Ecological Informatics 61 (2021) 101243

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Reconstruction of damaged herbarium leaves using deep learning techniques for improving classification accuracy

Burhan Rashid Hussein[a], Owais Ahmed Malik[a,b,*], Wee-Hong Ong[a,b], Johan Willem Frederik Slik[c]

[a] Digital Science, Faculty of Science, Universiti Brunei Darussalam, Gadong, Jalan Tungku Link, BE1410, Brunei Darussalam
[b] Institute of Applied Data Analytics, Universiti Brunei Darussalam, Gadong, Jalan Tungku Link, BE1410, Brunei Darussalam
[c] Environmental and Life Sciences, Faculty of Science, Universiti Brunei Darussalam, Gadong, Jalan Tungku Link, BE1410, Brunei Darussalam

ARTICLE INFO

Keywords:
Deep learning
Damaged leaves
Generative adversarial networks
Partial convolution
Plant species identification
Herbaria

ABSTRACT

Leaf is one of the most commonly used organs for species identification. The traditional identification process involves a manual analysis of individual dried or fresh leaf's features by the botanists. Recent advancements in computer vision techniques have assisted in automating the plants families/species identification process based on the digital images of leaves. However, most of the existing studies have focused on using datasets for fresh and intact leaves. A huge amount of data for preserved plants in the form of digitized herbaria specimens have not been effectively utilized for the task of automated identification because of the presence of damaged leaves in specimens. In this study, deep learning techniques have been proposed as a tool for reconstructing the damaged herbarium leaves in order to maximize the usefulness of the digitized specimens for automated plant identification task by increasing the number of individual samples of leaves. The reconstruction results of two different families of convolution neural networks (CNNs) have been compared for data from ten different plant families namely Anacardiaceae, Annonaceae, Dipterocarpaceae, Ebenaceae, Euphorbiaceae, Malvaceae, Phyllanthaceae, Polygalaceae, Rubiaceae and Sapotaceae. The performance of automated identification task was improved by more than 20% using the reconstructed leaves images as compared to using the original data (i.e. images of specimens with damaged leaves). This work evidently suggests that deep learning techniques can be utilized for reconstruction of damaged leaves even on a challenging herbarium leaves dataset.

1. Introduction

Herbaria collections worldwide have largely invested both in terms of time and money to preserve several hundreds of millions of specimens which serve as a crucial resource of plant data (Wäldchen et al., 2018). These collections consist of dried plants preserved on a special herbarium sheet. The specimens vary in colour, shape and texture even for the same species due to drying and pressing process (Tomaszewski and Górzkowska, 2016). Most of these specimens are yet to be identified at species level while others need to be updated following recent taxonomic knowledge (Carranza-Rojas et al., 2017). For traditional identification processes such as single-access and multi-access keys, visual similarity or difference can narrow down the search space for identification at different taxonomic level (Mata-Montero and Carranza-Rojas, 2016). These techniques require a complete specimen organ such as intact leaf to make correct decisions. Although this is desired, most of the herbarium leaves are damaged due to various factors such as preservation period, herbivores activities or accidental damage due to physical contact of the specimen which increase the complexity of identification process (Meineke et al., 2018).

The current digitization effort of these specimens is ongoing, however, most of the studies on plant species identification have largely focused on the application of computer vision techniques on data for fresh and intact plant leaves (Barré et al., 2017; Joly et al., 2014; Wäldchen and Mäder, 2018). This may be due to the fact that existing leaves of herbarium specimens are commonly found damaged to some extent. Nevertheless, most of the herbarium samples are of great importance to the botanical community although they are considered less useful by the computer vision community as they have lost their original shape, colour and texture due to damages (Carranza-Rojas et al., 2017).

An important factor that impacts the performance of machine

* Corresponding author at: Digital Science, Faculty of Science, Universiti Brunei Darussalam, Gadong, Jalan Tungku Link, BE1410, Brunei Darussalam.
E-mail address: owais.malik@ubd.edu.bn (O.A. Malik).

https://doi.org/10.1016/j.ecoinf.2021.101243
Received 11 September 2020; Received in revised form 22 January 2021; Accepted 22 January 2021
Available online 2 February 2021
1574-9541/

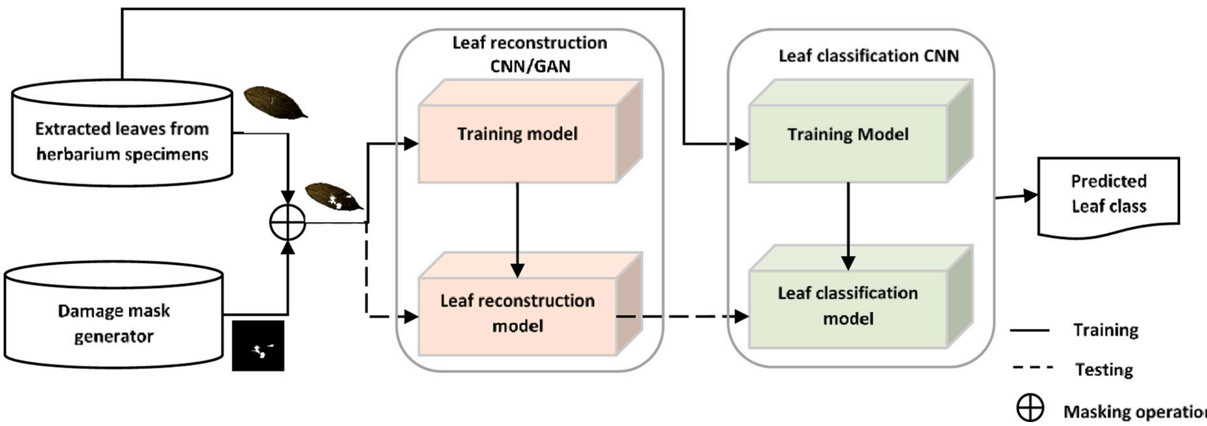

Fig. 1. Framework for training and evaluating damaged leaf reconstruction Model.

learning-based classifiers is the requirement of enough samples for each class in the data set. For automated plant species identification task, imbalanced species data generally deteriorate the overall performance of the machine learning models (Hussein et al., 2020a). Hence reconstruction and utilization of the existing damaged leaves could significantly improve the data samples for imbalance species especially for specimen found in herbarium (Carranza-Rojas et al., 2018; Hussein et al., 2020a). On the other hand, for herbarium species with damaged leaves yet to be identified, reconstructing the damaged leaf could be an important pre-processing step before the identification process. Studies such as (Carranza-Rojas and Mata-Montero, 2016; Corney et al., 2012; du Ji et al., 2006) have highlighted the impact of damaged leaves in degrading the performance of the identification system developed. Furthermore, the reconstruction of damaged leaves could also be used as a pre-processing step when building a herbarium leaf dataset repository (Mata-Montero and Carranza-Rojas, 2016).

Therefore, we approach this problem by proposing deep learning techniques as a tool for reconstructing the damaged herbarium leaves images. Deep learning techniques have successfully been used for image reconstruction in other domains (Isola et al., 2017; Liu et al., 2018). The reconstruction of damaged leaves images will help us to generate a repository of individual intact leaves which can be used for developing a more effective and accurate classifier for plant identification task. To the best of our knowledge, this is the first study to attempt using deep learning techniques for reconstructing damaged herbarium leaves images. The proposed method can also be adapted for reconstructing damaged leaves from fresh samples. In summary, the contribution of this work is as follows:

- We demonstrated the successful application of deep learning techniques for reconstructing the damaged leaves images of herbarium specimens and showed the improvement in the plant identification task using the reconstructed leaves data.
- We compared the performance of two state-of-the-art deep learning architectures, one being task-specific and the other being a general-purpose architecture for reconstructing the damaged leaves of herbarium specimens.
- We compiled a new dataset of individual herbarium leaves that can be used for training/testing different machine learning techniques.

2. Proposed model

This section provides a detailed explanation of the proposed model using deep learning technique for reconstructing damaged leaves of herbarium specimens. All the experiments performed for this study are discussed in the next section.

The main objective of this study is to develop a method that can be used in reconstructing damaged herbarium leaves images as a pre-processing step in building species identification system. Hence, we propose the use of deep learning techniques to reconstruct damaged leaf before passing it to a classifier. Two different families of convolution neural networks (CNNs) architectures are assessed by training and testing for developing damaged leaves reconstruction model. Fig. 1 depicts the framework used for training and evaluating the quality of the

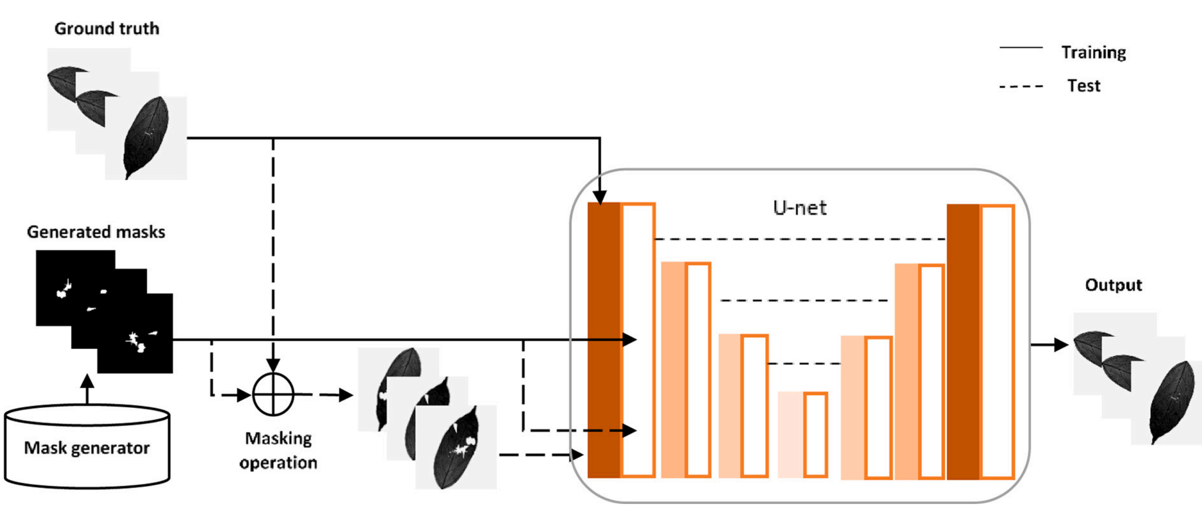

Fig. 2. Illustration of the PConv network.

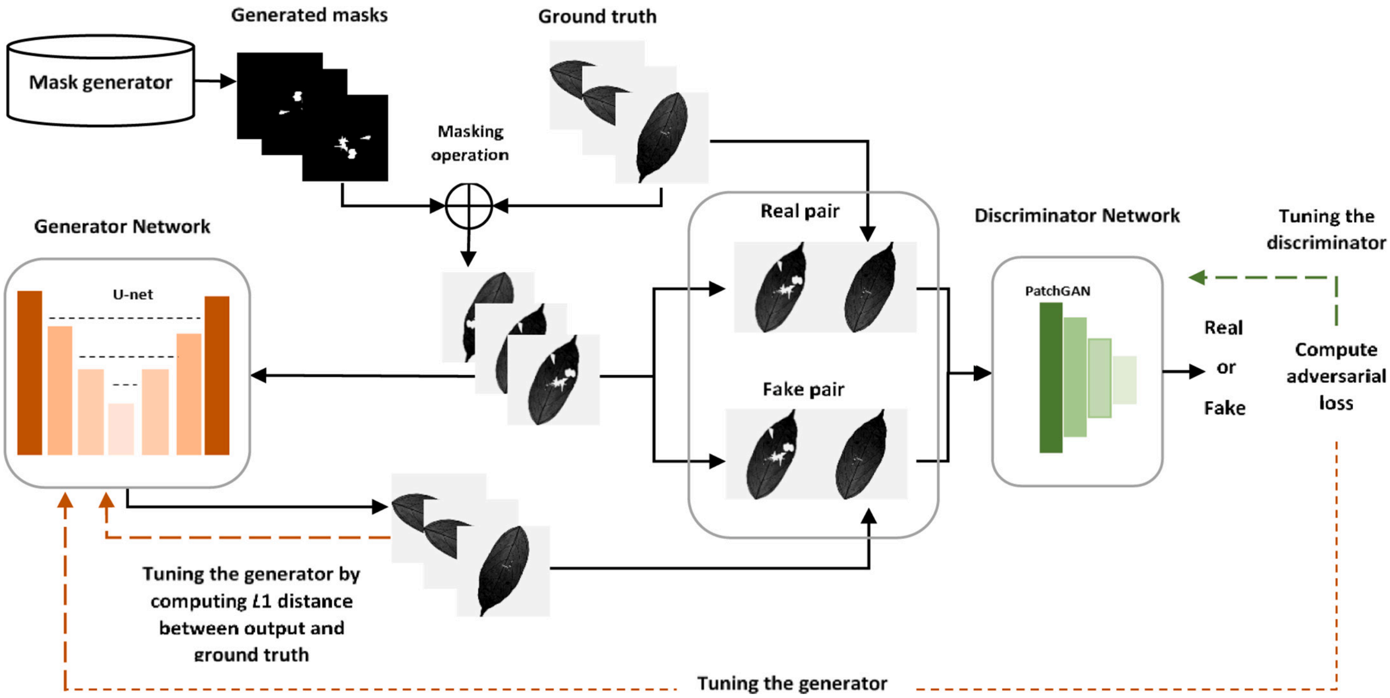

Fig. 3. Illustration of Pix2Pix framework.

reconstructed leaves. In order to artificially simulate the leaves damages, a simple algorithm is used to generate random masks which are employed with intact leaves for training/testing of reconstruction models. As a result, when a new image of a damaged leaf is given to a trained model, the model can successfully reconstruct the damaged part of the leaf. To quantitatively measure the performance of the reconstruction model, we trained a leaf classifier with intact leaves and then evaluated it using reconstructed leaves. All the steps involved are explained in the following sections.

2.1. Deep learning architectures for damaged herbarium leaves reconstruction

In this section, we provide a brief overview of the deep learning techniques used for image reconstruction and discuss their adaptation for our proposed framework.

2.1.1. Convolution neural networks

Convolution neural network has been widely adapted in herbarium settings due to their capability in learning hidden representation of images in various applications such as image classification, object detection and segmentation tasks (Hussein et al., 2020b; Mora-Fallas et al., 2019; Younis et al., 2018). Recent studies have proposed different modification of CNNs to account for image inpainting tasks but the results suffer from noise artefacts, overfitting and requirement of expensive post-processing to enhance model outputs (Liu et al., 2018).

In this study, we have applied a modified CNN called partial convolution (PConv) neural network that has achieved state-of-the-art results in various domains (Liu et al., 2018). We proposed a general framework for training and testing this architecture as illustrated in Fig. 2. This network uses a new convolutional layer called partial convolution layer which is designed to replace the normal convolution layers. The network uses a U-net like architecture which is adapted for the image inpainting task. During training, the network uses two inputs including the original image together with its mask indicating the region to be reconstructed. The inputs are passed to the network layers, composed of the two-components namely the partial convolution operation and the mask update function. This function is used to update the mask as the network progresses in deeper layers. Given an input image (damaged leaf image) with X being the current convolution window together with the mask M, the partial convolution operation at every location can be defined:

$$X' = \begin{cases} W^T(X^*M)\dfrac{sum(1)}{sum(M)} + b, if\ sum(M) > 0 \\ 0, otherwise \end{cases} \tag{1}$$

where * denotes the element-wise multiplication, '1' represents the same shape as M with all ones, and W and b represent the weights and bias of the convolution filter, respectively. The sum(1)/sum(M) is used as a scaling factor to adjust for a varying amount of valid (unmasked) input. The value '0' in eq. 1 indicates the input mask with no valid pixel i.e. a pixel with no information that can be used in inpainting other regions. From Eq. (1), it can be seen that the output of the partial convolution layer depends on the valid inputs of the image i.e. pixels of the image that can be used to in paint other masked regions. After each partial convolution operation, the mask is then updated by marking the location to be valid if the convolution was able to condition its output on at least one valid input values which are expressed as:

$$m' = \begin{cases} 1, if\ sum(M) > 0 \\ 0, otherwise \end{cases} \tag{2}$$

where m' is the updated mask. The application of partial convolution layer will ensure the whole mask will become ones if the input contains any valid pixel. During testing, the user needs to supply both leaf image and the mask showing a location of the leaf that needs to be reconstruction. The model uses various loss functions to target both per-pixel reconstruction and improve on how the predicted outputs for the inpainted region blends with the surrounding pixels.

2.1.2. Generative adversarial networks

The second technique evaluated for image reconstruction was the Generative Adversarial Networks (GANs). Recent progress in GANs has attracted new applications for herbarium specimen images (Villacis-llobet et al., 2019). Unlike traditional CNNs, GANs belong to the family of generative models that consist of two competing networks, a generator G and a discriminator D which are jointly trained where the generator tries to generate plausible images to fool the discriminator and the discriminator tries to distinguish the generated images as real or fake. GANs have successfully been used in various computer vision tasks such as image synthesis, image-to-image translation and image editing (Wang et al., 2020). For traditional GANs models, the generator uses only latent variable z to generate image y, G: z➔y. However, in the case of image-to-image translation tasks, the generated images need to resemble that of the source image. To solve this problem, conditional GANs (cGANs) takes additional information from the source domain x and from latent variable z to generate image y, G: {x, z}➔y.

In this framework, we considered the problem of reconstructing damaged leaves as an image-to-image translation task which takes a damaged leaf image as input and translates/reconstructs it as an undamaged leaf image. The proposed framework is based on the work proposed by Isola et al. (Pix2Pix) which is an extension of cGANs (Isola et al., 2017). Given a masked image as a source image x (damaged leaf image) and the ground truth y (undamaged leaf image), the objective of a cGAN can be expressed as shown in Eq. (3).

$$\mathscr{L}_{cGAN}(G, D) = \mathbb{E}x, y[logD(x, y)] + \mathbb{E}x, z[log(1 - D(x, G(x, z)))] \tag{3}$$

where generator G tries to minimize the loss against the discriminator D which is also trying to maximize. Hence, this ensures that G is trained to produce realistic images by an adversarially trained D which is also trained to do well in detecting fake images from the generator. Additionally, L1 loss is used by the generator to produce not only images that can fool the discriminator but also realistic images which are close to the ground truth images. The L1 loss is calculated as shown in Eq. (4).

$$\mathscr{L}_{L1}(G) = \mathbb{E}x, y, z[\|y - G(x, z)\|_1] \tag{4}$$

Combining the two objectives, the final equation becomes:

$$\frac{G^* = arg\ \ min\ \ max\ \ L_c\mathrm{GAN}(\mathrm{G}, D)}{GD + \lambda L_{L1}(G)} \tag{5}$$

Fig. 3 shows the general framework for training Pix2Pix model for reconstructing herbarium leaves. We used a simple algorithm to produce masks for simulating the artificial damage of the leaves. The produced mask is combined with a normal leaf to form a new damaged leaf image. For the training step, the model requires two inputs: one for the original image (damaged leaf) and other as its corresponding transformation (undamaged leaf). The generator network uses a U-net architecture with skips connections to allow sharing of information between layers. For the discriminator network, a CNN termed PatchGAN is used to classify a patch of the image as a real or fake rather than the whole image at once and the results from all patches are averaged to form the final score [15]. The weights of the generator are then updated via the adversarial loss together with the L1 loss between the generated image and the ground truth.

3. Experimental work

This section presents a detailed explanation of the experiments conducted based on the proposed models. All experiments were implemented in python using Keras with Tensorflow backend on a high-

Table 1
Dataset summary.

Family name	Training	Validation	Testing
Anacardiaceae	209	29	61
Annonaceae	199	28	58
Dipterocarpaceae	202	28	59
Ebenaceae	201	30	60
Euphorbiaceae	207	29	61
Malvaceae	207	29	61
Phyllanthaceae	210	30	60
Polygalaceae	200	28	58
Rubiaceae	205	29	60
Sapotaceae	200	28	58
Total	**2040**	**288**	**596**

performance computing facility equipped with an AMD Ryzen Threadripper 3960 × 24-Core Processor (48 CPUs)-3.8GHz, 64GB RAM and two NVIDIA GeForce RTX 2080 SUPER GPUs.

3.1. Dataset acquisition

The dataset of herbarium leaves was collected from Universiti Brunei Darussalam Herbarium (UBDH) which is hosting over 8000 plant species samples from the tropical region (Polgar et al., 2018). This dataset consists of intact herbarium leaves with a uniform background collected from ten different plant families namely Anacardiaceae, Annonaceae, Dipterocarpaceae, Ebenaceae, Euphorbiaceae, Malvaceae, Phyllanthaceae, Polygalaceae, Rubiaceae and Sapotaceae. The leaves were automatically extracted using a deep learning pipeline consisting of semantic segmentation model, connected component analysis and a single-leaf classifier trained on binary images to automatically detect intact leaves. The proposed models for reconstructing the damaged leaves will serve as an additional pre-processing step for developing herbarium leaf data repository. The summary of the dataset is presented in Table 1.

3.2. Generating artificial leaf damage

Collection of the actual damaged leaf structure in herbarium leaves can be a challenging task. This is because the damaged part of the leaves needs to be manually traced and then translated to the actual leaf for evaluating the proposed framework. On the other hand, simulation of artificial leaf damage that resembles the actual damage of the leaf is also not a trivial task. For these experiments, we opted for a simpler yet effective solution to generate training masks while keeping in mind realistic damages that appear in herbarium leaves. Various generated random shaped polygons were used to simulate the damaged area of the leaf. Since all leaves used for this study had a uniform background, the algorithm first segments the leaf pixels from the background using Otsu algorithm (Otsu, 1979) and then apply a flood-fill operation to ensure all the leaf pixels are enclosed within the leaf area. In the next step, a random point *c* is selected within a leaf area and a radius r is generated using Eq. (6). Using *c* as a center and *r* as a radius of a circle, random points p are then drawn around the circle by taking an angular step at each point using Eq. (8) while putting a point at random radius *r* using Eq. (9).

$$r = \sqrt{A/_{\pi}} \tag{6}$$

$$\delta\theta_i = U\left(\frac{2\pi}{n} - \epsilon, \frac{2\pi}{n} + \epsilon\right) \tag{7}$$

$$\theta_i = \theta_{i-1} + \frac{1}{k}\delta\theta_i, \textit{where } k = \sum \delta\theta_i \Big/ 2\pi \tag{8}$$

$$r_i = clip(\mathcal{N}(r, \sigma), 0, 2r) \tag{9}$$

where n represents the number of points to be used in formulating the polygon, θ_i gives the angular change, r_i is the radius of each point relative to the center c, U is a uniform distribution with minimum and maximum bound, $\mathcal{N}$ is a Gaussian distribution, and a *clip* function is used to threshold a value into a required range. The epsilon (ϵ) value in Eq. (7) is used to control the irregularity of the polygon while the sigma (σ) value in Eq. (9) is used to control the spikiness of the polygon points with both values ranging between 0 and 1. The values of σ, ϵ, r, c and n are used as a parameter to this algorithm. The generation of these random polygons is repeated multiple times until the desired leaf area is covered. An example of the generated mask together with the masked leaf is shown in Fig. 4.

3.3. Networks training process

Due to computation cost associated with training deep learning models, we experimented the application of the proposed framework with only two different image sizes i.e. 256 × 256 pixels and 512 × 512 pixels. The dataset was roughly partitioned into 70% training, 10% validation and the remaining 20% for testing as shown in Table 1. In order to demonstrate the effectiveness of the proposed approach for reconstruction of leaves images, we conducted experiments with different degrees of leaf damages when the leaves were damaged either 10% or 20% or 30% relative to their area. In total, 12 different reconstruction models were trained based on the different image sizes and damage thresholds. Besides, two classification models for plant identification were trained based on two different input image sizes. All the reported parameters were found to work well after preliminary experiments. Since all networks used present different architectural setups, different training procedures have been discussed. Each category of a CNN (i.e. PConv networks, Pix2Pix networks and classification networks) consisted of multiple trained models so the same training

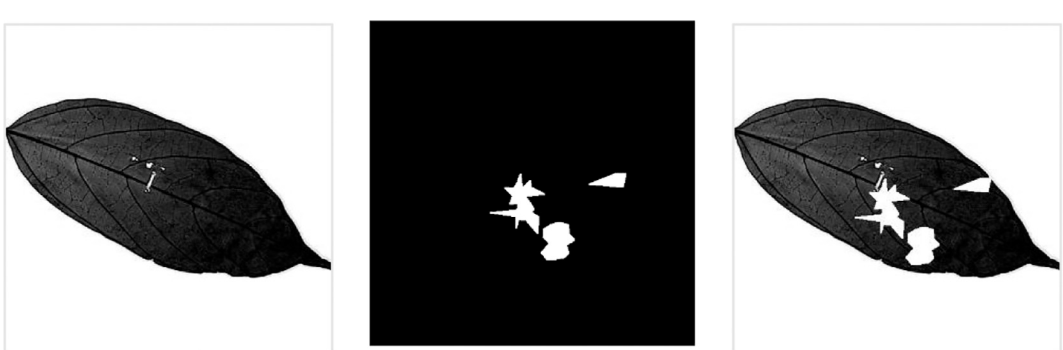

Fig. 4. (a)Sample image, (b) the generated mask and the masked image in (c) used for training reconstruction models.

parameters were used in each category unless otherwise specified. Where available, the original implementation reported earlier in literature were used (Isola et al., 2017; Liu et al., 2018).

3.3.1. PConv network

Instead of training the model from scratch, we opted for the fine-tuning process which helped in faster convergence, avoided overfitting and eliminated the limitations of the small dataset while adding performance gains (Yosinski et al., 2014). All networks were initialized with pre-trained weights which were computed and optimized using ImageNet dataset (Russakovsky et al., 2015). During the training process, the mask generated for simulating the damaged leaf area together with the masked image was used to train the network while an intact leaf was used as a ground truth. All input images were scaled between 0 and 1. In each iteration, images were augmented by flipping horizontally or vertically while a different mask was generated for that image. In other words, the same image received different mask during each iteration which was used as a form of regularization and hence enabled our models to generalize better.

The original implementation of PConv network required a two-phase training process (Liu et al., 2018). This is because the generated holes presented a problem for batch normalization layer when computing mean and variance of the hole pixels. In this study, all models were trained in a single process while freezing the batch normalization layer in the encoder part as the network used pre-trained weights with a small batch size of 1. The models were optimized using Adam optimizer with a learning rate of 0.0002 while using relu activation for encoder part and leakyRelu activation for the decoder part of the network. All models were trained for 50 epochs. For each training epoch, a model was evaluated using the validation set and was saved as the best model based on a validation loss.

3.3.2. Pix2Pix network

Unlike PConv networks, Pix2Pix models were trained from scratch as they have demonstrated to work well even of a small dataset (Isola et al., 2017). Network weights were randomly initialized from a gaussian distribution with a mean of 0 and a standard deviation of 0.02 with a batch size of 1. Both generators and discriminators were trained using Adam optimizer with a learning rate of 0.00009 and momentum of 0.5. To balance the training process, the discriminator losses were divided by 2 to slow down its learning speed relative to the generator networks as discriminators learn at a faster rate than the generators. For generator models, we set the weight of the L1 loss to 100 while keeping the GAN loss weight to 1. Other network parameters were left unchanged as reported in (Isola et al., 2017).

Training GANs is challenging as both generator and discriminator are competing and hence expected to reach at equilibrium (Pan et al., 2019). In this study, all Pix2Pix models were trained for 50 epochs. After every 500 iterations, the performance of the generator models were evaluated using both Structural Similarity Index Measure (SSIM) and the Peak Signal-To-Noise Ratio (PSNR) metrics and the model was saved as the best model if it has improved the performance on the validation set compared to previous iterations.

3.3.3. Classification

Pretrained classification models such as VGG16 have demonstrated to perform well on various plant species identification tasks (Pang and Lim, 2019). To train our classification model, we adapted the vgg16 network pre-trained on ImageNet dataset (Simonyan and Zisserman, 2014). The fully connected layers of the networks were replaced with a global max-pooling layer, a dropout layer with a dropout ratio of 0.5 as a network regularizer and single hidden layer with a total of 256 units for a larger image sized model(512 × 512) and 128 units for a smaller image size model (256 × 256). Both networks were trained for 20 epochs with a batch size of 32 and various data augmentation techniques such as height and width shift, flipping, zooming and brightness changes were applied. All input images were preprocessed by mean centering the image with ImageNet values and then rescaled between 0 and 1. Since our dataset was balanced, both networks were trained with a learning rate of 0.03, cross-entropy loss as a loss function optimized by Adam optimizer (Kingma and Ba, 2014). Table 2 summarizes the parameter used for training each model.

Table 2
Summary of the model parameters.

	PConv networks	Pix2Pix networks	Classification networks
Input dimension	256 × 256/512 × 512	256 × 256/512 × 512	256 × 256/512 × 512
Batch size	1	1	32
Learning rate	2e-4	9e-5	0.03
Optimizer	Adam optimizer	Adam optimizer	Adam optimizer
Loss function	per-pixel loss, perceptual loss, style loss and total variation loss	Adversarial loss and L1 loss	Cross-entropy loss
Epochs	50	50	20
Pre-trained network	Pre-trained on ImageNet	Trained from Scratch	Pre-trained on ImageNet

3.4. Performance evaluation

Quantitative evaluation of generative models is still a challenging task due to the existence of many possible solutions (Black et al., 2020). Similar to recent studies, all reconstruction models were evaluated using Peak Signal-To-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

3.4.1. Structural similarity index measure (SSIM)

SSIM is a perceptual metric which is used to quantify the similarity between two images provided that one of the images is the ground truth (Wang et al., 2004). Compared to other metrics such as mean square error between images, SSIM provides more detailed image quality assessment framework as it tries to mimic the visual perception of the human system. The metric is a weighted combination on the comparison of luminance, contrast, and structure between the two images and ranges between a minimum of 0 to a maximum of 1 where 1 been the perfect output as the ground truth. This metric was used as it takes the texture of the image into account which is important for our application domain.

3.4.2. Peak signal-to-noise ratio (PSNR)

PSNR represents the ratio between the maximum power of a signal and the power of distorting noise that affects the quality of representation. In term of images, the signal represents the original image data while the distorting noise is the error introduced by the reconstruction technique. PSNR tries to estimate the human perception on the quality of reconstruction of the signal (image) (Sara et al., 2019). Due to a wide range of many signal changes, the PSNR metric is commonly expressed in term of logarithmic decibel scale. We also reported the PSNR as it has commonly been used in different image-to-image translation to measure the quality of the generated image generated. The higher value of PSNR is desired to indicate the better quality of generated images.

4. Results and discussion

All reconstruction models were trained for 50 epochs with a total of 2040 leaves used for training, 288 for validation and the remaining 596 for testing. In general, all models performed consistently well based on the metrics used. While observing the performance of the models on a 256 × 256 image size, PConv model performed better with an average of 0.96 SSIM score across all damage ratios compared to Pix2Pix model which achieved an average SSIM score of 0.93 across the three-damage

Table 3
Performance comparison of the models for reconstructing damaged herbarium leaves for 256 × 256 image size. Mean ± Std. ↑[d] performance gain from the damaged leaf, ↓[a] performance loss from the actual leaves.

% of damage relative to the leaf area	PConv			Pix2Pix		
	SSIM	PSNR	Classification	SSIM	PSNR	Classification
10	0.971 ± 0.015	24.748 ± 0.603	0.238↑d, 0.012↓a	0.948 ± 0.019	28.821 ± 2.545	0.226↑d, 0.024↓a
20	0.96 ± 0.021	24.304 ± 0.821	0.39↑d, 0.022↓a	0.935 ± 0.027	27.639 ± 2.781	0.375↑d, 0.037↓a
30	0.95 ± 0.025	23.832 ± 0.978	0.461↑d, 0.035↓a	0.93 ± 0.032	27.63 ± 3.182	0.452↑d, 0.044↓a

Table 4
Performance comparison of the models for reconstructing damaged herbarium leaves for 512 × 512 image size.

% of damage relative to the leaf area	PConv			Pix2Pix		
	SSIM	PSNR	Classification	SSIM	PSNR	Classification
10	0.982 ± 0.011	29.427 ± 1.164	0.28↑d, 0↓a	0.972 ± 0.01	33.334 ± 3.012	0.268↑d, 0.012↓a
20	0.973 ± 0.014	28.386 ± 1.655	0.365↑d, 0.01↓a	0.945 ± 0.024	28.79 ± 3.398	0.308↑d, 0.067↓a
30	0.964 ± 0.02	27.337 ± 2.003	0.4↑d, 0.015↓a	0.937 ± 0.029	28.103 ± 3.274	0.315↑d, 0.1↓a

Mean ± Std. ↑[d] performance gain from the damaged leaf, ↓[a] performance loss from the actual leaves.

ratios on the test set. Despite this trend, the PSNR of the Pix2Pix model was observed to be consistently better than that of the Pconv models. However, while looking at the standard deviation, the Pconv models seemed to have more stable PSNR results compared to that of Pix2Pix model. The observed behavior depicts the challenging task of training GANs models.

Looking at the classification performance of both reconstruction networks on a 256 × 256 image size, there is a constant significance improvement of classification performance as the damaged area of the leaf increases (Table 3). On the other hand, the performance difference between PConv models and the Pix2Pix models compared to the performance of the actual leaf is marginally smaller with a drop of only 3–5% accuracy despite a 30% of the damaged leaf area been reconstructed. This indicates that both models have synthesized the leaf structure that is almost indistinguishable to the actual leaf when presented to a classification network and hence showing the robustness of both models in the reconstruction of the leaves.

When the image size was doubled, a similar pattern in the performance was observed for all models (Table 4). Both PConv and Pix2Pix models had an increase in performance in terms of PSNR compared to the smaller image size (256 × 256). This is likely caused by an increase in the number of pixels that models use during the training process. However, there is an increasing drop in classification performance as the damaged area of the leaf increases. This behavior can be noticed more while using Pix2Pix models which is also reflected by a higher standard deviation of PSNR compared to PConv models.

As depicted in Fig. 5, visual inspection indicates that the image generated by PConv model tends to better preserve the leaf shape which is not always the case for the Pix2Pix. This is expected as the Pconv model uses a mask so the shape of the leaf boundary is known prior to training. Pix2Pix model introduces new artefacts at the margin of the leaves which may be the explanation of having slightly lower performance than the PConv models for both SSIM and classification accuracy. On the other hand, the texture representation of Pconv models is also better than that of the Pix2Pix models. This may be attributed by the nature of the architecture which processes both the image and the mask, hence the model only tries to fill in the region of the mask given without distorting the other part of the image. However, the performance of the Pix2Pix models is only slightly lower in terms of SSIM and classification accuracy which is remarkable given that the model tries to learn all the leaf features explicitly without specific guidance.

As seen in Figs. 6 and 7, the performance of the damaged leaves was very low compared to both reconstructed leaves for all damage ratios. This is expected as damaged leaves have missing features that the classifier learned from undamaged leaves such as leaf shape and texture. The results show a significant improvement in classification performance when using the proposed models for reconstructed leaves compared to using damaged leaves. Tables 3 and 4 show that the results were still consistent with the other metrics used in this study. Although different parts of the leaves have been reconstructed, the images generated by both models still have more noisy texture and margin as compared to the ground truth image which may be the reason for not matching the performance with the ground truth leaves. Taking an example of leaf veins, both models have difficulties in capturing the vein patterns although visual inspection showed that the output of PConv models performs better than Pix2Pix models. Possible reasons could be the variability of the samples even on the same taxa which makes it a challenging task for models to learn these patterns when dealing with herbarium leaves. The problem needs to be further addressed by different ways including further tuning the loss weights to account for texture or devising new loss functions that will account for the problem in such kind of leaves.

4.1. Test on actual damaged leaves

The performance of the proposed framework was also tested on the actual damaged leaves. For demonstration purpose, both PConv and Pix2Pix models trained on 512 × 512 image size with 30% damage ratio were used as some of the damaged leaves tested had higher than 30% of the damaged leaf area. Since the PConv model requires to supply the leaf together with a mask, we manually set the leaf boundaries for the leaves which have been damaged on leaf edges (Fig. 8, first two rows). We then segmented the leaf using the Otsu algorithm followed by extraction of inner contours which were found inside the leaf boundary to extract the damaged part of the leaf.

As shown in Fig. 8, both PConv and Pix2Pix model reconstructed the damaged leaf when the leaf boundary was clear. Despite this, in some cases, the texture of the reconstructed part did not always blend well with the rest of the leaf. This was mostly caused by the diversity of leaf conditions even within the same taxa as the leaves have been collected and stored at different times and in different conditions. This problem can be alleviated by increasing the number of training samples for each taxon and perhaps developing more complex loss functions to penalize the model when producing different texture. Fig. 9 shows some failure cases of Pix2Pix model. Despite the model being mask-free as compared to PConv model, Pix2Pix model failed to capture some of the leaf boundaries. However, like PConv model, Pix2Pix model was able to reconstruct the inner parts of the damaged leaves. Improving leaf samples could enhance the model into better capturing of the leaf boundaries and further improve the overall quality of the synthesized texture.

Fig. 5. Sample output from the test set. The first row 10%, second row 20%, and third-row 30% of the leaf damage area.

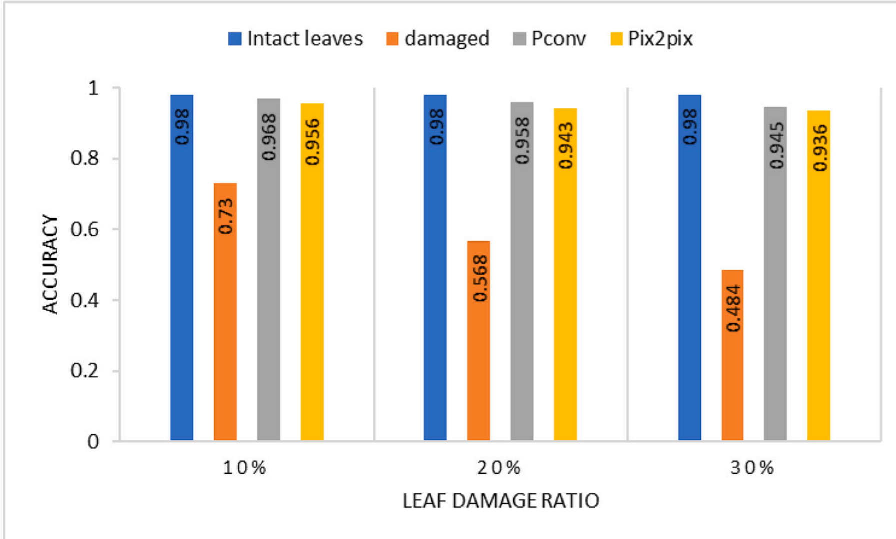

Fig. 6. Comparison of classification performance between intact, damaged, and reconstructed leaves for 256 × 256 image size.

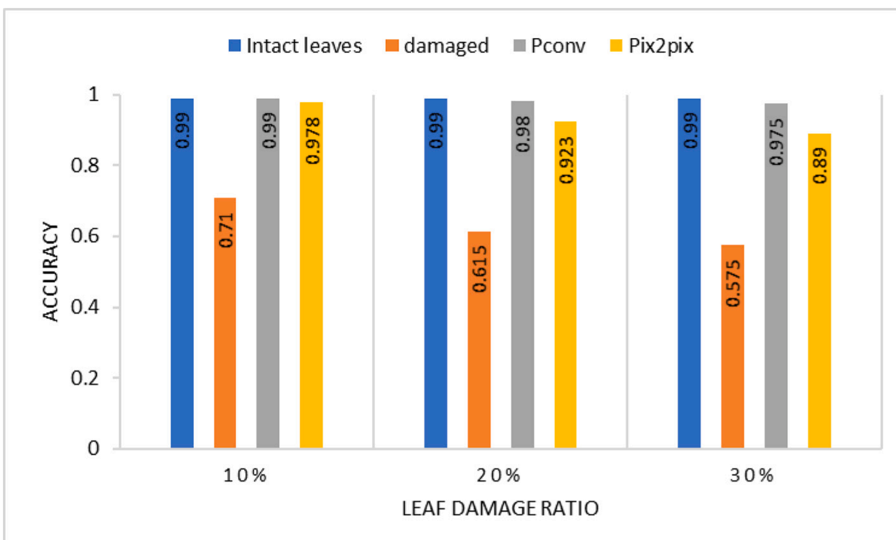

Fig. 7. Comparison of classification performance between intact, damaged, and reconstructed leaves for 512 × 512 image size.

5. Conclusions and future work

In this study, we proposed a deep learning solution for reconstruction of damaged leaves for digitized herbarium specimens. We used a challenging dataset comprising of a mixture of old and recently dried leaves which presented a perfect expected scenario from herbarium specimen images. The results of this study indicate that both PConv and Pix2Pix models can reconstruct the images of damaged leaves with decent accuracy. The performance of the plant's identification task has been found significantly improved using the reconstructed leaves images data set as compared to using the original data set containing the images of damaged leaves. These findings suggest the applicability of such techniques even in challenging herbarium specimens' images. These techniques could open up new research directions for the full utilization of herbarium specimens as most of them contain damaged leaves. It is worth noting that the proposed framework has been evaluated at a plants family-level classification which presents a mixture of different species within a single-family. Likely, the performance of the reconstruction model may slightly decrease due to the subtle changes within species from a single-family but that need to be further investigated.

Despite of a better performance of Pconv models compared to Pix2Pix models, generation of a mask can be challenging especially for the damage occurring on the leaf margins. This may require a user to know the prior shape of the leaf and, hence limiting its usability. Better techniques such as leaf shape template matching may be of great use to discover the prior shape of the leaf from existing leaf databases to automate the generation of leaf mask. In such cases, PConv models can be used in tasks related to leaf analysis where the texture of the leaf is needed to be reconstructed while the leaf boundaries are known in prior. Further improvements of Pix2Pix models could provide a better tool as the model automatically learns both leaf texture and shape which is more desired compared to a time-consuming task of generating a mask for PConv models.

In future, other image processing techniques will be investigated to improve the quality of generated images, specifically focusing on the texture, venation pattern and the shape as both PConv and Pix2Pix models have difficulties in recovering these features which are of great importance. Furthermore, an additional metric that will incorporate these features will be useful as the current existing metrics presents some difficulties in their quantitative interpretations. Automatically detecting

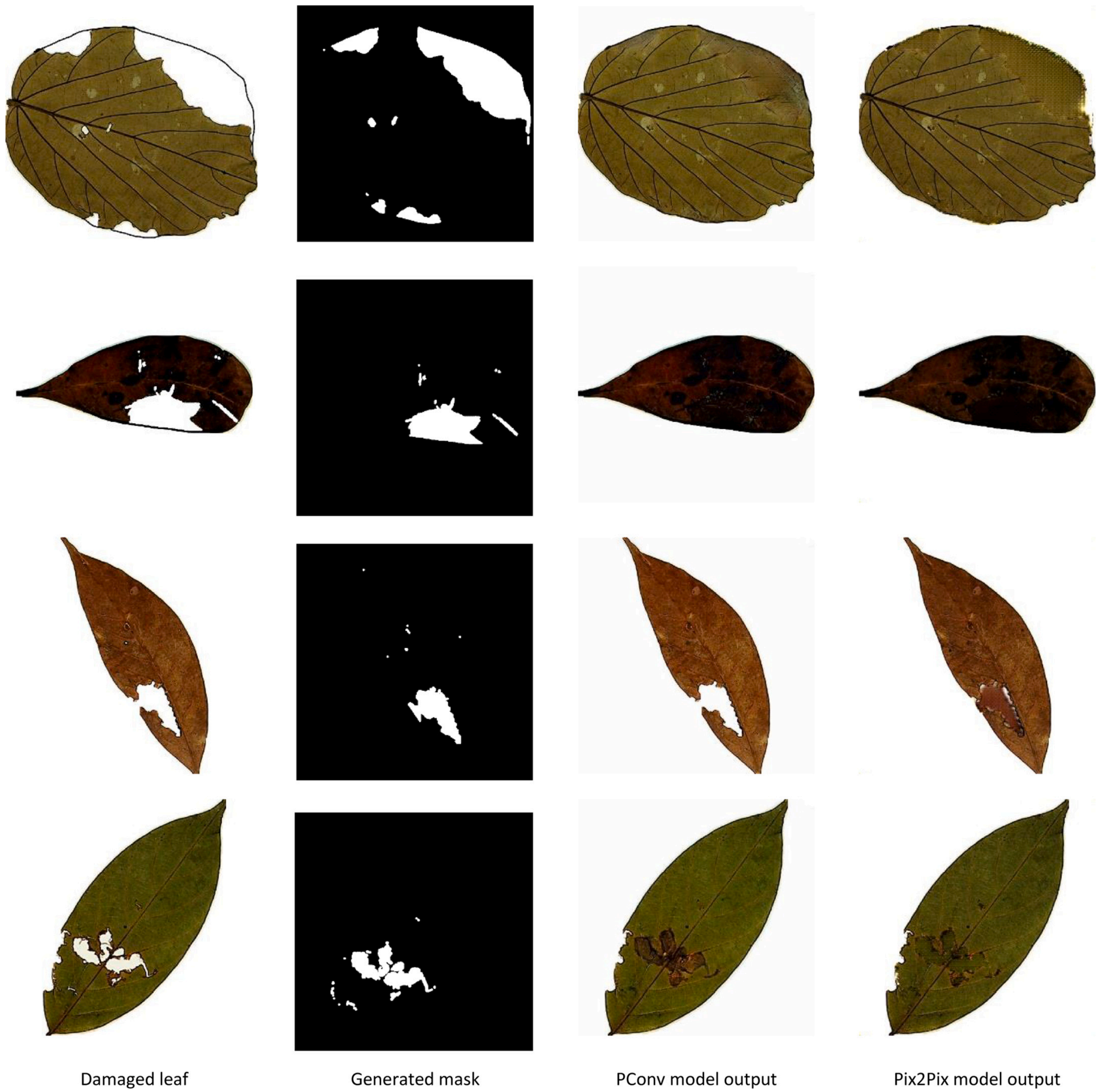

Fig. 8. Performance of reconstruction model on the actual damaged leaves.

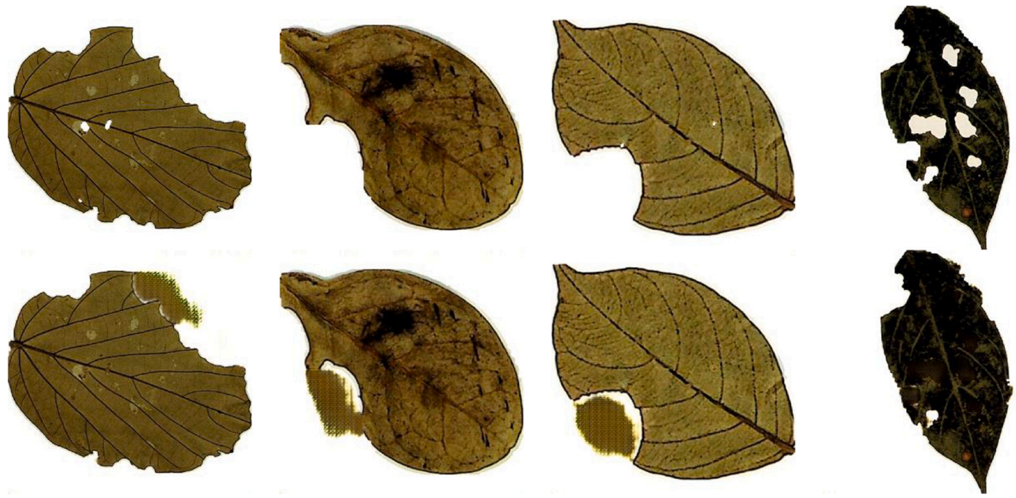

Fig. 9. Failure cases when the leaf boundary is not clear for Pix2Pix model.

damages will be useful to guide the model on leaf regions to reconstruct rather than distorting the intact part of the leaf. A better method for simulating the artificial leaf damages needs to be investigated to ensure the simulated damages resembles that of real damages occurring in herbarium leaves which could improve the generalization of reconstruction model. Finally, further analysis is required to assess the impact of the reconstructed leaves when used in herbarium plant species identification tasks ias the current experiment involved only three taxa.

Availability of the data and materials

The dataset used in this study is available at ubd-herbarium repository.

Declaration of Competing Interest

All authors of this paper have no potential conflict of interest.

Acknowledgements

Special thanks to Universiti Brunei Darussalam for sponsoring this research under the research grant UBD/RSCH/1.4/FICBF(b)/2018/011. We gratefully acknowledge the Institute of Applied Data Analytics, Universiti Brunei Darussalam for providing facilities and support.

References

Barré, P., Stöver, B.C., Müller, K.F., Steinhage, V., 2017. LeafNet: a computer vision system for automatic plant species identification. Ecol. Inform. 40, 50–56. https://doi.org/10.1016/j.ecoinf.2017.05.005.

Black, S., Keshavarz, S., Souvenir, R., 2020. Evaluation of Image Inpainting for Classification and Retrieval. In: IEEE Winter Conf. Appl. Comput. Vis, pp. 1060–1069.

Carranza-Rojas, J., Mata-Montero, E., 2016. Combining Leaf Shape and Texture for Costa Rican Plant Species. CLEI Electron. J. 19, 1–29. https://doi.org/10.19153/cleiej.19.1.7.

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A., 2017. Going deeper in the automated identification of herbarium specimens. BMC Evol. Biol. 17, 181. https://doi.org/10.1186/s12862-017-1014-z.

Carranza-Rojas, J., Joly, A., Goëau, H., Mata-Montero, E., Bonnet, P., 2018. Automated identification of herbarium specimens at different taxonomic levels. Multimed. Tools Appl. Environ. Biodivers. Inform. 151–167. https://doi.org/10.1007/978-3-319-76445-0_9.

Corney, D.P.A., Tang, H.L., Clark, J.Y., Hu, Y., Jin, J., 2012. Automating digital leaf measurement: the tooth, the whole tooth, and nothing but the tooth. PLoS One 7, 1–10. https://doi.org/10.1371/journal.pone.0042112.

du Ji, X., de Huang, S., Wang, X.F., Gu, X., 2006. Computer-Aided Plant Species Identification (CAPSI) based on leaf shape matching technique. Trans. Inst. Meas. Control. 28, 275–285. https://doi.org/10.1191/0142331206tim176oa.

Hussein, B.R., Malik, O.A., Ong, W.-H., Slik, J.W.F., 2020a. Automated classification of tropical plant species data based on machine learning techniques and leaf trait measurements. In: Comput. Sci. Technol. Springer, pp. 85–94.

Hussein, B.R., Malik, O.A., Ong, W.-H., Slik, J.W.F., 2020b. Semantic segmentation of herbarium specimens using deep learning techniques. In: Comput. Sci. Technol. Springer, pp. 321–330. https://doi.org/10.1007/978-981-15-0058-9_31.

Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, 2017. CVPR, pp. 5967–5976. https://doi.org/10.1109/CVPR.2017.632.

Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., Yahiaoui, I., Carré, J., Mouysset, E., Molino, J.F., Boujemaa, N., Barthélémy, D., 2014. Interactive plant identification based on social image data. Ecol. Inform. 23, 22–34. https://doi.org/10.1016/j.ecoinf.2013.07.006.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. Int. Conf. Learn. Represent. 1–15. http://arxiv.org/abs/1412.6980.

Liu, G., Reda, F.A., Shih, K.J., Wang, T.-C., Tao, A., Catanzaro, B., 2018. Image inpainting for irregular holes using partial convolutions. In: Proc. Eur. Conf. Comput. Vis, pp. 85–100.

Mata-Montero, E., Carranza-Rojas, J., 2016. Automated plant species identification: Challenges and opportunities. In: IFIP Adv. Inf. Commun. Technol, pp. 26–36. https://doi.org/10.1007/978-3-319-44447-5_3.

Meineke, E.K., Davis, C.C., Davies, T.J., 2018. The unrealized potential of herbaria for global change biology. Ecol. Monogr. 88, 505–525. https://doi.org/10.1002/ecm.1307.

Mora-Fallas, A., Goëau, H., Mazer, S., Love, N., Mata-Montero, E., Bonnet, P., Joly, A., 2019. Accelerating the automated detection, counting and measurements of reproductive organs in herbarium collections in the era of deep learning. Biodivers. Inf. Sci. Stand. 3, 4–6. https://doi.org/10.3897/biss.3.37341.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man. Cybern. 9, 62–66.

Pan, Z., Yu, W., Yi, X., Khan, A., Yuan, F., Zheng, Y., 2019. Recent progress on generative adversarial networks (GANs): a survey. IEEE Access. 7, 36322–36333. https://doi.org/10.1109/ACCESS.2019.2905015.

Pang, P.K., Lim, K.H., 2019. Review on automatic plant identification using computer vision approaches. IOP Conf. Ser. Mater. Sci. Eng 0–9. https://doi.org/10.1088/1757-899X/495/1/012032.

Polgar, G., Grafe, T.U., Pang, H.Y.K., Brahim, A., Cicuzza, D., Slik, J.W.F., 2018. The universiti Brunei darussalam biological collections: history, present assets, and future development. Raffles Bull. Zool. 66, 320–336.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Sara, U., Akter, M., Uddin, M.S., 2019. Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study. J. Comput. Commun. 07, 8–18. https://doi.org/10.4236/jcc.2019.73002.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. ArXiv Prepr 1556, 1–14. ArXiv1409. http://arxiv.org/abs/1409.1556.

Tomaszewski, D., Górzkowska, A., 2016. Is shape of a fresh and dried leaf the same? PLoS One 11, 1–14. https://doi.org/10.1371/journal.pone.0153071.

Villacis-llobet, J., Lucio-troya, M., Calvo-navarro, M., Calderon-Ramirez, S., Mata-Montero, Erick, 2019. A first glance into reversing senescence on herbarium sample images through conditional generative adversarial networks. In: Lat. Am. High Perform. Comput. Conf. Springer, pp. 438–447.

Wäldchen, J., Mäder, P., 2018. Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. Springer Netherlands. https://doi.org/10.1007/s11831-016-9206-z.

Wäldchen, J., Rzanny, M., Seeland, M., Mäder, P., 2018. Automated plant species identification—trends and future directions. PLoS Comput. Biol. 14, 1–19. https://doi.org/10.1371/journal.pcbi.1005993.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13, 600–612. https://doi.org/10.1109/TIP.2003.819861.

Wang, L., Chen, W., Yang, W., Bi, F., Yu, F.R., 2020. A state-of-the-art review on image synthesis with generative adversarial networks. IEEE Access. 8, 63514–63537. https://doi.org/10.1109/ACCESS.2020.2982224.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Adv. Neural Inf. Process. Syst, 27. Curran Associates, Inc, pp. 3320–3328. http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf.

Younis, S., Weiland, C., Hoehndorf, R., Dressler, S., Hickler, T., Seeger, B., Schmidt, M., 2018. Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks. Bot. Lett. 165, 377–383. https://doi.org/10.1080/23818107.2018.1446357.