

中国计算机视觉行业研究报告

2017年



海量行研报告免费读

摘要



深度学习主要提升的是计算机视觉领域分类任务的准确率；开源环境仅降低计算机视觉领域的入门技术门槛，前沿算法的技术壁垒依然存在；计算机视觉比赛成绩、论文成果不直接代表技术团队解决实际业务问题的能力。



2017年中国计算机视觉规模预期为40亿，凭借安防领域的爆发性增长，预期2020年将增长至725亿。



前端嵌入式智能系统的渗透率将逐步提升，与后端协同智能计算，加速产业智能升级。



算法迭代将不断提升限定场景识别准确率，加速渗透为各行业应用赋能。



前沿算法之外，计算机视觉公司的商业壁垒有赖于产品、服务、市场等综合建设。

计算机视觉技术概述

1

计算机视觉行业概况

2

计算机视觉的应用场景

3

计算机视觉典型公司案例

4

计算机视觉行业发展趋势

5

计算机视觉的概念与行业主要应用

计算机视觉横跨感知与认知智能，现阶段应用以感知为主

视觉使人类得以感知和理解周边的世界，人的大脑皮层大约有70%的活动在处理视觉相关信息。计算机视觉即通过电子化的方式来感知和理解影像，以达到甚至超越人类视觉智能的效果。从1966年学科建立（MIT：The Summer Vision Project）至今，尽管计算机视觉在感知与认知智能方向仍有大量难以解决、尚待探索的问题，但得益于深度学习算法的成熟应用（2012年，采用深度学习架构的AlexNet模型，以超越第二名10个百分点的成绩在ImageNet竞赛中夺冠），侧重于感知智能的图像分类技术在工业界逐步实现商用价值，助力金融、安防、互联网、手机、医疗、工业等领域智能升级。2017年下半年，数家计算机视觉公司单笔融资上亿美元，再次将计算机视觉推向人工智能领域最受关注的方向之一。本报告聚焦于计算机视觉技术现阶段在工业界的应用与研发，将在对相关技术热点及产业整体做概述性介绍的基础上，对典型应用的具体使用场景分领域论述，希望理清现状，写明征途，供产业界、投资界、政策制定者及关注人工智能领域的社会各界以参考。因报告研究对象以技术层创业公司为主，偏颇遗漏之处，敬请指正。

视觉感知小有所成，视觉认知道阻且长



计算机视觉的典型应用



安防领域
疑犯追踪 视频结构化



金融及互联网领域
刷脸认证



手机及娱乐领域
影像分类 影像处理 AR特效



零售领域
商品识别



广告营销领域
自动化挖掘影像内容广告位



工业领域
产品质量 3D分拣



医疗领域
医疗影像分析



自动驾驶领域
环境感知 高精地图 定位



无人机/机器人领域
环境感知 定位 自动避障

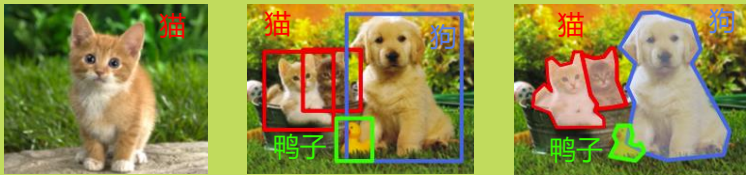
来源：艾瑞根据专家访谈、公开资料等研究绘制。

计算机视觉的学术研究任务

现阶段有较好商业化进展的主要为语义感知中的分类任务

与人类实时选择性处理视觉信息不同（如人在驾驶时不需在意公路边草地的纹理或形状，也不用知道每辆车的确切形状），计算机仍难以从实际需求出发自主选择性输入并计算影像信息，通常需要人类对具体任务进行分解并使用与之匹配的计算方法，建立完整理想的智能视觉系统仍有很大挑战。另外，与可结合常识做猜想和推理进而辅助识别的人类智能系统相比，现阶段的视觉技术往往仅能利用影像表层信息，缺乏常识以及对事物功能、因果、动机等深层信息的认知把握。

计算机视觉的学术研究任务（部分）

语义感（认）知 建立影像信息与语言描述的映射			去模糊、超分辨率等技术 可有效增强影像信息
		图像分类 物体检测 语义分割	
		分类：图像分类、物体检测、语义分割等，分别为图像整体、局部以及像素级的分类任务	
		对象关系识别：空间关系、动作关系、介词方位关系等	
3D目标感（认）知与重建		行为识别及预测、追踪	
		看图写话：场景理解，语句、段落等生成	
		3D目标分类及对象关系、姿势行为分析、相关场景理解	
定位与控制		3D重建：从影像信息中重建这个世界的几何模型，一般从二维影像出发得到场景的三维表示，可利用运动视差、双目立体视觉、多视图、纹理、明暗、轮廓、物体及场景几何结构等线索	
		基于影像信息（往往也结合惯性运动单元 IMU等其他传感器获取多维度信息） 机器人（包括自动驾驶汽车）在环境中对自身进行定位，进而导航	
		指导机器人操纵物体，拾起、抓住、转动等，自动驾驶领域则涉及横向、纵向及障碍物避让等控制 是一个需要不停观察、规划的连续决策问题	

来源：参考资料 Stuart Russell.人工智能：一种现代的方法[M]等。

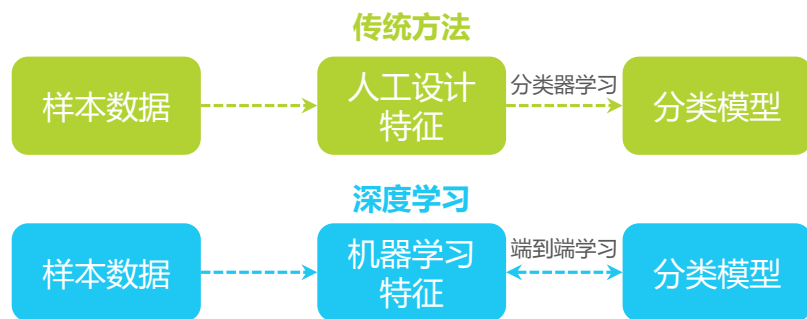
核心算法之深度学习

深度学习的概念及成就

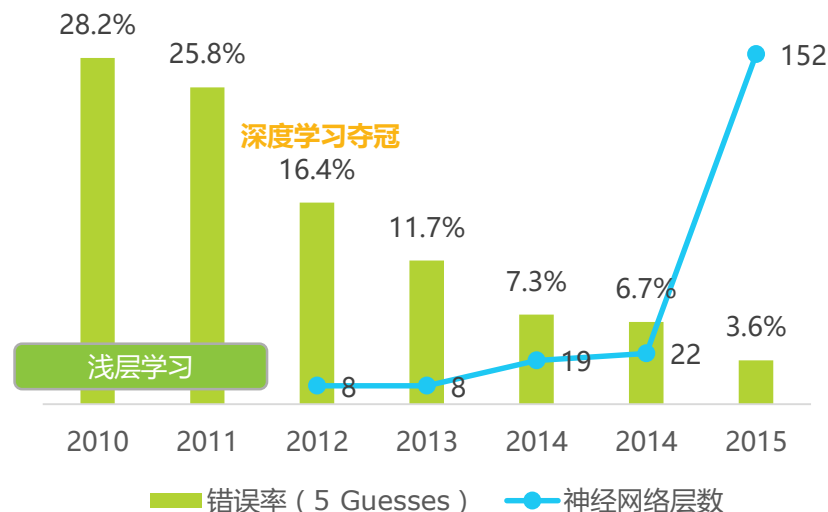
深度学习算法由多重非线性变换构成多个处理层，辅以复杂结构设计和各种梯度技术，通过对大量样本的输入与对应输出数据的抽象计算，拟合出一个可处理新输入信息的函数模型，解决其分类或预测问题。尽管神经网络（可理解为深度学习算法的前身或别称）的研究历史其实比计算机视觉的研究历史还要长，但在之前的研究过程中，深度神经网络的方法并未得到有效验证。直到2011年，语音识别领域凭借深度学习取得重大突破；2012年，AlexNet，一个8层的神经网络，以超越第二名10个百分点的成绩在ImageNet竞赛中夺冠（图像分类的Top 5错误率为16.4%），深度学习终迎来包括学术探索与工业应用中的热潮。不断提升的层数逐步提升计算机分类视觉的准确率，2015年夺冠的 ResNet 深达152层，以3.57%的错误率超越人类视觉的5.1%。

深度学习与传统方法的区别

传统方法，针对不同类别的物体，一般首先由研究员充分发挥聪明才智，手工定义不同的特征，然后利用不同的机器学习算法（分类器学习），这时的算法一般仅有两到三次非线性变换，学到的参数较少（浅层学习）；深度学习则通过机器学习自身来产生特征，因此特征和分类器学习不再有区分。如今深度学习的算法已可达上千层。



ImageNet Large Scale Visual Recognition 2010-2015



来源：参考资料 周志华.机器学习[M]等。

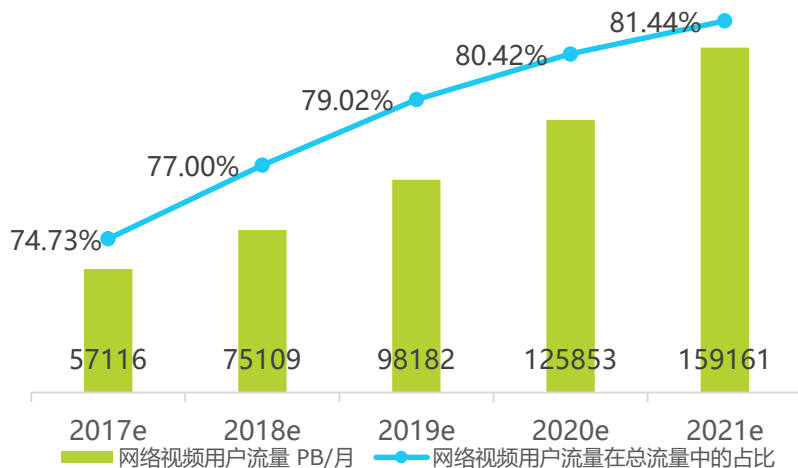
来源：参考资料 CS231n.Convolutional Neural Networks for Visual Recognition等。

核心算法之深度学习

数据与算力是深度学习的重要支撑

日益丰富的影像内容为深度学习算法提供了大量的数据支撑。据思科公司评估，2021年单月上传至全球网络的视频总时长将超过500万年，每秒将诞生1百万分钟的网络视频内容，网络视频流量将占据全球所有网络用户流量的81.44%。需要说明的是，现在的学习多为有监督学习（需要对数据进行充分标注），而且并非所有类型的影像数据都易得易标注（比如医疗影像数据需由专业医师标注病灶），业界领先的视觉公司一般会有数百人的标注团队（多为外包，但需专业培训，实时指导）。另一方面，深度学习学习过程中的“训练”与应用部署后的“推断”均涉及大量并行计算，传统CPU算力不足，而GPU、FPGA（现场可编程门阵列）、ASIC（TPU、NPU等AI专属架构芯片）等具有良好并行计算能力的芯片可提供数十倍乃至上百倍于CPU的性能，与云服务一起，大幅缩短计算过程（在过去，往往数周甚至数月才能跑出一组结果，然后调整模型架构，效率极低），易于短期调整多种模型架构，显著提升分类模型的进步速度。2010年以后，CPU内部晶体管数量的增长明显放缓，传统摩尔定律失效，而GPU类处理器依然保持着快速增长的势头（2016年GPU的计算力为10个TFLOP/S，2017年达到了120个TFLOP/S，TPU则实现了惊人的180个TFLOP/S），验证着AI时代的摩尔定律。

2017-2021年全球网络视频用户流量



并行计算能力有效提升模型计算效率



GPU、FPGA、ASIC等并行运算处理器
在云端服务器、用户终端等
助推模型训练与推断

来源：参考资料 Cisco Visual Networking Index: Forecast and Methodology 等。

来源：艾瑞咨询研究院自主研究绘制。

热门技术之人脸识别

人脸识别是当下视觉领域热门应用的重要技术支撑

人脸识别可看做语义感知任务中针对人脸影像的分类问题，也是当下视觉领域热门应用的重要技术支撑，各个环节都因深度学习算法的推进实现了更优的计算结果。泛金融领域的远程身份认证、手机领域的刷脸解锁一般属于人脸验证（技术相对成熟）；安防影像分析一般为人脸识别，刑侦破案对亿级甚至十亿级比对有刚性需求，目前技术仍有很大进步空间，更多新功能、新场景的解锁依赖于最先进的算法团队和相关业务领域开拓者的共同努力。

人脸识别的技术流程



人脸检测

检测出图像中人脸所在位置



人脸配准

定位出人脸的五官关键点

人脸提特征

将人脸图像转化为人脸特征（固定长度的数值串）

人脸比对

衡量人脸间的相似度



人脸属性识别

识别出人脸的
性别、年龄、种族、表情等

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_21354

