Spatio-temporal generation of morphological Plant features for yield prediction before harvest from Visual Image input using Progressively Growing GANs

Dhruv Sheth[1,1]

[1]EdgeImpulse Inc.

November 30, 2022

Abstract

Recent Innovations in Precision Agriculture (PA) are driven by Computer Vision and Data Processing systems to quantify plant parameters. Quantitative analysis of Plant Phenotyping in PA and monitoring morphological traits is a protracting process, precluding the objective and phenotyping pipeline. Greenhouses growing Genetically Modified (GM) crops need to be maintained at constant environmental and simulated conditions. Multiple parameters have to be controlled and regulated inside a greenhouse for effective growth of crops and yield maximisation. Not at all times are these factors derived and so, yield maximisation in greenhouse is an experimental approach to new varieties. For deduced environmental parameters and conditions for certain crops, few other biotic and abiotic factors can hinder or affect growth in certain ways that are not always factored in during calculating parameters conducive for plant growth. Such factors may not always be affecting parametral calculations, but transpose visual cues on plant growth environment such as spectral change in soil values, or minute changes like leaf reflectance or visible changes in plant stimuli to biotic factors. Plant growth is inclusive of multiple environmental variables, and yield maximisation approaches are experimental to finding the optimum derived value for these variables. Computer Vision provides a catalytic approach to predicting optimum parameters for yield maximization in phenomics. Computer Vision and Generative Adversarial Networks (GAN)'s offer a catalytic approach to the time-consuming process, providing a solution to the phenotyping bottleneck. This research proposes a concept of curating data of plant growth over time to predict conditional growth and responsive stimuli of the plant under different situations and how this can affect crop yield. The method proposed here is a non-invasive approach to the existing destructive biomass estimation methods and Frameworks. This methodology of the research focuses on utilizing image parameters modelled using a time series Progressively Growing Generative Adversarial Networks PGGAN to map plant growth patterns and progressive variance in biomass of plant in the Spatio-Temporal Domain. These Generative networks evaluate and predict based on merely raw pixel input excluding dependence on further constraints, feature vectors or parameters influencing data.

Spatio-temporal generation of morphological Plant features for yield prediction before harvest from Visual Image input using *Progressively Growing GANs*

February 8, 2022

1 Abstract

Recent Innovations in Precision Agriculture (PA) are driven by Computer Vision and Data Processing systems to quantify plant parameters. Quantitative analysis of Plant Phenotyping in PA and monitoring morphological traits is a protracting process, precluding the objective and phenotyping pipeline. Greenhouses growing Genetically Modified (GM) crops need to be maintained at constant environmental and simulated conditions. Multiple parameters have to be controlled and regulated inside a greenhouse for effective growth of crops and yield maximisation. Not at all times are these factors derived and so, yield maximisation in greenhouse is an experimental approach to new varieties. For deduced environmental parameters and conditions for certain crops, few other biotic and abiotic factors can hinder or affect growth in certain ways that are not always factored in during calculating parameters conducive for plant growth. Such factors may not always be affecting parametral calculations, but transpose visual cues on plant growth environment such as spectral change in soil values, or minute changes like leaf reflectance or visible changes in plant stimuli to biotic factors. Plant growth is inclusive of multiple environmental variables, and yield maximisation approaches are experimental to finding the optimum derived value for these variables. Computer Vision provides a catalytic approach to predicting optimum parameters for yield maximization in phenomics. Computer Vision and Generative Adversarial Networks (GAN)'s offer a catalytic approach to the time-consuming process, providing a solution to the phenotyping bottleneck. This research proposes a concept of curating data of plant growth over time to predict conditional growth and responsive stimuli of the plant under different situations and how this can affect crop yield. The method proposed here is a non-invasive approach to the existing destructive biomass estimation methods and Frameworks. This methodology of the research focuses on utilizing image parameters modelled using a time series *Progressively Growing Generative Adversarial Networks PGGAN* to map plant growth patterns and progressive variance in biomass of plant in the Spatio-Temporal Domain. These Generative networks evaluate and predict based on merely raw pixel input excluding dependence on further constraints, feature vectors or parameters influencing data.

1.1 Fast-Forward Session Slide

Dhruv Sheth
Research Intern - Harvard Medical School

Introduction:

Greenhouses and Vertical farms need to be maintained at constant environmental conditions for optimum plant growth and calculating these environmental variables is a slow process.

Research Aim:

To develop an automated solution to generate visual future plant growth frames months before harvest using just RGB image input and Generative Adversarial Networks (GAN).

Result:

Developed a GAN model on *Beta vulgaris* and *Arabidopsis thaliana* and showed 98% overall accuracy in predicting plant growth and suitable environment providing maximum plant yield in Greenhouses and Vertical Farms.
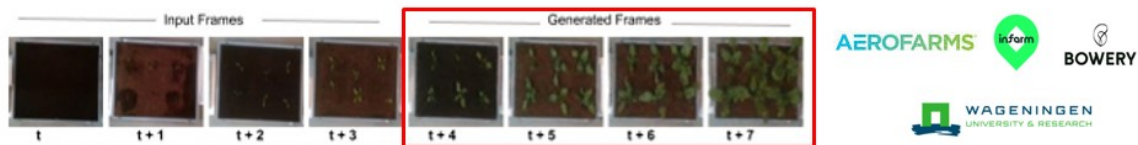
For an Input of 4 images in the *Beta vulgaris* model, it was able to generate 4 frames into future.

Email: **dhruvsheth.linkit@gmail.com**   Twitter: **@dhruvsheth_**

Spatio-temporal generation of morphological Plant features for yield prediction before harvest from Visual Image input using *Progressively Growing GANs*

DHRUV. M. SHETH[1,2,3]

[1] *Udayachal High School, Secondary Education, Grad - 2021*
[2] *Pace Junior College, Senior Secondary Education, Expected Grad - 2023*
[3] *Embedded Machine Learning Ambassador - EdgeImpulse Inc. San Jose, California*
[*] *Corresponding author: dhruvsheth.linkit@gmail.com*

Compiled November 7, 2021

Recent Innovations in Precision Agriculture (PA) are driven by Computer Vision and Data Processing systems to quantify plant parameters. Quantitative analysis of Plant Phenotyping in PA and monitoring morphological traits is a protracting process, precluding the objective and phenotyping pipeline. Computer Vision and Generative Adversarial Networks (GAN)'s offer a catalytic approach to the time-consuming process, providing a solution to the phenotyping bottleneck. This research proposes a concept of curating data of plant growth over time to predict conditional growth and responsive stimuli of the plant under different situations and how this can affect crop yield. The method proposed here is a non-invasive approach to the existing destructive biomass estimation methods and Frameworks. This methodology of the research focuses on utilizing image parameters modelled using a time series *Progressively Growing Generative Adversarial Networks PGGAN* to map plant growth patterns and progressive variance in biomass of plant in the Spatio-Temporal Domain. These Generative networks evaluate and predict based on merely raw pixel input excluding dependence on further constraints, feature vectors or parameters influencing data.

1. INTRODUCTION

A. Problem Statement and discussion

Phenomics is the systematic study of phenotypes coined by Steven A. Garan et al.[1] as emerging transdisciplinarity dedicated to the systematic study of phenotypes on a genome-wide scale. From decades quantifiable phenotyping and assessing parameters presented invasive methods and techniques. This task to a lot extent, was manual and hindered accurate as well as time-efficient measurement and assessment. All these approaches arguably present one common interdisciplinary goal, i.e increasing phenotyping throughput and maximizing yield quantitatively. However such approaches survey crops and present data analysis delayed enough to not procure any room for yield maximization. Multiple different approaches have claimed to observe and analyze crop parameters over complete growth period and establish a fine relation between temporal and yield estimation patterns, but such proposed methods obtain results over a large set of data points, generalized over a specific factor.

Fig. 1. An example of GM greenhouse with monitored and controlled environmental and spatial conditions source - The Independent

B. Importance and Relevance

Greenhouses growing Genetically Modified (GM) crops need to be maintained at constant environmental and simulated conditions. Multiple parameters have to be controlled and regulated inside a greenhouse for effective growth of crops and yield maximisation. Not at all times are these factors derived and so, yield maximisation in greenhouse is an experimental approach to new varieties. For deduced environmental parameters and conditions for certain crops, few other biotic and abiotic factors can hinder or affect growth in certain ways that are not always factored in during calculating parameters conducive for plant growth. Such factors may not always be affecting parametral calculations, but transpose visual cues on plant growth environment such as spectral change in soil values, or minute changes like leaf reflectance or visible changes in plant stimuli to biotic factors. Plant growth is inclusive of multiple environmental variables, and yield maximisation approaches are experimental to finding the optimum derived value for these variables. Computer Vision provides a catalytic approach to predicting optimum parameters for yield maximization in phenomics.

Fig. 2. An example of High Tunnel farming of crop under controlled environmental conditions source - Robin et al.

Another example to enabling high throughput phenotyping under controlled conditions is through high tunnel farming systems. Such systems allow all year farming of crops under controlled environmental conditions so crops can be grown later into fall, or earlier in spring and throughout the year as well. However while growing year around, the environmental variables need to be precisely maintained and contained to yield high throughput in crop harvest. As mentioned by Elizabeth et al. [2] in her research, Environmental conditions like temperature, relative humidity and light as well as other external factors affect crop growth, and tunnels under controlled conditions can affect the factors, but require adequate and optimum environmental values for high throughput phenotyping. In cur-

rent greenhouse and high tunnel environments, environmental variables are measured using hygrometer (humidity) and temperature sensors, which are not scalable across multiple plants, as well as do not incorporate visual traits possessed by plants that hinder growth such as leaf reflectance, temporal evaluation of LAI and inflorescence anthesis, and phenotyping stress effect on stems and foliage. Visual telemetry monitoring systems as proposed in this approach help solve the existing limitations to maximising yield, enabling autonomous modes of monitoring and comply to meet the second UN's Sustainable Development Goal (SDG), *to implement resilient agricultural practices to increase productivity and production*

C. Approach to the Method and Methodologies

Many of these proposed approaches are observed to be invasive/destructive and not testable over Real world factors. Computer Vision based phenomics has yielded significantly favorable results studied in several papers and it's objectively iterated to solutionise phenotyping bottlenecks. Various Deep Learning Methods and Convolutional Networks, Frameworks have claimed to procure high accuracy and efficiency in Fruit Detection, and occlusion based identification in different environmental conditions and plant types. Data analysis and Feature Extraction as well as Feature Engineering are integral components of phenomics. Early detection of plant diseases using GAN's presented by Ahmed A. Gomaa et al.[3] present a classification system based on real-time images for early identification of plant infection prior of onset of severe disease symptoms using CycleGAN networks which use Pix to Pix conversion to overlay predicted diseases on plant. However, the CycleGAN network solely predicts the generative description of disease on plant and does not factor-in other parameters influencing diseases, as well as is unable to predict a growth pattern on arbitrary images and latter disease formation outcome. This method does not produce descriptive outcome while evaluating *Leaf Area Index (LAI)*, and hence a Progressively Growing GAN (introduced later) is used. Another Deep Neural Network (DNN) method proposed by Inkyu et al. [4] involves subsampling cultivars using UAV for weed sampling and vegatation monitoring, which was accurate and reliable, however expensive and unscalable. Few other research methods imposed constriction in Real-time Yield Estimation, Predction, and Analysis. A small sample of Research papers [5–7] offer highly efficient phenotyping research and algorithms individually, but in discrete forms.

Klukas et al.[8] in his research accurately points out that *Phenotyping tools in common use are labor-intensive, time-consuming and costly, and require destruction of plants at fixed times or at particular phenological stages. The goal of current plant phenotyping is to raise the accuracy, precision, and throughput of phenotype inference at all levels of biological organization, while reducing costs and labor through mechanization, remote sensing, improving data integration, and experimental design.* Research in phenomics is yet to explore all the aspects of Temporal Generative prediction of plant growth. While Spatial evaluation of plant parameters have been proposed in various research models, temporal prediction of time-frame of plants in visual context is to be fully explored. Generative prediction of plant growth presents an outline of the different factors contributing to selective growth and environmental stresses hindering growth. Collectively, this technology enables crop improvement in response to future and present climatic conditions and demographics. Some constraints and stresses affect the plant phenotype in a specific manner, and such conditional variances if generatively modelled can help in

early understanding of factors involved in growth. Unmanned autonomous agricultural units can employ the use of such a methodological approach understand stress on plant preceding harvest, to inoculate biostimulants which influence phenotypic traits and improve yield by enhancing crop stress-tolerance.

D. Proposition and Introduction to solution

To enable spatio-temporal prediction of growth frames of input plant data, spatio-temporal 3d convolutions are used in all encoder and decoder modules in the GAN network. GANs, introduced by Goodfellow et al is indisputably one of the fastest emerging Network in Agricultural Domain. As elaborately discussed above, GANs entail an important task in Spatial and Temporal ingestion and objective prediction of plant traits to assess plant growth and conditional stimuli on different indices such as Leaf fresh weight (LFW), leaf dry weight (LDW), leaf area Index (LAI), Stem Diameter Variation(SDV). Biomass and other genotypic and phenotypic traits of plants under environmental stress or constraints. Current conventional approaches to GAN in precision agriculture were circumscribed to Synthetic data generation applications proposed by Ibis Prevedello et al.[9] or Pixel transpose approaches in disease prediction. Current systems and GAN architectures proposed by Marco Körner et al.[10] have enabled Temporal as well as spatial Convolution types with the help of Progressively Growing GANs otherwise known as PGGAN's to accelerate highly accurate, detailed Generative predictions. The aim of the proposed research is to enable adapting to the crop setup sample according to the observed predictions objectively through predicted spatial and temporal predictions for further crop frames. Subsequently, this system helps in prediction of further plant traits generatively for testing Genetically Modified (GM) crops under simulated conditions to get a predictive outcome comparable to established outcome couple of weeks prior to yield harvest. While the current system has been limited to development on different genotypes of Arabidopsis Thaliana plant species, such a concept can be used in quantitative prediction of multiple varieties of GM crops. GM crops have been developed with the objective of maximizing yield and also making the plant sustainable to various biotic and abiotic environmental constraints and stresses keeping in mind biotechnology practices. While such a system has been widely adopted, multiple crops fail to reproduce traits of crop origin under certain environmental conditions and such a cultivar is of no use. This testing system is a long exhaustive process and simulating environmental conditions for a long period solely for testing plant optimum to be produced under GM is not always sustainable. An autonomous system, similar to what is proposed above can not only help in predicting visual generative models of the plant in the form of temporal and spatial analysis, but also helps in reducing human intervention and maximize yield through prediction.

With the integration of autonomous data collection approaches such as motorized plates for translational data collection approach 3 and another robot navigation system in greenhouses by M. Pattinson et al. [11] (intially developed for pest monitoring) are efficient solutions for data collection systems to be ingested in GAN's to model yield growth, as proposed in the next few sections.

2. RELATED WORK

Recent Innovations in precision agriculture are devised with the use of multiple technological frameworks or Machine Learning

Fig. 3. Translation autonomous phenotyping methodology for data collection source - Paula Ramos et al.


Fig. 4. Autonomous telemetry system for data collection using high precision GNSS geolocationing system by M. Pattinson et al. [11]

and Robotic approaches. In the past decade, a boom in use of Computer vision approaches to DNN's and CNN's observed high accuracy in classification and object detection approaches. Temporal approaches to Plant parameters prediction were observed to base their approach on numerical time series data arrays. While these temporal predictions yielded accurate results, these approaches have been limited to prediction upon numerical data observed, measured and ingested, impeding the objective of autonomous precision agriculture and self sustaining systems. Observations on research methods proposed previously demonstrate numerical temporal data predictions influenced by factors such as Climate or Temperature(Degree Day Predictions), soil constituents [12], growth influenced by different stresses such as Nitrogen deficiency or water/weed stress. Few Raw methods propose predictions based on Change in Leaf Area or parameters like LFW, LDW, SDV et al. Such methods were followed by multiple different applications with Computer Vision. Frameworks such as Recurrent Convolutional Neural Networks(R-CNN), Mask R-CNN, Deep Neural Networks(DNN), Convolutional Neural Networks(CNN), Long Short term Memory Cells(LSTM) and Support Vector Machines(SVM) and Regression(SVR) have achieved commendable accuracy in object detection approaches in precision agriculture. Such applications include fruit detection and counting methods, plant disease classification and other similar applications(eg - My research on Apple plant disease estimation using Yolov4 framework) for Spatial Classification. PGGAN's propose generative combined method of Spatial and Temporal prediction from previous two methods. Compared to the previous methods, which used temporal analysis on one dimensional input data values and singular parameters, PGGAN's are 3 Dimensional

Convolutional Networks without relying on input parameters, take in raw pixel values and other exposed plant traits to predict future frames.

A. Previous Approaches to Spatial and Temporal GAN's

Time Series Generative Adversarial Networks or TGAN's was introduced in 2019 to generate sequential time series data.
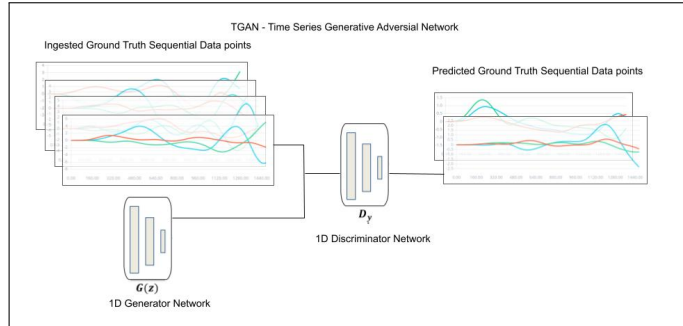
Fig. 5. An example of 1 Dimensional Temporal TGAN Network with ingested sequential IMU data points

An equation to describe the supervised loss function of TGAN in the figure 5 described can be written as,

$$L_{\text{supervised}} = \mathbb{E}_{s,x_{1:T} \sim p} \left[\sum_t \| h_t - g_X (h_S, h_{t-1}, z_t) \|_2 \right] \quad \textbf{(1)}$$

where the equation and functions in TGAN are elaborately explained by Eoin et al.[13] in his paper.

TGAN's can predict discrete or sequential data points (example - Acoustical Data) while respecting ingested data pattern. These TGAN's for a long time were applied to numerical data points and with PGGAN, as 3D convolutional networks to generate combined image pixels, with 2 Dimensional Generated Image data and another dimension addressed to Temporal prediction. Spatial GAN's are making their way into precision agriculture, with one core benefit that they learn and generate in an unsupervised manner which means that they do not require reference or evaluation data points or values to test and predict upon. Reference data, which available upon harvest is limited and GAN's offer meaningful generated predictions without the evaluation data points in an unsupervised manner. CycleGAN's consistently used in agriculture applications such as visual generation of plant disease spread on leaves or modelling spread or decay of bacteria on plants use pix2pix transpose methods to convert original pixels to generated pixels on a confined parametral region.

These Generative methods are also useful in creating near-identical but distinguishable image data points to help create synthetic data in the form of multi-modal output for model training. Such methods also help translate pixels in visual images to highlight certain pixels useful in calculation of certain plant indices through solely visual RGB inputs. However such GAN's are limited and confined to Spatial Generation without the variable of time, i.e fail to generate progressive variance in images temporally.

3. MATERIALS AND METHODS

This study comprised of experiments to dynamically utilize a State-of-the-art Generative Adversarial framework for spatio-
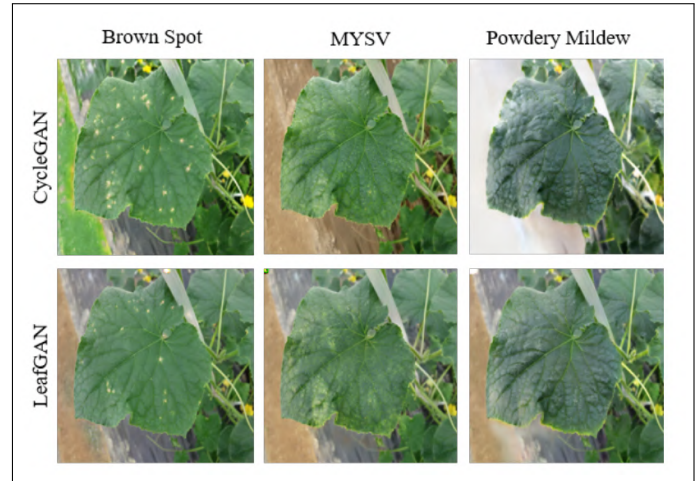
Fig. 6. An implementation of CycleGAN in the form, Leaf-GAN by Quan Huu et al. [14]

temporal generative mapping of plant phenomics and evaluate the efficacy of the generated data on series of Vegetative Indices, Structural Similarity Index Measures, Fréchet inception distance (FID) scores of the generated data under multiple environmental stresses. The deduced correlation will be used in evaluating the optimum stresses and environmental variables conducive for maximizing yield in temporally predicted frames.

A. Datasets

In the proposed paper, I use two phenotyping datasets differing in throughput and quantitatively using PGGAN's to generate predicted crop yields for the frames. With the pandemic widespread during research, it wasn't favorable to collect data from plant phenotyping specimens in the school laboratory under different controlled environments. Hence, choosing evaluated and properly curated datasets under monitored conditions from Open Access research papers made available for academic purposes was the best choice. The first dataset used in the research was of Arabidopsis Thaliana plants in a high throughput environment. The second dataset, were of Beta vulgaris (sugar beet plants) of the variety 'Samuela' in *KWS Suisse SA, Basel, Switzerland* under different stresses.

A.1. Arabidopsis Thaliana Dataset

Arabidopsis Thaliana is an annual weed which has been used in multiple research methods for phenotyping over the past decade and is an important plant model. One of the reasons for the growing interest in Thale Cress, Arabidopsis thaliana is due to the fact that it grows, reproduces and responds to external environments or stresses similar to other commercial plants and crops. This plant can easily be cultivated in a small confined lab or greenhouse, and is inexpensive as well as allows simulating conditional growth as observed in multiple other crops. With the ability to mutagenize easily and possessing a comparatively small genome, enables extensive genetic experiments and covariance in genotype-phenotype correlations.

Despite being recognized as a non-commercial crop and having a low economic value, it has proven to be a sustainable plant model to simulate phenotype and genotype growth patterns across multiple crops. Arabidopsis contains a diploid genome with under 30,000 protein encoding genes distributed over 5 chromosomes to enable scalable research over different genome
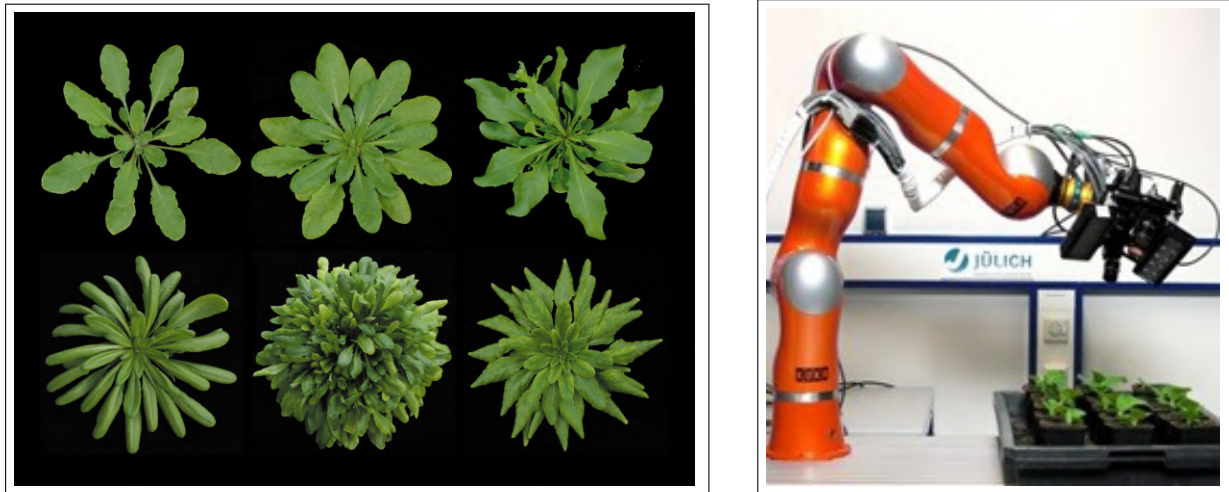
Fig. 7. (Left) Figure displays entropy in Arabidopsis Thaliana phenotype with variance in genome type and multiple genome discovery facilitated. **(Right)** Data Acquisition Method for capturing images of Arabidopsis Thaliana under controlled environment and Interference of light [15]

and phenotypical patterns[16]. For a a particular environmental condition and under different stresses, if the Arabidopsis Thaliana phenotype performs effective under these conditions and a corresponding genotype can be derived. With the high covariance in genotypical similarity, a similar gene can be found in other plant accelerating research in maximizing yield.

In this research paper, to accomplish the objective of proposed method, I utilize the Arabidopsis Thaliana Dataset curated by Hannah et al. [15] under the background of the EPSRC funded project "Dynamic Modelling of Plant Growth with Computer Vision". The dataset is acquired in a top-down view format in visual RGB imagery, in a high throughput growth environment. The dataset includes segmented area for the corresponding leaves and plant.

These Arabidopdis Thaliana plants, accession Columbia (Col-0) were grown in 180ml compost (Levington F2 + 205 grit sand) in PSI 6cm square pots, as described in Open Access Dataset. These plants were grown to 75% field capacity (region extent covering the field/pot) with gravimetric watering in a 10 hour day in a glasshouse regime maintained at 15°C to 20°C on the PSI platform. The seeds were sown and the plants were pricked after 10 days into weighted pots. 80 plants were grown, split between 4 trays. To lessen the probability of plants overlapping adjacent plants in trays at a later stage of growth, certain neighbouring plants were removed between 30 to 53 days after sowing and the plants were harvested at 56 days after sowing. Imaging for each plant started at 21 days after sowing and continued for a total of 35 days. The top-view image of the plant was captured each at an interval of 15 minutes, which started at 9am and continued till 8pm daily.

Since this dataset is curated such, and plants are grown in a high throughput phenotyping environment, no environmental constraints or stresses were introduced in growing the plant. The plant was intended to grow in ideal conditions, and the intent of the dataset used for the GAN model was to replicate growth in high throughput environment while simultaneously also testing the efficacy of another dataset under environmental and biotic stresses to ensure Generated temporal and spatial sequences maintain high accuracy even in biased, or continuously varying, data with introduced anomalies as well as in ideal conditions.

Table 1. Attributes and data of annotated plant images from Tray 31

Plants per image	Images	Total plants	Ecotype
20	12	240	Col-0
18	9	162	Col-0
16	8	128	Col-0
14	7	98	Col-0
12	4	48	Col-0
10	3	30	Col-0
Total	43	706	

Table 2. Attributes and data of annotated plant images from Tray 32

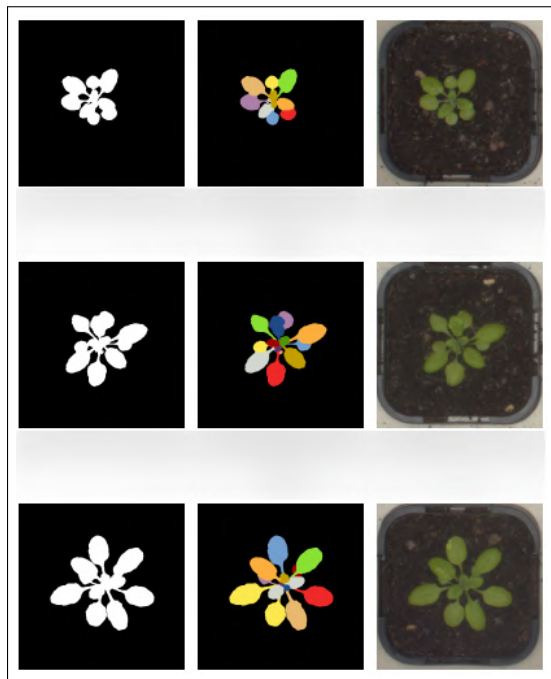| Plants per image	Images	Total plants	Ecotype	Days after sowing
20	4	80	Col-0	21,22,25,28
18	2	36	Col-0	31,34
16	2	32	Col-0	37,40
14	2	28	Col-0	43,46
12	2	24	Col-0	49,52
10	1	10	Col-0	55
Total	13	210		
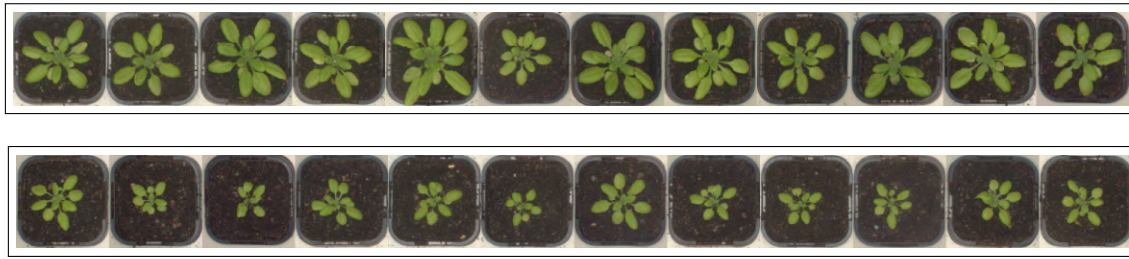
Fig. 8. Growth of different Arabidopsis Thaliana plant pots under different Temporal sequences

Training GAN's require an upwards of 100k images in general, but this solution maintains generation of high resolution images for training dataset under 700 and 500 images respectively for each dataset. Acquisition of images for plant phenotyping is usually limited for training of Neural Networks. Under this constraint, developing a GAN Network capable of generating similar images from small dataset has been one of the challenges, but also one of the aims of the project.

A.2. Sugar Beet plants (Beta vulgaris) Dataset

Contrary to the initial Dataset of Arabidopsis Thaliana Col-0 Ecotype grown under high throughput conditions, this dataset aims to replicate Real world multivariate environmental factors affecting crop growth in a controlled environment. The plant cultivars grown in this dataset are grown under multiple stress factors such as weeds, drought and nutrient deficiency, and other multiple combinations of such constraints containing weed induced nitrogen contraint or weed induced hydration deficiency. Such multiple environmental constraints on crops in controlled environment intend to simulate real conditions of crops where certain environmental variables and parameters aren't conducive for crop growth or prevent maximization of yield. This dataset also visually replicates the real world scenario under multiple aspects with distorted varied geometry in each image, varying illuminance, varying dimensions for Region of Interest (ROI) and certain measured errors in distance of captured images, and a smaller dataset compared to the first one with under 500 images. The dataset generated spatio-temporal outputs can be used to predicted the optimum growth environment from visual traits as well as deriving the soil pattern with optimum use of bio-fertilizers or other inputs to plants enabling Variable Rate Application (VRA) methodology on plant cultivars. All such visual, and environmental constraints make this a perfect dataset to replicate the Real world scenario and test the efficacy of the model in adverse conditions claiming the robustness of the utilized novel Generative Adversarial Network (GAN) architecture in training the model in a simulated setting.

**Fig. 9.** Images of Segmented leaves, plant and RGB Visual image in the Dataset

The curated dataset consists of prevalent Spatio-temporal visual, Infrared and Spectral data, each of the same set of plant cultivar captured of sugarbeet growth under optimal conditions, high throughput singleton condition with surplus Nitrogen fertilization and abundantly under varying environmental stresses, Nitrogen deficiency, low/medium/high weed stress and also covarying weed stress with Nitrogen and Water deficiency. Compared to the previous approach consisting of two trays and multiple plant arrays with images acquired every 15 minutes, this approach consists of 30 trays, with single plant and image acquired biweekly, reducing the overall set of images twice to a total of 432 images of different plants captured and 16 Spatio-temporal images of the same plant. The curated dataset contains over two months of data, wherin visual, stereo

Fig. 10. Dataset of Sugar Beet (Beta Vulgaris) used in training of PGGAN model. The image order in rows represents different periods of plant growth, in the order frame 3, frame 6, frame 9 and frame 16 (pre-harvest) respectively.
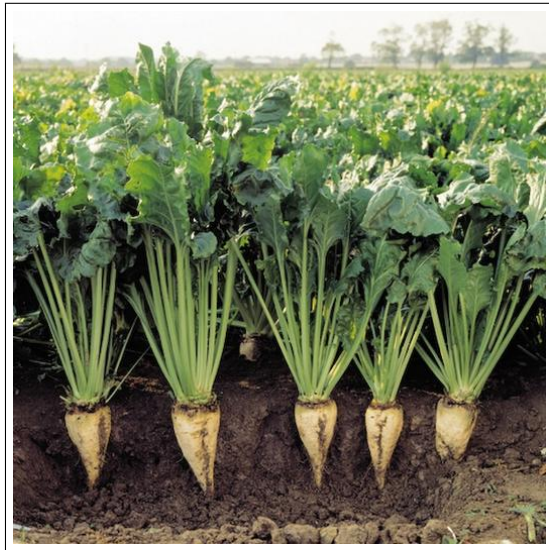
Fig. 11. Example of Sugar Beet grown in optimum conditions in a greenhouse to maximize yield. source - Agritech Portal

infra-red and multispectral images are biweekly collected. The Multispectral and infra-red pairs do not contribute to training of the PGGAN Neural Network, due to the reduced raw image features available. However, these multispectral and Infra-red images can be used to evaluate generated visual images of different parameters. During the time of the harvest, the dataset also provides images of obtained sugarbeet crop and the yield of sugarbeet. The proposed research will also establish correlation in Projected Leaf Area (PLA), deduced maximum yield for Optimum Environmental Variables and PLA Yield Contraction, Optimum Nitrogen N_{opt}, Optimum Hydration H_{opt}, Yield Contraction through Weed Stress Ws_{yc} generated by the *PGGAN*.

Plant Phenotyping and Dataset acquisiton was conducted at ETH research station for plant sciences in Lindau Eschikon, Switzerland by Raghav Khanna et al. [17] under Creative Commons License. These Sugar beet plants (Beta Vulgaris) of the specific variety *Samuela* from (KWS Suisse SA, Basel, Switzerland) were grown in a greenhouse chamber under simulated, but controlled environmental conditions with Temperature varying from 24°C to 12°C from day to night respectively, and the relative humidty varying from 50% to 80% with an average observed 60% throughout the day for a period of 2 months. 6 sugar beet plant cultivars each, were sown in 30 trays for a period of 2 months, using a peat substrate (Klasmann substrate 1 and 2). Depending on the allocated hydration levels to each plant under different stresses, regular watering was volume controlled, and applied manually. After the period of 2 months, the plants were harvested. Different stresses were introduced in these plants as discussed above, with severity increasing from low to high and weed induced secondary stresses were also introduced in selected samples. Table 3 elaborately explains corresponding constraints and their severity with crop samples and soil types and number of boxes being affected by each constraint.

Induced Weed Stress Ws$_{cstr}$: Replicating the climatic and environmental conditions or real world to test robustness of the PGGAN has been one of the aims of the project, and hence, a Weed Stress introduced in the dataset has been beneficial in testing systematic responsive generation by identification of anomalies in plant growth as weeds and systematically generating morphological traits was evaluated. The weed constraints introduced in Beta Vulgaris are two different monocytyl and di-

Table 3. Environmental constraints corresponding to box number and severity of constraint induced in Beta Vulgaris

Description	n(Boxes)	Soil Type	Water Input H_{prov}	Nitrogen Input N_{prov}	Weed Pressure Ws_{cstr}	Box Numbers
Control	3	2	Sufficient	High	None	1,2,3
Low N	3	1	Sufficient	Low	None	19,20,21
Med N	3	2	Sufficient	Medium	None	4,5,6
Med weeds	3	2	Sufficient	High	Medium	7,8,9
High weeds	3	2	Sufficient	High	High	10,11,12
Dry	3	2	Low	Medium	None	13,14,15
Weed Only Dicot	1	2	Sufficient	High	High	NaN
Weed Only Monocot	1	2	Sufficient	High	High	NaN
Weed Only Mixed	1	2	Sufficient	High	High	NaN
Low N - Med Weed	3	1	Sufficient	Low	Medium	22,23,24
Drying - Med N - High Weed	3	2	Low	Medium	High	16,17,18
Drying - Low N	3	1	Low	Low	None	25,26,27

cytyl weeds. For monocyte weeds, three different grass species were used - Poa pratensis L., Lolium perenne L. and Festuca rubra agg. L., in variable combinations.
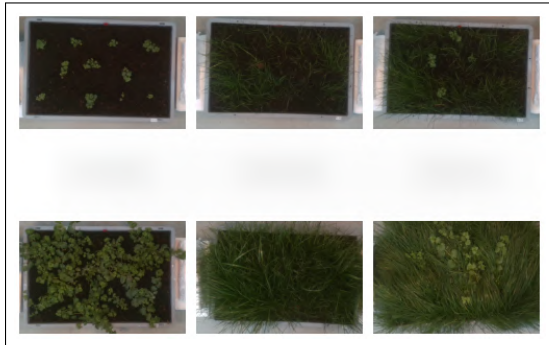
Fig. 12. Left shows Dicot weeds, **Centre** shows monocot weeds and **Right** shows a mixture of dicot and monocot weeds at intervals 4 weeks apart inducing stress on the plant growth. These weeds grown individually for representation, are planted in plant trays during the actual cycle.

The Dicot weeds used in the experiment are locally collected in the form, Stellaria media (L.) Vill. (common chickweed). During the experimentation, as mentioned in Table 3, different severity levels of weed stresses were used which affected yield contraction in different manners. Three different levels of weed density used were low weed stress, having few or none weeds affecting the crop, medium weed stress having 2 to 4 dicotyledonous weeds, and high weed stress having 4 to 8 dicotyledonous weeds and 2 to 4 monocotyledonous weeds, significantly higher. It was observed that Medium weed stress have almost negligible effect on the crop growth, and high weed stress to some extent had an effect but not on a deeper extent, as compared to Water constraint.
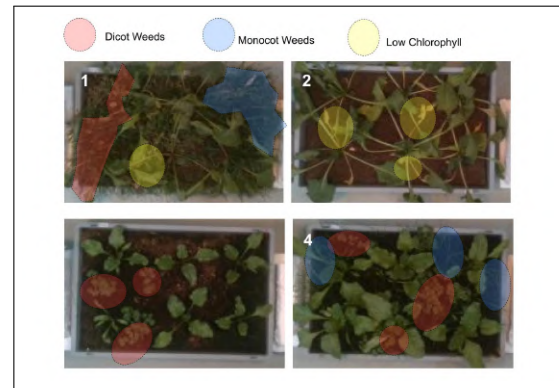

Fig. 13. Explaining environmental stresses in dataset and their temporal impact on visual traits

Drought Stress/Water constraint H_{cstr} : This stress constraint imposes the most visually observable contraction on plant and is on a higher severity scale compared to the Weed Stress. Two levels of severity of the water constraint were added to the plant, with limited (drying) plant boxes and the other one with relatively sufficient hydration, being irrigated on a regular interval of 2 to 3 days. The cultivars with water constraint did not have continuous long duration constraint which would cause wilting, rather alternative irrigation was carried out, where the plant was watered regularly for a duration of about 2-3 weeks and only after that was the water constraint imposed, which was continued for a specific period, after which proper irrigation was provided for 2 weeks preceding senescence.

Nitrogen Stress H_{cstr}: Following drought stress, Nitrogen stress had a high magnitude of contribution in inducing Yield Contraction (Y_{cntrn}). Plot 14 provides data that Medium Nitrogen levels in cultivar has nearly the same Y_{cntrn} as much as by Medium Weeds, but Lower Nitrogen levels in soil has significantly higher Y_{cntrn} as compared to high weeds in cultivar which explains
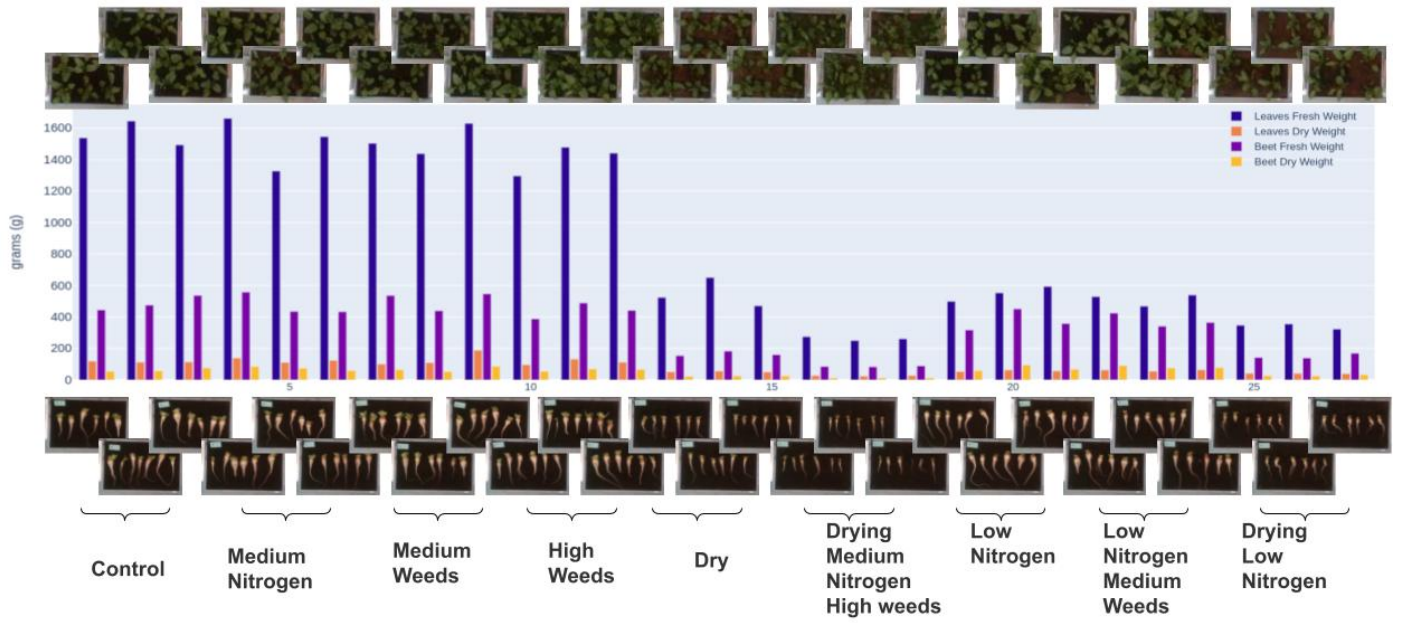
Fig. 14. Visual plot comparing Leaf Fresh Weight, Beet Fresh Weight, Leaf Dry Weight and Beet Dry Weight of the dataset of 30 samples with 3 samples corresponding to environmental stresses imposed on each cultivar curated in the experiment.

the magnitude of $\mathbf{Y}_{cntrn}$ of Nitrogen level co-variation proportionality with Fresh Weight of Leaves and Beet. 3 conditions for Nitrogen stress, deficient, medium and surplus were imposed on cultivars simulating conditions of 2, 4 and 8 x 10^{-3} kg/m^2. The low Nitrogen soil cultivar had bare-minimum supply to initiate initial plant growth, the medium Nitrogen cultivar had Nitrogen supply to intiate 1-2 months of growth and the surplus Nitrogen supply cultivar received additional Nitrogen each week.

There were also cultivars where multiple constraints were imposed. Example in containers from 16 to 18, High Weed Induced, Medium Nitrogen stress and drying were the constraints. With Drought Stress, Nitrogen Stress and Weed Stress as $\mathbf{H}_{cstr}$, $\mathbf{N}_{cstr}$, $\mathbf{Ws}_{cstr}$ respectively, Yield Contraction $\mathbf{Y}_{cntrn}$ can be expressed as a function of these variables. The maximum and minimum values of Fresh Leaf Weight $\mathbf{Leaf}_{FrWt}$ can be expressed in upper limit and lower limit of the observed values in cultivar 16-18 under same environmental stresses.

$$\mathbf{Y}_{cntrn}^{16-18} = f(\mathbf{Ws}_{high}, \mathbf{N}_{med}, \mathbf{H}_{low}) \quad 249.75g \leq \mathbf{Leaf}_{FrWt} \leq 274.54 \tag{2}$$

The Geometric mean for the $\mathbf{Leaf}_{FrWt}$ is calculated from the obtained three discrete values observed in cultivars 16 -18, which is then used for calculation of the approximate $\mathbf{Y}_{cntrn}^{16-18}$, considering the Leaf Fresh Weight in Optimal condition to have a $\mathbf{Y}_{cntrn}$ = 0

$$\mathbf{Leaf}_{FrWt} : \left(\prod_{c=16}^{c=18} Leaf_{FrWt} \right)^{\frac{1}{3}} = \sqrt[3]{\mathbb{L}_{FrWt}^{16} \mathbb{L}_{FrWt}^{17} \mathbb{L}_{FrWt}^{18}} = 261.54g \tag{3}$$

The Yield Contraction, measured with respect to Fresh Leaf Weight in Control cultivar (Tray 1 to 3). The Yield Contraction as a function of Environmental stresses 2, is a difference of Fresh Leaf weight measured (Tray 16 to 18) and Fresh Leaf Weight under Optimum Environmental parameters(Tray 1 to 3)

$$\mathbf{Y}_{cntrn}^{16-18} = \mathbf{Leaf}_{FrWt}^{opt} - \mathbf{Leaf}_{FrWt}^{meas} \tag{4}$$

The optimum Fresh Leaf Weight was found out to be 1557.418g using the Geometric Mean of tray 1, 2 and 3 from 3.

Hence the Yield Contraction for Tray 16 to 18 was derived to be 1295.878g under the constraints of three environmental stresses of low, medium and high severity. Yield Contraction would serve as an important index for comparing the optimum Projected Leaf Area (PLA) affected by different stresses and *n(stresses)* for generated frames by the *PGGAN* through Visual traits.

For comparison of the generated visual models by the PG-GAN of the dataset, initial and exhaustive visualization of the available data in the dataset is important. This helps to get a thorough understanding of the correlation and covariance of different feature vectors or ground truth observations with predicted generated traits in performance evaluation. Presenting a visual representation of different indices curated here, presents a benchmark for comparing indices and metrics evaluated in generated frames. Covariance in different ground truth feature vectors and predicted feature vectors were established using multivariant regression as well as linear regression models presented in further sections for evaluation. Certain anomalies in clusteral data points of different environmental stresses were observed in the ground truth biomass dataset. Such data points increase the overall loss functions in linear regression models which are solely dependent on feature vectors predictions. Such anomalies in data points are also visualised in tabulated clusteral plots to test the efficacy of the GAN model as discussed in the next section.

Figure 16 explains covariance in ground truth feature vectors extracted from the plant during the harvest. The Leaf Fresh Biomass and Box Numbers were used as input features in the **sample** linear regression model for the sole purpose of establishing correspondence in non-visual traits of plants. The output
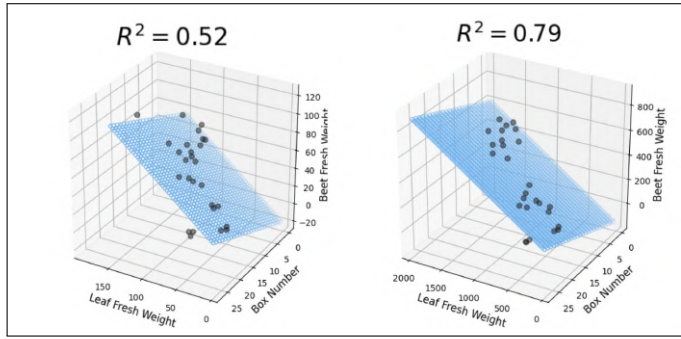
Fig. 15. Left X axis corresponding to Leaf Dry Weight, Y axis as Box no. and Z axis as Beet Dry Weight. **Right** 3 dimensional plot with 3 axes, X axis corresponding to Leaf Fresh Weight/Biomass, Y axis as Box No. and Z axis as Beet Fresh Weight.

feature vector was Beet Fresh Weight which was used to demonstrate covariance between the other features. Multiple Anomalies in feature correlation were observed and linear regression models could not explain correlation between non-visual traits of plants. These feature vectors were calculated by invasive and destructive means, but did not prove to make an impact in successful prediction of plant growth and correlation between different Biomass samples and environmental stresses in boxes.
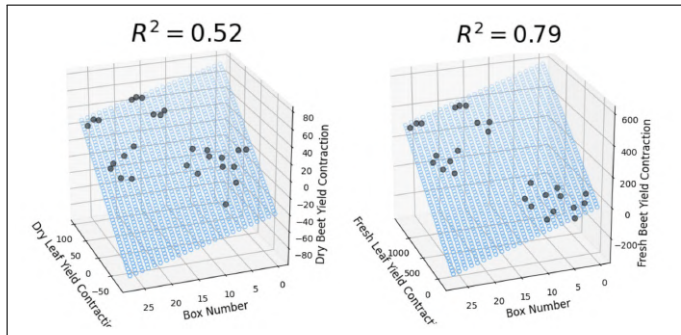
Fig. 16. Left X axis corresponding to Dry Leaf Yield Contraction, Y axis as Box no. and Z axis as Dry Beet Yield Contraction. **Right** 3 dimensional plot with 3 axes, X axis corresponding to Fresh Leaf Yield Contraction, Y axis as Box No. and Z axis as Fresh Beet Yield Contraction.

Similar to the Ground Truth Biomass Measurements, Yield Contraction Index 4 was used to evaluate decrease in yield of sample measurement in box with environmental stress compared to Geometric Mean of the plant cultivars in the 'control' Environmental sample to understand the impact of the Yield Contraction under different environmental stresses. Equation 3 was used to derive Mean Yield Contraction in the sample 16-18, under $\mathbf{Ws}_{high}$, $\mathbf{N}_{med}$, $\mathbf{H}_{low}$. The same approach was applied over all other samples during harvest and the leaf contraction was observed for all. To understand the covariance between Fresh Leaf Yield Contraction and Fresh Beet Yield Contraction, a statistical coefficient, coefficient of determination also known as R^2 was used. The covariance in Dry Yield Contraction was merely 0.52, indicating that the linear regression model was able to establish relation between 50% of the data points, while it was measured to be 0.79 in Fresh Yield Contraction.

$$R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{\sum (e_i^2)}{\sum (y_i - \bar{y})^2} \quad (5)$$

The reference index point to measure Yield Contraction in other samples was set by the Geometric Mean of the plant cultivar with controlled Environmental variables and the Geometric Mean of the reference samples were calculated in the form -

$$\mathbf{Leaf}_{FrWt} : \left(\prod_{c=1}^{c=3} Leaf_{FrWt} \right)^{\frac{1}{3}} = \sqrt[3]{\mathbb{L}_{FrWt}^1 \mathbb{L}_{FrWt}^2 \mathbb{L}_{FrWt}^3} = 1557.4g \quad (6)$$

$$\mathbf{Beet}_{FrWt} : \left(\prod_{c=1}^{c=3} Beet_{FrWt} \right)^{\frac{1}{3}} = \sqrt[3]{\mathbb{B}_{FrWt}^1 \mathbb{B}_{FrWt}^2 \mathbb{B}_{FrWt}^3} = 484.3g \quad (7)$$

$$\mathbf{Leaf}_{DryWt} : \left(\prod_{c=1}^{c=3} Leaf_{DryWt} \right)^{\frac{1}{3}} = \sqrt[3]{\mathbb{L}_{DryWt}^1 \mathbb{L}_{DryWt}^2 \mathbb{L}_{DryWt}^3} = 115.5g \quad (8)$$

$$\mathbf{Beet}_{DryWt} : \left(\prod_{c=1}^{c=3} Beet_{DryWt} \right)^{\frac{1}{3}} = \sqrt[3]{\mathbb{B}_{DryWt}^1 \mathbb{B}_{DryWt}^2 \mathbb{B}_{DryWt}^3} = 62.06g \quad (9)$$

The Geometric Mean of the Control cultivar samples were used as reference data points to measure the yield contraction of other cultivars.
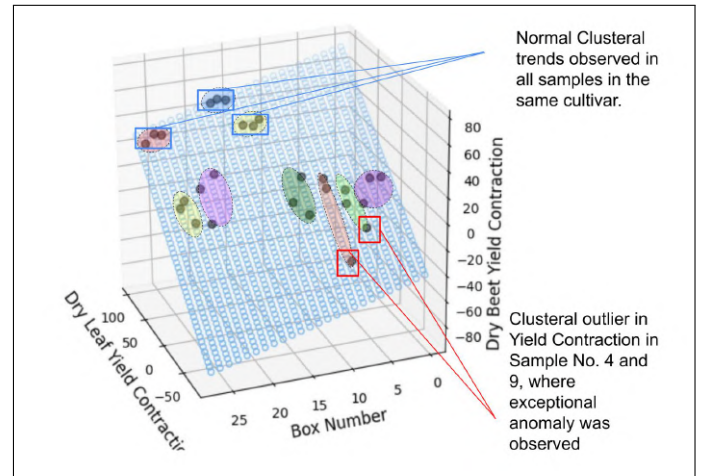

Fig. 17. Anomaly trends in Yield Contraction observed as cluster outliers. Negative Yield Contraction observed in certain samples indicating better performance of crops than Reference crops.

An overall trend comparison indicated that the sample of cultivar 17, with Low Hydration, Medium Nitrogen and High Weed stress seemed to perform the worst and had a high Yield Contraction Index. General comparative trends provided an outcome that cultivars following Box 16 tended to perform poor and the Yield Contraction $\mathbf{Y}_{cntrn}$ remained high in these samples. Certain Anomalies were also observed. Certain samples measured a negative Yield Contraction, indicating they performed
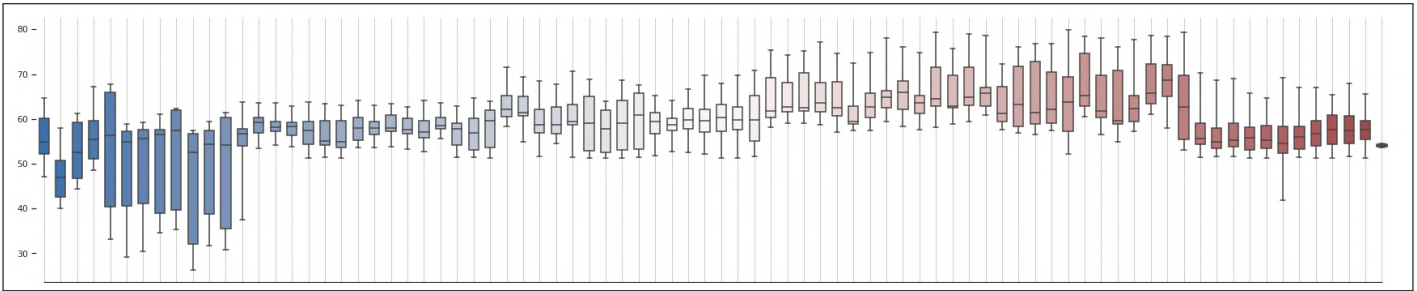
Fig. 18. Humdity measurement graph of greenhouse humidity maintained constant throughout cultivars
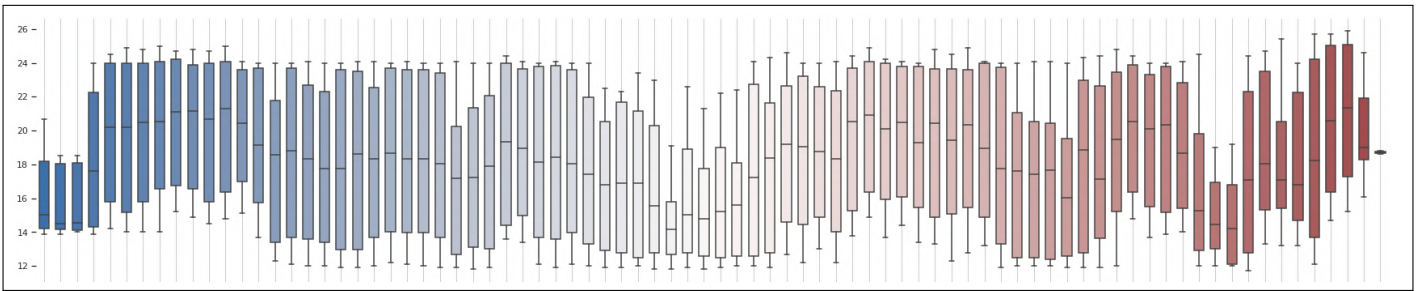
Fig. 19. Temperature measurement graph of Temperature conditions inside greenhouse kept constant during plant growth measurement

better than the cultivar under controlled environmental conditions.These Anomalies brought forward through these plots helps in assessing overall robustness of the PGGAN during evaluation, by observing performance on such samples showcasing anomalies in trends. Surprisingly a subset with two sample cultivars, Box Number 4 and 9 with Medium Nitrogen and Medium Weeds showed higher Fresh Leaf Biomass than Reference samples under optimum conditions.

These Anomalies in the form of feature vectors contribute in increasing the loss of a single dimensional linear regression model, but in the form of visual traits, these visual traits are generated with high accuracy in the PGGAN model explained elaborately in the proceeding subsection.

Multiple different metrics such as correlation of growth with Temperature and Humidity and Box Weight measurements temporally correlating with induced stress; were visualized before model training to thoroughly examine how the plant responds to environmental stimuli and responsiveness action to visual traits. On this basis, the robustness and efficacy was deduced and metrics such as 18, 19, 20 were further calculated.
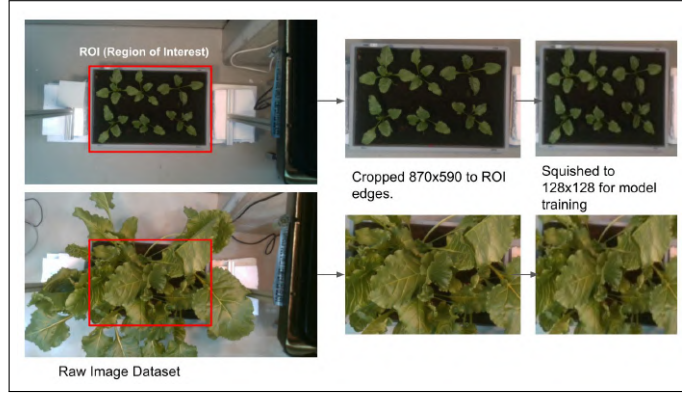
B. Data Preprocessing

Data preprocessing is one of the least time-consuming sections of the model training section, yet considered to be one of the most important. The sequential and topological order in which data is fed to the Neural Network, determines the type of generated output of the model. For the first dataset (Arabidopsis Thaliana), not a lot of data preprocessing had to be carried out because, the images in the dataset were organized and categorically sorted before made available to researchers. However the intial dataset was too large (61.6 GB) in size, and the aim of the project was to create a robust PGGAN model, which performs well on dataset with sample size below 1000 images. To meet this goal, the dataset was intially downloaded on Google Colaboratory (A Virtual Machine (VM) service offered by Google Research to
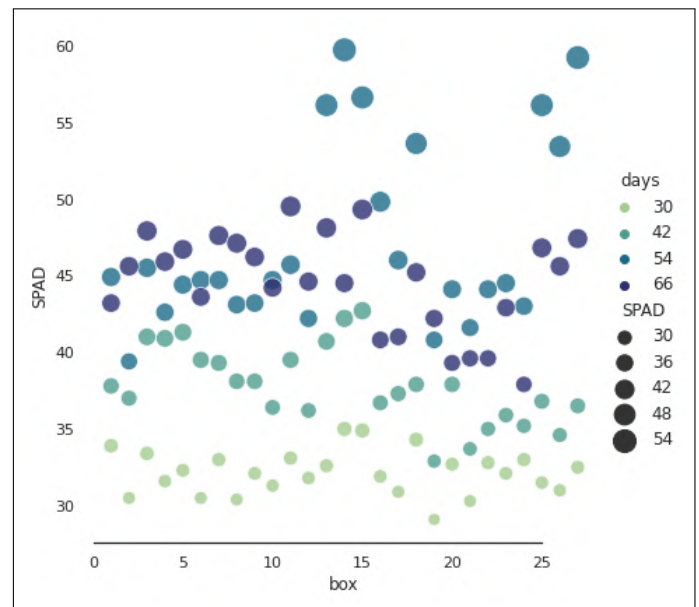
Fig. 20. SPAD Measurement covariance with Box Number and Days. With increasing days, The SPAD index also used to measure chlorophyll and Nitrogen content in plants depicted a smooth incremental trend, however some plants performed exceedingly well while some performed just average

enable Researchers carry computationally expensive tasks over the cloud). The downloaded Dataset was manually analyzed and a python script was later constructed to arbitrarily, but sequentially select random samples of Arabidopsis plant from the dataset and this dataset, with merely 619 plant images was used in the training of the model. The size of the images was reduced to a total of 128x128 using a python script and the size was truncated to 160 MB to train a light, and efficient GAN model.

Fig. 21. Data preprocessing for Beta Vulgaris Dataset and ROI extraction

The second dataset, of Beta Vulgaris required some additional preprocessing to ensure the model trains on an accurate set of temporal dataset and does not learn the gratuitous features in the dataset. The Raw dataset consisted of plant growth with significant amount of overlap of background. These background features were unnecessary for the model, and would also be a causation for model overfitting. To esnure extraction of relevant foreground of the plant including Vegetation and overlapping soil samples as well as Box Border as an indication of relative growth, a Region of Interest (ROI) had to be constructed around the images to ensure just the right amount of Area of images is used in the training process. A manual analysis method was used to determine prominent edges with high edge point density and rough co-ordinates were derived statistically measured over different samples. These co-ordinates were used to appropriately crop the 1920x1080 dimension images to 870x590. These images were squished to 128x128 by 'fitting the shortest width' image squash method. These images were then sorted into respective box types temporally finally ready for training.

4. PGGAN ARCHITECTURE AND MODEL TRAINING

A. Proposed Architecture

The most crucial part of this research was to understand the Network Architectures most suitable to fulfill the aim of the project. This research required the use of a Spatio-Temporal 3 Dimensional Generative Adversarial Network Architecture suitable under multiple settings and generates accurate and realistic predictions of Visual RGB images with a small dataset < 700 images. For this purpose the GAN constructed for the research was adopted from one of the initially proposed Progressively Growing Spatio-Temporal GANs (PGGAN) for generating Future Frames by Marco et al. [10]. Generative Adversarial Networks are an extension of Convolutional Neural Networks and are composed of two diverse Networks, Generator and Discriminator. Given a training set, initial approaches to standard GANs

learn to generate new data with statistical reference from training set. GANs initially were proposed to generate new data in unsupervised learning, however the rate of application of GANs in different domains has evolved this network to be utilized in supervised, semi-supervised and in reinforcement learning approaches. Two networks in a GAN, the Generator, as the name indicates Generates new, but statistically similar images as in the training set and the Discriminator evaluates the Generations in an unsupervised adversarial setting. These networks functions simultaneously, and the Generator maps data distributions in the constrained interest from the latent space and the generations are evaluated to test the analogous similarities in the generations to the training set in the latent space.

A developed approach to the same concept of GANs, a PG-GAN was introduced in this research to effectuate the goal of Spatio-Temporal 3 Dimensional data point generations. The proposed framework is based on the idea of training the Generator in an adversarial setting to predict sequential video frames based on input past frame. The discriminator is trained to distinguish between the generated realistic and fake samples. The proposed network employs discriminator in a way such that it is trained to discriminate between the generated sequence by the generator as well as from training dataset alternatively for a given set of temporal frames. Discriminator distinguishes between samples with a score ranging from 0 to 1, with 0 representing fake samples and 1 close to realistic. Based on the output score feedback representing the generated samples as realistic or fake, the Generator updates it's model weights after a given set of iterations to generate better samples with the feedback from the sample scores.

Training the GAN function model from features in latent space is broadly defined as,

$$V(D, G) = \mathbb{E}_{x P_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z P_z(z)}[\log(1 - D(G(z)))] \quad \textbf{(10)}$$

where D and G are Discriminator and Generator respectively, and Pdata(x) represents distribution of real data and P(z) represents distribution of generator. This network learns the visual traits pixel-by-pixel based on solely raw pixel input 2 Dimensional feature points and temporally predicts 3 Dimensional Frames with the third dimensional corresponding to time.

The Framework adopted from [10], the Discriminator Network D receives (eqn 11) no. of input frames from the ground Truth Training set and the Generator takes (eqn 12) no. of input and output frames.

$$x = (x_{t-t_{\text{in}}+1}, \ldots, x_{t+t_{\text{out}}}) \quad \textbf{(11)}$$

$$\widetilde{x} = (z, G(z)) = (x_{t-t_{\text{in}}+1}, \ldots, \widetilde{x}_{t+t_{\text{out}}}) \quad \textbf{(12)}$$

These input and output values are variables before the training process, but allotted a constant value right before initiation of the model training. These variable x and $\widetilde{x}$ were set to 6 for the first dataset (Arabidopsis Thaliana) Training and 4 for Beta Vulgaris dataset. This difference was set due to the difference in the training datasets, the first having 619 images and the second having 332 images to avoid underfitting of the model. This means that the first network took an input of 6 frames and generated 6 frames, while the second network took an input of 4 frames and generated 4 frames.

The proposed Architecture is a modification of the existing PGGAN network to scale the existing 2 Dimensional Convolutional Layers to 3 Dimensional Convolutional Layers to enable
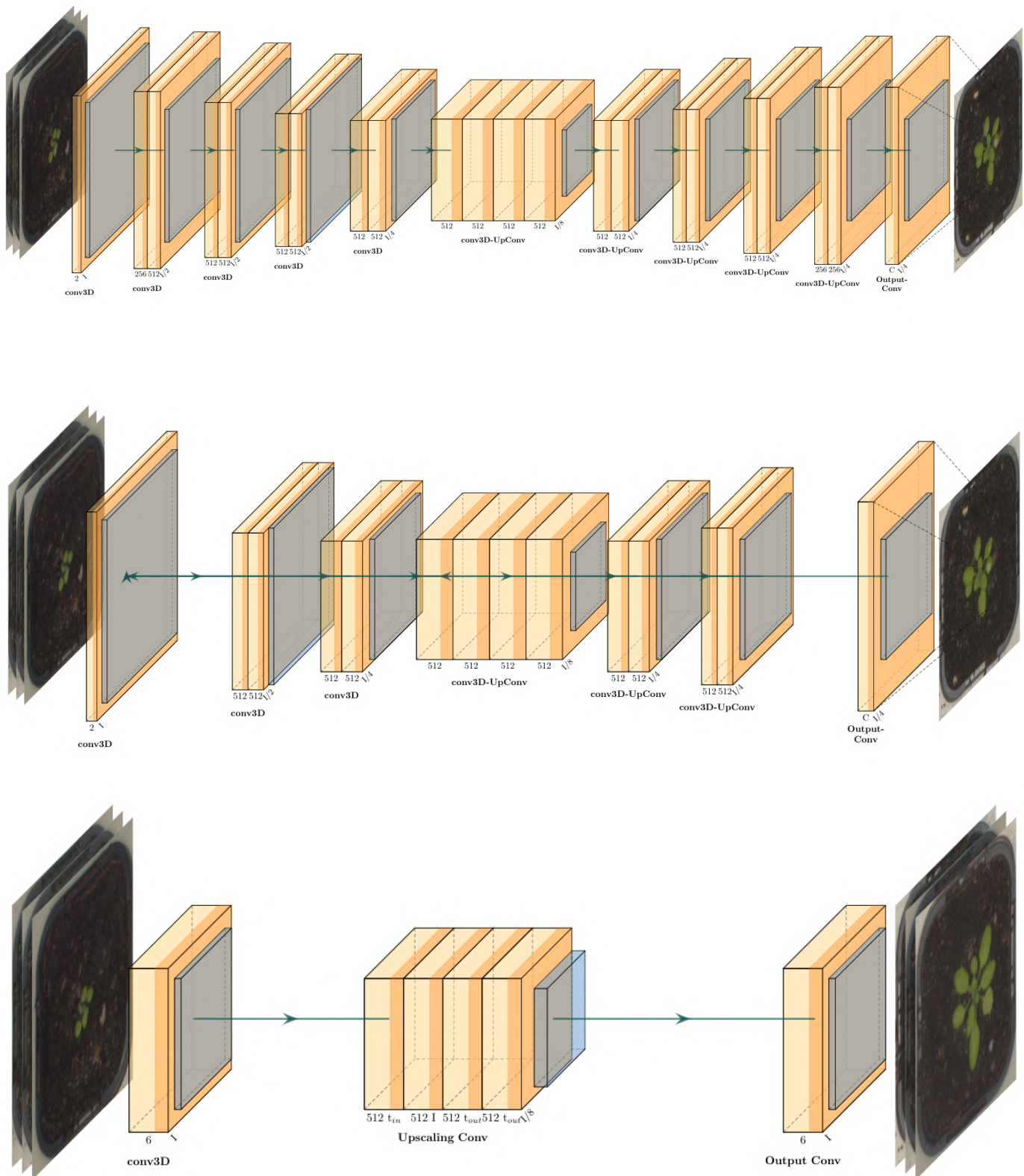
Fig. 22. Proposed PGGAN Generator Network Architecture. The model is initialized with an input and output of 4 x 4pixels. During the training, layers are added progressively to the Network Architecture to sequentially increase resolution. The Network Architecture resolution always corresponds to the input and Output node resolution. The first image architecture displays 64 x 64px, second 16x16px and third, 4x4px resolution nodes respectively. The model takes in 6 frames and generates temporal 6 frames for the First dataset. Same Architecture is followed for both datasets.

encoding and decoding of spatio-temporal components of the dataset. The PGGAN network replaces GANs and cGAN networks in sequential upsampling of spatial resolution of frames. Spatial upsampling initially starts with a lower resolution frame by learning the features from the latent space and thereby increasing resolution Geometrically to higher resolution capped at 128x128 as in the figure 22.

Encoder part of the Architecture is symmetrical with the decoder part except that the spatial upsampling layers replace the spatial downsampling layers. Spatial downsampling of layers is performed with the help of 3D convolutions of asymmetrical kernel sizes and stride layers. 2 Convolutional layers are further added in the encoder and decoder part to increase the Network resolution on top of the existing PGGAN Network. Generally in Generative Networks, Batch Normalization is used after Convolutional layers. But in Progressively Growing GANs, feature vector normalization is applied to each pixel to Normalize it to unit length before the convolutional network in order to restrict the arbitrarily changing value nature of training between the Generator and Discriminator. However there are no Feature Vector Normalization layers used in the proposed architecture, in contrast a mini-batch standard deviation layer is added to the end with the core intent of increasing variation in generators output and prevent mode collapse. This methodology is adopted since this PGGAN is built on top of concatenated 3D convolutional layers instead of 2D in PGGANs and is suited for highly accurate generations for datasets with < 700 images. This feature vector normalization layer is modified to be applied on Spatial as well as Temporal components of the generation and can be described as,

$$b_{x,y,z} = a_{x,y,z} / \sqrt{\frac{1}{n_f} a_{x,y,z}^\top a_{x,y,z} + \epsilon} \qquad \textbf{(13)}$$

$\epsilon = 10^{-8}$, n_f corresponds to the number of feature maps, $a_{x,y,z}$ is the original feature vector and $b_{x,y,z}$ is the Normalized feature vector, calculated for pixels(x,y,z)

These values over all features in the latent space and spatial locations is averaged to a scalar value. This generates an additional feature map in mini-batch over replication. Later, the original layer is modified to calculate temporal and spatial locations through the generated feature map which yields in incremental variation over the data points. The last layer in the model is a linear activation function following a Fully Connected layer. A LeakyReLU, non-linear activation function is made use at each layer to help address neuron deactivation and vanishing gradients.

In the Network, the Convolutional Layer takes in input 128 x 128px image, which is further processed through batch deviation and pooling layers till 4 x 4px, after which the Upsampling Convolutional layer (UpConv t_{out}) is initiated from 4 x 4px to 128 x 128px matching the input resolution. Finally an output 3D Convolutional Layer (Conv3D) is used to output the generated frames.

Complementary to the existing layers, to stabilize the training, a weight scaling layer is added on top of the layers. This layer helps in estimation of the standard deviation of the weights and Normalizes them.

For the evaluation of the model, the Wasserstein GAN with gradient penalty (WGAN-GP) loss function is employed. The Wasserstein GAN loss seeks to increase the gap between the real and generated image scores and is a function of critic score on real and generated images [18]. The WGAN-GP loss function ef-

fectively increases the quality of the Generated frames. There are two discrete WGAN-GP loss functions for the discriminator and generator. The loss function for the discriminator is explained as,

$$L_D(\boldsymbol{x}, \tilde{\boldsymbol{x}}, \hat{\boldsymbol{x}}) = \underbrace{\underset{\tilde{\boldsymbol{x}} \sim P_g}{E}[D(\tilde{\boldsymbol{x}})] - \underset{\boldsymbol{x} \sim P_r}{E}[D(\boldsymbol{x})]}_{\text{WGAN loss}}$$
$$+ \lambda \underbrace{\underset{\hat{\boldsymbol{x}} \sim P_{\hat{\boldsymbol{x}}}}{E}\left[(\|\nabla_{\hat{\boldsymbol{x}}}D(\hat{\boldsymbol{x}})\|_2 - 1)^2\right]}_{\text{gradient-penalty}} + \underbrace{\epsilon \underset{\boldsymbol{x} \sim P_r}{E} D(\boldsymbol{x})^2}_{\text{epsilon-penalty}} \qquad (14)$$

where, L_D is the Discriminator loss, P_r and P_b are the data distribution and model distribution coefficients respectively. ϵ is the epsilon penalty coefficient and λ is the gradient penalty coefficient adopted from [18].

Similarly, the WGAN-GP loss for optimizing the generator is defined as,

$$L_G(\tilde{\boldsymbol{x}}) = - \underset{\tilde{\boldsymbol{x}} \sim P_g}{E}[D(\tilde{\boldsymbol{x}})] \qquad \textbf{(15)}$$

also initially adopted from [18].

B. HyperParameter Tuning and Selection

To train the model, specific hyperparameters were analyzed and experimentally evaluated to yield best results on the training dataset. Eventually these hyperparameters were decided and used to train both the datasets. The Adam optimizer function was used in the model training with β_1 as 0.0 and β_2 as 0.99. The learning rate was a variable function, initially set to 0.001, and a learning rate decay was added at every resolution transition set as 0.87. The Batchsize was a dynamic variable set according to the available GPU RAM. The epsilon penalty ϵ was set to 0.001 and gradient penalty λ was set to 10. Feature Dimension of the Final layer of the discriminator and generator was set to 512. An additional noise vector of dimension 512 was added to the Network. Model Weights were updated every 5 epochs, and maximum resolution of the output image was set to 128 x 128px. The Arabidopsis Thaliana model was trained for approximately 21 hours (7500 iterations) in multiple sessions, while the Beta Vulgraris Dataset was trained for approximately 32 hours (14300 iterations) in multiple sessions as well. Pytorch Framework and Anaconda were used as supporting frameworks in training of the model. The model training was carried out on Google Colaboratory which is a Free platform for Students and Data scientists, offering Virtual Machines to conduct computationally expensive tasks such as Machine Learning model training. A Nvidia Tesla K80 GPU was used in the training of the model which took approximately 7-10 seconds per iteration.

5. MODEL EVALUATION AND PERFORMANCE

A. Loss function evaluation

The model loss function, gives as Wasserstein GAN with gradient penalty (WGAN-GP) 14 was used in mapping the loss function of the model. As mentioned in the earlier sections, a mini-batch standar deviation layer was used instead of the Feature Vector Normalization layer to increase the variations which results in a net increase in variation of the loss function as well. The loss function is also dependent on the data distribution coefficients calculated through latent space, which contribute in the overall increase of the loss function. The Arabidopsis Thaliana
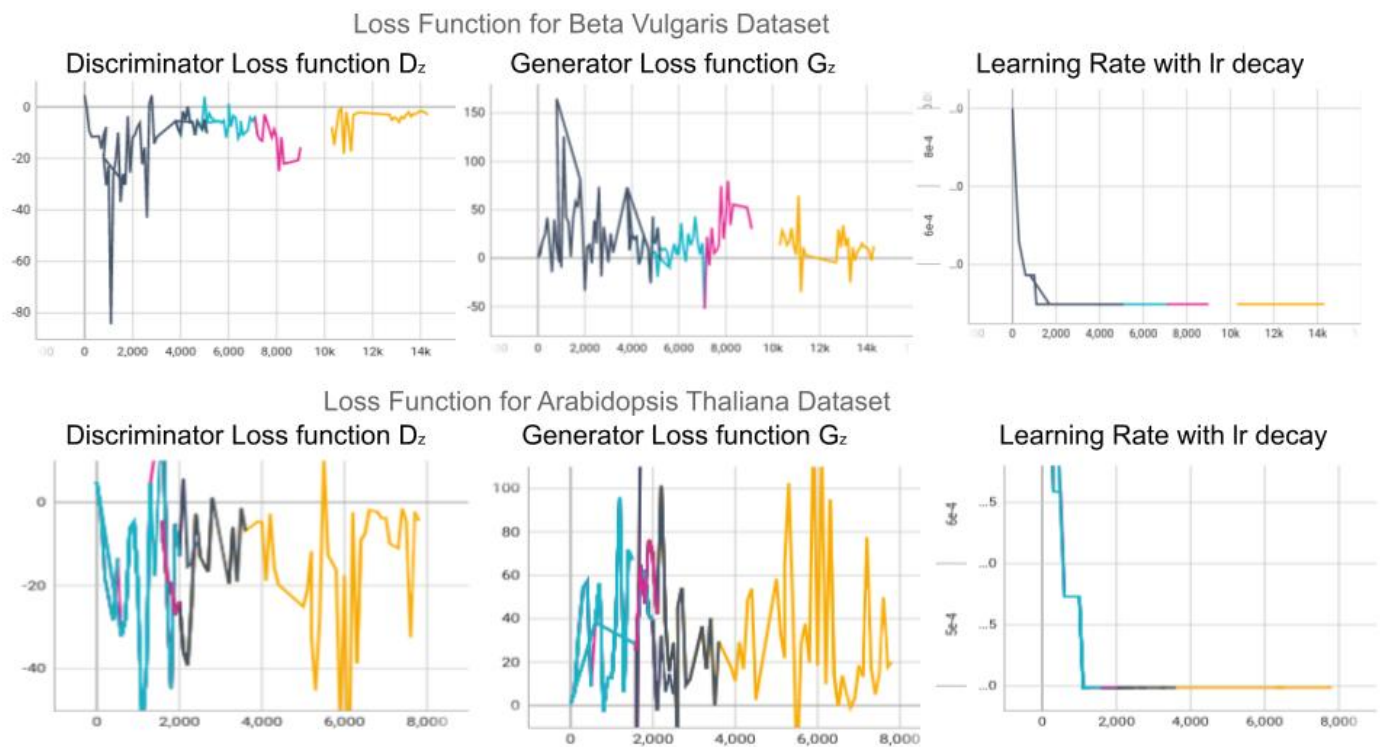
Fig. 23. Loss Function score plotted on tensorboard every 100 iterations. **Left** Discriminator Loss Function. **Centre** Generator Loss Function. **Right** Learning Rate with lr decay
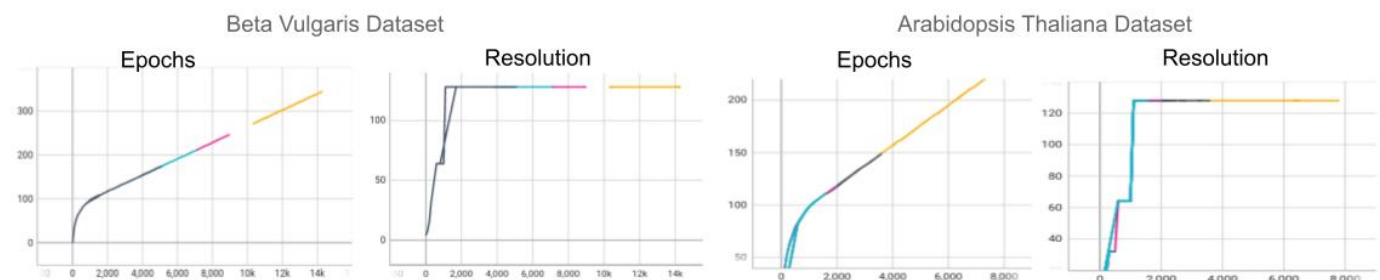

**Fig. 24.** Graph plotted on tensorboard for Epoch:Iterations and Resolution of Input/Output Frame:Iterations

dataset made use of 6 input frames and 6 output frames which increased the overall frame input and output nodes and feature distribution. This increases the variation in the loss and low power of loss convergence. However, this model performed exceedingly well on all other GAN evaluation metrics. The Beta Vulgaris model showed a standard convergence in loss observed in both, the Generator and Discriminator loss. The Discriminator loss was observed to be negative in both the Models. Since a Learning Rate Decay function (lr-decay) of 0.87 was added to the models upon transition in resolution, the learning rate of both the models decreased hierarchically until 128 x 128px resolution after which it remained constant. A gap was observed in the measured readings in the Beta Vulgaris Model for iterations from 9k to 10k, which occurred due to a minor glitch in the script. An additional python function was added to the main script to log all loss functions and other readings to tensorboard during training. The Tensorboard log file stored all tensor weights for log values during the training process. PGGANs and in general

all other GANs do not employ accuracy and precision scores as default scores to evaluate the performance of the GAN. Several other measures and Distance measurements are used to calculate the Simimilarity index between the generated frame and the ground truth frame.

During training in the generated samples, I noticed that in specific iteration gaps, different visual layers noticed certain biases. In some iteration layers, leaf details were intricately generated with high resolution while in few iteration samples, soil samples were better generated. This could be better explained by observing biases in different layers of the Progressively growing Generative Adversarial Network. The Encoder and Decoder imposed certain biases in specific networks arbitrarily due to change in variation magnitudes of Normalization values. When further analyzed, plots of these biases and weight magnitude distribution in intermediate layers revealed the average bias and weight magnitude distribution between encoder and decoder networks in Generator. This shift bias observed iteration-wise in
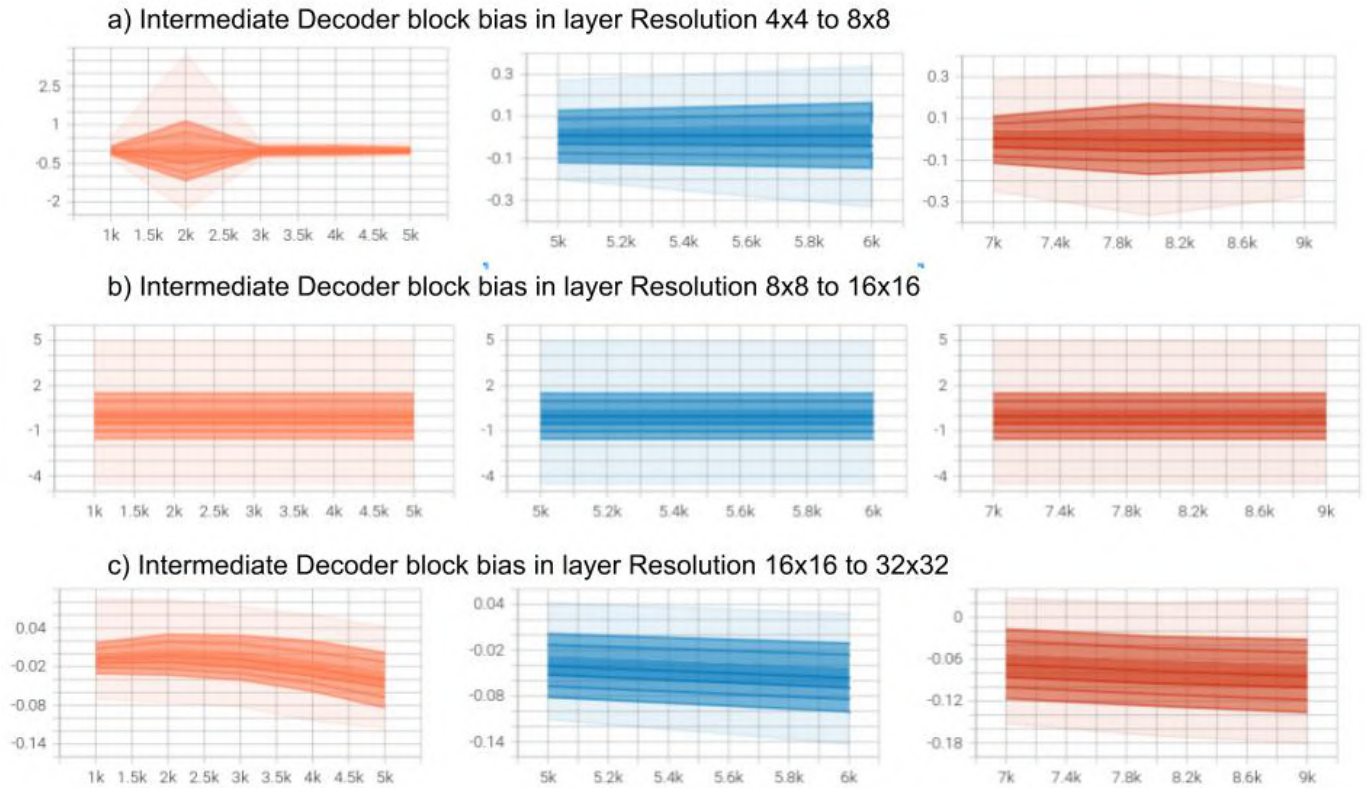
Fig. 25. Intermediate decoder block bias in layer resolution 4x4 to 32x32 (Generator)
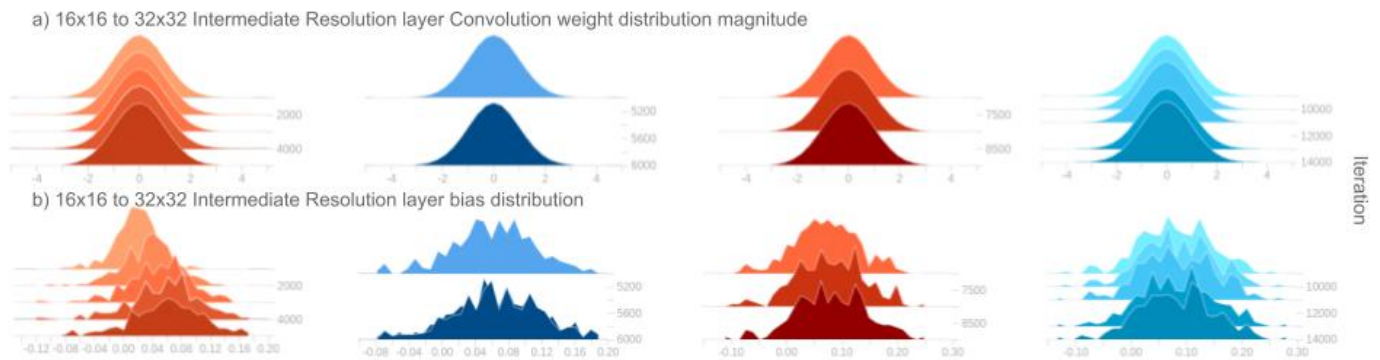
Fig. 26. Weight Distribution Magnitude and bias distribution in 16x16 to 32x32 Intermediate Resolution layer

plots could thus explain plausible visual change in feature distribution over plant networks. 25 and 26 concatenated iteration-wise Intermediate block layers reveal general bias distribution certain layer resolutions.

B. Evaluation and Similarity Metrics

Multiple Similarity, Ratio and Error evaluation metrics, apart from the loss function are further used to evaluate the model performance. The peak-signal-to-noise ratio (PSNR), structural similarity index measure (ssim) and mean squared error (mse) metrics were used to evaluate the model performance. The PSNR is defined as the ratio between the maximum possible power of a signal to the power of noise that affects the fidelity of signal representation. PSNR is based on Mean Squared Error formula. Generated images by GANs have a noise corrupting the signal representation, which is evaluated by PSNR. The Structural Similarity Index (SSIM) was also calculated which is a perceptual metric that quantifies the rate of image quality on a scale of 0 to 1, and is based on visible images structures. The MSE or the Mean squared deviation is the most commonly used evaluation metric in GANs which evaluates average squared difference between the predicted image and the ground truth image. These metrics can be defined as,

$$\text{PSNR} = 20\log_{10}\left(\frac{\max_f}{\sqrt{MSE}}\right) \tag{16}$$

$$\text{SSIM}(f,g) = \frac{\left(2\mu_f\mu_g + c_1\right)\left(2\sigma_{fg} + c_2\right)}{\left(\mu_f^2 + \mu_g^2 + c_1\right)\left(\sigma_f^2 + \sigma_g^2 + c_2\right)} \tag{17}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{18}$$

where, μ_x the average of x, μ_y the average of y;, σ_x^2 the variance of x,, σ_y^2 the variance of y;, σ_{xy} the covariance of x and y; $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ are two stabilizing variables in weak division with denominator. These metrics were evaluated for both the datasets and the results were plotted visually.?? 27

These results evaluated in the models were compared to other benchmark results for predicted frame types on different datasets. The compared results can be presented as,

Table 4. Proposed results compared to other PGGAN predicted results on different datasets. Results averaged over all frames.

Framework	MSE	SSIM	PSNR
FutureGAN	0.1603	0.7780	14.335
CopyLast	0.2580	0.679	11.723
fRNN	0.1854	0.7408	13.7345
MCNet	0.0048	0.8692	-
Beta Vulgaris[c]	0.007761	0.7286	27.4274
Arabidopsis Thaliana [c]	0.0098036	0.5620	26.3232

A quantitative Analysis of different initial proposals of Temporal PGGANs and a benchmark comparison for model performance evaluation was conducted. The above models excluding mine were trained on datasets « 6000 image samples compared

to mine > 700 image samples. Also, these models were trained for a minimum of 500 epochs compared to mine at 230 and 365 epochs for Arabidopsis Thaliana and Beta Vulgaris respectively. The model did not perform as well in SSIM metric, but outperformed the other models in MSE and PSNR by a significant difference. This was used to evaluate the initial efficacy of the model as a whole and how robust the model is compared to benchmark scores. [c] represents the scores of the "current" model used in the project.

## C. Visual Evaluation of the Generated Model Frames

The model evaluation was not only limited to these metrics but also how the model was visually represented and visual generated traits as well as generative morphology. These generated visual traits are further evaluated based on visual comparison methods and PLA (Projected Leaf Area) as well as on numerous Vegetation Indices, Fréchet inception distance (FID) and other soil traits and reflectance metrics.

### C.1. Generated Frames for Arabidopsis Thaliana Dataset

This dataset was modelled in a way such that the prediction of frames from the input 6 frames is at an hourly interval from the preceding frame. This approach had been kept experimental, which was later made concrete to generate Future n frames at an interval of 1 hour based on input frames at intervals of an hour, where n is 6. Contrary to this approach, the second dataset predicted the future frames on a biweekly basis so there was a significant generative gap between the input and generated frame. Both these PGGAN output node type models were adopted to evaluate the model efficacy in two different domains simultaneously increasing the overall robustness of the proposed system. An example of Ground Truth Input and Output frames compared to Generated output frames is presented in 28 and 29. The model training output feature correlation was not constant for each frame in each iteration. Since the GAN training is an iterative process, the model generated features were progressively improving over each epoch and model weights were updated after every 5 epochs. The generated frames by the GAN for Arabidopsis Thaliana dataset were realistic and corresponded visual traits as possessed by the Ground Truth frames at epoch 230 (final epoch of the training procedure). The position, growth structure, solid color and leaf count show high correspondence to the Ground Truth Frame. However, human biological correspondence based on visible plausible features is sometimes biased, and so I used a number of Vegetation Indices to establish correspondence with the predicted and ground truth frame over all generated frames. Prior to that, the proceeding Image cluster accurately demonstrates evolution of Generated frames from 4x4 resolution all the way upto 128x128 iteratively.

The iterative process of PGGAN generating images from epoch 0 to epoch 230, at which the training was halted revealed the heuristic evolution of generated features by the PGGAN model. The model picked up features from each plant images that had less complexity and variance. At first, the color of the soil samples and confined boundary of the sample were generated evenly for plants observed till epoch 98. However, an outlying spectral noise on generated soil samples could still be clearly observed. Till epoch 111, the noise was reduced and relatively the PSNR for generated soil values had increased with the decrease in noise. The model learnt leaf boundaries by epoch 126 which initially observed constraints in features. This point entailed an extended bias to features learn for leaves and petioles, proceeding the feature bias towards soil sample. By epoch 145,
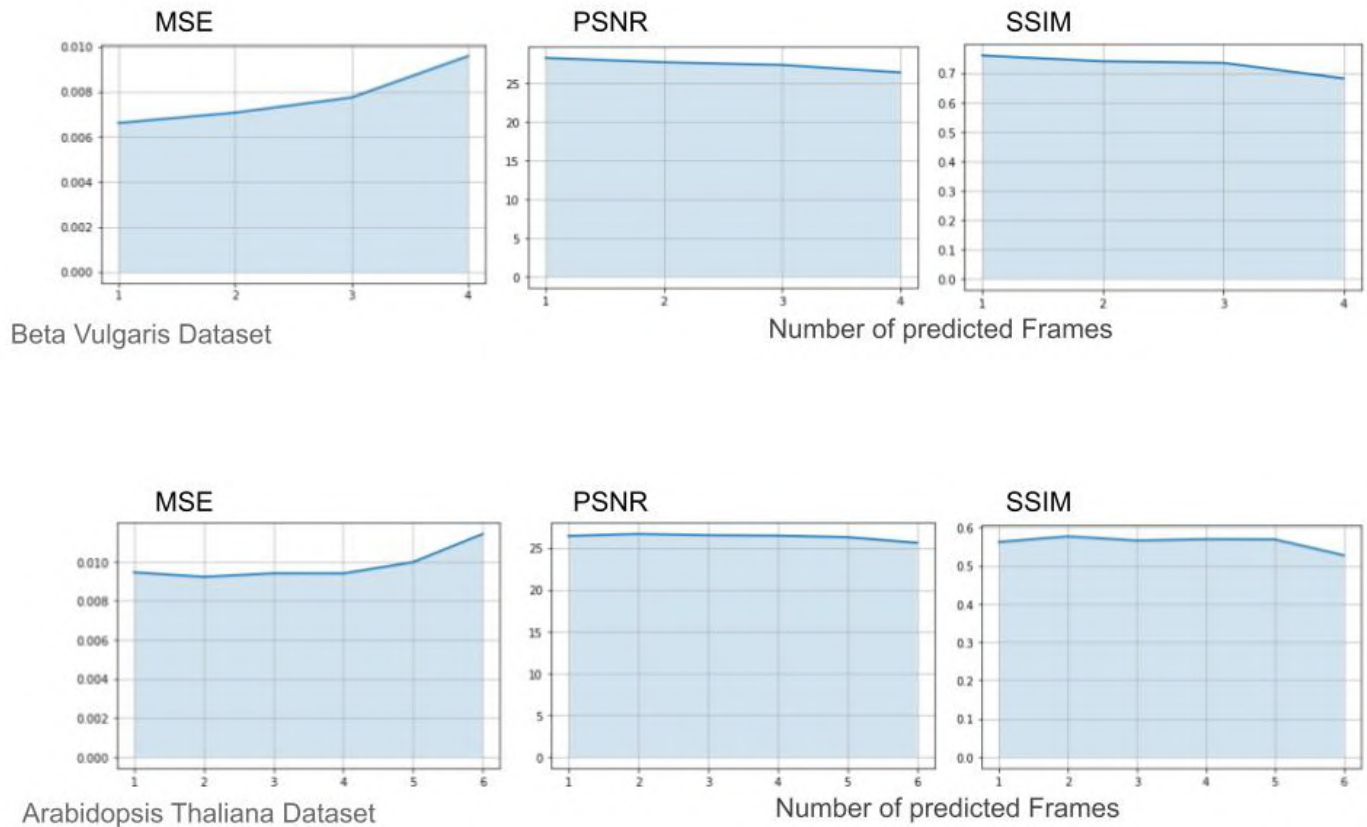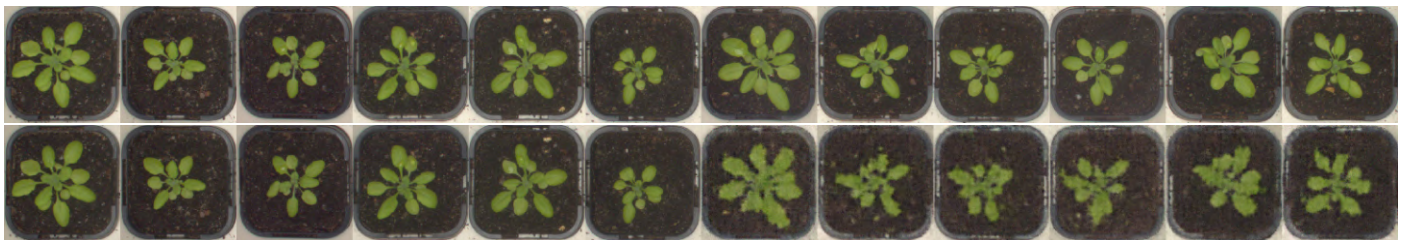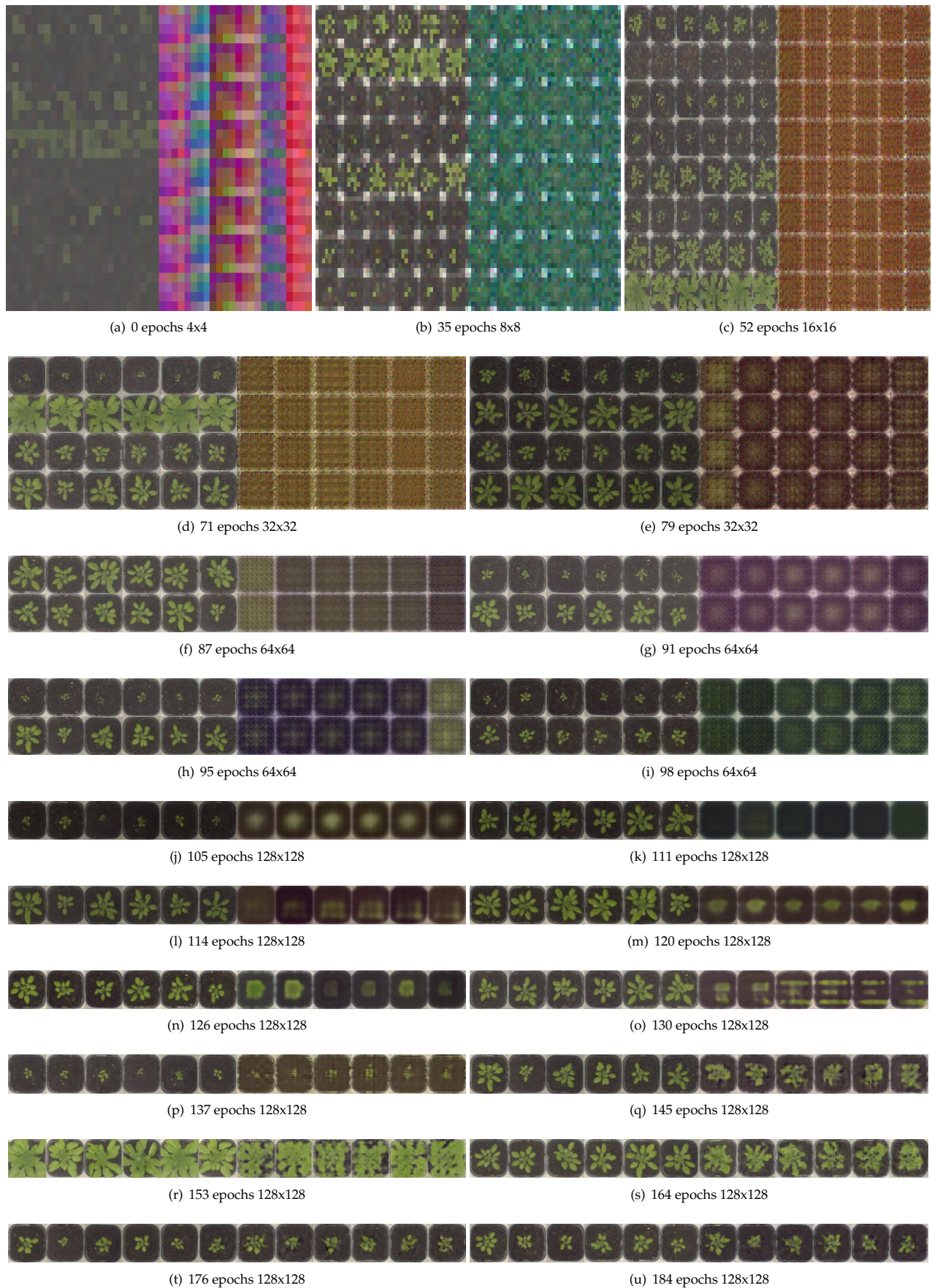
Fig. 27. Calculated metrics, MSE, SSIM, PSNR for Arabidopsis Thaliana and Beta Vulgaris Dataset
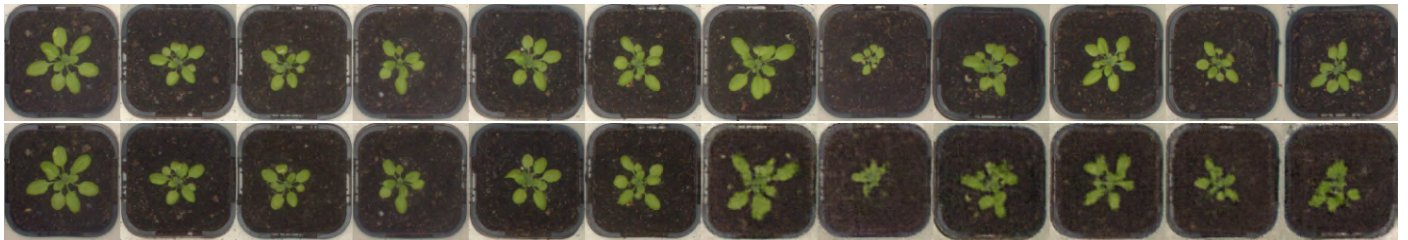
Fig. 28. Ground Truth Input frames compared to Generated proceeding frames by PGGAN network. **Top** Ground Truth Input 6 frames and Ground Truth expected prediction of 6 frames for comparison purposes. **Bottom Left** Ground Truth Input 6 frames to the PGGAN model. **Bottom Right** Generated Images by the PGGAN Network based on the images provided.


Fig. 29. Ground Truth Input frames compared to Generated proceeding frames by PGGAN network. **Top** Ground Truth Input 6 frames and Ground Truth expected prediction of 6 frames for comparison purposes. **Bottom Left** Ground Truth Input 6 frames to the PGGAN model. **Bottom Right** Generated Images by the PGGAN Network based on the images provided.
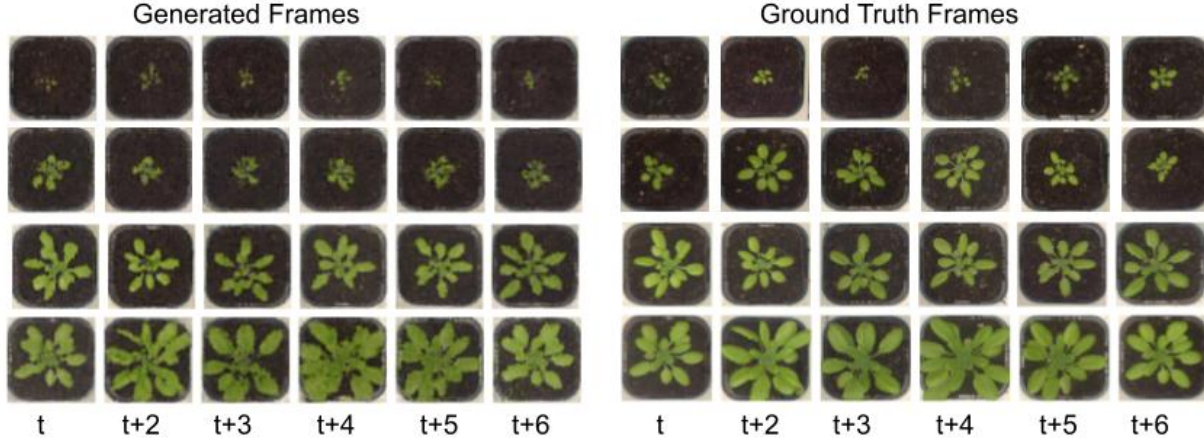
(a) 0 epochs 4x4

(b) 35 epochs 8x8

(c) 52 epochs 16x16

(d) 71 epochs 32x32

(e) 79 epochs 32x32

(f) 87 epochs 64x64

(g) 91 epochs 64x64

(h) 95 epochs 64x64

(i) 98 epochs 64x64

(j) 105 epochs 128x128

(k) 111 epochs 128x128

(l) 114 epochs 128x128

(m) 120 epochs 128x128

(n) 126 epochs 128x128

(o) 130 epochs 128x128

(p) 137 epochs 128x128

(q) 145 epochs 128x128

(r) 153 epochs 128x128

(s) 164 epochs 128x128

(t) 176 epochs 128x128

(u) 184 epochs 128x128

Fig. 30. Generated morphological Images of PGGAN from epoch 0 to 184 including resolution 4x4 to 128x128px

(a) 197 epochs 128x128

(b) 214 epochs 128x128

(c) 220 epochs 128x128

(d) 230 epochs 128x128
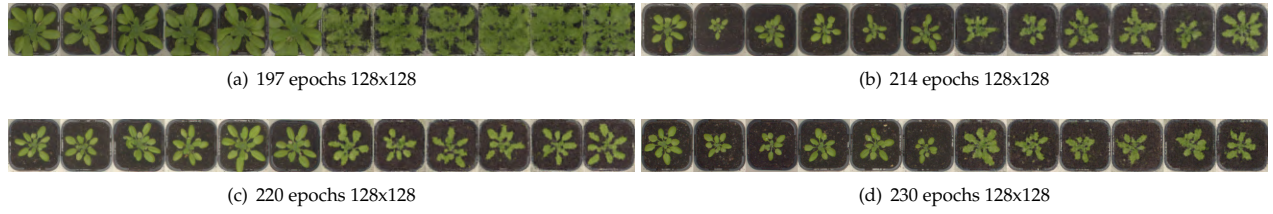
Fig. 31. Generated morphological images of PGGANs from epochs 197 to 230 at resolution 128x128

Fig. 32. left Generated frames by PGGAN for different growth stages of the plant. **Right** Ground Truth frames for comparison. Note: Input 6 frames aren't displayed in the image.
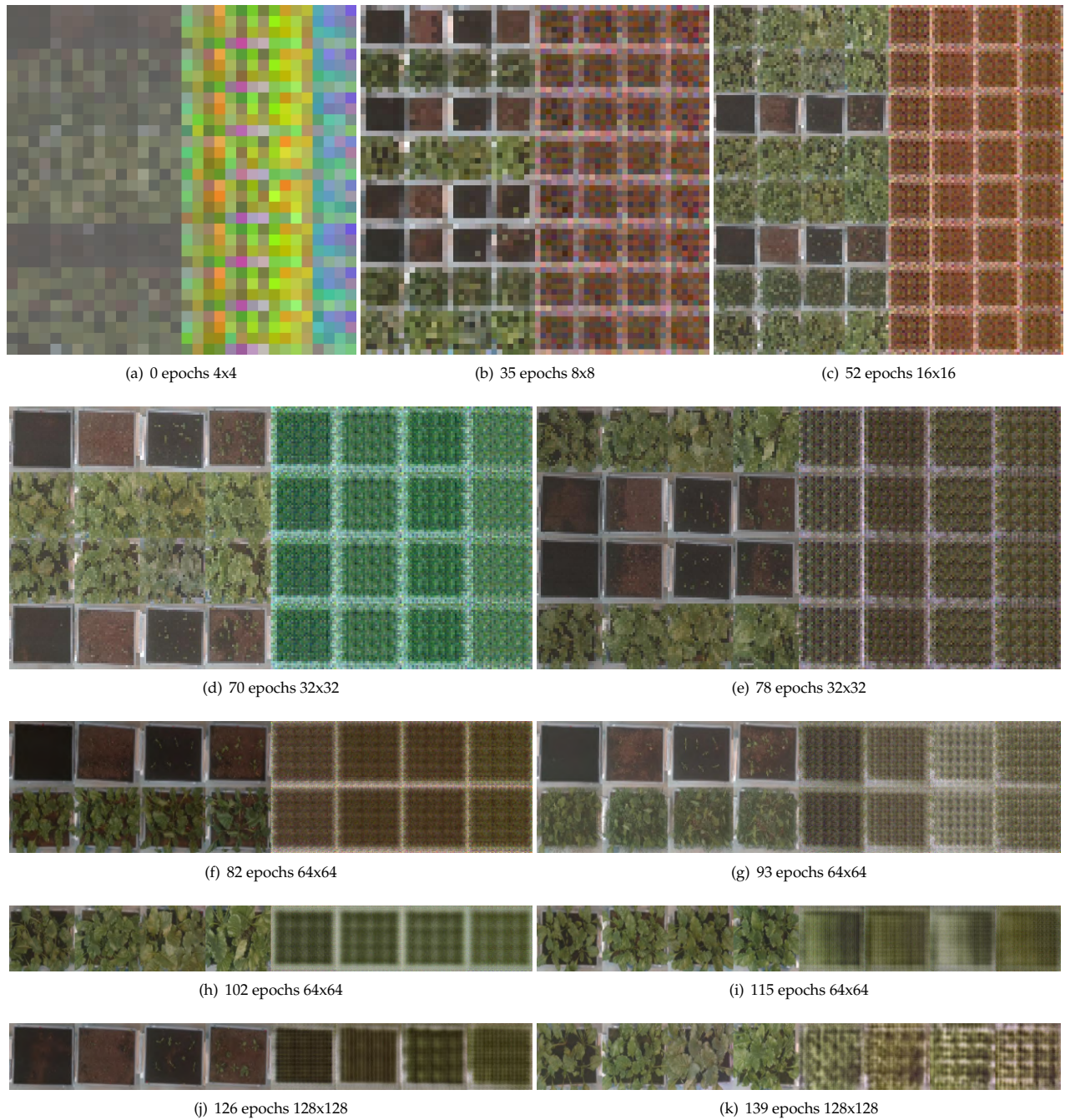
the GAN could generate a close representation of the ground truth frames, however it lacked in generating features necessary to meet the goal of the project. These features would differ for each plant sample and would contribute the most in evaluating the nature of plant growth and responsive stimuli. The PGGAN initially started to learn external visual features affecting of hindering growth patterns such as leaf orientation of soil type and color intensity as well as density of leaf count. These external features even though present more abundantly in the second dataset, this dataset observed growth patterns dependent on the external features as well. Features for these learn external representations showed initial appearance at 197 epochs and were completely developed by 230 epochs. Few generated samples still observed raw undefined leaf boundaries in certain arbitrary samples, however these samples showed close resemblance to visual traits important in deciding plant stimuli towards environmental variables or induced conditions. While existing frame generation approaches explained in "Related Work" aim to predict a single frame after t days, this approach predicts 6 frames till t^{th} day as well as the frame on t^{th} for a given defined interval between frames. This not only helps in understanding the harvest outcome on t^{th} day but also helps in understanding the pattern of growth between the final day from the initial frame. Fig 32 shows comparison of t to t+6 frames generated by PGGAN to ground truth frames. The generated frames show higher correspondence to Ground frames for higher Leaf Area occupied by the Arabidopsis Plant. For Rows 3 and 4, the predicted outcome is visually similar to ground truth.

The Generated morphological images in the spatio-temporal domain can be evaluated for structural resemblance from the ground truth frame in multiple different methods. Human biological mind has different expectations of generated frames in

terms of reliability, resemblance, generated stimulus of the plant, which corresponds to actual nature of growth in images, ensuring the robustness of the proposed system in multiple topologies. Alongside predicting growth pattern dependent on visual traits, the model is compared to have a realistic generation in terms of different aspects of the plants. Also, fig 32 ensures that plant growth generation is not limited to a certain time frame and is robust, and applicable over all growth stages of plant growth till harvest. Eventually the goal of this research is to implement autonomous monitoring systems to predict and generate plant growth stages before harvest in multiple varieties of plant to understand the positive and negative impacts of environmental variables in plants and yield contraction. Assessing plant harvest and growth pattern till harvest through visual traits of germinated seeds, which help in identifying early wilting or yield contraction or environmental stresses hindering growth.

C.2. Generated Frames for Beta Vulgaris Dataset

This dataset involves much higher feature complexity as compared to the previous dataset. Also, this dataset is considered to be the epicentre of testing the robustness of the system. The PGGAN model for this dataset was created to take an input of 4 frames and predict 4 frames with a duration of 4 days between two frames. This means the interval between the first input frame and the last predicted frame is approximately 28 days, indicating that the model can generate growth patterns of plants 28 days into the future. An experimental evaluation of both the models could conclude that this model had higher consistency in generating patterns of growth stages. Similar to the first dataset, this was an iterative process from 0 to 365 epoch for almost 15400 iterations which spanned twice as much as compared to the Arabidsopsis Thaliana dataset. The model weights were each

(a) 0 epochs 4x4   (b) 35 epochs 8x8   (c) 52 epochs 16x16

(d) 70 epochs 32x32   (e) 78 epochs 32x32

(f) 82 epochs 64x64   (g) 93 epochs 64x64

(h) 102 epochs 64x64   (i) 115 epochs 64x64

(j) 126 epochs 128x128   (k) 139 epochs 128x128

Fig. 33. Generated morphological Images of PGGAN from epoch 0 to 139 and progressively increasing image as well as layer resolution from 4x4px to 128x128px doubled after every updated weight.

(a) 148 epochs 128x128

(b) 156 epochs 128x128

(c) 167 epochs 128x128

(d) 174 epochs 128x128

(e) 191 epochs 128x128

(f) 196 epochs 128x128

(g) 206 epochs 128x128

(h) 215 epochs 128x128

(i) 228 epochs 128x128

(j) 239 epochs 128x128

(k) 245 epochs 128x128

(l) 254 epochs 128x128

(m) 263 epochs 128x128

(n) 274 epochs 128x128

(o) 293 epochs 128x128

(p) 308 epochs 128x128

(q) 322 epochs 128x128

(r) 341 epochs 128x128

(s) 348 epochs 128x128

(t) 365 epochs 128x128

Fig. 34. Generated morphological images of PGGANs from epochs 148 to 365 upto 15400 iterations at resolution 128x128 and constant learning rate as well as perpetual converging loss function.
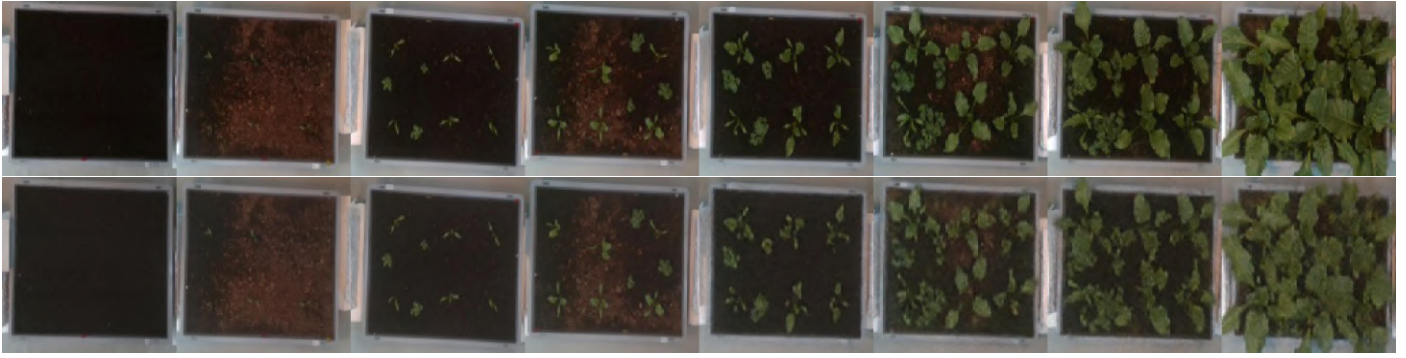
Fig. 35. Ground Truth Input frames compared to Generated proceeding frames by PGGAN network for Beta Vulgaris Dataset. **Top** Ground Truth Input 4 frames and Ground Truth expected prediction of 4 frames for comparison purposes. **Bottom Left** Ground Truth Input 4 frames to the PGGAN model. **Bottom Right** Generated 4 frame Images by the PGGAN Network based on the images provided.

updated after 5 epochs and the loss logs were saved after every 100 iterations. The Progressively Growing GAN started initially from 4x4 epochs to 128x128 epochs, and in this period the learning rate decreased sequentially due to lr decay to avoid model overfitting. It is worth to note that the images were cropped to size 870 x 590, which were reduced to 190 x 128 maintaining ratio aspect, further the model squishing the input image size to 128x128 as a standard to ensure standardized training of the frames using fitting the smallest width approach. Similar to the training procedure of the previous dataset, this model observed temporally varying bias towards complexity in features learnt, which would be observed in histogram plots in Figure 25 and 26. Uneven feature bias among layers of different resolution spread over encoders and decoders discretely allowed feature variation and feature priority over expected features ubiquitously over data points in the latent space. The feature bias was centred to feature distribution over soil variation till a certain iteration point of 126 epochs. Following this, the iteration distribution bias is more contributing towards learning leaf and venation features rather than soil features. The epoch evolution from 126 to 365 epochs, involve feature bias distribution to leaf generation variation towards environmental variables. Leaf features such as leaf density, color, saturation, and natural stimulus towards environmental elements such as photropism or hydrotropism observed through leaf orientation and direction. The model used 4 input frames and generated 4 output frames, while the dataset included 16 frames for each plant from germination till senescense in plants. Hence, the model split a plants growth stage in two parts with one part predicting frames 5 to 8, while the other predicting frames 13 to 16. The frame 5 to 8 show large variation in generated visual output. Example, weeds from germination become clearly visible pointing out low quality or soil sample. However, the effect of Nitrogen or Low Hydration on leaf wilting is not clearly visible in this frame sample. In frames 9 to 16, with 13 to 16 as predicted frames, the effect of weed stress "visually" on the plant reduces, while the most distinctive features, low Hydration followed by Nitrogen deficiency is observed in the trend. These environmental stresses induced by water and nitrogen become visually striking and observable in the latter part of the growth stage due to increased and prolonged exposure of observed stress. These stresses usually are "chronic" and follow the plant until an environmental variable is altered. Comparing it with weed stress, based on observed sample readings, it could be observed that weed stress plays

important role in inducing "visual" constraint on initial growth stages of plant. The complete observation over these samples points out the distributed strain imposed on plant in varying plant stages by Environmental stresses. The observed visual features were clearly visible on generated visual cues of images generated by the PGGAN model for Beta Vulgaris dataset. An extent of effect varying with time can be observed as.

$$3days > \mathbf{Ws}_{cstr} > 7days$$

$$6days > \mathbf{N}_{cstr} > 12days$$

$$7days > \mathbf{H}_{cstr} > 16days$$

These observations were manual observations as seen over all samples in the dataset. These constraints overlap growth duration, and some affect the crop growth simultaneously based on the cultivar where each constraint is imposed. Here, since the duration between two sample images captured was 4-5 days, an observable change in features were visible in analyzing the Generated Images. By epoch 148, the soil bias had nearly transitioned to leaf feature bias (bias corresponding to eccentric distribution of feature learning mechanism over samples in latent space). By epoch 206, the model achieved highly similar feature generation over all features but with lot of noise and few distortions. However, an additional bias in learning features for frame 1 to 8 - 9 to 16, the early bias from around 100 epochs to 220 epochs contributed in learning features for 1 to 8. These features, since most distinctively visible, could be learnt easily and generated images would contain less noise. Comparing that to frames in 9 to 16 growth stage, these features were not so distinct, were enigmatic to learn and hence took an additional 100 epochs more as compared to 1 to 8 frames, after which the features and visual morphology could be visually distinguished. The plants with higher intensity and number of stresses could observe wilting of leaves, color fade and low ExG (Excessive Green Vegetation) and more soil sample space. However, the cultivars with variables under control observed high leaf density, leaf overlapping and increased occlusion, higher distributed concentration of artificial light and anomalous leaf reflection over images. With increase in Projected leaf Area, and current Observable Leaf Area Index, the volume of the plant increased as well, which wasn't distinctively visible in the dataset. Sans these visual cues of depth in RGB datasets, the model could learn growth morphology with features correlating to specific

Drying, Medium Nitrogen and High Weed Beta Vulgaris Cultivar
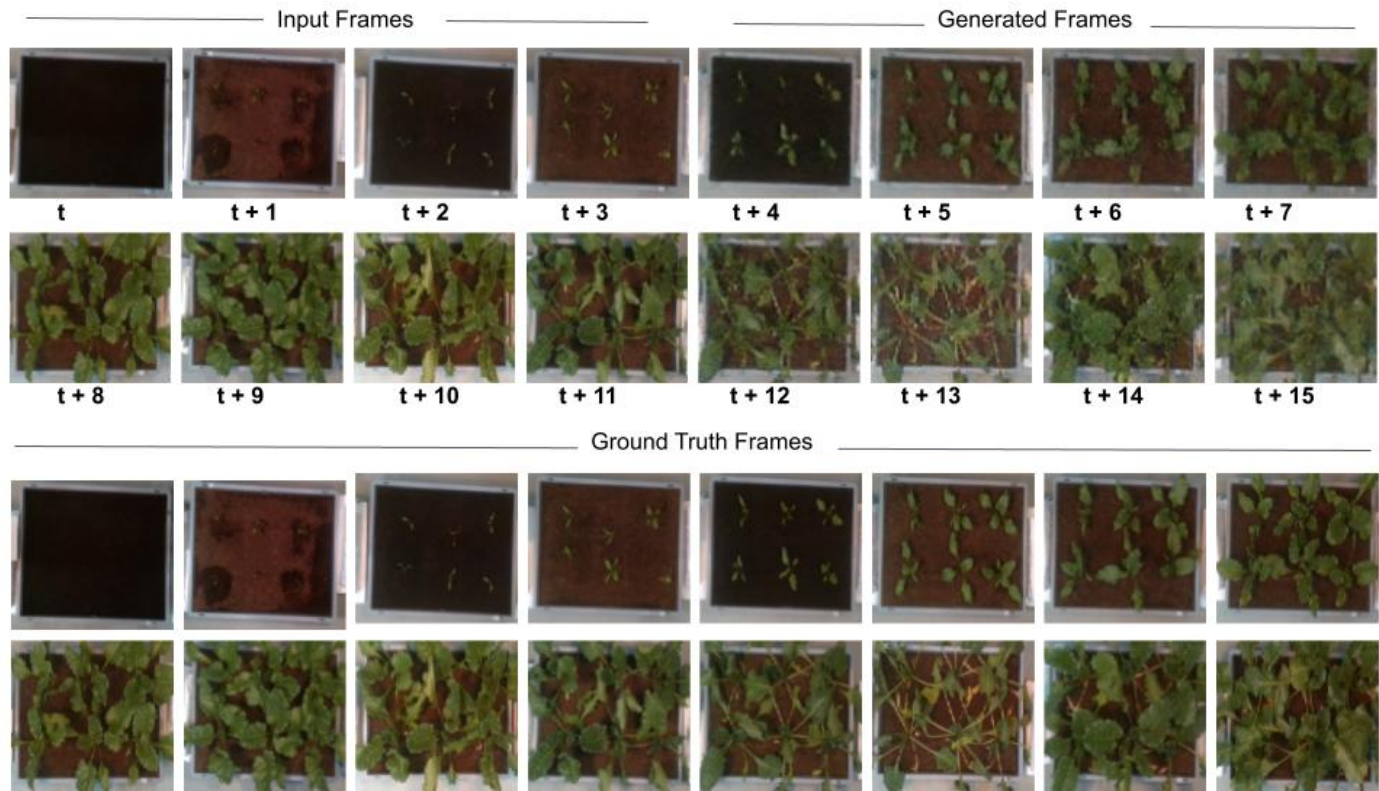
Fig. 36. Generated frames by the PGGAN model for beta vulgaris dataset for Drying, Medium Nitrogen and High Weed Cultivar

plant and produce accurate generation. Irrespective of the continuous increase in volume over all cultivars as a function of time, the visible leaf area in the ROI (Region of Interest) were not constant leading to certain anomalous trends in the latter frames. The comparision of the Vegetation Indices (ExG, ExGR, ExR and Area) is more deterministically measured for the initial frames (an upper limit of 9 frames) and the latter frames observe high entropy in volume change, but less observable leaf area change in the frames. For the results we tested on this dataset, despite the low predictive power of Leaf Area in the latter frames, the generated frames could observe similar trend in leaf area generation and hence the covariance between this Index for latter frame features could be established making the evaluation process in the model testing robust as well. An extensive analysation of the generated 36 explains how weed density contributes most to visual features in the frames 5 to 7, while the Hydration level affects the crop most in 13 to 15 the frame and is viusally visible thereafter. The model adapts to orientation of the cultivar, and also changes in dimension or illumination. With increase in overall dataset, the model can simultaneously learn augmented features or difference in feature concentration which is useful in serving the aim of identifying environmental variables conducive for plant growth and predicting harvest through visual features to test the best environment for hybrid or Genetically Modified (GM) plants (or in that case any) plant, weeks before harvest. This is beneficial to test resilience of the new breed or type of plant under different conditions with observed data for previous growth and generate spatio-temporal images to predict yield weeks before harvest to understand growth weeks before

harvesting the crop or days withing crop growth.

D. Testing model on Vegetation Indices and correlation of traits

In this subsection of the Model Evaluation section, I propose multiple Vegetation Indices and correlation coefficients to test the model performance on the Generated frames with respect to the Ground Truth frames, and most importantly draw conclusions (for Beta Vulgaris Dataset) for which environmental variables contract the yield and which of them propagate or induce maximum yield and growth, and how these conclusions deduced compare to the Ground Truth recorded measurements of favourable environment these crops grow in.

Analysing these Generated frames on multiple Vegetation Indices involved a thorough research on which Vegetation Indices can analyse generated frames without any bias over feature importance and are widely accepted indices on which these plants can compare upon- with other species in a database and draw meaningful conclusions interpretable over other plant experiments. Before finding the optimum Vegetation Indices, analyzing the data on which evaluation was to be done, was important. There were multiple caveats where implementing vegetation indices, estimating background feature contribution to threshold of the vegetation index, or artificial saturation and induced light contribution to determine the variable threshold- and many constraints contributing to selecting the optimum vegetation index.

Before coming across the optimum Vegetation indices, I tried to implement a few Adaptive threshold algorithm, one of them

being the Otsu's Threshold which can binarize discrete pixels and segment background from foreground in certain image types. As defined by the author, the threshold segments the foreground and background by maximizing the inter-class intensity variance or minimizing the intra-class intensity-variance, which in short terms binarizes the background and foreground. The Otsu's threshold algorithm can be defined as,

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \qquad (19)$$

where, ω_0 and ω_1 are probablities of two classes, and t is the threshold.

and the implementation of the algorithm in python is,

Algorithm 1. Otsu's Threshold

1: **otsu_threshold, image_result** = *cv2.threshold*(image, 30, 255, cv2.THRESH_BINARY + **cv2.THRESH_OTSU**)
2:

The formulated algorithm was implemented in a python code and deployed over a set of images to evaluate the segmentation results. There were several loopholes hindering the efficacy of the threshold.
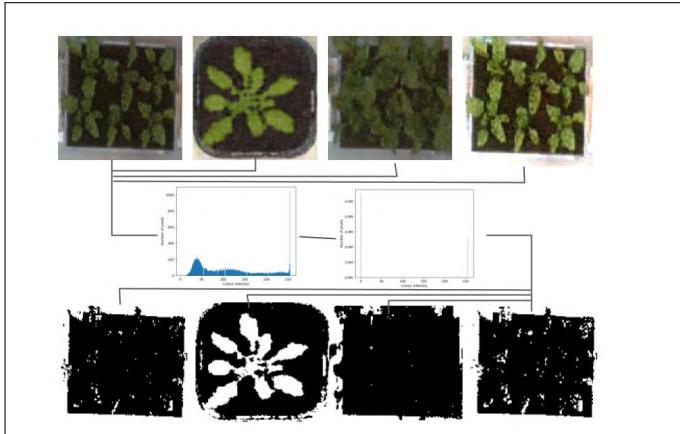
Fig. 37. Otsu's threshold pipeline and segmented frames using the Adaptive Threshold algorithm showing low efficacy.

The Arabiopsis Thaliana dataset and Beta Vulgaris dataset showed extremely low efficacy in segmentation using otsu's threshold. While the plant leaves and soil sample serve as distinctive foreground and background, it isn't possible to understand this difference in intensity due to complex features in the plant soil and pot sample. The Otsu's threshold algorithm didn't prove to be accurate in leaf segmentation from the soil sample.

After conducting sufficient research in evaluating vegetation, multiple vegetation indices proposed in [19] and PhenoBot were adopted to evaluate the efficacy of the model. Vegetation indices such as ExG (Excessive Green Vegetation Index). ExR (Excessive Red Vegetation Index) and ExGR (Difference of Excessive Green and Red) also described as ExG - ExR. The minimum, maximum and mean values of these Vegetation Indices was calculated as well to understand pixel-level feature vectors of these vegetation indices.

r, g and b chromatic color co-ordinates of an image were used to calculate the Vegetation Indices of the plant. The r, g and b co-ordinates were calculated as,

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \qquad (20)$$

$$R = \frac{R_c}{R_m}, G = \frac{G_c}{G_m}, B = \frac{B_c}{B_m} \qquad (21)$$

R, G and B pixel values were normalized from 0 to 1 and the r,g and b values were calculated to formulate the ExG, ExR and ExGR equation. Similarly, $\mathrm{R_m}$, $\mathrm{G_m}$, $\mathrm{B_m}$ are constants at 255, taken as maximum tonal intensity for each color channel inspired from [20]

ExG, ExR and ExGR were calculated as,

$$ExG = 2g - r - b$$
$$ExR = 1.4r - g$$
$$ExGR = ExG - ExR$$

Algorithm 2. Vegetation Indices

1: blue = img[:,:,2]
2: green = img[:,:,1]
3: red = img[:,:,0]
4: exg = 2*green - red - blue
5: exr = 1.4*red - green
6: exgr = exg - exr
7: exgr = np.where(exgr < 25, 0, exgr).astype('uint8')   ▷ Thresholding removes low negative values (noise)
8: exgr = exgr.astype(np.uint8)   ▷ convert back to uint8
9:

The Algorithm can be understood as 3 channels of the image, r,g and b area extracted and the pixel values as arrays are used to calculate the vegetation indices. For the final value of ExGR, additional noise is filtered and the arrays are converted back to uint8 image format to be further analysed for calculation of area. The ExGR converted image format is used to find the leaf area occupied by the plant, which shows high efficacy with ground truth segmented values and is used as standard index to calculate the Projected Leaf Area (PLA) and Leaf Area Index (LAI). Another experimental algorithm was used to find the visual NIR from RGB image. A red colormap was applied over the segmented image, split to three Red Color bands simulating the Red and NIR bands in IR images and the final NIR segmented image was obtained. However, since the process was clearly experimental and the algorithm was not peer reviewed, my algorithm hasn't been used for evaluation in this experiment. However, visual results for the algorithm are presented.

The algorithm can be understood as follows,

Algorithm 3. Visual NIR Algorithm

1: img_c   =   cv2.applyColorMap(exgr, cmapy.cmap('Reds')).astype(np.int)
2: _, R, NIR = cv2.split(img_c)
3:

Figure 40 shows high efficacy in segmenting leaves and plant features from the background through ExGR and a colormap is imposed on the set of segmented images to highlight the intensity of ExGR value through pixel color intensity. Clearly, the Arabidopsis Thaliana dataset shows high ExGR compared to Beta Vulgaris which might be visually evident observing the
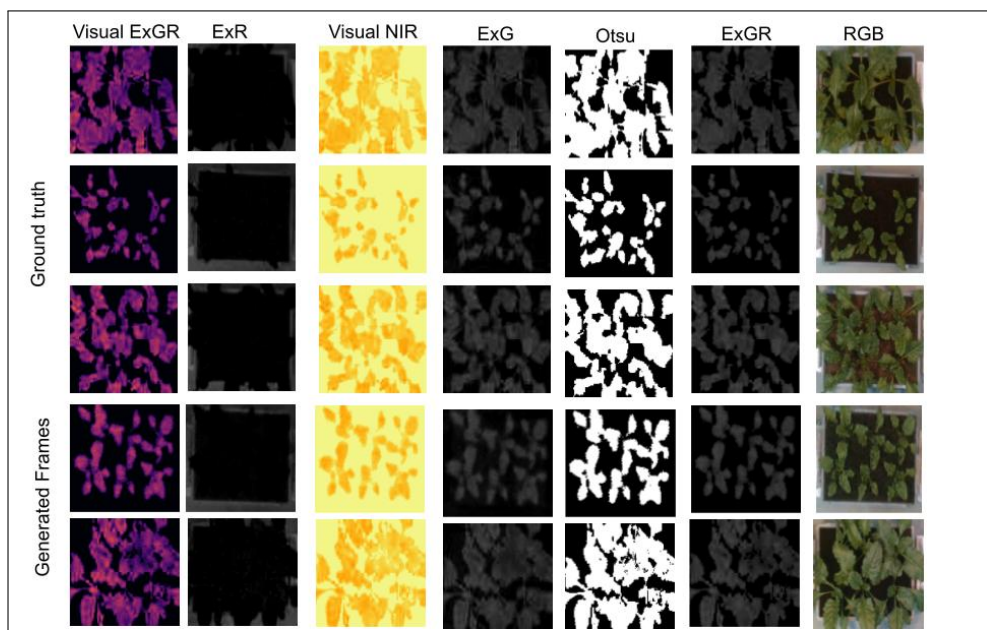
Fig. 38. Vegetation Indices including ExGR, ExR, ExG, Visual NIR and Visual ExGR (intensity of ExGR pixels) applied on Beta Vulgaris dataset.
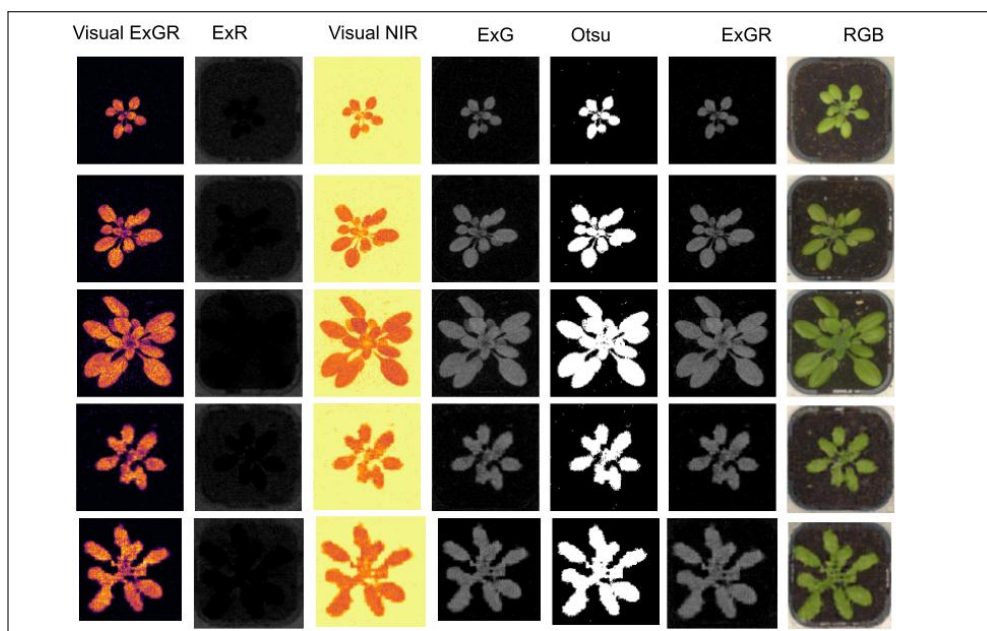
Fig. 39. Vegetation Indices including ExGR, ExR, ExG, Visual NIR and Visual ExGR (intensity of ExGR pixels) applied on Arabidopsis Thaliana Dataset
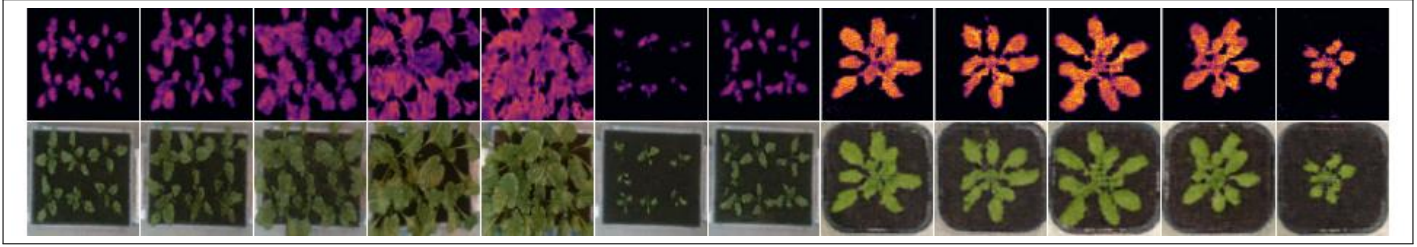
Fig. 40. Visual ExGR evaluated for the whole dataset. Truncated to 5-7 generated images for better interpretation. **Left** Visual ExGR on Beta Vulgaris dataset. **Right** Visual ExGR on Arabidopsis Thaliana Dataset. All images are generated by the PGGAN model.
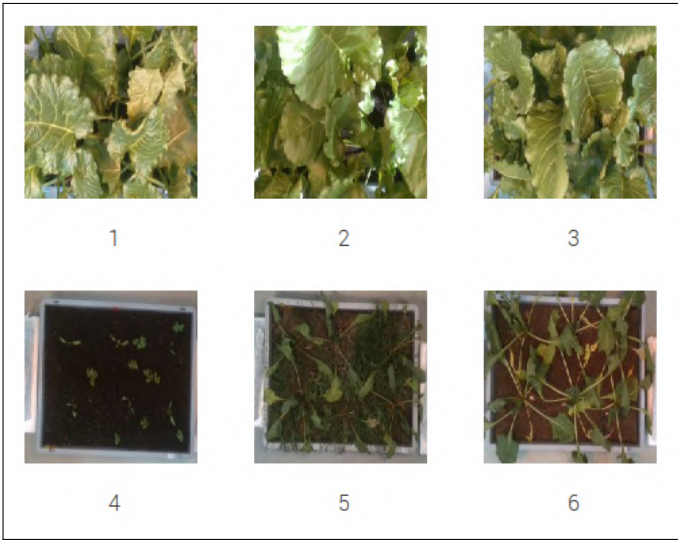
Fig. 41. Compared to Otsu's Threshold, the Visual ExGR performs exceedingly well in segmenting images with complex features, differing illumination and small visible plant regions as well as wilted or faded leaves.

color of the pixels and leaf reflectance over both images. Since the ExGR is merely used as a comparison metric to evaluate the accuracy of the generated images through feature vectors and ground truth observations, minor ExGR deviations over different sample subsets aren't altered and kept as is during evaluation.

A python script is constructed to enable autonomous segmentation of images in the Arabidopsis Thaliana Dataset and Beta Vulgaris Dataset. The images are segmented and autonomously stored in a folder and the numerical data points for the values of ExG, ExGR, ExR, area and min, max, mean functions of these indices with corresponding input frame is stored in a csv file. The area of the Visual ExGR is simply calculated by applying a threshold to the image, and using a function to calculate white pixels (of the leaves) threshold(ed) by the function.

Finally, the Vegetation indices from Generated Images were evaluated against Ground Truth image indices and high correlation among the generated image indices and ground truth indices was observed. This correlation was specifically much higher for the Beta Vulgaris dataset, in an overall comparison between all cultivars as well as in stress induced ($\mathbf{Ws}_{cstr}$, $\mathbf{N}_{cstr}$, $\mathbf{H}_{cstr}$) cultivars.

While evaluating the Generated values with Ground Truth, ExG and ExGR showed direct proportionality with the predicted area, while ExR showed inverse variation. the min and max val-

ues showed no specific correlation, because certain minimum or maximum values for these Vegetation Indices were observed to be arbitrarily high in certain samples. However, in Beta Vulgaris dataset, the max values for ExG and ExGR showed somewhat steady trend. Irrespective, the mean values were used as default indices to compare the model efficacy since a lot of entropy in data points over min and max values were seen which were from initial observation not used in comparison of the covariance scores.

$$ProjectedLeaf\,Area(PLA) \underset{\sim}{\propto} ExG$$

$$ProjectedLeaf\,Area(PLA) \underset{\sim}{\propto} ExGR$$

$$ProjectedLeaf\,Area(PLA) \underset{\sim}{\propto} ExR^{-1}$$

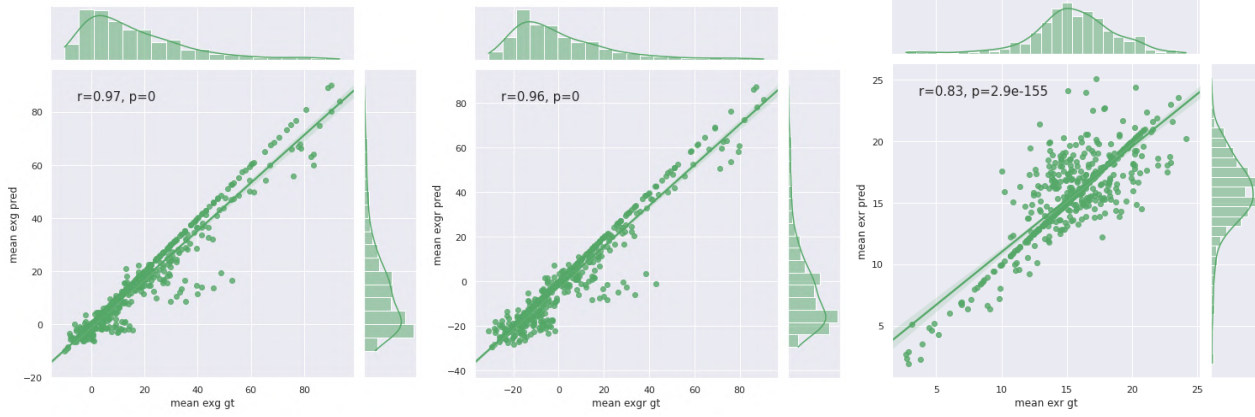### D.1. Covariance metrics for Arabidopsis Thaliana Dataset
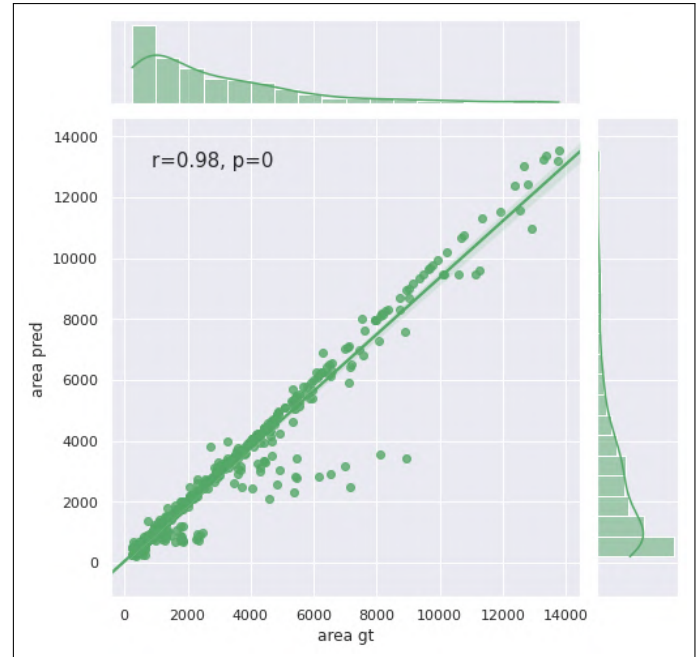
Fig. 42. Comparing Correlation between Generated image Vegetation Indices and Vegetation Indices observed in ground truth frames for Visual Leaf Area

The plot demonstrates correlation between the vegetation indices of Generated images and Ground Truth vegetation indices. The pearson's correlation coefficient (r) showed a value

Fig. 43. Comparing Correlation between Generated image Vegetation Indices and Vegetation Indices observed in ground truth frames for ExG, ExR, ExGR

of 0.98 whereas, the p-value was observed to e 0.0 indicating that all alternate hypothesis over data in latent space was accepted. The high correlation for vegetation indices in Generated frames by the PGGAN showed high accuracy in generating accurate frames with PLA of high covariance. For certain samples between 4000 - 8000 pixels in ground truth, the model underpredicted the data to less than 4000 pixels. The Vegetation indices for input frames in Generated subset as well as Ground Truth subset were included in the plot. The r-value for Mean ExG correlation followed a value of 0.97, Mean ExR correlation with a value of 0.83 and Mean ExGR correlation with a avlue of 0.96.

The Pearson's Correlation Coefficient (PCC) known as (r), is the ratio between the covariance between two variables and product of standard deviations. This coefficient measures the linear correlations between two datasets. Equation 22 shows pearson's coefficient and Equation 23 shows p-value.

The coefficient can be described in the following Equation,

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable $\quad$ **(22)**

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 \frac{(1-p_0)}{n}}} \quad \textbf{(23)}$$

$\hat{p}$ is the sample proportion, p_0 is the hypothesized proportion and n is the sample size.

The vegetation indices calculated for the generated frames were distributed over 4 different subsets for detailed evaluation under accuracy of each subset. The first subset included unseen arbitrary pattern of samples of the Arabidopsis plant taken from the dataset. The interval between two successive images, kept constant were altered arbitrarily and hence a different pattern was formed. The second subset included samples of low observed leaf area in the plant, third subset included samples of intermediate samples of observed leaf area while the fourth subset included sample of high observed leaf area. These were categorically separated to evaluate individual correlation among the Vegetation Indices and observe accuracy and predictive power.

ExR, with inverse proportionality to the calculated area, represents the content of Excessive Red in the plant sample. this has to do mainly with the area occupied by the Soil sample and Red shades in the background. The ExR, when subtracted from ExG, highlights a clear picture of the plant leaf sample, and is deduced in the form of ExGR. ExR alone, with inverse proportionality does not show high correspondence with the Ground Truth extracted vegetation ExR, because of the varying Red sample in the generated images. This is specifically valid for the Arabidopsis Thaliana plant, with the soil sample color varying with temporal changes in leaf evolution over time. Certain sample observe an excessive Green Index over the soil, alternatively observed as moss in a specific amount. The Generated sample observe occasional change in Excess of Red/Green in the soil with respect to other leaf features. ExR observed highest consistency and covariance in plant samples with high leaf density and projected area. Most of the occupied area was covered by leaf samples and hence, ExR did not vary a lot. Conversely, the ExR was measured as the highest Vegetation Index for plant samples of lower leaf area. Similar trend was found for intermediate leaf area samples and a pearson's coefficient (r) of 0.57 and 0.67 for both the samples respectively was observed. Certain unseen patterns of growth were randomly selected and curated with time deviation between two frames greater or less than what's selected for the whole dataset. The 'r' value for the unseen sample patterns were observed to be 0.65.

ExG and ExGR showed higher covariance with Ground Truth samples and Predicted samples. The highest correlation for ExG was observed in High and Intermediate Leaf Area Samples, with 'r' value of 0.96 and 0.95 and for lower leaf area samples and unseen samples, a 'r' value of 0.93 and 0.84. Similar trend for ExGR with 'r' value of 0.97 and 0.94 for high and intermediate leaf area samples and 0.84 and 0.80 for lower leaf area and unseen area samples. Cumulatively, all these Vegetation Indices contributed to calculating the Leaf Area Index of the plant per plant unit(128x128px of plant box cultivar). Since all these indices contributed cumulatively to calculation of the LAI, errors in the calculation was minimized and accuracy increased. High and Intermediate Plant leaf samples saw 'r' value of 0.99 each and low plant leaf area sample saw 'r' value of 0.98. However unseen pattern leaf area correlation observed 'r' value of 0.86. These samples over-predicted the Vegetation Indices and area values. This could be explained by the fact that unseen pattern in image samples showed inconsistency between growth duration
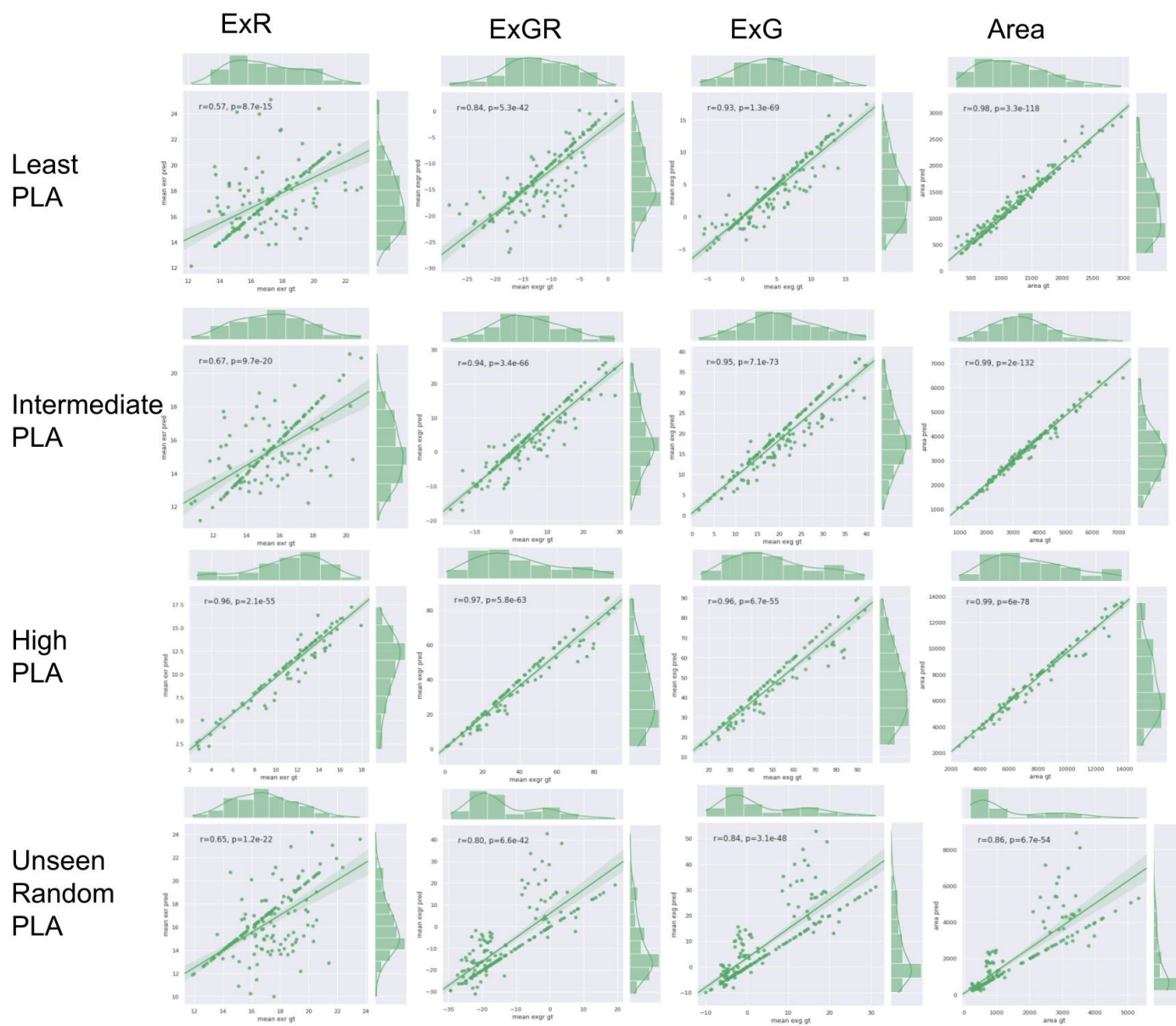
Fig. 44. Vegetation Index correspondence with different subsets of the observed dataset
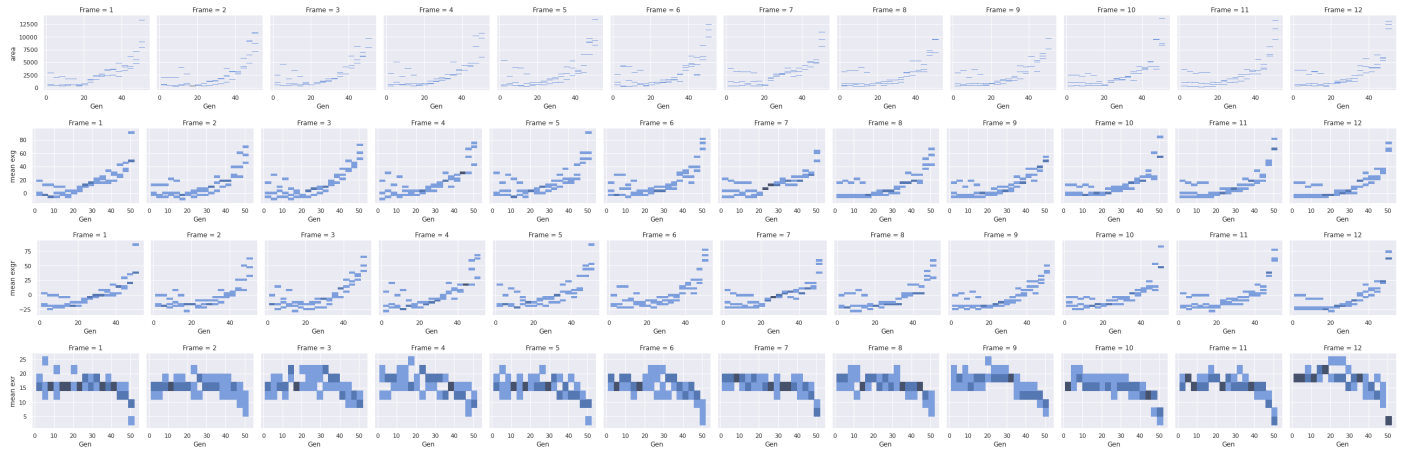
Fig. 45. Plot for progressive frames from 1 to 12 with x axis as Box Number starting from unseen Random samples to low leaf area samples, intermediate leaf area samples and high leaf area samples. Y axis corresponds to Ground Truth Vegetation Indices. **a** Ground Truth Leaf Area, **b** GT ExG Vegetation Index, **c** GT ExGR Vegetation Index, **d** GT ExR Vegetation Index.
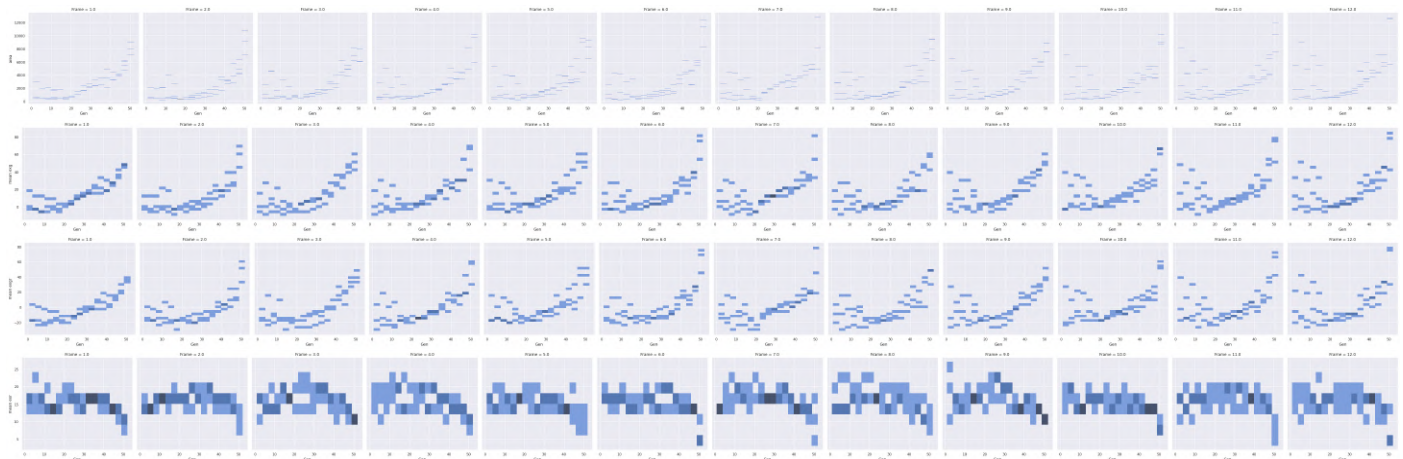
Fig. 46. Plot for progressive frames from 1 to 12 with x axis as Box Number starting from unseen Random samples to low leaf area samples, intermediate leaf area samples and high leaf area samples. Y axis corresponds to Vegetation Indices for Generated Samples. **a** Generated Leaf Area, **b** Generated ExG Vegetation Index, **c** Generated ExGR Vegetation Index, **d** Generated ExR Vegetation Index.

and so certain samples observed over-predicted area samples.

Figure 45 and 46 show plots for Frame-wise deviation in Vegetation Index Morphology with respect to Generated Frame type, initially with Unseen pattern in frames to high leaf area plant frames, from 0 to 601 frames (each in Ground Truth and Generated Image category). Clearly, the ExR frames show inverse correlation to the calculated Leaf Area Index for each unit of Plant Sample (128x128px). For Generated frames, ExG and ExGR were overestimated in unseen pattern of data input type. These graphs unveil an in-depth explanation of corresponding Area GT or predicted for each observed data in the latent space.

Figure 47 shows correlation between Vegetation Indices in Generated Frames as well as Ground Truth Frames. The Vegetation Indices Correlate with high accuracy to produce Visual Area segmentation. Generated Frames and Ground Truth Frames observe **R**^{2} of 0.98 and 0.98 respectively and the high correlation can also be observed in graphical trends spread out over frames.

While, the Arabidopsis Thaliana dataset contains lower complexity over soil and plant features and a comparatively more ideal trend with respect to Beta Vulgaris Dataset. A tabulated form of truncated dataset 5 was autonomously stored using a

python script to a csv file. For each image frame of Arabidopsis Thaliana dataset, the python script analyzed the frames, and segmented the plant leaf area using Vegetation Indices. The table shows collected leaf area for 12 samples in a generated sequence with an inteval of 15-30mins. The input video sequence consists of video frames with frame area from 304 and 625, similarly the predicted frames observe the descending pattern in area and Generate images with area from 319 to 402.

A barplot 48 is illustrated as a function of trend for different samples in the dataset distributed. With, the ExR Vegetation Index having the least correlation to the Area segmentation having the highest correlation over all samples, the highest absolute leaf area sample observed highest correlation for all indices. A larger set Generated images of Arabidopsis Thaliana dataset is included in the code provided.

Conclusions about the evaluation of this Dataset can be inferred through graphs and plots. While this spatio-temporal PG-GAN is trained on merely <700 images, it achieves great scores in Evaluation metrics like SSIM, PSNR, MSE as well as through first principal Generated image evaluation through Vegetation Indices. Not do these images provide relevant generations on

Table 5. Sample Truncated Table for Vegetation Indices captured from Generated Arabidopsis Thaliana Samples
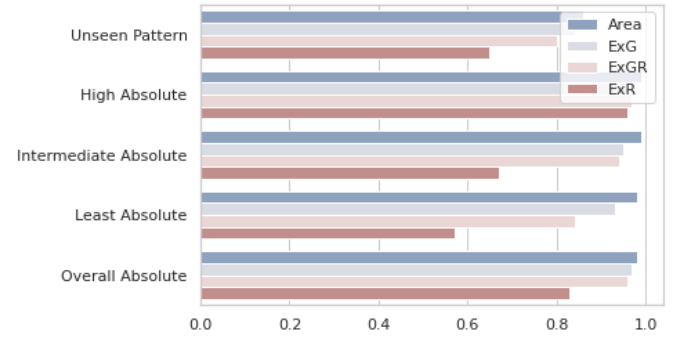
area	max exg	mean exg	min exg	max exr	mean exr	min exr	max exgr	mean exgr	min exgr	Gen	Frame
625	135	-3.036071777	-79	98	14.78167725	0	135	-17.81774902	-157	1	1
515	131	-3.44329834	-71	85	14.2623291	0	131	-17.70562744	-133	1	2
531	107	-4.777709961	-81	79	11.85357666	0	107	-16.63128662	-136	1	3
373	161	-8.67980957	-98	91	20.6975708	0	161	-29.37738037	-156	1	4
304	123	-1.790771484	-88	82	16.44232178	0	123	-18.23309326	-168	1	5
558	135	-2.464599609	-73	89	15.50317383	0	135	-17.96777344	-158	1	6
372	117	-4.964599609	-76	95	17.57275391	0	117	-22.53735352	-138	1	7
402	129	-5.661743164	-72	88	16.8972168	0	129	-22.55895996	-121	1	8
326	111	-4.611816406	-70	87	20.2142334	0	111	-24.8260498	-142	1	9
356	126	-3.625061035	-63	84	15.31286621	0	126	-18.93792725	-118	1	10
319	119	-5.202697754	-74	86	16.17230225	0	119	-21.375	-148	1	11
327	142	-6.21685791	-83	93	19.79919434	0	142	-26.01605225	-151	1	12
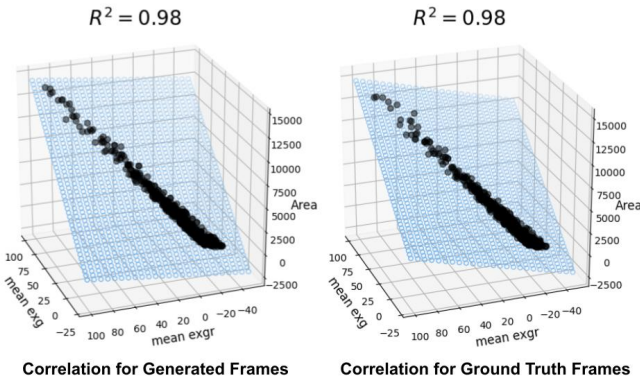
Fig. 47. 3 Dimensional plot covariance with ExG ExGR and Area over Ground Truth and Generated Frames.

Fig. 48. Barplot for correlation between Arabidopsis Generated samples and ground truth samples through pearson's coefficient (r) depicting efficacy distribution over the dataset.

plant and leaf morphology and responsive stimulus, but also they are accurate and reliable making this research novel and deployable over a variety of Agricultural domains. While the Arabidopsis Thaliana was the initial test of the PGGAN model framework with lesser complexity, the Beta Vulgaris model sets a benchmark by performing significantly better with much higher complexity under induces stresses also with varying artificial light and under unsteady capturing conditions. The Fréchet Inception Distance for both the models is calculated at the end of the subsection.

D.2. Covariance Metrics for Beta Vulgaris Dataset

With increased complexity and number of features in the sample and different visual traits contributing to growth modelling and morphology, and different sample types with induced stresses, this model had multiple complicated features in the latent space to learn from as compared to the previous model. For the same reason, this model was trained for 365 epochs rather than 230 in the first model. The latter 135 epochs took significantly larger amount of time than the initial epochs because of the larger weight size scaled up to 128x128 resolution. The lr decay helped

in increasing entropy and decreasing noise while acting as a caveat which increases the number of epochs required for training. Irrespective of all these flaws, the model showed high correlation between Ground Truth and Generated frame Vegetation Indices from Visual Area, ExG, ExGR to even ExR indicating that soil features show high correlation as well (An improvement from the first dataset). Overall Correlation for ExG, ExR, ExGR and Segmented Area for the Beta Vulgaris Generated samples with the ground truth samples maintained high consistency throughout the different samples in the dataset. The Ground Truth and Generated Samples observed near-stellar covariance illustrating that the Generated samples show high similarity with the Ground Truth Samples.

Speaking about the ExR, a measure of Excessive Red typically for the soil, the soil features generated in this model showed great similarity of r-value 1.00 -The p-value of 0 indicating that all samples in the dataset are statistically significant in the model. These Vegetation Indices are not a measure of pixel-level similarity (SSIM measures Structural Similarity) - instead these Vegetation Indices are a measure of the gradient of features present in the Generated samples. Talking about plant vegetation, the mean ExG measures distributed trend of Excessive Green ob-

Table 6. Sample Truncated Table for Vegetation Indices captured from Generated Beta Vulgaris Samples

area	max exg	mean exg	min exg	max exr	mean exr	min exr	max exgr	mean exgr	min exgr	Gen	Frame
1	25	1.892028809	-73	110	13.5345459	0	25	-11.64251709	-119	1	1
0		-8.922302246	-39	108	17.17071533	0	19	-26.09301758	-120	1	2
69	41	-1.775939941	-43	110	19.40899658	0	41	-21.18493652	-126	1	3
391	73	-4.355529785	-46	106	17.65063477	0	73	-22.00616455	-117	1	4
978	73	6.761779785	-27	110	18.25341797	0	73	-11.49163818	-118	1	5
3373	69	15.32513428	-31	106	17.65997314	0	69	-2.334838867	-118	1	6
5077	76	18.27893066	-26	116	14.61547852	0	76	3.663452148	-127	1	7
8392	72	25.87530518	-24	108	9.267700195	0	72	16.60760498	-109	1	8
10565	80	32.94207764	-31	99	5.315673828	0	80	27.62640381	-119	2	9
12582	94	43.27319336	-24	89	3.318237305	0	94	39.95495605	-93	2	10
11765	100	43.03289795	-69	143	8.996948242	0	100	34.03594971	-180	2	11
11914	110	50.86523438	-20	95	3.79498291	0	110	47.07025146	-96	2	12
12323	109	47.9375	-22	100	4.136169434	0	109	43.80133057	-122	2	13
13347	85	44.3470459	-32	96	4.110656738	0	85	40.23638916	-112	2	14
13621	112	48.36645508	-17	107	5.970092773	0	112	42.3963623	-106	2	15
12628	85	38.66821289	-44	94	4.686340332	0	85	33.98187256	-137	2	16

served throughout the plant sample and the density of leaf in a unit measured, correlating with the Excessive Green in Generated Samples. These Vegetation Indices help in better identifying the responsive and morphological overall growth and generation of feature type, rather than pixel level similarity. This helps in a holistic measurement of the plant visual traits generated rather than just understanding how similar was the Generation to the Ground Truth Image. Taking an example, if a weed is observed in the input samples and the Generated images similarly generate the weed samples carried over further frames, the Vegetation Indices would measure the density of the Generated weed or the features of the Weed in the samples rather than similarity measure as of identifying the exact position of the weed generation and correlating efficacy. This extensively helps in identifying if a model is Robust in it's true sense in real world samples or if they are conditional to outliers. **The model performing well in Correlation of Generated Samples in Vegetation Indices as well as SSIM,PSNR,MSE proves that not only is the model objectively capable of understanding outliers, but is also proved to generating similar and accurate real world samples, demystifying prediction of harvest through conditional visual features which extensively affect growth where existing methods fail upon.**

With the ideology being carried to evaluating the Vegetation Indices in Generated samples, the model is objectively as well as subjectively capable of generating frames till harvest with each of these frames (till 8th) contributing to pearson's coefficient (r) in evaluation. Mean ExR shows a r-value of 1.00, the highest in all Vegetation Indices with the Soil feature samples learning diversely all features from the visual input and outliers. Mean ExG and ExGR of r-value 0.99 dynamically indicate that leaf features throughout the sample covary in terms of leaf reflectance, density, venation features and color gradient in response to external

sunlight or internal nutritional factors, under stresses. Cumulatively, the segmented Area from Visual ExGR with r-value of 0.99, illustrating leaf generation (in terms of generative morphology, from high leaf area generated for control plant cultivars to low leaf area generated for stressed cultivars), dynamically morphing over all samples.

Similar to the Arabidopsis Thaliana dataset divided into samples with unknown pattern to high leaf area, the Beta Vulgaris dataset was divided into multiple samples, with corresponding stresses induced on the cultivars. For each sample type, the r-value of the samples was measured to individually understand the data samples which performed well and which of them did ot correlated well. With a pre-assumption that samples under multiple stresses would have a low correlation, the assumption was immediately proved wrong with all the samples cumulatively correlating with high r-value making the model robust to all samples and outliers in each one and bias distribution while learning the model equally favoured to features in each sample. Table 6 shows 16 truncated samples of extracted Vegetation Indices and features out of 500+ samples. The first and second image normally do not observe any segmented Area and observe a negative mean ExGR normalized from all pixels. After frame 8, the measurement of segmented area becomes unyielding to easily measure due to the selected consistent ROI for all images. After 10th sample, the leaves in the cultivar occupy the complete ROI and following that, the volume of the plant inside the ROI increases due to increase in height and the leaf area if to be calculated by the algorithm is observed to be nearly constant inside the ROI. However visually perceived, the morphological evolution of the plant can be distinctly made out by the model, and the same reason why this research is novel to using spatio-temporal generation method instead of feature vectors. This is why during evaluation, the 15th and 16th frame
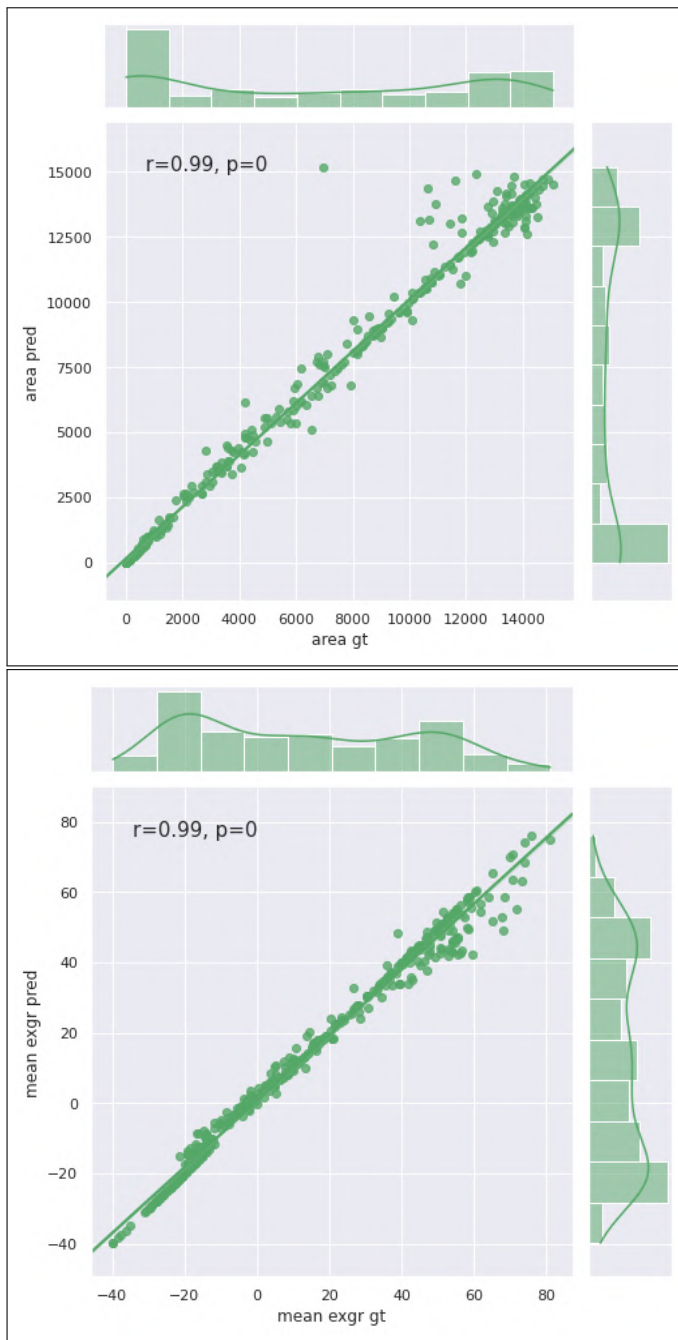
Fig. 49. Comparing Correlation between Generated image Vegetation Indices and Vegetation Indices observed in ground truth frames for Area and ExGR in Beta Vulgaris Dataset
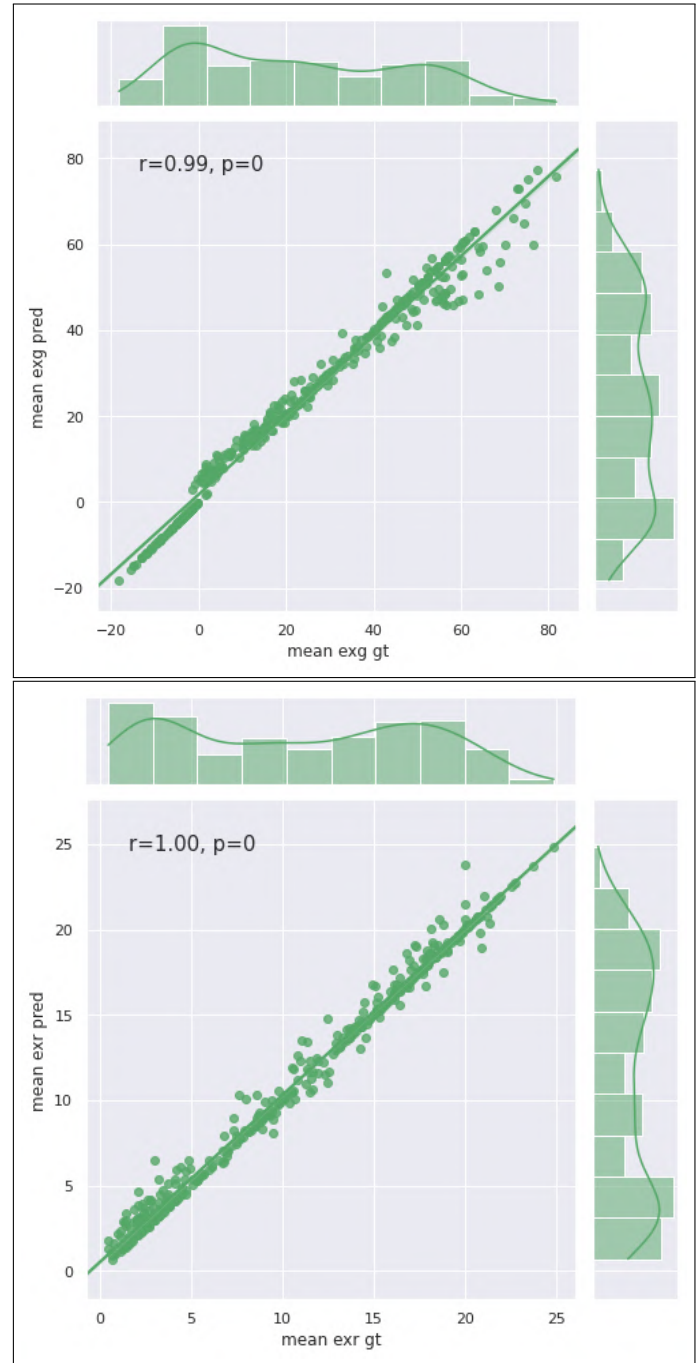
Fig. 50. Comparing Correlation between Generated image Vegetation Indices and Vegetation Indices observed in ground truth frames for ExG and ExR in Beta Vulgaris Dataset

Fig. 51. Comparing Correlation between Generated image Vegetation Indices and Vegetation Indices observed in ground truth frames for ExG,ExR, ExGR and Segmented Area in Beta Vulgaris Dataset for samples - "Control", "Low Nitrogen", "Medium Nitrogen", "Medium Weeds", "High Weeds", "Dry", "Low Nitrogen - Medium Weed", "Drying - Medium Nitrogen - High Weed"

observe increase in petiole and stem height and so certain leaves to be areally segmented are not eventually captured, merely the increase in height can be observed. For this reason, an anomaly in 15th and 16th frame which decreases the morphological trend can be noticed. This might hamper the "evaluation process a bit", but in reality it doesn't because the same caveat is observed in Generated frames as of Ground Truth frames. This isn't a loophole while generating the frames by the model, but is evident merely during evaluating model efficacy which can be neglected as a caveat.

The dataset correlation was distributed to different samples, "Control", "Low Nitrogen", "Medium Nitrogen", "Medium Weeds", "High Weeds", "Dry", "Low Nitrogen - Medium Weed", "Drying - Medium Nitrogen - High Weed". An individual correlation plot with pearson's coefficient (r) was measured individually for each plot and for each Vegetation Index, so a total of 32 plots. 48 image samples were used for each plot with 3 cultivars for each stress induced and 16 frames in each cultivar. Each graph presents a spatial correlation- with temporally morphed yield with frame progression. The lowest correlation for any individual plot was 0.97, which in itself is high, and the highest, observed for multiple plots was 1.00. As observed in Fig 51, all plots had high correlation in Generated to Ground Truth frames from 0.97 and 1.00, with a high jump from Arabidopsis Thaliana dataset. These individual evaluations for plots set a benchmark for yield prediction through or without generative methods, and signifies high reliability, high efficacy and objective robustness to outliers in visual inputs. The r-value(pearson's coefficient) and p-value of the plots was calculated using the standard algorithm 22 and 23.

$$0.97 \leq \textbf{Segmented Area} \leq 1.00$$

$$0.99 \leq \textbf{ExR} \leq 1.00$$
$$0.99 \leq \textbf{ExGR} \leq 1.00$$
$$0.99 \leq \textbf{ExG} \leq 1.00$$

The upper and lower limit of these observed Vegetation Indices are described in the equation. Considering "Control", "Low Nitrogen" and "Medium Nitrogen as cultivars in the control condition, an average r-value for Segmented Area is calculated to be **0.9967**, while for Stresses plants in the rest of the cultivars, is calculated to be **0.9920**. Comparing these calculated values from this Research to one of the most prominent methods to predict yield before harvest using Feature vectors and multiple Neural Network approaches to get the best outcome, it can be concluded that this novel approach in my paper to predicting yield using Spatio-Temporal generation achieves the best outcome. I compared my method results to those proposed by Christian Klukas et al. [21], one of the most prominent researcher in the field of plant phenotyping (with over 3450 citations). His research paper on "Predicting plant biomass accumulation from image-derived parameters", a State of the Art (SOTA) framework in plant phenotyping, used 2 dimensional extracted features in their network to predict yield. This approach made use of multiple imaging tools from Geometric and Color related in Visual to Fluorescent Imaging (Fluo) and Near Infrared (NIR), from top and side view for each imaging tool, so a Permutation of 6 sets of images captured for a single frame. For all these frames, 45 image related traits were extracted and modelled in 4 different Network Frameworks, support vector regression (SVR), multivariate linear regression (MLR), multivariate adaptive regression splines (MARS) and random forest (RF). On top of that, 3 experiments were conducted for each sample and finally the best results were proposed. 2 categories for plant samples, "Control" and "Stress" for each plant were also taken and analyzed.

The Average r-value for plants in "Control" sample was **0.8686** over all 3 experiments and Average r-value for plants in "Stress" Sample was **0.9179**, determined from samples with upper and lower limit,

$$0.8227 \leq \textbf{Controlled Biomass} \leq 0.9403$$
$$0.897 \leq \textbf{Stressed Biomass} \leq 0.9448$$

Table 7. Comparing Pearson's Coefficient in my Approach to existing SOTA approach

	Controlled plant	Stressed Plant
SOTA Framework	0.8686	0.9179
My Framework	0.9967	0.9920
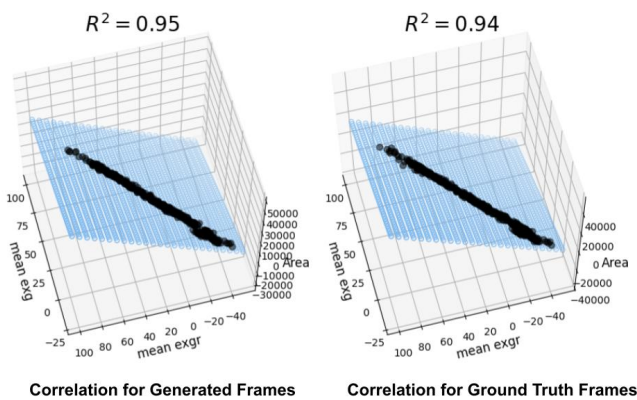
Fig. 52. 3 Dimensional plot covariance with ExG ExGR and Area over Ground Truth and Generated Frames for Beta Vulgaris Dataset.

Plot 52 shows intra-trait covariance and the correlation for ExG, ExGR and Segmented Area derived from these traits. The correlation is 0.95 for Generated samples and 0.94 for ground truth samples. This is an additional plot to show dependence between the traits.
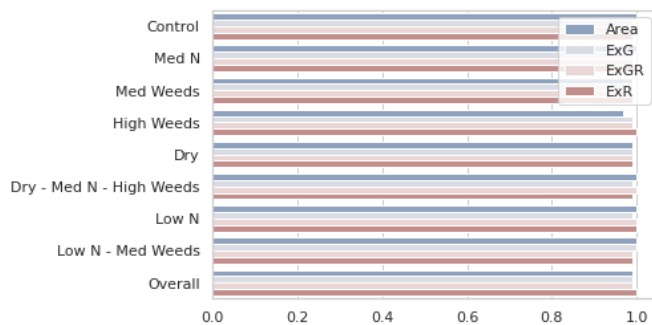
Fig. 53. A barplot to summarize existing correlations between Generated and Ground Truth Vegetation Indices using pearson's coefficient

Finally, a barplot to summarize the pearson's coefficient distribution over Generated and Ground Truth frames is presented.

With all the plots showing high correlation among the generated and ground truth samples, the robustness of the model over all samples is equidistributed and absolutely no bias over an individual sample cultivar is noticed.
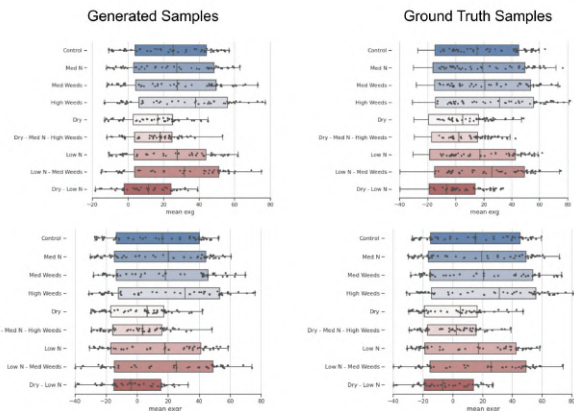
Fig. 54. Boxplot temporally analyzing lower and upper limit of mean ExGR and mean ExG over all frames and cultivar belonging to induced stress category.
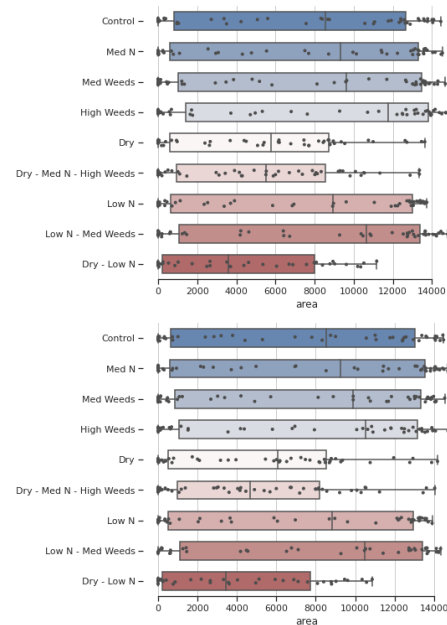

Fig. 55. Boxplot temporally analyzing lower and upper limit of segmented Area for Generated Samples (top) Ground Truth (down) frames and cultivar belonging to induced stress category.

To go on an in-depth analysis of the Generative growth of the plant and per frame correlation, a plot for frame-wise individual Vegetation Indices is presented. For each bar plot corresponding to each frame from 1 to 16, the X-axis of each plot contains the cultivar values predicting by the model. With the plot for Area constantly incrementing from the first frame, ExG, ExGR and progress slowly with few anomalous trends (These trends are anomalous due to Dry samples and stress induced samples). Finally, the ExR shows an exact inverse trend compared to the rest three plots. Overall, Beta Vulgaris dataset performed incredibly
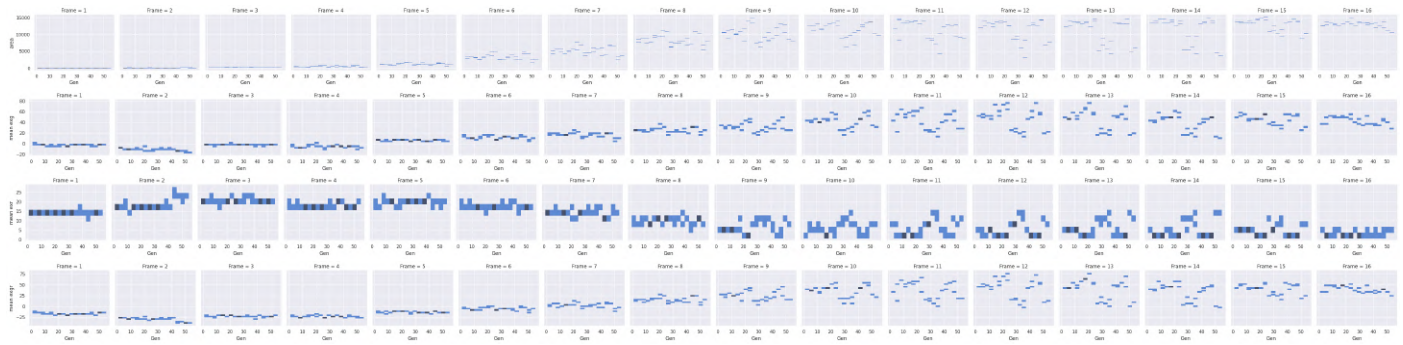
Fig. 56. Plot for progressive frames from 1 to 16 with x axis as Box Number and cultivar type classified in terms of induced stress. Y axis corresponds to Ground Truth Vegetation Indices. **a** Generated Leaf Area, **b** Generated ExG Vegetation Index, **c** Generated ExR Vegetation Index, **d** Generated ExGR Vegetation Index.
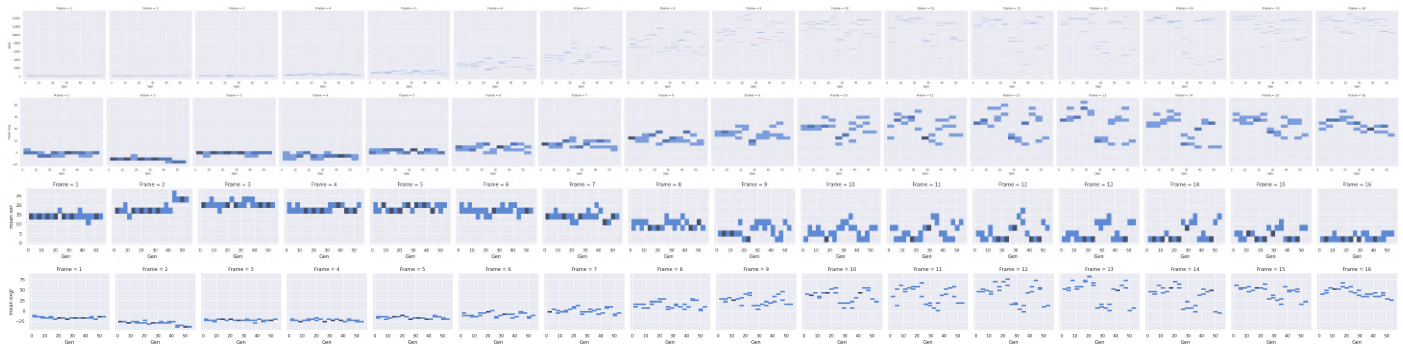
Fig. 57. Plot for progressive frames from 1 to 16 with x axis as Box Number and cultivar type classified in terms of induced stress. Y axis corresponds to Ground Truth Vegetation Indices. **a** Ground Truth Leaf Area, **b** GT ExG Vegetation Index, **c** GT ExR Vegetation Index, **d** GT ExGR Vegetation Index.

well over all sample types, over conditional outliers in visual samples and even analyzing correlation between the Vegetation Indices, showing high r-value and low p-value making all data points statistically significant in the model. The beta vulgaris dataset also sets a benchmark in the plant phenotyping category, with higher r-value and overall comparison with the observed SOTA framework (Note: SOTA models are not officially defined in this category, and hence the model most recognized and compared to perform better than existing models in that category is selected as SOTA). The Beta Vulgaris model, even with high complexity in features, multiple sample types and a small dataset observed high accuracy and similarity in generating images and most importantly, with robustness. This concludes that the model can be adopted by different Greenhouses or autonomous industries in predicting yield before harvest through solely RGB input frames and finding the optimum environmental variables for conducive growth through visual feature inputs.

D.3. Additional plots and Fréchet inception distance (FID)

After an exhaustive summary on correlation and covariance scores and results, this section understands general and overall trend over Generated and Ground Truth samples offering a graphical summary to the evaluated results. As explained in the Arabidopsis subsection, the unseen samples are underpredicted which can be graphically interpreted. Similarly, in evaluating other Vegetation Indices, the mean ExG, and ExGR show an upward trend while min and max Vegetation Indices shows near constant linear trend because that defines the max or min value for a pixel in the image. Mean values take a mean for the overall pixel values and so show an incremental growth. Mean ExR on the other hand decreases by a certain factors over both, the Ground Truth and Generated Samples. In the Beta Vulgaris Dataset, controlled crops or crops with nitrogen stress show a good overall trend over Vegetation Indices and segmented area, while dry crops and those with multiple stress induced show a low graphical trend. Comparing the Vegetation Indices individually, the overall noticeable trend for Ground Truth and generated frames is similar for Arabidopsis Thaliana Dataset as well as Beta Vulgaris Dataset. The Fréchet inception distance (FID) score was calculated for the Beta Vulgaris Dataset. The even numbered samples contained frames 9-16 for the cultivar and the odd numbered samples contained 1-8. A General trend with the FID being lower for 1-8 frames as compared to 9-16 frames was observed throughout the dataset. The lower the FID, better is the generated image. While FID is yet another parameter like SSIM or PSNR, it is used in multiple GAN evaluation frameworks and hence it was used. An Average FID of 116.7275 was observed over all samples. Minimum FID of 59.45 and Maximum of 198.81 was observed.
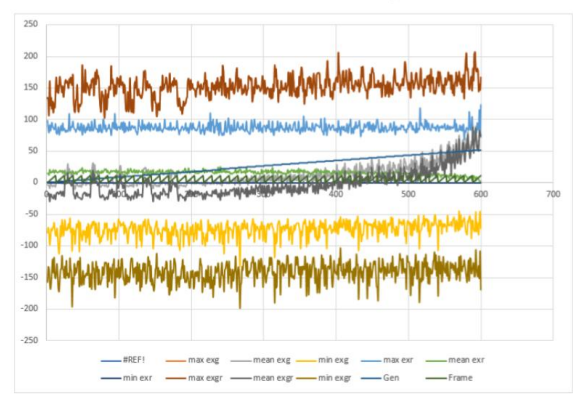
Fig. 58. Plot for progressive frames for Arabidopsis Samples with Area and other Vegetation Indices, starting from unseen pattern trend to low leaf area, intermediate leaf area and high leaf area.
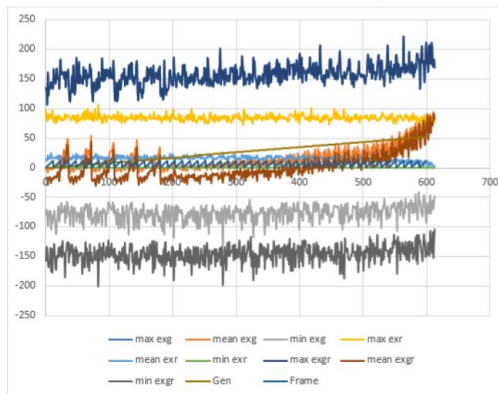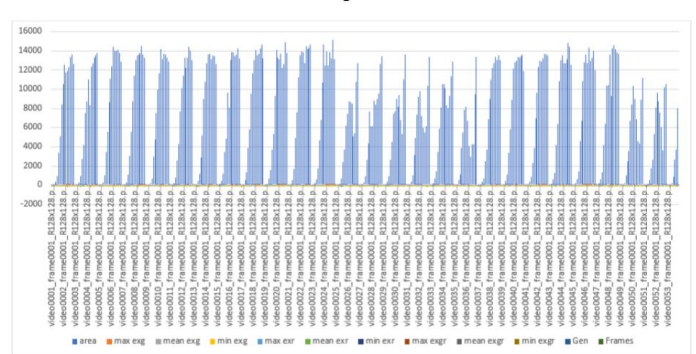

Fig. 59. Plot for progressive frames for Arabidopsis Samples with closer look on Vegetation Indices, starting from unseen pattern trend to low leaf growth, intermediate leaf growth and high leaf growth.
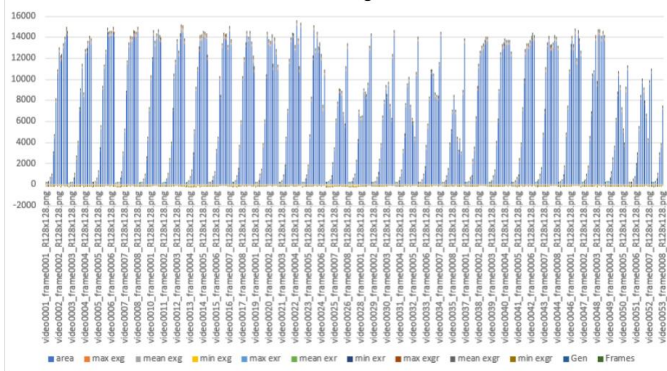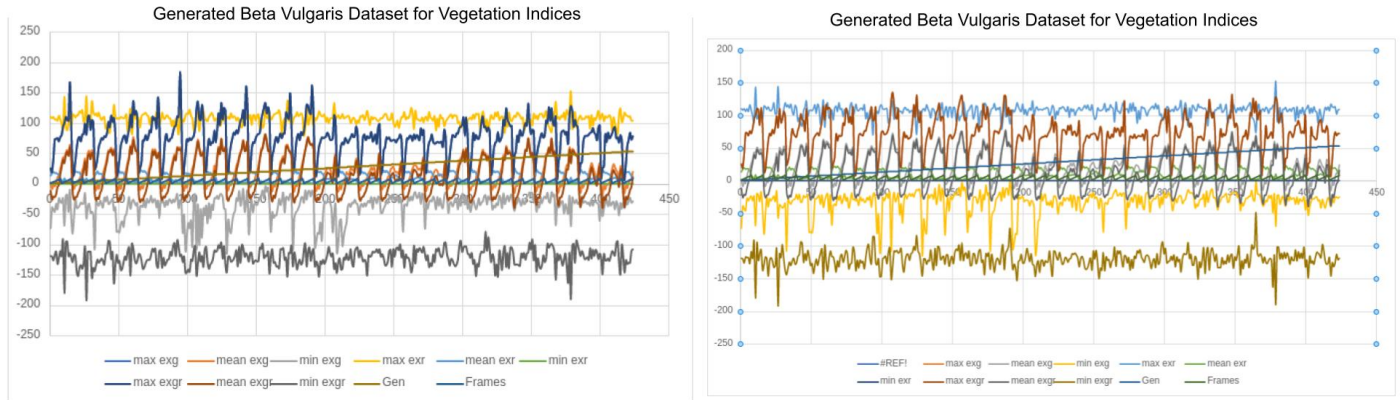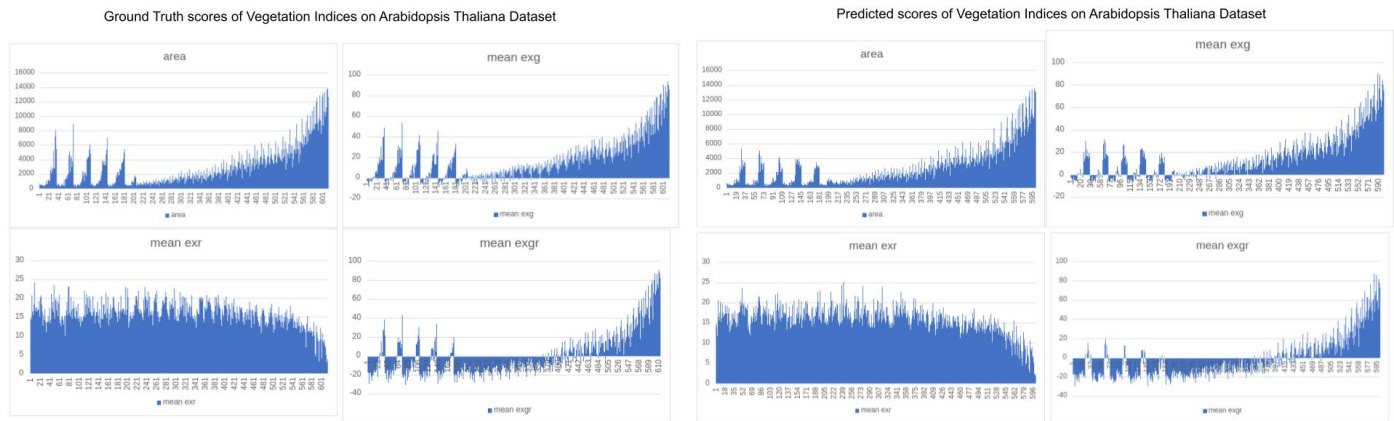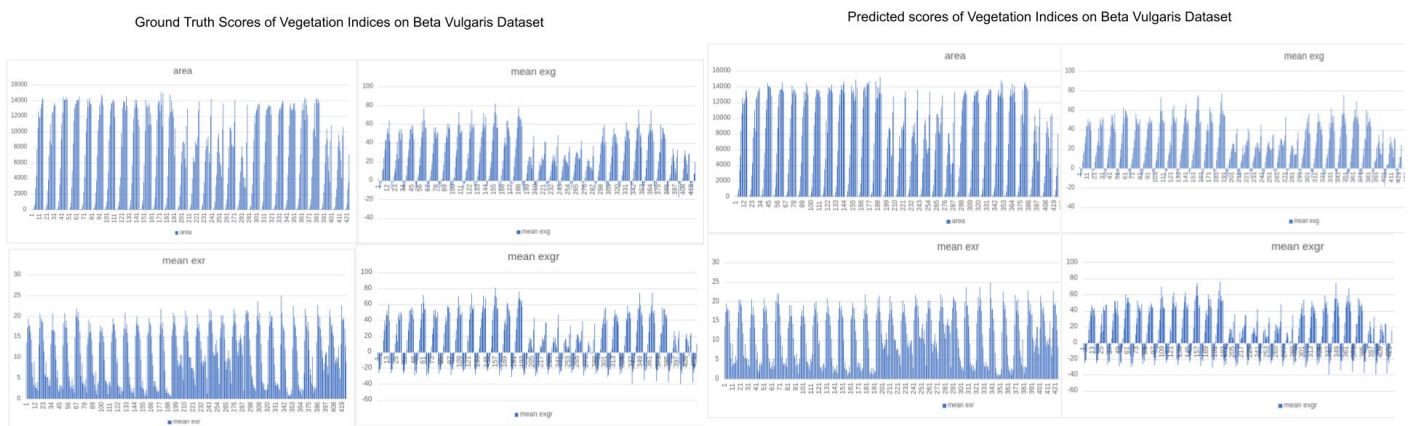

Fig. 60. Plot for progressive frames for Beta Vulgaris Samples with Area and other Vegetation Indices, starting from unseen pattern trend to low leaf area, intermediate leaf area and high leaf area.

Fig. 61. Plot for progressive frames for Beta Vulgaris Samples with closer look on Vegetation Indices, starting from unseen pattern trend to low leaf growth, intermediate leaf growth and high leaf growth.

Fig. 62. Plot for progressive frames for Arabidopsis Samples with Area and other Vegetation Indices, starting from unseen pattern trend to low leaf area, intermediate leaf area and high leaf area.

Fig. 63. Plot for progressive frames for Arabidopsis Samples with closer look on Vegetation Indices, starting from unseen pattern trend to low leaf growth, intermediate leaf growth and high leaf growth.
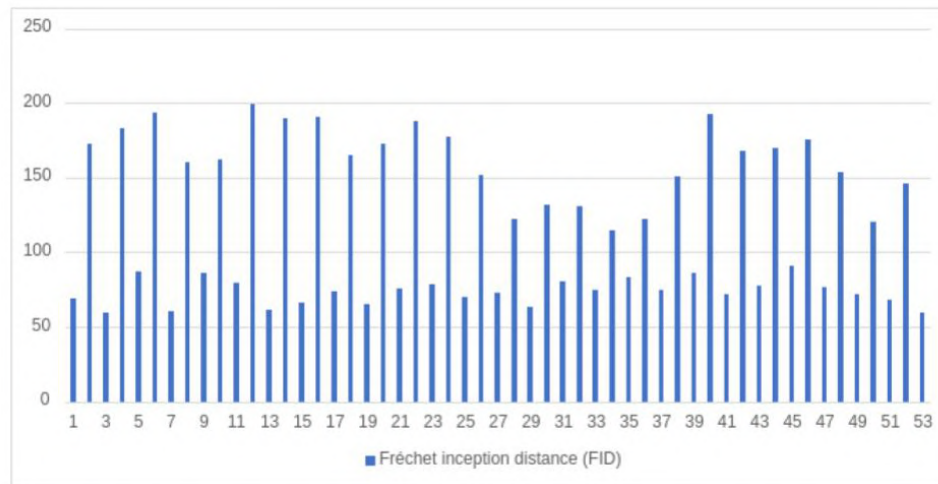
Fig. 64. Fréchet inception distance (FID) for Beta Vulgaris Plot

6. CONCLUSION

Understanding that the existing SOTA approach in plant phenotyping, depends on using 2 dimensional networks with numerical data input of 45 traits extracted from 6 frames of a plant with multiple imaging sensors. This in itself is expensive, requires more equipment, is time consuming to extract multiple traits and is not visually capable of perceiving outliers in image data. Simultaneously, the dependence of multiple traits reduces the ability to understand anomalous trends in data points. Eventually, using these approaches and Frameworks, the average r-value over all experiments is 0.867, compared to 0.9967 as in the proposed approach for control plants and 0.9179 as compared to 0.9920 in stressed plants. The proposed novel approach to generating yield not only achieves a high accuracy, but also requires solely an RGB image input of 4 frames and learns generative features over the raw pixel, enabling democratization of yield prediction algorithms. These robust models, if initially trained on a species of plant samples using proper labels and data format, these are directly assessable by Farmers using Visually generated modes of interpretation. A single image input format, and high similarity with robustness towards conditional outliers and anomalies in visual cues makes this model's results better and more assessable than the current SOTA framework relying on feature vector extraction and numerical data input method.

In the proposed research by [22] and [23], the Research portrays how real world environmental factors and stresses influence hybrid crop tolerance to these factors. My proposed research method can be implemented to accelerate these experimental processes to protracting phenotyping approaches and predict crop yield weeks before harvest through input of visually highlighted cues and features in RGB images. An autonomous system, similar to what is proposed above can not only help in predicting visual generative models of the plant in the form of temporal and spatial analysis, but also helps in reducing human intervention and maximize yield through prediction.

7. FIELD TESTING

Pre and post research observations, I conducted certain observations in Mumbai, India and Banglore, India on how different plant sample varieties respond to visual traits and environmental visual factors affecting plant growth. My analysis could conclude that different soil patterns and features can distinguish whether a plant is drying or wilting as well as the leaf color, venation, features and density overall affects the trend. ExG and ExGR as well as depth for certain samples was analyzed to understand height growth and leaf reflectance to conditional visual features and how they are affected in real world. On this basis, the aim as well as the algorithm for my project originated and was designed on how these visual traits play a key role in deciding plant growth.

8. DATA AND CODE AVAILABILITY

Code : All code and supporting files as well as csv data used for analyzing data. (log files and model structure provided for each model Resolution layer) https://github.com/dhruvsheth-ai/Spatio-Temporal-GAN

Data:

1) Beta Vulgaris Dataset by Raghav Khanna et al. - http://robotics.ethz.ch/ asl-datasets/2018_plant_stress_phenotyping_dataset/

2) Arabidopsis Thaliana Dataset by Hannah et al. - http://download.fz-juelich.de/ibg-2/CVPPP2017_LSC_training.zip

3) Generated Images as well as segmented Area - https://www.kaggle.com/dhruvsheth12345/spatiotemporal-gan Note: Gifs for input as well as generated image are available in each video frame. The dataset is navigable and content for each folder can be viewed.

Note: Unless not referenced by the author, all codes, figures and plots are created by the Author. All results are carefully evaluated, and standard protocols are followed throughout the research. Due acknowledgements to work used in this Research are given to Researchers.
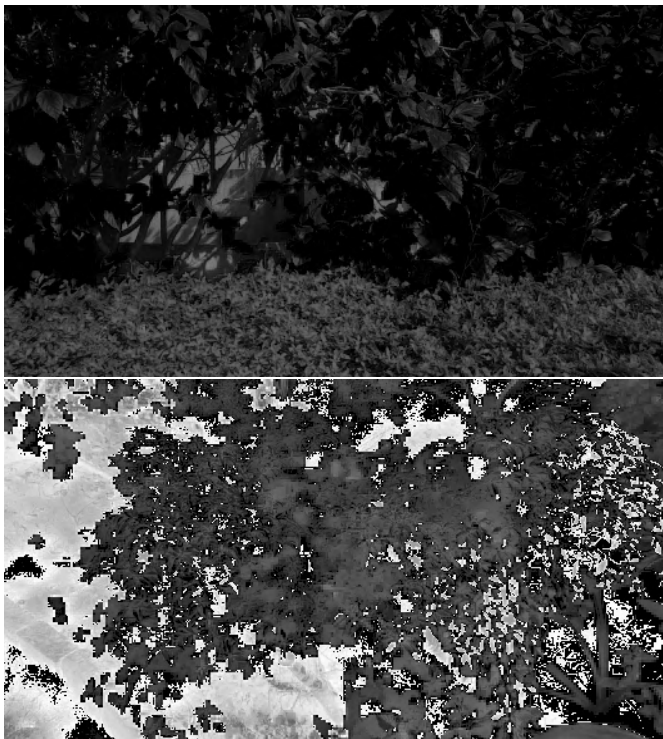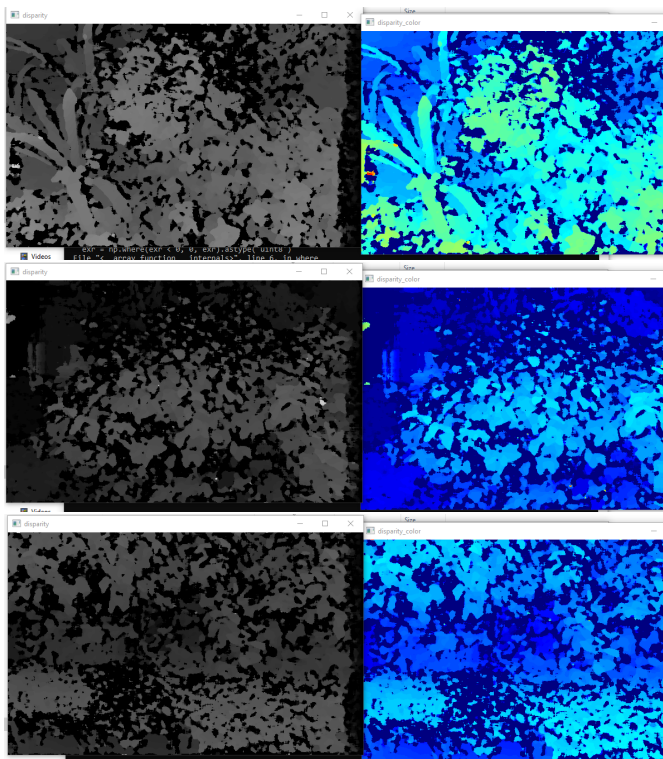
Fig. 66. Observational Analysis at Banglore, India

Fig. 65. Observational Analysis at Mumbai, India

Fig. 67. ExG and ExR obtained from observational analysis

Fig. 68. Depth samples collected to understand height for observational analysis of features

REFERENCES

1. R. M. Bilder, F. w. Sabb, T. D. Cannon, E. D. London, J. D. Jentsch, D. S. Parker, R. A. Poldrack, C. Evans, and N. B. Freimer, "Phenomics: The systematic study of phenotypes on a genome-wide scale," PMC **164**, 30–42 (2009).

2. E. Maynard and M. O'Donnell, "Managing the environment in high tunnels for cool season vegetable production," (2019).

3. A. A. Gomaa1 and Y. M. A. El-Latif, "Early prediction of plant diseases using cnn and gans," (IJACSA) Int. J. Adv. Comput. Sci. Appl. **12** (2021).

4. I. Sa, M. Popovi, R. Khanna, Z. Chen, P. Lottes, F. Liebisch, J. Nieto, C. Stachniss, A. Walter, and R. Siegwart, "Weedmap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming," Remote. Sens. MDPI pp. 30–42 (2018).

5. Z. Lingxian, X. Zanyu, X. Dan, M. Juncheng, C. Yingyi, and F. Zetian, "Phenomics: The systematic study of phenotypes on a genome-wide scale," Hortic. Res. **7** (2020).

6. R. Yasrab, J. Zhang, P. Smyth, and M. P. Pound, "Predicting plant growth from time-series data using deep learning," Remote. Sens. **13**, 331 (2021).

7. L. Drees, L. V. Junker-Frohn, J. Kierdorf, and R. Roscher, "Temporal prediction and evaluation of brassica growth in the field using conditional generative adversarial networks," arXiv. (2021).

8. R. MM, C. D, G. Z, K. C, and C. M, "Advanced phenotyping and phenotype data analysis for the study of plant growth and development." Front. Plant Sci. (2015).

9. M. Fawakherji, C. Potena, I. Prevedello, A. Pretto, D. D. Bloisi, and D. Nardi, "Data augmentation using gans for crop/weed segmentation in precision farming," in *2020 IEEE Conference on Control Technology and Applications (CCTA),* (2020), pp. 279–284.

10. "Futuregan: Anticipating the future frames of video sequences using spatio-temporal 3d convolutions in progressively growing gans," arXiv. (2018).

11. M. Pattinson, S. Tiwari, Y. Zheng, M. Campo-Cossio, R. Arnau, D. Obregón, A. Ansuategui, C. Tubío, I. Lluvia, O. Rey, J. Verschoore, L. Lenza, and J. Reyes, "Gnss precise point positioning for autonomous robot navigation in greenhouse environment for integrated pest monitoring," (2019).

12. M. Hakojärvi, M. Hautala, A. Ristolainen, and L. Alakukku, "Spatial and temporal yield variation in three different clay soil fields," IFAC Proc. Vol. **46**, 196–201 (2013). 4th IFAC Conference on Modelling and Control in Agriculture, Horticulture and Post Harvest Industry.

13. E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative adversarial networks in time series: A survey and taxonomy," (2021).

14. Q. Cap, H. Uga, S. Kagiwada, and H. Iyatomi, "Leafgan: An effective data augmentation method for practical plant disease diagnosis," IEEE Transactions on Autom. Sci. Eng. **PP**, 1–10 (2020).

15. H. Scharr, H. Dee, A. French, and S. Tsaftaris, "Special issue on computer vision and image analysis in plant phenotyping," Mach. Vis. Appl. **27** (2016).

16. D. Weigel and R. Mott, "The 1001 genomes project for arabidopsis thaliana," Genome Biol. **10**, 107 (2009).

17. R. Khanna, L. Schmid, A. Walter, J. Nieto, R. Siegwart, and F. Liebisch, "A spatio temporal spectral framework for plant stress phenotyping," Plant Methods **15**, 13 (2019).

18. I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," in *Proceedings of the 31st International Conference on Neural Information Processing Systems,* (Curran Associates Inc., Red Hook, NY, USA, 2017), NIPS'17, p. 5769–5779.

19. L. F. Sánchez-Sastre, N. M. S. Alte da Veiga, N. M. Ruiz-Potosme, P. Carrión-Prieto, J. L. Marcos-Robles, L. M. Navas-Gracia, and P. Martín-Ramos, "Assessment of rgb vegetation indices to estimate chlorophyll content in sugar beet leaves in the final cultivation stage," AgriEngineering **2**, 128–149 (2020).

20. D. M. Woebbecke, G. E. Meyer, K. V. Bargen, and D. A. Mortensen, "Color indices for weed identification under various soil, residue, and lighting conditions," Transactions ASABE **38**, 259–269 (1994).

21. D. Chen, R. Shi, J.-M. Pape, K. Neumann, D. Arend, A. Graner, M. Chen, and C. Klukas, "Predicting plant biomass accumulation from image-derived parameters," GigaScience **7** (2018).

22. K.-H. Nam, D. Y. Kim, Y. S. Moon, I. S. Pack, S.-C. Jeong, H. B. Kim, and C.-G. Kim, "Performance of hybrids between abiotic stress-tolerant transgenic rice and its weedy relatives under water-stressed conditions," Sci. Reports **10**, 9319 (2020).

23. H. A. Hussain, S. Men, S. Hussain, Y. Chen, S. Ali, S. Zhang, K. Zhang, Y. Li, Q. Xu, C. Liao, and L. Wang, "Interactive effects of drought and heat stresses on morpho-physiological attributes, yield, nutrient uptake and oxidative status in maize hybrids," Sci. Reports **9**, 3890 (2019).

AUTHOR BIOGRAPHIES

Dhruv. M. Sheth is a Junior (11th Grade) at Pace Junior Science College Powai. He conducts research on Spatial Perception and Depth Mapping Systems in Computer Vision and work towards enabling quantisation of Machine Learning Models deployed on Embedded Systems.

Currently, he works on developing Machine Learning quantised models for Embedded Systems at EdgeImpulse. He has also worked at Luxonis Corporation during a part of summer internship which involved working on DepthAI API to develop algorithms and integrate Spatial Perception to Computer Vision Models, using the OpenVINO model framework. During this period, he worked on multiple Computer Vision frameworks including Medianet's FaceMesh and Blazepose, Semantic Segmentation - Deeplab v3+, Image Classification - EfficientNet, MovileNet, Resnet and SqueezeNet, MaskRCNN and FRCNN, Multiple OCR Frameworks.

Whilst spending his 9th Grade Summer Vacation, he worked on and founded an E-commerce startup named Schemobotics, which is a platform for Robotics, Development boards, SoC's, Quadcopters, Robotic Arms, Bluetooth and LoRa modules, and various other SBC's and Embedded Boards.

His recent achievements include: 1) Research Paper "ElastiCL" graded core A* in SenSys Embedded Networks conference (one of the most prestigious conferences in Sensor Network. He co-authored this paper along with other Grad, Undergrad and PhD students, with 3 of them from NUI Galway, 2 from Harvard CS.

2) Grand Prize($1000) awarded by David Eagleman in Neosensory Feel the Future Contest.

3) 3rd Prize ($1000) in Arm DevSummit Competition, for application of UN's SDS's in the project.

4) Panelist at TinyML Asia 2021, for my Research, "Plant GrowthEstimation using quantized Embedded Regression models for high throughput phenotyping"

5) Hackster Impact Prize for best use of UN's SDG's

6) Runner Up in US China Young Maker Competition

7) Best Embedded ML model in ElephantEdge Contest, featured in Microsoft IoT (Project 15)

8) 2nd Prize, BuildWithAI hackathon (team project - 7 members)