8-19-2024 2:00 PM

Enhancing Strawberry Disease and Quality Detection: Integrating Vision Transformers with Blender-Enhanced Synthetic Data and SwinUNet Segmentation Techniques

Kimia Aghamohammadesmaeilketabforoosh, *The University of Western Ontario*

Supervisor: Pearce, Joshua M, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Engineering Science degree in Electrical and Computer Engineering

Abstract

Agricultural productivity in strawberry cultivation was enhanced through the application of machine learning in this study. Traditional methods for detecting diseases and assessing ripeness in strawberries were identified as labor-intensive and error-prone, which limited farming efficiency and reduced crop yields. To address these challenges, it was hypothesized that advanced machine learning models incorporating attention mechanisms could significantly improve these tasks. The objective was to evaluate the effectiveness of various models by optimizing them for specific agricultural applications. Two datasets of strawberry images were augmented, and three pretrained models—Vision Transformer (ViT), MobileNetV2, and ResNet18—were fine-tuned. Data quality was improved through background removal and noise reduction, and weighted training was employed to manage imbalanced class distributions. The robustness of the models was further tested using synthetic data generated via Blender to simulate data scarcity. The results indicated that the ViT model achieved the highest accuracy and precision in identifying diseases and assessing ripeness. The models' effectiveness was further enhanced by the integration of attention mechanisms, and the potential for real-world agricultural applications was validated through the use of synthetic data. This research demonstrated that crop monitoring could be significantly improved with advanced machine learning models, particularly ViT, offering a promising tool for more sustainable and efficient strawberry cultivation. Future studies were recommended to expand these methods to other crops and integrate them into broader agricultural practices to enhance productivity and sustainability.

Keywords

Computer vision, Machine learning, MobileNetV2, ResNet18, Strawberry, SwinUNet, Synthetic data, Transformers, Yield monitoring

Summary for Lay Audience

The application of machine learning and computer vision in agriculture, specifically in strawberry cultivation, has opened new avenues for precision farming, crop monitoring, and disease detection. These technologies provide critical, real-time data that enable farmers to make informed decisions, optimizing resource use and improving crop management. Traditional methods, which rely on manual inspection for disease detection, are labor-intensive and subject to human error. By contrast, computer vision offers a more efficient solution by automating the detection of diseases, pests, and other issues, leading to timely interventions and improved crop quality.

In addressing the challenges specific to strawberry cultivation, such as the lack of large, annotated datasets and the variability in environmental conditions, this research utilized advanced imaging techniques to address these obstacles. The study involved collecting two separate sets of strawberry images, which were then enhanced through resizing and augmentation, including background removal, implementing noise, etc., to train three pretrained models: Vision Transformer (ViT), MobileNetV2, and ResNet18. To counteract the issue of imbalanced class distribution, a weighted training approach was adopted, which equitably distributed the impact across all classes during the training process.

Moreover, the study explored the integration of models like Swin Transformer to tackle more complex segmentation tasks, overcoming the limitations of standard ViT models which lack necessary segmentation heads for pixel-level classification. Through the strategic use of synthetic data and machine learning algorithms, the study aimed to provide robust solutions for strawberry disease identification and ripeness classification, enhancing the capabilities of farmers to monitor and manage their crops effectively.

Co-Authorship Statement

This thesis incorporates content from two manuscripts. One is already published and the second is to be published.

Paper 1 : Aghamohammadesmaeilketabforoosh, K., Nikan, S., Antonini, G. and Pearce, J.M., 2024. Optimizing Strawberry Disease and Quality Detection with Vision Transformers and Attention-Based Convolutional Neural Networks. *Foods*, *13*(12), p.1869.

Author Contributions

Conceptualization, J.M.P.; methodology, K.A., S.N., G.A. and J.M.P.; software, K.A., S.N. and G.A.; validation, K.A. and G.A.; formal analysis, K.A., S.N., G.A. and J.M.P.; investigation, K.A. and G.A.; resources, S.N. and J.M.P.; data curation, K.A. and G.A.; writing—original draft preparation, K.A., S.N., G.A. and J.M.P.; writing—review and editing, K.A., S.N., G.A. and J.M.P.; visualization, K.A. and G.A.; supervision, S.N. and J.M.P.; project administration, J.M.P.; funding acquisition, S.N. and J.M.P.

Paper 2 : Kimia Aghamohammadesmaeilketabforoosh, Joshua Parfitt, and Joshua M Pearce. Blender-Enhanced Synthetic Image Generation: Advancing SwinUNet Segmentation Techniques through Controlled Data Environments, (to be published).

Author Contributions

Conceptualization, J.M.P.; methodology, K.A., J.P. and J.M.P.; software, K.A., and J.P.; validation, K.A. and J.P.; formal analysis, K.A., J.P. and J.M.P.; investigation, K.A. and J.P.; resources, J.M.P.; data curation, K.A. and J.P.; writing—original draft preparation, K.A., J.P. and J.M.P.; writing—review and editing, K.A., J.P. and J.M.P.; visualization, K.A. and J.P. supervision, J.M.P.; project administration, J.M.P.; funding acquisition, J.M.P.

Acknowledgments

I extend my deepest gratitude to everyone who has supported me throughout this remarkable journey. I am especially thankful to my advisor, Dr. Joshua Pearce, whose invaluable guidance and inspiration have greatly enriched my professional and academic development, significantly contributing to my growth.

I also wish to express my sincere appreciation to Dr. Soodeh Nikan for her steadfast support and insightful guidance on this project.

A heartfelt thank you to Professor Seyyed Hossein Amirshahi, my first academic supervisor at Amirkabir University of Technology, who ignited my passion for scientific research in computer vision.

My gratitude goes out to my colleagues at the Free Appropriate Sustainability Technology (FAST) Research Group at Western University—Joshua G, Joshua P, Catalina, Alex, and others—for their encouragement and support.

I am profoundly grateful to my family whom I miss a lot; my parents, Massie and Homauon, my beloved brother Kian. And a huge thank you to my dear friends Farsima, Hossein, and Fatemeh for listening to my pointless dramas and supporting me during difficult times which is invaluable.

I am deeply appreciative of this enriching experience.

Table of Contents

List of Tables

List of Figures

Acronyms

CNN	Convolutional Neural Network
CV	Computer Vision
DA	Data Augmentation
DDC	Deep Domain Confusion
DSC	Dice Similarity Coefficient
GS	Grid Search
HPO	Hyper Parameter Optimization
MLP	Multi-Layer Perceptron
ML	Machin Learning
mPA	Mean Pixel Accuracy
ReLU	Rectified Linear Unit
SETR	Segmentation Transformer
ViT	Vision Transformer

Chapter 1

1 Introduction

1.1 Introduction

Machine learning and computer vision have proven to be valuable tools for farmers to streamline their resource utilization to lead to more sustainable and efficient agricultural production. These techniques have been applied to strawberry cultivation in the past with limited success. To build on this past work, in this study, first, two separate sets of strawberry images, along with their associated diseases, were collected and subjected to resizing and augmentation. Subsequently, a combined dataset consisting of nine classes was utilized to fine-tune three distinct pretrained models: ViT, MobileNetV2, and ResNet18. To address the imbalanced class distribution in the dataset, each class was assigned weights to ensure nearly equal impact during the training process. To enhance the outcomes, new images were generated by removing backgrounds, reducing noise, and flipping them. The performances of ViT, MobileNetV2, and ResNet18 were compared after being selected. Customization specific to the task was applied to all three algorithms, and their performances were assessed. Throughout this experiment, none of the layers were frozen, ensuring all layers remained active during training. Attention heads were incorporated into the first five and last five layers of MobileNetV2 and ResNet18, while the architecture of ViT was modified.

In the evolving fields of agriculture and food production, the integration of advanced technologies like machine learning and computer vision has begun to reshape traditional practices, offering new methods for enhancing efficiency and accuracy. This thesis explores innovative approaches to monitoring and improving strawberry cultivation using state-of-the-art machine learning models, particularly focusing on disease detection and ripeness classification.

1.2 Overview of the Thesis

Following this introductory chapter, Chapter 2 provides a detailed analysis of the application of Vision Transformers to detect diseases and assess the ripeness of strawberries. Chapter 3 expands on the methodology by incorporating synthetic image generation to train segmentation models, demonstrating the effectiveness of these techniques in overcoming the limitations of real-world data. The thesis concludes with Chapter 4, where the findings are synthesized, and directions for future research are outlined, emphasizing the ongoing need for innovation in agricultural technology.

1.3 Significance of the Study

The necessity for more advanced monitoring techniques in agriculture is evident (1) due to various challenges such as disease management, and the need for sustainable practices. Traditional methods of crop monitoring and disease detection are labor-intensive and often lack precision(2). By integrating machine learning models capable of processing and generating complex visual data, this research aims to provide more reliable and efficient tools for farmers, thereby reducing waste, optimizing resource use, and improving crop management.

1.4 Methodology

A significant contribution of this thesis is the application of Vision Transformers and other advanced machine learning models to the field of agriculture. The research not only tests these models under controlled conditions but also evaluates their performance in real-world scenarios, providing a comprehensive assessment of their practical utility. Furthermore, the use of synthetic data to train these models addresses common challenges such as data scarcity and the high cost of data collection, presenting a solution for widespread application in data collection for agricultural technology.

Chapter 2

2 Optimizing Strawberry Disease and Quality Detection with Vision Transformers and Attention-Based Convolutional Neural Networks

This chapter[1] is adapted from the manuscript "Optimizing Strawberry Disease and Quality Detection with Vision Transformers and Attention-Based Convolutional Neural Networks" published in Foods, vol. 13, no. 12, Art. no. 12, Jan. 2024, doi: 10.3390/foods13121869

2.1 Introduction

The fields of machine learning (ML) and computer vision (CV) are rapidly expanding within agriculture, offering a multitude of applications, including precision farming, crop monitoring, and disease detection (1). These technologies provide real-time, precise data regarding agricultural yields and equip farmers and agribusinesses with the necessary insights to make informed decisions in crop management (3). Traditional techniques for disease detection in strawberries often involve manual inspection by experts, which is time-consuming, labor-intensive, and prone to human error. These methods include visual inspections for signs of disease such as leaf spots, discoloration, and mold growth (1).

Computer vision, on the other hand, encompasses optimizing irrigation and fertilization strategies (4). It proves invaluable in identifying issues such as diseases, pest infestations, and fruit damage (3,5), facilitating timely intervention and enhancing overall crop quality (6). Furthermore, it empowers farmers to streamline their resource utilization, encompassing water, fertilizer, and labor, ultimately leading to more sustainable and efficient agricultural practices (7).

One area in which these benefits have yet to reach their full potential is strawberry cultivation. There is limited study on strawberry cultivation and disease control through computer vision due to several challenges such as the scarcity of large and annotated datasets, variability in environmental conditions, and the complexity of disease symptoms. Additionally, the high computational requirements and integration costs pose significant barriers (8). Strawberries are a popular fruit in Canada, with the majority of the crop being grown in Quebec, Ontario, and British Columbia. In addition to being a tasty and nutritious

---

[1] A version of this chapter has been published in Foods MDPI journal. K. Aghamohammadesmaeilketabforoosh, S. Nikan, G. Antonini, and J. M. Pearce, "Optimizing Strawberry Disease and Quality Detection with Vision Transformers and Attention-Based Convolutional Neural Networks," *Foods*, vol. 13, no. 12, Art. no. 12, Jan. 2024, doi: 10.3390/foods13121869.

fruit, strawberries also have a significant economic impact on the Canadian agricultural industry (9).

There have been four core studies applying machine vision to strawberries. First, Zheng et al. (7) revealed that vegetable recognition and size detection could be effectively achieved using a stereo camera in conjunction with a key point detection method. Another study with a focus on vegetable health (10) aimed to detect diseases in vegetables through the utilization of a combination of K-means clustering and support vector machines (SVMs). Transitioning to the context of strawberries, Afzaal et al.(8) successfully detected diseases in strawberries and their leaves. This achievement was made possible through the application of classic deep learning techniques and the implementation of region-based convolutional neural networks (R-CNNs). Moreover, Puttemans et al.(11) conducted a separate investigation that employed object detection methods to distinguish between ripe and unripe strawberries. Their methodology not only facilitated the differentiation of strawberries based on ripeness but also enabled the individual isolation of each strawberry from its cluster(11). In this study, an additional set of data from StrawDI(12), introduced by Borrero, was utilized to complement the dataset from the study by Afzaal et al. (8), allowing for a broader comparison that extends beyond diseases. Necessary modifications were subsequently made, and three pre-existing classification models were specifically trained for this task and comparison. The primary objective of this study is to provide farmers with valuable guidance concerning the most effective method for classifying their images.

In this study, two separate sets of strawberry images, along with their associated diseases, were collected and subjected to resizing and augmentation. Subsequently, a combined dataset consisting of nine classes was utilized for fine-tuning three pretrained classification algorithms: vision transformers, MobileNetV2, and ResNet18. The available dataset is imbalanced. Class imbalance is a major issue in machine learning, causing biased classifiers and poor performance for minority classes. Traditional methods to address this include cost-sensitive learning, which creates synthetic instances and adjusts class weights. Recent advancements focus on handling class overlap and improving evaluation metrics. However, challenges like effective overlap handling and developing adaptive techniques continue to be active research areas (13,14).

To address the imbalanced class distribution in the dataset, each class was assigned weights using PyTorch's Weighted Random Sampler(15) to ensure nearly equal impact during the training process. To improve accuracy, augmentation and attention layers were employed, proving particularly effective in addressing major misclassifications. All three algorithms underwent task-specific customization, and their performance was compared at the conclusion of the study.

The objectives and major contributions of this study include five major tasks. First, the vision transformer, MobileNetV2, and ResNet18 models were fine-tuned to achieve higher accuracy in classifying strawberry diseases and quality. Next, two separate strawberry image datasets were combined and enhanced to create a robust and balanced dataset for the training and evaluation of the models. As the datasets were imbalanced, the weighted random sampler was employed. In addition, attention layers in CNNs were introduced to reduce misclassification and enhance model performance. Finally, the performance of the vision transformer, MobileNetV2, and ResNet18 models were assessed to determine the most effective method for detecting strawberry diseases.

2.2 Methodology

2.2.1 Dataset and Preparation

Two datasets were merged with the goal of identifying diseases in strawberries. The initial dataset, illustrated in Figure 2.1, comprises seven distinct types of strawberry diseases: (a) angular leaf spot, (b) anthracnose fruit rot, (c) blossom blight, (d) gray mold, (e) leaf spot, (f) powdery mildew fruit, and (g) powdery mildew leaf. Initially, the classes were separated in the training file. The number of images in each class was 287, 64, 146, 332, 452, 90, and 380, respectively. The images were RGB and 419 × 419 pixels in size. They were captured with a SAMSUNG Galaxy Note 5 under greenhouse lighting (8,16). The second dataset consisted of cluttered images of strawberries that required cropping and labeling for seamless integration. For the other two classes, images from the StrawDI(12) repository were used. These two classes, ripe and unripe, were included to enhance the completeness of the dataset. The number of images for each class in the training set was 202 for ripe and 208 for unripe. This process resulted in the creation of two distinct classes to distinguish between ripe and unripe strawberries, shown in Figure 2.2. Subsequently, the two datasets were merged to form a more comprehensive dataset.

Figure 2.1. Images of the seven diseases in strawberries and their leaves: (a) angular leaf spot;(b) anthracnose fruit rot; (c) blossom blight; (d) gray mold; (e) leaf spot; (f) powdery mildew leaf; (g) powdery mildew fruit.

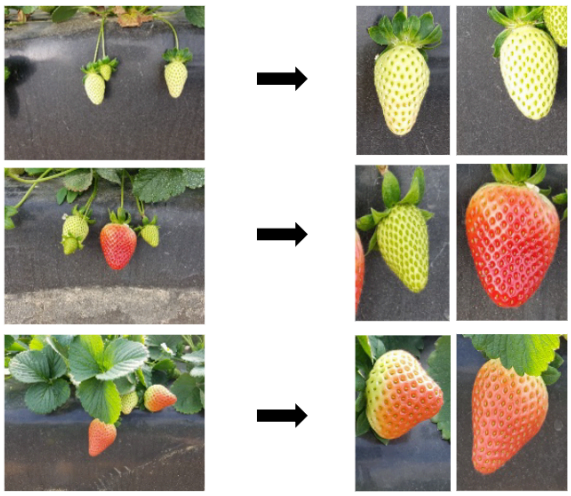


Figure 2.2. Strawberries dataset before and after cropping

For this dataset, transformations were applied to ensure uniformity in the images, including resizing to 256 × 256 pixels, converting to tensors, and normalizing their pixel values based on predetermined zero mean and unit standard deviation values to normalize them for the employed model. An imbalance in class distribution was initially observed in the dataset, as illustrated in Table 2.2. The first approach undertaken to address this issue involved generating additional images for classes with the fewest number of images, namely anthracnose and powdery mildew fruit. Sets of new images were generated through background removal, flipping, and blurring of the existing images using OpenCV functions. Due to the limited number of images, approximately 70, achieving a balanced dataset through image generation, however, was not feasible. The other attempt to address this challenge was to adopt a technique involving a weighted sampling function, where a higher representation will have a smaller weight (17). This approach assigns higher weights to the minority class samples and lower weights to the majority class samples during the model training process, amplifying the impact of the minority classes on prediction (17). First, the dataset was split into 0.8 and 0.2 for training and testing, respectively. Addressing the initial dataset's imbalance issue, weights were calculated and assigned to each class via the 'WeightedRandomSampler' in PyTorch(15,18).
Table 2.2 shows the calculated weights for each class and the details of the augmentation step. Figure 1.3 schematically shows the preprocessing steps in this study. The details of the augmentation are depicted in Table 2.2.

Table 2.1. Preprocessing Details

Preprocessing	Value
Resize	(256,256)
Center Crop	(224,224)
Normalize (mean)	[0.485, 0.456, 0.406]
Normalize (std)	[0.229, 0.224, 0.225]

Table 2.2. Distribution of training data before and after augmentation along with the assigned weights to each class.

Class Name	Angular Leaf Spot	Anthracnose Fruit Rot	Blossom Blight	Gray Mold	Leaf Spot	Powdery Mildew Fruit Rot	Powdery Mildew Leaf	Ripe Strawberries	Unripe Strawberries
No. of Original	245	54	117	255	382	80	319	230	243
After Addition	245	100	150	255	382	151	319	230	243
Class weights	0.8569	3.8847	1.7481	0.7438	0.5406	2.6757	0.6490	1.0724	1.0385

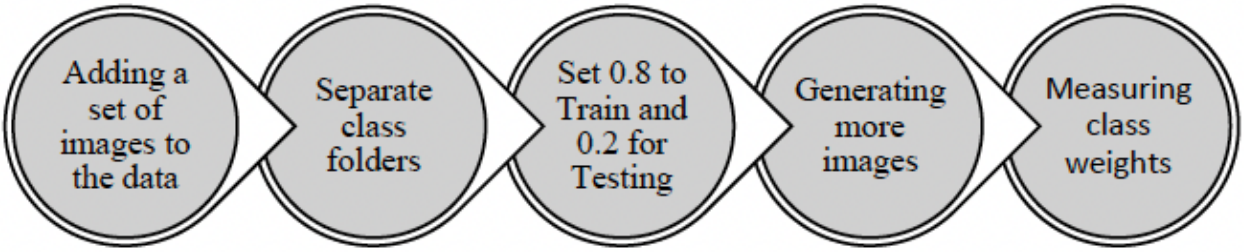


Figure 2.3. Preprocessing steps before training

2.2.2 Methods

The prepared data were fed to three distinct pretrained models: vision transformer (19), MobileNetV2(20), and ResNet18 (21). Each of these models underwent specific adjustments to make them suitable for the intended task, as elaborated in the paragraphs below. The vision transformer has demonstrated that models trained on large and varied datasets can effectively grasp fundamental visual concepts. This leads to better performance across various tasks and areas, showing improved adaptability and understanding (22). The streamlined structure of MobileNetV2 facilitates faster convergence in training, expediting both model development and deployment. Integrating MobileNetV2 into transfer learning leverages this accelerated training, resulting in more efficient utilization of computational resources (23). ResNet18, being a widely used algorithm, serves as a viable benchmark for comparison purposes in image classification

applications.
Table 1.3 presents the specifications of the models. It is important to highlight that, to ensure fair comparison among the models, the parameters are identical and are outlined in table1.3.

Table 2.3. Parameters chosen for the method

Parameter	Value
Optimizer	SGD
Batch size	32
Learning rate	0.001
Epoch	200
Momentum	0.9
Training GPU	Digital Alliance Canada (sharcnet) A100 (Google Colab)

2.2.2.1 Vision Transformer

A standard transformer architecture was used to process both token embeddings and 2D image data. Images were converted into sequences of flattened patches, which were then mapped to a fixed-size vector. Positional information was maintained using standard 1D position embeddings. The transformer encoder consists of alternating layers of self-attention and multi-layer perceptron (MLP) blocks, with layer normalization and residual connections. This setup allows the model to effectively represent and process both textual and image data (24). ViT differs from CNNs in its inductive bias, utilizing the two-dimensional neighborhood structure sparingly, primarily by dividing the image into patches. Adjustments to position embeddings were made during fine-tuning for images of different resolutions. Unlike CNNs, ViT's position embeddings contain no initial information about patch positions, requiring the model to learn spatial relationships between patches from scratch (19). Instead of raw image patches, the input sequence can be made from feature maps of a CNN. In this hybrid model, patches from the CNN feature map underwent patch embedding projection. Patches with a spatial size of 1×1 flattened the feature map's spatial dimensions, projecting it to the transformer dimension. Classification input and position embeddings were added as described (25). The encoder part of the original transformer architecture was employed by the ViT, and the decoder was not utilized. A sequence of embedded image patches, with a learnable class embedding prepended to the sequence, was taken as input to the encoder, which was augmented with positional information. The self-attention mechanism, a key component of the transformer architecture, was employed. Importance scores were assigned to patches by the model, allowing it to understand the relationship between different parts of an image and focus on the most relevant information. This aids in better comprehension of the image and enables the model to perform various computer vision tasks. Following this, a classification head attached to the output of the encoder received the value of the learnable class embedding

to output a classification label. Figure 2.4 illustrates all of these processes.

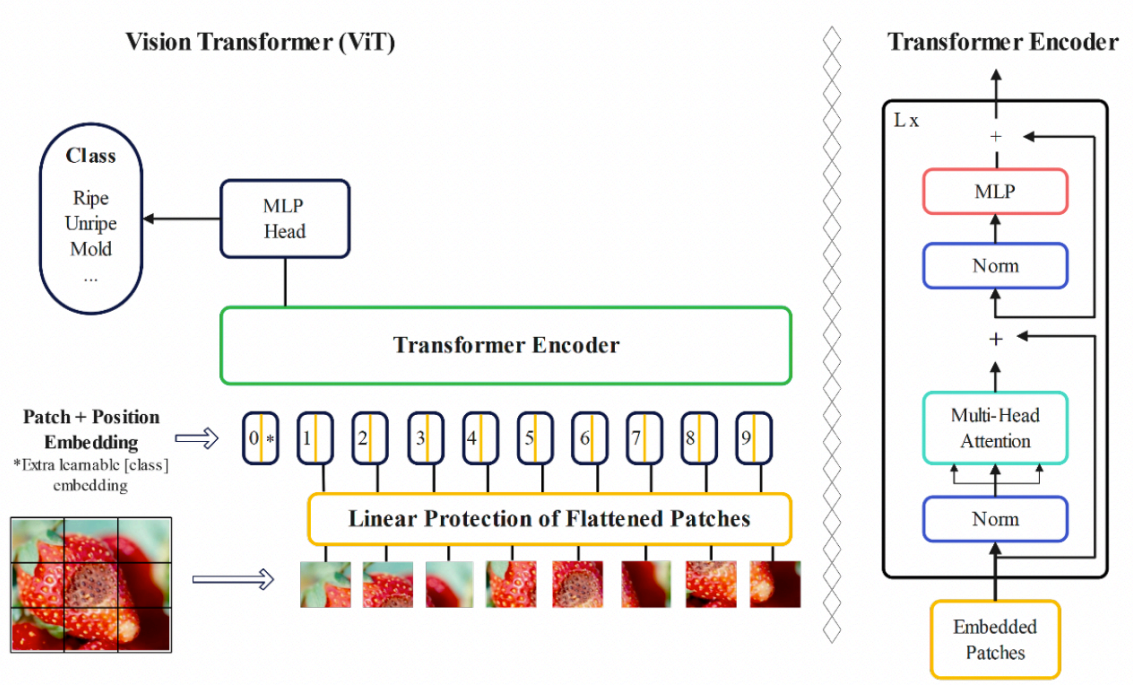


Figure 2.4. Architecture of a ViT

2.2.2.2 MobileNetV2

MobileNetV2 is a specialized type of CNN designed for a range of visual tasks, particularly useful in agriculture (20). Its standout feature is efficiency, crucial in scenarios with limited computational resources (26). Its ability to achieve high accuracy with a reasonable number of parameters makes it suitable for real-time applications such as crop monitoring and disease identification in agriculture. In MobileNetV2, two types of blocks were present, one being a residual block with astride of 1, and the other, a block with a stride of 2 for downsizing. Both types consisted of three layers each. First, a 1 × 1 convolution layer followed by ReLU6 activation was applied. For a standard configuration with a width multiplier of 1 and a resolution of 224x224, MobileNetV2 performs 300 million MAC operations, indicating the number of operations combining multiplication and accumulation performed during one forward pass, reflecting the model's computational demand. Variations in input resolutions and width multipliers adjust the computational cost up to 585 million MACs and parameters between 1.7 and 6.9 million, with the removal of ReLU6 at each bottleneck module output enhancing accuracy. However, the performance trade-offs were further explored for various input resolutions ranging from 96 to 224 and width multipliers from 0.35 to 1.4, leading to computational costs up to585 M multiply-adds and model sizes between 1.7 M and 6.9 M parameters. Notably, the removal of ReLU6 at the output of each bottleneck module resulted in improved accuracy. Additionally,

incorporating shortcuts between bottlenecks yielded better performance compared to shortcuts between expansions or those without any residual connections. Figure 2.5 illustrates the architecture of the original MobileNetV2.

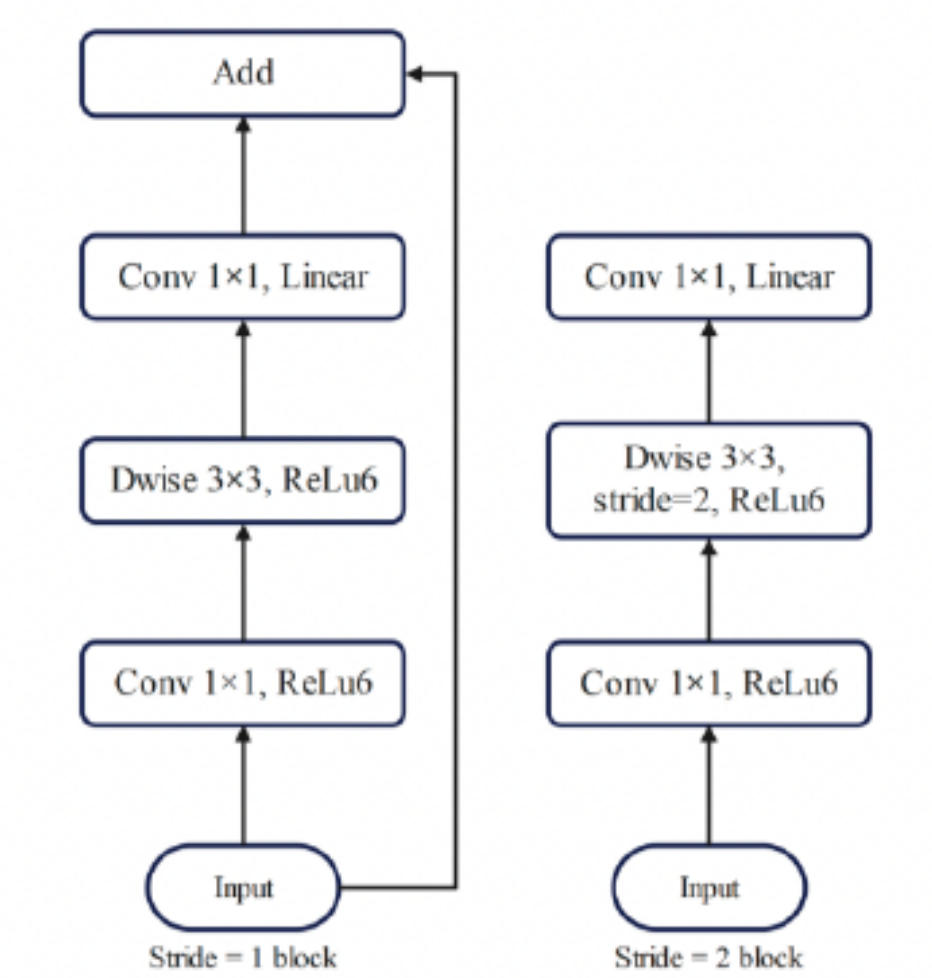


Figure 2.5. Original MobileNetV2's architecture.

2.2.2.3 ResNet18

ResNet18's key feature is its depth with many layers, which helps to extract distinctive features in complex image classification tasks (21). It deals with a common problem called "vanishing gradient", which can make training difficult. In the context of deep learning, a "vanishing gradient" problem occurs when the gradients, which are used during the backpropagation process to update the weights of the neural network, become increasingly smaller as they are propagated back through the network's layers. This issue becomes more pronounced in deeper networks with many layers, like ResNet18. As gradients become smaller, the updates to the weights in the earlier layers of the network become

insignificantly small, slowing down the learning process or stopping it altogether, making the network difficult to train. To address this, “residual connections” allow the network to skip certain layers during training, making it easier to train deep models (21). This approach demonstrated that deeper networks achieved improved optimization and accuracy (27). Traditional deep networks face difficulties in training, however, and increasing the number of layers does not guarantee better learning outcomes. As deeper networks converge, accuracy plateaus and then rapidly declines. Residual learning tackles this issue by learning residual mappings rather than direct mappings. The original ResNet-18 architecture comprised eighteen layers, known as residual including convolutional layers with 3 × 3 filters and down sampling layers with a stride of 2. Figure 6 illustrates the architecture of the original Resnet18. These blocks played a crucial role in improving how the network learns intricate features from input data. Throughout the network, residual shortcut connections were inserted between layers either maintaining the same dimensions or adjusting for dimensionality changes. Residual block operation can be expressed as follows:

$$F(x) = H(x) - x \tag{2.1}$$

where F(x) is the residual function to be learned, x is the input to the block, and H(x) is the underlying mapping. The residual connection facilitates the learning of the residual function, mitigating the vanishing gradient problem. Another aspect is the use of global average pooling, a method that simplifies information before making final predictions. This pooling helps to reduce the spatial dimensions of feature maps (21).

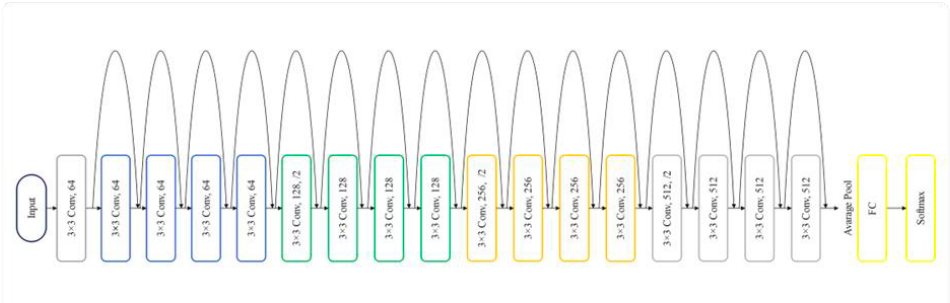


Figure 2.6. Original ResNet18’s architecture

2.2.3 Hyper Parameter Optimization and Attention Mechanism

Machine learning algorithms often rely on hyperparameters, which have to be chosen through automatic hyperparameter optimization (HPO) in order to obtain reliable and reproducible results (28). GridSearch (GS) is a method involving the systematic evaluation of hyperparameter combinations by discretizing their ranges. Numeric and integer hyperparameter values are typically evenly spaced within their specified constraints, with the number of distinct values per hyperparameter termed the grid’s resolution. The optimization of categorical hyperparameters involves considering either a subset or all

possible values. HPO methods streamline the process of finding optimal hyperparameter configurations, enhancing the performance and reproducibility of ML models. In this research, GridSearch (29) was used in every algorithm to select the best learning rate to fine-tune the model during validation. To determine the suitable learning rate, a gradual approach was taken. Initially, a low learning rate of 0.00001 was implemented for warm up purposes, followed by a gradual increase to 0.1. Subsequently, after optimization, a value of 0.001 was identified as the optimal learning rate, which was then employed consistently across all models to ensure fair comparison. Cross-entropy loss and the SGD optimizer were employed, with a learning rate of 0.001. The model underwent 5-fold cross validation to enhance the models' generalization and reduce the risk of overfitting. To understand feature maps and introduce attention mechanisms, it is essential to recognize that patterns and combinations of low-level features are captured by intermediate layers. A balance between low-level and high-level information is provided by extracting features from these layers. High-level semantic information, contributing to the understanding of more abstract concepts, is captured by later layers and residual blocks. Attention mechanisms were introduced to specific layers responsible for these misclassifications,(30) directing the model's focus to distinct parts of input images.

2.2.4 Computational Power

The implementations were performed using the library PyTorch (Torch 2.1) of Python with the support of the Digital Alliance of Canada (31) and Google Collaboratory, which provided the GPU resources to accommodate the enhanced processing demands posed by the extensive dataset and prolonged training epochs.

2.2.5 Evaluation Metrics

Accuracy measures the fraction of correctly classified instances in the total number of instances and is deemed effective when class distribution is balanced. Precision gauges the fraction of accurately classified positive instances in relation to all instances classified as positive, indicating how many of the predicted positive instances are truly positive. It is a suitable metric in scenarios where it is crucial to minimize the occurrence of false positives (32). Recall, also referred to as sensitivity or true-positive rate, assesses the fraction of accurately classified positive instances relative to all the positive instances, indicating the number of actual positive instances that were correctly identified as positive. It is advantageous in situations where missing a positive instance has significant consequences. On the other hand, the F1 score, which is the harmonic mean of precision and recall, is a combined metric that balances the values of precision and recall, thus providing a single score that is useful when both false positives and false negatives need to be considered. Accuracy, precision, recall, and F1 score, defined in the following equations, are

commonly employed evaluation metrics in machine learning for quantifying the performance of a classifier:

$$\text{Accuracy}(\%) = \frac{\text{True Negative} + \text{True Positive}}{\text{True Negative} + \text{False Positive} + \text{True Positive} + \text{False Negtive}} * 100 \quad (2.2)$$

$$\text{Precision}(\%) = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} * 100 \quad (2.3)$$

$$\text{Recall}(\%) = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} * 100 \quad (2.4)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.5)$$

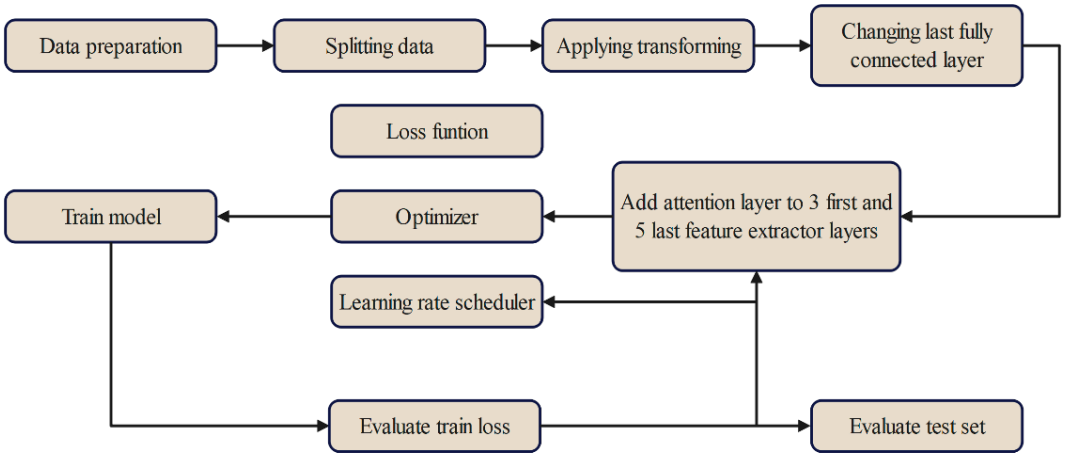


Figure 2.7. Procedure for the algorithms.

2.3 Results

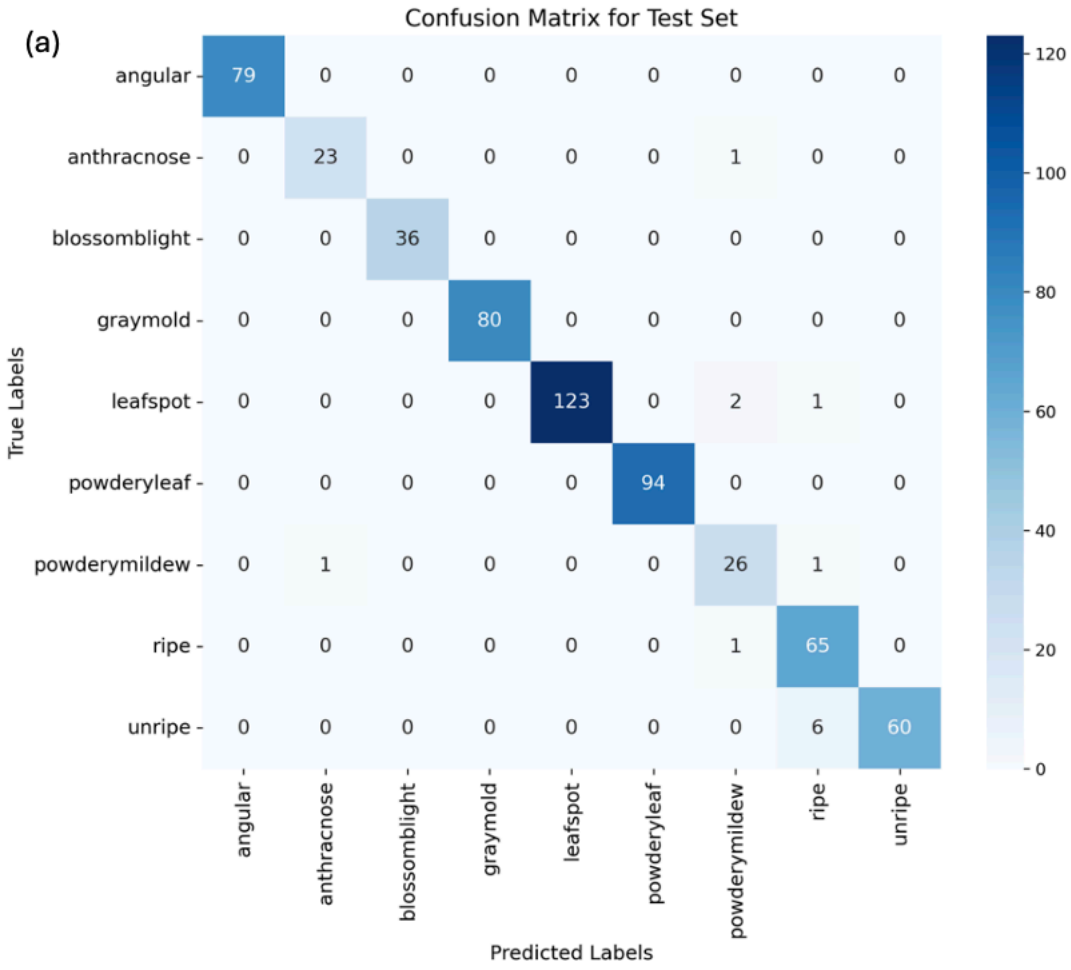
The results of the three models were compared using the accuracy, precision, recall, and F1 score metrics, as shown in Table 2.4. Based on the work of Afzaal et al. (8), an average precision of 82.43% was attained. In this investigation, the precision in each class was improved. Additionally, all three algorithms utilized were different from ResNet101, as employed in the original study.

Table 2.4. Evaluation Results for each model

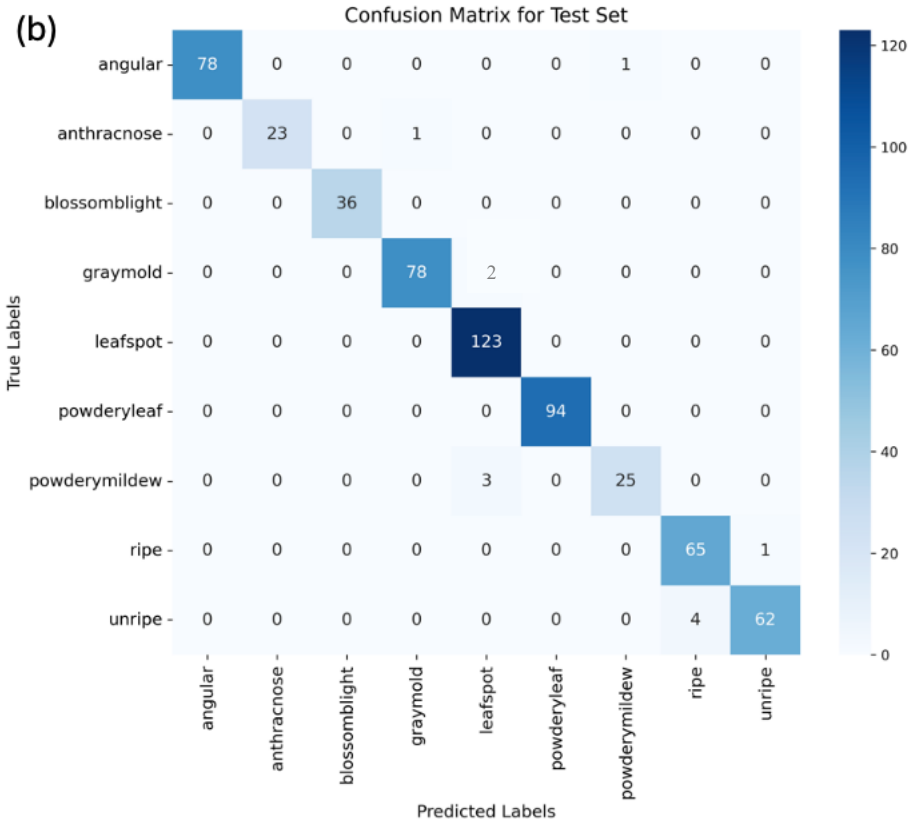
Model	Precision	Recall	F1 Score	Accuracy
Vision transformer	0.983	0.983	0.983	0.984
MobileNetV2	0.980	0.979	0.979	0.981
ResNet18	0.979	0.978	0.978	0.979

In this study, to enhance feature representation in models based on convolutional neural networks (MobileNetv2 and ResNet18), attention mechanisms were employed. Custom modules were employed for efficient feature extraction while minimizing computational complexity. Additionally, attention mechanisms were integrated into the CNN architecture through a module named Attention Module, allowing for the dynamic adjustment of feature importance. Attention Modules were embedded into key feature extraction layers of a pretrained CNN model, specifically the first five and last layers of convolutional layers. For the vision transformer, the ViT feature extractor was loaded to extract features from images. A collate function was defined to convert batches of data into tensors. The model was trained and evaluated, and metrics were logged. Optionally, a model card was created with information about the fine-tuning process, dataset, and tags. Each step contributed to the comprehensive process of loading, preparing, training, and evaluating the model for image classification. The models were evaluated on the 20% dataset used as the test set, with attention mechanisms impacting feature representation and the overall model performance being assessed through experimental validation.

Moreover, these metrics were utilized to analyze the confusion matrix of the predictions. The confusion matrix shows the number of true positives, false positives, true negatives, and false negatives for each class. Accuracy measures the overall correctness of a model's predictions, while precision assesses the correctness among positively labeled instances. In imbalanced datasets, precision is favored over accuracy because it highlights performance on the minority class and avoids misleading metrics due to the dominance of the majority class. Precision is especially valuable in scenarios like medical testing or fraud detection, where false positives have significant consequences. By analyzing the confusion matrix, we can identify which classes appeared to be more difficult to predict and focus on improving the performance of the model on those classes. Figure 8 displays the confusion matrices for ViT, MobileNetV2, and ResNet18.


(a)
Confusion Matrix for Test Set
True Labels
Predicted Labels
angular
anthracnose
blossomblight
graymold
leafspot
powderyleaf
powderymildew
ripe
unripe
79 0 0 0 0 0 0 0 0
0 23 0 0 0 0 1 0 0
0 0 36 0 0 0 0 0 0
0 0 0 80 0 0 0 0 0
0 0 0 0 123 0 2 1 0
0 0 0 0 0 94 0 0 0
0 1 0 0 0 0 26 1 0
0 0 0 0 0 0 1 65 0
0 0 0 0 0 0 0 6 60
120
100
80
60
40
20
0

(b)


Confusion Matrix for Test Set
True Labels
Predicted Labels
angular
anthracnose
blossomblight
graymold
leafspot
powderyleaf
powderymildew
ripe
unripe
78 0 0 0 0 0 1 0 0
0 23 0 1 0 0 0 0 0
0 0 36 0 0 0 0 0 0
0 0 0 78 2 0 0 0 0
0 0 0 0 123 0 0 0 0
0 0 0 0 0 94 0 0 0
0 0 0 0 3 0 25 0 0
0 0 0 0 0 0 0 65 1
0 0 0 0 0 0 0 4 62
120
100
80
60
40
20
0

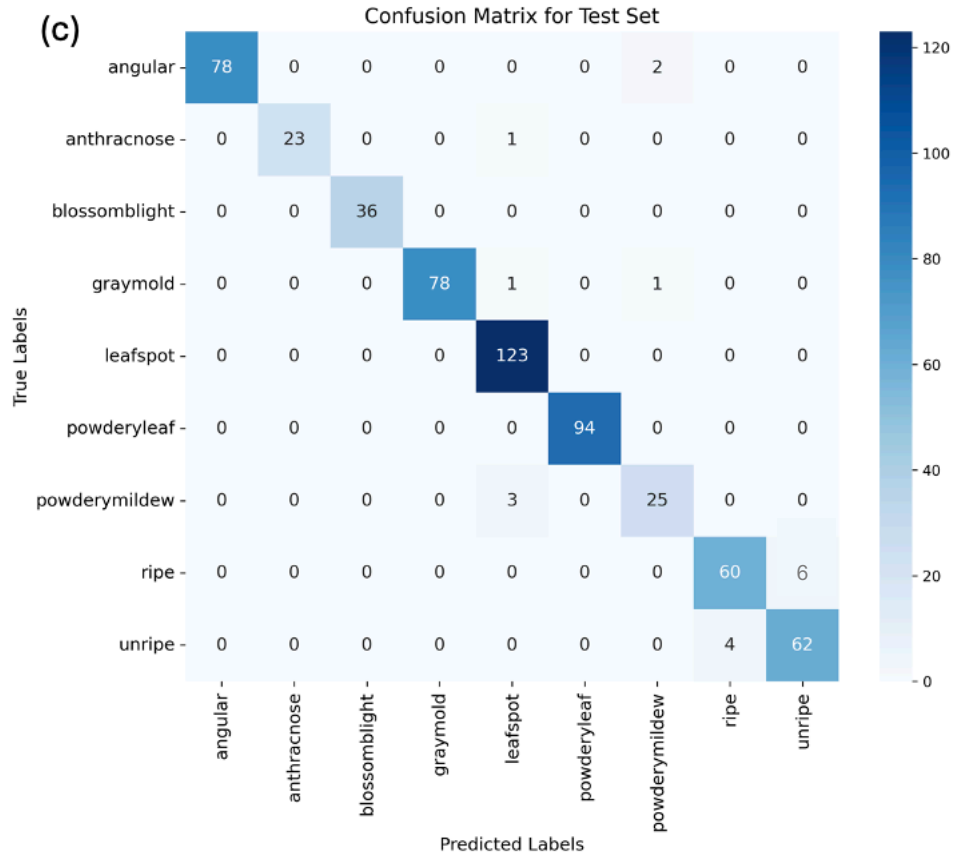


Figure 2.8. Confusion Matrix for (a) ViT, (b)MobileNetV2, and (c)ResNet18.

The confusion matrices for the ViT, MobileNetV2, and ResNet18 models highlight instances of misclassification, which requires more discussion. In the vision transformer model, anthracnose was once misclassified as blossom blight. Additionally, powdery mildew was once confused with gray mold, both presenting mold-like symptoms, which may make them difficult to distinguish, especially in the early stages. Unripe strawberries were misclassified as ripe strawberries six times, due to variability in color through their growth. The MobileNetV2 model, enhanced with attention layers, showed somewhat similar patterns of misclassification, with anthracnose and powdery mildew fruit rot being confused with gray mold once and three times, respectively, again due to overlapping visual features. Moreover, unripe strawberries were misclassified as ripe four times, suggesting challenges in differentiating ripeness stages due to the fact that images were mostly in the middle stages of growth and included both red and yellow colors. The attention layers improved focus on relevant features, but some subtle distinctions still posed challenges. For the ResNet18 model, which also included attention layers, anthracnose was once misclassified as a powdery mildew leaf. Unripe strawberries were classified as ripe four times, and ripe ones were misclassified as ripe five times. This shows that it is the weakest algorithm in terms of distinguishing ripe and unripe classes. The inclusion of attention layers helped to some extent, but further improvements are needed. Despite these minor misclassifications, the overall results are promising. The high accuracy rates

achieved by all three models indicate that they are effective in identifying and classifying strawberry diseases and ripeness stages. These results suggest that the models, particularly the vision transformer, are well suited for practical applications in agricultural settings. The minor misclassifications highlight specific areas for improvement but do not significantly detract from the models' overall performance.

C-curves (Arrow C Curves) were also used to evaluate and compare the performance of three models across different classes. The C-curve shows the cumulative distribution of F1 scores, providing a clear visual representation of how well each model performs across all categories. This plot is useful for identifying strengths and weaknesses in the models' classification abilities. In Figure 2.9, the ViT's C-curve shows high performance in most classes, with some variability indicating areas for improvement. The C-curves help highlight differences between the models, offering insights into their robustness and fairness in classification. This makes the plot especially valuable in understanding how each model performs in a multi-class setting and where enhancements might be needed.

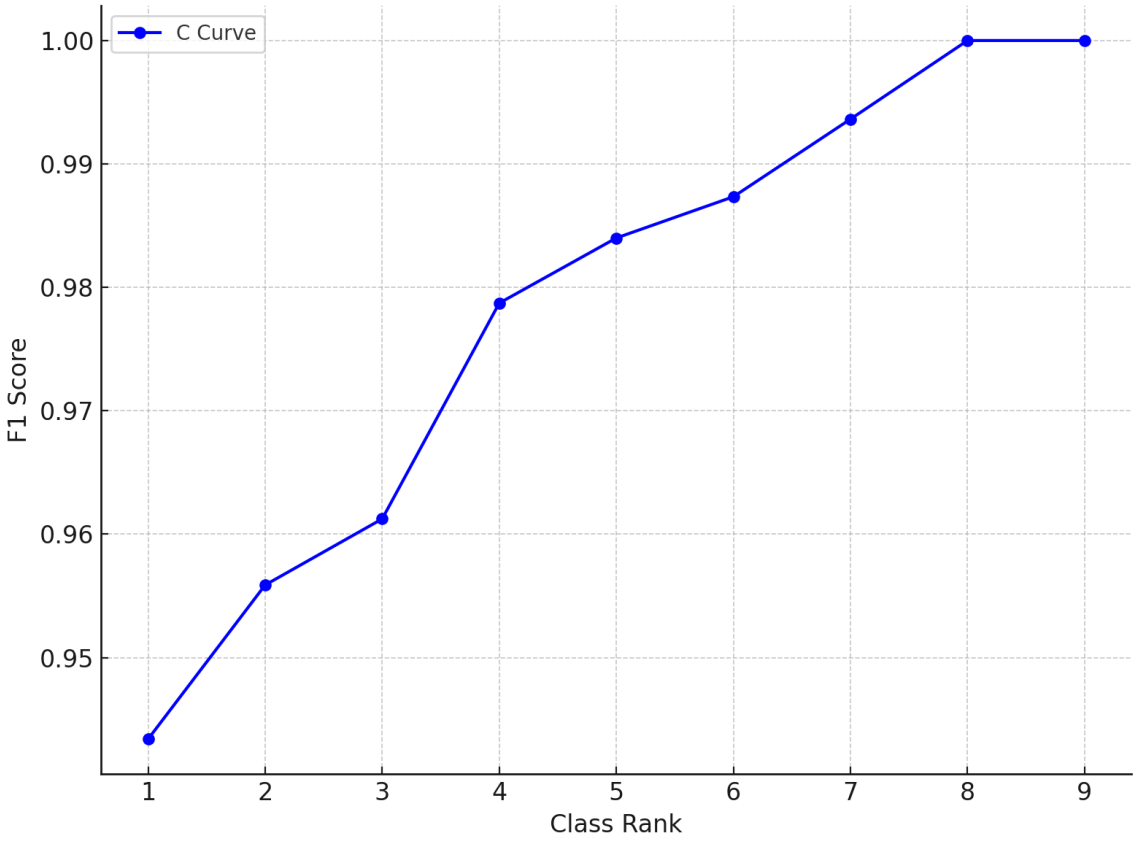


Figure 2.9 C-curve for the best model, ViT.

These misclassifications highlight the need for enhancing the dataset with more diverse samples to help the models learn subtle differences. The use of attention layers has shown promise in improving model performance by focusing on relevant features, but additional refinements are necessary. By addressing the issues, the models can better support farmers in accurately identifying and managing strawberry diseases, thereby promoting healthier crops and more efficient agricultural practices. Upon an analysis of the confusion matrices, it is evident that, for this specific task, after the modifications, the ViT is the more appropriate selection. The observation can be made that there are more true positives, and the occurrences of false negatives and positives are reduced. In order to ensure the model's ability to distinguish between ripe and unripe strawberries, an unripe image was deliberately placed in the ripe folder for testing purposes. The confusion matrix generated for the ViT demonstrates that only one unripe image was misclassified, providing insight into the model's performance in correctly identifying ripe and unripe strawberries. In terms of the efficiency of the weighted random sampler, the AGHRNet study (33) addressed class imbalance through a hybrid loss function that combines cross-entropy loss and dice loss, resulting in a segmentation accuracy of 77.79% mIoU (mean intersection over union) and 89.46% mPA (mean pixel accuracy). Another study (13)utilized data-level techniques such as oversampling and augmentation to balance the dataset, achieving improved segmentation accuracy, though specific accuracy metrics are not detailed. When compared to a weighted random sampler approach, which adjusts the sampling probability to balance the class distribution during training, this method provides a more comprehensive solution. The weighted random sampler helps to address class imbalance by ensuring that minority class samples are more likely to be selected, thereby mitigating bias. This, along with the augmentation of classes with a small number of images, helps to achieve better results.

2.4 Discussion

From the results shown above, the vision transformer model presents better performance overall. ViT, MobileNetV2, and ResNet18 reached their highest accuracy values of 98.4%, 98.1%, and 97.9%, respectively; in this particular context, however, where the class distribution is relatively not balanced, accuracy loses reliability. In spite of this, the precision, defined in Equation (3), which is used to determine how many of the predicted positive instances are truly positive, reached almost 98% with the ViT. It is a metric often used to minimize the occurrence of false positives (32), resulting, in this case, in a more reliable and less waste of good and healthy strawberries. Food waste is a major issue (34),and reducing food waste (35) could help feed many of those suffering from food insecurity and starvation unnecessarily (36,37).Overall, effective discrimination between strawberries and non-strawberries was achieved by all three of the algorithms. The classification task faced increased difficulty with images of diseased leaves due to their higher visual complexity and crowding. Additionally, addressing one of the challenges encountered in this project, the imbalanced and small-sized nature of the image dataset, necessitated careful consideration, augmentation, and the assignment of appropriate weights.

2.5 Future work

Finally, it should be pointed out that a more balanced dataset could potentially alter the results. Future work can investigate the impact of the dataset balance on the results. By leveraging a more extensive dataset featuring high-quality images, a larger and balanced dataset could be generated, enabling the implementation of more sophisticated algorithms with appropriate modifications to facilitate the generalization of the models. In addition, by fixing the distance between the camera and the strawberries, future work could also enable the determination of the size of the strawberries for automatic yield monitoring in both conventional (38,39) and agrivoltaics-based crop systems growing under solar panels (40–42). This would be the next step in a fully autonomous open-source system for strawberry harvesting (43).The work presented here represents the inception of a project aimed at the integration of machine learning into the quality control process for berries, particularly strawberries. In evaluating machine learning for strawberry disease detection, the approach used here built on and extended the methodologies of established studies. The application of Mask R-CNNby Afzaal et al. (8) and convolutional neural networks by Xiao et al. (44) highlights the efficacy of deep learning models in accurately identifying plant diseases, especially straw-berries. The research advances revealed in this study were obtained by combining these findings with integrating vision transformers, which Kamilaris and Prenafeta-Boldú (2) and Turkoglu et al. (45) suggested could further optimize disease detection in agricultural applications. The robustness of these models in handling varied and complex datasets is critical for achieving high accuracy, as demonstrated in the results and supported by Lee et al. (46). The objective of this experiment was to develop a model capable of advancing the current standards in the classification and recognition of strawberry status through images. This model can be used for both outdoor traditional strawberry farming as well as in controlled environment production systems designed for strawberry cultivation. The wall system organizes plants in rows and facilitates water circulation from a reservoir, moving along rails to the top before descending. Throughout the growth cycle on these vertical walls, daily images can be captured to monitor the progress and health of the strawberries. This can be accomplished with cameras on each wall or mobile cameras with robots on rails, wires, or mobile rolling robots. Additionally, through the utilization of cameras and the application of image preprocessing methods to mitigate the impact of sunlight, these algorithms can be effectively extended for outdoor farming applications. To further advance the application of machine learning in strawberry cultivation, the exploration of hybrid models that combine the strengths of CNNs and vision transformers is proposed. This approach is supported by the potential of advanced deep learning techniques for crop disease detection, as discussed by Xiao et al. [43]. Additionally, the enhancement of data preprocessing techniques to more effectively address class imbalances is noted by Buda et al. [14]. The integration of real-time disease detection systems into agricultural practices, envisioned by Xiao et al. [43] and Afzaal et al. [7], is expected to drastically improve operational efficiencies and crop health management.

Chapter 3

3 Blender-Enhanced Synthetic Image Generation: Advancing SwinUNet Segmentation Techniques through Controlled Data Environments

3.1 Introduction

Deep learning, particularly through the use of models with a large number of parameters, has become a pivotal technique for tackling complex problems in machine vision (47). These models require extensive training on diverse visual data, but the high costs and time required to acquire and annotate large datasets can be prohibitive (48). To circumvent these challenges, data augmentation (DA) is often employed to artificially enhance the size of training datasets. This involves altering existing data to simulate various real-world conditions—such as different viewing angles, object deformations, and camera distortions—while maintaining the original labels (49). Such techniques are particularly useful in scenarios where data is scarce, of poor quality, imbalanced, or difficult to obtain due to high costs or other restrictions (48,50,51).

Many computer vision tasks require specifically tailored data formats and annotations, which makes broadly-annotated, publicly available datasets often unsuitable for these specialized requirements. In such cases, creating custom training data from scratch is necessary (51). In fields requiring custom data formats and annotations, like object grasping (52), and manipulation, synthetic data generation using modern image synthesis techniques proves invaluable. These techniques, including 3D modeling, allow for scalable resolutions and customizable content, providing flexibility for specific use cases (47). For instance, synthetic images have been successfully applied in monitoring additive manufacturing processes (53).

Bridging these advancements in data creation, the integration of novel architectural innovations such as the Vision Transformer (ViT) represents a pivotal shift (54). The introduction of Transformers, particularly the ViT (54), has revolutionized fields beyond its initial application in natural language processing (NLP) (24) ViT adapts to various computer vision tasks such as image classification, semantic segmentation, and object detection by effectively processing long-range dependencies without built-in assumptions (54–57). This allows the ViT to effectively identify and process long-range dependencies in data, requiring fewer built-in assumptions. This characteristic improves the model's ability to independently learn from the input data. When the ViT is trained with a sufficient amount of data, it performs exceptionally well, surpassing the effectiveness of state-of-the-art CNNs (54). However, standard ViT models lack segmentation heads necessary for tasks like semantic pixel-level classification, prompting the use of architectures like SETR and Swin Transformer for more complex segmentation tasks (58,59). It is worth noting that both CNN-based and Transformer-based models are used for this purpose (56). There is a

limitation with standard ViT models; they are not directly applicable to segmentation tasks as they do not have segmentation heads, which are essential for pixel classification. Segmentation Transformer (SETR) (60) and Swin Transformer (61) architectures can be utilized for complex segmentation tasks in computer vision, contrasting them with simpler image classification tasks.

A critical characteristic of the ViT model is its substantial need for extensive data sets to facilitate effective training. Dosovitskiy et al. (54) demonstrate that the performance of ViT models scales positively with the increasing size of the training data. This dependency highlights a dual aspect of utilizing ViT in research and practical applications—while it demands large, diverse datasets, it also offers the potential for superior performance if these requirements are met. Such dynamics underscore the importance of resource allocation in data collection to fully exploit the capabilities of ViT models in advanced computational tasks. , there are instances where data is scarce or difficult to obtain for model training, and collecting data can also be costly (48).

This study aims at training a ViT model to identify ripe and unripe strawberries on growth walls using synthetic data for training to avoid the issue of data scarcity and compare the results with conventional data collection methods. The solution involved generating synthetic strawberry images along with their corresponding masks using an open-source software named Blender (62). This approach helped create a more robust dataset, allowing for effective training of the model. Subsequently, the augmented images were trained and evaluated using SwinUNet (63)as a method for transfer learning and Deep Domain Confusion (64) for domain adaptation. The trained model was then tested on real-time images captured by cameras from growth walls.

3.2 Backgrounds

In this section, the applications of semantic segmentation and Vision Transformers (ViTs) will be explored first. SwinUNet (63) will then be discussed as an alternative to ViTs. Data Augmentation will be introduced, along with its advantages and the benefits of using Blender (62) to generate images. Additionally, challenges such as domain adaptation and the use of the Dice Similarity Coefficient as a method will be examined.

3.2.1 Semantic Segmentation

Semantic segmentation, crucial for detailed visual scene analysis, offers pixel-level precision advantageous for medical imaging and autonomous driving due to its accuracy in object delineation (65). It surpasses simple detection by precisely outlining object shapes, crucial for complex interactions (59). It demands, however, substantial computational resources for training and execution and struggles with the scarcity of annotated datasets, especially in specialized fields (66). These factors make semantic segmentation both powerful for in-depth analysis and challenging to implement due to high resource demands and data acquisition challenges.

Semantic segmentation serves a pivotal role in various application areas, including remote sensing, medical imaging, and agriculture (67), each benefiting from its ability to identify and delineate distinct regions within images accurately. Remote sensing imagery, combined with computer vision and AI, is essential for analyzing complex features across large geographical areas (68). Neural networks facilitate the processing of this vast data, enabling precise object detection and semantic segmentation (69). Research has improved the processing of high-resolution remote sensing images for semantic segmentation by optimizing the ViT architecture (56).This optimization involves the strategic addition of layers and attention mechanisms to the models, enhancing their efficiency. Notable advancements include the development of models such as the Efficient Transformer and the Wide-Context Transformer (70) which have demonstrated superior performance in these image analysis tasks.

In agriculture, particularly, semantic segmentation is utilized to enhance precision farming by enabling robots to detect and classify crops and weeds effectively, facilitating targeted weeding actions(67). This technology relies on CNN-based models for real-time segmentation, distinguishing between elements like sugar beet plants, weeds, and background solely using RGB data.

Experts in various fields extensively analyze these images, a time-consuming process. To enhance efficiency, deep learning methods have been employed for automatic feature extraction, improving medical image analysis. Notably, the U-Net architecture (65), initially developed for medical purpose, has seen various enhancements and applications across different medical datasets, including heart, lesion, and liver segmentation. There have been recent advancements in the application of ViT architectures to the agricultural and medical sector, particularly through the introduction of TransUNet (71) and Swin-Unet(63). These hybrid Transformer models integrate features from the U-Net and have demonstrated enhanced accuracy in segmenting. It also points out a significant challenge: the datasets are generally smaller compared to extensive natural image datasets, which include millions of images of landscapes, people, animals, and vehicles (20,72). This disparity in dataset size can pose difficulties in effectively training and developing these sophisticated models. Employing open-source software such as Blender can be extremely beneficial for generating additional images, thus addressing the challenge of limited datasets in medical imaging (73). By creating synthetic images, Blender can help enhance the volume and variety of data available for training sophisticated models like TransUNet or Swin-Unet, leading to better performance and more accurate results.

3.2.2 Vision Transformers

A standard transformer architecture was employed to handle both token embeddings and 2D image data, transforming images into sequences of flattened patches that were converted into fixed-size vectors with positional embeddings to preserve spatial context (54). The architecture consists of alternating layers of self-attention and multi-layer perceptrons (MLP), complemented by layer normalization and residual connections, which facilitates the simultaneous processing of text and image data. Vision Transformers differ from traditional CNNs mainly in their limited use of inductive bias, choosing instead to learn spatial relationships among image patches from the ground up. Adjustments to the

positional embeddings are made to accommodate images of different resolutions. Through its self-attention mechanism, the model assigns importance to each patch, improving its understanding of the relationships within the image and focusing on the most relevant segments for performing various vision-based tasks. The classification outcome is determined by a classification head that interprets the encoder's output, effectively summarizing these processes in a coherent workflow. Vision transformers are extensively employed for detecting strawberries and monitoring their quality in the field. Zheng et al. [31] evaluated ViT models for strawberry quality classification, detailing a ViT-based method integrated with a Support Vector Machine (SVM) that achieves a recognition accuracy of 98.1%.; The use of Vision ViT was studied for detecting strawberry diseases, enhanced by transfer learning to improve precision. This approach classified diseases across seven categories of strawberry parts, achieving an accuracy and F1-score of 0.927 on the Strawberry Disease Detection dataset [32]; Also, LS-YOLOv8s integrated with the LW-Swin Transformer module, was developed to enhance the detection and classification of strawberry ripeness. The model significantly outperformed existing models, achieving 94.4% detection precision [33]. These methods have achieved high accuracies exceeding 95% while maintaining relatively low computational costs. Most recently in a study, the ViT was fine-tuned on augmented strawberry images, achieving 98.4% accuracy and nearly 99% precision in disease classification and ripeness detection, demonstrating its potential to enhance agricultural practices through precise crop monitoring (77).

3.2.2.1 SwinUNet

SwinUNet (63) is an innovative architecture that combines the strengths of Swin Transformers with the proven structure of the U-Net (65). The Swin Transformer is a variant of vision transformers that employs shifted windows for self-attention, effectively processing hierarchical and multi-scale image features to boost performance across diverse computer vision applications (61). Although SwinUNet is common in medical imaging, it has recently been used for agricultural uses (76). This hybrid model integrates the hierarchical Swin Transformers, known for their efficiency in processing vision tasks due to the shifted windowing technique in attention mechanisms (61). This integration allows SwinUNet to adeptly manage different scales of attention across its layers, enhancing its capability to maintain detailed context and achieve fine segmentation accuracy (61). SwinUNet has demonstrated exceptional performance, surpassing traditional CNN-based models in various segmentation tasks, such as organ and tumor delineation and fruit ripeness detection (63,76). It offers substantial improvements in terms of Dice scores and shows enhanced generalization across different imaging modalities. Studies (49) further validate its superiority, especially in handling complex images, where it outperforms standard U-Nets in both precision and computational efficiency (63). The flexibility of SwinUNet to adapt to various image sizes and types underscores its potential as a versatile tool in medical diagnostics, promising to boost diagnostic accuracy and thereby improve patient outcomes.

3.2.3 Data Augmentation

Contemporary computer vision techniques predominantly utilize CNN and deep learning, which depend on extensive amounts of labeled data and significant computational

power (73). The difficulty in gathering and labeling large datasets complicates the widespread application of computer vision, particularly in areas where data annotation is notably laborious (78). Additionally, the quality and diversity of the data are critical for developing robust computer vision models. Due to these challenges, there is increasing interest in synthetic image data as a more affordable and readily available option for training (48). There are various data augmentation methods utilized such as: 1) geometric data augmentation (79), 2) photometric techniques such as color jittering (80), lighting perturbation (81), and image denoising (82) modify the visual properties like contrast, brightness, and noise of images to enhance model robustness against such variations.

Generative AI methods offer the most promising potential for creating synthetic datasets for complex computer vision tasks (48). Generative modeling methods are deep learning techniques that utilize specialized neural architectures to learn and replicate the statistical distributions of target data, facilitating the creation of synthetic training data (51). The famous example of it is generative adversarial networks (GANs) (83). There are methods to synthesize data based on computer graphics modelling. Aforementioned methods normally approach data augmentation as a 2D transformation of an image within its own domain, using various techniques to generate variations of the original data in 2D space. These methods lack a semantic 3D context, making them potentially inadequate in representing true real-world variations (48). Utilizing a software such as Blender (62) can facilitate the synthesis of images in various shapes and forms, allowing for the generation of any desired number of samples. In crop disease and quality assessment, Blender provides the flexibility to design and incorporate any specific details required for the analysis.

Supervised learning in deep neural networks has some limitations, especially for tasks in computer vision and NLP that require extensive labeled data, which is challenging and expensive to obtain (69). It highlights that while transfer learning offers some relief, it falls short for specific applications like satellite imagery segmentation due to mismatches in data domains. Consequently, achieving high accuracy remains difficult. Self-supervised learning within Transformer architectures is an effective solution to these data-intensive challenges. This approach, which learns from unlabeled data, is made similar to human vision, offering a promising avenue to bypass the constraints of traditional supervised methods. In this study, the objective was to train a ViT model for segmentation without the necessity of acquiring external data. Previous studies (77) have indicated that a shortage of training data can lead to methods that incur high computational costs.

Blender is a free, open-source 3D graphics software that encompasses the entirety of the 3D process, including modeling, animating, rendering, and compositing (84). It offers several key advantages that make it ideal for generating synthetic datasets. First, Blender is equipped with a physics-based rendering engine (85), making it capable of producing high-quality, photorealistic images. A particular advantage of using Blender is its ability to precisely control the appearance of surfaces and materials through its shader nodes system (86). An extensive library of textures and backgrounds is also available, enabling a wide variety of realistic objects and environments to be created. Applying these resources can enhance the diversity and realism of synthetic datasets, making them more representative of the real world. The introduction of a node-based procedural workflow in version 2.92 has also significantly enhanced Blender's capabilities (87). This feature

facilitates the creation and manipulation of complex geometries without the need for manual modeling, allowing for a high degree of flexibility and control over the object. It also enables the randomization of an object's geometry within each frame, allowing for numerous object variations to be incorporated into a single scene animation. Additionally, Blender supports scripting and automation (88), helping to address issues related to data scarcity and imbalance by providing a scalable and cost-effective solution for generating large volumes of high-quality training data tailored to specific needs.

Previous research has demonstrated Blender's effectiveness in generating synthetic datasets for various computer vision applications across different domains (47,89). For example, in the realm of additive manufacturing, Blender has been employed to generate comprehensive datasets for semantic segmentation of 3D-printed parts, improving real-time failure analysis systems by accurately detecting various structural elements (53). In industrial applications, Blender has been used to create synthetic images for steel defect recognition, leading to improved performance in classifying and segmenting defects on steel slabs (90). Blender has also been instrumental in developing a quality inspection system for scaffolding, combining synthetic and real datasets to train models for assessing structural safety (91). In agriculture, Blender has been used to develop synthetic datasets for crop size estimation, effectively addressing challenges such as occlusions and perspective distortions (92). It has also enabled the creation of realistic datasets for object detection in sweet pepper cultivation through procedural generation, enhancing the training of deep learning models for both object detection and semantic segmentation (93)

3.2.4 Domain Adaptation

A trained deep learning model exhibits optimal performance on test data when the data shares the same distribution as the training set (94). For instance, datasets comprising images from mobile phones differ significantly from those captured with high-end cameras, which can lead to the failure of traditional transfer learning approaches. In such cases, each new dataset requires initial annotation followed by re-training of the model to accommodate the new data characteristics. Domain adaptation offers a solution to this challenge by enabling adjustments to a pre-trained model to enhance its performance on new datasets without the need for re-training (64). Domain Adaptation is a machine learning strategy that modifies models trained on one domain to achieve high performance on a different yet related domain. This approach is especially beneficial when there is a lack of labeled data in the target domain but an abundance in the source domain, as it aligns the data distributions across domains (95,96). A typical application is adapting image recognition models trained on controlled environment images to accurately recognize images taken under diverse real-world conditions (97). This approach conserves computational resources and, in the case of unsupervised domain adaptation, eliminates the need for labeling the new data. In this study, the source domain is the data generated by Blender and the target domain is the data received by the camera.

3.2.5 Deep Domain Confusion

In CNNs there is a problem of dataset bias in deep learning. A standard supervised deep CNN, even when trained on extensive datasets, fails to completely eliminate bias when

tested against benchmarks (64). Deep Domain Confusion (DDC) is a method developed to address domain shift in domain adaptation, where a model trained in one domain does not perform well in another. Introduced by Tzeng et al. (64), DDC works by integrating a domain confusion loss into the training process, encouraging the model to learn features that are invariant across domains. This approach helps in enhancing the generalizability of deep learning models, particularly in applications like image recognition and natural language processing where the training and application environments greatly differ (73). The strategy involves optimizing a combined loss function that accounts for both prediction accuracy and domain confusion, thereby improving the model's effectiveness across varied domains. For example, Kamnitsas et al.(98) demonstrated how unsupervised domain adaptation could be applied to MRI scans, where models trained on data from one set of MRI machines significantly improved their performance on scans from different machines, showcasing the approach's effectiveness in medical imaging contexts.

3.2.6 Dice Similarity Coefficient and Jaccard Index

The Dice Similarity Coefficient (DSC) (99) and the Jaccard Index (100) are commonly used to evaluate segmentation accuracy with. The DSC is calculated as 2True Positive/(2True Positive + False Positive + False Negative) , and the Jaccard Index is defined by True Positive/(True Positive + False Positive + False Negative). These metrics are essential for assessing how well segmentation algorithms, typically optimized using (weighted) cross-entropy, perform in practical applications. Cross entropy optimization, however, often leads to a discrepancy between the training loss and the evaluation metrics, which can adversely affect model performance (99). To address this issue, recent innovations in computer vision have introduced methods like soft-Dice, soft-Jaccard, and Lovász-softmax (101). These approaches aim to directly optimize the desired metric, aligning the training objectives more closely with the performance evaluation metrics. These methodologies have been explored to determine whether weighted cross-entropy can act effectively as a surrogate for the Dice or Jaccard metrics.

3.3 Methods

This section details the procedures used to assess the ripeness and health of strawberries, outlining how the data was collected, and the specific methods and algorithms employed in the analysis.

3.3.1 Dataset

3.3.1.1 Data Preparation

Petsiuk et al. (102) described a method for creating a procedural strawberry plant model using Blender geometry nodes in which parameters such as scale, rotation, and position of each plant part were precisely controlled through a node network, allowing for flexible adjustments to simulate natural plant variations. Their work forms the foundation for this study, and the following section details how their model was utilized and adapted to suit our specific application.

In this study, Blender version 4.0.2 was employed to generate the synthetic dataset of strawberry plants for computer vision training. The scene (shown in Figure 3.1) comprised a vertical grow wall, peat cups with soil for a strawberry plant, a single strawberry plant, a track for the camera, and the camera itself. It was intended to closely resemble the actual growing conditions of strawberries planted in an indoor vertical grow wall located in the agrivoltaic agrotunnel at the Western Innovation for Renewable Energy Deployment in London, ON, Canada. The vertical grow wall was carefully replicated within Blender to ensure the synthetic data would be representative of the actual testing conditions. Lighting within the scene was achieved through the ambient illumination provided by the default forest environment texture, which mimicked natural light conditions.

The random node (103) was further integrated into the geometry node network, introducing variability by randomizing parameters such as the size and orientation of leaves, the curvature of stems, and the color of strawberries. The randomization was carefully constrained within realistic limits to avoid non-realistic appearances, ensuring that each rendered frame was unique.

Several improvements were also made to the original leaf model. The shape of the leaf was refined to more closely align with a strawberry plant’s true leaf morphology by utilizing Blender's knife tool (104) to cut out the shape from an imported image-as-plane, based on a reference image (105). A shader nodes setup (86) was created that allows for variation in leaf color between lighter and darker green. Additionally, the ability to randomly vary the leaf's curl was introduced. Finally, the Principled BSDF node (106) was used to increase the roughness of the leaf’s surface, further improving its realism.

Object instancing (107) was used to manage the high number of individual components efficiently. Each instance was subjected to the same randomization parameters, maintaining consistency across different parts of the plant while ensuring variability between frames. The camera also was set on a track and animated to move along it, changing angles and perspectives in each frame. This method allowed for the creation of highly realistic yet diverse synthetic datasets.

Blender's compositing tool (108) was used to create masks for image segmentation. This tool assigned separate pixel values to each part of the plant. For each image, the strawberry was assigned its own value while the rest of the plant and background were assigned a different value. The cycles rendering engine (85) was used to render the images with the number of samples set to 256. Initially, images were rendered using a CPU, which was later switched to a GPU to decrease rendering time. Using a GPU, less than two days were required to generate 4,000 images. While the animation rendering process occasionally resulted in strawberries being covered by leaves, and object intersection, such as overlapping leaves, was a challenge, it generally did not detract from the realism and was often negligible.

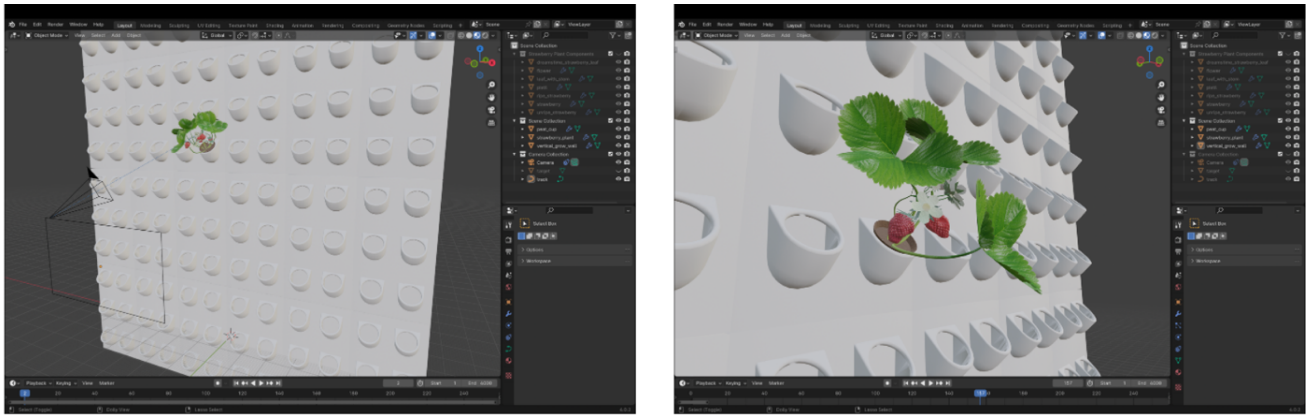


Figure 3.1. Scenes from the grow wall and strawberries in Blender

In this method, images were generated using the open source Blender (109), and specific pixel values were assigned to different elements within each image, including the background (comprising the wall and leaves), as well as the ripe and unripe strawberries. Figure 3.2 displays representative generated images of ripe and unripe strawberries alongside their corresponding masks.

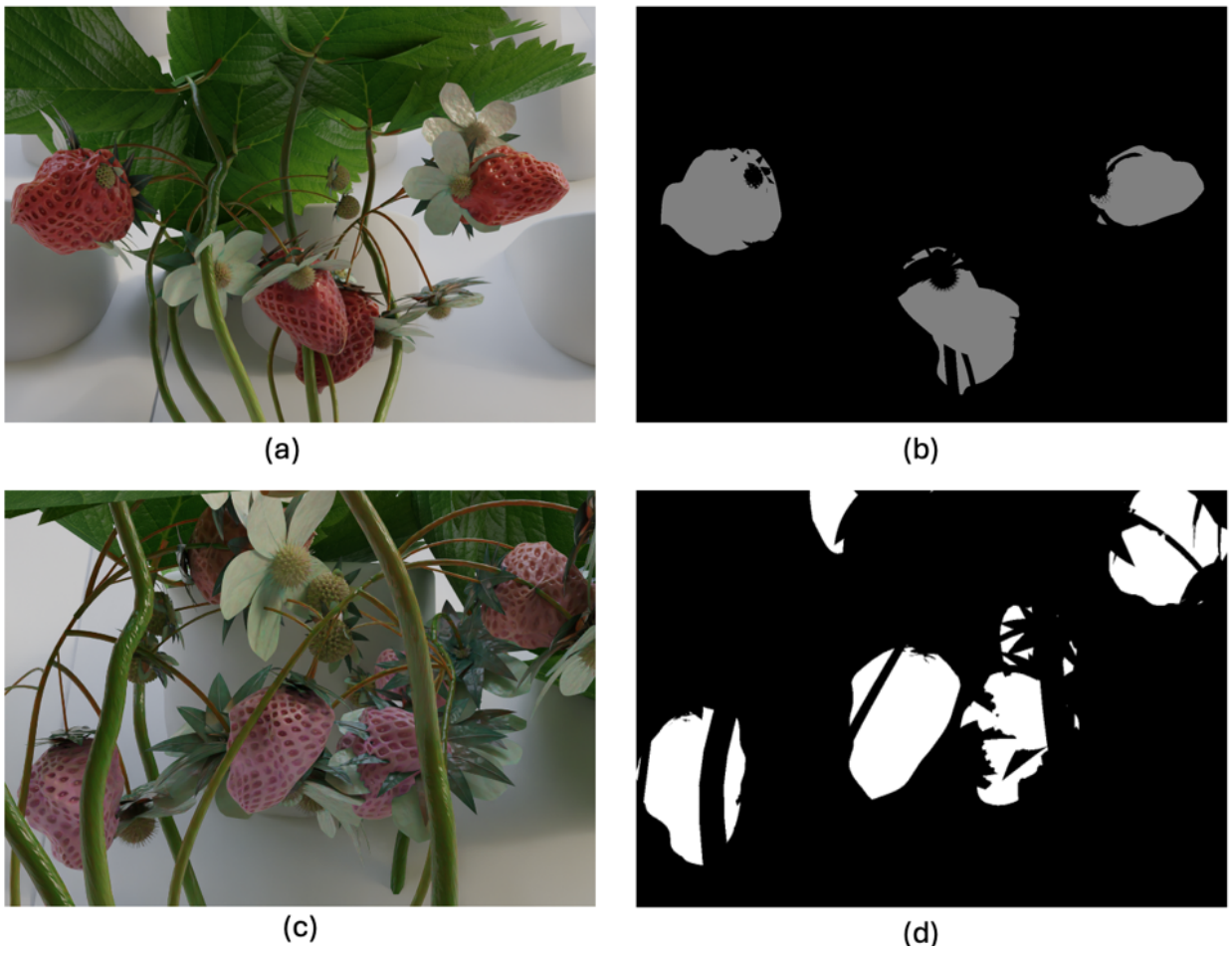


Figure 3.2. Images generated with Blender with their corresponding masks. (a) image of ripe strawberry with its mask (b). (c) is the image of unripe strawberry with its corresponding mask (d).

3.3.1.2 Data pre-annotation

The generated data was divided into training and validation subsets. To assess the model's performance in real-time and under real-world conditions, a collection of images taken in an agrivoltaics agricultural tunnel (110) was employed for testing. These images (representative example shown in Figure 3) lacked predefined masks, necessitating pre-annotation. For this experiment, to pre-annotate, the images were grayscaled and masked creation was carried out using Roboflow. (111)

Table 3.1 presents the quantity of images utilized for training and testing within each category. Also, along with the real images with corresponding masks that were fed to the saved model for testing.

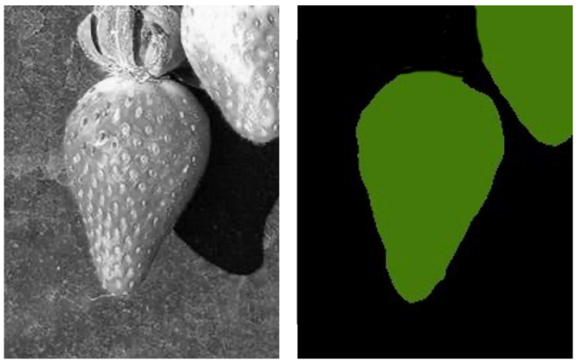


Figure 3.3. left (the image taken from the camera and grey scaled), right (pre-annotated image)

Table 3. 1 – Number of images in each data set.

	Ripe	Unripe
Training (Synthetic)	5000	5000
Evaluation (Synthetic)	1000	1000
Testing (Real)	300	300

Deep domain confusion was implemented, and the model was saved and tested on real-time images.

Following the preprocessing steps, the images were processed using SwinUNet for both training and validation, and the resultant model was evaluated on the real-time data with images captured from a Crealty (112) webcam. Deep domain confusion was implemented during evaluation to reduce the differences between the source and target domain.

The computational resources for this project were provided by the Digital Alliance of Canada, which included the use of an A100 GPU. All code used in this study were written in Python and have been made accessible on the Open Science Framework (OSF) (109). Detailed specifications of the model are presented in Table 3.2. DSC method was implemented to evaluate the accuracy of the pixels segmented on both the ground truth sample and the segmented one. The DSC ranges from 0 to 1, where a value of 1 indicates a 100% match between the segments in the ground truth sample and the segmented target, representing perfect agreement.

Table 3.2. Details of the architecture and algorithms used for this experiment.

Parameters	Value
Programming language	Python 3
Architecture	SwinUNet
Domain Adaptation Method	Deep Domain Confusion
Evaluation Metric	Dice Similarity Coefficient
Loss Function	Weighted Cross Entropy Loss
Source Domain	Blender Generated Images
Target Domain	Creality CRCC-S7 HD 1080 3D printer Web Camera
Computing Power	Digital Alliance of Canada (A100 GPU)
Epochs	400

3.4 Results

SwinUNet and deep domain confusion techniques were applied to a dataset comprising 5000 images per class augmented by Blender, categorized as ripe and unripe, each accompanied by corresponding masks. The DSC for both classes on the evaluation set, achieving 98% for ripe and 97.4% for unripe classes. The performance on the test set (real images) did not mirror these results, registering DSCs of 92% and 90% for ripe and unripe classes respectively. Despite increasing the number of images, the real-time data testing results continued to lag behind the evaluation set, underscoring the persistent challenge of

domain disparity. This pattern underscores the inherent limitations of domain adaptation methods in fully bridging the gap between varied data domains.

Table 3.3 Validation and test results using 5000 images for training.

	DSC (%) - validation	DSC (%) - test	Accuracy
Ripe	98	92	98.74
Unripe	97.4	90	97.2

In this study, the model's initial training involved 200 images per class, resulting in DSC of 58% for ripe strawberries and 56.4% for unripe ones on the test set. Subsequently, when the training set was expanded to 1,000, 4,000 and 5,000 per class, brought the DSC to 92% for ripe and 90% for unripe strawberries as shown in Figure 5. These results clearly demonstrate that within the same environment, increasing the number of training images substantially improves both the accuracy and the DSC of the model.

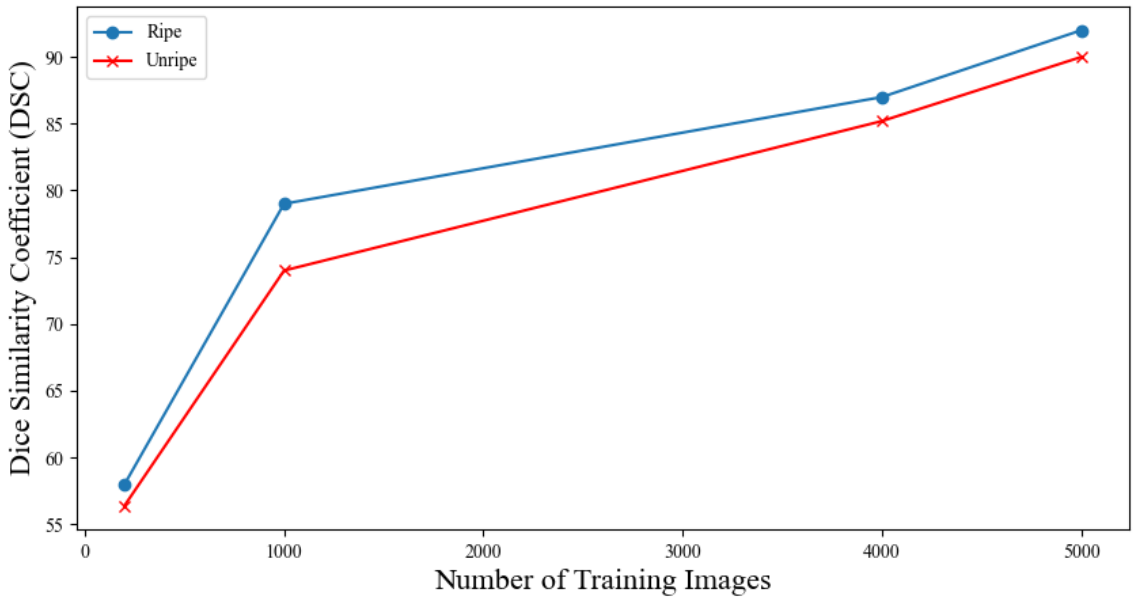


Figure 3.4. The rate of DSC increase when increasing the number of training images.

3.5 Discussion

The findings from this study clearly indicate that an increase in the volume of training data significantly enhances the accuracy for each class. In studies involving AlsmViT (113), designed for analyzing foods with similar shapes but differing nutritional values, the top performance using Swinv2 achieved a 93% accuracy rate. Additionally, real-time semantic segmentation of crops and weeds (67) achieved an 89.5% success rate. In projects focusing on strawberry disease identification (75) using vision transformer-based models, the original ViT model reached a 98% accuracy level. However, these projects were limited to identifying and classifying strawberries using a dataset provided by Afzaal et al. (8), with the results obtained solely from augmented data. In a prior study(77) concerning strawberry disease and ripeness, the datasets, albeit imbalanced, originated from authentic images and were consistent across both training and testing phases within the same domain. The implementation of ViTs alongside weighted sampling methodologies facilitated a classification precision of 98.9%. In contrast, the current study utilized an evaluation set from the same domain as the training data, achieving comparable DSC and accuracy levels in a segmentation task. Nevertheless, the test data, sourced from a different domain, yielded reliable yet slightly less accurate results compared to the aforementioned study.

With the current methodology, the generation of additional data effectively addresses the issue of data scarcity. A noticeable discrepancy remains between the accuracy on real-time data and that on the evaluation set. This gap may potentially be narrowed by enhancing the quality of the generated images, suggesting a focus on improving image generation processes as a means to boost real-time data performance.

3.6 Future Work

Looking ahead, there are several promising avenues for enhancing the functionality and accuracy of this image processing system. First, there is potential for improvement in the quality of the generated images to make them more closely resemble real images. This enhancement would likely lead to better model training outcomes and more accurate segmentation in practical applications. Second, in the future, the adoption of stereoscopic cameras (114) for capturing real-time images could significantly advance our capabilities. By utilizing such technology, it would be possible to accurately measure the size and volume of objects, such as strawberries. This dimensional data could provide valuable additional information to determine the optimal timing for harvesting.

Additionally, exploring various Transformer models could further enrich our understanding and effectiveness in segmentation tasks. Experimenting with models like the SETR (69) could provide insightful comparisons with the currently employed algorithms, potentially revealing strengths or weaknesses that could inform future improvements and adaptations in our approach.

Also, future research could explore the effects of training with zoomed-in images to determine if synthesizing images from a close distance influences accuracy. Additionally, generating images that contain both ripe and unripe strawberries within a single frame could be investigated, as this approach more closely mirrors real-world conditions.

3.7 Conclusions

In this research, synthetic data created using Blender was utilized to train a ViT model. Gathering data in agricultural fields can be challenging and time-consuming due to the extended growth periods required for crops. Originally, the ViT is not tailored for segmentation tasks; thus, Swin-UNet, a variant within the ViT family, was employed via transfer learning for this purpose. To evaluate the model's performance on real-world data, this data had to be pre-annotated, and masks generated, necessitating the use of domain adaptation techniques. The model achieved a high accuracy of 98.5% on a validation set. The performance on real-time data did not match this, which is expected, but it still maintained a reliable accuracy and DSC above 90%, proving effective for applications like fruit detection.

Chapter 4

4 Conclusion

This thesis has examined the application of advanced machine learning techniques, specifically Vision Transformers (ViTs) and synthetic data generation, to improve disease detection and ripeness classification in strawberry cultivation. The integration of these technologies into agricultural practices represents a stride towards more precise and automated crop management systems. This chapter synthesizes the findings from the conducted studies and discusses potential avenues for future research.

4.1 Summary of Results

4.1.1 Vision Transformers for Disease and Quality Detection

The first study, detailed in Chapter 2, focused on utilizing Vision Transformers to optimize strawberry disease detection and quality assessment. The results demonstrated that ViTs, enhanced with attention mechanisms, significantly outperform traditional convolutional neural networks in both accuracy and precision. The ViT model achieved a classification precision nearing 98.9%, underlining its capability to handle the nuanced variations in disease symptoms and fruit ripeness. This performance is attributed to the model's ability to process long-range dependencies and its robustness in learning from imbalanced datasets, enhanced by weighted sampling techniques.

4.1.2 Synthetic Data and SwinUNet for Advanced Segmentation

Chapter 3 explored the creation and utilization of synthetic datasets through Blender to train the SwinUNet, a model adapted from ViT for complex segmentation tasks. This approach addressed the challenge of data scarcity and allowed for the controlled simulation of diverse agricultural environments. The SwinUNet, trained on these synthetic datasets, proved to be highly effective, achieving Dice Similarity Coefficients that closely approached those obtained with real-world data. The model demonstrated acceptable generalizability, which is crucial for adapting to the variability existed in agricultural applications.

4.2 Implications of the Findings

The findings from this thesis underscore the potential of machine learning to revolutionize agricultural practices through enhanced disease detection and crop monitoring. By reducing reliance on manual labor and subjective assessments, these technologies can significantly improve the efficiency, accuracy, and timeliness of interventions. This not only contributes to better crop health and yield but also promotes sustainable agricultural practices by optimizing resource use.

4.3 Future Work

4.3.1 Expansion to Other Crops and Conditions

Expanding the application of these models to other types of crops and environmental conditions would also be valuable. This includes testing the models in different climates, soils, and with various types of crop diseases, which could help in developing more robust, universal models for agricultural applications.

4.3.2 Advanced Synthetic Data Generation

Advancements in synthetic data generation techniques hold significant potential for enhancing the training efficiency and effectiveness of machine learning models used in agricultural applications. By developing more sophisticated methods that more accurately mimic real-world variability, synthetic data can facilitate the training of models that are both highly adaptable and precise. Future research should focus on several promising strategies to improve the accuracy and functionality of the image processing systems in these settings. Enhancing the realism of synthesized images is expected to significantly boost both model training and segmentation accuracy. Furthermore, the integration of stereoscopic cameras could provide precise volumetric measurements of agricultural produce, such as strawberries, optimizing harvest timings. Additionally, exploring the effects of training models on zoomed-in imagery and generating composite images that include both ripe and unripe strawberries could enhance model robustness and better replicate complex field conditions. Together, these initiatives could greatly advance precision agriculture, improving decision-making processes and increasing the accuracy of crop yield predictions through refined technological applications.

Chapter 5

5 References

1. Shorif Uddin M, Bansal JC. Computer Vision and Machine Learning in Agriculture. Algorithms for Intelligent Systems [Internet]. Vol. 2. Springer Singapore; 2021. Available from: https://doi.org/10.1007/978-981-33-6424-0.

2. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture [Internet]. 2018 Apr 1 [cited 2024 Jun 9];147:70–90. Available from: https://www.sciencedirect.com/science/article/pii/S0168169917308803

3. Hadipour-Rokni R, Askari Asli-Ardeh E, Jahanbakhshi A, Esmaili paeen-Afrakoti I, Sabzi S. Intelligent detection of citrus fruit pests using machine vision system and convolutional neural network through transfer learning technique. Computers in Biology and Medicine [Internet]. 2023 Mar 1 [cited 2023 Dec 6];155:106611. Available from: https://www.sciencedirect.com/science/article/pii/S0010482523000768

4. Zhou C, Hu J, Xu Z, Yue J, Ye H, Yang G. A Novel Greenhouse-Based System for the Detection and Plumpness Assessment of Strawberry Using an Improved Deep Learning Technique. Frontiers in Plant Science [Internet]. 2020 [cited 2023 Dec 6];11. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2020.00559

5. Lello F, Dida M, Mkiramweni M, Matiko J, Akol R, Nsabagwa M, et al. Fruit fly automatic detection and monitoring techniques: A review. Smart Agricultural Technology [Internet]. 2023 Oct 1 [cited 2023 Dec 6];5:100294. Available from: https://www.sciencedirect.com/science/article/pii/S2772375523001235

6. Surianarayanan C, Lawrence JJ, Chelliah PR, Prakash E, Hewage C. A Survey on Optimization Techniques for Edge Artificial Intelligence (AI). Sensors [Internet]. 2023 Jan [cited 2023 Nov 8];23(3):1279. Available from: https://www.mdpi.com/1424-8220/23/3/1279

7. Chen Y, Lee WS, Gan H, Peres N, Fraisse C, Zhang Y, et al. Strawberry Yield Prediction Based on a Deep Neural Network Using High-Resolution Aerial Orthoimages. Remote Sensing [Internet]. 2019 Jan [cited 2023 Dec 6];11(13):1584. Available from: https://www.mdpi.com/2072-4292/11/13/1584

8. Afzaal U, Bhattarai B, Pandeya YR, Lee J. An Instance Segmentation Model for Strawberry Diseases Based on Mask R-CNN. Sensors [Internet]. 2021 Jan [cited 2023 Dec 6];21(19):6565. Available from: https://www.mdpi.com/1424-8220/21/19/6565

9. Daubney HA. Cultivated Berries. The Canadian Encyclopedia [Internet]. 2015 [cited 2024 May 16]; Available from: https://www.thecanadianencyclopedia.ca/en/article/cultivated-berries

10. Rahamathunnisa U, Nallakaruppan MK, Anith A, Kumar KS S. Vegetable Disease Detection Using K-Means Clustering And Svm. In IEEE; 2020 [cited 2023 Oct 24]. p. 1308–11. Available from: https://ieeexplore.ieee.org/document/9074434

11. Puttemans S, Tits L, Vanbrabant Y, Goedeme T. Automated visual fruit detection for harvest estimation and robotic harvesting. [cited 2023 Dec 6]; Available from: DOI:10.1109/IPTA.2016.7820996

12. Pérez-Borrero I, Marín-Santos D, Gegúndez-Arias ME, Cortés-Ancos E. A fast and accurate deep learning method for strawberry instance segmentation. Computers and Electronics in Agriculture [Internet]. 2020 Nov [cited 2023 Oct 24];178:105736. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168169920300624

13. Vuttipittayamongkol P, Elyan E, Petrovski A. On the class overlap problem in imbalanced data classification. Knowledge-Based Systems [Internet]. 2021 Jan 5 [cited 2024 May 16];212:106631. Available from: https://www.sciencedirect.com/science/article/pii/S0950705120307607

14. Ojo MO, Zahid A. Improving Deep Learning Classifiers Performance via Preprocessing and Class Imbalance Approaches in a Plant Disease Detection Pipeline. Agronomy [Internet]. 2023 Mar [cited 2024 May 16];13(3):887. Available from: https://www.mdpi.com/2073-4395/13/3/887

15. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks [Internet]. 2018 Oct [cited 2024 May 16];106:249–59. Available from: http://arxiv.org/abs/1710.05381

16. Ketabforoosh K. Strawberry Images [Internet]. Open Science Framework; 2023 [cited 2023 Dec 6]. Available from: https://osf.io/https://osf.io/n9kjq

17. Rezaei-Dastjerdehei MR, Mijani A, Fatemizadeh E. Addressing Imbalance in Multi-Label Classification Using Weighted Cross Entropy Loss Function. In: 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME) [Internet]. 2020 [cited 2023 Dec 6]. p. 333–8. Available from: https://ieeexplore.ieee.org/document/9319440

18. PyTorch documentation — PyTorch 2.1 documentation [Internet]. [cited 2023 Dec 6]. Available from: https://pytorch.org/docs/stable/index

19. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [Internet]. arXiv; 2021 [cited 2024 Mar 17]. Available from: http://arxiv.org/abs/2010.11929

20. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks [Internet]. arXiv; 2019 [cited 2023 Oct 24]. Available from: http://arxiv.org/abs/1801.04381

21. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. 2016 [cited 2023 Dec 6]. p. 770–8. Available from: https://ieeexplore.ieee.org/document/7780459

22. Caron M, Touvron H, Misra I, Jégou H, Mairal J, Bojanowski P, et al. Emerging Properties in Self-Supervised Vision Transformers [Internet]. arXiv; 2021 [cited 2024 Mar 17]. Available from: http://arxiv.org/abs/2104.14294

23. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [Internet]. arXiv; 2017 [cited 2024 Mar 17]. Available from: http://arxiv.org/abs/1704.04861

24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need [Internet]. arXiv; 2023 [cited 2024 Mar 17]. Available from: http://arxiv.org/abs/1706.03762

25. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation [Internet]. 1989 Dec [cited 2024 Mar 17];1(4):541–51. Available from: https://direct.mit.edu/neco/article/1/4/541-551/5515

26. Chen J, Zhang D, Suzauddola M, Zeb A. Identifying crop diseases using attention embedded MobileNet-V2 model. Applied Soft Computing [Internet]. 2021 Dec 1 [cited 2023 Dec 6];113:107901. Available from: https://www.sciencedirect.com/science/article/pii/S1568494621008231

27. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions [Internet]. arXiv; 2014 [cited 2024 Mar 24]. Available from: http://arxiv.org/abs/1409.4842

28. Andonie R. Hyperparameter optimization in learning systems. Journal of Membrane Computing [Internet]. 2019 Oct 16; Available from: https://digitalcommons.cwu.edu/compsci/105

29. Great Learning Team. Hyperparameter Tuning with GridSearchCV [Internet]. [cited 2023 Dec 6]. Available from: https://www.mygreatlearning.com/blog/gridsearchcv/

30. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing [Internet]. 2021 Sep 10 [cited 2023 Dec 6];452:48–62. Available from: https://www.sciencedirect.com/science/article/pii/S092523122100477X

31. Digital Research Alliance of Canada [Internet]. 2023 [cited 2023 Dec 6]. Digital Research Alliance of Canada. Available from: https://alliancecan.ca/en/node/10

32. Gad AF. Paperspace Blog. 2020 [cited 2023 Dec 6]. Accuracy, Precision, and Recall in Deep Learning. Available from: https://blog.paperspace.com/deep-learning-metrics-precision-recall-accuracy/

33. Zheng Z, Hu Y, Guo T, Qiao Y, He Y, Zhang Y, et al. AGHRNet: An attention ghost-HRNet for confirmation of catch-and-shake locations in jujube fruits vibration harvesting. Computers and Electronics in Agriculture [Internet]. 2023 Jul 1 [cited 2024 May 16];210:107921. Available from: https://www.sciencedirect.com/science/article/pii/S0168169923003095

34. Girotto F, Alibardi L, Cossu R. Food waste generation and industrial uses: A review. Waste management (New York, NY). 2015 Jun 27;45.

35. Dhiman S. Sustainable Social Entrepreneurship: Serving the Destitute, Feeding the Hungry, and Reducing the Food Waste. In 2020. p. 193–208.

36. Denkenberger D, Pearce J. Feeding Everyone No Matter What: Managing Food Security After Global Catastrophe. Feeding Everyone No Matter What: Managing Food Security After Global Catastrophe. 2014.

37. Meyer T, Pearce J. How Easy is it to Feed Everyone? Economic Alternatives to Eliminate Human Nutrition Deficits. Food Ethics. 2022 Nov 21;8.

38. Oğuz İ, Oğuz Hİ, Kafkas NE, Oğuz İ, Oğuz Hİ, Kafkas NE. Strawberry Cultivation Techniques. In: Recent Studies on Strawberries [Internet]. IntechOpen; 2022 [cited 2024 May 17]. Available from: https://www.intechopen.com/chapters/81392

39. Zacharaki K, Monaghan J, Bromley J, Vickers L. Opportunities and challenges for strawberry cultivation in urban food production systems. PLANTS, PEOPLE, PLANET. 2024 Jan 2;6.

40. Dinesh H, Pearce JM. The potential of agrivoltaic systems. Renewable and Sustainable Energy Reviews [Internet]. 2016 Feb 1 [cited 2024 May 17];54:299–308. Available from: https://www.sciencedirect.com/science/article/pii/S136403211501103X

41. Widmer J, Christ B, Grenz J, Norgrove L. Agrivoltaics, a promising new tool for electricity and food production: A systematic review. Renewable and Sustainable Energy Reviews [Internet]. 2024 Mar 1 [cited 2024 May 17];192:114277. Available from: https://www.sciencedirect.com/science/article/pii/S1364032123011358

42. Wydra K, Vollmer V, Busch C, Prichta S, Wydra K, Vollmer V, et al. Agrivoltaic: Solar Radiation for Clean Energy and Sustainable Agriculture with Positive Impact on Nature [Internet]. IntechOpen; 2023 [cited 2024 May 17]. Available from: https://www.intechopen.com/online-first/87330

43. Woo S, Uyeh DD, Kim J, Kim Y, Kang S, Kim KC, et al. Analyses of Work Efficiency of a Strawberry-Harvesting Robot in an Automated Greenhouse. Agronomy [Internet]. 2020 Nov [cited 2024 May 17];10(11):1751. Available from: https://www.mdpi.com/2073-4395/10/11/1751

44. Xiao JR, Chung PC, Wu HY, Phan QH, Yeh JLA, Hou MTK. Detection of Strawberry Diseases Using a Convolutional Neural Network. Plants (Basel) [Internet]. 2020 Dec 25 [cited 2024 Jun 9];10(1):31. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7823414/

45. TÜRKOĞLU M, HANBAY D. Plant disease and pest detection using deep learning-based features. Turkish Journal of Electrical Engineering and Computer Sciences [Internet]. 2019 Jan 1;27(3):1636–51. Available from: https://journals.tubitak.gov.tr/elektrik/vol27/iss3/6

46. Lee SH, Chan CS, Mayo S, Remagnino P. How Deep Learning Extracts and Learns Leaf Features for Plant Classification. Pattern Recognition. 2017 May 1;71.

47. Man K, Chahl J. A Review of Synthetic Image Data and Its Use in Computer Vision. Journal of Imaging [Internet]. 2022 Nov [cited 2024 Jul 27];8(11):310. Available from: https://www.mdpi.com/2313-433X/8/11/310

48. Mumuni A, Mumuni F, Gerrar NK. A survey of synthetic data augmentation methods in computer vision. Mach Intell Res [Internet]. 2024 Mar 20 [cited 2024 Jul 27]; Available from: http://arxiv.org/abs/2403.10075

49. Liu X, Ono K, Bise R. arXiv.org. 2023 [cited 2024 Jul 27]. Mixing Data Augmentation with Preserving Foreground Regions in Medical Image Segmentation. Available from: https://arxiv.org/abs/2304.13490v1

50. Moreno-Barea FJ, Jerez JM, Franco L. Improving classification accuracy using data augmentation on small data sets. Expert Systems with Applications [Internet]. 2020 Dec 15 [cited 2024 Jul 27];161:113696. Available from: https://www.sciencedirect.com/science/article/pii/S0957417420305200

51. Saini M, Susan S. Deep transfer with minority data augmentation for imbalanced breast cancer dataset. Applied Soft Computing [Internet]. 2020 Dec 1 [cited 2024 Jul 27];97:106759. Available from: https://www.sciencedirect.com/science/article/pii/S1568494620306979

52. Lin Y, Tang C, Chu FJ, Vela P. Using Synthetic Data and Deep Networks to Recognize Primitive Shapes for Object Grasping. 2020. 10494 p.

53. Petsiuk A, Singh H, Dadhwal H, Pearce JM. Synthetic-to-Real Composite Semantic Segmentation in Additive Manufacturing. Journal of Manufacturing and Materials Processing. 2024 Apr;8(2):66.

54. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 2021 Jun 3 [cited 2024 Jul 24]; Available from: http://arxiv.org/abs/2010.11929

55. Chen CF, Fan Q, Panda R. CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification [Internet]. arXiv; 2021 [cited 2024 Jul 24]. Available from: http://arxiv.org/abs/2103.14899

56. Xu Z, Zhang W, Zhang T, Yang Z, Li J. Efficient Transformer for Remote Sensing Image Segmentation. Remote Sensing [Internet]. 2021 Jan [cited 2024 Jul 24];13(18):3585. Available from: https://www.mdpi.com/2072-4292/13/18/3585

57. Park N, Kim S. How Do Vision Transformers Work? [Internet]. arXiv; 2022 [cited 2024 Jul 24]. Available from: http://arxiv.org/abs/2202.06709

58. Mumuni A, Mumuni F, Gerrar NK. arXiv.org. 2024 [cited 2024 Jul 27]. A survey of synthetic data augmentation methods in computer vision. Available from: https://arxiv.org/abs/2403.10075v2

59. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. 2015 [cited 2024 Jul 25]. p. 3431–40. Available from: https://ieeexplore.ieee.org/document/7298965

60. Zheng S, Lu J, Zhao H, Zhu X, Luo Z, Wang Y, et al. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers [Internet]. arXiv; 2021 [cited 2024 Jul 24]. Available from: http://arxiv.org/abs/2012.15840

61. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows [Internet]. arXiv; 2021 [cited 2024 Jul 25]. Available from: http://arxiv.org/abs/2103.14030

62. Foundation B. blender.org. 2024 [cited 2024 Jul 24]. Blender - Free and Open 3D Creation Software. Available from: https://www.blender.org/

63. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. 2021 May 12 [cited 2024 Jul 25]; Available from: http://arxiv.org/abs/2105.05537

64. Tzeng E, Hoffman J, Zhang N, Saenko K, Darrell T. Deep Domain Confusion: Maximizing for Domain Invariance [Internet]. arXiv; 2014 [cited 2024 Jul 25]. Available from: http://arxiv.org/abs/1412.3474

65. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [Internet]. arXiv; 2015 [cited 2024 Jul 25]. Available from: http://arxiv.org/abs/1505.04597

66. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J. A Review on Deep Learning Techniques Applied to Semantic Segmentation [Internet]. arXiv; 2017 [cited 2024 Jul 29]. Available from: http://arxiv.org/abs/1704.06857

67. Milioto A, Lottes P, Stachniss C. Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In: 2018 IEEE International Conference on Robotics and Automation (ICRA) [Internet]. 2018 [cited 2024 Jul 27]. p. 2229–35. Available from: https://ieeexplore.ieee.org/abstract/document/8460962?casa_token=oXohkq7i2_8AAAAA:FhMW2QQG3gCBlyo4tD1rWZH8P4WG6Ef9rPly2exWaF6W_pigRIOZsuyZ8T0X5B2HDAU53d-szdM

68. Boguszewski A, Batorski D, Ziemba-Jankowska N, Dziedzic T, Zambrzycka A. LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water and Roads from Aerial Imagery [Internet]. arXiv; 2022 [cited 2024 Jul 25]. Available from: http://arxiv.org/abs/2005.02264

69. Thisanke H, Deshan C, Chamith K, Seneviratne S, Vidanaarachchi R, Herath D. Semantic Segmentation using Vision Transformers: A survey. 2023 May 5 [cited 2024 Jul 23]; Available from: http://arxiv.org/abs/2305.03273

70. Ding L, Lin D, Lin S, Zhang J, Cui X, Wang Y, et al. Looking Outside the Window: Wide-Context Transformer for the Semantic Segmentation of High-Resolution Remote Sensing Images. IEEE Trans Geosci Remote Sensing [Internet]. 2022 [cited 2024 Jul 25];60:1–13. Available from: http://arxiv.org/abs/2106.15754

71. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation [Internet]. arXiv; 2021 [cited 2024 Jul 25]. Available from: http://arxiv.org/abs/2102.04306

72. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2012 [cited 2024 Jul 27]. Available from: https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

73. Atapour-Abarghouei A, Breckon TP. Real-Time Monocular Depth Estimation Using Synthetic Data With Domain Adaptation via Image Style Transfer. In 2018 [cited 2024 Jul 27]. p. 2800–10. Available from: https://openaccess.thecvf.com/content_cvpr_2018/html/Atapour-Abarghouei_Real-Time_Monocular_Depth_CVPR_2018_paper.html

74. Zheng H, Wang G, Li X. Identifying strawberry appearance quality by vision transformers and support vector machine. Journal of Food Process Engineering [Internet]. 2022 [cited 2024 Jul 28];45(10):e14132. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/jfpe.14132

75. Nguyen HT, Tran TD, Nguyen TT, Pham NM, Nguyen Ly PH, Luong HH. Strawberry disease identification with vision transformer-based models. Multimed Tools Appl [Internet]. 2024 Feb 8 [cited 2024 Jul 28]; Available from: https://doi.org/10.1007/s11042-024-18266-0

76. Yang S, Wang W, Gao S, Deng Z. Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer. Computers and Electronics in Agriculture [Internet]. 2023 Dec 1 [cited 2024 Jul 27];215:108360. Available from: https://www.sciencedirect.com/science/article/pii/S0168169923007482

77. Aghamohammadesmaeilketabforoosh K, Nikan S, Antonini G, Pearce JM. Optimizing Strawberry Disease and Quality Detection with Vision Transformers and Attention-Based Convolutional Neural Networks. Foods [Internet]. 2024 Jan [cited 2024 Jun 14];13(12):1869. Available from: https://www.mdpi.com/2304-8158/13/12/1869

78. Sakaridis C, Dai D, Van Gool L. Semantic Foggy Scene Understanding with Synthetic Data. Int J Comput Vis [Internet]. 2018 Sep [cited 2024 Jul 28];126(9):973–92. Available from: http://arxiv.org/abs/1708.07819

79. Ornek AH, Ceylan M. Comparison of Traditional Transformations for Data Augmentation in Deep Learning of Medical Thermography. In: 2019 42nd International Conference on Telecommunications and Signal Processing (TSP) [Internet]. 2019 [cited 2024 Jul 28]. p. 191–4. Available from: https://ieeexplore.ieee.org/document/8769068

80. Kim EK, Lee H, Kim JY, Kim S. Data Augmentation Method by Applying Color Perturbation of Inverse PSNR and Geometric Transformations for Object Recognition Based on Deep Learning. Applied Sciences [Internet]. 2020 Jan [cited 2024 Jul 28];10(11):3755. Available from: https://www.mdpi.com/2076-3417/10/11/3755

81. Sakkos D, Shum HPH, Ho ESL. Illumination-Based Data Augmentation for Robust Background Subtraction [Internet]. arXiv; 2019 [cited 2024 Jul 28]. Available from: http://arxiv.org/abs/1910.08470

82. Kotwal A, Bhalodia R, Awate SP. Joint desmoking and denoising of laparoscopy images: 2016 IEEE 13th International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2016. 2016 IEEE International Symposium on Biomedical Imaging [Internet]. 2016 Jun 15 [cited 2024 Jul 28];1050–4. Available from: http://www.scopus.com/inward/record.url?scp=84978427203&partnerID=8YFLogxK

83. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks [Internet]. arXiv; 2014 [cited 2024 Jul 28]. Available from: http://arxiv.org/abs/1406.2661

84. Foundation B. About [Internet]. blender.org. [cited 2024 Jul 29]. Available from: https://www.blender.org/about/

85. Cycles - Blender 4.2 Manual [Internet]. [cited 2024 Jul 29]. Available from: https://docs.blender.org/manual/en/latest/render/cycles/index.html

86. Shader Nodes - Blender 4.2 Manual [Internet]. [cited 2024 Jul 29]. Available from: https://docs.blender.org/manual/en/latest/render/shader_nodes/index.html

87. Geometry Nodes - Blender 4.2 Manual [Internet]. [cited 2024 Jul 29]. Available from: https://docs.blender.org/manual/en/latest/modeling/geometry_nodes/index.html

88. Quickstart - Blender Python API [Internet]. [cited 2024 Jul 29]. Available from: https://docs.blender.org/api/current/info_quickstart.html

89. Rohe DP, Jones EMC. Generation of Synthetic Digital Image Correlation Images Using the Open-Source Blender Software. Exp Tech. 2022 Aug 1;46(4):615–31.

90. Boikov A, Payor V, Savelev R, Kolesnikov A. Synthetic Data Generation for Steel Defect Detection and Classification Using Deep Learning. Symmetry. 2021 Jul;13(7):1176.

91. Kim A, Lee K, Lee S, Song J, Kwon S, Chung S. Synthetic Data and Computer-Vision-Based Automated Quality Inspection System for Reused Scaffolding. Applied Sciences. 2022 Jan;12(19):10097.

92. Dolata P, Wróblewski P, Mrzygłód M, Reiner J. Instance segmentation of root crops and simulation-based learning to estimate their physical dimensions for on-line machine vision yield monitoring. Computers and Electronics in Agriculture. 2021 Nov 1;190:106451.

93. Procedural Generation of Synthetic Dataset for Robotic Applications in Sweet Pepper Cultivation [Internet]. [cited 2024 Jul 29]. Available from: https://ieeexplore.ieee.org/document/9954643

94. Wang W, Li H, Ding Z, Wang Z. Rethink Maximum Mean Discrepancy for Domain Adaptation. 2020 Jul 1 [cited 2024 Jul 25]; Available from: http://arxiv.org/abs/2007.00689

95. Csurka G. Domain Adaptation for Visual Applications: A Comprehensive Survey [Internet]. arXiv; 2017 [cited 2024 Jul 29]. Available from: http://arxiv.org/abs/1702.05374

96. Ganin Y, Lempitsky V. Unsupervised Domain Adaptation by Backpropagation [Internet]. arXiv; 2015 [cited 2024 Jul 29]. Available from: http://arxiv.org/abs/1409.7495

97. Tzeng E, Hoffman J, Saenko K, Darrell T. Adversarial Discriminative Domain Adaptation. In 2017 [cited 2024 Jul 29]. p. 7167–76. Available from: https://openaccess.thecvf.com/content_cvpr_2017/html/Tzeng_Adversarial_Discriminative_Domain_CVPR_2017_paper.html

98. Kamnitsas K, Baumgartner C, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks [Internet]. arXiv; 2016 [cited 2024 Jul 30]. Available from: http://arxiv.org/abs/1612.08894

99. Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, et al. Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Cham: Springer International Publishing; 2019. p. 92–100.

100. Costa L da F. Further Generalizations of the Jaccard Index [Internet]. arXiv; 2021 [cited 2024 Jul 29]. Available from: http://arxiv.org/abs/2110.09619

101. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Medical Imaging [Internet]. 2015 Aug 12 [cited 2024 Jul 25];15(1):29. Available from: https://doi.org/10.1186/s12880-015-0068-x

102. Aliaksei Petsiuk, Xiang Li, Joshua M. Pearce. Procedural synthetic strawberry plant dataset generation for AI-based segmentation.

103. Random Value Node - Blender 4.2 Manual [Internet]. [cited 2024 Jul 29]. Available from: https://docs.blender.org/manual/en/latest/modeling/geometry_nodes/utilities/random_value.html

104. Knife Tool — Blender Manual [Internet]. [cited 2024 Jul 29]. Available from: https://docs.blender.org/manual/en/2.81/modeling/meshes/editing/subdividing/knife.html

105. Dreamstime [Internet]. [cited 2024 Jul 29]. Strawberry Leaf Isolated on White Background. the Texture of the Leaf and Streaks is Clearly Visible Stock Photo - Image of streaks, fruit: 118646616. Available from: https://www.dreamstime.com/strawberry-leaf-isolated-white-background-texture-leaf-streaks-clearly-visible-strawberry-leaf-isolated-image118646616

106. Principled BSDF - Blender 4.2 Manual [Internet]. [cited 2024 Jul 29]. Available from: https://docs.blender.org/manual/en/latest/render/shader_nodes/shader/principled.html

107. Instancing - Blender 4.2 Manual [Internet]. [cited 2024 Jul 29]. Available from: https://docs.blender.org/manual/en/latest/scene_layout/object/properties/instancing/index.html

108. Compositing - Blender 4.2 Manual [Internet]. [cited 2024 Jul 29]. Available from: https://docs.blender.org/manual/en/latest/compositing/index.html

109. Parfitt J, Ketabforoosh K. Blender generated strawberry images - unripe. 2024 Jul 24 [cited 2024 Jul 28]; Available from: https://osf.io/4g5nx/

110. Asgari N, Jamil U, Pearce JM. Net Zero Agrivoltaic Arrays for Agrotunnel Vertical Growing Systems: Energy Analysis and System Sizing. Sustainability. 2024 Jan;16(14):6120.

111. Roboflow: Computer vision tools for developers and enterprises [Internet]. 2024 [cited 2024 Jul 25]. Roboflow. Available from: https://roboflow.com/

112. Creality Canada [Internet]. 2024 [cited 2024 Jul 25]. CRCC-S7 HD 1080P Web Camera. Available from: https://store.creality.com/ca/products/crcc-s7-hd-1080p-web-camera

113. Gao X, Xiao Z, Deng Z. High accuracy food image classification via vision transformer with data augmentation and feature augmentation. Journal of Food Engineering [Internet]. 2024 Mar 1 [cited 2024 Jul 27];365:111833. Available from: https://www.sciencedirect.com/science/article/pii/S0260877423004314

114. Islam A, Asikuzzaman M, Khyam M, Noor-A-Rahim M, Pickering MR. Stereo vision-based 3D positioning and tracking. 2020 Jul 1 [cited 2024 Jul 29]; Available from: https://acquire.cqu.edu.au/articles/journal_contribution/Stereo_vision-based_3D_positioning_and_tracking/13416131/2

6 Curriculum Vitae

Name: Kimia Aghamohammadesmaeilketabforoosh

Post-secondary Education and Degrees:
Amirkabir University of Technology
Tehran, Tehran, Iran
2015-2020 B.Sc. in Chemical Engineering

The University of Western Ontario
London, Ontario, Canada
2022 - 2024 MESc. in Software Engineering

Related Work Experience
Research Assistant
The University of Western Ontario, FAST Lab
2022 – 2024

Teaching Assistant
The University of Western Ontario, ECE Department
2024

Publications:

K. Aghamohammadesmaeilketabforoosh, S. Nikan, G. Antonini, and J. M. Pearce, “Optimizing Strawberry Disease and Quality Detection with Vision Transformers and Attention-Based Convolutional Neural Networks,” *Foods*, vol. 13, no. 12, Art. no. 12, Jan. 2024, doi: 10.3390/foods13121869.