

가시성 분류 손실 함수 개선을 통한 부분 폐색 객체의 Visibility-aware Keypoint Detection 연구

Visibility-aware Keypoint Detection for Partially Occluded Objects via Enhanced Visibility Classification Loss

이지민¹ · 이나연² · 이윤결² · 장인훈^{3,*}

Jimin Lee · Nayeon Lee · Yungyeol Lee and Inhoon Jang

¹한경국립대학교 인공지능연구소

E-mail: ljim2000@hknu.ac.kr

²한경국립대학교 AI반도체융합학부

E-mail: lny040713@hknu.ac.kr, ruf4869@hknu.ac.kr

³한경국립대학교 AI반도체융합학부, 인공지능연구소장

E-mail: inhuns@hknu.ac.kr

요약

인공지능 기반의 로봇 수확 시스템에서 정밀한 파지점 결정과 충돌 회피 경로 계획을 위해서는 객체의 완전한 형상 정보가 필수적이다. 그러나 실제 과수 환경은 잎이나 가지에 의한 부분 폐색(Partial Occlusion)이 빈번하여, 가려진 객체의 정확한 포즈를 추정하는 데 한계가 있다. 특히 장애물을 회피하며 최적의 파지 지점에 접근하기 위해서는 가지 영역과 비가지 영역을 명확히 구분하는 정보가 요구된다. 본 연구에서는 이를 해결하기 위해 Keypoint 기반 Pose Estimation을 활용하여 과실의 전체 형상을 추정하고, 가시성을 분류하는 방법론을 제안한다. 제안된 모델은 과실의 키포인트를 탐지함과 동시에 각 점의 가시 여부를 판별할 수 있도록 기존 YOLO-Pose 모델의 Visibility 라벨 정보를 활용하여 가시성 분류를 위한 새로운 손실 함수를 도입하였다. 이를 통해 가려진 키포인트에 대한 식별력을 높였으며, 최종적으로 추정된 과실의 위치와 자세를 가시성 정보가 포함된 구조적 데이터로 표현함으로써 향후 폐색 객체에 대한 로봇 파지점 추정 및 경로계획의 정밀도를 높일 수 있는 기반을 마련하였다.

키워드 : Keypoint Detection, Pose Estimation, Occlusion, Apple Harvesting

1. 서론

수확 로봇은 과실의 손상을 최소화하면서 안정적인 파지점을 탐색해야 한다. 잎과 가지에 의해 부분적으로 가려지는 폐색(Partial Occlusion) 환경에서는 제한된 시각 정보로부터 과실을 정확하게 인식하는 것이 필요하다.

YOLO-Pose와 같은 Pose Estimation 모델은 가려진 부분의 keypoint 위치를 추정하여 과실의 형상을 나타낼 수 있다 [1]. 그러나 기존 방법은 keypoint 위치를 추정하기 때문에, 보이는 부분의 keypoint인지 나뭇잎이나 가지에 의해 가려진 부분의 keypoint인지 구분하지 못한다.

따라서 본 연구에서는 keypoint 기반의 가시성(visibility) 분류 학습 방법을 제안한다. 가시성이란 각 keypoint가 관측 가능한 상태(Visible keypoint)인지 또는 가려진 상태(Occluded keypoint)인지를 의미한다. 이를 통해 keypoint 단위의 과실 형상을 추정함과 동시에 각 keypoint에 대해 가시성을 구분할 수 있다.

2. 방법론

2.1 Visibility Loss Function 설계

YOLO-Pose 모델에서 keypoint 라벨 정보는 (x,y) 좌표와 Visibility Flag v 로 구성되며, 이때 v 는 세 가지 상태로 구분된다 [2].

- $v = 0$: 프레임 밖에 있어 데이터로 활용이 불가능한 경우
- $v = 1$: 과실이 잎 또는 다른 과실에 의해 가려진 상태
- $v = 2$: 과실이 관측 가능한 상태

YOLO-Pose의 전체 손실함수는 식 (1)과 같이 정의되며, 클래스, 경계박스, keypoint의 위치 및 존재 유무를 학습한다 [2]. Visibility Flag는 유효한 keypoint($v > 0$)를 선택하기 위한 필터링 기준으로 사용되며, 가시성을 구분하여 학습하지는 않는다.

$$L_{Total} = \sum (\lambda_{cls} L_{cls} + \lambda_{bbox} L_{bbox} + \lambda_{kpts} L_{kpts} + \lambda_{kpts\ conf} L_{kpts\ conf}) \quad (1)$$

제안하는 방법은 Visibility Flag를 새로운 학습 라벨로 정의하고, 이를 이미지 내에 존재하는 유효한 keypoint($v > 0$)에 대하여 가려짐(Occluded)과 보임(Visible) 상태를 식별하는 분류 문제로 정형화하였다. 네트워크가 keypoint의 가시성 여부를 효과적으로 학습할 수 있도록 전용 예측 헤드(head)를 추가로 설계하였다. 이를 최적화하기 위해 기존 손실함수 식 (1)에 Cross Entropy Loss 형태의 가시성 손실 항 L_{vis} 을 더하여 식 (2)와 같은 최종 손실 함수를 제안한다. 이때 추가된 가시성 손실 함수 L_{vis} 의 세부 수식은 식 (3)과 같다.

$$L_{proposedTotal} = \sum (\lambda_{cls} L_{cls} + \lambda_{bbox} L_{bbox} + \lambda_{kpts} L_{kpts} + \lambda_{kpts\ conf} L_{kpts\ conf} + \lambda_{vis} L_{vis}) \quad (2)$$

$$L_{vis} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=0}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (3)$$

여기서 N은 keypoint의 총개수, C는 가시성 클래스의 수, 즉 Occluded와 Visible을 나타내며 λ_{vis} 는 가시성 손실에 대한 가중치를 나타낸다.

3. 실험 방법 및 실험 결과

3.1 실험 방법

본 연구에서 정의한 과실의 keypoint는 총 15개이며, 각각 (x, y, v) 형태로 표현된다. 그림 1(a)과 같이, keypoint는 1개의 과실 꼭지(pedicle)와 14개의 외곽점으로 구성된다 [1].

학습 데이터셋은 Unity 기반의 시뮬레이터를 통해 취득하였다. 실제 사과 농장과 유사한 환경을 구현하기 위해 스캔된 사과 객체를 활용하였으며, 다양한 가려짐 환경을 반영한 이미지와 keypoint 라벨(위치 및 가시성)을 자동 생성하였다. 구축된 데이터셋은 총 19,000장으로 학습에는 17,200장, 테스트에 1,800장을 사용하였으며, 데이터셋의 예시는 그림 1(b)에 제시하였다.

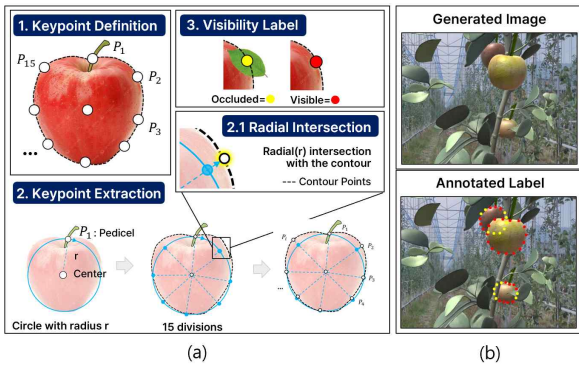


그림 1. (a) 사과 keypoints 정의 (b) 데이터셋 예시.

Fig 1. (a) Apple keypoints definition (b) Dataset examples.

본 실험에서는 YOLOv8s-pose를 기반으로 하되, keypoint 가시성을 효과적으로 학습하기 위해 전용 예측 헤드(head)를 추가 설계하고 식 (2)에서 제안한 가시성 손실 항 L_{vis} 을 손실 함수에 새롭게 도입하였다. 가시성 손실 함수의 가중치 λ_{vis} 는 분류 손실 가중치 λ_{cls} 와 동일한 0.5로, OKS Loss에서 각 keypoint에 적용되는 가중치 σ 는 $1/15=0.0667$ 로, 나머지 하이퍼파라미터는 YOLO 프레임워크의 기본 설정값을 적용하였다 [3].

3.2 실험 결과

테스트셋을 이용한 가시성 분류 실험 결과, keypoint는 OKS 지표를 활용하여 ground truth와 유사도를 측정하고, threshold 0.5를 기준으로 분류 성능 지표를 계산하였다 [3]. 제안 모델의 가시성 분류 성능을 평가한 결과, 표 1과 같이 평균적으로 recall 0.872, precision 0.900의 우수한 수

치를 기록하였다. 특히 가시성이 확보된 상태(Visible)에서 더욱 높은 정밀도를 보였으며, 가려진 상태(Occluded)에서도 안정적인 분류 성능을 유지함을 확인하였다.

표 1. 분류 성능 지표.

Table 1. Classification evaluation metrics.

	Occluded	Visible	Mean
Recall	0.812	0.931	0.872
Precision	0.883	0.915	0.900

다음은 객체별 keypoint 좌표의 위치 오차 정확도를 평가하기 위해, 정규화된 평균 오차(NME, Normalized Mean Error)를 산출하였다. NME는 true positive로 검출된 keypoint에 대해 유클리디안 거리를 구하고, 입력 이미지 해상도를 기준으로 정규화를 수행하였다. 그 결과, 평균 NME는 0.008로 도출되었으며, 안정적인 위치 추정 성능을 확인하였다.

그림 2는 제안한 방법을 통해 모형 사과의 keypoint를 예측하고 시각화한 결과로 제안한 모델이 과실의 형상 추정뿐만 아니라, 개별 keypoint의 가시성 여부를 실시간으로 동시에 식별할 수 있음을 확인하였다.

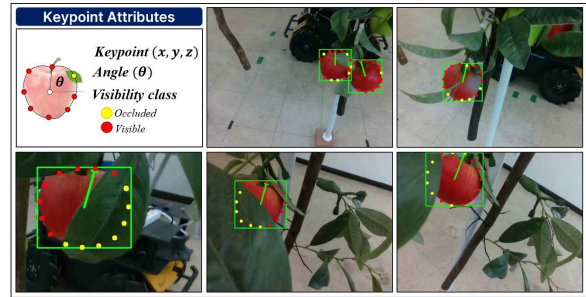


그림 2. 모형 사과 예측 결과.

Fig 2. Keypoint detection on synthetic apples.

4. 결론

본 연구에서는 keypoint 기반 과실의 가시성 정보를 구분할 수 있는 손실함수를 제안하였다. 실험 결과, 분류 성능과 위치 정확도 측면에서 안정적인 결과를 보였으며, 모형 사과 실험을 통해 폐색 환경에서 과실 keypoint의 위치 추정과 동시에 가시성을 분류할 수 있음을 검증하였다.

참 고 문 헌

- [1] J. Lee, J. Kim, "Occluded Region Restoration and Geometric Positioning of Fruits Using Deep Learning-Based Pose Estimation and Landmark Detection," *Journal of Korean Institute of Intelligent Systems*, vol. 35, no. 4, pp. 333-340, 2025.
- [2] D. Maji, S. Nagori, M. Mathew and D. Poddar, "YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss," , *arXiv preprint arXiv:2204.06806*, 2022.
- [3] G. Jocher, G. Qiu and A. Chaurasia, "Ultralytics YOLOv8", Available: <https://github.com/ultralytics/ultralytics>, [Accessed Date: Apr. 27, 2026].

† 교신저자

감사의 글 : 본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 고부가가치식품기술개발 사업의 지원을 받아 연구되었음(RS-2022-IP322054)