

# 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。

□ 课程详情请咨询

■ 微信公众号：小象

■ 新浪微博：ChinaHadoop



# 第1课 课程介绍

---

计算机视觉

Computer Vision

# 本章结构

---

- 课程简介
- 计算机视觉
- 主要研究问题
- 研究挑战
- 开源库介绍
- 环境搭建

# 课程简介

---

## □ 部分I：计算机视觉的基础

- 数据图像处理
- 图像特征及描述

## □ 部分II：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
- 图像问答 ← 迭代神经网络RNN
- 图像生成 ← 生成对抗网络GAN

## □ 部分III：图像检索

# 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

- OpenCV 3.2
  - ☐ Windows
  - ☐ Python 2.7
- TensorFlow 1.1（GPU/CPU）
  - ☐ Windows
  - ☐ Python 3.5

# 课程简介

---

## ☐ 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》

☐ 《OpenCV3编程入门》

☐ 《学习OpenCV（中文版）》

### ■ TensorFlow

☐ 《TensorFlow：实战Google深度学习框架》

☐ 《TensorFlow实战》

# 课程简介

---

## □ 部分I：计算机视觉的基础

- Python
- OpenCV

## □ 部分II：深度学习在计算机视觉中的应用

- Python
- Tensorflow

## □ 部分III：图像检索

- Python
- OpenCV/Lire

# 计算机视觉

---

图片信息量有多大？→ 100个词





# 计算机视觉

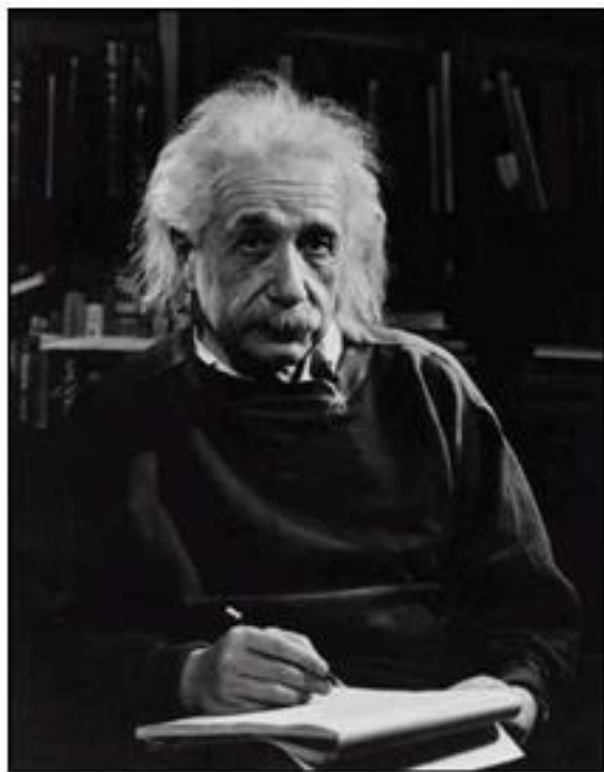
图片信息量有多大？ → 10,000 个词



# 计算机视觉

---

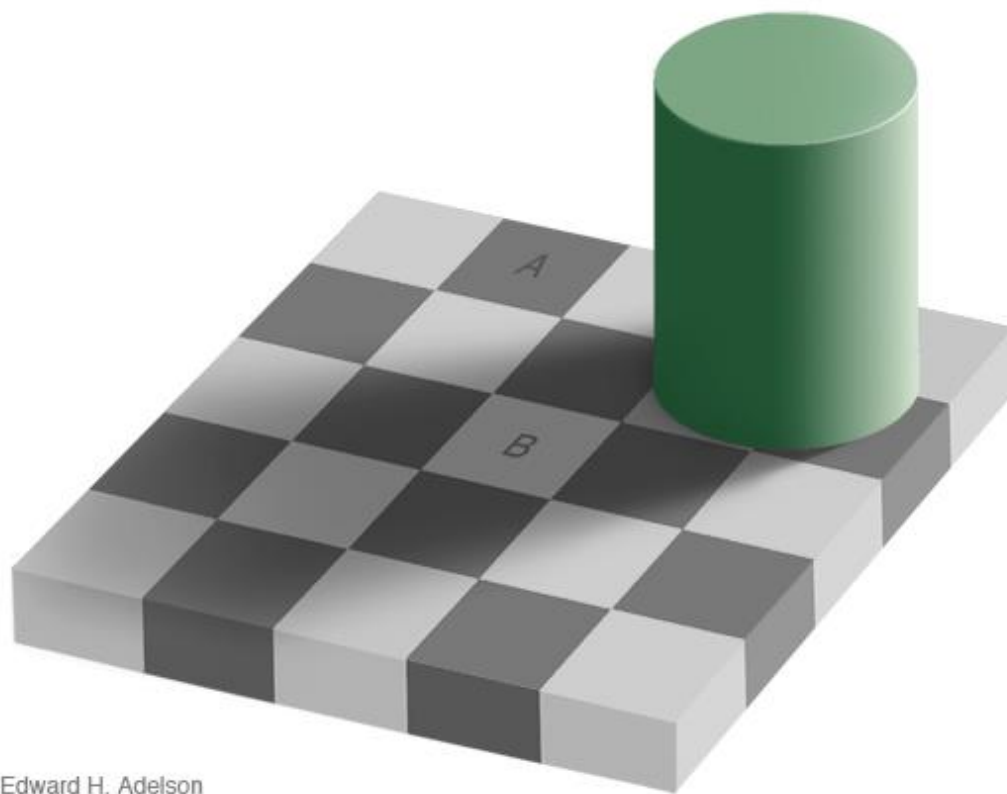
图片信息量有多大？ $\rightarrow$  1,000,000个词



# 计算机视觉

---

图片信息量有多大？ → 立体空间



Edward H. Adelson

# 计算机视觉

---

研究理论和应用：

- 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
- .....

# 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、磨轮两可
- 不善于长时间稳定的执行同一个任务
- .....

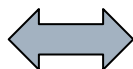
# 计算机视觉

图像数据  
(图片、视频、深度图片等)



解释

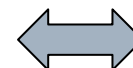
方向盘  
特斯拉  
地图  
无人驾驶  
.....



感知设备  
(摄像头)



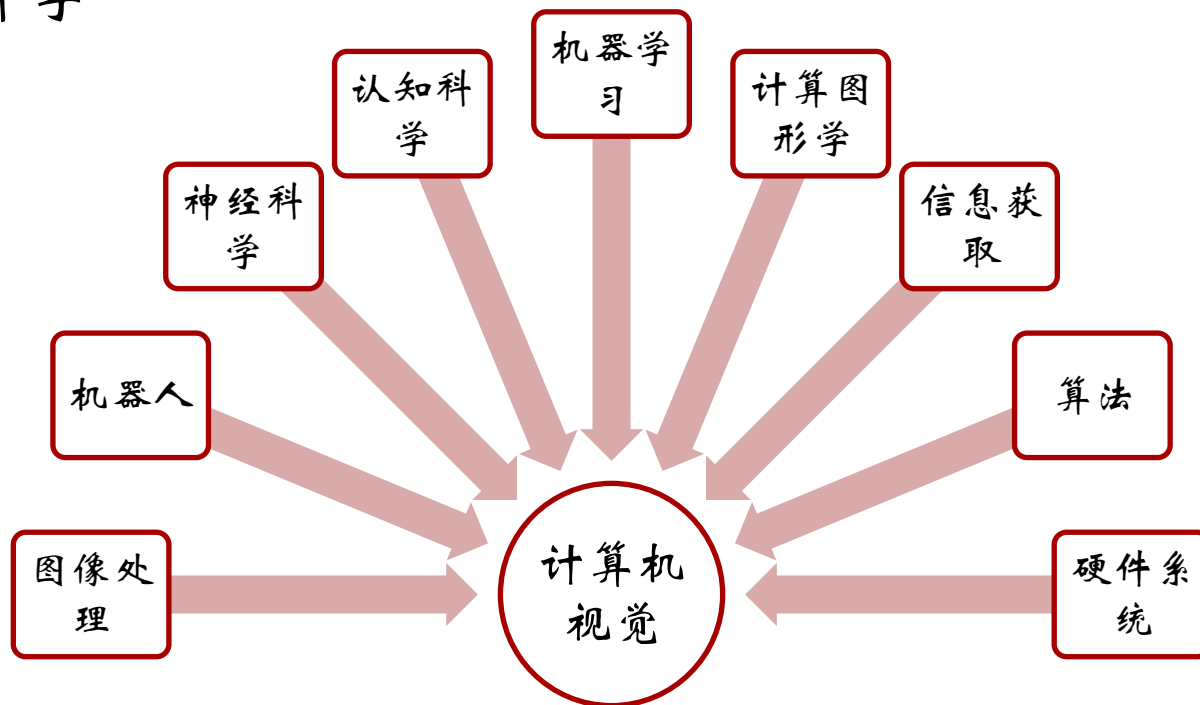
计算设备  
(CPU/GPU/FPGA)



# 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
- ...



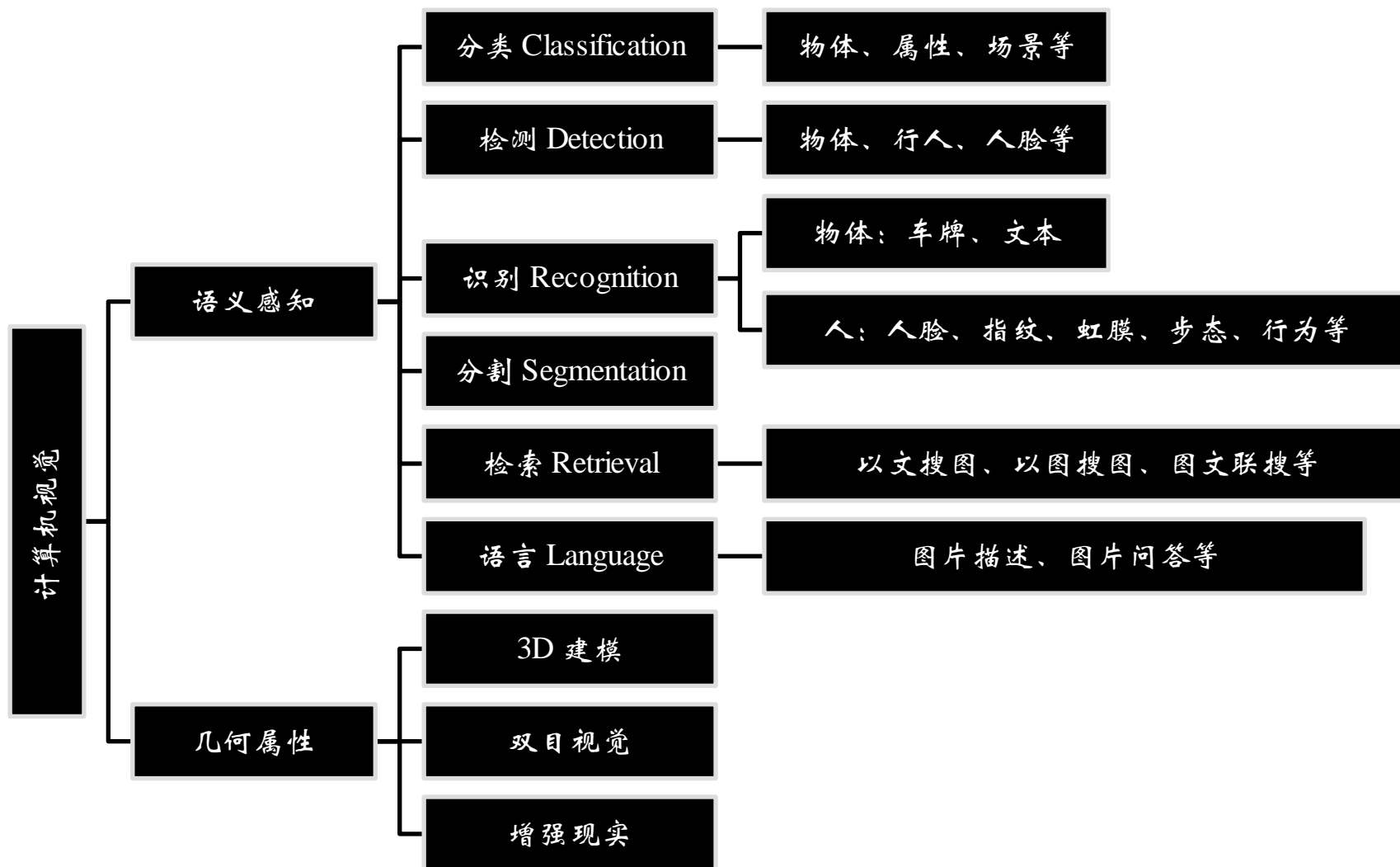


# 计算机视觉

---

- 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

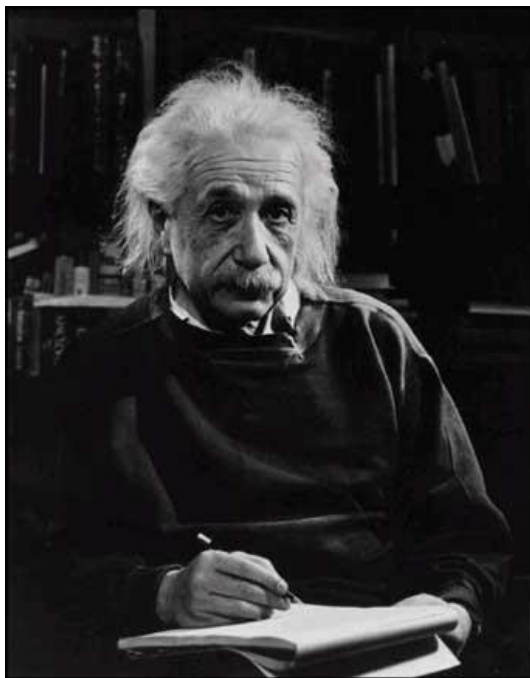
---

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

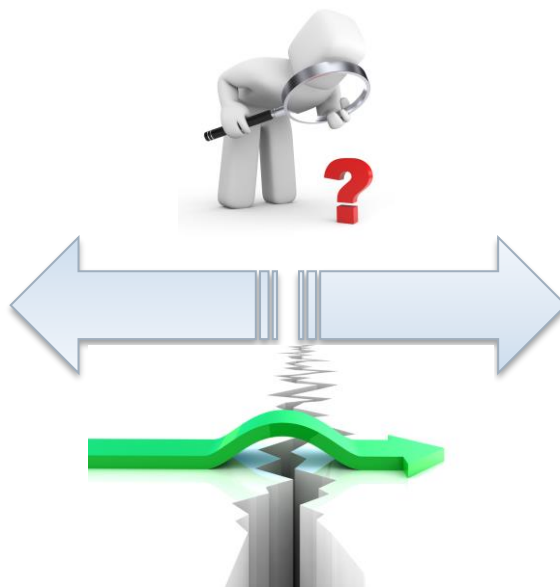
计算机

# 计算机视觉

- 人工智能目标
  - 解决”像素值”与”语义”之间的差距（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

# 主要研究问题

---

- 图像数据处理
  - 空域分析及变换
    - Sobel, 拉普拉斯, 高斯, 中值等
  - 频域分析及变换
    - 傅里叶 (Fourier) 变换
    - 小波 (Wavelet) 变换
  - 模板匹配, 金字塔, 滤波器组
  - 特征数据操作
    - 主成分分析/PCA
    - 奇异值分解/SVD
    - 聚类/Cluster

# 主要研究问题

---

- 图像特征及描述
  - 颜色特征
    - RGB, HSV, Lab等
    - 直方图
  - 几何特征
    - Edge, Corner, Blob等
  - 纹理特征
    - HOG, LBP, Gabor等
  - 局部特征
    - SIFT, SURF, FAST等

# 主要研究问题

- 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



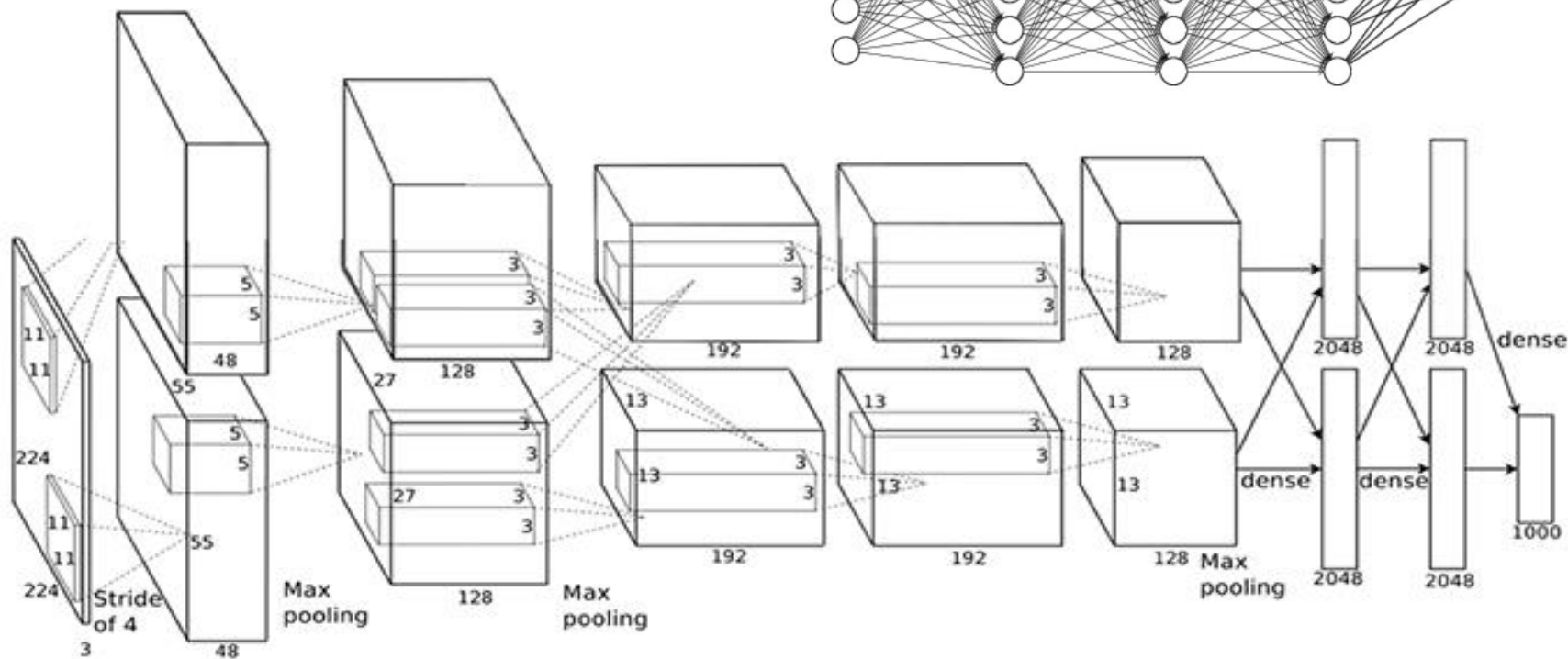
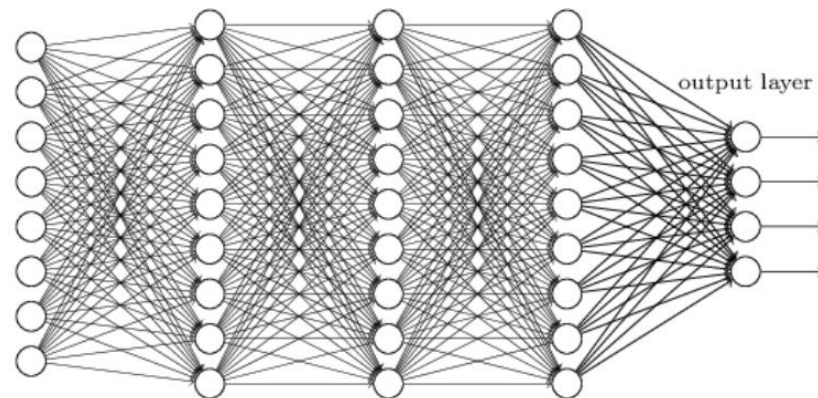
# 主要研究问题

---

- 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

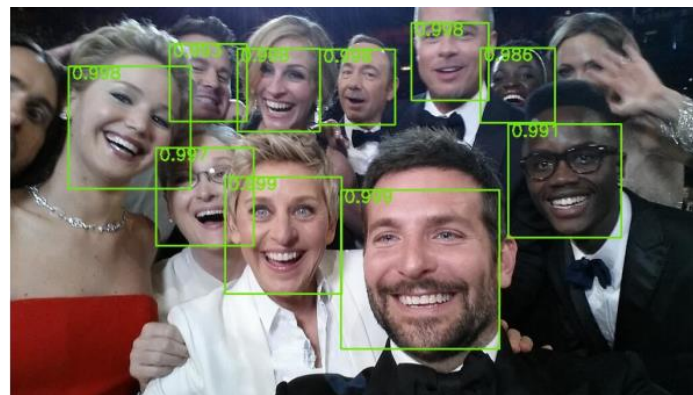
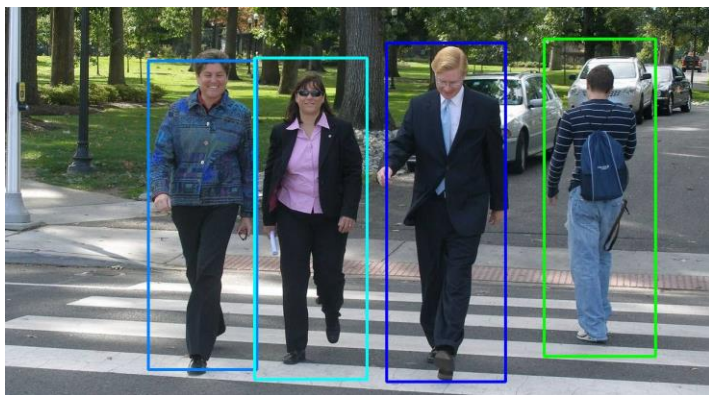
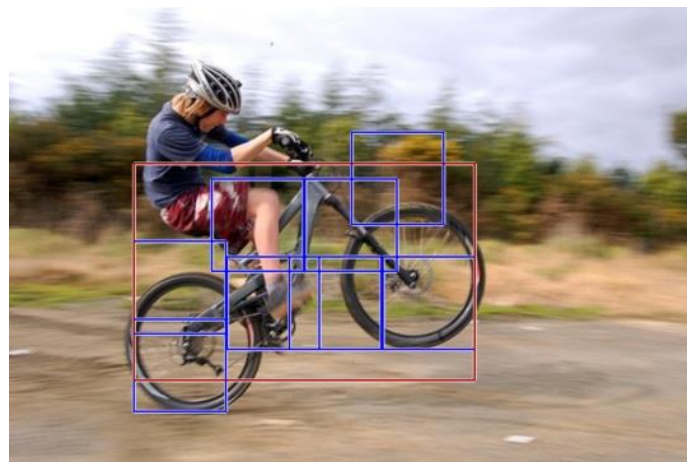
# 主要研究问题

## AlexNet卷积网络



# 主要研究问题

- 检测 → 目标在哪儿?
  - 物体检测
  - 人脸检测
  - 行人检测



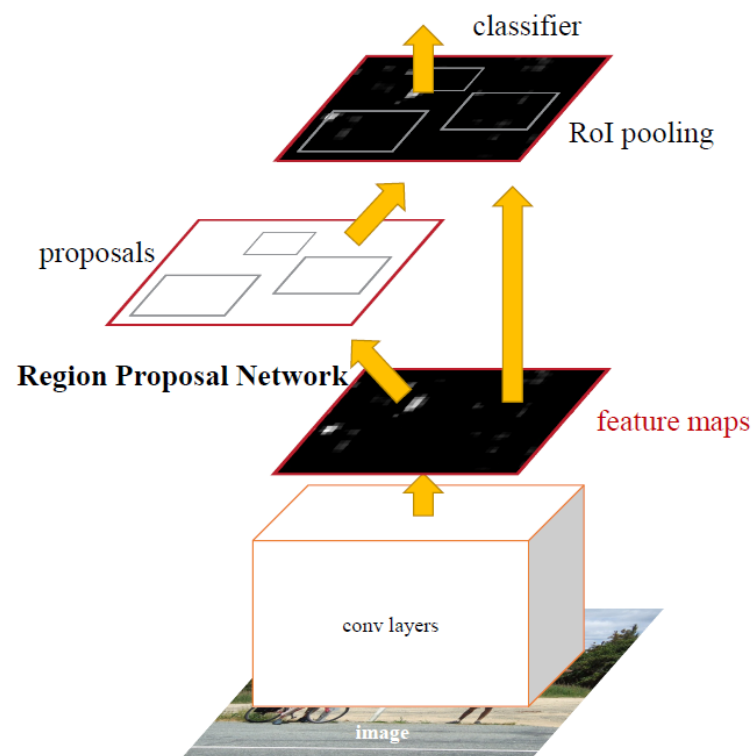
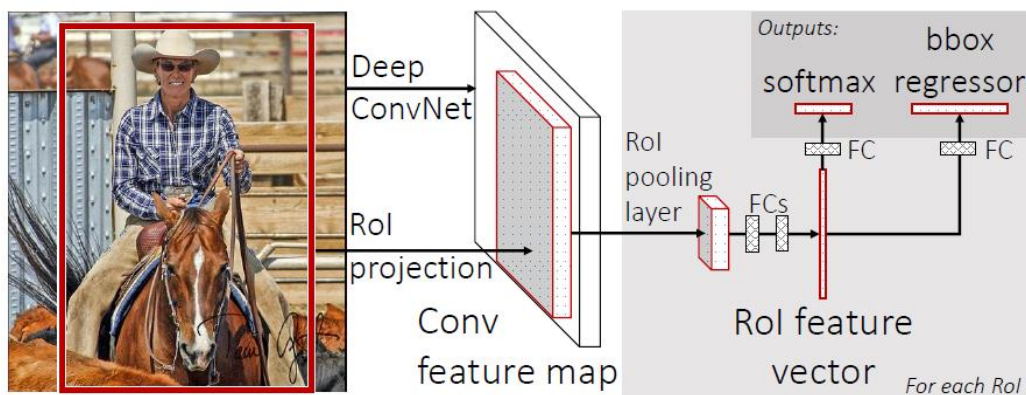
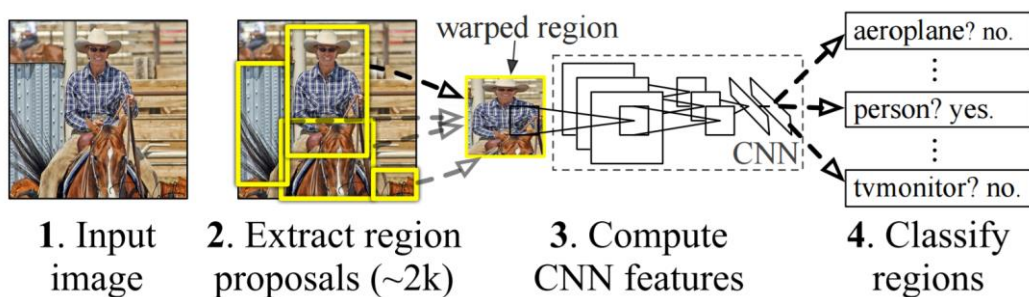
# 主要研究问题

---

- 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

- 图片分割 → 区域轮廓？
  - 前景分割
  - 语义分割



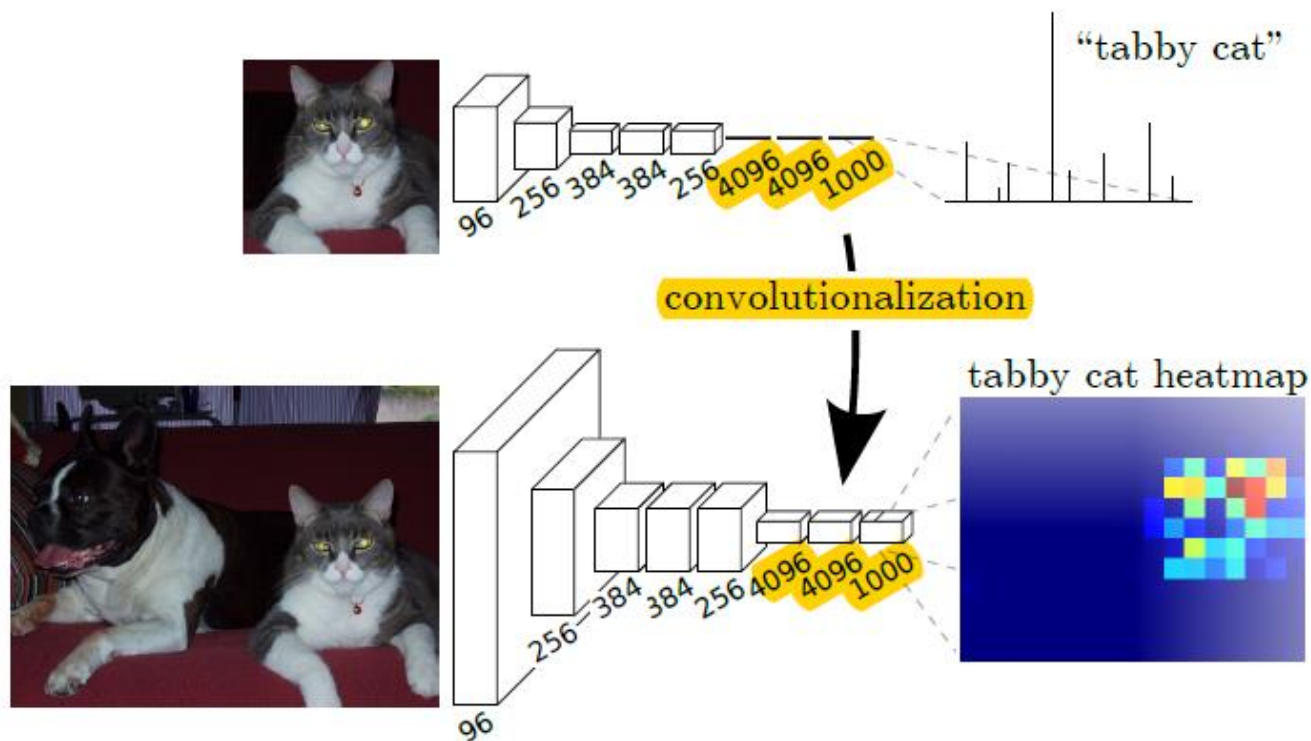
# 主要研究问题

---

- 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
    - .....

# 主流深度神经网络

- 全卷积网络FCN
  - 反卷积/转置卷积

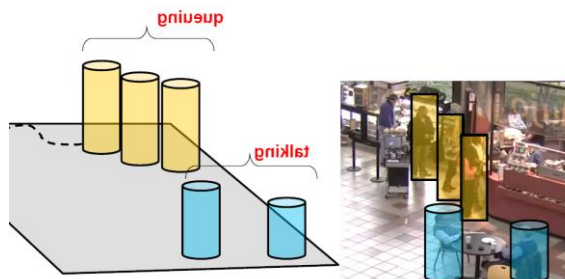
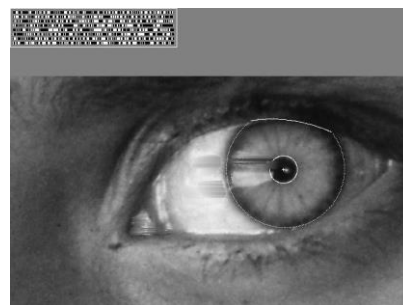




# 主要研究问题

- 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



- 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

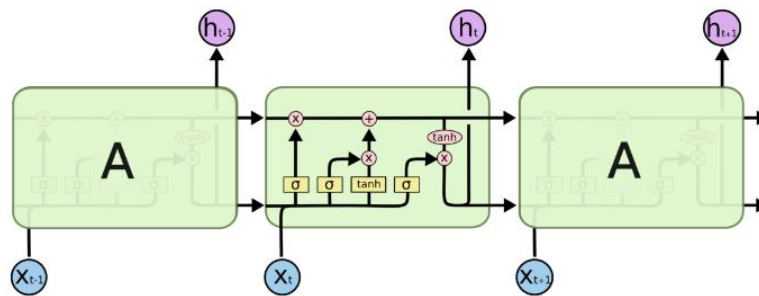
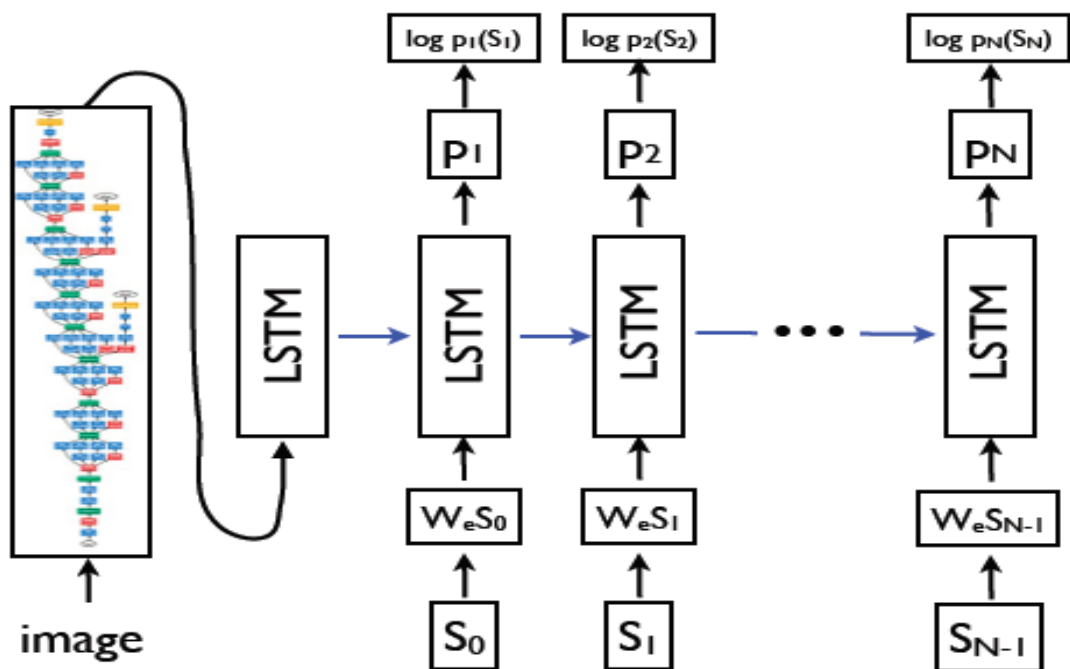
# 主要研究问题

---

- 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

bicubic  
(21.59dB/0.6423)



SRResNet  
(23.53dB/0.7832)



SRGAN  
(21.15dB/0.6868)



original



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



# 主要研究问题

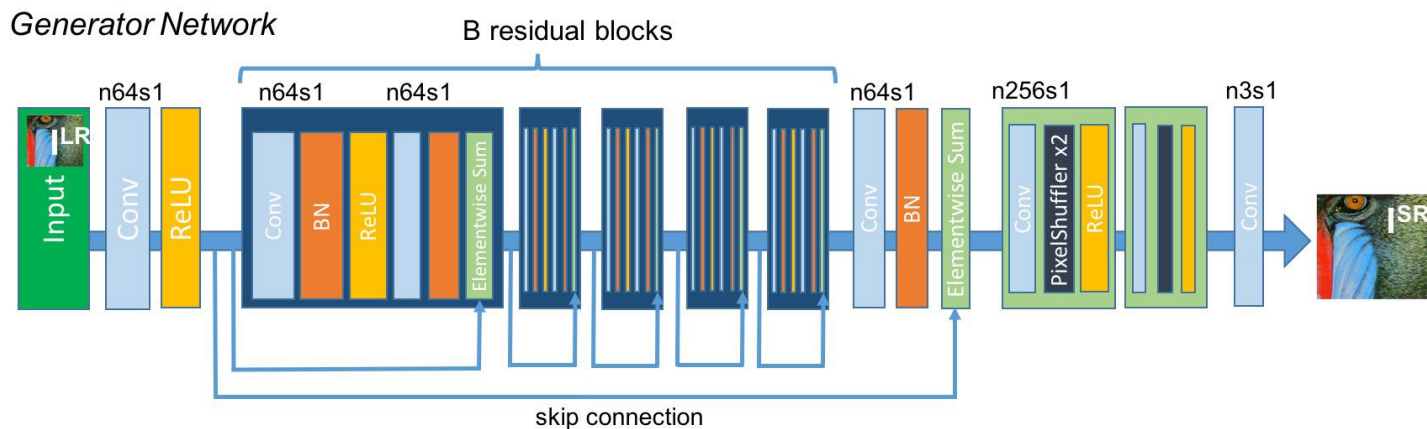
---

- 生成对抗网络GAN
  - 网络结构
    - 生成器网络 ( Generator )
    - 判别器网络 ( Discriminator )
  - 网络路线
    - 无监督:  $GAN \rightarrow DCGAN \rightarrow wGAN$
    - 有监督: SRGAN、SalGAN、RLA
  - 应用范围:
    - 样本数据分布 (生成) 学习
    - 半监督问题的数据增强
    - 有监督问题的优化 (Dynamic loss)

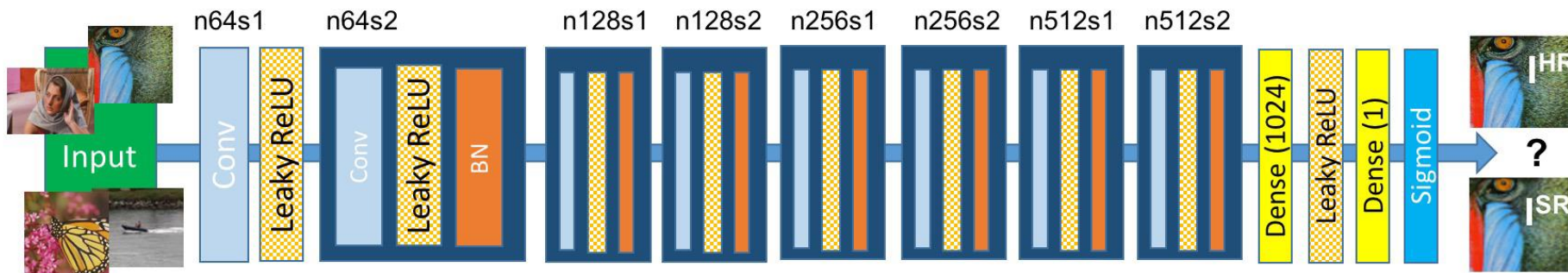


# 主要研究问题

- 超分辨率SRGAN网络

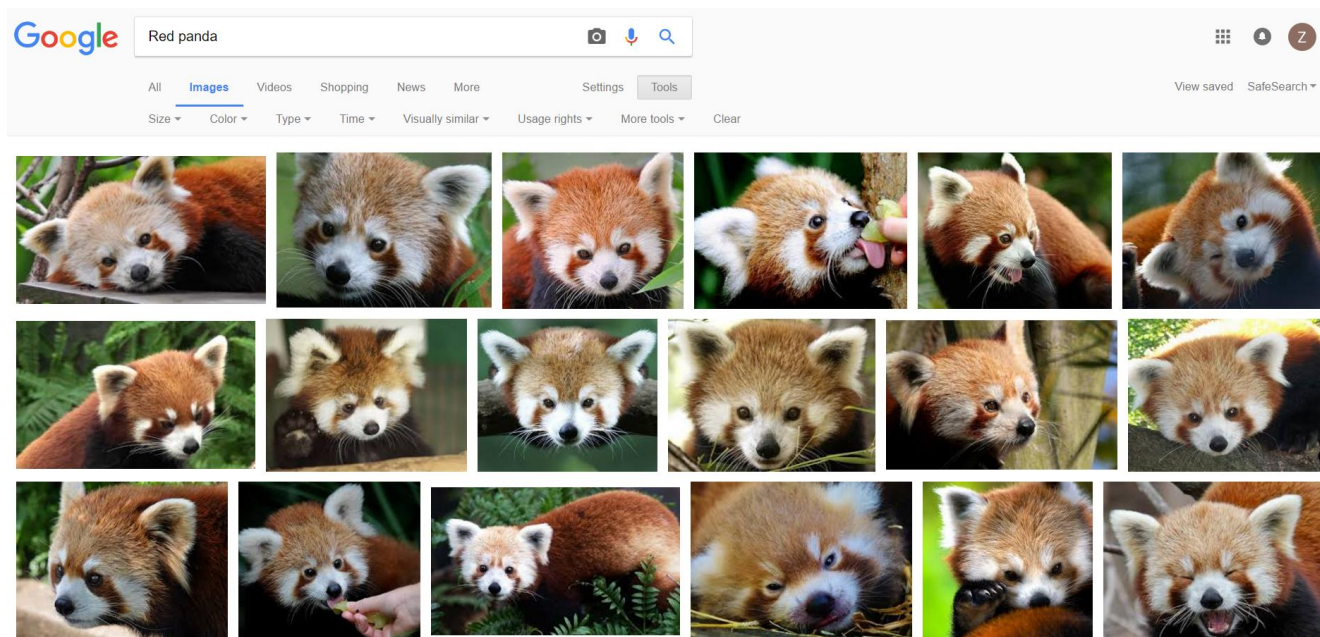


## Discriminator Network



# 主要研究问题

- 图片检索→相似图片





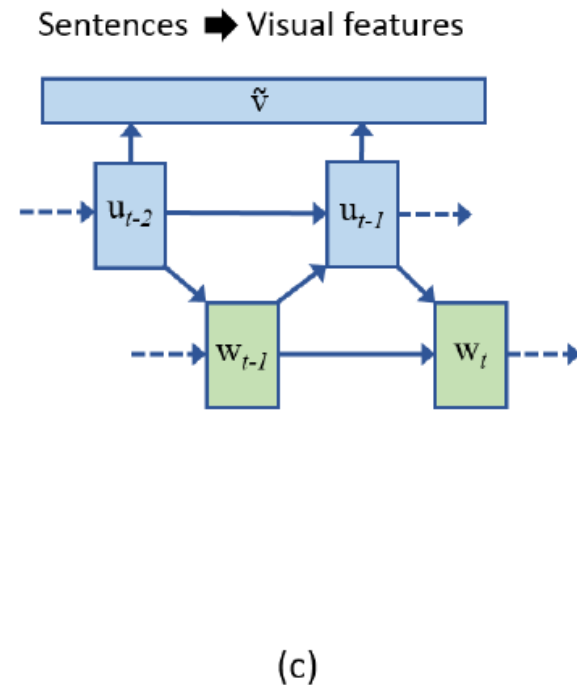
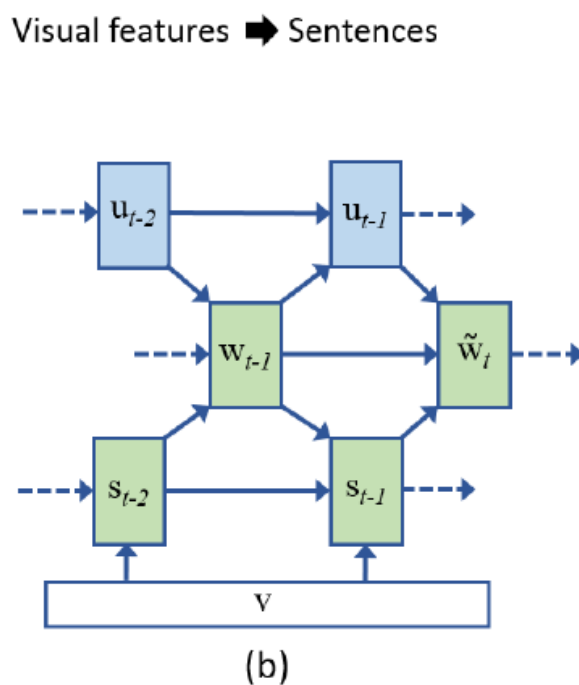
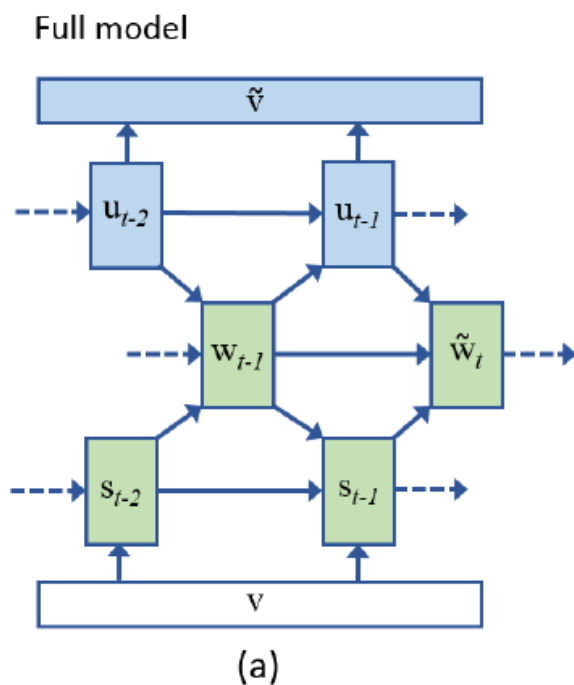
# 主要研究问题

---

- DNN网络的综合应用
  - 系统结构
    - 图片预处理
      - 显著性检测、区域分割
    - 图片特征提取
      - 手工设计、CNN类网络层特征抽取
    - 索引构建
      - 局部敏感性哈希 (LSH)
    - 特征相似度计算
      - 特征向量的距离计算
    - 相似度搜索
      - 相似距离排序

# 主要研究问题

- 视觉文本双向编码网络



# 研究挑战

---

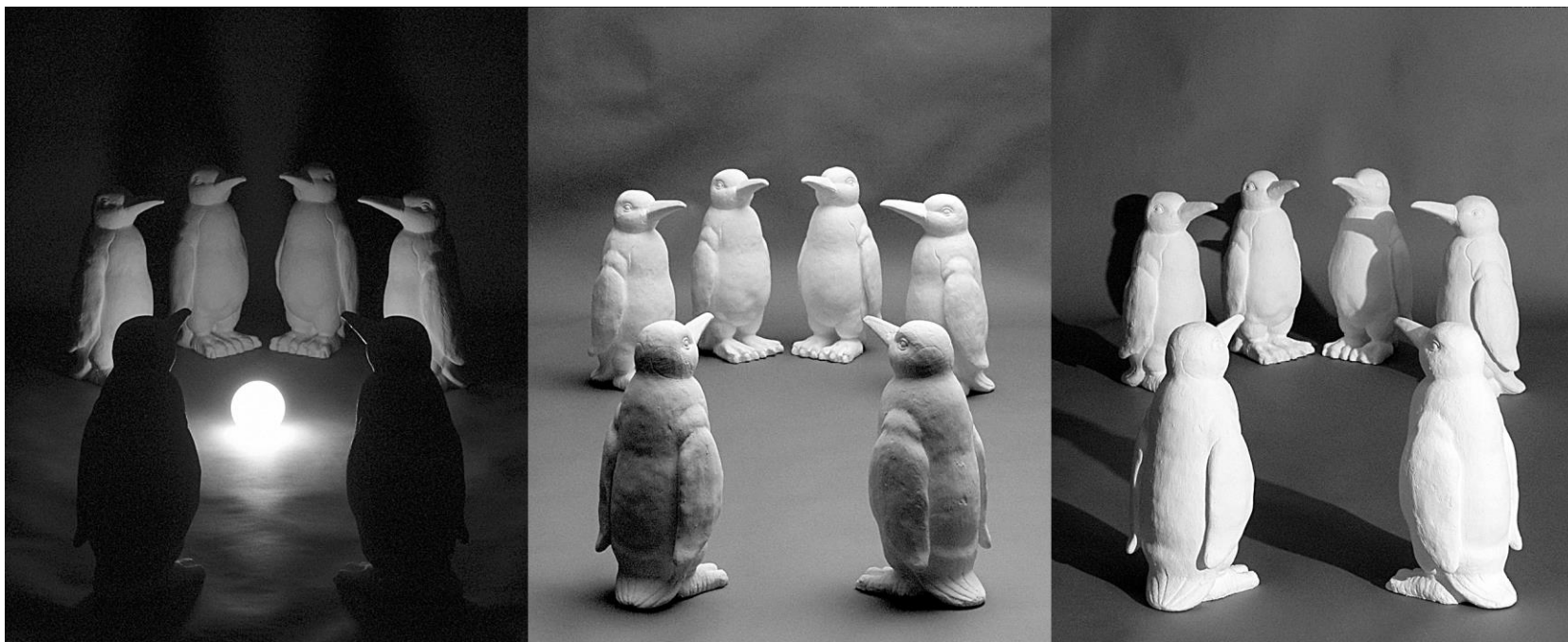
- 视角变化



# 研究挑战

---

- 光照变化



# 研究挑战

---

- 尺度变化



# 研究挑战

---

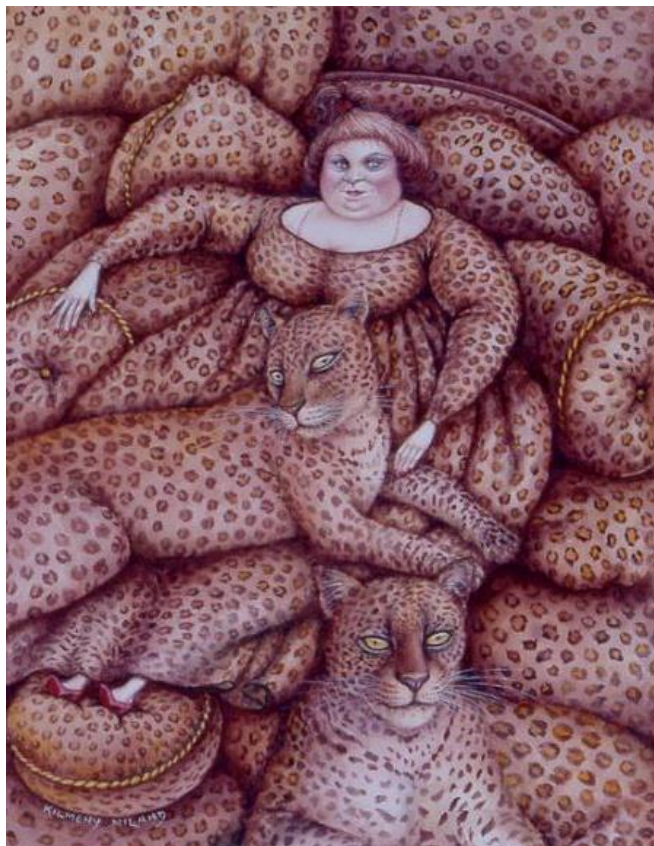
- 形态变化





# 研究挑战

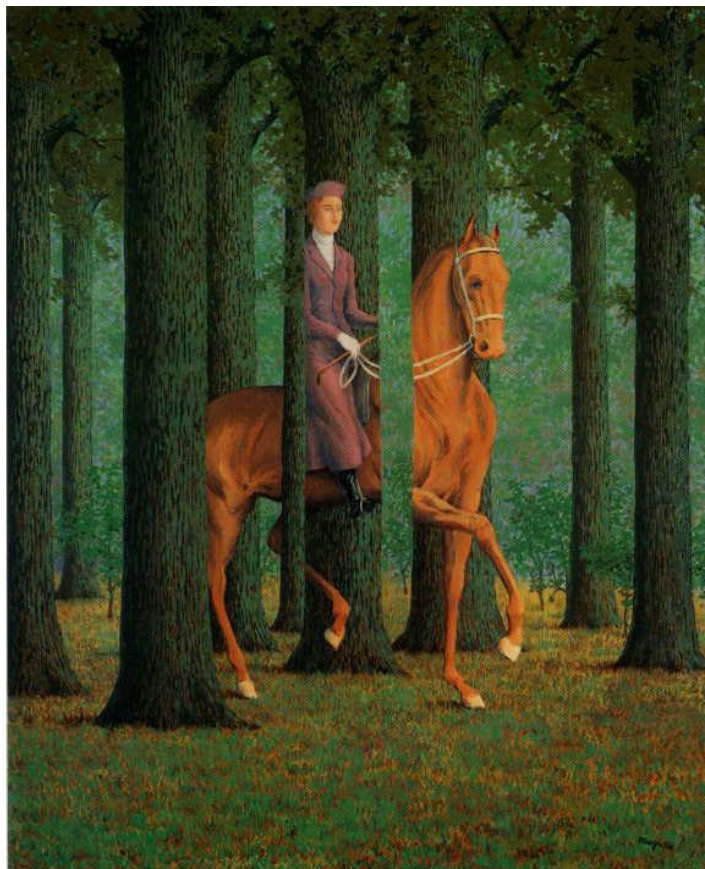
- 背景混淆干扰



# 研究挑战

---

- 遮挡





# 研究挑战

---

- 类内物体的外观差异



# 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
  - 系统：Ubuntu, Red Hat, OS X , Windows
  - 并行计算：CUDA
  - URL: <http://caffe.berkeleyvision.org>

# 开源库介绍

---

- Theano
  - 函数库：深度学习
  - 发布方：LISA Lab, 蒙特利尔工程学院
  - 编程语言：Lua(Python)
  - 系统：Ubuntu, Red Hat, OS X, Windows
  - 并行计算：CUDA
  - URL: <http://deeplearning.net/software/theano/>

# 开源库介绍

---

- Torch(PyTorch)
  - 函数库：深度学习
  - 发布方：Facebook
  - 编程语言：Lua(Python)
  - 系统：Ubuntu, Red Hat, OS X
  - 并行计算：CUDA
  - URL: [https://torch.ch\(http://pytorch.org/\)](https://torch.ch(http://pytorch.org/))

# 开源库介绍

---

- Tensorflow
  - 函数库：深度学习
  - 发布方：Google
  - 编程语言：Python
  - 系统：Ubuntu, Red Hat, OS X
  - 并行计算：CUDA
  - URL: <https://www.tensorflow.org/>
  - 特色：
    - 多机多GPU分布式
    - Tensorboard

# 深度学习库对比

深度框架	编程语言 /灵活性	主要应用场景	维护方	Pre-trained 模型	多GPU 模型并行	多机 分布式
(Py)Torch	Lua(Python) /Yes +	CNN/RNN(中等)/... 通用	Facebook	Yes ++	Yes	No
Theano	Python /Yes	CNN/RNN(支持)/... 通用	LISA Lab	Yes (Lasagne)	Tesing	No
Caffe	Python&C++ /No	CNN 图像分类、检测、分割	BVLC	Yes ++	No	No
<b>TensorFlow</b>	<b>Python /Yes +</b>	<b>CNN/RNN(Best)/... 通用</b>	<b>Google</b>	<b>Yes (Inception)</b>	<b>Best</b>	<b>Yes</b>

# 课程应用案例环境

---

- 编程语言
  - Python 2.7
  - Anaconda Python 3.5
- 函数库/计算框架
  - OpenCV 3.2
  - Tensorflow 1.1
- DNN模型实践
  - ResNet、Faster R-CNN、DeepLab
  - Show-and-Tell、VQA
  - SRGAN



# 演示环节

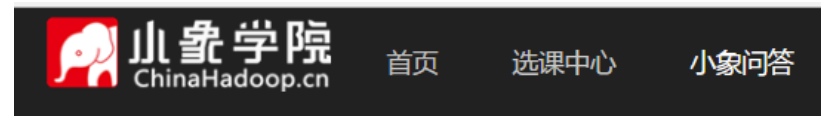
---

- 安装配置
  - Python
  - OpenCV 3.2
- 链接推荐
  - Windows 安装教程
    - [http://docs.opencv.org/3.2.0/d5/de5/tutorial\\_py\\_setup\\_in\\_windows.html](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](http://docs.opencv.org/3.2.0/d6/d00/tutorial_py_root.html)

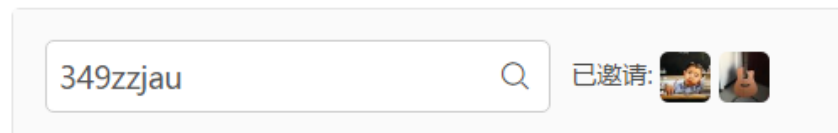
# 演示环节

- 小象问答

1. 登录[www.chinahadoop.cn](http://www.chinahadoop.cn)
2. 点击标题栏中的小象问答
3. 点击提问
4. 填写问题
5. 添加话题“计算机视觉”
6. 邀请老师：349zzjau



2017-05-12    添加评论    邀请 2    编辑    相关链接



# 疑问

---

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他人回复问题

## Q & A

小象账号：349zzjau

课程名：基于深度学习的计算机视觉

课后调查问卷：<http://cn.mikecrm.com/0D9JujS>

# 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
- 新浪微博：ChinaHadoop

