Estimation of the number of wheat spikes and spikelets based on deep learning and computer vision

Yusa Dursun, Vaibhav Jhajharia

Master's thesis
2024:E55

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematics

Centrum Scientiarum Mathematicarum

Estimation of the number of wheat spikes and spikelets based on deep learning and computer vision

Vaibhav Jhajharia
va2504jh-s@student.lu.se

Batuhan Dursun
yu4421du-s@student.lu.se

September 7, 2024

Master's thesis work carried out at
the Department of Mathematics, Lund University.

Supervisor: Anders Heyden, anders.heyden@math.lu.se

Examiner: Mikael Nilsson, mikael.nilsson@math.lth.se

Abstract

Computer vision and artificial intelligence have been used with various techniques for high-throughput plant phenotyping in research and industry. Among different plant parts, the yield has the highest importance. This study utilizes a deep learning approach for yield-based wheat phenotyping, presenting an algorithm that empowers farmers and agricultural researchers to assess crop yield performance in agricultural fields. We propose a methodology using combined classical image processing and advanced deep neural networks like YOLO, Fast-RCNN and SAM to segment and detect wheat heads from field images and count grains to make inference of the plant density in the field and average grain count per wheat head from the visual images.

Acknowledgements

We would like to express our sincere gratitude to our supervisor, Anders Heyden, for his consistent support and guidance throughout this project. We would also like to thank Marc Ahlse (Sony supervisor) for helping out with the technicalities of different camera sensors and for advising about the workflow throughout the project. We would also like to appreciate the support of Ajit Nehe for providing the agricultural setup access and test environment that made this work possible.

Contents

Chapter 1
Introduction

The growing global population and increased food demand have emphasized the necessity for advancements in agriculture. Plant phenotyping traditionally relied on manual methods involving labor-intensive measurements and observations of plant traits, which are time-consuming and prone to human error. However, with the emergence of AI and machine learning, automated phenotyping has become a game-changer. By utilizing cutting-edge technologies with sensors, drones, and computer vision algorithms, automated phenotyping enables rapid and accurate assessment of plant traits.

By harnessing AI algorithms, farmers can analyze vast datasets encompassing factors like soil quality, weather patterns, biomass and crop health to implement informed precision agriculture practices. This enables tailored interventions such as optimal irrigation scheduling, targeted pest management, and personalized fertilizer application, thereby maximizing yield while minimizing resource usage and environmental impact.

The goal of the thesis is to identify the characteristics of an optimal sensor setup for 2D/3D reconstruction of wheat plants to enable extraction of wheat plant phenotypic traits. The focus area would be the detection and estimation of wheat head counts and then further detecting and counting grains from the detected wheat heads. A robust system using the sensor setup and computer vision algorithms for the field or uncontrolled settings is to be developed to automate the data collection process in the agriculture research to aid the analysis of the wheat crop best practices.

Wheat is a major food resource, so various methods to estimate total yield have been in research using different sensory devices. [8] AI presents a transformative opportunity to optimize wheat farming, a cornerstone of global food security. With wheat being a major food source for billions worldwide, enhancing its cultivation through AI-powered solutions is paramount. AI-enabled phenotyping techniques facilitate the rapid evaluation of wheat varieties, diseases, photosynthetic performance, accelerating the breeding of high-yielding, resilient cultivars tailored to diverse growing conditions.

Identifying optimal sensor setup for 2D/3D reconstruction of wheat plants for capturing the detailed information suitable for extracting the phenotypic characteristics is critical for

further model development. Researchers can then streamline data collection and extract standardized phenotypic traits efficiently using automatized vision model pipelines. This accelerates crop breeding research and optimize agricultural best practices while maintaining the consistency of the information at large scale.

An RGB camera sensor with good resolution is useful for capturing structure, shape, or texture-based characteristics of a scene. Further enhancement of sensor setup for morphological traits such as leaf angle, leaf area, height of the plant is possible with 3D proximal sensors which add in the depth information for spatial analysis. Light Detection and Ranging (LiDAR), stereo cameras and time of flight (ToF) cameras are some of them. LiDAR sensors use laser to obtain a point cloud representation of the scene. However, they are costly, have low resolution and color information is not present. A combination with RGB camera is needed to obtain color information. On the other hand, a time of flight camera uses illumination and calculates the time taken by the light to reach the objects. ToF camera is sensitive to illumination and the performance of image acquisition is diminished under strong sunlight conditions. Stereo cameras utilize two separate cameras to compute depth by triangulation. The main drawbacks of stereo camera are finding the point correspondences for depth calculation and the depth interpretation errors due to overlapping leaves and the computation that needs to be done on every iteration to calculate the depth.

The major part of the phenotyping research has been done on indoor environment with controlled setups, without accounting for robustness against wind or sunlight effects for outdoor image acquisition. In the field, the acquisition of plant traits from general images is a challenging task due to environmental factors and overlapping of the plant parts in the field images.

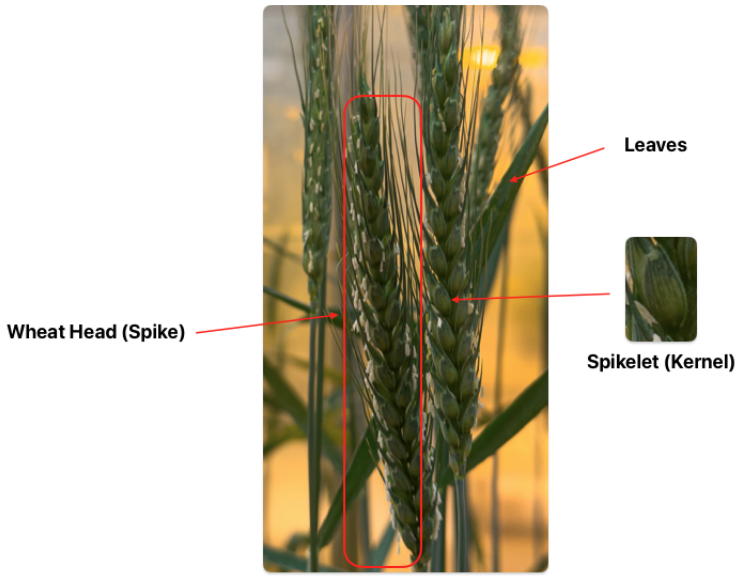


Figure 1.1: Morphological structure of a wheat plant containing wheat heads, leaves and grains

Plant segmentation allows for the precise isolation of individual plants and remove the background part within the images, facilitating detailed analysis of key morphological traits essential for understanding plant development and productivity. Detecting wheat heads

within segmented plants provides critical insights into reproductive characteristics such as spike density, spike length, and grain yield potential, offering valuable information for breeding programs and agronomic decision-making. Furthermore, counting grains within these detected wheat heads enables the quantification of crucial yield-related parameters, including grain number per head and grain size distribution, which are vital indicators of crop performance and economic value. By leveraging these phenotypic analyses, researchers can gain deeper insights into the genetic and environmental factors influencing wheat growth and yield, ultimately developing informed strategies to enhance crop productivity, resilience, and sustainability.

The incorporation of advanced phenotyping techniques into research holds significant long-term benefits for agriculture and research. Standardized protocols ensure consistency and facilitate data sharing, leading to large-scale analyses and meta-analyses that uncover genetic relationships. Automated phenotyping driven by AI enables rapid data processing, contributing to the development of comprehensive phenotypic databases useful for agriculture research at scale as well as to further grow the agriculture AI capabilities. Long-term phenotypic studies deepen our understanding of plant biology, adaptation, and response to stressors, ultimately informing the development of resilient and high-yielding crop varieties.

Chapter 2
Related Work

AI is revolutionizing agricultural research by providing innovative solutions to challenges such as precision agriculture, crop monitoring, genomics, supply chain optimization, robotic farming, and climate resilience. Through machine learning algorithms and data analytics, AI enables farmers to optimize resource use, maximize yields, and minimize environmental impact. Though the role of AI in agriculture is still in starting phase, the possibilities of the impact on the domain and food security are immense.

In this project, our area of interest is to detect and count estimate the wheat heads in an RGB image and then further estimate the total grains in the detected wheat head frames. Currently developed solutions for this task are using different image acquisition methods in a controlled environment or in field environment with limited information capture. The paper [9] explores the counting problem using wheat head images in the controlled settings. The annotation process is automated using watershed algorithm and Cb component of the YCbCr color space. These annotations are used to train a FasterRCNN neural network for detecting and counting the grains in the image.

A light-weight solution for counting wheat spikelets in field conditions [10] is implemented where they take the isolated image of a wheat head and then count the wheat grains in it by using YOLOv5s-T model and reducing the convolution operation in the YOLOv5s backbone network. On the other hand, segmenting plants in the outdoor settings [3] uses stereo-camera images to segment out the plants from the noisy field environment. A robust segmentation method is obtained by also adding HSV color information, helping to manage variable light conditions. Another segmentation work [13] has an approach in form of the established DeepLab V3+ architecture with custom adjustments to the training pipeline and mild changes to the architecture delivers good performance on semantic segmentation task. Differently, 3D reconstruction of small plants in a controlled imaging condition is implemented by extracting 3D point cloud data from multi-view images [11]. These type of projects take advantage of controlled environment setting eliminating the wind and light illumination effects. Parts per object count paper [6] approaches the counting problem in the field settings using a deep neural network trained on the annotated field data for detection

and counting. They are able to get good performance for the counting of banana and grapes in field settings but for the wheat grain counting, the performance is not good enough to be useful for the practical applications. Our approach is focused on detecting the grains with high precision and accuracy followed by estimation of the possible grains in the undetected part of the wheat head.

Apart from RGB sensor, other image acquisition methods are being utilized for extracting different phenotypic traits of the plants. X-ray imaging technique is used for grain counting [14] in controlled environment by creating 3D wheat models and finding grain centroid coordinates to merge the ones representing the same spikelet. Furthermore, hyperspectral and chlorophyll imaging successfully identifies frost and drought stress interactions prior to the commencement of visual symptoms [4]. In addition, using hyperspectral data, rapid quantification of enzyme of spring wheat under frost and drought stress conditions is provided. However these techniques are quite precise in the lab settings but are costly and not optimal for plant phenotyping in the field. Thus, the effectiveness of RGB camera sensor and the depth sensor setup is explored to find an optimal system suitable for the field analytics.

Although the projects above provide various solutions, manual intervention is still required and it acts as a bottleneck to the AI capabilities. Taking plant into a lab or controlled setting itself is a manual and labour intensive task to collect data at scale. So a robust method with less manual intervention and more automated execution is desired to improve and standardize the plant phenotyping in the field.

Availability of good annotated data for model training is a major restriction in the agriculture field. For the supervised counting methods, the annotations of object should be annotated accurately in order to have accurate counting operation and it requires more time to make inference. The counting problem has been handled with detection or regression approaches by researchers [12]. The detection algorithm learns from bounding box or a single dot annotations of all object instances. However, overlap situation of objects, illuminations and perspectives from different angles can reduce the estimation accuracy. Moreover, researchers tend to use density based approaches [5] to overcome this issue. Finally, unsupervised methods with texture based counting approach also seems promising, which relies on tracking motion similarities [2].

Since agriculture is a new emerging field of application for computer vision and artificial intelligence, there is a lot of scope in terms of contributing on the available solutions. Our work involves exploring the ways of developing a robust solution for detecting and counting wheat heads as well as counting grains in the controlled and uncontrolled environment settings. The optimal camera configuration is evaluated based on the ability of the system to capture the information to enable the AI models to perform in a robust way. RGB images captured using the optimal camera configuration are further used for extracting the phenotype characteristics of the wheat crop to aid the research and monitoring in the agriculture domain.

Chapter 3
Theory

Automating wheat plant phenotyping with AI streamlines the analysis of key traits like yield, height, disease resistance, and stress tolerance. Using advanced computer vision and machine learning, this approach rapidly assesses large-scale wheat fields, aiding breeders in selecting superior cultivars and optimizing agricultural practices. By expediting breeding cycles and enhancing resilience to environmental challenges, AI-driven phenotyping contributes to sustainable wheat production and food security.

Specifically for the wheat plant phenotype traits extraction, there are a wide range of options for image acquisition and then further plethora of computer vision algorithms to explore for the wheat head and grains estimation use case.

3.1 Image Acquisition

Image acquisition is significant for the agricultural research automation, ensuring that the extracted phenotype traits are accurate and robust. High-quality images are required because they provide the detailed information needed for precise analysis. The level of detail (spatial resolution) and the ability to monitor changes over time (temporal resolution) are important for tracking crop growth and environmental effects. Using specific spectral bands allows us to see details about plant health and stress. Good image acquisition enables development of accurate artificial intelligence systems for agricultural best practices. Automation with drones or satellites makes data collection faster and more efficient. However, the cost must be weighed against the benefits to get the best value.

Sensors for image acquisition have all their pros and cons and evaluating the sensor setup for the specific use case of plant phenotyping is a major step as it should be able to capture the traits we want to detect with good precision. Without proper setup the captured information might not be good enough for the AI models to detect something.

- **RGB Camera Sensors:** These are commonly used for capturing high-resolution 2D images in different configurations to optimize image quality. RGB cameras can provide

detailed visual information about crops, allowing for analysis of various traits such as color, shape, and texture. For tasks such as counting wheat heads and grains within them, 2D image-based segmentation and detection algorithms are employed. These algorithms analyze the 2D images captured by the cameras to identify and delineate specific features of interest, such as individual wheat heads or grains. Once detected, these features can be quantified and analyzed for various purposes, such as yield estimation or pest monitoring.

- **Time of Flight (ToF) and Stereo Cameras:** These are utilized for capturing 3D images, which offer depth information in addition to the 2D visual data. ToF camera, with a resolution of 640 x 480 pixels, measure the time it takes for light to travel to the object and back, allowing for the creation of a 3D model. Stereo cameras use the principle of triangulation to calculate depth by comparing images from two slightly offset cameras. 3D images are valuable for assessing spatial phenotypic characteristics such as length, angle, and volume of crops, which are crucial for various agricultural applications including yield estimation and plant breeding.

3.2 Plant segmentation

State of the art deep neural networks for variety of computer vision tasks have good performance on the general use cases, but they need further tuning to perform well on new specific use cases like agricultural research. In these scenarios, utilizing the domain specific knowledge and applying pre-processing steps on the new corpus helps to tune the state of the art models with limited amount of data available. There are a variety of classical as well as deep learning image processing methods to enhance the image data for better model performance.

1. **Classical method:** Otsu thresholding is utilized to segment out the plants to account for the changing lighting conditions. Otsu thresholding is a widely used technique in image processing for automatic image segmentation. It calculates an optimal threshold value based on the histogram of pixel intensities in the image. The goal is to maximize the inter-class variance between foreground and background pixels, effectively separating them into distinct regions. By determining this threshold dynamically, Otsu thresholding adapts to variations in image contrast and illumination, making it particularly useful for applications like object detection, edge detection, and image segmentation. This technique is computationally efficient and straightforward to implement, making it a popular choice in various fields, including medical imaging, computer vision, and remote sensing.

2. **Deep learning method:** Segment Anything Model (SAM) by Meta [7] is used for segmenting the plant from the background. The SAM paper adopts a prompt-based approach to perform image segmentation, aiming to predict a segmentation in the image based on a provided prompt. The prompt specifies what to segment in the image, it can be a point, box or text prompt. The project of building SAM includes three interconnected parts: a promptable segmentation task that produces a mask from prompts specifying what to segment, the task should produce a valid mask even from ambiguous prompts, the segmentation model, SAM, and a data engine for collecting the sizeable dataset.

3.3 Wheat head detection

Transfer learning for wheat head detection involves leveraging pre-trained deep learning models on general image datasets and fine-tuning them using the available wheat field images annotated with bounding boxes. This approach allows for the transfer of knowledge from the pre-trained model to the task of wheat head detection, even when the dataset is limited or lacks diversity. By initializing the model with learned features from a large and diverse dataset, and then fine-tuning it on the wheat field images, the model can adapt to the specific characteristics and variations present in the wheat head images. This process accelerates training and enhances the performance of the detection algorithm, enabling more accurate and efficient identification of wheat heads in agricultural fields.

For wheat head detection, employing YOLO (You Only Look Once) and Faster R-CNN models involves leveraging their architecture and pre-trained weights to facilitate efficient and accurate detection. YOLO, with its single forward pass and real-time processing capability, swiftly detects wheat heads by dividing the image into a grid and predicting bounding boxes and class probabilities directly. On the other hand, Faster R-CNN, with its two-stage approach involving region proposal and classification, offers precise localization and recognition of wheat heads by first generating region proposals and then refining them for accurate detection. Both models are fine-tuned using transfer learning on annotated wheat field images, enhancing their ability to detect wheat heads across diverse agricultural settings.

A brief comparison of YOLO and Faster R-CNN model architecture and performance:

- **YOLO model**:
  - **Speed**: YOLO is designed for real-time object detection and is significantly faster than Faster R-CNN. It processes images in a single pass through the network, making it capable of achieving high frame rates (e.g., 45 frames per second on a high-end GPU).
  - **Architecture**: YOLO divides the image into a grid and predicts bounding boxes and class probabilities directly from the full images in a single evaluation. This unified architecture makes it very efficient.
  - **Accuracy**: While YOLO is faster, its accuracy, especially for detecting small objects or objects that are close together, can be lower compared to Faster R-CNN. YOLO tends to have more localization errors.
  - **Use Cases**: YOLO is ideal for applications requiring real-time processing, such as video surveillance, self-driving cars, and live tracking systems. With the latest improvements, YOLOv8 is highly effective for a broader range of applications, including those that require higher accuracy in addition to real-time performance.

- **Faster R-CNN model**:

  - **Speed**: Faster R-CNN is slower compared to YOLO but faster than its predecessors like R-CNN and Fast R-CNN. It uses a region proposal network (RPN) to generate region proposals, which speeds up the detection compared to previous methods, but it still requires multiple stages.
  - **Architecture**: Faster R-CNN consists of two stages: the first stage proposes regions of interest (RoI) using the RPN, and the second stage classifies these regions and refines their boundaries. This two-stage process tends to be more computationally intensive but also more accurate.
  - **Accuracy**: Faster R-CNN generally offers higher accuracy, particularly for complex scenes with small objects or objects that are densely packed. The region proposal step helps in precisely locating objects and reducing false positives.
  - **Use Cases**: Faster R-CNN is suitable for applications where detection accuracy is more critical than speed, such as detailed image analysis, medical imaging, and automated inspection systems.

3.4 Wheat grain counting

For the estimation of grain count from the detected wheat head, the wheat head's orientation is converted to vertical orientation to create uniformity across all the detected wheat heads which might have different orientation due to wind or movement while capturing the images. A PCA (Principal Component Analysis) based alignment algorithm is implemented on the detected wheat head. The coordinates of the non-zero (white) pixels in the segmented wheat head image are found and PCA is computed on the coordinates of the non-zero pixels. Identified principal components help in finding the main axes of the wheat head and we pick the top two principal components to estimate the centre of wheat head and the orientation angle. Finally, the rotation matrix is computed using the calculated angle and the center of the wheat head. The wheat head is then vertically aligned using the calculated rotation matrix.

A bottom up segmentation technique is applied for the estimation of grain count from the vertically oriented wheat heads. Bottom-up segmentation is a fundamental approach in computer vision and image processing, aiming to partition an image into meaningful regions starting from the pixel level. This method contrasts with top-down approaches, which typically begin with a higher-level representation of the image and then subdivide it into smaller segments.

The process of bottom-up segmentation generally involves several key steps:

- **Pixel-Level Analysis**: The segmentation starts by analyzing individual pixels based on characteristics such as color, intensity, and texture.
- **Feature Extraction**: Features are extracted from each pixel to create a feature space, which may include color values, gradients, and texture descriptors.

- **Similarity Grouping**: Pixels with similar features are grouped together to form small regions. Clustering algorithms like K-means or graph-based methods are commonly used for this purpose.

- **Region Merging**: The initial small regions are progressively merged into larger segments based on similarity measures and spatial proximity. Techniques such as region growing, agglomerative clustering, and graph cuts are often employed.

- **Segmentation Refinement**: The resulting segments are refined to improve boundary accuracy and ensure that the segments are meaningful, often involving morphological operations or edge detection techniques.

Bottom-up segmentation is advantageous for capturing fine details and subtle variations in the image, making it suitable for applications requiring high precision. It relies on the inherent properties of the image data, allowing it to adapt to various types of images and content. However, it can be computationally intensive and sensitive to parameter choices, such as similarity thresholds and clustering criteria. Despite these challenges, bottom-up segmentation remains a powerful technique in computer vision, with broad applicability in fields such as medical imaging, object detection, and image editing.

SAM (Segment Anything Model) [7] developed by Meta with some tuned hyper-parameters is used to segment the wheat into small regions. These regions are treated as similarity groupings in the bottom-up segmentation approach.

SAM's sophisticated design enables it to adapt to new image distributions and tasks without prior knowledge, a capability known as zero-shot transfer. Trained on the extensive SA-1B dataset, which includes over 1 billion masks across 11 million meticulously curated images, SAM demonstrates remarkable zero-shot performance, often exceeding previous fully supervised results in numerous scenarios.

SAM uses a transformer based architecture and tries to predict valid masks for a given input prompt which can be a dot or a box for the object of interest. It also can predict masks all over the image assuming a grid of points across the image as an input prompt and then predicting valid masks for the image.

Meta has developed three different versions of the model with different backbone sizes for SAM to cater to various computational and performance needs. Each model size offers a different trade-off between computational efficiency and segmentation accuracy:

- **SAM Base (SAM-B)**: Optimized for efficiency and lower computational requirements, making it suitable for applications where processing power and memory are limited. It has 91 million parameters and 375 MB size.

- **SAM Large (SAM-L)**: Balances performance and computational demands, offering a middle ground that is ideal for most general-purpose applications. It has 308 million parameters and 1.25 GB size.

- **SAM Huge (SAM-H)**: Provides the highest accuracy and detail, optimized for environments where computational resources are abundant and high precision is critical. It has 636 million parameters and 2.39 GB size.

Chapter 4

Method

Our objective is to analyze and detect and count wheat heads and further grains in the wheat heads as phenotypic traits for the wheat crop from a side-view perspective of rows of wheat plants in both field and indoor settings. The initial aim is to identify an optimal sensor setup that allows for robust extraction of the above phenotypic traits. Then by applying plant segmentation techniques to isolate the area of interest in a visual frame, effectively removing background noise, and utilize that for further analysis. Subsequently, a wheat head detection model is employed to identify and filter out the wheat heads from the frame. Following this, image preprocessing techniques are applied to the identified wheat heads to enhance grain distinction and aid the grain detection model. Finally, a grain detection model is implemented to estimate the number of grains.

There is a big shortage of the labeled dataset for training big deep learning models for wheat plant phenotyping. Hence, a hybrid approach is used combining classical image processing algorithms with fine-tuning of state of the art pre-trained deep learning models on the limited corpus to develop more powerful ensemble approach for the detection and counting models.

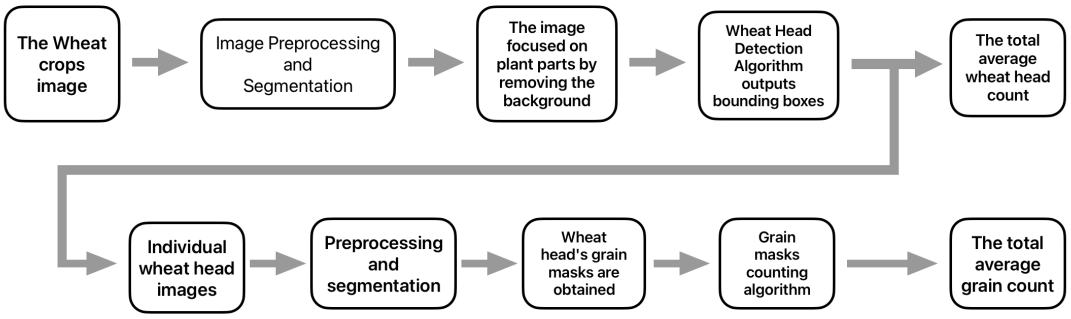


Figure 4.1: The overall architecture of our method

4.1 Image Acquisition

For data acquisition, different camera setups and sensors are considered to extract the required phenotypic information. The GoPro 10, an RGB camera sensor with 24 MP resolution wide-angle image capture and phone camera with 12 MP resolution focused capture are used for image acquisition to aid in object detection and segmentation tasks. Additionally, the LUCID Helios2 model ToF (Time of Flight) camera with a 0.3 MP resolution monochrome sensor and the Intel RealSense D415 stereo camera with a 2 MP resolution RGB sensor are tested to reconstruct 3D plant models. For the extraction of wheat head and grain estimation, the 2D RGB image based segmentation and detection algorithms are utilized. The 2D RGB data is sufficient for our phenotyping application while having the lower computational and cost requirements and faster operation, as well as the higher resolution and greater RGB information compared to 3D imaging methods. Additionally, 3D reconstruction can be prone to errors in outdoor environments. The use of 3D data can be more beneficial for tasks involving the spatial spatial phenotypic traits of different plant parts such as leaf angle, plant height, leaf width, etc.

4.2 Plant segmentation

After getting the plant images using the optimal sensor setup, image processing and segmentation techniques are applied to filter out the plants and remove extra noise and background from an image. Different methods have been tried to segment out the plants from the background. First, Otsu thresholding based segmentation is used to filter out the plant parts and removed the background. Second, using dilation on top of the Otsu segmented plants parts to prevent information leak in the plant image areas and the generated mask is then used to filter out the plant parts from the original image. Third, SAM based image segmentation which segments out the plants from the image and the SAM predicted masks are then used to filter out the region of interest without background from the original image.

Figure 4.2: Segmented wheat plants without the background noise

After removing the background with pre processing steps, the task of object recognition is simplified for the YOLO model. Instead of having to analyze the entire image and differentiate between wheat heads and background elements, the model only needs to focus on detecting wheat heads against a uniform background. This simplification improves the efficiency and accuracy of the detection process. Additionally, the likelihood of false positives can be reduced in the detection results. False positives occur when the model mistakenly identifies background elements as wheat heads. We eliminate this issue without a background.

4.3 Wheat head detection

The scarcity of access to diverse and annotated datasets containing wheat head images, poses a significant obstacle for developing and refining wheat head detection algorithms. For the wheat head detection model training, we have used a Kaggle competition's Global Wheat dataset [1] which consists of field images of wheat heads taken from multiple view angles. The dataset comprises images of wheat fields annotated with bounding boxes delineating each detected wheat head. However, not all images contain wheat heads or corresponding bounding boxes. These images were captured across diverse locations worldwide. Data augmentation techniques like rotation, scaling, cropping, mosaic and some others are applied for creating a diverse dataset and account for uncertainties in the real world implementation for the fine tuning of the YOLO and Faster R-CNN models on the wheat dataset.

Figure 4.3: Mosaic data augmentation method example with labels

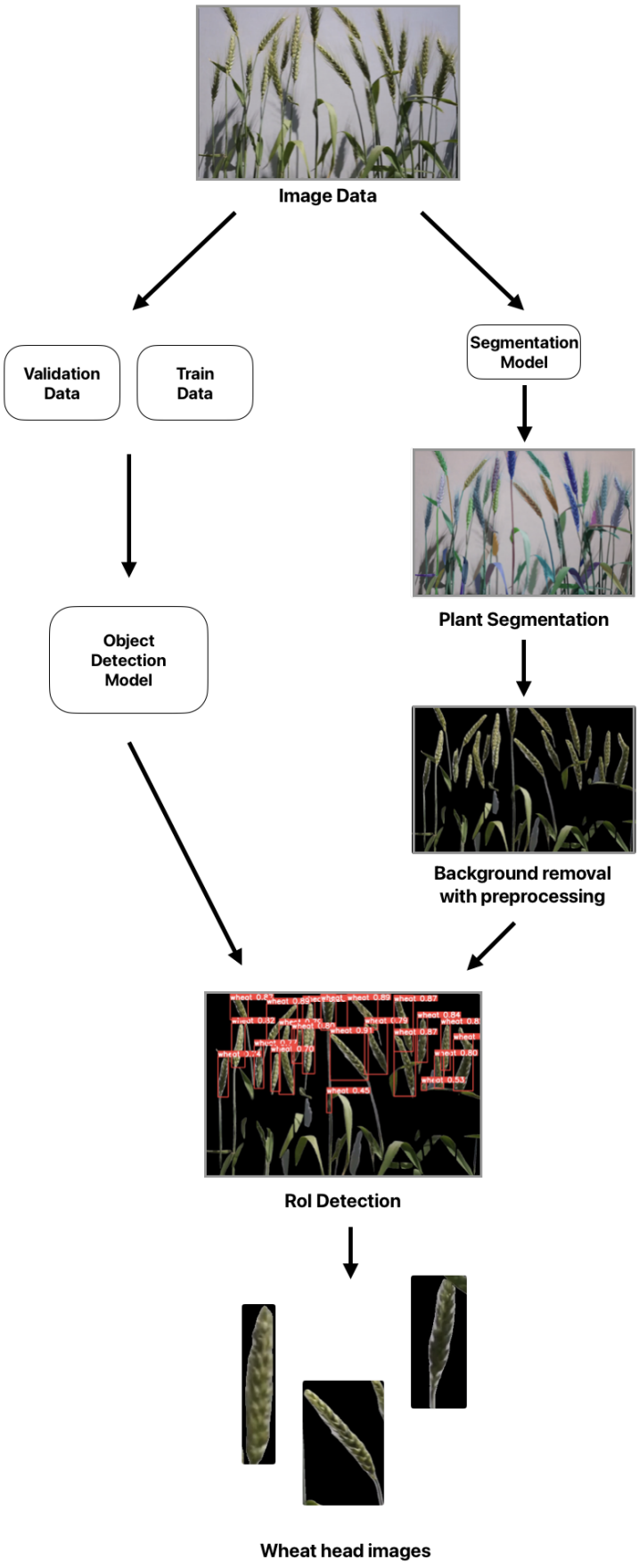


Figure 4.4: Wheat head detection module flowchart

Utilizing the YOLO and Faster R-CNN models, which have been fine-tuned on annotated wheat field images, for predicting on new images involves inference of the images using the developed models and extracting bounding box predictions around detected wheat heads. Subsequently, the number of bounding boxes within each frame is counted, representing the quantity of wheat heads. This streamlined process enables automated detection and counting of wheat heads in wheat field images providing valuable insights for crop assessment and management. After obtaining the bounding boxes from the developed wheat head detection model, the next step involves using these bounding boxes to identify regions of interest (ROI) within the wheat heads. This process entails cropping the original image based on the bounding box coordinates to extract individual regions containing the wheat heads. These extracted ROI areas serve as input for further analysis to count the grains within each detected wheat head. By focusing specifically on these regions of interest, the counting process becomes more precise and efficient, enabling accurate assessment of grain quantities within the wheat heads.

4.4 Wheat grain counting

Counting wheat grains is by itself a complex problem considering the granular level texture variations. Detecting and counting grains in field conditions increases the complexity of the problem which requires intricate analysis. Here, we have explored different grain detection methodologies like developing model using annotated data, classical image processing and region adjacent graphs based segmentation techniques, and utilizing SAM driven bottom up segmentation approach.

Identified wheat head region of interests in the images by the wheat head detection model are used as an input to the wheat grain counting module. Detected individual wheat heads are oriented in vertically upright position using PCA based rotation matrix calculation and using it to rotate the image. This step helps in creating consistency in the images for further analysis since in field conditions wind or any other movement might change the orientation of wheat head.

4.4.1 Wheat grain detection methods

A variety of methods, involving data based grain detection model development, image processing and graph based image segmentation, and using SAM to over-segment and use a bottom up segmentation approach to detect the grains, are explored.

- **Data-driven model development**: Using annotated grain data, which consists of good quality labeled images of wheat heads taken in controlled settings, YOLO model is fine tuned to detect grains. Image augmentation techniques like scaling, cropping and blurring are applied to account for the uncertainties in the field conditions. The trained model has good performance for the wheat head focused images taken in the controlled settings with good image capture. But the model's performance for general field images or images in uncontrolled settings is quite poor and it is not able to detect grains well.
- **Image processing and RAG (Region Adjacency Graph)**: To mitigate the shortcomings of the data driven model, classical image processing methods are tried to segment out

the grains. Combination of image pre-processing steps like corrosion, dilation, histogram equalization and contrast increase, K-means based color segmentation, shape matching, density based segmentation, etc. are evaluated to increase distinctive texture within wheat head for improved grain detection. RAGs based segmentation approach is also evaluated to account for region based grain texture segmentation. These approaches though can get good results for the specific type of images for which they are developed, they still fall short on generalizing on the field images. Thus, the classical approach may be used for getting grains in field settings but not suitable for field conditions with high variability in the image capture.

- **SAM based bottom-up segmentation**: A combination of the state of the art deep learning segmentation based SAM model for generating segmentation masks followed by classical image processing for getting individual grain masks is implemented. Once the SAM model has predicted the wheat grain areas for the wheat head image, we encounter instances where multiple masks may overlay a single grain, necessitating image processing techniques to resolve this challenge. This problem is resolved using the connected-components based approach to merge the masks representing the same grain area. This helps in combining the relevant sub-masks into a a single mask for the same grain structure similar to a bottom-up segmentation approach.

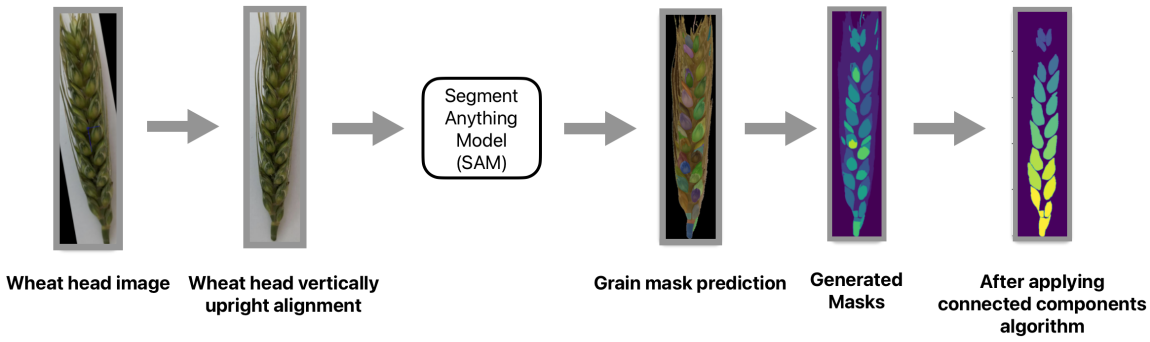


Figure 4.5: Wheat grain detection steps using SAM

SAM may overlook certain grain masks depending on the detected wheat head image quality. Grain estimation accuracy can be further improved by identifying potential regions likely to contain grains from the undetected grain area in the wheat head. A potential grain estimation algorithm is developed to account for missing grains in the undetected grains area of the wheat head.

4.4.2 Potential wheat grain estimation

After using SAM to segment out the grains, we have individual masks for the wheat head and the detected grains. SAM misses out the detection of some of the grains in the wheat head, so we develop a shape matching algorithm using the average wheat grain mask size to look for potential undetected grains in the remaining undetected area of the wheat head. The idea is to utilize the average grain shape obtained from the detected grain masks and then matching it over the wheat head mask without the grain masks.

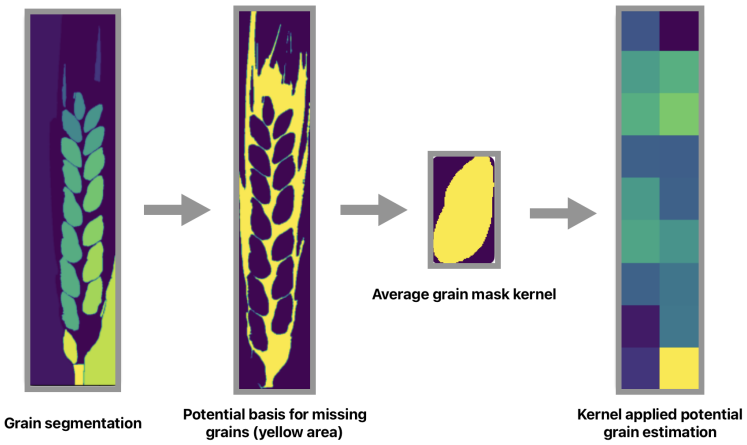


Figure 4.6: Detection algorithm for potential missing grains

Figure 4.6 shows the flow of the algorithm where the first image is of the detected grain masks on a wheat head using SAM. Second image represents the wheat head mask without grain areas which gives the area available for missing grains. Further, the grain mask is used to find the possibility of it being present in the potential grain area. Finally, the last image shows a grid generated for the possible grain locations. As seen in the image, the bottom right cell in the grid shows high possibility of having a grain which was undetected by SAM.

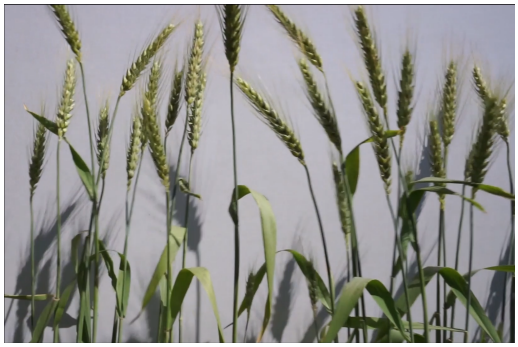
Chapter 5
Results

After training the models on public data, evaluating their performance in both controlled environments and field conditions is imperative to ascertain their practical utility. While initial training provides a foundation, real-world agricultural settings pose unique challenges that may not be fully captured in standardized datasets. Testing the models in controlled environments allows for benchmarking against ground truth data, providing insights into accuracy and efficiency under standardized conditions. However, field testing is essential to assess robustness and adaptability to the complexities of actual agricultural operations, such as varying lighting, terrain, and crop conditions. This evaluation process reveals how well the models generalize across different environments and identifies areas for improvement, guiding iterative refinement to ensure reliable performance in diverse agricultural applications. For the performance evaluation of the wheat head detection model, we have evaluated the model for performance on the raw images as well as pre-processed images from the GWHD dataset, controlled greenhouse environment and field environment. For the grain counting model, apart from good quality public dataset, controlled environment and field images, model is also evaluated across different resolution of the input images. Image resolution in grain counting plays a major role, since the wheat head is detected from an input image as a bounding box and might not have good enough resolution for the counting and segmentation of grains.

5.1 Segmentation of plants from the image

For 2D/3D reconstruction of the wheat plants, we utilize image pre-processing followed by image segmentation algorithms to get the plant parts out of the image and remove all the background and other un-related information. Good segmentation results are obtained using classical Otsu thresholding algorithm and deep learning based Meta's SAM (Segment Anything Model) model. Both of the approaches along with some other algorithms are tested for finding the optimum segmentation results irrespective of the lighting conditions or environ-

mental conditions.



(a) Original image

(b) Segmentation using SAM model

(c) Segmentation using Otsu thresholding

(d) Segmentation using Otsu thresholding with dilation

Figure 5.1: Wheat plants image and the segmentation results using different segmentation methods

Figure 5.1 shows the segmentation results obtained using different algorithms. SAM based segmentation on original image is able to segment out the plant parts quite crisply without extra background noise but occasionally misses some of the objects like wheat plant stem, wheat head and leaves. So even though very good segmentation results are obtained using SAM, it has its limitation of not being to detect all the plant parts consistently. Using Otsu thresholding for segmentation, all the plants parts are segmented out without any missing parts. But Otsu thresholding based image segmentation causes pixel information leakage and is not able to segment out the parts fully and might miss some information near the edges of the plant parts. So to mitigate for the loss of information, Otsu thresholding combined with dilation based approach gives quite good reconstruction of the plant structure.

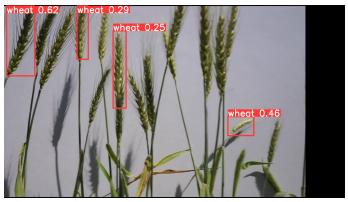
Thus, Otsu thresholding combined with dilation is utilized for plant reconstruction since it gives good enough plant reconstruction without losing information near the edges of the plants.

5.2 Performance of the wheat head detection model

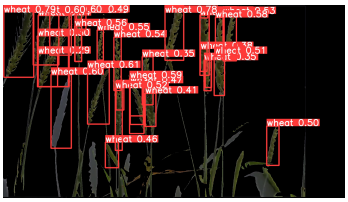
Wheat head model is developed with YOLOv8 model as backbone and then fine-tuned using the Global wheat detection (GWHD) competition dataset [1] consisting of 3605 images for training and 1448 images for validation. Model is evaluated on pre-processed and original images as well as on the GWHD dataset, images from field and controlled environments to evaluate the robustness of the model. For evaluation of the generalization of the developed wheat head detection model, 400 images taken in green house conditions, including movement and blur, are used. Figure 5.2a shows an example image for evaluation.

5.2.1 Performances on original vs pre-processed images

Due to limited availability of annotated data, some image pre-processing and segmentation techniques are utilized to remove the background noise to improve the performance of our wheat head detection model. The images in the training data are from different point of view and thus model performance on the different type of field images than the ones present in the training data.



(a) Detection for the original image



(b) Detection for the segmented image using Otsu thresholding

(c) Detection for the segmented image using Otsu thresholding and dilation

Figure 5.2: Wheat head detection for original vs pre-processed images using the YOLO model

Figure 5.2 shows the wheat head detection model's prediction on different versions of the same view. The first image shows the prediction for the original image and the model is not able to predict all of the wheat heads. In the second image with Otsu thresholding, the same model is able to detect almost all of the wheat heads in the image though some of the shadows of the wheat heads are also being detected as wheat heads occasionally. In the third image with Otsu thresholding and dilation based segmentation, the model's performance has improved further in detection of actual wheat heads. The model is not detecting shadows as wheat heads as it has more information available compared to the one with just Otsu thresholding.

So, Otsu thresholding with dilation helps in improving the performance of the model along with the segmentation results. It gives an optimal balance between background noise removal and region of interest information retention.

5.2.2 Model performance on Global Wheat Dataset

The YOLO model is preferred over the FastRCNN model in wheat head detection due to its high performance accuracy and faster inference rate. The developed YOLO model is evaluated on the GWHD validation dataset to check for the performance on the similar unseen data that the model has been trained on. The model accurately predicted 39,157 wheat head bounding boxes out of 47,624 ground truth bounding boxes in the validation dataset. The model achieved a precision of 0.92 and recall of 0.88 on the validation data.

Figure 5.3: Wheat head detections on GWHD dataset

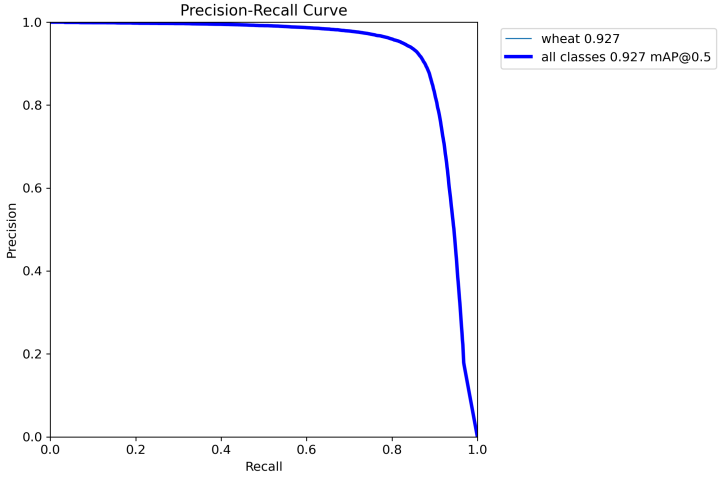


Figure 5.4: Precision-Recall Curve

5.3 Grain counting model performance

Wheat grain counting is a major step to estimate yield of the crop, we are using SAM based segmentation mask generation on detected wheat heads and then applying shape matching from the detected masks. For evaluation of the approach, we evaluate different versions of SAM for generating input masks and varying the resolution of the wheat head image. Further, image processing based grain prediction approach is used to account for the undetected grains from the SAM model to further improve the grain count prediction and evaluate the performance on public data[1], controlled environment [9] and field images of the wheat heads. The test data consists of 150 wheat head images with white background. On the other hand, we used 50 detected wheat head images in green house conditions to further evaluate the generalization in uncontrolled environment.

5.3.1 Grain segmentation with different versions of SAM model

Figure 5.5 shows the segmentation results from different SAM models. The SAM-Huge model outperforms grain segmentation among other variants. This version of the model benefits from a more complex architecture and larger number of parameters as well as it being trained on large dataset with diverse masks, allowing it to capture finer details and make more precise predictions. The improved accuracy in detecting individual grains within a wheat head comes at the cost of more computation power requirements. SAM-Large model also performs good but for fine grain textures lags compared to the huge variant of SAM. For cost optimal system setup, large variant of SAM can be preferred at the expense of decreased accuracy of the grain segmentation robustness. SAM-Small model is good for high level segmentation but does not perform well on the small and detail oriented segmentation tasks. The enhanced capability of the huge SAM model makes it particularly suitable for the grain segmentation task where precision is critical, albeit at the cost of increased computational resources and processing time.

(a) Wheat head image

(b) Segmentation using SAM-Base model

(c) Segmentation using SAM-Large model

(d) Segmentation using SAM-Huge model

Figure 5.5: Grain segmentation using different versions of SAM

5.3.2 Grain segmentation performance evaluation for different image resolutions

SAM-Huge model is used to segment out the grain masks for all the images at different resolutions. Resolution and image quality are key factor in distinguishing grains within a wheat head. We evaluated wheat head images at different resolutions by down sampling the image captured at high resolution. Evaluation is done on high quality image taken in controlled setting as well as image taken in outdoor settings.

- **Controlled environment wheat head image**: For controlled environment setting a focused and good quality image of wheat head is taken up close and then down-sampled by reducing the resolution of the image by half each time. Figure 5.6 shows the effect of decreasing the resolution for the same wheat head image.

(a) Original wheat head image

(b) Segmentation of original image

(c) Segmentation of image with half down-sampled resolution

(d) Segmentation of image with twice half down-sampled resolution

Figure 5.6: SAM-Huge model segmentation results with reduced resolution for the same image

- **Uncontrolled environment wheat head image**: As a second experimental setup, we acquired some images from greenhouse by capturing wheat plant images using a 12 MP camera from approximately 1 m distance. The images were captured at random covering multiple plants at once similar to a field setting. Figure 5.9 shows one such greenhouse image.

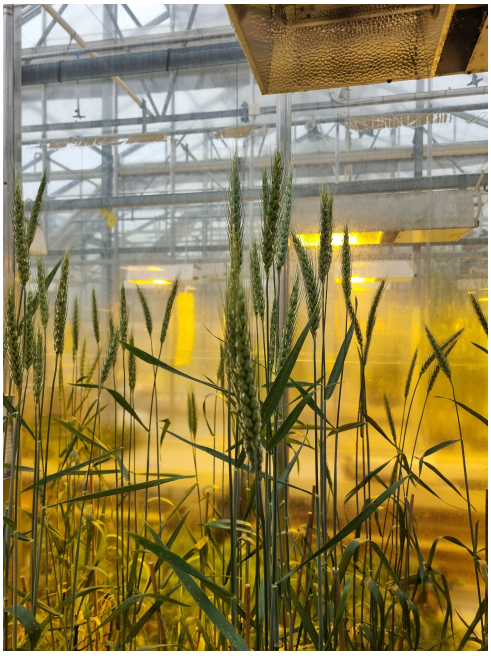
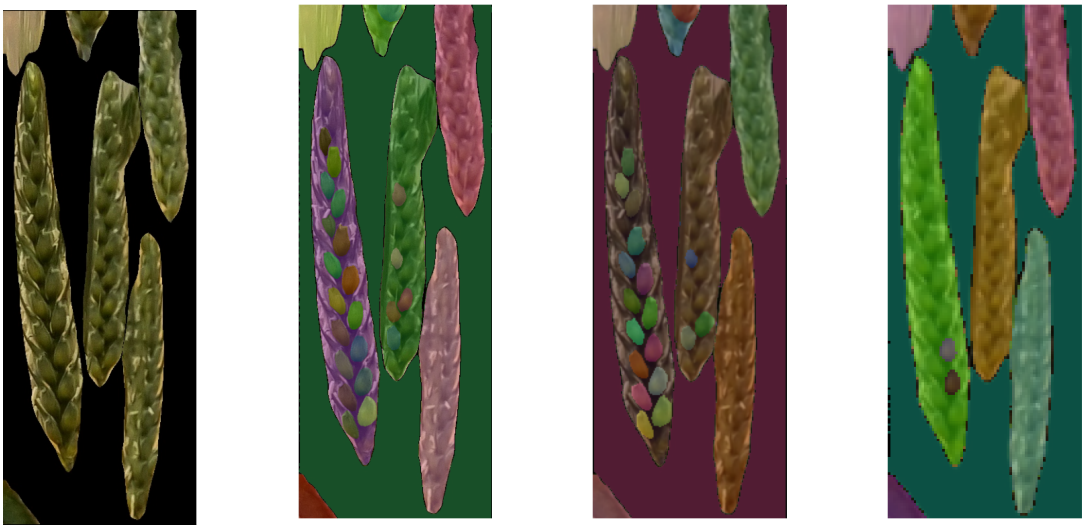


Figure 5.7: Greenhouse wheat plants image

Further, wheat head region of interest identified using wheat head detection model are detected and figure 5.8 shows one of the identified regions and the detected grain masks using SAM-Huge model on original image as well as down-sampled images with lower resolution.



(a) Original wheat head image

(b) Segmentation of original image

(c) Segmentation of image with half down-sampled resolution

(d) Segmentation of image with twice half down-sampled resolution

Figure 5.8: SAM-Huge model segmentation results with down-sampled resolution for the same greenhouse image

5.3.3 Performance analysis and evaluation of wheat head count estimation

We use the YOLO model with Otsu thresholding and dilation steps for background removal from the raw images as our preferred setup for the task of estimating the wheat plant density phenotypic traits of the wheat crop. The setup is evaluated on 255 images acquired in uncontrolled settings. Some of the images have blurred regions, plant shadows and overlap of the plant parts.

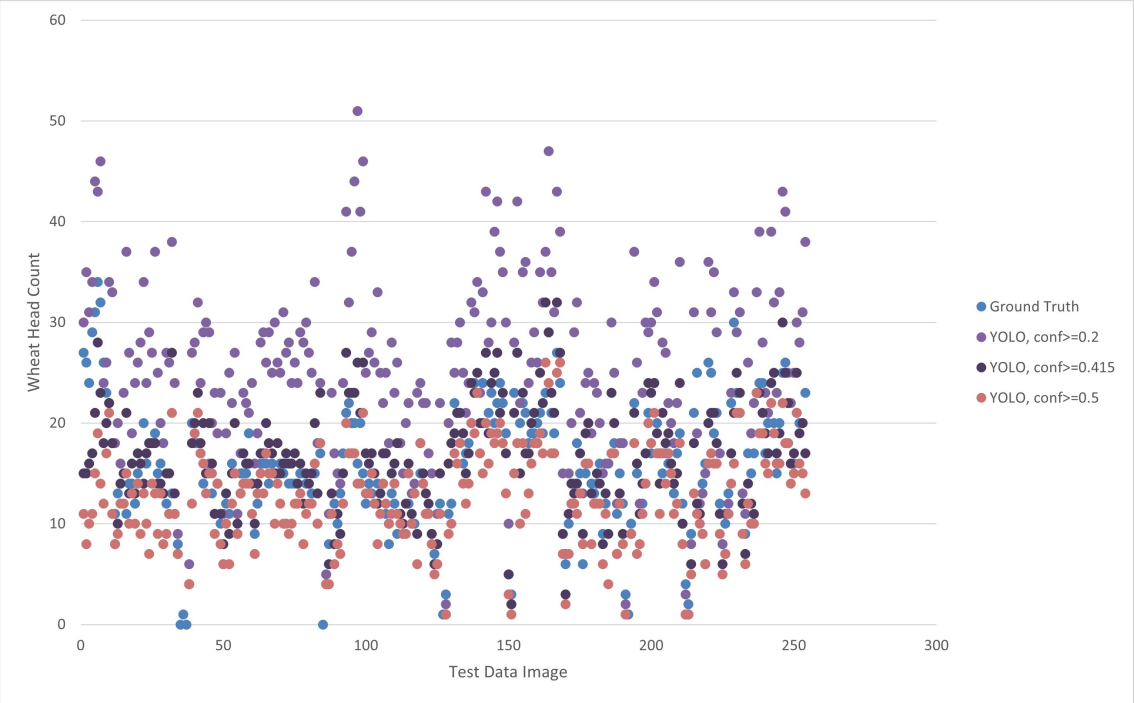


Figure 5.9: Distribution of ground truth wheat head count with model predictions using different confidence cutoffs for individual images.

In figure 5.9, the model predictions with confidence cutoff of 0.415 follow the ground truth more closely as compared to the other cutoffs. Due to uncontrolled conditions, we can see some outliers in the scatter plot. So, 0.415 confidence cutoff is fixed for the detection model to identify a wheat head in the image. We observe the mean absolute error (MAE) of 2.45 wheat head count per frame with respect to ground truth. Furthermore, in comparison to the average ground truth wheat head count of 15.57 per frame, the model achieves average wheat head count of 15.82 .

5.3.4 Performance analysis and evaluation of wheat grain count estimation

SAM model is used to segment out the grains from the wheat head images and then further estimation of potential grains in the remaining undetected wheat head area is done to improve the total grain count for the wheat head. A dataset consisting of 151 wheat head images

[9] in controlled settings is used to evaluate the model setup by counting grains manually for the ground truth.

Model performs well achieving 1.39 MAE for grain count per wheat head using just the SAM based grain segmentation. Moreover, for the 17.97 average number of grains per head in the ground truth, the SAM model based estimation of grain counts reaches 16.65 grains per wheat head.

To further improve the grain estimation, the potential grain areas are identified in the remaining undetected part of the wheat head to account for the grains missed by the SAM model. The grain estimation improved by potential grain area identification approach achieves 17.9 average number of grains per wheat. Thus, the combined estimated grain count gets closer to the expected ground truth. Also, MAE of grain estimation per wheat head is reduced to 1.35 as compared to the 1.39 MAE using only SAM for grain estimation.

Chapter 6

Discussion

Image acquisition

Image acquisition and the sensor quality plays a critical role for the model development and inference. Specifically, for the grain estimation on the detected wheat heads the image quality is critical for the accurate segmentation of the grains. We figured out from our experiments that wheat head image with resolution of at least 0.2 MP and good quality is required to detect grains effectively. Wheat head detection is comparatively an easier problem and thus a sensor setup with resolution good enough for grain estimation would be sufficient for the wheat head detection. The image acquisition setup can capture images with closer view having narrow depth of field and limited image level resolution while ensuring the wheat head level resolution or with wide angle capture with higher resolution for full coverage of plant parts.

Wheat head detection model

The GWHD dataset has a one class wheat head data set that are created by taking images from top view angle. Our problem has a side view angle analysis and we have to train our model on top view due to limitations in wheat datasets. Although with image pre-processing the wheat head detection model is able to achieve good results, model level accuracy and performance can be further improved by collecting and annotating more use case specific training data.

Grain estimation model

The proposed SAM segmentation-potential grain area detection approach achieves significantly good performance in wheat grain estimation. Several studies have focused on this problem using Faster-RCNN, density maps using dot annotation, textural feature tracking with deep learning etc. The current approaches in the agriculture research domain [9] [6] are

able to achieve satisfactory performance on the training data but they can not generalize well on the new data. Our SAM segmentation-potential grain area detection approach is able to outperform them and generalize well on well-captured images without being limited by the training data and could be scaled further for better estimation and generalization.

The algorithm for potential grain area detection can be further improved by considering an average grain mask's shape and accounting for the size and structural pattern of a grain distribution in a wheat head. The precise segmentation of the wheat head can also help in further improvement of the potential grain area identification compared to a softer wheat head segmentation which has excess area along the edges of the wheat head due to the wheat awns which may decrease the accuracy of potential grain estimation.

Wheat phenotype extraction

Using the estimated count of wheat heads per frame, we can get average number of wheat heads per unit length for a fixed setup. This would help in estimating the wheat plant density for agricultural research. Further, having average grain count per wheat head can also aid in estimating the quality of the wheat breeds and evaluating the expected yield of a crop combining plant density with the grains per wheat head.

Chapter 7
Conclusion and Future Work

Accurate estimation of wheat head and spikelets from both controlled and uncontrolled environments wheat crop images is crucial for assessing crop yield traits, yet this task is complicated by the variability, density, and occlusion of plant parts in the real-world scenario. In this study, the developed wheat head detection model is able to generalize well in controlled settings as well as uncontrolled environment. It gives good performance for the estimation of the plant density of the crop. Experiments confirmed YOLO's ability to accurately detect dense and tiny wheat heads, exhibiting excellent adaptability and generalization. This makes the YOLO model a valuable tool for wheat head detection and counting in complex field environments, providing technical support for agricultural wheat phenotype monitoring and yield prediction. The model's counting accuracy was evaluated under challenging conditions such as low illumination, blur, and overlapping occlusion, showing a stronger correlation with manual counting.

SAM segmentation - potential grain area estimation approach to estimate number of spikelets in a wheat head gives significantly improved results with better generalization compared to the other research papers currently in the domain. Proper image acquisition setup is essential for the estimation of the grains from the detected wheat heads in the image since detected bounding boxes occupy a small space of the whole image. So ensuring the image quality at the small wheat head area level can ensure accurate estimation of the grain count.

The presented approach can be extended to further extrapolate the phenotypic trait extraction for other characteristics of the wheat plant or other crops. Better image acquisition setup and creation of annotated dataset can aid in further increasing the reliability and accuracy of the models. Wheat head model could improve with the newly created dataset with real field images. In wheat head recognition, treating all wheat heads in the GWHD dataset as a single category disregards the differences between varieties. To address this, we can employ a multi-task learning approach. This method allows the model to perform both wheat head detection and variety classification simultaneously, thus improving its ability to differentiate between various wheat varieties.

Future research could explore integrating wheat ear recognition models with various sens-

ing modalities, including optical, thermal, infrared, and hyperspectral images. Such multimodal fusion aims to enhance the accuracy and robustness of wheat-ear recognition by capturing a comprehensive representation of wheat ears. They enable visualising extra details that are unseen by RGB CMOS camera sensors. As a result, eliminating the disadvantages of light illumations or decreasing the need for high image resolution. This would enable the extraction of other phenotypic traits of the crops such as plant diseases, leaf angle, leaf length by using spatial data with LIDAR or stereo camera.

References

[1] Global wheat dataset. https://www.global-wheat.com/gwhd.html.

[2] Narendra Ahuja and Sinisa Todorovic. Extracting texels in 2.1d natural textures. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, Oct 2007.

[3] Sébastien Dandrifosse, Arnaud Bouvry, Vincent Leemans, Benjamin Dumont, and Benoît Mercatoris. Imaging wheat canopy through stereo vision: Overcoming the challenges of the laboratory to field transition for morphological features extraction. *Frontiers in Plant Science*, 11, 2020.

[4] Irsa Ejaz, Wei Li, Muhammad Asad Naseer, Yebei Li, Weilong Qin, Muhammad Farooq, Fei Li, Shoubing Huang, Yinghua Zhang, Zhimin Wang, Zhencai Sun, and Kang Yu. Detection of combined frost and drought stress in wheat using hyperspectral and chlorophyll fluorescence imaging. *Environmental Technology Innovation*, 30:103051, 2023.

[5] Hongyu Guo. Wheat head counting by estimating a density map with convolutional neural networks, 2023.

[6] Faina Khoroshevsky, Stanislav Khoroshevsky, and Aharon Bar-Hillel. Parts-per-object count in agricultural images: Solving phenotyping problems via a single deep neural network. *Remote Sensing*, 13(13), 2021.

[7] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.

[8] Zhenyu Ma, Rakiba Rayhana, Ke Feng, Zheng Liu, Gaozhi Xiao, Yuefeng Ruan, and Jatinder S. Sangha. A review on sensing technologies for high-throughput plant phenotyping. *IEEE Open Journal of Instrumentation and Measurement*, 1:1–21, 2022.

[9] Ruicheng Qiu, Yong He, and Man Zhang. Automatic detection and counting of wheat spikelet using semi-automatic labeling and deep learning. *Frontiers in Plant Science*, 13, 2022.

[10] Lei Shi, Jiayue Sun, Yuanbo Dang, Shaoqi Zhang, Xiaoyun Sun, Lei Xi, and Jian Wang. Yolov5s-t: A lightweight small object detection method for wheat spikelet counting. *Agriculture*, 13(4), 2023.

[11] Sheng Wu, Weiliang Wen, Wenbo Gou, Xianju Lu, Wenqi Zhang, Chenxi Zheng, Zhiwei Xiang, Liping Chen, and Xinyu Guo. A miniaturized phenotyping platform for individual plants using multi-view stereo 3d reconstruction. *Frontiers in Plant Science*, 13, 2022.

[12] Haipeng Xiong, Zhiguo Cao, Hao Lu, Simon Madec, Liang Liu, and Chunhua Shen. Tasselnetv2: in-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods*, 15(1):150, Dec 2019.

[13] Radek Zenkl, Radu Timofte, Norbert Kirchgessner, Lukas Roth, Andreas Hund, Luc Van Gool, Achim Walter, and Helge Aasen. Outdoor plant segmentation with deep learning for high-throughput field phenotyping on a diverse wheat dataset. *Frontiers in Plant Science*, 12, 2022.

[14] Hu Zhou, Andrew B. Riche, Malcolm J. Hawkesford, William R. Whalley, Brian S. Atkinson, Craig J. Sturrock, and Sacha J. Mooney. Determination of wheat spike and spikelet architecture and grain traits using x-ray computed tomography imaging. *Plant Methods*, 17(1):26, Mar 2021.

Master's Theses in Mathematical Sciences 2024:E55
ISSN 1404-6342

LUTFMA-3550-2024

Mathematics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/