

Term project proposal

8조

나건주, 강효준, 곽제우

주제 : YOLO v5를 이용한 도로 상황 인식

요약 : 현재 자율주행시스템에 관한 많은 연구가 진행되고 있고, 특히 자율주행 자동차, 드론, 서비스 로봇과 같은 실생활의 변화를 가져올 수 있는 분야에 대한 많은 투자 및 연구가 진행되고 있다고 한다. 이에 따라 자율주행에 관련된 연구도 계속해서 진행 중인데, 높은 정확도로 자율주행을 하기 위한 핵심 기술은 영상처리 및 인식기술이라고 볼 수 있다. 자율주행을 위해 객체를 탐지하는 기술은 크게 2가지로 나뉜다. Lidar 센서를 이용하여 객체의 위치와 깊이를 탐지하는 기술과 카메라를 이용하여 딥러닝으로 물체를 탐지하는 기술이다. 이번 프로젝트를 통해서 딥러닝 기술을 활용해 카메라만으로도 객체탐지를 정확하게 잘 수행하는지 알아보고자 진행하게 되었다. 본 8조에서는 자동차, 사람, 자전거, 표지판, 신호등, 나무 등을 클래스로 구분하고 각 객체의 사진을 YOLO v5를 이용해 학습시킨 후 도로주행 영상에서 지정한 객체가 실시간으로 검출되는지 확인할 계획이다.

데이터 :

Aihub에서 제공하는 도로주행 영상 데이터를 사용한다. YOLOv5 모델은 bbox를 탐지하는 모델이기 때문에 해당 데이터에서 bbox로 제공하는 20만 frame의 이미지 + XML 데이터 셋을 이용한다. 데이터의 라벨은 총 33개의 Label Map이 있는데 자동차, 보행자, 신호등, 표지판, 노면 표시, 노면 화살표 등을 Label로 제공한다. 해당 데이터에서는 bbox가 (xmin, ymin, xmax, ymax)로 제공되며 yolo의 경우 (x_center, y_center, w, h) 형식을 따르므로 변경해야 한다.

이미지는 1920 * 1080의 데이터가 주어지고, 모델은 640 or 1280을 선호하며, 작은 개체가 많을수록 더 높은 해상도로 훈련하는 것이 선호되기 때문에 1280 크기로 학습할 계획이다.

학습 데이터가 매우 크기 때문에 GCP 혹은 Colab Pro를 이용하여 가장 큰 데이터 셋인 YOLOv5x -P6를 통해 학습을 진행할 계획이다.

후에 결과가 부족하다면 Hyper Parameter를 수정하거나 Background image를 넣어 다시 시도할 계획이다.

기존 연구 사례 조사 :

1. Two stage detector - Fast R-CNN

Two stage detector에서는 위치를 찾는 Region Proposal, 분류하는 Region Classification으로 나누어 detection을 실행한다. 대표적인 Two stage detector로 Fast R-CNN이 있다. 전체적인 프로세스는 다음과 같다.

1-1. R-CNN에서와 마찬가지로 Selective Search를 통해 RoI를 찾는다.

1-2. 전체 이미지를 CNN에 통과시켜 feature map을 추출한다.

2. Selective Search로 찾았던 RoI를 feature map크기에 맞춰서 projection시킨다.

3. projection 시킨 RoI에 대해 RoI Pooling을 진행하여 고정된 크기의 feature vector를 얻는다.

4. feature vector는 FC layer를 통과한 뒤, 구 브랜치로 나뉘게 된다.

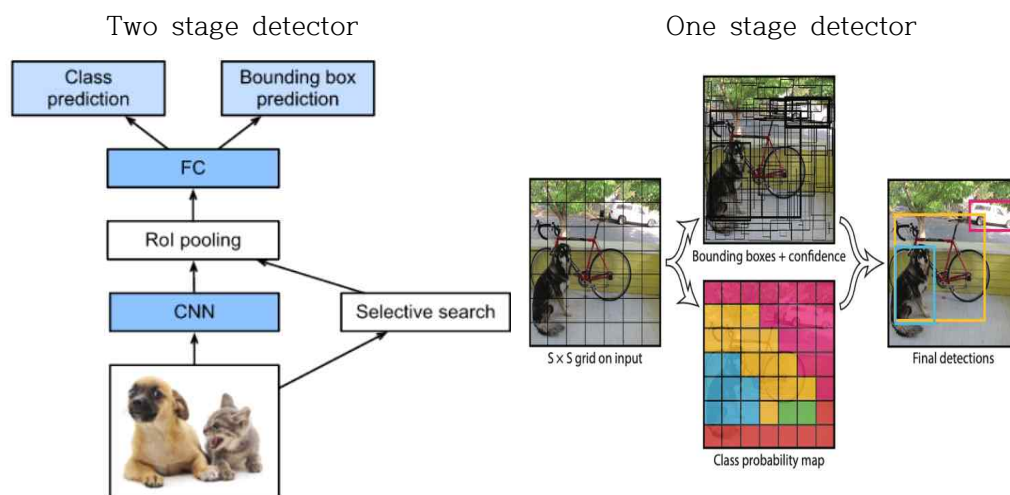
5-1. 하나는 softmax를 통과하여 RoI에 대해 object classification을 한다.

5-2. bounding box regression을 통해 selective search로 찾은 box의 위치를 조정한다.

2. One stage Detector -YOLO

기존 two stage detector의 경우, 객체를 검출하는 정확도 측면에선 굉장히 좋은 성능을 냈지만, 속도(FPS) 측면에서는 느리다. 속도 문제를 해결하기 위해 grid 개념을 도입하여 region proposal과 classification을 동시에 하는 one stage detector가 제안되었다. One stage detector의 대표적인 모델로 YOLO 모델이 있다. 전체적인 프로세스는 다음과 같다.

- 입력 이미지를 S x S grid로 나눈다.
- 만약에 객체의 중심이 grid cell에 들어가면, 해당 grid cell은 그 객체를 탐지하는 역할
- 각 grid cell은 B개의 bbox를 예측한다.
- 해당 bbox에 대한 confidence score(해당 bbox안에 물체가 있는지 없는지 확률값)를 예측한다.
- bbox : x, y, width, height, confidence → 5가지 예측한다.



- YOLO v5 small과 Faster R-CNN 성능 비교

아래 표는 YOLO v5 small과 Faster R-CNN ResNet 50을 5가지 항목으로 나누어 비교한 표이다. 평가에 사용한 영상은 도로주행 영상, 농구 경기 영상, 사람들이 봄비는 지하철역 영상으로 3개의 비디오이고, 품질과 속도에 대해서 평가를 시행하였다.

	YOLO v5	Faster RCNN
Inference Speed	✓	
Detection of small or far away objects	✓	
Little to no overlapping boxes	✓	
Missed Objects	✗	✗
Detection of Crowded objects	✓	✓

- YOLO v5의 실행 속도는 52.8 FPS이고, Faster R-CNN 모델은 21.7 FPS로 YOLO v5 모델이 두 배 이상 빨랐다.

- 신호등과 멀리 떨어진 차량과 같이 작은 물체는 YOLO v5 모델에서만 검출됐다.

- Faster R-CNN 모델의 경우에는 바운딩 박스 겹침이 심하게 나타났다. 반면 YOLO v5 모델은 바운딩 박스 겹침이 거의 나타나지 않았다.

- 두 모델 모두 멀리 있는 사람들을 검출하지 못했지만, 사람들이 가까이 오면 겹쳐있는 사람들까지 인식하였다.

One stage detector 대표 모델 YOLO V5와 Two stage detector 대표 모델 Faster R-CNN을 비교한 결과 YOLO 모델이 더 높은 성능을 보였다. 따라서 본 8조는 YOLO V5 모델을 사용하여 물체 인식을 수행할 계획이다.

- 모델 성능평가

1) FPS

FPS는 Frame Per Second의 약자이다. 즉 1초당 몇 이미지 Frame이 우리에게 보이는지 이다. 실시간으로 객체를 검출하기 위해서는 통상 30 fps를 넘겨야 한다.

2) IoU

IoU는 Intersection over unit으로 예측한 바운딩 박스가 정답 바운딩 박스의 교집합 영역의 넓이/ 합집합 영역의 넓이를 계산해 나온 값이다. IoU는 모델이 예측한 바운딩 박스가 정답 바운딩 박스와 겹치는 부분이 많을수록 높은 reward를 준다. 그리고 평가된 IoU 값을 기준으로 물체가 옳게 검출되면 precision과 recall을 계산한다.

3) mAP

mAP는 Mean average precision으로 mAP가 높을수록 정확하다. mAP는 PR곡선에서 아래 면적에 해당하는 AP를 평균값을 계산하여 구한다. multi_object detection을 수행해야 하므로 하나의 클래스가 아닌 모든 클래스의 AP 평균을 구하는 mAP를 사용한다.

목표 결과 : 32종류의 객체를 실시간으로 인식하고 각 객체를 정확히 분류한다.

	목표치
인식 속도(FPS)	30 fps
인식 정확도(mAP)	80%