What can we get from the bounding box?

Jie Shao (邵杰)
jieshao.mail@gmail.com

公安部第三研究所

物联网技术研发中心简介

- **中心使命：**支撑“智慧警务”
- 参与部、省、市、县各级公安机关视频监控信息化顶层设计及应用建设
- **视频解析服务体系：**从“处理、分析、挖掘、评价”等环节出发，实现对海量视频资源的深度应用，促进整个视频监控产业实现从监控到理解的转型

The Third Research Institute Of Ministry Of Public Security - Shanghai Jiao Tong University
公安部第三研究所－上海交通大学
智能视频评测联合实验室
Joint Laboratory of Intelligent Video Evaluation

基于视频结构化描述的视频语义分析系统

- 可描述车辆颜色、车型、品牌等， 车型类别 ＞ 1200类
- 个性化检索、以图搜图等
- 参与重大案件侦破数十起：桂林爆炸案、苏州抓捕案、亚信反恐…

车牌识别

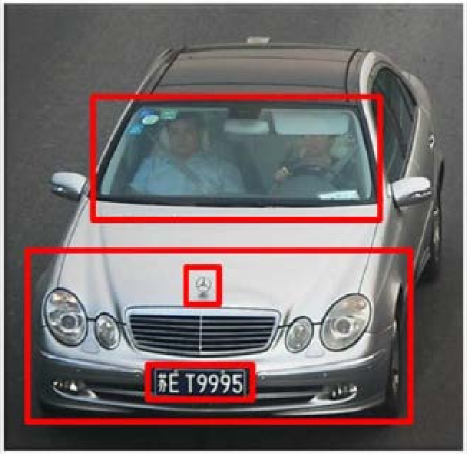
车颜色识别

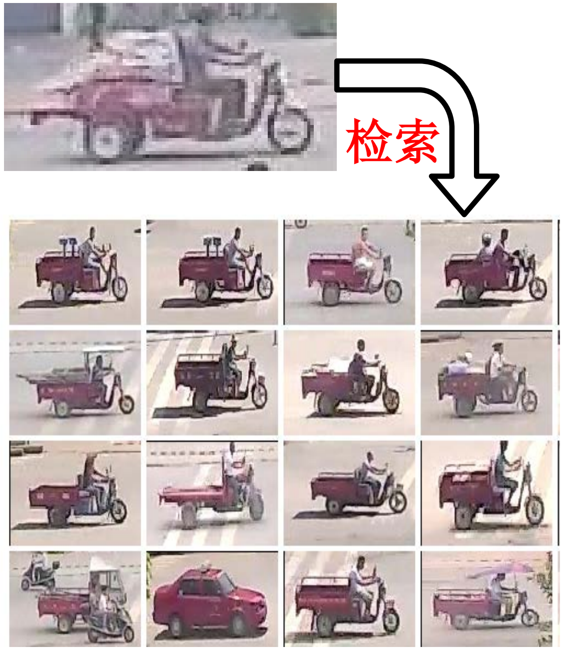
车品牌识别

车型识别

车窗识别

年检标识别





非机动车识别

视频检测

图像检索

DECEMBER 11-18, 2015
ICCV15
International Conference on Computer Vision

TRIMPS

Trimps-Soushen(搜神) at ILSVRC2015

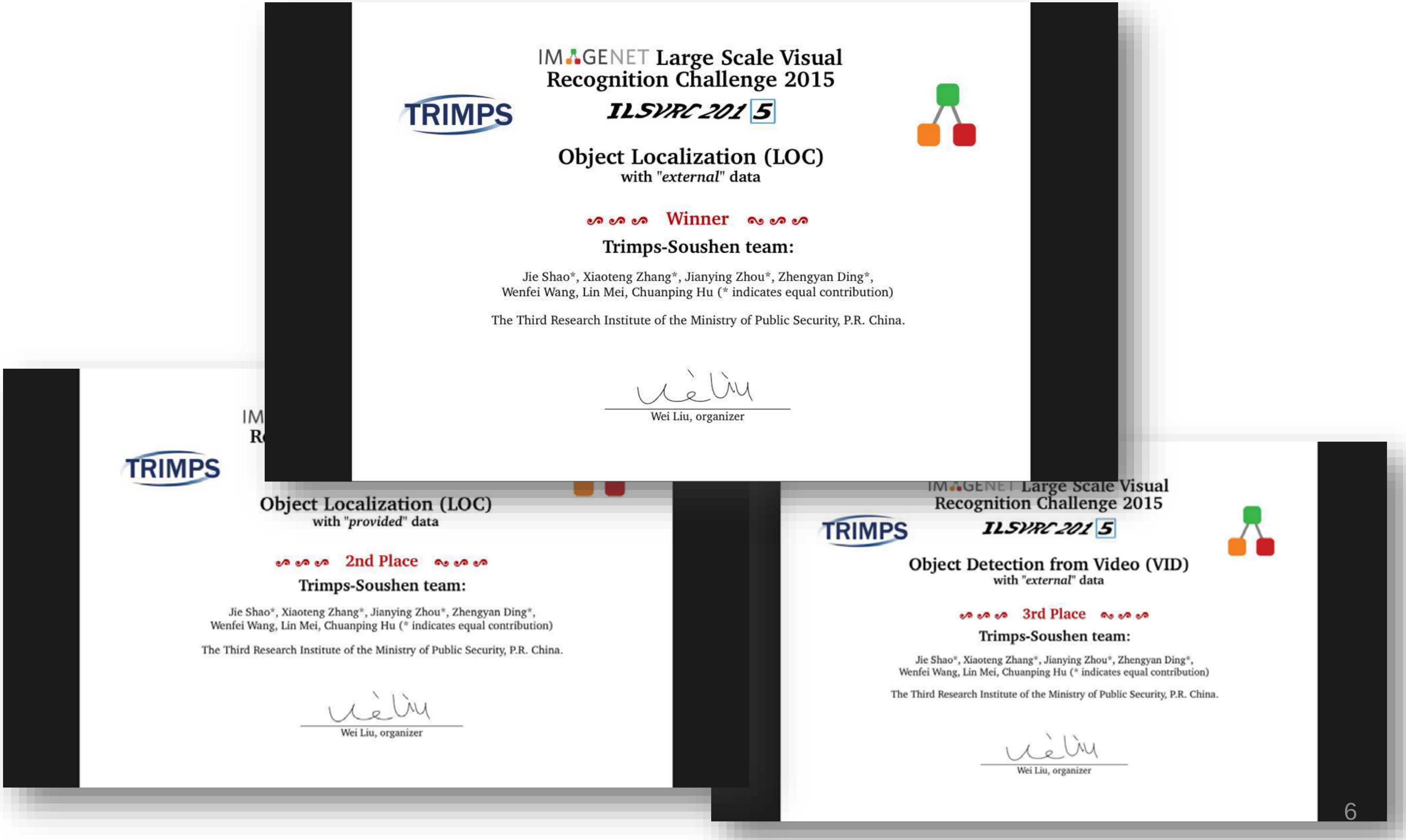
Jie SHAO, Xiaoteng ZHANG, Jianying ZHOU, Zhengyan DING, Wenfei WANG, Lin MEI, Chuanping HU

The Third Research Institute of the Ministry of Public Security, P.R. China.

Summary of Trimps Submission

- **Object localization**
  - 2nd place, 12.29% error (1st place with extra data)
- **Object detection from video (VID)**
  - 4th place, 0.461 mAP (3rd place with extra data)
- **Scene classification**
  - 4th place, 17.98% error
- **Object detection**
  - 7th place, 0.446 mAP (4th place with extra data)

ILSVRC2015 official certification



TRIMPS

IMAGENET Large Scale Visual Recognition Challenge 2015

ILSVRC2015

Object Localization (LOC)
with "*external*" data

Winner

Trimps-Soushen team:

Jie Shao*, Xiaoteng Zhang*, Jianying Zhou*, Zhengyan Ding*,
Wenfei Wang, Lin Mei, Chuanping Hu (* indicates equal contribution)

The Third Research Institute of the Ministry of Public Security, P.R. China.

Wei Liu, organizer

TRIMPS

IM
R

Object Localization (LOC)
with "*provided*" data

2nd Place

Trimps-Soushen team:

Jie Shao*, Xiaoteng Zhang*, Jianying Zhou*, Zhengyan Ding*,
Wenfei Wang, Lin Mei, Chuanping Hu (* indicates equal contribution)

The Third Research Institute of the Ministry of Public Security, P.R. China.

Wei Liu, organizer

TRIMPS

IMAGENET Large Scale Visual Recognition Challenge 2015

ILSVRC2015

Object Detection from Video (VID)
with "*external*" data

3rd Place

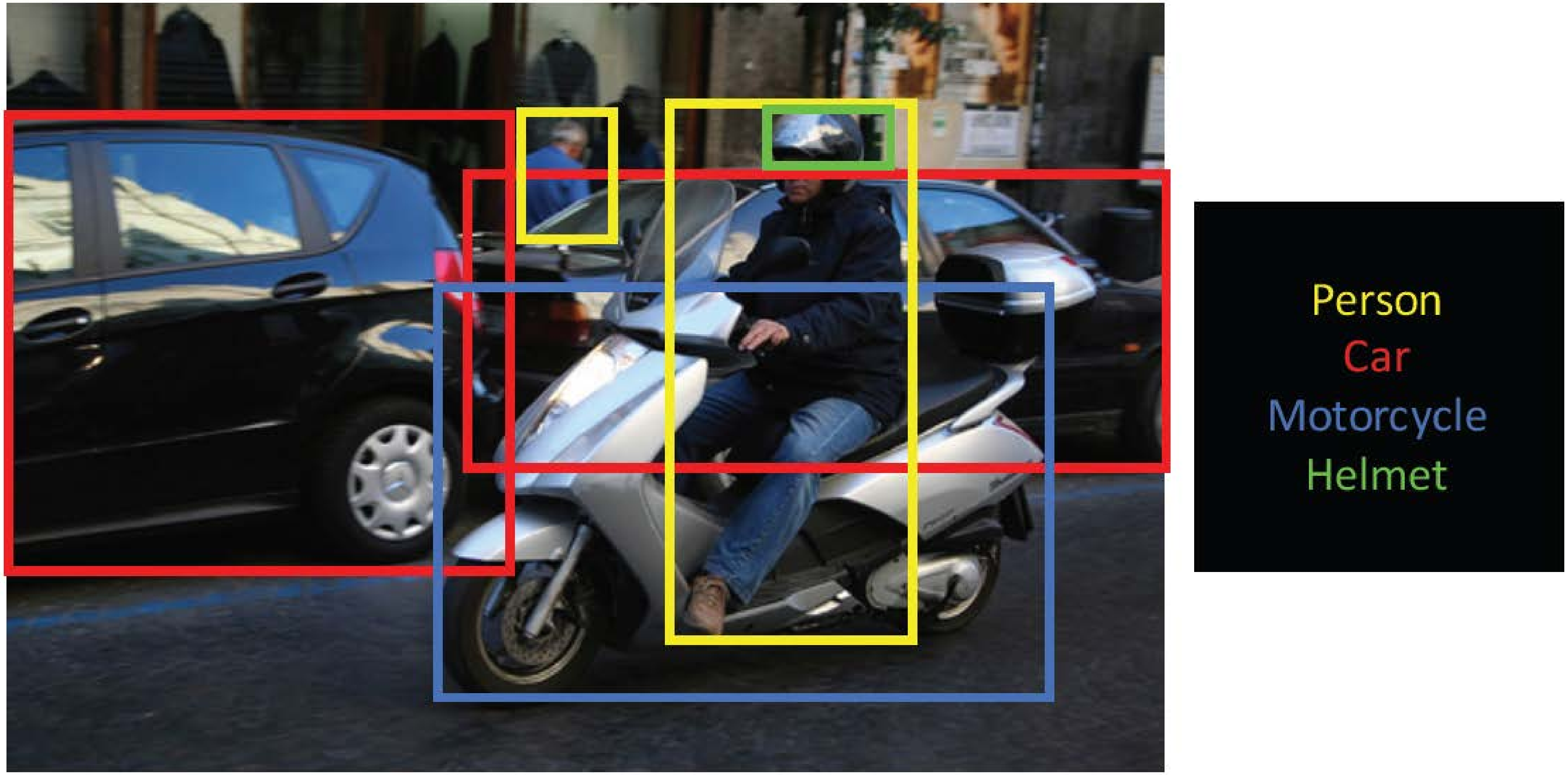
Trimps-Soushen team:

Jie Shao*, Xiaoteng Zhang*, Jianying Zhou*, Zhengyan Ding*,
Wenfei Wang, Lin Mei, Chuanping Hu (* indicates equal contribution)

The Third Research Institute of the Ministry of Public Security, P.R. China.

Wei Liu, organizer

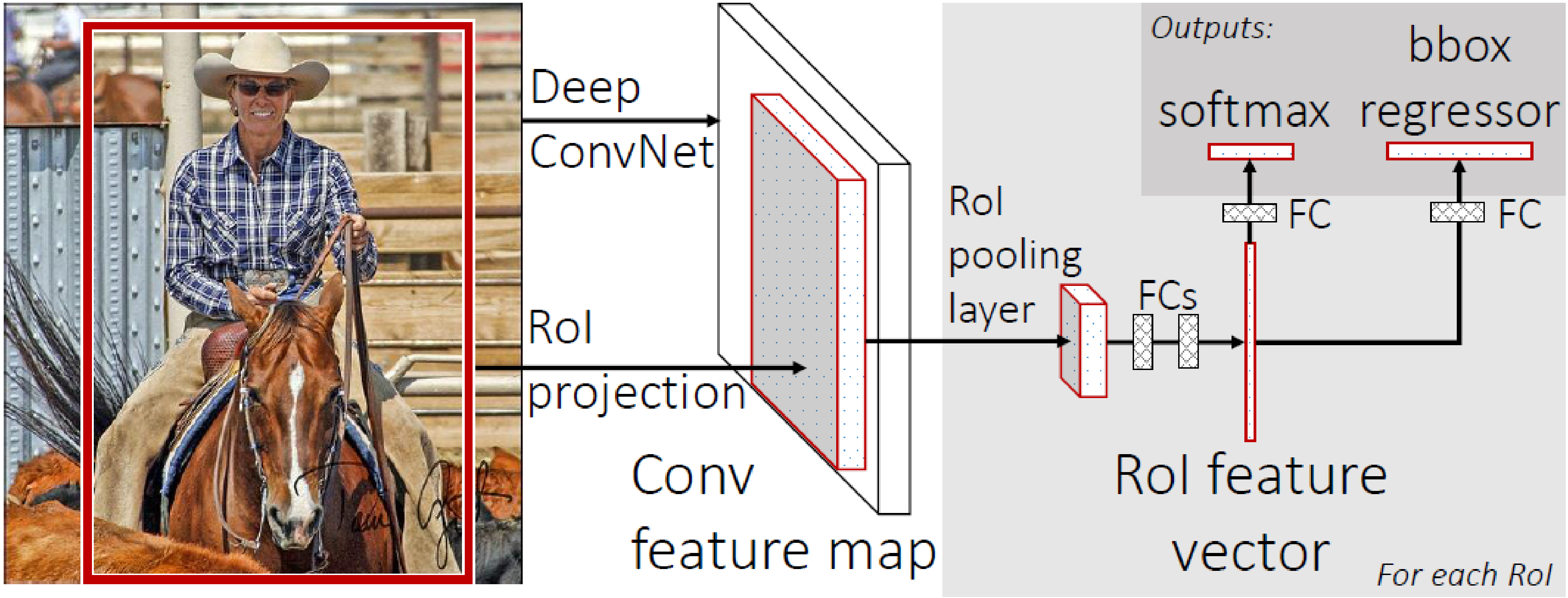
Bounding box annotations



What can we get?

- **Objectness**
- **Negative categories**
- **Bounding box voting**

Region-based detection pipeline

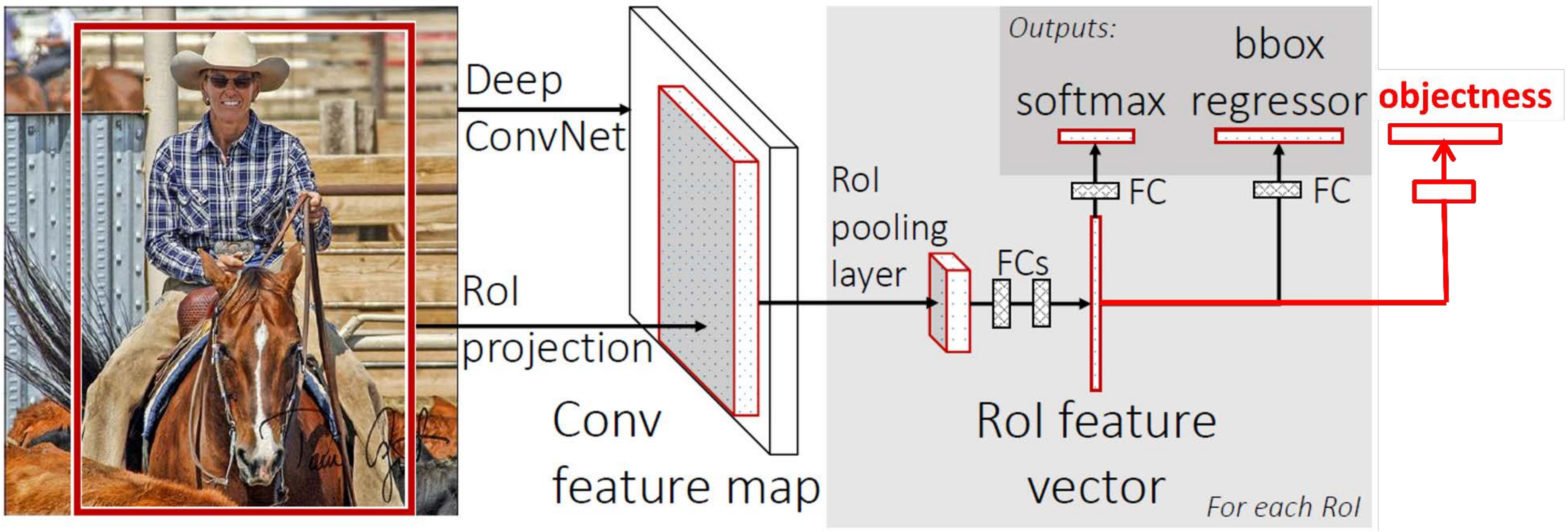


Fast R-CNN, Ross Girshick, IEEE International Conference on Computer Vision (ICCV), 2015

Objectness

- **Motivation**
  - Positive samples must be object first
  - Put objectness in an end-to-end pipeline
- **Related works**
  - BBox rejection
  - DeepBox
  - Region proposal networks(RPN)

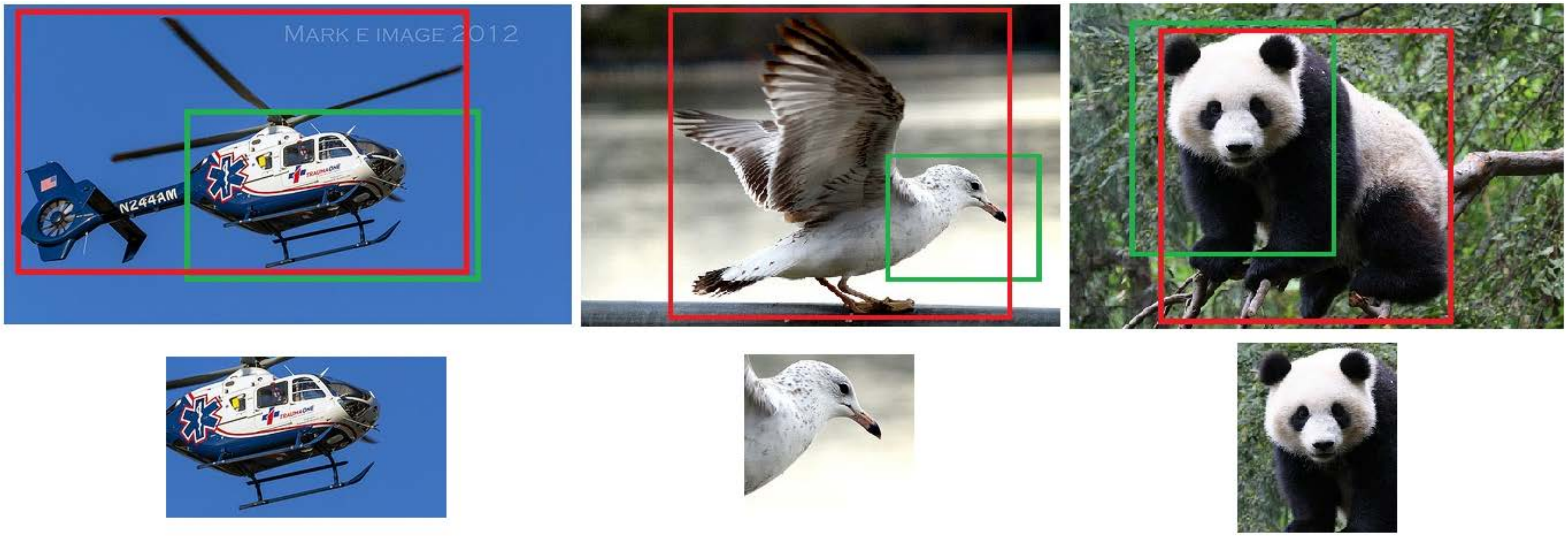
Objectness – our approach



- – Regions with iou>=0.5 label as **1**, otherwise **0**
- – Only use in training stage
- – Most improved on val: +2.2% mAP (avg 1.1%)

Negative categories

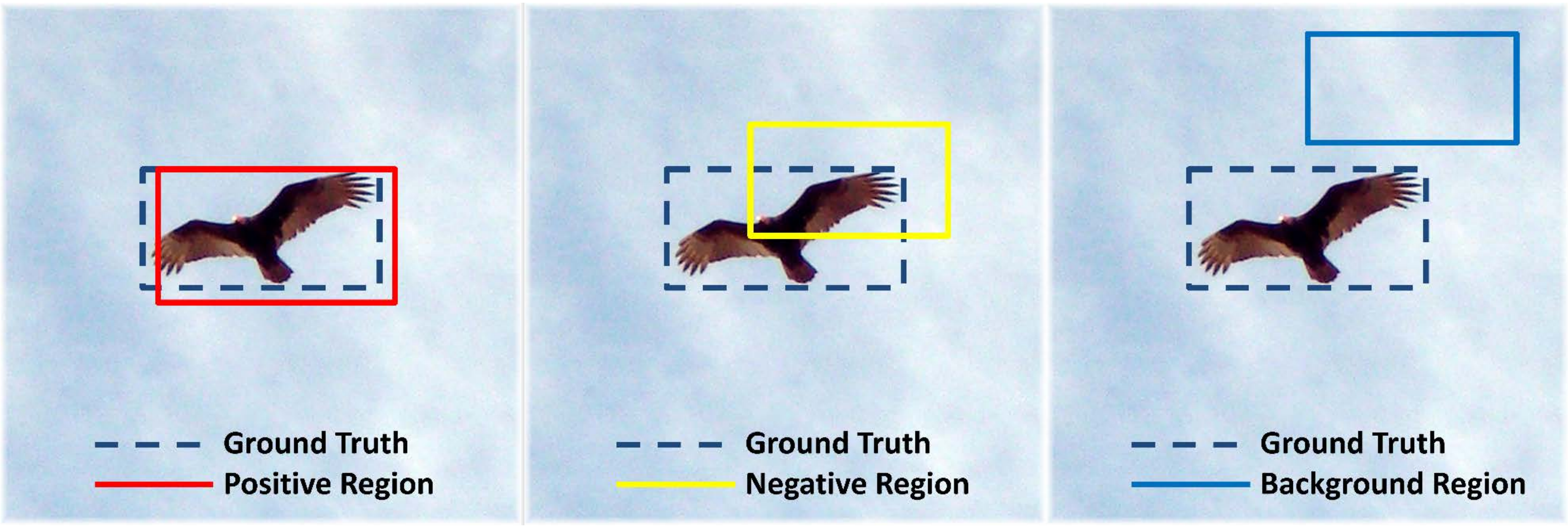
- **Motivation**
  - Set all IOU<0.5 regions to be same categories is **NOT reasonable**
  - Harder task always helps



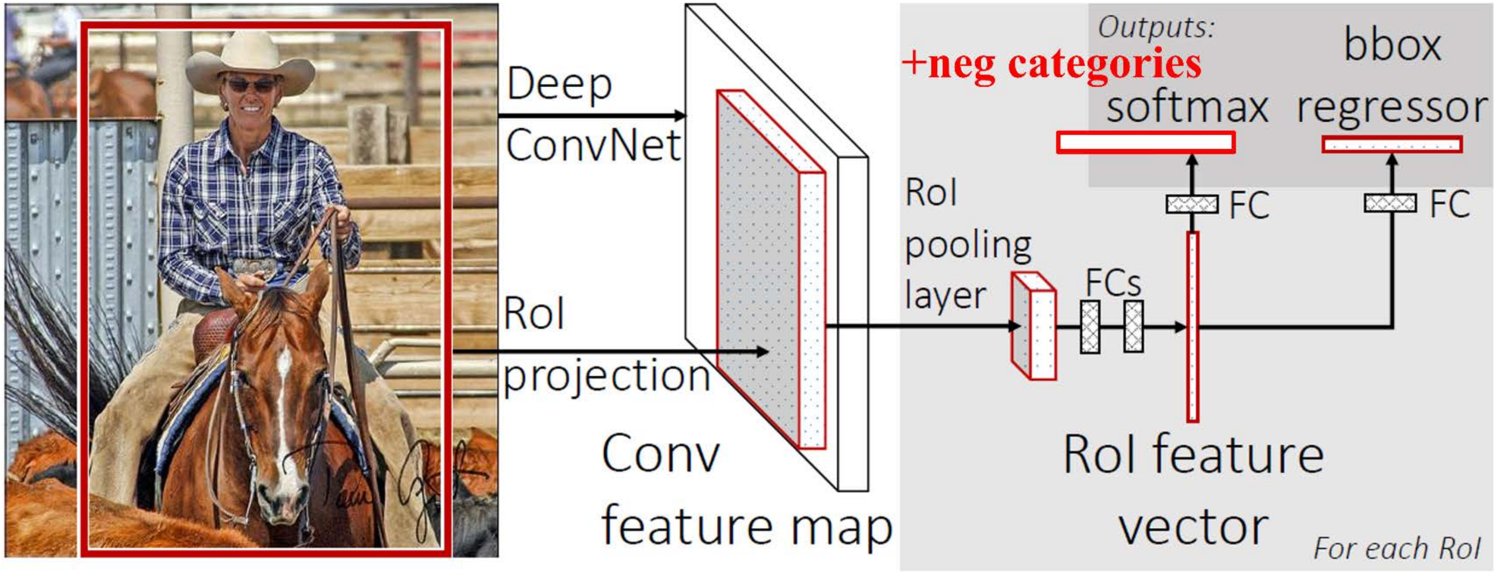
Negative categories – our approach

- **More categories**

  Positive: IOU>=0.5, **Negative**: 0.2<=IOU<0.5, Background: others



Negative categories – our approach



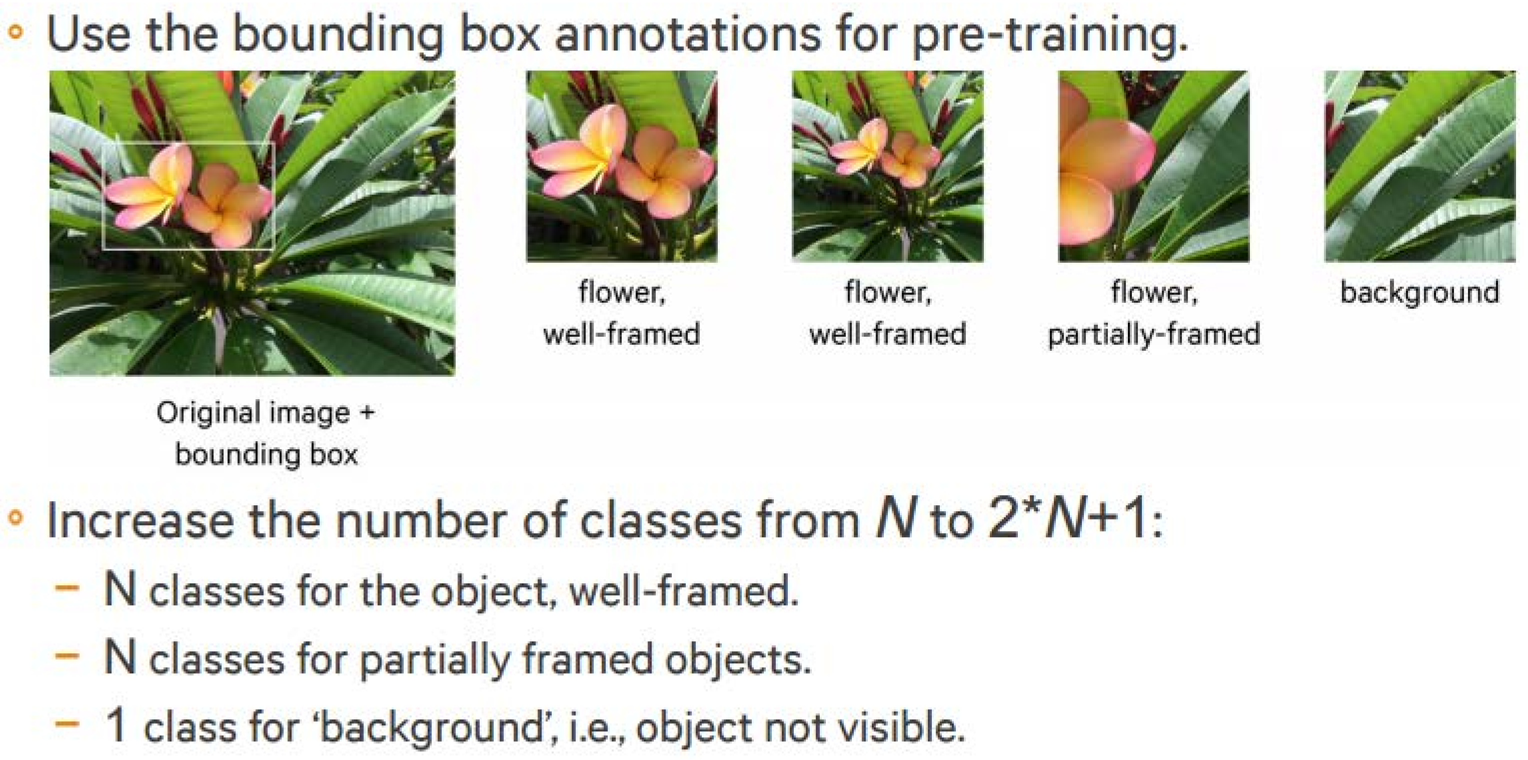
– 401 categories in total

– Regressor trained on pos regions

– Most improved on val: +3.2% mAP (avg 2.2%)

Negative categories – Similar works

- **“Object centric pre-training” by Qualcomm Research**

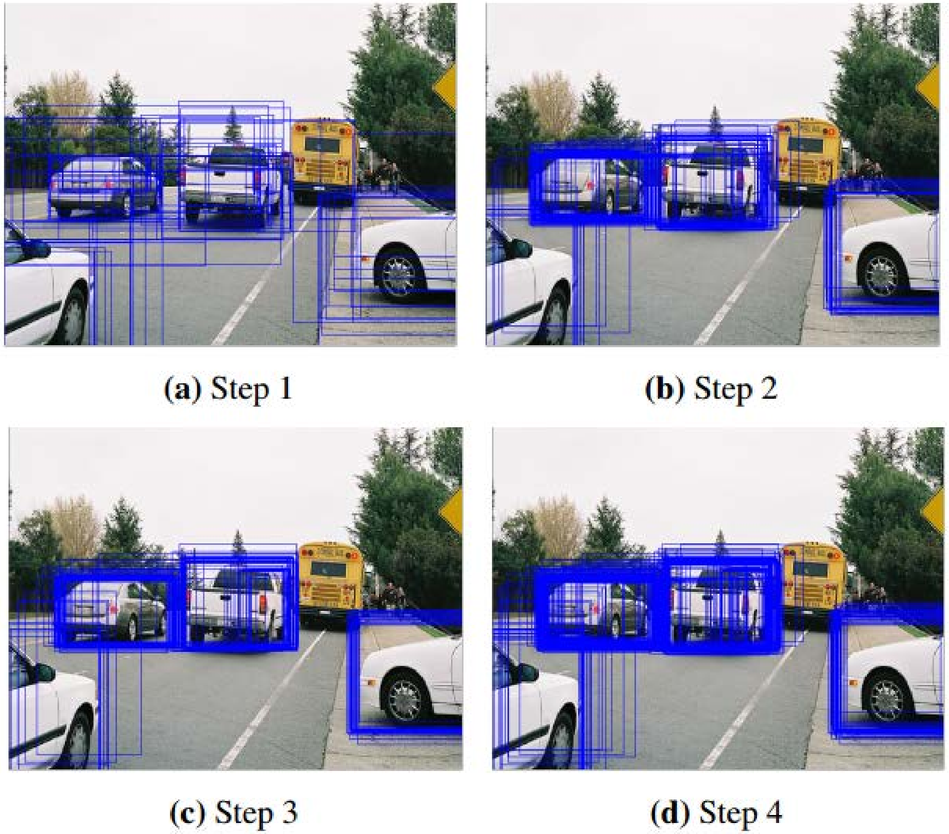


- Use the bounding box annotations for pre-training.

- Increase the number of classes from N to $2*N+1$:
  - N classes for the object, well-framed.
  - N classes for partially framed objects.
  - 1 class for ‘background’, i.e., object not visible.

Bounding box voting

- After standard NMS, keep region *b* with highest score in local area
- Select regions *R,* IOU>=0.5
- Voting using *R*U*b*,
  - $Box = \frac{\sum_{i=1}^{k} score_i * bbox_i}{\sum_{i=1}^{k} score_i}$
  - $Keep\ highest\ score$


(a) Step 1

(b) Step 2

(c) Step 3

(d) Step 4

Object detection via a multi-region & semantic segmentation-aware CNN model, Gidaris S, Komodakis N. 2015

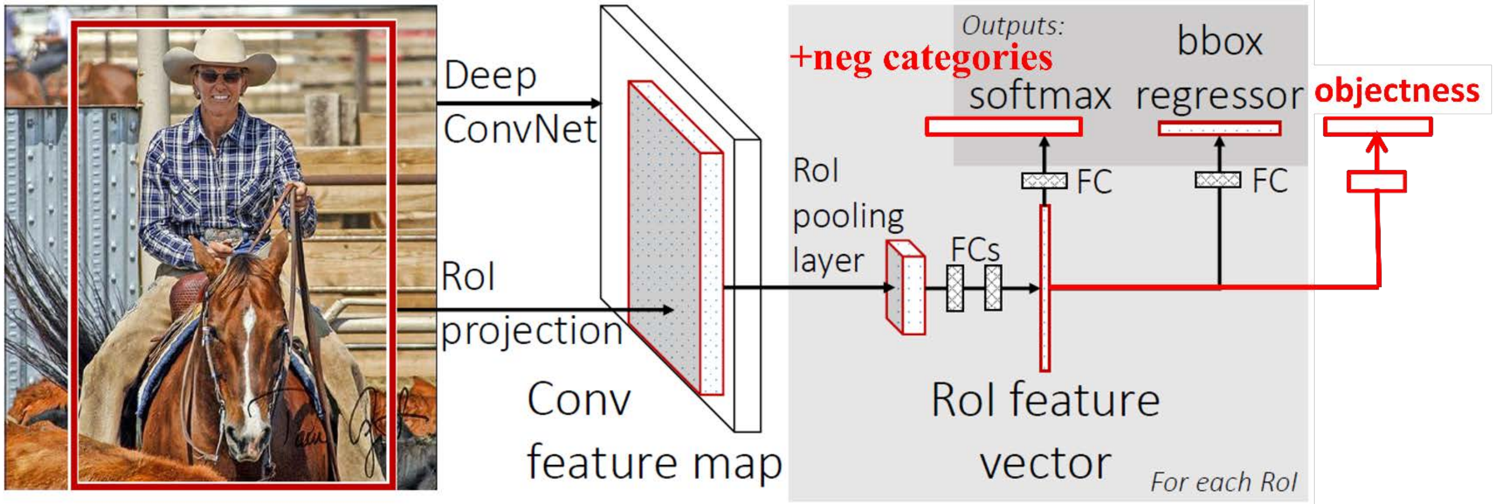
More details

- **Edge Boxes for regions extraction**
- **Pre-train model**
  - VGG16, VGG19, pooling->conv
  - 489 non-overlap subcategories
- **COCO data used in some models**
  - 43 categories with more data
- **Faster-rcnn model**
  - 5x4=20 anchors, ratios(0.2,0.4,1,2,5) and scales(2,3,4,5)
  - Negative categories and objectness in fast stage

Detection results (val set)

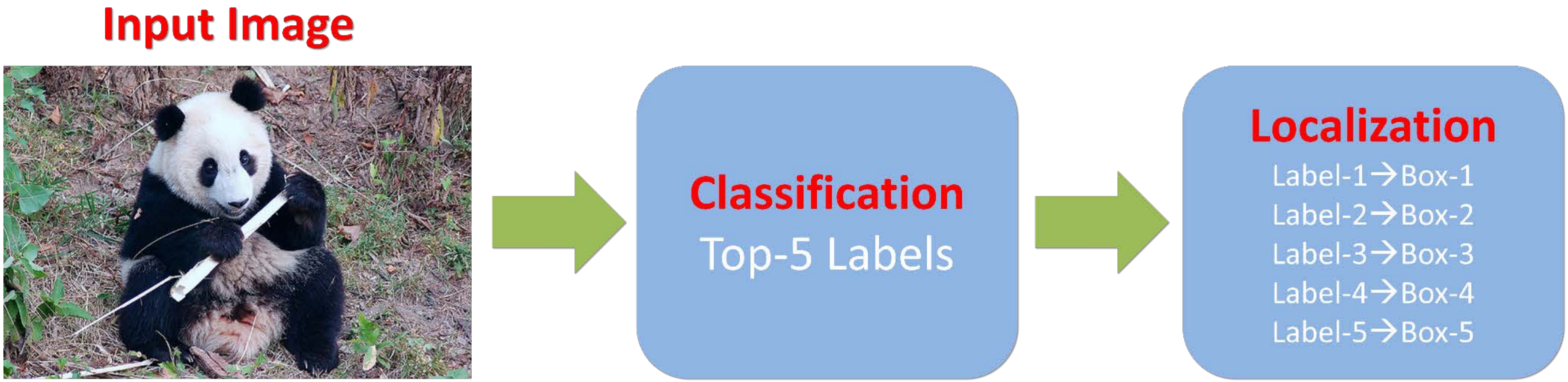
- **Baseline:** VGG16 pre-trained on CLS data

Model	baseline	+obj	+neg	+obj +neg
mAP	43.0	45.2	46.2	46.9



Object Localization

- **Simple pipeline**



Input Image

Classification

Top-5 Labels

Localization

Label-1→Box-1

Label-2→Box-2

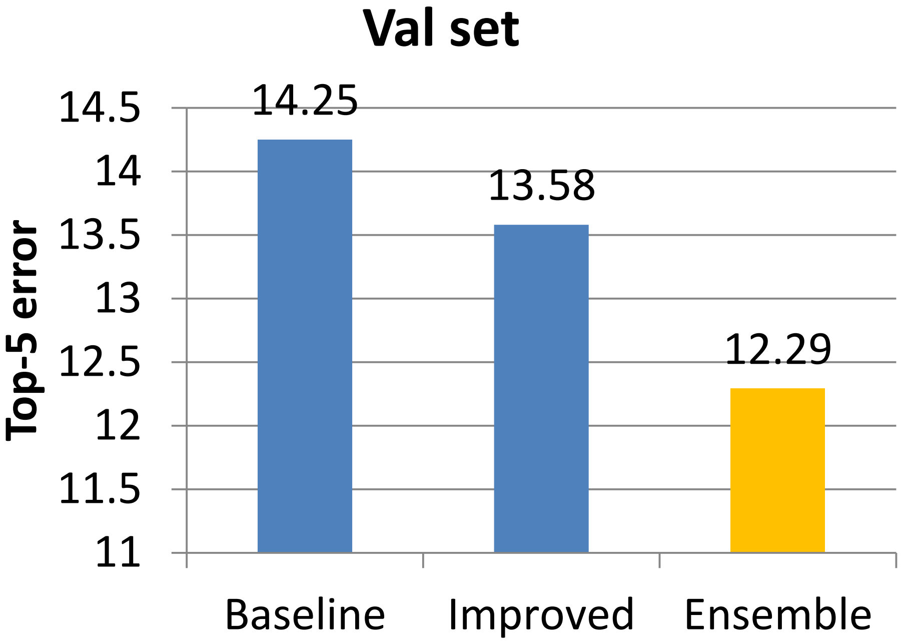
Label-3→Box-3

Label-4→Box-4

Label-5→Box-5

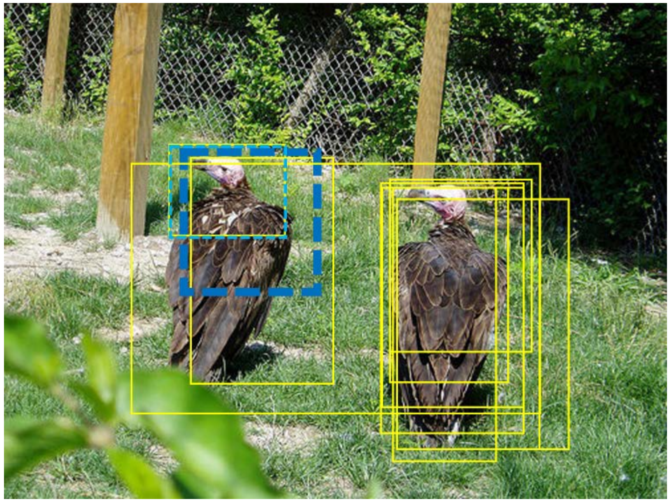
Object Localization

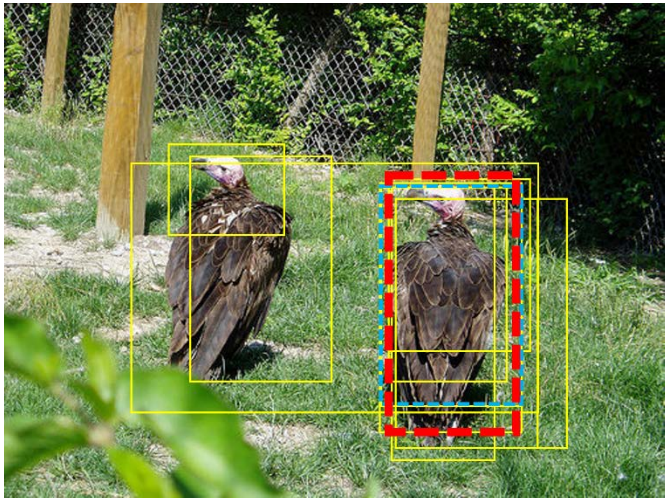
- **Single model improvements**
  - – Objectness loss
  - – Negative categories
  - – Bounding box voting
- **Multi-model**



Object Localization

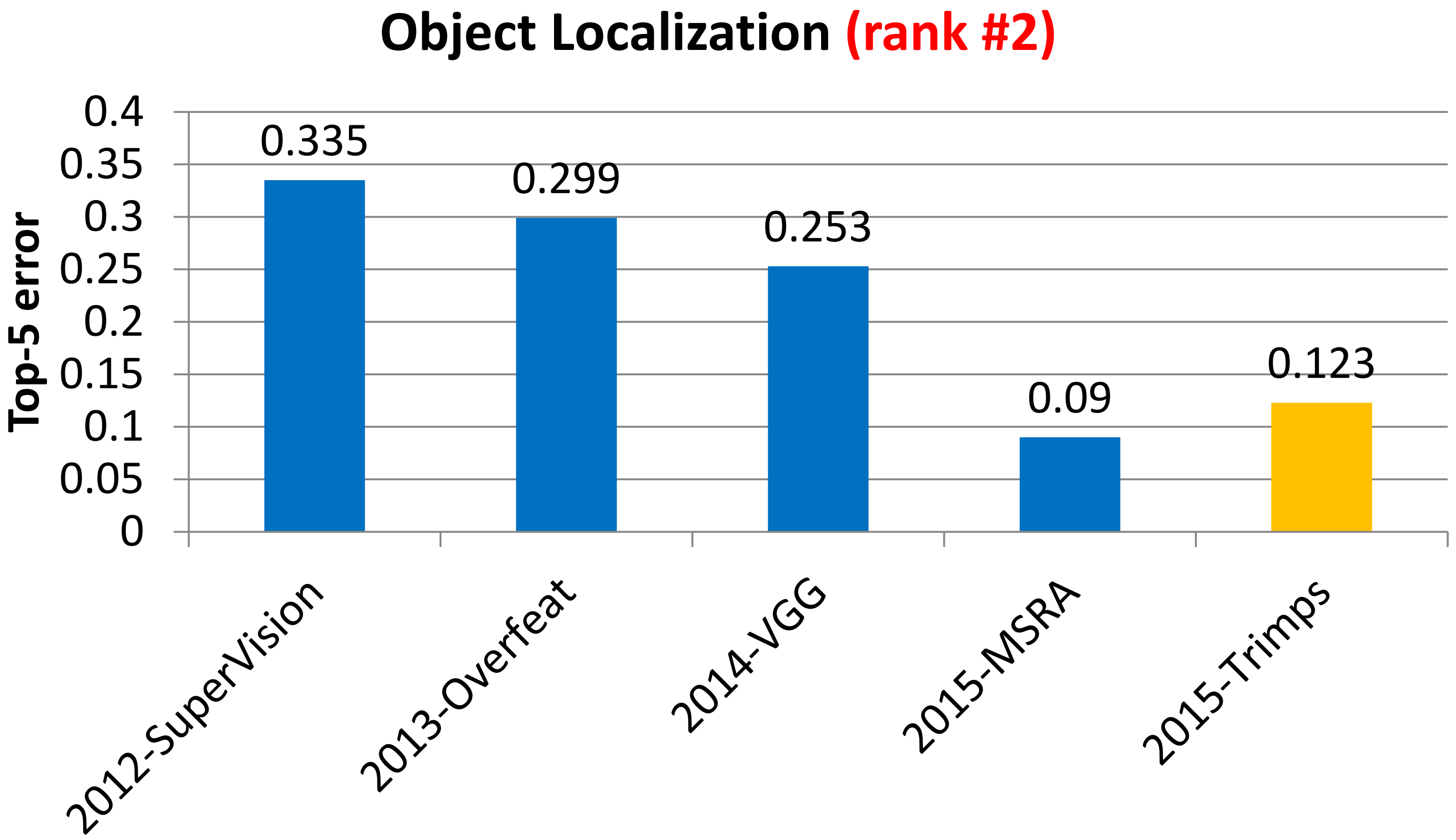
- **Multi-model ensemble (testing)**
  - Bounding box voting (+0.3% vs best single model)
  - Most crowded (not highest scored, +1.4%)





Object Localization

- **Top-5 localization error (test set)**

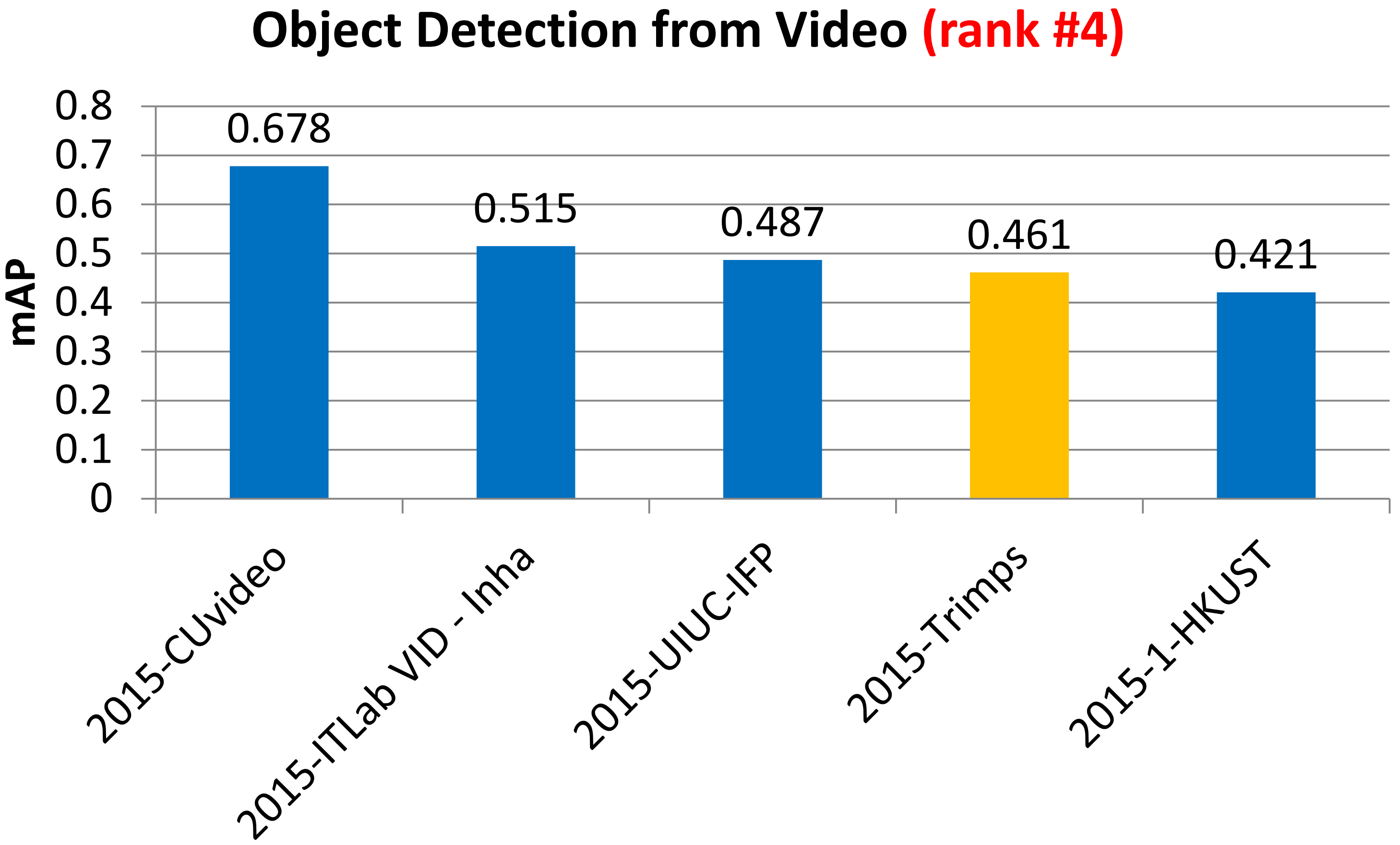


Object Detection from Video

- **From 200 to 30**
  - Using models from object detection task
  - Using video data for fine-tuning
  - Tracking (not finished)

Object Detection from Video

- **Results**



视频图像分析技术挑战赛(筹)

- **组织单位：** 创新论坛组委会主办，公安部第三研究所-上海交通大学智能视频评测联合实验室承办
- **比赛目标：** 提高智能**视频图像**分析技术的研究水平，促进公安实战中的应用
- **任务设置：** 视频图像目标**检测**、视频图像目标**检索**
- **比赛时间：** **2016.09**
- **详细信息将稍后公布**

http://sist.shanghaitech.edu.cn/racv2016/

Thank you!

Q&A