



QCon 全球软件开发大会
INTERNATIONAL SOFTWARE
DEVELOPMENT CONFERENCE

BEIJING 2018

未来都市 - 智慧城市与基于深度学习的 机器视觉

演讲者 / 陈宇恒



基于实践经验总结和提炼的品牌专栏
尽在【极客时间】



重拾极客时间，提升技术认知

GTLC
GLOBAL
TECH LEADERSHIP
CONFERENCE

全球技术领导力峰会

通往**年薪百万**的CTO的路上，
如何打造自己的技术**领导力**？

扫描二维码了解详情



概要

- 我们是谁
- 智慧城市中机器视觉应用
- 我们是如何构建城市级AI+智慧城市系统
- 大规模深度学习实战系统的几点经验



- 商汤科技联合创始人，架构师
- C++/Go/Rust/Ruby开发者
- 多个开源项目贡献者
- NIPS国际会议论文作者

 @chyh1990

软银孙正义收购Google旗下的机器人公司Boston Dynamics和Schaft



2017.6

2017.3



Intel 153亿美元收购无人驾驶技术公司 Mobileye

软银孙正义设立1000亿美元人工智能基金，320亿美元收购芯片架构公司ARM



2016.7

2016.8

赢得ImageNet 5项主要比赛中的3项世界冠军

通用 10亿美元 收购无人驾驶技术初创公司Cruise Automation



2016.3

2016.3



DeepMind AlphaGo 围棋AI击败李世石

首次中国公司在ImageNet竞赛夺冠，视频分析技术登顶

2015.11

2014.11



Google 6.6亿美元收购深度学习公司 DeepMind

人脸识别大幅提高精度，商汤科技首次突破人类肉眼识别准确率，领先于Facebook

2014.6

2013.12



Facebook 建立人工智能实验室并聘用 Yann LeCun

Google 5000万美元招入Hinton，发布基于深度学习的搜索引擎



2013.3

2012.12



深度学习鼻祖 Hinton 赢得ImageNet图像识别竞赛

Microsoft 深度学习驱动的语音识别大幅提升精度



2011年中

公司简介

20年
科研经验

800余位
技术研发人员

150余位
人工智能博士

唯一
深度学习
平台公司

历史业绩

累计融资
全球最大

核心技术
全球领先

商业营收
市占率
行业第一

400余家
大型客户伙伴
共同发展

领先技术



AI+金融



AI+智慧城市

AI

AI+智能手机



AI+芯片



AI+无人驾驶

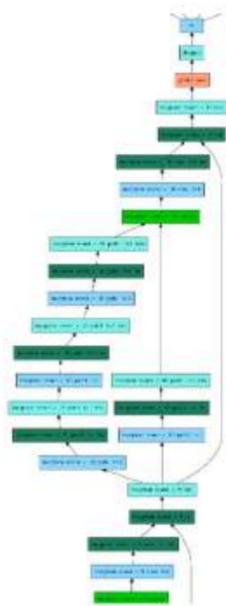


AI+医疗图像



AI+智慧城市：以智能安防场景为例

特征向量



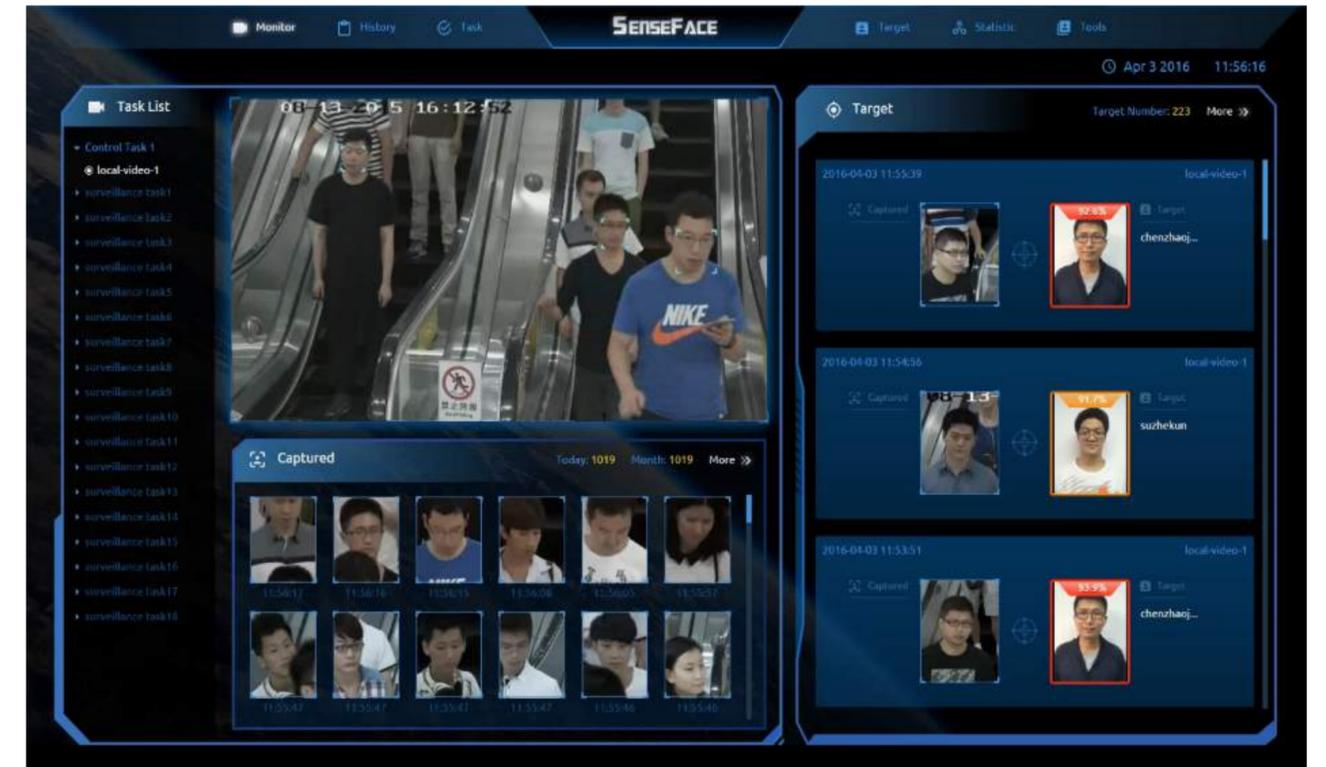
AI+智慧城市

2015-2017

- 单机、简易分布式人脸检测、跟踪、比对平台
- 处理数十路到数百路监控摄像头数据
- 千万级别深度学习特征检索
- 行业试水

2018-2019

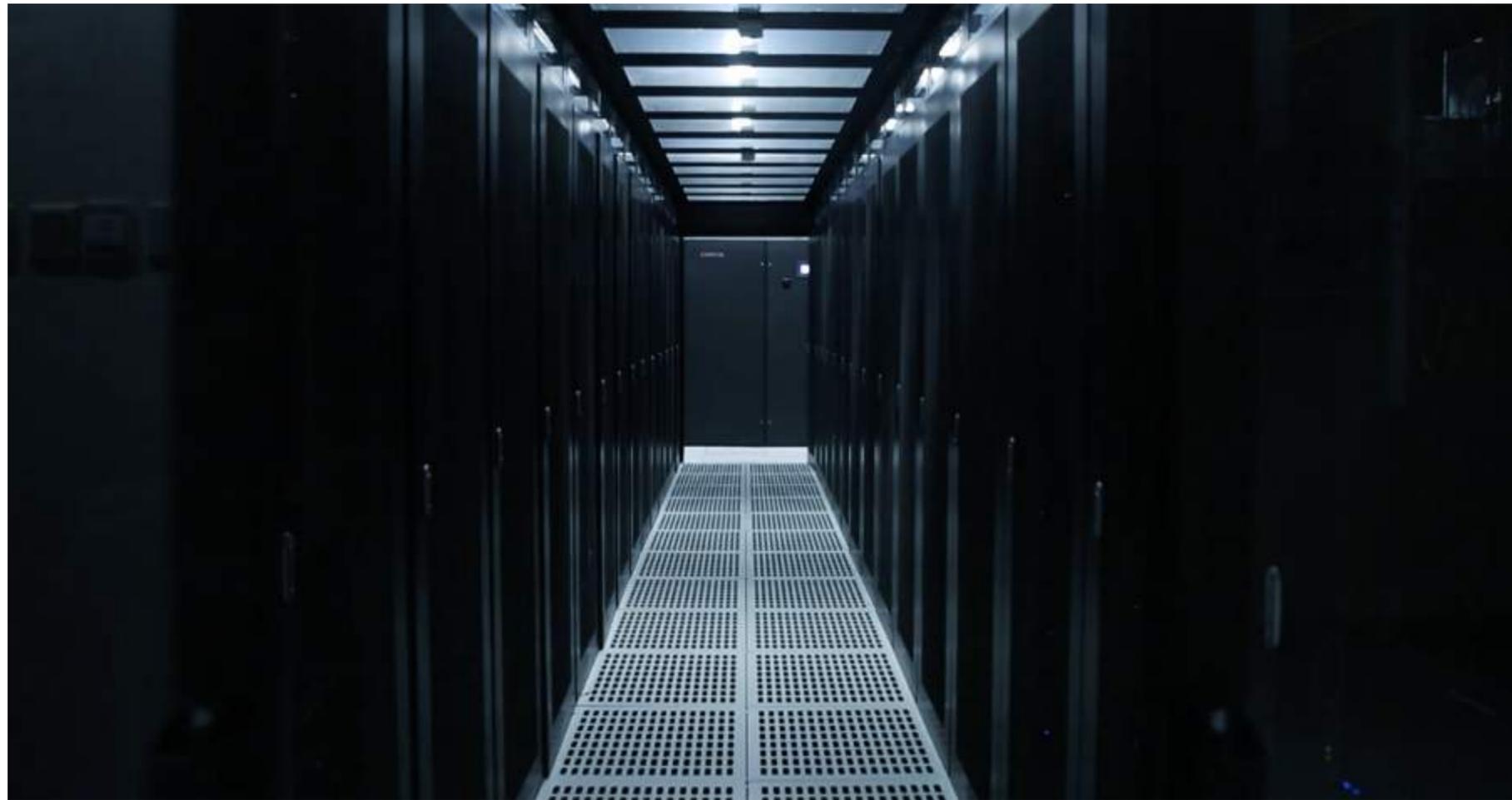
- 云原生Cloud-Native超大规模视图存储、处理、检索
- 处理**数万到数十万**路，城市范围级别监控、门禁摄像头数据
- **10-100 Billion**级别深度学习特征检索
 - PB以上级别数据库存储
 - 100PB级别抓拍图片存储
 - 每秒万次并发检索请求
- 大规模推广应用
- 某种程度上说，城市内所有市民都是系统的用户



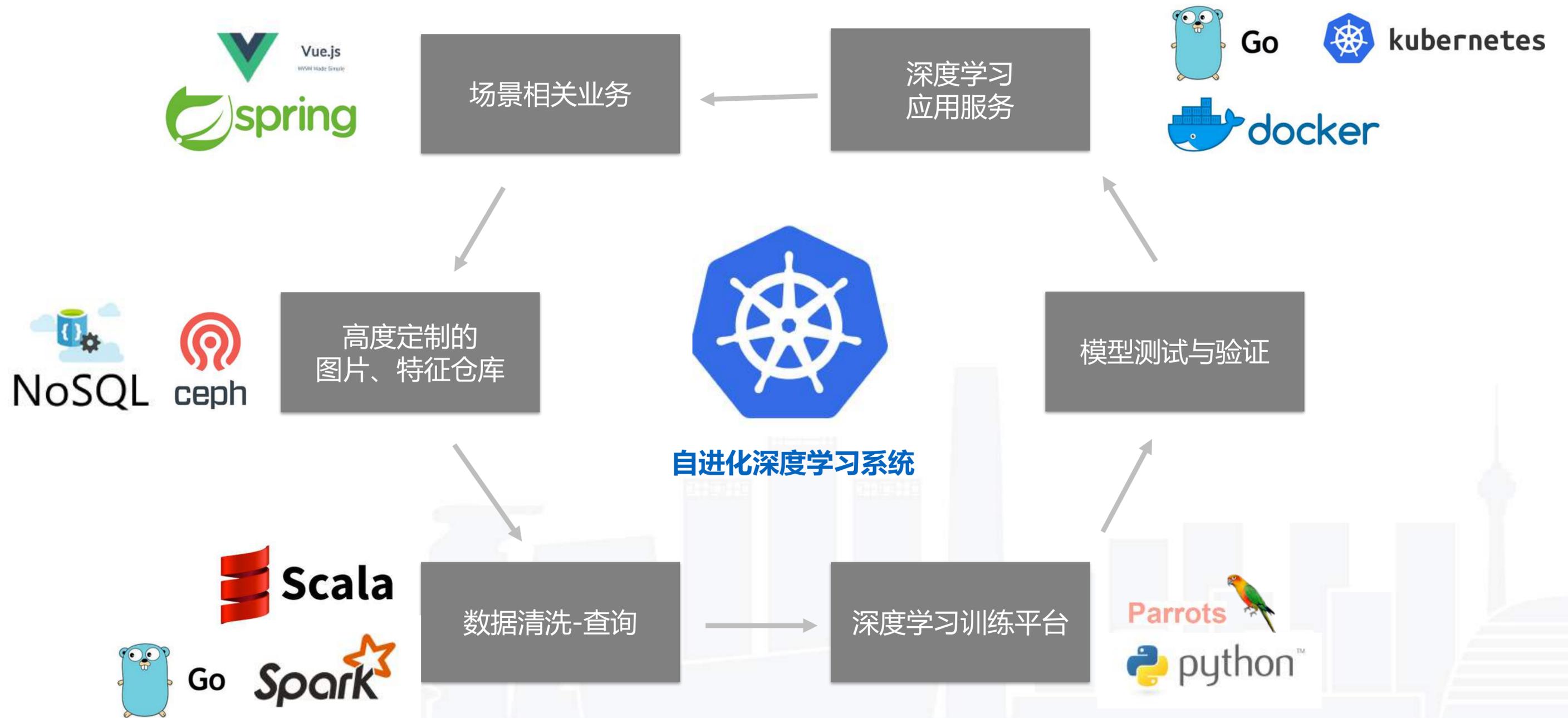
深度学习算法发展为平台系统赋能



自我演化的异构人工智能云



云原生的深度学习数据闭环



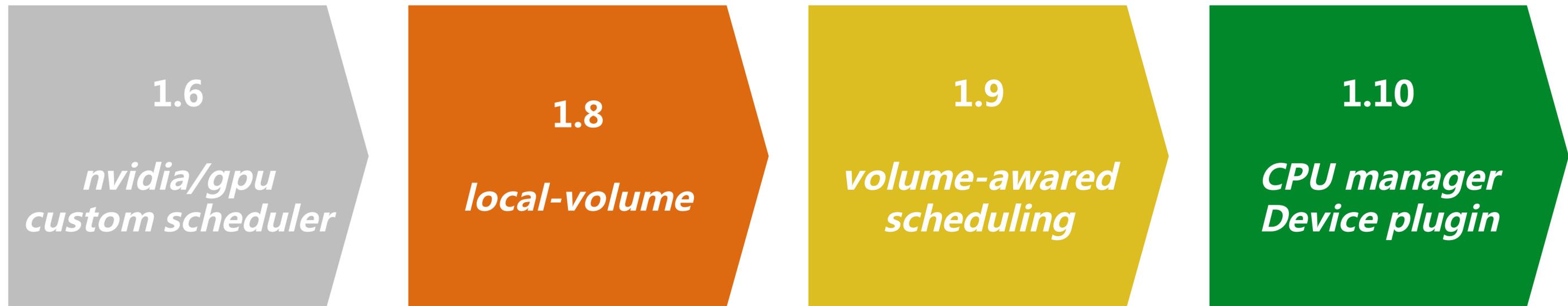
深度学习算法在产品应用中的挑战

- 深度学习算法也需要“深度”学习业务需求
 - 处理特殊输入，如模糊、黑白照片
 - 适配具有不同特征的数据源
 - 在严肃应用中，客户追求100%准确率，算法性能提升永无止境
- 深度学习模型需要在准确率和速度上做均衡
 - 使用更加精巧的模型和Operator设计
 - 使用模型压缩算法，在基本保障准确率的情况下大幅提升速度
 - 利用最新的硬件特性，如GPU TensorCore/int8



*示意图来自互联网

Kubernetes在异构系统调度中的挑战



- Kubernetes版本发布快，新特性更新频繁，对异构调度的支持不断加强；但配套设施落后（e.g. Spark on K8s, GitlabCI）
- 容器系统调用栈深，需要仔细验证操作系统，内核及异构设备驱动的兼容性
- Kubernetes对NUMA、异构计算、存储设备的调度能力待加强

Go语言在高性能系统中的实践经验

- 为什么用Go
 - 比起C++，更易于实践各种**并发**模式
 - 比起Java，更加简洁，更易于与**C/C++交互**
 - 比起脚本语言，**类型和内存安全**，保证重构效率与产品质量
 - 完善的配套工具，如go test, gofmt, go lint, race-detector

```

import (
    "sync"
)

// FooPipeline ...
type FooPipeline struct {
    stage1 chan TaskStage1
    stage2 chan TaskStage2
    sink   chan Sink

    wg sync.WaitGroup
}

func (f *FooPipeline) stage1() {
    defer f.wg.Done()
    for e := range f.stage1 {
        f.stage2 <- someProcess1(e)
    }
    close(f.stage2)
}

func (f *FooPipeline) stage2() {
    defer f.wg.Done()
    for e := range f.stage2 {
        f.sink <- someProcess2(e)
    }
    close(f.sink)
}

```

```

// Run ...
func (f *FooPipeline) Run() {
    defer f.wg.Done()

    f.wg.Add(1)
    go f.stage1()
    f.wg.Add(1)
    go f.stage2()

    f.wg.Add(1)
    for e := range f.sink {
        output(e)
    }
}

// Stop graceful exit the whole pipeline
func (f *FooPipeline) Stop() {
    close(f.stage1)
    f.wg.Wait()
}

```

Go语言在高性能系统中的实践经验

- Go在开发高性能应用上也有一些不足，对比C++：
 - 无法直接控制操作系统线程，CUDA调用需要特殊处理
 - 部分标准库实现依赖reflect，性能较差
 - GC的带来的开销，如在Go Heap上构建百万以上级别的对象缓存，需要仔细优化

```
import (  
    "bytes"  
    "encoding/binary"  
)  
  
func toBytes(fs []float32) []byte {  
    var buf bytes.Buffer  
    binary.Write(&buf, binary.BigEndian, fs)  
    return buf.Bytes()  
}
```

百倍慢于等价的C实现！

回顾

- 智慧城市中，在智能安防领域机器视觉有着爆发式应用
- 我们使用基于深度学习的机器视觉技术，构建了超大规模的自我演化的分布式智能系统
- 在构建这个规模的系统中，我们广泛使用了Kubernetes、Go等流行技术，“那些年踩过的坑”



关注QCon微信公众号，
获得更多干货！

Thanks!



主办方 **Geekbang** > **InfoQ**
极客邦科技

GMITC 2018

全球大前端技术大会

—— 大前端的下一站 ——



<<扫码了解更多详情>>

关注 ArchSummit 公众号

获取国内外一线架构设计

了解上千名知名架构师的实践动向



Apple • Google • Microsoft • Facebook • Amazon 腾讯 • 阿里 • 百度 • 京东 • 小米 • 网易 • 微博

深圳站：2018年7月6-9日

北京站：2018年12月7-10日

QCon

上海站

全球软件开发大会【2018】

2018年10月18-20日

7折

预售中, 现在报名立减2040元

团购享更多优惠, 截至2018年7月1日



极客邦科技
企业培训与咨询

Geekbang

扫码关注

获取更多培训信息

