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

JifengDai , YiLi , KaimingHe , JianSun

Conference on Neural Information Processing Systems (NIPS 2016)

PRESENTER: 林仕閔

Abstract

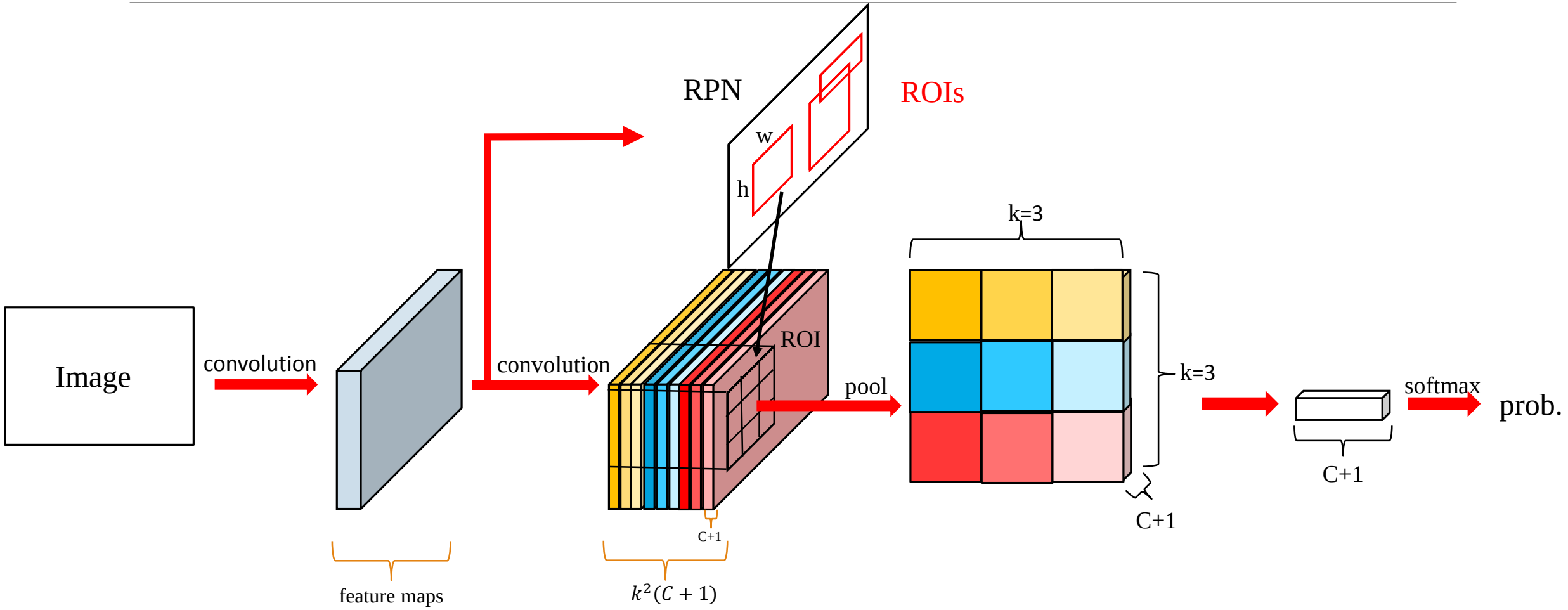
• Region-based, fully convolutional networks(based on ResNet-101) ➡ classify object categories

•Target : predict objects in the image

input

output

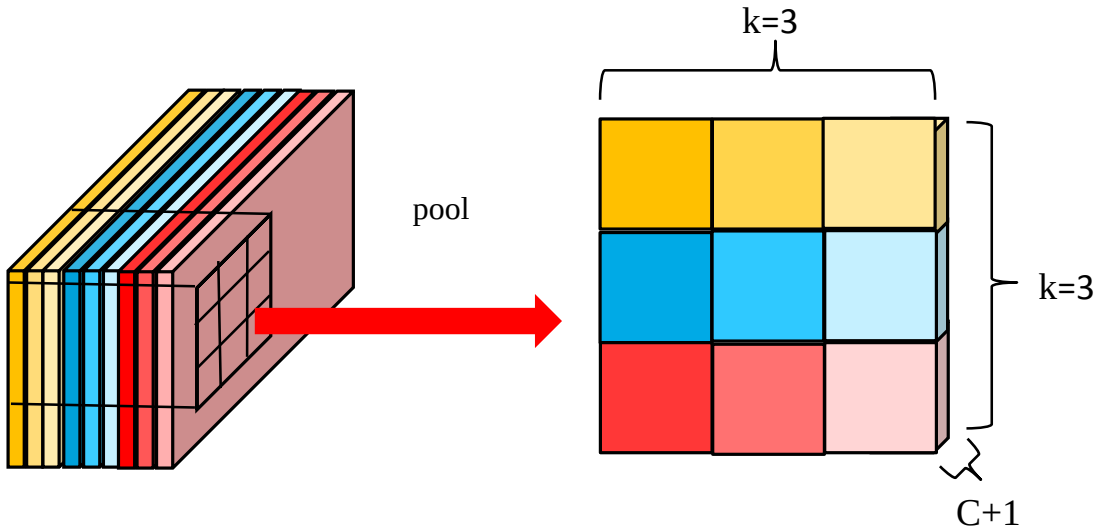
Framework


RPN
ROIs
w
h
Image
convolution
convolution
ROI
pool
k=3
k=3
C+1
C+1
softmax
prob.
feature maps
$k^2(C+1)$

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

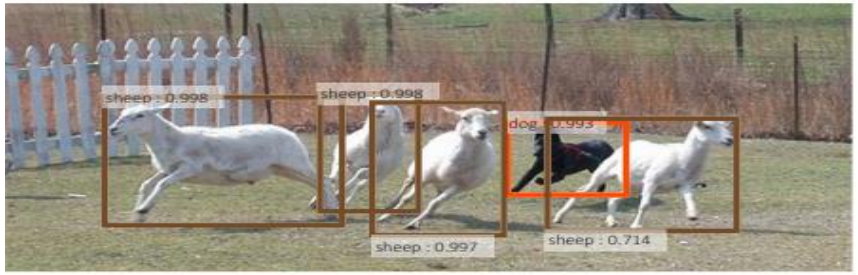
n : number of pixels in the bin

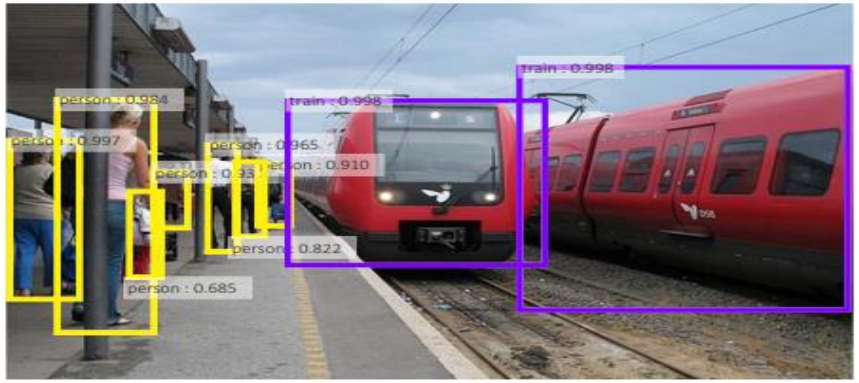
Objective Function

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

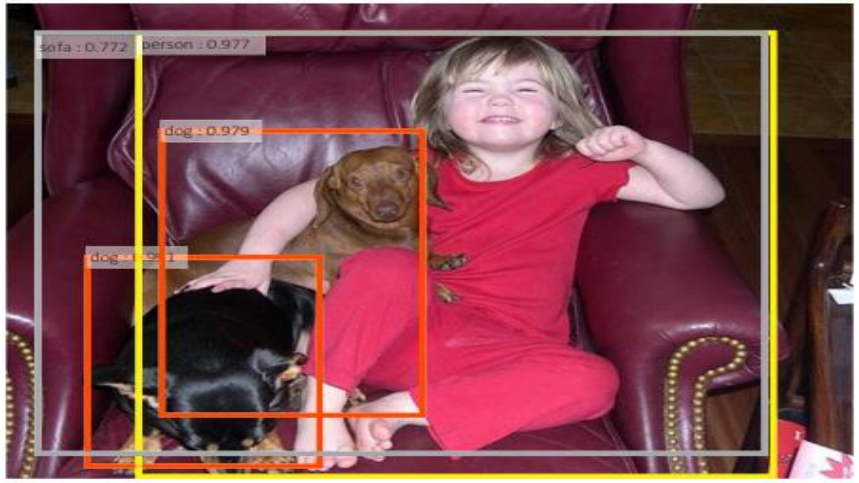
- $L_{cls}(s_{c^*})$ = -log(s_{c^*})
- $s_{c^*} = \frac{\exp(r_c(\Theta))}{\sum_{c'=0}^{C}\exp(r_{c'}(\Theta))}$, $\begin{cases} s_{c^*} : \text{softmax responses across categories} \\ r_c(\Theta) : c^{th}\ total\ score \\ \Theta : \text{all learnable parameters of the network} \end{cases}$
- $\lambda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

Result


sheep : 0.998
sheep : 0.998
sheep : 0.997
sheep : 0.714


train : 0.998
train : 0.998
person : 0.910
person : 0.822
person : 0.685


aeroplane : 0.979
aeroplane : 0.959


sofa : 0.772
person : 0.977
dog : 0.979

Reference

[1] **Rich feature hierarchies for accurate object detection and semantic segmentation**

[2] **Faster R-CNN**

[3] **Fully convolutional networks for semantic segmentation**