

로컬 그룹화 및 차별적 특징 어텐션 학습 기반의 조밀한 해충 카운팅

손 창 환*

Dense Pest Counting based on Local Grouping and Discriminative Feature Attention Learning

Chang-Hwan Son*

본 성과물은 농촌진흥청 연구사업(과제번호: PJ016303)의 지원에 의해 이루어진 것임

요 약

트랩에서 촬영된 해충 이미지는 해충이 조밀하게 분포하거나 군집을 형성하기 때문에 폐색 문제 및 다양한 포즈 변화를 야기한다. 이러한 걸림돌을 해결하기 위해서는 객체 영역과 배경 영역을 분리하여 로컬 특징의 구별력을 강화할 수 있는 어텐션 메커니즘을 설계하는 것이 필수적이다. 이를 위해 본 연구에서는 로컬 그룹화 및 스케일 유도 어텐션을 멀티스케일 센터넷 프레임워크에 통합하는 새로운 방안을 제시한다. 특히 스택 백본의 첫 번째에서 예측된 히트맵 정보, 즉 해충의 중심점을 활용하여 로컬 어텐션을 설계하는 방법을 제안한다. 실험을 통해 제안된 모델은 로컬 그룹화 및 차별적 특징 어텐션을 통해 특징 구별력을 향상하였고, 폐색 및 포즈 변화 문제를 극복하는 데 효과적인 것으로 입증되었다. 무엇보다 제안된 모델은 기존 객체 검출 기반 카운팅 모델보다 약 24.6%의 정확도 향상을 달성하였다.

Abstract

Pest images captured from traps cause occlusion problems and various pose changes because the pests are densely distributed or form clusters. To solve these obstacles, it is essential to model an attention mechanism that can enhance the discrimination of local features by separating the object region and the background region. To this end, in this study, we present a new model that integrates local grouping and scale-induced attention into a multi-scale CenterNet framework. In particular, we propose a method to model local attention by utilizing the predicted heatmap information, i.e., the center point of the pest, from the first stack backbone. Through experiments, the proposed model has been proven to be effective in improving feature discrimination through local grouping and discriminative feature attention modeling, and in overcoming occlusion and pose changes. Above all, the proposed model achieves an accuracy improvement of approximately 24.6% over the existing object detection-based counting model.

Keywords

digital trap, pest counting, stacked backbone, attention, heatmap, object detection

* 국립군산대학교 소프트웨어학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0001-7077-3074>

· Received: Sep. 02, 2024, Revised: Sep. 23, 2024, Accepted: Sep. 26, 2024
· Corresponding Author: Chang-Hwan Son
Department of Software Science and Engineering, Kunsan National University, South Korea
Tel.: +82-63-469-8915, Email: cson@kunsan.ac.kr

I. 서 론

해충은 작물의 품질 저하와 경제적 손실을 초래하며, UN 식량 농업 기구에 따르면 매년 작물 생산량의 약 40%가 해충으로 손실된다[1]. 이를 해결하기 위해, 최근 디지털 트랩을 노지에 설치하여 해충 수를 파악하여 조기 방제를 적용할 수 있는 무인 관찰시스템이 도입되고 있다. 하지만, 트랩에 포획된 해충이 조밀하게 분포하여 해충 간에 서로 가려지는 폐색 문제, 포즈의 다양화, 해충 외관의 유사성 문제는 딥러닝 모델을 개발하는데 걸림돌로 작용하고 있다. 해충 카운팅을 위해, 바운딩 박스를 검출하는 기존의 객체 검출 모델을 적용할 수 있다. 대표적인 모델로는 Faster RCNN[2], YOLO[3] 등이 있다. 하지만 기존의 객체 검출 모델은 스파스하게 분포하는 객체를 대상으로 개발되었다. 즉, 이미지 내에 포함된 객체의 수가 적고 객체 간의 가려짐의 정도가 낮다. 또한 자동차, 보행자와 같이 객체 스케일이 적절하며 객체 간의 외관의 유사도가 낮은 편이다. 하지만, 본 연구에서 다룰 해충 이미지는 조밀하게 분포하는 해충의 개수를 카운팅하는 문제로 기존의 객체 검출 문제와는 확연히 다르다. 그림 1은 조밀한 해충 카운팅 문제점을 보여주고 있다. 그림에서 해충들의 가려짐이 심하고 해충의 포즈도 다양하며 외관도 아주 유사한 것을 볼 수 있다. 따라서 기존의 객체 검출은 조밀한 해충 카운팅 문제에는 적합하지 않다.



그림 1. 조밀하게 분포하는 해충들의 예시
Fig. 1. Sample of densely distributed pests

본 연구에서는 조밀한 해충 카운팅 문제를 해결할 수 있는 새로운 해충 카운팅 모델을 제안하고자

한다. 이를 위해, 최근 조밀한 해충 카운팅 문제에 최적화된 멀티스케일 센터넷 모델[4]을 개선하는 작업을 수행하고자 한다. 멀티스케일 센터넷 모델은 기존의 객체 검출 모델과는 달리 고해상도의 출력 텐서를 제공한다. 이는 다운샘플링 변환에 따른 최종 예측 텐서에서의 객체 위치의 샘플링 오차를 줄일 수 있음을 의미한다. 또한 스택 백본 간의 스킵 연결을 통해, 저해상도 특징 벡터와 고해상도 특징 벡터를 퓨전함으로써, 저해상도 특징 벡터를 부가 정도를 활용하여 고해상도의 특징 벡터의 구별력을 강화할 수 있었다. 하지만, 퓨전 과정에서 객체와 배경 특징을 구분하지 않고 전역적인 어텐션 과정을 적용하였다. 따라서 본 연구에서는 기존의 멀티스케일 센터넷에서 객체와 배경 특징을 분리하여 로컬 어텐션을 적용할 수 있는 조밀한 해충 카운팅 모델을 개발하고자 한다. 실험결과를 통해, 제안한 조밀한 해충 카운팅 모델은 해충의 특징 구별력을 강화하여 폐색 및 포즈의 가변성에 강인하게 대처할 수 있음을 확인하였다. 그리고 기존의 객체 검출 모델보다 약 24.6% 정확도를 개선하여 해충 카운팅 분야의 괄목할만한 성과를 달성하였다.

II. 멀티스케일 센터넷

최근 멀티스케일 센터넷(Multiscale CenterNet)[4]이 조밀한 해충 카운팅 문제를 해결하기 위해, 전역 어텐션(Global attention) 모델링에 기반한 멀티스케일 퓨전 기법을 소개하였다. 그림 2는 LR(Low-Resolution) 및 HR(High-Resolution) Hourglass라 불리는 서로 다른 스케일의 백본과 전역 어텐션을 포함하는 멀티스케일 센터넷의 핵심 모듈을 보여준다. 최종 객체 검출을 위해, HR Hourglass의 출력 결과인 고해상도 히트맵(Heatmap), 오프셋 맵(Offset map), 바운딩 박스 맵(Bounding box map)이 사용된다. 이러한 HR 예측치는 다운샘플링에 따른 이산적인 오차를 줄여 해충의 폐색 문제에 더 강인하게 대처할 수 있다. 또한 LR 특징 정보를 토대로 HR 특징의 어텐션 기능을 강화할 수 있다. 특히, 전역 어텐션 모델링은 LR Hourglass와 HR Hourglass에서 추출된 서로 다른 스케일의 LR 및 HR 특징을 스케일-내적(Scale-dotproduct) 통해 구현되었다.

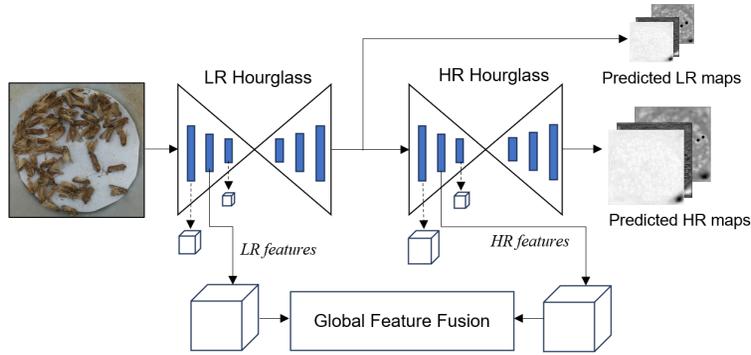


그림 2. 밀집한 해충 카운팅을 위한 기존의 multiscale CenterNet
Fig. 2. Conventional multiscale CenterNet for dense pest counting

$$Q = V = \Upsilon(\Phi_{HR}^l) \quad (1)$$

$$K = \Upsilon(\Phi_{LR}^l) \quad (2)$$

$$Atten(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

Φ_{HR}^l 과 Φ_{LR}^l 은 각각 HR과 LR Hourglass의 l 번째 잔차 블록(Residual block)에서 추출된 특징 맵을 의미한다. 그리고 Υ 은 특징 임베딩을 위한 선형 변환이다. 식 (1)과 (2)에서 HR 특징은 쿼리(Q)와 값(V)으로 사용되고 LR 특징은 키(K)로 사용된다. 즉, LR 특징이 HR 특징의 구별력을 제어하는 키로 작동한다. 식 (3)은 스케일-내적 연산에 해당하며, d_k 는 차원수이며 $softmax$ 는 입력값을 [0-1]로 사상하는 소프트맥스 함수를 나타낸다. QK^T 는 유사도 행렬이며 V 의 선형 결합에 필요한 가중치 값을 저장하고 있다. 따라서, LR 특징은 HR 특징 맵에서 어떤 특징을 강화하고 억제해야 되는지에 관한 제어 역할을 담당한다. 이를 통해, 특징 구별력을 제고하여 최종 해충 카운팅의 정확도를 개선할 수 있다.

III. 조밀한 해충 카운팅을 위한 제안한 로컬 그룹화 및 스케일 유도 어텐션 모델

멀티스케일 센터넷은 밀집된 해충 카운팅 및 검출에 최적화된 모델이지만, 여전히 개선 사항이 존재한다. 기존의 멀티스케일 센터넷은 배경과 해충 객체에 대한 그룹화 과정 없이 전역적인 어텐션을

수행한다. 하지만 조밀하게 군집을 형성하는 해충을 세밀하게 분리하기에는 한계가 있을 수밖에 없다. 따라서 조밀한 해충 카운팅 문제에서는 객체와 배경을 분리하여 로컬 어텐션을 모델링하는 과정이 필수적이다. 따라서 이 연구에서는 로컬 그룹화 및 스케일 유도 어텐션을 멀티스케일 센터넷에 통합할 수 있는 새로운 아키텍처 설계를 제시하고자 한다.

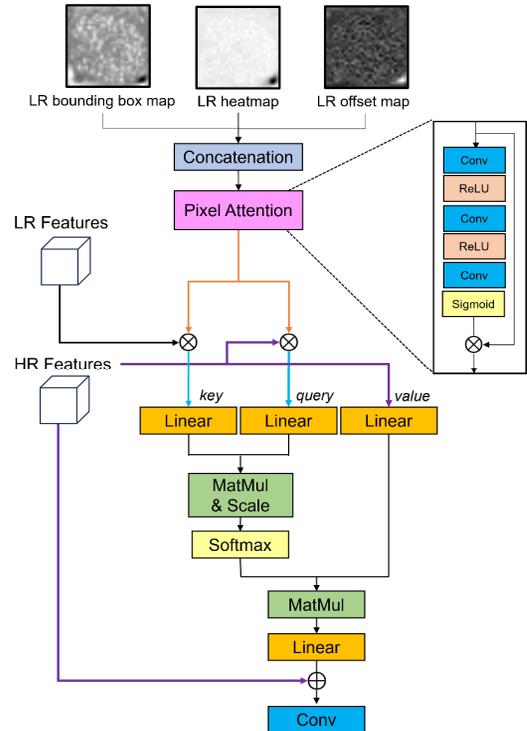


그림 3. 제안된 로컬 어텐션 모듈의 아키텍처
Fig. 3. Architecture of the proposed local attention module

그림 3은 제안된 로컬 어텐션 모듈의 아키텍처이다. 제안한 로컬 어텐션 모듈은 HR 특징의 구별력을 강화하는 것을 목적으로 하며 이를 위해 첫 번째 Hourglass의 예측 맵, 즉 히트맵, 읍셋 맵, 바운딩 박스 맵을 활용하는 간단하지만 효과적인 로컬 어텐션 모델링 설계 방식을 제안한다. 그림 3의 상단에서 예측된 LR 히트맵은 해충의 중심점을 포함하고 있기 때문에 객체와 배경을 쉽게 분리할 수 있다. 이는 부가적인 클러스터링이나 전경 분리와 같은 복잡한 모델링 작업을 대체할 수 있다는 것을 말한다. LR 히트맵은 그 자체로 로컬 그룹화 정보를 포함하고 있기 때문에 로컬 클러스터링을 위한 초기 솔루션을 제공할 수 있다. 이를 통해, HR 배경 특징을 억제하고 HR 객체 특징을 보다 강화할 수 있는 로컬 어텐션 구현할 수 있다.

3.1 학습 가능한 공간 어텐션 맵 변환

제안된 로컬 어텐션 모듈은 그림 3의 상단에서 알 수 있듯이 예측 맵을 입력으로 필요로 한다. 특히, 히트맵은 해충의 중심점을 포함하고 있기에 로컬 그룹화를 위한 초기 시작점으로 활용될 수 있다. 하지만, 히트맵 자체는 예측치로써 오차를 포함할 수 있으며 희소한 분포를 가지기 때문에 히트맵을 바로 사용하는 것은 바람직하지 않다. 따라서 제안한 로컬 어텐션 모듈에서는 히트맵을 포함한 예측맵을 학습 가능한 공간 어텐션(Spatial attention) 맵으로 변환하고자 한다. 이를 위해, SqueezeNet[-]에서 도입된 픽셀 어텐션 모듈(PAM, Pixel Attention Module)을 적용하고자 한다. PAM 구조는 그림 2의 오른쪽 상단과 같이 합성곱(Conv), 정류 선형 유닛(ReLU, Rectified Linear Unit) 및 시그모이드(Sigmoid)와 같은 세 가지 유형의 계층으로 구성된다.

$$F = \text{Concat}(\hat{Y}^{LR}, \hat{S}^{LR}, \hat{O}^{LR}) \quad (4)$$

$$W = f^s \circ f^c \circ f^a \circ f^c \circ f^a \circ f^c(F) \quad (5)$$

$$M = F \otimes W \quad (6)$$

식 (4)의 *Concat*은 연결(Concatenation) 계층이며 f^s , f^c , f^a 는 각각 시그모이드, Conv, ReLU 계층을 나타낸다. 기호 \circ 및 \otimes 는 각각 합성 함수와 요소별 곱셈을 나타낸다. 그리고 \hat{Y}^{LR} , \hat{S}^{LR} , \hat{O}^{LR} 는 각각 예측된 히트맵, 바운딩 박스 맵, 읍셋 맵을 의미한다. 따라서 식 (4)는 이 세 가지 예측 맵을 채널 축을 따라 쌓아서 PAM의 입력 텐서 F 를 생성한다. 식 (5)는 입력 텐서 F 에서 공간적으로 어떤 영역을 더 강조하고 억제할지를 제어하기 위한 2차원 가중치 맵 W 를 생성한다. 그리고 식 (5)의 마지막 계층은 입력 값을 0과 1로 압축하기 위한 시그모이드 계층임을 알 수 있다. PAM의 최종 출력은 식 (6)에 해당하며, 입력 F 와 가중치 맵 W 의 요소별 곱셈으로 결정된다. 이 수식은 로컬 그룹화의 초기 솔루션인 예측 맵을 학습 가능한 공간 어텐션 맵으로 변환한 것을 의미한다. 즉, 입력 F 는 학습된 공간 어텐션 맵인 W 에 따라 가변될 수 있다. 제안된 학습 가능한 공간 어텐션 맵 변환은 로컬 특징 그룹화를 보다 유연하고 정확한 방향으로 유도할 수 있다.

3.2 로컬 그룹화와 스케일 유도 어텐션

트랜스포머에서 소개된 자가 어텐션(Self attention) 모델[5]은 의미 정보를 추출하고 배경 영향을 줄이는 데 아주 탁월한 모델로 인정받고 있다. 특히 입력 토큰 간의 전역적인 종속성을 스케일 내적 연산으로 구현하여 탁월한 성능을 입증하였다. 따라서 제안한 모델에서도 스케일 내적 연산을 기반으로 로컬 어텐션을 모델링하고자 한다. 즉, 제안된 로컬 어텐션은 기존 스케일 내적 어텐션의 고도화된 버전으로 간주할 수 있다. 본 연구에서는 객체 특징을 추출하기 위해, 주파수 영역에서 곱셈으로 표현된 필터링 개념을 적용하고자 한다. 즉, 학습 가능한 공간 어텐션 맵인 M 을 입력 LR 및 HR 특징에 곱하여 객체 영역을 추출하고자 한다.

$$Q_F = \Upsilon(\Phi_{HR}^l) \otimes f^{d(l)}(M) \quad (7)$$

$$K_F = \Upsilon(\Phi_{LR}^l) \otimes f^{d(l)}(M) \quad (8)$$

$$V = \Upsilon(\Phi_{HR}^l) \quad (9)$$

Φ_{HR}^l 과 Φ_{LR}^l 은 각각 HR과 LR hourglass의 l 번째 잔차 블록에서 추출한 HR과 LR 특징을 나타낸다. 그리고 r 는 임베딩 계층이며 $f^{d(i)}$ 는 다운샘플링을 의미한다. $f^{d(i)}$ 는 M 의 크기를 HR과 LR 특징 크기로 변경하기 위해 필요하며 선형 보간법으로 구현되었다. 식 (7)과 (8)에서 결국 학습 가능한 공간 어텐션 맵인 M 이 각각 HR 특징과 LR 특징에 원소별로 곱해진 것으로 해석할 수 있다. 즉, 주파수 영역에서의 이미지 필터링처럼 작용하여 객체 특징을 추출하고 배경 특징을 제거하는 역할을 한다. 그리고 트랜스포머에서 사용된 자가 어텐션과는 달리, 식 (7)-(9)에서 쿼리, 키, 값이 다른 형태를 취한다. 필터링된 HR과 LR 객체 특징은 Q_F 와 K_F 로 사용되고, HR 특징은 V 로 사용된 것을 볼 수 있다.

$$SIM_F = Q_F K_F^T \quad (10)$$

$$Atten(Q_F, K_F, V) = softmax\left(\frac{SIM_F}{\sqrt{d_k}}\right) V \quad (11)$$

여기서 SIM_F 은 유사도 행렬을 의미하고, 필터링된 HR 특징 Q_F 와 필터링된 LR 특징 K_F 가 얼마나 유사한지 그 상관성을 측정하여 저장하고 있다. 이 유사도 행렬은 식 (11)에서와 같이 HR 특징 V 의 선형 결합(Linear combination)을 위해 가중치 값으로 사용된다. 따라서, 유사도 행렬은 서로 다른 스케일의 객체 특징 간의 내적 결과이므로 로컬 그룹화가 적용된 어텐션이라 볼 수 있다. 또한 필터링된 객체 LR 특징은 어떤 HR 특징을 더 강조되는지 제어 변수 역할을 수행하기 때문에 스케일 기반 어텐션이라 볼 수 있다.

최종 HR 특징은 다음과 같이 스킵 연결을 통해 업데이트된다.

$$\Phi_{HR}^{(l+1)} = f^c(\Phi_{HR}^l + Atten(Q_F, K_F, V)) \quad (12)$$

로컬 어텐션이 적용된 결과는 스킵 연결을 통해 입력 HR 특징과 원소별 덧셈을 통하여 퓨전되고 최종 합성곱 계층을 통과한다. 이 HR 특징 $\Phi_{HR}^{(l+1)}$ 은 다음 HR Hourglass 계층으로 전달된다. 이 프로

세스는 스택 Hourglass 간의 스킵 연결마다 반복된다. 이를 통해, 제안한 로컬 어텐션 모듈은 HR 특징 구별력을 향상시켜 궁극적으로 밀집된 해충 카운팅의 성능 개선으로 이어질 수 있다.

IV. 실험 결과

조밀 해충 카운팅 실험을 위해, 본 연구에서는 1에서 200마리까지 거의 균일한 분포를 갖는 톱다리 노린재 해충 데이터셋을 구성했다. 디지털 트랩에 표본을 넣어서 총 3,000장의 이미지를 촬영했으며 실제 환경을 반영하기 위해, 트랩을 흔들어서 해충 간 가려짐 현상이나 군집을 형성하도록 유도했다. 데이터 증강을 위해 이미지 반전(Image Flipping) 기법을 적용했으며 학습 데이터와 테스트 데이터셋의 비율은 7:3로 설정했다. Adam[6] 최적화 기법을 적용했고 배치 크기는 4로 설정했다. 에폭 수는 100이며 학습률은 0.001로 설정하였고 PyTorch 프레임워크로 사용했다. 입력 해충 이미지는 512×512의 고정 크기로 리사이징 되었고 스케일 조절을 위한 스트라이드는 4로 설정되었다. 보다 자세한 환경 설정은 멀티스케일 센터넷 논문에 설명되어 있다.

4.1 정량적 평가

제안한 로컬 어텐션 모델을 평가하기 위해, RetinaNet[7], RepPoints[8], CenterNet[9], Multiscale CenterNet[4], YOLOX[10], YOLOv7[11], Faster RCNN[2]과 비교하였다. 정량적 지표를 위해, 평균 절대 오차(MAE, Mean Absolute Error), 평균 제곱근 오차(RMSE, Root Mean Absolute Error), 평균 정밀도(AP, Average Precision)를 계산하였다 [12]. 여기서 MAE와 RMSE 지표는 해충 카운팅의 정확도를 반영할 수 있고 AP 지표는 바운딩 박스 검출의 정확도를 측정할 수 있다.

표 1은 해충 카운팅 모델에 대한 정확도 결과를 보여준다. 표 1에서 기존의 Faster RCNN, RedPoints, 및 YOLO 계열 모델의 성능이 예상외로 저조한 것을 볼 수 있다. 이는 기존 객체 검출 기반 카운팅 모델이 조밀한 해충 카운팅 문제에는 적합하지 않음을 의미한다.

표 1. 해충 카운팅 모델의 정량적 평가
Table 1. Quantitative evaluation of pest counting models

Counting models	MAE(↓)	RMSE(↓)	AP(↑)
RetinaNet[7]	39.589	52.454	0.595
Faster RCNN[2]	29.397	42.553	0.521
RepPoints[8]	28.404	43.095	0.678
YOLOX[10]	11.612	14.717	0.676
YOLOv7[11]	4.843	16.573	0.609
CenterNet[9]	6.015	11.031	0.653
Multiscale CenterNet[4]	3.334	9.719	0.693
Proposed model	2.538	6.666	0.920

즉, 일반 객체 검출 모델은 폐색 문제, 포즈 가변성, 외관 유사도 문제 등으로 인접한 해충을 식별하는데 한계점을 보인다. 이는 기존 객체 검출 기반 모델이 조밀한 해충 카운팅에는 적합하지 않으며 새로운 해충 카운팅 모델 개발의 필요성을 뒷받침한다. 표 1에서 멀티스케일 센터넷과 제안한 모델이 기존의 RCNN 및 YOLO 계열보다 3가지 지표에서 모두 우수한 것을 볼 수 있다. 특히, 제안된 모델은 MAE 평가에서 Faster RCNN, YOLOv7, Multiscale

CenterNet에 비해 각각 1058.2%, 90.8%, 31.3% 정확도 향상을 달성했다. AP 평가에서도 35.3%, 33.8%, 24.6%의 정확도 개선했다. 이는 조밀 해충 카운팅 분야에서 괄목할만한 성과로 볼 수 있다.

4.2 해충 검출 시각화 및 정성적 평가

그림 4는 바운딩 박스 검출 결과를 시각화한 이미지다. 각각의 이미지 위에는 정답 해충 마릿수와 예측된 마릿수가 적혀 있다. 제안한 모델은 기존 멀티스케일 센터넷을 고도화한 버전이므로, 기존 멀티스케일 센터넷의 바운딩 박스 검출 결과와 비교하였다. 이를 통해, 제안한 로컬 어텐션 모델의 효과를 검증할 수 있다. 그림 4의 해충 이미지에서 해충의 수가 180마리에 달하여 해충 간 폐색, 포즈 변화, 유사도 문제 등이 발생하는 것을 볼 수 있다. 연두색 박스, 즉 멀티스케일 센터넷으로 검출된 확대된 영역에서 해충이 미검출된 것을 볼 수 있다. 이는 기존 멀티스케일 센터넷은 해충들이 거의 인접하게 붙어 있거나 포즈 및 색상 변화가 동반될 경우 강인하게 대처하지 못함을 의미한다.

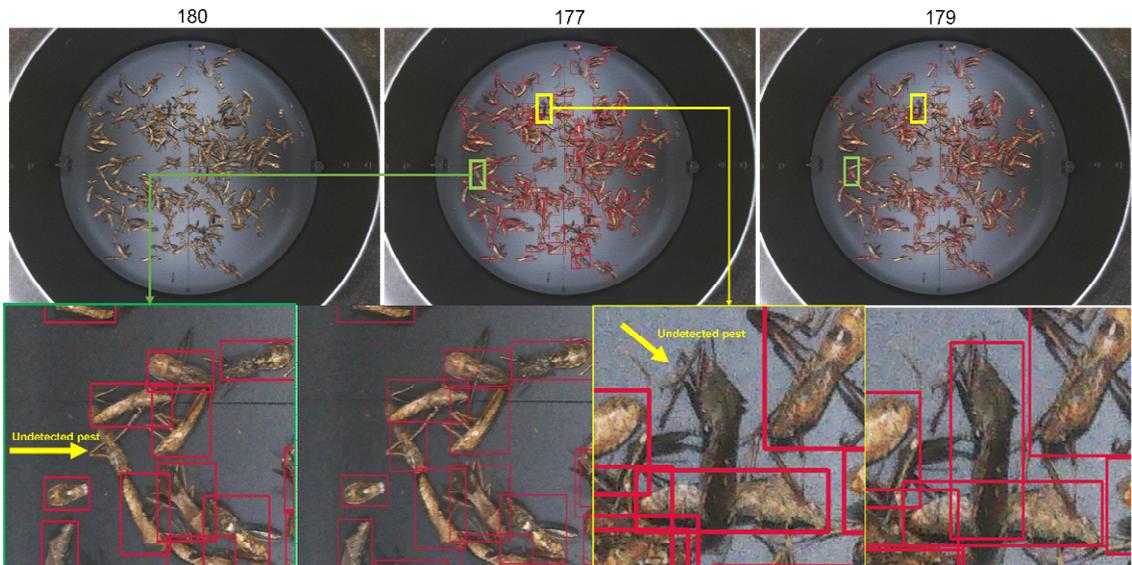


그림 4. 바운딩 박스 검출 예시; (상단) 입력 해충 이미지 (왼쪽), 기존 멀티스케일 센터넷으로 검출된 결과 (중앙), 제안한 로컬 어텐션 모델로 검출된 결과 (오른쪽), (하단) 확대 영상

Fig. 4. Examples of detected bounding boxes: (upper part) input pest image, detected result with conventional multiscale CenterNet, detected result with the proposed local attention model, (lower part) enlarged images

반면, 제안한 방법은 인접한 해충을 세밀하게 구분하고 포즈 및 색상 변화에도 민감하지 않고 해충을 정확하게 검출할 수 있었다. 이는 제안한 로컬 어텐션 모델이 폐색, 포즈 변화 문제 등에 효과적으로 대응할 수 있는 모델임을 말해준다. 그리고 Hourglass에서 출력된 예측 맵이 로컬 특징 그룹화를 위한 좋은 솔루션이 되어 객체의 특징을 강화하고 배경 특징을 억제할 수 있음을 뒷받침해준다.

그림 4는 다른 해충 이미지에 대한 예측된 마릿수와 시각화된 바운딩 박스 결과를 보여주고 있다. 그림에서 제안한 기법이 기존의 멀티스케일 센터넷보다 예측 마릿수가 정답에 더 가까운 것을 볼 수 있다. 실제 바운딩 박스 검출 결과에서도 더 정확한 결과를 얻을 수 있었다. 원본 이미지의 노란색 박스를 확대한 영상에서 제안한 모델은 군집을 형성한 해충들을 정확하게 구분하여 바운딩 박스를 검출한 것을 볼 수 있다.

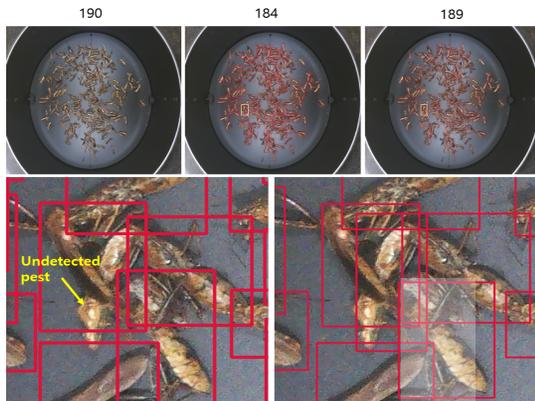


그림 5. 바운딩 박스 검출 예시; (상단) 입력 해충 이미지, 기존 멀티스케일 센터넷으로 검출된 결과, 제안한 로컬 어텐션 모델로 검출된 결과 (좌에서 우), (하단) 확대된 영상, 기존 멀티스케일 센터넷(좌측)과 제안한 모델(우측)
Fig. 5. Examples of detected bounding boxes; (upper part) input pest image, detected result with conventional multiscale CenterNet, detected result with the proposed local attention model (left to right), (lower part) enlarged images, multiscale CenterNet(left), proposed model(right)

V. 결 론

본 연구는 디지털 트랩에서 포획된 조밀하게 분포된 해충 개수를 예측하는 새로운 문제를 제시하

였다. 기존의 객체 검출과는 달리, 조밀한 해충 카운팅은 심각한 폐색 문제, 다양한 포즈 변화, 외관의 유사성 문제를 야기한다. 이를 해결하기 위해, 이 연구에서는 로컬 그룹화와 스케일 유도 어텐션을 적용한 새로운 해충 카운팅 모델을 제안하였다. 특히, 제안한 모델은 클러스터링이나 전경 분리와 같은 복잡한 작업을 수행하지 않고, 예측맵을 초기치로 시작해서 학습 가능한 공간 어텐션 맵 변환을 통해 간단하지만 효과적인 로컬 어텐션 설계 방안을 제시하였다. 실험 결과를 통해, 제안된 로컬 어텐션 모델은 기존 최첨단 모델을 큰 격차로 능가하였고 조밀 해충 카운팅 분야에서 괄목할만한 성과를 거두었다.

References

- [1] W. Zhang, H. Huang, Y. Sun, and X. Wu, "AgriPest-YOLO: A rapid light-trap agricultural pest detection method based on deep learning", *Frontiers in Plant Science*, Vol. 13, Dec. 2022. <https://doi.org/10.3389/fpls.2022.1079384>.
- [2] S. 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. <https://doi.org/10.1109/TPAMI.2016.2577031>.
- [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] J.-H. Lee and C.-H. Son, "Trap-based pest counting: Multiscale and deformable attention CenterNet integrating internal LR and HR joint feature learning", *Remote Sensing*, Vol. 15, No. 15, pp. 3810, Jul. 2023. <https://doi.org/10.3390/rs15153810>
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit,

- L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need", In Proc. Advances in Neural Information Processing Systems, Long Beach, USA, Dec. 2017. <https://doi.org/10.48550/arXiv.1706.03762>.
- [6] 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>.
- [7] 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, Venice, Italy, pp. 2999-3007, Oct. 2017. <https://doi.org/10.1109/ICCV.2017.324>.
- [8] 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, Seoul, Korea, pp. 9657-9664, Aug. 2019. <https://doi.org/10.1109/ICCV.2019.00975>.
- [9] 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>.
- [10] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021", arXiv:2107.08430 [cs.CV], Jul. 2021. <https://doi.org/10.48550/arXiv.2107.08430>.
- [11] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors", arXiv:2207.02696 [cs.CV], Jul. 2022. <https://doi.org/10.48550/arXiv.2207.02696>.
- [12] J.-H. Lee, C.-H. Son, and H. Yi, "Multiscale CenterNet for pest detection and counting", The Journal of Korean Institute of Information Technology, Vol. 20, No. 7, pp. 111-121, Jul. 2022. <https://doi.org/10.14801/jkiit.2022.20.7.111>.

저자소개

손 창 환 (Chang-Hwan Son)



2002년 2월 : 경북대학교
전자전기공학부(공학사)

2004년 2월 : 경북대학교
전자공학과(공학석사)

2008년 8월 : 경북대학교
전자공학과(공학박사)

2017년 4월 ~ 현재 :

국립군산대학교 소프트웨어학과 부교수
관심분야 : 컴퓨터 비전, 영상처리, 기계학습, 딥 러닝