ORIGINAL ARTICLE

QuanCro: a novel framework for quantification of corn crops' consistency under natural field conditions

Fatimah Islam[1] · Muhib Ullah[1] ⬥ · Abdul Bais[1]

Abstract

Crop population and inter-plant spacing in corn farms can provide useful insight into plant phenotypic analysis and informed establishment decisions, improving crop productivity. Traditionally, farmers have relied on manual inspection to assess crop consistency, such as counting plant stands and manually estimating plant spacing. This assessment is carried out in small predefined areas, leading to insufficient crop consistency analysis on the entire field. Moreover, alternative computer vision techniques computing only one or two key parameters also prove insufficient for accurate crop consistency assessment. This research presents a framework called QuanCro that utilizes red–green–blue images from field machinery to analyze crop parameters such as plant stands counting, plant emergence rate, and plant spacing. It utilizes the state-of-the-art object detection network—You Only Look Once version 7 (YOLOv7), to locate and count corn plants combined with our proposed semantic segmentation model, Small Pyramid-UNet (SP-UNet) architecture, to determine leaf area index. This architecture is designed to be memory efficient and computationally less expensive than similar networks, such as HRNet_Mscale (72.1M) and SegNet (34.65M), as it has approximately 21M parameters. The SP-UNet is further integrated with the Zhang–Suen thinning technique and progressive probabilistic Hough transform for crop row detection and plant spacing information. QuanCro accurately estimates crop densities and identifies inconsistent crop areas. The method is tested using 8000 images and shows a mean average precision of 0.976 for identifying plant stands. The SP-UNet achieves intersection over union scores of 0.973, 0.924, and 0.926 for crops, rows, and backgrounds, respectively.

Keywords Plant spacing statistics · Crop row detection · Progressive probabilistic Hough transform · Perspective transformation

1 Introduction

Corn is a vital crop in North America that requires constant monitoring to ensure good production. Analyzing plant density is crucial in assessing crop consistency and providing valuable information for forecasting crop productivity [1]. However, to fully understand the consistency of crops across large fields, it is important to consider other parameters, such as inter-plant distances, percentage of crop coverage, and population of plants per row. This comprehensive analysis offers valuable insights for farmers and agronomists to implement variable rate technologies, enabling efficient management and control of crop growth [2].

The traditional method of assessing crop consistency involves manually counting crops and finding inter-plant distance on predefined sample areas in the field. This limits the view of the crops' status and becomes laborious and expensive when applied on a large scale. Developing automated techniques for crop consistency analysis can overcome these problems and enable farmers to make informed decisions about crop establishment and early intervention, leading to optimized crop production and reduced wastage. These techniques can continuously

Fatimah Islam, Muhib Ullah, and Abdul Bais have contributed equally to this work.

✉ Muhib Ullah
muo145@uregina.ca

Fatimah Islam
fib236@uregina.ca

Abdul Bais
abdul.bais@uregina.ca

[1] Faculty of Engineering and Applied Science, University of Regina, 3737 Wascana Pkwy, Regina, SK S4S 0A2, Canada

monitor crops and analyze various characteristics, such as size, shape, and color. Several agricultural applications use computer vision (CV) techniques, such as convolutional neural networks (CNNs), which can count corn plants [3], classify leaf diseases [4, 5], segment leaves [6], and estimate weed density [7]. Deeper convolutional feature representations enhance the capability of these models to extract semantic information, resulting in better generalization in diverse crop and weed conditions [8, 9]. Various architectures have been proposed to obtain plant densities and estimate their distribution across a corn field [10–12].

Kitano et al. use the deep CNN architecture UNet to segment corn from the background and a blob detection methodology (based on CV) to quantify the number of plants in the image [10]. However, this method lacks to provide precise locations for the detection. For on-ear corn kernel counting, Khaki et al. utilize truncated VGG-16 for feature extraction in their proposed DeepCorn framework, reporting a mean absolute error (MAE) and root mean squared error (RMSE) of 41.36 and 60.27, respectively [13]. However, this method performs poorly on images from different fields and uses more trainable parameters (26.61M), making it inadequate for large-scale data handling. Mota et al. use CNNs for detecting and counting corn plants in the presence of weeds [14]. They base their research on YOLOv4 [15] and YOLOv5 [16] architectures. Similarly, Guo et al. introduce a two-step machine learning-based image processing technique to detect and count sorghum heads using high-resolution images from unmanned aerial vehicles (UAVs) [17]. However, the robustness of these methods in different field conditions is undetermined, and assessing only plant stands counting is not sufficient to accurately evaluate the consistency of crops on large fields. Therefore, there is a need to develop an algorithm that considers multiple parameters to analyze crop consistency efficiently and can adapt to natural field conditions.

The identification of crop rows is a crucial factor in the analysis of crop consistency as it provides valuable data on crop emergence and inter-plant spacing. This information is used to monitor crop growth per row, which plays a significant role in identifying potential issues in crop development. Several CV techniques are developed to detect straight crop rows in an image, including the conventional Hough transform, linear regression, and green pixel accumulation [18–20]. The Hough transform is a popular technique that transforms the image into Hough space, detecting lines and geometric shapes in the original image. Linear regression fits a straight line to a set of points in the image, while green pixel accumulation counts the number of green pixels in a region of the image to infer the presence of straight crop lines. However, these techniques can misclassify weeds as crops, leading to false row detection [21]. They rely on certain assumptions about the characteristics of crops, such as their shape, size, and color, and are sensitive to variations in lighting and other environmental conditions. To overcome these limitations, fully convolutional networks (FCNs)-based semantic segmentation architectures have been introduced for row detection and crop coverage estimation tasks [22, 23]. These models can identify the pixels in an image corresponding to crops and weeds and assign each pixel to the appropriate class [23, 24].

Silva et al. introduce a method for crop row extraction using a UNet-based CNN model that performs semantic segmentation prior to post-processing steps [25]. They report an accuracy of 89.36%, which is lowered by 4.93% under difficult field conditions, such as high weed pressure and sparse crop content in misaligned crop rows. On the other hand, Tang et al. develop a sequence of image processing procedures to identify and estimate the distance between corn plants in crop row images [26]. Their algorithm utilizes various crop information, such as plant color, morphological features, and the center line of the crop rows. They report a spacing RMSE error of 1.7 cm and a coefficient of determination (R^2) of 0.96 cm. This method provides better results in estimating the spacing between crops during the initial growth stages, but the error increases in later stages and when the crops are damaged. To date, no studies have been found that employ FCN architectures to evaluate consistency based on plant spacing estimation, crop coverage area, and density per row.

We propose a Small Pyramid-UNet (SP-UNet) network that uses FCN to segment crop leaves and rows. Like UNet [27], our network has an encoder–decoder structure, with ResNet34 [28] as the backbone. In the decoder section, we have incorporated the feature pyramid network (FPN) architecture to extract rich, multi-scale features. By adding more context and detail, this feature extraction helps to improve the model's performance. Additionally, our architecture aims to reduce the number of parameters, which helps to prevent overfitting and results in a more lightweight network. Despite having only 21.85M trainable parameters, our model has comparable prediction performance to more complex segmentation models like HRNet_Mscale (72.1M) and SegNet (34.65M). We observe that RNet_Mscale and SegNet exhibit poor performance in row extraction, especially in scenarios where plants within a row are sparsely distributed, which is frequently seen in corn crops. We utilize the Zhang–Suen thinning algorithm [29] to reduce the segmented row to a single-pixel structure and integrate it with the progressive probabilistic Hough transform (PPHT) [30], which assigns a line to the center of each row.

After the crop segmentation task, we estimate crop coverage through leaf area index (LAI) calculation. This task identifies extracted pixels in the image corresponding to leaves and measures the total leaf area. While many studies support semantic segmentation models for estimating LAI, they do not establish a correlation with other key elements that contribute to the analysis of crop consistency [31]. For example, in [32], the authors use the UNet architecture to estimate LAI but do not present other consistency parameters such as crop row detection, plant stand count per row, and inter-row and inter-plant distances.

We use the latest object detector, YOLOv7 [33], to accurately determine corn stand count. To avoid redundant results, we apply the Distance-IOU Non-Maximum Suppression (DIoU-NMS) technique [34]. This method considers both the detection boxes' overlapping areas and center points, enabling better detection of target plants that may be obstructed. Our algorithm evaluates the number of plants per row, their spacing, and LAI to quantify consistency, as shown in Fig. 1. Our framework is effective for crops under biotic stress and at progressive growth stages. We believe that no other existing framework in the literature can accurately assess multiple crop parameters to quantify crop consistency.

This research paper presents three main contributions as follows:

- A novel architecture called SP-UNet is proposed for segmenting crops and crop rows in the RGB images captured by field machinery-mounted cameras.
- Implementation of the novel method for quantifying corn crop consistency under natural field conditions.
- Establishing a correlation between inter-plant distance, plant stand count per row, and crop coverage percentage for crop consistency analysis.

The notations utilized in this work are summarized in Table 1.

2 Materials and methods

In this section, we present detailed information on the dataset and the QuanCro framework, which is utilized to evaluate the consistency of corn crops.

2.1 Data collection and its challenges

A dataset of 8000 images of corn fields at various growth stages is being utilized for this study. These images are taken using SWATCAM, a system based on an Android phone, mounted on field machinery as depicted in Fig. 2. The images measure 1440×1080 pixels and are taken at intervals of 15.24–21.33m with a count of 17–20 images per acre. The dataset presents a range of difficulties,
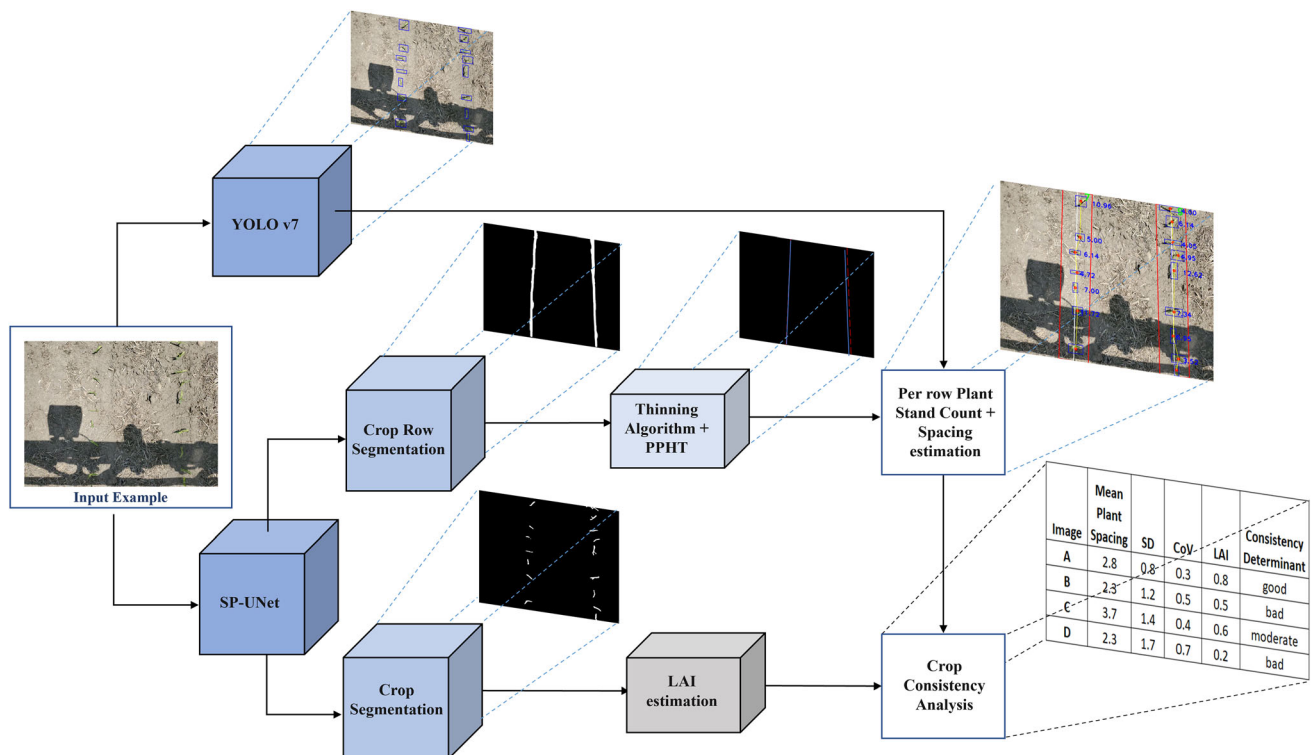

Fig. 1 Workflow of the proposed algorithm

Table 1 Summary of notations

Symbols	Definition
x	Width coordinate of bounding box
y	Height coordinate of bounding box
$\mathcal{D}$	Highest score of detection box
s_i	Classification score
ε	NMS threshold
R_{DIoU}	Distance-IoU penalty term
P_i	Lower score of detection box
P_n	Number of pyramids
θ	Angle between the optical projection axis
φ	Angle between the horizon and optical axis
μ	Camera's angular aperture
$k \times g$	Image resolution
ϑ	Projection weighting factors
ρ	Projection weighting factors
D	Shortest distance in pixels
m	Gradient
M_{crop}	Pixels corresponding to crops
N_{all}	Pixels corresponding to crops/background/weeds
Pr	Predicted segmentation mask
Gr	Ground-truth segmentation mask
cL	Object class
R^2	Coefficient of determination

Fig. 2 SWATCAM cameras mounted on each sprayer boom of the field machinery (https://swatmaps.com)

including low light or nighttime shots, motion blur, tilted angles, and varying boom heights. Additionally, there are instances of mid-stage corn plants and weeds overlapping, which can create further complications. It is crucial to carefully select the data used to train the model to ensure accurate detection of corn plants in diverse conditions. Some examples from the dataset are depicted in Fig. 3. The images in Fig. 3a show early-stage crops surrounded by weeds. Figure 3b exhibits shots of crops under different lighting conditions, with the crops overshadowed by field machinery in the left image. In contrast, the right image is captured under an auxiliary lighting source. Figure 3c displays images with mid-stage crop in an overlapping state.

2.1.1 Data augmentation

Data augmentation increases the data by generating new data points from the existing data. This is achieved by implementing reasonable modifications to the original images, which are then added to the training dataset as new augmented copies of the images. The semantic segmentation dataset is modified through various geometric and photometric augmentations, such as adjusting the saturation and hue of the image's color channels and applying random scaling, translation, rotation, and cropping. In addition, for plant stands detection, the dataset is augmented by applying CutOut [35], CutMix [36], MixUp [37], and Mosaic [15] augmentation techniques. These augmentation techniques involve applying modifications to multiple images at once. In the MixUp technique, a pair of images are superimposed onto each other through a weighting operation to create a new sample for which a new corresponding label is also generated. The CutOut method slices away a part of images that the model may rely heavily on during training. This gives the model a platform to learn all the elements in the images thoroughly. CutMix substitutes a slice of another image onto the small patch of zero-pixel areas generated by CutOut. CutOut and CutMix are both predecessors to the Mosaic augmentation method. This method combines images of different sizes into mosaic patterns and corresponding annotations. Subsequently, a random portion of this entire grid is selected and cropped to create a new augmented image, which is then utilized for training the network.

2.1.2 Data annotation

This experiment uses two different labeling tools to prepare the dataset. Bounding boxes for each corn plant are generated using the LabelImg labeling tool, and the rows and plants are segmented using the Segments.ai platform (https://segments.ai). An example image for labeling corn plants using bounding boxes is shown in Fig. 4a. Only pixels belonging to the plants are annotated for semantic segmentation, while the crop rows are segmented using a polygon tool, as illustrated in Fig. 4b. The row extraction takes account of the plants and the sparse spacing between each plant.

(a) Crops under high weed pressure

(b) Crops under different lighting conditions

(c) Overlapping of the crops in mid-stage

Fig. 3 Illustration of various types of challenges in the dataset

2.2 Methodology

In this research paper, a novel approach called QuanCro is presented, which leverages a semantic segmentation model to identify crop rows and estimate the percentage of crop leaves. Additionally, an object detection algorithm is employed to locate and count individual plant stands. The segmentation model is further integrated with the Zhang–Suen thinning algorithm and the PPHT to identify crop rows accurately. Using these parameters, the plant density
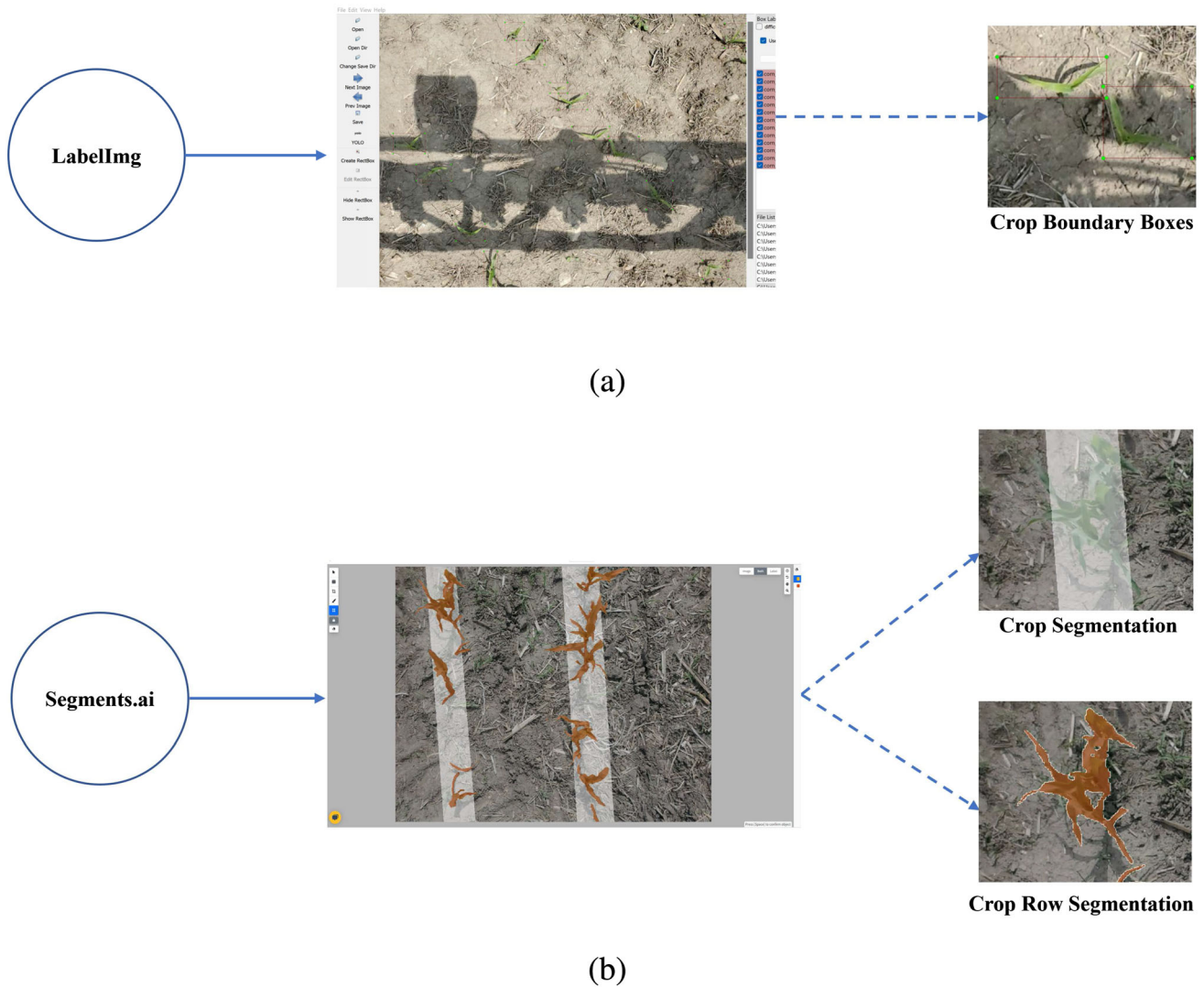
(a)

Crop Boundary Boxes

Crop Segmentation

Crop Row Segmentation

(b)

Fig. 4 Demonstration of different labeling techniques

per row and inter-plant spacing are estimated, and further calculations are performed to quantify the consistency of crops in a corn field.

2.2.1 Plant stand count using YOLOv7

The YOLOv7 is the latest addition to the YOLO series. This network is a single-stage detector that predicts all bounding boxes and class probabilities with just one pass through a CNN without requiring pre-generated region proposals. The YOLOv7 architecture consists of three main modules: a backbone, a neck, and a head. The backbone uses extended efficient layer aggregation networks (E-ELANs) to extract key features from input images. The E-ELAN uses shuffle and merge cardinality in the final layer aggregation to improve the feature learning capabilities of the network across layers, resulting in increased

accuracy. The neck module collects feature maps from different stages and utilizes FPN for enhancement. The head module makes the final detection, and a modified
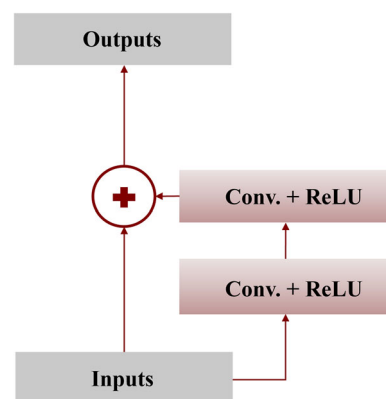

Fig. 5 Building block of residual unit

post-processing non-maximum suppression (NMS) suppresses multiple detection boxes of the same plant, resulting in a reduced number of false positives.

To improve the inference speed, the architecture incorporates a new model re-parameterization technique that combines multiple computational elements into a single module. This technique is separated into two groups: model-level and module-level ensemble. The YOLOv7 uses a concatenation-based method for scaling the model, generating models of varying sizes, and ensuring the structure remains optimal during the execution stage. The YOLOv7 is trained on a corn crops dataset for plant stand counting. The trained model is used to locate corn plants by predicting bounding boxes, which are utilized to count the number of plants in the images.

The classical NMS technique is the last post-processing step in the detection algorithms [38]. It uses a probability-bound criterion system to filter out single detection boxes out of many overlapping boxes. These overlapping boxes have false-positive predictions that need to be suppressed. Another alternative is soft NMS which applies a continuous function based on IoU to reduce the detection score of neighboring bounding boxes. This results in a gentler and more resilient suppression than the classical NMS approach [39]. However, both of these techniques have shortcomings in occlusion scenarios. They encounter challenges in detecting objects from the same class in proximity and tend to delete the true predictions. This problem is addressed by adopting the DIoU-NMS method [34], which performs well in clustered scenes and effectively handles highly overlapping target objects. This method suppresses boxes by
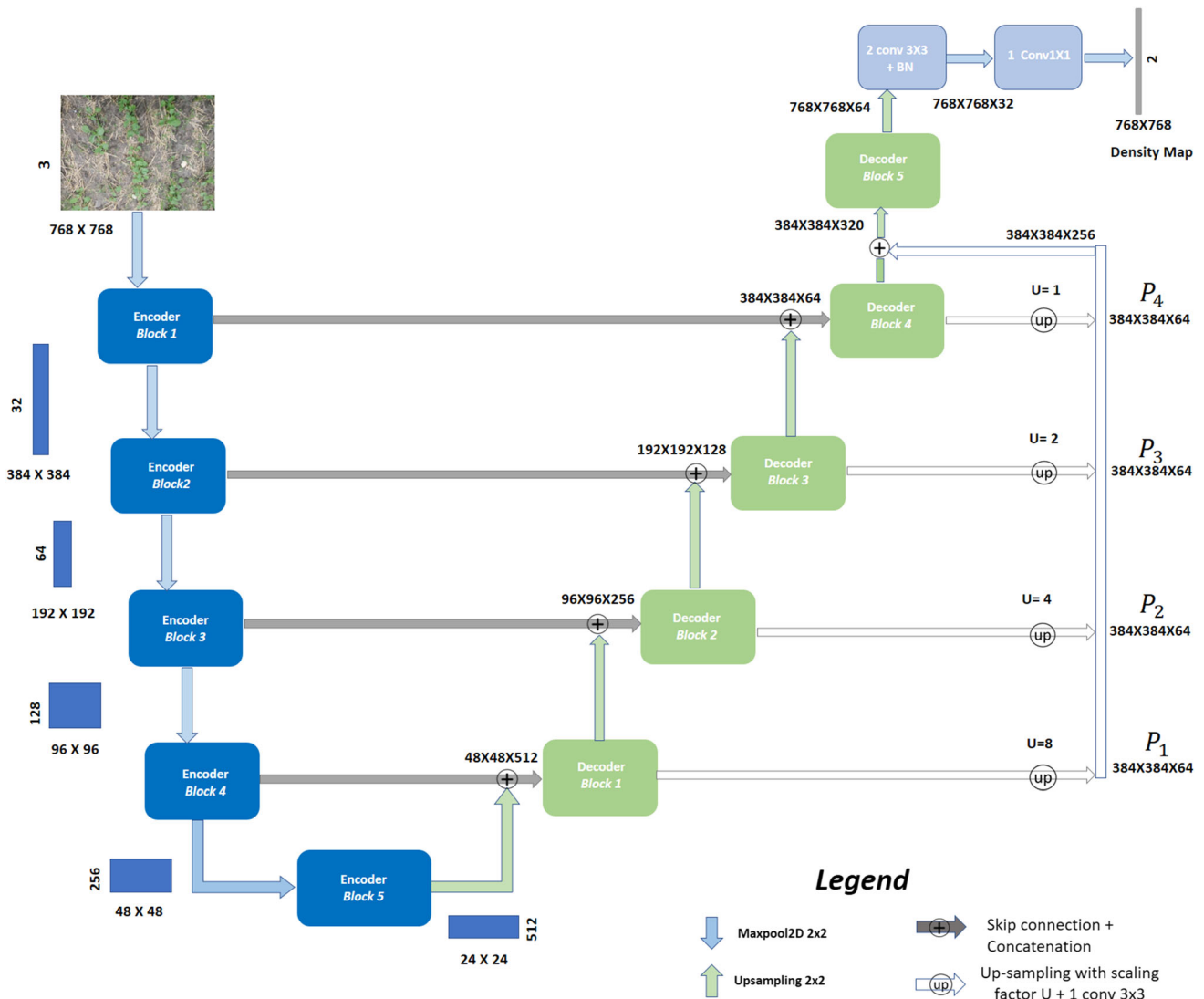

Fig. 6 SP-UNet network architecture. The number of channels is represented by the vertical numbers on the left side, while the numbers below each map represent the size of the feature map

considering the distance between the center point of the corresponding detection boxes and the overlapping area. The distance-IoU penalty term denoted as R_{DIoU} equipped with an adaptive threshold facilitates score penalization of the false neighboring boxes to achieve a refined detection box.

$$s_i = \begin{cases} s_i, & IoU - \mathcal{R}_{\text{DIoU}}(\mathcal{D}, P_i) < \varepsilon \\ 0, & IoU - \mathcal{R}_{\text{DIoU}}(\mathcal{D}, P_i) \geq \varepsilon \end{cases} \tag{1}$$

where $\mathcal{D}$ represents the highest score of the detection box, while s_i and ε are denoted as the classification score and NMS threshold, respectively. Based on a predefined threshold value obtained using the suppression criterion, the lower score of the P_i box is eliminated.

2.2.2 Proposed semantic segmentation network: SP-UNet

The proposed SP-UNet follows a U-shaped encoder–decoder structure, similar to the original UNet [40]. The encoder contains convolutional and max pooling layers, while the decoder has upsampling and convolutional layers. The U-shape of the network facilitates feature localization in the input image and preserves spatial resolution in the segmentation output. The encoder produces feature maps of various scales by convolving and down-sampling the input, which is then concatenated with the corresponding upsampling layers of the decoder using skip connections. The skip connections enable the decoder to access and combine high-level global and low-level local features, which is crucial for accurate image segmentation, particularly for enhancing edge features.

To increase the effectiveness of feature extraction during the down-sampling process, the ResNet34 [28] module is added to the UNet encoder. The ResNet34 architecture comprises multiple residual blocks, each containing several convolutional and batch normalization layers, and has shortcut connections between each block's input and output. The residual connections allow the gradients to flow more easily through the network, making it possible to train much deeper networks without the vanishing gradient problem. Integrating the ResNet34 architecture within the UNet encoder enhances the feature extraction capability of the model by utilizing the shortcut connections present in the ResNet34. This is particularly beneficial for image segmentation tasks as it allows for extracting high-level semantic information and fine-grained details from the input image, which is crucial for precise object recognition and classification [41]. To balance the number of trainable parameters and the model's ability to extract features, only the first five blocks of the ResNet34 are used.

Fig. 7 Illustration of the detected and extended line

The decoder is made up of five upsampling blocks. Each block contains transpose convolutions followed by two convolutions, batch normalization, and ReLU activation. This increases the resolution of feature maps while reducing their depth. Using FPN improves the network's performance by enabling it to incorporate multi-scale contextual information from different network levels. This is particularly useful in tasks like image segmentation, where the spatial layout of objects is important. For efficient crop row segmentation, the network needs to detect large row segments rather than specific vegetation regions in images. Utilizing an FPN in the decoder part of the network allows the network to consider global context and local details in the image. This leads to increased model performance with fewer parameters and reduced computation time (Fig. 5).

The output features of the first four decoder blocks are up-sampled to a fixed size of 384×384 and a depth of 64 (channels), creating four pyramids of the feature maps: $P_1, P_2, P_3,$ and P_4. These pyramids with 64 channels are combined, resulting in a feature map with 256 channels. This is then concatenated with the output of the fourth upsampling layer and fed to the last fifth upsampling layer. This increases its resolution and matches the input size before being processed by the final block of the network. The final block consisting of three 3×3 convolutions is applied to the output of the fifth upsampling layer, resulting in a 768×768 feature map with 32 channels. The final step applies a 1×1 convolution filter to the feature map, reducing the number of channels to match the number of classes (crop leaves and rows). This final feature map is then used to predict the class for each pixel in the image. A
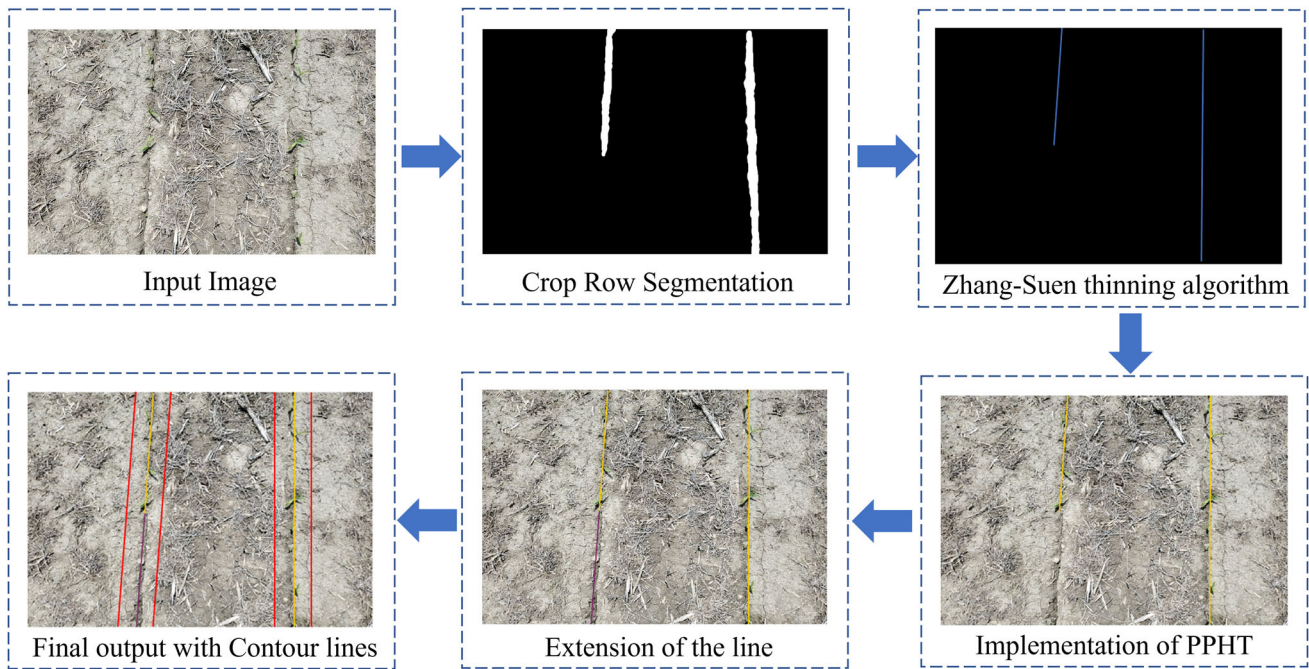
Fig. 8 Proposed crop row extraction workflow

graphical representation of the model's structure is shown in Fig. 6.

2.2.3 Crop row detection algorithm

The SP-UNet is used to segment the corn crop rows from the background, which includes the soil and weeds. The Zhang–Suen thinning method [42] is then applied to the segmented rows to obtain a minimal/pixel-wide representation, skeleton, of the crop rows [29]. This thinning method is an iterative process consisting of two passes. During each iteration, it examines whether it satisfies the following conditions before eliminating pixels from the

outer sets of pixel regions that do not belong to the skeleton.

- First pass

  1. The chosen pixel f is black and is surrounded by eight neighboring pixels.
  2. It has a minimum of two and a maximum of six black neighboring pixels.
  3. The connectivity value of this pixel is one.
  4. At least one pixel should be white in either east, north, or south position from pixel f.
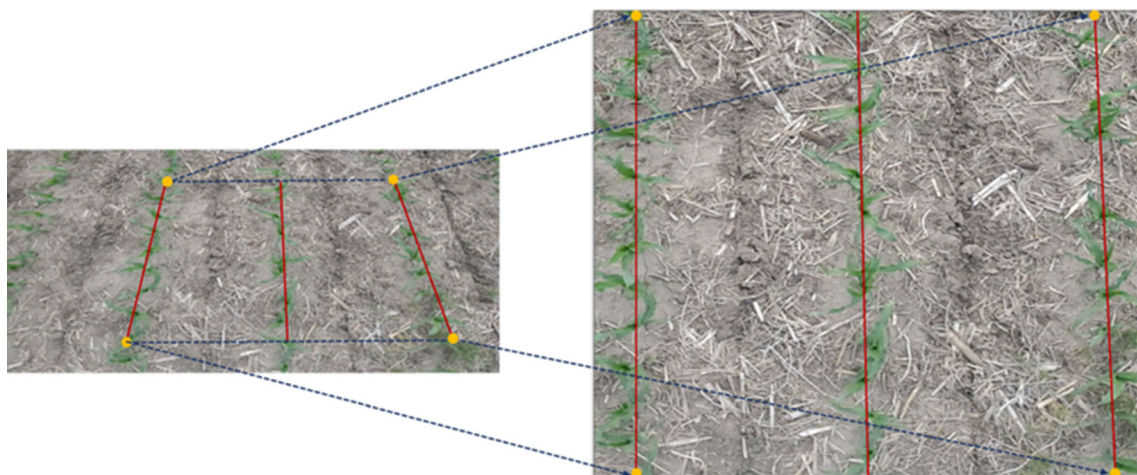 5. At least one pixel should be white in either west, east, or south position from pixel f.

Fig. 9 Demonstration of perspective view image and Bird's-eye view image
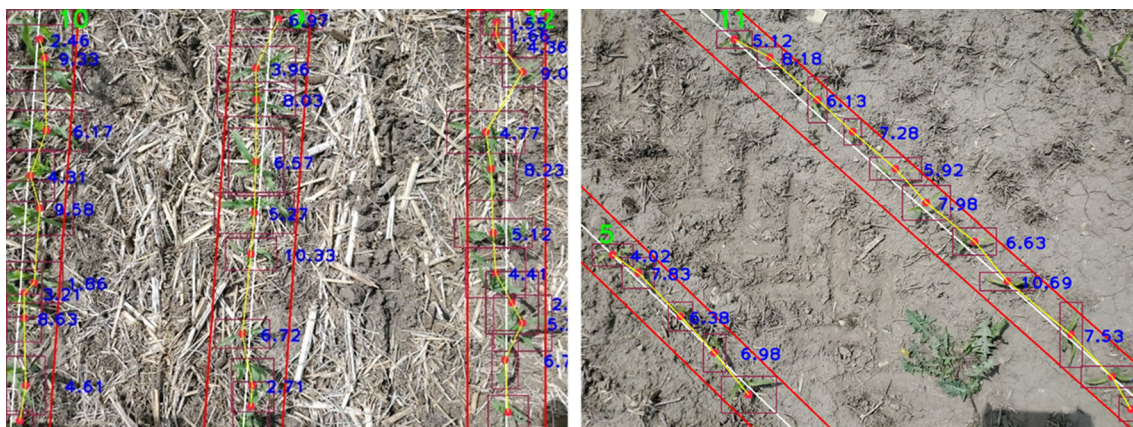
Fig. 10 Distance estimation between each crop in their respective row shown for both mid- and early-stage crops images

- Second pass

  1. The chosen pixel f is black and is surrounded by eight neighboring pixels.
  2. It has a minimum of two and a maximum of six black neighboring pixels.
  3. The connectivity value of this pixel is one.
  4. At least one pixel should be white in either east, north, or west position from pixel f.
  5. At least one pixel should be white in either west, north, or south position from pixel f.

This process uses a 3×3 window to make calculations across the entire image in a clockwise direction. Through each pass, the pixel that satisfies the conditions will be eliminated and will continue until no more pixels are removed, and a more definitive structure is obtained. There are other alternatives for thinning algorithms, such as the Guo–Hall [43]. Guo–Hall deletes redundant pixels and produces thinner skeletons compared to the Zhang–Suen algorithm but does not at times retain the structure of the component accurately [44].

Following the implementation of the thinning algorithm, a Gaussian filter with a small standard deviation (SD) is employed to minimize noise. A CV technique called PPHT is adopted at the final stage to obtain a precise center line of each crop row. This adaptation of the probabilistic Hough transform (PHT) [45] performs the standard Hough transform [46] on only a pre-selected fraction of the points. The PPHT exploits the difference in the distribution of the votes required to accurately perform line detection with the help of the different number of supporting points. This results in a significant reduction in the computation required for PPHT in contrast to PHT, where prior knowledge and the information on the fraction of points utilized for voting are needed and hence is time-consuming.

Because of the sparse inter-plant spacing at the edges of the images, an incomplete row length is obtained, leading to incorrect plants per unit calculations (plants/length of row). The detected lines are further extended to the edges to address this problem, as shown in Fig. 7. In addition to line detection, contour lines are incorporated around the row's center line. This allows the plant detection and numerical calculations to be restricted within the contour lines margin to avoid detecting and counting voluntary plants that grow outside the row. The row detection process is depicted in Fig. 8.

2.2.4 Inverse perspective transformation

The images are fundamentally two-dimensional (2D) projections of three-dimensional (3D) objects. This perspective projection leads to distortion and fails to capture information on an object's precise spacing and location from an RGB image [47]. The projection of parallel lines seems to converge and intersect at a point (vanishing point) when represented in a linear perspective. Some of the images in our dataset have crop rows converging at the vanishing point, which contrasts with the field plane, where the rows are parallel. This makes the field-scale calculation for estimating the distance between rows incorrect. The number of pixels at the image's upper edge (near the vanishing point) corresponding to a certain physical

Table 2 Comparing the accuracy and efficiency of various YOLO object detection architectures

| Model	Precision	Recall	mAP
YOLOv3	0.948	**0.902**	0.867
YOLOv4	0.943	0.869	0.894
YOLOv5	0.963	0.901	0.929
YOLOX	0.916	0.878	0.903
YOLOv7	**0.981**	0.888	**0.976**

Bold values in the table indicate statistically significant results

Table 3 Comparing various methods for detecting and classifying corn crops: an in-depth performance analysis

Research focus	Data acquisition	Method platform	Method evaluation	Reference
Screening corn ear	RGB	VGG-16 and ResNet50	Classification accuracy = 97.23%	[50]
Corn kernel counting	RGB	Semi-supervised deep learning	RMSE = 60.27 MAE = 41.36	[13]
Early-stage corn detection	UAS/RGB	Supervised learning	Classification accuracy = 96%	[2]
Corn crop stand	UAV	EXG method	Mean accuracy = 0.46	[48]
Corn crop stand	UAV	k-means clustering—segmentation	Mean accuracy = 0.91	[48]
Maize seedlings stand	RGB	Faster R-CNN with VGG19	Precision > 97.71%	[49]
Corn plant identification	RGB	Image processing algorithms	Mean misidentification ratio = 3.7%	[26]
Corn crops within the rows	RGB	**YOLOv7**	**mAP = 0.976**	**Our implementation**

Bold values in the table indicate statistically significant results

Fig. 11 Comparison of counts from NMS and DIoU-NMS with manual counts
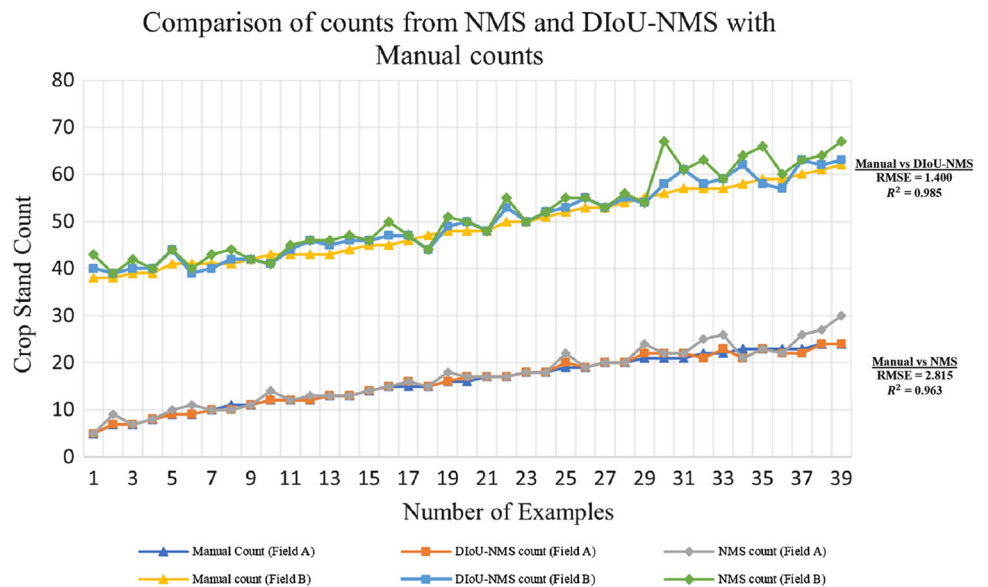
distance differs from those at the lower edge. To eliminate these non-linear perspective effects, it is important to apply inverse perspective mapping. This technique takes every pixel from a 2D projection of 3D objects and remaps it to a new position, subsequently generating a new image representation on an inverse 2D plane [47]. The equations to support this mapping are formulated as follows:

$$F = \{(x, y, z)\} \in Es^3 \tag{2}$$

$$I = \{(u, v)\} \in Es^2 \tag{3}$$

$$u(x, 0, z) = \frac{\rho(x, 0, z) - (\theta - \mu)}{\frac{2\mu}{k-1}} \tag{4}$$

$$v(x, 0, z) = \frac{\vartheta(x, 0, z) - (\varphi - \mu)}{\frac{2\mu}{g-1}} \tag{5}$$

$$\rho = \tan^{-1}\left(\frac{Z}{x}\right) \tag{6}$$

$$\vartheta = \tan^{-1}\left(\frac{h}{\sqrt{x^2 + z^2}}\right) \tag{7}$$

Table 4 Comparison of existing row detection models with our proposed SP-UNet on corn field

Architecture	Method performance	Reference
Adaptive Multi-ROI	Accuracy = 95.3%	[53]
Mask Scoring R-CNN	mAP = 82.8%	[52]
Improved Hough Transform	98% correct detection	[51]
SP-UNet	**mIoU = 0.924**	**Ours**

Bold values in the table indicate statistically significant results

The coordinates of the field plane are translated onto the image plane using (4) and (5) where θ indicates the angle between the optical projection axis on the horizontal plane and ϕ denotes the angle between the horizon and optical axis. The μ and $k \times g$ represent the camera's angular aperture and the image resolution, respectively. Lastly, the angles ρ and ϑ can be termed projection weighting factors for the pixels that change horizontally and vertically during the remapping process. In Fig. 9, the left image demonstrates the original field image where the crop rows' dimensions are non-parallel, and its vertices are represented in a trapezoid shape. The right image is the bird's-eye view resulting from the inverse perspective transformation.

2.2.5 Plant spacing calculations

Once the plant detection model detects individual plant stands and determines the plant population across each row, the spacing between the adjacent plant stands is calculated. This numerical estimation is based on the distance calculation between the plants' centroids (x,y). The Euclidean distance method is the most appropriate distance metric to determine the spacing within the neighboring plants in the rows [48]. Before this calculation, a distance matrix ensures no multiple distance estimation from each plant to the rest of the row. Similarly, another triangular matrix is generated to discard the repeated values that only initiate the distance calculation when the neighboring plant is closest to the next corresponding plant. The equation for the distance (in pixels) estimation is given by (8):

$$\text{Distance } (F, G)^2 = (F_x - G_x)^2 + (F_y - G_y)^2 \tag{8}$$

The subscripts x and y represent the centroid of the bounding box corresponding to plants F and G. The distance estimated in (8) is then translated onto field-scale (actual) values using inter-row distance obtained by (9):

$$D = \frac{\|C_2 - C_1\|}{\sqrt{1 + m^2}} \tag{9}$$

$$Y_1 = mx + c_1 \tag{10}$$

$$Y_2 = mx + c_2 \tag{11}$$

The line equations associated with each crop row are (10) and (11). D represents the shortest distance in pixels between them. As the crop rows are generally 0.3048m apart in the corn field, (12) calculates the scale for the actual distance corresponding to each pixel. This scale is then multiplied with each inter-plant distance, Distance (F, G) in (8), to obtain the distance in inches.

$$scale = 0.3048/D \tag{12}$$

The calculation based on the precise distance estimation between each plant provides a relative analysis of the consistency of the crops. The projected distance estimation on both mid- and early-stage crop images is demonstrated in Fig. 10. The blue fonts are used to denote the distance in pixels. Following the distance calculation, the coefficient of variation (CoV) is obtained, providing insight into determining the consistency of the crops.

2.2.6 LAI estimation from crop segmentation

LAI characterizes the structural attributes of a crop quantitatively to analyze the crop emergence at a particular stage. The SP-UNet model is trained to segment images into different classes, such as crops and rows. The segmented crop content calculates the total leaf area over a large field. The formulation of calculating the percentage of LAI is given as follows:

$$\text{Percentage of crop LAI} = \frac{\sum M_{\text{crop}}}{\sum N_{\text{all}}} \tag{13}$$

M_{crop} represents all the pixels allocated in crop content while $\sum N_{\text{all}}$ indicates the summation of all the pixels of the image, including crops/background/weeds.

The CoV value, along with the LAI estimate, is used as the deciding factor for assessing the consistency of the crops at a particular growth stage. These details are elaborately discussed in the next section.

Table 5 Comparison of segmentation models with our proposed model on corn field

Model	Parameters	Evaluation metrics				
		mIoU	IoU$_{\text{Background}}$	IoU$_{\text{Crop}}$	IoU$_{\text{Row}}$	Accuracy$_{\text{Val}}$
HRNet_Mscale	72.10M	**0.915**	0.975	0.850	0.920	**98.3%**
Simple UNet	31.04M	0.867	**0.986**	0.748	0.885	98.1%
DeepLabV3+ with ResNet50	24.81M	0.892	0.967	0.865	0.882	98.5%
SegNet50 with ResNet50	34.65M	0.878	0.978	0.778	0.739	95.3%
SP-UNet (ours)	**21.85M**	0.913	0.950	**0.867**	**0.924**	96.5%

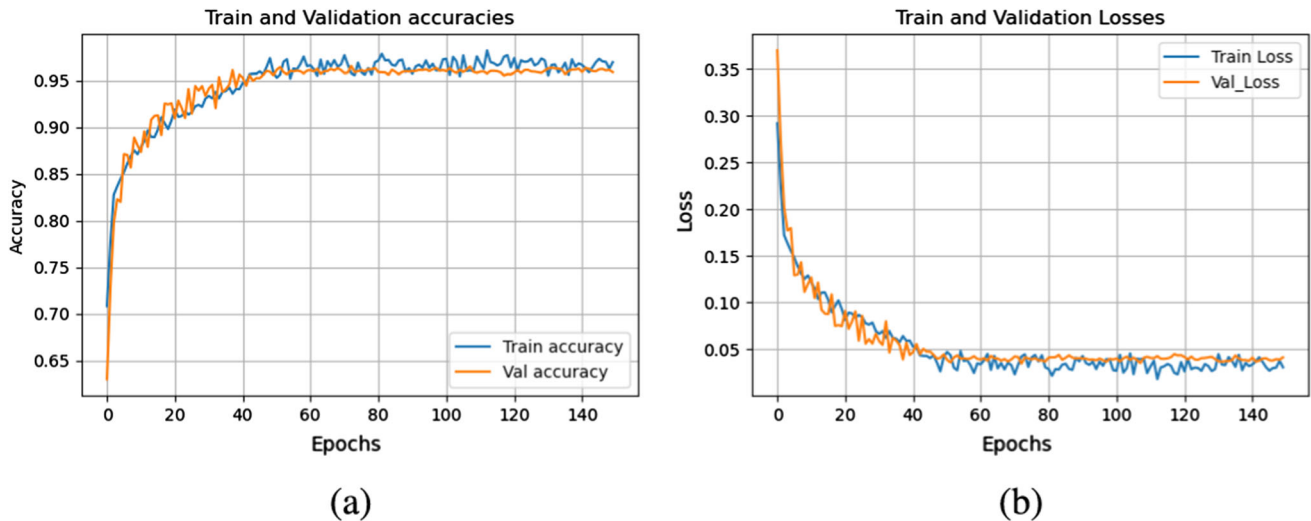Bold values in the table indicate statistically significant results

Fig. 12 Demonstration of different labeling techniques

3 Results

We conduct experiments to test the effectiveness of our proposed QuanCro method. We evaluate the accuracy of our plant counting model by comparing it to manual counts on a separate test dataset. We also compare the performance of different detection networks, including YOLOv3, YOLOv5, and YOLO7, based on their mAP and F1 scores. We visually assess the detection results to determine which model performs the best. Additionally, we analyze several segmentation architectures, such as SegNet, UNet, HRNet_Mscale, and DeepLabV3+, with different backbones like ResNet and VGG and compare them to our proposed SP-UNet model. Finally, we calculate numerical estimations to determine the consistency of corn crops and establish their correlation with the LAI from the SP-UNet.

3.1 Evaluation metrics

Several metrics can be used when evaluating a segmentation model's performance. One of the most commonly used and effective methods is IoU, the Jaccard index. This metric is calculated by determining the ratio of the intersection of the predicted segmentation mask (Pr) and the ground-truth segmentation mask (Gr) to their union. It ranges from 0 to 1, with a value of 1 indicating a perfect segmentation, given as:

$$\text{IoU} = \frac{\text{Pr} \cap Gr}{\text{Pr} \cup Gr} \tag{14}$$

Mean IoU (mIoU) is the average of the IoU scores across all classes and provides an overall measure of the model's performance. The formula for mIoU is as follows:

$$\text{mIoU} = \frac{1}{c} \times \sum \text{IoU}_i \tag{15}$$

where c is the number of classes and IoU_i is the IoU score of the ith class, calculated using the formula in (14). Another popular metric is the Dice coefficient (DC), which is calculated as the ratio of the intersection of the predicted and ground-truth segmentation masks to the sum of the sizes of the masks. It also ranges from 0 to 1, with a value of 1 indicating perfect segmentation.

$$DC = 2 \times \frac{overlapping\ pixels}{total\ pixels\ in\ both\ Pr\ and\ Gr} \tag{16}$$

Furthermore, accuracy is used as an additional metric to evaluate and compare our model with the baseline/reference models.

Mean average precision (mAP) is commonly used in object detection and measures the average precision (AP) across different object classes, providing an overall measure of the model's performance. To determine the mAP, the AP for each class is computed and then the APs are averaged across all classes. The AP score for a specific class evaluates the model's P and recall (R) when detecting instances of that class. It is determined by creating a P–R curve and taking the average of the P values at various R levels.

$$\text{AP} = \int_0^1 P(R)\text{d}R \tag{17}$$

$$\text{mAP} = \frac{1}{\text{cL}} \sum_{i=0}^{\text{cL}} \text{A P}_i \tag{18}$$

By employing these evaluation criteria, we ensure a comprehensive and rigorous assessment of our frameworks' performance in semantic segmentation and object detection
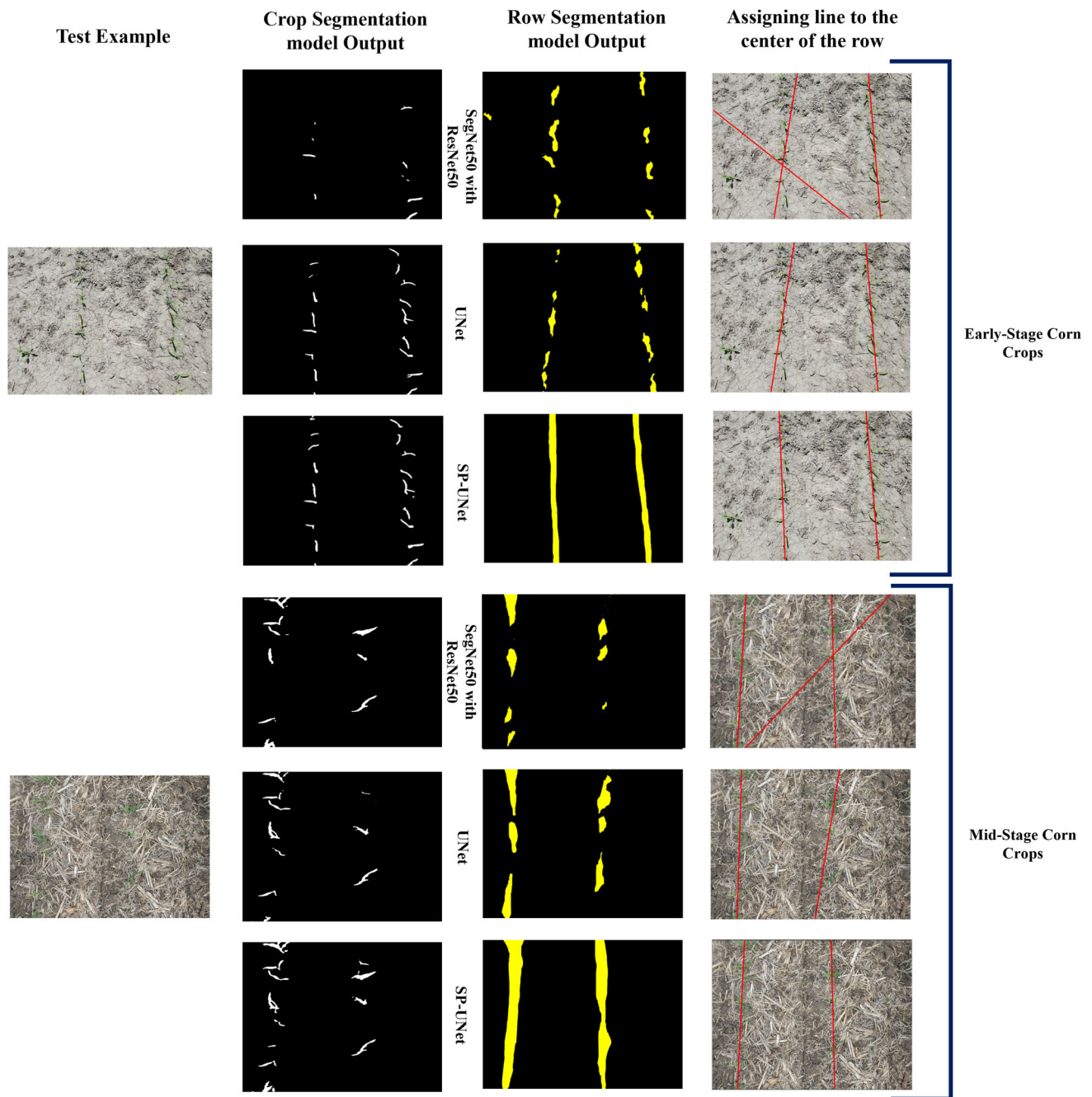
Fig. 13 Demonstration of different CNN models for segmenting crops and rows in corn crop images, and the final center-line association to the rows

tasks. These metrics allow for fair comparisons with other state-of-the-art networks.

3.2 Evaluation of plant detection model

In this study, we evaluate the performance of different object detection models in the YOLO pipeline by comparing their results on our dataset of corn crops. To ensure a fair analysis, we use consistent hyperparameter settings, including a learning rate of 1×10^{-2}, a batch size of 16, and 120 training epochs for all models. During training, we utilize the Adam optimizer and binary cross-entropy loss function for all models. According to Table 2, YOLOv7 outperforms all other models with the highest precision and mAP values, which are 5.06% and 1.87% higher than YOLOv5. While YOLOv3 has the highest recall value of 0.902, its mAP value is lower than YOLOv5 and YOLOv7.
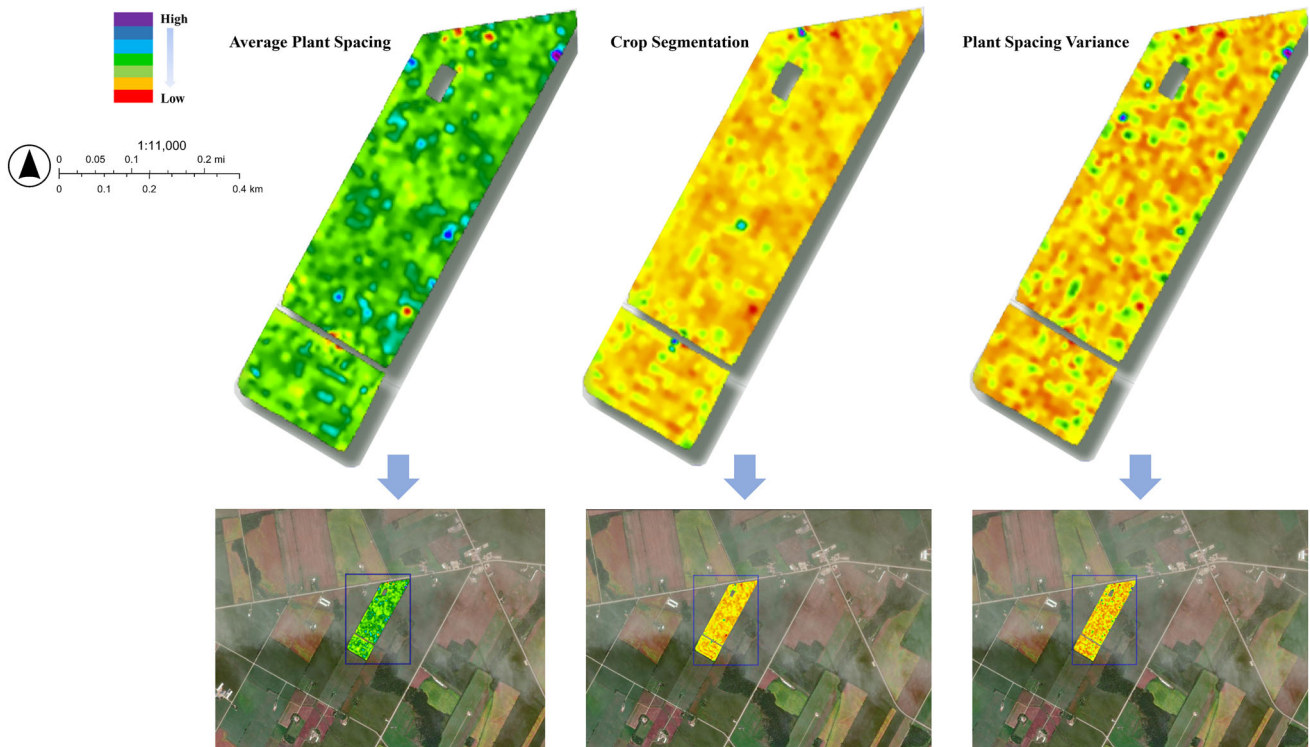
Fig. 14 Heatmaps of average plant spacing, variance, and crop segmentation overlaid on top of an RGB satellite image background

Table 6 Establishment of the correlation between the crop variables with a consistency determinant to evaluate crop consistency

Ex. images	Total count	Variance spacing	Mean spacing	SD	CoV	LAI	CD
Image 1	18	2.835	0.726	0.852	0.301	0.802	Good
Image 2	16	3.476	1.940	1.393	0.401	0.626	Moderate
Image 3	17	3.706	2.623	1.620	0.437	0.597	Moderate
Image 4	15	3.334	2.145	1.465	0.439	0.623	Moderate
Image 5	15	2.330	1.550	1.245	0.535	0.553	Bad
Image 6	14	2.041	1.229	1.109	0.543	0.515	Bad
Image 7	12	2.949	2.300	1.717	0.747	0.274	Bad
Image 8	11	2.655	2.513	1.630	0.648	0.222	Bad

Among all the models evaluated in this study, YOLOX has the worst performance.

Table 3 compares networks used to identify and count corn stands in both RGB and UAV imagery. While semi-supervised and supervised DL methods have been successful in detecting and counting early-stage corn crops, their effectiveness on mid-stage crops has not been reported yet [2, 13]. The table highlights that the faster R-CNN method integrated with VGG19 achieves an impressive 97.71% precision rate for detecting corn crops. However, its performance decreases as the crops grow and in varying lighting conditions [49]. On the other hand, our YOLOv7-based corn plant stand counting approach achieves high precision, outperforming other networks used for the same task.

In Fig. 11, we compare the actual manual crop count in test Fields A and B at different growth stages with the predicted count obtained from YOLOv7. We use two NMS post-processing techniques: NMS and DIoU-NMS. We evaluate 38 examples randomly selected from each field. Field A primarily consists of mid-stage crops, while Field B represents early-stage crops. The number of crop stands in Field A is higher than in Field B.

The results demonstrate that the correlation between the manual and predicted count using the DIoU-NMS is stronger than the NMS. We observe similar trends in Field B, but with slightly higher variations between the actual and predicted count using both techniques. This is due to the higher crop emergence at later growth stages with overlapping crops and the resemblance of weeds to the
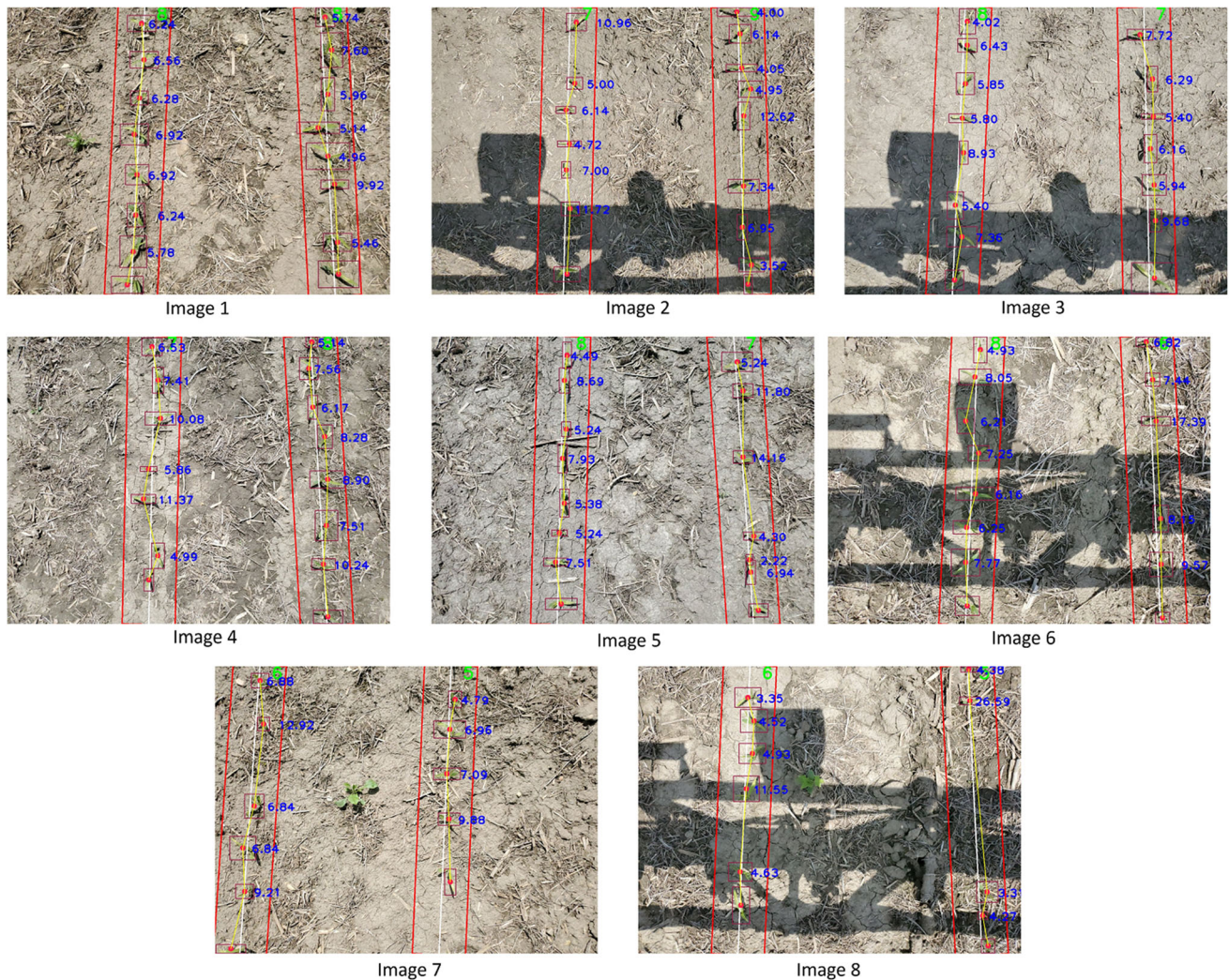
Fig. 15 Representation of the consistency of crop images, showing examples of both well-consistent and poorly consistent crops

narrow structure of the leaves. Despite this, the correlation between the actual and predicted counts using DIoU-NMS post-processing is superior. It achieves a score of 1.400 for RMSE, which is 2.3% better than the RMSE achieved by the model when utilizing the NMS technique. The R^2 value using the DIoU-NMS is 0.985.

3.3 Crop row extraction networks

Based on corn fields in Table 4, the effectiveness of the proposed SP-UNet for extracting crop rows is evaluated using existing methods. However, a modified technique using the Hough transform yields only a 98% correct detection rate on a limited dataset of 100 images, which is insufficient for determining overall performance in different scenarios [51]. Furthermore, this technique is sensitive to noise and low contrast in the data. Another method, mask scoring R-CNN, has a high mAP of 82.8%, but its

accuracy decreases with lower spatial resolution [52]. Although the Adaptive Multi-ROI method [53] has a high accuracy of around 95%, it is computationally more expensive than the proposed SP-UNet.

3.4 Evaluation of segmentation models and their comparison with SP-UNet

The performance of various models is compared in Table 5 using several evaluation metrics, including mIoU, validation accuracy, and the number of trainable parameters for each model. Among all the models, DeepLabV3+ with ResNet50 has the lowest number of trainable parameters and is computationally faster than the others. However, it has a lower mIoU score of 2.6% and 2.4% compared to HRNet_Mscale and SP-UNet, respectively. SegNet50, which is SegNet with a ResNet50 backbone, has the lowest IoU score of 0.739 for crop rows and requires 58.6% more

parameters than SP-UNet. Although HRNet_Mscale has the best mIoU score, it is computationally complex and requires many parameters. Additionally, its high validation accuracy is not a reliable metric for evaluating segmentation models due to imbalanced class data. In contrast, the SP-UNet model has better crop and row IoU scores than HRNet_Mscale and uses 239% fewer parameters. Its faster processing speed and minimal impact on IoU scores make it a more suitable choice for various applications. Furthermore, The SP-UNet obtains the second-best mIoU score, employing only 21.85M parameters.

The SP-UNet is trained using the Adam optimizer and a learning rate of 1×10^{-2}. The accuracy and loss graphs for the train and validation sets are shown in Fig. 12. In Fig. 12a, it is evident that the accuracy fluctuates frequently throughout the training epochs. The accuracy for both sets reaches a cut-off point at approximately the 40th epoch where it reaches 95%. After this point, the validation accuracy follows a steady trend, while the training accuracy shows sporadic fluctuations. As the training continues, the model's performance on the validation dataset improves, reaching an accuracy of 96.5%. At the same time, the training accuracy reaches 98.6%. The loss graphs in Fig. 12b show that the model generalizes well on unseen data. This is supported by the rapid convergence of the training and validation loss, which eventually reaches values of 0.04 and 0.02, respectively, at the 72nd epoch.

The picture shown in Fig. 13 compares various segmentation models for crop row extraction and crop segmentation used for LAI estimation. This analysis is being conducted during the early and mid-stages of crop growth, and the results are being demonstrated using two randomly selected images from each stage. The first column displays the test images, the second and third columns show the output of crop and row segmentation, respectively, and the final column shows the center-line association with the crop rows. Among the segmentation models being tested, the SegNet50-based model performs the worst, with incorrect row segmentation leading to false positives during early-stage crop growth and misaligned rows during mid-stage growth. Additionally, it is missing some crop pixels, leading to incorrect LAI calculations. The UNet architecture performs better in segmenting rows but misclassifies some background/weed pixels as crops, leading to incorrect row alignment. It also has small misses in crop segmentation, especially in poor field conditions. Our proposed SP-UNet model effectively overcomes these challenges and accurately segments both rows and crops, which is beneficial for LAI and determining the distance between crops within the predicted rows.

3.5 Quantification of crops' consistency

Extensive analysis is carried out on corn cropland spanning a thousand acres to determine the correlation between crop emergence percentage, average inter-plant spacing, and variance. The field area has been presented in a heatmap format using the ArcGIS platform [54], as shown in Fig. 14. This online mapping application utilizes KMZ (Keyhole Markup Language Zipped) files to analyze geographical data. KMZ files are compressed folders containing all the necessary files and KML (Keyhole Markup Language) format files representing the location coordinates. These location coordinates are represented in XML notation. The left-most map demonstrates the average plant spacing obtained through distance estimation between consecutive plant stands. The middle map showcases the proposed SP-UNet model's crop segmentation, while the right-most image displays the variance obtained from the plant spacing. The field heatmaps exhibit a significant degree of dispersion, indicating a considerable variation in crop spacing. A correlation between mean plant spacing and variance can be observed in the left and right maps, pinpointing key locations with high or low spacing. Further examination of these patterns can be conducted through crop segmentation techniques. Regions with high crop density display a lower variance, whereas areas with low crop density exhibit a higher variance. At an intermediate level of crop density, the dispersion appears to be moderate.

The LAI value provides information about the percentage of crops in an image, while the CoV is derived from the SD and mean spacing. These variables have a strong correlation that can give insight into the consistency of crops. The correlation between them is established numerically in Table 6, and a consistency determinant is assigned based on these quantifications. To assess consistency, eight images are randomly selected from the test data. The table shows that as the spacing variance increases with the CoV value, the LAI decreases due to missed crops. This suggests an inversely proportional relationship. When CoV is around 30%, and LAI is around 80%, the crop is uniform across the area with minimal sparsity in the rows. When CoV is around 70%, and LAI is around 25%, the crop is either in a stressed condition or the field has very low-density areas. The expression "moderate" indicates the acceptable range of variations in crop consistency. As CoV increases, the estimation of crop LAI tends to decrease, as observed in Fig. 15. Comparing Image 1, which shows good consistency in crop spacing and an LAI of 0.802, to Image 8, which exhibits poor consistency and a 261.3% lower LAI value. While the plant stand count is a valuable metric, it is insufficient to evaluate crop consistency fully.

This becomes apparent when observing Images 6 and 8, where one section of a row contains multiple crops close together, while the other crops in the next row are widely scattered. The CoV value is helpful in these circumstances. Utilizing the CoV with LAI can provide a more comprehensive understanding of crop health. A lower LAI value can indicate stress in the crops.

4 Conclusion

This research presents a novel vision-based framework—QuanCro to assess crops' consistency by estimating four crop parameters; plant stands count, inter-plant distance, coefficient of variation, and crop coverage in corn fields. Accurate estimation of these parameters is crucial for operational decisions, such as replanting and post-emerge herbicide applications, and assessing spatial production variability. QuanCro utilizes RGB imagery captured using a top-mounted camera on the field machinery. It uses the state-of-the-art object detector YOLOv7 to determine the number and location of plant stands. Additionally, this framework uses our proposed semantic segmentation network, SP-UNet encoder–decoder architecture, for segmenting crops and crop rows. The SP-UNet has fewer parameters (21.85M) than other semantic segmentation models such as SegNet (34.65M), HRNet_Mscale (72.10M), and Simple UNet (31.04M). Despite having fewer parameters, the SP-UNet achieves a higher segmentation of mIoU 0.913 with IoU scores of 0.94 and 0.924 for crop and crop rows segmentation, respectively. The segmented rows are further processed using the Zhang–Suen thinning and progressive probabilistic Hough transform to assign a line to each row's center. The segmented crop provides leaf area index estimation.

This whole framework is promising, and it holds significant potential for future research applications in assessing the consistency of various crop types and adapting it to diverse agricultural settings. One area of interest for upcoming studies is exploring the integration of alternative sensor modalities, such as multispectral or hyperspectral imaging, to enhance QuanCro's adaptability and robustness in different field conditions. It is also crucial to acknowledge the limitation of this proposed framework, particularly when dealing with late growth stages of crops. At this stage, individual plants tend to cluster closely together, posing a challenge for the framework's object detector to detect and distinguish individual plants accurately. The overlapping nature of these clusters can lead to erroneous inter-plant spacing calculations, ultimately affecting the overall assessment of crops' consistency.

Data availability The data that support the findings of this study are available from Croptimistic Technology Inc. Still, restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of Croptimistic Technology Inc.

Declarations

Conflict of interest The authors state no conflict of interest in this manuscript's subject matter or material.

References

1. Ren X, Sun D, Wang Q (2016) Modeling the effects of plant density on maize productivity and water balance in the loess plateau of China. Agric Water Manag 171:40–48. https://doi.org/10.1016/j.agwat.2016.03.014

2. Varela S, Dhodda PR, Hsu WH, Prasad PV, Assefa Y, Peralta NR, Griffin T, Sharda A, Ferguson A, Ciampitti IA (2018) Early-season stand count determination in corn via integration of imagery from unmanned aerial systems (UAS) and supervised learning techniques. Remote Sens 10(2):343. https://doi.org/10.3390/rs10020343

3. Che Y, Wang Q, Zhou L, Wang X, Li B, Ma Y (2022) The effect of growth stage and plant counting accuracy of maize inbred lines on lai and biomass prediction. Precis Agric 66:1–27. https://doi.org/10.1007/s11119-022-09915-1

4. Ahila Priyadharshini R, Arivazhagan S, Arun M, Mirnalini A (2019) Maize leaf disease classification using deep convolutional neural networks. Neural Comput Appl 31:8887–8895. https://doi.org/10.1007/s00521-019-04228-3

5. Nagaraju M, Chawla P (2022) Maize crop disease detection using npnet-19 convolutional neural network. Neural Comput Appl 66:1–25. https://doi.org/10.1007/s00521-022-07722-3

6. Deb M, Garai A, Das A, Dhal KG (2022) Ls-net: a convolutional neural network for leaf segmentation of rosette plants. Neural Comput Appl 34(21):18511–18524. https://doi.org/10.1007/s00521-022-07479-9

7. Mishra AM, Harnal S, Gautam V, Tiwari R, Upadhyay S (2022) Weed density estimation in soya bean crop using deep convolutional neural networks in smart agriculture. J Plant Dis Prot 129(3):593–604. https://doi.org/10.1007/s41348-022-00595-7

8. Russel NS, Selvaraj A (2022) Leaf species and disease classification using multiscale parallel deep cnn architecture. Neural Comput Appl 34(21):19217–19237. https://doi.org/10.1007/s00521-022-07521-w

9. Bouguettaya A, Zarzour H, Kechida A, Taberkit AM (2022) Deep learning techniques to classify agricultural crops through UAV imagery: a review. Neural Comput Appl 34(12):9511–9536. https://doi.org/10.1007/s00521-022-07104-9

10. Kitano BT, Mendes CC, Geus AR, Oliveira HC, Souza JR (2019) Corn plant counting using deep learning and UAV images. IEEE

Geosci Remote Sens Lett. https://doi.org/10.1109/LGRS.2019.2930549

11. Buzzy M, Thesma V, Davoodi M, Mohammadpour Velni J (2020) Real-time plant leaf counting using deep object detection networks. Sensors 20(23):6896. https://doi.org/10.3390/s20236896

12. Vong CN, Conway LS, Feng A, Zhou J, Kitchen NR, Sudduth KA (2022) Corn emergence uniformity estimation and mapping using UAV imagery and deep learning. Comput Electron Agric 198:107,008. https://doi.org/10.1016/j.compag.2022.107008

13. Khaki S, Pham H, Han Y, Kuhl A, Kent W, Wang L (2021) Deepcorn: a semi-supervised deep learning method for high-throughput image-based corn kernel counting and yield estimation. Knowl Based Syst 218:106,874. https://doi.org/10.1016/j.knosys.2021.106874

14. Mota-Delfin C, López-Canteñs GdJ, López-Cruz IL, Romantchik-Kriuchkova E, Olguín-Rojas JC (2022) Detection and counting of corn plants in the presence of weeds with convolutional neural networks. Remote Sens 14(19):4892. https://doi.org/10.3390/rs14194892

15. Bochkovskiy A, Wang CY, Liao HYM (2020) Yolov4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934

16. Jocher G, Stoken A, Borovec J, NanoCode012, ChristopherSTAN, Changyu L, Laughing, tkianai, Hogan A, lorenzomammana, yxNONG, AlexWang1900, Diaconu L, Marc, wanghaoyang0106, ml5ah, Doug, Ingham F, Frederik, Guilhen, Hatovix, Poznanski J, Fang J, Yu L, changyu98, Wang M, Gupta N, Akhtar O, PetrDvoracek, Rai P (2020) Ultralytics/yolov5: v3.1—bug fixes and performance improvements. Zenodo. https://doi.org/10.5281/zenodo.4154370

17. Guo W, Zheng B, Potgieter AB, Diot J, Watanabe K, Noshita K, Jordan DR, Wang X, Watson J, Ninomiya S et al (2018) Aerial imagery analysis-quantifying appearance and number of sorghum heads for applications in breeding and agronomy. Front Plant Sci 9:1544. https://doi.org/10.3389/fpls.2018.01544

18. Guerrero JM, Guijarro M, Montalvo M, Romeo J, Emmi L, Ribeiro A, Pajares G (2013) Automatic expert system based on images for accuracy crop row detection in maize fields. Expert Syst Appl 40(2):656–664. https://doi.org/10.1016/j.eswa.2012.07.073

19. García-Santillán I, Peluffo-Ordoñez D, Caranqui V, Pusdá M, Garrido F, Granda P (2018) In: International conference on information technology & systems. Springer, pp 355–366. https://doi.org/10.1007/978-3-319-73450-7_34

20. Winterhalter W, Fleckenstein FV, Dornhege C, Burgard W (2018) Crop row detection on tiny plants with the pattern Hough transform. IEEE Robot Autom Lett 3(4):3394–3401. https://doi.org/10.1109/LRA.2018.2852841

21. Bah MD, Hafiane A, Canals R (2017) In: 2017 Seventh international conference on image processing theory, tools and applications (IPTA). IEEE, pp 1–6. https://doi.org/10.1109/IPTA.2017.8310102

22. Bah MD, Hafiane A, Canals R (2019) Crownet: deep network for crop row detection in UAV images. IEEE Access 8:5189–5200. https://doi.org/10.1109/ACCESS.2019.2960873

23. Ullah HS, Bais A (2022) Evaluation of model generalization for growing plants using conditional learning. Artif Intell Agric 6:189–198. https://doi.org/10.1016/j.aiia.2022.09.006

24. Doha R, Al Hasan M, Anwar S, Rajendran V (2021) In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining, pp 2773–2781. https://doi.org/10.1145/3447548.3467155

25. de Silva R, Cielniak G, Gao J (2021) Towards agricultural autonomy: crop row detection under varying field conditions using deep learning. arXiv preprint arXiv:2109.08247

26. Tang L, Tian LF (2008) Plant identification in mosaicked crop row images for automatic emerged corn plant spacing measurement. Trans ASABE 51(6):2181–2191. https://doi.org/10.13031/2013.25381

27. Long J, Shelhamer E, Darrell T (2015) In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440. https://doi.org/10.1109/cvpr.2015.7298965

28. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778. https://doi.org/10.1109/cvpr.2016.90

29. Sudarma M, Sutramiani NP (2014) The thinning Zhang–Suen application method in the image of Balinese scripts on the papyrus. Int J Comput Appl 91(1):9–13. https://doi.org/10.5120/15844-4726

30. Galamhos C, Matas J, Kittler J (1999) In: Proceedings of the 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149), (IEEE, 1999), vol 1, pp 554–560. https://doi.org/10.1109/CVPR.1999.786993

31. Baar S, Kobayashi Y, Horie T, Sato K, Suto H, Watanabe S (2022) Non-destructive leaf area index estimation via guided optical imaging for large scale greenhouse environments. Comput Electron Agric 197:106,911. https://doi.org/10.1016/j.compag.2022.106911

32. Asad MH, Bais A (2020) Crop and weed leaf area index mapping using multi-source remote and proximal sensing. IEEE Access 8:138,179-138,190. https://doi.org/10.1109/ACCESS.2020.3012125

33. Wang CY, Bochkovskiy A, Liao HYM (2022) Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696

34. Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D (2020) In: Proceedings of the AAAI conference on artificial intelligence, vol 34, pp 12993–13000. https://doi.org/10.1609/aaai.v34i07.6999

35. DeVries T, Taylor GW (2017) Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552

36. Yun S, Han D, Oh SJ, Chun S, Choe J, Yoo Y (2019) In: Proceedings of the IEEE/CVF international conference on computer vision, pp 6023–6032. https://doi.org/10.1109/iccv.2019.00612

37. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D (2017) mixup: beyond empirical risk minimization. arXiv preprint arXiv:1710.09412

38. Hosang J, Benenson R, Schiele B (2017) In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4507–4515. https://doi.org/10.1109/cvpr.2017.685

39. Bodla N, Singh B, Chellappa R, Davis LS (2017) In: Proceedings of the IEEE international conference on computer vision, pp 5561–5569. arXiv:1704.04503

40. Ronneberger O, Fischer P, Brox T (2015) In: International conference on medical image computing and computer-assisted intervention (Springer, 2015), pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

41. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J (2017) A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857

42. Zhang TY, Suen CY (1984) A fast parallel algorithm for thinning digital patterns. Commun ACM 27(3):236–239. https://doi.org/10.1145/357994.358023

43. Guo Z, Hall RW (1992) Fast fully parallel thinning algorithms. CVGIP Image Underst 55(3):317–328. https://doi.org/10.1016/1049-9660(92)90029-3

44. Jain H, Kumar AP (2017) A sequential thinning algorithm for multi-dimensional binary patterns. arXiv preprint arXiv:1710.03025

45. Stephens RS (1991) Probabilistic approach to the Hough transform. Image Vis Comput 9(1):66–71. https://doi.org/10.1016/0262-8856(91)90051-p

46. Hart PE, Duda R (1972) Use of the Hough transformation to detect lines and curves in pictures. Commun ACM 15(1):11–15. https://doi.org/10.1145/361237.361242

47. Muad AM, Hussain A, Samad SA, Mustaffa MM, Majlis BY (2004) In: 2004 IEEE Region 10 conference TENCON 2004. IEEE, pp 207–210. https://doi.org/10.1109/TENCON.2004.1414393

48. Shirzadifar A, Maharlooei M, Bajwa SG, Oduor PG, Nowatzki JF (2020) Mapping crop stand count and planting uniformity using high resolution imagery in a maize crop. Biosyst Eng 200:377–390. https://doi.org/10.1016/j.biosystemseng.2020.10.013

49. Quan L, Feng H, Lv Y, Wang Q, Zhang C, Liu J, Yuan Z (2019) Maize seedling detection under different growth stages and complex field environments based on an improved Faster R-CNN. Biosyst Eng 184:1–23. https://doi.org/10.1016/j.biosystemseng.2019.05.002

50. Zhang J, Ma Q, Cui X, Guo H, Wang K, Zhu D (2020) High-throughput corn ear screening method based on two-pathway convolutional neural network. Comput Electron Agric 175:105,525. https://doi.org/10.1016/j.compag.2020.105525

51. Zhang L, Grift TE (2010) In: 2010 Pittsburgh, Pennsylvania, June 20–June 23, 2010 (American Society of Agricultural and Biological Engineers, 2010), p 1. https://doi.org/10.13031/2013.29834

52. Pang Y, Shi Y, Gao S, Jiang F, Veeranampalayam-Sivakumar AN, Thompson L, Luck J, Liu C (2020) Improved crop row detection with deep neural network for early-season maize stand count in UAV imagery. Comput Electron Agric 178:105766. https://doi.org/10.1016/j.compag.2020.105766

53. Zhou Y, Yang Y, Zhang B, Wen X, Yue X, Chen L (2021) Autonomous detection of crop rows based on adaptive multi-ROI in maize fields. Int J Agric Biol Eng 14(4):217–225. https://doi.org/10.25165/j.ijabe.20211404.6315

54. Tateosian L, Tateosian L (2015) Arcgis and python. Python For ArcGIS, pp 77–94. https://doi.org/10.1007/978-3-319-18398-5_5