

如何打造一流的视觉AI技术

邓亚峰 格灵深瞳

计算机视觉是人工智能领域最重要的方向之一

人工智能应用领域

自动驾驶

智慧城市

智慧医疗

机器人

增强现实

人工智能技术领域

计算机
视觉

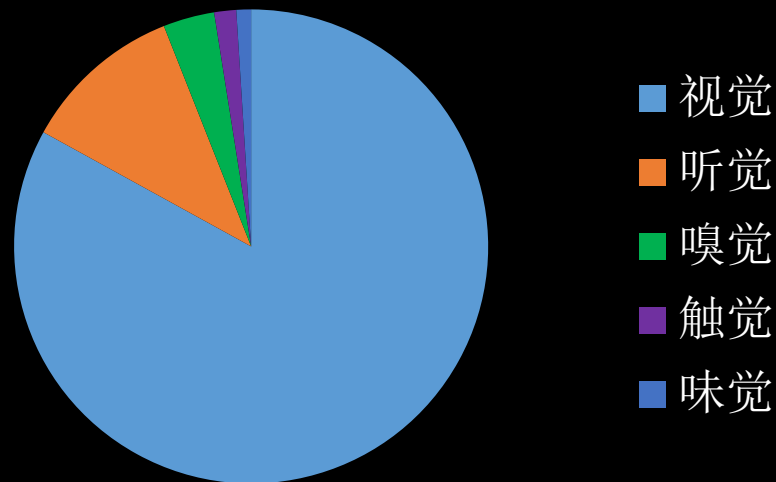
语音识别

自然语言
理解

核心算法：机器学习

深度学习

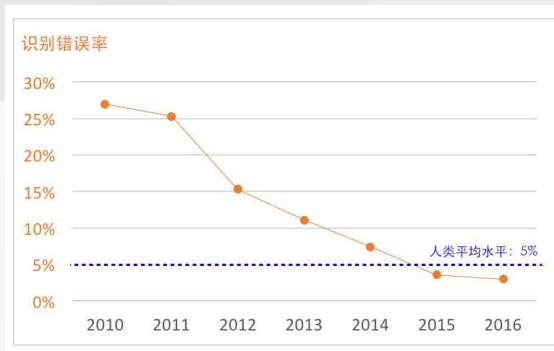
各感觉器官接受信息比例



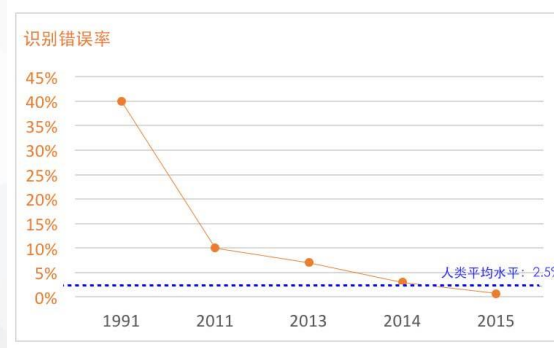
深度学习引发了计算机视觉技术变革



使得用更大量数据训练更深的网络
成为可能



图像识别错误率下降（以ImageNet为例）



人脸识别错误率下降（以LFW数据集为例）

在某些限定场景下，图像识别算法超越了
人类的能力

智能时代趋势 - 计算机视觉应用爆炸性增长



移动互联网



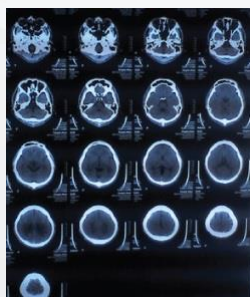
智慧城市



机器人



自动驾驶



智慧医疗



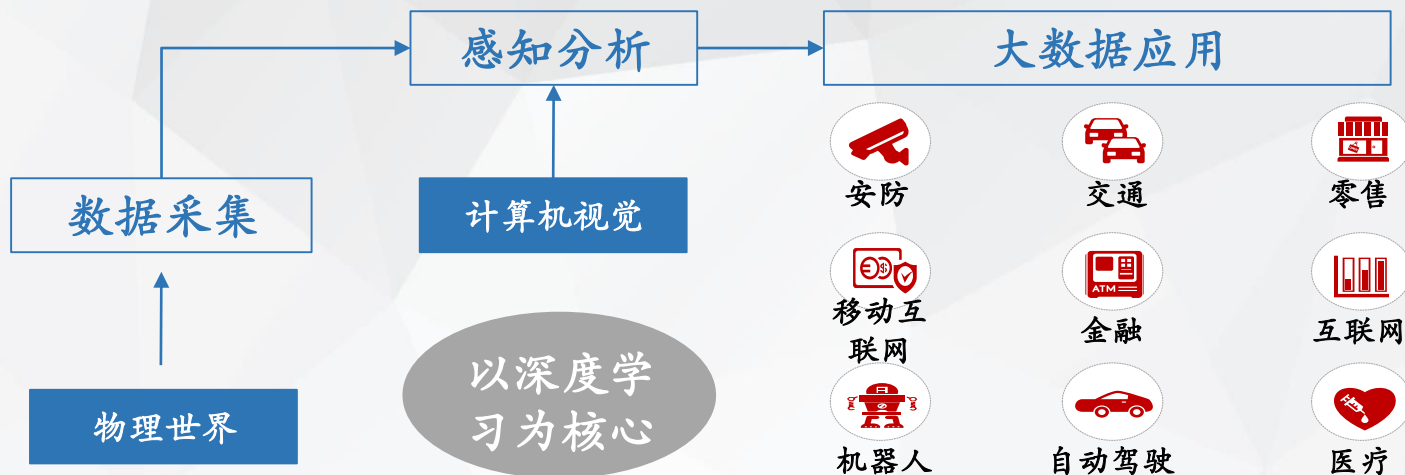
工业视觉



增强现实 (AR)

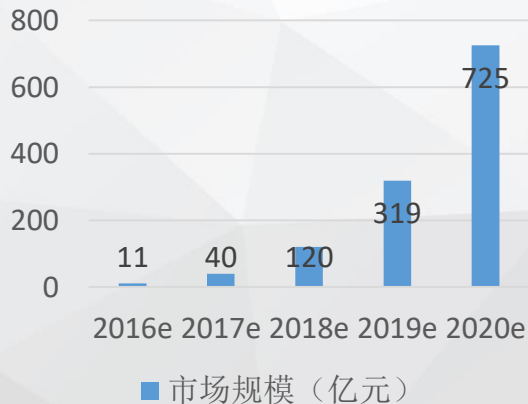
计算机视觉是智慧物联网时代最大的数据入口，拥有广阔的应用场景

计算机视觉
技术
推动视频大
数据广泛应
用



视频大数据
行业前景广
阔

中国计算机视觉市场规模



中国大数据应用市场规模

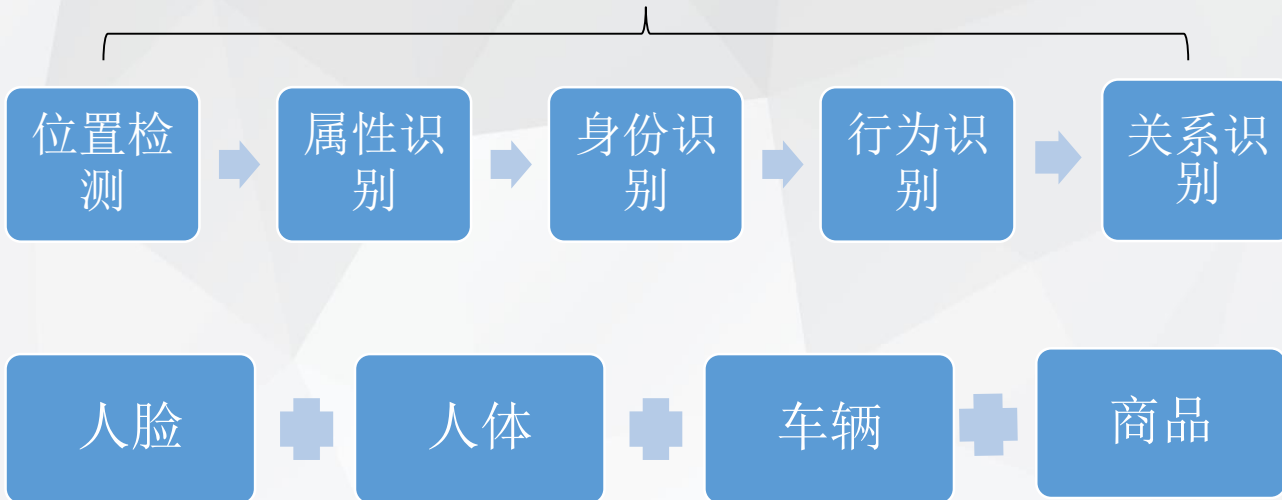


广义上，让计算机看懂世界；

狭义上，通过丰富的视觉传感器，理解物理世界中每个物体的位置、属性、身份以及行为等信息；

计算机视觉技术愿景

感知



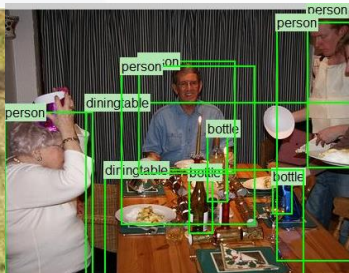
功能更加
丰富

目标类别
更全

理解粒度
更细



分类



检测



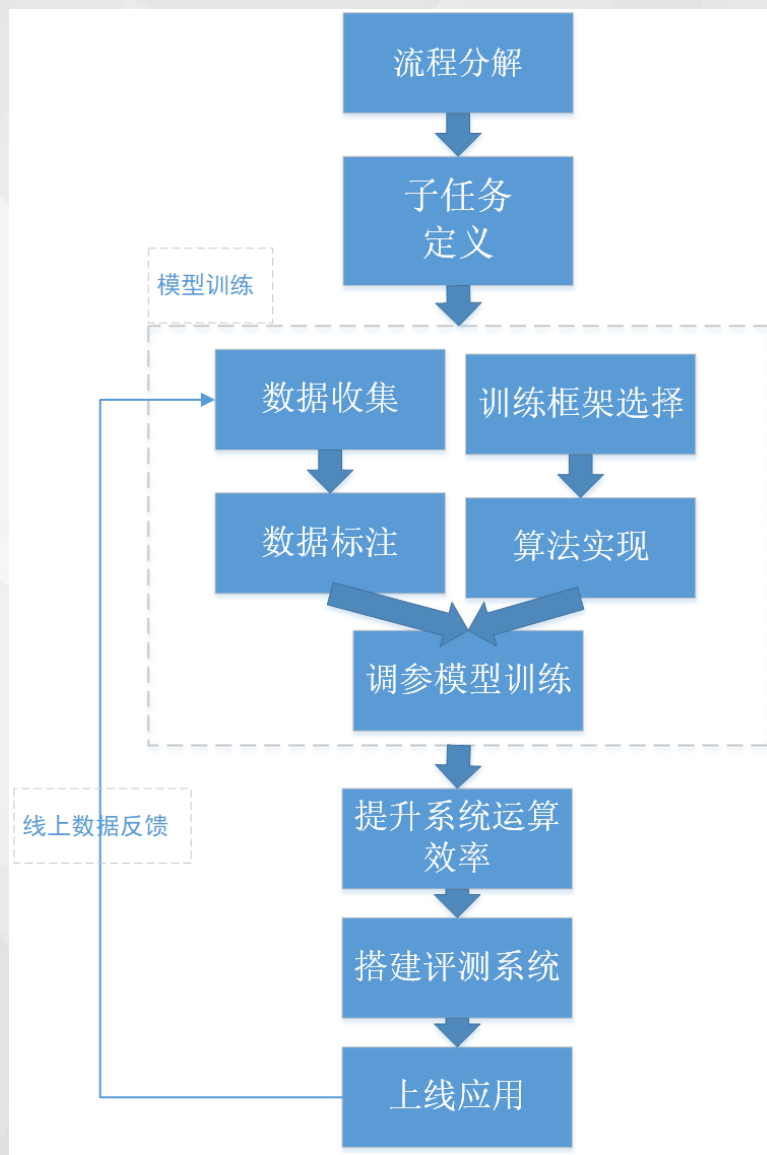
关键点检测



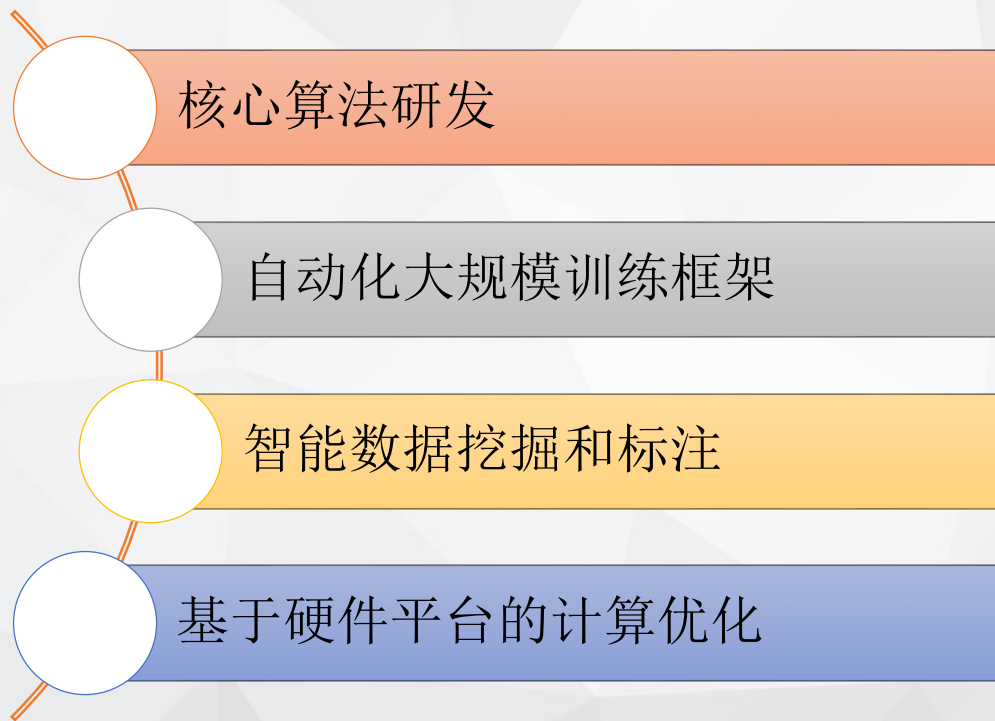
语义分割

精度更高、速度更快、成本更低、规模更大

构建计算机视觉系统的基本流程



构建视觉计算系统的关键因素



算法需要极致打磨

对于每一个算法，都需要尝试多种改进方向：

- 数据以及处理方式
- 不同参数
- 模型结构
- Loss函数
- 模型蒸馏
- 系统流程划分&关联其它任务

人脸识别算法Loss函数演进（举例）

Image1-> CNN->特征1
Image2-> CNN->特征2 } 是否是一个人

$$\text{Verif}(f_i, f_j, y_{ij}, \theta_{ve}) = \begin{cases} \frac{1}{2} \|f_i - f_j\|_2^2 & \text{if } y_{ij} = 1 \\ \frac{1}{2} \max(0, m - \|f_i - f_j\|_2)^2 & \text{if } y_{ij} = -1 \end{cases},$$

Pair-wise Loss(Siamese network)

$$\mathcal{L}_{\text{softmax}} = -\frac{1}{N} \sum_{i=1}^N \log \left(\frac{e^{\mathbf{w}_{y_i}^T \mathbf{f}_i + \mathbf{b}_{y_i}}}{\sum_{j=1}^K e^{\mathbf{w}_j^T \mathbf{f}_i + \mathbf{b}_j}} \right).$$

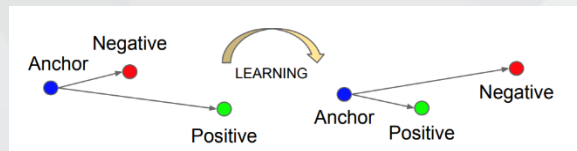
Softmax

$$\begin{aligned} \mathcal{L}_S &= -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{s \tilde{\mathbf{w}}_{y_i}^T \tilde{\mathbf{f}}_i}}{\sum_{j=1}^n e^{s \tilde{\mathbf{w}}_j^T \tilde{\mathbf{f}}_i}} \\ &= -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{-\frac{s}{2} \|\tilde{\mathbf{f}}_i - \tilde{\mathbf{w}}_{y_i}\|_2^2}}{\sum_{j=1}^n e^{-\frac{s}{2} \|\tilde{\mathbf{f}}_i - \tilde{\mathbf{w}}_j\|_2^2}}, \end{aligned}$$

$$\mathcal{L}_T = \max(0, m + \|\tilde{\mathbf{f}}_i - \tilde{\mathbf{f}}_j\|_2^2 - \|\tilde{\mathbf{f}}_i - \tilde{\mathbf{f}}_k\|_2^2), \quad c_i = c_j, c_i \neq c_k,$$

限制条件 $\|\mathbf{f}\| = \text{const}$ 、 $\|\mathbf{W}\| = \text{const}$ 和 $b=0$

NormFace/CocoLoss



The loss that is being minimized is then $L =$

$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+ . \quad (3)$$

Triplet Loss

$$\mathcal{L} = \mathcal{L}_S + \lambda \mathcal{L}_C$$

$$= -\sum_{i=1}^m \log \frac{e^{W_{y_i}^T \mathbf{x}_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T \mathbf{x}_i + b_j}} + \frac{\lambda}{2} \sum_{i=1}^m \|\mathbf{x}_i - \mathbf{c}_{y_i}\|_2^2$$

Center Loss

$$L_7 = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}}, \quad (9)$$

subject to

$$W_j = \frac{W_j}{\|W_j\|}, \mathbf{x}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}, \cos \theta_j = W_j^T \mathbf{x}_i. \quad (10)$$

InsightFace/ArcFace

人脸识别系统影响因素（举例）

- 训练数据：数据量、纯度、数据分布
- 数据增强：克服分辨率、光照、模糊等因素
- 检测/Alignment精度提升
- 人脸识别Loss函数设计
- 超大规模模型训练框架
- 模型加速(Model Distilling、量化等)
- 部件模型融合
- 视频人脸跟踪
- 人脸质量评估
- 视频人脸多帧融合
- 利用数据Fine-Tune来充分利用数据提升场景适应性
-

模型维护复杂度

所有模型 \times N

- 服务器平台模型
- 嵌入式平台模型
- ...

如何在算法人员有限、数据资源有限的情况下，打造一流的核心算法？

传统算法开发模式

传统算法工程师工作时间分配：

60%在处理数据

20%在调参

15%在实现算法

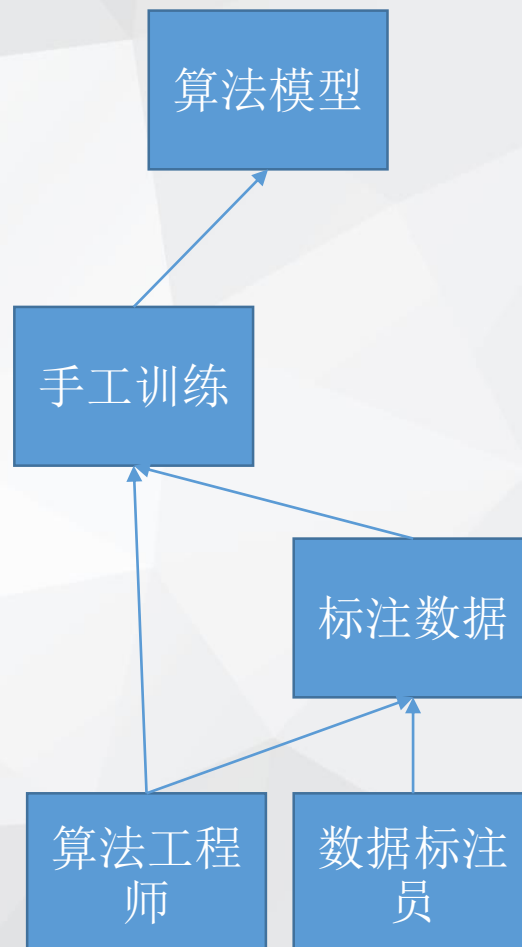
5%在思考算法

缺点：

数据标注成本高

数据依赖于算法工程师，无法持续

GPU利用率低



改进算法开发模式

新型算法工程师工作时间分配：

20%在沟通确定数据处理流程

10%在调参

50%在实现算法

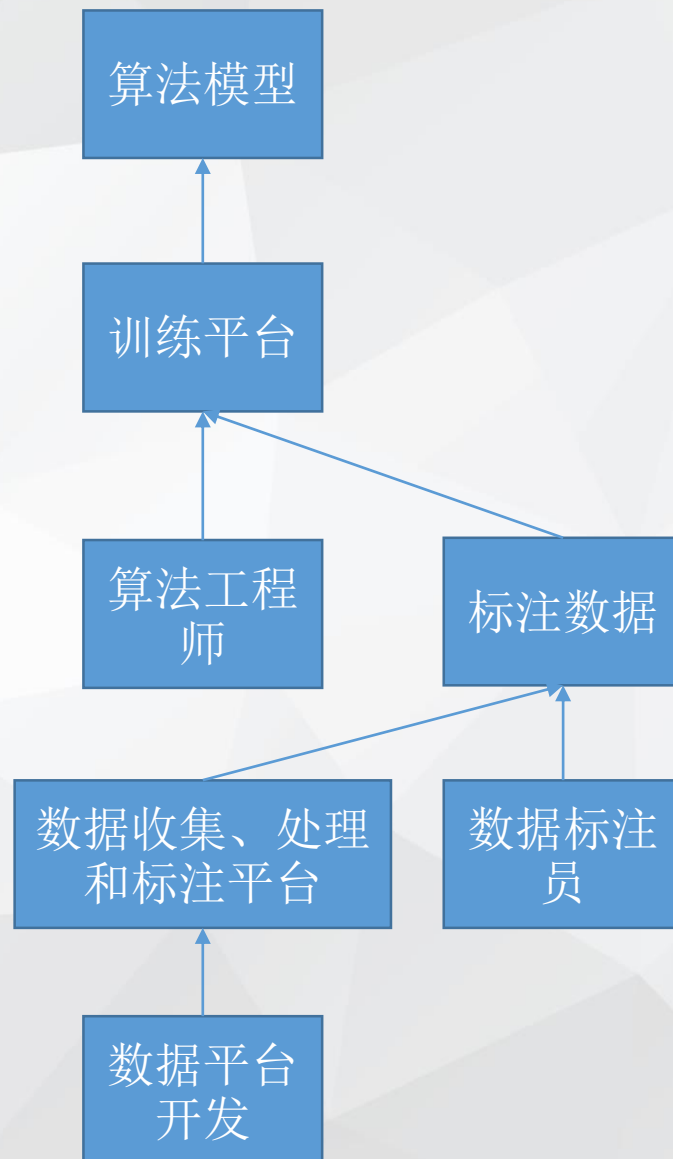
20%在思考算法

优点：

GPU利用率提升

数据费用下降

数据量持续增加



打造高效的自动化训练管理平台

目标:

提升算法人员产出效率, 节省人力

提高训练资源的利用率

训练任务管理:

训练集管理

代码/参数管理

环境管理

计算资源管理

模型评测和选择

训练方式:

自动化模式

虚拟机模式

云主机模式



拥有大规模集群训练能力

已有开源框架：Caffe、MxNet、TensorFlow、PyTorch、PaddlePaddle等

- 完全重复造轮子并不必要的：

开源框架有行业里面最好的工程师的贡献；

对员工而言，技能更加熟练，容易上手和适应；

可以利用更多的开源社区的开源算法，更容易和业界交流；

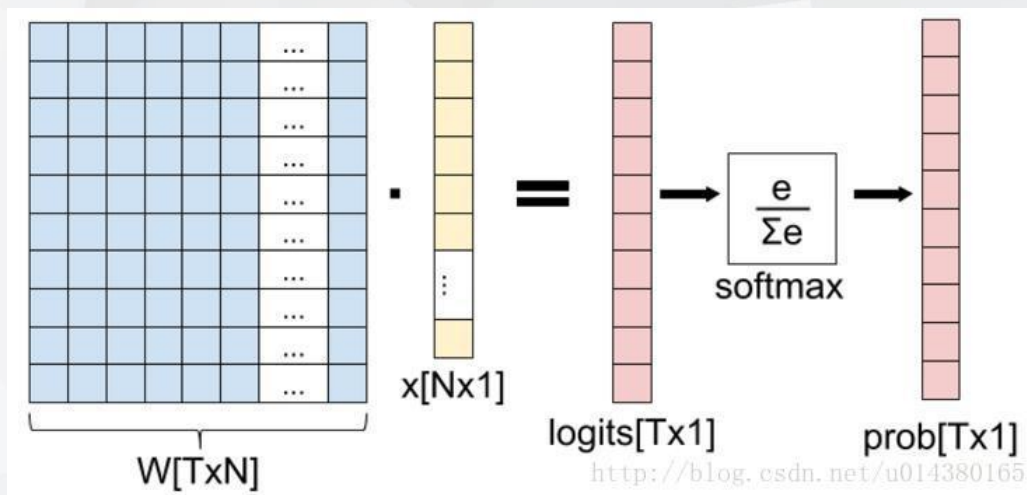
- 但要有基于开源框架构建大规模训练框架的能力：

现有框架并行度做得不好

超大规模问题需要自己解决

打造大规模训练集群

问题：如何解决数亿类别的分类任务？



网络结构

Loss函数:

$$L_i = -\log(p_{y_i})$$

$$p_k = \frac{e^{f_k}}{\sum_j e^{f_j}}$$

$$\frac{\partial L_i}{\partial f_k} = p_k - \mathbb{1}(y_i = k)$$

当类别数是1亿时，特征长度是128维时，参数矩阵参数数据量约为51G

打造大规模训练集群 - 数据并行+模型并行

步骤一

N计算节点	每节点M卡	Batch数据并行	数据并行特征计算
Machine_1	GPU_1	batch_1_1	$X_{1_1} = \text{CNN}(\text{batch}_{1_1})$
	GPU_2	batch_1_2	$X_{1_2} = \text{CNN}(\text{batch}_{1_2})$

	GPU_M	batch_1_M	$X_{1_M} = \text{CNN}(\text{batch}_{1_M})$
Machine_2	GPU_1	batch_2_1	$X_{2_1} = \text{CNN}(\text{batch}_{2_1})$
	GPU_2	batch_2_2	$X_{2_2} = \text{CNN}(\text{batch}_{2_2})$

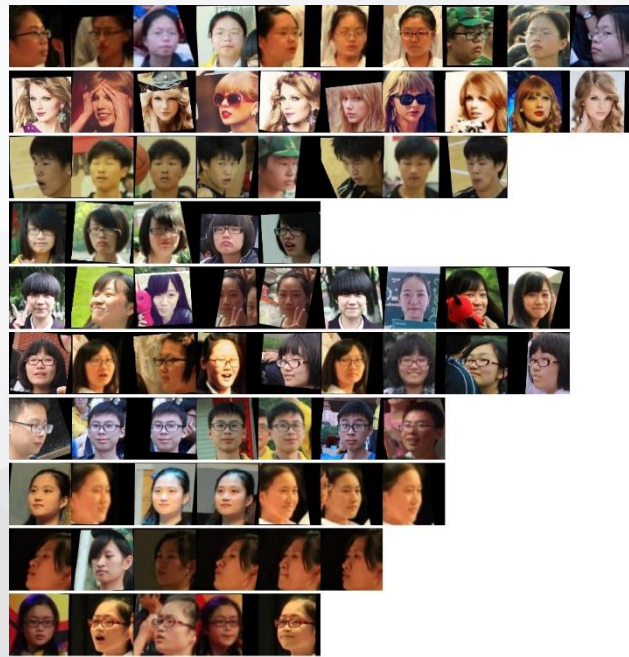
	GPU_M	batch_2_M	$X_{2_M} = \text{CNN}(\text{batch}_{2_M})$
...

步骤二

N计算节点	特征汇聚	模型并行	模型并行FC计算	通信优化(FW+BP)	梯度汇聚
Machine_1	X	W_1_1	$X * W_{1_1}$	grad_X_1_1, grad_W_1_1	grad_X
		W_1_2	$X * W_{1_2}$	grad_X_1_2, grad_W_1_2	
		
		W_1_M	$X * W_{1_M}$	grad_X_1_M, grad_W_1_M	
Machine_2		W_2_1	$X * W_{2_1}$	grad_X_2_1, grad_W_2_1	
		W_2_2	$X * W_{2_2}$	grad_X_2_2, grad_W_2_2	
		
		W_2_M	$X * W_{2_M}$	grad_X_2_M, grad_W_2_M	
...

通过数据平台持续高效获取海量训练数据

- 如何便捷、低成本的获取数据
 - 数据源
 - 选择标注工作量小的有先验约束的数据源
 - 例如：
 - 相册数据
 - 人脸1：1人证数据
 - 处理流程
 - 全部人工标注吗？
 - 隐私问题
 - 成本问题
 - 人工标注的准确性



相册数据源

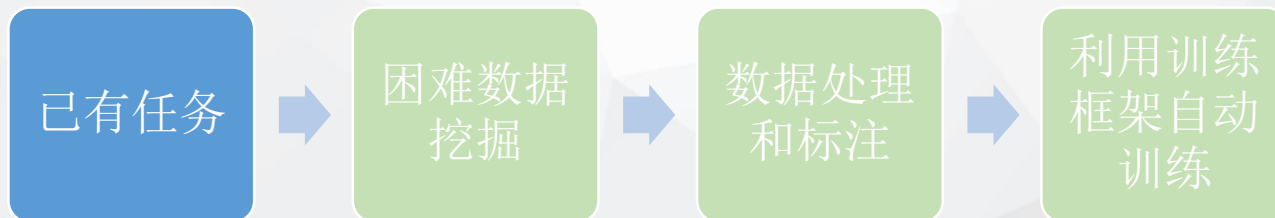
人脸自动数据标注流程



数据团队的工作模式

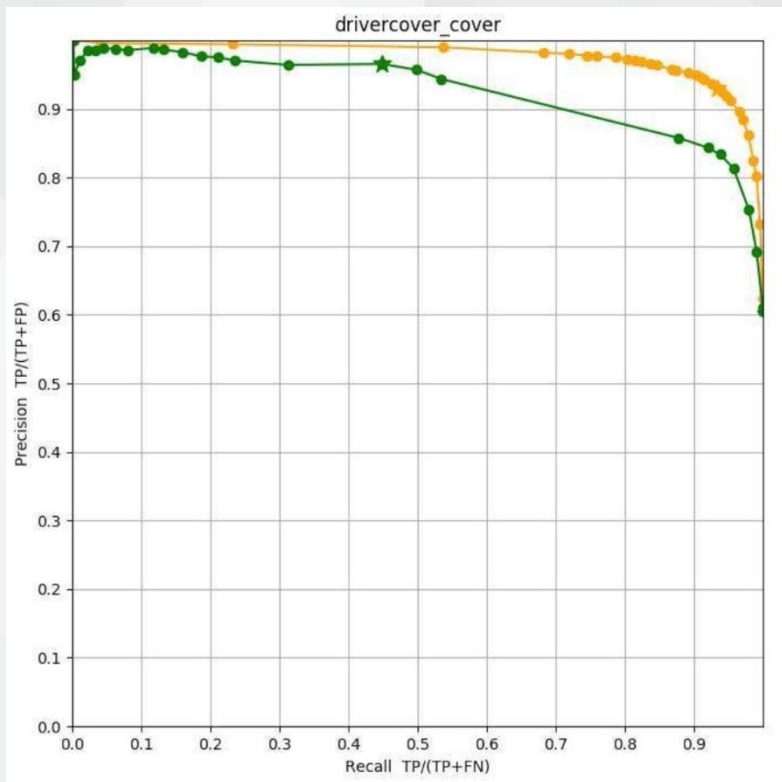


传统工作模式

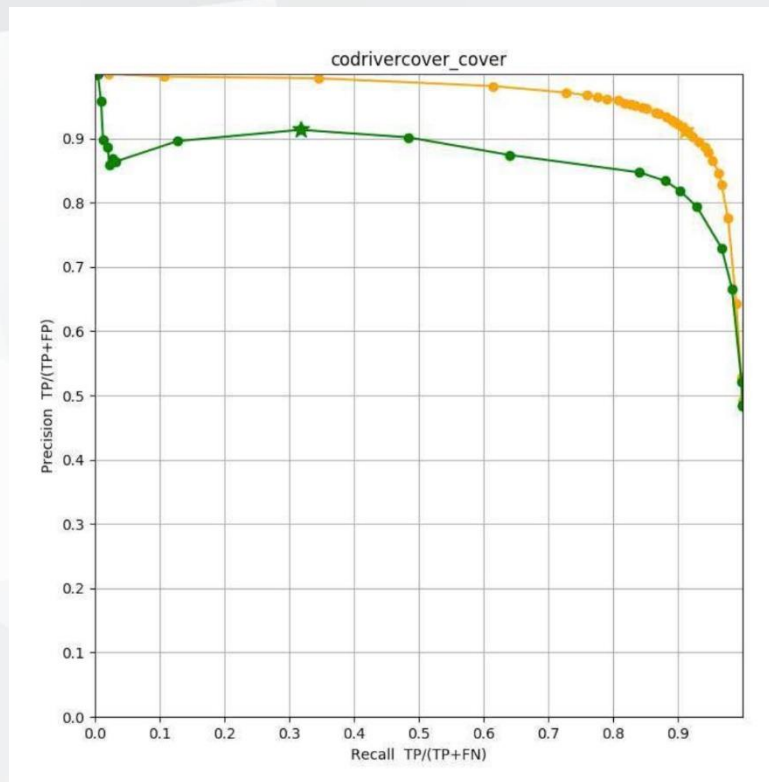


数据驱动工作模式

数据驱动模型效果提升

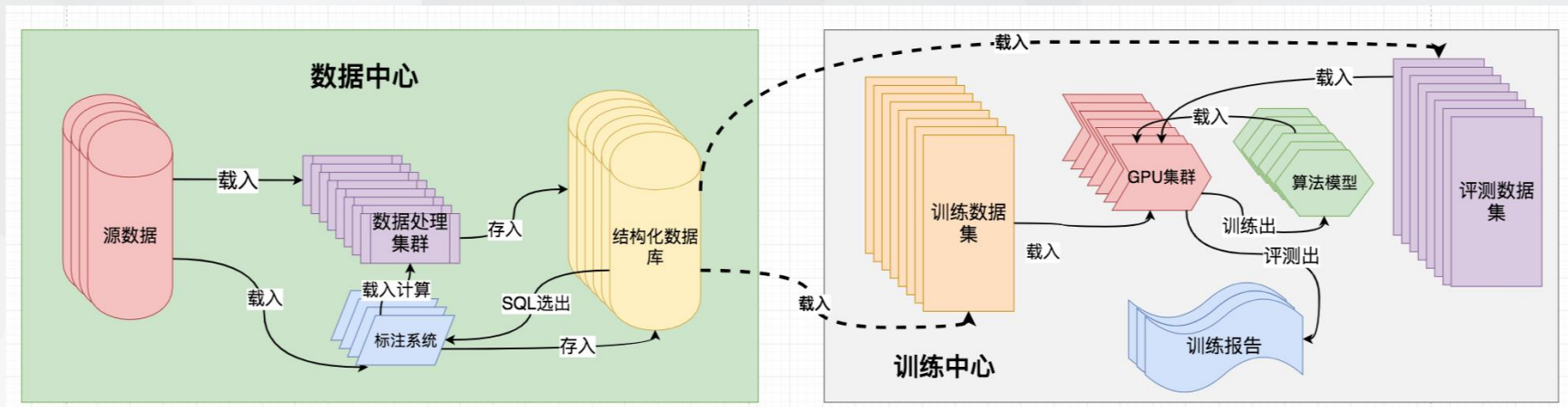


驾驶员面部遮挡



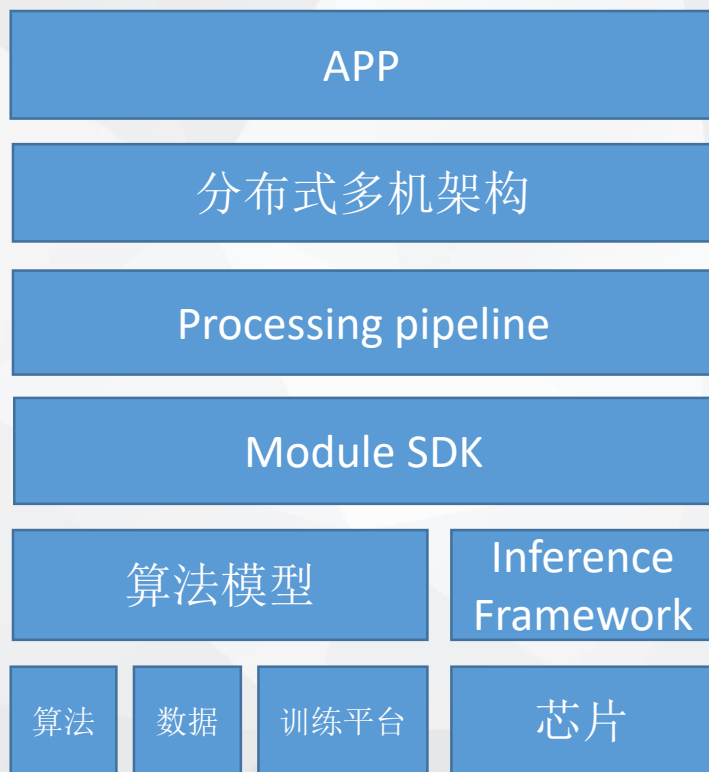
副驾人员面部遮挡

深瞳大脑：构建 数据->标注->算法->训练->产品->数据 正循环



- 数据总量：数十亿人脸、数千万人体、数亿车辆
- 训练数据总量：来自数千万人的数亿人脸、数百万人体、数千万车辆
- 训练框架可支持上百卡的GPU集群训练
- 训练架构可支持上亿类别任务训练
- 训练架构可支持多达上百层、数十亿参数的深度学习模型

视觉计算系统预测架构



提升视觉计算系统运算效率的方式

• 硬件平台选择

- GPU（1080TI、P4、2080TI、TX1、TX2等）
- Arm（海思平台、RockChip平台等）
- FPGA
- DSP（Movidius、Ceva等）
- ASIC（通用：海思3559A等；专用：？）

• 基于网络结构改变的加速策略

- 简单修改成熟网络（减少filter channel number、减少层数等）
- 手工设计新网络结构（Depth-Wise、ShuffleNet）
- 模型参数裁剪（模型参数压缩；如何使小网络更好收敛：Mimick）
- NAS（例如：基于强化学习的模型结构搜索）

• 软件优化

- 模型定点化
- 优先将模型用TensorRT运行
- GPU平台其它运算用CUDA实现，少使用CPU
- ARM平台，利用NEON指令
- 合理组Pipeline提高系统资源利用率（系统级优化）

深度神经网络定点化

◆ 权重定点化;

将训练好的模型，利用具有代表性的足量标定样本，用 KL-divergence 度量每层输出的 float32 和 int8 结果分布的相似度，然后确定参数和数据输入的 scale，将各层参数进行定点化;

◆ 预测时利用定点运算来处理;

◆ 基于 NCNN 预测框架，用 NEON 指令进行实现;

◆ 优点:

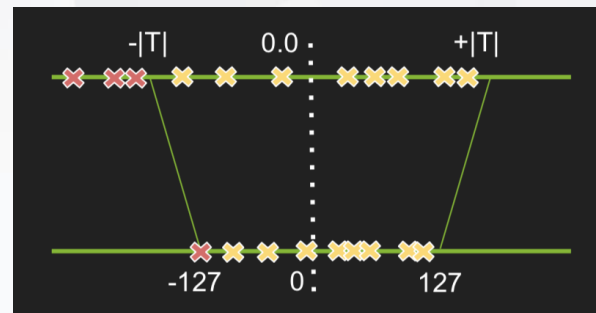
- **效率高:** ARM NEON 汇编优化，计算速度极快; 支持多核并行计算加速，ARM big.LITTLE CPU 调度优化;
- **精简:** 精细的内存管理和数据结构设计，内存占用极低; 整体库体积小于 500K，并可轻松精简到小于 300K;
- **适应性广:** 支持卷积神经网络，支持多输入和多分支结构，可计算部分分支; 可扩展的模型设计，支持 8bit 量化和半精度浮点存储，可导入 caffe/pytorch/mxnet/onnx 模型;

从 float32 到 int8 的量化:

$$\text{weight} * x + \text{bias}$$

$$q_weight = \text{scale}_w * \text{weight}$$

$$q_x = \text{scale}_x * x$$

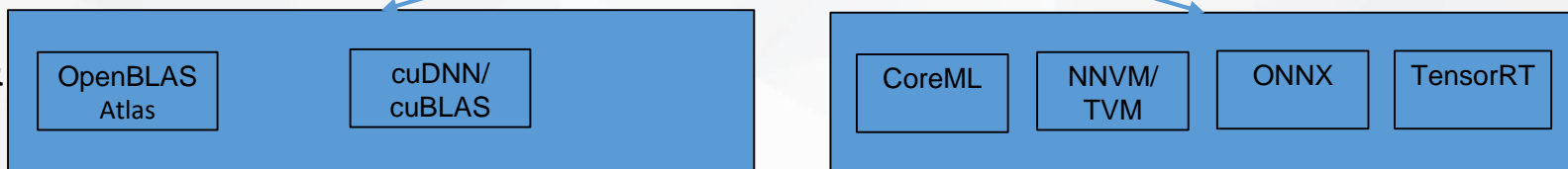


预测框架

预测框架



神经网络&矩阵运算



系统指令集层



人工智能芯片的蓬勃发展



NVIDIA



GOOGLE TPU2



地平线



深度学习芯片的核心竞争力



芯片的竞争不仅是IC Design、工艺、优化算法等技术的竞争，更是产业链生态的竞争

公司简介

- 公司于2013年4月，由创始人赵勇在北京成立
- 现有团队全职员工三百名，主要来自人工智能，智能硬件，安防监控，互联网等领域



赵 勇 创始人&CEO

国家千人特聘专家

前Google X Lab资深图像识别科学家 Google Glass发明者之一

专攻计算机视觉与运算影像学十余年

布朗大学 (Brown University) 计算机工程博士

公司愿景

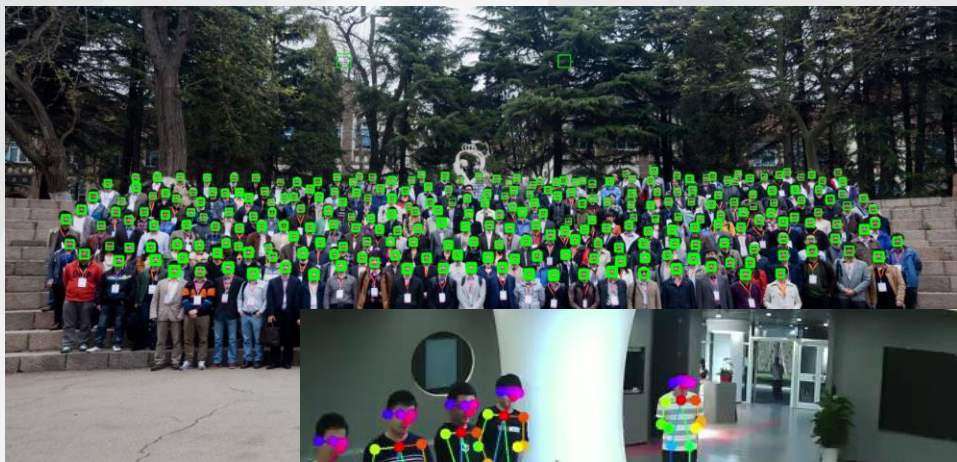
格灵深瞳是一家全国领先的人工智能和智慧物联网（AIoT）科技企业；我们的愿景是让计算机理解世界，让AI使生活更美好。

我们专注于研发包括计算机视觉在内的人工智能核心算法，并把先进的人工智能科技转化为具备低成本、大规模部署能力的产品和服务，并深度结合应用场景，为用户提供高性能、可靠实用的智慧解决方案。

自主研发的深瞳技术在人脸识别、人和车的检测、跟踪、识别以及检索等方面都居于世界领先水平。我们的产品包括从云到端的一系列产品，诸如机器人、智能抓拍相机、行为分析仪、智能感知云等。目前，格灵深瞳在智慧城市、智慧零售、智慧银行等领域，为遍布全国和全世界的客户提供包含智能传感器、智能识别、智能云计算和服务机器人的综合智能解决方案和服务。



人脸/人体/车辆 检测、跟踪、定位



人脸检测



人脸关键点定位

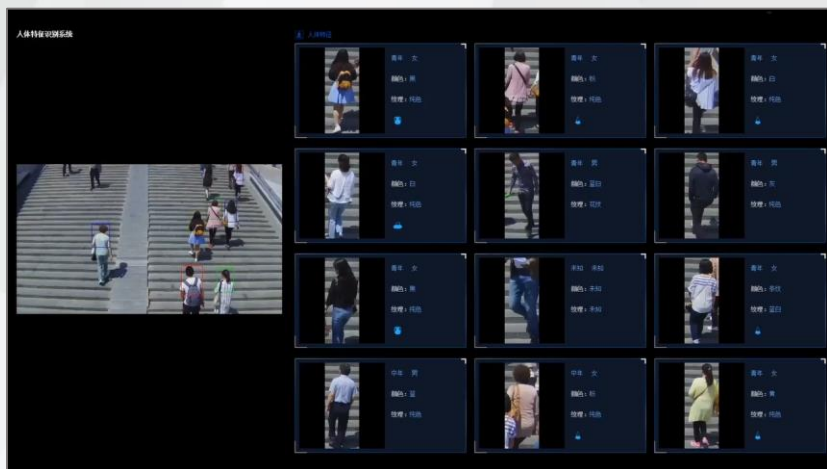


人体检测



车辆检测、跟踪

人脸/人体/非机动车 属性识别



支持近50种人体属性识别

- 年龄、性别
- 头部特征: 眼镜、口罩、帽子
- 上身特征: 颜色、纹理、长短袖
- 下身特征: 颜色、长短裤、裙子
- 其它特征: 手提包、单肩包、双肩包、行李箱、推车、拎其它物品



支持45种属性非机动车属性识别

- 车型: 自行车、二轮车、三轮车
- 车辆特征: 车身颜色、带棚、封闭驾驶舱
- 驾驶员特征:
 - 性别
 - 头部特征: 眼镜、口罩、帽子、头盔
 - 上身特征: 颜色、纹理
 - 其它特征: 单肩包、双肩包、雨衣

您上传的原始图片是:



显示Json

ID	属性	属性值	概率
0	性别	male	0.892027
	是否微笑	normal	0.994582
	美丑程度		80.529
	race	white	0.888218
	age	32.9781	
1	性别	female	0.784077
	是否微笑	normal	0.960947
	美丑程度		61.9601
	race	white	0.813042
	age	38.9629	

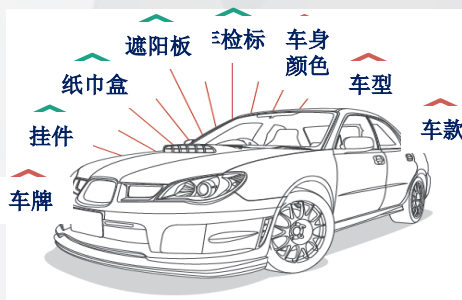
人脸属性识别

车辆属性识别



识别 **10** 大特征

抓拍车辆的车型、车款、车身颜色、车牌、年检标、遮阳板、纸巾盒、挂件、摆件、安全带等细分特征



识别 **19** 种车型

轿车、面包车、皮卡、越野车/SUV、商务车/MPV、轻型客车、中型客车、大型客车、公交车、校车、微型货车、轻型货车、三轮车、二轮车、行人、其他等共计19种



识别 **20000** 多种车款

常见的200多种车辆主品牌，4000多种子品牌及年款



识别 **12** 种车身颜色

白色、橙色、粉色、黑色、红色、黄色、灰色、蓝色、绿色、银色、紫色、棕色等共计12种



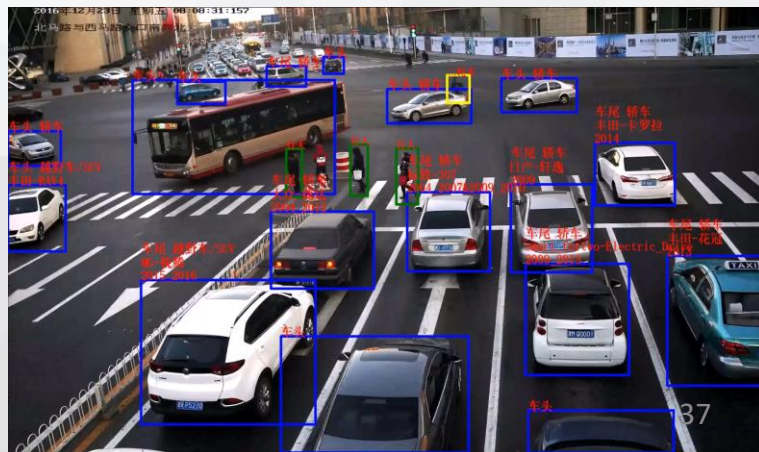
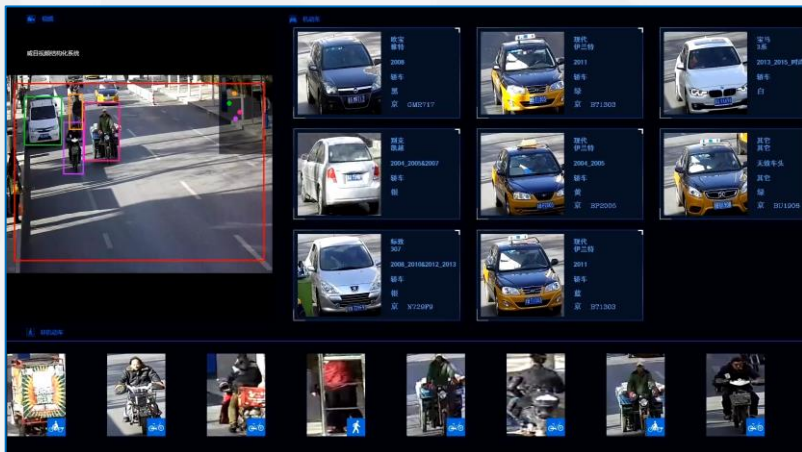
识别 **5** 种号牌

车牌号码，蓝牌、黄牌、警牌、军牌、武警牌、农用车牌；
车牌颜色，包括蓝、黄、白、黑、绿5种；
车牌类型，包括单行、双行2种；



识别 **6** 种标志物

年检标、遮阳板、纸巾盒、挂件、摆件、安全带等细分特征



人/车 身份识别



是谁?



Query (查询图片)

Gallery (库图片)

人脸识别

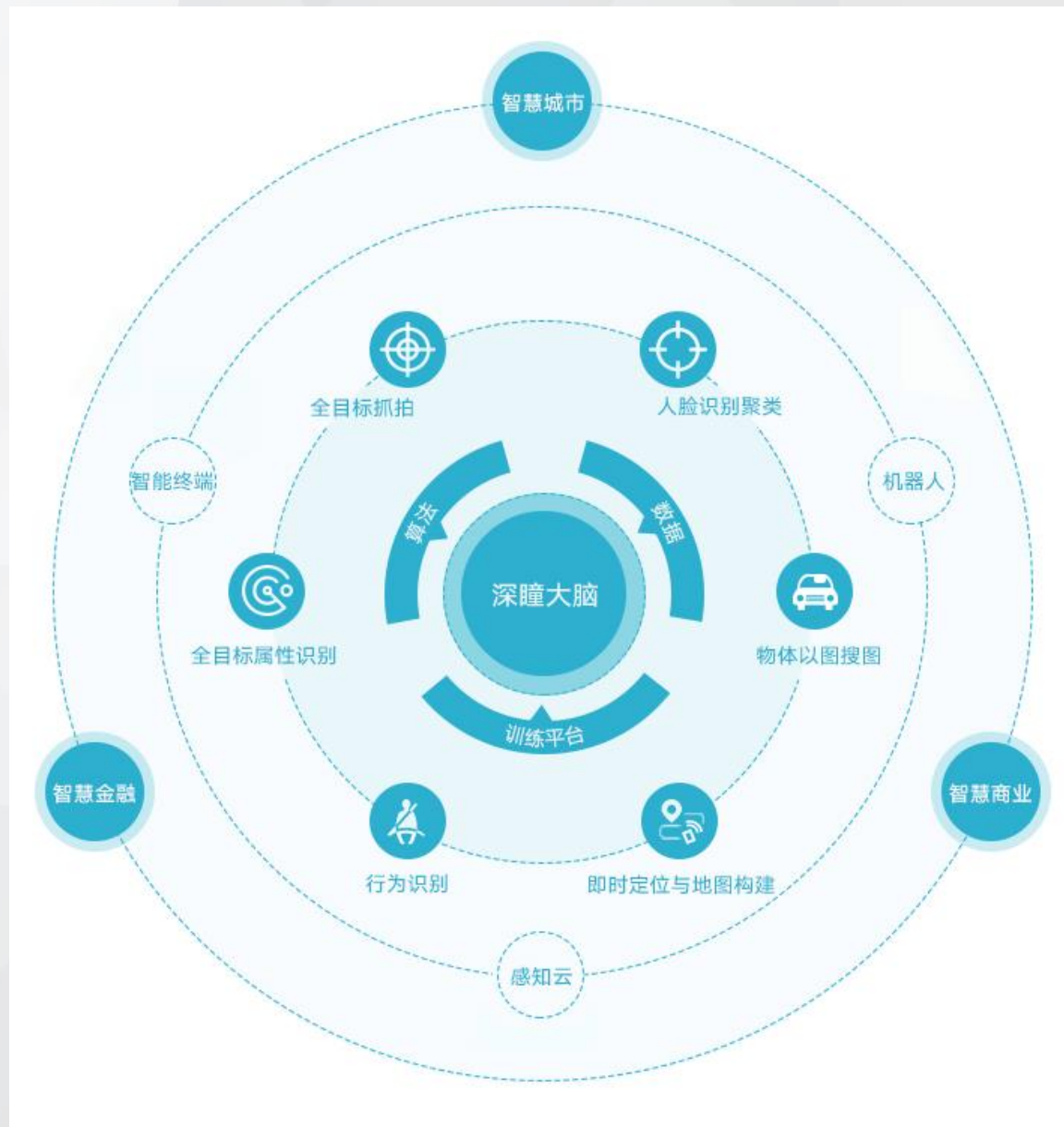


人体再识别 (ReID)

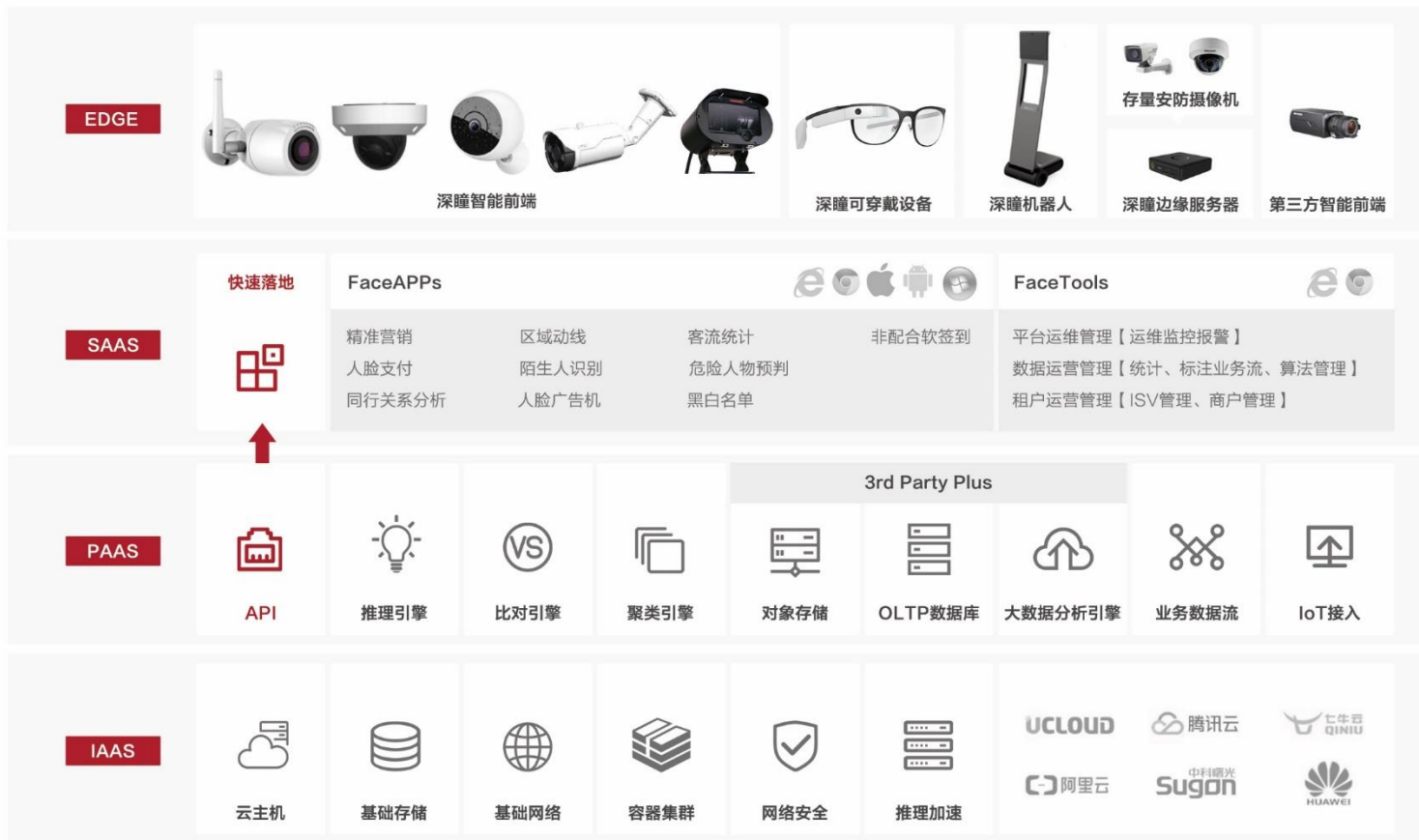


车辆以图搜图 (ReID)

以深瞳大脑为核心的技术产品研发架构



打造端+云的视觉计算系统



目标抓拍识别 - 深瞳人眼摄像机



深瞳人眼摄像机动态人脸识别系统

人脸识别率: 0.84

2017-02-27 下午 12:46:05 星期一
深瞳人眼摄像机

抓拍图

特征库

相似度: 85%
乐爱兵

相似度: 86%
王一舒

相似度: 91%
潘伟民

相似度: 85%

格灵深瞳威目人脸识别系统

实时监控

设备总数: 4

抓拍图

14:23:44 14:23:41 14:23:16 14:23:14 14:23:02 14:23:53 14:23:53 14:23:52

14:27:10 14:27:10 14:25:44 14:25:14 14:25:02 14:23:33 14:23:33 14:23:31 14:23:31

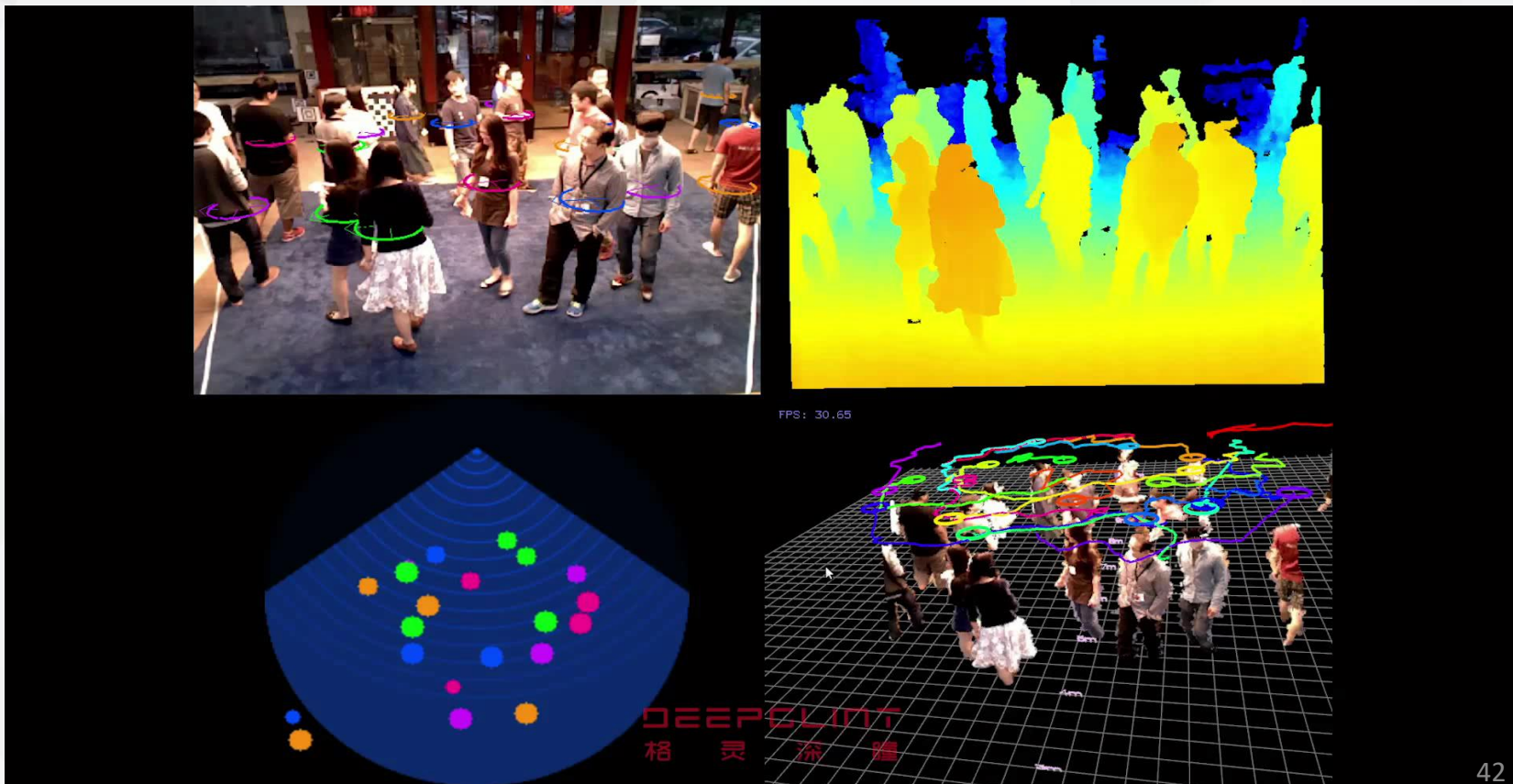
41

行为分析 - 皓目行为分析仪



银行、零售、司法、博物馆领域

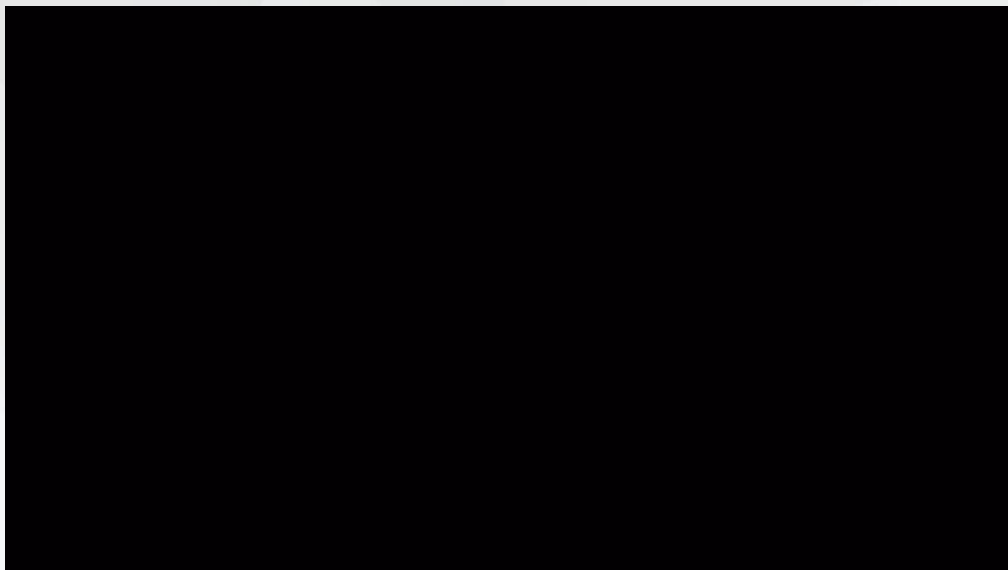
创新性地采用三维计算机视觉技术，突破了传统二维图像处理方法的限制，基于RGBD数据，利用运动在时空的连续性等线索，实时而精确地实现对人的检测、追踪和识别。



由感知到行动（感知、决策、控制）



机器人



自动驾驶汽车



驭势联合创始人秦岩1月份路演（北理工）
AI JIANG, UNIFEI.COM/UNIFEI

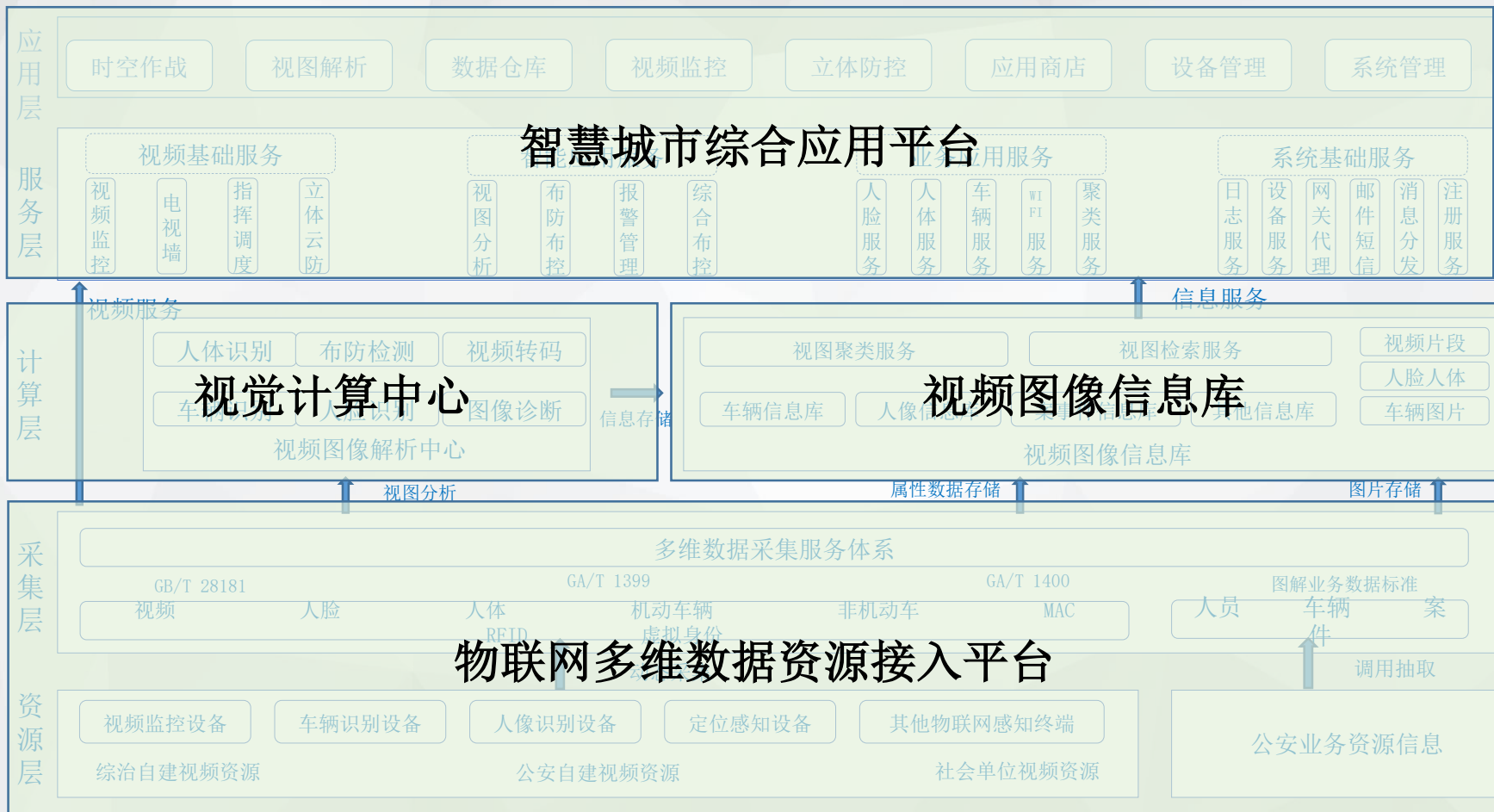
应用领域



AI+



智慧城市解决方案



统一认证体系

智慧城市应用 - 找人新篇章



1

以人体背影找正面

2



65分
人脸质量

3

以不清晰人脸找清晰人脸



确认身份

4

人脸信息识别



轨迹刻画

每页显示 5 1 2 4 下一页 刷新

[案件] 盗窃罪
盗窃罪 201412207 刑事案件 盗窃 入户 盗窃 刑事 昆明市
公安局 手动录入 2014-12-08 16:14:35 未侦破 昆明 有
人入室盗窃 2014-12-08 16:18:27
上传者: lizhuo

[案件] 20150207重庆江山假日停车场砸车盗窃案
20150207重庆江山假日停车场砸车盗窃案
J201502071100234 刑事案件 盗窃 盗窃汽车 刑事 重庆
市公安局 手动录入 2015-02-07 00:00:00 未侦破 重庆江
山假日停车场 20150207重庆江山假日停车场砸车盗窃
上传者: admin

[案件] 合肥市阳信县盗窃案
合肥市阳信县盗窃案 H00012 刑事案件 盗窃 盗窃工业原
材料 刑事 合肥市公安局 手动录入 2014-10-09 15:31:00
未侦破 阳信县厂区 2014年10月9日厂区工作人员上班后
发现厂区少了20袋工业原料 C14120914531374372
上传者: chenchen

[案件] 测试案件管理
测试案件管理 A
重庆市公安局 报警
星光三路 2015-
P1507221019
上传者: wuwu



人货场关联，打造完整数据链

- 利用人脸识别、属性识别、商品识别等技术，将人与关注、购买记录关联起来，从而在商业场景将物理世界数字化，并利用线上分析手段提升商业效率



- 商品认知
- 人脸认知
- 人货关联



Cam物联网
传输向量



客流总计 (人)
1654

人脸虚拟ID技术

通过人脸识别将同一个人多次到店关联起来

人员信息

抓拍结果 2018/07/11

性别: 男
年龄: 26 - 35
出现次数: 259
最近出现时间: 2018/07/11 22:58:06
最近抓拍来自设备: B区向东
抓拍设备ID: 1428014358

抓拍结果 2018/07/11

时间: 22:57:23 — 22:58:06
设备组: 公司测试

抓拍结果 2018/07/11

时间: 21:40:51 — 22:08:28
设备组: 公司测试

利用人脸将商品购买记录关联起来，并形成动线和热力图

华北区

- 奥体店
- 育知东路店**
- CBD店

华南区

- 金鹰广场店

☆ 当前店铺监控点 (3)

- 进门
- 糖果区
- 乳制品
- 日化
- 收银台

当前在店人员 (14人)

全部人员

上次购买的清单:

- 荷兰豆
- 佳洁士牙膏钻白薄荷...
- 徐福记凤梨酥 (340...

00:13:34 历史 >

上次购买的清单:

- 美汁源果粒橙
- 佳洁士牙膏钻白薄荷...
- 徐福记凤梨酥 (340...

00:13:34 历史 >

上次购买的清单:

- 飘逸长发洗发水
- 如草般生发洗发露
- 发如丝生发剂

00:13:34 历史 >

上次购买的清单:

- 荷兰豆

新会员

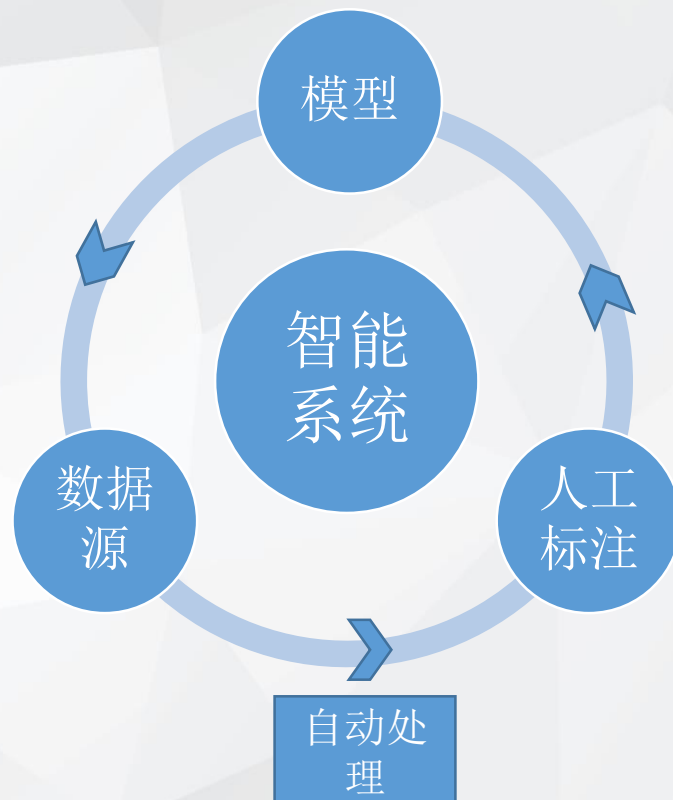
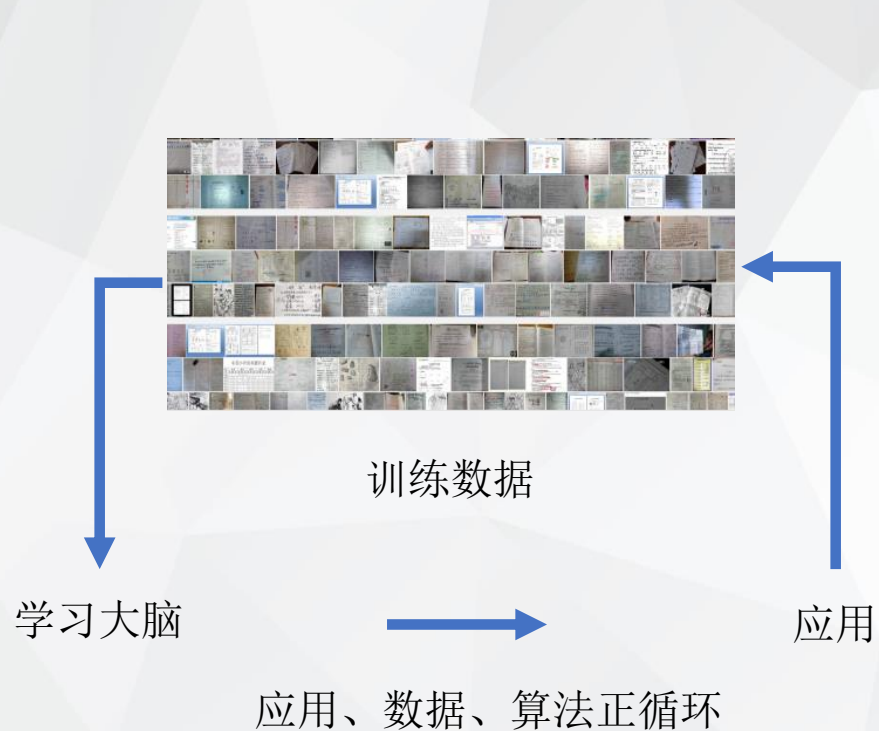
上次购买的清单:

- 郭志远饮料


未来趋势

算法、芯片、数据三级火箭将推动计算机视觉技术在准确率、成本、规模、功能等方面取得显著进步，并以边缘计算和智能感知云的产品形态，在智慧城市、智慧商业、工业自动化、智慧医疗、自动驾驶等多个垂直场景逐步落地。

构建自我持续改进的视觉智能系统

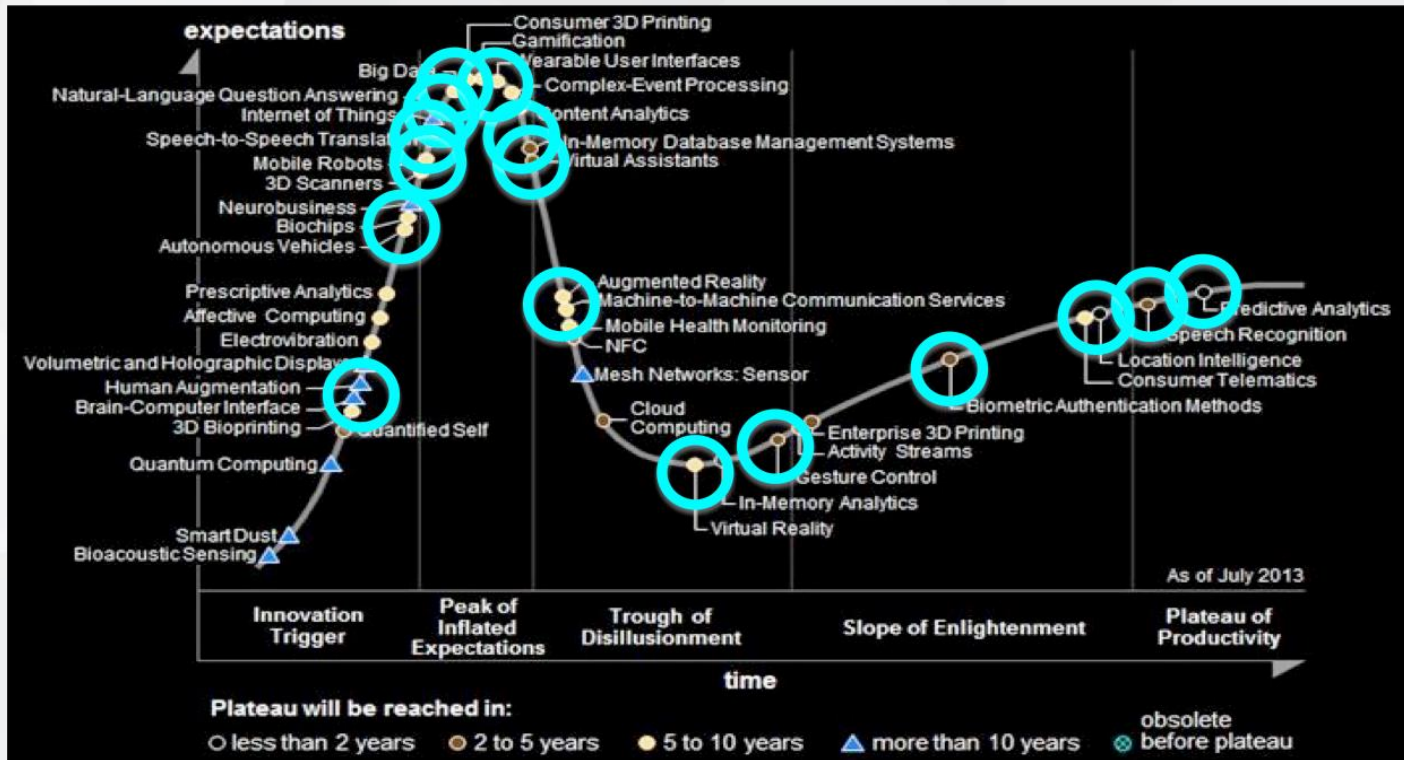


持续学习的智能系统

A wide-angle landscape photograph featuring a two-lane asphalt road that recedes into the distance. The road is flanked by golden-brown grassy fields. In the background, there are rolling hills, a small white building on the left, and a bright sun setting on the right, creating a dramatic sky with dark, heavy clouds and a bright glow. The overall mood is contemplative and serene.

依然在路上

人工智能必然是一次长跑



人们往往高估一项技术的短期表现，
而低估一项技术的长期表现

做有温度的人工智能



谢谢大家!

| 格灵深瞳，让计算机理解世界，让AI使生活更美好 |