

그룹명 : 66기 비전팀

(1)차 주간보고서

활동 현황

작성자	정성실	장소	온라인
모임일자	2022년 09월 18일 일요일	모임시간	10:00 ~ 11:30 (총 90 분)
참석자	여윤기, 이근호, 정성실	결석자	

학습 내용

학습주제 및 목표

Object detection 모델 중 빠른 추론 속도에 비해 높은 정확도를 가진 YOLOv7 모델에 대한 기본적인 개념 및 성능 분석

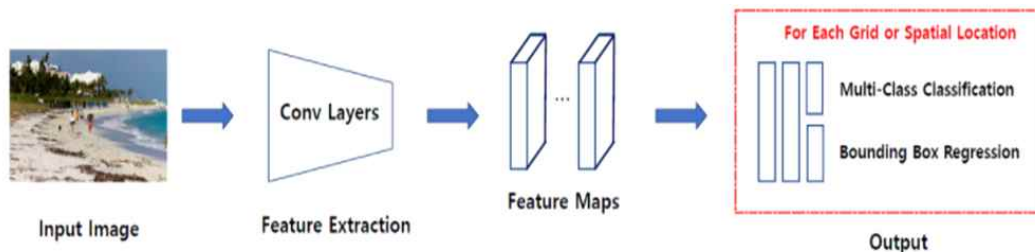
학습내용

◇ Object detection 이란?

- 1) 한 이미지에서 객체와 그 경계 상자(bounding box)를 탐지
- 2) 객체 탐지 알고리즘은 일반적으로 이미지를 입력으로 받고, 경계 상자와 객체 클래스 리스트를 출력
- 3) 경계 상자에 대해 그에 대응하는 예측 클래스와 클래스의 신뢰도(confidence)를 출력



- 여러 물체에 대해 어떤 물체인지, 어디에 위치해 있는지 탐지
- Classification + Localization
- 기존 전통적인 모델들은 2-Stage Detector로, Regional Proposal과 Classification이 순차적으로 진행



- Object가 있을 법한 위치의 후보들을 뽑아내고(Region Proposal), 이후 실제 Object를 Classification 수행 및 정확한 Bounding Box를 구하는 Regression 수행하는 단계가 분리되어 있음
-> R-CNN 계열 등

- 문제점 : 느리고, Region Proposal된 영역만을 이용하여 예측하기 때문에 Background Error 비율이 높은 등 최적화하기가 어려움

◇ Object detection 모델의 성능을 나타내는 지표란?

IOU(Intersection Over Union)

실측값(Ground Truth)과 모델이 예측한 값이 얼마나 겹치는지를 나타내는 지표
IOU가 높을수록 잘 예측한 모델

- 중복값 처리 용도

NMS(Non-Maximum Suppression, 비최대값 억제)

확률이 가장 높은 상자와 겹치는 상자들을 제거하는 과정
최대값을 갖지 않는 상자들을 제거

• 과정

- 1) 확률 기준으로 모든 상자를 정렬하고 먼저 가장 확률이 높은 상자를 취함
- 2) 각 상자에 대해 다른 모든 상자와의 IOU를 계산
- 3) 특정 임계값을 넘는 상자는 제거

정밀도(Precision)와 재현율(Recall)

- 일반적으로 객체 탐지 모델 평가에 사용되지는 않지만, 다른 지표를 계산하는 기본 지표 역할을 함

True Positives(TP) : 예측이 동일 클래스의 실제 상자와 일치하는지 측정

False Positives(FP) : 예측이 실제 상자와 일치하지 않는지 측정

False Negatives(FN) : 실제 분류값이 그와 일치하는 예측을 갖지 못하는지 측정

- 모델이 안정적이지 않은 특징을 기반으로 객체 존재를 예측하면 거짓긍정(FP)이 많아져서 정밀도가 낮아짐
- 모델이 너무 엄격해서 정확한 조건을 만족할 때만 객체가 탐지된 것으로 간주하면 거짓부정(FN)이 많아져서 재현율이 낮아짐

AP(Average Precision, 평균 정밀도)와 mAP(mean Average Precision)

- 곡선의 아래 영역에 해당

항상 1x1정사각형으로 구성되어 있음. 즉, 항상 0~1 사이의 값을 가짐

단일 클래스에 대한 모델 성능 정보를 제공

- 전역 점수를 얻기 위해서 mAP를 사용

예를 들어, 데이터셋이 10개의 클래스로 구성된다면 각 클래스에 대한 AP를 계산하고, 그 숫자들의 평균을 다시 구함

◇ YOLO란?

- 가장 빠른 객체 검출 알고리즘 중 하나
- 256x256 사이즈의 이미지
- 파이썬, 텐서플로 기반 프레임워크가 아닌 C++로 구현된 코드 기준 GPU사용 시, 초당 170 프레임 (170 FPS. frames per second)
- 작은 크기의 물체를 탐지하는데는 어려움
- Single Network 하나만을 사용하여 속도가 매우 빠름

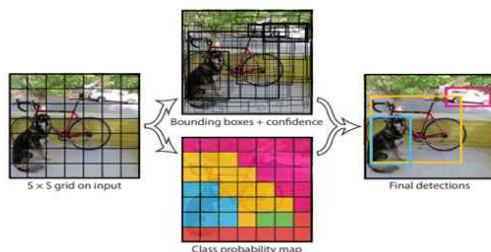
- Less false positives on background – 배경에 대해 물체가 있다고 인식하는 오류율이 적음
- 일반적인 표현을 잘 학습하므로, 새로운 Domain을 갖는 데이터에도 쉽게 적용 가능함.
- 단, 작은 크기의 물체를 탐지하는데는 어려움이 있다는 단점이 있음(50픽셀 이하)

YOLO 아키텍처

- 백본 모델(backbone model) 기반
- 특징 추출기(Feature Extractor)라고도 불림
- YOLO는 자체 맞춤 아키텍처 사용

◇ YOLOv1의 특징

1. Resizes Image
 - 입력 이미지는 S*S grid로 분할
2. 객체의 중심이 Grid cell에 맞으면, 해당 Grid cell은 객체를 탐지했다고 표현
3. 각 Grid cell은 Bounding box와, 각 Bounding Box에 대한 Confidence Score 예측
4. 각 Grid cell은 Bounding box 이외에도 Class 확률을 예측
5. Non-max suppression을 거쳐 최종 Bounding box 선정
 - * 각각의 bounding box의 IoU 값 비교를 통해 여러 개의 bounding box 중 하나만 선택되게 함



$$\Pr(\text{Class}_i | \text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}} \quad (1)$$

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


The diagram shows two overlapping rectangles. The top rectangle is light blue and the bottom rectangle is dark blue. The intersection of the two rectangles is shaded in a darker blue, representing the 'Area of Overlap'. The union of the two rectangles is the total area covered by both, representing the 'Area of Union'.

◇ YOLOv7의 특징 (논문에서 제시하는 다른 Object detection과 차별화 된 부분)

논문 - YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors

<https://arxiv.org/abs/2207.02696>

깃허브 - <https://github.com/wongkinyiu/yolov7>

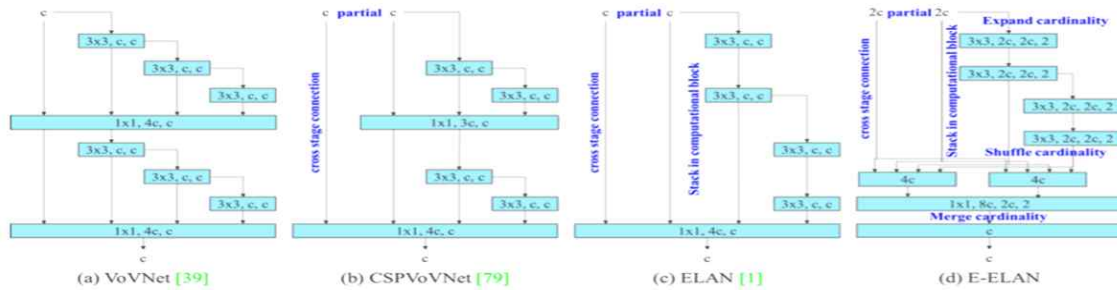


Figure 2: Extended efficient layer aggregation networks. The proposed extended ELAN (E-ELAN) does not change the gradient transmission path of the original architecture at all, but use group convolution to increase the cardinality of the added features, and combine the features of different groups in a shuffle and merge cardinality manner. This way of operation can enhance the features learned by different feature maps and improve the use of parameters and calculations.

- 1) YOLO v7은 5~160 FPS 범위의 속도와 정확도 측면에서 현재까지 나온 모든 Object Detector의 성능을 능가
- 2) V100 GPU에서 30 FPS 이상의 현재까지 알려진 Real-time Detector 중 가장 높은 정확도인 56.8% AP를 달성
- 3) YOLOv7-E6는 Transformer 기반 검출기인 SWIN-L Cascade-Mask R-CNN 보다 속도는 509%, 정확도 2%를 증가하며, Convolution 기반 검출기인 ConvNeXt-XL Cascade-Mask R-CNN 보다 속도는 551%, 정확도는 0.7% 앞섬

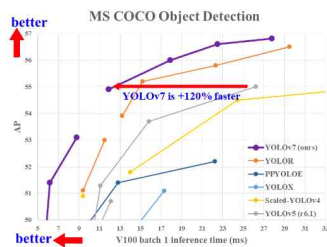
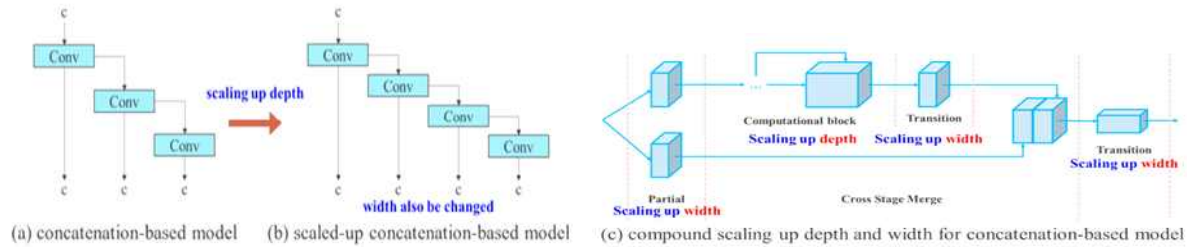


Figure 1: Comparison with other real-time object detectors, our proposed methods achieve state-of-the-arts performance.

그림 1. MS COCO 데이터셋에 대하여 다른 모델과 비교한 YOLOv7의 객체 탐지 결과를 나타낸 것.
MS COCO(Microsoft Common Objects in Context)는 대규모의 객체 탐지, 세그멘테이션 및 캡션 데이터셋을 지칭

◇ YOLOv7의 Architecture

- 모델학습 시 디바이스의 memory cost나 계산량을 줄이기 위해 어떤 네트워크 방식이 가장 효율적인가?
-> 본 논문에서는 YOLO v7의 baseline 구조인 E-ELAN 제안
- 기존에 제시된 Net 구조들과 비교하여 computainal block을 마음껏 쌓아도 학습이 잘 되도록 하기 위해서 E-ELAN 제안
- Expand -> Shuffle -> Merge 과정을 통해 original gradient path의 파괴 없이 네트워크의 학습능력을 지속적으로 향상(computational block을 계속 쌓아도 안정적으로 학습 정확도 향상 시킬 수 있는 효과가 있음)
 1. group conv g 값을 파라미터로 사용하여 computational block의 채널수를 확장
 2. 각 computational block에서 계산된 feature map은 g값에 따라 그룹으로 섞인 다음 함께 연결됨
 3. g개의 feature map group을 서로 합쳐서 Merge 과정 적용



- Concatenation-based 의 Model Scaling 기법 제안
- 모델 Scaling의 목적 -> 높은 정확도, Inference time 감소 등 보다 높은 효율을 추구하고자 모델의 일부 attribute(모델의 width, depth, resolution 등)를 조정하며 다양한 scale의 모델을 생성
- 앞서 제안한 E-ELAN 네트워크 방식으로 모델을 학습시킨다면, scaling 비율 조정 시 Transition Layer도 변동이 생기게 됨
 - > transition layer의 input/output channel이 모두 변경되고 output channel이 변경되면 그 다음 layer의 Input channel 또한 변경되어, 결국 scaling factor에 대한 분석이 불가능하게 됨.
- 이 문제점을 해결하기 위해 YOLO v7은 Compound model scaling method 제안
 1. Computation Block에 depth scaling up 적용 -> Output channel 변경
 2. Output channel 변경 -> Transition layer의 width scaling up
 3. partial 부분의 input feature map은 transition layer의 output feature 이므로 똑같이 width scaling up 시켜준다.

◇ YOLOv7의 성능

1. 실시간 객체 검출기(real-time object detection)이면서 추론비용(inference cost)을 증가시키지 않고도 정확도를 향상시킬 수 있는 trainable bag-of-freebies 방법을 제안
2. re-parameterized module이 original module을 대체하는 방법과 different output layers에 대한 dynamic label assignment strategy 방법 + 여기서 발생하는 어려움을 해결하기 위한 방법 제안
3. 매개변수(parameter)와 계산을 효과적으로 활용할 수 있는 실시간 객체 검출기(real time object detector)를 위한 "확장(extend)" 및 "복합 스케일링(compound scaling)" 제안
4. 본 논문에서 제안한 방법은 SOTA(State-of-the-art) 방법 보다 parameter 수를 40%, 계산량을 50% 감소시킬 수 있으며 더 빠른 inference time과 더 높은 accuracy를 달성

그룹 운영 기록사항

-

다음 모임 계획

모임일자	2022년 09월25일 일요일	모임시간	20:00~21:00 (총 60 분)
역할분담	개발 환경 셋팅	장 소	온라인