

# 人工智能应用技术初步



机器智能与先进计算  
教育部重点实验室

郑伟诗

<http://www.isee-ai.cn/~zhwshi/>

中山大学

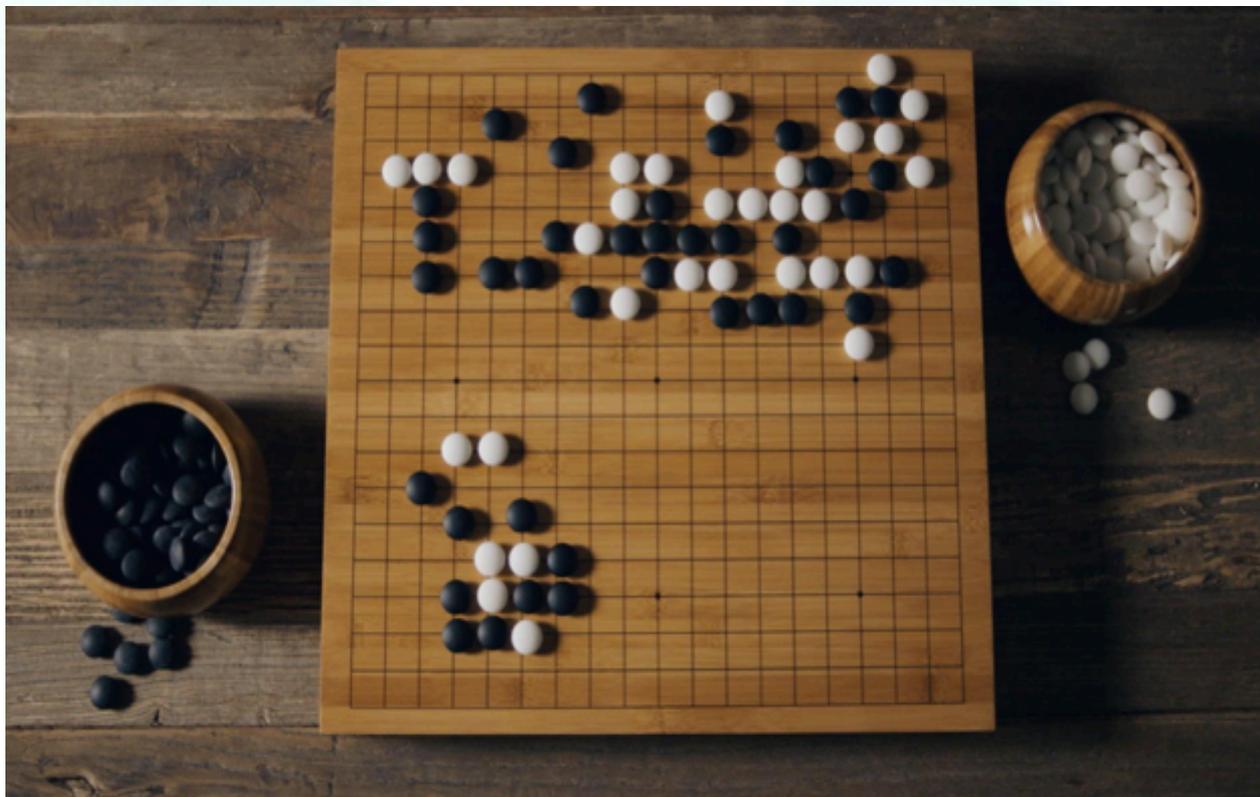
声明：该PPT只供非商业使用，也不可视为任何出版物。由于历史原因，许多图片尚没有标注出处，如果你知道图片的出处，欢迎告诉我们 at [wszheng@ieee.org](mailto:wszheng@ieee.org).



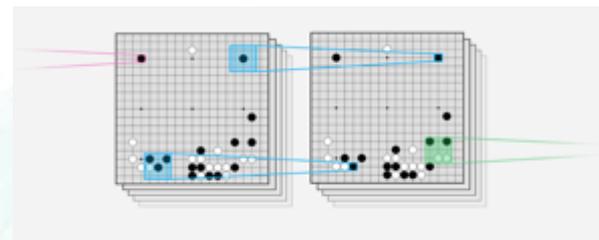
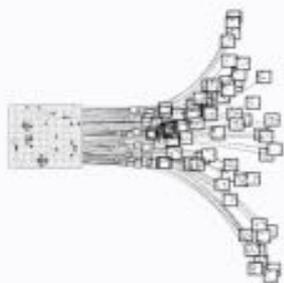
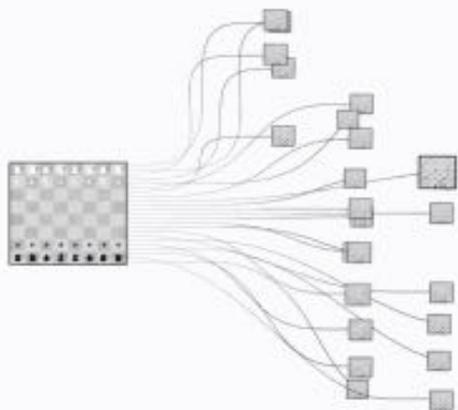
# 故事从围棋开始：这个也可以用视觉算法？

## 围棋的搜索空间太大

- 过于复杂（甚至难以求解，譬如求解一个多元不可导函数的最小值），用搜索或者动态规划等常规方法很难在多项式时间中求解出问题的解。

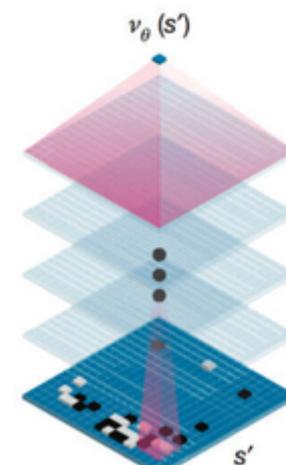
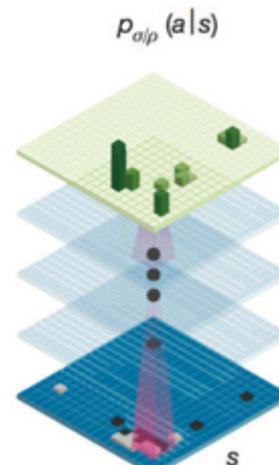


# AlphaGo



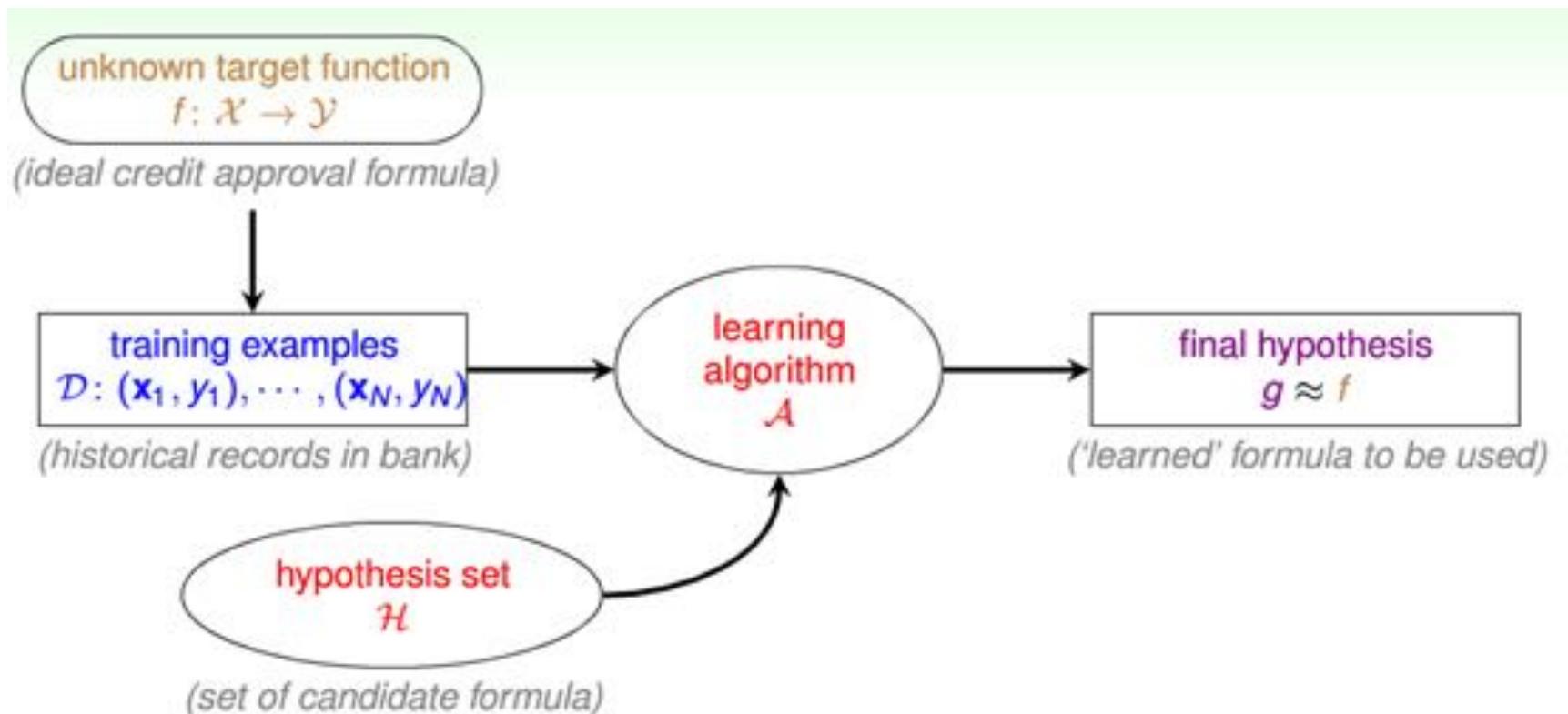
Policy network

Value network



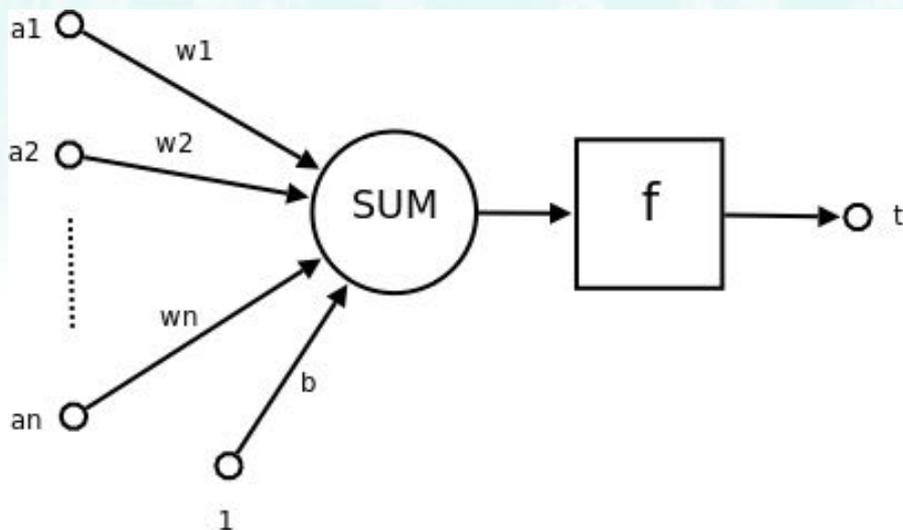


# 机器学习的算法框架



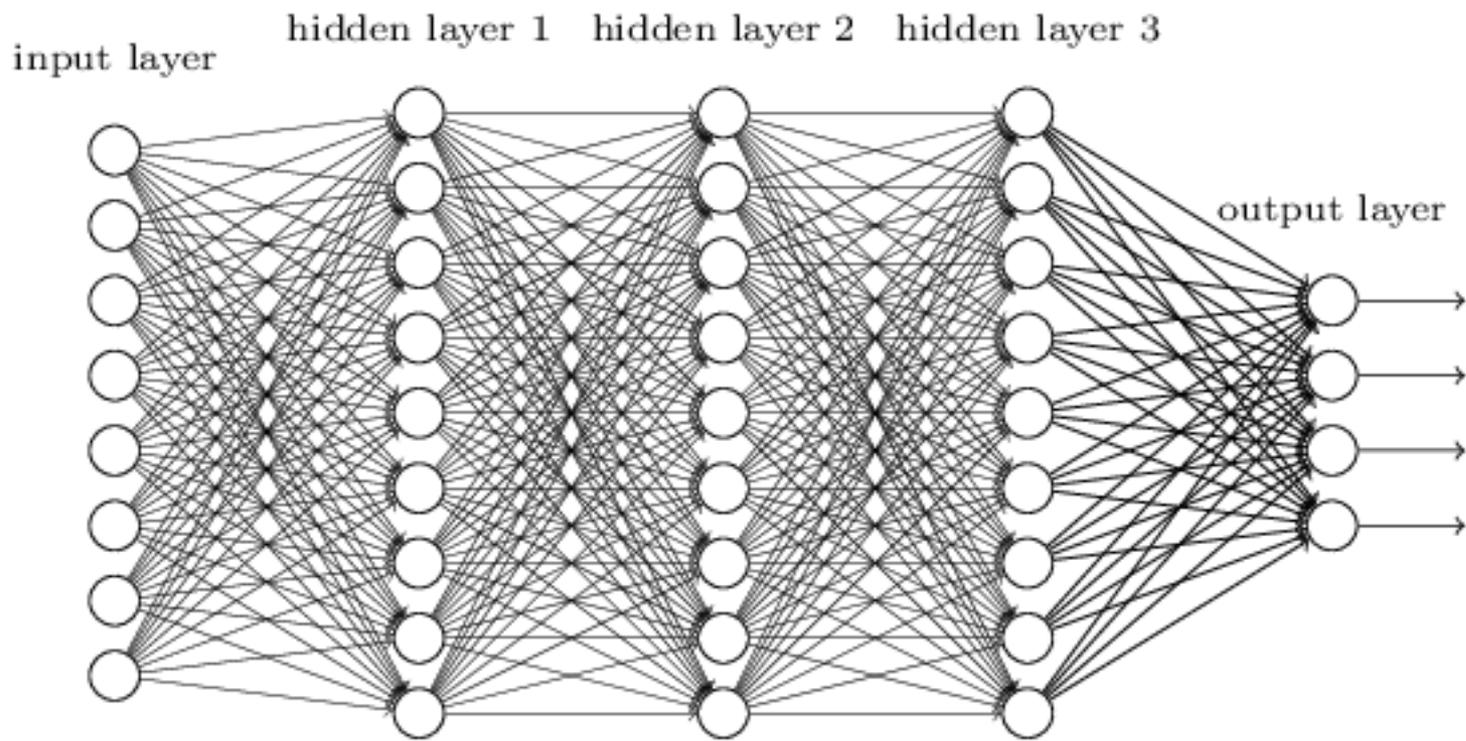
# 人工神经网络

- 许多问题复杂，我们难以直接确定模型的形式（譬如，使用二次函数，三次函数，还是十次函数？）
- 如果使用一个线性函数表示一个神经元，一个函数的输出与另一个函数的输入模拟神经网络的连接，能否模拟出一个复杂的模型？
- 此时便有了神经网络模型，它最基本的神经元是一种线性模型。



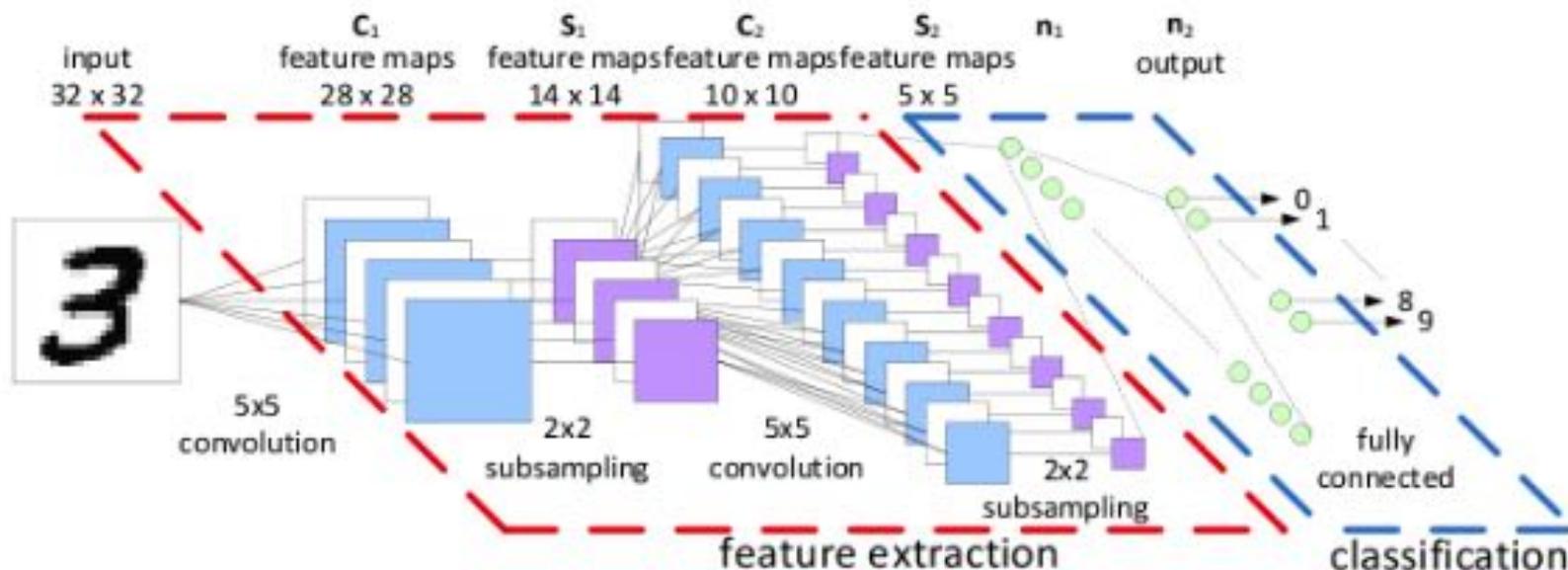
# 人工神经网络

- 神经网络模型将多个线性模型（及更复杂的模型）层层叠加，构成了一个复杂的假设模型。

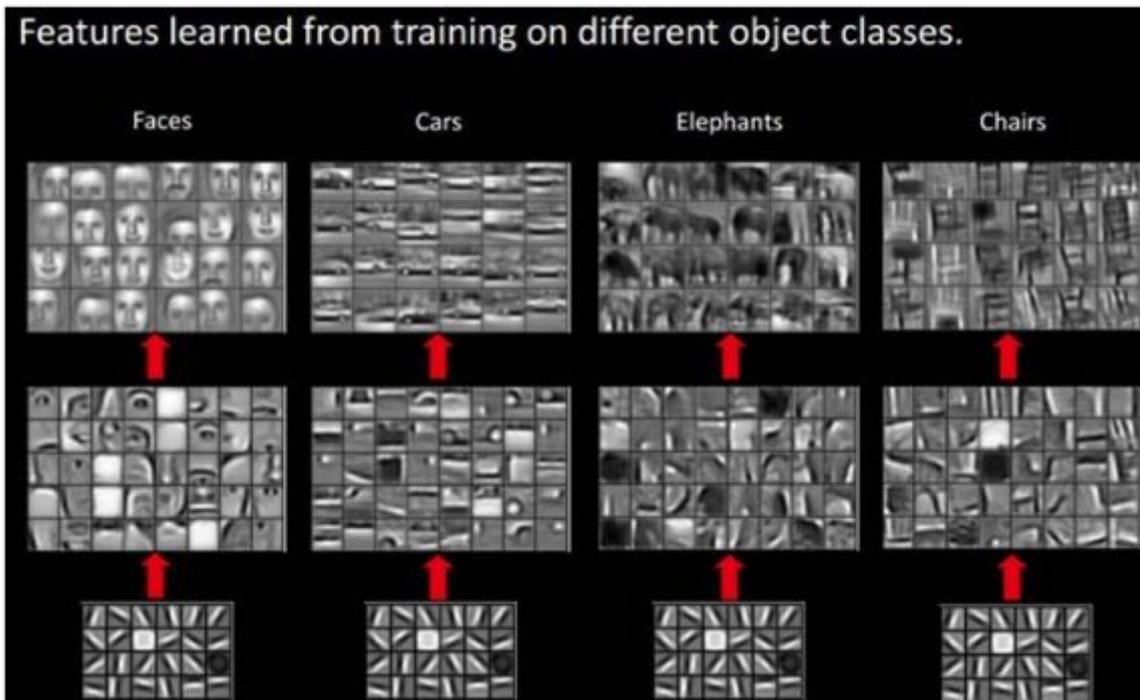
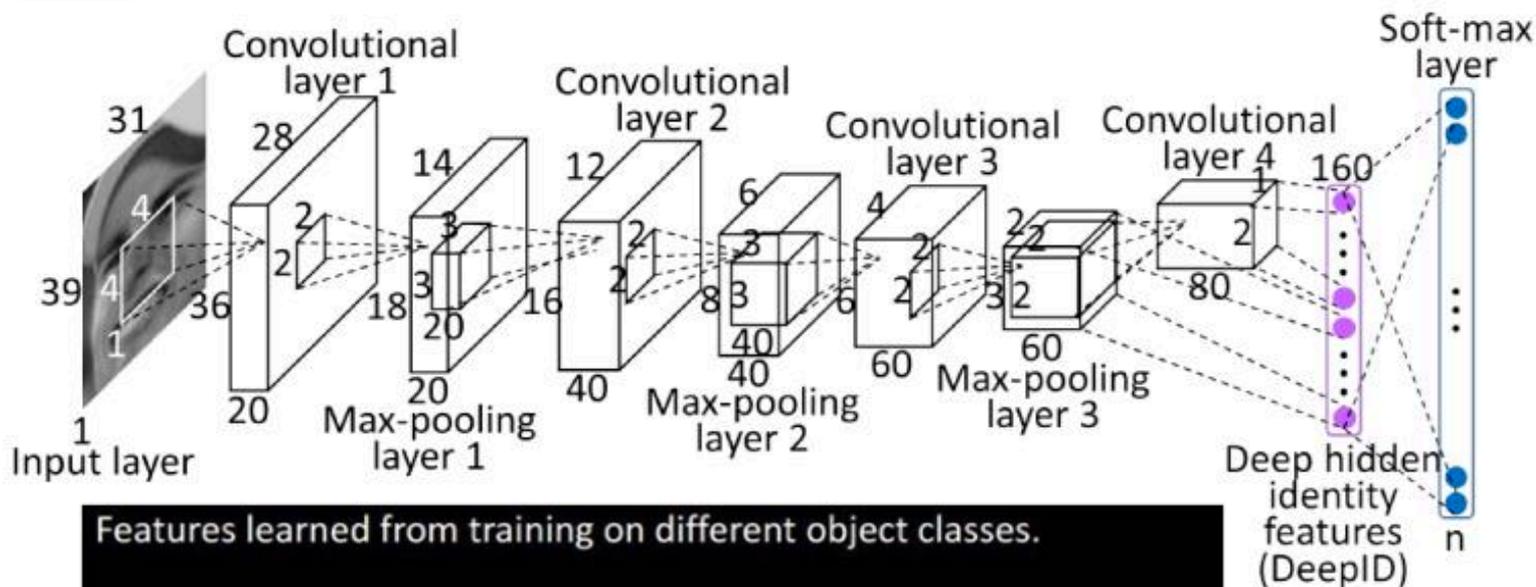


# 深度学习

- 深度学习已被应用计算机视觉、语音识别、自然语言处理、音频识别与生物信息学等领域并获取了极好的效果。



# 卷积神经网络

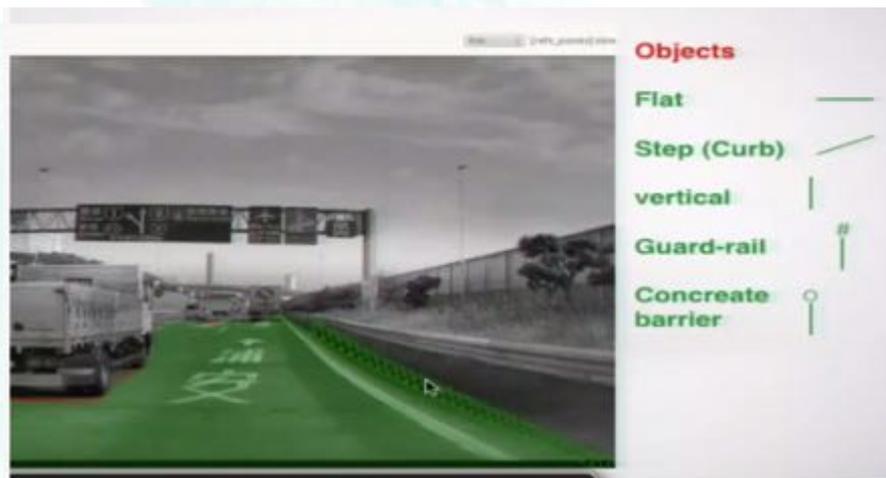


# 自动驾驶

- ❑ 计算机视觉在无人驾驶中应用
- ❑ 可行驶区域 (free space) 检测
  - 利用深度学习做场景理解
  - 绿色为可行驶区域

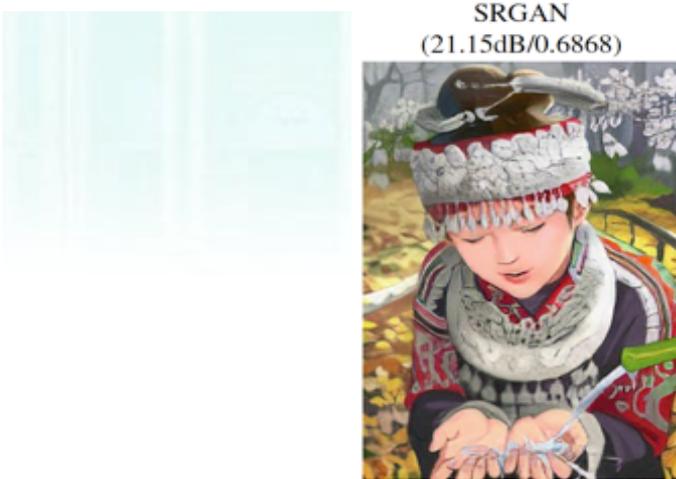
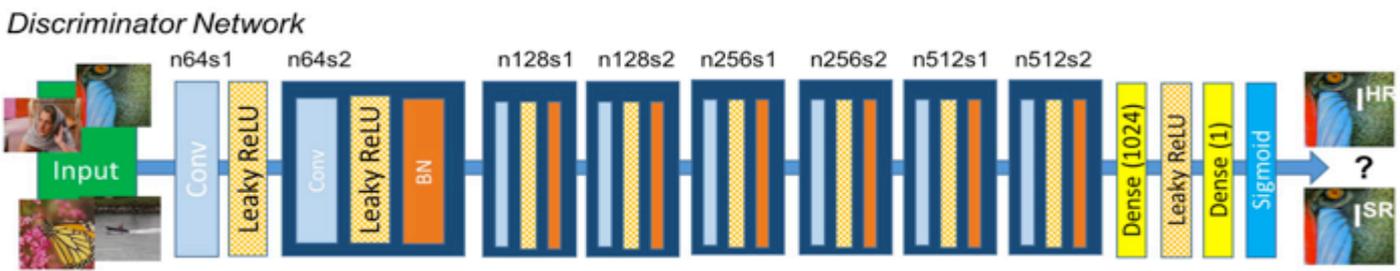
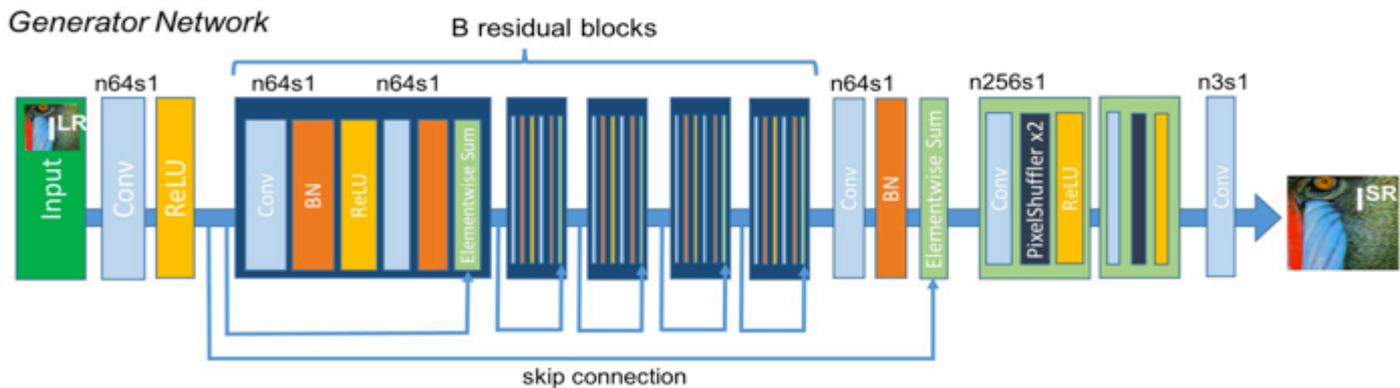


场景理解

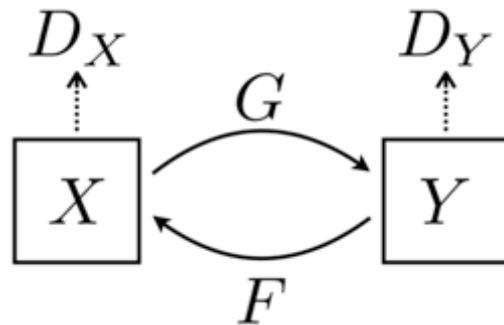


可行驶区域检测

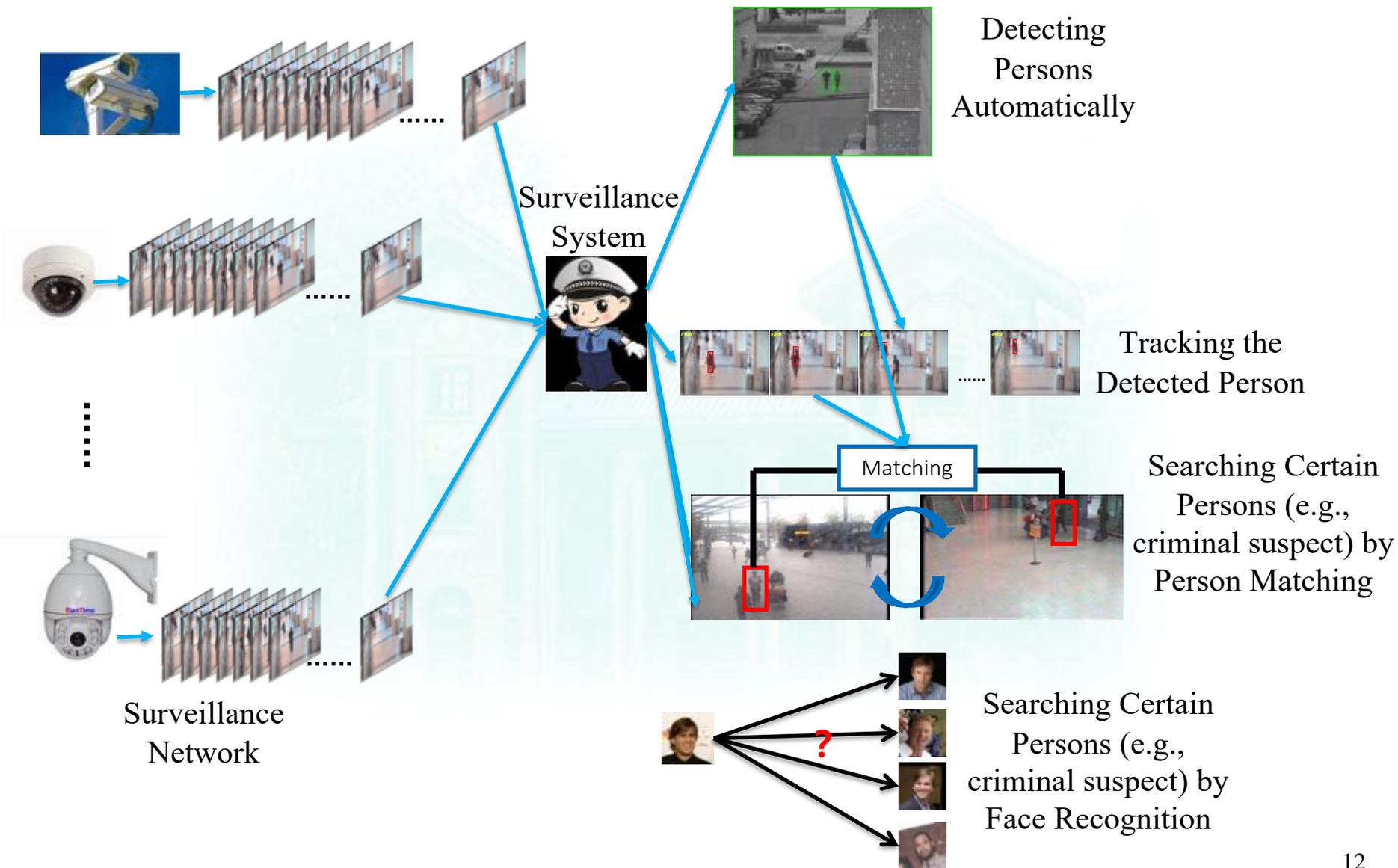
# 超分辨率



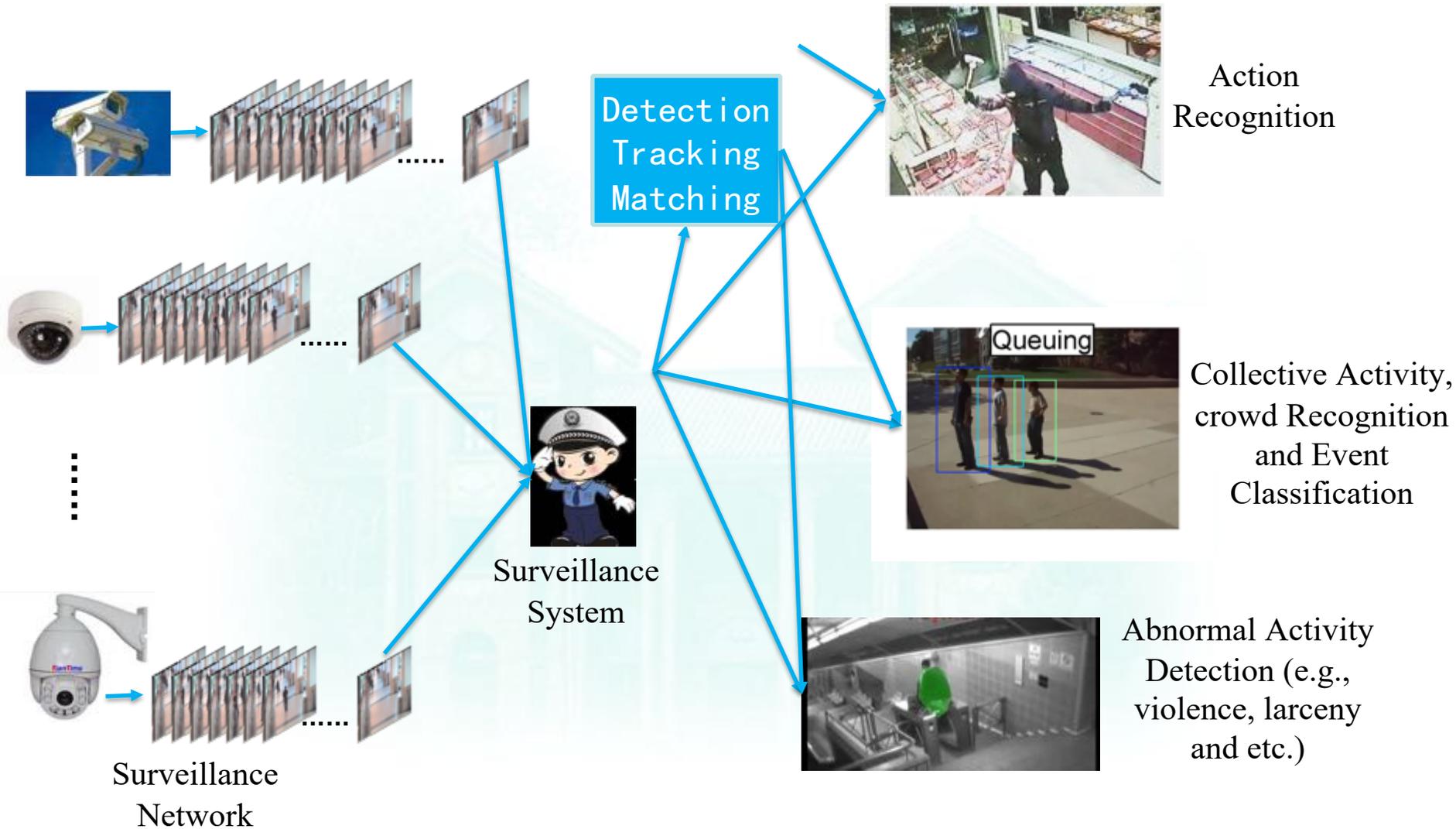
# 梦想成真



# 智能视频监控系统：你去过哪里

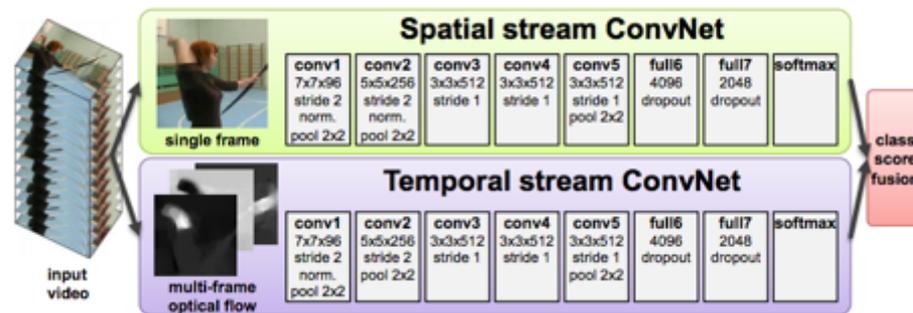


# 智能视频监控系统：你做啥？



# 行为识别

## □ 双通道神经网络



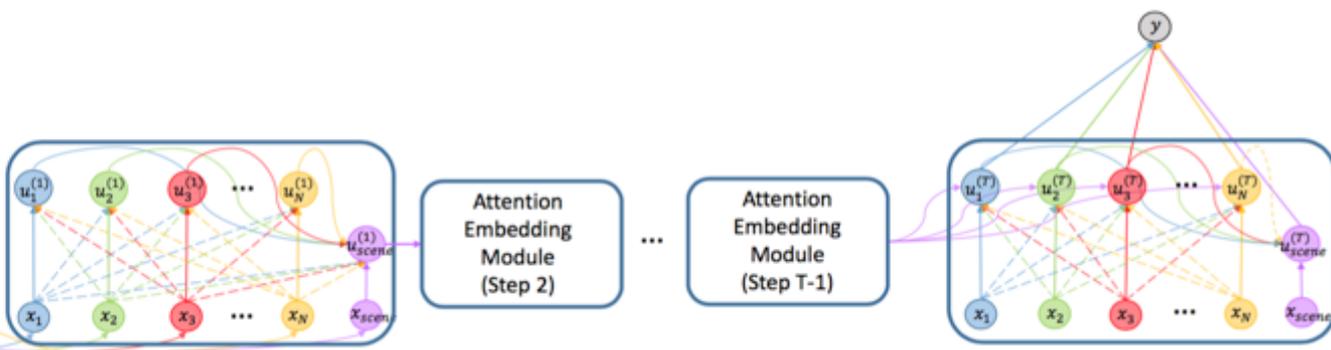
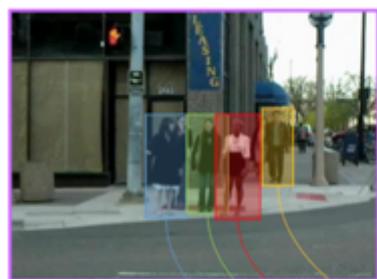
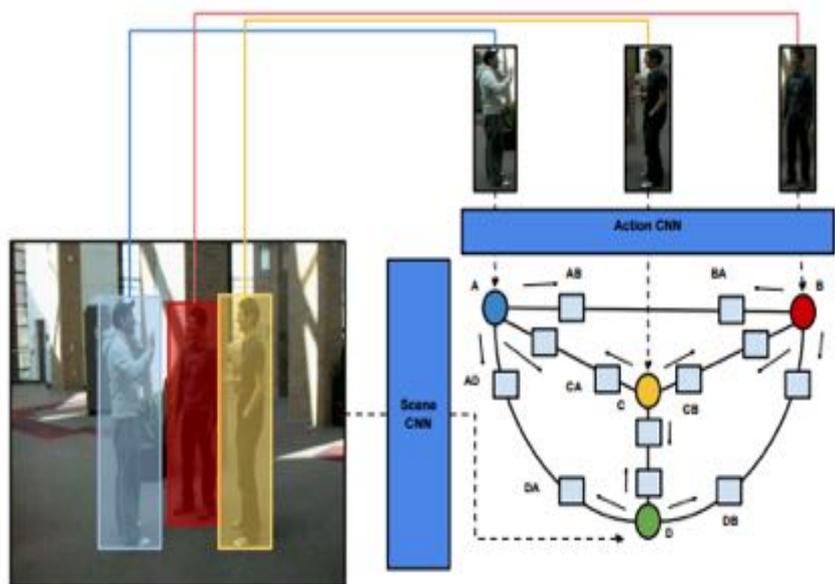
Activity Example: playing phone



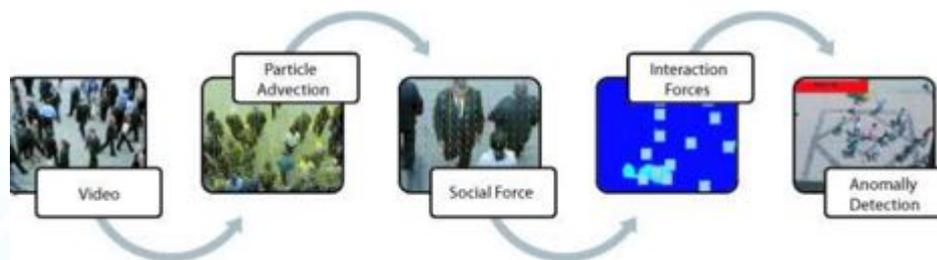
Recognition Comparison Results

- |                   |                         |
|-------------------|-------------------------|
| drinking          | sitting chair           |
| pouring           | moving chair            |
| calling phone     | taking out wallet       |
| playing phone     | taking sth. from wallet |
| wearing backpacks | mopping                 |
| packing backpacks | sweeping                |

# 群体行为识别

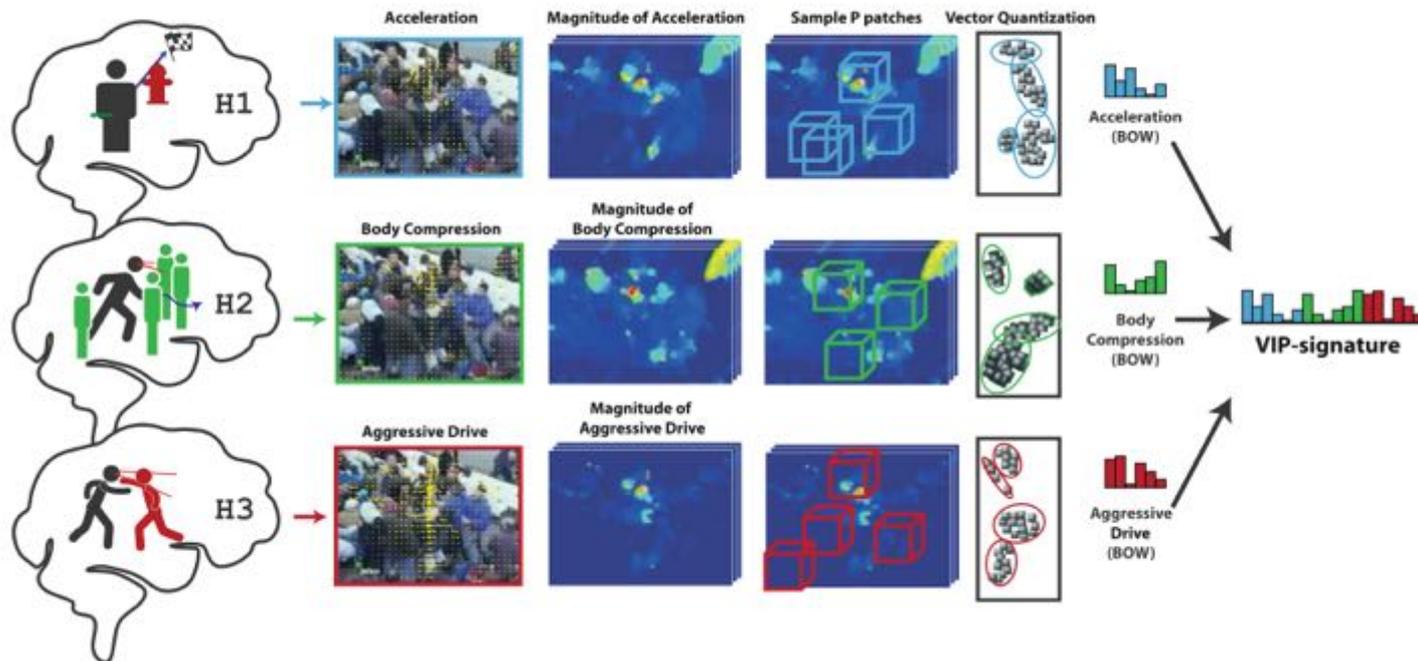


# 暴力行为检测



## Behavioral Heuristics

## Approximation and Estimation from Video



# 智慧医疗



**Corporate Wellness Programs Get Boost From Wearables**

**SAMSUNG**



**Machine Learning Set to Dominate This Year's Healthcare Trends**

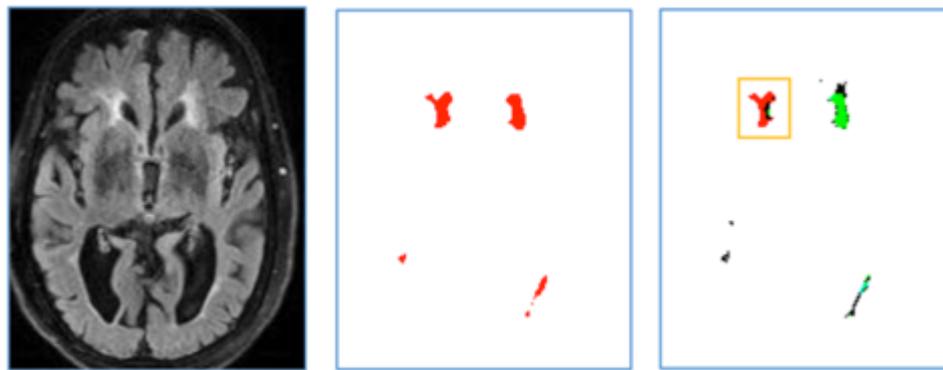
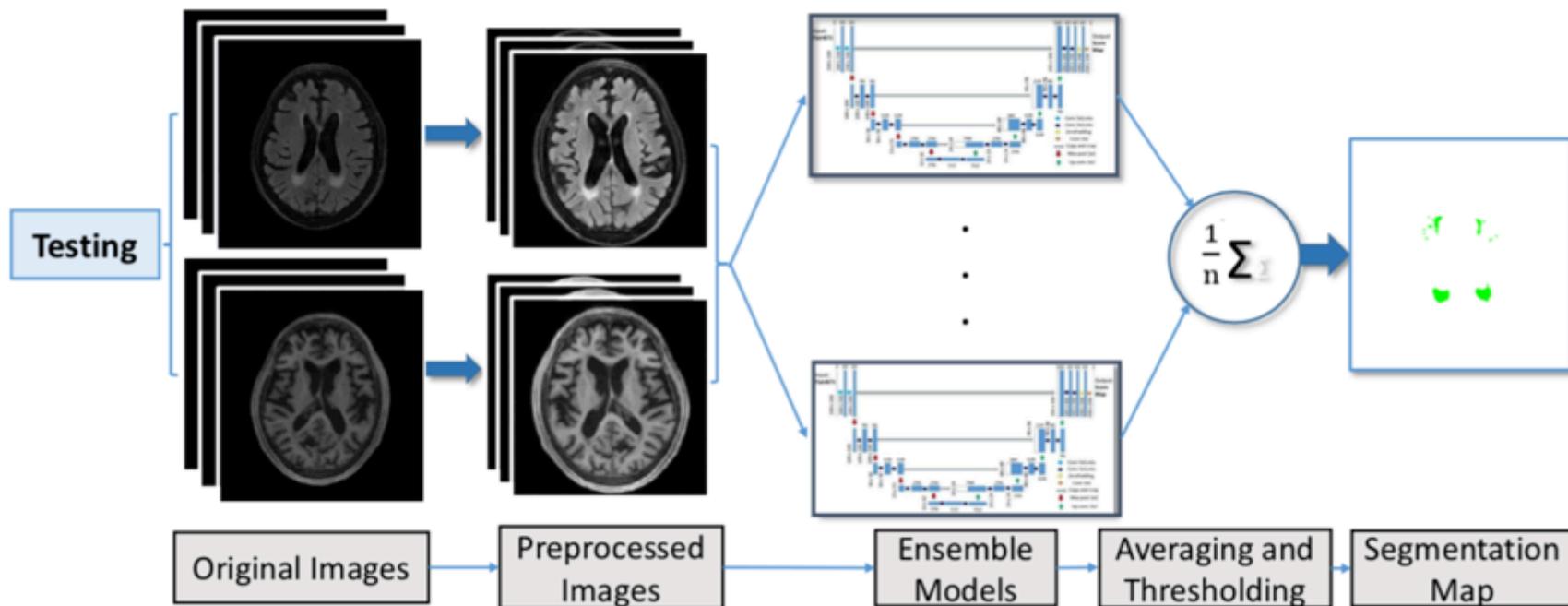


**Next Step for Remote Patient Monitoring: Virtual Physical Therapy**



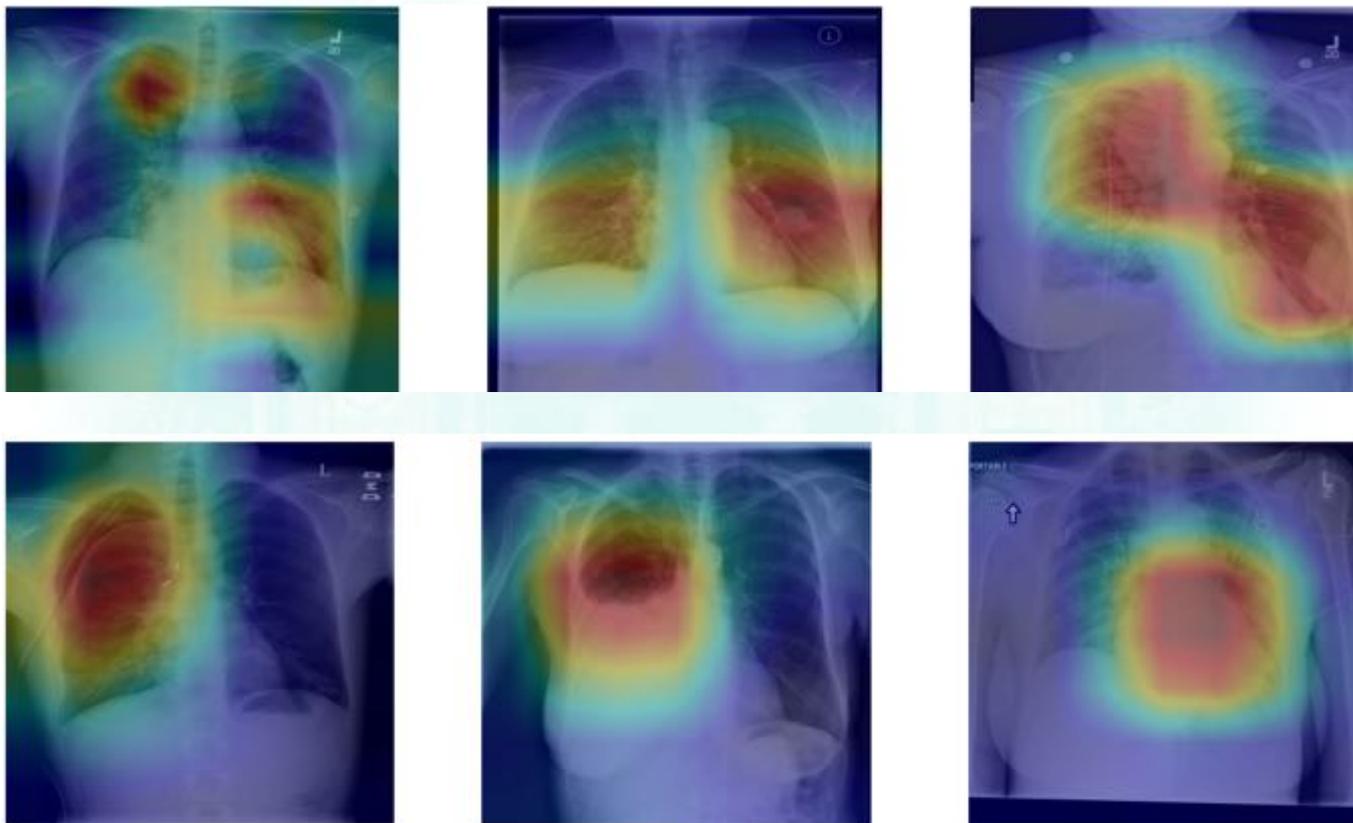
**Medical Robots Step Into the Healthcare Spotlight**

# 智慧医疗



Ground Truth

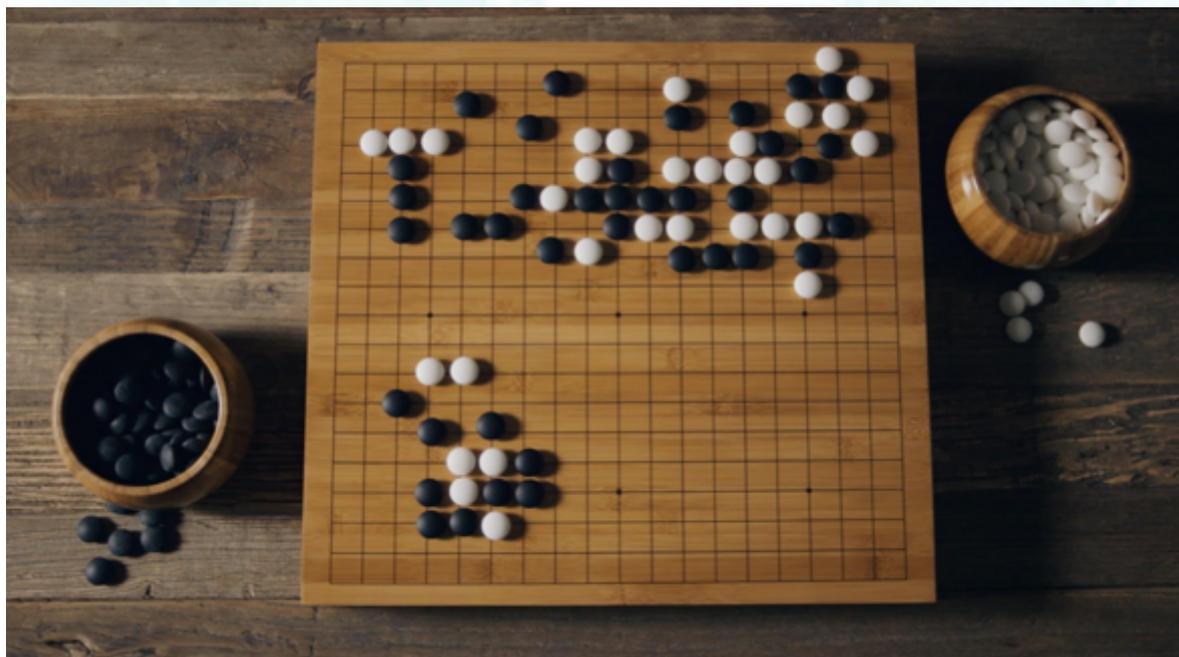
如何增加AI辅助诊断的可解释性/透明性  
AI模型关注的区域与病变区域一致！



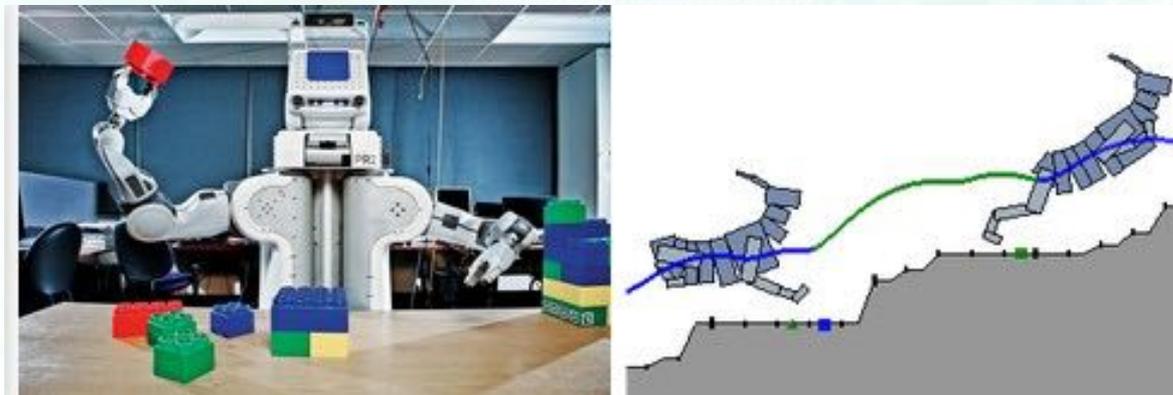
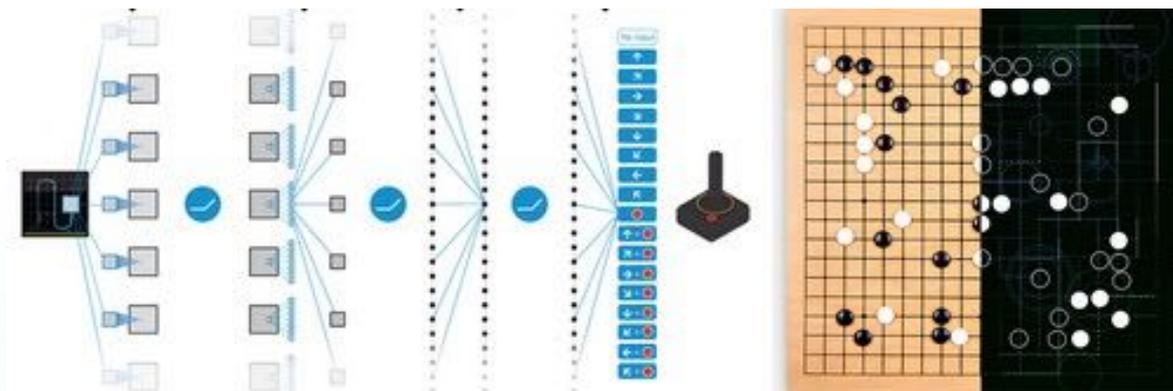
# 让我们再回到AlphaGo



# AlphaGo



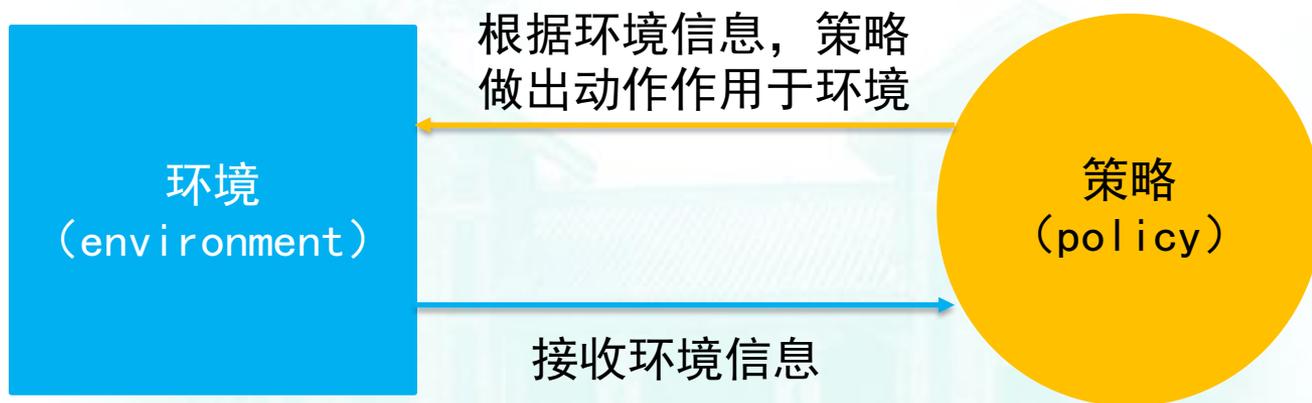
# 强化学习



现实中我们往往没有办法对每个时刻做出的动作的好坏进行标注，但是我们却知道我们想要的结果，强化学习告诉我们有这些就够了。

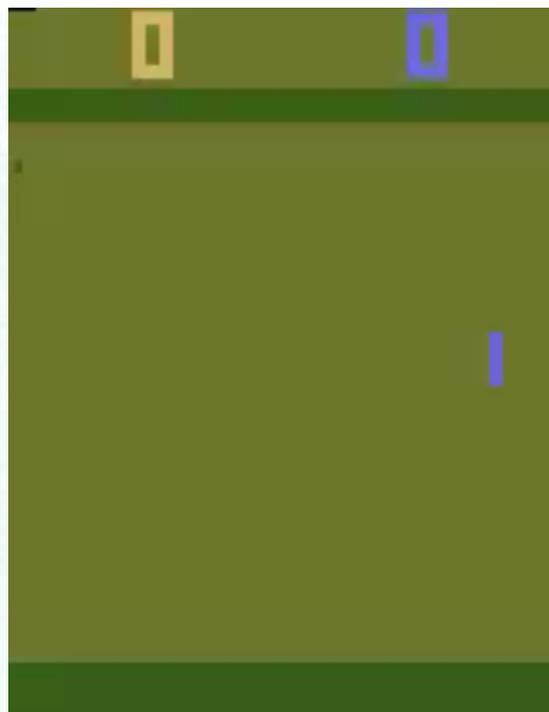
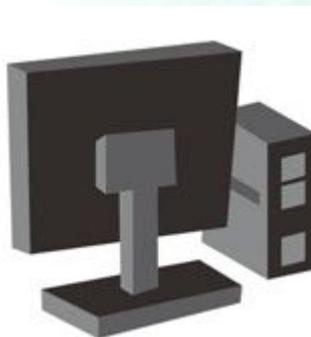


# 强化学习

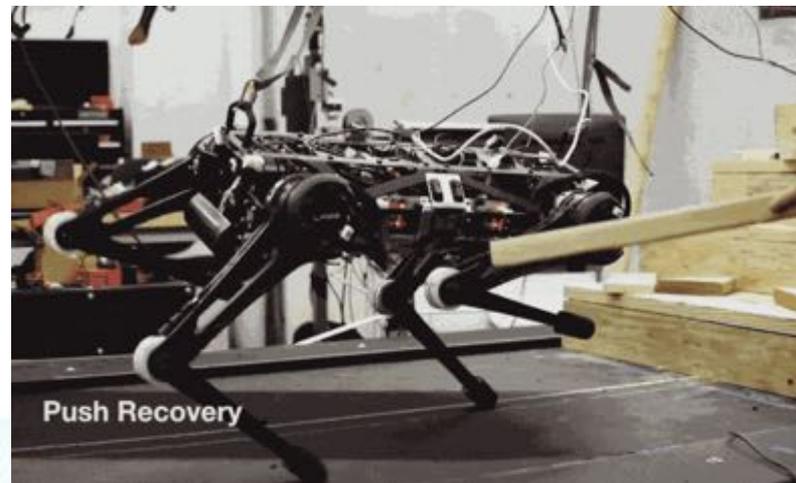


# 机器学习打游戏

训练一种策略 (policy)，这种策略能够根据当前游戏画面 (state)，做出动作 (action: up or down)



# 让机器学习我们



# 人机互动





## 第二部分：基本技术介绍



# 技术路线

- 贝叶斯建模
- 线性分类
- 非线性分类
- 深度神经网络
- 集成学习
- 无监督学习
- 泛弱监督学习
- 数据降维与模型压缩



# 分类的本质是贝叶斯建模

## □ 贝叶斯定理

通常事件A在事件B已发生的条件下发生的概率, 与B在A已发生的条件下发生的概率是不一样的。然而两者是有确定关系的, 贝叶斯定理就是这种确定关系的描述。

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

其中  $P(A|B)$  称作A的后验概率,  $P(A)$  称作A的先验概率。



# 分类的本质是贝叶斯建模

## □ 分类模型——朴素贝叶斯 (Naïve Bayes)

理论上，概率分类器是一个条件概率模型：

$$P(C | X_1, \dots, X_n)$$

其中独立变量 $C$ 有若干类别，条件依赖于若干特征变量  $X_i (i = 1, \dots, n)$



# 分类的本质是贝叶斯建模

## □ 分类模型——朴素贝叶斯 (Naïve Bayes)

根据贝叶斯公式可以得到：

$$P(C | X_1, \dots, X_n) = \frac{P(X_1, \dots, X_n | C) * P(C)}{P(X_1, \dots, X_n)}$$

即是：

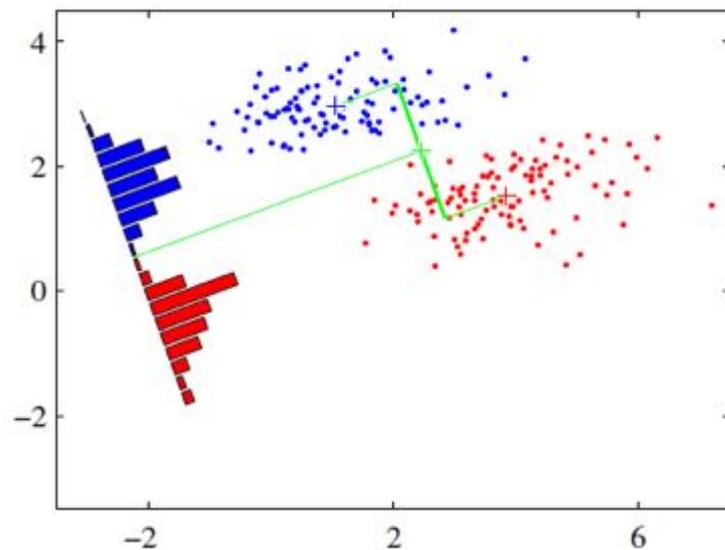
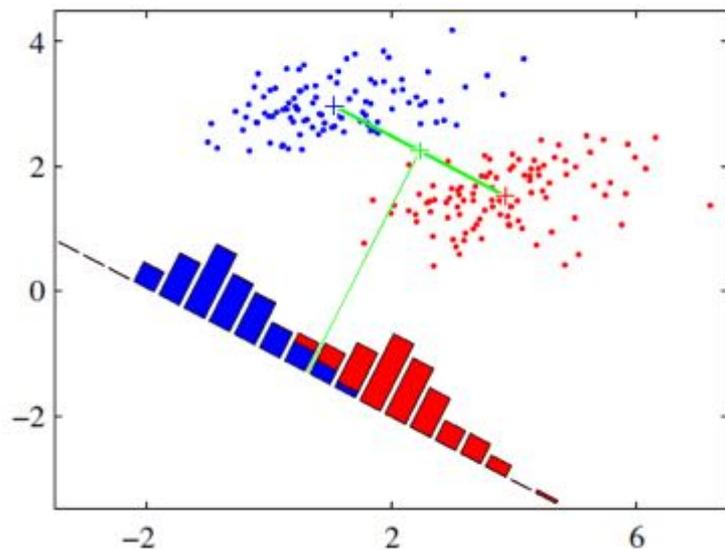
$$P(\text{类别} | \text{特征}) = \frac{P(\text{特征} | \text{类别}) * P(\text{类别})}{P(\text{特征})}$$

分类结果可表示为：

$$\text{Classifier}(X_1, \dots, X_n) = \underset{C}{\operatorname{argmax}} P(C | X_1, \dots, X_n)$$

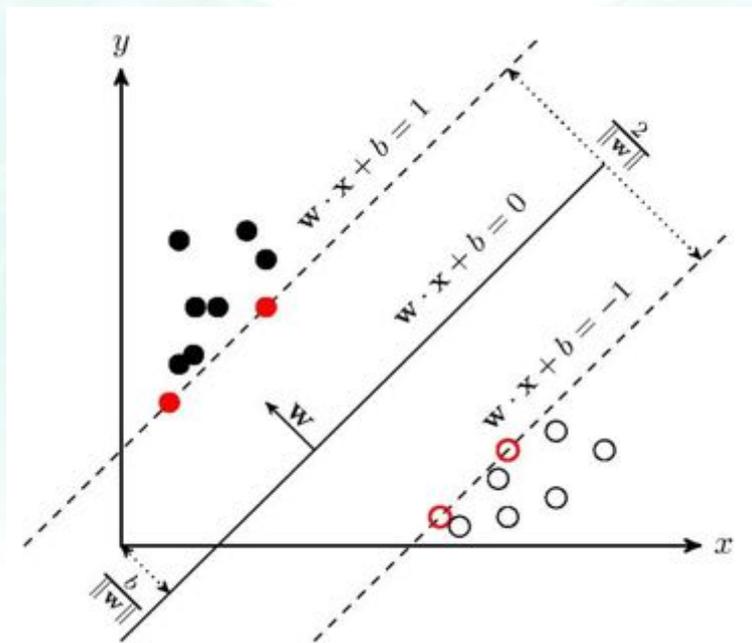
# 线性分类是最直观的

- 线性判别分析 (Linear Discriminant Analysis, LDA) 将数据在低维度上进行投影，投影后希望每一种类别数据的投影点尽可能的接近，而不同类别数据的类别中心之间的距离尽可能的大



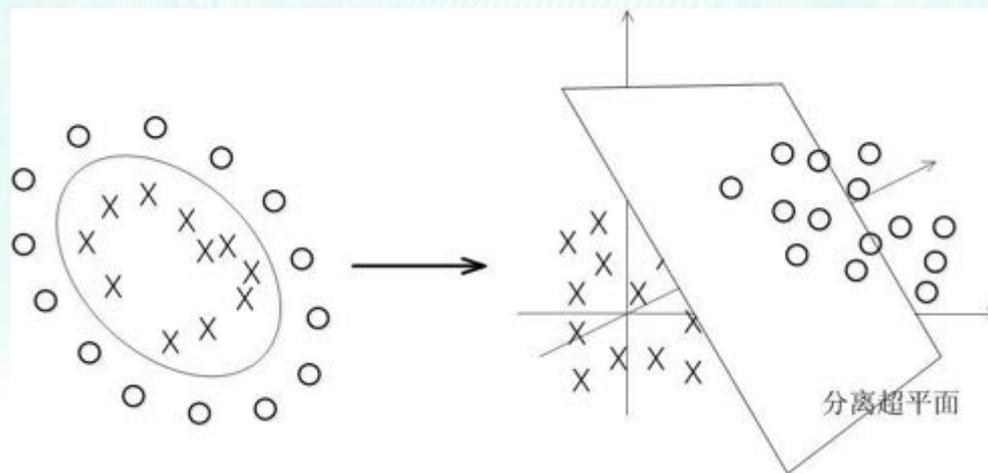
# 线性分类是最直观的

- 支持向量机（support vector machines, SVM）是一种二分类模型，基本想法是求解能够正确划分训练数据集并且几何间隔最大的分离超平面



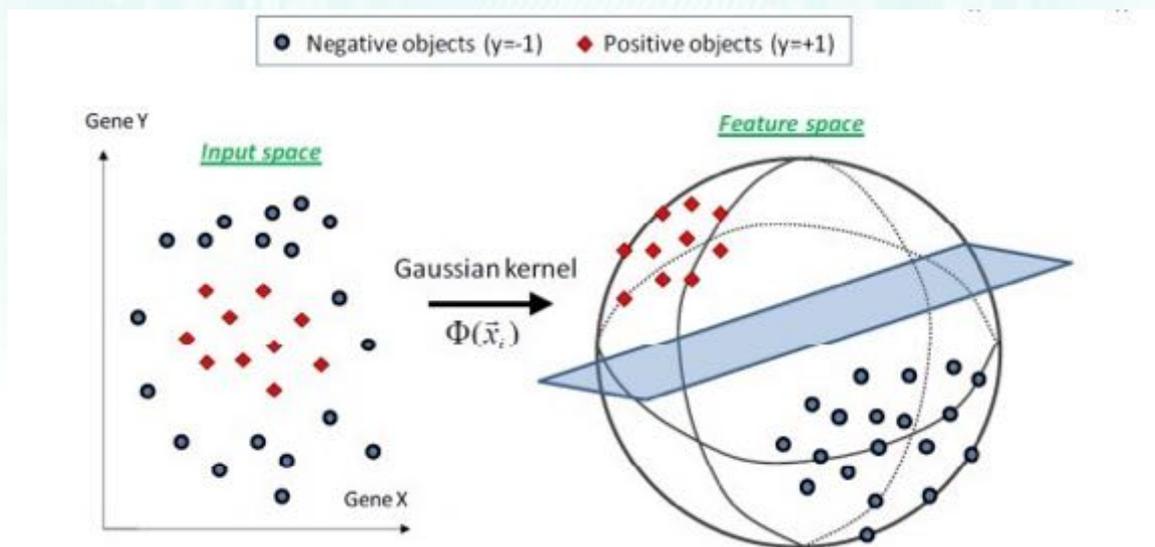
# 非线性分类是普遍现象

- 大部分时候数据并不是线性可分的，这个时候不存在满足条件的超平面。非线性SVM 通过使用核函数，将数据映射到高维空间，来解决在原始空间中线性不可分的问题



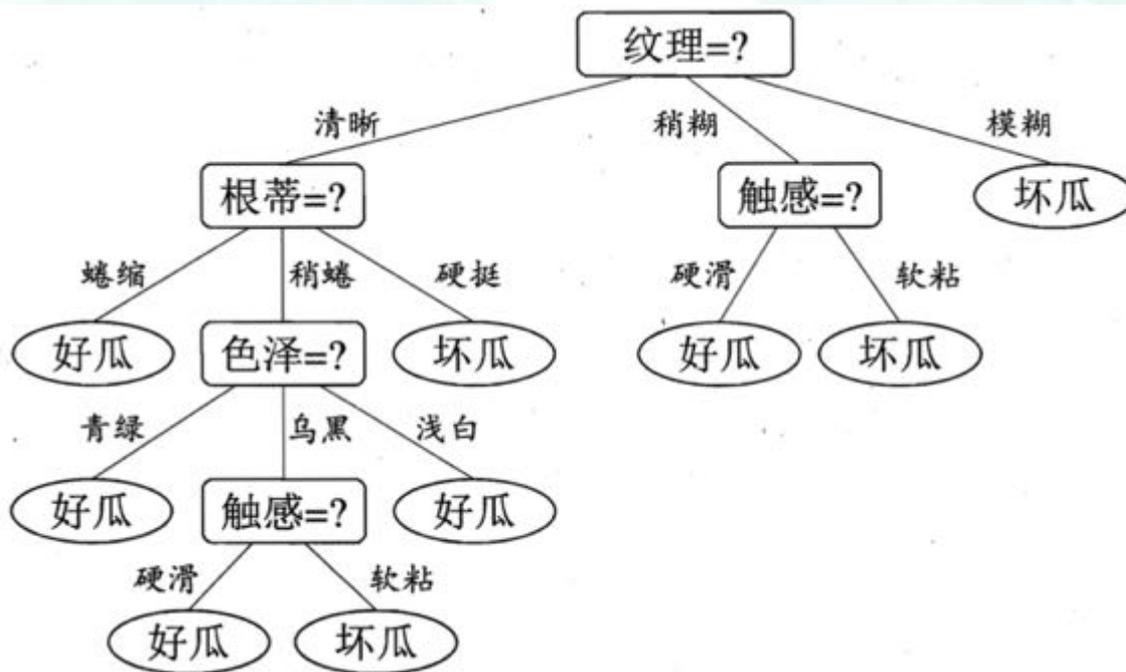
# 非线性分类是普遍现象

- 核函数是计算两个向量在隐式映射过后的空间中的内积的函数，将特征进行从低维到高维的转换，避免了直接在高维空间中的复杂计算。
- 常用核函数：高斯核（Gaussian Kernel），指数核（Exponential Kernel）等



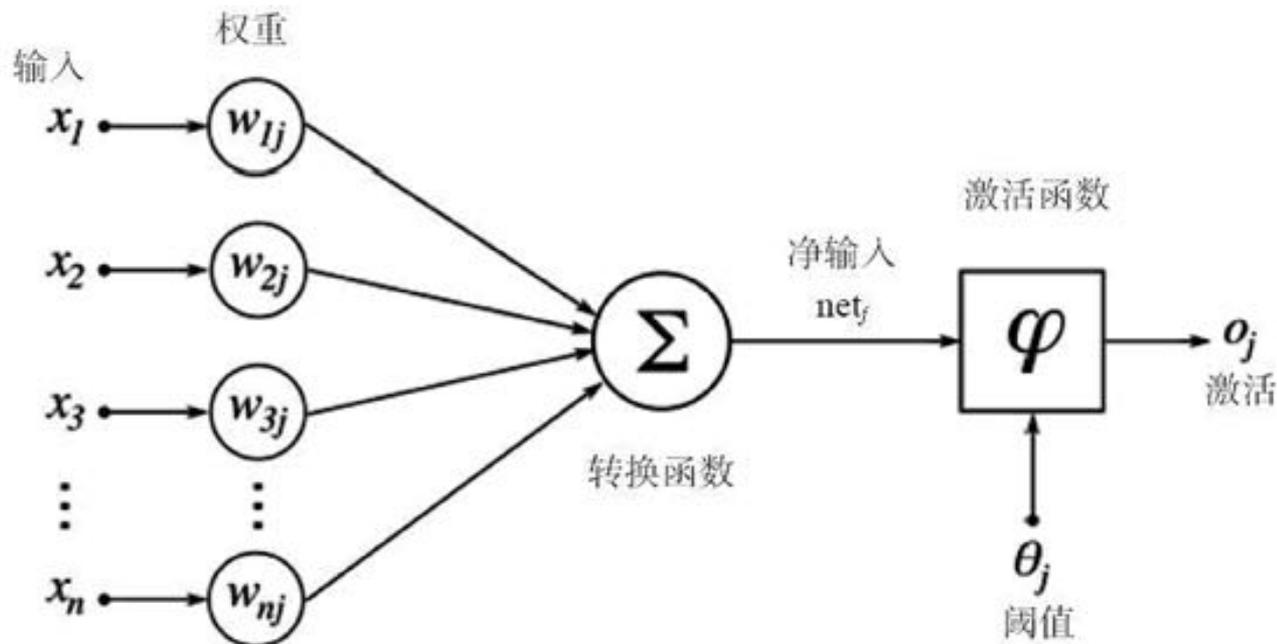
# 非线性分类是普遍现象

- 决策树 (decision tree) 是一个树结构，它从根节点开始，测试待分类项中相应的特征属性，并按照其值选择输出分支，直到到达叶子节点，将叶子节点存放的类别作为决策结果



# 非线性函数建模利器-深度神经网络

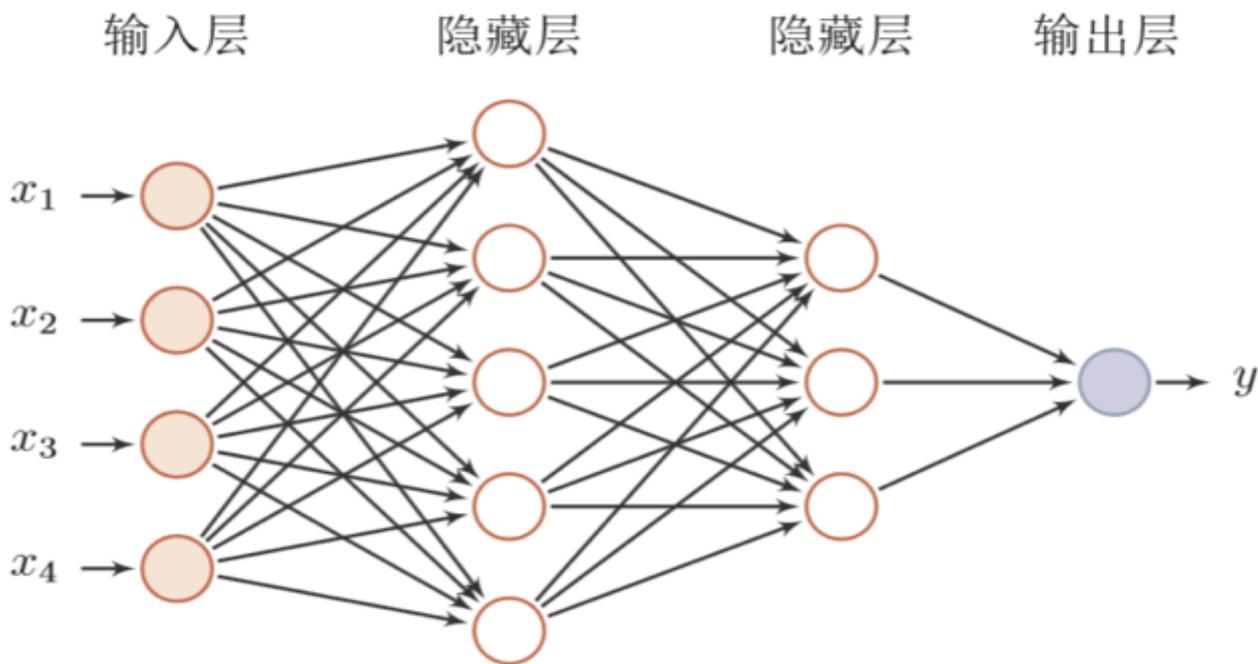
## 典型的神经元结构



- 当只有一层这样的神经元存在时，它被称为感知机。
- 单层感知机有一个严重的问题，即它对稍复杂一些的函数都无能为力（比如最为典型的“异或”操作）。

# 非线性函数建模利器-深度神经网络

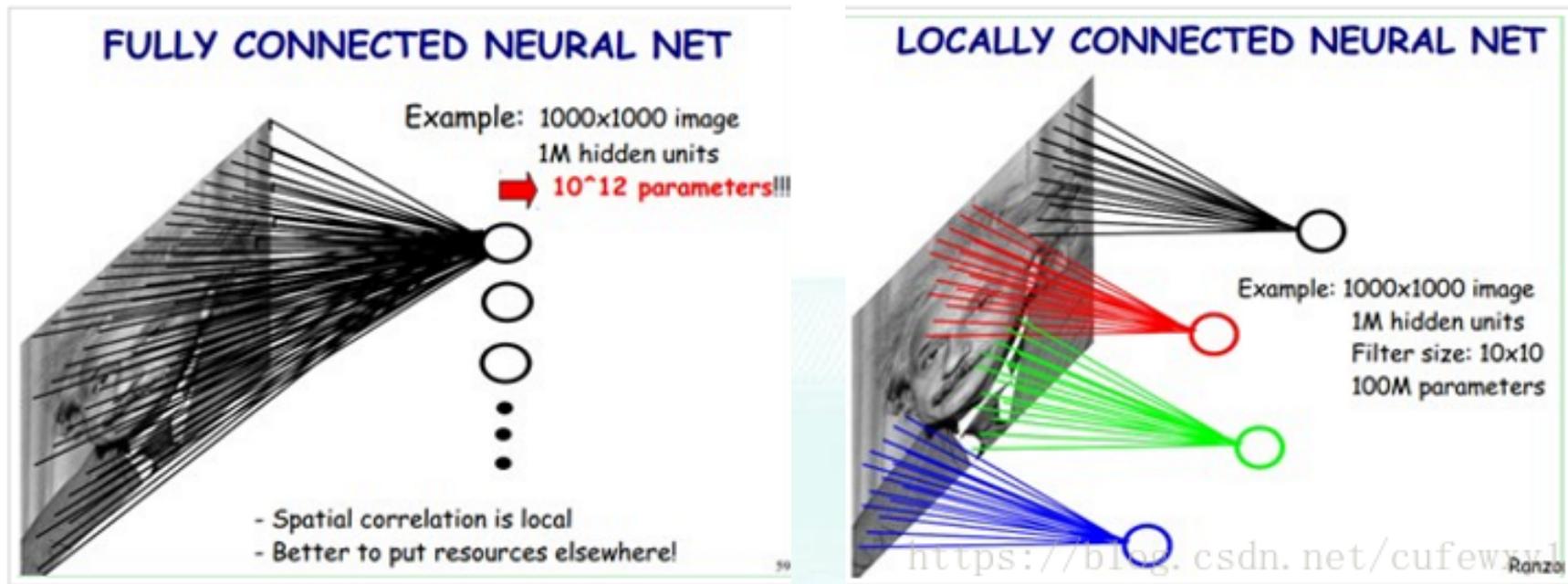
## 多层感知机



多层感知机解决了之前无法模拟异或逻辑的缺陷，同时更多的层数也让网络更能够刻画现实世界中的复杂情形。

# 非线性函数建模利器-深度神经网络

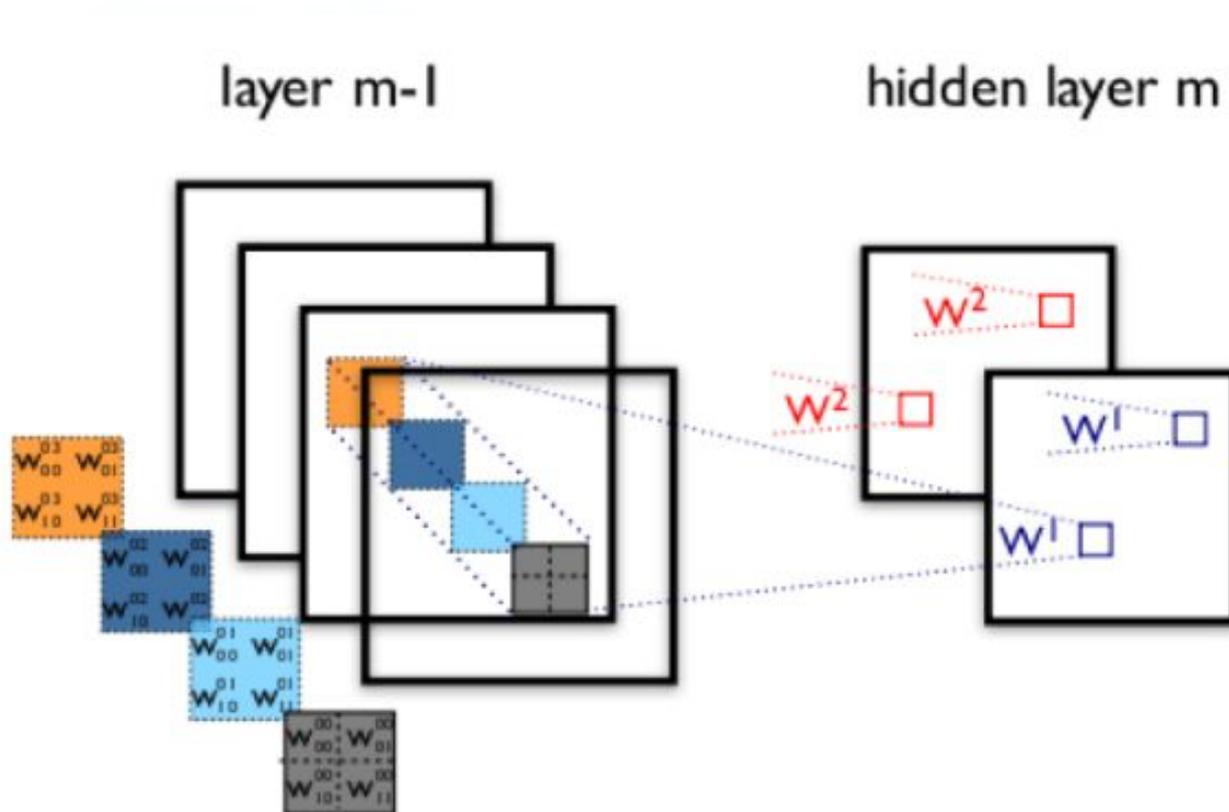
## 全连接层和卷积层对比



- 全连接DNN的结构里下层神经元和所有上层神经元都能够形成连接，带来的潜在问题是参数数量的膨胀。
- CNN通过“卷积核”作为中介。同一个卷积核在所有图像内是共享的，图像通过卷积操作后仍然保留原先的位置关系。

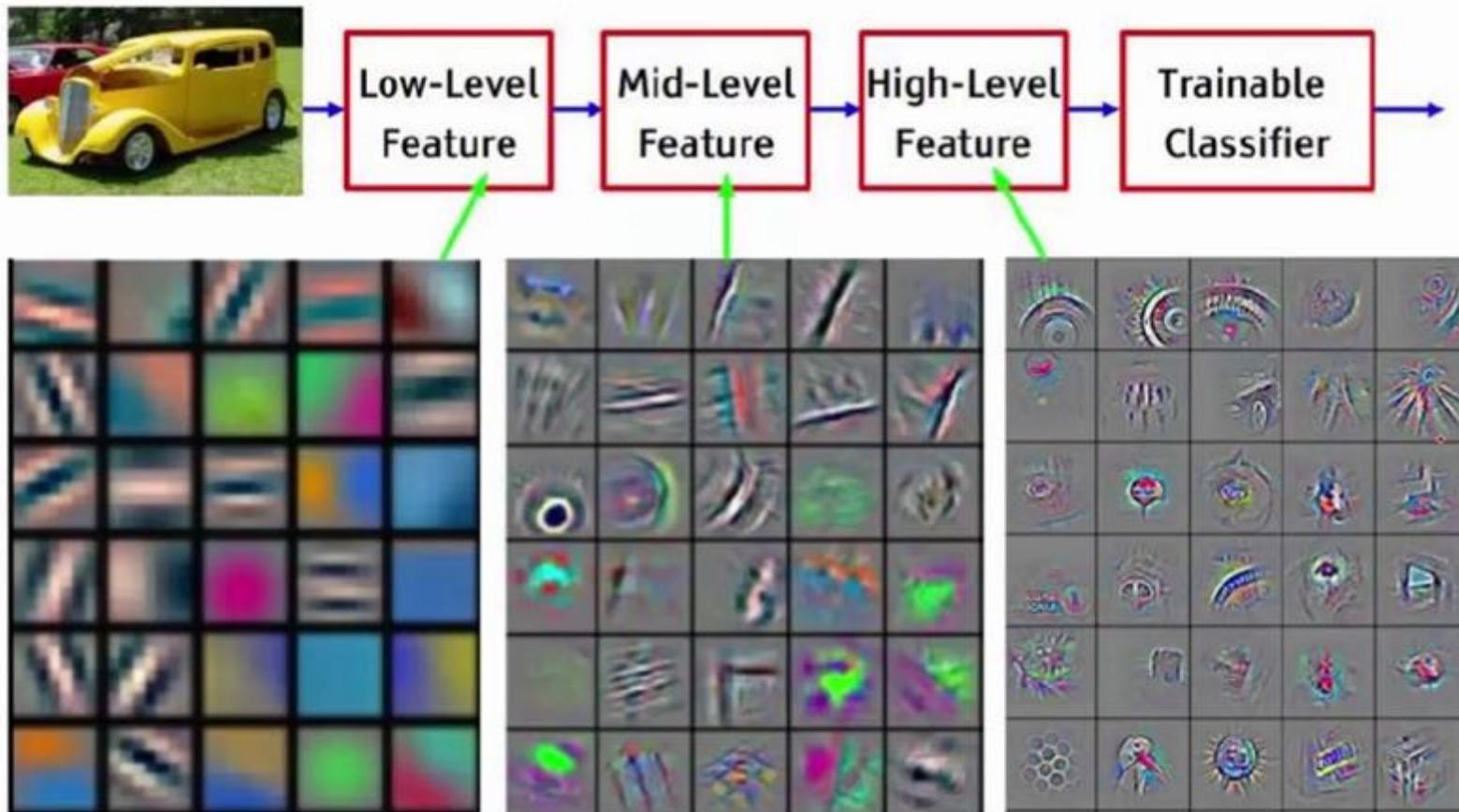
# 非线性函数建模利器-深度神经网络

## 卷积神经网络隐含层



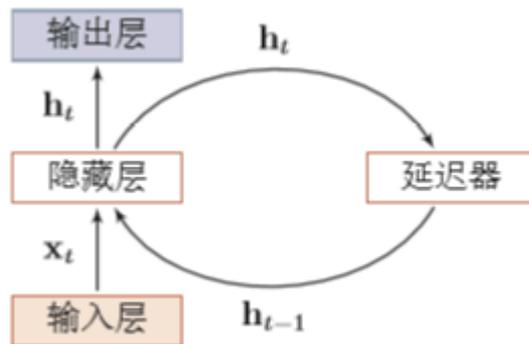
# 非线性函数建模利器-深度神经网络

## 卷积特征提取过程

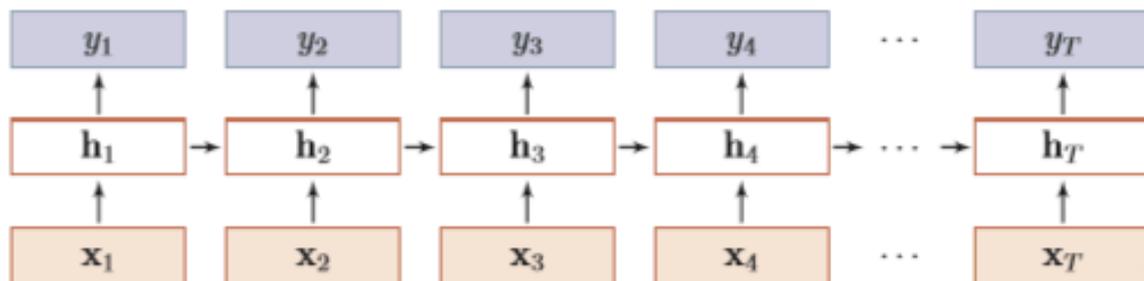


# 非线性函数建模利器-深度神经网络

## 循环神经网络

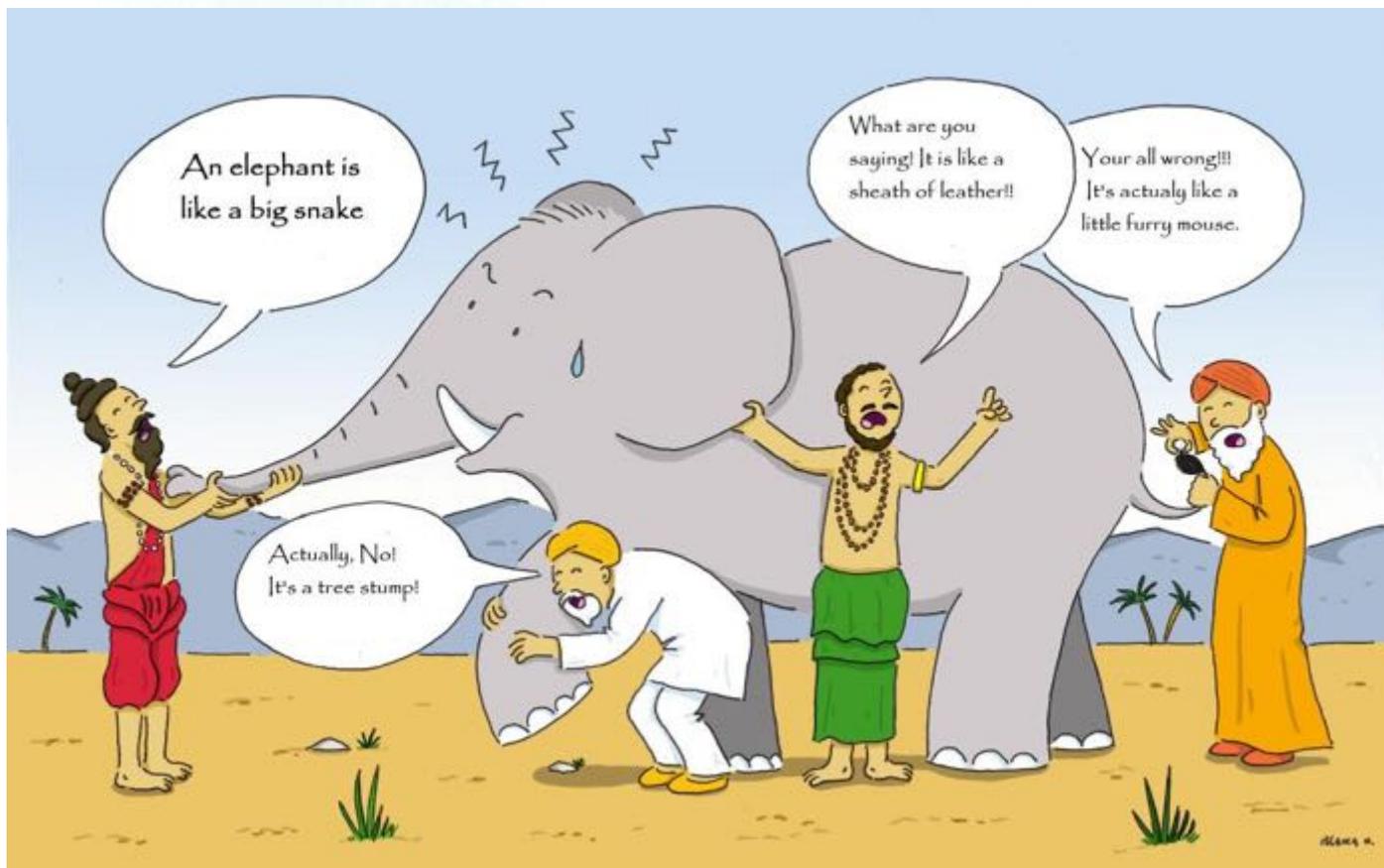


按时间展开



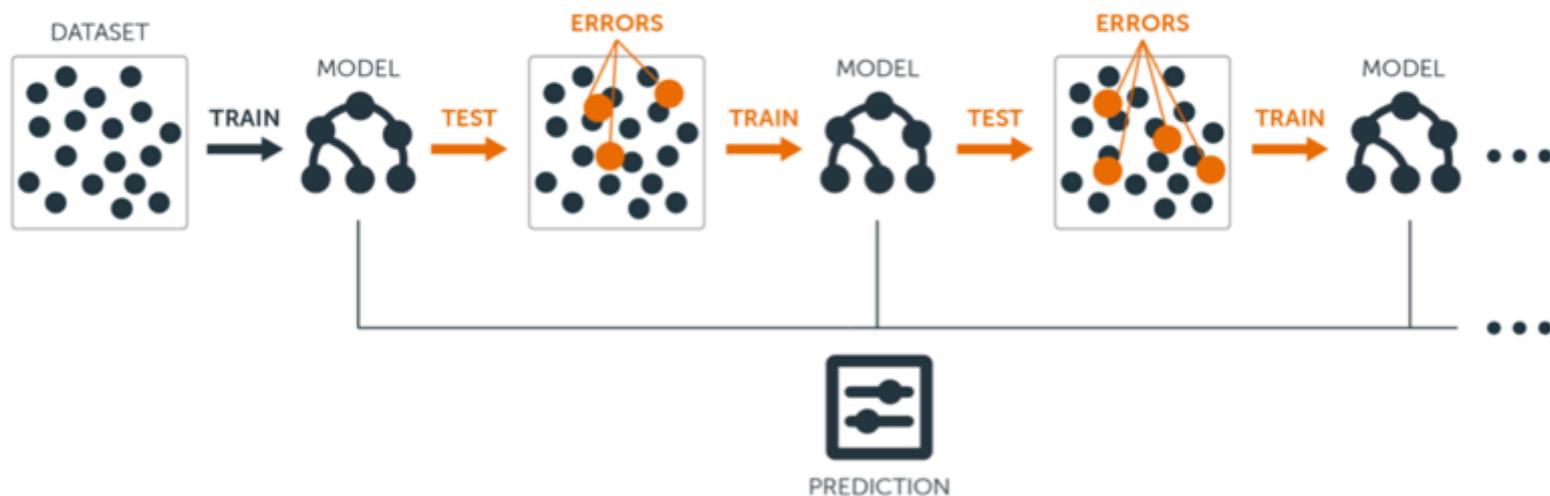
# 三个臭皮匠顶一个诸葛亮-集成学习

- 集成学习 (ensemble learning) 通过构建并结合多个学习器来完成学习任务。



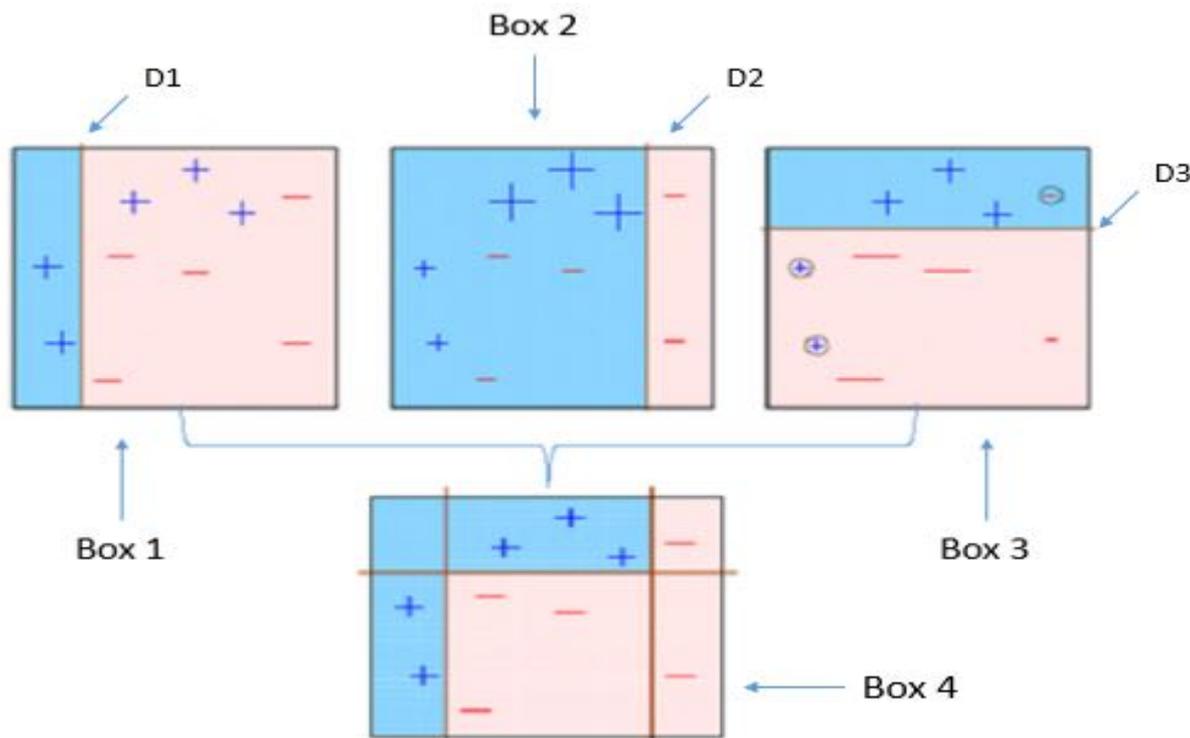
# 三个臭皮匠顶一个诸葛亮-集成学习

- Boosting方法基于串行策略，每一次迭代时训练集的选择与前面各轮的学习结果有关，而且每次是通过更新各个样本权重的方式来改变数据分布。



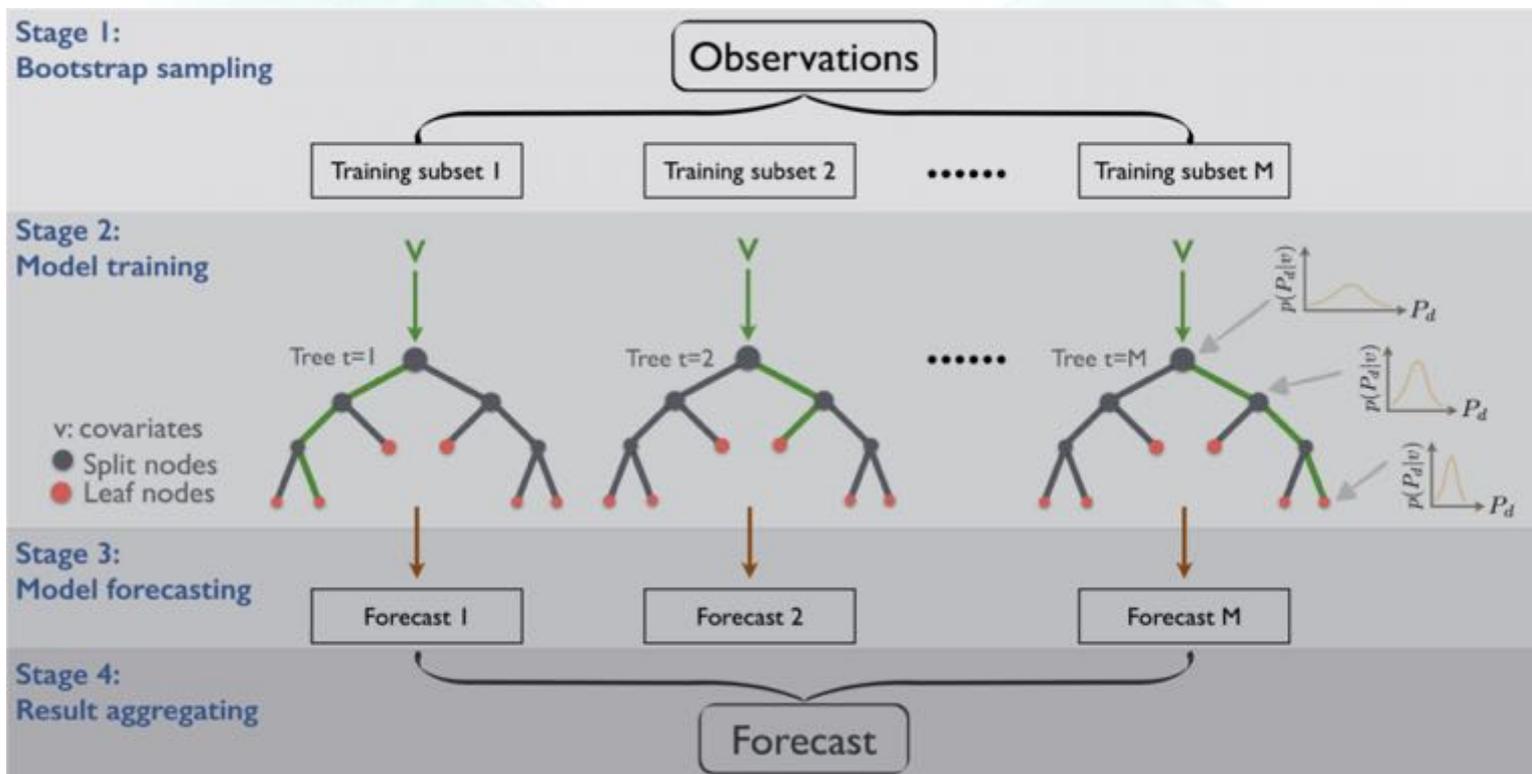
# 三个臭皮匠顶一个诸葛亮-集成学习

- AdaBoost算法通过给基学习器预测错误的样本更高的权重，使得基学习器误判的训练样本在后续受到更多关注的方式来弥补已有模型的不足。



# 三个臭皮匠顶一个诸葛亮-集成学习

- Bagging方法基于并行策略，基学习器之间不存在依赖关系，可实现并行训练。每一次迭代前，采用有放回的随机抽样来获取训练数据。



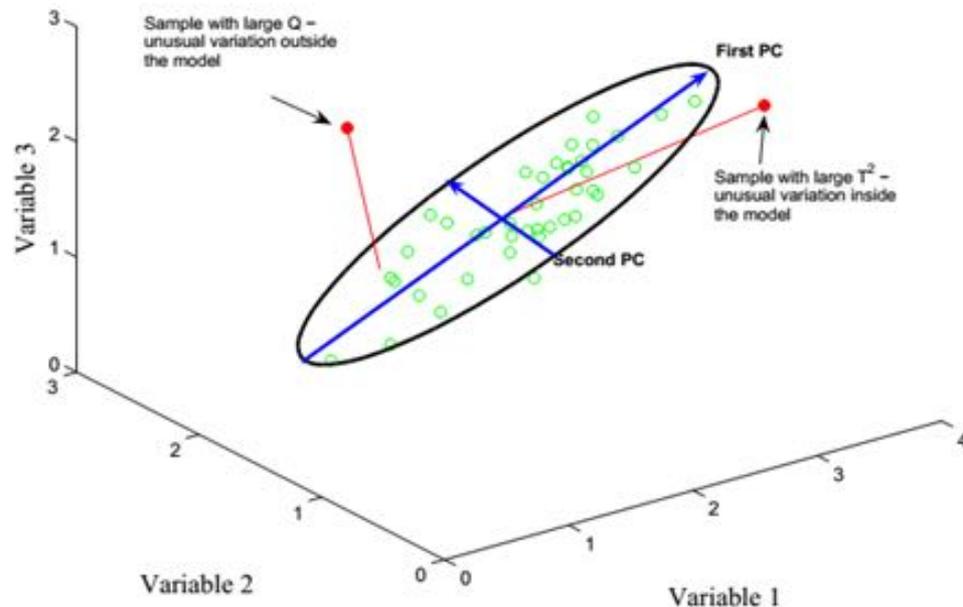
# 三个臭皮匠顶一个诸葛亮-集成学习

- 随机森林 (Random Forest) 算法构造多棵相互独立的CART决策树，形成一个森林，利用这些决策树共同决策输出类别。



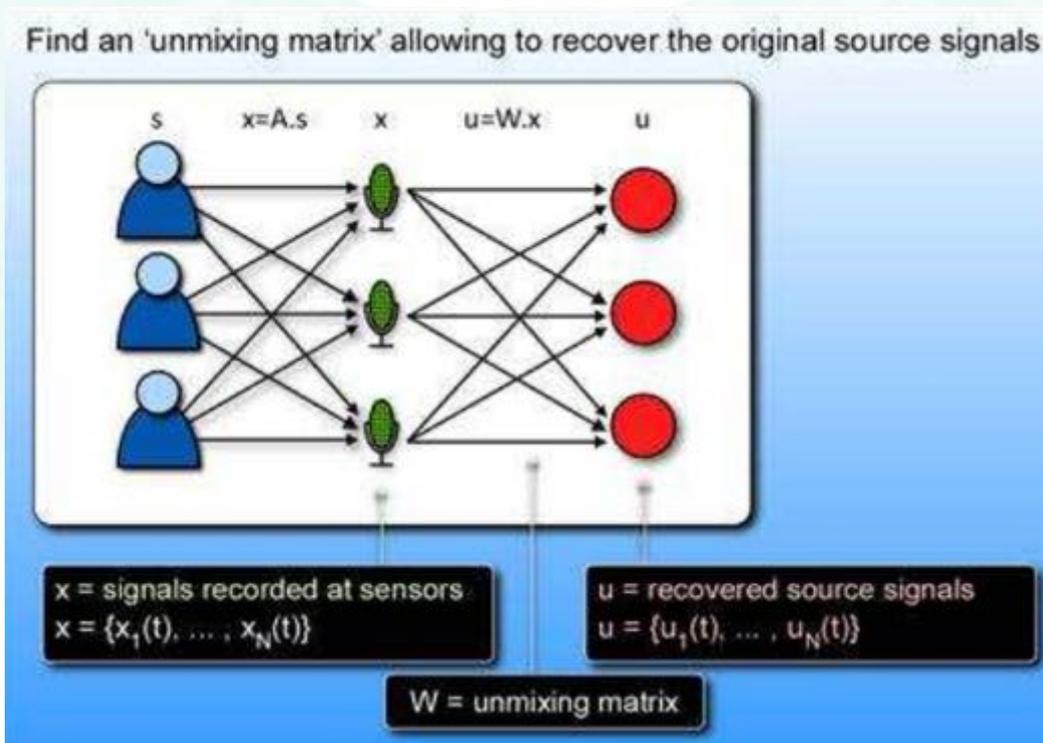
# 没有标注不可怕-无监督学习

- 主成分分析（Principal Component Analysis, PCA）是最常用的降维方法之一，通过正交变换将一组可能存在相关性的变量数据转换为一组线性不相关的变量（即主成分）。



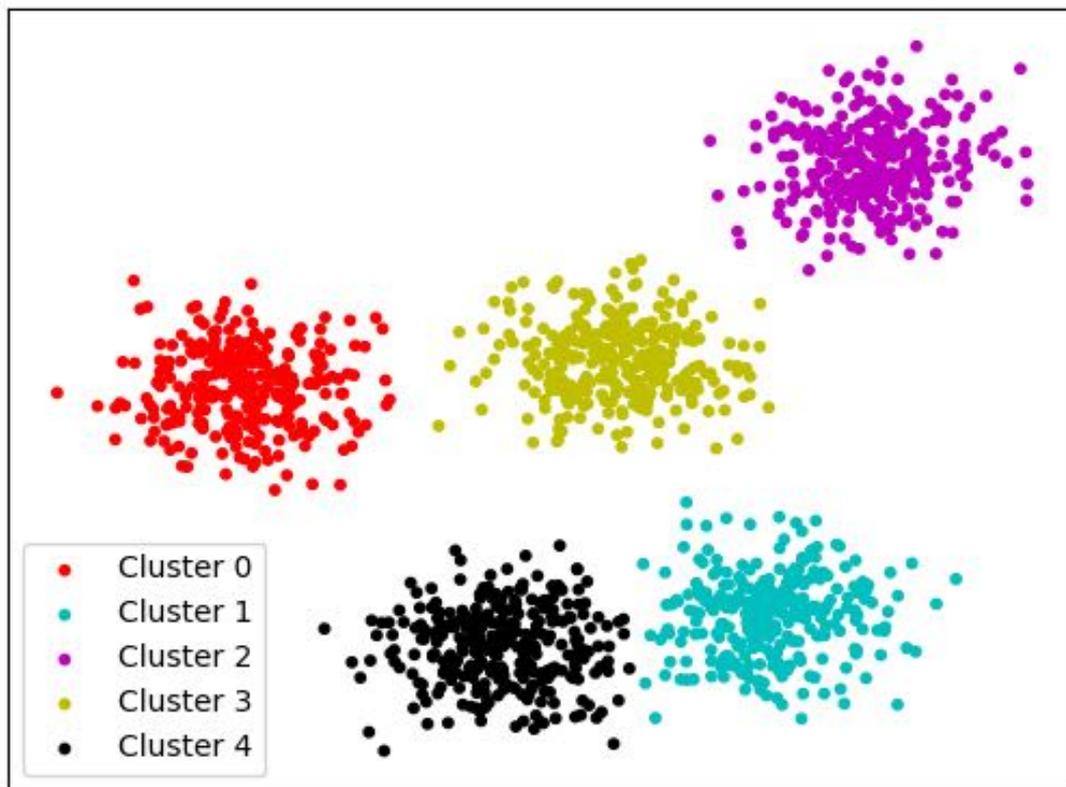
# 没有标注不可怕-无监督学习

- 独立成分分析（Independent Components Analysis, ICA）是一种线性变换，它把混合的数据或信号分离成统计独立的非高斯的信号源的线性组合。



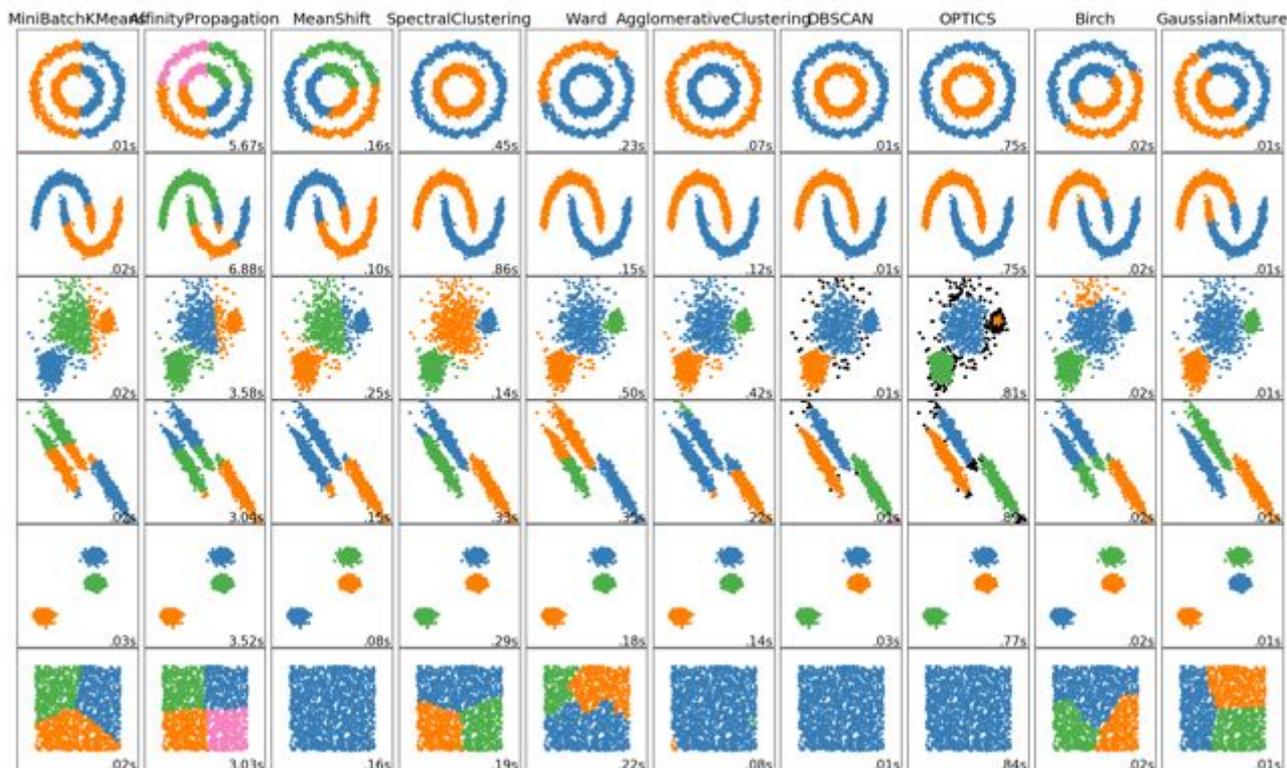
# 没有标注不可怕-无监督学习

- 聚类（Clustering）是将无标签样本集合中相似的样本分配到相同的类，不相似的样本分配到不同的类。



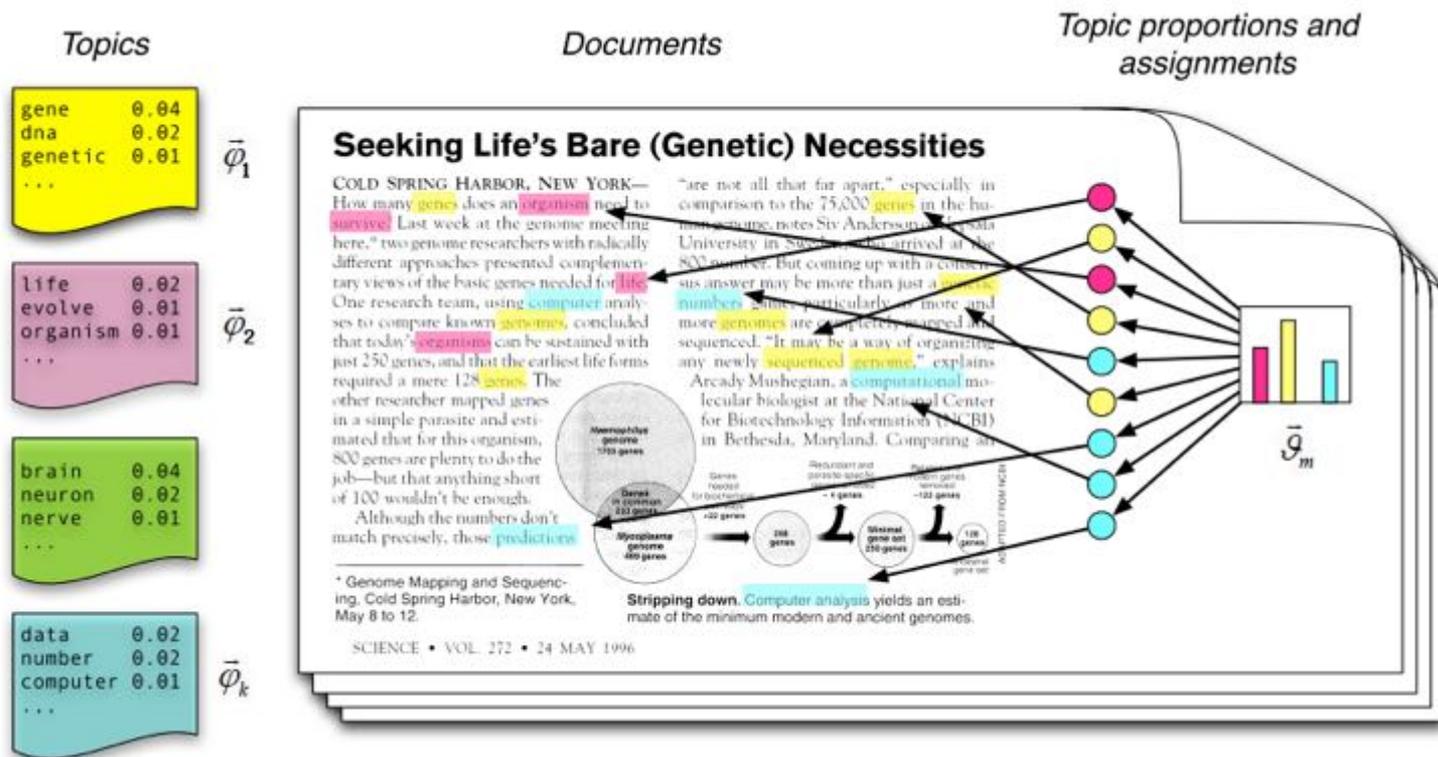
# 没有标注不可怕-无监督学习

- 常用的聚类方法有层级聚类（Agglomerative clustering），K均值聚类（K-means），基于密度的聚类（DBSCAN）等。



# 没有标注不可怕-无监督学习

- 话题模型 (Topic Models) 是对文本数据进行分析建模的一种技术。话题分析旨在发现文本中的话题，而话题则由单词的集合表示。



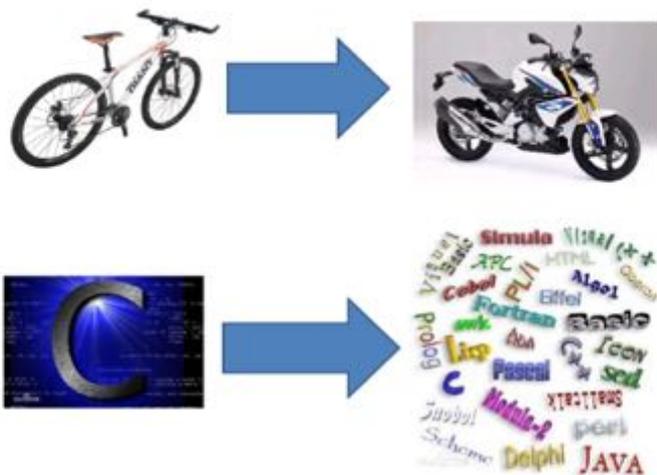


# 没有标注不可怕-无监督学习

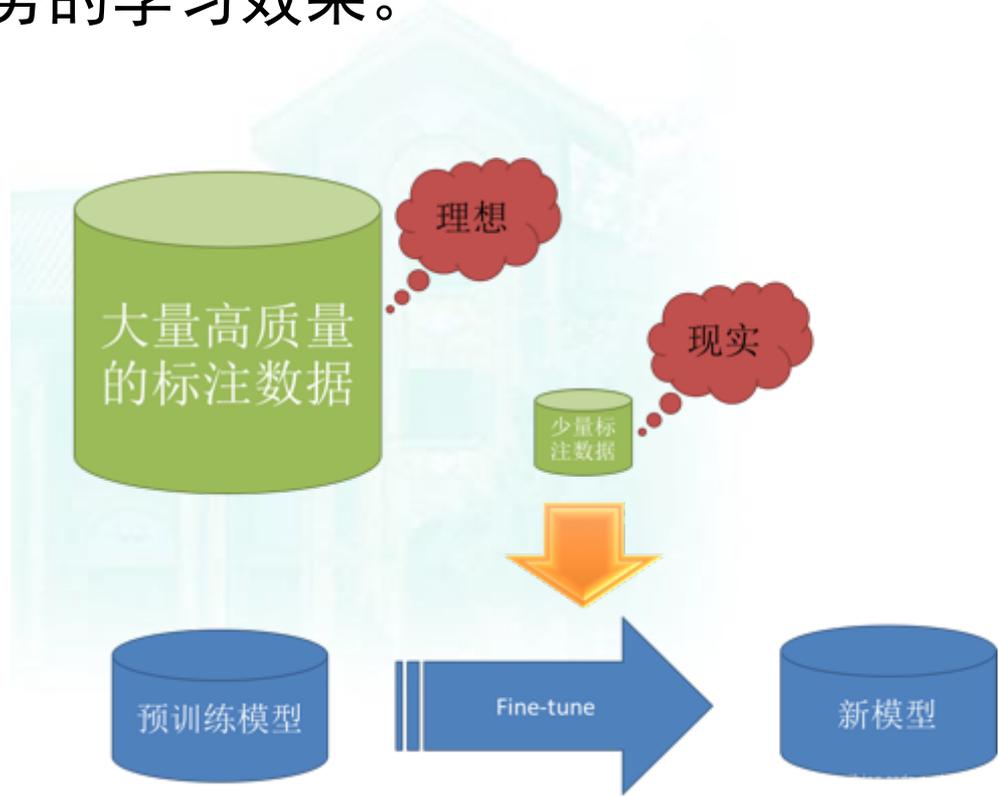
- 经典的Topic Models:
- LSA: 潜在语义分析模型, Latent Semantic Analysis
- PLSA: 概率潜在语义分析模型, Probabilistic Latent Semantic Analysis
- LDA: 隐狄利克雷分布模型, Latent Dirichlet Allocation

# 藏龙卧虎-泛弱监督学习

- 迁移学习 (Transfer Learning) 目标是将某个领域或任务上学习到的知识或模式应用到不同但相关的领域或问题中。主要思想是从相关领域中迁移标注数据或者知识结构、完成或改进目标领域或任务的学习效果。



能否像人类一样举一反三?



# 藏龙卧虎-泛弱监督学习

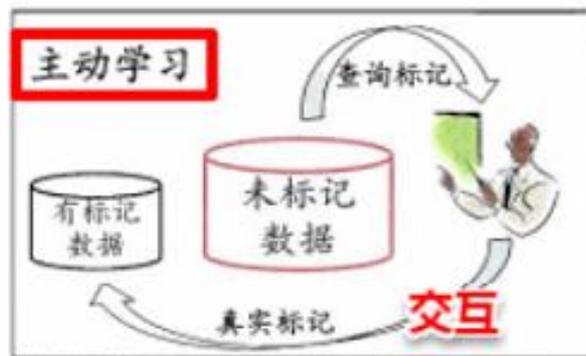
- 自监督学习 (Self-supervised learning) 自监督学习是没有人工标注标签的监督学习，可以将它看作没有人类参与的监督学习。标签仍然存在（需要有东西来监督学习过程），但它们是从输入数据中生成的，通常使用启发式算法生成的。



Let machines learn like humans

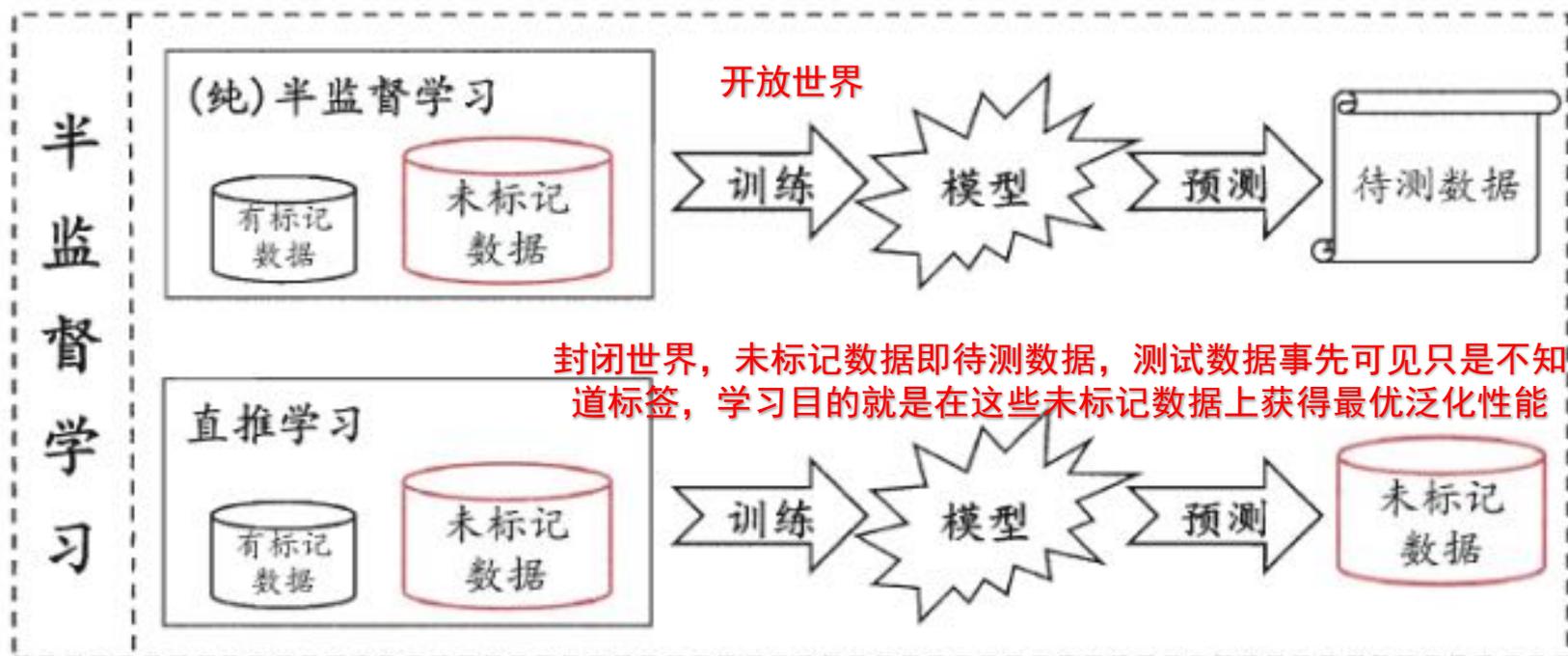
# 藏龙卧虎-泛弱监督学习

- 主动学习 (Active learning) 先使用有标记的样本数据训练出一个学习器，再基于该学习器对未标记的样本进行预测，从中挑选出不确定性高或分类置信度低的样本来咨询专家并进行打标，最后使用扩充后的训练集重新训练学习器。目标是使用尽量少的/有价值的咨询来获得更好的性能。



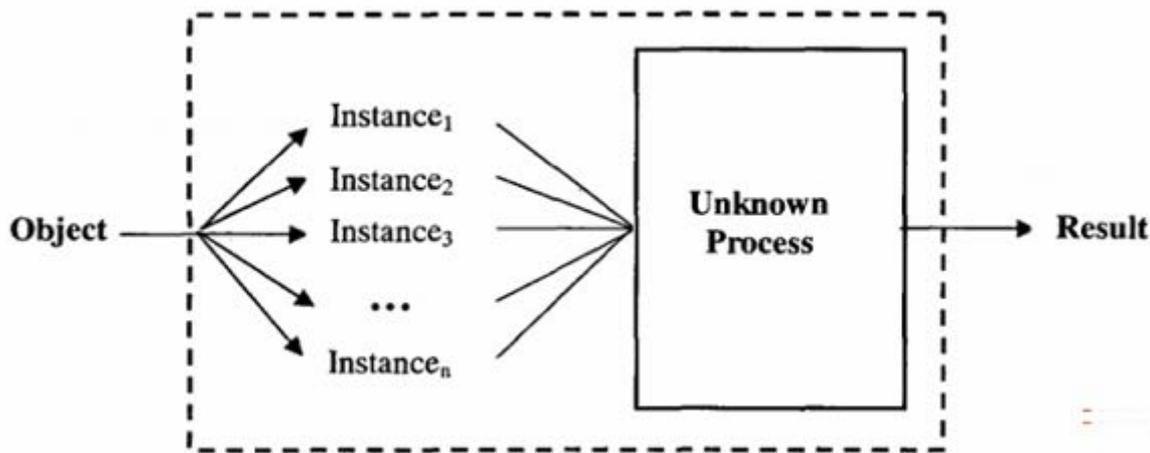
# 藏龙卧虎-泛弱监督学习

- 半监督学习 (Semi-supervised learning) 是指让学习器不依赖外界交互 (与主动学习不同), 自动地利用未标记样本提升学习性能的方法, 即训练集同时包含标记数据和未标记数据。



# 藏龙卧虎-泛弱监督学习

- 多示例学习 (Multiple-Instance learning) 针对数据只有粗粒度的标记信息的时候。将数据 (图片, 标签已知) 看成一个包含多个示例 (图中的物体, 标签未知) 的包。多示例学习的过程就是通过模型对包及其包含的多个实例进行分析预测得出包的标签。





# 藏龙卧虎-泛弱监督学习

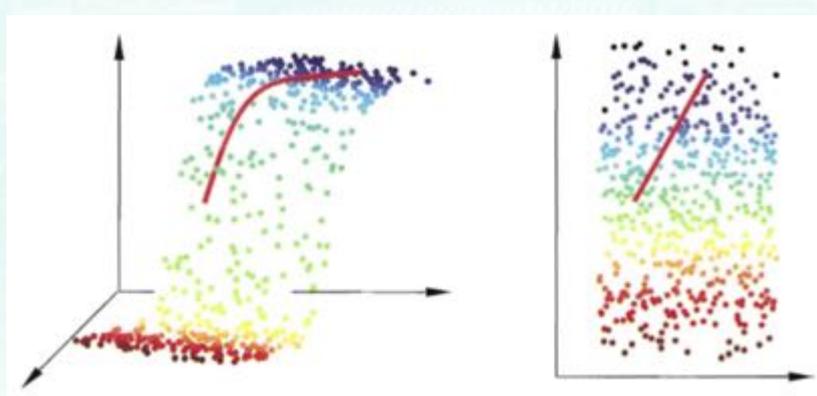
- 弱监督学习 (weakly-supervised learning) ---小结
  - 不完全监督：训练数据中只有一小部分数据有标签，而大部分数据没有标签，且这一小部分有标签的数据不足以训练一个好的模型 -----半监督学习、主动学习、迁移学习。
  - 不确切监督：即训练样本只有粗粒度的标签。比如一张图只有整张图的类别标注，而对于图片中的各个实体则没有标注。-----多示例学习
  - 不准确监督：给定的标签并不总是真值，比如人工标注的时候标错了。
  - 完全没有人工标注的标签-----自监督学习

# 边缘智能-数据降维与模型压缩

## □ 数据降维

降维（dimension reduction）是通过某种数学变换将原始高维属性空间转变为一个低维“子空间”（subspace）。

如下图所示，原始高维空间中的样本点，在这个低维嵌入子空间中更容易学习。



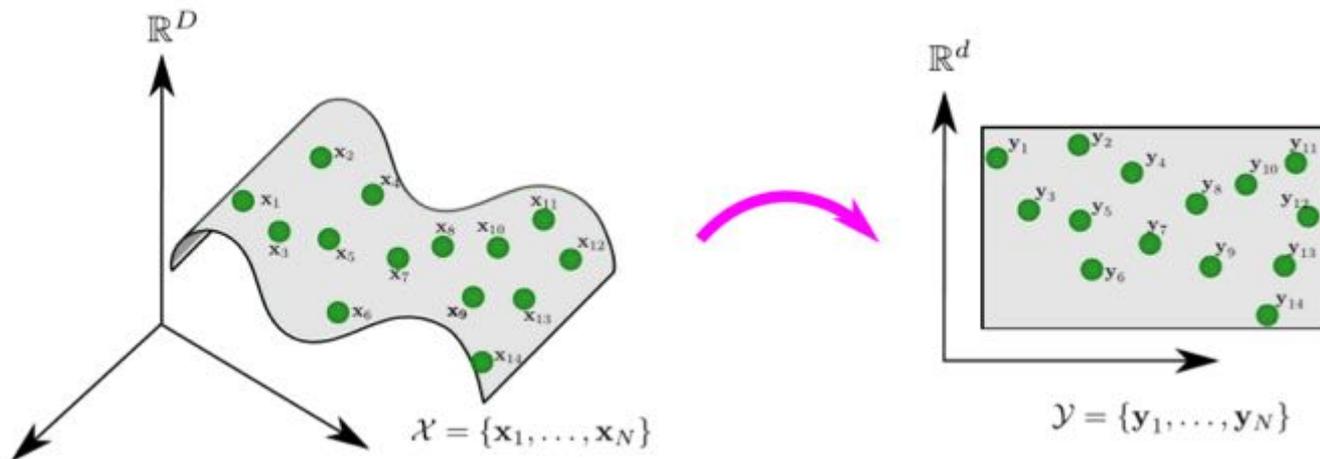
(a) 三维空间中观察到的样本点 (b) 二维空间中的曲面

低维嵌入示意图

# 边缘智能-数据降维与模型压缩

## □ 流形学习

流形学习 (manifold learning) 是一类借鉴了拓扑流形概念的降维方法。流形 (manifold) 在局部具有欧式空间的性质, 能用欧式距离来进行距离计算, 这给降维方法带来了很大的启发。



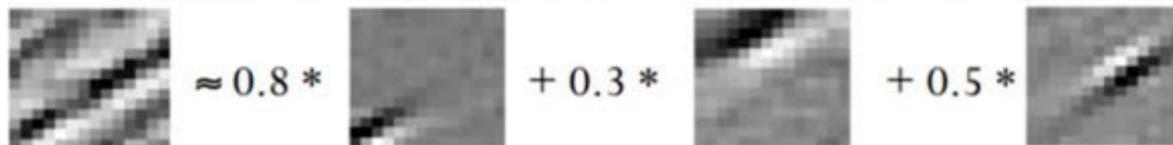
示意图：三维空间的数据与其对应的二维流形

# 边缘智能-数据降维与模型压缩

## □ 稀疏编码

稀疏编码 (sparse coding) 是一个线性因子模型, 已作为一种无监督特征学习和特征提取机制得到了广泛研究。

稀疏编码用于表示的形象描述如下:



Represent  $\mathbf{x}_i$  as:  $a_i = [0, 0, \dots, 0.8, 0, \dots, 0, 0.3, 0, \dots, 0, 0.5, \dots]$



# 边缘智能-数据降维与模型压缩

## □ 模型压缩

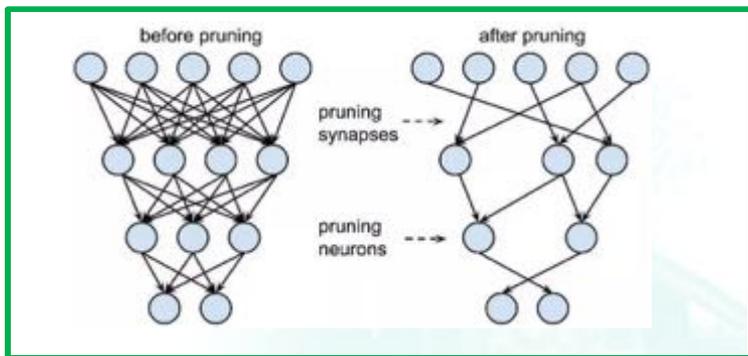
模型压缩的主要目的是降低网络模型存储和计算成本。

综合现有的深度模型压缩方法，它们主要分为四类：

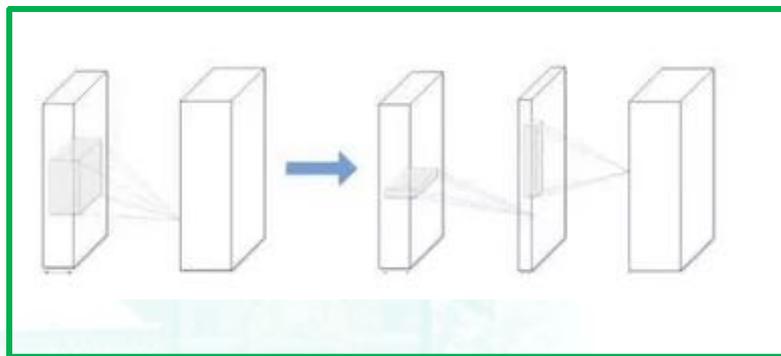
方法名称	描述	应用场景
参数修剪和共享 (parameter pruning and sharing)	删除对准确率影响不大的参数	卷积层和全连接层
低秩因子分解 (low-rank factorization)	使用矩阵对参数进行分解估计	卷积层和全连接层
转移/紧凑卷积滤波器 (transferred/compact convolutional filters)	设计特别的卷积核来保存参数	只有卷积层
知识蒸馏 (knowledge distillation)	训练一个更紧凑的网络来从大的模型蒸馏知识	卷积层和全连接层

# 边缘智能-数据降维与模型压缩

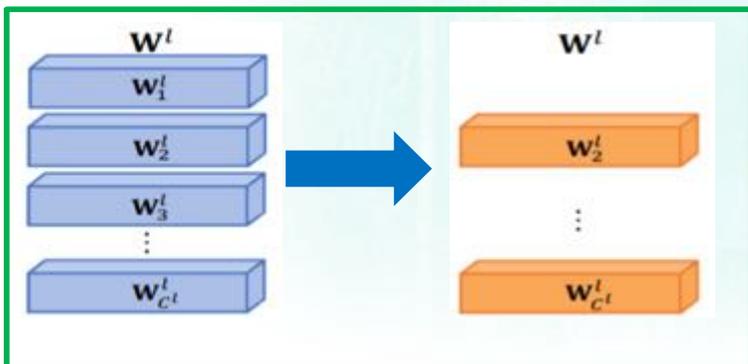
## 四种模型压缩方法简略示意图



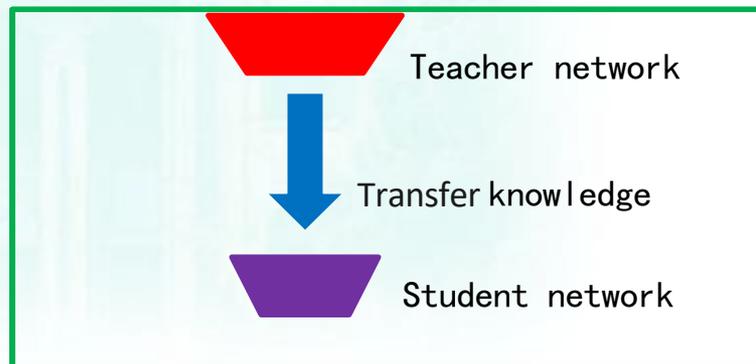
(a) 剪枝和共享



(b) 压缩2D卷积层的典型低秩方法



(c) 紧凑卷积滤波器

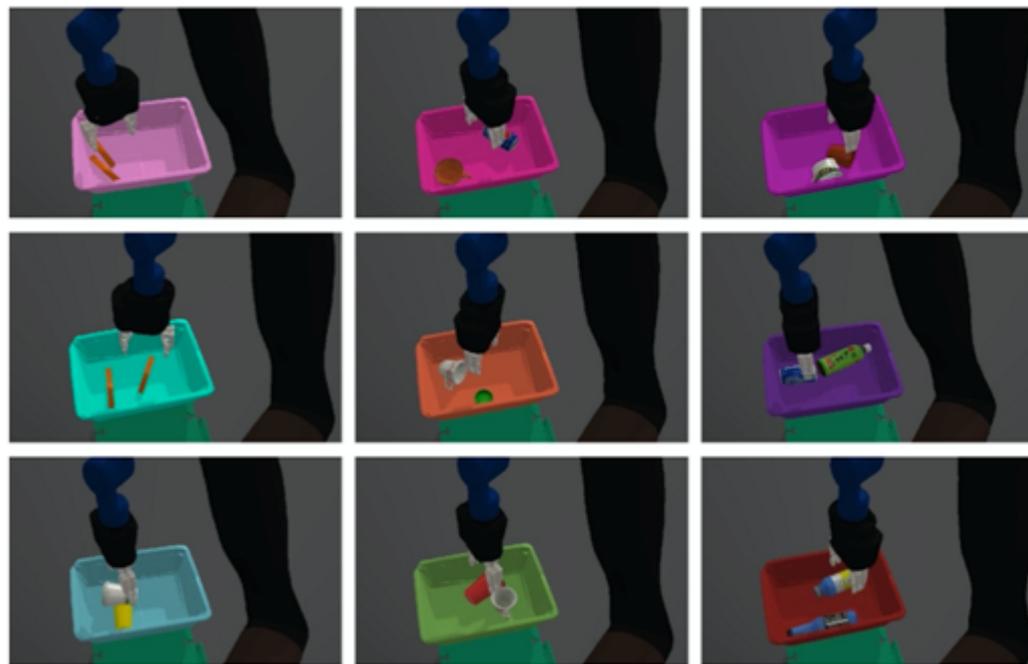
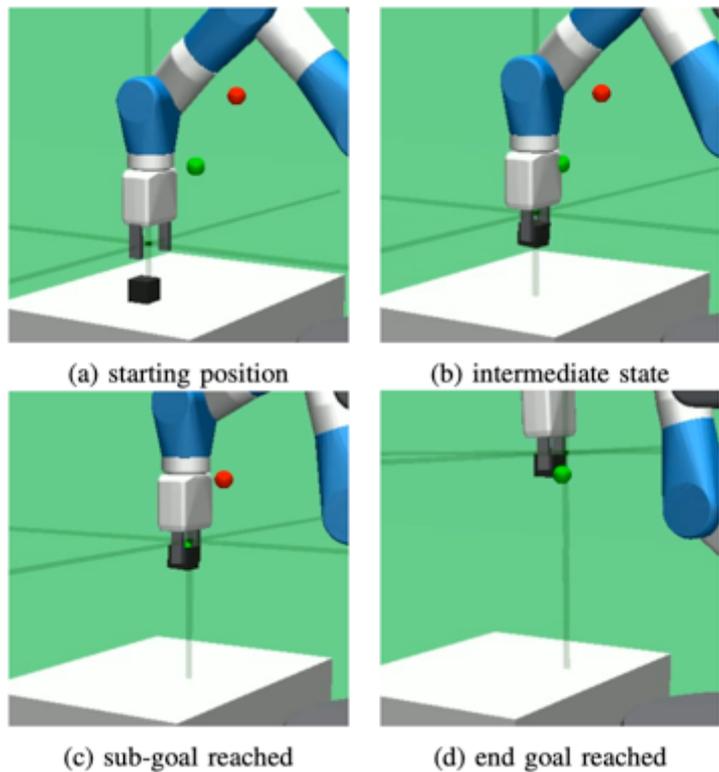


(d) 知识蒸馏



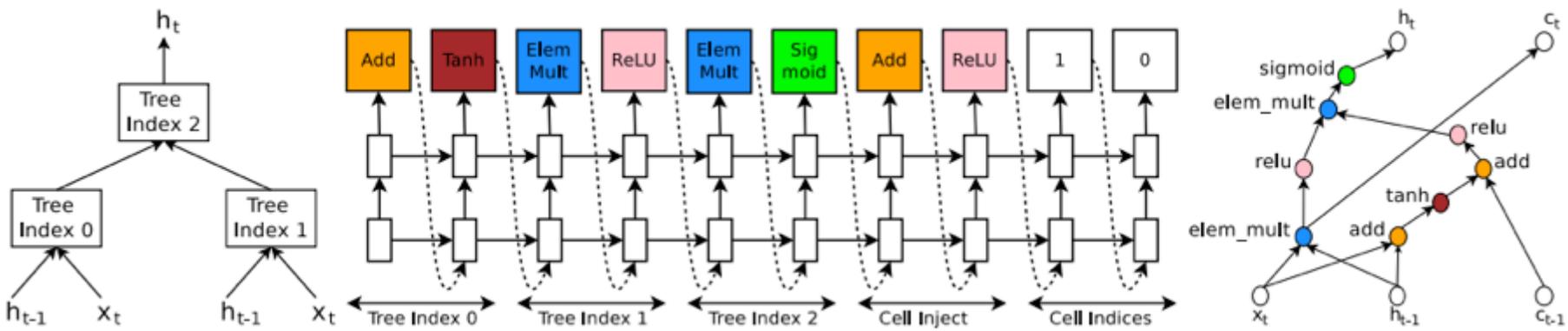
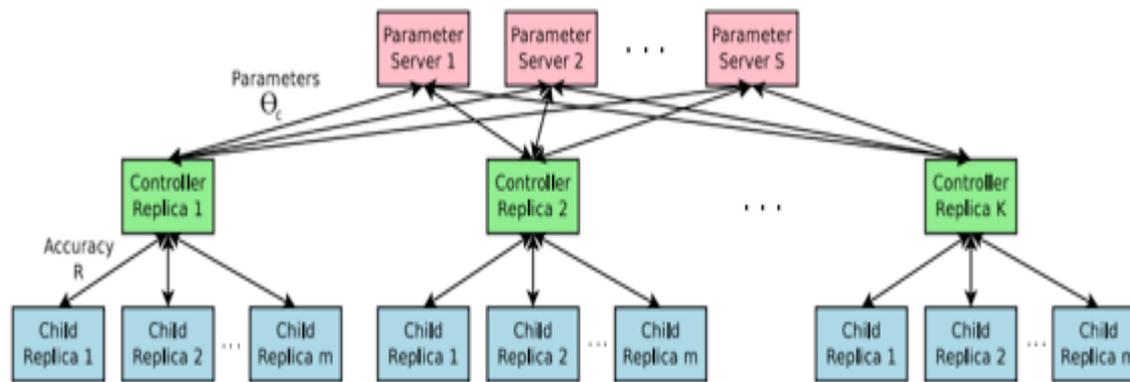
# 新的机器示范学习

## □ 模仿学习算法



# AutoML

## Network Architecture Search (NAS)



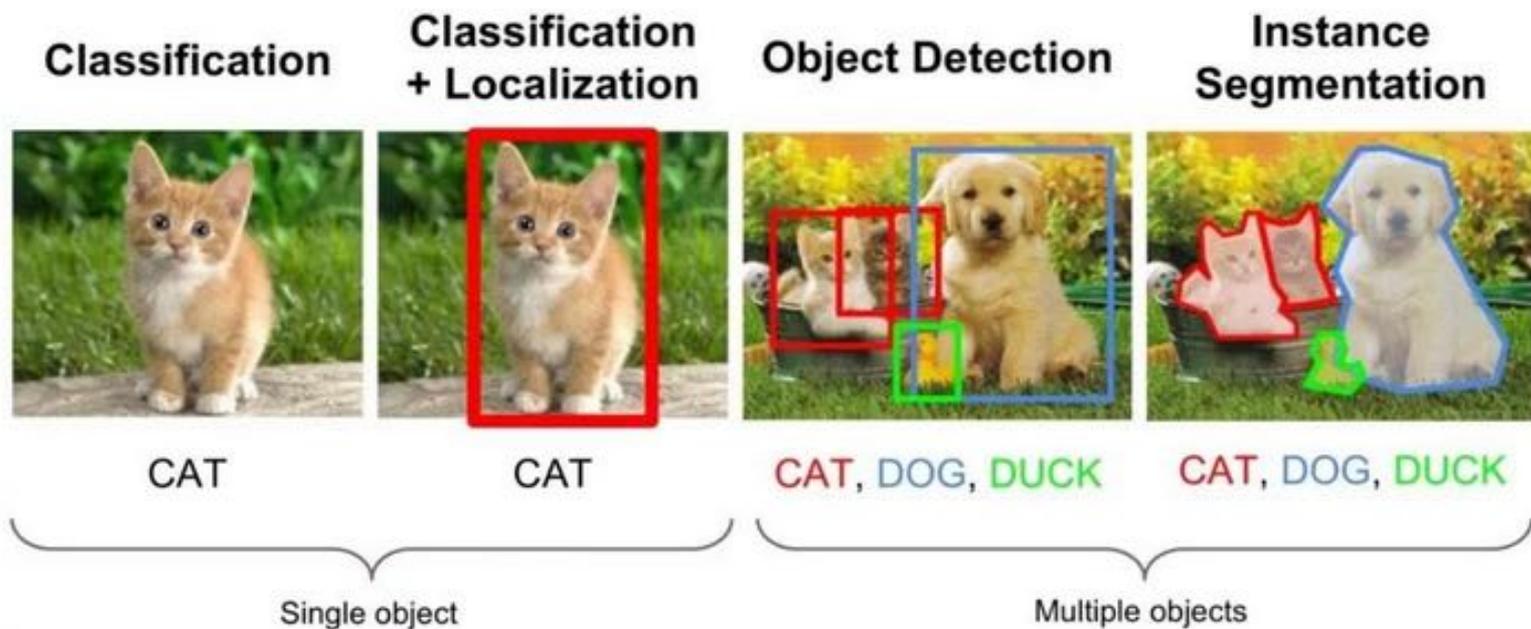


## 第三部分：应用技术

# 计算机视觉

## □ 检测

检测-Detection: 解决“是什么？在哪里？”的问题，即定位出感兴趣的物体的位置并判定目标的类别。



# 计算机视觉

## □ 分割

分割 -Segmentation：分为语义分割（Semantic Segmentation）和实例分割（Instance-level），解决“每个像素属于哪个目标物或者场景”的问题。

- 语义分割：语义分割在语义上理解图像中每个像素的角色



# 计算机视觉

## □ 分割

### ○ 实例分割

实例分割将不同的实例进行分类。



# 计算机视觉

## □ 追踪

- 追踪-Tracking: 在给定的视频序列初始帧的目标大小和位置的情况下，标注出后续帧中该目标的大小和位置。



# 计算机视觉

## 识别和检索

识别-Recognition: 即判断给定图像中物体的类别。

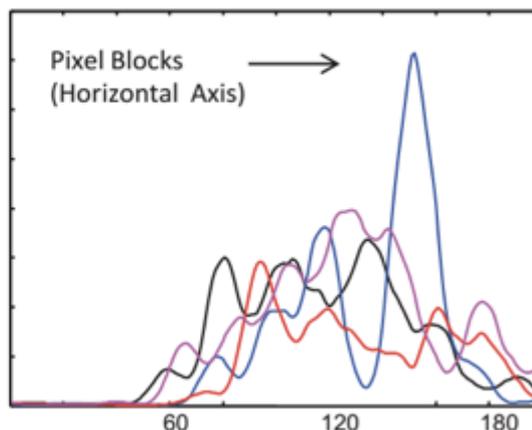
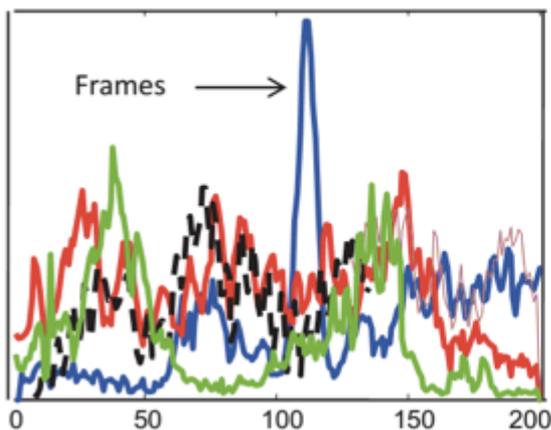
检索-Retrieval: 即从图像检索数据库中检索出与查询图片相似的图片。



# 数据挖掘

## □ 异常检测 (Anomaly detection)

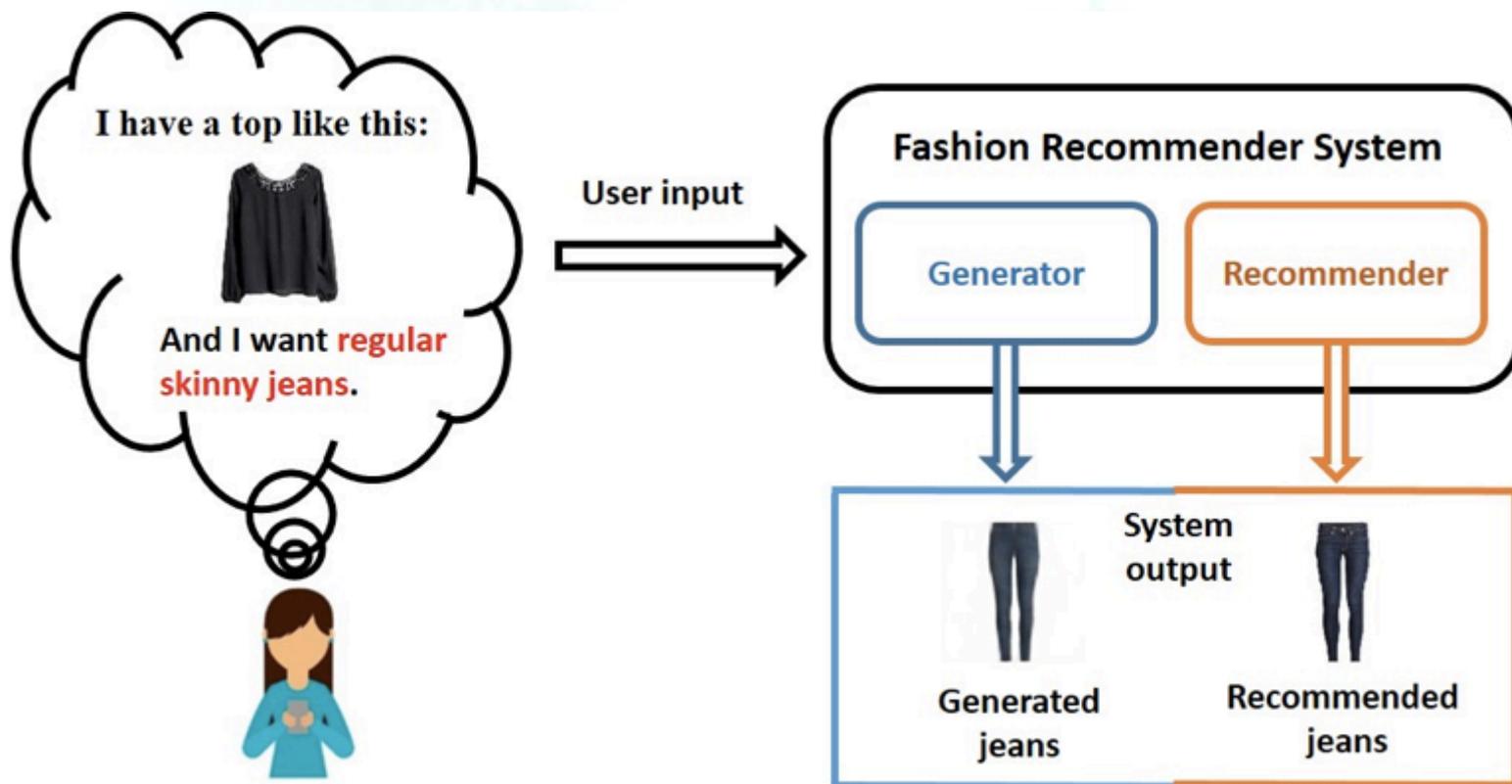
- 识别不寻常的数据记录，错误数据需要进一步调查。



# 数据挖掘

## □ 推荐系统 (Recommend System)

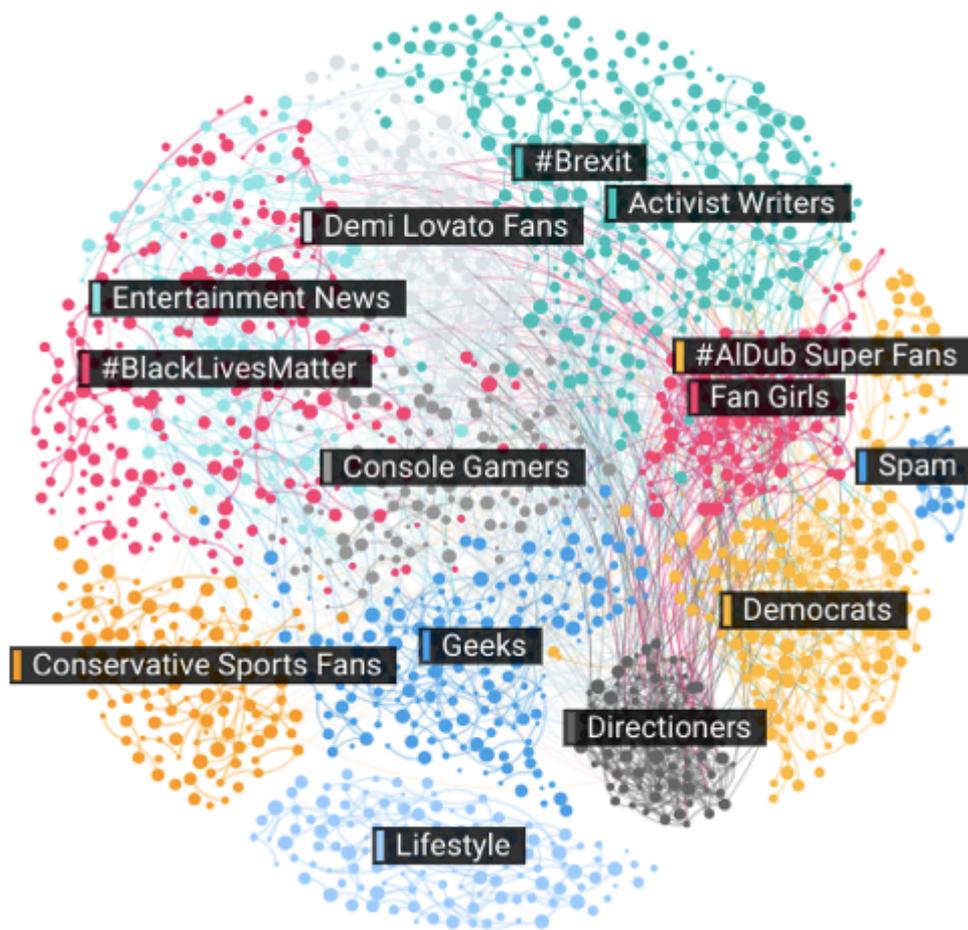
- 推荐系统是一种信息过滤系统，用于预测用户对物品的“评分”或“偏好”。



# 数据挖掘

## □ 聚类 (Cluster)

- 是在未知数据的结构下，发现数据的类别与结构。

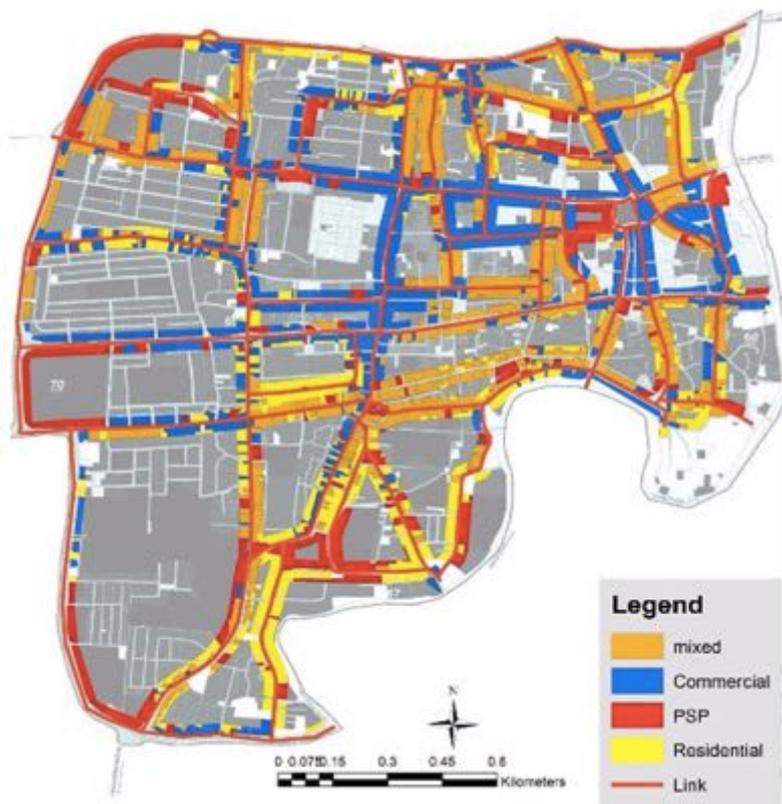




# 数据挖掘

## □ 预测 (Prediction)

### ○ 交通流量预测



# 自然语言处理

## □ 机器翻译



机器翻译，从一种语言到另一种语言的文本或语音的自动翻译，是NLP最重要的应用之一。



# 自然语言处理

## □ 信息检索

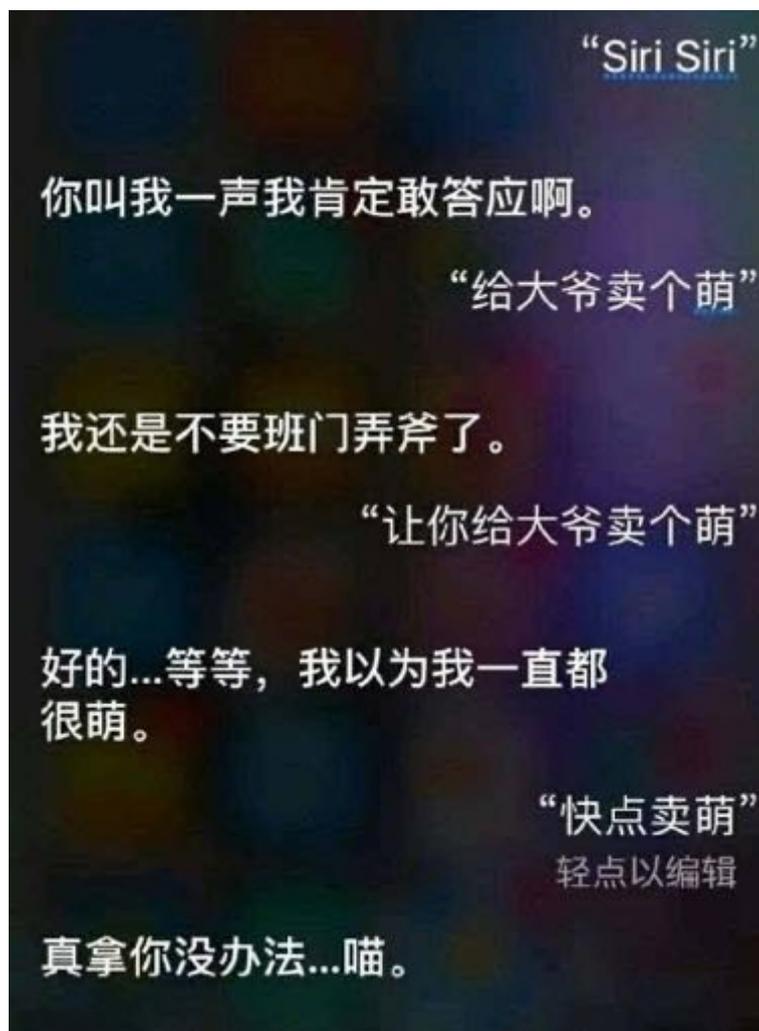


信息检索的基本原理是将用户输入的检索关键词与数据库中的标引词进行对比，当二者匹配成功时，检索成功。



# 自然语言处理

## □ 自动问答



自动问答是指利用计算机自动回答用户所提出的问题以满足用户知识需求的任务。自动问答系统在回答用户问题时，首先要正确理解用户所提出的问题，抽取其中关键的信息，在已有的语料库或者知识库中进行检索、匹配，将获取的答案反馈给用户。



# 自然语言处理

## 评论观点抽取

- 味道超级好吃，服务也不错~
- 服务员服务热情，味道不错，下次还来。
- 他家环境不错，服务也可以，味道也好!
- 是我家最近的一个价格非常实惠的餐厅了!
- .....

味道好(10) 服务不错(8) 环境还行(4) 价格便宜(8)

- 味道超级好吃，服务也不错~
- 服务员服务热情，味道不错，下次还来。
- 他家环境不错，服务也可以，味道也好!
- 是我家最近的一个价格非常实惠的餐厅了!
- .....

自动分析评论关注点和评论观点，并输出评论观点标签及评论观点极性，包括美食、酒店、汽车、景点等，可帮助商家进行产品分析，辅助用户进行消费决策。

# 自然语言处理

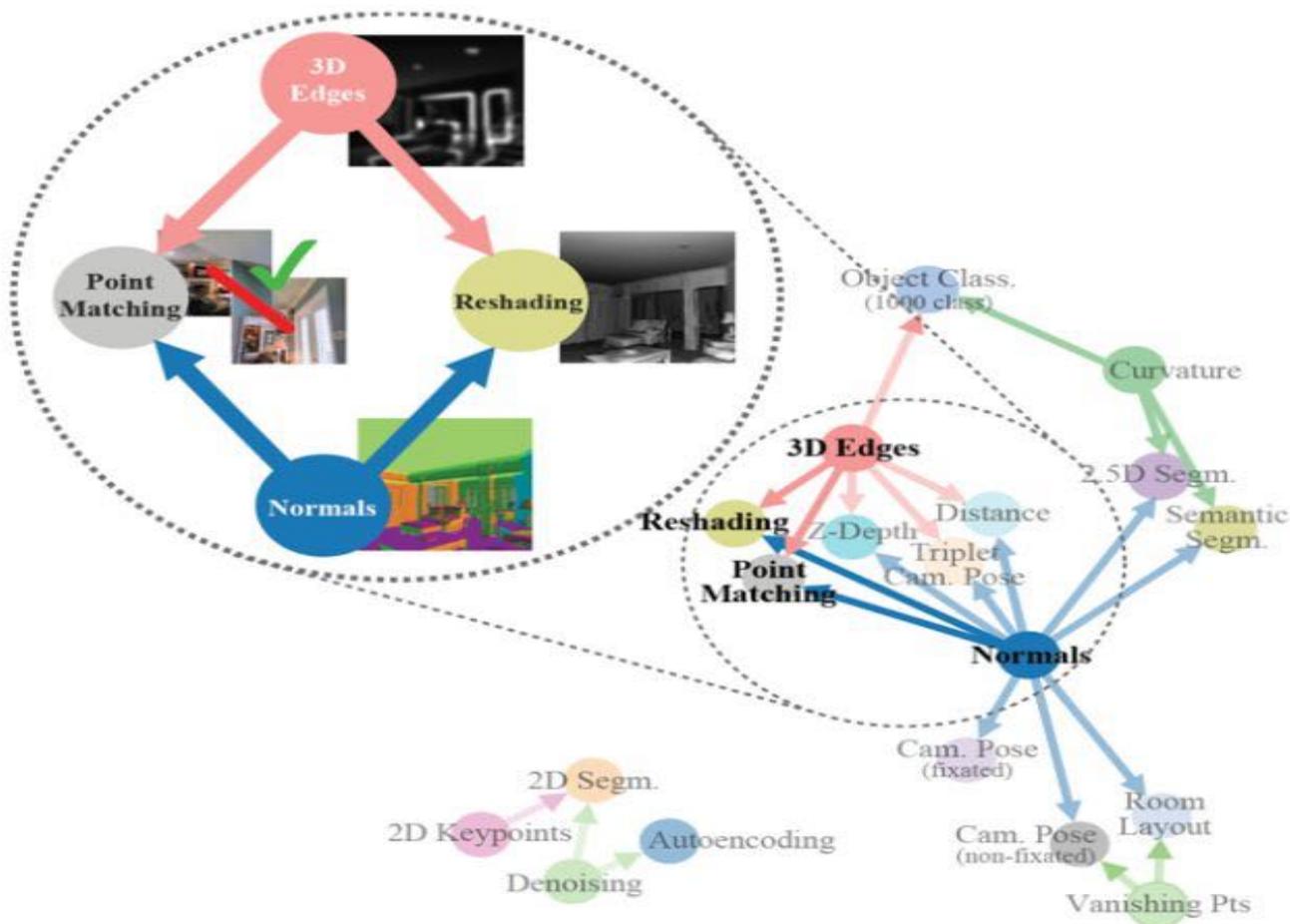
## 情感倾向分析



情感分析又称意见挖掘，是指通过计算技术对文本的主客观性、观点、情绪、极性的挖掘和分析，对文本的情感倾向做出分类判断。

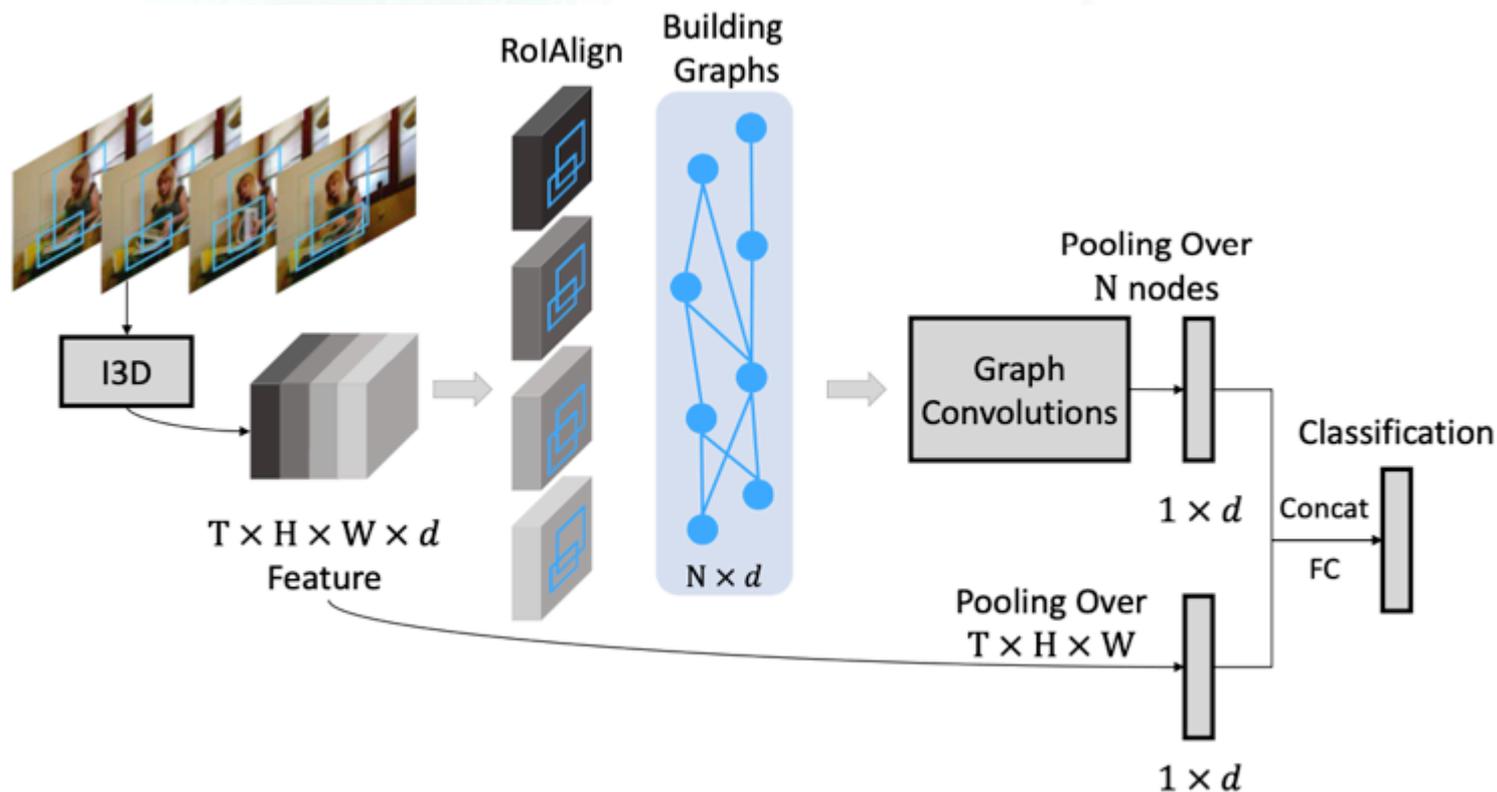
# 融汇共通

Taskonomy 量化不同视觉任务之间关联、并利用这些关联来最优化学习策略的研究。



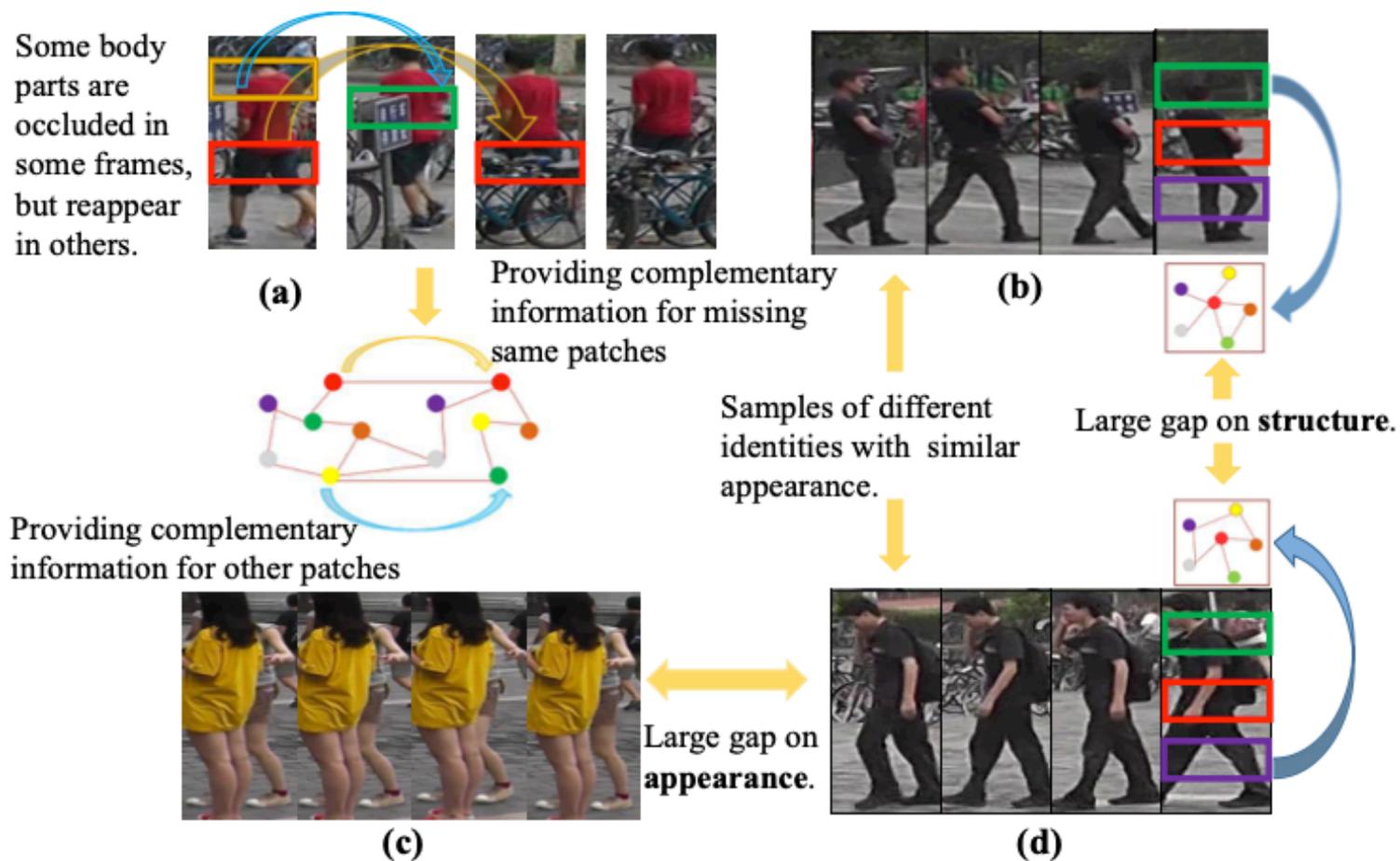
# 关联交互：人物交互

- 用图卷积模型，把不同视觉对象关联起来，形成推理网络



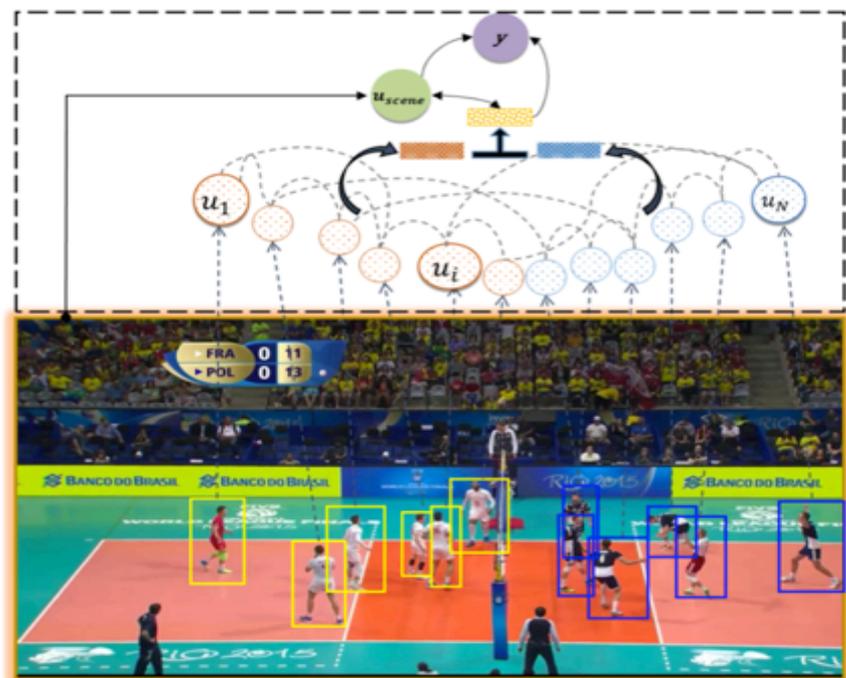
# 关联交互：身份识别

- 用图卷积模型，把不同视觉对象关联起来，形成推理网络



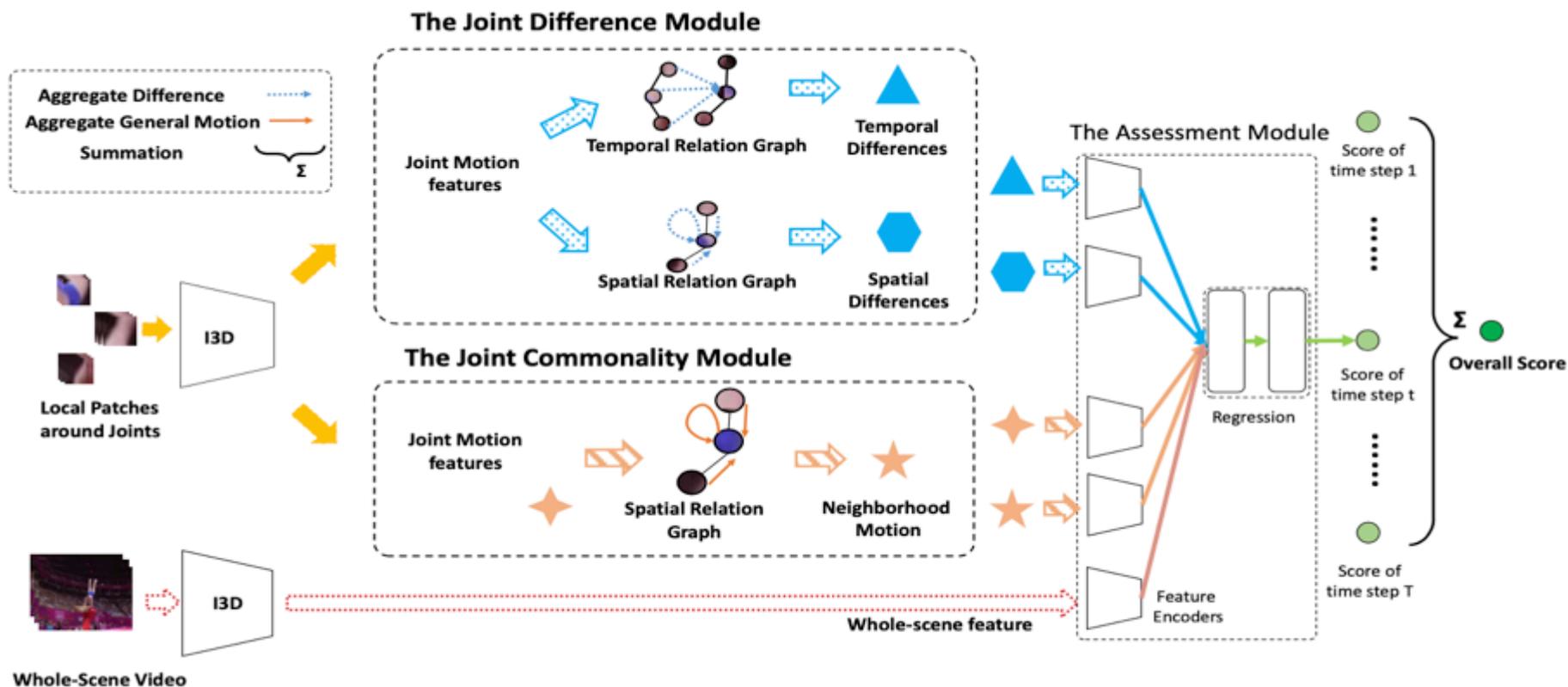
# 关联交互：复杂行为理解

- 用图卷积模型，把不同视觉对象关联起来，形成推理网络



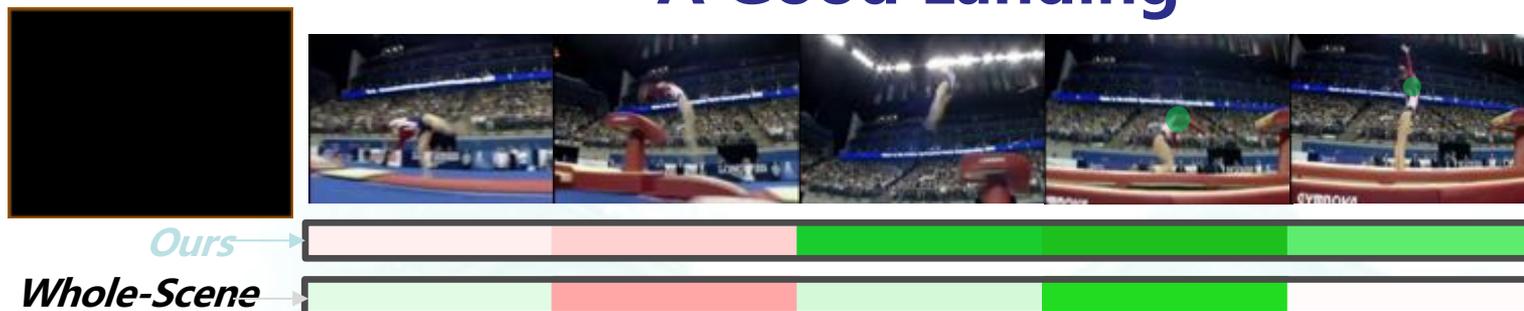
# 关联交互：行为评估

- 用图卷积模型，把不同视觉对象关联起来，形成推理网络

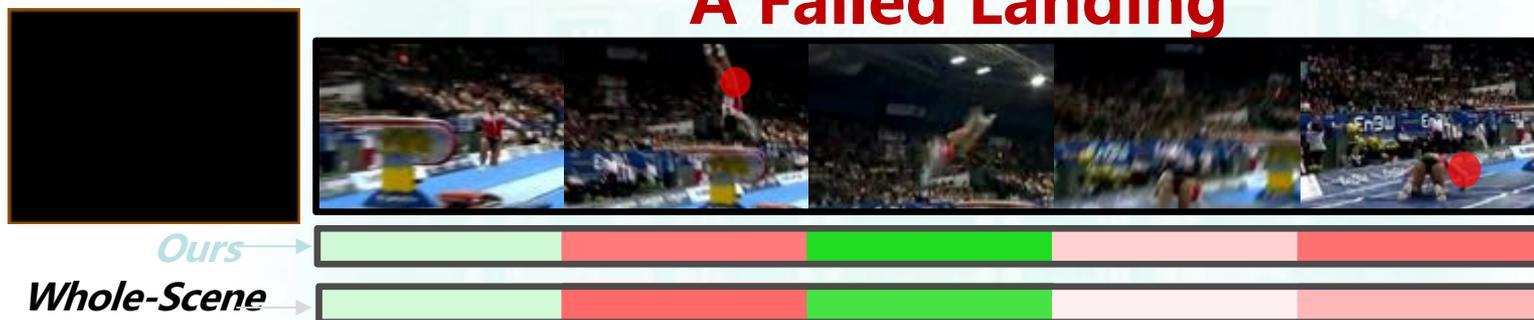


# 关联交互：行为评估

## A Good Landing



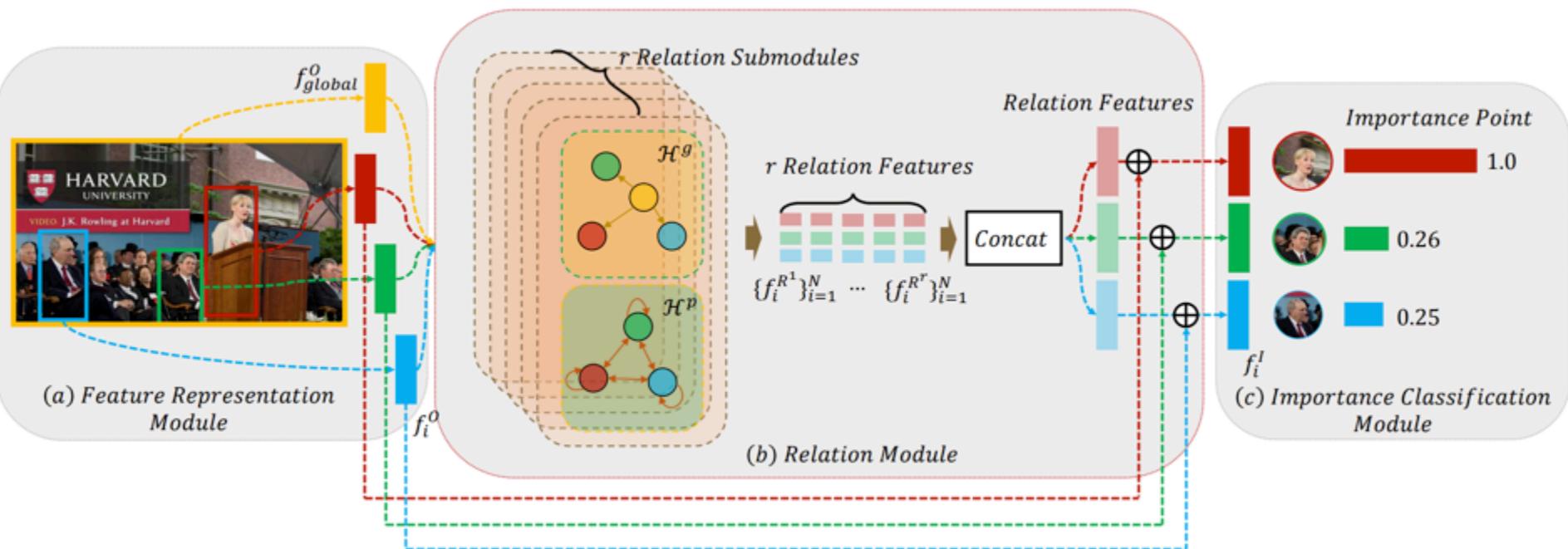
## A Failed Landing



Better

Worse

# 关联交互：理解焦点





# 参考用书

- ❑ 基础教材
  - R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 机械工业出版社
- ❑ 计算机视觉：
  - Computer Vision: Models, Learning, and Inference (作者: Simon J.D. Prince, 出版时间: 2012)
  - Computer Vision: A Modern Approach (作者: David A. Forsyth, Jean Ponce)
- ❑ 数据挖掘: 斯坦福大学的教材: Mining of Massive Datasets, <http://mmds.org/>
- ❑ Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999.

# 拥抱未来



如发现PPT观点和事实错误，请发我邮箱告知：[wzheng@ieee.org](mailto:wzheng@ieee.org)