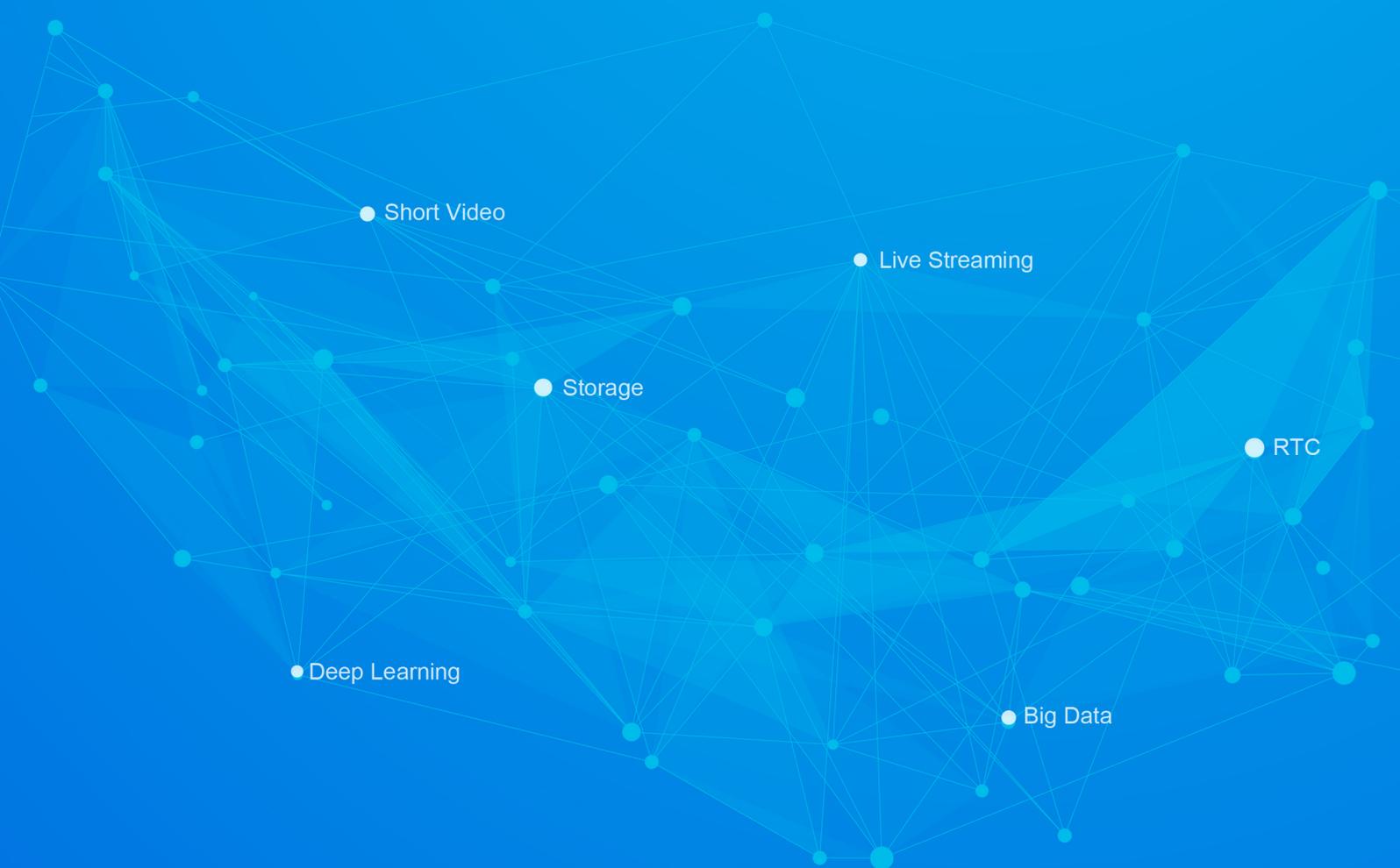


新一代智能视频云的五个关键因素

七牛云 & Gartner 联合发布



一、导读

过去十年里，中国移动互联网的高速发展，带动了各个产业 IT 基础架构的扩展和升级：

- 2017 年，大陆地区用户平均带宽达到 52M，已经追平甚至超越了许多发达国家水平；
- 移动互联网终端数超过 14 亿台，其中超过 60% 的设备接入了 4G 网络；
- 5G 时代也即将到来，视频的内容和展现形式将更加多元化，在直播、点播等基础场景上，又增加了短视频和实时音视频互动的新浪潮，同时视频技术也在安防、医疗、教育、司法、广电等领域有着全新的应用。

本文介绍各行业在视频应用领域的升级过程中，主要面临的技术痛点和挑战，并提出了构建新一代视频云的 5 个关键要素和相关参考标准，描述了智能视频云如何以更便捷的服务和更低的成本帮助客户加速升级。未来视频的应用的场景和领域可能会成为企业产品或营销服务中最核心的环节，而且它本身的规模变化也非常符合互联网中的马太效应，因此企业都需要提前为自己未来海量的富媒体素材做好充分准备，避免数据失控。

二、各行业在视频时代的新兴场景

1、安防监控

根据 Markets&Markets 的调查结果，2017 年至 2022 年期间，全球监控视频市场年复合增长率将达 15.4%，2022 年将增至约 756 亿美元的规模。视频监控广泛应用于各类场合，包括：

- 道路交通监控
- 城市安全监控
- 公共区域监控
- 家庭安防监控

最近两年来，以幼儿园、学校代表的公共区域场所监控提出了更高的要求：

- 过去无需外网访问的监控，现在需要做到全网实时观看；
- 过去无需长期储存监控画面，现在需要做到可随时查阅、长期存档。

而道路交通和城市安全的监控，除了传统的车辆违章监控和识别，行人的违章也逐步纳入到了监控的体系中，如：

- 智能识别画面中行人的违章行为；
- 实时识别人脸信息，快速确定违章行人身份；
- 自动录制违章画面，保存证据，上传至云端保存。

由此可以看到，视频监控领域正在面临着产业升级，如何做到公网稳定地访问，产生的海量的图像和视频数据如何更好地存储、分析和检索，是一项巨大的挑战。

2、在线教育

近年来，在线教育发展异常火热，利用互联网音视频技术，解决了优质教育资源传递的时空限制，主要体现形式如下：

- 直播课堂：直播授课，讲师可以更自然地提供在线辅导；
- 实时互动：低延时音视频传输，帮助教师实现和学生的实时在线交流；搭配使用聊天室功能：文字、语音、图片、包括自定义消息等，有效地提升了课堂氛围；
- 点播回放：通过云端录制以及在线回放，学生可以随时查看录播视频，对知识点进行查缺补漏，加深对知识的理解。

如何解决学生观看直播的卡顿问题，如何进一步降低视频互动的延时，提高老师和学生的体验，显得至关重要。另外，随着人工智能的发展，视频技术如何更好地借助 AI 技术，也成为在线教育领域大家更为关注的方向，如：

- 智能视频标签：根据学生上课状态，设置奖杯、小红花等奖励元素，截屏记录孩子优秀表现的状态，并保存，集锦视频分享给家长；
- 智能视频推荐：智能推荐适合学生的学习内容，包括教学视频、讲义、习题及习题讲解、测试题等。

3、广电新媒体

随着互联网直播兴起，各类广电平台想要做出一档可看且互动性强的直播节目，除了要精心编排节目内容，还要特别重视音视频及互动方式的呈现。显然，在新媒体时代下，传统的广电方案具有一定的局限性：

- 传统广电接入互联网，既要混编传统的音视频流，又要混编 PPT 文档共享等新兴数据流，保证多路画面实时切换；
 - 采用专属线路传输，三网合一方案进展缓慢；
 - 分辨率与码率较低，难以满足日渐普及的 4K 电视需求；
 - 线性播出，内容无法回放，互动形式较为单一；
 - 缺乏对视频内容的分析能力，仅能通过固定群体样本，推断出收视率等较为粗放的参考数据。
-

面对这些局限，广电行业迫切需要全新的视频系统，为消费者提供高画质、强互动的视频娱乐体验和可量化的精确数据管理：

- 在直播导播方面，云端进行音画混编服务，并在云端切换视频，快速形成导播效果；
- 在媒资管理方面，可对图像、视频进行智能拆条、智能审核和智能编目等，实现内容汇聚的生产、审核、管理等全套环节的智能处理，极大提升内容制作效率；
- 在内容运营方面，利用大数据能力和算法，实现用户行为数据标签管理，实现内容与广告的千家千面投放，提升广告价值；
- 在终端播放方面，传输层基于宽带网络的 4K 超高清视频传输，保证电视画面的自定义时移、回看，开放场景下的也可跨屏互动、多屏互通，提升终端用户和电视交互的便捷性和趣味性。

如何实现导播内容的实时切换，如何保证广电内容实时传输，如何让广电传媒的广告价值最大化，如何制作低成本、高质量的节目，这对广电新媒体来说均是急需解决的难题。

4、智慧法庭

2016 年 7 月 1 日起，最高人民法院所有公开开庭案件均有网上直播，所有直播视频均有视频存档，公众可在线观看。截至 2018 年 3 月，全国范围内的庭审直播已累计 66 万余场，累计访问近 50 亿人次。智慧法庭充分运用互联网、大数据、云计算、人工智能等先进信息技术，支持全业务网上办理、全流程依法公开、全方位智能服务：

基于视频及文书，结合人工智能计算机视觉技术，读取、分析电子化卷宗，抓取重要因素，并分类标注。例如把犯罪动机、犯罪时间、犯罪工具等一一以不同颜色打上标签，并予以比对分析。

建设智慧法庭，如何保障开庭审讯直播视频的实时传输，如何存储海量直播视频用于点播回放，如何基于广泛的视频内容进行智能分析，对视频基础设施的可靠性提出了更高的要求和挑战。

5、远程医疗

当前医疗资源还存在着地区分布不均问题，医疗专家可以通过在线直播、实时音视频进行跨区域互动会诊：

- 临床交互式会诊：利用视频会议系统，患者可以和医生“面对面”交流，医生现场实时解答患者的各种问题；
- 远程影像会诊：患者通过主视频与医生交流，同时可通过视频传送病历资料和数据，包括放射检查影像、病理检查图像、心电图、血压图、化验单或保存的录像等，还原线下会诊真实场景；
- 远程医疗培训：通过远程医疗培训系统开展专题讲座，传授最新医疗信息和诊疗经验，现场解答各种疑难问题，帮助听课医生拓展诊疗思路，提高下级医院整体从业水平。

远程医疗这一概念被提出后，已经被广泛应用。但是，如何提高视频传输性能，如何确保家庭、基层医疗机构和户外应急的远程医疗快速接入，是当前的远程医疗业务系统面临的主要挑战。

从以上列举的几个行业新兴场景可以看出，在产业升级的过程中，他们都面临着巨大的技术和资源挑战。大部分企业无法在短时间内构建有效的相关视频服务能力，因此如何选择和利用公有云的相关视频服务，快速满足业务的升级需求，显得至关重要。

三、智能视频云的关键要素

为满足各行业在视频时代新的需求和挑战，智能视频云至少需要拥有以下 5 个关键要素：

1. 稳定的网络传输与分发：直播延迟不高于 1s，互动延迟不高于 150ms
2. 可扩展的海量存储服务：可以实现业务层无感知 PB 级扩容
3. 云端媒体的编辑处理：可以在数秒内创建独立计算实例，执行自定义媒体编辑操作
4. 视频内容的智能分析：视频内容识别率高于 95%
5. 完备的权限控制：彻底杜绝非法拷贝，秒级封禁盗链流量

要素 1: 稳定的网络传输与分发：提供流畅的观看体验与低延迟的互动

- 内容分发网络 (CDN) 的优化：在传统 CDN 基础上融合优化，建立高质量全球节点，实现直播、点播内容的加速分发，进一步提供秒级首开、低延迟的视频播放体验。
- 实时流网络 (LiveNet)：针对复杂的网络环境、高成本的跨运营商服务和边远地区基础设施落后等客观因素，基于全球节点搭建全球化实时流网络，采用软件定义网络的方式，动态制定线路组合并决策调度最优线路。
- 具备完善的客户端 SDK：智能视频云需要一整套含推流、短视频、播放器等多种 SDK 开发套件，帮助用户完成视频的生产、编辑和消费，同时还能降低移动端应用的开发难度和时间成本。
- 利用最新编解码技术和传输协议，减少对传输网络带宽和质量的要求，降低卡顿率，如 P2P 通信，H.265 编解码，QUIC 协议等。
- 支持标准的 WebRTC 协议栈，降低端到端的延时，实现百毫秒级别的音视频互动体验。

要素 2: 可扩展的海量存储服务：高可靠易扩展的数据安全保障

- 稳健可靠的对象存储：通过纠删码存储、跨数据中心的副本冗余等技术手段，提供高达 16 个 9 的数据可靠性，保证每年服务不可用时间低于 30s，确存储数据的超高可用性。

- 易扩展的技术架构：存储系统支持动态扩充存储节点，实现存储需求的弹性收缩，保证 PB 级别的存储内容可动态扩容，确保业务系统聚焦业务增长，无后顾之忧。
- 利用边缘计算和边缘存储：将运算和存储分散在靠近数据源的近端设备中处理，无需把数据实时回传云端处理，减少云平台工作量。大幅提高效率，降低延迟，成为云平台的有效补充和优化。
- 支持低频存储特性：将拥有高吞吐量，较高持久性和较低访问延迟的数据存储在云端，极大降低海量存储数据的企业运营成本，成本比常规方案降低 60%，访问延时可控制在 50ms 以内。

要素 3: 云端媒体的编辑处理：快捷多用途云端视频编辑

- 快捷轻量化处理：视频编辑的任务，本质上是针对视频文件的数据运算与处理，通过部署灵活的容器化平台，视频编辑时需要的物理资源利用率可从 40% 提高到 70% 以上，发布效率提升 5 倍，突发业务实现秒级响应。
- 丰富的多媒体编辑能力：提供视频转码、截图、水印、旋转、切片等多媒体数据处理服务，满足直播、点播等多种实时节目制作场景，快速完成效果制作，输出专业制播内容。

要素 4: 视频内容的智能分析：结合人工智能发挥视频数据的最大价值

- 智能多媒体内容识别：通常包含内容审核、OCR、场景识别、人脸识别、音视频处理、图片处理等丰富功能。同时，面对每天高速增长的数据处理请求，弹性扩缩容的内容识别平台能保障服务器免受巨大压力。
- 深度学习平台的数据整理：一个高性能的深度学习平台框架可以轻松完成每日撰写迭代训练脚本、新数据的增删和管理、增量学习和迭代学习、搭建半监督打标系统、模型的比较和融合等工作，综合下来，可以减少 70% 的重复劳动。
- 海量媒资知识库体系由视频结构化模块，知识图谱模块和大数据检索模块三者组成。视频结构化模块完成对视频中基础元素和内容的提取和整理，将线性的视频解构成可以被碎片化使用的组件；知识图谱用于将视频结构化得到的事件、人物、物体、场景等信息的有序梳理，并以便于检索和关联的形式进行存储和呈现；大数据检索模块在前两者的基础上，提供海量媒资内容的高效检索，可

以根据人物特征、人脸特征、图像特征、视频特征及更复杂的组合结构，快速地提供视频数据检索服务。

要素 5: 完备的权限控制：杜绝非法拷贝与盗链

- 完备的防盗链机制：无论是直播还是点播，对视频内容的访问，都需要有完备的防盗链机制，常见的防盗链方式有 referer 防盗链、时间戳防盗链、回源鉴权等多种，基本上能有效地降低盗链风险。同时视频云还应当对突发性的盗链流量具有感知能力与报警机制，可以快速阻止盗链访问。
- 可靠的 DRM 数字版权保护机制：除了防盗链，内容的版权保护也是视频云需要提供的能力。常见的方式是对上传的视频文件进行转码和加密，输出加密后的视频文件再进行网络分发，由终端播放时再解密播放，这样即可真正实现对内容的版权保护，防止非法拷贝。

根据以上 5 个关键要素，七牛云认为一套完整的智能视频云需要由以下模块组成：



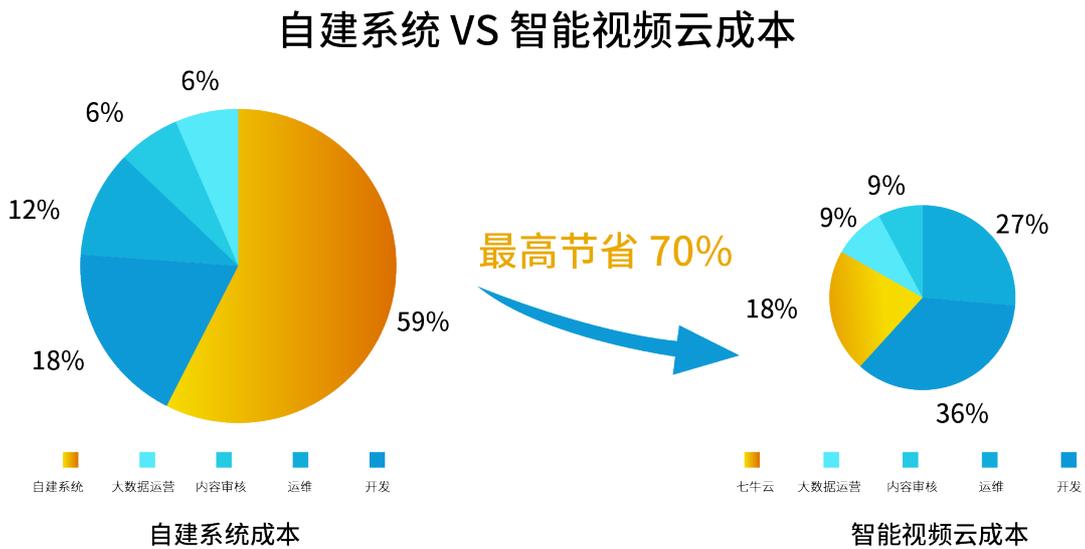
直播 SDK 产品设计可参考：www.qiniu.com/products/pili

短视频 SDK 产品设计可参考：www.qiniu.com/products/plsv

播放器 SDK 产品设计可参考：www.qiniu.com/products/player

四、智能视频云的成本优势

智能视频云不仅在技术上能够完全满足各行业在视频时代的新需求，相比自主研发，还能为企业大幅节省研发成本和运营成本。



面对高昂成本，视频云服务在提供了丰富产品的同时，还具有简单易用、可弹性扩展、维护成本低的特点。智能视频云服务通过提供通用且易于根据具体业务定制的技术系统，大大降低了行业应用的开发周期与成本开销，同时视频云的多个模块均可私有或混合部署，在保障企业数据安全的同时，获得和公有云一样的稳定性、可靠性和灵活性。

五、AI 在智能视频云的增值应用

在智能视频云的技术和成本优势背后，人工智能，尤其是计算机视觉技术的深度应用发挥了巨大的作用。在七牛云整个智能视频云系统中，计算机视觉技术在诸多环节取代了人工操作，极大地提高了视频内容处理的效率。通过完全不同于传统数据分析的角度，实现了先前难以想象的数据分析应用。

1、视频内容识别：自动识别视频画面信息，从标签库中匹配对应标签

作为计算机视觉基础模型层中最重要的几项技术，人脸识别、物体识别、场景识别等技术在安防、广电、教育等领域的众多场景中有广泛的应用场景。

例如在安防行业，拥有人脸识别、动作追踪等功能的高清摄像头，能根据监控范围内人物的动作并判断其行为，对具有可疑行为的人触发自动报警。当智能摄像头与公安部门的在逃嫌犯资料库匹配之后，能在机场、火车站等人流密集场所配合警方识别可疑人员，大幅提升破案、抓捕效率。

2、视频内容结构化：可以像搜索文本文件一样搜索视频文件中的信息

计算机视觉视频结构化相比人工标注具有识别范围广、准确性好、学习模型不断迭代、GPU 机器效率高、成本低等一系列明显优势。经标注后的视频能够在远程医疗、在线教育、广电等行业发挥巨大作用。

例如，远程医疗行业拥有的存量 and 新增影像资料远远超出人工标注的能力范围，想要发挥医疗视频资料的最大价值，需要根据不同类别将视频和图片进行分类。通过人工智能高效完成视频精确分类后，就能像搜索文本文件一样搜索视频文件中的关键信息，进行更有效的利用，真正将医疗大数据变为医疗知识图谱。

3、视频内容审核：从视频中识别敏感信息，提升审核效率

如今，图片和视频取代文字成为主流的传播方式，针对图片和视频的内容审核正变得越来越重要。然而人工审核不仅会造成高企的人力成本，其效率和准确性也很难满足当下巨大的视频数据量带来的审核需求。

例如在广电行业，视频内容鉴黄、鉴暴恐、政治人物识别等工作过去均由人工完成。随着计算机视觉技术的日渐成熟，利用机器代替人工进行绝大部份的内容审核工作，成为了视频时代的提高审核效率的不二之选。凭借计算机视觉技术带来的审核效率革命，鉴黄、鉴暴恐、政治人物识别将不再是广电行业的难题。

体验智能内容审核产品可访问：www.qiniu.com/products/censor

4、视频内容推荐：可根据用户消费视频习惯，实现精细化运营

除了能够高效完成视频结构化和内容审核工作，计算机视觉技术还能在业务层面成为内容运营的创新引擎，实现更多个性化产品需求。

同样在广电行业，例如在完成视频内容结构化之后，运营者可以根据用户收看行为记录进行智能内容推荐，甚至在视频的特定时间段和画面的特定位置实现广告定向投放，最大化提升广告转化效果。智能视频内容推荐能够帮助内容运营人员以最高效率实现高水平的用户精细化运营。

六、借助灵活的智能视频云 从容面对未来产业升级的未知挑战

未来，很少有企业能脱离互联网而独立存在，因此企业数据的总量会不断增加，对应产生的数据价值甚至负担也会同步提升。每个企业都需要拥有对于文件和富媒体素材（包含海量图片，视频，音频）的使用和存储的弹性能力，但只有极少数企业才需要拥有自建视频云的能力和资源。大部分企业需要的是一套稳定且可持续升级的视频云平台，迎接未来产业中不断变化和升级的未知挑战。

本文涉及相关资料

Gartner 白皮书原文：

www.gartner.com/technology/media-products/newsletters/qiniu/1-4UWO347/index.html

直播 SDK：www.qiniu.com/products/pili

短视频 SDK：www.qiniu.com/products/plsv

播放器 SDK：www.qiniu.com/products/player

智能内容审核：www.qiniu.com/products/censor

2018 年 7 月

七牛云出品

七牛云是国内领先的以视觉智能和数据智能为核心的企业级云计算服务商，同时也是国内最有特色的智能视频云服务商，累计为 70 多万家企业提供服务，覆盖了国内 80% 网民。围绕富媒体场景推出了对象存储、融合 CDN 加速、容器云、大数据平台、深度学习平台等产品，并提供一站式智能视频云解决方案。为各行业及应用提供可持续发展的智能视频云生态，帮助企业快速上云，创造更广阔的商业价值。