

해충 및 과수화상병 분류를 위한 클래스 활성화 맵 기반의 비전 트랜스포머

손창환*, 신정희**

Vision Transformer based on Class Activation Map for Pest and Fire Blight Classification

Chang-Hwan Son*, Jung-Hee Shin**

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

요 약

최근 비전 트랜스포머 모델이 이미지 분류기의 대표적인 모델로 자리매김하고 있다. 하지만, 배경과 객체의 색상이 유사하거나 배경의 텍스처가 복잡할 경우 모델의 정확도가 저하되는 문제가 발생한다. 이를 해결하기 위해, 관심 영역 기반의 딥러닝 모델이 개발되었지만, 배경이 복잡한 경우 이미지 분할 라벨링 작업은 사실상 어렵다고 볼 수 있다. 또한 이미지 분할 맵은 이진화된 값으로 표현되기 때문에 정보량이 부족한 단점이 있다. 따라서 본 연구에서는 이미지 분할 맵 대신에 클래스 활성화 맵을 사용한 비전 트랜스포머를 제시하고자 한다. 실험 결과를 통해, 제안한 모델이 배경이 복잡한 이미지에 대해서도 관심 영역을 잘 포착할 수 있었고 해충 및 과수화상병 데이터셋에 대해서 기존 분류 모델보다 더 우수한 정확도를 달성할 수 있었다.

Abstract

Recently, the vision transformer model has become a representative model for image classifiers. However, if the background and object colors are similar or the background texture is complex, the model's accuracy deteriorates. To solve this problem, a deep learning model based on the region of interest has been developed, but when the background is complex, the image segmentation labeling task can be considered difficult in practice. Additionally, image segmentation maps have the disadvantage of insufficient information because they are expressed as binary values. Therefore, in this study, we would like to present a vision transformer that uses a class activation map instead of an image segmentation map. Through experimental results, the proposed model was able to capture regions of interest well even in images with complex backgrounds and achieved better accuracy than existing classification models for pest and fire blight datasets.

Keywords

image classification, vision transformer, class activation map, saliency

* 국립군산대학교 컴퓨터소프트웨어학부 교수(교신저자)

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

** 국립군산대학교 컴퓨터소프트웨어학부 학사과정

- ORCID: <https://orcid.org/0009-0003-6872-9404>

• Received: Jul. 04, 2024, Revised: Jul. 24, 2024, Accepted: Jul. 27, 2024

• Corresponding Author: Chang-Hwan Son

School of Computer Science and Engineering, Kunsan National
University, South Korea

Tel.: 82+63-469-8915, Email: cson@kunsan.ac.kr

1. 서 론

해충과 과수화상병은 농작물에 심각한 피해를 주는 주요한 요인이다. 따라서 해충의 종류와 발생량 및 과수화상병을 조기에 발견하여 방제하는 것이 중요하다. 이를 위해, 최근 인공지능 기술을 적용하여 촬영된 이미지로부터 해충 및 과수화상병 종류를 자동으로 진단하는 기법이 도입되고 있다. 종래의 육안 판별 기법보다 인력비용과 감별 시간을 감축할 수 있었지만, 정확한 진단을 하기에는 아직 넘어야 할 산이 많다. 특히, 객체의 크기가 아주 작거나 배경과 객체의 텍스처 및 색상이 유사할 때 정확도가 떨어지는 문제가 발생한다. 본 논문에서는 이런 문제점을 해결하기 위해, 클래스 활성화 맵(CAM, Class Activation Map)[1]이라 불리는 돌출 맵(Saliency map)을 활용한 비전 트랜스포머(ViT, Vision Transformer)를 제안하고자 한다.

기존의 해충 종류 및 과수화상병 분류는 크게 CNN(Convolutional Neural Network) 기반과 ViT 기반으로 나눌 수 있다. CNN은 컨볼루션 연산을 통하여 국부 특징을 추출하는 반면 ViT는 전체 영역에서 패치 토큰 간의 의존성을 모델링한다. 초기 CNN 모델은 기울기 소실 문제를 해결하고 깊은 레이어를 쌓을 수 있는 아키텍처를 설계하는 데 중점을 두었다[2]. 그리고 CNN 내부에 관심 영역을 주의 깊게 관찰할 수 있는 자가 어텐션 메커니즘을 개발하는 방향으로 전환되었다[3]. 한편, ViT 모델은 원래 자연어처리 분야에서 문맥 이해를 위해 개발된 트랜스포머에 기반하지만, 최근 이미지 분야에서도 CNN에 버금가는 성능을 발휘하고 있다. 이미지는 텍스트와는 달리, 출력 결과가 고해상도를 원하는 태스크가 있기에 원본 ViT 모델은 이미지 처리에 적합하지 않다. 이런 문제를 해결하기 위해, 원본 ViT를 피라미드 구조로 개조한 PVT(Pyramid Vision Transformer) 모델[4]이 개발되었다. 그리고 객체 스케일 대응과 특징 구별력 강화를 위해 멀티스케일[5] 및 크로스 어텐션[6] 모델들도 개발되고 있다.

상기에서 언급된 일반화된 분류 모델은 해충 및 과수화상병 분류 문제에도 적용 가능하고 우수한

인식률을 획득할 수 있다. 그러나 여전히 개선할 사항들이 남아 있다. 예를 들면, 특정 클래스 데이터 취득의 어려움으로 발생하는 클래스 불균형 문제[7], 학습 시 미리 정의된 부류가 아닌 테스트 이미지의 출현[8], 객체 크기가 아주 작거나 복잡한 배경의 영향으로 특징 추출의 어려움[9]이 현안으로 떠오르고 있다.

본 연구에서는 복잡한 배경의 영향을 극복할 수 있는 CAM 기반의 ViT 모델을 제시하고자 한다. CAM은 기존의 어텐션 모델에서 사용되는 이미지 분할 맵[10]과는 달리 라벨링 작업을 요구하지 않는다. 특히 과수화상병 데이터셋과 같은 경우는 복잡한 배경이 존재하므로 이미지 분할 작업은 사실상 불가능하다. 또한 이미지 분할 맵은 이진화된 값으로 표현되기 때문에 특징 추출을 위한 정보량이 부족하다. 반면 CAM은 히트맵 형식으로 표현되기 때문에 이미지 분할 맵 보다는 더 많은 정보량을 포함할 수 있다. 따라서 라벨링 작업을 요구하는 이미지 분할 맵보다 사전 학습된 모델에서 특정 클래스에 반응하는 돌출 맵을 바로 생성할 수 있는 CAM을 활용하는 것이 데이터 구축과 학습 측면에서 더 효과적이다. 이러한 CAM 정보를 효과적으로 활용하기 위해, 본 연구에서는 이중 분기로 구성된 ViT 아키텍처를 도입하고자 한다. 각 분기에서는 해충 이미지와 CAM 이미지를 입력으로 받아서 크로스 어텐션을 통해 정보를 서로 상호 교환하여 특징 추출 능력을 강화한다. 실험 결과를 통해, 제안한 ViT 모델이 배경이 복잡한 해충 및 과수화상병 이미지에서도 객체 인식 능력이 우수함을 확인하였다. 그리고 제안한 모델이 초기 CAM 이미지를 더 정확한 형태로 교정할 수 있음도 확인하였다.

II. 관련 연구

2.1 관심 영역 기반 CNN 모델

복잡한 배경에서 객체 분류를 하기 위한 핵심 관건은 관심 영역을 추출하는 것이다. 왜냐하면 배경 영역에서는 객체에 대한 정보가 없고 관심 영역에만 존재하기 때문이다.

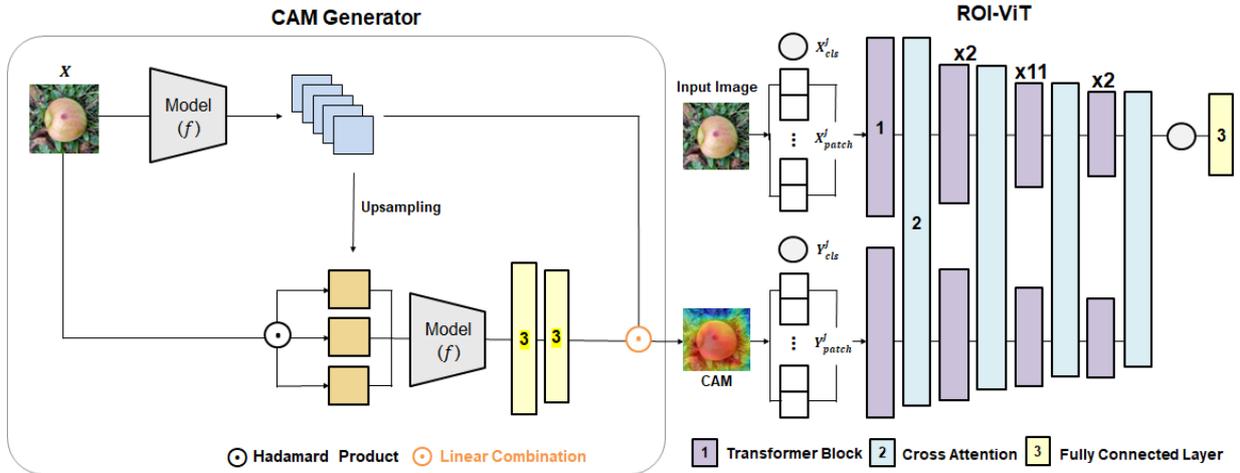


그림 1. 제안한 클래스 활성화 맵 기반의 비전 트랜스포머
 Fig. 1. Proposed vision transformer based on class activation map

즉, 배경은 잡음과 같이 객체 특징 추출을 방해하는 요인으로 작용한다. 이런 문제를 해결하기 위해, 이미지 분할 맵[10]이나 슈퍼 픽셀[11] 등을 관심 영역으로 활용한 CNN 모델이 개발되었다. 초기 모델은 관심 영역을 입력 이미지와 함께 쌓아서 기존 CNN 모델을 학습하기 때문에 이미지 분할 맵과 CNN 모델을 따로 학습해야 하는 번거로움이 있다. 이런 문제를 해결하기 위해, 이미지 분할 맵과 CNN 모델을 어텐션으로 연결하여 종단 간 학습을 완성한 LSA-Net[10]이 등장했다. LSA-Net은 사과나무 잎사귀 질병 분류에서 기존의 이미지 분류와 PVT 모델을 능가하는 결과를 보여주었다. 하지만, 서론에서 언급했듯이, 과수화상병과 같은 데이터셋은 복잡한 배경을 지닌 학습 데이터가 많기 때문에 이미지 분할 라벨링 작업을 하기가 만만치 않다. 또한 아키텍처 측면에서도 채택된 백본이 오래되었고 멀티스케일 구현이 누락되어 스케일에 취약하다. 그리고 어텐션도 단순 연결 기반으로 초기 및 후기 퓨전만을 적용했기 때문에 특징의 상호 보완 능력도 떨어질 수밖에 없다. 따라서 최신 ViT 백본과 특징 상호 보완 능력을 강화할 수 있는 새로운 아키텍처가 필요하다.

2.2 비전 트랜스포머 모델

최근 이미지 분류 모델은 CNN 모델에서 ViT 모델로 전환하는 추세이다. ViT는 컨볼루션 연산이

없고 오직 자가 어텐션 메커니즘을 기반으로 토큰 간의 의존성을 모델링한 아키텍처이다. ViT 모델은 입력 이미지를 패치로 분리한 후, 자가 어텐션으로 구현된 트랜스포머 레이어를 거쳐 특징을 추출하고 최종 헤더를 추가해서 분류한다. 원래 ViT 모델은 입력 패치와 출력 패치가 동일하기 때문에 고해상도 출력에는 적합하지 않다. 이를 해결하기 위해, 어텐션 연산시 풀링이나 컨볼루션을 통해 특징의 공간 해상도를 줄이고 채널 개수를 증가하는 멀티스케일 ViT 모델[4][5]이 개발되었다. 그리고 이중 분기 구조로 해상도가 다른 패치들을 퓨전함으로써, 특징의 상호 보완 능력을 강화한 크로스 어텐션 ViT 모델[6]도 개발되었다. 이외에도 로컬 특징을 그룹핑하여 ViT[12]를 설계하거나 시맨틱 토큰을 활용한 이중 분기 기반의 ViT 모델[13]도 있다. 이런 ViT 모델은 최근 해충 분류나 과수화상병 분류에도 적용되고 있다. 하지만 ViT 모델과 거의 흡사하여 전이학습에 가깝다고 할 수 있다. 최근에는 멀티모달 퓨전[14]이나 클래스 불균형[7] 문제를 해결하기 위한 ViT 모델도 개발되고 있다.

III. 제안한 CAM 기반 ViT 모델

3.1 제안한 모델 아키텍처

그림 1은 제안한 CAM 기반 ViT 모델의 아키텍처이다. 제안한 모델은 크게 CAM 생성기와 이중

분기 구조 기반의 ROI-ViT 모듈로 구성된다. CAM 생성기는 돌출 맵을 생성하는 부분이고 ROI-ViT는 CAM 이미지와 입력 이미지의 특징 정보를 크로스 어텐션을 통해 특징을 강화하고 최종 분류기를 학습한다. 그리고 ROI-ViT는 입력 이미지의 패치 토큰을 입력으로 받는 이미지 분기와 CAM 패치 토큰을 입력으로 받는 관심 영역 분기로 구성된다. 그리고 CLS(Class Classification Token) 토큰은 모든 패치 토큰과 상호 작용하여 전역적인 정보를 담고 있는 특징 벡터이다. 따라서 완전연결계층(Fully connected layer)으로 구현된 분류기는 CLS 토큰만을 입력으로 받는다. 그림 1에서 트랜스포머 블록은 기존의 ViT 모델에서 널리 사용되는 모듈이다. 자세한 내용은 관련 문헌에 나와 있다[4][5]. 본 논문에서는 제안한 모델의 핵심인 크로스 어텐션과 CAM 생성기에 대해 자세히 설명하고자 한다.

3.2 CAM 생성기

본 연구에서는 전역 풀링이 필요 없이 일반화된 모델에서도 돌출 맵을 추출할 수 있는 Score-CAM[1]을 사용하였다. Score-CAM은 크게 두 단계로 구성된다. 첫 번째 단계는 왼쪽 상단처럼 입력 이미지를 모델(f)에 넣어 특징 맵을 추출하고 두 번째 단계에서는 왼쪽 하단과 같이 추출된 각각의 특징 맵에 대한 기여도를 계산한다.

$$CAM^c = ReLU\left(\sum_k \alpha_k^c A_l^k\right) \quad (1)$$

여기서 A_l^k 는 추출된 특징 맵으로 모델 f 의 l 번째 레이어에서 추출된 k 번째 맵을 말한다. α_k^c 는 추출된 특징 맵에 대한 가중치로써 기여도를 나타낸다. 그리고 $ReLU$ 는 정류선형유닛(Retified Linear Unit)이다. c 는 입력 이미지의 부류 값이고 CAM^c 는 생성된 최종 CAM 이미지를 의미한다.

$$\alpha_k^c = f^c(s(up(A_l^k)) \circ X) - f^c(X) \quad (2)$$

기여도를 나타내는 식 (2)에서 X 는 입력 이미지

이고 f^c 는 모델 f 의 c 번째 노드에서의 예측 값이다. \circ 는 원소별 곱이고 up 와 s 는 각각 업샘플링과 스케일링 과정을 의미한다. 즉, up 는 A_l^k 를 입력 이미지 X 와 동일한 크기로 맞추는 역할을 하고 s 는 업샘플링 된 A_l^k 를 0~1로 변환한다. 스케일링(s)는 해당 특징 맵에서 최소와 최대 구간으로 선형적으로 매핑해서 정규화한다. 따라서 식 (2)에서 α_k^c 는 입력 이미지 X 를 A_l^k 와 곱하여 해당 모델이 예측한 값으로 볼 수 있다. 식 (2)의 α_k^c 는 소프트맥스 함수를 통해 최종 0~1로 정규화된다.

3.3 크로스 어텐션

입력 이미지와 CAM 이미지의 특징 정보를 서로 상호 교환하여 입력 이미지의 특징 구별력을 제고하기 위해 크로스 어텐션 연산을 수행한다. 크로스 어텐션은 전역 정보를 함축한 CLS 토큰을 사용해서 수행한다.

$$CLS = X_{ds}^j + Y_{ds}^j \quad (3)$$

여기서 X_{ds}^j 와 Y_{ds}^j 는 각각 크로스 어텐션에 입력되는 이미지 분기와 관심 영역 분기의 CLS 토큰에 해당한다. j 는 그림 1의 크로스 어텐션 레이어의 인덱스 값이다. 크로스 어텐션이 총 4번 수행하므로 j 는 1에서 4까지의 정수로 표현된다. 단, 입력 레이어의 CLS 토큰은 j 가 0으로 할당된다. 수식 (3)에서 제안한 크로스 어텐션은 각 분기의 CLS 토큰을 더함으로써 특징을 상호 교환하고 보완한다. 이렇게 갱신된 CLS 토큰은 식 (4)와 (5)와 같이, 나머지 패치 토큰과 결합하여 다음 레이어로 전송된다.

$$X^{j+1} = [CLS \parallel X_{patch}^j] \quad (4)$$

$$Y^{j+1} = [CLS \parallel Y_{patch}^j] \quad (5)$$

여기서 \parallel 는 연결(Concatenation) 연산을 의미한다. X_{patch}^j 와 Y_{patch}^j 는 각 분기의 CLS 토큰에 대응되는 패치 토큰이다.

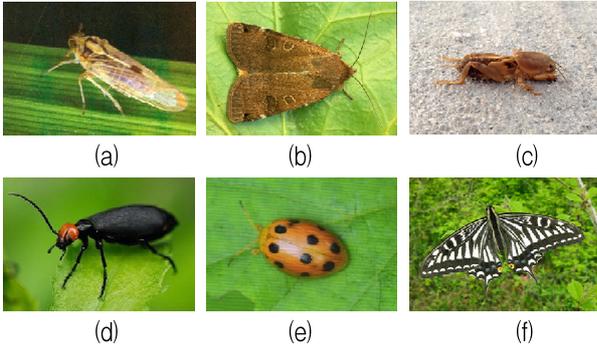


그림 2. 해충 이미지 데이터 예시

(a) 벼멸구, (b) 나방유충, (c) 거세미 나방, (d) 물집 딱정벌레, (e) 진드기종, (f) 매미충과

Fig. 2. Example of pest images (a) Brown plant hopper, (b) Grub, (c) Large cutworm, (d) Lytta polita, (e) Brevipalpus lewisi mcgregor, (f) Cicadellidae

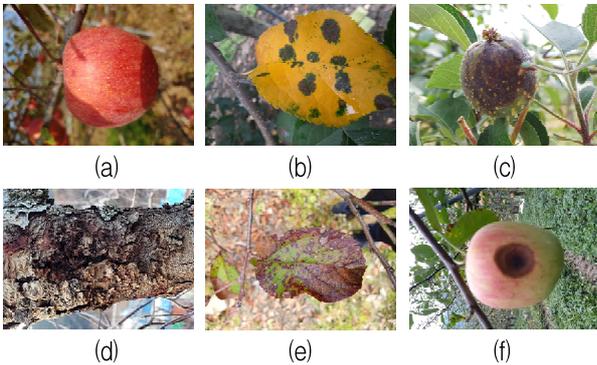


그림 3. 과수화상병 이미지 데이터 예시

(a) 정상, (b) 사과갈색무늬병, (c) 사과과수화상병, (d) 사과부란병, (e) 사과점무늬낙엽병, (f) 사과탄저병

Fig. 3. Example of fire blight images

(a) Healthy, (b) Marssonia blotch, (c) Citrus canker, (d) Valsa canker, (e) Alternaria leaf spot, (f) Rot

IV. 실험 및 결과

4.1 실험 환경

제안한 모델의 성능 평가를 위해 두 종류의 데이터셋을 활용하였다. 하나는 해충 이미지 데이터셋이고 다른 하나는 과수화상병 데이터셋이다. 해충 이미지 데이터셋은 IP102 공개 데이터셋[15]이고 총 102개의 클래스를 가진다. 본 연구에서는 해충이 아닌 애벌레 등과 같은 이미지를 삭제하고 총 86개 부류로 구성된 35,296장의 이미지를 학습 데이터로 구성했다. 과수화상병 데이터셋은 AI 허브 사이트 [16]에 공개된 이미지 데이터셋이고 총 6종류의

41,715장의 사과 이미지로 구성되어 있다. 그림 2와 3은 실험에서 사용된 해충 이미지와 과수화상병 이미지 예시이다. 학습을 위해 훈련과 테스트 데이터셋의 비율은 7:3으로 분할하였다. 배치 크기는 10, 에폭은 80, 학습률은 0.0001로 설정하였다. 최적화 기법은 Adam[15]이고 Pytorch 딥러닝 프레임워크를 사용했다. 그리고 모델 f 를 위해 MViT[5] 모델을 적용하였다.

표 1. 해충 데이터셋에 대한 정량적 평가

Table 1. Quantitative evaluation for pest dataset

Models	Correct recognition rate
ResNet50 [2]	73.64%
CrossViT [6]	76.81%
PVT [4]	76.95%
MViT [5]	80.06%
Proposed model	80.24%

표 2. 과수화상병 데이터셋에 대한 정량적 평가

Table 2. Quantitative evaluation for fire blight dataset

Models	Correct recognition rate
ResNet50 [2]	98.66%
CrossViT [6]	99.46%
PVT [4]	91.90%
MViT [5]	99.52%
Proposed model	99.56%

4.2 정량적 평가

모델 성능 평가를 위해, CNN 계열인 ResNet50[2] 과 최신 ViT 모델인 CrossViT[6], PVT[4], MViT[5] 를 비교했다. 평가척도로는 정인식률을 사용했다. 정인식률은 전체 테스트 이미지 개수에서 올바르게 분류한 이미지 개수의 비율을 말한다. 표 1과 2는 각각 해충 데이터셋과 과수화상병 데이터셋에 대한 비교 대상 모델의 정인식률 결과이다. 표에서 보듯이, 제안한 CAM 기반 ViT 모델이 두 종류의 데이터셋에 대해 정확도가 가장 우수함을 알 수 있다. 특히 최신 MViT[5] 모델보다 정확도가 0.18%와 0.04% 증가하였다. 이는 제안한 CAM 기반 ViT 모델이 특징의 구별력을 강화하여 최종 이미지 분류의 정확도를 제고할 수 있음을 말해준다.

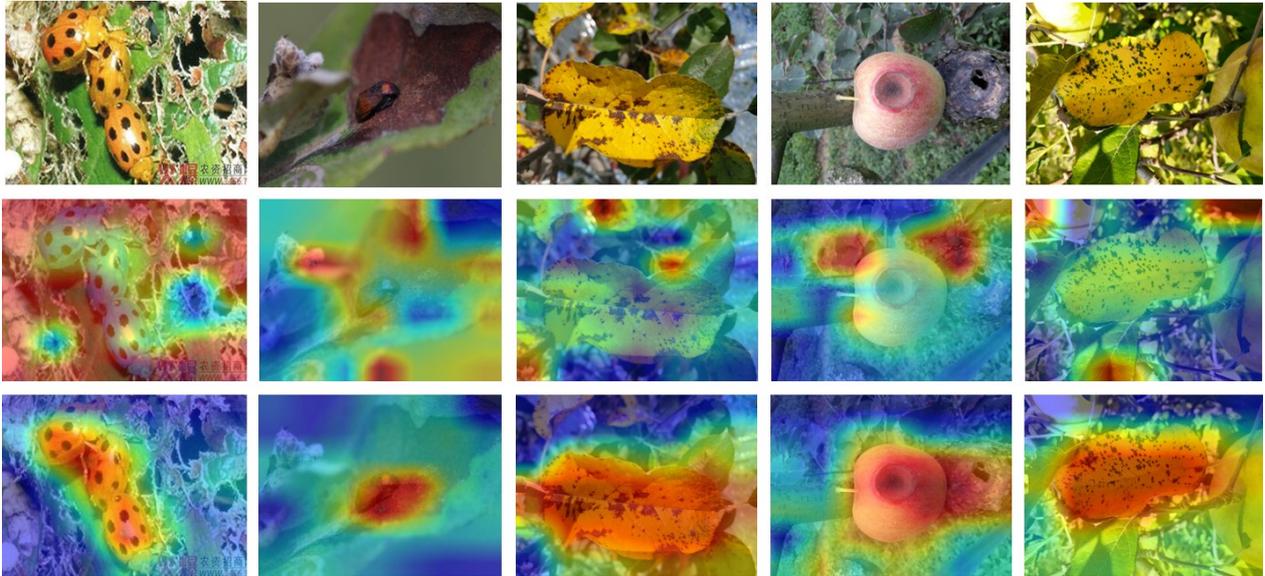


그림 4. 입력 이미지 (첫 번째 행), 초기 CAM 이미지 (두 번째 행), 제안한 모델로 시각화된 CAM 이미지(마지막 행)
 Fig. 4. Input images (first row), initial CAM (second row), visualized CAM with proposed model (third row)

즉, CAM 이미지와 입력 이미지를 크로스 어텐션을 통해서 CLS 토큰을 퓨전함으로써, 최종 분류기에 입력되는 CLS 특징 구별력을 강화할 수 있음을 말해준다.

4.3 CAM 시각화

제안한 모델이 복잡한 배경을 지닌 테스트 이미지에 대해 돌출 맵을 잘 찾는지를 검증하기 위해 Score-CAM을 적용하였다. CAM 이미지에서 붉은색 계통은 해충이나 질병이 존재하는 관심 영역이며 푸른색 계통은 배경 영역에 해당한다. 그림 4는 초기 CAM 이미지가 제안한 모델을 통과한 후, CAM 이미지가 어떻게 갱신되는지를 보여준다. 그림에서 보듯이, 초기 CAM 이미지는 복잡한 배경이 있을 경우, 해충이나 질병이 존재하는 영역을 검출하지 못하는 것을 볼 수 있다. 즉, 초기 CAM 이미지는 정확도가 떨어지는 것을 볼 수 있다. 그러나 제안한 모델을 통과한 후, CAM 이미지가 해충이나 질병이 존재하는 과수 영역을 정확히 포착하는 것을 볼 수 있다. 이는 제안한 모델이 복잡한 배경에서도 객체의 특징 구별력을 제고할 수 있음을 입증해 준다.

V. 결 론

본 논문에서는 해충 및 과수화상병 분류를 위한 CAM 기반의 ViT 모델을 제안하였다. 실험 결과를 통해, 두 종류의 해충 및 과수화상병 데이터셋에 대해 제안한 모델이 가장 우수한 정확도를 달성할 수 있었다. 특히 최신 MViT[5] 모델보다 정확도가 0.18%와 0.04% 증가한 것을 확인하였다. 또한, 복잡한 배경을 가진 테스트 이미지에 대해서도 제안한 모델이 이미지 분류를 위한 관심 영역을 잘 포착할 수 있음을 확인하였다. 그리고 CAM 이미지 시각화를 통해, 제안한 모델이 부정확한 초기 CAM 이미지의 성능을 향상할 수 있음을 확인하였다. 이를 통해, 제안한 모델은 크로스 어텐션을 통해 CLS 클래스 토큰의 특징 구별력을 강화하여 최종 분류기 성능을 제고할 수 있다고 결론지을 수 있다.

References

- [1] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-CAM: Score-weighted visual explanations for convolutional neural networks", In Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, pp. 111-119, Jun. 2020. <https://doi.org/10.1109/cvprw.50498.2020.00020>.

- [2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, pp. 770-778, Jun. 2016. <https://doi.org/10.1109/cvpr.2016.90>.
- [3] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks", In Proc. IEEE Conference on Computer Vision, Salt Lake, USA, pp. 7132-7141, Jun. 2018. <https://doi.org/10.1109/cvpr.2018.00745>.
- [4] 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, Montreal, QC, Canada, pp. 548-558, Oct. 2021. <https://doi.org/10.1109/iccv48922.2021.00061>.
- [5] H. Fan, B. Xiong, K. Mangalam, Y. Li, Z. Yan, J. Malik, and C. Feichtenhofer, "Multiscale vision transformers", In Proc. IEEE International Conference on Computer Vision, Montreal, QC, Canada, pp. 6804-6815, Oct. 2021. <https://doi.org/10.1109/iccv48922.2021.00675>.
- [6] C. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-attention multi-scale vision transformer for image classification", In Proc. IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, pp. 347-356, Oct. 2021. <https://doi.org/10.1109/iccv48922.2021.00041>.
- [7] C. Wang, J. Zhang, J. He, W. Luo, X. Yuan, and L. Gu, "A two-stream network with complementary feature fusion for pest image classification", Engineering Application of Artificial Intelligence, Vol. 124, Sep. 2023. <https://doi.org/10.1016/j.engappai.2023.106563>.
- [8] A. Zareian, K. D. Rosa, D. H. Hu, and S.-F. Chang, "Open-vocabulary object detection using captions", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Nashville, USA, pp. 14393-14402, Jun. 2021. <https://doi.org/10.1109/cvpr46437.2021.01416>.
- [9] G.-E. Kim and C.-H. Son, "Multiscale cross attention vision transformer for pest image classification", The Journal of Korean Institute of Information Technology, Vol. 21, No. 7, pp. 77-84, Jul. 2023. <https://doi.org/10.14801/jkiit.2023.21.7.77>.
- [10] H.-J. Yu and C.-H. Son, "Leaf spot attention network for apple leaf disease identification", in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, pp. 229-237, Jun. 2020. <https://doi.org/10.1109/cvprw50498.2020.00034>.
- [11] E. C. Tetila, et al., "Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks", IEEE Geoscience and Remote Sensing Letters, Vol. 17, No. 5, pp. 903-907, May 2020. <https://doi.org/10.1109/LGRS.2019.2932385>.
- [12] C.-F. Chen, R. Panda, and Q. Fan, "Regionvit: Regional-to-local attention for vision transformer", in Proc. International Conference on Learning Representation, Virtual, Sep. 2022.
- [13] T. Yao, Y. Li, Y. Pan, Y. Wang, X.-P. Zhang, and T. Mei, "Dual vision transformer", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 45, No. 9, pp. 10870-10882, Sep. 2023. <https://doi.org/10.1109/tpami.2023.3268446>.
- [14] Y. Zhang, L. Chen, and Y. Yuan, "Multimodal fine-grained transformer model for pest recognition", Electronics, Vol. 12, No. 12, Jun. 2023. <https://doi.org/10.3390/electronics12122620>.
- [15] X. Wu, C. Zhan, Y.-K. Lai, M.-M. Cheng, and J. Yang, "IP102: A large-scale benchmark dataset for insect pest recognition", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, pp. 8787-8796, Jun. 2019. <https://doi.org/10.1109/cvpr.2019.00899>.
- [16] <https://www.aihub.or.kr/> [accessed: Jul. 22, 2024]

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 컴퓨터소프트웨어학부 부교수
관심분야 : 컴퓨터 비전, 영상처리, 기계학습, 딥 러닝

신 정 희 (Jung-Hee Shin)



2022년 3월 ~ 현재 :
국립군산대학교
컴퓨터소프트웨어학부 학사과정
관심분야 : 컴퓨터 비전, 영상처리,
기계학습, 딥 러닝