

# 특징 적응적 프롬프트 선택 및 특징 강화를 활용한 밀도 맵 기반의 해충 카운팅 모델

김경태\*, 손창환\*\*

## Density-based Pest Counting Model with Feature-Adaptive Prompt Selection and Feature Enhancement

Kyoung-Tae Kim\*, Chang-Hwan Son\*\*

이 성과물은 2026년도 정부(과학기술정보통신부)의 재원으로 연구개발특구진흥재단-(군산) 2026년 강소특구 이노테크 기업 육성 사업의 지원을 받아 수행된 연구임(RS-2026-25545611)

### 요 약

최근 해충 증가로 인한 농작물 피해를 줄이기 위해 효과적인 해충 모니터링 기술의 중요성이 커지고 있다. 본 연구에서는 디지털 트랩 환경에서 촬영된 해충 이미지의 마릿수를 정확하게 예측하기 위해 밀도 맵 기반 해충 카운팅 모델을 제안한다. 특히, 밀도 맵과 벡터 프롬프트 간의 의미적 정렬 학습 방법을 도입하여 두 특징 간의 연관성을 효과적으로 학습하도록 설계하였다. 또한 특징 적응적 프롬프트 선택 모듈과 특징 강화 모듈을 적용하여 사전 학습된 벡터 프롬프트를 입력 해충 이미지 특징에 적응적으로 결합함으로써 밀도 맵 추정 정확도를 향상시켰다. 실험 결과, 제안한 해충 카운팅 모델은 기존 객체 검출 기반 방법과 군중 카운팅 모델에 비해 평균 절대 오차 지표에서 더 우수한 성능을 보였다. 이러한 결과는 제안한 프롬프트 기반 특징 강화 방식이 해충이 밀집된 환경에서의 카운팅 정확도를 향상시키는 데 효과적임을 보여준다.

### Abstract

Recent increases in pest populations have heightened the importance of effective pest monitoring technologies to mitigate crop damage. In this study, we propose a density map-based pest counting model to accurately estimate the number of pests captured in images from digital trap environments. Specifically, we introduce a semantic alignment learning method between density maps and vector prompts to effectively capture the correlation between these features. In addition, a feature-adaptive prompt selection module and a feature enhancement module are applied to adaptively integrate pre-trained vector prompts with input pest image features, thereby improving the accuracy of density map estimation. Experimental results demonstrate that the proposed model outperforms conventional object detection-based methods and crowd counting models in terms of mean absolute error. These findings indicate that the proposed prompt-based feature enhancement approach is effective for improving counting accuracy in densely populated pest environments.

### Keywords

pest counting, prompt learning, stacked hourglass, crowd counting, density estimation

\* 국립군산대학교 소프트웨어학과 학사과정

- ORCID: <https://orcid.org/0009-0003-3512-4783>

\*\* 국립군산대학교 소프트웨어학과 교수(교신저자)

- ORCID: <https://orcid.org/0000-0001-7077-3074>

· Received: Mar. 24, 2026 Revised: Apr. 07, 2026, Accepted: Apr. 10, 2026

· Corresponding Author: Chang-Hwan Son

Dept. of Software Science and Engineering, Kunsan National University,  
South Korea

Tel.: 82+63-469-8915, Email: [cson@kunsan.ac.kr](mailto:cson@kunsan.ac.kr)

## I. 서 론

최근 무역의 확대와 급격한 기후변화로 인해 다양한 외래해충과 돌발해충의 유입이 증가하고 있다. 이러한 해충은 농작물의 품질 저하와 수확량 감소를 초래하여 농가 소득에 직접적인 피해를 발생시킨다. 따라서 효과적인 방제 전략을 수립하기 위해서는 해충의 발생 현황을 조기에 신속하게 파악하는 것이 중요하다. 최근에는 노지 환경에 디지털 트랩을 설치하여 해충을 유인한 후, 내장된 카메라로 촬영된 이미지에서 해충 개체 수를 자동으로 카운팅하는 기술이 활용되고 있다[1]. 이러한 지능형 카운팅 시스템은 해충 발생량을 주기적으로 모니터링할 수 있도록 하여 보다 효율적인 방제 전략 수립을 가능하게 하며, 궁극적으로 농작물의 안정적인 품질 관리와 농가 소득 증대에 기여할 수 있다.

해충 카운팅 모델은 크게 객체 검출(Object detection) 기반과 군중 카운팅(Crowd counting) 기반으로 구분될 수 있다[2]. 객체 검출은 이미지 내에서 해충과 같은 객체의 위치와 클래스 정보를 동시에 추정하는 기법으로, 일반적으로 객체의 위치는 바운딩 박스 형태로 표현된다. 그러나 기존 객체 검출 모델은 주로 자연 이미지에서 수십 개 수준의 객체를 대상으로 하며, 객체 간 폐색이나 자세 변형이 상대적으로 심하지 않은 환경을 가정하고 개발되어 왔다. 반면 디지털 트랩에 포획된 해충의 경우, 동일한 영상 내에 수백 마리에 이르는 개체가 밀집하여 존재할 수 있으며 심한 폐색과 다양한 자세 변형이 빈번하게 발생한다. 이러한 특성으로 인해 기존 객체 검출 기반 방법을 해충 카운팅에 직접 적용하는 데에는 한계가 있다.

이를 보완하기 위한 대안으로 군중 카운팅 기법을 해충 카운팅에 적용할 수 있다. 군중 카운팅 모델은 실내 콘서트홀이나 대형 광장과 같이 다수의 인파가 밀집한 환경에서 사람의 수를 추정하기 위해 개발된 방법으로, 높은 밀도의 객체가 존재하고 폐색이 빈번한 상황에서도 효과적으로 개체 수를 추정할 수 있다. 이러한 특성은 디지털 트랩에서 포획된 해충을 카운팅하는 문제와 유사한 특성을 가지므로, 기존의 군중 카운팅 기법을 해충 카운팅 분

야에 활용할 수 있다.

한편, 최근 이미지와 텍스트 간의 의미적 정렬을 학습할 수 있는 비전-언어(Vision-language) 모델이 활발히 연구되고 있다. 대표적인 모델로는 CLIP(Contrastive Language-Image Pretraining)[3]이 있다. CLIP은 이미지 인코더와 텍스트 인코더를 각각 활용하여 시각 특징과 텍스트 특징을 추출한 후, 대조 학습(Contrastive learning)을 통해 두 특징 간의 정렬을 학습한다. 그러나 기존 CLIP 모델은 입력 프롬프트의 표현 방식에 따라 성능이 크게 달라지는 한계를 가진다. 이러한 문제를 해결하기 위해 최근에는 벡터 기반 프롬프트 학습 기법이 제안되고 있으며, 대표적으로 CoOp(Context Optimization)[4] 과 같은 방법이 연구되고 있다.

본 연구에서는 CLIP과 같은 이미지-텍스트 특징 정렬 방식을 활용하여 밀도 맵 기반 해충 카운팅 모델을 설계하고자 한다. 기존의 CLIP 및 CoOp과 같은 벡터 기반 프롬프트 학습은 주로 이산적인 클래스 분류 문제에 적용되어 왔다. 그러나 해충 카운팅 문제의 경우 해충의 개체 수가 연속적인 값으로 표현되기 때문에 기존 방법을 그대로 적용하기에는 한계가 있다. 이를 해결하기 위해, 본 연구에서는 해충 카운팅 문제에 적용 가능한 정렬 학습 방식을 제안한다. 구체적으로, 이미지 인코더에는 해충 이미지 대신 밀도 맵을 입력으로 사용하고, 텍스트 인코더에는 해충 개체 수에 대응하는 클래스 정보를 벡터 기반 프롬프트 형태로 구성하여 입력한다. 이를 통해 해충 카운팅에 적합하도록 CoOp 구조를 확장한 새로운 아키텍처를 설계하고, 학습된 프롬프트를 밀도 맵 기반 카운팅 모델과 통합하는 방식을 제안한다. 실험 결과를 통해, 제안한 방법이 해충 카운팅 문제에서 효과적인 벡터 프롬프트 표현을 학습하며, 최종적으로 해충 카운팅의 정확도를 향상시킬 수 있음을 확인하였다.

## II. 관련 연구

### 2.1 군중 카운팅 방법

디지털 트랩에서 촬영된 해충 이미지는 수십 마

리에서 최대 수백 마리에 이르는 해충을 포함할 수 있다. 이러한 환경에서는 해충 간의 가려짐이 빈번하게 발생하며, 날개의 접힘이나 펼쳐짐과 같은 다양한 자세 변형이 나타날 수 있다. 이러한 특성은 대형 광장이나 공연장과 같이 밀집된 인파에서 사람의 수를 추정하는 군중 카운팅 문제와 유사한 특성을 가진다.

기존 군중 카운팅 기술은 크게 입력 이미지로부터 사람의 수를 직접 예측하는 회귀 기반 방식과, 입력 이미지로부터 밀도 맵을 추정하는 방식으로 구분될 수 있다. 밀도 맵은 각 픽셀이 해당 위치의 객체 밀도를 나타내며, 전체 픽셀 값을 합산함으로써 이미지에 포함된 총 객체 수를 추정할 수 있다. 최근 군중 카운팅 분야에서는 밀도 맵 추정 기반 방법이 주로 활용되고 있으며, 다중 스케일 특징 학습[5], 적응적 커널 생성[6], 그리고 손실 함수 모델링[7]과 같은 다양한 접근 방법들이 제안되고 있다. 대표적인 군중 카운팅 딥러닝 모델에는 MCNN[5], KDMG[6], DM-Count[7], ICCNet[8], SANet[9] 등이 있다.

## 2.2 밀도 맵 생성 방법

군중 카운팅 모델은 지도 학습 기반으로 학습되며 밀도 맵을 예측하기 때문에 정답 밀도 맵이 필요하다 [5]. 일반적으로 밀도 맵은 객체의 중심점을 흰색으로 표시한 도트 맵을 생성한 후, 각 중심점 위치에 가우시안 커널을 배치하여 생성된다. 이때 서로 인접한 중심점에 배치된 가우시안 커널은 공간적으로 중첩되며 그 값이 누적되어 최종 밀도 맵을 형성한다. 이러한 밀도 맵은 각 객체 중심점에 위치한 가우시안 함수들의 합으로 표현될 수 있다. 먼저, 도트 맵은 식 (1)과 같이 표현된다.

$$H(x) = \sum_{i=1}^N \delta(x - x_i) \quad (1)$$

여기서,  $H(x)$ 는 도트 맵이고  $\delta$ 는 디랙 델타 함수(Dirac delta function)이다. 그리고  $x$ 는 이미지 픽셀 인덱스이고,  $x_i$ 는  $i$ 번째 객체 중심 좌표의 인덱스에 해당한다. 도트 맵이 주어진 경우, 밀도 맵은 식 (2)

와 같이 계산된다.

$$D(x) = H(x) \otimes G(x) = \sum_{i=1}^N G(x - x_i) \quad (2)$$

여기서  $\otimes$ 는 컨볼루션 연산을 의미하므로, 최종 밀도 맵  $D(x)$ 는 도트 맵  $H(x)$ 와 가우시안 커널  $G(x)$ 의 컨볼루션 결과로 표현된다. 그리고  $H(x)$ 는 디랙 델타 함수로 구성된 도트 맵이므로, 이는 각 객체의 중심 위치에서 가우시안 커널을 배치한 후, 그 값을 공간적으로 중첩하여 누적한 것과 동일한 의미를 갖는다.

## III. 제안한 특징 적응적 프롬프트 선택 및 강화에 기반한 해충 카운팅 모델

### 3.1 밀도 맵과 벡터 프롬프트 간 정렬 학습

기존의 CLIP[3] 모델은 입력 이미지와 자연어 기반의 텍스트 프롬프트를 각각 이미지 인코더와 텍스트 인코더의 입력으로 사용한다. 그러나 해충 카운팅 문제에서 해충의 마릿수는 원본 이미지보다 밀도 맵과 더 직접적인 연관성을 가진다. 이는 밀도 맵이 객체의 중심점을 기반으로 생성되며, 모든 픽셀 값을 합산하면 이미지에 포함된 객체의 총 개수를 나타내기 때문이다. 따라서 본 연구에서 제안하는 이미지-텍스트 의미적 정렬 학습 방식에서는 원본 이미지 대신 밀도 맵을 이미지 인코더의 입력으로 사용한다. 또한 CLIP 모델에서 사용되는 자연어 기반 텍스트 프롬프트는 사용자의 표현 방식에 따라 문장이 달라질 수 있으며, 이러한 차이는 모델의 최종 성능에 민감한 영향을 미칠 수 있다. 이를 완화하기 위해 본 연구에서는 CoOp[4]과 같은 벡터 기반 프롬프트 학습 방법을 도입한다. 아울러 기존 CLIP 기반 방법에서 클래스는 일반적으로 이미지의 객체 유형을 기준으로 정의되지만, 본 연구에서는 해충 이미지의 종류가 아닌 해충의 마릿수를 기준으로 클래스를 정의한다. 구체적으로, 해충 이미지에 포함된 개체 수를 기준으로 총  $M$ 개의 범주로 그룹화하여 클래스를 구성한다. 그림 1은 제안한 밀

도 맵과 벡터 프롬프트 간 정렬 학습 구조를 보여준다. 그림 1의 좌측 하단에는 전체 해충 마릿수를 세 개의 그룹으로 나눈 예시가 제시되어 있다. 또한 좌측 상단과 같이, 각 그룹에 대응하는 학습 가능한 벡터 프롬프트를 설계하였다. 그림 1의 밀도 맵 그룹과 벡터 프롬프트 간의 구체적인 정렬 과정은 다음과 같다.

먼저, 입력 해충 이미지로부터 밀도 맵을 생성한 후, 식 (3)과 같이 이미지 인코더를 통해 시각 특징 벡터로 변환된다.

$$v_i = f(D_i^c) \quad (3)$$

여기서  $f$ 는 이미지 인코더이고,  $v_i$ 는  $i$ 번째 밀도 맵  $D_i^c$ 에 대응하는 시각 특징 벡터이다.  $c$ 는 밀도 맵이 속한 그룹을 의미한다. 각 그룹은 해충 마릿수를 기준으로 등간격으로 구간을 나누어 정의되며, 전체 그룹의 개수는  $M$ 개로 구성된다.

둘째, 텍스트 특징 벡터는 식 (4)와 같이 벡터 프롬프트를 텍스트 인코더에 입력하여 생성한다.

$$t_c = g(P^c) \quad (4)$$

여기서  $P^c$ 는 클래스  $c$ 에 대응하는 학습 가능한 벡터 프롬프트를 나타낸다. 즉,  $P^c$ 는  $[V_1, V_2, \dots, V_{16}]$ 와

$[CLASS]$ 로 구성된 벡터 시퀀스이다.  $g$ 는 텍스트 인코더를,  $t_c$ 는 텍스트 특징 벡터를 의미한다.

셋째, 텍스트 특징 벡터와 시각 특징 벡터 간의 의미적 정렬을 위해 두 특징 간의 코사인 유사도를 계산한다.

$$s_{i,c} = \frac{v_i \cdot t_c}{\|v_i\| \|t_c\|} \quad (5)$$

식 (5)는 코사인 유사도를 계산하는 수식으로, 두 특징 벡터인  $v_i$ 와  $t_c$ 를 내적( $\cdot$ )하여 각각의 벡터의 크기(norm)으로 나눈 값으로 정의된다.

마지막으로, 정렬 학습을 위해 유사도 값을 소프트맥스 함수에 통과시켜 확률 분포를 계산한 후, 교차 엔트로피 손실 함수를 적용하여 학습을 수행한다. 확률 변환 과정은 식 (6)과 같다.

$$p_{i,c} = \frac{\exp(s_{i,c}/\tau)}{\sum_{k=1}^M \exp(s_{i,k}/\tau)} \quad (6)$$

여기서  $\tau$ 는 소프트맥스 함수의 모양을 제어하기 위한 파라미터이고,  $p_{i,c}$ 는  $i$ 번째 밀도 맵에 대응하는 시각 특징 벡터와  $c$ 번째 클래스에 대응하는 텍스트 벡터 간의 유사도를 확률로 표현한 값이다.

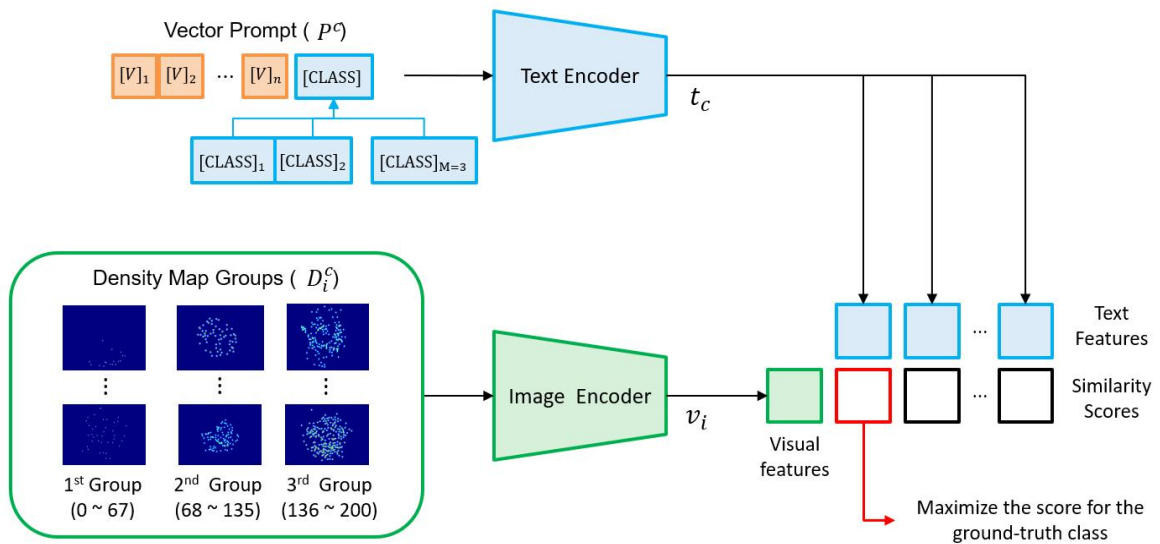


그림 1. 제안한 밀도 맵 그룹과 벡터 프롬프트 간의 정렬 학습의 개요

Fig. 1. Overview of the proposed alignment learning between density map groups and vector prompts

이러한  $p_{i,c}$ 는 교차 엔트로피 손실 함수를 통해 학습되며, 동일한 클래스에 해당하는 특징 벡터 간의 유사도는 높아지고 서로 다른 클래스 간의 유사도는 낮아지도록 최적화된다. 이를 통해, 밀도 맵 특징과 벡터 프롬프트 특징 간의 의미적 정렬 관계를 학습할 수 있다.

### 3.2 제안한 특징 적응적 프롬프트 선택 및 특징 강화 모듈 기반의 해충 카운팅 구조

그림 2는 제안한 해충 카운팅 모델의 전체 구조를 보여준다. 모델은 크게 모래시계(Hourglass)[10] 백본을 스택 형태로 쌓은 구조이다. 또한, 사전 학습된 벡터 프롬프트를 활용하여 입력 해충 이미지 특징에 적응적으로 최적의 프롬프트를 선택 및 조합함으로써 특징을 강화한다. 이러한 벡터 프롬프트는 밀도 맵과의 정렬 학습을 통해 획득된 것이므로, 최종 정답 밀도 맵을 보다 정확하게 추정하는 데 도움을 줄 수 있다.

첫째, 특징 적응적 프롬프트 선택 모듈 (Feature-adaptive feature selection module)은 모래시계 백본 내부에서 생성된 입력 해충 이미지 특징  $F$ 에 적응적인 최적의 벡터 프롬프트를 재구성한다. 식

(7)은 특징 적응적 프롬프트 선택 모듈의 연산 과정을 나타낸다.

$$P = FAPS(F) = \omega_1 P^1 + \omega_2 P^2 \dots + \omega_c P^c \quad (7)$$

여기서  $FAPS$ 는 특징 적응적 프롬프트 선택 모듈을 의미하고 재구성된 벡터 프롬프트  $P$ 는 사전 학습된 클래스별 벡터 프롬프트와 가중치와의 선형 결합으로 표현된다. 이때 가중치  $\omega$ 는 입력 해충 이미지의 특징  $F$ 에 따라 적응적으로 결정된다. 즉, 전역 평균 풀링(Global Average Pooling)과 컨볼루션, 그리고 시그모이드 함수를 통과하여 [0-1] 사이의 가중치 값으로 변환된다.

둘째, 특징 강화 모듈 (Feature Enhancement Module)은  $FAPS$ 로 재구성된 벡터 프롬프트  $P$ 를 컨볼루션 연산을 적용하여 특징 표현의 유연성을 강화한다. 이후, 모래시계 백본의 내부 특징인  $F$ 와 융합되어 밀도 맵 추정에 적합한 특징 표현으로 변환된다. 식 (8)은 특징 강화 모듈의 연산 과정을 나타낸다.

$$\hat{F} = FE(F, P) = f_{c,k} f_{c,k} f_{ViT} f_{c,k} (Concat(F, P)) \quad (8)$$

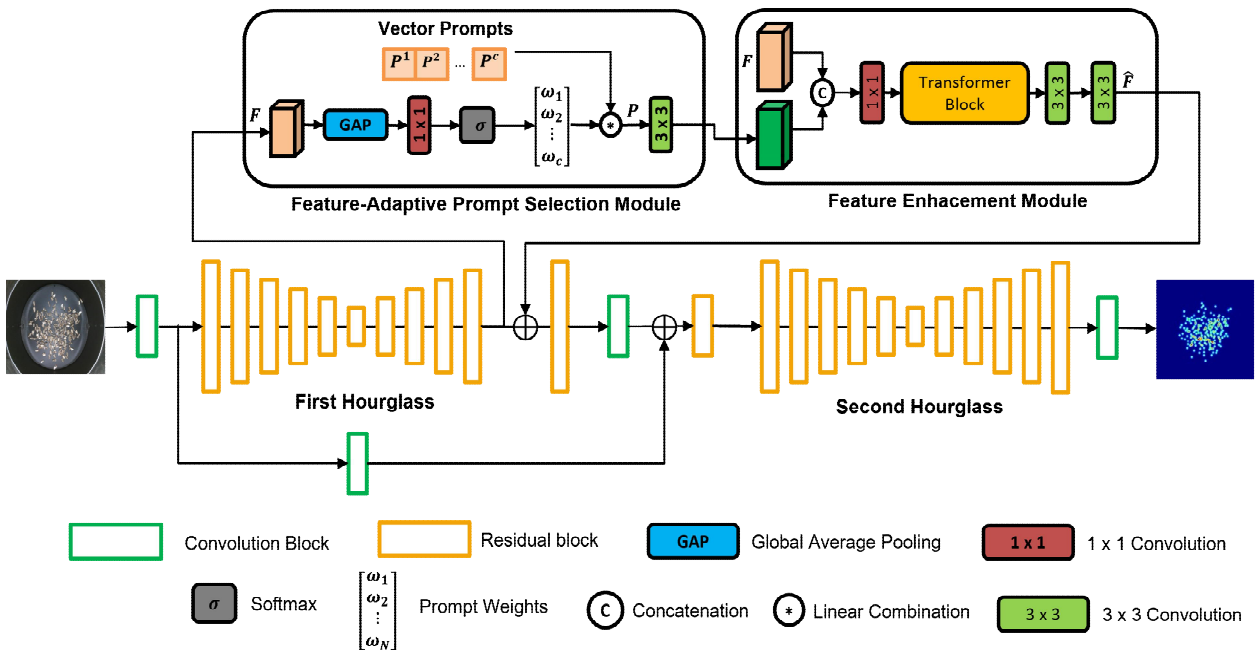


그림 2. 제안한 특징 적응적 프롬프트 선택과 특징 강화 모듈 기반 해충 카운팅 모델

Fig. 2. Proposed pest counting model with feature-adaptive prompt selection and feature enhancement modules

여기서, *Concat*는 연결(Concatenation) 계층으로서 채널 축 방향으로 입력 특징들을 결합한다. 그리고  $F_{c,k}$ 는 필터 크기  $k$ 를 갖는 컨볼루션 계층이고  $F_{ViT}$ 는 어텐션 연산을 포함한 트랜스포머 블록이다. 따라서, 특징 강화 모듈을 통해, 입력 해충 이미지의 특징  $F$ 는 밀도 맵과 정렬된 프롬프트 정보와 융합함으로써, 최종 밀도 맵을 추정하는데 더욱 효과적인 특징  $\hat{F}$ 으로 변환된다.

### 3.3 손실 함수

학습을 위해, 그림 2의 제안한 해충 카운팅 구조에서 각 모래시계의 마지막 단계 손실 계층을 추가하였다. 손실 함수는 추정한 밀도 맵과 정답 밀도 맵 간의 오차로 정의되며, 본 연구에서는 평균 제곱 오차(Mean Squared Error, MSE)를 사용하였다. MSE는 식 (9)와 같이 정의된다.

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N \| \hat{D}_i - D_i \|^2 \quad (9)$$

여기서  $\hat{D}_i$ 는 모델이 예측한 밀도 맵이며,  $D_i$ 는 정답 밀도 맵이다. 수식에서 보듯이, 평균 제곱 오차는 두 밀도 맵 간의 픽셀 오차의 합을 계산함으로써, 픽셀 수준에서 오차를 측정함을 알 수 있다.

## IV. 실험 및 결과

### 4.1 실험 환경

학습 데이터는 디지털 트랩에서 촬영된 해충 이미지 8,504장을 사용하였으며, 이를 7:3의 비율로 랜덤하게 분할하여 훈련 집합과 테스트 집합으로 구성하였다. 학습 파라미터 갱신을 위해 아담 옵티마이저(Adam Optimizer)[11]를 사용했고 배치 크기는 5, 에폭은 100회, 학습률은  $1.25e-4$ 로 설정했다. 제안한 모델의 성능을 평가하기 위해, 기존 객체 검출과 군중 카운팅 모델인 RetinaNet[12], Faster RCNN[13], DM-Count[7], KDMG[6], Stacked Hourglass[10]와 비교 실험을 수행하였다. 제안한 해충 카운팅의 모래

시계 백본은 입력 이미지를  $512 \times 512$ 의 동일한 크기로 변환한다. 실험은 PyTorch 프레임워크 기반으로 구현되었으며, NVIDIA GeForce RTX 4090 GPU 환경에서 수행되었다.

### 4.2 정량적 평가

정량적 평가를 위해, 군중 카운팅 분야에서 널리 사용되는 평균 절대 오차(MAE: Mean Absolute Error)와 평균 제곱근 오차(RMSE: Root Mean Squared Error)를 사용하였다. MAE는 실제 해충 개수와 예측된 해충 개수 간의 절대 오차의 평균을 의미하며, RMSE는 실제 해충 개수와 예측된 해충 개수 간의 오차 제곱의 평균에 대한 제곱근으로 정의된다. RMSE는 오차가 큰 샘플에 더 큰 가중치를 부여하므로, 해충 개수가 많은 이미지에서 카운팅 정확도를 보다 엄격하게 평가할 수 있다. MAE와 RMSE는 각각 식 (10)과 식 (11)로 정의된다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{p}_i - p_i| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{p}_i - p_i)^2} \quad (11)$$

여기서  $n$ 은 테스트 이미지 개수,  $\hat{p}_i$ 와  $p_i$ 는 각각 모델이 예측한 해충 개수와 사람이 계측한 해충의 개수이다. 수식에서 보듯이, 두 지표 모두 수치가 작을수록 카운팅의 정확도가 높다는 것을 의미한다. 그리고  $\hat{p}_i$ 는 모델에서 추정된 밀도 맵의 모든 픽셀 값을 합산해서 획득된다.

표 1은 해충 카운팅의 정량적 평가 결과를 보여준다. 표에서 확인할 수 있듯이, 바운딩 박스를 기반으로 객체를 검출하는 Faster RCNN과 RetinaNet 모델은 MAE 값이 상대적으로 높게 나왔으며, 해충이 조밀하게 밀집된 환경에서의 카운팅 예측에는 적합하지 않음을 알 수 있다. 반면, 기존 군중 카운팅 모델인 DM-Count, KDMG, Stacked Hourglass는 객체 검출 기법에 비해 상대적으로 더 정확한 예측 성능을 보여 주었다. 이러한 결과를 통해 디지털 트



랩에서 획득된 해충 이미지의 마릿수 예측에는 군중 카운팅 기반 접근 방법이 보다 적합함을 알 수 있다. 또한 제안한 모델은 MAE와 RMSE 평가지표 모두에서 기존 군중 카운팅 모델보다 우수한 성능을 보였다. 특히 제안한 모델은 Stacked Hourglass 구조를 기반으로 특징 적응적 프롬프트 선택 모듈과 특징 강화 모듈을 추가한 구조로 볼 수 있다. Stacked Hourglass 모델과 비교했을 때 MAE와 RMSE가 모두 감소하는 것을 확인할 수 있으며, 이는 밀도 맵과 사전 정렬된 벡터 프롬프트가 입력 해충 이미지 특징을 효과적으로 강화하는 데 기여함을 의미한다.

표 1. 해충 카운팅을 위한 정량적 평가

Table 1. Quantitative evaluation for pest counting

| Counting models        | MAE   | RMSE  |
|------------------------|-------|-------|
| Faster RCNN[13]        | 26.4  | 16.3  |
| RetinaNet[12]          | 34.6  | 48.7  |
| DM-Count [7]           | 2.316 | 3.902 |
| KDMG [6]               | 2.368 | 4.256 |
| Stacked hourglass [10] | 2.310 | 4.510 |
| Proposed model         | 2.012 | 3.675 |

그림 3은 제안한 밀도 맵과 벡터 프롬프트 간 정렬 학습 결과를 t-SNE(t-distributed Stochastic Neighbor Embedding)[14]를 이용하여 시각화한 것이다. 해충 마릿수를 세 개의 클래스로 구분한 경우, 각 클래스에 대응하는 프롬프트 벡터들이 서로 유사한 영역에 군집을 형성하는 것을 확인할 수 있다. 이는 프롬프트 벡터와 밀도 맵 간의 의미적 정렬이 효과적으로 학습되었음을 보여준다.

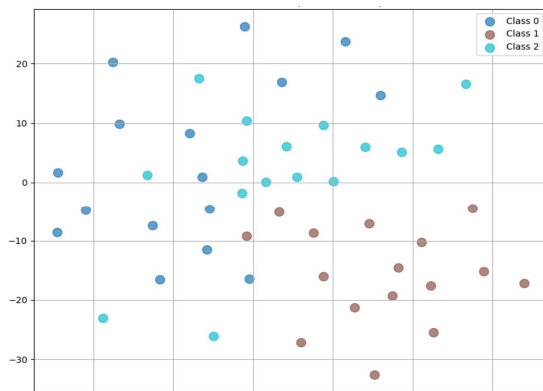


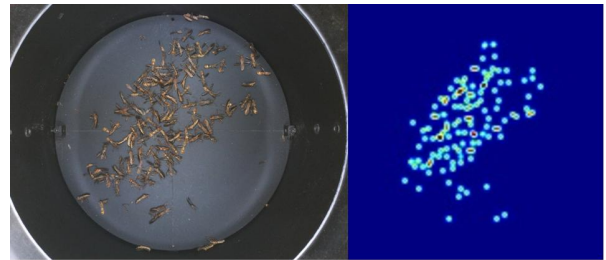
그림 3. 사전 학습된 벡터 프롬프트 시각화  
Fig. 3. Visualization of the pretrained vector prompts

#### 4.3 밀도 맵 시각화

그림 4는 제안한 해충 카운팅 모델을 통해 예측된 밀도 맵의 예시를 보여준다. 각 해충 이미지의 하단에는 육안으로 측정된 실제 해충 개수와 모델이 예측한 해충 개수가 함께 제시되어 있다. 그림에서 확인할 수 있듯이, 해충 이미지는 개체 간 가려짐 현상이 심하고 다양한 자세를 보이는 특징이 있다. 그럼에도 불구하고 그림 하단의 결과에서 실제 해충 개수와 예측된 해충 개수가 거의 일치하는 것을 확인할 수 있다. 이러한 결과는 군중 카운팅 기법이 해충 카운팅 문제에 효과적으로 적용될 수 있음을 보여준다. 특히 제안한 방법은 밀도 맵과 사전 정렬된 벡터 프롬프트를 입력 해충 이미지 특징에 적응적으로 결합함으로써, 해충 밀집 환경에서도 높은 카운팅 정확도를 유지할 수 있음을 확인할 수 있다.



Measured pest count: 199  
Predicted pest count : 199.532



Measured pest count: 120  
Predicted pest count : 120.016

그림 4. 입력 해충 이미지와 예측된 밀도 맵  
Fig. 4. Input pest images and their predicted density maps

#### 4.4 애블레이션 실험

클래스 개수에 따른 해충 카운팅 정확도의 변화를 분석하기 위해 애블레이션 실험을 수행하였다. 표 2에서 확인할 수 있듯이, 클래스 개수가 5일 경

우 RMSE 값이 가장 낮게 나타났으며, 클래스 개수가 7일 경우 MAE 값이 가장 낮게 나타났다. 그러나 클래스 개수가 5 이상인 경우 전반적인 성능 차이는 크지 않은 것으로 나타났다. 이러한 결과는 프롬프트 클래스의 수가 일정 수준 이상 증가할 경우, 밀도 맵과 프롬프트 간의 정렬 학습이 안정적으로 이루어져 카운팅 성능이 유사한 수준으로 수렴함을 의미한다.

표 2. 클래스 개수에 따른 해충 카운팅 정확도

Table 2. Pest counting accuracy according to the number of classes

| Class number | MAE   | RMSE  |
|--------------|-------|-------|
| 3            | 2.012 | 3.675 |
| 5            | 1.967 | 3.433 |
| 7            | 1.964 | 3.592 |
| 10           | 1.967 | 3.593 |

## V. 결 론

본 연구에서는 디지털 트랩 환경에서 촬영된 해충 이미지의 마릿수를 정확하게 예측하기 위해 밀도 맵 기반 해충 카운팅 모델을 제안하였다. 제안한 방법은 밀도 맵을 이미지 인코더의 입력으로 활용하고, 해충 마릿수에 따라 정의된 클래스 기반 벡터 프롬프트를 텍스트 인코더에 입력하여 두 특징 간의 의미적 정렬을 학습하도록 설계하였다. 또한 특징 적응적 프롬프트 선택 모듈과 특징 강화 모듈을 도입하여 사전 학습된 벡터 프롬프트를 입력 해충 이미지 특징에 적응적으로 결합함으로써 밀도 맵 추정 성능을 향상시켰다. 실험 결과, 제안한 방법은 기존 객체 검출 기반 방법뿐만 아니라 기존 군중 카운팅 모델에 비해서도 MAE와 RMSE 지표에서 우수한 성능을 보였다. 특히, 기반 모델인 모래시계 카운팅 기법 대비 약 12.9%의 성능 향상을 보였다. 또한 프롬프트 벡터의 t-SNE 시각화를 통해 밀도 맵과 프롬프트 간의 의미적 정렬이 효과적으로 학습되었음을 확인하였다. 이러한 결과는 제안한 프롬프트 기반 특징 강화 방식이 해충이 밀집된 환경에서도 카운팅 정확도를 향상시키는 데 효과적임을 보여준다.

그러나 본 연구는 디지털 트랩에 포획된 해충의 종류를 단일 클래스라고 가정한 한계가 있다. 실제 환경에서는 다양한 종류의 해충이 동시에 포획될 수 있으므로, 이러한 복잡한 상황을 충분히 반영하지 못한다. 따라서 향후 연구에서는 제안한 방법을 다중 클래스 해충 카운팅 문제로 확장하고자 한다. 또한 밀도 맵과 벡터 프롬프트 간의 정렬 학습을 더욱 정교하게 수행하기 위해, 다층 프롬프트 및 앙상블 프롬프트 기반의 확장된 접근 방법을 개발할 계획이다.

## Acknowledgment

'2025년도 한국정보기술학회 추계종합학술대회에서 발표한 논문(다중 계층 프롬프트 학습을 이용한 해충 밀도맵 기반 카운팅 모델)[15]을 확장한 것임'

## References

- [1] 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>.
- [2] 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, Art. no. 3810, Jul. 2023. <https://doi.org/10.3390/rs15153810>.
- [3] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision", Proc. International Conference on Machine Learning, Online, pp. 8748-8763, Feb. 2021. <https://doi.org/10.48550/arXiv.2103.00020>.
- [4] K. Zhou, J. Yang, C. C. Loy, and Z. Liu, "Learning to prompt for vision-language models",



- International Journal of Computer Vision, Vol. 130, No. 7, pp. 2337-2348, Jul. 2022. <https://doi.org/10.1007/s11263-022-0165>.
- [5] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network", Proc. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 589-597, Jun. 2016. <https://doi.org/10.1109/CVPR.2016.70>.
- [6] 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>.
- [7] B. Wang, H. Liu, D. Samaras, and M. Hoai, "Distribution matching for crowd counting", Proc. Neural Information Processing Systems, Vancouver, Canada, pp. 1595-1607, Dec. 2020. <https://doi.org/10.48550/arXiv.2009.13077>.
- [8] X. Cao, Z. Wang, Y. Zhao, and F. Su, "Scale aggregation network for accurate and efficient crowd counting", Proc. European Conference on Computer Vision, Munich, Germany, pp. 757-773, Oct. 2018. [https://doi.org/10.1007/978-3-030-01228-1\\_45](https://doi.org/10.1007/978-3-030-01228-1_45).
- [9] V. Ranjan, Hieu. Le, M. Hoai, "Iterative crowd counting", Proc. European Conference on Computer Vision, Munich, Germany, pp. 278-293, Oct. 2018. [https://doi.org/10.1007/978-3-030-01234-2\\_17](https://doi.org/10.1007/978-3-030-01234-2_17).
- [10] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation", Proc. European Conference on Computer Vision, Amsterdam, Netherlands, pp. 483-499, Sep. 2016. <https://doi.org/10.48550/arXiv.1603.06937>.
- [11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimzation", Proc. International Conference on Learning Representation, San Diego, USA, May 2015. <https://doi.org/10.48550/arXiv.1412.6980>.
- [12] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection", Proc. IEEE International Conference on Computer Vision, Venice, Italy, pp. 2999-3007, Oct. 2017. <https://doi.org/10.1109/ICCV.2017.324>.
- [13] 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>.
- [14] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE", Journal of Machine Learning Research, Vol. 9, pp. 2579-2605, 2008. <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- [15] K.-T. Kim and C.-H. Son, "Pest counting model using density maps based on multi-level prompt learning", Proc. of KIIT Conference, Jeju Island, South Korea, pp. 841-844, Dec. 2025.

## 저자소개

### 김 경 태 (Kyoung-Tae Kim)



2026년 2월 : 국립군산대학교  
소프트웨어학과(학사)  
관심분야 : 컴퓨터 비전, 영상처리,  
기계학습, 딥 러닝

### 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 교수  
관심분야 : 컴퓨터 비전, 영상처리, 기계학습, 딥 러닝