



# SSD

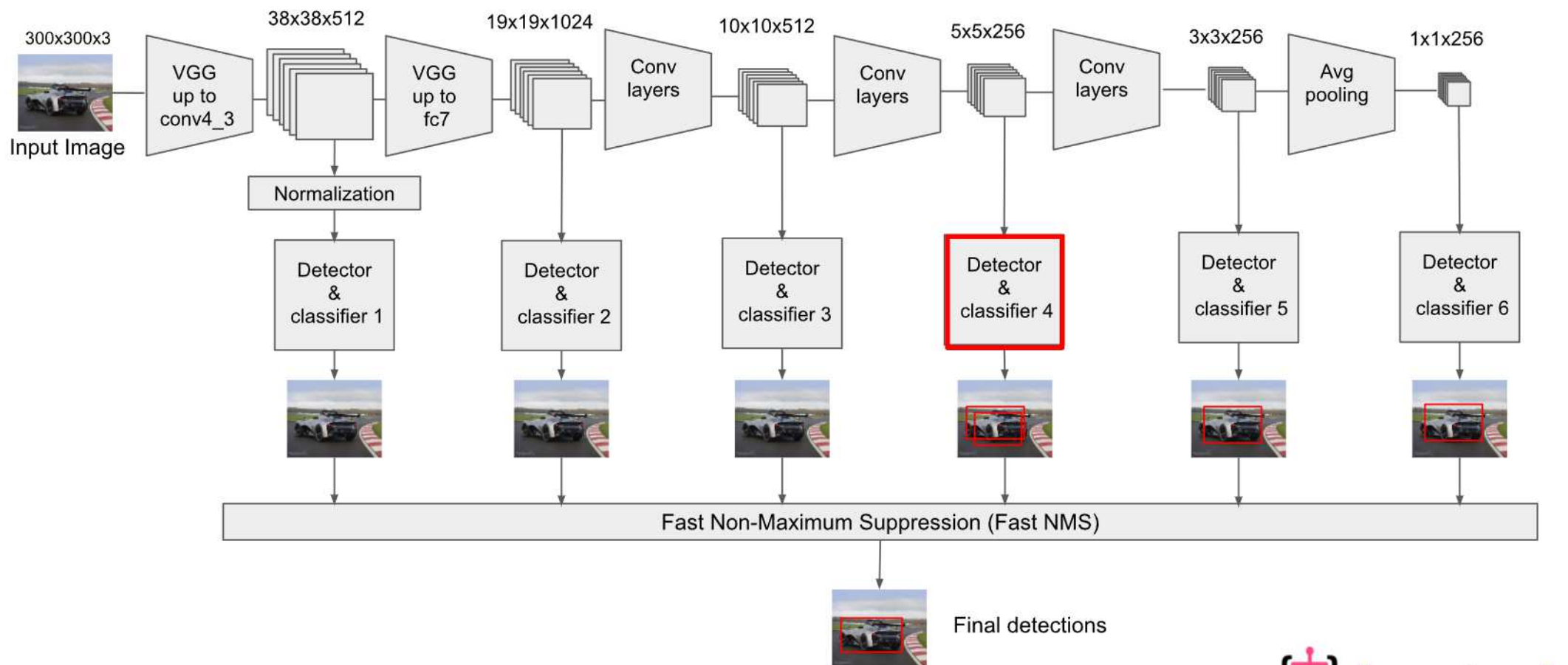
## Single Shot MultiBox Detector

박 철

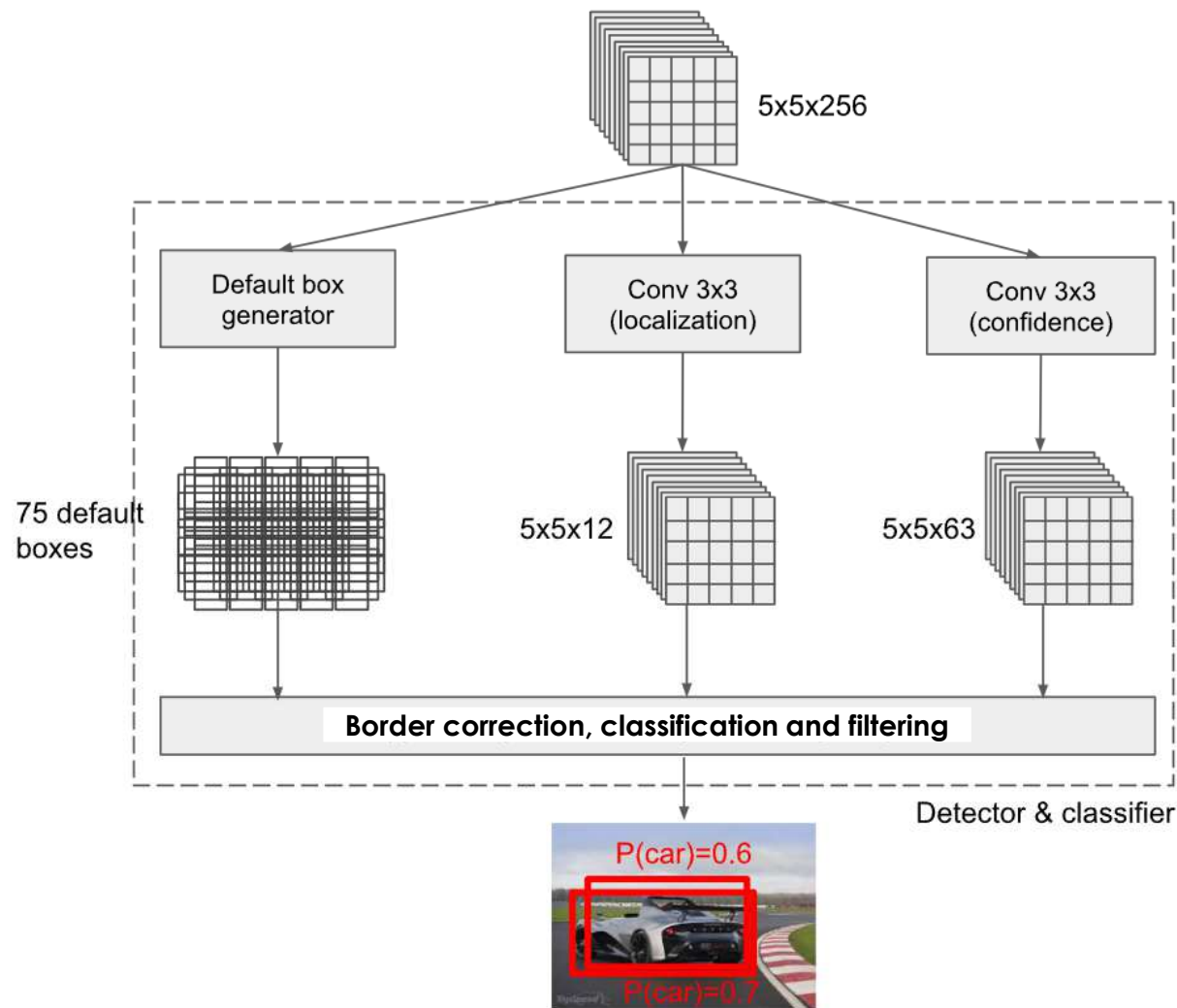
# SSD 300 architecture

base Network : VGG-16 conv4\_3까지 적용

multi feature maps : 38x38, 19x19, 10x10, 5x5, 3x3, 1x1 피쳐맵



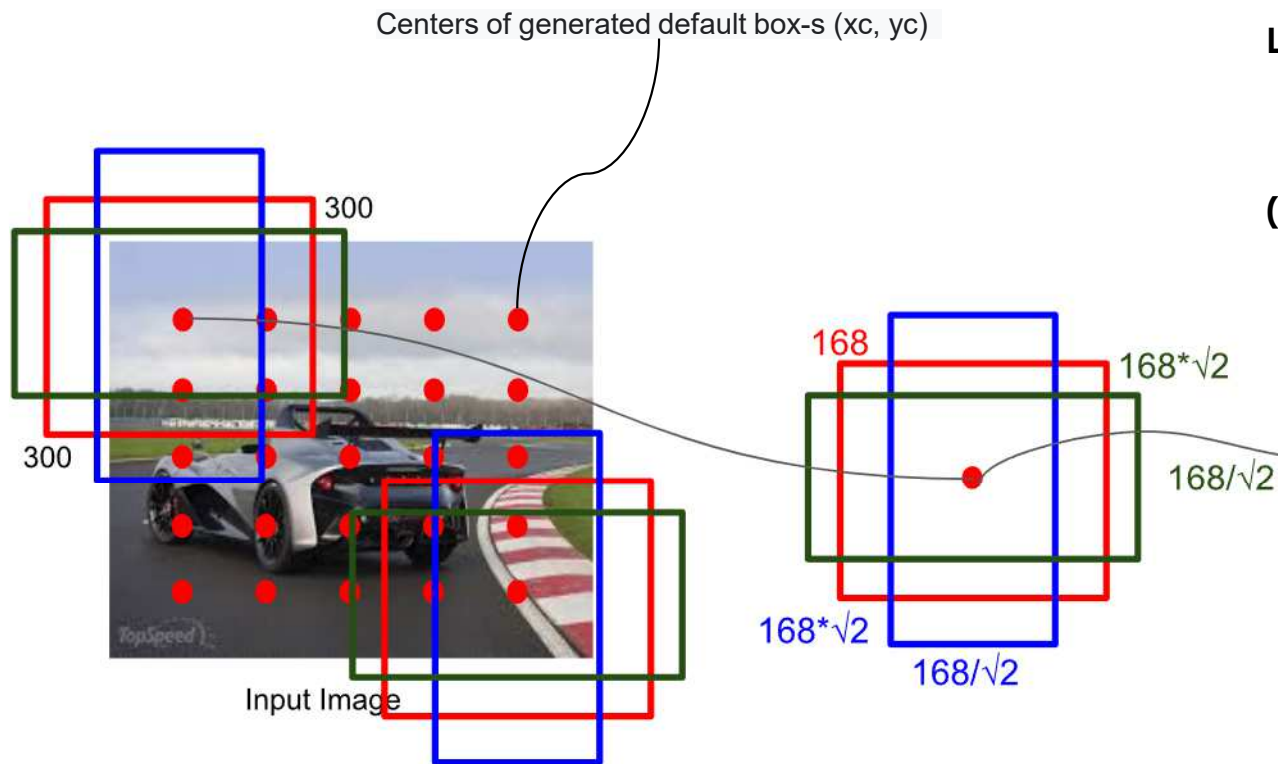
# SSD 300 Detector & classifier architecture



Let the following parameters be given:

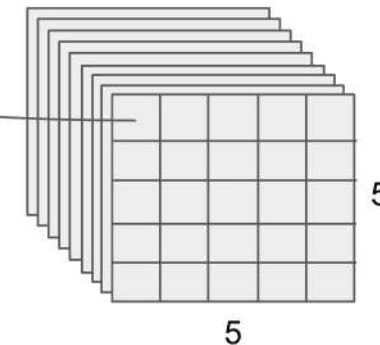
- The size of the original image (300 x 300)
- Feature maps dimension (5 x 5 x 256)
- #default boxes = 3

# Generating default boxes



Let the following parameters be given:

- The size of the original image (300 x 300)
  - Feature maps dimension (5 x 5 x 256)
  - #default boxes = 3
- (there are 3 rectangles per point in feature maps)
- min size = 168
  - aspect\_ratio = 2

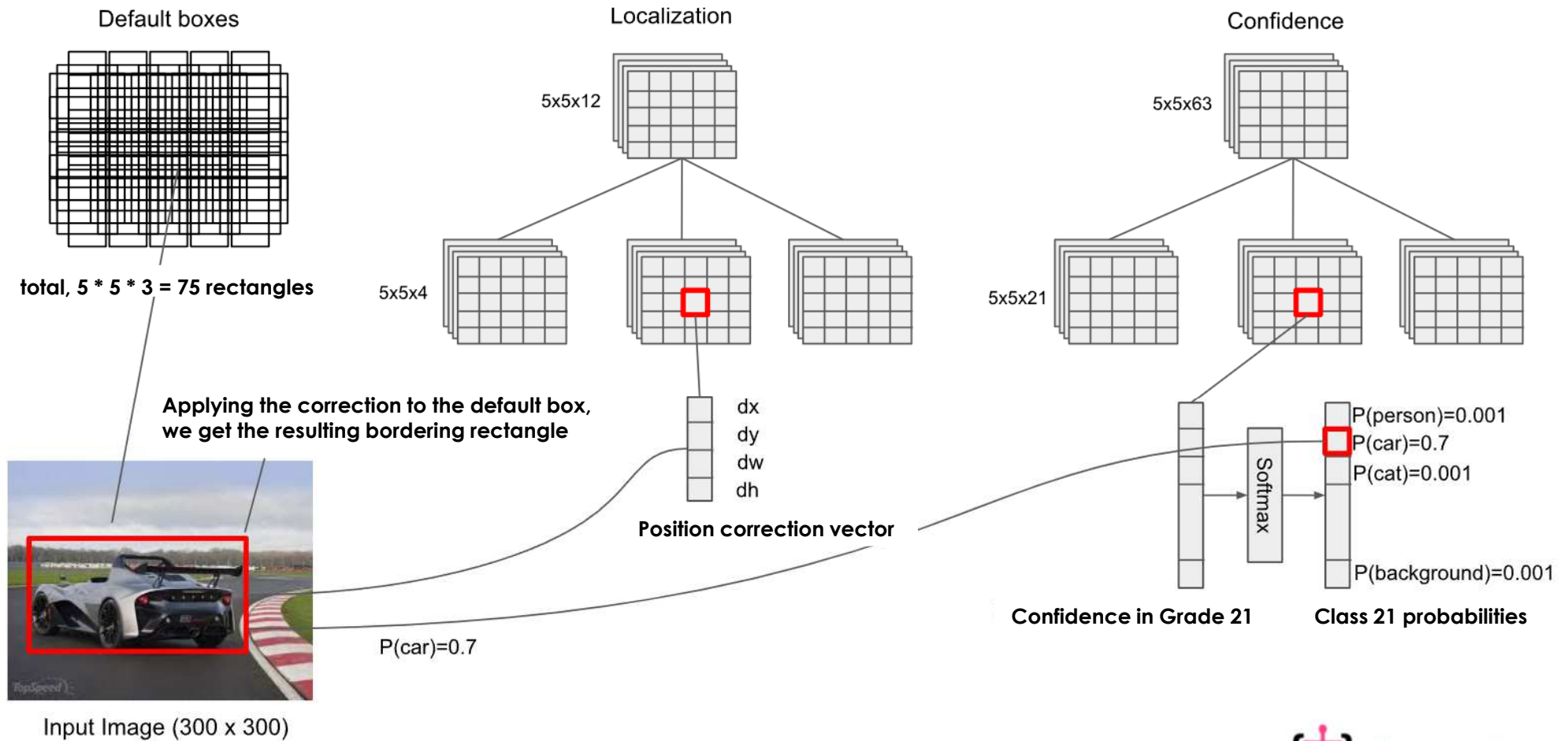


The default box is set with the following values:

- xc, center of the rectangle by x
- yc, - the center of the rectangle by y
- w - the width of the rectangle
- h - the height of the rectangle

In total,  $5 * 5 * 3 = 75$  rectangles will be generated

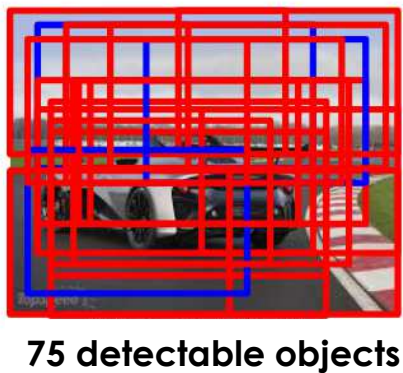
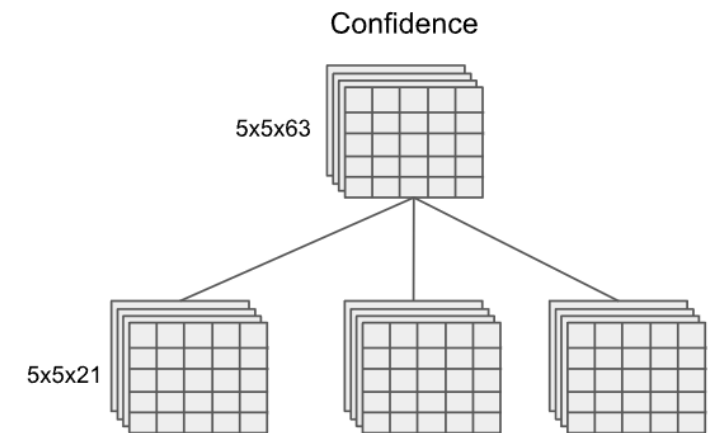
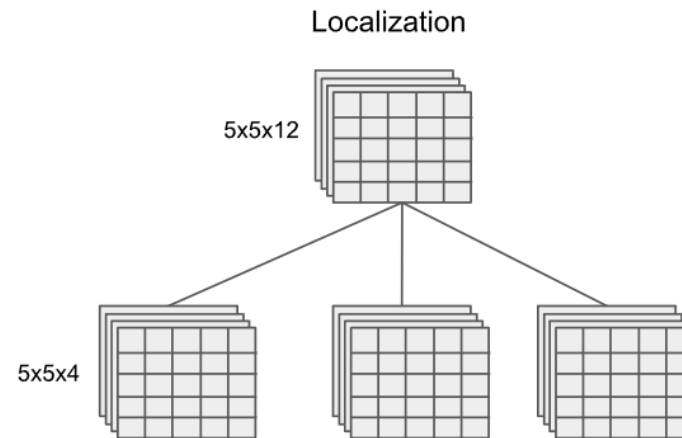
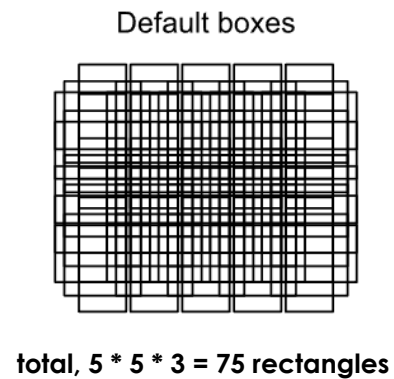
# Border correction, classification and filtering 1



Result : 75 children bound to the class



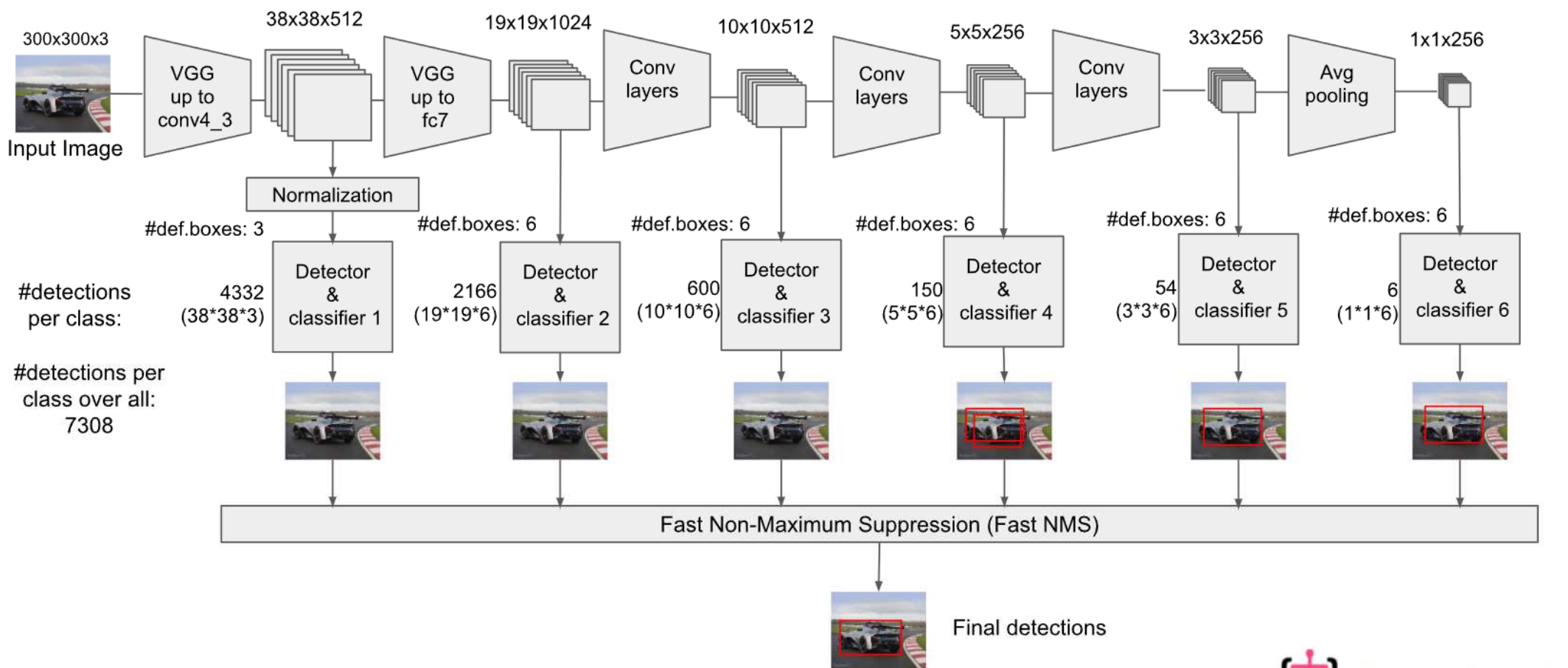
# Border correction, classification and filtering 2



Filtering by confidence



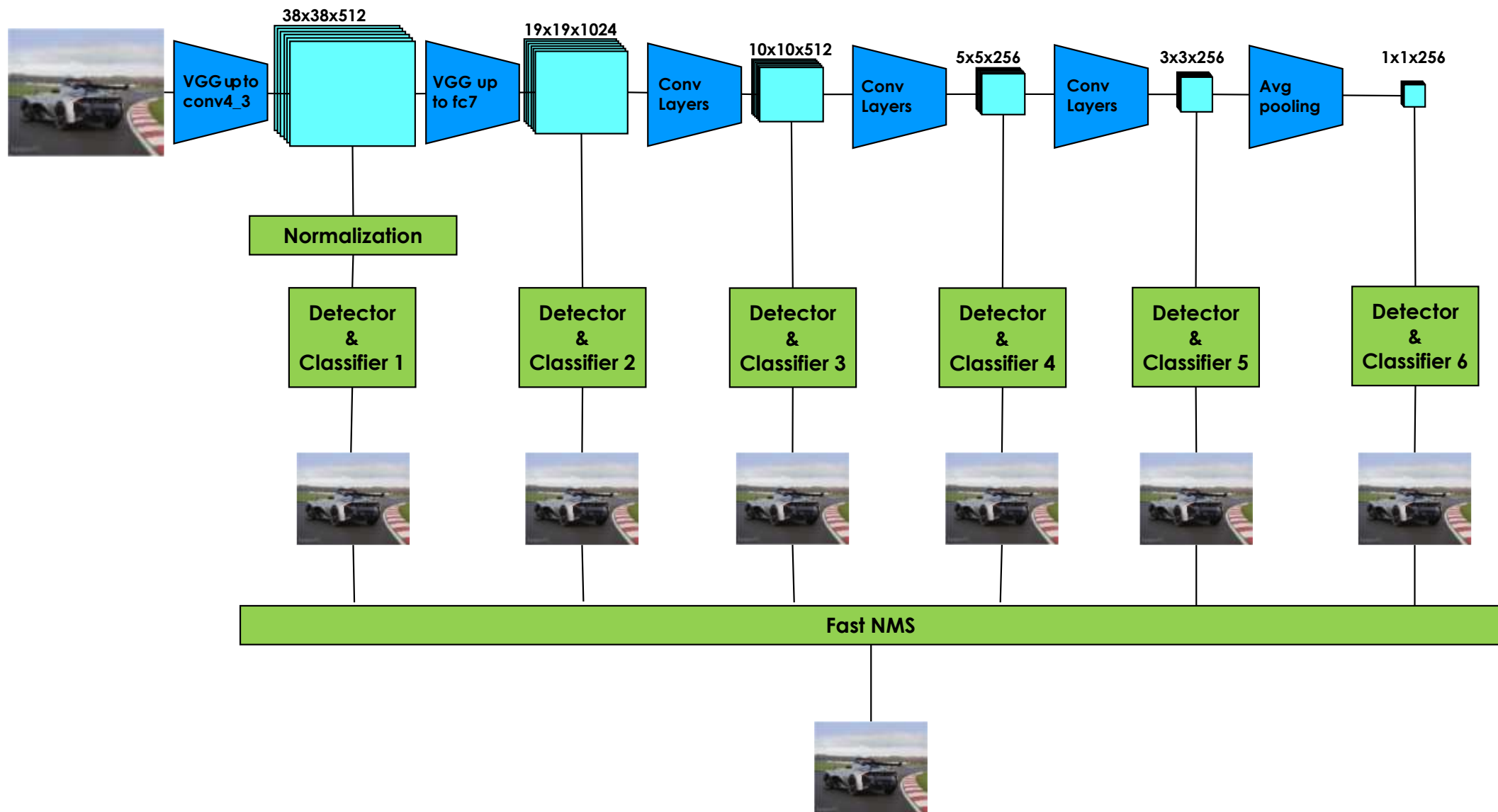
# SSD 300 architecture

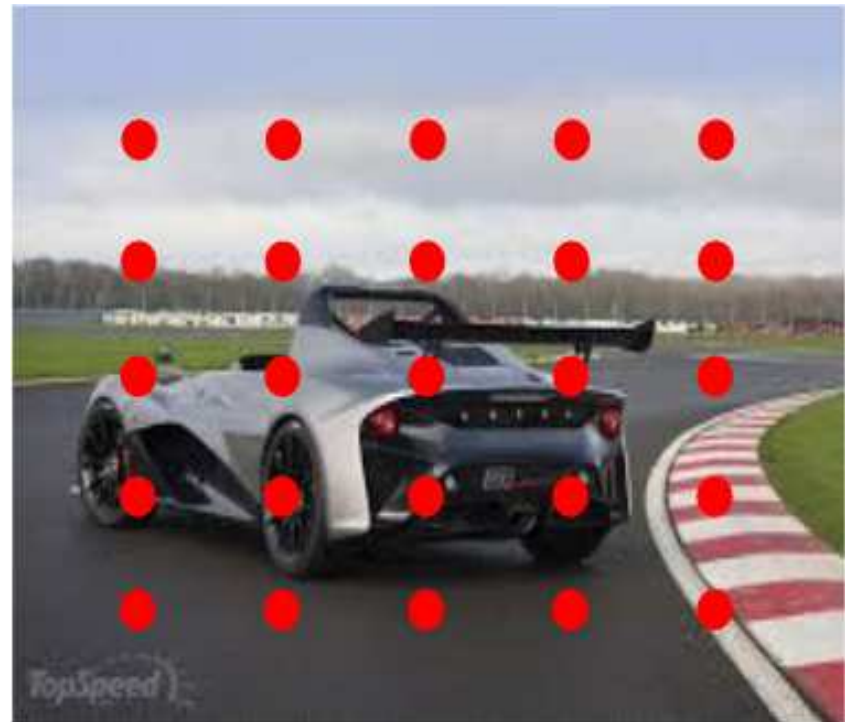


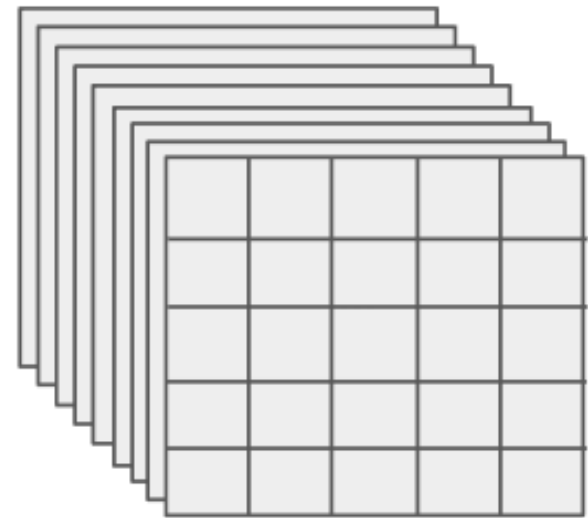
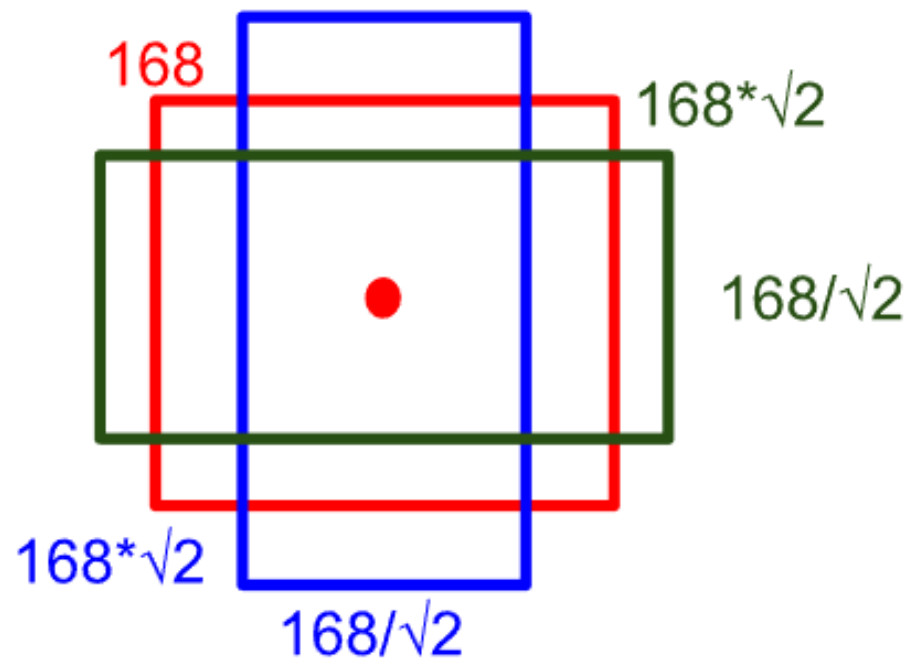
# Key Points

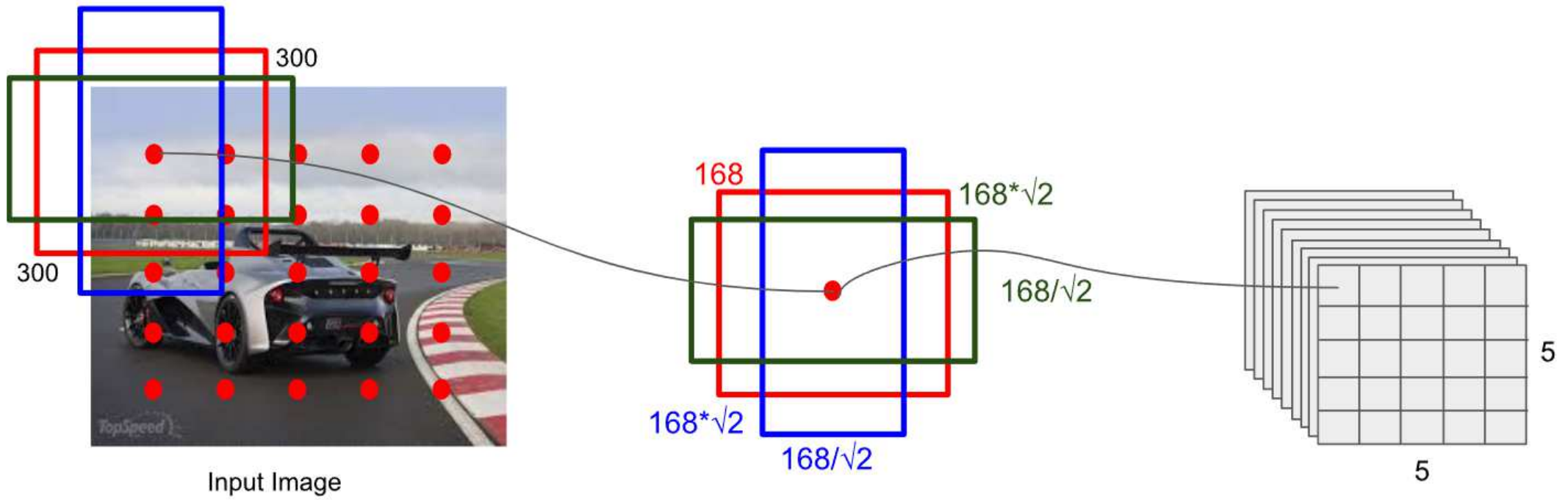
- **SSD architecture enables real-time object detection**
  - **The quality of work is close to Faster R-CNN**
  - **Detection occurs at different scales, which allows you to localize objects of different sizes**
  - **A large number of default boxes are used, covering the input image at different scales**
  - **At the Inference stage, the SSD 300 architecture detects 7308 objects, most of which are subsequently filtered.**
-











# 과정 개요



아웃풋을 만드는 공간을 나눈다(multi feature map). 각 피쳐맵(아웃풋맵)에서 다른 비율과 스케일로 **default box**를 생성하고 모델을 통해 계산된 좌표와 클래스값에 **default box**를 활용해 최종 **bounding box**를 생성한다.

정확도 향상을 위해 서로 다른 피쳐맵에서 서로 다른 스케일의 예측을 할 수 있게 했다. (Yolo v1은 최종 아웃풋은 한 피쳐맵이고, 각 그리드 셀당 2개의 바운딩 박스를 예측하는데 비해 SSD는 여러가지의 그리드셀(피쳐맵)을 가지고 각 피쳐맵당 여러가지의 (보통 6개)바운딩 박스를 가지기 때문) 이 역시 2.1 Model과, 2.2 Training에서 자세히 살펴볼 것.