

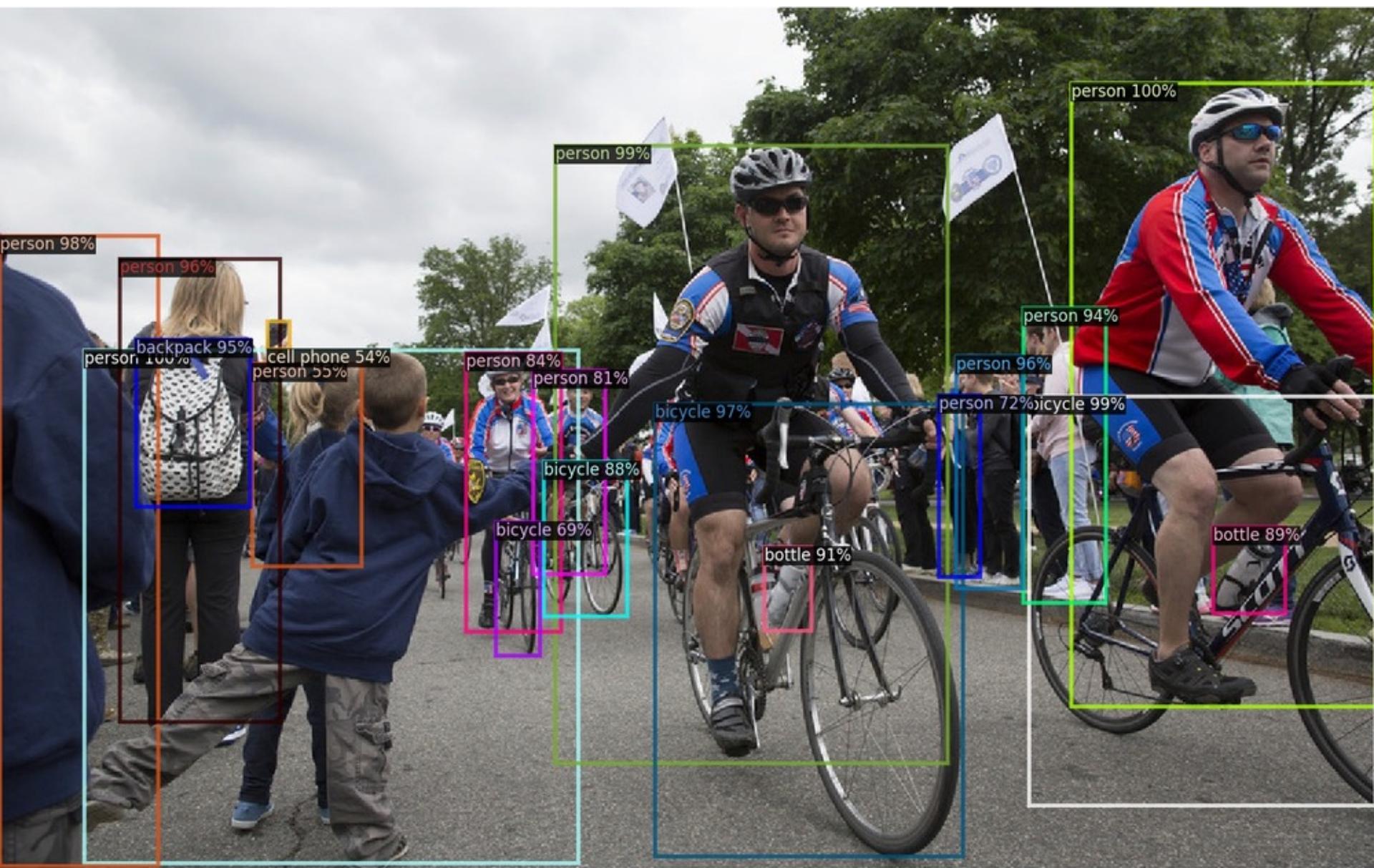
Object Detection

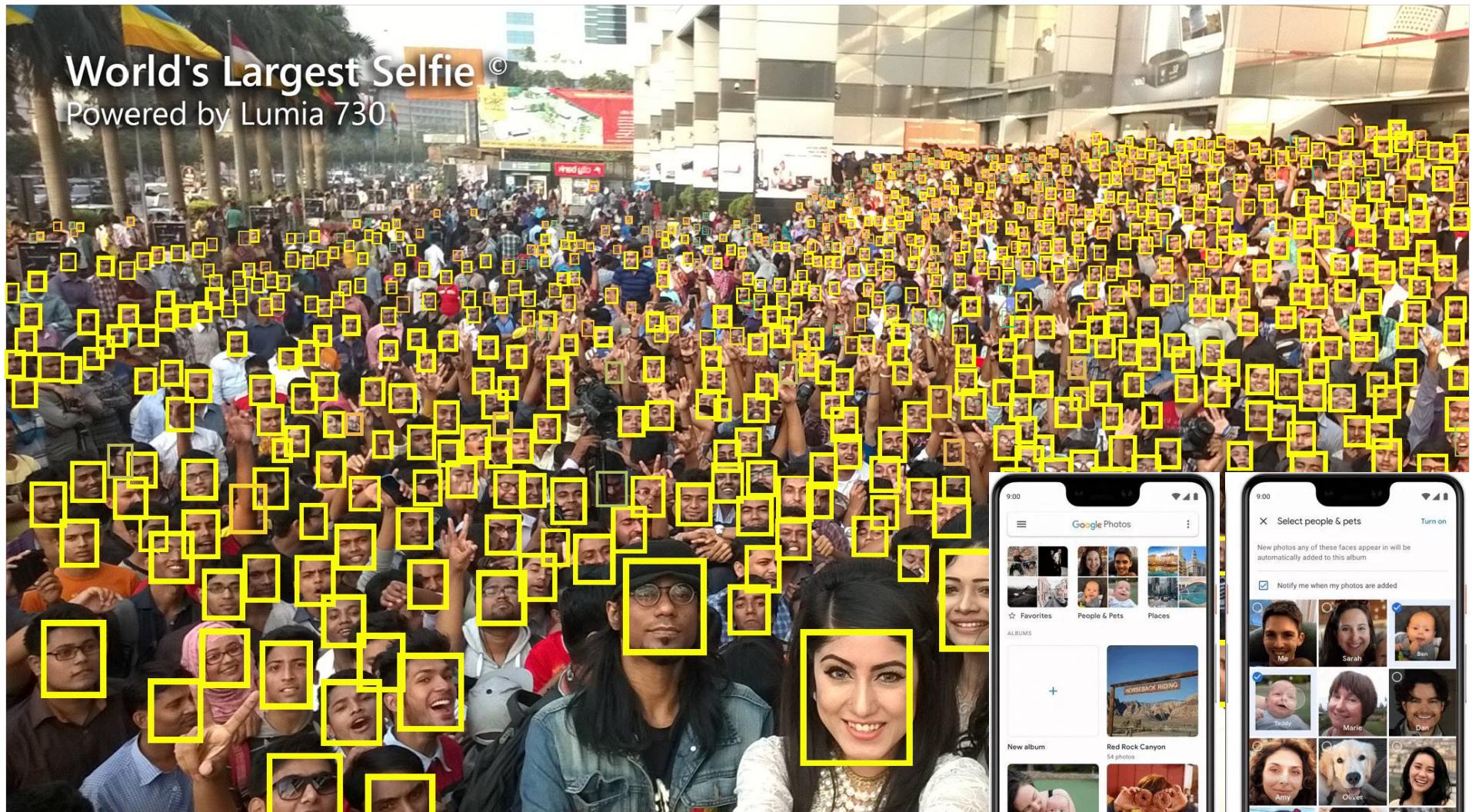
KCCV 2022 Tutorial
2022.02.24

김현우 (MLV Lab)
고려대학교 기계학습 및 비전 연구실
(hyunwoojkim@korea.ac.kr)

인공지능은 물체를 어떻게 이해할까?

- 소주제 1: 물체 탐지 원리
- 소주제 2: 물체 탐지 기법
- 소주제 3: 물체간 관계 추론





Hu, Peiyun, and Deva Ramanan. "Finding tiny faces." CVPR 2017.

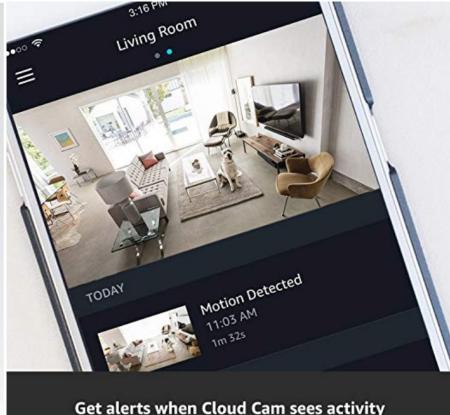
Smart Home & Fashion

amazon cloud cam

1080p HD | Night Vision | Two-Way Audio



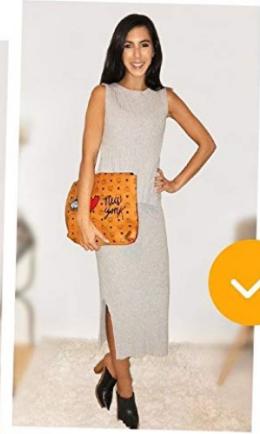
Roll over image to zoom in

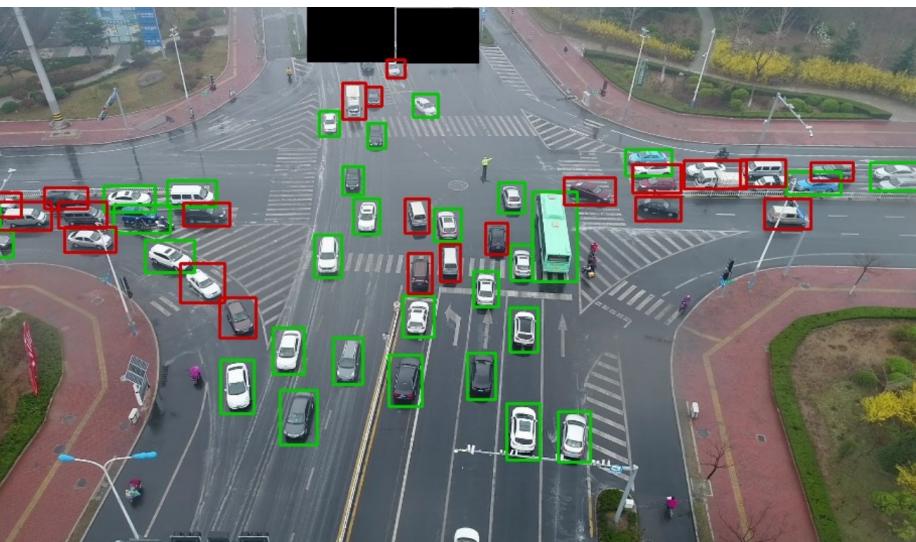
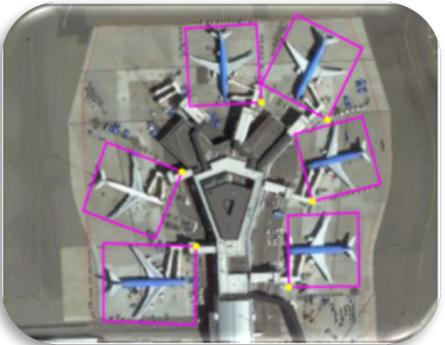


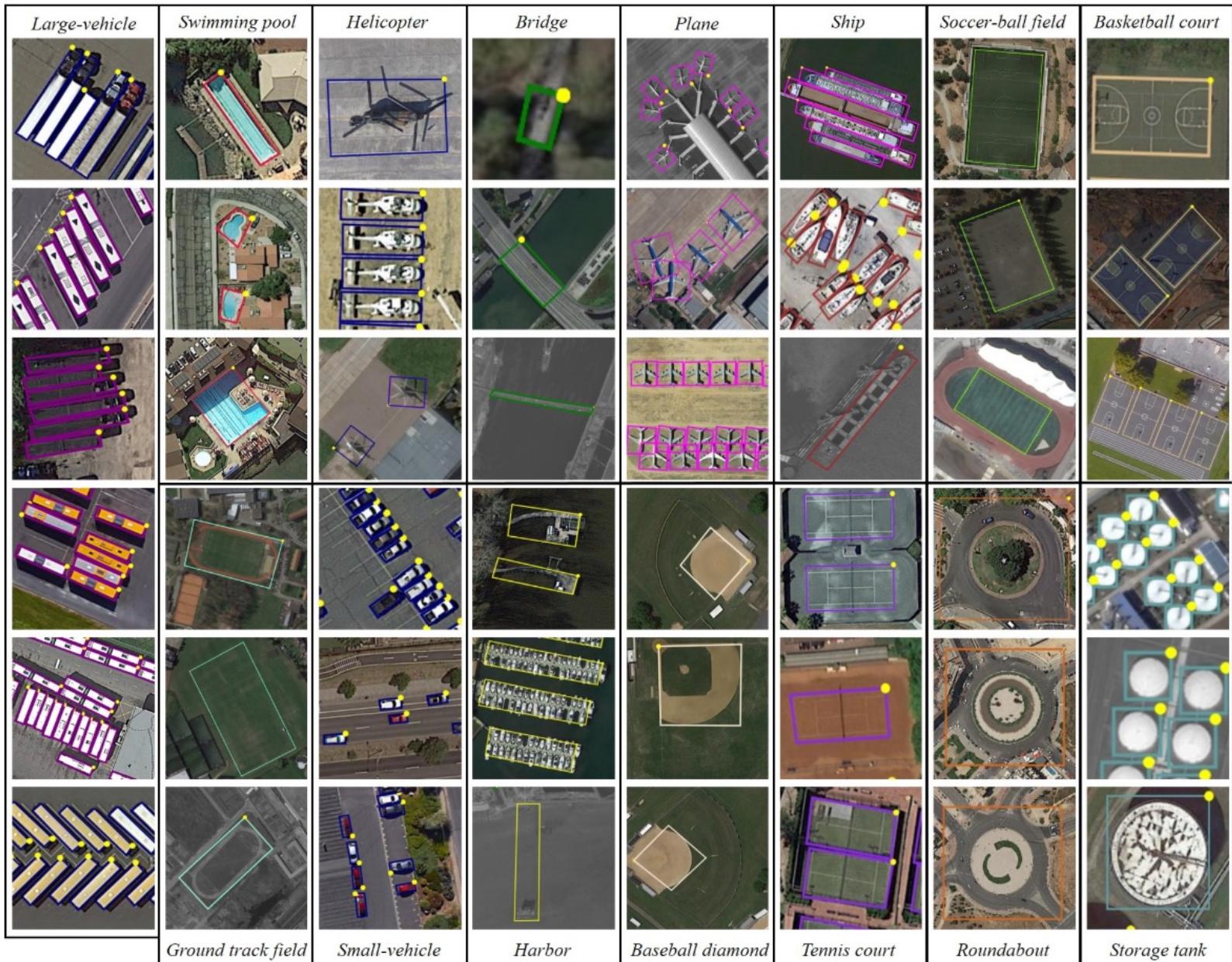
Roll over image to zoom in

echo look

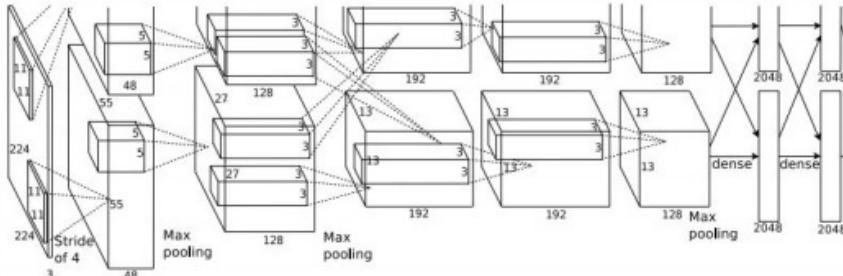
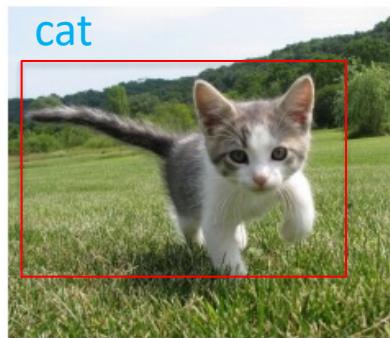
Love your look. Every day







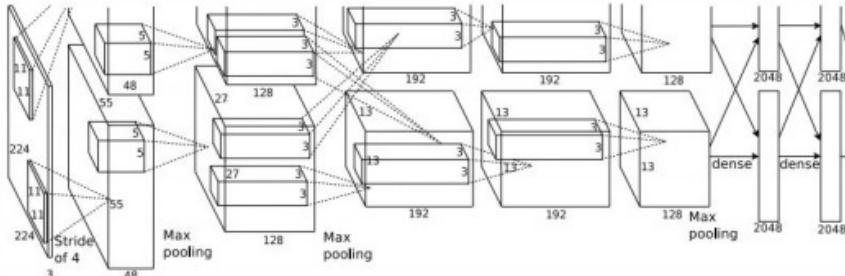
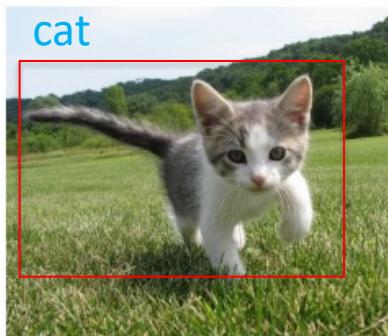
물체 탐지=위치 추정 + 분류



위치추정: localization
(x,y,w,h)

분류: classification
고양이

물체 탐지=위치 추정 + 분류



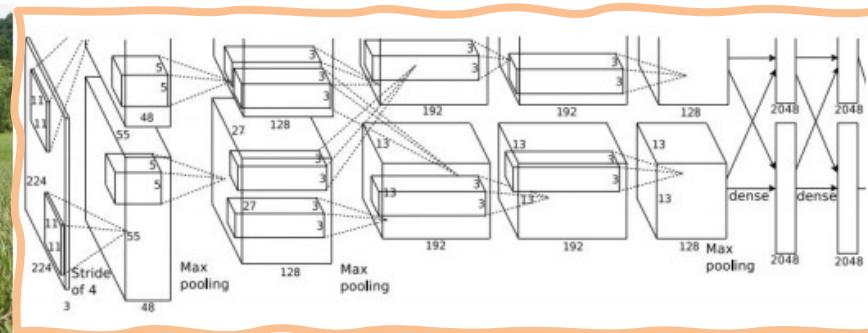
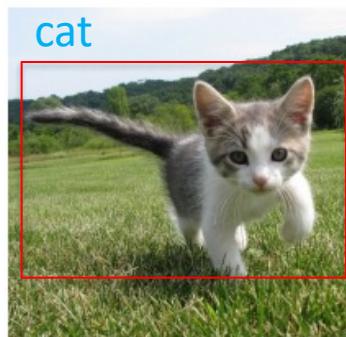
다중 태스크 학습
다중 손실 함수

위치추정: localization
(x, y, w, h)
L2 손실함수

+

분류: classification
고양이
softmax 손실함수

물체 탐지: 사전학습 (Pre-training)



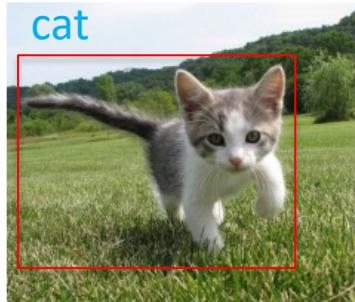
기저 네트워크 (Base Network): 특성값 추출
사전 학습 (Pre-training): ImageNet

위치추정: localization
(x, y, w, h)
L2 손실함수

+

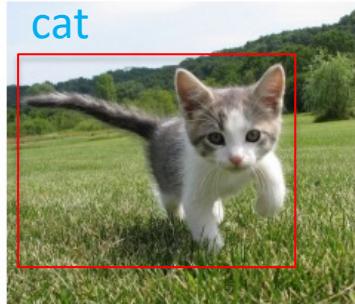
분류: classification
고양이
softmax 손실함수

물체 탐지: 다수 물체 탐지 ?



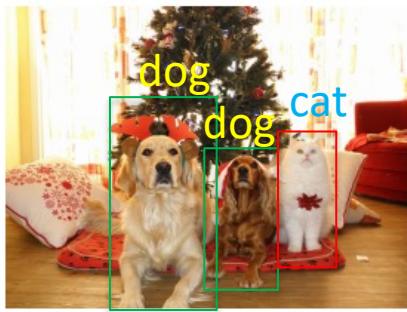
Cat: (x, y, w, h)

물체 탐지: 다수 물체 탐지 ?



Cat: (x, y, w, h)

20 클래스 + 4개 위치정보 = 24 숫자

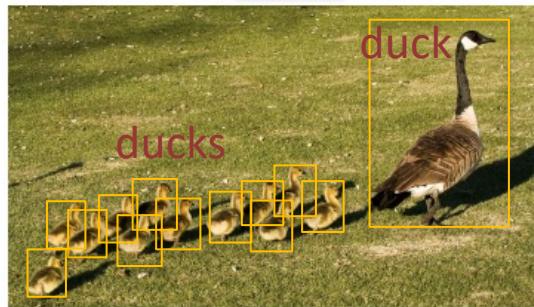


Dog: (x, y, w, h)

Dog: (x, y, w, h)

Cat : (x, y, w, h)

24 숫자 x 3



Duck : (x, y, w, h)

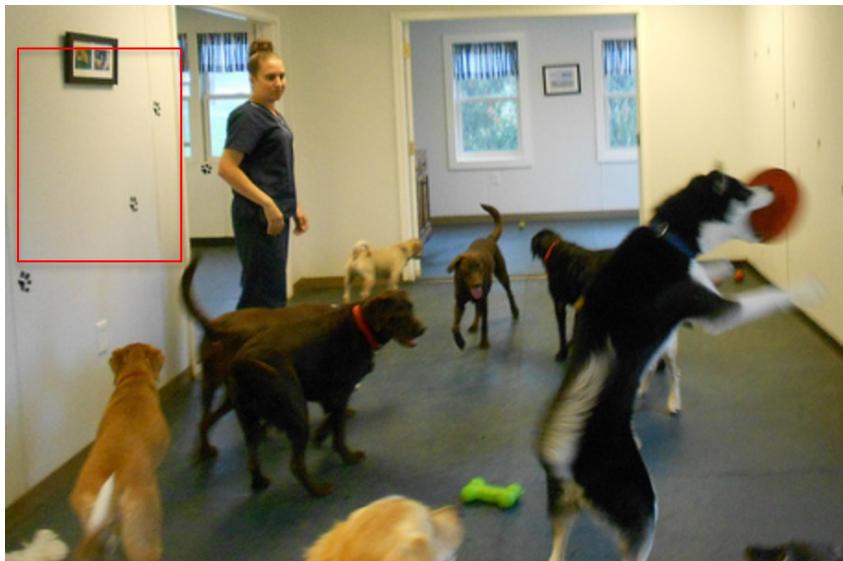
Duck : (x, y, w, h)

Duck : (x, y, w, h)

...

24 숫자 x ?

물체 탐지: 다수 물체 탐지 ?



Background : Yes

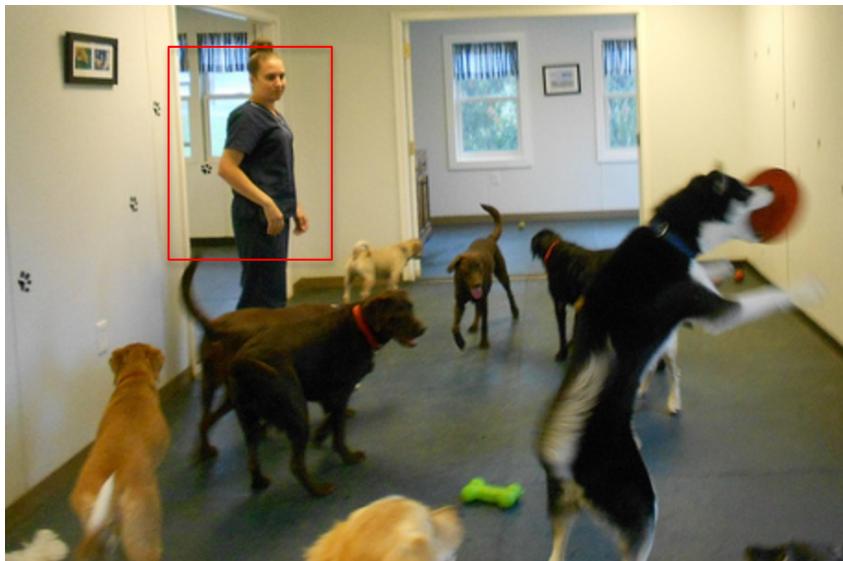
Dog : No

Frisbee : No

Person : No

...

물체 탐지: 다수 물체 탐지 ?



Background : No

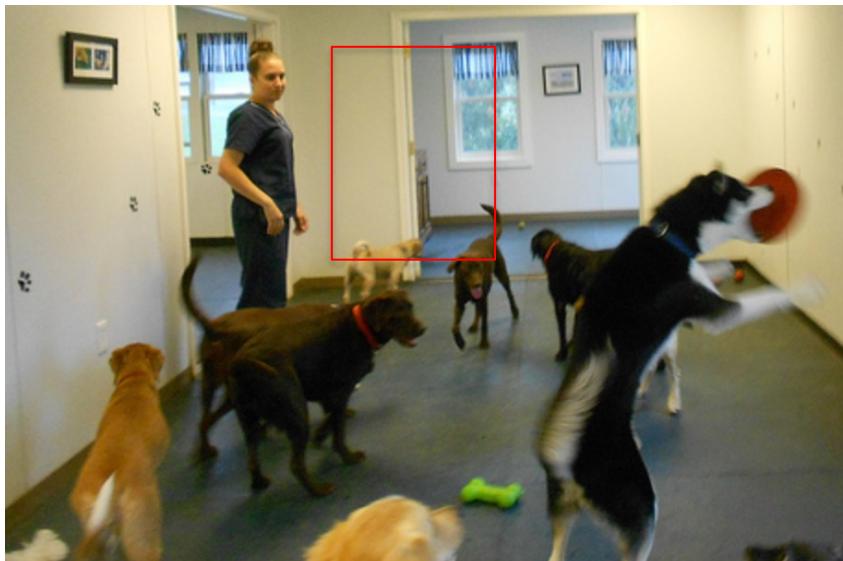
Dog : No

Frisbee : No

Person : Yes

...

물체 탐지: 다수 물체 탐지 ?



Background : Yes

Dog : No

Frisbee : No

Person : No

...

물체 탐지: 다수 물체 탐지 ?



Background : No

Dog : Yes

Frisbee : No

Person : No

...

물체 탐지: 다수 물체 탐지 ?



Background : No

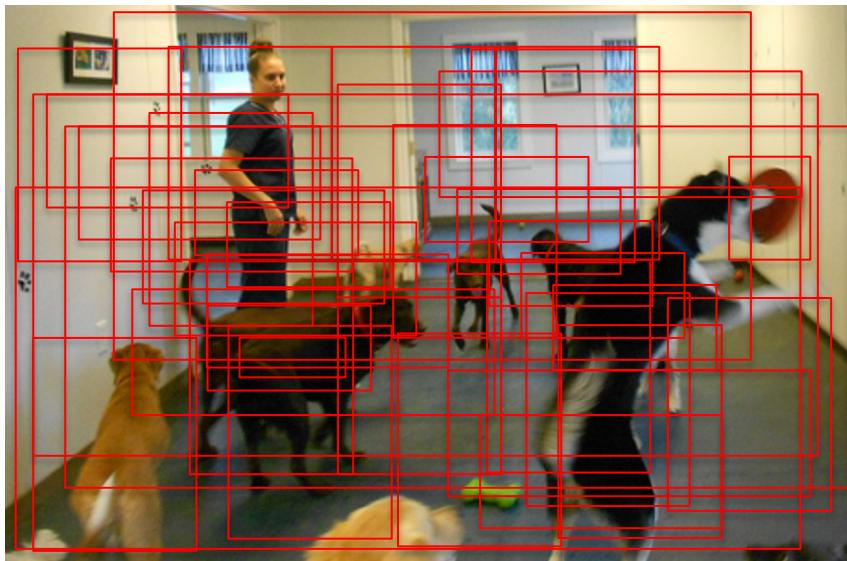
Dog : No

Frisbee : Yes

Person : No

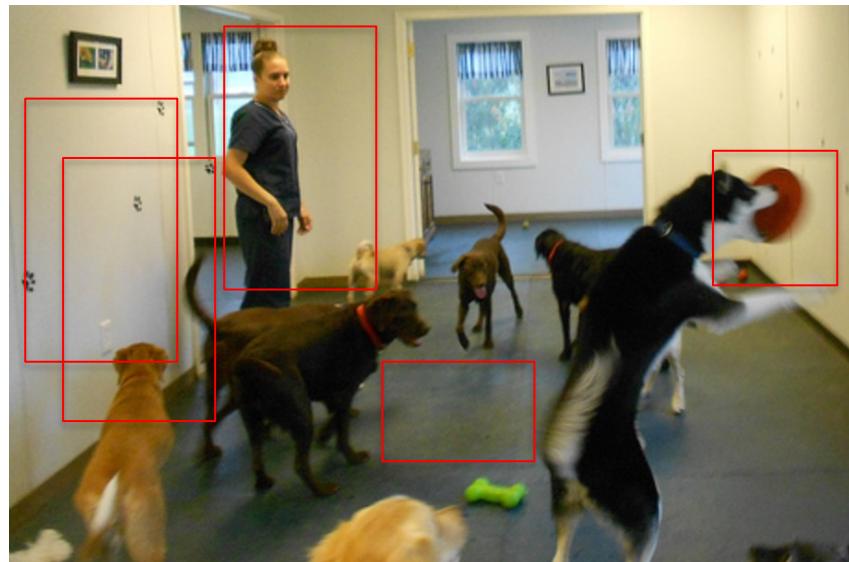
...

물체 탐지: 다수 물체 탐지 ?



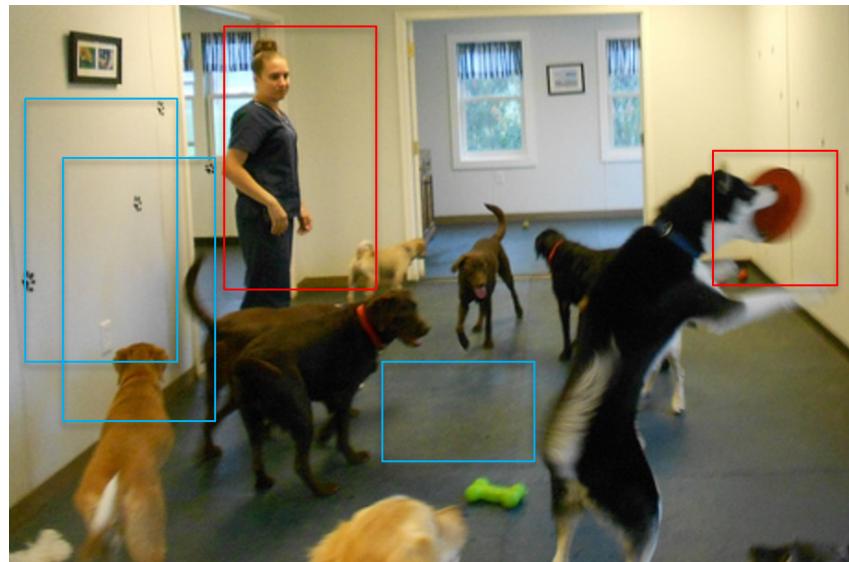
- 다양한 모양의 후보가 필요
- 너무 많은 후보 상자
(bounding box or anchor box)
- 비효율적 검색

선택적 검색 (Selective Search)



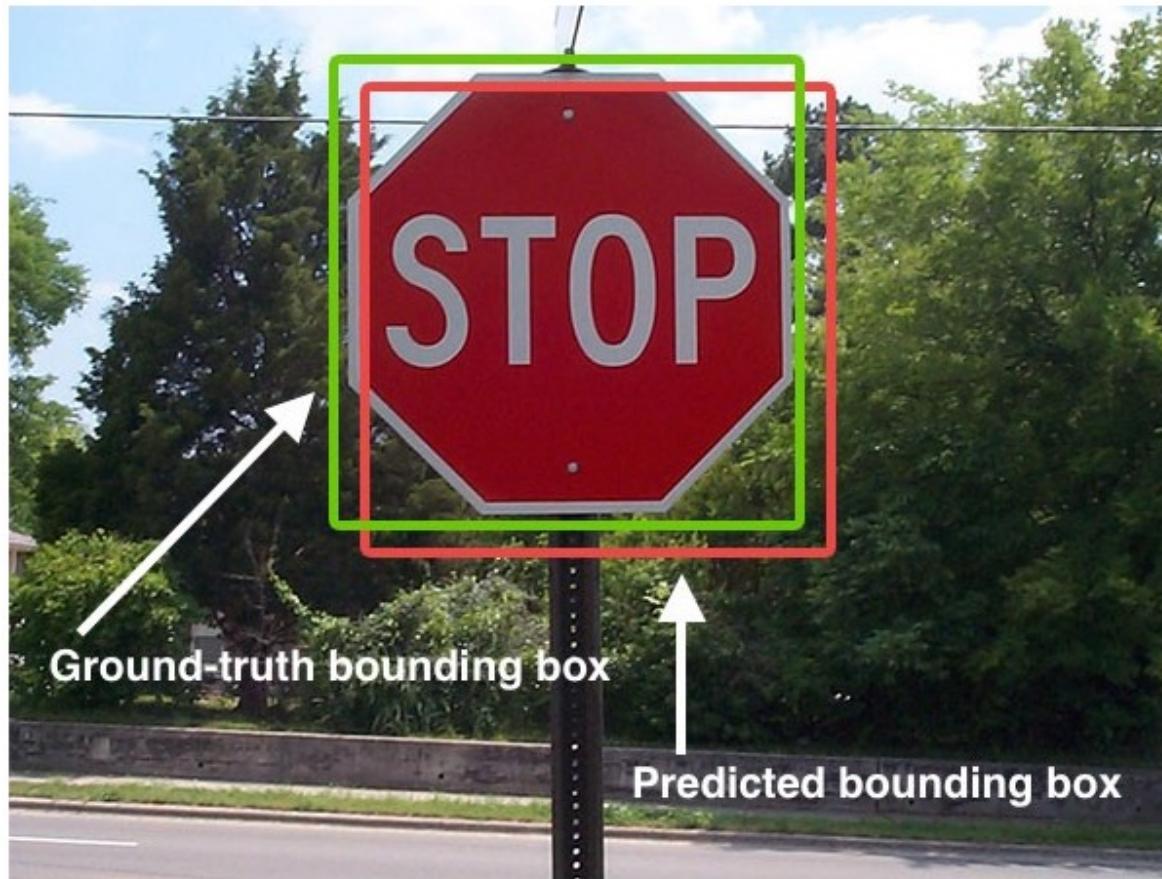
- 물체 유무 우선 판별

선택적 검색 (Selective Search)



- 물체 유무 우선 판별
- 물체 있는 박스만 다시 정밀 분석.

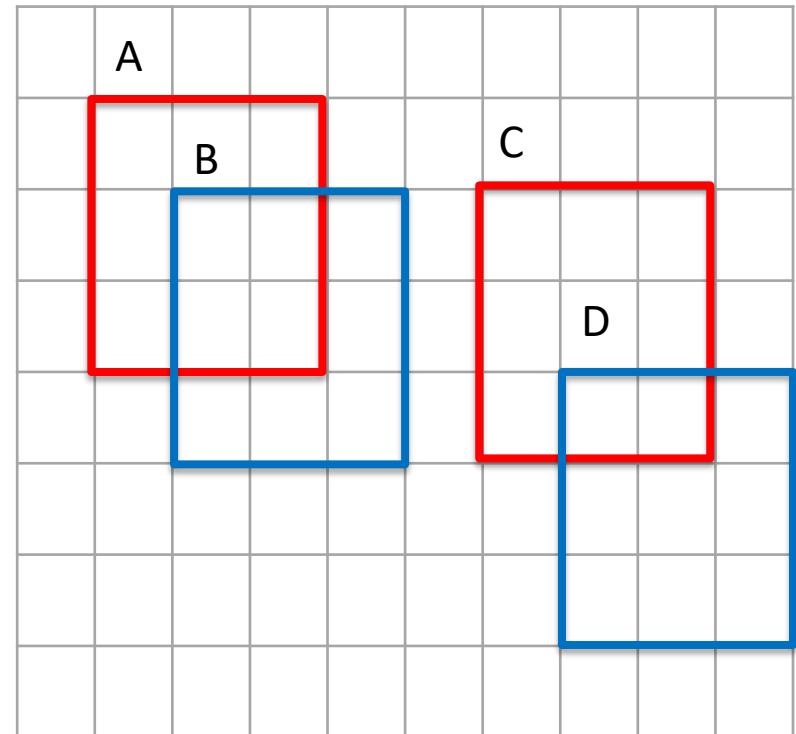
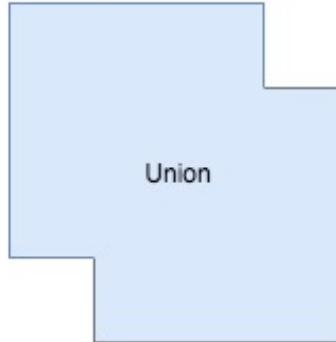
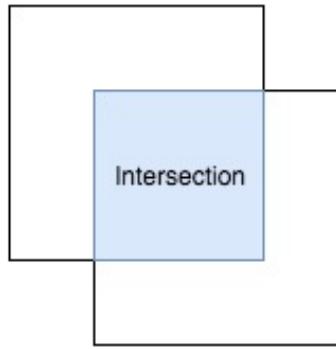
중복 정도 평가/정확도 측정



<https://towardsdatascience.com/pedestrian-detection-using-non-maximum-suppression-b55b89cef6>

교집합/합집합 (IoU)

$$\text{IoU} = \frac{\text{Intersection}}{\text{Union}}$$



$$\text{IoU}(A,B) = 4/14$$

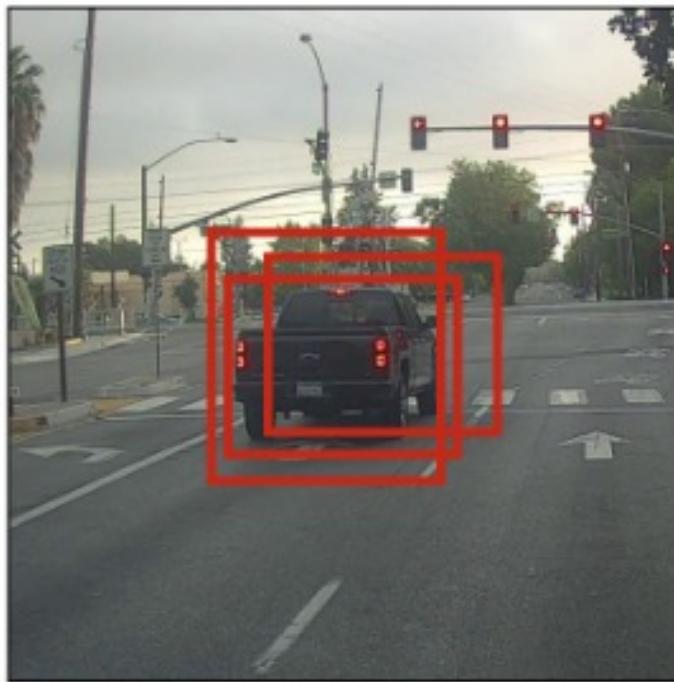
$$\text{IoU}(C,D) = 2/16$$

Intersection over Union

<https://towardsdatascience.com/non-maximum-suppression-nms-93ce178e177c>

중복 탐지 결과 제거 NMS (non-maximum suppression)

Before non-max suppression



Non-Max
Suppression

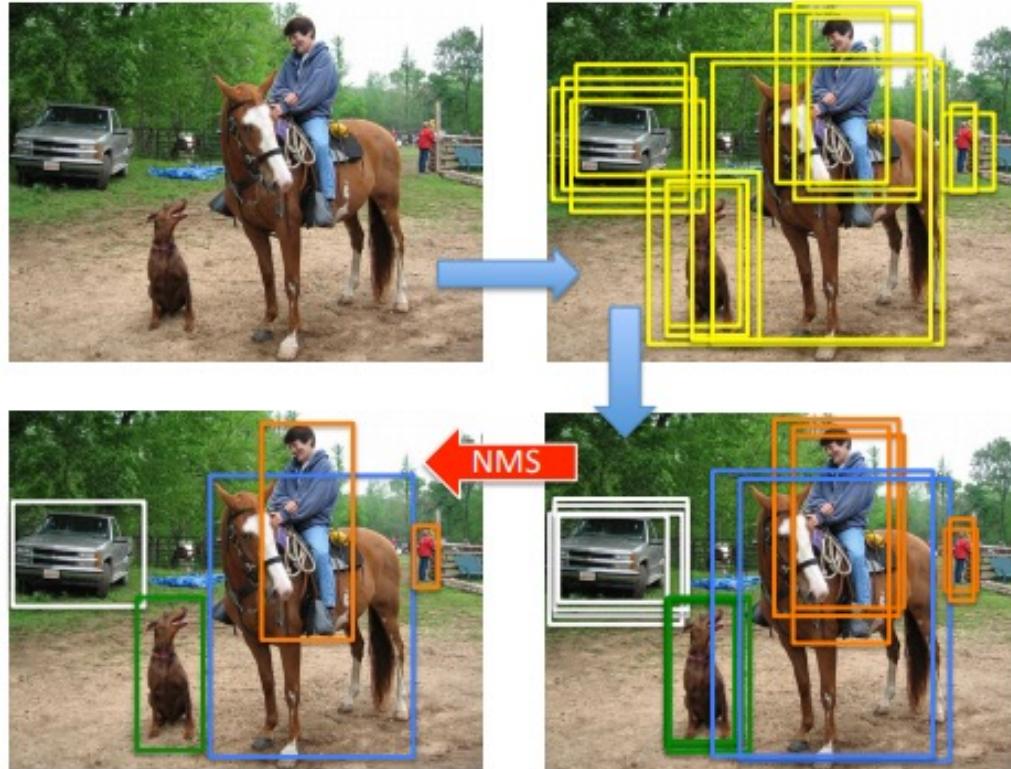


After non-max suppression



<https://towardsdatascience.com/non-maximum-suppression-nms-93ce178e177c>

중복탐지제거 NMS (non-maximum suppression)



- 가장 확신이 높은 탐지 결과 순으로 정렬
- 심각하게 겹친 같은 클래스 탐지 결과 삭제

소주제 1: 물체 탐지 원리 요약

- 물체 탐지 = 위치 추정 (x, y, w, h) + 분류
- 다중태스크 학습, 다중 손실함수
- ImageNet 등에 사전 학습된 기저네트워크 사용
- 후보 박스 (bounding box)
- 선택적 검색: 물체 있는 bounding box 탐지
- IoU (Intersection over Union)를 통한 중복 정도 측정/정확도 측정
- NMS (Non-maximum Suppression) 중복탐지결과 제거

Object Detection

소주제 2: 물체 탐지 기법

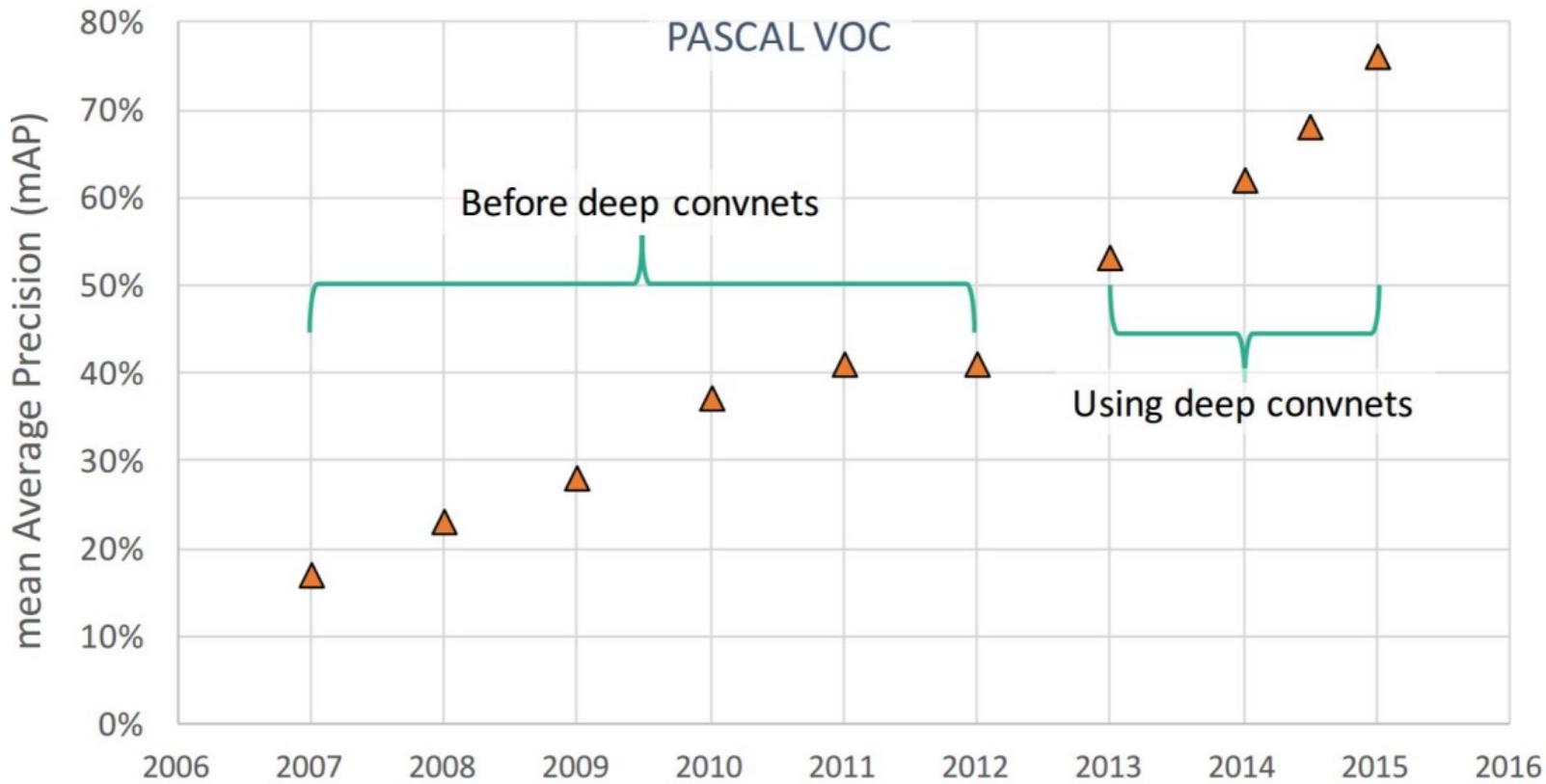
김현우 (hyunwoojim@korea.ac.kr)

(정보대학 컴퓨터학과 기계학습 및 비전 연구실)

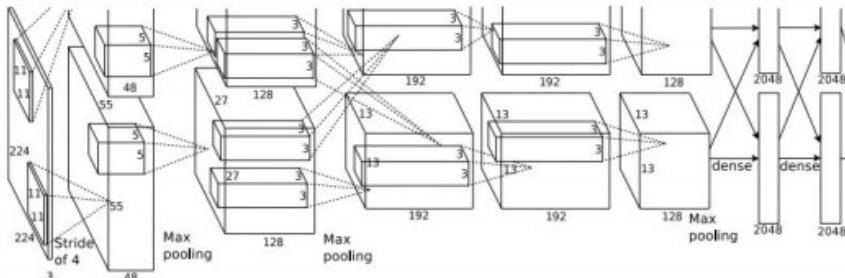
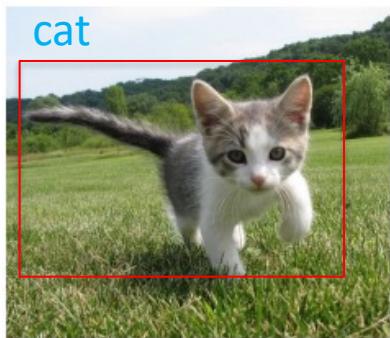
인공지능은 물체를 어떻게 이해할까?

- 소주제 1: 물체 탐지 원리
- 소주제 2: 물체 탐지 기법
- 소주제 3: 물체간 관계 추론

딥러닝 기반 물체 탐지 기법



물체 탐지=위치 추정 + 분류



다중 태스크 학습
다중 손실 함수

위치추정: localization
(x, y, w, h)
L2 손실함수

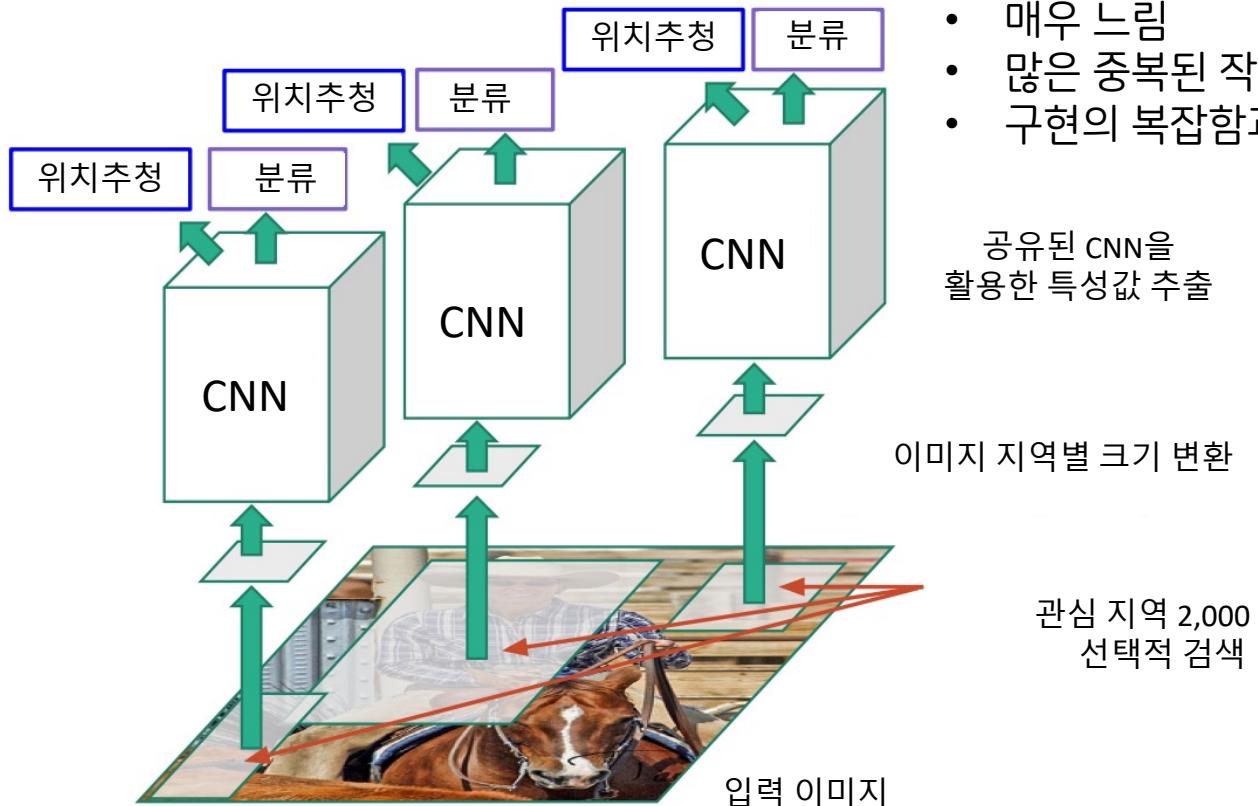
+

분류: classification
고양이
softmax 손실함수

다단계 물체 탐지 기법

- 다단계 물체 탐지 기법
- RCNN
- Fast-RCNN
- Faster-RCNN
- Mask-RCNN
- Detectron2

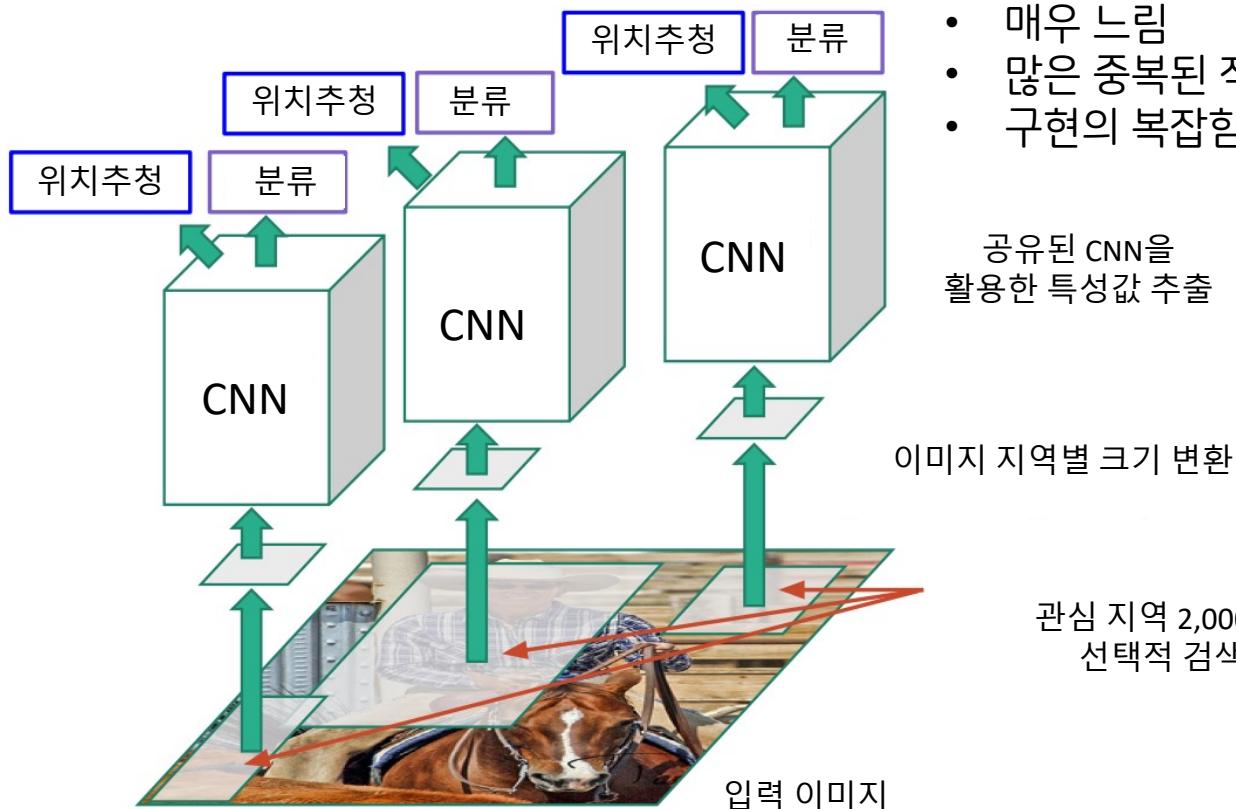
RCNN: 초기 모델



- 위치추정: 회기분석 (dx, dy, dw, dh)
- 분류: support vector machine (SVM)
- 매우 느림
- 많은 중복된 작업 존재 2,000 번 특성값 추출
- 구현의 복잡함과 비효율적 GPU 사용

Girshick et al., CVPR 2014

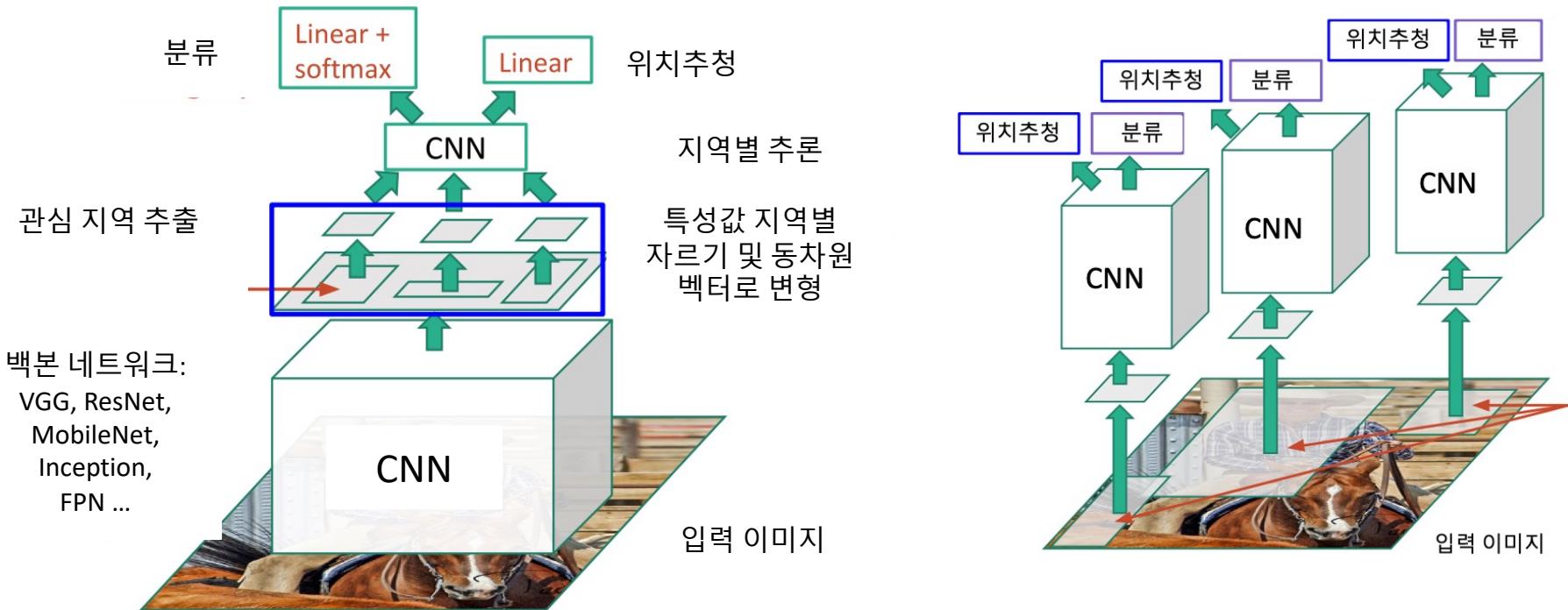
RCNN: 초기 모델



- 위치추정: 회기분석 (dx, dy, dw, dh)
- 분류: support vector machine (SVM)
- 매우 느림
- 많은 중복된 작업 존재 2,000 번 특성값 추출
- 구현의 복잡함과 비효율적 GPU 사용

Girshick et al., CVPR 2014

Fast RCNN



지역별 동차원 특성값 추출

A

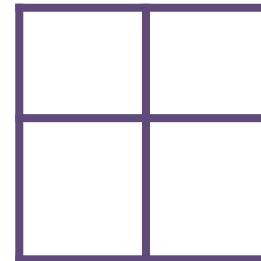
1	2	3	2	5	6	7	4	8	2
5	2	5	6	7	4	10	3	4	8
3	3	8	7	4	7	8	2	5	5
2	9	4	10	2	2	3	2	7	2
8	7	3	9	2	8	7	9	2	10
5	8	3	5	7	5	10	5	1	5
8	4	2	6	10	9	1	8	2	9
9	3	6	1	3	1	8	1	6	2

B

A: 8 차원
B: 24 차원

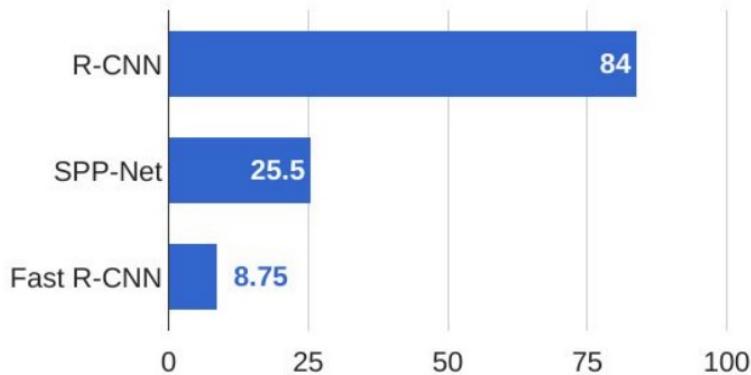
어떻게 같은 차원으로
만들 수 있을까요?

ROI Pooling (샘플링)
각 세부구역 최대치 사용



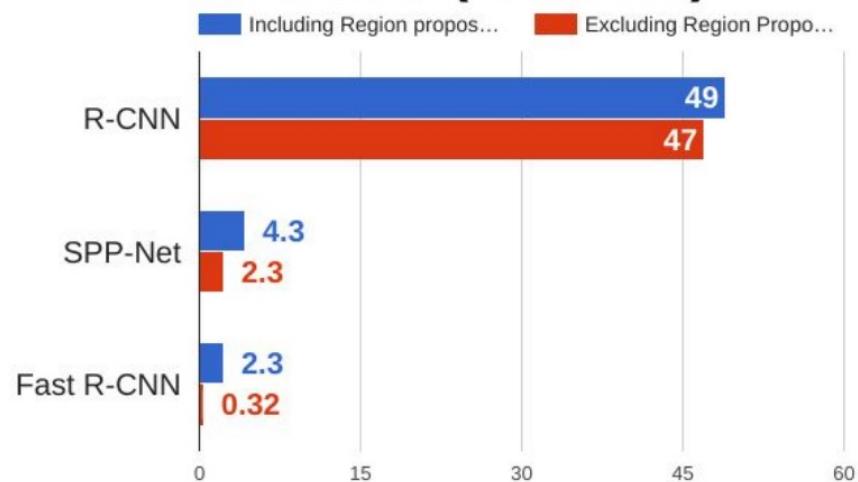
학습 및 추론 속도 개선

Training time (Hours)



약 10배 향상

Test time (seconds)

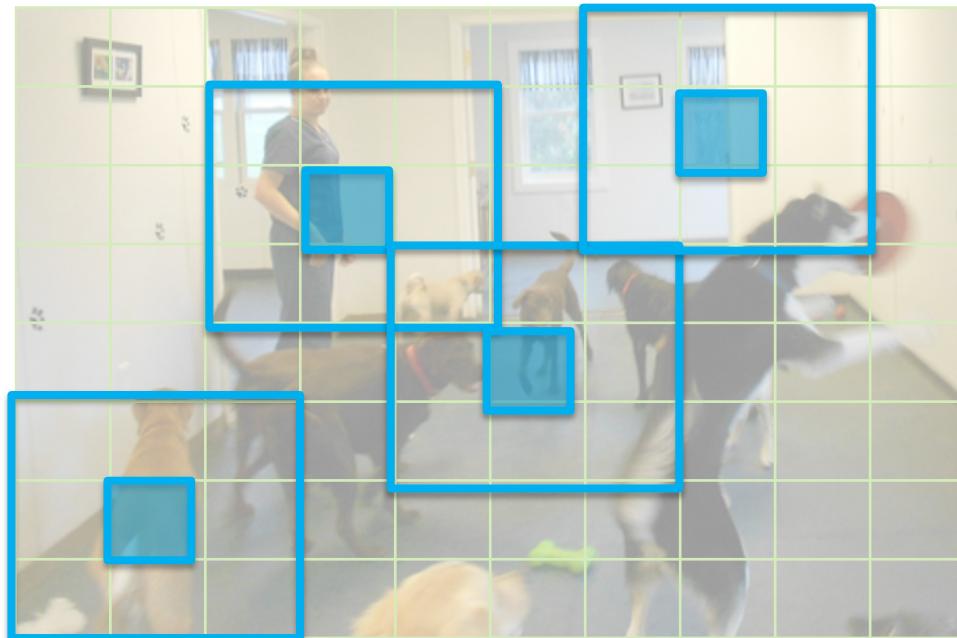


약 21배 향상

관심 지역 추출 2 sec!

관심지역 추출 네트워크 (RPN)

RPN CNN: 관심지역 추출

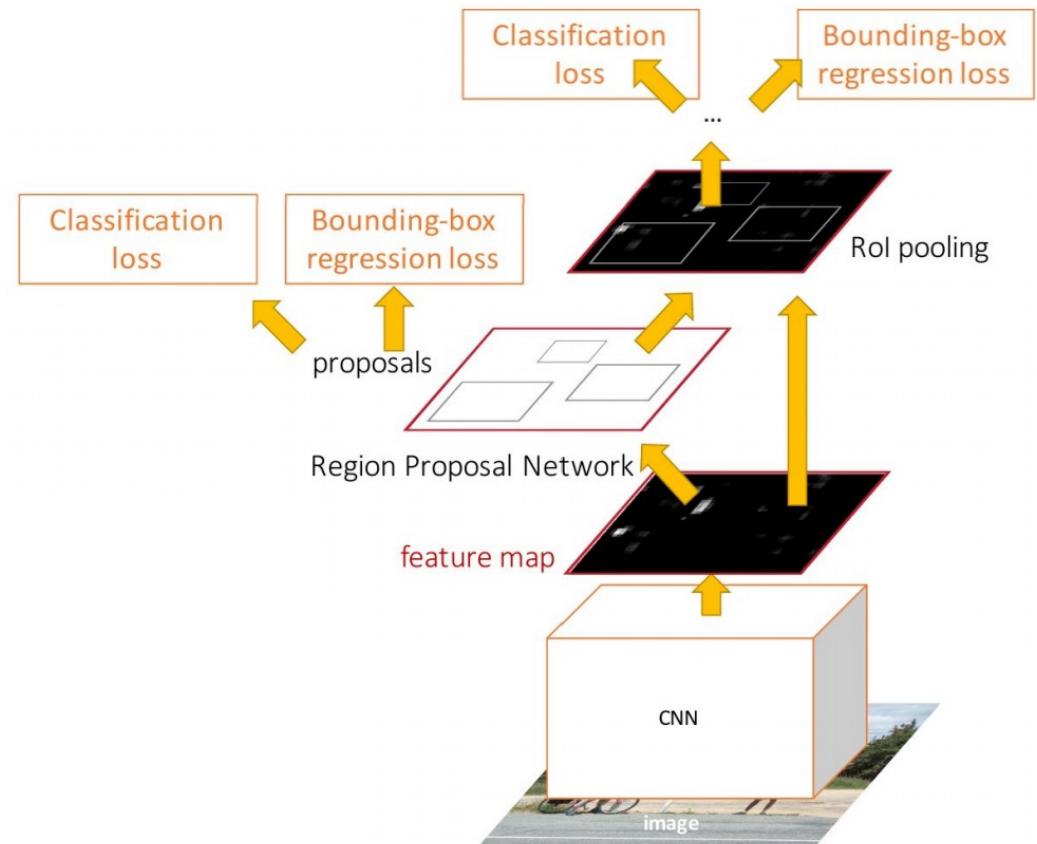


Backbone CNN: 특성값 추출

- 동일한 크기의 고정 박스 (Anchor box)
- 각 고정 박스별 물체 포함 가능성 1차 판별 (딥러닝 기반 선택적 검색)
- 이미지 전체 특성값 재활용

Faster RCNN

- 딥러닝 기반 선택적 검색
- 특성값 재사용



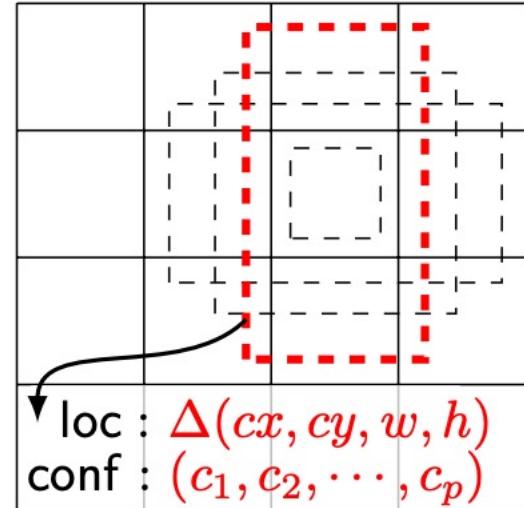
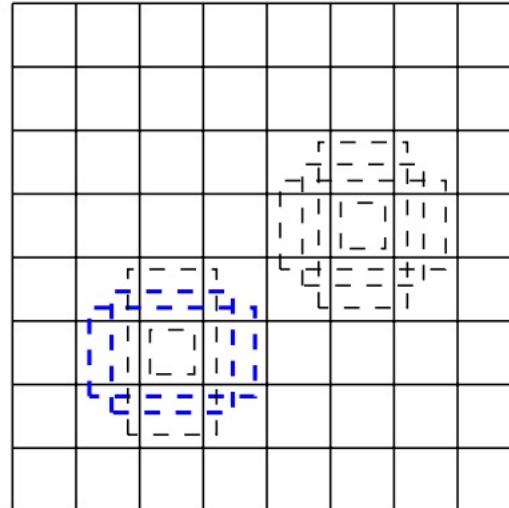
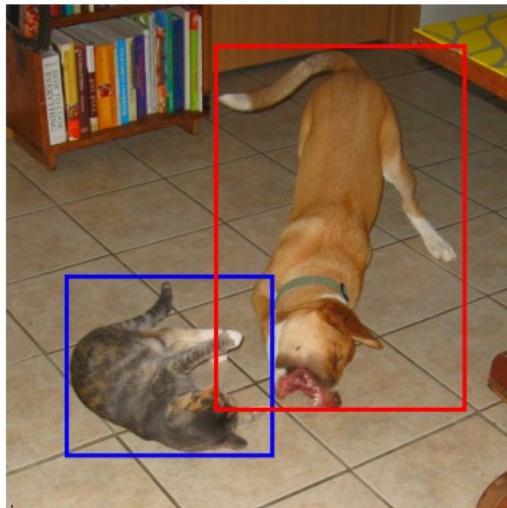
Ren, Shaoqing, et al. NeurIPS 2015

단단계 vs 다단계 물체 탐지 기법

- 다단계 물체 탐지 기법 (Multi-stage Object Detection)
 - RCNN, Fast-RCNN, Faster-RCNN …
- 단단계 물체 탐지 기법 (Single-stage OD)
 - Single-shot object detector, YOLO, RetinaNet…

Single-Shot Object Detector

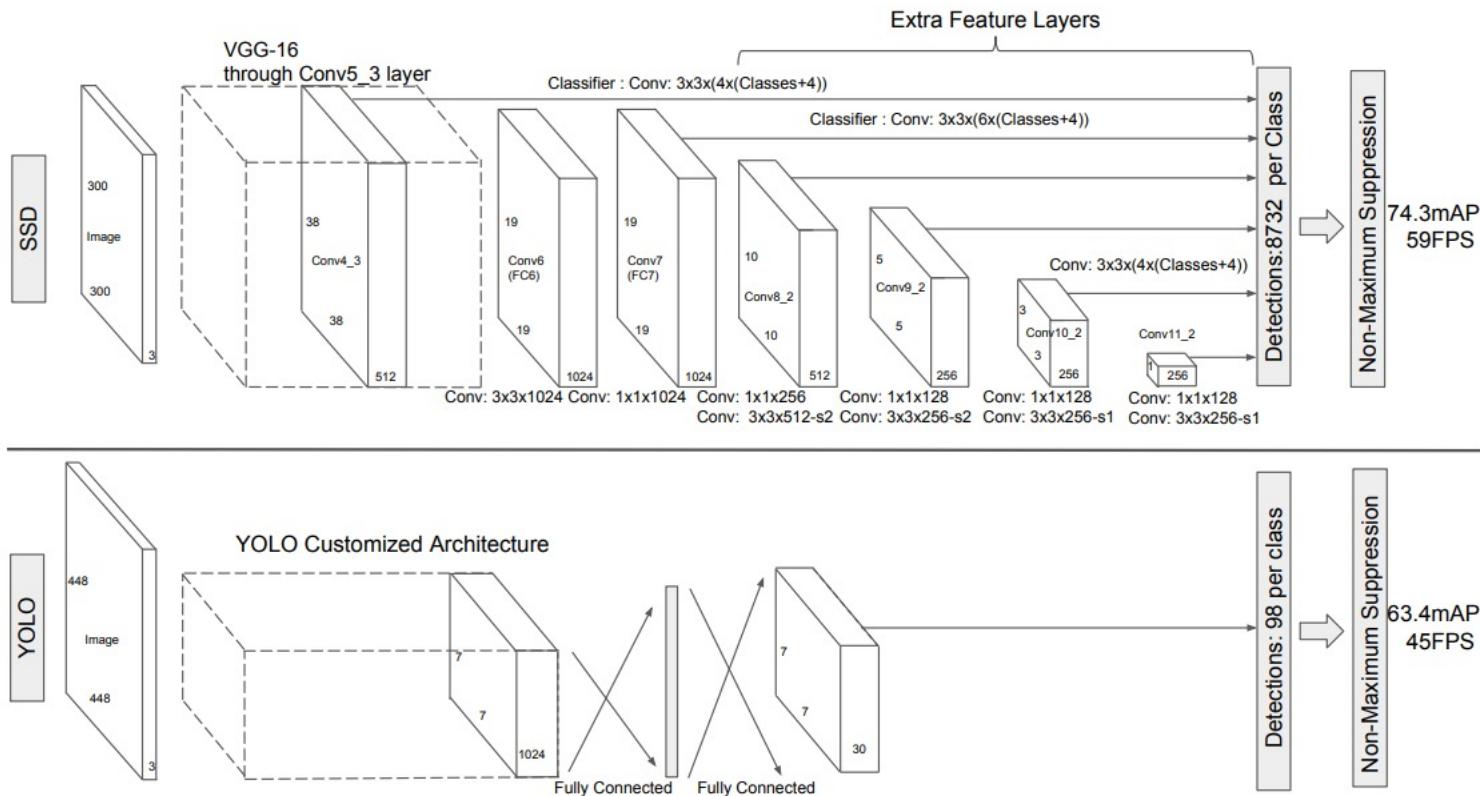
- SSD ~ 선택적 검색 (X)
- GPU 한번에 추론
- 구현이 상대적으로 덜 복잡함



Liu, Wei, et al. "Ssd: Single shot multibox detector." ECCV 2016.

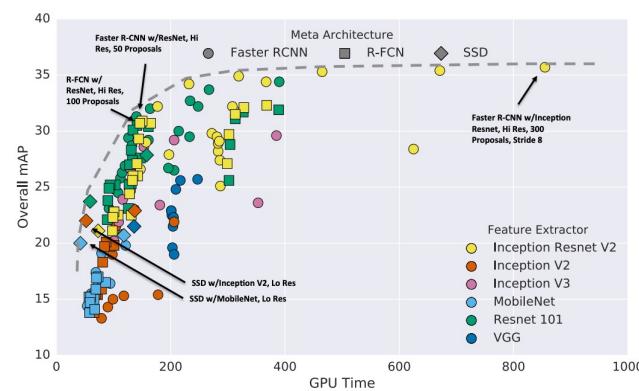
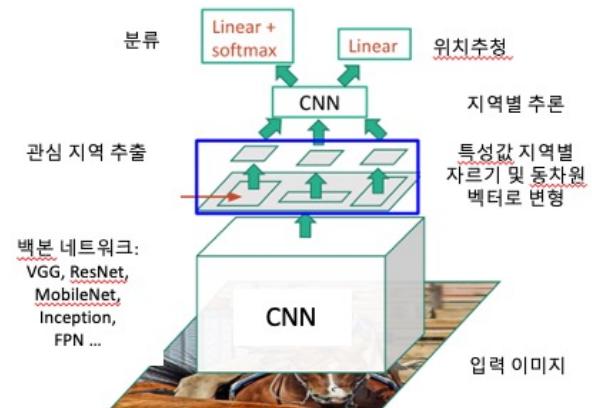
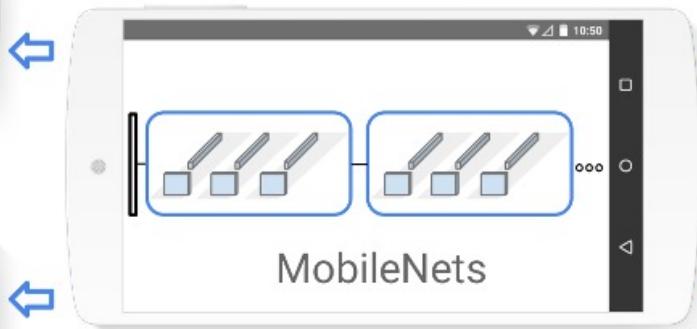
Single-Shot Object Detector

- 멀티 스케일 분석 (Multi-scale analysis)



Liu, Wei, et al. "Ssd: Single shot multibox detector." ECCV 2016.

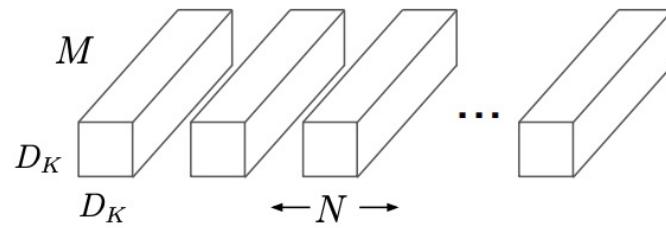
기반 네트워크 (backbone)



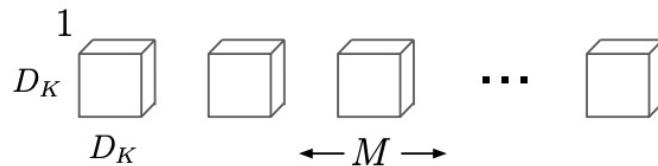
<https://ai.googleblog.com/2017/06/mobilenets-open-source-models-for.html>

Huang, Jonathan, et al. " CVPR 2017.

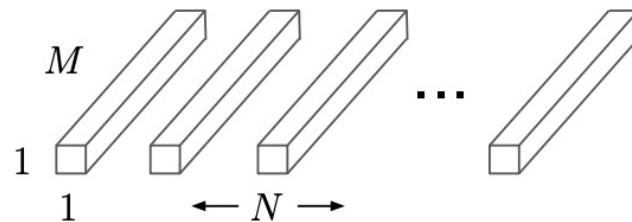
Depthwise Separable Convolution



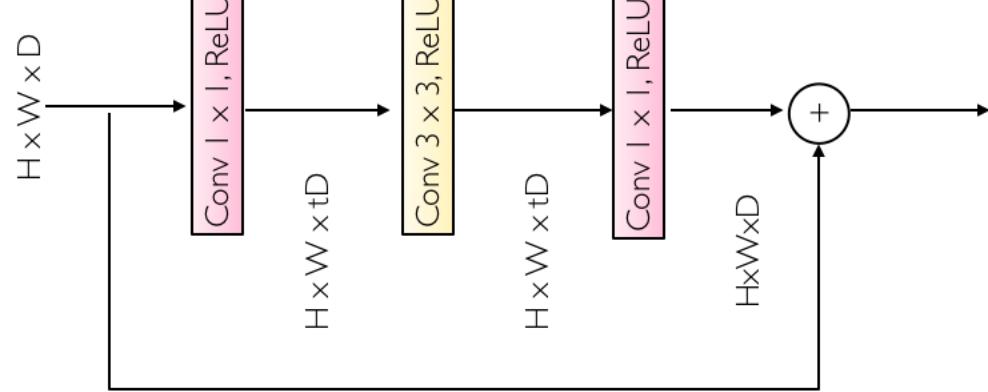
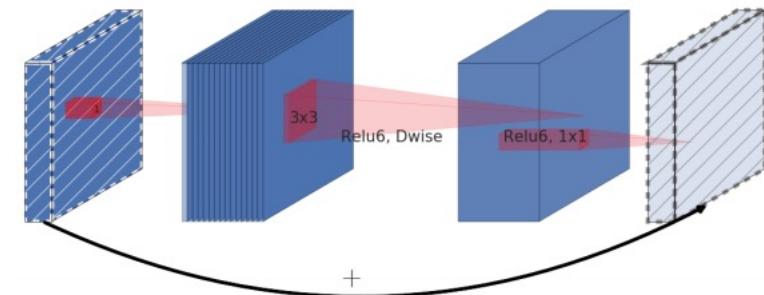
(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters



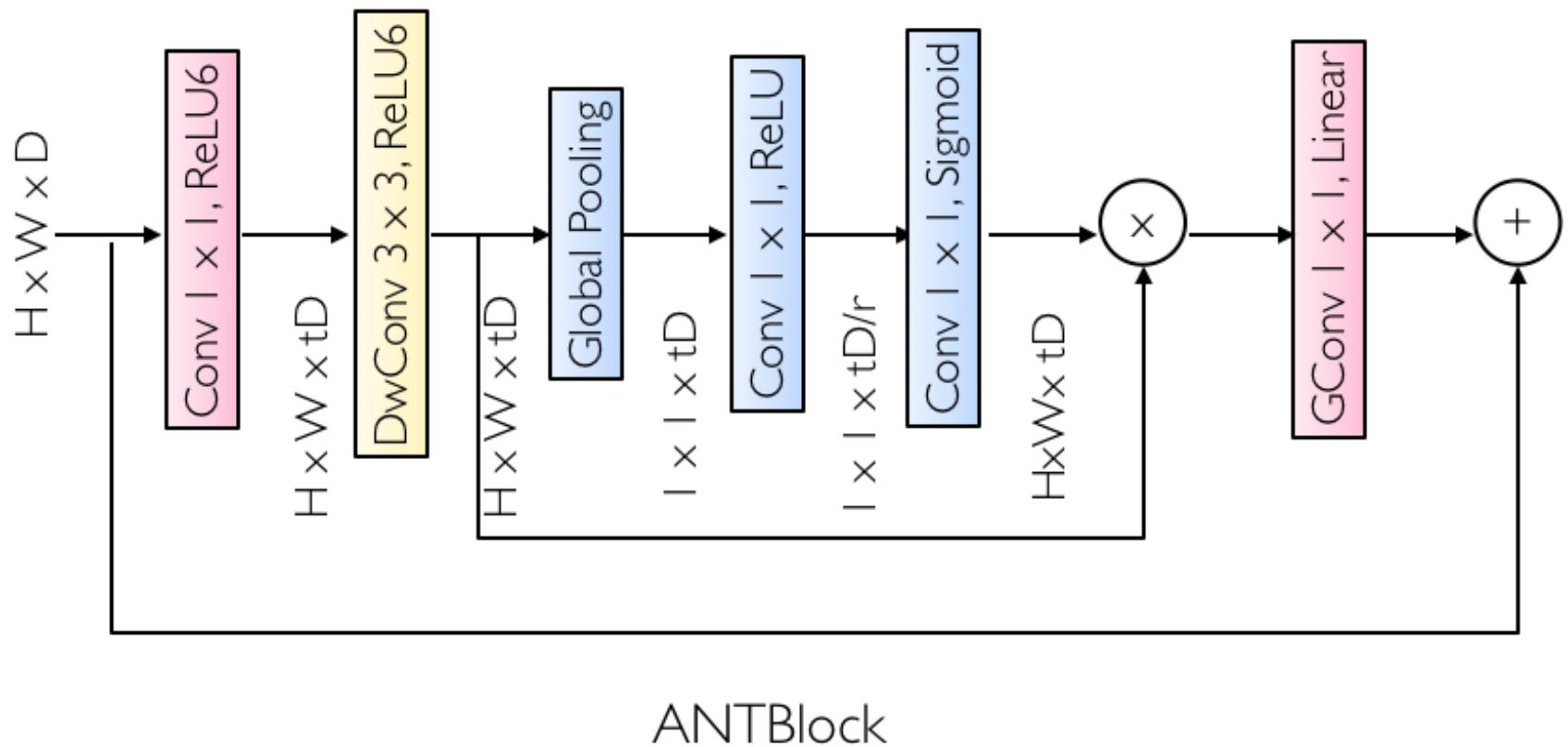
MobileNet V1



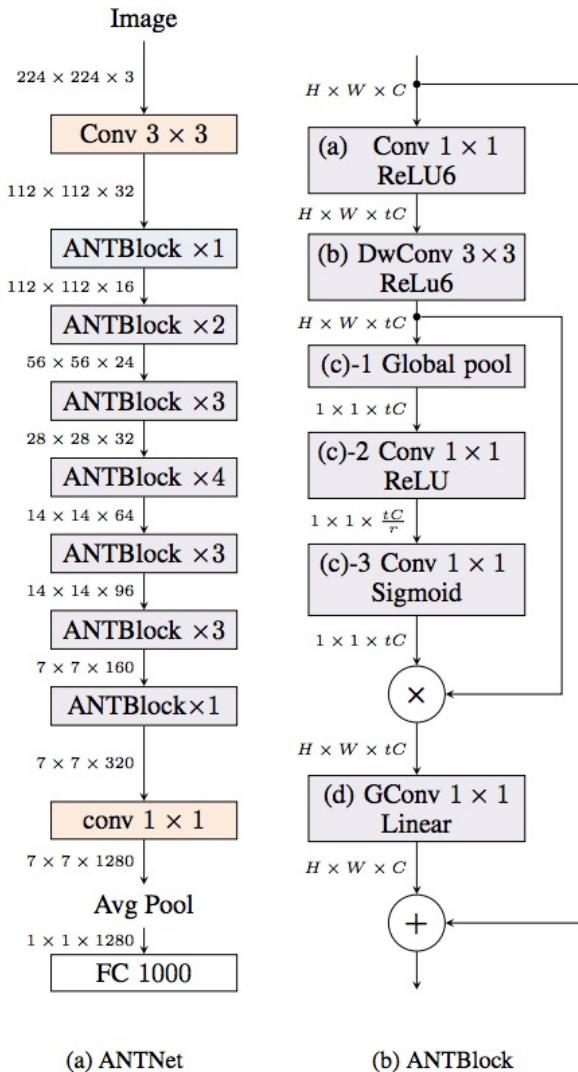
Inverted Residual Block

MobileNet V2

ANTBlock with Channel Attention



ANTNets



정확도 향상:
Inverted Residual Block + Squeeze-N-Excitation
(MobileNet V2 Block) (Channel Attention)

속도 향상:
Group Convolution (2 groups, 10-20% speed up)
Location: before projection
The highest number of channels

CIFAR 100

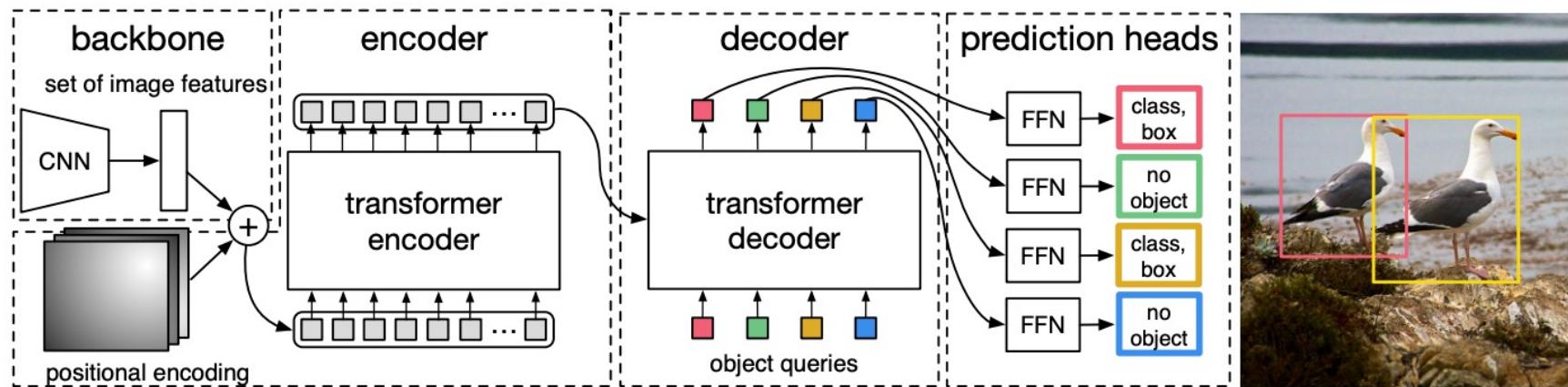
	Network	Top-1 Accu.	Top-5 Accu.
Google	MobileNetV2	74.2	93.3
Ours	se-MobileNetV2	74.1	92.8
	c-ANTNet	73.4	93.3
	ANTNet-c	74.4	93.5
Ours	ANTNet (proposed)	75.7	93.6

+1.5

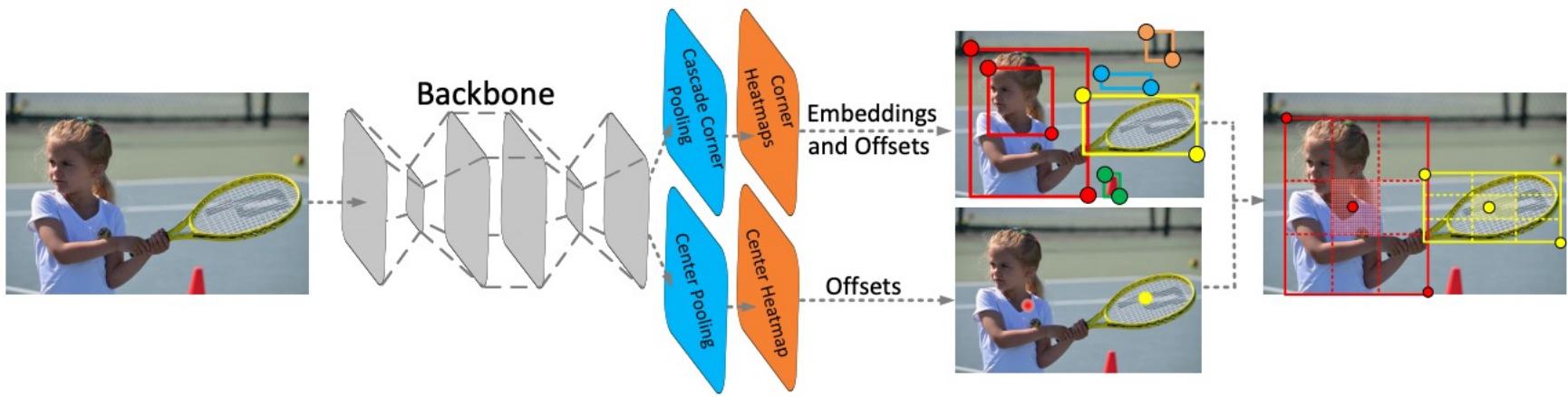
[CVPRW 2019] ANTNets: Mobile Convolutional Neural Networks for Resource Efficient Image Classification,

어텐션 기반 물체 탐지

End-to-End Object Detection with Transformers, ECCV 2020

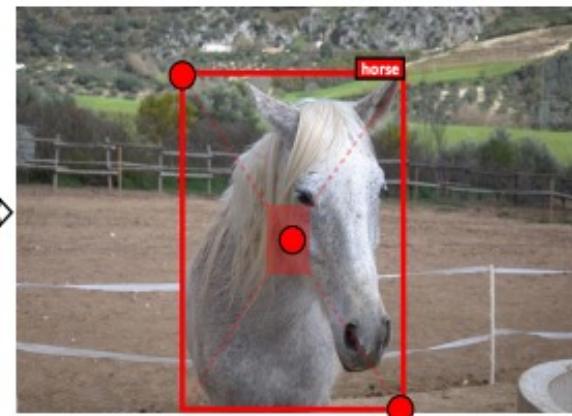
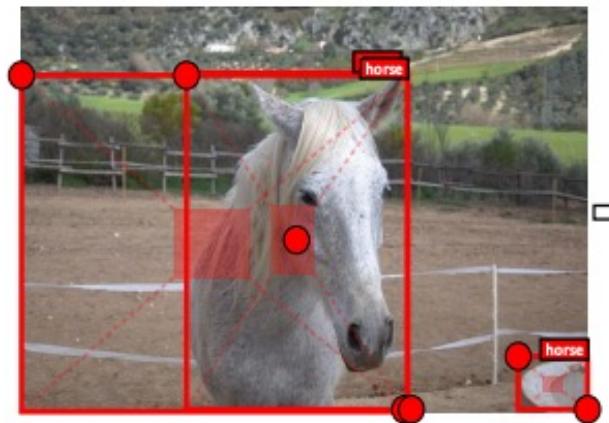
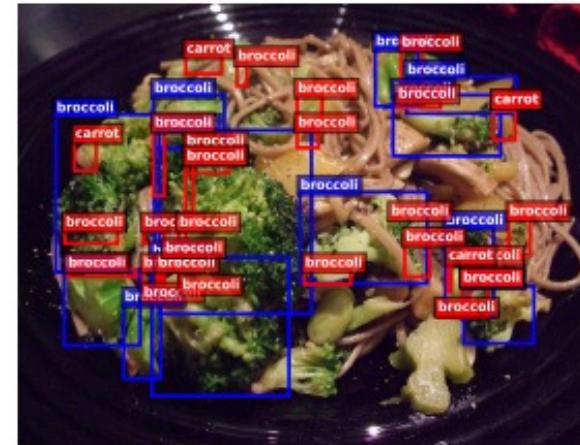
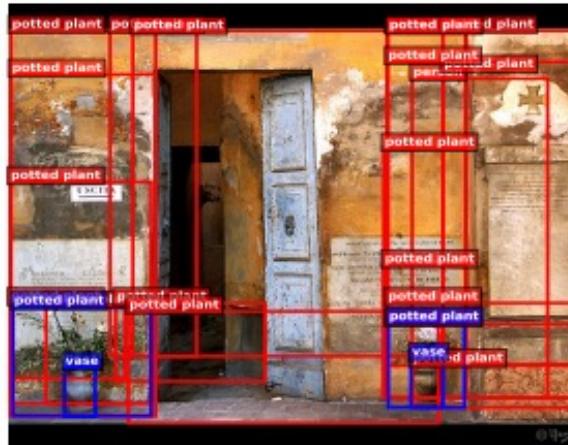


CenterNet: 키포인트 기반 물체 탐지



Duan, Kaiwen, et al. "Centernet: Keypoint triplets for object detection." ICCV 2019.

CornerNet vs CenterNet



Tensorflow Object Detection API

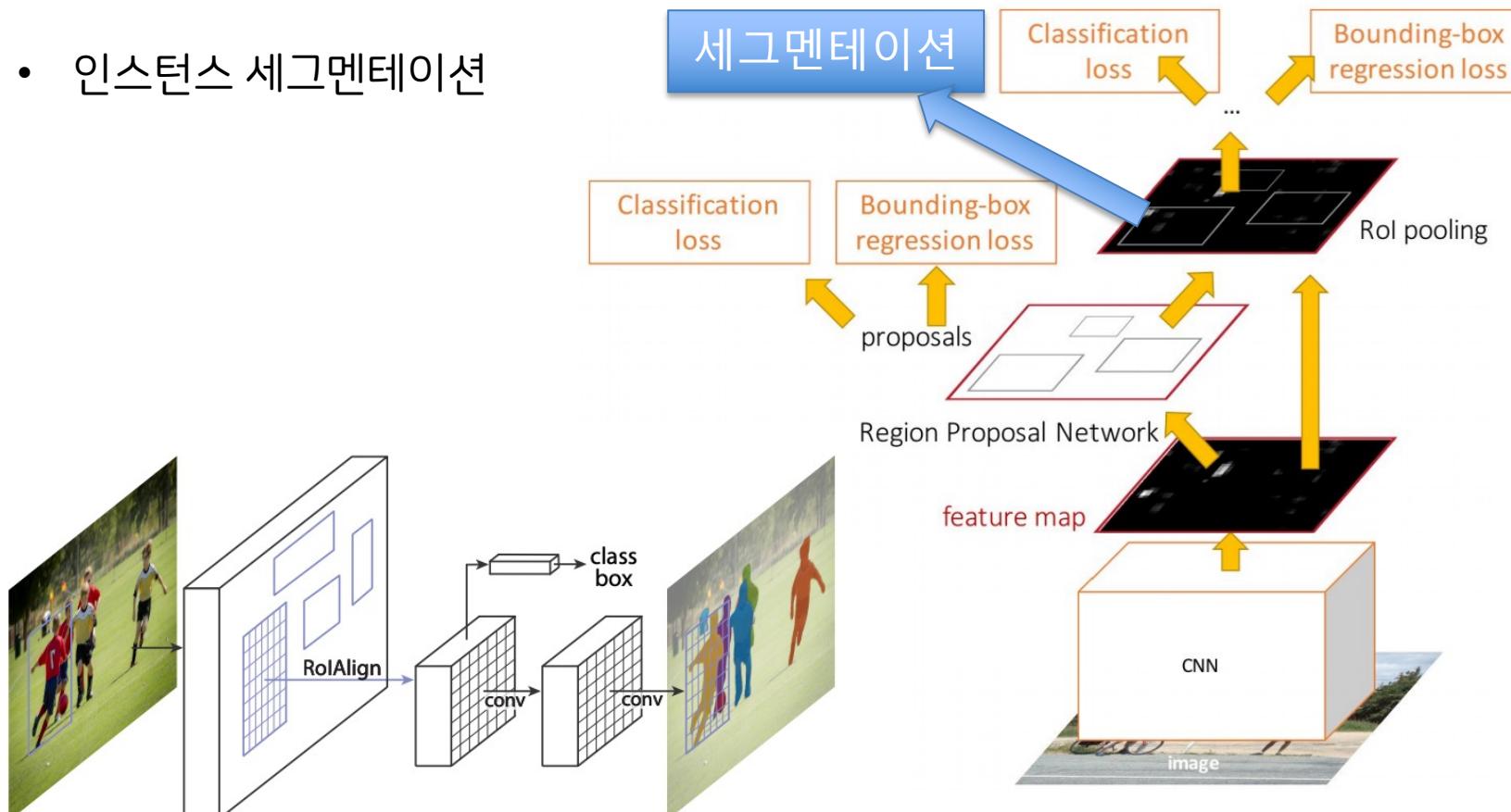
Creating accurate machine learning models capable of localizing and identifying multiple objects in a single image remains a core challenge in computer vision. The TensorFlow Object Detection API is an open source framework built on top of TensorFlow that makes it easy to construct, train and deploy object detection models. At Google we've certainly found this codebase to be useful for our computer vision needs, and we hope that you will as well.



https://github.com/tensorflow/models/tree/master/research/object_detection

Mask-RCNN

- 인스턴스 세그멘테이션

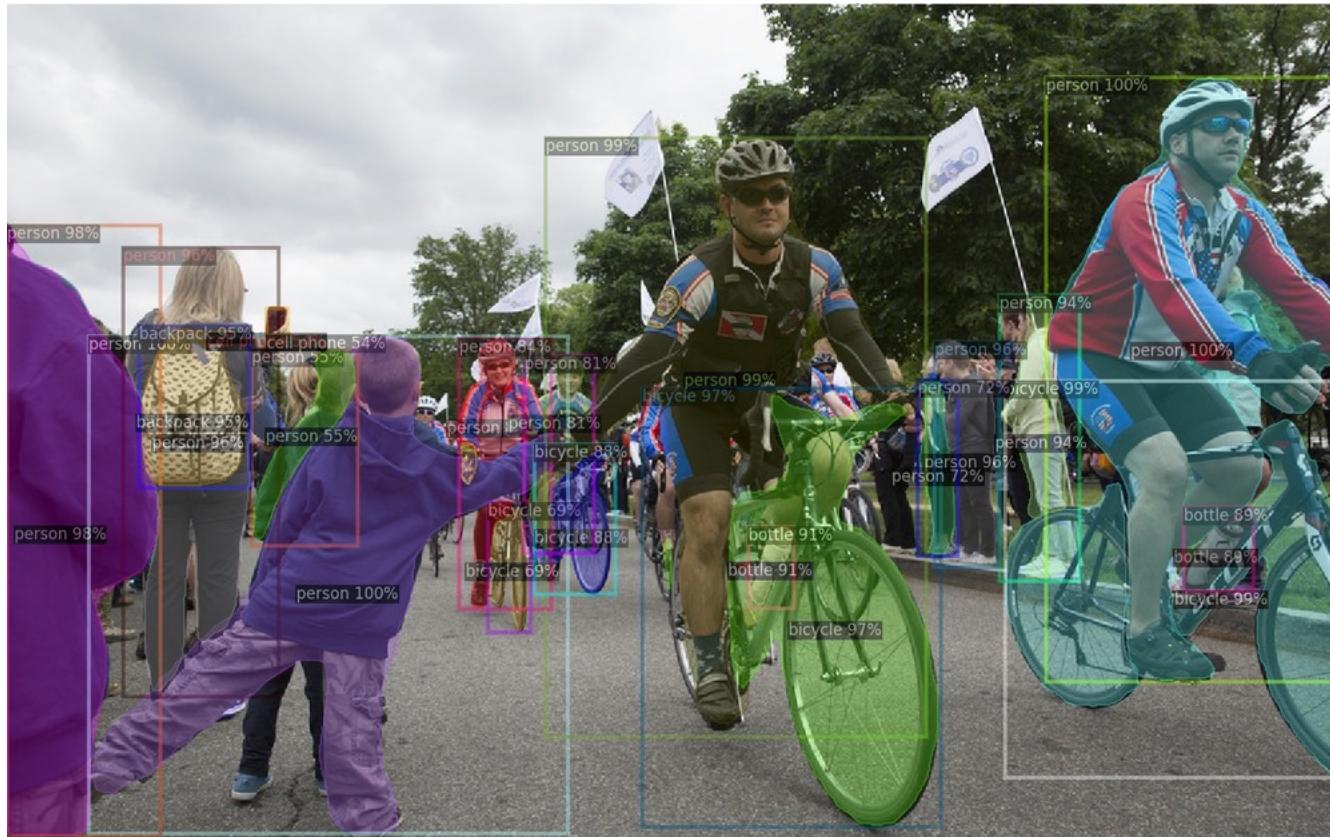


Ren, Shaoqing, et al. NeurIPS 2015

인스턴스 세그멘테이션 (Instance Segmentation)

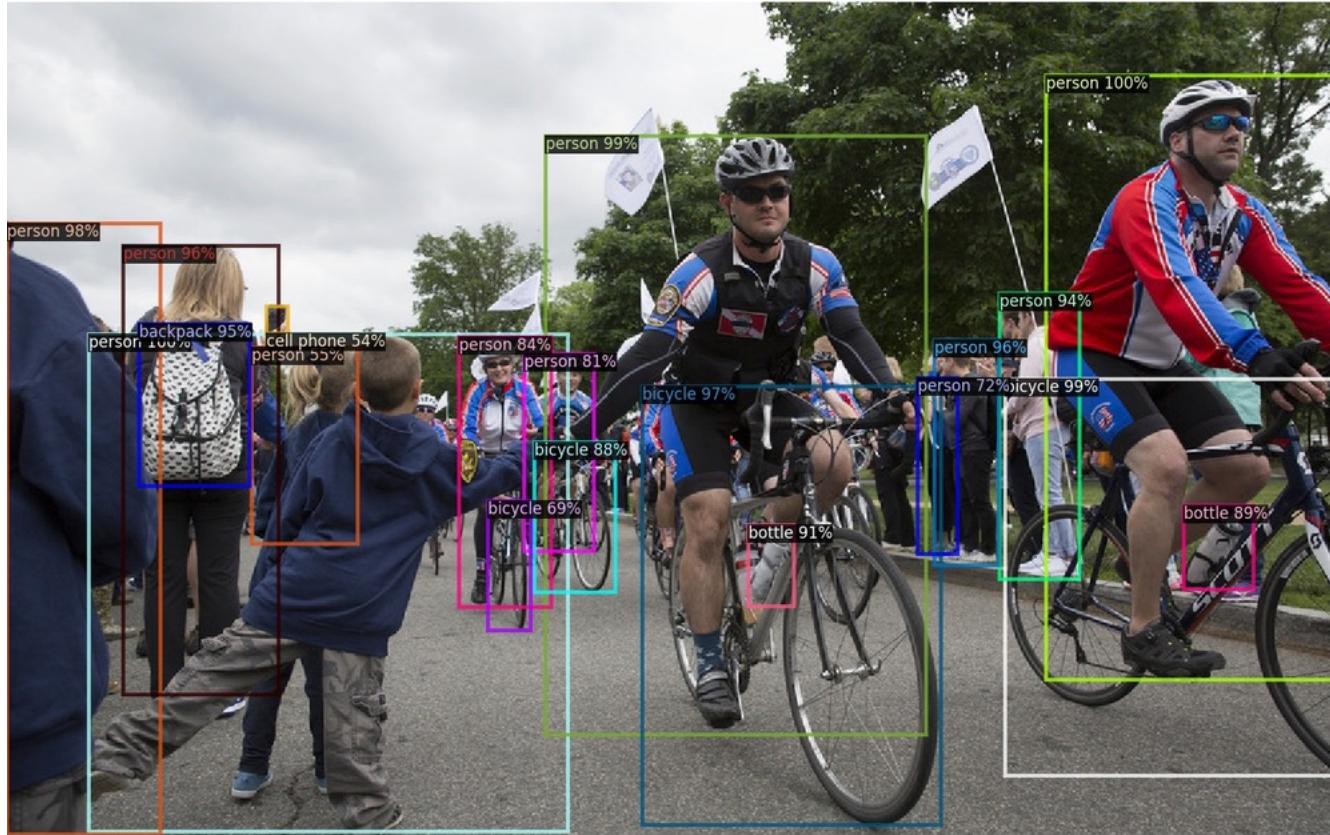


Detectron2 is Facebook AI Research's next generation software system that implements state-of-the-art object detection algorithms. It is a ground-up rewrite of the previous version, [Detectron](#), and it originates from [maskrcnn-benchmark](#).



물체 탐지 + 인스턴스 세그멘테이션

Detectron 2



물체 탐지

Detectron 2



판옵틱 세그멘테이션

Detectron 2



포즈 추정 (pose estimation)

Detectron 2



포즈 추정 (dense pose estimation)

소주제 2: 물체 탐지 기법 요약

- 물체 탐지 기법
- 다단계 탐지 기법: RCNN, Fast-RCNN, Faster-RCNN
- 단단계 탐지 기법: SSD, YOLO
- RPN 딥러닝 기반 선택적 검색
- ROI Pooling 활용 지역별 특성값 동일 차원 벡터로 변형
- 고정 박스(Anchor box) 없는 최신 기법: DETR, CenterNet
- 다중 태스크 물체 이해: 세그멘테이션, 포즈 추정