




目标检测与相关算法



第三组

2024-03-07



CONTENTS

- 目标检测简介
- 重要算法：RCNN、YOLO
- 算法评估指标
- 难点与挑战
- 总结



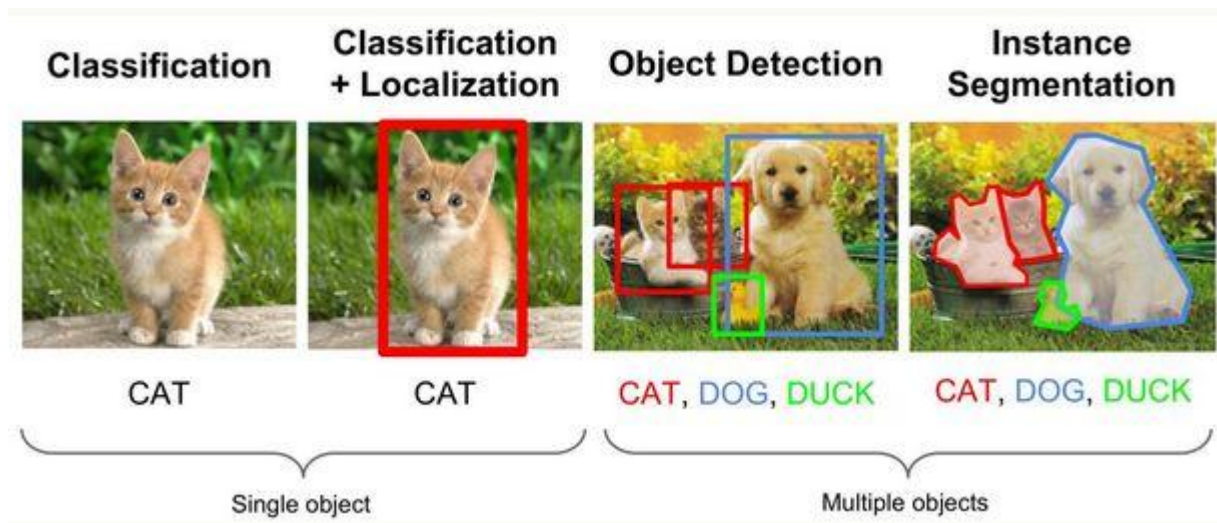
目标检测简介

目标检测问题简介

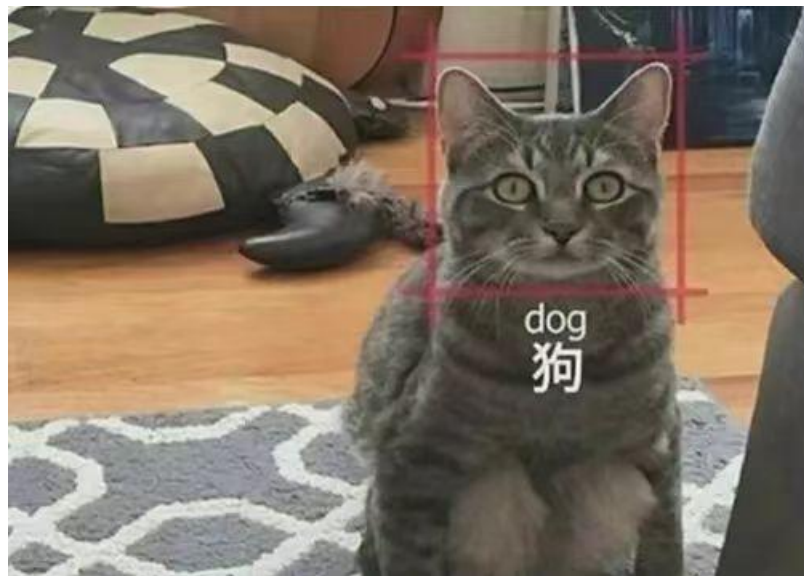
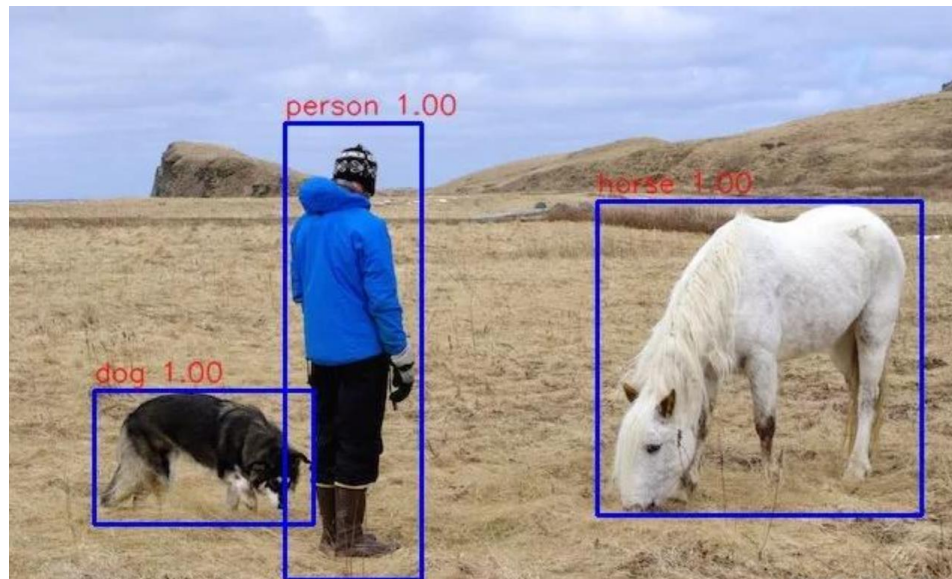


目标检测是什么？

目标检测(Object Detection)是计算机视觉(Computer Vision, CV)领域的一项热门关键任务，旨在从图像或视频中定位和识别特定目标。目标检测在许多领域中具有广泛的应用，包括自动驾驶、视频监控、人脸识别等。



目标检测问题简介



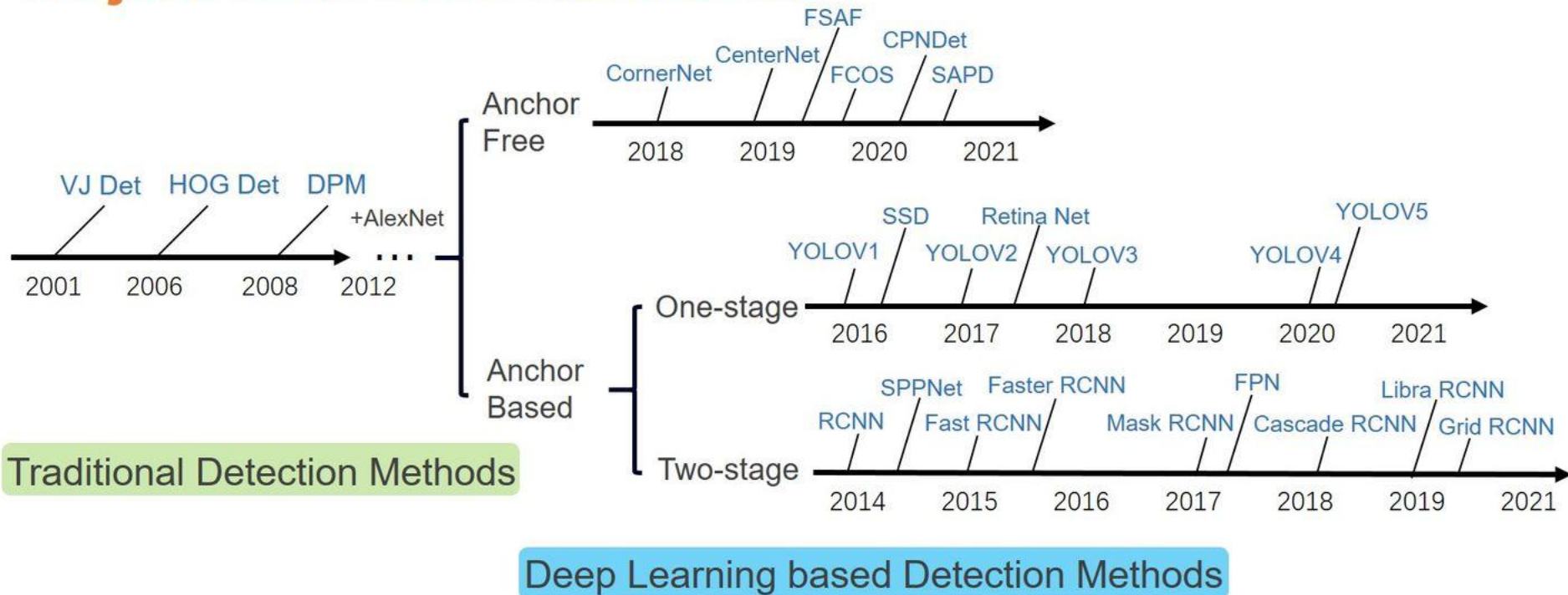


重要算法

目标检测算法发展历程



Object Detection Milestones



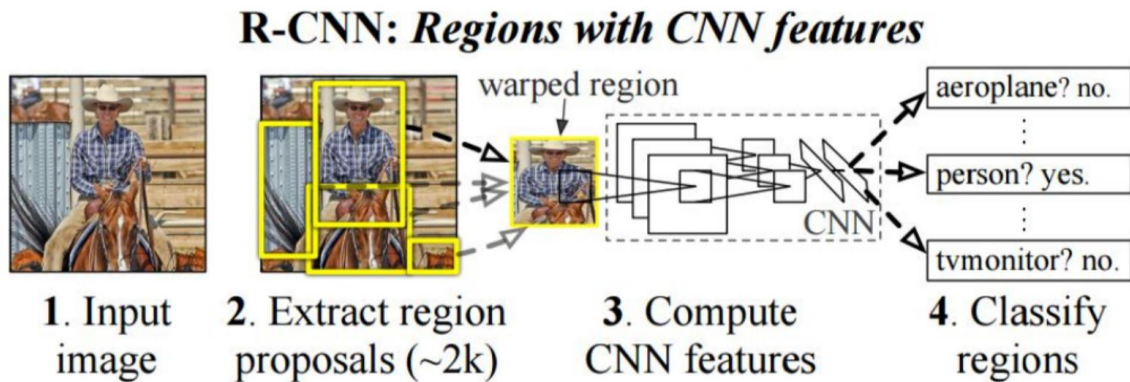
RCNN系列算法：RCNN

二阶段(Two-Stage)检测算法主要分为以下两个阶段

Stage1: 从图像中生成候选区域(region proposals)

Stage2: 从候选区域生成最终的物体边框

RCNN于2014年提出，它首先通过选择性搜索算法(Selective Search)生成可能的对象候选框，然后将候选框中的图像缩放变形到某一固定尺寸的图像，并喂入到预训练的CNN模型中(如上页提到的AlexNet)提取特征，最后将提取出的特征送入到SVM分类器来预测该对象框中的图像是否存在待检测目标，从而进一步预测该检测目标具体属于哪一类。

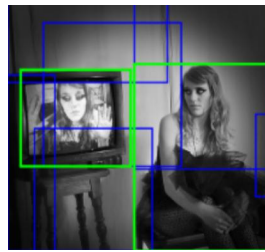


RCNN系列算法：RCNN



什么是选择性搜索算法？

根据一些预定的规则，对图片中的每一个像素点进行聚类：相邻的相同颜色、相同纹理的像素是一个类别。像素点聚集起来后形成块状斑，斑与斑之间的差异较大，斑内的像素之间的差异小。斑的外接矩形就是该算法生成的候选框。由此生成的候选框极有可能是我们需要检测的目标。

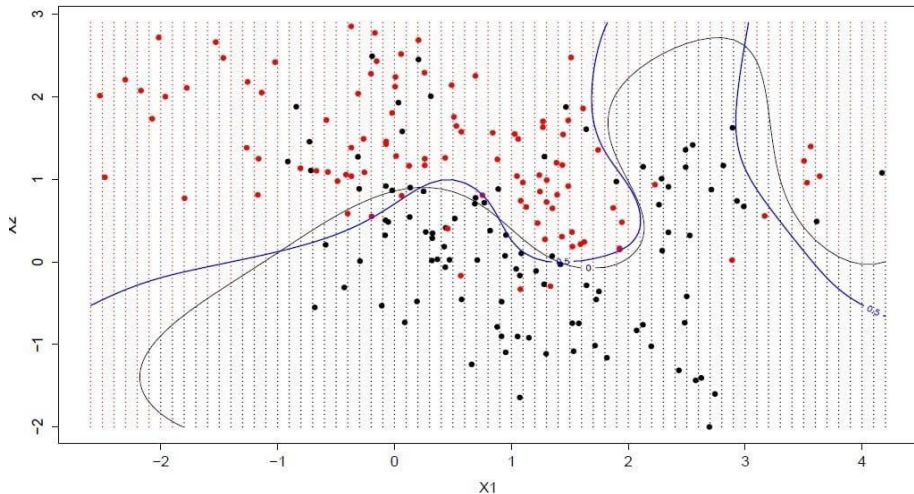


什么是NN？什么是CNN？

课本第八章。卷积神经网络（Convolutional Neural Networks）。

什么是SVM分类器？

支持向量机（support vector machines, SVM）是用于求解能够正确划分训练数据集并且几何间隔最大的分离超平面以解决二分类问题的算法。

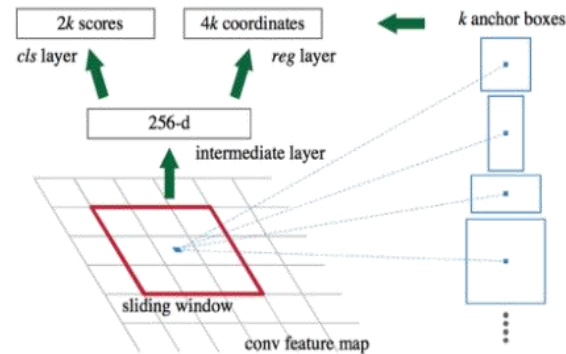
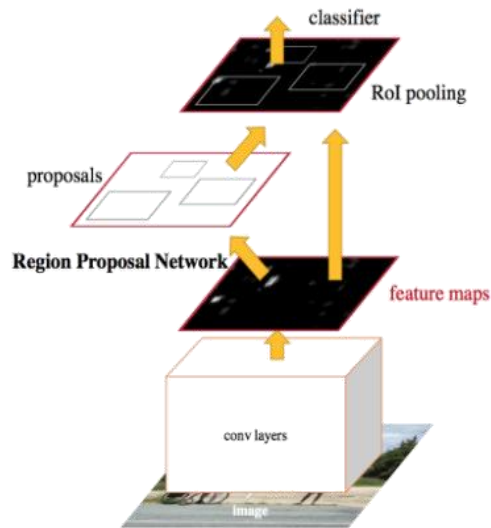


RCNN系列算法：Faster RCNN

RCNN于2014年提出，它首先通过选择性搜索算法(Selective Search)生成可能的对象候选框，然后将候选框中的图像缩放变形到某一固定尺寸的图像，并喂入到预训练的CNN模型中(如 AlexNet)提取特征，最后将提取出的特征送入到 SVM分类器来预测该对象框中的图像是否存在待检测目标，从而进一步预测该检测目标具体属于哪一类。

改进：可以使用CNN模型来直接生成候选框吗？

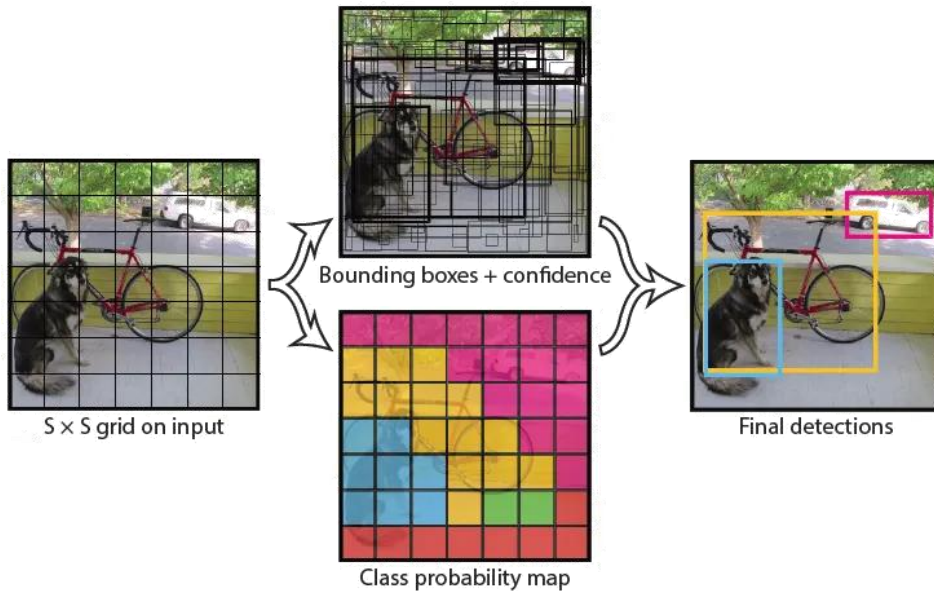
用RPN神经网络代替RCNN中的Selective Search是Faster RCNN的核心内容。由于RPN与RCNN权值共享，目标检测的速度被大幅度提升。



YOLO系列算法

一阶段(One-Stage)目标检测算法不需要region proposal阶段，直接产生物体的类别概率和位置坐标值，经过一个阶段即可直接得到最终的检测结果，因此有着更快的检测速度。

YOLO(You Only Look Once)算法提出于2016年，其名得自只使用一个神经网络的核心思想。算法首先将一幅图像分成 $S \times S$ 个网格，如果某个目标的中心落在这个网格中，则这个网格就负责预测这个目标。对于每个网格，都用CNN神经网络预测出B个边框，这些边框用 $(x,y,w,h,confidence)$ 的五元组表示，并且计算该网格属于各类别的条件概率。最后将上一步得到的 $S \times S \times B$ 个边框用非极大值抑制算法(Non-Maximum Suppression, NMS)将重复冗余的边框合并或删除。

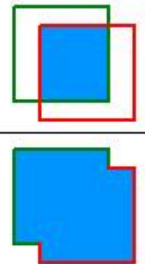




算法评价指标

算法评价指标：IoU、准确率、召回率

交并比(IoU)是度量两个检测框（程序结果框与真实结果框(基准真相, Ground Truth, GT)）的交叠程度的指标，当其大于一定阈值时可以认为检测到了目标。

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{\text{area of intersection}}{\text{area of union}}$$


TP: IoU > 阈值的检测框数量（同一GT只计算一次）

FP: IoU ≤ 阈值的检测框数量 + 同一个GT的多余检测框的数量

FN: 没有检测到的GT的数量

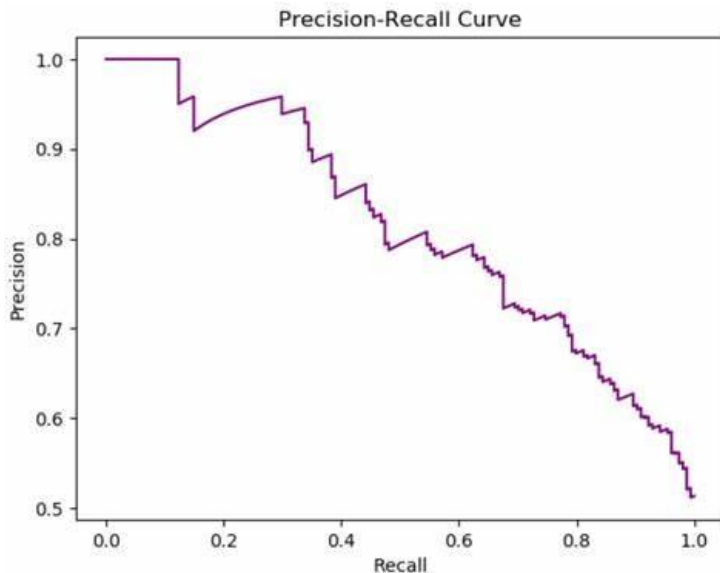
$$\text{准确率 Precision} = \frac{TP}{TP + FP} = \frac{TP}{\text{all detections}}$$

$$\text{召回率 Recall} = \frac{TP}{TP + FN} = \frac{TP}{\text{all ground truths}}$$

算法评价指标：AP、mAP、FPS

AP(Average Precision) 用于评估算法在图像中识别和定位特定类别对象的能力。对某一待检测的类别，根据不同的阈值，计算其准确率与召回率，可以画出一条PR曲线(Precision-Recall Curve)。对PR曲线进行插值平滑处理后，计算该曲线下的面积大小，即为AP。

mAP(mean Average Precision): 计算在所有待检测类别下的AP，然后求其平均值。



算法评价指标



FPS(Frame Per Second)指的是模型一秒钟能检测图片的数量。

不同的检测模型往往会有不同的mAP和检测速度。

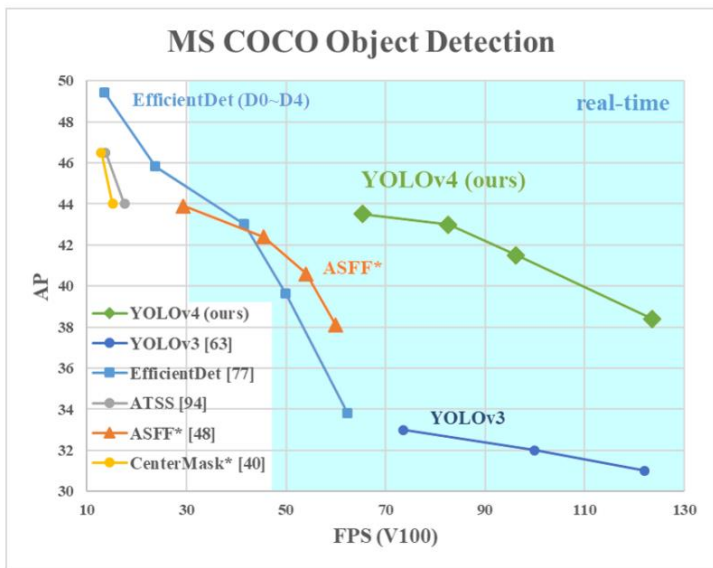


表 不同目标检测算法在 COCO 数据集上的测试结果

Method	Backbone	Input size	FPS	mAP(%)
Faster R-CNN ^[33]	ResNet-101	-	-	59.1
RepPoints ^[35]	ResNet-101-DCN	-	-	65
RetinaNet ^[44]	ResNet-101	800*800	5.1	57.5
SSD ^[42]	VGG-16	512*512	22	48.5
DSSD ^[43]	ResNet-101	-	-	53.3
CenterNet ^[46]	DLA-34	384*384	28	60.3
YOLOv3 ^[40]	Darknet53	416*416	35	55.3
YOLOv4 ^[41]	CSPDarknet53	608*608	33	65.7
YOLOv5	CSPDarknet53	832*832	40	69.6



难点与挑战

难点与挑战：人脸检测



1.人脸姿态变化大

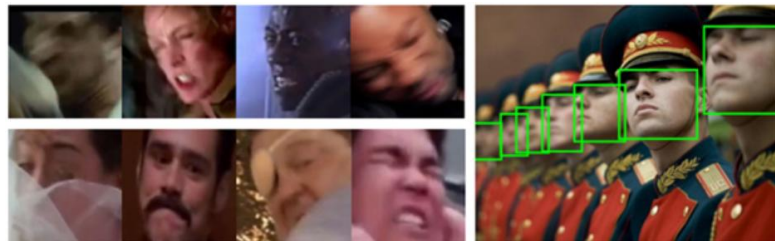
人类脸部可能出现各种变化，如表情，肤色，姿势和运动等变化。

2.脸部遮挡问题

人脸可能被其它物体所遮挡。

3.多尺度检测问题

在检测大尺寸人脸的时候，也经常要检测小尺寸的人脸，因此存在多尺度检测挑战。



(a)

(b)



(c)

难点与挑战：行人检测



1. 小尺寸行人

距离摄像头较远的行人目标往往尺寸较小，其高度可能小于30个像素点，难以检测。

2. 困难负样本

由于场景图像中一些背景在视觉上与行人非常相似，这导致网络很难将负样本进行正确分类，往往会将一些很像行人的背景物体预测成为行人。

3. 密集且遮挡情况

由于图像中多存在行人密集且行人遮挡情况，这导致模型难以精准检测出被遮挡的行人。



(a)



(b)



(c)



总结

总结



1.目标检测简介

从图像中定位和识别特定目标

2.重要算法

2.1 RCNN系列

二阶段算法：选择性搜索/RPN→CNN特征提取→二分类器

2.2 YOLO系列

一阶段算法：网格分割→边框预测→概率计算→NMS

3.算法评估指标

IoU、准确率、召回率、PR曲线、mAP、FPS

4.难点与挑战

两个例子

如何评价人工智能？



A decorative pattern of interlocking 3D cubes in a light teal and a golden-yellow color, arranged in a staggered grid. The cubes are rendered with perspective, giving them a three-dimensional appearance. They occupy the left half of the image.

THANKS