Rich feature hierarchies for accurate object detection and semantic segmentation

SKT Fellowship

LEE JINKYU

Department of Civil, Environmental and Architectural Engineering, Korea University

January 29, 2023

SKT AI Fellowship

Introduction

About R-CNN

1. About Object Detection

2. Object Detection Metric

- IoU
- Pecision & Recall (AP, mAP)

3. Process of R-CNN

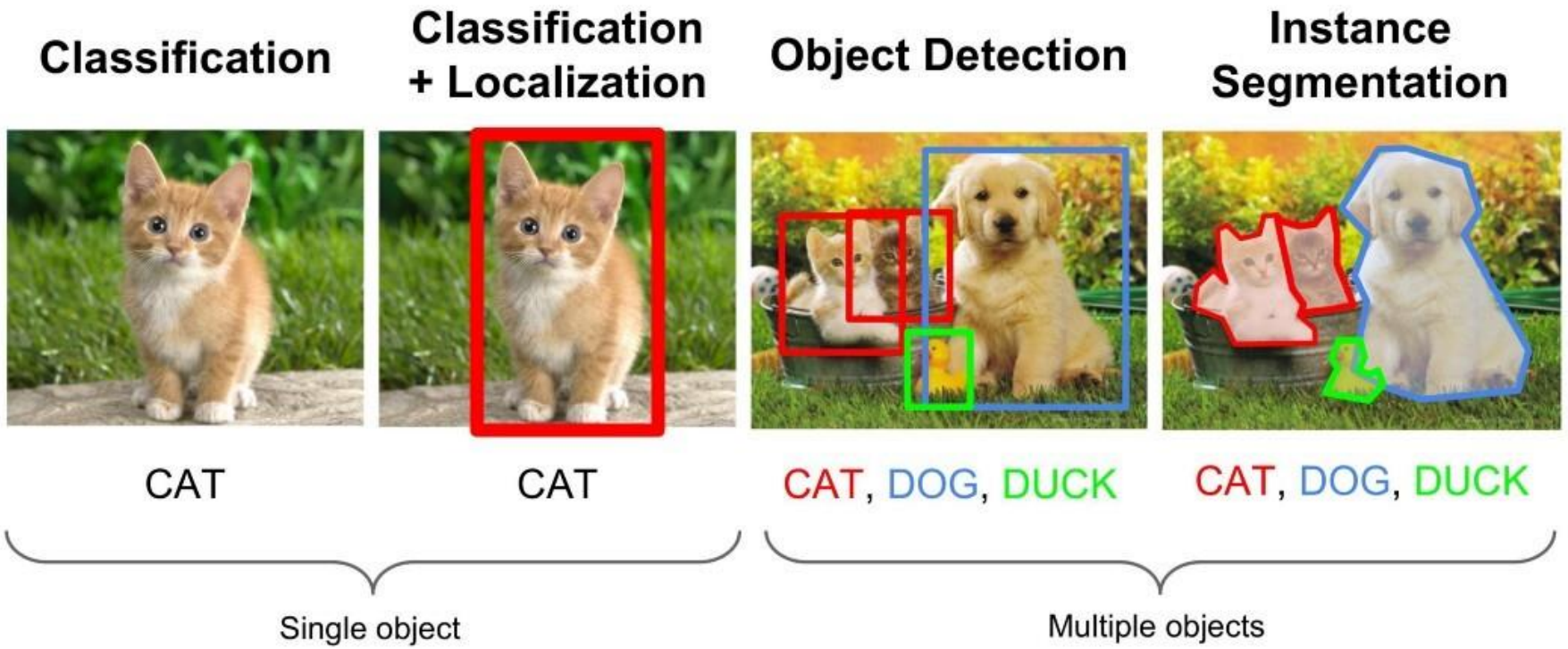
- Overall process
- Selective search
- image warpping
- Linear SVM
- Bounding Box regressor
- Non Maximum Suppression

About Object Detection

What is Object Detection

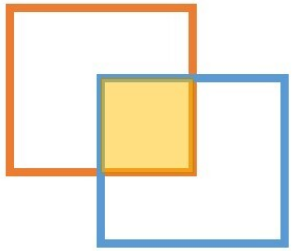
Object Detection :

- [물체의 위치 인식 (Bounding Box) + 물체 판별 (Classification)] 과정이 Multiple objects에 대해서 수행되는 알고리즘
- Bounding Box의 좌표와 각 Box의 Confidence score를 출력한다
- 1 stage, 2 stage 알고리즘이 각각 존재하며 R-CNN의 경우 2 stage 알고리즘이다



Object Detection Metric - IoU

Intersection over Union



$$Intersection\ over\ Union\ (IoU) = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

IoU
- 객체의 위치 추정 정확도를 평가하는 평가 지표이다
- IoU = 두 영역의 교집합의 크기 / 두 영역의 합집합의 크기

→ Object Detection의 경우 IoU를 기반으로 정밀도(Precision), 재현율(Recall)을 구한다

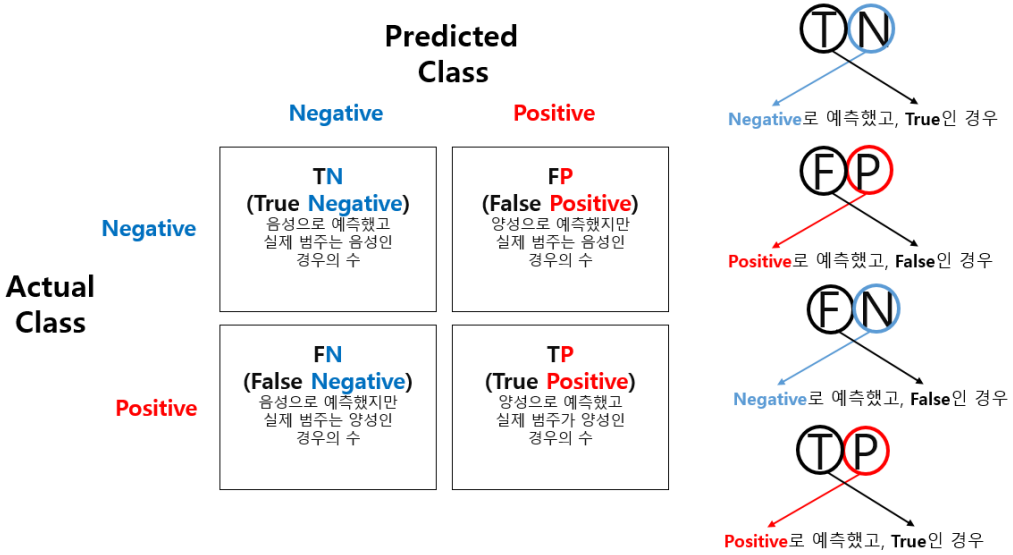
Object Detection Metric – AP, mAP

Defined by Precision & Recall

Precision (정밀도) = 예측 값 1 중 실제 값 1 / 예측 값 1
Recall (재현율) = 실제 값 1 중 예측 값 1 / 실제 값 1

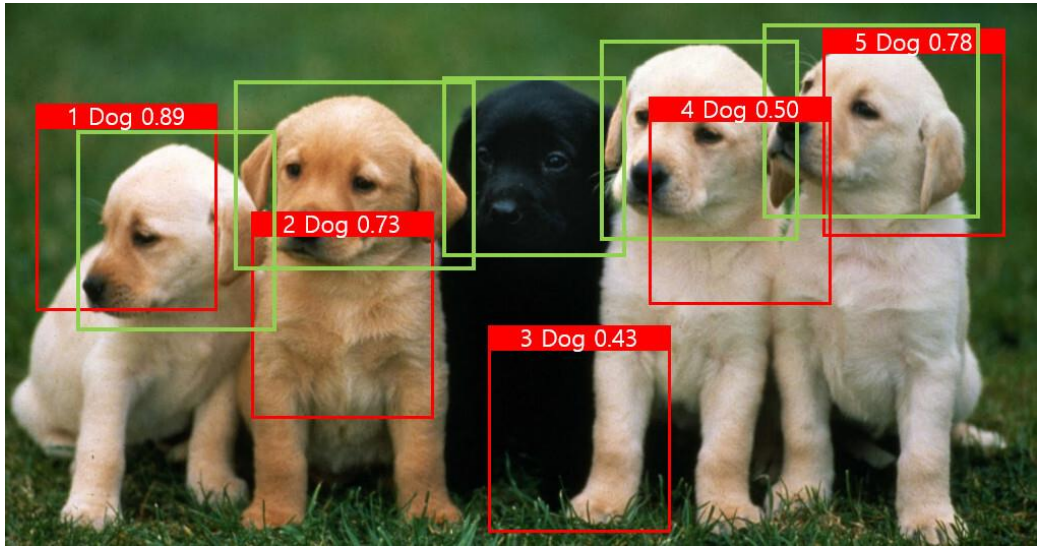
→ IoU에 따른 PR-Curve를 그리고 이를 바탕으로 AP, mAP 평가 metric를 정의한다

*AP : PR-Curve 의 밑 면적 넓이
*mAP : 여러 객체에 대한 AP의 평균치



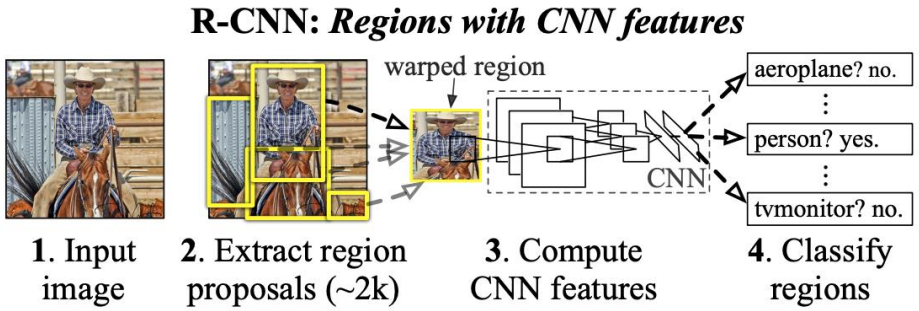
TP = 3 (1,4,5 Bounding Box)
FN = 1 (3번 째 강아지)
FP = 2 (2,3 Bounding Box)

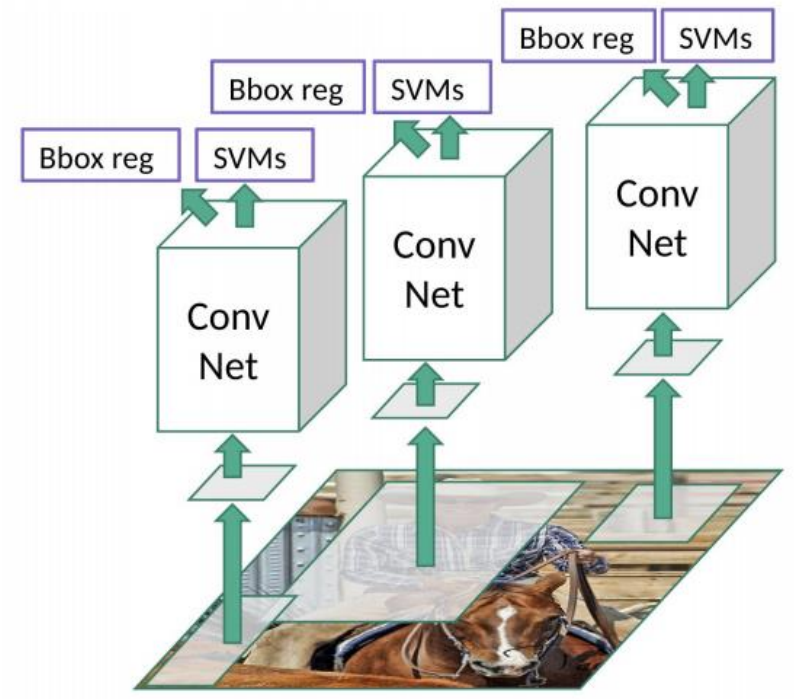
→ Precision = TP / (TP + FP) = 0.6
→ Recall = TP / (TP + FN) = 0.75
+ IoU Threshold에 따라 변화한다



Process of R-CNN

Overall Process





R-CNN
- 2 stage Detector (Localization → Classification)

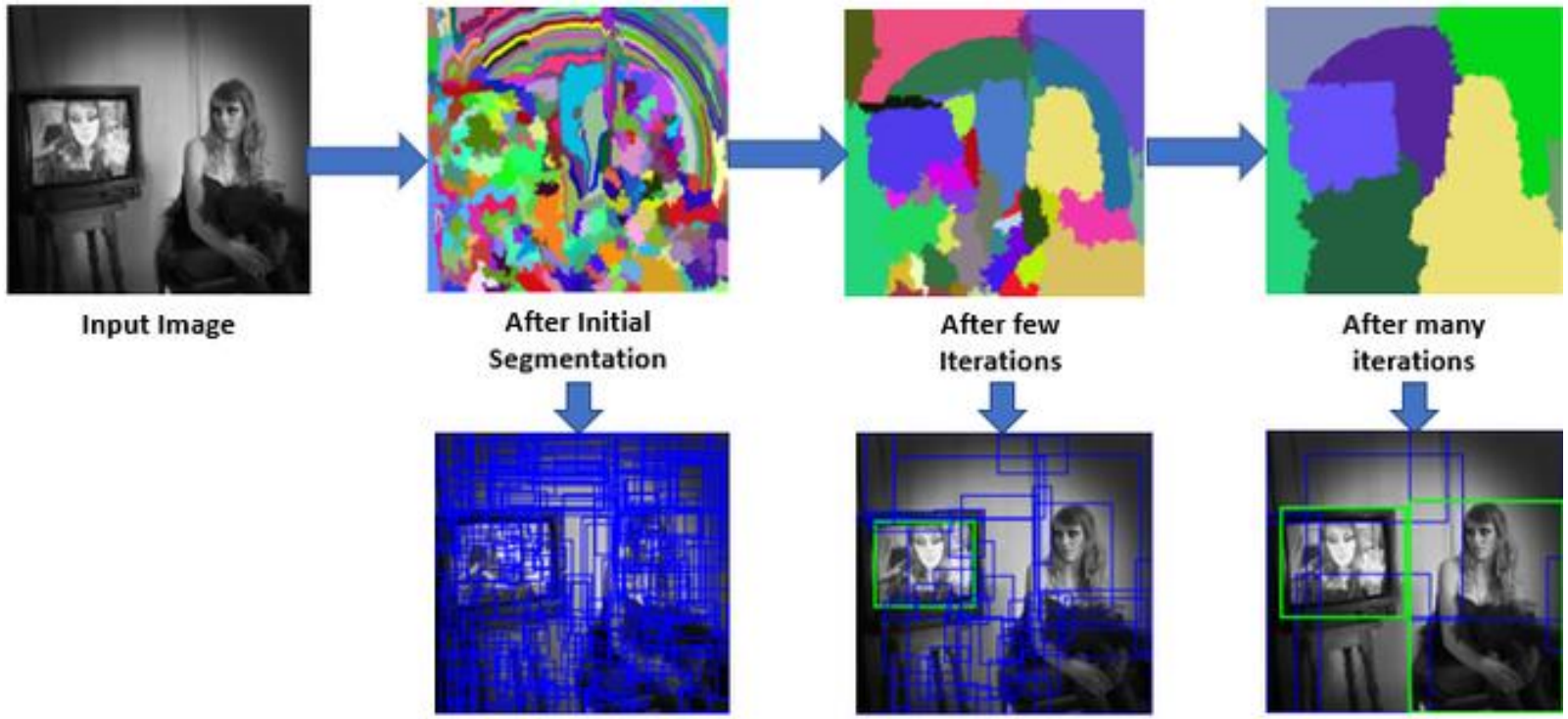
1. Selective Search를 통해 후보 영역 (2000개)를 추출한다. 이 후, 227 x 227로 이미지를 Warp한다
2. 이미지를 Fine tunning 된 Alexnet에 넣어 4096 사이즈의 feature vector를 추출한다
3. Feature vector를 각각 linear SVM / Bounding Box regressor에 넣어 Confidence와 Bounding Box의 좌표를 반환한다
4. Non Maximum Suppression을 통해 최적 Bounding Box를 생성한다

R-CNN – Selective Search

What is selective search

Selective Search

- Object 인식이나 검출을 위한 가능한 후보 영역을 제공하는 것
- 인식하고 싶은 객체에 대한 information(size)를 모르는데 fixed window를 사용하는 것이 맞는가에 대해서 착안해 만들어진 알고리즘 → 이미지를 먼저 segment하고 이를 바탕으로 window 후보 생성

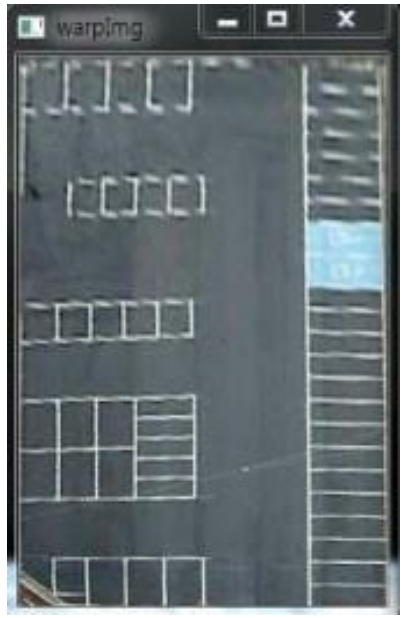


R-CNN – Image Warpping

What is image warpping

기하학적 변형의 한 종류로써 (x,y) → (x*,y*)로 대응시키는 알고리즘
이미지 사이즈 조절, 이미지 관측 시점 조절 등 이미지 가공을 위해 다양하게 사용된다

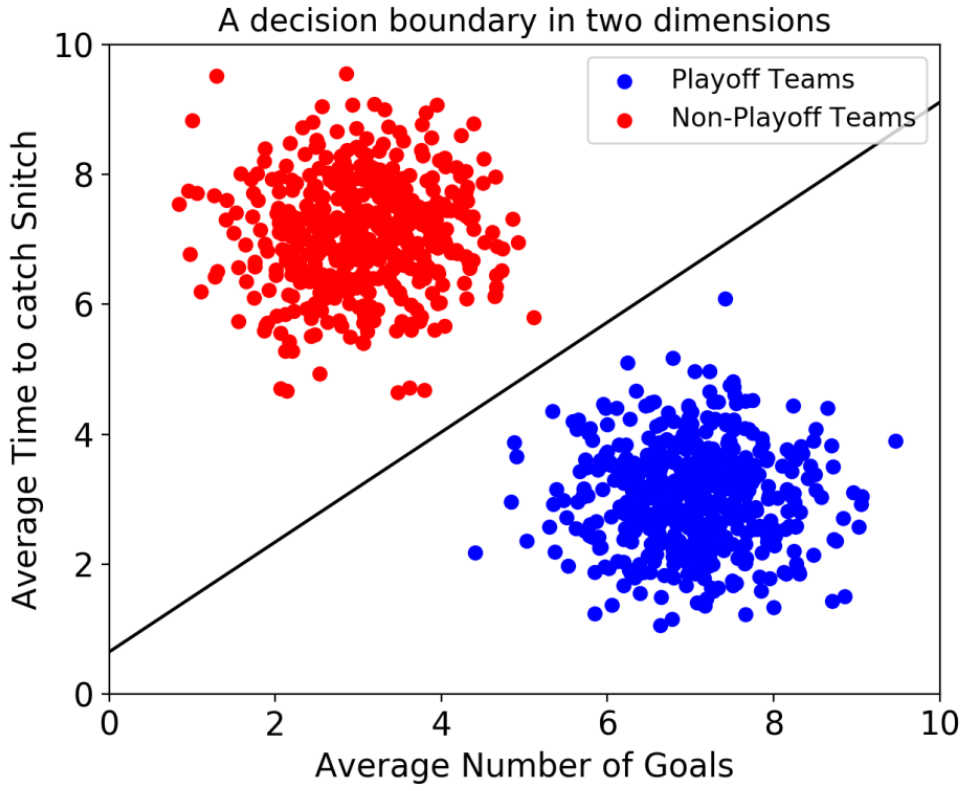
Image Warpping

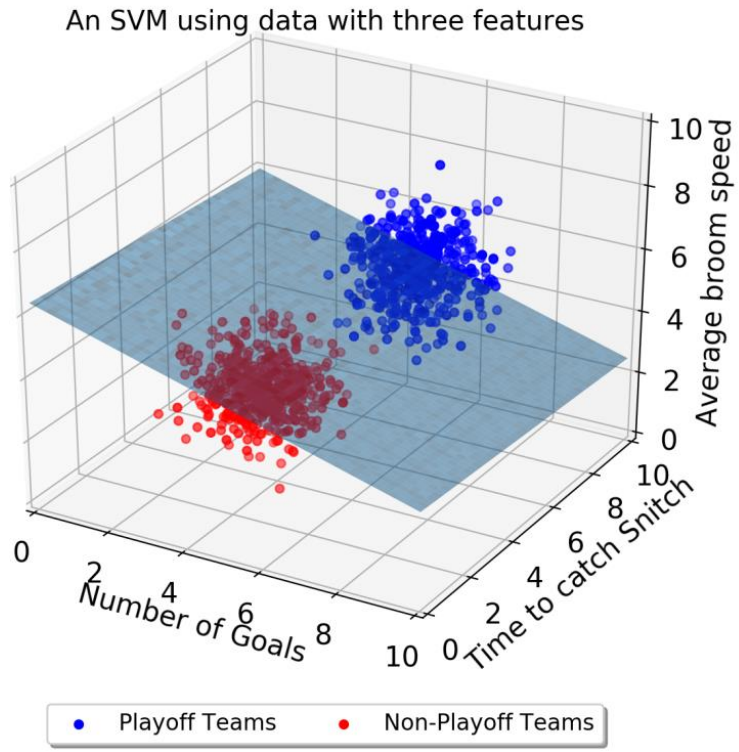


R-CNN – LinearSVM

Linear Support Vector Machine

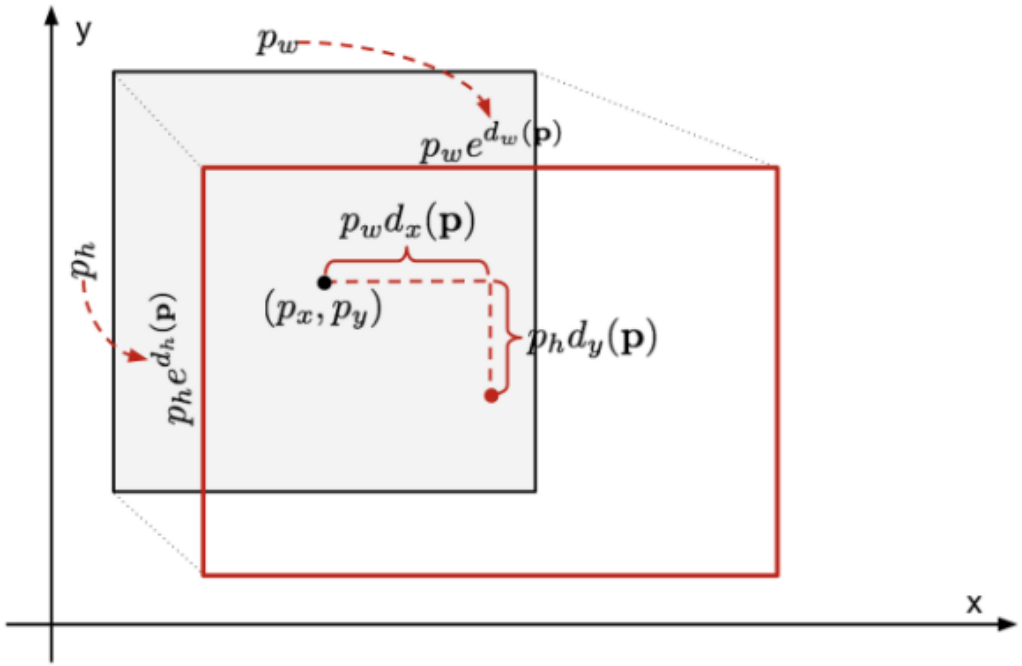
결정 경계 (Decision Boundary)를 생성하는 알고리즘 (분류)
보통 분류를 위해서 Softmax를 사용하지만 R-CNN의 경우 LinearSVM이 더 효과적이었다고 함





R-CNN – Bounding Box regressor

Adjust Bound Box



Bounding box regressor

$$\hat{g}_x = p_w d_x(\mathbf{p}) + p_x$$
$$\hat{g}_y = p_h d_y(\mathbf{p}) + p_y$$
$$\hat{g}_w = p_w \exp(d_w(\mathbf{p}))$$
$$\hat{g}_h = p_h \exp(d_h(\mathbf{p}))$$

$$t_x = (g_x - p_x)/p_w$$
$$t_y = (g_y - p_y)/p_h$$
$$t_w = \log(g_w/p_w)$$
$$t_h = \log(g_h/p_h)$$

$$\mathcal{L}_{\text{reg}} = \sum_{i \in \{x,y,w,h\}} (t_i - d_i(\mathbf{p}))^2 + \lambda \|\mathbf{w}\|^2$$

예측된 Box와 Ground Truth Box의 차이를 최소화

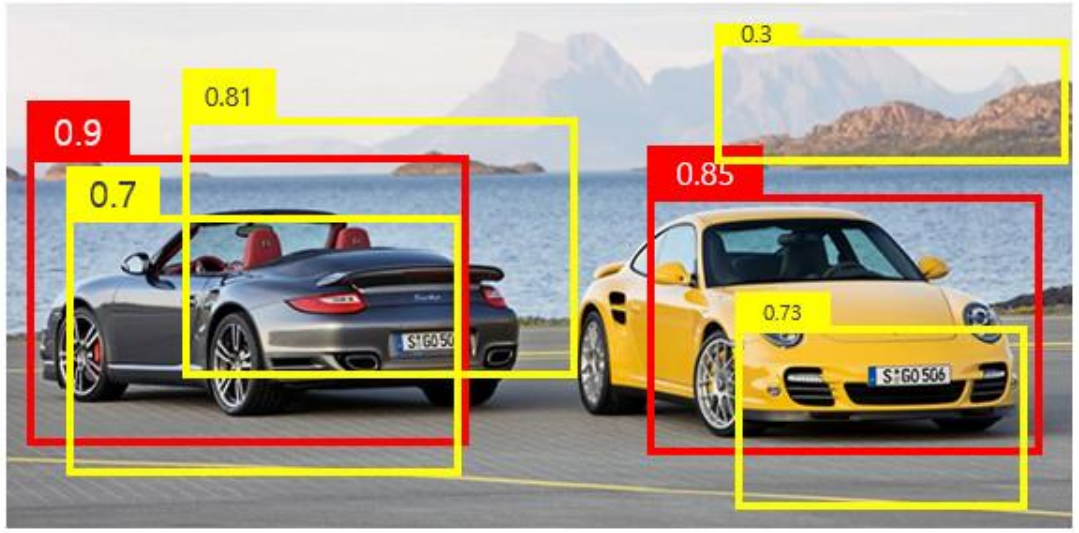
R-CNN – Non Maximum Suppression

For optimal bounding box

이 전 단계에서 얻은 2000개의 Bounding box를 표시할 경우 하나의 객체에 대해 지나지게 많은 bounding box가 겹침 , 정확도 하락

→ Non Maximum Suppression 알고리즘으로 optimal bounding box만 남기고 나머지 제거

1. bounding box별로 지정한 confidence score threhold 이하의 box를 제거합니다.
2. 남은 bounding box를 confidence score에 따라 내침차순으로 정렬합니다. 그 다음 confidence score가 높은 순의 bounding box부터 다른 box와의 IoU 값을 조사하여 IoU threshold 이상인 box를 모두 제거합니다.
3. 남아있는 box만 선택

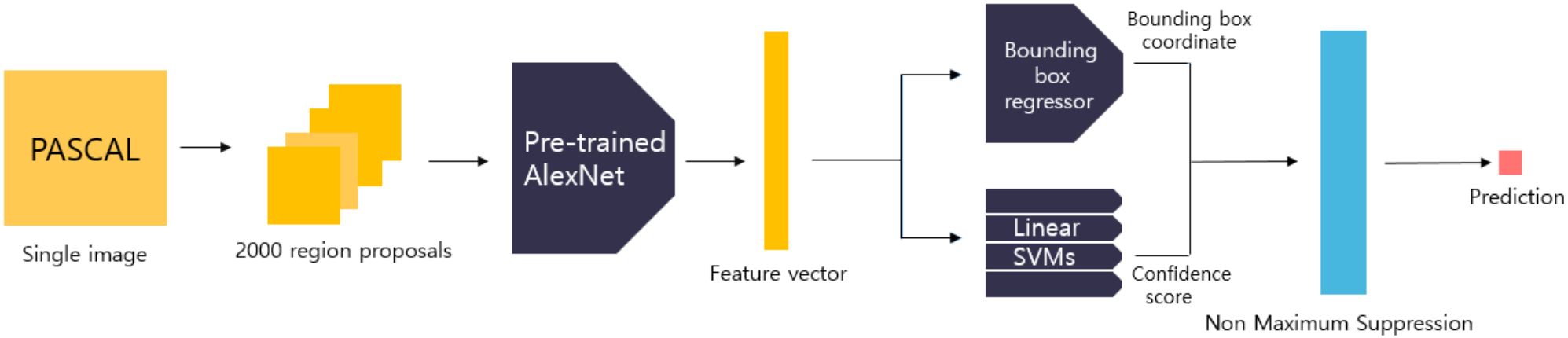


Before Non Maximum Suppression

After Non Maximum Suppression

R-CNN

Total Process



R-CNN 2013.11 → **OverFeat** ICLR' 14 → MultiBox CVPR' 14 → SPP-Net ECCV' 14 → MR-CNN ICCV' 15 → DeepBox ICCV' 15 → AttentionNet ICCV' 15 →

Fast R-CNN ICCV' 15 → DeepProposal ICCV' 15 → **Faster R-CNN** NIPS' 15 → **OHEM** CVPR' 16 → **YOLO v1** CVPR' 16 → G-CNN CVPR' 16 → AZNet CVPR' 16 →

Inside-OutsideNet(ION) CVPR' 16 → HyperNet CVPR' 16 → CRAFT CVPR' 16 → MultiPathNet(MPN) BMVC' 16 → **SSD** ECCV' 16 → GBDNet ECCV' 16 →

CPF ECCV' 16 → MS-CNN ECCV' 16 → **R-FCN** NIPS' 16 → PVANET NIPSW' 16 → DeepID-Net PAMI' 16 → NoC TPAMI' 16 → DSSD arXiv' 17 → TDM CVPR' 17 → **YOLO v2** CVPR' 17 →

Feature Pyramid Net(**FPN**) CVPR' 17 → RON CVPR' 17 → DCN ICCV' 17 → DeNet ICCV' 17 → CoupleNet ICCV' 17 → **RetinaNet** ICCV' 17 → DSOD ICCV' 17 →

Mask R-CNN ICCV' 17 → SMN ICCV' 17 → **YOLO v3** arXiv' 18 → SIN CVPR' 18 → STDN CVPR' 18 → **RefineDet** CVPR' 18 → MLKP CVPR' 18 → Relation-Net CVPR' 18 →

Cascade R-CNN CVPR' 18 → RFBNet ECCV' 18 → CornerNet ECCV' 18 → PFPNet ECCV' 18 → Pelee NIPS' 18 → HKRM NIPS' 18 → R-DAD AAAI' 19 → **M2Det** AAAI' 19 ···