



# Day 9

## 视频AI相关应用场景和技术简介

# 目录

- 视频识别场景和技术
- 视频理解场景和技术
- 客户使用案例

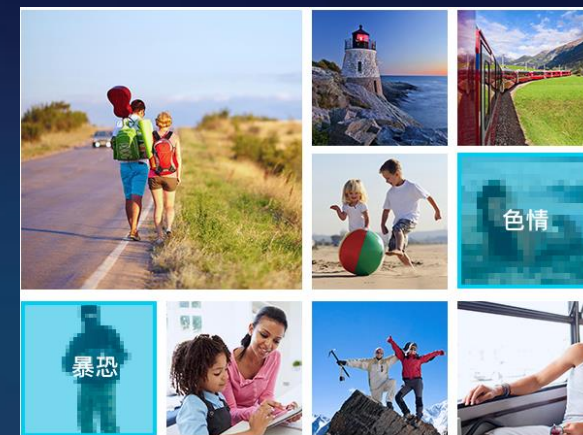
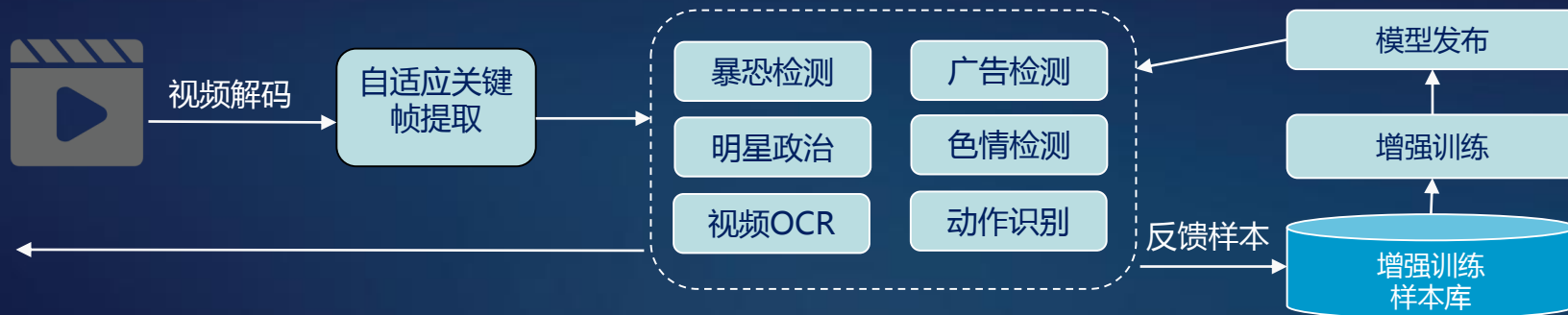
# 视频内容审核Video Content Moderation



## 视频内容审核

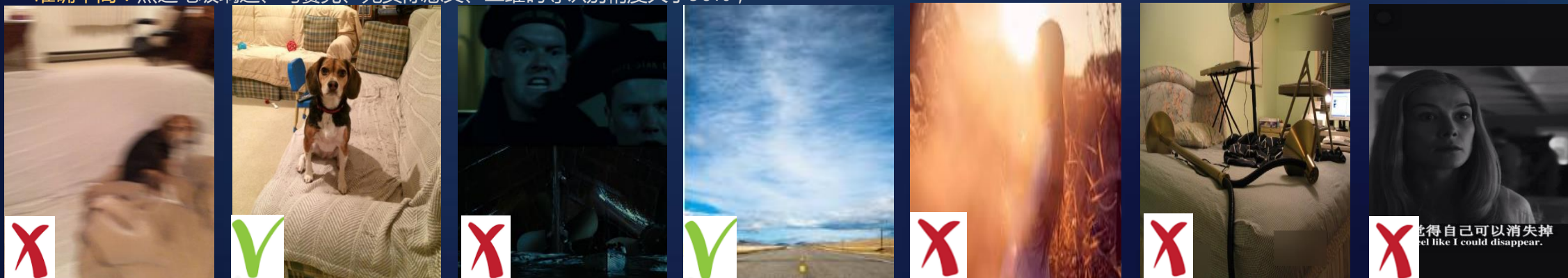
为各类视频平台提供视频中涉黄、涉恐、涉暴内容自动审核，有效遏制风险与释放审核人力，提升效率。

- **多模态审核**：支持同时对视频字幕、声音与画面多维度智能核查
- **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**：分钟级短视频秒级可以处理完成。

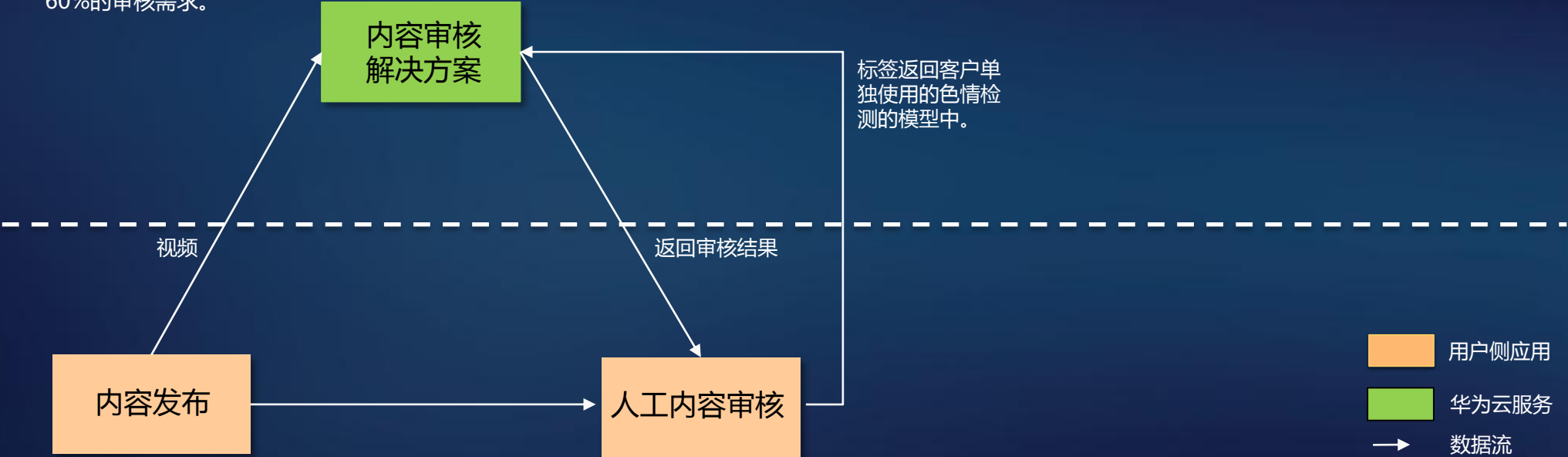


## 视频质量审核

- **10+质量审核指标**：模糊图像检测、光照强度检测、曝光检测、黑/白边检测、低分辨率检测、马赛克检测、二维码检测等。
- **华为独家**
- **准确率高**：黑边毛玻璃边、马赛克、无实际意义、二维码等识别精度大于90%；



如色情检测，服务本身的模型提供最基础的色情检测的正常，性感，色情三大类（包括如正常类中一般正常，卡通正常，雕塑像，名画，亲密行为，性玩具）以满足客户60%的审核需求。



如色情检测，不同客户对色情检测的定义无法完全一致，例如教育类与小视频类应用，且不同时期对色情检测的力度也不一致，所以通常在人工内容审核中，会打上更细致的标签。例如，比基尼、维密秀、wwe女女生赛事、性感女主播跳舞、海滩性感风（比基尼类除外）。客户审核人员打得标签会返回客户单独使用的色情检测模型中，以天为单位，加入训练。最终达到越用越准，满足客户80%的审核需求。



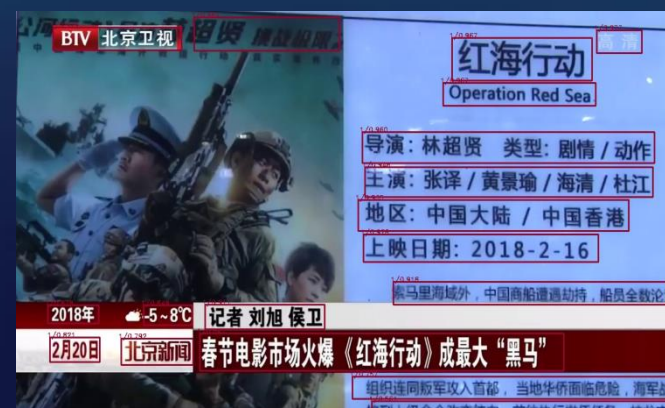
# 视频内容标签 Video Content Tagging



视频内容标签（Video Content Tagging），是一种对视频进行场景分类、人物识别、语音识别、文字识别等多维度分析，形成层次化分类标签的服务。

**层次标签：**支持七千多种标签，覆盖物体、场景、动物等多个类别；

**多维分析：**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。



# 视频内容标签

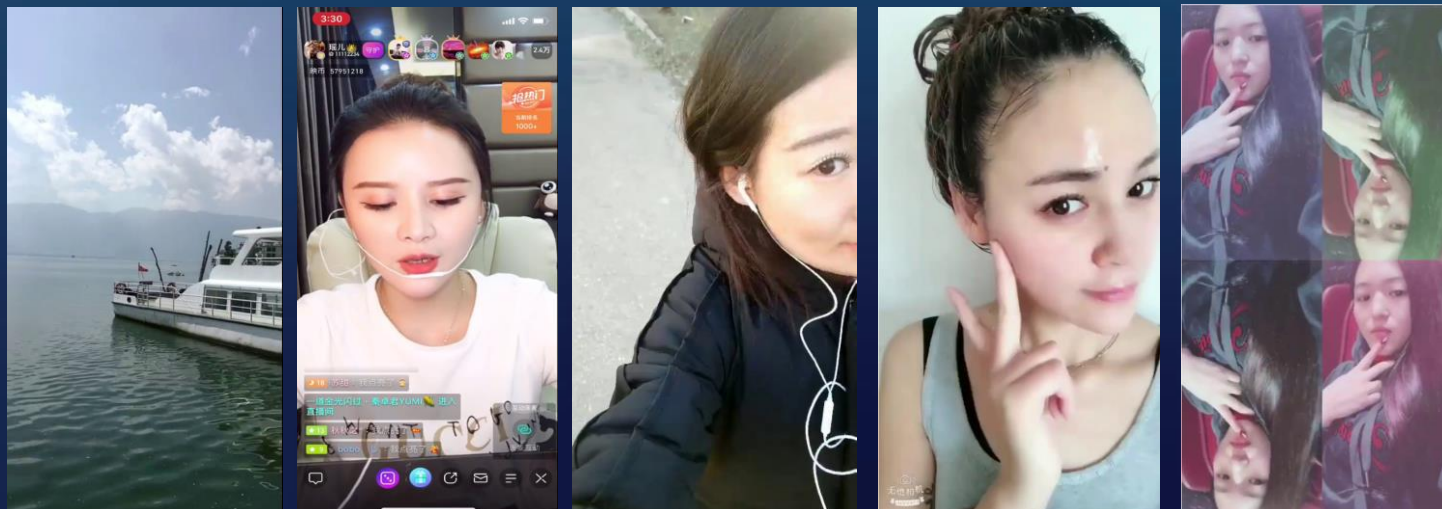


## X短视频

与XX图库联创，使用XX图库视频训练，标签模型具有约7500个标签的**通用标签**。7500个标签为通用标签，当前对用户提供的视频封面进行打标签。标签用于推荐系统，客户当前对标签效果满意。

## X直播

主播服饰/风格与场景标签正在开发中。室内室外场景分类了已演示过。下次演示时间未定。与通用场景不同，将来可抽象成**娱乐场景的标签库**



## 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



# 视频动作识别

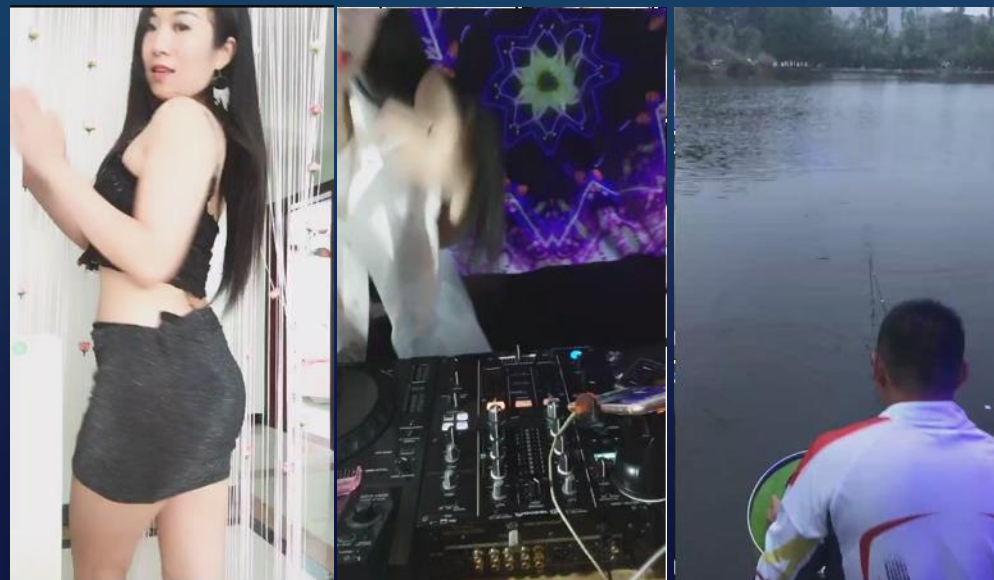
## 物流

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，客户对算法的效果满意。



## 直播

主播实时才艺识别正在开发中。通过视频行为分析，实时识别直播中的**唱歌、跳舞等9种主播才艺**，可用于实时推荐、用户行为分析等。





# 视频封面(进行中)



场景描述：基于互联网在线视频的内容理解，快速输出最具代表性和吸引力的精彩封面

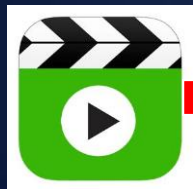
商业价值：智能选取代表性和吸引力的视频封面，相比人工选取封面速度提高20+倍，大幅降低运营成本；同时精彩的封面能够吸引用户的注意，大幅提高视频点击率。

原视频封面 VS 新视频封面：



技术特色：

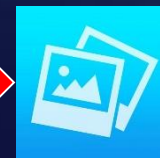
- 自动选图：提供端到端的服务，输入视频即可自动输出得到精彩封面，替代人工选图工作
- 提升视觉体验：基于视频内容理解的精彩封面，吸引用户眼球，提升视觉体验，提高视频点击率



视频帧过滤

关键帧提取

视频封面选取



# 视频摘要



场景描述：基于视频的内容相关度、精彩画面，提取场景片段制作视频摘要

商业价值：智能选取视频高质、精彩、代表性的摘要，相比人工生成摘要速度提高20+倍，大幅降低运营成本；同时高质量的视频摘要提升视频点击率与用户体验。

原视频 VS 视频摘要：



技术特色：

- 自动提取摘要：提供端到端的服务，输入视频即可自动生成视频摘要
- 提升视觉体验：基于视频内容理解的精彩摘要，吸引用户眼球，提升视觉体验，提高视频点击率



# 视频拆条



场景描述：基于人物、场景、语音、OCR字幕等多模态信息分析技术，快速精确地把长视频分割成不同主题的片段

商业价值：面向海量多媒体视频，大幅提高视频识别、剪辑、检索等处理的效率，降低人工运营的成本。整体效率相对人工提升10+倍，新闻拆条准确率超过95%，特定场景准确率超过98%

我国乡村振兴战略全面开局



经济高质量发展势头良好



上海：打响文化品牌，彰显城市精神



## 技术特色：

- 简单易用：提供端到端的服务，输入视频即可得到拆分结果
- 识别精度高：海量视频数据训练、分析、建模，精确拆分





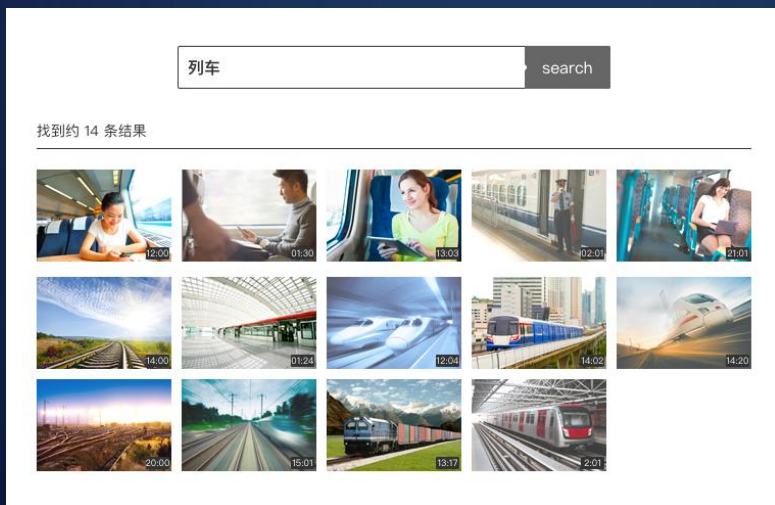
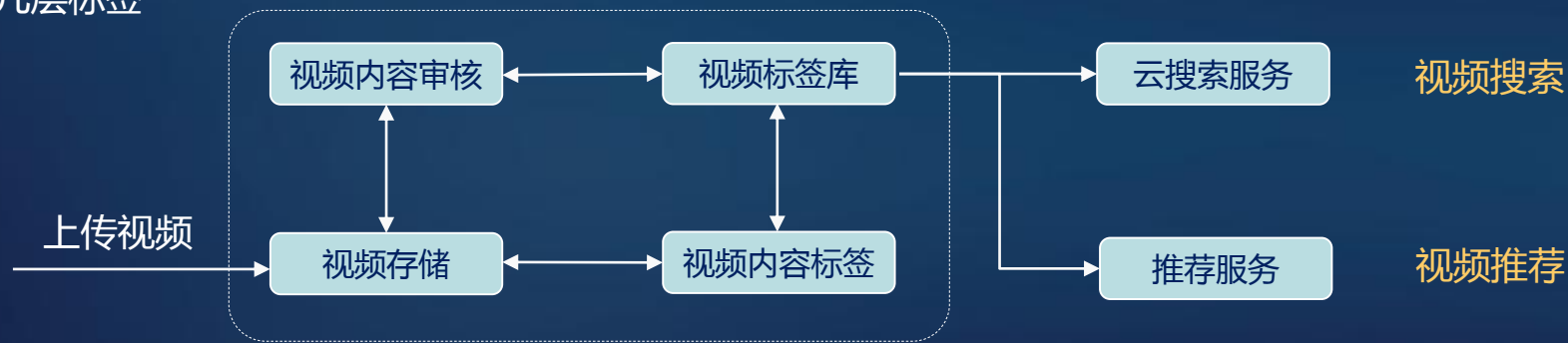
# 短视频网站



XX科技拥有大量的短视频，视频长度一般小于5分钟；每天1-2分钟长度的视频大约50万个，2-5分钟10万左右，目前是采用100人左右的团队人工审核。预计可减少50人以上。

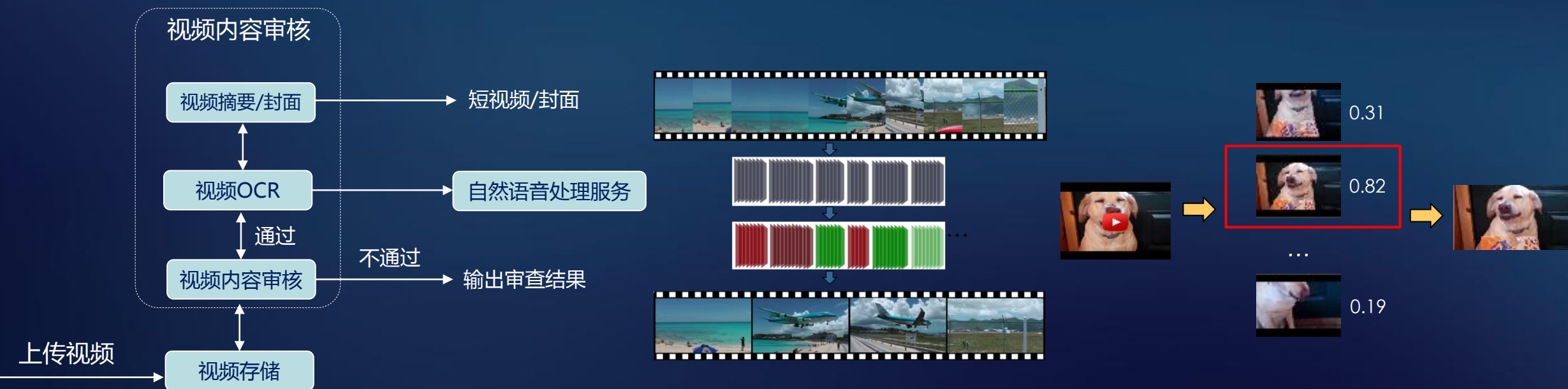
视频审核的流程，分5轮审核：

- 初审：根据视频的封面、关键帧等进行初审，过滤掉分辨率过低、画面有黑白白边、其他公司logo、涉黄、涉恐、涉政、广告、音画不同步等内容；
- 继续审核过滤视频，同时人工会对视频进行分类和打一些标签。标签分从基础（比如狗、猫等等）到高层语义（比如晚宴、聚会、旅游等等）的几层标签



## 帮助广电做视频处理与分析、内容审查

- **涉黄，政治人物，暴恐场景等**：由于他们是电视台客户，其实绝对违规的视频比较少，他们对鉴定的精度要求很高。
- **视频OCR**：识别新闻下面的标题文字，目前他们是对新闻视频写了个模板自动框出文字行，然后调用第三方服务转成文字；视频OCR、图文混排内容OCR；后期NLP提取出的文字内容进行自动摘要、标签、自动分类等能力，后续再提供更智能的语义分析能力。
- **视频封面和视频精彩片段摘要**：长视频中精彩场景的识别、以及精彩场景提取生成gif的能力。



# 直播



XX直播是实时、高清、快捷、流畅的视频直播平台,中国全新的视频社交媒体。

直播内容识别：

**人脸标签：**性别，年龄，颜值，长发/短发，脸型，眉形，单眼/双眼.....

**场景标签：**宿舍、教室、餐厅、酒店、办公室、休闲娱乐场所、商铺、直播间、健身房、交通工具、运动场、医院、理发店和户外。

**动作标签：**唱歌、跳舞、钢琴、古筝、鼓、吉他、健身、瑜伽.....

审核（底线过滤）：

涉黄/涉恐/涉政。

视频质量：清晰度审核。



唱歌检测：检测主播是否有唱歌行为



弹吉他检测：检测主播是否有弹吉他行为





THANK YOU

