SEMANTIC SEGMENTATION

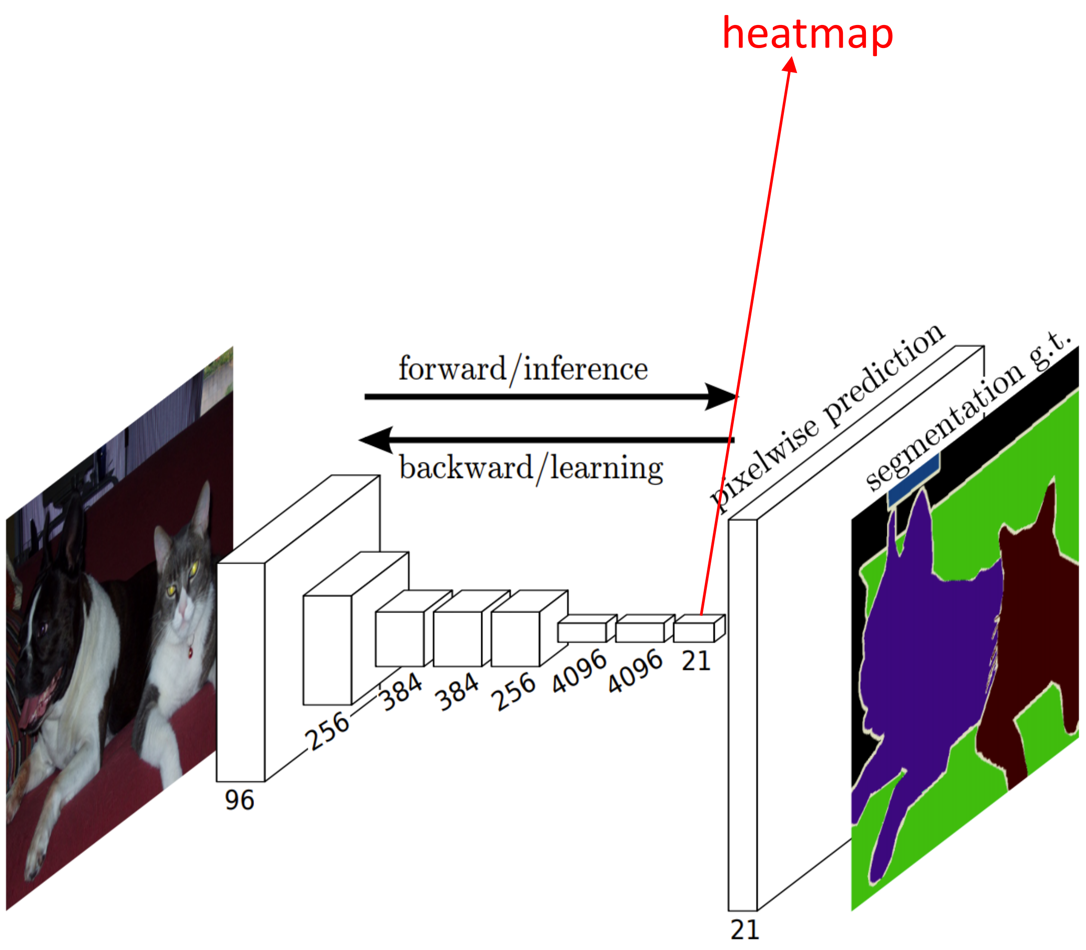
焦思宇

CONTENTS

01 Fully Convolutional Networks

01 FCN

- Convolutional—卷积化
- Upsample—上采样
- Skip Layer—跳跃结构



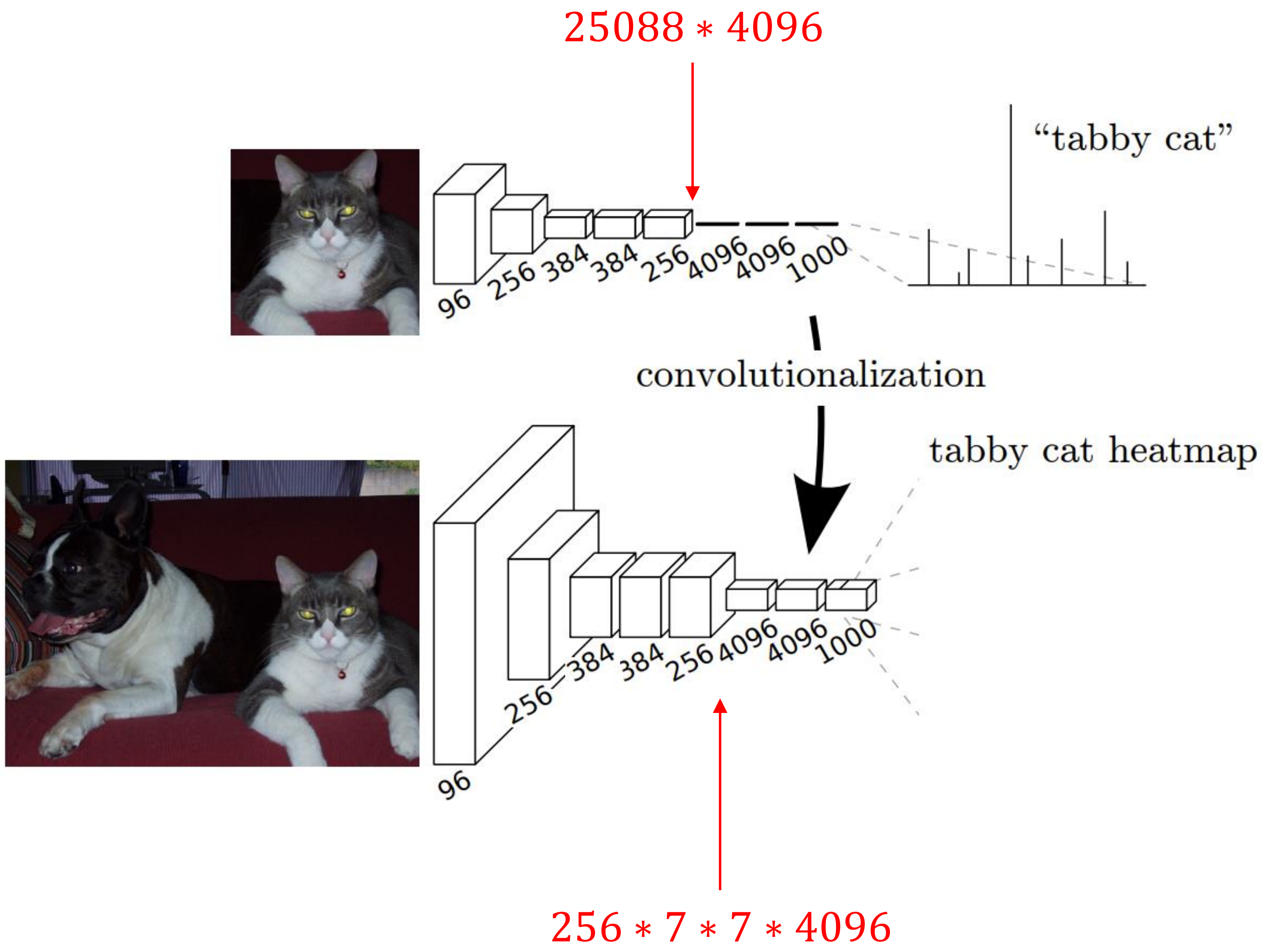
01 FCN

- Convolutional

$224 * 224 \longrightarrow 7 * 7$

$w * h \longrightarrow \frac{w}{32} * \frac{h}{32}$

1、输入图片大小不受限
2、Fine-turning

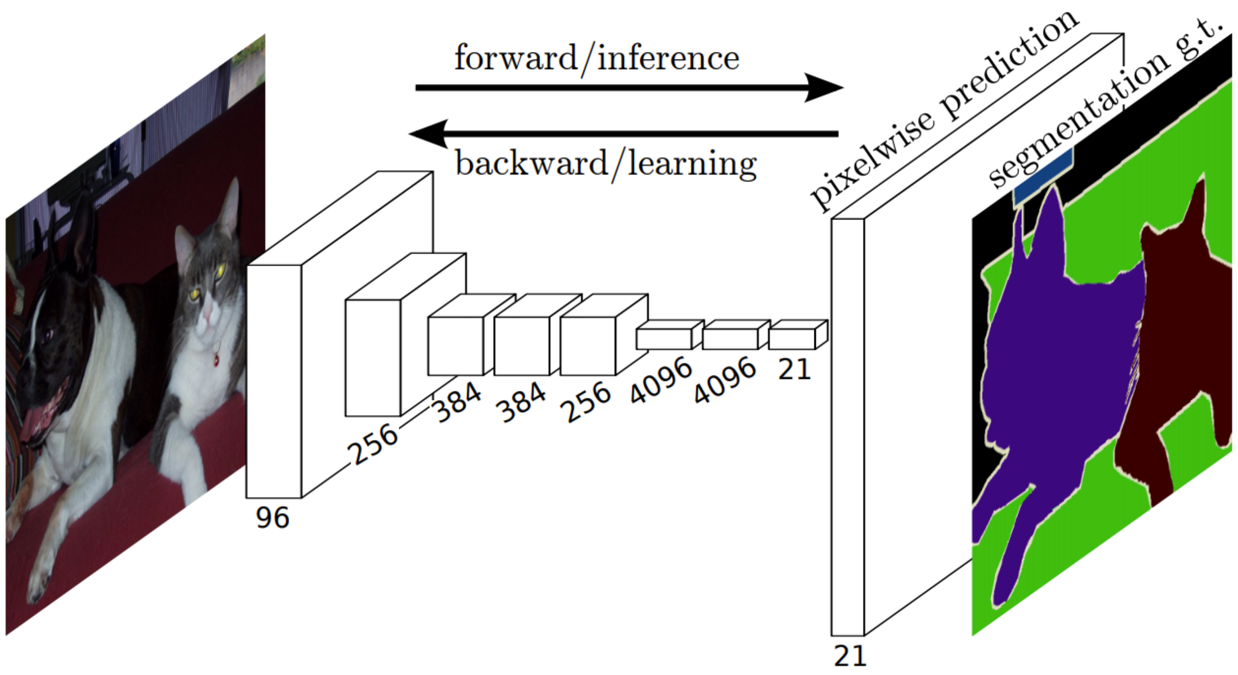


01 FCN

- Upsample

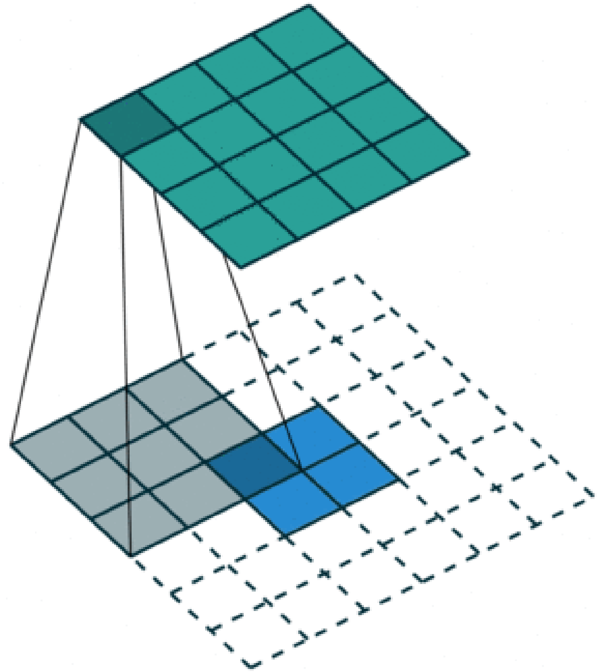
  - Deconvolution

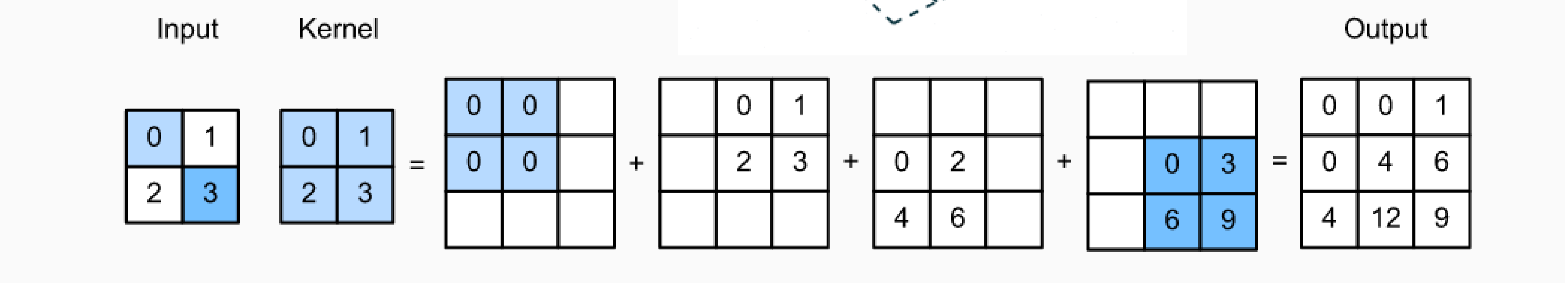
  - Bilinear interpolation



01 FCN

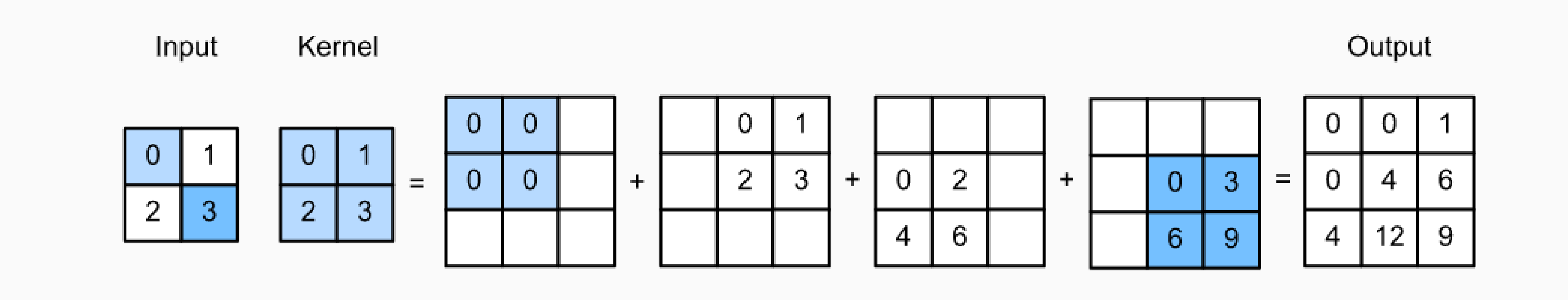
- Upsample
    - Deconvolution



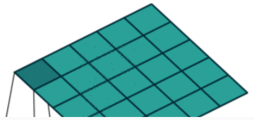


01 FCN

- Upsample
  - Deconvolution



- Strides=2



Because $(320-64+16\times2+32)/32=10$ and $(480-64+16\times2+32)/32=15$, we construct a transposed convolution layer with a stride of 32 and set the height and width of the convolution kernel to 64 and the padding to 16. It is not difficult to see that, if the stride is s, the padding is $s/2$ (assuming $s/2$ is an integer), and the height and width of the convolution kernel are $2s$, the transposed convolution kernel will magnify both the height and width of the input by a factor of s.

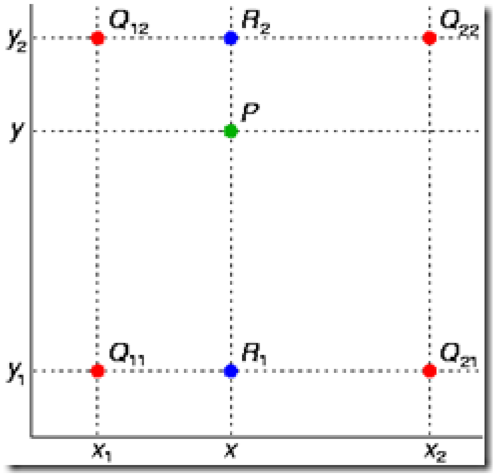
- Padding=1

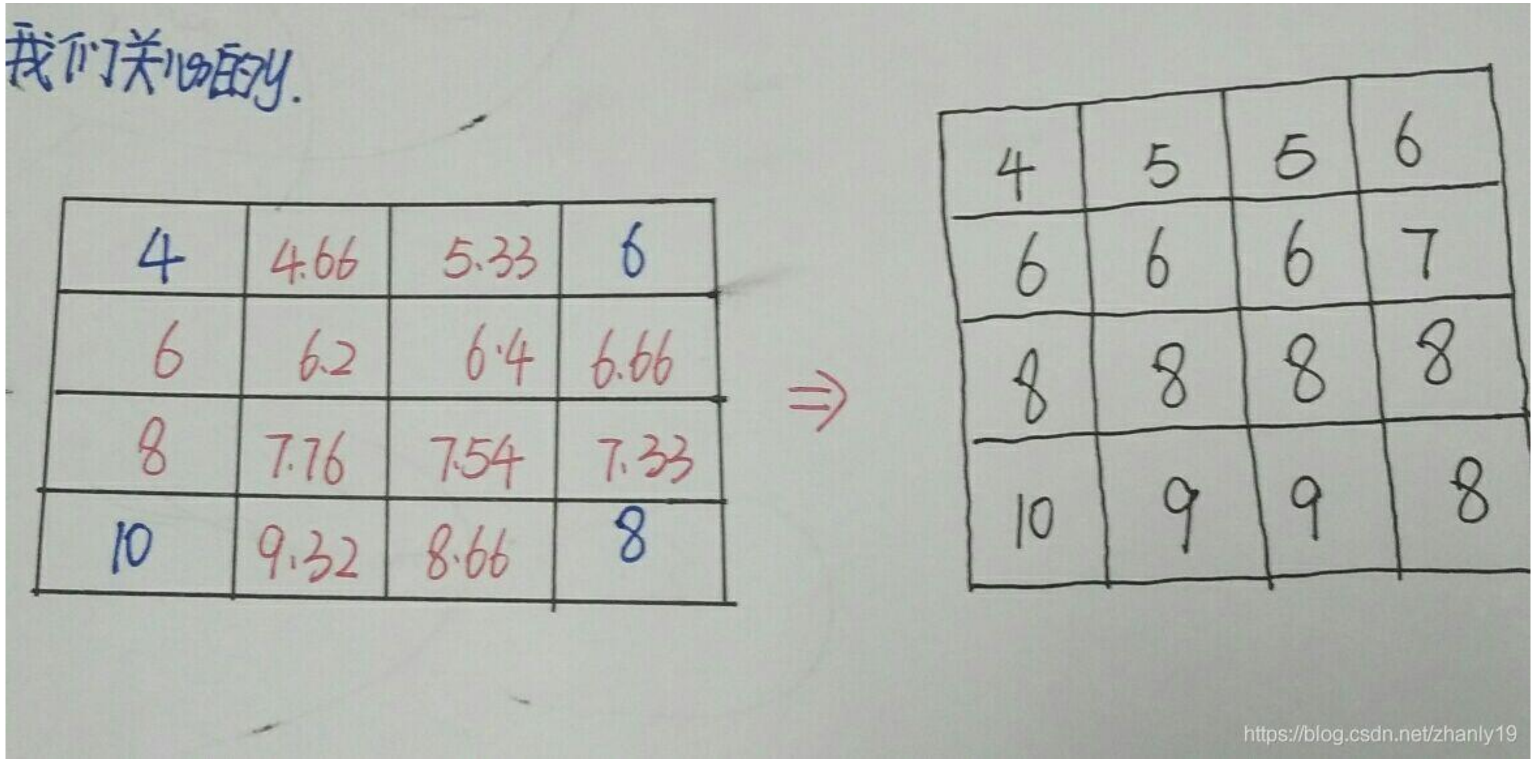
```
array([[[[4.]]]])
```

01 FCN

- Upsample

  - Bilinear interpolation 双线性差值





01 FCN

- Upsample
    - Deconvolution 的初始化

Kernel=3

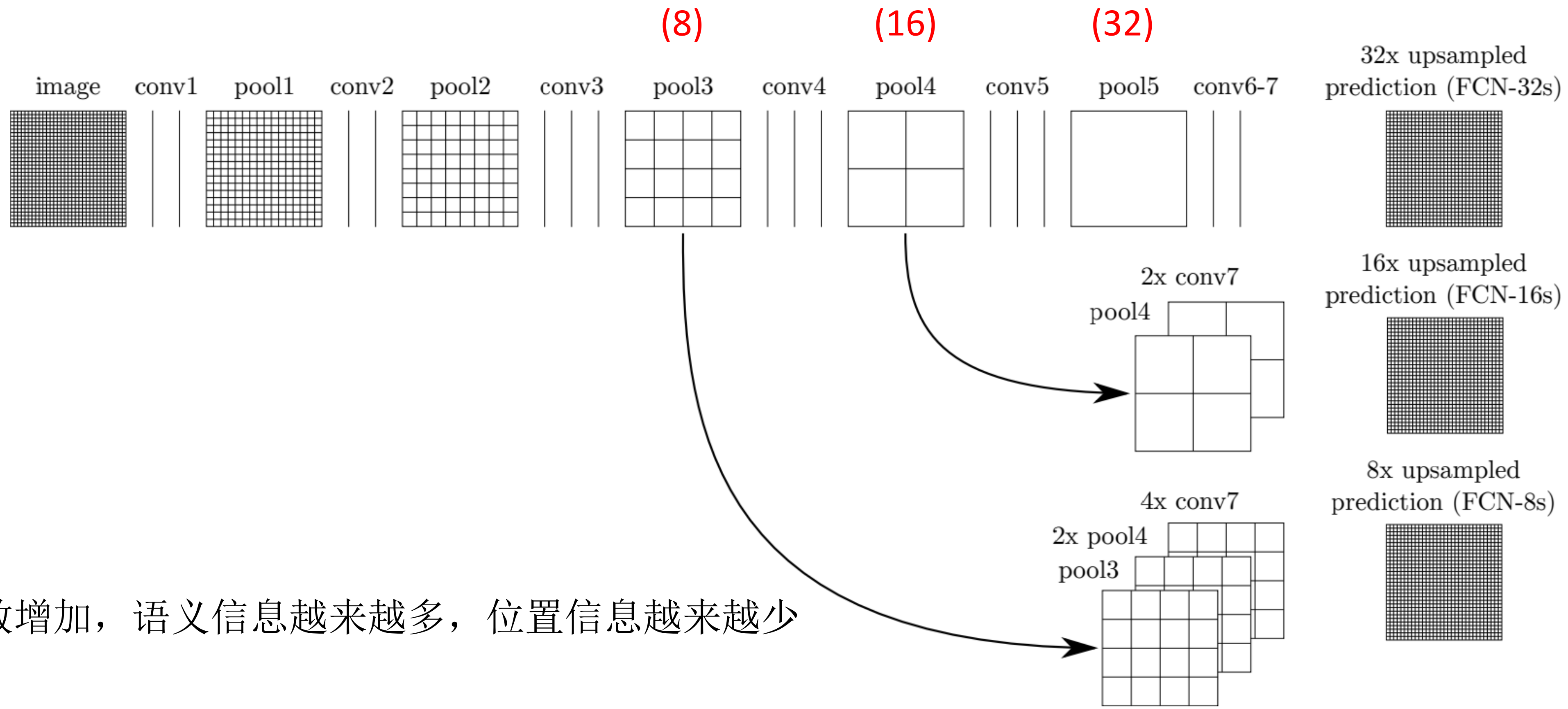
```
[[[[0.25 0.5  0.25]
   [0.5  1.   0.5 ]
   [0.25 0.5  0.25]]]]
```

Kernel=5

```
[[[[0.1111111 0.2222222 0.3333333 0.2222222 0.1111111]
   [0.2222222 0.4444444 0.6666666 0.4444444 0.2222222]
   [0.3333333 0.6666666 1.        0.6666666 0.3333333]
   [0.2222222 0.4444444 0.6666666 0.4444444 0.2222222]
   [0.1111111 0.2222222 0.3333333 0.2222222 0.1111111]]]]
```

01 FCN

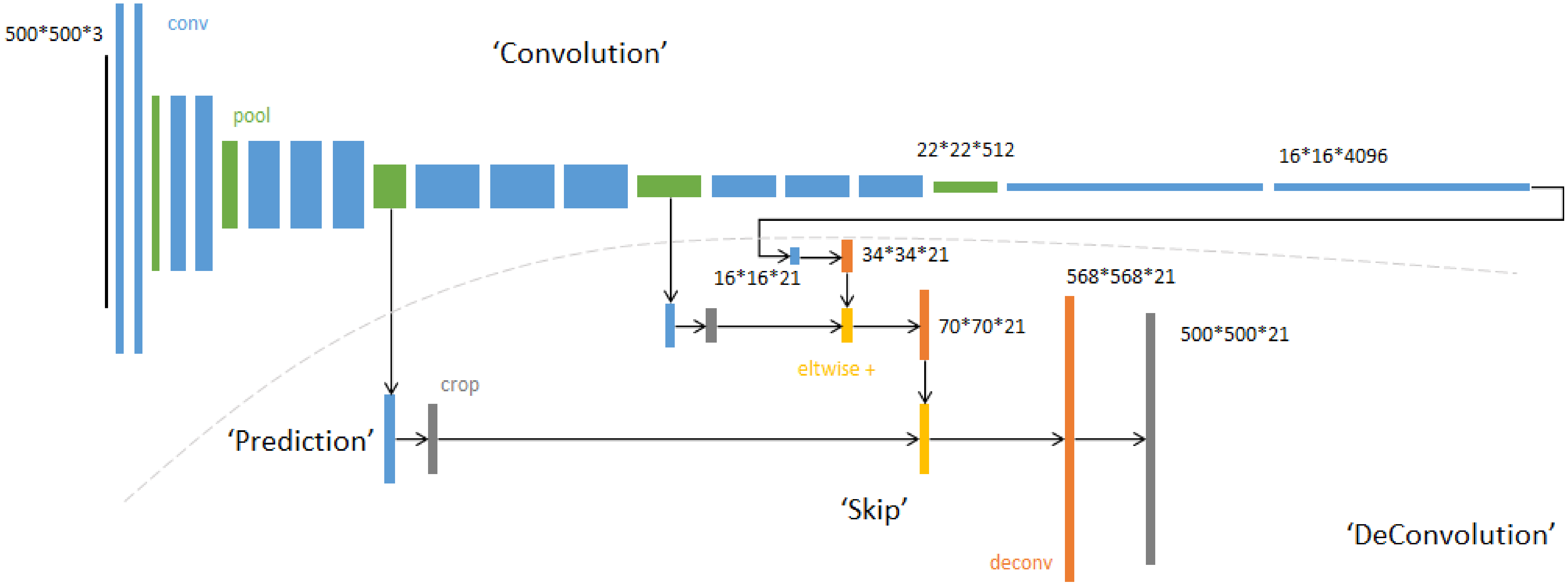
- Skip Layer



- 随着网络层数增加，语义信息越来越多，位置信息越来越少
- Multiscale

01 FCN

- Skip Layer



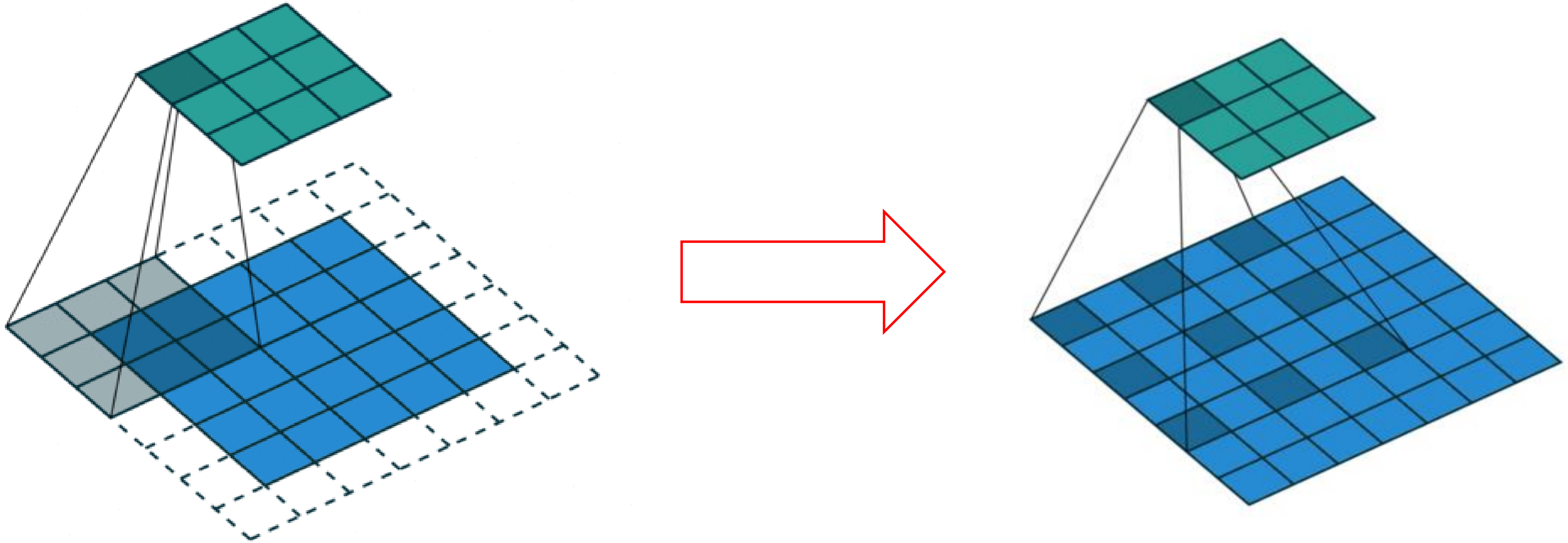
02 DeepLab

02 DeepLab

Problems	→	Solution
• Downsampling导致最终输出特征图的分辨率明显降低		• atrous convolution 空洞卷积
• 多尺度目标		• ASPP 空洞空间卷积池化金字塔
• DCNN空间不变性导致定位不准确		• Fully Connected CRF 条件随机场

02 DeepLab

- atrous convolution



Rate=1

Rate=2

02 DeepLab

- ASPP —>Atrous convolution + SPP

  - SPP

SPP 显著特点

1) 不管输入尺寸是怎样，产生固定大小的输出
2) 使用多个窗口
3) 结构独立

Output=$(16+4+1)\times 256-1D$

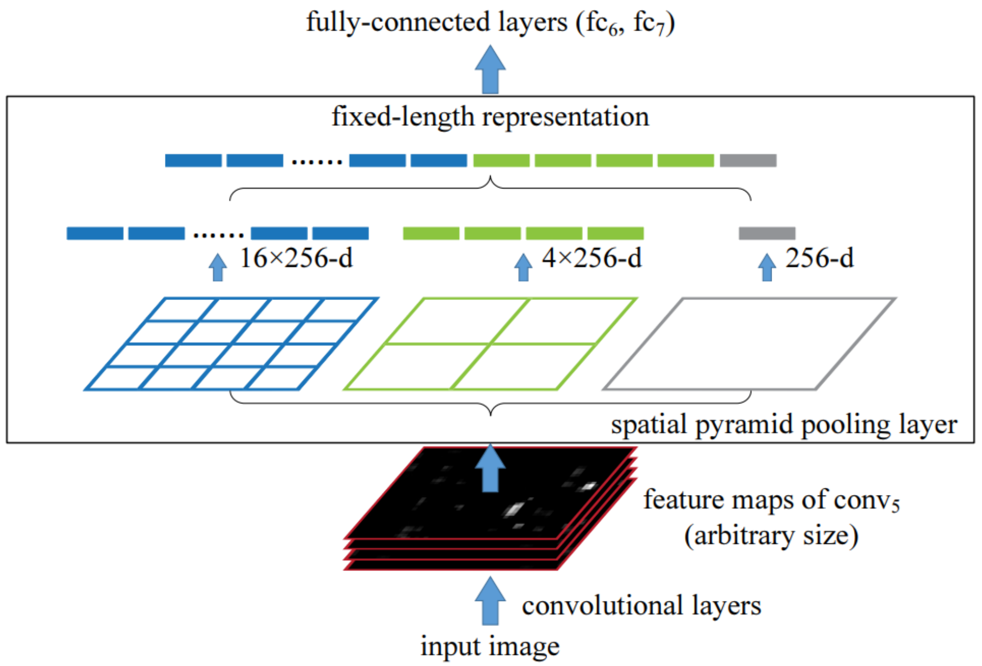
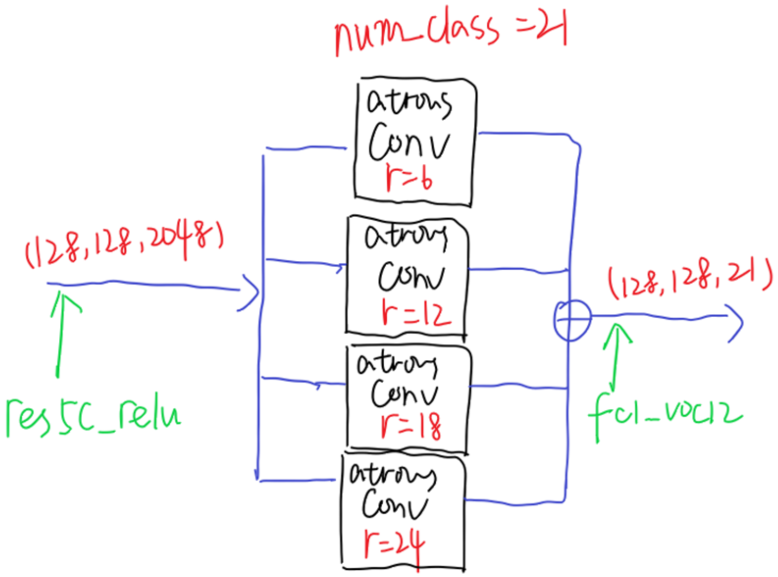
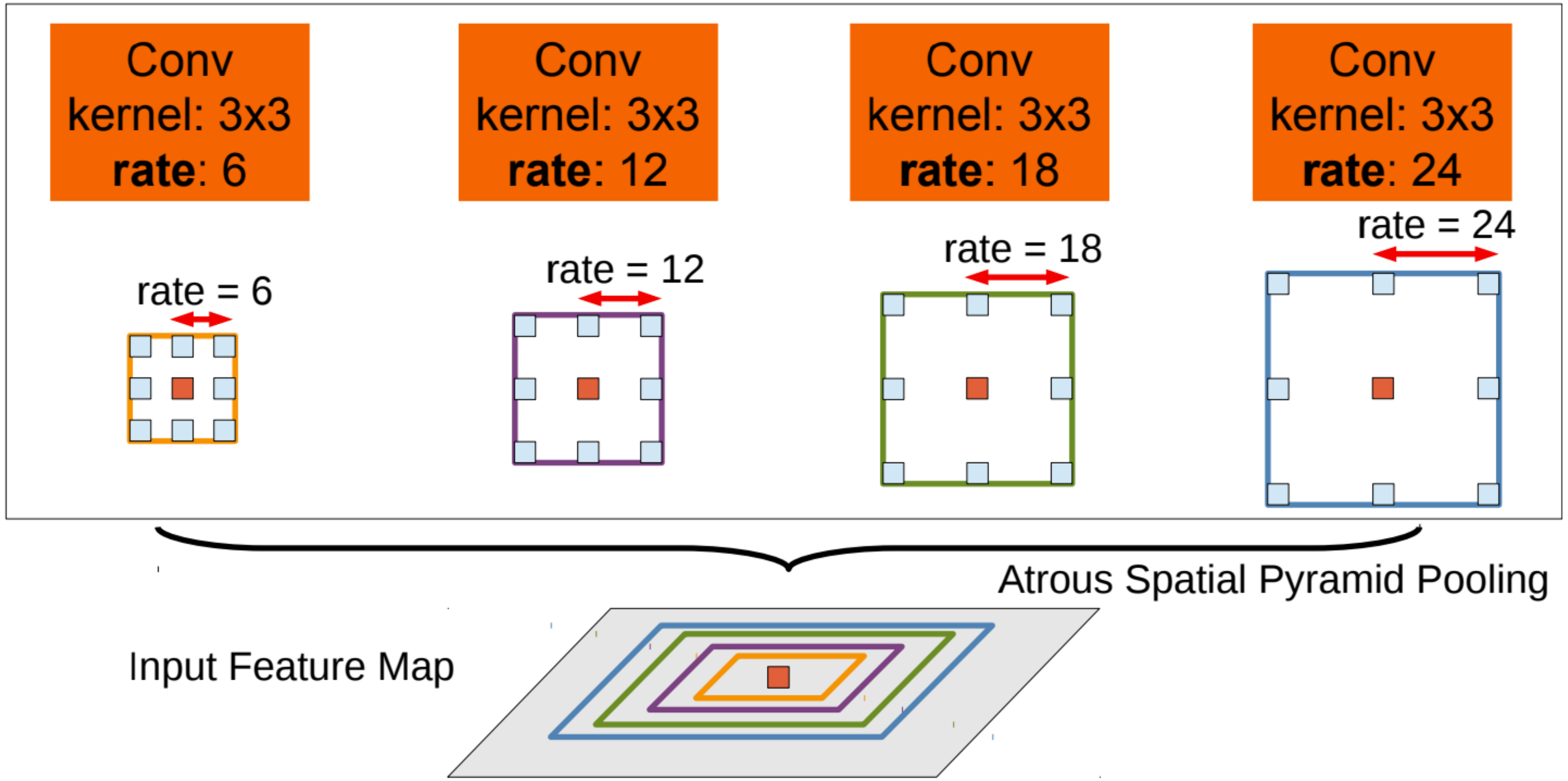


Figure 3: A network structure with a **spatial pyramid pooling layer**. Here 256 is the filter number of the $conv_5$ layer, and $conv_5$ is the last convolutional layer.

02 DeepLab

- ASPP
  - Atrous convolution + SPP

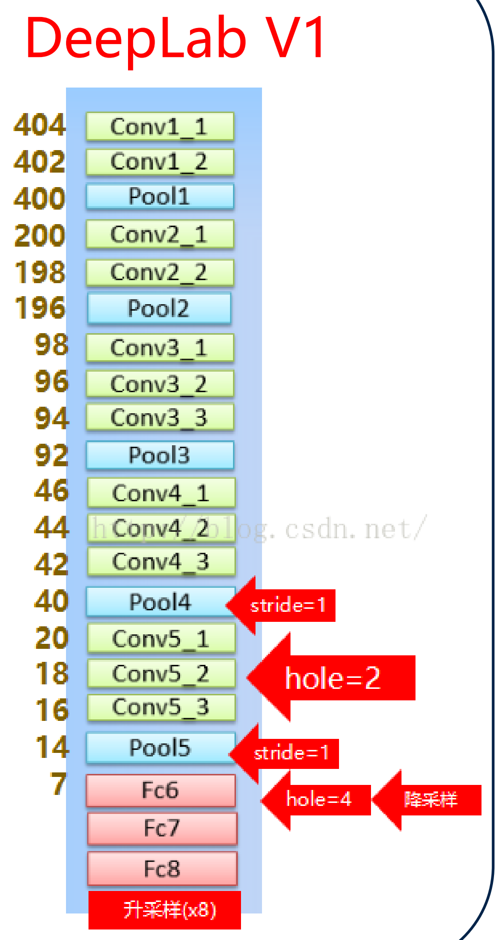


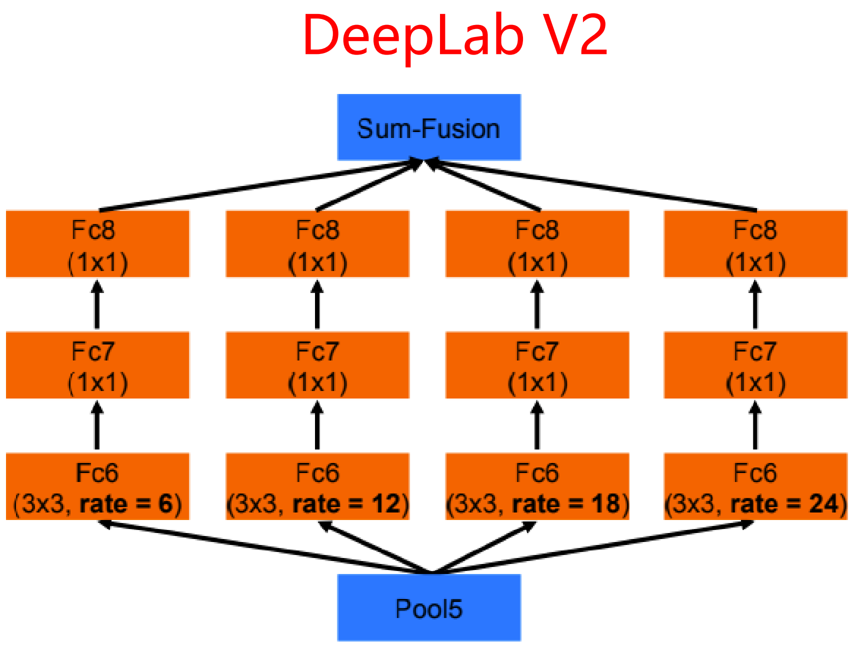


02 DeepLab

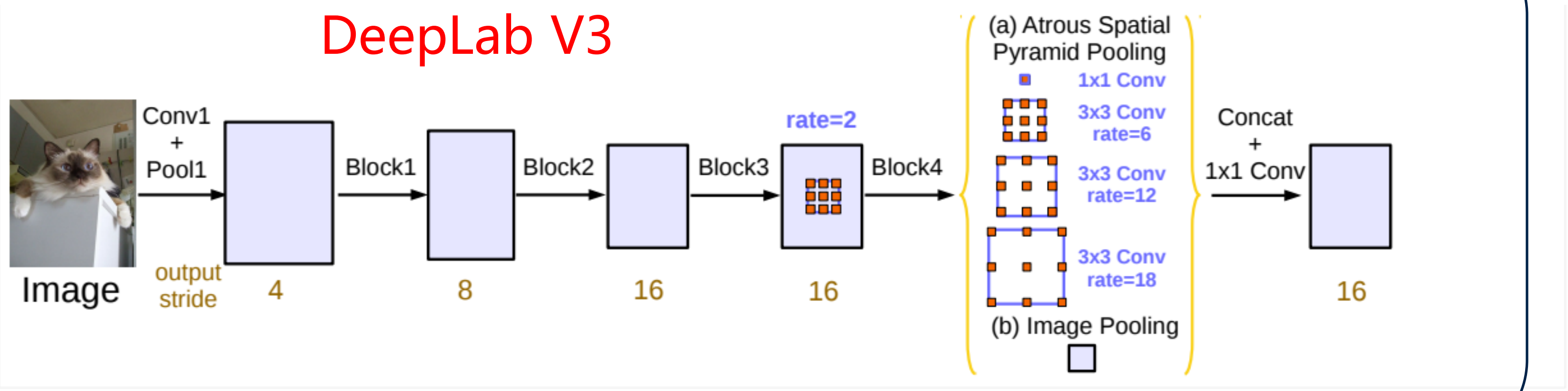
解决了：（1）Downsampling导致最终输出特征图的分辨率明显降低（2）多尺度目标

- 结构对比 VGG16





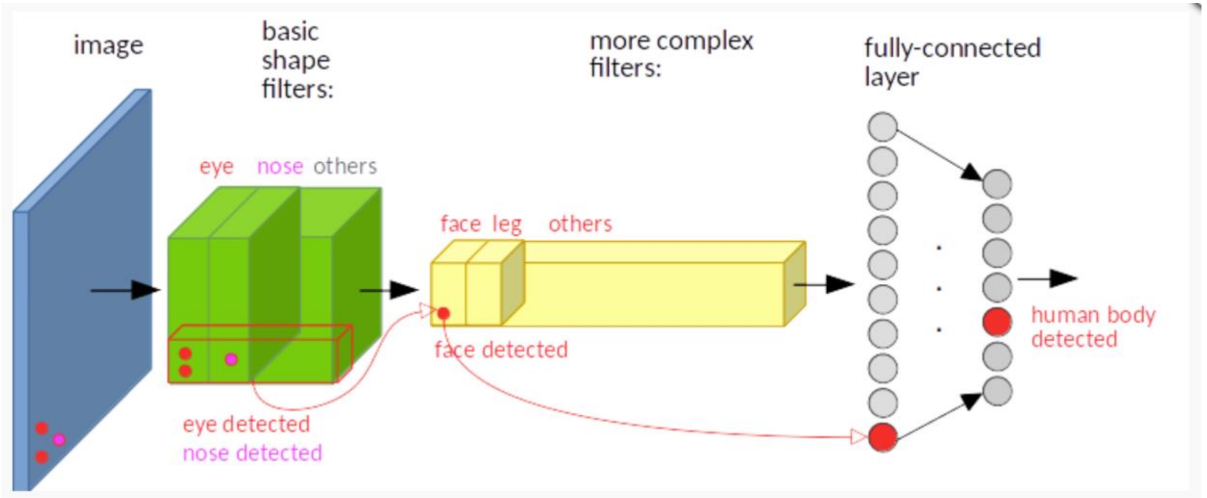
采用双线性差值上卷积，因为空洞卷积，stride小，感受野大

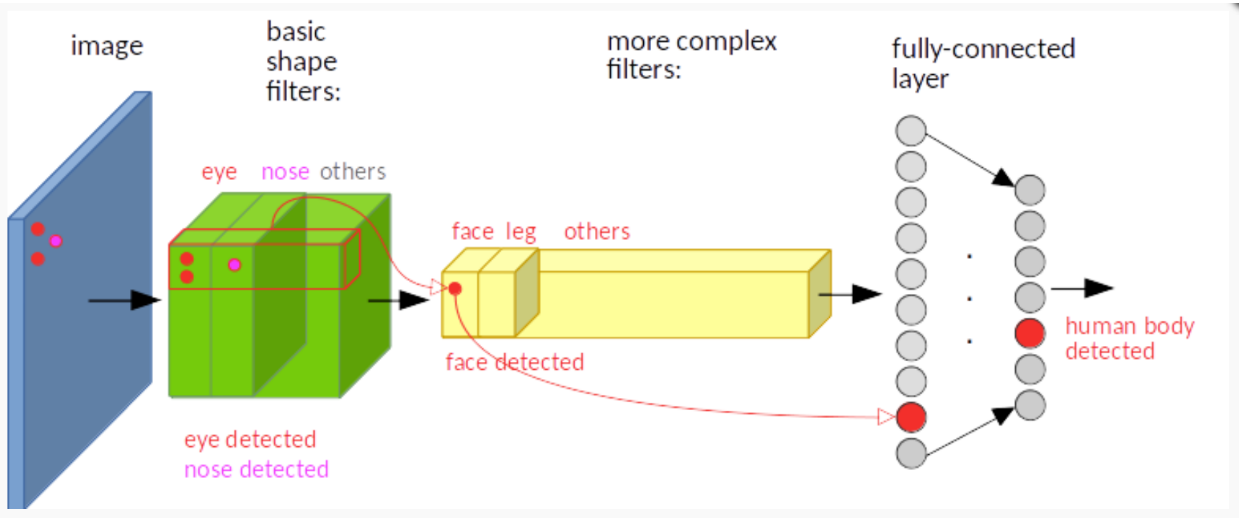


02 DeepLab

- Fully Connected CRF 条件随机场

Problem3： DCNN空间不变性





02 DeepLab

说下笔记本最后
的圈1圈2

- Fully Connected CRF

像素概率

鼓励相似像素分配相同的标签，
而相差较大的像素分配不同标签，

$$E(\boldsymbol{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$

$$\theta_i(x_i) = -\log P(x_i)$$

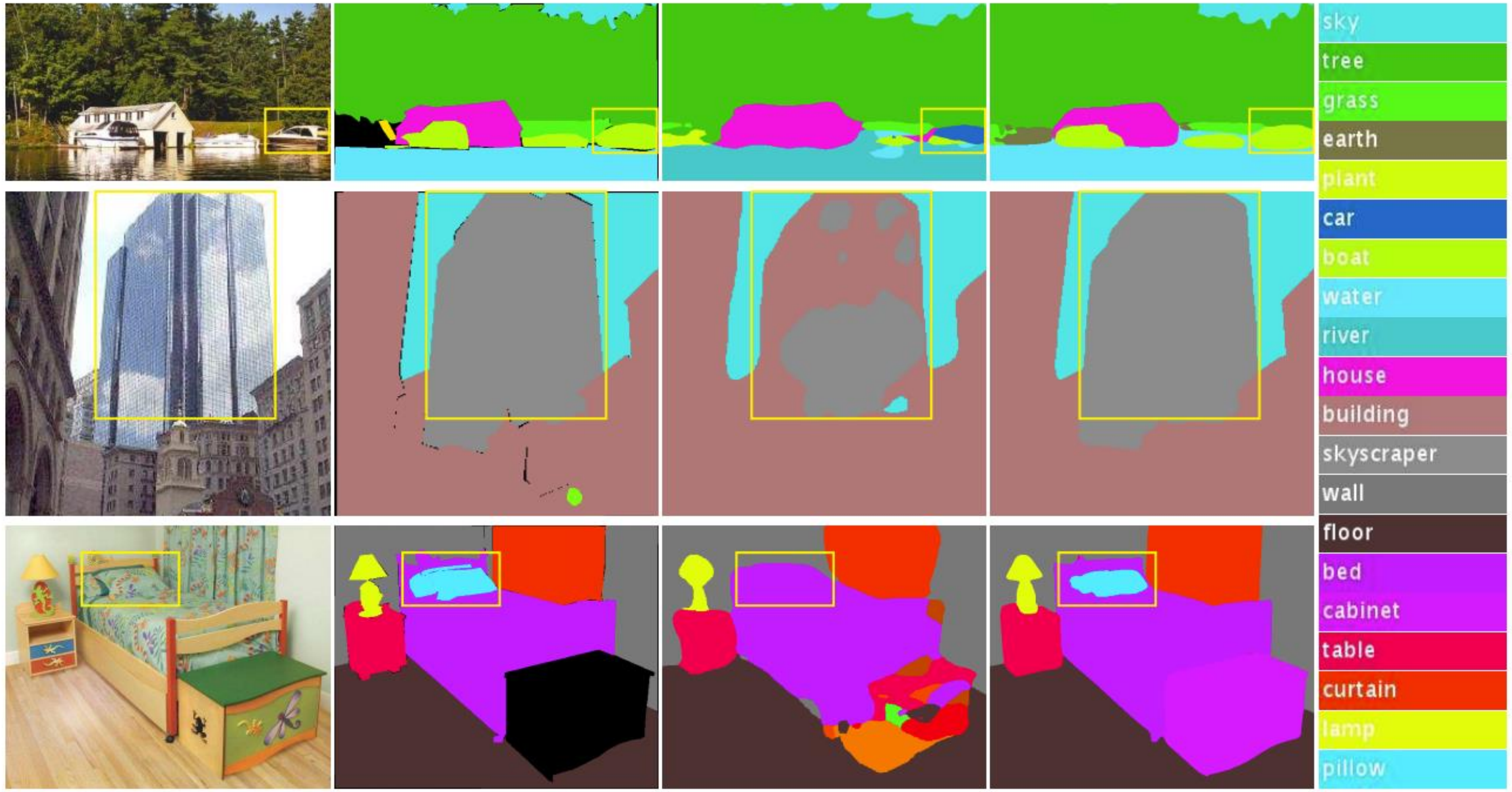
$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j)\left[w_1 \exp\left(-\frac{||p_i - p_j||^2}{2\sigma_\alpha^2} - \frac{||I_i - I_j||^2}{2\sigma_\beta^2}\right) + w_2 \exp\left(-\frac{||p_i - p_j||^2}{2\sigma_\gamma^2}\right)\right]$$

$$minE(x)$$

03 PSPNet

contextual relationship and global information
全局信息和语境关系

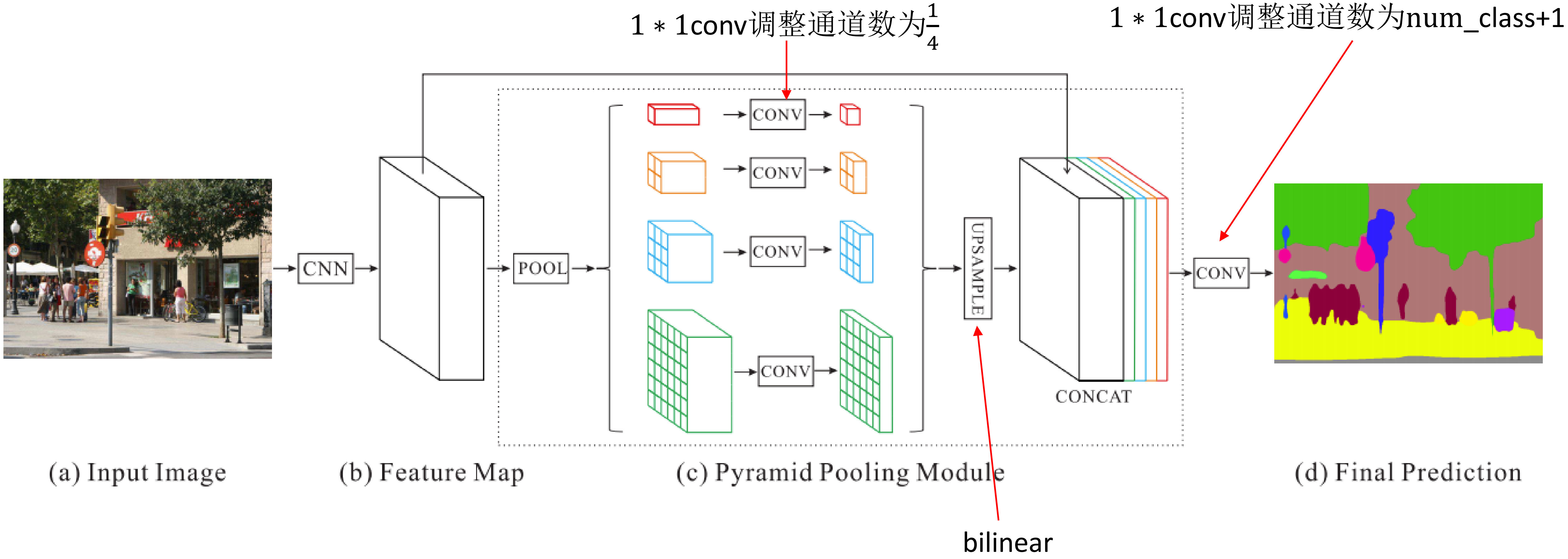
- Mismatched Relationship
  语义信息不匹配
- Confusion Categories
  类别混淆
- Inconspicuous Classes
  不明显类别



(a) Image (b) Ground Truth (c) FCN (d) PSPNet (e) ColorMap

03 PSPNet

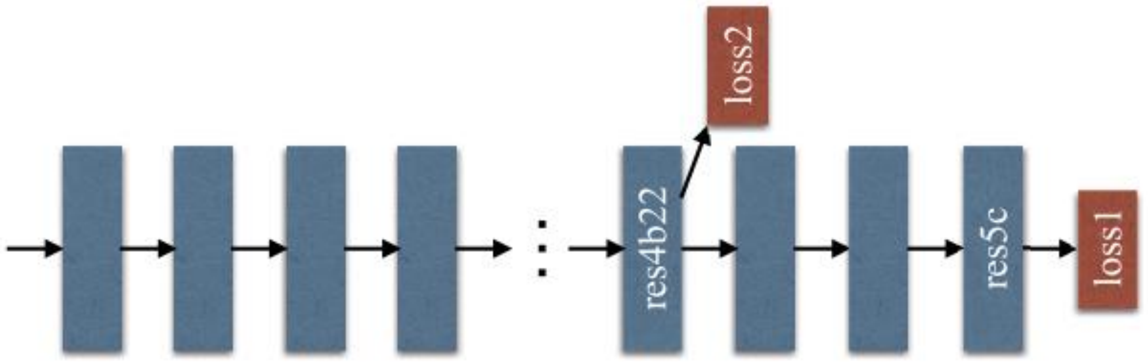
- Pyramid Pooling Module 金字塔池化



03 PSPNet

- Train

两个loss一起传播，使用不同的权重，共同优化参数



04 总结

- 两大方向
  - Multiscale Combining
    - Skip layer :不同层信息融合
    - ASPP：不同rate的空洞卷积并行进行融合
    - PSPNet:金字塔池化，将局部信息和全局信息融合
  - Structure Predition
    - CRF