Computers and Electronics in Agriculture 219 (2024) 108795

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

# Automated segmentation of individual leafy potato stems after canopy consolidation using YOLOv8x with spatial and spectral features for UAV-based dense crop identification

Hanhui Jiang [a], Bryan Gilbert Murengami [a], Liguo Jiang [a], Chi Chen [a], Ciaran Johnson [f], Fernando Auat Cheein [f], Spyros Fountas [e], Rui Li [a,*], Longsheng Fu [a,b,c,d,*]

[a] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
[c] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling, Shaanxi 712100, China
[d] Northwest A&F University Shenzhen Research Institute, Shenzhen, Guangdong 518000, China
[e] Agricultural University of Athens, Athens 11855, Greece
[f] UK National Robotarium, Edinburgh Centre for Robotics, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK

ARTICLE INFO

*Keywords:*
Deep learning
Instance segmentation
Potato phenotyping
Spectral feature
UAV imagery

ABSTRACT

High throughput phenotyping of potatoes after canopy consolidation is crucial to crop breeding and management. A prior step is to segment their leafy potato stems, which is challenging after canopy consolidation because potato stems are dense and intertwined. Current methods for dense crop segmentation are manual. This study equipped unmanned aerial vehicles with a high-resolution RGB sensor in ultra-low flight as a high-throughput alternative. An end-to-end method was proposed to segment their leafy potato stems using YOLOv8x and five kinds of band combinations, i.e., RGB, RGB-DSM, RGB-CHM, RGB-DSM × 3, RGB-ExG. The YOLOv8x model with the RGB-DSM combination achieved superior performance with *F*1 *score* of 0.86 and *Intersection over Union* (*IoU*) of 0.83. Both *F*1 *score* and *IoU* improved by more than 16 %, when adding DSM or CHM to RGB images. Results demonstrated that height mutation at the edge of leafy potato stems played a crucial role in improving the segmentation of leafy potato stems. Millimeter-level ground sampling distance facilitates high throughput phenotyping of potatoes. The accuracy and efficiency of YOLOv8x has great potential for guiding the phenotypic automation of potatoes as well as other arable crops through remote sensing.

## 1. Introduction

Phenotypes of the above-ground plants after canopy consolidation are critical for crop breeding and crop management of potatoes. Potatoes (*Solanum tuberosum* L.) are a dominant non-grain crop worldwide and are vital in feeding the increasing world population (Johnson and Auat Cheein, 2023). Potato production has increased by $1.03 \times 10^8$ tons over the past 60 years and is expected to keep growing (UN Food & Agriculture Organization, 2023). An urgent need for crop management and breeding is underscored by such growth, with adequate phenotyping being a potential tool. Phenotypes of above-ground potato plants directly affect radiation perception by crops, thus impacting tuber yield. Tuber yield is an important parameter in crop management and breeding which guides the selection of desirable traits (Silva-Díaz et al., 2020). Especially after canopy consolidation, intercepted radiation is converted mainly to potato tubers. Thus, it's worth assessing phenotypes of above-ground plants after canopy consolidation.

Unmanned aerial vehicles (UAVs) with remote sensors are a flexible and nondestructive method that supports in-field phenotyping research. The advantages of UAVs in terms of spatial and real-time monitoring capabilities meet the demand for high-throughput phenotypic analysis, especially when other remote sensing platforms such as satellites and ground-based remote sensing (Liu et al., 2022). Notably, ultralow-altitude UAVs have been showing promising prospects in accessing precise phenotype traits of low vegetation (Siebring et al., 2019; Xiao et al., 2023). UAVs can be equipped with various sensors, usually RGB sensor, hyperspectral sensor, multispectral sensor, Light Detection and Ranging (LiDAR) and Synthetic Aperture Radar (SAR), depending on

* Corresponding authors.
*E-mail addresses:* ruili1216@nwafu.edu.cn (R. Li), fulsh@nwafu.edu.cn (L. Fu).

https://doi.org/10.1016/j.compag.2024.108795
Received 16 December 2023; Received in revised form 28 January 2024; Accepted 26 February 2024
Available online 1 March 2024

specific field phenotyping task (Lin et al., 2023; Nie et al., 2023). Among them, the RGB sensor is most widely utilized due to its capacity to produce high-spatial-resolution RGB images while being lightweight and relatively inexpensive (Liu et al., 2022). Extensive studies have successfully employed RGB sensors with UAVs to assess the phenotypes of above-ground plants, including canopy cover, crop height and stem density (Li et al., 2019; Liu et al., 2021b, 2022; Mhango et al., 2022).

Extracting clear and distinct individual plants from potato canopies is required before analyzing the potato plant phenotypes. However, direct segmentation of individual plants is challenging after canopy consolidation because potato plants are too occluded and highly similar in texture to be distinguished on a plant-by-plant basis. Instead, it is feasible to segment stems with their growing leaves as fundamental units that correlate with potato tuber yield (Mhango et al., 2022). Segmenting the leafy stems not only facilitates acquisition of individual plants but also help analyze phenotypes from an organ-level perspective. Potato stems grow leaf primordia and complete leaves, which have distinctive folding and color features that are beneficial in segmentation tasks. Besides, height mutation happens in outer edges of the leafy potato stems, which offers a credential for segmentation. Thus, both spatial and spectral features are important to segment the leafy potato stems.

A few researches are focusing on combination of spatial and spectral features. Point clouds are utilized to describe plant spatial structure and are mostly treated as the targets of mostly current organ-level segmentation methods including clustering or deep learning-based techniques (Malambo et al., 2019; Qiao et al., 2023). Point cloud-based methods focus on learning spatial features at the expense of spectral features, which requires different information about of the leafy potato stems. Band combinations such as RGB, crop height model (CHM), digital surface model (DSM), and vegetation indices were applied to describe other vegetations (Bittner et al., 2018; Dai et al., 2018; Hao et al., 2021; Li et al., 2022; Yang et al., 2021, 2022; Zheng et al., 2022). However, plants in the above studies do not suffer from the smaller more highly occluded stems found in the potato canopy. Thus, besides combination descriptions, small ground sampling distances (GSDs) are expected, requiring high-resolution sensors and a suitable flight altitude that is low enough while ensuring that plants are not affected by wind disturbance from UAV's wings (Hu et al., 2019).

Recent advancements in convolutional neural networks (CNN) have made notable progress in vegetation segmentation using band combinations. A mask region-based convolutional neural network (Mask R-CNN) was adopted and accepted band combination as input for tree segmentation, which achieved an Intersection over Union (IoU) of 91.27 % (Hao et al., 2021). Meanwhile, strawberry plant canopies from images of band combination of RGB and near-infrared were segmented using Mask R-CNN and obtained an IoU of 98.45 % (Zheng et al., 2022). U-Net was applied to segment wheat regions in combined RGB and DSM images with IoU of 88.99 % (Yang et al., 2021). All above mentioned approaches belong to CNN architecture, which has shown potential in feature extraction using band combination for vegetation segmentation. You look only once version eight extra-large (YOLOv8x) is a state-of-the art network based on CNN architecture and supports segmentation tasks, which has been mostly applied to the segmentation of three band images with red, green and blue.

An automated segmentation method of leafy potato stems was developed herein for the phenotyping of above-ground potato plants. YOLOv8x was applied to the orthoimages of band combinations to segment the leafy potato stems. Meanwhile GSD is considered as small as possible. In summary, our study has the following specific contributions: (1) A ratio of one for RGB and DSM is the most suitable band combination for segmentation of leafy potato stems; (2) Millimeter-level ground sampling distance satisfied segmentation of leafy potato stem. The remaining sections of this article are organized as follows. In Section 2, the materials and methods are described. In Section 3, the results are presented and discussed. Lastly, in Section 4, conclusions of this work are described.

## 2. Materials and methods

### 2.1. Experimental site

Our research was conducted at a trapezoidal field in Gaojia village, Wuquan town, Shaanxi province, China (34.292°N, 107.973°E), as shown in Fig. 1a. The field has a continental monsoon climate with an average annual precipitation of 649.5 mm and average annual temperature of 12.9 °C, which is suitable for the cultivation of potatoes. The elevation of the field ranges from 487.1 m to 493.6 m. Potatoes were planted on February 25th, 2023, with the variety Wotu 5. Plant patterns are as follows: ridge height is 20 cm, plant spacing is 30 cm, ridge spacing is 120 cm, and ridge width is 90 cm, in a double-row planting.

### 2.2. Data collection

The UAV imagery was collected in May 2023 using DJI M300 RTK (SZ DJI Technology Co., Shenzhen, China), an industrial-grade drone renowned for a sophisticated GNSS system that supports GPS, GLONASS, BeiDou, and Galile. The M300 RTK also features a Real-Time Kinematic (RTK) module that, when enabled and fixed, provides a positioning accuracy of 1 cm + 1 ppm horizontally and 1.5 cm + 1 ppm vertically, ensuring precise navigation and positioning. The Zenmuse P1 camera, equipped with a full-frame sensor (35.9 × 24 mm), a 50 mm lens with a field of view (FoV) of 46.8°, an aperture range of f/2.8 - f/16, and an ISO range of 100——25600, was used in conjunction with the drone. The aerial survey was conducted in the morning (10:30 AM to 12:00 PM) under clear sky and windless conditions. The flight altitude needs to be kept as low as possible while ensuring that plants are not disturbed visibly by wind from UAV's wings. Therefore, three altitudes were set before the formal experiment, namely 10, 15 and 20 m. Among them, 15 m was the most suitable flight altitude with a GSD of 0.19 cm per pixel. Flight patterns were programmed to achieve an 80 % forward overlap rate and 70 % side-lap rate. RTK positioning of UAVs was applied to obtain highly accurate UAV waypoint positions in flight. A total of 623 images with 8192 × 5460 pixels were obtained and saved in JPG image format. The images are available at https://github.com/fu3lab/Leafy_potato_stem_segmentation.

### 2.3. Data preprocessing and data extraction

Both DSM and CHM observe height mutation of the outer edge in leafy potato stems, as shown in Fig. 1d,e. Orthomosaic RGB images and DSMs of the site were generated using structure from motion (SfM) algorithm and DJI Terra software (Shenzhen, China). CHM was calculated based on Eq. (1), where DSM represents the absolute height of the object and digital elevation model (DEM) represents absolute height of bare ground under potato canopy. DEM was estimated by interpolating values extracted from the adjacent bare soil using Arcmap 10.2, and the binary kriging method (Li et al., 2020). The distances between interpolating points becomes a key factor in generating the DEM after the bare soil was obtained by the maximum likelihood classification method (Bolstad and Lillesand, 1991). Thus, three distances, i.e., 1 m, 0.1 m, and 0.01 m, were set to select the optimal one. In addition, the Excess green (ExG) index which is sensitive to green plants and is usually used for crop segmentation was implemented (Li et al., 2019). The ExG index was calculated according to Eq. (2).

$$\mathrm{CHM} = \mathrm{DSM} - \mathrm{DEM} \quad (1)$$

$$\mathrm{ExG} = \frac{2R - G - B}{R + G + B} \quad (2)$$

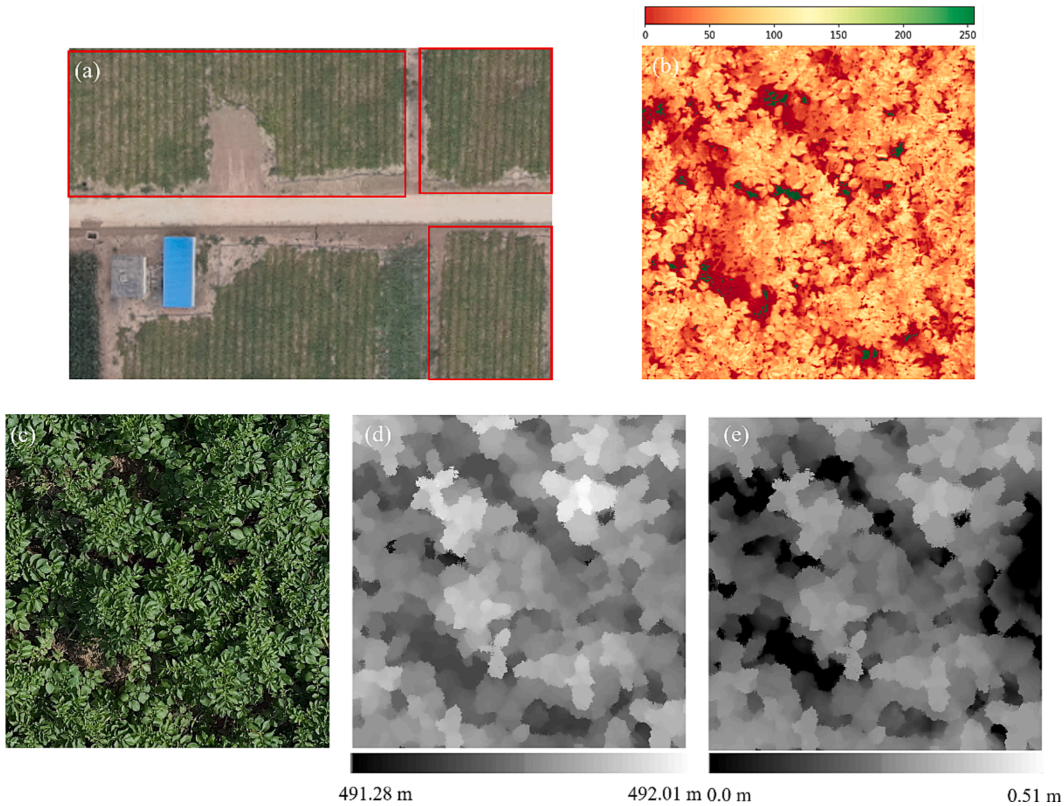


**Fig. 1.** UAV imaging area which drew in red rectangles (a) and data examples of ExG (b), RGB (c), DSM (d), and CHM (e).

### 2.4. Model construction

#### 2.4.1. Multiple channel input of band combinations

Bands are combined and fed as inputs to the YOLOv8x model. Five band combinations were selected, as shown in Table 1. RGB images are the most utilized input for plant segmentation models, and as such are used in every combination (Fernandez-Gallego et al., 2020; Ma et al., 2021; Majeed et al., 2020; Zheng et al., 2022). The ExG was combined with RGB to segment leafy potato stems from bare soil, while DSM or CHM were combined with RGB to provide spatial features in addition to spectral features. In the final combination, the DSM was added three times when combining with the RGB data to keep the balance of channel attention while RGB occupies three channels.

#### 2.4.2. Dataset preparation

Images of the combined bands were used to make five datasets before feeding the YOLOv8x model. The images of the whole site were cropped to size of 1280 × 1280 pixels with 10 % overlap. Leafy potato stems were annotated using AnyLabeling software (https://github.com/vietanhdev/anylabeling) with two kinds of labels, i.e., centerb (representing potato stems with leaf primordia) and branches (representing potato stems with complete leaves). The two kinds of labels exhibit distinct features, as illustrated in Fig. 2. The range of leafy potato stems were annotated based on DSM images, followed by classes given by RGB images. The datasets were split at a ratio of 4:1 for model training and testing. In addition, the training data were augmented by varying brightness (with coefficients of 0.8, 0.9, 1.1, and 1.2, respectively), saturation (with coefficients of 0.8 and 1.3, respectively), and contrast (with coefficients of 0.7 and 1.2, respectively), and rotating 90°, 180°, and 270° from the original orientation. This helps compensate for the limited number of training datasets while increasing model robustness. In summary, five datasets were obtained for this study each containing 7007 images.

**Table 1**
Band combinations of input images.

| Band combinations | Image layers | Layer descriptions |
|---|---|---|
| RGB | 3 | Red, Green, Blue |
| RGB-DSM | 4 | Red, Green, Blue, DSM |
| RGB-ExG | 4 | Red, Green, Blue, ExG |
| RGB-CHM | 4 | Red, Green, Blue, CHM |
| RGB-DSM × 3 | 6 | Red, Green, Blue, DSM, DSM, DSM |

#### 2.4.3. YOLOv8x training implement

The five datasets were used to train the YOLOv8x model. The models were trained on a Window 10 desktop computer running on an Intel (R) Xeon (R) Gold 5218 CPU and an Nvidia GeForce RTX 4090 GPU with 128 GB of RAM. PyTorch deep learning framework version 2.0.1 was utilized for implementation. The models were trained for a total of 200 epochs with batch size of four. The training process is conducted using the Adam optimizer with the linear learning rate decay strategy, which allows for efficient convergence. An initial learning rate was set to 0.01 and the initial model weights were randomized. A non-maximum suppression algorithm was used to remove the overlapping identification, where IoU threshold was set to 0.7. Confidence threshold was set to 0.1 to trust identification for validation.

### 2.5. Field survey data

The heights of leafy potato stems were manually measured to select suitable CHMs. The CHMs are influenced by DEM, which is determined by the distance between interpolating points. Therefore, a field survey was immediately carried out after the acquisition of the UAV imagery. Potatoes are known to form main and multiple secondary stems originating below the ground (Mhango et al., 2021). Stem heights were measured vertically from the part exposed from the soil to the top of the stem using a tape measure. A total of 40 leafy potato stems were picked randomly and their height of minimum and maximum are 20.7 cm and 47.2 cm, respectively. The distribution of stem heights from field


**Fig. 2.** An annotation example of leafy potato stems. Potato stems with complete leaves were shaped using green polygons and labeled as branches while potato stems with leaf primordia were drawn in red polygons and labeled as centerb.

measurements is as shown in Fig. 3.

### 2.6. Evaluation indicators

Stem height was evaluated to determine the optimal CHM. The *mean absolute percentage error* (*MAPE*) and the *rooted mean squared error* (*RMSE*) are used for error estimation between manual measured stem height and estimated stem height, which was defined in Eq. (3) and (4), respectively. Segmentation performance of the YOLOv8x model was evaluated using *precision* (*P*), *recall* (*R*), *F*1 *score*, and *IoU* (Hao et al., 2021).

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{4}$$

where $n$ is the number of leafy potato stems (40 in this study). $y_i$ is measured height manually of stem $i$. $\hat{y}_i$ is estimated height of stem $i$ by the CHM.

## 3. Results and discussion

### 3.1. Crop height estimation

Accurate CHMs qualify for comparison with DSMs in terms of spatial feature input to the YOLOv8x model. The CHMs (Fig. 4d-f) were generated based on DEMs (Fig. 4a-c), respectively. The DEMs interpolated by 0.01 m (Fig. 4a) and 0.1 m (Fig. 4b) reflects the shape of ridge while the DEM interpolated by 1 m (Fig. 4c) reflects rough terrain. Estimated crop height calculated by Eq. (1) was compared with ground truth data and showed that the CHM generated by 0.01 m interpolation had the lowest error with 18.7 % of *MAPE* and 7.89 cm of *RMSE* in Table 2. This is consistent to the estimated RMSE of 7.24 cm for potato height reported by Li et al. (2020).

The use of subtraction between DSM and DEM for out-of-ridge leafy potato stems causes the predicted crop height to be larger than the ground truth data, as shown in Fig. 5b. Estimating crop height at a plant level after canopy consolidation is difficult. Consequently, crop height estimation is performed at the plot level in previous studies, improving robustness as plant heights are averaged or maximized across the plot (Han et al., 2018; Iqbal et al., 2017). In this work, although we could obtain stem height on a plant level, the estimated height of out-of-ridge

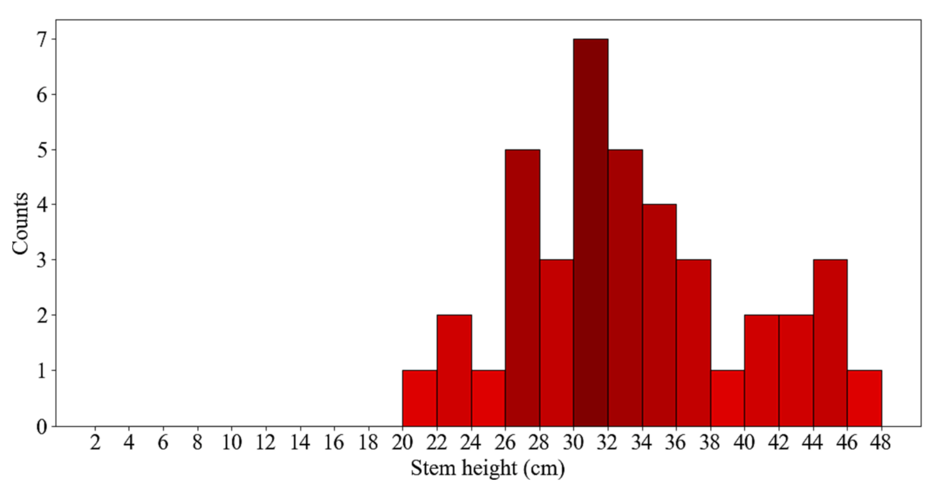


**Fig. 3.** The distribution of stem heights from field measurements.

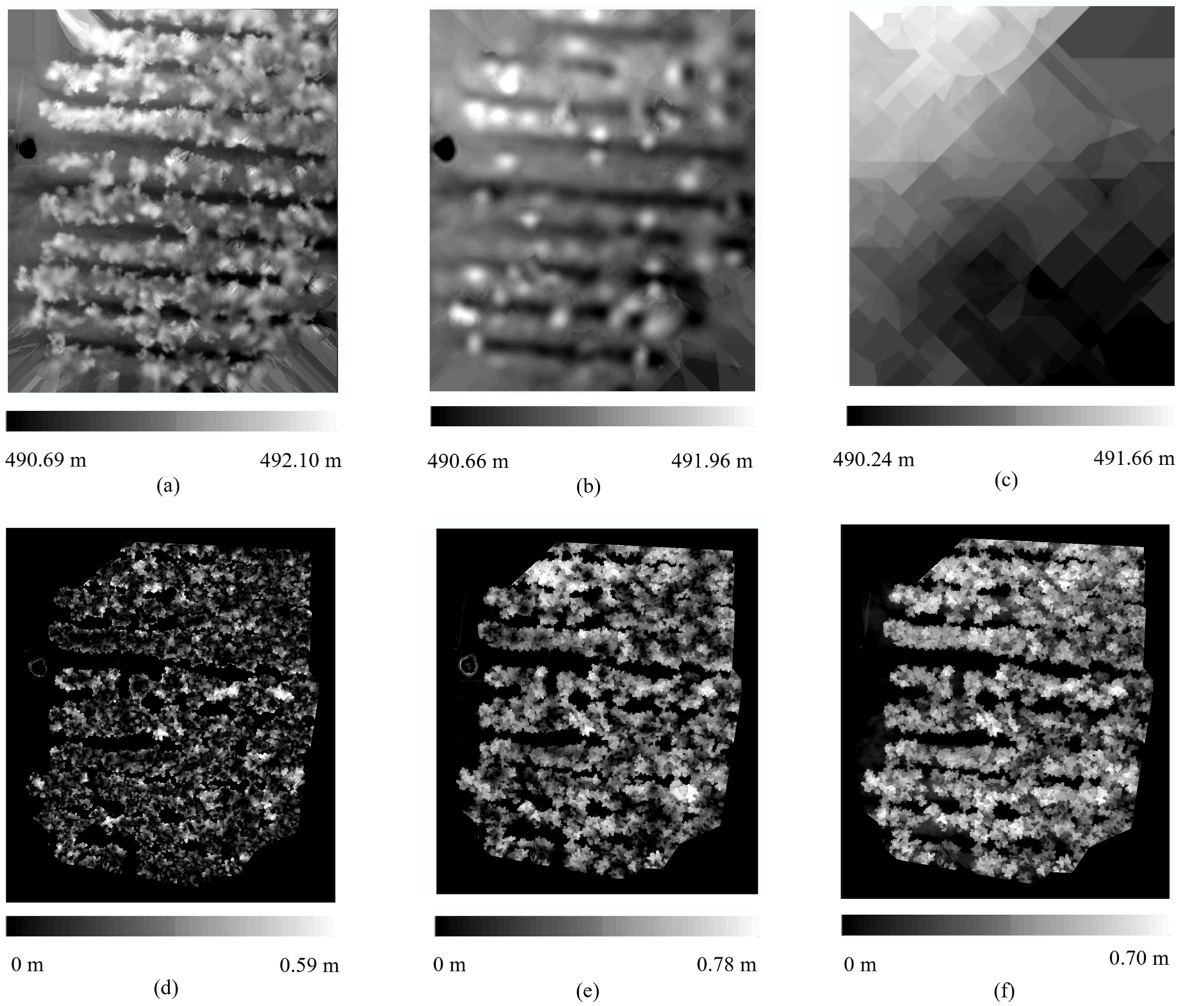


**Fig. 4.** Sample DEM images in interpolating distances of 0.01 m (a), 0.1 m (b), and 1 m (c) and their corresponding CHM images of 0.01 m (d), 0.1 m (e), and 1 m (f).

**Table 2**
CHM estimation results on different interpolating distances.

| Parameters | 1 m | 0.1 m | 0.01 m |
|---|---|---|---|
| *MAPE* (%) | 26.67 | 18.74 | 37.56 |
| *RMSE* (cm) | 10.88 | 7.89 | 15.74 |

leafy potato stems was larger than the ground truth. As shown in Fig. 5c, h1 + h2 is estimated height of out-of-ridge leafy potato stems while h1 is the ground truth, which makes h2 is the estimated error of stem height. This error only is specific to out-of-ridge leafy potato stems, which is helpful to obtain accurate stem height if the error can be diminished. Therefore, a correction factor $k_j$ was given to amend crop height estimation, as defined in Eq. (5). Based on our observation and experience, the factor was set as half of ridge height. The amended crop height estimation method showed lower *RMSE* of 5.19 cm compared with 7.89 cm of Fig. 5b. It is concluded that the correction factor $k_j$ is important for estimating height of crops grown on ridges such as potatoes on plant level.

$$CHM_j = DSM_j - DEM_j - k_j \times \mathrm{rh}/2 \quad (5)$$

where the value of $k_j = 1$ when pixel $j$ belongs to potato stems out of ridge, otherwise $k_j = 0$ and rh represents ridge height.

It is worth noting that DEMs require suitable number of interpolated points because their interpolating distances define the formation of the DEM. Interpolation methods were widely applied to generate DEMs while interpolating distances were simply utilized after considering image spatial resolution and computing resources, which is based on balancing of treating plots as a unit (Han et al., 2018; Iqbal et al., 2017; Schirrmann et al., 2017; Tilly et al., 2014). Further considerations on interpolating distances are needed for segmenting potatoes planted in ridges. CHMs generated from a dense or sparse number of interpolation points are not accurate enough to estimate crop height. The DEM generated with an interpolation distance of 0.01 m shaped the ridge but predicted the crop height poorly due to inaccurate DEM covered by vegetation.

### 3.2. *YOLOv8x model performance on segmentation of leafy potato stems*

Combination of spectral and spatial data showed the best segmentation performance of leafy potato stems, as shown in Table 3. Both RGB-DSM and RGB-CHM enhanced segmentation compared with only RGB. Although the segmentation performance of RGB-CHM is slightly lower than that of RGB-DSM, both supplied crucial spectral and spatial details, especially height mutation, necessary for differentiating leafy potato stems, as shown in Fig. 6. The input was overloaded with spatial data when 3x DSM data were included, leading to a decline in segmentation quality, which is not suitable to segment leafy potato stems. This suggested that the excessive spatial feature could overshadow the spectral data. Therefore, the ratio of DSM and RGB was set to 1 since segmentation performance of RGB-DSM is better than that of RGB-DSM × 3. Moreover, it was observed that the RGB-EXG yielded slightly better results than using RGB alone, indicating that EXG index amplifies the

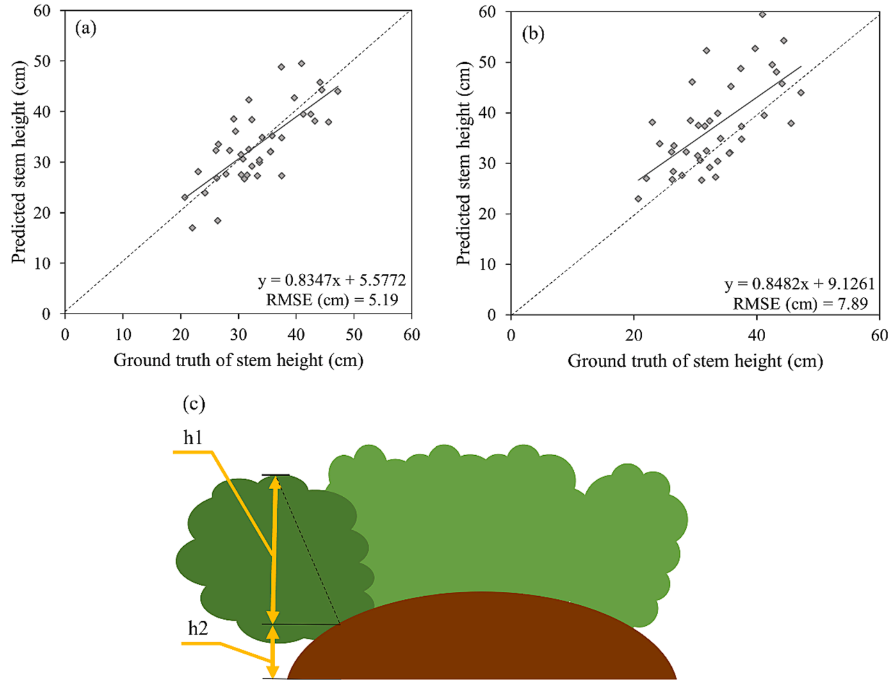


**Fig. 5.** Linear regressions of stem height between field measurement and CHM prediction interpolating with distance of 0.1 m. (a) $k_j \neq 0$; (b) $k_j = 0$. And a side view of potato crops (c).

**Table 3**
Segmentation performance on leafy potato stems using different combinations.

| Parameters | *P* | *R* | *IoU* | *F1 score* |
|---|---|---|---|---|
| RGB | 0.65 | 0.64 | 0.63 | 0.65 |
| RGB-DSM | **0.86** | **0.80** | **0.83** | **0.86** |
| RGB-CHM | 0.84 | 0.74 | 0.79 | 0.81 |
| RGB-ExG | 0.66 | 0.65 | 0.65 | 0.66 |
| RGB-DSM × 3 | 0.50 | 0.37 | 0.43 | 0.52 |

presence of green leafy potato stems, which is beneficial for distinguishing potato stems from soil. But it was negligible compared with RGB-DSM or RGB-CHM, which demonstrated that distinguishing potato stems from other stems is more important than from soil.

Height mutation at the edge of leafy potato stems plays a crucial role in improving segmentation. There was almost identical difference in terms of improved segmentation when adding DSM and CHM. DSM indicates the height of the surface from a reference point (usually sea level), while CHM indicates plant height from a reference point with ground level. Both showed the same information in terms of relative height although their absolute heights are different, which indicates height mutation well. This yielded good results on leafy potato stems after canopy consolidation, especially for intertwine upper and lower levels.

### 3.3. *Discussion*

#### 3.3.1. *Innovation of study*

An automated method was proposed for segmenting leafy potato stems on a fine scale. Existing studies have employed manual segmentation to deal with challenges posed by low-lying and dense crops that resemble potato leaf stems, such as their similar textures and high density (Liu et al., 2021a; Xiao et al., 2023). Therefore, the fusion of multimodal data is required to address these challenges. One approach is to utilize colored point clouds, which involves high costs, multiple camera calibration, and point effectiveness while the other approach is a two-stage process, where spectral segmentation is performed first, followed by projection onto the three-dimensional space (Dai et al., 2018; Gené-Mola et al., 2020). Focusing on potatoes, although combinations of spectral and spatial features have been applied to estimate above-ground biomass, the features were manually extracted to multiple stepwise regression (Liu et al., 2022). Compared with the above methods, an end-to-end method of deep learning was applied to the fuse the multimodal data to improve the segmentation of the leafy potato stems.

Existing literature focuses on fusing multimodal data using deep learning for crop classification, citrus segmentation, and crop segmentation (Fan and Lu, 2021; Niu et al., 2022; Osco et al., 2021). However, these approaches are either field crops covering a large area or high vegetation, which is different to leafy potato stems (individual and low vegetation in a finer scale). Thus, besides of information fusion, it is essential to consider the GSD for the leafy potato stems. Research has demonstrated that smaller GSD leads to better segmentation results, which are determined by the flight altitude and image resolution (Hu et al., 2019; ten Harkel et al., 2020). By leveraging deep learning and the small GSD, our method achieves improved segmentation results in leafy potato stems.

#### 3.3.2. *Potential application*

This work is helpful to high throughout phenotyping at a scale more granular than the typical plot level. In previous studies, automated measurements of individual crop phenotypes could only be replaced with average values of crop phenotypes in plots, or extracted manually (Li et al., 2020). Specifically, this method lays the groundwork for precise potato plant localization, which is the first step towards comprehensive plant phenotype (Mhango et al., 2021). Moreover, segmentation objects are categorized into potato stems with complete leaves and primordia. Potato stems with primordia are located at the center of the stem and serves as a proxy to stems for yield prediction, which was studied by Mhango et al. (2022). Besides potato stems with primordia, potato stems with complete leaves were extracted from our study, which is beneficial to analyzing potato phenotypes. The method also has the potential to be adaptable to other dense crops with leaves that are characterized by height mutation at the edges, which is common

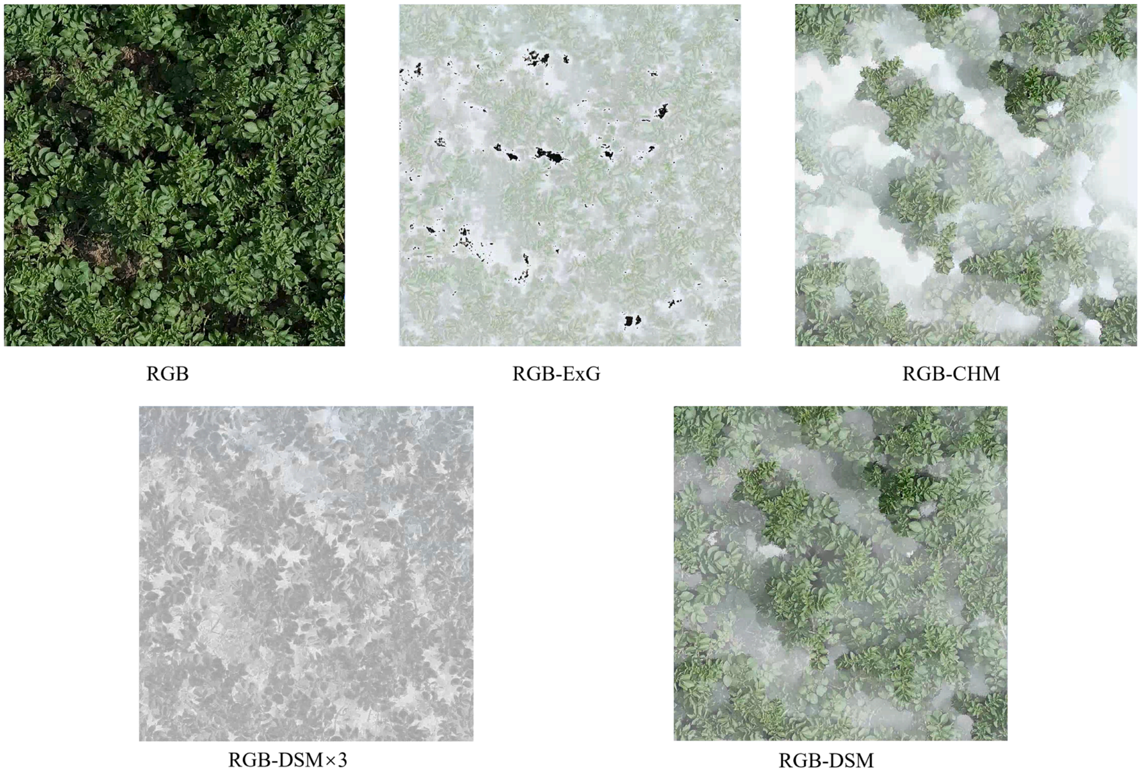


**Fig. 6.** Example images of band combinations, i.e., RGB, RGB-ExG, RGB-CHM, RGB-DSM × 3, and RGB-DSM.

in many field crops.

YOLOv8x is competent in segmenting leafy potato stems on the top of canopy. However, some potato stems located under the canopy were obscured in the images, resulting in an incomplete potato phenotype. A possible solution is to perform fitting algorithms based on stem growth and its shape feature. In addition, our study found that the mended factor is crucial for estimating crop height in ridge-planted crops like potatoes on plant level. Therefore, automating classification of potato stems outside the ridge is further needed. Moreover, GSD of 0.19 cm per pixel qualifies to segment the leafy potato stems with a flight altitude of 15 m and an RGB sensor captured with 8192 × 5460 resolution. The flight altitude and image resolution in our study are capable of segmenting potato stems while the relationship between flight altitude and image resolution can be further investigated considering the flight efficiency and cost.

## 4. Conclusions

This work presented a novel method for the segmentation of leafy potato stems using YOLOv8x with band combinations. YOLOv8x with spatial and spectral features was shown to be capable of stem segmentation. Height mutation is beneficial to segment the leafy potato stems or other dense crops with leaves. Moreover, it was found that 0.1 m is a suitable interpolating distance to generate accurate CHM. Millimeter-level ground sampling distance is promising to obtain high throughput phenotyping, which requires ultra-low flight altitude and high image resolution. The accuracy and efficiency of the YOLOv8x has great potential for guiding phenotypic automation of potatoes and other arable crops through remote sensing.

### CRediT authorship contribution statement

**Hanhui Jiang:** Writing – original draft, Methodology, Investigation, Conceptualization. **Bryan Gilbert Murengami:** Writing – review & editing, Methodology, Conceptualization. **Liguo Jiang:** Writing – review & editing, Methodology, Conceptualization. **Chi Chen:** Writing – review & editing, Methodology, Conceptualization. **Ciaran Johnson:** Conceptualization, Writing – review & editing. **Fernando Auat Cheein:** Conceptualization, Writing – review & editing. **Spyros Fountas:** Writing – review & editing, Supervision, Methodology. **Rui Li:** Writing – review & editing, Supervision, Methodology. **Longsheng Fu:** Writing – review & editing, Supervision, Methodology, Conceptualization.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data will be made available on request.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (32171897); Open Project of Key Laboratory of Agricultural Equipment for Hilly and Mountainous Areas in Southeastern China (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, China (QSKF2023002); National Foreign Expert Project, Ministry of Science and Technology, China (DL2022172003L, QN2022172006L).

### References

Bittner, K., Adam, F., Cui, S., Körner, M., Reinartz, P., 2018. Building footprint extraction from VHR remote sensing images combined with normalized DSMs using fused fully convolutional networks. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 11, 2615–2629. https://doi.org/10.1109/JSTARS.2018.2849363.

Bolstad, P.V., Lillesand, T.M., 1991. Rapid maximum likelihood classification. Photogramm. Eng. Remote Sens. 57, 67–74.

Dai, W., Yang, B., Dong, Z., Shaker, A., 2018. A new method for 3D individual tree extraction using multispectral airborne LiDAR point clouds. ISPRS J. Photogramm. Remote Sens. 144, 400–411. https://doi.org/10.1016/j.isprsjprs.2018.08.010.

Fan, C., Lu, R., 2021. UAV image crop classification based on deep learning with spatial and spectral features. In: IOP Conference Series: Earth and Environmental Science, Vol. 783, No. 1. IOP Publishing, p. 012080. https://doi.org/10.1088/1755-1315/783/1/012080.

Fernandez-Gallego, J.A., Lootens, P., Borra-Serrano, I., Derycke, V., Haesaert, G., Roldán-Ruiz, I., Araus, J.L., Kefauver, S.C., 2020. Automatic wheat ear counting using machine learning based on RGB UAV imagery. Plant J. 103, 1603–1613. https://doi.org/10.1111/tpj.14799.

Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J.R., Morros, J.R., Ruiz-Hidalgo, J., Vilaplana, V., Gregorio, E., 2020. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. Comput. Electron. Agric. 169, 105165 https://doi.org/10.1016/j.compag.2019.105165.

Han, L., Yang, G., Yang, H., Xu, B., Li, Z., Yang, X., 2018. Clustering field-based maize phenotyping of plant-height growth and canopy spectral dynamics using a UAV remote-sensing approach. Front. Plant Sci. 9, 1638. https://doi.org/10.3389/fpls.2018.01638.

Hao, Z., Lin, L., Post, C.J., Mikhailova, E.A., Li, M., Chen, Y., Yu, K., Liu, J., 2021. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (mask R-CNN). ISPRS J. Photogramm. Remote Sens. 178, 112–123. https://doi.org/10.1016/j.isprsjprs.2021.06.003.

Hu, P., Guo, W., Chapman, S.C., Guo, Y., Zheng, B., 2019. Pixel size of aerial imagery constrains the applications of unmanned aerial vehicle in crop breeding. ISPRS J. Photogramm. Remote Sens. 154, 1–9. https://doi.org/10.1016/j.isprsjprs.2019.05.008.

Iqbal, F., Lucieer, A., Barry, K., Wells, R., 2017. Poppy crop height and capsule volume estimation from a single UAS flight. Remote Sens. 9, 24–27. https://doi.org/10.3390/rs9070647.

Johnson, C.M., Auat Cheein, F., 2023. Machinery for potato harvesting: a state-of-the-art review. Front. Plant Sci. 14, 1156734. https://doi.org/10.3389/fpls.2023.1156734.

Li, Y., Deng, T., Fu, B., Lao, Z., Yang, W., He, H., Fan, D., He, W., Yao, Y., 2022. Evaluation of decision fusions for classifying karst wetland vegetation using one-class and multi-class CNN models with high-resolution UAV images. Remote Sens. 14 https://doi.org/10.3390/rs14225869.

Li, B., Xu, X., Han, J., Zhang, L., Bian, C., Jin, L., Liu, J., 2019. The estimation of crop emergence in potatoes by UAV RGB imagery. Plant Methods 15, 15. https://doi.org/10.1186/s13007-019-0399-7.

Li, B., Xu, X., Zhang, L., Han, J., Bian, C., Li, G., Liu, J., Jin, L., 2020. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. ISPRS J. Photogramm. Remote Sens. 162, 161–172. https://doi.org/10.1016/j.isprsjprs.2020.02.013.

Lin, Y., Li, S., Duan, S., Ye, Y., Li, B., Li, G., Lyv, D., Jin, L., Bian, C., Liu, J., 2023. Methodological evolution of potato yield prediction: a comprehensive review. Front. Plant Sci. 14, 1214006. https://doi.org/10.3389/fpls.2023.1214006.

Liu, F., Hu, P., Zheng, B., Duan, T., Zhu, B., Guo, Y., 2021a. A field-based high-throughput method for acquiring canopy architecture using unmanned aerial vehicle images. Agric. For. Meteorol. 296, 108231 https://doi.org/10.1016/j.agrformet.2020.108231.

Liu, Y., Huang, J., Sun, Q., Feng, H., Yang, G., Yang, F., 2021b. Estimation of plant height and above ground biomass of potato based on UAV digital image. Natl. Remote Sens. Bull. 25, 2004–2014. https://doi.org/10.11834/jrs.20210419.

Liu, Y., Feng, H., Yue, J., Li, Z., Yang, G., Song, X., Yang, X., Zhao, Y., 2022. Remote-sensing estimation of potato above-ground biomass based on spectral and spatial features extracted from high-definition digital camera images. Comput. Electron. Agric. 198, 107089 https://doi.org/10.1016/j.compag.2022.107089.

Ma, B., Du, J., Wang, L., Jiang, H., Zhou, M., 2021. Automatic branch detection of jujube trees based on 3D reconstruction for dormant pruning using the deep learning-based method. Comput. Electron. Agric. 190, 106484 https://doi.org/10.1016/j.compag.2021.106484.

Majeed, Y., Zhang, J., Zhang, X., Fu, L., Karkee, M., Zhang, Q., 2020. Deep learning based segmentation for automated training of apple trees on trellis wires. Comput. Electron. Agric. 170, 105277 https://doi.org/10.1016/j.compag.2020.105277.

Malambo, L., Popescu, S.C., Horne, D.W., Pugh, N.A., Rooney, W.L., 2019. Automated detection and measurement of individual sorghum panicles using density-based clustering of terrestrial lidar data. ISPRS J. Photogramm. Remote Sens. 149, 1–13. https://doi.org/10.1016/j.isprsjprs.2018.12.015.

Mhango, J.K., Harris, E.W., Green, R., Monaghan, J.M., 2021. Mapping potato plant density variation using aerial imagery and deep learning techniques for precision agriculture. Remote Sens. 13, 2705. https://doi.org/10.3390/rs13142705.

Mhango, J.K., Grove, I.G., Hartley, W., Harris, E.W., Monaghan, J.M., 2022. Applying colour-based feature extraction and transfer learning to develop a high throughput inference system for potato (*Solanum tuberosum* L.) stems with images from unmanned aerial vehicles after canopy consolidation. Precis. Agric. 23, 643–669. https://doi.org/10.1007/s11119-021-09853-4.

Nie, P., Qian, C., Qin, R., Deng, S., Sun, C., He, Y., 2023. Development status and trends of space-air-groound integrated information sensing and fusion technology. J. Intell. Agric. Mech. 4 (2), 1–11. https://doi.org/10.12398/j.issn.2096-7217.2023.02.001.

Niu, B., Feng, Q., Chen, B., Ou, C., Liu, Y., Yang, J., 2022. HSI-TransUNet: a transformer based semantic segmentation model for crop mapping from UAV hyperspectral imagery. Comput. Electron. Agric. 201, 107297 https://doi.org/10.1016/j.compag.2022.107297.

Osco, L.P., Nogueira, K., Marques Ramos, A.P., Faita Pinheiro, M.M., Furuya, D.E.G., Gonçalves, W.N., de Castro Jorge, L.A., Marcato Junior, J., dos Santos, J.A., 2021. Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery. Precis. Agric. 22, 1171–1188. https://doi.org/10.1007/s11119-020-09777-5.

Qiao, Y., Liao, Q., Zhang, M., Han, B., Peng, C., Huang, Z., Wang, S., Zhou, G., Xu, S., 2023. Point clouds segmentation of rapeseed siliques based on sparse-dense point clouds mapping. Front. Plant Sci. 14, 1188286. https://doi.org/10.3389/fpls.2023.1188286.

Schirrmann, M., Hamdorf, A., Giebel, A., Gleiniger, F., Pflanz, M., Dammer, K.H., 2017. Regression kriging for improving crop height models fusing ultra-sonic sensing with UAV imagery. Remote Sens. 9, 665. https://doi.org/10.3390/rs9070665.

Siebring, J., Valente, J., Franceschini, M.H.D., Kamp, J., Kooistra, L., 2019. Object-based image analysis applied to low altitude aerial imagery for potato plant trait retrieval and pathogen detection. Sensors 19, 5477. https://doi.org/10.3390/s19245477.

Silva-Díaz, C., Ramírez, D.A., Rinza, J., Ninanya, J., Loayza, H., Gómez, R., Anglin, N.L., Eyzaguirre, R., Quiroz, R., 2020. Radiation interception, conversion and partitioning efficiency in potato landraces: how far are we from the optimum? Plants 9, 787. https://doi.org/10.3390/plants9060787.

ten Harkel, J., Bartholomeus, H., Kooistra, L., 2020. Biomass and crop height estimation of different crops using UAV-based LiDAR. Remote Sens. 12, 17. https://doi.org/10.3390/RS12010017.

Tilly, N., Hoffmeister, D., Schiedung, H., Hütt, C., Brands, J., Bareth, G., 2014. Terrestrial laser scanning for plant height measurement and biomass estimation of maize, international archives of the photogrammetry, remote sensing and spatial. Inf. Sci. 181–187. https://doi.org/10.5194/isprsarchives-XL-7-181-2014.

UN Food & Agriculture Organization, 2023. Production of potatoes by the world. https://www.fao.org/faostat/en/#data. Accessed October 11, 2023.

Xiao, S., Ye, Y., Fei, S., Chen, H., Cai, Z., Che, Y., Wang, Q., Ghafoor, A., Bi, K., Shao, K., Wang, R., 2023. High-throughput calculation of organ-scale traits with reconstructed accurate 3D canopy structures using a UAV RGB camera with an advanced cross-circling oblique route. ISPRS J. Photogramm. Remote Sens. 201, 104–122. https://doi.org/10.1016/j.isprsjprs.2023.05.016.

Yang, R., Dai, Q., Cheng, H., Zhang, Y., Chen, N., Wang, L., 2022. Improving semantic segmentation performance by jointly using high resolution remote sensing image and NDSM. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 77–83. https://doi.org/10.5194/isprs-Annals-V-3-2022-77-2022.

Yang, B., Zhu, Y., Zhou, S., 2021. Accurate wheat lodging extraction from multi-channel uav images using a lightweight network model. Sensors 21, 6286. https://doi.org/10.3390/s21206826.

Zheng, C., Abd-Elrahman, A., Whitaker, V.M., Dalid, C., 2022. Deep learning for strawberry canopy delineation and biomass prediction from high-resolution images. Plant Phenomics 2022, 9850486. https://doi.org/10.34133/2022/9850486.