

해충 이미지 분류를 위한 멀티스케일 교차 주의집중 비전 트랜스포머

김가은*, 손창환**

Multiscale Cross Attention Vision Transformer for Pest Image Classification

Ga-Eun Kim*, Chang-Hwan Son**

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

요 약

해충은 작물을 가해함으로써 작물 생산성 감소와 품질 저하를 수반한다. 따라서 해충 종류를 신속·정확하게 파악하여 적합한 방제 작업을 적기에 실시하는 것이 중요하다. 기존의 해충 감별은 전문가의 육안 판단에 의해서 수행되기에 많은 시간과 인력 비용이 요구된다. 따라서 이 논문에서 해충 예찰 무인 자동화를 실현하기 위해, 해충 이미지로부터 해충 종류를 자동으로 판별하기 위한 딥러닝 모델을 제시하고자 한다. 특히, 본 연구에서는 관심 영역 기반의 교차 주의집중 모듈을 내장한 멀티스케일 비전 트랜스포머 모델을 제안하고자 한다. 제안한 모델은 이중 분기로 설계되었고 해충 분기와 관심영역 분기 간의 클래스 토큰과 패치 토큰을 상호교환하여 주의집중 기능을 강화하였다. 실험 결과를 통해서, 제안한 교차 주의집중 모듈이 특징 추출 능력을 향상하여 최종 분류 모델의 정확도를 약 1.3% 개선할 수 있음을 확인하였다.

Abstract

Pests damage crops, resulting in reductions in crop productivity and quality. Therefore, it is important to promptly and accurately identify pest types and perform appropriate control work in a timely manner. Existing pest identification requires a lot of time because it is performed by visual judgment of experts. Therefore, in this paper, in order to realize the unmanned automation of pest prediction, we propose a deep learning model for automatically discriminating pest types from pest images. In particular, in this study, we propose a multiscale vision transformer model with a built-in cross-attention module based on the region-of-interest(ROI). The proposed model is designed as a dual branch, and the attentional function is strengthened by exchanging class tokens and patch tokens between the pest branch and the ROI branch. Through the experimental results, it was confirmed that the proposed cross-attention module can improve the feature extraction ability and improve the accuracy of the final classification model by about 1.3%.

Keywords

pest image classification, region-of-interests, cross attention, vision transformer

* 군산대학교 소프트웨어학부 소프트웨어학전공 학사과정 · Received: May 24, 2023, Revised: Jun. 11, 2023, Accepted: Jun. 14, 2023
- ORCID: <https://orcid.org/0009-0007-1059-7364> · Corresponding Author: Chang-Hwan Son
** 군산대학교 소프트웨어학부 소프트웨어학전공 교수 · Software Science and Engineering Major, School of Software, Kunsan National University, Korea
(교신저자) · Tel.: +82-63-469-8915, Email: cson@kunsan.ac.kr
- ORCID: <https://orcid.org/0000-0001-7077-3074>

1. 서 론

스마트팜은 전통적인 농사법에 정보통신 기술을 접목하여 작물의 생육환경을 적절하게 유지·관리할 수 있는 농업을 말한다. 인공지능, 빅데이터, 사물인터넷, 센서, 드론 등을 활용하여 관련 데이터를 모니터링하고 분석하여 작물의 생산량과 품질을 최적화하고 노동력과 비료와 같은 자원을 최소화할 수 있다[1].

해충은 작물의 잎, 줄기, 꽃 등을 먹거나 오염시켜 작물의 생산성 감소와 품질 저하를 수반한다. 따라서 해충 종류에 따른 적절한 방제를 적기에 실시하여 작물 피해를 최소화할 필요가 있다. 기존의 해충 감별 방식은 숙련된 농부나 전문가의 육안 판단으로 수행되기 때문에 많은 시간과 인력 비용이 요구된다. 또한 최근에는 농산물 교역 증가에 따른 외래 및 돌발해충이 발생하기 때문에 유입된 해충에 대한 감별이 어려울 수 있다. 따라서 이 논문에서는 카메라로 촬영된 해충 이미지로부터 자동으로 해충 종류를 파악할 수 있는 이미지 분류(Image recognition) 기술을 개발하고자 한다.

고전적인 이미지 분류 기술은 SIFT[2], HOG[3]와 같은 수작업 기반의 특징 추출 모델을 설계하는 것에 중점을 두었다. 그러나 특징의 다양성과 구별력의 한계로 인해, 최근에는 순수 데이터에서 고수준의 특징을 학습 과정을 통해 자동으로 추출할 수 있는 딥러닝 모델이 도입되고 있다. 초창기 딥러닝 모델은 VGG[4], ResNet[5]과 같은 CNN 모델에 기반을 두었지만 최근에는 자가주의집중(Self-attention) 모델인 비전 트랜스포머(Vision transformer)가 대세를 이루고 있다. 대표적인 비전 트랜스포머에는 DeiT[6], CrossViT[7], PVT[8], MViT[9] 등이 있다.

기존의 자가주의집중 기반 비전 트랜스포머는 다중스케일 문제와 주의집중 연산량 문제를 주로 개선해 왔다. 하지만 해충 분류의 경우, 해충이 존재하는 관심 영역과 배경 영역이 확연히 구분된다. 따라서 관심 영역에 대한 중요도를 비전 트랜스포머의 주의집중 모델에 반영할 수 있다면, 비전 트랜스포머의 특징 구별력을 높일 수 있으며 자가주의집중 모델의 학습 능력을 강화할 수 있다. 이를 위해,

본 논문에서는 해충 감별을 위한 관심 영역 기반 교차 주의집중 비전 트랜스포머(RCA-ViT, ROI-based Cross Attention ViT)를 제시하고자 한다. 제안한 RCA-ViT는 이중 분기(Dual branch) 네트워크로 구성되었으며 관심영역 분기(ROI branch)에서 추출된 특징 정보를 해충 분기(Pest branch)에 활용함으로써 해충 분기의 자가주의집중 능력을 강화하고자 한다. 특히, 교차 주의집중 모델을 설계하여 관심영역 분기와 해충 분기 간의 정보를 상호 교환함으로써 학습 능력을 제고하고자 한다. 그리고 실험 결과를 통해, 제안한 RCA-ViT가 자가주의집중 모델의 성능을 개선할 수 있음을 검증하고자 한다.

II. 비전 트랜스포머 모델

비전 트랜스포머[10]는 입력 영상을 패치로 분할하고 패치 단위로 연산을 수행한다. 비전 트랜스포머는 크게 패치 임베딩(Patch embedding), 인코더(Encoder), 분류기(Classifier)로 구성된다. 패치 임베딩 과정은 패치를 특징 벡터 공간으로 변환하고 인코더는 주의집중 연산을 수행하여 구별력 높은 특징을 생성하며 분류기는 최종 추출된 특징을 활용해서 분류 과정을 수행한다. 패치 임베딩 과정에서 전역 특징 벡터(Global feature vector)를 학습하기 위해 클래스 토큰(Class token)을 새로 추가하고 패치의 위치 정보를 활용하기 위해 위치 임베딩(Position embedding)을 적용한다. 그리고 분류기에서는 최종 인코더를 통해 생성된 클래스 토큰을 다층신경망(Multi-Layer neural network)에 통과시켜 분류기를 학습한다.

그림 1의 맨 좌측은 비전 트랜스포머 인코더의 구조를 보여주고 있다. 그림에서 보듯이, 인코더는 레이어 정규화(Layer normalization), 주의집중 모듈(Attention module), 순방향 신경망(Feedforward neural network)로 구성된다. 여기서 주의집중 모듈이 최근 중점적으로 연구되고 있다. 대표적인 주의집중 모듈로는 MHA(Multi-Head Attention), SRA(Spatial Reduction Attention), MHPA(Multi-Head Pooling Attention) 등이 있다.

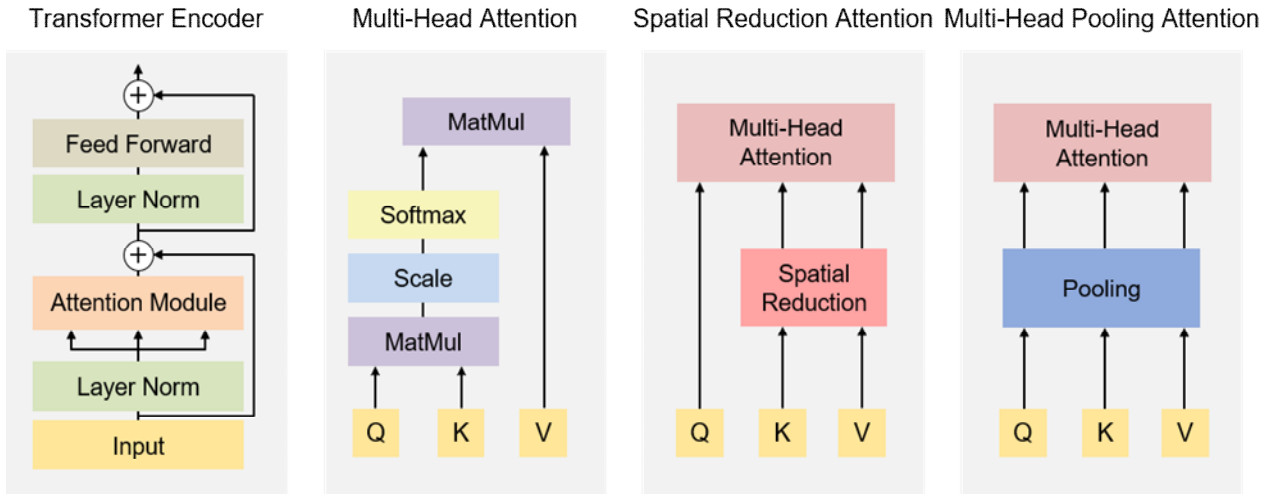


그림 1. 비전 트랜스포머 모델; 트랜스포머 인코더, 멀티 헤드 주의집중, 공간 축소 주의집중, 멀티 헤드 풀링 주의 집중(왼쪽에서 오른쪽)

Fig. 1. Vision transformer model; Transformer encoder, Multi-Head attention, Spatial reduction attention, Multi-Head pooling attention(left to right)

2.1 DeiT 모델

DeiT[6]는 비전 트랜스포머 기반 이미지 분류 모델로서, 데이터 증강과 지식 증류(Knowledge distillation) 통해 상대적으로 적은 수의 학습 데이터를 사용하여 학습 효율을 개선하였다. 지식 증류는 미리 학습된 교사 모델(Teacher model)의 지식, 즉 라벨 정보를 학생 모델(Student model)이 학습하는 방식을 말한다. 특히, 증류 토큰(Distillation token)을 클래스 토큰과 함께 사용해 학생 모델이 교사 모델의 라벨 정보를 예측하도록 유도하였다.

2.2 CrossViT 모델

CrossViT[7]는 기존 비전 트랜스포머 모델[10]에 멀티스케일과 교차 주의집중을 추가한 모델이다. 멀티스케일 구현을 위해, 두 종류의 스케일을 학습할 수 있는 이중 분기 구조를 채택하였다. 그리고 스케일 간 정보 교환을 위해, 교차 주의집중 모듈을 설계하였다. 교차 주의집중 모듈은 전역 특징 정보를 담고 있는 클래스 토큰을 상호 교환함으로써 다른 스케일 정보를 반영하여 최종 멀티스케일 기반 이미지 인식 성능을 제고하였다.

2.3 PVT 모델

PVT[8]는 기존의 단일 스케일 기반 비전 트랜스포머 구조[10]를 피라미드 형태로 변경한 모델로서, 고해상도를 요구하는 다양한 컴퓨터 비전 분야에 적용할 수 있도록 개선하였다. 특히 FPN(Feature Pyramid Network)[11]처럼, 피라미드 구조에서 출력된 멀티스케일 특징들을 결합해서 다목적용 백본을 완성하였다. 또한 피라미드 구조에서 발생할 수밖에 없는 주의집중 연산량 문제를 해결하기 위해, 그림 1의 SRA 주의집중 모듈을 새롭게 제시하였다.

2.4 MVIT 모델

MViT[9]도 PVT[8] 모델처럼 멀티스케일 비전 트랜스포머의 일종이다. 단지 차이점은 MViT는 주의집중 모듈에서 풀링(Pooling) 연산을 통해 멀티스케일을 구현하는 반면 PVT는 패치 임베딩 과정을 통해 수행한다. 또한 MViT는 주의집중 모듈을 구현하기 위해 그림 1의 MHPA 모델을 적용한다. 이를 통해 연산량 감소뿐만 아니라 멀티스케일 구조를 실현함으로써 이미지 분류 성능을 크게 향상시켰다.

III. 관심 영역 기반 멀티스케일 교차 주의집중 비전 트랜스포머

3.1 제안한 접근 방법

제안한 기법의 접근 방법은 기존의 CNN 백본에 관심 영역 정보를 결합해서 잎사귀 질병 분류 분야에서 큰 성과를 거둔 LSA-Net[1][12]에 기반한다. 잎사귀 질병 분류에서 질병이 위치하는 곳은 배경 영역이 아닌 관심 영역, 즉 잎사귀 내부이다. 따라서 잎사귀 내부 영역에 대한 정보를 CNN에 제공함으로써 CNN의 반점(Spot)을 검출하는 능력을 향상할 수 있다. 이와 유사하게 해충 감별도 해충의 특성을 대표할 수 있는 텍스처가 배경 영역이 아닌 관심 영역, 즉 해충 객체 내부에 존재한다. 따라서 배경 영역과 해충이 존재하는 관심 영역에 대한 정보를 비전 트랜스포머에 제공함으로써 분류 정확도를 제고할 수 있다. 기존의 LSA-Net과 제안한 RCA-ViT와의 차이는 관심 영역의 정보를 비전 트랜스포머에서 구현했다는 것이다. 특히, ROI 분기와 해충 분기 간의 관심 영역의 정보를 주의집중 모듈에서 상호 교환함으로써 학습 성능을 강화하였다.

그림 2는 제안한 RCA-ViT의 구조를 보여준다.

그림에서 보듯이, 제안한 RCA-ViT는 이중 분기 네트워크로써 관심 영역 특징 추출기를 포함한 해충 분기와 관심영역 분기로 구성된다. 관심영역 분기는 해충 이미지에서 관심 영역 특징을 멀티스케일 방식으로 추출하고 해충 분기는 멀티스케일 기반으로 관심영역 분기의 특징과 상호 보완하여 최종 해충 분류를 수행한다.

3.2 관심 영역 특징 추출

관심영역 분기는 그림 2에서 보듯이, 입력 해충 이미지가 아닌 관심 영역 특징을 입력으로 받는다. 관심 영역 특징을 초기화하기 위한 간단한 방법은 이미지 분할 맵을 활용하는 것이다. 따라서 본 연구에서는 기존의 멀티스케일 기반 이미지 분할 모델 [13]을 도입하였다. 그림 2의 좌측은 이미지 분할 모델의 구조로써, 크게 ResNet50[5] 백본과 멀티스케일 기반 해상도 변환 모듈로 구성된다. 해상도 변환 모듈은 백본에서 추출된 고수준의 특징에서 고해상도의 이미지 분할 맵을 생성한다. 특히 해상도 증가를 위해 멀티스케일 기반 업샘플링(Upsampling) 레이어를 적용했다. 즉, 다양한 풀링 크기를 적용하여 멀티스케일 해상도 변환을 구현하였다.

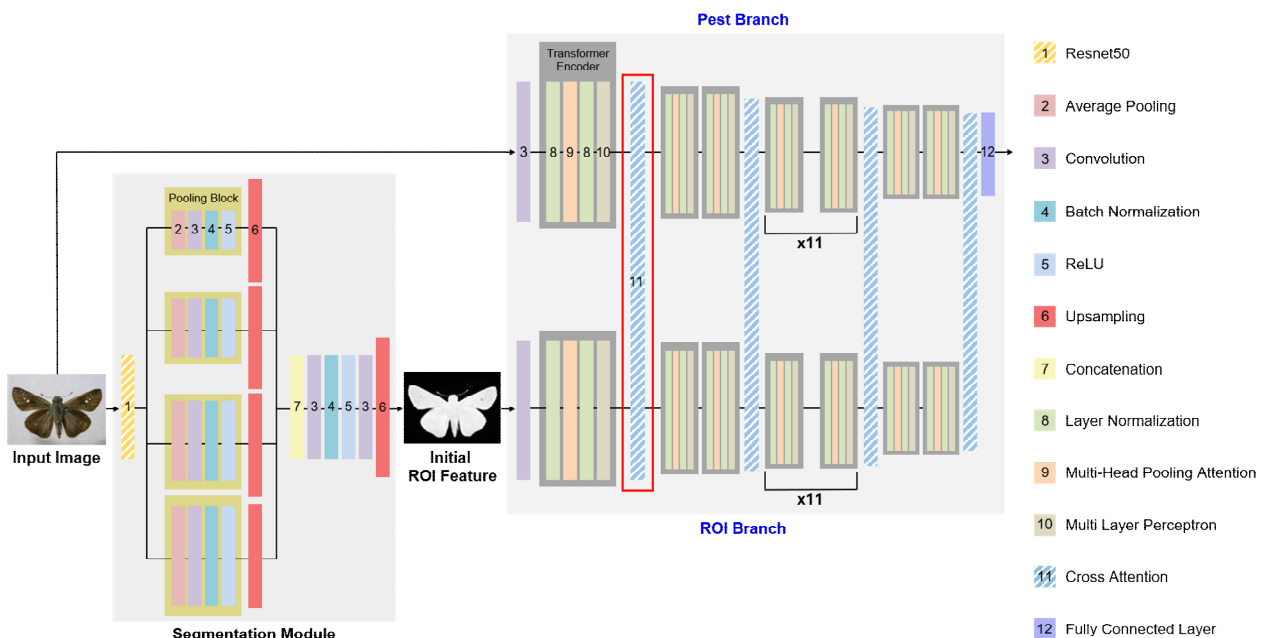


그림 2. 제안한 전체 네트워크 구조
Fig. 2. Architecture of the proposed entire network

이미지 분할 모델을 학습하기 위해 IP102[14] 데이터셋에서 15,855장의 해충 영상을 추출하여 훈련 데이터로 사용하였다. 모델 학습을 위해, 배치 크기는 10, 에폭은 300, 학습률은 0.0001로 설정하였다. 손실 함수는 교차 엔트로피를 사용하였고 최적화 기법으로는 Adam[15] 옵티마이저를 적용하였다. 참고로 이미지 분할 모델에 사용된 데이터는 해충 분류 학습 데이터에서 제외되었다. 그림 3은 이미지 분할 모델의 결과 영상을 보여준다. 그림에서 보듯이, 정답 라벨링 영상과 추정된 라벨링 영상이 거의 유사한 것을 볼 수 있다. 따라서 이 추정된 이미지 분할 맵을 제안한 RCA-ViT의 관심영역 분기의 초기치로 활용할 수 있다.

3.3 관심 영역 기반 교차 주의집중 모듈

제안한 RCA-ViT의 이중 분기는 각각 해충 이미지와 초기화된 관심 영역 특징을 입력으로 받는다.

제안한 RCA-ViT 모델에서 이중 분기는 멀티스케일 특징 추출을 위해 기존의 MViT[9]로 구현되었다. 해충 분기와 관심영역 분기는 해충 이미지와 분할 맵을 각각의 트랜스포머 인코더를 통과시켜 특징에 대한 중요도를 스스로 학습한다. 그리고 교차

주의집중(Cross attention) 모듈에서 각 분기에서 생성된 특징 정보를 상호 교환하여 자가주의집중 학습 능력을 강화한다.

그림 4는 교차 주의집중 모듈로써 입출력 과정을 보여준다.

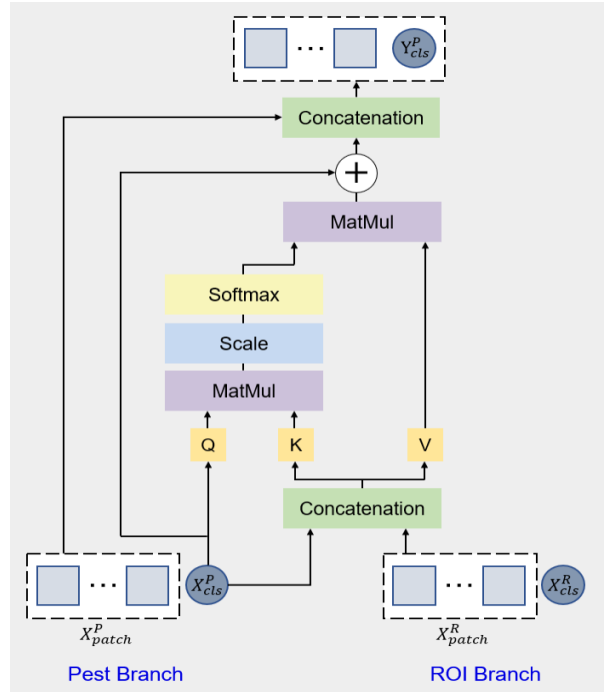


그림 4 교차 주의집중 모듈
Fig. 4. Cross attention module

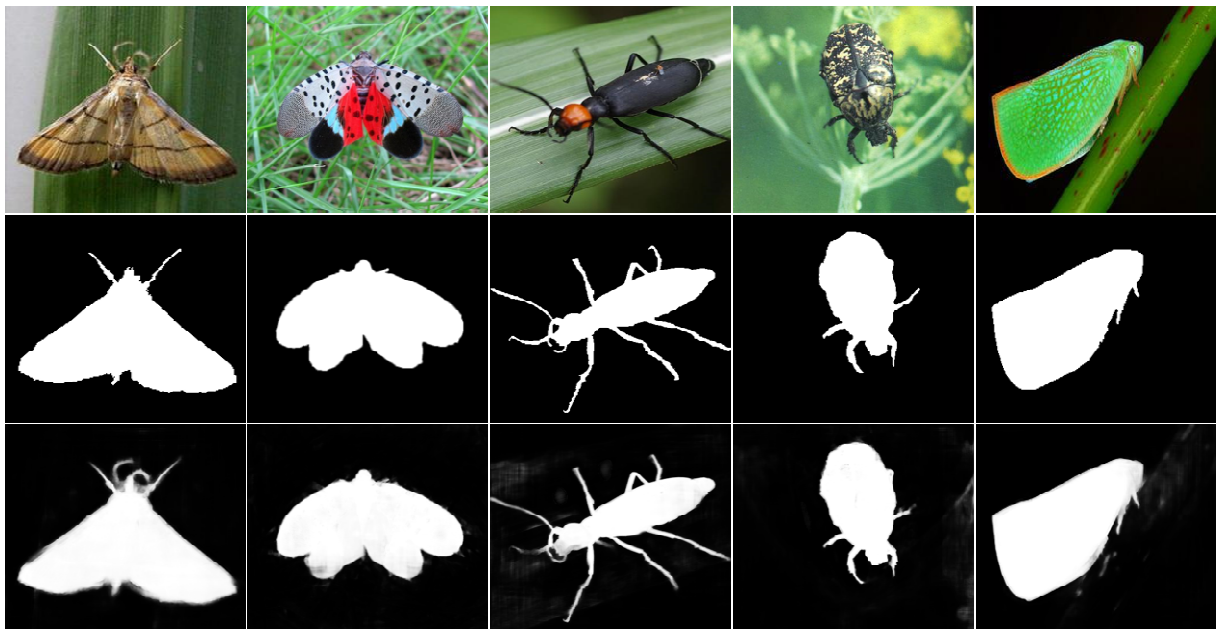


그림 3. 입력 해충 이미지(첫 번째 행), 라벨링 이미지(두 번째 행), 분할 맵(세 번째 행)
Fig. 3. Input pest images(first row), labeled images(second row), segmented maps(third row)

먼저, 해충 분기의 클래스 토큰과 관심영역 분기의 패치 토큰과 결합한다. 컴퓨터 비전 분야에서 토큰이란 이미지 또는 특징 맵의 패치에 해당한다. 그리고 패치 토큰은 입력 패치에 대응하는 국부 특징(Local features)에 대응하며 반면 클래스 토큰은 전역 특징(Global features)을 추출하기 위해 추가적으로 사용된 토큰이다. 이중 분기의 서로 다른 클래스 토큰과 패치 토큰의 결합은 다음과 같이 처리된다.

$$X'^P = f_c(X_{cls}^P, X_{patch}^R) \quad (1)$$

여기서 X_{cls}^P , X_{patch}^R 는 각각 해충 분기의 클래스 토큰과 관심영역 분기의 패치 토큰에 해당한다. f_c 는 두 종류의 토큰을 결합하기 위한 연결 레이어(Concatenation layer) 함수이다. 그리고 교차 주의집중 연산은 아래와 같이 연산 된다.

$$Q = X_{cls}^P W_q, \quad K = X'^P W_k, \quad V = X'^P W_v \quad (2)$$

$$Y_{cls}^P = \text{Softmax}\left(\frac{QK^T}{\sqrt{C/h}}\right)V \quad (3)$$

여기서 W_q , W_k , W_v 는 학습 파라미터로써 선형 변환 행렬을 의미한다. C 는 특징 맵의 채널 수이고 h 는 헤드(Head) 개수이다. 식 (2)는 선형 변환 행렬을 사용해서 쿼리(Query), 키(Key), 값(Value)을 만드는 과정에 해당한다. 여기서 주목할 점은 해충 분기의 클래스 토큰인 X_{cls}^P 을 사용해서 쿼리를 도출하고 키와 값은 관심영역 분기의 패치 토큰을 포함한 X'^P 을 활용하여 생성한다는 것이다. 그리고 식 (3)의 Q 와 K 의 내적 연산을 통해 유사도 행렬을 만들고 소프트맥스 함수를 통과하여 가중치 값으로 최종 변환한다. 유사도 행렬과 값과의 행렬-벡터 연산을 통해, 해충 분기의 클래스 토큰과 관심영역 분기의 패치 토큰을 융합할 수 있다. 즉, 관심영역 분기 패치 토큰의 정보를 해충 분기의 클래스 토큰에 반영할 수 있다.

$$Z^P = f_c(Y_{cls}^P, X_{patch}^P) \quad (4)$$

교차 주의집중 연산을 통해 갱신된 해충 분기 토큰인 Y_{cls}^P 는 식 (4)와 같이, 최종 해충 분기의 패치 토큰인 X_{patch}^P 와 다시 결합하여 해충 분기의 입력 특징을 갱신한다. 즉, 클래스 토큰만을 갱신한다. 동일한 방식으로 그림 4에서 관심영역 분기의 클래스 토큰을 갱신하기 위해서는 그림 4의 좌측과 우측의 분기 위치를 바꾸면 된다.

제안한 RCA-ViT는 상기에서 설명한 것처럼, 해충 분기와 관심영역 분기의 클래스 토큰을 갱신하기 위해 다른 분기의 패치 토큰을 활용한다. 이는 교차 주의집중 모듈의 학습 능력을 강화할 수 있고 해충 분기의 특징 구별력을 더 향상할 수 있다.

참고로 해충 분기의 클래스 토큰만을 해충 분기의 입력으로 사용한다. 즉, 관심영역 분기의 패치 토큰은 해충 분기의 학습 능력을 강화하기 위해 사용되었다. 그리고 모델 용량을 경량화하기 위해 두 분기의 파라미터는 공유하였다.

IV. 실험 및 결과

4.1 실험 환경

제안한 RCA-ViT 모델을 학습하기 위해, IP102[14] 공개 데이터셋을 사용했다. IP102는 총 102개의 부류로 구성되고 총 65,534장의 해충 이미지를 포함하고 있다. 그러나 본 연구에서는 텍스트와 같이 부적합한 이미지를 필터링하여 총 86개 부류로 구성된 총 35,296장의 해충 이미지를 사용했다. 훈련과 테스트 데이터셋의 비율은 7:3으로 분할하였고 배치 크기는 10, 에폭은 80, 학습률은 0.0001로 설정하였다. 최적화 기법은 Adam[15] 옵티마이저이고 Pytorch 딥러닝 프레임워크를 사용했다.

이중 분기는 4단계로 구성하였고 단계마다 1개, 2개, 11개, 2개의 트랜스포머 인코더를 사용했다. 또한 각 단계의 마지막에 1개의 교차 주의집중 모듈을 적용하여 총 4개의 교차 주의집중 모듈을 사용했다.

4.2 정량적 평가

제안한 RCA-ViT 성능을 평가하기 위해, 최신 비전 트랜스포머 모델인 DeiT[6], CrossViT[7], PVT[8], MViT[9]와 비교하였다. 그리고 평가척도로 정인식률(Correct recognition rate)을 사용했다. 정인식률은 전체 테스트 영상 개수에서 올바르게 분류한 영상 개수의 비율을 말한다. 표 1은 평가 대상 모델의 정인식률 수치를 보여주고 있다. 표 1에서 보듯이, 제안한 RCA-ViT 모델이 가장 성능이 좋은 것을 볼 수 있다. 제안한 RCA-ViT 모델은 기존 MViT 백본을 이중 분기로 구현해서 교차 주의집중 모델을 강화한 것으로 볼 수 있다. 즉, 제안한 RCA-ViT 모델은 백본으로 MViT를 사용했고 이중 분기의 가중치를 서로 공유했기 때문에 MViT와의 가장 큰 차이점은 교차 주의집중 모듈 적용에 있다. 따라서 이 실험 결과를 통해, 관심영역 분기의 패치 토큰을 해충 분기의 클래스 토큰을 갱신하는데 활용함으로써 교차 주의집중 모델이 최종 인식률을 개선하는데 효과적임을 확인할 수 있다. 또한 오류율에 대한 분석 결과 해충 이미지에서 해충의 크기가 작은 경우 오류율이 높은 것을 확인할 수 있었다.

표 1. 정량적 평가

Table 1. Quantitative evaluation

Models	Correct recognition rate
DeiT [6]	77.59%
CrossViT [7]	76.10%
PVT [8]	76.24%
MViT [9]	80.34%
Proposed RCA-ViT model	81.61%

V. 결 론

본 논문에서는 해충 분류를 위한 관심 영역 기반 멀티스케일 교차 주의집중 비전 트랜스포머 모델을 제안하였다. 특히 해충 분기와 관심영역 분기로 구성된 제안한 RCA-ViT 모델에서는 이중 분기 간의 클래스 토큰과 패치 토큰을 서로 교환하여 유사도를 반영함으로써 각 분기의 전역 특징을 대표하는

클래스 토큰을 갱신하였다. 실험 결과를 통해, 동일한 백본을 사용한 MViT에 비해 정인식률이 약 1.3%가 증가할 수 있었다. 이는 제안한 관심영역 분기와 해충 분기의 교차 주의집중 모듈이 주의집중 학습을 강화하고 정확도를 개선하는데 효과적임을 말해준다. 또한 제안한 RCA-ViT 모델은 IP102 공개 해충 데이터셋을 사용한 실험에서 기존의 최첨단 비전 트랜스포머 모델보다도 더 우수한 성능을 달성할 수 있었다.

References

- [1] H.-J. Yu, C.-H. Son, and D. H. Lee, "Apple Leaf Disease Identification through Region-of-Interest-Aware Deep Convolutional Neural Network", *Journal of Imaging Science and Technology*, Vol. 64, No. 2, pp. 20507-1-20507-10, Jan. 2020. <https://doi.org/10.2352/j.imagingsci.technol.2020.64.2.020507>.
- [2] D. G. Lowe, "Distinct Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, Nov. 2004. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [3] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, USA, pp. 886-893, Jun. 2005. <https://doi.org/10.1109/CVPR.2005.177>.
- [4] K. Simonyan and A. Zisserman, "Very Deep Convolution Networks for Large-Scale Image Recognition", In *Proc. International Conference on Learning Representation*, Sep. 2014. <https://doi.org/10.48550/arXiv.1409.1556>.
- [5] 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>.

- [6] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, "Training Data-Efficient Image Transformers & Distillation through Attention", Proc. of the 38th International Conference on Machine Learning, pp. 10347-10357, Jul. 2021.
- [7] 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>.
- [8] 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>.
- [9] Y. Li, C. Wu, H. Fan, K. Mangalam, B. Xiong, J. Malik, and C. Feichtenhofer, "MViTv2: Improved Multiscale Vision Transformers for Classification and Detection", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, pp. 4794-4804, Jun. 2022. <https://doi.org/10.1109/CVPR52688.2022.00476>.
- [10] 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>.
- [11] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 936-944, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.106>.
- [12] H. -J. Yu and C. -H. Son, "Leaf Spot Attention Network for Apple Leaf Disease Identification", 2020 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>.
- [13] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 6230-6239, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.660>.
- [14] 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/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, pp. 8787-8796, Jun. 2019. <https://doi.org/10.1109/CVPR.2019.00899>.
- [15] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", In Proc. International Conference on Learning Representation, San Diego, USA, Dec. 2014. <https://doi.org/10.48550/arXiv.1412.6980>.

저자소개

김 가 은 (Ga-Eun Kim)



2020년 3월 ~ 현재 : 군산대학교
소프트웨어학부 학사과정
관심분야 : 컴퓨터 비전, 영상처리,
딥 러닝

손 창 환 (Chang-Hwan Son)



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