

# 트랩 기반 해충 카운팅을 위한 멀티스케일 및 변형 가능한 밀도 어텐션 모델

이원도\*, 손창환\*\*, 박성웅\*\*\*

## Multiscale and Deformable Density Attention Model for Trap-based Pest Counting

Won-Do Lee\*, Chang-Hwan Son\*\*, and Sung-Woong Park\*\*\*

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

### 요약

본 연구에서는 트랩에 포획된 해충의 마릿수를 자동으로 측정하기 위한 해충 카운팅 모델을 제안하고자 한다. 해충은 농작물에 피해를 주기 때문에 조기 방제가 아주 중요하다. 이를 위해, 최근 해충 발생량을 시기별로 측정하기 위해 디지털 트랩이 활용되고 있다. 하지만 고전적인 카운팅 방식은 사람의 육안으로 측정하기 때문에 상당한 시간과 비용이 요구되었다. 이를 해결하기 위해, 본 연구에서는 기존 군중 카운팅 기술을 해충 카운팅 분야에 적용 및 확장하는 방안을 소개하고자 한다. 특히, 해충 카운팅의 걸림돌이 되는 해충 포즈의 다양성과 스케일 및 폐색 문제를 극복할 수 있는 새로운 해충 카운팅 모델을 제안하고자 한다. 실험 결과를 통해서, 제안한 해충 카운팅 모델이 트랩에서 포획된 해충 개수를 정확히 추정할 수 있음을 검증하였다. 그리고 기존의 해충 카운팅 모델들과 비교해서, 평균 제곱 오차에서 26% 이상의 정확도를 개선할 수 있었다.

### Abstract

In this study, we propose a pest counting model to automatically measure the number of pests caught in a trap. Pests damage crops, so early control is very important. To this end, digital traps have recently been utilized to measure the amount of pests generated over time. However, since the classical counting method is measured with the human eye, considerable time and cost are required. To solve this problem, this study introduces a method to apply and expand the existing crowd counting technology to the field of pest counting. In particular, we would like to propose a new pest counting model that can overcome the problems of diversity of pest poses, scale, and occlusion, which are obstacles to pest counting. Through the experimental results, it was verified that the proposed pest counting model can accurately estimate the number of pests captured in the trap. Also, compared to the existing pest counting models, it was possible to improve the accuracy of more than 26% in the mean square error.

### Keywords

deep learning, pest counting, crowd counting, density estimation, deformable convolution, attention

\* 군산대학교 소프트웨어학부 소프트웨어학전공 학사과정 · Received: Aug. 17, 2023, Revised: Sep. 07, 2023, Accepted: Sep. 10, 2023  
- ORCID: <https://orcid.org/0009-0003-7067-465X> · Corresponding Author: Chang-Hwan Son  
\*\* 군산대학교 소프트웨어학부 교수(교신저자) · Software Science and Engineering Major, School of Software, Kunsan National University, Korea  
- ORCID: <https://orcid.org/0000-0001-7077-3074> · Tel.: 82+63-469-8915, Email: [cson@kunsan.ac.kr](mailto:cson@kunsan.ac.kr)  
\*\*\* 비엔에스코퍼레이션(주) 연구소장  
- ORCID: <https://orcid.org/0009-0004-1587-8024>

## 1. 서론

해충은 농작물의 잎, 줄기, 꽃, 열매 등에 직접적인 피해를 주고 병원성 질병을 전파한다. 이는 해충 발생량이 농작물 수확량에 아주 큰 영향을 미칠 수 있음을 나타낸다. 시기별 해충 발생량을 파악하면 특정 시기에 발생하는 해충에 대한 조기 방제가 가능하며 농작물 피해 감소와 생산량 증대를 유도할 수 있다[1]. 해충 발생량은 농부나 전문가가 트랩에 포집된 해충의 개수를 직접 카운팅 함으로써 파악할 수 있다. 그러나, 이런 육안 기반의 측정 방식은 시간과 비용 측면에서 비효율적이고 정확도 관점에서 개인 편차가 발생하므로 해충 개수를 자동으로 예측할 수 있는 해충 카운팅 기술이 필요하다. 따라서 본 연구에서는 주광성 기반의 디지털 트랩에 포집된 해충 이미지로부터 해충 개수를 자동으로 측정할 수 있는 해충 카운팅 딥러닝 모델을 제안하고자 한다.

기존의 객체 카운팅(Object counting) 기술은 크게 객체 검출(Object detection)과 밀도 추정(Density estimation) 방식으로 나눌 수 있다. 객체 검출 방식은 이미지에서 객체가 존재하는 위치를 바운딩 박스(Bounding box)로 검출하는 방식을 말하며 검출된 총 바운딩 박스의 개수가 바로 해당 객체의 개수가 된다. 하지만 트랩 내부에 해충이 많을수록 폐색 문제가 발생하기 때문에 객체 검출 성능이 저하될 수 있다. 반면 밀도 추정 방식은 밀도 맵을 추정하여 객체의 개수를 산출하기 때문에 객체의 개수가 많고, 인접해 있는 경우에 아주 적합하다. 예를 들면, 군중 카운팅(Crowd counting) 분야를 들 수 있다.

해충 카운팅의 경우, 해충 발생량에 따라 트랩에 포획된 해충들이 서로 인접하거나 폐색될 수 있다. 이런 경우 해충 이미지는 기존의 군중 이미지와 유사한 현상을 띄게 된다. 즉, 해충 카운팅 문제가 기존의 군중 카운팅 문제와 유사해진다. 따라서 본 연구에서는 군중 카운팅에서 사용되는 밀도 추정 방식을 해충 카운팅에 적용 및 확장하고자 한다. 특히, 해충 카운팅에서 주요 걸림돌인 폐색 문제, 스케일 변화, 포즈의 다양성, 색상과 텍스처의 유사도 문제를 극복할 수 있는 멀티스케일 및 변형 가능한 밀도 어텐션 해충 카운팅 모델(MaD2a-PCM, Multiscale and Deformable Density Attention Pest Counting

Model)을 소개하고자 한다.

제안한 해충 카운팅 모델은 기존 해충 카운팅 모델과 다음과 같은 차이점이 있다. 첫째, 제안한 모델에서는 해충 스케일 문제를 해소하기 위해, 멀티스케일 특징 추출을 위한 저해상도 모래시계(Low-resolution hourglass[2])와 고해상도 모래시계(High-resolution hourglass) 두 종류의 백본을 사용한다. 둘째, 고해상도 모래시계 백본의 특징 추출력을 강화하기 위해, 모래시계 백본 간의 특징 정보를 전송할 수 있도록 모래시계 간 스킵 연결(Between-hourglass skip connection)을 고안하였다. 셋째, 해충의 포즈 및 폐색 문제를 해결하기 위해, 변형 가능한 합성곱 계층(Deformable convolution layer)[3]을 활용한다. 특히, 제안한 모델에서는 모래시계 간 스킵 연결상에서 저해상도 모래시계 백본에서 생성된 저해상도 특징에 변경 가능한 합성곱 연산을 적용한 방안을 제시하였다. 넷째, 최종 밀도 맵 추정의 정확도를 개선하기 위해, 모래시계 간 스킵 연결에서 전송된 저해상도 특징을 쿼리(Query)로 사용하고, 고해상도 모래시계 내부에서 생성된 특징을 키(Key)와 값(Value)으로 활용한 어텐션 기능을 모델링하였다.

본 연구를 통해, 기존 군중 카운팅 기술에 해충 카운팅 분야에 응용 및 확장함으로써 해충 카운팅 문제를 해결할 수 있음을 확인하고자 한다. 그리고 제안한 해충 카운팅 모델이 기존의 첨단 군중 카운팅 기법보다 더 우수함을 검증하고자 한다.

## II. 군중 카운팅 관련 연구

기존 군중 카운팅 기술은 입력 이미지에서 회귀 모델을 통해서 객체 수를 직접 추정하는 방식[4]과 밀도 맵을 추정하여 간접적으로 객체 수를 산출하는 두 종류의 방식이 있다. 하지만 최근 군중 카운팅 기술 개발 추세는 밀도 추정 방식으로 전환되고 있다. 밀도 추정 기반 군중 카운팅 기술은 밀도 맵을 추정하여 객체 수를 산출하기 때문에 정답 밀도 맵 생성 과정이 요구된다. 밀도 맵 생성 과정은 먼저 입력 이미지에서 객체가 존재하는 중심점에 화이트 색상을 할당하여 도트 맵(Dot map)을 생성한다.

그리고 가우시안 필터링(Gaussian filtering)을 적용하여 최종 밀도 맵으로 변환한다. 이때, 가우시안 필터의 합은 1로 제약되므로 밀도 맵의 총 픽셀의 합은 객체의 개수와 동일하게 된다. 밀도 맵 생성 과정은 수식적으로 다음과 같이 표현된다.

$$Y = D * k_{\sigma} \quad (1)$$

여기서  $D$ 는 도트 맵,  $k_{\sigma}$ 는 표준편차가  $\sigma$ 인 가우시안 커널이며  $*$ 는 합성곱 연산이다. 실제 수식 (1)의 구현은 가우시안 필터를 도트 맵의 화이트 색상 위치에 오버레이 합으로써 구현된다.

정답 밀도 맵 생성 후, 밀도 추정 기반 군중 카운팅 기술은 학습 이미지를 모델에 통과한 후, 추정된 밀도 맵과 정답 밀도 맵과의 오차를 계산한다. 주로 평균 제곱 오차(MSE, Mean Squared Errors)가 손실 함수로 사용된다. 그리고 역전파와 최적화 기법을 적용해서 모델의 파라미터를 갱신한다.

기존의 대표적인 밀도 추정 기반 군중 카운팅 모델로는 MCNN, ICCNet[6], SANet[7], KDMG[8] 등이 있다. 그림 1은 가장 초기 군중 카운팅 모델인 MCNN의 아키텍처를 보여주고 있다. MCNN은 군중

이미지에서 객체의 스케일 변화 문제를 효율적으로 대응하기 위해 개발된 모델이다. 그림 1에서 보듯이, MCNN은 삼중 분기로 구성된다. 그리고 각 분기의 서브네트워크에서는 합성곱 블록으로 구현되고 서로 다른 크기의 합성곱 필터를 적용한 것을 볼 수 있다. 그림에서 각 열이 하나의 분기에 해당한다. Conv\_k\_c\_s는 합성곱 연산을 뜻하며 k는 커널의 크기, c는 채널 수, s는 스트라이드를 의미한다. 따라서 MCNN의 각 서브네트워크는 서로 다른 객체의 스케일을 담당하는 역할을 한다.

### III. 해충 카운팅을 위한 제안한 변형 가능한 멀티스케일 밀도 어텐션 모델

#### 3.1 제안한 접근 방법

본 연구에서는 군중 이미지와 해충 이미지에서 객체가 분포하는 양상이 유사하다는 것에 착안해서 밀도 추정 기반의 군중 카운팅 기술을 해충 카운팅 분야에 적용하고자 한다. 그림 2는 군중 이미지와 해충 이미지를 비교해 준다.

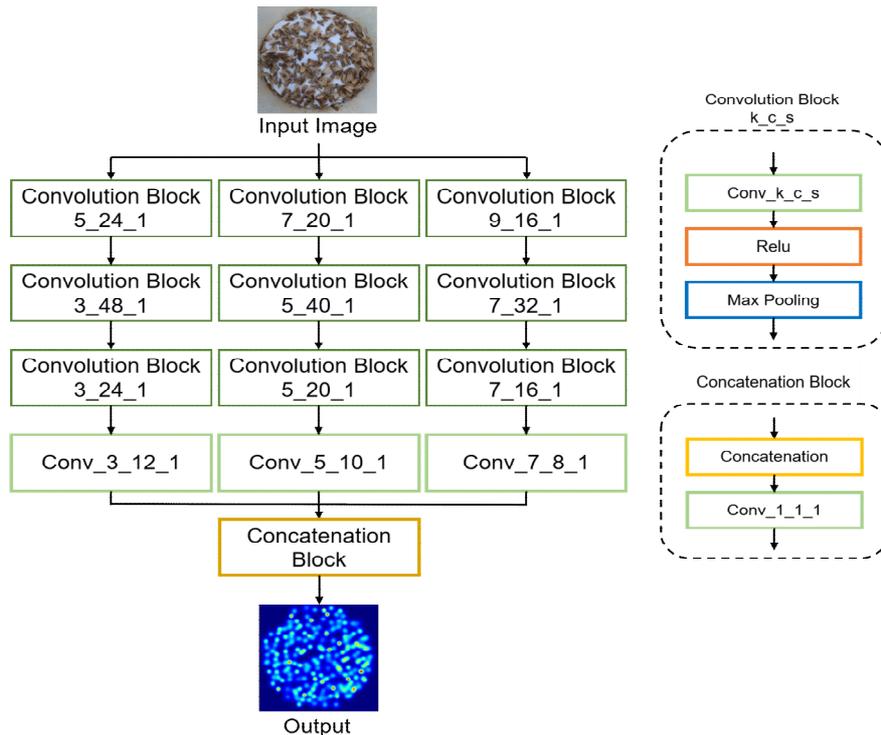


그림 1. MCNN 아키텍처[5]

Fig. 1. Architecture of MCNN[5]

군중 이미지는 기존 객체 검출을 위한 데이터셋과는 달리, 이미지 내에서 상당히 많은 객체를 지닌다. 이는 객체가 서로 인접해 있음을 의미하고 폐색 문제가 발생할 확률이 높다는 것을 의미한다. 따라서 기존의 바운딩 박스 검출 기반의 접근 방식은 군중 카운팅에는 적합하지 않다. 해충 이미지의 경우도 그림에서 보듯이 이미지 내에서 해충의 개수가 상당히 많은 것을 볼 수 있다. 따라서 본 연구에서는 이러한 객체 분포 양상의 유사성을 토대로 기존 군중 카운팅 기술을 해충 카운팅 분야에 응용하고자 한다.

그림 2의 해충 이미지 샘플에서 보듯이, 해충 카운팅의 주요 이슈는 다음과 같다. 첫째, 해충의 모양이 상당히 다양하다. 예를 들어, 해충의 날개가 접힌 수도 펼쳐질 수도 있다. 또한 트랩 포획 당시 해충의 날개가 잘려서 해충의 모양이 왜곡될 수 있다. 이는 포즈의 다양화로 해충 카운팅의 성능이 저하될 수 있음을 의미한다. 둘째, 해충 발생 시기별로 해충의 개수가 상당히 가변적으로 변할 수 있다. 따라서 해충 개수에 따른 스케일 변화에 적응적인 모델 설계 방안이 필요하다. 셋째, 군중 이미지보다 폐색 문제가 더 심각하다. 군중 이미지에서는 사람에 의해서 가려지는 형태가 비슷하지만 해충 이미지에서는 해충의 다양한 포즈로 인해 폐색 모양도 더 다양하고 폐색의 난이도 또한 더 높다.



그림 2. 군중 이미지(좌측) 대 해충 이미지(우측)  
Fig. 2. Crowd images(left) vs. pest images(right)

본 연구에서는 상기에서 언급된 해충 카운팅의 주요 방해 요소를 극복하기 위한 해충 카운팅 전용 모델 개발하고자 한다. 첫째, 스케일 문제, 즉 해충 개수에 따른 스케일 문제를 해소하기 위한 저해상도 및 고해상도 모래시계 백본 기반의 멀티스케일 접근 방식을 제안하였다. 둘째, 해충의 폐색 문제와 포즈의 다양화를 대응하기 위한, 변형 가능한 멀티스케일 어텐션 모델링을 제시하였다. 특히, 저해상도와 고해상도 모래시계 백본 간의 특징 정보를 교환할 수 있는 멀티스케일 모래시계 간 스킵 연결을 고안하였다. 그리고 이 스킵 연결 위에 변형 가능한 합성곱 계층을 적용함으로써, 저해상도 모래시계의 변형된 특징 정보를 고해상도 모래시계의 백본으로 전달하여 멀티스케일 어텐션 기능을 구현하고 학습 능력을 강화하였다.

### 3.2 제안한 해충 카운팅 모델의 아키텍처

그림 3은 제안한 MaD2a-PCM 모델의 아키텍처를 보여주고 있다. 제안한 모델의 주요한 특징점은 저해상도 모래시계(LR hourglass) 백본, 고해상도 모래시계(HR hourglass) 백본, 입력 특징 퓨전 모듈(Input feature fusion module), 멀티스케일 모래시계 간 스킵 연결(Skip connection between multiscale hourglass), 변형 가능한 합성곱과 멀티스케일 어텐션 퓨전(Deformable convolution and multiscale attention fusion)을 포함한다는 것이다.

#### 3.2.1 입력 특징 퓨전 모듈

제안한 모델은 해충 카운팅을 위해, 512x512 크기의 해충 이미지( $I_{HR}$ )를 입력으로 받아서 저해상도 모래시계 백본을 통과한 후, 128x128 크기의 저해상도 밀도 맵( $\hat{Y}_{LR}^p$ )을 추정한다. 그리고 고해상도 모래시계 백본의 입력 특징 정보를 보다 강화하기 위해, 입력 특징 퓨전 모듈을 적용한다. 즉, 글로벌 잔차 스킵 연결(GRKC, Global Residual Skip Connection)을 새롭게 도입하여 입력단의 해충 특징 정보를 추출하고 저해상도 모래시계의 출력 특징을 업샘플링 블록(Upsampling block)을 통과하여 특징의 다양성을 확보한다.

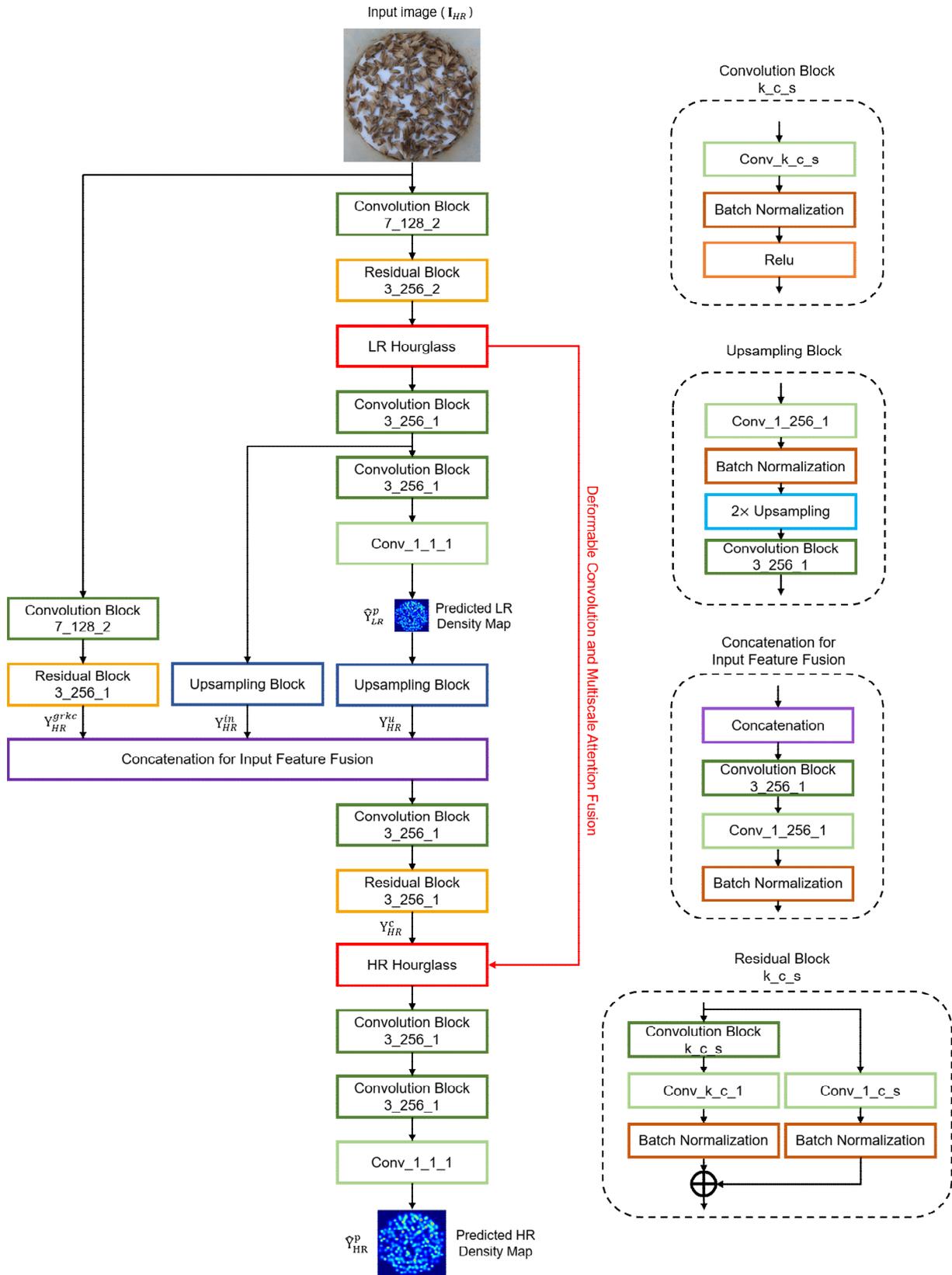


그림 3. 트랩 기반 해충 카운팅을 위한 제안된 변형 가능한 멀티스케일 밀도 어텐션 모델  
 Fig. 3. Proposed multiscale and deformable density attention model for trap-based pest counting

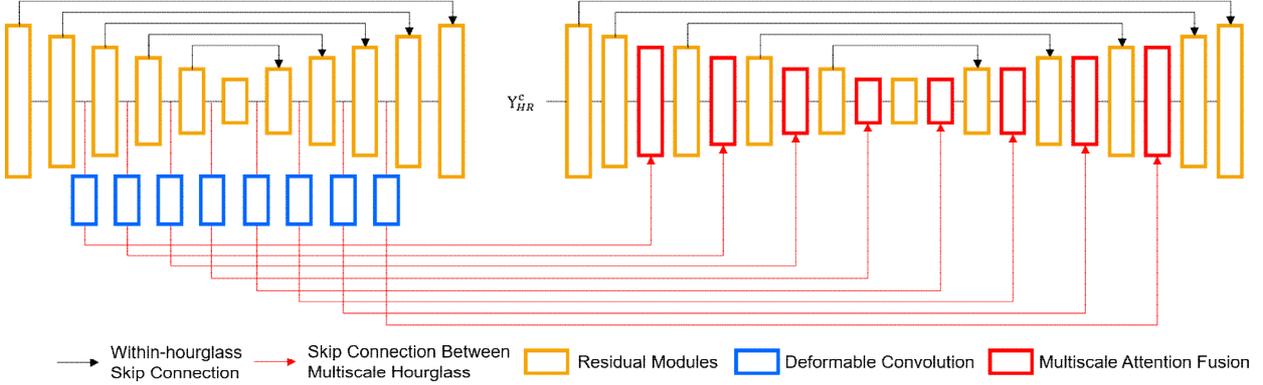


그림 4. 제안된 변형 가능한 멀티스케일 어텐션 퓨전 모듈  
Fig. 4. Proposed deformable and multiscale attention fusion module

그림에서 보듯이 입력 특징 퓨전 모듈은 3종류의 특징을 입력으로 받아서 퓨전 한다. 즉, 글로벌 잔차 스킵 연결의 출력 결과( $Y_{HR}^{grkc}$ ), 저해상도 모래시계의 특징 결과( $\hat{Y}_{HR}^{in}$ ), 그리고 저해상도 밀도 맵 결과( $Y_{HR}^u$ )를 연결 계층을 통해 퓨전 한다. 수식으로 표현하면 아래와 같다.

$$Y_{HR}^c = f_{RB} \circ f_{CB} \circ f_{CONCAT}(Y_{HR}^{grkc}, Y_{HR}^{in}, Y_{HR}^u) \quad (2)$$

여기서  $f_{RB}$ ,  $f_{CB}$ ,  $f_{CONCAT}$ 는 각각 잔차 블록, 합성곱 블록, 연결 계층을 의미한다. 입력 특징 퓨전 모듈의 결과인  $Y_{HR}^c$ 는 두 번째 고해상도 모래시계의 백본으로 입력된다. 따라서 고해상도 모래시계 백본은 저해상도 모래시계에서 예측된 저해상도 밀도 맵을 추가 정보로 활용하기 때문에 최종 출력 결과인 고해상도 밀도 맵을 더 정확하게 추정할 것으로 기대된다.

### 3.2.2 멀티스케일 모래시계 간 스킵 연결

입력 특징 퓨전 모듈을 통과한 특징은 두 번째 고해상도 모래시계 백본을 통과하여 최종 256x256 크기의 고해상도 밀도 맵  $\hat{Y}_{HR}^p$ 을 생성한다. 이때 저해상도 모래시계 백본과 고해상도 모래시계 백본은 멀티스케일 모래시계 간 스킵 연결을 통해 서로 연결된다. 이를 통해, 저해상도 모래시계 백본에서 추정된 특징 정보를 고해상도 모래시계 백본에 전달

함으로써, 고해상도 모래시계 백본의 특징 추출 능력을 강화할 수 있다. 즉, 제안한 멀티스케일 모래시계 간 스킵 연결은 스케일이 다른 두 백본을 효율적으로 공동 학습(Joint learning)할 수 있는 새로운 특징 추출 방식이다.

그림 4는 멀티스케일 모래시계 간 스킵 연결이 구체적으로 어떻게 설계되었는지 묘사해 주고 있다. 그림에서 보듯이, 저해상도 모래시계 백본에서 추출된 특징 정보는 멀티스케일 스킵 연결을 통해 고해상도 모래시계 백본으로 전송된다. 이때, 해충 포즈의 다양성과 폐쇄 문제를 해소하기 위한 방안으로 변형 가능한 합성곱 계층을 멀티스케일 스킵 연결에 적용하였다.

### 3.2.3 변형 가능한 합성곱

변형 가능한 합성곱은 기존의 표준 합성곱과는 달리, 기하학적인 변형을 모델링할 수 있다. 표준 합성곱 연산은 고정된 필터에서 합성곱 연산을 해서 특징을 추출한다. 따라서 규칙적으로 나열된 그리드 위치에서 국부 특징 추출은 가능하지만, 변형된 객체의 특징을 추출하기에는 한계가 있다. 따라서 본 연구에서는 해충 이미지에서 해충의 다양한 포즈의 기하학적 변형을 모델링하기 위해 변형 가능한 합성곱을 적용하고자 한다. 변형 가능한 합성곱의 수식은 아래와 같다.

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (3)$$

여기서  $R$ 은 규칙적으로 나열된 그리드 영역을 의미하고  $w$ 와  $x$ 는 각각 필터와 입력 이미지이다. 표준 합성곱과 달리, 옵셋  $\Delta p_n$ 을 추가해서 고정된 위치에서 특징을 추출하는 것이 아니라 전역적인 공간에서 객체 검출에 더 유용한 특징을 찾는다.

제안한 모델에서 변형 가능한 합성곱 연산은 멀티스케일 스킵 연결에 적용된다는 점이 특이한 사항이다. 즉, 저해상도 모래시계에서 추출된 저해상도 특징에 변형 가능한 합성곱 연산을 적용한다. 이는 고해상도 모래시계 백본이 특징을 추출할 때, 이미기하학적 모델링이 반영된 저해상도 특징을 추가 정보로 활용하기 위함이다.

### 3.2.4 멀티스케일 어텐션 퓨전

제안한 멀티스케일 어텐션 퓨전은 그림 4에서 보듯이, 고해상도 모래시계 백본에서 생성된 특징과 저해상도 모래시계 백본에서 전달된 특징을 서로 퓨전한다. 따라서 제안한 어텐션 퓨전은 멀티스케일 어텐션 기능을 모델링했다고 볼 수 있다. 제안한 멀티스케일 어텐션 퓨전의 피연산자인 쿼리( $Q$ ), 키( $K$ ), 벨류( $V$ )는 아래 수식을 통해 생성된다.

$$Q^{(l)} = V^{(l)} = f_{fc}(H_{HR}^{(l)}) \quad (4)$$

$$K^{(l)} = f_{fc}(f_{dc}(H_{LR}^{(l)})) \quad (5)$$

여기서  $f_{fc}$ ,  $f_{dc}$ 는 각각 선형 변환, 변형 가능한 합성곱을 나타내고  $H_{LR}^{(l)}$ ,  $H_{HR}^{(l)}$ 은 각각 저해상도, 고해상도 모래시계 내부의  $l$ 번째 잔차 블록에서 출력된 특징을 의미한다.

고해상도 모래시계는 인코더와 디코더로 구성되기 때문에 멀티스케일 어텐션 퓨전은 인코더와 디코더에 따라 조금 다르게 모델링된다. 첫째, 인코더 부분에서 고해상도 모래시계의 멀티스케일 어텐션 퓨전은 아래와 같이 계산된다.

$$Attention_E(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

여기서  $d_k$ 는 키의 차원 수를 나타내고  $softmax$ 는 0~1 사이의 가중치 값으로 변환하기 위한 소프트맥스 함수이다. 수식에서  $K$ 는 식 (5)에서 계산된 키에 해당한다. 식 (6)에서, 제안한 어텐션 모델에서는 저해상도 모래시계에서 추출된 특징인  $K$ 를 키로 활용해서  $softmax$  함수의 입력인 유사도 행렬을 최종 계산하는 것을 볼 수 있다.

둘째, 디코더 부분에서의 멀티스케일 어텐션 퓨전은 인코더와 거의 유사하게 모델링된다.

$$Attention_D(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V + S \quad (7)$$

여기서  $S$ 는 모래시계 내부의 스킵 연결을 통해 전달된 인코더 특징 정보를 나타낸다. 따라서 인코더와는 달리, 특징 정보  $S$ 가 추가적으로 활용된 것을 볼 수 있다.

제안한 멀티스케일 어텐션 퓨전은 최종 밀도 맵 추정에 있어서 중요한 특징을 선별하는 역할을 한다. 즉, 자기학습(Self attention)을 통해 특징 중요도를 결정한다. 본 연구에서는 저해상도 모래시계 백본에서 추출된 저해상도 특징을 쿼리로 활용하고 고해상도 모래시계 백본에서 추출된 고해상도 특징을 키와 값으로 활용한다는 점이 다른 어텐션 모델 [9][10]과는 큰 차이가 있다. 즉, 두 종류의 모래시계 백본을 쌓아서 서로 간의 특징 정보를 교환하여 더 강화된 특징을 공동 학습할 수 있는 장점이 있다. 그리고 고해상도 모래시계 백본의 인코더와 디코더에 따라서도 어텐션 수식이 다르게 적용된다.

### 3.2.5 손실 함수

추정한 밀도 맵과 정답 밀도 맵과의 오차를 측정하기 위해 평균 제곱 오차와 구조적 유사성(SSIM, Structural Similarity)[11]을 사용했다. 평균 제곱 오차 손실은 다음과 같이 계산된다.

$$L_{MSE} = \sum_{i=1}^N \|\hat{Y}_i - Y_i\|^2 \quad (8)$$

여기서  $\hat{Y}_i$ 는 모델이 예측한 밀도 맵이며,  $Y_i$ 는 수식 (1)을 통해 생성된 정답 밀도 맵이다. 수식에서 보듯이, 평균 제곱 오차는 두 밀도 맵 간의 픽셀 오차의 합을 계산함으로써, 픽셀 수준에서 오차를 측정함을 알 수 있다. 이에 반해, 구조적 유사성은 예지와 같은 이미지 구조의 유사도를 평가할 수 있다.

$$L_{SSIM} = \sum_{i=1}^N (1 - SSIM(\hat{Y}_i, Y_i)) \quad (9)$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (10)$$

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (11)$$

$$\sigma_x = \left( \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \quad (12)$$

SSIM은 식 (10)에서 보듯이, 휘도 평균 항목과 표준편차 평균 항목으로 구성된다. 휘도 평균은 두 패치 간의 밝기가 얼마나 유사한지를 측정하기 위한 항목이고, 표준편차 평균 항목은 두 패치 간의 밝기의 편차를 모델링할 수 있다. 이를 통해, 두 밀도 맵에 포함된 모든 패치 간의 휘도와 표준편차의 오차를 계산하여 이미지 구조의 유사도를 평가할 수 있다. 식 (10)에서  $C_1$ 과  $C_2$ 는 분모가 제로가 되는 상황을 막기 위해 사용된 아주 작은 상숫값이다. 그리고 SSIM의 값은 0~1 사이에 속하며 1에 가까울수록 두 이미지가 유사하다는 것을 의미한다. 수식 (11)와 (12)에서  $x$ 는 이미지에서 추출된 하나의 패치에 해당한다. 패치는 입력 이미지에서 래스터 스캐닝 방향으로 겹침 없이 추출할 수 있다.

최종 제안한 모델의 손실함수는 다음과 같다.

$$L_{ALL} = \lambda_{MSE} L_{MSE} + \lambda_{SSIM} L_{SSIM} \quad (13)$$

본 연구에서는 평균 제곱 오차와 구조적 유사성 손실을 1:1의 비율로 세팅했다. 즉,  $\lambda_{MSE} = \lambda_{SSIM} = 1$ 이다.

## IV. 실험 및 결과

### 4.1 실험 환경

학습 데이터는 디지털 트랩에서 촬영된 해충 이미지 4,462장을 사용하였으며 7:3의 비율로 랜덤하게 훈련 집합과 테스트 집합으로 나누었다. 학습 파라미터 갱신을 위해 아담 옵티마이저(Adam Optimizer)[12]를 사용했다. 그리고 배치 크기는 1, 에폭은 100회, 학습률은  $1.25e-4$ 로 설정했다. 제안한 모델의 성능을 평가하기 위해 기존 군중 카운팅 모델인 MCNN[5], ICCNet[6], SANet[7], KDMG[8]를 비교하였다.

### 4.2 밀도 맵 추정 결과

그림 5는 입력 해충 이미지와 제안한 MaD2a-PCM 모델로 추정된 밀도 맵 결과를 보여준다. 그리고 해충 이미지 하단에는 사람의 육안으로 측정된 해충 개수가 적혀 있다. 즉, 정답 해충 개수이다. 그리고 예측된 밀도 맵 하단에도 추정된 해충 개수가 적혀 있다. 여기서 추정된 해충 개수란 밀도 맵의 모든 픽셀을 더한 값을 말한다. 그림에서 보듯이, 측정된 해충 개수와 예측된 해충 개수가 거의 동일한 것을 볼 수 있다. 이는 기존의 군중 카운팅 기술 기반의 제안한 모델이 해충 카운팅 분야에도 잘 작동된다는 것을 보여준다. 특히 해충의 개수가 적거나 많은 경우에도 정확히 잘 추정하는 것을 볼 수 있다. 물론 해충의 개수가 많은 경우는 폐색 문제 등으로 인해 카운팅의 정확도가 상대적으로 떨어질 수 있다.

### 4.3 정량적 평가

본 연구에서는 군중 카운팅 분야에서 많이 사용되는 평균 절대 오차(MAE, Mean Absolute Error)를 평가 척도로 사용했다. 평균 절대 오차는 측정된 해충 개수와 예측된 해충 개수의 오차 값의 평균을 계산한다.

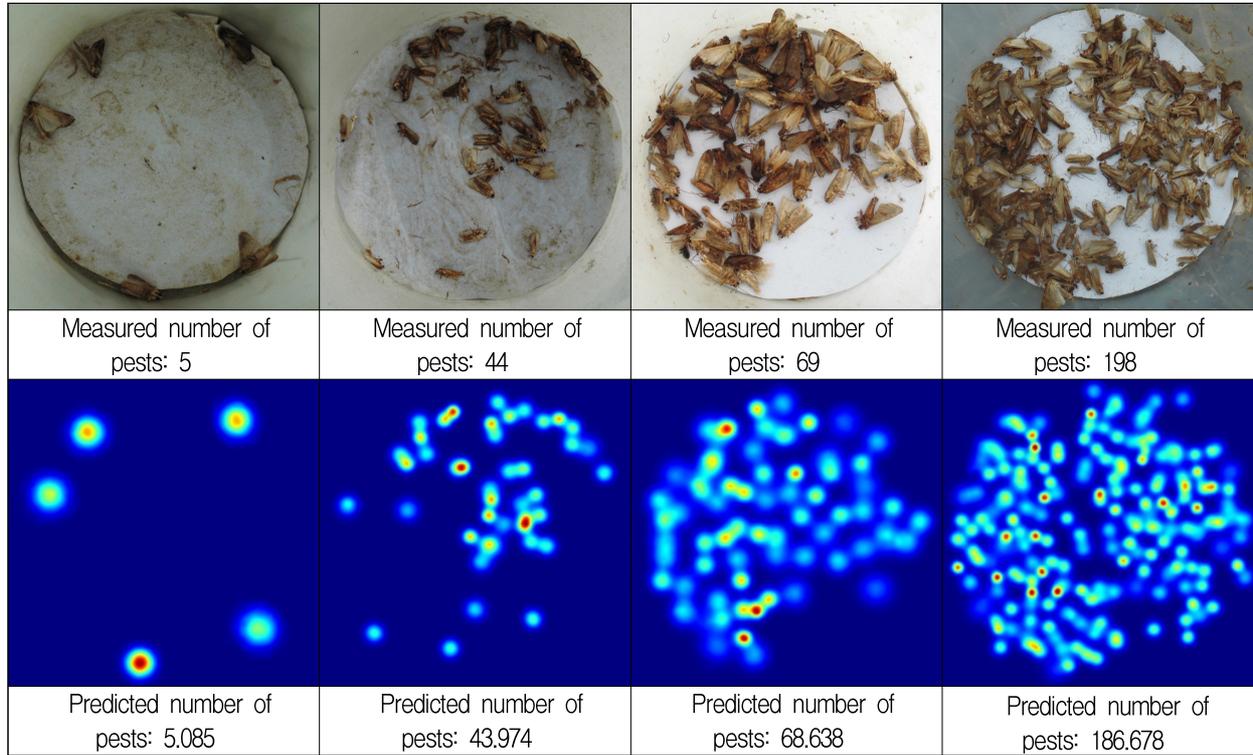


그림 5. 실험결과; 입력 해충 이미지(첫 번째 행), 예측된 밀도 맵(두 번째 행)  
 Fig. 5. Experimental results: input pest image(first row), predicted density map(second row)

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (14)$$

여기서  $n$ 은 테스트 이미지의 개수,  $\hat{y}_i$ 는 모델이 예측한 해충 개수,  $y_i$ 는 측정된 해충 개수이다. 따라서 평균 절대 오차는 값이 작을수록 모델의 성능이 우수함을 나타낸다.

표 1. 해충 카운팅을 위한 정량적 평가  
 Table 1. Quantitative evaluation for pest counting

Pest counting models	MAE
MCNN [5]	4.104
ICCNNet [6]	9.710
SANet [7]	2.114
KDMG [8]	1.273
<b>Proposed MaD2a-PCM</b>	<b>0.941</b>

표 1은 제안한 모델과 기존 모델에 대한 평균 절대 오차 값을 보여준다. 표에서 보듯이, 제안한 해충 카운팅 모델이 평균 제곱 오차가 가장 작은 것

을 확인할 수 있다. 이는 제안한 모델이 비교 모델들보다 정확도가 가장 우수하다는 것을 의미한다. 또한, 제안한 모델은 기존의 최고 성능 모델인 KDMG 보다 MAE 측면에서 약 26% 정도 개선된 결과를 보인다.

#### 4.4 애플리케이션 실험

제안한 멀티스케일 모래시계 백본, 변형 가능한 합성곱, 멀티스케일 어텐션 퓨전 모듈 추가 따른 성능 효과를 검증하기 위해 평균 제곱 오차를 측정하였다. 표 2는 제안한 모듈 추가에 따른 평균 제곱 오차 결과를 보여준다. 표에서 보듯이, 제안한 각각의 모듈이 추가됨에 따라 정확도가 향상되는 것을 볼 수 있다. 이는 제안한 멀티스케일 모래시계 백본, 변형 가능한 합성곱 기반의 멀티스케일 어텐션 퓨전 모듈이 해충 카운팅 성능을 향상할 수 있음을 입증한다. 즉, 멀티스케일 모래시계 백본은 해충 스케일에 적응적으로 대응할 수 있으며 변형 가능한 합성곱은 해충의 기하학적 변환을 모델링할 수 있다.

그리고 멀티스케일 어텐션 모듈은 밀도 맵 추정  
에 중요한 특징을 선별하고 특징 구별력을 강화하  
는 역할을 한다. 이를 통해, 최종 해충 카운팅의 정  
확도를 상당히 제고할 수 있었다.

표 2. 제안한 모듈 추가에 따른 MAE 평가

Table 2. MAE evaluation according to the addition of the  
proposed modules

Multiscale hourglass backbone	Multiscale attention fusion	MA
-	-	1.306
○	-	1.023
○	○	<b>0.941</b>

## V. 결론

본 연구에서는 트랩에서 포획된 해충 마릿수를  
예측하는 해충 카운팅 기술을 제안하였다. 특히, 기  
존의 군중 이미지와 해충 이미지와의 유사성에 기  
반해서 밀도 추정 기반의 군중 카운팅 기술을 해충  
카운팅 분야에 응용 및 확장하는 방안을 제시하였  
다. 제안한 모델은 기존의 군중 카운팅 모델과는 달  
리, 해충 스케일 문제에 대응할 수 있는 멀티스케일  
모래시계 백본을 설계하였으며 그리고 해충 포즈의  
다양성과 폐색 문제를 해결하기 위해 변형 가능한  
합성곱 모델을 모래시계 간 스킵 연결에 적용하는  
방안을 소개하였다. 또한 밀도 맵 추정에 중요한 특  
징을 선별하고 강화할 수 있는 멀티스케일 어텐션  
퓨전 모델을 개발하였다. 실험 결과를 통해, 제안한  
밀도 추정 기반의 해충 카운팅 기술이 트랩에서 포  
획된 해충 개수를 정확히 추정할 수 있음을 보였다.  
그리고 애플리케이션 연구를 통해서 제안한 모듈이  
해충 카운팅의 정확도를 제고하는데 효과적임을 확  
인하였다.

## References

[1] J. H. Lee, C. H. Son, and H. J. Yi, "Multiscale  
centernet for pest detection and counting", *Journal  
of KIIT, JeJu, Korea*, Vol. 20, No. 7, pp.  
111-121, Jul. 2022.

[2] A. Newell, K. Yang, and J. Deng, "Stacked  
hourglass networks for human pose estimation", in  
*Proc. European Conference on Computer Vision*,  
Amsterdam, Netherlands, Vol. 9912, pp. 483-499,  
Sep. 2016. [https://doi.org/10.1007/978-3-319-46484-8\\_29](https://doi.org/10.1007/978-3-319-46484-8_29).

[3] 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, Oct. 2017.

[4] V. Lempitsky and A. Zisserman, "Learning to  
count objects in images", in *Proc. Conference on  
Neural Information Processing Systems*, Vancouver,  
Canada, pp. 1324-1332, Dec. 2010.

[5] 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 Nevada, pp. 589-597, Jun.  
2016.

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

[7] 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. 757-773,  
Oct. 2018.

[8] 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>.

[9] 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>.

[10] W. Wang, E. Xie, X. Li, D. P. Fan, K. Song,

D. Liand, T. Lu, P. Luo, and L. Shao, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions", in Proc. IEEE/CVF International Conference on Computer Vision, pp. 548-558, 2021.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, Vol. 13, No. 4, pp. 600-612, Apr. 2004. <https://doi.org/10.1109/TIP.2003.819861>.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", In Proc. International Conference on Learning Representation, San Diego, USA, May 2015. <https://doi.org/10.48550/arXiv.1412.6980>.

## 저자소개

이 원 도 (Won-Do Lee)



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

손 창 환 (Chang-Hwan Son)



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

박 성 웅 (Sung-Woong Park)



2018년 1월 ~ 현재 :  
비엔에스코퍼레이션(주) 연구소장  
관심분야 : 식물보호학, 응용곤충,  
트랩 및 유인등 개발설계