



计算机视觉：导论

叶山 中国地质大学（北京）

yes@cugb.edu.cn

关于这门课

这是一门关于计算机视觉的课程，也是一门关于
深度学习的课程。

关于这门课

课号： DR043229

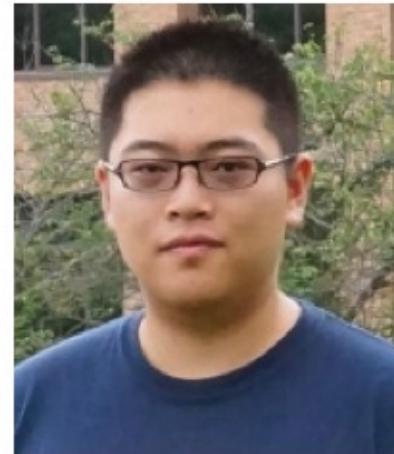
计算机视觉

叶山（信工学院计算机系）

邮箱： yes@cugb.edu.cn

办公室： 教三楼221

辅助工具： 学习通



主讲人背景

本科：宾夕法尼亚州立大学（地理信息）& 密歇根大学（地球和环境科学）

硕士：塔尔萨大学（地质学）

博士：威斯康辛大学麦迪逊分校（地学大数据）

关于这门课



关于这门课



平时成绩：Pop Quiz（课堂“突击”测验）

- 使用腾讯问卷答题（请带手机或平板电脑），每次5道选择题，每题2分，满分10分。
- 内容为上一次Pop Quiz以来的知识点。
- 每道题答对得2分，答错得1分，缺席记0分。
- 随机进行5次。学期结束后，取历次的平均分为平时成绩。

关于这门课

Lab作业（一共8次）

- 使用Python（3.x）语言在Jupyter Notebook中完成代码编写。8次平均分为学期的Lab成绩。
- 建议：安装并熟悉Anaconda及虚拟环境。

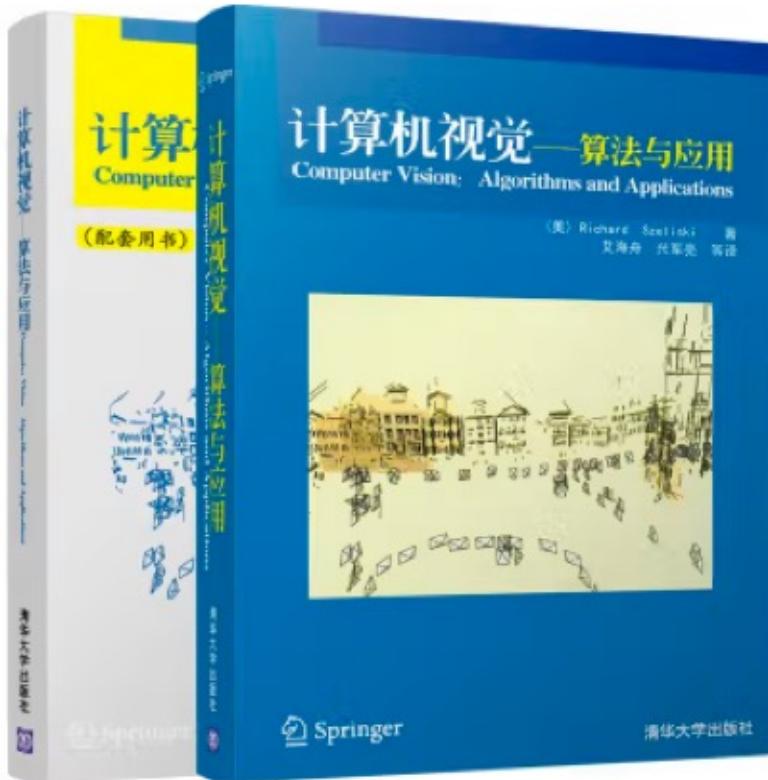
关于这门课

期末考试（闭卷）

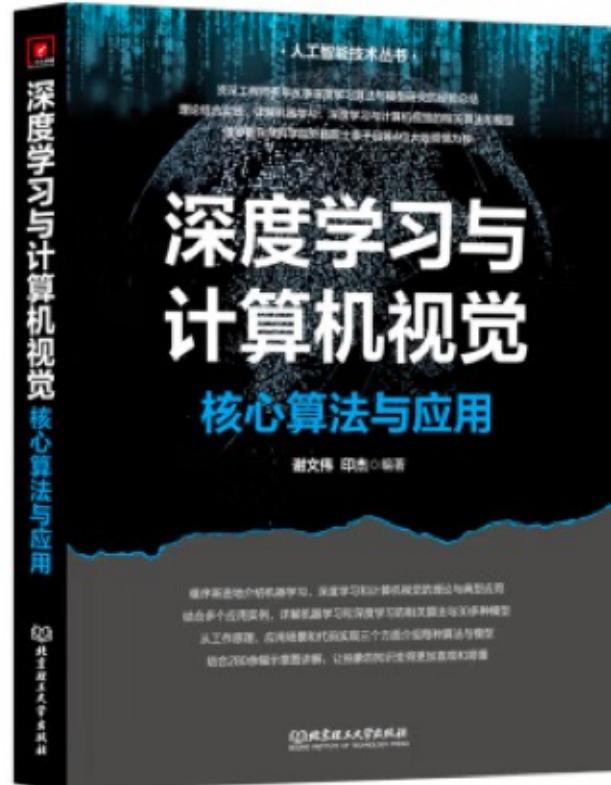
- 考查内容：对课上涉及概念的掌握及运用能力。
- 考试时间：期末考试周（教务系统会通知）。

关于这门课

推荐参考书（非必须）



Richard Szeliski 著
艾海舟、兴军亮 等译
清华大学出版社



谢文伟、印杰 著
北京理工大学出版社

关于这门课

图像识别任务的难点与研究现状。



经典的计算机视觉任务：
分类、检测、图像分割等。



和计算机视觉相关的神经
网络。



Python实战（OpenCV）。

关于这门课

参考课程（致谢）：

- 斯坦福大学 CS 231n (李飞飞团队的经典课程)
- 密歇根大学 EECS 442 & EECS 498 (Justin Johnson)
- 威斯康辛大学麦迪逊分校 CS 539 & CS 766



计算机视觉举例

查找图片来源

中国地质大学 (北京)

2.6 ★★★★☆

大学

G 搜索

w Wikipedia File:CUG Beijing Gate.jpg - 维基百科
查看完全匹配的... >

头条百科 中国地质大学附属中学 - 抖音百科
Baidu 北京地质学院_百度百科

CSDN 北京旅游网 中国地质大学
武汉计算机大学排名 2018, 2018年中国...

Sohu 中国地质大学 (北)

谷歌图片搜索

计算机视觉举例



(a)



(b)



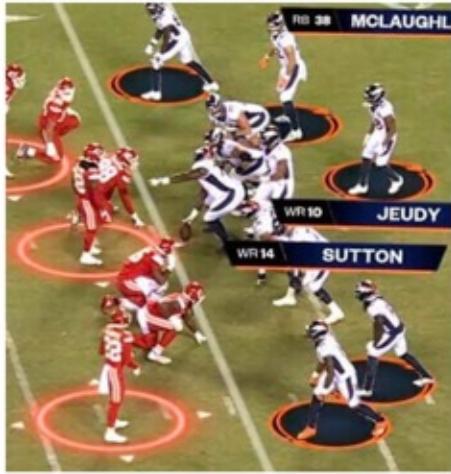
(c)



(d)

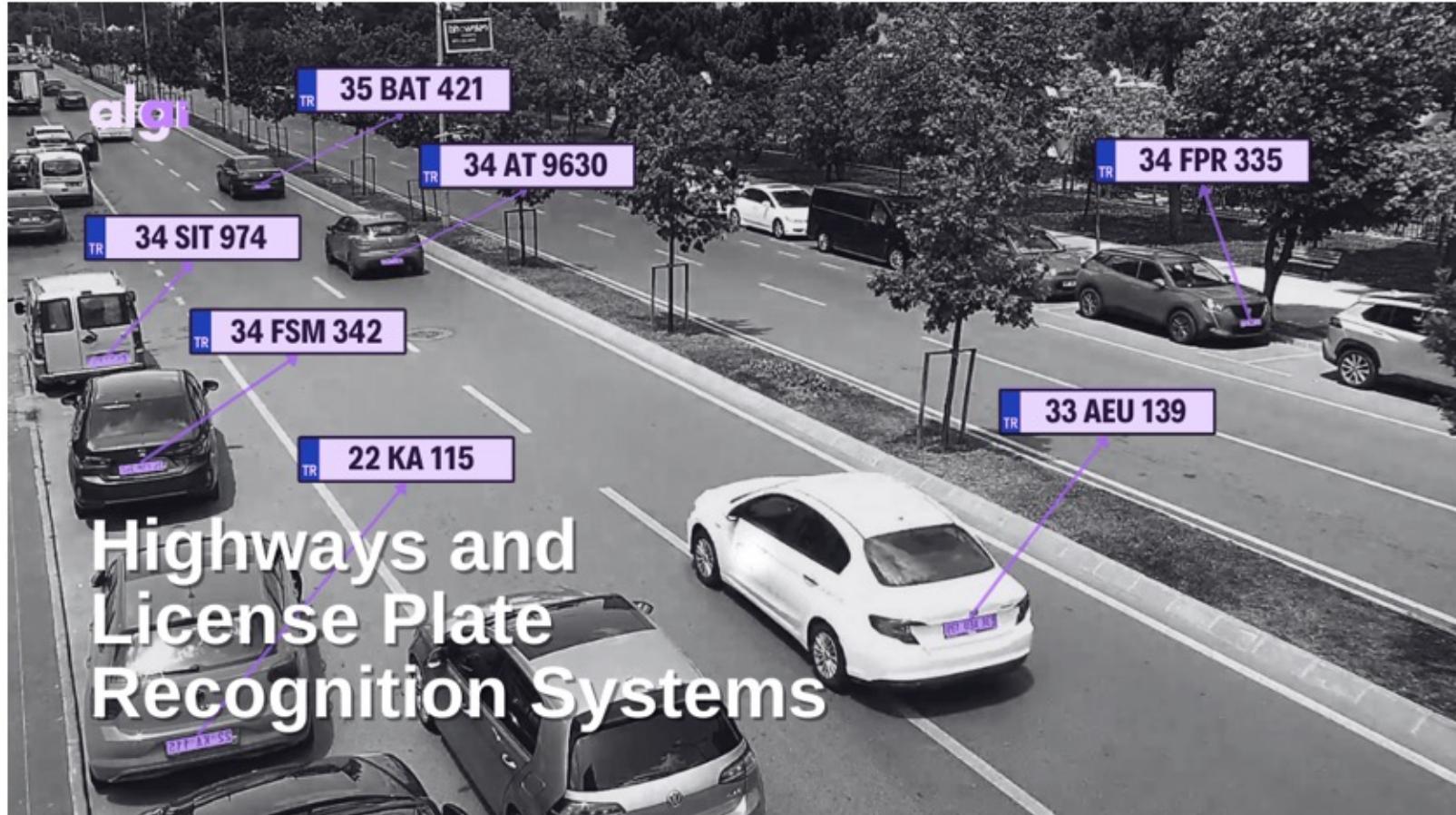
画面除雾操作

计算机视觉举例



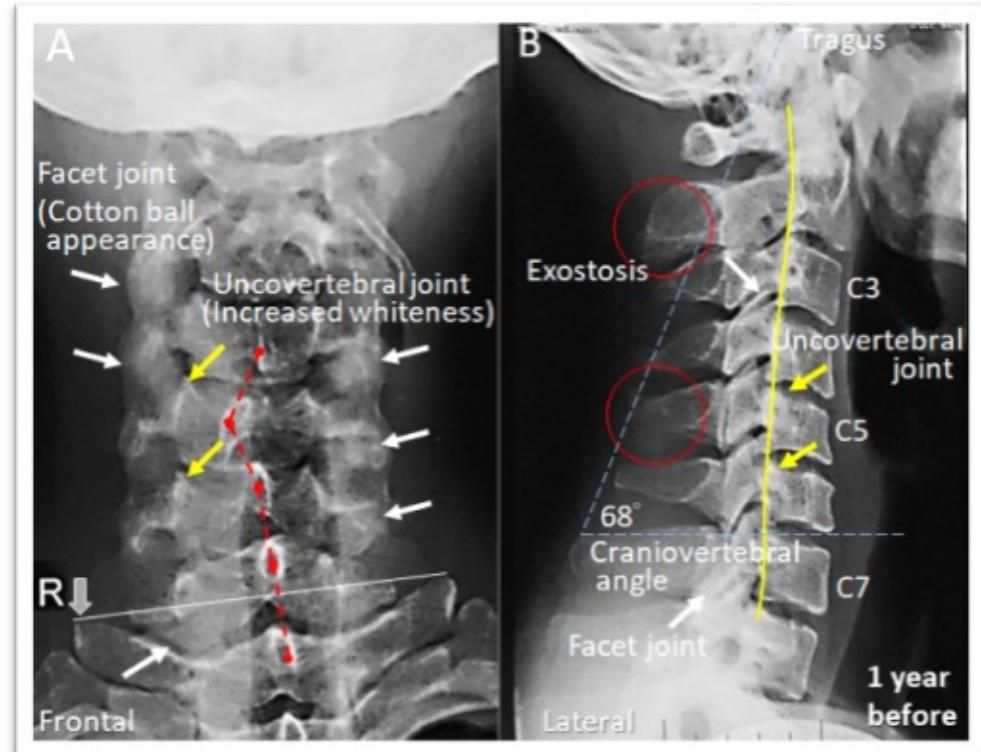
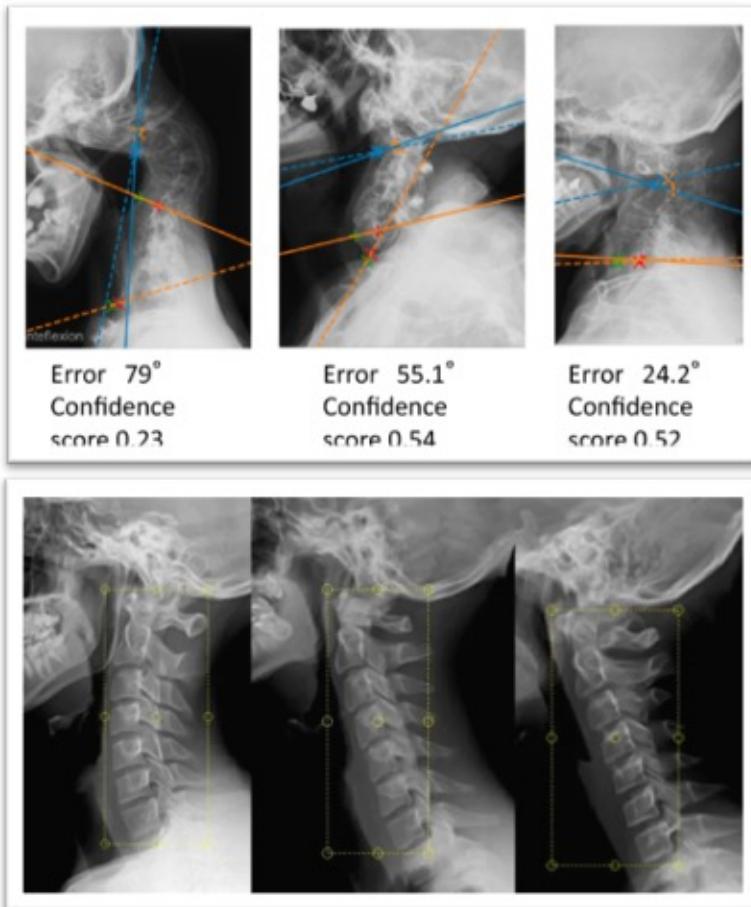
视频助理裁判

计算机视觉举例



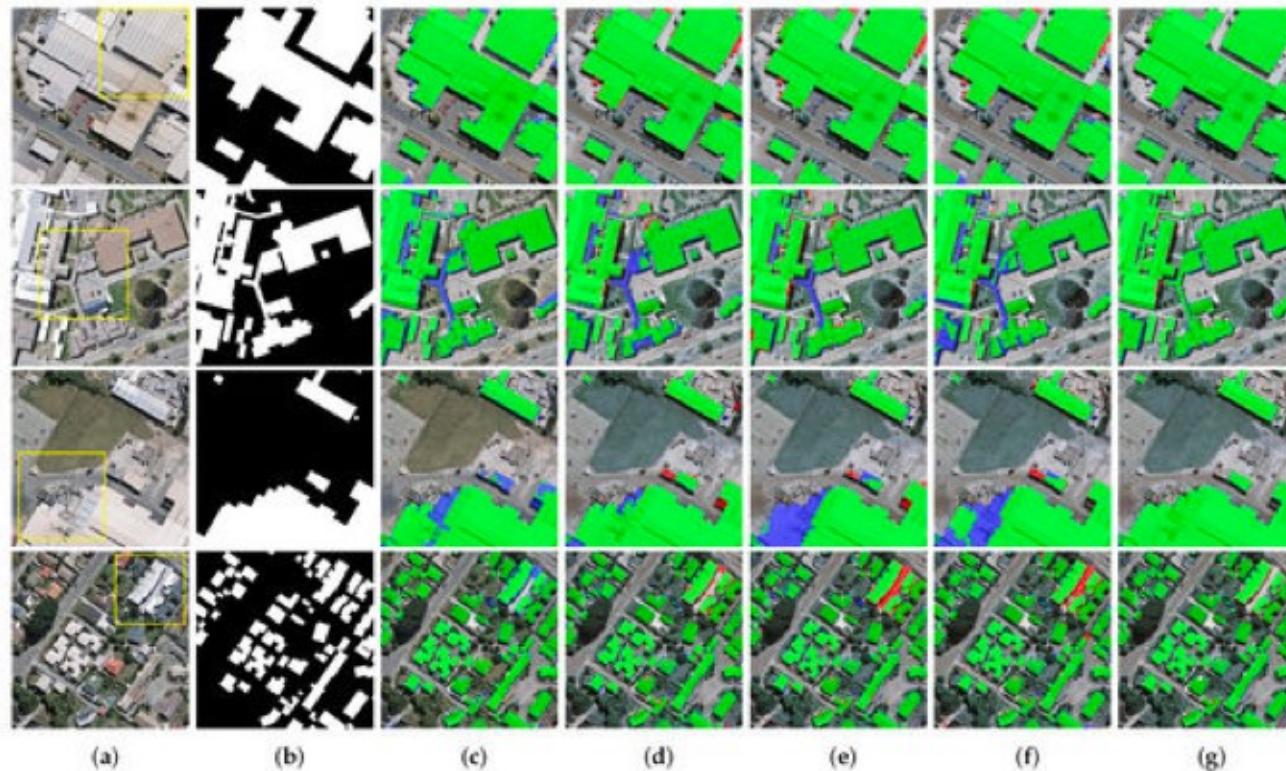
车牌识别系统

计算机视觉举例



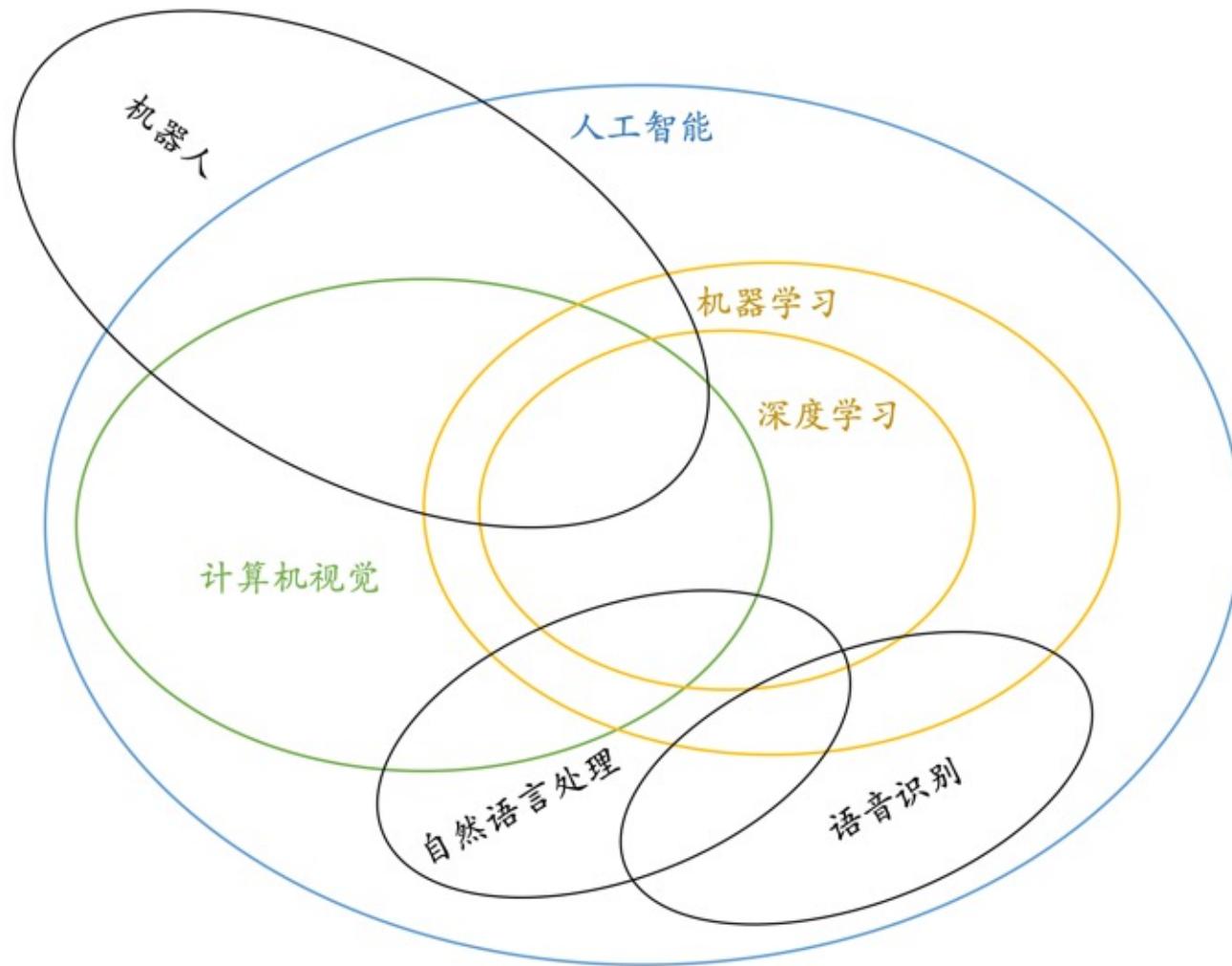
医用图像识别

计算机视觉举例



遥感图像识别

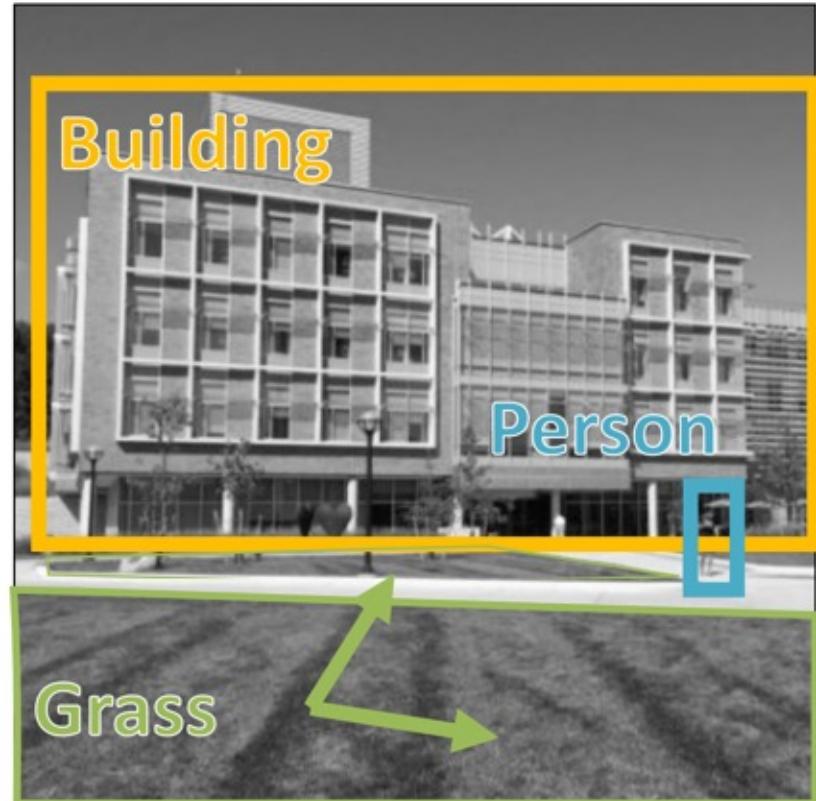
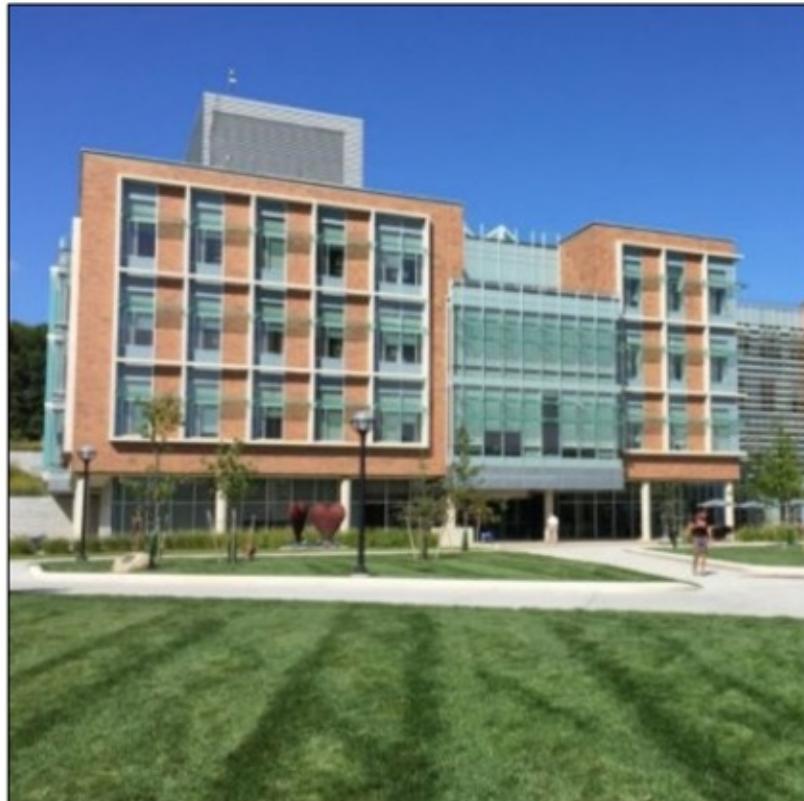
计算机视觉和其他学科的关系



计算机视觉和其他学科的关系



计算机视觉和其他学科的关系



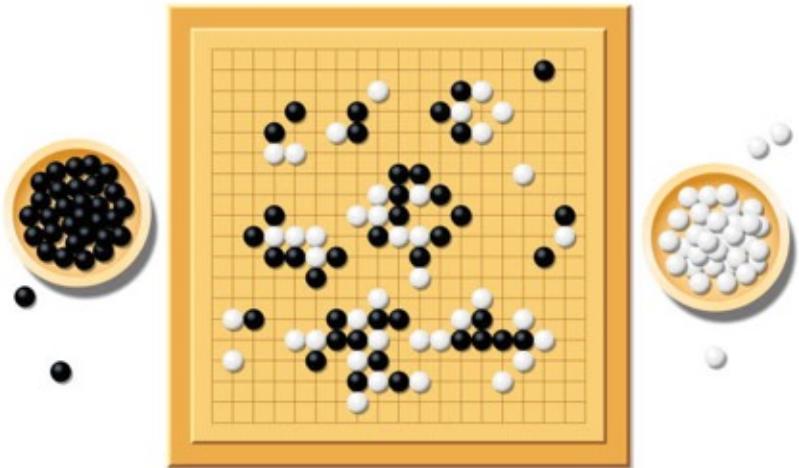
输出标签

计算机视觉和其他学科的关系



特斯拉的人形机器人

视觉相关的智能



SIMPLE SCORING READY RECKONER (Non Vul)											
CONTRACT DEFEATED - 50 pts per trick short (regardless of denomination).											
Contract Made			Contract Made			Contract Made			Contract Made		
BID	TRKS	SCORE	BID	TRKS	SCORE	BID	TRKS	SCORE	BID	TRKS	SCORE
1♦	7	70	2♣	8	90	3♥	9	110	5♣	10	140
1♦	8	90	2♣	9	110	3♥	10	130	5♣	11	170
1♦	9	110	2♦	10	130	3♠	11	200	5♦	12	220
1♦	10	130	2♦	11	150	3♠	12	230	5♦	13	440
1♦	11	150	2♦	12	170	3♠	13	260	5♥	11	450
1♦	12	170	3♦	13	190	3NT	9	400	5♥	12	480
1♦	13	190	3♦	8	110	3NT	10	430	5♦	13	510
1♥	7	80	2♥	9	140	3NT	11	460	5	11	460
1♥	8	110	2♥	10	170	3NT	12	490	5	12	490
1♥	9	140	2♦	11	230	3NT	13	520	NT	13	520
1♥	10	170	2♦	12	260	4♣	10	130	NT	10	130
1♦	11	200	3♦	13	260	4♣	11	150	6♣	12	920
1♦	12	230	3♦	8	120	4♣	12	170	6♦	13	940
1♦	13	260	3♦	9	150	4♦	13	190	6♥	12	980
1♦	7	90	2NT	10	180	4♦	10	420	6♦	11	1010
1♦	8	120	2NT	11	210	4♦	11	450	6♦	12	990
1♦	9	150	3♣	12	240	4♦	12	480	6♦	13	1020
1♦	10	180	3♣	9	110	4♣	13	510	NT	13	1020
1♦	11	210	3♣	10	130	4♣	10	430	7♦	13	1440
1♦	12	240	3♣	11	150	4♣	11	460	7♦	13	1510
1♦	13	270	3♦	12	170	4♣	12	490	7NT	13	1520
1NT	13	270	3♦	13	190	4♣	13	520	7NT	13	1520



视觉相关的智能



Deep Blue vs. Kasparov (1997)



AlphaGo vs. 李世石 (2016)

视觉相关的智能

- 机器智能：计算机系统通过学习和推理等方式，**模拟**人类智能行为的能力。这种智能行为包括**自主学习、适应性强、高效决策**等特点。
- 与传统的程序编写不同，机器智能可以通过**机器学习、大数据分析和模式识别**等技术，从数据中学习并提取规律，从而实现自主决策和智能行为。
- 目前，机器智能和真正的人类智能还相差甚远，它只能在某些**限定环境**下运行（专用人工智能）。

视觉相关的智能



图像

感知设备

- 人类视觉：眼睛
- 机器视觉：摄像头

解释器

- 人类视觉：大脑
- 机器视觉：算法模型

解释结果

- 晴朗的天气
- 屋子前有两只猫

视觉相关的智能



```
[[105 112 108 111 104 99 106 99 96 103 112 119 104 97 93 87]
 [ 91 98 102 106 104 79 98 103 99 105 123 136 118 105 94 85]
 [ 76 85 98 105 128 105 87 96 95 99 115 112 106 103 99 85]
 [ 99 81 81 93 120 131 127 100 95 98 102 99 96 93 101 94]
 [106 91 61 64 69 91 88 85 101 107 109 98 75 84 96 95]
 [[114 108 85 55 55 69 64 54 64 87 112 129 98 74 84 91]
 [133 137 147 103 65 81 88 65 52 54 74 84 102 93 85 82]
 [128 137 144 148 109 95 86 70 62 65 63 63 60 73 86 101]
 [125 133 148 137 119 121 117 94 65 79 80 65 54 64 72 98]
 [127 125 131 147 133 127 126 131 111 96 89 75 61 64 72 84]
 [115 114 109 123 150 148 131 118 113 109 100 92 74 65 72 78]
 [ 89 93 90 97 108 147 131 118 113 114 113 109 106 95 77 80]
 [ 63 77 86 81 77 79 102 123 117 115 117 125 125 130 115 87]
 [ 62 65 82 89 78 71 88 101 124 126 119 101 107 114 131 119]
 [ 63 65 75 88 89 71 62 81 120 138 135 105 81 98 110 118]
 [ 87 65 71 87 106 95 69 45 76 130 126 107 92 94 105 112]
 [118 97 82 86 117 123 116 66 41 51 95 93 89 95 102 107]
 [164 146 112 88 82 128 124 104 76 48 45 66 88 101 102 109]
 [157 170 157 120 93 86 114 132 112 97 69 55 78 82 99 94]
 [138 128 134 161 139 100 108 118 121 134 114 87 65 53 69 86]
 [128 112 96 117 150 144 120 115 104 107 102 93 87 81 72 79]
 [123 107 96 86 83 112 153 149 122 109 104 75 80 107 112 99]
 [122 121 102 80 82 86 94 117 145 148 153 102 58 78 92 107]
 [122 164 148 103 71 56 78 83 93 103 119 139 102 61 69 84]]
```

计算机视觉的目标：跨越语义隔绝（语义鸿沟），建立像素到语义的映射。

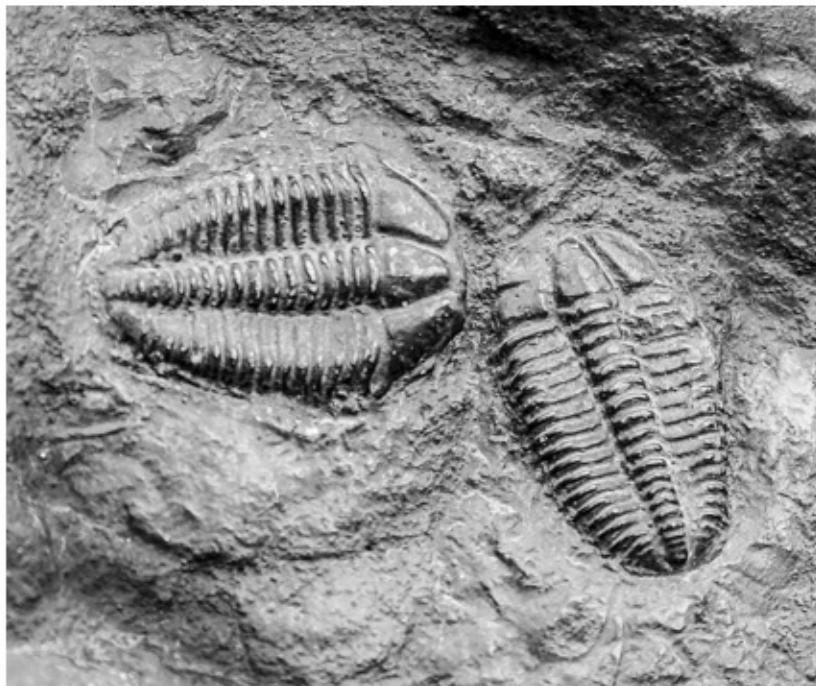
人类视觉

生命视觉的演化

Cambrian Explosion

寒武纪生命大爆发（5.41亿年前）

潜在原因之一：视觉的进化

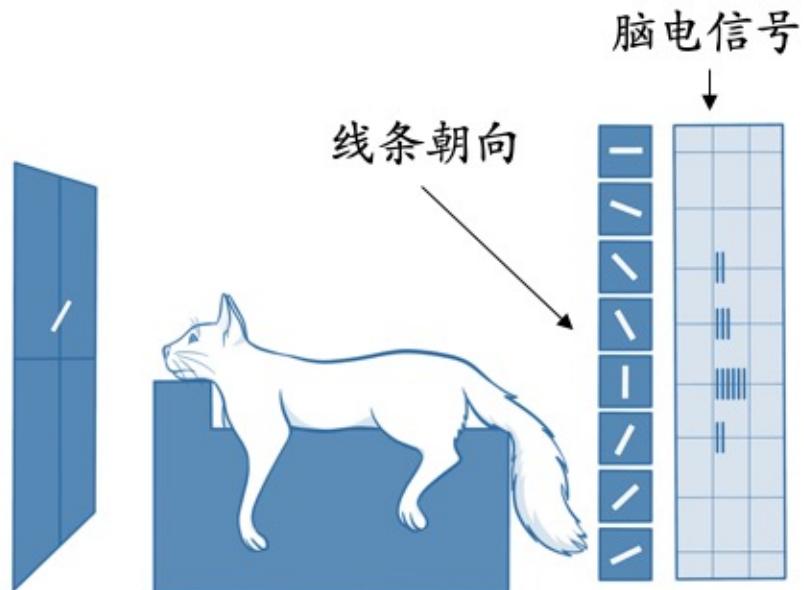


三叶虫



奇虾

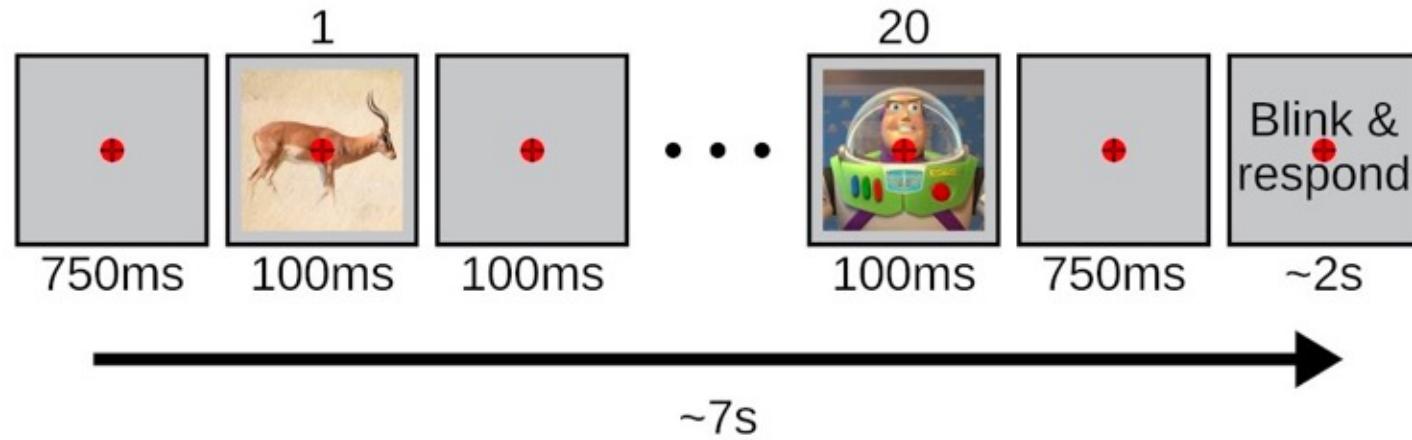
人类视觉



David Hubel (左) & Torsten Wiesel (右)

人类视觉：多层次的信息处理过程。

人类视觉



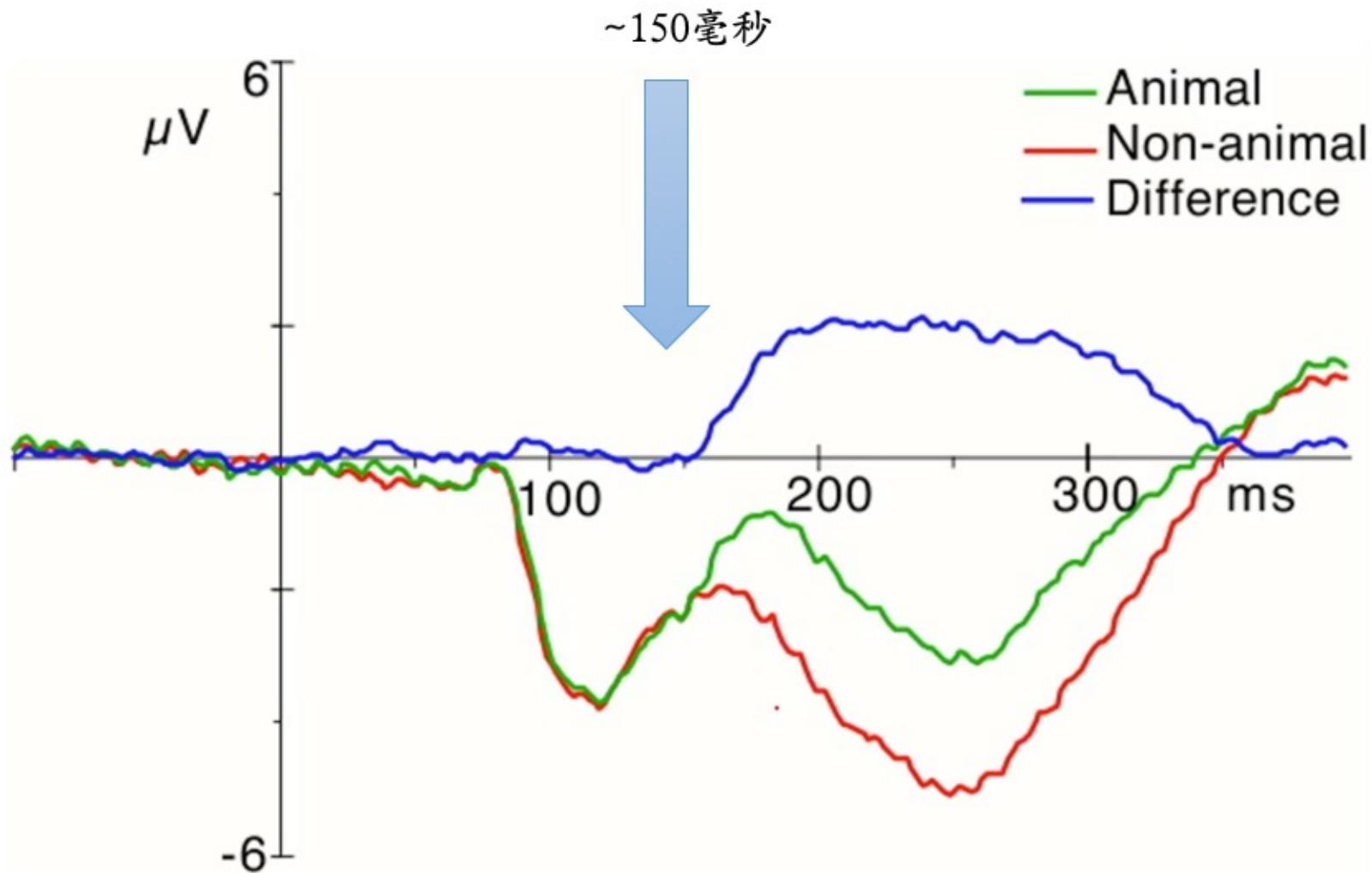
快速系列视觉感知 (RSVP/ rapid serial visual perception) : 人类视觉系统的高效性



人类视觉



人类视觉



Thorpe et al., 1996

人类视觉

- 变化视盲（Change Blindness）：也被称为无意视盲，是一种心理现象，指的是观察者无法探测到物体或场景中的显著变化。
 - 当我们能够对周围世界中出现的变化进行觉察、识别和定位时，这被称为变化探测（change detection）。然而，在某些情况下，我们却未能觉察到这种变化，这一现象即被称为变化视盲。
 - 这种现象的产生可能与大脑处理信息的方式有关。大脑处理信息有自上而下和自下而上两个路径，其中自下而上的处理是指处理传入的感官信息，而自上而下的处理则是由认知驱动的知觉。
 - 当我们的注意力被其他事物所吸引，或者对某一事物形成了固定的认知模式时，就可能会出现变化视盲现象，因为我们的大脑会应用它所知道的和它所期望的东西，并填补空白，从而忽略了实际发生的变化。

人类视觉



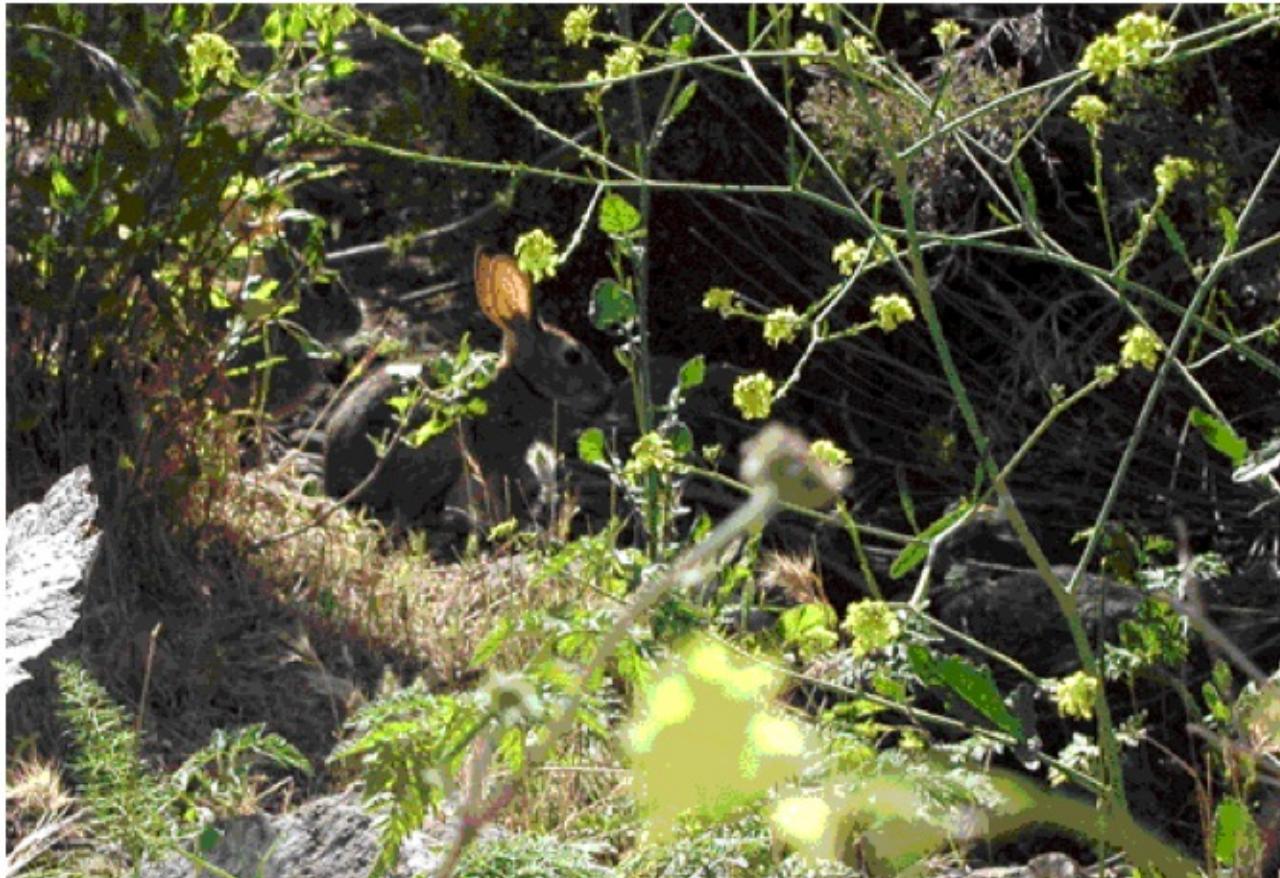
变化视盲

人类视觉



变化视盲

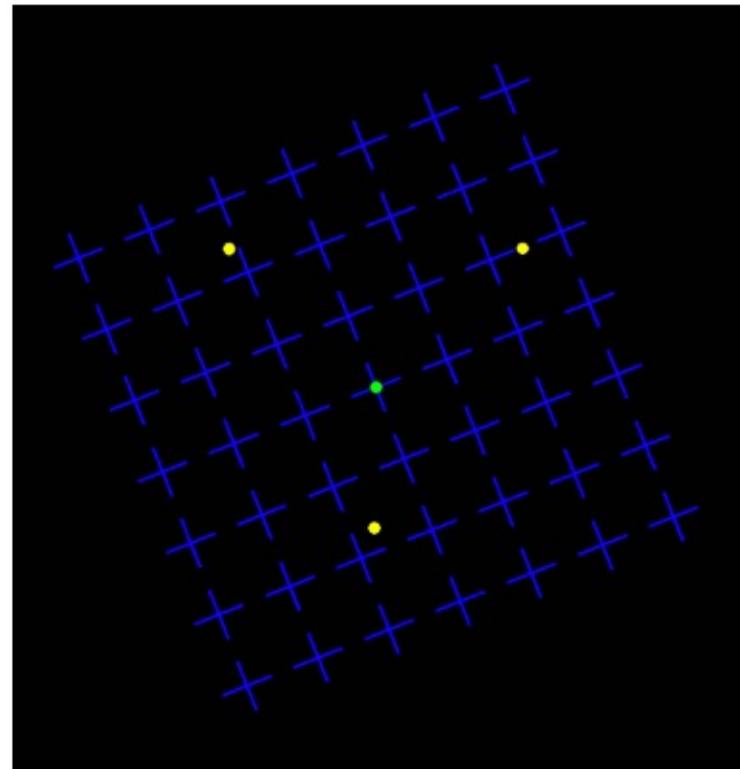
人类视觉



变化视盲

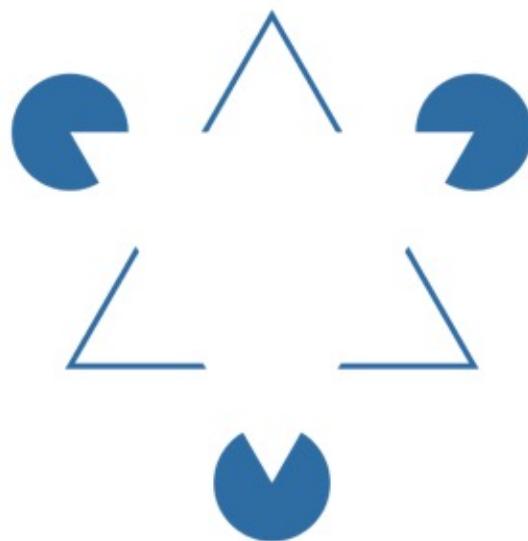
人类视觉

- 运动视盲（motion-induced blindness）：突显的静止目标被一定空间范围内的运动背景所覆盖时，人类视觉会感受到静止目标消失再重现的现象。

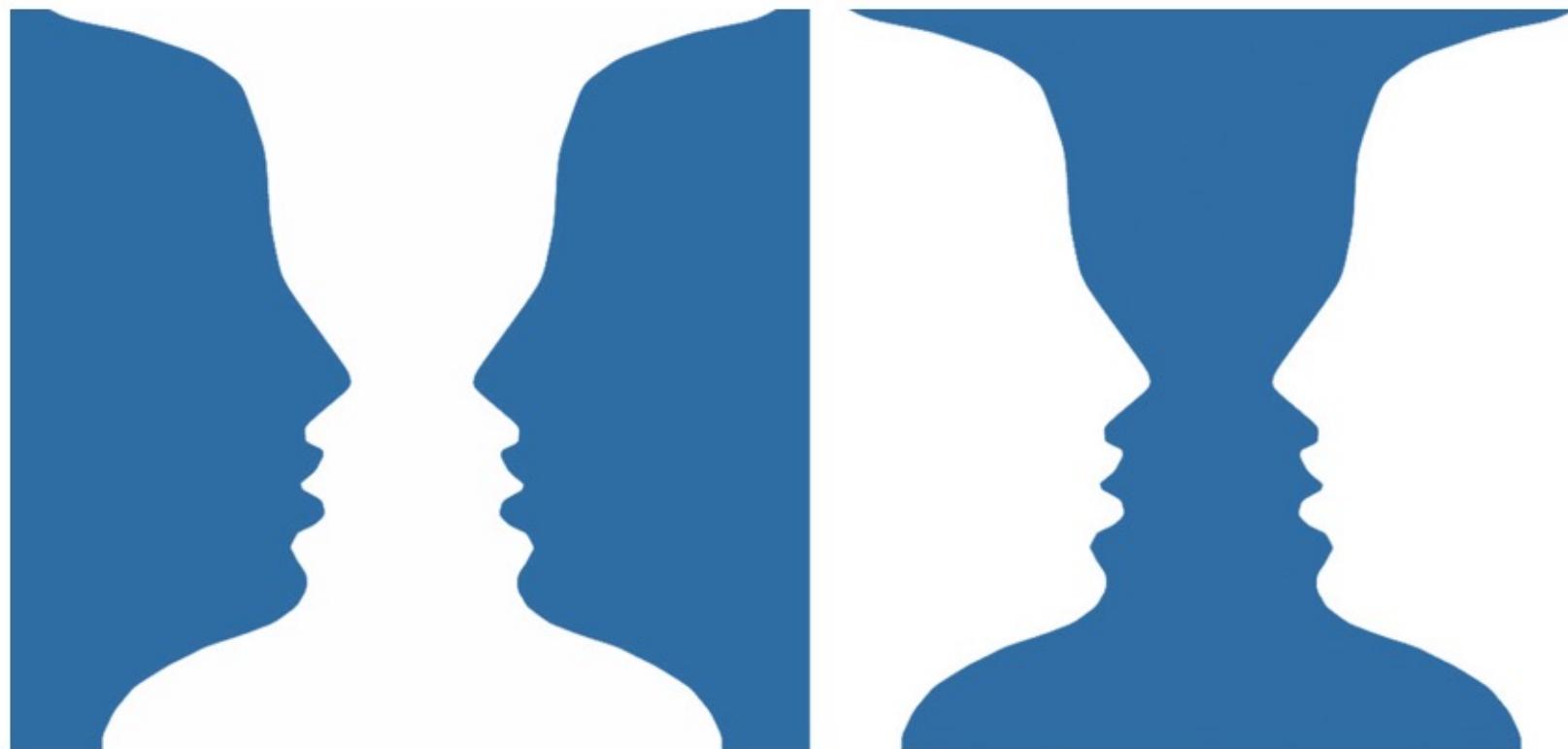


人类视觉

- 视觉涌现（emergence）：从一系列简单的图案中，识别出一些相对高阶的模式和场景。这是格式塔理论（Gestalt）的一部分。



人类视觉



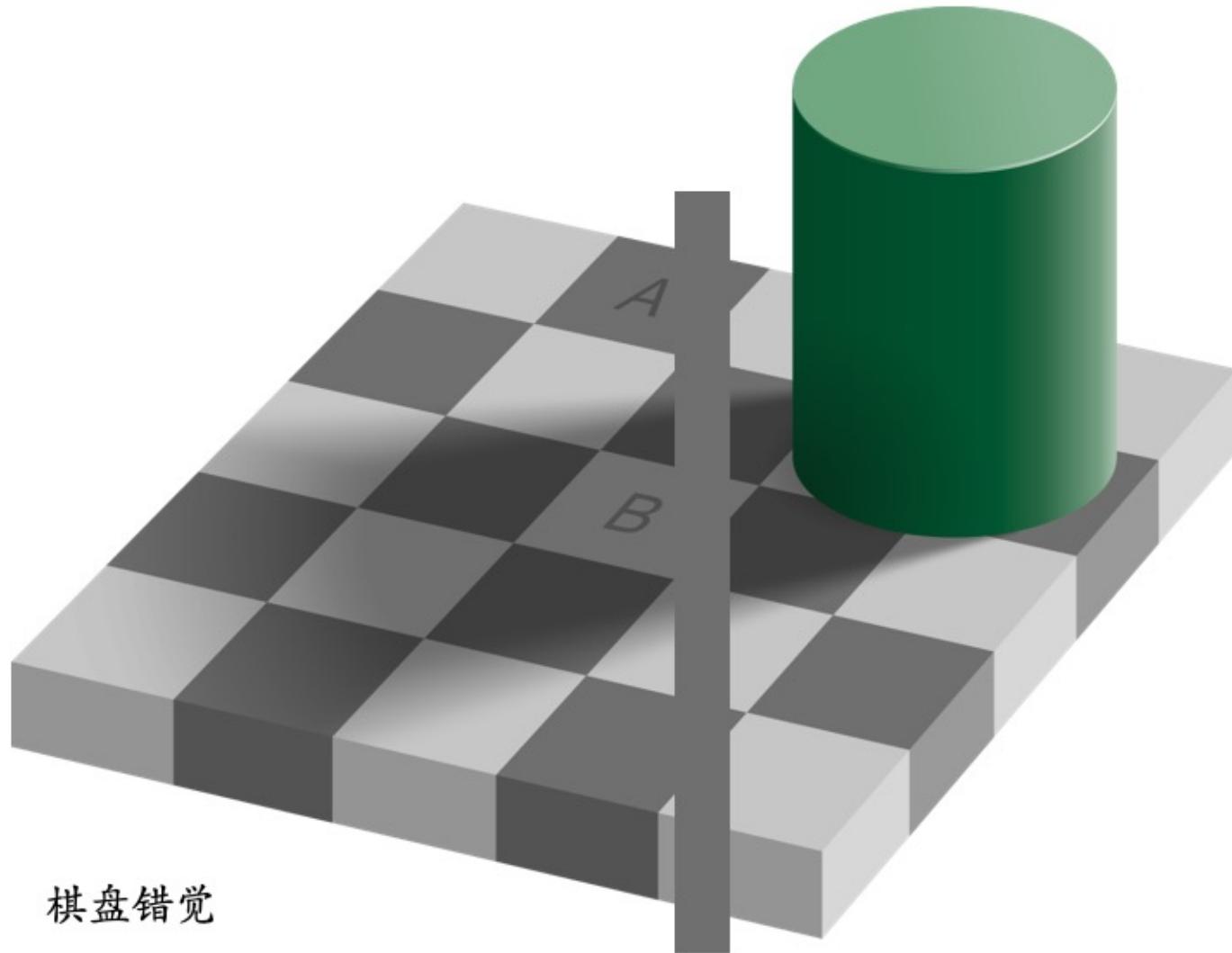
根据经验上下文来理解图像的语义，从而进行图像分割。

人类视觉



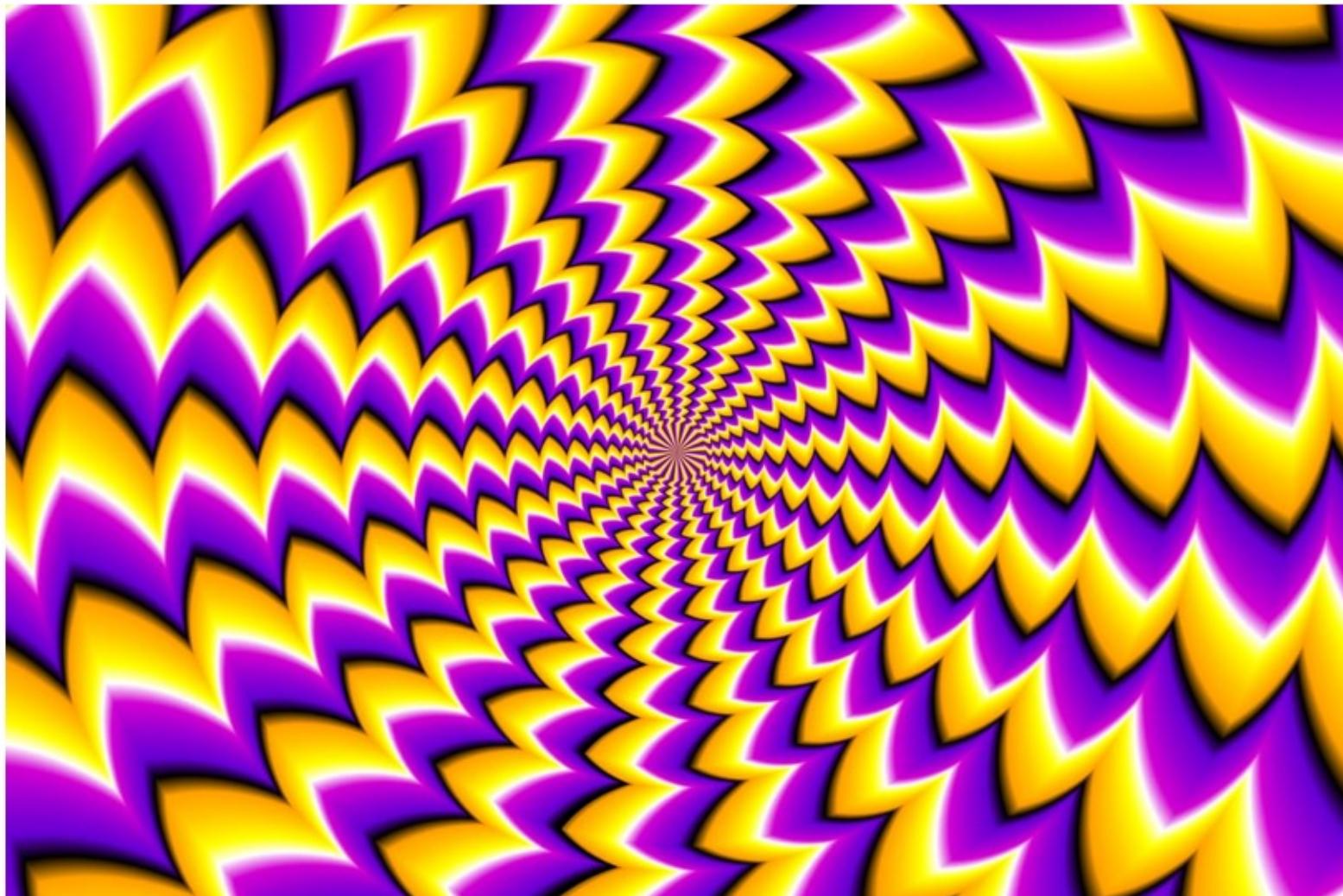
- 图片里都有谁?
 - 90年代的实验：多数被测试的美国人第一直觉认为图中是总统克林顿和副总统戈尔。
 - 但实际上，两张脸都是克林顿。

人类视觉



棋盘错觉

人类视觉



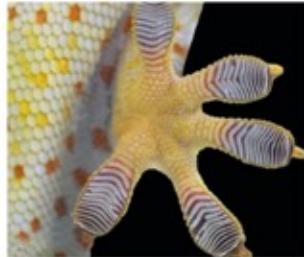
人类视觉



翠鸟头部 → 高速列车 (风阻降噪)



龙虾的眼部结构 → 太空探测器的X射线感应器



壁虎的脚掌 → 超强粘胶

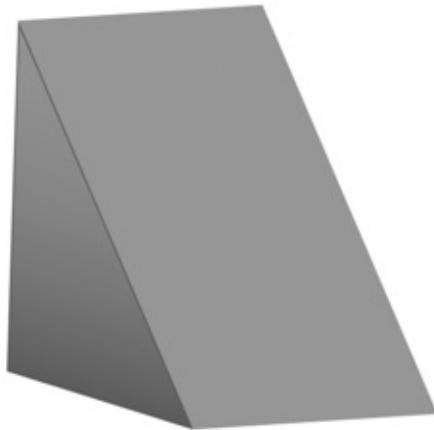


啄木鸟头部减震结构 → 飞机黑匣子

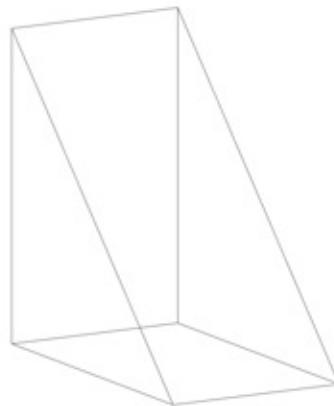


蚯蚓 → 盾构机

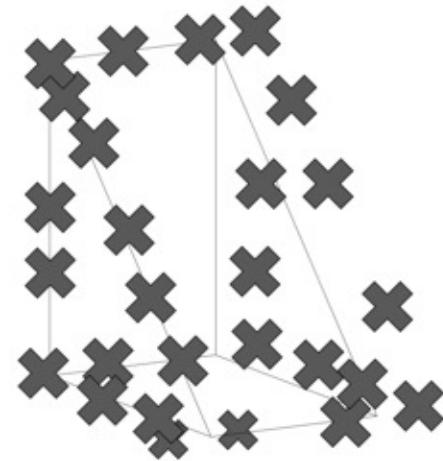
计算机视觉的发展



(a) Original picture



(b) Differentiated picture



(c) Feature points selected



Lawrence (Larry) Gilman Roberts
“Machine Perception of Three-Dimensional Solids”
1963 – 麻省理工学院 博士毕业论文
计算机视觉领域的第一篇论文

计算机视觉的发展

麻省理工学院1966年夏季机器视觉课题

- MIT Summer Vision Project, 由Minsky发起, 导师为Papert, 团队为Sussman带领的10个本科生。
- 试图让电脑描述出摄像机拍摄的画面。
- 向世人展示了计算机视觉问题的复杂性, 使得计算机视觉成为了独立的研究方向。



Marvin Minsky



Seymour Papert



Gerald Sussman

计算机视觉的发展



David Marr
计算机视觉之父

核心思想：如果要让计算机视觉达到人类视觉的效果，其算法设计需符合人类的视觉原理。

- 表象/表征 (representation)：能把某些信息表达清楚的一种形式化系统，以及说明该系统如何行使其职能的规则。
- 描述 (description): 某实体在表象中的表达结果。

计算机视觉的发展

Marr为计算机视觉划分了3个层次。对一切事物的描述，都必须处于一定的层次之下。

计算理论

- 计算的目的是什么？
- 计算的逻辑是什么？
- 问题有哪些已知的约束和边界条件？

表象和算法

- 如何实现计算理论？
- 输入输出的表象是什么？
- 如何实现表象间的变换（中间信息）？

硬件的实现

- 在物理上如何实现这种算法？
- 需要怎样设计和使用芯片等硬件？

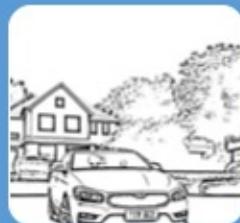
最本质

最重要

计算机视觉的发展



输入图像 (Raw image)



2D 基素图/边缘图 (2D edge features)

- 表明了原始图像中强度变化剧烈的位置，及其几何分布和组织结构。
- 基元：斑点、端点、边界等



2.5D要素图 (2.5D sketch)

- 以观察者为中心的坐标系中，可见表面的法线方向、深度(遮挡)信息、不连续轮廓等。
- 包含了深度(遮挡)的信息，但还不是真正的三维表示，所以得名2.5维图。



3D模型表象 (3D model representation)

- 在以物体为中心的坐标系中，用含有体积基元（即表示形状所占体积的基元）和面积基元的模块化分层次表象，描述形状及其空间组织形式。

Marr: 表达的3个阶段



早期视觉：识别原始二维图像中的重要信息并清楚地表示出来。

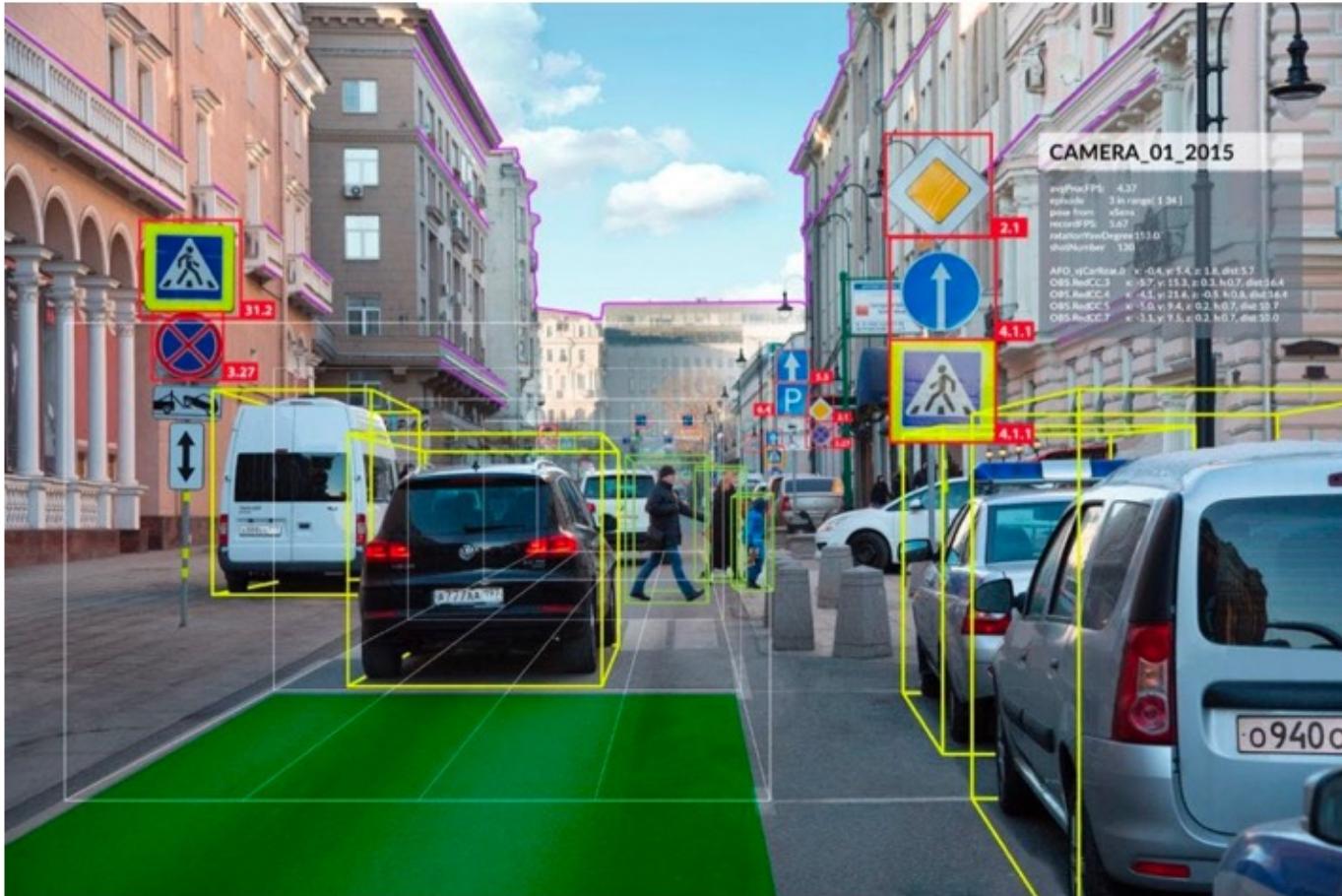
中期视觉：通过符号处理，将线条和点以不同的方式组织起来而获得2.5维图。

后期视觉：由输入图像、基素图、2.5维图综合获得物体的三维表示。

计算机视觉的发展

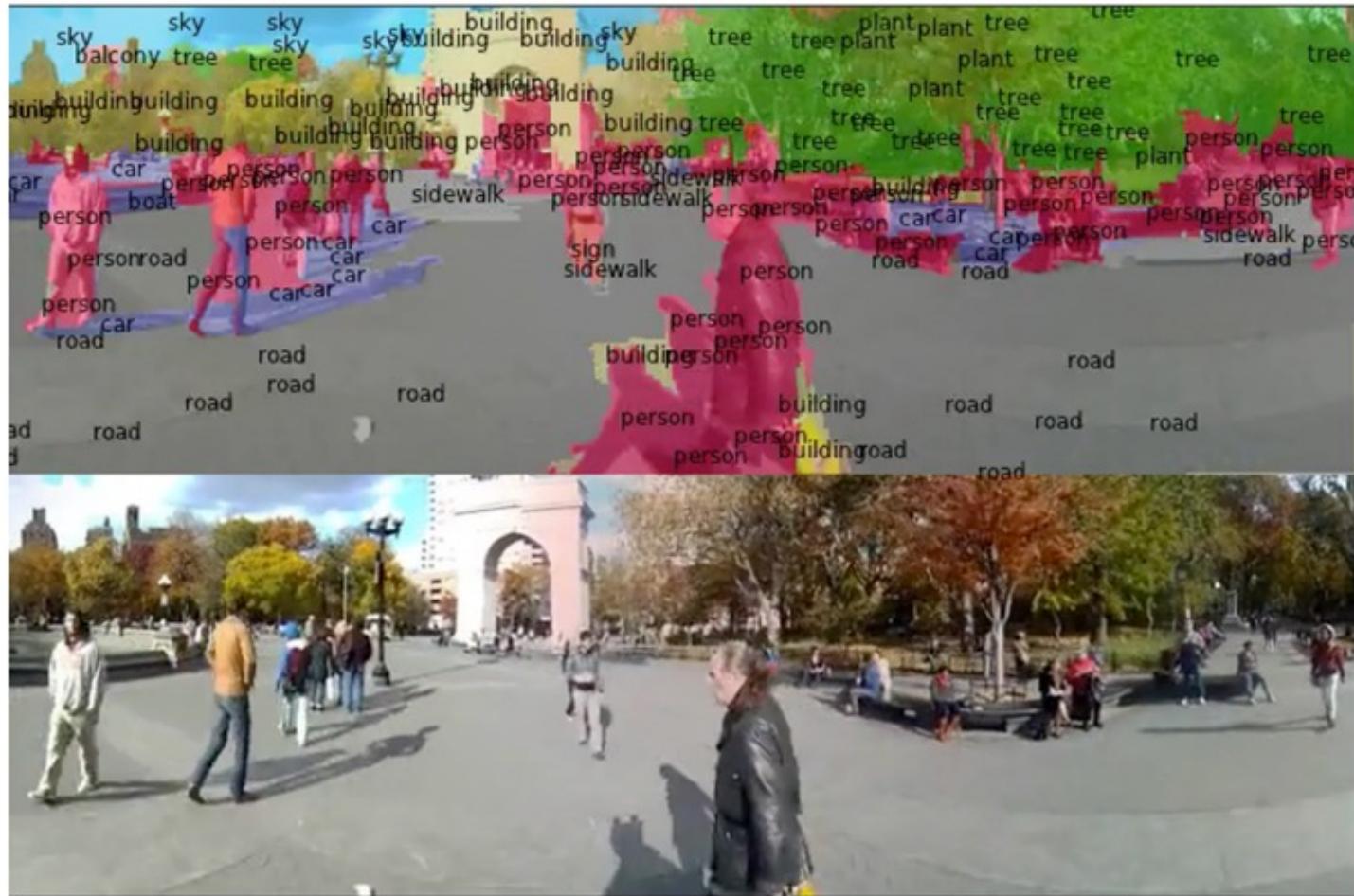
计算机视觉的目标：跨越语义隔绝
(Semantic Gap/语义鸿沟)

计算机视觉的发展



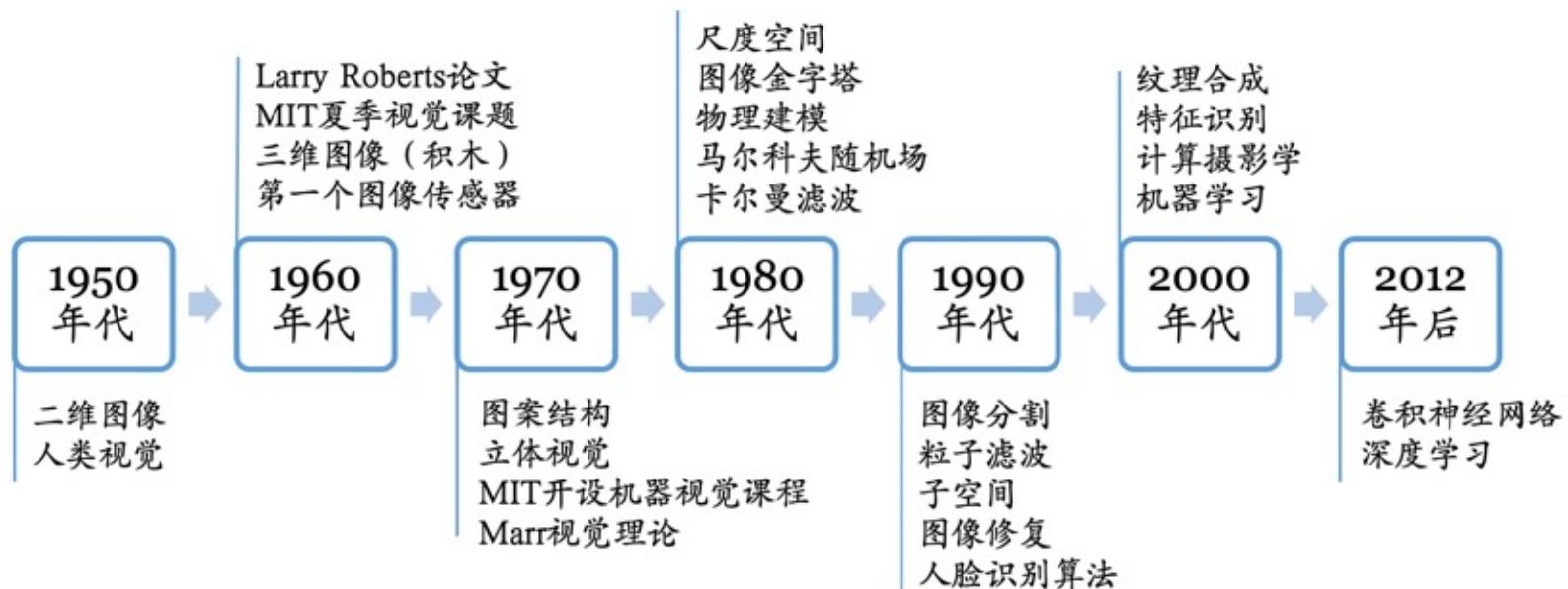
三维场景的结构信息 → 语义信息

计算机视觉的发展

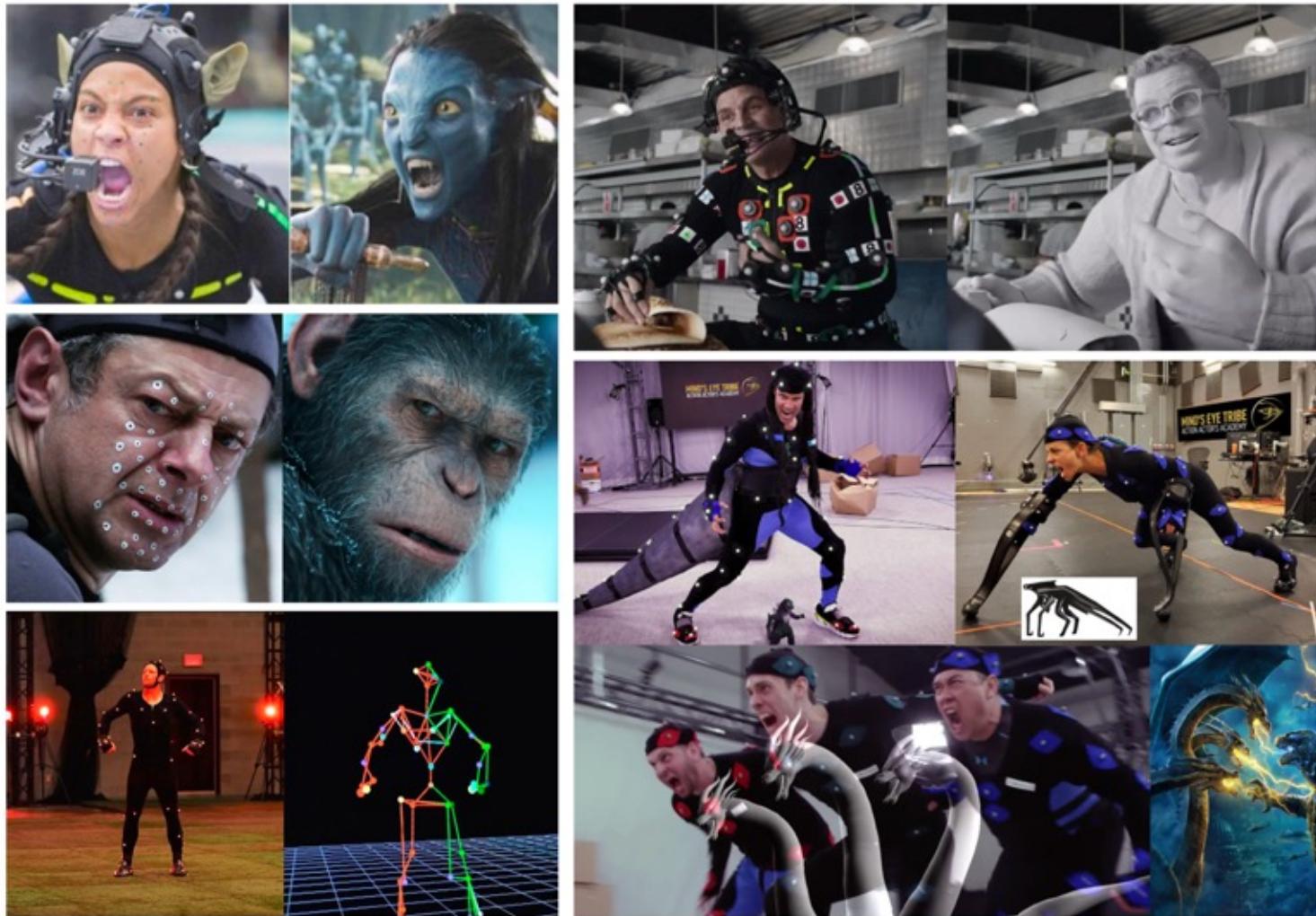


三维场景的结构信息 → 语义信息

计算机视觉的发展

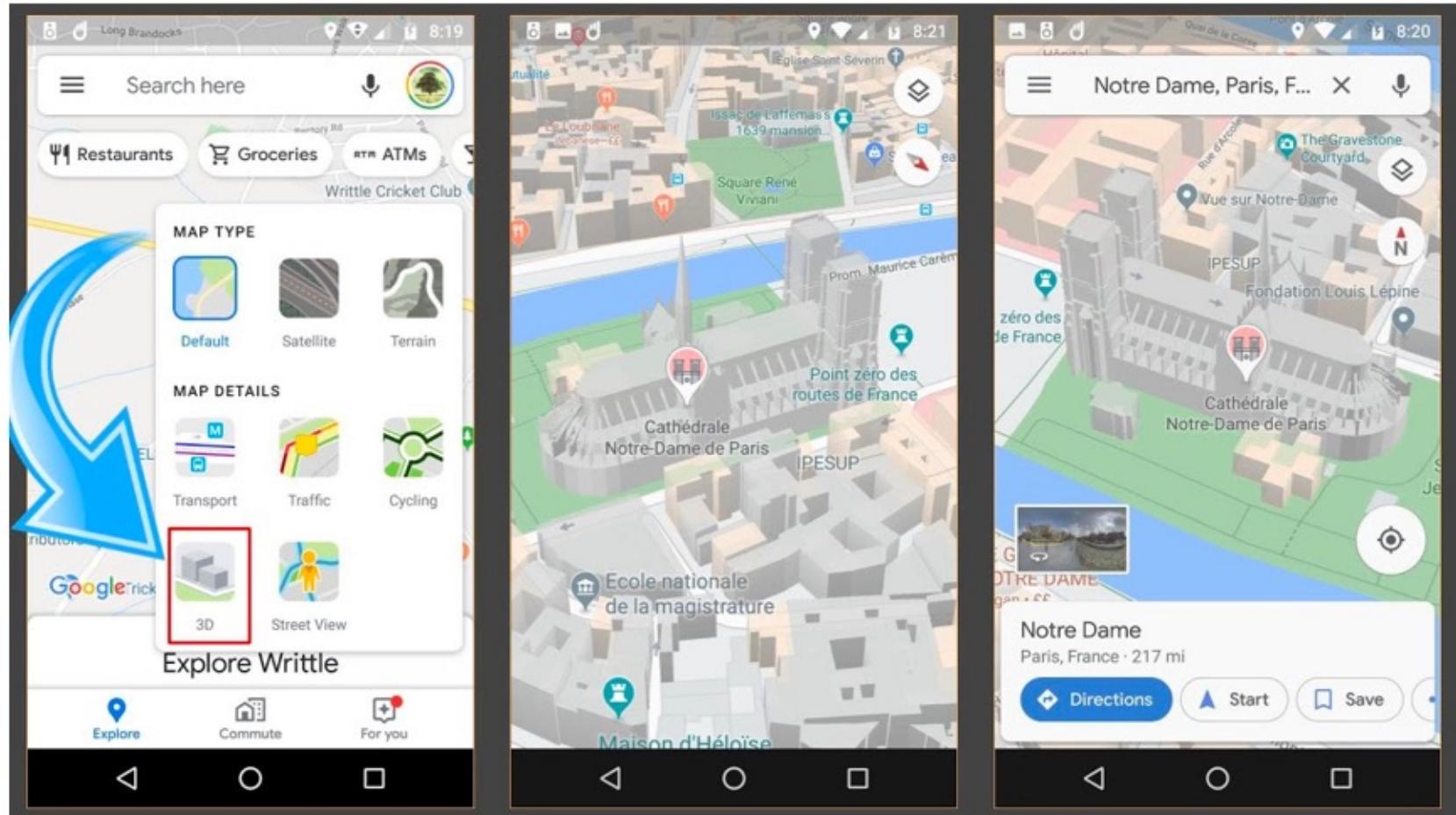


当前进展



影视、动画、游戏产业——动作捕捉 (motion capture)

当前进展

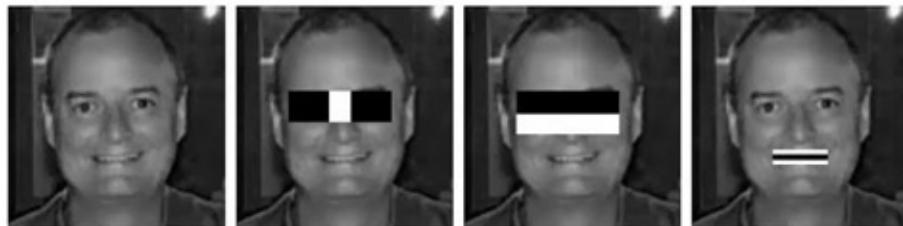


谷歌地图-3D建模

当前进展



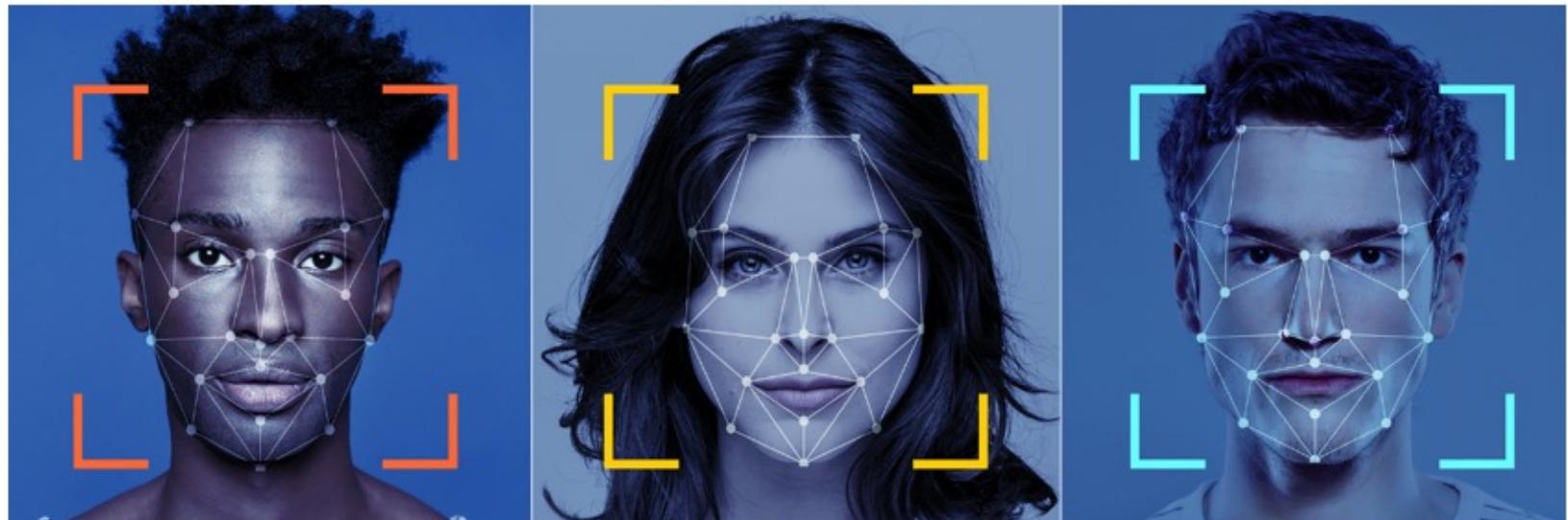
$$\Delta = \text{dark} - \text{white} = \frac{1}{n_{\text{dark}}} \sum_{\text{dark}}^n I(x) - \frac{1}{n_{\text{white}}} \sum_{\text{white}}^n I(x)$$



维奥拉-琼斯目标检测框架 (Viola-Jones object detection framework) - AdaBoost 算法

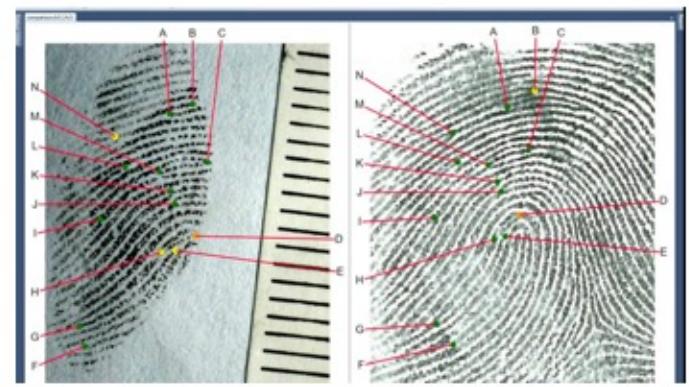
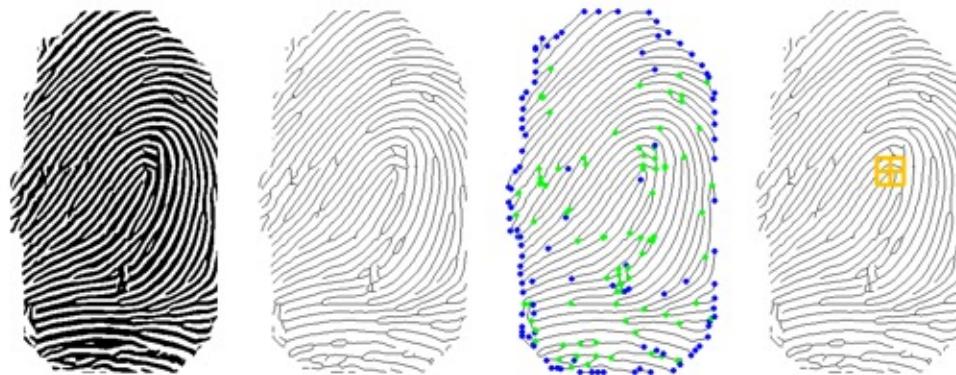
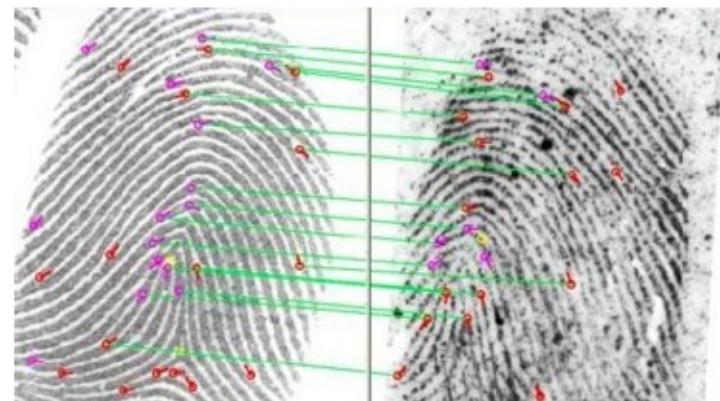
Paul Viola

Michael Jones



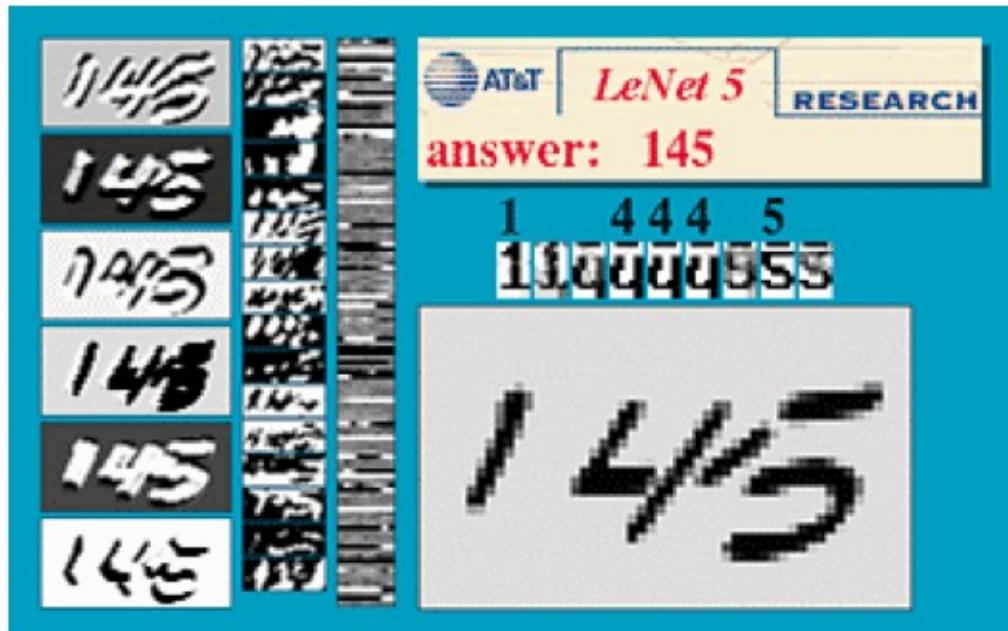
人脸识别

当前进展

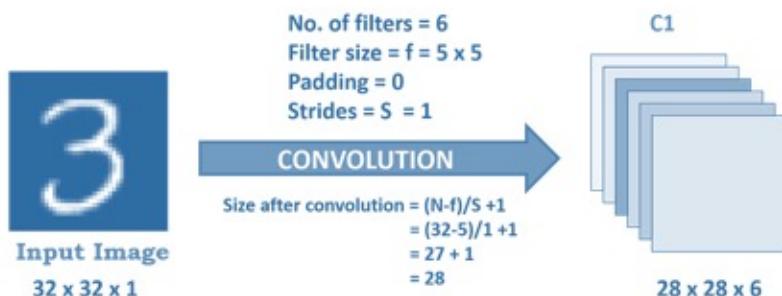


基于计算机视觉的生物识别技术，包括指纹识别、虹膜识别等。

当前进展



LeNet-5 手写字体识别网络



光学字符识别 (OCR)



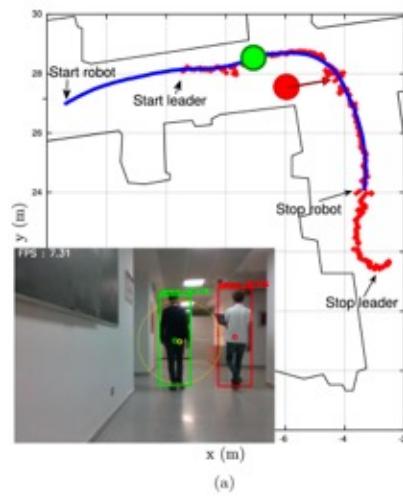
Yann LeCun 杨立坤



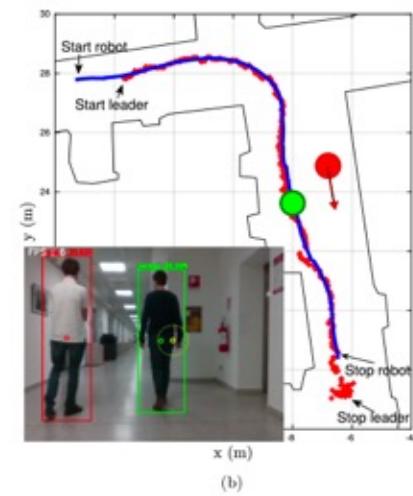
1993年的LeNet网络Demo

Optical Character Recognition

当前进展



(a)



(b)

机器人视觉

当前进展



天文学图像识别



谷歌眼镜



Esri vGIS (AR/增强现实)

当前进展

Mordvinsev et al., 2015
Gatys et al., 2016



图像风格迁移

当前进展



图像分类

			
mite mite black widow cockroach tick starfish	container ship container ship lifeboat amphibian fireboat drilling platform	motor scooter go-kart moped bumper car golfcart	leopard leopard jaguar cheetah snow leopard Egyptian cat
			
grille convertible grille pickup beach wagon fire engine	mushroom agaric mushroom jelly fungus gill fungus dead-man's-fingers	cherry dalmatian grape elderberry ffordshire bulterrier currant	Madagascar cat squirrel monkey spider monkey titi indri howler monkey

图像分类



ImageNet是一个用于视觉对象识别研究的大型图像数据库，是目前世界上图像数量最多的数据库。它包含了超过1400万张图像（附有标签），分属于约2万个类别。

2010-2017年，每年举办一次基于ImageNet的图像分类识别比赛（ILSVRC），从ImageNet抽取140万张图片（1000个类别），用不同的参赛模型进行分类识别。

图像分类



Top 1 结果:

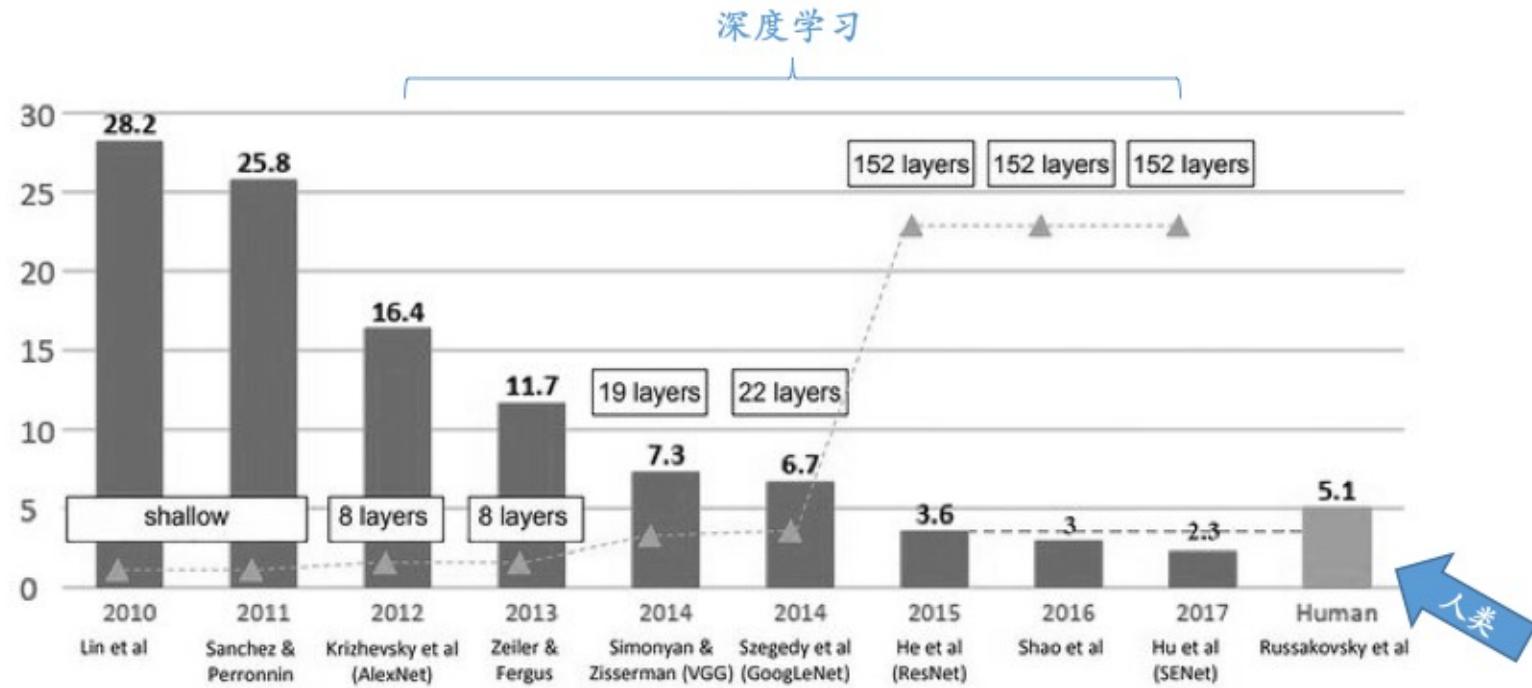
- 快艇 ✗



Top 5 结果:

- 快艇 (0.35)
- 飞机 (0.32) ✓
- 渡轮 (0.19)
- 潜水艇 (0.08)
- 海豚 (0.02)

图像分类



ILSVRC 历年冠军的Top 5分类结果错误率