

# **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks**

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun

박태우

# Introduction

- Selective Search와 같은 region proposal 알고리즘들은 여전히 많은 시간이 소요됨.
- 게다가 Selective Search는 cpu를 사용하는 단점이 존재.
- 이러한 문제점을 개선하기 위해 region proposal을 NN에 넣는 방식을 제안함.
- 그래서 RPN( Region Proposal Network )를 사용하게 됨.
- RPN + Fast R-CNN의 형태를 가짐.

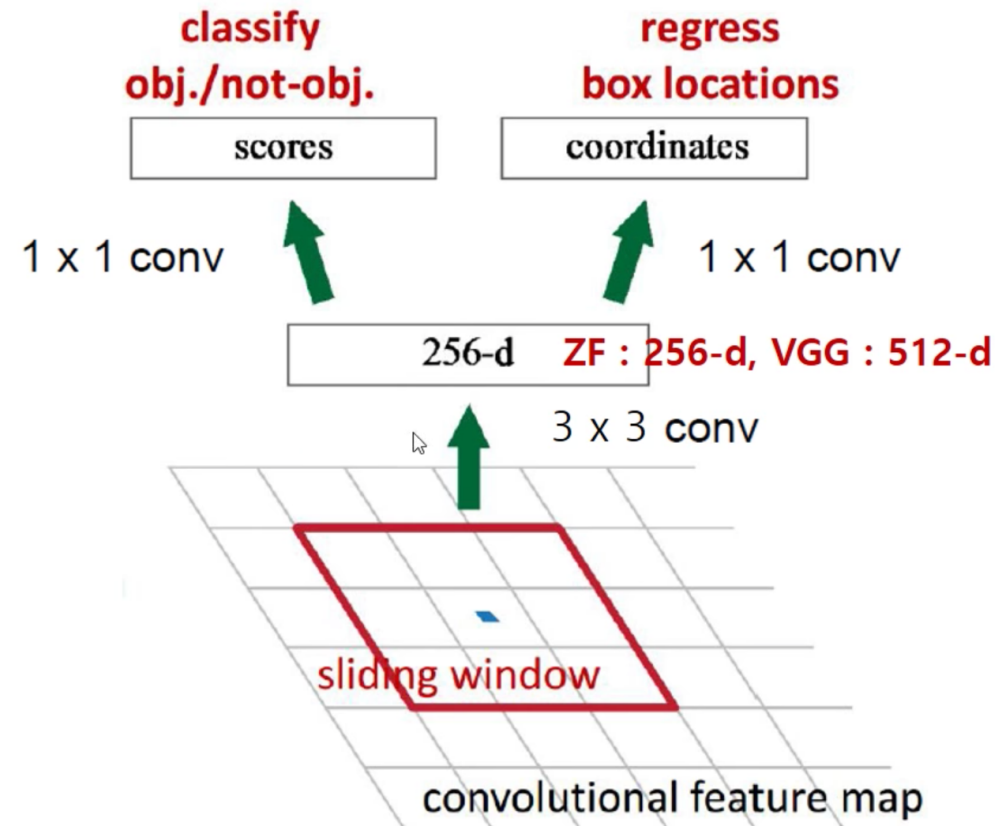
# RPN

- Feature map으로부터  $N \times N$  conv를 사용.

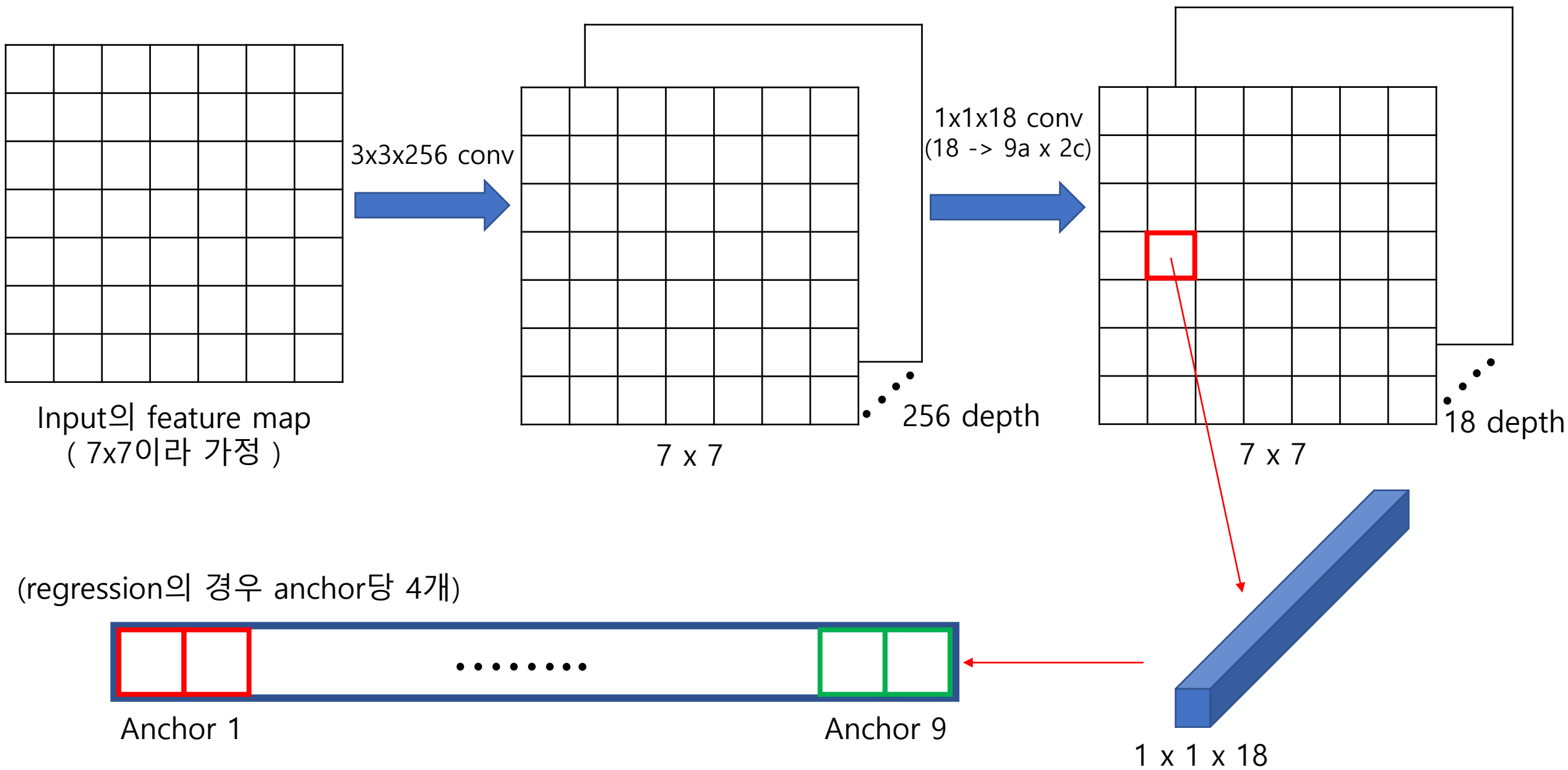
( 해당 논문에서는  $3 \times 3$  사용 )

- conv에 의해  $3 \times 3 \times 256$  feature가 나오고  $1 \times 1$  conv를 다시 한번 적용해 classification, regression 정보를 얻음.

- 주의할 점은 RPN에서 classification은 해당 지점이 object인지 아닌지 판별한다.

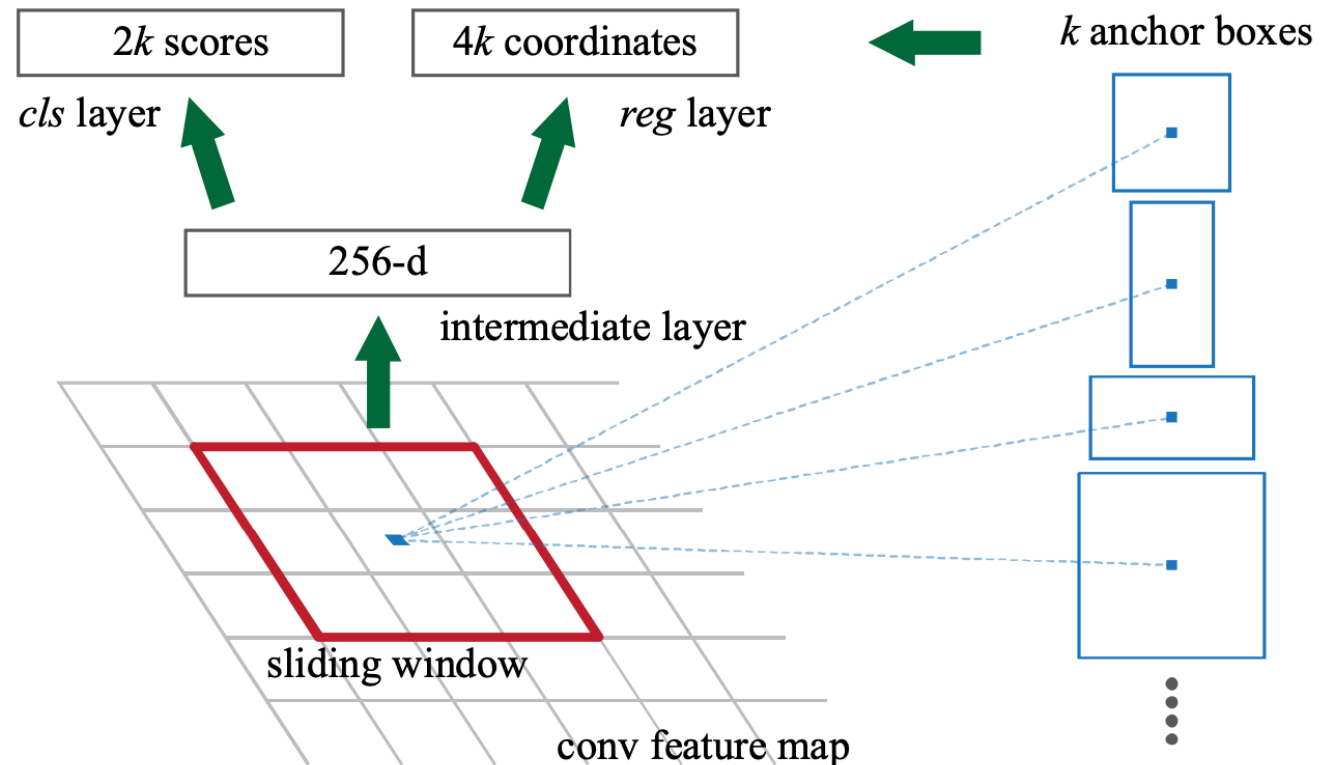


# RPN



# RPN – Anchor box

- Bbox regression 하기에 앞서 **anchor box**들을 먼저 위치시킴.
- Anchor box : 효율적인 학습을 하기 위해 첫 학습을 하기 전, 미리 정해놓은 Bbox.
- 해당 논문에서는 anchor box 의 수  $k$ 를 9로 정했음.
- Scale 3가지 ( 128x128, 256x256, 512x512 ), ratio 3가지 (1:1, 2:1, 1:2)로 총 9가지



## Positive label

1) gt와의 IOU가 가장 높은 anchor

2) gt와의 IOU가 0.7이상인 anchor

\*하나의 gt에 여러 anchor가 positive label로써 할당될 수 있음.

## Negative label

1) gt와의 IOU가 0.3 이하인 anchor

- positive, negative 이외의 anchor는 training에 참여하지 않는다.

- Mini-batch로써 256개의 anchor를 사용하고 positive, negative의 비율은 1:1로 함.

## Loss function

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) \\ + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*).$$

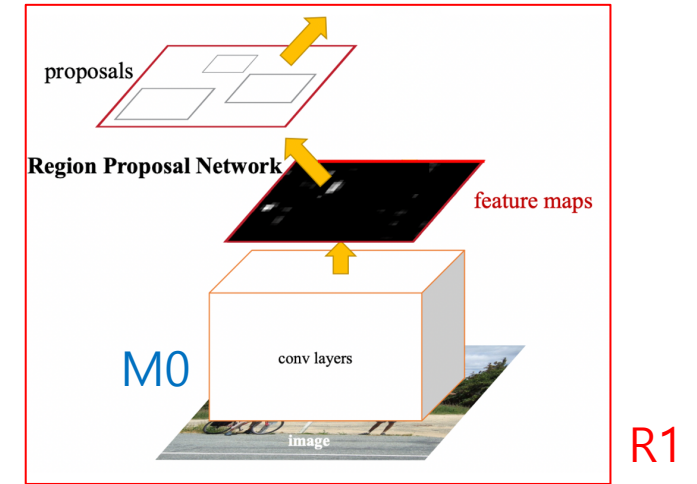
gt objectness label 물체가 있으면 1, 없으면 0.  
즉 물체가 없을 땐 regression loss는 0.

# Faster R-CNN Training

## 4-set training algorithm

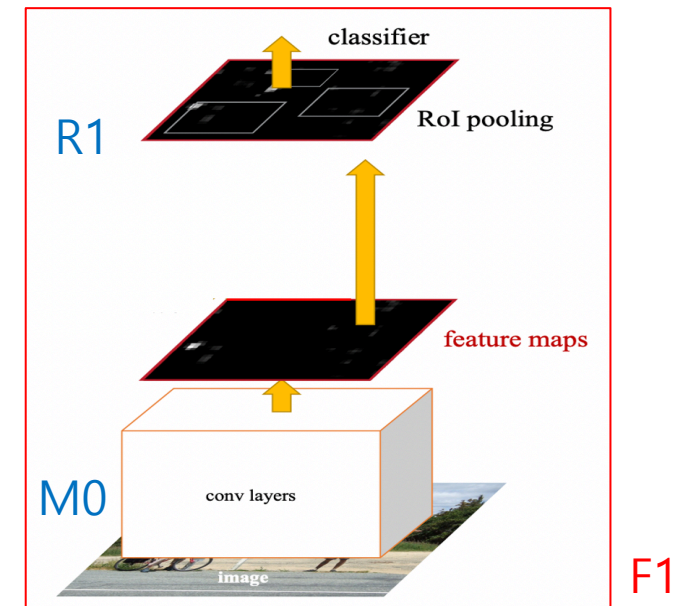
M0 : pretrained network based an ImageNet.

1) M0로 RPN을 학습시켜 RPN model R1을 얻음.



2) M0로 Fast R-CNN을 학습, region proposal은 R1을 이용.

( R1과 F1는 conv layers가 서로 다른 상태)

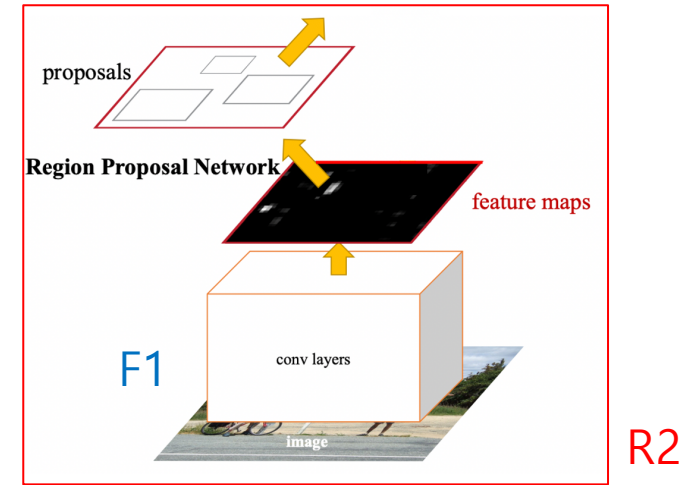




# Faster R-CNN Training

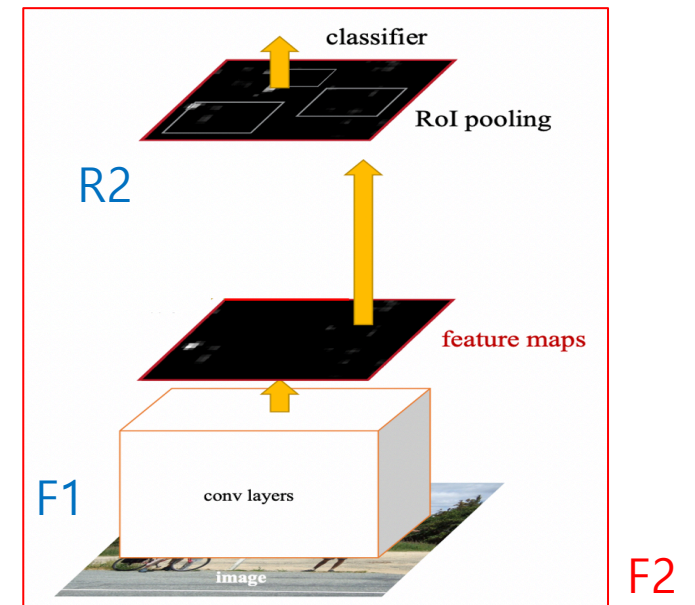
3) F1로 RPN을 다시 학습시켜 R2를 얻음. 이 때 conv layers는 학습시키지 않음.

( F1, R2는 conv layers가 서로 같은 상태 )



4) F1로 Fast R-CNN을 학습, region proposal은 R2를 이용.

이 때 conv layers는 학습시키지 않음.



# Result

	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image (with proposals)	50 seconds	2 seconds	<b>0.2 seconds</b>
(Speedup)	1x	25x	<b>250x</b>
mAP (VOC 2007)	66.0	<b>66.9</b>	<b>69.9</b>

model	system	conv	proposal	region-wise	total	rate
VGG	SS + Fast R-CNN	146	1510	174	1830	0.5 fps
VGG	RPN + Fast R-CNN	141	<b>10</b>	47	<b>198</b>	<b>5 fps</b>
ZF	RPN + Fast R-CNN	31	<b>3</b>	25	<b>59</b>	<b>17 fps</b>

## 참고 링크

<https://www.slideshare.net/JinwonLee9/pr12-faster-rcnn170528>

<https://89douner.tistory.com/91>

<https://yeomko.tistory.com/17>