

You Only Look Once: Unified, Real-Time Object Detection

* Base YOLO model은 real-time에서 초당 45개의 프레임의 이미지들을 처리&네트워크의 작은 버전인 Fast YOLO는 다른 real-time detector의 mAP의 두 배를 달성하면서 초당 155개의 프레임들을 처리함

* 다른 SOTA 탐지 시스템에 비교하여 YOLO는 localization error를 더 만들지만 background에서 false positives를 예측할 확률이 적음

* YOLO는 object의 아주 일반적인 representation을 학습함

* YOLO는 natural image를 다른 도메인에 일반화할 때 DPM과 R-CNN을 포함한 다른 객체 탐지 기법보다 좋은 성능을 보임

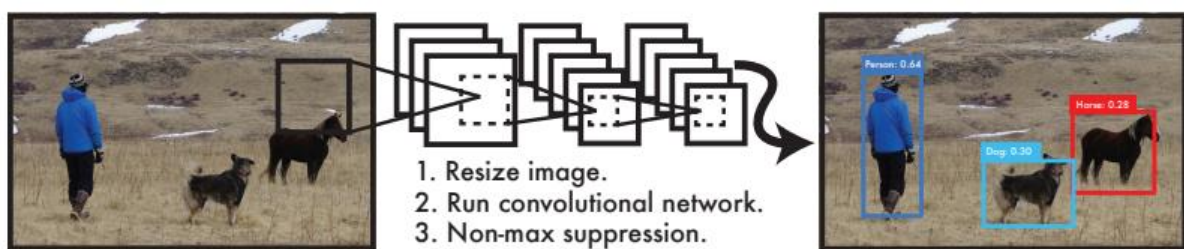
* R-CNN과 같은 최근 접근들은 region proposal method을 사용하여 이미지에서 잠재적인 bounding box를 생성하고 이 제안된 box들에서 classifier를 작동함. 분류 후에 post-processing은 bounding box를 정제하고, 중복된 것들을 제거하고, scene에서 다른 object들을 기반으로 box들을 rescore함

→ 이러한 복잡한 파이프라인은 느리고 최적화하기 힘들(각 구성 요소가 각각 훈련되어야 하기 때문에)

* 논문의 저자들은 객체 탐지를 single regression problem으로 생각함

→ YOLO를 이용하면 이미지를 단 한 번 만을 보고 이미지에 어떤 객체들이 존재하고 그것들이 무엇인지 예측함

*



Input image를 48 X 48로 resize, 이미지에서 single convolutional network를 처리하고, model's confidence에 따라 resulting detection의 기준점이 됨

→ single convolutional network이 multiple bounding boxes와 그러한 boxes들의 class probability들을 동시에 예측함

→ YOLO는 전체 이미지를 훈련하고 감지 성능을 직접적으로 최적화함

→ 이러한 통일된 모델은 기존 전통적인 객체 탐지 모델보다 장점들이 있음

① YOLO는 굉장히 빠름

-> 또한 YOLO는 다른 real-time 시스템보다 평균 정확성의 2배를 얻음

② 예측을 만드는 동안 YOLO는 이미지에 대해 전역적으로 판단

-> YOLO는 클래스들과 그것들의 모습에 대한 contextual information을 함시적으로 인코딩하기 위해서 훈련과 시험 시간 동안 전체 이미지를 봄

③ YOLO는 일반화 가능한 객체의 representation을 배움

* YOLO는 아직 SOTA 감지 시스템에 비해 정확도가 떨어짐

* 네트워크는 각 bounding box를 예측하기 위해 전체 이미지의 특성을 이용

* YOLO는 grid cell 당 여러 개의 bounding box를 예측한다

-> 훈련 시간동안 한 bounding box predictor가 각 object에 대해 책임지기 원함(어떤 예측이 가장 큰 현재 IOU를 가지고 있는지에 따라)

* YOLO는 각 grid cell들이 2개의 box들을 예측하고 오직 하나의 클래스만을 가지고 있기 때문에 bounding box 예측들에 강한 공간 제약을 부과함

-> 이 공간 상의 제약은 모델이 예측할 수 있는 근처 객체들의 수를 제한함

* Deformable parts models -> 모든 이질적인 부분들을 하나의 CNN으로 대체함

* R-CNN과의 유사한 점: 각 grid cell은 잠재적인 bounding box를 제안하고 이러한 box들을 convolutional 특성을 이용하여 점수를 매김

→ 그렇지만 YOLO는 공간의 제약을 grid cell proposal에 넣어 같은 객체에 대해 완화된 다양한 탐지를 도움

* 결론

- 객체 탐지를 위한 unified model

- 만들기 쉽고 전체 이미지에 대해 직접적으로 훈련할 수 있음

- classifier-based 접근들과 달리 YOLO는 detection 성능과 직접적으로 부합하는 loss function에 대해 훈련하고 전체 모델은 합동으로 훈련함

- Fast YOLO는 literature에서 가장 빠른 general-purpose object detector임

- YOLO는 새로운 도메인에 일반화가 잘 됨 -> 빠르고 튼튼한 객체 탐지에 의존하는 애플리케이션들에 이상적임