



Day 11

视频理解应用场景和使用案例

目录

- 视频拆条
- 视频封面
- 视频浓缩
- 视频摘要

视频拆条



服务背景：巨量多媒体视频的互联网时代，快速对视频的理解与处理，需要先把视频拆分成不同主题的片段，这样才能大幅提高视频识别、剪辑、检索等处理的效率，降低人工运营的成本。

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，快速精确智能地把长视频分割成不同主题的片段。

例如在新闻领域，一条条不同主题新闻放在一起组成一个完整的新闻节目。智能拆条具有效率高精度的特点，能快速对新闻条目进行拆分，智能拆条服务有很大的市场前景。

产品优势：

简单易用：提供符合RESTful的API访问接口,使用方便,输入视频即可得到拆分结果

精准拆分：采用深度卷积网络与海量视频数据训练、分析，精确拆分

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

我国乡村振兴战略全面开局



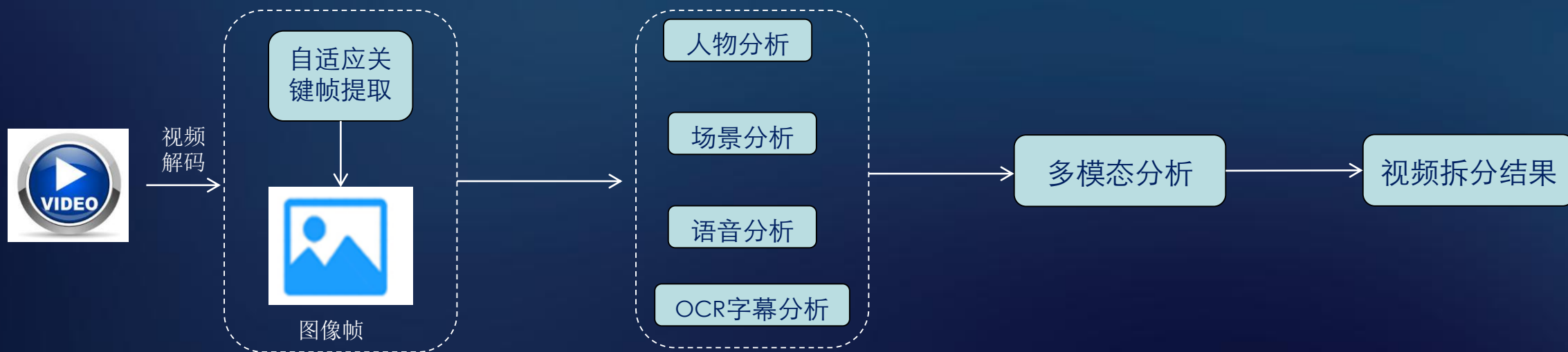
经济高质量发展势头良好



上海：打响文化品牌，彰显城市精神



- 视频拆条服务是通过深度卷积神经网络算法，对视频中的人物、场景、语音、以及OCR字幕进行分析，基于主持人分析、场景切换、话题转换、字幕内容转换等多模态信息将新闻节目智能地拆分成不同主题的片段。
- 智能拆条服务能大幅提高视频识别、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%



拆条服务支持符合RESTful的API访问接口，输入视频即可得到拆分结果。

拆条服务也支持IVS平台测试。将视频放入obs，指定输入目录，创建拆条作业，系统会输出json结果到指定输出路径。

输出结果：视频拆分结果以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：

fps: 视频帧率

result: 拆条结果数组

frame: 拆分点帧号

time: 拆分时间点

```
{
  "fps": 25,
  "result": [
    {
      "frame": 1382,
      "time": "00:00:55:07"
    },
    {
      "frame": 3469,
      "time": "00:02:18:19"
    },
    {
      "frame": 4179,
      "time": "00:02:47:04"
    }
  ]
}
```

视频封面

场景描述

移动互联网大行其道的今天，视频已经在娱乐、生活、生产内容中占据很大的比率。而视频封面能够快速向用户传递视频的主要内容，承担着吸引用户注意力、提升视频点击率的重要作用。

短视频、娱乐视频等内容生产速度快到已经无法人工设置视频封面。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。

技术特色

(1) 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，智能人工选图。

(2) 故事性

基于内容理解，结合图像美学分析，选择最能表征视频内容的静态封面，吸引用户注意力，提升点击率；

(4) 稳定高效

结合视频场景理解、关键帧分析以及深度网络模型压缩，提升视频内容处理速度；适用于多种视频编码和封装格式，功能稳定高效，有效降低成本；

视频封面

技术路线

(1) 基于图像分析

场景检测、分割、聚类

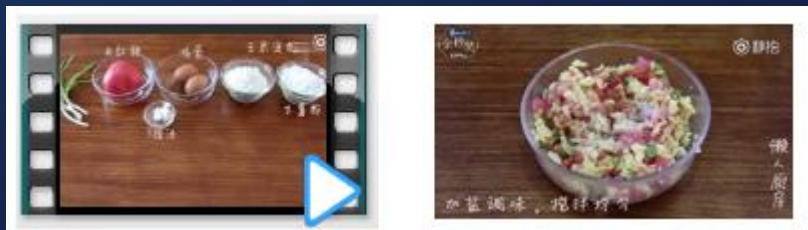
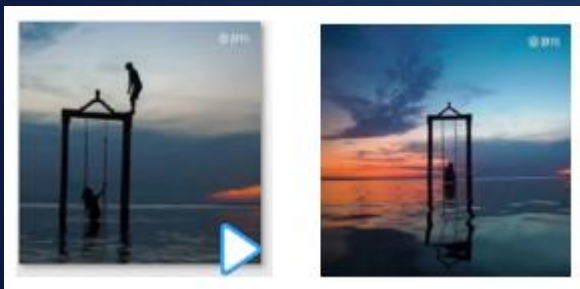
关键帧提取

低质量帧排除（模糊、色偏等判断）

(2) End2end DL based

主题事件关键帧提取 + 图像美学优选、排序

CNN based end2end cover selection



视频浓缩

视频浓缩

从原始视频中提取关键的信息组成缩略视频，让用户能够快速浏览整个视频内容。

场景描述

- (1) 视频浓缩，是对视频内容中事件/活动的一个简单概括，以自动或半自动的方式，先通过运动目标分析，提取运动目标，然后对各个目标的运动轨迹进行分析，将不同的目标拼接到一个共同的背景场景中，并将它们以某种方式进行组合。
- (2) 在安防、教育监控等领域，越来越多的使用到监控摄像头，导致爆炸监控视频数据量巨大。
- (3) 监控视频可以在许多实际应用中提供有用的信息作为安全监控和刑事证据。由于大多数监控视频在很长一段时间内很少有活动或者事件发生，这种极稀疏的视频人工实时审查效率低下、非常耗时。

视频浓缩

技术特色

(1) 高准确性

背景建模，使用背景生成及GAN技术，生成符合原始视频内容的背景图像序列；

轨迹优化，避免摘要后同一帧中的物体相互间不覆盖，轨迹不重合；

目标与背景的融合，使用泊松融合、深度神经网络等技术对目标与背景进行融合，生成自然连续的摘要视频片段。

(2) 稳定高效

融合目标检测和跟踪，缩短网络时间开销，提升目标检测的准确率和效率；

适用于多种视频编码和封装格式，功能稳定高效，有效降低成本；

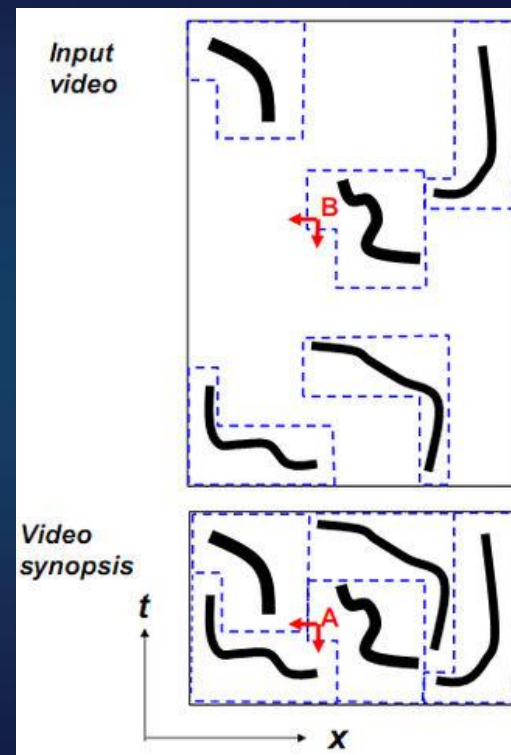
视频浓缩

技术路线

- (1) 前背景分离, 背景建模 (GMM等)
- (2) 运动目标检测和跟踪 (Kalman, KCF等)
- (3) 运动轨迹优化

浓缩后同一帧中的目标物尽量不重合或者最小化的减少重合覆盖问题

- (4) 目标、背景融合



视频摘要

视频摘要

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。将这些关键帧组合成新的视频。将原视频内容压缩成较短的视频摘要的形式，快速展现原视频中的主要内容。

视频摘要可用于视频内容快速浏览，也可用于视频的动态封面，向用户快速展示视频的精彩内容。

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。

技术优势

- (1) 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；
- (2) 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；
- (3) 适用于多种视频编码和封装格式，功能稳定高效，有效降低成本；

视频摘要

技术路线

(1) 传统非监督方法

- 人工设计的准则（相关性、重要性、多样性等）
- 主要生成方法
 - 简单规则
 - 基于视觉信息
 - 融合多模态信息（视觉、语音、字幕等）
 - 基于视频的句法语义结构
- Common base:
 - 视频分割（场景分割和聚类）
 - 关键帧提取

(2) 基于DL的有监督/无监督方法

- DPP+LSTM (end2end key-frame selection)

