Title (English)	Estimating vegetation index for outdoor free-range pig production using YOLO
Running Title (English)	Estimating vegetation index for outdoor free-range pig production using YOLO
Author	Sang-Hyon OH [first_author] [1], Hee-Mun Park[2], Jin-Hyun Park[2]
Affiliation	[1] Division of Animal Science, College of Agriculture and Life Science, Gyeongsang National University, Jinju 52725, Korea, Republic of [2] School of Mechatronics Engineering, Engineering College of Convergence Technology, Gyeongsang National University, Jinju 52725, Korea, Republic of
Corresponding Author	Jin-Hyun Park (uabut@gnu.ac.kr, office: , mobile:)
ORCID	Sang-Hyon OH (https://orcid.org/0000-0002-9696-9638) Hee-Mun Park (https://orcid.org/0000-0001-5182-1739) Jin-Hyun Park (https://orcid.org/0000-0002-7966-0014)
Conflict of interest	No potential conflict of interest relevant to this article was reported.
Funding information	-
Acknowledgements	
Availability of data and material	
Author Contribution	Conceptualization: OH SH, Park JH Data curation: OH SH, Park JH Formal analysis: Park HM, Park JH Methodology: Park HM, Park JH Software: Park HM, Park JH Validation: OH SH, Park JH Investigation: OH SH, Park JH Writing - original draft: OH SH Writing - review & editing: OH SH, Park HM, Park JH
IRB/IACUC approval	The present experiment was reviewed and approved by the Institutional Animal Care and Use Committee of North Carolina A&T University (IACUC: 12-003.0).

Estimating vegetation index for outdoor free-range pig production using YOLO

Sang-Hyon OH[1†], Hee-Mun Park[2†] and Jin-Hyun Park[2*]

[1]Division of Animal Science, College of Agriculture and Life Science, Gyeongsang National University, Jinju 52725, South Korea

[2]School of Mechatronics Engineering, Engineering College of Convergence Technology, Gyeongsang National University, Jinju 52725, South Korea

[†] Both authors contributed equally to this manuscript.

***Corresponding author**: uabut@gnu.ac.kr

ORCID

Sang-Hyon Oh https://orcid.org/0000-0002-9696-9638

Hee-Mun Park https://orcid.org/0000-0001-5182-1739

Jin-Hyun Park https://orcid.org/0000-0002-7966-0014

Title of the manuscript: Estimating vegetation index for outdoor free-range pig production using YOLO

ABSTRACT

The objective of this study was to quantitatively estimate the level of grazing area damage in outdoor free-range pig production using a UAV with an RGB image sensor. Ten corn field images were captured by a UAV over approximately two weeks, during which gestating sows were allowed to graze freely on the corn field measuring 100×50 m^2. The images were corrected to a bird's-eye view, and then divided into 32 segments and sequentially inputted into the YOLOv4 detector to detect the corn images according to their condition. The 43 raw training images selected randomly out of 320 segmented images were flipped to create 86 images, and then these images were further augmented by rotating them in 5-degree increments to create a total of 6,192 images. The increased 6192 images are further augmented by applying three random color transformations to each image, resulting in 24,768 datasets. The occupancy rate of corn in the field was estimated efficiently using YOLO. As of the first day of observation (day 2), it was evident that almost all the corn had disappeared by the ninth day. When grazing 20 sows in a 50×100 m^2 cornfield (250 m^2/sow), it appears that the animals should be rotated to other grazing areas to protect the cover crop after at least five days. In agricultural technology, most of the research using machine and deep learning is related to the detection of fruits and pests, and research on other application fields is needed. In addition, large-scale image data collected by experts in the field are required as training data to apply deep learning. If the data required for deep learning is insufficient, a large number of data augmentation is required.

Keywords: outdoor, pig, production, vegetation index, image analysis

INTRODUCTION

Free-range outdoor pig production is steadily increasing in the United States and Europe due to the niche market strategy for small farmers, consumer antipathy to factory farm products, and the trends towards environmentally friendly and animal welfare practices. Ongoing research is also being conducted to support this trend [1-4].

One advantage of free-range outdoor pig production is that it can be operated with a small capital investment. However, one of the disadvantages is that the soil can become depleted due to the natural rooting behavior of pigs. If not appropriately managed, it can lead to groundwater eutrophication. Accordingly, the USDA (United States Department of Agriculture) requires that outdoor free-range pig production systems have at least 75% of the outdoor area covered in vegetative cover, such as crops or grass [5], which is to help prevent soil erosion, improve soil quality, and reduce the risk of nutrient runoff into nearby water sources.

As farmers who cannot accurately calculate the area covered by crops or vegetation may resort to using a sacrifice area to maintain the required 75% vegetative cover, it is not uncommon for pigs to be concentrated in a small area of the outdoor space while the rest is left unused. However, this can lead to overgrazing and soil damage in that area, increasing the risk of groundwater contamination from waste products. Therefore, it is important for farmers to implement good management practices, such as rotational grazing to minimize the environmental impact of outdoor pig production [6-8].

With the advancement of technology and science in recent years, photographing on Unmanned Aerial Vehicles (UAV) is no longer a difficult and expensive task [9]. If this technology were applied to outdoor free-range pig production to monitor the condition of grazing areas, it would greatly help producers maintain grazing areas at recommended levels without leaving them to degrade beyond repair. It may also be possible to estimate how much grass a pig has consumed in a particular grazing area by comparing the color changes in the captured images and the amount of pre- collected dry matter. This is particularly useful as it can be challenging to gauge the amount of grass consumed by pigs as they may cause damage to the grazing area.

YOLO (You Only Look Once) is an object detection technique utilizing deep learning in images and was proposed by Redmon et al. [10], which is a system that can recognizes the objects in an image and their locations at once, meaning it only needs to look at the image once. Compared to the classifier-based approach of CNN (Convolutional Neural Network), YOLO's network architecture is relatively simple as it directly learns the loss function that has a significant impact on detection performance. YOLO also has the ability to perform real-time object detection, which has been widely used in many research areas [11-14]. Figure 1 shows the schematic structure of the YOLOv4 object detection system.

The objective of this study was to develop an algorithm to quantitatively predict the extent of damaged grazing area in outdoor free-range pig production using a UAV with an RGB image sensor.

MATERIALS AND METHODS

Animal care

The present experiment was reviewed and approved by the Institutional Animal Care and Use Committee of North Carolina A&T University (IACUC: 12-003.0).

Animals, diets, and experimental design

The images used for the analysis were taken at a swine unit located within the University Farm of North Carolina A&T State University (Greensboro, NC, USA; 36°4′16.63″N, 79°43′33.02″E). A 50×100 m^2 grazing area was established for twenty pregnant sows that were allowed to graze pasture two weeks prior to their expected delivery date. The grazing area was planted with corn crops. The climate in this location is classified as humid subtropical climate (Köppen climate classification), with hot and humid summers and mild winters. The average annual precipitation is around 107 cm. The sows were given access to slightly less than standard National Research Council balanced rations (2-3kg/day) considering the consumption of corn in the pasture, but water *ad libitum*.

Data collection

The UAV used in this study is the Phantom 2 Vision model manufactured by DJI® with a quad-rotor system consisting of four propellers. Including a camera, the maximum takeoff weight is 1.3 kg, and it can fly for about 25 minutes using a 5,200 mAh lithium polymer battery (Table 1). It has a remote-control range of up to 300 m and is equipped with a high-resolution camera sensor of 14 Megapixels and 1/2.3″ size, with a fixed-focus wide-angle lens of 120° FOV and a focal length of 28 mm. It is equipped with an automatic flight control device, and a 2.4 GHz wireless remote controller was used for takeoff and landing as well as manual control of the aircraft.

Ten aerial images were taken using the UAV from a height that allowed the entire grazing area to be captured in a single image, from September 1st to September 13th, 2015, excluding days with rain. Also, the images were captured around 10:00 AM without additional lighting, with an effort made to minimize the effect of shadows caused by the sun. We tried to maintain the same altitude and position using the GPS attached to the UAV. Figure 2 shows the images captured by the UAV over two weeks after releasing the pigs. Each image has a size of 4,384 × 3,288 pixels.

Image analysis

This study aims to use only ten images captured by the UAV over a two-week period to numerically represent the process of cornfield degradation caused by gestation sows, using the degree of corn occupancy. Therefore, data augmentation is essential to utilize a small number of image data for deep learning. Data augmentation should be designed with consideration for the characteristics of images captured by the UAV. The YOLO network, which is one of the deep learning algorithms, was used with the augmented data to predict the occupancy level of cornfield in the images.

Correcting training images

The cornfield images in Figure 2 show two types of distortion. The first distortion is a convex fish-eye image caused by the wide-angle lens of the camera. The second distortion is due to the camera not being able to capture the cornfield at the exact center position and height, resulting in unequal sizes of the cornfield on the left and right sides. Therefore, it was necessary to correct for the distortions to accurately compare the extent of corn occupancy in the ten images. The fish-eye distortion was corrected using the method proposed by Scaramuzza [15].

By the way, the external and internal parameters of the camera had to be obtained to connect 3D world coordinates to a 2D image. World coordinate points were selected in the distorted fish-eye image and converted into camera coordinates using the external parameters. The camera coordinates are then mapped onto the image plane using the internal parameters. The distorted images captured from an inaccurate position and height of the UAV were solved by converting them into bird-eye views which are created using inverse perspective mapping to generate a 2D image of the scene. Figure 3 represents the process of correcting the image. Figure 3(a) shows the distorted original image, Figure 3(b) is an example of converting the fish-eye image into an undistorted image, and Figure 3(c) represents the result of correction using the bird's eye view with the first corrected image (Figure 3(b)). However, it was difficult to achieve perfect image correction due to the uncertainty of the camera's internal and external parameters. The corrected images were cropped to a resolution of 3,520×1,760 pixels to facilitate image comparison.

Training data

Ten corrected images are very insufficient to train a deep learning network. Deep learning systems based on deep artificial neural networks are highly dependent on the number of training data for their performance. The large number of training data prevents overfitting of prediction performance, and improves the generalization capability of the model, thereby improving object detection performance. Geometric methods, such as flipping and rotating images, and color adjustment methods are the most commonly used techniques for data augmentation in deep learning systems [16-17].

Although the number of images obtained through aerial photography is very small, the image resolution is still very high at 3,584×1,792 pixels even after image correction. If a high-resolution image is input to the deep learning network, the number of input parameters of the network increases, requiring a very long training and processing time. In addition, despite the high resolution of the images, the corn plants, which are our object of interest, have very small pixel sizes, making it very difficult to select the objects accurately. Therefore, it is useful to divide the high-resolution images into appropriate sizes for network training, and then reassemble the network's results for the segmented images for further processing. Therefore, the ten corrected images were segmented into sizes suitable for deep learning in this study.

The actual size of the experimental subject, the corn field, is 100×50 m^2. Therefore, it was divided into eight parts horizontally and four parts vertically at intervals of 12.5m, resulting in 32 images with a resolution of 448×448 pixels, as shown in Figure 4. Figure 5 shows 43 raw training images selected randomly out of 320 segmented images, each with different degrees of corn devastation.

Data labels

Data labels are required for training deep learning networks. The images of the cornfield were labeled into three categories based on the state of the corn: $Corn_I$, $Corn_D$, $Corn_S$.

$Corn_I$ refers to the preserved state of corn that had not been eaten or damaged by sows. This state is characterized by a clear green color of the corn, without any bending caused by sow movement. $Corn_D$ refers to the state of corn that had been damaged by sows, with corn lying at an angle or in a withering state. $Corn_S$ refers to the severely damaged state of corn where sows had almost completely eaten the corn, leaving only the cob. Table 2 defines these three labels.

The raw training images were converted into data using the three defined labels based on the state and size of the corn, as determined by human observation. The sample in Table 2 shows an example of the 43 images. The definition for each labeled bounding box is as shown in Equation (1).

$$Box_{ij}(Bx_{ij}, By_{ij}, Bw_{ij}, Bh_{ij}) \tag{1}$$

where, $\boldsymbol{i}$ denotes the label index, $\boldsymbol{j}$ denotes the bounding box number, (Bx_{ij}, By_{ij}) represents the coordinates of the top-left corner of the bounding box, and (Bw_{ij}, Bh_{ij}) represents the width and height of the bounding box.

Data augmentation

Data augmentation is a method of increasing the size of a dataset by generating new data that reflects the characteristics of the original data, especially in cases where the original dataset is limited. Although we have created 43 basic datasets for image segmentation and data labeling, it is still a very small number for training deep learning networks. Images obtained from the UAV are particularly advantageous for data augmentation techniques such as rotating or flipping images to increase data. In general, small angles are commonly used when performing data augmentation by rotation transformation. For example, an image of a person rotated by 180 degrees is not needed as a training image. On the other hand, it is irrelevant even if the image is rotated by 180° for corn images captured by UAVs. Furthermore, flipped images (both horizontally and vertically) can also be used as training images.

To effectively increase the number of training images, the 43 original data images were flipped to create 86 images, and then these images were further augmented by rotating them in 5-degree increments to create a total of 6,192 images. The increased 6192 images are further augmented by applying three random color transformations to each image, resulting in a total of 24,768 datasets. Figure 5 represents this process described above.

YOLOv4 Object Detection and Network Training

Figure 7 shows the YOLOv4 object detector used in this study to recognize the degree of corn devastation. In this study, ResNet50 was used as a backbone for detecting object characteristics, and SPP (Spatial Pyramid Pooling) and PANet (Path Augmented Network) were applied to the neck. The head was the same as YOLOv3. The output of the head represents the position and size of the bounding box, the probability of confidence score on the object, and the probability of class. The final output of YOLOv4 selects the final bounding boxes by applying the output values of the head and the non-maximum suppression (NMS). The input of the YOLOv4 detector is an image with 448×448 pixels. After correction, the image is divided into 32 (4×8) segments and inputted into the YOLOv4 detector. YOLOv4 extracts the features of the corn image when the image is inputted, and outputs the position and size of the corn image as well as the probability of each class.

The YOLOv4 network used in this study was provided by Matlab® [18], and the backbone network was changed by modifying the input layer of the network to match the augmented dataset using ResNet50. For the training parameters, the initial training rate was set to 0.001, and Adam Optimizer was used for the training method. The network was trained for a maximum of 30 epochs with a mini-batch size of 32. The hardware used in the study includes the Intel i9-12900 central processing unit and NVIDIA RTX-A6000 graphics accelerator.

The YOLOv4 network uses anchor boxes with specific heights and widths of predefined bounding boxes to improve the efficiency and object detection performance of the network, which also has a significant impact on training time. To determine the number of specific bounding boxes, the average IoU (Intersection over Union) value was calculated for all the bounding boxes in the prepared dataset, and the optimal value was selected. Figure 8 shows the average IoU value for all bounding boxes in the dataset by the number of specific boxes. The average IoU value is high at 0.86 when the number of specific boxes is 4.

The total loss function used for training the YOLOv4 network is equation (2), where the object classification loss and object confidence loss are computed using binary cross-entropy, and the bounding box localization error is computed using the Root Mean Square Error (RMSE).

$$TotalLoss = \mathrm{a} \times clsloss + \mathrm{b} \times objloss + \mathrm{c} \times boxloss \quad (2)$$

where, $[a, b, c] = [1, 1, 1]$ represents the weights for each loss term, where $clsloss$ is the object classification loss, $objloss$ is the object confidence loss, and $boxloss$ is the bounding box localization error.

The YOLOv4 network reached a RMSE of 0.21 after 30 epochs of training. Figure 8 shows some of the results of YOLOv4 network after training on 24,768 images.

Estimating the distribution and occupancy of corn

The proposed system aims to estimate the distribution and occupancy of corn for a specific date using a YOLOv4 network trained on a dataset of 24,768 images generated through data augmentation. Figure 10 shows an overview of the proposed overall system using a corn image for a specific date. As the input of the trained YOLOv4 detector is a 448×448 pixel image, the captured image for a specific date is corrected to 3,584×1,792 pixels and then sequentially inputted into the YOLOv4 detector by dividing it into 32 (4×8) segments. When an image is inputted into YOLOv4, it extracts features of the corn image and calculates the location and size of the corn in the image, as well as the probability of an object existing and the class probability. Objects with a probability of existence and a class probability above a certain threshold are selected for bounding boxes by NMS. For each segmented image, the number and area of labels detected by YOLOv4 are accumulated and calculated. As YOLOv4 outputs the location and size of corn in the image, the occupancy rate is calculated using Equation (3) by setting weights based on the three states of corn.

$$Occupancy_i = \left(\begin{array}{c} w_1 \times \sum_{j=1}^{32} \sum ACorn_{I,ij} + w_2 \times \sum_{j=1}^{32} \sum ACorn_{D,ij} \\ + w_3 \times \sum_{j=1}^{32} \sum ACorn_{S,ij} \end{array} \right) / Max_{ACorn} \qquad (3)$$

where, $i = 1,2,\ldots,10$ represents the index of the corn field image and $j = 1,2,\ldots,32$ represents the segmented image. w_1, w_2, w_3 are the weights assigned to each state of corn. $ACorn_{I,ij}$ represents the area of $Corn_I$, $ACorn_{D,ij}$ represents the area of $Corn_D$, and $ACorn_{S,ij}$ represents the area of $Corn_S$. Max_{ACorn} represents the maximum area of corn occupancy.

RESULTS

Figure 11 shows the detection results of the images captured 10 times in chronological order. Figure 11(a) shows the number and total area of intact corn objects represented by $Corn_I$ without being damaged by sows. It can be seen that it decreases exponentially over time. Figure 11(b) represents the number and total area of corn plants $Corn_D$. It can be seen that it linearly increases until the fourth day, and then decreases afterward. Figure 11(c) represents the number and total area of corn plants $Corn_S$. It can be seen that it sharply increases until the fourth day and gradually decreases afterward similar to the results of $Corn_D$. Figure 11(d) represents the occupancy rate of corn plants calculated using Equation (3). The weights for corn plant conditions were set as $[w_1, w_2, w_3] = [1,\ 0.5,\ 0.2]$, and the date with the largest area of land was set as the second day because no image was taken on the first day when the sow was released into the pasture. It can be seen that the occupancy rate of corn plants decreases very rapidly over time.

As a result, the occupancy rate of corn in the field was estimated efficiently using YOLO. As of the first day of observation (day 2), it was evident that almost all the corn had disappeared by the ninth day. When grazing 20 sows in a 50×100 m^2 cornfield (250 m^2/sow), it appears that the animals should be rotated to other grazing areas to protect the cover crop after at least five days.

DISCUSSION

YOLO object detection system

The input image was divided into grid cells through CNN, and objects are detected by generating anchor boxes and class probabilities for each cell section to predict the object's location and size [19]. Anchor boxes are boundary boxes with predefined height and width, and they are much faster than other detection systems because they do not use a separate network to extract candidate regions, unlike two-stage detectors. The YOLO object detection system has been improved by many researchers, and YOLOv4 demonstrates faster and more accurate detection rates among various versions by incorporating state-of-the-art deep learning techniques such as Weighted Residual Connections (WRC), Cross Stage Partial Connections (CSP), and the Complete Intersection over Union (CioU) loss [20]. The YOLOv4 network consists of a backbone network and a neck to detect object features, and the head outputs the object's position, the probability of being on the object, and the class probabilities. The final objects were detected by applying this.

Recently, image and video processing techniques have been widely applied in various fields, especially in the field of computer vision, where there has been significant research on image classification, object detection, and multiple object detection within images. As a classical image processing method, the image processing-based approach classifies and recognizes objects based on their direct features such as color, texture, and edges. This approach often results in significantly different output in object recognition within images due to lighting conditions, shadows, and camera settings.

For several years, object detection research in image recognition using machine and deep learning techniques has demonstrated significant advantages in computer vision tasks, resulting in significant improvements in object detection and recognition performance compared to traditional approaches [21]. This progress has been made possible by the utilization of big data, advances in high-performance hardware such as Graphic Processing Units (GPUs), and the development of useful learning algorithms for deep learning training, which has led to the evolution of practical and useful technologies.

The CNN(Convolutional Neural Network) is the most widely used deep learning algorithm for object detection research and was developed by LeCun in the late 1990s, which has a very high accuracy in object detection compared to traditional image processing methods

[22]. In addition to CNN, YOLO is widely used in object detection research due to its fast-processing time and high accuracy. Many studies have been conducted on YOLO in object recognition [10]. However, CNN requires algorithms such as Region-CNN (RCNN) to recognize the exact location of objects within an image in addition to object detection [23]. However, while RCNN has improved the accuracy of object detection, it requires a lot of computational time compared to traditional image processing methods and has an extremely high complexity of network training and algorithm.

On the other hand, YOLO has fast object detection and high accuracy. Machine and deep learning-based farming technologies are mainly applied for fruit detection and ripeness classification, as well as predicting pests and diseases in fruits [19]. In early machine learning research, Quiang et al. [24] identified fruits and tree branches using an SVM (Support Vector Machine) trained in the RGB color space. While it showed superior performance compared to previous threshold-based methods, it is still heavily affected by lighting conditions. Zhao et al. [25] applied a combination of the AdaBoost classifier and color analysis for tomato detection, but real-time processing was difficult due to the slow processing speed. Luo et al. [26] also suggested an AdaBoost and color feature-based framework for grapefruit detection, but it was affected by weather conditions and changes in lighting such as leaf covering.

Traditional machine learning research has greatly improved image processing-based methods, but the design of proposed methods is complicated and only adaptable to some specific conditions, resulting in poor flexibility. Deep learning has overcome the limitations of traditional machine learning by being more abstract and generalizable, particularly through the use of CNNs. Additionally, the utilization of big data has made it possible to apply these technologies to a range of agricultural problems, including image processing. Sa et al. [27] applied Faster R-CNN [28] to RGB and near-infrared images for fruit detection and showed better performance than previous methods. Mota-Delfin et al. [11] used YOLO to detect corn effectively in a weed-rich background using images captured by RPAS (Remotely Piloted Aerial Systems) and predicted the yield [10].

The basic data augmentation techniques include image processing methods that preserve the characteristics of the original image while maintaining diverse features of the objects. There are image processing techniques such as flipping, rotating, cropping images, and adjusting their brightness and color using various methods [16].

CONCLUSION

In agricultural technology, most of the research using machine and deep learning is related to the detection of fruits and pests, and research on other application fields is needed. In addition, large-scale image data collected by experts in the field are required as training data to apply deep learning. However, collecting training data takes a lot of time and effort. If there are few images for training, the effort and time for acquiring training images can be reduced while increasing training images through image segmentation and data augmentation (flip, rotation, brightness, color adjustment conversion) as in the proposed method. In addition, calculating the occupancy level of the whole image after calculating the occupancy level of each segmented image, as in the proposed method, is very effective. It is an excellent and effective technique to classify the status of corn ($Corn_I$, $Corn_D$, $Corn_S$) by date using the YOLO network. Therefore, the proposed method can be easily applied to many other fields and guarantees high precision.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

REFERENCES

1. Jang JC, Oh SH. Management factors affecting gestating sows' welfare in group housing systems - A review. Anim Biosci 2022;35(12): 1817-1826.

2. Choi W, Nassif N, Whitley NC, Oh SH. Comparison of temperature susceptibility for three types of outdoor farrowing huts. Applied Engineering in Agriculture 2014;30(2): 241-247.

3. Park H, Min B, Oh SH. Research trends in outdoor pig production. Animal Bioscience 2017a;30(9):1207-1214.

4. Park H, Oh SH. Seasonal variation in growth of Berkshire pigs in alternative production systems. Animal Bioscience 2017b;30(5):749-754.

5. NRCS. 2007. Conservation planning guidelines for outdoor swine operations.

6. Pietrosemoli S, Green J, Bordeaux C, Menius L, Curtis J. 2012. Conservation practices in outdoor hog production systems: Findings and recommendations from the Center for Environmental Farming Systems. Center For Environmental Farming Systems, North Carolina State University.

7. Whitley N, Hanson D, Morrow WEM, See MT, Oh SH. Comparison of pork quality and sensory characteristics for antibiotic free Yorkshire crossbreds raised in hoop houses. Animal Bioscience 2012a;25(11):1634-1640.

8. Whitley N, Morrow WEM, See MT, Oh SH. 2012b. Comparison of growth performance in antibiotic-free Yorkshire crossbreds sired by Berkshire, Large Black, and Tamworth breeds raised in hoop structures. Animal Bioscience 2012b;25(10):1351-1356.

9. Lee JM, Lee YH, Choi NK, Park H, Kim HC. Deep-Learning-based plant anomaly detection using a drone. Journal of the Semiconductor & Display Technology 2021;20(1):94-98.

10. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016;779-788. 10.1109/CVPR.2016.91

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

12. Du J. Understanding of object detection based on CNN family and YOLO. J Physics: Conference Series. 2018. IOP Publishing.

13. Viswanatha V, Chandana R, Ramachandra A. Real time object detection system with YOLO and CNN Models: A Review. J Xi'an Univ Archit Technol. 2022;14(7):144-151.

14. Tabelini L, Berriel R, Paixao TM, Badue C, De Souza AF, Oliveira-Santos T. Keep your eyes on the lane: Real-time attention-guided lane detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

15. Scaramuzza D, Martinelli A, Siegwart R. A toolbox for easily calibrating omnidirectional cameras. Proceeding to IEEE International Conference on Intelligent Robots and Systems. IEEE. 2006.

16. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. Journal of big data, 2019;6(1):1-48.

17. Taylor L, Nitschke G. Improving deep learning with generic data augmentation. in 2018 IEEE Symposium Series on Computational Intelligence (SSCI). 2018.

18. Works TM. Lidar object detection using complex-YOLO v4 Network. 2022.

19. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture. 2018;147:70-90.

20. Bochkovskiy A, Wang CY, Liao HYM. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.

21. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018;7068349.

22. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998;86(11):2278-2324.

23. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 2014;580-587. 10.1109/CVPR.2014.81

24. Qiang L, Jianrong C, Bin L, Lie D, Yajing Z. Identification of fruit and branch in natural scenes for citrus harvesting robot using machine vision and support vector machine. Int J

Agric Biol Eng. 2014;7(2):115-121.

25. Zhao Y, Gong L, Zhou B, Huang Y, Liu C. Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis. Biosyst Eng. 2016;148:127-137.

26. Luo L, Tang Y, Zou X, Wang C, Zhang P, Feng W. Robust grape cluster detection in a vineyard by combining the AdaBoost framework and multiple color components. Sensors. 2016;16(12):2098.

27. Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C. Deepfruits: A fruit detection system using deep neural networks. Sensors. 2016;16(8):1222.

28. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;39(6):1137-1149. 10.1109/TPAMI.2016.2577031 28.

Table 1. Specifications of the UAV platform used in the study.

Airframe	DJI Phantom 2 Quad-rotor
Dimensions	43.2 x 20.6 x 31.75 cm
Battery	3S LiPo 5200mAh, 11.1V
Takeoff Weight	≤1300g
Maximum Flight Time	25min (approx.)
Signal Frequency	2.4GHz ISM
Diagonal Length	350mm

Table 2. Data labels

Label	Label index	Color	Corn description	Sample
$Corn_I$	0	Blue	Intact corn	
$Corn_D$	1	Yellow	Damaged corn	
$Corn_S$	2	Red	Corn with stubble	

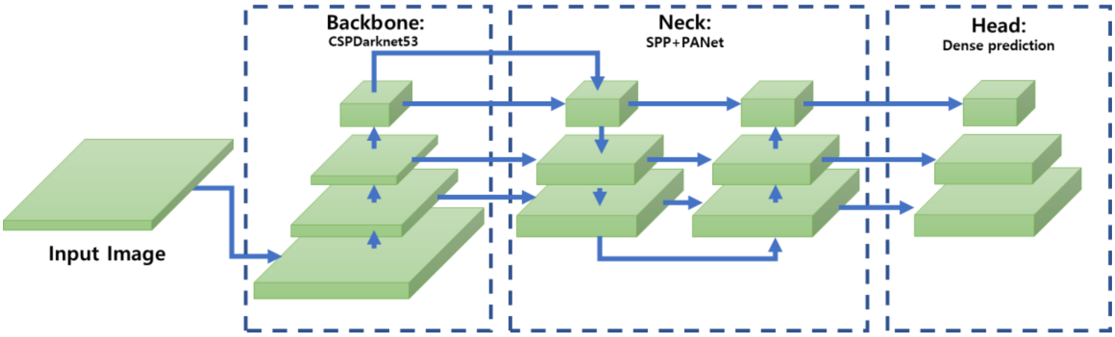


Figure 1. The YOLO Detection System

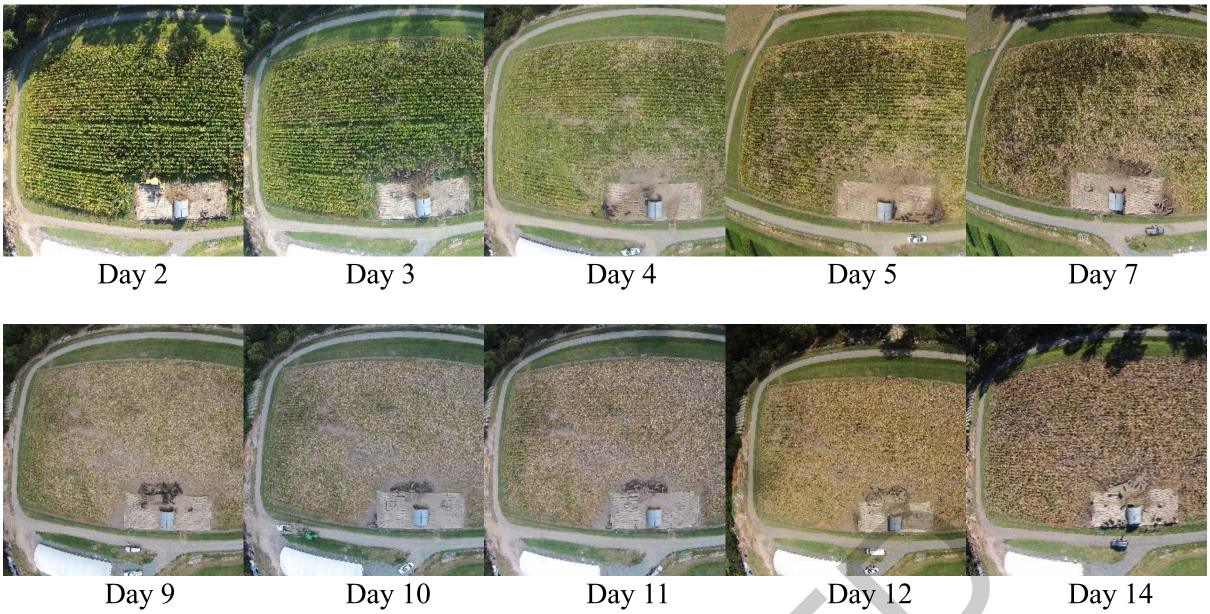


Figure 2. Original images used for the analysis

(a) Distorted image (b) Undistorted image (c) Bird's eye view and cropping

Figure 3. Example of image correction

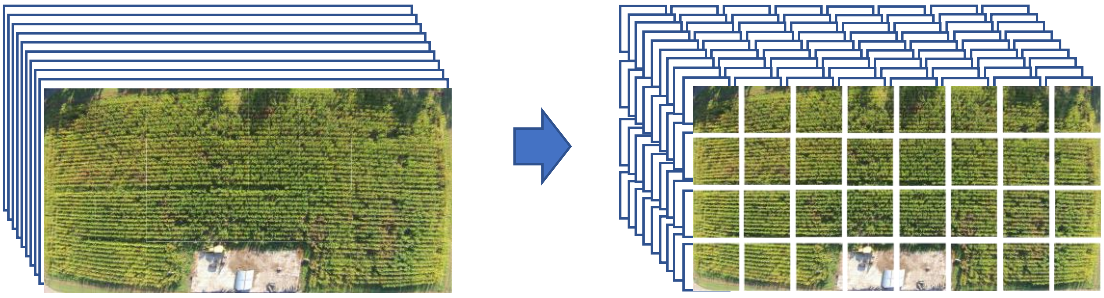


Figure 4. 4×8 split images

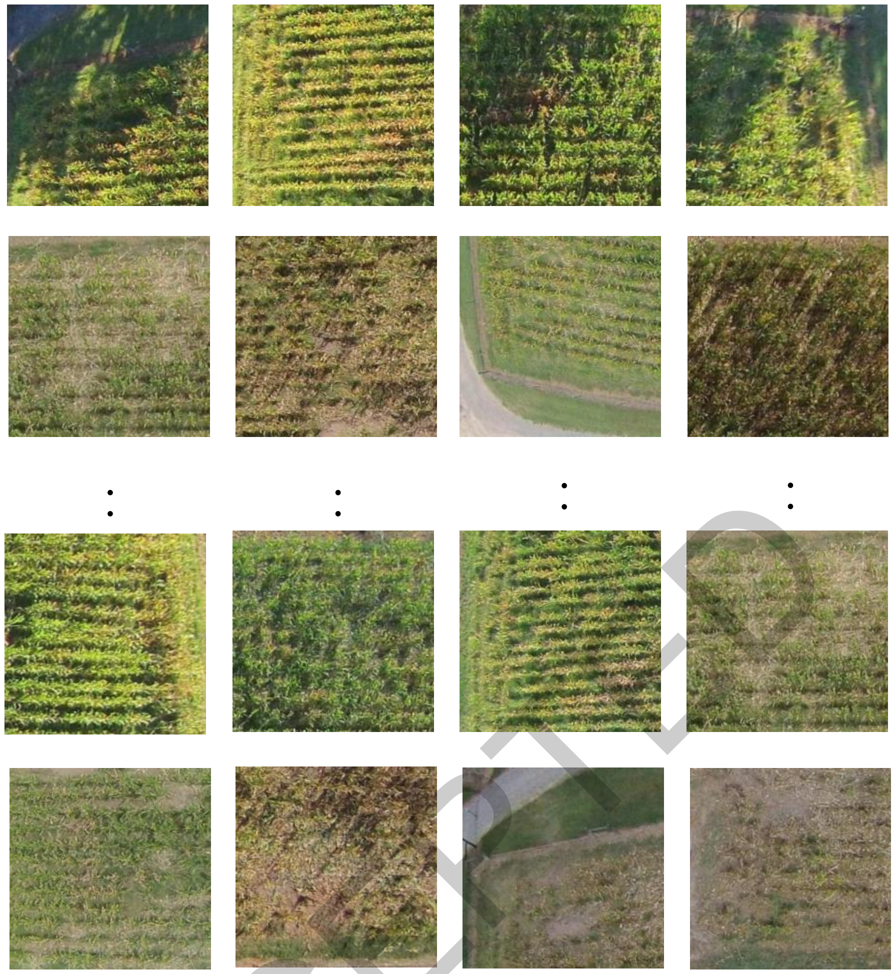
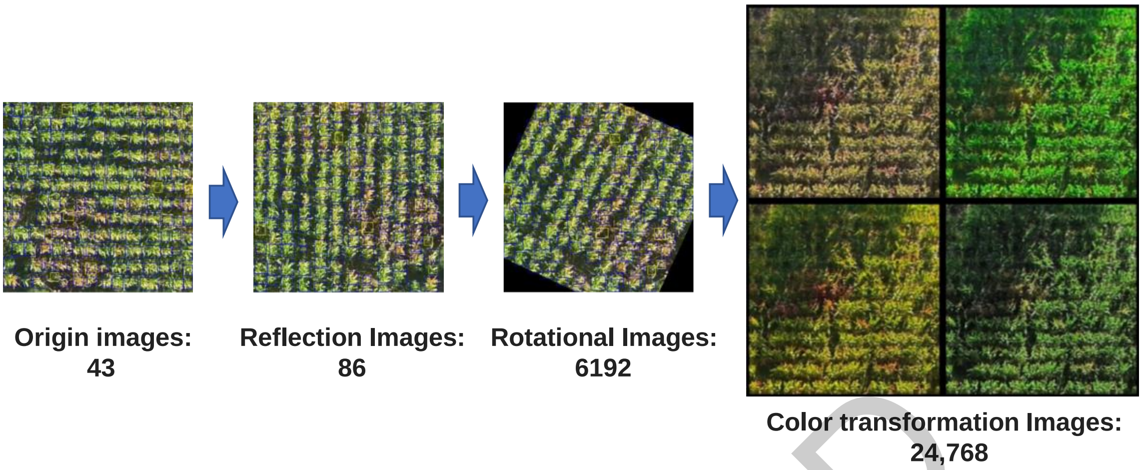


Figure 5. Raw training images

469



470

471 **Figure 6.** Data augmentation

472

473

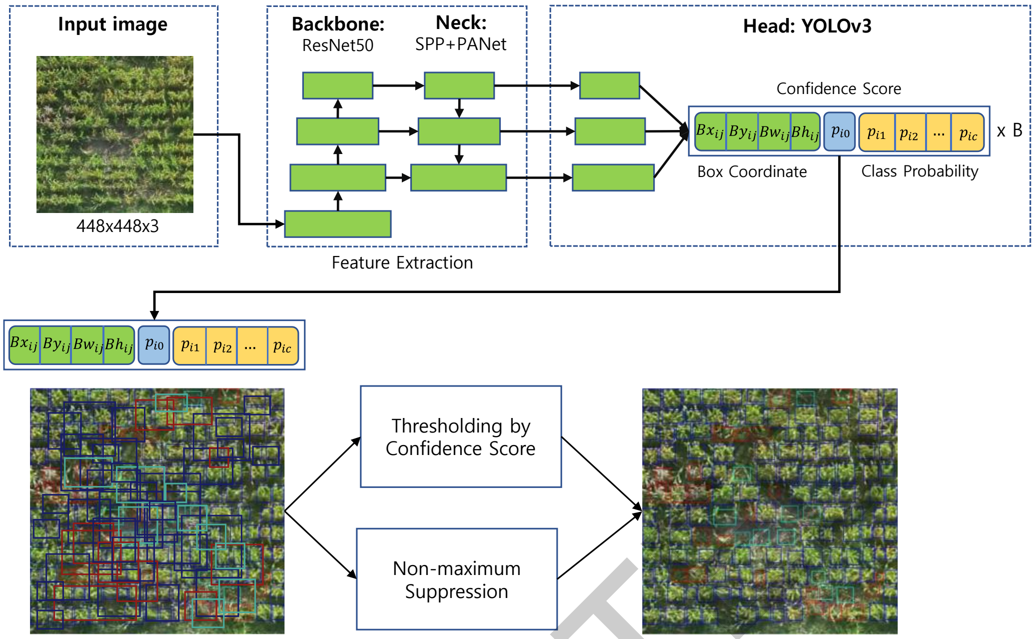


Figure 7. YOLO Detection

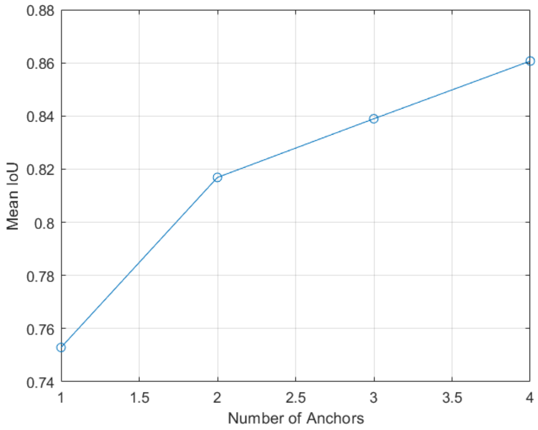


Figure 8. Number of Anchors vs. Mean IoU

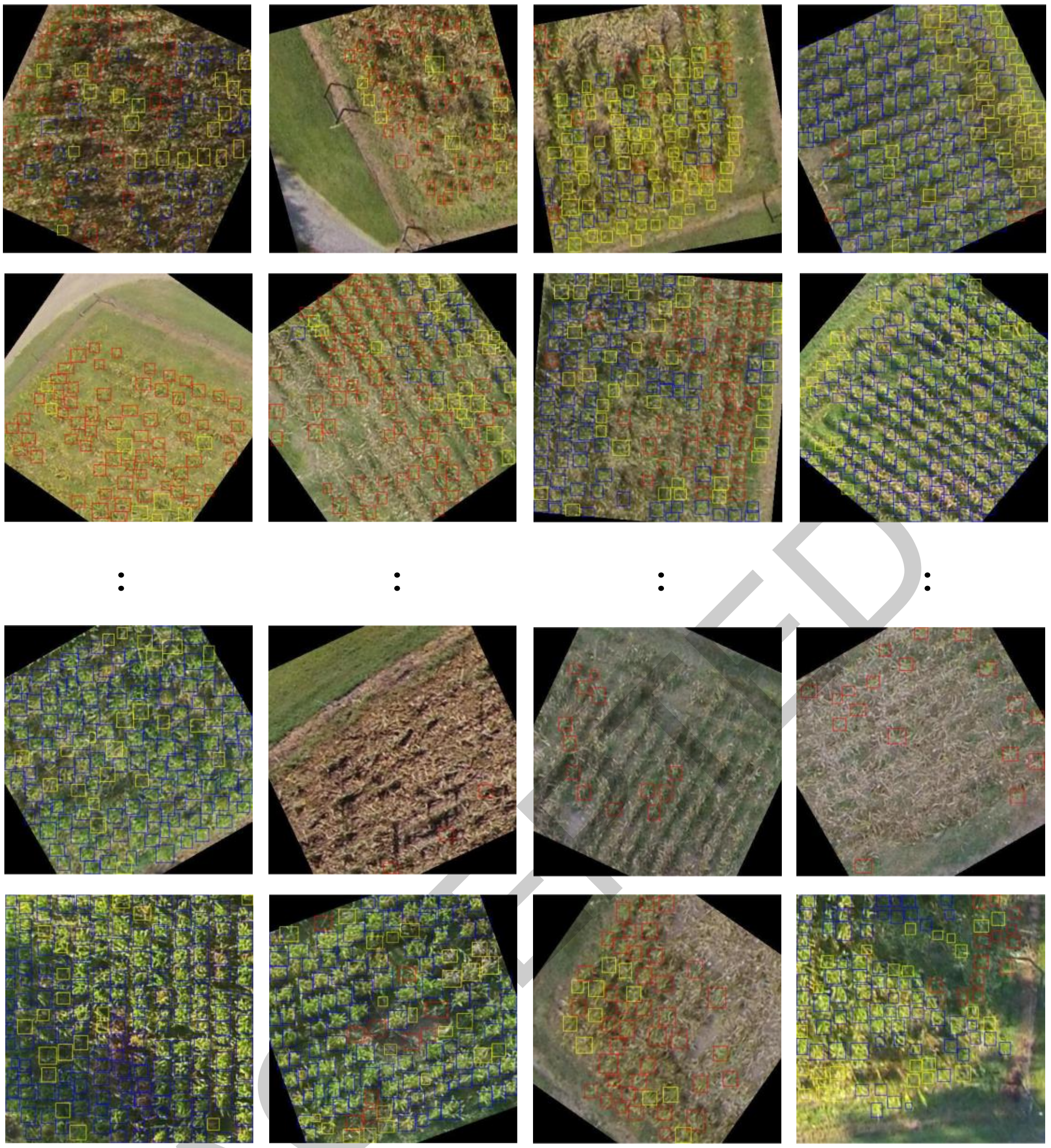


Figure 9. Recall results after training

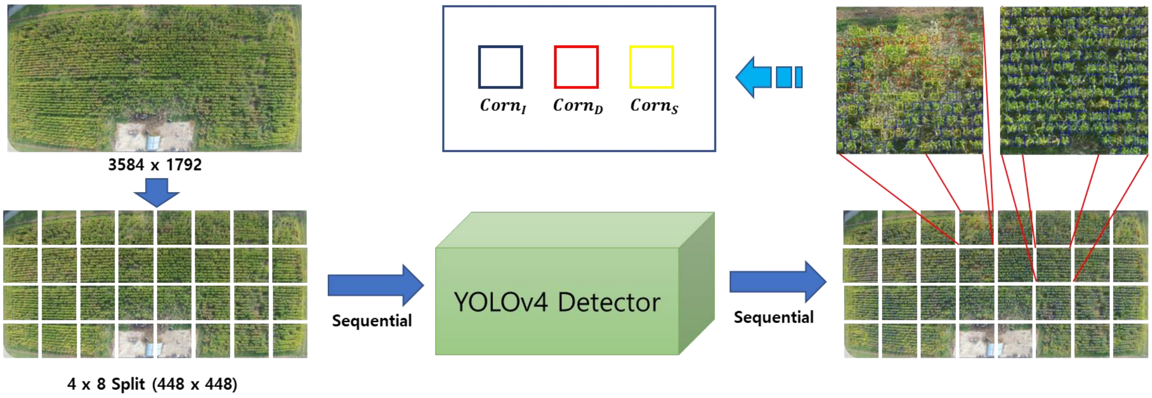
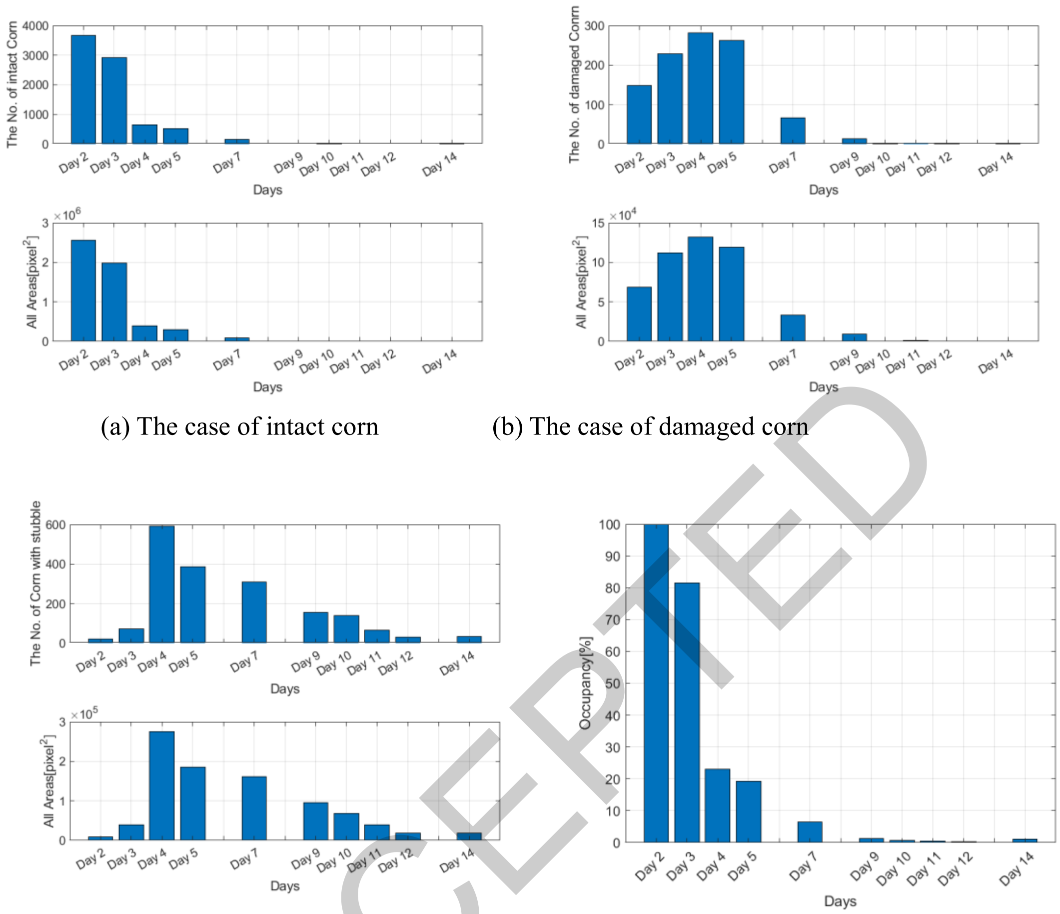


Figure 10. System configuration



(a) The case of intact corn (b) The case of damaged corn

(c) The case of corn with stubble (d) The occupancy rate of corn

Figure 11. The degree of occupancy of corn by date