R-FCN: Object Detection via Region-based Fully Convolutional Networks

JifengDai , YiLi , KaimingHe , JianSun

Conference on Neural Information Processing Systems (NIPS 2016)

SPEAKER: 林仕閔

Target

• Region-based, fully convolutional networks(based on ResNet-50) ➡ classify object categories

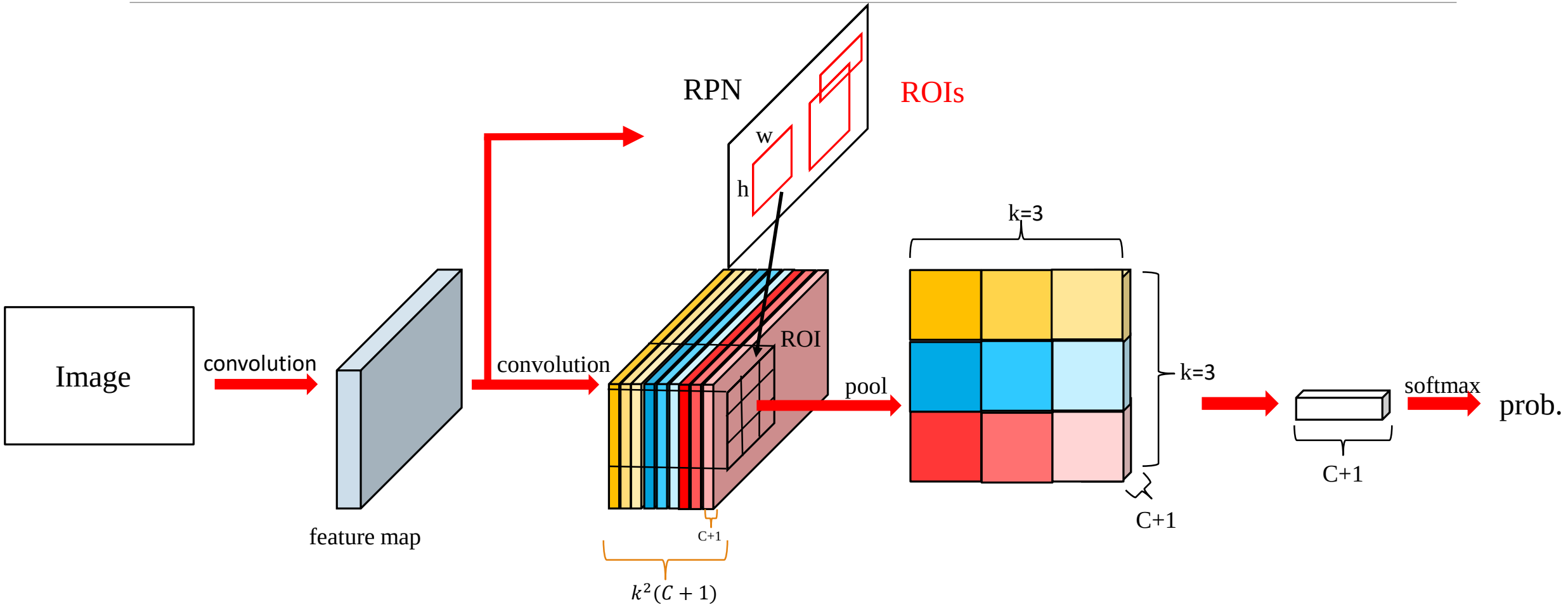
•Target : predict objects in the image

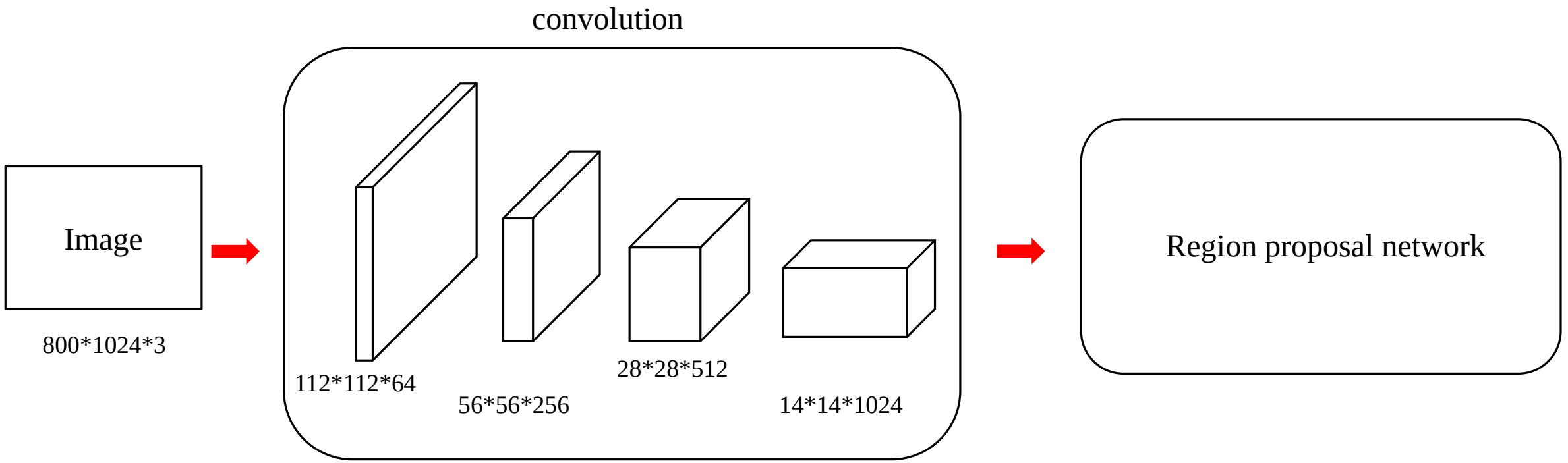
input

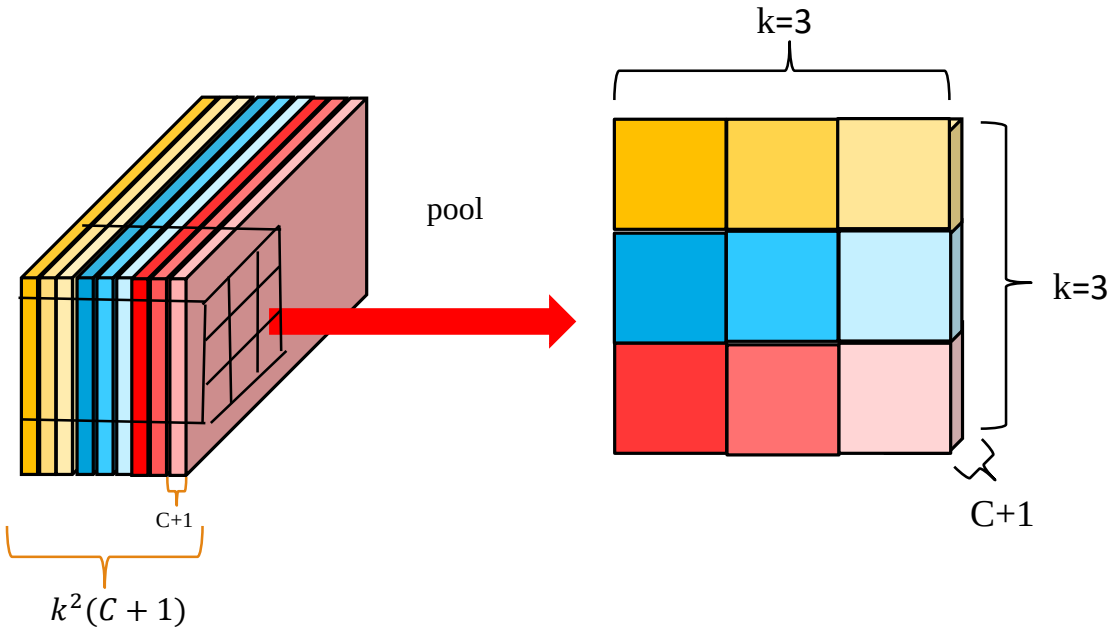
output

Framework

S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.




convolution
Image
800*1024*3
112*112*64
56*56*256
28*28*512
14*14*1024
Region proposal network

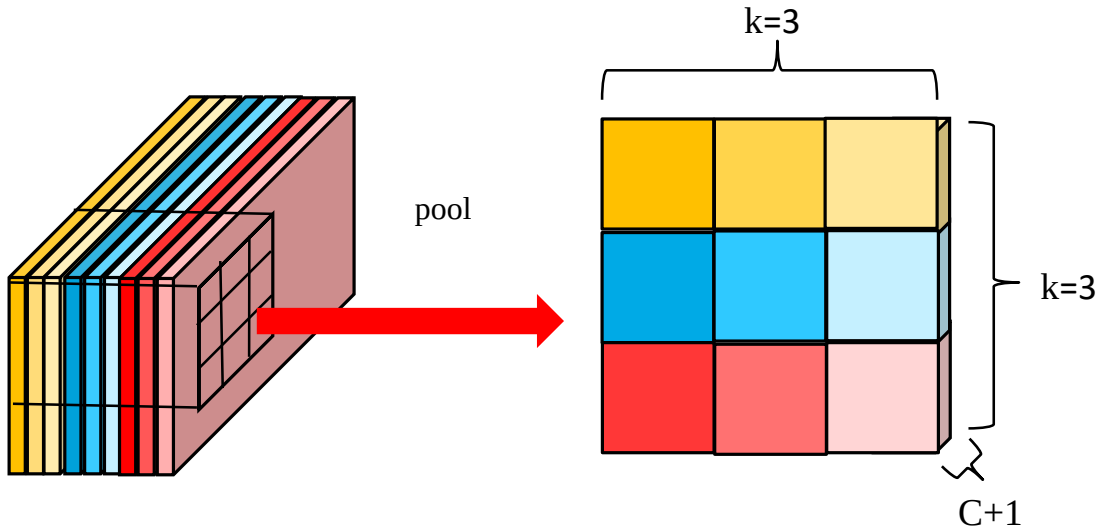


每個顏色的部分為物體對應到的位置，像是偵測人的話，黃色部分可能是人的頭，藍色部分可能為人的手，紅色部分可能是腳

Pool

$$r_c(i,j \mid \Theta) = \sum_{(x,y)\in bin(i,j)} \frac{z_{i,j,c}(x+x_0, y+y_0)}{n}$$

$$r_c(\Theta) = \sum_{(i,j)} r_c(i,j \mid \Theta)$$



c : category

(i,j) : ROI pooling bin $(1 \le i,j \le k)$

Θ : all learnable parameters of the network

$z_{i,j,c}(x+x_0, y+y_0)$: score

(x_0, y_0) : the top-left corner of an RoI

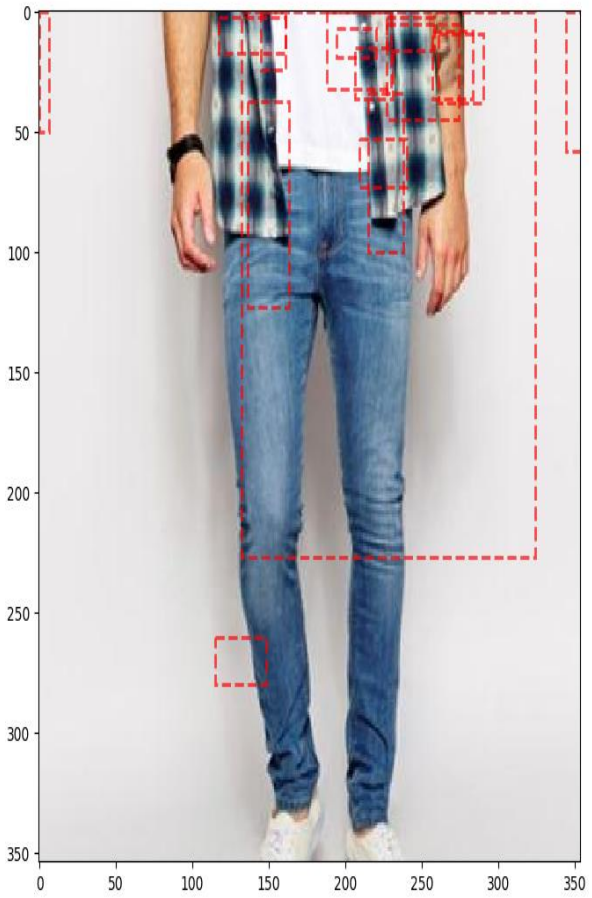
n : number of pixels in the bin

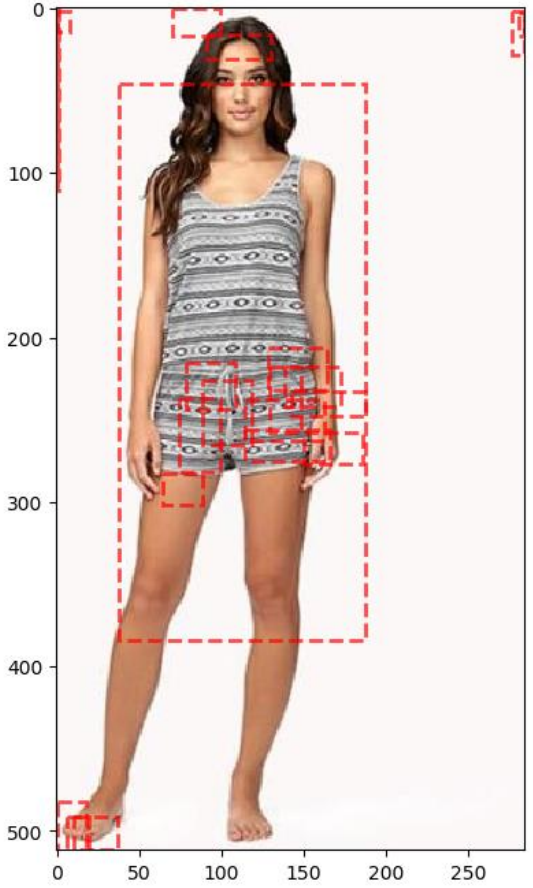
Objective Function

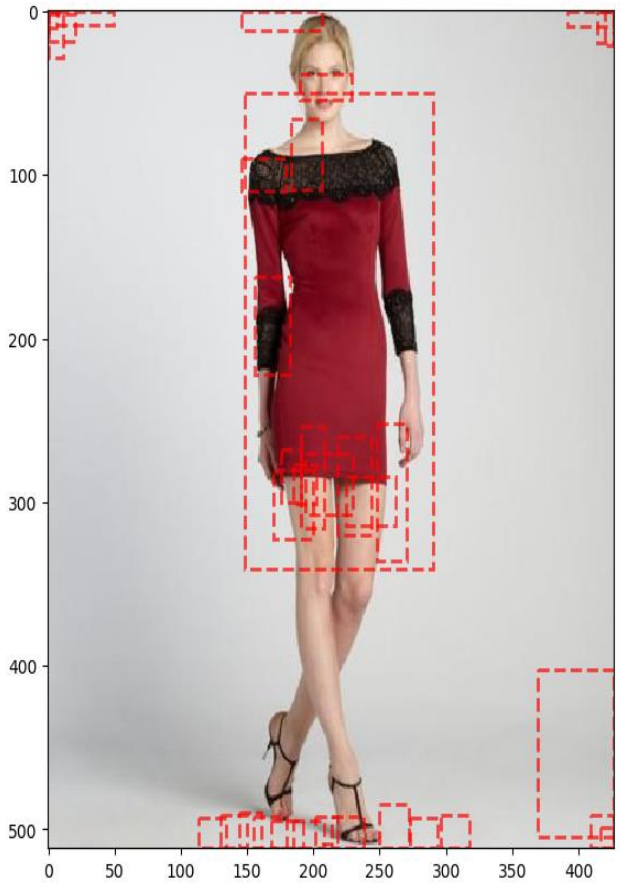
Objective Function : $L_{(s,t_{x,y,w,h})} = L_{cls}(s_{c^*}) + \lambda[c^*>0]L_{reg}(t, t^*)$

- $L_{cls}(s_{c^*})$ = -log(s_{c^*})
- $s_{c^*} = \frac{\exp(r_c(\Theta))}{\sum_{c'=0}^{C}\exp(r_{c'}(\Theta))}$, $\begin{cases} s_{c^*} : \text{softmax responses across categories} \\ r_c(\Theta) : c^{th}\ total\ score \\ \Theta : \text{all learnable parameters of the network} \end{cases}$
- $\lambda = \begin{cases} 0\ , & background \\ 1\ , & otherwise \end{cases}$
- $L_{reg}(t, t^*) = \sum_{i\in(x,y,w,h)} smooth_{L1}(t - t^*)$, $smooth_{L1}(x) = \begin{cases} 0.5x^2, & if\ |x| < 1 \\ |x| - 0.5, & otherwise \end{cases}$

Result


0
50
100
150
200
250
300
350
0
50
100
150
200
250
300
350


0
100
200
300
400
500
0
50
100
150
200
250


0
100
200
300
400
500
0
50
100
150
200
250
300
350
400