

模式识别与计算机视觉

绪论

张振宇
南京大学智能科学与技术学院

2025



课程介绍

课程群和教学立方

✓ 课程群（昵称改为班级-姓名-学号）

✓ 助教：

- 郭俊文 (jwg9967@smail.nju.edu.cn)
- 周智勤 (zin@smail.nju.edu.cn)

✓ 教学立方（作业用）邀请码是**HDVR59H6**

- 1.请同学们在微信中搜索“教学立方”公众号并关注
2.点击公众号底部“学生”菜单 3.如未登录，请先注册/登录 4.点击课程列表页面中的“加入课程”输入邀请码即可加入。



群聊：2025春-模式识别与计算机视觉



该二维码7天内(2月24日前)有效，重新进入将更新

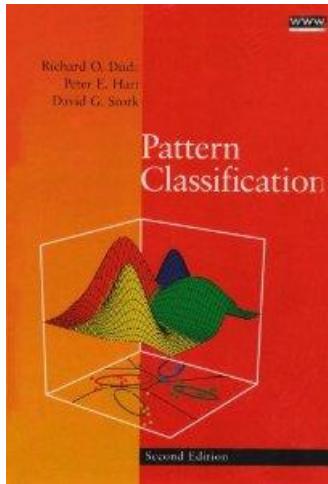
课程目标

- ✓ 理解、记忆模式识别和计算机视觉中的基本概念、
步骤和方法
- ✓ 对重要方法，要能实际应用，并能理解其前提条件、
应用范围、应用注意事项和方法原理及推导
- ✓ 对讲授的其他方法，要能理解其含义和使用环境
- ✓ 对模式识别&计算机视觉的前沿领域有感性的认识

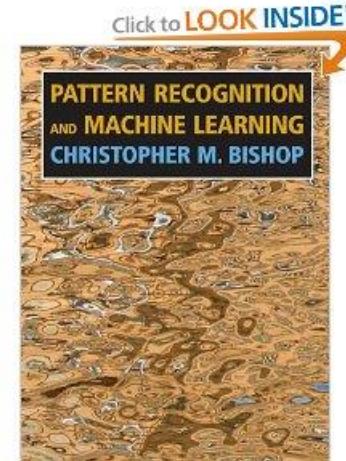
- ✓ 提高目标
 - 进一步能通过独立阅读和实践掌握较深入的问题和方法，并能应用到学习、研究中遇到的问题中去

参考书

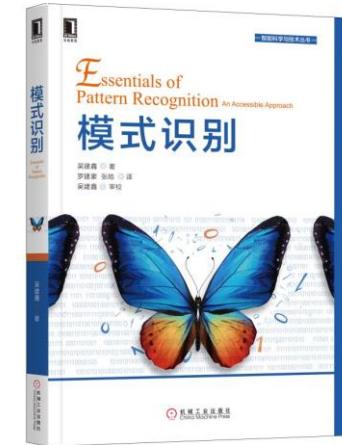
DHS



PRML



PR



参考书中译本：<http://item.jd.com/10057422.html>

实验环境

- ✓ 本课程中用到的编程，很多在Matlab环境中最方便
 - <http://cn.mathworks.com/help/matlab/>
 - Matlab简明教程：请自行搜索Matlab的tutorial
- ✓ Python当然是可以的
 - Scikit-learn
 - Numpy
 - OpenCV
 - 深度学习库

数学背景与框架

1. 模式识别的基本理论方法

- 前置课程：高等数学，线性代数，概率统计
- 需要的更多的数学知识会在授课中随时补充

2. 计算机视觉的核心方法和技术

- 帮助对有志CV研究的同学，熟悉和认识相关领域问题方法

3. 实践与前沿拓展

- 每章会提供关于进一步阅读的资源指南
- 通用资源：
 - 教科书
 - 参考书（DHS、PRML）
 - 维基（Wiki）
 - Wolfram MathWorld (<http://mathworld.wolfram.com/>)

课程总体内容

- 缇论
- 基础数学工具
- 最近邻方法
- 主成分分析
- 特征归一化和线性判别分析
- 概率方法
- 神经网络与优化方法
- 稀疏表示和数据对齐
- 图像滤波和特征提取
- 角点检测和2D变换
- 相机模型
- 图像分类、检测、分割
- 双目视觉
- 光流&跟踪
- 神经场方法
- 生成模型

课程考试考核要求

1. 平时成绩 (10%)

- 教学立方的加入和活跃情况
- 点名、签到情况

2. 作业 (40%)

- 习题
- 实验性练习

3. 考试 (50%)

- 基础知识点
- 跟作业相关的

PR & CV简介

什么是模式识别？

什么是计算机视觉？

什么是模式识别?

- ✓ “**Pattern recognition** is a field in machine learning.”
 - 取自2013年12月25日，2016年2月9日仍然如此
 - Pattern recognition (**psychology**), identification of faces, objects, words, melodies, etc.
 - [http://en.wikipedia.org/wiki/Pattern_recognition_\(disambiguation\)](http://en.wikipedia.org/wiki/Pattern_recognition_(disambiguation))
- ✓ In machine learning, pattern recognition is the **assignment** of a **label** to a given **input value**.
 - http://en.wikipedia.org/wiki/Pattern_recognition
- ✓ PRML: **automatic discovery** of **regularities** in **data** ... and with **the use of these regularities** to take actions such as classifying the data into different categories

模式识别的特征

✓ 输入： 数据

- data, input value, ...

✓ 输出：“模式” (Pattern)

- faces, objects, words, melodies, (classification) label, categories, regularity, ...

✓ 输入转换到输出：

- 由计算机系统完成(automatic, algorithm, assignment ...)
- 是一个困难但有趣的过程(discovery)

什么是计算机视觉？



人眼观察世界所看到的

什么是计算机视觉？

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

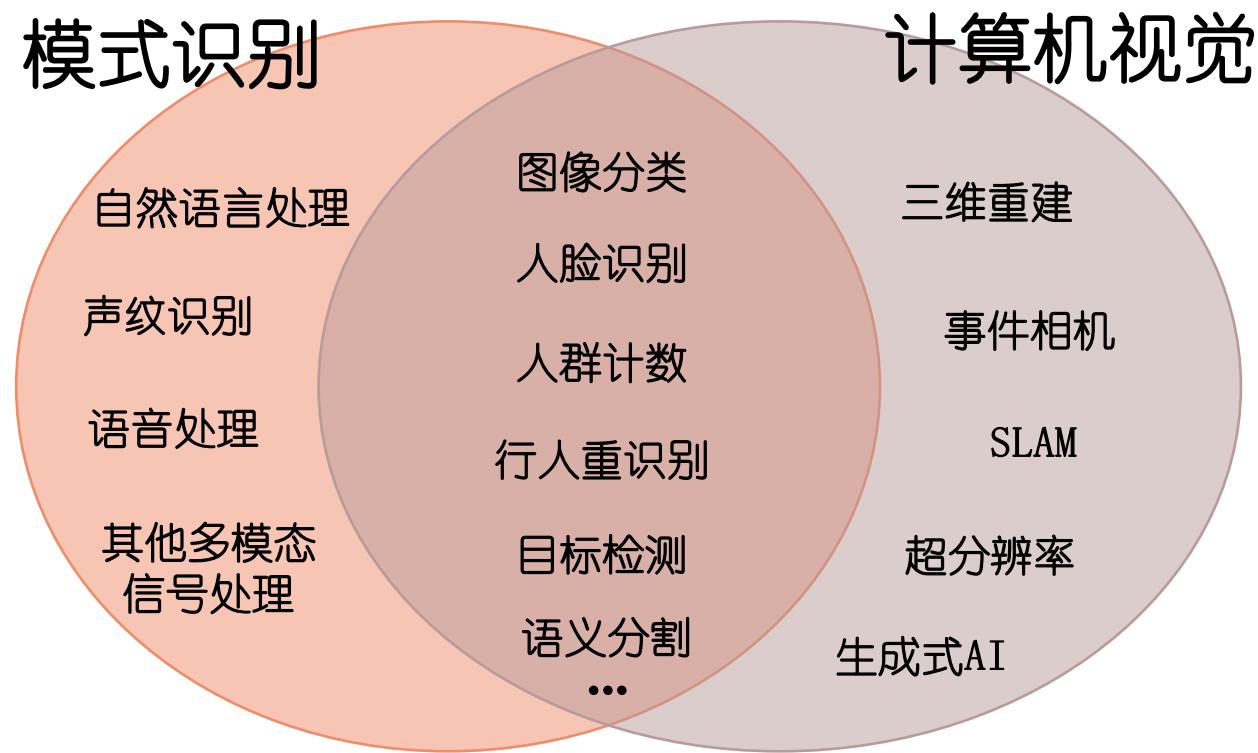
计算机所“看”到的

什么是计算机视觉？

- ✓ 计算机视觉的目标就是赋予计算机（超越）人类水平的感知能力。

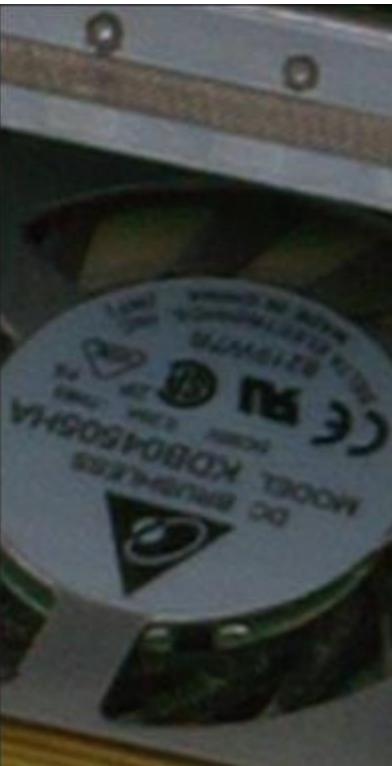
The screenshot shows the InternVL AI interface. At the top left is the logo and name "InternVL". Below it is a button "+ New". The main area has a heading "你可以描述一下这个场景吗" (Can you describe this scene?) followed by a truncated text "can you describe this sc...". To the right, there is a question "你可以描述一下这个场景吗" (Can you describe this scene?) above a generated image of a building facade with bicycles and flowers. Below the image is a detailed description of the scene, starting with a blue AI icon and the text: "这张图片展示了一个装饰着花卉的建筑外墙, 墙上有两扇绿色百叶窗, 并且窗户下面和旁边的墙上摆放了许多红色和粉色的花朵。墙壁是红砖色与米黄色相间的设计。在前景中, 可以看到三辆自行车停靠在一起, 这些自行车分别有不同的颜色和款式, 其中一辆带有前篮子。左边的背景里有一个小门, 门旁有一个石阶通向更高的地面或平台, 在这道门的背后是一片绿意盎然的小花园区域, 显得非常生机勃勃。总体来说, 这幅图像给人一种宁静、舒适的感觉, 花坛和自行车的存在使得整个画面充满了生活气息。"/>

模式识别、计算机视觉的相互关系



模式识别vs.计算机视觉

- ✓ 模式识别与计算机视觉(computer vision)的研究和应用有**非常多的重合**
 - 识别recognition是计算机视觉中最重要的问题之一
 - 模式识别中很大部分输入是图像
- ✓ 模式识别包含很多视觉以外的问题
 - 音频、雷达、文本、...
- ✓ 计算机视觉包括很多识别以外的问题
 - 如，超分辨率super-resolution
 - 三维重建3D reconstruction



[http://upload.wikimedia.org/wikipedia/commons/e/ed/An example of super resolution with still RAW photo.jpg](http://upload.wikimedia.org/wikipedia/commons/e/ed/An_example_of_super_resolution_with_still_RAW_photo.jpg) (链接已失效)

<http://grail.cs.washington.edu/rome/>

模式识别vs.计算机视觉

✓ 早期依赖模式识别

- 20世纪的计算机视觉主要依赖手工设计特征（如边缘检测）加分类器，与模式识别高度重叠。

✓ 深度学习时代的分化与融合

- 深度学习（如CNN）在视觉中取得突破后，计算机视觉逐渐形成独立的研究体系，但其底层技术（如特征学习）仍属模式识别的演进。

模式识别vs.计算机视觉

✓ 模式识别是方法论

- 作为通用技术，服务于多个领域（如语音、文本、视觉）

✓ 计算机视觉是应用领域

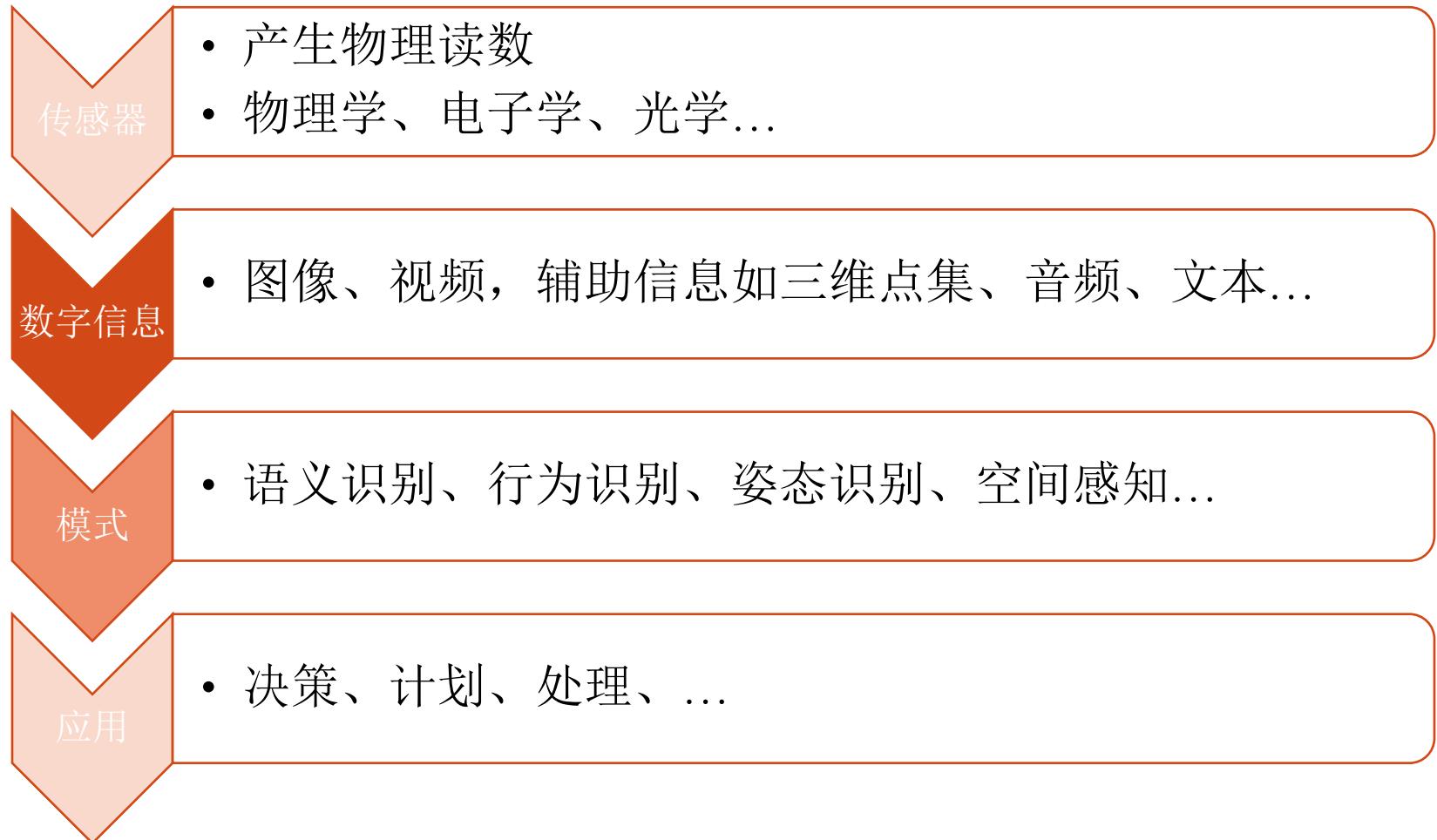
- 融合模式识别、图像处理、图形学、神经科学等多学科知识，目标是实现“视觉智能”。

PRCV系统举例：Autopilot, FSD

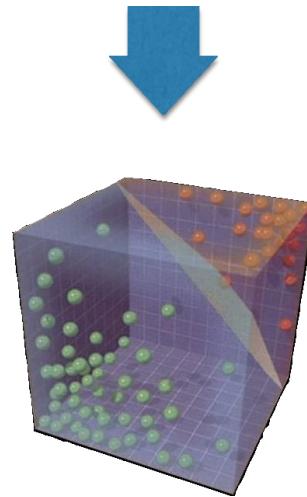


- ✓ 输入？ 视频来源
- ✓ 输出？ <https://www.tesla.cn/models>
- ✓ 困难？ Autopilot主页
<https://www.tesla.cn/autopilot?redirect=no>

从数据到模式 From data to pattern



从数据到模式 From data to pattern **representation**



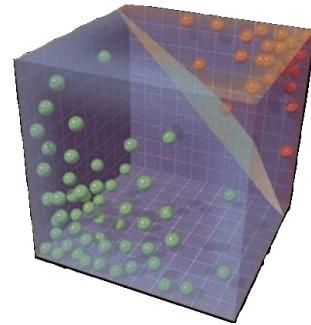
'fancy math'



output

从数据到模式 From data to pattern

representation

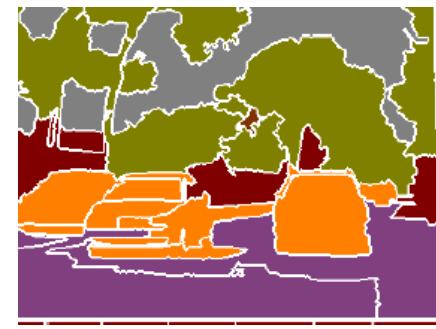


'fancy math'



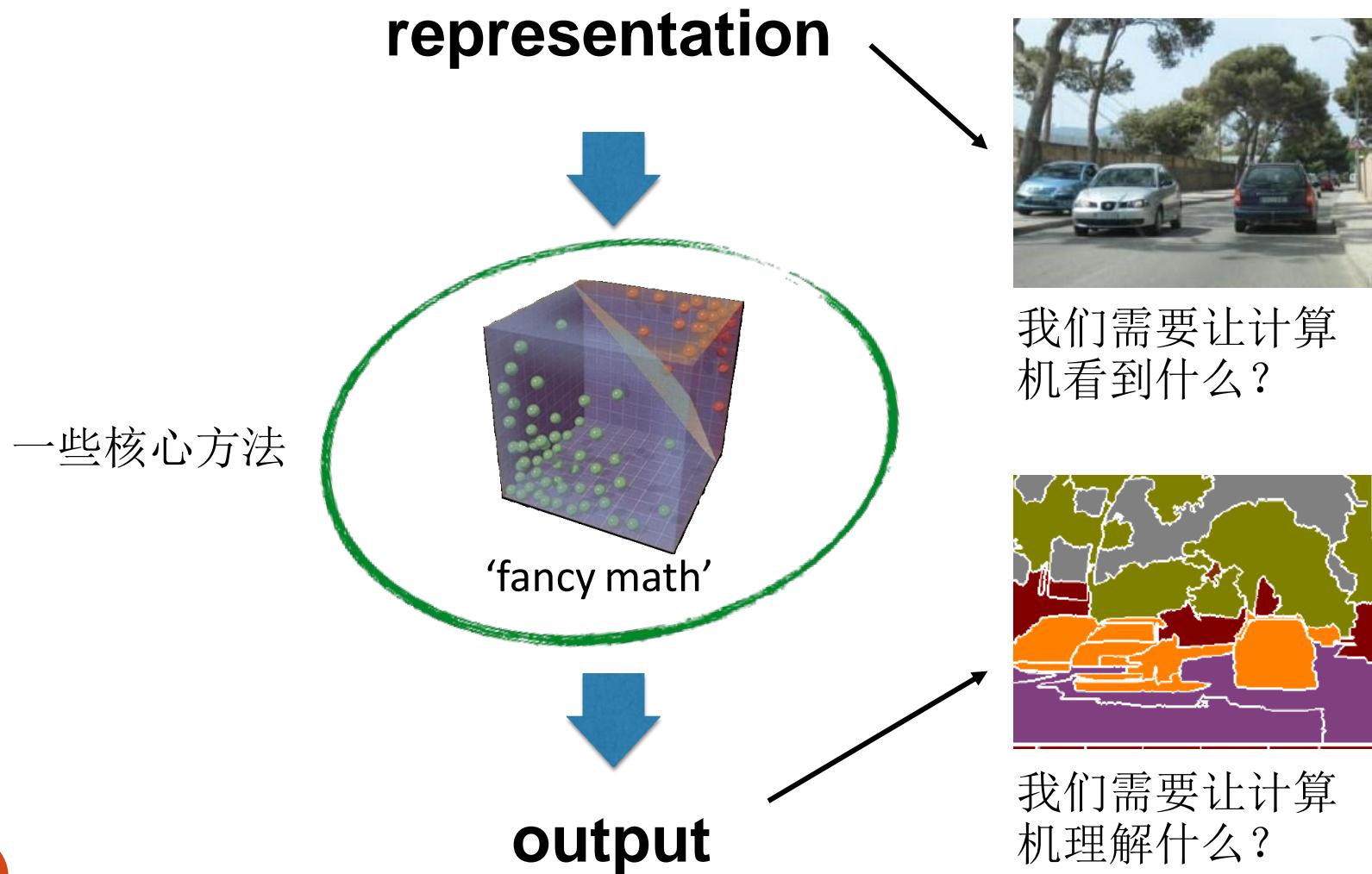
我们需要让计算
机看到什么？

output



我们需要让计算
机理解什么？

从数据到模式 From data to pattern



(虚拟例子) 性别识别：特征

- ✓ 选取什么**数据**来判别一个人的性别?
 - 长相? 喉结?
 - 身高?
 - 体重?
 - 声音?
 - 行为?
 - ...
- ✓ 被**使用**的数据或从原始输入数据(raw input data)中提取(extract)的数据称为**特征**(feature)
 - 为什么要选取、提取、或者**学习**特征?

(虚拟例子) 性别识别：评估

✓ 怎么知道模式识别的结果好坏？

- 评估或评价(evaluate)
- 通常是把模式识别系统的输出(称为预测, prediction)
与真实值(groundtruth)进行比较

✓ 真实值从哪里来？

- 女性的性染色体 $XX \rightarrow X$ (来自母亲) + X (来自父亲)
- 男性的性染色体 $XY \rightarrow X$ (来自母亲) + Y (来自父亲)

✓ 如何进行比较？

- 以后会讲。现在先想一想？

(虚拟例子) 性别识别：转换

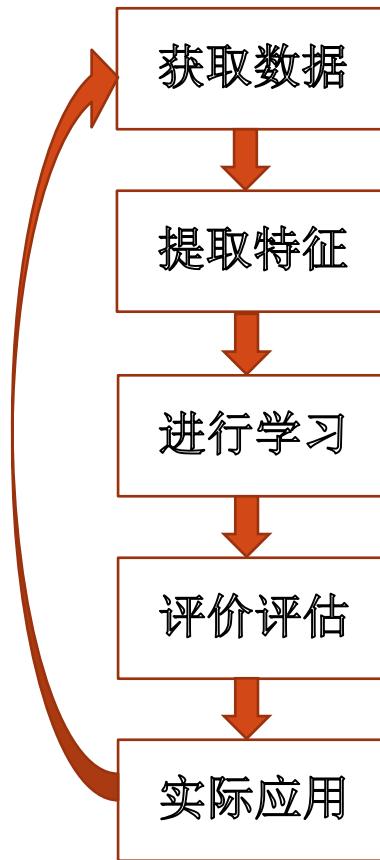
- ✓ 如何构建一个模式识别系统将输入转换为输出
 - 常见方法：机器学习(machine learning)
 - 从过去的经验学习Learning from experience
 - 训练集(training set): 已搜集的数据
 - 模型(model): 从训练集得到的规律，表现形式多样
 - 测试(prediction, testing): 将这些规律应用到新的例子以得到针对该例子的结果（模式识别的输出）
 - 测试集(testing set): 为了评估这些规律搜集的数据
 - 那么，对训练集和测试集应该有什么要求？

(虚拟例子) 性别识别：应用

- ✓ 得到输出后，有什么用？
 - 通常在更大的系统中起作用
 - iOnRoad, Siri, Kinect, ...

- ✓ 那么，PRCV系统要做到多好呢？
 - 由那个“更大的系统”决定
 - 是不是越高越好？

小结：PRCV的步骤



问题：

- ✓ 是这样吗？
 - 需要多次反馈(feedback)、尝试、修改
- ✓ 什么步骤最重要？
- ✓ 深度学习在图中什么位置？

模式识别为什么困难？

✓ 多种原因

- 语义鸿沟semantic gap, 如, 一副 2000×1000 的图像是什么?
 - 对人的眼睛和大脑: 教师、桌椅、讲台、黑板、人、...
 - 对计算机: $3 \times 2000 \times 1000 = 6,000,000$ 个数字
- 计算能力computational power
 - 很多算法在台式机上要数年或更长时间
 - 大数据, 存储和计算都成为问题
- 数据的获取data acquiring
 - 很多时候数据难以获取(如, 医学图像medical imaging)
 - 或者虽然容易获取, 但是难以标注(如, 对google图像的精确标注annotation)

我的书架_个人中心_起点 > lsrc2010 - Google 搜索 > ImageNet Tree View > www.image-net.org/synset?wnid=n03028079

Number of synsets: (the number of synsets in the subtree).
ImageNet 2011 Fall Release (21841)
- animal, animate being, beast, brute, c
- plant, flora, plant life (3775)
- person, individual, someone, somebo
- fungus (298)
- natural object (551)
- artifact, artefact (7894)
- paving, pavement, paving materia
- sheet, flat solid (45)
- restoration (0)
- ready-made (0)
- plaything, toy (17)
- padding, cushioning (12)
- opening (18)
- mystification (0)
- marker (2)
- fine (26)
- sphere (0)
- square (0)
- building material (41)
- weight (3)
- way (98)
- track (2)
- thing (2)
- surface (37)
- structure, construction (1238)
- - airdock, hangar, repair shed (0)
- corner, quoin (0)
- cross (0)
- deathtrap (0)

TreeMap Visualization Images of the Synset Downloads

imageNet 2011 Fall Release > Place of worship, house of prayer, house of God, house of worship > Cathedral

Cathedral Kirk Abbey Basilica

www.image-net.org/synset?wnid=n03028079#panel © 2010 Stanford Vision Lab, Stanford University, Princeton University support@image-net.org Copyright infringement

一个例子

- ✓ ILSVRC2010: Large Scale Visual Recognition Challenge 2010
 - 图像分类: 1,000个类别, 1,200,000训练图像
 - 原始输入 (图像) 约100+GB
 - 有效的图像表示: 每个图像表示为262,144的向量, 需要1258GB空间来存储特征
 - 设计新的机器学习算法, 使得存储开销缩小到约40G
 - 用SVM^{light} 或LIBSVM要很多年
 - 用LIBLINEAR要大概2个月
 - 使用新的深度学习(deep learning)方法, 用GPU辅助计算, 可以得到更好的精度accuracy
 - 尝试一下用你的笔记本电脑测试这个网络?

各步骤的（实践）重要性

1. 提取特征

- 设想一下，告诉你一个人的年纪，需要判别其性别？
- 或者，告诉你一个人的染色体组成，需要判别其性别？

2. 数据获取

- 设想一下，根据图像判断性别，所有训练图像都只包含男子、而不包含任何女子？
- 或者，图像中有男有女，但是所有都标记为女？

3. 机器学习

- 有很多理论上重要的问题，对实践效果也非常重要

和其他领域的关系

PRCV vs. 机器学习

- ✓ 机器学习在模式识别、计算机视觉中有非常重要的作用
- ✓ 但是， PRCV具有更多的“系统”性
 - 数据获取
 - 提取特征
 - ...
- ✓ 主要的区别是：“数据” vs.“特征”
 - 但是，在深度学习中，这个区别不明显了

PRCV与多媒体(multimedia)

- ✓ 都可能牵涉多种媒体
- ✓ 但多媒体更具有“系统”性
 - 多媒体不特别强调单个模块的性能
 - 更注重整个大系统的成功
 - 比模式识别更强调多种媒体之间的配合
 - 就算每个模块都采用了已有的技术，但是一个科学利用现有技术和多种媒体的系统仍然可以是很大的成功

小结：PRCV vs. 其他学科

- ✓ PRCV与很多学科有广泛的联系
 - 机器学习、多媒体、可视化(visualization)、图形学(graphics)、 ...
- ✓ 需要很多学科的支持
 - 物理学、电子学、 ...
- ✓ PRCV与相关学科是有区别的
 - 但是，现在各个学科呈现越来越融合的趋势
 - 所以，不要强调区别，让我们更注重共性、合作、以及更重要的—解决问题！

进一步的阅读

- ✓ 各相关领域的最新进展可以参考其重要会议和期刊
 - 机器学习：JMLR, ICML, NeurIPS
 - 计算机视觉：TPAMI, IJCV, CVPR, ICCV, ECCV
 - 多媒体：ACM Multimedia
- ✓ 有用的资源：
 - 计算机视觉、图像处理OpenCV: <http://opencv.org/>
 - Matlab: 若干工具箱toolbox
 - 深度学习软件包（百花齐放）！