

## UAV 항공 영상에서의 딥러닝 기반 잣송이 검출

## Deep Learning Based Pine Nut Detection in UAV Aerial Video

김 규 민<sup>1</sup> · 박 성 준<sup>1</sup> · 황 승 준<sup>1</sup> · 김 희 영<sup>2</sup> · 백 중 환<sup>1\*</sup>

<sup>1</sup>한국항공대학교 항공전자정보공학부

<sup>2</sup>링크투

Gyu-Min Kim<sup>1</sup> · Sung-Jun Park<sup>1</sup> · Seung-Jun Hwang<sup>1</sup> · Hee Yeong Kim<sup>2</sup> · Joong-Hwan Baek<sup>1\*</sup>

<sup>1</sup>School of Electronics and Information Engineering, Korea Aerospace University, Gyeonggi-do 10540, Korea

<sup>2</sup>LinktoTo Co. Ltd, Gyeonggi-do 16897, Korea

### [요 약]

잣은 우리나라 대표적인 견과류 임산물이자 수익형 작물이다. 그러나 잣송이는 사람이 직접 나무 위로 올라가 수확하기 때문에 위험성이 높다. 이러한 문제를 해결하기 위해서 로봇 또는 UAV(unmanned aerial vehicle)를 이용한 잣송이 수확이 필요하다. 본 논문에서는 UAV를 이용한 잣송이 수확을 위해 UAV 항공 영상에서 딥러닝(deep learning) 기반의 잣송이 검출 기법을 제안한다. 이를 위해, UAV를 이용하여 실제 잣나무 숲에서 동영상 촬영했으며, 적은 수의 데이터 보안을 위해 데이터 증강기법을 사용했다. 3D 검출을 위한 데이터로는 Unity3D를 이용하여 가상 잣송이 및 가상환경을 3D 모델링 하였으며 라벨링은 좌표계의 3차원 변환법을 이용해 구축했다. 잣 분포 영역 검출, 잣 객체에 대한 2D 및 3D 검출을 위한 딥러닝 알고리즘은 DeepLabV3, YOLOv4, CenterNet을 각각 이용하였다. 실험 결과, 잣송이 분포 영역 검출률은 82.15%, 2D 검출률은 86.93%, 3D 검출률은 59.45%이었다.

### [Abstract]

Pine nuts are Korea's representative nut forest products and profitable crops. However, pine nuts are harvested by climbing the trees themselves, thus the risk is high. In order to solve this problem, it is necessary to harvest pine nuts using a robot or an unmanned aerial vehicle(UAV). In this paper, we propose a deep learning based detection method for harvesting pine nut in UAV aerial images. For this, a video was recorded in a real pine forest using UAV, and a data augmentation technique was used to supplement a small number of data. As the data for 3D detection, Unity3D was used to model the virtual pine nut and the virtual environment, and the labeling was acquired using the 3D transformation method of the coordinate system. Deep learning algorithms for detection of pine nuts distribution area and 2D and 3D detection of pine nuts objects were used DeepLabV3+, YOLOv4, and CenterNet, respectively. As a result of the experiment, the detection rate of pine nuts distribution area was 82.15%, the 2D detection rate was 86.93%, and the 3D detection rate was 59.45%.

**Key word** : Unmanned aerial vehicle, Pine nut, Segmentation, Detection.

<https://doi.org/10.12673/jant.2021.25.1.115>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 31 January; Revised 1 February 2021

Accepted (Publication) 26 February (28 February 2021)

\*Corresponding Author: Joong-Hwan Baek

Tel: +82-2-300-0125

E-mail: [jhbaek@kau.ac.kr](mailto:jhbaek@kau.ac.kr)

## 1. 서 론

잣은 Korean pine nut이라 불리며, 우리나라의 대표적인 견과류 임산물에 속할 뿐만 아니라 경기도를 대표하는 관광 테마이자 수익형 작물이다. 잣 수확은 대개 8월 말부터 11월까지 이루어지며 잣나무 한그루에서는 보통 3년에 한 번 수확한다. 그러나 높이 20~30 미터의 잣나무 꼭대기에서 생육하는 잣 열매를 채취한다는 것은 사람이 직접 나무 위로 올라가 장대로 떨어뜨리는 방법 이외에는 현재 다른 방법이 없는 것이 문제이며, 인명사고 발생 시 사고처리에도 큰 비용이 발생한다. 이러한 수작업의 위험성을 줄이고 효율적으로 수확하기 위해서는 잣의 분포도 생성 및 과실 검출 기술이 필요하다. 최근 스마트농업 분야에서 드론을 이용하여 농지와 작물의 생육상태를 진단하여 재배 계획 및 수확량을 예측하는 연구가 활발히 진행 중이다 [1], [2]. 또한, 농업용 로봇을 이용하여 작물의 생육환경에 대한 모니터링부터 제초, 방제, 살포, 수확, 이식, 착유 등 다양한 분야에서도 활용되고 있다 [3]. 스마트농업에서 CNN (convolutional neural network) 기반 과실 분포도 생성과 검출은 자동화를 위한 중요한 구성 요소이다. 기존 연구로 블록 분류를 기반으로 비닐봉지에 사과를 넣어 인식하는 방법으로 워터셰드 (watershed) 알고리즘을 채택하여 원본 이미지를 불규칙한 블록으로 분할 한 다음 이러한 블록을 VSM (variance shadow map)에 의해 과일 블록과 비 과일 블록으로 분류하는 방법을 제안했다 [4]. Choi는 깊이 영상에서 3차원 정보를 이용하여 지도에서 과실 영역을 식별하고 CNN 기반의 네트워크를 이용하여 분류했다 [5]. Tian은 YOLOv3 네트워크의 성능을 향상하여 사과의 성장 과정을 분류했다 [6]. 이외에 가려진 과실 검출을 위해 Mask-RCNN을 사용하며 좋은 성능을 보였다 [7].

최근 인공지능 기술이 딥러닝과 같은 높은 성능을 내는 방법들로 인해서 비약적으로 발전하고 있으며, 이러한 인공지능 기술을 통해 카메라로부터 얻어진 2차원 영상을 이용해 3차원 정보를 정확하게 인지하는 기술들이 활발하게 연구되고 있다 [8]-[10]. 이러한 카메라 영상은 다른 센서들에 비해 상대적으로 비용이 저렴하고 간단하며, 스마트폰이나 블랙박스 등의 보급으로 인해 누구나 쉽게 영상을 얻을 수 있고, 인터넷에 영상 기반의 방대한 데이터가 존재한다. 정확한 과실 수확을 위해서는 대상 객체에 대한 정확한 3차원 정보를 얻는 것이 중요하다. 정확한 3차원 정보 획득을 위해 라이다/레이더 센서, RGB+D 카메라가 이용된다 [11], [12]. 그러나 위와 같은 카메라 또는 센서들은 깊이 정보가 비교적 정확하나 비용이 많이 들고 복잡하다. 또한, 직접 과실 수확을 위해 데이터를 구축해야 하는 경우, 데이터를 취득하고 라벨링을 하는 데 많은 시간과 비용이 드는 문제가 있어 실제 데이터를 많이 모으기 힘들다.

본 논문에서는 잣 수확을 위해 딥러닝 알고리즘을 활용한 UAV 항공영상 기반의 잣 분포 영역 검출, 잣 객체에 대한 2D 검출, 그리고 가상환경에서의 3D 검출 및 포즈 추정 기법을 제안한다. 제안하는 방법은 그림 1에 보인다. UAV를 이용해 잣

무 숲 지역을 촬영하고, 입력되는 영상에서 영상분할 및 2D 검출 기법을 이용하여 객체의 경계와 영역을 추출하여 잣송이의 분포도를 생성한다. 3D 검출을 통해서는 깊이 및 크기를 추출하여 향후 로봇 또는 UAV를 통한 잣송이 수확을 위한 3D 자세를 추정한다. UAV를 이용하여 잣송이 검출할 경우 수확 시기에 잣 열매가 많이 포진된 지역을 바로 알 수 있어 빠르게 이동 가능하며 사람이 직접 나무 위로 올라가 수확하는 위험성을 줄일 수 있다. 본 논문에서 제안하는 방법을 구현하기 위해 축령산 잣나무 숲에서 UAV를 통해 잣나무 항공 영상을 취득하였으며 딥러닝 기반의 객체 검출 기법인 DeepLabV3, YOLOv4, CenterNet [13]-[15]을 기반으로 잣송이 분포 영역, 2D 및 3D 검출을 수행하였다. 검출의 성능 및 부족한 데이터를 위해 여러 데이터 증강기법을 적용하였으며 3D 검출은 Unity3D 게임 엔진을 활용하여 실제 잣나무 숲과 흡사한 가상공간을 모델링 하여 딥러닝 학습 및 평가에 사용하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 학습과 평가에 사용되는 데이터 생성 및 구축 방법을 설명하고, 3장에서는 잣송이 수확을 위한 딥러닝 기반 검출 기법을 제안한다. 4장에서는 2, 3장에서 제안한 방법을 적용하여 잣 분포 영역, 2D 및 3D 검출 실험 결과 및 고찰을 보인다. 마지막 5장에서는 결론을 맺으며 향후 연구 방향을 제시한다.

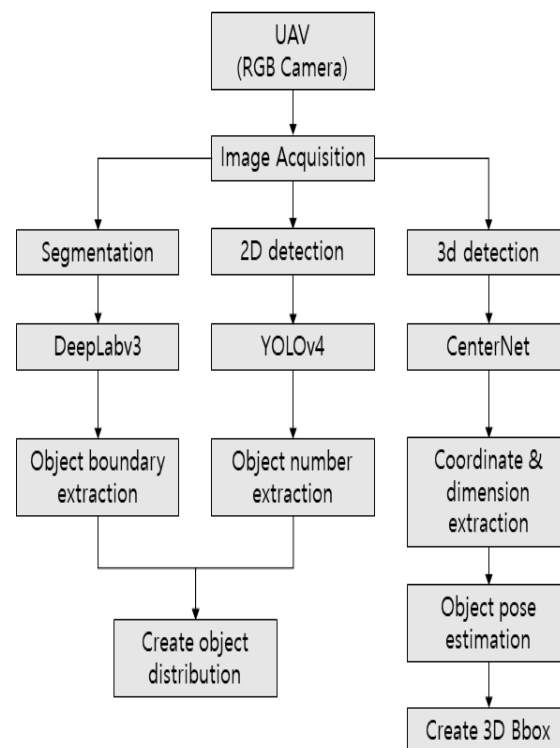


그림 1. 잣송이 검출 기법의 전체적인 구조도

Fig. 1. Overall structure diagram of pine nut detection.

## II. 데이터 생성

### 2-1 항공 영상 수집

본 논문에서는 잣송이 분포 영역 및 2D 검출 학습자료와 정확도 검증을 위한 데이터 확보를 위해 DJI사에서 제작한 Phantom3 Pro기체(그림 2)를 사용하였다. Phantom3 Pro는 회전 날개 기체로 조작이 쉽고 2축 짐벌을 탑재하고 있어 레저용 및 영상 촬영용으로 많이 사용되는 장비이다. 수동비행으로 잣나무 숲 꼭대기를 기준으로 고도 10~30m에 호버링(hovering)하여 여러 방향으로 동영상을 촬영하였다. Phantom 3 Pro 기체의 제원은 Table. 1과 같다.

잣송이 분포 영역 생성 및 2D 검출 학습을 위한 영상 촬영 지역은 경기도 가평에 있는 축령산 잣나무 숲에서 진행했다. UAV를 통해 수집한 잣나무 숲 항공 영상은 그림 3과 같다.



그림 2. 팬텀 3 프로  
Fig. 2. Phantom 3 Pro.

표 1. 팬텀 3 프로 제원

Table 1. Specifications of Phantom 3 Pro.

Type	Rotary Wing(Quad copter)
Size	350mm
Weight	1,280g
Accuracy of GPS hovering	Vertical: 1.5m Horizontal: 0.1m
Receiver Coverage	5km
Max. speed	57.6km/h
Max. altitude	6,000m
Max. flight time	23 minutes
Camera FOV	94° 20
Number of Video Pixels	4096 x 2,160p



그림 3. UAV를 이용한 잣나무 숲 영상 예시  
Fig. 3. Example of video acquired using UAV.

### 2-2 데이터 증강(Data Augmentation)

잣나무의 학명이 Korean Pine임에도 불구하고, 현재 우리나라의 잣나무와 잣나무 숲 그리고 잣송이에 대한 학습데이터가 부족하다. UAV를 통해 취득한 잣나무 영상은 비슷할뿐더러 적은 수의 데이터 세트로 인해 모델이 과적합(overfitting) 되는 현상을 볼 수 있다. 또한, 취득한 영상의 잣나무에서 잣송이 1개의 픽셀은 평균 19x19픽셀로 4K 영상에서 0.03%의 비율을 차지하며 딥러닝 모델이 검출하기 작은 크기이다. 이와 같은 문제점을 보완하기 위해 그림 4와 같이 자르기, 확대, 반전, 회전, 이동, 이미지 병합, 그리고 밝기조절의 데이터 증강방법을 수행했다.

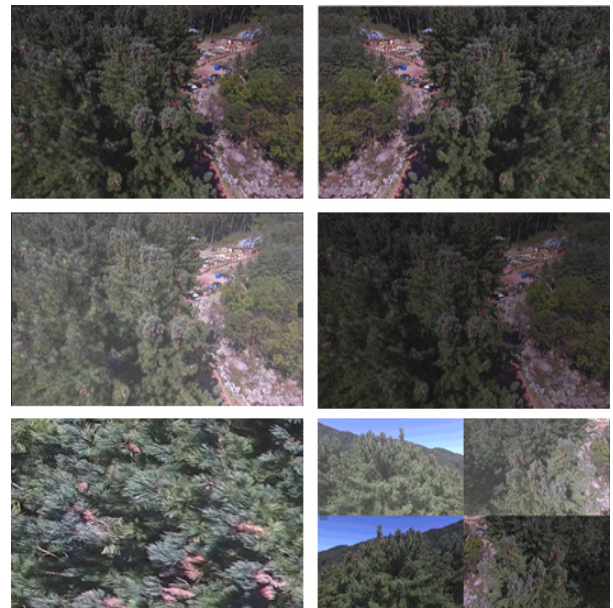


그림 4. 잣나무 숲 영상 데이터 증강 예시  
Fig. 4. Examples of pine nut forest data augmentation.



## 2-3 가상 데이터 생성

갯송이 3D 검출을 위한 딥러닝 모델 학습 시 객체에 대한 annotation 파일이 필요하다. Annotation 파일에는 이미지에서 물체에 대한 BBox(bounding box) 정보와 라이다로 취득한 물체의 3차원 거리 값, 크기 값, 회전 값 그리고 카메라 캘리브레이션 값이 요구된다. 그러나 UAV로 취득한 영상에서는 라이다 데이터가 부족하여 확보하기 어렵고 레이블링을 포함한 데이터 정리 및 준비에 많은 시간이 소요된다. 이러한 문제점을 해결하기 위해 Unity3D 게임엔진을 이용하여 실제 갯송이 및 갯나무 숲과 흡사하게 3D 모델링 하였다. 가상에서 생성한 갯송이 및 갯나무 숲 환경은 그림 5와 같다.

생성한 가상공간의 정보를 이용하여 카메라 좌표계에서의 거리, 크기, 폭은 월드 좌표계(world coordinate system), 카메라 좌표계(camera coordinate system), 픽셀 좌표계(pixel image coordinate system), 정규 좌표계(normalized image coordinate system)의 3D 변환법을 통해 취득할 수 있다[16]. 월드 좌표계는 실제 공간에서의 사물의 위치를 표현할 때 기준으로 삼는 좌표계이다. 좌표의 단위는 미터(meter)로 측정했으며  $(X, Y, Z)$ 로 표기한다.

카메라 좌표계는 그림 6과 같이 카메라를 기준으로 한 좌표계이다. 카메라의 초점을 원점, 카메라의 정면 광축 방향을  $Z$ 축, 카메라 아래 방향을  $Y$ 축, 오른쪽 방향을  $X$ 축으로 잡는다. 카메라 좌표계 단위 또한 미터로 측정했으며  $X_c, Y_c, Z_c$ 로 표기한다. 픽셀 좌표계는 실제 눈으로 보는 영상에 대한 좌표계로 이미지의 왼쪽 상단 모서리를 원점, 오른쪽 방향을  $X$ 축 증가방향, 아래쪽 방향을  $Y$ 축 증가 방향으로 한다. 좌표계의 단위는 픽셀(pixel)이며  $x, y$ 로 표기한다.

정규 좌표계는 카메라의 내부 파라미터(intrinsic parameter)의 영향을 제거한 이미지 좌표계로 카메라 초점과의 거리가 1인 가상의 이미지 평면을 정의하는 좌표계이다. 카메라 캘리브레이션은 사진 좌표와 이에 대응하는 일직선상의 실세계 좌표와의 관계를 통해 내부표정 요소(IOP: interior orientation parameters)를 결정하는 것이다.



그림 5. 갯송이 및 갯나무 숲 가상데이터 생성 예시  
Fig. 5. Examples of creating virtual pine nut and pine forest.

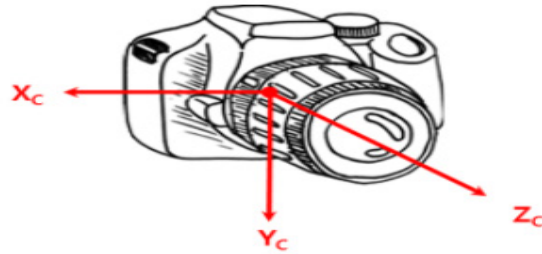


그림 6. 카메라 좌표계  
Fig. 6. Camera coordinate system.

내부표정 요소는 카메라의 초점거리, 사진의 주점, 렌즈 왜곡 계수 등으로 이루어져 있으며, 이를 통해 사진 내의 객체에 대한 정확한 3차원 좌표를 결정할 수 있다. 카메라 캘리브레이션을 위해 Unity 3D에서 존재하는 가상카메라를 이용했으며 변환 관계는 다음 식 (1)과 같다.

$$s \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & skew\_cf_x & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (1)$$

식의 첫 번째 행렬은 카메라 내부파라미터로  $f_x, f_y$ 는 초점거리,  $skew\_cf_x$ 는 비대칭 계수,  $c_x, c_y$ 는 주점이다. 두 번째 행렬은 카메라 외부 파라미터(extrinsic parameters)로 카메라 좌표계와 월드 좌표계 사이의 변환 관계를 설명하는 파라미터이다. 두 좌표계 사이의 회전 및 평행이동 변환으로 표현되며 가상카메라 위치와 방향에 따라 변환된다.

카메라 깊이 정보 추출에 대한 3차원 공간상의 한점에 대해 카메라 좌표와 월드 좌표 사이의 변환 관계는 식 2와 같다.

$$\begin{aligned} P_c &= RP_w + t \\ P_w &= R^t(P_c - t) \end{aligned} \quad (2)$$

이때 3차원 공간의 한 점  $P$ 에 대한 카메라 좌표는  $P_c$ , 월드 좌표는  $P_w$ 로  $R$ 은 식 (2)에서 카메라 외부파라미터이다.

카메라 상대 위치에 대한 깊이 및 크기 취득을 위해서는 지면 좌표가 필요하며 카메라의 내부파라미터에 대한 영향을 제거하기 위해 픽셀 좌표를 식 (3)과 같이 정규좌표로 변환한다.

$$\begin{aligned} u &= (x - c_x) / f_x \\ v &= (y - c_y) / f_y \end{aligned} \quad (3)$$

식 (3)을 이용해 취득한 정규좌표는 3차원 카메라 좌표로 해석하면  $(u, v, 1)$ 이 되며 식 (2)를 통해 취득한  $P_c$ 와 동일하다. 최종적으로 식 (2)와 식 (3)을 통해 취득한  $P_c$ 와  $P_w$ 를 통해 카메라의 상대적 크기 및 폭을 취득할 수 있다.

이처럼 가상데이터를 생성해서 모델 학습 시에 필요한 실제

값(ground truth) 정보를 라이다/레이더 센서 없이 얻을 수 있으며, 다양한 시점과 환경요인을 고려하여 필요한 만큼의 학습데이터를 자동으로 생성할 수 있다.

### III. 딥러닝 기반 잣송이 검출 기법

#### 3-1 잣송이 분포 영역 검출

잣 수확의 동선 및 효율적 수확 일정을 위해 잣나무 숲에 대한 지도 및 지도상에 잣송이의 분포를 표시하기 위해 의미론적 분할 기법을 시행한다. 의미론적 분할(semantic segmentation)은 객체 검출 기법의 하나이며 이미지 내에 있는 물체들을 같은 영역으로 분할하여 표시하며, 한 영상 내에 다수의 특징을 구분하는 것이 가능하다. 특정 분류는 픽셀 단위로 이루어지며, 영상 내 각 픽셀이 분류된 특징 정보를 가진다. 잣 분포 영역 검출을 위하여 여러 의미론적 분할 알고리즘 중 높은 정확도를 나타내고 있는 DeepLabv3 [13] 알고리즘을 채택하였다.

DeepLabv3는 아래 그림 7과 같이 인코더와 디코더 형태의 구조로 되어 있다. 인코더 부분에서는 Xception을 인코더로 채택하고 있으며 컨벌루션을 진행할 때 다른 네트워크와는 다른 atrous convolution을 사용하여 다양한 크기의 특징들을 추출한다. 디코더는 디코더 과정의 upsampling 네트워크 출력값과 인코더 전체 출력값을 concat하여 진행한다. 인코더와 디코더 구조로 객체의 경계를 뚜렷하게 담을 수 있으며 backbone은 ResNet으로 기본적인 입력 영상의 특징 추출기로 사용한다.

#### 3-2 잣송이 2D 검출

객체 검출(object detection)은 특정 객체가 어느 위치에 존재하는지 알아내는 탐색(localization) 문제와 그 객체가 무슨 클래스인지 알아내는 인식(recognition)을 동시에 추론하는 방법이다.

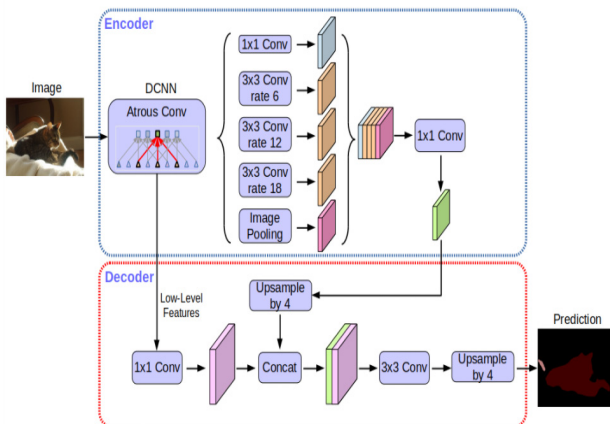


그림 7. DeepLabv3 알고리즘 네트워크 구조도  
Fig. 7. Network architecture of DeepLabv3 algorithm.

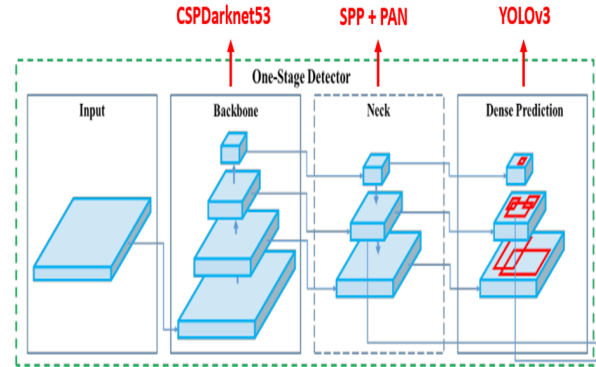


그림 8. YOLOv4 알고리즘 네트워크 구조도  
Fig. 8. Network architecture of YOLOv4 algorithm.

딥러닝 기반의 객체 검출 방법은 크게 1-stage, 2-stage로 2가지 방향으로 나눌 수 있다. 탐색영역(region proposal)과 해당 영역을 분류(detection)하는 두 과정이 순차적으로 수행되는 방법은 2-stage이며, 두 가지 과정을 한 번에 처리하는 방법은 1-stage로 불린다. 2-stage는 1-stage보다 비교적 느리지만 정확하며 반대로 1-stage는 2-stage에 비해 빠르지만 부정확하다는 특징을 가지고 있다. 본 논문에서는 UAV 영상을 이용한 잣 수확을 위한 것으로 실시간 소형 객체 검출을 요구한다. 이에 1-stage object detector인 기존의 YOLO(you only look at once)의 성능을 개선한 YOLOv4 [14]를 잣송이 검출을 위하여 채택하였다.

YOLOv4 알고리즘은 그림 8과 같이 기존의 YOLOv3 버전을 일부 수정하여 작은 객체도 검출할 수 있게 배포한 모델이다. YOLOv4는 Backbone, Neck, Head 세 부분으로 구성된다. Backbone은 특징지도 추출을 위해 활용되는 네트워크이며, Neck은 Backbone의 다른 단계에서 다른 특징지도들을 추출하기 위해 Backbone과 Head 사이에 존재하는 계층이다. Head는 실제로 객체 검출을 담당하는 부분이다. YOLOv4의 Backbone으로는 CSPNet기반의 CSPDarkNet53을 사용함으로써 학습할 때 중복으로 사용되는 기울기 정보를 제거하며 연산량을 줄이면서 성능을 높이는 특징을 가진다. Neck은 SPP(spatial pyramid pooling)와 PAN(path aggregation network)을 적용함으로써 CNN에서 fixed-size image를 요구하는 것을 제거할 수 있는 장점이 있다. 즉, 입력 영상의 크기에 상관없이 고정된 크기의 output을 생성할 수 있다. 마지막 Head는 기존의 YOLOv3를 사용함으로써 경계상자의 객체 점수와 클래스를 분류한다.

#### 3-3 잣송이 3D 검출

본 논문에서는 잣송이 3D 검출을 위한 알고리즘으로 단안 카메라에서의 빠른 속도와 높은 정확도를 수행하는 CenterNet [15]을 잣송이 3D 검출 알고리즘으로 채택하였다. CenterNet의 알고리즘 구조도는 그림 9와 같다. CenterNet 알고리즘의 backbone으로는 DLA34를 사용하며 다양한 스케일의 객체를

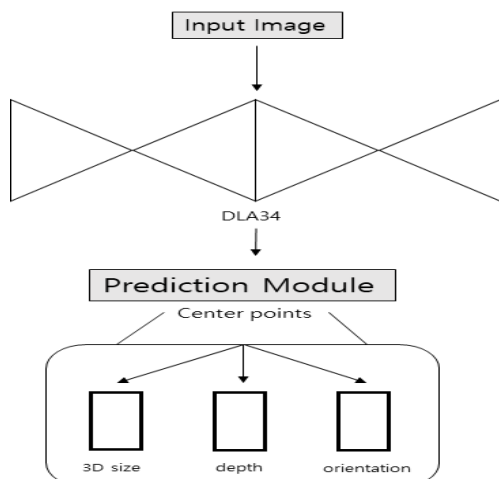


그림 9. CenterNet 알고리즘 네트워크 구조도  
Fig. 9. Network architecture of CenterNet algorithm.

수용하는 장점을 가지며, 객체마다 단 하나의 keypoint인 중심점(center point)을 출력으로 가진다. 객체들을 모두 하나의 점으로 표현하기 때문에 한 개의 anchor box를 사용함으로써 기존의 1-stage object detector 모델이 사용하는 NMS(non-maximum-suppression)의 후처리 과정이 필요 없어 빠른 속도를 특징으로 가진다. Backbone에서 출력으로 나온 객체 중앙에 놓인 point의 특징값을 기반으로 객체의 3D 크기, 깊이 그리고 방향을 추정하여 3D BBox를 예측한다.

## IV. 실험

### 4-1 학습환경

본 논문의 실험에 사용된 컴퓨터의 하드웨어 규격은 프로세서 Intel Core i9-9900K, 그래픽 카드 RTX 2080-TI, RAM 32GB이다. 딥러닝 학습을 위한 프레임워크로 Pytorch를 이용했으며 잣송이 3D 검출을 위한 3D 모델링 및 가상공간 생성은 Unity3D 2019.3.6. f1 엔진을 이용했다.

### 4-2 항공 영상 기반 검출 결과

드론을 이용하여 수집된 잣나무 숲 영상으로부터 잣송이 분포 영역 및 2D 검출 모델의 학습과 성능 평가를 수행했다. 수집한 영상을 기반으로 1초에 7장씩 frame crop을 하였으며 잣송이 구분이 불가능한 영상과 중복되는 영상은 제외했다. 추출한 원본 데이터의 크기는 총 500장으로 2장에서 언급한 데이터 증강 방법을 통해 총 25,000장의 데이터를 대상으로 실험을 수행하였으며, 학습데이터 20,000장, 테스트 데이터 5,000장으로 구성했다. 딥러닝 기반의 객체 의미론적 분할 및 검출을 위해 DeepLabv3와 YOLOv4 알고리즘을 사용하였으며, 모델의 입력

영상은 608×608픽셀로 정규화하였다.

본 논문에서 사용한 잣송이 분포 영역 검출 성능평가 방법은 의미론적 분할 기법에서 가장 많이 사용되는 방식인 mIoU(mean intersection over union)를 측정하여 성능을 평가하였다. 식(4)는 클래스별 IoU를 계산하는 방법이다.

$$IoU = \frac{B_{gt} \cap B_p}{B_{gt} \cup B_p} \quad (4)$$

이때,  $B_{gt}$ 는 정답 영상에서 해당 클래스의 영역을 의미하며,  $B_p$ 는 예측한 결과에서 해당 클래스의 영역을 의미한다. 제안한 모델에 의한 검출 결과는 그림 10에 보이며, 표 2는 분포 영역에 대한 검출정확도를 나타낸다. 잣송이 분포 영역 검출 정확도는 82.15%로 기존의 PASCAL VOC 2012 데이터셋 [17]에서의 DeepLabv3 [13]의 검출정확도 85.7%보다 3.55% 낮게 측정되었다. 이는, UAV 항공 영상에서 촬영한 잣송이 1개는 4K 영상에서 0.03%의 비율을 차지하기 때문에 상대적으로 낮은 정확도로 검출하는 것으로 분석된다.

잣송이 2D 검출 결과는 그림 11에 보이며 표 3은 검출 결과에 대한 오차 행렬이다. 표 3에서 행은 테스트 데이터에 대한 ground truth를 의미하며, 열은 모델에 의해 예측된 분류 클래스를 의미한다. 검출 결과, 전체 잣송이 객체 243,972개 중, 오분류를 포함하여 검출된 객체의 수는 208,693개(85.53%)인 것으로 나타났다. 실험 영상 분석 결과, 대부분의 잣송이는 검출하지만, 잣송이가 모여있는 부분들은 검출 자체를 못하는 경우가 발생하는 것을 볼 수 있다. 이는 나뭇잎과 잣송이 사이의 경계가 모호한 일부 영상에서 발생하는 것으로 분석되었다. 잣송이 2D 검출에 대한 성능평가는 정확도(precision)와 재현율(recall)로 수행하였다. 정확도는 시스템에 의해 검출된 객체 중 정확하게 인식된 객체의 비율로 식 5와 같다.

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{All\ Detections} \quad (5)$$

재현율은 전체 객체 중 모델에 의해 정확하게 인식된 객체의 비율로 식 6과 같다.

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{All\ Ground\ Truths} \quad (6)$$

성능평가 결과는 표 5와 같다. 정확도 평가 결과, 잣송이 객체에 대한 정확도와 재현율은 각각 86.93%, 78.04%로 상대적으로 재현율이 낮은 것은 앞서 언급한 것과 같이 영상에서 나뭇잎과 잣송이 사이의 경계가 모호한 일부 영상에서 분류된 것에서 야기한 결과로 분석된다.



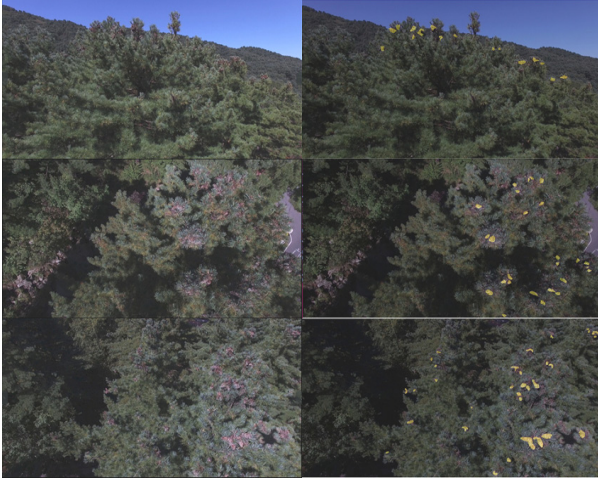


그림 10. 잣송이 분포 영역 결과  
Fig. 10. Pine nut segmentation result.

표 2. 잣송이 분포 영역 검출 정확도  
Table 2. Pine nut segmentation accuracy.

Segmentation(mIoU)	82.15%
--------------------	--------

표 3. 잣송이 2D 검출 오차 행렬  
Table 3. Pine nut 2D detection confusion matrix.

Classes	Pine	none
Pine	181,427	62,545
none	27,266	—

표 4. 잣송이 2D 검출 정확도 결과  
Table 4. Pine nut 2D detection.

Classes	No.object	Detected	Non-Detected	Precision	Recall
Pine	267,412	181,427	27,266	86.93%	78.04%



그림 11. 잣송이 2D 검출 결과  
Fig. 11. Pine nut 2D detection result.

#### 4-3 가상 데이터 기반 잣송이 검출 결과

Unity3D를 이용하여 만든 가상 잣나무 숲 영상으로부터 제안한 3D 검출 모델 학습과 성능평가를 진행하였다. 총 11,200장의 가상데이터를 만들었으며 학습은 7,500장, 테스트는 3,700장으로 구성하였다. 모델 학습 후, 본 논문에서 사용한 잣송이 3D 검출은 3D 객체 검출에서 가장 많이 사용되는 방식인 KITTI Evaluation Kit을 이용했으며 3D BBox의 IoU 임계값은 0.5, Mode는 Easy를 기준으로 성능을 평가하였다. 그림 12는 잣송이 3D 검출 결과이며 표 5는 3D 검출정확도이다. 가상데이터에서의 잣송이 3D 검출률은 59.45%로 표 6에서의 단안 카메라 기반 KITTI Benchmark [18]의 검출률보다 30% 이상 좋은 성능을 보인다. 이러한 결과는 그림 12와 같이 잣송이가 나뭇잎에 가려지지 않은 테스트 셋이 대부분으로 이루어져 있어, 정확도에 영향을 미치는 것으로 보인다.

표 5. 잣송이 3D 검출정확도  
Table 5. Pine nut 3D detection accuracy.

3D detection AP(IoU@0.5)	59.45%
--------------------------	--------



그림 12. 가상공간에서의 잣송이 3D 검출 결과  
Fig. 12. Pine nut 3D detection result in virtual space.

표 6. KITTI Benchmark 사이트에서의 KITTI 3D 테스트 셋에 대한 AP 점수

Table 6. AP scores on KITTI 3D test set from KITTI Benchmark website.

Method	Easy	Moderate	Hard
StereoFENet	29.14%	18.41%	14.20%
Kinematic3D	19.07%	12.72%	9.17%
D4LCN	16.65%	11.72%	9.51%
AM3D	16.50%	10.74%	9.52%
RTM3D	14.41%	10.34%	8.77%
MonoPair	13.04%	9.99%	8.65%
SMOKE	14.03%	9.76%	7.84%
M3D-RPN	14.76%	9.71%	7.42%

## V. 결 론

본 논문에서는 UAV를 이용한 잣 수확을 위해 항공 영상에서 딥러닝 기반 잣송이 검출 기법을 제안했다. 본 연구에서 제안한 방법을 검증하기 위하여 UAV를 통해 잣나무 숲 동영상을 취득하였으며 3D 검출은 Unity3D를 이용한 가상공간에서 잣송이 검출을 수행하였다. 학습 모델로는 딥러닝 기반의 검출 기법인 DeepLabv3, YOLOv4, CenterNet을 이용했으며 분포 영역 검출률은 82.15%, 2D 검출률은 86.93%, 3D 검출률은 59.45%의 정확도를 달성하였다. 본 연구를 통해 딥러닝 객체 검출 알고리즘을 이용하여 향후 드론을 이용한 잣송이 수확의 가능성을 확인하였다. 추후 연구로는 멀리 있는 잣송이에 대한 검출률과 정확도를 높이기 위한 딥러닝 기반의 객체 검출 알고리즘 성능 고도화를 진행하고, 추가로 의미론적 분할 알고리즘을 통한 매핑 기술에 대한 연구와 로봇 팔을 이용한 잣송이 수확 기술에 관한 연구를 진행할 예정이다.

## Acknowledgments

본 연구는 경기도 지역협력 연구센터 사업의 일환으로 수행하였음.[GRRC항공2017-B04, 지능형 인터랙티브 미디어 및 공간 융합 응용 서비스 개발].

## References

- [1] W. J. Kim and J. H. Kang, "ToA-based localization algorithm for mitigating positioning error in NLoS channel," *Journal of Institute of Control, Robotics and Systems*, Vol. 24, No. 11, pp. 1043-1047, Oct. 2010.
- [2] C. Ju and H. I. Son, "Design of supervisors for the control of agricultural multiple UAV systems: A preliminary result," *Journal of Institute of Control, Robotics and Systems*, Vol. 24, No. 11, pp. 1005-1013, 2018.
- [3] U. R. Mogili and B. B. V. L. Deepark, "Review on application of drone systems in precision agriculture," *Procedia Computer Science*, Vol. 133, pp. 502-509, July. 2018.
- [4] Liu, W. Jia, C. Ruan, D. Zhao, Y. Gu, and W. Chen, "The recognition of apple fruits in plastic bags based on block classification," *Precision Agriculture*, Vol. 19, No. 4, pp. 735-749, Aug. 2018.
- [5] D. Choi, W. S. Lee, J. K. Schueller, R. Ehsani, F. Roka, and J. Diamond, "A performance comparison of RGB, NIR, and depth images in immature citrus detection using deep learning algorithms for yield prediction," in *ASABE Annual*

- International Meeting*, Kentucky: KY, pp. 1, 2017.
- [6] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Computers and Electronics in Agriculture*, Vol. 157, pp. 417-426, Jan. 2019.
- [7] Y. Yu, K. Zhang, L. Yang, and D. Zhang, "Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN," *Computers and Electronics in Agriculture*, Vol. 163, Jun. 2019.
- [8] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, and R. Urtasun, "Monocular 3d object detection for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas: LAS, pp. 2147-2156, 2016.
- [9] Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, "Deep ordinal regression network for monocular depth estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City: SLC, pp. 2002-2011, 2018.
- [10] Ku, A. D. Pon and S. L. Waslander, "Monocular 3d object detection leveraging accurate proposals and shape reconstruction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, California: CA, pp. 11867-11876, 2019.
- [11] Z. Yang, Y. Sun, S. Liu, X. Shen, and J. Jia, "Std: sparse-to-dense 3d object detector for point cloud," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, California: CA, pp. 1951-1960, 2019.
- [12] Y. You, Y. Wang, W. L. Chao, D. Garg, G. Pleiss, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-lidar++: accurate depth for 3d object detection in autonomous driving," in *International Conference on Learning Representations*, Addis Ababa: ADD, pp. 1-22, 2020.
- [13] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv preprint arXiv:1706.05587, 2017.
- [14] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
- [15] X. Zhou, D. Wang and P. Krähenbühl, "Objects as points," arXiv preprint arXiv:1904.07850, 2019.
- [16] Y. S. Kim, B. R. Min, W. Kim, D. W. Kim, K. W. Seo, H. T. Kim and D. W. Lee, "Development of image processing system for tracking location of moving objects," *Proceedings of the Korean Society for Agricultural Machinery Conference*, pp. 292-297, 2002.
- [17] S. Shetty, "Application of convolutional neural network for



image classification on Pascal VOC challenge 2012 dataset,” arXiv preprint arXiv:1607.03785, 2016.

[18] Karlsruhe Institute of Technology. The KITTI Vision

Benchmark Suite [Internet]. Available:

[http://www.cvlibs.net/datasets/kitti/eval\\_object.php?obj\\_benchmark=3d](http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=3d).



**김 규 민 (Gyu-Min Kim)**

2013년 2월 : 한국항공대학교 항공전자정보공학부 (공학사)

2020년 3월 ~ 현재 : 한국항공대학교 항공전자정보공학 석사과정

※ 관심분야 : 영상처리, 컴퓨터 비전, 가상현실



**박 성 준 (Sung-Jun Park)**

2019년 2월 : 한국항공대학교 전자및항공전자공학 (공학사)

2019년 ~ 현재 : 한국항공대학교 항공전자정보공학 석사과정

※ 관심분야 : 컴퓨터비전, 영상처리



**황 승 준 (Seung-Jun Hwang)**

2012년 2월 : 한국항공대학교 정보통신공학 (공학사)

2014년 2월 : 한국항공대학교 일반대학원 정보통신공학 (공학석사)

2014년 3월 ~ 현재 : 한국항공대학교 일반대학원 항공전자정보공학 박사과정

※ 관심분야 : 영상처리, 패턴인식



**김 희 영 (Hee Yeong Kim)**

1986년 2월 : 고려대학교 경영학 학사

1988년 2월 : 한국외국어대학교 경영정보학 석사

2014년 8월 : 한국외국어대학교 경영학 박사

2020년 7월 ~ 현재 : 한국정보시스템감사통제협회(ISACA Korea) 회장

2016년 1월 ~ 현재 : (주) 링크투 대표이사

※ 관심분야 : 빅데이터, 인공지능, 스마트농업, SW신뢰안전성



**백 중 환 (Joong-Hwan Baek)**

1981년 2월 : 한국항공대학교 항공통신공학 (공학사)

1987년 7월 : 오글라호마주립대학원 전기 및 컴퓨터공학 (공학석사)

1991년 7월 : 오글라호마주립대학원 전기 및 컴퓨터공학 (공학박사)

1992년 ~ 현재 : 한국항공대학교 항공전자정보공학부 교수

※ 관심분야 : 영상처리, 패턴인식, 멀티미디어, 가상현실