



YOLO-v5를 이용한 등검은말벌집 탐색 기술 연구

정유석^{1,3}, 전문석^{1,3}, 김수배², 김동원², 유승화¹, 김경철¹, 이시영¹, 이창우³, 최인찬^{1,*}

¹국립농업과학원 농업공학부, ²국립농업과학원 농업생물부, ³군산대학교 소프트웨어학부

Study on the Technology for Searching *Vespa Velutina* Nest Using YOLO-v5

Yu Seok Jeong^{1,3}, Moon-Seok Jeon^{1,3}, Su-bae Kim², Dongwon Kim², Seung-Hwa Yu¹, Kyung-Chul Kim¹, Siyoung Lee¹, Chang-Woo Lee³ and Inchan Choi^{1,*}

¹Department of Agricultural Engineering, National Institute of Agricultural Sciences, Jeonju, Republic of Korea

²Department of Agricultural Biology, National Institute of Agricultural Sciences, Jeonju, Republic of Korea

³School of Computer Information and Communication Engineering, Kunsan National University, Gunsan, Republic of Korea

Abstract

Vespa velutina, such as ecological disturbance wildlife introduced into Busan through Shanghai, China in 2003, is an ecological disturbance wildlife and difficult to find with the naked eye due to the nature of creating wasp nests in high places in the forest. In addition, its population increases every year, causing the economical damage to beekippers (about 170 billion won). Therefore, the purpose of this study is to acquire images and explore the location of *Vespa velutina* based on artificial intelligence (AI) to easily remove *Vespa velutina* using drones. Differently from the ground images, aerial images have the characteristic that the objects become very small as the altitude rises, so it is necessary to study the appropriate image size that can be mounted on drones and can search *Vespa velutina* with AI. The AI model YOLO-v5 using four image sizes (640×384 px, 1,280×736, 1,920×1,088, 3,840×2,176) was applied in this study and the original size of the image (3,840×2,160) was used for learning and verification. When confidence was higher than 0.7, F1 score of YOLO-v5s-default (640) learned with image size 640×384 was 2.4%, YOLO-v5s-1280 was 36.5%, YOLO-v5s-1920 was 64.2% and YOLO-v5s-3840 obtained the best detection performance at 96.1%. In addition, the performance was confirmed when the verification image size was 3,840×2,176 in four AI learned with different image sizes. In this case, YOLO-v5s-default (640) had F1 score of 4.4%, YOLO-v5s-1280 was 7.1% and YOLO-v5s-1920 was 16.4% with no performance improvement. Therefore, this study found that the networks that are learned and validated using the aerial images larger than 640 showed the better performance. Continuous research on real-time information sharing and estimation of the location of the occurrence of *Vespa velutina* through data collection using drone systems will be carried out.

Keywords

Artificial intelligence, *Vespa velutina* nest, Detection

서 론

인간이 먹기 위해 기르는 작물 중의 75%가 꿀벌이나 나

비와 같은 화분 매개 동물의 수분에 의존함에 따라 매개 동물인 꿀벌은 농업에 큰 비중을 차지하고 있어 개체수 관리가 중요하다. 2003년 중국 상하이를 통해 부산으로 유

입된 등검은말벌은 말벌과 곤충으로 열대 아시아가 원산지인 아열대성 말벌로 털보말벌과 함께 국내 꿀벌 및 매개 동물을 위협하며 많은 피해를 주고 있다. 등검은말벌에 따른 연간 피해액은 약 1,700억 원으로 추정되며, 국내 말벌들과 다르게 꿀벌에 대한 사냥 성공률이 높고 사냥에 걸리는 시간도 짧은 편이다. 이에, 등검은말벌은 2019년 대한민국의 생태계 교란 야생생물로 지정되었으며, 전국에서 채집한 말벌 개체의 72%를 차지하는 우점종으로 번식력이 강하며 도시환경에 적응성이 좋다. 또한, 1개의 벌집 군체가 최대 10,000~12,000마리에 이르고, 번식기에는 벌집 하나당 출방하는 여왕벌 수가 300~500마리에 육박하기 때문에 확산성 역시 뛰어나 매년 피해가 늘어나고 있다. 활동 시기는 3~5월 월동한 여왕벌들이 출현하기 시작하며 4~5월 주간 평균 최고 온도가 27~28°C일 때 교미, 애벌레 육아 등의 활동이 활발해진다. 6월 하순부터 개체 수가 증가하고 8~9월에 활동이 가장 왕성하며 말벌집 크기를 키워가는 시기이다. 10월 중순까지 활동하고 월동을 하여 다음해 봄에 다시 출현한다. 등검은말벌은 보통 10m 이상의 나무에 많이 위치하고 생활 반경은 말벌집 기준으로 1 km 내외이기 때문에 양봉장과 인근 마을의 주변 숲 속에 생성되어 사람의 눈으로 발견하여 제거하기에는 어려움이 크다(Abrol, 1994; Jung *et al.*, 2007a, 2007b; Jung, 2012; Sim *et al.*, 2014).

최근, 말벌 탐지 연구는 수행되었으나 말벌 피해 발생의 중심인 말벌집 탐지 연구는 부족하며 기존 지상 영상 경우와는 다르게 넓은 범위와 높은 시야에서 매우 작게 변하는 객체들에 대한 인공지능 신경망 연구는 부족한 실정이다

(이철희 등, 2019; 김경철 등, 2021). 그러나 고해상도의 항공 영상과 인공지능 기술을 이용한 원격탐사 연구가 활발히 수행되고 있다. 이는 넓은 면적을 효율적으로 관리 및 모니터링할 수 있다는 장점이 있다. 이러한 장점으로 등검은말벌집보다 높은 위치에서 영상 촬영이 가능하다. 또한 등검은말벌집 크기는 월동 전 가장 커지며 나뭇잎이 떨어져서 비교적 발견이 수월한 이점과 가장 군세가 클 때 제거하여 이듬해에 발생하는 월동한 개체를 줄일 수 있다. 따라서 본 연구에서는 항공 영상과 인공지능 알고리즘을 활용하여 조기에 등검은말벌집의 검출이 가능한 모델을 개발하고자 하였다. 이를 위해, 드론(DJI Mavic2)을 사용하여 전북 부안군과 완주군에서 등검은말벌집 영상 데이터를 획득하였으며 YOLO-v5 (Jocher *et al.*, 2020; Zhu *et al.*, 2021)를 사용하여 등검은말벌집 검출을 위한 인공지능 학습을 진행하였다.

재료 및 방법

1. 데이터 수집 및 가공

등검은말벌집 영상 데이터 10월부터 월동 전까지 전북 부안군과 완주군에 위치한 양봉장 일대 19곳에서 드론에 설치된 RGB 카메라를 이용하여 해발고도 65~150 m에서 취득하였다. 드론은 DJI Mavic2 Enterprise (DJI, China)를 사용하였고, RGB 카메라의 해상도는 3,840×2,160이며, 초당 30 프레임(frame)으로 데이터가 수집된다. 등검은말벌집 영상 데이터 크기에 따른 특징점을 얻어 인공지능 학



Fig. 1. Pro-processing of image data to get labeled data using the 'labellmg' program.



Fig. 2. Image data used for validation.

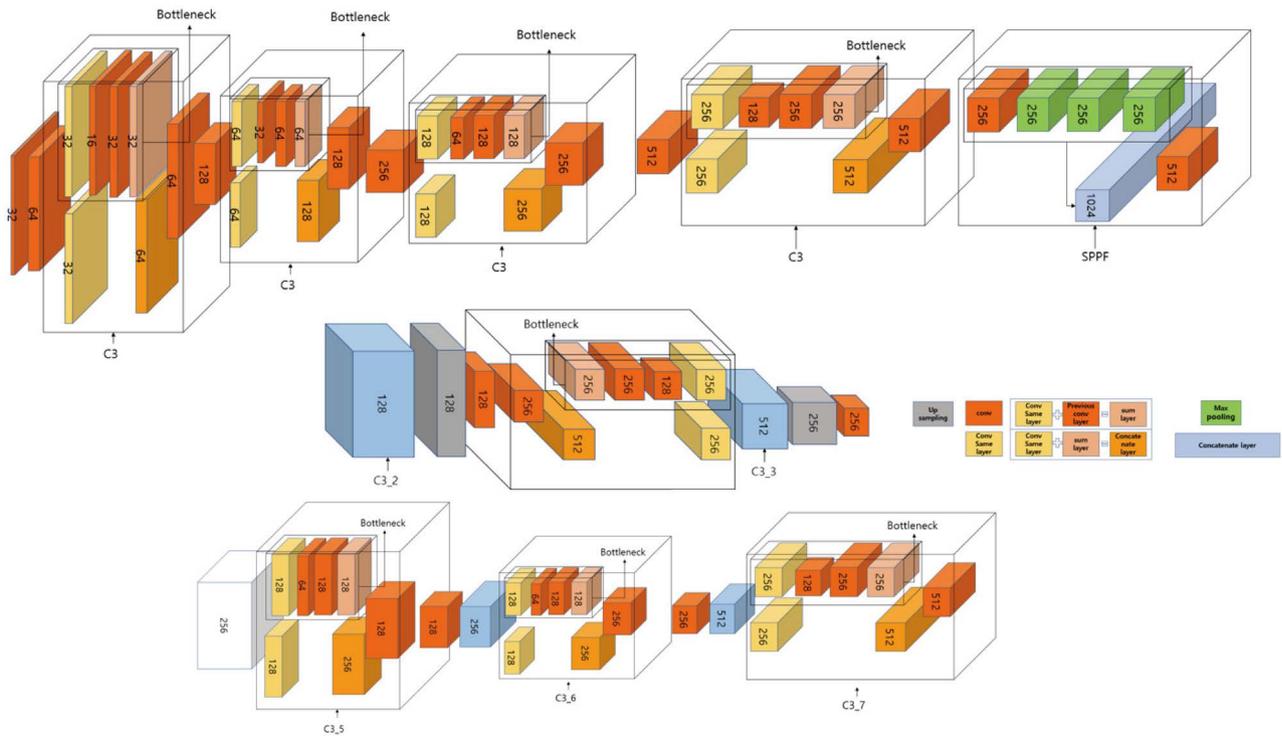


Fig. 3. Structure of Artificial Neural Network (YOLO-v5s).

습에 사용하였다. 학습에 사용된 데이터는 총 11,838장이며, 10,095개의 영상 데이터는 등검은말벌집의 좌표정보가 포함되고 좌표정보 지정 프로그램(labelImg, MIT)을 사용하여 데이터를 전처리하였다(Fig. 1).

검증 데이터는 해발고도 110 m 높이에서 GPS 자율 비행으로 취득한 영상을 사용하였다. 카메라 각도를 드론의 정중앙 하단부로 고정하였으며 전라북도 완주군에서 초당 30프레임으로 3분 58초 동안 촬영된 7,158장이다

(Fig. 2). 학습과 검증에 사용한 영상 데이터는 각 11,838장, 7,158장으로 비율은 약 1:0.6이다.

2. YOLO-v5 학습 하드웨어 및 설정

인공지능 모델의 학습을 위해 중앙처리장치(CPU)는 Intel i7-11700k 3.6 GHz, 메모리(RAM) 128.0 GB, 그래픽 NVIDIA GeForce RTX 3080Ti 12 GB를 사용하였다. 본 연

Table 1. Training setting of Yolo-v5 for the four image sizes

Model	Batch size	Image size	Multi-scale	Image processing
(a)	16	640 × 384	×	O
(b)	8	1,280 × 736	O	O
(c)	4	1,920 × 1,088	O	O
(d)	1	3,840 × 2,176	O	O

(a) YOLO-v5s-default, (b) YOLO-v5s-1280, (c) YOLO-v5s-1920, (d) YOLO-v5s-3840

구에서는 YOLO-v5 기반의 인공지능 모델을 사용하였으며, YOLO-v5s 모델에서 파라미터 변화에 따른 탐색 성능을 비교하였다(Fig. 3).

학습에 있어서 YOLO-v5는 입력 영상 크기(input image size)에 따라 검출 상자(Anchor box) 크기가 달라지며 이에 따라서 검출할 수 있는 객체의 최대, 최소 크기가 다르다. 따라서 입력 영상 크기 변경, 검출 상자 크기 변환, 입력 영상 크기 50% 범위에서 무작위 조정(Multi-scale), 학습에서 사용하는 영상의 최소 개수(batch size), 파라미터를 변경하여, 검출할 수 있는 객체의 크기 조정을 통해 항공 영상에서의 검출 성능 향상을 목적으로 학습하였다. 또한, 날씨에 따른 다양한 영상 변화와 위치에 따른 변화된 데이터를 얻기 위해 사각 무늬 형태(Image Mosaic), 밝기 대비(Brightness Contrast) 등의 영상처리 기법을 활용하여 학습에 적용하였다(Redmon and Farhadi, 2017, 2018) (Table 1).

3. 영상 크기 학습 및 검증

등검은말벌집 탐색을 위해 YOLO-v5에 영상 크기가 다른 4종류의 데이터를 사용하여 학습과 검증하였다. YOLO-v5s-default는 YOLO-v5의 정형화된 파라미터 설정을 그대로 사용하며 입력 영상 크기는 640 × 384이다. 또한 YOLO-v5s-1280의 영상 크기는 1,280 × 736, YOLO-v5s-1920은 1,920 × 1,088, YOLO-v5s-3840은 3,840 × 2,176 영상 크기로 정형화된 파라미터 설정값이 아닌 입력 영상 크기에 따라 학습하였다. 검증은 전라북도 완주군에서 획득한 초당 30프레임으로 3분 58초 동안 촬영한 영상을 프레임 단위로 나누어 총 7,158 프레임 데이터를 사용하였다. 검증 데이터 영상 크기는 학습에 사용된 영상 크기와 동일한 640 × 384, 1,280 × 736, 1,920 × 1,088, 3,840 × 2,176를 사용하였다. 또한 앞에서 학습된 4개의 신경망을

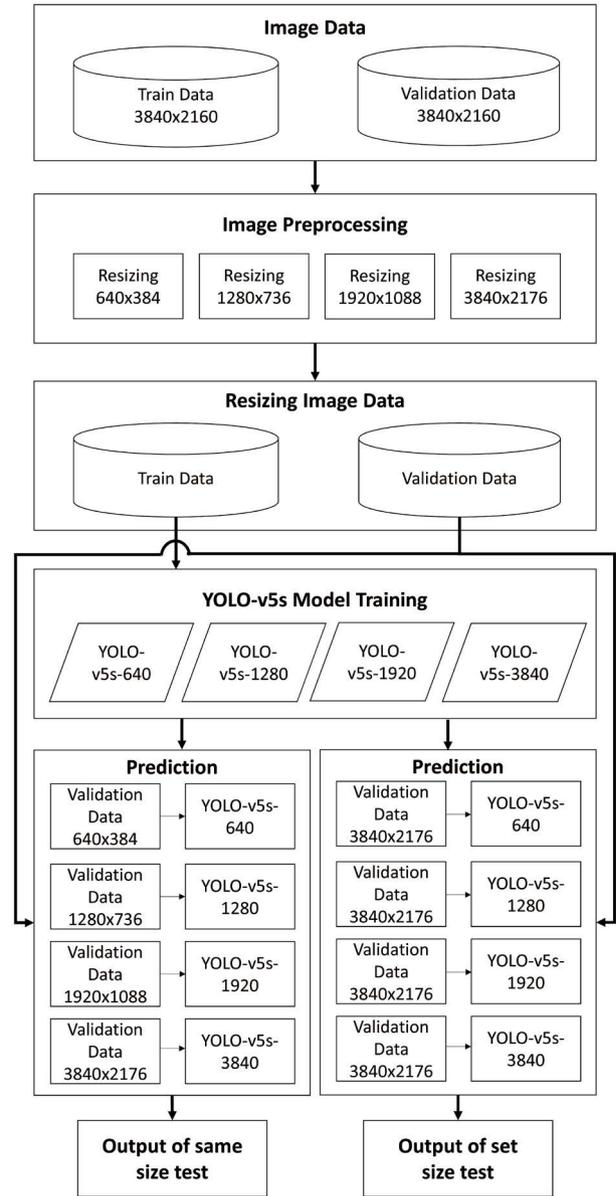


Fig. 4. Process for the two tests (same and set images).

이용하여 검증 영상 크기를 3,840 × 2,176으로 고정된 후 성능을 비교 분석하였다(Fig. 4).

4. 등검은말벌집 탐색 알고리즘 성능 평가

인공지능 학습 성능은 정밀도(Precision), 재현율(Recall) 및 F1 수치(F1 Score)를 계산하여 분석한다. 정밀도는 신경망이 정답이라고 분류한 것 중에서 실제 정답의 비율을 의미하고, 재현율은 실제 정답 중에 신경망이 정답이라고 한 비율, F1 수치는 정밀도와 재현율의 조화평균을 이용

한 수치다. 혼동 행렬은 True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN)로 4가지 수치가 있다. 정밀도(Precision, 식 1), 재현율(Recall, 식 2)과 F1 수치(F1 Score, 식 3)는 혼동 행렬 수치에 의해 결정되며 식은 다음과 같다(Stehman, 1997).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1\ Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (3)$$

결과 및 고찰

1. 영상 크기에 따른 학습 및 검증

인공지능 학습은 각각의 학습 손실 그래프(Train loss graph)가 오차의 변화가 증가하지 않을 때까지 진행하였다

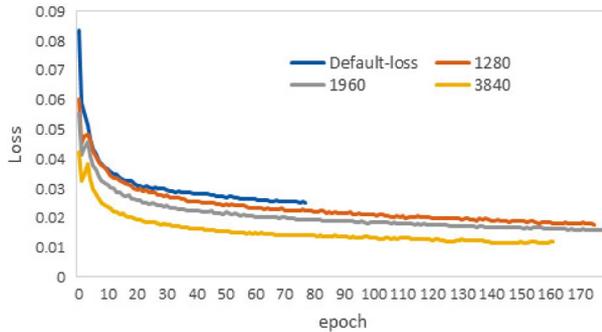


Fig. 5. Train loss graph for the four modules.

(Fig. 5). 학습 손실 그래프의 최소값을 찾는 최적 수식은 Adaptive Moment Estimation (Adam)을 사용하였다(Kingma and Ba, 2015). 영상 크기에 대한 YOLO-v5s-default의 epoch는 77이고 Loss 0.025이며, YOLO-v5s-1280은 175와 0.017, YOLO-v5s-1920은 179와 0.015, YOLO-v5s-3840은 161와 0.011의 결과를 얻었다(Table 2).

본 연구에서 TP는 실제 말벌집을 말벌집이라 판단한 수치이며, FP는 말벌집이 아닌 것을 말벌집이라 판단, FN는 말벌집을 말벌집이 아니라고 판단, TN는 말벌집이 아닌 것을 말벌집이 아니라고 판단한 것을 의미한다(Table 3).

검증 영상 7,158 프레임 중에서 등검은말벌집은 1개가 존재하며 총 229 프레임에 등장한다. YOLO-v5s-default는 7 프레임, YOLO-v5s-1280은 214 프레임, YOLO-v5s-1920은 213 프레임, YOLO-v5s-3840은 212 프레임에서 등검은말벌집을 검출했다(Fig. 6). 검증 영상에서 YOLO-v5s-default는 332 프레임, YOLO-v5s-1280은 726 프레임, YOLO-v5s-1920은 220 프레임, YOLO-v5s-3840은 0 프레임에서 오검출이 발생했다(Fig. 7).

검증 영상에서 등장하는 등검은말벌집 229 프레임 중 한 번이라도 등검은말벌집을 검출한다면 위치를 추정할 수 있는 데이터를 도출할 수 있다. 따라서 4개의 신경망 모두 등검은말벌집 위치 추정이 가능하다는 결과를 얻었다. 하지만 FP의 수치가 높을수록 위치 추정에 혼란이 생기기 때문에 FP 수치가 낮을수록 더 정확한 등검은말벌집 위치

Table 2. Training result for the image sizes

Model	Epoch	Loss
YOLO-v5s-default	77	0.025
YOLO-v5s-1280	175	0.017
YOLO-v5s-1920	179	0.015
YOLO-v5s-3840	161	0.011

Table 3. Validation table according to training image sizes

Model	Confusion matrix				Precision	Recall	F1 Score
	TP	FP	FN	TN			
(a)	7	332	222	6,597	2.0%	3.0%	2.4%
(b)	214	726	15	6,203	22.7%	93.4%	36.5%
(c)	213	220	16	6,709	49.1%	93.0%	64.2%
(d)	212	0	17	6,929	100%	92.5%	96.1%

(a) YOLO-v5s-default (640), (b) YOLO-v5s-1280, (c) YOLO-v5s-1920, (d) YOLO-v5s-3840

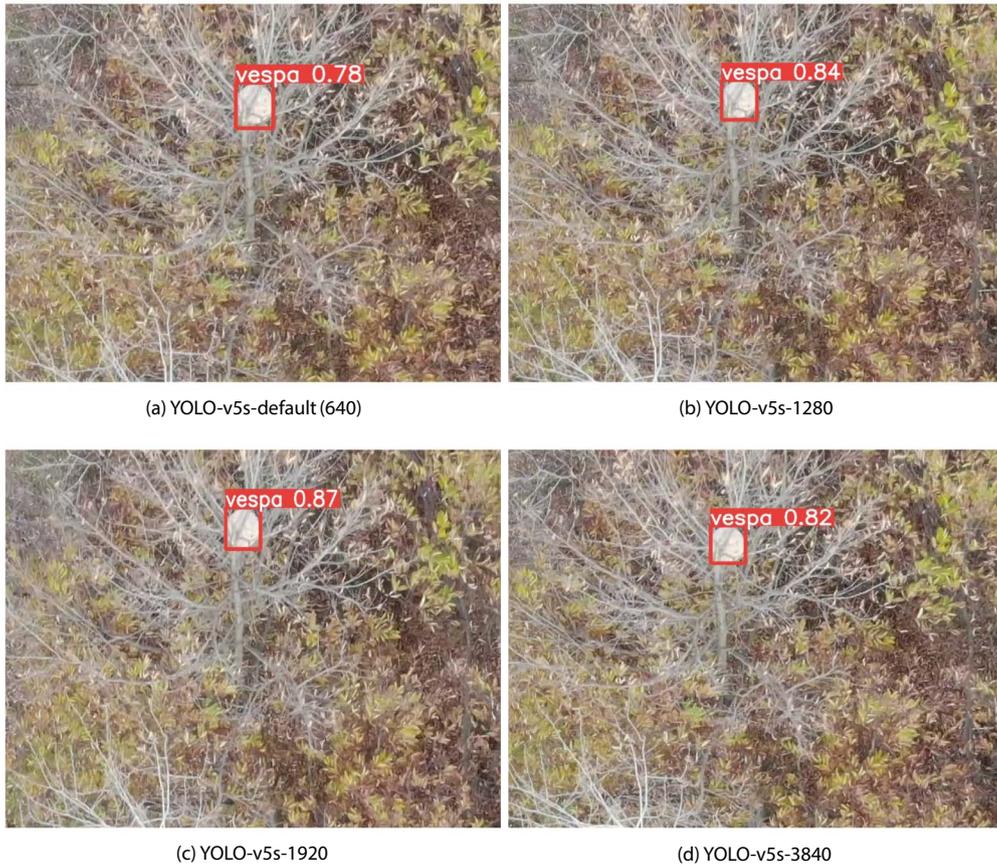


Fig. 6. Detection of Velutina nest in image data.

를 추정할 수 있다고 판단하였고 Table 3과 같이 YOLO-v5s-3840이 정밀도, 재현율 및 F1 수치가 100%, 92.5%, 96.1%로, 다른 학습 신경망들과 비교하여 좋은 성능을 나타내었다. 학습 및 검증 영상 크기가 원본 크기에 근접할 수록 등검은말벌집 검출 성능 증가 효과를 얻었다.

2. 고해상도 검증 영상 데이터에 대한 성능 평가

학습 영상 크기와 무관하게 검증 영상의 크기에 따라 검출 성능이 좋아질 수 있다고 생각할 수 있다. 따라서 Table 1에서 학습된 4개의 신경망을 이용하여 검증 영상 크기를 $3,840 \times 2,176$ 으로 고정한 후 비교 분석하였다. YOLO-v5s-3840을 제외한 3개의 신경망은 등검은말벌집 TP가 상승한 수치를 얻을 수 있었다. 하지만 오검출률도 같이 상승하여 정밀도와 F1 수치가 낮아졌다(Table 4).

검증 영상의 크기가 커질수록 더 높은 검출 성능을 얻을 수 있지만, 학습 영상 크기와 차이에 따라 오검출 또한 증가한다는 지표를 얻었다. 검증 영상 크기에 따른 성능

증가는 없었다.

기존 인공지능 학습에서 보편적으로 사용되는 영상 크기인 640은 지상을 기준으로 객체 탐지를 위해 설정된 크기이므로 항공 영상에서 작은 객체를 탐지하기엔 부족한 점이 있다. 그래서 드론으로 획득한 영상에서 등검은말벌집 객체 탐지를 위한 인공지능 학습은 기본보다 영상 크기를 높여 고해상도로 학습을 진행하는 것이 적합하다는 결론을 얻었다.

고 찰

본 연구에서는 생태교란 야생생물인 등검은말벌집을 탐색하기 위해 드론과 인공지능 알고리즘을 사용하였다. 높은 위치에 말벌집을 생성하는 등검은말벌의 특성으로 인해 육안으로는 발견하기 어렵다. 이러한 이유로, 드론을 사용하여 항공 영상을 취득하고, 인공지능 알고리즘을 통해 등검은말벌집을 검출하고자 하였다. 등검은말벌집 검



Fig. 7. False detection images of the three models.

Table 4. Validation table for the 3,840 image size

Model	Confusion matrix				Precision	Recall	F1 Score
	TP	FP	FN	TN			
(a)	192	8,094	37	3,002	2.3%	83.8%	4.4%
(b)	222	5,756	7	3,661	3.7%	96.9%	7.1%
(c)	223	2,247	6	5,087	9.0%	97.3%	16.4%
(d)	212	0	17	6,929	100%	92.5%	96.1%

(a) YOLO-v5s-default (640), (b) YOLO-v5s-1280, (c) YOLO-v5s-1920, (d) YOLO-v5s-3840

출을 위해 YOLO-v5를 사용하였으며 지상 영상과는 다르게 높이에 따라 객체들의 크기가 작게 변하는 항공 영상의 특성이 있다. 따라서 가장 적합한 인공지능 신경망을 얻기 위해 학습 영상 크기를 기본 설정인 640×384 와 $1,280 \times 736$, $1,920 \times 1,088$, $3,840 \times 2,176$ 의 크기로 학습을 진행하고 검출 성능을 비교 분석하였다. 인공지능 신경망 검증 시 신뢰도(Confidence) 0.7 이상일 때 가장 좋은 성능을 보인 YOLO-v5s-3840의 정밀도(Precision)는 100%, 재현율(Recall)은 92.5%, F1 수치(F1 Score)는 96.1%이다. 또한 학습 영상 크기와 무관하게 검증 영상 크기에 영향을 받는지 확인하고자 영상 크기를 3,840으로 고정하여 비교하였다. 검증 영상의 크기가 커질수록 검출과 오검출이 증가하는 지표를 얻었고 성능 증가 효과는 없었다. 항공 영상에서 작은 객체를 찾기 위해 고해상도 영상을 인공지능 학습에 사용하여 등검은말벌집을 찾을 수 있는 신경망을 얻을 수 있었다. 이를 활용하여 드론을 활용한 실시간 등검은말벌집 탐지, 제거를 위한 실시간 정보 공유, 데이터 수집을 통한 등검은말벌집 발생 위치 추정 등 추가적인 연구를 진행할 것이다.

적 요

2003년 중국 상하이로 통해 부산으로 유입된 생태교란 야생생물인 등검은말벌은 말벌집을 숲속 높은 위치에 생성하는 특성 때문에 육안으로 발견이 힘들며 매년 개체 수가 증가하여 양봉농가에 연간 1,700억 원 피해를 주고 있다. 이에 발견이 어려운 등검은말벌집을 쉽게 제거하여 양봉장 피해를 줄일 수 있도록 드론을 사용하여 영상을 획득하고 인공지능 기반으로 등검은말벌집 위치를 탐색하고자 한다. 지상 영상과는 다르게 항공 영상은 고도가 올라갈수록 객체가 매우 작아지는 특징을 가지기 때문에 드론에 장착하고 인공지능으로 등검은말벌집을 탐색할 수 있는 적합한 영상 크기 연구가 필요하다. 그래서 인공지능 모델 YOLO-v5에서 4개의 영상 크기(640×384 , $1,280 \times 736$, $1,920 \times 1,088$, $3,840 \times 2,176$)로 전처리하고 학습하였으며 성능을 비교하였다. 영상의 원본 크기는 $3,840 \times 2,160$ 이고 학습과 검증에는 동일한 영상 크기를 사용한다. 신뢰도(Confidence) 0.7 이상일 때 영상 크기 640×384 으로 학습된 YOLO-v5s-default (640)는 F1 수치(F1 Score) 2.4%, YOLO-v5s-1280은 36.5%, YOLO-

v5s-1920은 64.2%였고 YOLO-v5s-3840은 96.1%로 가장 좋은 검출 성능을 얻었다. 또한 다른 영상 크기로 학습한 4개의 인공지능에서 검증 영상 크기가 $3,840 \times 2,176$ 경우의 성능을 확인하였다. 이때 YOLO-v5s-default (640)는 F1 수치 4.4%, YOLO-v5s-1280은 7.1%, YOLO-v5s-1920은 16.4%로 성능증가는 없었다. 따라서 항공 영상에서 작은 객체 탐지를 위해서는 기존 인공지능 학습에서 보편적으로 사용되는 영상 크기 640보다 높은 해상도로 학습하고 검증하는 네트워크가 더 높은 성능 결과를 얻었다. 이를 활용하여 추가적으로 탐지 및 제거를 위한 드론 시스템, 실시간 정보 공유, 데이터 수집을 통한 등검은말벌집 발생 위치 추정 등의 연구를 진행할 것이다.

감사의 글

본 연구는 농촌진흥청(Rural Development Administration) 연구사업(PJ01497002)의 지원을 받아 수행되었습니다.

인용 문헌

- 김경철, 서다솜, 최인찬, 홍영기, 김국환, 권경도. 2021. 딥러닝 기반 등검은말벌 모니터링 시스템 개발. 한국산학기술학회 논문지 22(10): 31-36.
- 이철희, 정윤주, 김태호, 박재현, 박성빈, 정철의. 2019. 말벌 영상인식을 위한 심층 합성곱 신경망의 성능 평가. J. Apic. 34(3): 207-215.
- Abrol, D. P. 1994. Ecology, Behaviour and Management of Social Wasp, *Vespa velutina* Smith (Hymenoptera: Vespidae), Attacking Honeybee Colonies. J. Apic. 9(1): 5-10.
- Jocher, G. et al. 2020. YOLOv5, Assessed in <https://docs.ultralytics.com> on 18 May 2020.
- Jung, C. E. 2012. Initial Stage Risk Assessment of an Invasive Hymenoptera, *Vespa velutina nigrithorax* Buysson (Hymenoptera: Vespidae) in Korea. J. Apic. 27(2): 95-104.
- Jung, C. E., M. S. Kang, D. W. Kim and H.S. Lee. 2007a. Vespidae Wasps (Hymenoptera) Occurring around Apiaries in Andong, Korea I. Taxonomy and life history. J. Apic. 22(1): 53-62.
- Jung, C. E., M. S. Kang and D. W. Kim. 2007b. Vespidae Wasps (Hymenoptera) Occurring Around Apiaries in Andong, Korea II. Trap catches and seasonal dynamics. J. Apic. 22(1): 63-70.
- Kingma, D. P. and J. Ba. 2015. Adam: A Method for Stochastic Optimization. In Proceeding of International Conference on Learning Representations. Poster presentation.

- Redmon, J. and A. Farhadi. 2017. YOLO9000: Better, Faster, Stronger. In Proceeding of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7263-7271.
- Redmon, J. and A. Farhadi. 2018. YoloV3. Computer Vision and Pattern Recognition. Accessed in <https://doi.org/10.48550/arXiv.1804.02767> on 8 Apr 2018.
- Sim, H. S., M. L. Lee, Y. S. Choi, H. Y. Kim, I. P. Hong, S. O. Woo, K. U. Byeon and M. Y. Lee. 2014. Pattern of Emergence of *Vespa velutina nigrithorax* Buysson (Hymenoptera: Vespidae) on Spring in South Part of Korea. *J. Apic.* 29(4): 353-358.
- Stehman, S. V. 1997. Selecting and interpreting measures of thematic classification accuracy. *RSE* 62(1): 77-89.
- Zhu, X., S. Lyu, X. Wang and Q. Zhao. 2021. Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. *International Conference on Computer Vision*. pp. 2778-2788.