



第九届中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2017

深度学习在图像审核的应用

腾讯优图实验室

谭国富

<http://open.youtu.qq.com>

关于优图实验室

优图团队立足于社交网络大平台，借助社交业务积累的海量人脸、图片、音乐等数据，专注在人脸、图像、音乐、语音、机器学习等领域开展技术研究，并积极推动研究成果在业务中落地产生价值。



人脸识别



图像识别



音频识别

- 2017年4月，在国际MegaFace海量人脸识别数据库刷新世界记录
- 2017年3月，在国际LFW人脸数据库上刷新世界纪录
- 2017年3月，在国际ICDAR 2015文本检测项目中刷新世界纪录
- 2016年，优图实验室获得“腾讯行业贡献奖”
- 2015年，优图哼唱识别技术获得“腾讯年度微创新奖”
- 2015年10月，在国际音频比赛MIREX的哼唱识别比赛中，取得总成绩世界第一，并刷新其中一项世界纪录
- 2015年4月，在国际Pascal VOC2012物体分类赛刷新世界纪录
- 2014年11月，腾讯优图人脸检测刷新FDDB世界纪录
- 2014年，人脸识别获得腾讯公司年度“重大技术突破奖”
- 2014年，联合上海交通大学获得上海市科技进步二等奖
- 2013年，优图压缩获得腾讯公司年度“重大技术突破奖”
- 2012年，优图电商联合团队获得腾讯公司级别“卓越运营奖”

目录

- 00 / 图像审核的行业背景
- 01 / 腾讯优图内容审核能力介绍
- 02 / 深度学习技术介绍
- 03 / 内容审核的扩展和延伸

内容审核 - 行业现状

不良信息泛滥，监管猝不及防

➢ 随着互联网的飞速发展和信息量的猛增，大量的色情图片、暴力等不良信息夹杂其中，严重影响着互联网的健康发展。



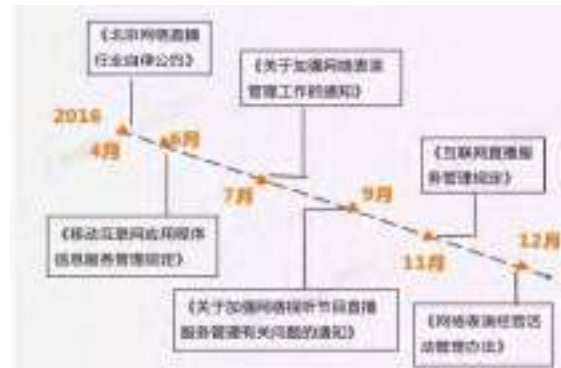
微信朋友圈日上传图片10亿张，视频播放20亿次

➢ 直播行业的快速兴起，使得视频中不良信息含量更加迅猛增长，色情暴力等不雅视频频繁流出，导致各网络直播平台面临危机。



4000亿QQ空间存量图片，每天空间相册新增6亿张上传图片

➢ 内容监管日趋严格，2017年上半年，各大直播行业协会相应成立，行业平台自我规范的同时，网信办、文化部等国家部门对于直播行业监管也越发严格，几乎所有知名的直播平台均被有关部门点名查处过，特别2017年月中旬，黄鳝事件引爆网络，让色情直播再度被推上舆论浪尖。



内容审核 - 痛点和诉求

业务痛点：面对越来越爆发的安全风险，解决办法门槛高，成本高；迫切需要技术解决方案

技术诉求：自动识别图片或视频中出现的文字、二维码、logo等内容以及违规人像、淫秽、血腥、暴力、极端主义、恐怖主义图像等，方便平台进行违规处理和风险管控。

自建识别模型

昂贵的专业机器、AI专家，
样本不足导致识别模型漏
过模型调优难度大

默默承受

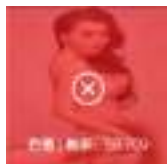
一旦出现严重违规平台
面临停业整顿风险



加大审核人力

人力审核疲劳容易发生漏过，人力招聘、
管理需要耗费不小成本





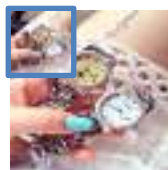
图像内容审核技术



图像识别技术



OCR技术



图像分割以及超分辨率技术

优图图像技术还包括：图像分类、图像增强、艺术滤镜、图片去水印、图像融合、图像修补等。

内容审核 - 图片鉴黄解决方案

区分图像中的色情、性感和正常内容

DeepEye可给出图片属于色情、性感和正常的概率，并结合三者概率给出综合分，通过分数所属区间判断图片性质。

- **色情图片**：包含露点或不雅行为的图片，可直接打击；
- **性感图片**：又称疑似图片，不含直接色情内容但有一定的诱惑性，建议进行人工审核；
- **正常图片**：不含不良内容的正常图片。

色情图片



性感图片



内容审核 – 图像暴恐内容识别

● 暴恐识别技术

- 对于输入的图片，系统将会通过对其内容的识别分析给出其属于武装份子、管制刀具、枪支弹药、人群聚集、火灾、血腥、极端主义或恐怖主义标识的概率，通过其概率最大的类型，判断其图片性质属于属于暴恐还是正常。
- 高准确率：在内部业务上测试，准确率97%，覆盖80%以上的案例

● 识别应用：腾讯云，微云，QQ群

- 腾讯云，承担每天数亿的图像审核，已经累计支持上百家客户。
- 微云，QQ群，支持视频识别的解决方案，成熟灵活的产品方案，帮助业务扫描掉互联网暴力、恐怖内容，有效的降低业务风险。



武装份子



管制刀具



枪支弹药



极端主义、恐怖主义标识



人群聚集



火灾



血腥

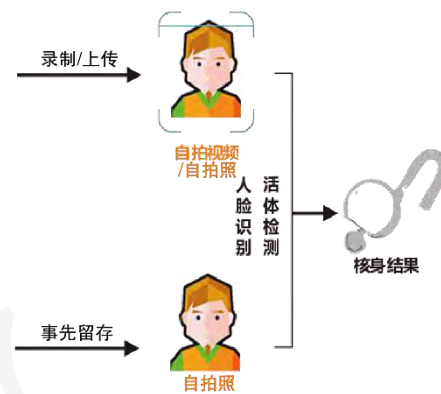
内容识别 – 人脸识别

● 人脸识别核身技术

- 上亿级别的人脸检索，秒级的检索速度从黑名单，白名单数据库中返回目标人脸信息。
- 技术指标：优图人脸识别通过传统方法和深度学习技术结合，以空间面孔墙和微众银行远程核身为基础，在性能上达到LFW 99.80%。

● 政治敏感人物识别，直播，视频等场景

- QQ，微云等：非法设置领导人头像，公众人物，明星等等他人肖像。
- 直播，游戏视频等，非法植入领导人，政府国际公众人物，明星等等。



政治人物识别



检索结果：奥巴马
相似度：98%
政治人物
不在黑名单

无版权人物识别



检索结果：林志玲
相似度：99%
在版权库

图片场景识别

● 图片场景识别技术

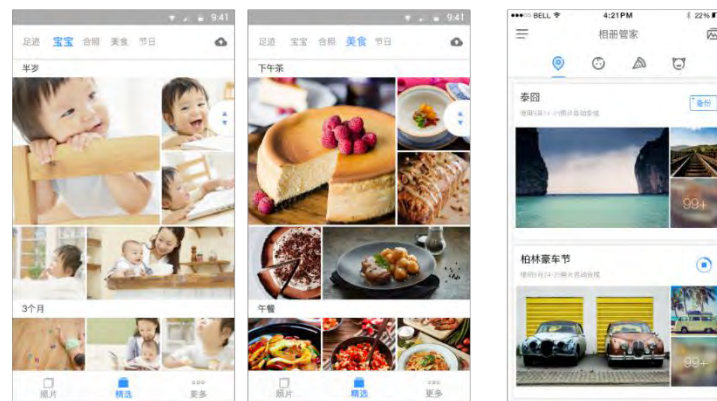
- 标签体系：面向社交领域的热词标签**200余种**，涵盖人物、风景、人造物、建筑、动植物、食物等**9个大类**。
- 技术指标：20个类别平均准确率MAP>90%以上，200种MAP>63%以上，性能CPU上约200ms/张。



识别标签效果

● 社交图像分类应用：微云，相册管家

- 微云，相册管家新版本推出智能全自动图片分类，通过上百种常用图片标签，实现对所有照片的自动识别分类。
- 用户上传图片之后即可被智能分类，各大类下包括小类如人物大类下有合影、女孩、男孩、聚会等小标签。只需要输入或点击标签即可获得对应类别的图片。



微云相册，相册管家

OCR识别 – 证件类

● 证件类OCR识别

- 优图OCR识别技术支持数字识别和超过7000个常用汉字的识别
- 在国际ICDAR 2015文本检测项目中刷新世界纪录
- 技术指标：名片91.4%，驾驶证91.5%，行驶证85.5%
- 应用场景：身份证、驾驶证、行驶证、营业执照、银行卡、车牌、名片等等多个垂直场景

● 落地应用

- 手Q名片识别，广点通营业执照识别：在手Q的扫一扫入口中，可以体验。
- Webank身份证识别，主播实名认证：方便用户快速的输入证件信息。



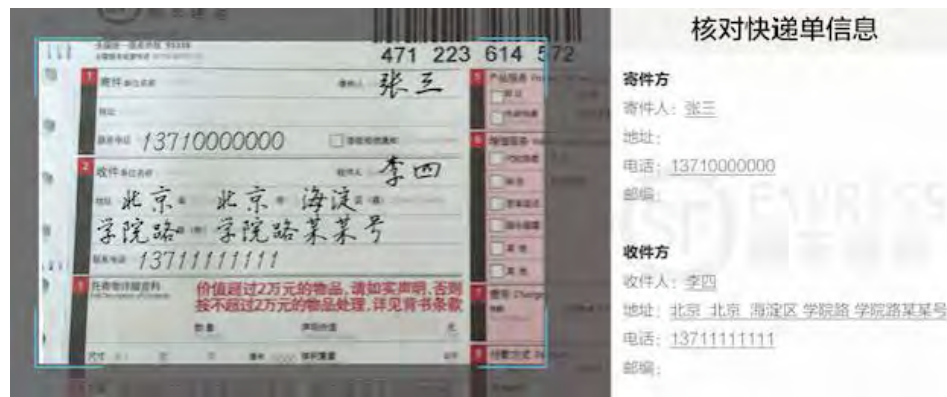
OCR识别 – 通用场景和手写

● 手写体OCR以及通用OCR识别

- 手写体手机/电话识别准确率可达99%以上。突破业界复杂手写体识别的难题。
- 通用场景准确率和召回率均在88%以上。
- 应用场景：快递单据，广告识别等。

● 落地应用：

- 微云相册识别，广点通广告识别，顺丰快递单据识别，微云文本识别等



更多图像定制识别功能

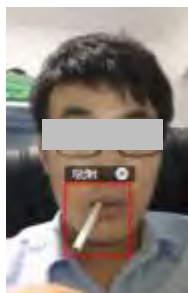
● 图标识别

- 图标形式：水印、二维码、条形码、logo等。并可根
据业务需求定制。
- 基于团队自研的深度识别引擎，核心算法与基础模型
经过多轮优化，可以准确快速地识别图片中所包含的
各类图标



● 烟雾，吸烟识别

- 基于视频直播监管需求，提供吸烟，烟雾，涉嫌吸毒
等场景的识别能力





深度学习介绍



图像海量数据的积累



深度网络训练选择



加快训练 - 分布式训练系统

RPN

加快计算 - 深度学习算法加速

深度学习 - 带动的AI浪潮

第1阶段：人工智能起步期
(1956-1980s)

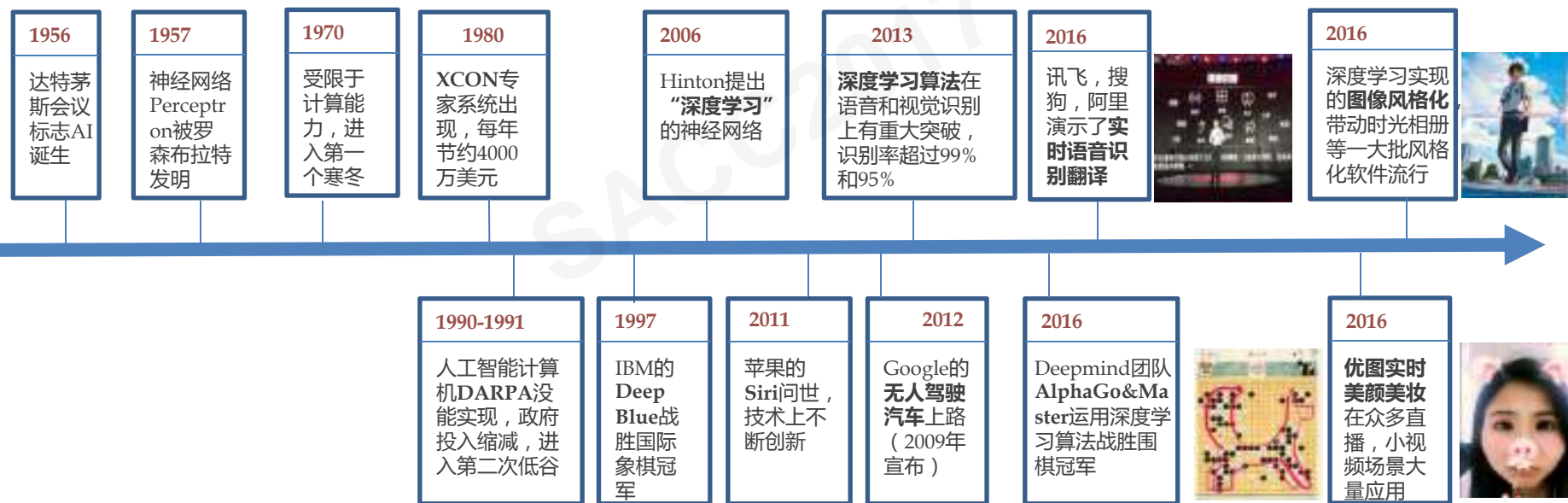
第2阶段：专家系统推广
(1980s-1990s)

第3阶段：深度学习
(2000s-至今)

深度学习全面爆发
2016 -

技术发展

应用突破



深度学习 – 相对于传统机器学习方法的突破

- 传统方法： $F(\cdot)$ =特征提取+学习算法



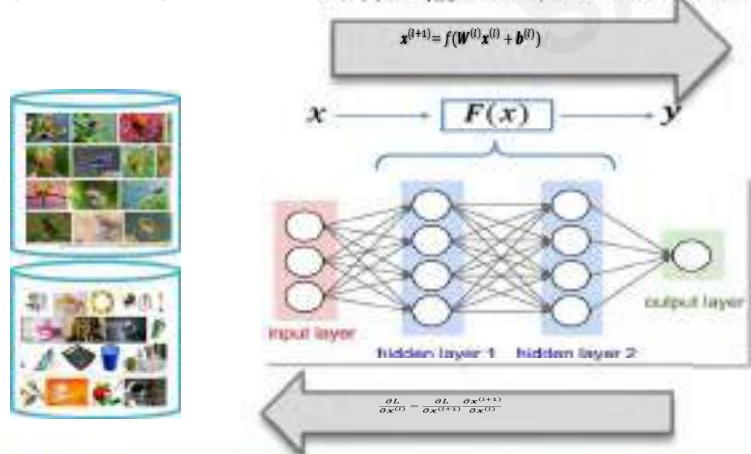
图像表示：Gabor, SIFT, HOG, LBP, POEM, LGBP, LPQ
 图像集表示：Manifold, GMM, Covariance



模式分类：SVM, Boosting, 决策树, 子空间学习, 度量学习:
 PCA/LDA, Manifold, LMNN, NCA.....
 词典学习&稀疏编码

$F(\cdot)$

- 深度学习：多层神经网络对输入进行非线性变换的学习技术



鸟类

正常

Towards Better Analysis of Deep Convolutional Neural Networks

Mengchen Liu, Jiaxin Shi, Zhen Li, Chongxuan Li, Jun Zhu, Shixia Liu

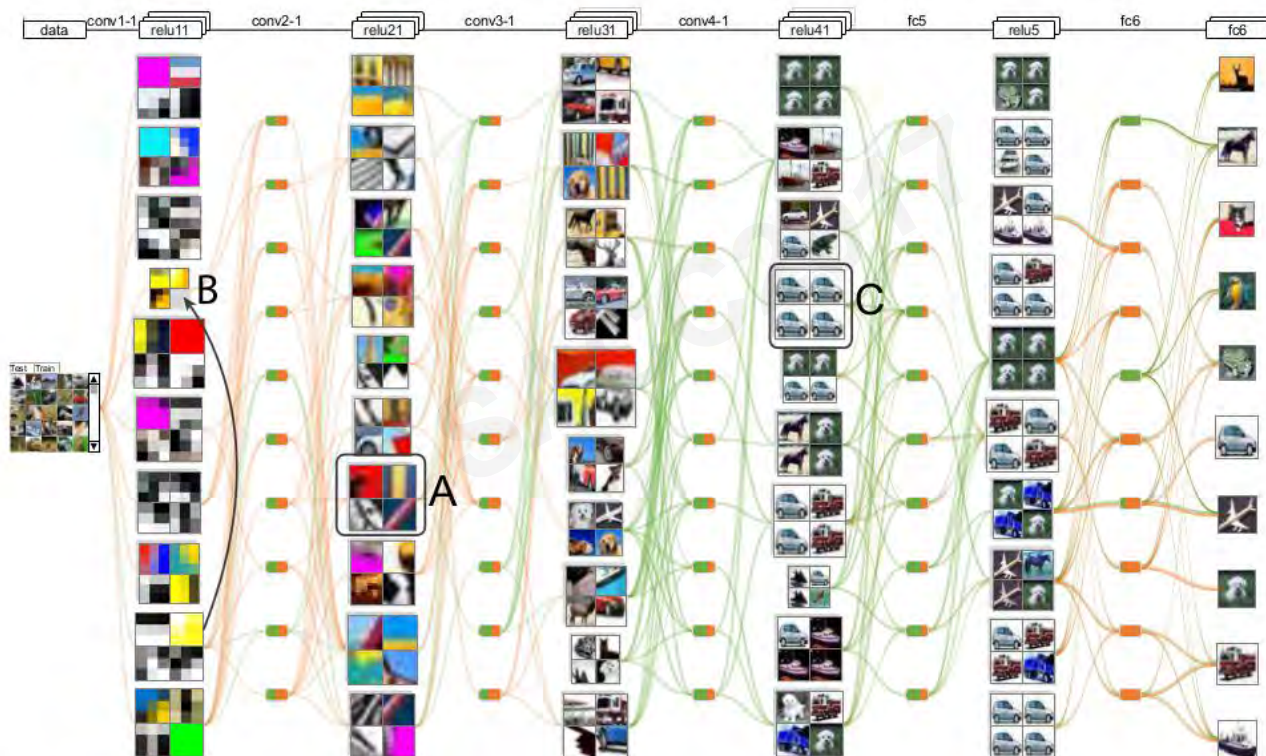
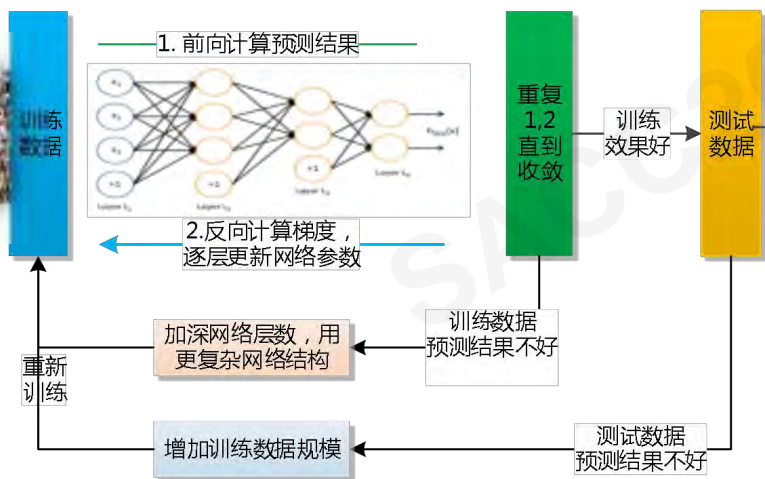


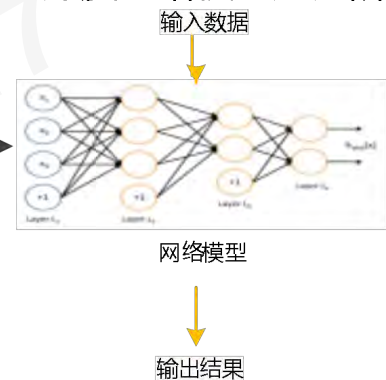
Fig. 1. CNNVis, a visual analytics toolkit that helps experts understand, diagnose, and refine deep CNNs.

深度学习 – 解决问题的统一框架

1. 深度网络模型训练阶段



2. 深度网络模型应用阶段



深度学习 - 如何设计网络结构

整体结构：深度大幅增加，宽度适当增加：

AlexNet, 8层, I型



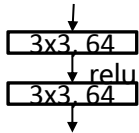
VGGNet, 19层, I型



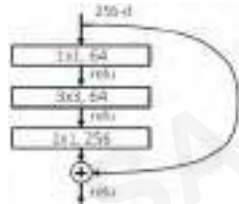
GoogleNet, 22层, W型



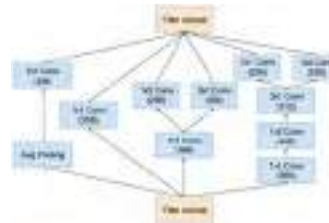
局部结构：直线型→分支型



直线型 (I型)



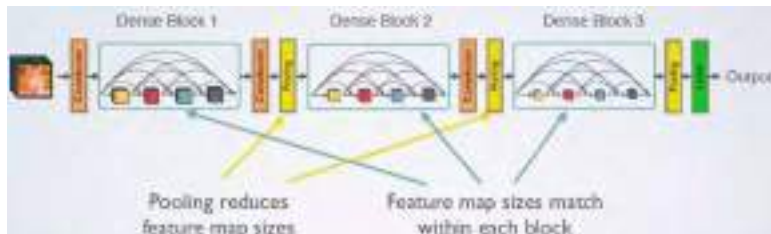
局部双分支型 (V型)



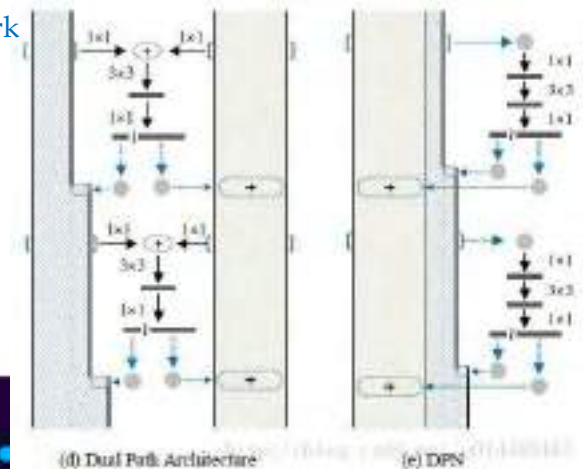
局部多分支型 (W型)

稠密连接：分支连接→多层连接

DenseNet

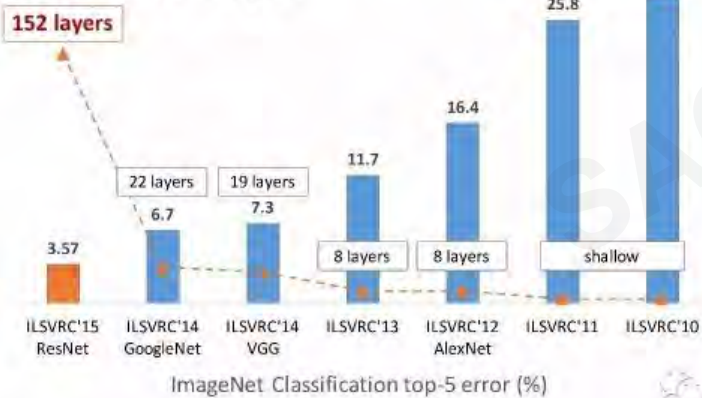


Dual Path Network

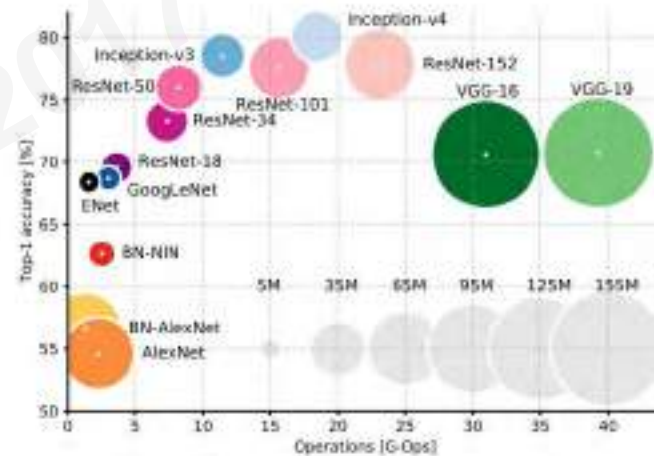


深度学习 - 更深更复杂的网络带来效果提升，计算量提升

Revolution of Depth



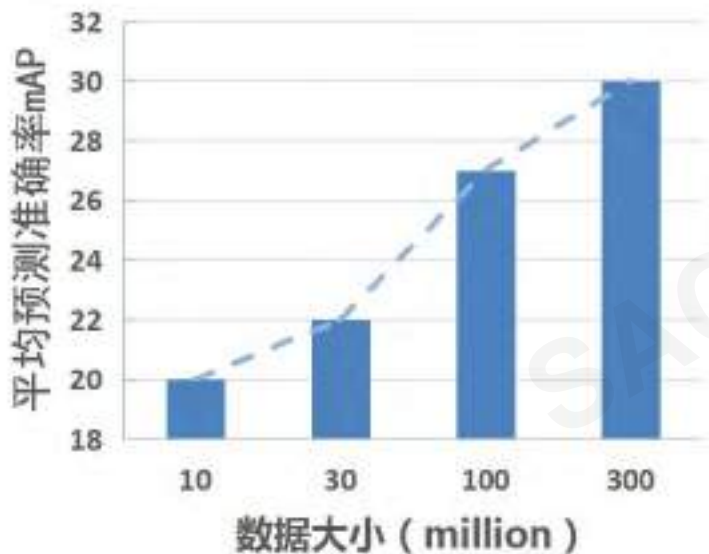
Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun, "Deep Residual Learning for Image Recognition", CVPR, 2016.



An Analysis of Deep Neural Network Models for Practical Applications
<https://arxiv.org/abs/1605.07678>

深度学习 - 更多数据带来效果质的提升

数据越多，效果越好



数据除了人工以外的获取方法

- 扩增数据 – 各种图像增强，加噪声
- 非监督学习 - 聚类
- 迁移学习 – 利用相似任务训练好的网络
- 生成样本数据 – 深度生成对抗网络

论文链接：**Revisiting Unreasonable Effectiveness of Data in Deep Learning Era**
<https://arxiv.org/abs/1707.02968>

深度学习 训练框架 和 硬件选择

不同场景，不同框架

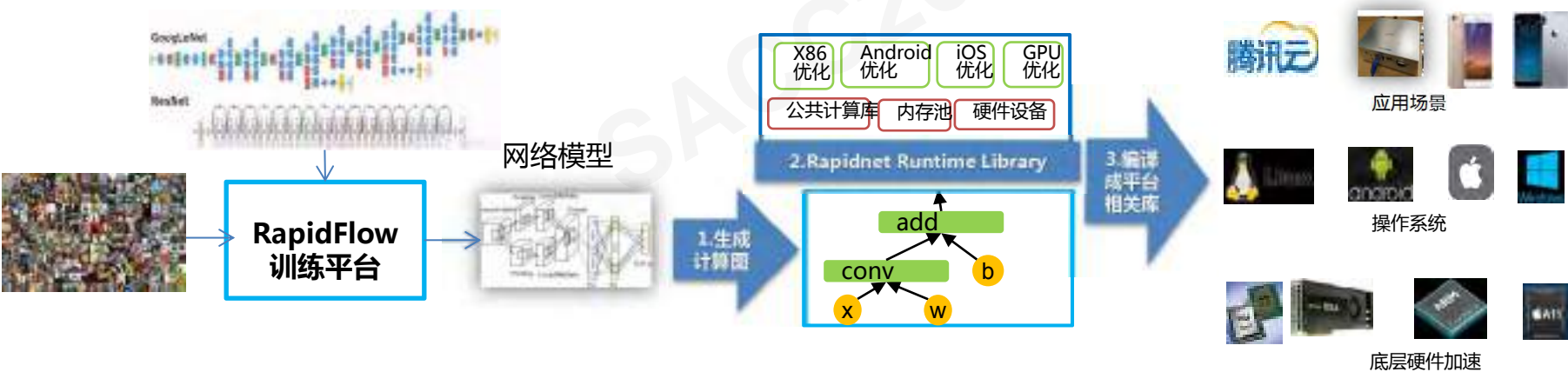
	Languages	Tutorials and training materials	CNN modeling capability	RNN modeling capability	Architecture: easy-to-use and modular front end	Speed	Multiple GPU support
Theano	Python, C++	++	++	++	+	++	+
Tensor-Flow	Python	+++	+++	++	+++	++	++
Torch	Lua, Python (new)	+	+++	++	++	+++	++
Caffe	C++	+	++		+	+	+
MXNet	R, Python, Julia, Scala	++	++	+	++	++	+++
Neon	Python	+	++	+	+	++	+
CNTK	C++	+	+	+++	+	++	+

预算多V100，预算少1080 TI

	特性	GTX - 1080TI	G7-P40	PCIe-V100
GPU核心	GPU微架构	Pascal	Pascal	Volta
	核心代号	GP104	GP102	GV100
	Tensor Cores	NA	NA	640
	CUDA核数量	3456	3840	5120
	处理器制程	-	16nm FinFET	12nm FinFET
	Core Clock(<=)	1621MHz	1531MHz	1450MHz
GPU显存	显存类型	GDDR5X	GDDR5	HBM2
	显存位宽	384-bit	384-bit	4096-bit
	显存带宽	480 GB/s	346 GB/s	900GB/s
	显存容量	12GB	24GB	16G
性能	FP32 (TFLOPS)	10.6	12	14
	FP16 (TFLOPS)	NA	NA	113
	Tensor (TFLOPS)	NA	NA	112
	TDP	250W	250W	250W

深度学习 – 打通训练和应用的闭环

- 越来越多的数据量，图像从百万到千万，数据从图像到视频
- 越来越复杂的网络结构，从Resnet，ResNeXt，DenseNet，DPN，SENET
- 越来越多的应用场景，云服务，Android，iOS，闸机嵌入式
- 越来越复杂的限制条件，内存，功耗，延迟



加快训练速度 - RapidFlow

任务监控与自动重启

- 分布式多机训练，不可避免遇到由于硬件/网络波动引起的异常
- 监控任务运行状况，当任务发生异常时，选择不同的重启策略

集群管理与监控

- 节点心跳异常告警
- 运维工具化，快速屏蔽/启动异常机器

灵活的资源分配

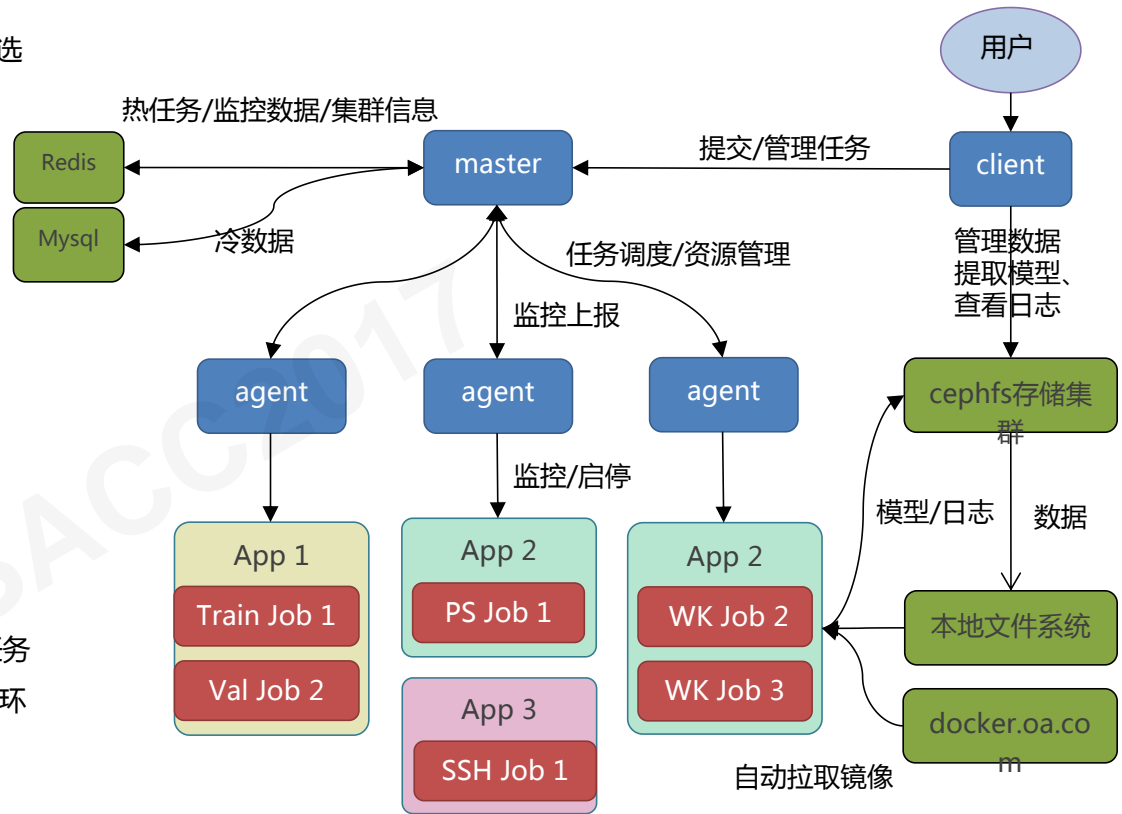
- 支持以 GPU 或节点为粒度进行资源分配
- 用户配置任务所需最小资源
- 自动扩缩容，最大化资源使用率

支持不同计算框架

- 调度与任务松耦合，用户可以灵活定义任务
- 支持配置 docker 镜像，完全自定义运行环境

良好的用户体验

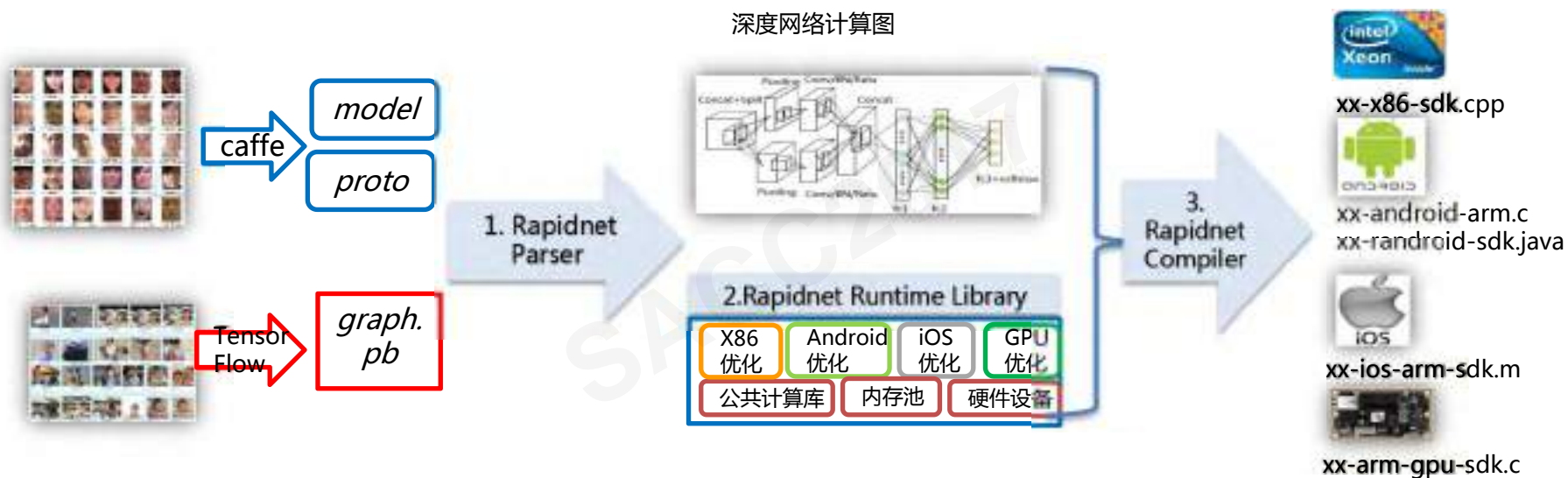
- 完善的客户端工具
- 任务进度微信提醒



加快应用速度 - RapidNet

Rapidnet : 深度网络应用的解决方案

- 将深度网络SDK生成，分为解析，编译，运行三个阶段
- 一键生成深度学习SDK，一个模型到处应用



Ncnn : 移动端前向网络开源框 <https://github.com/tencent/ncnn>

- 针对移动端优化版本
- 开源建设，2.6k+ stars



Deep
Eye

优图-腾讯云 天御内容识别解决方案

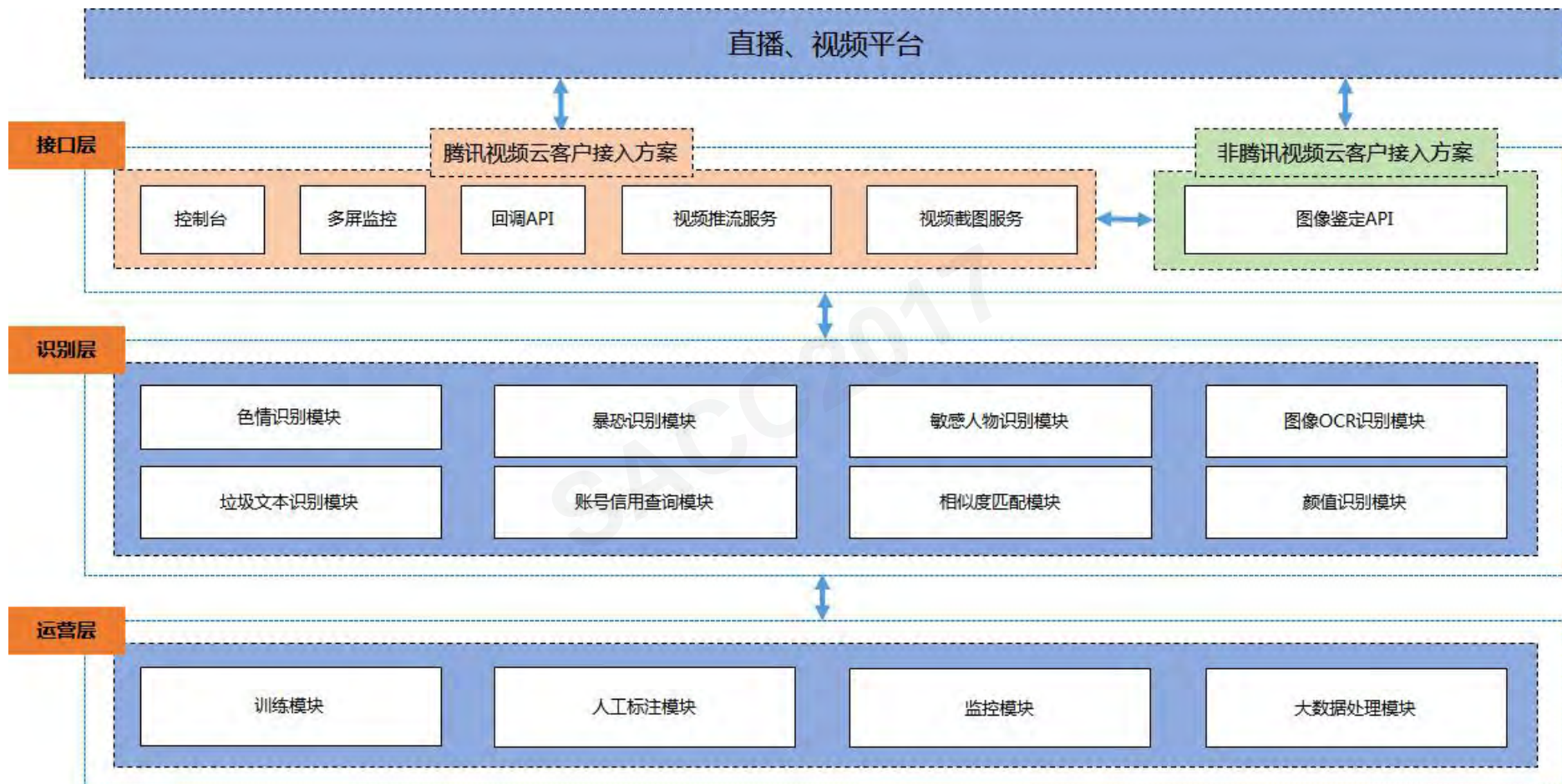


从静到动：结合视频识别能力



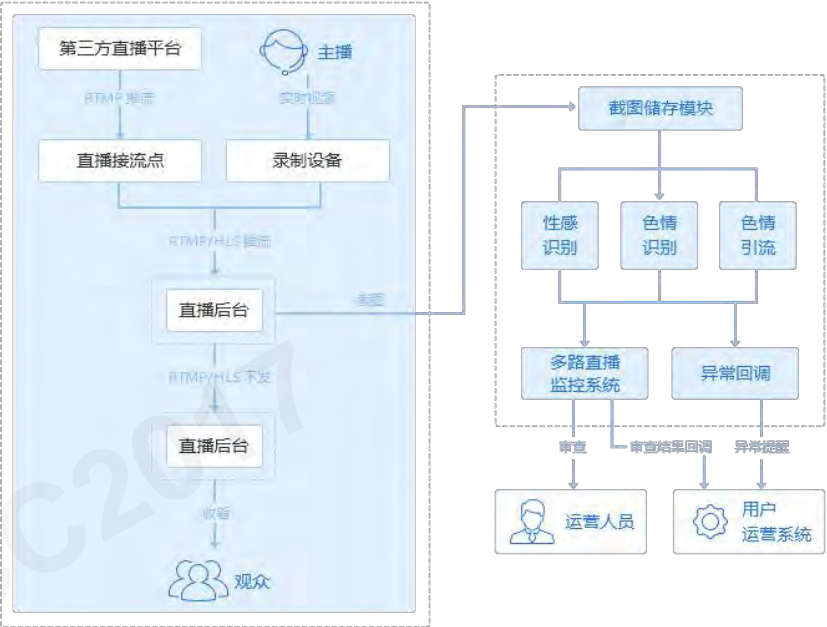
从图像到声音：音频识别

腾讯优图-腾讯云天御 内容审核解决方案



针对直播 – 视频鉴黄解决方案

- 在部署了DeepEye视频直播鉴黄解决方案后，系统对直播房间的视频流按指定的时间间隔（用户可配置）进行截图，通过鉴黄引擎给该图片进行鉴别，并将可疑图片和对应的房间信息回调给开发者，开发者可以根据返回的结果信息优先给审核人员进行审核，进行封停等进一步处理。经过审核没有问题的内容再呈现给观看者的屏幕。

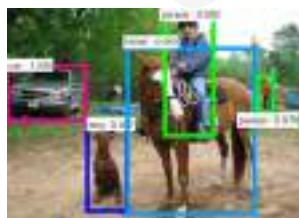


从静到动：结合视频识别能力

- 结合视频监控场景，在图像序列中，识别图像物体，识别人体，人群密度，人体属性等各种信息
- 融合图像，动作识别能力，提升审核准确率，召回率，增加更多场景风险控制



监控场景人体检测



多物体检测



监控场景人体属性

女性
青年
长发
背面
背包
灰色衣服
青色裤子



实际：
106人

预测：
113.4人

人群密度估计

从图像到声音：音频识别

声纹识别



小语种声音分类



QQ 音乐- 音乐检索



● 优图原音音频识别系统



腾讯优图 部分合作伙伴



云智未来^{9th}

第九届中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2017

Thanks & QA



腾讯优图 AI开放平台：<http://open.youtu.qq.com>

官方邮箱：youtu@tencent.com

腾讯优图公众号：腾讯优图

腾讯云-天御：<https://cloud.tencent.com/product/pf>