Senescence Reversion in Plant Images using Perception and Unpaired Data

Esteban A. Esquivel-Barboza, *Student Member, IEEE* and Jose Carranza-Rojas

***Abstract*—Recent work on using herbarium images for automatic plant identification, and in particular to do domain adaptation to field images, has been promising. A potential way to address such domain adaptation problem is generative: Hallucinate how a herbarium image would have looked like when it was in the field, and use such synthetic, fresh and green image for plant identification purposes. Such generative task, called senescence reversion, has been poorly explored. To our knowledge, it has been studied only in terms of paired data, meaning, with pairs of dry and fresh images of the *same specimen*. Such paired data is hard to produce, curate and find. In this work we explore herbarium senescence reversion via unpaired data. The lack of pairs at specimen level presents its own challenges, as capturing the intricacies of each species depends on images of different specimens from different domains. We explore learning a mapping from a herbarium image to leaf plant images and vice-versa, aligning two models, one for each herbarium and leaf domain. We experiment against the state-of-the-art paired baseline Pix2Pix which yields a SSIM of 0.8986, compared to our unpaired approach that yields 0.8865, showing very similar results on reconstruction metrics regardless of our approach not having the luxury of pairs by specimen. Additionally, we apply perceptual loss in order to improve the natural look of the synthetic images. The balance of how much perception is good to avoid reconstruction problems is also studied. Lastly, using a new unpaired dataset built by ourselves, our results show that using a low λ value from 0.025 to 0.05 for perceptual loss, helps getting lower Fráchet Inception Distances and higher Inception Scores. To our knowledge, no other work has focused on reverting senescence of herbarium sheet images based on unpaired data.**

***Index Terms*—Image-to-image translation, unpaired data, CycleGAN, herbaria, senescence reversion.**

I. Introduction

There are around 375 million specimens captured in herbaria institutions around the world [1]. Many have been digitized, allowing their usage directly for automatic plant identification [2]. Furthermore, such herbarium images have also been used for domain adaptation, helping on the identification of species that lack lots of field images. The latest PlantCLEF challenges [3] introduced a new domain adaptation task from herbaria images to field images in this same line of work. However, very few studies in the plant identification domain use generative models to help on the plant classification tasks. Generative models could help on classification tasks by synthesizing images for those species that have less fresh images for model training. Another use of herbaria images in a generative way, could be to hallucinate how extinct plants would have looked like.

In terms of data augmentation, using Generative Adversarial Networks (GANs) as a mean to do data augmentation for classifiers has shown to be effective in some other domains [4], [5]. In particular in the plant domain, converting one herbarium image into a fresh looking one would require, in its basic supervised form, pairs of images of the same specimen, such as the fresh looking one from the time when it was collected, and the herbarium one when the same plant specimen was pressed and dried. This would smooth the learning process, as the shape and other visual properties of the specimen would be present in both herbarium and fresh images. However, collecting such paired dataset seems very unlikely. Building paired datasets of plant images is costly, time-consuming, and even sometimes impossible, such as the case of already extinct plants. It is also bound to have very few species and specimens, given the former constraints.

On the other hand, creating unpaired datasets from both herbarium and fresh looking images of the same species (not specimen) is easier. There are datasets with herbaria images in services such as iDigBio, GBIF and iNaturalist. Regarding fresh looking plant images, finding datasets is simple as well, with applications such as Pl@ntNet [6]. This means finding ways to do such unpaired image-to-image translation from a herbaria dataset to a fresh dataset, using the intersection of the species from both dataset types, would enable existing datasets to be used in an additional machine learning task for biodiversity conservation. We refer to unpaired image-to-image translation as the task of converting an image into another one, without the need of specific pairs of images and their converted counterparts.

In general, publications of generative work with plant datasets are very limited. Style transfer and image-to-image translation on plant datasets have being mostly unexplored. Previous work on leaf reconstruction, or filling holes in leaves, can be found in [7]. It uses, however, paired images for the reconstruction. Concerning senescence reversion, to our knowledge, there is only one previous work in this domain. [8] explores reversing the senescence of a small, paired dataset of only 3 species of plants. To our knowledge, no other paired studies have been done towards senescence reversion, let alone unpaired data senescence reversion.

General unpaired image-to-image translation using GANs has gained traction since the release of CycleGAN [9]. Also, the use of perceptual losses has proven to be important to generate better looking, more realistic images [10]. It has also been shown there is a need to keep a balance between

Esteban A. Esquivel-Barboza is with the Computer Engineering Department, Costa Rica Institute of Technology e-mail:eesquivel@estudiantec.cr

Jose Carranza-Rojas is with Microsoft Research and Computer Engineering Department, Costa Rica Institute of Technology e-mail: jose.m.carranza@microsoft.com

reconstruction loss and perceptual loss strength, to balance reconstruction accuracy and good perceptual feeling from the image [11].

In this work we do senescence reversion by using unpaired images of plants. We use existing herbaria and fresh-looking images without pairs of the same specimen in both domains. We also employ perceptual loss, and measure in which quantity the best results are yield, to balance out both reconstruction and perception.

In general, our contributions in this work are the following:

1) We curate a plant dataset with images from 195 species that contains images of herbarium sheets and green leaves on a white background. Such dataset is unpaired at specimen level.
2) We pioneer senescence reversion work using unpaired data, meaning, without images of the same specimen of both herbarium and fresh domains.
3) We propose the use of perception losses with a CycleGAN model to help preserve the plants natural looking when doing the unpaired image-to-image translation.
4) We measure how classifiers are affected when using hallucinated images as additional data during training, but tested with real images.

The rest of the document is as follows: Section II contains the related work. Section III depicts the datasets, loss functions and architectures used. Section IV contains a description of the experiments with their respective results. Section V contains the discussion over our results, and Section VI depicts our conclusions. Finally, Section VII lists potential future work in this line of research.

II. Related Work

Senescence inversion is, at its core, an image-to-image translation task. Literature on image-to-image is extensive, but very few studies have focused on reverting senescence of plant images. In the broad image-to-image translation literature, both paired and unpaired approaches are studied [12], [12].

A. Paired Herbarium Image Translation

One of the flagship works to explore image-to-image tasks using paired image translation techniques is known as Pix2Pix [12], that extends the original GAN work [13]. Both the generator G and the discriminator D sample an image x from the "origin" domain. Then another image y from the "target" domain is used jointly with x only on the discriminator, and the random vector z, jointly with x, is used as input to G. [12] define their GAN loss function as seen in (1), where p_x is the distribution of samples from domain X, p_y is the distribution of samples from domain Y, and p_z is a random distribution to sample noise

$$\begin{aligned}\mathcal{L}_{\mathrm{cGAN}}(G,D) =&\mathbb{E}_{x\sim p_x,y\sim p_y}[logD(x,y)]\\ &+\mathbb{E}_{x\sim p_x,z\sim p_z}[log(1-D(x,G(x,z)))].\end{aligned} \tag{1}$$

[12] also noticed that the generator learned to ignore the noise, so the noise vector is not sampled. Instead, they sample only the input x from p_x. Also, they added a Manhattan distance (l^1-norm, see (2)) for the GAN objective, to measure the distance of the generated image from the ground truth. This can only be done in the paired data domain

$$\mathcal{L}_{\mathrm{L1}}(G) = \mathbb{E}_{x\sim p_x,y\sim p_y,z\sim p_z}[||y-G(x,z)||_1]. \tag{2}$$

The final loss function used in Pix2Pix is found in (3), where λ is an hyper-parameter to weight the L^1 regularization

$$\mathcal{L}_{\mathrm{Pix2Pix}}(G,D) = \mathcal{L}_{\mathrm{cGAN}}(G,D) + \lambda\mathcal{L}_{\mathrm{L1}}(G). \tag{3}$$

To the best of our knowledge, the work of [8] is the only one to address the problem of reversing the senescence of plants. They trained a Pix2Pix [12] model with a new handmade dataset named *LeavesDryFresh*. The dataset consists of dried and fresh leaves from three species (*P. lutea*, *L. camara* and *C. roseus*), 20 photos of dry leaves per species and 20 photos of fresh leaves per species, for a total of 120 images. The dataset only contains pictures from the frontal part of the leaves. In their experiments they achieved an average Structure Similarity Index Measure (SSIM) [14] score of 0.9.

The work by [8] could be expanded in many fronts. It was based on paired images, and to our knowledge, no other work has focused on unpaired images. Additionally the dataset is limited, containing only samples from three species and only images from the frontal side of the leaves. Using herbarium images and green leaves is a much complicated problem, but would enable the usage of herbarium images for fresh plant identification, among other tasks.

Other line of work that explores generative tasks in herbarium data is [7]. They compare a Pix2Pix model and a fine-tuned U-Net [15] for the task of damaged leaf reconstruction. After such reconstruction, they test the reconstructed leaf images with a classifier of dry leaves based on VGG16 [16]. They collected a new herbarium dataset from the University Brunei Darussalam Herbarium consisting of $2,040$ images of intact herbarium leaves from 10 families. They simulate holes in the leaves using the algorithm of [17]. In all instances they got more than 0.93 on SSIM and more than 23.82dB on Peak Signal-to-Noise Ratio (PSNR).

B. Unpaired Image Translation

A more challenging problem is translating image-to-image without pairs of corresponding images for supervision [12]. In the context of herbarium data it means we do not have the same specimen in both herbarium and fresh domains. We did not find any unpaired image translation work on herbarium datasets.

Nowadays the main approaches to tackle this task use an adversarial loss [18]. One tries to find an intermediate latent space between both domains [4], or two style and one content latent spaces are used [19]. Another approach is by using two GANs and enforcing cycle consistency [9], [20], [21]. It consists on passing an image through one generator, and then the resulting image through the other generator, getting back the original one. This idea was popularized by [9] and we use it as our base model. Recent work of [22] studies a new approach by using a contrastive loss based on [23]. This approach does not rely on the cycle consistency concept, lowering the quantity of parameters by half.

1) CycleGAN: The authors propose to not only learn the mapping function $G: X \rightarrow Y$, but a second mapping function $F: Y \rightarrow X$, essentially training two generators and two discriminators D_X and D_Y.

They applied the same adversarial loss from [13], but added a new term to the loss function, they called "cycle consistency loss". This loss consists on the l_1-norm of the original image from the output of a cycle ($F(G(x))$ and $G(F(y))$), as shown in (4)

$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G, F) =& \mathbb{E}_{x\sim p_x}[||F(G(x)) - x||_1] \\ &+ \mathbb{E}_{y\sim p_y}[||G(F(y)) - y||_1].\end{aligned} \tag{4}$$

This cycle consistency loss is weighted with a hyper-parameter λ in the final loss function, as shown by (5)

$$\begin{aligned}\mathcal{L}_{\text{CycleGan}}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X) \\ &+ \lambda\mathcal{L}_{\text{cyc}}(G, F) \\ &+ \lambda\mathcal{L}_{\text{L1}}(G) + \lambda\mathcal{L}_{\text{L1}}(F).\end{aligned} \tag{5}$$

2) Perceptual Loss: Introduced in [24], the perceptual loss concept complements the idea of pure pixel based losses, also known as reconstruction losses. It consists on using high-level features extracted from a pretrained networks for loss calculation. To extract this high-level features, an image is passed through a neural network and activations on specific layers are used to compare two images. The intuition behind these high-level features is to attempt to generate images that make sense at a high, human level [11], [24], [25].

C. Datasets

Most of the existing datasets are focused on identification of the species, and very few on generative tasks. The following are the most important datasets relevant to our generative research:

- Herbarium255 [2]: Dataset used on pioneer work on herbarium sheet images. This dataset is focused on plant identification, with 255 species from Costa Rica.
- CRLeaves [26]: Includes a total of 255 species from the Central Plateau in Costa Rica. It consists of $7,262$ images of leaves with uniform backgrounds.
- PlantCLEF 2020 [27]: Dataset focused on domain adaptation for classification. Uses a mix of $320,000$ herbarium sheets and more than $3,000$ field images. The field images are taken in the wild, leading to higher levels of complexity.
- LeavesDryFresh [8]: The most closely related dataset to our work. This dataset consists of 60 images of dry leaves and 60 images of fresh leaves, for a total of 120 images from 3 different species.

III. Methods

The following section depicts the methodology used, concerning datasets, models and loss functions.

A. Datasets

For our experiments we used two different datasets. We used them in an unpaired fashion, regardless if the original dataset contained pairs of images of the same specimen. The datasets are:

- LeavesDryFresh: We did not make any modifications to this dataset, aside from sampling in unpaired fashion. We used it to compare our unpaired approach with the previous paired approach [8]. Section II-C contains further details of this dataset.
- Unpaired Leaves Image Dataset (UnLID): A subset of species from the intersection of Herbarium255 [2] and CRLeaves [26]. The dataset has both herbarium sheet images and green fresh leaf images of the same species, but not same specimens. We removed the species that did not contain images in both domains, as our training regime requires pairs of images at species level, not at specimen level. The dataset consists of $7,514$ images of green leaves and $4,868$ images of herbarium sheets. For those species with different amount of images on each domain, we repeat the images of the domain with fewer images, yielding a total of $9,539$ images across 195 species. Fig. 1 shows a small sample of this dataset.

B. Models

We use a CycleGAN [9] as the base model. We learn mapping functions from/to two different domains. For further details of such model, please refer to Section II-B1. We also make use of a VGG16 model trained on ImageNet [28], used as perceptual loss to help on keeping a balance between cycle consistency and perception.

The overall approach is shown in Fig. 2, which depicts the learning process. We randomly sample two unpaired images from the herbarium and the leaves domain, which match at species level (but not at specimen). We then pass them through the CycleGAN model, which learns mappings for each domain, given the two generators. Additionally, two discriminators also attempt to discriminate both domain's real versus fake images. We apply the "cycle consistency loss" explained in Section II-B1 using the output of both generators. Finally, we use VGG16 to extract the activations of the layer `relu2_2`, and use it as perceptual loss.

C. Losses

1) Perceptual loss: We use a pretrained VGG16 [16] model on the ImageNet dataset [28] as the loss network. We use VGG high-level features comparisons as an additional loss function for the generator networks. We compute the weighted and squared distance of the activation map from both target and generated images, as shown in (6), where y and $\hat{y}$ are the target and generated image respectively, $\phi_j(x)$ is the activation map of x in the layer j, C is the number of channels, and H is the height and W is the width of the activation map

$$l_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j}||\phi_j(\hat{y}) - \phi_j(y)||_2^2. \tag{6}$$

(a) (b) (c)

Fig. 1. Samples of the UnLID. From left to right: *Barleria Oenotheroides* (a), *Malvaviscus Arboreus* (b) and *Piper Tuberculatum* (c).

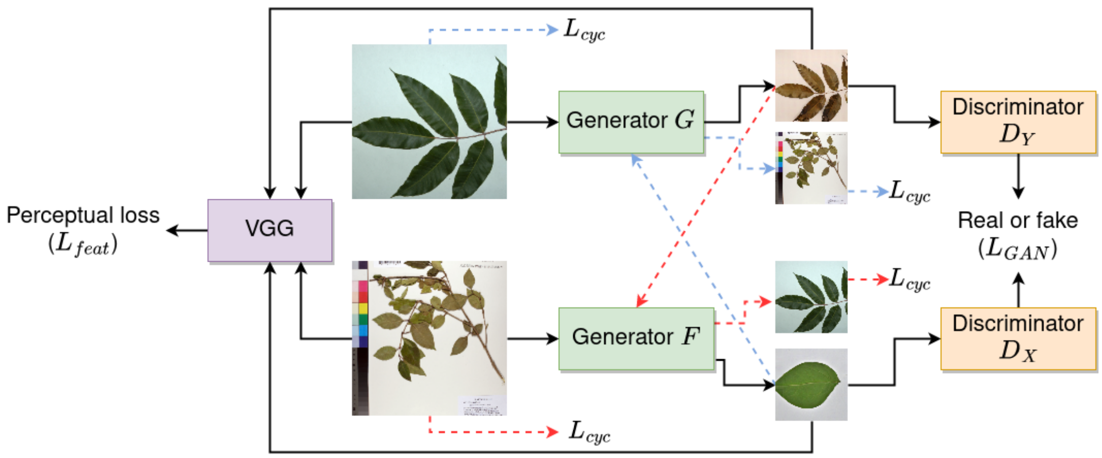


Fig. 2. Diagram of the proposed architecture using CycleGAN and perceptual loss.

2) Final loss: Lastly, we combine both the perceptual loss with the losses used in CycleGAN. We argue that using perceptual loss will help preserving the perceptual quality on the generated images beyond reconstruction. To balance the perceptual loss we add a weight λ_p to its loss term. The final loss used in this work is presented in the (7), where $\mathcal{L}_{\text{feat}}$ is the perception loss

$$\begin{aligned}\mathcal{L}_{\text{Final}}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{CycleGan}}(G, F, D_X, D_Y) \\ &+ \lambda_p \mathcal{L}_{\text{feat}}(G) + \lambda_p \mathcal{L}_{\text{feat}}(F).\end{aligned} \tag{7}$$

IV. Experiments and Results

This section describes the different senescence reversion experiments, from quantitative to qualitative perspectives. First a benchmarking experiment against a paired approach is executed to measure reconstruction metrics. Then a experiment with our own unpaired dataset UnLID is studied, to measure non-reference metrics as well as the influence of perceptual loss. Finally, a third experiment measures the performance of a classification model when hallucinated leaf images are used during training.

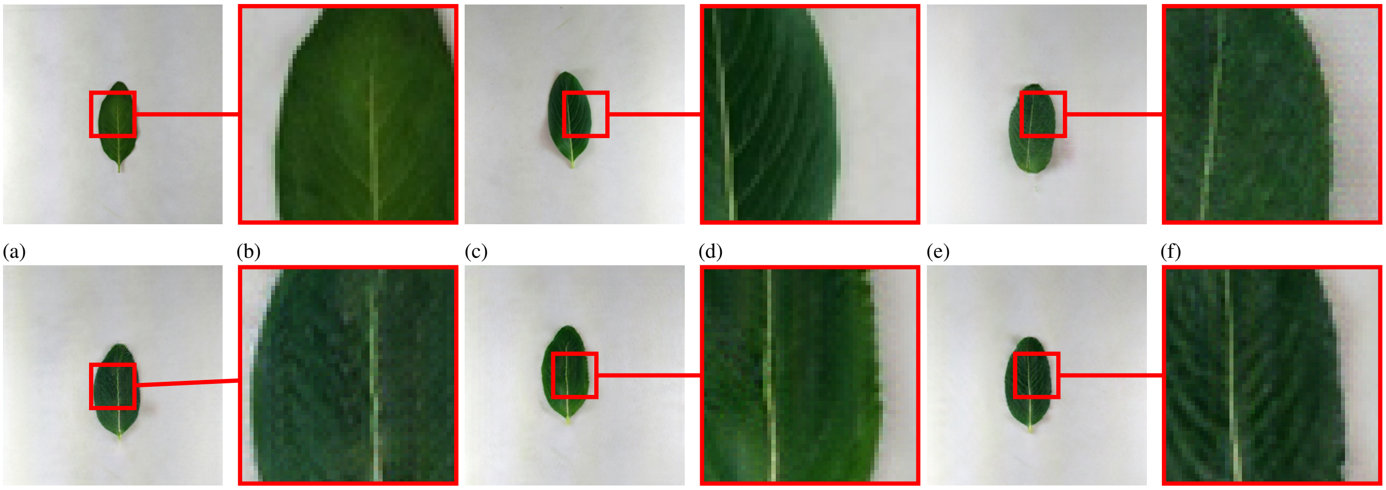
A. Benchmarking with LeavesDryFresh

To test the effects and the feasibility of using an unpaired approach for the senescence reversion task, we decided to use LeavesDryFresh [8] to test our unpaired approach, even though it is a paired dataset. This, in order to compare our unpaired approach with the paired baseline reported previously. For our approach, we change the sampling method and picked random samples from a specific species on each iteration, guaranteeing unpaired sampling. The number of epochs and learning rate were kept the same as [8].

Table I shows the quantitative results using the SSIM metric. The results show that the previously reported paired Pix2Pix [8] had an average higher score than vanilla CycleGAN by 0.0208. Also, we observe that, perceptual loss helped the CycleGAN model to obtain higher scores compared to the vanilla CycleGAN. Additionally, the scores were similar on all instances, with a smallest average distance being of 0.0121 between the original Pix2Pix and the CycleGAN with perceptual loss.

Notice how CycleGAN together with perceptual loss are able to obtain similar results to the Pix2Pix paired baseline [8], regardless of not using additional information of pairs of images of the same specimen. This suggests the feasibility of using unpaired approaches to translate plant images.

In Fig. 3, we show qualitative results from this experiment on the species *C. roseus*. Fig. 3(k) displays a zoom-in into a synthetic image of said species, which shows the level of detail on the veins of the leaf, captured by the our unpaired generative model.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

Fig. 3. Benchmarking with LeavesDryFresh results on *C. roseus*. (a) is the input image (dry leaf). (c) is the ground truth (green leaf). (e) is the generated image by Pix2Pix using an 80/20 split. (g) is the generated image by CycleGAN using an 80/20 split. (i) is the generated image by CycleGAN using a 90/10 split. (k) is the generated image by CycleGAN using a 90/10 split and perception losses. Notice how the usage of perceptual loss with CycleGAN seems to provide better details on venation.

TABLE I
SSIM RESULTS FROM BENCHMARK EXPERIMENT. PL STANDS FOR PERCEPTUAL LOSS.

Run	Split	PL	*C. roseus*	*L. camara*	*P. lutea*	Average
Pix2Pix	80/20	✗	**0.9172**	0.9047	**0.8739**	**0.8986**
CycleGAN	80/20	✗	0.8845	0.9075	0.8416	0.8778
CycleGAN	90/10	✗	0.8995	**0.9173**	0.8343	0.8837
CycleGAN	90/10	✓	0.8892	0.9093	0.861	0.8865

B. Measuring the Effects Perception Loss on Herbarium Unpaired Data

After assessing the feasibility of using unpaired data with the experiment explained in Section IV-A, we test the model in a more difficult task.

As seen in Fig. 1, UnLID shows more complex features such as different amount of herbarium leaves, black squares, different background colors and even herbarium captions inside the image.

We trained the model with the intersection of species of the dataset as explained in Section III and tested different perceptual loss weights, including zero perceptual loss. We use two metrics to asses the quality of the trained models: Fréchet Inception Distance (FID) [29], which is the squared Wasserstein metric between two multidimensional Gaussian distributions, to compare the distributions of the real and hallucinated images, and Inception Score (IS) [30] to measure the quality of the hallucinated images. Lower FID is better and higher IS is better.

The results obtained in this experiment are meant to define a baseline for future unpaired senescence reversion studies. To our knowledge, this has not been attempted before. We study the consequences of using perceptual loss as well as cycle consistency loss to learn a mapping between the herbarium and green leaves domains.

Quantitative results are shown in Table II. We observe that the run with a weight of 0.025 had the lowest FID, 24.75 points lower than without perceptual loss. Similarly, the run with a weight of 0.05 had the highest IS, 0.797 points higher than without perceptual loss, suggesting a favorable presence of perceptual loss.

Fig. 4 and 5 show synthetic leaf images created out of herbarium sheet images and vice-versa. In both instances we observe different samples where hallucinations are able to keep the shape, such as Fig. 4(e) and 5(d), but some others where shape is lost, such as Fig. 4(e) and 5(b).

TABLE II
SCORES OBTAINED WITH DIFFERENT PERCEPTION LOSS WEIGHTS (λ_p) ON THE TEST SET. METRICS WITH (A) REFER TO THE GENERATOR G AND METRICS WITH (B) REFER TO THE GENERATOR F.

λ_{pl}	FID (a)	FID (b)	FID	IS (a)	IS (b)	IS
0.5	198.875	194.875	196.875	**3.928**	3.895	3.91
0.05	199.625	177.75	188.75	**3.928**	**4.258**	**4.094**
0.025	**79.063**	**90.125**	**84.625**	2.875	3.783	3.328
0	124.5	94.188	109.375	3.068	3.523	3.297

C. Impact on Classification Performance

In order to measure the impact of the hallucinated images on classification tasks, we compared the results of a baseline ResNet50 trained only on real data, against another ResNet50 with hallucinated images added to the training set. We used the same testing set based on only real images in both cases. The real images for training consisted on $7,514$ in total, from 195 species. We added a total of $4,868$ hallucinated images to the training set, based on availability from the herbarium dataset.

Fig. 6 shows the distribution of all the 195 species with their respective differences in accuracy between baseline and hallucinated. Negative values in the difference shows an improvement for the classifier compared to the baseline, positive

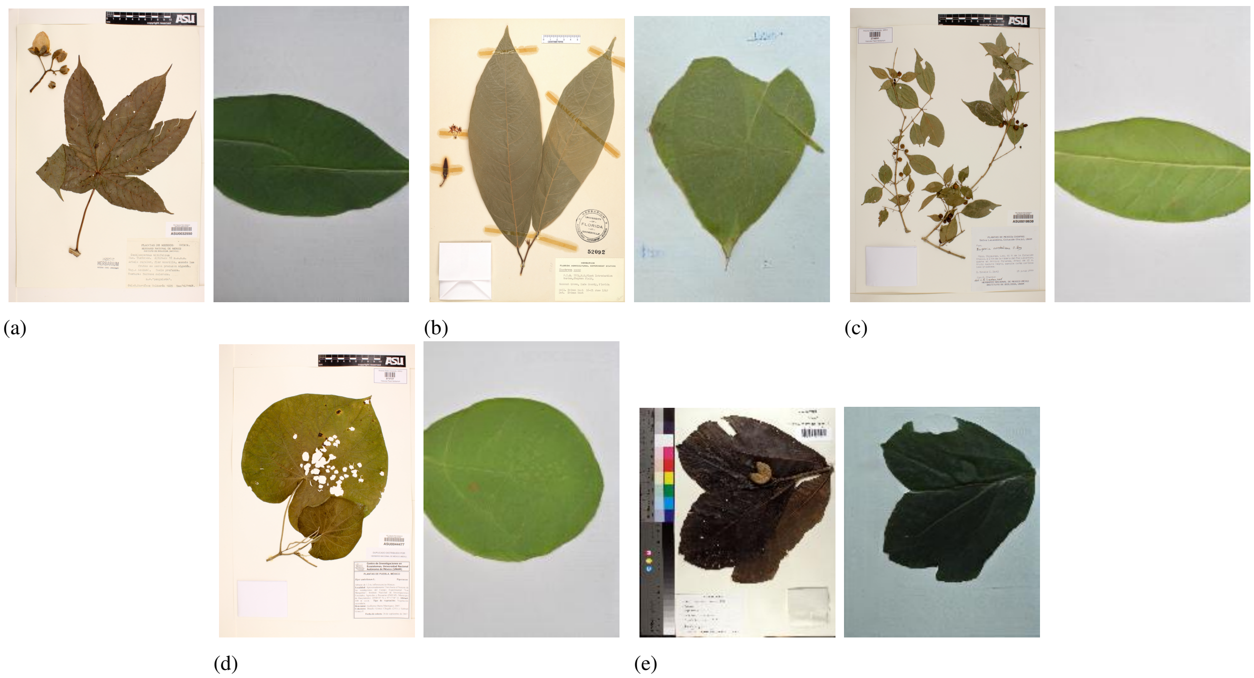


Fig. 4. Leaf images hallucinated from real herbarium sheet images. *Cochlospermum vitifolium* (a), *Theobroma cacao* (b), *Eugenia oerstediana* (c), *Piper umbellatum* (d), *Genipa americana* (e). Notice that for some of the herbarium images, several leaves are mapped to a single leaf in the green domain. Also, some of the holes in the herbarium images are fixed automatically by the senescence reversion process. Finally, herbarium sheet labels are properly erased by the model.

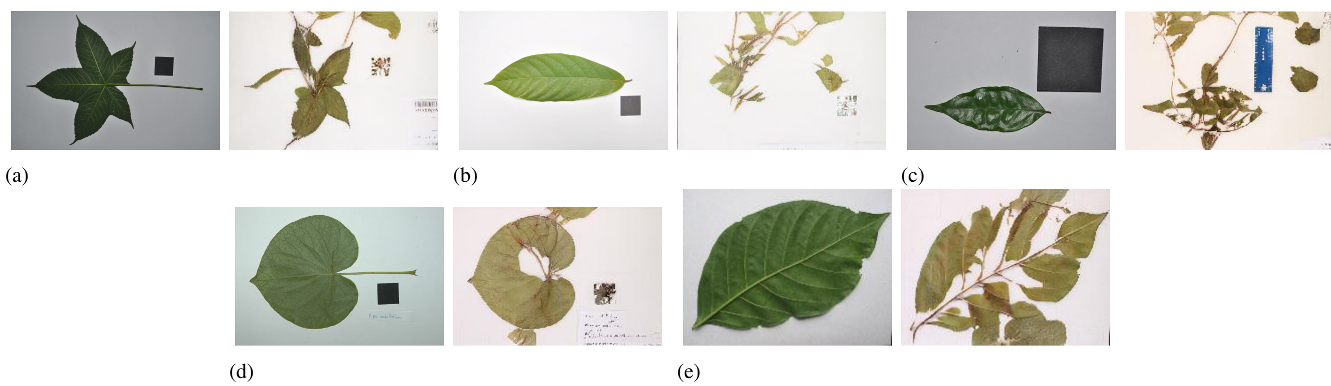


Fig. 5. Herbarium images hallucinated from real leaf images. Each sub-figure depicts the real image to the left, and the hallucinated image to the right. *Cochlospermum vitifolium* (a), *Theobroma cacao* (b), *Eugenia oerstediana* (c), *Piper umbellatum* (d), *Genipa americana* (e).

values show a degradation, and a difference of 0.0 shows no difference. The balanced accuracy of all 195 species decreased by 0.154 compared to the baseline, going from 0.970 to 0.816.

The number of species that improved accuracy is 10, with *Luehea speciosa* getting the biggest accuracy improvement of 0.399. A total of 68 species show a difference of 0.0, which indicate they could be potentially used for data augmentation proposes in absence of real images, as they did not degrade the classification. If we are more permissive on degradation, at lower differences than 0.2, the total amount of useful species is 144, which accounts for the 73.8% of all the dataset. Notice that the amount of hallucinated images (the size of the points) did not influence heavily on degradation of the accuracy.

V. Discussion

Dealing with Unpaired Plant Images: Our experiment with LeavesDryFresh shows the potential of using unpaired data and unpaired models to do senescence reversion. While paired methods gave slightly overall better results, unpaired methods are not far from the paired counterparts. The SSIM obtained by previous, paired state-of-the-art was 0.8986, which is comparable with our unpaired method's SSIM of 0.8865.

Furthermore, the quantitative and qualitative results of using CycleGan in our own, bigger and unpaired UnLID dataset shows the possibility of reverting senescence in real herbarium data. Some problems still remain, such as shape preservation. As noted in Fig. 4(b), 5(b) and 5(c) such problem rises when the shape of the herbarium and leaf images differs greatly, in particular when the herbarium image captures the whole plant. This suggests that working with field images is a harder

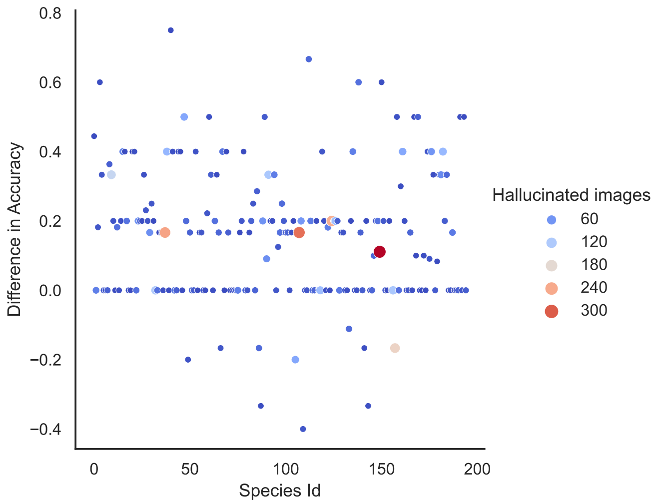


Fig. 6. Distribution of the 195 species based on the accuracy difference between baseline and hallucinated runs. Negative numbers account for accuracy improvement, positive for degradation, and 0.0 for neither. Species with difference of accuracy lower or equal than 0.0 suggest usability for data augmentation proposes.

problem, where additional guidance to the model may be needed.

Shape Problems: In some instances, such as for species *T. cacao* and *E. oerstediana*, the herbarium and the green leaf images may differ considerably. This is due to the nature of the dataset, where the green leaf images contain only a single leaf, compared to their herbarium counterpart which has complete plants. This causes a shape preserving problem, where the generative model attempts to translate the multiple organ plant into a single leaf, causing visual problems such as Fig. 4(a). A shape preserving methodology to avoid drastic content changes between herbarium images and the green leaf images seems to be needed.

Effects of Perceptual Loss: Based on the results in Section IV, we noticed that the inclusion of perception losses helps to close the gap between paired and unpaired models. Quantitative results hint that there is an optimal weight for the perceptual loss term, since using a too much weight tends to collapse the training into two auto-encoders, and too little presents a more notorious issue with preserving features such as shape. This issue was not present when using the LeavesDryFresh as unpaired dataset, since both domains present images of a single leaf in vertical position, therefore obtaining similar results than the paired Pix2Pix [8].

Impact of hallucinated images in classification: The results in Section IV-C show a large number of species with neither degradation nor improvement on classification accuracy. Such difference of 0.0 in accuracy suggests that, for those species, hallucinated images could be interchangeably used for data augmentation in the absence of real images. If we are more permissive and allow a degradation up to 0.2, more than 73.8% of the species become useful. Additionally, the amount of hallucinated images does not seem to affect heavily on degradation. Nevertheless, more exhaustive experiments may be needed to confirm these findings.

VI. Conclusions

Our method for unpaired senescence inversion showed comparable performance with the paired counterpart, even under the disadvantage of lacking image pairs. Furthermore, using our more complex UnLID dataset, our approach also has shown the possibility of reverting senescence in real herbarium data. A large number of species did not show degradation on a classification task's accuracy, suggesting that hallucinated images could be interchangeably used for data augmentation in the absence of real images. Additionally, if the herbarium and the field domains differ greatly regarding the shape of the plant, our approach may produce undesired visual results. This could be be addressed with new ways to control the shape of the generated plant inside the hallucinated image, perhaps with shape-preserving loss functions. Finally, the inclusion of the perception loss helps to close the gap between paired and unpaired models, depending on the strength of the perception itself.

VII. Future Work

The following future directions are interesting to improve upon our senescence reversion work. Trying other models on unpaired image to image translation, such as [19], may yield better translation results. Exploring models for multiple image generation may allow to produce a lot of images from the same specimen or species. Working on contrastive learning in patches may allow to keep the content unchanged, preserving shape [23]. Additionally, experimenting with other perceptual losses besides VGG16 is worth investigating. A user study with taxonomists in a classification task will asses the impact of the hallucinated images on taxonomic quality. Automatic classifiers can also be used for such study. Lastly, exploring field images beyond leaf images is an even harder problem, but worth exploring.

References

[1] M. Yerman, "Herbarium Explorations,"

[2] J. Carranza-Rojas, H. Goeau, P. Bonnet, E. Mata-Montero, and A. Joly, "Going deeper in the automated identification of Herbarium specimens," vol. 17, no. 1, p. 181.

[3] H. Goeau, P. Bonnet, and A. Joly, "Overview of LifeCLEF Plant Identification task 2020," p. 15.

[4] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised Image-to-Image Translation Networks.".

[5] Y. S. Chen, Y. C. Wang, M. H. Kao, and Y. Y. Chuang, "Deep Photo Enhancer: Unpaired Learning for Image Enhancement from Photographs with GANs," pp. 6306–6314.

[6] A. Affouard, H. Goëau, P. Bonnet, J.-C. Lombardo, and A. Joly, "Pl@ntNet app in the era of deep learning," p. 7.

[7] B. R. Hussein, O. A. Malik, W.-H. Ong, and J. W. F. Slik, "Reconstruction of damaged herbarium leaves using deep learning techniques for improving classification accuracy," vol. 61, pp. 101243–101243.

[8] J. Villacis-Llobet, M. Lucio-Troya, M. Calvo-Navarro, S. Calderon-Ramirez, and E. Mata-Montero, "A First Glance into Reversing Senescence on Herbarium Sample Images Through Conditional Generative Adversarial Networks," in *Communications in Computer and Information Science*, vol. 1087 CCIS, pp. 438–447, Springer.

[9] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks."

[10] H. Talebi and P. Milanfar, "NIMA: Neural Image Assessment," vol. 27, no. 8, pp. 3998–4011.

[11] Y. Blau and T. Michaeli, "The Perception-Distortion Tradeoff."

[12] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks."

[13] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks,"
[14] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," vol. 13, no. 4, pp. 600–612.
[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation.".
[16] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition.".
[17] M. Ounsworth, "Anticipatory movement planning for quadrotor visual servoeing."
[18] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least Squares Generative Adversarial Networks.".
[19] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal Unsupervised Image-to-Image Translation.".
[20] M. Amodio and S. Krishnaswamy, "TraVeLGAN: Image-to-image Translation by Transformation Vector Learning.".
[21] T. Wang and Y. Lin, "CycleGAN with Better Cycles," p. 8.
[22] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive Learning for Unpaired Image-to-Image Translation.".
[23] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations.".
[24] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution."
[25] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric."
[26] E. Mata-Montero and J. Carranza-Rojas, "A texture and curvature bimodal leaf recognition model for identification of costa rican plant species," in *2015 Latin American Computing Conference (CLEI)*, pp. 1–12, 2015.
[27] G. Hervé, B. Pierre, and J. Alexis, "Lifeclef 2020 plant identification challenge," 2 2020.
[28] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255.
[29] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.".
[30] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs."

Esteban A. Esquivel-Barboza is a MSc in Computer Science student at the Costa Rican Institute of Technology where he received his B.Eng in Computer Engineering. He interned at Universidade da Coruña, Spain, working in computer vision modules for educational smartphone-based robots. He currently works at Ex^2 Outcoding as a Machine Learning Engineer.

Dr. Jose Carranza-Rojas obtained his Doctorate of Engineering at the Costa Rican Institute of Technology with advisorship from INRIA and CIRAD, France on automatic plant identification. He worked as a Research Scientist at Rakuten Institute of Technology, Japan, on image generative modeling. He currently works as an Associate Professor at Costa Rica Institute of Technology and as a Data & Applied Scientist at Microsoft, in topics related to Natural Language Processing, Computer Vision and Biodiversity Informatics.