속도 및 정확도 비교 (Faster R-CNN, R-FCN, SSD, FPN , RetinaNet, YOLOv3)

성능에 영향을 미치는 다른 사항 (알고리즘의 성능이아니라 트레이닝 진행시 영향을 끼치는 것들)

-기능추출기(VGG16, ResNet, Inception, MobileNet).

- 추출기에 대한 출력 보폭

- 입력 이미지 해상도

- 매칭 전략 및 IoU 임계값(1) (손실계산에서 예측이 제외되는 방법)

- 최대 억제 IoU 임계값

- 어려운 마이닝 비율( 양수대 음수 앵커 비율)

- 제안되는수 와 예측의 수

- 바운딩박스 인코딩

- 데이터 확대

- 트레이닝 데이터 셋

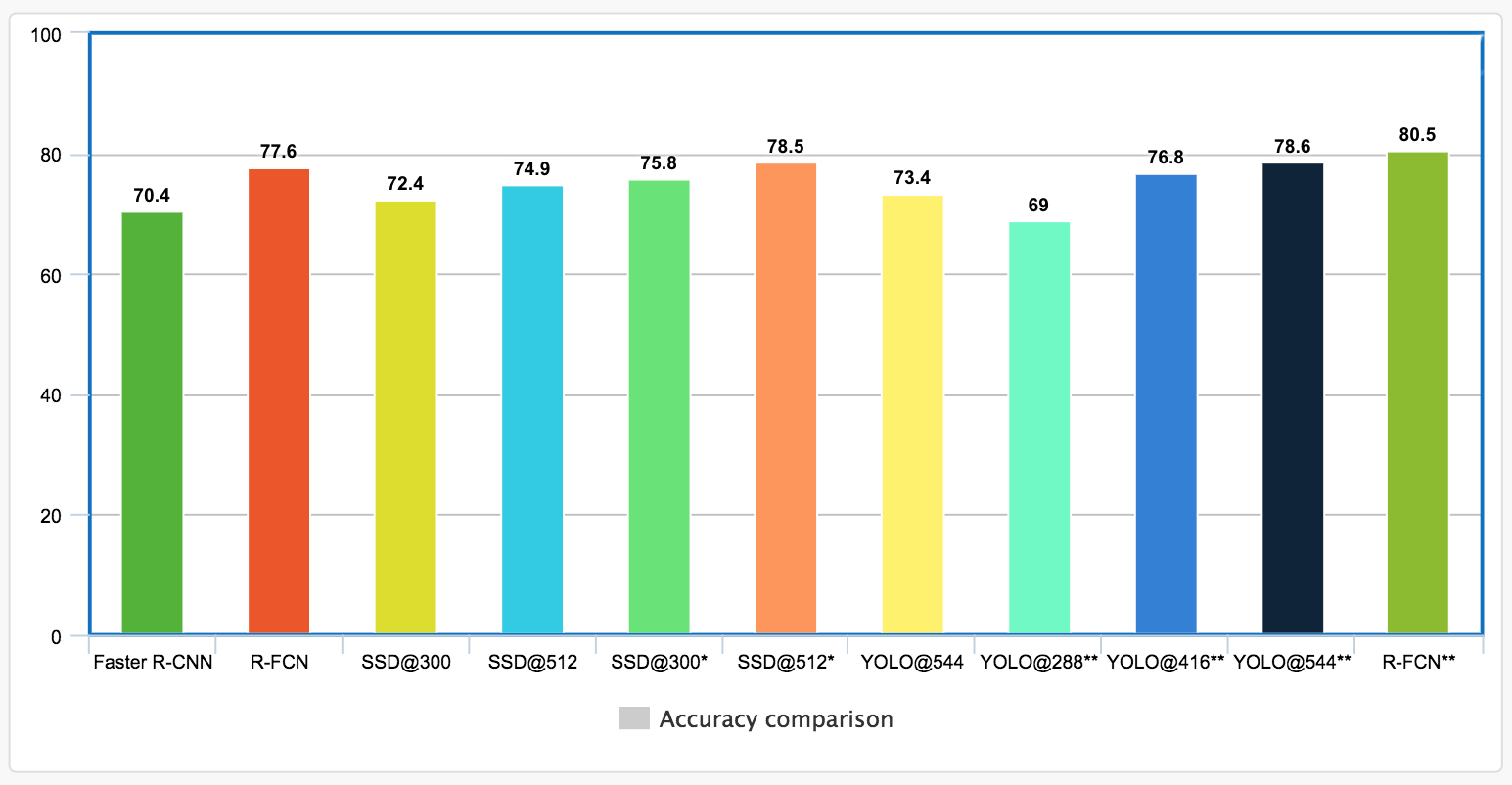
- 교육 또는 테스트에 멀티스케일이미지를 사용하는 것

- 객체 감지를 위한 기능 맵 레이어

- 딥러닝 소프트웨어를 쓰는 플랫폼 환경

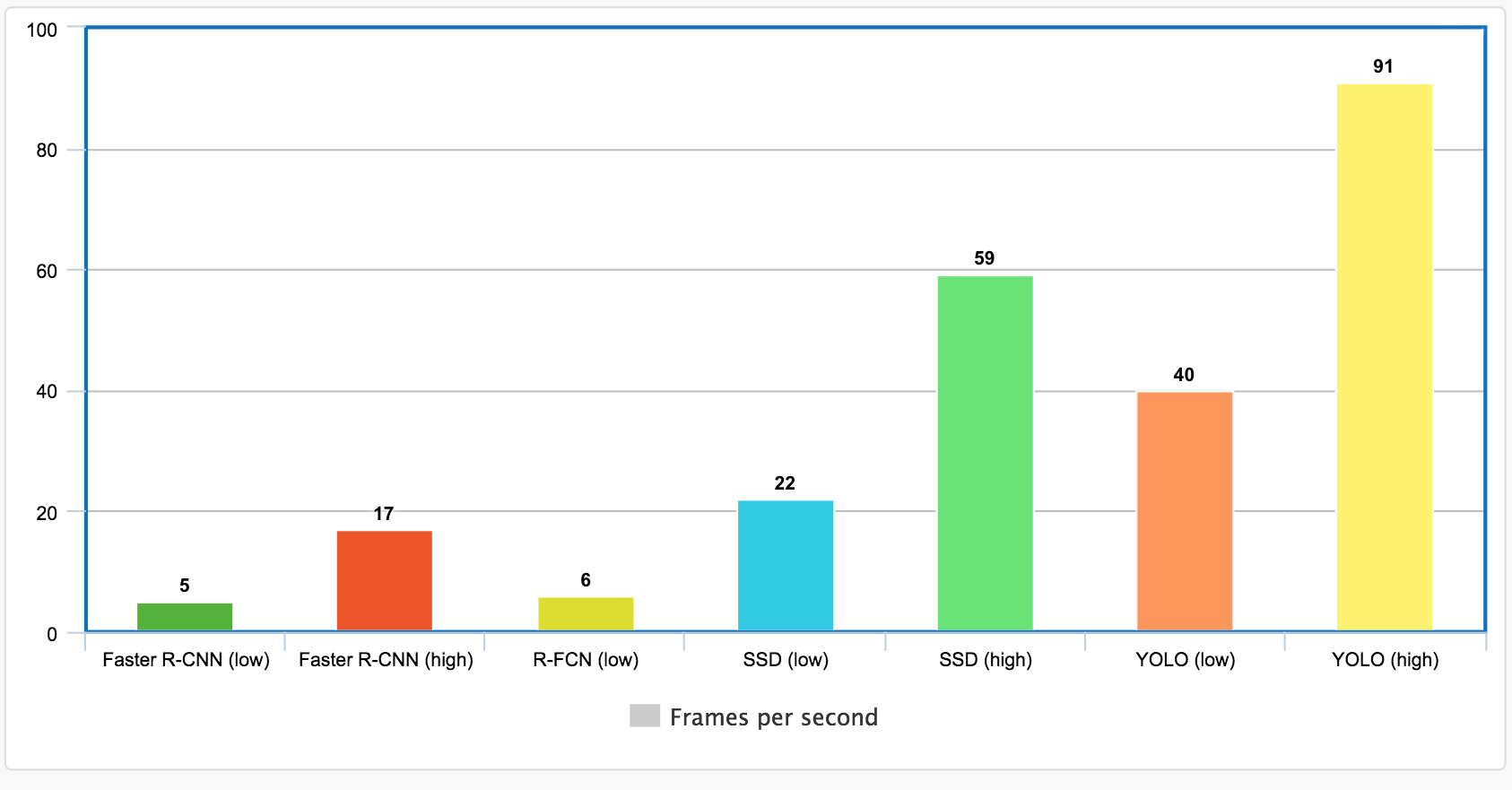
- 배치크기, 입력이미지 크기조정, 학습속도 및 학습속도 감소를 포함한 교육 구성

출처 : <https://medium.com/@jonathan_hui/object-detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359>



정확도의 경우 69~ 80까지 거의 70대에서 머문다고 할 수 있다. PASCAL VOC 2012 데이터를 훈련한 결과.

\*\*PASCAL VOC데이터는 비행기, 자전거, 사람, 새, 동물, 보트, 병, 버스, 자동차 등 우리 사회에서 흔히 볼 수 있는 여러가지의 객제들의 이미지 데이터



FPS는 이미지의 객체들을 분류 처리를 할 수 있는 초당 프레임 수

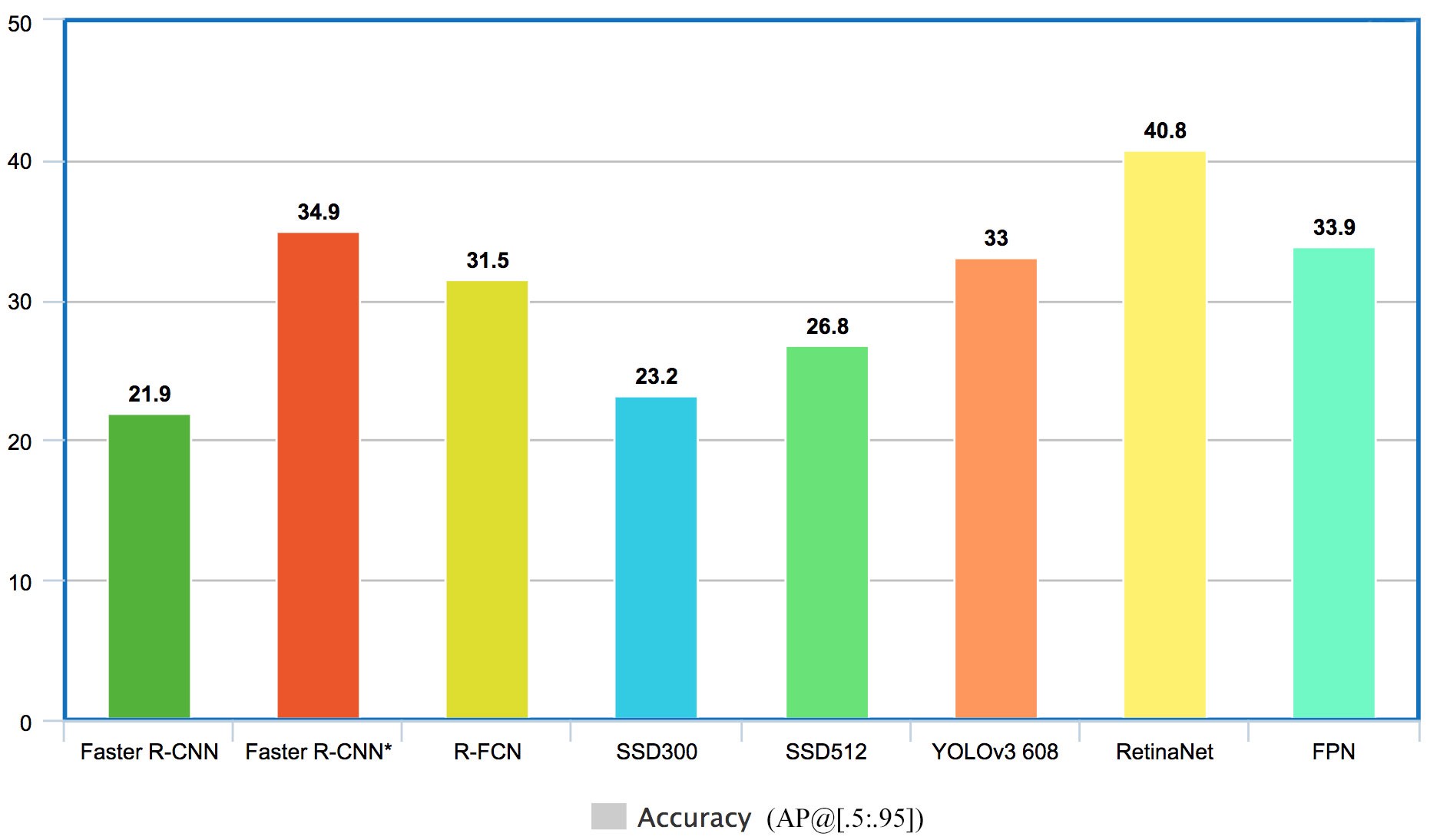
즉, YOLO가 가장 빠른 성능을 보여 주고 있음 가장 느린 Fater R-CNN과 정확도 차이는 -1%~+3%

그러나 FPS차이는 80프레임 이상의 차이를 보여준다.

(참고, 표.Titan X GPU 환경에서 트레이닝 시켜 얻은 결과)

\*\* mAP는 처리하여 객체가 있다고 판단해서 그려주는 바운딩 박스별 밀집도

|  |  |  |  |
| --- | --- | --- | --- |
| 종류 | mAp | Fps (Titan X GPU) | 우선순위 |
| YOLO V2 | 76.8 | 67 Fps | 1 |
| SSD 500 | 75.5 | 58 Fps | 2 |
| SSD 300 | 72.1 | 58 Fps | 3 |
| Faster R-CNN | 73.2 | 7 Fps | 4 |
| Fast R-CNN | 68.4 | - | 5 |



COCO dataset을 훈련시켜 얻은 알고리즘 별 객체 분류 정확도

\*\*COCO dataset은 PASCAL과 비슷하게 일반적으로 볼 수 있는 객체들의 이미지이지만 일반적으로 물체 감지가 어렵고 감지기들의 정확도가 훨씬 낮게 측정이 된다.,

물체 감지가 어려운 COCO dataset에서도YOLO는 알고리즘들 사이에서 중간정도의 성능을 보여주고 있다.

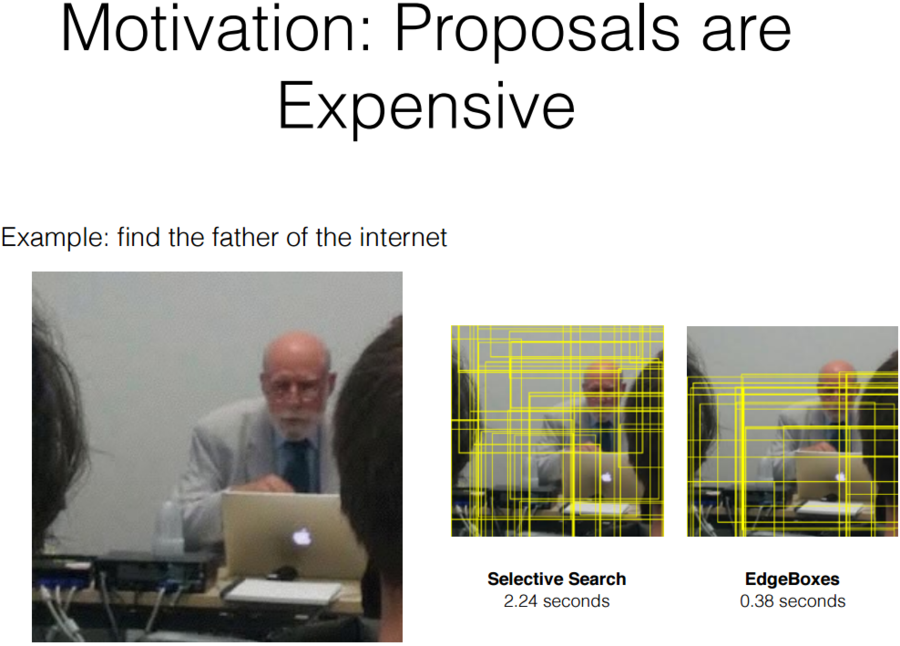
Faster R-CNN이나 SSD R-FCN의 경우 정확도가 좋다고 회자되지만 훈련시키는 이미지 1장에 1초정도 걸립니다.

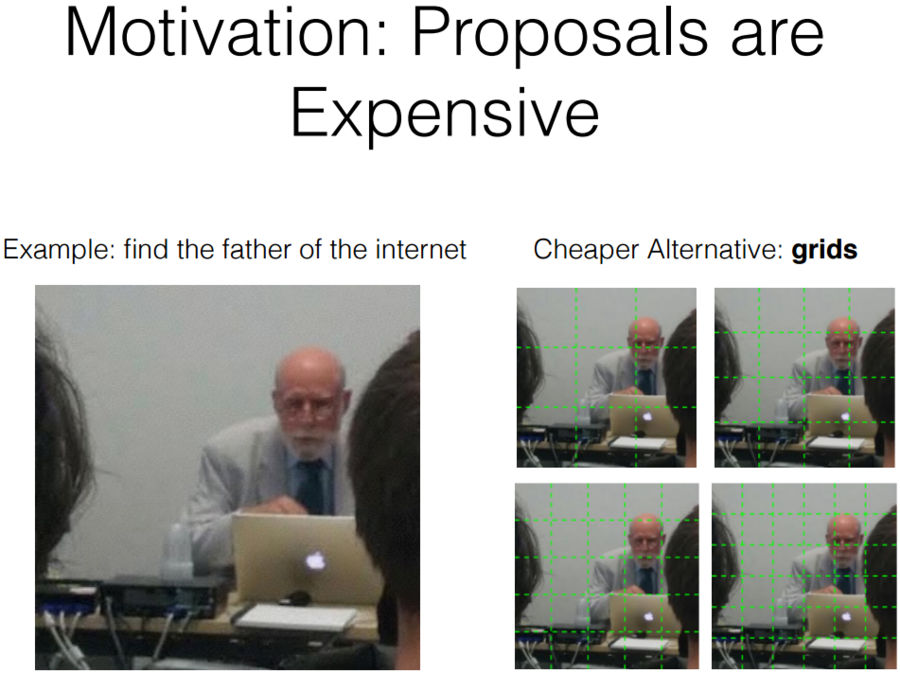
만약 1만개의 부품이 부품당 8장의 사진을 가지고 있다면 훈련시키는 시간이 24시간이 걸리므로 현실적으로 한달간의 프로젝트로는 훈련시키다가 기간이 끝날 수가 있음.

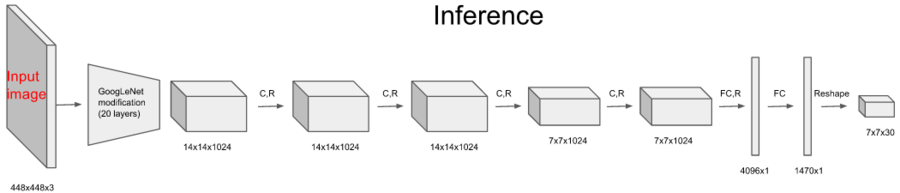
또한 1초가 걸리는 이미지의 해상도가 낮은편이므로 해상도가 높거나 크기가 커지면 속도가 더 느려진다.

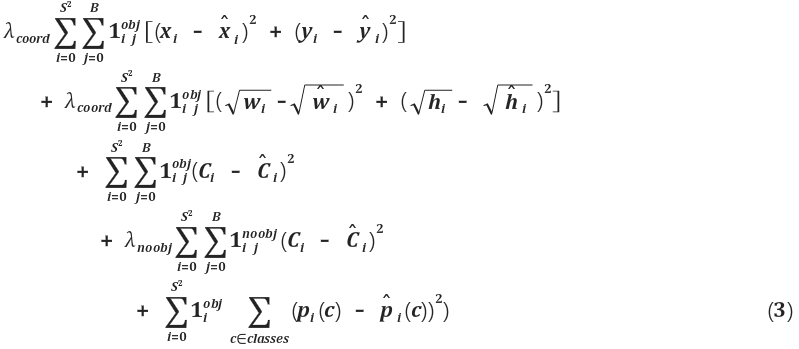
그러므로 일정 수준의 성능을 보여주며 UI제작을 통해 일종의 프로토타입을 만들어 내는 것이 목표이므로

YOLO알고리즘을 선택했다.











**한계**

YOLO는 영상을 7x7 의 그리드셀로 분할하여 각 그리드 셀을 중심으로 하는 각종 크기의 오브젝트에 대해서 경계박스 후보를 2개 예측한다.

R-CNN 계열은 후보를 1천개 이상 제안하는것에 비해 YOLO는 총 7x7x2 = 98개의 후보를 제안하므로 이로 인해 성능이 떨어진다.

그래서 한 오브젝트 주변에 여러개의 오브젝트가 있을 때 검출을 잘 못한다.

예를 들면 새떼처럼 조그만 오브젝트들이 모여 있는 경우이다.

영상에서 작게 나타난 새 크기에 비해 그리드 셀은 상대적으로 너무 크다.

같은 이유로 셀 하나 안에 오브젝트가 여러개 있으면 최대 2개까지밖에 예측을 못한다.