

七牛深度学习平台支持 亿级审核系统

李朝光 lichaguang@qiniu.com

极客时间VIP年卡

每天6元, 365天畅看全部技术实战课程

- 20余类硬技能, 培养多岗多能的混合型人才
- 全方位拆解业务实战案例, 快速提升开发效率
- 碎片化时间学习, 不占用大量工作、培训时间



Agenda



Overview



七牛云内容审核介绍



七牛云深度学习平台介绍



深度学习平台助力亿万级审核系统

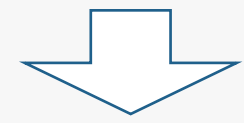


未来发展展望

人工智能在七牛

公有云接入

私有化部署



业务Portal , API接入



通用硬件 , 交换机 , 网络

Agenda



Overview



七牛云内容审核介绍



七牛云深度学习平台介绍



深度学习平台助力亿万级审核系统



未来发展展望

内容安全面临挑战

海量数据安全风险大、管理成本高

网络视频与直播用户已分别超过 5 亿和 4 亿，随之而来的是大量违规内容的产生和高企的管理成本。

内容安全已成为各大平台的生命线，只有在完善的管理和审核下，才能防范违规避免价值受损。

色情



当前无法显示该图像。

暴恐



当前无法显示该图像。

敏感



当前无法显示该图像。

低俗



当前无法显示该图像。

内容安全面临诸多挑战

审核效率低下

采用人审机制的 UGC 平台，审核时间长达 15-60 分钟。严重影响用户体验。

人力成本高

审核人员招聘、雇佣、培训、场地等费用带来主营业务外大量支出，成本动辄高达百万。



违规内容损害企业根本

政府执行最严内容监管标准，一旦违规，立即惩罚，甚至直接关停。负面舆论损害正常运营与企业形象。

审核标准滞后

最新审核标准难以第一时间在审核团队完全同步，带来违规内容审查疏漏，造成重大损害

七牛审核优势

当前
无法

对象存储云服务商
海量的图片、视频存储资源
完善的图片、视频处理服务

当前
无法显示
该图

完善的迭代流程
Dora - AVA - Lego结构化

当前
无法

基于高效AVA深度训练平台打造
直接基于云端数据训练模型
审核结果直接反馈到云端

当前

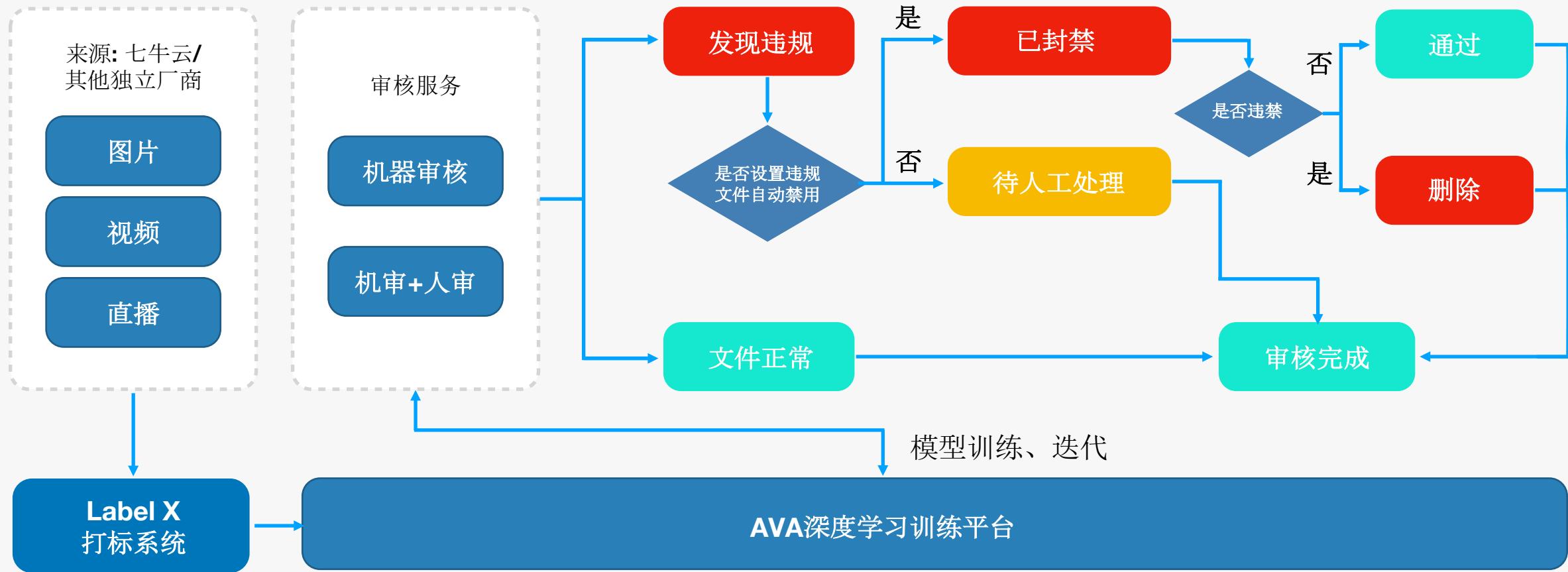
一键内容审核保护
鉴黄、鉴报、鉴恐，准确率高

审核流程

存储服务

审核、推理服务

审核服务（业务流程）



标注服务

训练服务

Agenda



Overview



七牛云内容审核介绍



七牛云深度学习平台介绍



深度学习平台助力亿万级审核系统



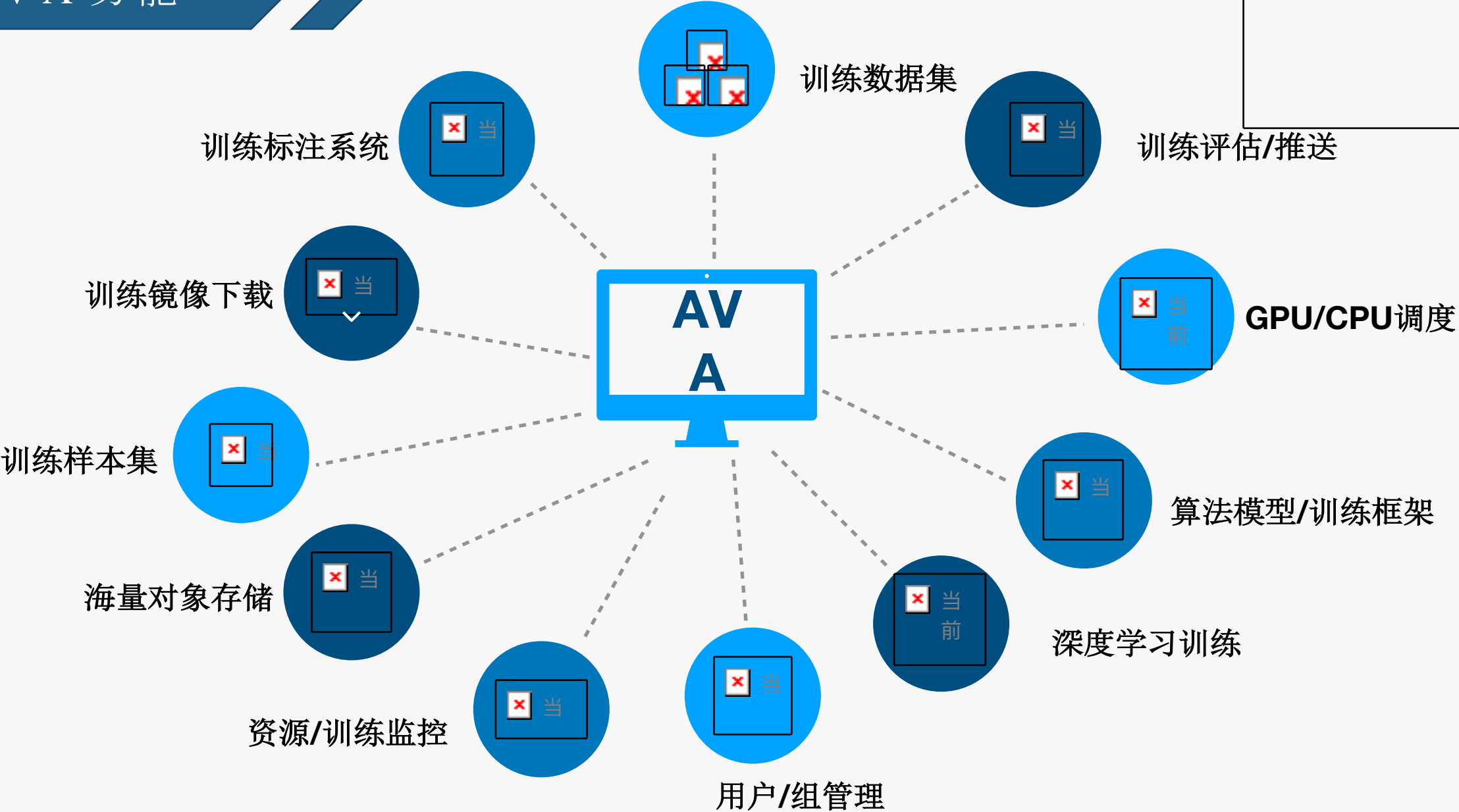
未来发展展望

AVA 概 览



当前无法显示该图像。

AVA 功能

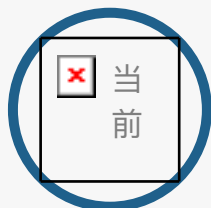


AVA 特点



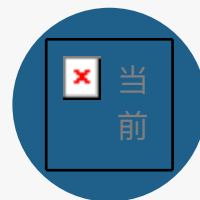
简单易用

大部分功能集成Portal页面，轻点鼠标即启动训练



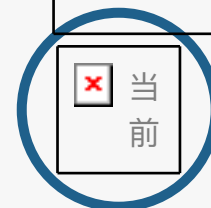
集成标注系统

Label X 打标系统，支持视频、图像分类、检测



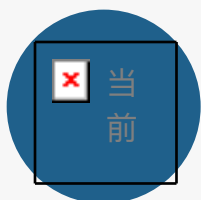
云上数据集管理

数据支持主要云运营商；云上数据轻松管理



模型管理方便

一键发布，支持快速迭代



灵活使用GPU资源

多机多卡分布式训练，多种调度策略



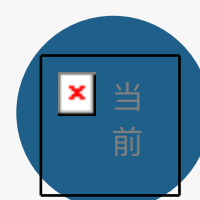
支持多种框架

MxNet、Pytorch、Caffe、Tensorflow



部署灵活

支持公、私有云部署，数据可不出岛



业务调度

业务task批量调度同步、异步高效调度

深度训练平台



Agenda



Overview



七牛云内容审核介绍



七牛云深度学习平台介绍



深度学习平台助力亿万级审核系统



未来发展展望

一组数据

- ◆ 平均日审核图片：1亿张，峰值2亿+
- ◆ 日均人工审核：10万+
- ◆ 个人日均审核图片：1200张
- ◆ 平均图像模型训练样本：800万张
- ◆ 模型平均训练时间：4天
- ◆ 训练平均需要GPU卡数：4张
- ◆ 对象存储bucket平均图片数：500万张
- ◆ 中型图像深度训练存储带宽：1GB/S

✖ 当前无法显示该图像。

✖ 当前无法显示该图像。

✖ 当前无法显示该图像。

✖ 当前无法显示该图像。

审核标注系统

 当前无法显示该图像。

 当前无法显示该图像。

标注过程和标注团队管理：

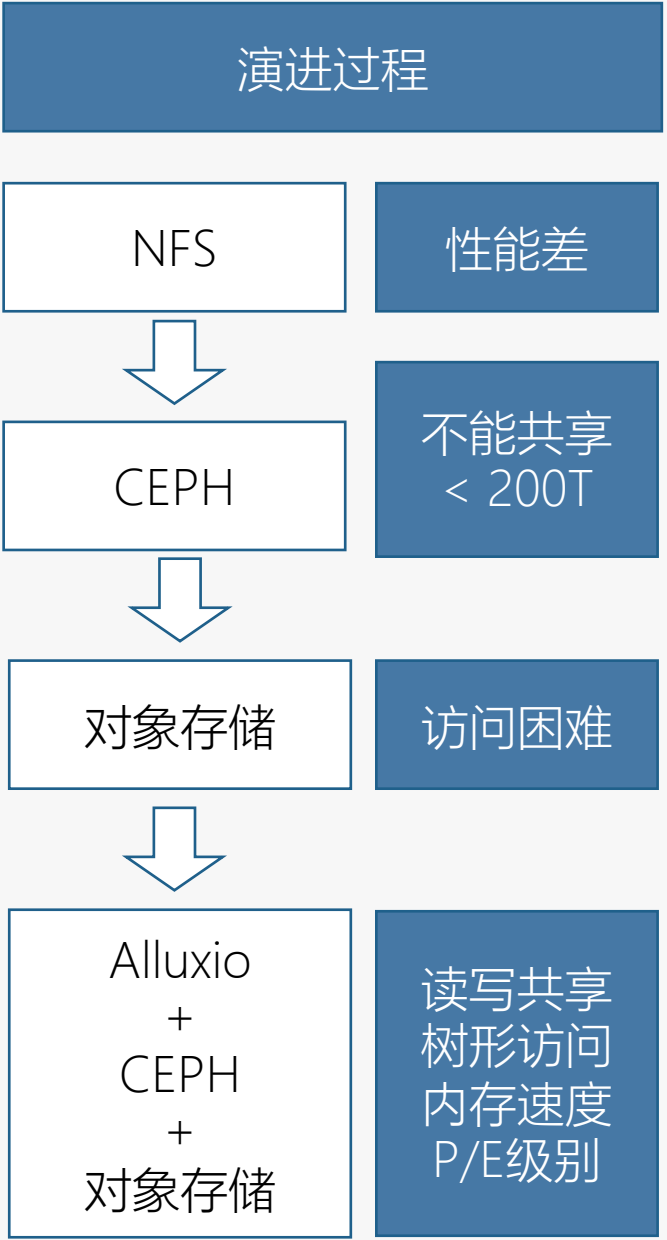
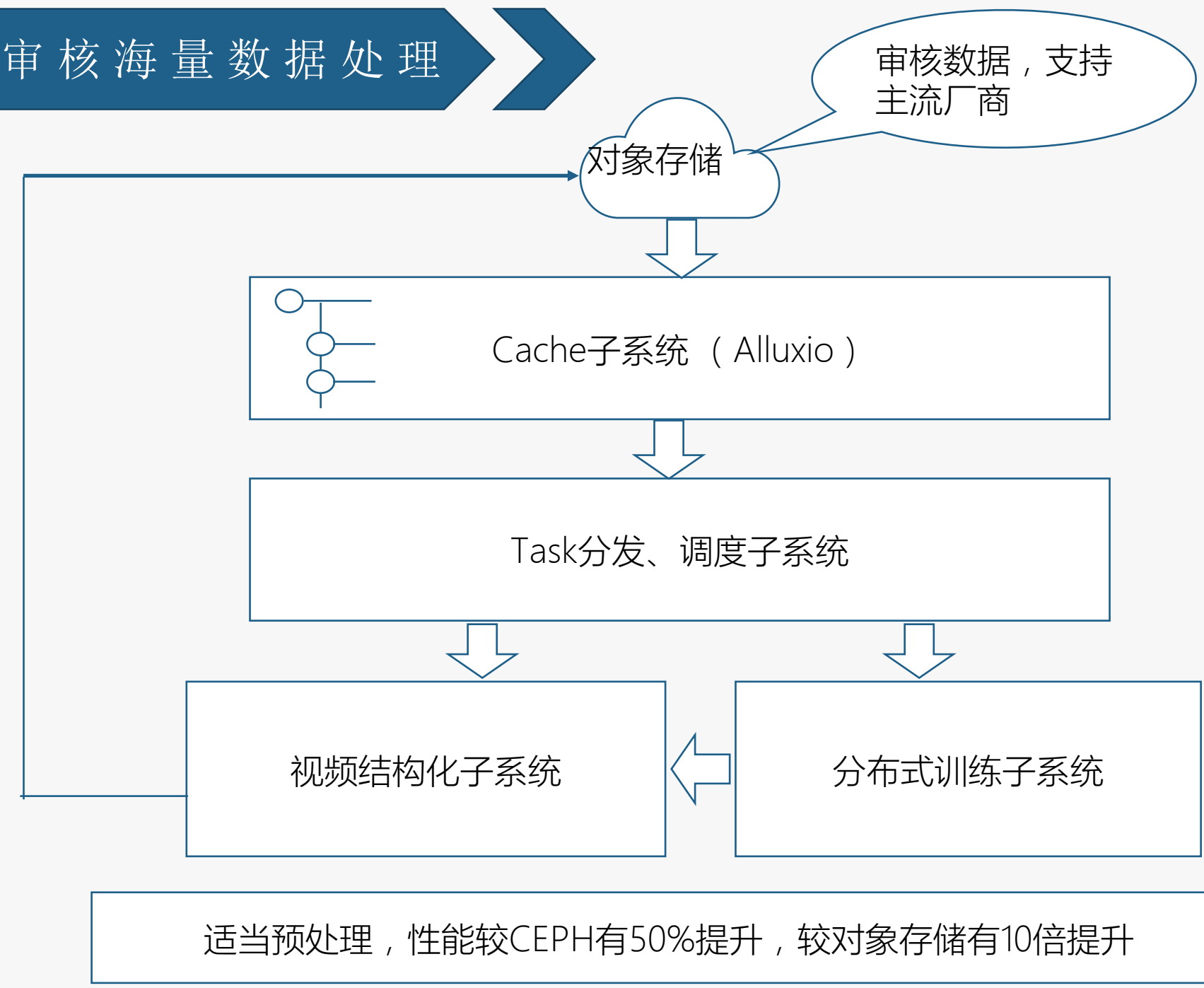
- * 分配标注任务
- * 查看标注进度
- * 接受、拒接标注结果
- * 创建标注数据集
- * 计费管理

 当前无法显示该图像。

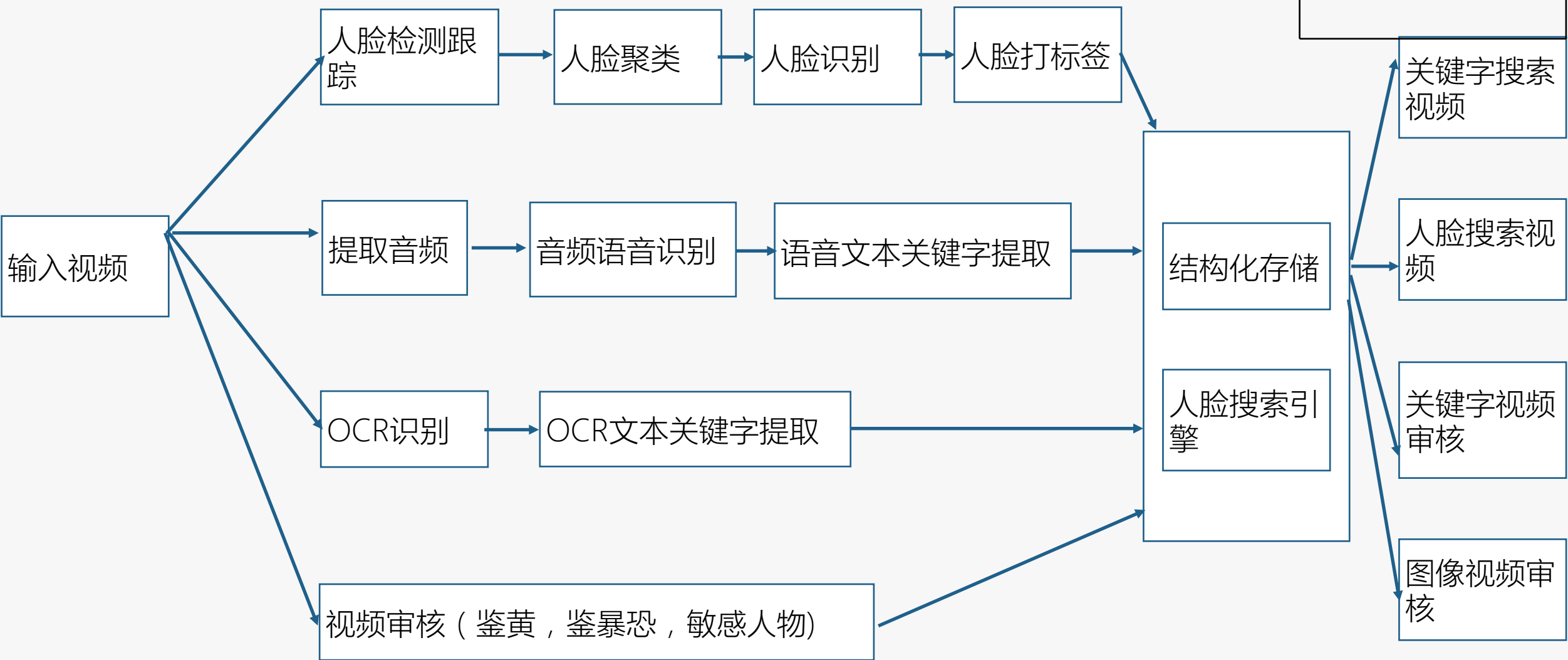
 当前无法显示该图像。

标注工具集

审核海量数据处理



视频结构化



```
graph LR
    subgraph Parallel_Task_Processing [并发任务处理]
        direction TB
        A[任务定义 json] --> B[任务生成]
        B --> C[(MQ)]
        C --> D[图片视频  
下载预热]
        D --> E[图片视频  
检测]
        E --> F[结果聚合]
    end

    subgraph Task_Scheduling [任务调度]
        direction TB
        G[Python SDK] --- H[任务编排]
        H --- I[任务调度  
查询]
        J[Orchestration  
client] --- K[Orchestration  
Server]
    end

    B --> D
    E --> F

    subgraph Task_Stack [任务栈]
        direction TB
        L[云对象存储]
        M[并发任务执行引擎]
        N[审核任务]
        O[审核任务]
        P[深度训练平台]
        Q[分布式调度系统]
        R[K8S, 硬件管理]
    end
```

并发任务处理

任务定义(json) → 任务生成 → MQ

任务生成 → 图片视频下载预热

图片视频下载预热 → 图片视频检测

图片视频检测 → 结果聚合

任务调度

Python SDK

SDK

任务编排

Orchestration client

任务调度查询

Orchestration Server

任务栈：

云对象存储

并发任务执行引擎

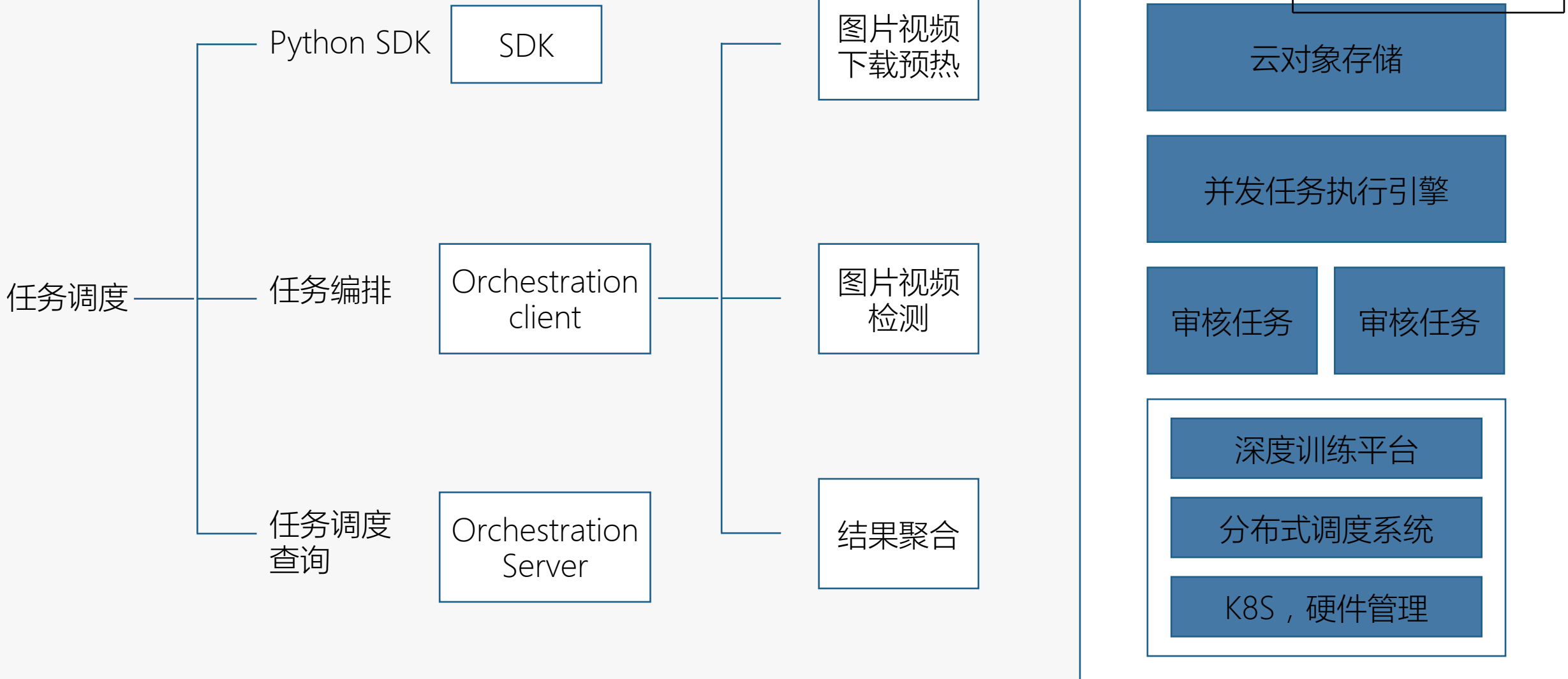
审核任务

审核任务

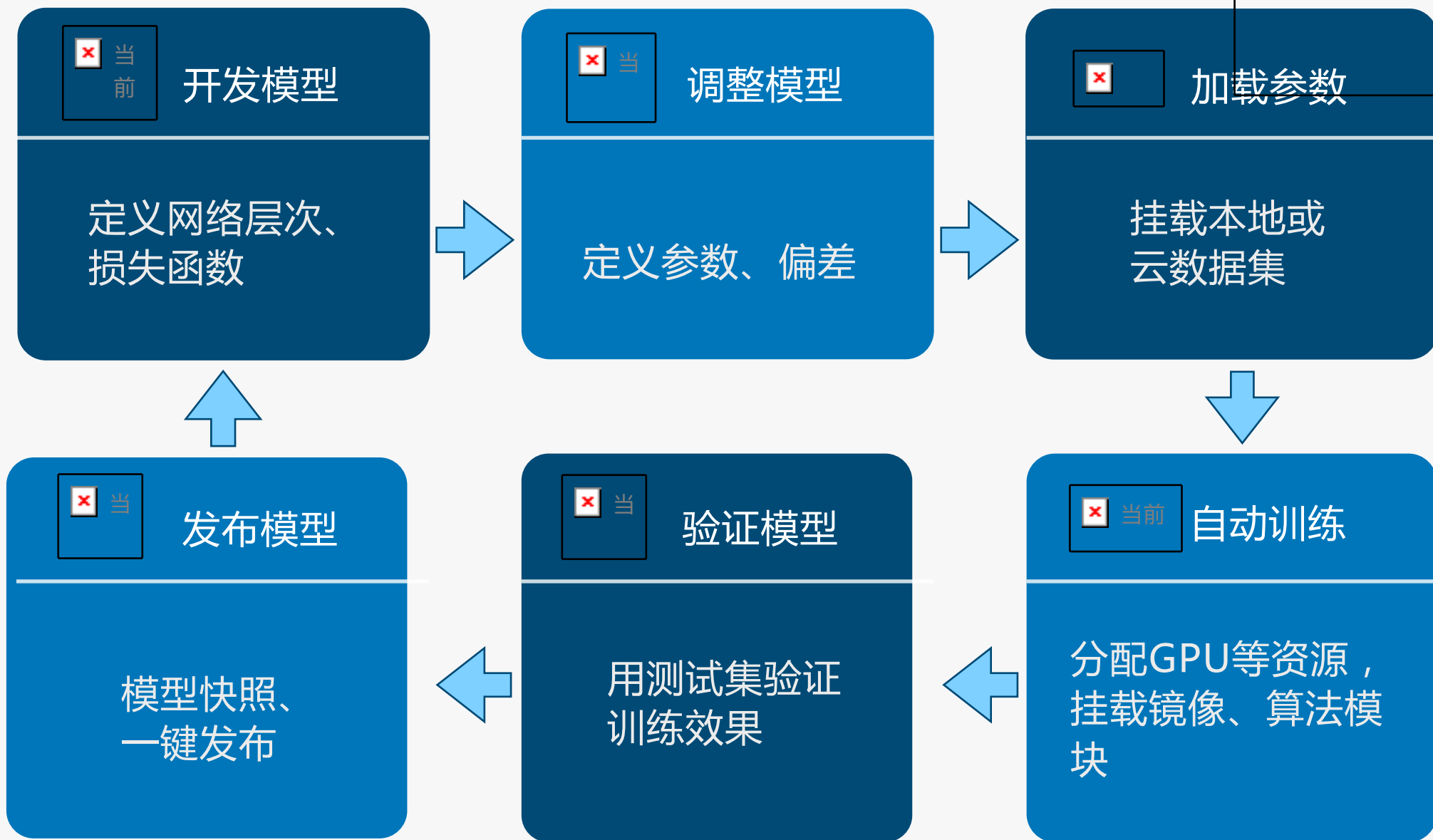
深度训练平台

分布式调度系统

K8S，硬件管理



完整开发流程

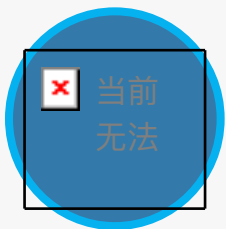


平台就近部署

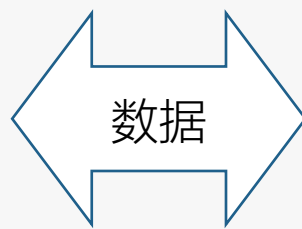
场景：

1. 应对业务突发
2. 存量数据不好迁移

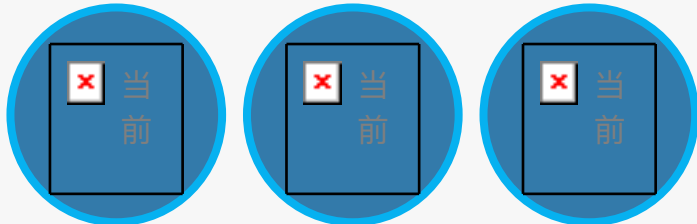
七牛云对象存储平台



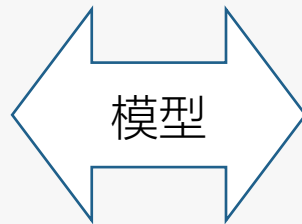
友商对象存储平台



七牛云审核系统



七牛云深度训练平台



七牛云深度训练平台



友商GPU运算力



内容审核服务

服务万级客户，日均活跃客户数千，日均调用量数亿

服务类型	服务类目	客户价值
智能机器审核	<div><div>图片鉴黄</div><div>图片鉴暴恐</div><div>图片鉴政</div><div>图片三鉴</div><div>视频鉴黄</div><div>视频鉴暴恐</div><div>视频鉴政</div></div>	降低人工审核工作量，提高及时性，节省人力成本。
智能机器审核 + 七牛云人工复核	<div><div>图片鉴黄</div><div>图片鉴暴恐</div><div>图片鉴政</div><div>图片三鉴</div><div>视频鉴黄</div><div>视频鉴暴恐</div><div>视频鉴政</div><div>人工复审</div></div>	减少审核团队的管理成本（例如，招聘、培训、绩效考核、人员管理和办公室成本）； 专业的人工审核员，及时处理违规图片视频。
全量托管服务	<div><div>图片鉴黄</div><div>图片鉴暴恐</div><div>图片鉴政</div><div>图片三鉴</div><div>视频鉴黄</div><div>视频鉴暴恐</div><div>视频鉴政</div><div>人工精审</div><div>文本过滤</div><div>语音审核</div><div>合规咨询</div><div>危急应对</div></div>	独家提供“全量托管服务”，提供文本、图片、视频、语音等内容的全量审核服务，并提供合规咨询和危急应对，让用户完全不用关心审核业务，彻底规避政策风险。

色情识别综合准确率行业领先

基于AVA深度学习平台不断快速迭代，优化算法，鉴黄产品无论是准确率还是召回率都要比竞品高。

	综合准确率	色情召回率	色情准确率	性感召回率	性感准确率	正常召回率	正常准确率
七牛云	96.48%	93.28%	96.52%	87.32%	65.66%	97.67%	98.75%
T 友商	91.82%	86.16%	93.56%	63.4%	34.86%	94.61%	97.33%
A 友商	93.12%	85.22%	87.52%	47.06%	50.44%	97.35%	96.45%
TP 友商	92.13%	74.34%	89.59%	38.76%	70.54%	98.89%	93.17%
B 友商	94.32%	87.17%	93.98%	58.1%	60.76%	97.8%	96.07%

支持最全暴恐类型

与上海网安成立联合实验室，支持更多贴合真实需求的鉴暴恐类型。

七牛云	血腥	爆炸	斩首	游行集会	打架斗殴	警民冲突	刀具	枪支	旗帜	敏感文字	军事场景	敏感着装	敏感蒙面
A 友商	血腥	爆炸	斩首	游行集会	打架斗殴	警民冲突	刀具	枪支	旗帜	敏感文字	军事场景	敏感着装	敏感蒙面
T 友商	血腥	爆炸	斩首	游行集会	打架斗殴	警民冲突	刀具	枪支	旗帜	敏感文字	军事场景	敏感着装	敏感蒙面
B 友商	血腥	爆炸	斩首	游行集会	打架斗殴	警民冲突	刀具	枪支	旗帜	敏感文字	军事场景	敏感着装	敏感蒙面
TP 友商	血腥	爆炸	斩首	游行集会	打架斗殴	警民冲突	刀具	枪支	旗帜	敏感文字	军事场景	敏感着装	敏感蒙面
WY 友商	血腥	爆炸	斩首	游行集会	打架斗殴	警民冲突	刀具	枪支	旗帜	敏感文字	军事场景	敏感着装	敏感蒙面
ST 友商	血腥	爆炸	斩首	游行集会	打架斗殴	警民冲突	刀具	枪支	旗帜	敏感文字	军事场景	敏感着装	敏感蒙面

政治敏感人物识别准

政治人物识别本质上是人脸识别，但是互联网上的人脸数据多种多样，图片各种清晰度和人脸角度增大了识别的难度。而七牛云采用云平台上的各种数据预处理的功能，预先处理图片，增加政治人物识别准确率。在近期浙江省公安厅、甘肃省公安厅的互联网使用场景中，我们多次完胜旷视政治敏感人物识别。



政治及敏感人物



准确率

Agenda



Overview



七牛云内容审核介绍



七牛云深度学习平台介绍



深度学习平台助力亿万级审核系统



未来发展展望

审核与平台发展规划

目标：

1. 日均10亿+审核量，服务客户百万+
2. 计算资源利用效率90%以上



视频审核
标注，海量存储、计算



图像、视频
审核算法优化



训练、迭代，推理
模型自动上线



端到端
调度、调试系统



异构资源使用
分布式、RDMA支持



多集群部署
互相打通

CNUTCon [上海]
全球运维技术大会

主办方 **Geekbang** **InfoQ**
极客邦科技

50+ 年末充电

开发&运维技术干货大盘点

容器

Kubernetes

DevOps

全链路压测

Severless

自动化运维

Service Mesh

Elasticsearch

微服务

使用折扣码 「QCon」 优惠报名 咨询电话：13269078023



扫码锁定席位

AiCon

2018.12.20-23 / 北京·国际会议中心

AI商业化下的技术演进实战干货分享

京东：智能金融

景驰科技：自动驾驶

阿里巴巴：NLP

清华人工智能研究院：机器学习

今日头条：机器学习

Twitter：搜索推荐

AWS：计算机视觉

Netflix：机器学习



扫码了解详情

Thank you !

