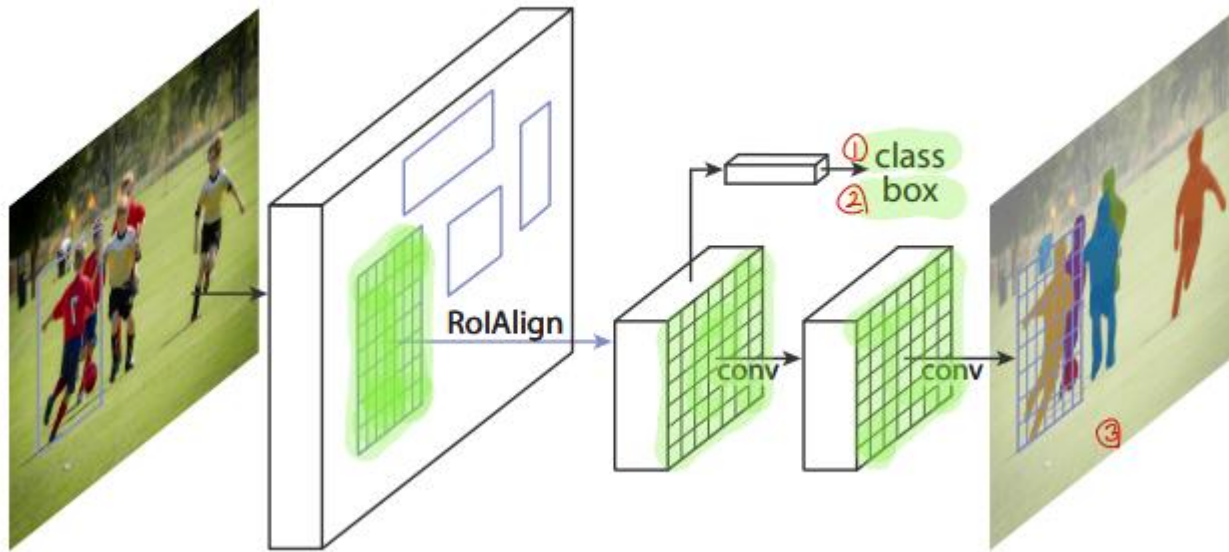


Mask R-CNN

Kaiming He
Georgia Gkioxari
Piotr Dollar
Ross Girshick
Facebook AI Research (FAIR)

◆ Introduction

- Faster R-CNN에 Mask Prediction branch를 추가.
- Class, box, segmentation mask prediction이 parallel 하게 작동함.



◆ Introduction

- Faster R-CNN의 feature extraction 할 때 coarse한 spatial quantization이 일어나는 문제점을 개선함 (RoI Align)
- Mask 와 class prediction을 독립적으로 작동하게 함 (segmentation을 먼저 한 후 class prediction을 하면 중간중간에 선이 생기는 문제 발생(Fig.1))

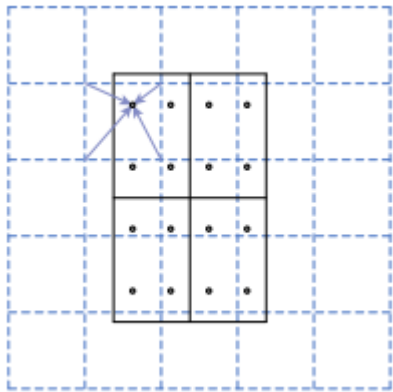


Fig.1

◆ Mask R-CNN

- RoI로 부터 K개의 class에 대해 $m \times m$ 해상도의 binary masks를 생성
- 그 중 ground class에 대한 mask 만 Loss에 기여.
- RoIAlign 도입

Featuremap으로부터 RoI를 sample 할 때, bilinear interpolation을 이용하여 quantization이 일어나지 않도록 함. (공간적인 정보를 보존함)



◆ RoIAlign

- RoIPooling은 RoI로부터 7x7 feature map을 extract하기 위한 기존의 방식이다. 그러나, MaxPooling에 의해 quantization이 일어나고, quantization에 의해 RoI와 extract된 feature 사이에 misalignments가 생긴다. 이를 RoIAlign의 bilinear interpolation으로 해결하였다.

◆ Network Architecture

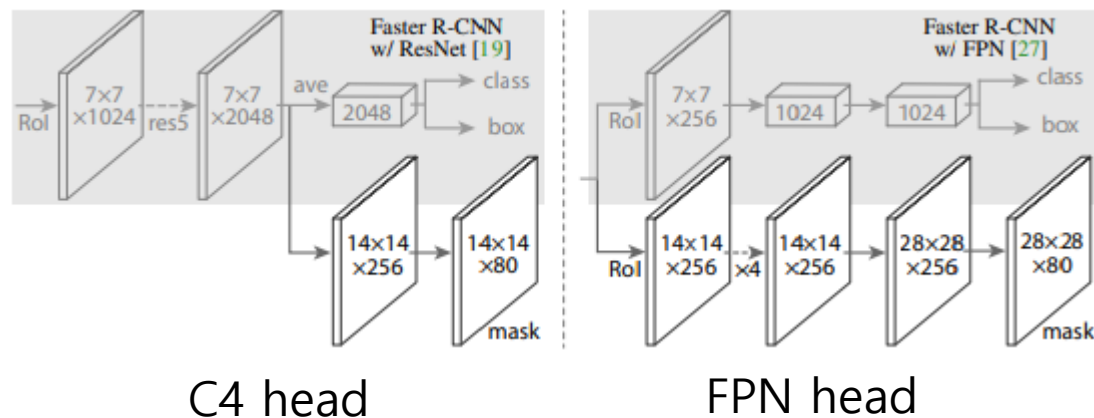
- Backbone Architecture

전체 이미지에 대한 feature extraction 할 때 사용.

- Head architecture

Backbone에서부터 나온 RoI에 대해 classification, box regression, mask prediction을 하는 데 사용.

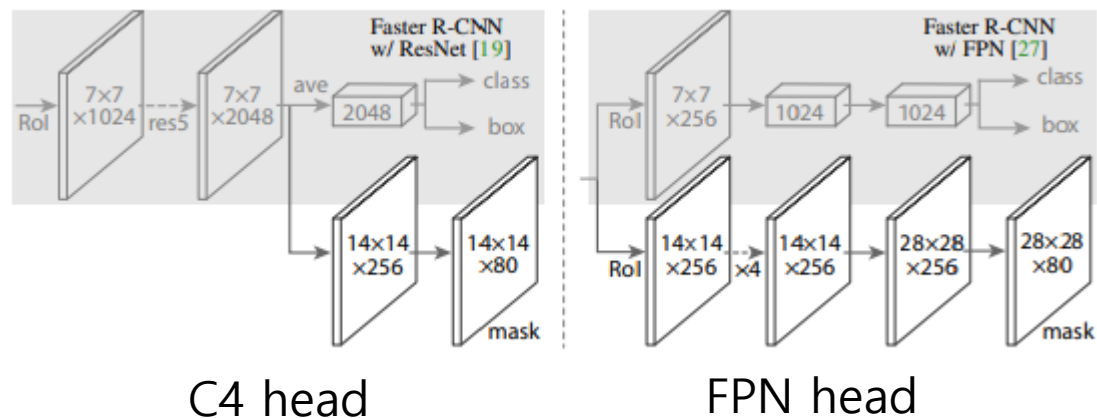
◆ Head Architecture



C4 : ResNet의 4번째 stage의 마지막 Convolutional layer로부터 extract된 feature

FPN : Feature pyramid로로부터 extract된 feature

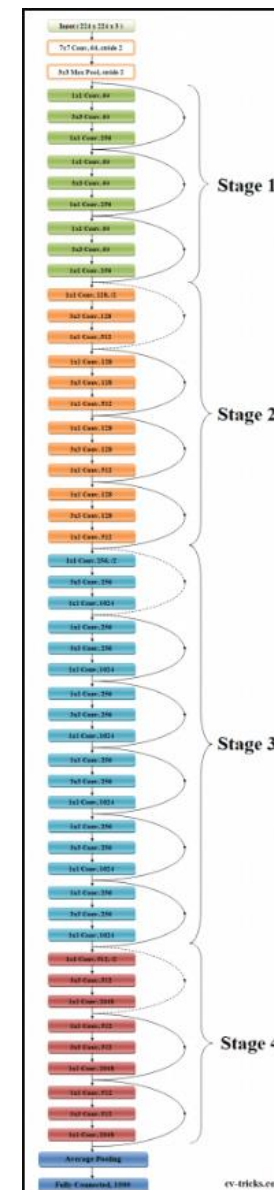
◆ Head Architecture



ResNet 50

C4 : ResNet의 4번째 stage의 마지막 Convolutional layer로부터 extract된 feature

FPN : Feature pyramid로로부터 extract된 feature



◆ Implement

- C4 backbone에서 300개의 proposal이 나오고, (FPN에선 1000개) box prediction에서 NMS를 실행한 후, mask branch가 100 detection box에 대해 실행됨.
- RoI마다 k개의 mask를 predic할 수 있지만, classification branc로 부터 나온 class에 관한 mask만 사용한다. 그 후, $M \times M$ mask output은 RoI size로 resized 된다.