周工作报告
DNN for Object Detection

刘四平

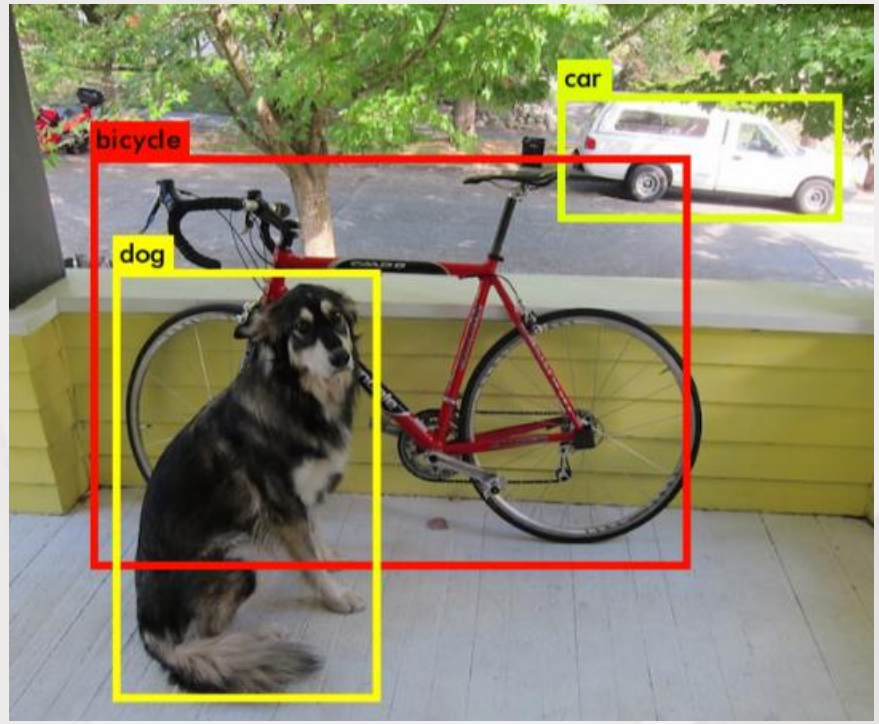
2017.9.8

基于深度学习的目标检测

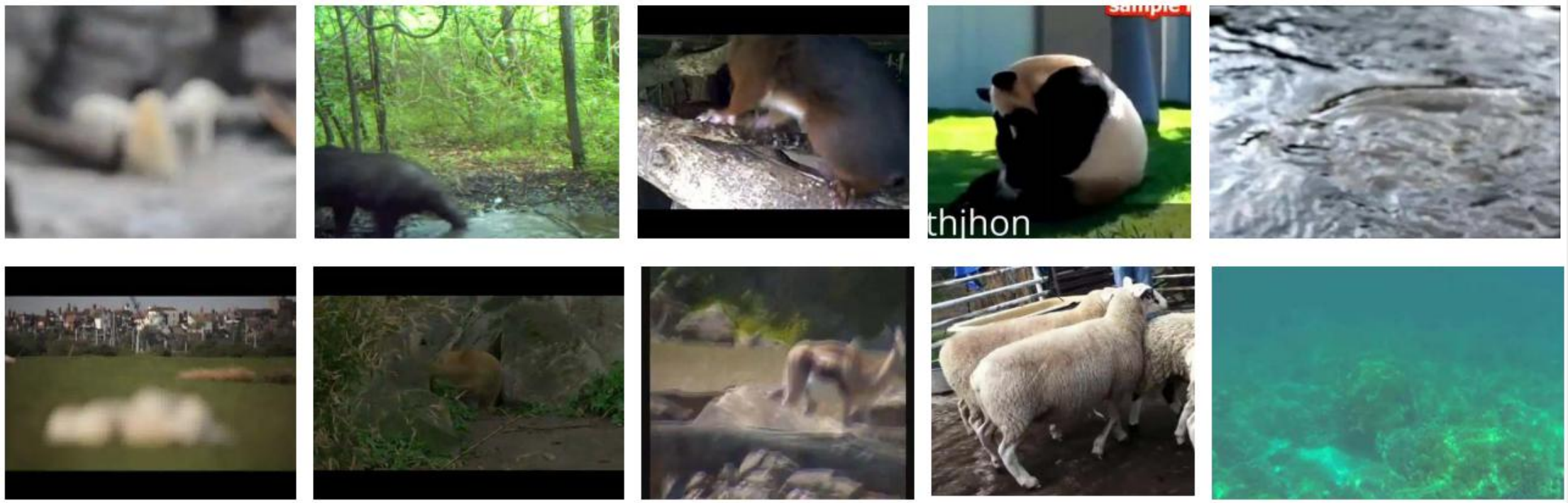
- 利用卷积神经网络(CNN)提取图像特征
- 识别图像中的目标的类别和位置

- 视频分析：监控领域，无人驾驶
- 特点：时间序列的图像， 帧的相关性

- 目的：进行实时目标检测



Challenges of Image/Video



camera defocus partial occlusion motion blur crowded instance background confusion

问题和方法

- 嵌入式 / 移动终端智能应用环境（Fog/Edge Computing）
- 将深度模型应用于视频中的每一帧图像？耗时，代价大

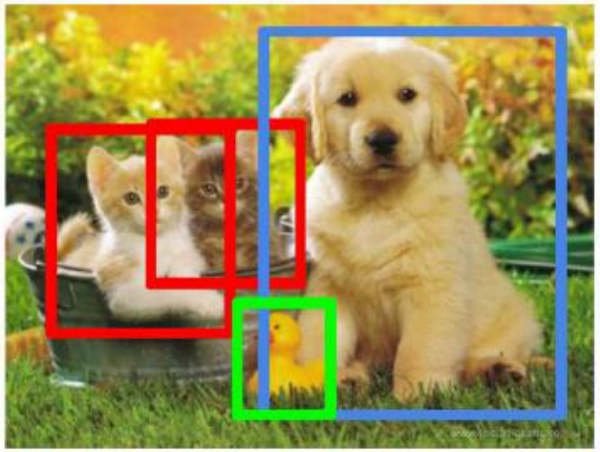
方法

- 利用时间-空间相关信息加速视频的实时分析，并提高准确率
- 权衡准确率和速度

Image Recognition

Classification Localization Object Detection Instance Segmentation






Single object, single class Multiple objects and classes Pixels level

ILSVRC2017-imagenet

Image Classification

Team name	Entry description	**Classification error**	Localization error
WMW	Ensemble C [No bounding box results]	0.02251	0.590987

Object localization

Team name	Entry description	**Localization error**	Classification error
NUS-Qihoo_DPNs (CLS-LOC)	[E3] LOC:: Dual Path Networks + Basic Ensemble	0.062263	0.03413

Object detection

Team name	Entry description	**mean AP**	Number of object categories won
BDAT	submission3	0.732227	65

Object detection from video

Team name	Entry description	**Number of object categories won**	mean AP
IC&USYD	provide_submission3	15	0.817265

source: http://image-net.org/challenges/LSVRC/2017/results

Object Detection from Video

- Object detection from video with provided training data

Rank 1# mAP: 81.8309%

mAP: 80.8292% (2016 NUIST)

- Object detection from video with additional training data

Rank 1# mAP: 81.9339%

- Object detection/tracking from video with provided training data

Rank 1# mAP: 64.1474%

mAP: 55.8557% (2016 CUVideo)

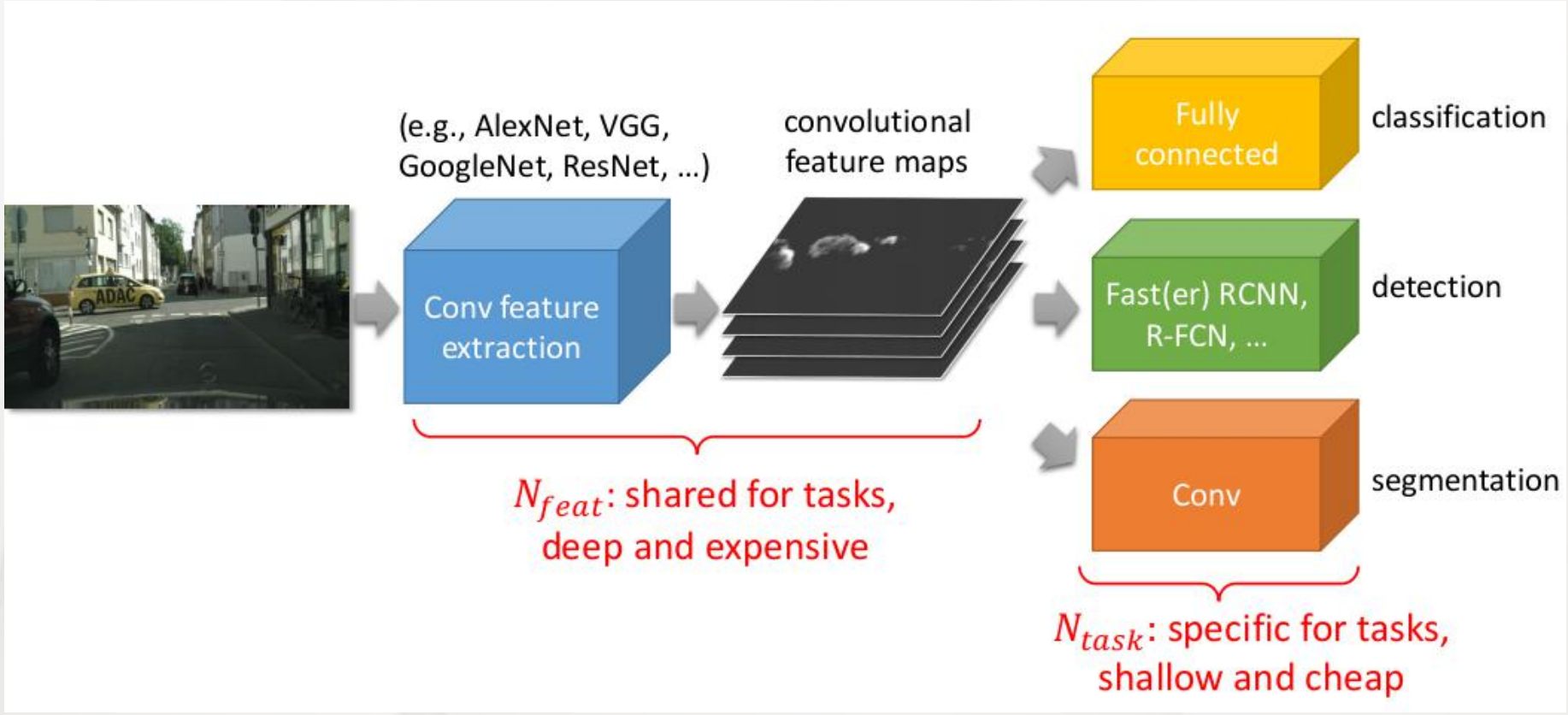
- Object detection/tracking from video with additional training data

Rank 1# mAP: 64.2935%

source: IC&USYD

General Design

- 基础网络结构：VGG, GoogleNet(Inception), ResNet，ResNeXt
- 对象检测方法：Region-Based, Regression



目标检测方法分类

- 传统方法：DPM(Deformable Part Model)
- 基于区域候选+分类的方法：R-CNN，R-FCN
- 基于回归的方法：YOLO，SSD

YOLO: You Only Look Once

实时目标检测

- **You Only Look Once: Unified, Real-Time Object Detection, CVPR 2016**
- **YOLOv2/YOLO9000 : Better, Faster, Stronger, CVPR 2017**

更加快速（faster）

Darknet-19网络在 ImageNet 1000类分类数据集训练

网络去掉了最后一个卷积层，加上了三个3*3卷积层，每个卷积层有1024个Filters，每个卷积层紧接着一个1*1卷积层

多分类的改进（Stronger）

结合词向量树（wordtree），使检测种类扩充到了9000类。融合**多个数据集(COCO+Imagenet)**的标签信息, 使用联合训练方法，并进行细粒度的标记，比如狗这一类就包括“哈士奇”、“金毛狗”等。

YOLOv2网络结构

精度的改进（Better）

Batch Normalization:采用正则化方法对网络进行优化，提高收敛性。

High Resolution Classifier:采用448×448分辨率的ImageNet数据使网络适应高分辨率输入；然后用于目标检测任务finetune。

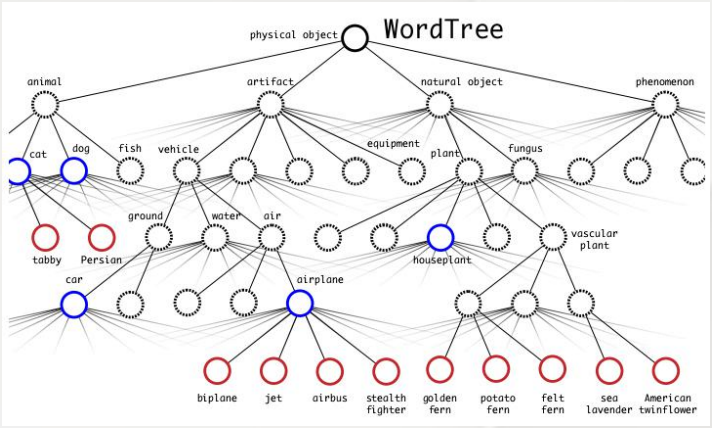
Convolutional With Anchor Boxes:之前的YOLO利用全连接层的数据完成边框的预测，导致丢失较多的空间信息，定位不准。借鉴Faster R-CNN中的Anchor思想。

Dimension Clusters:在训练集Bounding Boxes下进行k-means聚类，找值。

Fine-Grained Features:引入passthrough layer，作用是将上一层特征图的相邻像素都切除一部分组成了另外一个通道。将26*26*512的特征图变为13*13*2048的特征图，有利用小目标物的检测。

Multi-Scale Training:使模型对不同尺寸图像具有鲁棒性。

Type	Filters	Size/Stride	Output
Convolutional	32	3 × 3	224 × 224
Maxpool		2 × 2/2	112 × 112
Convolutional	64	3 × 3	112 × 112
Maxpool		2 × 2/2	56 × 56
Convolutional	128	3 × 3	56 × 56
Convolutional	64	1 × 1	56 × 56
Convolutional	128	3 × 3	56 × 56
Maxpool		2 × 2/2	28 × 28
Convolutional	256	3 × 3	28 × 28
Convolutional	128	1 × 1	28 × 28
Convolutional	256	3 × 3	28 × 28
Maxpool		2 × 2/2	14 × 14
Convolutional	512	3 × 3	14 × 14
Convolutional	256	1 × 1	14 × 14
Convolutional	512	3 × 3	14 × 14
Convolutional	256	1 × 1	14 × 14
Convolutional	512	3 × 3	14 × 14
Maxpool		2 × 2/2	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	512	1 × 1	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	512	1 × 1	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	1000	1 × 1	7 × 7
Avgpool		Global	1000
Softmax			



准确率和速度对比

Detection Frameworks	Train	mAP	FPS
Fast R-CNN [5]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[15]	2007+2012	73.2	7
Faster R-CNN ResNet[6]	2007+2012	76.4	5
YOLO [14]	2007+2012	63.4	45
SSD300 [11]	2007+2012	74.3	46
SSD500 [11]	2007+2012	76.8	19
YOLOv2 288 × 288	2007+2012	69.0	91
YOLOv2 352 × 352	2007+2012	73.7	81
YOLOv2 416 × 416	2007+2012	76.8	67
YOLOv2 480 × 480	2007+2012	77.8	59
YOLOv2 544 × 544	2007+2012	**78.6**	40

- FPS(Frames Per Second)
- mean AP
- VOC 2007+2012

本学期目标

细化研究问题和方向

综述报告

阅读论文：

- CNN网络结构设计: DPN, MobileNet, ShuffeNet, SqueezeNet
- 目标检测方法（实时、准确率、能耗）: SSD, RON, SqueezeDet

关注最新动态：

UCB, Stanford

MSRA, Google, Facebook; 商汤科技, Face++, 深鉴科技

谢谢

请各位老师和同学批评指正

Reference

- Speed/Accuracy Trade-offs for Object Detection from Video, IC&USYD
- YOLO9000: Better, Faster, Stronger
- You Only Look Once: Unified, Real-Time Object Detection
- Large Scale Visual Recognition Challenge 2017 (ILSVRC2017)
- Flow Based Video Recognition, Yichen Wei, MSRA
- Faster r-cnn: Towards real-time object detection with region proposal networks. S. Ren, K. He, R. Girshick, and J. Sun.
- SSD: single shot multibox detector. W. Liu.