



Day 10

视频识别应用场景和使用案例

HUAWEI TECHNOLOGIES CO., LTD.

www.huawei.com



目录

- 视频动作识别
- 视频名人识别
- 视频OCR
- 视频内容标签
- 视频质量识别

视频动作识别

背景：

快递行业普遍存在快递员由于对自身要求不高，经常性出现暴力分拣的现象，这既带来了货损，也产生了额外的管理成本。



暴力分拣行为识别：

通过对监控视频进行实时/离线行为分析，识别拣货员在拣货过程中出现的扔、抛、推倒、用力踢等暴力分拣行为，该服务可以有效降低暴力分拣行为的发生，从而减少由于暴力分拣带来的货损。



视频动作识别（暴力分拣行为识别）

服务优势：

● 准确率高

采用RGB+光流的深度学习网络与海量训练数据，模型识别准确率高。

● 识别速度快

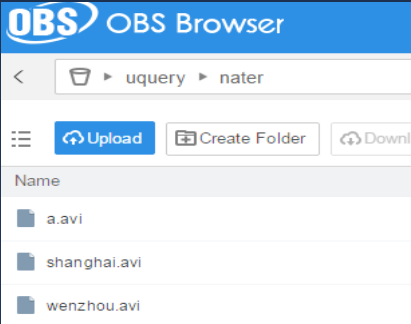
支持实时4路对视频进行暴力分拣检测识别。

● 支持复杂场景

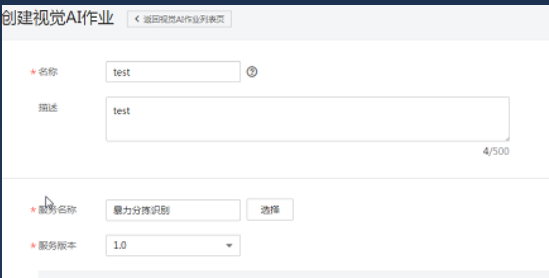
支持识别大雾、强光、弱光等复杂条件下的暴力分拣检测识别。

服务样例：

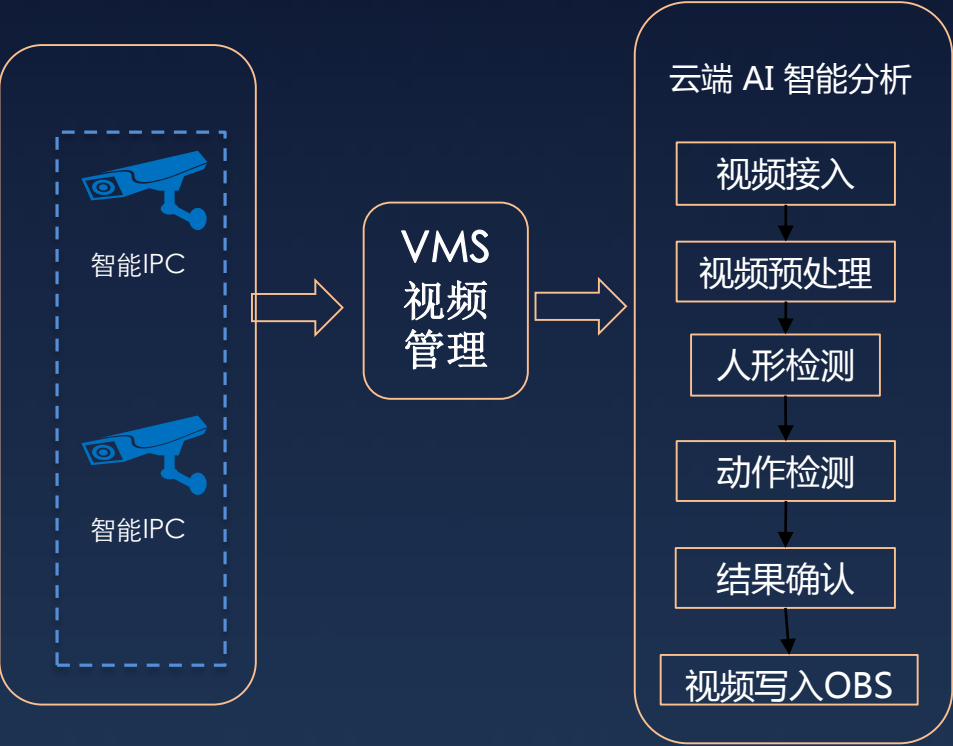
方式一：文件自动检测，
结果存OBS



方式二：华为云手动创建
视频作业



算法流程：



视频名人识别



服务背景：巨量多媒体视频的互联网时代，敏感内容的识别和处理非常重要，人工实时审核几乎不可能。自动精准检测是否包含敏感人物内容，大大降低企业违规风险，同时降低人工审核的成本。

场景描述：利用深度神经网络模型，准确识别视频中的政治人物、明星网红等名人，帮助业务规避违规风险。

例如在新闻视频或者娱乐视频领域，基于视频内容进行人脸检测，快速判断是否有政治人物、明星网红等敏感信息。识别并预警视频中的敏感内容，帮助客户快速定位处理，降低业务违规风险

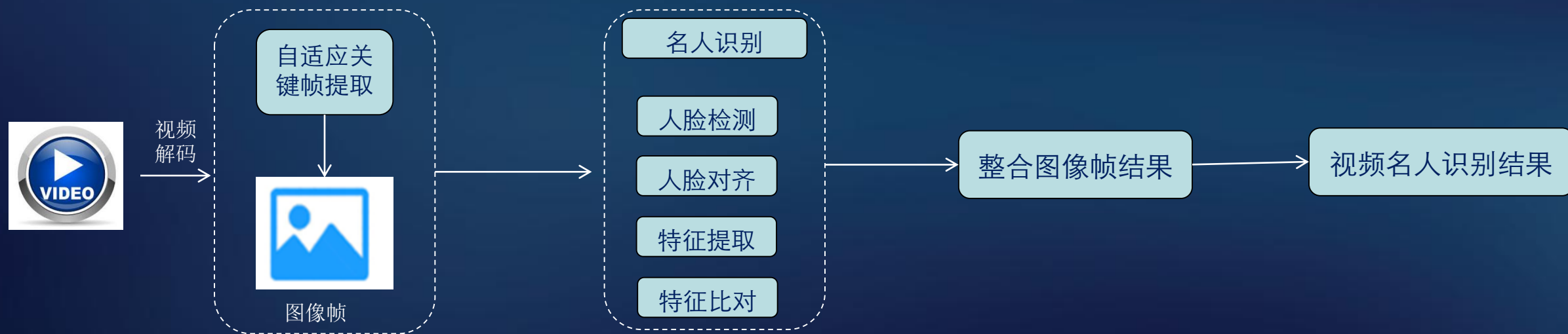
视频截帧检测结果:



视频名人识别



视频名人识别服务利用深度神经网络模型对视频内容进行检测，提取关键帧后，进行人脸检测、关键点对齐、特征提取，并与当前的名人底库进行特征比对，最后整合每一帧的识别结果，输出视频的名人识别结果。



视频名人识别服务能大幅降低业务违规风险，同时降低人工运营的成本。

产品优势:

简单易用：提供RESTful规范的API接口,使用方便,输入视频即可得到识别结果

精准检测：采用深度学习技术与大量样本库，涉政人物支持地市级、省部级和国家领导人，测试准确率达到95%

稳定高效：适用于多种格式视频，功能稳定高效，有效降低成本

安全可靠：快速提供安全、灵活的应用环境，确保系统稳定运行

视频名人识别



名人识别服务支持符合RESTful的API访问接口，输入视频即可得到识别结果。

名人识别服务也支持IVS平台测试。将视频放入obs，指定输入目录，创建识别作业，系统会输出json结果到指定输出路径。

输出结果：视频名人识别结果以json文件的方式保存在用户指定的输出路径下，如下图所示：

其中，各字段：

suggestion:检测结果是否通过。

block: 包含敏感信息，不通过

pass: 不包含敏感信息，通过

confidence: 置信度，取值0-1

frames:名人识别结果数组

url: 包含敏感信息的图像帧的obs保存路径

offset: 包含敏感信息的帧号

result: 检测出的名人信息

label:检测出的人物名称

confidence:置信度，取值0-1

face_detail:人物面部信息，包括4个值：**h:** 区域高度 **w:** 区域宽度 **x:** 区域左上角坐标**y:** 区域左上角坐标

```
{
  "frames": [
    {
      "url": "obs-dushubao/wxh_hjx/zhengzhirenwu/zhongguozuhe.mp4.capture/462.jpg",
      "offset": 18,
      "result": [
        {
          "label": "周强",
          "confidence": 0.8355730772018433,
          "face_detail": {
            "h": 78,
            "w": 57,
            "x": 144,
            "y": 178
          }
        }
      ]
    },
    {
      "url": "obs-dushubao/wxh_hjx/zhengzhirenwu/zhongguozuhe.mp4.capture/2002.jpg",
      "offset": 80,
      "result": [
        {
          "label": "李克强",
          "confidence": 0.7355591058731079,
          "face_detail": {
            "h": 163,
            "w": 124,
            "x": 319,
            "y": 138
          }
        }
      ]
    }
  ],
  "suggestion": "block",
  "confidence": 0.8721793293952942
}
```


场景描述:

利用OCR技术，精确识别视频画面中包括字幕、标题、弹幕、自然场景等内容，用以内容审核、视频归类、将视频内容存档；提取的内容做进一步后处理，可以用到涉政审核、广告内容识别、场景识别等大量应用场景。同时可以跟NLP技术结合，进一步理解视频内容，智能化分析视频。



技术特色:

- 速度快：现已达到可以同时处理8路视频，并持续优化
- 简单易用：提供端到端的服务，输入视频即可得到文字内容
- 识别精度高：海量视频数据训练、分析、建模，并能识别自然场景文字，竖直文字
- 通用性强：可以支持新闻、电影、电视剧、动漫、综艺等各种视频场景

调用方式：

- 上传视频至obs，IVA平台选择相关视频，输出包含所有视频文字的json文件
- 将视频拆帧，逐张图片调用，返回对应图片包含的文字



```
{
  "name": "obs-wxh/dianningpianduan.mp4",
  "fps": 15,
  "contents": [
    {
      "time_start": "00:00:00",
      "time_end": "00:00:01",
      "content": [
        "猫扑视频",
        "东方头条",
        "我正打算去健身中心"
      ]
    },
    {
      "time_start": "00:00:01",
      "time_end": "00:00:02",
      "content": [
        "猫扑视频",
        "东方头条",
        "去强健一下自己身体呢"
      ]
    }
  ]
}
```



视频内容标签

- 背景：**短视频、直播的兴起使视频的积累量剧增。然而海量的视频数据无法人工一一辨别视频内容。视频内容标签通过为视频打上标签，为用户搜索相关视频提供前提。



视频内容标签

- **技术模块:** 采用ResNet深度学习网络训练，支持七千多种标签分类。



- **服务优势:** 1. 支持七千多种标签，覆盖物体、场景、动物等多个类
2. 从声音、动作、图像、文字等分析视频，全方位深度理解视频内容，输出结果更加全面。
- **调用过程:** 用户直接将视频放入obs，指定目录，框架会自动读取视频，输出json结果。



```
{
  "result": {
    "tags": [{
      "level0": "Clothing",
      "confidence": 0.185079
    },
    {
      "level0": "Person",
      "level1": "Boy",
      "confidence": 0.305595
    },
    {
      "level0": "Person",
      "level1": "Football player",
      "confidence": 0.206600
    },
    {
      "level0": "Person",
      "level1": "Player",
      "confidence": 0.200861
    }
  ]
}
```


视频质量识别



服务背景：短视频、视频直播的兴起使得互联网公司视频的积累量剧增。然而用户上传的视频质量鱼龙混杂，良莠不齐，视频公司需要通过视频质量审核来保证优质的视频输出来吸引客户。随之而来的问题是传统的人工审核筛选出高质量视频存在难度，由于个人的偏好和标准不同，且质量审核考虑指标较多，导致筛选结果存在差异且精度低。



清晰？模糊？



低光照？模糊？

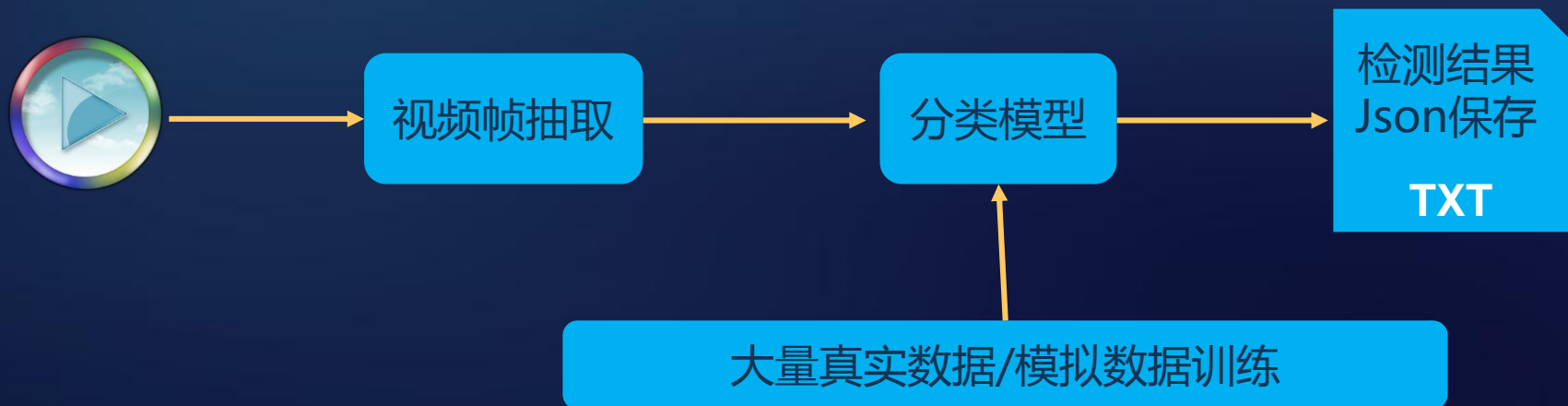
服务需求：有效地帮助视频公司筛选出符合质量标准的视频，进行自动化地过滤和筛选，保留清晰美观的视频，符合大众的审美，相比于人工审核，能够有效地提升审核效率。

视频质量识别

分类解决的10+质量类别：

- 抖动和重影：相机抖动或者视频中物体人物运动导致的重影
- 失焦：分辨率低、失焦、聚焦模糊（主体清晰就可以，不需要所有细节清晰；以手机屏幕大小来判断是否清晰）
- 低光照（光线暗的图像，拍的高质量的夜晚月亮图片除外）
- 过曝光
- 黑边（黑边、灰色的边、毛玻璃边缘等等）
- 白边
- 噪点（椒盐噪声）
- 马赛克（涂抹、马赛克、局部模糊等）
- 无实际意义（纯色背景、只有几行字+纯色背景、单调背景超过二分之一）
- 二维码
- 清晰但不美观
- 竞品logo

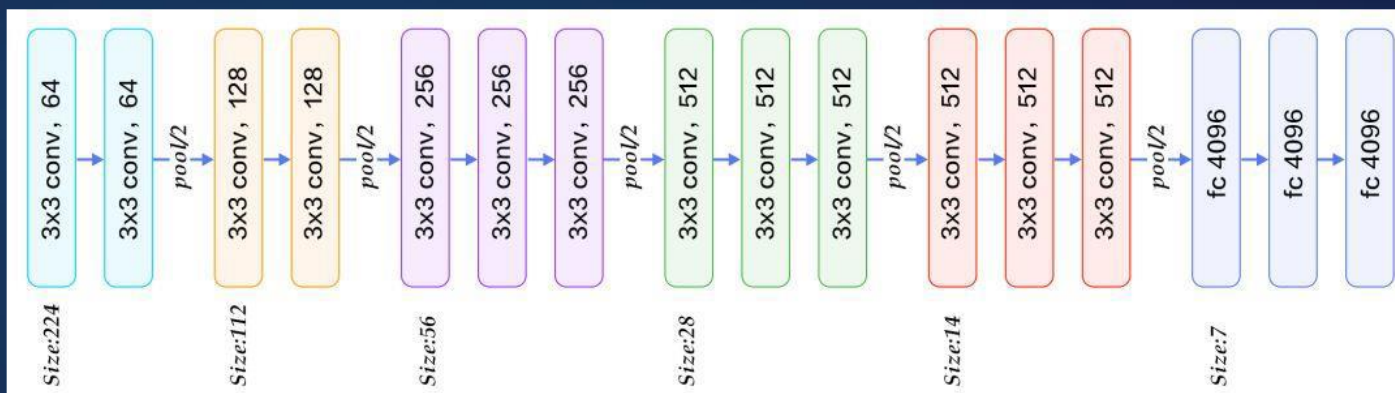
算法服务流程：



视频质量识别

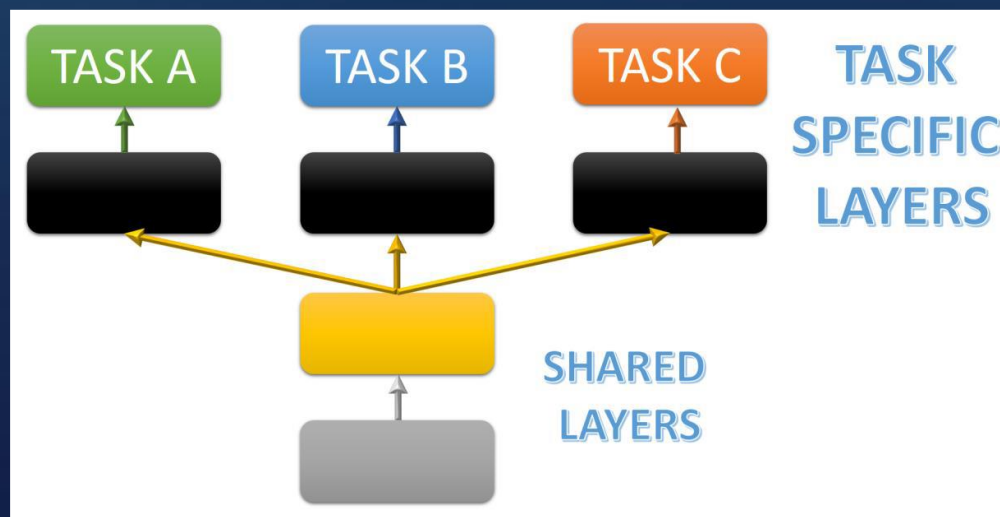
技术模块：

1. 采用VGG卷积神经网络结构进行图像特征提取。



2. 多任务学习。

- 加快质量检测的速度
- 减少需要的模型数量和模型大小



视频质量识别

服务优势：

- 准确率高

采用深度卷积神经网络与海量训练数据，模型识别准确率高。

- 识别速度快

实时对视频进行质量识别，快速识别低质量视频。

- 智能高效

质量算法全面识别13种评判指标，通过阈值把控快速识别、归类并过滤低质量视频。

服务样例：

输入 视频



输出 结果

```
{ "result": "Bad", "confidence": 0.790404,
  "details": [ { "tag": "mosaic", "probability": 0.732323 } ] }
```

THANK YOU

