

멀티스케일 로컬 어텐션 기반의 해충 카운팅 모델

정재민*, 손창환**, 홍서연***

Pest Counting Model based on Multiscale Local Attention

Jae-Min Jeong*, Chang-Hwan Son**, and Seoyeon Hong***

본 성과물은 농촌진흥청 연구사업(과제번호: PJ01630303)의 지원에 의해 이루어진 것임

요약

본 논문에서는 해충 카운팅을 위한 로컬 어텐션 기반의 MaDa-CenterNet(v2)를 소개하고자 한다. 기존의 MaDa-CenterNet은 객체와 배경 영역을 분리하지 않고 어텐션 퓨전을 적용하여 특징 구별력을 약화할 수 있다. 이러한 단점을 보완하기 위해, 본 연구에서는 객체와 배경 영역을 분리하기 위한 객체 및 배경 마스킹 모듈과 이를 기반으로 객체 영역의 특징 구별력을 강화하기 위한 멀티스케일 로컬 어텐션 퓨전 모듈을 제안하고자 한다. 특히 객체 및 배경 마스킹 모듈에서는 저해상도 모래시계에서 추정된 히트맵을 사용하여 마스크 맵을 생성하고 멀티스케일 로컬 어텐션 퓨전 모듈에서는 생성된 마스크를 기반으로 객체와 배경 영역에 따라 멀티스케일 로컬 어텐션을 모델링한다. 실험 결과를 통해, 제안한 해충 카운팅 모델이 기존의 MaDa-CenterNet의 성능을 개선할 수 있었으며 제안한 로컬 어텐션 퓨전 모델이 성능 개선에 효과적임을 입증하였다.

Abstract

In this paper, we would like to introduce MaDa-CenterNet(v2) based on local attention for pest counting. The existing MaDa-CenterNet applies attention fusion without separating object and background regions, thereby weakening feature discrimination. In order to compensate for these shortcomings, this study proposes an object and background masking module to separate the object and background areas, and based on this, a multi-scale local attention fusion module to enhance feature discrimination of the object area. In particular, the object and background masking module generates a mask map using the heatmap estimated from the low-resolution hourglass, and the multiscale local attention fusion module models multiscale local attention according to the object and background regions based on the generated mask. Through experimental results, it was proven that the proposed pest counting model was able to improve the performance of the existing MaDa-CenterNet and that the proposed local attention fusion model was effective in improving performance.

Keywords

pest counting, digital trap, deep learning, CenterNet, local attention

* 군산대학교 소프트웨어학부 소프트웨어학전공 학사과정 · Received: Jan. 08, 2024, Revised: Jan. 25, 2024, Accepted: Jan. 28, 2024
- ORCID: <https://orcid.org/0009-0004-1134-3305> · Corresponding Author: Chang-Hwan Son
** 군산대학교 소프트웨어학부 소프트웨어학전공 교수 · Software Science and Engineering Major, School of Software,
(교신저자) · Kunsan National University, Korea
- ORCID: <https://orcid.org/0000-0001-7077-3074> · Tel.: +82-63-469-8915, Email: cson@kunsan.ac.kr
*** 국립식량과학원 생산기술개발과 농업연구사
- ORCID: <https://orcid.org/0009-0006-2013-813X>

1. 서 론

해충은 작물의 잎, 줄기, 열매 등에 직접적인 손상을 입혀 농작물의 품질 저하를 수반하고 경제적 손실을 유발한다. 따라서 해충의 종류와 발생량을 신속히 파악해서 해충 방제를 적기에 실시해야 한다[1]. 이를 위해, 농가에서는 노지에 디지털 트랩을 설치해서 해충의 발생 시기, 종류 및 마릿수를 예측하였다. 즉, 디지털 트랩에 광원이나 페로몬과 같은 유인물질을 장착하여 해충을 포획한 후, 내장된 디지털 카메라를 활용해서 이미지를 촬영하고 학습된 기계학습이나 딥러닝 모델을 적용하여 해충의 발생량을 예측하였다. 최근 디지털 트랩 기반으로 해충 마릿수를 예측하는 해충 카운팅 기술이 활발히 연구되고 있다. 하지만 해충이 많은 경우, 해충 일부가 가려지는 폐색 문제나 해충의 날개가 접히거나 퍼진 경우와 같은 해충 자세의 다양성으로 해충 카운팅이 어려움을 겪고 있다.

기존의 해충 카운팅 기술은 크게 객체 검출 기반 해충 카운팅(Pest counting by object detection)과 밀도 맵 추정 기반 해충 카운팅(Pest counting by density map estimation)으로 구분할 수 있다[2][3]. 객체 검출 기반 해충 카운팅 기술은 해충이 존재하는 영역의 바운딩 박스를 검출하여 마릿수를 예측하는 방식이고 밀도 맵 추정 기반 해충 카운팅 기술은 밀도 맵을 추정하여 마릿수를 예측하는 방식이다. 원래 밀도 맵 추정 기반 해충 카운팅 기술은 군중 카운팅(Crowd counting) 기술과 동일한 것으로 해충이 많고 작은 경우 바운딩 박스 추정이 어렵기 때문에, 이를 해결하기 위해 밀도 맵을 생성하고 픽셀 값을 합산하여 마릿수를 예측할 수 있는 새로운 접근 방식에 해당한다. 따라서 트랩에서 포획된 해충의 마릿수가 아주 많아서 폐색이 심할 경우는 밀도 맵 추정 방식이 효과적일 수 있고 해충의 마릿수가 적당하고 트랩에 골고루 퍼져 있는 경우에는 객체 검출 카운팅 방식이 더 효과적일 수 있다. 그리고 밀도 맵 추정 방식은 이미지 크기의 밀도 맵을 추정해야 되므로 모델 용량과 계산 비용이 증가하는 단점이 있다. 반면 객체 검출 방식은 비교적 작은 사이즈의 특징 맵에서 바운딩 박스를 추정하기 때

문에 상대적으로 계산 비용과 모델 용량이 적을 수 있다. 대표적인 객체 검출 방식의 해충 카운팅 기술에는 RedPoint[4], RetinaNet[5], Faster RCNN[6], CenterNet[7] 등이 있고, 밀도 맵 추정 방식의 해충 카운팅 기술에는 MCNN[8], ICCNet[9], KDMG[10] 등이 있다.

최근 CenterNet[7]의 단점을 보완하여 해충 카운팅 성능을 고도화한 Mada-CenterNet[3]이 개발되었다. 이 연구에서 처음으로 객체 검출 및 밀도 추정 기반의 해충 카운팅 방식을 소개하였고 트랩에서 촬영된 해충 이미지를 대상으로 두 종류의 카운팅 모델의 성능을 평가하였다. Mada-CenterNet은 기존의 CenterNet 모델에 변형 가능한 컨볼루션(Deformable convolution)을 적용하여 폐색 문제나 포즈 다양성 문제를 해결하였다. 그리고 스케일이 다른 스택 방식의 백본을 도입하여 저해상도 특징과 고해상도 특징을 멀티스케일 어텐션 퓨전으로 결합하여 특징 구별력을 제고할 수 있었다. 하지만 MaDa-CenterNet은 멀티스케일 어텐션 모델 과정에서 해충이 존재하는 객체 영역과 배경 영역을 분리하지 않고 전체 영역에 대해 어텐션을 적용하였다. 이는 해충 영역과 배경 영역에 대한 특징 구별력을 개선할 수 있다는 것을 의미한다. 따라서 이 논문에서는 해충 영역과 배경 영역을 구분하여 어텐션 모델을 적용한 고도화된 MaDa-CenterNet(v2)을 제안하고자 한다. 특히, 객체 및 배경 마스킹 모듈을 설계하는 과정과 이를 활용하여 로컬 어텐션을 모델링하는 과정을 새롭게 제시하고자 한다. 그리고 실험 결과를 통해, 제안한 로컬 어텐션 기반의 MaDa-CenterNet(v2)가 기존의 CenterNet과 Mada-CenterNet보다 해충 카운팅 정확도 평가에서 더 우수함을 입증하고자 한다.

II. 기존의 MaDa-CenterNet

그림 1은 기존의 MaDa-CenterNet의 아키텍처를 보여준다. 그림에서 보듯이, MaDa-CenterNet은 스택으로 쌓은 모래시계(Hourglass)를 백본으로 사용해서 특징을 추출한다. 그리고 해충 마릿수가 발생 시기에 따라 가변적이므로 스케일 문제가 발생한다.

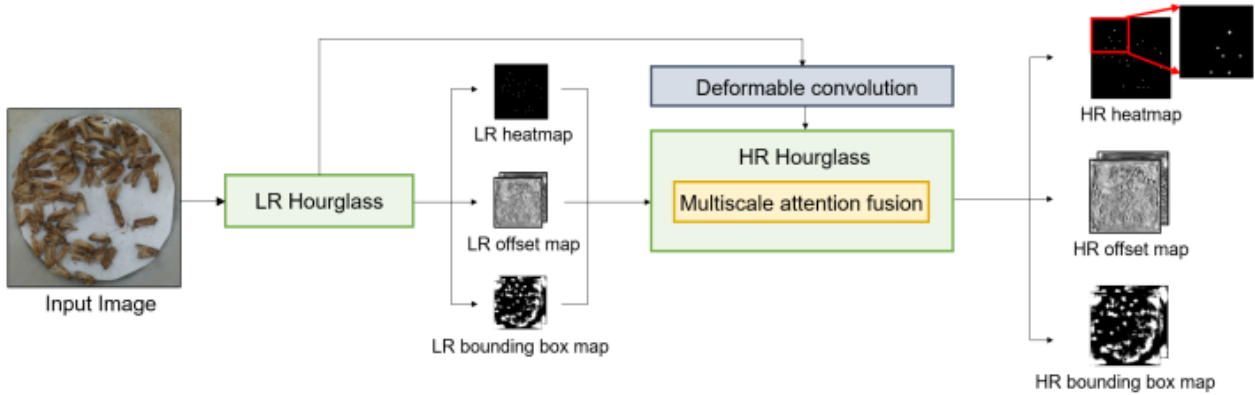


그림 1. 트랩 기반 해충 카운팅을 위한 MaDa-CenterNet의 아키텍처[3]
Fig. 1. Architecture of MaDa-CenterNet for trap-based pest counting[3]

이를 위해, 스케일이 다른 두 종류의 저해상도 모래시계(LR hourglass)와 고해상도 모래시계(HR hourglass)를 적용하였다. 또한 고해상도 모래시계의 특징 구별력을 제고하기 위해, 저해상도 모래시계의 출력 결과인 히트맵(Heatmap), 바운딩 박스 맵(Bounding box map), 오프셋 맵(Offset map)을 고해상도 모래시계의 입력으로 사용하였다. 그리고 저해상도 모래시계 내부에서 생성되는 특징 정보를 고해상도 모래시계의 특징 정보와 결합하였다. 즉, 저해상도 모래시계에서 추출된 특징 맵을 키(Key)로 활용하고 고해상도 모래시계에서 추출된 특징 맵은 쿼리(Query)와 값(Value)로 활용하여 멀티스케일 어텐션을 모델링하였다. 참고로 그림 1의 히트맵의 경우, 대부분이 배경 영역을 의미하는 검은색으로 표시되기 때문에 해충이 존재하는 흰색 영역이 잘 보이지 않는다. 이를 위해, 그림 1의 우측에 히트맵 일부분을 확대하여 키포인트 위치를 확인할 수 있게 하였다.

한편, 트랩에 포획된 해충이 많은 경우, 서로 가려지는 폐색 문제나 다양한 자세 변화가 발생하므로 기하학적인 변환을 고려할 필요가 있다. 이를 위해, MaDa-CenterNet에서는 변형 가능한 컨볼루션 연산을 적용하였다. 변형 가능한 컨볼루션 연산은 표준 컨볼루션 연산과는 달리, 고정된 국부 영역에서 필터링을 하는 것이 아니라 학습 가능한 오프셋 파라미터를 도입해서 필터링 영역을 확장할 수 있다. 표준 컨볼루션과 변형 가능한 컨볼루션 연산 과정은 다음과 같다[11].

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \quad (1)$$

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (2)$$

식 (1)은 컨볼루션 연산으로 w 와 x 는 각각 필터와 입력 이미지를 나타낸다. R 은 필터의 크기이며 컨볼루션 연산이 수행되는 국부 영역을 결정한다. 영상처리에서 표준 컨볼루션 연산은 에지나 텍스처와 같은 저수준의 특징을 추출할 수 있다. 따라서 식 (1)의 표준 컨볼루션 연산은 국부 영역에서 이미지 특징 추출이 가능하다. 하지만 필터의 영역이 제한적인 한계가 있다. 이를 극복하기 위해서, 변형 가능한 컨볼루션은 식 (2)와 같이 학습 가능한 오프셋 파라미터(Δp_n)를 추가로 도입했다. 이를 통해, 특징 추출 영역을 유연하게 확장하여 해충 카운팅에 더 효과적인 특징의 위치를 추정할 수 있다.

III. 제안한 해충 카운팅을 위한 로컬 어텐션 기반의 MaDa-CenterNet(v2)

3.1 제안한 접근 방법

기존의 MaDa-CenterNet은 객체와 배경 영역을 분리하지 않고 전체 특징 맵에 어텐션을 적용하였다. 따라서 배경 영역의 특징이 객체 영역의 특징에 영향을 줄 수 있고 해충 특징 구별력이 저하될 수 있다. 이러한 문제점을 보완하기 위해, 본 논문에서는 객체 영역과 배경 영역을 분리해서 특징 구별력을 강화할 수 있는 고도화된 MaDa-CenterNet(v2)를 제시하고자 한다.

4 멀티스케일 로컬 어텐션 기반의 해충 카운팅 모델

특히, 객체와 배경 영역을 분리하기 위한 마스킹 모듈 설계 과정과 이를 기반으로 객체와 배경 영역에 대해 각각 어텐션을 적용할 수 있는 로컬 어텐션 설계 과정을 자세히 소개하고자 한다. 이를 통해, 특징 맵에서 해충이 존재하는 위치의 중요도를 반영하여 해충 카운팅의 정확도를 제고하고자 한다.

3.2 제안한 해충 카운팅을 위한 로컬 어텐션 기반의 MaDa-CenterNet(v2)

그림 2는 제안한 로컬 어텐션 기반의 MaDa-CenterNet(v2)의 아키텍처를 보여준다. 그림에서 보듯이 기존의 MaDa-CenterNet과의 차이점은 객체와 배경 마스킹 모듈(8번 레이어)과 고해상도 모래시계에 추가로 삽입된 멀티스케일 로컬 어텐션 퓨전 모듈(9번 레이어)에 있다. 즉, MaDa-CenterNet의 멀티스케일 퓨전 모듈을 개선한 점이 가장 큰 차이점이다. 참고로 그림 2의 나머지 부분은 그림 1의 MaDa-CenterNet과 동일하다. 다만 보다 자세한 아키텍처를 제공하기 위해 상세하게 표현하였다.

그림 3은 고해상도 모래시계의 멀티스케일 로컬 어텐션 퓨전 모듈에 대한 상세한 구조를 보여준다. 먼저, 객체 및 배경 마스킹 모듈은 저해상도 모래시계에서 추정된 저해상도 히트맵(LR heatmap)을 필요로 한다. 히트맵은 객체가 존재하는 곳에는 화이트 색상에 가까운 값을 저장하고 있다. 따라서 객체 및 배경 마스킹 모듈은 히트맵의 값을 활용해서 객체와 배경 영역을 구분할 수 있는 마스크를 생성한다. 또한 고해상도 모래시계와 저해상도 모래시계는 인코더와 디코더 형태의 구조를 가진다. 따라서 동

일한 레이어 위치에서 특징 맵을 각각 추출한 후, 멀티스케일 로컬 어텐션 퓨전 모듈을 통과해서 저해상도 및 고해상도 특징 맵을 결합할 수 있다. 마지막으로 이 결합된 특징 맵은 고해상도 모래시계의 다음 레이어로 입력이 된다. 이를 통해, 멀티스케일 기반으로 객체와 배경에 대한 특징 구별력을 강화할 수 있다.

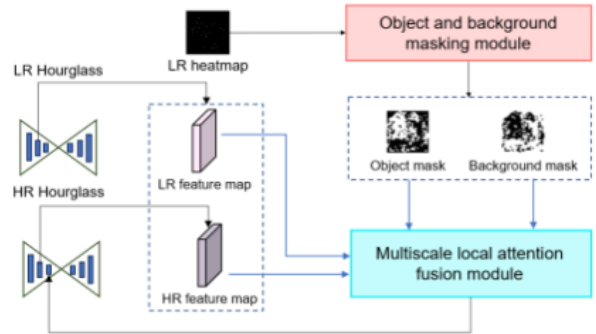


그림 3. 고해상도 모래시계의 상세한 구조
Fig. 3. Detailed architecture of HR Hourglass

3.3 객체 및 배경 마스킹 모듈

그림 4는 그림 3의 객체 및 배경 마스킹 모듈의 자세한 과정을 보여준다. 추정된 저해상도 히트맵은 [0-1] 사이의 값으로 표현되고 객체가 존재하는 위치에서는 1의 값에 가깝다. 즉, 객체 존재 유무를 확률로 표현한 맵으로 생각하면 된다. 따라서 이 히트맵을 활용하면 객체와 배경 영역을 분리할 수 있다. 먼저, 히트맵의 크기와 추출된 특징 맵의 크기를 정합하기 위해 다운 샘플링(Down-sampling)을 적용한다. 그리고 히트맵에서 임계값을 기준으로 마스크 맵을 생성하였다.

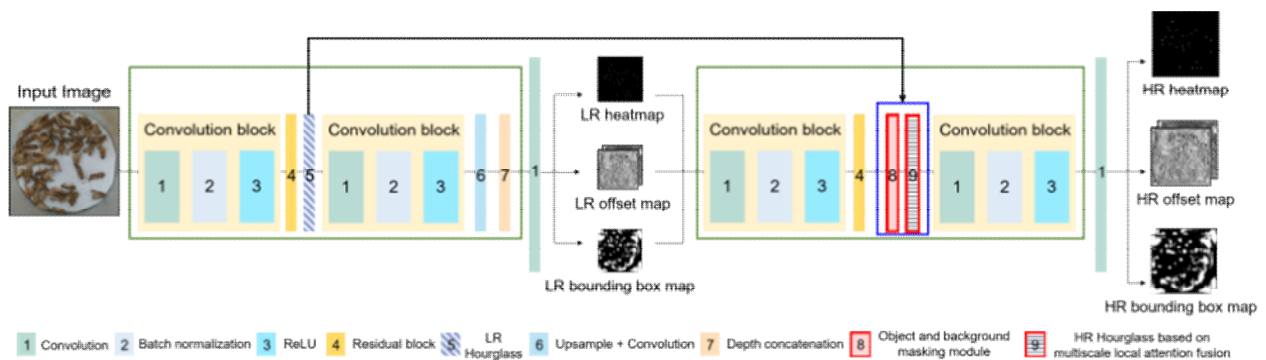


그림 2. 해충 카운팅을 위한 제안한 멀티스케일 로컬 어텐션 기반의 MaDa-CenterNet(v2)
Fig. 2. Proposed MaDa-CenterNet(v2) based on multiscale local attention for pest counting

배경 영역에서 히트맵의 값이 너무 작은 경우, 학습이 잘되도록 임계값을 경험적으로 설정하였다. 객체와 배경 마스크 생성 과정은 다음과 같다.

$$f(x,y) = \begin{cases} 1, & \text{if } Heatmap(x,y) \geq TH \\ 0.01, & \text{otherwise} \end{cases} \quad (3)$$

$$g(x,y) = \begin{cases} 1, & \text{if } Heatmap(x,y) < TH \\ 0.01, & \text{otherwise} \end{cases} \quad (4)$$

f 와 g 는 각각 객체와 배경 마스크를 나타내며, x 와 y 는 높이와 너비를 나타낸다. 객체 마스크는 히트맵의 값이 임계값 TH 보다 크면 1로 작으면 0.01로 할당한다. 반대로 배경 마스크는 임계값 TH 보다 작으면 1로 크면 0.01로 할당한다. 본 연구에서는 TH 값을 0.3으로 설정하였다. 배경 값을 0으로 설정하지 않은 이유는 컨볼루션 연산에서 대부분의 값이 0이 되기 때문에 특징 정보가 손실된다. 이는 학습 능력 저하를 유도한다. 따라서 본 연구에서는 경험적인 실험을 통해서 0.01과 0.3의 값을 도출했다.

그림 4의 맨 아래에는 생성된 객체 및 배경 마스크 맵의 예시를 보여주고 있다. 그림에서 보듯이, 객체 마스크는 객체가 존재하는 영역을 흰색으로 배경 마스크는 배경 영역을 흰색으로 표현한 것을 볼 수 있다.

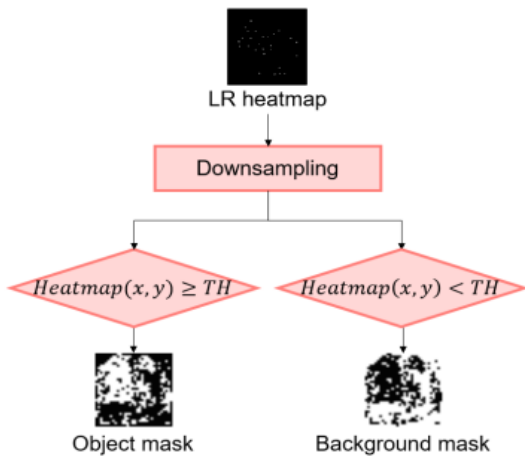


그림 4. 객체 및 배경 마스크 모듈
Fig. 4. Object and background masking module

3.4 멀티스케일 로컬 어텐션 퓨전 모듈

그림 5는 제안한 멀티스케일 로컬 어텐션 퓨전의 상세한 구조를 보여주고 있다. 그림에서 보듯이, 제

안한 로컬 어텐션 모듈은 크게 객체 영역에 대한 어텐션 과정과 배경 영역에 대한 어텐션 과정으로 나뉘어 있다. 객체 영역의 어텐션 과정은 객체 마스크 맵을 각각 저해상도와 고해상도 모래시계에서 생성된 특징 맵에 원소별로 곱해준다. 이때 객체 영역의 특징 값은 그대로 유지되고 배경 영역의 특징 값은 작은 값으로 변경된다. 그리고 마스크된 저해상도 특징맵은 키로, 마스크된 고해상도 특징맵은 쿼리와 값으로 각각 어텐션 과정에서 사용된다. 마찬가지로 배경 영역에서의 어텐션 과정은 배경 마스크를 각각 저해상도와 고해상도 특징맵에 원소별로 곱해준다. 배경 영역의 특징 값은 원본 그대로 유지되는 반면 객체 영역의 값은 아주 작은 값으로 매핑된다. 이렇게 마스크된 저해상도 특징맵과 고해상도 특징맵은 어텐션 과정에서 각각 키와 쿼리로 사용된다.

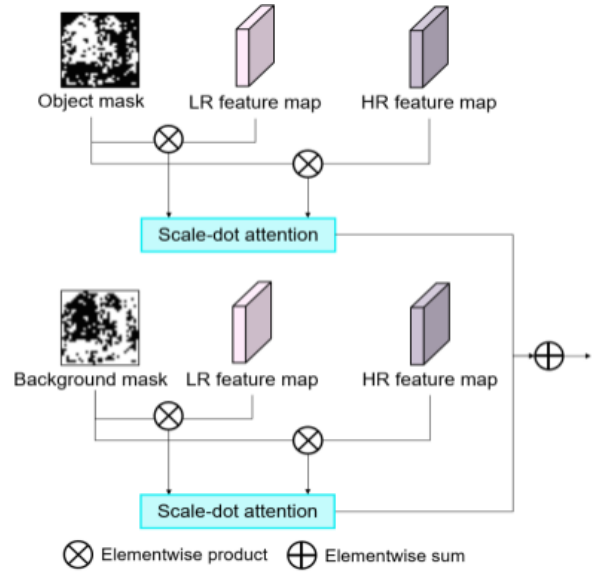


그림 5. 객체 및 배경 마스크 모듈
Fig. 5. Object and background masking module

본 연구에서는 로컬 어텐션을 구현하기 위해, 스케일 도트 어텐션(Scale-dot attention) 모델을 적용했다[12]. 자세한 계산과정은 다음과 같다.

$$F_{LR}^M = M \otimes F_{LR}, \quad F_{HR}^M = M \otimes F_{HR} \quad (5)$$

$$Q_M = F_{LR}^s W_q, \quad K_M = F_{HR}^s W_k, \quad V_M = F_{HR}^s W_v \quad (6)$$

$$Attention_M(Q, K, V) = softmax\left(\frac{Q_M K_M^T}{\sqrt{d}}\right) V_M \quad (7)$$

식 (5)에서 행렬 F 는 저해상도 또는 고해상도 모래시계에서 추출된 특징맵을 의미하고 M 은 객체 마스크 f 또는 배경 마스크 g 에 해당한다. 그리고 \otimes 는 원소별 곱 연산을 의미한다. 따라서 식 (5)는 마스크 처리 과정을 나타낸 수식이다. 식 (6)에서 F_{LR}^s 와 F_{HR}^s 는 각각 F_{LR}^M 와 F_{HR}^M 의 (x,y) 위치의 특징 벡터를 가로 방향으로 저장한 행렬에 해당한다. 그리고 W_q , W_k , W_v 는 각각 쿼리, 키, 값에 대한 임베딩 레이어를 구현하기 위한 행렬이다. 여기서 임베딩 레이어란 특징 차원을 변경하기 위한 계층으로써, 행렬 연산을 통해 공간 변환을 수행한다. 식 (7)에서 softmax 함수는 0-1사이의 값으로 매핑하기 위한 함수이고 행 방향으로 정규화 과정을 수행한다. 그리고 $Q_M K_M^T$ 는 벡터 간의 유사도를 측정하기 위한 벡터 내적 연산을 수행한다. d 는 키의 차원수를 나타내며 스케일링 과정을 통해 어텐션 연산을 안정화하며 학습을 원활하게 만든다.

그림 5에서 객체 및 배경 영역에 대해 어텐션을 수행한 후, 출력 결과는 원소별 덧셈을 통해서 융합된다. 그림에서 \oplus 는 원소별 덧셈을 의미한다. 융합된 특징 결과는 고해상도 모래시계의 다음 레이어로 전송되어 동일한 과정을 반복한다. 이는 저해상도 및 고해상도 특징 맵의 객체와 배경 영역의 특징 구별력을 개선하여 어텐션 학습 능력을 강화할 수 있음을 의미한다.

IV. 실험 및 결과

4.1 실험 환경

본 연구에서는 제안한 로컬 어텐션 기반의 MaDa-CenterNet(v2)를 학습하기 위해 트랩에서 촬영한 해충 데이터를 사용했다. 총 4,462장으로 구성된 데이터셋을 7:3의 비율로 랜덤하게 나누어 훈련 집합과 테스트 집합을 구성했다. 학습에 사용된 배치 크기는 8, 에폭은 100 그리고 학습률은 0.0001로 설정했다. 최적화 기법은 아담 옵티마이저[13]를 사용했고 파이토치 딥러닝 프레임워크로 제안한 모델을 구현하였다. 모델 성능 평가를 위해, 객체 검출 기반과 밀도 맵 추정 기반인 RetinaNet[5], Faster RCNN[6],

RepPoints[4], CenterNet[7], Multiscale CenterNet[14], MaDa-CenterNet[3], KDMG[10]을 비교하였다.

4.2 정량적 평가

정량적 평가를 위해, 본 논문에서는 해충 마릿수를 측정할 수 있는 평가 척도를 도입했다. 특히 기존의 군중 카운팅 분야에서 널리 활용되는 평균 절대 오차(MAE, Mean Absolute Error)와 평균 제곱근 오차(RMSE, Root Mean Squared Error)를 해충 카운팅의 정량적 평가 도구로 사용했다. MAE와 RMSE 둘 다 실제 해충 마릿수와 모델의 예측 마릿수와의 오차를 계산한다. 차이점은 MAE는 오차의 절대 값을 적용하며 RMSE는 오차의 제곱을 적용한다. 따라서 RMSE는 오차가 크면 클수록 페널티를 더 부과하는 효과를 지닌다. MAE와 RMSE 둘 다 값이 작을수록 해충 카운팅의 정확도가 우수하다는 것을 의미한다.

표 1은 기존 모델과 제안한 모델과의 MAE와 RMSE의 결과를 보여준다. 표 1에서 보듯이, 밀도맵 추정 기법인 KDMG보다 객체 검출 방식인 CenterNet 모델이 더 우수하다는 것을 알 수 있다. 그리고 제안한 MaDa-CenterNet(v2)가 기존의 MaDa-CenterNet보다 MAE와 RMSE 오차를 개선했음을 알 수 있다. 또한 기존 모델보다 훨씬 더 우수한 정확도를 달성할 수 있었다. 이는 해충 영역과 배경 영역에 대한 로컬 어텐션이 해충 카운팅 성능 개선에 효과적임을 말해준다. 참고로 밀도 추정 방식과 객체 검출 성능 방식은 데이터셋의 해충 비율에 따라 그 성능이 달라질 수 있다.

표 1. 정량적 평가

Table 1. Quantitative evaluation

	MAE	RMSE
RetinaNet[5]	4.634	10.449
Faster RCNN[6]	3.312	10.297
RepPoints[4]	1.471	3.436
CenterNet[7]	0.766	1.981
Multiscale Centernet[14]	0.752	2.066
MaDa-CenterNet[3]	0.696	1.806
KDMG[10]	1.273	2.625
Proposed MaDa-CenterNet(v2)	0.640	1.602





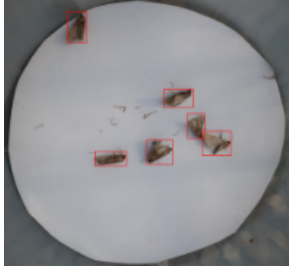
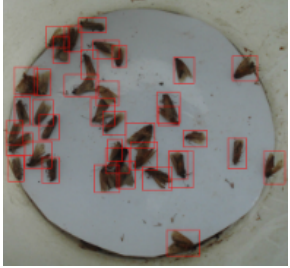
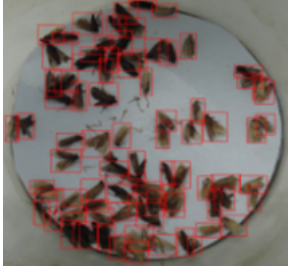

Measured number : 6	Measured number : 38	Measured number : 80	Measured number : 124
			
Predicted number : 6	Predicted number : 38	Predicted number : 82	Predicted number : 122
			

그림 6. 실험 결과; 해충 이미지와 측정된 해충 마리수(첫 번째 행), 예측된 해충 마리수와 바운딩 박스(두 번째 행)
 Fig. 6. Experimental results; input pest images and measured pest number (first row), predicted pest number and bounding boxes (second row)

4.3 해충 카운팅 결과 예시

그림 6은 제안한 로컬 어텐션 기반의 MaDa-CenterNet(v2)를 사용해서 추정된 해충 카운팅의 결과 예시이다. 제안한 모델은 객체 검출 방식에 해당하므로 결과 그림에 예측된 마리수와 검출된 바운딩 박스 결과를 같이 기재하였다. 그림 6에서 측정된 해충 마리수는 육안으로 직접 카운팅한 결과이다. 그림 6에서 보듯이, 해충의 마리수가 작은 경우나 큰 경우에도 측정된 해충 마리수와 예측된 해충 마리수가 거의 유사한 것을 볼 수 있다. 이는 제안한 객체 및 배경 마스크 모듈과 로컬 어텐션 퓨전 모듈이 해충 카운팅에 효과적임을 뒷받침해 준다.

V. 결 론

본 논문은 해충 카운팅을 위한 로컬 어텐션 기반의 MaDa-CenterNet(v2)를 제안하였다. 기존의 MaDa-CenterNet은 특징맵에서 해충 영역과 배경 영역을 나누지 않고 어텐션 모델을 구현하였다. 따라서 해충과 배경 영역에 대한 특징 구별력이 저하될 수 있다. 이를 해결하기 위해, 본 연구에서는 객체 및 배경 마스크 모듈과 로컬 어텐션 모듈을 설계하는 과정을 새롭게 제시하였다. 특히 저해상도 히트맵의

결과를 사용해서 객체 및 배경 마스크 맵을 생성하는 과정과 이를 활용하여 객체와 배경 영역에 각각 로컬 어텐션을 적용할 수 있는 아키텍처를 개발하였다. 실험 결과를 통해, 제안한 로컬 어텐션 기반의 MaDa-CenterNet(v2)가 기존의 MaDa-CenterNet의 성능을 개선할 수 있었으며 제안한 로컬 어텐션 모델이 해충 카운팅에 효과적임을 입증하였다.

References

- [1] S.-H. Oh, S.-W. Yang, H.-C. Kim, D.-H. Kim, and Y.-H. Doh, "Development of automated quantitative spray control system for high quality crop cultivation", The Journal of The Institute of Webcasting, Internet and Telecommunication, Vol, 17, No. 3, pp. 267-274, Jun. 2017. <https://doi.org/10.7236/JIIBC.2017.17.3.267>.
- [2] J.-W. Song, T.-K. An, M.-H. Kim, and Y.-S. Hong, "Measurement of the crowd density in outdoor using neural network", The Journal of The Institute of Webcasting, Internet and Telecommunication, Vol, 12, No. 2, pp. 103-110, Apr. 2012.

[3] 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>.

[4] 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, Oct. 2019.

[5] 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.

[6] 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>.

[7] 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>.

[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, H. Le, and M. Hoai, "Iterative crowd counting", in *Proc. European Conference on Computer Vision*, Munich, Germany, pp. 278-293, Oct. 2018.

[10] J. Wan, Q. Wang, and A. B. Chan, "Kernel-based density map generation for dense object counting", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 3, pp. 1357-1370, Mar. 2022. <https://doi.org/10.1109/TPAMI.2020.3022878>.

[11] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks", *Proc. of the IEEE International*

Conference on Computer Vision (ICCV), Venice, Italy, pp. 764-773, 2017.

[12] A. Dosovitskiy, et al., "An image is worth 16x16 words: Transformers for image recognition at Scale", *arXiv:2010.11929[cs.CV]*, Oct. 2020. <https://doi.org/10.48550/arXiv.2010.11929>.

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimzation", *arXiv:1412.6980[cs.LG]*, Dec. 2014. <https://doi.org/10.48550/arXiv.1412.6980>.

[14] 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

정재민 (Jae-Min Jeong)



2018년 3월 ~ 현재 : 군산대학교
소프트웨어학부 학사과정
관심분야 : 컴퓨터 비전, 영상처리,
기계학습, 딥 러닝

손창환 (Chang-Hwan Son)



2017년 4월 ~ 현재 : 군산대학교
소프트웨어학부 부교수
관심분야 : 컴퓨터 비전, 영상처리,
기계학습, 딥 러닝

홍서연 (Seoyeon Hong)



2020년 1월 ~ 현재 :
국립식량과학원 생산기술개발과
농업연구사
관심분야 : 해충 무인트랩,
방제기술