
机器学习绪论

陈飞宇

fchen@cqu.edu.cn

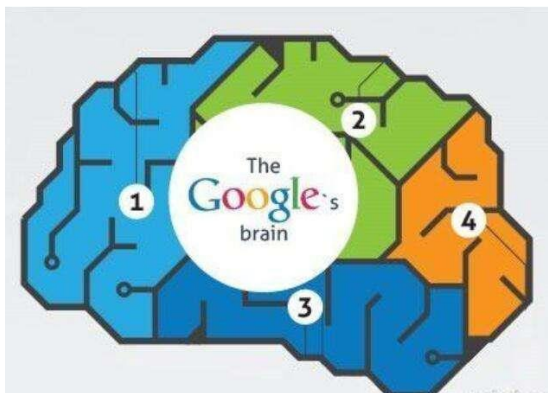
办公室：软件学院529室

Why Study Machine Learning?

- 时机成熟
 - GPU技术发展
 - 大数据
 - 云计算
 - 丰富的机器学习方法积淀

Why Study Machine Learning?

- 人工智能浪潮



□ 政府：“人工智能”被写入《政府工作报告》！

□ 创业潮：云从科技、中科视拓、商汤科技、旷世科技、格林深瞳…

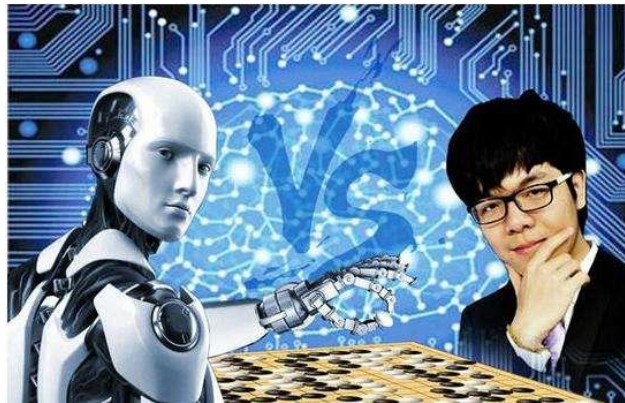
Why Study Machine Learning?

- 无限应用前景 (Better Engineering, more advanced Industry)

安防



机器人+游戏博弈



医疗



金融



交通



● ● ● ● ● ●

Why Study Machine Learning?

- 科学研究
 - 认知科学
 - 仿生学
 - 生物学
 - 医学
 - 工程学
 - 金融学
 -



Why Study Machine Learning?

● 人才需求

中国人工智能人才缺口超500万 供求比例仅为1:10

2017-07-14 18:03

人工智能

公众号 | 深眼看经济

王兴军是一位数据挖掘高级工程师,按照时下大众的说法,他是最受青睐的人工智能人才,是传说中的“人工智能小圈子”里的一分子。在关闭招聘网站自己简历的情况下,王兴军每个月仍会收到4到5个猎头的电话,游说他去其它公司,但都被他谢绝。

人工智能人才缺口超500万 互联网企业不惜百万年薪挖人

2017-05-20 16:24

未来网(www.k618.cn)北京5月19日电(见习记者 王嘉佩)日前,全球移动互联网大会北京站开幕。卡内基梅隆大学计算机科学学院机器学习系主任 Tom Mitchell,就人工智能的发展趋势做了分析研究,人工智能人才成为了关注的焦点。

AI人才短缺薪资高,你准备好读人工智能专业了吗?

2017-08-30 15:14

这是一个属于人工智能的时代。当前,人工智能是一颗闪耀的“明星”,已经成为国际竞争的新焦点,世界多国都在加紧人工智能发展布局,以至于提到了战略高度的地位。日前,国务院发布了《新一代人工智能发展规划》,对完善人工智能领域学科布局,设立人工智能专业,推动人工智能领域一级学科建设提出了指导意见。

2018 届互联网校招高薪清单

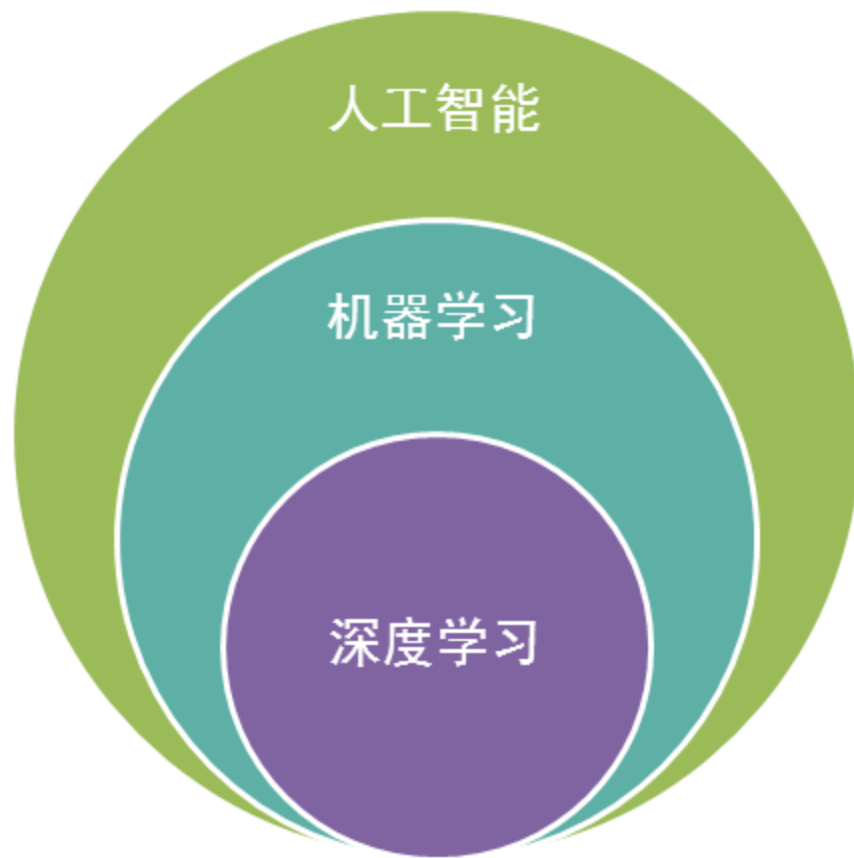
公司	岗位	年薪(人民币)
谷歌中国	人工智能(ssp)	56w
微软	算法工程师	51w
google	算法工程师	50w
腾讯	基础应用研究(ssp)	45-50W
腾讯	腾讯云后台研发工程师	32.4w
大疆	算法工程师	30w-40w
海康	算法工程师	30w
华为	研发工程师(硕士ssp)	30w+
Works Applications	软件工程师	36w
网易游戏	游戏研发工程师	32.8w
网易游戏	算法工程师(sp)	36w+
阿里巴巴	算法工程师	38.4
美团	研发工程师(sp)	30w
大疆	后台开发工程师	30-40w
OPPO	AI视觉(sp)	31-32w
滴滴	算法工程师(新锐)	30.4w-37.5W
51信用卡	Java工程师(sp)	30w
好未来	数据挖掘工程师	35w
百度	算法工程师	32W
今日头条	算法工程师(博士sp)	33w
28所	论证设计师	32W+11W安家费
京东	算法工程师(ssp)	31w
亚马逊	软件工程师SDE	31w+股票
深信服	软件研发(ssp)	30w

数据来源于校招薪水公众号

Why Study Machine Learning?

排序方式： 默认 最新	月薪： 不限	工作性质： 不限
高级机器学习 [北京·苏州街] 11:07发布	25k-40k 经验不限 / 本科	深度学习 自然语言处理
机器学习 [北京·西直门] 09:50发布	30k-50k 经验3-5年 / 本科	企业服务 搜索 算法 推荐 NLP
机器学习/数据挖掘工.. [北京·上地] 2017-10-30	20k-40k 经验1-3年 / 本科	算法 深度学习 推荐 NLP 人工智能
机器学习工程师 [深圳·科技园] 1天前发布	20k-40k 经验1-3年 / 本科	算法 深度学习 自然语言处理
机器学习 [北京·朝阳区] 1天前发布	18k-30k 经验3-5年 / 本科	hadoop 算法 数据挖掘 深度学习 大数据
排序方式： 默认 最新	月薪： 不限	工作性质： 不限
软件工程师 [深圳·龙岗区] 08:38发布	10k-15k 经验3-5年 / 本科	旅游 软件开发
系统软件工程师 [北京·五道口] 15:06发布	20k-40k 经验3-5年 / 本科	硬件制造 嵌入式 linux C++
软件工程师 [广州·大沙] 09:15发布	7k-13k 经验不限 / 本科	医疗健康 中级 C#/.NET Java c#
程序员 [厦门·莲前] 1天前发布	8k-15k 经验3-5年 / 本科	企业服务 PHP
开发工程师 [成都·锦江区] 10:28发布	6k-10k 经验3-5年 / 大专	微信

What is Machine Learning?

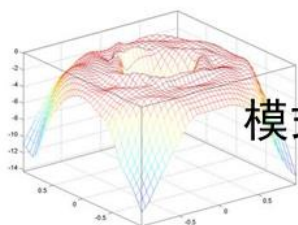


机器学习是人工智能的一个子领域，是人工智能的核心。

深度学习是机器学习的一个子领域，是目前最火的方向。

机器学习是从数据通往智能的技术途径，是现代人工智能的本质。

What is Machine Learning?



模式识别

计算机视觉



数据挖掘



机器学习

语音识别



统计学习



自然语言处理

What is Machine Learning?

“Learning is any process by which a system improves performance from experience.” -- Herbert Simon

Machine learning aims to study **Approaches** which improve the **Performance** of a machine at a specific **Task** with **Experiences**.

机器学习 = 任务 + 方法 + 经验 + 性能

What is Machine Learning?

- 任务-T: 机器学习要解决的问题
- 方法-A: 各种机器学习方法
- 经验-E: 训练模型的数据, 实例
- 性能-P: 方法针对任务的性能评估准则

任务是机器学习的研究对象;

方法是学习的核心内容;

经验是学习的动力源泉;

性能是学习的检验指标。

机器学习例子

- T: 人脸识别
 - A: 线性回归
 - E: 已标定身份的人脸图片数据
 - P: 人脸识别准确率
- T: 象棋博弈
 - A: 人工神经网络
 - E: 指令化棋谱
 - P: 对随机对手的获胜比率
- T: 股价预测
 - A: 多项式回归
 - E: 不同股票近三年各交易日股价数据
 - P: 估价误差（方差）

Tasks

- 机器学习基本任务：
 - 回归 (Regression)
 - 分类 (Classification)
 - 聚类 (Clustering)
 - 表征 (Representation)

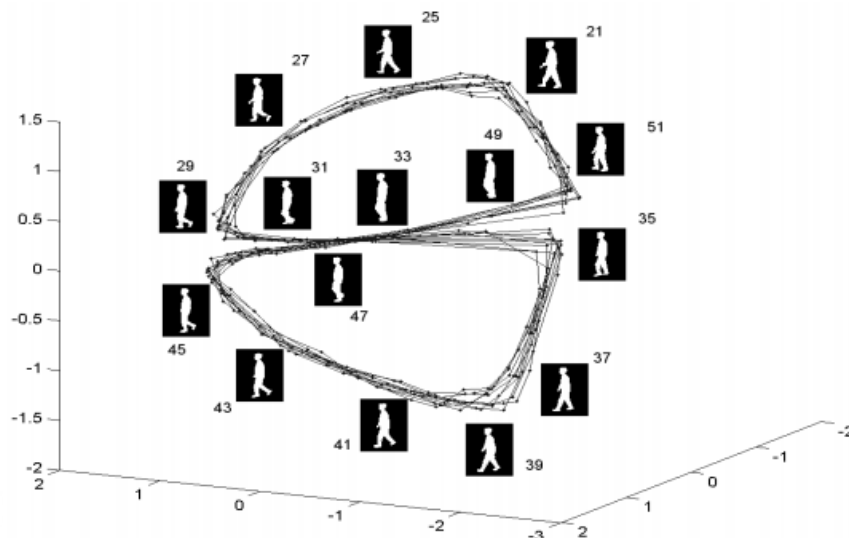
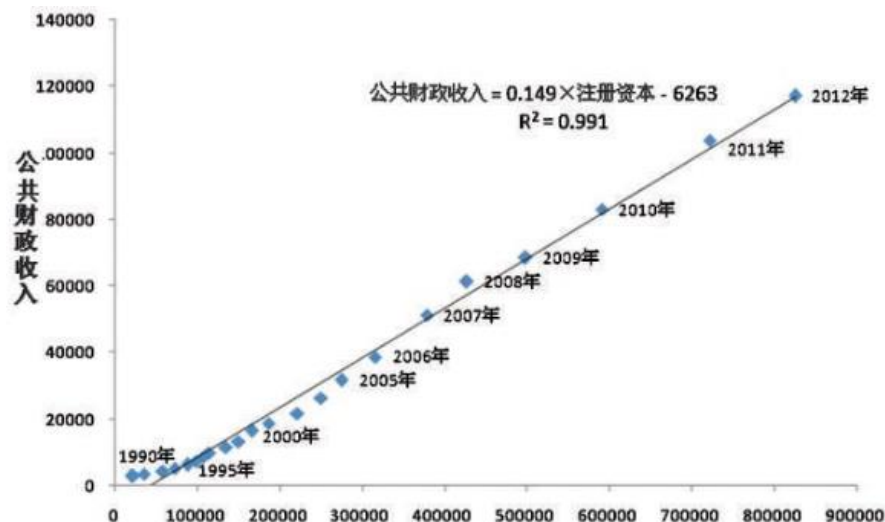
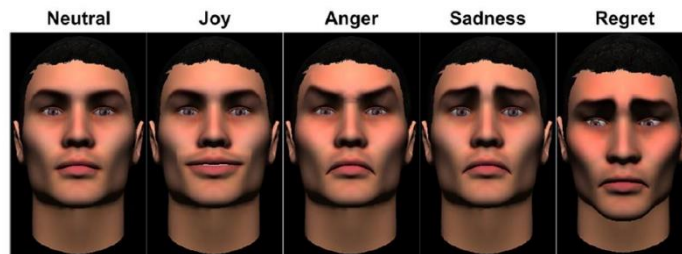
Tasks

- 回归(Regression):

- 能解决分类、预测问题

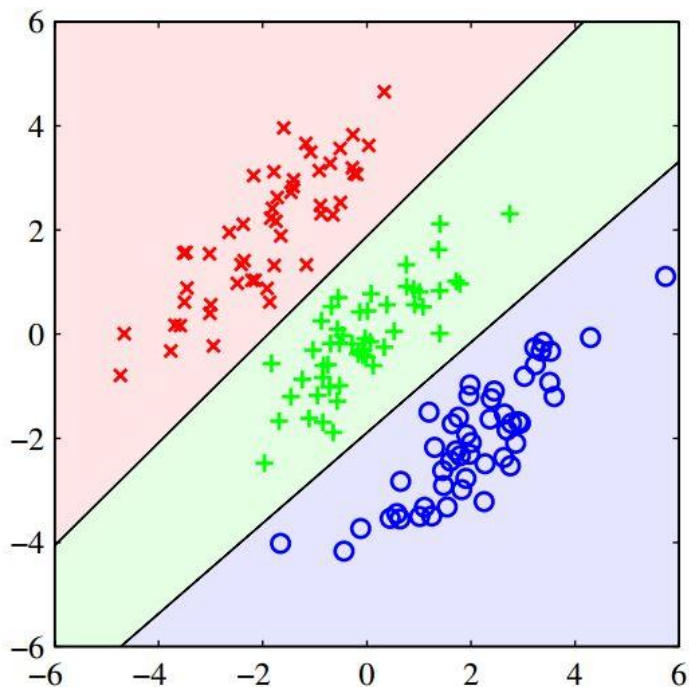
- 年龄估计

- 形状分析、表情分析、运动分析等



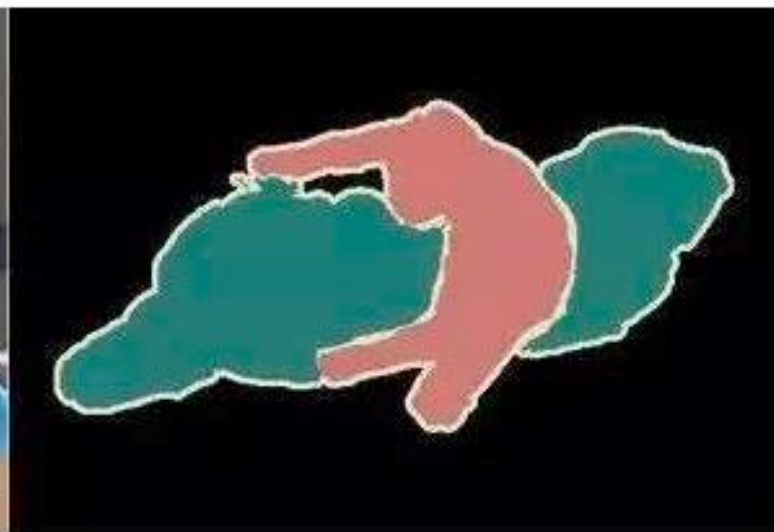
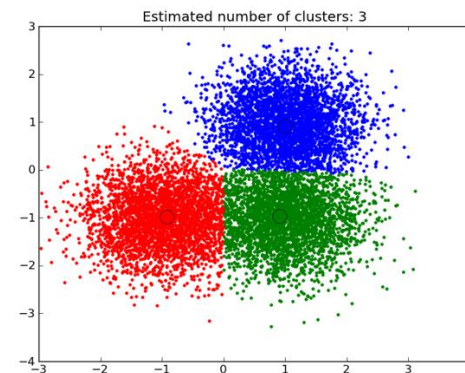
Tasks

- 分类(Classification):
 - 分类（图像、视频、文本.....）
 - 识别（语音、人脸、指纹.....）
 - 检测（行人、车辆、军事目标.....）



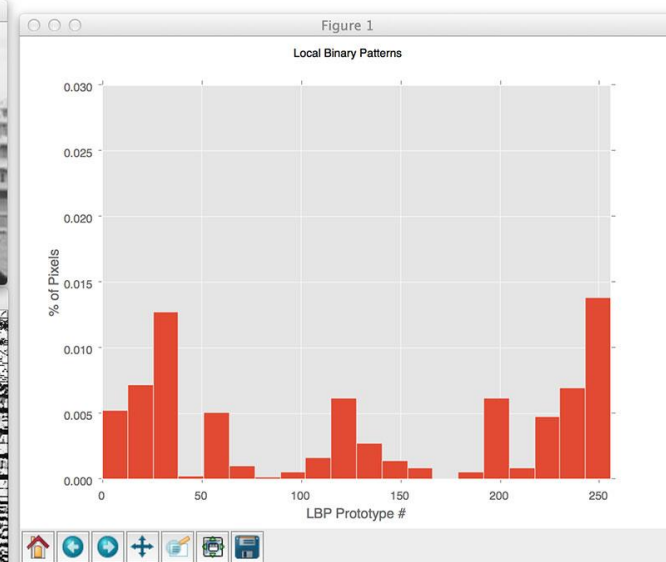
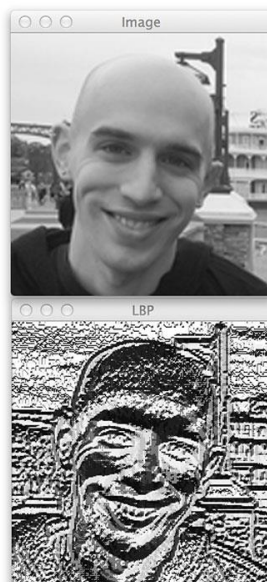
Tasks

- 聚类(Clustering):
 - 分割（图像、视频）、背景建模
 - 数据挖掘、数据恢复
 - 字典学习（视觉信息，文本）



Tasks

- 表征(Representation):
 - 特征提取（便于前三个任务解决）
 - 数据重构
 - 信息检索



Approaches

- 方法分类（根据学习形式）
 - 有监督学习（Supervised Learning）
 - 数据都有明确的标签，根据机器学习产生的模型可以将新数据分到一个明确的类或得到一个预测值。
 - 无监督学习（Unsupervised Learning）
 - 数据没有标签，机器学习出的模型是从数据中提取出来的模式（提取决定性特征或者聚类等）
 - 半监督学习（Semi-supervised Learning）
 - 部分数据有明确的标签，根据机器学习产生的模型可以将新数据分到一个明确的类或得到一个预测值。

Approaches

- 典型方法
 - Supervised Learning
 - 支持向量机、贝叶斯分类器、决策树、线性判别分析.....
 - Unsupervised Learning
 - K均值、Meanshift、主成分分析、典型相关分析.....
 - Semi-supervised Learning
 - 图直推学习、超图直推学习.....

部分机器学习算法可以通过设置采用以上任何一种学习形式！

Experience

- 数据例子1-人脸识别与人脸验证(face recognition & face verification)



Labeled faces in Wild (LFW)
<http://vis-www.cs.umass.edu/lfw/>

Experience

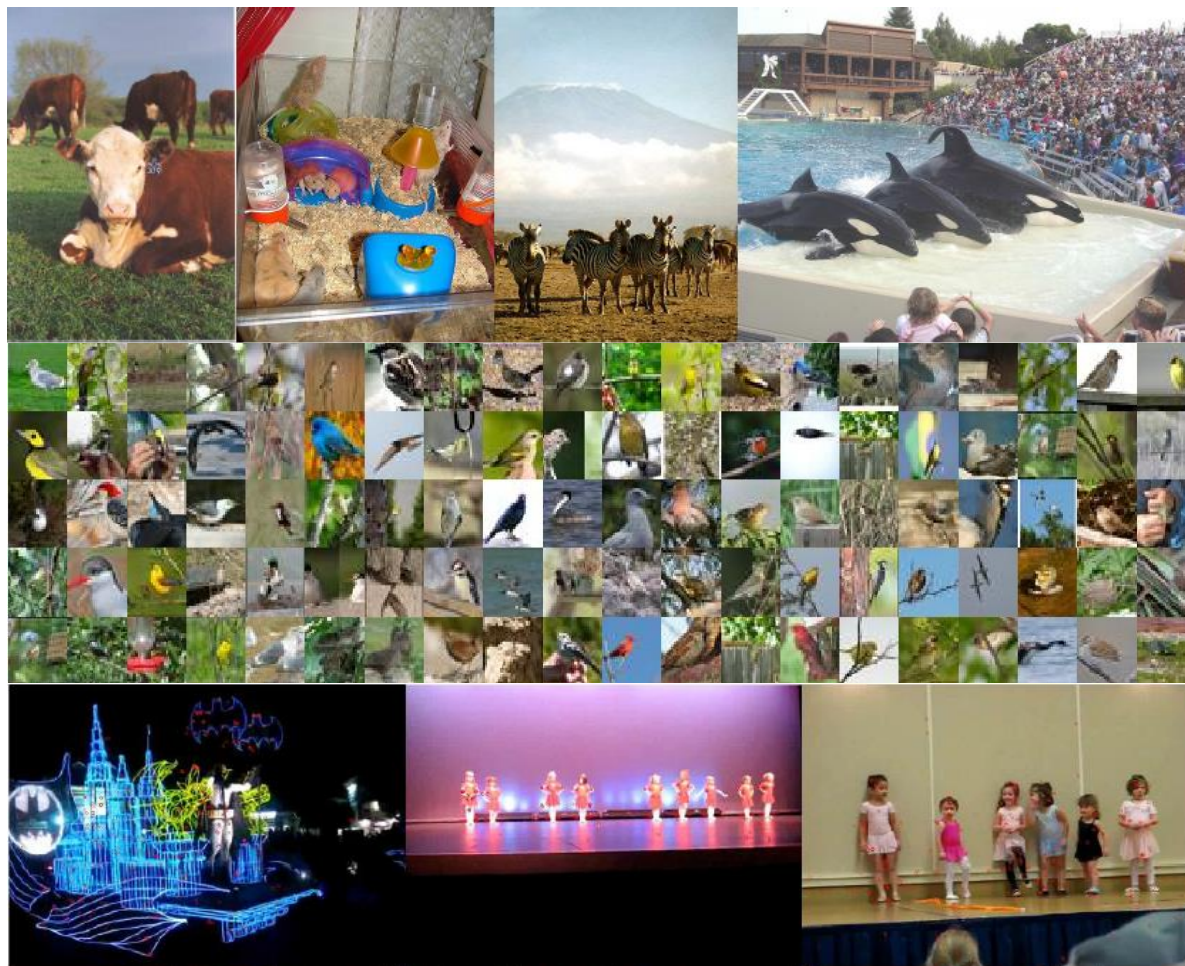
- 数据例子2-物体识别(Object Recognition)



ImageNet Dataset-100万+张的图片构成物体识别数据库,
<http://www.image-net.org/>

Experiences

- Other Examples:



AwA图像数据库，
用于属性学习（
Attribute Learning）

CUB图像数据库，用
于细粒度分类（Fine-
Grained Classification）

USAA视频数据库，用
于活动识别（Activity
Recognition）