

RESEARCH ARTICLE

주요지점 간 상대 오프셋 추정을 통한 폐색 오이의 2차원 자세추정

성백겸, 이태신, 이대현*

충남대학교 스마트농업시스템기계공학과

2D pose estimation of occluded cucumbers using relative offset regression between keypoints

Baek-Gyeom Sung, Tae-Sin Lee, Dae-Hyun Lee*

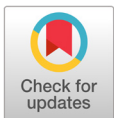
Department of Smart Agriculture Systems Machinery Engineering, Chungnam National University, Daejeon, Republic of Korea

*Corresponding author: leedh7@cnu.ac.kr

Abstract

In agricultural harvesting robots, fruit pose information is essential for determining the grasping point. However, in greenhouse environments, frequent occlusion caused by leaves and stems considerably degrades the accuracy of existing models. This study aims to improve cucumber pose estimation performance in occluded scenes. To achieve robust estimation performance under various occlusion conditions, we designed an approach that infers the relative offsets of occluded keypoints from visible keypoints and reconstructs the entire pose. The training dataset consisted of images collected from an actual cucumber greenhouse under diverse capturing angles, while the test dataset was constructed by imposing various occlusion conditions using leaf patches to emulate realistic field distribution. In addition, a similarity-based offset loss was introduced to encourage consistent relative-offset regression among keypoints. For practical deployment in real-world applications, a YOLO (you only look once)-based model was adopted to simultaneously detect cucumbers, estimate keypoints, and regress the relative offsets between keypoints within a single architecture. Comparative experiments demonstrated that the proposed method achieved higher accuracy than existing approaches across various occlusion scenarios. These results indicate that the approach is capable of compensating for missing pose components by reasoning from partially visible regions. With continued research, this method is expected to enhance the precision of vision pipelines for cucumber harvesting robots operating under real greenhouse conditions.

Keywords: Cucumber, Green house, Harvesting robot, Occlusion, Pose estimation



OPEN ACCESS

Journal of Agricultural Machinery Engineering
5(4):127-137

DOI: <https://doi.org/10.12972/jame.2025.5.4.3>

Received: November 19, 2025

Revised: December 12, 2025

Accepted: December 14, 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

시설원예기술의 발달로 오이(*Cucumis sativus*)의 연중 상시 공급이 가능해졌지만, 과채류 중에서 짧은 재배기간과 빠른 노화속도로 인해 반복적으로 노동집약적인 수확 작업이 요구된다. 최근 농업 노동력이 감소하고 있는 상황에서 이를 해결하기 위해 자동 수확 로봇에 대한 관심이 높아지고 있다(RDA, 2021).

수확로봇은 기존 인력 기반 방식보다 일관된 기준과 속도로 작업을 수행하여 높은 효율성을 제공한다는 장점이 있으며, 온실과 같이 고밀도로 정형화되어 있는 시설에서 반복적인 작업에 적용될 수 있는 잠재력이 높다. 일반적으로 오이 수확 로봇의 자동 수확과정은 온실 장면에서 과실을 찾고, 자세를 추정한 뒤 파지 지점을 엔드이펙터(end-effector)로 파지하여 수확하는 순서로 구성되어 있다. 따라서 성공적인 수확작업을 위해서는 이러한 인식 기술들의 우선적인 고도화가 요구된다(Kim et al., 2023a; Choi et al., 2025).

지난 몇 년간 컴퓨터 하드웨어(hardware) 및 심층학습(deep learning)의 발전으로 수확 로봇을 위한 컴퓨터 비전(computer vision) 기술은 온실에서 배경과 과실을 구분하고, 자세를 추정할 수 있게 되었다. 예를 들어, Lee et al. (2025)는 YOLO (you only look once)-v8-segmentation 모델을 활용하여 딸기의 과실, 꽃받침, 줄기를 분할(segmentation)하고 이를 헝가리안 알고리즘을 통해 그룹화하여 딸기의 자세를 추정하였다. 관련된 연구들은 토마토(Kim et al., 2023b), 참외(Kang et al., 2025) 등 다양한 작물에 걸쳐 높은 자세 추정 정확도를 보여주었지만, 모두 공통적으로 폐색 상황에서 한계를 보고하였다.

실제 수확로봇이 운용되는 온실환경은 잎과 줄기에 의한 폐색이 빈번하게 발생된다(Kim et al., 2023). 폐색 상황에서의 강건한 추정 성능을 위해 폐색 복원이나 깊은 신경망을 이용한 연구들이 수행되었지만 고성능 컴퓨터 자원과 긴 추론시간이 요구되어, 수확로봇의 인식 기술로는 실용성이 부족한 실정이다(Son et al., 2025). 오이의 경우 일반적으로 줄기 근처에서 수직 방향으로 자라나도록 유도하는 수직 유인 재배 방식으로 재배되기 때문에 어느정도 일관된 형태를 가진다(Lee et al., 2004). 따라서 보이는 부위와 가려진 부위간 공간적 유사도 관계를 학습한다면 폐색상황에서도 자세 추정이 가능할 것으로 기대된다.

따라서 본 연구에서는 폐색 상황에서 오이 수확로봇의 강건한 자세 추정 성능을 위해 주요지점(keypoint)간 상대적 관계를 학습하여 폐색 오이의 2차원 자세 추정을 수행하였다. 제안된 방법은 YOLO-v8-Pose를 검출 모델로 사용하여, 기존 모델에서 출력되는 보이는 주요지점 좌표로부터 가려진 주요지점까지의 상대 오프셋(offset)을 예측함으로써, 오이의 전체 자세를 추정할 수 있어 수확로봇의 실용적인 폐색 상황 대응 도구로 활용 될 수 있는 가능성을 제시한다.

Materials and Methods

오이 자세 추정 개요

실제 수확로봇에서의 적용을 위해서 인식 모델은 실시간으로 오이 위치 검출, 자세 추정 등을 수행할 수 있어야 한다. YOLO는 다수의 연구에서 높은 검출 정확도와 실시간성으로 실용성이 입증되었다(Maji et al., 2022). 본 연구에서도 Fig.1과 같이 YOLOv8-Pose 모델을 기반으로 오이의 자세 추정을 수행하였다 (Jocher G et al., 2023). 모델은 입력 영상(image)에서 오이의 경계 상자(bounding box), 클래스(class), 주요지점의 위치(keypoint location), 주요 지점의 존재가능성(keypoint objectness), 주요지점들간 상대 오프셋(relative offset)을 예측한다. 이를 통해 온실 장면에서 오이 검출, 객체 내 주요지점 기반 자세 추정, 가려진 주요지점을 추정하기 위한 단서 제공을 동시에 수행할 수 있다. 본 연구에서는 오이의 전체 자세 추정을 위한 오이의 주요 지점을 꽃 끝(top), 과실 중앙(center), 꽃 자루(bottom)으로 정의하였으며, 상대 오프셋은 이 주요지점들간 벡터차를 의미한다. 이 중 과실 중앙은 수확 로봇의 파지 지점에 해당하고, 꽃 끝과 꽃 자루는 오이의 길이, 방향, 기울기를 정의하

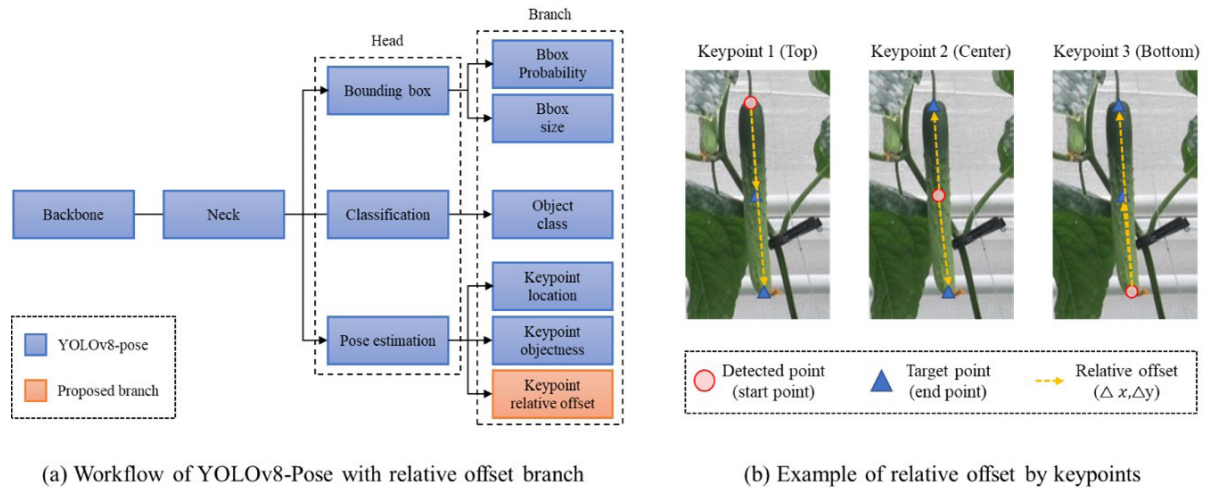


Fig. 1. Overview of pose estimation method used in this study: (a) workflow of YOLOv8-Pose with relative offset branch, and (b) Example of relative offset by keypoints.

는 기준으로 사용된다. 이러한 주요 지점 기반 자세 추정 은 곡물, 길이, 기울기가 객체별로 상이한 오이에서도 안정적인 그리퍼(gripper) 접근을 가능하게 한다. 이어지는 문단에서는 사용된 데이터, 학습 모델에 대한 설명, 상대 오프셋을 최적화 하기 위해 제안된 손실함수에 대해 설명하고 주요 결과 및 결론에 대한 논의가 진행된다.

데이터 구성

Fig. 2는 본 연구에서 사용된 데이터셋(dataset)의 수집과정 및 주석(annotation) 예시이다. 사용된 이미지 데이터는 경상북도 상주 스마트팜 혁신밸리에 위치한 오이 유리 온실($36^{\circ} 28' 17.5'' \text{N}$ $128^{\circ} 13' 31.1'' \text{E}$)에서 수집되었으며, 오이 품종은 '백다다기'로 설정되었다. 데이터 수집은 스마트폰(iPhone 15 Promax, Apple, CA, USA)를 활용하여 수행되었으며, 수확로봇 시점을 가정하여 카트에서 다양한 각도로 수집되었다. 수집된 이미지는 3024×4032 해상도를 가진 117장의 RGB 이미지로 이들은 모델 입력을 위해 $640 \times 640 \times 3$ 크기로 조정된 뒤, 학습용(training), 검증용(validation), 평가용(test)으로 각각 70장, 20

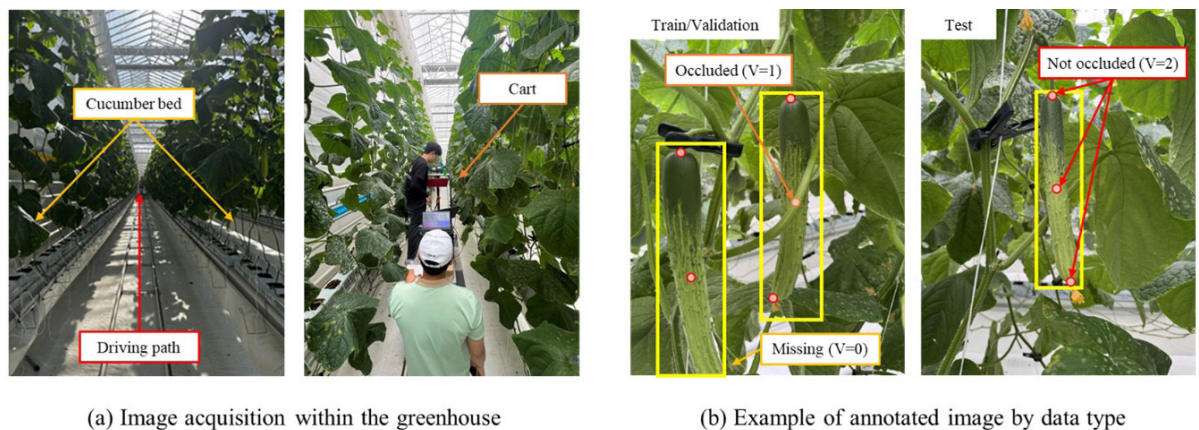


Fig. 2. Data collection in the greenhouse: (a) image acquisition within the greenhouse, and (b) Example of annotated image by data type.

장, 27장씩 구분되었다. 이때, 평가용 데이터는 세 주요지점이 모두 보이는 이미지로만 선별되어 구성되었다. 학습을 위한 주석은 YOLO 학습 데이터 형식에 맞게 클래스, 경계상자, 2차원 좌표와 가시성을 포함한 주요지점(p_x, p_y, V)으로 선정되었다. 이때 주요지점의 가시성(V)은 이미지밖으로 나갔거나 전혀 유추가 불가능한 경우 $V = 0$, 유추가 가능하게 가려진 경우 $V = 1$, 보이는 경우 $V = 2$ 로 설정되었다(Dwyer B et al., 2025; Dibenedetto et al., 2025).

자연적인 주요지점 폐색 상황에서 오이의 정확한 주요지점 정답(ground truth)를 만드는 것은 불가능하다. 따라서 본 연구에서는 선행연구를 따라 이미지 처리기법을 활용하여 인공적인 폐색을 생성하였다(Kim et al., 2023). Fig. 3은 본 연구에서 사용된 푸아송 블렌딩(poisson blending) 과정으로 나뭇잎 패치(patch)의 색상 변화율(gradient)을 유지하면서, 타겟 이미지(target image) 내에서 패치가 부착될 패치 경계내부의 색상을 시각적으로 자연스럽게 재구성하기 위해 푸아송 방정식이 활용된다. 여기서 나뭇잎 패치는 평가용 이미지에서 사전에 수동으로 추출된 40장에서 무작위로 1장씩 선택되었으며, 무작위 각도로 회전되어 사용되었다. 패치의 크기는 사전에 정의된 경계상자 크기의 20~50% 수준으로 설정되었으며, 오이의 꽃 끝, 과실 중앙, 꽃 자루 위치에 생성함으로써 폐색 조건(occlusion condition)을 세가지로 정의하였다. 또한, 본 연구에서는 이러한 인공 폐색에 대한 모델의 과적합을 방지하기 위해 사전에 분류된 27장의 평가용 샘플(sample)에만 폐색 증강을 적용하여 총 81장의 평가용 이미지로 구성하였다.

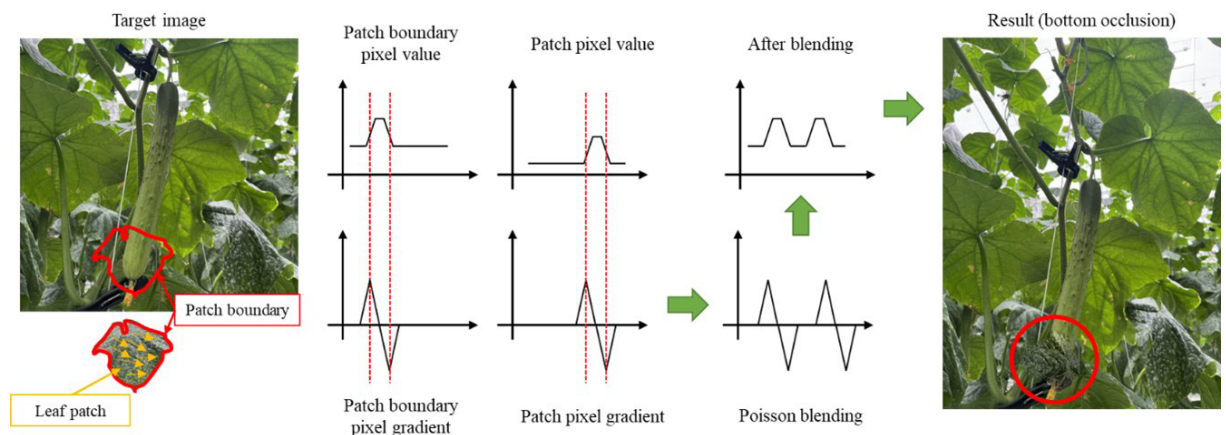
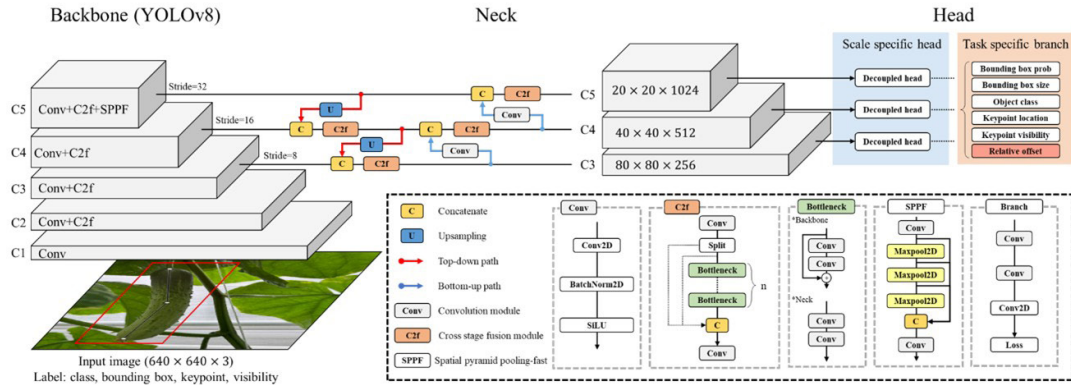


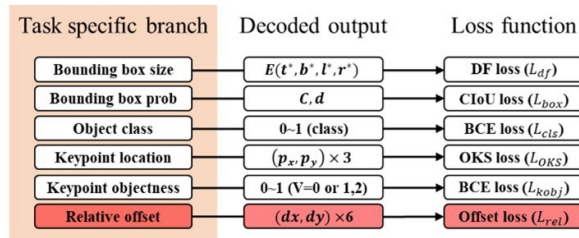
Fig. 3. Process of image processing (poisson blending) for synthetic occlusion condition using leaf patch.

오이 자세 추정 모델 및 학습

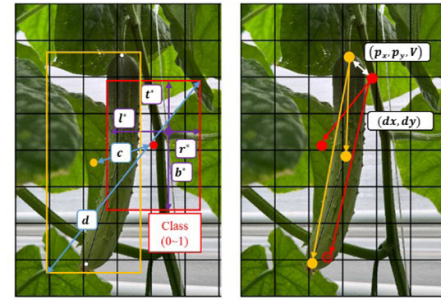
Fig. 4는 본 연구에서 사용된 YOLOv8-Pose 모델로, feature pyramid network 구조로 설계되어 있으며 백본(backbone), 넥(neck), 헤드(head)로 구성되어 있다(Bochkovskiy et al., 2020). 백본은 입력 이미지로부터 3가지 해상도의 특징맵(feature map)을 추출하고, 넥은 상향식(bottom-up), 하향식(top-down) 경로를 통해 특징들을 결합하여 3개의 특징 맵을 출력한다. 헤드는 입력된 다중 해상도 특징 맵으로부터 목적 별 브랜치(branch)를 통해 경계 상자 회귀, 클래스 분류, 그리고 세 개 주요 지점의 좌표 및 존재가능성에 대한 예측을 출력하며, 이러한 예측 결과는 디코딩 과정을 거쳐 손실 함수 계산에 사용되고, nom maximum suppression을 통해 최종 출력함으로써 객체의 크기에 상관없이 정밀한 예측을 가능케한다. 여기서 주요지점의 존재 가능성은 가시성을 기반으로 존재하지 않을 확률($V=0$)과 존재할 확률($V=1$ 또는 $V=2$)을 나타낸다. 본 연구에서는 이러한 기존 구조에서 주요 지점별 상대 오프셋(dx, dy)을 계산하여, 총 6쌍의 상대 오프셋을 출력하는 브랜치를 추가적으로 병렬 배치하여 학습 및 추론을 수행한다.



(a) Architecture of YOLOv8-Pose model with proposed branch



(b) Model output by branch with loss function



(c) Example of model output

Fig. 4. Network architecture for cucumber pose estimation used in this study: (a) architecture of YOLOv8-Pose model with proposed branch, (b) model output by branch with loss function, and (c) example of model output (yellow) and ground-truth (red).

학습에는 YOLO의 기존 구성을 따라 브랜치별로 다른 손실 함수가 사용된다. 경계 상자 위치에 대해서는 예측된 경계 상자와 정답 경계 상자 간의 중심거리(C)와 두 경계상자를 모두 포함하는 최소 크기의 외접 상자의 대각선 길이(d)를 기반으로 계산되는 CIoU (complete intersection over union) loss가 사용되며(Tian et al., 2022), 상자 크기에 대해서는 경계 상자내 객체 중심점으로부터 예측된 상자의 상하좌우까지의 거리(t^*, b^*, l^*, r^*)를 기반으로 DF (distribution focal) loss가 사용된다(Li et al., 2020). 클래스에 대해서는 BCE (binary cross entropy) loss를 사용하여 데이터 내 모든 클래스 각각에 속할 확률에 대한 손실을 계산하며, 예측된 주요 지점의 존재가능성에 대해서도 BCE loss를 통해 존재할 경우에만 손실을 계산하여 모델의 잘못된 주요 지점 좌표 예측을 방지하면서 추정된 주요 지점의 신뢰도를 나타낼 수 있다. 주요 지점 좌표에 대해서는 식 (1)과 같은 OKS (object keypoint similarity) loss가 사용되었는데, 여기서 OKS는 예측된 주요지점과 정답 주요지점 간의 유사도 척도로, 인간 자세추정 분야에서 평가에 사용되며 식(2)와 같이 계산된다(Maji et al., 2022). 본 연구에서는 이러한 OKS loss의 원리를 차용한 오프셋 유사도 손실을 통해 상대 오프셋을 학습하였다. 제안된 손실함수는 기존 모델에서 예측된 주요 지점의 좌표를 위치벡터(position vector)로 취급하여 식 (3)과 같이 L2 거리 기반의 오프셋을 구한 뒤, OKS와 동일하게 L2정규화를 통해 식(4)와 같이 오프셋 벡터의 방향성만을 남김으로써 안정적으로 유사도를 평가한다.

$$L_{OKS} = 1 - OKS = 1 - \frac{\sum_{i=1}^{N_{kpts}} s_i \delta_i(v_i > 0)}{\sum_{i=1}^{N_{kpts}} \delta_i(v_i > 0)} \quad (1)$$

$$s_i = \exp\left(-\frac{\|p_i - g_i\|_2^2}{2s^2(\sigma_i^2)}\right) \quad (2)$$

where N_{kpts} denotes the number of keypoints considered for each object, p_i denotes the estimated two-dimensional coordinate of the i -th keypoint, while g_i denotes its corresponding ground-truth coordinate, s denotes object scale factor calculated from the bounding box area, σ_i is a pre-defined constant specific to each keypoint type (COCO dataset). v_i is the visibility label of the i -th keypoint in the annotation, and δ_i is the visibility indicator, returning 1 when $v_i > 0$ (i.e., the keypoint is labeled) and 0 otherwise.

$$e_{ij} = \exp\left(-\frac{\|\hat{r}_{ij} - r_{ij}\|_2^2}{2\|r_{ij}\|_2^2(\tau_i^2)}\right), \hat{r}_{ij} = p_{i,s} - p_{j,e}, r_{ij} = g_{i,s} - g_{j,e} \quad (3)$$

$$L_{rel} = 1 - \frac{\sum_{i=1}^{N_{pairs}} e_{ij} \delta(v_{i,s} > 0, v_{j,e} > 0)}{\sum_{i=1}^{N_{pairs}} \delta(v_{i,s} > 0, v_{j,e} > 0)} \quad (4)$$

where $g_{i,s}$ and $g_{j,e}$ denote the ground-truth position vectors of the starting and ending keypoints, and $p_{i,s}$ and $p_{j,e}$ denote their predicted counterparts. \hat{r}_i denotes the estimated offset of the i -th keypoint pair, and r_i denotes the corresponding ground-truth offset. τ_i is a predefined constant associated with each keypoint type (set to 0.33 in our experiments). N_{pairs} is the number of predefined keypoint pairs, which is at most six in this study. $\delta(\bullet)$ denotes the visibility indicator for a keypoint pair, returning 1 when both keypoints in the pair are visible.

최종적으로 본 연구의 모델은 식 (5)의 손실함수를 사용하여 학습을 수행하며, 이때 각 손실함수의 가중치는 YOLO의 기본 설정을 사용하였다. 이를 통해 모델이 기존 출력들 외에도 오이의 구조적 유사성을 추가적으로 학습하도록 장려하여, 오이 영상에서 보이는 주요지점으로부터 가려진 주요지점의 위치를 추론할 수 있도록 한다. 폐색으로 인해 기존 모델이 검출하지 못한 주요지점에 대한 추론은 식 (6)과 같이 예측된 주요 지점으로부터 검출되지 못한 주요지점까지의 예측된 오프셋을 주요지점 신뢰도 기반 가중치 합하여 계산된다.

$$L_{total} = w_{df} \cdot L_{df} + w_{box} \cdot L_{box} + w_{cls} \cdot L_{cls} + w_{OKS} \cdot L_{OKS} + L_{kobj} + L_{rel} \quad (5)$$

where L_{total} denotes the final loss function used in this study. The loss weights were set as $w_{df} = 1.5$, $w_{box} = 7.5$, $w_{cls} = 0.5$, $w_{OKS} = 12.0$, while all remaining loss terms were assigned a weight of 1.0. These settings follow the standard YOLO loss design.

$$\hat{k}_m = \frac{\sum_{i \in V} c_i (\hat{k}_i + d_{i \rightarrow m})}{\sum_{i \in V} c_i} \quad (6)$$

where \hat{k}_i denotes the estimated two-dimensional coordinate of the visible keypoint i , $d_{i \rightarrow m}$ denotes the predicted offset from keypoint i to the occluded keypoint m . c_i is the confidence score associated with keypoint i , and V is the set of visible keypoints with confidence score above a threshold (set to 0.3 in our experiments).

본 연구에서 학습 모델은 500 epoch동안 학습되었으며, 과적합을 방지하기 위해 조기종료(early stop)를 사용하였다. 따라서, 모델의 파라미터(parameter)는 학습 중 가장 낮은 검증 손실(validation loss)을 달성한 시점으로 선정되었다(Kim et al., 2021). 모델에 입력되는 학습 데이터의 배치 크기(batch size)는 16으로 설정되었으며, 모델의 파라미터 업데이트를 위한 최적화 알고리즘(optimizer)은 AdamW가 사용되었다. 학습 중 학습률(learning rate)은 2×10^{-3} , 가중치 감쇠(weight decay)는 5×10^{-4} 로 설정되었다. 모델 및 학습 구현은 Python 3.10.11 및 PyTorch 2.4.1을 통해 수행되었으며, 결과 시각화 및 이미지 전 처리에는 OpenCV 4.10.0이 사용되었다. 학습은 Intel Core i7-12700F CPU (Intel Corporation, Santa Clara, CA, USA)와 NVIDIA

GeForce RTX 4090 GPU (NVIDIA Corporation, Santa Clara, CA, USA)환경에서 수행되었다.

성능평가

본 연구의 방법과 자세 추정 성능 비교를 위한 Baseline으로는 Table 1에 제시된 것과같이 일반적인 YOLOv8-Pose 모델이 선정되었으며, 제안된 유사도 기반 손실이 아닌 오프셋간 L1 및 L2 손실을 이용하여 학습을 수행한 방법들과도 비교되었다. 이때, 비교에 사용된 방법들은 모두 동일한 설정 및 환경에서 학습이 수행되었으며, 평가용 데이터셋에서 비교되었다.

Table 1. Comparison of pose estimation methods, their respective offset loss functions, and their model parameters.

Method	Offset loss	Parameters (M)
YOLOv8n-Pose (Baseline)	None	3.08
YOLOv8n-Pose-L1 offset (With L1)	L1 loss	3.15
YOLOv8n-Pose-L2 offset (With L2)	L2 loss	3.15
YOLOv8n-Pose-OKS offset (Our method)	Similarity loss	3.15

모델의 주요지점 검출 성능 평가는 예측 값과 정답 값 간의 관계를 나타내는 혼동행렬(confusion matrix) 기반의 지표를 사용하였다. 사용된 지표는 mAP (mean average precision)로, 모델이 정답으로 분류한 샘플 중 실제 정답의 비율인 정밀도 (precision)와 실제 정답 중 모델이 정답으로 분류한 샘플의 비율인 재현율(recall)로 그려지는 정밀도-재현율 곡선(precision-recall curve)의 아래면적으로 계산된다(Lee et al., 2025). 이때, 정답으로 분류하는 임계값(threshold)으로는 자세추정 분야에 서 일반적으로 사용되는 OKS가 사용되었으며(Maji et al., 2022), 본 연구에서는 50%, 70%, 50%–95% 수준으로 설정되었다.

모델을 통해 검출된 주요지점의 정확도 평가는 인간 자세 추정에서 사용되는 PCK (percentage of correct keypoints) 지표가 사용되었다. PCK는 식 (7)과 같이 예측된 주요지점과 정답 주요 지점간 L2 거리로 계산된다. 이때, 임계값으로는 오이에 맞게 조정하여(Kang et al., 2025), 꽃 끝부터 꽃 자루까지 길이의 5%, 10%, 20% 수준으로 비교되었다.

$$PCK@threshold = \frac{\sum_{i=1}^{N_{kpts}} \frac{\|p_i - g_i\|_2}{l_c} < threshold}{N_{gt,kpts}} \quad (7)$$

where l_c denotes the length of the cucumber in the image, computed as the Euclidean distance between the top and bottom keypoints, and the thresholds are set to 0.05, 0.10, and 0.20.

Results and Discussion

자세 추정 학습 결과

Fig. 5는 방법 별 학습 후 추론 결과로, 시각화는 꽃 끝, 과실 중앙, 꽃 자루 각각에 대해 대표적인 폐색 샘플에서 제시되었다. 결과에서 각 행은 폐색 상황을 구분하며, 각 열은 방법을 나타낸다. 폐색을 위한 나뭇잎 패치의 위치는 노란색 원을 통해 제시되었다. 빨간색 점은 예측된 주요 지점이며, 초록색 점은 정답 주요지점을 나타낸다. 이때, 분홍색 선과 흰색 선은 각각의 주요지점을 잇는 오이의 자세이다. With L1과 With L2는 특정 지점에 대해 불안정한 추론 결과를 보여주었으며, Baseline과 제안된 방법은 안정적인 지점별 추론 결과를 보여주었다. 하지만 대부분의 결과에서 전체적인 자세 추정이 가능함을 보여주어 모든 방법에서 학습이 정상적으로 수행된 것으로 판단된다.

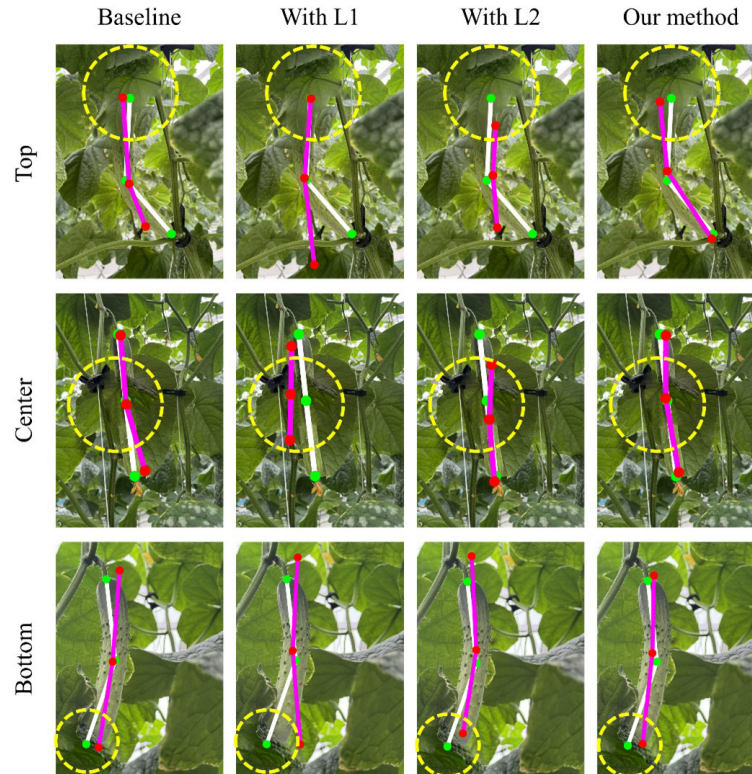


Fig. 5. Comparison of pose estimation result by different methods. The visualization is presented on representative occluded samples for the top, center, and bottom keypoint of cucumber. Rows correspond to different occlusion conditions, and columns represent the compared methods. The location of the leaf patch is additionally highlighted by a yellow circle. The red point is estimated keypoint and green point is its ground-truth keypoint. The pink line represents the predicted cucumber pose reconstructed from the estimated keypoints, and the white line represents the ground-truth pose constructed from the ground-truth keypoints.

자세 추정 성능 비교

Table 2는 방법 별 자세 추정 성능을 제시한 것이다. 제안된 방법은 모든 수준의 mAP 지표에서 다른 방법들보다 높은 성능을 달성하였으며, 이는 Baseline보다 1.2%~5.4% 높은 수준이었다. With L1과 With L2는 Baseline보다도 낮은 정확도를 보여주었는데, 이는 오프셋 학습에서 비교적 단순한 L1 및 L2 기반 오프셋 손실은 안정적인 관계정보를 제공하지 못함을 보여준다. 이에 반해 제안된 유사도 기반의 오프셋 손실은 주요지점 간 상대적 방향 일관성 및 공간적 구조를 유지할 수 있어 더 효과적인 학습이 가능한 것으로 판단된다. 하지만 제안된 방법을 포함한 오프셋 손실을 사용하는 방법들은 PCK 지표에서 Baseline보다 낮은 성능을 보여주었다. 이는 오프셋 손실이 모델이 전체적인 구조에 집중하게 하여 폐색 상황에서의 전반적인 추정 품질을 향상시키지만 주요지점 자체의 좌표 추정 정밀도는 감소시키는 것으로 해석된다. 그러나 제안된 방법에서는 PCK의 임계값이 높아질수록 Baseline과의 성능차이가 점진적으로 줄어드는 경향을 보여주어, 오프셋 보정 전략의 안정화 등 추가적인 개선이 가능할 것으로 판단된다.

Table 2. Performance evaluation of pose estimation on test datasets.

Method	Evaluation metrics					
	mAP@0.5	mAP@0.75	mAP@0.5 : 0.95	PCK@0.05	PCK@0.1	PCK@0.2
Baseline	0.92	0.92	0.85	0.45	0.79	0.95
With L1	0.57	0.47	0.42	0.19	0.41	0.75
With L2	0.74	0.71	0.62	0.22	0.48	0.83
Our method	0.97	0.96	0.86	0.31	0.73	0.94

폐색 조건 별 성능 분석

Fig. 6은 폐색 조건 별 자세추정 성능을 평균값과 비교한 것이다. 비교는 각 지표에서 대표적으로 사용되는 임계값인 $mAP@0.50:0.95$ 와 $PCK@0.1$ 에서 수행되었다. mAP 지표에서 제안된 방법을 제외한 다른 방법들은 과실 중앙이 폐색되었을 때 성능감소가 두드러졌다. PCK 지표에서는 오프셋 손실을 사용하는 방법들이 꽃 자루가 가려졌을 때 성능이 하락하는 공통적인 경향을 보여주었다. 성능 하락 요인을 분석하기 위해 정답 주요지점과 정답 상대 오프셋들의 가시성 라벨을 확인하였으며, 그 분포는 Table 3에 제시되었다. 확인 결과, 꽃 자루에 대한 데이터가 가장 부족하였고, 과실 중앙에 대한 데이터가 가장 많았다. 이에 따라, 제안된 방법을 제외한 다른 방법들의 mAP 감소는 학습 과정 중 가장 많은 정답이 존재하는 과실 중앙 정보에 의존하여 과실 중앙이 가려졌을 때 성능 감소가 발생한 것으로 판단되며, 오프셋 손실 기반 방법들의 꽃 자루가 가려졌을 때의 PCK 감소는 주요 지점에 대한 정보가 부족한데도 오프셋 정보가 많기 때문인 것으로 해석된다. 따라서 폐색 조건 별 성능 편향은 학습 데이터 내 라벨 불균형으로 인한 것으로 판단되며, 이를 완화시키기 위한 방법을 적용한다면 편향을 감소시킬 수 있을 것으로 기대된다.

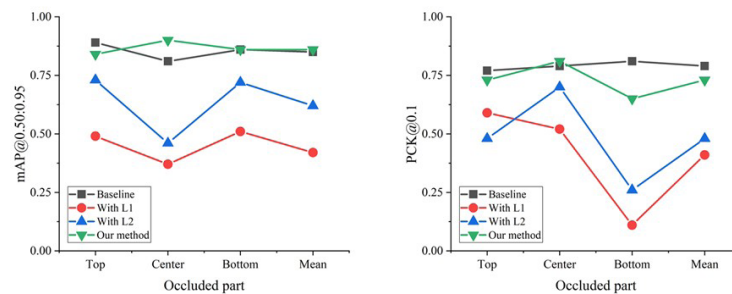


Fig. 6. Comparison of pose estimation performance by different occlusion conditions. The comparison is shown for two evaluation metrics: $mAP@0.50:0.95$ (left) and $PCK@0.1$ (right).

Table 3. Distribution of keypoint visibility and corresponding relative offsets.

Visibility (V)	Keypoints			Relative offset pairs		
	Top	Center	Bottom	T-C ^x	T-B ^y	C-B ^z
Missing (0)	21	2	20	2	12	8
Occluded (1)	13	14	36	50	66	82
Visible (2)	139	157	117	250	186	214

^{x, y, z}Relative offset pairs between the corresponding keypoint pairs (Top–Center, Top–Bottom, Center–Bottom).

Fig. 7은 학습 데이터 내 라벨 불균형을 완화하기 위해 앞서 언급한 인공 폐색 증강을 학습용 데이터에도 적용하여 학습한 뒤, 방법별로 비교한 결과이다. 인공 폐색 증강은 각 해당 데이터세트 분할에서 수동으로 추출된 나뭇잎 패치를 사용하여, 기존과 동일하게 꽃 끝, 과실 중앙, 꽃 자루에 대해 각각 적용하였으며, 방법 별 비교 또한 기존 평가용 데이터세트에서 진행되었다. Baseline은 증강된 데이터로 학습하여 성능 변화가 관찰되지 않았는데 이는 폐색정보에 대한 별도의 표현학습을 하지 못하는 것으로 판단된다. 이에 반해 오프셋 손실 기반 방법들은 두 지표 모두에서 성능 향상을 보여주었다. 이는 학습용 데이터의 폐색 정보 분포가 완화되어 특정 주요지점과 관계에 의존되지 않고 폐색 정보에 대한 학습을 수행한 것으로 판단된다. 특히, 제안된 방법은 모든 지표에서 가장 높은 정확도를 달성하여, 상대 오프셋을 학습하는데 다른 방법들보다 유사도 손실이 더 효과적임을 보여주었다.

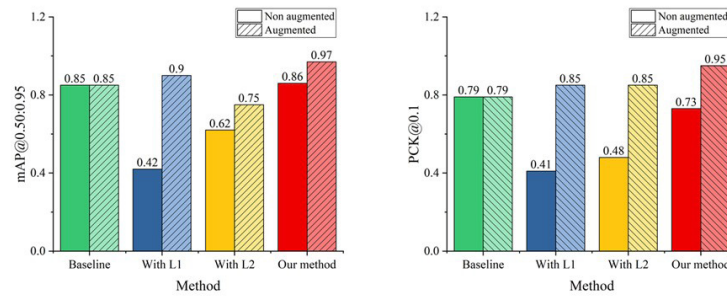


Fig. 7. Performance variation across methods after applying occlusion augmentation on training dataset. The comparison is shown for two evaluation metrics: mAP@0.50:0.95 (left) and PCK@0.1 (right).

Conclusion

본 연구는 온실 환경 내 폐색 상황에서 강건한 자세 추정 성능을 확보하기 위한 오이 자세 추정 방법을 제안하였다. 제안된 방법은 YOLOv8-Pose 모델을 기반으로 하여 주요 지점 검출과 상대 오프셋 회귀를 통합적으로 수행함으로써, 일부 주요 지점이 폐색된 경우에도 전체 자세를 복원할 수 있도록 설계되었다. 또한 오이의 구조적 특성을 반영하기 위해, 주요 지점 간 관계의 일관성을 강화하는 유사도 기반 상대 오프셋 손실 함수를 도입하였다. 제안된 방법은 기존 방법인 YOLOv8-Pose를 비롯하여 일반적으로 사용되는 L1 및 L2거리를 오프셋 손실로 사용하는 방법론들과 비교되었다. 제안된 접근법은 다양한 폐색 상황에서 다른 방법들보다 우수한 추정 성능을 보여주었지만 특정 지표에서는 기존 방법 대비 성능 감소를 보여주었으며, 폐색 조건 별 성능 편향이라는 한계를 보여주었다. 하지만 이는 소규모 데이터셋에서 비롯된 학습 데이터 분포 불균형에서 야기된 것으로 추가적인 폐색 증강 실험을 통해 분포를 완화시켜줄 시, 다른 방법들보다 더 폐색에 강인한 표현학습이 가능한 것으로 확인되어 개선 가능성을 확보하였다. 따라서 향후 연구에서는 다양성을 포함한 대규모 데이터셋을 구축할 필요가 있으며, 상대 오프셋 외에도 형태학적 특성 등을 학습단계에서 고려하여 성능 향상을 도모해야 한다. 이를 통해 더 높은 수준의 정밀도와 강건성을 확보한다면 다양한 폐색 조건에 대응 가능한 오이 수확로봇의 인식 시스템으로 발전할 수 있을 것으로 기대된다.

Acknowledgements

본 연구는 농림축산식품부의 재원으로 농림식품기술기획평가원의 스마트팜다부처패키지혁신 기술개발사업의 지원을 받아 수행되었음 (RS-2025-02219360).

References

- Bochkovskiy A, Wang CY, Liao HYM. 2020. Yolov4: Optimal speed and accuracy of object detection arXiv preprint arXiv:2004.10934.
- Choi TH, Lee TS, Kang SW, Sung BG, Lee DH. 2025. 3D pose estimation of strawberry using 2D keypoint detection and regression. *Journal of Agricultural Machinery Engineering* 5(3):99-110. [in Korean]
- Dwyer B, Nelson J, Hansen T. 2025. Roboflow (Version 1.0) [Software]. Available from <https://roboflow.com>. Computer vision
- Dibenedetto G, Sotiropoulos S, Polignano M, Cavallo G, Lops P. 2025. Comparing human pose estimation through deep learning approaches: An overview. *Computer Vision and Image Understanding* 252:104297.

- Jocher G, Chaurasia A, Qiu J. 2023. Ultralytics YOLOv8 (version 8.0. 0). Accessed in <https://github.com/ultralytics/ultralytics>
- Kim WS, Lee DH, Kim T, Kim G, Kim H. Sim T, Kim YJ. 2021. One-shot classification-based tilled soil region segmentation for boundary guidance in autonomous tillage. *Computers and Electronics in Agriculture* 189:106371.
- Kim S, Hong SJ, Ryu J, Kim E, Lee CH, Kim G. 2023. Application of amodal segmentation on cucumber segmentation and occlusion recovery. *Computers and Electronics in Agriculture* 210:107847.
- Kim T, Lee DH, Kim KC, Choi T, Yu JM. 2023a. Tomato maturity estimation using deep neural network. *Applied Sciences* 13(1):412.
- Kim T, Lee DH, Kim KC, Kim YJ. 2023b. 2D pose estimation of multiple tomato fruit-bearing systems for robotic harvesting. *Computers and Electronics in Agriculture* 211:108004.
- Kang SW, Kim KC, Kim YJ, Lee DH. 2025. Real-time pose estimation of oriental melon fruit-pedicle pairs using weakly localized fruit regions via class activation map. *Computers and Electronics in Agriculture* 237:110734.
- Lee JN, Lee UH, Kim WB, Ryu SY, Yong YR. 2004. Establishment of arch training method in summer season cultivation of cucumber under rain-shelter in highlands. *Journal of Bio-Environment Control* 13(1):26-32. [in Korean]
- Li X, Wang W, Wu L, Chen S, Hu X, Li J, Tang J, Yang J. 2020. Generalized Focal Loss: Learning qualified and distributed bounding boxes for dense object detection. *Advances in neural information processing systems* 33:21002-21012.
- 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]
- Maji D, Nagori S, Mathew M, Poddar D. 2022. YOLO-Pose: Enhancing YOLO for multi person pose estimation using object keypoint similarity loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 2637-2646.
- RDA (Rural Development Administration). 2021. Agricultural technology guide 107: cucumber. [in Korean]
- Son JH, Song H, Song C, Ha M, Kang D, Ha YS. 2025. Condition-based synthetic dataset for amodal segmentation of occluded cucumbers in agricultural images. *Computers and Electronics in Agriculture* 238:110800.
- Tian D, Han Y, Wnag S, Chen X, Guan T. 2022. Absolute size IoU loss for the bounding box regression of the object detection. *Neurocomputing* 500:1029-1040.