The International Society of Precision Agriculture presents the
15th International Conference on
Precision Agriculture
26-29 JUNE 2022
Minneapolis Marriott City Center
Minneapolis, Minnesota USA

A Generative Adversarial Network-based method for High Fidelity Synthetic Data Augmentation

Jesse Bier, Srinivas Sridharan, Sudhir Sornapudi,
Qiao Hu, Siva P. Kumpatla
Department of Data Science and Bioinformatics, Corteva Agriscience™
7000 NW 62nd Avenue, Johnston, IA 50131, United States

June 26 – 29, 2022
Minneapolis, Minnesota, United States

Abstract.

Digital Agriculture has led to new phenotyping methods that use artificial intelligence and machine learning solutions on image and video data collected from lab, greenhouse, and field environments. The availability of accurately annotated image and video data remains a bottleneck for developing most machine learning and deep learning models. Typically, deep learning models require thousands of unique samples to accurately learn a given task. However, manual annotation of a large dataset will either take a long time if done by a single annotator or drive-up costs significantly if done by many expert annotators. To provide some relief to the data bottleneck, alternative approaches like automatic augmentation algorithms are needed, however, traditional techniques such as rotation, cropping, resampling, adjusting colors, white balance, and contrast are too simplistic and generic due to which they have significant limitations. For example, if the orientation of the objects in the image is important, then augmentation by rotation would not be an option. Clearly a better approach for data augmentation is required to address these issues and maintain the original properties of the data. Generative Adversarial Network (GAN), a recent development in deep learning, offers a promising solution where two neural networks, one to generate data and the other to detect fake/synthetic data, compete against each other to improve the accuracy of both. A carefully trained GAN model can generate images with high fidelity that can be used as a tool for massive data augmentation. We propose a novel method for data augmentation leveraging NVIDIA's GauGAN model and the Mask R-CNN model to generate unique, synthetic data that are indistinguishable from the real data. A few hundred manually annotated samples are used to train a Mask R-CNN model to generate additional semantic segmentation masks. These generated masks and their corresponding real images are used to train the GauGAN model. Once the GauGAN has been trained, a virtually unlimited quantity of diverse training samples and annotations can be generated. To improve the quality of the segmentation masks the Mask R-CNN model is retrained with the GAN-generated samples. In addition, we also present "Toodle", a Python Dash web application integrated with the GauGAN model for interactive synthetic data generation on lab images collected to study insect damage to maize and soy leaf samples.

Keywords.

Data augmentation, Generative Adversarial Networks, Image Segmentation, GauGAN, Mask R-CNN

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 15th International Conference on Precision Agriculture. EXAMPLE: Lastname, A. B. & Coauthor, C. D. (2018). Title of paper. In Proceedings of the 15th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

1 Introduction

Advances in digital technologies have revolutionized many industries by transforming the product development and business processes while paving the way for disruptive innovations. Digital agriculture is the use of digital tools and technologies for gathering and analyzing spatial and temporal data of crops, soils, and the environment, and leveraging the insights towards implementing appropriate management practices. Such data- and insight-driven management measures are improving farm productivity and profitability.

While the throughput of molecular technologies such as DNA sequencing and genotyping has advanced rapidly, plant phenotyping is the rate-limiting step in performing full-fledged analytics towards crop improvement decisions and for in-field interventions to improve productivity. Images, collected through diverse proximal and remote sensing platforms, and associated analytics are addressing this bottleneck and significantly increasing the spatial coverage and throughput of plant phenotyping (Fahlgren et al., 2015). Powerful tools such as Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) are providing insights into crop growth and health, a prerequisite for site-specific management or precision farming (Benos et al., 2021; Chandra et al., 2020). Satellites, drones, mobile devices, LIDAR, microscopes, hyperspectral/fluorescence imagers are being used to collect images and videos in labs, greenhouses, and field. Computer vision methods are then used to analyze these data to perform image classification, multiple object detection, semantic segmentation. A critical requirement for improving the predictive power of DL models is the availability of a large amount of image data, in the absence of which they could overfit the training data. In situations where insufficient number of images are available, data augmentation is a promising technique to increase the volume of images either by creating slight variations of existing images using simple techniques such as rotating, cropping, flipping, color variations, and noise injection or by *de novo* creation of synthetic images (Shorten & Khoshgoftaar, 2019).

Generative Adversarial Network (GAN), first developed by Ian Goodfellow (Goodfellow et al., 2014), is a class of neural network architecture that performs generative modeling using Convolutional Neural Networks (CNNs). It works by training two competing networks, a generator and a discriminator, where the generator creates new data, and the discriminator classifies those data samples as real or fake. GANs have been used for several image and video applications such as image-to-image translation, photo colorization, generating new human poses, photo blending, super resolution, photo inpainting, video prediction, and 3D object generation (Brownlee, 2019). In addition to this, GANs have also become attractive options for generating synthetic images and, therefore, could be leveraged for data augmentation purposes (Perez & Wang, 2017). Since the images generated by GANs are realistic looking, they do not have the drawbacks of those generated by traditional techniques such as rotation and noise additions, therefore, can improve the generalization ability and predictive power of models like CNNs (Fountas et al., 2020; Marchesi, 2017).

The remainder of this paper is organized as follows: background is presented in section 2, materials and methods are in section 3, results and discussion are presented in section 4, and the paper concludes in section 5 with a summary of the contributions and potential avenues of future research.

2 Background

Artificial intelligence has been used to solve a vast number of problems in diverse domains (LeCun et al., 2015), especially computer vision techniques such as image classification (Simonyan & Zisserman, 2014; Szegedy et al., 2015; He et al., 2016), image segmentation (Chen et al., 2014; Long et al., 2015), and object detection (Girshick et al., 2014; Ren et al., 2015; W. Liu et al., 2016; Redmon et al., 2016) have been used to analyze images and videos. DL models have been shown to perform significantly better than traditional computer vision and machine learning algorithms (Krizhevsky et al., 2012). Due to their accuracy and efficiency, these models found applications in several key areas including agriculture and were even shown to surpass certain aspects of human vision (He et al., 2015; Russakovsky et al., 2015).

In agriculture, DL models are trained on image and video data from satellites, UAVs, farm rovers, smartphones, and microscopes, to identify objects of interest at different scales. Example applications include land cover classification (Rebetez et al., 2016; Kussul et al., 2017; Lu et al., 2017), crop phenology (Khaki et al., 2022; Z. Lin & Guo, 2020; Zendler et al., 2021), crop or plant detection (Sa et al., 2016; Bargoti & Underwood, 2017), weed detection (Milioto et al., 2017; Potena et al., 2016), and disease identification (Mohanty et al., 2016; B. Liu et al., 2017; Ferentinos, 2018). Among DL models, Mask R-CNN (He et al., 2017) is extensively used to detect and count objects in images and videos. It has been used in agriculture for plant or stock counting (Machefer et al., 2020; Xu et al., 2020), crop detection (S. Wang et al., 2021), fruit detection ((Yu et al., 2019)), etc.

Although DL models can achieve outstanding performance, they still require voluminous data and ground-truth to learn key features and generalize better on unseen images. While abundant raw image data are available for several cases in agriculture, the cost and time required to annotate these data is a major challenge. This is due to the shortage in the number of domain experts needed, or due to the cumbersome nature of capturing ground-truth information. Hence there is a need for a preprocessing step to augment data and corresponding annotations to train DL models with high performance and with reasonable cost and time. GANs have been explored to solve this issue by training a model that can automatically generate synthetic data and associated annotations.

GAN networks were first developed by Ian Goodfellow (Goodfellow et al., 2014). Since the development of original algorithm by Goodfellow et al. several GAN networks have come into existence such as pix2pix (Isola et al., 2017), CycleGAN (Zhu et al., 2017), StyleGAN (Karras et al., 2019), and GauGAN (Park et al., 2019). More recently, data augmentation via synthetic data generation has seen numerous approaches with GAN-based architectures. These techniques range from generation of novel samples from the learned distribution implied by data (DCGAN; (Radford et al., 2016), image-to-image translation (pix2pix), and hybrid methods such as SimGAN which combine 3D rendering with GAN (Shrivastava et al., 2017). The hybrid technique SimGAN by Apple applies realistic textures to 3D rendered eyes for training the iPhone X face unlock. Using this approach, they were able to generate realistic training data with perfect annotations. Since annotation is typically a large expense in both time and money, the ability to automatically generate annotations is crucial for any data augmentation task. Data augmentation with only synthetic data (Jahanian et al., 2022) is a very recent development by MIT. In this paper, we employ both real and synthetic data for training our downstream model.

In this paper, we used the GauGAN model to generate synthetic images for data augmentation. These data were used to train a Mask R-CNN model to accurately generate segmentation maps. The main contribution of the paper is the development of a GAN-based method combined with a procedural algorithm to generate high fidelity synthetic image data for augmentation. The next section discusses in detail the materials and methods used in the project.

3 Materials and Methods

In this paper, we propose a novel framework (see Figure 1) for high fidelity synthetic data augmentation using generative adversarial networks. This framework consists of a Mask R-CNN (He et al., 2017) segmentation model trained on exceedingly small, annotated leaf tissue dataset. A GauGAN model is trained with Mask R-CNN data to generate photo-realistic leaf disc images. The images generated by the GAN model were then used to re-train the Mask R-CNN model and improve its accuracy compared to traditional or no augmentation. Section 3.1 provides information on the image and ground-truth datasets, Section 3.2 explains the Mask R-CNN segmentation model, Section 3.3 describes the GauGAN training process, Section 3.4 describes the retraining of Mask R-CNN with augmentation, and finally Section 3.5 presents the *"Toodle",* a Python web application that converts doodles into realistic leaf disc images.

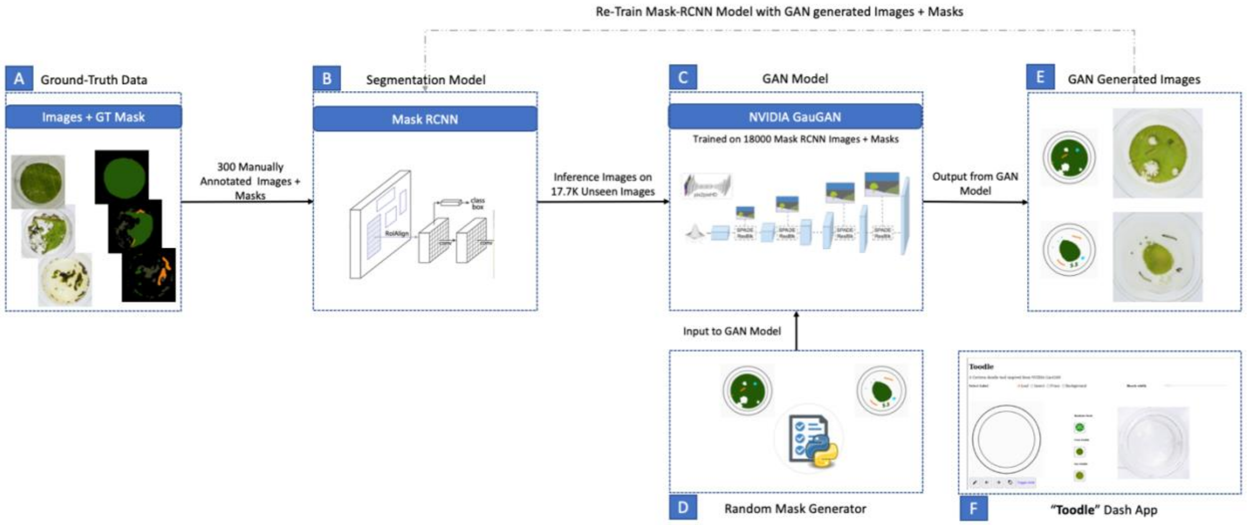


Figure 1. Overall workflow for a GAN-based method for high fidelity synthetic data augmentation. Panel (A) shows sample soy and corn images and the corresponding human-annotated masks. The experts annotated 300 images (150 each for soy and corn) to identify leaf tissue, insect, and frass at a pixel level. These data were used to train an image segmentation model such as Mask R-CNN as shown in panel (B). The Mask R-CNN was used to generate masks for an additional 17700 images. The 18000 images were then trained using NVIDIA GauGAN as shown in panel (C). The GauGAN model was evaluated with a random mask generator that procedurally created masks that resembled the ground-truth annotations (D). The GAN generated images for a given mask are shown in panel (E). We also developed "Toodle" a python dash web app tool to draw custom user generated masks as shown in panel, however the app was not used for training as shown in panel (F).

3.1 Datasets

Plants containing insect resistant traits or crop protection insecticides are evaluated using high-throughput insect feeding assays in labs. Multiple corn and soy circular leaf tissues are infested with fall armyworm, corn earworm, European corn borer, soybean looper, and velvet bean caterpillar larvae using plate-based assays. These plates are later imaged to evaluate the effectiveness of the corresponding trait or insecticide compared to controls. We used computer vision and deep learning models to automatically quantify damage from these images. To accurately estimate damage, a Mask R-CNN model was trained to identify pixels that are leaf, insect, frass, and background. Figure 2 below shows a sample image of corn and soy leaf tissue. The segmentation model must accurately assess the tissue damage for estimating the impact of the trait or insecticide.

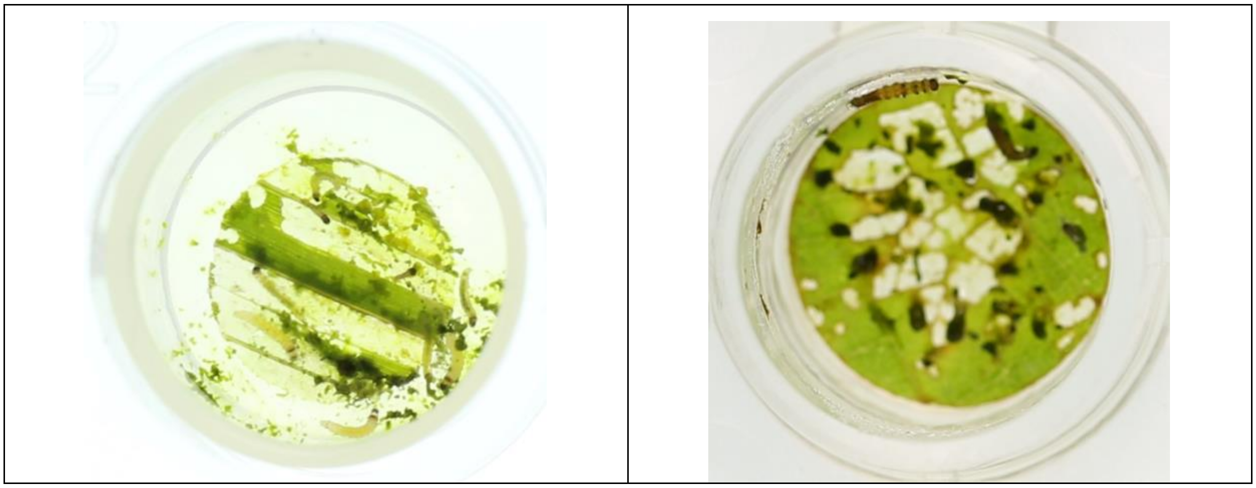


Figure 2 – Sample leaf tissue images used to train the Mask-RCNN and GauGAN model. Image on left contains corn leaf tissue infested with European corn borer larvae. Image on right contains soy leaf tissue infested with corn earworm.

We used 200 soy and corn images each with varied amounts of tissue damage in them. Experts manually annotated the images to obtain pixel-level ground-truth information. The 300 annotated images were used to train the Mask R-CNN segmentation model and 100 images were set aside as test dataset. We then extracted some statistics from the annotated images that were used to procedurally generate random mask images for the GauGAN model. Tables 1 and 2 below show the image statistics extracted from these masks.

Table 1: Image statistics for leaf tissue, insect, and frass for corn assay

Image Stats	Leaf	Insect	Frass
Mean Count	4	4	8
(Min,Max) Count	(1, 38)	(1, 5)	(1, 66)
Mean Pixel Area (%)	24	0.24	0.11
(Min, Max) Pixel Area (%)	(22, 27)	(0.01, 0.97)	(0, 2.34)

Table 2: Image statistics for leaf tissue, insect, and frass for soy assay

Image Stats	Leaf	Insect	Frass
Mean Count	16	1	61
(Min,Max) Count	(1, 88)	(1, 3)	(1, 185)
Mean Pixel Area (%)	38	0.82	0.05
(Min, Max) Pixel Area (%)	(20, 50)	(0.01,3.14)	(0, 0.5)

In the tables above pixel area is defined as the ratio of total number of pixels for each class (leaf, insect, frass) to the total number of image pixels. Another 17,700 unannotated leaf disc images were used to generate segmentation maps using the Mask R-CNN model. The GauGAN model was trained using these images and their corresponding masks.

3.2 Mask R-CNN

Mask R-CNN, a deep learning model for instance segmentation, aims to accurately detect objects in an image and simultaneously generate masks for each instance. It is basically an extension of Faster R-CNN where a mask predictor branch is added in parallel to a bounding box predictor.

Mask R-CNN consists of two stages. Stage I is a Feature Pyramid Network (FPN) (T.-Y. Lin et al., 2017) along with Region Proposal Network (RPN) for identifying possible region of interest (RoI) or anchors in an image. FPN is built in conjunction with a backbone feature extractor model (ResNet-50). The hierarchical features at different scales from the intermediate layers of ResNet-50 are fed to the FPN to improve the generic feature extraction. Stage II extracts small feature maps using RoIAlign, in contrast to RoIPool in Fast R-CNN, from each candidate RoI and performs classification, bounding box regression, and pixel-level segmentation for each RoI. The candidate RoIs generated by RPN are usually filtered by non-maximum suppression algorithm to eliminate any redundant anchors and find the coordinates of an optimum object detection bounding box.

The ResNet-50 model in stage I is pre-loaded with ImageNet weights and then finetuned with the leaf disc dataset. Fine-tuning is considered as an efficient scheme while training deep CNNs with sparse dataset. The Mask R-CNN model was trained for 200 epochs. The initial learning rate was set to 0.01 and the learning rate decay was chosen to reduce by 0.1 for every 30 epochs. The loss for each sampled RoI is a summation of classification, bounding box regression, and mask losses. We used stochastic gradient descent with momentum parameter to minimize the loss function. The inference time per image was 0.21 seconds on a NVIDIA Tesla T4 GPU.

3.3 GauGAN Training

GauGAN (Park et al., 2019) is a model developed by NVIDIA which is named after post-impressionist painter Paul Gauguin. The NVIDIA GauGAN model generates a photo-realistic image of a user drawn doodle. GauGAN is built on the NVIDIA's pix2pixHD model (T.-C. Wang et al., 2018), which itself is an improvement on the original pix2pix model to enable higher resolution images. It improves on the efficiency of training a pix2pix model by incorporating a novel variant of normalization called SPADE layers. These layers maintain semantic information between convolutional layers and normalization operations. GauGAN is also designed to use an optional variational auto encoder (VAE) network in addition to the semantic mask as an input to the generator. We chose not to employ the VAE in our work because the number of classes is low and very specific to type of object we wished to render.

The GauGAN model was trained on approx. 9000 annotated images generated by Mask R-CNN that were split evenly between corn and soy. We utilized the training script which came with the GauGAN repository without any modification. The model was initially trained for 50 epochs. Since the quality of the corn leaf texture was insufficient, we continued training for 25 more epochs with an additional 9000 images (split evenly between corn and soy). Further training improved the quality of the corn leaf texture rendered by the model. The first 50 epochs were trained for roughly 48 hours on an NVIDIA A100 GPU. The remaining 25 epochs were trained for roughly 48 hours on an NVIDIA V100 GPU. We assessed the quality of the model based solely on visual inspection of the generated output.

3.4 Mask R-CNN Retraining

The Mask R-CNN model was initially trained with 150 images each for corn and soy. To evaluate the GauGAN model we re-trained the Mask R-CNN with the masks and GAN generated photo-realistic leaf disc images. We compared Mask R-CNN model trained only on the original data with Mask R-CNN model trained on traditional and GAN-based augmented images.

3.4.1 Traditional Augmentation

Traditional augmentation approaches are easy and cost less to generate transformed views of an image. These usually include flipping, rotating, cropping, and tuning images (Buslaev et al., 2020; Papakipos & Bitton, 2022). We generated 34 transformed views of each image to extrapolate the 150 original images to generate almost 5000 new images per assay. The transformations were chosen such that the resulting image retains the key characteristics of the original image.

The leaf disc dataset contains a leaf tissue cut in a disc shape and placed at center of the well in each plate. Insects are confined to this well to feed on the leaf, and they leave waste, also known as frass. The damage can be on any part of the leaf and of any shape. To account for this, the images were flipped along vertical and horizontal axes and cropped randomly with a given height and width. The images were rotated at various angles {45^{o}, 60^{o}, 90^{o}, 135^{o}, 180^{o}, 225^{o}, 270^{o}, 315^{o}} with zero padding and cropping. The images were randomly tuned for varied brightness with a factor ranging [0.5, 2] and varied contrast with factor [0.5, 0.7]. The image signal was processed with a high pass filter to induce noise mean of {12, 25, 50} and standard deviation of {8, 17, 35}, respectively; and with low pass filter to apply median and uniform blur with kernel sizes {5, 15}. Finally, Gaussian blur was also applied with sigma ranging [1, 15]. All the values were chosen randomly, and ranges were chosen to limit the variations.

3.4.2 GAN-based Augmentation

We utilized two methods to generate masks as input for GauGAN. We generated 5000 images per assay with each method. The first method augments labeled masks via procedural generation (see Figure 3). We used statistics calculated from the expert annotated mask images (see Table 1 and Table 2) to parameterize the generation. We also used other features not listed in the tables such as orientation, location, and the length of objects. All these features are assumed to be of Gaussian distribution to generate shapes that closely resemble the objects in real masks.

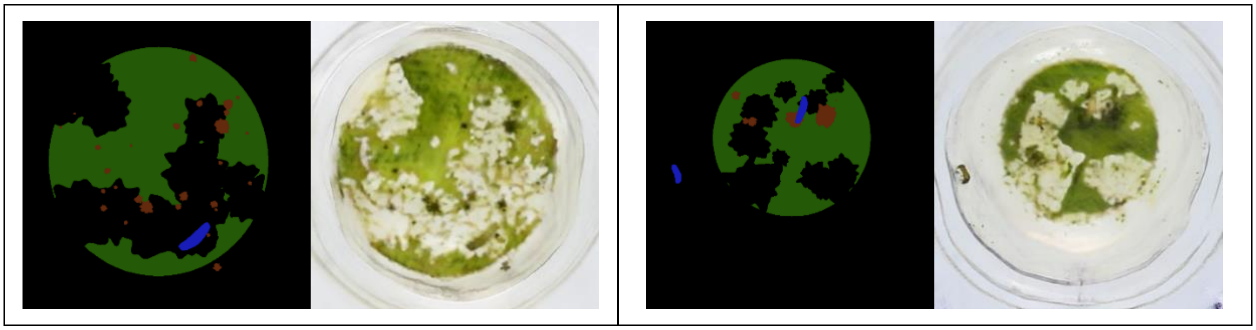


Figure 3. An example of procedurally simulated masks and corresponding GauGAN-generated soy (left) and corn (right) leaf disc images.

The second method extracts objects from labeled data and randomly re-arrange them to simulate masks. This ensures that the statistical properties such as size, shape, and scale are not adversely altered. The only change to the labeled object is its location and/or orientation. To achieve this each labeled object is individually rotated in a random fashion about the center of the image by an angle between 0 and 359 degrees. A morphological operation such as dilation, erosion, opening, or closing is randomly applied to this object. This ensures we create minute variations of the object in each mask image. The leaf tissue object may contain disconnected regions; hence all regions are rotated by the same angle to preserve the shape of the tissue object. The tissue regions were then rendered to the semantic mask to ensure damage consistency. The masks and their corresponding GAN generated leaf disc images are shown in Figure 4.

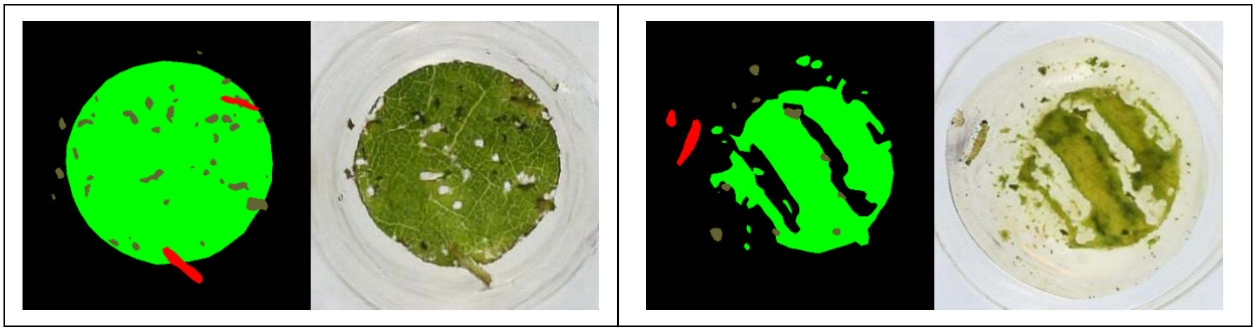
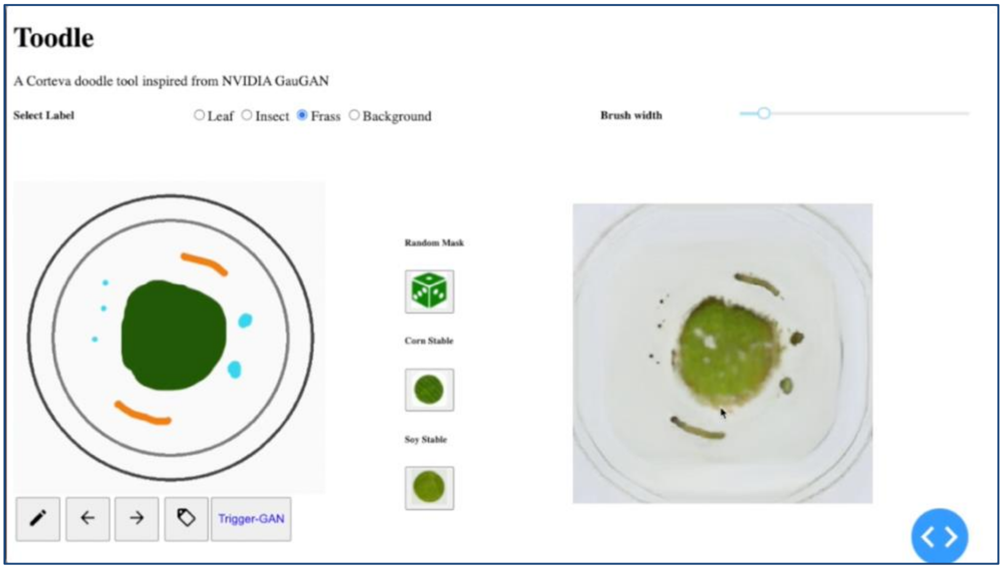


Figure 4. An example of randomly re-arranged masks and corresponding GauGAN-generated soy (left) and corn (right) leaf disc images.

3.5 Toodle: Python-Dash Web Tool

Finally, we developed a web application using Python Dash that can convert doodles created on a canvas into a photo-realistic leaf disc image. Figure 5 below shows "***Toodle",*** a Corteva doodle tool, that is inspired from NVIDIA GauGAN. Using this tool, the user can draw leaf, insect, and frass on the canvas by selecting the appropriate label and brush size. The user can then press the "Trigger GAN" button to generate photo-realistic leaf disc image that will be displayed on the right. The user can toggle between corn and soy leaf tissues by selecting the appropriate buttons in the middle. The tool also provides an option to generate random mask images procedurally and display their corresponding GAN generated leaf disc images.



Toodle

A Corteva doodle tool inspired from NVIDIA GauGAN

Select Label ○ Leaf ○ Insect ● Frass ○ Background

Brush width

Figure 5 – Screenshot image of Toodle web application tool that can convert doodles into a photo-realistic leaf disc image. The user can draw leaf tissue, insect, and frass on the canvas (left), and the corresponding GauGAN generated photo-realistic image can be viewed (right).

4 Results and Discussion

We performed a detailed comparison among three Mask R-CNN models each trained on the original images and their corresponding masks with no augmentation, traditional augmentation, and GAN-based augmentation. In this section, we report the results in standard metrics defined in COCO challenge (T.-Y. Lin et al., 2015) for evaluation of object detection and instance segmentation masks. These include mean Average Precision (mAP) and mean Average Recall (mAR). The mAP metrics are tabulated as average precision, with mean over ten 0.50:0.95 Intersection-Over-Union (IoU) thresholds, AP_{50} (IoU=0.50), AP_{75} (IoU=0.75) and AP_S, AP_M, AP_L (mAP at different small, medium, and large scales). The mAR metrics have average recall, AR_1, AR_{10}, AR_{100} (based on number of detections per image) and AR_S, AR_M, AR_L (mAR at different small, medium, and large scales).

We evaluated these three models on two test datasets. First, we analyzed the Mask-RCNN models using 50 original unseen test images and then, we added an additional 50 unseen GAN-generated images to the former test dataset (total of 100 unseen test images). The Mask R-CNN models are assay specific, that is, separate models are trained for Soy and Corn stable images because the leaf texture and the insects used are significantly different.

The results from the first test dataset containing the fifty original images are shown in Tables 3-6. We observed that the Mask R-CNN from traditional augmentation performed better than the GAN-based augmentation and no augmentation models. This is because the images represented by the traditional augmentation belong to the same distribution as the original training dataset. This indicates that the images generated by the GauGAN model are not photo-realistic enough as compared to the original images.

Table 3. Soy assay bounding box AP and AR

Augmentation	Average Precision						Average Recall					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
None	25.8	51.3	22.7	23.4	24.6	63.6	15.2	22.7	33.2	30.6	34.2	65.3
Traditional	30.8	55.6	30.7	27.3	30.3	81.6	19.6	29.2	39.3	35.8	40.2	82.5
GAN	28.5	53.3	26.0	25.3	26.6	72.4	18.6	28.1	37.2	34.9	37.6	72.9

Table 4. Soy assay instance segmentation mask AP and AR

Augmentation	Average Precision						Average Recall					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
None	22.2	49.5	18.0	17.2	30.7	45.0	13.1	20.0	29.2	25.8	33.2	45.4
Traditional	25.7	54.1	22.4	19.6	33.8	65.4	16.2	25.0	34.0	30.2	36.7	66.0
GAN	21.9	51.5	16.3	16.2	31.6	59.9	14.5	22.5	30.5	26.7	33.6	60.2

Table 5. Corn assay bounding box AP and AR

Augmentation	Average Precision						Average Recall					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
None	36.5	59.8	38.4	29.6	24.4	74.5	17.2	41.1	43.1	37.1	27.2	80.9
Traditional	39.3	66.2	41.5	30.9	37.3	80.7	16.9	43.1	47.9	41.2	41.1	83.4
GAN	35.3	56.4	37.4	26.0	26.7	83.5	17.6	39.3	41.1	33.1	28.8	84.5

Table 6. Corn assay instance segmentation mask AP and AR

Augmentation	Average Precision						Average Recall					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
None	28.5	59.7	21.3	22.8	20.1	60.7	14.4	33.8	35.3	29.7	21.5	72.0
Traditional	30.7	66.2	21.9	23.3	31.5	69.9	14.2	34.8	38.8	32.7	33.7	74.5
GAN	29.2	57.1	24.0	21.7	20.6	70.8	15.0	33.5	35.2	28.9	22.2	72.0

These results were later confirmed on the second test dataset containing fifty original and fifty GAN-generated images as shown in Tables 7-10. The Mask R-CNN model trained on GAN augmentation outperforms the traditional and no augmentation models. This indicates that the GAN augmentation model's data distribution does not precisely resemble with that of the real data. A potential reason for this observation is that the GauGAN model was trained for a small number of epochs with weakly labelled noisy data. This can be avoided by training GauGAN for more iterations on a larger dataset with accurate ground truth masks to generate more photo-realistic leaf disc images. However, the GAN-generated images are visually better-looking and closely resemble the original corn and soy leaf tissue images.

Table 7. Soy assay bounding box AP and AR

Augmentation	Average Precision						Average Recall					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
None	20.9	43.6	16.9	18.4	23.4	53.9	11.4	20.5	30.2	26.3	37.0	58.9
Traditional	24.1	47.2	21.4	19.8	30.9	69.3	15.5	24.4	33.2	27.8	41.9	73.9
GAN	38.0	62.8	38.7	33.2	56.2	71.1	22.7	35.3	44.0	39.3	61.2	72.3

Table 8. Soy assay instance segmentation mask AP and AR

Augmentation	Average Precision						Average Recall					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
None	18.2	42.0	14.0	14.4	29.0	35.6	10.1	18.3	27.0	23.5	36.5	37.7
Traditional	20.9	45.5	17.7	15.2	34.7	56.7	13.8	21.9	29.8	25.2	40.3	59.2
GAN	31.2	61.2	31.1	24.8	54.3	64.8	19.6	30.2	37.9	33.1	55.5	65.6

Table 9. Corn assay bounding box AP and AR

Augmentation	Average Precision						Average Recall					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
None	28.4	47.5	30.5	20.6	25.4	68.0	16.1	34.5	35.5	27.9	31.4	75.8
Traditional	29.9	49.4	31.8	21.0	30.7	77.7	16.7	35.7	38.1	29.7	39.5	82.3
GAN	44.8	68.0	48.0	34.1	48.7	88.5	21.4	48.0	49.2	39.5	51.8	89.5

Table 10. Corn assay instance segmentation mask AP and AR

Augmentation	Average Precision						Average Recall					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
None	21.8	46.8	16.3	16.6	21.9	44.9	13.0	28.8	29.4	23.6	26.2	57.3
Traditional	22.7	48.9	17.0	16.2	27.5	54.0	13.5	29.1	31.1	24.5	34.2	63.3
GAN	37.4	68.4	36.4	28.6	44.1	70.2	18.2	41.1	42.3	34.3	46.4	73.7

Figures 6 and 7 show a sample leaf tissue image (left-right) that overlays the segmentation mask with expert annotations, a mask generated by the Mask R-CNN model trained on dataset without any augmentation, a mask generated by the Mask R-CNN model trained on dataset with traditional augmentation, and a mask generated by the Mask R-CNN model trained on dataset with GAN-based augmentation. The GAN-based and traditional augmentation models generate masks that look visually similar. The regions where they differ happen primarily around the pixels of damaged leaf tissue. This issue could be resolved by re-training the GAN model with more leaf tissue images with damage.

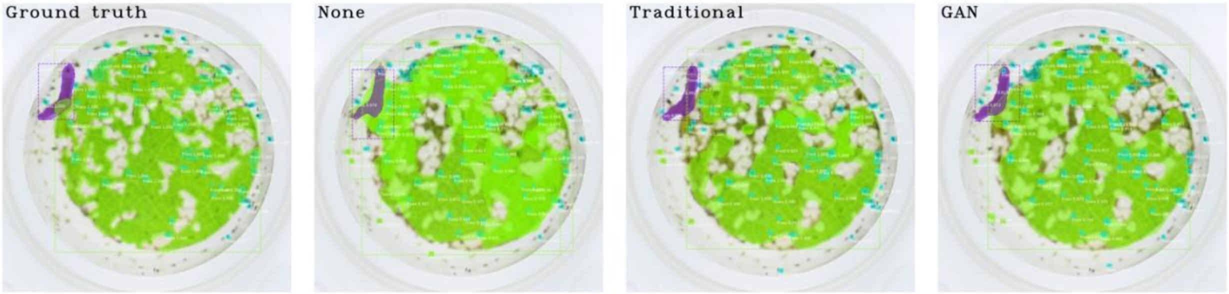


Figure 6. A sample soy leaf tissue image showing (left-right) the expert annotated ground-truth mask, mask generated by Mask-RCNN model trained with no augmentation, with traditional augmentation, with GAN-augmentation

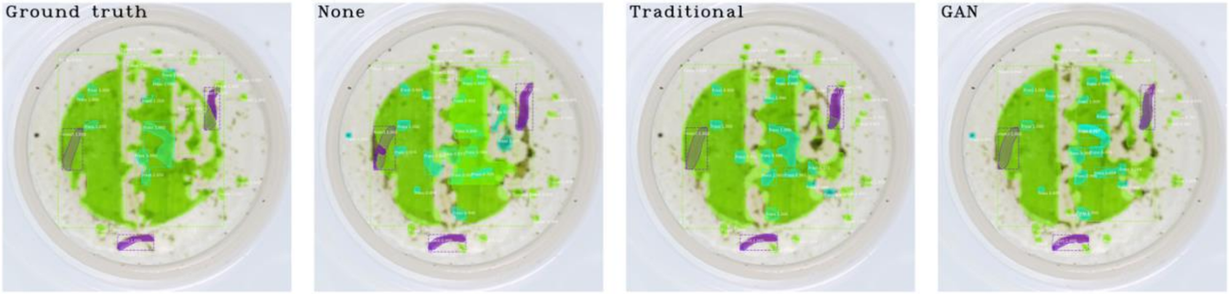


Figure 7. A sample corn leaf tissue image showing (left-right) the expert annotated ground-truth mask, mask generated by Mask-RCNN model trained with no augmentation, with traditional augmentation, with GAN-augmentation

Our results indicate that it is possible to train a Mask R-CNN model with only 150 annotated images (per assay) and train a GauGAN model to generate new photo-realistic images. These data are then used to re-train the Mask R-CNN segmentation model using GAN-based augmentation to accurately assess leaf tissue damage.

5 Conclusion and Future Work

In this paper, we proposed a novel workflow to improve the performance of a Mask R-CNN segmentation model trained on a small, annotated leaf tissue dataset by performing data augmentation using a generative adversarial network (GauGAN). This is a four-fold approach. First, we train a Mask R-CNN with data as small as 150 images. Second, we predict segmentation masks using trained Mask R-CNN on 9000 unseen images. Third, we train a GauGAN model with these noisy predicted masks to generate photo-realistic leaf disc images. Fourth, we retrain the Mask R-CNN on new expanded dataset which contains the leaf disc dataset and generated leaf disc images. While our accuracy on our validation set didn't show improvement, accuracy on the validation set plus unseen generated images showed improvement. This suggests with improvement to the GAN training, our method could substantially increase the accuracy and recall of downstream models by a significant margin.

6 Acknowledgements

We would like to thank Gilda Rauscher, Greg Yeckel, Airene Millsap, Sefriyani Johan, and Evy Santiago Martinez for providing data and scientific expertise on corn and soy leaf disc assays, Andrea Arias, Manny Ruidiaz, Jochen Scheel, Steve Callistein, Nicole Janovick and Janae Lehman Bell for review and approval of the manuscript. We extend a special thanks to Charles Hitt from NVIDIA for helpful discussions and approval to use GauGAN model for this project.

7 References

Bargoti, S., & Underwood, J. (2017). Deep fruit detection in orchards. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 3626–3633.

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors*, *21*(11), 3758. https://doi.org/10.3390/s21113758

Brownlee, J. (2019). *Generative adversarial networks with python: Deep learning generative models for image synthesis and image translation*. Machine Learning Mastery.

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and Flexible Image Augmentations. *Information*, *11*(2), 125. https://doi.org/10.3390/info11020125

Chandra, A. L., Desai, S. V., Guo, W., & Balasubramanian, V. N. (2020). Computer Vision with Deep Learning for Plant Phenotyping in Agriculture: A Survey. *Advanced Computing and Communications*. https://doi.org/10.34048/ACC.2020.1.F1

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. *ArXiv Preprint ArXiv:1412.7062*.

Fahlgren, N., Gehan, M., & Baxter, I. (2015). Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology*, *24*, 93–99.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311–318.

Fountas, S., Espejo-García, B., Kasimati, A., Mylonas, N., & Darra, N. (2020). The Future of Digital Agriculture: Technologies and Opportunities. *IT Professional*. https://doi.org/10.1109/MITP.2019.2963412

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, *27*. https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). *Image-To-Image Translation With Conditional Adversarial Networks*. 1125–1134. https://openaccess.thecvf.com/content_cvpr_2017/html/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.html

Jahanian, A., Puig, X., Tian, Y., & Isola, P. (2022). *Generative Models as a Data Source for Multiview Representation Learning*. 22.

Karras, T., Laine, S., & Aila, T. (2019). *A Style-Based Generator Architecture for Generative Adversarial Networks*. 4401–4410. https://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html

Khaki, S., Safaei, N., Pham, H., & Wang, L. (2022). WheatNet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. *Neurocomputing*, *489*, 78–89. https://doi.org/10.1016/j.neucom.2022.03.017

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*.

Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), 778–782.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. *ArXiv:1612.03144 [Cs]*. http://arxiv.org/abs/1612.03144

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). Microsoft COCO: Common Objects in Context. *ArXiv:1405.0312 [Cs]*. http://arxiv.org/abs/1405.0312

Lin, Z., & Guo, W. (2020). Sorghum Panicle Detection and Counting Using Unmanned Aerial System Images and Deep Learning. *Frontiers in Plant Science*, *11*, 534853. https://doi.org/10.3389/fpls.2020.534853

Liu, B., Zhang, Y., He, D., & Li, Y. (2017). Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry*, *10*(1), 11.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. *European Conference on Computer Vision*, 21–37.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Lu, H., Fu, X., Liu, C., Li, L., He, Y., & Li, N. (2017). Cultivated land information extraction in UAV imagery based on deep convolutional neural network and transfer learning. *Journal of Mountain Science*, *14*(4), 731–741.

Machefer, M., Lemarchand, F., Bonnefond, V., Hitchins, A., & Sidiropoulos, P. (2020). Mask R-CNN refitting strategy for plant counting and sizing in uav imagery. *Remote Sensing*, *12*(18), 3015.

Marchesi, M. (2017). Megapixel Size Image Creation using Generative Adversarial Networks. *ArXiv:1706.00082 [Cs]*. http://arxiv.org/abs/1706.00082

Milioto, A., Lottes, P., & Stachniss, C. (2017). REAL-TIME BLOB-WISE SUGAR BEETS VS WEEDS CLASSIFICATION FOR MONITORING FIELDS USING CONVOLUTIONAL NEURAL NETWORKS. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *4*.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, *7*, 1419.

Papakipos, Z., & Bitton, J. (2022). *AugLy: Data Augmentations for Robustness*. https://doi.org/10.48550/ARXIV.2201.06494

Park, T., Liu, M.-Y., Wang, T.-C., & Zhu, J.-Y. (2019). *Semantic Image Synthesis With Spatially-Adaptive Normalization*. 2337–2346. https://openaccess.thecvf.com/content_CVPR_2019/html/Park_Semantic_Image_Synthesis_With_Spatially-Adaptive_Normalization_CVPR_2019_paper.html

Perez, L., & Wang, J. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *ArXiv:1712.04621 [Cs]*. http://arxiv.org/abs/1712.04621

Potena, C., Nardi, D., & Pretto, A. (2016). Fast and accurate crop and weed identification with summarized train sets for precision agriculture. *International Conference on Intelligent Autonomous Systems*, 105–121.

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *ArXiv:1511.06434 [Cs]*. http://arxiv.org/abs/1511.06434

Rebetez, J., Satizábal, H. F., Mota, M., Noll, D., Büchi, L., Wendling, M., Cannelle, B., Perez-Uribe, A., & Burgos, S. (2016). Augmenting a convolutional neural network with local histograms-A case study in crop classification from high-resolution UAV imagery. *ESANN*.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *28*.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., & Bernstein, M. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors*, *16*(8), 1222. https://doi.org/10.3390/s16081222

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., & Webb, R. (2017). *Learning From Simulated and Unsupervised Images Through Adversarial Training*. 2107–2116. https://openaccess.thecvf.com/content_cvpr_2017/html/Shrivastava_Learning_From_Simulated_CVPR_2017_paper.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015).

Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9.

Wang, S., Sun, G., Zheng, B., & Du, Y. (2021). A Crop Image Segmentation and Extraction Algorithm Based on Mask RCNN. *Entropy*, *23*(9), 1160.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). *High-Resolution Image Synthesis and Semantic Manipulation With Conditional GANs*. 8798–8807. https://openaccess.thecvf.com/content_cvpr_2018/html/Wang_High-Resolution_Image_Synthesis_CVPR_2018_paper.html

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., & Schneider, D. (2020). Automated cattle counting using Mask R-CNN in quadcopter vision system. *Computers and Electronics in Agriculture*, *171*, 105300.

Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, *163*, 104846.

Zendler, D., Malagol, N., Schwandner, A., Töpfer, R., Hausmann, L., & Zyprian, E. (2021). High-throughput phenotyping of leaf discs infected with grapevine downy mildew using shallow convolutional neural networks. *Agronomy*, *11*(9), 1768.

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). *Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks*. 2223–2232. https://openaccess.thecvf.com/content_iccv_2017/html/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.html