Article

Not peer-reviewed version

Overcoming Domain Shift in Neural Networks for Accurate Plant Counting in Aerial Images

Javier Rodriguez-Vazquez * , Miguel Fernandez-Cortizas , David Perez-Saura , Martin Molina , Pascual Campoy

Posted Date: 3 February 2023

doi: 10.20944/preprints202302.0070.v1

Keywords: deep learning; aerial imagery; precision agriculture; plant detection; domain adaptation; unsupervised learning; self-supervision; adversarial learning; domain shift; tropical crops

Article

Overcoming Domain Shift in Neural Networks for Accurate Plant Counting in Aerial Images

Javier Rodriguez-Vazquez [1,2,*], Miguel Fernandez-Cortizas [1], David Perez-Saura[1], Martin Molina[2] and Pascual Campoy [1]

[1] Computer Vision and Aerial Robotics Group, Centre for Automation and Robotics (C.A.R.), Universidad Politécnica de Madrid (UPM-CSIC), Madrid, Spain
[2] Computer Vision and Aerial Robotics Group, Department of Artificial Intelligence, Universidad Politécnica de Madrid (UPM), Madrid, Spain
* Correspondence: javier.rodriguez.vazquez@upm.es

Abstract: This paper presents a novel approach for accurate counting and localization of tropical plants in aerial images that is able to work in new visual domains in which the available data is not labeled. Our approach uses deep learning and domain adaptation, designed to handle domain shift between the training and test data, which is a common challenge in this agricultural applications. This method uses a source dataset with annotated plants and a target dataset without annotations, and adapts a model trained on the source dataset to the target dataset using unsupervised domain alignment and pseudolabeling. The experimental results show the effectiveness of this approach for plant counting in aerial images of pineapples under significative domain shift, achieving a reduction up to 97% in the counting error when compared to the supervised baseline.

Keywords: deep learning; aerial imagery; precision agriculture; plant detection; domain adaptation; unsupervised learning; self-supervision; adversarial learning; domain shift; tropical crops

1. Introduction

Precision agriculture relies heavily on the ability to accurately count and locate plants in crop fields via aerial imagery. This task is crucial in optimizing the use of resources [1], such as water [2], fertilizers, and pesticides [3], by targeting specific areas and reducing waste. Furthermore, accurate plant detection can aid in improving crop yields by identifying and addressing issues such as pests or diseases [4]. Additionally, it can minimize the environmental impact of agriculture by reducing the use of chemicals and the risk of pollution [5], ultimately enhancing food security and sustainability [6].

However, one of the major challenges in this field is the domain gap between different crops, as each crop possesses its own distinct characteristics, such as leaf shape, color, light conditions, or even soil, making it difficult to generalize plant detection models from one crop to another. Traditional methods of plant detection, which require manual annotation of images, are both time-consuming and costly due to the labor-intensive nature of the process, the need for a large number of labels, and the risk of error and inconsistency. Furthermore, current approaches do not account for the domain gap, necessitating the generation of a dataset that encompasses all possible domains, thus increasing the cost significantly.

This study tackles the challenge of accurate plant counting and localization in crop fields despite domain shifts, by introducing a new semi-supervised method. Our approach represents a breakthrough over previous methods, as it utilizes dot annotations instead of bounding boxes, thus decreasing the cost of labeling and enabling the use of unlabeled data from new unseen shifted domains in an unsupervised manner. Our method generalizes to new domains by utilizing only unlabeled data. Our method comprises of two key mechanisms: (1) unsupervised adversarial domain alignment of intermediate features, and (2) self-supervision on the target domain through the inclusion of a novel pseudolabeling loss.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

To validate the effectiveness of our system, we conducted several experiments on a dataset of pineapple crops that we created, which is composed of multiple sub-datasets, each representing a crop from a distinct geographical region. As depicted in Figure 1, there is a striking domain shift between datasets due to factors such as lighting conditions, growth stage, soil type, etc., which makes generalization from one dataset to another extremely challenging with traditional fully-supervised methods. To the best of our knowledge, this research is the first to tackle the problem of plant detection using dot annotations in a semi-supervised manner while addressing domain shifts. To gauge the impact of our contributions, we compared the proposed method with a fully-supervised baseline method that only utilizes the available labels as input. Our approach demonstrates a significant enhancement in localization and counting accuracy on the target domain.

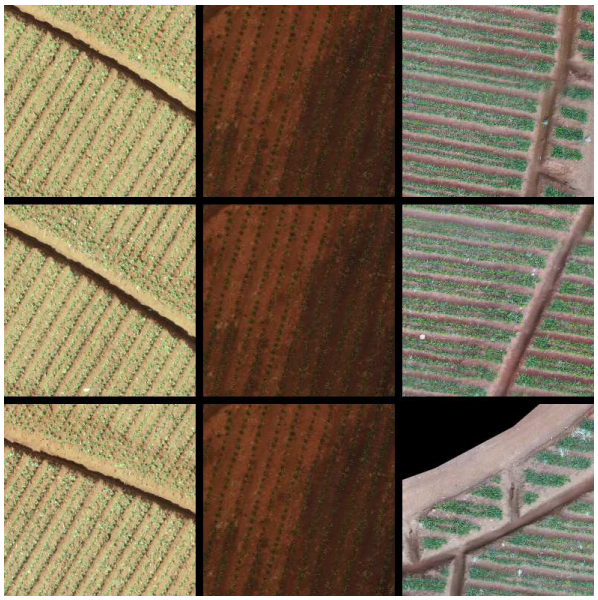

Figure 1. Domain gap between different crop domains in the pineapple dataset. The images in each column belong to a different crop domain, characterized by different lighting conditions, plant growth stage, soil type, and other factors. The significant variations between domains pose a challenge for traditional fully supervised methods, which struggle to generalize across domains.

In the forthcoming sections, we will first review the related literature in the field of crop counting and localization from aerial imagery. Following that, we will elaborate on our proposed approach in detail, including the network architecture and semi-supervised training procedure. We will then demonstrate the efficacy of our approach through the results of our experiments, which exhibit its versatility across a diverse range of crops and conditions. Finally, we will conclude with a discussion of the implications of our work and potential avenues for future research. Our key contributions include: (1) the introduction of a novel counting method from aerial images using dot annotations, and its application to precision agriculture, (2) the presentation of an unsupervised domain adaptation method, which enables the model to leverage information from unlabeled domains to improve its generalization capabilities, and (3) the proposal of a new research direction on semi-supervised methods for crop counting robust to domain gaps, with the goal of reducing costs in agriculture. In summary, our approach represents a significant advancement in the field of crop counting and localization from aerial imagery. By harnessing the capabilities of deep learning and introducing a novel semi-supervised training procedure, we are able to generalize to new domains using only unlabeled data. This makes our approach highly desirable for real-world applications, where labeled data may be scarce. This research has the potential to accelerate the implementation of precision agriculture practices and make a positive impact on the efficiency and sustainability of agricultural operations.

 doi:10.20944/preprints202302.0070.v1

2. Related work

2.1. Crop monitoring using aerial images

Crop monitoring is a vital aspect of precision agriculture, and with the advent of deep learning, it has become increasingly efficient and accurate. Unmanned aerial vehicles (UAVs) have played a crucial role in crop monitoring [7], providing high-resolution images and enabling fastmonitoring of crops [8]. The use of UAVs equipped with different cameras, such as RGB, thermal, and hyperspectral cameras, has opened up new possibilities for crop monitoring.

RGB cameras are the most widely used cameras for crop monitoring [9,10], providing high-resolution images that are useful for identifying plant growth stages, identifying pests and diseases, and estimating crop yield. Thermal cameras, on the other hand, can detect temperature variations in the crop canopy, providing useful information about plant stress [11] and water uptake [12]. Hyperspectral cameras, which can capture images across a wide range of wavelengths, can provide detailed information about the chemical composition of crops, such as chlorophyll content and water content [13].

The utilization of deep learning models in crop monitoring has been widespread, with object detection and semantic segmentation being the most prevalent approaches. Research has been conducted utilizing object detection to identify individual plants in various crops, such as mango [14,15], banana [16] or citrus tree [17,18]. Object detection models, such as YOLO [19], require input data in the form of large sets of bounding boxes. While these datasets are relatively inexpensive to acquire, using dot annotations and only labeling the center of the object can reduce the amount of input data required by half. In contrast, semantic segmentation approaches enable the segmentation of pixels into different regions, such as leaves, stems, and background, providing detailed information about the structure of the crop [20–22]. However, it is worth noting that datasets for semantic segmentation are very expensive.

2.2. Object counting from dot annotations

The task of accurately counting objects within images can be approached through a variety of methods, including individual object detection [23], direct count estimation [24], and the generation of intermediate density maps. Our approach aligns with the method of individual object detection as proposed by [23], as it allows for the preservation of object position information. This is achieved by generating a proximity map of each pixel to the center of the objects, utilizing dot annotations placed near the centers.

Recent advancements in object counting methods have emphasized the utilization of density map estimation, first introduced by [25]. This approach utilizes linear regression on SIFT features to estimate the density map of the desired objects. Subsequent developments in this line include the implementation of regression forests in place of linear regression [26], modification of the data generation procedure [27], or the application of postprocessing techniques to eliminate low confidence detections [28].

The application of convolutional neural networks for the estimation of density maps was first introduced in [29], as a means of circumventing the need for handcrafted features. Building upon this concept, the method proposed in [30] incorporates a redundant counting technique by utilizing square kernels, enabling neurons to count the number of objects within their receptive field.

Recent progress in object counting has also been made through modifications to neural network architecture, such as the introduction of upsampling layers for enhanced counting resolution and improved centroid localization, as proposed by [31]. Additionally, techniques such as those employed by [32,33] address the challenge of varying object sizes through the implementation of multiresolution methods. Furthermore, [34] proposed a channel attention module, adaptable to a wide range of neural networks, that enhances counting accuracy.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

Efforts have also been made to address the challenge of errors arising from uniform background regions in object counting, such as incorporating self-attention modules [35], background segmentation [36,37], or designing region-based loss functions that specifically consider background regions [38].

2.3. Unsupervised domain adaptation

Unsupervised domain adaptation addresses the challenge of applying a model trained on a specific source distribution to a related but distinct target distribution. While traditional "shallow" domain adaptation methods focus on reweighting source samples and learning a shared feature space between the source and target datasets [39], the utilization of deep neural networks (DNNs) in deep domain adaptation has proven to yield more transferable representations. This is due to the tendency of DNNs to learn highly transferable features in the lower layers, with decreasing transferability in higher layers. Therefore, the goal of deep domain adaptation is to leverage this property of DNNs.

One popular approach to deep domain adaptation is the Deep Adaptation Network (DAN) [40], which utilizes weighting techniques to match the different domain distributions and improve feature transferability. Additionally, DAN employs an optimal multi-kernel selection method to further reduce domain discrepancy.

Another approach, Deep CORAL [41], is an unsupervised method that utilizes a non-linear transformation to align the correlations of layer activations in DNNs. The use of a non-linear transformation in Deep CORAL enables the capturing of complex relationships between layers, resulting in improved performance compared to linear transformations used by other methods.

Deep domain confusion [42] is a technique for creating a representation that is both semantically meaningful and invariant across different domains. This is achieved by introducing an adaptation layer into the CNN architecture and implementing an additional loss function referred to as "domain confusion loss". This allows the model to learn representations that are not biased towards any particular domain, making it more generalizable when applied to new contexts.

Another promising approach is CoGAN (Coupled Generative Adversarial Networks) [43], which can learn a joint distribution of multi-domain images without requiring tuples of corresponding images in different domains in the training set. To accomplish this, CoGAN uses samples drawn from the marginal distributions and enforces a weight-sharing constraint to favor the joint distribution solution over the product of marginal distributions.

Finally, the DANN method [44] works by augmenting a feed-forward model with standard layers and a novel gradient reversal layer. This enables the model to learn deep features that are both specific to the source domain and applicable to the target domain. The gradient reversal layer promotes adaptation behavior, allowing for successful transfer across different domains when trained using standard backpropagation.

3. Method

3.1. Overview

Our proposed method for crop counting and localization from aerial imagery comprises of two distinct stages: (1) A convolutional neural network (CNN) is utilized to predict the probability of the presence of the center of each plant in the input image, and (2) a blob detector is employed to localize each plant.

To address the challenge of domain shifts between different crops, we propose a semi-supervised training procedure that incorporates two key mechanisms: an adversarial framework and pseudolabeling. In the adversarial framework, we utilize a domain discriminator (D_{dom}) to learn to differentiate between samples from two datasets that are similar but diverge due to domain shifts (e.g., different soils, growth stages of plants, lighting conditions, etc.). This forces the main network to only utilize relevant features that are present in both domains, aligning the intermediate feature representations of both domains.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

However, this approach only focuses on making the domains indistinguishable at the feature level, which could result in the loss of semantic information in the target data. To circumvent this, we introduce a pseudolabeling mechanism that reinforces the confident outputs of the network and prevents forgetting as training progresses. This enables us to incorporate samples from a different source domain during training in an unsupervised manner, while still preserving the semantic information present in the target domain.

3.2. Supervised Baseline model

We adopt the methodology presented in [45] as our supervised baseline. This approach seeks to achieve count and localization of objects by dividing the problem into two primary stages.

Initially, a deep neural network G maps the input image I to a new image C, representing the probability of the presence of the center of an object at each pixel. To aid in this objective, a second neural network D_{image} is utilized to discriminate between ground truth target images and those generated by G through an adversarial training procedure.

Subsequently, each individual object is detected using the Laplacian of Gaussian (LoG) [46], enabling the detection of objects that are very close or even overlapping.

The following sections provide a more comprehensive understanding of the baseline method and the modifications we have made to it. For a thorough overview of the method, we direct the reader to the original publication [45].

3.2.1. Target label construction

The aim of the generator network G is to learn to create a map that shows the probability of finding the center of an object at each pixel of the image. This is done by defining a Gaussian map G_m for each pixel x using the following equation.

$$G_m(x, P) = e^{-\frac{DT(x,P)^2}{2\sigma^2}}$$

where P is the set of point annotations in the image, and σ is a configurable parameter that determines the width of the blobs in the map. To calculate G_m, we need to first calculate the distance transform DT of the annotation set:

$$DT(x, P) = min_{y \in P}; dist(x, y)$$

This map represents the distance of each pixel x to the closest point y from the annotated set P. We use the Euclidean distance to calculate the distances:

$$dist(x, y) = \sqrt{(x_i - y_i)^2 + (x_j - y_j)^2}$$

To avoid detecting objects that are too close together as a single one, we emphasize the frontiers between objects by setting the values of those pixels to 0. This is done using the following equation:

$$T(x, P) = \begin{cases} 0 & \exists p_x \exists p_y, |dist(x, p_x) - dist(x, p_Y)| \leq t_d \\ G_m(x, P) & otherwise \end{cases} \quad (1)$$

where t_d is a distance threshold that determines the thickness of the frontiers, and p_x and p_y are annotations in P. This threshold is set to 2 for all experiments. The use of these frontiers encourages the neural network to learn to divide objects that are too close together into different blobs, making detection easier.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

3.2.2. Network architecture

The baseline method uses a generative adversarial network (GAN) architecture to avoid blurry results when minimizing the Euclidean distance between pairs of pixels [47]. The GAN consists of two networks: a generator G that learns to map input images to intermediate representations, and a discriminator D_{image} that learns to distinguish between the generated outputs from G and the ground truth. The generator and discriminator compete with each other in an adversarial fashion, with the generator trying to produce outputs that are indistinguishable from the ground truth, and the discriminator trying to accurately identify the generated outputs. This competition drives both networks to improve, ultimately resulting in more accurate center maps.

We use a modified Up-net [45] architecture as the generator network. This design is based on the work presented in [48], and combines the advantages of U-net [49] and fully convolutional networks (FCNs) [50]. U-nets are effective at extracting rich semantic features at the bottleneck layer and recovering high-frequency information at higher layers using skip connections, while FCNs use skip connections to propagate information throughout the network without increasing the total number of parameters. By combining these two approaches, our modified U-net architecture is able to extract useful features from the input image and reconstruct it accurately, without significantly increasing the number of parameters in the network. The architecture is shown in Figure 2

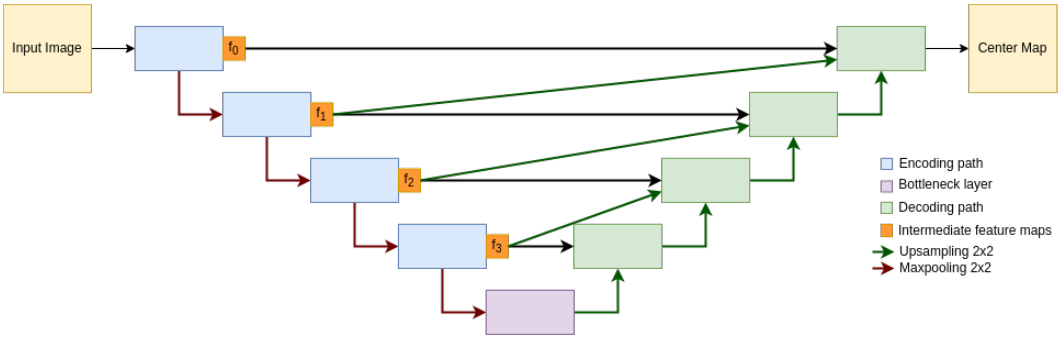

Figure 2. Selected Up-Net architecture [45] for the generator network. The network has 4 main parts, (1) the encoding path generates rich features to represent the input image decreasing the resolution, (2) the bottleneck layer, (3) the decoding path increases the resolution of the generated features and generates the final output, (4) the skip connections provide high spatial resolution to the decoding path. Each convolutional block is composed of three convolutions (with kernel size 3, each one followed by a Batch Normalization layer [51] and with ReLU activation). Green arrows depict the upsampling layers, which are composed of a first bicubic upsampling of the feature maps that doubles the resolution and followed by a convolutional layer that halves the number of channels, Batch Normalization andd ReLU activation.

The discriminator architecture follows the PatchGAN [47] design, as outlined in Table 1. By analyzing patches of the input image rather than the entire image, the network is able to lower computational costs and improve efficiency.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

Table 1. The architecture of the discriminator used in our model is presented in this table. We employ LeakyReLU activation functions in all layers except for the output layer, where we use the hyperbolic tangent function. With a stride of 32 pixels, this discriminator is able to effectively analyze and distinguish structures of up to that size.

Layer	Input channels	Output channels	Kernel size	Stride	Activation
Conv2D	1	64	4	2	LeakyReLU(0.2)
BatchNormalization					
Dropout					
Conv2D	64	128	4	2	LeakyReLU(0.2)
BatchNormalization					
Dropout					
Conv2D	128	256	4	2	LeakyReLU(0.2)
BatchNormalization					
Dropout					
Conv2D	256	512	4	2	LeakyReLU(0.2)
BatchNormalization					
Dropout					
Conv2D	512	512	4	2	LeakyReLU(0.2)
BatchNormalization					
Dropout					
Conv2D	512	256	4	1	LeakyReLU(0.2)
BatchNormalization					
Conv2D	256	1	4	1	Tanh

Inspired by the work of Ganin et al. [44], we added a gradient reversal layer (GRL) between the generator and the discriminator. This change allows us to train both networks jointly in the same forward pass, reducing the training complexity and computational costs while maintaining opposing objectives in each network. As proposed in [44], we scale the gradients flowing from the discriminator to the generator inversely proportional to the current training step, to overcome the early instabilities of adversarial training. Figure 3 provides a visual overview of the training procedure.

During training, the generator and discriminator networks are optimized using a combination of adversarial and reconstruction losses (Equations 3 , 2). The adversarial loss is used to encourage the generator to produce outputs that are indistinguishable from the ground truth, while the reconstruction loss is used to encourage the generator to accurately reconstruct the input image. We adopted a least square GAN [52] objective. These losses are combined and used to update the weights of the generator and discriminator networks, ultimately leading to more accurate results.

$$\begin{aligned} \mathcal{L}_{Baseline}(G, D_{image}) &= \mathcal{L}_{supervised}(G) + \lambda_{Adv}\mathcal{L}_{Adv}(G, D_{image}) \\ \mathcal{L}_{supervised}(G) &= \mathbb{E}_{x,y}\left[||y - G(x)||_1\right] \end{aligned} \tag{2}$$

$$\mathcal{L}_{Adv}(G, D_{image}) = \mathbb{E}_{x,y}\left[||1 - D_{image}(y)||_2\right] + \mathbb{E}_{x,y}\left[||-1 - D_{image}(G(x))||_2\right] \tag{3}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

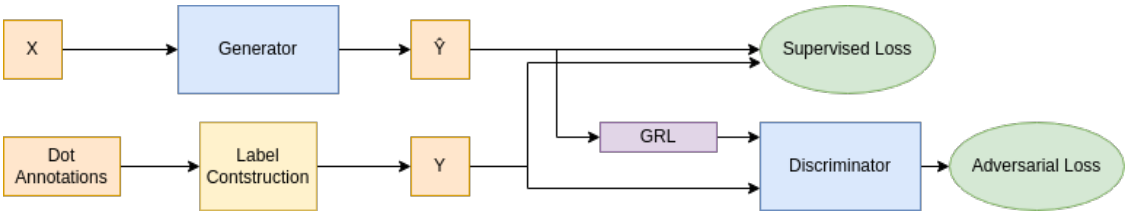

Figure 3. The baseline method uses two neural networks, G and D_{image}, which are trained together in an adversarial manner. G attempts to map input images to center maps, while D_{image} tries to distinguish between ground truth and generated outputs. The gradient reversal layer (GRL) allows both networks to be trained together, even though they have opposing objectives, by reversing the sign of the gradient and scaling it when it flows from D_{image} to G. This allows the networks to be trained in a single pass.

3.3. Semi supervised training under domain distribution shifts

In this work, we aim to address the challenge of domain shift and improve the generalization of a base model by incorporating unlabeled data from the target domain into the training process. To achieve this, we propose a novel approach that combines two key mechanisms: adversarial alignment of intermediate features between the two domains and pseudo-labeling of the target domain data. Our adversarial alignment strategy involves training a neural network to perform a task while also learning domain-invariant features through an adversarial training process. This helps the model generalize better to the target domain by learning features that are common across both domains. Our pseudo-labeling approach involves using the model's own predictions as labels for the target domain data, thereby preserving the richness and meaning of the target domain features and allowing the model to capture the unique characteristics of the target domain. By combining these two mechanisms, we are able to improve the performance of the base model on the target domain, enabling it to generalize to new domains.

3.3.1. Multilevel adversarial domain align

In order to improve the generalization of our base model to the target domain, we draw inspiration from the DANN method [44]. This method involves training a secondary neural network, called the domain discriminator (D_{domain}), to distinguish between samples from the source and target domains based on the intermediate feature representation produced by the main network. The main network is then trained in an adversarial manner, using the gradients from the domain classification loss to update the network weights and force the encoder path to extract features that are invariant across domains.

However, in our case, we are using a U-Net-like architecture with skip connections. Enforcing feature invariance at a single point (e.g., at the bottleneck layer) is not sufficient for aligning the feature spaces of the two domains due to the flow of information between different levels of the network enabled by the skip connections. To address this issue, we propose a new multilevel domain discriminator that takes as input the features at each skip connection level, aligning the domains at each level and ensuring that all features used by the decoder path are aligned.

Our multilevel discriminator architecture consists of four discriminator blocks and a final block, as illustrated in Figure 4. Each discriminator block takes as input the features at the current level, as well as the output of the previous block (except for the first block). This hierarchical representation of the features allows the network to extract and combine features at each skip connection level, enabling more effective alignment of the feature spaces of the two domains. The final block of the discriminator aggregates all of this information and uses it to determine, at the patch level, whether the features are from the source or target domain, following the PatchGan architecture proposed in [47]. Additionally, we have added residual connections [53] to improve the propagation of gradients and facilitate the training process.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

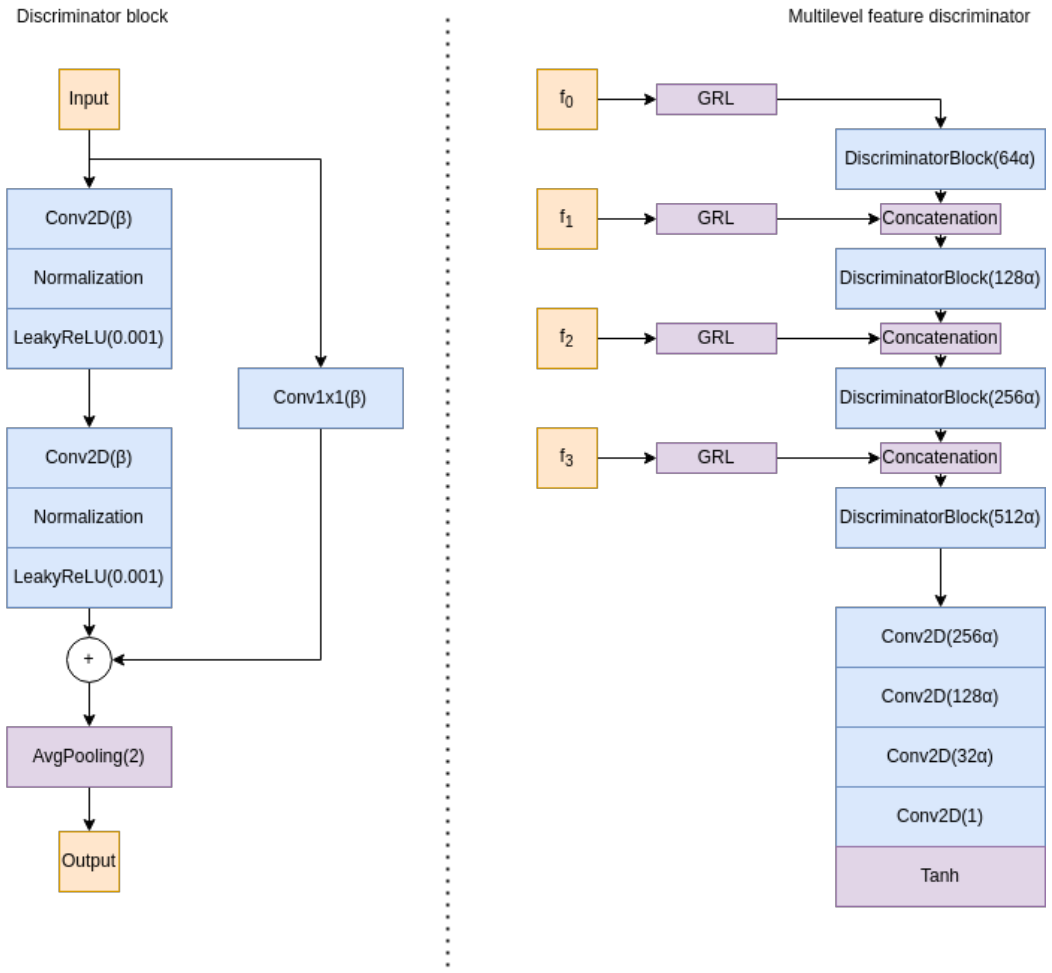

Figure 4. Multilevel discriminator architecture. The aim of this design is to to adapt features at various levels ($f_0 - f_3$). The architecture consists of five main blocks, with the first four blocks taking as input the features at the current skip connection level and the output of the previous block in order to build a hierarchical representation of the features. The last block is used to determine whether the features come from a source or target sample. The use of a Gradient Reversal layer at each input enables both networks to be trained simultaneously, even if they have opposing objectives. It is important to note that each discriminator block includes a residual skip connection to prevent already indistinguishable features at higher levels from hindering the adaptation of more shallow ones.

We denote the domain label d as an indicator, with $d = -1$ indicating that a sample is drawn from the source domain and $d = 1$ indicating that it is drawn from the target domain.

To train the domain discriminator, we follow a Least Squares Generative Adversarial Network (LSGAN) [52] objective, which leads to the following loss term:

$$\mathcal{L}_{Domain}(E, D_{Domain}) = \mathbb{E}x, y\,[||d - D_{Domain}(E(x))||_2] \tag{4}$$

Here, E represents the encoder part of the generator network G, as shown in Figure 2. The domain discriminator D_{Domain} takes as input the intermediate feature representation at all skip levels produced by the encoder E, and produces a prediction of the domain label d for that sample.

3.3.2. Selective Confidence Pseudolabeling

While adversarial alignment can ensure the statistical alignment of intermediate features, it does not guarantee semantic alignment. As a result, it is possible that the resulting intermediate

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

representations in the target domain, while conforming to the same data distribution as the source domain, may not be useful for detecting plants.

To address this issue, we observed that at early stages of training, when the network is not fully adapted to the source domain, it outputs some highly confident predictions that are accurate. However, as training progresses, the network becomes better suited to the provided data and forgets these confident outputs, resulting in an inability to detect any of the plants in the target data. To capitalize on this phenomenon, we have developed a selective confidence pseudolabeling technique with the goal of avoiding the forgetting of these early accurate outputs.

To compute the pseudolabel, we first gather the confident coordinates. This is achieved by smoothing the output of the network, $\hat{y}$, with a Gaussian filter and then identifying local maxima in the output. We only consider highly confident outputs with two thresholds: an adaptive threshold $t_{adaptive}$, set at 0.9 of the maximum value of the current output, and a hard absolute threshold $t_{absolute}$, typically set at 0.5. Additionally, we filter out maxima that are too close together using a threshold $t_{distance}$, set at 2σ, where σ is a configurable parameter determining the size of the blobs in the baseline method. The pseudolabel is then computed using only these dot annotations, similar to the baseline method.

Since the network does not detect all objects at the beginning of training, we do not want to train the network using negative pseudolabels. To address this, we mask the loss in pixels where the pseudolabel is less than a threshold t_{mask}, typically set at 0.2. Finally, we compute the loss between $\hat{y}$ and the pseudolabel $\tilde{y}$ using an $L2$ (MSE) loss.

To further improve the robustness of our approach, we use an adaptive scaling term, β_{scale}, to multiply the loss term. This term is scheduled to be very small at the beginning of training and gradually increases as training progresses. This helps to better gather confident pseudolabels as the network becomes more confident.

Overall, our masked selective confidence pseudolabeling approach allows us to leverage the confident outputs of the network at early stages of training and avoid the forgetting of these outputs as training progresses. This helps to improve the semantic alignment of the intermediate representations in the target domain and improve the ability of the network to detect plants in the target data.

$$\mathcal{L}_{pseudolabel}(G) = \begin{cases} 0 & \tilde{y} < t_{mask} \\ \beta_{scale}\mathbb{E}x,y\left[||(E(x)) - \tilde{y}||_2\right] & otherwise \end{cases} \quad (5)$$

Being $\tilde{y}$ the pseudolabel generated by Algorithm 1.

Algorithm 1 Proposed selective confidence pseudolabeling

procedure COMPUTEPSEUDOLABEL($\hat{y}$)
  $\hat{y}_{smooth} \leftarrow$ MedianFilter($\hat{y}$)
  $M \leftarrow$ FindLocalMaxima($\hat{y}_{smooth}$)
  $\hat{P} \leftarrow$ FilterMaxima($M, t_{adaptive} = 0.9, t_{absolute} = 0.5$)
  $\hat{P} \leftarrow$ FilterCloseMaxima($\hat{P}, t_{distance} = 2\sigma$)
  $\tilde{y} \leftarrow T(\hat{P})$
  return $\tilde{y}$
end procedure

4. Results

In the following section, we present the results obtained from validating our proposed method.

4.1. Experimental setup

All the proposed method is implemented using the frameworks Pytorch [54] and Pytorch Lightning [55]. The GPU used for training has been an Nvidia GeForce RTX 2080 Ti. For training all networks we use Adam [56] optimizer with 0.0001 learning rate.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

To increase the generalization capabilities of the model we use RandAugment [57] with 3 steps in all tests.

4.1.1. Dataset

In this study, we employ an aerial imagery dataset of pineapple crops from various geographical regions to demonstrate the effectiveness of our proposed method in handling domain shift. The dataset comprises of a diverse set of images that exhibit significant variations in lighting conditions, plant growth stage, soil type, and other factors. It is composed of several sub-datasets, each belonging to a crop from a different geographical area, as illustrated in Figure 1.

The images in the dataset were labeled using dot annotations, which mark the center of each plant. In total, the dataset comprises 2944 images, with a total of 33280 plants. The images have a resolution of 256x256 pixels and are in the RGB color space.

To evaluate the effectiveness of our proposed method, we divided the dataset into three domain folds: A, B, and C. Each fold corresponds to a single crop with distinct characteristics. This enabled us to assess the generalization ability of our method across different domains. Folds A and B are roughly the same size, with 1408 and 1280 images, respectively, while fold C contains only 256 images. This imbalanced distribution allows us to test the robustness of our method against uneven domain sizes.

Furthermore, we have gathered a separate dataset specifically for testing the domain generalization abilities of our system. This dataset is composed of data from 9 distinct domains (from different geographical areas), each with a unique set of characteristics. To ensure a thorough evaluation, we have gathered a total of 4224 images across all domains,reserving 64 from each domain for testing purposes. Although the test dataset is balanced, the training dataset is strongly unbalanced across domains, with domains represented as low as 3% while others account for 20% of the representation, conforming an additional challenge. In total, this dataset comprises 45782 plants. Figure 5 depicts a sample of the dataset.

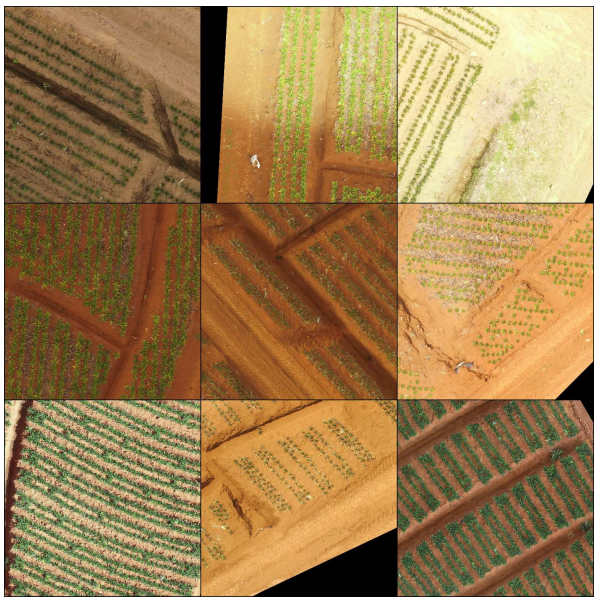

Figure 5. Sample of the General Domain dataset depicting 9 diverse domains with unequal representation. The distribution of characteristics poses a challenge for generalizing to all domains simultaneously, as some have a greater prominence over others.

4.2. Experiments

In this study, we aimed to investigate the impact of each component of our proposed method for domain adaptation in aerial images of pineapple crops. To evaluate the performance of our method, we selected the relative Mean Absolute Error (rMAE) of crop count estimates as the evaluation metric. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

rMAE is defined as the ratio of the absolute error to the true count (presented in percentaje). For each trial, we randomly selected a 70% of the dataset for training and validation, and used the remaining images as the test set. To estimate the mean and standard deviation of the rMAE, we repeated this procedure ten times.

4.2.1. Ablation study

We conducted an ablation study to understand the individual contributions of each component of our method to the final performance. Our ablation study consisted of five experimental trials, in which we systematically introduced and removed components of our method. The results of the ablation study are presented in Table 2. The first experiment employed the baseline method without any additional modifications. In the second experiment, we introduced the adversarial domain alignment mechanism and observed an improvement in performance on the target domain. However, we encountered convergence issues when training both discriminators simultaneously, so in the third experiment, we disabled the adversarial branch of the baseline method to investigate its impact. The fourth experiment examined the pseudo-labeling approach without any domain alignment, and we observed that the pseudo-labeling approach relies on accurate and confident network outputs. Without domain alignment, the model began to reinforce incorrect pseudo-labels, leading to a significant decrease in performance. Finally, in the fifth experiment, we incorporated both pseudo-labeling and domain alignment mechanisms and observed a significant reduction in error.

Table 2. Ablation study results, showing the impact of each component of our proposed method on rMAE.

$\mathcal{L}_{supervised}$	$\mathcal{L}_{adv}$	$\mathcal{L}_{domain}$	$\mathcal{L}_{pseudolabel}$	$rMAE(\%)$
✓	✓			59.39 ± 21.31
✓	✓	✓		24.63 ± 19.55
✓		✓		12.32 ± 7.29
✓			✓	27.42 ± 23.75
✓		✓	✓	2.44 ± 1.54

4.2.2. Domain adaptation experiments

We also evaluated the domain adaptation capabilities of our method by testing the adaptation from one source domain to another. For this evaluation, we utilized datasets A, B, and C and compared the results of our proposed approach to those of two fully-supervised methods: a baseline method that can only access source domain labels and an oracle method that had access to both source and target domain labels. Table 3 summarizes the results of our domain adaptation experiments. The results illustrate that our method consistently demonstrates proficiency in adapting between domains, resulting in a mean reduction in error up to 97%. However, there was a single instance where the reduction in error was limited to 10% only. Nonetheless, when our method successfully performed the adaptation, the error margins were closely aligned with those of the oracle method.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0070.v1

Table 3. Results of our unsupervised domain adaptation approach in rMAE(%). We demonstrate the effectiveness of our method in adapting across three distinct domains (A, B, and C). Each row shows the results of the models trained with one source dataset and tested on another one. The final column represents the performance of a fully supervised model that has access to both source and target domain labels, serving as an upper bound for comparison.

Experiment	Only source	Our method	Oracle
$A \rightarrow B$	59.39 ± 21.31	2.44 ± 1.54	2.39 ± 1.12
$A \rightarrow C$	56.93 ± 24.53	5.94 ± 11.45	6.43 ± 1.35
$B \rightarrow A$	48.60 ± 16.43	1.42 ± 2.97	1.39 ± 1.73
$B \rightarrow C$	87.12 ± 16.77	6.24 ± 4.59	3.44 ± 2.65
$C \rightarrow A$	91.79 ± 15.31	7.85 ± 4.90	1.42 ± 1.76
$C \rightarrow B$	82.95 ± 14.19	74.85 ± 12.60	2.16 ± 1.92

4.2.3. Domain generalization experiments

To measure the ability of our method to generalize across various domains, we performed experiments that evaluated its performance on the domain generalization dataset, with source domains A, B, and C. The results of these experiments are summarized in Table 4. Our findings indicate that, while our unsupervised approach consistently outperforms the supervised baseline, achieving a mean reduction in error of 61%, there is still room for improvement in this setting if we compare the results obtained with the oracle.

Table 4. Results on domain generalization of our approach in rMAE(%). We demonstrate the effectiveness of our method in generalizing to several domains at the same time in an unsupervised manner. We show the results on the generalized dataset training with just one source dataset in each column. The final row depicts the performance of a fully supervised model that has access to all labels, serving as an upper bound for comparison.

Source	Only source	Our method	Oracle
A	58.65 ± 26.71	21.61 ± 11.27	3.43 ± 4.56
B	70.19 ± 25.18	25.73 ± 15.20	2.97 ± 9.17
C	68.97 ± 37.27	29.46 ± 17.49	5.371 ± 5.74

5. Discussion

In this study, we present a novel semi-supervised approach for precise plant counting in aerial images of crop fields, which effectively addresses the challenge of domain shift between training and test data commonly faced in agricultural applications. Our method integrates deep learning and domain adaptation techniques to adapt a model trained on a labeled source dataset to an unlabeled target dataset through unsupervised domain alignment and pseudo-labeling.

The experimental results show that our approach excels in handling significant domain shifts in a one-to-one adaptation setting, reducing error by up to 97% compared to a supervised baseline, remaining very competitive with respect to an oracle model with access to all labels. However, the reliance on a confidence-based pseudolabeling approach can result in failure when the domain gap is significant. In such cases, false positive pseudolabels can cause the model to diverge in the target domain, leading to inability to recover. To overcome this limitation, developing mechanisms to detect such cases could be beneficial. In the domain generalization setting, our approach reduces error by an average of 61%, but there is still room for improvement as the gap with respect to oracle remains large. The confidence-based pseudolabeling approach can lead to early, confident outputs dominating the adaptation, resulting in underrepresentation of domains with large distances from the main domain. To address this, redesigning the adversarial framework to consider multiple target domains or creating subdomains in an unsupervised manner could detect underperforming domains and increase the weight of such domains to alleviate the underrepresentation issue.

 doi:10.20944/preprints202302.0070.v1

Future work in this field could focus on addressing the limitations identified in this study. One approach could be to enhance the pseudolabeling mechanism to ensure more accurate label predictions and prevent model divergence in the target domain. This could be achieved by implementing voting systems for pseudolabels or incorporating a history of pseudolabels. Another area for improvement is the adaptation framework, which could be modified to consider multiple target subdomains to better handle diverse domains.

Additionally, developing unsupervised techniques for early stopping and hyperparameter tuning would be beneficial, as these mechanisms currently rely on access to target validation data. The current method also requires retraining from scratch for every new domain and access to the source dataset. To overcome these limitations, exploring source-free retraining methods that only require access to target data samples and developing online domain adaptation techniques to continuously adapt to new domains without the need for retraining would be valuable avenues for future research.

6. Conclusions

In conclusion, our novel semi-supervised approach is a significant improvement for plant counting in aerial images of tropical crops. It effectively addresses the challenge of domain shift by combining deep learning and domain adaptation techniques through unsupervised domain alignment and pseudo-labeling. The results of our experiments demonstrate the potential of our approach in reducing error up to 97% compared to a supervised baseline and remaining competitive with respect to an oracle model with access to all labels.

Our approach has the potential to improve efficiency and sustainability in the agricultural sector, reducing the cost of crop monitoring and minimizing the use of resources such as water, fertilizers, and pesticides. However, there are limitations that must be addressed, such as the reliance on confidence-based pseudolabeling and the need for retraining for each new domain.

The findings of this paper provide a solid foundation for further research and have the potential to have a significant impact on the agricultural industry.To facilitate building upon of our work and encourage further research in this area, we are releasing the code used in this paper. The code can be accessed at https://github.com/cvar-upm/tropical_plant_counting_UDA.

Author Contributions: Conceptualization, J.R.V , M.M. and P.C.; methodology, J.R.V.; software, J.R.V., D.P.S. and M.F.C.; validation, D.P.S. and M.F.C.; formal analysis, J.R.V. and D.P.S; investigation, J.R.V.; resources, P.C.; data curation, D.P.S. and M.F.C; writing—original draft preparation, J.R.V.; writing—review and editing, D.P.S, M.F.C, M.M and P.C.; supervision, M.M. and P.C.; project administration, P.C.; funding acquisition, P.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the project COPILOT ref. Y2020/EMT6368 "Control, Monitoring and Operation of Photovoltaic Solar Power Plants by means of synergic integration of Drones, IoT and advanced communication technologies", funded by Madrid Government under the R&D Synergic Projects Program and partially funded by the project INSERTION ref. ID2021-127648OBC32, "UAV Perception, Control and Operation in Harsh Environments", funded by the Spanish Ministry of Science and Innovation under the program "Projects for Knowledge Generation". The work of the third author is supported by the Spanish Ministry of Science and Innovation under its Program for Technical Assistants PTA2021-020671.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Indigo Drones for proving the aerial images for this study, and to Adrián Álvarez-Fernández and Joaquín Fernández-Zafra for helping in the labeling process of the dataset.

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

References

1. Bongiovanni, R.; Lowenberg-DeBoer, J. Precision agriculture and sustainability. *Precision agriculture* **2004**, *5*, 359–387.
2. Lu, Y.; Liu, M.; Li, C.; Liu, X.; Cao, C.; Li, X.; Kan, Z. Precision Fertilization and Irrigation: Progress and Applications. *AgriEngineering* **2022**, *4*, 626–655.
3. Talaviya, T.; Shah, D.; Patel, N.; Yagnik, H.; Shah, M. Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. *Artificial Intelligence in Agriculture* **2020**, *4*, 58–73.
4. Li, W.; Chen, P.; Wang, B.; Xie, C. Automatic localization and count of agricultural crop pests based on an improved deep learning pipeline. *Scientific reports* **2019**, *9*, 1–11.
5. Roberts, D.P.; Short, N.M.; Sill, J.; Lakshman, D.K.; Hu, X.; Buser, M. Precision agriculture and geospatial techniques for sustainable disease control. *Indian Phytopathology* **2021**, *74*, 287–305.
6. Cohen, A.R.; Chen, G.; Berger, E.M.; Warrier, S.; Lan, G.; Grubert, E.; Dellaert, F.; Chen, Y. Dynamically Controlled Environment Agriculture: Integrating Machine Learning and Mechanistic and Physiological Models for Sustainable Food Cultivation. *ACS ES&T Engineering* **2021**, *2*, 3–19.
7. Barbedo, J.G.A. A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stresses. *Drones* **2019**, *3*, 40.
8. Hafeez, A.; Husain, M.A.; Singh, S.; Chauhan, A.; Khan, M.T.; Kumar, N.; Chauhan, A.; Soni, S. Implementation of drone technology for farm monitoring & pesticide spraying: A review. *Information Processing in Agriculture* **2022**.
9. Bouguettaya, A.; Zarzour, H.; Kechida, A.; Taberkit, A.M. A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images. *Cluster Computing* **2022**, pp. 1–21.
10. Bouguettaya, A.; Zarzour, H.; Kechida, A.; Taberkit, A.M. Deep learning techniques to classify agricultural crops through UAV imagery: a review. *Neural Computing and Applications* **2022**, pp. 1–26.
11. Pineda, M.; Barón, M.; Pérez-Bueno, M.L. Thermal imaging for plant stress detection and phenotyping. *Remote Sensing* **2020**, *13*, 68.
12. Stutsel, B.; Johansen, K.; Malbéteau, Y.M.; McCabe, M.F. Detecting plant stress using thermal and optical imagery from an unoccupied aerial vehicle. *Frontiers in plant science* **2021**, p. 2225.
13. Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote sensing* **2017**, *9*, 1110.
14. Xiong, J.; Liu, Z.; Chen, S.; Liu, B.; Zheng, Z.; Zhong, Z.; Yang, Z.; Peng, H. Visual detection of green mangoes by an unmanned aerial vehicle in orchards based on a deep learning method. *Biosystems engineering* **2020**, *194*, 261–272.
15. Koirala, A.; Walsh, K.; Wang, Z.; McCarthy, C. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'. *Precision Agriculture* **2019**, *20*, 1107–1135.
16. Neupane, B.; Horanont, T.; Hung, N.D. Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PloS one* **2019**, *14*, e0223906.
17. Ampatzidis, Y.; Partel, V. UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sensing* **2019**, *11*, 410.
18. Osco, L.P.; De Arruda, M.d.S.; Junior, J.M.; Da Silva, N.B.; Ramos, A.P.M.; Moryia, É.A.S.; Imai, N.N.; Pereira, D.R.; Creste, J.E.; Matsubara, E.T.; others. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, *160*, 97–106.
19. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
20. Yang, M.D.; Tseng, H.H.; Hsu, Y.C.; Tsai, H.P. Semantic segmentation using deep learning with vegetation indices for rice lodging identification in multi-date UAV visible images. *Remote Sensing* **2020**, *12*, 633.
21. Song, Z.; Zhang, Z.; Yang, S.; Ding, D.; Ning, J. Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging. *Computers and Electronics in Agriculture* **2020**, *179*, 105812.

22. Kitano, B.T.; Mendes, C.C.; Geus, A.R.; Oliveira, H.C.; Souza, J.R. Corn plant counting using deep learning and UAV images. *IEEE Geoscience and Remote Sensing Letters* **2019**.
23. Xie, Y.; Xing, F.; Kong, X.; Su, H.; Yang, L. Beyond classification: Structured regression for robust cell detection using convolutional neural network. International conference on medical image computing and computer-assisted intervention. Springer, 2015, pp. 358–365.
24. Seguí, S.; Pujol, O.; Vitria, J. Learning to count with deep object features. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 90–96.
25. Lempitsky, V.; Zisserman, A. Learning to count objects in images. *Advances in neural information processing systems* **2010**, *23*, 1324–1332.
26. Fiaschi, L.; Köthe, U.; Nair, R.; Hamprecht, F.A. Learning to count with regression forest and structured labels. Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). IEEE, 2012, pp. 2685–2688.
27. Jiang, N.; Yu, F. A Cell Counting Framework Based on Random Forest and Density Map. *Applied Sciences* **2020**, *10*, 8346.
28. Jiang, N.; Yu, F. A refinement on detection in cell counting. 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE). IEEE, 2021, pp. 306–309.
29. Xie, W.; Noble, J.A.; Zisserman, A. Microscopy cell counting and detection with fully convolutional regression networks. *Computer methods in biomechanics and biomedical engineering: Imaging & Visualization* **2018**, *6*, 283–292.
30. Paul Cohen, J.; Boucher, G.; Glastonbury, C.A.; Lo, H.Z.; Bengio, Y. Count-ception: Counting by fully convolutional redundant counting. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 18–26.
31. Rad, R.M.; Saeedi, P.; Au, J.; Havelock, J. Cell-net: Embryonic cell counting and centroid localization via residual incremental atrous pyramid and progressive upsampling convolution. *IEEE Access* **2019**, *7*, 81945–81955.
32. He, S.; Minn, K.T.; Solnica-Krezel, L.; Anastasio, M.A.; Li, H. Deeply-supervised density regression for automatic cell counting in microscopy images. *Medical Image Analysis* **2021**, *68*, 101892.
33. Jiang, N.; Yu, F. A Two-Path Network for Cell Counting. *IEEE Access* **2021**, *9*, 70806–70815.
34. Jiang, N.; Yu, F. Cell Counting with Channels Attention. 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP). IEEE, 2020, pp. 494–498.
35. Guo, Y.; Stein, J.; Wu, G.; Krishnamurthy, A. SAU-Net: A Universal Deep Network for Cell Counting. Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 2019, pp. 299–306.
36. Jiang, N.; Yu, F. A Foreground Mask Network for Cell Counting. 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC). IEEE, 2020, pp. 128–132.
37. Arteta, C.; Lempitsky, V.; Zisserman, A. Counting in the wild. European conference on computer vision. Springer, 2016, pp. 483–498.
38. Jiang, N.; Yu, F. Multi-column network for cell counting. *OSA Continuum* **2020**, *3*, 1834–1846.
39. Mehrkanoon, S.; Blaschko, M.; Suykens, J. Shallow and deep models for domain adaptation problems. *Proceedings ESANN 2018* **2018**, pp. 291–299.
40. Long, M.; Cao, Y.; Wang, J.; Jordan, M. Learning transferable features with deep adaptation networks. International conference on machine learning. PMLR, 2015, pp. 97–105.
41. Sun, B.; Saenko, K. Deep coral: Correlation alignment for deep domain adaptation. European conference on computer vision. Springer, 2016, pp. 443–450.
42. Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; Darrell, T. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474* **2014**.
43. Liu, M.Y.; Tuzel, O. Coupled generative adversarial networks. *Advances in neural information processing systems* **2016**, *29*.
44. Ganin, Y.; Lempitsky, V. Unsupervised domain adaptation by backpropagation. International conference on machine learning. PMLR, 2015, pp. 1180–1189.
45. Rodriguez-Vazquez, J.; Alvarez-Fernandez, A.; Molina, M.; Campoy, P. Zenithal isotropic object counting by localization using adversarial training. *Neural Networks* **2022**, *145*, 155–163.

46. Wang, G.; Lopez-Molina, C.; De Baets, B. Automated blob detection using iterative Laplacian of Gaussian filtering and unilateral second-order Gaussian kernels. *Digital Signal Processing* **2020**, *96*, 102592.
47. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.
48. Sampedro, C.; Rodriguez-Vazquez, J.; Rodriguez-Ramos, A.; Carrio, A.; Campoy, P. Deep learning-based system for automatic recognition and diagnosis of electrical insulator strings. *IEEE Access* **2019**, *7*, 101283–101308.
49. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
50. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
51. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* **2015**.
52. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least squares generative adversarial networks. Proceedings of the IEEE international conference on computer vision, 2017, pp. 2794–2802.
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
54. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Wallach, H.; Larochelle, H.; Beygelzimer, A.; d Alche-Buc, F.; Fox, E.; Garnett, R., Eds.; Curran Associates, Inc., 2019; pp. 8024–8035.
55. Falcon, W. PyTorch Lightning. *GitHub. Note: https://github.com/PyTorchLightning/pytorch-lightning* **2019**, *3*.
56. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.
57. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. Randaugment: Practical automated data augmentation with a reduced search space. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020, pp. 702–703.