Segmentation of Tomato Growing Truss a Depth Image Conversion Model Based on CycleGAN

Dae-Hyun Jung
Korea Institute of Science and Technology
Cheoul Young Kim
Yonsei University
Taek Sung Lee
Korea Institute of Science and Technology
Soo Hyun Park (ecoloves@kist.re.kr)
Korea Institute of Science and Technology https://orcid.org/0000-0002-9826-2695

Research

Keywords: Deep learning, Generative adversarial networks, Convolutional neural network, Robot platform
Abbreviations, GAN, generative adversarial networks

Posted Date: December 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1146999/v1

Abstract

Background: The truss on tomato plants is a group or cluster of smaller stems where flowers and fruit develop, while a growing truss is the most extended part of the stem. Because the state of the growing truss reacts sensitively to the surrounding environment, it is essential to control the growth in the early stages. With the recent development of IT and artificial intelligence technology in agriculture, a previous study developed a real-time acquisition and evaluation method for images using robots. Further, we used image processing to locate the growing truss and flowering rooms to extract growth information such as the height of the flower room and hard crab. Among the different vision algorithms, the CycleGAN algorithm was used to generate and transform unpaired images using generatively learning images. In this study, we developed a robot-based system for simultaneously acquiring RGB and depth images of the tomato growing truss and flower room groups.

Results: The segmentation performance for approximately 35 samples was compared through the false negative (FN) and false positive (FP) indicators. For the depth camera image, we obtained FN as 17.55±3.01% and FP as 17.76±3.55%. Similarly, for CycleGAN, we obtained FN as approximately 19.24±1.45% and FP as 18.24±1.54%. As a result of image processing through depth image, IoU was 63.56 ± 8.44%, and when segmentation was performed through CycelGAN, IoU was 69.25 ± 4.42%, indicating that CycleGAN is advantageous in extracting the desired growing truss.

Conclusions: The scannability was confirmed when the image scanning robot drove in a straight line through the plantation in the tomato greenhouse, which confirmed the on-site possibility of the image extraction technique using CycleGAN. In the future, the proposed approach is expected to be used in vision technology to scan the tomato growth indicators in greenhouses using an unmanned robot platform.

1. Introduction

In crops, the growing tip and the roots where cell division occurs are sensitive to the surrounding environment. In particular, the hypertrophy of early reproductive growth of crops can be determined from the state of the growing truss [1], which can also affect the quality of flowers and fruits. Although experts can determine hypertrophy with the naked eye, it makes the collection of accurate numerical data difficult, and realizes various disadvantages in setting the crop management standards. While studies are being actively conducted on analyzing crop diseases using digital imaging studies conducted on tomato crops, not many have measured the indicators related to tomato growth. the case of the growing truss, it is difficult to collect numerical information from the information obtained from the image to determine a label value considering the lack of video images for reference.

The development of a future-oriented agricultural robot platform is expected to reduce the challenges in acquiring image data comprising growth information. Singh et al. (2020) developed a mechanical robot arm with a high degree of freedom and an intelligent control unit that moves the arm by judging the

captured image. However, research on recognizing objects in a diversified view based on images by placing the robot arm in a more advantageous position is currently underway [2, 3]. Chang et al., 2015 reported the use of image processing techniques such as color space transformations, morphological operations, and 3D localization to identify objects and grippers in captured images and estimate their relative positions using the computer vision area as the novel algorithm for extracting the object before determining the movement of the robot arm. In agriculture, measuring the growth using computer vision has been in progress for a relatively long time [5, 6]. In particular, robots are used in harvesting, and various image processing techniques have been applied to extract fruit and determine the ripeness [7, 8]. Zhuang et al. (2019) proposed a computer-vision-based method for locating acceptable picking points for litchi clusters, and the image processing algorithm was used to track the location of the fruit while considering the agronomic characteristics of the picking point.

Although it is necessary to apply image processing techniques by identifying the characteristics of the crop, no image segmentation method has been developed to determine the growing truss in tomato plants. However, the segmentation of tomato stem and leaves were not able to distinguish the overlap of other surrounding objects in the RGB images. Because a tomato cultivation environment inside a greenhouse is dense, classifying stems or leaves using images is difficult [10, 11]. As a related study, Xiang, (2018) performed crop segmentation through a simplified pulse coupled neural network by measuring 385 tomato images at night time. The best results obtained through the segmentation technique confirmed that the best and false rates are 59.22% and 13.77%, respectively. However, there was a limitation in that it could be performed through a specific light at night for light correction, which would require more mechanical devices and technical improvements to measure the growing truss of tomatoes. Zhang and Xu, (2018) reported method for improving the accuracy of image segmentation in the middle stage and late stage of the fruit growth by using unsupervised method. However, the segmentation of tomato stem and leaf did not distinguish the overlap of other surrounding objects, so it did not show the possibility in RGB images. Many studies have been conducted on the fruit of the main target for identification of tomatoes using RGB images, and there are many reported results about the possibility in tomato cultivation, but segmentation studies on tomato stems at growing points have yet to be successfully reported.

In order to solve this problem, there is a potential possibility to use a 3D camera capable of segmentation according to distance or image processing techniques that are affected by solar light in greenhouse. 3D depth cameras are widely used in image acquisition platforms for recognizing objects in various industries, including agriculture [14–16]. It has been reported that a technology that combines depth and color image information through a stereo camera, one of the 3D camera technologies, can be presented, and segmentation of objects can be performed on real images recorded with a stereo camera [17, 18]. Unlike conventional 2D cameras, 3D depth cameras can be distributed to the field and used to calculate the depth value of each pixel in an image, whereas research on growth measurement is underway to determining growth measurement using 3D cameras.

Deep learning image processing technology has advanced in recent years. For instance, in image recognition and classification, studies using convolutional neural networks (CNN) are effectively applied to various industrial fields [19–22]. The use of Mask-RCNN, which recognizes objects at high speed and is specialized for segmentation, is expected. As a related study, Afonso et al., (2020) used Mask-RCNN for tomato fruit recognition and confirmed its potential inside a greenhouse. The structure of such a CNN has the form of general supervised learning, and annotation of the region of interest (ROI) is required in all image data, and the accuracy of the model is contributed to some extent by the quantity and quality of the data obtained. Therefore, it is important not only to develop a robot platform to extract accurate images in an automated greenhouse, but also to apply an algorithm that can self-learning with an appropriate number of images.

Generative advertising networks (GANs) have particularly gained wide attention [24, 25]. The basic GAN configuration comprises a deep learning technology that learns the delimiter and generator model simultaneously to obtain the target image from the generator, showing endless possibilities in unsupervised learning. The recently devised CycleGAN has been trained to avoid switching between images by learning two unpaired images by circulating the two generators and identifiers [26, 27]. A representative application example of the CycleGAN is a study wherein the zebra pattern was converted to that of an ordinary horse. Researches have reported (28) that this technology is capable of switching the patterns of two images, that is, a photo with depth information and a general image with RGB information. Furthermore, unlike other CNN algorithms, CycleGAN is a learning process while generating images by self-learning, and the number of labeled image data required is relatively small. This is expected to enable efficient algorithm application using relatively little data in environments where data acquisition is difficult, such as in a greenhouse environment.

Considering these points, we can say that the current research lacks detection technology for determining the tomato growing point, and it is necessary to systematically secure the related study. For image acquisition using an unmanned robot, extraction of the tomato growing truss must be performed on-site, which requires segmentation using depth image information. Therefore, the specific objectives of this study are :

1) To identify the tomato growing truss image, a monitoring system that can measure the height of the growing truss was built based on a robot, and based on this, the RGB image and the depth image of the growing truss were secured.

2) Using the obtained image, a conversion model between RGB and Depth is created through CycleGan, and this is combined with image processing techniques to segment the growing truss of tomatoes growing in the greenhouse without overlapping.

2. Materials And Methods

2.1 Greenhouse environment and image acquisition device

The experiment was conducted in a greenhouse facility where tomatoes are grown. A 2000 m2 interlocking Venlo greenhouse was utilized, wherein the insides comprised sensors and control systems to manage the level of carbon dioxide at constant temperature and humidity. The location of the greenhouse is at latitude 37.7986 and longitude 128.8575. We used Dafnis tomatoes for the experiment, and images of the harvested tomatoes were collected approximately 180 days after plantation. Tomatoes are grown in a greenhouse drip irrigation-based hydroponics system, and the nutrient solution is supplied through solar proportional irrigation control. The roots of tomatoes are established in the rock-wool substrate, and the substrate and the gutter supporting it are located at a height of about 1.3 m from the ground. The growing truss of tomatoes is located 1.6~2.5m from the gutter using the inducer lines, and this location is determined by the line works of the farmer.

To acquire the images, we used a vehicle placed on a robot platform capable of driving automatically in a greenhouse, and a 5-joint UR5 (Universal Robots, Odense, Denmark) was used as a menu plater to fix the photographing unit at the position of the tomato growing truss. The menu plate operation was manually adjusted in the field, and the position of the photographing camera was kept constant at the center of the line. The image acquisition unit comprising a Realsense 435i camera (Intel, Santa Clara, CA, USA) acquired RGB and depth images. The maximum resolution of the two camera is 1600 by 800. The measured images were collected on a mini-Windows PC (NUC, Intel, Santa Clara, CA, USA) and saved using Python programming. Figure 1 shows a photograph of the robot platform and the measurement module used.

2.2 CycleGAN implementation for segmentation of the tomato growing points

2.2.1 The CycleGAN structure

The GAN is said to be successful when an adversarial loss makes the generated image indistinguishable from the actual photo. This loss is particularly powerful for image-creation tasks considering most computer graphics aim at achieving optimization [27]. The objective of the CycleGAN model is to learn the mapping functions between two domains X and Y using the given training samples $\left\{x_i\right\}_{i=1}^{N}$ where $x_i \in X$ and $\left\{y_j\right\}_{j=1}^{M}$ where $y_j \in Y,$ which can be expressed as data distribution as $x\tilde{p}_{data}(x)$ and $y\tilde{p}_{data}(y)$. Zhu et al. introduced two cycle consistency losses [Figure.3(a)], indicating that the starting position of x must be reached when converting from one domain to another and vice versa. The forward cycle consistency loss is given as: x → G (x) → F (G (x))) ≈ x [Figure. 2(b)] and the reverse cycle consistency loss is given as y → F (y) → G (F (y)) ≈ y [Figure 2(c)].

2.2.2 Application of CycleGAN for tomato depth image transaction

As seen in Figure 3, the RGB and depth images were obtained from the robot platform and the acquisition unit. As seen in Figure 4 (left), a normal RGB image is similar to an image obtained from a normal camera. Figure 4 (right) shows a picture with the depth technology applied, and the location information between the camera and the object in the video is displayed in a color table.

Using CycleGAN learning, we constructed a model that converts RGB images to depth images and vice versa, as seen in Figure 4. The model was configured using approximately 356 samples images of the tomato growing truss pictures at the fruit growing stage acquired from the image acquisition device. Of the 356 sample images, 276 were used to train the CycleGAN model and 80 samples were used for testing.

Each CycleGAN generator comprises three sections: The encoder, transformation, and decoder. Figure 5 shows the components of each generator section. The 1600 × 900 pixels image used in this study was obtained as a raw value and resized to 512 × 512 pixels. First, the resized input image is fed directly to the encoder comprising three convolutional layers to increase the number of channels and decrease the representation size. The activated result is passed to the transformation, which is a series of eight resnet blocks that efficiently transfer information in the CNN structure. Therefore, it can be used as an optimal algorithm for the transformer layer. The transformation result is then expanded by the decoder comprising two transpose convolutions that enlarges the representation size and one output layer that produces the final RGB image. Although each layer is followed by an instance normalization and ReLU layer, it has been omitted from this study for simplicity. Furthermore, we built a discriminator that captures images and predicts whether it is real or fake. The image corresponding to real is an actual RGB or depth image, and fake means an image generated by CycleGAN. The generator can be visualized in the following image:

2.2.3 Image processing and evaluation methods for the extraction of growth points

The obtained depth image was subjected to pre-processing to extract the parts corresponding to the crop. First, a general RGB-based picture of the crop showing the tomato growing points were converted into a depth image. In the depth image, it exists as an image that can be divided through color change by recognizing the distance as a boundary line. Here, the growing truss of the closest part of the image we want has a relatively red color, therefore, we extracted the area through HSV. Although the extraction performance was better than the RGB-based method, the process was optimized with a trial-and-error method. In addition, morphological operations were performed for the filling of the remaining small fragments and the extracted area.

We used the model developed as a CycleGAN in this study. The image was pre-processed by applying the HSV and Otsu thresholds. As seen in Figure 6, the three HSV ranges were applied to the image preprocessing model in the development stage under the following three conditions: (a), H: 0 to 65 S: 150 to 255 V: 150 to 255, (b) H: 0 to 30 S: 180 to 245 V: 250 to 255, and (c) H :0 to 30 S: 248 to 255 V: 240 to

255. As a result, the HSV range corresponding to (c), showing the best growing truss, was applied. The image designated by the HSV was then converted to be further binarized using the Otsu threshold.

Figure 6. Comparison of crop extraction area using HSV range under three conditions: (a) H: 0 to 65 S: 150 to 255 V: 150 to 255, (b) H: 0 to 30 S: 180 to 245 V: 250 to 255, and (c) H: 0 to 30 S: 248 to 255 V: 240 to 255

The contour of the crop was determined using the morphology EX algorithm, which can perform advanced morphological transformations using basic erosion and dilation operations that can be performed in place. In multichannel images, each channel is processed independently. The edge was detected from the contour obtained, and erosion was performed in one iteration using a 3 × 3 kernel. Finally, erosion was performed to remove small objects that are independent and correspond to noise. Although this process can be applied universally in tomato greenhouses, it is difficult to use in general outdoor areas and places where the distance of the plantation from the camera keeps changing. The results of the entire image processing are shown in Figure 7.

We compared the accuracy between the obtained image from image processing and from which the growth point was manually extracted using 80 test samples. For the manual image extraction, a method of creating polygons and leaving ROI areas was intuitively determined by a person.

The image extracted by hand was assumed to be the actual region of interest. The extracted growing truss from image processing and the actual region of interest of the same size were overlapped, and the extracted image value at the same coordinate as the position of the actual growing truss was eliminated. The error rate was then calculated based on the number of pixels in the remaining images. Two indicators were calculated for the error rate: The residual ratio of image after removing the predicted pixels from the actual image was designated as false negative (FN), and after removing the actual image pixel from the predicted image pixel was designated as false positive (FP). In addition, as a standard segmentation method, intersection over union (IoU) was calculated for evaluating an image segmentation method. Figure 8 shows the specific calculation method for FN and IoU using the resulting image.

$$\text{Residual ratio } (\%) = \frac{The number of pixels in the remaining image}{Total number of pixels in the ROI of crop} \times 100 \backslash$$

$$\text{Intersection over union (IoU) } (\%) = \frac{Area of overlap}{Area of union} \times 100$$

2.3 Continuous measurement of images of robots for field applicability of CycleGAN

We conducted a field applicability test to examine the possibility of measuring the desired tomato stem section in the greenhouse crop bed driving. The vehicle was driven between the planting spaces in the greenhouse in a straight line and continuously scanned pictures of a particular location. We only

collected the RGB images from the RealSense camera, which were then converted using the previously developed CycleGAN. Further, an image processing technique was applied to extract the area of interest from the image. The RGB images were captured continuously at intervals of 1 min by advancing approximately 5 m for every 2 s by fixing the forward speed of the robot to 0.5 m/s. We simultaneously performed the image conversion and extraction of the region of interest on the stem. Figure 9 shows the performance of the growth measurement experiment inside an actual greenhouse.

3. Results

3.1 Training results of the CycleGAN

The growing truss of tomato was collected through the camera attached to the vehicle-based robot arm proposed above. A total of 350 pairs of images were collected, and CycleGAN learning was performed through this. This data can be confirmed through supplementary data. Figure 10 shows the collected data, the shape of the tomato growing point, and the greenhouse cultivation environment.

The CycleGAN was trained for approximately 9600 iterations in five batches using approximately 276 training samples. At this time, the changes in the loss of the generator and discriminators X and Y can be confirmed, as seen in Figure 11. First, the generator loss was observed to have converged in the half at a certain level, although there was some loss in the beginning. The discriminator gradually converged to 0.5 for Dx, but further converged to approximately 0.55 for Dy. For depth to RGB generation, an error with the actual sample was confirmed. However, the learning performance, which was mainly used in this study, seemed to have been secured in the RGB to depth image to an extent.

Figure 12 (a) shows the RGB-to-depth learning process. It was confirmed that the generator results obtained at 8800 iterations clearly depicted the appearance of crops as compared to initial iterations in the initial learning period. In addition, the RGB color differed based on the size and shape of the crop, and a similar pattern was observed in the depth images. Conversely, for the depth to RGB image, a low-quality crop image was obtained considering the input image could not generate high-quality images, as seen in Figure 12(b). Although the appearance, characteristics, and color of the crops were simulated like real RGB images, it was difficult to grasp specific features with the naked eye.

3.2 Accuracy of image extraction

The conversion from an RGB image to a depth image was mainly for the segmentation of target crops, and we verified the accuracy of FN and FP as an evaluation method. From the previously developed CycleGAN models, the results were inferred using 8,800 iterations, whereas the image pre-processing and growing truss extraction image processing methods were the same. We obtained results as seen in Figure 13 by comparing the results based on 80 images that were not used for model training. When the FN value was calculated using the image obtained from the depth camera, we obtained an approximate value of 17.55% ± 3.01%, and FP was 17.76±3.55%. Similarly, on converting the image using CycleGAN,

FN was approximately 19.24% ± 1.45% and FP was 18.24% ± 1.54%. In terms of error probability, although CycleGAN and depth images were compared with the actual extracted region and crossed segmentation values through IoU as shown in Figure 14. Among the total test samples, when using depth image, the IoU was 63.56 ± 8.44%, and when segmentation was performed through CycelGAN, the IoU was 69.25 ± 4.42%. As additional data, analysis result samples for each algorithm were additionally presented through the attached file IoU samples.

3.3 Field application results for continuous detection

Because this study aims to extract the growing truss of tomato crops in greenhouse using CycleGAN and image processing technology, the possibility was confirmed by field application experiments. Figure 15(a) shows the result of continuously acquiring and matching images with a height of approximately 3.5 m, wherein the growing truss can be confirmed while the image acquisition vehicle advances inside the greenhouse. After advancing for 5 m, and it was confirmed that approximately six crops were unevenly distributed. Figure 15(b) shows the result of converting the image into a depth image using the developed CycleGAN model. Similar to the actual depth image, the image showed the object to be segmented. The depth in the image was indicated in red for the closer crops, and in blue for the farther crops. Finally, the result of extracting the growing truss, that is, the stem and leaves of the tomato plant using the image processing technique, can be confirmed from Figure 15(c).

4. Discussion

While past researches have focused on applying the vision of crops in fruit-oriented research, in case of stem plants, such as tomatoes, the state of the growth point, which grows continuously, can be used as an important indicator to determine the future yield. Therefore, we conducted research on image processing techniques to identify the growing truss and used deep learning to acquire highly efficient results. We first devised an image processing technique for segmentation of the part that could be a growing truss through CycleGAN using a depth image and a simple RGB image converted into a depth image. Given that the CycleGAN is useful in image conversion, it was advantageous in recognizing objects that existed in two images, which was the growing truss to be extracted. Furthermore, it was possible to convert the color of the growing truss to the color of the depth. The color of the depth of the prepared training set was red, and the main learning factor. Owing to the CycleGAN method, both transformations can be advantageously applied, which has already been proven in a previous study that demonstrated the horse and zebra transformation attention [24, 25]. If we compare the purpose and approach of the existing segmentation studies on tomato images(Xiang, 2018; Zhang and Xu, 2018), there is a difference in the direction that many studies have focused on the analysis of tomato fruit. It is very difficult to classify the stems or leaves of the tomato we want because the growing environment is very dense. On the other hand, the possibility of an approach using depth imaging was confirmed in this study.

Although the identification error rate was lower on using the depth image, as seen in Figure 13, the average error rate was less than 20% in the two techniques, which indicated that the segmented object was not another region of interest. This was a result of the assumption that the error rate could not be reduced considering the ground truth was determined manually. However, in standard deviation, CycleGAN confirmed the result with a minor deviation value, which could have been due to the depth camera image being applied to the field considering the tendency of the camera to lose focus at proximity with a 10% probability, as seen in Figure 15 (c). However, because this was related to the applicability field of the camera, it was not considered in this study. Objects that remain unrecognized by the depth camera are termed as a failure case, as seen in Figure 16, which can cause problems will be in field applications in the future. However, this problem did not occur in the depth image converted using CycleGAN considering it was already being used in preparing the training set stage.

Additionally, it is often necessary to prepare annotated image samples as training data for artificial intelligence algorithms that recognize objects by judging through human intellectual contribution, and a large amount of data samples is required to verify the accurate performance. On comparing these points, the results of CycleGAN and the image processing method proposed in this study have confirmed that preparing the annotated image samples is not required.

In the future, robots would be required in agriculture to automatically measure plantation growth. However, to choose the desired growing truss, the robot must accurately recognize the growing truss to establish a menu plating strategy. In this study, we adopted CycleGAN, an artificial intelligence image conversion technique, as the first step for the robot to recognize the growing truss. As a result, the robot was able to effectively extract the growing truss using the matched image even in field applications. In the field applicability verification experiment, the moving robot matched several images and finally converted the image using CycleGAN. The result was verified by only extracting the growing truss from the image. However, an irregular connection of images was observed during the registration, and the CycleGAN structure used when converting to depth applies to only 512 × 512 images, making a grid shape inside the images. As this applies to all images using deep learning, it is necessary to solve the problem using an algorithm for the flexible application of the input layer structure. Nevertheless, the result indicated that the application of unmanned robots in agriculture in the future has been well considered.

In future research, we will consider using a method of acquiring optimal images by menu plating the robot arm once the growing truss is recognized. Automated robots and systems are likely to be delivered to systems. In addition, the result of converting the depth image to an RGB image, although not addressed in this paper, is worthy of discussion as a future study (Figure 12). Virtual reality has a high potential for human contribution [30], and creative results can be achieved when fused with artificial intelligence.

5. Conclusions

In this study, we developed a technique for extracting the growing truss in tomato plantation in a greenhouse using image processing techniques based on the image information obtained by a robot platform and images of the growing truss captured by a depth camera. Furthermore, a study was conducted to convert the characteristics of two images, that is, converting RGB images into depth images, using the CycleGAN algorithm. Discriminators X and Y used in the loss of learning process converged to 0.43 and 0.65, respectively. The image information converted using CycleGAN was further used to compare the performance of the extraction of growing truss. The false negative value based on the images from the depth camera was approximately 17.55% ± 3.01%, and the false positive value was 17.76 ± 3.55%. Similarly, using CycleGAN, the false negative was approximately 19.24% ± 1.45% and the false positive was 18.24% ± 1.54%. When using depth image, the IoU was 63.56 ± 8.44%, and when segmentation was performed through CycleGAN, the IoU was 69.25 ± 4.42%. In terms of error probability, CycleGAN exhibited a higher value. Finally, we performed field application tests to determine the growing truss of tomatoes, wherein the continuously scanned image information was converted into depth images using CycleGAN. In the future, the proposed approach is expected to be used in vision technology to scan the tomato growth indicators in greenhouses using an unmanned robot platform.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors consent on the submission of this paper for publication in Plant Methods.

Availability of data and materials

The raw RGB and Depth images of tomato growing truss used in this study, CycleGAN code, executable .py file, and the model generator H5 file were shared at the supplementary file.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was supported by the Industrial Fundamental Technology Development Program (Grant No. 20004055) funded by the Ministry of Trade, Industry & Energy (MOTIE) of Korea.

Authors' contributions

DHJ is expected to have made substantial contributions to the conception, design of the work, the acquisition, analysis, interpretation of data, and the creation of new software used in the work. CYK and TSL designed image acquisition devices and analyzed the data. SHP have substantively revised the manuscript. All authors have agreed to the submission of the manuscript for publication. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank all KIST staff for their assistance in collecting image data.

References

1. Araus JL, Kefauver SC. Breeding to adapt agriculture to climate change: affordable phenotyping solutions. Curr Opin Plant Biol. 2018;45:237–47.
2. Chu C-R, Lan T-W, Tasi R-K, Wu T-R, Yang C-K. Wind-driven natural ventilation of greenhouses with vegetation. Biosyst Eng [Internet]. 2017;164:221–34. Available from: http://www.sciencedirect.com/science/article/pii/S1537511017307183
3. Yang C, Wu H, Li Z, He W, Wang N, Su C. Mind Control of a Robotic Arm With Visual Fusion Technology. IEEE Trans Ind Informatics. 2018;14(9):3822–30.
4. Chang J-W, Wang R-J, Wang W-J, Huang C-H. Implementation of an Object-Grasping Robot Arm Using Stereo Vision Measurement and Fuzzy Control. Int J Fuzzy Syst [Internet]. 2015;17(2):193–205. Available from: https://doi.org/10.1007/s40815-015-0019-2
5. Popa C. Adoption of Artificial Intelligence in Agriculture. Bull UASVM Agric [Internet]. 2011 [cited 2019 Feb 2];68(1). Available from: http://journals.usamvcluj.ro/index.php/agriculture/article/viewFile/6454/5747
6. Yaguchi H, Nagahama K, Hasegawa T, Inaba M. Development of an autonomous tomato harvesting robot with rotational plucking gripper. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2016. p. 652–7.
7. Kondo N, Yata K, Iida M, Shiigi T, Monta M, Kurita M, et al. Development of an End-Effector for a Tomato Cluster Harvesting Robot. Eng Agric Environ Food [Internet]. 2010;3(1):20–4. Available from: https://www.sciencedirect.com/science/article/pii/S1881836610800072
8. Ling X, Zhao Y, Gong L, Liu C, Wang T. Dual-arm cooperation and implementing for robotic harvesting tomato using binocular vision. Rob Auton Syst [Internet]. 2019;114:134–43. Available from: https://www.sciencedirect.com/science/article/pii/S092188901830808X
9. Zhuang J, Hou C, Tang Y, He Y, Guo Q, Zhong Z, et al. Computer vision-based localisation of picking points for automatic litchi harvesting applications towards natural scenarios. Biosyst Eng [Internet]. 2019;187:1–20. Available from: https://www.sciencedirect.com/science/article/pii/S1537511019308116

10. Wan P, Toudeshki A, Tan H, Ehsani R. A methodology for fresh tomato maturity detection using computer vision. Comput Electron Agric. 2018;146:43–50.
11. Ling X, Zhao Y, Gong L, Liu C, Wang T. Dual-arm cooperation and implementing for robotic harvesting tomato using binocular vision. Rob Auton Syst. 2019;114:134–43.
12. Xiang R. Image segmentation for whole tomato plant recognition at night. Comput Electron Agric [Internet]. 2018;154:434–42. Available from: https://www.sciencedirect.com/science/article/pii/S0168169918309268
13. Zhang P, Xu L. Unsupervised Segmentation of Greenhouse Plant Images Based on Statistical Method. Sci Rep [Internet]. 2018;8(1):4465. Available from: https://doi.org/10.1038/s41598-018-22568-3
14. Vitzrabin E, Edan Y. Changing Task Objectives for Improved Sweet Pepper Detection for Robotic Harvesting. IEEE Robot Autom Lett. 2016;1(1):578–84.
15. Osman HI, Hashim FH, Zaki WMDW, Huddin AB. Entryway detection algorithm using Kinect's depth camera for UAV application. In: 2017 IEEE 8th Control and System Graduate Research Colloquium (ICSGRC). IEEE; 2017. p. 77–80.
16. Battisti F, Bosc E, Carli M, Le Callet P, Perugia S. Objective image quality assessment of 3D synthesized views. Signal Process Image Commun. 2015;30:78–88.
17. Ottonelli S, Spagnolo P, Mazzeo PL, Leo M. Improved video segmentation with color and depth using a stereo camera. In: 2013 IEEE International Conference on Industrial Technology (ICIT). 2013. p. 1134–9.
18. Leens J, Piérard S, Barnich O, Van Droogenbroeck M, Wagner J-M. Combining color, depth, and motion for video segmentation. In: International Conference on Computer Vision Systems. Springer; 2009. p. 104–13.
19. Ubbens JR, Stavness I. Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks [Internet]. Vol. 8, Frontiers in Plant Science . 2017. p. 1190. Available from: https://www.frontiersin.org/article/10.3389/fpls.2017.01190
20. Jung D-H, Kim NY, Moon SH, Jhin C, Kim H-J, Yang J-S, et al. Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering. Animals. 2021;11(2).
21. Hershey S, Chaudhuri S, Ellis DPW, Gemmeke JF, Jansen A, Moore RC, et al. CNN architectures for large-scale audio classification. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017. p. 131–5.
22. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
23. Afonso M, Fonteijn H, Fiorentin FS, Lensink D, Mooij M, Faber N, et al. Tomato Fruit Detection and Counting in Greenhouses Using Deep Learning [Internet]. Vol. 11, Frontiers in Plant Science . 2020. p. 1759. Available from: https://www.frontiersin.org/article/10.3389/fpls.2020.571299
24. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Advances in neural information processing systems. 2014. p. 2672–80.

25. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. Med Image Anal. 2019;58:101552.
26. Hiasa Y, Otake Y, Takao M, Matsuoka T, Takashima K, Carass A, et al. Cross-modality image synthesis from unpaired data using CycleGAN. In: International workshop on simulation and synthesis in medical imaging. Springer; 2018. p. 31–41.
27. Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 2223–32.
28. Kwak D, Lee S. A Novel Method for Estimating Monocular Depth Using Cycle GAN and Segmentation. Vol. 20, Sensors . 2020.
29. Xiang R. Image segmentation for whole tomato plant recognition at night. Comput Electron Agric. 2018;154:434–42.
30. Yu F, Zhang J-F, Zhao Y, Zhao J-C, Tan C, Luan R-P. The research and application of virtual reality (VR) technology in agriculture science. In: International Conference on Computer and Computing Technologies in Agriculture. Springer; 2009. p. 546–50.

Figures

Figure 1

(Left) Robot platform for image acquisition in greenhouse. (Right) End effector for RGB depth image acquisition

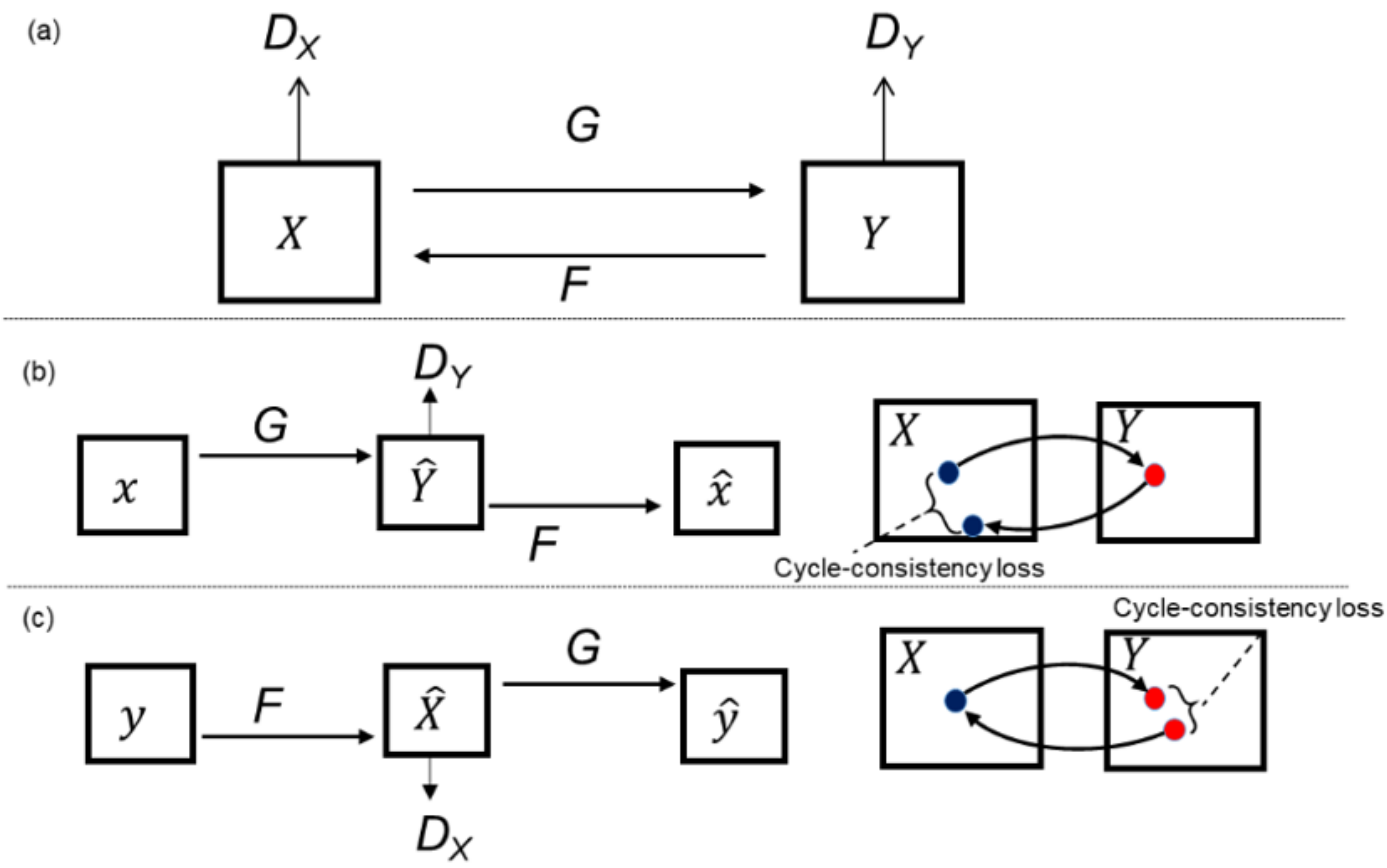

Figure 2

The CycleGAN structure. (a) Two mapping functions G: X → Y and F: Y → X. (b) Forward cycle-consistency loss. (c) Reverse cycle-consistency loss.

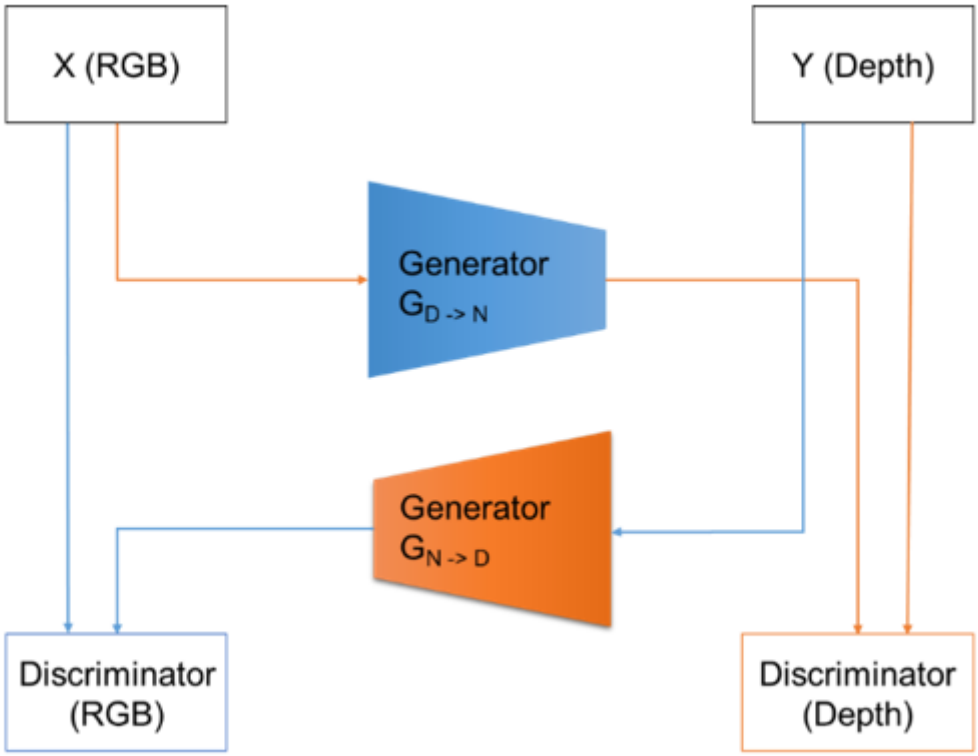

Figure 3

Schematic representation of the two cyclical generators of the CycleGAN.

Figure 4

Relationship between the images generated from the X and Y generators and the image data to be extracted.

Figure 5

The CycleGAN generator architecture

Figure 6

Comparison of crop extraction area using HSV range under three conditions: (a) H: 0 to 65 S: 150 to 255 V: 150 to 255, (b) H: 0 to 30 S: 180 to 245 V: 250 to 255, and (c) H: 0 to 30 S: 248 to 255 V: 240 to 255

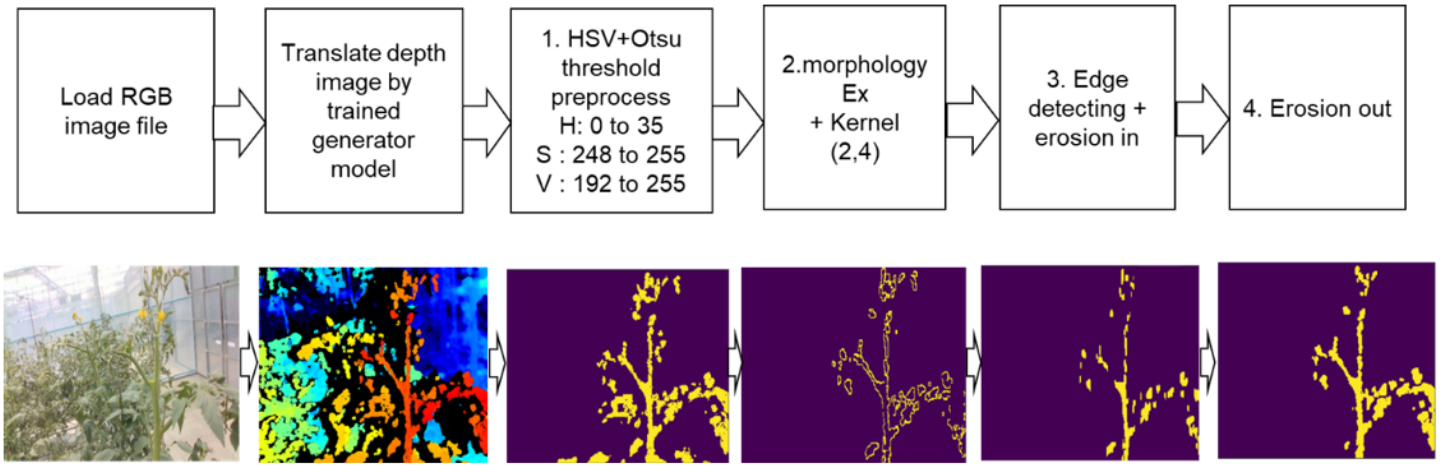

Figure 7

The entire image processing after CycleGAN conversion.

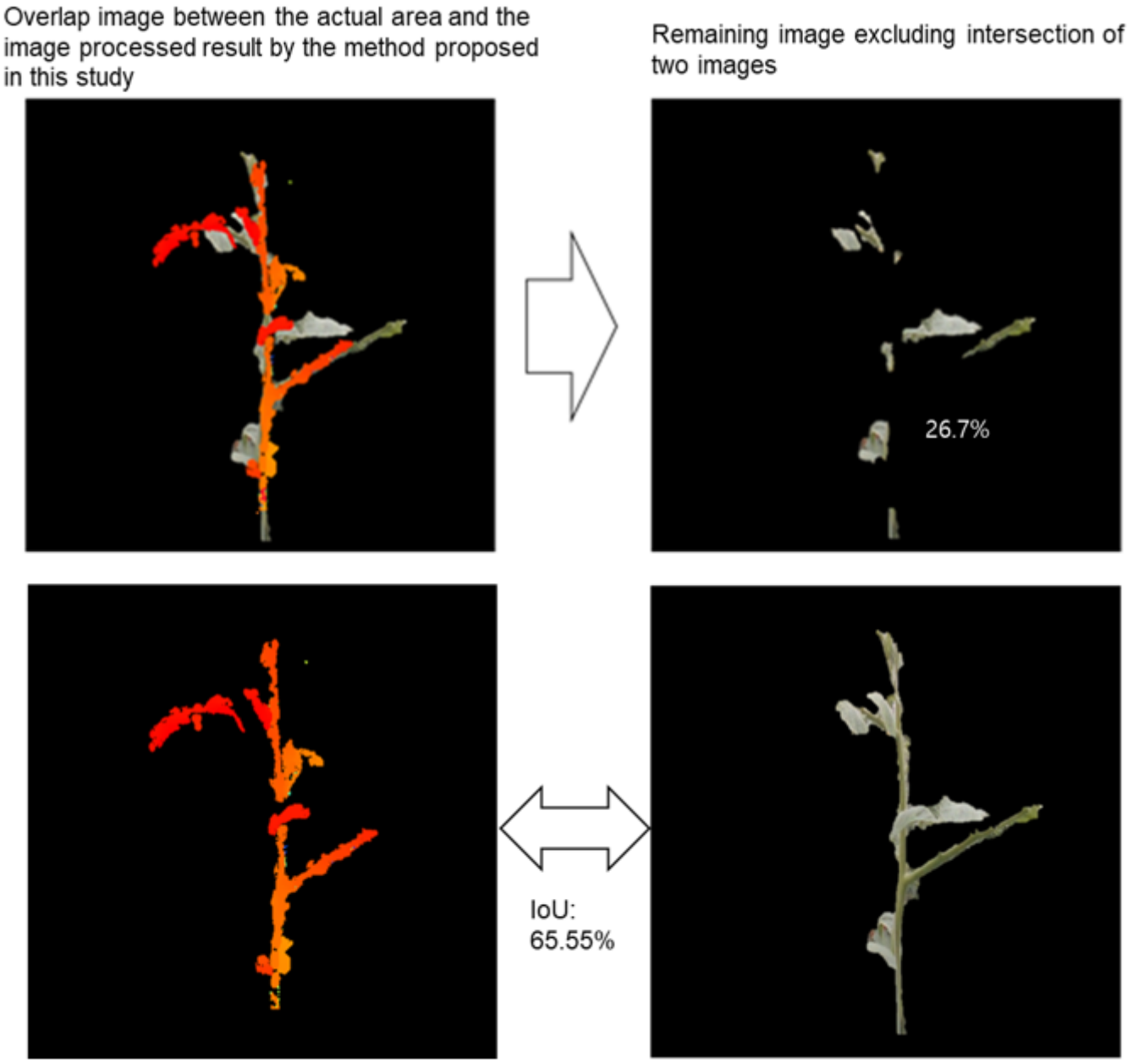

Figure 8

Ratio of the remaining image after removing the predicted pixel from the actual image, designated as false negative (FN), and when the actual image pixel is removed from the predicted image pixel, designated as false positive (FP).

Figure 9

Schematic representation of the experiment for applying the developed algorithm to sequentially captured and registered RGB images. (a) Front view of the scanning area of the robot. (b) Experimental schematic. (c) Actual robot running direction.

Figure 10

RGB and depth image to be used in the acquired data set (top : RGB, bottom : depth)

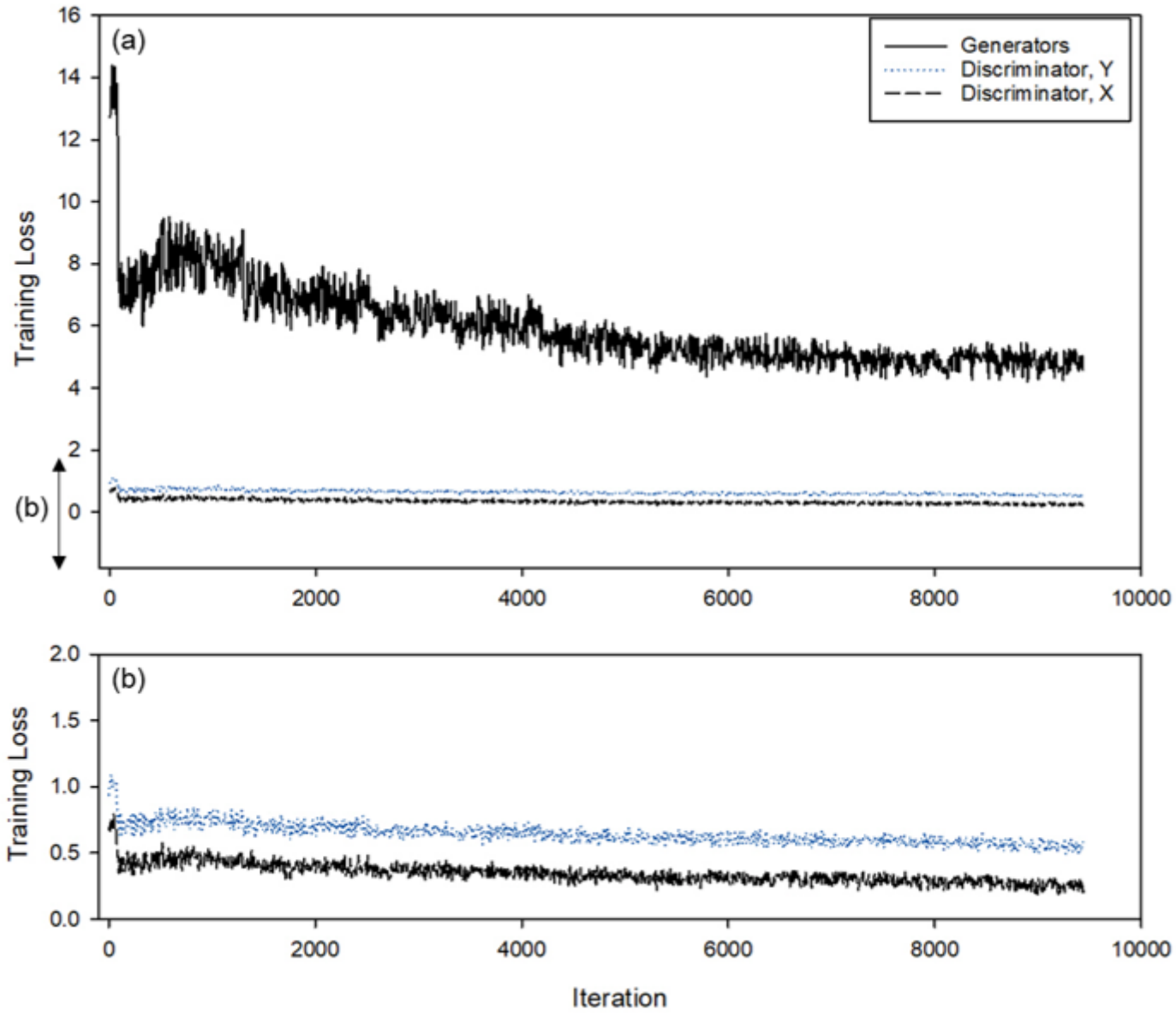

Figure 11

Changes to the overall loss. (a) Change in loss of discriminator X and Y. (b) On training the proposed CycleGAN structure.

Figure 12

Results of RGB to depth image conversions (a) and realizing the depth in RGB (b) through CycleGAN's 8800 iteration learning.

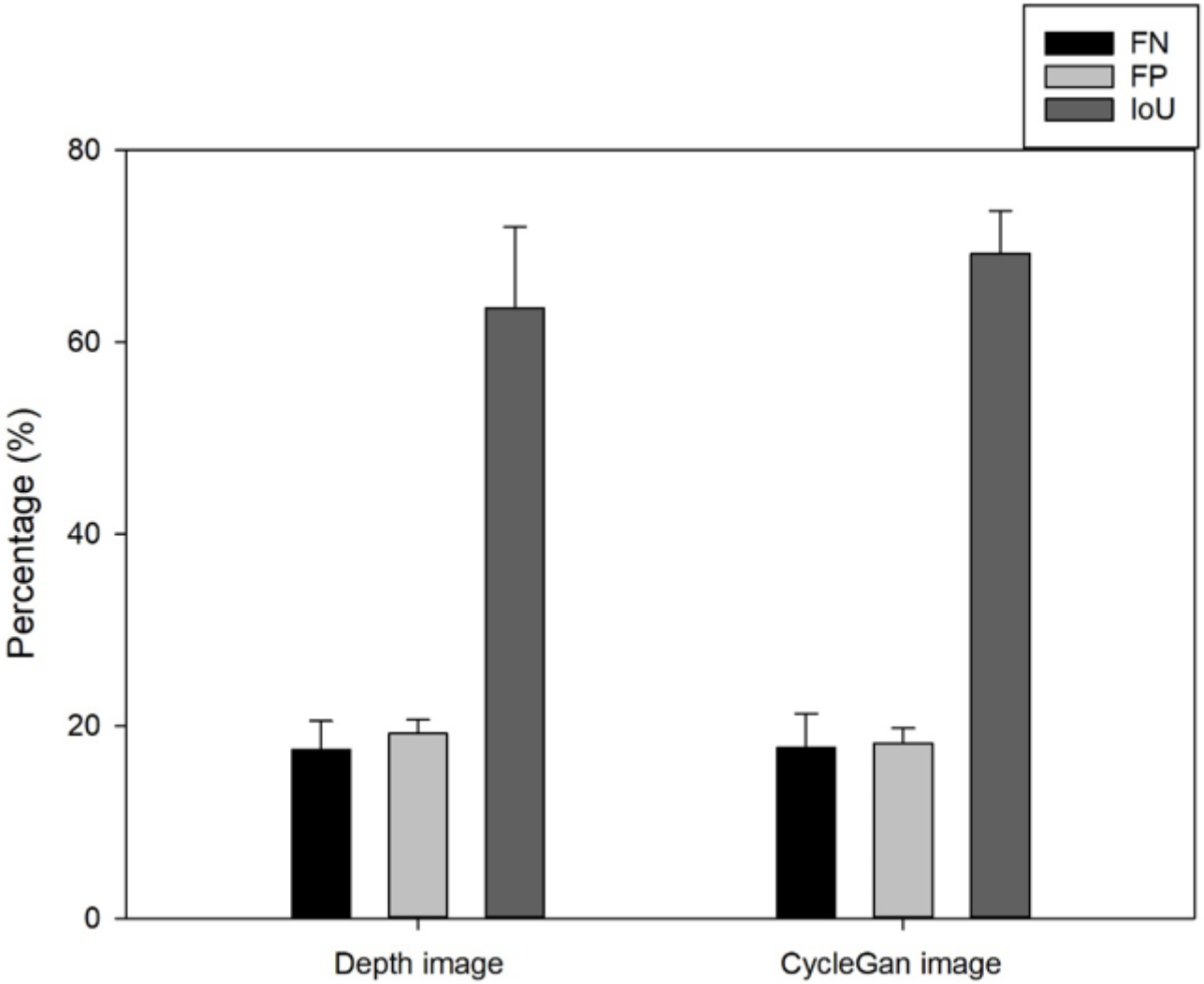

Figure 13

Comparison between depth image and CycleGAN image with ROI specified by hand through image processing technique, and the FP, FN, and IoU values in pixel units.

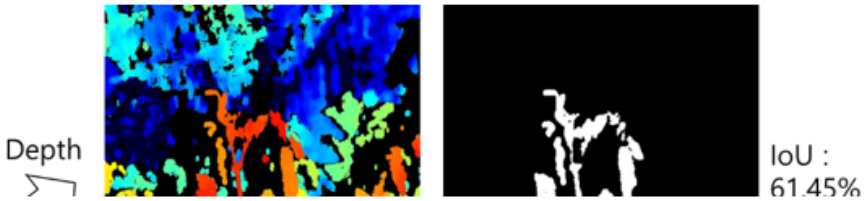

Figure 14

Comparison result of CycleGAN and Depth Image with the actual extracted area and cross-segmentation values through IoU

Figure 15

RGB image acquired using on-site robot platform(a), Depth image created using CycleGAN. (b) view underneath the growth point location extracted through image processing (c).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PMsupplementarydata.zip