

AnoLeaf: Unsupervised Leaf Disease Segmentation via Structurally Robust Generative Inpainting

Swati Bhugra^{+†}, Vinay Kaushik^{+*}, Amit Gupta^{*}, Brejesh Lall[†] and Santanu Chaudhury[§]

^{*}DVIEW, [†]IIT Delhi and [§]IIT Jodhpur

[†]{eez138301,brejesh}@ee.iitd.ac.in, ^{*}{vkaushik,amitg}@dview.ai, [§]schaudhury@gmail.com

Abstract

Plant diseases severely limits agriculture production, necessitating the high-throughput monitoring of plant leaves. Currently, this is formulated as an automatic disease segmentation task addressed via deep learning frameworks. These deep leaning frameworks trained with leaf image data in a supervised paradigm have few limitations, mainly: (1) training datasets are heavily imbalanced towards healthy leaf images, (2) disease region annotation is labour-intensive and (3) due to the heterogeneity of disease symptoms, these frameworks lacks generalisability. In this paper, we reformulate disease segmentation as an anomaly localisation task. Specifically, we introduce a novel unsupervised framework (AnoLeaf) based on an edge-guided inpainting that optimises the learning of contextual attention on only healthy leaf images. The network utilisation on diseased leaf images results in reconstruction of its healthy counterparts, generating an inpainting error. The contextual attention maps reinforce the inpainting error to effectively localise the disease. Thus, AnoLeaf alleviates the acquisition and annotation of rare disease images. Additional experiments on MVTec anomaly detection dataset further demonstrate its generalisability.

1. Introduction

Plant diseases pose a significant threat to agricultural production, highlighting the significance of selecting disease resistant plants [39]. Mostly, the disease symptoms are reflected on plant leaves as changes in its texture and colour. Thus, plant researchers primarily rely on visual leaf monitoring for the selection of disease resistant plants. This continuous monitoring of symptoms at different stages of disease progression is a highly subjective and time-consuming task. With acquisition of leaf image data, the aforementioned task has emerged as an application domain for com-

puter vision algorithms and is currently formulated as an automatic disease segmentation. In this context, Otsu's thresholding and k -means clustering in excess green and excess red image space has been widely adopted [36]. However, these methods struggle to deal with: (1) non-uniform contrast that leads to mild symptoms undetected and (2) curved surfaces and veins results in edges misclassified as disease regions. In addition, the priors utilised for disease symptom characterisation vary among disease classes, necessitating its repetitive formulation for accurate disease segmentation. Figure 1 highlights the heterogeneity of disease symptoms.

In contrast to the traditional methods, supervised deep learning frameworks have also been proposed for disease segmentation. For example, Dechant et al. [11], used different convolutional neural networks (CNNs) combinations to generate a heat map of lesion probabilities. These probabilities were obtained from the classification of small patches using sliding window over maize leaf images. In another study, Cruz et al. [10] proposed an improved LeNet model for disease segmentation. It is to be noted that the accuracy of these methods heavily rely on large training images with pixel-level annotations. However, limited annotated leaf datasets are available with high imbalance towards the healthy leaf class [35]. In addition, widely used ImageNet weights are obtained with image data that semantically differs from diseased leaf images [17].

As previously mentioned, pixel-level annotations of disease regions is a time consuming task. Thus, most of the leaf disease datasets are weakly labeled i.e. disease class, indicating the type of disease at an image level. Consequently, deep learning based methods have primarily focused on disease classification with limited studies on disease segmentation [39]. Disease segmentation permits disease forecasting and understanding relationships between disease symptoms and the environment [11]. This accurate assessment of disease intensity is not offered in a disease classification paradigm. Thus, disease segmentation is an important and non-trivial task. Few methods have been proposed that train deep learning networks in a weakly supervised paradigm to generate an explainable map show-

Thanks to Hwy Haul: <https://www.hwyhaul.com/>

⁺Equal Contribution

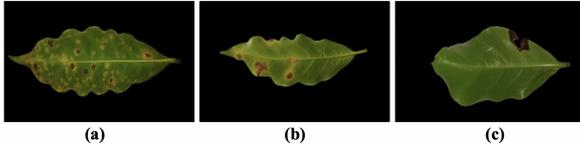


Figure 1. Sample leaf images (Coffee dataset) with contrast disease symptoms.

ing pixels that determined the classification results [26, 27] and facilitating the localisation of disease spots. Authors in [31, 41] used AlexNet and GoogLeNet architectures for disease classification and visualisation activation in the first layers to indicate the disease regions. However, with a large number of feature maps it still remains challenging to select a specialised activation that precisely localise these symptoms.

The disease progression can be considered as a degradation of healthy (normal) parts of plant leaves [39]. Thus, we reformulate leaf disease segmentation as an anomaly localisation task, with its training paradigm dependent only on healthy (normal) leaf images. This reformulation alleviates the task of generating data based on acquisition of rare disease images and corresponding pixel-level annotations. In this context, we present *AnoLeaf*, a deep inpainting based framework that learns an optimal reconstruction of healthy leaf images. The current state-of-the-art inpainting networks [48] utilize edge information, model self-attention, incorporate texture, structural and semantic cues [18, 34, 47, 52] in a coarse-to-fine set-up. In contrast, we propose to utilize structural information by developing an adversarial edge prediction framework and integrate it with the coarse inpainted image via contextual attention refinement framework (colour, shape). The network utilisation on diseased leaf images results in the reconstruction of only its healthy counterparts. Based on this reconstructed image, a widely adopted strategy is to compute its difference with the input diseased image using metrics such as Euclidean distance and structural similarity index (SSIM) [24]. However, each pixel in an inpainted region is assigned the same metric value, resulting in a coarse disease segmentation. Considering the fact, that an inpainting network has learnt to restore the healthy counterparts, we propose to utilise an attention map that highlights the regions not utilised for inpainting. This novel strategy permits precise disease localisation with robustness to disease symptoms heterogeneity (shown in Figure 1). In summary, the paper has the following contributions:

- To the best of our knowledge, we are the first to reformulate plant disease segmentation as an unsupervised anomaly localisation task.
- We propose a novel two stage framework (*AnoLeaf*) based on an edge-guided inpainting network for im-

proved inpainting. To further refine the localisation of diseased regions, a novel strategy of utilising learned contextual attention maps is implemented.

- We also conduct additional experiments on the MVTEC anomaly detection dataset (MVTEC AD) [4] that further shows the generalisability of the proposed framework.

The rest of the paper is organized as follows: Section 2 gives an overview of the relevant disease segmentation literature, Section 3 presents the methodology, experimental results and analysis are included in Section 4 and Section 5 concludes the paper.

2. Related Work

The relevant literature on leaf disease segmentation can be broadly divided into three categories: (i) traditional methods (ii) supervised convolutional neural networks (CNNs) based methods and (iii) CNNs visualisation based methods. The following subsections provides an overview of the aforementioned categories.

2.1. Traditional methods

Traditional methods extract hand-crafted features such as colour, texture etc. to characterise disease regions (also termed as lesion spots) based on prior knowledge of plant diseases. For example, authors in [28] utilised an interactive region growing method based on three channel colour components (1) Excess Red Index (ExR), (2) H component of HSV colour space and (3) b component of Lab colour space. Whereas, Xiong et al. [46] proposed to use Grab-Cut algorithm for automatic disease segmentation. Several other researchers [1, 2, 22, 28, 51] employed Otsu's thresholding on Lab colour image space, HSV colour space and Excess Green space (ExG). In these methods, the accuracy of disease segmentation is highly dependent on a threshold value. Authors in [29] investigated these methods and reported misclassification of illuminated regions as lesion spots. k -means clustering of Lab colour space was investigated in [40, 50] whereas, Bai et al. [3] proposed an improved fuzzy c -means based clustering technique for disease segmentation. As shown in Figure 1, the disease regions (clusters) vary depending on the disease type and its degree of progression. However, dynamically determining these clusters is time-consuming [1, 42].

In contrast to these traditional methods, the proposed framework (*AnoLeaf*) do not rely on disease priors i.e disease symptoms (colour, texture etc.) or disease clusters.

2.2. Supervised CNNs based methods

Supervised CNNs based methods primarily utilise CNN architecture to automatically learn features for disease region characterisation. For example, Ha et al. [19] employed

a hybrid model based on VGG-CNN and k -means clustering method to segment disease regions, whereas Wang et al. [53] constructed a regional disease detection network (RD-net) with traditional VGG16 model. Authors in [13, 25] presented UNet based semantic segmentation model and [14] designed SegNet based encoder-decoder architecture for disease region segmentation. Another study [16] compared Feature Pyramid Network (FPN), UNet and DeepLabV3+ with Xception backbone for pixel-level predictions of disease regions.

These methods were trained on diseased leaf datasets with pixel-level annotations exhibiting specific disease symptoms. Due to the high inter-class variance among disease classes, these methods may not be generalisable and require retraining. In contrast, *AnoLeaf* do not rely on pixel-level annotations of diseased leaf datasets.

2.3. CNNs visualisation methods

Ghosal et al. [15] presented a deep CNN framework for disease classification and its top-K high-resolution feature maps to isolate the disease symptoms. Several other studies [27, 31, 41] also utilised visualisation of features in each block. These methods primarily transform internal activations to visualise lesion spots. In contrast, authors in [6] proposed to use saliency map for visualising disease regions. Occlusion method investigated in [7], is another method to localise disease symptoms. The primary idea was to introduce occlusions to a portion of the image and investigate CNN's sensitivity to these occlusions with respect to its output. However, occlusion approach is sensitive to shape, size and the displacement stride of occlusion regions. These factors limits its application on leaf images with several disease regions. Also, in this scenario occlusion of one disease region does not affect the network decision resulting in undersegmentation of disease regions. Another study [32], investigated various visualisation methods namely smooth-Grad, boot back-propagation, depth Taylor decomposition, integration gradient layered associated transmission and gradient time input. The authors reported that these visualisation methods detected the lesion spots for some of the correctly classified disease categories. In this respect, authors [8] concluded that the heat maps based on these methods are noisy to interpret. To overcome this limitation, a Teacher/Student architecture was presented that handles the deconstruction and reconstruction of the image to remove noise and extract salient disease features.

Previously discussed studies either rely on disease priors or image/pixel-level disease annotations. Due to the heterogeneity of disease symptoms, unsupervised deep learning is worthy of being exploratory in the application of plant disease segmentation. This permits disease localisation with no domain knowledge. In this context, authors [21] re-

cently proposed to utilise a conditional adversarial network for colour reconstruction of healthy leaf images corresponding to its grayscale image. The inaccurate reconstruction of colours for disease affected regions was subsequently used for disease classification i.e. healthy and disease. In contrast to this study, *AnoLeaf* yield precise disease localisation with no assumption for reconstruction of normal (healthy) leaf regions. Additional experiments on MVTEC anomaly detection dataset demonstrates this generalisability.

3. Methodology

Plant disease segmentation is a non-trivial task with challenges such as: (1) variations in visual symptoms with respect to shape and disease clusters, (2) multiple simultaneous disease manifestation on the same leaf, (3) similarities in the visual symptoms of different diseases and (4) dependence on domain experts for acquisition of annotated plant disease data. Thus, we model leaf disease segmentation as an anomaly localisation task. Specifically, anomaly localisation approaches are generally dependent on only modelling normal (healthy) images [5, 9, 33, 45]. The deviation of learnt features is used to detect anomalies (disease) during inference, eliminating the dependency on large pixel-level annotated datasets.

Anomaly localisation algorithms can be primarily categorised into (i) discriminative and (ii) reconstruction based approaches. Since the accuracy of discriminative anomaly localization approaches [5, 33] rely on the spatial size of the regions utilised for modeling normality, these approaches are not suitable for disease segmentation task. Specifically, large spatial size miss small disease regions and small spatial size is sensitive to non-disease local changes.

In this respect, reconstruction based anomaly localisation algorithms provides a suitable solution for this task. For example, variational autoencoder (VAE) [12] was utilised to obtain the nearest normal image by iteratively updating the VAE inputs via gradient descent. In another study, *AnoGAN* proposed in [37] was trained only on normal images and defects were localised by computing the differences between test image and its nearest normal generated image. These methods depend on an iterative procedure to find the nearest normal image, thus decreasing its computational efficiency. In a recent study [24], inpainting network was trained on normal images, this results in restoring defect regions into its normal counterpart. Similar to the previously mentioned algorithms the defect is localised based on the difference between input test and inpainted image. It is to be noted that the defect localisation is dependent on accurate inpainting output. In this respect, authors in another study [48] conducted extensive experiments that highlighted the limitation of inpainting network to reproduce sharp edges and complex textures for anomaly localisation. To overcome this limitation, we propose a novel

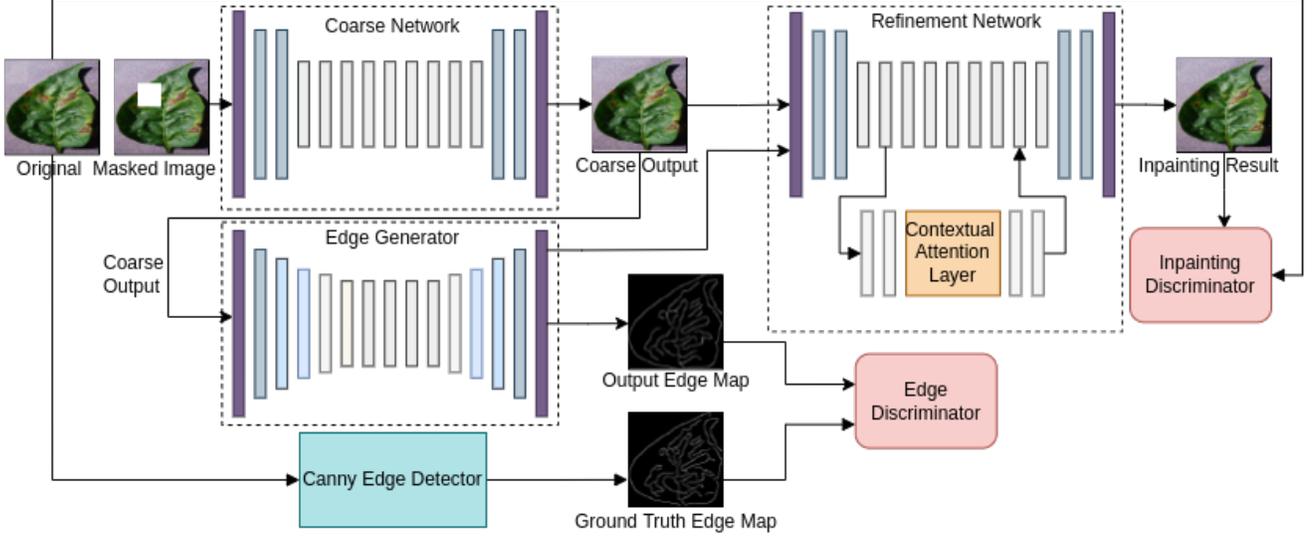


Figure 2. Workflow of the proposed inpainting network.

inpainting framework motivated by the current state-of-the-art inpainting network [47].

The proposed framework (*AnoLeaf*) comprises of two modules (a) an inpainting network that learns to restore only normal (defect/disease free) images and (b) a novel strategy to utilise contextual attention maps for precise defect/disease localisation. We elucidate these modules in the following subsections.

3.1. Generative Inpainting Network

The inpainting network trained on synthesising the normal (healthy) regions of the image is utilised for localising the anomaly (disease) regions. This is accomplished via computing the difference between the input test image and the corresponding restored image. Thus, the accuracy of this localisation is highly dependent on the inpainted output. However, current state-of-the-art inpainting networks fails to reconstruct details and structures. This is because, they employ a coarse to fine refinement architecture [47] with refinement stage dependent on the inpainted output of the coarse network. We propose to facilitate the refinement stage with an edge map in conjunction with the already available coarse information. Figure 2 shows the overview of the proposed inpainting network consisting of three components.

3.1.1 Component: Coarse Network

The first module is a dilated convolutional network (A_1) that is trained to produce a coarse inpainted output (I_{coarse}) of an input image ($I_{original}$) with binary mask image (I_{mask}), white pixels indicate missing regions. The module follows the same architecture employed in [47] with ℓ_1

reconstruction loss.

3.1.2 Component: Edge GAN

The second module predicts the fine grain edge image (O_{edge}) corresponding to the coarse input (I_{coarse}). The architecture of this component comprises of an encoder (A_2) with downsampling followed by eight residual blocks (dilated convolutions) and a decoder for upsampling [34].

$$O_{edge} = A_2(I_{coarse}, I_{edge}) \quad (1)$$

We use I_{edge} and O_{edge} conditioned on I_{coarse} as inputs of the discriminator that predicts whether or not an edge map is real. The edge network is trained with an objective based on an adversarial loss and feature-matching loss [43].

$$\min_{A_2} \max_{B_1} \mathcal{L}_{A_2} = \min_{A_2} \left(\lambda_{adv} \max_{B_1} (\mathcal{L}_{adv}) + \lambda_{FM} \mathcal{L}_{FM} \right) \quad (2)$$

where, λ_{adv} and λ_{FM} are regularization parameters. The adversarial loss is defined as

$$\mathcal{L}_{adv} = \mathbb{E}_{(I_{edge}, I_{coarse})} [\log B_1(I_{edge}, I_{coarse})] + \mathbb{E}_{I_{coarse}} \log [1 - B_1(O_{edge}, I_{coarse})] \quad (3)$$

The feature-matching loss \mathcal{L}_{FM} compares the activation maps in the intermediate layers of the discriminator. This stabilizes the training process by forcing the generator to produce results with representations that are similar to real images. The feature matching loss \mathcal{L}_{FM} is defined as

$$\mathcal{L}_{FM} = \mathbb{E} \left[\sum_{i=1}^L \frac{1}{N_i} \left\| D_1^{(i)}(C_{gt}) - D_1^{(i)}(C_{pred}) \right\|_1 \right] \quad (4)$$

where L is the final convolution layer of the discriminator, N_i is the number of elements in the i 'th activation layer, and $D_1^{(i)}$ is the activation in the i 'th layer of the discriminator. Spectral normalisation [30] scales down weight matrices with their largest singular values to limit the network's Lipschitz constant to one, stabilising training. Edge GAN module is trained on edge images trained with ground truth edge image (I_{edge}) computed from $I_{original}$ using canny edge detection algorithm [34].

3.1.3 Component: Refinement Network

This generated edge image (O_{edge}) and the coarse image (I_{coarse}) is subsequently concatenated ($I_{combined}$) and fed as an input to the refinement network (A_3) to predict the output inpainted image (I_{out}). The combined image ($I_{combined}$) provides complementary information to the third component, specifically fine structural details via I_{edge} and approximate details (colour, shape etc.) via (I_{coarse}). This permits an improved restoration of ($I_{original}$) in contrast to only utilising I_{coarse} . A_3 consists of two parallel encoder networks i.e. convolutional network and contextual attention network [47] integrated into a single auto-encoder and its discriminator (B_2) learns to distinguish between I_{out} and $I_{original}$. Both coarse network (first component) and refinement network (third component) are trained with spatially discounted reconstruction loss and WGAN-GP loss [47]. We follow the same configuration as in [47] for the refinement network.

At the inference step, the trained inpainting module is utilised for input test and mask image pair to restore its missing regions. Multiple mask images with patches of fixed size are generated for the test image to achieve complete image traversal. The difference between inpainted and input test image (I_{diff}) for each mask image is computed to obtain a defect/disease map. It is to be noted that metric utilised for I_{diff} computation should display large variance between normal and abnormal inpainted regions. It was empirically observed that the Euclidean distance, structural similarity index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) failed to provide a robust threshold for distinguishing the normal and anomaly regions. Thus, perceptual similarity metric [49] was employed to produce coarse localisation of the defect/disease regions.

3.2. Attention Map Module

Disease/defect regions exhibit size variance and each pixel in an inpainted patch is assigned the same metric value, this provides a coarse disease/defect map. A similar methodology was presented in [24] that employed super-pixels regions from normal images to train the inpainting network and an anomaly heat map was extracted based on SSIM.

Algorithm 1: Anomaly Disease Segmentation Pseudo Code.

Result: Disease segmentation mask $I_{disease}$
Given: Input image I , mask images (patch size N) ;
Queue $q = \{I_m \mid I_m \in \text{mask images}\}$;
Disease segmentation = \emptyset ;
while $q.size() > 0$ **do**
 $I_m = q.pop()$;
 $I_{test} = I \cdot I_m$;
 $I_{inpainted}, I_{attention} = \text{Inpaint}(I_{test}, I_m)$;
 $I_{error} = \text{perceptual}(I, I_{inpainted})$;
end
 $I_{error\ all} = \text{binarise}(\sum I_{error})$;
 $I_{attention\ all} = \text{binarise}(\sum I_{attention})$;
 $I_{not\ attention} = \sim I_{attention\ all}$;
 $I_{disease} = I_{not\ attention} \cdot I_{error\ all}$;

Function $\text{Inpaint}(I_{test}, I_m)$:
Given: Trained model, f_θ ;
 $I_{inpainted} = f_\theta(I_{test})$;
foreach $(x_f, y_f) \in I_m$ **do**
 foreach $(x_b, y_b) \notin (I_m)$ **do**
 $\text{attention}_{x_b, y_b} = f_{x_f, y_f} \cdot b_{x_b, y_b}$;
 end
 $x, y = \min(\text{attention}_{x_b, y_b})$;
 $I_{pixel\ attention}(x, y) = 255$;
end
 $I_{attention} = \sum I_{pixel\ attention}$;
return $I_{inpainted}, I_{attention}$;
End Function

\triangleright where, f_{x_f, y_f} is the feature vector patch $[3 \times 3]$ of $I_{inpainted}$ at the contextual layer centered at (x_f, y_f) and b_{x_b, y_b} is the feature vector patch at the contextual layer at (x_b, y_b) . The \cdot measures the similarity between the aforementioned patches. The $\text{perceptual}()$ function measures the perceptual metric. The $\text{binarise}()$ function refers to Otsu's threshold.

Figure 3. Pseudo Code of the proposed framework *AnoLeaf*.

For precise disease/defect localisation, we propose to generate an activation map from the contextual layer [47]. The activation maps represent long-term correlations between distant contextual information and the missing regions based on high-dimensional features. Specifically, it highlights the related spatial regions for each missing region that is subsequently utilised for inpainting. As previously discussed, the first component generates the blurry normal counterparts for each missing patch. Based on this input, the contextual layer does not attend to abnormal (disease/defect) image content. This leads to overlooked abnormal regions in the corresponding activation maps. Figure 4(b) shows the overlay of the activation map indicating the regions employed for inpainting and over-looking disease regions. Thus, the computed map provides additional pixel-level information of the normal region. This information is subsequently integrated with (I_{diff}) further facilitating the precise disease/defect localisation.

Method	Precision	Recall	Jaccard Index	F-value
Excess Green Threshold [44]	0.3875	0.7724	0.3269	0.4519
Student Teacher Network [8]	0.1131	0.8066	0.1099	0.1896
AnoGAN (Integrated) [37]	0.0818	0.2851	0.0679	0.1169
AnoGAN (CIEDE 2000) [37]	0.0846	0.0178	0.0147	0.0270
AnoGAN (Perceptual) [37]	0.5214	0.6443	0.4092	0.5359
RIAD (Integrated) [24]	0.1398	0.5605	0.1169	0.1976
RIAD (CIEDE 2000) [24]	0.1801	0.1040	0.0617	0.1064
RIAD (Perceptual) [24]	0.4746	0.4730	0.3031	0.4321
Pix2Pix GAN (CIEDE 2000) [21]	0.8655	0.5910	0.5281	0.6592
Pix2Pix GAN (Perceptual) [21]	0.9158	0.4319	0.4041	0.5354
AnoLeaf (Ours)	0.9644	0.8862	0.7730	0.9236

Table 1. Quantitative leaf disease segmentation results on Coffee dataset via traditional segmentation method (shown in dark gray), weakly supervised segmentation method (shown in light gray), unsupervised segmentation methods (shown in blue) and *AnoLeaf*.

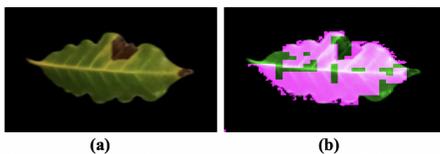


Figure 4. Attention map (b) based on sample Coffee leaf image.

4. Experiments and Results

In this section, we show the results and evaluation of the proposed framework on publicly available datasets:

- Coffee dataset [13]: The dataset consists of top-down views of healthy/disease leaves with a total of 372 images.
- MVTec Abnormal Detection dataset [4]: MVTec AD is a benchmark for industrial inspection anomaly detection methods. The dataset consists of 5000 images in 15 object and texture categories. Each category includes defect-free training images and test images with/without defects.
- PlantVillage dataset [20]: The dataset contains 54,303 images of healthy/disease leaves categorised based on plant species and diseases.

Coffee and MVTec AD datasets also comprises of annotated images, where a binary label is assigned to disease/defect regions differentiating it from healthy/normal pixels. Thus, for quantitative analysis the aforementioned datasets are utilised. In contrast, PlantVillage dataset was employed only for qualitative results (shown in Figure 2), since no pixel-level annotations are available.

Module	Precision	Recall	Jaccard Index	F-value
Contextual Attention	0.3875	0.7724	0.3269	0.4519
Edge GAN	0.9186	0.5076	0.4940	0.5813
AnoLeaf (Ours)	0.9644	0.8862	0.7730	0.9236

Table 2. Ablation Study

4.1. Training Details

We elucidate the training details of three components in the proposed inpainting framework (see section 3.1). During training, we resize the image to 256×256 with fixed rectangle patches of size 64×64 (shown in Figure 8 (b)). The learning rate of Adam optimiser was set to 0.0001, β_1 to 0.5 and β_2 to 0.9 with a batch size of 12. We first train the coarse network and then use the results of the trained coarse model as an input to EdgeGAN. We choose $\lambda_{adv} = 1$ and $\lambda_{FM} = 10$ as hyperparameters for training EdgeGAN. The refinement network is trained as the final step to obtain an inpainted image. The discriminator to generator learning ratio was set to 0.1 for EdgeGAN and 0.5 for the refinement network. It is important to note that although the edge network takes a coarse image as input, it was able to predict a fine-grained edge map that was very similar to the edge map computed using its ground truth image. The proposed inpainting framework was trained for 200k and 300k iterations on coffee dataset and MVTec AD dataset respectively. At the inference stage, the input test image is resized as mentioned previously and image traversal is achieved with fixed patch size 64×64 . The inpainted output for each patch is utilised to generate the corresponding reconstructed image. Pseudo code for disease segmentation based on the reconstructed image is shown in Figure 3. To evaluate the performance of the proposed framework, following metrics are utilised:

- Precision [37]: computes the ratio of true positives (TP) and all the positive predictions (TP+FP).
- Recall [37]: computes the ratio of true positive predictions (TP) and all the actual positives (TP+FN).
- Jaccard Index [37]: computes the percent overlap between the ground truth mask (GT) and predicted mask.
- F1-value [37]: measures the weighted average of precision and recall.

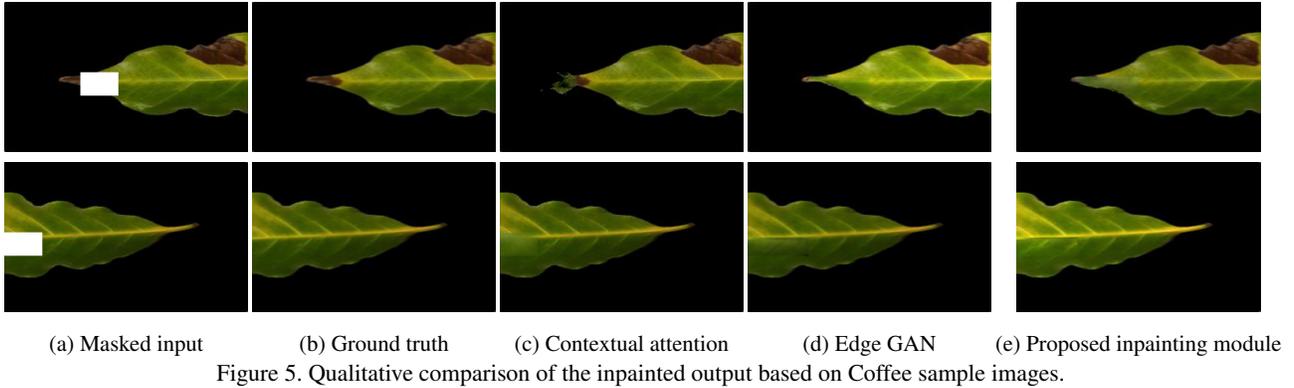


Figure 5. Qualitative comparison of the inpainted output based on Coffee sample images.

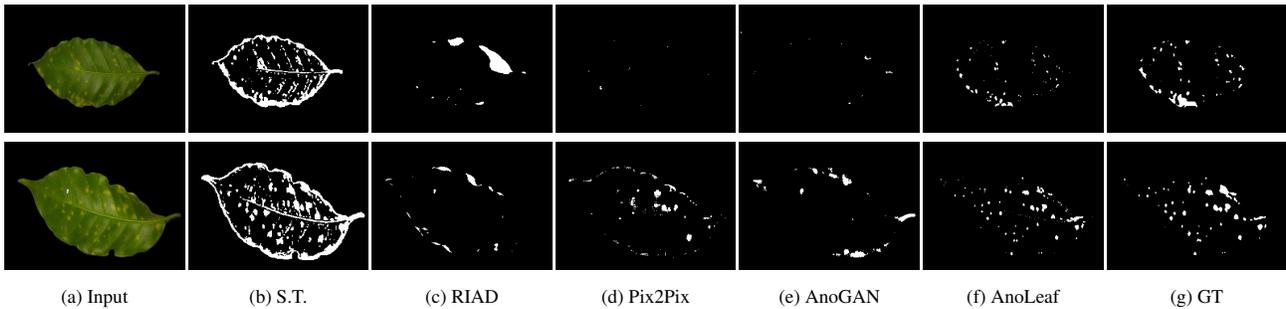


Figure 6. Qualitative comparison of the disease segmentation results based on Coffee sample images, here S.T. denotes student-teacher method and GT denotes the ground truth.

For disease (anomaly) segmentation, TP refers to correctly classified disease (anomaly) pixels, FP refers to healthy (normal) pixels classified as disease (anomaly) pixels, FN refers to disease (anomaly) pixels classified as healthy (normal) pixels. The best value of these metrics is 1.

4.2. Disease segmentation baselines

As evident from the related work (section 2), the problem of disease segmentation has been addressed via (a) traditional methods, (b) supervised CNNs based methods and (c) CNNs visualisation methods (weakly supervised). For a fair comparison, the baseline is formulated with widely adopted traditional excess green method [44] and student-teacher based visualisation method [8]. Since, we train *Anoleaf* in a reconstruction based anomaly paradigm, state-of-the-art anomaly localisation methods such as AnoGAN [37] and RIAD [24] have also been included as baselines. It is to be noted that for disease localisation, the difference between the reconstructed and input image is computed differently in AnoGAN and RIAD. Since, this step is critical for accurate disease localisation, additional baselines based on (1) integrated metric used in AnoGAN and RIAD, (2) CIEDE 2000 colour difference anomaly score [38] and (3) state-of-the-art perceptual metric score [49] have been included. We also extended an unsupervised disease classification framework based on colour reconstructibility of pix2pix GAN [21] for disease localisation. CIEDE 2000 and perceptual

metric was used for computing the difference between the input and reconstructed test image.

These baselines permits an exhaustive comparison of *Anoleaf* with leaf disease segmentation methods and reconstruction based anomaly localisation methods using various metrics. We also perform an ablation that highlights the contribution of each component (section 3.1) in the proposed inpainting framework.

4.3. Disease segmentation results

Figure 5(e) show the output images on two sample leaf images (Figure 5(a)) based on the proposed inpainting framework. The utilisation of inpainting network without edge information leads to blurring and loss of vein structures (Figure 5(c)). EdgeGAN [34] improved the inpainting of the veins but still missed to reconstruct it completely (Figure 5(d)) as compared to the proposed inpainting framework. This is because the predicted coarse image with fine-grained edge map robustly predict a structurally consistent inpainted image. Table 2 presents the quantitative segmentation results based on this ablation.

We also present qualitative comparison of *Anoleaf* based on the reconstructed leaf image with the anomaly frameworks in Figure 7. It is evident that AnoGAN [37] results in smooth leaf textures and fails to replicate the leaf shape. This results in leaf contour being falsely misclassified as disease regions (shown in Figure 6(e)). RIAD

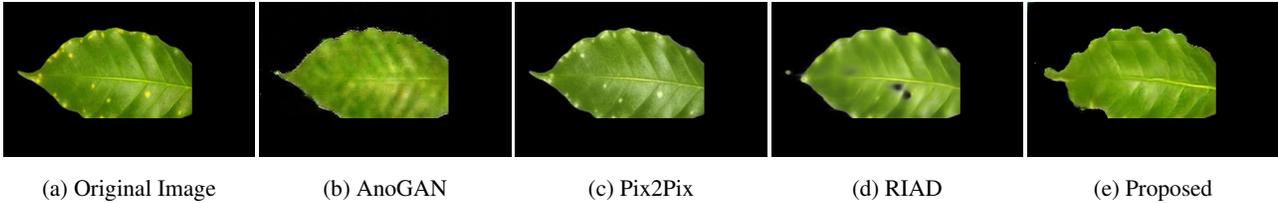


Figure 7. Qualitative comparison of the reconstruction results based on Coffee sample image.

[24] suffers from shape reconstruction errors and random noise (shown in Figure 7 (d)). It is to be noted that although pix2pix [21] shows better reconstruction compared to AnoGAN [37] and RIAD [24], this method rely on an assumption that disease progression only leads to changes in leaf colour and this simplistic assumption may prevent its applicability on diseases with textural and shape changes. In contrast, *AnoLeaf* do not rely on disease characterisation priors and was also be utilised for anomaly localisation on MVTec AD dataset (see next subsection).

The quantitative comparison of *AnoLeaf* with the baselines (subsection 4.2) is presented in Table 1 and qualitative results are shown in Figure 6. It is evident in Figure 6(b) that the weakly supervised method results in misclassification of veins as disease regions (lowest precision). The experiments also revealed that the performance of perceptual similarity metric is consistently superior to the integrated and CIEDE 2000 metrics for reconstruction based anomaly localisation methods.

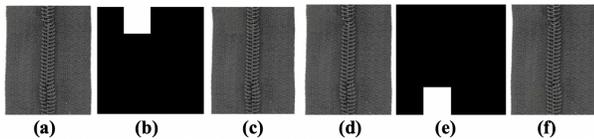


Figure 8. Inpainting results (c) & (f) on MVTec AD sample images (a) & (d) from proposed inpainting framework.

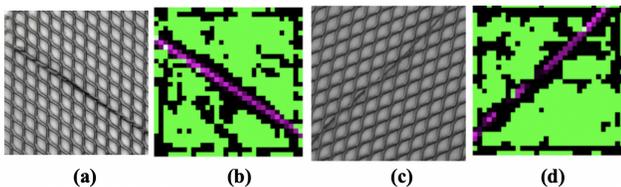


Figure 9. Attention maps (b) & (d) on MVTec AD sample images (a) & (c) respectively.

4.4. MVTec Anomaly Results

Figure 8 shows the output image from the proposed inpainting network with missing patches at different spatial locations. It can be observed that the missing patch on the defected area (Figure 8(e)) results in the reconstruction of a defect-free zipper adhering to texture and shape details. Table 3 demonstrates the effectiveness of the proposed approach compared to the existing anomaly localisation methods i.e. AnoGAN [37] and Autoencoders (AEs) with L2 and SSIM [23]. The baseline framework corre-

Method	Precision	Recall
Autoencoder L_2 [23]	0.65	0.63
Autoencoder SSIM [23]	0.88	0.89
AnoGAN [37]	0.92	0.91
Baseline Framework	0.90	0.89
Proposed Framework	0.93	0.92

Table 3. Quantitative comparison of anomaly localization results on MVTec AD dataset.

sponds to only utilising inpainting module for disease localisation (w/o contextual attention strategy). It is to be noted that AnoGAN performs iterative backpropagation to estimate the latent space at the inference stage, in contrast the proposed framework is computationally efficient. As previously mentioned, the contextual attention maps indicates regions employed for inpainting. Figure 9(b) and 9(d) empirically shows the overlay of the activation map (shown in green) and ground truth anomaly region (shown in pink). This yield missing regions in the generated dense activation maps corresponding to the disease regions. In this context, Figure 10 highlights the efficacy of the proposed framework to segment small anomaly regions (shown in green).

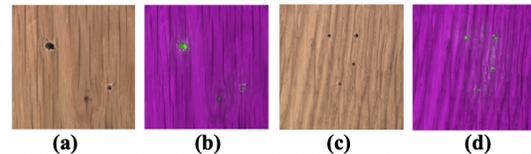


Figure 10. Anomaly segmentation results (b) & (d) on MVTec AD sample images (a) & (c) using the proposed framework.

5. Conclusion

We propose a novel unsupervised framework for leaf disease segmentation. *AnoLeaf* trained in an anomaly localisation paradigm, integrates structurally consistent reconstructed image with contextual attention map for precise segmentation. With an improved segmentation, subtle differences among the plant species may be revealed. Since, no priors regarding disease characterisation was utilised in *AnoLeaf*, additional experiments on MVTec AD dataset demonstrates its generalisability.

References

- [1] Aliyu Muhammad Abdu, Musa Mohd Mokji, and Usman Ullah Sheikh. Automatic vegetable disease identification ap-

- proach using individual lesion features. *Computers and Electronics in Agriculture*, 176:105660, 2020.
- [2] XD Bai, Zhi-Guo Cao, Y Wang, ZH Yu, XF Zhang, and CN Li. Crop segmentation from images by morphology modeling in the $cie\ l^* a^* b^*$ color space. *Computers and electronics in agriculture*, 99:21–34, 2013.
 - [3] Xuebing Bai, Xinxing Li, Zetian Fu, Xiongjie Lv, and Lingxian Zhang. A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images. *Computers and Electronics in Agriculture*, 136:157–165, 2017.
 - [4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019.
 - [5] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4183–4192, 2020.
 - [6] Mohammed Brahimi, Marko Arsenovic, Sohaib Laraba, Srdjan Sladojevic, Kamel Boukhalfa, and Abdelouhab Moussaoui. Deep learning for plant diseases: detection and saliency map visualisation. In *Human and machine learning*, pages 93–117. Springer, 2018.
 - [7] Mohammed Brahimi, Kamel Boukhalfa, and Abdelouhab Moussaoui. Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence*, 31(4):299–315, 2017.
 - [8] Mohammed Brahimi, Saïd Mahmoudi, Kamel Boukhalfa, and Abdelouhab Moussaoui. Deep interpretable architecture for plant diseases classification. In *2019 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages 111–116. IEEE, 2019.
 - [9] Diego Carrera, Giacomo Boracchi, Alessandro Foi, and Brendt Wohlberg. Scale-invariant anomaly detection with multiscale group-sparse models. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3892–3896. IEEE, 2016.
 - [10] Albert C Cruz, Andrea Luvisi, Luigi De Bellis, and Yiannis Ampatzidis. Vision-based plant disease detection system using transfer and deep learning. In *2017 asabe annual international meeting*, page 1. American Society of Agricultural and Biological Engineers, 2017.
 - [11] Chad DeChant, Tyr Wiesner-Hanks, Siyuan Chen, Ethan L Stewart, Jason Yosinski, Michael A Gore, Rebecca J Nelson, and Hod Lipson. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology*, 107(11):1426–1432, 2017.
 - [12] David Dehaene, Oriel Frigo, Sébastien Combexelle, and Pierre Eline. Iterative energy-based projection on a normal data manifold for anomaly localization. *arXiv preprint arXiv:2002.03734*, 2020.
 - [13] José GM Esgario, Renato A Krohling, and José A Ventura. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture*, 169:105162, 2020.
 - [14] Junfeng Gao, Jesper Cairo Westergaard, Ea Høegh Riis Sundmark, Merethe Bagge, Erland Liljeroth, and Erik Alexandersson. Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning. *Knowledge-Based Systems*, 214:106723, 2021.
 - [15] Sambuddha Ghosal, David Blystone, Asheesh K Singh, Baskar Ganapathysubramanian, Arti Singh, and Soumik Sarkar. An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences*, 115(18):4613–4618, 2018.
 - [16] Juliano P Goncalves, Francisco AC Pinto, Daniel M Queiroz, Flora MM Villar, Jayme GA Barbedo, and Emerson M Del Ponte. Deep learning architectures for semantic segmentation and automatic estimation of severity of foliar symptoms caused by diseases or pests. *Biosystems Engineering*, 210:129–142, 2021.
 - [17] Ronja Güldenring and Lazaros Nalpantidis. Self-supervised contrastive learning on agricultural images. *Computers and Electronics in Agriculture*, 191:106510, 2021.
 - [18] Xiefan Guo, Hongyu Yang, and Di Huang. Image inpainting via conditional texture and structure dual generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14134–14143, 2021.
 - [19] Jin Gwan Ha, Hyeonjoon Moon, Jin Tae Kwak, Syed Ibrahim Hassan, Minh Dang, O New Lee, and Han Yong Park. Deep convolutional neural network for classifying fusarium wilt of radish from unmanned aerial vehicles. *Journal of Applied Remote Sensing*, 11(4):042621, 2017.
 - [20] David Hughes, Marcel Salathé, et al. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*, 2015.
 - [21] Ryooya Katafuchi and Terumasa Tokunaga. Image-based plant disease diagnosis with unsupervised anomaly detection based on reconstructability of colors. *arXiv preprint arXiv:2011.14306*, 2020.
 - [22] Muhammad Attique Khan, Tallha Akram, Muhammad Sharif, Majed Alhaisoni, Tanzila Saba, and Nadia Nawaz. A probabilistic segmentation and entropy-rank correlation-based feature selection approach for the recognition of fruit diseases. *EURASIP Journal on Image and Video Processing*, 2021(1):1–28, 2021.
 - [23] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
 - [24] Zhenyu Li, Ning Li, Kaitao Jiang, Zhiheng Ma, Xing Wei, Xiaopeng Hong, and Yihong Gong. Superpixel masking and inpainting for self-supervised anomaly detection. In *BMVC*, 2020.
 - [25] Ke Lin, Liang Gong, Yixiang Huang, Chengliang Liu, and Junsong Pan. Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Frontiers in plant science*, 10:155, 2019.
 - [26] Jiang Lu, Jie Hu, Guannan Zhao, Fenghua Mei, and Changshui Zhang. An in-field automatic wheat disease diagnosis system. *Computers and electronics in agriculture*, 142:369–379, 2017.
 - [27] Yang Lu, Shujuan Yi, Nianyin Zeng, Yurong Liu, and Yong Zhang. Identification of rice diseases using deep convo-

- lutional neural networks. *Neurocomputing*, 267:378–384, 2017.
- [28] Juncheng Ma, Keming Du, Lingxian Zhang, Feixiang Zheng, Jinxiang Chu, and Zhongfu Sun. A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. *Computers and Electronics in Agriculture*, 142:110–117, 2017.
- [29] Giuliano L Manso, Helder Knidel, Renato A Krohling, and Jose A Ventura. A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust. *arXiv preprint arXiv:1904.00742*, 2019.
- [30] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.
- [31] Sharada P Mohanty, David P Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in plant science*, 7:1419, 2016.
- [32] Koushik Nagasubramanian, Asheesh K Singh, Arti Singh, Soumik Sarkar, and Baskar Ganapathysubramanian. Usefulness of interpretability methods to explain deep learning based plant stress phenotyping. *arXiv preprint arXiv:2007.05729*, 2020.
- [33] Paolo Napoletano, Flavio Piccoli, and Raimondo Schettini. Anomaly detection in nanofibrous materials by cnn-based self-similarity. *Sensors*, 18(1):209, 2018.
- [34] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Z Qureshi, and Mehran Ebrahimi. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv preprint arXiv:1901.00212*, 2019.
- [35] Haseeb Nazki, Sook Yoon, Alvaro Fuentes, and Dong Sun Park. Unsupervised image translation using adversarial networks for improved plant disease recognition. *Computers and Electronics in Agriculture*, 168:105117, 2020.
- [36] Santanu Phadikar, Jaya Sil, and Asit Kumar Das. Classification of rice leaf diseases based on morphological changes. *International Journal of Information and Electronics Engineering*, 2(3):460–463, 2012.
- [37] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer, 2017.
- [38] Gaurav Sharma, Wencheng Wu, and Edul N Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 30(1):21–30, 2005.
- [39] Asheesh Kumar Singh, Baskar Ganapathysubramanian, Soumik Sarkar, and Arti Singh. Deep learning for plant stress phenotyping: trends and future perspectives. *Trends in plant science*, 23(10):883–898, 2018.
- [40] Aditya Sinha and Rajveer Singh Shekhawat. Olive spot disease detection and classification using analysis of leaf image textures. *Procedia Computer Science*, 167:2328–2336, 2020.
- [41] Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic. Deep neural networks based recognition of plant diseases by leaf image classification. *Computational intelligence and neuroscience*, 2016, 2016.
- [42] Kai Tian, Jiuhao Li, Jiefeng Zeng, Asenso Evans, and Lina Zhang. Segmentation of tomato leaf images based on adaptive clustering number of k-means algorithm. *Computers and Electronics in Agriculture*, 165:104962, 2019.
- [43] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [44] David M Woebbecke, George E Meyer, Kenneth Von Bargen, and David A Mortensen. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, 38(1):259–269, 1995.
- [45] Xianghua Xie and Majid Mirmehdi. Texture exemplars for defect detection on random textures. In *International Conference on Pattern Recognition and Image Analysis*, pages 404–413. Springer, 2005.
- [46] Yonghua Xiong, Longfei Liang, Lin Wang, Jinhua She, and Min Wu. Identification of cash crop diseases using automatic image segmentation algorithm and deep learning with expanded dataset. *Computers and Electronics in Agriculture*, 177:105712, 2020.
- [47] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5505–5514, 2018.
- [48] Yu Zeng, Zhe Lin, Huchuan Lu, and Vishal M Patel. Cr-fill: Generative image inpainting with auxiliary contextual reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14164–14173, 2021.
- [49] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.
- [50] Shanwen Zhang, Xiaowei Wu, Zhuhong You, and Liqing Zhang. Leaf image based cucumber disease recognition using sparse representation classification. *Computers and electronics in agriculture*, 134:135–141, 2017.
- [51] Shanwen Zhang, Yihai Zhu, Zhuhong You, and Xiaowei Wu. Fusion of superpixel, expectation maximization and phog for recognizing cucumber diseases. *Computers and Electronics in Agriculture*, 140:338–347, 2017.
- [52] Wendong Zhang, Junwei Zhu, Ying Tai, Yunbo Wang, Wenqing Chu, Bingbing Ni, Chengjie Wang, and Xiaokang Yang. Context-aware image inpainting with learned semantic priors. *arXiv preprint arXiv:2106.07220*, 2021.
- [53] W Zhen, Z Shanwen, and Z Baoping. Crop diseases leaf segmentation method based on cascade convolutional neural network. *Comput. Eng. Appl.*, 56(15):242–250, 2020.