

해충 검출 및 카운팅을 위한 멀티스케일 CenterNet

이재현*, 손창환**, 이휘종***

Multiscale CenterNet for Pest Detection and Counting

Jae-Hyeon Lee*, Chang-Hwan Son**, and Hwijong Yi***

본 성과물은 농촌진흥청 연구사업(과제번호: PJ01630302)의 지원에 의해 이루어진 것임

요 약

해충의 조기 발견은 농작물의 피해를 감소하고 생산성 향상을 도모할 수 있다. 본 연구에서는 해충 발생량을 조기에 파악하고자 디지털 트랩에서 포획된 해충의 마릿수를 카운팅하는 기법을 제안하고자 한다. 디지털 트랩에서 포획된 해충의 마릿수는 가변적으로 변할 수 있다. 해충의 마릿수가 많은 경우에는 해충이 서로 인접하게 붙어 있거나 다른 해충에 의해 가려질 수 있다. 이는 해충 카운팅을 어렵게 하는 원인이다. 이러한 문제를 해결하기 위해서는 멀티스케일 기반의 객체 검출이나 객체 카운팅 기법이 필요하다. 제안한 방법에서는 기존의 단일스케일 기반 CenterNet 객체 검출 모델을 개선하여 멀티스케일 CenterNet 아키텍처를 새롭게 제안하고자 한다. 또한 제안한 멀티스케일 CenterNet을 변형하여 해충 카운팅을 위한 밀도 추정 네트워크도 소개하고자 한다. 실험 결과를 통해, 제안한 멀티스케일 CenterNet과 멀티스케일 밀도 추정 네트워크가 기존의 최첨단 해충 검출 및 해충 카운팅 모델보다 더 우수함을 확인하고자 한다.

Abstract

Early detection of pests can reduce crop damage and improve productivity. In this study, we propose a method of counting the number of pests captured by digital traps, in order to identify the amount of pests at an early stage. The number of pests captured by digital traps varies widely. In the case of a large number of pests, the pests may stick to each other or be obscured by other pests. This makes pest counting difficult. To solve this problem, a multiscale-based object detection or object counting technique is needed. In the proposed method, we try to propose a new multiscale CenterNet structure by improving the existing single-scale CenterNet object detection model. In addition, we introduce a density estimation network for pest counting by modifying the proposed multiscale CenterNet. Through the experimental results, it is confirmed that the proposed multiscale CenterNet and multiscale density estimation network are superior to the existing state-of-the-art pest detection and pest counting models.

Keywords

pest counting, deep learning, object detection, crowd counting

* 군산대학교 소프트웨어학과
- ORCID: <https://orcid.org/0000-0002-7655-9690>
** 군산대학교 소프트웨어학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0001-7077-3074>
*** 국립식량과학원 생산기술개발과
- ORCID: <https://orcid.org/0000-0002-7570-4585>

· Received: May 03, 2022, Revised: Jun. 23, 2021, Accepted: Jun. 26, 2021
· Corresponding Author: Chang-Hwan Son
Department of Software Science & Engineering, Republic of Korea
Tel.: 82+63-469-8915, Email: cson@kunsan.ac.kr

I. 서 론

스마트팜은 정보통신 기술을 활용해 작물의 생육 환경을 관측하고 최적의 상태로 관리하는 과학 기반의 농업 방식을 말한다. 인공지능 및 빅데이터 기술과 결합해 최적화된 생산·관리를 위한 데이터 기반의 의사결정이 가능하다. 또한 최적화된 생육환경을 제공하여 수확 시기와 수확량을 예측하고 작물의 품질과 생산량을 한층 더 높일 수 있다[1].

해충은 작물의 생육 과정에서 잎, 줄기, 뿌리 등을 갉아 먹거나 즙액을 빨아들여 농작물에 직접적인 해를 가하여 농작물의 품질 저하와 생산량 저하를 수반한다. 따라서 해충의 피해를 줄이기 위해서는 해충 방제를 조기에 실시할 필요가 있다. 기존의 해충 방제는 해충 방제 전문가가 해충의 종류, 출몰 지역, 해충의 예상 경로 및 서식지를 조사하여 해충 방제를 실시한다. 하지만 해충의 발생 시기와 장소는 해마다 변동되고 또한 해외에서 유입되는 외래 해충 때문에 해충 발생량을 파악하기가 힘들다.

따라서 본 논문에서는 주광성 기반의 디지털 트랩을 사용해서 해충의 발생량을 파악하고자 한다. 디지털 트랩에 유입된 해충을 카메라 센서로 촬영한 후, 해충 카운팅 딥러닝 모델을 적용하여 해충의 발생량을 예측하고자 한다. 기존의 해충 카운팅 딥러닝 모델은 밀도 맵(Density map) 추정 기반과 바운딩 박스 검출 기반으로 나뉠 수 있다. 원래 밀도 맵 기반 추정 기법은 군중 카운팅(Crowd counting)을 위한 목적으로 개발되었다. 디지털 트랩에서 포획된 해충의 개수가 많은 경우에는 밀도 맵 기반의 군중 카운팅 방식이 효과적일 수 있다. 그리고 바운딩 박스 기반 검출 방식은 기존의 객체 검출(Object detection)을 말한다. 디지털 트랩에서 포획된 해충을 하드웨어 제어를 통해 접착면에 떨어지는 해충의 마릿수를 제어할 수 있다면, 기존의 객체 검출 방식이 더 효과적일 수 있다.

본 연구에서는 밀도 맵 추정 기반과 바운딩 박스 검출 기반에 모두 적용 가능한 멀티스케일 기반 CenterNet 아키텍처를 제안하고자 한다. CenterNet[2]은 바운딩 박스를 바로 검출하는 기존의 YOLO[3]나 RCNN[4] 객체 검출과는 달리, 객체의 센터에만

흰색 밝기 값을 갖는 히트맵(Heatmap)을 찾아서 최종 바운딩 박스를 검출하는 방식을 채택하고 있다. 여기서 사용된 히트맵의 생성 과정은 군중 카운팅에 사용되는 밀도 맵과 유사한 의미를 가진다. 따라서 본 연구에서는 이러한 점을 착안해서 밀도 맵 추정 기반과 객체 검출 기반에 모두 활용될 수 있는 멀티스케일 기반 CenterNet을 제안하고자 한다.

제안한 멀티스케일 CenterNet과 기존 CenterNet과의 차이는 멀티스케일 특징 추출 여부에 있다. 기존의 CenterNet은 특징 추출을 위해 Stacked hourglass [5]를 사용하여 단일스케일의 히트맵만 추정한다. 하지만 해충 영상의 경우, 해충의 개수가 시기별로 달라질 수 있기 때문에 단일스케일 기반 히트맵 추정 방식은 정확도에서 떨어질 수 있다. 따라서 포획된 해충의 마릿수에 유연하게 대처할 수 있는 멀티스케일 기반의 객체 검출 또는 객체 카운팅 기술이 요구된다. 본 연구에서는 멀티스케일 기반 CenterNet 아키텍처를 새롭게 제시하고자 한다. 또한 실험 결과를 통해 제안한 해충 카운팅 딥러닝 모델이 기존의 객체 검출과 카운팅 기술보다 정확도 측면에서 더 우수함을 검증하고자 한다.

II. 관련 연구

2.1 밀도 맵 추정 기반 해충 카운팅

기존의 군중 영상(Crowd images)은 객체가 너무 인접하거나 객체가 다른 객체를 가리는 폐색(Occlusion) 문제 때문에 바운딩 박스 기반의 객체 검출 방식은 군중 카운팅에는 적합하지 않다. 이를 위해, 밀도 맵 기반의 군중 카운팅 기술이 개발되었다. 고전적인 군중 카운팅 기술은 특징 추출과 회귀 모델 파라미터 학습으로 이루어진다[6]. 최근에는 딥러닝 모델을 사용해서 입력 영상에서 군중의 개수를 바로 추정하는 방식이 대세를 이루고 있다.

대표적인 군중 카운팅 모델로는 CrowdNet[7], MCNN(Multi-Column Convolutional Neural Networks) [8], ICCNet(Iterative Crowd Counting Network)[9], SANet(Scale Aggregation Network)[10] 등이 있다. 따라서 기존의 군중 카운팅 기법은 해충카운팅에도 그대로 적용될 수 있다.

그림 1은 MCNN 모델의 아키텍처를 보여주고 있다. MCNN은 객체 스케일 문제를 해결하기 위해, 필터의 크기가 서로 다른 3개의 합성곱 신경망(Convolutional neural networks)을 사용한다. 그림에서 보듯이, 입력 영상은 3개의 합성곱 신경망을 거쳐 최종 밀도 맵을 추정한다. 각 서브네트워크의 출력 맵은 연결 레이어(Concatenation layer)를 거쳐 퓨전된다. 여기서 밀도 맵은 각 객체의 중심점에 가장 밝은 픽셀 값을 저장한다. 흰색 밝기의 픽셀 값은 1로 정규화 되어 있다고 가정한다. 따라서 밀도 맵의 전체 픽셀 값을 합하면 객체의 총 개수가 된다.

학습 단계에서는 추정된 밀도 맵과 정답 밀도 맵과의 오차를 계산하기 위해 평균 제곱 오차(Mean squared error) 손실함수를 사용한다. 테스트 단계에서는 입력 영상을 학습된 네트워크를 거쳐 밀도 맵을 추정한 후, 전체 픽셀 값을 더함으로써 최종 객체 개수를 추정한다.

그림 1의 맨 우측에는 밀도 맵의 예시를 보여준다. 밀도 맵은 실제 그레이 영상이지만 시각화를 위해 칼라 영상으로 표시하였다. 그리고 밀도 맵을 생성할 때, 흰색 밝기를 갖는 픽셀에서 가우시안 커널을 한번 적용한다. 따라서 밀도 맵 생성 과정은 수식적으로 다음과 같이 표현된다[8].

$$Y = \sum_{x_i \in S} \delta(x - x_i) * G_\sigma \tag{1}$$

여기서 Y 는 밀도 맵이고 G_σ 는 가우시안 커널이다. S 는 객체의 중심 위치인 픽셀 인덱스(x_i)를 저장하고 있는 집합이다. δ 는 델타 함수(Delta function)로써 객체 중심 위치에서만 1의 값을 갖는다. 즉, 픽셀 인덱스인 x 가 x_i 가 되는 위치에서만 1의 값을 가진다. 그리고 가우시안 커널의 합은 1이기 때문에 Y 의 픽셀 값을 모두 더하면 객체의 개수가 된다.

2.2 바운딩 박스 검출 기반 해충 카운팅

해충 카운팅을 위한 또 다른 접근 방식은 기존의 객체 검출을 그대로 적용하는 것이다. 객체 검출은 입력 영상에서 객체가 위치한 곳의 바운딩 박스를 찾는 기법을 말한다[3][4]. 총 검출된 바운딩 박스의 개수가 바로 영상 내의 객체의 개수와 동일하다. 그림 2는 바운딩 박스 개수를 추정함으로써 영상 내의 객체의 개수를 추정하는 방식을 보여 주고 있다.

전통적인 객체 검출 기법은 슬라이딩 윈도우(Sliding window) 방식과 영역 제안(Region proposal) 방식으로 나뉠 수 있다[11].

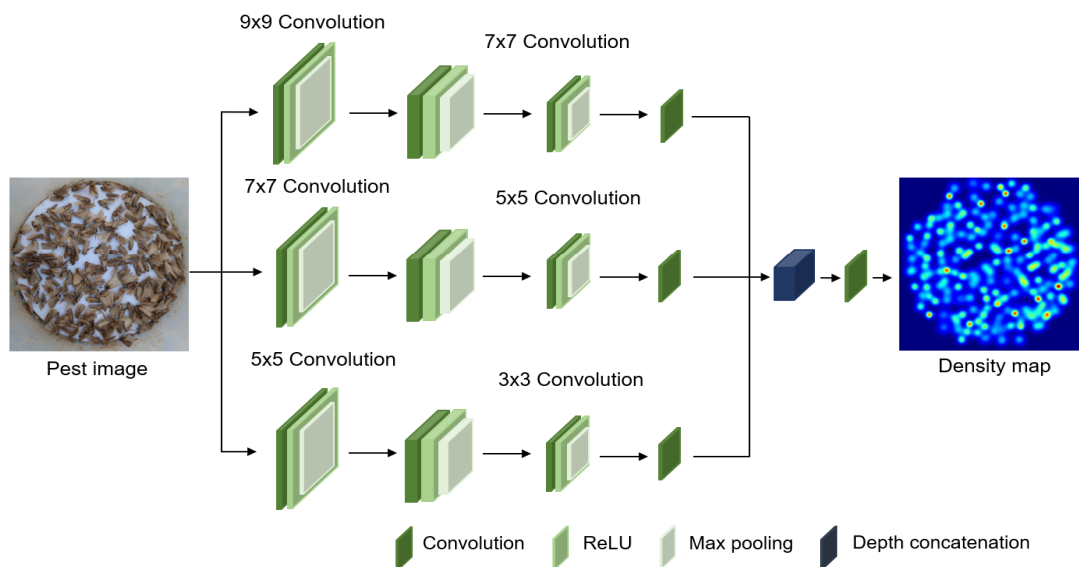


그림 1. 밀도 맵 추정 기반 해충 카운팅
Fig. 1. Pest counting based on density map prediction

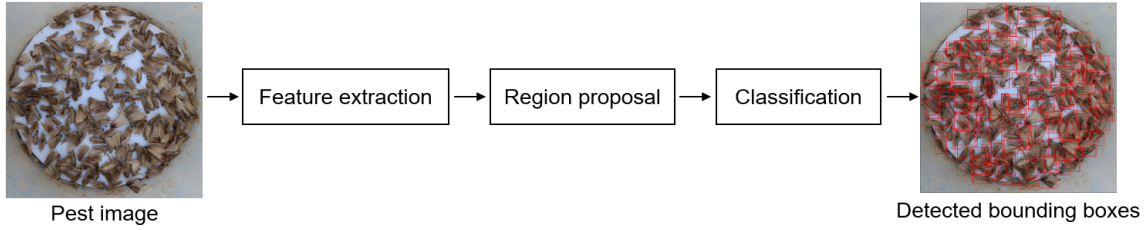


그림 2. 바운딩 박스 검출 기반 해충 카운팅
Fig. 2. Pest counting based on bounding box detection

하지만, 최근에는 딥러닝 기반의 객체 검출 기법이 주류가 되었다. 딥러닝 기반 객체 검출은 크게 RCNN 계열[4]과 YOLO 계열[3]로 구분될 수 있다. RCNN 계열과 YOLO 계열의 가장 큰 차이는 영역 제안과 분류 단계의 이원화 구조에 있다. RCNN 계열은 영역 제안과 분류기 학습이라는 두 단계로 구성되는 반면 YOLO 계열은 한 단계로 객체 검출이 이루어진다. 따라서 RCNN 계열은 영역 제안과 분류기 학습을 세분화하여 객체 검출을 진행하기 때문에 상대적으로 정확도가 높은 편이다. 반면 YOLO 계열은 RCNN 계열보다는 속도가 빠르다는 장점이 있다. 또한 검출 성능도 RCNN 계열과 견줄만하다.

최근에는 검출 속도와 정확도를 동시에 향상할 수 있는 CenterNet 객체 검출 모델이 제안되었다[2]. 특히 CenterNet은 기존의 객체 검출 기법과는 달리, 히트맵 기반으로 바운딩 박스를 예측한다. 따라서, CenterNet은 기존의 밀도 맵 추정 기반의 군중 카운팅 기술로의 변형이 가능하다. 즉, CenterNet의 히트맵 추정 부분을 개선한다면 해충 검출뿐만 아니라 해충 카운팅까지도 적용 가능하다. 이러한 점에 착안하여 본 연구에서는 해충 카운팅을 위한 객체 검출 모델로써 CenterNet을 기반 모델로 활용하고자 한다. 하지만, 기존의 CenterNet에서 히트맵 추정 모듈은 단일스케일 기반의 Stacked hourglass[5]로 구성되어 있다. 이러한 단일스케일 기반 Stacked hourglass는 해충의 개수가 가변적으로 변할 때 스케일에 강인한 히트맵 추정을 보장할 수 없다. 따라서 영상에서 해충의 개수에 따라 스케일에 강인한 멀티스케일 CenterNet 개발이 필요하다.

III. CenterNet 모델

3.1 CenterNet 학습 및 추론 과정

CenterNet은 객체의 중심점을 저장하고 있는 히트맵을 활용하여 바운딩 박스를 예측한다.

CenterNet은 너비(W)와 높이(H)를 갖는 입력 영상이 주어지면 0과 1의 값을 갖는 히트맵을 추정한다.

$$I \in R^{W \times H \times 3} \quad (2)$$

$$\hat{Y} \in [0, 1]^{R \times \frac{W}{R} \times \frac{H}{R} \times C} \quad (3)$$

여기서 R 은 스트라이드(Stride)를 의미하고 출력 영상의 스케일을 결정한다. 원문에서는 4로 설정하여 원본 크기의 1/4 크기의 히트맵을 추정한다. C 는 객체 부류(Class)에 대한 개수이다. 히트맵에서는 키포인트(Keypoint), 즉 객체의 중심점에 해당하는 위치에서는 1의 값을 가지며 나머지는 0의 값을 가진다. 즉, $\hat{Y}_{x,y,c} = 1$ 이며 배경은 $\hat{Y}_{x,y,c} = 0$ 이다. 객체의 중심점은 바운딩 박스의 중심에 대응한다. 참고로 실제 히트맵을 생성할 때는 1과 0으로 구성된 \hat{Y} 를 사용하는 것이 아니라 가우시안 커널을 적용한 블러링된 히트맵을 사용한다.

\hat{Y} 를 예측하기 위해, 클래스 불균형 문제를 해소할 수 있는 로지스틱 회귀(Logistic regression) 기반의 초점 손실(Focal loss)을 사용한다.

$$L_k = \frac{-1}{N} \sum_{xyc} \begin{cases} (1 - \hat{Y}_{xyc})^\alpha \log(\hat{Y}_{xyc}) & \text{if } Y_{xyc} = 1 \\ (1 - \hat{Y}_{xyc})^\beta (\hat{Y}_{xyc})^\alpha & \text{otherwise} \end{cases} \quad (4)$$

여기서 N 은 입력 영상에서 객체 키포인트의 개수이다. 그리고 α 와 β 는 하이퍼 파라미터이고 각각 2와 4로 설정되었다. 식 (4)에서 $Y_{x,y,c} = 1$ 인 경우는, 추정된 $\hat{Y}_{x,y,c}$ 가 1에 가까워야 손실이 작아짐을 알 수 있다. 반면에 $Y_{x,y,c} \neq 1$ 경우는 $\hat{Y}_{x,y,c}$ 가 0

에 근접해야 됨을 알 수 있다.

식 (2)-(4)는 CenterNet을 통해 객체의 키포인트와 부류 값을 저장한 \hat{Y} 만을 추정할 수 있다. 따라서 최종 바운딩 박스를 검출하기 위해서는 바운딩 박스의 가로와 세로 값을 저장한 바운딩 박스 맵 (Bounding box map)을 추정할 필요가 있다.

$$L_s = \frac{1}{N} \sum_{k=1}^N \left| \hat{S}_{p_k} - s_k \right| \quad (5)$$

여기서 $s_k = (w_k, h_k)$ 는 k 번째 해층의 바운딩 박스의 가로와 세로 크기를 저장하고 있다. \hat{S}_{p_k} 는 키포인트 위치(p_k)에 대한 추정된 바운딩 박스 맵을 의미한다. \hat{S}_{p_k} 의 크기는 \hat{Y} 의 가로 및 세로 크기만 동일하고 채널의 개수는 2개이다. CenterNet에서는 L_k 와 L_s 손실외에 오프셋 맵(Offset map)을 추정하기 위한 손실이 추가적으로 사용된다. 식 (3)에서 키포인트의 좌표 값도 R 만큼 감소함을 알 수 있다.

$$\tilde{p} = \left\lfloor \frac{p}{R} \right\rfloor \quad (6)$$

여기서 p 는 원본 영상의 키포인트 값을 의미하고 $\lfloor \cdot \rfloor$ 는 정수 값으로 변환하는 연산자이다. 디지털화 과정 때문에 키포인트 값이 정수 값으로 변환되므로 이에 대한 보정이 필요하다.

$$L_{off} = \frac{1}{N} \sum_p \left| \hat{O}_{\tilde{p}} - \left(\frac{p}{R} - \tilde{p} \right) \right| \quad (7)$$

여기서 $\hat{O}_{\tilde{p}}$ 는 디지털화 과정으로 발생한 오프셋 값을 보정하기 위한 맵이다. 즉, 실수 값의 원본 키포인트인 p/R 과 추정된 정수 값의 키포인트인 \tilde{p} 와의 오차를 저장하고 있다. 따라서 최종 손실은 L_k, L_s, L_{off} 이며 다음과 같다.

$$L = L_k + \lambda_s L_s + \lambda_{off} L_{off} \quad (8)$$

여기서 λ_s, λ_{off} 는 손실 함수의 가중치 값이며 각각 1, 0.1로 설정되었으며 내리막경사법과 같은 최적화 기법을 사용해서 네트워크의 파라미터를 갱신한다.

테스트 단계에서는 CenterNet을 통해 추정된 히트맵에서 3×3 최대 풀링(Max pooling) 적용하여 키포인트를 추출한다. 이 키포인트 값에 추정된 오프셋 맵과 바운딩 박스 맵의 정보를 활용해서 최종 바운딩 박스를 검출한다.

3.2 CenterNet 모델의 아키텍처

그림 3은 CenterNet의 아키텍처를 보여 준다. CenterNet은 크게 합성곱 레이어(Convolution : Convolution layer), 배치 정규화 레이어(Batch normalization : Batch normalization layer), 정류 선형 유닛 레이어(ReLU : Rectified linear unit layer), 잔차 블록(Residual block), 원소 합(Elemental-wise sum), Stacked hourglass로 구성된다. 여기서 Stacked hourglass는 Hourglass를 쌓은 구조를 말한다. Hourglass 모듈은 대칭 구조를 가지며 잔차블록으로 구성된다. 자세한 Hourglass 구조는 참고문헌 [5]에 기술되어 있다.

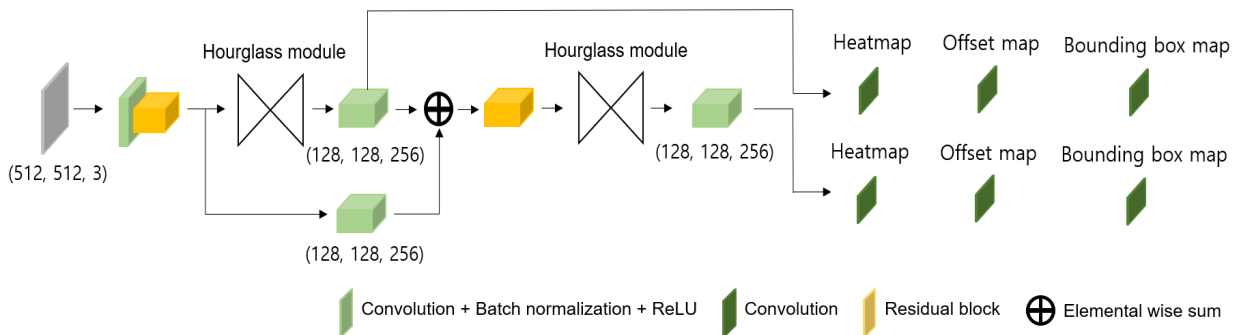


그림 3. 객체 검출을 위한 CenterNet 아키텍처
Fig. 3. Architecture of the CenterNet for object detection

CenterNet은 512×512 입력 영상을 받아 합성곱 계층과 잔차 블록을 통해 128×128 크기의 특징 맵을 생성한다. 그리고 첫 번째 Hourglass 네트워크를 통과한 후, 128×128 크기의 특징 맵을 출력한다. 그리고 최종 히트맵, 읍셋 맵 그리고 바운딩 박스 맵을 추정하기 위해, 1×1 필터 크기로 합성곱 연산을 적용한다. 두 번째 Hourglass도 첫 번째 Hourglass와 동일한 크기의 해상도를 갖는 출력 맵을 생성한다. 그리고 이 출력 맵을 활용하여 최종 히트맵, 읍셋 맵, 바운딩 박스 맵을 추정한다.

CenterNet은 그림 3에서 보듯이, 단일스케일의 히트맵, 읍셋 맵, 바운딩 박스 맵을 추정한다. 입력 영상에서 해충의 개수가 거의 동일한 경우에는 단일스케일 기반의 해충 검출이 잘 동작할 수 있다. 하지만, 해충의 개수가 가변 될 경우에는 오검출률이 높아질 수 있다. 이런 문제를 보완하기 위해서는 멀티스케일 기반의 CenterNet이 필요가 있다.

IV. 제안한 멀티스케일 CenterNet

4.1 제안한 접근 방법

기존의 CenterNet은 단일 스케일의 특징 맵을 추출하기 때문에 입력 영상에서 해충의 개수가 많을 경우 오검출될 수 있다. 그림 4는 디지털 트랩에서 실제로 촬영된 해충 영상을 보여 주고 있다. 해충의 개수가 많은 경우 해충이 인접하게 붙어 있거나 다른 해충에 의해 가려질 수 있다. 이런 경우, 출력 맵의 해상도가 고정되기 때문에 객체 검출의 정확

도가 떨어질 수 있다. 따라서 본 논문에서는 멀티스케일 특징 맵을 추출하여 최종 해충 카운팅의 성능을 개선할 수 있는 멀티스케일 CenterNet을 구현하고자 한다.

4.2 제안한 멀티스케일 CenterNet

그림 5는 제안한 멀티스케일 CenterNet의 아키텍처를 보여 주고 있다. 그림 5와 그림 3의 가장 큰 차이점은 멀티스케일 적용 여부이다. 제안한 네트워크에서는 멀티스케일을 실현하기 위해 첫 번째 Hourglass와 두 번째 Hourglass의 출력 맵의 해상도를 다르게 설정하였다. 즉, 첫 번째 Hourglass의 특징 맵은 128×128 의 크기를, 두 번째 Hourglass의 특징 맵은 256×256 의 크기를 갖는다.

또한, 첫 번째 Hourglass의 출력 맵으로부터 추정된 바운딩 박스 맵, 읍셋 맵, 히트맵을 업샘플링(Upsampling)한 후, 두 번째 Hourglass의 입력으로 활용한다. 이는 첫 번째에서 예측된 저해상도의 결과를 두 번째 Hourglass의 고해상도 출력 예측에 활용함으로써, 히트맵, 바운딩 박스, 읍셋의 정확도를 제고할 수 있다.

또한, 두 번째 Hourglass의 입력은 기존의 CenterNet과는 달리, 3종류의 특징 맵을 사용한다. 즉, 첫 번째 Hourglass의 출력 맵과 업샘플링된 히트맵, 바운딩 박스, 읍셋, 그리고 입력 영상에서 추출된 특징 맵을 퓨전해서 입력으로 사용한다. 퓨전을 위해 연결 레이어(Concatenation layer)를 사용하였다.



그림 4. 디지털 트랩에서 포획된 해충 영상 예시
Fig. 4. Example of pest images captured by digital trap

기존의 CenterNet은 퓨전을 위해 원소 합을 사용했지만, 3종류의 특징 맵은 서로 다른 정보를 담기 때문에 연결 레이어를 선택하였다. 3종류의 특징 퓨전은 추출된 특징의 정보력과 구별력을 높일 수 있다. 특히 추정된 히트맵, 오프셋 맵, 바운딩 박스의 재 활용은 멀티스케일 특징 설계를 구현했다고 할 수 있다.

4.3 해충 카운팅을 위한 제안한 밀도 추정 기반 멀티스케일 CenterNet 변형

해충 카운팅을 위한 접근 방법으로는 바운딩 박스 검출 기반과 밀도 맵 추정 기반이 있다. 그림 5는 바운딩 박스 검출 기반의 해충 카운팅 기법에 해당한다. 만약 해충이 군중 카운팅처럼 디지털 트랩에 가득 채워진다면 바운딩 박스 기반의 해충 검출은 적합지 않을 수 있다. 따라서 밀도 맵 기반의 해충 카운팅 방식도 필요하다.

제안한 멀티스케일 CenterNet은 히트맵을 추정한다. 히트맵은 해충이 존재하는 객체의 바운딩 박스의 중심점을 저장하기 때문에 거의 밀도 맵의 개념과 유사하다. 이러한 점을 착안하여 제안한 멀티스케일 CenterNet을 밀도 맵 추정 기반의 해충 카운팅 모델로 수정 가능하다. 즉, 제안한 멀티스케일 CenterNet에서 히트맵을 밀도 맵으로 수정하고 오프셋 맵과 바운딩 박스 맵만 제거하면 된다. 히트맵과 밀도 맵의 차이는 단지 가우시안 커널의 중심 값을 1로 설정하는 것에 따라 좌우된다.

그림 6은 제안한 밀도 맵 추정 기반의 해충 카운팅을 위한 멀티스케일 CenterNet의 변형이다. 그림 5와 6의 차이점은 그림 5는 최종 바운딩 박스 예측이기 때문에 바운딩 박스와 오프셋 예측이 필요하지만 그림 6은 최종 밀도 맵을 예측하기 때문에 바운딩 박스 맵과 오프셋 맵이 필요 없다. 나머지는 멀티스케일 특징 추출에 해당하는 부분이므로 동일하다.

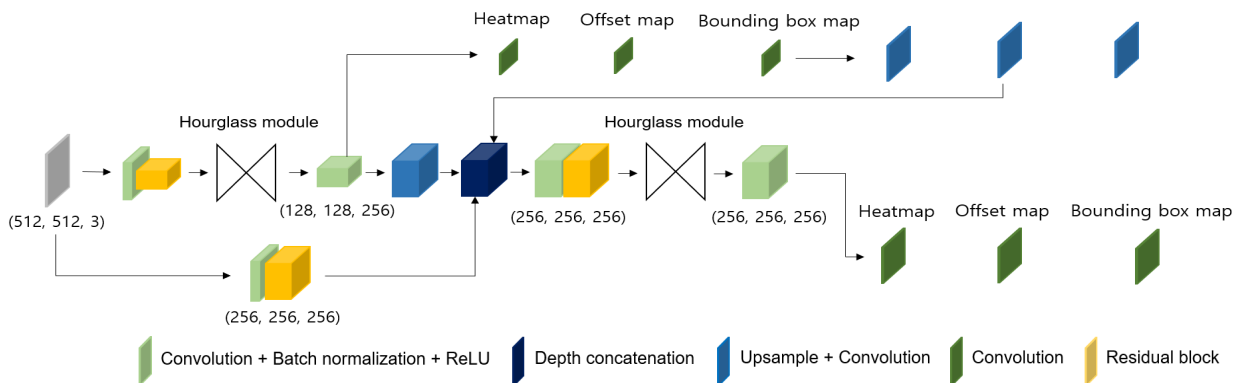


그림 5. 해충 검출을 위한 제안한 멀티스케일 CenterNet
Fig. 5. Proposed multiscale centerNet for pest detection

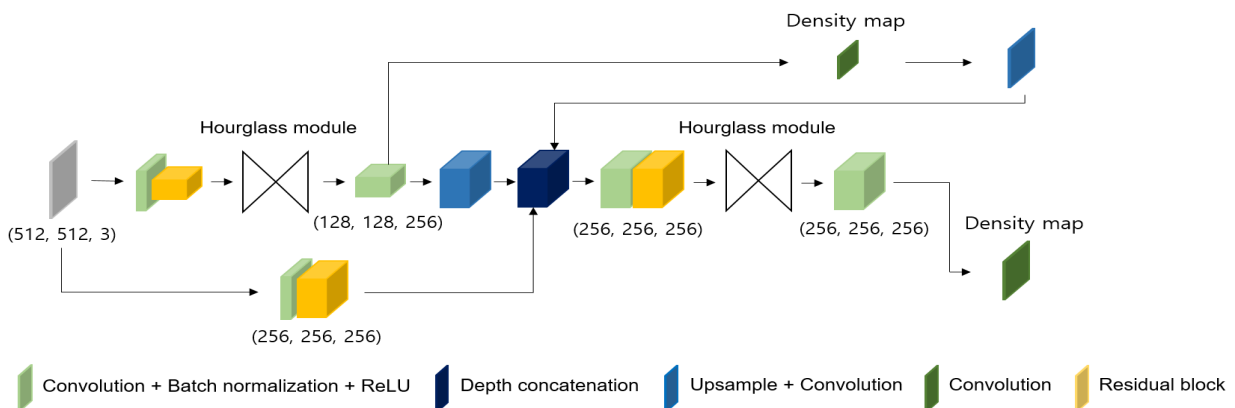


그림 6. 해충 카운팅을 위한 제안한 멀티스케일 밀도 추정 네트워크
Fig. 6. Proposed multiscale density estimation network for pest counting

그림 6에서 추정된 밀도 맵의 오차를 계산하기 위해 다음과 같은 손실함수를 정의한다.

$$L = \sum_{i=1}^N \|\widehat{M}_i - M_i\|^2 + \lambda(1 - CS(\widehat{m}_i', m_i')) \quad (9)$$

손실함수는 두 개의 항목으로 구성된다. 첫 번째 항목은 밀도 맵 간의 유사도를 측정하고 두 번째 항목은 벡터 방향에 대한 유사도를 계산한다. 수식에서 \widehat{M}_i 은 추정된 밀도 맵이며 M_i 는 정답 밀도 맵이다. \widehat{m}_i' 와 m_i' 는 \widehat{M}_i 과 M_i 행렬을 벡터로 표현한 것이며 CS 는 코사인 유사도를 의미한다. 그리고 λ 는 코사인 유사도 손실에 대한 가중치이며 이 연구에서는 10으로 설정되었다.

V. 실험 및 결과

5.1 실험 환경

학습 데이터는 디지털 트랩에서 촬영한 2501장을

사용했다. 해충의 종류는 파밤나방이며 훈련 집합과 테스트 집합은 7:3의 비율로 랜덤하게 나누었다. 최적화 기법은 아담 옵티마이저[12](Adam optimizer)를 사용했으며 배치 크기는 6, 에폭은 100회, 학습률은 0.0001로 설정하였다.

소스코드는 PyTorch 프레임워크로 작성되었고 윈도우 운영체제에서 실행되었다. 제안한 모델의 성능을 평가하기 위해, 두 가지 실험을 수행했다. 하나는 기존의 객체 검출 모델의 성능을 평가하였고 다른 하나는 기존의 군중 카운팅 모델과의 성능을 평가했다. 객체 검출을 위한 비교 대상으로는 RetinaNet[13], RepPoints[14], CenterNet[2], Faster RCNN[15]을, 군중 카운팅 모델로는 MCNN[8], ICCNet[9], SANet[10], Stacked hourglass[5]를 선택했다.

5.2 바운딩 박스 및 밀도 맵 추정 결과

그림 7은 제안한 멀티스케일 CenterNet 및 멀티스케일 밀도 추정 네트워크를 사용해서 추정된 바운딩 박스와 밀도 맵을 보여준다.




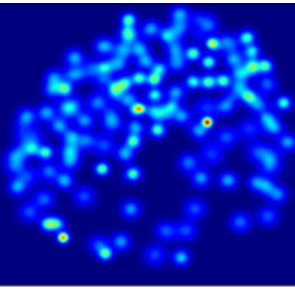
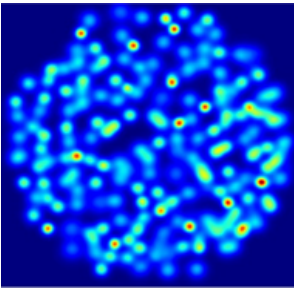
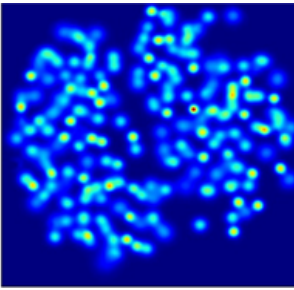
측정된 해충 개수: 124 Measured number of pests: 124	측정된 해충 개수: 192 Measured number of pests: 192	측정된 해충 개수: 197 Measured number of pests: 197
		
(a) 예측된 해충 개수: 116 (a) Predicted number of pests: 116	(b) 예측된 해충 개수: 181 (b) Predicted number of pests: 181	(c) 예측된 해충 개수: 189 (c) Predicted number of pests: 189
		
(d) 예측된 해충 개수: 124.04 (d) Predicted number of pests: 124.04	(e) 예측된 해충 개수: 185.94 (e) Predicted number of pests: 185.94	(f) 예측된 해충 개수: 180.67 (f) Predicted number of pests: 180.67

그림 7. 실험 결과 : 추정된 바운딩 박스(a,b,c), 추정된 밀도 맵(d,e,f)
Fig. 7. Experimental results : Estimated bounding boxes(a,b,c), Estimated density maps(d,e,f)

그리고 각각의 결과 그림 위에는 실제 해충의 개수가 적혀 있다. 정답 값인 해충의 개수는 육안으로 카운팅 되었다. 추정된 밀도 맵에서 파란색은 배경이고 노란색과 빨간색은 해충이 존재하는 곳이다.

그림 7에서 해충이 인접하게 붙어 있을지라도 해충의 개수가 비교적 정확하게 추정된 것을 알 수 있다. 이전 절에서 언급했지만, 해충 카운팅은 두 가지 방법으로 접근할 수 있다. 즉, 바운딩 박스 검출을 통한 해충 카운팅과 밀도 맵 추정을 통한 해충 카운팅이 있다. 본 연구에서는 수집된 해충 데이터 셋에 대해, 그림 5와 6에 제시된 멀티스케일 CenterNet과 멀티스케일 밀도 추정 네트워크가 둘 다 우수한 결과를 얻을 수 있었다.

5.3 정량적 평가

해충 카운팅을 위한 평가 척도(MAE, Mean Absolute Error)로 평균 절대 오차를 사용했다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (10)$$

여기서 n 은 테스트 영상의 개수이며 y_i 는 측정된 해충의 개수이고 y'_i 은 모델에 의해 추정된 해충의 개수이다.

표 1은 두 가지 접근 방식에 대한 MAE 결과를 보여준다. 하나는 객체 검출 기반의 해충 카운팅이고 다른 하나는 밀도 추정 기반의 해충 카운팅이다. 결과에서 볼 수 있듯이, 제안한 멀티스케일 CenterNet과 멀티스케일 밀도 추정 네트워크의 MAE 값이 가장 작은 것을 알 수 있다. 즉, 평가 대상에서 성능이 가장 우수하며 정답 값과 예측 값의 오차가 약 1.1정도로 우수한 성능을 달성한 것을 알 수 있다.

그리고 제안한 멀티스케일 CenterNet이 제안한 밀도 추정 네트워크보다 성능이 약간 더 우수한 것을 볼 수 있다. 이는 테스트 영상에 대해서 바운딩 박스가 잘 작동될 수 있는 허용 범위를 말해 준다. 만약 트랩에서 포획된 해충의 마릿수가 더 증가하면 멀티스케일 밀도 추정 네트워크가 더 효과적으

로 동작할 수 있다. 향후 더 많은 해충 데이터가 수집되면 이 두 가지 접근 방식의 장단점을 비교 분석하고자 한다.

표 1. 해충 카운팅을 위한 정량적 평가

Table 1. Quantitative evaluation for pest counting

Methods	Models	MAE
Density-based pest counting	MCNN [8]	2.924
	ICCNNet [9]	2.233
	SANet [10]	1.397
	Stacked hourglass [5]	1.904
	Proposed multiscale density network	1.170
Detection-based pest counting	RetinaNet [13]	5.885
	Faster RCNN [15]	6.230
	RepPoints [14]	2.039
	CenterNet [2]	1.435
	Proposed multiscale centernet	1.049

그리고 제안한 기법은 기존의 CenterNet 모델을 멀티스케일로 변경한 개선된 모델이다. 표 1에서 기존의 CenterNet 모델과 비교했을 때, 해충 카운팅의 성능이 개선됨을 알 수 있다. 이는 제안한 멀티스케일 특징 설계가 효과적임을 말해준다. 즉, 멀티스케일 접근 방식은 저해상도의 결과를 활용해서 고해상도의 결과를 예측하기 때문에 히트맵 또는 밀도 맵의 정확도를 개선할 수 있다.

VI. 결 론

본 논문에서는 해충 카운팅을 위한 멀티스케일 CenterNet을 제안하였다. 디지털 트랩에서 실제 포획된 영상을 대상으로 실험을 하였고 제안한 기법이 기존의 최신 객체 검출 및 군중 카운팅 기법보다 더 우수한 성능을 달성하였다. 특히, CenterNet에 멀티스케일 특징 설계를 구현함으로써, 저해상도 결과를 고해상도의 결과 예측에 활용함으로써 해충 카운팅의 정확도를 개선할 수 있었다. 향후 연구로는 해충의 마릿수가 더 많은 경우에 대해 제안한 멀티스케일 CenterNet과 멀티스케일 밀도 추정 네트워크의 성능을 비교 분석해 볼 계획이다. 그리고 수집된 날짜와 해충 개수와의 연관성을 모델링할 수 있는 멀티모달 CenterNet 모델도 개발할 계획이다.

References

- [1] H. J. Yu, D. H. Lee, and C. H. Son, "Apple leaf disease identification through region-of-interest-aware deep convolutional neural network", *Journal of Imaging Science and Technology*, Vol. 64, No. 2, pp. 20507-1-20507-10, Jan. 2020. <https://doi.org/10.48550/arXiv.1903.10356>.
- [2] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points", *arXiv:1904.07850v2 [cs.CV]*, Apr. 2019. <https://doi.org/10.48550/arXiv.1904.07850>.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, realtime object detection", in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 779-788, Jun. 2016. <https://doi.org/10.48550/arXiv.1506.02640>.
- [4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA, pp. 580-587, Jun. 2014.
- [5] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation", in *Proc. European Conference on Computer Vision*, Amsterdam, Netherlands, pp. 483-499, Sep. 2016.
- [6] K. Chen, C. C. Loy, S. Gonh, and T. Xiang, "Feature mining for localized crowd counting", in *Proc. British Machine Vision Conference*, Jan. 2012.
- [7] L. Boominathan, S. S. Kruthiventi, and R. V. Babu, "CrowdNet: A deep convolution network for dense crowd counting", in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 640-644, Aug. 2016. <https://doi.org/10.1145/2964284.2967300>.
- [8] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network" in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 589-597, Jun. 2016.
- [9] V. Ranjan, Hieu. Le, and M. Hoai, "Iterative crowd counting", in *Proc. European Conference on Computer Vision*, Munich, Germany, pp. 270-285, Oct. 2018.
- [10] X. Cao, Z. Wang, Y. Zhao, and F. Su, "Scale aggregation network for accurate and efficient crowd counting", in *Proc. European Conference on Computer Vision*, Munich, Germany, pp. 734-750, Oct. 2018.
- [11] T. M. Choi, C. H. Son, and D. Lee, "Attention-enhanced YOLO model via region of interest feature extraction for leaf diseases detection", *The Journal of Korean Institute of Information Technology*, Vol. 19, No. 4, pp. 83-93, Apr. 2021.
- [12] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization", in *Proc. International Conference on Learning Representations*, San Diego, USA, May 2015. <https://doi.org/10.48550/arXiv.1412.6980>.
- [13] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection" in *Proc. IEEE International Conference on Computer Vision*, pp. 2999-3007, Oct. 2017.
- [14] Z. Yang, S. Liu, H. Hu, L. Wang, and S. Lin, "RepPoints: Point set representation for object detection," in *Proc. IEEE International Conference on Computer Vision*, pp. 9657-9664, Aug. 2019.
- [15] Sh. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137-1149, Jun. 2017.

저자소개

이 재 현 (Jae-Hyeon Lee)



2017년 3월 ~ 현재 : 군산대학교
소프트웨어학과 학사과정
관심분야 : 컴퓨터 비전, 영상처리,
딥 러닝

손 창 환 (Chang-Hwan Son)



2002년 2월 : 경북대학교
전자전기공학과(공학사)
2004년 2월 : 경북대학교
전자공학과(공학석사)
2008년 8월 : 경북대학교
전자공학과(공학박사)
2017년 4월 ~ 현재 : 군산대학교

소프트웨어학과 부교수

관심분야 : 컴퓨터 비전, 영상처리, 기계학습, 딥 러닝

이 휘 중 (Hwijong Yi)



2007년 2월 : 충북대학교
농생물학과(공학사)
2015년 2월 : 충북대학교
농생물학과(공학석사)
2019년 2월 : 충북대학교
농생물학과(공학박사)
2011년 11월 ~ 현재 :

국립식량과학원 생산기술개발과 농업연구사

관심분야 : 곤충생태, 해충방제, 딥 러닝, 영상처리