

Sparse Model and Ensemble Learning with Convolution Neural Network

唐胜 副研究员

ts@ict.ac.cn

中国科学院计算技术研究所多媒体计算研究组

MCG-ICT-CAS: Multimedia Computing Group, ICT, CAS

2016年1月27日

团队介绍



跨媒體計算課題組
Multimedia Computing Group, ICT-CAS

- 中科院计算所多媒体计算研究组
 - 成立于2000年4月，员工12人、博士生和硕士生50多人
 - 李锦涛研究员(副所长)、张勇东研究员(2015国家杰青)
 - 副研究员5人、高工1人
 - 主要研究方向：<http://mcg.ict.ac.cn/>
 - 计算机视觉与模式识别（深度学习）
 - 多媒体内容分析与检索、内容过滤
 - 音视频编解码、计算摄影学
 - 获得授权专利40余项，发表相关研究论文200余篇
 - 2008年国际视频检索评测TRECVID**视频检索**任务**第一**
 - 2009年国际视频检索评测TRECVID**拷贝检测**任务**第一**
 - 2006年、2014年**北京市科学技术奖一等奖**
 - 2012度首届中国**计算机学会科学技术奖**



ILSVRC2015参赛团队介绍



唐胜:总体



张蕊:分类



肖俊斌:分类



曹智:分类



李瑜:定位



邓立曦:定位



李灵慧:检测



王斌:检测

提纲

- ILSVRC介绍
- 分类定位任务 (CLS_LOC)
- 目标检测任务 (DET)
- 总结与展望

ILSVRC介绍

- 2015年比赛任务:
 - Main (传统任务)
 - 目标检测 (Object Detection, DET)
 - 使用规定数据集✓ (第一次参加)
 - 使用外部数据集
 - 目标定位 (Object Localization, CLS-LOC)
 - 使用规定数据集✓ (第一次参加)
 - 使用外部数据集
 - Taster (新任务)
 - 视频目标检测(Object Detection from Video)
 - 场景分类 (Scene Classification)

大规模视觉识别挑战赛成绩

- 分类定位任务(CLS-LOC)
 - **23**个参赛队伍中，错误率低于去年最好结果(25.30%)的前十名

名次	参赛队伍	Error (%)
1	MSRA (微软)	9.01
2	Trimps-Soushen(公安部三所)	12.29
3	Qualcomm Research (高通)	12.55
4	MCG-ICT-CAS	14.69
5	Lunit-KAIST (韩国科学技术院)	14.73
6	Tencent-Bestimage (腾讯)	15.55
7	ReCeption (Google)	19.58
8	Cuimage (香港中文大学)	21.21
9	CIL (Samsung Electronics)	23.19
10	VUNO (美国公司)	25.30

大规模视觉识别挑战赛成绩

- 目标检测任务(DET)

- **20**个参赛队伍中，平均精度MAP高于去年最好结果(43.9%)的前七名

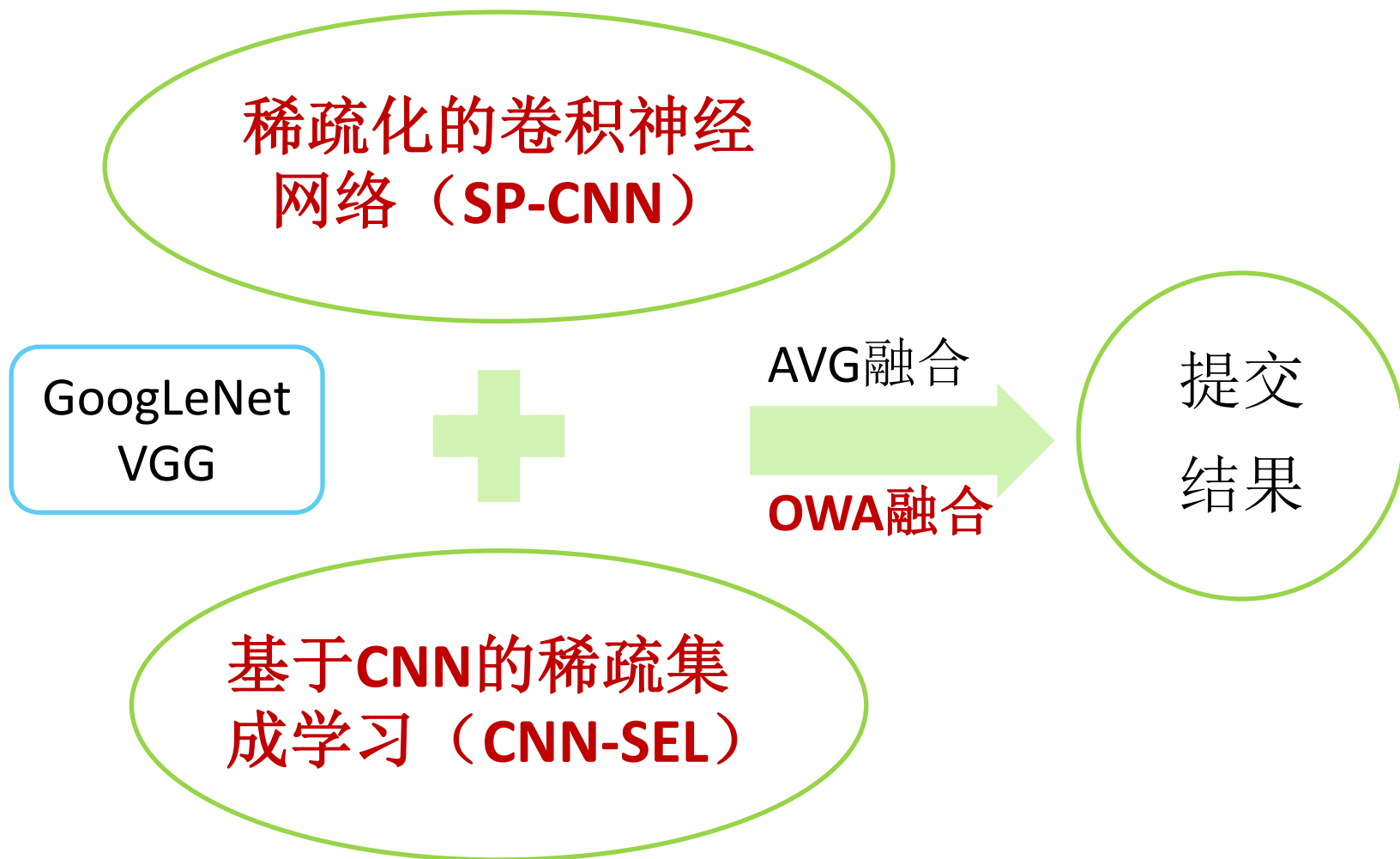
名次	参赛队伍	MAP (%)
1	MSRA (微软)	62.07
2	Qualcomm Research (高通)	53.57
3	Cuimage (香港中文大学)	52.71
4	The University of Adeleide	51.44
5	MCG-ICT-CAS	45.36
6	DROPLET-CASIA (中科院自动化所)	44.86
7	Trimps-Soushen(公安部三所)	44.64

以上排名均来源于：<http://image-net.org/challenges/LSVRC/2015/results>

提纲

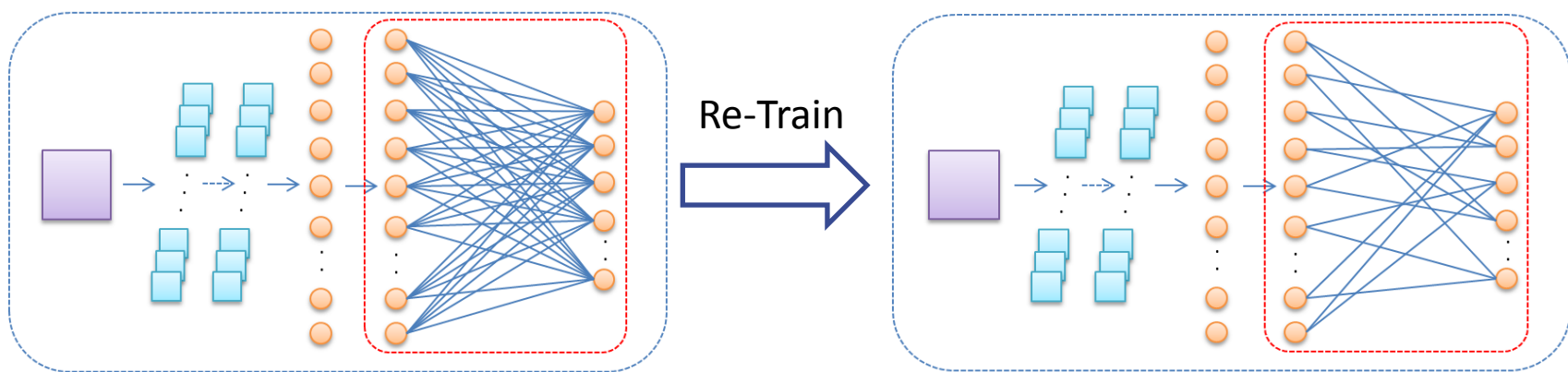
- ILSVRC介绍
- 分类定位任务 (CLS_LOC)
- 目标检测任务 (DET)
- 总结与展望

目标分类方案



稀疏化的卷积神经网络(SP-CNN)

- 神经科学研究表明^[1]：神经元之间
 - 稀疏激活(Sparse Activity)
 - 稀疏连接(Sparse Connectivity)
- 一个类别可以用类别基元稀疏表达
- 在全连接层中自适应地断开不重要的连接，仅保留重要连接
 - 去除不重要的连接引入的干扰噪声
 - 精简模型，更适合于移动终端和嵌入式设备



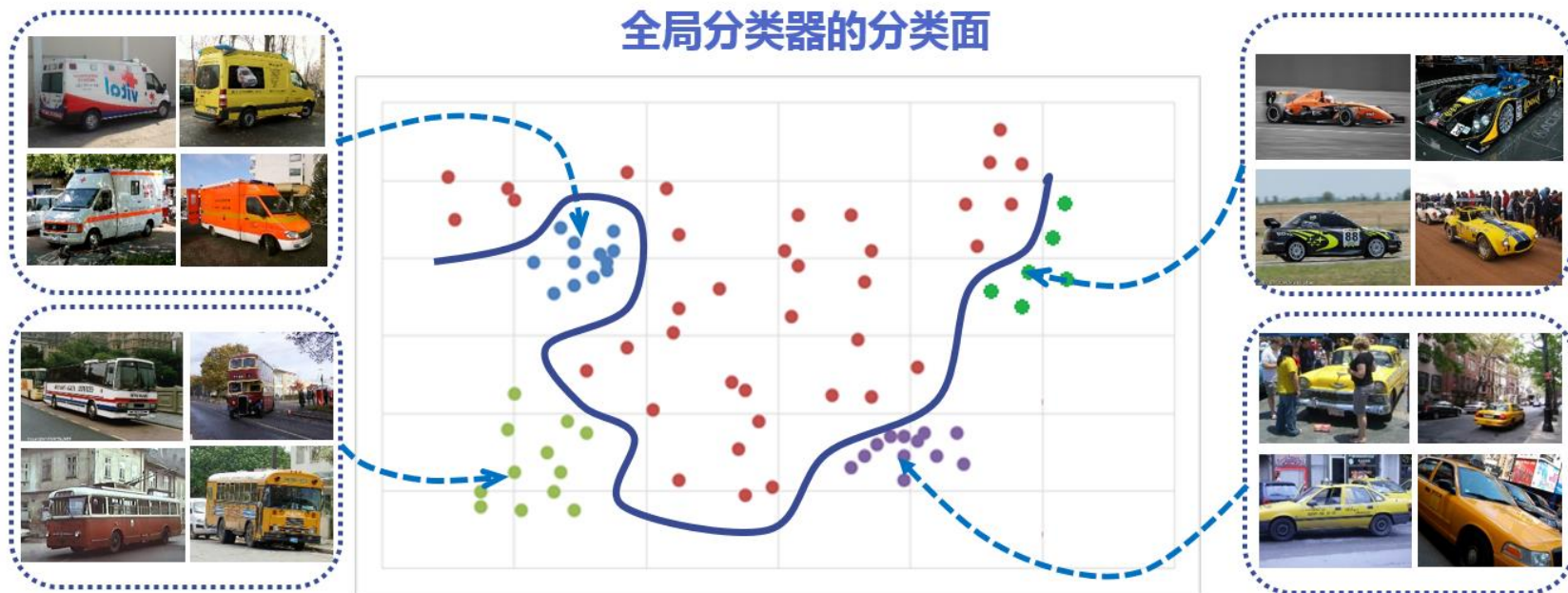
SP-CNN：验证集结果

VGG 模型	Top-1 准确率		Top-5 准确率	
	原始	稀疏化	原始	稀疏化
256	73.05%	74.32%	91.17%	91.91%
384	74.32%	74.88%	91.89%	92.07%
512	72.09%	72.98%	90.49%	90.74%
融合	75.38%	76.1%	92.55%	92.87%

在最后一个全连接层参数数量仅为原模型**9.12%**的情况下:Top-1和Top-5准确率分别提高了**0.72%**、**0.32%**。

CNN-SEL : Motivation

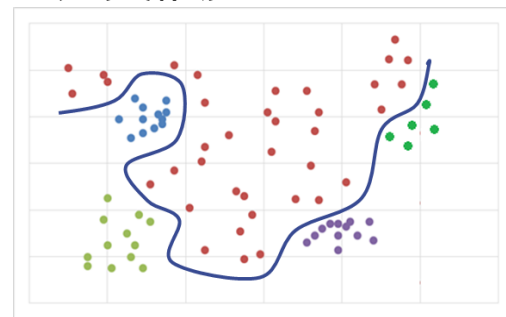
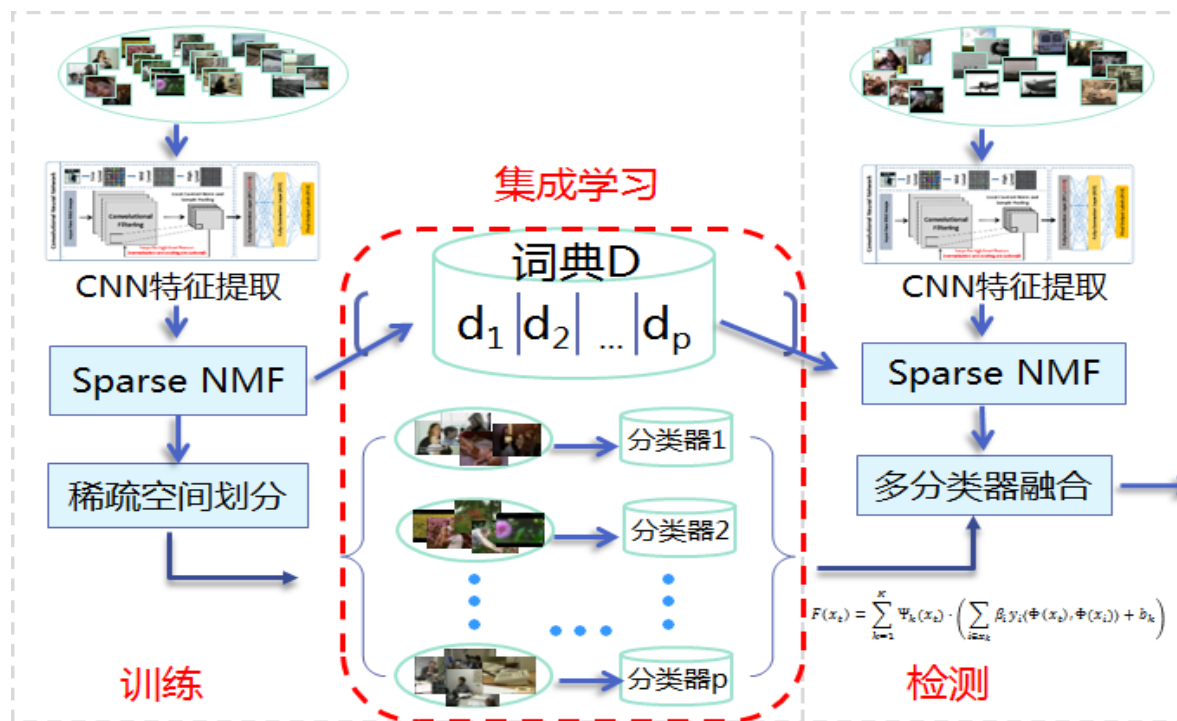
- 语义多态性
 - 类内差异大，需要大量样本，以尽可能囊括可能的样本情况
- 全局分类方法
 - 训练复杂度高，训练RBF核SVM复杂度 $O(n^2)$ - $O(n^3)$
 - 分类面复杂，模型大，识别效率低



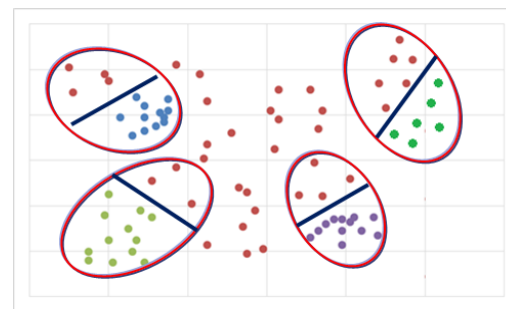
CNN-SEL : 系统框架

- 基于CNN特征的稀疏集成学习^[2]

- 稀疏划分：训练时用稀疏编码划分子空间，大幅提高训练效率
- 稀疏融合：测试时用稀疏编码进行多分类器融合，提高测试效率
- 子分类面简单、激发的子分类器个数少、互为补充、提高分类精度



单一全局分类器分类面

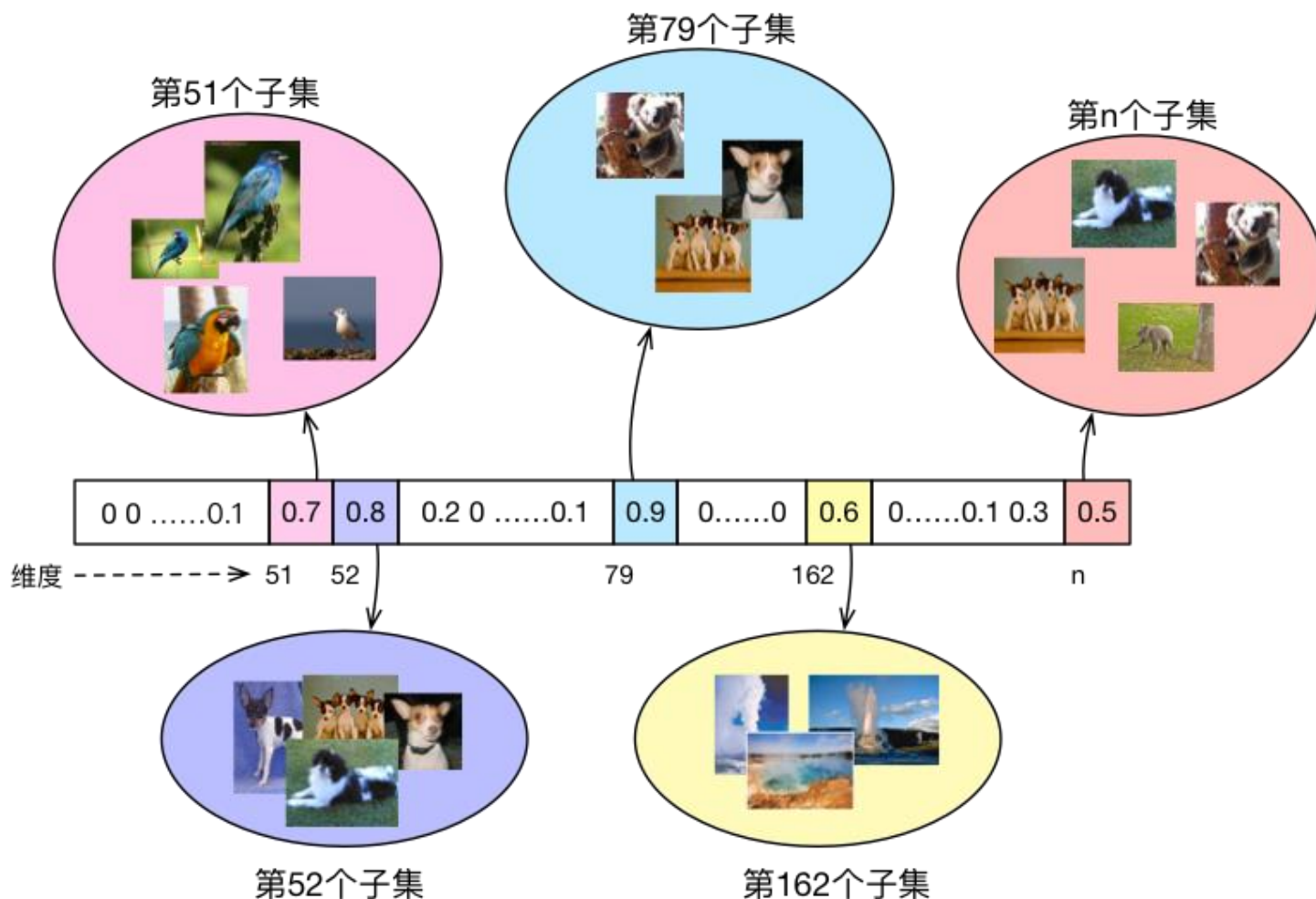


稀疏集成分类器分类面

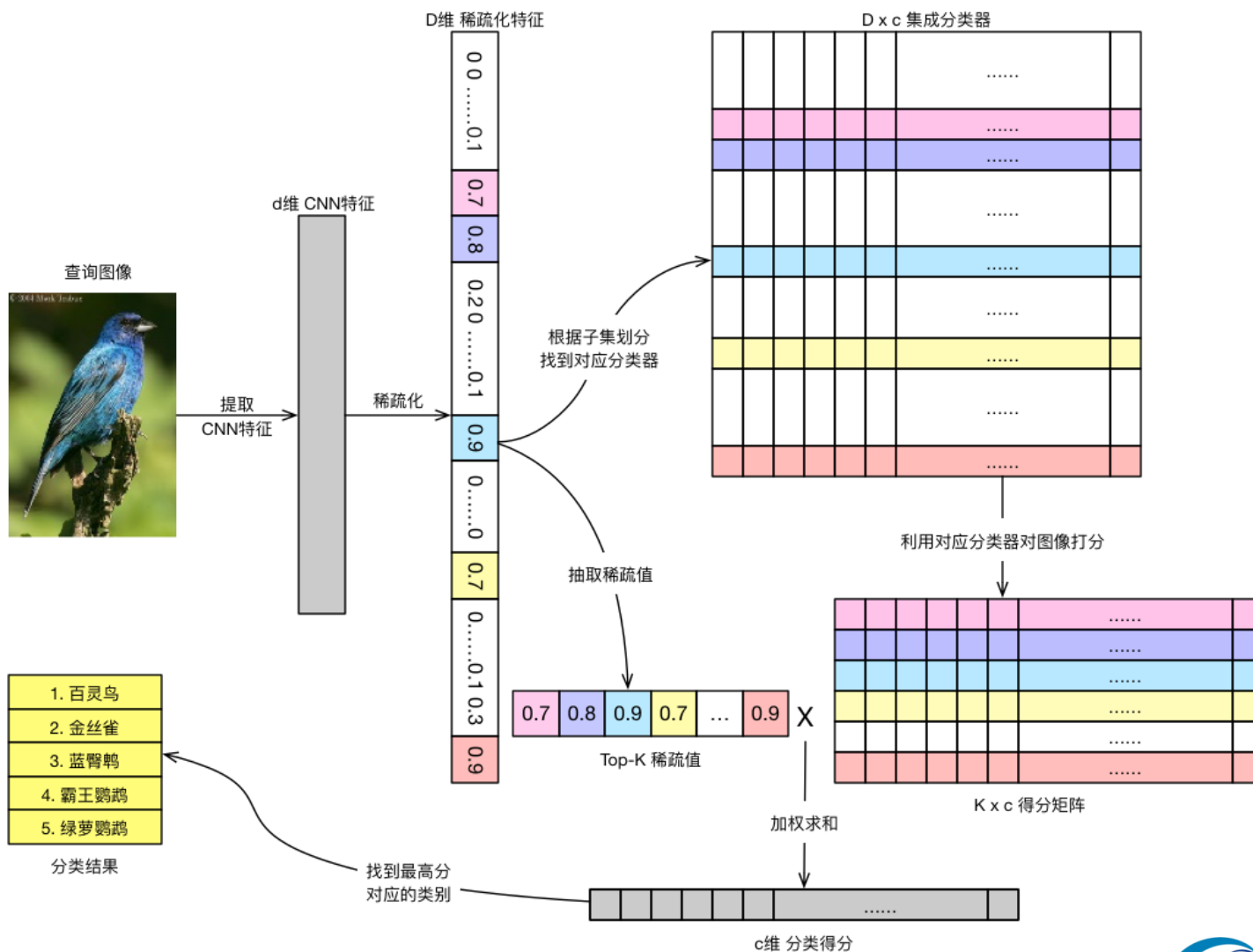
[2] Sheng Tang, Yan-Tao Zheng, Yu Wang, Tat-Seng Chua, "Sparse Ensemble Learning for Concept

第 13 页 Detection", IEEE Transactions on Multimedia, 14 (1): 43-54, February 2012

CNN-SEL : 稀疏划分



CNN-SEL : 稀疏融合



CNN-SEL：验证集实验

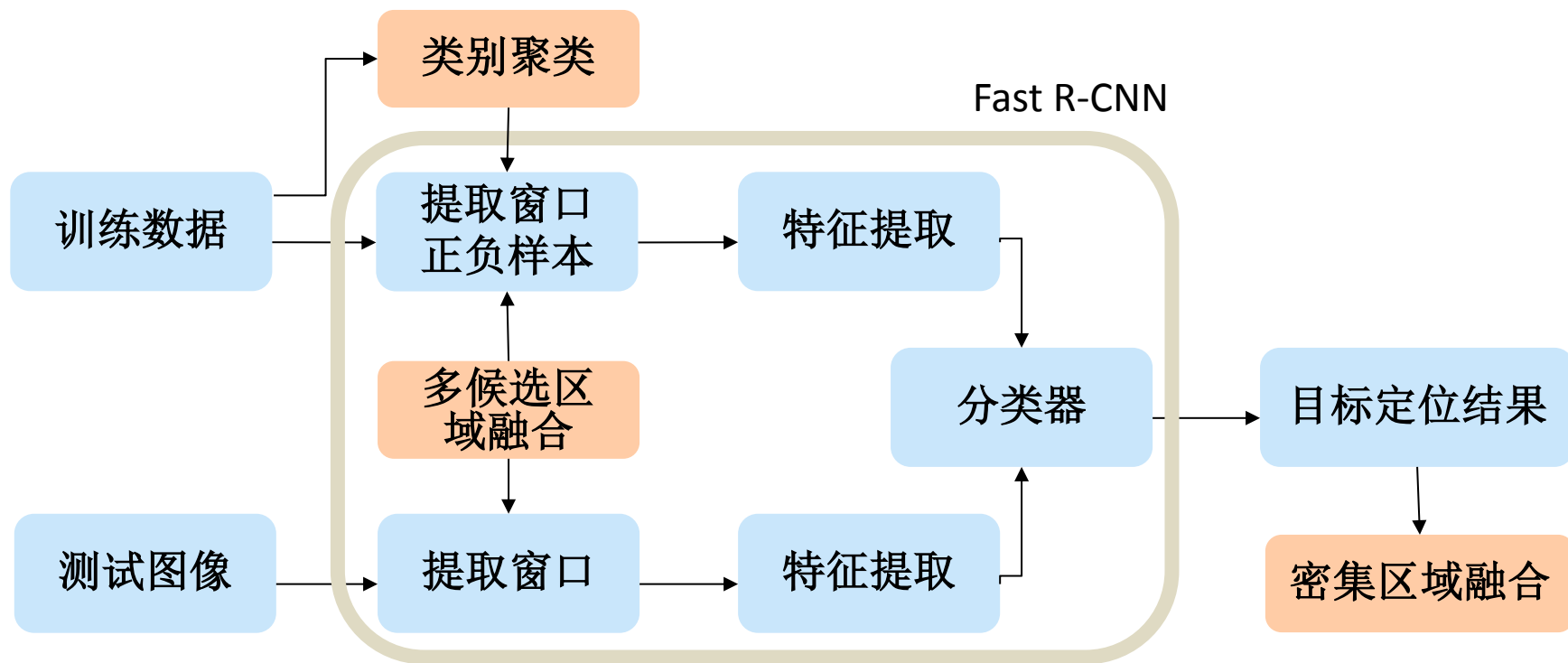
模型	Top-1 准确率	Top-5 准确率
CaffeNet-512D (无数据扩增)	53.83%	77.92%
CNN-SEL-512D (无数据扩增)	55.76% ↑ 1.93%	78.69% ↑ 0.77%
VGG(测试14Crop)	73.63%	91.47%
CNN-SEL-4KD (训练VGG-CNN特征 5Crop,测试14Crop)	73.75% ↑ 0.12%	91.09% ↓ 0.38%

在CaffeNet能明显提高准确率，在VGG上因为训练SVM子分类器只用了5个不同位置的Crop，没有考虑更为复杂的数据扩增方法，如多尺度、镜像翻转、对比度和颜色变化等，因而提高不明显。

CNN-SEL：模型融合结果

模型		验证集 Top-5 准确率	测试集 Top-5 准确率
单 模 型	VGG16	92.55%	-
	VGG19	92.66%	-
	GoogLeNet	90.73%	-
	SP-CNN	92.87%	-
	CNN-SEL	91.09%	-
多 模 型 融 合	VGG16 + VGG19 + GoogLeNet	93.28%	93.30%
	VGG16 + VGG19 + GoogLeNet + SP-CNN	93.44%	93.59%
	VGG16 + VGG19 + GoogLeNet + SPCNN + CNN-SEL(AVG)	93.51%	93.68%
	VGG16 + VGG19 + GoogLeNet + SPCNN + CNN-SEL(OWA)	93.72%	93.69%

目标定位方案

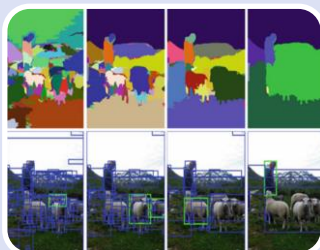


候选区域：多种候选区域（Region Proposals）融合

密集区域融合：利用最终结果的密集区域对目标位置进行回归

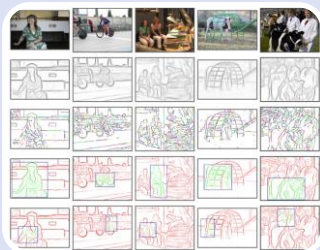
类别聚类：提出了使用CNN特征类别聚类的目标定位方法

目标定位: 多候选区域融合



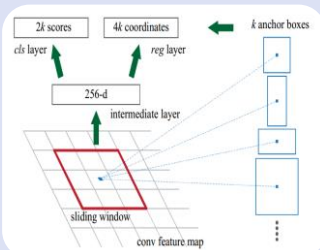
Selective Search (SS)

- 图像分割 -> 聚集
- 2000个窗口



Edge Boxes (EB)

- 边缘信息->计算轮廓个数->评分
- 3000个窗口



Region Proposal Networks (RPN)

- 使用CNN来训练预选窗口
- 300个窗口

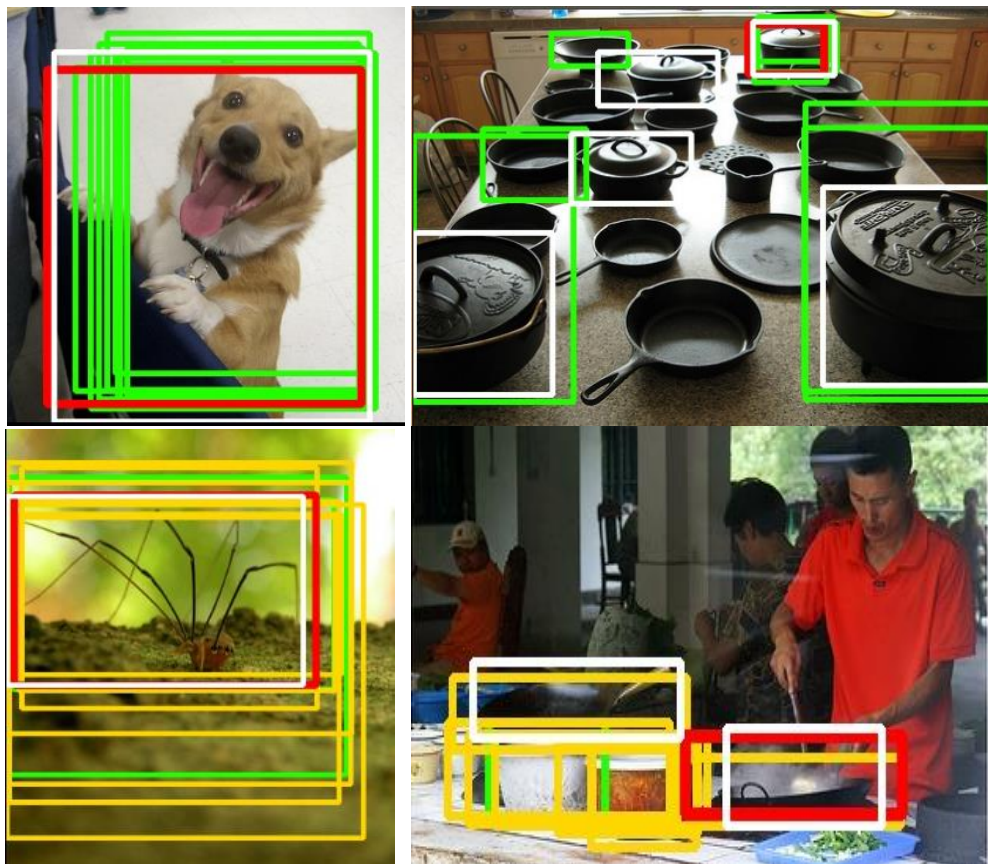
多候选区域融合验证集实验

- 为快速验证，从1000个类别随机选取50个类别
- 为避免分类影响，假设分类准确率为100%，即取验证集上的真实类别标签(Ground truth)
- **多候选区域融合有效提高了目标定位任务的准确率**

训练候选区域	测试候选区域	平均准确率
EB	SS	89.29%
SS	SS	89.53%
RPN	SS	89.11%
SS+EB	SS	90.28%
SS+RPNs	SS	90.20%
SS+EB+RPNs	SS	90.37%

准确率最
高

密集区域融合：Motivation



1. 得分最高不一定是目标
2. 在目标周围有大量高分窗口
3. 很多窗口都是目标位置的微小偏移

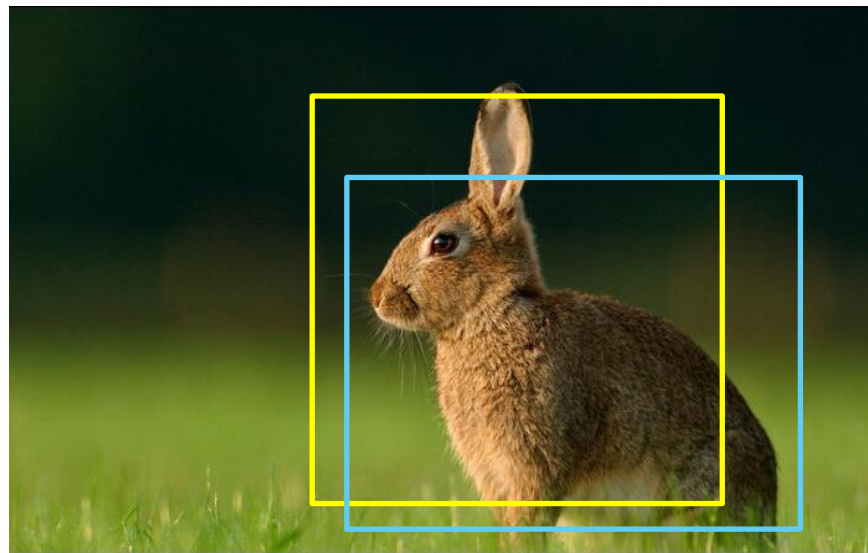


密集区域融合

白框为标注窗口
红框为密集融合窗口
绿色框得分大于等于0.6
黄色框得分低于0.6大于0.3

LOC: 密集区域融合

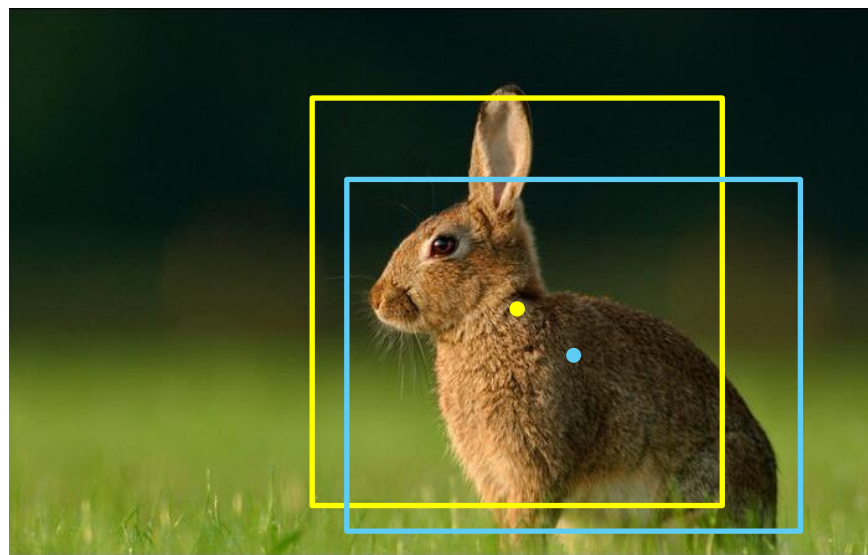
- 密集区域指在同一个目标附近大量类似的候选区域，可以看作同一目标位置的偏移。
- 密集区域的判别函数通过计算两个窗口的中心点距离和窗口之间的IoU来确定。



$$F(B_1, B_2) = \begin{cases} 0, & \text{Dis}(B_1, B_2) > \gamma \text{ 且 } \text{IoU}(B_1, B_2) < \vartheta \\ 1, & \text{Dis}(B_1, B_2) < \gamma \text{ 且 } \text{IoU}(B_1, B_2) > \vartheta \end{cases}$$

LOC: 密集区域融合

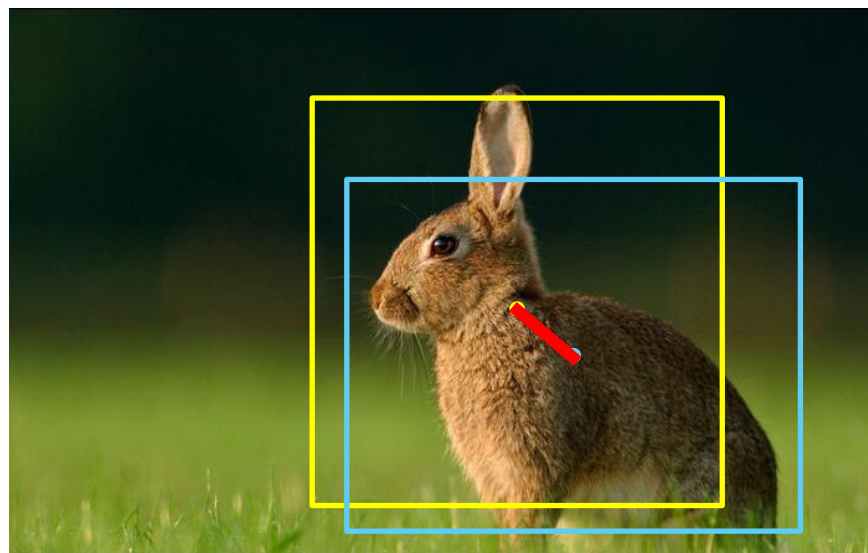
- 密集区域指在同一个目标附近大量类似的候选区域，可以看作同一目标位置的偏移。
- 密集区域的判别函数通过计算两个窗口的中心点距离和窗口之间的IoU来确定。



$$F(B_1, B_2) = \begin{cases} 0, & \text{Dis}(B_1, B_2) > \gamma \text{ 且 } \text{IoU}(B_1, B_2) < \vartheta \\ 1, & \text{Dis}(B_1, B_2) < \gamma \text{ 且 } \text{IoU}(B_1, B_2) > \vartheta \end{cases}$$

LOC: 密集区域融合

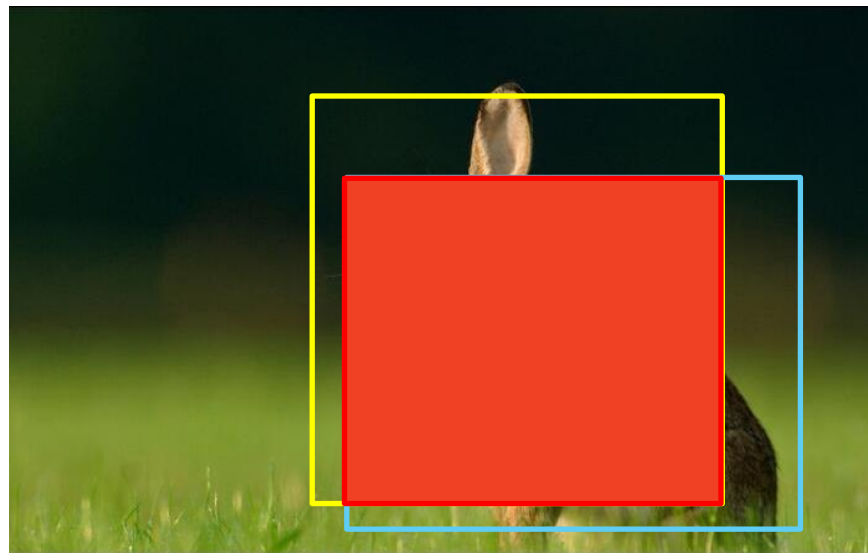
- 密集区域指在同一个目标附近大量类似的候选区域，可以看作同一目标位置的偏移。
- 密集区域的判别函数通过计算两个窗口的中心点距离和窗口之间的IoU来确定。



$$F(B_1, B_2) = \begin{cases} 0, & \text{Dis}(B_1, B_2) > \gamma \text{ 且 } \text{IoU}(B_1, B_2) < \vartheta \\ 1, & \text{Dis}(B_1, B_2) < \gamma \text{ 且 } \text{IoU}(B_1, B_2) > \vartheta \end{cases}$$

LOC: 密集区域融合

- 密集区域指在同一个目标附近大量类似的候选区域，可以看作同一目标位置的偏移。
- 密集区域的判别函数通过计算两个窗口的中心点距离和窗口之间的IoU来确定。



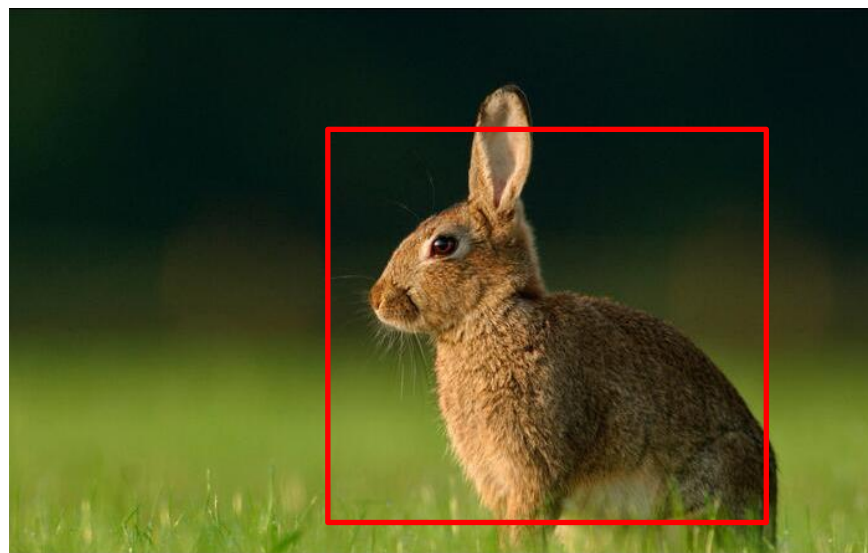
$$F(B_1, B_2) = \begin{cases} 0, & \text{Dis}(B_1, B_2) > \gamma \text{ 且 } \text{IoU}(B_1, B_2) < \vartheta \\ 1, & \text{Dis}(B_1, B_2) < \gamma \text{ 且 } \text{IoU}(B_1, B_2) > \vartheta \end{cases}$$

LOC: 密集区域融合

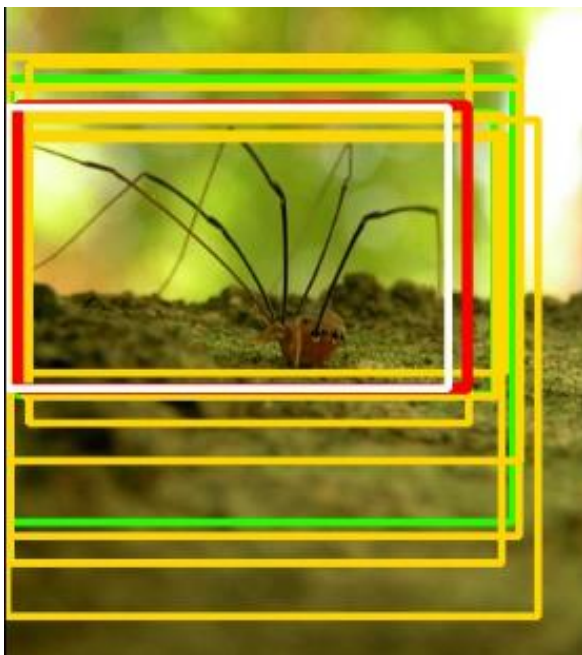
- **密集区域融合(DRF)**

Dense Region Fusion

- 区域融合：将两个区域的坐标取平均值。
- 得分融合：高得分窗口分数加上低得分窗口分数的一半。



LOC: 密集区域融合实验结果



原始EB1000EB
准确率87.27%



DRF准确率
87.99%

0.72%

原始EB1000EB
IoU:74.2%



DRF IoU
77.3%

3.1%

1. 准确率提高，排除了高得分非目标窗口
2. 重叠率提高，目标位置的偏移得到回归

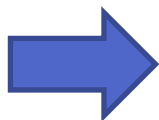
LOC: 基于CNN聚类的目标定位



黑腹滨鹬、
长嘴半蹼鹬
红脚鹬、
翻石鹬



赤蠵龟、
棱皮龟、
淡水龟、
水龟



1. 类别之间可能非常相似
2. 聚类后使得正例增多
3. 降低了时间复杂度



使用CNN特征进行类别聚类

LOC: 基于CNN聚类目标定位



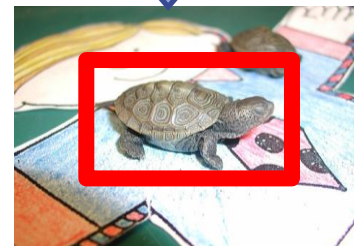
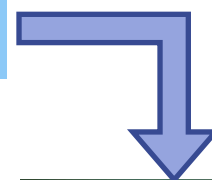
鸻



龟

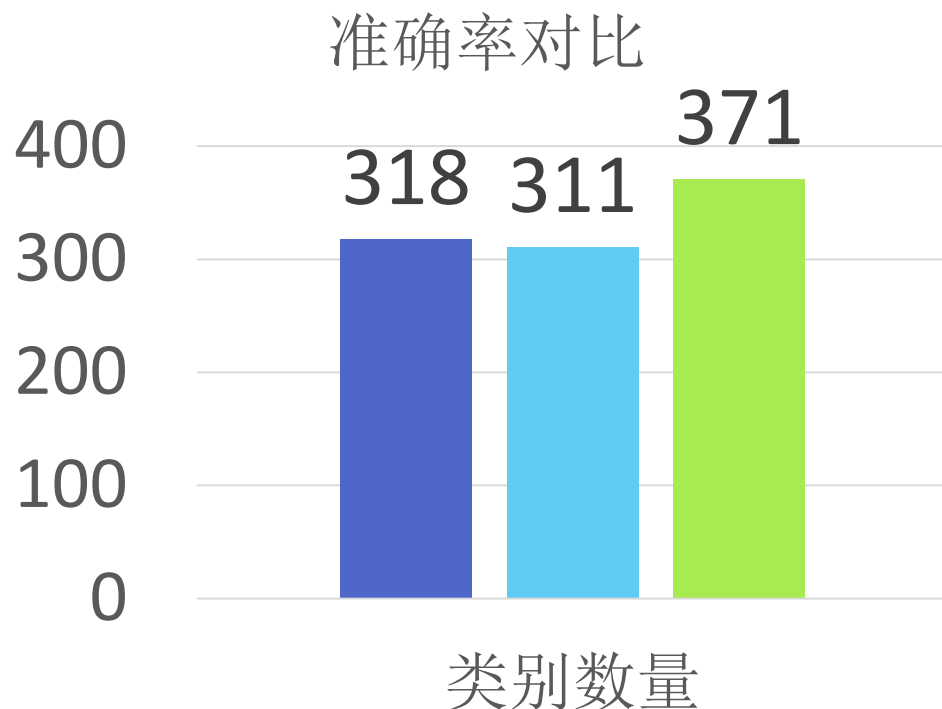


Fast R-CNN



将聚类后的类别作为一个大类
训练Fast R-CNN模型
得到定位结果

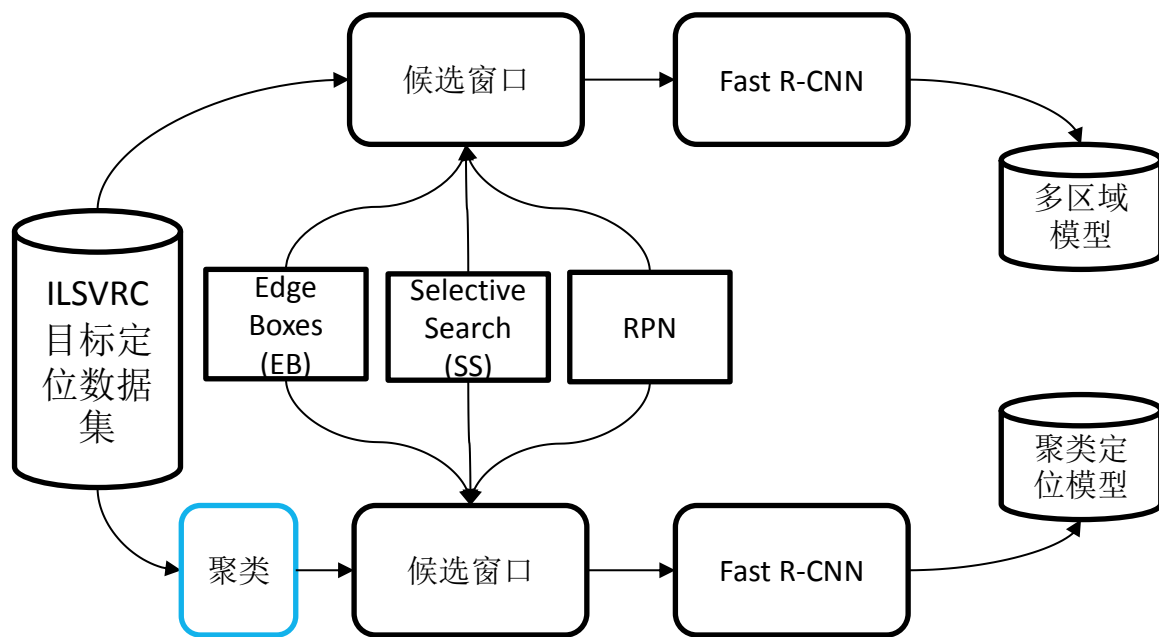
LOC: 基于CNN聚类的目标定位



- 聚类200准确率高
- 准确率相同
- 原始1000类准确率高

- 类别聚类的定位方法在部分类别上有更高的准确率
- 聚类方法平均训练时间为0.6秒/迭代，原始方法为1.2秒/迭代，相比较减少了一半的时间，提高了定位效率

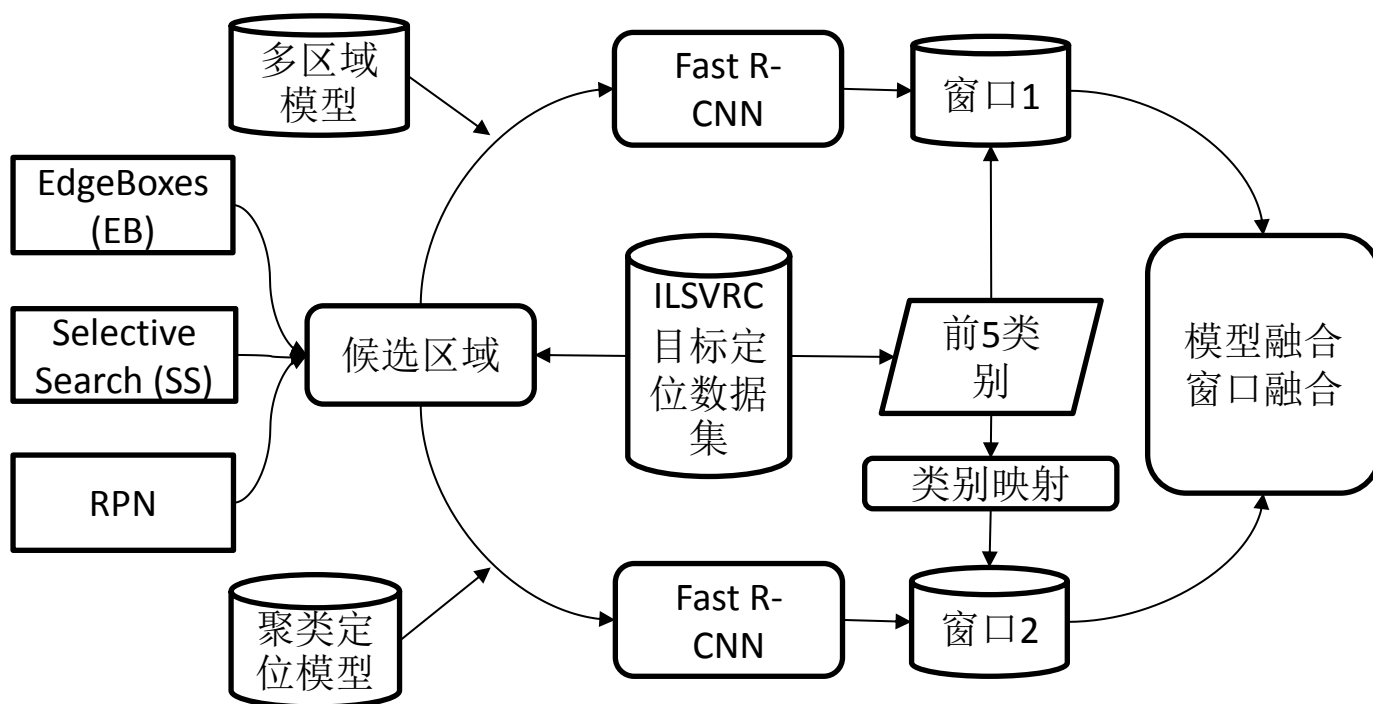
目标定位系统框架-训练



训练

1. 多区域模型：基于多候选区域融合的定位方法
2. 聚类定位模型：基于CNN特征特征类别聚类的定位方法

目标定位系统框架-测试



测试

1. 模型融合：两个模型在不同类别上准确率不同，根据验证集上图片类别的结果，选用准确度高的模型
2. 窗口融合：使用密集区域融合对定位位置进行回归

LOC-验证集结果

[GoogleNet+VGG](93.28%)情况下的定位结果

模型	SSEB1000EB	EB200RPN	SSEB1000EB +EB200RPN	SSEB1000EB +EB200RPN +DRF
TOP5准确率	83.42%	82.62%	84.56%	84.96%

- 1. 聚类和多区域模型互补：**尽管聚类定位模型准确率低于多区域定位模型，但是通过模型融合，准确率比前两者分别提高了**1.14%**、**1.94%**。
- 2. 密集窗口融合，**准确率提高了**0.40%**

LOC-测试集结果

Run	分类方法	定位方法	分类错误率	定位错误率
Run5	[GoogleNet+VGG+SPCN N+SEL]OWA	EB200RPN+SSEB1000EB	6.31%	14.69%
Run4	[GoogleNet+VGG+SPCN N+SEL]AVG	EB200EB+SSEB1000EB	6.32%	14.82%
Run3	[GoogleNet+VGG+SPCN N]	EB200RPN+SSEB1000EB	6.41%	14.73%
Run2	[GoogleNet+VGG+SPCN N]	EB200RPN+SSEB1000RPN	6.41%	14.95%
Run1	[GoogleNet+VGG]	EB200EB+EB1000EB	6.70%	15.82%

Category Aggregation Among Region Proposals For Object Detection

唐胜 副研究员

ts@ict.ac.cn

中国科学院计算技术研究所多媒体计算研究组
MCG-ICT-CAS: Multimedia Computing Group, ICT, CAS

2016年1月27日

提纲

- ILSVRC介绍
- 分类定位任务 (CLS_LOC)
- 目标检测任务 (DET)
- 总结与展望

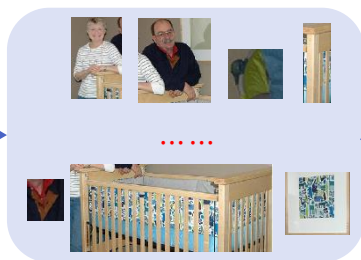
基于类别聚合的目标检测方案

引入类别信息：减少假阳性（False Positive），提高检测精度和效率

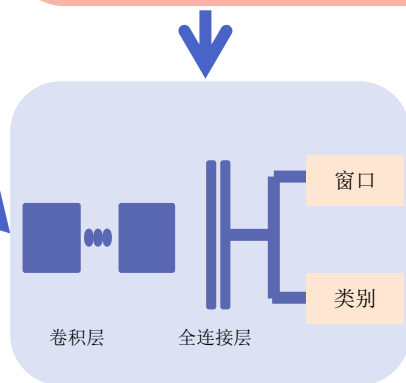
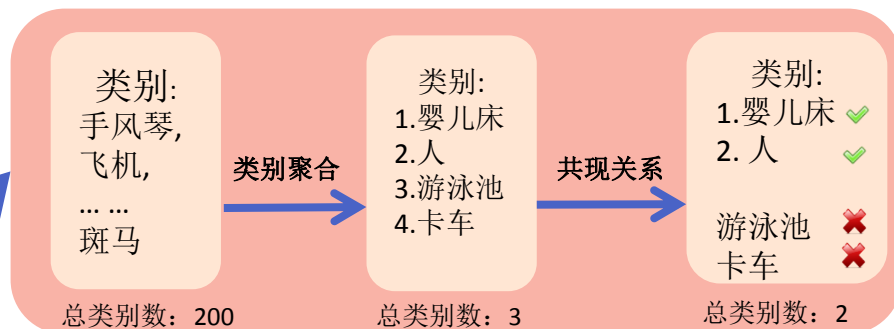
1. 测试图像



2. 预选区域



3. 类别聚合与共现关系

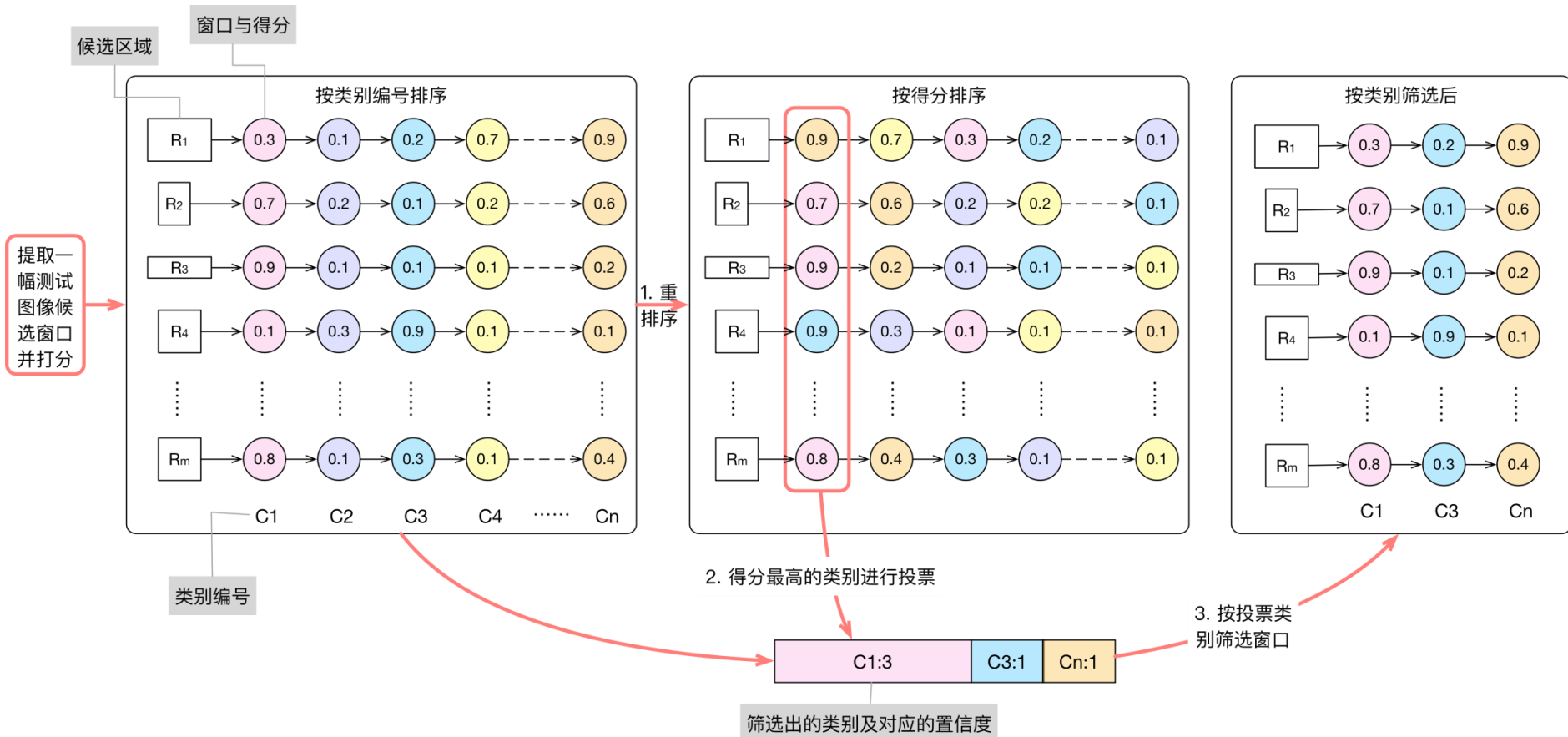


4. 分类模块

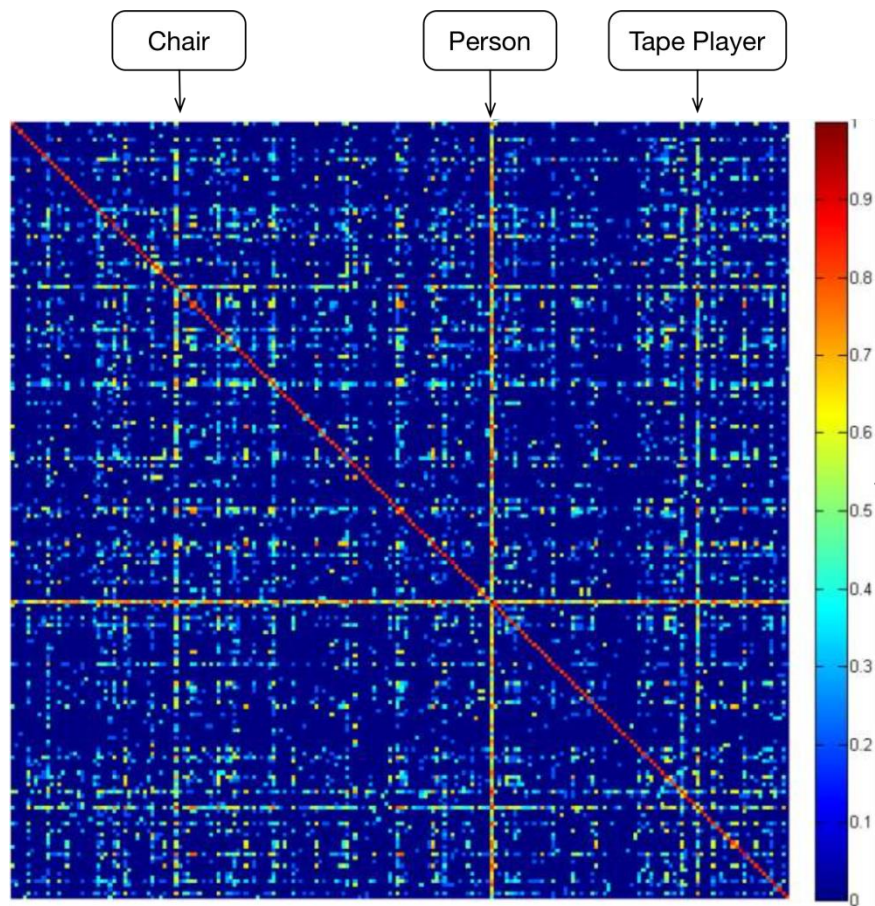


5. 检测融合

目标检测：类别聚合(CA)

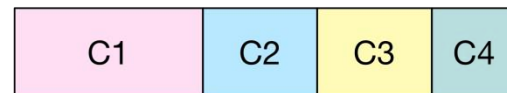


目标: 共现关系(CO)



类别共现矩阵

通过类别聚合得到的类别



1. 获取
共现表

C1	C2 C4
C2	C1 C4
C3	C4
C4	C1 C2 C3

类别共现表

2. 查询
共现表



利用共现关系得到的最终类别

DET: 实验结果

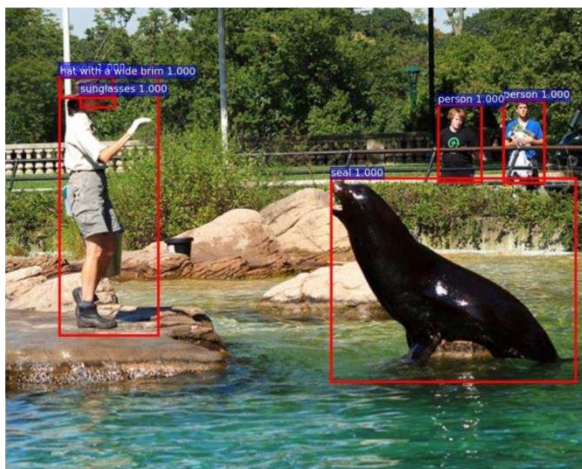
Run	Run方法说明	验证集 MAP	测试集 MAP
Run5	CA+CO{[SS+EB]+(SS)}+{[RPN] +(RPN)}	49.00%	45.36%
Run4	CA{[SS+EB]+(SS)}+{[RPN]+(RPN)}	48.60%	45.19%
Run3	{[SS+EB]+(SS)}+{[RPN]+(RPN)}	48.10%	45.36%
Run2	CA+CO[RPN]+(RPN)	47.60%	44.04%
Run1	CA+CO[SS+EB]+(SS)	47.40%	44.49%

类别聚合: **Run4 VS. Run3**, 验证集上提高**0.9%**, 但测试集上略低, 可能是测试集和训练集在数据分布上存在差异。

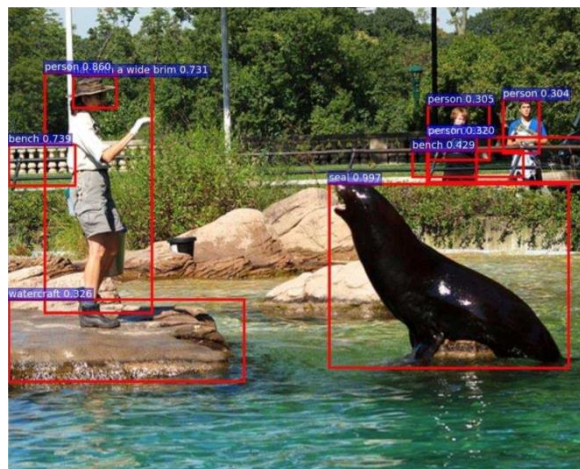
共现关系: **Run5 VS. Run4**, 在验证集和测试集上均提高

多窗口融合有效: **Run5 VS. Run1 / Run2**, 在验证集和测试集上均提高

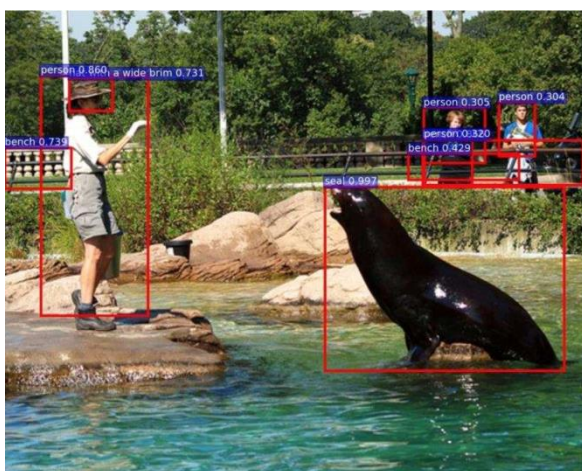
DET: 结果展示



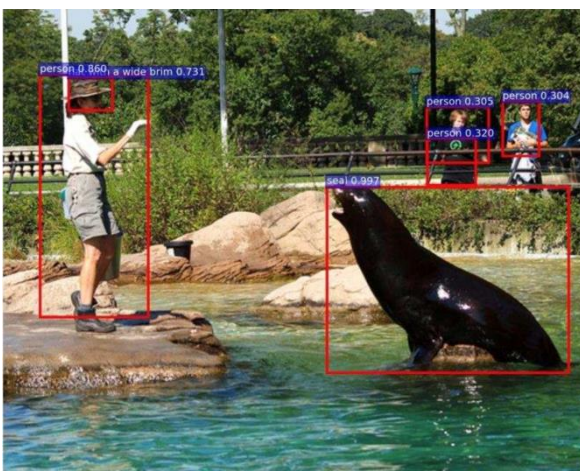
Ground Truth



Fast R-CNN



CA (去掉Watercraft)



CA+CO(去掉Bench)

谢 谢!

Web: <http://mcg.ict.ac.cn/people/shengtang.htm>

