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（上）

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

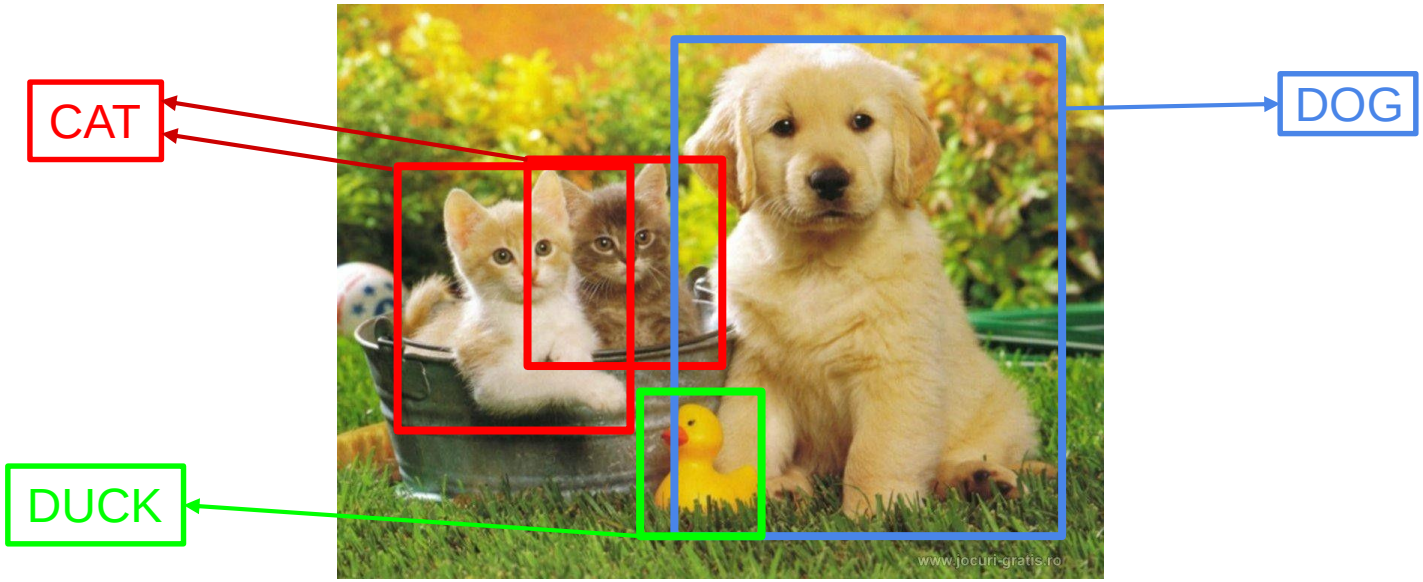
本章结构

- 物体检测（Object detection）
- ILSVRC竞赛
- 区域卷积神经网络（R-CNN）系列
- 行人检测&人脸检测
- 应用案例：
  - 人脸检测的Faster R-CNN应用

物体检测

检测图片中所有物体的

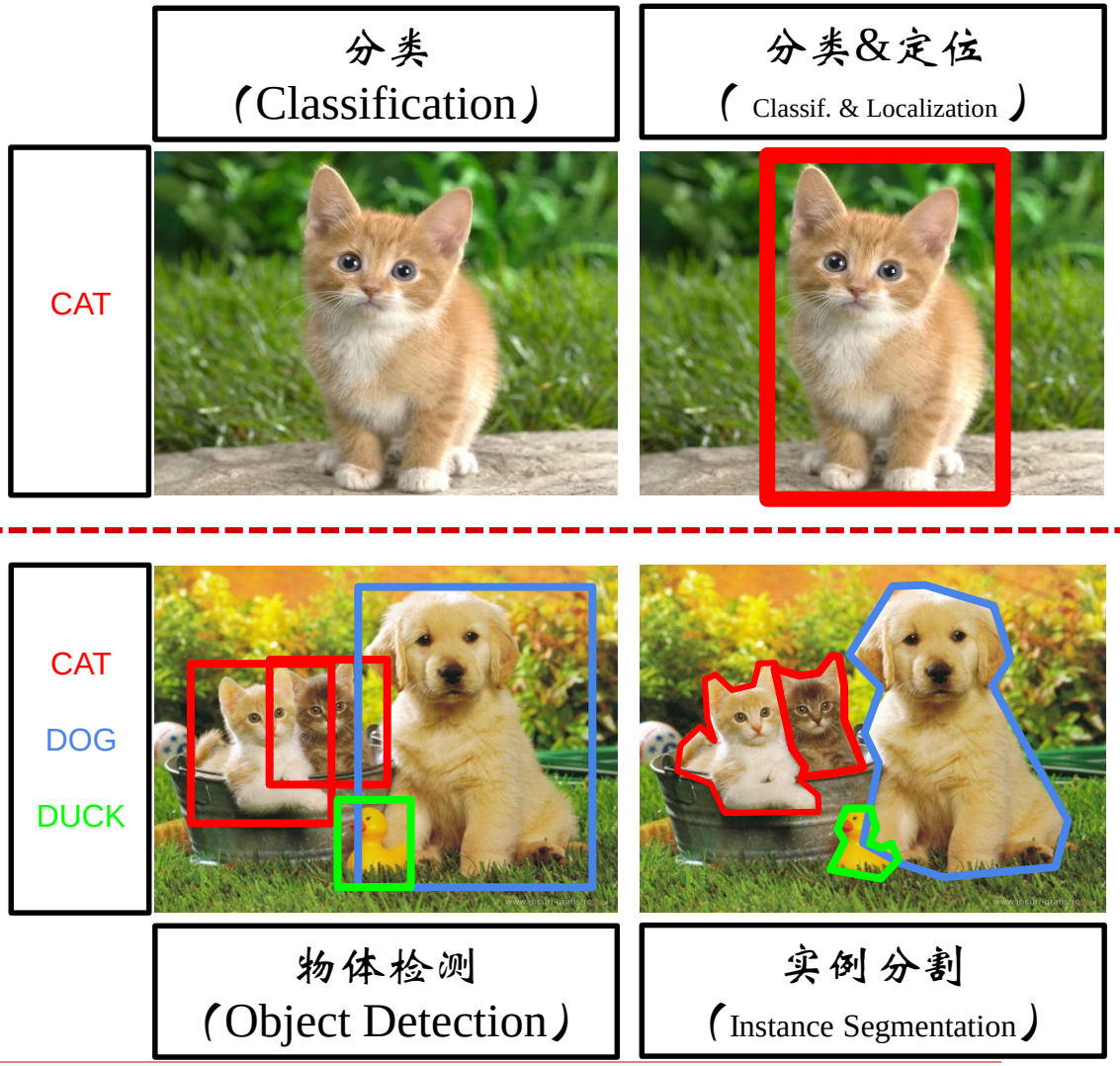
- 类别标签（Category label）
- 位置（最小外接矩形/Bounding box）



物体检测

与其他任务的区别

- 单例任务
  - 分类
  - 分类&定位
- 多例任务
  - 物体检测
  - 实例分割



ILSVRC竞赛

Imagenet Large Scale Visual Recognition Challenge

- 物体检测（Object Detection）
  - 竞赛历史：2013-2017
  - 物体类别：200
  - 每个图片多组标签
    - 类别+Bounding box(x, y, w, h)
- 其他任务
  - 图像分类（Image Classification ）
  - 场景分类（Scene Classification）
  - 物体定位（Object Localization）
  - 场景解析（Scene parsing）
- URL： http://image-net.org/challenges/LSVRC/2017/index

ILSVRC竞赛

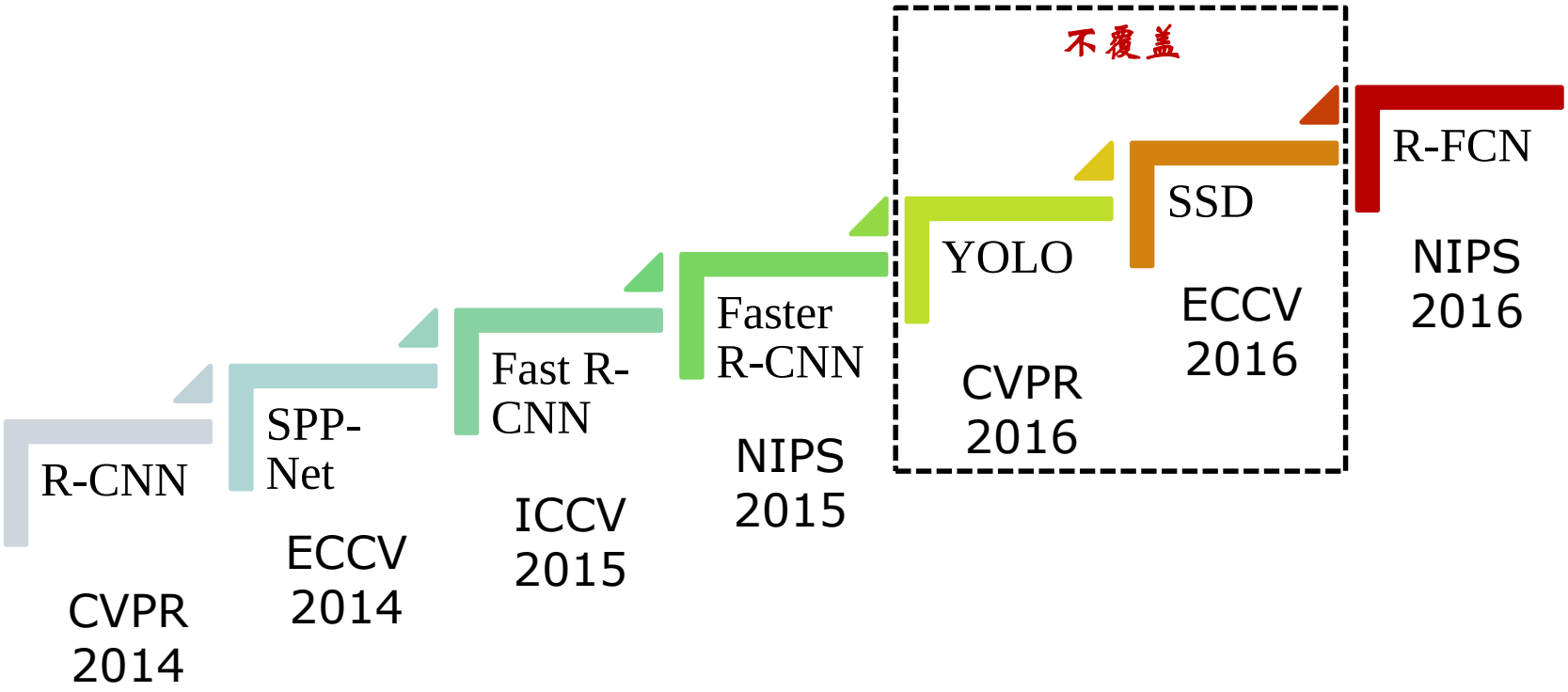
PASCAL VOC

- 竞赛历史: 2005-2012
- URL: http://host.robots.ox.ac.uk/pascal/VOC/

		PASCAL VOC 2012	ILSVRC
	物体类别	20	200
训练集 Training	图片数量	5717	456567
	标注数量	13609	478807
验证集 Validation	图片数量	5823	20121
	标注数量	13841	55502
测试集 Testing	图片数量	10991	40152
	标注数量	-	-

区域卷积神经网络R-CNN

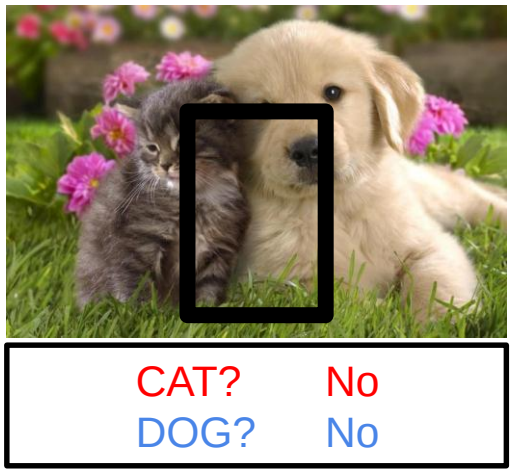
模型进化

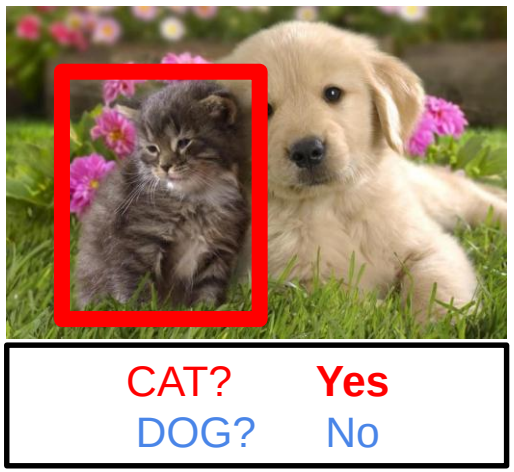


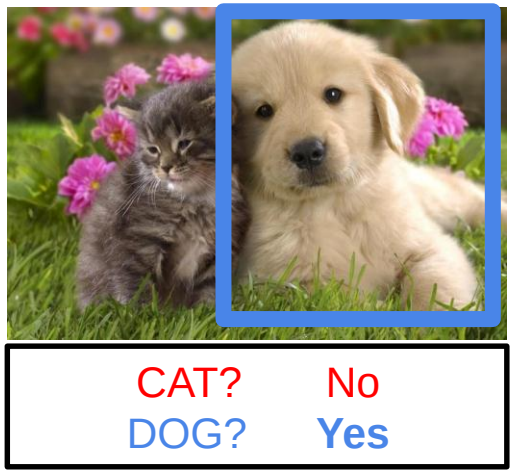
区域卷积神经网络R-CNN

模型结构

- 按分类问题对待
  - 模块1：提取物体区域（Region proposal）
    - 不同位置，不同尺寸，数量很多
  - 模块2：对区域进行分类识别（Classification）
    - CNN分类器，计算量大



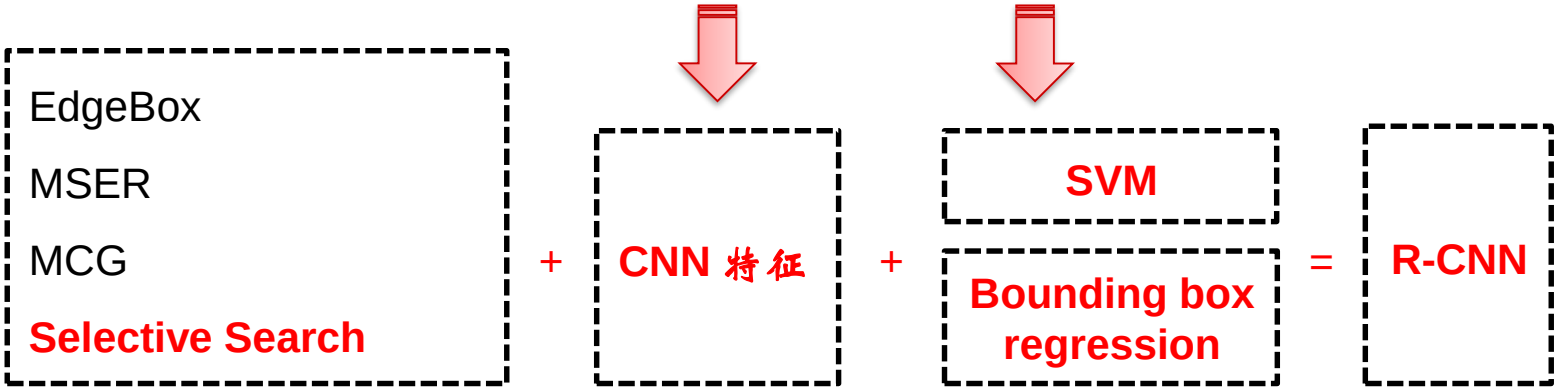


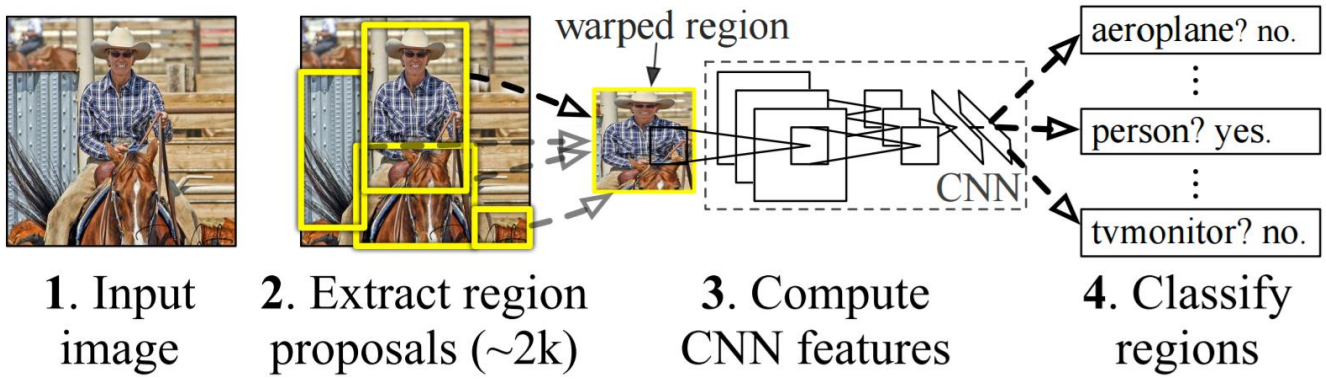


区域卷积神经网络R-CNN

传统方法 → R-CNN

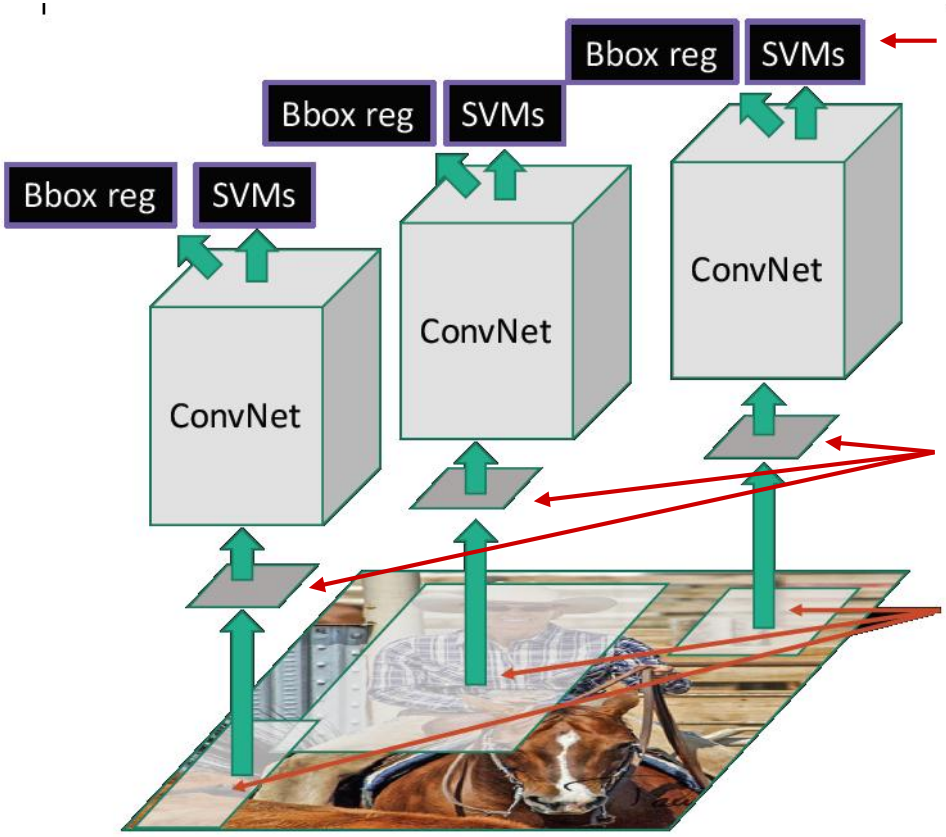
- Region proposals + 手工特征 + 分类器





区域卷积神经网络R-CNN

R-CNN结构



模块4：Bounding box回归模型
- 对SS提供的区域进行精化
- 基于CNN特征
- 每个分类一个SVM

模块3：线性SVMs分类器
- 对CNN特征（4096）进行分类
- 每个分类一个SVM

模块2：AlexNet网络
- 对所有区域进行特征提取（Fc7）
- fine-tune

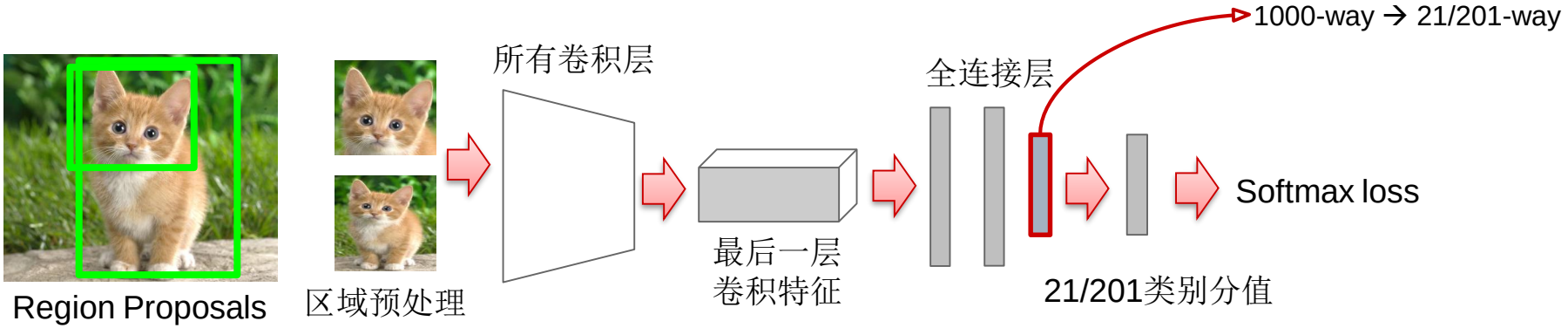
区域预处理
- Bounding box膨胀（16p）
- 尺寸变换成227x227

模块1：Selective Search(SS)获取区域
- ~2000个区域Region proposals
- 跟分类无关，包含物体

区域卷积神经网络R-CNN

R-CNN训练流程

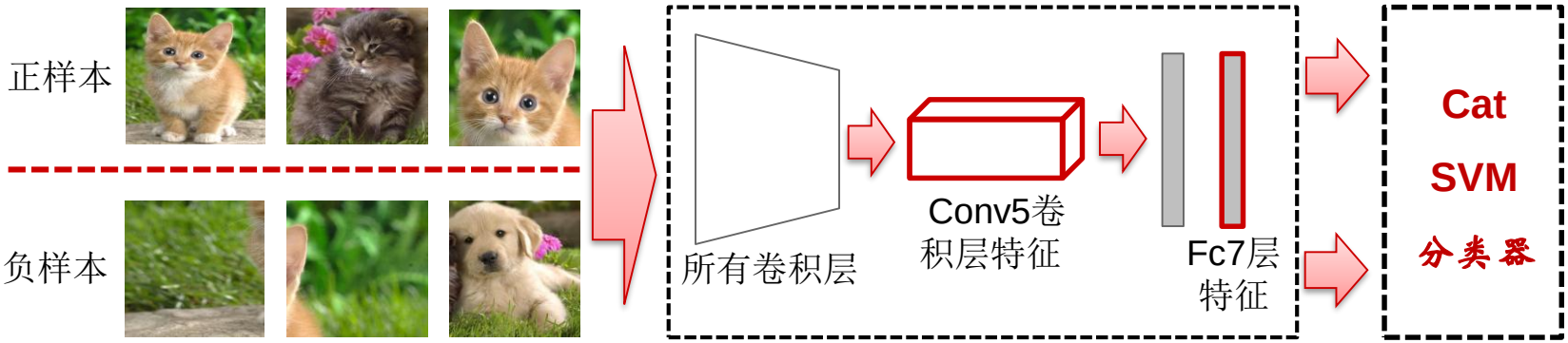
- M ← 在ImageNet上对CNN模型进行pre-train
- M'← 在SS生成的所有区域上对M进行fine-tune
  - Log loss
  - Softmax层改成(N+1)-way，其余不变
  - 32个正样本（N类）：跟Ground-truth重合IoU >= 0.5
  - 96个负样本（1 类）：IoU < 0.5



区域卷积神经网络R-CNN

R-CNN训练流程

- C←在M’的Fc7特征上训练线性SVMs分类器
  - Hinge loss
  - 每个类别（N类）对应一个SVM分类器
  - 正样本：所有Ground-truth区域
  - 负样本：跟Ground-truth重合IoU < 0.3的SS区域



区域卷积神经网络R-CNN

R-CNN训练流程

- R ← 在M'的Fc7特征上训练Bounding box回归模型
  - 提升定位性能（Bounding box的准确性）
  - 每个类别（N类）训练一个回归模型
    - 将SS提供的Bounding box做重新映射P→G
    - 训练输入
      - Bounding box对 $\{(P^i, G^i)\}_{i=1,\ldots,N}$
        - 中心位置（x, y）
        - 宽高尺寸（w, h）
      - CNN的Conv5特征 $\phi_5(P)$

$$P^i = (P^i_x, P^i_y, P^i_w, P^i_h)$$
$$G = (G_x, G_y, G_w, G_h)$$

$$\hat{G}_x = P_w d_x(P) + P_x$$
$$\hat{G}_y = P_h d_y(P) + P_y$$
$$\hat{G}_w = P_w \exp(d_w(P))$$
$$\hat{G}_h = P_h \exp(d_h(P))$$
$$d_\star(P) = \mathbf{w}_\star^{\mathrm{T}} \phi_5(P)$$

    - P的IoU > 0.6
    - Squared loss

$$\mathbf{w}_\star = \underset{\hat{\mathbf{w}}_\star}{\operatorname{argmin}} \sum_i^N (t^i_\star - \hat{\mathbf{w}}_\star^{\mathrm{T}} \phi_5(P^i))^2 + \lambda \|\hat{\mathbf{w}}_\star\|^2$$

$$t_x = (G_x - P_x)/P_w$$
$$t_y = (G_y - P_y)/P_h$$
$$t_w = \log(G_w/P_w)$$
$$t_h = \log(G_h/P_h).$$

    - 测试阶段
    - 参数w已经训练好

区域卷积神经网络R-CNN

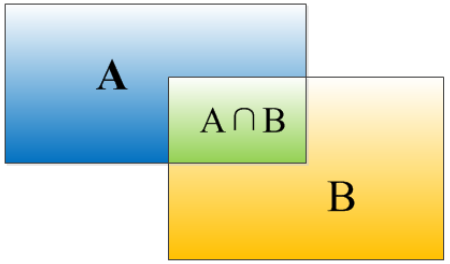
R-CNN测试阶段

- Selective Search（fast mode）提取~2000区域/图片
- 将所有区域膨胀+缩放到227x227
- 使用fine-tune过的AlexNet计算2套特征
  - 为每个类别执行
    - Fc7特征 → SVM分类器 → 类别分值
    - 使用非极大值抑制（IoU>=0.5）获取无冗余的区域子集
      - 所有区域按分值从大到小排序
      - 剔除冗余：与最大分值区域IoU>=0.5的所有区域
      - 保留该最大分值区域，剩余区域作为新候选集
    - Conv5特征 → Bounding box回归模型 → Bbox偏差
    - 使用Bbox偏差修正区域子集

区域卷积神经网络R-CNN

R-CNN性能评价

- mAP@0.5（mean Average Precision）
  - 给每一类分别计算AP，然后做mean平均
    - AP是Precision-Recall Curve下面的面积
    - 准确率precision: TP/(TP+FP)
    - 召回率recall: TP/(TP+FN)
      - True Positive区域：与Ground truth区域的IoU >= 0.5
      - False Positive区域： IoU < 0.5
      - False Negative区域：遗漏的Ground truth区域
      - IoU = Intersection over Unit
        = (A∩B)/(A∪B)
        = SI/(SA+SB-SI)



区域卷积神经网络R-CNN

R-CNN性能

- mAP大幅提升

数据集	Best baseline	R-CNN
PASCAL VOC 2010	35.1%	53.7%
ILSVRC	24.3%	31.4%

- 问题
  - 训练时间很长（84小时）
    - Fine-tune（18）+ 特征提取（63）+ SVM/Bbox训练（3）
  - 测试阶段很慢：VGG16一张图片47s
  - 复杂的多阶段训练

区域卷积神经网络R-CNN

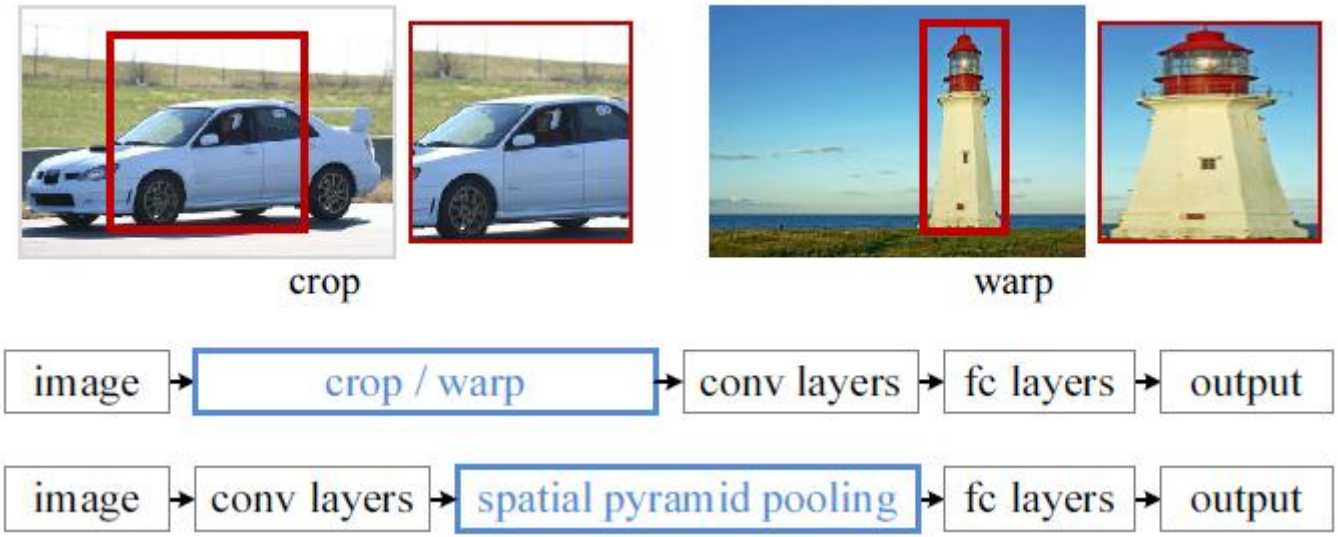
SPP-Net网络

- R-CNN速度慢的一个重要原因
  - 卷积特征重复计算量太大
    - 每张图片的~2000区域都会计算CNN特征
- 2大改进
  - 直接输入整图，所有区域共享卷积计算（一遍）
    - 在Conv5层输出上提取所有区域的特征
  - 引入空间金字塔池化（Spatial Pyramid Pooling）
    - 为不同尺寸的区域，在Conv5输出上提取特征
    - 映射到尺寸固定的全连接层上

区域卷积神经网络R-CNN

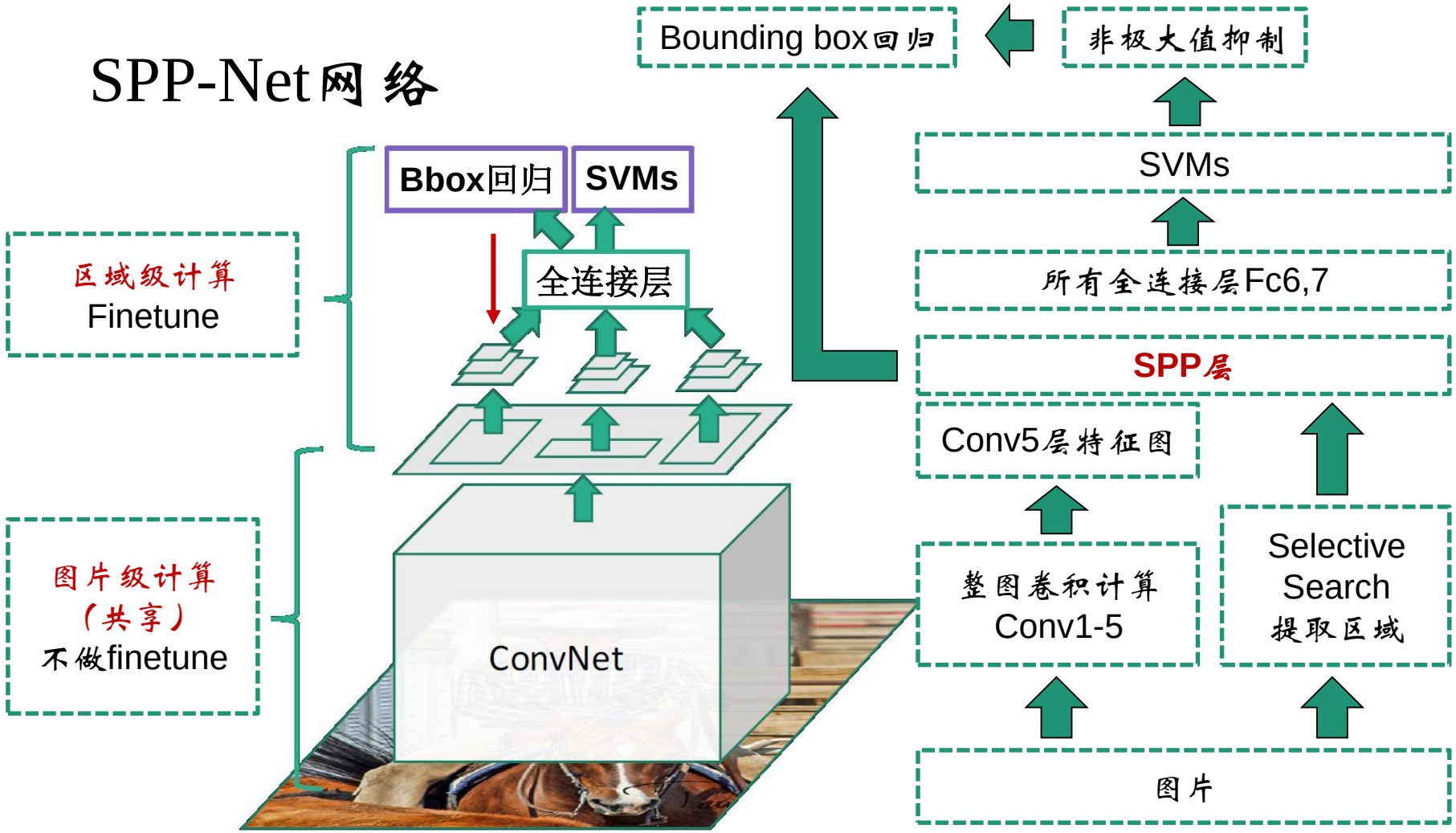
SPP-Net网络

- 使用SPP技术实现了
  - 共享计算
  - 适应不同输入尺寸



区域卷积神经网络R-CNN

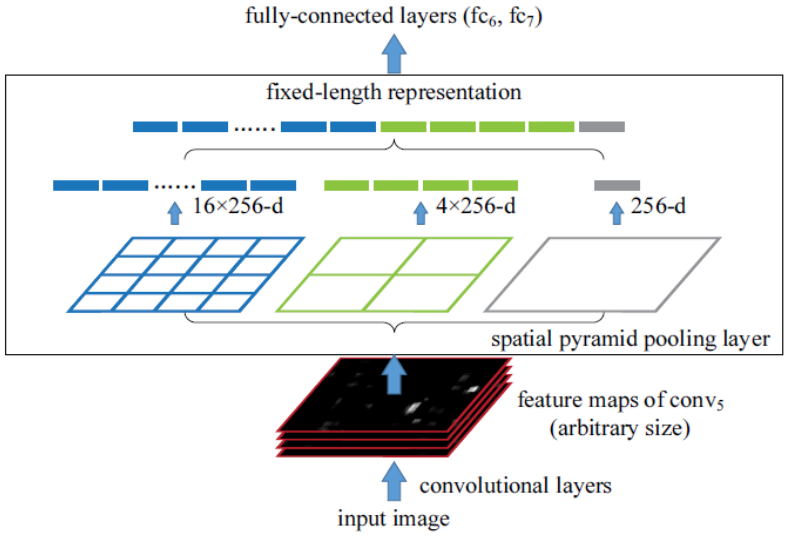
SPP-Net网络

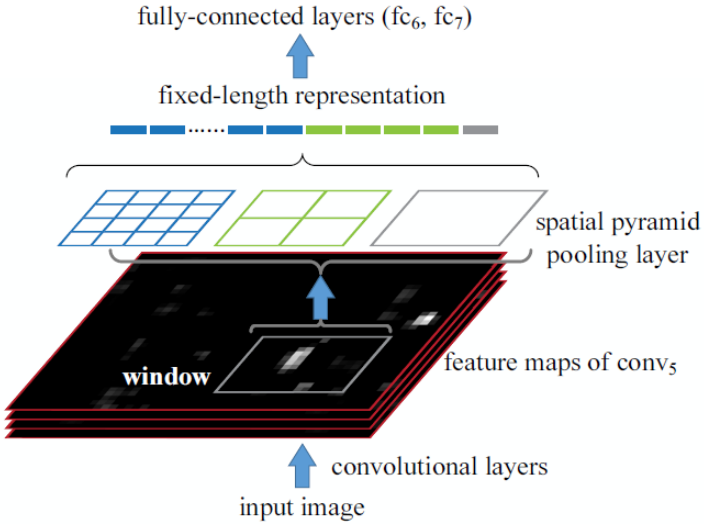


区域卷积神经网络R-CNN

SPP-Net网络

- 空间金字塔池化
  - 替换Conv5的Pooling层
  - 3个level和21个Bin：1x1, 2x2, 4x4
  - Bin内使用Max pooling





区域卷积神经网络R-CNN

SPP-Net训练流程

- M ← 在ImageNet上对CNN模型进行pre-train
- F ← 计算所有SS区域的SPP特征
- M’ ← 使用F特征finetune新fc6→fc7→fc8层

与R-CNN区别
- SPP特征←Pool5特征
- 只finetune全连接层

- F’ ← 计算M’的fc7特征
- C ← 使用F’特征训练线性SVM分类器
- R ← 使用F特征训练Bounding box回归模型

区域卷积神经网络R-CNN

SPP-Net问题

- 继承了R-CNN的剩余问题
  - 需要存储大量特征
  - 复杂的多阶段训练
  - 训练时间仍然长（25.5小时）
    - Fine-tune（16）+ 特征提取（5.5）+ SVM/Bbox训练（4）
- 新问题
  - SPP层之前的所有卷积层参数不能finetune

区域卷积神经网络R-CNN

Fast R-CNN网络

- 继承了R-CNN的剩余问题
  - 需要存储大量特征
  - 复杂的多阶段训练
  - 训练时间仍然长（25.5小时）
    - Fine-tune（16）+ 特征提取（5.5）+ SVM/Bbox训练（4）
- 新问题
  - SPP层之前的所有卷积层参数不能finetune

区域卷积神经网络R-CNN

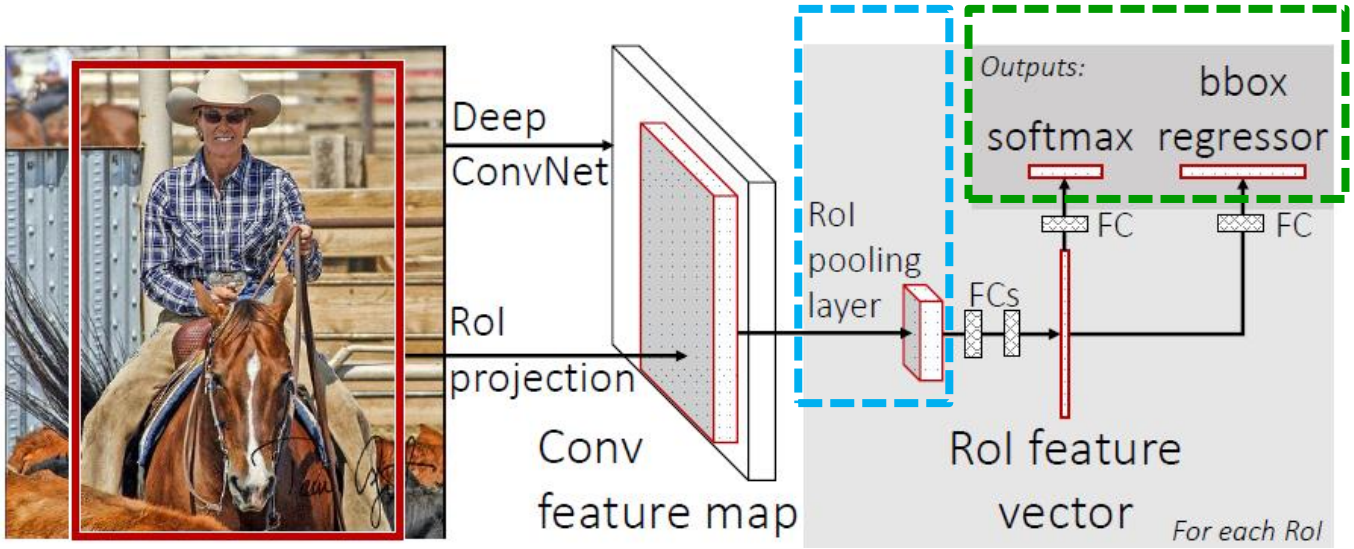
Fast R-CNN网络

- 改进
  - 比R-CNN，SPP-Net更快的trainng/test，更高的mAP
  - 实现end-to-end（端对端）单阶段训练
    - 多任务损失函数（Multi-task loss）
  - 所有层的参数都可以finetune
  - 不需要离线存储特征文件

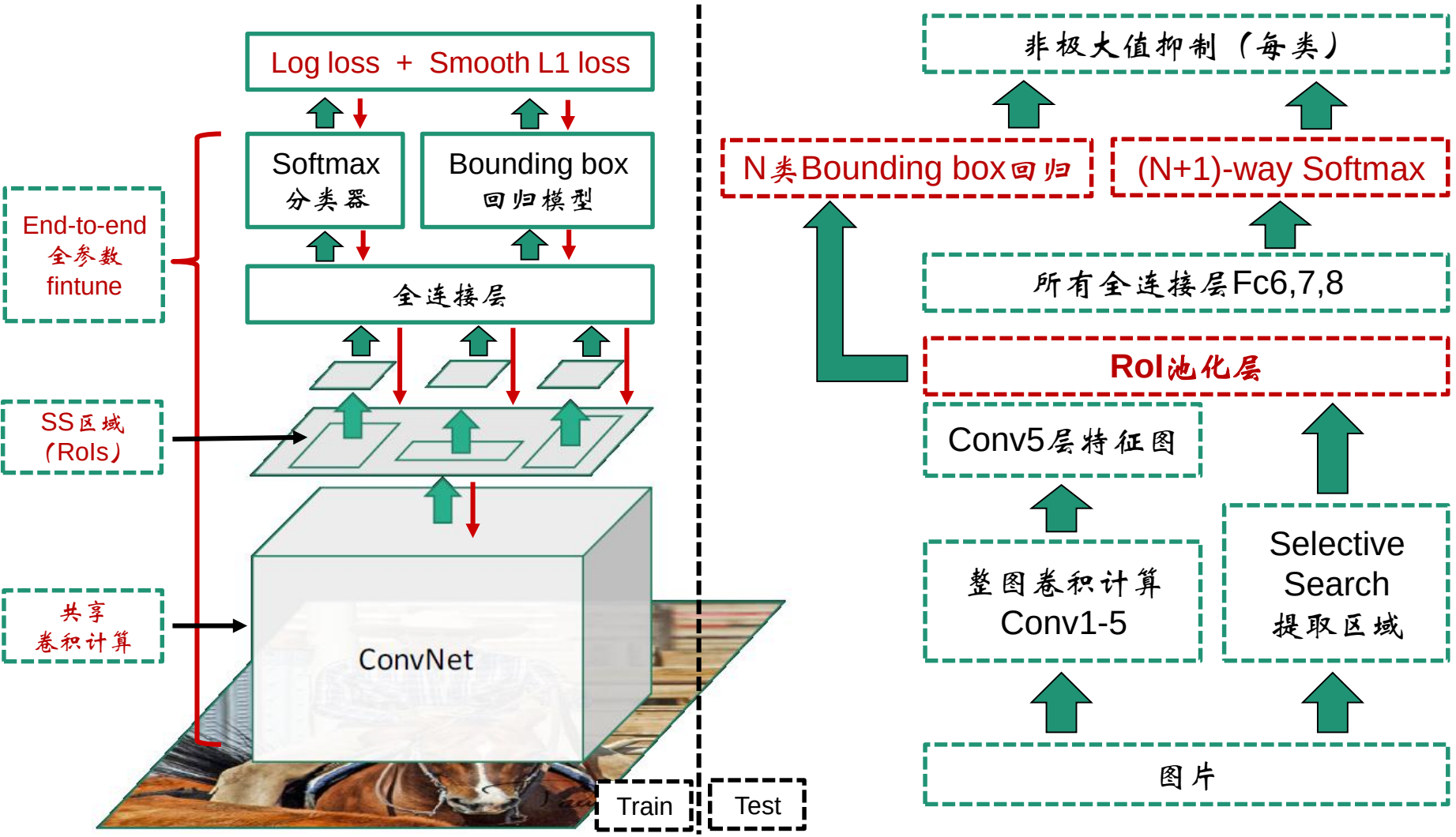
区域卷积神经网络R-CNN

Fast R-CNN网络

- 在SPP-Net基础引入2个新技术
  - 感兴趣区域池化层（RoI pooling layer）
  - 多任务损失函数（Multi-task loss）



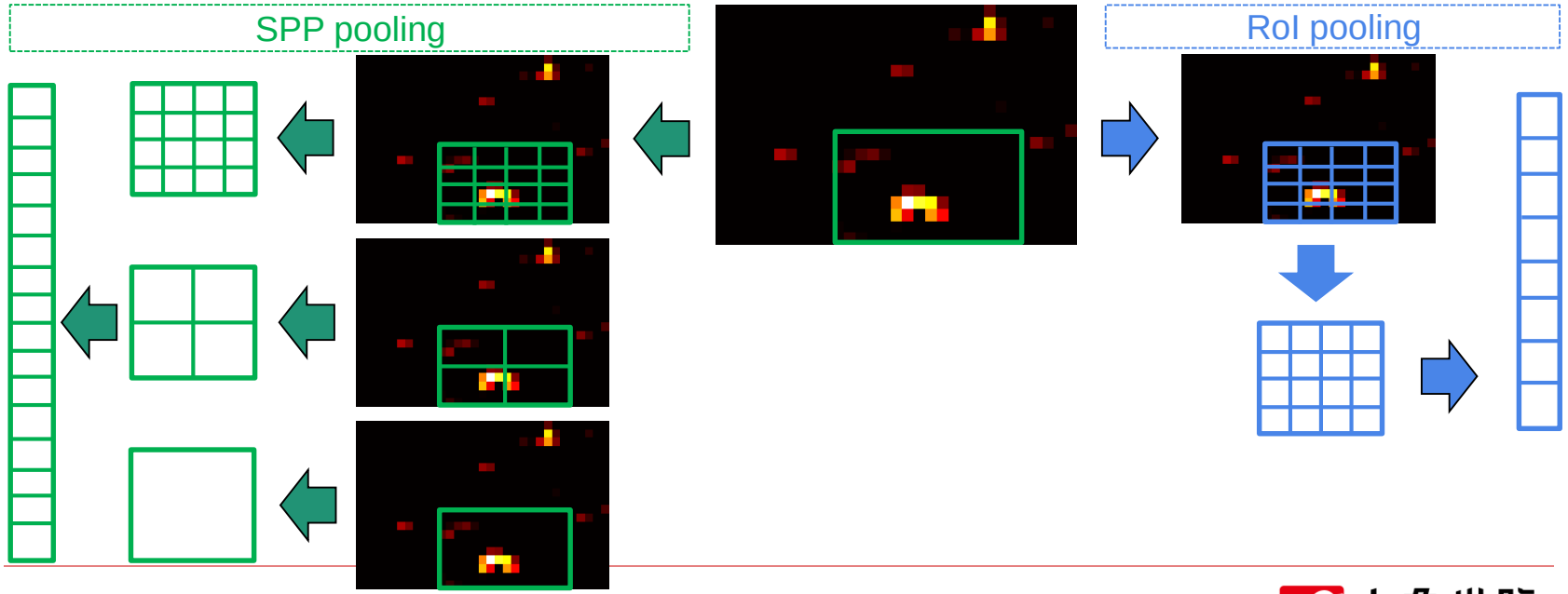
区域卷积神经网络R-CNN



区域卷积神经网络R-CNN

Fast R-CNN网络

- 感兴趣区域池化（RoI pooling）
  - 空间金字塔池化（SPP pooling）的单层特例
    - 将RoI区域的卷积特征拆分成H x W网格（7x7 for VGG）
    - 每个Bin内的所有特征进行Max pooling



区域卷积神经网络R-CNN

Fast R-CNN网络

- 感兴趣区域池化（RoI pooling）
  - 非重叠区域：类似Max pooling
  - 重叠区域
    - 多个区域的偏导之和

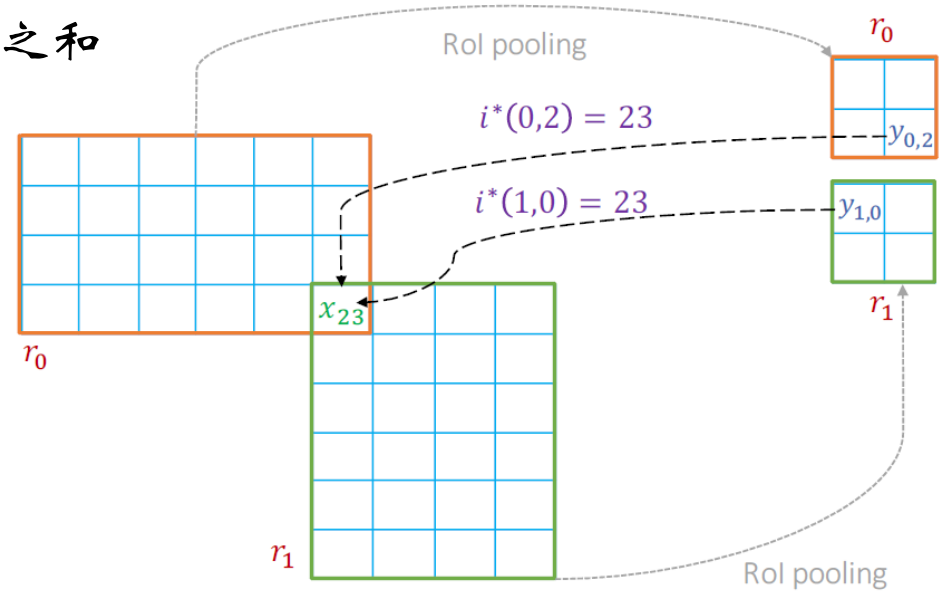
$$\frac{\partial L}{\partial x_i} = \sum_r \sum_j [i = i^*(r,j)] \frac{\partial L}{\partial y_{rj}}$$

1 if r, j "pooled" input i; 0 o/w

Partial for x_i

Over regions r, locations j

Partial from next layer



区域卷积神经网络R-CNN

Fast R-CNN网络

- 多任务损失（Multi-task loss）

$$L(p, u, t^u, v) = L_{\text{cls}}(p, u) + \lambda[u \geq 1]L_{\text{loc}}(t^u, v)$$

  - 分类器loss：$L_{\text{cls}}(p, u) = -\log p_u$
    - 每个RoI的概率分布 $p = (p_0, \ldots, p_K)$
    - Ground truth类别 u
  - Bounding box回归L1 loss: $L_{\text{loc}}(t^u, v) = \sum_{i \in \{x,y,w,h\}} \text{smooth}_{L_1}(t_i^u - v_i)$
    - 每个RoI共有N个loss（per-class）
    - 偏差目标 $v = (v_x, v_y, v_w, v_h)$
    - 预测偏差 $t^u = (t_x^u, t_y^u, t_w^u, t_h^u)$
    - 指示函数 $[u \geq 1]$
      - 物体类别：1，有回归loss
      - 背景类别：0，没有回归loss

高斯（0,1）归一化

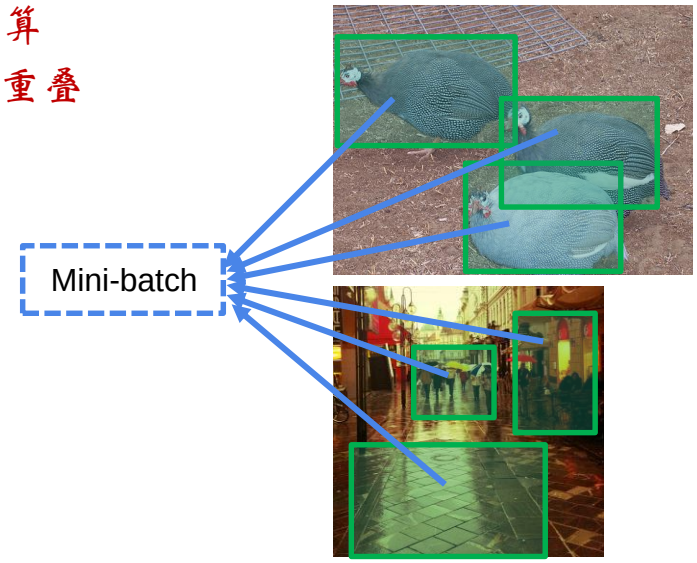
$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

$$\begin{aligned} t_x &= (G_x - P_x)/P_w \\ t_y &= (G_y - P_y)/P_h \\ t_w &= \log(G_w/P_w) \\ t_h &= \log(G_h/P_h). \end{aligned}$$

区域卷积神经网络R-CNN

Fast R-CNN网络

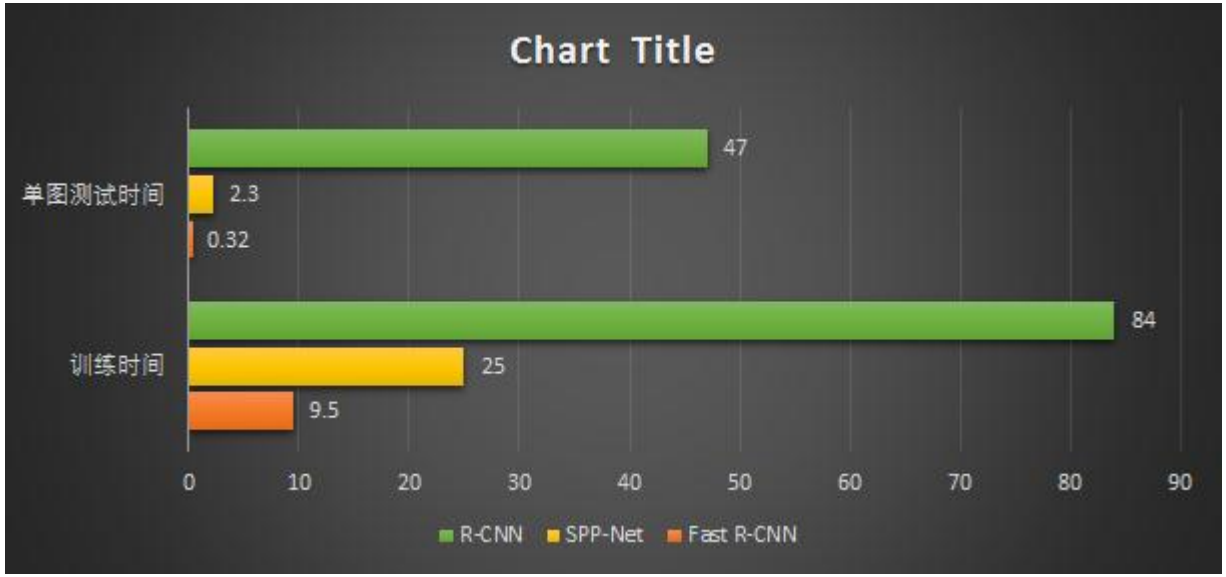
- 训练：在pre-trained模型上做finetune
  - Mini-batch sampling抽样
    - 分级抽样法（Hierarchical sampling）
      - Batch尺寸(128) = 每个batch的图片数量(2) x 每个图片的RoI数量(64)
      - 充分利用卷积层的共享计算
    - RoI分类基于与Ground truch的重叠
      - 物体：IoU >= 0.5
      - 背景：IoU在[0.1, 0.5)中
- 初始模型VGG16



区域卷积神经网络R-CNN

Fast R-CNN性能提升

	Fast R-CNN	SPP-Net	R-CNN
训练时间	9.5（8.8x）	25（3.4x）	84
单图测试时间	0.32s（146x）	2.3s（20x）	47.0s
mAP	66.9%	63.1%	66.0%



区域卷积神经网络R-CNN

Fast R-CNN准确性提升

- 端对端（End-to-end）训练

Finetune层数	Fc6层之后	Conv3层之后	Conv2层之后
VOC 2007 mAP	61.4%	66.9%	67.2%
单图测试时间	0.32s	0.32s	0.32s

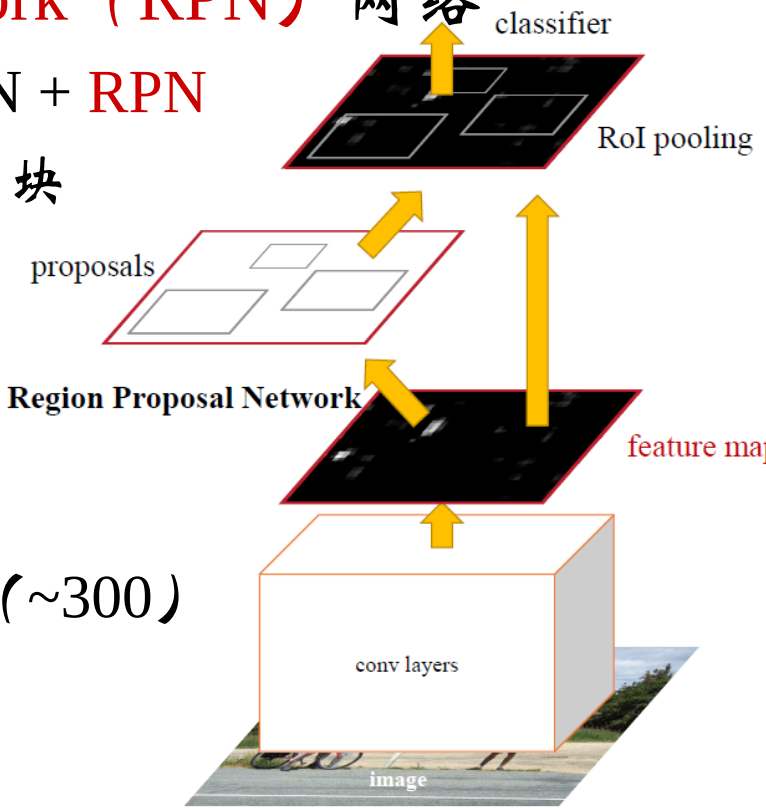
- 多任务（Multi-task）训练

训练类型	Fast R-CNN （VGG16）			
多阶段训练			Y	
Mutil-task训练		Y		Y
测试阶段Bbox回归			Y	Y
VOC 2007 mAP	62.6%	63.4%	64.0%	66.9%

区域卷积神经网络R-CNN

Faster R-CNN网络

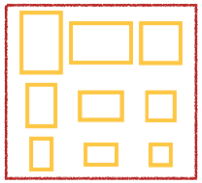
- 集成Region Proposal Network（RPN）网络
  - Faster R-CNN = Fast R-CNN + RPN
  - 取代离线Selective Search模块
    - 解决性能瓶颈
  - 进一步共享卷积层计算
  - 基于Attention注意机制
    - 引导Fast R-CNN关注区域
  - Region proposals量少质优（~300）
    - 高precision，高recall

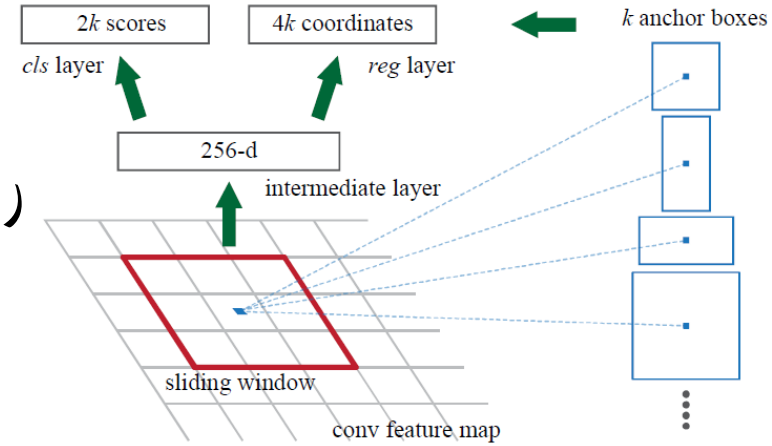


区域卷积神经网络R-CNN

Faster R-CNN网络

- Region Proposal Network（RPN）网络
  - 一种全卷积网络（Fully Convolutional Network）
    - 3x3, 256-d卷积层 + ReLU←输入图片的Conv5特征（WxH）
    - 1x1， 4k-d卷积层→输出k组proposal的offsets（r, c, w, h）
    - 1x1， 2k-d卷积层→输出k组（object score, non-object score)
  - 参考框Anchor box类型k=9
    - 中心跟=卷积核中心
    - 3个尺度scale（128, 256, 512）
    - 3个宽高比ratio
      - 1:1, 1:2, 2:1
  - Anchor数量WHk





区域卷积神经网络R-CNN

Faster R-CNN网络

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*).$$

- RPN网络的loss
  - 分类 L_{cls} 覆盖2类：object & non-object
  - 回归 $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$ 使用smooth L1
    - x, x_a, x^* 分别是对应预测框、anchor框、ground truth框
    - y, w, h类似
    - 不同的特征输入：3x3卷积
  - 指示函数 p_i^*
  - 训练样本mini-batch
    - 单个图片
    - 128个正样本：IoU > 0.7的anchor框 （或最大IoU）
    - 128个负样本：IoU < 0.3的anchor框

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

区域卷积神经网络R-CNN

Faster R-CNN网络

- 4步训练流程
  - Step1 – 训练RPN网络
    - 卷积层初始化←ImageNet上pretrained模型参数
  - Step2 - 训练Fast R-CNN网络
    - 卷积层初始化←ImageNet上pretrained模型参数
    - Region proposals由Step1的RPN生成
  - Step3 - 调优RPN
    - 卷积层初始化←Fast R-CNN的卷积层参数
    - 固定卷积层，finetune剩余层
  - Step4 - 调优Fast R-CNN
    - 固定卷积层，finetune剩余层
    - Region proposals由Step3的RPN生成

卷积层无共享

卷积层共享

区域卷积神经网络R-CNN

Faster R-CNN性能提升

	Faster R-CNN	Fast R-CNN	R-CNN
单图测试时间	0.198s	2.0s	50.0s
PASCAL VOC 07 mAP	66.9%	66.9%	66.0%

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/h5chJQt

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