**region proposal 기반의 object detection 모델**

stage 1 selective search를 통해 이미지에서 2000개의 region을 proposal함

stage 2 CNN detection을 수행함 feature extrator(AlexNet) -> feature Map -> FC layer -> SVM classifier

AlexNet의 이미지 크기는 277 \* 277 이어야 해서 proposal된 이미지를 crop과 warp를 적용시켜야 함

**Bounding Box Regression과 RCNN 장단점**

selective search는 있을 만한 곳을 찾아 주는 것이기 때문에 후에 bouding box regression을 다시 수행함

장점: 동시대의 다른 알고리즘과 비교해서 detection 정확도가 매우 높음

단점: detection시간이 너무 느리고 아키텍쳐 및 학습 프로세스가 복잡함

- 하나의 이미지마다 selective search를 수행하여 2000개의 region 영역 이미지를 검출함

- 2000개의 영역 이미지마다 CNN feature map을 생성함

- 구성요소가 selective search, CNN feature map, SVM, bounding box regressor로 구성되어 복잡함

RCNN을 통해 object detection 성능을 입증

region proposal 기반 성능 입증

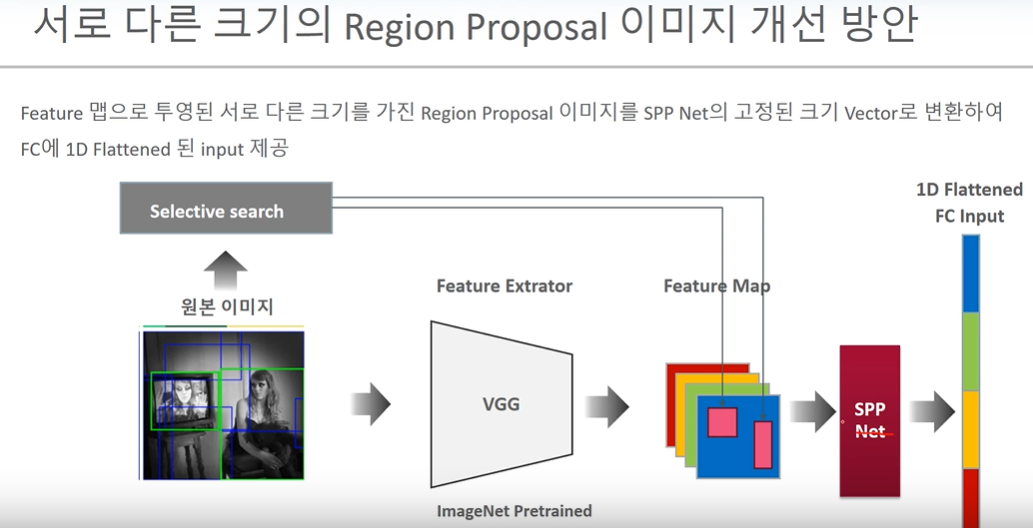
detection 수행 시간을 줄이고 분리된 개별 아키텍처를 통합할 수 있는 방안을 연구하기 시작

**SPPNet(spatial pyramid pooling)의 이해**

2000개의 region image가 CNN으로 입력되어 수행시간이 오래걸리고 region image가 crop/warp되는 문제점을 해결하기 위해 나옴

2000개의 region proposal image를 CNN으로 입력하지 않고 원본 이미지만 CNN으로 feature map을 생성한 후 추천된 영역의 이미지만 fature map으로 매칭하여 추출한다

하지만 이미지의 크기가 다르다는 문제가 있는데 이를 SPP를 통해 해결한다

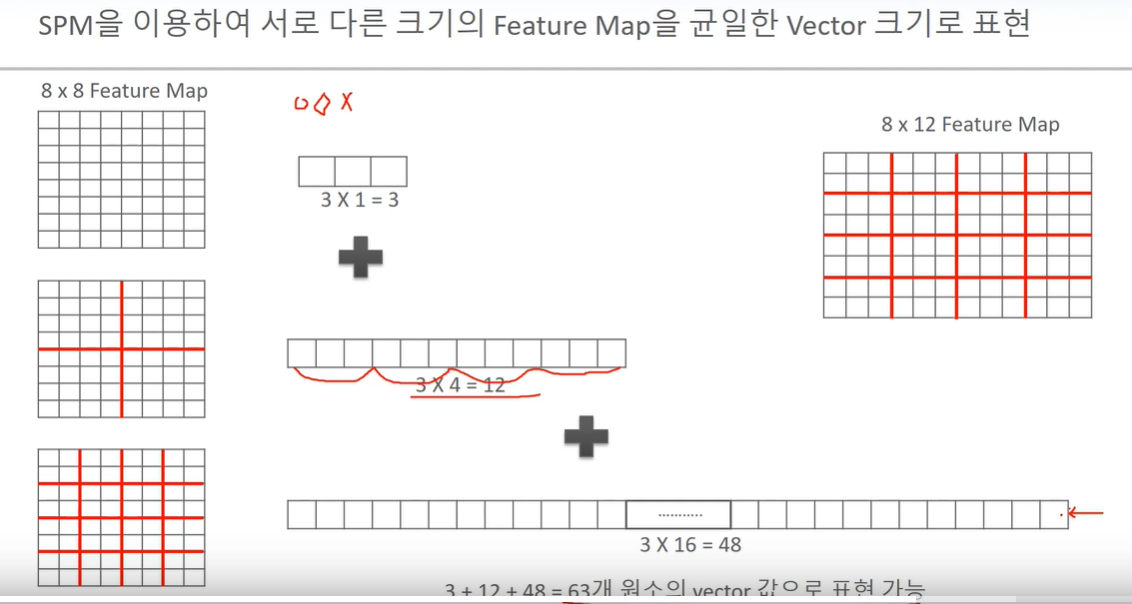


어떠한 크기의 이미지가 들어오더라도 spp를 거치면 고정된 크기로 변환을 시킴

bag of visual words

이미지의 feature를 추출해 각 feature의 빈도수를 통해 이미지를 분류함

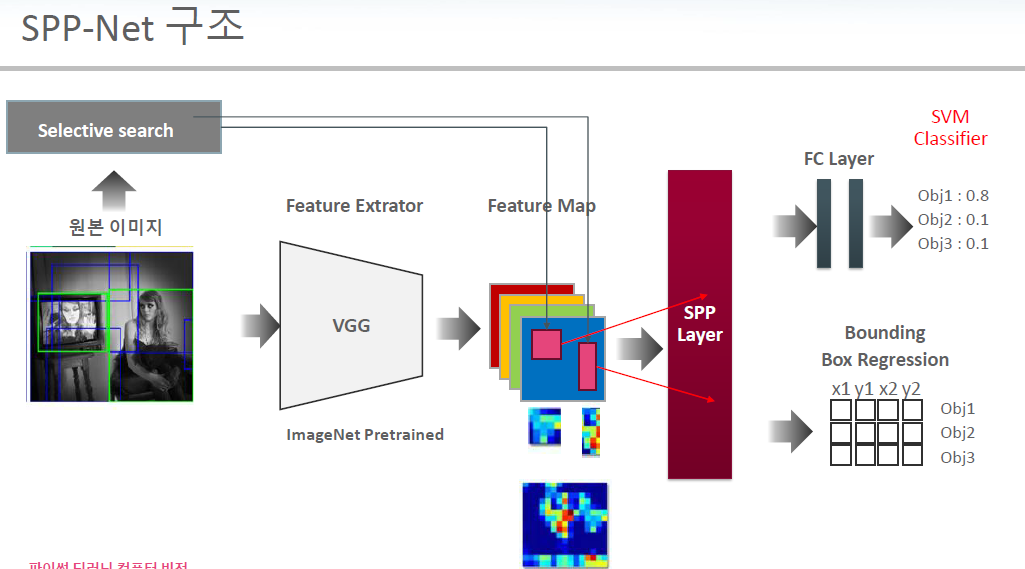
spm(spatial pyramid matching)을 통해 이미지를 여러 개로 나누어 각 이미지들 마다 feature들의 히스토그램을 추출해 균일한 vector크기로 표현한다



다음과 같이 크기가 다른 이미지더라도 이미지를 나누는 방법만 같다면 vector의 원소의 개수를 같게 만들수 있음

spatial pyraimd pooling은 이미지의 분면을 나누어 maxpooling을 적용하기 때문에 이미지의 크기와 상관없이 고정된 벡터값으로 나타낼 수 있다

selectvie search를 통해 propose된 이미지를 feature map에 투영할 때 이미지의 크기가 다르지면 SPP layer를 거쳐 고정된 크기의 flattened FC input을 만들수 있다.



**Fast RCNN의 이해**

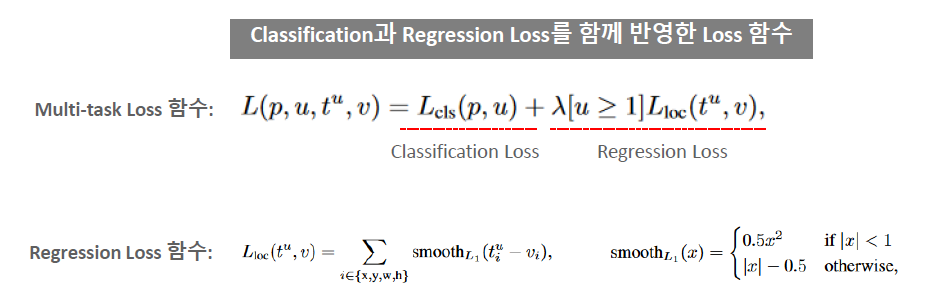
주요 특징

SPP layer를 ROI Pooling layer로 바꿈

- 균일한 크기의 vector로 표현할 때 7 \* 7로 나누어 만듬

End – to – End network learning

- SVM을 softmax로 바꾸고, multi-task loss 함수로 classification 과 regression을 함께 최적화함



**Faster RCNN의 이해**

RPN(region proposal network)과 Anchor Box

faster RCNN = RPN + Fast RCNN

RPN은 CNN을 거친 Feature map에서 object가 있을 만한 곳을 neural network 구조를 통해propose하는 것임

RPN을 하기 위해 주어지는 feature는 pixel값, targetㅇㄴ ground truth bounding box인데 selective search 수준의 proposal을 위해서 anchor box가 도입됨

anchor box는 object가 있을법한 후보 box이고 총 9개의 서로다른 크기, ratio로 구성됨

anchor box는 feature map의 pixel마다 생성이 되어 object의 유무를 판단함