

# YY在直播场景的图像识别应用

梁炎/欢聚时代 ( YY Inc. )



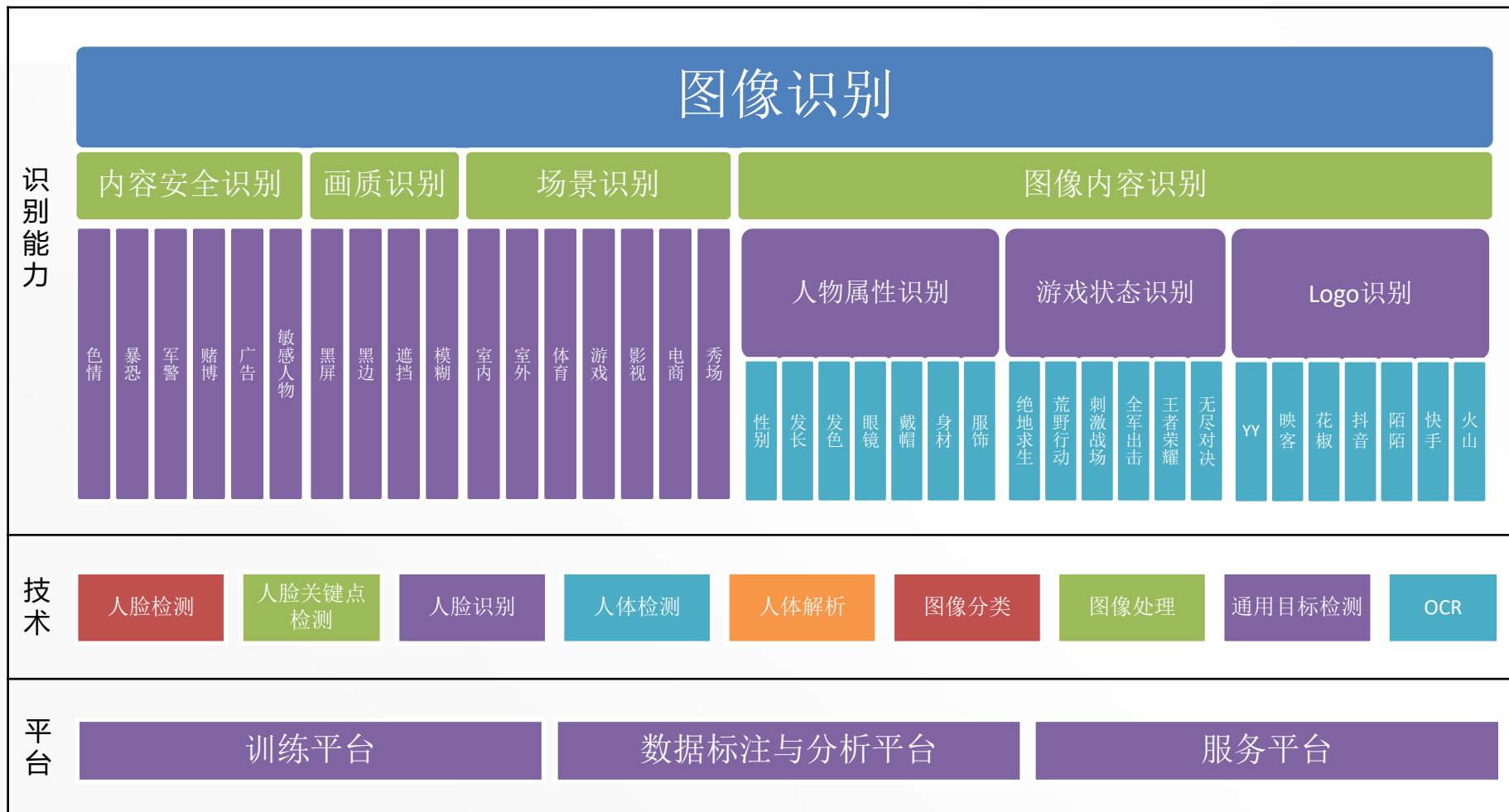
# 个人介绍

- 2002-2006 中山大学-数学与应用数学-学士
- 2006-2012 中山大学-应用数学-博士
  - 研究方向：图像超分辨率、人脸识别、子空间学习、稀疏表达
- 2013-2015 华为2012实验室中央研究院-计算机视觉工程师
  - 工作内容：人脸识别、图像质量评估、OCR等前沿技术预研
- 2015-现在 欢聚时代 ( YY Inc. ) AI技术部-机器学习工程师
  - 工作内容：图像内容分析与理解，并推动在YY直播各场景的应用

# 内容大纲

- YY图像识别应用概览
- 应用实例1——内容安全识别
- 应用实例2——图像文字识别
- 应用实例3——图像质量识别
- 应用实例4——人物属性识别
- 总结与展望

# YY图像识别应用概览



- 标签的丰富性——超过100个直播相关内容标签
- 结果的准确性——识别准确率 ( Accuracy ) > 95%
- 推理的实时性——耗时 < 10ms
- 服务的可靠性——每日不间断响应，总处理量达十亿级别
- 训练的及时性——最快模型迭代周期 < 1天

➤ 训练数据量大

| Dataset    | Images |
|------------|--------|
| LFW        | 13233  |
| WDRRef     | 99773  |
| CelebFaces | 202599 |
| VGG        | 2.6M   |
| FaceBook   | 4.4M   |
| Google     | 200M   |

➤ 训练时间长

| GPU    | network   | Batch size | Samples/s | Total times (hours) |
|--------|-----------|------------|-----------|---------------------|
| K80*3  | Resnet-50 | 256        | 205       | 160                 |
| V100*8 | Resnet-50 | 256        | 1155      | 29                  |

数据来自mxnet官方

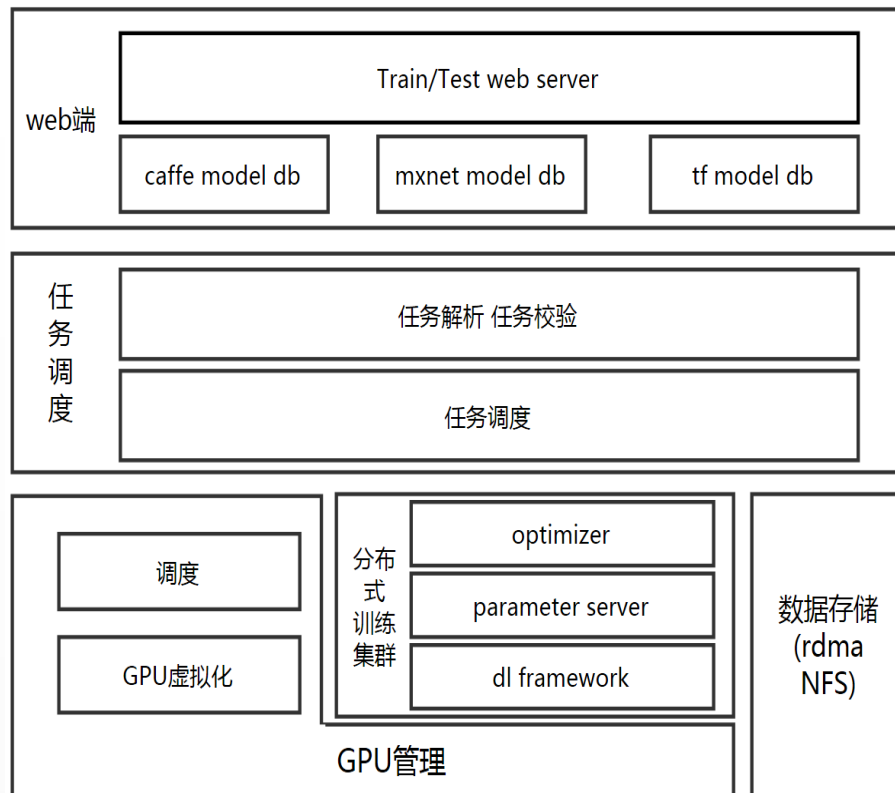


➤ 现有主流框架的分布式能力

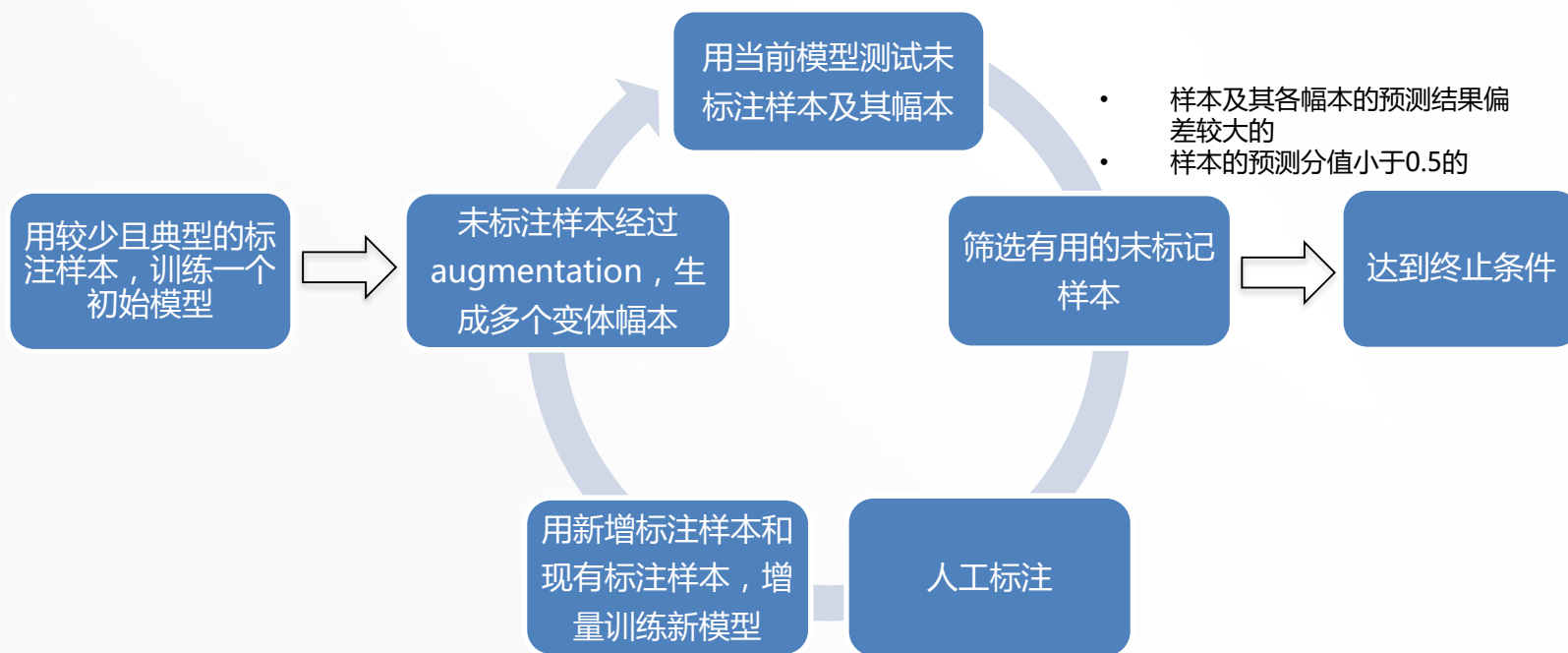
| Framework  | 分布式 | 同步/<br>异步 | Overlap | 梯度压缩 | RDMA |
|------------|-----|-----------|---------|------|------|
| caffe      | x   | x         | x       | x    | x    |
| caffe2     | √   | √         | √       | x    | √    |
| mxnet      | √   | √         | √       | √    | x    |
| tensorflow | √   | √         | √       | x    | √    |

## ➤ 分布式训练平台架构

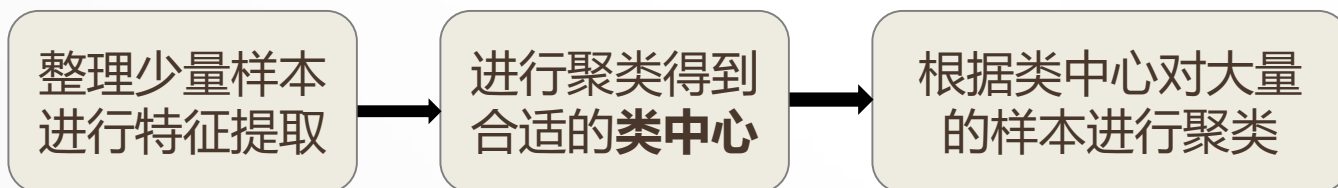
- 统一接口
- 统一训练数据和模型管理
- 通用的parameter-server组件
- 基于nccl库实现两级通信
- 基于RDMA的通信改造
- 半精度训练和梯度压缩



## ➤ 主动学习

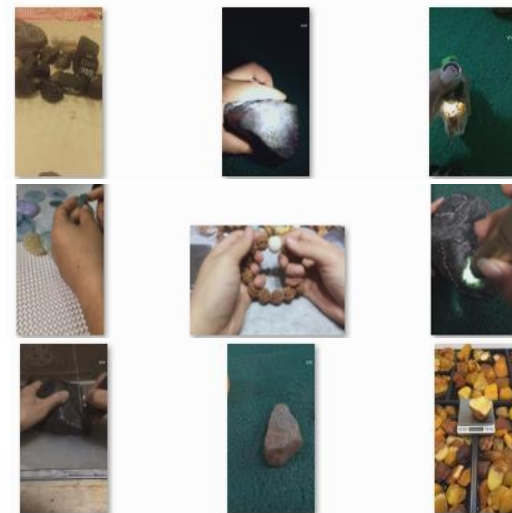


## ➤ 聚类分析

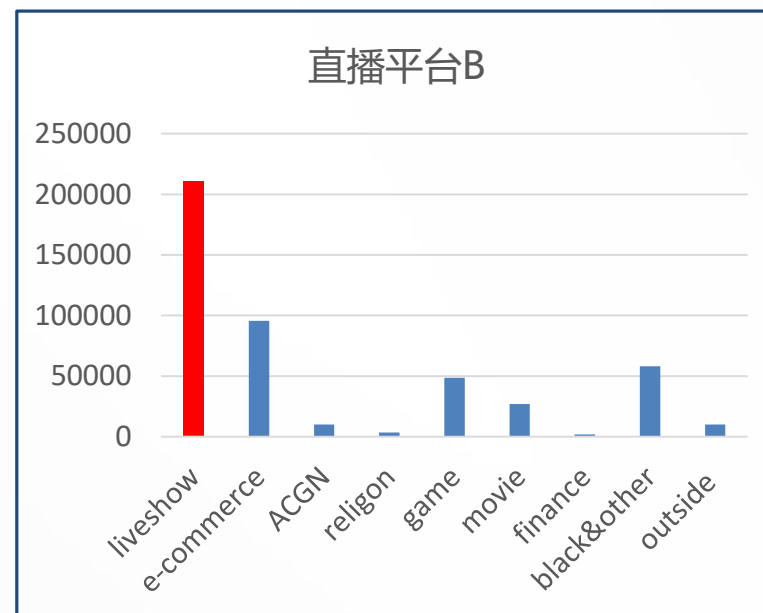
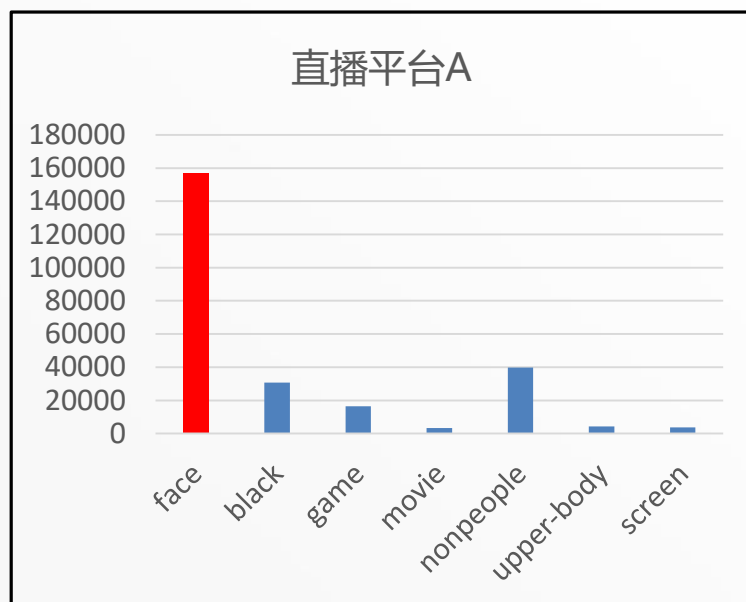


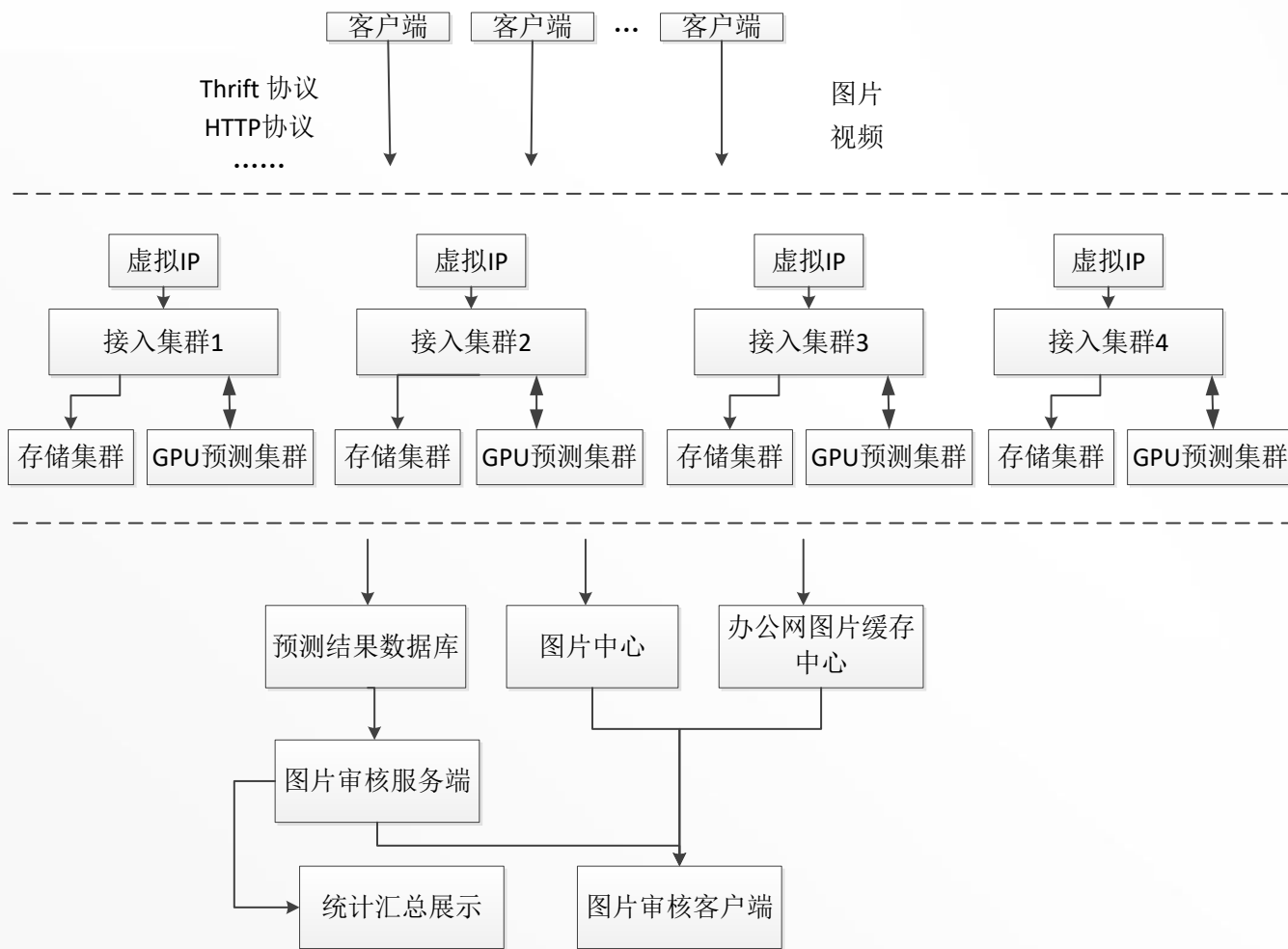
- 经过CNN模型提取的图像特征含有丰富的语义信息，采用模型输出的特征进行聚类可以得到较好的同类合并效果
- 聚类后可以清晰了解样本类型以及分布情况，便于后续分析出现的问题
- 通过聚类算法对未打标签的数据可以进行快速整理，帮助加速人工标注

## ➤ 采用聚类算法后的类别示意图



## ➤ 聚类后对样本分布的分析例子





应用实例-1

# 内容安全识别



## ➤ 色情



## ➤ 暴恐

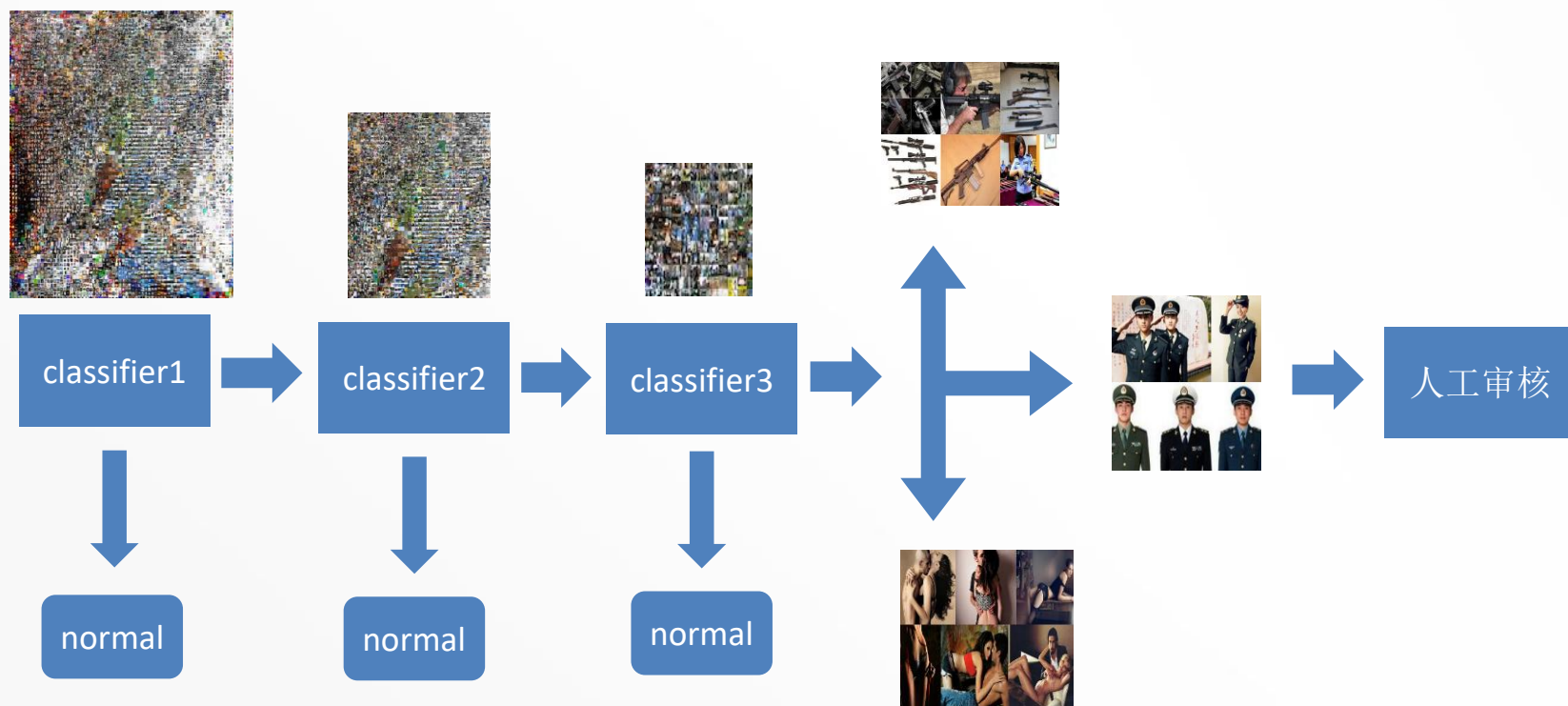


## ➤ 其他（包括宗教、玉石、游戏等等）



● ● ●

## ➤ 内容安全识别的基本流程

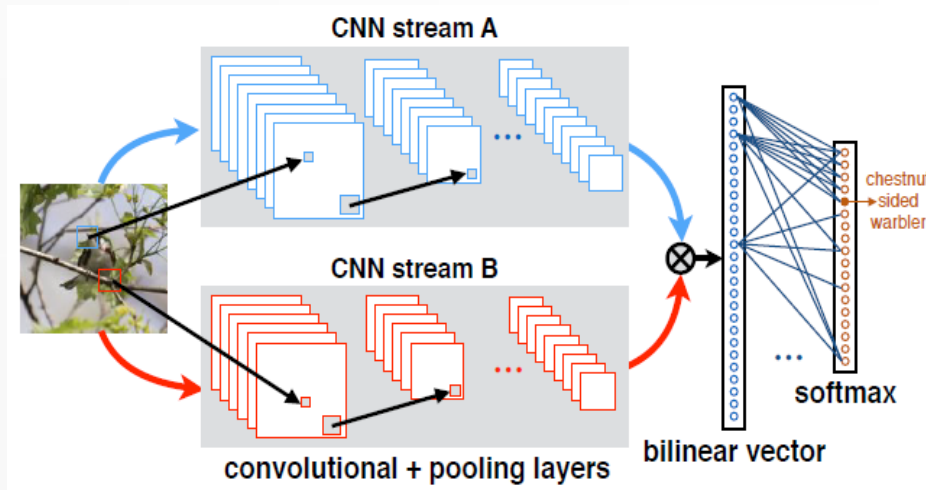


## ➤ 内容安全识别的挑战

- 鉴图量大
  - 密集截图，每天超过3亿鉴图量级，每秒平均鉴图超过4000张
- 类型众多
  - 色情、血腥、暴力、广告、标语.....
- 场景复杂
  - 覆盖所有直播场景，包括成像质量差、角度变化大、画中画、违规区域占比过小等
- 标准难把握
  - 擦边球

## ➤ 双线性CNN模型

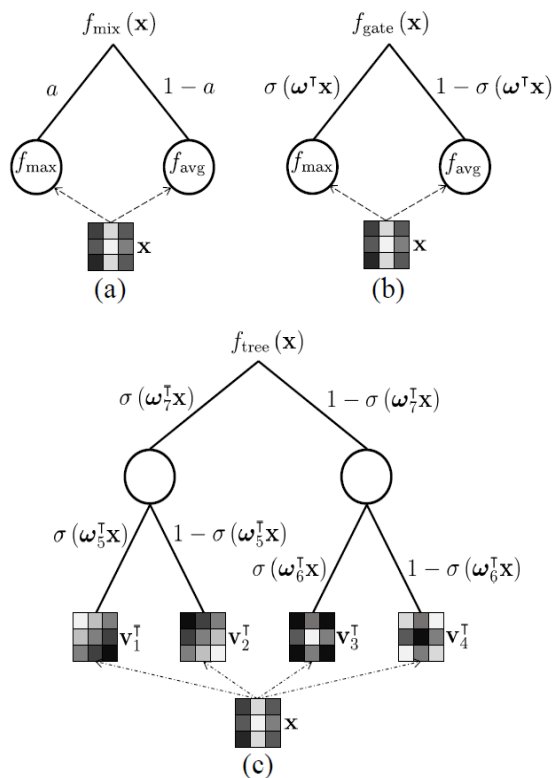
- 包含两个特征提取器，其输出经过外积相乘、池化后获得图像描述子
- 优点：
  - 该架构能够以平移不变的方式，对局部的对级（pairwise）特征交互进行建模，适用于细粒度分类
  - 能够泛化顺序无关的特征描述子,增加了特征的非线性，将低维特征映射到更高维，丰富了特征表达
- 缺点：
  - 模型的得到的特征维度较高，那么得到的参数数目较多，计算量较大，存储和读取开销较大



[ Bilinear cnn models for fine-grained visual recognition ]

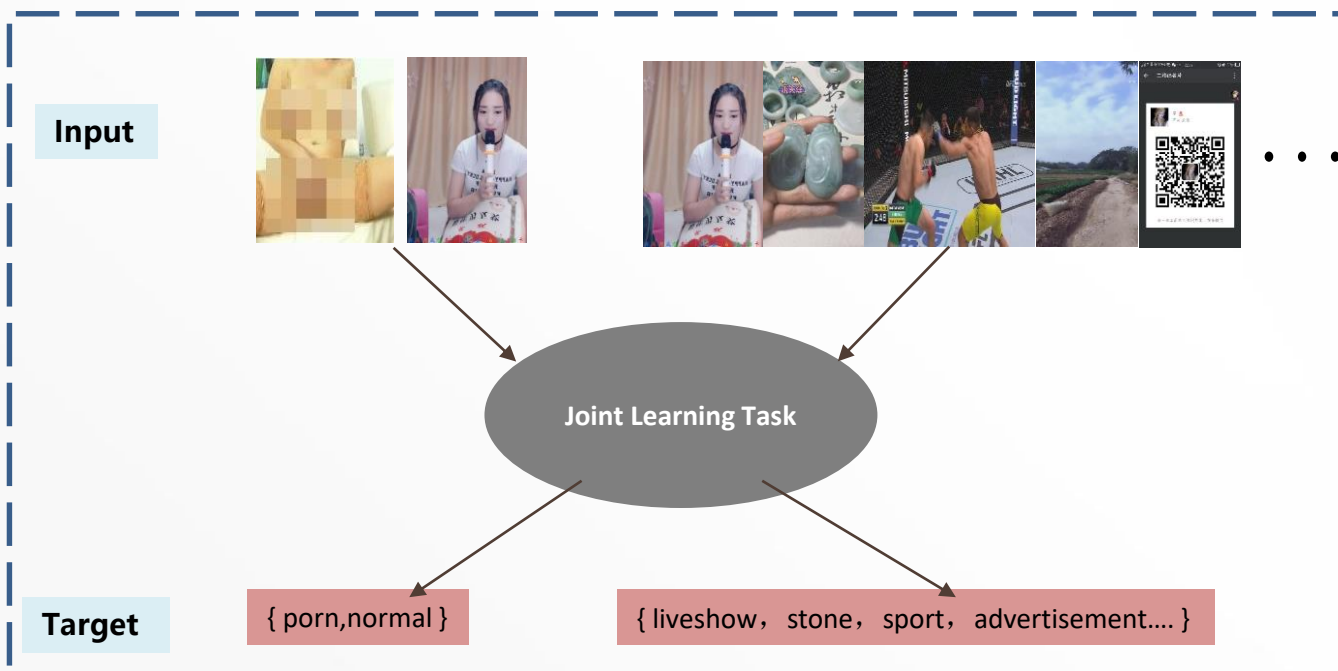
## ➤ Generalizing Pooling Functions

- 常规使用到的pooling方法有Max、Avg Pooling
- 将pooling层也加入参数学习，学习一种更加适合相应任务的pooling方法
- Gate\_pooling: 对输入进行卷积，归一化后产生权重，对max、avg pooling的结果进行加权输出
- Tree\_pooling: 用两个不同的卷积替换掉 Gate\_pooling方法中的max和avg pooling，右图c, tree为3层的示意
- 将部分pooling层替换为右侧几种pooling的结构，对最终的效果有小幅提升

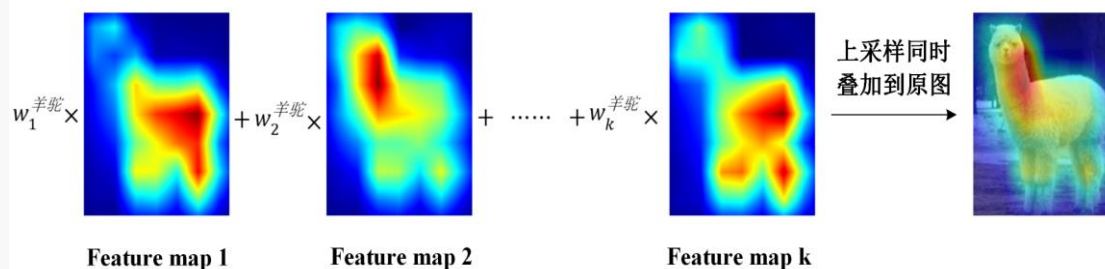
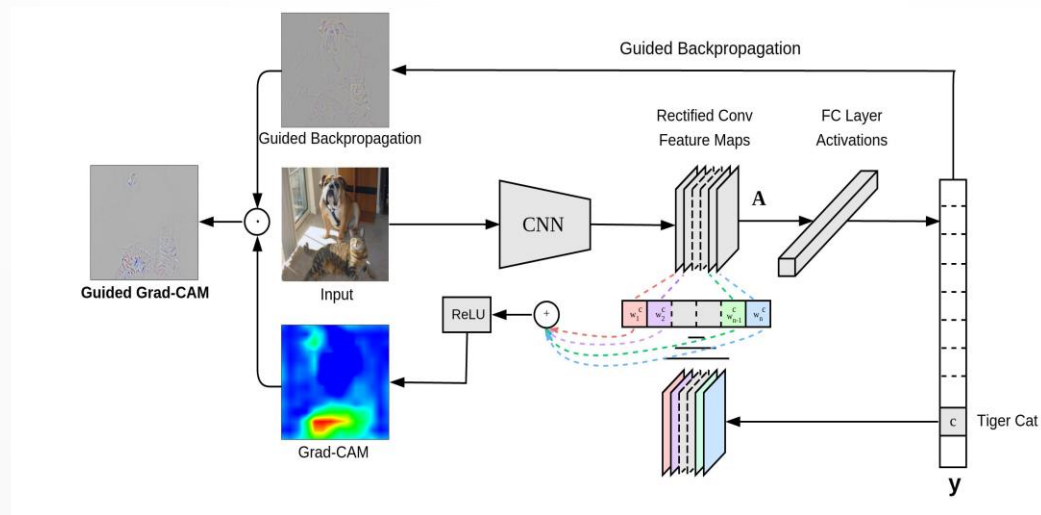


## ➤ 多任务学习

- 迁移学习算法的一种，其基本目标是提高泛化性能
- 通过相关任务训练的特定信息来提高泛化能力，利用共享的网络并行训练，学习多个任务



## ➤ 模型可视化



[ Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization ]



## ➤ 弱监督定位

## ➤ 小目标二次识别

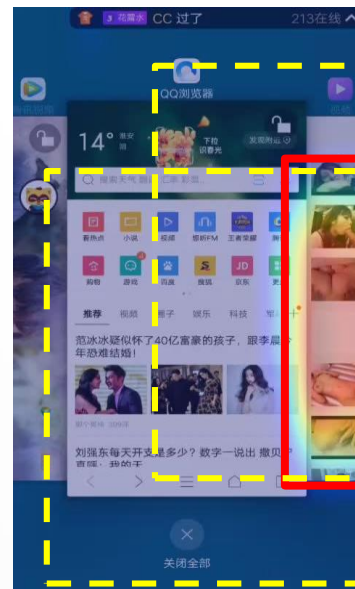
- 对于小的违规目标，因其范围较小，总权值有可能小于正常类特征总权值（区域较大），但热图定位能定位，故采用二次识别方法
- 先基于热图定位到可能违规区域，对包含框放大1.5倍，裁剪下来，送给模型进行二次识别，可将较小目标召回





## ➤ 增强裁剪

- 预训练一个分类模型对违规类进行识别，使用热图定位方法，计算出热图区域，归一化后，设定一阈值进行二值化处理，得到最大包含矩形，从而得到某类别图像的矩形框
- 将此框坐标记到文本，修改训练代码，使其在随机比例裁剪时包含此区域，避免太随机而造成的违规图像内容丢失问题

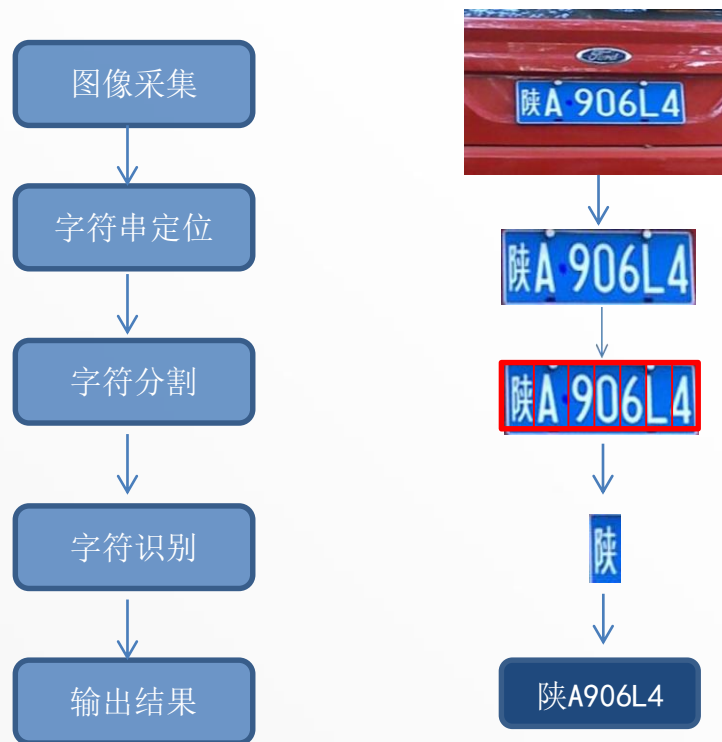


应用实例-2

# 图像文字识别

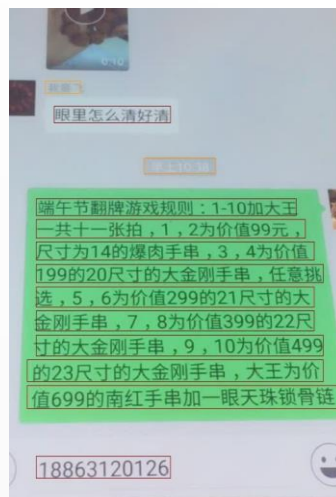
### ➤ 特定场景的文字识别

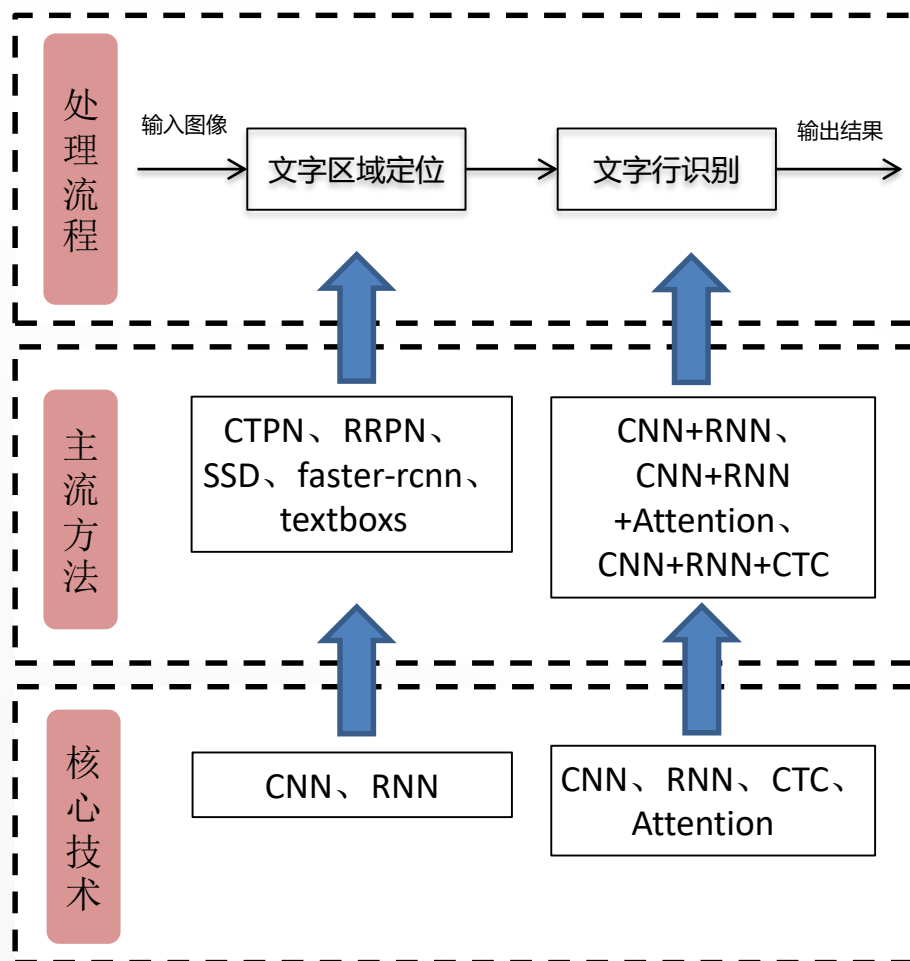
- 包括身份证、车牌、钞票号码、银行卡、驾驶证、行驶证、营业执照、彩票、发票、拍题等特定场景下的文字识别



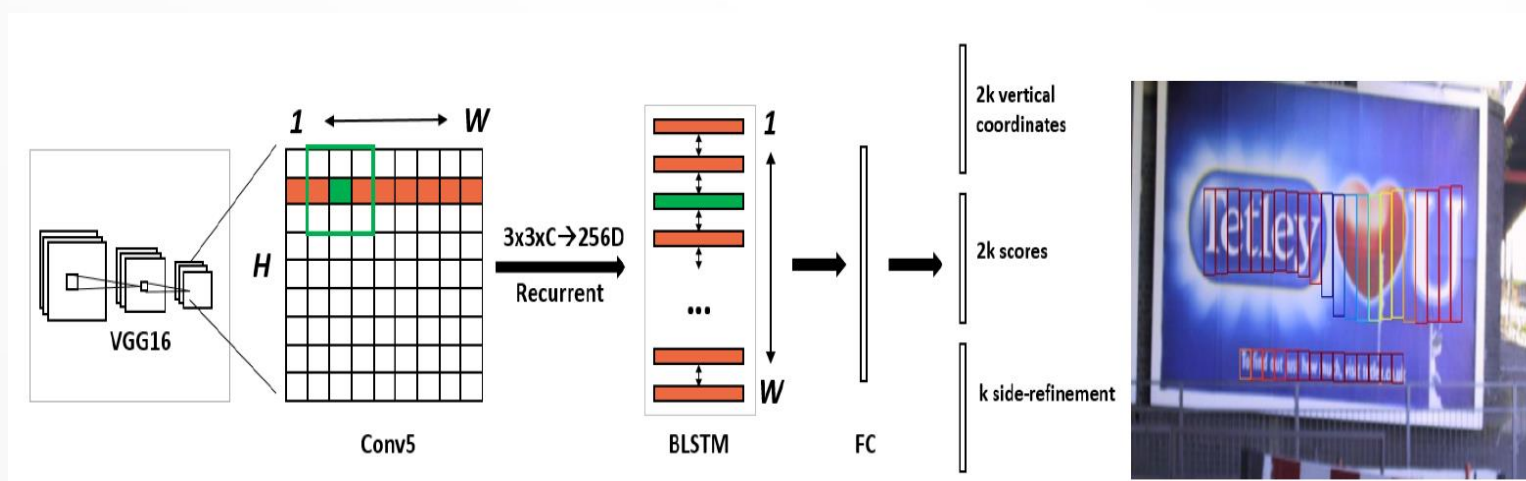
### ➤ 任意场景的文字识别

- 目标图像来源不可控，属于非配合条件下采集的图像
- 目标图像文字的字体、大小、颜色、排版、字数等特征不可控
- 目标图像文字背景不可控



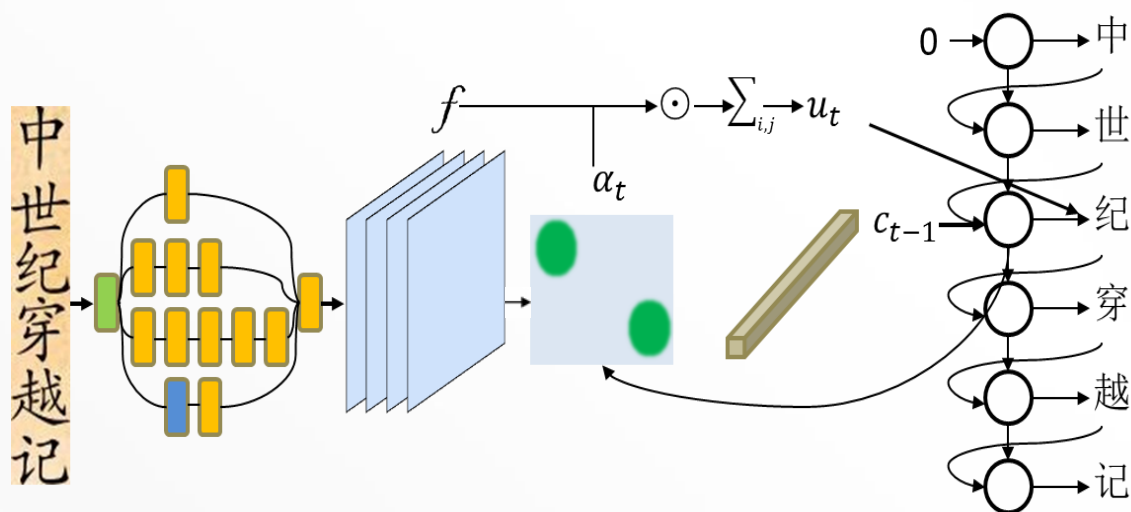


### ➤ 文字区域定位



[ Detecting Text in Natural Image with Connectionist Text Proposal Network ]

## ➤ 文本行识别





### ➤ 自然场景样本生成







我是盛树宝-业务经  
我为欢乐支付代言

识别二维码,手机微信秒  
变POS机,一清秒到账,  
银行监管,安全快捷,注  
册后实名认证即可免费使  
用,刷卡养卡不求人,自  
用省钱分享赚钱,联系电  
话:13764527008

回三口  
面一二

长按识别二维码进入公众号

全国客服热线:0591-87519992

我是盛树宝-业务经  
我为欢乐支付代言  
我为欢乐支付代言

话:13764527008  
话:13764527008  
长按识别进入公众号  
长按识别进入公众号

识别二维码,手机微信秒  
变POS机,一清秒到账,  
变POS机,一清秒到账,  
银行监管,安全快捷,注  
册后实名认证即可免费使  
用,刷卡养卡不求人,自  
用省钱分享赚钱,联系电  
用省钱分享赚钱,联系电

全国客服热线:0591-87519992  
全国客服热线:0591-87519992

## ➤ 游戏状态识别



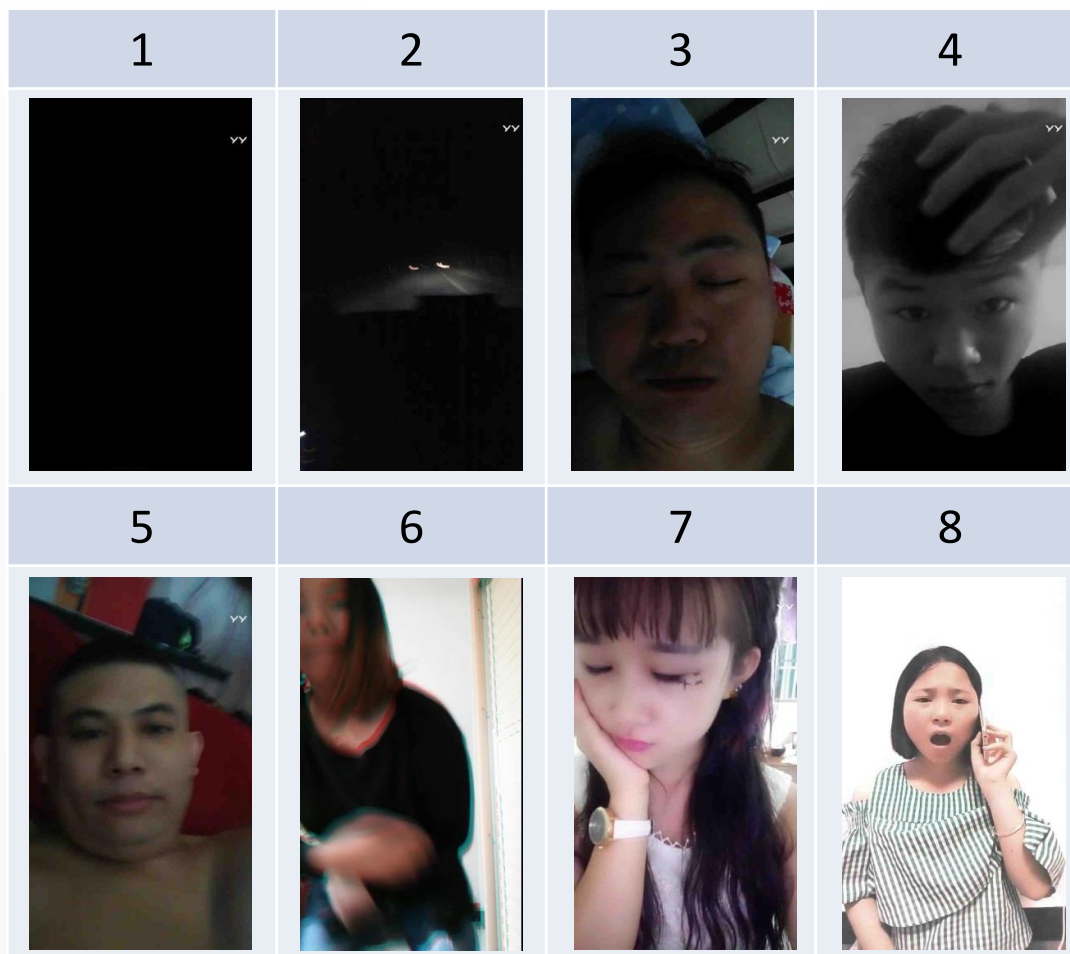
|       |        |
|-------|--------|
| 杀人    | 5      |
| 存活    | 18     |
| 枪械    | SCAR-L |
| 剩余    | 4      |
| 行进    | 3490   |
| 物资    | 30     |
| 击杀    | 2      |
| 死亡    | 7      |
| 助攻    | 7      |
| ..... | .....  |



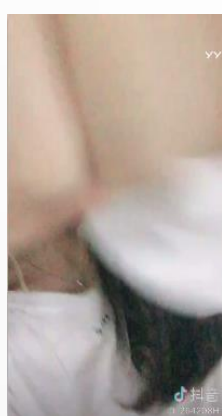
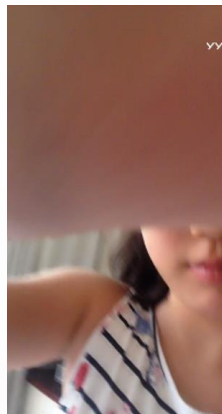
应用实例-3

# 图像质量识别

## ➤ 亮度识别

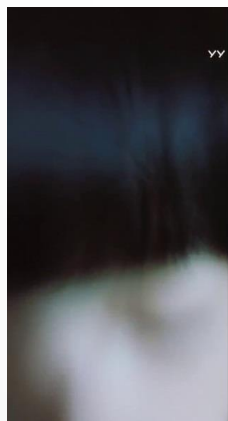
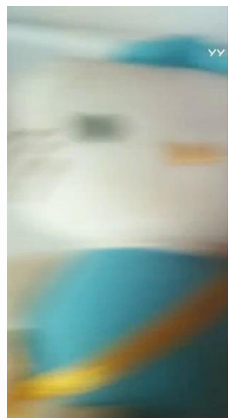
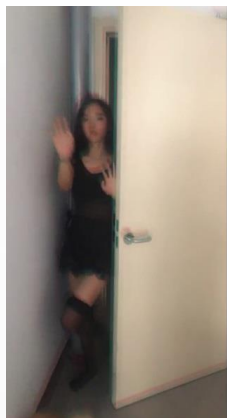


### ➤ 遮挡识别





### ➤ 模糊识别



## 图像退化形式多样



原图



模糊



局部模糊

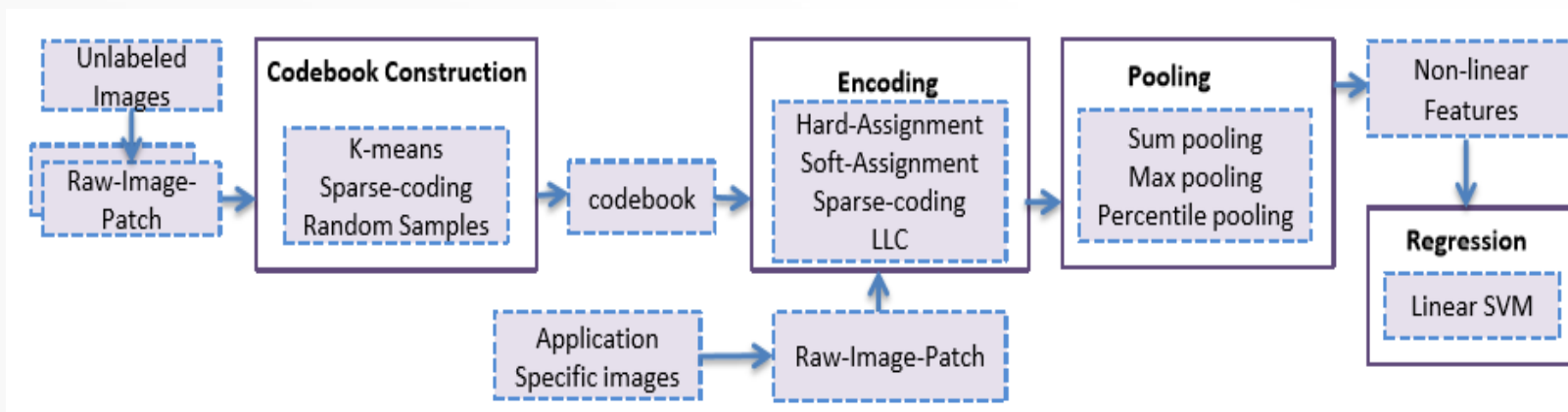


压缩失真



噪声

- 传统基于梯度、频率、熵等方法难以处理各种退化情况
- 基于特征学习的无参考图像质量评估



### ➤ 开源数据集

- LIVE IQA database: A statistical evaluation of recent full reference image quality assessment algorithms. 2006.
- TID2008 database: Tid2008 - a database for evaluation of full-reference visual quality assessment metrics. 2009.
- CSIQ database: Most apparent distortion: full-reference image quality assessment and the role of strategy. 2010.

[ Unsupervised Feature Learning Framework for No-reference Image Quality Assessment ]

应用实例-4

# 人物属性识别



不戴眼镜

长直发

女性

斜刘海

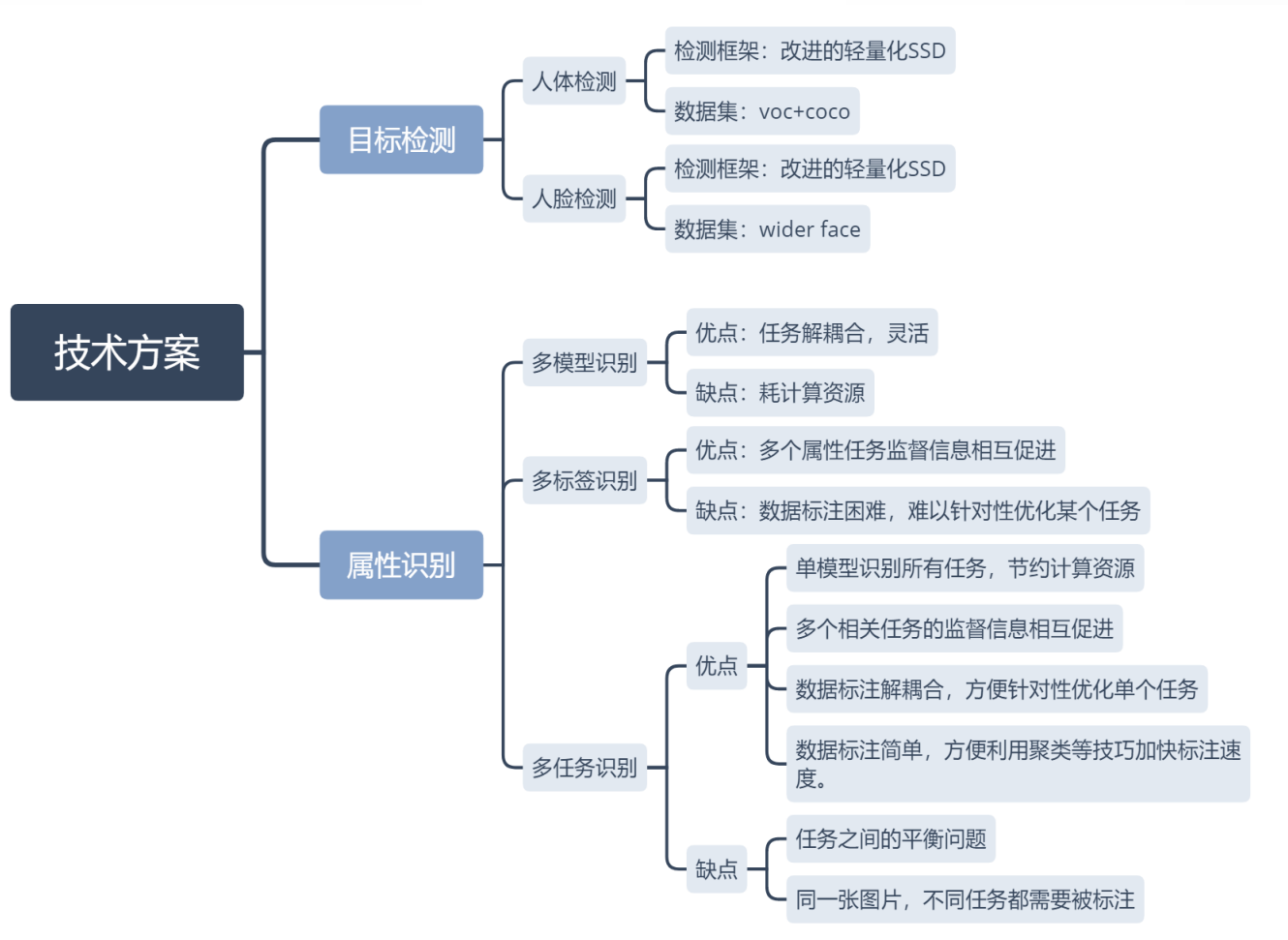
不戴帽

苗条

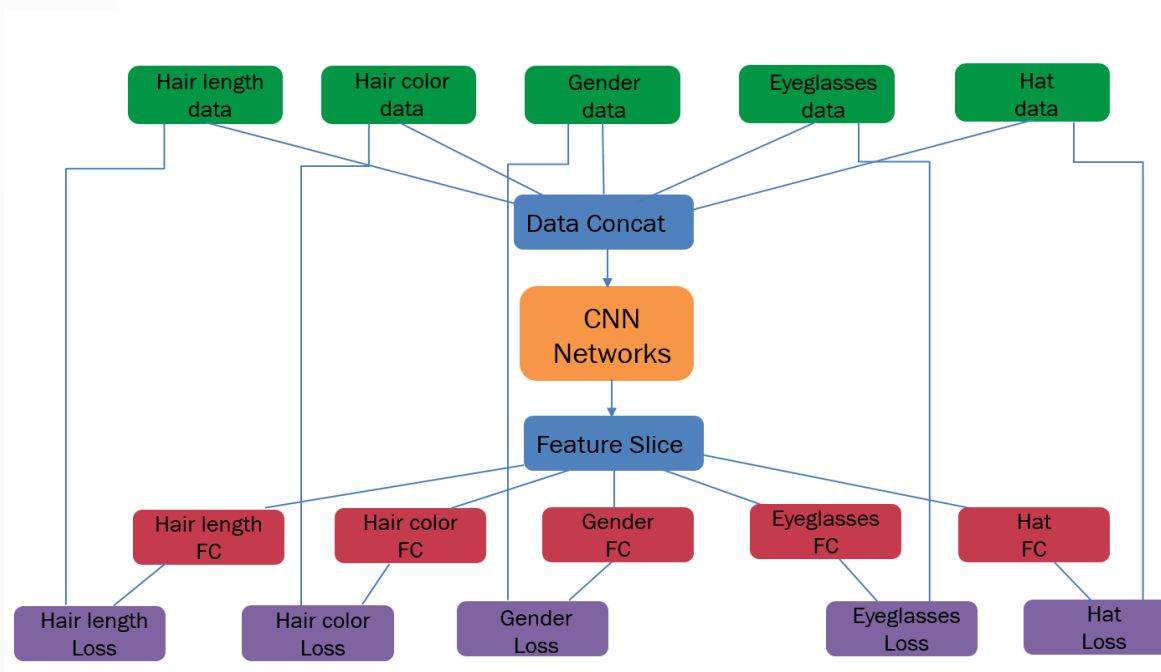


棕色头发

T恤

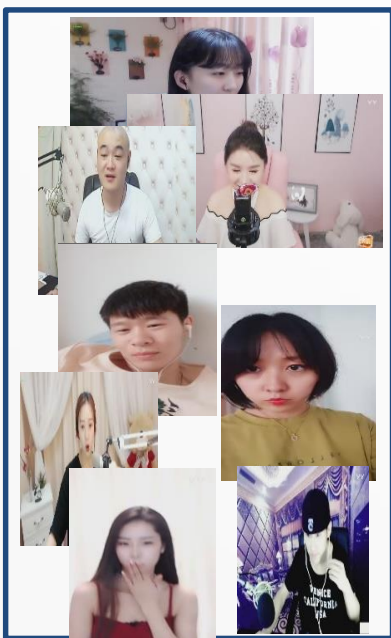


## ➤ 网络结构

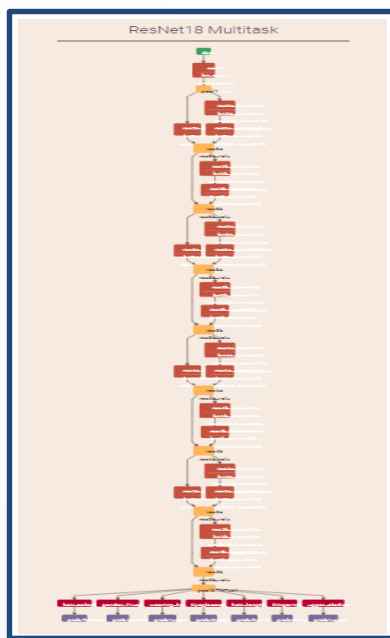


## ➤ 人物属性识别基本流程

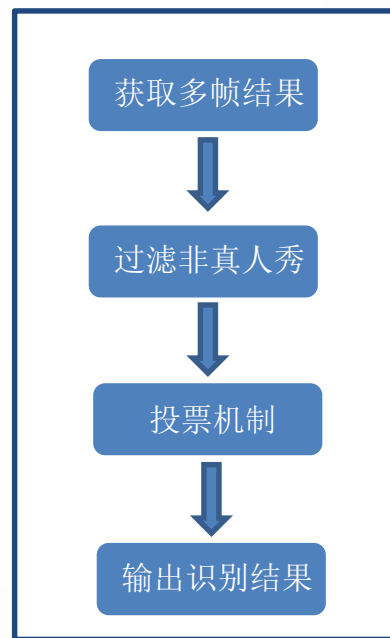
线上主播截图



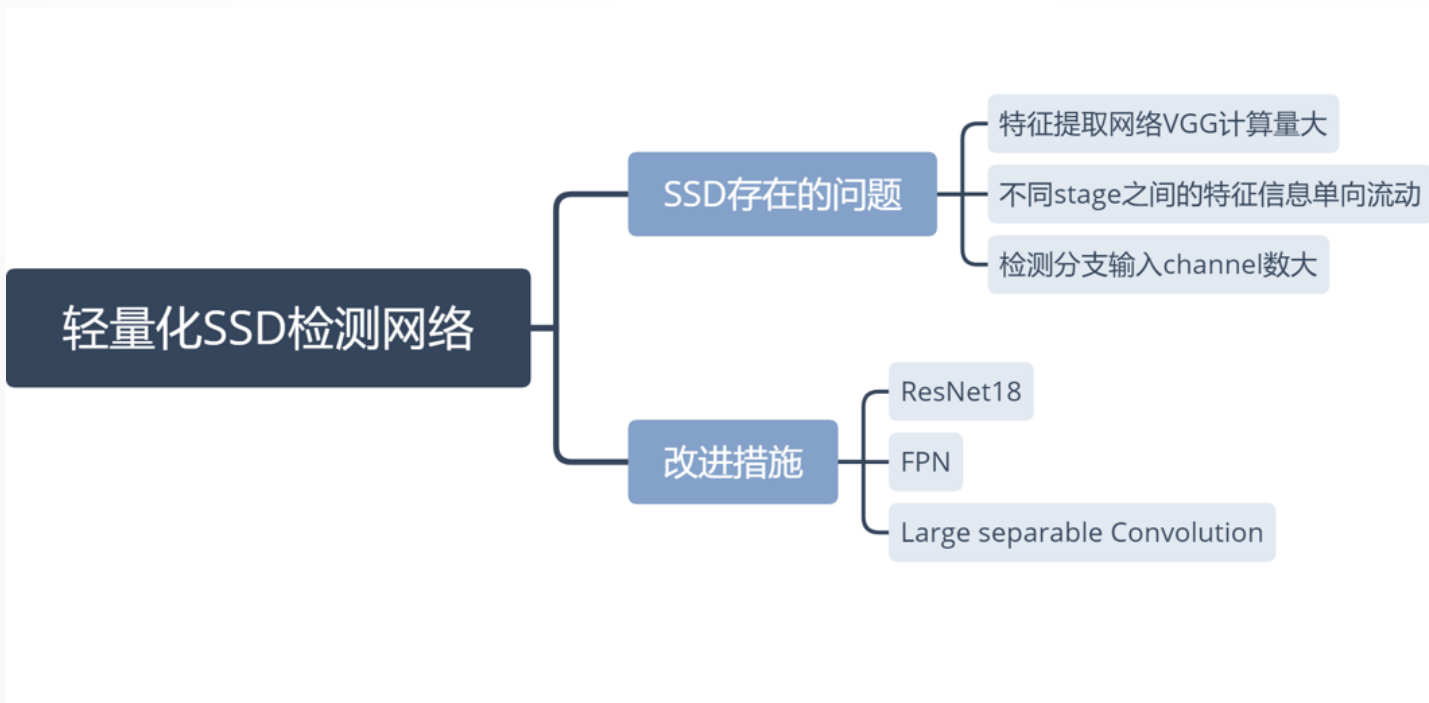
多任务属性识别算法



后端逻辑处理



### ➤ 检测算法

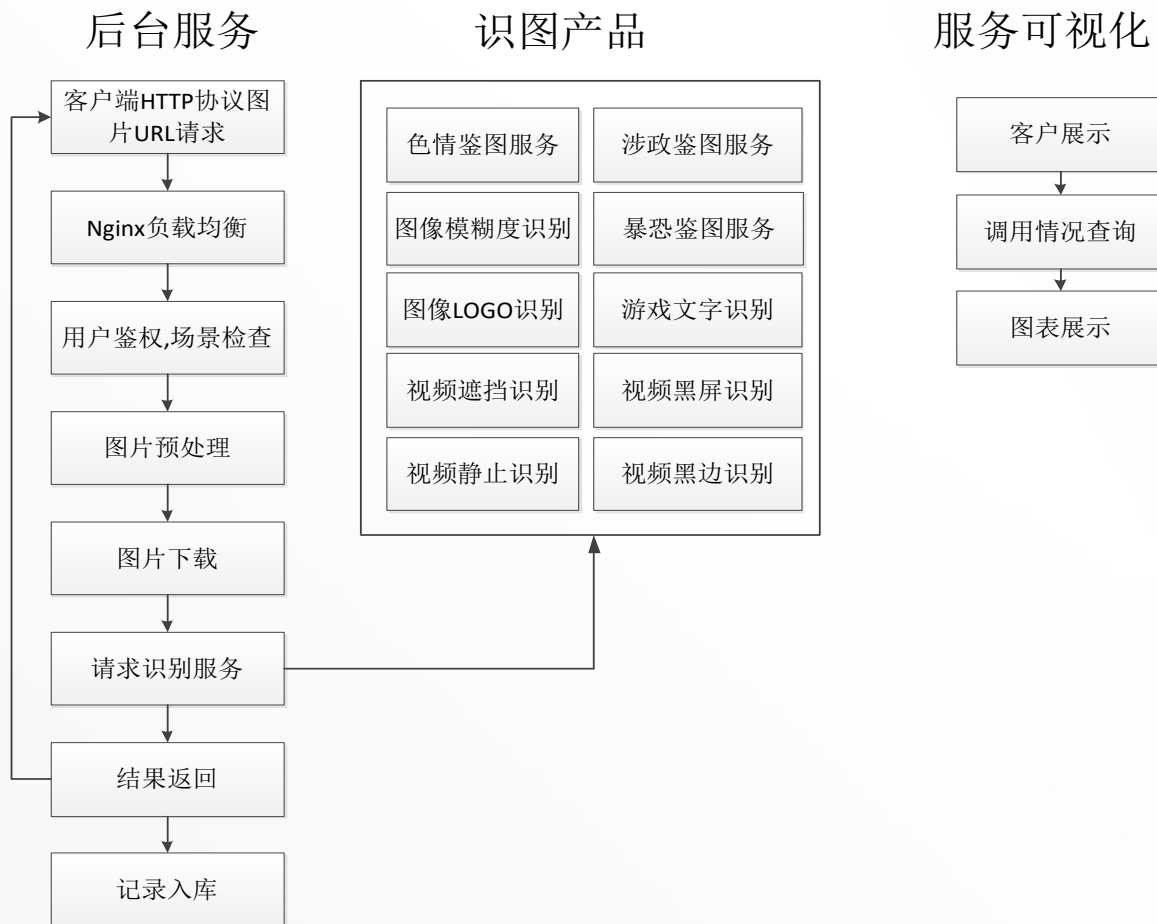


# 总结与展望

### ➤ 人工智能正循环



## ➤ 开放服务应用平台







<http://ai-image.yy.com>

# THANK YOU!

## 广州公司总部

地址：广州番禺区兴南大道368号万达广场B-1栋23层-39层  
邮编：511442  
电话：(+8620) 82120000 传真：(+8620) 82120801

## 广州公司分部

地址：广州黄埔大道中309号羊城创意产业园3-08栋  
邮编：510655

## 北京分公司

地址：北京市海淀区知春路7号致真大厦C座11层  
邮编：100085

## 珠海公司总部

地址：珠海市香洲区唐家湾镇高新区前湾三路总部基地A栋  
邮编：519000

## 上海分公司

地址：上海市卢湾区淮海中路381号中环广场2330室  
邮编：200021