

# Anchor-Free 기반 3D 객체 검출을 이용한 과수 꼭지 검출 시스템 구현

이영욱\*, 노동희\*, 이혜민<sup>o</sup>

## Implementation of a Fruit Stem Detection System Using Anchor-Free-Based 3D Object Detection

Yeong-wook Lee\*, Dong-hee Noh\*, Hea-min Lee<sup>o</sup>

### 요약

본 논문은 과수 꼭지 제거를 위한 Anchor-Free 기반 3D 객체 검출 모델을 이용한 과수 꼭지 검출 시스템을 제안한다. 과수 꼭지는 과수의 위생, 식품 안전, 품질, 신선도 등에 영향을 주는 중요한 부분이다. 기존의 수작업에 의존하는 꼭지 제거 작업을 자동화하여 인력과 시간을 절감할 수 있으며, 관련 농업 분야에서의 효율성을 높일 수 있다. FCAF3D 모델은 3D 객체 검출 알고리즘으로, 과수의 몸체와 꼭지를 각각 예측하여 과수 꼭지의 작은 크기에도 높은 검출 성능을 보인다. 모델의 네트워크 구조는 ResNet, GSDN, FCOS 네트워크로 구성되어 있으며, FPN 구조를 통해 다양한 객체 스케일을 처리한다. 본 논문에서는 사과를 예시로 모델 학습을 하였고 학습된 모델은 사과 데이터셋에서 높은 정확도를 보여주었으며, 테스트 결과의 바운딩 박스 좌표값은 과수 꼭지 제거 시스템에 활용할 수 있다. 실험 결과를 통해 FCAF3D 모델은 과수 꼭지 검출에 높은 성능을 보이는 것으로 나타났다.

**키워드** : 딥러닝, 디지털 농업, 3D 객체 검출, 과수 꼭지 검출

**Key Words** : Deep learning, Digital agriculture, 3D Object detection, Fruit stem detection

### ABSTRACT

This paper proposes a fruit stem detection system using Anchor-Free-based 3D object detection model for fruit stem removal. The fruit stem is an important part that affects the hygiene, food safety, quality, and freshness of fruits. It can save manpower and time by automating existing manual-dependent tap removal operations, and increase efficiency in related agricultural fields. The FCAF3D model is a 3D object detection algorithm that predicts the skin and stem of the fruit, respectively, showing high detection performance even in the small size of the fruit stem. The network structure of the model consists of ResNet, GSDN, and FCOS networks, which handle various object scales through the FPN structure. In this paper, model training was performed using apples as an example, and the learned model showed high accuracy in the apple dataset, and the bounding box coordinate value of the test results can be used in the fruit stem removal system. Experimental results showed that the FCAF3D model showed high performance in detecting fruit stem.

※ 본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 노지분야 스마트농업기술 단기 고도화사업의 지원을 받아 연구되었음(122032-03)

♦ First Author : Korea Electronics Technology Institute, thisis206@keti.re.kr, 정희원

° Corresponding Author : Korea Electronics Technology Institute, lee10849@keti.re.kr, 정희원

\* Korea Electronics Technology Institute, dheenoh@keti.re.kr, 정희원

논문번호 : 202307-014-0-SE, Received July 14, 2023; Revised September 13, 2023; Accepted October 24, 2023;

## 1. 서 론

과일은 전세계적으로 일상에서 떼어놓을 수 없는 영양 식품이다. 그 중 특히 사과는 우리나라에서 가장 많이 소비되는 과일 중 하나다. 2021년 기준 우리나라 총 과수 생산량 2,109t 중 사과는 516t으로, 636t인 감귤 다음으로 생산량이 많다<sup>[1]</sup>. 또한 사과는 1인당 연간 과수 소비량에서 상당한 부분을 차지한다. 이처럼 사과는 평상시 섭취뿐만 아니라 제사, 명절 선물 등 많은 상황에서 소비가 된다. 사과와 마찬가지로 대부분의 과수는 수확할 때 과수의 꼭지를 따서 수확하고 가공 및 처리 과정에서 과수 꼭지를 제거한다.

과수 꼭지 제거는 과수를 가공 또는 소비하기 전에 필요한 작업 중 하나로, 과수의 품질과 식품 안전을 유지하기 위해 중요하며 식품 가공 및 소비 과정에서 다양한 문제를 야기할 수 있다.

위생 및 식품 안전: 과수 꼭지는 과수에 미생물이나 먼지 등의 오염 물질이 존재할 수 있는 부분이다. 또한, 과수 꼭지 주변에는 토양에 사용된 비료나 살충제 등의 잔류물이 존재할 수도 있다. 따라서, 과수 꼭지를 제거함으로써 위생 및 식품 안전을 보장할 수 있다.

품질 및 신선도: 과수 꼭지를 제거하지 않고 유통할 경우 유통 과정에서 과수 꼭지가 다른 과수의 외과피에 긁힘 등의 손상을 줄 가능성이 있다. 외과피가 긁혀 흠집이 난다면 이는 품질 저하를 일으키고 내부로 산소, 물 또는 흙 등이 유입되어 신선도가 감소하고 부패의 위험이 증가한다.

과수 꼭지 제거는 위에서 언급한 이유로 인해 중요한 작업이다. 하지만 농업 분야에서 과수 수확과 관련된 안정적인 파지 방법<sup>[2]</sup>이나 그리퍼<sup>[3]</sup> 개발 등의 연구는 과거부터 진행되고 있지만 지금까지 과수의 꼭지 제거와 관련된 기계 장치 개발은 전무하며 전적으로 수작업에 의존해야 하는 수확, 꼭지 제거 작업은 단기간 내 노동 집약도가 가장 높아 시급히 기계화, 자동화를 통해 인력 투입을 줄여야 하는 분야이다. 2012년 한국농수산대학에서 수행한 연구 용역 결과에 따르면 수확 후 꼭지 절단을 위한 인건비는 여성 노임으로 계산하였을 경우 전국 재배면적 기준 연간 약 200억원에 달하는 결과가 나왔으며 생산 현장에서는 수확기 노동력의 35%가 꼭지 절단 작업에 투입되어 많은 시간을 소요한다.<sup>[4]</sup> 우리나라 과원은 아직 인력 의존적 생산체계에 머물러 있는 상황이고 급격하게 상승하는 인건비와 빠르게 진행되는 농촌 인구 고령화로 인한 인력 수급 문제로 노동력 절감은 필수적이다.

따라서, 과수 꼭지를 정확하게 검출하고 제거하는 시

스템을 구현하여 과수의 시각적 외관을 개선하고, 위생과 식품 안전을 유지하며, 가공 및 조리 과정에서의 편의성을 높일 뿐만 아니라 수확 후 처리 과정에서 소요되는 많은 노동 시간을 줄임으로써 작업 효율을 증대시킬 수 있다.

그림 1은 과수 꼭지 제거 시스템의 개요도이다. 과수 꼭지를 제거하기 위한 로봇을 개발하기 위해 우리는 3D 객체 검출 알고리즘을 사용하여 딥러닝 학습을 한다. 본 논문에서는 사과를 대표적인 예로 들어 실험하였다. 깊이 카메라로 사과 데이터셋을 학습시켜 사과의 꼭지를 검출해 낼 수 있는 시스템을 구현한다. 사과의 꼭지 부분을 인식하기 위해서는 매우 작은 물체에 대한 검출을 해야한다.

본 논문에서는 FCAF3D(Fully Convolutional Anchor-Free 3D Object Detection)<sup>[5]</sup> 모델을 이용하여 과수 꼭지 검출 시스템을 구현한다. FCAF3D는 객체 검출을 위한 기술 중 하나로, Anchor-Free라는 개념을 기반으로 하여 기존의 Anchor 기반 방식과 달리 객체의 위치와 크기를 미리 정의된 Anchor Boxes 없이 직접 예측하여 다양한 객체 모양에 맞게 유연하게 생성할 수 있으며 비교적 단순한 예측 프로세스로 훈련하기가 쉽고 유연한 바운딩 박스 생성의 특징으로 작은 객체에 대한 성능이 우수하다. 또한 과수 몸체와 꼭지를 각각 예측했을 때 바운딩 박스가 겹쳐서 발생하는 문제를 해결할 수 있다.

본 논문에서는 매우 작은 크기의 사과 꼭지의 특징을 고려한 데이터셋을 구성하고 모델을 학습시킨 후 테스트 데이터에 대해 검출 성능을 평가한다. 논문의 구성은 다음과 같다. 제 II장에서는 FCAF3D 모델을 이용한 과수 꼭지 검출 시스템에 관하여 설명하고 제 III장에서는 실험환경 및 실험방법에 관하여 기술하고 실험 결과를 이용하여 정확도 및 손실함수를 도출하고 특징을 비교, 분석한다.

그리고 제IV장에서 결론을 도출해내고 실제 응용 분야

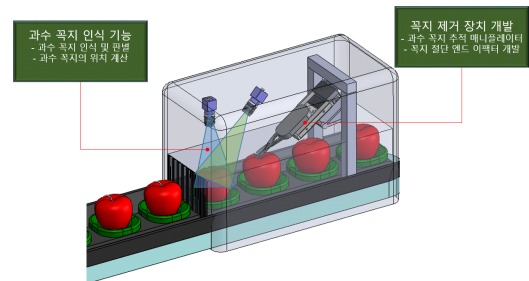


그림 1. 과수 꼭지 제거 시스템  
Fig. 1. Fruit stem removal system

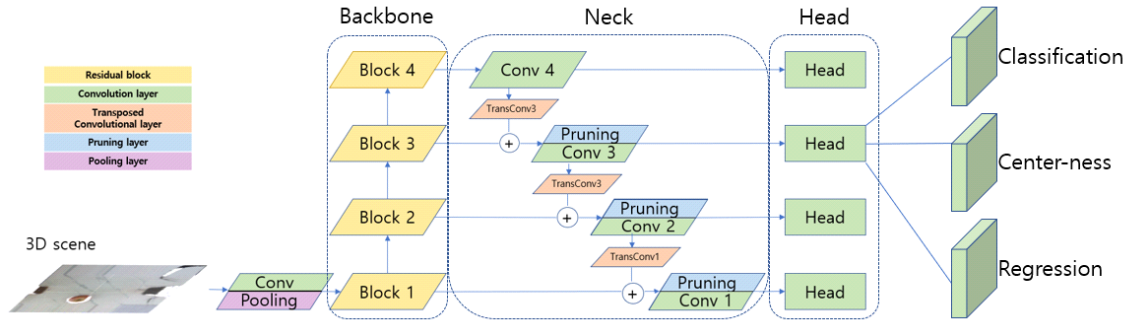


그림 2. FCAF3D 네트워크 구조  
Fig. 2. FCAF3D Network structure

에서의 활용 가능성 및 향후 연구 방향에 대해 논의한다.

## II. 본 론

FCAF3D의 네트워크는 그림 2와 같다. FCAF3D는 백본, 넥, 헤드로 구성되며 각 네트워크는 다음과 같이 구성되어있다.

백본에서는 ResNet<sup>[6]</sup>을 사용하였다. ResNet은 일반적인 2D CNN 구조에서 Residual 블록을 추가하여 스킵 연결을 통해 잔차를 학습하도록 만들어진 인공 신경망으로, 일반적인 딥러닝 모델보다 예측 정확도가 높다.

넥에서는 GSDN<sup>[7]</sup> 모델을 사용하여 입력으로 복셀화 된 3D 데이터에 대해 Convolution 계층 생성 및 Convolution 연산을 통해 다운샘플링을 수행한다. Convolution 연산할수록 계층은 높아지며 계층이 높아질수록 더 큰 영역을 예측한다. 이후 Transposed Convolution<sup>[8]</sup>을 통한 업샘플링으로 3D 복셀 공간을 생성하여 이전 단계에서 발생한 정보 손실을 최소화한다. 이후 Pruning 기법으로 불필요한 복셀을 제거하고 Center-ness 후보 영역만 남기고 제거되지 않은 모든 점에 대해 바운딩 박스를 예측한다. 불필요한 복셀을 제거함으로써 연산량을 줄이고 예측 프로세스를 간소

화할 수 있다.

헤드에서는 FCOS(Fully Convolutional One-stage Object Detection)<sup>[9]</sup> 네트워크를 사용하였다. FCOS는 대표적인 Anchor-Free 네트워크로 GSDN과 마찬가지로 FPN(Feature Pyramid Network)<sup>[10]</sup> 구조를 도입하여 넥 네트워크에서 계층별로 구분된 3D 데이터의 분류 결과와 회귀 값, Center-ness를 출력한다. Center-ness는 객체의 중심에서 멀리 있는 값들의 영향을 줄이기 위해 도입되었으며 중심점과 가까울수록 0~1의 값을 가지며 중심과 멀리 있는 점은 NMS(Non-Maximum Suppression)를 통해 제거된다.

FCAF3D 모델에서 사용된 전체적인 손실함수는 식 (1)과 같다.

$$L = \frac{1}{N_{pos}} \sum_{x,y,z} (L_{cls}(\hat{p}, p) + \mathbb{1}_{\{p_x, \hat{p}_x \neq 0\}} L_{reg}(\hat{b}, b) + \mathbb{1}_{\{p_x, \hat{p}_x \neq 0\}} L_{centr}(\hat{c}, c)). \quad (1)$$

분류 손실  $L_{cls}$  은 focal loss, 회귀 손실  $L_{reg}$  은 IoU loss, center-ness 손실은 binary cross-entropy loss를 사용하였다.

본 논문에서는 FCAF3D 모델을 이용하여 과수 꺾기

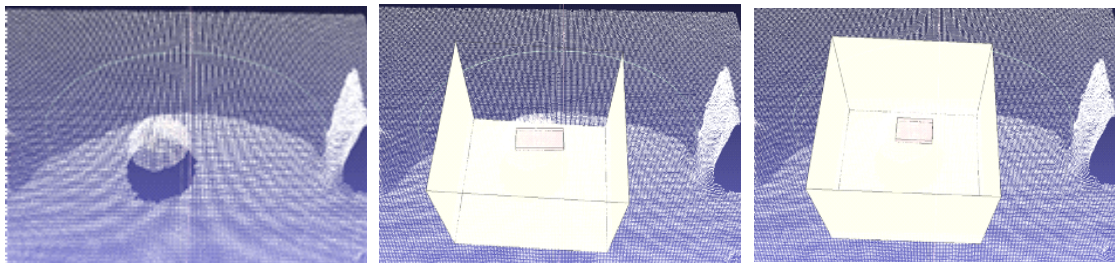


그림 3. 학습 결과 시각화(좌: Input data, 중: Ground truth, 우: Predicted value)  
Fig. 3. Visualization training result(Left: Input data, Middle: Ground truth, Right: Predicted value)

검출 시스템을 구현한다. 데이터로 사용된 사과는 사과와 몸체와 꼭지, 두 클래스로 나누어 라벨링을 실시한다. 사과 몸체의 포인트 개수는 비교적 많지만 사과 꼭지의 경우 객체 크기가 작기 때문에 포인트의 개수도 매우 적다. 그러므로 동일한 Convolution 계층에서 연산하여 학습한다면 꼭지 객체의 검출 정확도가 매우 낮아지는 문제가 발생한다. 이를 해결하기 위해 FCAF3D 모델 내에서 FPN 구조를 도입하여 객체의 스케일을 계층별로 나누어 연산하고 학습하여 작은 객체인 꼭지 또한 검출 정확도를 향상한다.

### III. 실험

#### 3.1 실험방법

본 논문에서는 FCAF3D 모델을 통해 과수 꼭지 검출 시스템을 구현하고 모델의 특징과 정확도를 이용한 분석을 진행하였다. 실험 샘플은 과수의 몸체와 꼭지를 검출하기 가장 용이한 사과로 선정하였으며 실험에 사용할 데이터셋을 수집하기 위해 미국 Intel사의 Realsense D405 깊이 카메라를 통해 사과 데이터를 수집하였고 제품에 대한 상세 정보는 그림 4에 나타나있다. 사과 데이터는 암실 박스 내에서 촬영하였고 조명을 통해 조도를 정밀하게 조절하여 최적의 포인트를 감지할 수 있도록 환경을 구성하고 카메라의 Calibration 데이터를 모두 동일하게 생성하였다. 사과 데이터 수집 환경에 대한 예시는 그림 5에 나타나있다.

Calibration 데이터는 다음과 같이 두 개의 3x3 행렬로 구성되어 있으며 각각은 카메라가 기울어진 정도, 카메라 intrinsic 값을 나타낸다.

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

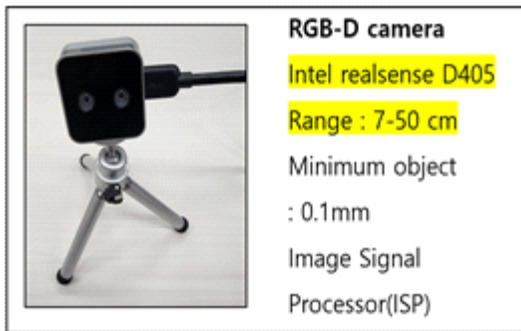


그림 4. Realsense D405 깊이 카메라 정보  
Fig. 4. Realsense D405 depth camera information

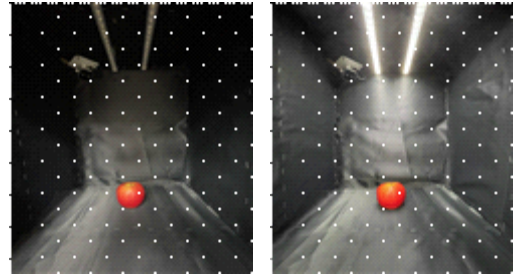


그림 5. 데이터셋 수집 환경  
Fig. 5. Dataset collection environment

$$Y = \begin{bmatrix} 0.491 & 0.784 & 0.501 \\ 0.490 & -0.060 & 0.069 \\ 0 & 0 & -0.0022 \end{bmatrix} \quad (3)$$

또한 깊이 카메라를 이용하여 2D 이미지 데이터와 3D 포인트 클라우드 데이터를 수집하였다. 수집된 포인트 클라우드 데이터는 라벨링 툴을 이용하여 사과와 몸체, 꼭지 두 클래스로 나누어 라벨링을 실시하였다. 준비된 데이터셋을 트레인셋과 테스트셋을 8:2의 비율로 나누었고 두 클래스 모두 약 2,200개의 데이터셋으로 구성하여 FCAF3D 모델에 학습하였다.

#### 3.2 실험환경

실험에서 사용한 카메라는 Realsense D405 깊이 카메라를 사용하였으며 그림 4와 같이 7~50cm 범위를 촬영하기에 적합하고 FCAF3D 모델을 학습한 GPU는 미국 NVIDIA사의 NVIDIA RTX Titan이며, 제품의 자세한 정보는 표 1에 나타나있다.

우리는 MMDetection3D<sup>[11]</sup> 프레임워크를 이용하여 FCAF3D를 구현하였으며 학습 절차는 기본 MMDetection 체계를 따른다. 총 12회에 걸쳐 학습을

표 1. GPU 상세정보  
Tabel 1. GPU detail

GPU Memory	24GB GDDR6
Memory Interface	384-bit
Memory Bandwidth	Up to 672 GB/s
NVIDIA CUDA Cores	4,608
NVIDIA Tensor Cores	576
NVIDIA RT Cores	72
NVIDIA NVLink	Connects 2 TITAN RTX GPUs
NVIDIA NVLink Bandwidth	100GB/s
Power Comsumption	280W



진행하였으며 시작 학습률은 0.001, 8번째, 11번째 학습에서 학습률을 0.0001로 감소시켰다. 하이퍼파라미터는 표 2와 같이 설정하였다.

표 2. 모델 하이퍼파라미터  
Table 2. Model hyperparameter

Hyperparameter	Value
$N_{pts}$	1,000,000
$N_{vox}$	0.001
$IOU_{threshold}$	0.5
$NMS$	1

### 3.3 실험결과

FCAF3D 모델을 이용한 사과 몸체와 사과 꺾지에 대한 학습 결과는 표 3에 나타나 있다. ‘Skin’ 클래스는 mAP와 mAR이 각각 0.98, 1.0으로 나타났고 ‘Stem’ 클래스는 0.76, 0.73으로 나타났다. ‘Skin’ 클래스의 경우, 사과의 몸체부분을 나타내며 대부분의 데이터에서 객체의 형태, 포인트의 수 등이 매우 유사한 형태이기 때문에 높은 정확도와 응답률을 보였다. 반면 ‘Stem’ 클래스는 사과의 꺾지 부분으로, 객체의 크기가 ‘Skin’ 클래스에 비해 상당히 작고 포인트의 수 또한 적다. 두 클래스의 포인트 수 차이 및 라벨링 예시는 그림 6에 나타나 있다. ‘Skin’ 클래스에 비해 ‘Stem’ 클래스의 mAP와 mAR의 수치가 낮은 이유는 객체의 크기가 작고 학습에 사용할 수 있는 포인트의 수가 한정적이기 때문에 검출하기 어렵기 때문이다. 손실함수는 classification loss, bbox loss, centerness loss로 나누어 결과를 도출하였으며 그림 8에 나타나 있다. 손실함수를 통해 iteration이 진행될수록 가중치가 최적화 되는 것을 확인할 수 있다. NMS 파라미터를 설정하기 전후 결과는 그림 7에 나타나 있다. NMS 파라미터를 설정함으로써 예측 바운딩박스가 과도하게 생성되는 경우를 줄여줄 수 있다. 또한 학습된 모델을 바탕으로 테스트를 한 후 예측된 바운딩 박스의 좌푯값을 출력하였다. 출력된 좌푯값은 그림 9에 나타나 있다. 그림 9에서 확인할 수 있듯이 두 클래스의 예측된 바운딩 박스 좌표(x, y, z)와 RGB 값을 출력하고 바운딩 박스의 각 꼭짓점을

표 3. 클래스별 mAP  
Table 3. mAP by class

Class	mAP	mAR
Skin	0.98	1.0
Stem	0.76	0.73



그림 6. 사과 클래스와 꺾지 클래스의 포인트 비교  
Fig. 6. Point comparison of skin class and stem class

도출해내었다. 이 좌푯값은 향후 ROS(Robot Operation System)과 연동하여 좌표를 전송하여 과수 꺾지의 위치를 추정하는 연구로 진행할 수 있다.

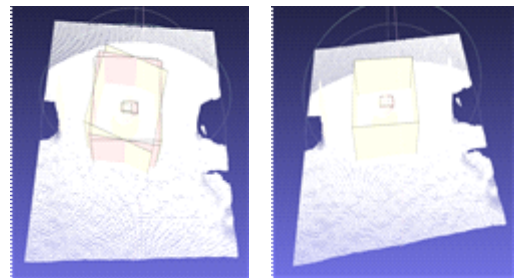


그림 7. NMS 파라미터 조정(좌: NMS\_pre=1000, 우: NMS\_pre=1)  
Fig. 7. Adjusting NMS Parameter(Left: NMS\_pre=1000, Right: NMS\_pre=1)

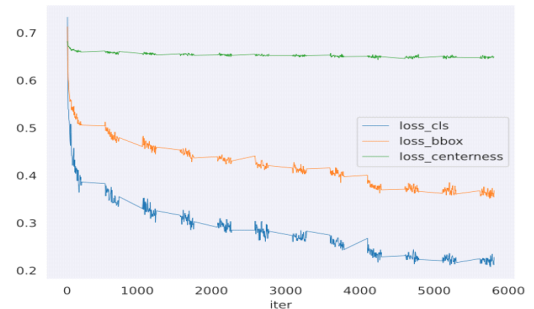


그림 8. 손실 그래프  
Fig. 8. Loss graph

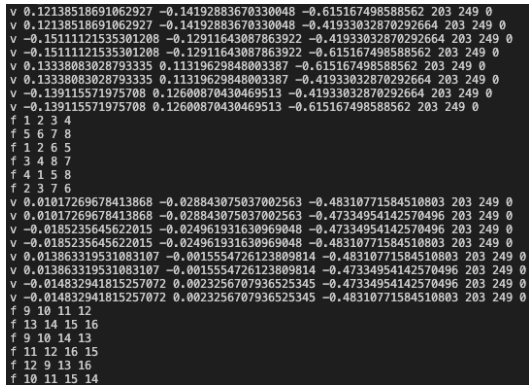


그림 9. 예측된 바운딩 박스 좌표  
 Fig. 9. Predicted bounding boxes location

#### IV. 결 론

본 논문에서는 3D 객체 검출 모델을 이용하여 과수 꼭지 검출 시스템을 구현하기 위해 사과와의 2D 이미지, 3D 맵스 이미지, 라벨링 데이터, calibration 데이터, 총 4가지 형식으로 데이터셋으로 구성하고 모델의 하이퍼 파라미터를 조정하여 학습하였고 손실함수와 정확도를 도출하였다. 과수 중 가장 보편적인 예로 사과를 선정하였고 사과 몸체가 사과 꼭지에 비해 더욱 높은 정확도를 보여주었지만 포인트의 수가 상당히 적은 사과 꼭지 객체의 특징을 감안하면 사과 꼭지의 정확도는 우수한 성능을 보여주는 것으로 확인되었다. 또한 NMS 파라미터를 조정하여 가장 높은 점수의 바운딩 박스만 출력할 수 있도록 설정하였고 학습된 모델을 이용하여 테스트를 진행하고 좌푯값을 도출해내었다. 출력된 좌푯값은 최종적인 사과 꼭지 제거 로봇 시스템 구현에 필요한 정보로, 향후 과수 꼭지 검출 시스템을 이용한 꼭지 제거 로봇 연구에 많은 기여를 할 것이다. 또한 추후 연구로 Anchor-Free 기반 3D 객체 감지 모델 중 FCAF3D 모델뿐만 아니라 GroupFree, TR3D 모델 및 Anchor 기반 모델들 또한 구현하고, 사과 외에 감귤, 복숭아, 배와 같은 과수에 대한 실험도 추가적으로 진행하여 과수 꼭지 검출에 가장 적합한 모델을 연구할 예정이며 이는 수작업에 의존하고 있는 우리나라 농업 분야의 발전에 기여할 것이다.

#### References

[1] Major Statistics of Agriculture, *Food and Rural Affairs 2022*, [https://www.nihhs.go.kr/farme/r/statistics/statistics.do?t\\_cd=0203](https://www.nihhs.go.kr/farme/r/statistics/statistics.do?t_cd=0203)

[2] D. S. Choi, et al., “Improved design for enhanced grip stability of the flexible gripper in harvesting robot,” *J. Korea Robotics Soc.*, vol. 15, no. 2, pp. 107-114, 2020. (<https://doi.org/10.7746/jkros.2020.15.2.107>)

[3] H. S. Kang, et al., “Robotic gripper with flexible structure for harvesting fruits. in institute of control,” *Robotics and Syst. ICROS 2019*, pp. 127-128, Gyeongju, Korea, May 2019.

[4] S. J. Ko, *Distribution without cutting the tip of the apple...Expectations to reduce production costs and maintain freshness*, 12, 24, 2021, <http://www.agrinet.co.kr/news/articleView.html?idxno=305876>

[5] D. Rukhovich, A. Vorontsova, and A. Konushin, “Fcaf3d: Fully convolutional anchor-free 3d object detection,” *Eur. Conf. Comput. Vision*. Cham: Springer Nature Switzerland, 2022.

[6] K. He, et al. “Deep residual learning for image recognition,” in *Proc. IEEE Conf. CVPR* 2016.

[7] J. Y. Gwak, C. Choy, and S. Savarese, “Generative sparse detection networks for 3d single-shot object detection,” *Computer Vision - ECCV 2020: 16th Eur. Conf.*, Glasgow, UK, Aug. 2020.

[8] H. Gao, et al., “Pixel transposed convolutional networks,” *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 42, no. 5, pp. 1218-1227, 2019.

[9] Z. Tian, et al., “Fcos: Fully convolutional one-stage object detection,” in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, 2019.

[10] T.-Y. Lin, et al., “Feature pyramid networks for object detection,” in *Proc. IEEE Conf. CVPR*, 2017.

[11] M. Contributors, *MMDetection3D: OpenMML abnext-generation platform for general 3D object detection*. <https://github.com/open-mmlab/mmdetection3d> (2020)

**이 영 옥 (Yeong-wook Lee)**



2020년 2월 : 금오공과대학교 전  
자공학부 졸업  
2022년 2월 : 금오공과대학교 IT  
융복합공학과 석사  
2022년 9월~현재 : 한국전자기  
술연구원 IT응용연구센터 연  
구원

<관심분야> 스마트팜, 딥러닝, 영상처리

[ORCID:0000-0003-2180-6161]

**이 혜 민 (Hea-min Lee)**



2014년 2월 : 금오공과대학교 전  
자공학 학사  
2016년 2월 : 금오공과대학교 전  
자공학 석사  
2021년 2월 : 경북대학교 전자공  
학 박사 수료  
2016년 2월~2021년 12월 : 한국

전자통신연구원 대경권연구센터

2021년 1월~현재 : 한국전자기  
술연구원 IT응용연구센  
터 연구원

<관심분야> 제어공학, 임베디드 시스템, 자율주행

[ORCID:0000-0001-9573-2500]

**노 동 희 (Dong-hee Noh)**



2013년 2월 : 금오공과대학교 전  
자공학 학사  
2015년 2월 : 금오공과대학교 전  
자공학 석사  
2018년 4월~현재 : 한국전자기  
술연구원 IT응용연구센터 선  
임연구원

<관심분야> 스마트팜, 빅데이터, 디지털 트윈, 로봇 제  
어 시스템

[ORCID:0000-0001-6931-2377]