Journal of Apiculture

2023.38(3) : 223-232
DOI: 10.17519/apiculture.2023.09.38.3.223

Original research article

학습 데이터셋에 따른 딥러닝 모델의 말벌 검출 정확도 비교

곽희진, 권영재, 이철희*

안동대학교 컴퓨터공학과

Comparison of Vespa Detection Accuracy for Deep Learning Models According to Training Dataset

Heejin Gwak, Yeongjae Kwon and Cheolhee Lee*

Department of Computer Engineering, Andong National University, Andong 36729, Republic of Korea

Abstract Training data configuration is critical for object detection based on supervised deep learning. Namely, the characteristics of training data should be very similar to an actual test environment to raise expected inference accuracy. However, a Vespa object size in pixels in a natural capture environment is not constant and is smaller than the Vespa object size of images in a training set. This study compares the inference accuracy of deep learning models YOLOv5, YOLOX, and YOLOv7 according to training datasets. As training data, the three types of datasets were prepared as follows. First, A basic dataset, composed of five species of Vespa and one bee, is produced for the Vespa training data set. Next, The 0.3% dataset, in which Hornet object size is approximately 0.3% ratio to the whole image size in pixels, is prepared using the basic dataset for tiny object detection. Finally, images are selected from the basic and the 0.3% datasets in the same proportions in a Mixed dataset. After configuring three types of training datasets, the three deep learning models above were trained using the three training datasets, the basic, the 0.3%, and the Mixed dataset, and calculated the training and test mAP. In cases where the training and test data environments are similar, YOLOv7 demonstrated the highest mAP at 95.4%. However, in a test result experiment for actual environments using trained weights by the basic dataset, the mAP@50 scores are 30%, 14%, and 85% for YOLOv5, YOLOv7, and YOLOX, respectively. That is, YOLOX, an Anchor-free model, is overwhelmingly excellent. The organization of the training dataset is essential to match the inputs of the actual detection environment to obtain the best accuracy in object detection, and YOLOv7 is the best model for a tailored training dataset among state of the art models.

Keywords Deep learning, Training dataset, Vespa, YOLO

서 론

국내에 서식하는 여러 종의 말벌로 인해 매년 전국의 양봉장은 심각한 피해를 입는다(Jeong *et al*., 2016). 이러한 말벌에 의한 피해를 줄이기 위해 다양한 대책이 시도되고 있다. 그중 딥러닝 기술을 이용한 객체 탐지(Object detection)에 관련된 연구도 다양하게 시도되고 있다. 객체 탐지 기술을 적용한 사례로 YOLOX를 이용한 실시간 말벌 모니터링 시스템(Jeong *et al*., 2022a), YOLOv5를 이용한 등검은말벌집 탐색 기술 연구(Jeong *et al*., 2022b), YOLOv3를 이용한 등검은말벌 모니터링 시스템 개발(Kim *et al*., 2021) 등이 있다. 객체 탐지란 영상 내에 존재

http://journal.bee.or.kr/

Received 15 August 2023; Revised 22 September 2023; Accepted 22 September 2023
*Corresponding author. E-mail: chlee@anu.ac.kr

하는 객체의 위치 예측 및 분류를 동시에 수행하는 것을 의미한다. 객체 탐지에 사용되는 딥러닝 알고리즘은 객체의 위치와 분류 과정을 동시에 처리하는 1단계 검출 방식과 객체의 위치와 분류 과정이 개별적으로 이루어지는 2단계 검출 방식이 있다. 1단계 검출기로는 SSD (Liu *et al*., 2016), YOLO (Redmon *et al*., 2016), YOLO9000 (Redmon and Farhadi, 2017), YOLOv3 (Redmon and Farhadi, 2018), YOLOv4 (Bochkovskiy *et al*., 2020), YOLOv5, YOLOX (Ge *et al*., 2021), YOLOv7 (Wang *et al*., 2023) 등이 있다. 그리고 2단계 검출기로는 RCNN (Girshick *et al*., 2014), Fast-RCNN (Girshick, 2015), Faster-RCNN (Ren *et al*., 2015), Mask-RCNN (He *et al*., 2017) 등이 있다. 두 가지 방식을 비교할 때 속도면에서는 1단계 방식이 더 유리하고 정확도 면에서는 2단계 방식이 더 유리하나 최근에는 1단계 방식의 정확도 개선이 많이 이루어져 YOLO (You Only Look Once) 기반 모델이 실시간 객체 탐지에 널리 활용되고 있다(Park, 2020).

전술한 딥러닝 모델은 공통적으로 학습 데이터를 기반으로 모델을 학습하고 추론하는 지도 학습(supervised learning) 기반의 알고리즘이다. 지도 학습에 속하는 딥러닝 모델에서는 최적 모델의 선택과 아울러 학습 데이터의 구성이 매우 중요하다. 즉, 추론하고자 하는 대상에 가장 적합한 딥러닝 모델을 선택하고 실제 적용하고자 하는 문제 영역에서 얻은 학습 데이터를 통해 추론하고자 하는 대상의 특성을 학습을 통해 모델링하여 이를 바탕으로 실제 문제 영역을 추론한다는 전제가 들어가 있다. 따라서 객체의 특성을 잘 표현하기 위해서는 적합한 해상도를 갖는 영상이 필요하고 영상 내에서도 객체의 크기가 영상 전체 면적에 비해 충분한 크기 이상으로 존재하여야 한다.

DORI (Detection, Observation, Recognition, Identification)는 감시 카메라의 감지, 관찰, 인식, 식별을 얻기 위해 제공해야 하는 PPM (Pixels Per Meter)의 정보를 정의하는 IEC 단체가 지정한 업계 표준이다. PPM의 제공에 따라 카메라의 성능을 정의하는 벤치마크 및 등급 시스템을 알 수 있다. DORI에 따르면 객체를 탐지하기 위해서는 영상의 세로 크기 중 객체의 높이는 영상의 세로 높이의 10%, 인식을 위해서는 20%가 필요하다(Ozge *et al*., 2019). 그러나 말벌 탐지의 경우 현장에서 카메라로 입력되는 말벌의 크기는 일정하지 않고 유동적이며 상대적으로 영상의 면적 대비 1% 미만의 객체 크기를 갖는 영상 데이터가 많다. 따라서 비교적 큰 객체를 많이 포함하는 학습 데이터로 객체의 면적이 작은 영상을 추론할 경우 추론의 정확도가 낮아지는 현상이 발생한다. 따라서 본 연구에서는 학습 데이터와 딥러닝 모델에 따른 추론 성능을 객관적으로 평가하고 이를 통해 말벌과 같이 작은 객체에 대한 추론 성능을 높이기 위한 학습 데이터의 구성과 효과적인 딥러닝 모델을 제시하고자 한다. 이를 위해, 딥러닝 모델로는 현재 가장 많이 활용되는 YOLOv5, YOLOX, YOLOv7을 선택하였다. 학습 데이터는 일반적 크기의 말벌 학습 데이터인 기본 학습 데이터셋(Basic Training Dataset), 기본 학습 데이터셋으로부터 영상 내 객체인 말벌의 면적 비율이 전체 영상 크기의 0.3%를 차지하도록 재구성한 0.3% 학습 데이터 집합(0.3% Training Dataset) 그리고 기본 학습 데이터 집합, 3% 학습 데이터셋 그리고 0.3% 학습 데이터셋을 동일한 비율로 무작위로 섞은 혼합 학습 데이터셋(Mixed Training Dataset)을 구성하였다. 이러한 3개의 학습 데이터셋에 대해 각 모델별로 학습을 시행한 후, 학습 데이터 3종류와 같은 방식으로 만든 3종류의 테스트 데이터를 이용하여 각 딥러닝 모델의 특성에 따른 성능의 장단점 및 학습 데이터에 따른 추론 성능(test mAP)을 비교함으로써 말벌 탐지에 가장 적합한 딥러닝 모델 및 학습 데이터를 제시하고자 한다.

재료 및 방법

1. 딥러닝 모델

YOLO란 1단계 검출 방식으로 빠른 인식 속도를 보여주는 객체 인식 딥러닝 모델이다. 입력된 이미지를 일정한 크기로 n×n 개의 그리드를 생성한 후, 각 그리드마다 바운딩 박스(bounding box)를 생성하고 해당 객체의 클래스를 예측한다. YOLO는 2015년 발표 당시, 2단계 검출 방식인 RCNN과 비교했을 때 압도적으로 빠른 추론 속도와 유용한 정확도를 보여주었으며 YOLO에 기반한 모델들은 지금까지 실시간 객체 탐지에 넓게 활용되고 있다. 실험에 사용한 YOLO 모델은 최근 가장 많이 활용되는 YOLOv5, YOLOX, YOLOv7이다. Table 1은 각 YOLO 모델의 세부 구조를 나타낸 것이다.

YOLOv5는 Pytorch 기반으로 구현한 YOLO 시리즈 중 하나이며 복합 스케일링 기법(Tan and Le, 2019)을

적용하여 모델의 성능과 연산량에 따라 small, medium, large, extreme 네 종류로 나누었으며, 활성화 함수로 hard swish (Howard *et al*., 2019)를 사용하여 학습 및 추론 속도를 개선하였다. YOLOX는 PAFPN (Lin *et al*., 2017)을 통해 Multi-Scale Feature Map을 얻는다. 낮은 해상도의 Feature Map에서는 큰 객체를 추출하고 높은 해상도의 Feature Map에서는 작은 객체를 추출하는 neck을 갖고 있다. Backbone과 neck은 YOLOv3와 동일한 SPP (He *et al*., 2015) 방식이 적용되어 있지만, head를 분리하여 기존 모델과 차별점을 주었다. YOLOv7은 보다 우수한 실시간 객체 검출 성능을 달성하고자 여러 계층의 가중치가 더 많은 기능을 학습하기 위해 경로를 효율적으로 제어하는 구조인 E-ELAN을 사용했다. RepVGG (Ding *et al*., 2021) YOLOv5와 YOLOv7은 Anchor-based 방식을 사용하며 YOLOX는 Anchor-free 방식을 사용한다(Zhang *et al*., 2020).

Table 1. Architectures of YOLO family models

Model	Architecture		
	Backbone	Neck	Head
YOLOv5	CSP-Darknet53	SPP, PANet	YOLOv3
YOLOX	CSP-Darknet53	PAFPN	Decoupled Head
YOLOv7	E-ELAN	CSPSPP + (E-ELAN)	YOLOR

두 방식의 차이점을 Fig. 1에 상세히 나타내었다. Fig. 1(a)의 Anchor-based 방식은 사전에 정의된 크기와 종류의 사각형 영역인 앵커(anchor)를 사용해 데이터셋 이미지 내 객체의 위치와 크기를 예측한다. 앵커는 이미지 내에서 다양한 종횡비와 크기를 가진다. 딥러닝 모델은 앵커를 기반으로 물체의 위치를 예측하며 예측한 위치와 실제 객체의 위치 사이의 IoU (Intersection over Union)를 계산한다. IoU는 두 개의 경계 상자가 얼마나 겹치는지 측정하는 값이다. 앵커 박스와 실제 객체의 영역을 표시한 상자 간 IoU의 값이 특정 임계값보다 크다면 정답이라고 간주한다. IoU의 값이 클수록 객체 위치에 대한 검출 능력이 높다는 것을 의미한다. Anchor-based는 학습데이터를 바탕으로 사전에 정의된 n개의 앵커 박스를 이용하여 학습을 통해 객체의 영역을 추론한다. Fig. 1의 (b)는 Anchor-free 방식

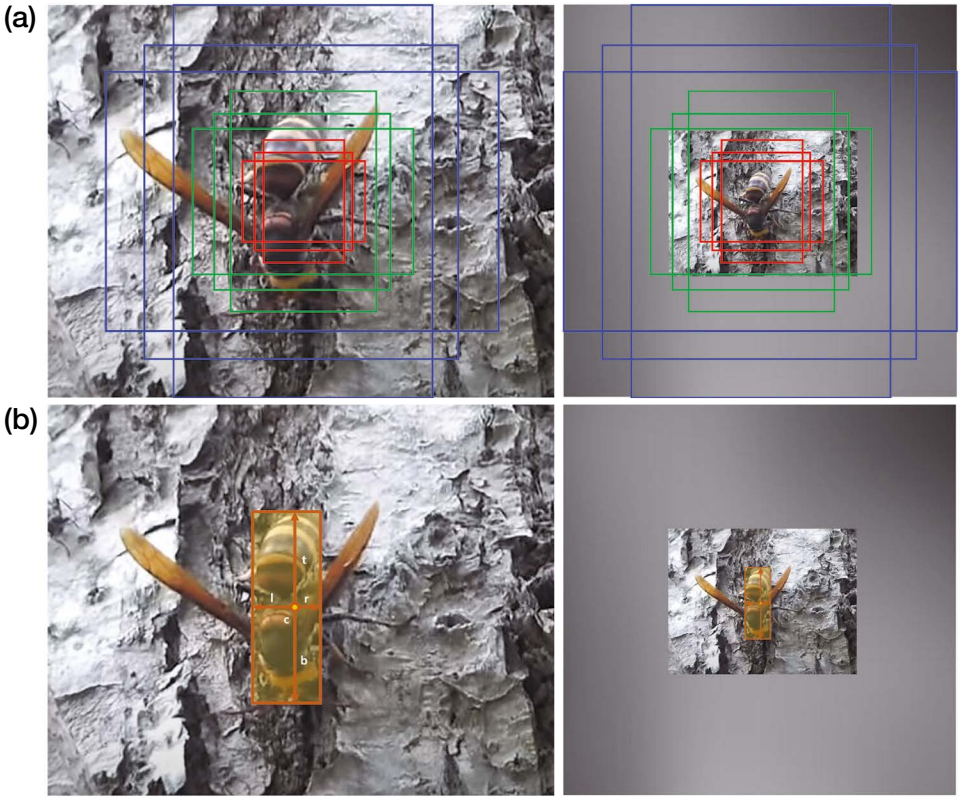

Fig. 1. Comparison between anchor-based method and anchor-free method. (a) Anchor boxs of the learning dataset according to predetermined size and aspect ratio. (b) Anchor-free method predicts the center point.

을 묘사했다. Anchor-free는 사전 정의된 앵커에 의존하지 않고 객체의 위치와 크기를 예측하는 방식이다. 객체의 중심점 위치와 크기에 대해 예측값을 생성하거나 객체의 주요 특징을 감지하여 위치를 파악한다. 앵커를 사용하지 않아 객체 감지를 좀 더 유연하게 할 수 있다. Anchor-based는 학습 데이터 기반으로 높은 정확도를 보여준다는 장점이 있으나 객체의 특징(비율)이 학습된 것에서 벗어나면 예측 성능이 크게 떨어진다. 다만 Anchor-free는 미리 정의한 앵커 없이 객체를 탐지하는 방식이기 때문에 학습이 되지 않은 다양한 객체의 크기에도 예측 성능이 크게 떨어지지 않는다. 딥러닝 모델에 따른 객체 탐지의 성능은 mAP (Mean Average Precision)를 평가 지표로 사용한다. mAP는 딥러닝 모델의 예측 결과를 기반으로 정밀도/재현율 곡선을 생성한다. 정밀도는 모델이 예측한 객체 중 실제 객체로 올바르게 감지한 비율이며 재현율은 실제 객체 중 모델이 올바르게 감지한 객체의 비율이다. 정밀도/재현율 곡선 아래의 면적을 계산한 값의 평균이 mAP이다. mAP를 통해 모든 클래스에 대해 얼마나 효과적으로 객체를 탐지하는지 나타내는 종합적인 성능 지표를 얻는다. 이때 IoU는 0.5를 기준으로 한다. 즉 추론된 객체와 Ground Truth간에 겹치는 비율이 50% 이상인 결과에 대하여 mAP를 계산하여 추론의 정밀도를 평가한다.

2. 데이터셋 구성

Fig. 2는 기본 학습 데이터셋에 포함된 말벌의 종류이다. 기본 데이터셋은 말벌류 5종(등검은말벌, 말벌, 꼬마장수말벌, 장수말벌, 털보말벌)과 양봉꿀벌로 총 6종이다. Table 2와 같이 말벌 영상이 학습용으로 5,705장, 유효성 검사용으로 900장, 검증을 위해 900장으로 이루어진 데이터셋이다.

기본 데이터셋은 객체가 이미지 내에서 많은 면적 비율을 차지하는 일반적인 학습용 데이터셋이다. 추가적으로 작은 객체에 대한 검출 성능을 향상시키기 위한 학습용 데이터를 제작했다. 기본 데이터셋을 기반으로 실제 현장에서 촬영된 카메라 영상 속 말벌의 면적 비율과 흡사하도록 데이터셋을 수정한다. 말벌과 카메라의 거리를 40 cm로 제한할 때 대부분 말벌은 크기에 따라 차이가 있지만 영상 전체 면적 대비 말벌 객체의 크기는 대략 1% 미만의 비율을 차지한다. 따라서 학습 데이터도 실제 면적 비율을 고려하여 임의로 객체의 면적이 영상 전체 면적의 0.3%가 되도록 제작했다. 0.3% 데이터셋의 수는 기본 데이터셋과 동일하다. 즉, 기본 데이터셋을 이용하여 상기에 설명한 방법으로 학습용, 유효성 검사용, 테스트용으로 동일한 수의 0.3% 데이터셋을 생성한다.

Table 2. Training and test dataset

Dataset	Training	Validation	Test
Basic	5705	900	900
0.3%	5705	900	900
Mixed	5705	900	900

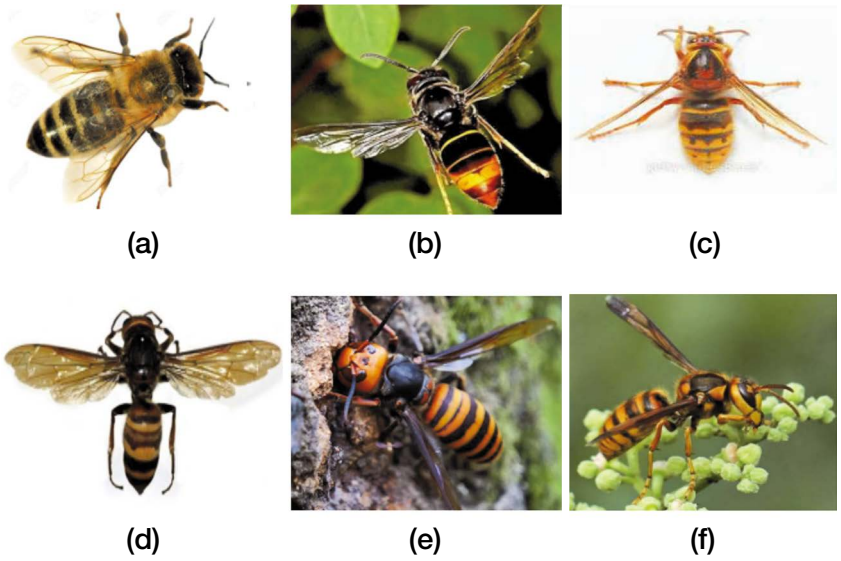

Fig. 2. Vespa dataset. (a) *Apis mellifera*, (b) *V.velutina nigrithorax*, (c) *V. crabro*, (d) *V. ducalis*, (e) *V. mandarina*, (f) *V. silmilima*.

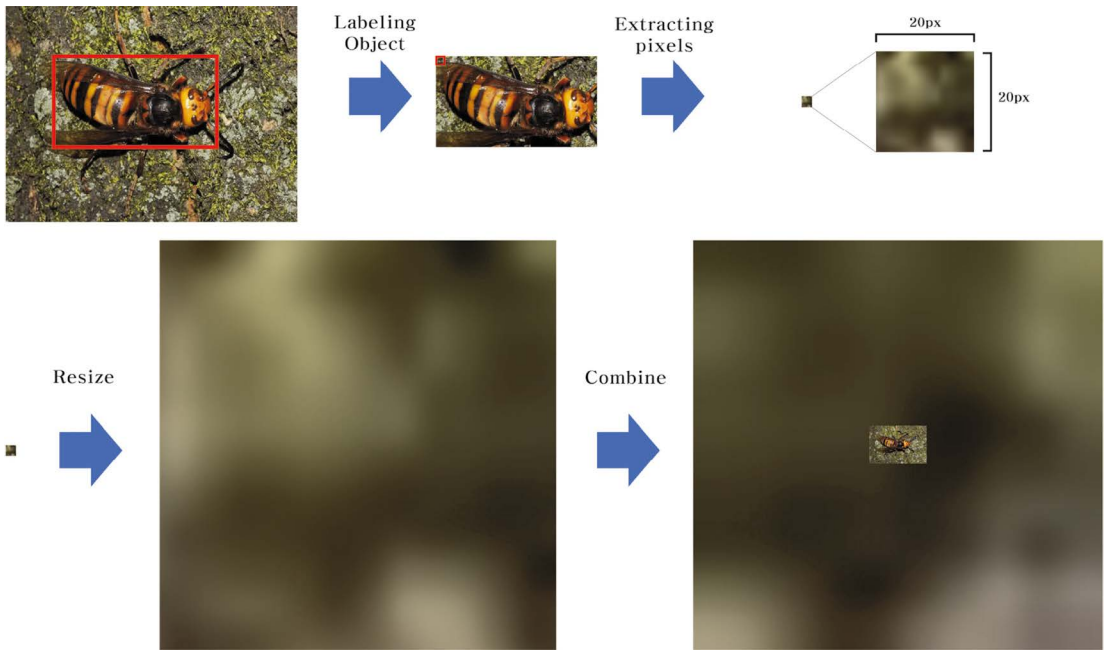

Fig. 3. Process to produce the 0.3% dataset.

Fig. 3은 0.3% 데이터셋을 생성하는 과정이다. 먼저 기본 데이터셋 내에 존재하는 말벌 영상 중 배경을 제외하고 객체(말벌) 영역만을 라벨링한다. 라벨링된 객체 영상 중 왼쪽 위에서부터 가로 및 세로 20픽셀씩 추출 후 저장한다. 추출한 배경은 라벨링된 객체의 비율이 전체 영상 내 면적의 0.3%의 비율을 차지하도록 넓이를 재구성한다. 말벌 영역만을 라벨링한 가로 및 세로 길이는 일정하지 않고 같은 종류의 말벌이라도 영상마다 길이가 다양하다. 가로와 세로의 길이의 비율을 유지하면서 배경의 넓이를 재구성하기에는 어려움이 있다. 따라서 객체가 배경의 0.3%의 면적 비율을 차지하도록 하는 재구성된 배경의 넓이를 쉽게 제작하기 위해 정사각형 모양으로 추출하였다. 라벨링된 객체의 영상의 가로와 세로 길이를 통해 넓이를 구하고 그 넓이의 마지막으로 재구성한 배경 중앙에 원본 말벌 영상을 결합한다.

혼합(Mixed) 데이터셋은 두 가지를 제작했다. 하나는 기본 데이터셋과 0.3% 데이터셋에서 무작위로 선출하여 기본 데이터셋과 개수를 맞춘 데이터셋이다. 다른 하나는 객체의 크기가 기본 데이터셋, 3% 데이터셋, 0.3% 데이터셋으로 총 3단계로 작아지는 데이터셋으로 제작했다. 마찬가지로 기본 데이터셋과 개수를 동일하게 조절하였다.

Fig. 4는 학습용으로 사용되는 기본 데이터셋, 3% 데이터셋, 0.3% 데이터셋에 포함된 영상의 예시이다. 전술한 바와 같이 현장에서 카메라로 촬영하는 말벌은 대략 영상 면적 대비 1% 미만으로 획득된다. 따라서 말벌의 면적 비율을 차례대로 일반적인 데이터셋의 비율인 기본 데이터셋부터 3% 데이터셋 그리고 0.3% 데이터셋까지 3단계로 비율이 작아지도록 혼합 데이터셋을 제작하여 YOLO 모델 학습에 이용했다.

3. 실험 방법

YOLOv5, YOLOX, YOLOv7을 이용하여 준비한 데이터셋 3가지를 각 모델별로 3번의 학습을 각각 진행한다. 학습을 진행한 하드웨어 환경은 GPU NVIDIA RTX4090이며 소프트웨어 환경은 Ubuntu 18.04, CUDA 11.6, CUDNN 8.9.0, OpenCV 4.8.0, Python 3.8.10이다.

Fig. 5는 3가지 YOLO 모델을 3가지 데이터셋으로 학습을 거친 후 테스트하는 과정이다. YOLO 모델을 학습하면 가중치(weight) 데이터가 여러 개 생성된다. 가중치를 이용해 영상 내에 존재하는 객체를 탐지하고 분류할 수 있다. 생성된 가중치 중에서 평가 지표가 가장 높은 best.pth를 테스트에 사용할 가중치로 선정했다. 최종적으로 YOLO 모델 3개마다 데이터셋 3종류를 학습하여 총 9개의 가중치를 이용하여 테스트를 진행한다. 테스트를 마치면 각 테스트의 결과로 mAP가 산출된다. 학습된 데이터셋 종류와 YOLO 모델을 비교하여 mAP가 가장 높은 경우를

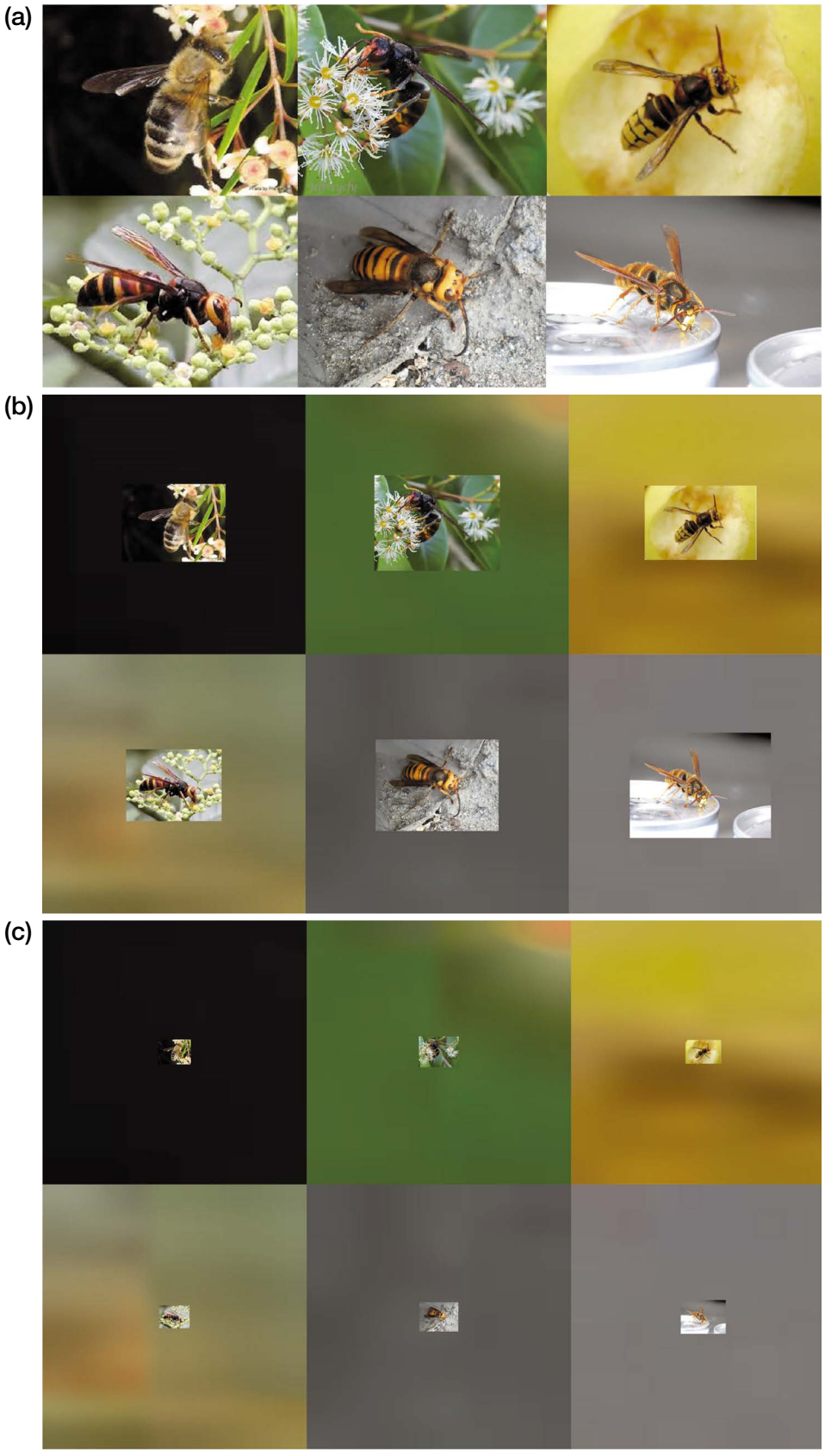

Fig. 4. Configuration of vespa training dataset. (a) Basic training dataset, (b) 3% training dataset, (c) 0.3% training dataset.

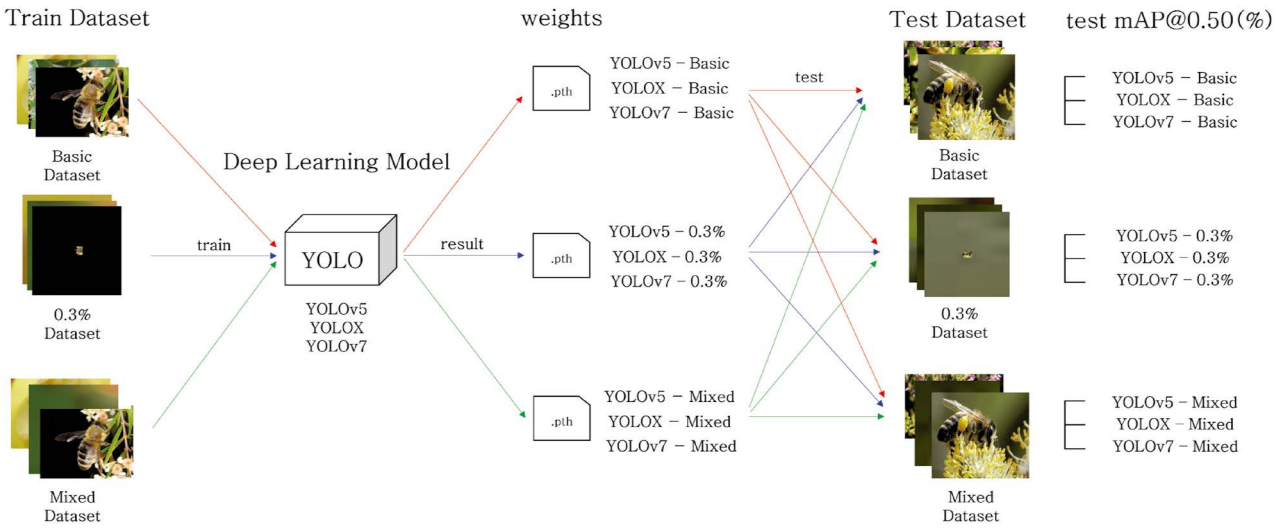

Fig. 5. Process of training and test for 3 YOLO models using the 3 datasets.

찾아 검출 결과가 가장 좋은 YOLO 모델과 효과적인 학습 데이터셋으로 결정한다.

결과 및 고찰

1. 학습 및 테스트 mAP 측정

Table 3은 각 모델을 각 데이터셋에 학습한 후 얻은 가중치에 대해 가장 보편적으로 활용되는 척도 IoU 0.5를 기준으로 학습 mAP를 측정한 결과이며 mAP를 통해 모델의 실질적인 성능 평가가 가능하다. 선행 연구(Everingham, 2010)를 참고하여 IoU 값을 0.5로 설정하였다. 예측한 영역과 실제 객체의 영역의 IoU 값이 0.5보다 크다면 Positive, 그렇지 않으면 Negative로 분류되어 mAP를 계산한다. mAP가 높을수록 딥러닝 모델이 객체에 대해 정확하게 감지하고 바운딩 박스를 예측한다는 것을 나타낸다. 학습 mAP는 모델이 훈련 데이터셋에 대해서 얼마나 잘 학습되었는지를 나타내는 지표이다. YOLOv7이 Basic, 0.3%, Mixed 3가지 데이터셋 학습에서 각각 96%, 97.8%, 96.1%로 YOLO 모델 중 모두 높은 학습 mAP를 나타내는 것을 확인할 수 있다. YOLOX는 3가지 데이터셋을 학습한 결과로 각각 95%, 92%, 93.5%가 나왔으며 특히 0.3% 데이터셋을 학습했을 때 YOLO 모델 중 가장 낮은 mAP를 나타냈다. YOLOv5의 3가지 학습 결과는 순서대로 95%, 95%, 95.9%를 나타냈으며 가장 학습 결과가 높은 YOLOv7과 mAP가 3% 미만 차이로 비슷한 수준의 학습 결과를 보였다.

Table 3. Training mAPs for the YOLO models.

Model	Train/val mAP@0.50 (%)			
	Basic dataset	0.3% dataset	Mixed dataset	*Mixed dataset
YOLOv5	95.0	95.0	95.6	95.9
YOLOX	95.0	92.0	93.2	96.1
YOLOv7	96.0	97.8	96.6	93.5

학습을 통해 얻은 9가지 가중치를 4가지의 테스트 데이터셋을 이용해 추론 성능을 평가했다. 학습이 완료된 YOLO 모델을 테스트 데이터셋에 포함된 말벌 영상을 이용해 테스트를 진행했다. 테스트는 학습 데이터와 달리 새로운 데이터를 통해 수행했다. 테스트 mAP는 모델이 실제 환경에서 얼마나 잘 탐지할 수 있는지 반영하며 이를 통해 최종적으로 말벌을 잘 탐지할 수 있다는 객관적인 성능 정보를 얻는다. 즉, mAP가 높을수록 말벌을 탐지하는 성능이 우수하다는 객관적인 지표이다. Table 4는 각 가중치와 테스트 데이터셋을 이용하여 추론했을 때의 mAP이다.

테스트 데이터는 기본 데이터셋, 0.3% 데이터셋, 기본과 0.3%를 합친 Mixed 데이터셋, 기본과 3%와 0.3%를 동일한 비율로 구성한 *Mixed 데이터셋을 사용했다. 테스트 학습 데이터와 테스트 데이터의 성질이 유사한 경우에서 YOLOv5와 YOLOv7이 mAP가 가장 높은 성능을 보

여주었다. 예시로 Basic 데이터셋으로 학습한 YOLOv5와 YOLOv7으로 Basic 데이터셋으로 테스트한 결과, 각 mAP가 93.1%, 92.9%를 나타냈다. 다만, 학습 데이터의 성질과 테스트 데이터의 성질이 다르다면 성능이 크게 저하되는 것을 볼 수 있었다. 그 예로 Basic 데이터셋으로 학습한 모델의 가중치는 실제 환경과 유사한 조건으로 맞춘 0.3% 데이터셋으로 테스트를 시행한 결과를 보면 YOLOv5가 30.8%, YOLOv7이 14.7%로 현저히 낮은 mAP가 측정되었다. 하지만, YOLOX는 85.2%로 YOLO 모델 중 압도적으로 우수함을 보여줬다. 두 가지 Mixed 데이터셋으로 학습한 YOLO 모델들은 테스트 데이터셋에 관계없이 준수한 mAP를 보여주었다. 마지막으로 YOLO 모델을 두 가지 Mixed 데이터셋으로 테스트를 진행한 결과는 큰 차이가 없이 비슷한 수준의 mAP를 보여주었다. Fig. 1과 같이 YOLOX는 Anchor-free로 인해 학습 데이터와 테스트 데이터의 객체 크기가 달라도 영향을 크게 받지 않는 특성이 그 이유이다. 다시 YOLOv5와 YOLOv7을 Basic 아닌 0.3% 데이터셋으로 학습한 후 0,3% 데이터셋으로 테스트하면 각각 94%, 95.4%의 결과를 보여준다. Anchor-based 기반인 YOLO 모델을 추론 환경과 유사한 학습 데이터로 학습을 한 후 추론한다면 일반적인 데이터셋으로 추론했을 때보다 높은 성능을 보여준다는 것을 실험 결과를 통해 알 수 있다.

Table 4. Resultant inference mAPs according to datasets in each trained YOLO model

Model weight	Resultant mAP@0.50(%) for each test dataset			
	Basic dataset	0.3% dataset	Mixed dataset	*Mixed dataset
YOLOv5 with Basic	93.1	30.8	71.6	78.3
YOLOv5 with 0.3%	6.16	94.0	63.5	53.5
YOLOv5 with Mixed	92.7	92.0	92.0	91.6
YOLOv5 with *Mixed	91.5	89.5	90.1	91.9
YOLOX with Basic	87.1	85.2	84.6	88.3
YOLOX with 0.3%	0.5	89.0	41.6	28.7
YOLOX with Mixed	87.1	89.9	87.3	87.2
YOLOX with *Mixed	85.6	88.4	85.8	87.0
YOLOv7 with Basic	92.9	14.7	59.5	70.0
YOLOv7 with 0.3%	2.16	95.4	44.8	30.0
YOLOv7 with Mixed	93.5	90.6	91.3	89.5
YOLOv7 with *Mixed	90.5	90.1	90.0	91.1

2. 테스트 mAP 결과에 대한 추론 영상 예시

본 연구의 목적은 현장에서도 잘 탐지할 수 있는 YOLO

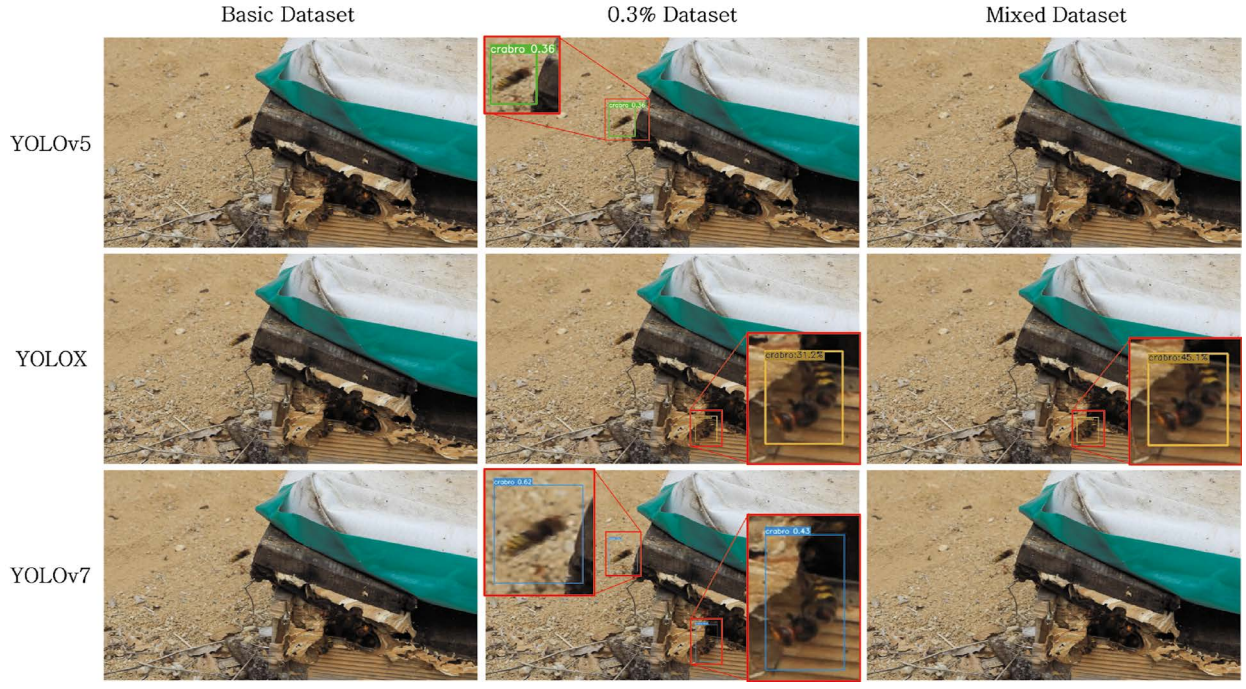

Fig. 6. Results of Inference by 3 trained-YOLO models for a Vespa video.

모델을 알아보기 위한 것이다. 사진으로 이루어진 데이터셋으로 테스트를 진행한 결과인 mAP에 대한 시각적 자료 예시로 말벌을 촬영한 동영상을 탐지하는 추가적인 평가를 수행하였다.

Fig. 6은 모델과 가중치별로 같은 초 단위 동영상의 한 지점을 스크린샷한 사진이다. 영상은 학습한 5종의 말벌류(등검은말벌, 장수말벌, 꼬마장수말벌, 말벌, 털보말벌) 중에서 말벌(crabro)이 촬영된 영상이다. 말벌을 인식하면 그 위치에 바운딩 박스가 생기며 말벌의 종류와 바운딩 박스가 말벌을 포함하고 있을 가능성인 confidence score를 보여준다. Confidence score를 보기 쉽도록 바운딩 박스 부분을 편집했다. Fig. 6에서 YOLOv5는 Basic 데이터셋과 Mixed 데이터셋으로 학습한 가중치로 추론한 결과로 말벌을 인식하지 못했으나 0.3% 데이터셋으로 학습한 YOLOv5로 추론하면 말벌의 confidence score가 36%로 인식한 것을 볼 수 있다. Basic 데이터셋으로 학습한 YOLOX는 말벌을 탐지하지 못했으나, 0.3% 데이터셋과 Mixed 데이터셋으로 학습한 후 영상을 추론하면 각각 31.2%, 45.1%로 말벌을 탐지한 것을 확인했다. 마지막으로 YOLOv7 중에서 Basic 데이터셋과 Mixed 데이터셋으로 학습한 가중치로 영상을 추론했을 때 말벌을 인식하지 못하였다. 0.3% 데이터셋으로 학습한 YOLOv7은 두 마리의 말벌을 각각 62%, 43%로 탐지하는 것을 확인했다.

YOLO 모델을 말벌 추론 영상을 실험한 결과로 YOLOv5와 YOLOv7은 0.3% 데이터셋으로 학습했을 때 Basic과 Mixed 데이터셋으로 학습한 결과보다 탐지 능력이 우수한 것을 확인했다. YOLOX는 다른 두 모델과 비교했을 때 학습 데이터셋의 차이로 인한 탐지 능력의 차이가 크게 보이지 않았다. Anchor-based 기반인 YOLO 모델로 탐지를 하기 위해서는 현장과 비슷한 환경으로 맞춰진 데이터셋으로 학습을 하면 그렇지 않은 데이터셋으로 학습하는 것보다 탐지 결과가 비교적 우수한 것을 확인할 수 있다. Anchor-free 기반 YOLO 모델은 학습 데이터셋에 크게 영향을 받지 않는 것을 확인했다.

본 논문에서는 YOLO 모델 중 YOLOv7를 선택하여 실제 추론 환경과 유사하도록 적절한 데이터셋인 0.3% 데이터셋으로 학습한 후 객체 추론을 한다면 일반적인 데이터셋으로 학습하여 추론하는 방법보다 우수한 성능을 보인다는 것을 입증한다.

적 요

말벌은 양봉 산업에 피해를 주는 요소 중 하나이다. 이 피해를 줄이기 위해서는 말벌 탐지를 기반으로 한 방제시스템의 개발이 필요하다. 본 연구에서는 말벌과 같이 작은 객체의 탐지를 위한 딥러닝 모델을 선택할 때 학습용 데이터의 구성에 따른 모델의 성능을 비교하여 말벌 탐지에서 유용한 딥러닝 모델 및 학습용 데이터 구성에 대한 방법론을 제시하였다. 딥러닝 모델은 가장 많이 활용되는 YOLOv5, YOLOX, YOLOv7을 선택했다. YOLOv5와 YOLOv7은 Anchor-based, YOLOX는 Anchor-free 기반 YOLO 모델이다. 그리고 학습용 데이터셋으로는 기본 말벌 데이터셋과 객체(말벌)가 영상 내 차지하는 면적 비율을 전체 영상 면적 비율에서 0.3%를 차지하도록 조절한 0.3% 데이터셋, Mixed 데이터셋으로 총 3가지를 준비했다. YOLO 모델 3가지와 데이터셋 3가지를 조합하여 9가지 실험을 수행한 후, 성능 평가 지표로서 mAP를 활용했다. 말벌을 학습한 YOLOv5, YOLOX, YOLOv7을 테스트 데이터셋으로 테스트한 결과로 기본 데이터셋으로 학습한 가중치를 0.3% 테스트 데이터셋을 실험한 결과로 YOLOv5와 YOLOv7은 mAP가 5% 미만을 나타냈다. 다만 YOLOX는 학습 데이터셋 내의 객체의 크기에 영향을 크게 받지 않는 Anchor-free 특성으로 인해 85%로 Anchor-based 모델에 비해 우수한 성능을 보여주었다. 0.3% 데이터셋으로 학습한 YOLOv5, YOLOX, YOLOv7을 0.3% 테스트 데이터셋으로 테스트하면 각각 94%, 89%, 95.4%의 결과를 보여주었다. 즉, 말벌 영상을 활용한 추론 실험에서는 YOLOv5와 YOLOv7이 테스트 환경과 유사하도록 설정한 데이터셋으로 학습되었을 때 mAP가 90% 이상으로 높은 탐지 능력을 보여주었다. 특히 YOLOv7이 95.4%로 가장 높은 mAP를 보여주었다. YOLOX는 데이터셋의 차이와 상관 없이 모든 테스트에서 mAP가 80% 이상으로 준수한 성능을 나타냈다. 이 연구는 YOLO 모델을 학습할 때 데이터셋 선택의 중요성에 대해 탐구하며, 실험에 사용된 YOLO 모델 중 YOLOv7이 준비된 데이터셋 중 0.3% 데이터셋으로 학습한 후 말벌을 탐지할 때 가장 우수하다는 것을 확인하였다.

감사의 글

본 논문은 농촌진흥청 공동연구사업(과제번호: PJ01476

103)으로 수행된 결과이며 농촌진흥청의 지원에 깊은 감사를 드립니다.

인 용 문 헌

Bochkovskiy, A., C. Y. Wang and H. Y. M. Liao. 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Ding, X., X. Zhang, N. Ma, J. Han, G. Ding and J. Sun. 2021. Repvgg: Making vgg-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 13733-13742.

Everingham, M., L. Van Gool, C. K. Williams, J. Winn and A. Zisserman. 2010. The pascal visual object classes (voc) challenge. IJCV 88(2): 303-338.

Ge, Z., S. Liu, F. Wang, Z. Li and J. Sun. 2021. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

Girshick, R. 2015. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision 1440-1448.

Girshick, R., J. Donahue, T. Darrell and J. Malik, 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 580-587.

He, K., G. Gkioxari, P. Dollr and R. Girshick. 2017. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision 2961-2969.

He, K., X. Zhang, S. Ren and J. Sun. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37(9): 1904-1916.

Howard, A., M. Sandler, G. Chu, L. C. Chen, B. Chen, M. Tan and H. Adam. 2019. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision 1314-1324.

Jeong, S. M., C. Y. Lee, D. W. Kim and C. E. Jung. 2016. Questionnaire Study on the Overwintering Success and Pest Management of Honeybee and Damage Assessment of Vespa Hornets in Korea. Korean J. Apic. 31(3): 201-210.

Jeong, Y. J., H. S. Hwang, Y. J. Kwon and C. H. Lee. 2022a. Design and Implementation of Real-time Wasp Monitoring System using Edge Device. J. Korea Multimed. Soc. 25(12): 1826-1839.

Jeong, Y. S., M. S. Jeon, S. B. Kim, D. W. Kim, S. H. Yu, K. C. Kim and I. C. Choi. 2022b. Study on the Technology for Searching Vespa Velutina Nest Using YOLO-v5. Korean J. Apic. 37(3): 255-263.

Kim, K. C., D. S. Seo, I. C. Choi, Y. K. Hong, G. H. Kim and K. D. Kwon. 2021. Development of *Vespa velutina* Monitoring System Based on Deep Learning. JKAIS. 22(10): 31-36.

Lin, T. Y., P. Dollar, R. Girshick, K. He, B. Hariharan and S. Belongie. 2017. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2117-2125.

Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu and A. C. Berg. 2016. Ssd: Single shot multibox detector. In Computer Vision - ECCV: 14th European Conference, Amsterdam, The Netherlands 21-37.

Ozge, U. F., B. O. Ozkalayci and C. Cigla. 2019. The power of tiling for small object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 0-0.

Park, H. J. 2020. Trend Analysis of Korea Papers in the Fields of 'Artificial Intelligence', 'Machine Learning' and 'Deep Learning'. Journal of KIIECT 13(4): 283-292.

Redmon, J. and A. Farhadi. 2017. YOLO9000: better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition 7263-7271.

Redmon, J. and A. Farhadi. 2018. YoloV3. Computer Vision and Pattern Recognition. Accessed in https://doi.org/10.48550/arXiv.1804.02767 8 Apr. 2018.

Redmon, J., S. Divvala, R. Girshick and A. Farhadi. 2016. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 779-778.

Ren, S., K. He, R. Girshick and J. Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in NeurIPS, 28.

Tan, M. and Q. Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on machine learning 6105-6114.

Wang, C. Y., A. Bochikovskiy and H. Y. M. Liao. 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 7464-7475.

Zhang, S., C. Chi, Y. Yao, Z. Lei and S. Z. Li. 2020. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9759-9768.