

R-CNN Review

[참고 블로그 링크1](#) [참고 블로그 링크2](#)

1. Introduction

1. classification
2. Object Detection
3. Image Segmentation
4. Visual Relationship

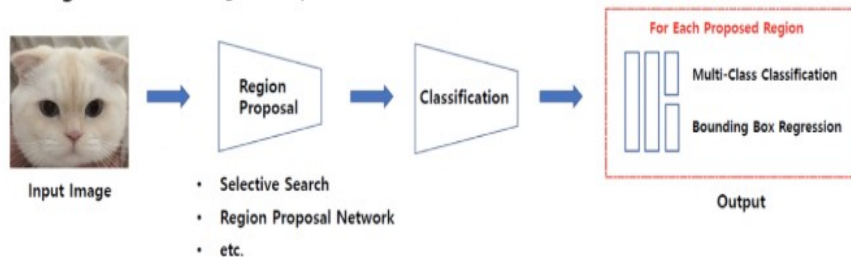
CV(Computer Vision)은 위에 4가지로 분류가 된다.

- Classification : Single object에 대해서 object의 클래스를 분류하는 문제
- Classification + Localization : Single Object에 대해서 Object의 위치를 Bounding Box로 찾고 (Localization) + 클래스를 분류하는 것(Classification)
- Object Detection : Multiple objects에서 각각의 object에 대해서 Classification + Localization을 수행하는 것
- Instance Segmentation : Object Detection과 유사하지만, 다른점은 object의 위치를 bounding box가 아닌 실제 edge로 찾는 것

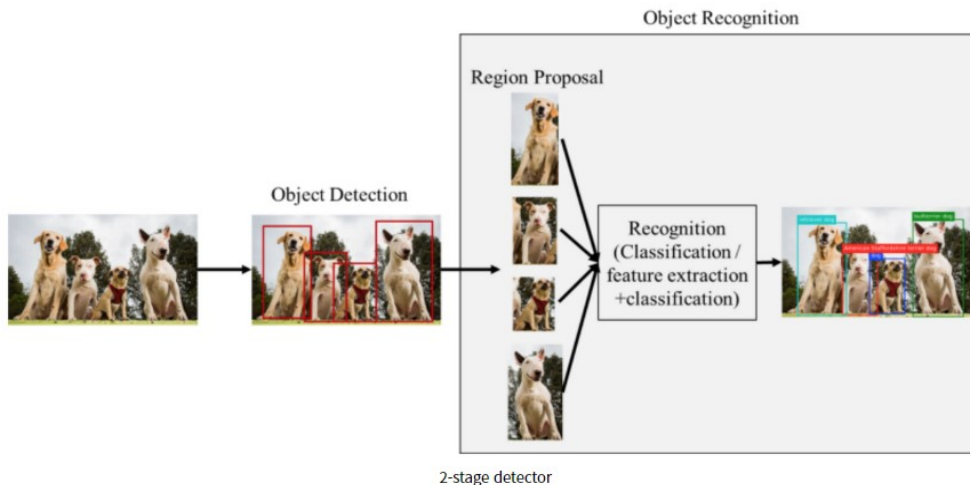
1-1). 2-Stage-Detector

따라서 2-stage detector에서 regional proposal과 classification이 순차적으로 이루어진다는 것은 classification과 localization문제를 순차적으로 해결한다는 것이다.

2-Stage Detector - Regional Proposal와 Classification이 순차적으로 이루어짐.



R-CNN계열 (R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN ...)



2-stage detector

- R-CNN계열(R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN etc.) - 2-Stage Detector는 Region Proposal network등과 같은 알고리즘 및 네트워크를 통해 객체가 있을 영역(Region of Interest, RoI)을 추출해낸다. 이후 Classification과 Localization을 진행한다.

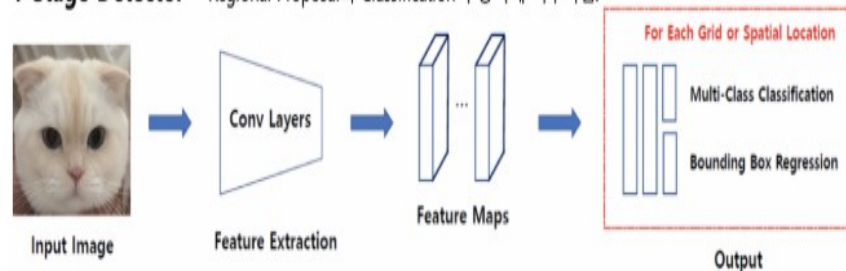
1-2). 1-Stage-Detector

<1-Stage Detector>

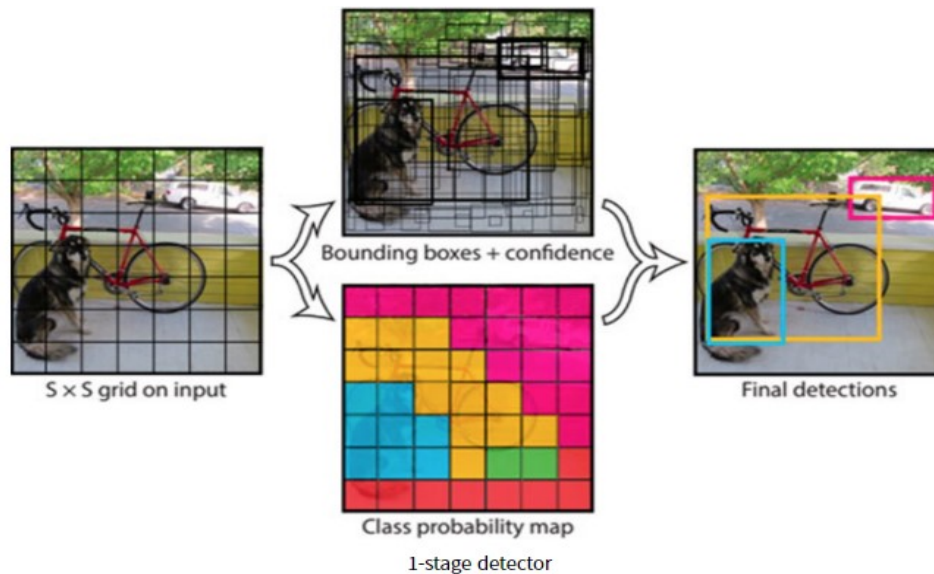
2-stage detector와 반대로 regional proposal과 classification이 동시에 이루어진다.

즉, classification과 localization문제를 동시에 해결하는 방법이다.

1-Stage Detector - Regional Proposal와 Classification이 동시에 이루어짐.



YOLO 계열, SSD 계열 (SSD, RetinaNet, RefineDet ...)



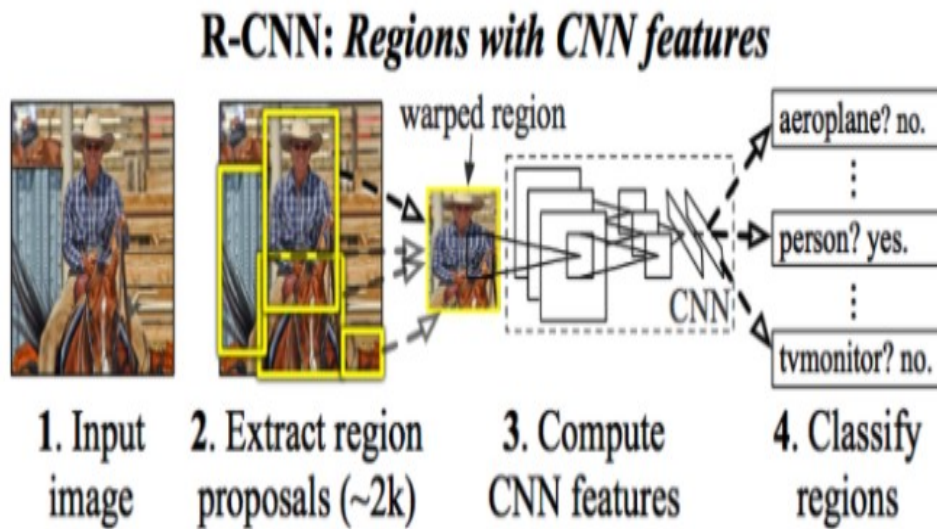
- YOLO, SSD계열(SSD, RetinaNet, RefineDet etc.) - 1-Stage Detector는 RoI를 추출하지 않고 전체 이미지에 대해 Convolution Network로 Classification과 Localization을 진행한다.

[참고 사진1](#) [참고 사진2](#)

1-3). 2-Stage-Detector와 1-Stage-Detector의 장단점

- 2-Stage-Detector장점 & 단점
 - 속도가 빠르지만 정확도가 떨어진다.
- 1-Stage-Detector장점 % 단점
 - 정확도가 높지만 속도가 느리다.

2. R-CNN



참고 사진2

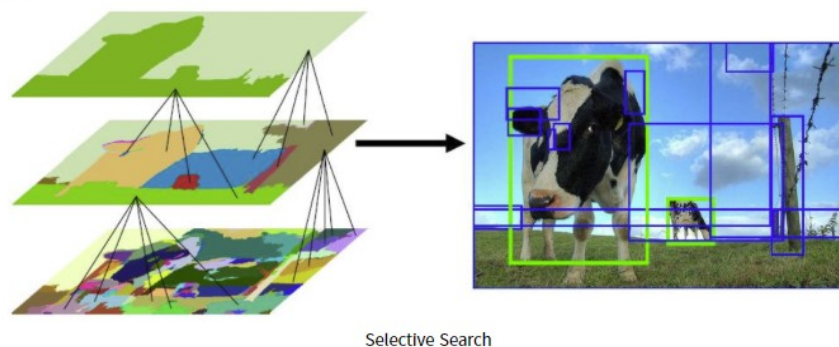
- R-CNN은 위에서 보았던, 2-Stage-Detector라고 볼 수 있다.
- R-CNN의 진행과정은 다음과 같다.
 1. 이미지 입력
 2. Selective search알고리즘을 사용하여 대략 2,000개의 regional proposal output추출 후 동일한 input size를 만들기 위해 warp해준다.(마지막 FC layer에서 input size는 고정이기 때문에 Convolution Layer에 대한 output size도 동일해야 해서 input size에서 부터 동일한 사이즈를 넣어 주어야 한다.)
 3. 대략 2,000개의 warped image를 CNN모델에 투입
 4. 각각의 Convolution 결과를 기반으로 Classification(SVM)을 진행 후 결과 추출

2-1). Region Proposal

- Sliding window(이미지의 맨 위에서 부터 쪽 읽으면서 탐색)은 탐색이 너무 느리고 비효율적이어서 Selective search알고리즘을 사용하게 됨

Selective search

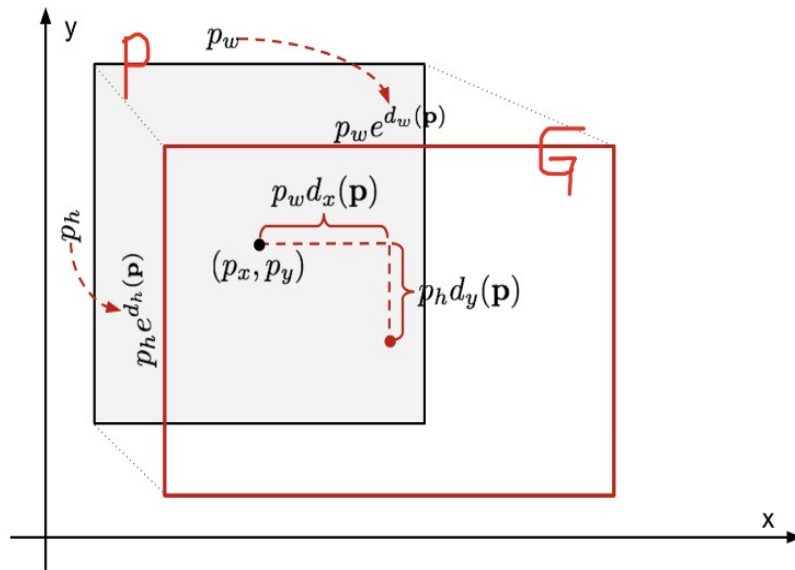
Selective search



- Selective search의 진행과정
 1. 색상, 질감, 영역크기 등을 이용하여 Non-object-based segmentation수행

2-4). Bounding Box Regression

- Bounding Box Regression은 처음에 y로 줬던 레이블과, CNN을 통과해서 나온 Bounding Box, 이 두 가지의 차이를 구해서 차이를 줄이도록 조정하는 선형회귀 모델 절차



참고 사진5

- 위 그림에서 P가 처음에 region proposal을 통해서 제안된 Bounding Box이고, G는 Ground Truth Bounding Box(정답 박스) 여기서 P를 정답 G에 맞추도록 transform 하는 것을 학습하는 것이 바로 Bounding Box Regression의 목표

3. 정리

- 이미지 하나의 임의의 영역들로 잘게 잘라서 merge 하고 warping 한 후,
- CNN의 input으로 통과하여 Feature vecture를 추출하고,
- 적용된 label들을 이용해 SVM 분류와 Bounding Box Regression을 수행해서 영역 위치까지 예측을 하게 되는 모델이 바로 R-CNN