Object Detection

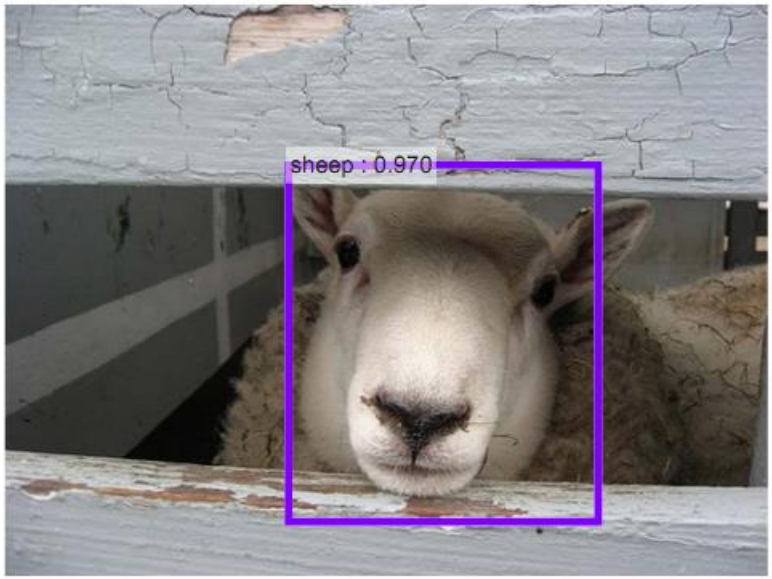
——From RCNN to Faster-RCNN

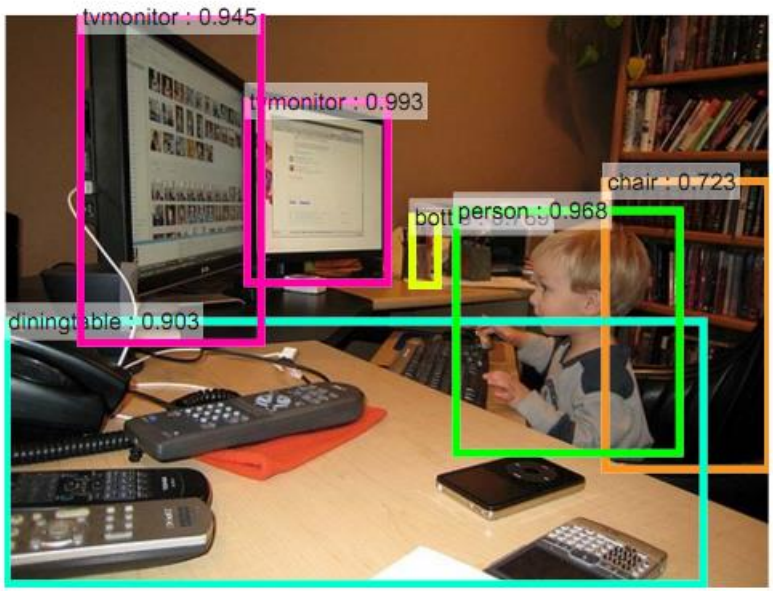
Zhengxia Zou
10/14/2018

- **Background**
- **RCNN**
- **SPPNet**
- **Fast-RCNN**
- **Faster-RCNN**

Background

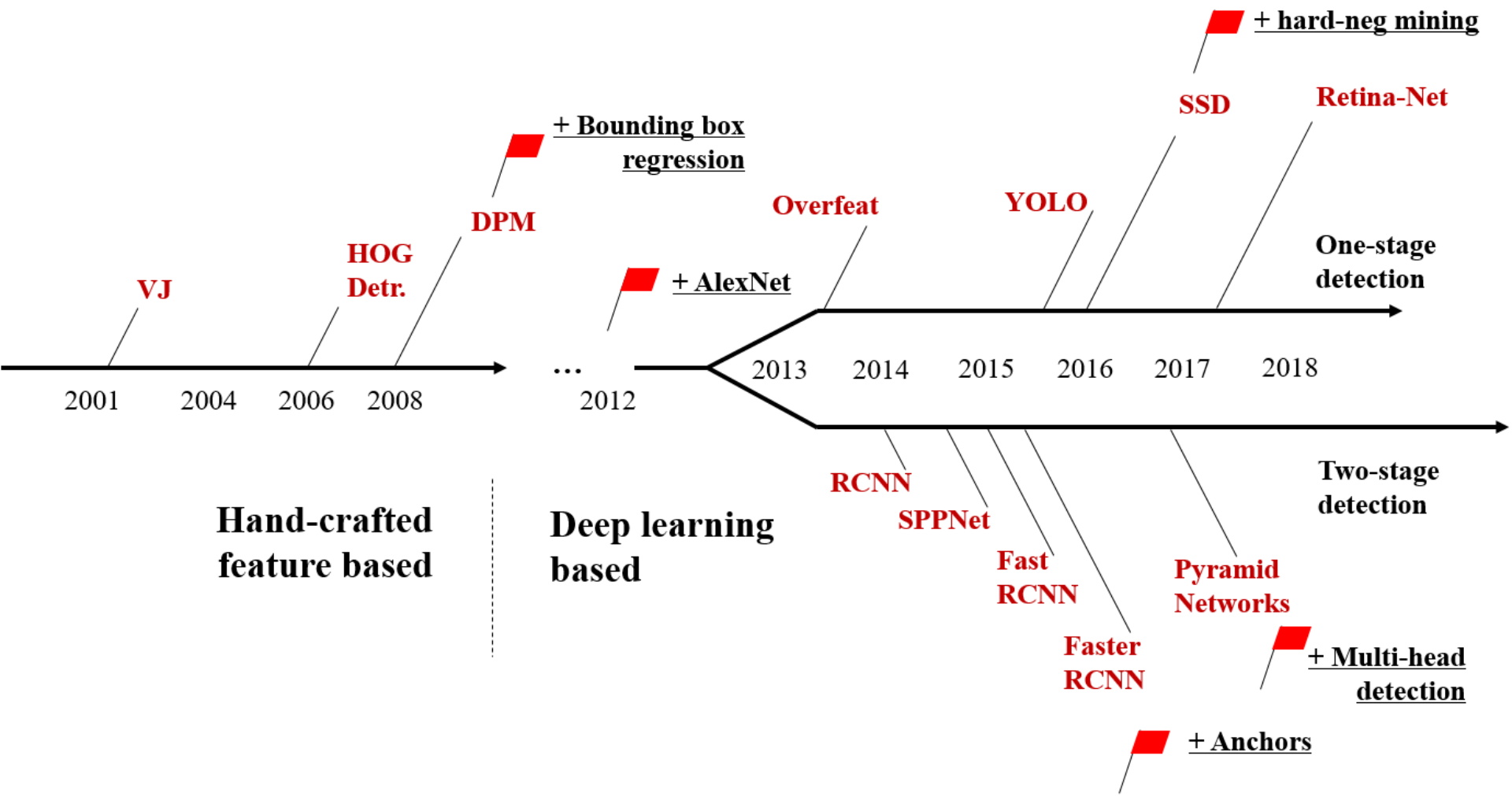
What is ‘Object Detection’?
-- Object bounding box: location and size
-- Object category.





@NIPS15-Faster RCNN

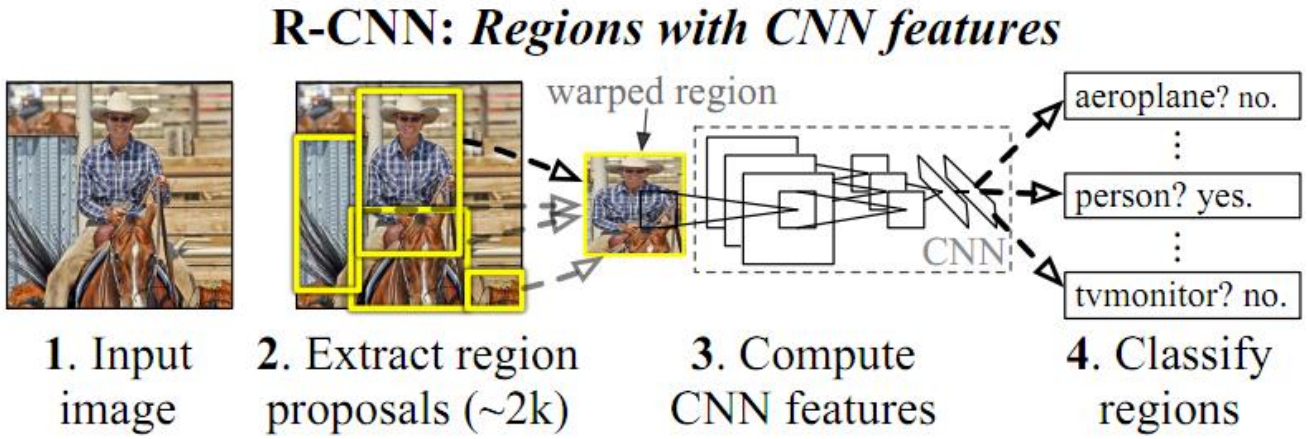
发展历程


+ hard-neg mining
SSD
Retina-Net
+ Bounding box regression
Overfeat
YOLO
DPM
HOG Detr.
VJ
+ AlexNet
One-stage detection
2001
2004
2006
2008
...
2012
2013
2014
2015
2016
2017
2018
RCNN
SPPNet
Fast RCNN
Pyramid Networks
Two-stage detection
Faster RCNN
+ Multi-head detection
+ Anchors
Hand-crafted feature based
Deep learning based

RCNN

R-CNN: Regions with CNN features (arXiv13, CVPR14)

Ross B. Girshick



- Object Proposal+CNN features
- BB Reg.
- Fine tuning
- VOC07 mAP: 33.7→58.5

Time:
14s/image on a GPU

@RBG-cvpr2014

RCNN

R-CNN: Regions with CNN features (arXiv13, CVPR14)

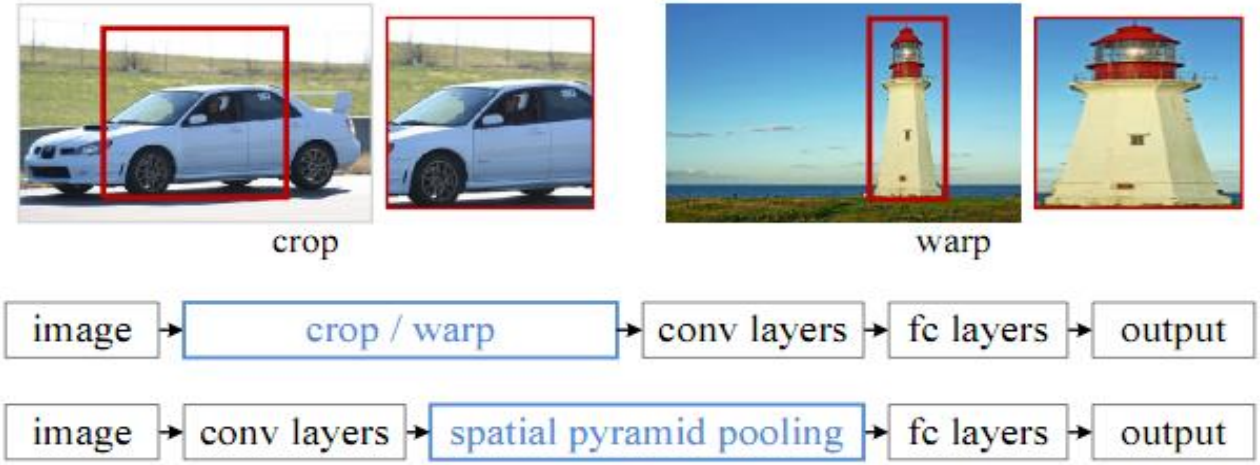
不足

- 训练是多阶段的，很繁琐；
- 训练耗时，占用的磁盘空间过大，特征文件动辄几百G；
- 计算冗余（2000+候选区域），GPU每帧14s。

SPPNet

SPPNet: Spatial Pyramid Pooling (arXiv14, ECCV14)

Kaiming He et al.



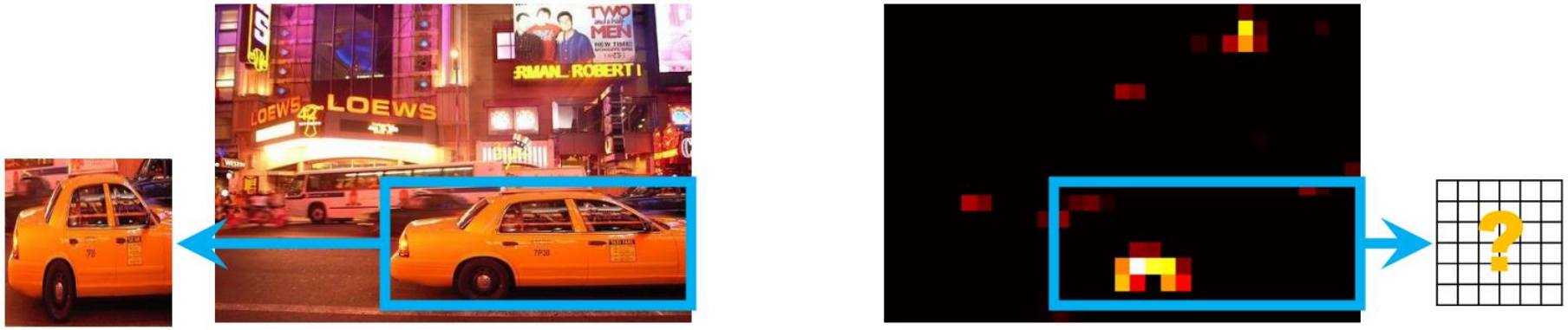
@He-eccv2014

- 共享计算
- SPP→Roipooling (Fast和Faster的基础)
- Fine tuning时只对全连接微调
- VOC07 mAP: 58.5→59.2

Time:
0.38s/image on a GPU

SPPNet

Regions on Feature Maps



- Fixed-length features are required by fully-connected layers or SVM
- But how to produce a fixed-length feature from a feature map region?
- Solutions in traditional compute vision: Bag-of-words, SPM…

ICCV15 International Conference on Computer Vision

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition". ECCV 2014.

@He-iccv2015 tutorial

SPPNet

Microsoft Research

Bag-of-words & Spatial Pyramid Matching

SIFT/HOG-based feature maps

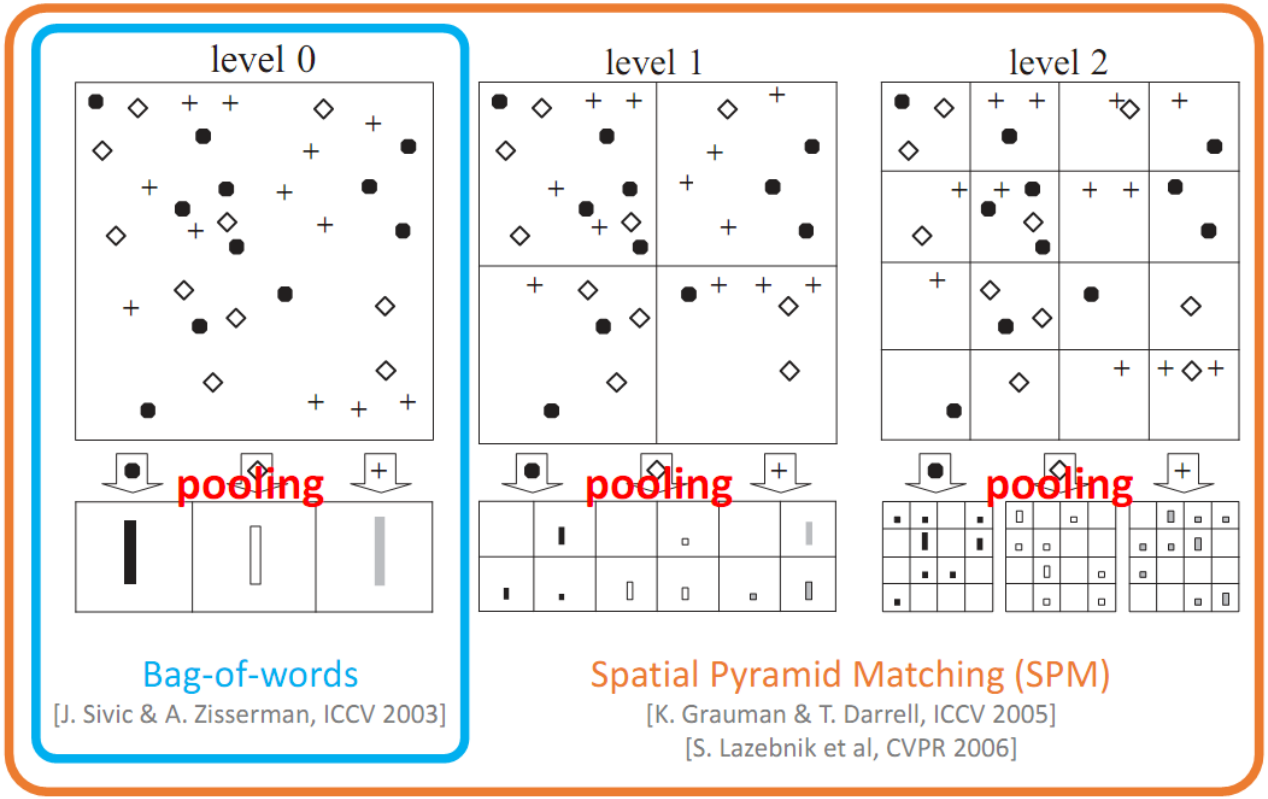


figure credit: S. Lazebnik et al.

ICCV15 International Conference on Computer Vision

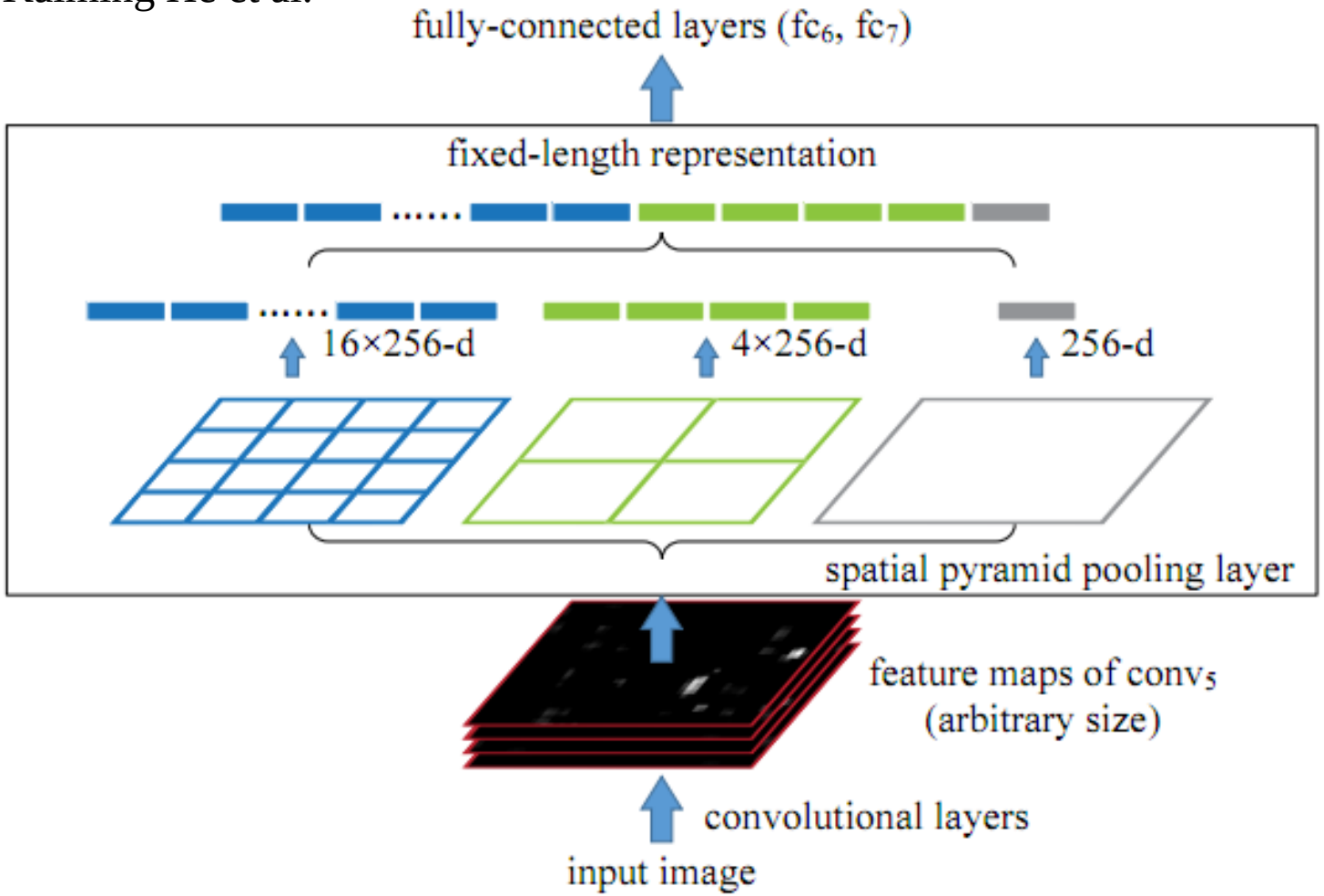
Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition". ECCV 2014.

@He-iccv2015 tutorial

SPPNet

SPPNet: Spatial Pyramid Pooling (arXiv14, ECCV14)

Kaiming He et al.



@He-eccv2014

SPPNet

SPPNet: Spatial Pyramid Pooling (arXiv14, ECCV14)

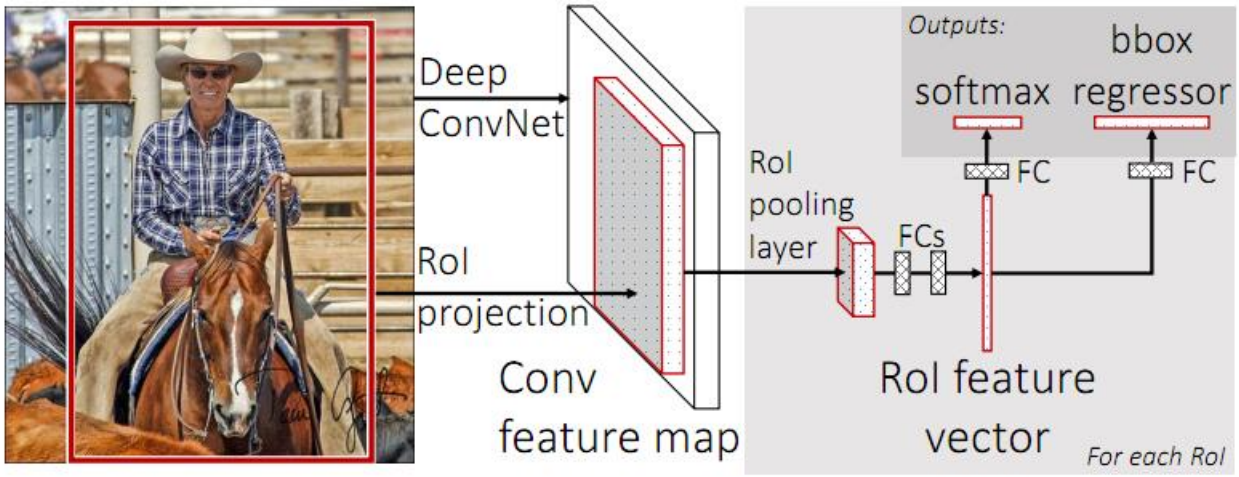
不足

- 训练依然是多阶段的；
- SPPNet在微调网络时，固定了卷积层的参数，只对全连接层进行了微调。

Fast RCNN

Fast R-CNN (arXiv15, ICCV15)

Ross B. Girshick



@RBG-ICCV2015

- ROI Pooling
- Multi-task loss (Clc. + BB Reg.)
- BP through RoI pooling layers
- VOC07 mAP: 58.5→70.0

Time:
0.32s/image on a GPU

@RBG-ICCV2015

Fast RCNN

Fast R-CNN (arXiv15, ICCV15)

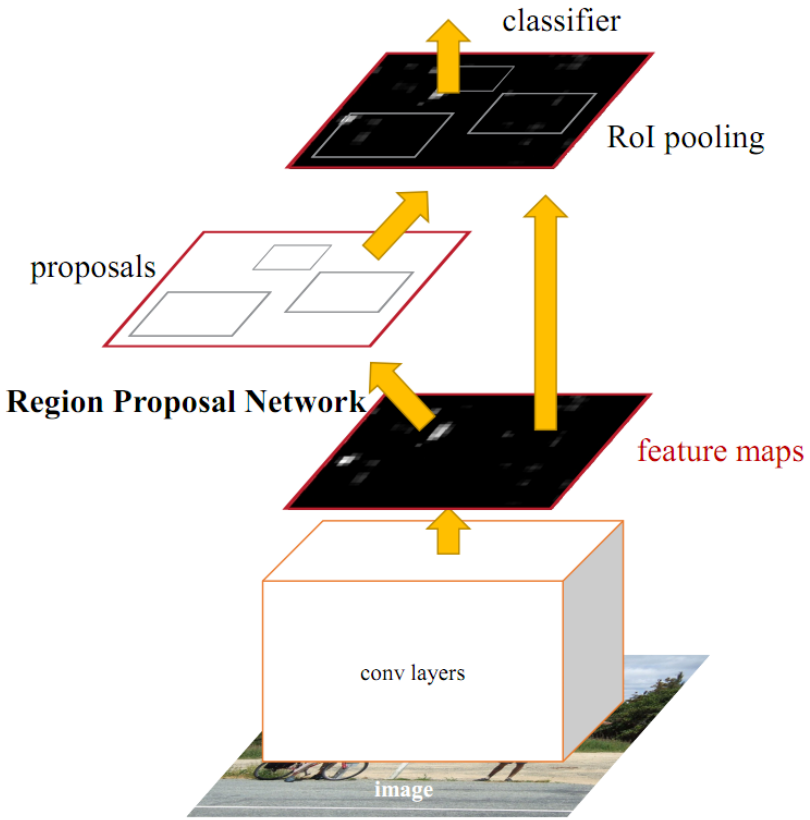
不足

- 不是端到端处理过程
- 需要事先使用Selective Search或Edge Boxes来提取目标后选框（速度瓶颈）。

Faster RCNN

Faster R-CNN (arXiv15, NIPS15)

Shaoqing Ren, Kaiming He, Ross B. Girshick, Jian Sun



- Region Proposal Network
- Detection Network
- Sharing Features
- VOC07 mAP: 70.0→78.8

Time:
17 fps on a GPU

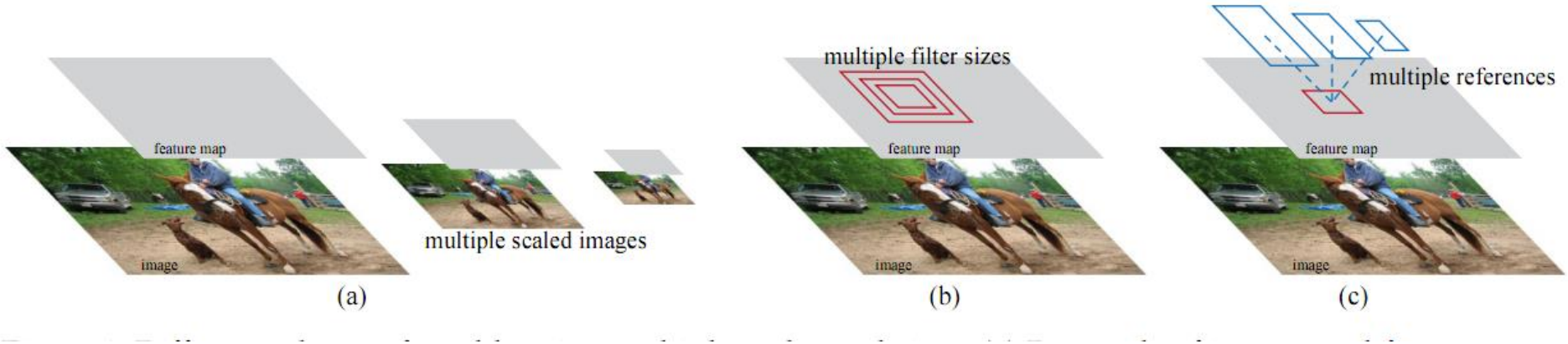
@RBG-NIPS2015

Faster RCNN

Faster R-CNN (arXiv15, NIPS15)

Shaoqing Ren, Kaiming He, Ross B. Girshick, Jian Sun

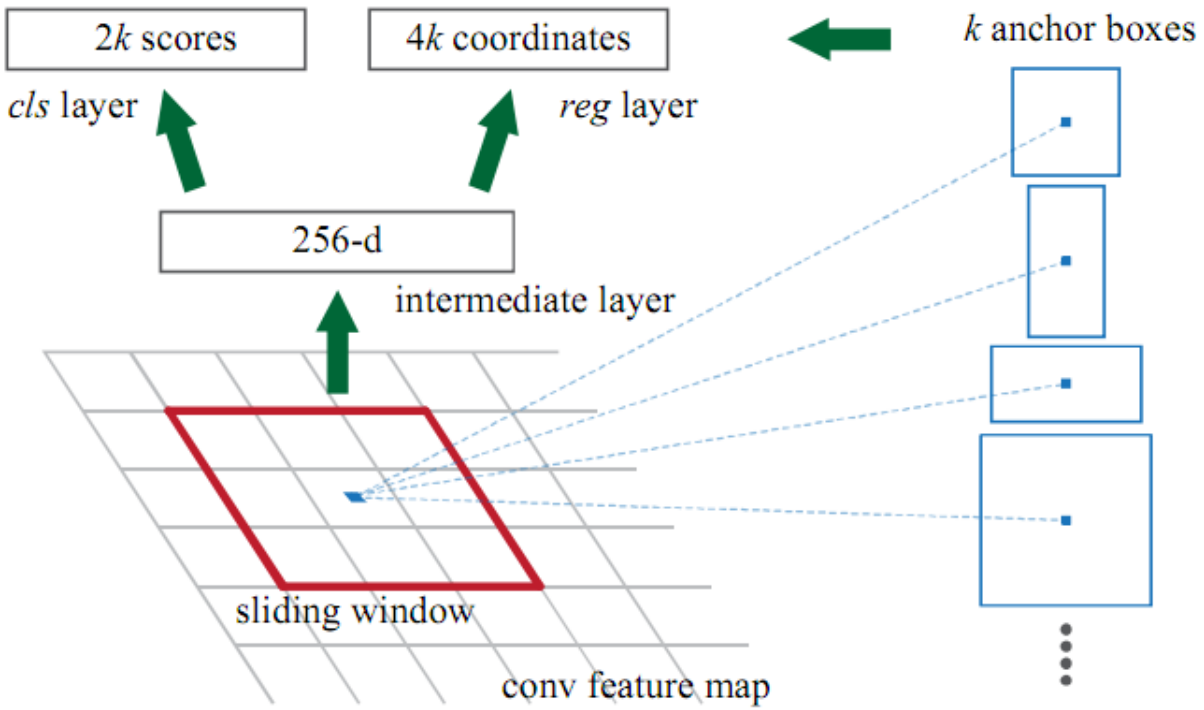
Fast R-CNN 如何处理RPN中的多尺度问题？



@RBG-NIPS2015

Faster RCNN

Anchors (Multi-reference Detection)



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

Faster RCNN

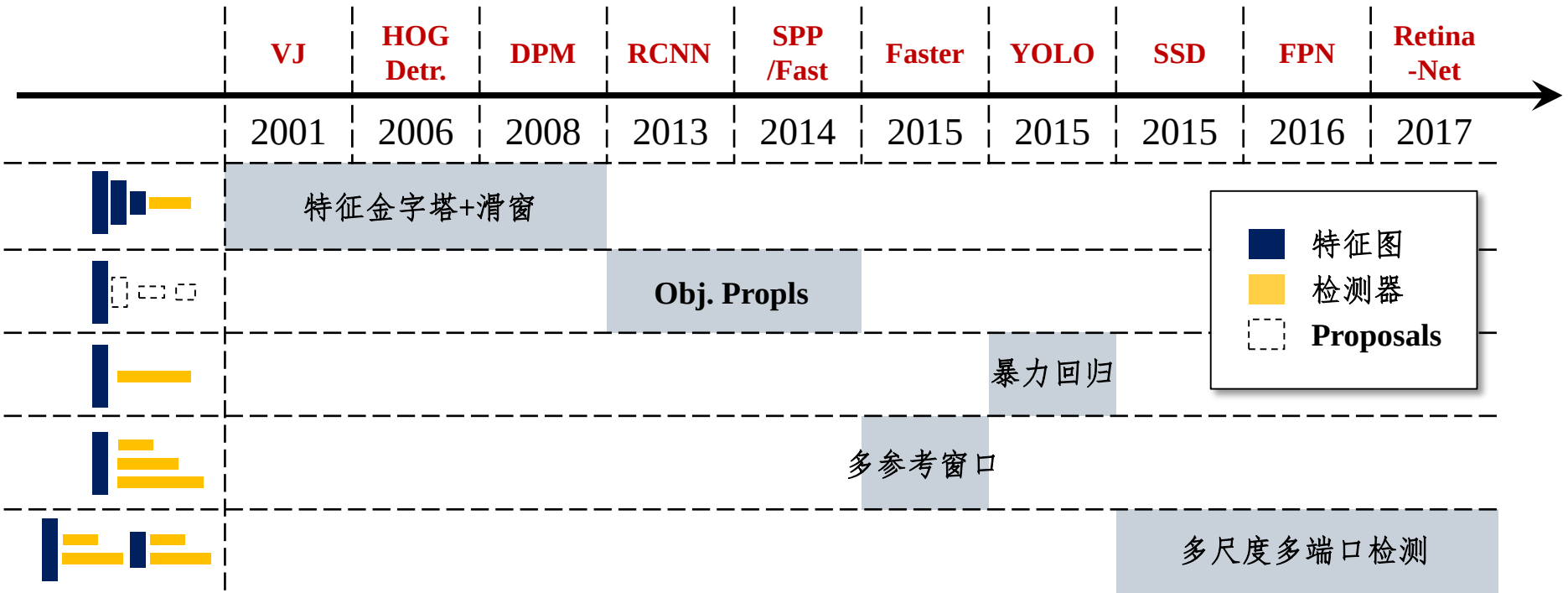
Faster R-CNN (arXiv15, NIPS15)

不足

- 没有结合多尺度特征
- 没有结合Context info
- Roi-pooling时仍然有计算冗余！

目标检测中的关键技术——多尺度检测

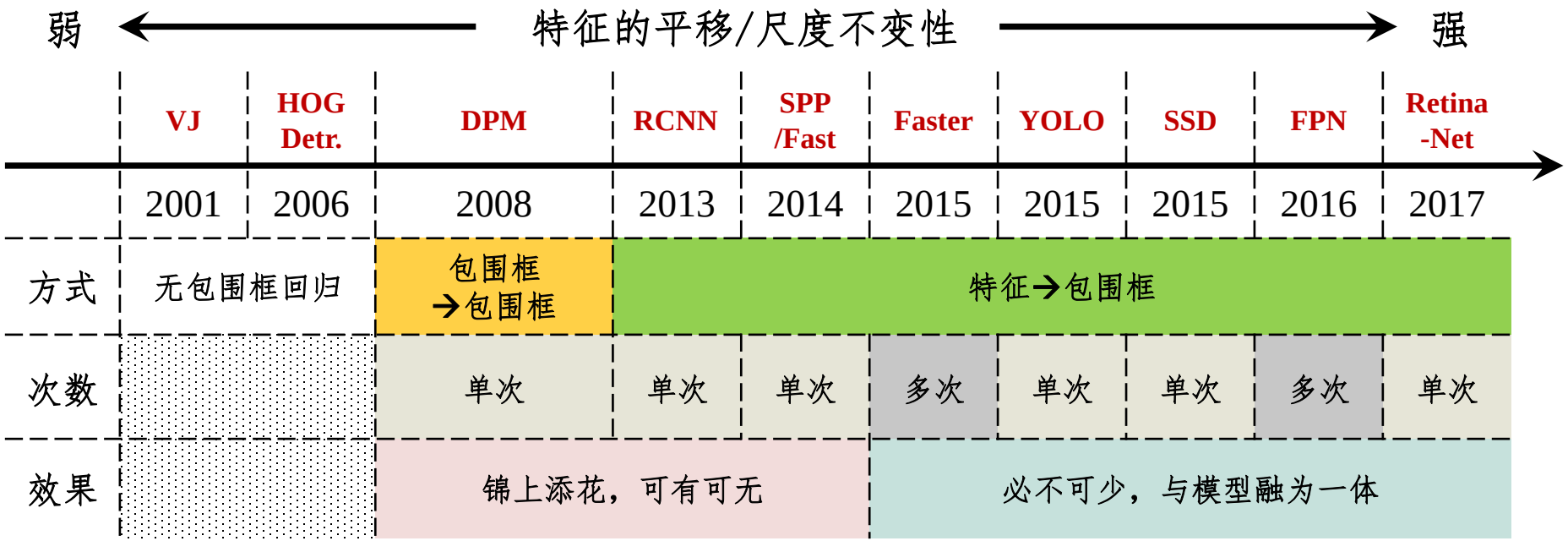
为了检测不同大小、不同长宽比的目标，早期算法大多采用“特征金字塔+滑窗”，后期发展出一系列效率更高的方法。



自然图像目标检测算法多尺度检测方式一览

目标检测中的关键技术——包围框回归

包围框回归经历了从无到有，从“锦上添花，可有可无”到最后“必不可少，与模型融为一体”的发展历程。



自然图像目标检测中包围框回归发展一览

目标检测中的关键技术——加速策略

手工特征时期（早期）的算法的加速策略丰富多样，而深度学习时期（后期）的加速模式较为固定。

	VJ	HOG Detr.	DPM	RCNN	SPP /Fast	Faster	YOLO	SSD	FPN	Retina -Net
	2001	2006	2008	2013	2014	2015	2015	2015	2016	2017
积分图特征	✔	✔	✘	✘	✘	✘	✘	✘	✘	✘
Cascades	✔	✔	✔	✘	✘	✘	✘	✘	✘	✘
决策树	✔	✔	✔	✘	✘	✘	✘	✘	✘	✘
特征图共享	✔	✔	✔	✘	✔	✔	✔	✔	✔	✔
一体化检测	✔	✔	✔	✘	✘	✘	✔	✔	✘	✔

自然图像目标检测算法加速策略一览