

法律声明

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。



关注 小象学院

第1讲 课程介绍

计算机视觉

主讲人：叶梓

上海交通大学博士

主要研究方向：机器学习、深度学习、人工智能

本章内容

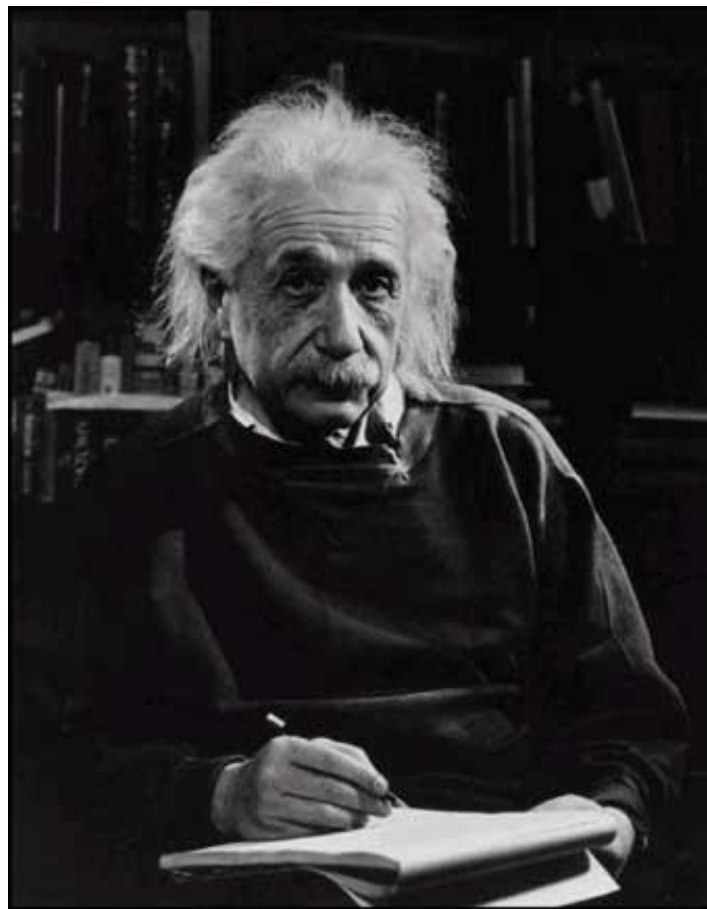
- 计算机视觉概述
- 研究挑战
- 课程简介
- 开源库与应用环境介绍
- 参考书
- 环境搭建

计算机视觉

图片信息量有多大？

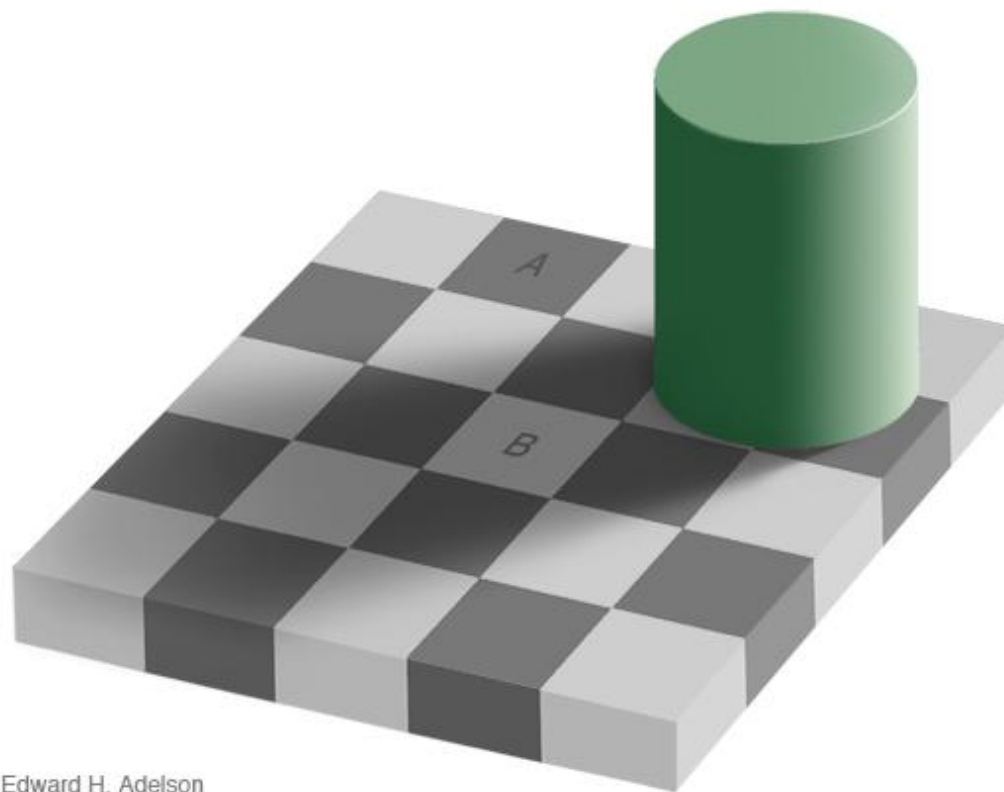
→ 1000个词？

→ 一图胜千言



计算机视觉

错觉→ A和B的颜色一样吗？



Edward H. Adelson

计算机视觉

研究理论和应用：

- 研究如何使机器^{视觉}“看”的科学
- 让计算机具有人类视觉的所有功能
- 让计算机从图像数据中，提取有用的信息，并解释
- 重构人眼；重构视觉皮层；重构大脑剩余部分

计算机视觉

模拟人类视觉的优越能力：

- 识别人、物体、场景
- 估计立体空间、距离
- 躲避障碍物进行导航
- 想象并描述故事
- 理解并讲解图片

计算机视觉

弥补人类视觉的缺陷：

- 关注显著内容、容易忽略很多细节
- 不在乎、不擅长精细感知
- 容易受幻觉干扰
- 描述主观、模棱两可
- 不善于长时间稳定的执行同一个任务
- 当然，还有视觉障碍人士

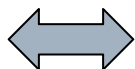
计算机视觉

图像数据
(图片、视频、深度图片等)



解释

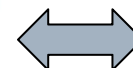
方向盘
特斯拉
地图
无人驾驶
.....



感知设备
(摄像头)



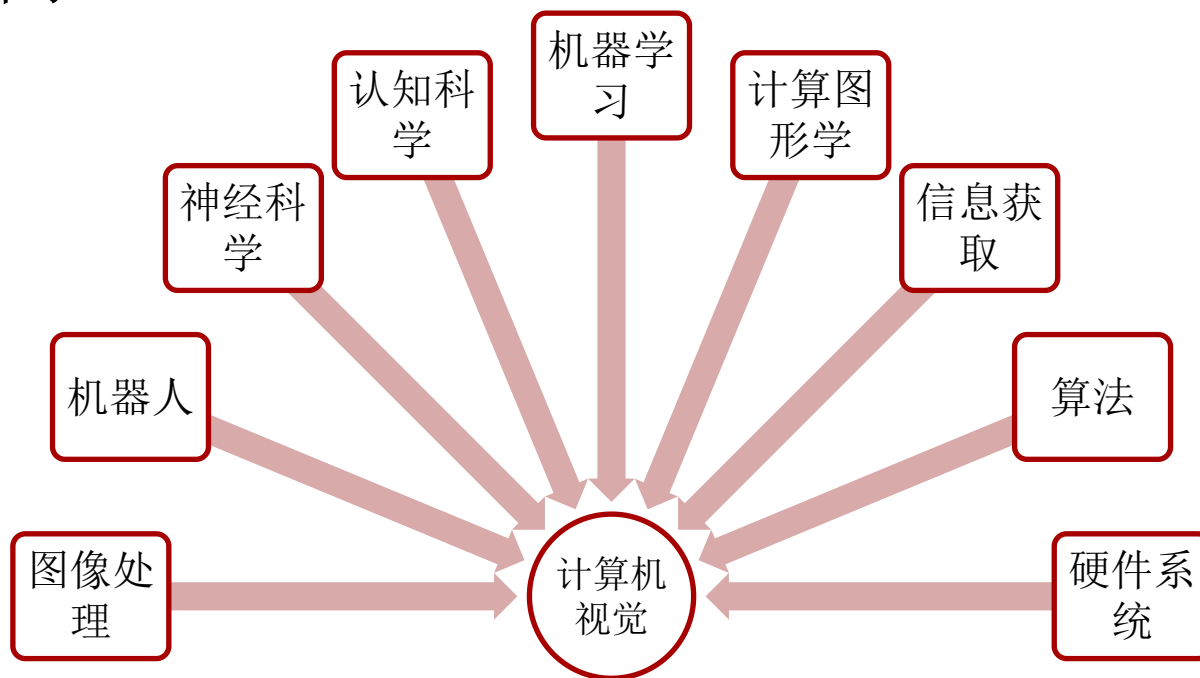
计算设备
(CPU/GPU/FPGA)



计算机视觉

高度复合学科

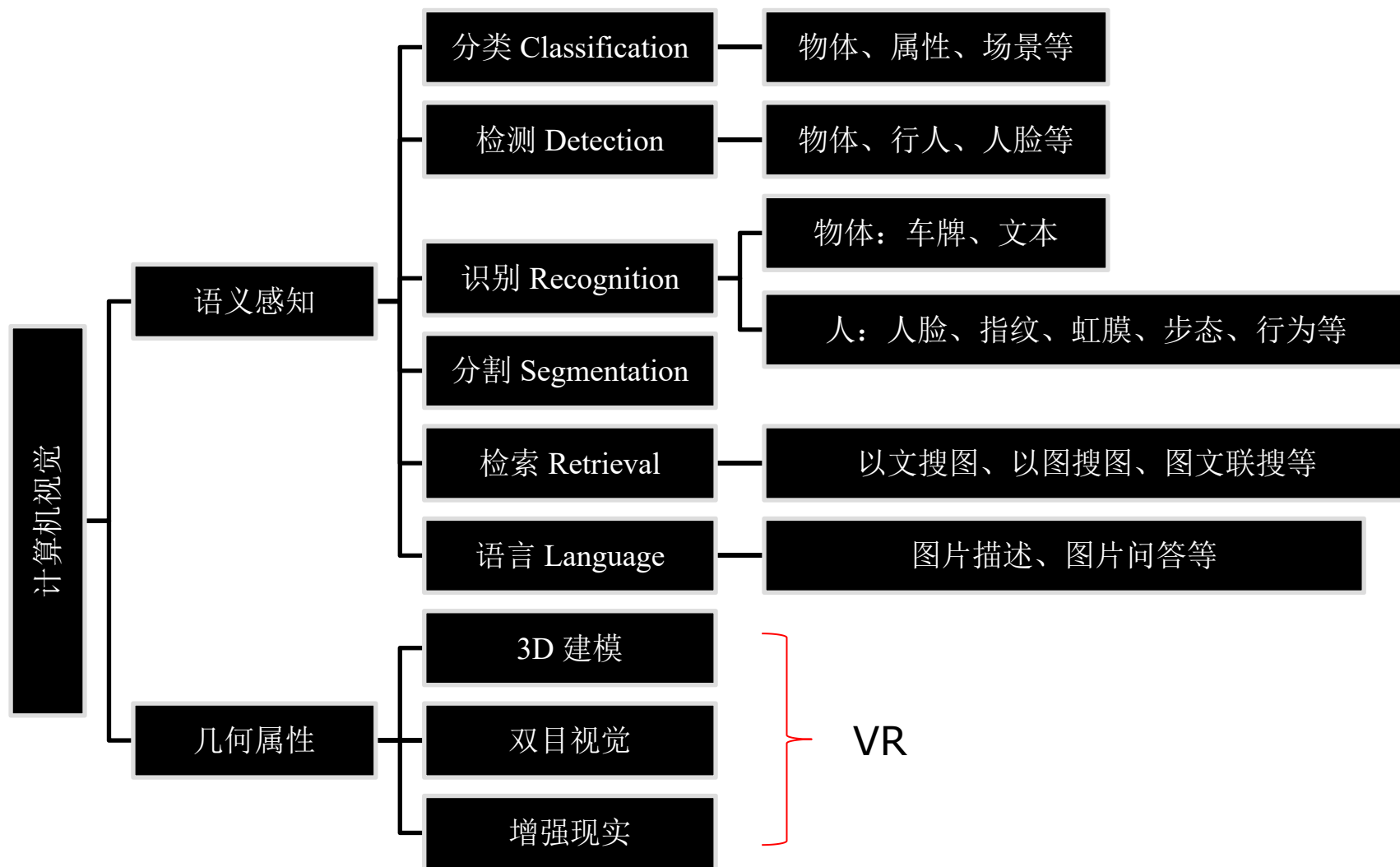
- 工程
- 计算机科学
- 数学
- 心理学
- 生物学
- ...



计算机视觉

- 2个主要研究维度
 - 语义感知（semantic）
 - 几何属性（Geometry）
- 赶上人类视觉了吗？
 - 迅猛发展（特定领域已赶上）
 - 任重道远（ not yet ）

计算机视觉



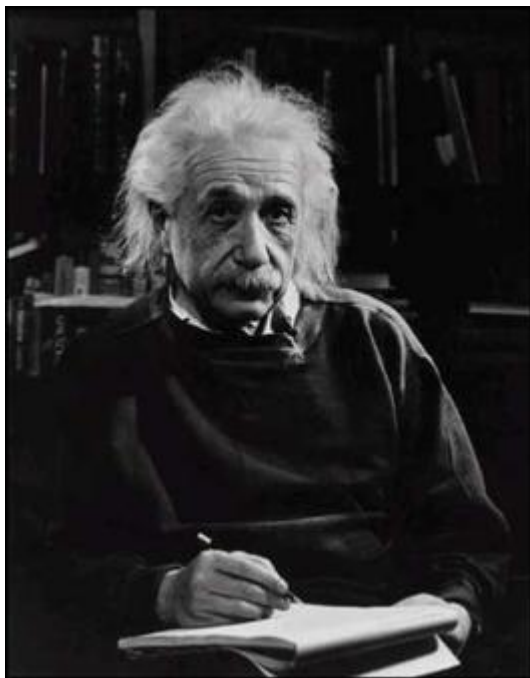
计算机视觉

语义感知在人工智能中是主角：

- 场景理解 ← 最终形态
 - 视觉描述（Visual Captioning）
 - 视觉问答（Visual Question Answering）
- 基础理解 ← 根本
 - 分类、识别、检测、分割、显著性等
 - 得益于深度学习网络发展，性能大幅提升

计算机视觉

感知上的本质差异



人类

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

计算机

计算机视觉

- 人工智能目标
 - 解决”像素值”与”语义”之间的鸿沟（Gap）



人眼所看到的



0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

计算机所看到的

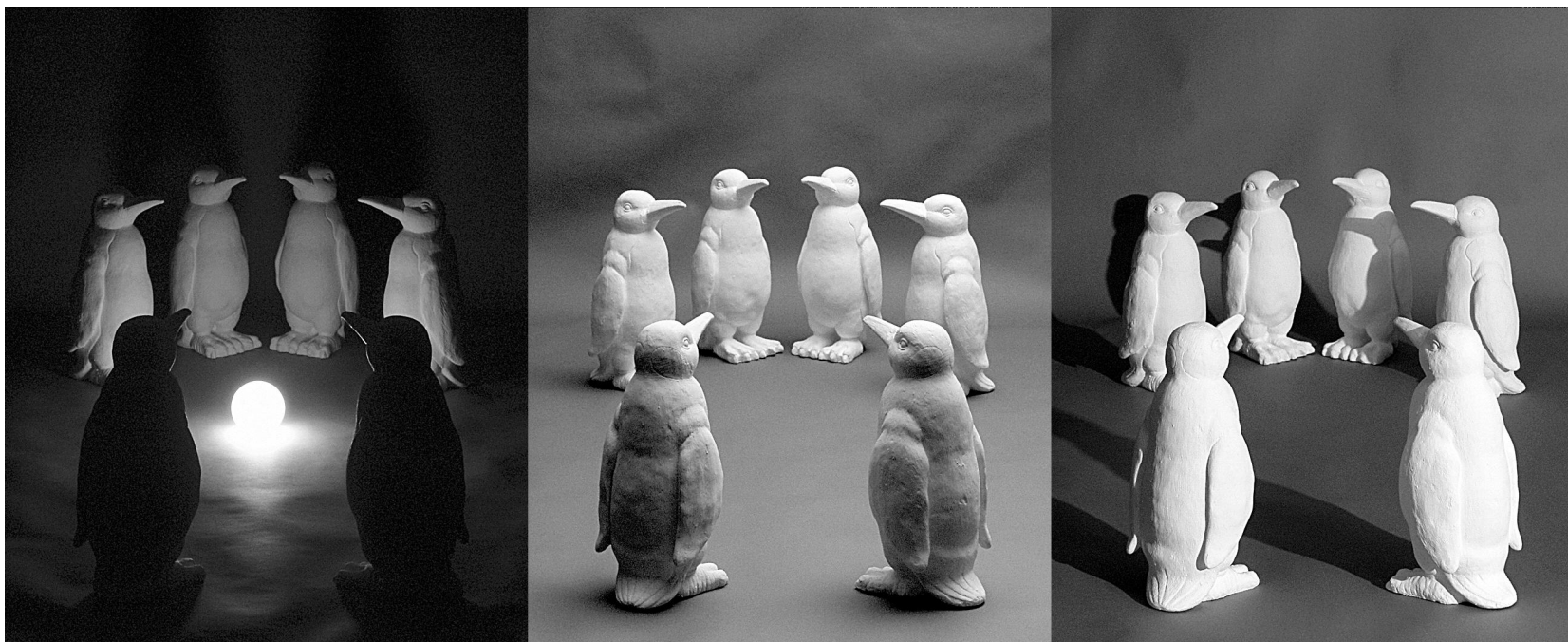
研究挑战

- 视角变化



研究挑战

- 光照变化



研究挑战

- 尺度变化



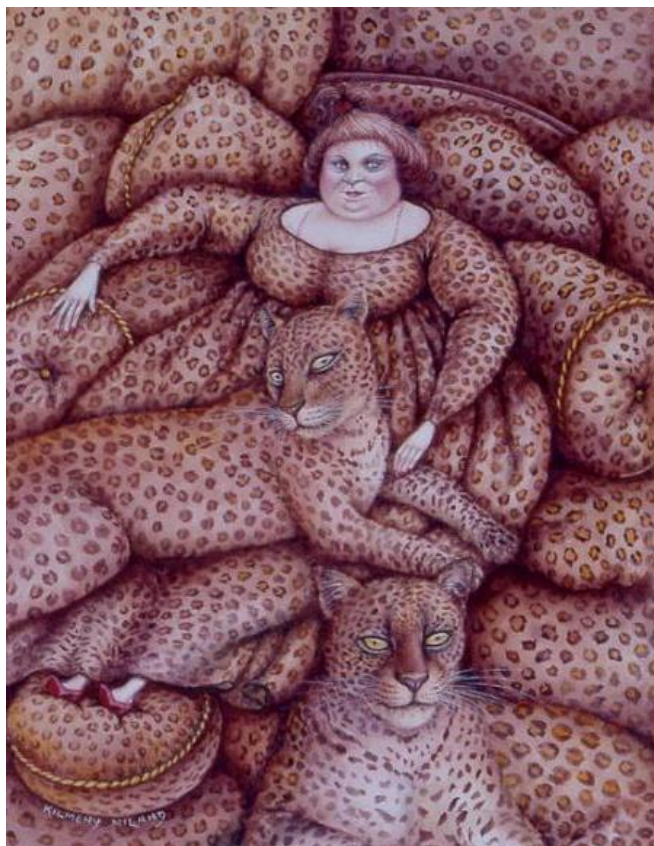
研究挑战

- 形态变化



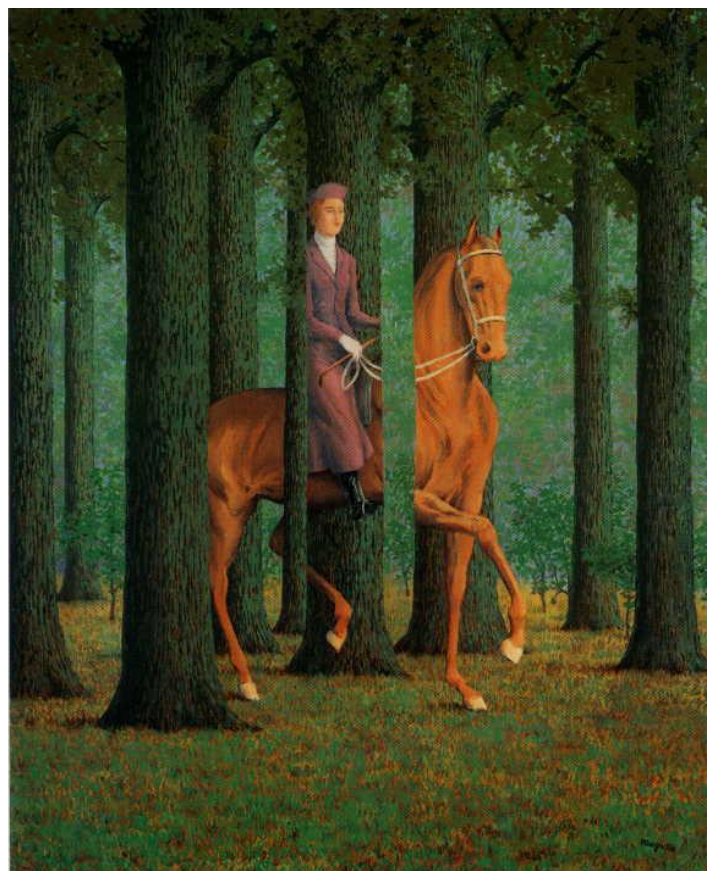
研究挑战

● 背景混淆干扰



研究挑战

- 遮挡



研究挑战

- 类内物体的外观差异



课程简介

□ 部分I：计算机视觉的基础

- 图像预处理
- 图像特征及描述
- 未有深度学习之前

□ 部分II：深度学习的理论基础

- BP神经网络详解
- 深度学习基础

□ 部分III：深度学习在计算机视觉中的应用

- 图像分类 ← 卷积神经网络CNN
- 图像检测 ← 区域卷积神经网络R-CNN
- 图像分割 ← 全卷积神经网络FCN
- 图像描述 ← 迭代神经网络RNN
- 图像生成 ← 生成对抗网络GAN

图像检索

主要研究问题

□ 图像预处理

- 图像显示与存储原理
- 图像增强的目标
- 点运算：基于直方图的对比度增强
- 形态学处理
- 空间域处理：卷积
- 卷积的应用（平滑、边缘检测、锐化等）
- 频率域处理：傅里叶变换、小波变换

主要研究问题

□ 图像特征及描述

■ 颜色特征

□ 量化颜色直方图

□ 聚类颜色直方图

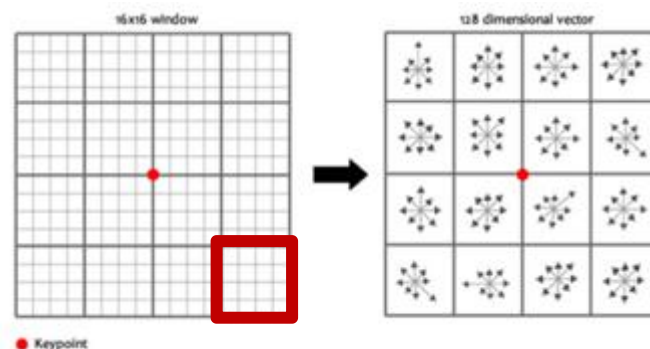
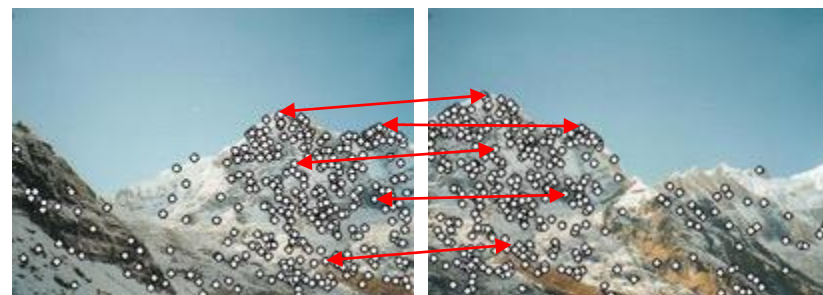
■ 几何特征

□ Edge, Corner, Blob

■ 基于关键点的特征描述子

□ SIFT, SURF, ORB

■ 其他特征提取（LBP、Gabor）



主要研究问题

□ 深度学习之前的方法

■ 图像分割

□ 基于阈值、基于边缘

□ 基于区域、基于图论

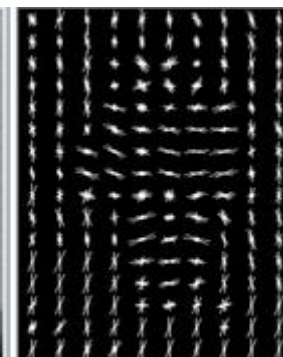
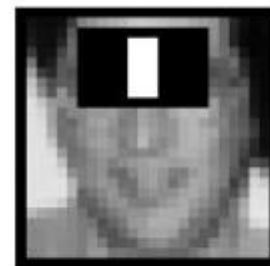
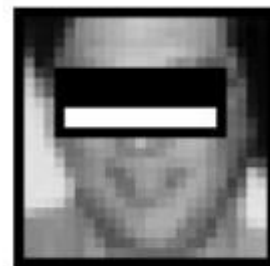
■ 人脸检测

□ Haar-like特征+级联分类器

■ 行人检测

□ HOG+SVM

□ DPM

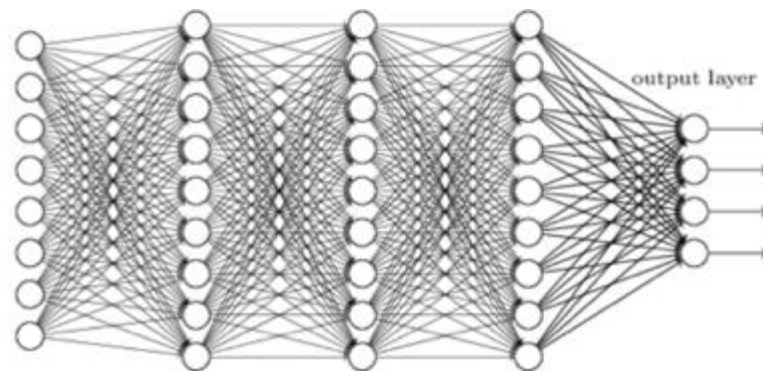


主要研究问题

□ 神经网络与深度学习基础

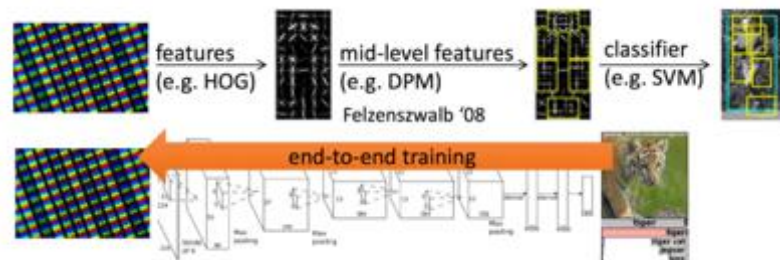
■ 神经网络

- 神经元
- 前馈网络
- 梯度下降
- 误差反向传播



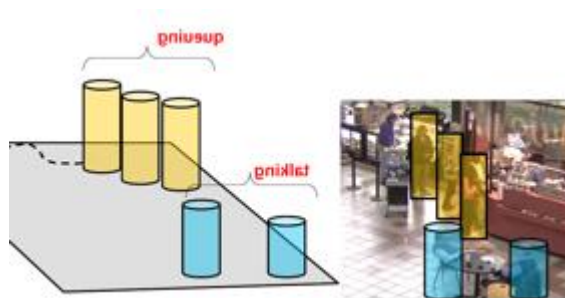
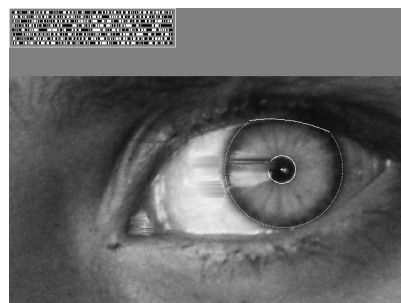
■ 深度学习

- 与传统神经网络的区别
- 目标函数
- 改进的梯度下降
- 避免过适应



主要研究问题

- 识别→内容是什么？
 - 人脸识别
 - 字符识别
 - 车牌识别
 - 行为识别
 - 步态识别



主要研究问题

- 基于深度学习的方法
- 图片分类→有没有？

Input Image



Thank you. After analysis, our system describes the image as follows:

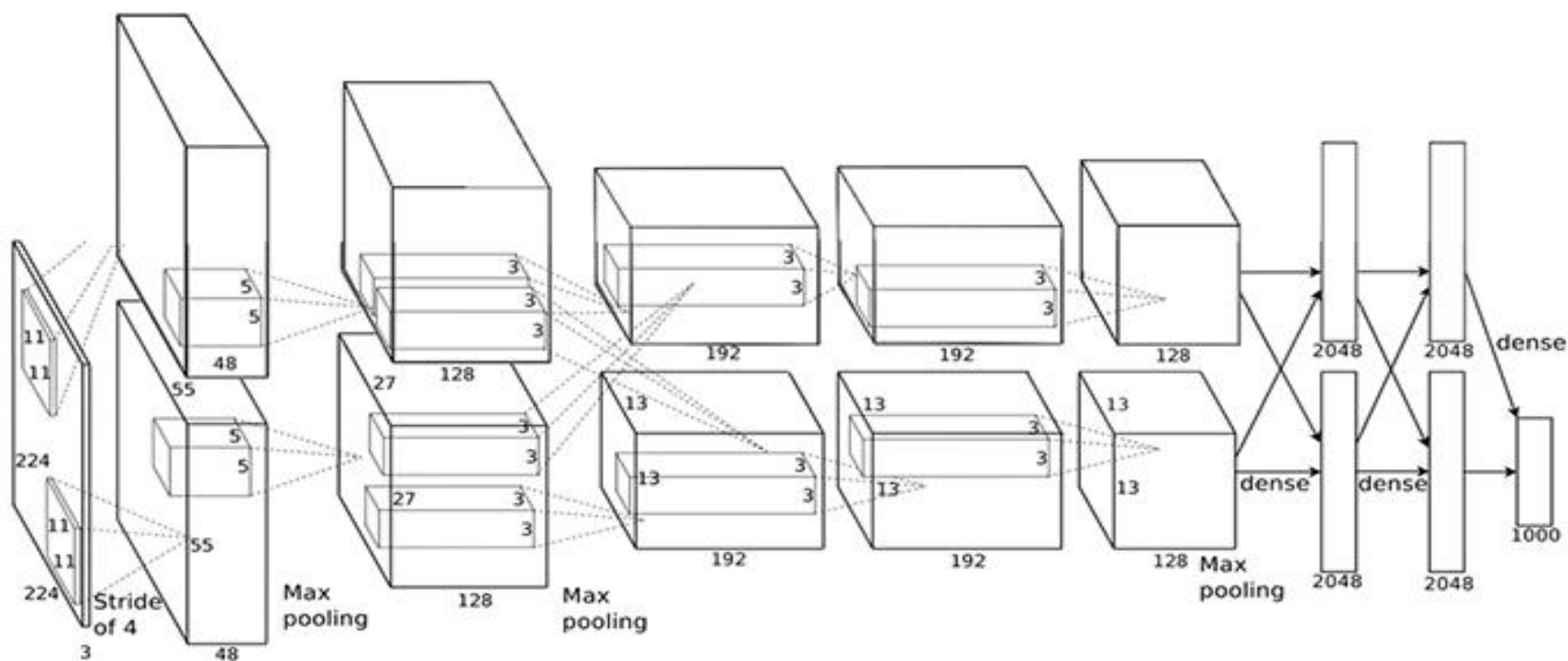
butterfly	3.49354
lepidopterous insect	3.42100
insect	2.50084
arthropod	2.01440
nymphalid	1.97341
invertebrate	1.73042
lycaenid	0.40975
admiral	0.38683

主要研究问题

- 卷积神经网络CNN
 - 计算机视觉中的基础网络
 - 有监督深度模型时代的起点
 - AlexNet→VGG→GoogLeNet→ResNet→ResNeXt
 - GoogLeNet
 - Inception V1 → V2 → V3 → V4
 - Inception ResNet V1 → V2
 - 结构趋势
 - 更深（Depth）：8层→1000+层
 - 更宽（Width）：1分支→4+分支
 - 更多基数（Cardinality）：1→32

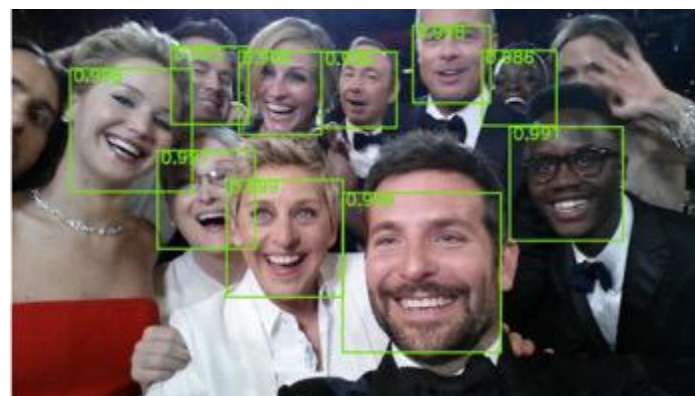
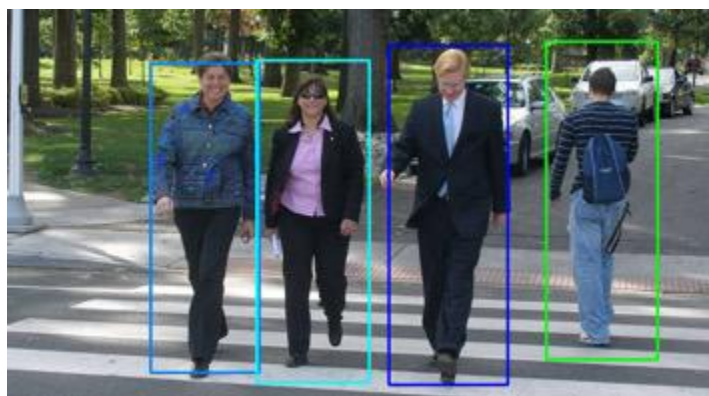
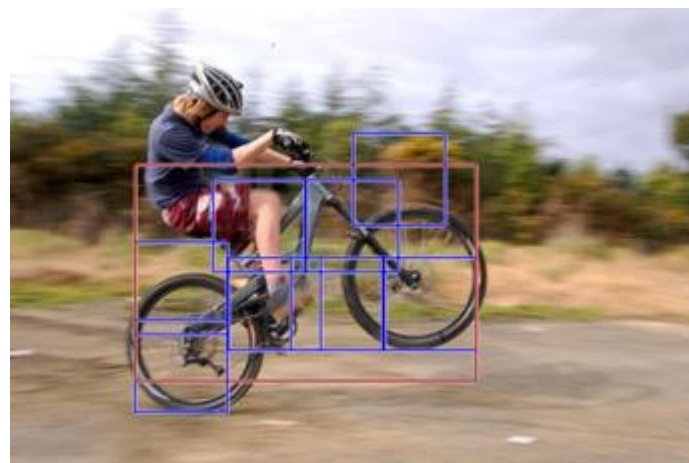
主要研究问题

AlexNet卷积网络



主要研究问题

- 基于深度学习的方法
- 检测→目标在哪儿？
 - 物体检测
 - 人脸检测
 - 姿态检测

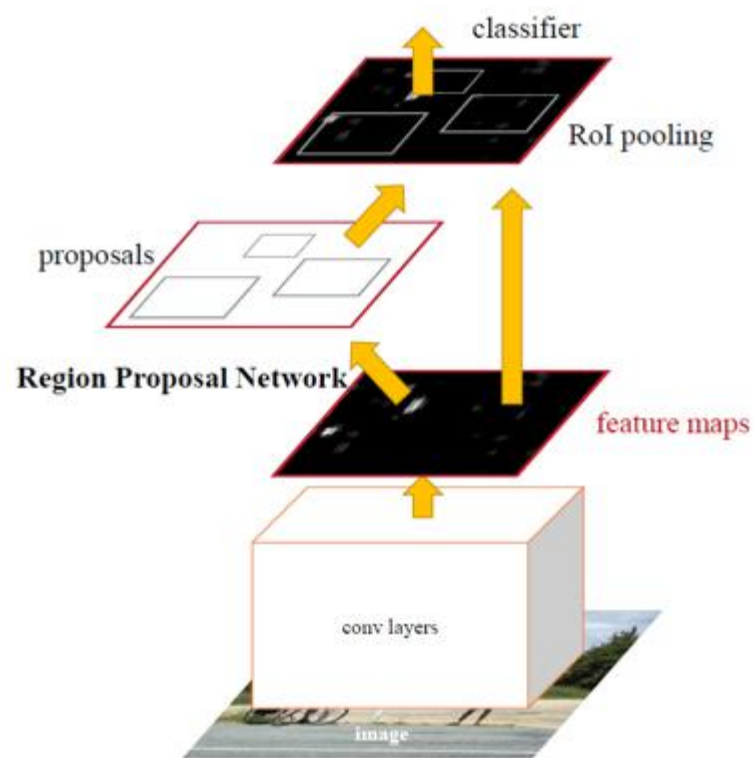
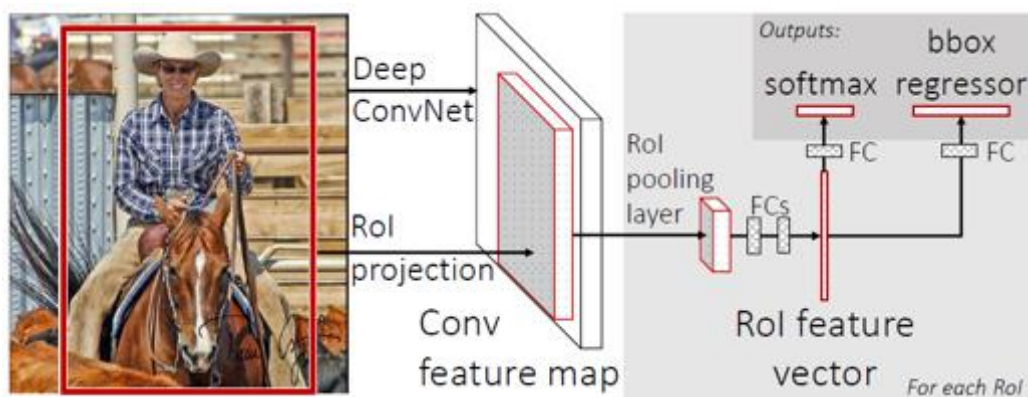
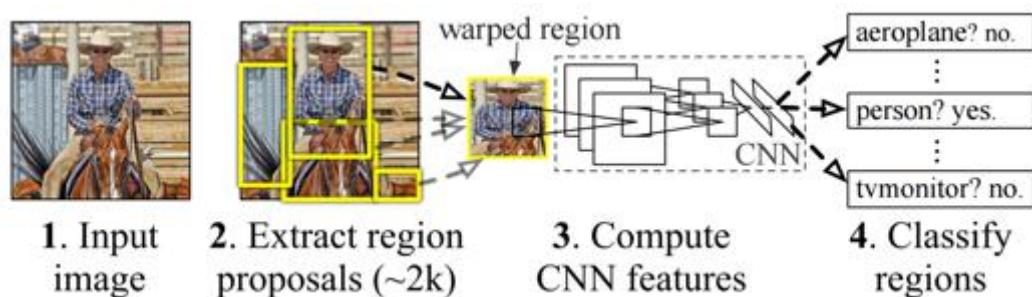


主要研究问题

- 区域卷积神经网络R-CNN
 - 让基础网络具备区域输出能力
 - 第1阶段：R-CNN→SPP-Net→Fast/Faster R-CNN
 - 第2阶段：YOLO→SSD→R-FCN
 - 目的
 - 检测更快、更准确
 - 工业应用
 - 智能监控
 - 辅助驾驶

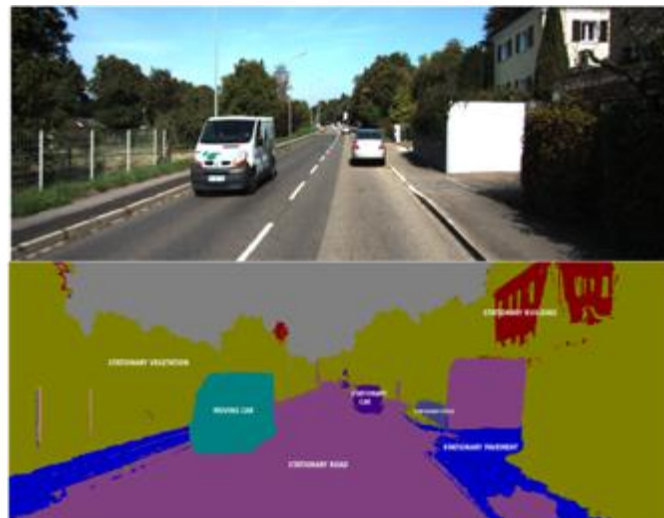
主要研究问题

R-CNN → Fast → Faster



主要研究问题

- 基于深度学习的方法
- 图片分割→区域轮廓?
 - 前景分割
 - 语义分割

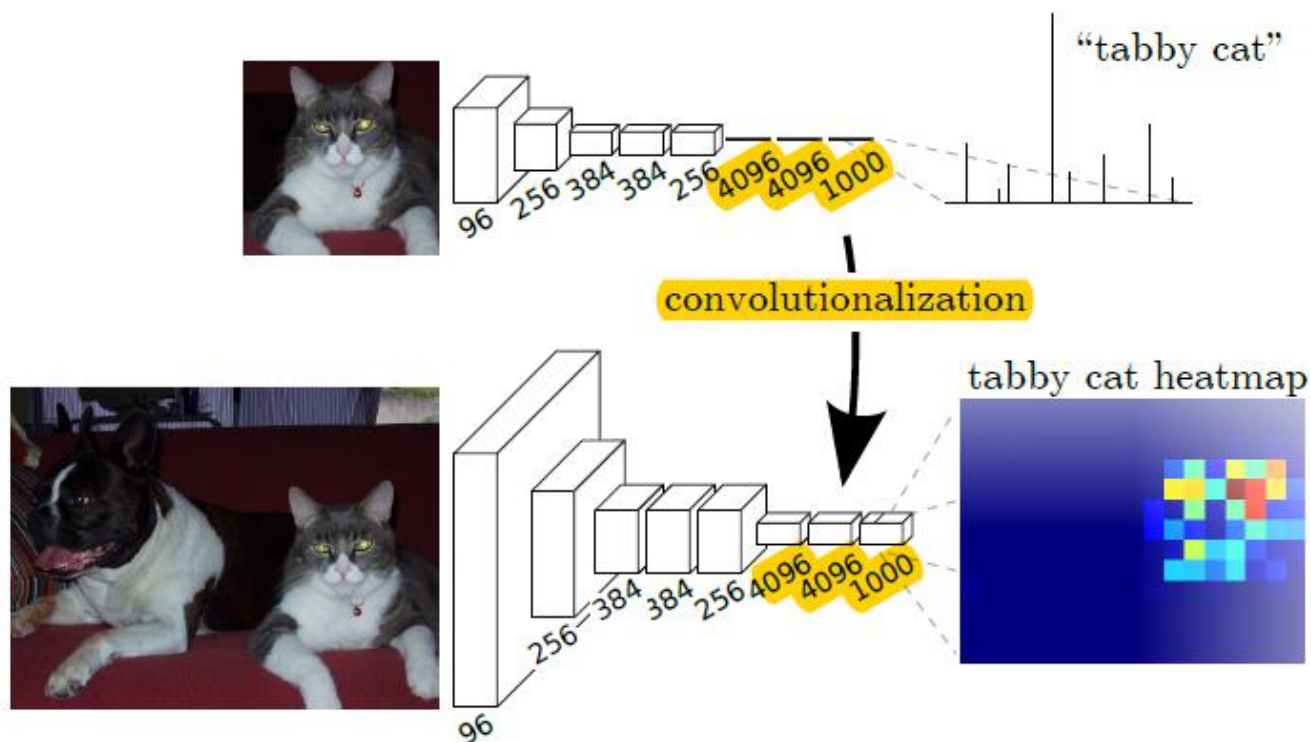


主要研究问题

- 全卷积神经网络FCN
 - 让基础网络做像素输出
 - FCN→SegNet/DeconvNet→DeepLab
 - 目的
 - 语义推断
 - 分割更精确
 - 工业应用
 - 辅助驾驶
 -

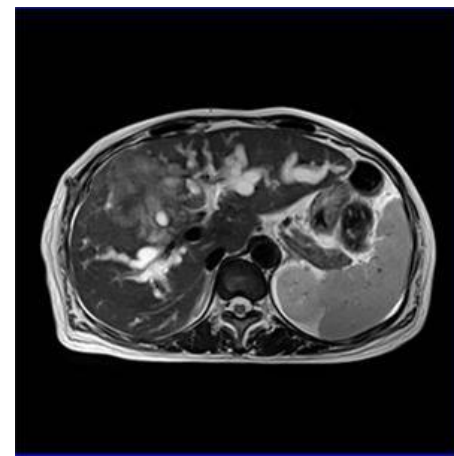
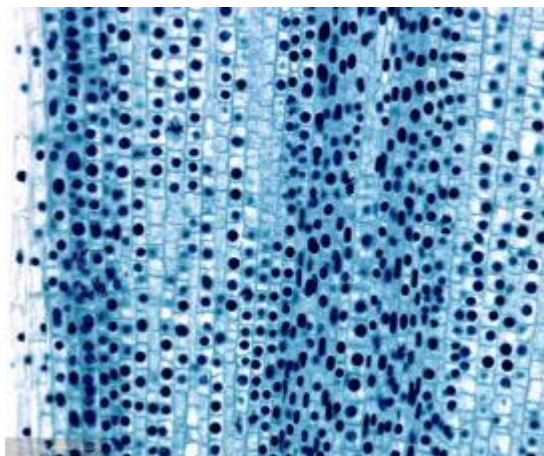
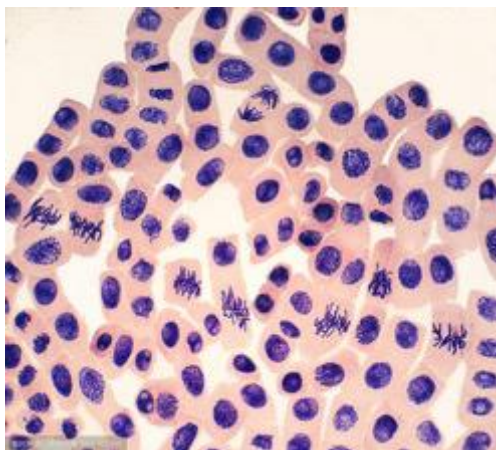
主要研究问题

- 全卷积网络FCN
 - 反卷积/转置卷积



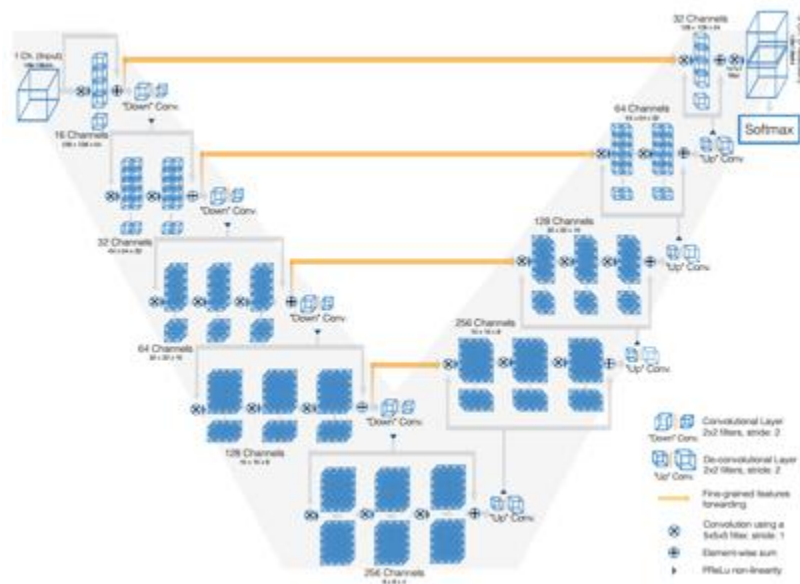
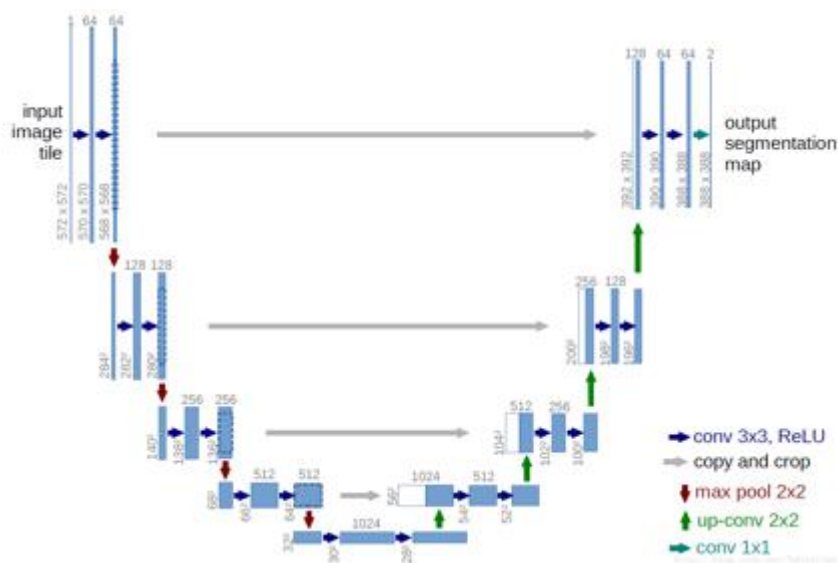
主要研究问题

- 对医学、生物影像中的器官、组织或细胞进行分割，即判断出影像中的每一个像素点是否属于器官、组织或细胞。



主要研究问题

- 生物医疗领域的图像分割
- U-net和V-net



主要研究问题

- 视觉&语言
 - 图片描述 (Captioning)
 - 图片问答 (Question Answering)
- Caption
 - 一个人在用自动驾驶系统开车
- QA
 - Q: 这个人为什么双手离开了方向盘
 - A: 因为他启动了无人驾驶功能

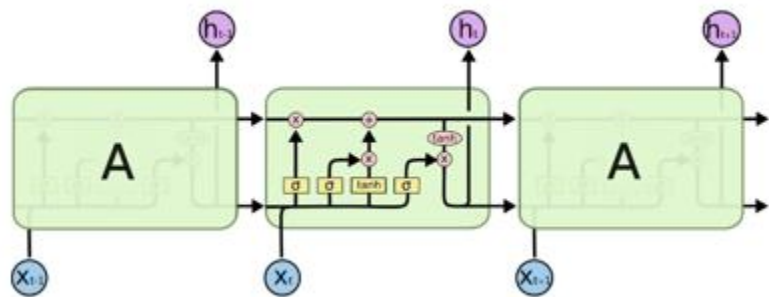
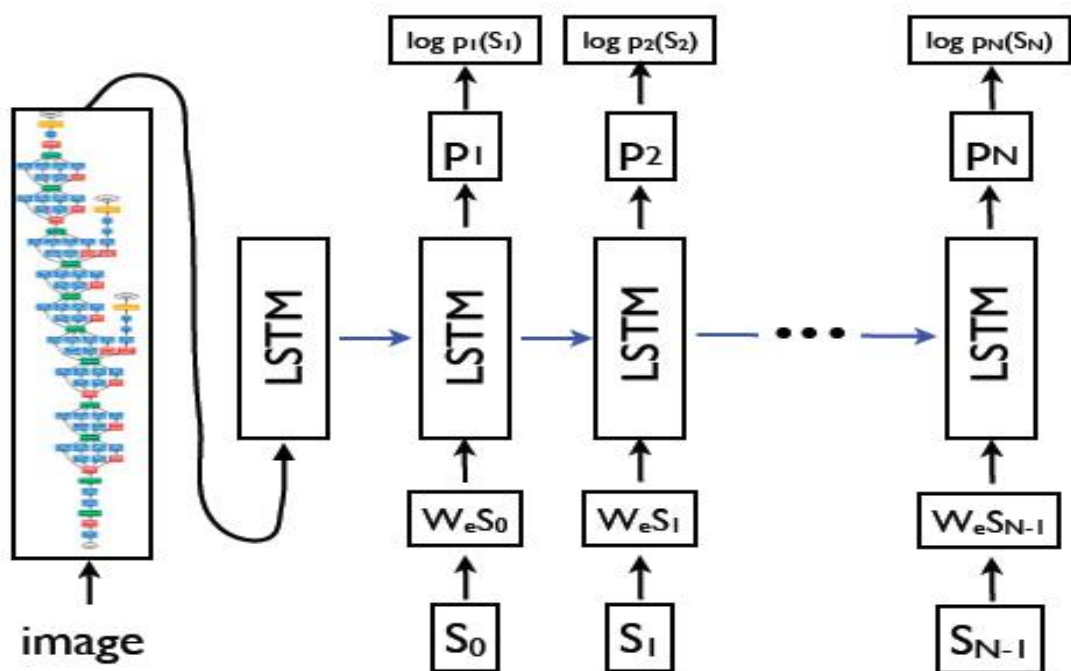


主要研究问题

- 递归神经网络RNN
 - 具有记忆功能，构建不定长序列数据的模型
 - Vanilla RNN→LSTM→GRU
 - 应用范围
 - 文本序列
 - 区域序列
 - 视频序列
 - 研究问题
 - 图片描述、问答
 - 机器翻译

主要研究问题

- 图说模型
 - CNN+LSTM



主要研究问题

- 图片生成
 - 样本分布学习
 - 超分辨率
 - 以图生文
 - 语义分割



this small bird has a pink breast and crown, and black primaries and secondaries.

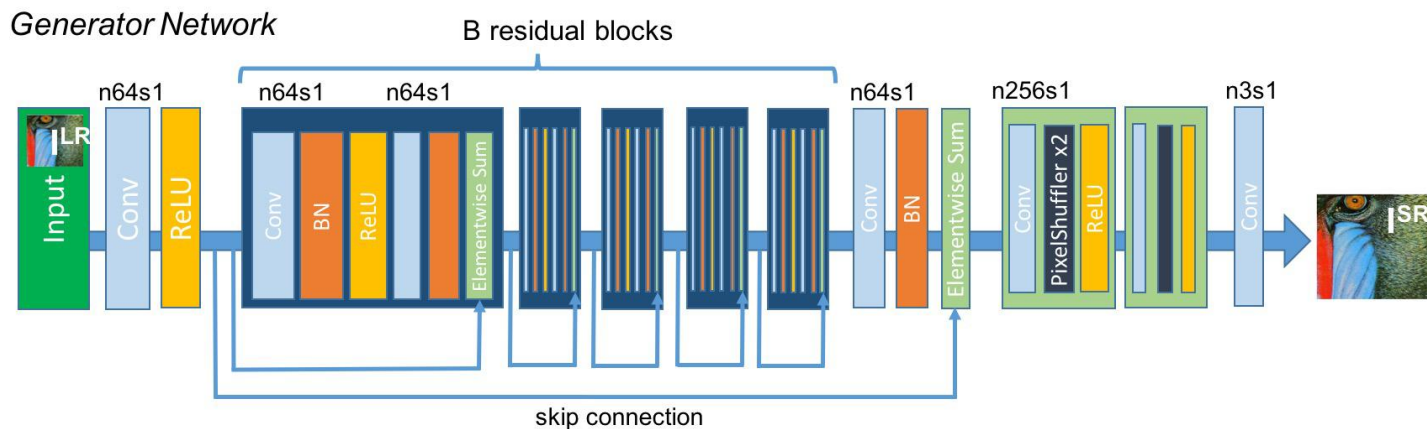


主要研究问题

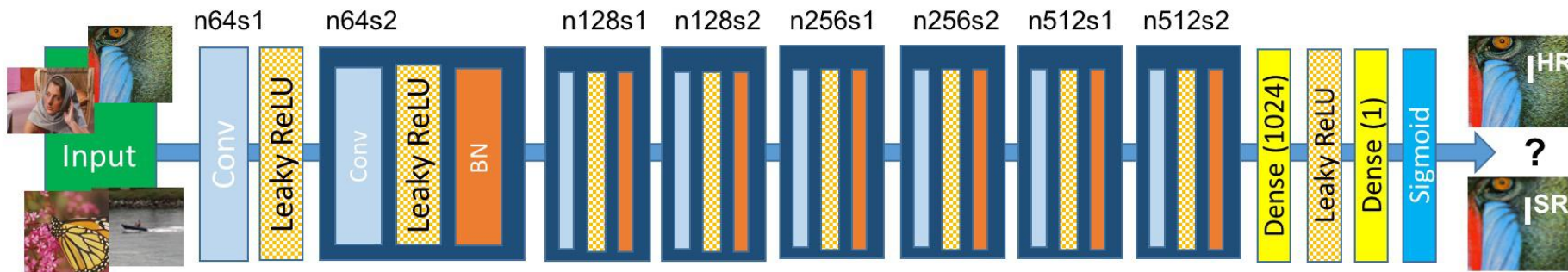
- 生成对抗网络GAN
 - 网络结构
 - 生成器网络（ Generator ）
 - 判别器网络（ Discriminator ）
 - 网络改进
 - GAN→CGAN → DCGAN→ wGAN
 - SRGAN
 - 应用范围：
 - 样本数据分布（生成）学习
 - 有监督问题的优化

主要研究问题

- 超分辨SRGAN网络

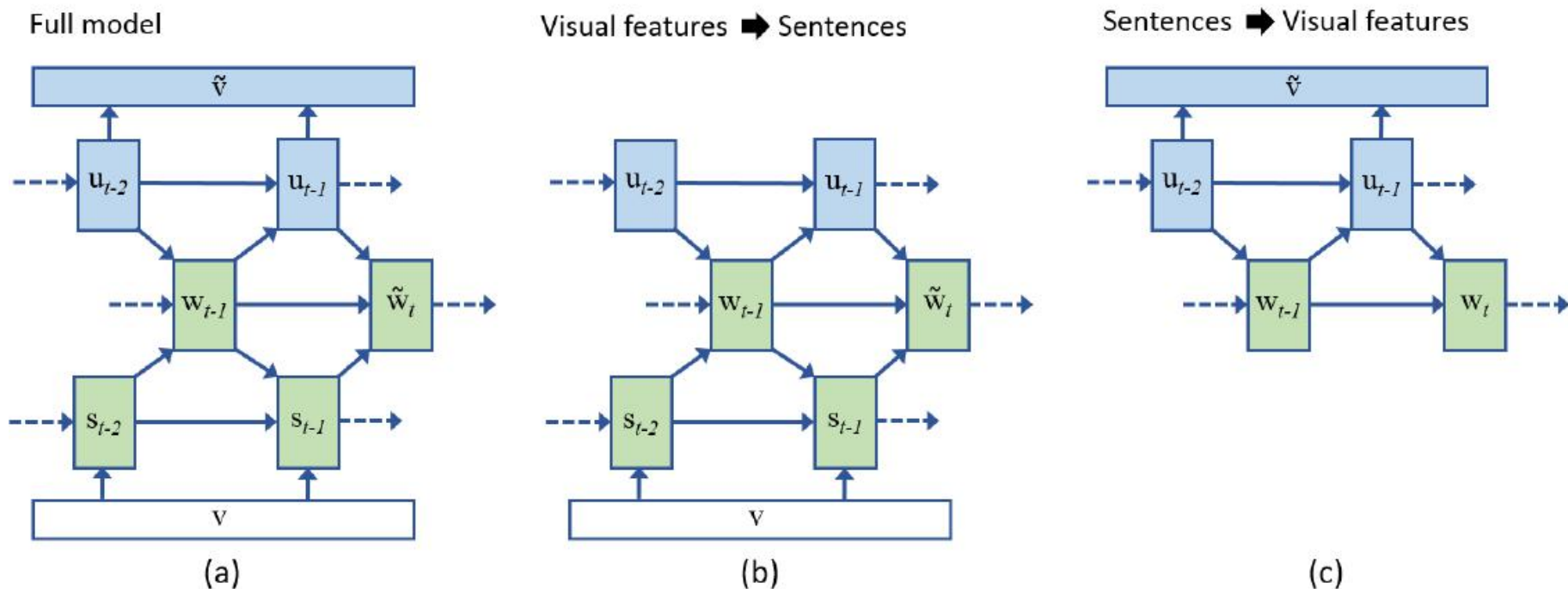


Discriminator Network



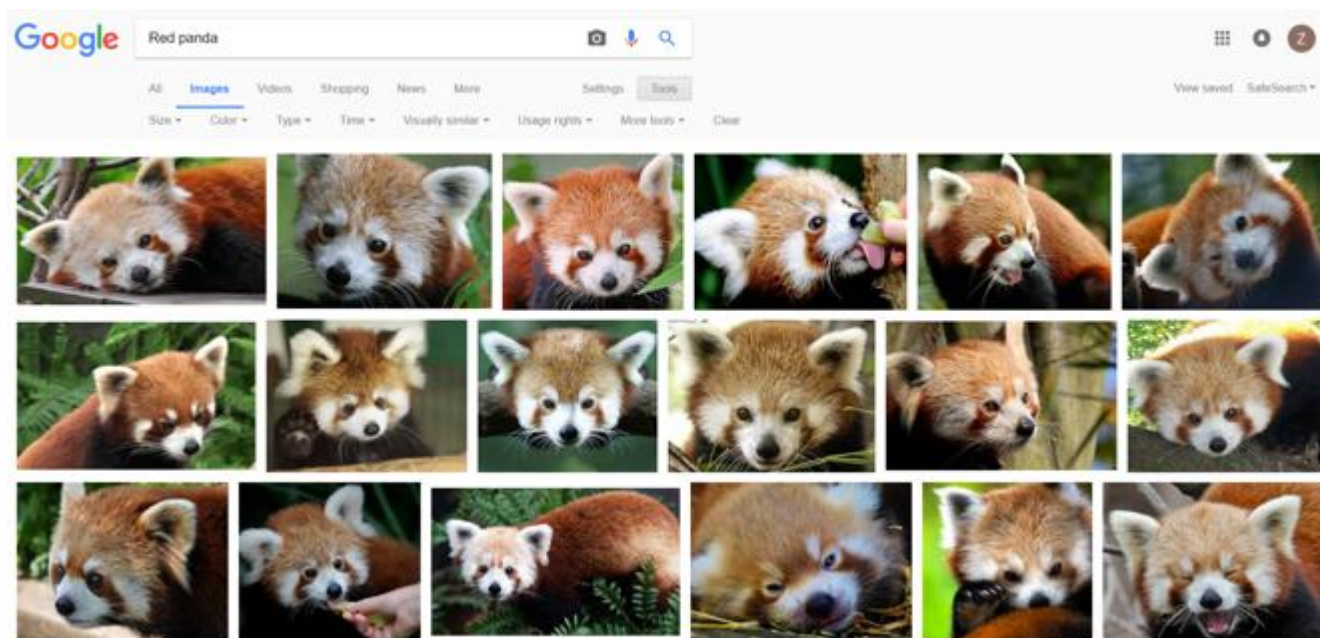
主要研究问题

- 视觉文本双向编码网络



主要研究问题

- 图片检索→相似图片



课程简介

☐ 重点&难点

- 各种深度神经网络模型（DNN）的理解
- Tensorflow的使用

☐ 实践环境

- Windows
 - ☐ OpenCV 3.4
 - ☐ Python 3.6+
 - ☐ TensorFlow 1.9+
 - ☐ Keras等

开源库介绍

- OpenCV
 - 函数库：计算机视觉，机器学习
 - 优化算法：2500+
 - 编程语言：C, C++, Java, Python, Matlab
 - 系统：Windows, Linus, Mac OS, iOS, Android
 - 并行计算：CUDA, OpenCL
 - URL: <http://opencv.org/>

开源库介绍

- Caffe
 - 函数库：深度学习
 - 发布方：Berkeley Vision and Learning Center (BVLC)
 - 编程语言：C++, Python
 - 并行计算：CUDA
 - URL: <http://caffe.berkeleyvision.org>
 - 特色：
 - 在配置中定义模型并优化，不需要硬编码
 - 代码的可扩展性强
 - 运行速度非常快

开源库介绍

- Torch(PyTorch)
 - 函数库：深度学习
 - 发布方：Facebook
 - 编程语言：Lua(Python)
 - 并行计算：CUDA
 - URL: [https://torch.ch\(http://pytorch.org/\)](https://torch.ch(http://pytorch.org/))
 - 特色：
 - 使用动态计算图，提供了更快速的方法
 - 提供了运行在 GPU/CPU 之上、基础的张量操作库；
 - 支持共享内存的多进程并发（multiprocessing）库

开源库介绍

- Tensorflow
 - 函数库：深度学习
 - 发布方：Google
 - 编程语言：Python
 - 并行计算：CUDA
 - URL: <https://www.tensorflow.org/>
 - 特色：
 - 多机、多GPU分布式
 - 适用于超大模型，超大数据集
 - Tensorboard将训练过程可视化

开源库介绍

- Keras
 - 函数库：深度学习
 - 发布方：Franois Chollet
 - 编程语言：Python
 - URL: <https://keras.io>
 - 特色：
 - 兼容 Theano 和 Tensorflow 的深度学习高级包
 - 用户友好，高度模块化，易扩展性

课程应用案例环境

- 编程语言
 - Anaconda Python 3.6
- 函数库/计算框架
 - OpenCV 3.2+
 - Tensorflow 1.9+
- DNN模型实践
 - VGG、ResNet、
 - Faster R-CNN、YOLO
 - DeepLab、U-NET
 - Show-and-Tell
 - GAN

参考书

□ Python

- Python基础教程（第2版 修订版）

□ OpenCV

- 《OpenCV 3计算机视觉：Python语言实现》

□ TensorFlow

- 《TensorFlow实战》

□ 深度学习

- 《深度学习》（goodfellow的“花书”）
- 《深度学习轻松学 核心算法与视觉实践》

关于课程的说明

□ 课程包括：

- 详细介绍深度学习在CV领域的常见模型
- 提供模型的源码
- 讲解模型源码的关键内容以便理解

□ 能帮到您的：

- 了解计算机视觉和深度学习的基本概念
- 了解模型演化的历程及改进的思路
- 节省阅读原版论文或教材的时间
- 打一个很薄、但相对比较全面的基础

演示环节

- 安装配置
 - Python
 - OpenCV 3.2
 - TensorFlow
- 链接推荐
 - Windows安装教程
 - http://docs.opencv.org/3.2.0/d5/de5/tutorial_py_setup_in_windows.html
 - 学习教程
 - http://docs.opencv.org/3.2.0/d6/d00/tutorial_py_root.html

问答互动

在所报课的课程页面，

- 1、点击“全部问题”显示本课程所有学员提问的问题。
- 2、点击“提问”即可向该课程的老师 and 助教提问问题。



联系我们

小象学院：互联网新技术在线教育领航者

— 微信公众号：**小象学院**

