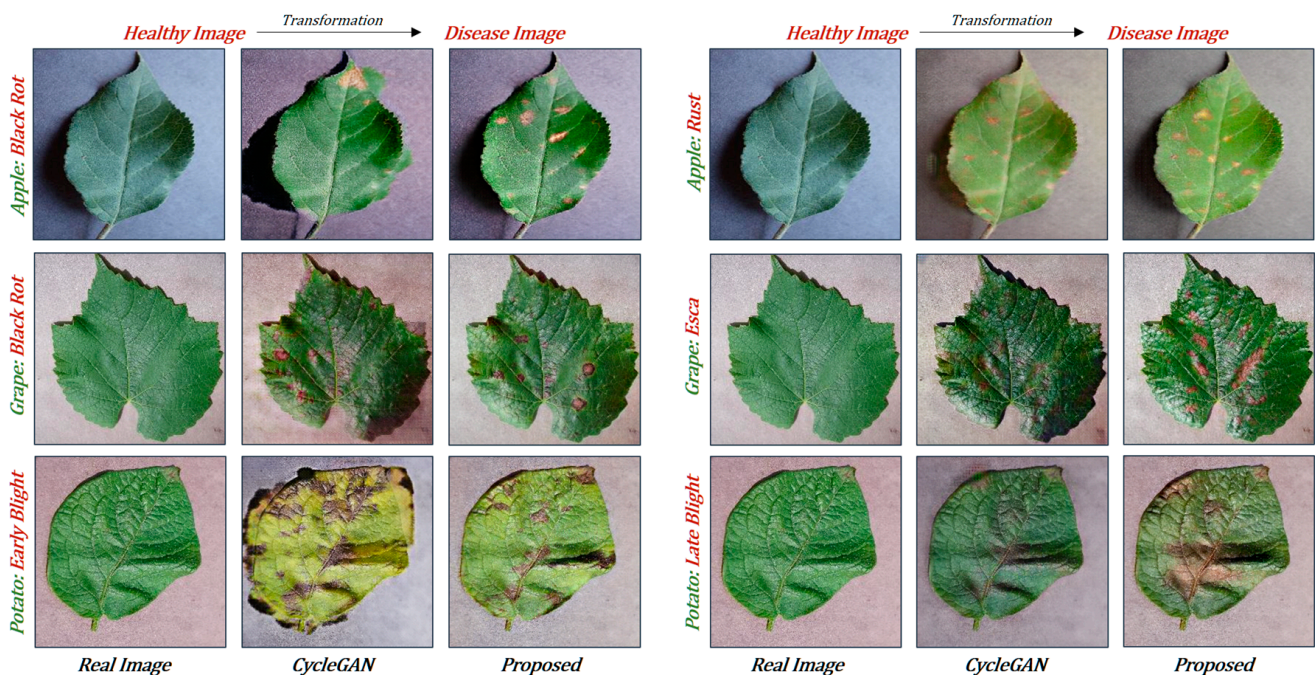


Figure 7. Comparison of disease-infected leaf image results generated by the proposed method and CycleGAN.

Quantitative evaluation of images generated by GANs is a very difficult problem. Initially, to evaluate the quality of the generated image, it was verified whether it follows the distribution of the training image well through Average Log-likelihood, but this is an indicator that is far from visual validity. Moreover, the generated images are high-dimensional data, making direct comparison difficult. To solve this problem, the proposed Frechet Inception Distance (FID) score calculates the Wasserstein-2 distance between two distributions through the average vector and covariance matrix of the feature vectors of the real image and the generated image. Therefore, the lower the better, and 0 corresponds to the highest score. Tables 10–12 are the results of evaluating the samples generated through these FID scores, and the samples generated through the proposed method as a whole are low. In

addition, the Black rot class in grape leaves showed the largest difference at 13.53 compared to the previous results.

Table 10. Comparison of generated Apple leaf image FID scores.

Class	CycleGAN	Our	Diff
Black rot	48.78	43.63	5.15
Scab	61.54	58.39	3.15
Rust	68.53	62.18	6.35

Table 11. Comparison of generated Grape leaf image FID scores.

Class	CycleGAN	Our	Diff
Black rot	39.36	25.83	13.53
Esca	44.59	38.63	5.960
Blight	59.29	52.98	6.310

Table 12. Comparison of generated Potato leaf image FID scores.

Class	CycleGAN	Our	Diff
Early Blight	68.10	58.70	9.4
Late Blight	70.37	63.30	7.07

Table 13 shows the performance of the classification models according to each augmented technique for the apple leaf data. Overall, models trained with augmented data through the proposed method and CycleGAN show better results than models trained with imbalanced data. This indicates that augmented disease images can improve actual generalization performance. In the overall experiment, each model generally showed excellent F1 scores when trained over augmented datasets via the proposed method, and the best results using EfficientNet and our method achieved 0.9995 F1 scores.

Table 13. Comparison of test results of Apple leaf classification models.

Model	Augmentation	Precision	Recall	F1 Score
ResNet 18	Imbalance	0.9768	0.9743	0.9755
	CycleGAN	0.9791	0.9762	0.9774
	Proposed	0.9793	0.9787	0.9790
DenseNet 161	Imbalance	0.9879	0.9859	0.9867
	CycleGAN	0.9912	0.9899	0.9905
	Proposed	0.9903	0.9926	0.9914
MobileNet v2	Imbalance	0.9855	0.9852	0.9853
	CycleGAN	0.9868	0.9845	0.9856
	Proposed	0.9882	0.9888	0.9885
EfficientNet b0	Imbalance	0.9969	0.9968	0.9968
	CycleGAN	0.9996	0.9985	0.9991
	Proposed	0.9998	0.9992	0.9995

Tables 14 and 15 show the performance of the classification models according to each augmentation technique for the grape and potato leaf data. Unlike the experiments in Table 13, the two datasets used for the experiments in Tables 14 and 15 lack samples from health classes. Although the results of learning augmented datasets via CycleGAN in some

classification models were excellent, the best were those utilizing EfficientNet and our method, which achieved F1 scores of 0.9994 and 0.999, respectively. We have experimentally confirmed that training a model on an augmented dataset usually solves the overfitting problem and improves overall performance. In addition, we confirmed that EfficientNet showed excellent performance even on imbalanced datasets and achieved the best results in all experiments when used with the proposed method in this paper.

Table 14. Comparison of test results of Grape leaf classification models.

Model	Augmentation	Precision	Recall	F1 Score
ResNet 18	Imbalance	0.9768	0.9743	0.9755
	CycleGAN	0.9791	0.9762	0.9774
	Proposed	0.9793	0.9787	0.9790
DenseNet 161	Imbalance	0.9879	0.9859	0.9867
	CycleGAN	0.9912	0.9899	0.9905
	Proposed	0.9903	0.9926	0.9914
MobileNet v2	Imbalance	0.9855	0.9852	0.9853
	CycleGAN	0.9868	0.9845	0.9856
	Proposed	0.9882	0.9888	0.9885
EfficientNet b0	Imbalance	0.9969	0.9968	0.9968
	CycleGAN	0.9996	0.9985	0.9991
	Proposed	0.9998	0.9992	0.9995

Table 15. Comparison of test results of Potato leaf classification models.

Model	Augmentation	Precision	Recall	F1 Score
ResNet 18	Imbalance	0.9768	0.9743	0.9755
	CycleGAN	0.9791	0.9762	0.9774
	Proposed	0.9793	0.9787	0.9790
DenseNet 161	Imbalance	0.9879	0.9859	0.9867
	CycleGAN	0.9912	0.9899	0.9905
	Proposed	0.9903	0.9926	0.9914
MobileNet v2	Imbalance	0.9855	0.9852	0.9853
	CycleGAN	0.9868	0.9845	0.9856
	Proposed	0.9882	0.9888	0.9885
EfficientNet b0	Imbalance	0.9969	0.9968	0.9968
	CycleGAN	0.9996	0.9985	0.9991
	Proposed	0.9998	0.9992	0.9995

Figure 8 shows the result of the two-dimensional visualization of feature vectors extracted from EfficientNet, which learned an augmented apple leaf dataset through the proposed augmentation method, through the t-distributed static neighbor embedding (t-SNE) technique. We find that the distribution of feature vectors extracted from the augmented 425 Rust samples and the real 75 Rust samples overlap, meaning that the generated image reflects the clear disease features of the real image class. Accordingly, the learned classification model may well extract features distinguished from other classes from the generated image.

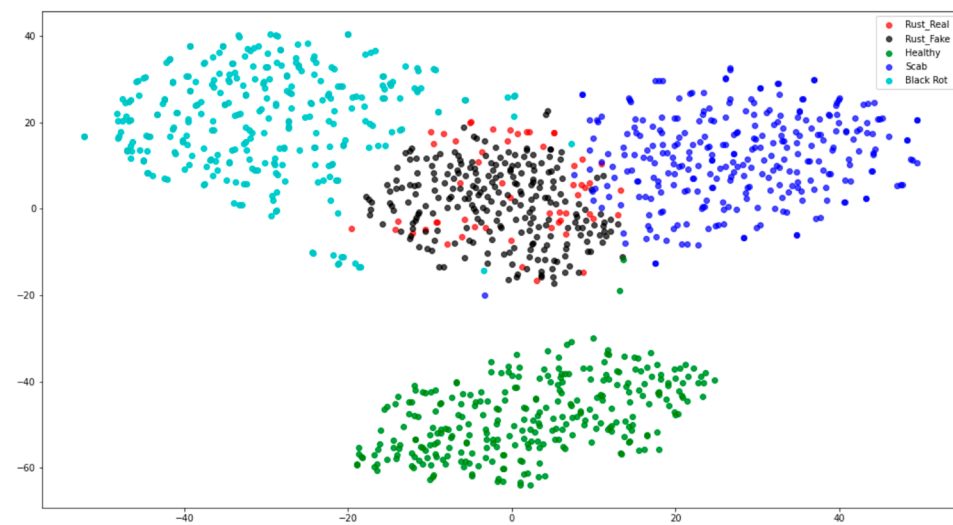


Figure 8. t-SNE visualization of feature vectors extracted from EfficientNet trained on the augmented apple rust dataset.

5. Discussion

Data generation research through GAN is actively being conducted in various domains. However, it is very difficult to train GAN models well, and usually requires large amounts of data for successful results. In this study, we focused on improving the expressiveness of the model through a concentration mechanism, not a method to increase the amount of data. Through this method, by fixing the number of data required for learning and partially modifying the model, it was possible to generate a clearer image than the result of the existing CycleGAN. The advantage of the proposed method is that it preserves the shape of the input image very well. These results show very clear transformation results compared to the results of GAN-based plant leaf augmentation techniques proposed by Wu et al. [24] and the Conditional GAN-based plant leaf augmentation techniques proposed by Abbas et al. [40]. However, our proposed method has a limitation in that it can only be augmented by the number of healthy leaf images as a 1:1 mapping function. To address these issues, there is room for improvement by conducting research through additional models such as StyleGAN [41] and StarGANv2 [42]. Also, some diseases can significantly damage the shape of healthy input leaves. Since the proposed method regulates maintaining the original input shape, it does not reflect these disease shapes. Therefore, it is necessary to proceed with future research on the generative model that considers the morphological transformation of diseased leaves. In addition to the disease texture, future research can be conducted through the UGATIT [43] model when considering additional transformations related to morphology. The UGATIT model is a model that can perform translation considering the shape of the target domain.

6. Conclusions

Recently, weather changes caused by climate change have had a great impact on crop productivity. If climate change continues as it is now, it is expected to be very difficult to maintain current crop production. To address this, deep learning-based plant early diagnosis studies are actively underway, but some rare disease samples have the problem of increasing the bias of learning models because they are difficult to collect.

Therefore, in this study, we study a data augmentation technique based on the Cycle-consistent Generative Adversarial Network to increase the diversity of rare disease samples by solving the above problem. The main content of the study is to address the problem of imbalances in plant disease data by generating diseased leaf images to be converted from healthy plant leaf images that are commonly collectible. We also propose a Pre-activation Residual Attention Block to improve the representation ability of the model,

which allows the generator to select features that should be considered more important during domain transformation. Furthermore, by learning the background loss function, the unnecessary transformation between the input leaf image and the fake leaf image generated from it was suppressed. Experiments were conducted to verify whether the plant leaf image generated from the proposed method could further improve the performance of the classification model for early diagnosis of plants, and Precision, Recall, and F1 Score were evaluated by indicators. The models used in the experiment were ResNet-18, DenseNet-161, MobileNet-v2, and EfficientNet-b0, all of which were fine-tuned using pre-trained weights. Experimental results confirmed that for all three plant leaves, the highest F1 score was achieved when the EfficientNet and the proposed data augmentation method were used together. In addition, improved FID scores and classification performance were achieved compared to CycleGAN as a whole. Through this, it was confirmed that the data augmentation method proposed in this study improves the generalization performance of various classification models.

Author Contributions: Conceptualization, B.M., T.K. and D.S. (Dongkyoo Shin); methodology, B.M., T.K. and D.S. (Dongkyoo Shin); software, B.M. and T.K.; validation, T.K., D.S. (Doingil Shin) and D.S. (Dongkyoo Shin); formal analysis, B.M., T.K. and D.S. (Dongkyoo Shin); investigation, B.M.; resources, B.M. and T.K.; data curation, B.M., T.K., D.S. (Dongil Sgin) and D.S. (Dongkyoo Shin); writing—original draft preparation, B.M.; writing—review and editing, T.K., D.S. (Doingil Shin) and D.S. (Dongkyoo Shin); visualization, B.M.; supervision, D.S. (Dongkyoo Shin); project administration, D.S. (Dongkyoo Shin); funding acquisition, D.S. (Dongkyoo Shin). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (421005-04).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The PlantVillage data set is available at the following link: <https://github.com/spMohanty/PlantVillage-Dataset/tree/master/raw> (accessed on 21 November 2022).

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

References

1. Malhi, G.S.; Kaur, M.; Kaushik, P. Impact of climate change on agriculture and its mitigation strategies: A review. *Sustainability* **2021**, *13*, 1318. [CrossRef]
2. Hunjan, M.S.; Lore, J.S. Climate change: Impact on plant pathogens, diseases, and their management. In *Crop Protection under Changing Climate*; Springer: Cham, Switzerland, 2020; pp. 85–100.
3. Ficke, A.; Cowger, C.; Bergstrom, G.; Brodal, G. Understanding yield loss and pathogen biology to improve disease management: Septoria nodorum blotch—a case study in wheat. *Plant Dis.* **2018**, *102*, 696–707. [CrossRef] [PubMed]
4. Wang, G.; Sun, Y.; Wang, J. Automatic image-based plant disease severity estimation using deep learning. *Comput. Intell. Neurosci.* **2017**, *2017*, 2917536. [CrossRef] [PubMed]
5. Tiwari, D.; Ashish, M.; Gangwar, N.; Sharma, A.; Patel, S.; Bhardwaj, S. Potato leaf diseases detection using deep learning. In Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 13–15 May 2020; pp. 461–466.
6. Agarwal, M.; Gupta, S.K.; Biswas, K.K. Development of Efficient CNN model for Tomato crop disease identification. *Sustain. Comput. Inform. Syst.* **2020**, *28*, 100407. [CrossRef]
7. Zhao, S.; Peng, Y.; Liu, J.; Wu, S. Tomato leaf disease diagnosis based on improved convolution neural network by attention module. *Agriculture* **2021**, *11*, 651. [CrossRef]
8. Barbedo, J.G.A. Plant disease identification from individual lesions and spots using deep learning. *Biosyst. Eng.* **2019**, *180*, 96–107. [CrossRef]
9. Zhou, C.; Zhou, S.; Xing, J.; Song, J. Tomato leaf disease identification by restructured deep residual dense network. *IEEE Access* **2021**, *9*, 28822–28831. [CrossRef]
10. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef] [PubMed]

11. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [\[CrossRef\]](#)
12. Hassan, S.M.; Jasinski, M.; Leonowicz, Z.; Jasinska, E.; Maji, A.K. Plant disease identification using shallow convolutional neural network. *Agronomy* **2021**, *11*, 2388. [\[CrossRef\]](#)
13. Lee, S.H.; Goëau, H.; Bonnet, P.; Joly, A. Attention-based recurrent neural network for plant disease classification. *Front. Plant Sci.* **2020**, *11*, 601250. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Longadge, R.; Dongre, S. Class imbalance problem in data mining review. *arXiv* **2013**, arXiv:1305.1707.
15. Barua, S.; Islam, M.M.; Yao, X.; Murase, K. MWMOTE—majority weighted minority oversampling technique for imbalanced data set learning. *IEEE Trans. Knowl. Data Eng.* **2012**, *26*, 405–425. [\[CrossRef\]](#)
16. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
17. Guo, M.H.; Xu, T.X.; Liu, J.J.; Liu, Z.N.; Jiang, P.T.; Mu, T.J.; Zhang, S.H.; Martin, R.R.; Cheng, M.M.; Hu, S.M. Attention mechanisms in computer vision: A survey. *Comput. Vis. Media* **2022**, *8*, 331–368. [\[CrossRef\]](#)
18. Jaderberg, M.; Simonyan, K.; Zisserman, A. Spatial transformer networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 2017–2025.
19. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
20. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual attention network for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3156–3164.
21. Park, J.; Woo, S.; Lee, J.Y.; Kweon, I.S. Bam: Bottleneck attention module. *arXiv* **2018**, arXiv:1807.06514.
22. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–18 September 2018; pp. 3–19.
23. Hughes, D.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv* **2015**, arXiv:1511.08060.
24. Wu, Q.; Chen, Y.; Meng, J. DCGAN-based data augmentation for tomato leaf disease identification. *IEEE Access* **2020**, *8*, 98716–98728. [\[CrossRef\]](#)
25. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
26. Deng, H.; Luo, D.; Chang, Z.; Li, H.; Yang, X. RAHC_GAN: A Data Augmentation Method for Tomato Leaf Disease Recognition. *Symmetry* **2021**, *13*, 1597. [\[CrossRef\]](#)
27. Cap, Q.H.; Uga, H.; Kagiwada, S.; Iyatomi, H. Leafgan: An effective data augmentation method for practical plant disease diagnosis. In *IEEE Transactions on Automation Science and Engineering*; IEEE: Piscataway, NJ, USA, 2020.
28. Xu, M.; Yoon, S.; Fuentes, A.; Yang, J.; Park, D.S. Style-Consistent Image Translation: A Novel Data Augmentation Paradigm to Improve Plant Disease Recognition. *Front. Plant Sci.* **2021**, *12*, 773142. [\[CrossRef\]](#) [\[PubMed\]](#)
29. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 630–645.
31. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv* **2016**, arXiv:1607.08022.
32. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv* **2018**, arXiv:1802.05957.
33. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
34. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6626–6637.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
36. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
37. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
38. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 6105–6114.
39. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
40. Abbas, A.; Jain, S.; Gour, M.; Vankudothu, S. Tomato plant disease detection using transfer learning with C-GAN synthetic images. *Comput. Electron. Agric.* **2021**, *187*, 106279. [\[CrossRef\]](#)

41. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8110–8119.
42. Choi, Y.; Uh, Y.; Yoo, J.; Ha, J.W. Stargan v2: Diverse image synthesis for multiple domains. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8188–8197.
43. Kim, J.; Kim, M.; Kang, H.; Lee, K. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. *arXiv* **2019**, arXiv:1907.10830.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.