

(2차)새로운 접근 방식 모색 (YOLO+SAM+ConvNeXt)

새로운 접근 방식 모색 - 인간의 시각적 인지 방식 참조

간의 시각적 인지 능력은 다양한 상황에서 높은 효율성을 보이며, 복잡한 환경에서도 빠르고 정확하게 객체를 인식하는 능력을 가집니다. 이를 AI 모델에 적용하기 위해서는 인간의 시각적 인지 메커니즘을 모델링하고, 이를 기반으로 CNN 모델의 한계를 극복할 수 있는 전략을 제시할 수 있습니다.

1. 인간의 시각적 처리 방식 모델링

인간은 시각 정보를 처리할 때, 단순히 이미지를 픽셀 단위로 분석하는 것이 아니라, 다양한 **추상화 수준**에서 정보를 처리합니다. 예를 들어, 인간은 **컨텍스트**와 **전역적 맥락**을 고려하여 이미지를 해석하며, 이는 CNN이 특정 로컬 특성에 지나치게 의존하는 것과 차별화됩니다. 따라서, **전역적 특성**을 반영하고, **다양한 스케일의 정보**를 종합적으로 분석하는 모델을 개발하는 것이 중요한 접근이 될 수 있습니다. 이를 위해서는 CNN 구조에 **다양한 규모의 컨볼루션 층**을 추가하거나, **멀티스케일 학습**을 적용하는 방법이 있을 수 있습니다.

2. 시각적 주의(Visual Attention) 메커니즘

인간은 중요한 정보를 선택적으로 처리하고, 불필요한 세부사항은 무시하는 능력을 가지고 있습니다. 이러한 **주의 메커니즘**(attention mechanism)은 이미 일부 자연어 처리(NLP) 모델에서 큰 성과를 거두었으며, 시각적 정보 처리에서도 중요한 역할을 합니다. 최근에는 CNN에 **주목(attention) 모듈**을 통합하여, 중요한 부분에 더 집중하도록 유도하는 방식이 적용되고 있습니다. 예를 들어, **Spatial Attention**이나 **Channel Attention** 기법을 통해 모델이 중요한 특징을 보다 잘 추출하도록 할 수 있습니다.

3. 인간의 시각적 학습 방식 적용

인간은 이미지를 학습할 때, 주어진 객체를 여러 관점에서 보고 학습하며, 다양한 ****변형** (rotate, scale, shift 등)**에도 잘 대응할 수 있습니다. 이를 기반으로, **데이터 증강** 기법을 넘어서는 새로운 방식으로 학습 데이터를 변화시키는 방법을 고려할 수 있습니다. 예를 들어, **강화 학습**을 통해 모델이 시각적 환경에서 스스로 중요한 변형을 학습하고, 이를 기반으로 모델의 일반화 성능을 높이는 방향으로 연구할 수 있습니다. 또한, **자기 지도 학습(self-supervised learning)** 기법을 활용하여, 라벨이 없는 데이터를 통해 모델이 중요한 특징을 독립적으로 학습하도록 할 수 있습니다.

4. 다양한 감각의 융합

인간은 시각 외에도 청각, 촉각 등 다양한 감각을 통합하여 세상을 인식합니다. 이는 단일 정보에 의존하지 않고 **다양한 인지적 관점**을 고려한다는 점에서 중요한 시사점을 제공합니다. 최근에는 **멀티모달 학습**이 활발히 연구되고 있으며, 이미지와 텍스트, 혹은 이미지와 소리 등을 결합하여 더욱 풍부한 특징을 추출하는 접근법이 사용되고 있습니다. CNN 모델을 다양한 종류의 센서 데이터와 결합하는 방법을 고려하여, 보다 풍부한 정보를 바탕으로 시각적 인지 성능을 개선할 수 있습니다.

5. 인간의 인지적 계층화 모델 적용

인간은 사물을 인식할 때, **상위 개념**을 빠르게 파악하고, 이를 바탕으로 하위 세부 정보를 해석하는 능력을 가지고 있습니다. 이를 CNN에 적용하려면, **계층적 학습** 방식을 고려할 수 있습니다. 예를 들어, 낮은 수준의 특성 추출 후, 이를 종합하여 상위 개념을 인식하는 구조를 모델에 구현할 수 있습니다. **시각적 계층적 네트워크(hierarchical vision networks)**는 이러한 관점을 반영하여, 세부적인 정보에서 점진적으로 복잡한 개념으로의 변환을 모델링하는 방법이 될 수 있습니다.

새로운 접근 방식 모색 - YOLO, SAM, ConvNeXt 결합 방안

이미지 인식 모델 배경 지식

이미지 인식 기술에서 YOLO, SAM, ConvNeXt 모델은 각각 고유한 기능과 특징을 가진 중요한 딥러닝 모델입니다.

1. YOLO (You Only Look Once) 모델은 실시간 객체 검출에 특화된 모델로, 이미지 내 객체의 위치를 바운딩 박스 형태로 찾아내고 해당 객체의 클래스를 예측합니다 3. 이 모델의 주요 장점은 빠른 처리 속도와 전체 이미지를 한 번에 분석할 수 있는 능력입니다. 그러나 작은 객체 검출이나 밀집된 객체 인식에는 한계가 있습니다 17.
2. SAM (Segment Anything Model)은 프롬프트(점, 박스 등)를 기반으로 이미지 내 객체를 마스크 형태로 정확하게 분할하는 것을 목표로 합니다 3. 이 모델은 다양한 객체를 세분화할 수 있는 뛰어난 능력을 가지고 있지만, 입력 프롬프트의 정확성에 크게 의존한다는 단점이 있습니다 16.
3. ConvNeXt 모델은 이미지 분류에 특화된 모델로, ImageNet과 같은 대규모 데이터셋에서 학습된 모델입니다 17. 복잡한 이미지 특징을 학습하는 데 뛰어나지만, 특정 도메인의 데이터셋에 적용할 때 성능 저하가 발생할 수 있습니다.

이러한 모델들은 서로 밀접하게 연결되어 있어, 한 모델의 성능이 다른 모델의 성능에 직접적인 영향을 미칩니다 3. 따라서 각 모델의 고유한 특성과 한계를 이해하고, 적절하게 활용하는 것이 중요합니다.

단, **YOLO**나 **SAM** 같은 모델은 **일반적인 객체 검출**이나 **일반적인 세그멘테이션**을 위한 데이터셋(예: COCO, VOC 등)에서 훈련되었기 때문에 본 과제 내 대상 데이터 셋으로 잡혀있

는 내용과 사전학습 이미지의 내용이 이치 하지 않음. 특정 특징이 다르거나, 검출할 객체나 세그멘테이션 형태가 모델의 사전 학습 데이터와 달라서 성능이 떨어질 수 있습니다.

이에 **ImageNet**과 같은 대규모 이미지 데이터셋을 활용하여 모델이 잘 작동할 수 있는 내용의 테스트 방향은 한정합니다. 대상데이터셋 중 "kritikseth_fruit-and-vegetable"과 사전 학습 데이터 셋 중 "이미지넷"에는 바나나를 포함한 다양한 과일들이 포함되어 있기 때문에, **바나나** 구분에 대한 ConvNeXt, YOLO와 SAM 모델을 활용한 접근은 충분히 가능하고, 초기 실험을 위한 좋은 시작점으로 테스트를 행 합니다.

1. 사전 학습 모델 활용

- **ImageNet**은 다양한 과일들을 포함하고 있으며, 바나나 역시 그 중 하나입니다. 따라서 **YOLO, SAM, ConvNeXt** 모델을 **ImageNet**에서 학습된 버전을 사용하여 초기 실험을 수행할 수 있습니다.
- **kritikseth_fruit-and-vegetable** 데이터셋은 과일과 채소를 포함하는 이미지들이기 때문에, 바나나를 포함한 이미지를 다룰 수 있는 좋은 실험 데이터셋이 됩니다.

2. YOLO 모델을 사용한 바나나 객체 검출

- **YOLO**는 객체 검출에 특화된 모델로, 바나나와 같은 특정 객체를 **이미지 내에서 정확히 검출**하는 데 사용될 수 있습니다.
- **실험 방식:**
 1. **YOLO 모델 준비:** 사전 학습된 YOLO 모델을 사용하여 바나나가 포함된 이미지를 처리합니다.
 2. **객체 검출:** 바나나가 있는 위치를 **bounding box**로 검출합니다.
 3. **성능 평가:** 검출된 바나나의 **정확도(Precision)**, **재현율(Recall)**, **IoU(Intersection over Union)** 등을 평가합니다.

3. SAM 모델을 사용한 바나나 세그멘테이션

- **SAM** 모델은 객체의 경계를 정확히 추출하고, 바나나와 같은 객체를 **세그멘테이션**하여 마스크를 생성할 수 있습니다.
- **실험 방식:**
 1. **SAM 모델 준비:** ImageNet에서 학습된 SAM 모델을 사용하여 바나나의 **세그멘테이션**을 수행합니다.
 2. **세그멘테이션 결과 분석:** 바나나가 포함된 이미지에서 **마스크**를 생성하여, 바나나의 정확한 형태를 추출합니다.

3. **성능 평가:** 세그멘테이션의 정확도와 마스크 생성의 품질을 **시각적으로 확인**하고, **정밀도와 회수율** 등을 측정합니다.

4. ConvNeXt 모델을 사용한 바나나 분류

- **ConvNeXt**는 최신 **컨볼루션 신경망** 모델로, 이미지 분류에서 좋은 성능을 보입니다. 이를 바탕으로 **바나나 분류** 작업을 테스트할 수 있습니다.
- **실험 방식:**
 1. **ConvNeXt 모델 준비:** 사전 학습된 ConvNeXt 모델을 활용하여 바나나 이미지에 대한 분류 작업을 진행합니다.
 2. **이미지 분류:** 바나나 이미지를 모델에 입력하여 **바나나** 클래스로 분류하는 작업을 수행합니다.
 3. **성능 평가:** 모델이 바나나를 얼마나 정확히 분류하는지 평가하며, **정확도** (Accuracy)와 **F1 score** 등을 측정합니다.

연구 보고서 분석 결과 정리

분석 코드 및 결과 파일 첨부

YOLO 샘플 코드

FindBanana_YOLO.py

yolo_banana_result.zip

YOLO+SAM 샘플 코드

FindBanana_YOLO_SAM.py

yolo_sam_result.zip

YOLO+SAM+ConvNeXt 테스트 코드

[FindBanana_YOLO_SAM_ConvNeXt.py](#)

▼ 테스트 결과.

0: 640×576 (no detections), 24.3ms

Speed: 4.1ms preprocess, 24.3ms inference, 19.6ms postprocess per image at shape (1, 3, 640, 576)

selenium_images/_banana/banana_0.jpg: 바나나가 검출되지 않았습니다.

0: 576×640 (no detections), 25.6ms

Speed: 1.6ms preprocess, 25.6ms inference, 0.3ms postprocess per image at shape (1, 3, 576, 640)

selenium_images/_banana/banana_1.jpg: 바나나가 검출되지 않았습니다.

0: 480×640 1 banana, 28.3ms

Speed: 1.0ms preprocess, 28.3ms inference, 56.2ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_10.jpg: 바나나일 확률: 93.26%

0: 384×640 3 bananas, 25.0ms

Speed: 0.8ms preprocess, 25.0ms inference, 1.1ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_11.jpg: 바나나일 확률: 79.85%

이미지를 로드할 수 없습니다: selenium_images/_banana/banana_12.jpg

0: 640×640 1 banana, 4.0ms

Speed: 1.3ms preprocess, 4.0ms inference, 0.8ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_13.jpg: 바나나일 확률: 0.08%

0: 448×640 1 bench, 26.5ms

Speed: 1.3ms preprocess, 26.5ms inference, 0.9ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_14.jpg: 바나나일 확률: 0.22%

0: 448×640 1 surfboard, 1 toilet, 9.8ms

Speed: 2.4ms preprocess, 9.8ms inference, 1.2ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_15.jpg: 바나나일 확률: 89.55%

0: 352×640 1 frisbee, 24.5ms

Speed: 1.6ms preprocess, 24.5ms inference, 1.1ms postprocess per image at shape (1, 3, 352, 640)

selenium_images/_banana/banana_16.jpg: 바나나일 확률: 94.00%

0: 640×640 1 vase, 1 toothbrush, 15.2ms

Speed: 2.0ms preprocess, 15.2ms inference, 1.1ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_17.jpg: 바나나일 확률: 90.84%

0: 384×640 (no detections), 13.2ms

Speed: 1.9ms preprocess, 13.2ms inference, 0.4ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_18.jpg: 바나나가 검출되지 않았습니다.

0: 512×640 (no detections), 25.5ms

Speed: 1.8ms preprocess, 25.5ms inference, 0.4ms postprocess per image at shape (1, 3, 512, 640)

selenium_images/_banana/banana_19.jpg: 바나나가 검출되지 않았습니다.

0: 480×640 1 banana, 10.9ms

Speed: 2.1ms preprocess, 10.9ms inference, 1.0ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_2.jpg: 바나나일 확률: 90.90%

0: 416×640 1 toothbrush, 24.2ms

Speed: 1.6ms preprocess, 24.2ms inference, 1.0ms postprocess per image at shape (1, 3, 416, 640)

selenium_images/_banana/banana_20.jpg: 바나나일 확률: 92.28%

0: 416×640 (no detections), 13.9ms

Speed: 2.2ms preprocess, 13.9ms inference, 0.8ms postprocess per image at shape (1, 3, 416, 640)

selenium_images/_banana/banana_21.jpg: 바나나가 검출되지 않았습니다.

O: 640×640 (no detections), 18.0ms

Speed: 2.2ms preprocess, 18.0ms inference, 0.4ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_22.jpg: 바나나가 검출되지 않았습니다.

O: 448×640 1 frisbee, 11.6ms

Speed: 3.5ms preprocess, 11.6ms inference, 1.1ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_23.jpg: 바나나일 확률: 93.17%

O: 448×640 1 frisbee, 12.7ms

Speed: 1.7ms preprocess, 12.7ms inference, 1.4ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_24.jpg: 바나나일 확률: 93.17%

O: 384×640 (no detections), 13.4ms

Speed: 1.7ms preprocess, 13.4ms inference, 0.4ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_25.jpg: 바나나가 검출되지 않았습니다.

O: 480×640 1 cup, 15.2ms

Speed: 1.9ms preprocess, 15.2ms inference, 1.1ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_26.jpg: 바나나일 확률: 89.78%

O: 640×640 1 frisbee, 1 surfboard, 2 scissorss, 14.7ms

Speed: 2.1ms preprocess, 14.7ms inference, 1.1ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_27.jpg: 바나나일 확률: 87.72%

O: 384×640 1 frisbee, 12.9ms

Speed: 3.1ms preprocess, 12.9ms inference, 1.1ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_28.jpg: 바나나일 확률: 90.01%

O: 640×608 (no detections), 28.0ms

Speed: 2.2ms preprocess, 28.0ms inference, 0.4ms postprocess per image at shape (1, 3, 640, 608)

selenium_images/_banana/banana_29.jpg: 바나나가 검출되지 않았습니다.

0: 512×640 (no detections), 13.5ms

Speed: 1.9ms preprocess, 13.5ms inference, 0.4ms postprocess per image at shape (1, 3, 512, 640)

selenium_images/_banana/banana_3.jpg: 바나나가 검출되지 않았습니다.

0: 512×640 1 bowl, 15.7ms

Speed: 1.8ms preprocess, 15.7ms inference, 1.2ms postprocess per image at shape (1, 3, 512, 640)

selenium_images/_banana/banana_30.jpg: 바나나일 확률: 87.63%

0: 608×640 1 surfboard, 1 knife, 26.4ms

Speed: 1.9ms preprocess, 26.4ms inference, 1.1ms postprocess per image at shape (1, 3, 608, 640)

selenium_images/_banana/banana_31.jpg: 바나나일 확률: 92.28%

0: 512×640 1 scissors, 13.2ms

Speed: 2.5ms preprocess, 13.2ms inference, 1.0ms postprocess per image at shape (1, 3, 512, 640)

selenium_images/_banana/banana_32.jpg: 바나나일 확률: 89.39%

0: 480×640 2 vases, 16.3ms

Speed: 1.7ms preprocess, 16.3ms inference, 1.1ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_33.jpg: 바나나일 확률: 46.87%

0: 640×640 1 cake, 14.7ms

Speed: 2.2ms preprocess, 14.7ms inference, 1.0ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_34.jpg: 바나나일 확률: 92.92%

0: 480×640 4 vases, 12.8ms

Speed: 1.9ms preprocess, 12.8ms inference, 1.1ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_35.jpg: 바나나일 확률: 89.57%

0: 640×640 1 baseball glove, 15.1ms

Speed: 2.2ms preprocess, 15.1ms inference, 1.0ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_36.jpg: 바나나일 확률: 93.69%

0: 640×480 1 person, 50.9ms

Speed: 3.4ms preprocess, 50.9ms inference, 1.1ms postprocess per image at shape (1, 3, 640, 480)

selenium_images/_banana/banana_37.jpg: 바나나일 확률: 89.60%

0: 448×640 (no detections), 12.9ms

Speed: 1.8ms preprocess, 12.9ms inference, 0.4ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_38.jpg: 바나나가 검출되지 않았습니다.

0: 448×640 1 frisbee, 12.8ms

Speed: 1.8ms preprocess, 12.8ms inference, 1.2ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_39.jpg: 바나나일 확률: 89.37%

0: 384×640 1 banana, 16.7ms

Speed: 1.9ms preprocess, 16.7ms inference, 1.1ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_4.jpg: 바나나일 확률: 94.75%

0: 416×640 (no detections), 14.9ms

Speed: 1.8ms preprocess, 14.9ms inference, 0.4ms postprocess per image at shape (1, 3, 416, 640)

selenium_images/_banana/banana_40.jpg: 바나나가 검출되지 않았습니다.

0: 448×640 1 bench, 13.0ms

Speed: 1.4ms preprocess, 13.0ms inference, 1.1ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_41.jpg: 바나나일 확률: 0.01%

0: 480×640 (no detections), 14.3ms

Speed: 1.9ms preprocess, 14.3ms inference, 0.4ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_42.jpg: 바나나가 검출되지 않았습니다.

O: 448×640 1 person, 1 banana, 12.5ms

Speed: 3.5ms preprocess, 12.5ms inference, 1.9ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_43.jpg: 바나나일 확률: 92.54%

O: 384×640 1 frisbee, 14.2ms

Speed: 1.7ms preprocess, 14.2ms inference, 2.1ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_44.jpg: 바나나일 확률: 89.87%

O: 640×640 1 scissors, 15.6ms

Speed: 3.0ms preprocess, 15.6ms inference, 1.2ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_45.jpg: 바나나일 확률: 94.52%

O: 384×640 (no detections), 17.2ms

Speed: 1.8ms preprocess, 17.2ms inference, 0.4ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_46.jpg: 바나나가 검출되지 않았습니다.

O: 384×640 2 vases, 1 toothbrush, 10.1ms

Speed: 1.7ms preprocess, 10.1ms inference, 2.4ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_47.jpg: 바나나일 확률: 92.13%

O: 384×640 2 vases, 1 toothbrush, 10.6ms

Speed: 1.7ms preprocess, 10.6ms inference, 1.3ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_48.jpg: 바나나일 확률: 92.13%

O: 640×640 (no detections), 18.6ms

Speed: 2.2ms preprocess, 18.6ms inference, 1.5ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_49.jpg: 바나나가 검출되지 않았습니다.

O: 448×640 (no detections), 110.7ms

Speed: 2.2ms preprocess, 110.7ms inference, 3.1ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_5.jpg: 바나나가 검출되지 않았습니다.

0: 448×640 1 cup, 1 bowl, 7.0ms

Speed: 1.7ms preprocess, 7.0ms inference, 1.1ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_50.jpg: 바나나일 확률: 0.06%

0: 448×640 1 bird, 2 bananas, 13.6ms

Speed: 2.1ms preprocess, 13.6ms inference, 1.5ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_51.jpg: 바나나일 확률: 90.77%

0: 448×640 1 bird, 2 bananas, 12.9ms

Speed: 1.7ms preprocess, 12.9ms inference, 2.2ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_52.jpg: 바나나일 확률: 90.77%

0: 448×640 (no detections), 12.0ms

Speed: 1.7ms preprocess, 12.0ms inference, 0.4ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_53.jpg: 바나나가 검출되지 않았습니다.

0: 640×640 1 scissors, 14.3ms

Speed: 2.3ms preprocess, 14.3ms inference, 1.8ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_54.jpg: 바나나일 확률: 90.90%

0: 480×640 1 scissors, 12.2ms

Speed: 2.2ms preprocess, 12.2ms inference, 1.1ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_55.jpg: 바나나일 확률: 91.99%

0: 448×640 1 banana, 11.1ms

Speed: 1.9ms preprocess, 11.1ms inference, 1.1ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_56.jpg: 바나나일 확률: 93.40%

0: 384×640 1 vase, 16.4ms

Speed: 3.2ms preprocess, 16.4ms inference, 1.1ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_57.jpg: 바나나일 확률: 91.64%

0: 480×640 1 suitcase, 10.9ms

Speed: 1.8ms preprocess, 10.9ms inference, 3.1ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_58.jpg: 바나나일 확률: 83.49%

0: 640×448 (no detections), 24.8ms

Speed: 2.6ms preprocess, 24.8ms inference, 0.4ms postprocess per image at shape (1, 3, 640, 448)

selenium_images/_banana/banana_59.jpg: 바나나가 검출되지 않았습니다.

0: 512×640 1 scissors, 13.9ms

Speed: 2.2ms preprocess, 13.9ms inference, 1.2ms postprocess per image at shape (1, 3, 512, 640)

selenium_images/_banana/banana_6.jpg: 바나나일 확률: 13.66%

0: 512×640 1 banana, 11.0ms

Speed: 1.8ms preprocess, 11.0ms inference, 2.5ms postprocess per image at shape (1, 3, 512, 640)

selenium_images/_banana/banana_60.jpg: 바나나일 확률: 85.14%

0: 512×640 (no detections), 10.8ms

Speed: 2.0ms preprocess, 10.8ms inference, 0.5ms postprocess per image at shape (1, 3, 512, 640)

selenium_images/_banana/banana_61.jpg: 바나나가 검출되지 않았습니다.

0: 352×640 1 cup, 13.4ms

Speed: 1.6ms preprocess, 13.4ms inference, 1.1ms postprocess per image at shape (1, 3, 352, 640)

selenium_images/_banana/banana_62.jpg: 바나나일 확률: 89.61%

0: 512×640 2 scissorss, 11.4ms

Speed: 2.4ms preprocess, 11.4ms inference, 1.1ms postprocess per image at shape (1, 3, 512, 640)

selenium_images/_banana/banana_63.jpg: 바나나일 확률: 87.37%
0: 640×448 (no detections), 11.7ms
Speed: 2.0ms preprocess, 11.7ms inference, 0.4ms postprocess per image at shape (1, 3, 640, 448)

selenium_images/_banana/banana_64.jpg: 바나나가 검출되지 않았습니다.
0: 448×640 (no detections), 11.8ms
Speed: 1.7ms preprocess, 11.8ms inference, 0.4ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_65.jpg: 바나나가 검출되지 않았습니다.
0: 448×640 (no detections), 9.7ms
Speed: 1.7ms preprocess, 9.7ms inference, 0.4ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_66.jpg: 바나나가 검출되지 않았습니다.
0: 288×640 (no detections), 29.5ms
Speed: 1.2ms preprocess, 29.5ms inference, 0.5ms postprocess per image at shape (1, 3, 288, 640)

selenium_images/_banana/banana_67.jpg: 바나나가 검출되지 않았습니다.
0: 448×640 1 banana, 11.3ms
Speed: 3.1ms preprocess, 11.3ms inference, 1.1ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_68.jpg: 바나나일 확률: 86.90%
0: 480×640 3 kites, 13.6ms
Speed: 2.1ms preprocess, 13.6ms inference, 1.0ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_69.jpg: 바나나일 확률: 0.12%
0: 448×640 1 toilet, 12.7ms
Speed: 2.0ms preprocess, 12.7ms inference, 1.5ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_7.jpg: 바나나일 확률: 90.86%
0: 640×640 1 bed, 15.1ms

Speed: 1.9ms preprocess, 15.1ms inference, 1.0ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_70.jpg: 바나나일 확률: 96.64%

0: 448×640 1 vase, 10.7ms

Speed: 1.8ms preprocess, 10.7ms inference, 1.1ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_71.jpg: 바나나일 확률: 90.08%

이미지를 로드할 수 없습니다: selenium_images/_banana/banana_72.jpg

0: 640×608 (no detections), 14.0ms

Speed: 2.3ms preprocess, 14.0ms inference, 0.4ms postprocess per image at shape (1, 3, 640, 608)

selenium_images/_banana/banana_73.jpg: 바나나가 검출되지 않았습니다.

0: 448×640 (no detections), 13.0ms

Speed: 1.8ms preprocess, 13.0ms inference, 0.4ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_74.jpg: 바나나가 검출되지 않았습니다.

0: 448×640 1 knife, 11.9ms

Speed: 1.8ms preprocess, 11.9ms inference, 1.3ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_76.jpg: 바나나일 확률: 95.76%

0: 640×640 1 vase, 12.9ms

Speed: 3.1ms preprocess, 12.9ms inference, 1.1ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_77.jpg: 바나나일 확률: 49.07%

0: 608×640 (no detections), 14.5ms

Speed: 2.3ms preprocess, 14.5ms inference, 0.4ms postprocess per image at shape (1, 3, 608, 640)

selenium_images/_banana/banana_78.jpg: 바나나가 검출되지 않았습니다.

0: 384×640 1 bird, 10.2ms

Speed: 1.6ms preprocess, 10.2ms inference, 1.1ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_79.jpg: 바나나일 확률: 86.73%
0: 384×640 1 frisbee, 1 vase, 14.2ms
Speed: 1.8ms preprocess, 14.2ms inference, 1.1ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_8.jpg: 바나나일 확률: 91.95%
0: 384×640 1 bird, 12.8ms
Speed: 1.8ms preprocess, 12.8ms inference, 1.1ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_80.jpg: 바나나일 확률: 86.73%
0: 448×640 1 surfboard, 1 toothbrush, 11.1ms
Speed: 2.0ms preprocess, 11.1ms inference, 1.2ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_81.jpg: 바나나일 확률: 91.68%
0: 384×640 2 bananas, 2 scissorss, 10.4ms
Speed: 1.7ms preprocess, 10.4ms inference, 1.0ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_82.jpg: 바나나일 확률: 91.64%
0: 640×640 (no detections), 12.6ms
Speed: 2.3ms preprocess, 12.6ms inference, 0.5ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_83.jpg: 바나나가 검출되지 않았습니다.
0: 448×640 (no detections), 14.0ms
Speed: 1.4ms preprocess, 14.0ms inference, 0.4ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_84.jpg: 바나나가 검출되지 않았습니다.
0: 480×640 1 dining table, 10.7ms
Speed: 1.7ms preprocess, 10.7ms inference, 1.0ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_85.jpg: 바나나일 확률: 49.95%
0: 448×640 (no detections), 10.8ms

Speed: 1.9ms preprocess, 10.8ms inference, 0.4ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_86.jpg: 바나나가 검출되지 않았습니다.

0: 352×640 3 surfboards, 1 banana, 10.0ms

Speed: 1.4ms preprocess, 10.0ms inference, 1.0ms postprocess per image at shape (1, 3, 352, 640)

selenium_images/_banana/banana_87.jpg: 바나나일 확률: 71.51%

0: 384×640 (no detections), 10.9ms

Speed: 2.3ms preprocess, 10.9ms inference, 0.4ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_88.jpg: 바나나가 검출되지 않았습니다.

0: 480×640 1 cup, 11.4ms

Speed: 1.8ms preprocess, 11.4ms inference, 1.9ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_89.jpg: 바나나일 확률: 90.86%

0: 384×640 (no detections), 12.5ms

Speed: 1.8ms preprocess, 12.5ms inference, 0.4ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_9.jpg: 바나나가 검출되지 않았습니다.

0: 640×480 1 surfboard, 12.7ms

Speed: 4.3ms preprocess, 12.7ms inference, 1.1ms postprocess per image at shape (1, 3, 640, 480)

selenium_images/_banana/banana_90.jpg: 바나나일 확률: 90.99%

0: 640×448 (no detections), 11.5ms

Speed: 2.2ms preprocess, 11.5ms inference, 0.3ms postprocess per image at shape (1, 3, 640, 448)

selenium_images/_banana/banana_91.jpg: 바나나가 검출되지 않았습니다.

0: 416×640 1 cake, 13.3ms

Speed: 1.6ms preprocess, 13.3ms inference, 1.0ms postprocess per image at shape (1, 3, 416, 640)

selenium_images/_banana/banana_92.jpg: 바나나일 확률: 89.66%

0: 352×640 1 cup, 10.7ms

Speed: 1.5ms preprocess, 10.7ms inference, 1.1ms postprocess per image at shape (1, 3, 352, 640)

selenium_images/_banana/banana_93.jpg: 바나나일 확률: 0.02%

0: 448×640 (no detections), 10.9ms

Speed: 1.9ms preprocess, 10.9ms inference, 0.4ms postprocess per image at shape (1, 3, 448, 640)

selenium_images/_banana/banana_94.jpg: 바나나가 검출되지 않았습니다.

0: 640×640 (no detections), 11.9ms

Speed: 1.8ms preprocess, 11.9ms inference, 0.5ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_95.jpg: 바나나가 검출되지 않았습니다.

0: 352×640 (no detections), 10.4ms

Speed: 2.4ms preprocess, 10.4ms inference, 0.4ms postprocess per image at shape (1, 3, 352, 640)

selenium_images/_banana/banana_96.jpg: 바나나가 검출되지 않았습니다.

0: 480×640 2 bananas, 10.3ms

Speed: 1.8ms preprocess, 10.3ms inference, 1.0ms postprocess per image at shape (1, 3, 480, 640)

selenium_images/_banana/banana_97.jpg: 바나나일 확률: 90.34%

0: 384×640 1 frisbee, 1 surfboard, 11.1ms

Speed: 1.7ms preprocess, 11.1ms inference, 1.0ms postprocess per image at shape (1, 3, 384, 640)

selenium_images/_banana/banana_98.jpg: 바나나일 확률: 91.46%

0: 640×640 (no detections), 12.6ms

Speed: 2.3ms preprocess, 12.6ms inference, 0.4ms postprocess per image at shape (1, 3, 640, 640)

selenium_images/_banana/banana_99.jpg: 바나나가 검출되지 않았습니다.

총 62개의 이미지 처리 완료. 평균 바나나 확률: 78.29%

결론:

본 연구에서는 ImageNet 데이터셋으로 사전 학습된 YOLO, SAM, 그리고 ConvNeXt 모델을 활용하여 특정 바나나 이미지 데이터셋에 대한 객체 검출, 세분화 및 분류 테스트를 진행했습니다. 분석 결과, 최신 딥러닝 모델임에도 불구하고 기본적인 CNN 모델인 ResNet50보다 낮은 바나나 분류 성능을 보이는 것으로 나타났습니다. 이는 YOLO 및 SAM 모델에 대한 사용자의 이해 부족으로 인해 부적절한 테스트 기획이 이루어진 근본적인 원인으로 판단됩니다.

상세 분석:

애초에 본 테스트 기획이 잘못 설정된 주요 원인은 다음과 같습니다.

1. 모델의 목표 불일치:

- **YOLO (You Only Look Once):** YOLO 모델은 실시간 객체 검출(Object Detection)을 주요 목표로 설계된 모델입니다. 이미지 내에 존재하는 객체의 위치를 바운딩 박스 형태로 찾아내고, 해당 객체의 클래스를 예측하는 데 특화되어 있습니다. 본 실험에서는 YOLO 모델의 검출 결과를 활용하여 SAM 모델의 입력으로 사용하고, 최종적으로 ConvNeXt 모델의 분류 성능을 평가하는 방식으로 진행되었습니다. 그러나 YOLO 모델 자체는 이미지 내에 '바나나' 객체가 존재하는지 여부를 높은 정확도로 판단하는 것을 목표로 하지 않습니다. 객체의 위치를 정확하게 파악하는 것이 더 중요한 성능 지표입니다. 따라서 YOLO 모델이 바나나를 검출하지 못하거나 잘못된 위치에 바운딩 박스를 생성하는 경우, 후속 단계인 SAM 및 ConvNeXt 모델의 성능 저하를 야기할 수밖에 없습니다.
- **SAM (Segment Anything Model):** SAM 모델은 프롬프트(점, 박스 등)를 기반으로 이미지 내의 어떠한 객체든 마스크 형태로 분할(Segmentation)하는 것을 목표로 합니다. 본 실험에서는 YOLO 모델이 예측한 바운딩 박스를 SAM 모델의 프롬프트로 사용하여 바나나 영역을 추출했습니다. SAM 모델은 주어진 프롬프트에 따라 마스크를 생성하는 데 탁월하지만, 입력 프롬프트 자체가 부정확하다면 잘못된 세그멘테이션 결과를 초래합니다. 즉, YOLO 모델의 오검출 또는 부정확한 바운딩 박스는 SAM 모델이 바나나가 아닌 영역을 세그멘테이션하거나, 바나나의 일부분만을 잘라내는 결과를 낳아 ConvNeXt 모델의 분류 성능에 부정적인 영향을 미칩니다.

2. 테스트 파이프라인의 의존성:

본 실험은 YOLO의 객체 검출 결과에 SAM의 세분화 성능이 의존적이고, 최종적으로 ConvNeXt의 분류 성능은 앞선 두 모델의 결과에 모두 영향을 받는 구조로 설계되었습니다. 이러한 파이프라인 구조에서는 하나의 모델에서 발생한 오류가 후속 모델로 그대로 전파되어 전체적인 성능 저하를 초래할 가능성이 매우 높습니다. 특히, YOLO 모델의 낮은 바나나 검출 성능은 SAM 모델이 적절한 입력 프롬프트를 받지 못하게 하여 정확한 바나나 영역 추출을 어렵게 만들고, 이는 ConvNeXt 모델이 바나나 특징을 제대로 학습하지 못한 이미지 조각을 입력받아 분류 오류를 발생시키는 주요 원인이 됩니다.

3. 모델의 사전 학습 데이터와 대상 데이터셋의 차이:

YOLO, SAM, 그리고 ConvNeXt 모델은 COCO, ImageNet과 같은 대규모의 일반적인 이미지 데이터셋으로 사전 학습되었습니다. 이러한 데이터셋에는 다양한 객체와 장면이 포함되어 있지만, 본 연구에서 사용된 특정 바나나 이미지 데이터셋의 특징(예: 이미지 품질, 바나나의 종류 및 형태, 촬영 환경 등)과는 차이가 있을 수 있습니다. 특히, YOLO 모델은 사전 학습 시 특정 형태의 바나나 객체에 익숙해져 있을 수 있으며, 대상 데이터셋의 바나나가 다른 특징을 가질 경우 검출 성능이 저하될 수 있습니다. 마찬가지로, ConvNeXt 모델 또한 ImageNet의 바나나 클래스를 학습했지만, 세분화 과정에서 노이즈가 섞이거나 바나나의 온전한 형태를 갖추지 못한 이미지를 입력받을 경우 분류 정확도가 떨어질 수 있습니다.

4. 평가 지표의 부적절성:

본 연구의 최종 목표는 바나나 '분류' 성능을 평가하는 것이었으나, 테스트 파이프라인은 객체 '검출'과 '세분화' 단계를 포함하고 있습니다. 만약 YOLO 모델이 바나나를 정확하게 검출하지 못하면, 후속 단계는 아예 진행되지 않거나 잘못된 대상을 처리하게 됩니다. 이러한 상황에서 최종 분류 성능이 낮게 나오는 것은 당연한 결과이며, 이는 YOLO와 SAM 모델 자체의 문제라기보다는 테스트 설계의 문제로 해석해야 합니다.

결론적으로, 본 연구에서 YOLO와 SAM 모델에 대한 사용자의 이해 부족으로 인해 객체 검출 및 세분화에 특화된 모델을 바나나 분류 성능 평가를 위한 파이프라인의 중간 단계로 부적절하게 활용하였습니다. 이는 기본적인 CNN 모델보다 낮은 분류 성능이라는 예상치 못한 결과를 초래했으며, 향후 연구에서는 모델의 목표와 특성을 정확히 이해하고, 평가하고자 하는 목표에 부합하는 적절한 모델 선택 및 테스트 설계를 수행해야 할 것입니다.