JAME Journal of Agricultural Machinery Engineering

pISSN : 2799-8673
eISSN : 2799-8819

RESEARCH ARTICLE

딸기 수확 로봇을 위한 완숙 딸기의 자세 추정

이태신, 강승우, 이대현*
충남대학교 농업기계공학과

Pose estimation of ripe fruit for strawberry harvesting robot

Tae-Sin Lee, Seung-Woo Kang, Dae-Hyun Lee*

Department of Agricultural Machinery Engineering, Chungnam National University, Daejeon, Republic of Korea

*Corresponding author: leedh7@cnu.ac.kr

Abstract

Domestic agriculture is facing a decline in strawberry cultivation area due to the aging population, necessitating the development of automation technologies for agricultural tasks. Research on agricultural robots has largely focused on clearly defined harvesting tasks. In automating these processes, effective detection technologies play a crucial role. However, because a strawberry consists of multiple components including the fruit, calyx, and stem precise pose estimation of these elements is essential for determining the optimal cutting point. Therefore, in this study we suggested a pose estimation method to provide detection information for strawberry harvesting robots. Our suggested method defines three classes essential for strawberry harvesting (fruit, calyx, and stem) and detects each object using a deep learning model. Subsequently, a matching algorithm is employed to form coherent groups by associating the detected fruit, calyx, and stem that belong to a single strawberry, and pose estimation is performed based on these groups. The evaluation of the detected classes and keypoints was conducted using mean average precision (mAP), keypoint pixel error, and the percentage of detected keypoint (PDK) index. Results showed an average 0.89 mAP@0.5, an average pixel error of 16.00 ± 10.42 for the x-coordinate and 13.18 ± 6.92 for the y-coordinate, and an average 59.33% PDK@0.5. Although some false positives occurred for the calyx and stem in certain images, the proposed method successfully detected most objects and estimated their poses. Results of this study, the proposed approach is expected not only to detect the fruit but also to comprehensively estimate the poses of all strawberry components, thereby contributing to the development of robotic detection systems capable of precise manipulation and efficient harvesting.

Keywords: strawberry, harvesting robot, pose estimation, Hungarian algorithm

Check for updates

OPEN ACCESS

Journal of Agricultural Machinery Engineering 5(1):13-24

DOI: https://doi.org/10.12972/jame.2025.5.1.2

Received: February 27, 2025
Revised: March 31, 2025
Accepted: March 31, 2025

Copyright: © 2025 Korean Society for Agricultural Machinery

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http: //creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

국내 농업은 65세 이상 고령 인구 비율의 꾸준한 증가와 젊은 세대의 농촌 이탈로 심각한 노동력 부족 문제를 겪고 있다. 이는 농업 생산성 감소에 직접적인 영향을 미치고 있으며(KREI, 2023), 이를 해결하기 위해 자율 주행 시스템, 농업용 드론, 농업로봇 등 다양한 농작업 자동화 기술이 활발히 개발되고 있다.

그중 농업로봇은 농업 생산과 가공, 유통, 소비의 전 과정에서 자율적인 동작과 지능화된 작업을 수행함으로써 기존 농업 인력을 대체할 수 있는 자동화 기술 중 하나로 주목받고 있으며(Park et al., 2017), 특히 수확 작업을 중심으로 주로 안정된 환경의 시설에서 재배되는 과채류를 대상으로 활발히 연구되고 있다(Kang et al., 2024). 자동 수확을 위해서는 로봇이 수확 대상 작물을 정확히 인식하고 최적의 절단 지점을 결정할 수 있어야 하며, 이를 위해 정밀한 인식 기술은 우선적으로 갖추어야 할 핵심 요소이다.

딸기는 국내에서 재배되는 시설 원예 작물 중 생산액이 가장 높은 작물로, 농가 소득 비중이 매우 높을 뿐만 아니라 주요 수출량이 지속적으로 증가되고 있다(KOSIS, 2022; Lee et al., 2016). 이에 따라 시설재배 딸기를 대상으로 수확 로봇을 위한 인식 연구가 다수 수행되었으며, 최근에는 심층학습(deep learning)을 이용한 방법이 높은 성능을 보여주고 있다. 특히, YOLO (you look only once) 모델이 높은 실시간성으로 딸기의 생장 상태 모니터링(An et al., 2022), 실시간 성장 단계 인식(Li et al., 2023a, Li et al., 2023b) 등 많은 연구에 활용되고 있으며, 최근에는 LW-Swin Transformer와 YOLOv8 모델을 동시에 사용한 완숙 딸기 검출(Liu et al., 2021; Yang et al., 2023)도 진행되었다. 이와 같이 딸기를 대상으로 다양한 인식 연구가 진행되었으나 대부분 성숙 단계 판별에 한정되어 있으며, 실제 자동 수확을 위해서는 완숙 판별 외 과실, 줄기 등 작업부에 대한 기구학적 정보가 추가적으로 요구된다.

또한, 딸기와 같이 과육이 부드러운 작물은 수확 과정에서 과실 표면의 손상을 유발할 가능성이 있어 줄기를 절단 혹은 열을 가해 과실을 분리하는 방식으로 수확이 이루어진다(Dimeas et al., 2015; Yu et al., 2020). 따라서, 수확에 필요한 정보 제공을 위해서는 대상 작물의 과실과 줄기의 정확한 검출이 필수적으로 수행되어야 한다(Chai et al., 2024). 그러나 실제 농업 환경에서는 다수의 과실과 줄기가 한 화면 내에 동시에 존재할 수 있어, 단순한 객체 검출만으로는 각 과실과 연결된 줄기를 명확히 식별하기 어려워 정확한 자세 추정을 위해서는 각 과실과 해당 줄기를 명확히 매칭하는 기술이 필요하다. 이에 따라, 연관성 결정을 위해 이분 그래프 매칭(bipartite graph matching)을 이용한 방법이 사용되고 있다. 그중 헝가리안 알고리즘은 서로 다른 두 그룹 간의 최적의 대응 관계를 찾기 위한 최적화 알고리즘으로(Munkres, 1957), 일반적으로 객체 간 최적의 연결 관계를 찾는 데 사용되며, 사람의 자세 추정을 위한 관절 간 연결(Cao et al., 2017), 최근에는 농업환경에서 토마토의 과실-줄기 매칭(Jang and Hwang, 2024) 등 다양한 자세 추정 연구에서 적용되고 있다. 이처럼, 헝가리안 알고리즘을 통한 연관성 매칭과 딥러닝 기반의 검출 모델을 결합함으로써, 다수의 객체가 존재하는 환경에서도 효과적인 자세 추정이 가능하다.

본 연구에서는 시설딸기 자동 수확 시 수확 대상체의 기구학적 정보 제공을 위해 딸기의 과실-줄기쌍의 위치와 자세를 자동 인식하였다. 검출 모델은 YOLOv8-segmentation을 이용하였으며, 검출된 객체 구성 요소 간 매칭은 헝가리안 알고리즘을 이용하여 결정되어 자세 추정이 수행되었다.

Materials and Methods

딸기 자세 인식 개요

본 연구에서 딸기의 자세 인식 개요는 Fig.1과 같이 세분화, 매칭, 자세 추정의 세 가지 단계로 구성된다. 세분화 단계에서

는 농업 환경에서 촬영된 이미지에서 과실, 꽃받침, 줄기 객체를 동시에 검출하기 위해 딥러닝 기반의 인스턴스 세분화 모델을 사용하였으며, 이를 통해 영상 내 개별 객체가 분리되어 각 요소가 독립적으로 검출된다. 매칭 단계에서는 검출된 객체 간의 연결 관계를 명확히 파악하기 위해 헝가리안 알고리즘을 활용하였으며, 과실과 연결된 꽃받침 및 줄기의 그룹화가 수행되어 이미지 내 존재하는 딸기 객체들이 각각의 그룹으로 구분된다. 최종적으로 자세 추정 단계에서는 그룹화된 딸기 객체들 중 완숙 딸기 그룹의 각 요소 중심점을 직선으로 연결함으로써 딸기의 자세가 결정된다.

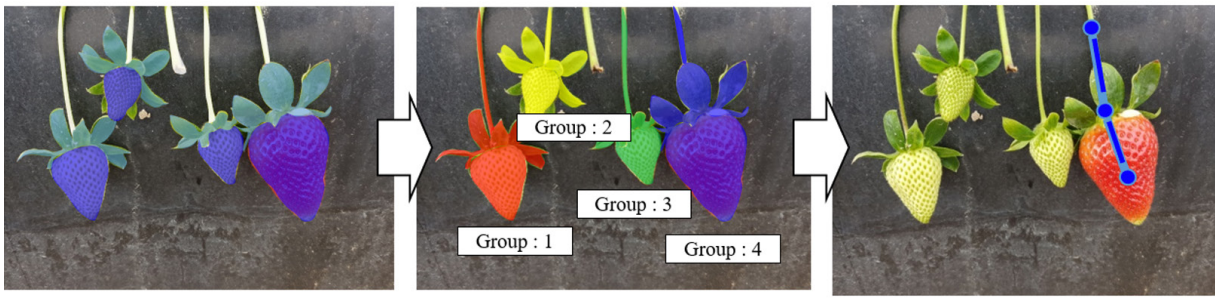


Fig. 1. Overview process of strawberry pose estimation proposed in this study.

딸기 주요 객체 세분화 모델

본 연구에서 과실, 꽃받침, 줄기 객체를 세분화하기 위해 실시간 객체 인식에서 높은 성능을 보이는 YOLOv8-segmentation 모델이 사용되었다(Jocher et al., 2023). 모델의 전체적인 구조는 Fig.2와 같이 backbone, neck, head의 3개의 과정으로 구성되며, 각 구성 요소는 특징 추출, 다중 해상도 특징 결합, 최종 검출 수행을 담당한다. backbone은 입력 영상에서 주요 특징을 추출하는 합성곱 신경망(convolutional neural network, CNN) 기반의 CSPDarknet53을 구조를 사용하며, cross stage partial network (CSPNet) 기법을 적용함으로써 연산량을 줄이면서도 정보 손실을 최소화하는 역할을 한다. 이를 통해 깊은 네트워크 구조에서도 효과적인 특징 학습이 가능하며, 고차원적 표현 학습을 통해 객체 검출의 정확도를 향상시킨다. neck은 다양한 크기에서 추출된 저차원과 고차원의 특징을 양방향으로 결합하여, 정보 손실을 최소화하고 다양한 규모에서의 객체 감지 성능을 향상시키는 역할을 수행한다. 이 과정에서 C2f 모듈이 사용되며, 이는 CNN, skip-connection, bottleneck 모듈을 포함하여 특징 전달과 학습을 최적화한다. 이를 통해 서로 다른 해상도의 특징이 효과적으로 조합되며, 작은 객체의 세부 구조와 큰 객체의 전역적 특성이 모두 반영되어 다양한 크기와 형태의 객체를 정밀하게 검출할 수 있게 된다. head는 neck에서 전달된 각각의 특징을 바탕으로 객체의 위치, 클래스 확률, 신뢰도 등을 예측하며, 전달된 다중 해상도의 특징을 이용하여 최종 객체 검출이 수행된다. 이를 위해, 객체의 바운딩 박스 중심점이 위치한 픽셀을 긍정 샘플(positive sample)로 정의하고, 이 픽셀을 중심으로부터 객체의 경계까지의 거리를 예측함으로써 객체의 정확한 위치와 크기를 결정하는 anchor-free 방식을 적용함으로써, 객체의 크기나 형태의 제약을 받지 않고 작은 객체의 검출 성능이 향상된다.

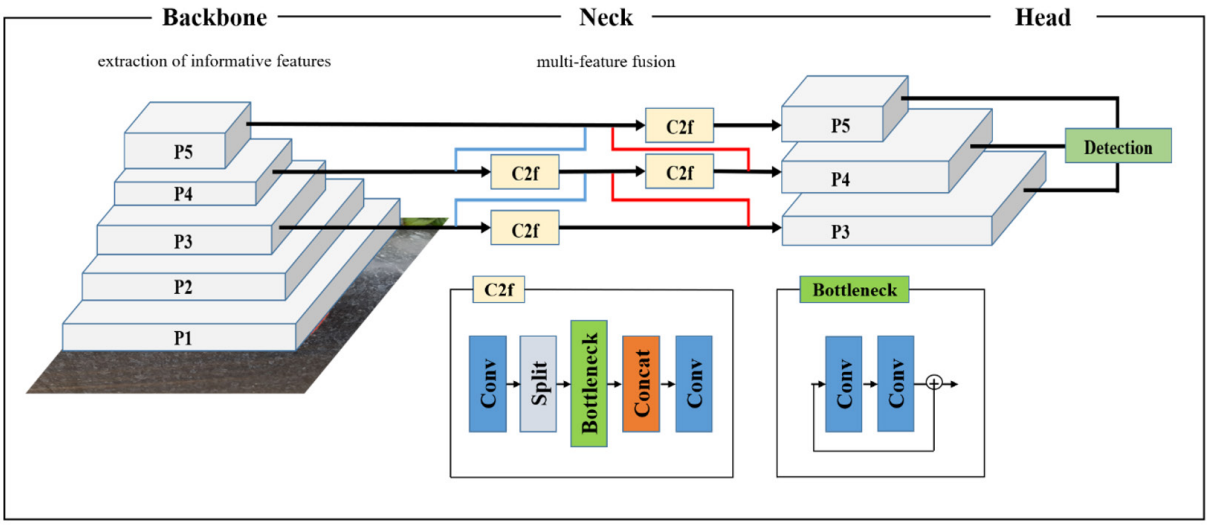
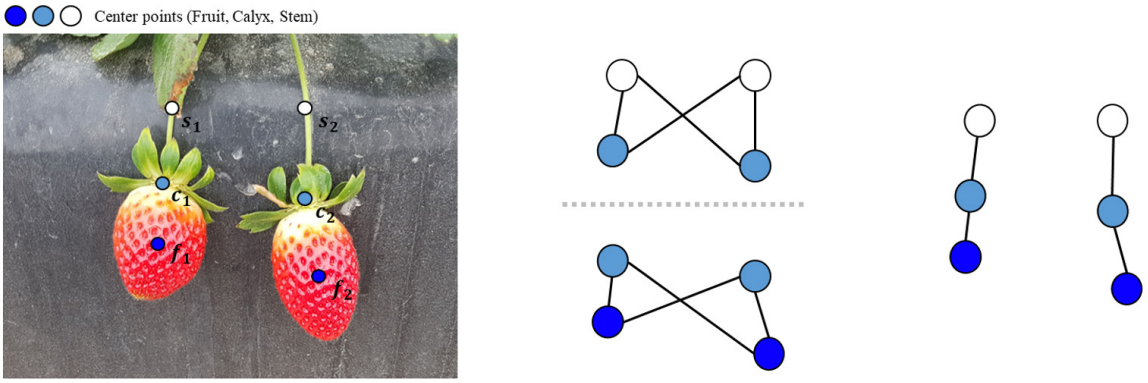


Fig. 2. Architecture of YOLOv8-segmentation model.

헝가리안 알고리즘 기반 객체 군집화

과실에 대한 꽃받침 및 줄기 객체 간의 연관성 매칭은 Fig.3과 같이 헝가리안 알고리즘을 이용하여 수행되었다. 본 연구에서는 객체 간 연관성을 효과적으로 매칭하기 위해 검출 모델로부터 얻어진 각 객체의 마스크(mask) 영역 픽셀 좌표의 평균값을 계산하여 객체의 중심점을 정의하였다. 계산된 중심점은 해당 객체의 기하학적 중심(geometric center)을 나타내며, 이를 기반으로 과실-꽃받침, 꽃받침-줄기 간의 매칭을 수행하였다. 연관성 매칭을 위한 객체 집합은 다음과 같이 정의된다. 식(1)은 과실 객체 집합, 식(2)은 꽃받침 객체 집합, 식(3)은 줄기 객체 집합을 나타내며, 각 객체의 중심점 좌표를 이용하여 유클리디안 거리(euclidean distance) 기반 비용 행렬을 생성한 후, 헝가리안 알고리즘을 적용하여 매칭이 수행된다.



(a) Object representative center points (b) Bipartite matching (c) Parsing result

Fig. 3. Object matching process based on Hungarian algorithm.

$$F = \{f_1, f_2, \dots, f_m\} \tag{1}$$

$$C = \{c_1, c_2, \dots, c_n\} \tag{2}$$

$$S = \{s_1, s_2, \dots, s_p\} \tag{3}$$

Where, F, C and S denote the sets of fruit, calyx, and stem objects, respectively, and f_m, c_n and s_p represent the individual elements within these sets.

매칭 비용을 위한 객체들 간의 유클리디안 거리는 다음과 같이 계산된다. 과실과 꽃받침 객체의 중심점 사이의 거리는 식(4), 꽃 받침과 줄기 객체의 중심점 사이의 거리는 식(5)와 같이 계산된다. 계산된 매칭 비용은 헝가리안 알고리즘을 위한 비용 행렬을 구성하는 데 사용되며, 과실과 꽃 받침 간 비용 행렬은 식(6), 꽃 받침과 줄기 간 비용 행렬은 식(7)로 구성된다.

$$d(f_i, c_j) = \sqrt{\left(x_{f_i} - x_{c_j}\right)^2 + \left(y_{f_i} - y_{c_j}\right)^2} \tag{4}$$

$$d(c_k, s_l) = \sqrt{\left(x_{c_k} - x_{s_l}\right)^2 + \left(y_{c_k} - y_{s_l}\right)^2} \tag{5}$$

Where, $d(f_i,c_j)$ denotes the euclidean distance between the center points of fruit object f_i and calyx object c_j. here x_{fi} and y_{fi} represent the x and y coordinates of the center of f_i, respectively, and x_{cj} and y_{cj} denote those of c_j. similarly, $d(c_k,s_l)$ represents the euclidean distance between the center points of calyx object c_k and stem object s_l, with x_{ck}, y_{ck} and x_{sl}, y_{sl} being the corresponding coordinates.

$$C_{FC} = \begin{bmatrix} d(f_1, c_1) & d(f_1, c_2) & \cdots & d(f_1, c_n) \\ d(f_2, c_1) & d(f_2, c_2) & \cdots & d(f_2, c_n) \\ \vdots & \vdots & \ddots & \vdots \\ d(f_m, c_1) & d(f_m, c_2) & \cdots & d(f_m, c_n) \end{bmatrix} \tag{6}$$

$$C_{CS} = \begin{bmatrix} d(c_1, s_1) & d(c_1, s_2) & \cdots & d(c_1, s_p) \\ d(c_2, s_1) & d(c_2, s_2) & \cdots & d(c_2, s_p) \\ \vdots & \vdots & \ddots & \vdots \\ d(c_n, s_1) & d(c_n, s_2) & \cdots & d(c_n, s_p) \end{bmatrix} \tag{7}$$

Where, C_{FC} is the cost matrix for matching fruit and calyx objects. each entry $C_{FC} = (i,j) = d(f_i, c_j)$ is the euclidean distance between the center of the fruit object f_i and the center of the pedicel object c_j. similarly, C_{CS} is the cost matrix for matching calyx and stem objects, where each element $C_{CS}(k,l) = d(c_k, s_l)$ represents the euclidean distance between the center of the pedicel object c_k and the center of the stem object s_l.

매칭 알고리즘을 통해 과실과 연관된 꽃받침 및 줄기 그룹화가 수행되며, 비용 행렬 C_{FC}와 C_{CS}를 기반으로 최소 매칭 비용을 찾아 최적화된 비용 함수는 식(8)과 같이 표현된다. 최적화된 비용 함수를 기반으로 매칭된 그룹 내 과실, 꽃받침, 줄기의 중심점을 직선으로 연결함으로써 자세 추정이 수행된다.

$$Z = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij} \cdot x_{ij} + \sum_{k=1}^{n}\sum_{l=1}^{p} C_{kl} \cdot y_{kl} \tag{8}$$

Where, C_{ij} denotes the matching cost between fruit object i and calyx object j, and C_{kl} represents the matching cost between calyx k object and stem object l. here, x_{ij} is a binary variable that indicates whether fruit object i and calyx object j are matched (i.e., x_{ij} = 1 if they are matched, and 0 otherwise), and y_{kl} is a binary variable indicating the matching status between calyx object k and stem object j (i.e., y_{kl} = 1 if they are matched, and 0 otherwise).

학습 및 평가 방법

모델 학습을 위한 데이터셋은 공개데이터인 StrawDI (strawberry digital images)를 사용하여 구축되었으며(Pérez-Borrero et al., 2020), 이때 모든 영상은 1,008×756 픽셀의 해상도를 가진다. 데이터는 총 500장의 딸기 영상으로 구성되어 있으며, 각 영상 내에는 복수의 딸기 객체가 1~8개의 범위로 포함되어 있다. 전체 이미지는 모델 학습, 검증, 평가를 위해 7:2:1로 비율로 분리하였으며, 이때 이미지 수는 각각 350장, 100장, 50장이다. 지도학습을 위한 각 이미지의 주석 작업은 Anylabeling을 통해(Nguyen, 2024) 객체별로 polygon 형식으로 수행하였다.

Table 1은 구현에 사용된 학습 환경 및 하이퍼 파라미터를 정리한 것이다. 모델 학습은 300 에폭(epoch) 동안 수행하였으며, 과적합(overfitting)을 방지하기 위해 검증 데이터의 손실값이 최소가 되는 시점의 가중치를 최종 모델로 선택하였다. 학습률(learning late)은 0.01로 설정하였으며, 배치 크기는 16으로 설정하였다. 모델 학습에는 stochastic gradient descent (SGD) 최적화 알고리즘을 사용하였으며, 딥러닝 프레임워크인 Pythorch 1.7.1을 기반으로 구현하였다. 학습은 GPU (GeForce RTX 4060ti, Nvidia, Santa Clara, CA, USA) 환경에서 수행되었다.

Table 1. Specification of implementation environment.

Item	Specification	
Hardware	CPU	Gen Intel(R) Core™ i5-13400F
	Memory	16.0GB
	GPU	NVIDIA GeForce RTX 4060Ti
Software	Program language	Python 3.8
	ANN framework	Pytorch 1. 7. 1
	Image processing	OpenCV 4.7.0
Hyper parameter	Learning rate	0.01
	Batch size	16
	Epochs	300
	Optimizer	SGD

모델의 검출 성능 평가는 검출 모델의 예측 결과와 실제 값 간의 관계를 나타내는 지표인 혼동 행렬(confusion matrix)을 기반으로 수행되었으며, true positive (TP), false positive (FP), true negative (TN), false negative (FN)의 4가지 항목으로 구성된다. 이를 기반으로 정밀도(precision), 재현율(recall) 및 mAP (mean average precision)를 각각 식(9), (10), (11)와 같이 계산하였으며, 이때 정밀도는 모델이 검출한 객체 중 실제 객체인 비율을 나타내며, 재현율은 실제 객체 중 모델이 정확히 검출한 비율을 평가하는 지표이다. mAP는 검출 성능을 종합적으로 평가하는 지표로, 모델의 각 부류에 대한 평균 정밀도를 구한 후, 이를 부류의 수 N으로 평균하여 최종적인 성능을 평가하였다.

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

Where, TP is the number of correctly detected instances, FP is the number of incorrect detections, and FN is the number of missed detections.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{11}$$

Where, AP_i is the average precision for the i-th class and N is the total number of classes.

검출된 핵심 지점의 성능 평가는 예측된 지점과 실제 지점의 x, y의 픽셀 오차에 대해 평가하였으며, 식(12)와 같이 계산된다.

$$Pixel_{error} = |P_{Pred} - P_{GT}| \tag{12}$$

Where, P_{Pred} and P_{GT} denote the predicted and ground truth positions of a keypoint, respectively.

검출된 지점의 정략적인 평가는 percentage of detected keypoint (PDK) 방법을 사용하였다(Zhang et al., 2022). 각 핵심 지점의 검출 여부를 판단하기 위한 임계값은 식(13)과 같이 정의하였다. 최종적으로, 임계값 내에 포함되는 예측된 핵심 지점의 비율을 전체 실제 핵심 지점의 수에 대해 산출하여 해당 임계값에서의 PDK를 계산하였으며, 이는 식(14)와 같이 표현된다.

$$Threshold = \alpha \times L \tag{13}$$

Where, α ranges from 0 to 1, and L represents the reference length (e.g., the average distance between the calyx and stem).

$$PDK@Threshold = \frac{1}{N}\sum_{i=1}^{N} r_i \times 100\% \tag{14}$$

Where, N is the total number of ground truth keypoints, and r_i is a binary variable equal to 1 if the euclidean distance between the i-th predicted keypoint and the corresponding ground truth keypoint is less than or equal to *Threshold*, and 0 otherwise.

Results and Discussion

모델 학습

모델의 학습 횟수에 따른 손실 결과는 Fig. 5과 같다. 학습 손실(train loss)은 학습 반복에 따라 점차 감소하여 0.7 수준으로 수렴하는 경향을 보이는 반면, 검증 손실(validation loss)은 50회 반복 시점에서 0.9 수준을 유지하였다. 따라서, 검증 데이터의 손실이 최소인 50회 반복 시의 모델 가중치를 최종적으로 선정하였으며, 이를 이용하여 딸기의 주요 지점인 과실, 꽃받침, 줄기 검출을 수행하였다.

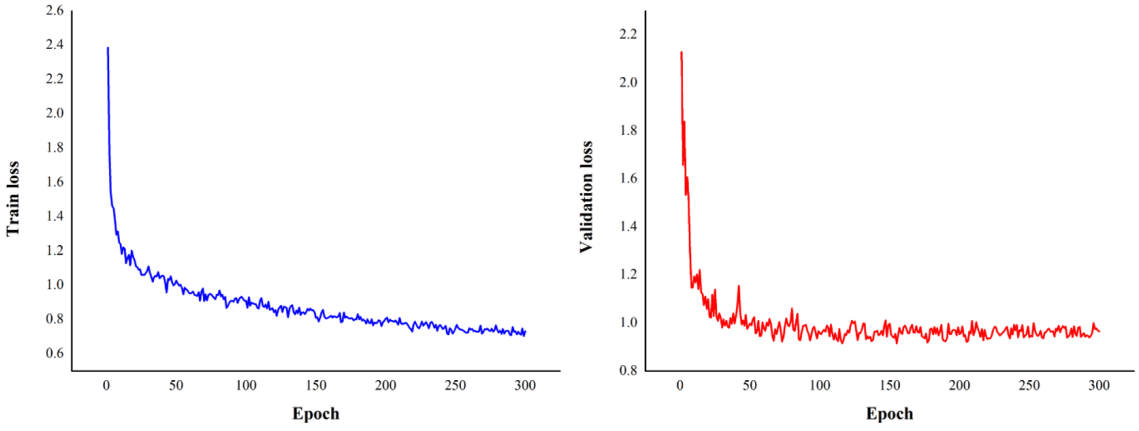


Fig. 5. Loss curves for training and validation sets by epoch

연관성 매칭 및 자세 추정

Fig.6은 이미지 내에서 과실, 꽃받침, 줄기 객체의 구분 및 동일 객체 내 부류 간 매칭 결과를 보여준다. 완숙 딸기의 검출은 매칭이 수행된 여러 딸기 그룹 중, 과실 영역을 색상 범위(hue), 채도(saturation), 명도(value) 색 공간으로 변환했을 때, 빨간색을 나타내는 색상 범위를 만족하는 픽셀이 해당 과실 마스크 영역 내에서 전체 픽셀의 70% 이상을 차지하는 그룹만을 시각적으로 나타냈다. 이를 바탕으로, 각 부류의 중심점을 직선으로 연결하여 완숙 딸기의 자세를 효과적으로 추정하였다. 그러나 매칭 과정에서 객체 간 경계가 모호하거나 객체들이 겹치는 경우 일부 매칭 오류가 관찰되었다. 이는 본 연구에서 중심점 간의 거리 정보만을 사용하여 매칭을 수행한 결과로, 객체의 모양이나 방향 등 추가적인 시각적 정보를 반영하지 않은 데 기인한 것으로 판단된다.

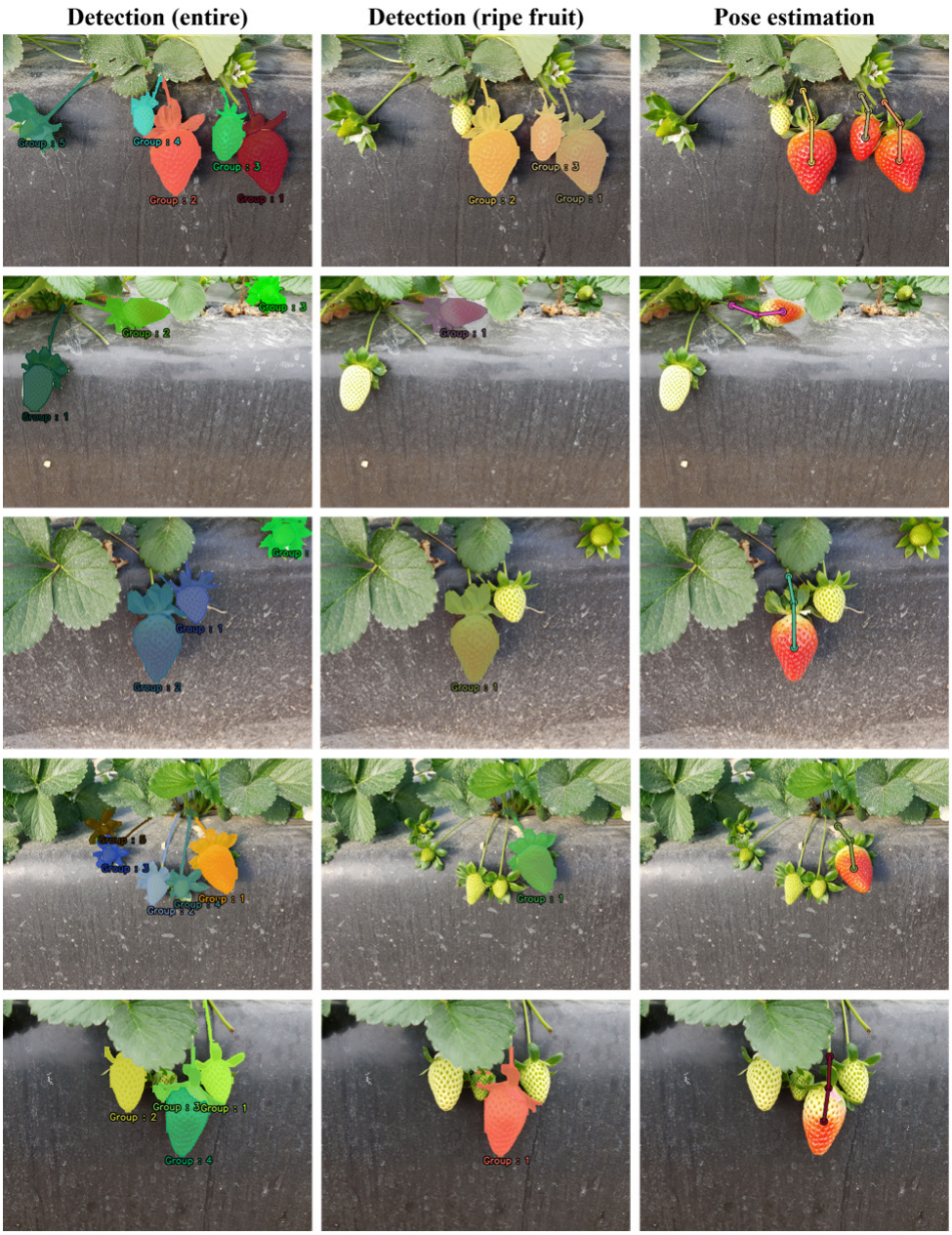


Fig. 6. Strawberry detection results using the matching algorithm

성능 평가

Table 2는 과실, 꽃받침, 줄기 객체의 개별적인 검출 성능 평가 결과이다. 과실 객체에 경우 정밀도 0.94, 재현율 0.96로 가장 높은 검출 정확도를 나타냈으며, AP@0.25에서 0.98, AP@0.75에서는 0.98로 관찰되어 모델에서 가장 높은 성능을 보였다. 꽃받침 객체는 정밀도 0.88, 재현율 0.89로 과실 객체보다는 다소 낮은 성능을 보였으나, AP@0.25에서 0.92, AP@0.5에서는 0.92로 여전히 높은 성능을 보였다. 반면, 줄기 객체는 정밀도 0.76, 재현율 0.78로 나타났으며, AP@0.25에서 0.80, AP@0.75에서는 0.78로 나타났다. 전체 객체에 대한 평균 검출 성능은 mAP@0.25에서 0.90, mAP@0.75에서는 0.88로 대부분의 객체를 검출할 수 있었다. 그러나, 줄기 객체의 경우 다른 객체들보다 현저히 낮은 검출 정확도를 보였으며, 이는 줄기가 다른 객체들에 비해 상대적으로 작은 크기와 객체들 간의 겹침으로 인해 발생한 성능 저하로 판단된다.

Table 2. Detection performance evaluation of strawberry components.

	Precision	Recall	AP@0.25	AP@0.5	AP@0.75
Fruit	0.94	0.96	0.98	0.99	0.98
Calyx	0.88	0.89	0.92	0.92	0.89
Stem	0.76	0.78	0.80	0.79	0.78
All	0.88	0.89	0.90	0.89	0.88

Table 3는 예측된 과실, 꽃받침, 줄기 3지점에 대한 픽셀 좌표의 x, y 오차를 평균 표준편차로 나타낸 결과이다. 오차 결과, 과실의 경우 x축 오차가 4.00±2.55 픽셀, y축 오차가 6.11±4.27 픽셀로 나타나, 비교적 낮은 오차를 보였다. 반면, 꽃받침의 경우 x축 오차는 22.74±35.53 픽셀, y축 오차는 19.93±37.32 픽셀로 관찰되었으며, 줄기는 x축 21.27±37.38 픽셀, y축 13.50±18.59 픽셀로 측정되었다. 전체 평균 오차는 x축에서 16.00±10.42 픽셀, y축에서 13.18±6.92 픽셀로 나타났다. 과실 영역에서는 상대적으로 안정적인 성능을 보였지만 꽃받침과 줄기 영역에서는 상대적으로 오차가 큰 것으로 확인되었다. 그러나, 위치의 세부적인 부분에서 불확실성이 존재할 수 있으나, 예측된 결과는 각 부류의 대략적인 지점의 검출은 가능하였다.

Table 3. Estimated pixel error of coordinates X and Y points.

Class	X pixel	Y pixel
Fruit	4.00 ± 2.55[x)]	6.11 ± 4.27
Clayx	22.74 ± 35.53	19.93 ± 37.32
Stem	21.27 ± 37.38	13.50 ± 18.59
Average	16.00 ± 10.42	13.18 ± 6.92

[x]Mean±standard deviation

Fig.7은 검출 지점의 정략적 평가를 수행한 PDK 결과이다. 과실의 경우 PDK@0.2에서 22.09%, PDK@0.3에서는 31.4%, PDK@0.5는 34.88%의 검출률을 보였다. 특히, PDK@1.0에서는 80.62%까지 상승함으로써, 영상 내 과실이 다른 객체에 비해 명확하게 검출됨을 확인할 수 있었다. 반면, 꽃받침은 PDK@0.2가 10%, PDK@0.3에서 12.86%, PDK@0.5에서 24.29%의 낮은 성능을 보였고, 최종적으로 PDK@1.0에서는 39.05%로 관찰되었다. 줄기의 경우도 PDK@0.2에서 8.33%, PDK@0.3에서는 25%, 최종적으로 PDK@1.0에서 58.33%의 검출률로 나타났으며, 평균적으로 PDK@0.5에서는 59.33%을 보였다. 이러한 결과는, 과실의 중심부의 경우 상대적으로 노출이 많아 검출 성능이 우수한 반면, 꽃받침 및 줄기는 복잡한 배경 및 객체 간 겹침 등으로 인해 검출률이 낮게 나타난 것으로 판단된다. 따라서, 실제 수확 환경에서는 객체 중심 위치를 우선적으로 활용하고, 주변 환경에 의한 노이즈나 폐색 문제에 대해서는 추가적인 보완이 필요하다고 판단된다.

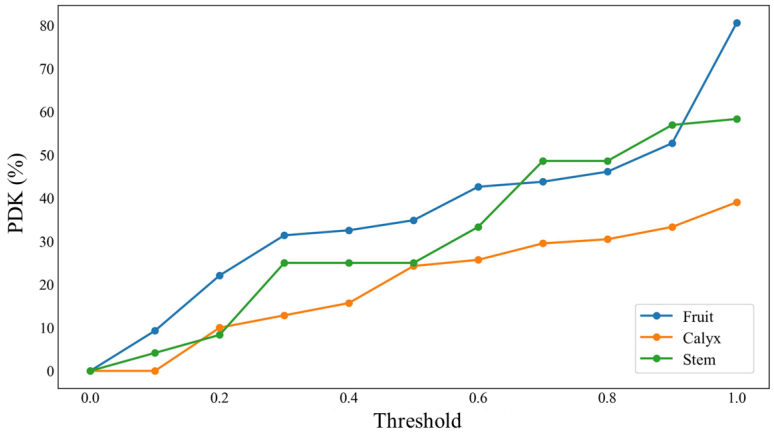


Fig. 7. PDK curves by threshold for fruit, calyx, and stem.

Conclusion

본 연구에서는 딸기 수확 로봇에 활용 가능한 인식 정보 제공을 위해 딥러닝 기반의 YOLOv8-segmentation 모델과 헝가리안 알고리즘을 이용하여 딸기의 과실과 연관된 꽃받침, 줄기 매칭을 통해 자세 추정을 수행하였으며, 모델의 전반적인 검출 성능과 핵심 지점의 픽셀 오차 및 정략적 성능에 대해 평가하였다.

학습에 사용된 데이터는 StrawDI를 이용하여 딸기의 과실, 꽃받침, 줄기 검출이 수행되었다. 모델의 학습 후 헝가리안 알고리즘을 통해 딸기의 각 요소를 매칭하였으며, 이미지 처리 기술을 통해 완숙 딸기의 그룹만을 시각화 하여 그룹 내 각 부류의 중심점을 직선으로 연결함으로써 딸기의 자세 추정이 수행되었다.

전체 객체 검출 성능은 mAP@0.5에서 0.89로 나타났으며, 검출된 지점의 픽셀 오차는 x, y 좌표에서 각각 평균 16.00±10.42, 13.18±6.92로 관찰되었다. 또한, 정략적 평가 지표로 사용한 PDK는 평균 59.33%의 검출률을 나타냈다. 전반적으로 대부분의 객체를 성공적으로 검출하였으나, 꽃받침과 줄기의 경우 일부 영상에서 오검출이 발생하였다. 이는 과실의 중심부가 상대적으로 명확하게 노출되어 정밀한 검출이 가능했던 반면, 꽃받침 및 줄기는 복잡한 배경 및 주변 객체와의 겹침 등으로 인해 오검출 및 누락이 발생하여 상대적으로 검출 성능이 낮아진 것으로 판단된다. 또한, 연구에서 제안한 매칭 알고리즘은 객체 간의 거리 정보를 이용하여 과실과 꽃받침 및 줄기를 그룹화하는 방식으로, 과실과 줄기가 서로 교차하거나 근접하여 복잡한 형태로 얽힌 경우가 많은 경우, 중심점 간 거리만으로는 정확한 대응 관계를 매칭하는 데 한계가 존재한다.

따라서, 주변 노이즈 및 폐색으로 인한 검출 저하 문제 해결을 위해 향후 연구에서는 꽃받침과 줄기의 검출 정확도 향상을 위한 추가적인 데이터 수집과 공간적 거리 외에도 형태학적 특성, 방향성 정보 또는 객체 간의 구조적 연결 정보를 추가적으로 고려하는 매칭 알고리즘의 개선 및 3D 정보 활용 등을 통해 실제 자동 수확을 위한 인식 기술 개발에 기여할 수 있을 것으로 판단된다

Acknowledgements

이 연구는 2024년도 산업통상자원부 및 한국산업기술기획평가원(KEIT) 연구비 지원에 의한 연구임(RS-2024-00443366)

References

An Q, Wang K, Li Z, Song C, Tang X, Song J. 2022. Real-time monitoring method of strawberry fruit growth state based on YOLO improved model. IEEE Access 10:124363-124372.

Cao Z, Simon T, Wei SE, Sheikh Y. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition 7291-7299.

Chai S, Xianjun Y, Fanzhe D. 2024. Research on strawberry fruit detection based on YOLOv8. 2024 3rd International Conference on Artificial Intelligence and Computer Information Technology IEEE. 1-4.

Dimeas F, Sako DV, Moulianitis VC, Aspragathos NA. 2015. Design and fuzzy control of a robotic gripper for efficient strawberry harvesting. Robotica 33(5):1085-1098.

Jocher G, Chaurasia A, Qiu J. 2023. Ultralytics YOLOv8 (version 8.0. 0). Accessed in https://github.com/ultralytics/ultralytics.

Jang M, Hwang Y. 2024. Tomato pose estimation using the association of tomato body and sepal. Computers and Electronics in Agriculture 221:108961.

KOSIS (Korean Statistical Information Service). 2022. Agricultural and forestry production statistics. [in Korean]

KREI (Korea Rural Economic Institute). 2023. Agricultural outlook 2023:Innovation and Future of Agriculture and Rural Areas 1. [in Korean]

Kang SW, Yun JH, Jeong YS, Kim KC, Lee DH. 2024. Key-point detection of fruit for automatic harvesting of oriental melon. Journal of Drive and Control 21(2):65-71. [in Korean]

Lee DH, Im DH, Kim GH, Lee JS, Hong YG, Kim SG, Han SR. 2016. A study on rail type robot system for automatic harvesting in strawberry greenhouse. Korean Society for Agricultural Machinery 21(2):173-173. [in Korean]

Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Guo B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision 10012-10022.

Li Y, Xue J, Zhang M., Yin J, Liu Y, Qiao X, Li Z. 2023a. YOLOv5-ASFF: A multistage strawberry detection algorithm based on improved YOLOv5. Agronomy 13(7):1901.

Li J, Zhu Z, Liu H, Su Y, Deng L. 2023b. Strawberry R-CNN: recognition and counting model of strawberry based on improved faster R-CNN. Ecological Informatics 77:102210.

Munkres J. 1957. Algorithms for the assignment and transportation problems. Journal of the Society for Industrial and Applied Mathematics 5(1):32-38.

Nguyen VA. 2024. AnyLabeling - effortless data labeling with AI support. Accessed in https://github.com/vietanhdev/anylabeling.

Park SJ, Ju CY, Son HI. 2017. Agricultural robotics and automation technology analysis and domestic and international trends. Magazine of the Korean Society of Agricultural Engineers 59(3):17-27. [in Korean]

Pérez-Borrero I, Marín-Santos D, Gegúndez-Arias ME, Cortés-Ancos E. 2020. A fast and accurate deep learning method for strawberry instance segmentation. Computers and Electronics in Agriculture 178:105736.

Yu Y, Zhang K, Liu H, Yang L, Zhang D. 2020. Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot. in IEEE Access 8:116556-116568.

Yang S, Wang W, Gao S, Deng Z. 2023. Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer. Computers and Electronics in Agriculture 215:108360.

Zhang F, Gao J, Zhou H, Zhang J, Zou K, Yuan T. 2022. Three-dimensional pose detection method based on keypoints detection network for tomato bunch. Computers and Electronics in Agriculture 195:106824.