



中国科学院大学
University of Chinese Academy of Sciences

深度学习应用（计算机视觉）



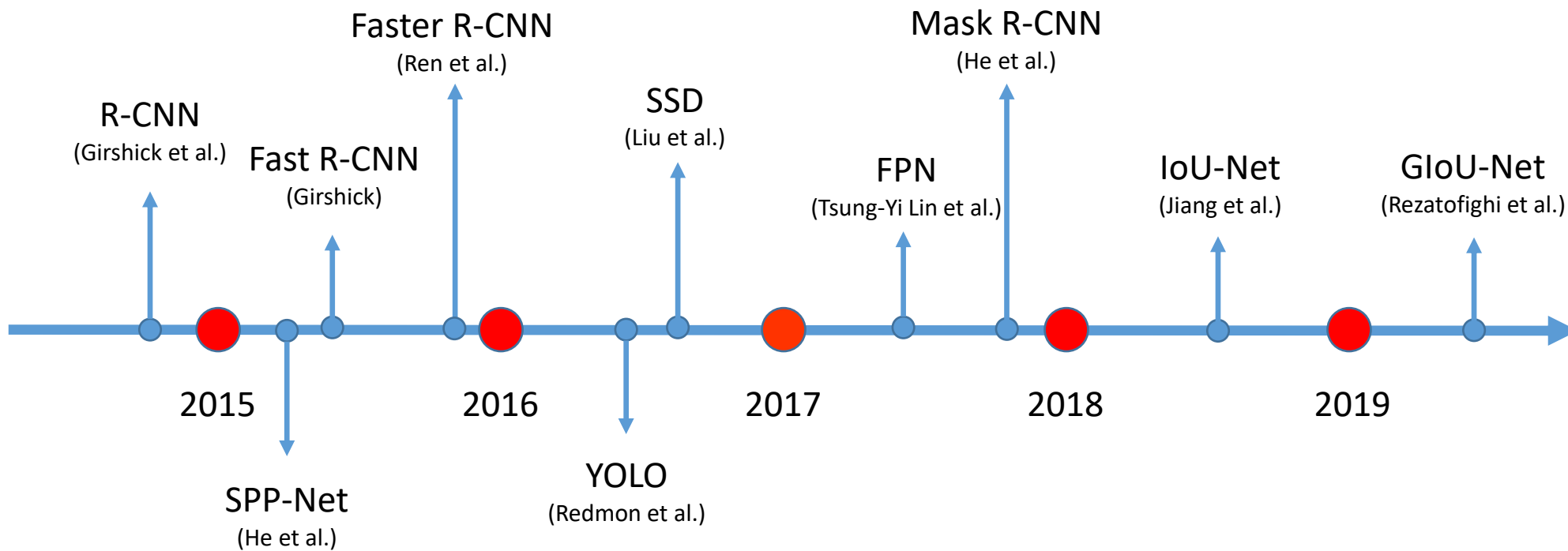


目标检测

AI DISCOVERY



目标检测发展历程



AI DISCOVERY



目标检测



AI DISCOVERY

两阶段方法

R-CNN、SPP-Net、Fast R-CNN、
Faster-RCNN、FPN、Mask-RCNN

一阶段方法

YOLO、SSD

最新进展

IoU-Net、GloU



AI DISCOVERY

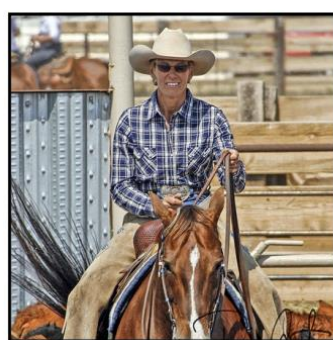


两阶段方法

AI DISCOVERY

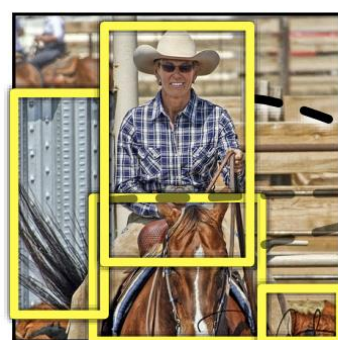
1、R-CNN

R-CNN: *Regions with CNN features*



1. Input image

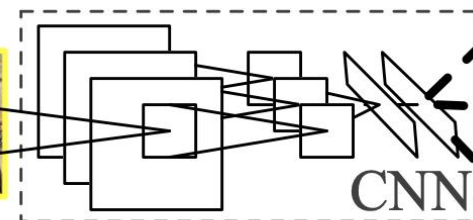
输入图像



2. Extract region proposals (~2k)

提取候选检测框
(约2000个)

warped region



CNN

3. Compute CNN features

为每个候选检测框提取CNN特征

aeroplane? no.

⋮

person? yes.

⋮

tvmonitor? no.

4. Classify regions

为每个候选检测框进行分类

AI DISCOVERY



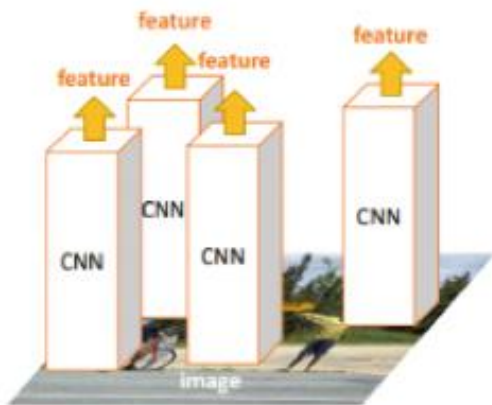
两阶段方法

AI DISCOVERY



2、SPP-Net

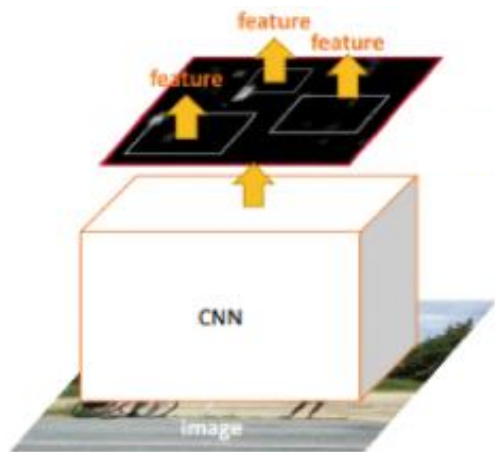
R-CNN缺点：为每个region proposal提取特征，花费大量的计算时间和存储空间。



R-CNN

2000 nets on image regions

逐个提取特征



SPP-NET

1 net on full image

一次性提取特征

SPP-Net不再单独提取每个候选区域的特征，而是一次性提取整个图像的特征，再在特征图上取出对应于不同region proposal的区域



减少了提取特征的时间，和用来存储特征的空间。

AI DISCOVERY

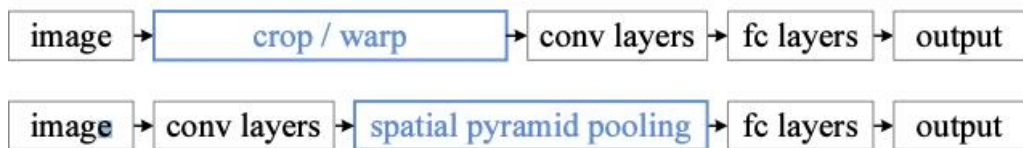


两阶段方法

AI DISCOVERY

2、SPP-Net

R-CNN缺点：使用Warp，为每个region proposal统一尺寸，严重影响CNN提取特征的质量。

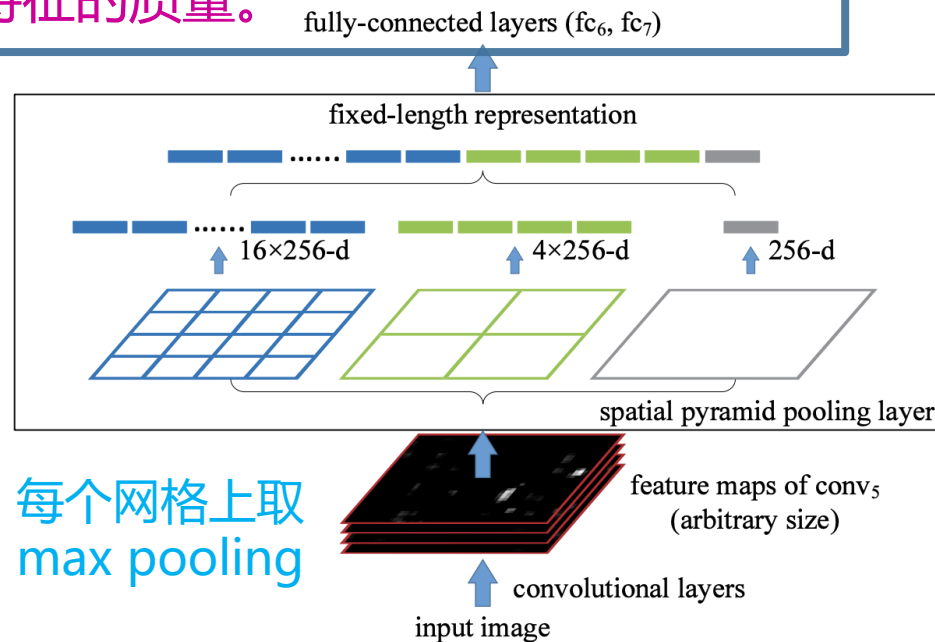


丢失信息/形态

通过spatial pyramid pooling将任意大小RoI特征统一成相同的尺寸



可以输入任意大小的候选区域，不再需要warp输入图像，提升了CNN提取的特征的质量，使特征更鲁棒



每个网格上取
max pooling

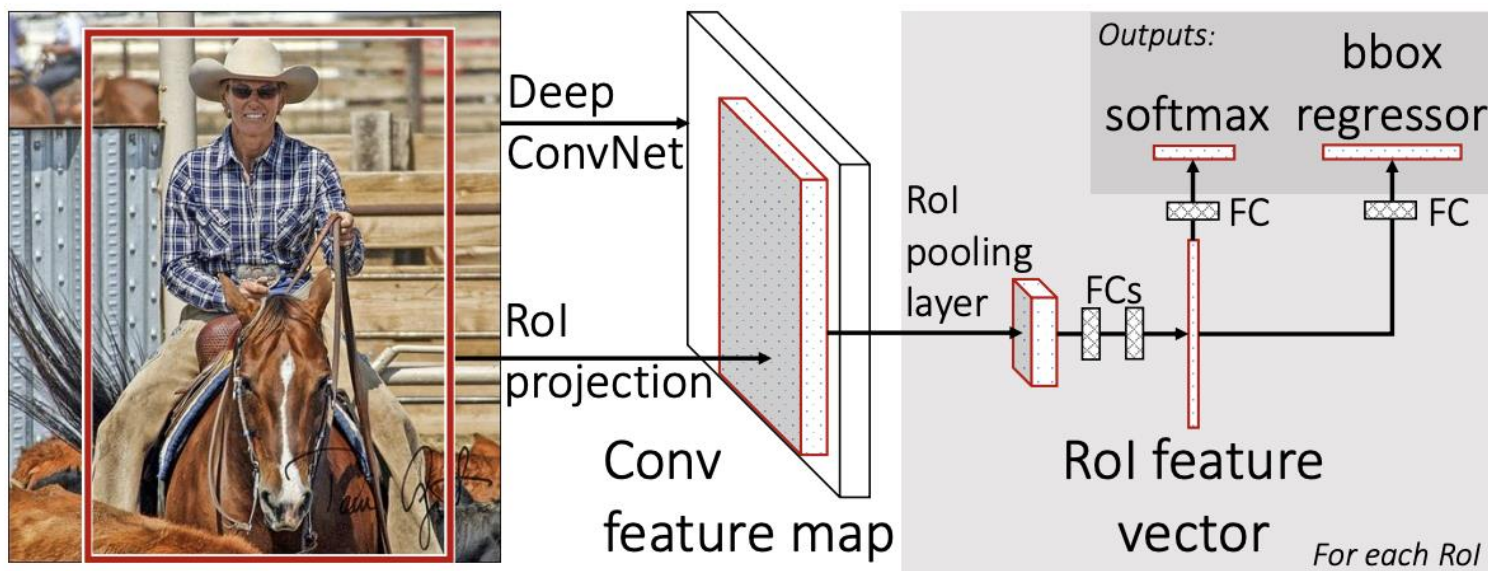
AI DISCOVERY



两阶段方法

AI DISCOVERY

3、Fast R-CNN：将分类损失和回归损失统一在同一个框架中



✓ 图像的warp->特征的warp

✓ 损失: svm+regressor->多任务损失
(softmax+regressor)

通过SS在图像中提取RoI->卷积网络提取特征->RoI Pooling->全连接层->分类/边界框回归

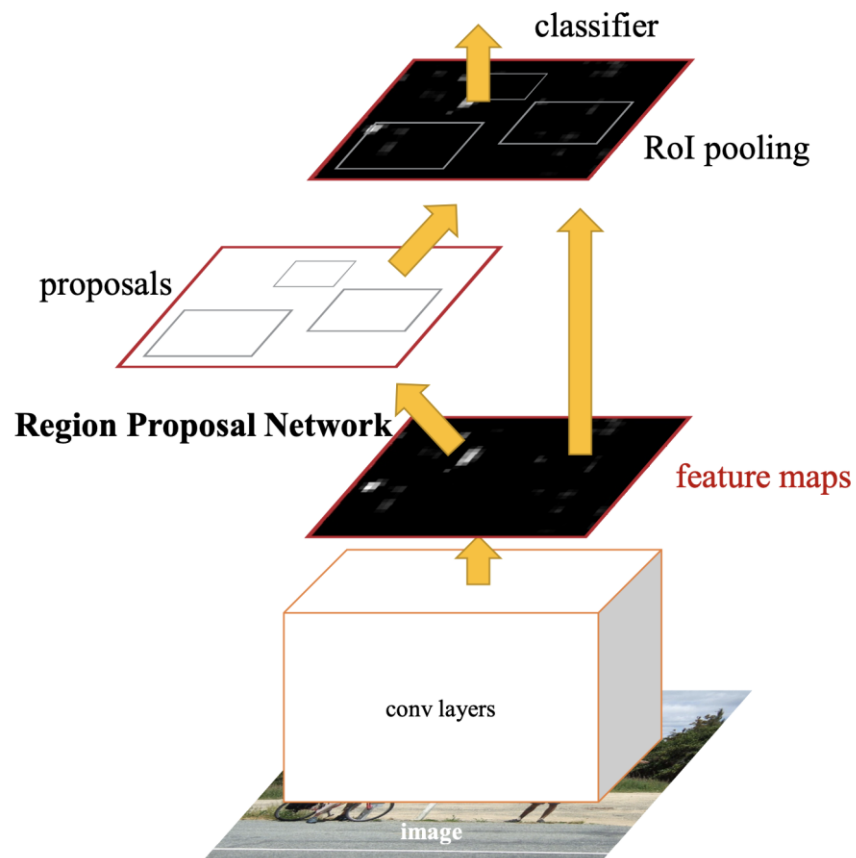
AI DISCOVERY



两阶段方法

AI DISCOVERY

4、Faster R-CNN：端到端检测网络，极大提升了检测速度



◆ 如何实现：

- ✓ 在最后一个卷积层后面添加Region Proposal Network (RPN)
- ✓ 候选区域由RPN网络直接生成，不再依靠额外的候选区域生成算法（如SS）
- ✓ 由RPN网络提取候选区域后，类似Fast RCNN，使用RoI Pooling，再使用两个分支分别计算类别和边界框回归

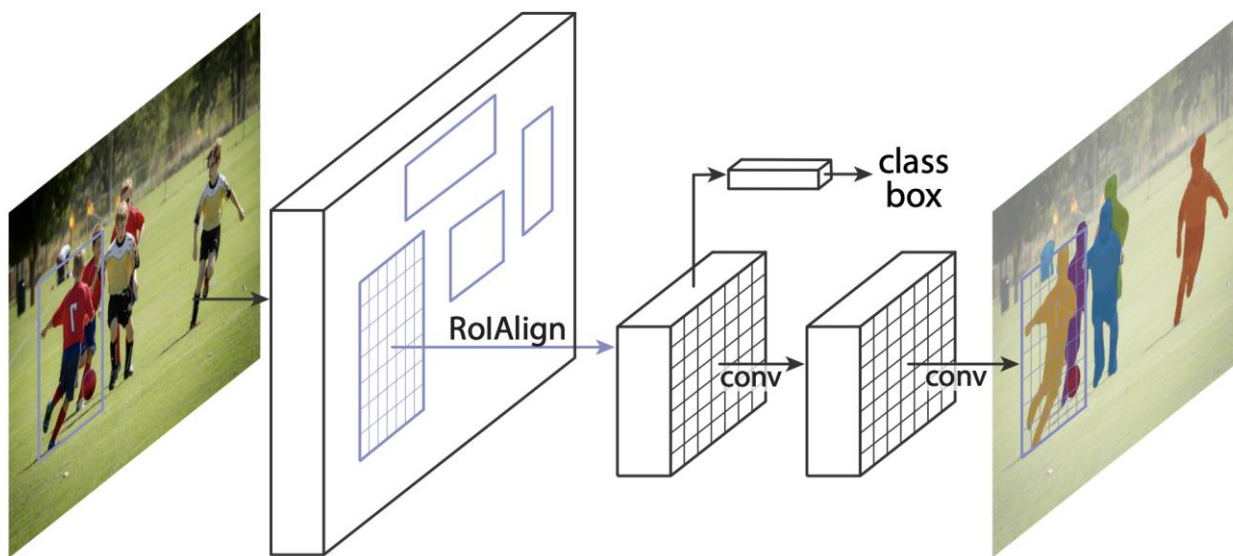
AI DISCOVERY



两阶段方法

AI DISCOVERY

6、Mask-RCNN



◆ 主要贡献:

- ✓ 强化的基础网络：通过**ResNet-101+FPN**用作特征提取网络，达到state-of-the-art的效果。
- ✓ 采用**RoI Align**，解决**misalignment**
- ✓ 处理分割任务的分支使用**全卷积网络**

AI DISCOVERY



目标检测



AI DISCOVERY

两阶段方法

R-CNN、SPP-Net、Fast R-CNN、Faster-RCNN、FPN、Mask-RCNN

一阶段方法

YOLO、SSD

最新进展

IoU-Net、GloU



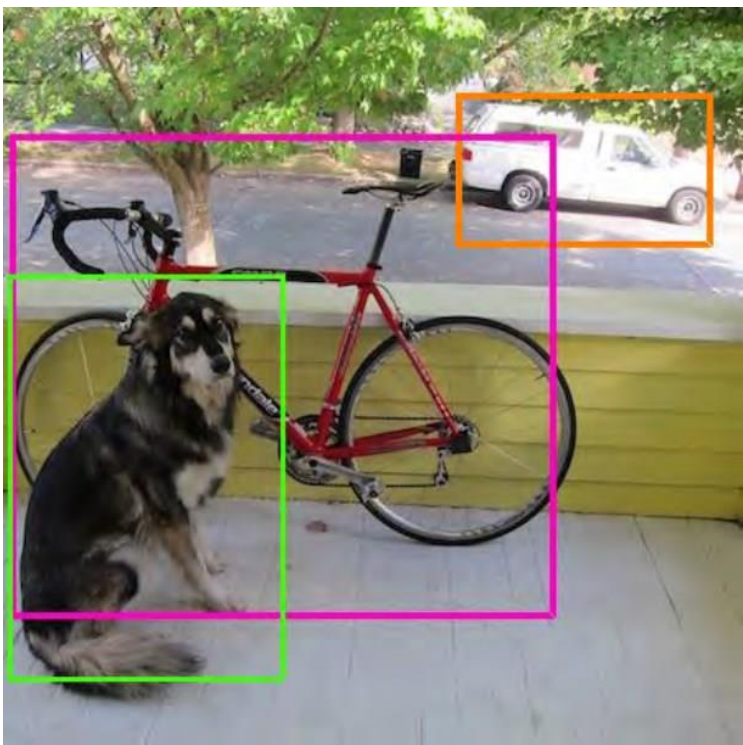
AI DISCOVERY



一阶段方法

You Only Look Once:
Unified, Real-Time Object Detection

1、YOLO: 实时检测



- ✓ YOLO的**检测速度非常快**。在Titan X上，不需要经过批处理，标准版本的YOLO系统可以每秒处理45张图像；YOLO的极速版本可以处理150帧图像。
- ✓ YOLO在做预测的时候，使用的是全局图像。与sliding window和region proposal这类方法不同，**YOLO一次“看”一整张图像**，所以它可以将物体的整体（contextual）的类别信息（class information）以及外观信息（appearance information）进行编码，**背景错误分类为目标概率低**。
- ✓ YOLO学到物体更泛化的特征表示。当在自然场景图像上训练YOLO，再在艺术风格化图像的数据集上去测试YOLO时，YOLO的表现相比其他网络要好。**YOLO模型能更好的适应新的领域**。

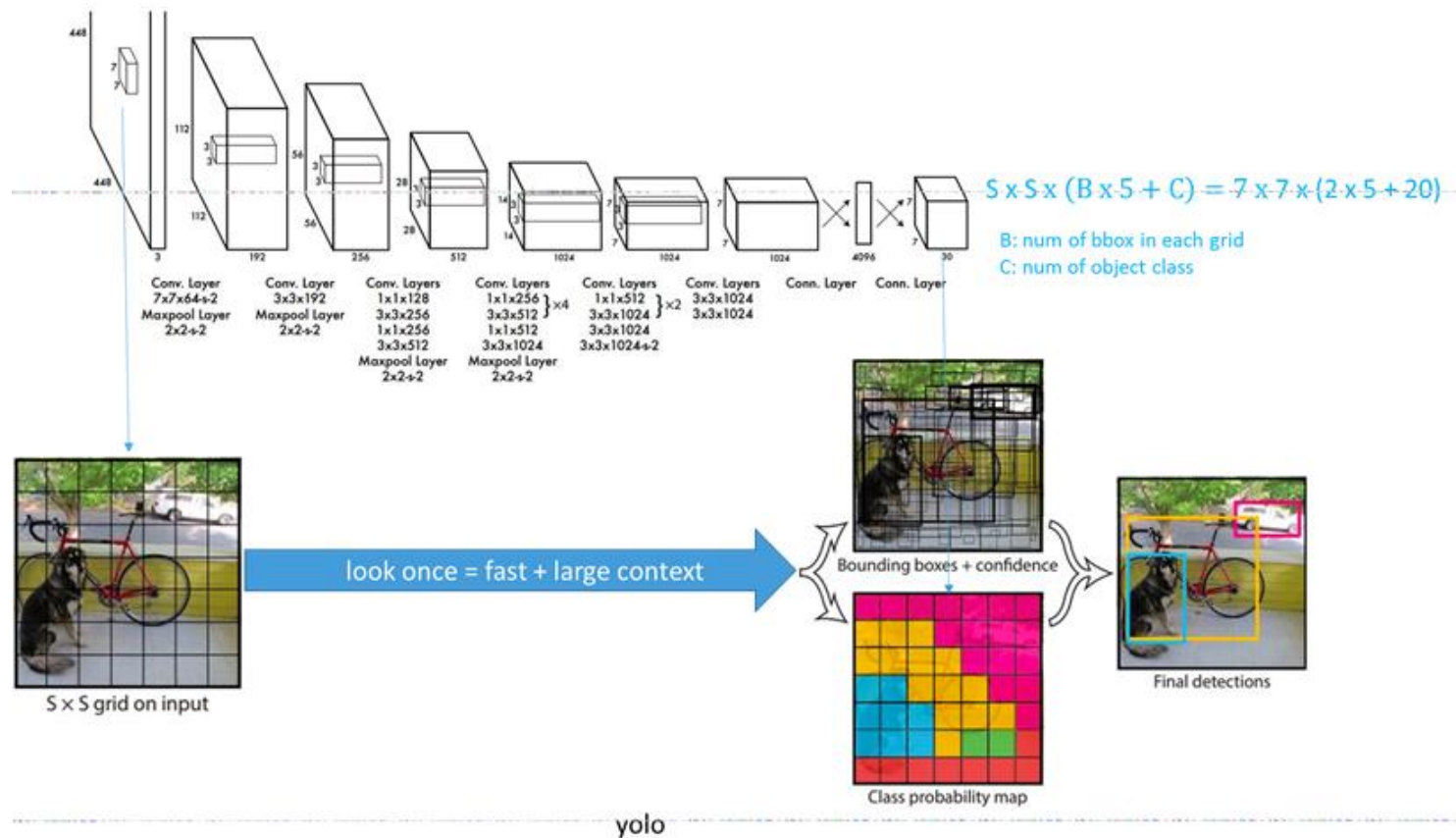


一阶段方法

1、YOLO：框架

◆ 算法流程

- ✓ 将输入的图像划分成 $S \times S$ 个网格
($S=7$)
- ✓ 每个网格预测 B 个边界框和这个边界框是物体的概率
(Objectness);具体的, 每个边界框会预测出5个值: x, y, w, h 和置信度 $\text{Pr}(\text{Object}) * \text{IoU}(\text{truth} \& \text{pred})$
- ✓ 每个网格预测 C 个类的概率



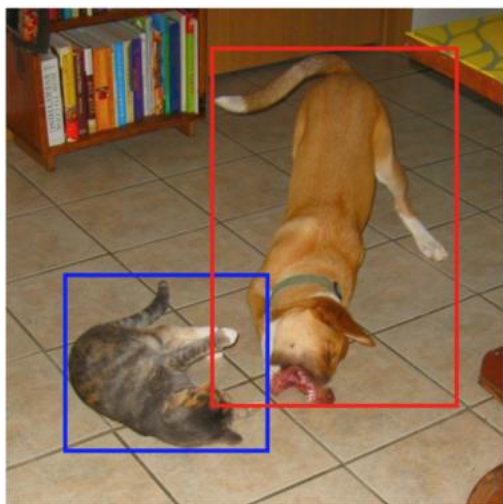
训练时, 有物体时 $\text{Pr}(\text{Object})=1$, 否则 $\text{Pr}(\text{Object})=0$



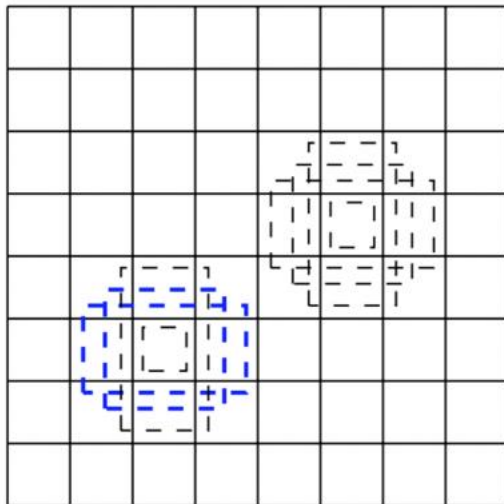
一阶段方法

AI DISCOVERY

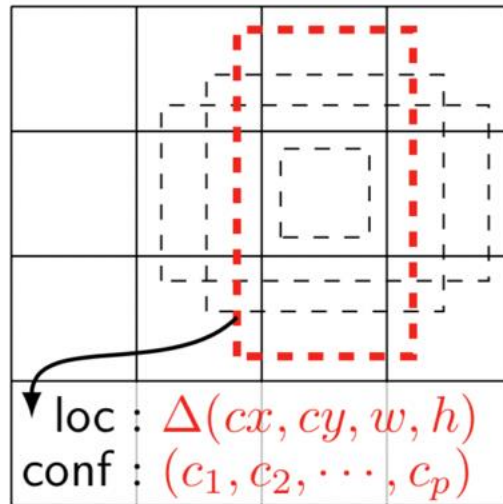
1、YOLO default box: 速度快, 精度高



(a) Image with GT boxes



(b) 8×8 feature map



(c) 4×4 feature map

- ✓ 对于一张特征图, 在每一个位置上提取预设数量的default box。
- ✓ 直接在特征图上密集的提取proposal进行预测, 使网络不需要先提取候选目标区域, 速度大幅度提升。

AI DISCOVERY

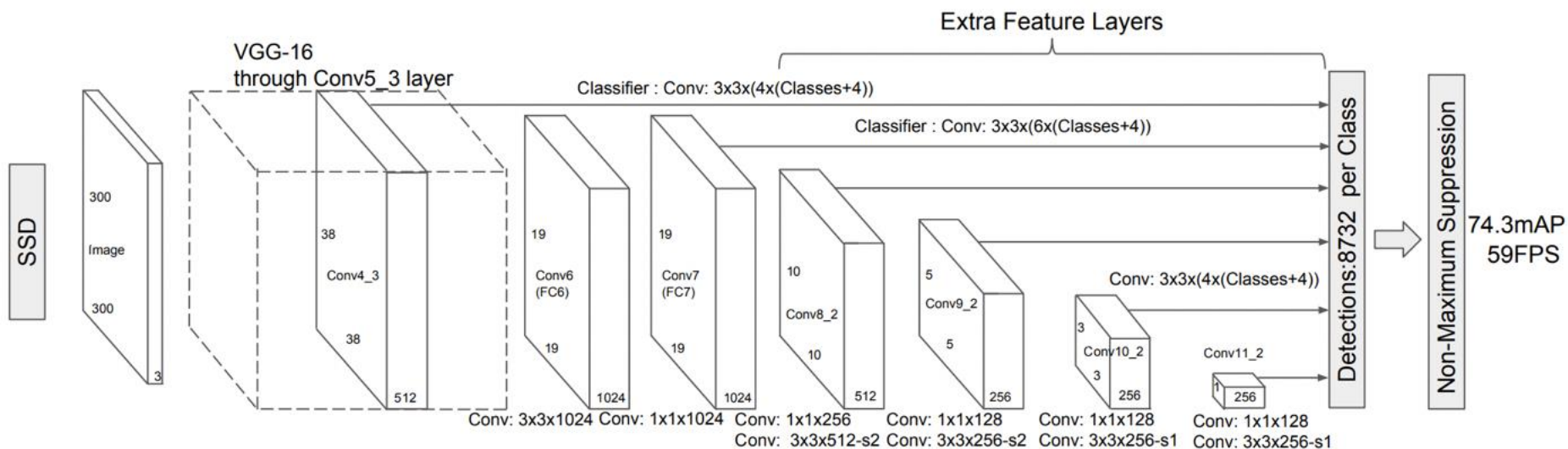


一阶段方法



AI DISCOVERY

2、SSD 网络结构：速度快，精度高



✓ 在不同尺度的特征图，直接提取预设数目default box，进行预测

✓ 提高检测精度（尤其在小目标有提升）



AI DISCOVERY

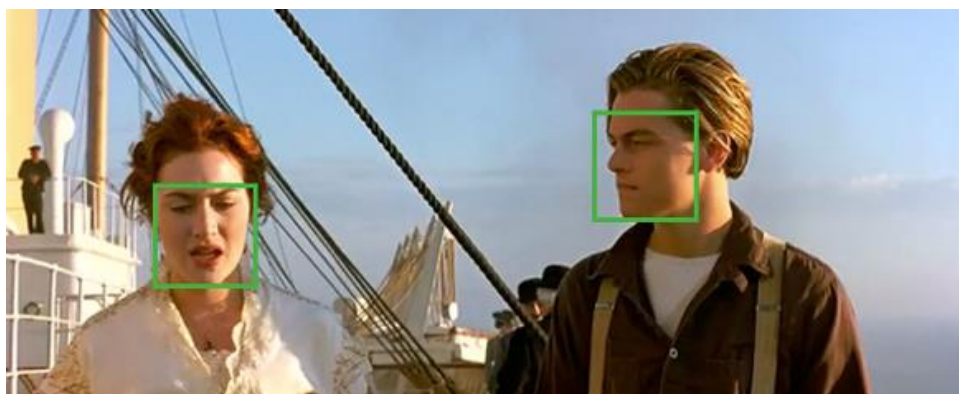
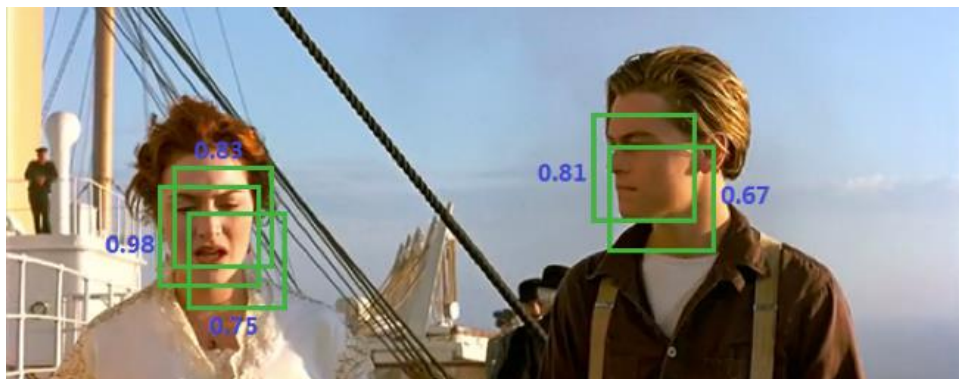


最新进展



AI DISCOVERY

1、IoU-Net



◆ 什么是NMS（非极大值抑制）

- ✓ 将**同一类**的所有的检测框，按照**分类置信度**排序
- ✓ 将与**分类置信度最高**的检测框的**重叠面积 (IoU)** 大于一定阈值的检测框删除
- ✓ 从未处理的框中，再选出一个分类置信度最高的检测框，重复上一步操作

通过非极大值抑制，消除冗余的检测框



AI DISCOVERY

