JAME Journal of Agricultural Machinery Engineering

pISSN : 2799-8673
eISSN : 2799-8819

RESEARCH ARTICLE

2차원 주요 지점 검출과 회귀를 이용한 딸기 3차원 자세 추정

최태현[1], 이태신[2], 강승우[2], 성백겸[2], 이대현[2*]
[1]센서아이㈜, [2]충남대학교 농업기계공학과

3D pose estimation of strawberry using 2D keypoint detection and regression

Tae-Hyun Choi[1], Tae-Sin Lee[2], Seung-Woo Kang[2], Baek-Gyeom Sung[2], Dae-Hyun Lee[2*]

[1]Sensoreye Co. Ltd., Daejeon, Republic of Korea
[2]Department of Agricultural Machinery Engineering, Chungnam National University, Daejeon, Republic of Korea

*Corresponding author: leedh7@cnu.ac.kr

Abstract

Strawberry harvesting is a repetitive and labor-intensive task, with yield quality heavily dependent on the operator's skill. To address labor shortages, improve productivity, and ensure consistent harvest quality, the development of harvesting robots is essential. In particular, implementing an autonomous harvesting system requires advanced perception technology capable of accurately identifying the position and condition of target fruit. However, simple object detection alone cannot provide sufficient kinematic information for automation, making keypoint detection of strawberries necessary. This study proposes a 3D pose estimation pipeline that combines a deep learning based keypoint detection model with a multi-layer perceptron (MLP) regression model to provide kinematic information applicable to strawberry harvesting robots. The centers of the fruit, calyx, and pedicel were defined as keypoints, with their 2D positions detected and corresponding depth values predicted using MLP regression. Performance evaluation included percentage of correct keypoints (PCK) for 2D accuracy, and mean per joint position error (MPJPE) and PCK for 3D depth estimation. Experimental results showed average 2D errors of mean PCK@5 pixels of 77.17%. for depth estimation, the average MPJPE was 16.63 mm and PCK@50 mm reached 91.3%. The proposed pipeline demonstrated stable detection of keypoint locations while preserving the overall 3D structure, indicating its potential contribution to the development of perception technologies for autonomous strawberry harvesting.

Keywords: Strawberry, 3D pose estimation, Harvesting robot, Keypoint detection, Multi-layer perceptron, Depth estimation

Journal of Agricultural Machinery Engineering 5(3):99-110

DOI: https://doi.org/10.12972/jame.2025.5.3.3

Received: August 21, 2025
Revised: September 19, 2025
Accepted: September 25, 2025

Introduction

딸기는 국내 시설 원예 작물 중 생산액이 가장 높은 작물 중 하나로 농가 소득에서 차지하는 비중이 매우 크며, 주요 수출량 또한 지속적으로 증가하고 있다(Lee et al., 2025). 그러나, 국내 농업 분야 전반에서 노동력 부족 문제가 심화되고 있으며, 특히 딸기 수확 작업은 반복적이고 고강도의 노동이 요구되며(Xiong et al., 2020), 과육의 연약한 특성으로 인해 수확 과정 중 손상이 빈번히 발생하여 품질 저하를 초래한다. 이로 인한 손상은 상품성을 저하시켜 경제적 손실을 유발하므로, 과실과의 직접 접촉을 최소화하는 방식으로 수확이 이루어지고 있다(Yu et al., 2020). 전통적으로 딸기 수확은 꽃자루(pedicel)를 엄지와 검지 사이에 가볍게 끼워 당기는 방식으로 시행되나, 이 방법은 작업자의 숙련도에 크게 의존하고 노동 강도가 높아 수확 자동화를 위한 기술 개발이 요구된다(RDA, 2019).

최근 스마트 농업 기술의 발전과 함께 딸기 수확 자동화에 대한 관심이 높아지고 있으나, 실제 재배 현장에서는 여전히 수작업에 의존하고 있는 실정이다(Dai et al., 2025). 따라서, 수확 로봇의 개발은 노동력 부족 문제 해결과 생산성 향상 및 수확 품질의 일관성을 확보하기 위한 핵심 과제로 주목받고 있다. 수확 로봇은 크게 인식, 주행, 파지 등 복합적인 기술로 구성되며, 그중 인식 기술은 자동 수확 시스템 구현을 위해 필수적으로 선행되어야 하는 핵심 요소이다(Kang et al., 2024). 최근 합성곱 신경망 기반의 심층학습(deep learning) 기술을 활용하여 실시간 딸기 검출(Zhang et al., 2022), 성숙도 분류(Wang et al., 2024), 생육 모니터링(An et al., 2022), 주요 지점 검출(Ma et al., 2025) 등 다양한 인식 기술 연구에서 우수한 성능을 보이고 있다. 특히, 수확에 필요한 정보 제공을 위해 대상 작물의 주요 지점 검출이 자세 추정(pose estimation) 연구에서 활발히 적용되고 있다.

자세 추정은 특정 관절의 지점을 찾는 문제로, 사람과 같이 관절의 정의가 명확한 객체에서 주로 적용되고 있다. 그러나, 농작물에 경우 품종 및 생장 단계에 따라 형상이 다양하고 주요 지점의 정의가 모호하여 자세 추정의 적용이 어려운 문제가 존재한다. 반면, 과채류는 줄기부와 과실부가 쌍을 이루는 구조를 가지므로 주요 지점의 정의가 비교적 용이하여 기술 적용이 시도되고 있다(Kang et al., 2024). 과채류 자세 추정의 선행 연구로는 다중 토마토 객체의 과실-꽃자루 쌍을 동시에 검출을 통한 토마토의 전반적인 자세 추정(Kim et al., 2023), encoder-decoder 기반의 딥러닝 모델을 사용하여 참외의 주요 지점 검출을 통한 자세 추정(Kang et al., 2025), 실시간 영상에서 딸기의 수확 지점을 결정하기 위해 R-YOLO 모델 개발을 통한 딸기의 자세 추정(Yu et al., 2020)과 YOLO 모델을 개선하여 딸기 꽃자루의 자세를 추정한 연구(Meng et al., 2025) 등이 있다. 이러한 여러 선행 연구를 통해 과채류의 정밀한 자세 추정이 가능함이 확인되었으나, 실제 로봇의 수확 동작을 위해서는 3차원의 기구학적 정보가 요구된다. 기존의 연구들은 대부분 2차원 주요 지점 검출에 국한되거나 단일 깊이 센서 기반 좌표 산출에 의존하여, 폐색(occlusion)과 깊이 정보의 왜곡 또는 누락이 발생될 경우 안정적인 3차원 정보를 제공하지 못하는 한계가 존재한다. 인간의 자세 추정 분야에서는 2차원 관절 좌표로부터 3차원 깊이 정보 추정을 위해 다층 퍼셉트론(multi-layer perceptron, MLP) 기반의 선형 회귀 방식이 적용되고 있으며(Martinez et al., 2017), 과채류 또한 과실-줄기의 쌍이 선형적으로 연결된 구조적 특징을 가지고 있어 회귀를 통한 깊이 정보 추정이 가능하다.

따라서, 본 연구에서는 2차원 주요 지점 검출 모델과 다층 퍼셉트론 모델을 이용하여 과실-줄기 쌍의 3차원 자세추정을 수행하였다. 제안된 방법은 검출된 주요 지점의 2차원 좌표를 이용해 깊이 값을 추정함으로써, 깊이 정보의 누락이나 폐색 환경에서도 강인한 장점을 가지며, 좌표 기반 오차 지표를 통해 정량적 성능을 평가하고 분석하였다.

Materials and Methods

딸기 자세 추정 개요

Fig. 1은 본 연구의 3차원 자세 추정 방법을 단계별로 나타낸 그림으로 크게 2차원 주요 지점 검출 단계와 3차원 깊이 예측 단계로 구성된다. 먼저, RGB 영상으로부터 딥러닝 기반의 주요 지점 검출 모델을 이용하여 과실, 꽃받침, 꽃자루 세 지점의 2차원 중심 좌표를 검출하고, 이를 카메라 내부 파라미터에 따라 단위 좌표계로 정규화한다.

3차원 깊이 예측 단계에서는 정규화된 세 지점의 좌표(x, y)를 연결하여 형성된 6차원 입력 벡터(pose vector)가 회귀 모델의 입력으로 사용된다. 학습 단계에서 과실 좌표를 기준점(reference point)으로 설정하고, 꽃받침과 꽃자루의 좌표가 깊이 축에서 갖는 상대적 위치 차이를 상대 깊이(relative depth) 값으로 회귀한다. 추론 단계에서는 예측된 상대 깊이를 과실의 깊이와 결합하여 세 주요 지점의 절대 깊이를 복원하였다.

이러한 방식은 과실-꽃받침-꽃자루가 일정한 구조적 배치를 이루는 딸기의 형태적 특성을 반영한 것으로, 꽃자루는 일반적으로 과실의 상부에, 꽃받침은 과실과 꽃자루 사이에 위치하는 구조적 제약 조건이 2차원 좌표 패턴에 투영된다는 점에 근거한다. 따라서, 학습을 통해 MLP는 2차원 좌표계에서 관찰되는 상대적 배치를 근사함으로써 깊이 정보를 추정할 수 있다. 이를 통해, 센서의 한계나 깊이 왜곡 또는 누락이 발생하더라도 깊이 정보를 예측함으로써 수확 로봇에 요구되는 3차원 자세 정보를 안정적으로 제공할 수 있다.

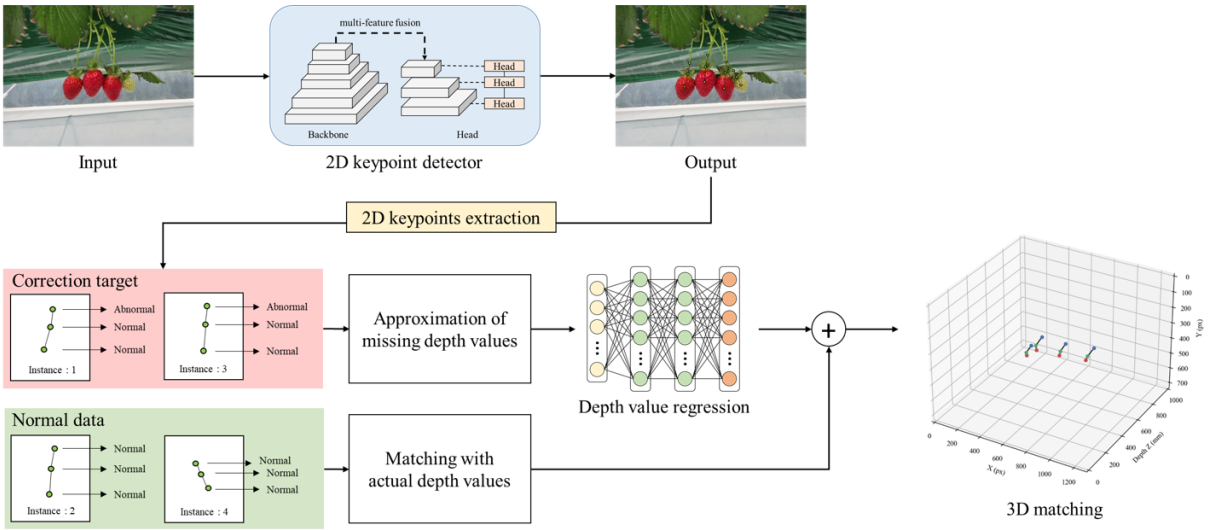

Fig. 1. Proposed 3d pose estimation of strawberry fruit-pedicel pair in this study.

학습 데이터

본 연구에서 사용된 딸기 영상 데이터는 2025년 4월 10일부터 4월 16일까지 충남대학교 농업생명과학대학에 위치한 딸기 시설 온실에서 수집되었다. 수집 대상은 국내 주요 품종 중 하나인 설향(*Seolhyang*)이며, 재배 환경은 Fig. 2(a)와 같이 수경재배 기반의 고설재배 시스템으로 구성되어 있다. 영상 수집 장치는 RGB-D 카메라(Femto bolt, Orbbec, Shenzhen, Guangdong, China)를 활용하여 Fig. 2(b)와 같이 RGB 영상과 함께 정합된 깊이 정보를 동시에 수집하였으며, 카메라의 주요

사양은 Table. 1과 같다. 영상 촬영 조건은 실제 로봇 수확 시점을 고려하여, 카메라는 배지 높이를 기준으로 지면으로부터 약 90 cm의 위치, 배지 상단에서 수직 방향으로 약 50~60 cm 거리에서 영상을 수집하였다. 각 영상에는 딸기 객체를 1~10개의 범위를 포함하고 있으며, 해상도 1280×720의 영상이 총 556장 수집되었다.

(a) Strawberry greenhouse (b) Acquired images

Fig. 2. Experimental conditions for collecting the images: (a) strawberry green house and (b) acquired images.

Table 1. Specification of RGB-D camera.

Item	Specification
Depth technology	Time of flight
Depth range	0.25~5.46m
Depth resolution/FPS	640×576@30fps
Depth FOV	H 120˚ V120˚
RGB resolution/FPS	1280 × 720@30fps
RGB FOV	H 80˚ V 51˚

학습 데이터 구축

딸기의 주요 지점 검출을 위해 Fig. 3과 같이 과실 중심, 꽃받침 중심, 꽃자루 중간 부분을 기하학적 지점으로 선정하였다. 학습 데이터셋 구축을 위해 각 이미지에서 정의된 지점에 대해 keypoint 형식으로 수동 주석작업을 수행하였다. 주석작업은 오픈 라이브러리 Roboflow를 사용하였으며(Dwyer et al., 2025), 모델 학습을 위해 학습, 검증, 평가 비율을 7:2:1로 각각 389, 111, 56장으로 분할하였다. MLP 모델의 학습 데이터는 단일 과실 객체를 입력으로 사용하므로, 각 분할 된 데이터셋에서 이미지의 과실 수만큼 각각 1775, 434, 204개의 개별 인스턴스로 추출하여 구성하였다.

주요 지점 검출 모델

딸기의 주요 지점 검출을 위한 모델은 여러 과채류 자세 추정에서 실시간 성능과 정확도에서 높은 성능을 보인 YOLOv8-pose 모델을 사용하였다(Jocher et al., 2023). 모델은 단일 단계(single-stage) 회귀 기반의 자세 추정 구조로 Fig. 4와 같이 backbone, neck, head로 구성되며, 각 구성 요소는 특징 추출, 특징 결합, 최종 검출을 담당한다. 먼저 backbone 에서는 C2f 블록, SPPF (spatial pyramid pooling-fast)를 순차적으로 적용하여 다양한 스케일의 특징 맵을 추출한다. 이 특징 맵들은 neck에서 상향 및 하향 경로로 결합 및 분리되어, 세부 공간 정보와 고차원의 정보를 통합한 다중 해상도 표현으로 융합된

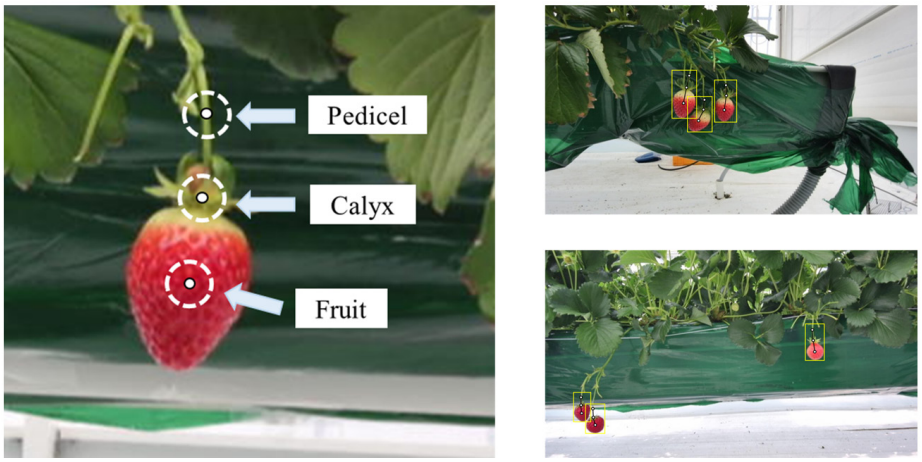

Fig. 3. Representative geometrical points of strawberry and annotated images.

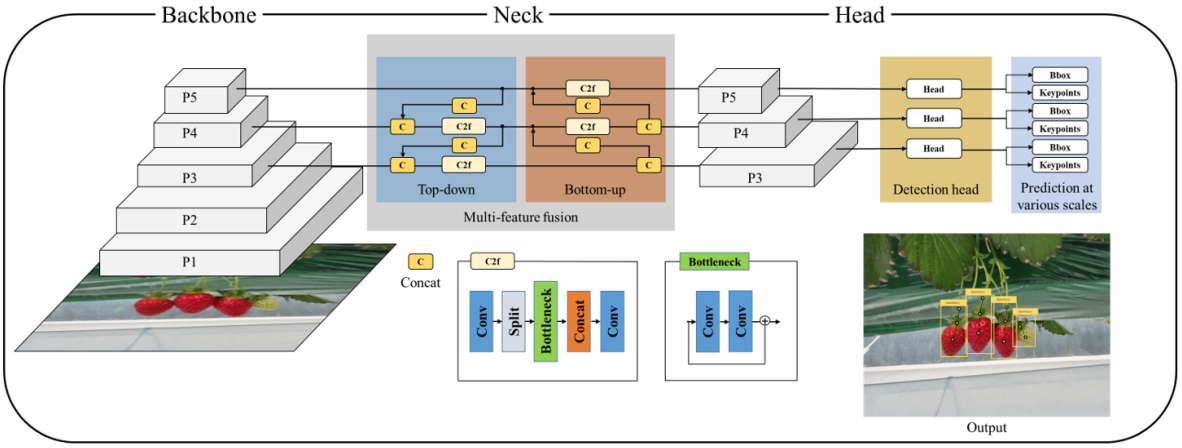

Fig. 4. Architecture of YOLOv8-pose used for estimation the 2d strawberry pose.

다. head는 detection head와 pose head로 분리되며, detection head 에서는 앵커프리(anchor-free) 방식으로 경계 상자(bounding box, bbox) 좌표, 객체 신뢰도, 클래스 확률을 회귀하고, pose head는 검출된 각 bbox를 기준으로 각 키포인트의 x, y 좌표와 신뢰도 값을 직접 회귀하여 절대 픽셀 좌표를 산출한다. 후처리 단계에서는 NMS (non-maximum suppression)를 통해 bbox를 정제하고, 해당 객체의 pose head 출력을 최종 keypoint로 사용함으로써 회귀 기반의 실시간 keypoint detection이 수행된다.

3차원 회귀 모델

딸기 주요 지점의 3차원 깊이 예측을 위하여 본 연구에서는 인간의 깊이 정보 추정을 위해 사용되었던 다층 퍼셉트론 구조의 선형 회귀 모델인 Simple Baseline MLP를 활용하였다(Martinez et al., 2017). 모델의 구조는 Fig. 5와 같이 선형 변환(linear layer), 배치 정규화(batch normalization), ReLU 활성화 함수, dropout을 순차 적용한 residual 블록으로 구성되며, 두 블록의 출력을 입력과 더하는 residual 구조가 세 차례 반복한다. 또한, 입력 직후 및 출력 직전에 각각 하나의 선형 계층이 추가된 총 8개의 선형 계층으로 네트워크 구성되었다.

영상으로부터 검출된 과실, 꽃받침, 꽃자루의 2차원 픽셀 좌표를 카메라 내부 파라미터를 이용하여 식(1)과 같이 카메라 평면상 단위 좌표로 정규화 하였다. 여기서, u, v는 영상 좌표계에서의 픽셀 좌표를 의미하며, c_x, c_y는 카메라 중심 좌표, f_x, f_y

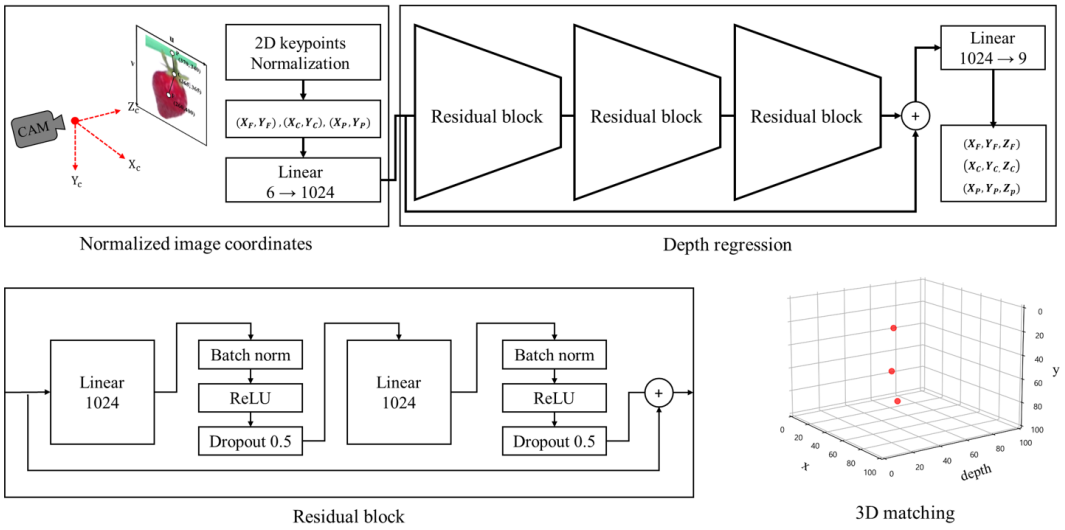

Fig. 5. Architecture of depth estimation model for strawberry keypoints.

는 카메라 초점 거리(focal length)를 나타낸다. 이를 통해 영상 좌표를 카메라 평면 상 단위 좌표계로 변환함으로써, 카메라 해상도 및 시점 변화에 무관한 표준화된 2차원 좌표 X, Y를 얻을 수 있다. 정규화된 세 주요 지점(F, C, P)의 좌표는 식(2)와 같이 6차원 입력 벡터로 구성되며, 이는 MLP 모델의 입력으로 사용된다.

$$X = \frac{u - c_x}{f_x}, \quad Y = \frac{v - c_y}{f_y} \tag{1}$$

Where, u, v are pixel coordinates in the image coordinate system; c_x, c_y are the principal point (camera center) coordinates; f_x, f_y are focal lengths along the x and y axes in pixel units; X, Y are normalized camera plane coordinates.

$$Normalized\ 2d\ keypoints = \left[X_F, Y_F, X_C, Y_C, X_p, Y_p\right]^T \tag{2}$$

Where, F, C, P denote the fruit, calyx, and pedicel keypoints, respectively; T denotes the transpose.

학습 단계에서는 과실 관절을 기준으로 깊이를 고정하여, MLP 모델이 꽃받침과 꽃자루의 상대 깊이를 회귀하도록 하였다. 이를 위해 각 관절의 정규화 좌표 차이를 계산하여 과실 기준 상대좌표를 정의하였으며, 식(3)과 같이 표현된다. 여기서, X_j, Y_j는 꽃받침 또는 꽃자루의 정규화 좌표를, X_F, Y_F는 과실의 정규화 좌표를 의미한다.

$$\Delta X_j = X_j - X_F, \ \Delta Y_j = Y_j - Y_F \quad (j \in \{C, P\}) \tag{3}$$

Where, ΔX_j, ΔY_j are fruit-referenced coordinate offsets used as inputs to the MLP.

모델 학습

2차원 주요 지점 검출의 모델 학습은 500회 동안 반복 수행하였으며, 과적합(overfitting) 방지를 위해 검증 데이터의 손실

값이 최소가 되는 시점의 가중치를 최종 모델로 선택하였다. 학습률(learning late)는 0.01로 설정하였으며, 배치 크기는 16으로 설정하였다. 모델학습에는 SGE (stochastic gradient descent) 알고리즘을 통해 최적화되었다. 3차원 깊이 예측 모델의 학습은 200회동안 수행하였으며, 손실함수 MSE (mean squared error)를 통해 최적화되며 식(4)같다. 여기서, $\widehat{\Delta Z_i}$는 모델이 예측한 상대 깊이, ΔZ_i는 실제 상대 깊이 값, N은 배치 내 인스턴스 수를 의미한다. 이 과정을 통해 동일한 픽셀 이동량 근처에서는 카메라로부터 실제 깊이 값이 평균적으로 유사하다는 통계적 패턴이 학습된다. 추론단계에서는 예측된 상대 깊이에 기준 깊이를 가산하여 각 관절의 절대 깊이가 복원된다. 딥러닝 프레임워크는 pytorch 1.7.1 기반으로 구현되었으며, 학습은 GPU (GeForce RTX 4060ti, Nvidia, Santa Clara, CA, USA) 환경에서 수행되었다.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\Delta Z_i} - \Delta Z_i\right)^2 \tag{4}$$

Where, $\Delta\widehat{Z_{i}}$ is the predicted relative depth; ΔZ_i is the ground truth relative depth; N is the number of instances in a mini-batch.

성능평가

모델의 검출 성능 평가는 혼동 행렬(confusion matrix)에 기반하여 TP (true positive), FP (false positive), TN (true negative), FN (false negative) 4가지 항목을 통해 정밀도(precision), 재현율(recall), mAP (mean average precision) 지표를 각각 식(5), (6), (7)과 같이 계산하였다. 여기서 정밀도는 모델이 양성으로 예측한 샘플 중 실제 양성의 비율이며, 재현율은 실제 양성 샘플 중 모델이 정확히 검출한 비율을 나타낸다. mAP는 각 클래스 별로 계산된 평균 정밀도를 클래스의 수 N으로 평균한 값으로, 전체 클래스에 대한 검출 성능을 종합적으로 평가하는 지표이다.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{7}$$

Where, TP is the number of true positives; FP is the number of false positives; FN is the number of false negatives; N is the number of classes; AP_i is the precision for class N.

주요 지점 검출의 정량적인 평가를 위해 예측된 지점의 좌표와 실제 지점의 x, y의 픽셀 오차를 식(8)과 같이 계산하였으며, 이때 P_{pred}와 P_{GT}는 각각 예측된 관절의 x, y 좌표와 실제 좌표를 의미한다.

$$Pixel_{error} = \left|P_{pred} - P_{GT}\right| \tag{8}$$

Where, P_{pred} is the predicted 2d keypoint in pixels, P_{GT} is the ground truth 2d keypoint in pixels.

자세 추정에서 대표적인 성능 평가 지표인 PCK (percentage correct keypoint) (Andriluka et al., 2014), MPJPE (mean per joint position error) (Lonescu et al., 2014), OKS (object keypoint similarity) (Kim et al., 2023)를 사용하여 평가를 수행하였으며, 각각 식(9), (10), (11)과 같이 계산하였다. 여기서, PCK는 예측된 관절의 위치가 임계값 T 이내에 존재하는 비율을 나타내며, MPJPE는 전체 관절 위치 오차의 평균을 나타내는 지표로 사용된다. 이때, $\hat{P}_i$와 P_i는 각각 예측 및 실제 관절의 좌표를 나타내며, N는 전체 관절 수이다. OKS는 예측된 관절과 실제 관절 사이의 거리 유사도를 계산하는 지표로, 객체 검출에서 IoU (intersection over union)과 유사한 역할을 한다. 여기서, i는 총 관절의 수, d_i는 예측 관절과 실제 관절 사이의 유클리드 거리, s는 객체의 크기, k_i는 관절의 허용 오차, v_i는 관절의 가시성을 의미한다.

$$PCK@threshold = \frac{1}{N}\sum_{i=1}^{N} 1(\left\|\hat{P}_i - P_i\right\| \leq T) \tag{9}$$

$$MPJPE = \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{P}_i - P_i\right\|_2 \tag{10}$$

Where, N is the number of evaluated keypoints; $P_i^{\wedge}$ and P_i are the predicted and ground truth 2d coordinates of keypoint i.

$$OKS = \frac{\sum_i exp\,(-\frac{d_i^2}{2s^2k_i^2}) \cdot v_i}{\sum_i v_i} \tag{11}$$

Where, d_i is the Euclidean distance between the predicted and ground truth location of keypoint i; s is an object scale; k_i is a keypoint specific tolerance; v_i indicates the visibility of keypoint i.

Results and Discussion

딸기 주요 지점 검출

Fig. 6은 자세 추정의 대표적인 결과로 입력 영상에 객체와 함께 자세를 표현한 시각적 결과와 이를 3차원 자세로 표현한 그래프를 보여준다. 2차원 주요 지점 검출의 경우 과실, 꽃받침, 꽃자루의 예측 좌표가 대부분 실제 위치와의 높은 일치도를 보여주었다. 3차원 매칭 단계에서는 검출된 각 부류의 2차원 x, y 좌표 상 위치에 깊이 정보를 결합하여 주요 지점의 공간적 배치를 시각화 하였다. 그 결과, 꽃자루의 깊이 예측에서 일정 수준의 편차가 관찰되었으나, 대략적인 위치의 검출이 가능하였으며, 전체적인 3차원 구조가 적절히 유지됨을 보여주었다.

성능 평가

Table 2는 자세 추정의 성능 평가를 나타낸 결과이며, 평가는 2차원 주요 지점 검출 및 3차원 깊이 예측 결과를 종합하여 수행하였다. 2차원 주요 지점 검출 모델은 딸기 객체에 대해 정밀도 0.89, 재현율 0.91, mAP@0.5에서 0.85, OKS@0.5는 0.94로 관찰되어 안정적인 검출이 가능함을 보여주었다. OKS의 결과는 수치가 0.72인 토마토의 자세추정에서 OpenPose 모델을 사용한 결과와(Kim et al., 2023), HRNet 모델을 사용한 인간과 동물의 자세 추정의 경우 각각 0.86, 0.93으로 딸기 자세 추정과 비슷한 성능을 보였다(Li et al., 2024; Yu et al., 2021). 인간과 동물의 자세 추정에서 관절의 구조는 더 복잡하지만, 이 연

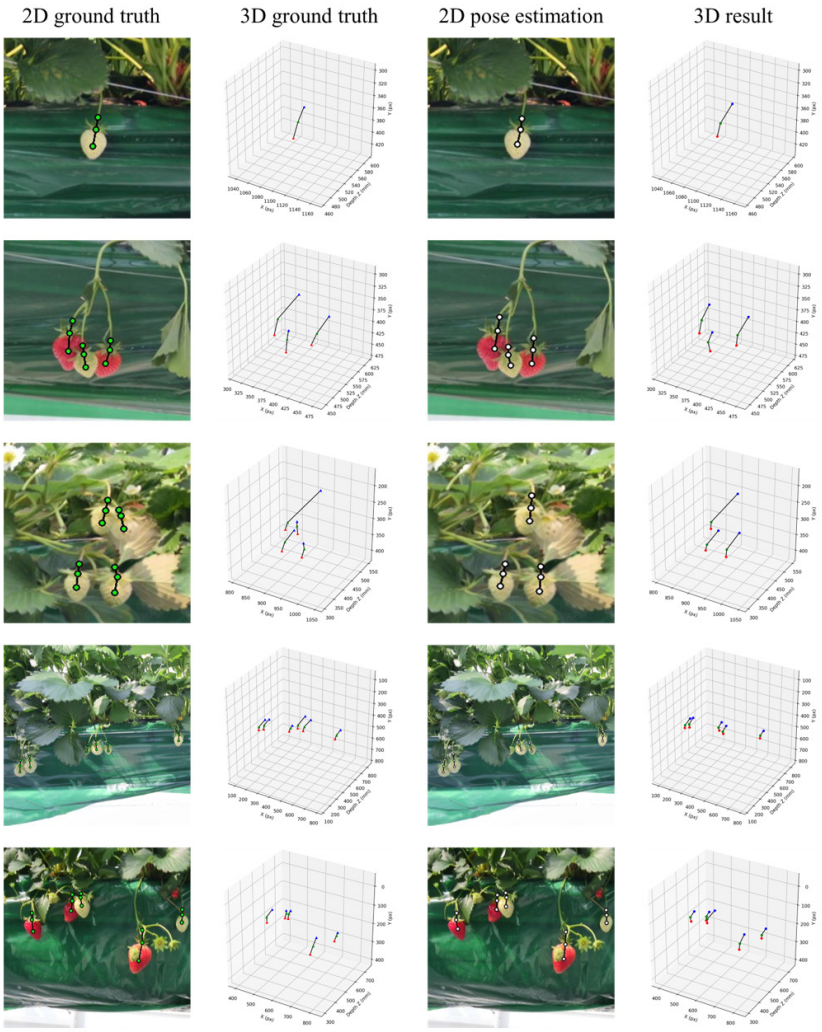

Fig. 6. Representative results of strawberry pose estimation inference.

Table 2. Performance evaluation for three components of strawberry

Class	2d coordinate			Depth	
	x pixel	y pixel	PCK@5 (pixel)	MPJPE (mm)	PCK@50(mm)
Fruit	1.62 ± 1.22[x)]	2.31 ± 2.52	86.96%	0.02 ± 0.01	100%
Calyx	2.10 ± 1.74	2.09 ± 2.51	83.15%	13.40 ± 16.31	97.3%
Pedicel	2.91 ± 2.84	3.71 ± 6.33	61.41%	36.48 ± 5.90	76.5%
Average	2.21 ± 2.12	2.70 ± 4.26	77.17%	16.63 ± 27.33	91.3%

[x]Mean±standard deviation

구의 결과는 기초 데이터가 부족하고 객체 간의 구별이 어려운 딸기에서도 높은 성능을 달성했음을 보여주며, 제안된 모델의 실용 가능성을 보여주었다.

관절별 오차 분석 결과, 과실의 평균 픽셀 오차는 x, y 픽셀에서 각각 1.62, 2.31로 가장 낮았으며, 꽃받침은 각각 2.10, 2.09 픽셀, 꽃자루의 경우 2.91, 3.71 픽셀의 오차를 보였다. 이들 관절에 대한 PCK결과 임계값 5 픽셀에서 과실 86.9%, 꽃받침 83.1%, 꽃자루 61.4%로 나타났다. 3차원 깊이 예측에서는 과실 관절의 깊이를 고정하여 예측했으므로 과실에 대해서는 오

차가 발생하지 않았으며, MPJPE의 경우 꽃받침은 13.40 mm, 꽃자루는 36.48 mm를 보였다, PCK의 경우 임계값 50 mm에서 꽃받침 97.3%, 꽃자루 76.5%로 나타났다. 특히, 꽃자루 깊이 예측에서 관찰된 높은 오차는 단순히 가는 구조적 특성과 픽셀 이동량 기반의 특징만으로는 깊이 변화에 따른 미세 패턴을 충분히 포착하기 어려운 데에서 기인한다. 또한, 영상 해상도에서 꽃자루가 차지하는 영역이 매우 작아 학습 데이터의 유효 표본이 제한적이고, 잎과 배경과의 색상 유사성으로 인해 검출된 좌표 주변에서 특징 분리가 어려워 주변 관절 대비 공간적 상관관계가 뚜렷하지 못한 성능 저하로 판단된다.

Fig. 8은 2차원 주요 지점 검출 및 3차원 깊이 예측 결과에 대한 PCK 곡선을 시각화한 결과로, 2차원 주요 지점 검출은 5픽셀 임계값에서 평균적으로 약 80% 이상의 검출률을 보였으며, 3차원 깊이 예측은 50 mm 임계값에서 평균적으로 약 90% 이상의 정확도를 유지함을 보여주었다. 50 mm를 초과하는 깊이 오차는 과실-꽃받침-꽃자루의 상대 배치가 짧은 길이 척도에서 결정된다는 점을 고려할 때 이 구간의 실패는 대체로 정삭적 변동이 아닌 왜곡 또는 정보 누락으로 인한 것으로 판단된다.

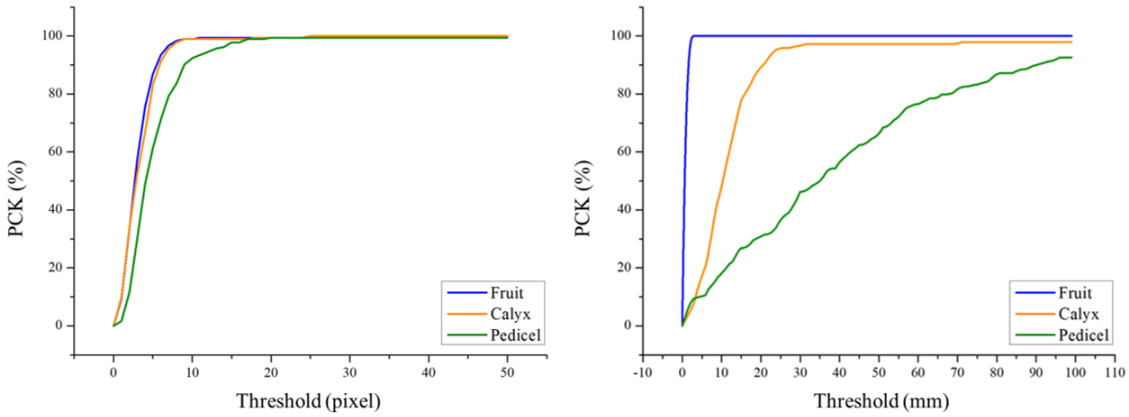

Fig. 8. The PCK curve of 3 keypoints and average value; (Left: 2D keypoint detection, Right: depth estimation)

Conclusion

본 연구에서는 RGB-D 환경에서 주요 지점을 3차원으로 복원할 수 있는 방법을 제시함으로써, 수확 로봇에 인식 정보 제공을 위한 3차원 자세 추정을 수행하였다. 먼저, 주요 지점 검출 모델을 이용하여 딸기의 과실, 꽃받침, 꽃자루 주요 지점의 2차원 픽셀 좌표를 검출하였으며, 검출된 좌표를 카메라 내부 파라미터에 기반한 단위 좌표계로 정규화 하였다. 이후, 검출된 좌표를 입력으로 MLP 회귀 모델을 통해 각 관절의 상대 깊이를 회귀 예측하여 최종 절대 깊이를 추정하였으며, 모델의 검출 성능과 주요 지점의 정량적 성능에 대해 평가하였다.

딸기의 검출 성능은 mAP@0.5에서 0.85로 나타났으며, OKS@0.5는 0.94로 관찰되었다. 세 주요 지점의 평균 픽셀 오차는 x, y 픽셀에서 각각 2.21, 2.70로 나타났으며, PCK@5 픽셀은 평균 77.1%로 관찰되었다. 3차원 깊이 예측에서는 과실의 경우 관절 깊이를 고정하여 입출력이 같아 오차가 없었으며, MPJPE의 결과는 평균 16.63 mm으로, PCK@50 mm는 전체 평균 91.3%으로 관찰되었다.

본 연구는 단일 품종 이미지와 제한된 온실 환경에서 수집된 데이터셋을 기반으로 하여, 환경적 다양성이 부족하다는 한계가 존재한다. 따라서, 향후 연구에서는 다양한 품종, 조도, 재배 환경을 포함한 대규모 데이터셋 구축 및 구조적 특징으로 인한 검출 저하 문제 해결을 위해 공간적 거리 외에도 형태학적 특성, 방향성 정보 또는 객체간 구조적 연결 정보를 추가적으로 고려할 계획이다. 또한, 제안된 자세 추정 방법을 실제 수확 로봇 플랫폼과 통합하여 파지 및 절단 성공률을 정량적으로 검증함으로써 실제 딸기 자동 수확을 위한 인식 기술 개발에 기여할 수 있을 것으로 기대된다.

Acknowledgements

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

References

Andriluka M, Pishchulin L, Gehler P, Schiele B. 2014. 2D human pose estimation: new benchmark and state of the art analysis. In Proceedings of the IEEE Conference on computer Vision and Pattern Recognition 3686-3693.

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.

Dai S, Tao B, Yunjie Z. 2025. Keypoint detection and 3D localization method for ridge-cultivated strawberry harvesting robots. Agriculture 15(4):372.

Dwyer B, Nelson J, Hansen T. 2025. Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. Computer vision.

Ionescu C, Papava D, Olaru V, Sminchisescu C. 2014. Human3.6M: large scale datasets and predictive methods for 3D human sensing in natural environments. IEEE transactions on pattern analysis and machine intelligence 36(7):1325-1339.

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

Kim T, Lee DH, Kim KC, Kim YJ. 2023. 2D pose estimation of multiple tomato fruit-bearing systems for robotic harvesting. Computers and Electronics in Agriculture, 211:108004.

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]

Kang SW, Kim KC, Kim YJ, Lee DH. 2025. Real-time pose estimation of oriental melon fruit-pedicel pairs using weakly localized fruit regions via class activation map. Computers and Electronics in Agriculture 237:110734.

Li R, Yan A, Yang S, He D, Zeng X, Liu H. 2024. Human pose estimation based on efficient and lightweight high-resolution network (EL-HRNet). Sensors, 24(2).

Lee TS, Kang SW, Lee DH. 2025. Pose estimation of ripe fruit for strawberry harvesting robot. Journal of Agricultural Machinery Engineering. 5(1):13-24. [in Korean]

Martinez J, Hossain R, Romero J, Little JJ. 2017. A simple yet effective baseline for 3d human pose estimation. In Proceedings of the IEEE international conference on computer vision 2640-2649.

Ma Z, Dong N, Gu J, Cheng H, Meng Z, Du X. 2025. STRAW-YOLO: A detection method for strawberry fruits targets and key points. Computers and Electronics in Agriculture 230: 109853.

Meng Z, Du X, Sapkota R., Ma Z, Cheng H. 2025. YOLOv10-pose and YOLOv9-pose: Real-time strawberry stalk pose detection models. Computers in Industry 165:104231.

RDA (Rural Development Administration). 2019. Agricultural technology guide 40: strawberry. [in Korean]

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.

Yu H, Xu Y, Zhang J, Zhao W, Guan Z, Tao D. 2021. Ap-10k: A benchmark for animal pose estimation in the wild. arXiv preprint arXiv:2108.12617.

Wang C, Wang H, Han Q, Zhang Z, Kong D, Zou X. 2024. Strawberry detection and ripeness classification using yolov8+ model and image processing method. Agriculture 14(5):751.

Xiong Y, Ge Y, Grimstad L, From PJ. 2020. An autonomous strawberry‐harvesting robot: Design, development, integration, and field evaluation. Journal of Field Robotics 37(2):202-224.

Zhang Y, Yu J, Chen Y, Yang W, Zhang W, He Y. 2022. Real-time strawberry detection using deep neural networks on embedded system (rtsd-net): An edge AI application. Computers and Electronics in Agriculture 192:106586.