Chapter 3-1

目标检测导论

An introduction to Object Detection

主讲人 安泓郡

ACoTAI Lab, Dalian Maritime University
ICDC department, Dianhang Association

an.hongjun@foxmail.com

Chapter 3-1

目录

CONTENTS

第一部分

BASIC TASK

从图像分类到目标检测

回顾：图像分类任务

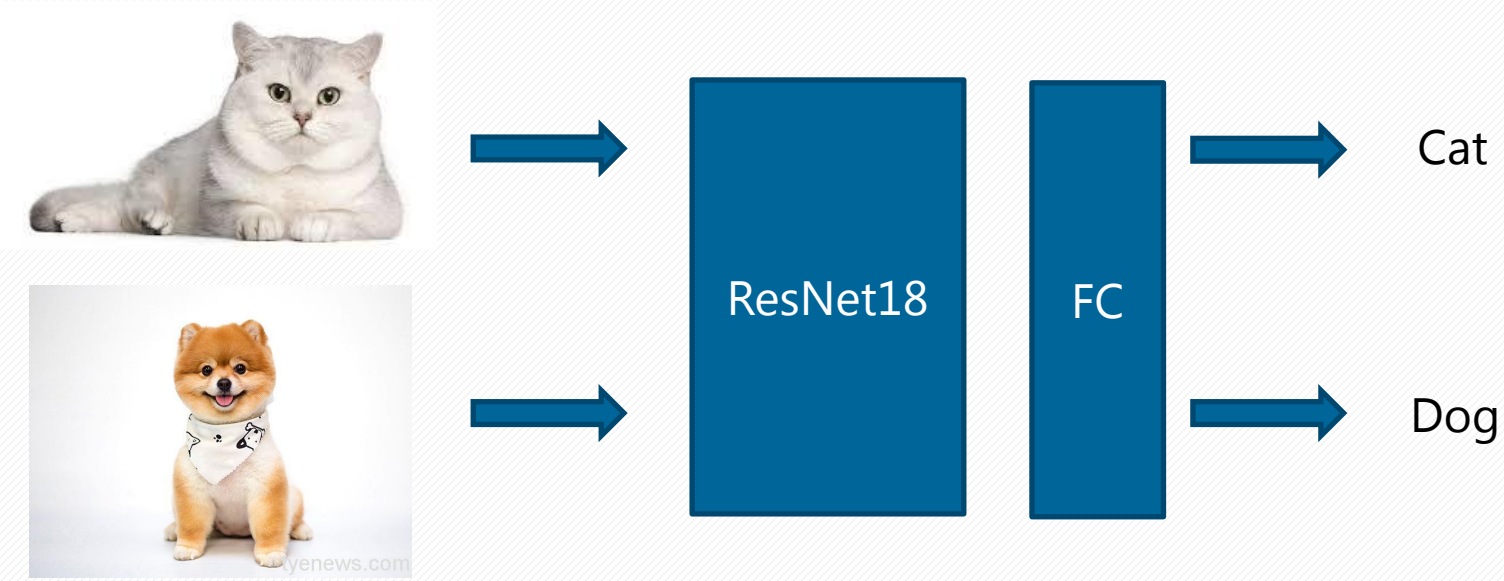
特征提取器

分类器

Cat

Dog

回顾：图像分类任务



回顾：图像分类任务

BackBone

Head

Cat

Dog

ResNet,VGG,
GoogLeNet,.etc

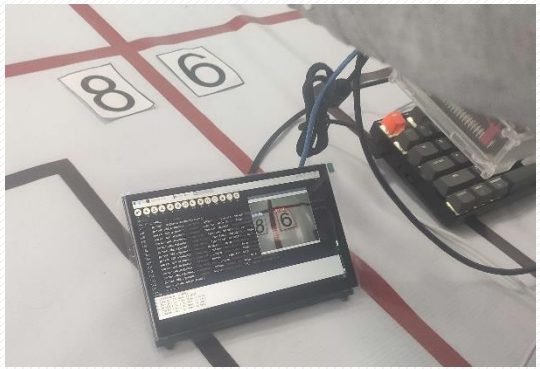
FC(MLP),
SVM

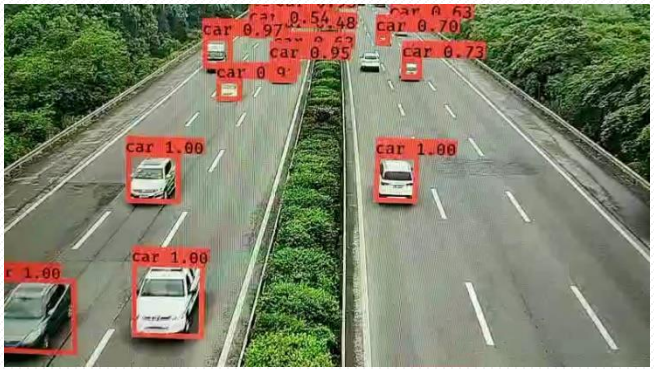
目标检测

图像分类：知道图片里有什么

目标检测：知道图片里有什么，并且知道在哪

应用场景





现代目标检测方法

- 以R-CNN系列为代表的双阶段目标检测算法
  - RCNN (2014, Ross Girshick et al.)
  - Fast-RCNN (2015, Ross Girshick et al.)
  - Faster-RCNN (2016, Shaoqing Ren et al.)

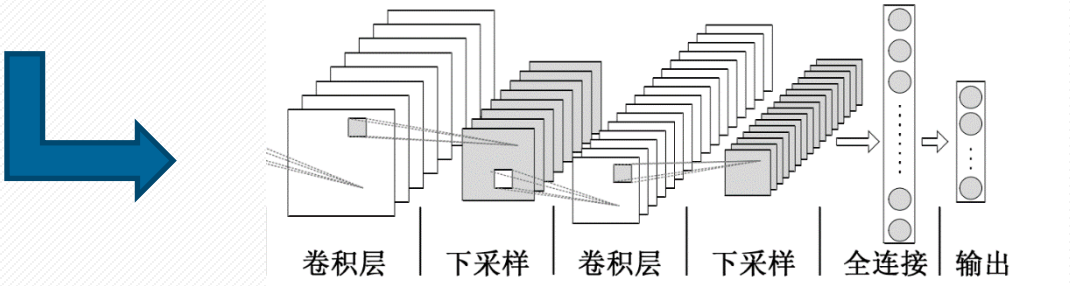
- 以YOLO系列为代表的单阶段目标检测算法
  - **YOLO v1 (2015, Joseph Redmon et al.)**
  - **YOLO v2 (2016, Joseph Redmon et al.)**
  - **YOLO v3 (2018, Joseph Redmon et al.)**
  - YOLO v4
  - YOLO v5
  - YOLO X

第二部分

Two-Stage Det

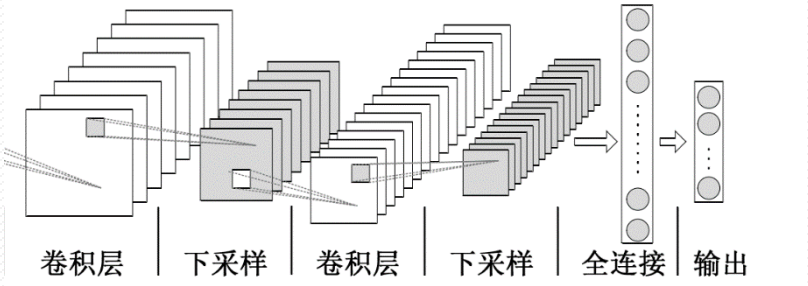
双阶段目标检测器

滑动窗口目标检测

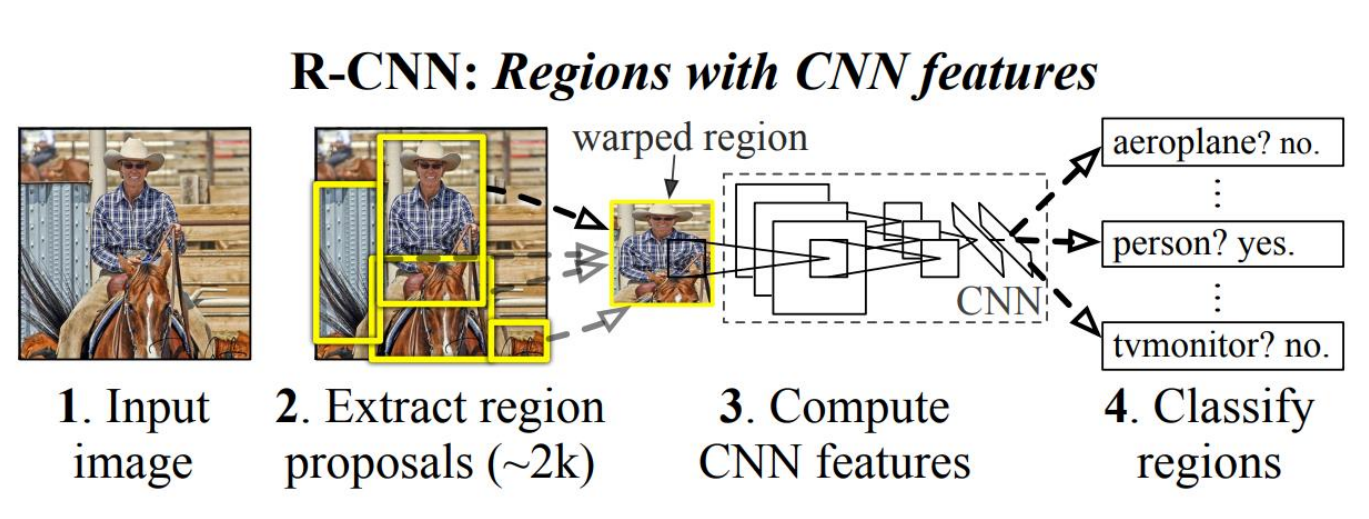


滑动窗口目标检测

- 算法实现简单
- 较高的召回率
- 难以接受的时间复杂度
- 较低的精确率



R-CNN：2014, Ross Girshick et al.



Fast R-CNN：2015, Ross Girshick et al.

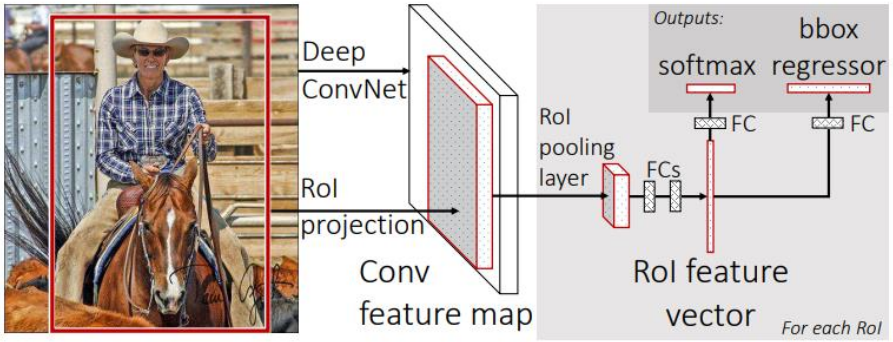


Figure 1. Fast R-CNN architecture. An input image and multiple regions of interest (RoIs) are input into a fully convolutional network. Each RoI is pooled into a fixed-size feature map and then mapped to a feature vector by fully connected layers (FCs). The network has two output vectors per RoI: softmax probabilities and per-class bounding-box regression offsets. The architecture is trained end-to-end with a multi-task loss.

Faster R-CNN：2016, Shaoqing Ren et al.

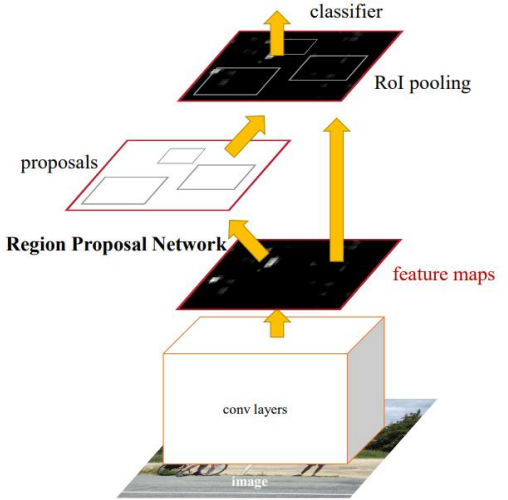


Figure 2: Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.

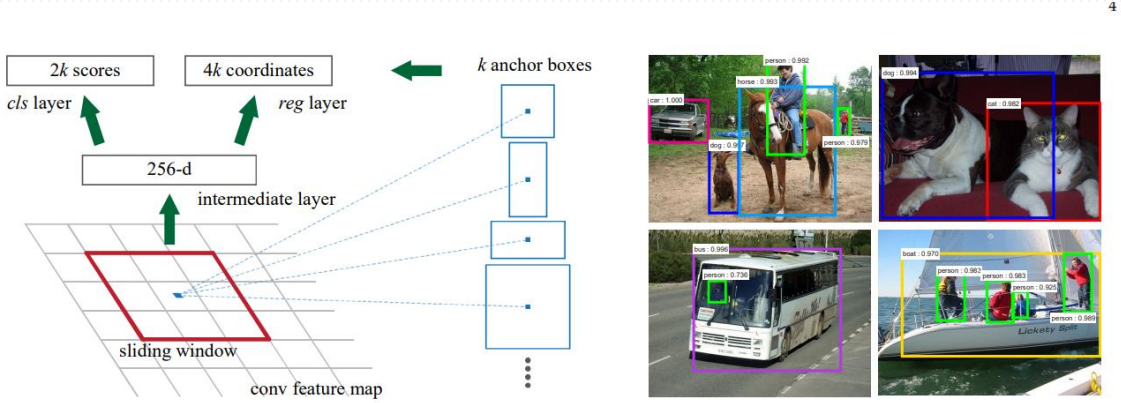


Figure 3: **Left**: Region Proposal Network (RPN). **Right**: Example detections using RPN proposals on PASCAL VOC 2007 test. Our method detects objects in a wide range of scales and aspect ratios.

第二部分

One-Stage Det

单阶段目标检测器

YOLO：2015, Joseph Redmon et al.

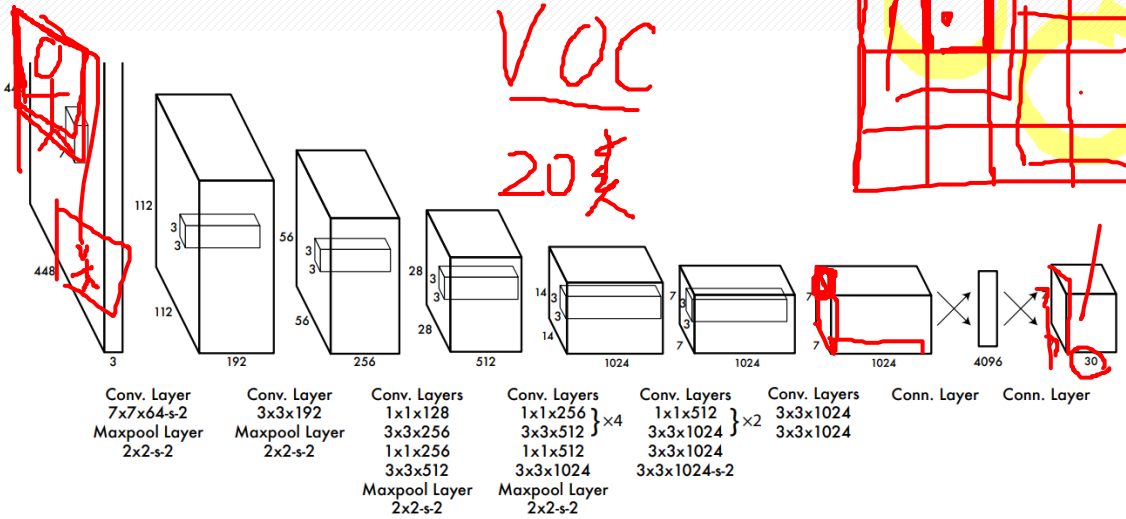


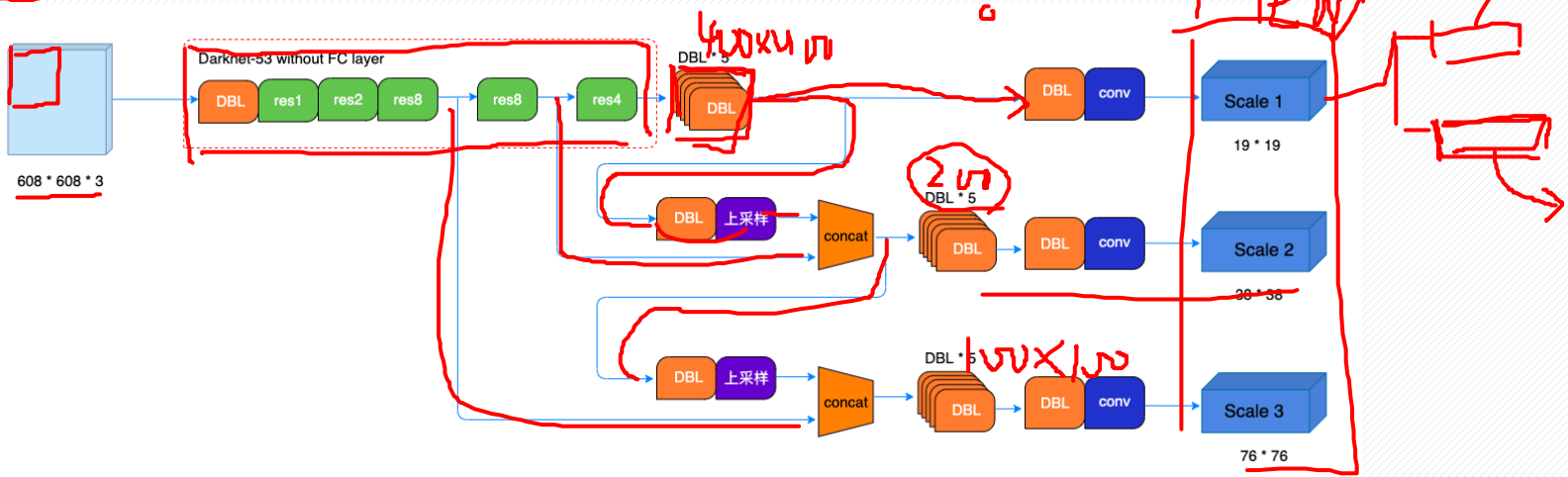
Figure 3: The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1×1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224×224 input image) and then double the resolution for detection.

YOLOv2 (YOLO 9000) : 2016, Joseph Redmon et al.

- Batch Normalization
- Anchor Boxes
- Demension Cluster
- ……

	YOLO							YOLOv2
batch norm?		✓	✓	✓	✓	✓	✓	✓
hi-res classifier?			✓	✓	✓	✓	✓	✓
convolutional?				✓	✓	✓	✓	✓
anchor boxes?				✓	✓			
new network?					✓	✓	✓	✓
dimension priors?					✓	✓	✓	✓
location prediction?					✓	✓	✓	✓
passthrough?						✓	✓	✓
multi-scale?							✓	✓
hi-res detector?								✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8	**78.6**

YOLOv3: 2016, Joseph Redmon et al.



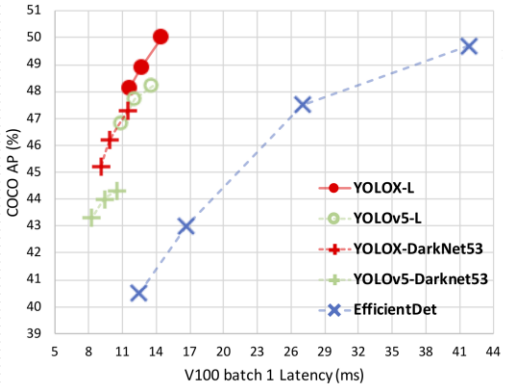
YOLOv4: 2020, Alexey Bochkovskiy et al.

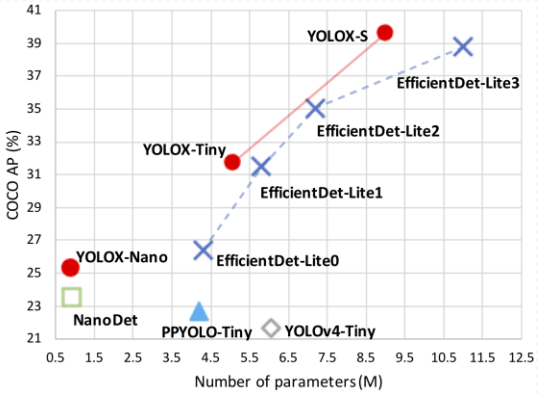
YOLOv5: ????

YOLOX: 2021, Ge Zheng et al.

YOLOX

Exceeding YOLO series in 2021





THANKS!

主讲人 安泓郡

ACoTAI Lab, Dalian Maritime University
ICDC department, Dianhang Association

an.hongjun@foxmail.com