The first step of this research focuses on acquiring essential environmental data, including aerial imagery and the corresponding global coordinates of the camera at the time of image capture. This data is critical for converting local pixel positions in the images into global coordinates, a process necessary for accurate spatial analysis. Several camera systems, such as the Sequoia, RedEdge, Micasense, and DJI Zenmuse X7, were considered based on their technical specifications and the resources available for this study. After careful evaluation, the GoPro Hero 4 was chosen due to its ability to capture high-resolution RGB images at a high frame rate. Although this camera lacks near-infrared (NIR) and thermal imaging capabilities, its RGB imagery is well-suited to the image processing algorithms employed in this research. In addition to aerial images, global positioning data is acquired using a Ublox-Neo-6m GPS module, which transmits the camera’s location to the processing unit. To further streamline the system, the camera’s altitude is predefined, rather than measured and transmitted in real-time, allowing for a more straightforward setup without compromising the precision needed for the study.

In response to the need for live waypoint generation in agricultural fields, a hardware setup is required to provide the necessary environmental information. This system has the role of capturing and transmitting data to the software unit, where the processing will take place. In the context of this project, a lightweight camera is crucial to take high-resolution aerial images of the target field which will be the main data source for the image processing algorithm. Several cameras including Sequoia b9-b4-b7-b6, Micasense RedEdge b9-b14-, and DJI Zenmuse X7 b5 were utilized in similar projects. Considering the specifications of these cameras and crop row detection algorithms developed based on their aerial images, it was concluded that having high-resolution RGB images meets the needs of this project. Although the NIR band contains valuable information about the field and can be used for vegetation segmentation, they are too sensitive to environmental conditions such as temperature and might lead to the poor performance of the software. b-5. In conclusion, the Gopro HERO4 camera was acquired for this project. The specifications of this camera including its lightweight, high-quality RGB images, and durability make it a suitable choice for aerial imagery.

Besides the camera, a radio transmission system consisting of power supplies, a transmitter, and a receiver was developed to provide live data transfer between the mounted camera on the UAV and the computer with an image processing program.

Also, the last step of this project, which is the conversion of pixel coordination on the images to global coordination, requires the global coordination of the camera at the time of photo capturing as a reference. A Ublox-Neo-6m GPS module must be mounted on the camera UAV to provide this information.

This section focuses on the crucial task of crop detection within the captured images, which is essential for identifying crop rows and determining the path for subsequent operations. The process begins with a preprocessing step that converts the raw images into a format compatible with the models used for detection. Two crop detection methods, K-means and Unet, are then implemented, with a detailed explanation of each approach. The performance of these models is compared to select the most effective one for this application. Additionally, insights from previous studies using these algorithms inform the analysis.

The main contribution of this paper lies within this section, to develop an image-processing pipeline to determine the position of pixels which are located in the path between the crops. The Figure shows the procedure proposed in this paper. Initially, the aerial images captured with the camera are split into sub-images, and then during a semantic segmentation process, a binary mask is generated for each sub-image showing the positions of vegetation and background. In the next step, using Hough Transform, crop rows are detected. After defining the path line, a desired number of equally spaced points are selected on this line, and their position is recorded for later steps. Finally, the sub-images and their local waypoint coordinates are reconstructed and form the original input image.

High-resolution aerial images covering large areas are typically too large to use in machine learning algorithms directly. To address this, the images are initially split into equal-sized sections, making them manageable for algorithmic processing. The size of these sections is determined by the specific requirements of the algorithms used, ensuring compatibility and optimized performance. Maintaining uniform image sizes across all sections enhances processing speed and efficiency.

High-resolution aerial images covering large areas are typically too large to be used in the image processing algorithms and networks directly. Besides, due to the variation in terrain topologies, the non-uniform shape of crop rows -including curved or irregular shapes- is common in some agricultural fields.

Given this information, the calculation of a straight line for an entire crop row in the original image is not an accurate approach for crop row detection.

To address this issue, the aerial image is initially split into equal-sized sections with a static image size of 512\*512 pixels. These sub-images are named accordingly to be reconstructed after the image-processing procedure is finished and their chosen size ensures compatibility with the algorithms and networks designed in the later steps of the image-processing section.

This project's most computationally intensive aspect is crop detection, which is crucial in ensuring accurate outcomes for subsequent procedures. A well-defined layer mask for field images is essential, as the crop detection system must effectively distinguish plants and crops from other objects, including background soil and unwanted vegetation such as grass and weeds.

Semantic segmentation is the most computationally intensive section of this project, while its outcome has a direct impact on the overall accuracy of the system. Given the RGB sub-images, it is important to identify the exact locations where crops exist, which is done by processing the image and generation of a binary mask. In the following section, two semantic segmentation methods are introduced, and implemented and their results are assessed.

By analyzing aerial images of crop fields, the idea emerged of applying color filtering to isolate areas within a specific range of green, corresponding to vegetation. Given that the input images are in RGB format, the initial approach involved using only the green channel to filter out pixels based on their green intensity. However, this method proved unsatisfactory for several reasons. First, white pixels containing high green intensity could not be effectively filtered using just the green channel. Additionally, areas with a more yellowish hue were incorrectly excluded by this filter. To address these issues, the RGB images were converted to the HSV (Hue, Saturation, and Value) format, which allows for more precise color filtering based on hue. The vegetation color range was then defined and applied to the images, resulting in improved segmentation of crops from the soil. The output of this process was a binary black-and-white mask, where white pixels represented areas containing plants.

The initial idea for semantic segmentation of crops in agricultural fields was to apply color filtering methods to isolate the areas having a green color. This method was first applied by analyzing the green channel of the image but performed poorly due to a few reasons. First, the green channel filtering method is not able to filter white pixels, since they also include a large amount of green color. Second, there are sections of crops with color tending to yellow, and they were unintentionally removed from the layer mask. The two issues mentioned were later resolved using a prefiltering for white pixels removal and also by converting the RGB image to HSV format (Hue, Saturation, and Value) so that it would be possible to filter the pixels based on their color considering their Vegetation Index (VI).

Next, to further highlight the areas including crops, a K-means clustering method was applied to the color-based filtered image. The clustering helped in removing sparse pixels identified as crops and uniting the areas where crops exist. The Parameters of the K-means algorithm were defined in an iterative approach, to identify the best-performing setting,

Without a high computational cost, this method can generate an accurate binary mask indicating the areas where vegetation exists. However, examining this method on the images taken from different fields with a variation of crops, it was observed that the method fails when there exists unwanted vegetation like grass and weeds in the field. Since the model filters the pixels based on their color, it cannot isolate crops without including the other vegetation. The same results in b5. FIGURE shows the performance of this method on different sub-images.

Although the color filtering method demonstrated satisfactory performance, it struggled to distinguish crops from other vegetation, such as grass or weeds, present in the field. This limitation arises because the algorithm passes all vegetation through the color filter, disrupting crop detection. These challenges highlight the inadequacy of color filtering for this task and underscore the need to utilize machine-learning models that intelligently differentiate crops from other vegetation.

The limited performance of color-based image segmentation in crop fields indicates the need for an intelligent model that can recognize the crops using more complicated methods. Utilization of machine learning methods to train semantic segmentation models has proven to be a reliable solution for crop detection. In this approach, a Unet network was chosen as the main image segmentation tool and was trained on a dataset developed by b5. This dataset consists of UAV-based aerial images taken from three vineyards with corresponding ground truth images, indicating the location of vines in the vineyard.

To use the annotated aerial images, the original and ground truth images were initially split into sub-images, using a sliding window of size 512 by 512 pixels. The step size of sliding was chosen to be 50 pixels, which resulted in more sample data for training. The resulting dataset consists of 5089 sub-images, which are divided into train, test, and validation sets with the ratio of 80%,10%,10%.

The Unet model was trained in 50 epochs, which reached the early stopping condition at epoch 7 with an accuracy of 96% and a loss value of 0.09. the predictions of this model on test data prove its ability in semantic segmentation of crops in different conditions, even at the present of unwanted vegetations.

Access to a suitable dataset is essential to train a machine learning model. For this purpose, the vineyard aerial images dataset [REF] was employed, which includes annotated layer masks indicating the positions of plants within the images. The images were initially split into equal-sized segments of 128 by 128 pixels to ensure consistency in the input to the model, eliminating the need for resizing during processing. Subsequently, a Unet model was defined and compiled based on the architecture outlined in [REF]. The model achieved a final accuracy of 96\%\, which was obtained after seven epochs, with early stopping criteria applied. This result confirms that the Unet model can effectively capture image features and accurately segment the crop areas.

Crop row detection involves assigning a line with a defined slope and intercept to each crop row. Based on the binary mask of segmented crops obtained in the previous step, lines must be fit to the pixels to minimize the least square error from the line. This study analyzed and implemented two approaches—linear regression and the Hough Transform—for this purpose.

The crop row detection process aims to assign a straight line with a defined slope and intercept to each group of white pixels annotated as crops in the binary mask. This study analyzed and implemented two approaches—linear regression and the Hough Transform—for this purpose.

This experiment used the binary layer mask of segmented crops to assign each pixel to its corresponding crop row. The first step involved determining the optimal angle of the crop rows through an iterative algorithm, assuming that all rows are parallel and share the same slope. Once the slope was defined, the x and y coordinates of the white pixels were rotated to align the crop rows vertically, simplifying the clustering process.

The idea for implementing linear regression in crop row detection is based on the similarity of this problem to line fitting to data points. Given the white pixels related to each crop row, it is possible to find a line passing through each group of white pixels that has the least square distance from all the pixels. Fitting this line, its slope and intercept can be stored for further steps.

To implement this method,

At this stage, the K-means algorithm was applied with a customized loss function designed to minimize the least square error of the distances between the white pixels and a vertical line. Unlike standard loss functions that measure the distance from a centroid point, this approach considered each cluster centered around a line, not a point.

Given the variability in the number of clusters across different images, the clustering algorithm is initiated by grouping nearby pixels. When a pixel was too distant from the existing group, it was treated as the center of a new cluster. However, this approach led to the formation of numerous unwanted clusters. To address this, adjacent clusters were merged in a subsequent step. Finally, the initial rotation of the pixel coordinates was reversed to visualize the resulting clusters. As illustrated in the results, even after merging close clusters, the method failed to accurately identify crop rows, leading to complications in the later stages of the project. To conclude the experiment, linear regression was applied to fit a line to the pixels within each cluster. However, the results indicated that this clustering and linear regression approach is ineffective for defining crop rows, necessitating exploring alternative methods.

A widely used solution for crop row detection is the Hough Transform. This computer vision technique is primarily designed to identify geometric shapes in images, making it particularly effective for detecting lines and curves. In this study, the Hough Transform was employed as an alternative to the linear regression model for predicting crop rows. Implementing this method requires significantly less preprocessing and image manipulation, mainly due to the availability of related packages in the OpenCV Python package. Applying the Hough Transform to the images, multiple lines are detected for each crop row, as illustrated in the results. It is observed that these lines successfully cover the crop rows, indicating the method's effectiveness in line detection. To refine the results and assign a single line to each crop row, the detected lines were merged using their average slope and intercept. The final results are presented in FIGURE.

Defining waypoints becomes straightforward once the crop rows have been established. Initially, a path is defined as a line parallel to the crop rows and equidistant from two neighboring crop rows. A predefined number of equally spaced points are selected along each path line. The coordinates of these points are recorded for use in subsequent steps. FIGURE

As stated in previous steps, the most computational parts of this study are image segmentation and crop row detection. The determination of crop rows as lines with defined slope and intercept makes it straightforward to calculate a line indicating the path. The path is defined as a line parallel to two neighboring crop rows at an equal distance from each.

Waypoints can be defined by choosing a specific number of equally spaced points selected along each path. The local coordination of these waypoints is stored to be used in the later steps of the study.

Following the image splitting described in section B.1, the image processing tasks outlined earlier will be applied to each segment of the leading aerial image. The goal is to determine and record the waypoint coordinates within each segment. Once the waypoint coordinates are computed, the initial image must be reconstructed and assembled. Subsequently, the waypoint coordinates need to be converted to align with the coordinates of the newly assembled image.