法律声明

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

- 课程详情请咨询
  - 微信公众号：小象
  - 新浪微博：ChinaHadoop

第4课　图像检测（下）

Image Detection

主讲人：张宗健

悉尼科技大学博士

主要研究方向：　计算机视觉、视觉场景理解、图像&语言、深度学习

图像检索CbIR、Human ReID等

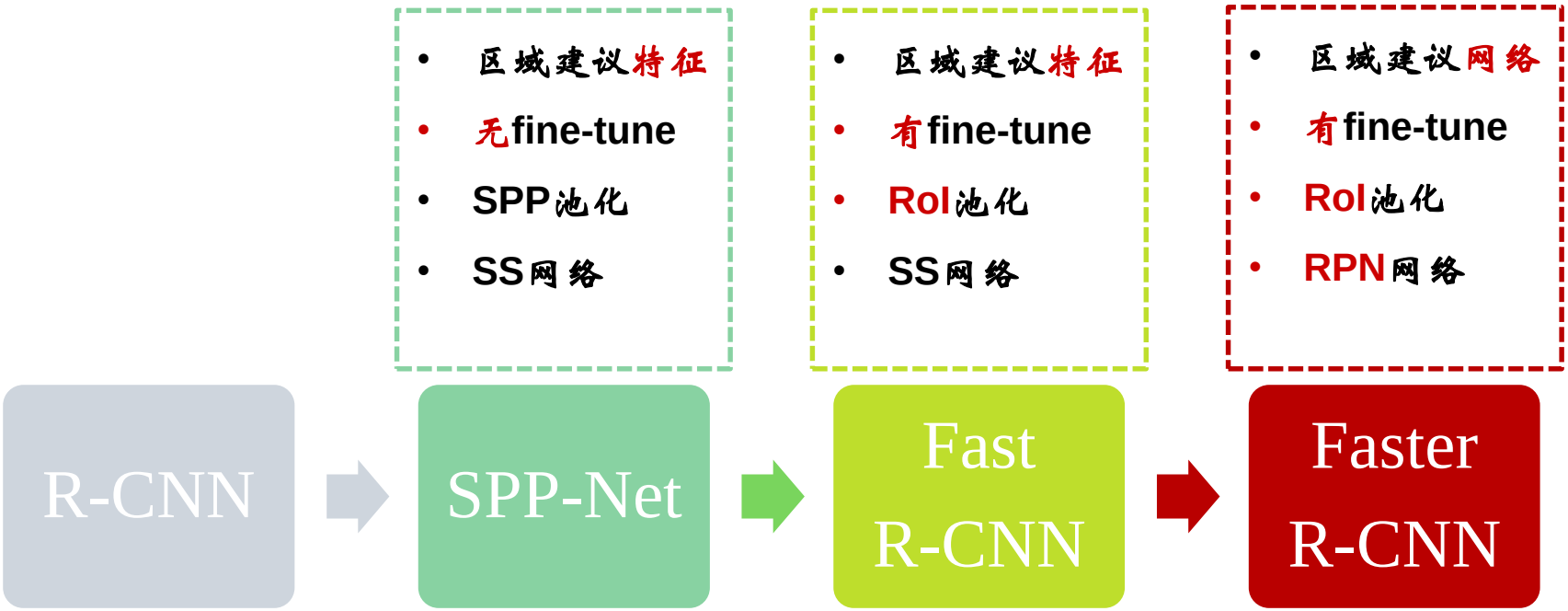
本章结构

- 区域卷积神经网络（R-CNN）系列
  - R-FCN
- 应用案例：
  - Faster R-CNN

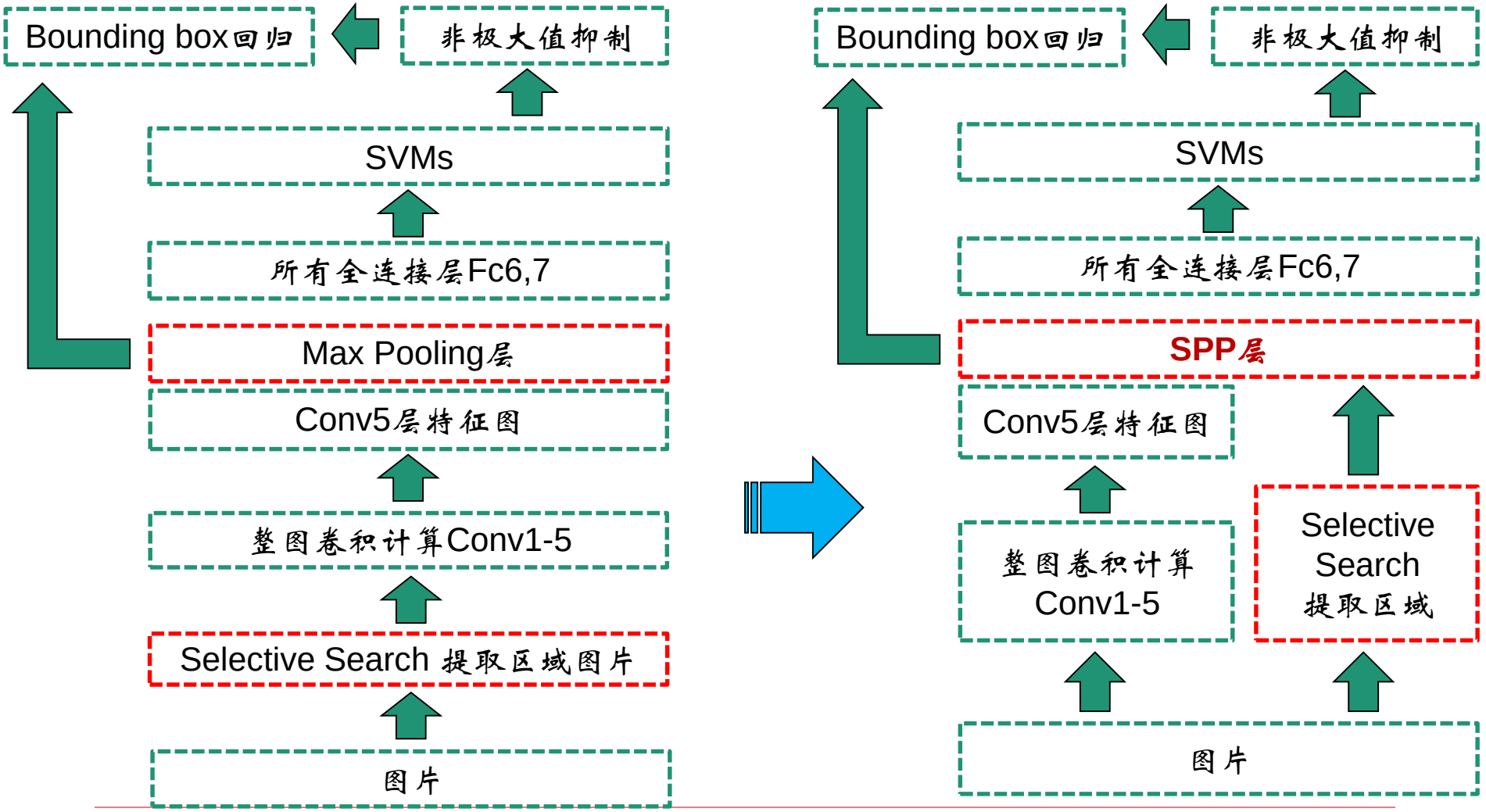
R-CNN回顾

模型进化

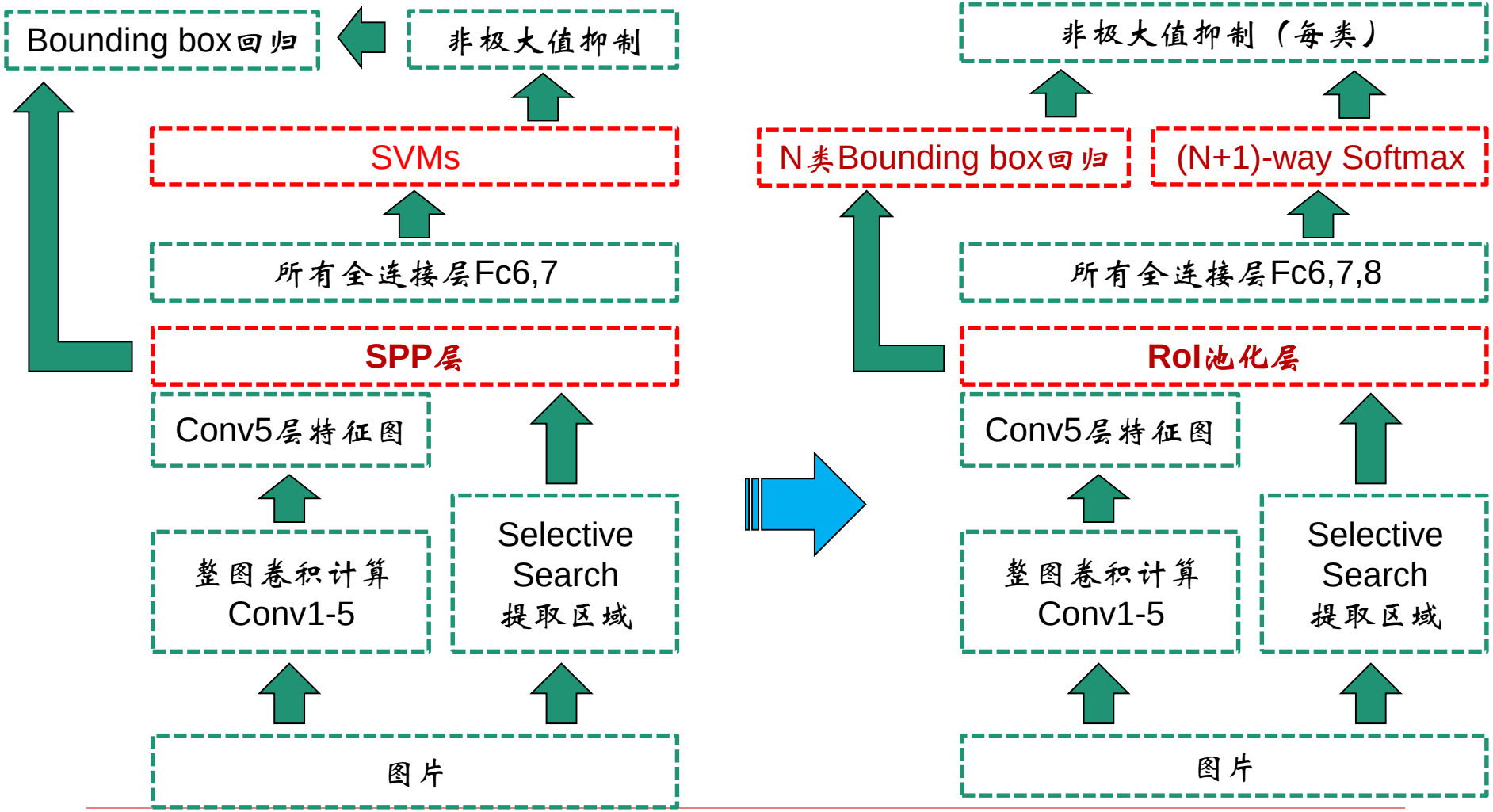
- 共享卷积层计算



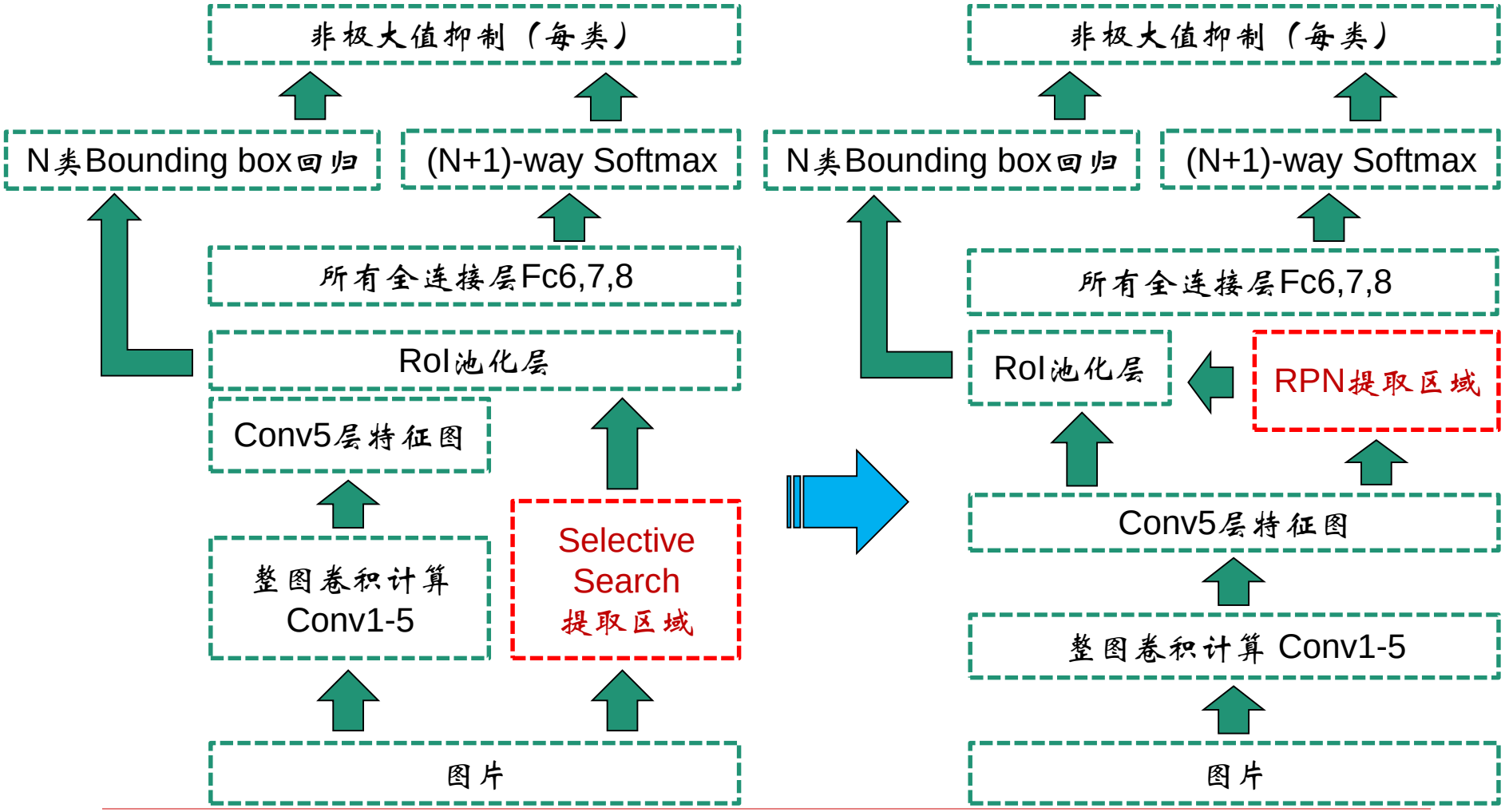
R-CNN vs SPP-Net



SPP-Net vs Fast R-CNN



Fast R-CNN vs Faster R-CNN



区域全卷积神经网络R-FCN

R-CNN系列的结构

- 基于旧形态CNN的结构（AlexNet，VGG）
  - 全卷积子网络
  - 全连接子网络
- 相对应的结构设计
  - 全卷积子网络（5层/组）
    - 独立于RoI
    - 计算共享
  - RoI-wise子网络（3层）
    - 计算无法共享

区域全卷积神经网络R-FCN

R-FCN（Fully Convolutional Network）

- CNN的全卷积化趋势（ResNet，GoogLeNet）
  - 只剩1个全连接层（2048→1000）
- 相应的，基于旧结构设计的R-CNN会出现问题
  - 结构：RoI-wise子网络无隐含层
  - 性能：检测性能跟分类性能不一致
  - 应用两难：
    - 检测网络的变换敏感性（Translation variance）
    - 分类网络的变换不变性（Translation invariance）
      - 卷积层越深，不变性越强，对变换越不敏感
  - 不适应设计：ResNet-101→ Conv91 + RoI池化 + Conv10
    - 准确率提升，但速度下降

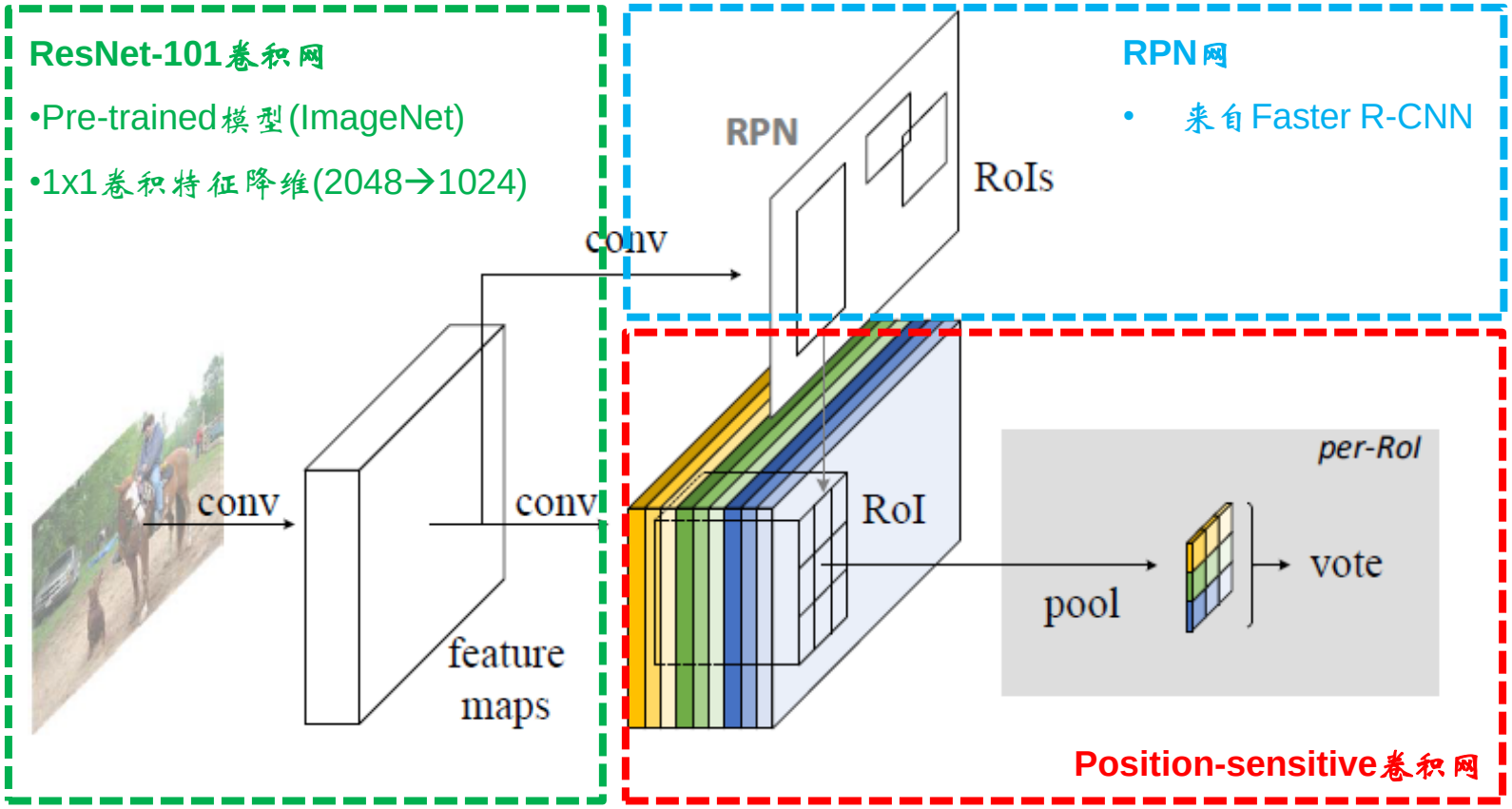
区域全卷积神经网络R-FCN

R-FCN（Fully Convolutional Network）

- 适应全卷积化CNN的结构，提出全卷积化设计
  - 共享ResNet的所有卷积层
  - 引入变换敏感性（Translation variance）
    1. 位置敏感分值图（Position-sensitive score maps）
       - 特殊设计的卷积层
       - Grid位置信息 + 类别分值
    2. 位置敏感池化（Position-sensitive RoI pooling）
       - 无训练参数
       - 无全连接网络的类别推断

区域全卷积神经网络R-FCN

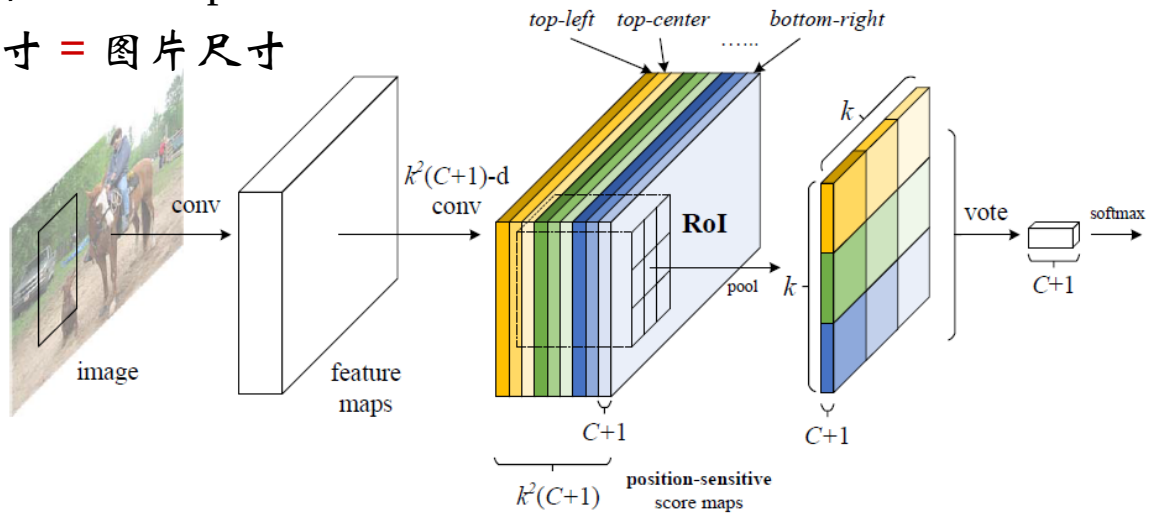
R-FCN结构



区域全卷积神经网络R-FCN

R-FCN的位置敏感卷积层

- 使用$k^2(C+1)$个通道对（位置，类别）组合进行编码
  - 类别：C个物体类+1个背景类
  - 相对位置：k x k个Grid（k=3）
  - 位置敏感分值图（Position-sensitive score maps）
    - 每个分类k^2个score map
    - score map尺寸 = 图片尺寸



区域全卷积神经网络R-FCN

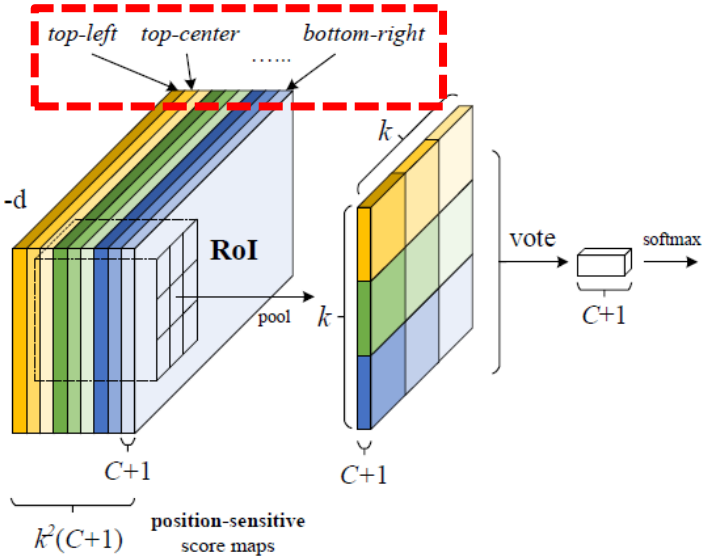
R-FCN的位置敏感RoI池化层

- 显式地编码相对位置信息
  - 将 $w \times h$ 尺寸的RoI拆分成k x k个 $\frac{w}{k} \times \frac{h}{k}$ 尺寸的bin
  - 不同(颜色)bin对应不同(颜色)通道层（score map）
  - Bin内做均值池化
  - 输出尺寸 k x k x (C+1)

$$r_c(i,j \mid \Theta) = \sum_{\substack{(x,y)\in \text{bin}(i,j) \\ (0 \le i,j \le k-1)}} z_{i,j,c}(x + x_0, y + y_0 \mid \Theta)/n$$

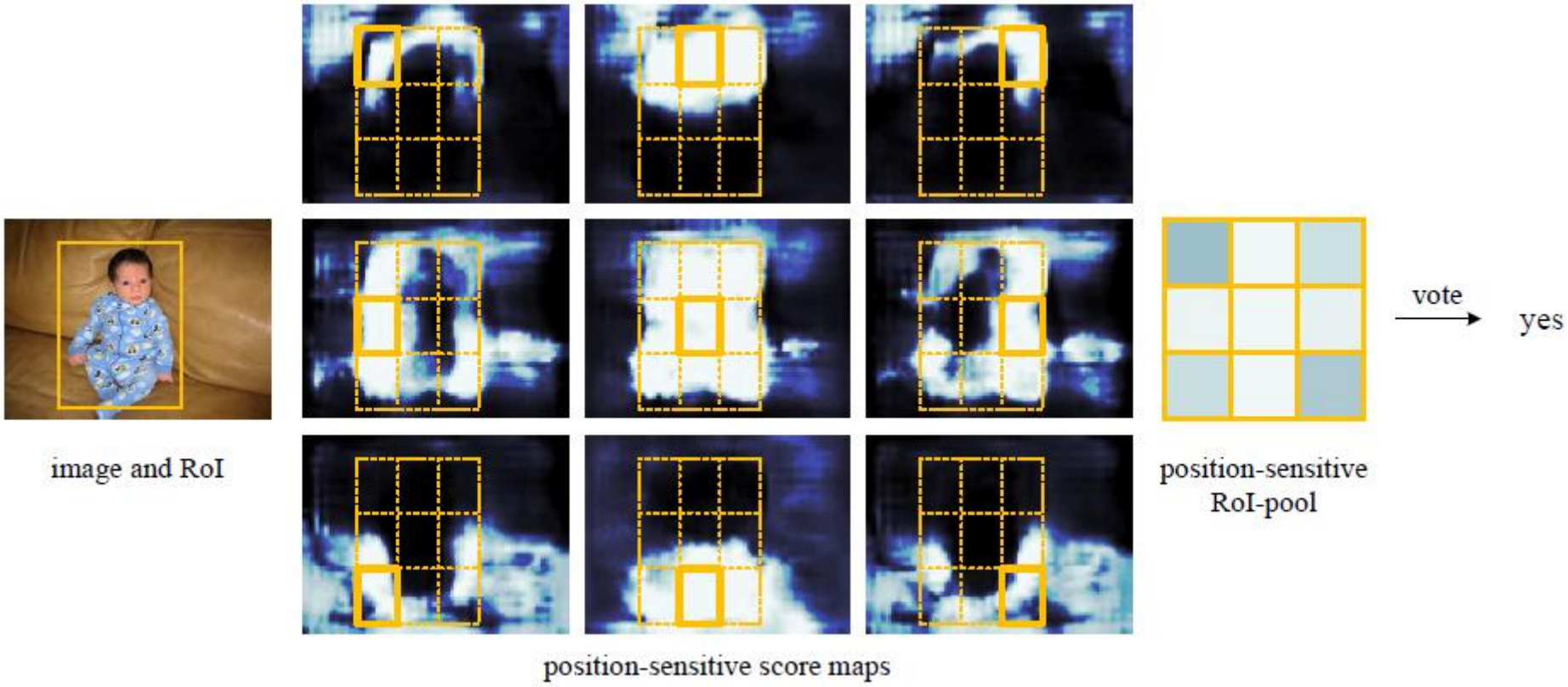
(x_0, y_0) 为RoI左上角坐标

$\lfloor i\frac{w}{k} \rfloor \le x < \lceil (i+1)\frac{w}{k} \rceil$ and $\lfloor j\frac{h}{k} \rfloor \le y < \lceil (j+1)\frac{h}{k} \rceil$



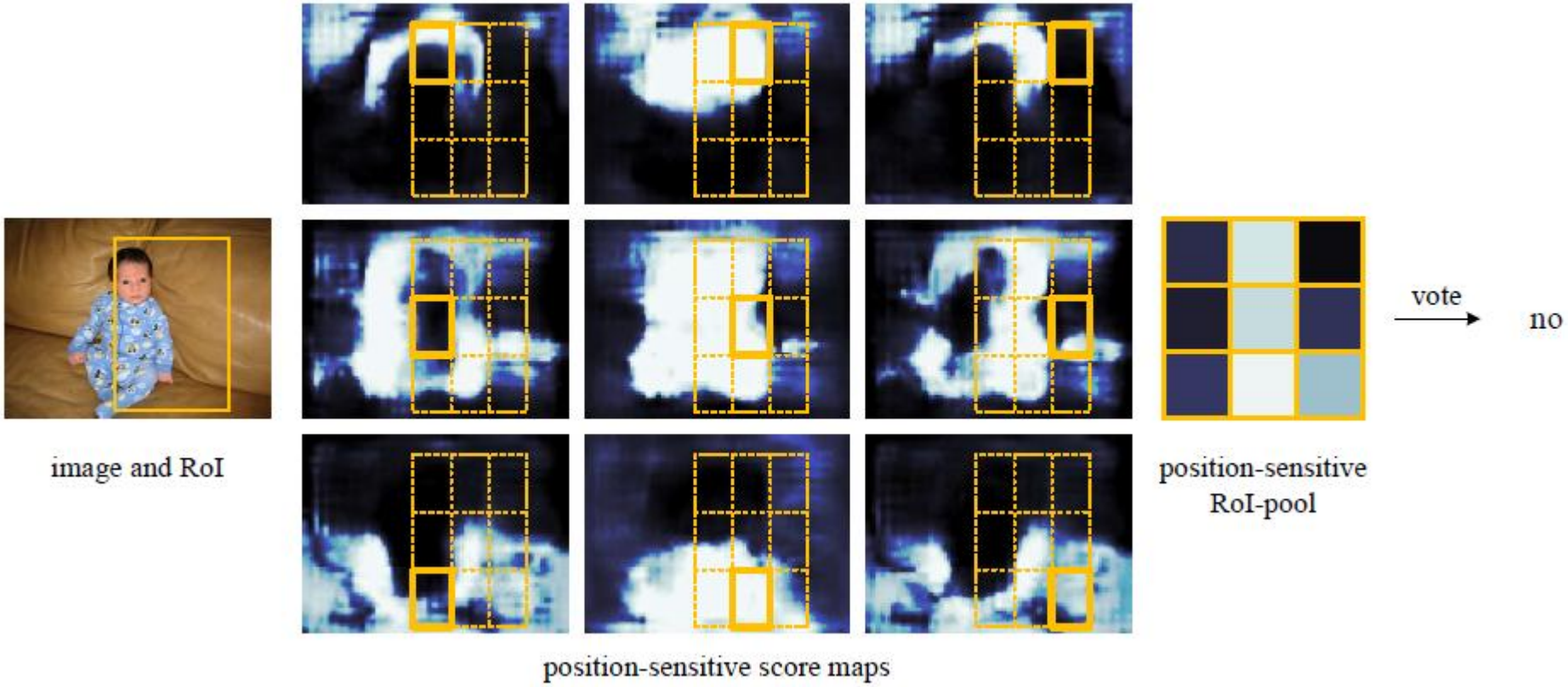
区域全卷积神经网络R-FCN

R-FCN的Score map可视化（person类别）



区域全卷积神经网络R-FCN

R-FCN的Score map可视化（person类别）

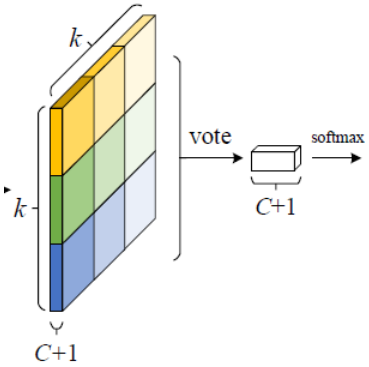


区域全卷积神经网络R-FCN

R-FCN的多任务损失函数

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

- 分类损失函数 $L_{cls}(s_{c^*}) = -\log(s_{c^*})$
  - 对池化输出计算k^2区域上的均值投票 $r_c(\Theta) = \sum_{i,j} r_c(i, j \mid \Theta)$
  - Softmax归一化 $s_c(\Theta) = e^{r_c(\Theta)} / \sum_{c'=0}^{C} e^{r_{c'}(\Theta)}$
- Bounding box回归损失函数 $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$
  - $4k^2$通道的位置敏感卷积层
  - 使用位置敏感RoI池化获取4 x k x k输出
  - 在k x k区域上均值投票 $t = (t_x, t_y, t_w, t_h)$
    - x 对应预测框
    - x_a 对应anchor框
    - x^* 对应ground truth框
    - y, w, h同理
  - 每个分类一个回归模型



$$t_x = (x - x_a)/w_a, \quad t_y = (y - y_a)/h_a,$$
$$t_w = \log(w/w_a), \quad t_h = \log(h/h_a),$$
$$t_x^* = (x^* - x_a)/w_a, \quad t_y^* = (y^* - y_a)/h_a,$$
$$t_w^* = \log(w^*/w_a), \quad t_h^* = \log(h^*/h_a),$$

区域全卷积神经网络R-FCN

R-FCN训练

- OHEM（Online Hard Example Mining）
  - 1个图片→1个Batch→1个GPU
    - 一个图片生成N个区域建议
    - 使用当前网络计算所有区域的loss
    - 根据loss从大到小排序区域建议
    - 使用前B=128个作为Batch数据
  - 8GPU并行→8x Batch size
- Faster R-CNN的4步训练法
  - 2轮：RPN跟R-FCN交替训练

区域全卷积神经网络R-FCN

R-FCN位置敏感的性能

- 基于ResNet-101
- 位置敏感性带来大幅提升

method	RoI output size ($k \times k$)	mAP on VOC 07 (%)
naïve Faster R-CNN	1×1	61.7
	7×7	68.9
R-FCN (w/o position-sensitivity)	1×1	*fail*
R-FCN	3×3	75.5
	7×7	**76.6**

区域全卷积神经网络R-FCN

性能对比R-FCN vs Faster R-CNN

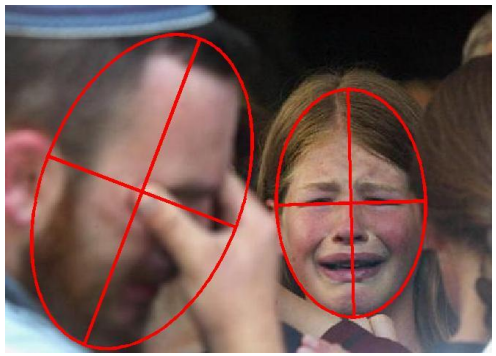
- 基于ResNet-101
- OHEM没有带来时间开销
- k x k = 7 x 7
- Test时间提升2.5x

	depth of per-RoI subnetwork	training w/ OHEM?	train time (sec/img)	test time (sec/img)	mAP (%) on VOC07
Faster R-CNN	10		1.2	0.42	76.4
R-FCN	0		0.45	0.17	76.6
Faster R-CNN	10	✓ (300 RoIs)	1.5	0.42	79.3
R-FCN	0	✓ (300 RoIs)	0.45	0.17	**79.5**
Faster R-CNN	10	✓ (2000 RoIs)	2.9	0.42	*N/A*
R-FCN	0	✓ (2000 RoIs)	0.46	0.17	79.3

人脸检测

FDDB - Face Detection Data set and Benchmark

- 2845张图片/5171张人脸
- 椭圆标注
- 灰度图/彩色图
- 检测难点
  - 遮挡（Occlutions）
  - 不同姿态（Different poses）
  - 低分辨率（Low resolution）
  - 失焦（Out-of-focus）
- Url
  - http://vis-www.cs.umass.edu/fddb/



人脸检测

WIDER FACE

- 32,203张图片/393,703张人脸
- 来自61个事件类
- 检测难点
  - 不同尺度（Different scales）
  - 遮挡（Occlutions）
  - 不同姿态（Different poses）
- Url
  - http://mmlab.ie.cuhk.edu.hk/projects/WIDERFace/

人脸检测

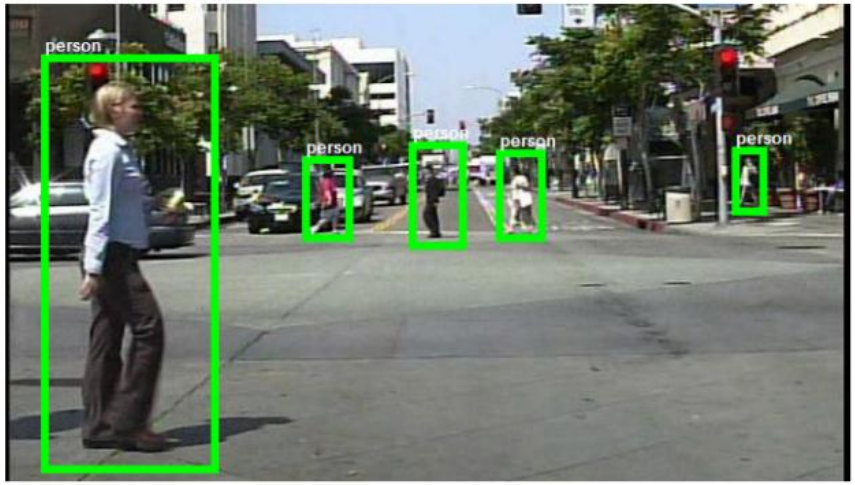
IJB-A - IARPA Janus Benchmark A

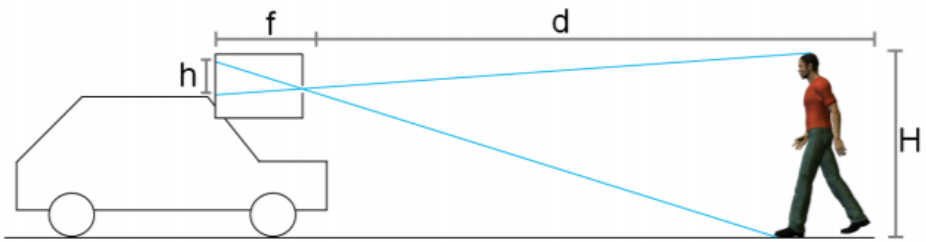
- 5,712张图片/2,085个视频/49,759张人脸
- 识别
  - 500个人
  - Meta data: 性别、肤色、姿态等
- 无约束的人脸数据集
- Url
  - https://www.nist.gov/itl/iad/image-group/ijba-dataset-request-form

行人检测

Caltech

- 10小时640x380视频/ 250,000帧
- 车载摄像头拍摄
- 350,000行人bounding box
- Url
  - http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/





演示环节

- Github
  - https://github.com/349zzjau
- 百度网盘
  - http://pan.baidu.com/s/1gfpCCwj

疑问

- 问题答疑： http://www.xxwenda.com/
  - 可邀请老师或者其他人回答问题

Q & A

小象账号：349zzjau

课程名：基于深度学习的计算机视觉

课后调查问卷http://cn.mikecrm.com/JtIE8KR

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：小象
- 新浪微博：ChinaHadoop