



人工智能与信息社会

人工智能的应用

2019.11.11

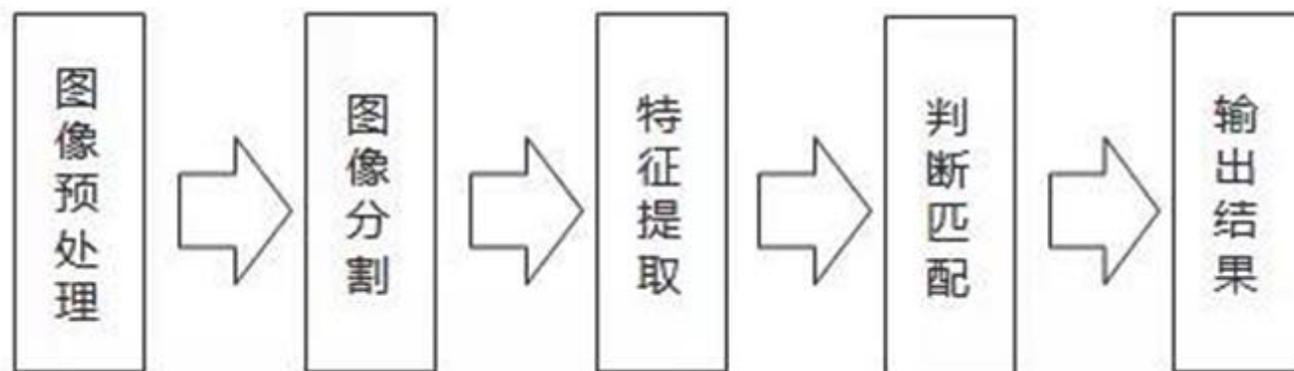
目录

- 图像识别与分类
- 医学影像分析
- 语音识别
- 人脸识别和情感计算
- 自动驾驶



什么是图像识别

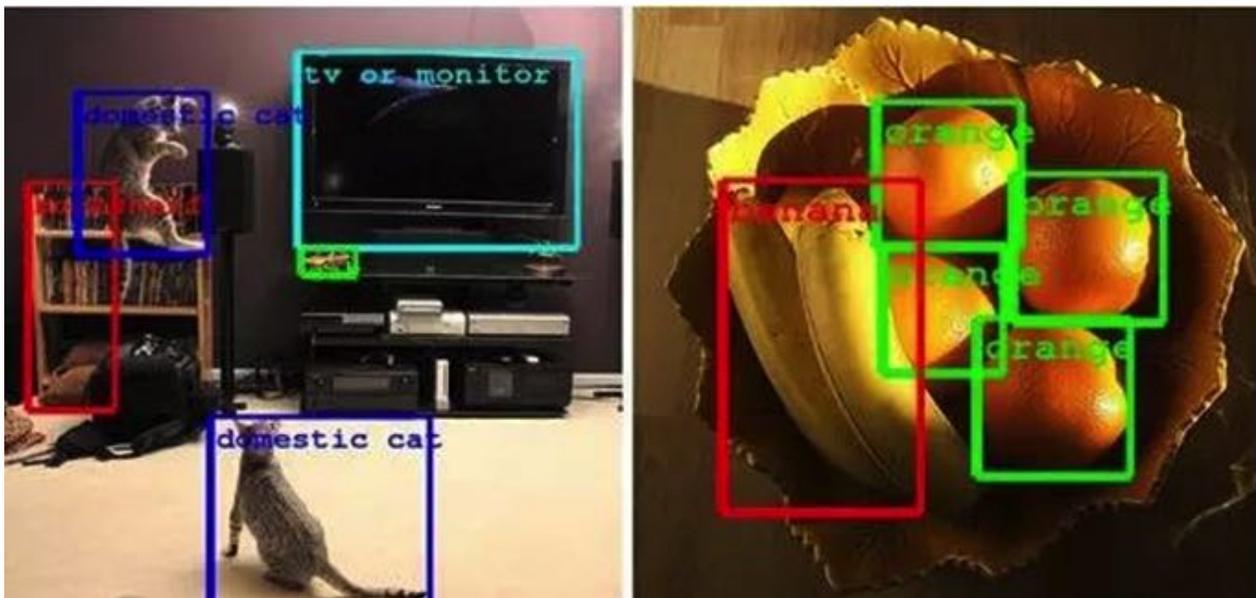
- 一般而言，传统图像识别系统主要由图像分割、图像特征提取以及图像识别分类构成。
- 图像分割将图像划分为多个有意义的区域，然后将每个区域的图像进行特征提取，最后根据提取的图像特征对图像进行分类。



图像识别系统图

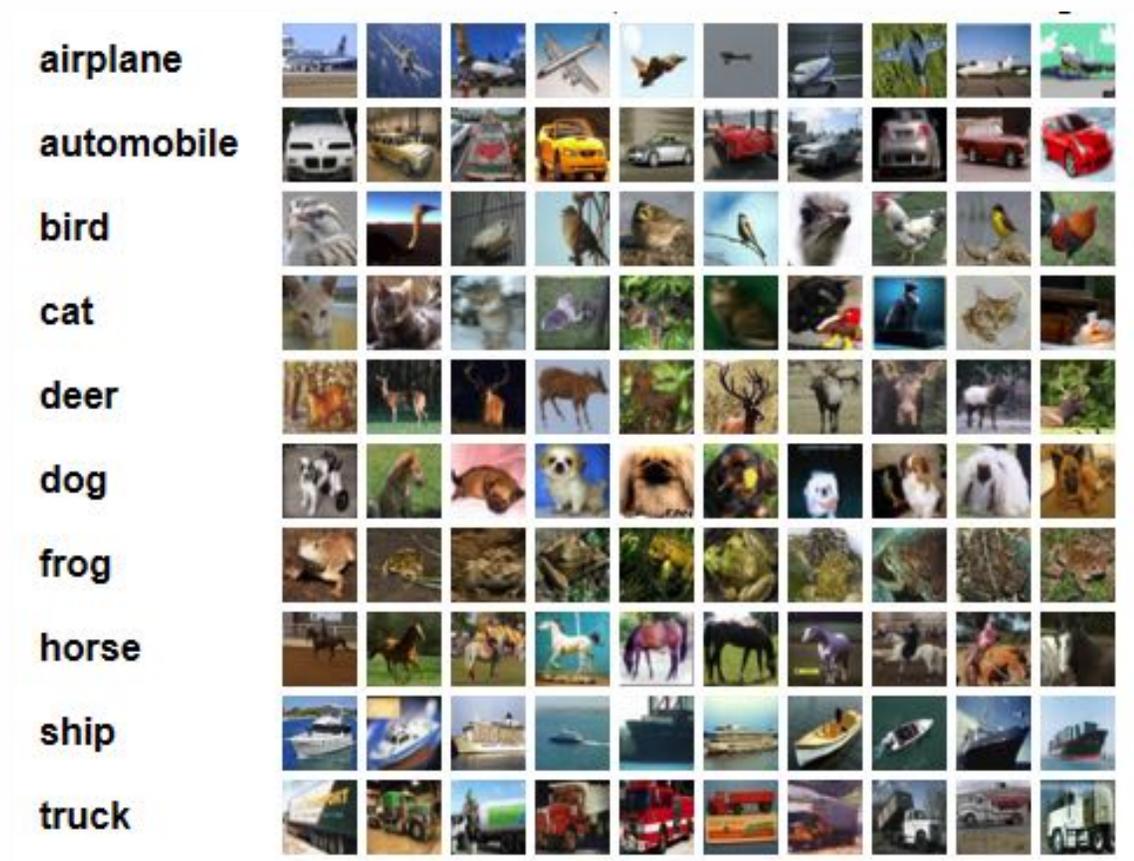
图像识别发展

- 从文字识别到数字图像处理与识别、物体识别，高性能芯片、摄像头和深度学习算法的进步都为图像识别技术发展提供了源源不断的动力。



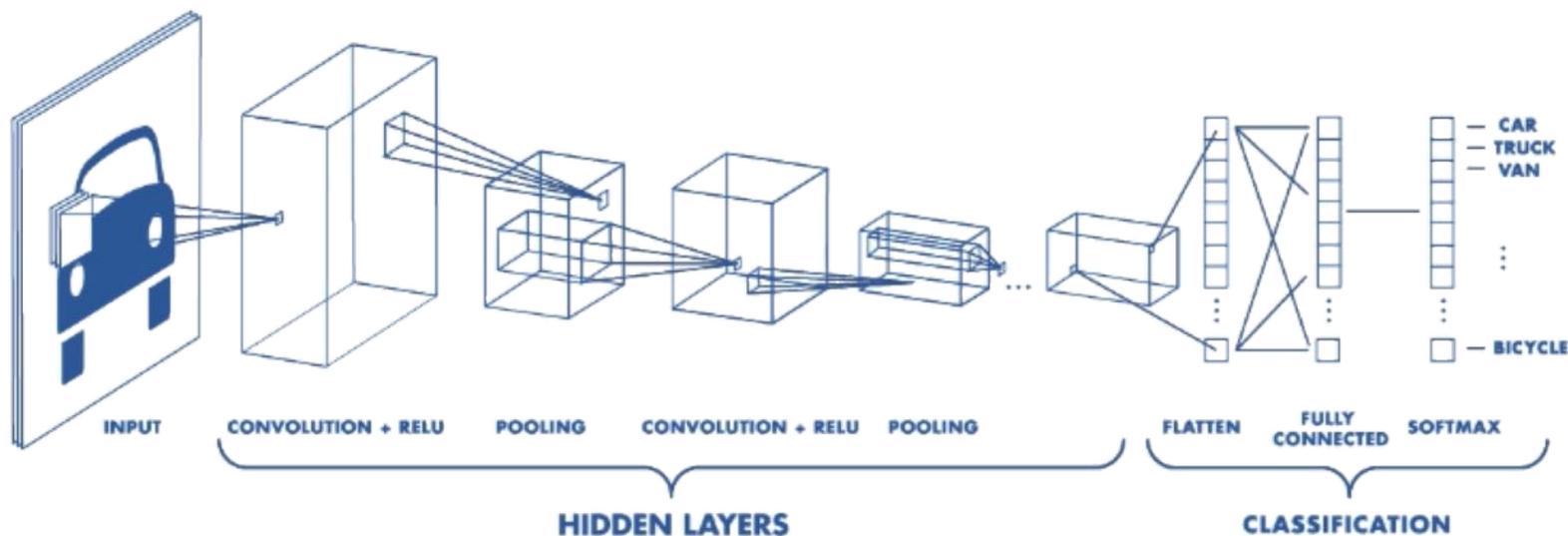
ImageNet数据集

- 包含1400多万幅图片，涵盖2万多个类别。
- 关于图像分类、定位、检测等研究工作大多基于此数据集展开，几乎成为了目前深度学习图像领域算法性能检验的“标准”数据集。



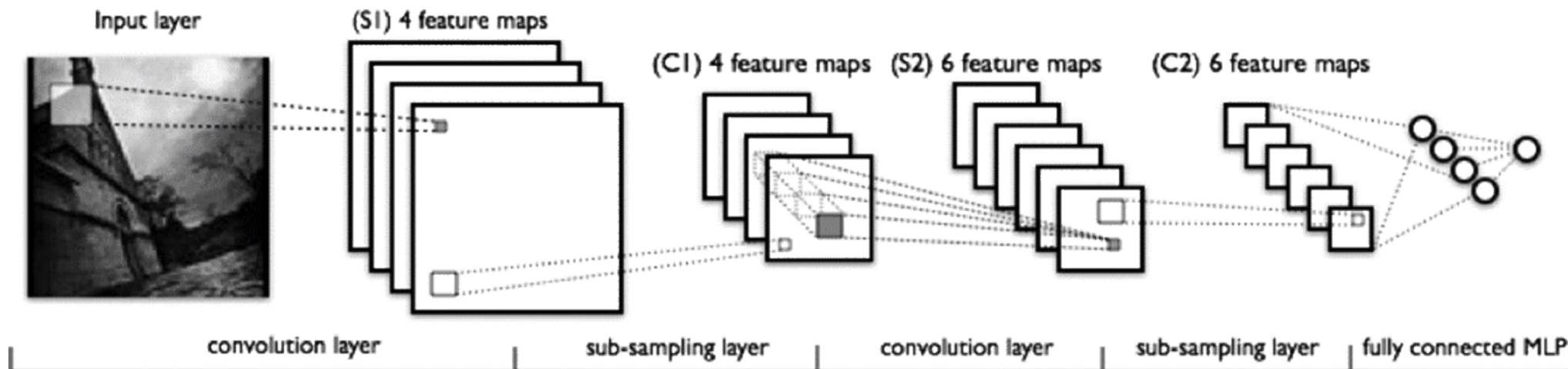
利用卷积神经网络进行图像识别

- 卷积神经网络是一种为了处理二维输入数据而特殊设计的多层人工神经网络。
- 不仅关注了全局特征，更利用了图像识别领域非常重要的局部特征，将局部特征抽取的算法融入到了神经网络中。



卷积神经网络的工作过程

- 网络对输入的真实图像运用不同的算子进行扫描提取不同的特征。并通过采样进行压缩。
- 进行多次特征提取和采样。
- 全连接层为每个节点的输出指定一个标签。



图像识别的主要应用

- 日渐成熟的图像识别技术已开始探索在各类行业的应用。



- ✓ 智能家居
- ✓ 金融
- ✓ 安防
- ✓ 医疗
- ✓ 交通系统

智能家居



- 在智能家居领域，通过图像识别技术识别出摄像头获取的图像内容。
- 若发现是可疑的人或物体，就及时报警给户主；而如果图像和主人的面部匹配，则会主动为主人开门。

安防



- 图像识别在安防领域应用较多，尤其在视频监控领域能直接帮助用户从视频画面中提取出“人”的信息。
- 大大提升监控系统的价值，使之成为打造智慧城市的核心环节。

金融



- 在金融领域，身份识别和智能支付将提高身份安全性与支付的效率和质量。
- 通过人脸识别进行一系列的验证、匹配和判定，从而快速完成身份核实。

医疗



- 将图像识别技术应用到医疗领域，可以更精准更快速地分辨X光片、MRI和CT扫描等图片。
- 既能诊断预防癌症，又能加速发现治病救命的新药。

交通系统

- 图像识别技术被广泛应用于交通运输领域：交通违章监测、交通拥堵检测、信号灯识别
- 提高交通管理者的工作效率，更好的解决城市交通问题。



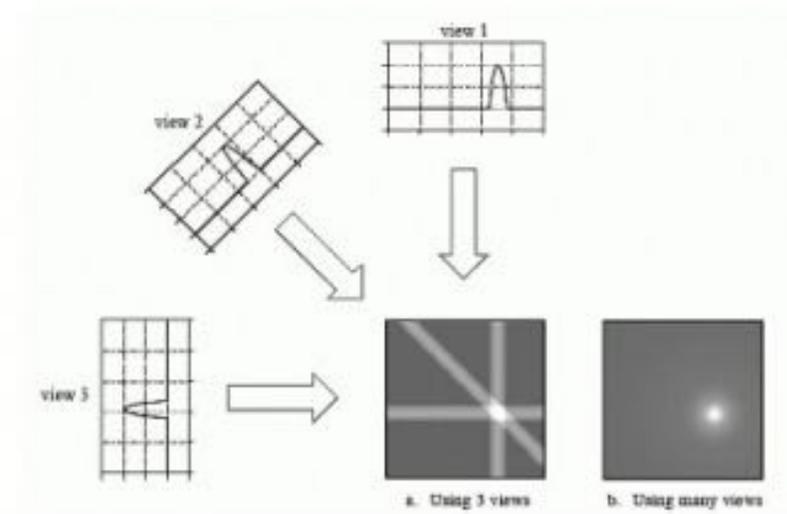
目录

- 图像识别与分类
- **医学影像分析**
- 语音识别
- 人脸识别和情感计算
- 自动驾驶



医学影像基础

- 硬件发展的突飞猛进，包括MR、CT等硬件的发展，这些成像技术让我们得到了很好的影像；
- 复杂数学工具の利用，通过这些方式可以对医学影像进行重建、分析与处理，从而得到清晰可见的医学图像



人工智能+医学影像



- 在医学影像的基础上，通过深度学习与大数据技术等，完成对影像的分类、目标检测、图像分割和检索工作。
- 是协助医生完成诊断、治疗工作的一种辅助工具。

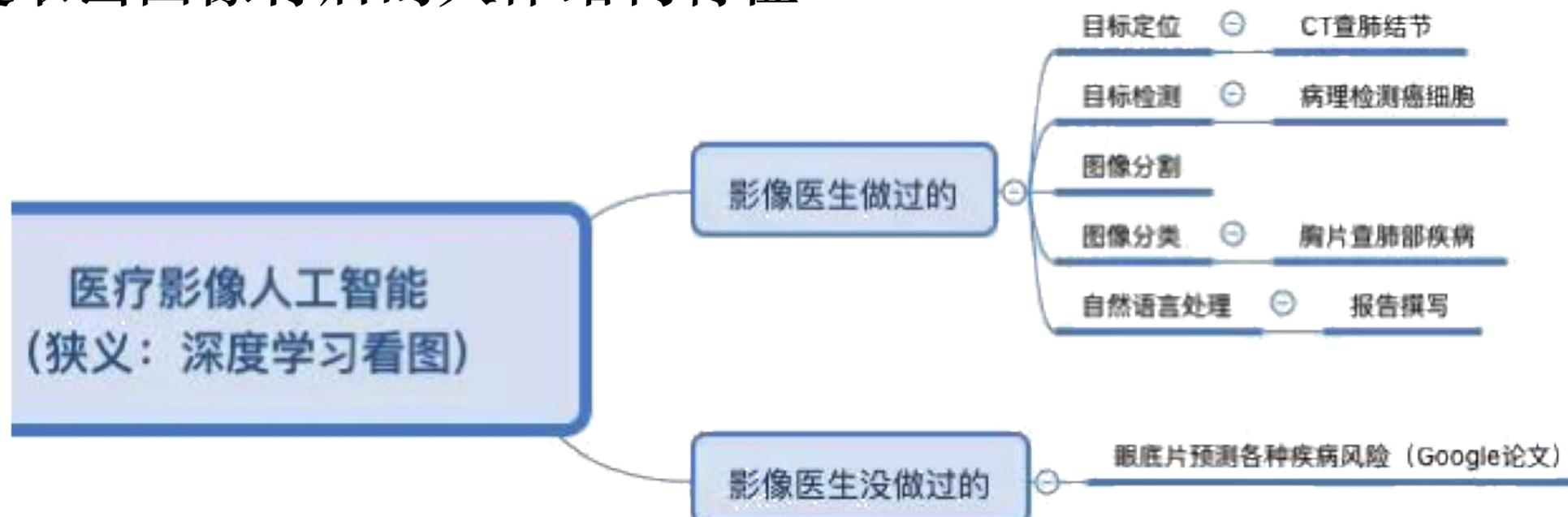
作用及优点



- 医学影像的解读需要长时间专业经验的积累，放射科医生的培养周期相对较长。
- 人工智能在对图像的检测效率和精度两个方面，都做得比专业医生更快，减少人为操作的误判率，提升医生看病效率和准确率。

深度学习方法

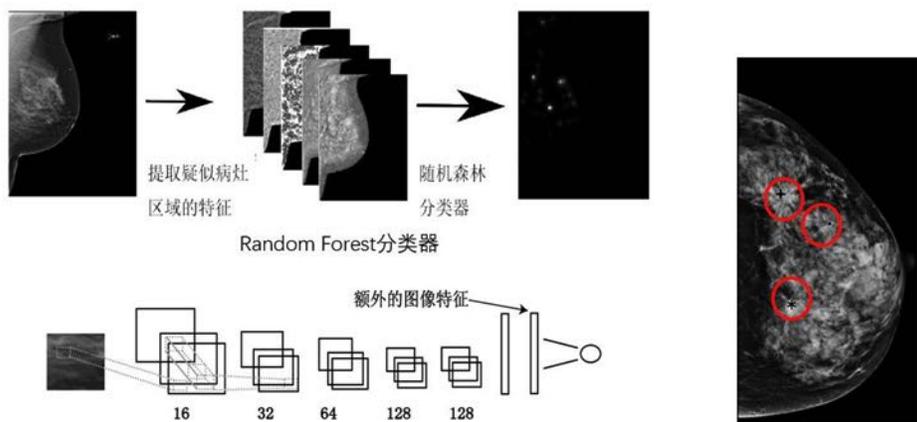
- 医学图像包含来自不同组织、不同形态的人体器官，深度学习包含多层感知器，可以通过组合低层特征形成更加抽象的高层特征，提取出图像背后的人体结构特征。



应用：计算机辅助诊断

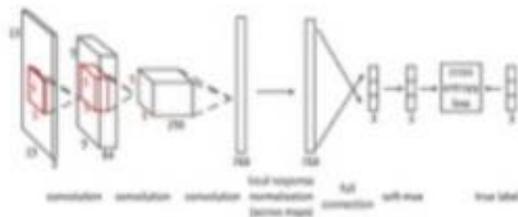
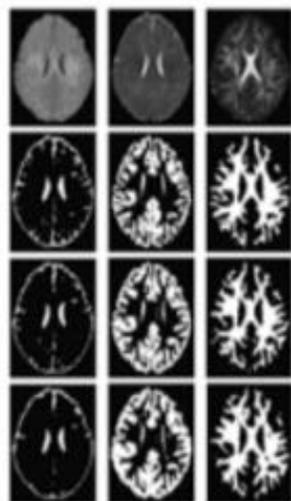
- 病灶检测，对可疑病灶进行识别和勾画。
- 病灶量化诊断，帮助医生鉴别疾病良恶性、分型分期等。
- 治疗决策，通过相关性分析，支持临床医生进行科学合理的治疗决策。

乳腺病灶识别方法



应用：图像分割

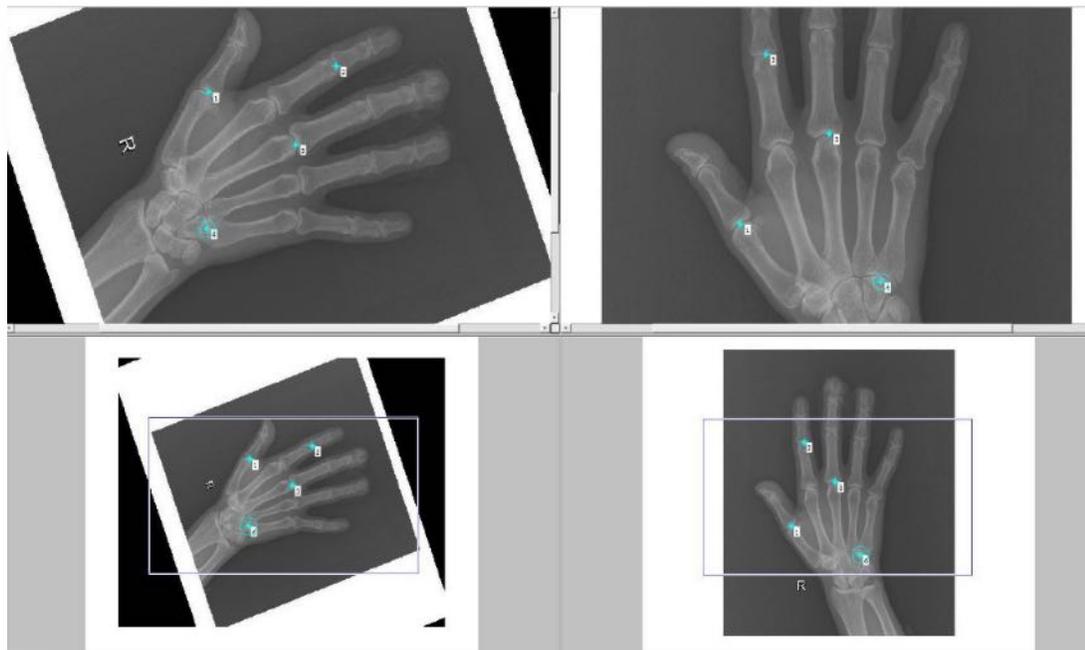
- 主要是对身体组织做明确分割，精度比医生手动分割更高，从而更加精准地定量评价治疗前后的效果。
- 2015年的一篇文章提到利用卷积神经网络CNN，自动将大脑灰质、白质、脑脊髓翼自动分割，从而分析大脑的病变。



		Sub. 1	Sub. 2	Sub. 3
CSF	CNN	0.8323	0.8314	0.8304
	RF	0.8192	0.8135	0.8323
	SVM	0.7409	0.7677	0.7733
	CLS	0.8064	0.8152	0.732
	MV	0.7072	0.6926	0.6826
GM	CNN	0.8531	0.8572	0.8848
	RF	0.8288	0.8482	0.8772
	SVM	0.7933	0.7991	0.8294
	CLS	0.8298	0.8389	0.8498
	MV	0.849	0.8442	0.8525
WM	CNN	0.8798	0.8116	0.8824
	RF	0.8612	0.7816	0.8687
	SVM	0.8172	0.7404	0.7623
	CLS	0.8383	0.8054	0.7998
	MV	0.8631	0.8002	0.8504

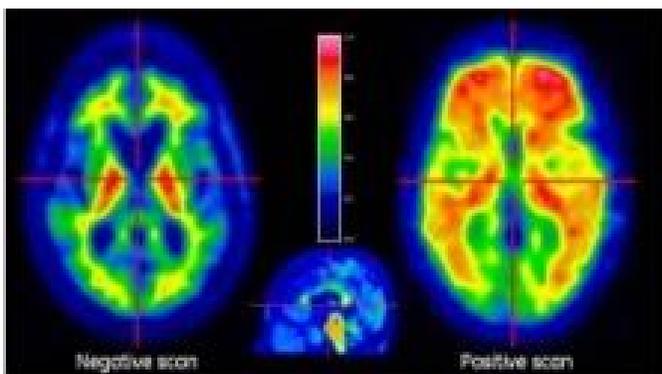
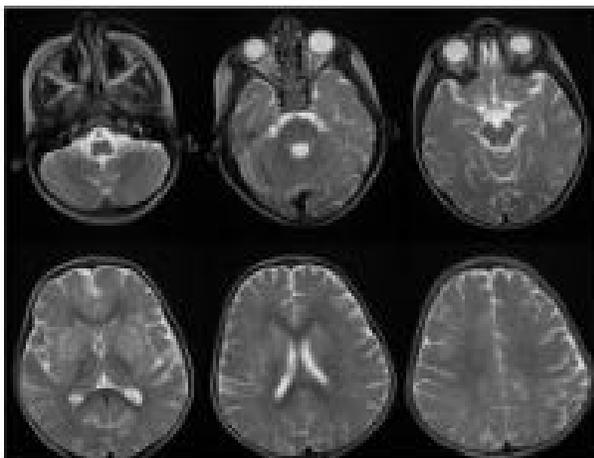
W. Zhang et. al. Deep Convolutional Neural Networks for Multi-modality Isointense Infant Brain Image Segmentation, NeuroImage, 2015

应用：图像配准



- 在对不同模式医学图像或多参数医学图像进行图像融合前，必须对图像进行精确配准。
- 一般采用非监督学习方法提取图像特征，接着采用卷积神经网络回归的方法来进行2D或3D图像配准。

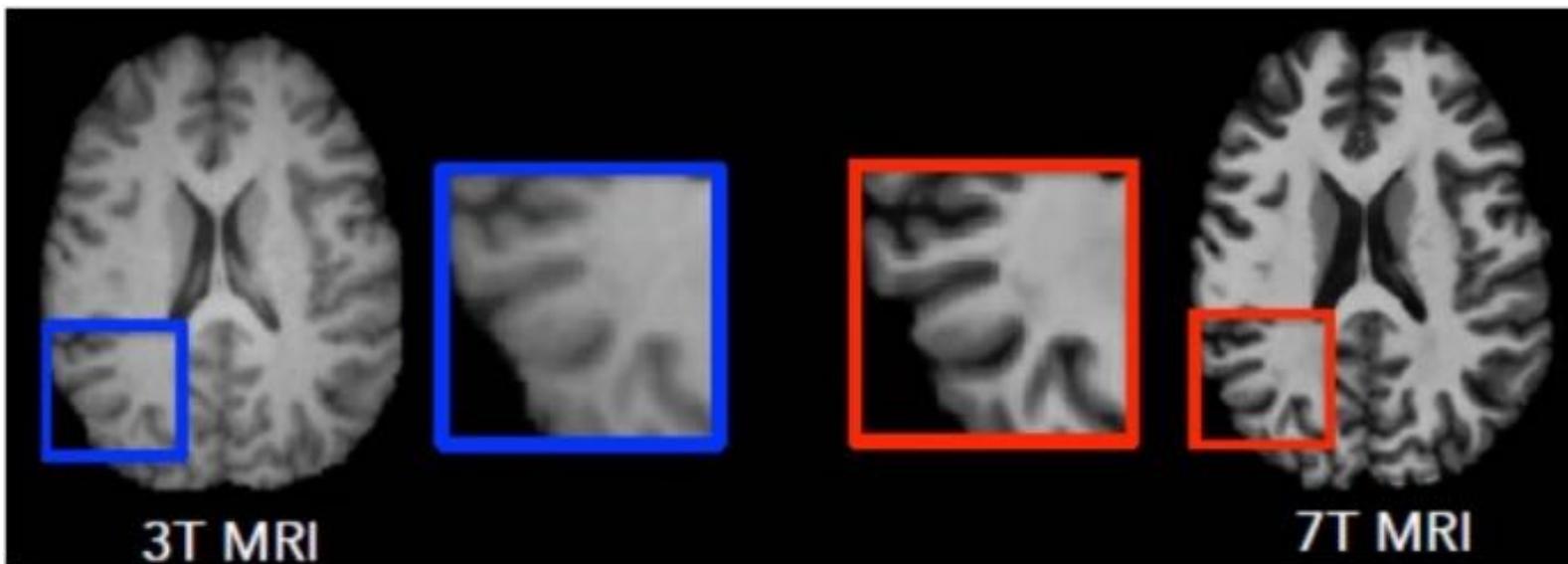
应用：图像融合



- 图像有结构性与功能性之分。
- 结构性：可以得到组织的结构性特征，但无法看到生物有机代谢的情况。
- 功能性图像，它可以提示代谢的衰变与下降，或功能性的疾病，但图像空间解析度差。
- 需要影像融合，将不同类型的图像结合在一起，这样可以了解到组织与器官的病变。

应用：图像重建

- 医学中常见的核磁机器是1T（特斯拉，即磁场感应强度）和1.5T，而7T的设备昂贵，需要很多费用，但7T图像的性噪比强
- 通过深度学习将3T变7T图像。



人工智能+医学影像



- 基于影像的医学诊断是目前人工智能关注较多的领域，
- “AI+医学影像”被多位业内人士认为最有可能率先实现商业化。

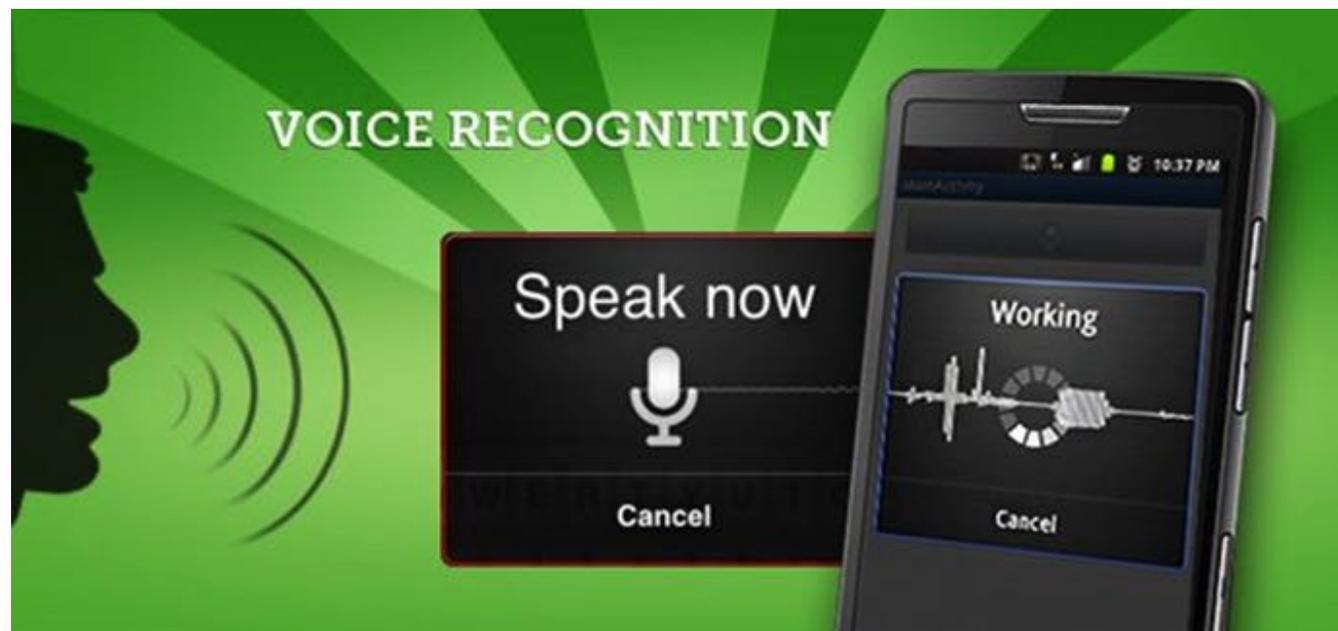
目录

- 图像识别与分类
- 医学影像分析
- **语音识别**
- 人脸识别和情感计算
- 自动驾驶



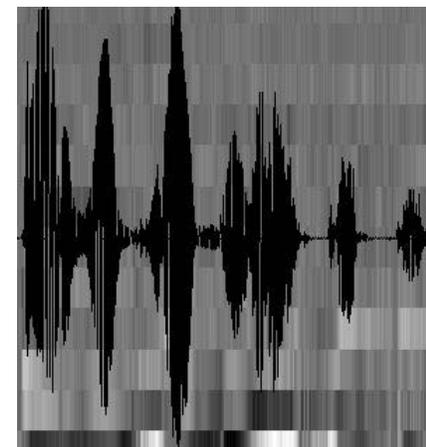
语音识别技术

- 语音识别技术，也被称为自动语音识别Automatic Speech Recognition(ASR)，其目标是将人类的语音中的词汇内容转换为相应的文字。



语音识别基本原理

- 我们知道声音其实是一种波，如果要对于声音进行分析，就需要对于声音进行分帧。
- 也就是把声音按照时间切成若干个小段，每一小段称为一帧。



语音识别基本原理

元音	12个单元音		长元音	[a:]	[ɔ:]	[i:]	[ɔ:]	[u:]	OC课堂			
			短元音	[ʌ]	[ə]	[ɪ]	[ʊ]	[u]	[æ]	[e]		
	8个双元音		[aɪ]	[eɪ]	[oɪ]	[iə]	[eə]	[uə]	[aʊ]	[əʊ]		
辅音	10对	清辅音	[p]	[t]	[k]	[f]	[s]	[ʃ]	[tʃ]	[tr]	[θ]	[ts]
		浊辅音	[b]	[d]	[g]	[v]	[z]	[ʒ]	[dʒ]	[dr]	[ð]	[dz]
	3个鼻音		[m]	[n]	[ŋ]							
	3个似拼音		[h]	[r]	[l]							
	2个半元音		[w]	[j]								

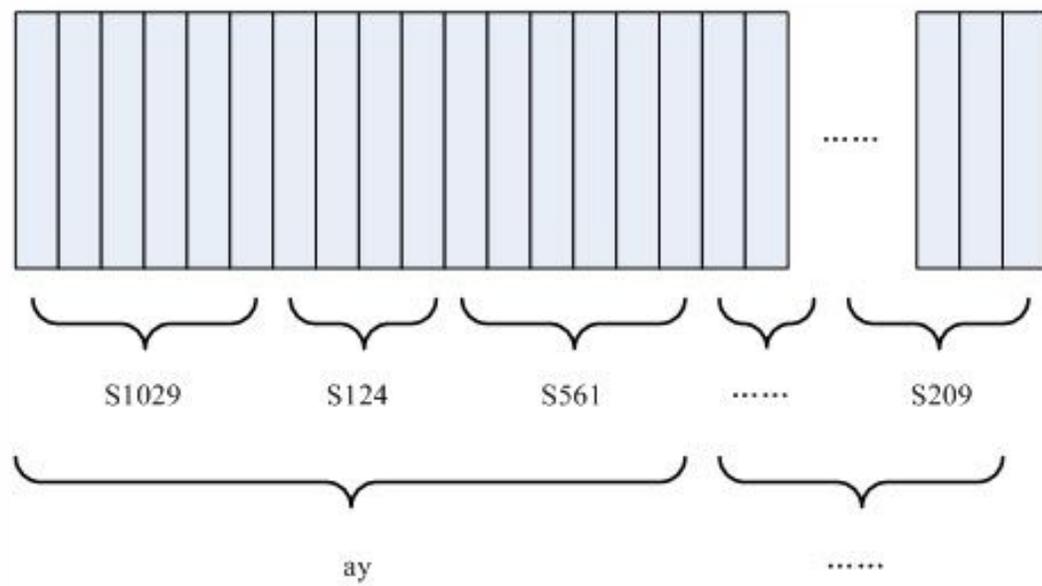
- 单词的发音由音素构成。音素是语音中的最小的单位，依据音节里的发音动作来分析，一个动作构成一个音素。
- 对英语，一种常用的音素集是卡内基梅隆大学的一套由39个音素构成的音素集。汉语一般直接用全部声母和韵母作为音素集。
- 状态：比音素更细致的语音单位。通常把一个音素划分成3个状态。

语音识别基本原理



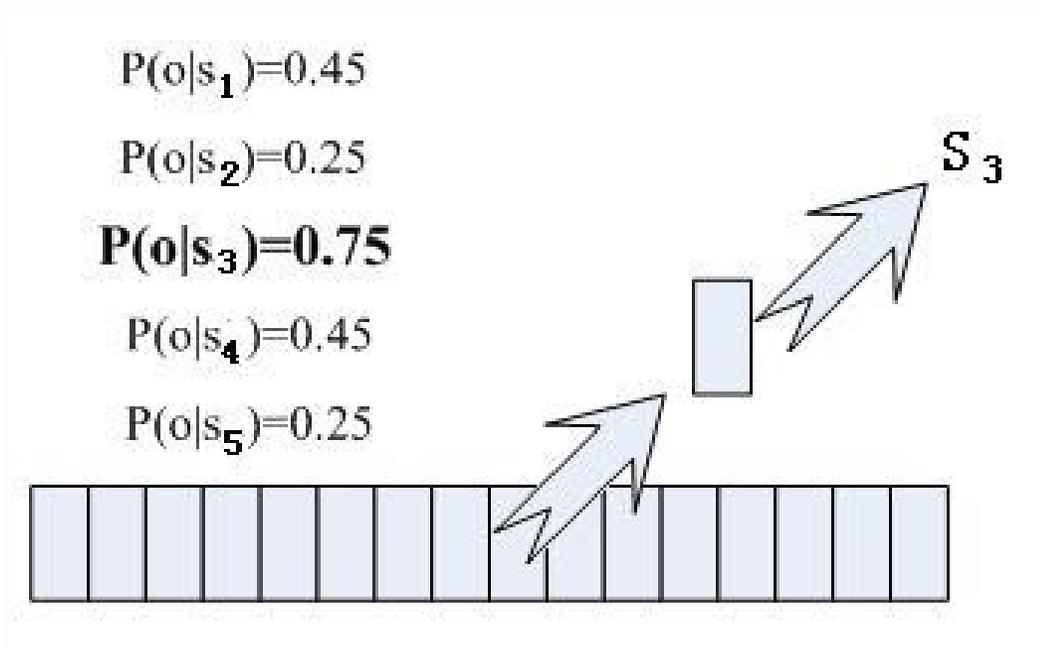
- 1、语音分帧
- 2、声学特征提取
- 3、把帧识别成状态
- 4、把状态组合成音素
- 5、把音素组合成单词

语音识别基本原理



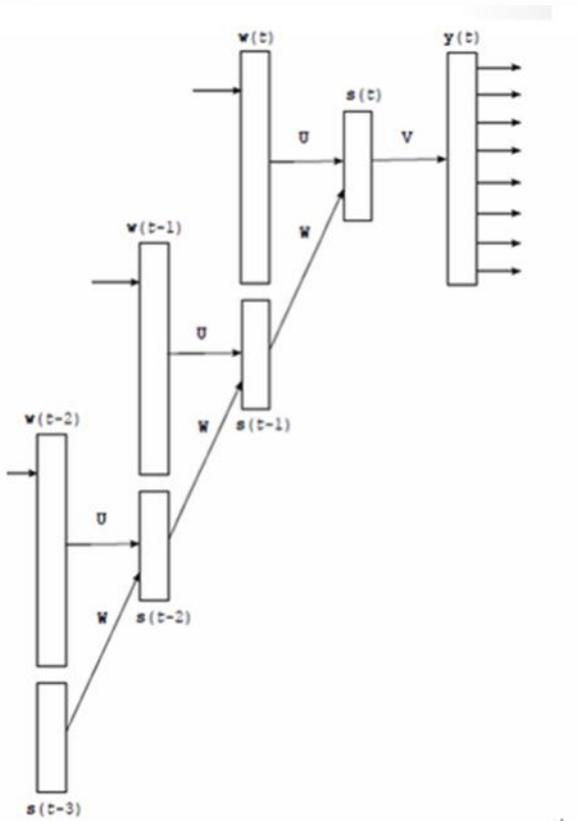
- 每个小竖条代表一帧，若干帧语音对应一个状态，每三个状态组合成一个音素，若干个音素组合成一个单词。
- 一般认为某帧对应哪个状态的概率最大，那这帧就属于哪个状态。

神经网络在语音识别中的应用



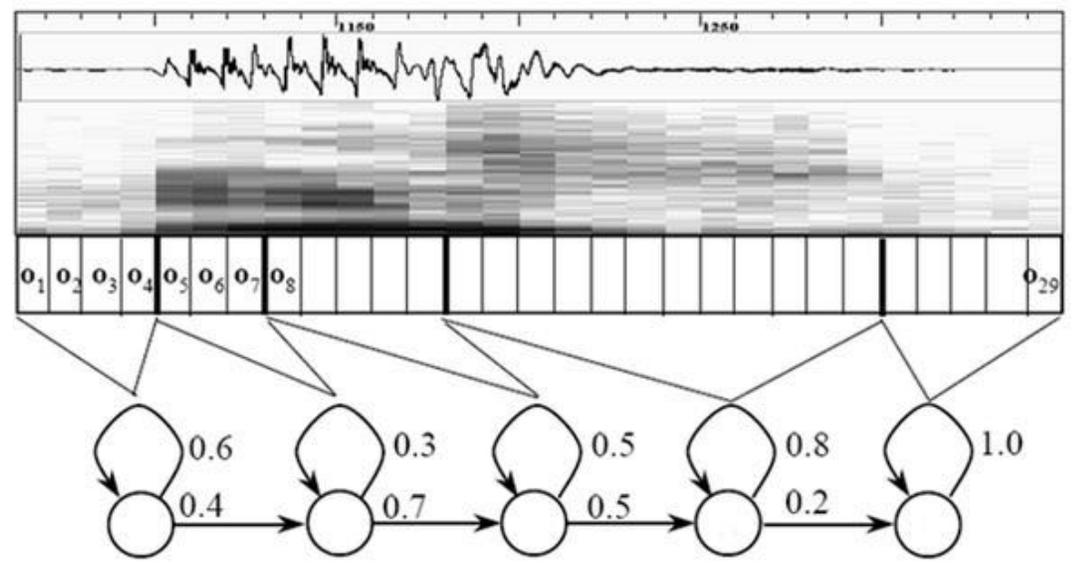
- 通过巨大数量的语音数据，经过神经网络监督训练得到“声学模型”。
- 通过声学模型里的参数，就可以知道帧和状态对应的概率。

神经网络在语音识别中的应用



- 神经网络不仅可以编码最近的几个词，还可以把前文中的所有词（称作“历史”）中的各种信息都作为输入特征。
- 由于历史是一个序列，可以采用递归神经网络（RNN）来建立声学模型。

神经网络在语音识别中的应用



- 遇到问题：每一帧都会得到一个状态，最后整个语音就会得到一堆乱七八糟的状态。
- 解决方法：是使用隐马尔可夫模型。第一步，构建一个状态网络。第二步，从状态网络中寻找与声音最匹配的路径。
- 这样就把结果限制在预先设定的网络中，避免了刚才说到的问题。

语音识别应用



- 语音识别技术的应用包括语音拨号、语音导航、室内设备控制等。
- 语音识别技术与其他自然语言处理技术如机器翻译及语音合成技术相结合，可以构建出更复杂的应用，例如同声传译。

语音识别应用：语音输入法



- 通过语音识别输入文字，最高速度能够达到1分钟400字，比普通键盘输入更加高效。
- 科大讯飞的语音输入，不仅支持中文录入、中文转英文等功能，还支持粤语、四川话、东北话、上海话、闽南语等多种方言输入。

语音识别应用：个人助理



语音识别应用：个人助理

- 小冰是微软推出的一个人工智能聊天机器人，已经可以创作诗歌、撰写新闻、主持节目，已在北京人民广播电台开播节目。
- 无论从用户数量、活跃度还是交互流量来看，微软小冰均是目前全球最大规模流量的对话式人工智能产品。

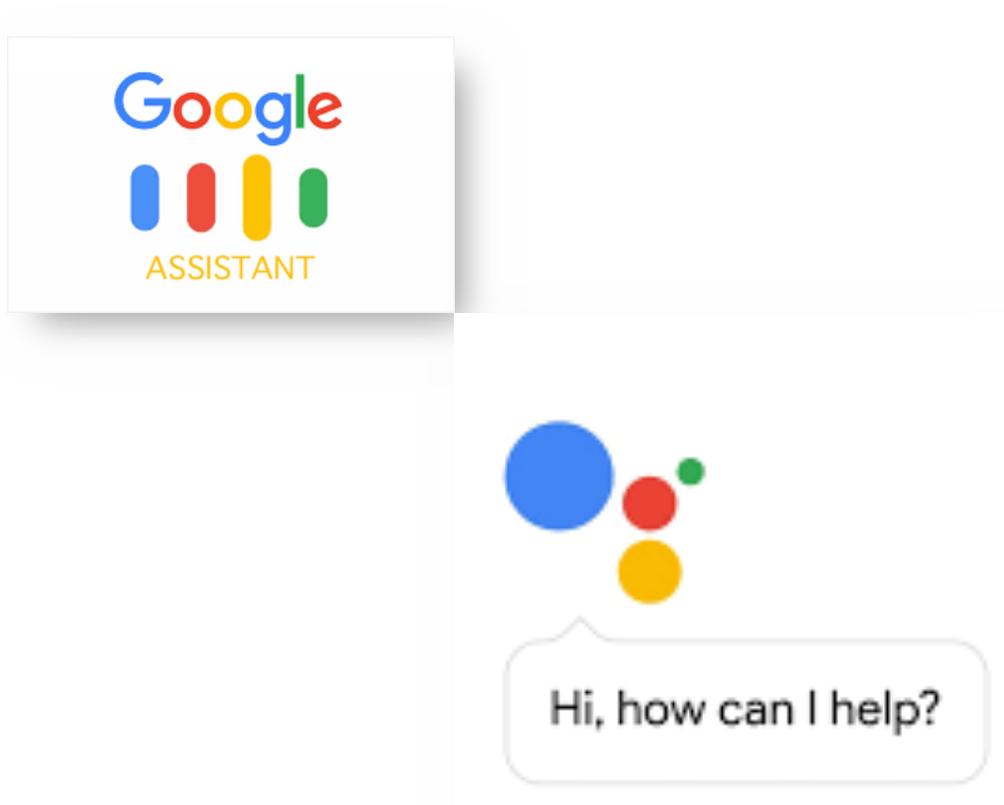


语音识别应用：个人助理



- Siri是一款内置在苹果iOS系统中的人工智能助理软件。
- 利用自然语言处理技术，用户可以使用自然的对话与手机进行交互，完成搜索数据、查询天气、设置手机日历、设置闹铃等许多服务。

语音识别应用：个人助理



- Google Assistant 结合了谷歌积累多年的技术，其“持续性对话”功能让机器与人的交流更为自然。
- 智能语音助手和搜索引擎是相辅相成的，更聪明的搜索逻辑能够更快的帮助用户找到答案。

目录

- 图像识别与分类
- 医学影像分析
- 语音识别
- 人脸识别和情感计算
- 自动驾驶



人脸识别的定义

- 人脸识别(Face Recognition)是一种依据人的面部特征(如统计或几何特征等)，自动进行身份识别的一种生物识别技术。

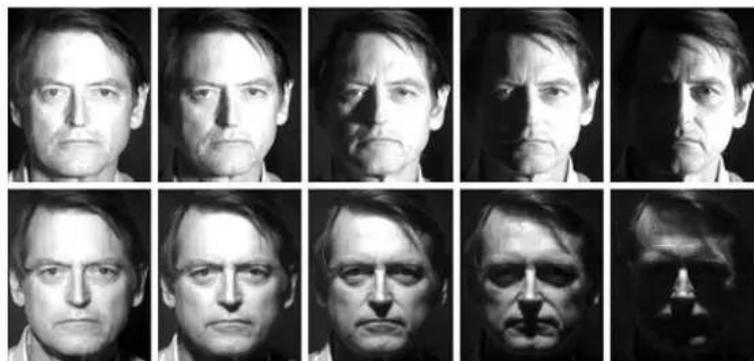


人脸识别一般流程

- 图像采集、人脸检测、特征提取（三维建模）、模型对比、结果输出。



人脸采集



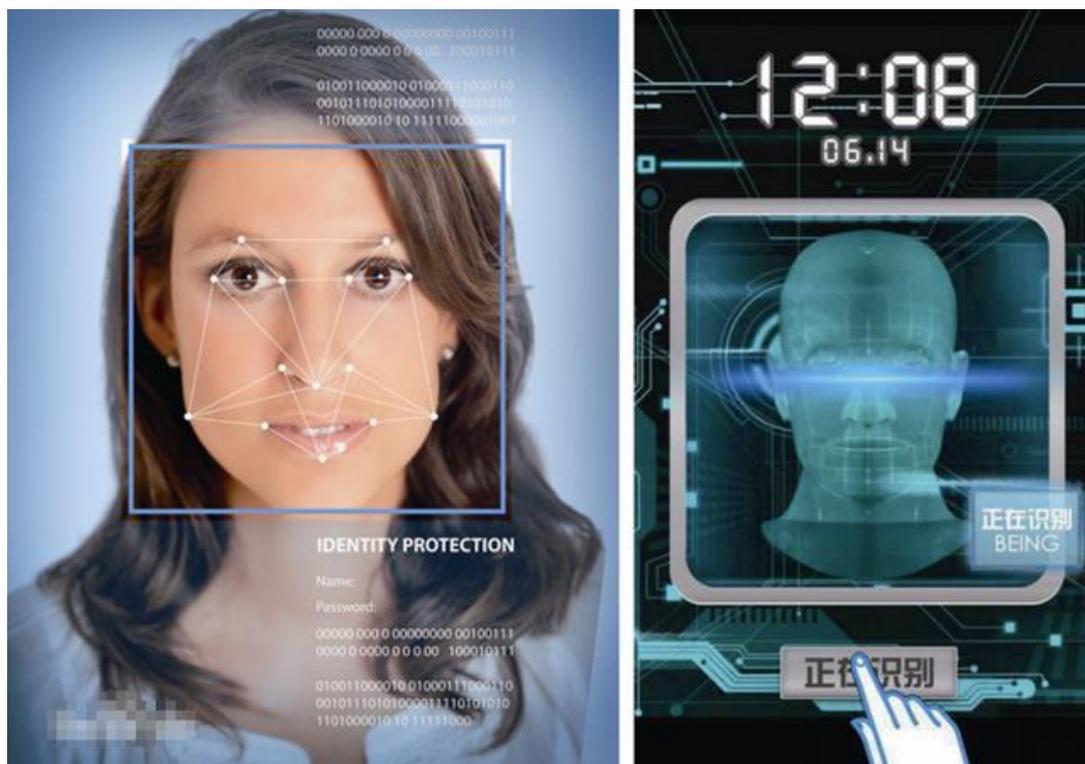
光照问题



遮挡问题

- 影响因素有：
- 图像大小
- 图像分辨率
- 光照环境
- 模糊程度
- 遮挡程度
- 采集角度

人脸检测



- 在图像中准确标定出人脸的位置和大小，并把其中有用的信息挑出来。
- 如直方图特征、颜色特征、模板特征、结构特征等
- 然后利用这些信息来达到人脸检测的目的。

人脸检测

- 基于检测出的特征采用综合分类学习算法（Adaboost 算法），挑选出代表性特征。
- 按照加权的方式将弱分类器构造为一个强分类器，再将训练得到的若干强分类器串联组成一个层叠分类器，有效地提高检测速度。



图像预处理



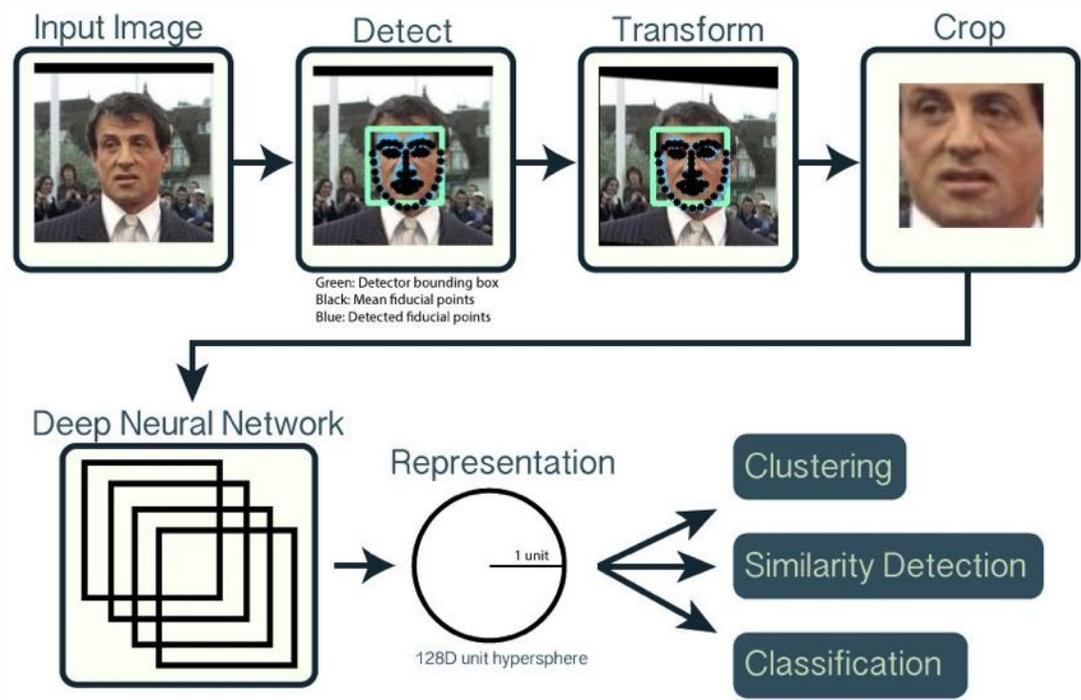
- 系统获取的原始图像由于受到各种条件的限制和随机干扰，往往不能直接使用。
- 需要在图像处理的早期阶段对它进行灰度矫正、噪声过滤等图像预处理。

人脸特征提取：传统基于知识表征方法



- 根据人脸器官的形状描述以及它们之间的距离特性来获得有助于人脸分类的特征数据。
- 其特征分量通常包括特征点间的欧氏距离、曲率和角度等对人脸进行特征建模。

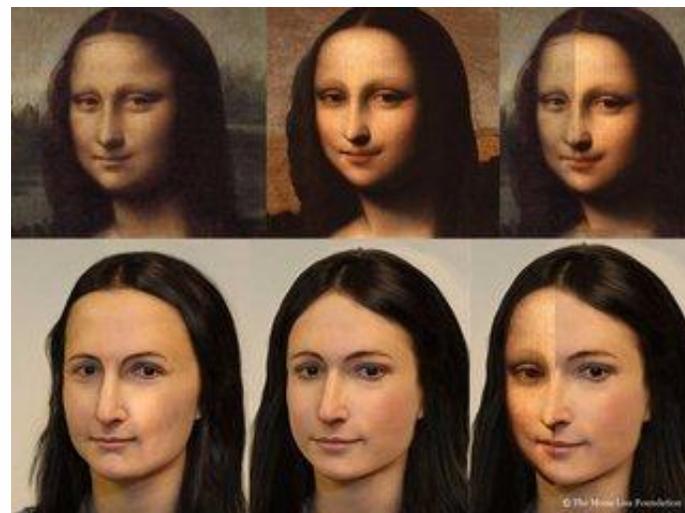
人脸特征提取：基于神经网络方法



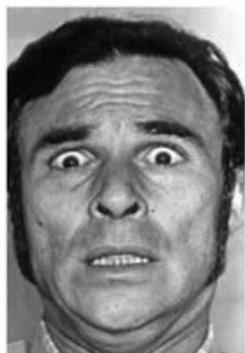
- 基于神经网络人脸进行特征建模
- 使用深度卷积网络，将输入的人脸图像，转换成一个向量的表示。
- 在理想的状况下，我们希望“向量表示”之间的距离就可以直接反映人脸的相似度。

匹配与识别

- 提取的人脸特征值数据与数据库中存贮的特征模板进行搜索匹配。
- 通过设定一个阈值，将相似度与这一阈值进行比较，来对人脸的身份信息进行判断。



情感计算：人脸表情识别



Fearful



Angry



Sad



Happy



Disgusted



Surprised

- 人类主要有六种基本情感：愤怒、高兴、悲伤、惊讶、厌恶、恐惧。
- 感情表达包括言词、声音、面部表情，其中一半以上可以从面部表情看出。
- 大多数表情识别是基于这六种情感及其拓展情绪实现的。

情感计算



- 表情识别的四个步骤中人脸检测、人脸配准、特征提取与人脸识别中的部分类似。
- 最后的表情分类则要根据提取的图像特征判断该表情所属的基本感情类别。

主要困难点

- 表情的精细化程度划分：每种情绪最微弱的表现是否需要分类。
- 表情类别的多样化：是否还需补充其他类别的情绪。
- 六种情绪在一些场景下远不能识别人类的真实情绪，因此还有精细表情识别、混合表情识别、非基本表情识别等细致领域的研究。

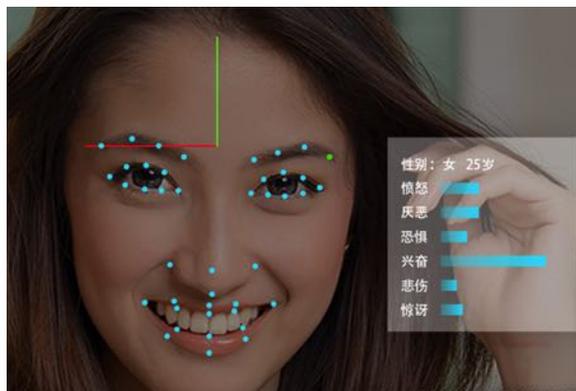
表情分析工具：FACS

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
 Inner Brow Raiser *AU 41	 Outer Brow Raiser *AU 42	 Brow Lowerer *AU 43	 Upper Lid Raiser AU 44	 Cheek Raiser AU 45	 Lid Tightener AU 46
 Lid Droop	 Slit	 Eyes Closed	 Squint	 Blink	 Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
 Nose Wrinkler AU 15	 Upper Lip Raiser AU 16	 Nasolabial Deepener AU 17	 Lip Corner Puller AU 18	 Cheek Puffer AU 20	 Dimpler AU 22
 Lip Corner Depressor AU 23	 Lower Lip Depressor AU 24	 Chin Raiser *AU 25	 Lip Puckerer *AU 26	 Lip Stretcher *AU 27	 Lip Funneler AU 28
 Lip Tightener	 Lip Pressor	 Lips Part	 Jaw Drop	 Mouth Stretch	 Lip Suck

- (Facial Action Coding System)
- 人类在表达同一情感时，面部肌肉运动具有一定的规律，可以基于运动单元给出面部动作编码系统。
- FACS系统致力于通过特定的符号来“详细描述内心情感与面部表情的关系”。

情感计算的应用场景

- 将帮助有自闭症的人群更好的融入社会。
- 可以迅速定位那些需要帮助或有学习障碍的学生，并可由此考核教师的素质。
- 用于心理医生判断病人是否明白其指示和病人的真实情感，以便更好的治疗。



目录

- 图像识别与分类
- 医学影像分析
- 语音识别
- 人脸识别和情感计算
- **自动驾驶**



人工智能与自动驾驶



- 当前人工智能的主要细分技术，包括机器视觉，深度学习，强化学习，传感器技术等均在自动驾驶领域发挥着重要的作用。
- 自动驾驶行业发展的瓶颈主要在于这些人工智能底层技术上能否实现突破。

自动驾驶现状

代表公司	时间	主要技术	主要产品
谷歌	最早在2009年曝光自动驾驶原型车	谷歌无人驾驶汽车依靠激光测距仪、视频摄像头、车载雷达、传感器等获得环境感知和识别能力，确保行驶路径遵循谷歌街景地图预先设定的路线	谷歌无人驾驶汽车
特斯拉	2003年	主要采用常规的雷达、相机、传感器、摄像机等进行环境感知和识别，通过基于车联网的协同式辅助驾驶技术进行智能信息交互	搭配Autopilot功能的特斯拉Model系列车型
Uber	2016年5月	自动驾驶汽车是由福特和沃尔沃XC90越野车改装而成，汽车上配备了数十个感应器和摄像头、激光、雷达和GPS信号接收器	Uber测试自动驾驶系统是搭配在沃尔沃SUV上。
奔驰	***	利用GPS、雷达，以及摄像头来识别交通信号、行人，以及其他障碍物。	自动公交“FutureBus”
亚马逊	2013年提出无人机送货服务	无人机	无人机送货
京东	***	自动驾驶	无人配送车

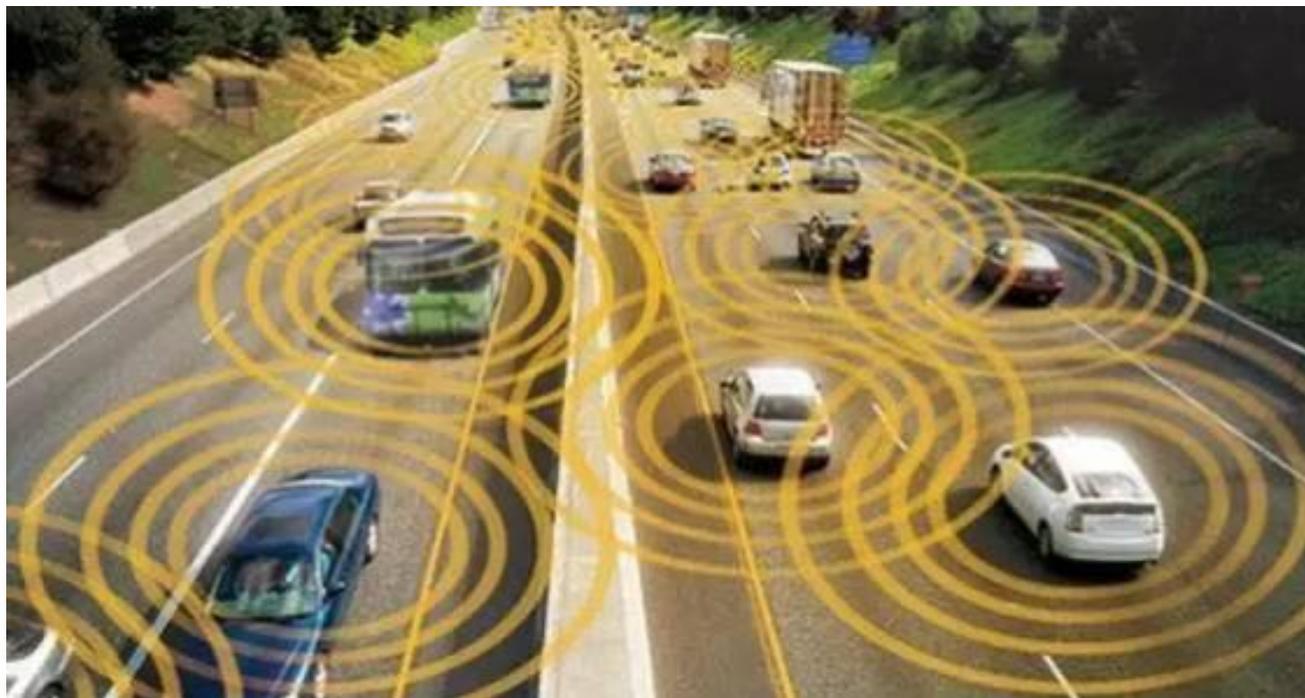
- 海内外各大企业争相加大人工智能在汽车领域应用的研发投入。尤其是非传统的汽车厂商，包括各大IT和互联网公司如Google，Tesla，中国的蔚来汽车等。

自动驾驶的SAE分类标准

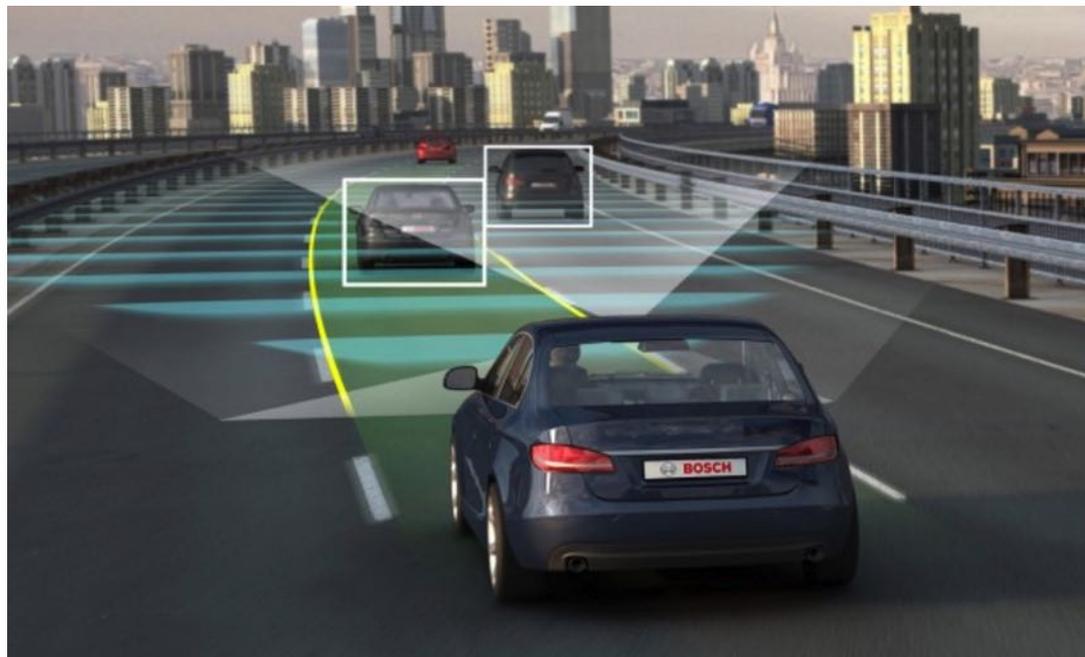
- Level 0 人工驾驶
- Level 1 辅助驾驶
- Level 2 半自动自动驾驶
- Level 3 高度自动驾驶
- Level 4 超高度自动驾驶
- Level 5 全自动驾驶
- 就目前来说，还在进行L3、L4级的研发

模块组成

- 环境感知模块、驾驶行为决策模块、运动控制模块。
- 自动驾驶中最重要、最具挑战的模块就是行为决策模块。



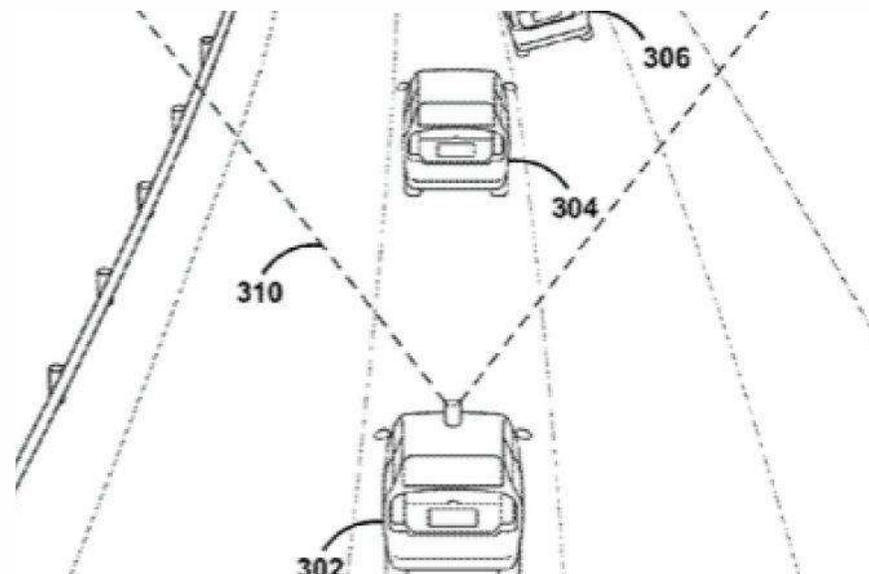
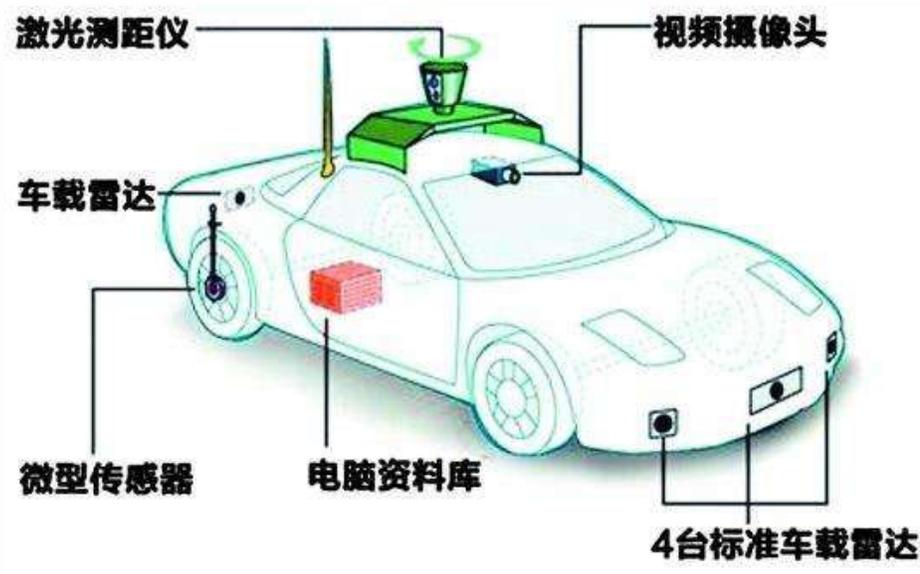
决策模型



- 无人驾驶中的行为决策需要根据实时路网信息、交通环境信息和自身驾驶状态信息，产生的安全快速的自动驾驶决策。
- 深度学习可以用来做环境的感知，而强化学习可以用来做控制行为的决策模型，这样就可以构成一个完整的自动驾驶系统。

障碍物识别

- 这个问题的解决方案是传感器融合算法，利用多个传感器所获取的关于环境全面的信息。
- 通过人工智能的融合算法来实现障碍物识别与跟踪和躲避。

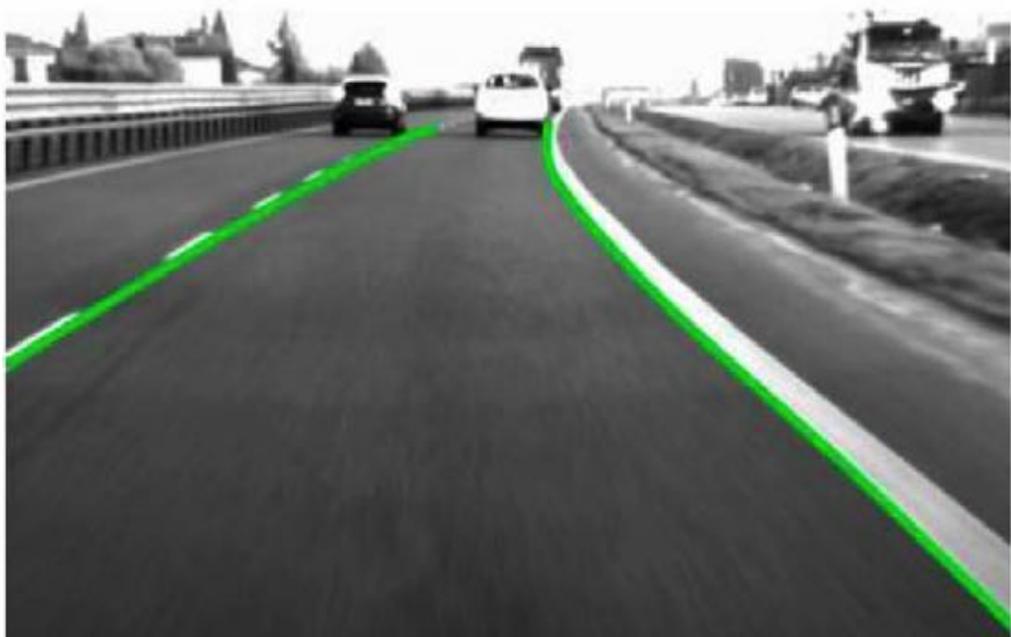


交通标志识别



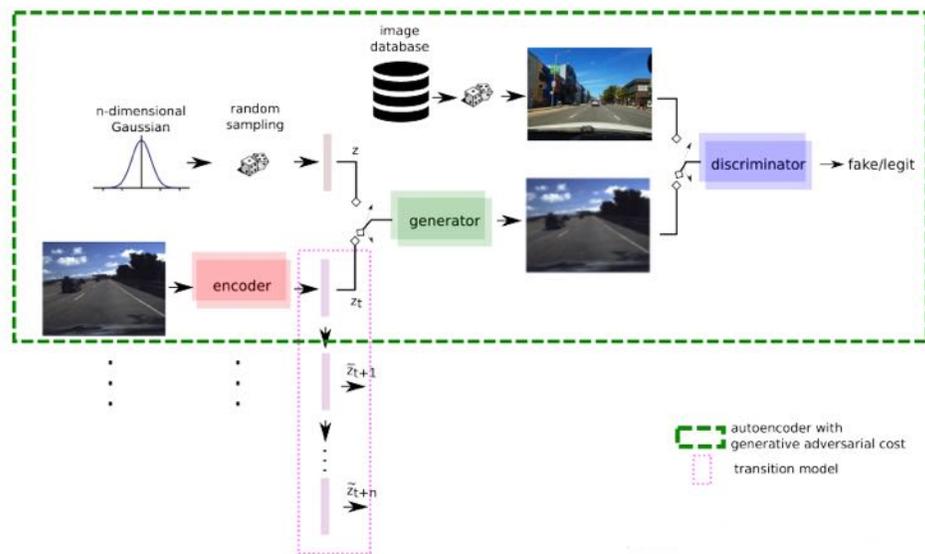
- 无人车也是要懂得交通规则的，所以识别交通标志并根据标志的指示执行不同指令也非常重要。
- 这也是个计算机视觉问题，可以用深度学习（卷积神经网络）的方法来完成。

车道识别



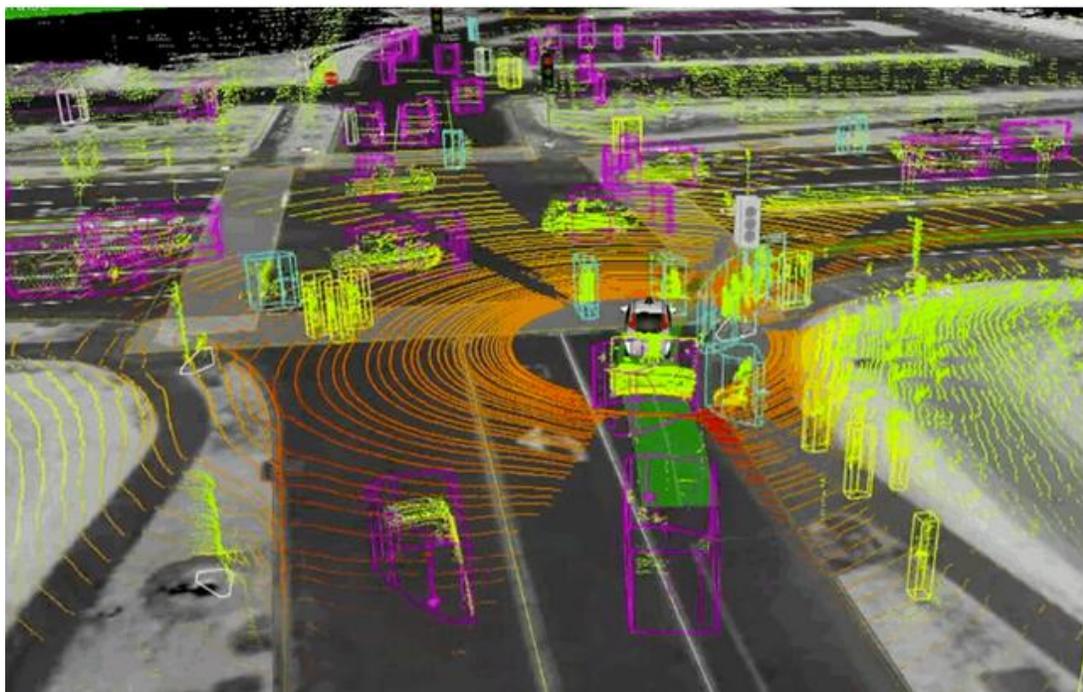
- 车道识别也是计算机视觉问题，高级的道路线检测需要计算相机校准矩阵和失真系数，对原始图像的失真进行校正。
- 这个过程也要使用神经网络的图像处理方法。

典型方案



- Comma.ai提出的端到端方法：
- 以摄像头的原始图像作为输入，
- 利用Autoencoder编码，
- 然后用一个循环神经网络（RNN）来从人类驾驶数据中进行转换学习
- 来拟合逼近最优驾驶策略。

未来趋势



- 就自动驾驶来说，诸如：
 - 车载深度学习芯片开发、
 - 传感器的融合/替代方案、
 - 高精度地图的制作、
 - 决策与控制系统的研发、
 - 安全保障技术等
- 会是未来的研究热点。

未来趋势

- 相信在可预见的将来，无人驾驶会把人类从低效、重复的驾驶中解放出来。

