사람의 시각 체계는 빠르고 정확하여, 운전 중에도 빠르게 주변 환경을 인식할 수 있습니다. 이와 같은 능력을 가진 객체 검출 모델을 개발하면 자율주행차 기술이 크게 발전할 것입니다. 기존 객체 검출 모델은 분류기를 검출기로 재정의하여 사용합니다. 분류는 이미지를 보고 그것이 무엇인지 판단하는 것이고, 객체 검출은 이미지 내에서 객체의 위치를 판단하는 것입니다. 기존 모델인 DPM과 R-CNN은 각각 슬라이딩 윈도 방식과 region proposal 방식을 사용합니다. 그러나 R-CNN은 복잡한 절차와 느린 속도가 단점입니다.

YOLO 연구진은 객체 검출을 하나의 회귀 문제로 보고 절차를 단순화했습니다. 이미지의 픽셀로부터 bounding box의 위치와 클래스 확률을 구하는 과정을 하나의 회귀 문제로 재정의하여, 빠르게 객체를 검출할 수 있게 했습니다. 이름 그대로 YOLO(You Only Look Once)는 이미지를 한 번만 보면 객체를 검출할 수 있습니다.

YOLO는 단순하면서도 효율적인 객체 검출 모델입니다. 하나의 컨볼루션 네트워크가 여러 bounding box와 클래스 확률을 동시에 계산하며, 이미지 전체를 학습하여 검출 성능을 최적화합니다. YOLO의 주요 장점은 다음과 같습니다:

1. \*\*속도\*\*: YOLO는 객체 검출을 하나의 회귀 문제로 단순화하여 매우 빠릅니다. 기존의 복잡한 파이프라인이 필요 없으며, 새로운 이미지를 신경망에 넣기만 하면 객체를 검출할 수 있습니다. 기본 네트워크는 Titan X GPU에서 1초에 45 프레임을, 빠른 버전(Fast YOLO)은 1초에 150 프레임을 처리합니다. 이는 실시간 동영상 처리에 충분한 속도입니다.

2. \*\*정확성\*\*: YOLO는 다른 실시간 객체 검출 모델보다 2배 이상의 mAP(mean average precision)를 자랑합니다.

YOLO는 객체 검출의 모든 요소를 하나의 신경망으로 통합한 모델입니다. 작동 방식은 다음과 같습니다:

- 이미지를 S x S 그리드로 나눕니다.

- 각 그리드 셀은 B개의 bounding box와 그에 대한 confidence score를 예측합니다.

- class-specific confidence score는 각 bounding box에 특정 클래스 객체가 있을 확률과 bounding box의 정확성을 나타냅니다.

- 최종 예측 결과는 (7 x 7 x 30) 크기의 텐서로 표현됩니다.

YOLO의 훈련 과정과 개선점은 다음과 같습니다:

1. \*\*사전 훈련\*\*: ImageNet 데이터셋으로 YOLO의 앞쪽 20개 컨볼루션 계층을 먼저 훈련시킵니다.

2. \*\*계층 추가\*\*: 사전 훈련된 계층 뒤에 4개의 컨볼루션 계층과 2개의 전결합 계층을 추가합니다.

3. \*\*활성화 함수\*\*: 마지막 계층에는 선형 활성화 함수를, 나머지 계층에는 leaky ReLU를 사용합니다.

4. \*\*개선점\*\*:

1. localization loss의 가중치를 증가시킵니다.

2. 객체가 있는 그리드 셀의 confidence loss 가중치를 높입니다.

3. bounding box의 너비와 높이에 제곱근을 적용한 값을 loss function으로 사용합니다.

5. \*\*과적합 방지\*\*: 드롭아웃과 데이터 증강을 사용합니다.

YOLO의 단점은 다음과 같습니다:

- 작은 객체들이 밀집해 있는 경우 검출 성능이 떨어집니다.

- 훈련 때 보지 못한 종횡비(aspect ratio)를 가진 객체를 테스트할 때 어려움을 겪습니다.

- 큰 bounding box와 작은 bounding box의 loss에 동일한 가중치를 부여합니다.