

# 딥러닝을 이용한 온실의 환경에서 토마토 화방 및 과실 객체 검출 모델 연구

서다솜\*, 김경철<sup>o</sup>, 이명훈\*, 권경도\*\*, 김국환\*\*

## Research on Tomato Flowers and Fruits Object Detection Model in Greenhouse Environment Using Deep Learning

Dasom Seo\*, Kyoung-Chul Kim<sup>o</sup>, Meonghun Lee\*, Kyung-Do Kwon\*\*, Gookhwan Kim\*\*

### 요약

농업에서 작물의 품질은 매우 중요하다. 스마트팜은 정보통신기술을 활용해 재배부터 수확까지 모든 과정에서 최적의 환경을 유지하고 작물의 품질을 향상시키며 자동화를 통해 인력 낭비를 줄이는데 유용한 농업 시스템이다. 본 연구에서는 토마토의 수확물의 품질에 영향을 주는 생육지표 중 하나인 화방, 과실의 영상 데이터를 수집하여 데이터 셋을 구축한다. 구축된 데이터 셋을 기반으로 객체 검출 모델을 활용한 토마토 데이터 셋 평가와 최적의 모델을 찾는 실험을 한다. 사용된 모델은 SSD, Yolo V3, V4를 이용하였으며 Yolo V4의 mAP@0.50가 92.82%로 가장 높은 결과를 나타냈다. 앞으로 이를 활용하여 토마토의 화방 위치, 개수를 찾고 과실의 위치, 성숙도 등을 예측하는 시스템 구축이 가능해질 것이다.

**키워드** : 토마토, 이미지 처리, 객체 검출, 딥러닝 기계학습

**Key Words** : Tomato, Image Processing, Object Detection, Deep Learning, Machine Vision

### ABSTRACT

In agriculture, the quality of the crop is very important. Smart farm is an agricultural system that is useful for maintaining the optimal environment in all processes from cultivation to harvesting, improving the quality of crops, and reducing manpower waste through automation using information and communication technology. A data set was constructed by collecting image data of flowers and fruits, which are one of the growth indicators that affect the quality of tomato crops. Find a model. A data set is constructed by collecting image data of flowers and fruits, which are one of the growth indicators that affect the quality of tomato crops. Based on the constructed data set, we conduct an experiment to evaluate the tomato data set using an object detection model and find the optimal model. The mAP@0.50 of those models(SSD, Yolo V3 and V4. Yolo V4) showed the highest result with 92.82%. In the future, it will be possible to use this to find the location and number of tomato' flower, and to build a system that predicts the location and maturity of tomatoes fruit.

※ 본 연구는 농림축산식품부 및 과학기술정보통신부, 농촌진흥청의 재원으로 농림식품기술기획평가원과 재단법인 스마트팜연구개발사업단의 스마트팜다부처패키지혁신기술개발사업의 지원을 받아 연구되었음(420131-04)

• First Author : Division of Smart farm Development, National Institute of Agricultural Sciences, ssomda21@korea.kr, 정희원

<sup>o</sup> Corresponding Author : Division of Smart farm Development, National Institute of Agricultural Sciences, kkemole@korea.kr, 정희원

\* Major of Smart Agricultural, Sunchon National University, leemh5544@gmail.com, 정희원

\*\* Division of Smart farm Development, National Institute of Agricultural Sciences, kwonkd89@korea.kr; meceng93@korea.kr, 정희원

논문번호 : 202109-231-0-SE, Received August 19, 2021; Revised September 28, 2021; Accepted October 25, 2021

## I. 서 론

최근 스마트 팜에 관한 연구가 급속도로 이루어지고 있으며<sup>[1-3]</sup> 그중 온실에서 취득된 빅데이터를 기반으로 인공지능을 활용한 연구가 주목받고 있다. 이러한 빅데이터는 주로 환경 및 제어 데이터가 주를 이룬다. 하지만 이러한 수치화된 빅데이터는 각각 작물에 대한 정보가 아니라 군집 단위로 취득되었기 때문에 한 작물의 생산량 및 품질 예측에는 한계가 있다. 따라서 본 연구에서는 환경 및 제어 데이터를 제외하고 토마토 작물에 대한 영상을 가지고 인공지능 모델을 활용하여 토마토의 화방, 과실 위치 및 개수 검출을 위한 모델을 찾는다.

인공지능 모델에 활용할 데이터인 화방과 과실의 데이터는 따로 촬영 장비를 준비해서 수집한 것이 아니라 개인이 가지고 있는 스마트폰, 태블릿을 활용하여 데이터를 수집한다. 수집된 데이터는 RGB 영상으로 취득되며 토마토의 생육 지표인 화방과 과실은 따로 Bounding Box를 만들어 주석 파일을 만든다. 활용된 인공지능은 최근 딥러닝 기술의 발전으로 비약적으로 성능이 향상된 객체 검출(Object Detection) 기술이며 이는 영상에서 사람, 물체 등의 세부적인 영상 정보를 추출하는 기술이다. 이러한 객체 검출 기술 중 두 단계로 이루어진 R-CNN 모델<sup>[4]</sup> 계열이 아니라 한 단계로 이루어진 Yolo(You only look once)<sup>[5]</sup>, SSD(Single Shot Detector)<sup>[6]</sup> 모델을 활용한다.

본 연구의 구성은 다음과 같다. 제2장에서는 관련 연구에 대해 살펴보고, 제3장에서는 토마토 생육지표 데이터 및 인공지능 모델 실험 결과에 대해 알아본다. 마지막으로 제4장에서는 결론 및 향후 연구계획을 기술한다.

## II. 관련 연구

객체 검출이란 영상 분류(Image Classification)보다 한 단계 더 나아가 영상 내의 사물의 위치를 예측하는 회귀(Regression)가 추가된 문제이다. 단순히 한 장의 영상에 대하여 예측하는 것이 아니라 한 영상 안의 한 개 혹은 여러 개의 물체를 찾아 인식하고 해당 물체의 Bounding Box를 찾는다. 즉, 객체 검출은 물체의 위치를 찾는 Localization 문제와 물체를 식별하는 Classification 문제를 합한 문제이며, 크게 1-stage Detector와 2-stage Detector로 나뉜다. 1-stage Detector는 위의 두 문제를 동시에 병행하는 방법이고, 2-stage Detector는 객체 후보 영역(ROI: Region

of Interest)을 찾는 단계와 후보들의 클래스 분류 및 Bounding Box Regression 작업을 하는 단계로 Localization, Classification 문제를 차례로 해결하는 방법이다.<sup>[4,7]</sup> 1-stage Detector는 비교적 속도는 빠르나 정확도는 낮으며 반대로 2-stage Detector는 속도는 느리나 정확도는 비교적 높은 편이다. 본 장에서는 두 가지 방법 중 1-stage Detector에 속하는 모델들인 SSD<sup>[6]</sup>와 Yolo V3<sup>[7]</sup>, V4<sup>[8]</sup>로 총 3가지 모델에 대해 살펴본다.

### 2.1 Single Shot Detector(SSD)

SSD는 1-stage Detector에 속하는 모델 중 하나로 VGG<sup>[9]</sup> 등 특징 추출 레이어(Feature extraction layer)의 뒤에 연결되어 물체의 위치를 찾는 Localization 문제와 물체를 식별하는 Classification 문제를 해결한다<sup>[6]</sup>. SSD의 핵심은 자주 발견되는 객체 모양과 비슷한 default box를 여러 feature map에 적용해 score와 box 좌표를 예측하는 것이다. 또한, 객체 검출의 정확도를 높이고 다양한 크기의 물체를 찾기 위해 서로 다른 크기의 feature map로부터 여러 크기의 예측을 수행하고 비율 또한 다르게 적용한다. 또한, end-to-end 학습이 가능하며 저해상도 영상에서도 높은 정확도를 가진다.

### 2.2 Yolo V3, V4

Yolo도 SSD처럼 1-stage Detector에 속하는 모델 중 하나이다<sup>[5,6]</sup>. 하지만 SSD보다 작은 개수의 박스만 사용하고 구조 역시 단순하므로 계산량이 상대적으로 적다. 또한 Darknet이라는 공식 코드가 구현되어있어 쉽게 접근하여 사용할 수 있다<sup>[5]</sup>.

Yolo는 처음 v1<sup>[5]</sup>부터 최근 v5<sup>[10]</sup>까지 지속적인 연구가 진행되어왔다. v2<sup>[11]</sup>에서는 Darknet19, anchor box 등 여러 방법을 사용하여 모델의 성능을 올렸으며, v2와 함께 발표한 YOLO9000은 검출할 수 있는 클래스가 9,000개인 모델을 선보였다. 그 이후, v3에서는 Bounding Box를 찾는 문제에서 파라미터들을 개선하여 성능을 높이고, SSD와 비슷한 방법으로 다양한 크기의 특징 맵을 최종 레이어에 연결하고, 특징을 추출하는 레이어에서 Residual Connection을 추가하여 성능을 더욱 끌어올렸다. 다음으로 v4는 Backbone, Neck, Head로 구성되어있는데 간단히 말해 Backbone은 입력된 영상을 받아 특징 맵으로 변형시켜주는 부분이며 ImageNet<sup>[12]</sup>으로 pre-trained 시킨 VGG16<sup>[9]</sup>, ResNet-50<sup>[13]</sup> 등이 대표적이다. Neck은 Backbone과 Head를 연결하는 부분으로 특징 맵을 정

제(refinement), 재구성(reconfiguration)하며 FPN(Feature Pyramid Network)<sup>[14]</sup>, PAN(Path Aggregation Network)<sup>[15]</sup> 등이 존재한다. Head는 네트워크 모델이 1-stage, 2-stage Detector를 결정짓는 Dense Prediction과 Sparse Prediction으로 나뉜다. V4에서는 Backbone에서는 CSP-Darknet53<sup>[16]</sup>, Neck은 SPP(Spatial Pyramid Pooling)<sup>[17]</sup>와 PAN을 사용했으며, Head에서는 YoloV3 모델을 사용했다. v4에서는 CutMix<sup>[18]</sup>, Mosaic Data Augmentation<sup>[8]</sup> 등 최신 딥러닝 기법들을 적용해 성능을 향상했다. 마지막으로 Yolo는 기본 모델뿐만 아니라 모델의 계층을 줄여 계산량을 줄이고 성능을 포기하는 대신 속도가 빠른 경량화 모델인 tinyYolo를 함께 배포했다<sup>[5]</sup>.

### III. 토마토 데이터 및 실험 결과

본 장에서는 토마토의 화방 및 과실 데이터를 구성하는 절차와 여러 객체 검출 모델을 이용하여 실험한 결과를 서술한다.

#### 3.1 토마토 생육지표 데이터

토마토 생육지표를 취득하는 이유는 과실의 크기, 당도 등의 토마토의 품질과 가격을 높이기 위해서 과실과 관련이 있는 데이터들을 취득하여 분석하기 위함이다. 생육지표에는 여러 가지 데이터가 있지만 직접 연관이 높은 화방과 과실 데이터를 기반으로 영상을 구축했다. 취득된 데이터는 그림 1과 같다.

객체 검출 모델을 학습시키기 위해서는 영상과 함께 Ground Truth를 입력으로 주어 학습시킨다. Ground Truth는 영상 속의 찾고자 하는 물체의 박스 좌표를 (x\_min, y\_min, x\_max, y\_max)로 나타낸 값들이다. 따라서 학습에 주어지는 입력 값은 영상파일과 주석 처리된 파일을 쌍으로 주어야 한다. 주석파일은 입력 영상의 경로, 좌표 값, 클래스로 이루어져 있다.

토마토 데이터셋은 화방 영상 6,408장과 과실 영상

19,692장으로 토마토 영상은 총 26,100장을 취득했다. 또한, 주석은 각각 19,178개와 57,921개로 총 77,099개의 Bounding Box를 만들었고 이는 한 장의 영상당 2~3개의 박스를 갖는다. 인공지능 모델의 더 높은 정확도를 위해서 데이터 증강을 했다. 데이터 증강 기법으로는 좌우 대칭, 영상 회전을 섞어 1장의 영상마다 추가적인 7장의 영상을 획득했다. 그림 2는 원본 영상과 데이터 증강이 된 결과이다. 왼쪽 위의 영상이 원본 영상이며 나머지 7장의 사진들은 좌우 반전, 상하 반전, 90도, 270도 회전을 조합해서 증강한 영상이다. 토마토의 영상을 수집하다 보니 좌, 우뿐만 아니라 다른 토마토 과실이나 화방에 밀려 거꾸로 자라는 영상이 종종 발견되기에 상하 반전도 데이터 증강 목록에 추가했다. 표 1은 초기 영상의 수와 주석의 수, 증강된 영상의 수와 주석의 수를 표로 나타낸 것이다. 추가된 영상의 수는 208,800장이며 주석은 616,792개이다.



그림 2. 토마토 과실 영상에 대한 데이터 증강  
Fig. 2. Data augmentation for tomato fruit images



그림 1. 토마토의 화방 및 과실에 대한 영상  
Fig. 1. Tomato Images of flowers and fruits

표 1. 증강 전·후 영상과 주석 수  
Table 1. The number of data set after augmentation

증강 전/후	화방/과실	영상(장)	주석(개)
증강 전 데이터 셋	화방	6,408	19,178
	과실	19,692	57,921
	총합	26,100	77,079
증강 후 데이터 셋	화방	51,264	153,424
	과실	157,536	463,368
	총합	208,800	616,792

3.2 객체 검출 모델 실험 결과

본 장에서는 구축된 토마토 화방, 과실 데이터셋의 활용 가능성 검토 및 가장 성능이 좋은 객체 검출 모델을 분석하고자 실험을 수행했다. 학습에는 Tesla V100을 사용하였다. 여러 객체 검출 모델을 이용하여 많이 사용되는 평가지표인 Mean Average Precision(mAP)와 Recall@K를 사용했다. 마지막으로 데이터 셋은 학습과 테스트셋으로 나누고 구성 비율은 8:2로 구성했다. 우선 실험에 사용된 알고리즘은 Yolo V3<sup>[7]</sup>, Yolo V4<sup>[8]</sup>, SSD<sup>[6]</sup>로 1-stage Detector에 속하는 모델들을 사용했다. 표 2는 객체 검출 모델의 mAP값을 보여준다. mAP@0.50은 SSD가 75.30%로 가장 낮은 수치를 보였으며, 반대로 Yolo V4가 92.82%로 가장 좋은 결과를 나타냈다. 또한, 대체로 화방의 데이터셋 개수보다 과실의 데이터셋 개수가 약 3배가량 많았기 때문에 모든 모델에서 화방보다 과실의 결과가 좋게 나타났다.

그림 3, 4는 각각 Yolo V3와 V4의 loss와 mAP@0.50을 차례로 나타낸 그래프이다. 우선 V3는 loss는 낮고 빠르게 줄어들었으나 mAP는 좋은 성능

표 2. 객체 검출 모델에 따른 정확도  
Table 2. Evaluation result of detecting model

	화방	과실	mAP@0.50
SSD	72.33%	78.27%	75.30%
Yolo V3	76.85%	84.15%	80.50%
Yolo V4	89.45%	96.18%	92.82%

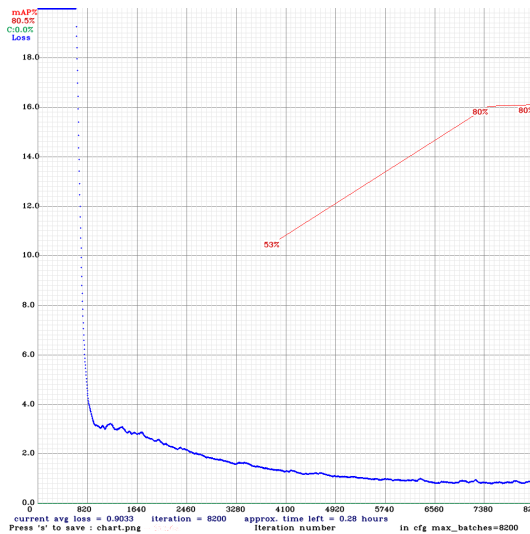


그림 3. Yolo V3의 Loss, mAP@0.50 그래프  
Fig. 3. Yolo V3 Loss and mAP@0.50 graph

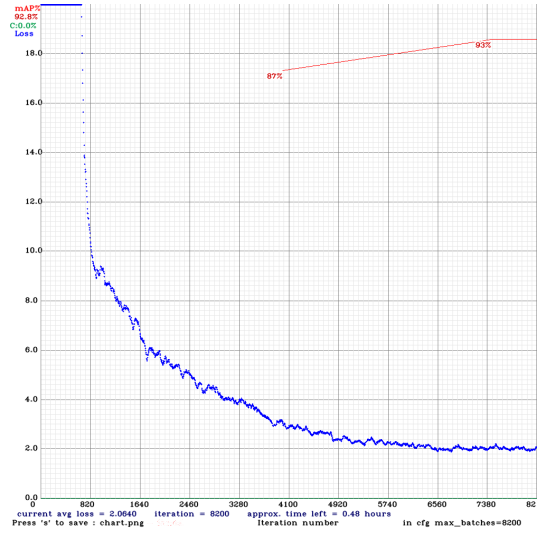


그림 4. Yolo V4의 Loss, mAP@0.50 그래프  
Fig. 4. Yolo V4 Loss and mAP@0.50 graph

을 보여주지 못했다. 반면 V4는 loss는 천천히 줄어들고 2.0보다 낮게 줄어들지 못했지만 87%에서 93%까지 좋은 결과를 나타냈다.

그림 5, 6은 가장 좋은 mAP를 가진 Yolo V4의 모델의 테스트 영상이다. 꽃 같은 경우는 꽃이 작을 때



그림 5. 토마토 화방에 대한 Yolo V4 결과  
Fig. 5. Yolo V4 results for tomato flowers





그림 6. 토마토 과실에 대한 Yolo V4 결과  
 Fig. 6. Yolo V4 results for tomato fruits

부터 완전히 났을 때, 꽃이 졌을 때까지 잘 잡아내는 모습을 볼 수 있었으며 과실은 가려진 경우도 뒤에 과실을 잘 찾아내는 모습을 볼 수 있다.

#### IV. 결 론

본 연구에서는 토마토의 생육지표 중 하나인 화방과 과실 데이터 셋을 구축하고, 구축된 데이터 셋을 기반으로 최적의 인공지능 모델을 찾는다. 기존 26,100장의 토마토 데이터 셋을 데이터 증강을 통하여 8배 증가시킨 208,800장과 616,792개의 주석을 가지는 데이터 셋으로 증강 시켰다. 이러한 데이터 셋은 인공지능 검출 모델에 대해서 합리적인 품질의 데이터 셋으로 구성되며 평균 1장당 3개의 Bounding Box로 적절한 비율을 가진다. 또한, 구민준축된 데이터 셋을 기반으로 인공지능 모델을 훈련한 결과 YoloV4가 mAP@0.50에서 92.8%로 가장 성능이 좋았으며 이를 활용하여 토마토 화방의 위치와 개수를 찾아 과실이 맺힐 화방의 수를 구하고, 과실의 개수를 파악하여 수확량 예측, 더 나아가 성숙도를 예측하고 토마토 자동 수확 로봇 개발 연구로 나아갈 수 있다.

#### References

- [1] M. Lee, J. Hwang, and H. Yoe, "Agricultural production system based on IoT," *2013 IEEE 16th Int. Conf. Comput. Sci. and Eng.*, pp. 833-837, Sydney, NSW, Australia, 2013.
- [2] S. Kim, M. Lee, and C. Shin, "IoT-Based strawberry disease prediction system for smart farming," *Sensors*, vol. 18, no. 11, 2018.
- [3] S. Lee, S. Choi, Y. Shin, "A smart farm real-time control platform using deep learning prediction algorithm," in *Proc. of KICS*, pp. 789-790, Aug, 2020.
- [4] R. Girshick, J. Donahue, T. Darrel, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. CVPR*, pp.580-587, 2014.
- [5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. CVPR*, pp. 779-788, 2016.
- [6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. ECCV 2016*, Springer, Cham, 2016.
- [7] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
- [8] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934 ,2020.
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
- [10] G. Jocher, *YOLOv5*(2020), Retrieved Jan., 20, 2021, from <https://github.com/ultralytics/yolov5/tree/v6.0>
- [11] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proc. IEEE Conf. CVPR*, pp. 7263-7271, 2017.
- [12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc.*

CVPR2009, Miami, Florida, USA, Jun. 2009.

- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. CVPR*, pp. 770-778, 2016.
- [14] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. CVPR*, pp. 2117-2125, 2017.
- [15] S. Liu, et al., "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. CVPR*, 2018.
- [16] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE Conf. CVPR Wkshps.*, 2020.
- [17] K. He, et al., "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 37, no. 9, pp. 1904-1916, 2015.
- [18] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE Int. Conf. Computer Vision*, 2019.

**서 다 숨 (Dasom Seo)**



2019년 2월: 전북대학교 컴퓨터공학전공 (공학석사)  
 2019년 6월~현재: 국립농업과학원 전문연구원  
 <관심분야> 인공지능, 컴퓨터 비전, XAI  
 [ORCID:0000-0002-4654-9846]

**김 경 철 (Kyoung-Chul Kim)**



2015년 8월: 전북대학교 정밀기계공학과 (공학박사)  
 2016년 7월~2019년 1월: 농업기술실용화재단 연구원  
 2019년 2월~현재: 국립농업과학원 연구사  
 <관심분야> 농업 로봇, 농업

인공지능

[ORCID:0000-0001-6699-881x]

**이 명 훈 (Meonghun Lee)**



2011년 2월: 순천대학교 정보통신공학 (공학박사)  
 2010년 4월~2013년 11월: 한국전자통신연구원 선임연구원  
 2017년 2월 2021년 2월: 국립농업과학원 연구사  
 2021년 3월~현재: 순천대학교

스마트농업전공 조교수

<관심분야> 농업 ICT융합, 통신표준, 인공지능

[ORCID:0000-0001-8863-7580]

**권 경 도 (Kyung-Do Kwon)**



2017년 2월: 충남대학교 농업기계공학과 (공학석사)  
 2020년 9월~현재: 농촌진흥청 국립농업과학원 농업연구사

<관심분야> 영상처리, 인공지능

[ORCID:0000-0003-1794-3438]

**김 국 환 (Gookhwan Kim)**



2010년 2월: 경희대학교 기계공학과 (공학박사)  
 2010년 3월~2014년 12월: 국립농업과학원 PostDoc.  
 2015년 1월~현재: 국립농업과학원 연구사

<관심분야> 농업 로봇, 자율항법기술

[ORCID:0000-0002-7278-3476]