

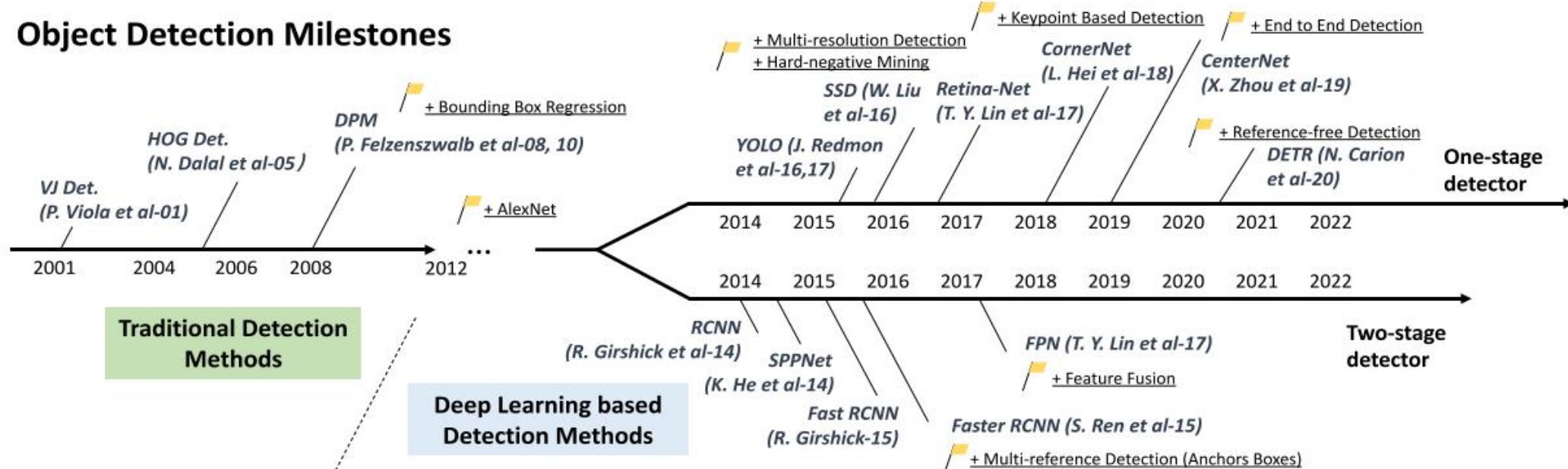
# 1. 背景概述



北京交通大学

## 目标检测

- ▲ 基本定义：是计算机视觉领域的一个重要任务，旨在从视觉数据中定位和识别感兴趣的目标物体。比目标识别更进一步，即不仅要识别目标类别，还要确定目标的位置和边界框。
- ▲ 简单分类：基于传统图像处理的方法和基于深度学习的方法。
- ▲ 研究意义：是理解图像高层语义信息的重要基础，从检测识别物体逐渐延伸到全面场景理解。
- ▲ 发展历程：

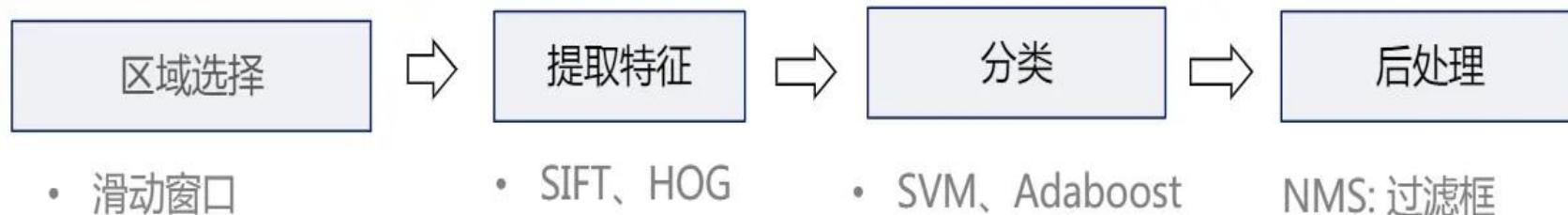


## 2.1 基于传统图像处理的目标检测方法



北京交通大学

### ▲ 主要步骤：



**选取感兴趣的候选区域——>提取候选区域的特征——>对特征进行分类**

#### (1) 选取感兴趣的候选区域：

例如有： a. 滑动窗口法； b. 选择性搜索算法

#### (2) 提取候选区域的特征：

基于手工特征构建，常用的手工特征方法例如有： a. HOG； b. SIFT； c. Harris .....

#### (3) 对特征进行分类：

使用机器学习分类器进行分类，例如有： a. SVM； b. Adaboost； c. NN .....

### ▲ 经典方法：

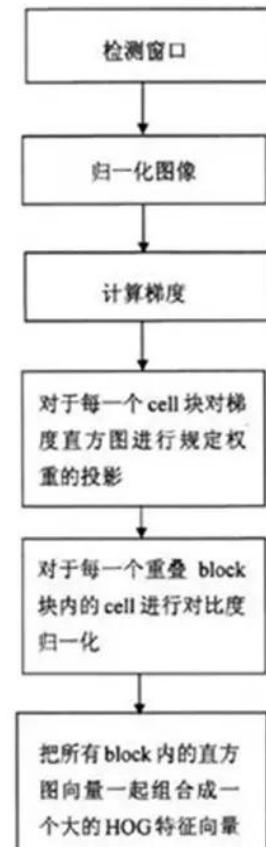
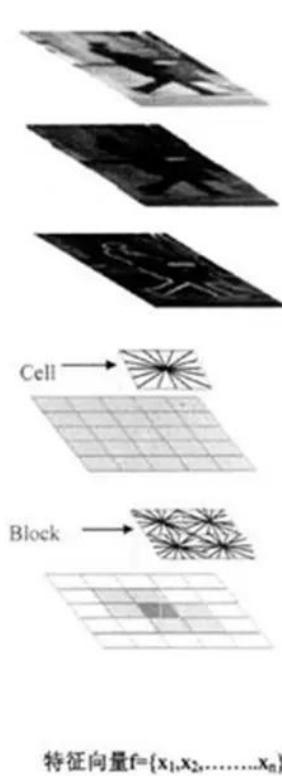
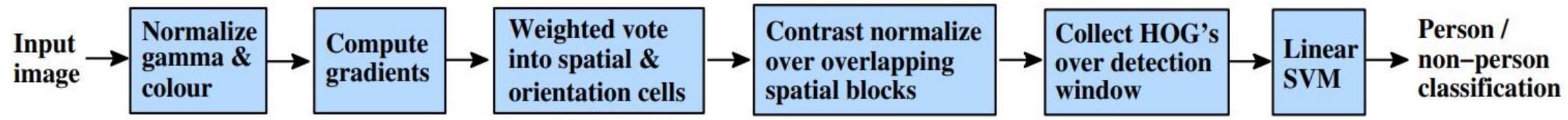
Viola Jones Detectors； HOG Detector； Deformable Part-based Model(DPM)等 .....

# 2.1 基于传统图像处理的目标检测方法

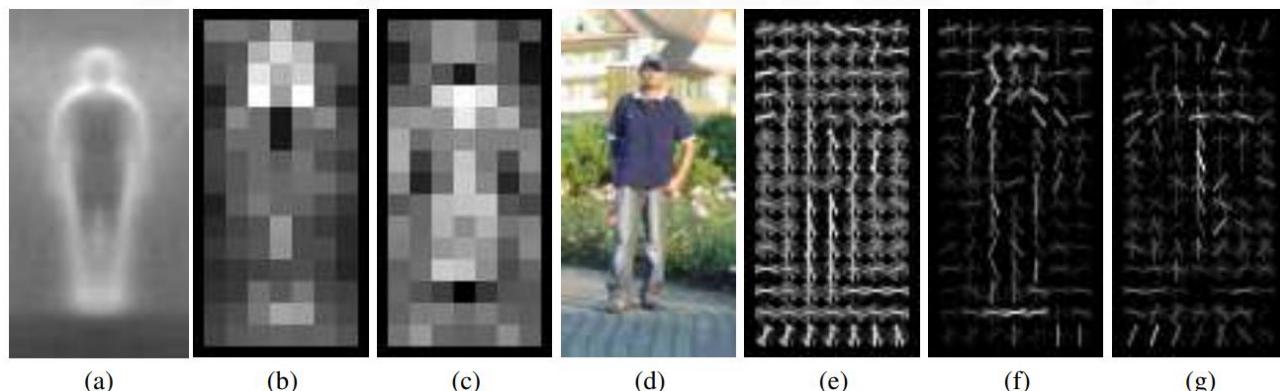


北京交通大学

## ▲ HOG目标检测方法



- (1) 对输入图进行灰度处理和**gamma矫正**；
- (2) 计算图像中每个像素的**梯度大小**和**梯度方向**；
- (3) 为每个单元格中的梯度构建**梯度直方图**；
- (4) 以区域块为单位获取特征；
- (5) 获取整个图像的梯度直方图得到HOG特征；
- (6) 使用线性分类器SVM进行识别分类。



## 2.2 基于深度学习的目标检测方法



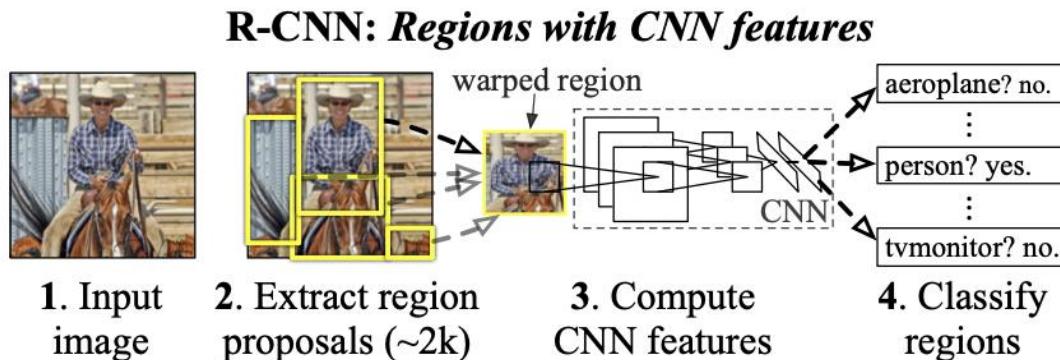
北京交通大学

### Two-stage检测方法

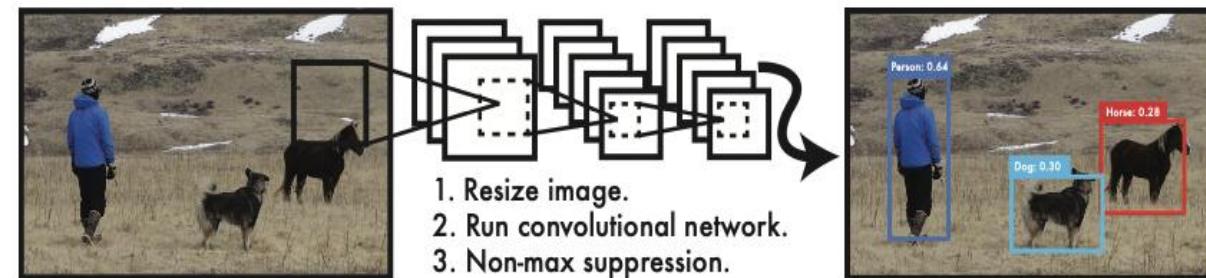
VS

### One-stage检测方法

- ▲ 两个步骤：  
首先**生成候选区域**(region proposals);  
然后对候选区域进行分类。
- ▲ **精度较高**, 适用于对准确性要求较高的场景。
- ▲ 常见的two-stage检测算法有：  
**R-CNN、SPPNet、FPN、Fast R-CNN、Faster R-CNN等**。



- ▲ 直接在图像上密集采样一系列候选框，判断每个候选框内是否存在目标，并直接对目标进行分类。
- ▲ **简单高效、实时性好**, 适用于实时应用场景。
- ▲ 常见的one-stage目标检测算法有：  
**YOLO系列、SSD、Retina-Net、CornerNet、CenterNet、DETR等**。



## 2.2 基于深度学习的目标检测方法

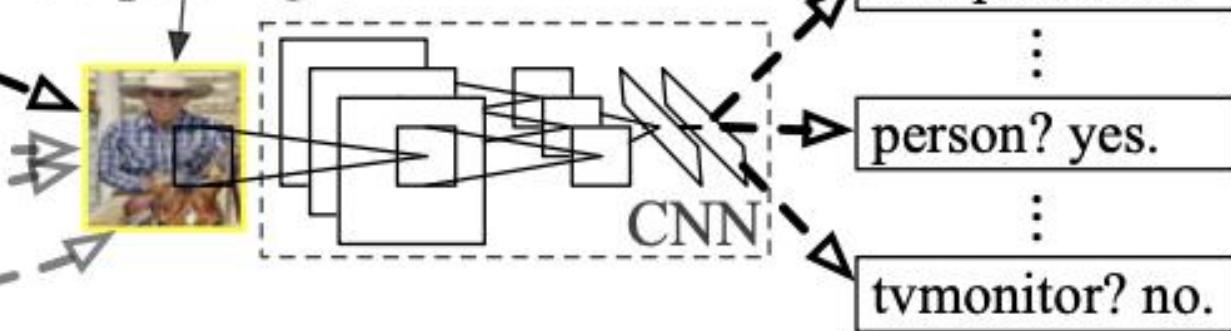


北京交通大学

### ▲ R-CNN (两阶段方法)



warped region



1. Input  
image

2. Extract region  
proposals (~2k)

3. Compute  
CNN features

4. Classify  
regions

生成候选区域 → 通过CNN提取特征 → 使用分类器对特征分类 → 边界框回归

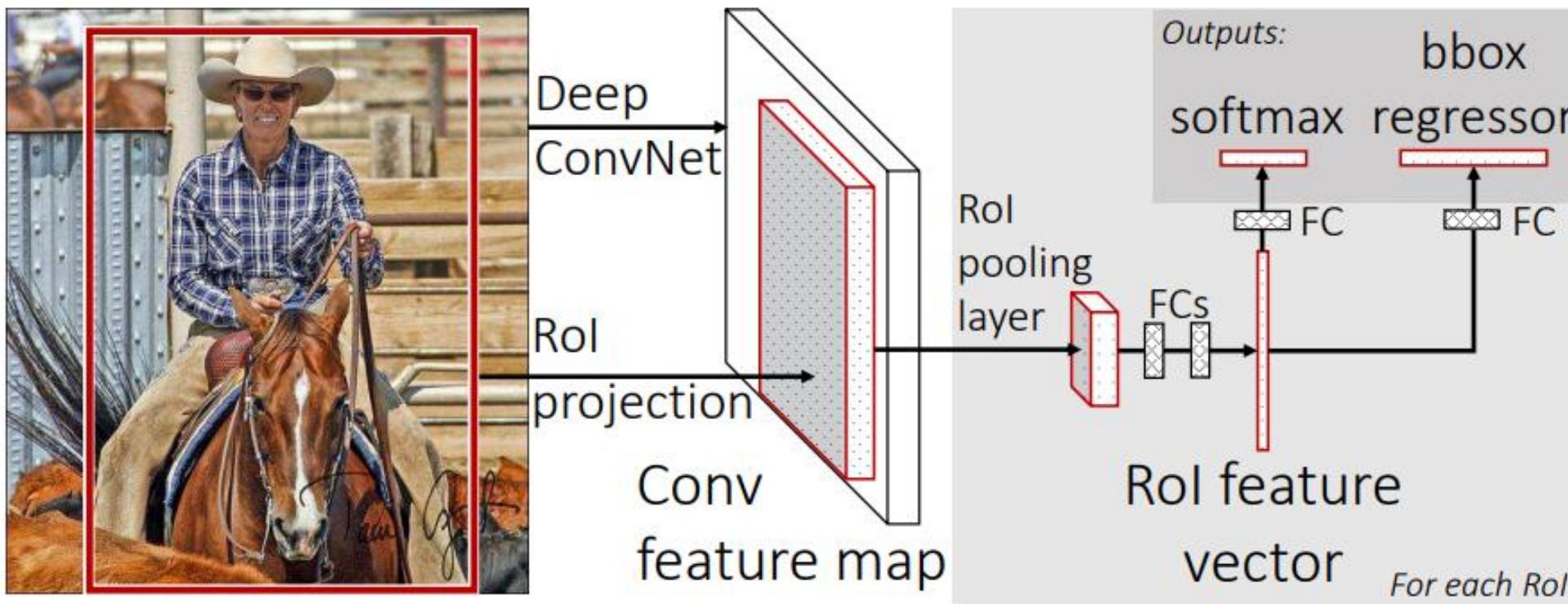
- (1) 生成候选区域：选择性搜索算法(SS)
- (2) 通过CNN提取特征：AlexNet模型
- (3) 特征分类：线性支持向量机(SVM)
- (4) 边界框回归：非极大值抑制(NMS)+候选区域坐标调整

## 2.2 基于深度学习的目标检测方法



北京交通大学

### ▲ Fast R-CNN (两阶段方法)

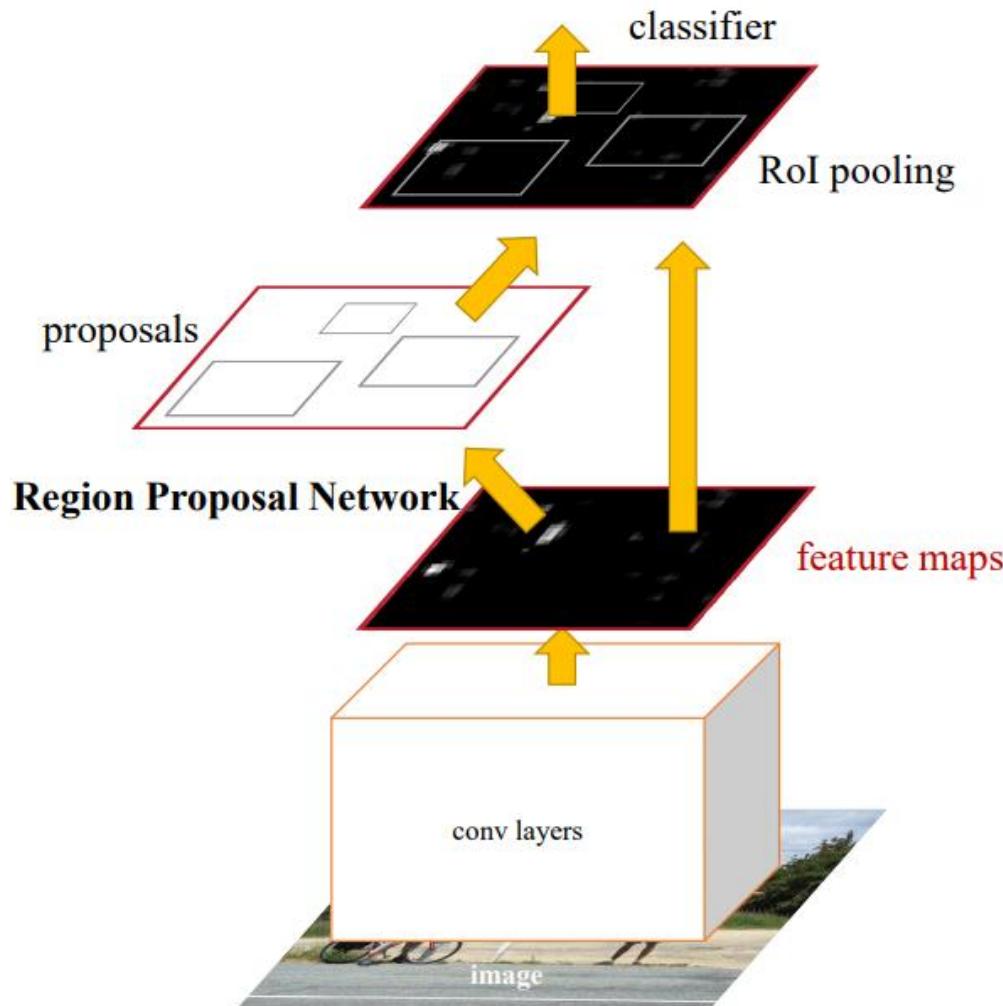


基于R-CNN做出如下改进：

- (1) 直接将原始图像输入CNN进行特征提取，再将候选区域映射到特征图中；
- (2) ROI Pooling层对感兴趣区域进行统一尺寸调整；
- (3) 分别通过Softmax和Regressor进行类别分类与边界框回归。

## 2.2 基于深度学习的目标检测方法

### ▲ Faster R-CNN (两阶段方法)

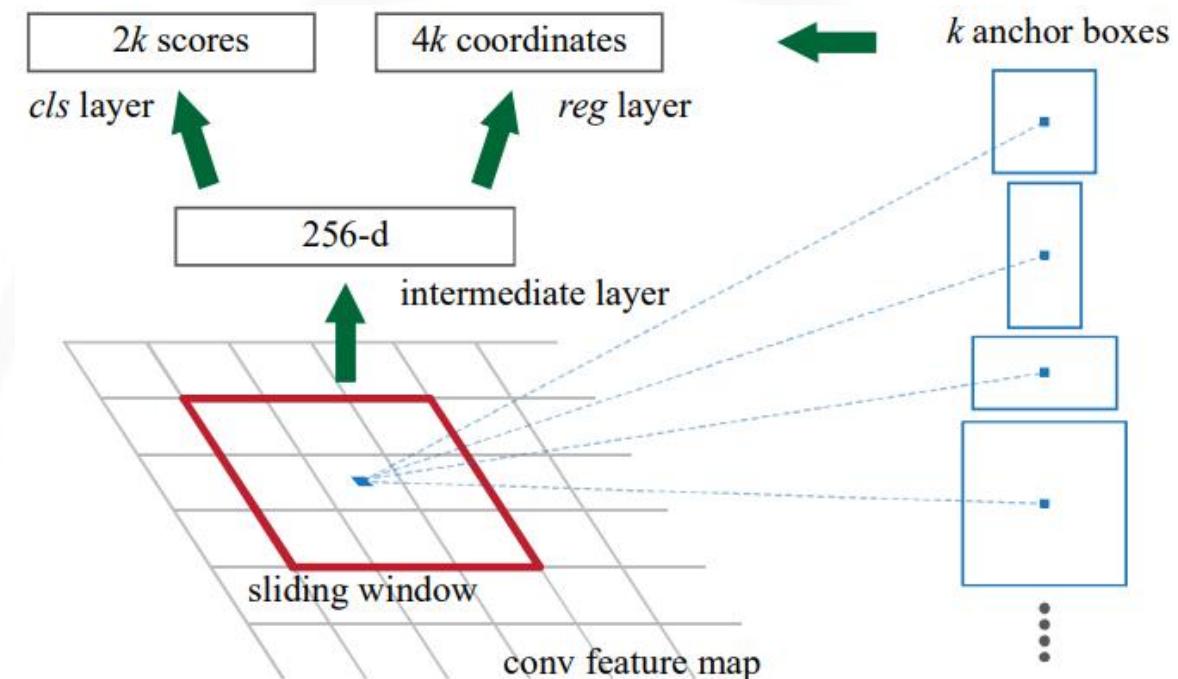


Faster R-CNN结构示意图

可以简单理解为：

**Faster R-CNN = 区域生成网络(RPN) + Fast R-CNN**

Faster R-CNN用RPN代替Fast R-CNN中的选择性搜索算法。



RPN结构示意图

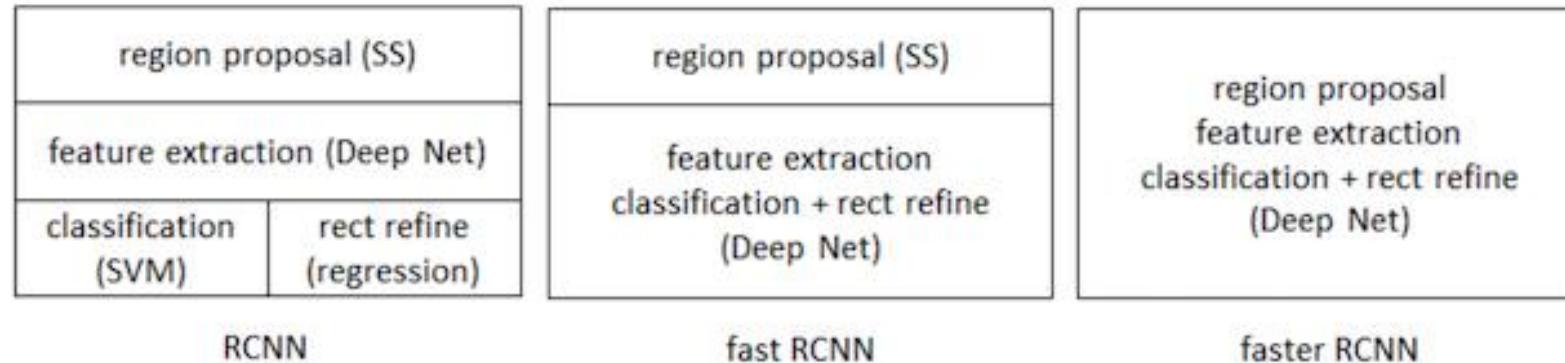


## 2.2 基于深度学习的目标检测方法

### ▲ 两阶段方法小结



北京交通大学



#### ➤ R-CNN缺点:

- (1) 训练阶段多：步骤繁琐，微调网络+训练SVM+训练边框回归器；
- (2) 占用空间大：例如5000张图像产生几百G的特征文件；
- (3) 处理速度慢：使用GPU加速, VGG16模型处理一张图像需要47s；
- (4) 图像变形：候选区域要经过裁剪等方式固定大小，无法保证图像不变形。

#### ➤ Fast R-CNN缺点:

- (1) 仍然使用选择性搜索算法提取候选区域，没有实现真正意义上的端对端；
- (2) 处理与训练仍比较耗时。

#### ➤ Faster R-CNN优缺点:

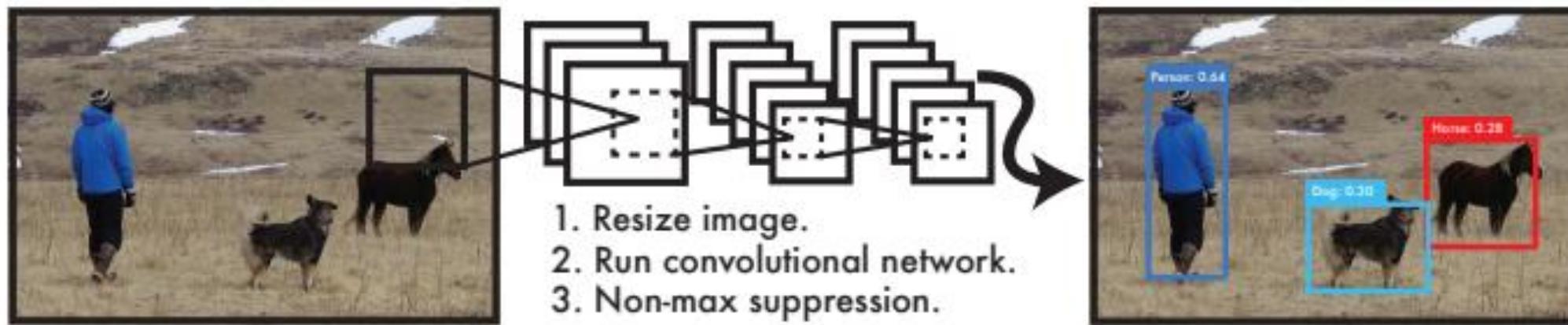
- (1) 优点：使用了RPN网络；**端到端网络模型**；
- (2) 缺点：训练参数过大；
- (3) 改进点：可以考虑提升速度，如YOLO系列算法移除了RPN，**直接对候选框进行分类回归**。

## 2.2 基于深度学习的目标检测方法

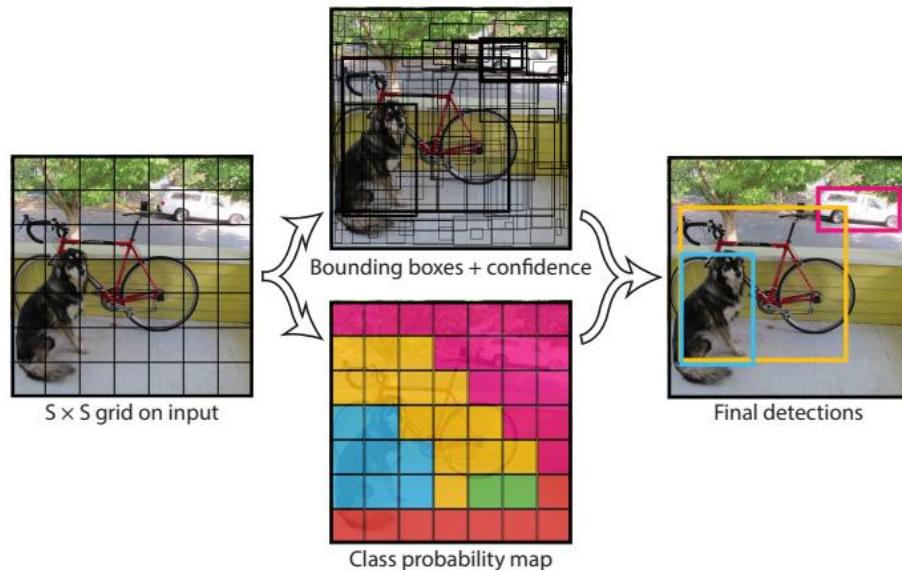


北京交通大学

### ▲ YOLOv1 (单阶段方法)



统一图像尺寸——> 通过CNN提取特征——> 通过FC层进行分类与回归——> 通过NMS筛选最终结果

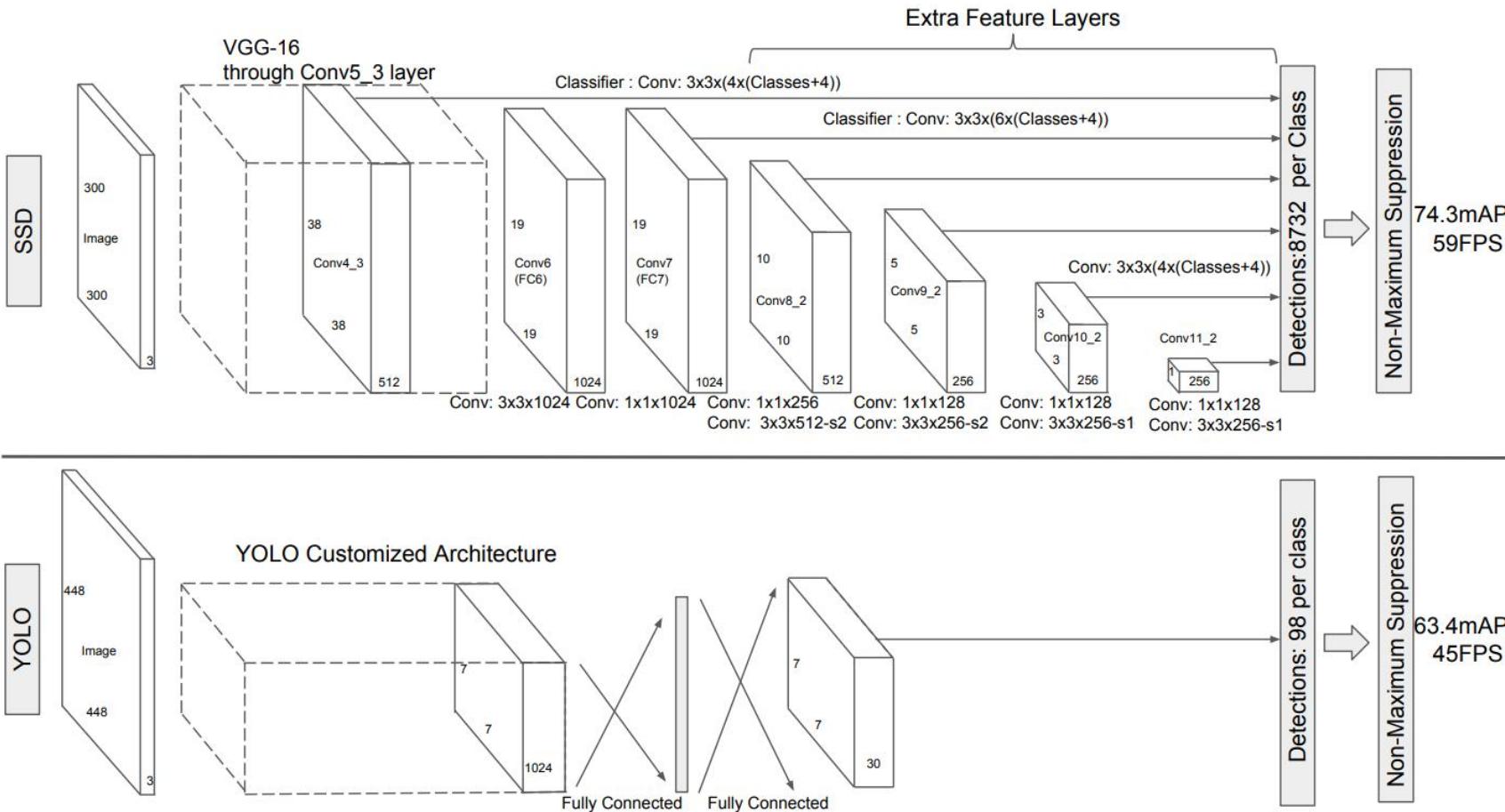


- 与Faster R-CNN相比，YOLOv1没有提取候选区域，直接对整张图像检测，使用CNN同时预测边界框的位置与类别
- 将图像划分为 $S \times S$ 个网格，每个网格内预测B个边界框、对应置信度，及C个类别的概率。这些预测信息被编码成一个张量，大小为：  
 **$S \times S \times (B \times 5 + C)$**   
注：此处的5包括 $x, y, w, h, confidence$

## 2.2 基于深度学习的目标检测方法



### ▲ SSD (单阶段方法)



SSD也是直接对整张图像进行检测，但与YOLOv1相比，具有如下特点：

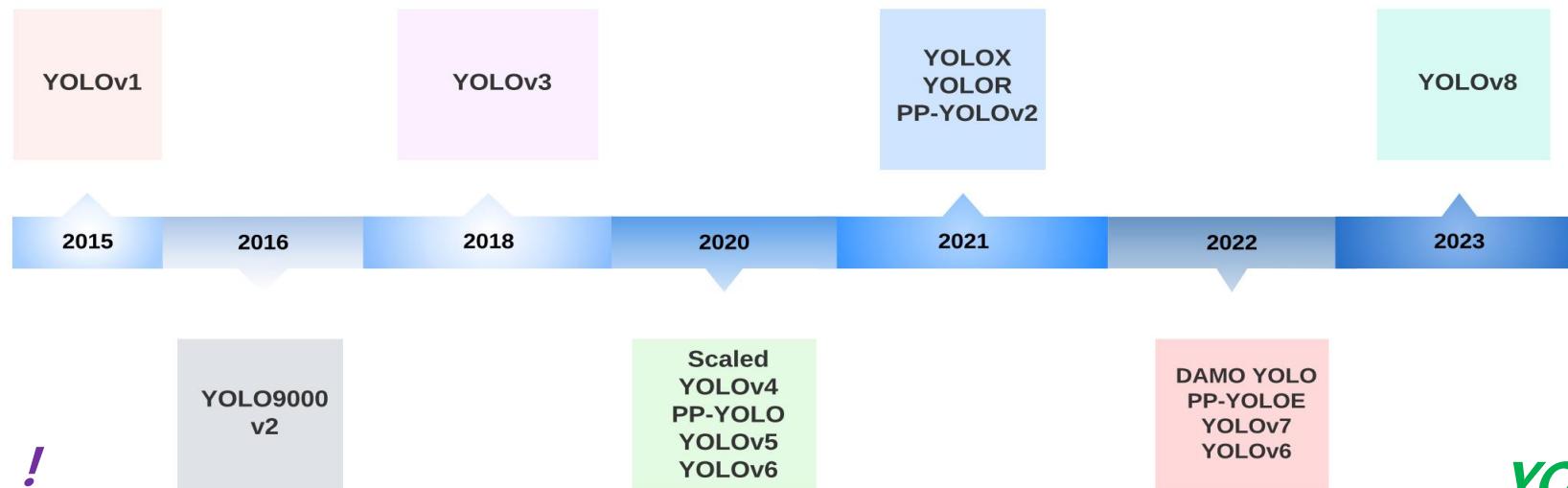
- (1) 通过**多尺度特征图**进行检测
- (2) 使用**卷积替代全连接层**进行结果输出；
- (3) 预设**先验框**辅助**多尺度检测**

## 2.2 基于深度学习的目标检测方法

### ▲ YOLO系列 (单阶段方法)

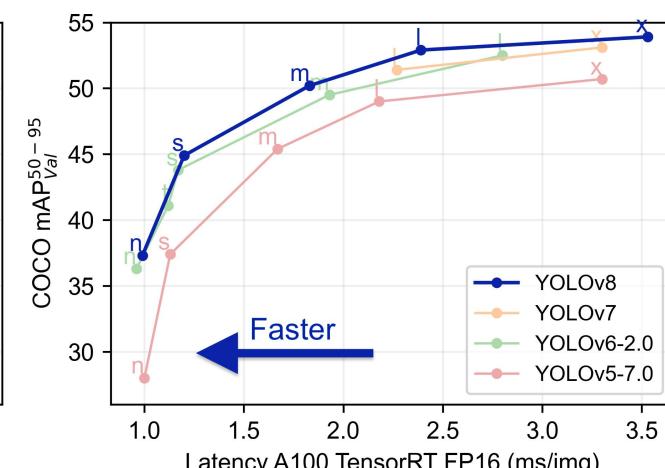
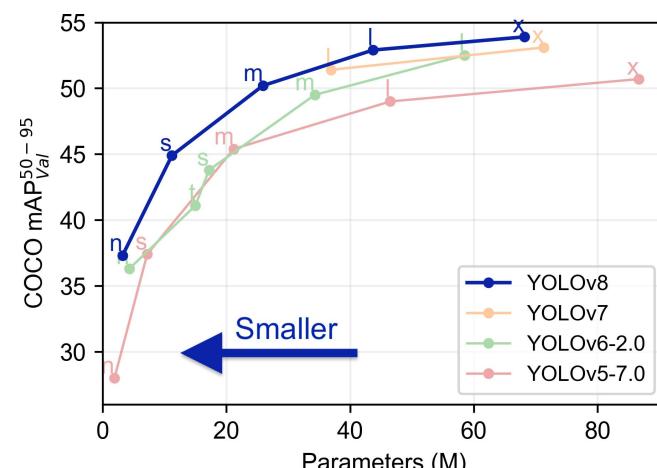
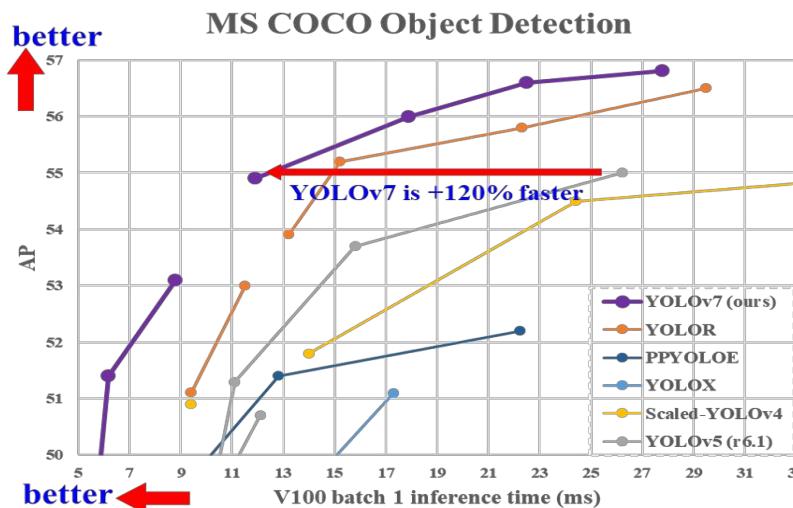


北京交通大学



YOLO 真香！！！

YOLO tql！！！



# 3.1 小目标检测



北京交通大学

▲ 小目标检测(**Small Object Detection**)：旨在检测识别出那些与尺寸较小、面积较小的目标物体。

▲ 何为小目标：

- (1) 若目标的尺寸小于整个图像尺寸的十分之一，可以认为是相对小目标；
- (2) 在COCO数据集中，若目标的尺寸小于 $32 \times 32$ 像素，则认为是绝对小目标。

▲ 难点：特征信息微弱；易受噪声干扰；易受预测边界框扰动影响等.....



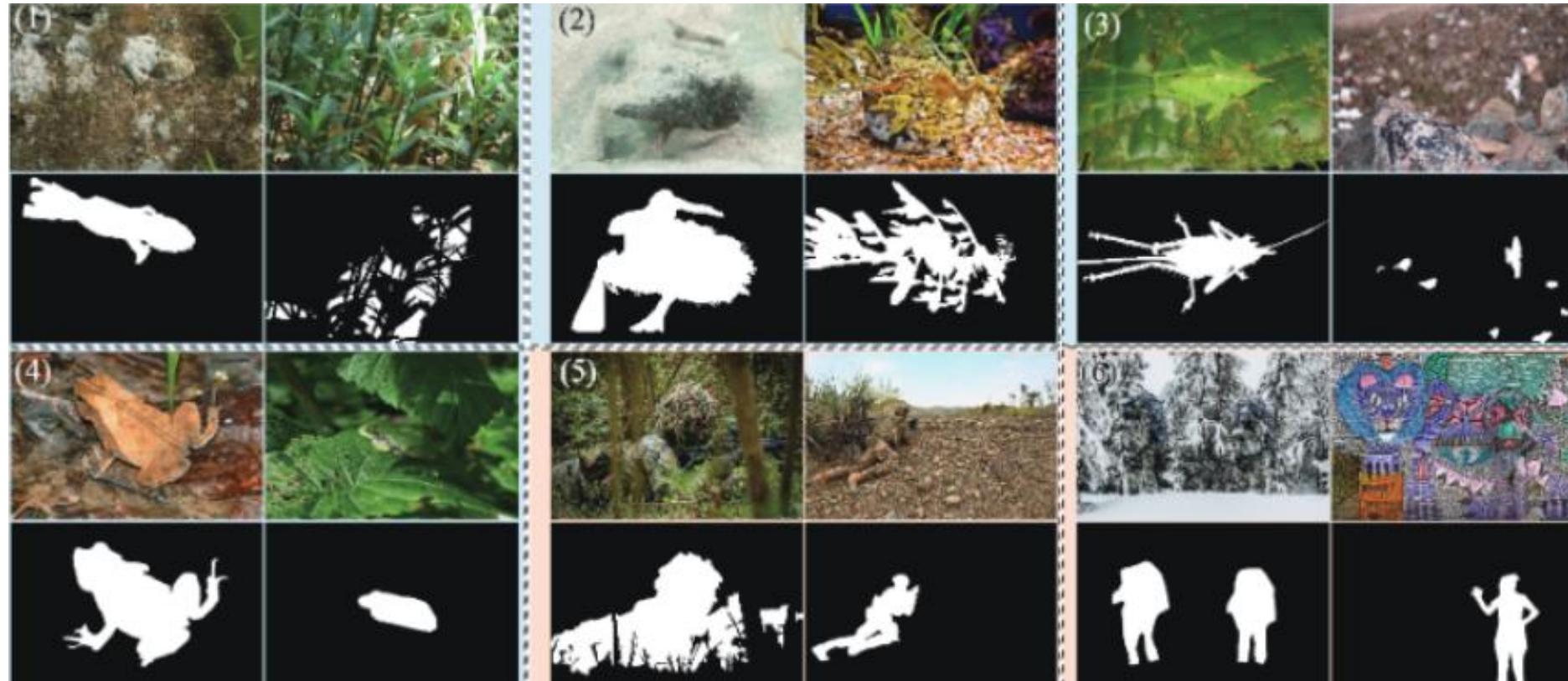
## 3.2 伪装目标检测



北京交通大学

▲ 伪装目标检测(Camouflaged Object Detection): 旨在检测识别出那些与周围环境“无缝”嵌入融合的目标。

▲ 难点: 伪装目标在纹理、颜色、形状等与背景高度相似, 且其边界与周围环境的视觉辨识度极低, 导致检测伪装目标更具挑战性.....

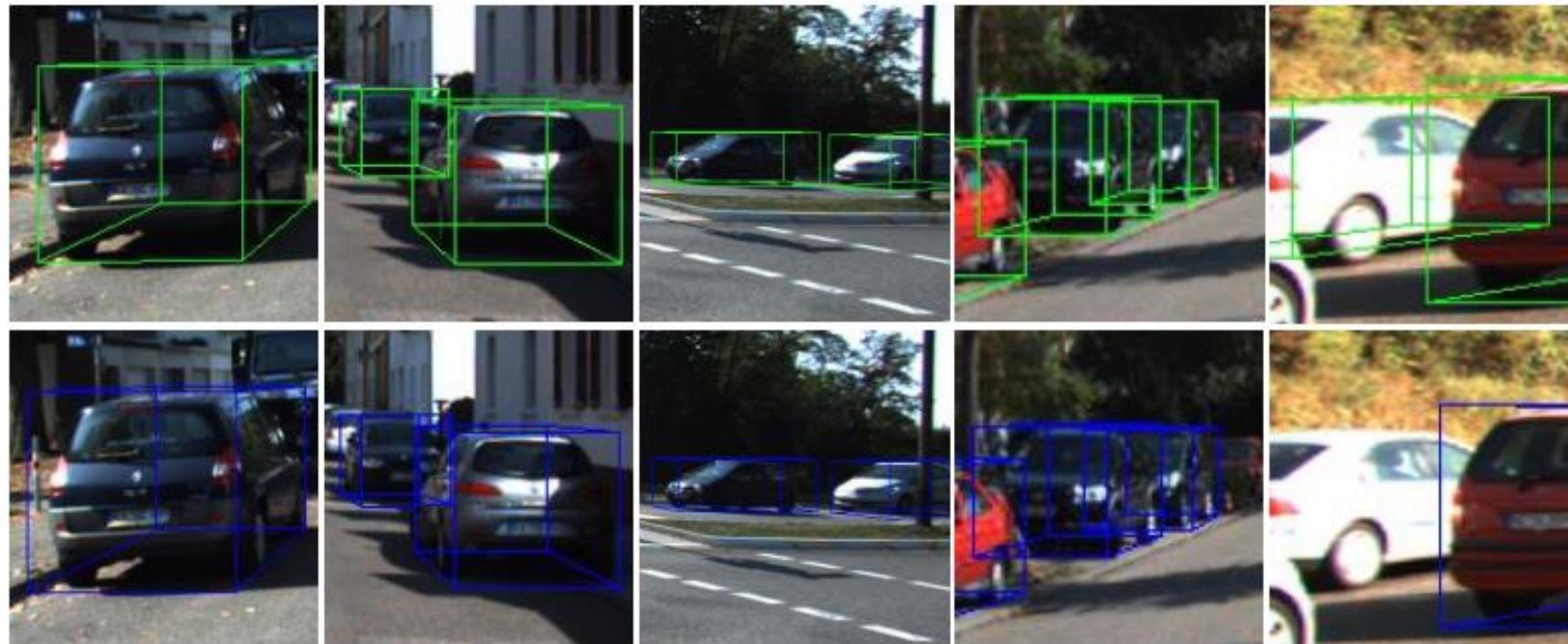


### 3.3 3D目标检测



北京交通大学

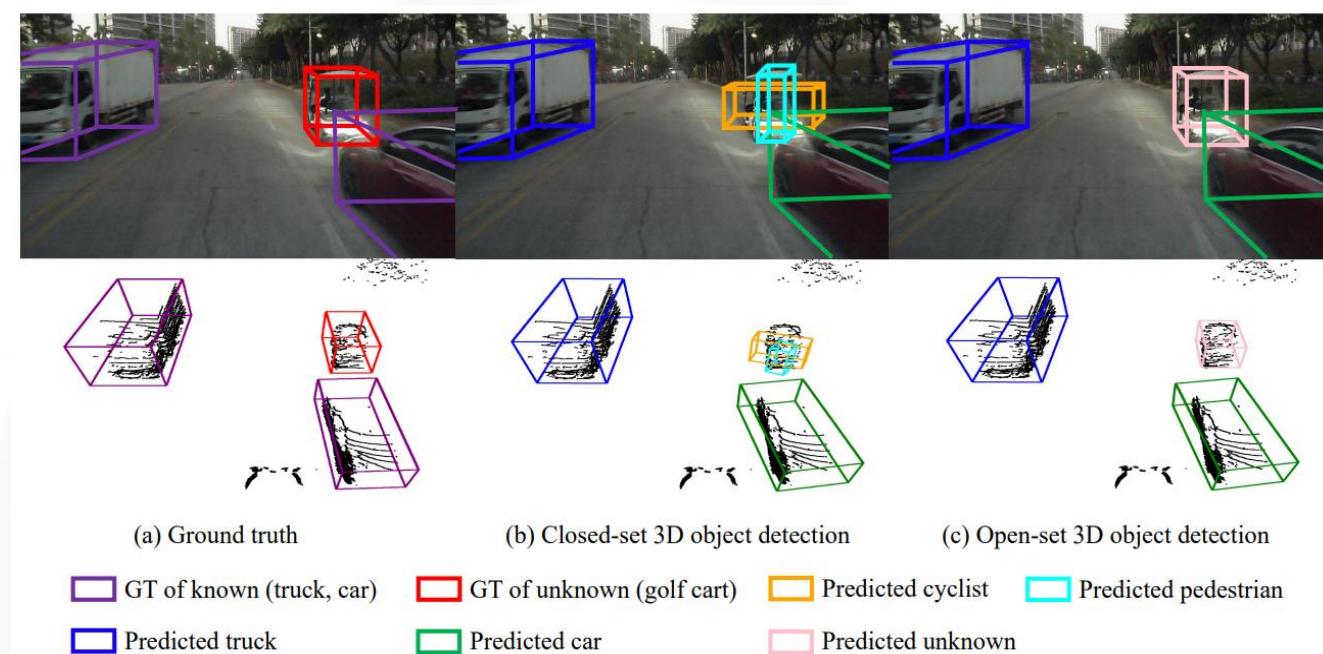
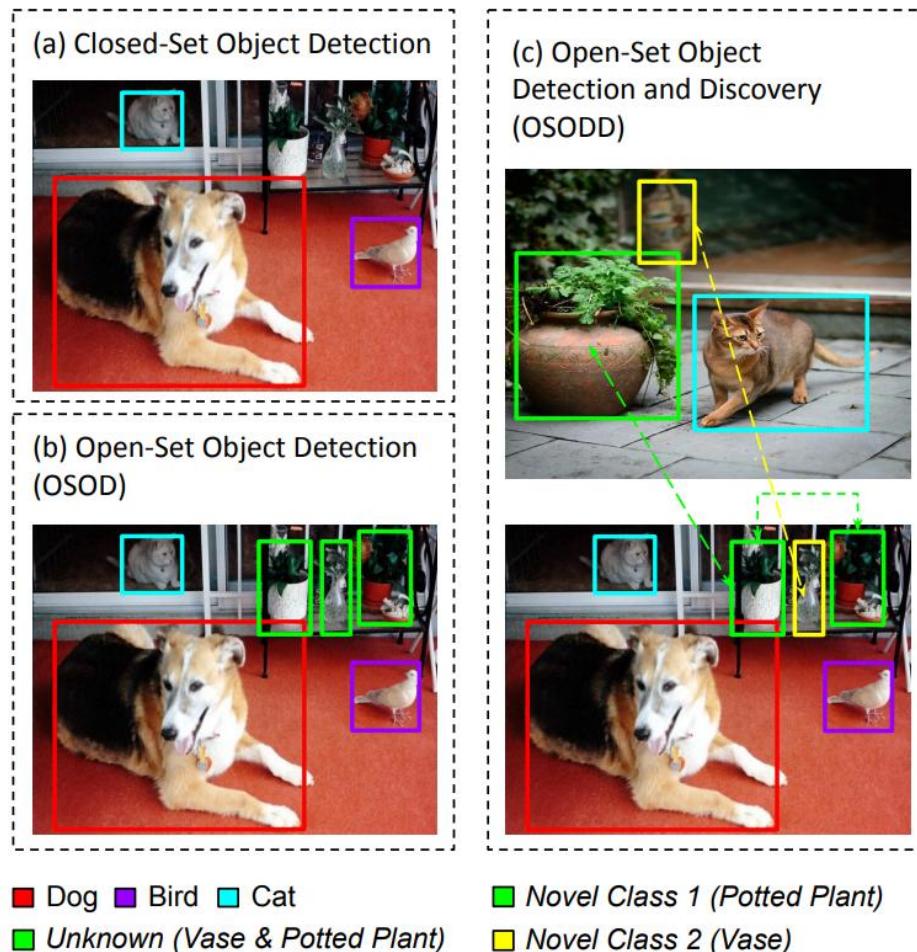
- ▲ **3D目标检测(3D Object Detection)**: 旨在预测3D空间中关键目标的位置、大小和类别。
- ▲ **与2D目标检测的区别**: 2D目标检测仅在图像上生成2D边界框，并忽略目标与观测点的实际距离信息，而3D目标检测更侧重于**对真实世界3D坐标系中目标的定位和识别**。
- ▲ **难点**: 物体遮挡、截断、周围环境的干扰；实时性欠佳等.....



## 3.4 开放集目标检测



- ▲ **开放集目标检测(Open-set Object Detection)**: 不仅要检测已知类别的目标，还要检测未知类别的目标，通过学习从已知类别到未知类别的映射关系，来实现对未知目标的检测。
- ▲ 难点：样本类别不平衡；未知类别边界判定；数据标注问题等.....



# 4. 研究数据

DOTA (<https://captain-whu.github.io/DOTA/index.html>) ,

遥感航空图像检测的常用数据集。包含了2806张航空图像，共包含15个类别，共计188282个实例。

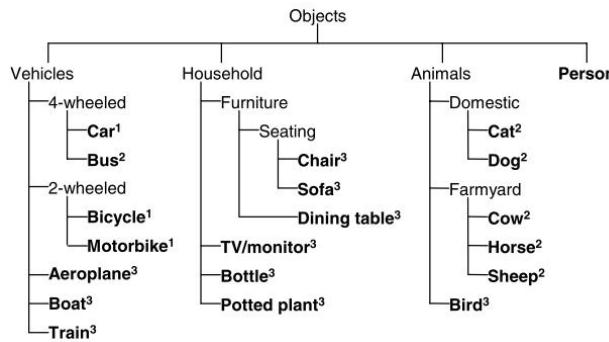


北京交通大学

▲ 相关经典数据集举例：航空图像区别于传统数据集：尺度变化性更大；密集的小目标检测；检测目标的不确定性。

(1) PASCAL VOC: (<http://host.robots.ox.ac.uk/pascal/VOC/>) ,

共有两个常用版本：VOC2007和VOC2012。2007版由5k张训练图像和12k个被标注的目标组成，2012版由11k张训练图像和27k个带注释的目标组成。从2007年开始，PASCAL VOC每年的数据集都是包含四个大类，总共20个小类。



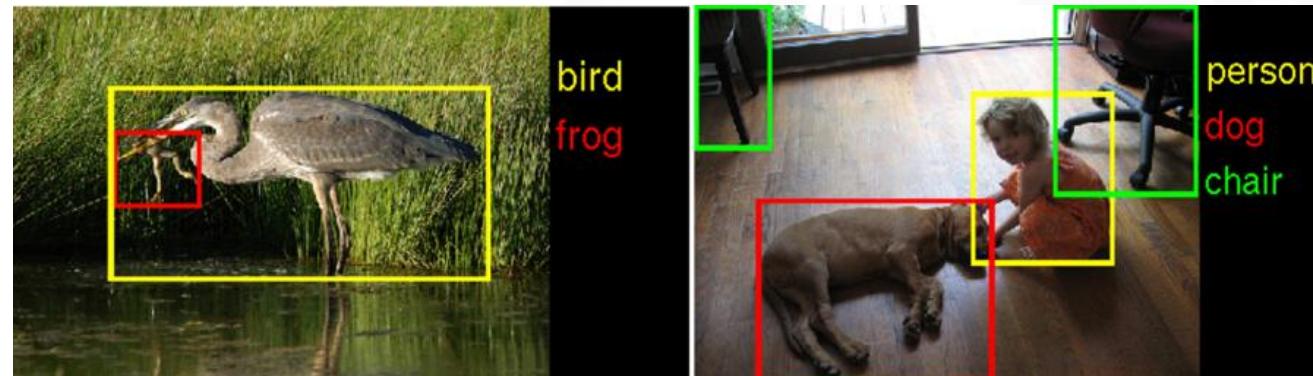
(2) ImageNet: (<https://image-net.org/>) ,

是深度学习领域中图像分类、检测、定位的最常用数据集之一。基于ImageNet的一个比赛名为ILSVRC，该词有时也用来特指该比赛使用的数据集，即ImageNet的一个子集，其中最常用的是ILSVRC2012，包含1000个分类，每个分类约有1000张图像。训练集图像总数约为120万张，验证集图像总数约为5万张，测试集图像约为10万张。

COD10K

目前规模最大的伪装目标数据集，共包含了10个超类和78个子类，共10000张伪装图像。

伪装目标类别包含有自然伪装中陆地，海洋，飞行，两栖生物。目标纬度包括大，中，小三个纬度。



## 5. 评价指标



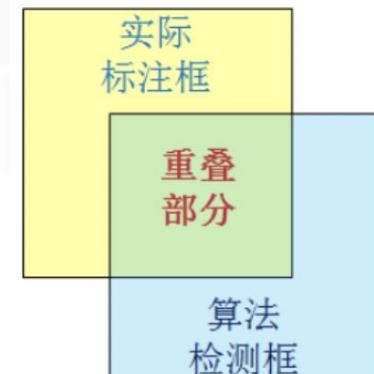
北京交通大学

### ▲ 常用的评价指标：(包括但不限于以下这些~)

- Accuracy准确率
- Precision精度
- Recall召回率
- AP(Average Precision)平均精度
- mAP(Mean Average Precision)平均精度均值
- IoU(Intersection over Union)交并比

- AP : 使用积分的方式表示为Precision-Recall曲线与坐标轴围成的面积。根据置信度对检测结果排序，逐步执行二分策略，计算每个类别的Precision-Recall曲线，根据曲线计算AP值。
- mAP: 由所有类别的AP值加权平均得到，AP值仅针对一个类别，而mAP用于衡量训练出来的模型在所有类别上的检测能力。
- IoU : 衡量模型预测的边界框与真实边界框之间的重叠程度。通过计算预测框与真实框的交集与并集之间的比例以衡量重叠程度。

$$mAP = \sum_{\text{class}_i}^n AP_i$$



# 7. 未来展望



北京交通大学

目标检测领域的研究前景与发展趋势非常广阔，以下内容仅个人观点，供参考批判~

智慧交通：自动驾驶；交通流量监测；交通安全检测；智能停车系统

➤ 提高精度与效率：目标检测将不断追求更高的检测精度和更高的运行效率。

可以通过引入更先进的网络结构、特征提取方法和优化算法，  
实现更准确、更快速的目标识别和定位。

智能安防：智能安检；入侵检测；视频监控预警；

➤ 小样本学习策略：目标检测将更好地适应小样本和少标注数据的情况。智能机器人：寻路避障；人机交互；  
通过引入元学习、迁移学习和增量学习等技术，能够在具有小样本情况下  
进行准确检测，实现对新类别和新场景的快速适应。

医学诊断：病灶检测；辅助手术；病理分析；

➤ 多模态数据融合：目标检测将更关注于多模态数据融合。航空遥感：城市规划；地图制作；环境监测；  
通过结合图像、语音、文本等多种数据信息，可以更全面理解和识别复杂  
环境中的目标，提高检测的准确性和鲁棒性。

工业检测：表观缺陷检测；产品组装；

➤ 跨域自适应学习：目标检测将更聚焦于跨域场景下的自适应学习。  
通过在不同领域、不同视角和不同感知模态之间进行知识迁移和特征共享，  
从而实现在新场景中的快速适应和准确检测。

➤ 改善检测实时性：实时性是目标检测中的重要问题，在许多应用中具有重要的价值和需求，  
如自动驾驶、视频监控、机器人导航等，对实时性有极高的要求。  
目标检测将更关注模型轻量化、分布式计算、数据预处理与优化等方面。