

이해도평가_올로

문제1) YOLO는 이미지를 여러 번 스캔하는 대신 단 한 번의 forward pass로 객체를 탐지한다.

정답: O

이미지 전체를 그리드로 나눔

단 한 번의 forward pass로 모든 그리드 셀에서 동시에:

Bounding box 좌표

Confidence score

Class probability

→ 를 한꺼번에 예측

문제2) IOU는 객체 탐지 모델의 성능을 평가할 때 사용되는 중요한 지표로, 예측된 바운딩 박스(predicted bounding box)와 실제 바운딩 박스(ground truth bounding box) 간의 겹치는 영역을 두 박스의 합집합 영역으로 나눈 값이다. 일반적으로 IOU 값이 0.5 이상이면 올바른 탐지로 간주하며, COCO 데이터셋과 같은 경우에는 더 엄격한 기준인 0.75나 0.95까지도 사용한다.

IOU 값이 1에 가까울수록 예측이 정확하다는 의미이며, 0이면 두 박스가 전혀 겹치지 않는다는 의미이다

정답: O

IoU 공식:

$$IoU = \frac{\text{합집합영역} \cap \text{교집합영역}}{\text{합집합영역}}$$

범위: 0 ~ 1

0: 겹침 없음

1: 완벽히 일치

임계값: 0.5(일반), 0.75/0.95(COCO 엄격 기준)

문제3) YOLO는 작은 객체들이 밀집해 있는 경우에도 R-CNN만큼 정확하게 탐지할 수 있다

정답: X

YOLO의 한계:

이미지를 그리드로 나누고, 각 셀당 제한된 수의 박스만 예측

작은 객체가 밀집 → 한 셀에 여러 객체 → 일부 누락

R-CNN 계열:

Region Proposal 방식으로 더 세밀하게 탐지

작은/밀집 객체에 더 강함

YOLO는 속도에서 우세, 작은 객체 밀집 탐지 정확도는 R-CNN이 우

문제4) YOLO는 이미지를 그리드(grid)로 나누고, 각 그리드 셀이 바운딩 박스와 클래스 확률을 예측한다

정답: O

YOLO 작동 방식:

이미지를 $S \times S$ 그리드로 분할 (예: 7×7)

각 그리드 셀이 예측:

B개의 바운딩 박스 ($x, y, w, h, \text{confidence}$)

C개의 클래스 확률

문제5) FP(False Positive)는 모델이 객체가 있다고 예측했지만 실제로는 없는 경우를 의미하고, FN(False Negative)은 실제로 객체가 있지만 모델이 탐지하지 못한 경우를 의미한다

문제6) YOLO 모델을 학습시킬 때 레이블 파일은 각 객체마다 클래스 이름(예: "person", "car", "dog")을 텍스트로 직접 저장하며, 바운딩 박스 좌표는 픽셀 단위의 절대값($x_{\min}, y_{\min}, x_{\max}, y_{\max}$)으로 표시한다

정답: X

YOLO 레이블 파일 실제 형식:

정수 인덱스 (0, 1, 2...) / 정규화된 값 (0~1) / $x_{\text{center}}, y_{\text{center}}, \text{width}, \text{height}$

0 0.5 0.4 0.3 0.6 → 클래스 0, 중심(0.5, 0.4), 너비 0.3, 높이 0.6

$0.5 - (0.3/2) \Rightarrow x_{\text{시작점}} * \text{이미지너비}, \quad 0.4 - (0.6/2) \Rightarrow y_{\text{시작점}} * \text{이미지높이}$

문제7) YOLO는 각 그리드 셀에서 하나의 클래스만 예측할 수 있으므로, 한 그리드 셀 안에 서로 다른 클래스의 객체가 여러 개 겹쳐 있으면 모든 객체를 정확하게 탐지할 수 있다. 이것이 YOLO가 빠른 속도를 유지하면서도 높은 정확도를 달성하는 핵심 원리이다

정답: X

논리 오류:

각 그리드 셀이 **하나의 클래스만 예측** → 맞는 사실

따라서 여러 클래스 객체가 겹치면 "**모든 객체를 탐지할 수 있다**" → 틀림

실제:

한 셀에 다른 클래스 객체가 겹치면 → **일부 누락** (YOLO의 한계점)

이것은 정확도 향상 원리가 아니라 **YOLO의 단점**

빠른 속도의 **대가**로 밀집/겹침 상황에서 정확도가 떨어짐

문제8) YOLO는 two-stage detector(R-CNN 계열)보다 일반적으로 탐지 속도가 느리다

정답: X

YOLO가 빠른 이유:

단 한 번의 forward pass로 탐지 완료

"You Only Look Once" = 한 번만 본다

YOLO는 R-CNN보다 **훨씬 빠름** (반대로 설명됨)

문제9) Non-Maximum Suppression(NMS) 과정에서 IOU는 중복된 바운딩 박스를 제거하는 기준으로 사용되는데, 같은 객체에 대해 여러 개의 바운딩 박스가 예측되었을 때 confidence score가 가장 높은 박스를 기준으로 다른 박스들과의 IOU를 계산한다. 만약 IOU가 설정된 임계값(threshold)보다 높으면 해당 박스들을 중복으로 간주하여 제거하고, 이 과정을 반복하여 각 객체당 하나의 최적 바운딩 박스만 남긴다. 이때 IOU 임계값을 너무 낮게 설정하면 서로 가까이 있는 별개의 객체들을 하나로 합쳐버릴 수 있다.

정답: O

Confidence 가장 높은 박스 선택

다른 박스들과 IoU 계산

IoU > 임계값 → 중복 제거

반복하여 최적 박스만 남김

문제10) Image Classification 모델은 이미지에서 여러 객체를 동시에 인식할 수 있지만 위치를 알 수 없고, Object Detection 모델은 한 번에 하나의 객체만 찾을 수 있지만 정확한 위치를 바운딩 박스로 표시할 수 있다.

정답: X

Image Classification: 이미지 → "이건 고양이이다" (위치 , 단일 클래스)

Object Detection: 이미지 → "여기 고양이, 저기 개, 거기 사람" (위치 , 다중 객체)

두 모델의 설명이 서로 뒤바뀌어 있음