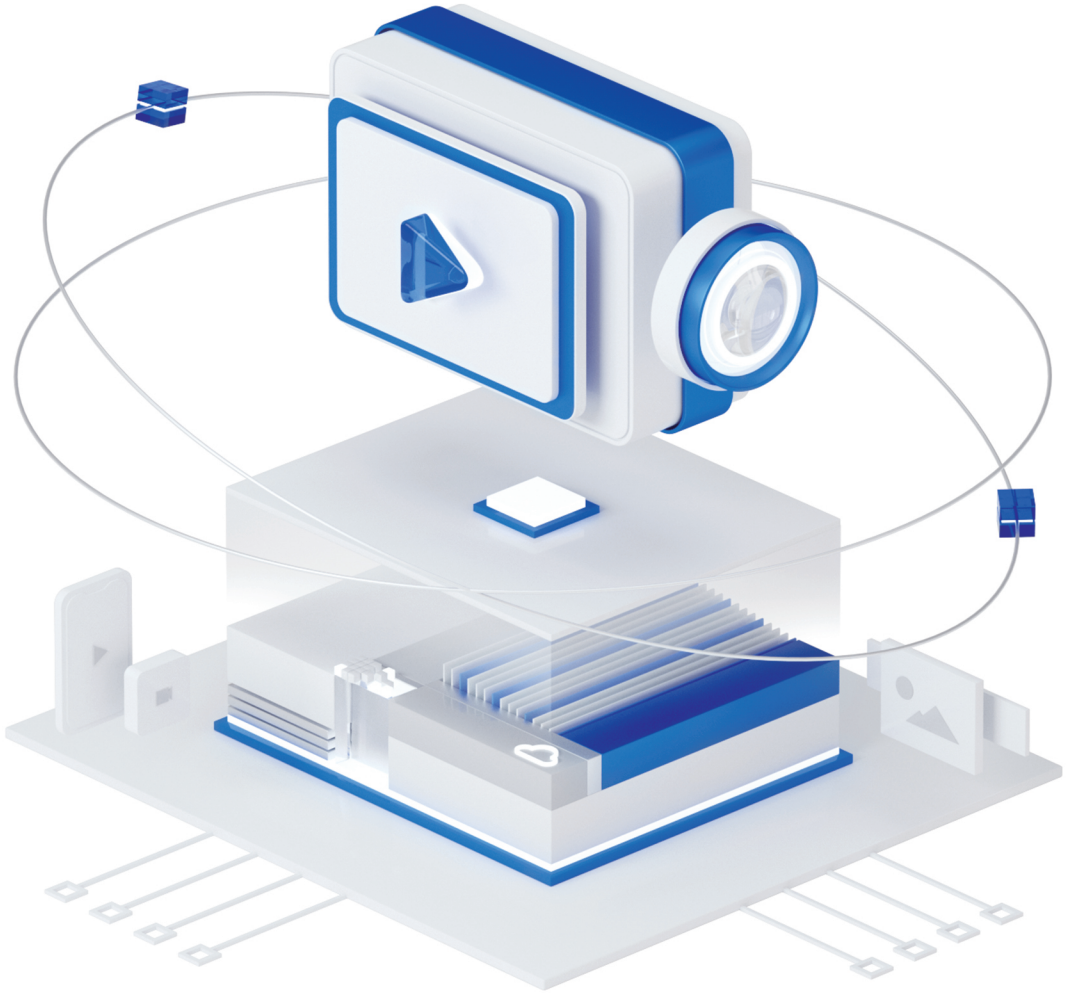
百度智能云

“云智一体”技术与应用解析系列白皮书

全场景智能视频篇



目录 CONTENS

一、云智一体，促进数字化转型与智能化升级一步到位 …… 02

1.1 产业智能化，数字经济发展新阶段 …… 03

1.2 百度智能云战略升级 …… 03

1.3 云智一体的百度智能云架构 2.0，加速产业智能化 …… 04

二、云智技术大发展让视频应用与体验跃入新阶段 …… 06

三、云智一体的全场景视频技术解析 …… 09

3.1 5G+AI 驱动下的视频行业发展趋势 …… 10

3.1.1 消费类视频的云智一体化诉求 …… 10

3.1.2 产业类视频的云智一体化诉求 …… 11

3.2 智能视频云 3.0 全景图 …… 12

3.2.1 智能视频 3.0 的特点 …… 14

3.2.2 智能视频 3.0 背后的端边云一体化基础架构 …… 14

3.2.3 基于云智一体的双平台 …… 15

3.3 智能视频云 3.0 中的“云”能力 …… 17

3.3.1 视频分发基础设施 CDN …… 17

3.3.2 智能编码 …… 18

3.3.3 智感超清 …… 19

3.3.4 实时音视频通讯 …… 21

3.3.5 低延时直播 …… 23

3.3.6 智能视频 SDK 矩阵 …… 24

3.4 智能视频 3.0 的“智”能力 …… 25

3.4.1 智能理解分析 …… 25

3.4.2 智能视频生产 …… 27

3.4.3 视频交互生成 …… 27

四、智能视频云 3.0 的三大行业应用场景 ····· 31
4.1 云智一体的视频技术推动泛媒体产业升级 ····· 32
4.1.1 泛媒体产业格局：应对新挑战、把握新动能 ····· 32
4.1.2 泛媒体产业演进的三条增长曲线 ····· 33
4.1.3 智能媒体解决方案全景图 ····· 34
4.1.4 泛媒体客户案例解析 ····· 36
4.2 云智一体的视频技术推动泛互联网产业升级 ····· 39
4.2.1 智能小视频解决方案：快速上线视频业务 ····· 39
4.2.2 全链路编解码优化方案：降低成本、提升画质 ····· 41
4.2.3 新零售直播 SaaS 解决方案：助力商家裂变获客、流量变现 ····· 42
4.2.4 在线社交解决方案 ····· 43
4.2.5 泛互联网客户案例解析 ····· 44
4.3 云智一体的视频技术推动泛产业的视频场景升级 ····· 45
4.3.1 产业类视频与消费类视频的区别 ····· 45
4.3.2 视联网发展的三大阶段 ····· 46
4.3.3 百度智能云视联网感知平台 ····· 47
4.3.4 视联网感知平台架构的三大特点 ····· 48
4.3.5 产业类客户案例解析 ····· 52
五、更多行业案例参考 ····· 54
六、关于本报告 ····· 55

云智一体，
促进数字化转型与智能化升级
一步到位 >>

1.1 产业智能化，数字经济发展新阶段

当前，我们正处在科技创新和产业发展最好的时代，以人工智能、大数据、云计算、5G 等为代表的新一代信息技术蓬勃发展，驱动新一轮科技革命和产业变革。

在产业应用新兴技术和数据资源转型升级的过程中，数字化是基础；而随着人工智能等新一代信息技术的发展，以及数据的爆发式增长，产业加快应用智能技术，从海量数据中发现规律、训练模型、提炼知识，促进产出增加和效率提升，实现企业生产经营的智能化，进而实现整个产业的智能化升级。

因此，数字经济既包括数字化转型，也包括智能化升级，产业智能化是数字经济发展的新阶段。

1.2 百度智能云战略升级

百度是拥有强大互联网基础的领先 AI 公司。从成立的第一天起，百度搜索引擎就与人工智能、云计算等技术息息相关。经过多年的技术积累和产业实践，百度已形成了全面布局，从基础的算力和数据技术、深度学习算法及框架，到语音、视觉、自然语言处理等感知、认知技术，以及飞桨深度学习开源开放平台等，具备云智一体的独特优势。

基于百度的技术优势和对产业发展的洞察，经过跟合作伙伴的共同成长与产业实践，百度智能云的战略升级为：以“云计算为基础”支撑企业数字化转型，以“人工智能为引擎”加速产业智能化升级，云智一体“赋能千行百业”，促进经济高质量发展。

云智一体的“云”为数字化转型提供安全、稳定、灵活的数字化底座，“智能化引擎”为智能化升级提供领先的创新技术和平台。我们推荐客户数字化转型与智能化升级一起做，一步到位，这样效率更高，效果更好。

同时，如果客户只做数字化转型，我们有丰富的云计算产品可以支持；或者客户已经有了数字化基础，我们的智能化引擎也可以助力智能化应用。

百度智能云的“云”，不仅为数字化转型提供安全稳定、弹性灵活的云计算服务，同时也是适合跑 AI 的云，我们称之为“AI 原生云”，继承云原生的优势，面向 AI 场景，提供极致弹性的高性能异构算力，打造简洁、高效的 AI 应用开发架构。

“云智一体”的智，是百度将深耕十余年的 AI 技术，以及 AI 赋能产业的实践经验，沉淀为助力产业智能化升级的技术、平台及方法论，是以加速产业智能化升级为己任、懂场景的 AI。

1.3 云智一体的百度智能云架构 2.0，加速产业智能化

基于百度智能云的战略升级，云智一体的百度智能云架构 2.0 同时发布，包含数字化底座、智能化引擎和全场景应用（如图 1.3-1 所示）。

数字化底座，包括基础云、数据库、物联网、边缘计算、区块链等基础平台，以及视频云、大数据、云原生开发和地图服务等；同时还有安全模块，为技术创新和产业转型升级保驾护航。

在数字化底座之上是智能化引擎，百度自主研发的飞桨深度学习平台为核心，软硬一体 AI 大生产平台 -- 百度大脑为支撑，包括 AI 中台和知识中台等。智能化引擎与行业深度融合，帮助企业建设自己的智能化中台，助力智能化升级。

基于数字化底座和智能化引擎，百度智能云深入行业场景，打造智能应用，同时与生态伙伴一起，在制造、能源、城市、金融、医疗、媒体等领域，助力数字化转型和智能化升级一步到位。

（图1.3-1 百度智能云全新架构2.0）

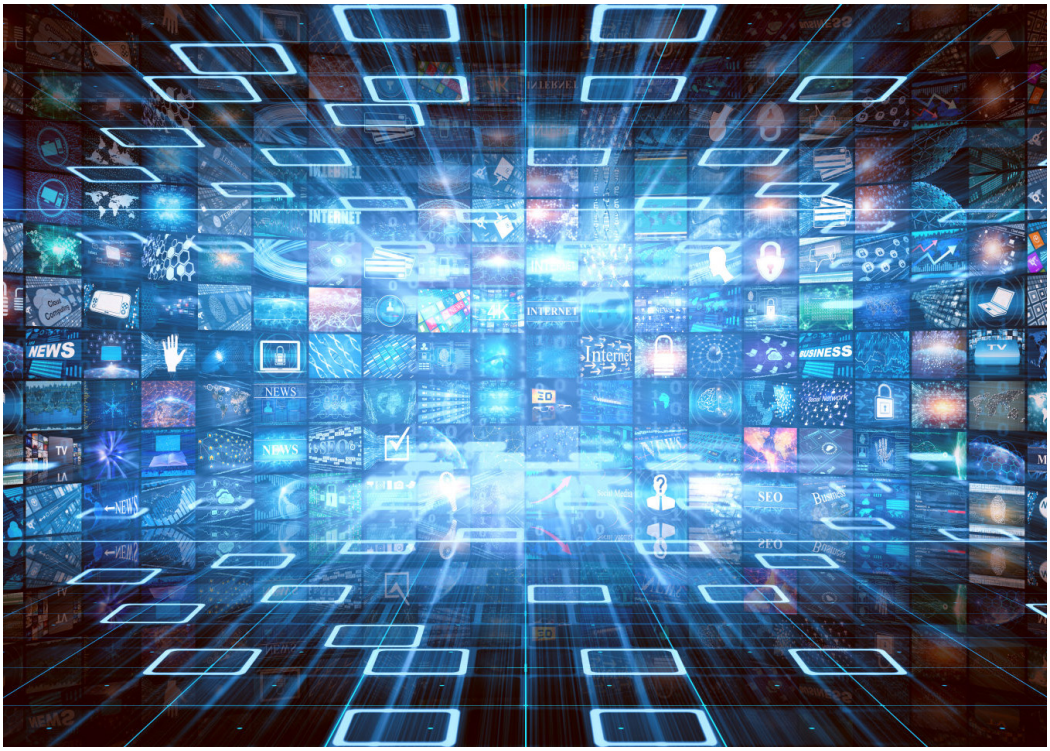
我们希望以百度智能云为纽带，联合技术和产业生态伙伴，通过开源开放平台降低 AI 开发的门槛，加快人工智能融入传统产业，帮助越来越多的行业大步快跑，加入产业智能化的大潮。

云智技术大发展
让视频应用与体验跃入新阶段 >>

视频作为人类最直接、最容易感知世界的一种方式，已广泛应用在日常生活和工作中。从整个视频发展史来看，过去的很长一段时间，最喜闻乐见的视频应用场景，主要集中在传媒业、娱乐业等大众消费领域。

而在最近的 10 年内，随着长视频、短视频、直播等行业应用的高速发展，加速了互联网内容的视频化趋势。人们在互联网上获取信息和娱乐的方式也开始更多的基于视频资源，但这些应用还是更多集中在消费互联网领域，属于一种视频化业务。

但在最近两三年，视频应用领域进一步扩大，进入到越来越多的行业日常业务中，更多业务正在视频化。例如在智慧城市领域，基于视频数据进行的污染治理、消防预警等；在电力领域，基于视频信息进行的线路自动巡检、风险预判；在交通领域，基于视频信息进行的拥堵治理和优化；在教育领域，特别是疫情出现后，很多课程以线上视频形式进行，企业内部沟通及知识分享培训也基于直播或在线视频进行。



这种发展趋势的技术驱动力是云计算和人工智能技术的空前发展，推动视频的应用和体验进入到新阶段。该阶段的核心特点如下：

内容高清化

人们对于内容高清化的需求越来越高，甚至可以说是永无止境；随着 5G 及超分技术的发展，视频正在逐步从高清升级到 4K/8K 超高清阶段。

体验沉浸化

最新的 CDN、边缘计算、低时延通信、RTC、编解码、AR/VR、5G 和 AI 等技术的发展，让人们在观看视频时，真正可以身临其境，实现与视频内容的全方位互动。全景沉浸式视频正在电视、直播及游戏等更多场景中得到应用。

分析智能化

相比文字和图片数据分析，视频数据的分析有着更高的复杂度，过去对于视频内容的理解分析往往依赖人工。随着人工智能技术的发展，对于视频这类非结构化数据的分析处理也上了新台阶，可以实现自动化对视频内容进行结构内容提取、审核及知识挖掘等。

视频是一个“重”数据应用，“重”在两块：一方面，是对海量视频数据的采集、存储、传输和分发提出了极高的诉求；另一方面，是需要强大的数据智能化分析与处理能力，对语音、图像、OCR、知识图谱等全面的智能化技术非常依赖。视频数据作为一种专门类型的数据资产，在企业中占比越来越高，为了更好的管理好视频数据资产，挖掘其中的业务价值，更好的帮助业务发展，企业开始统一规划与构建云智一体的视频平台。

而百度在视频领域有着长期丰富的业务实践，在长视频、短视频、直播，以及自动驾驶、智能交通等业务领域都有相应的技术积累。同时，百度在长期实践中构建了业界最全面、最领先的 AI 能力，再加上优秀的云、边、端等资源和能力，百度已然具备云智一体的完整能力，是帮助企业实现业务视频高清化、沉浸化、智能化的最佳合作伙伴。

百度智能云期望能够和企业、开发者及合作伙伴一起，在视频领域共同创新、共同进步，加速产业的智能化转型。

云智一体的
全场景视频技术解析 >>

3.1 5G+AI 驱动下的视频行业发展趋势

3.1.1 消费类视频的云智一体化诉求

如今在消费类视频应用领域，沉浸式视频体验和智能化数字孪生成为主流趋势。在趋势的裹挟下，消费类视频面临着诸多挑战：

一是随着超高清时代的来临， 4K/8K 等超高清视频的市场需求每年都在万小时量级以上，而目前量级只有千小时级别，视频内容的数量和质量都无法满足超高清时代的需求。

二是随着视频内容的海量增长，传统的人工编目效率根本无法跟上视频增长速度，视频内容的生产、管控和播放等环节都迫切需要更智能化、自动化和更丰富的技术手段来支持。

三是针对视频内容，需要解析更多的结构化信息，才能实现更精准的算法推荐。

面对上述挑战，只有依托云的大规模算力，用 AI 技术带动内容生产与管播等环节的智能化升级，消费类视频才能实现内容质量与数量上质的飞跃。目前，很多行业互联网企业都已经意识到并开始积极应对这些挑战，但离成功依然遥远。

（图3.1.1-1 消费类视频应用的云智一体化诉求）

3.1.2产业类视频的云智一体化诉求

视频应用在产业类服务场景的外延化趋势明显。早期在传统产业里，“视频 + 联网”主要是在监控场景中使用。如今，视频应用已经深入到生产、工作、管理等产品场景的方方面面，工艺优化、质量检测、安全巡检等都是目前比较成熟的产品类视频应用场景。

除了服务场景的外延化，视频的产业参与度也越来越深。传统的视频监控，只解决了视频的接入问题，后续仍然需要由人来查看和判断。而随着“业务流程数字化与智能化”概念的提出，视频正在逐渐成为一种对物理世界进行智能化改造的手段：通过对设计、生产、运维与分析决策等全场景进行视觉感知，生成海量视频图片等多维度数据并接入边云平台，由云边端算力来承载和调度，由人工智能技术处理视频内容的结构化，对视频语音语义进行自动化理解，实现智能化决策。

“业务流程数字化与智能化”的实现，需要大带宽、低延时、高安全存储和高并发算力的“云”来承载，还需要端智能检测、多模搜索、辅助分析与自动化决策的“智”来引领。

（3.1.2-1 产业类视频应用的云智一体化诉求）

3.2 智能视频云 3.0 全景图

随着时代的发展，全面视频化的时代线条越来越清晰，为了应对海量数据接入并提升应用效率，视频数据的数字化与智能化成为必然趋势。

首先，从技术角度来看，数据密集型和人工智能密集型的工作场景已然大规模出现。但由于终端设备型号各异、数据采集规模异常巨大，故而要求设备接入、数据采集与解析等环节都需要更高的异构融合性、算力高并发性以及自动化等能力，这对端边云的承载和大规模算力都提出了严峻挑战。而视频数据要实现自动化解析、智能化建模与应用，也需要强大的算力和传输能力做支撑。最终，这些双向需求推动了端、边、云、智的充分融合，端边云逐渐一体化、云智逐渐一体化。

其次，从业务流程来看，人工智能技术已经可以作用到业务流程的各个阶段。有了颗粒度更细的场景算力与人工智能算法做加持，问题的解决能力也会持续提升，逐渐由只能解决通用场景问题，发展到可以解决更细分、更个性化的场景问题。可以说云智技术的发展，一方面扩展了视频应用的外延范围，另一方面也加深了视频在传统应用领域的作用和价值。

需要注意的是，虽然视频领域有了更多的新技术，但企业却面临着如何低成本、高效率的利用这些技术去实现业务创新的问题。因为，视频技术的种类非常多、复杂程度高，企业盲目创新会造成研发成本和资源投入的高企，普通企业从时间、成本、资源等方面都无法承受。

那么，如何消除架在企业视频创新需求与资源成本投入间的鸿沟呢？百度智能云给出了答案：智能视频云 3.0——云智一体的、端到端的、全场景的智能视频平台级解决方案。

为什么说是智能视频云 3.0 呢？1.0 时代解决的是视频传输问题，2.0 时代更多的提供针对单一场景的功能化组件。而 3.0 时代是视频技术与能力的全面升级，百度智能云希望用深入到传统与新兴视频领域的、云智一体的、一站式平台方案，推动企业视频应用的可持续创新。

（图3.2-1 智能视频云3.0全景图）

3.2.1智能视频3.0的特点

云智一体的 3.0：百度智能云开放了智能编码处理、智感超清、低延时直播等 10 余种云智产品组件、100 多个 API，以及视频创作分发平台与视联网感知平台等两大场景平台，希望借助全场景与平台化的智能视频技术，为企业提供更完整丰富、一站式的视频平台开发与应用集成能力。

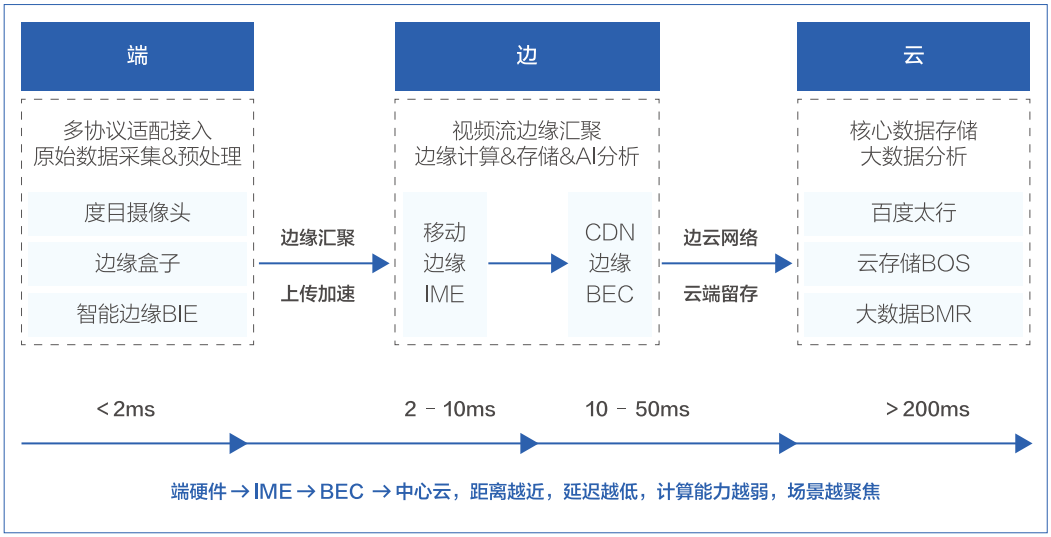
生态开放的 3.0：百度智能云携手众多合作伙伴，共同打造“智能视频云 3.0”，用“平台 + 应用”的模式，为企业提供端到端的解决方案。目前，已形成覆盖互动娱乐、实时通讯、生产管理，安全监管等 10 多个视频场景的解决方案，助力企业快速、高效的实现视频应用的落地和交互。

3.2.2 智能视频3.0背后的端边云一体化基础架构

构建智能视频 3.0 平台的核心驱动力是数据和应用的分布式需求，不同时延和算力层次也对基础架构提出了高要求。

为满足以上需求，百度智能云构建了完整的端边云一体化基础架构体系：端侧，提供轻量级终端，如 AI 摄像头和边缘盒子，并提供轻量级的算力环境；边缘侧，提供小算力环境，满足长距离场景的低时延需求，如移动边缘 IME 和 BEC（CDN 边缘产品）；云侧，提供全栈的大数据分析、AI 计算、高性能存储等产品矩阵。

这样的一整套算力分布式基础架构，既可满足端侧预处理与边缘侧汇聚计算的需求，也可确保视频数据在云中心的各种复杂运算。



（图3.2.2-1 端边云一体的基础架构）

3.2.3 基于云智一体的双平台

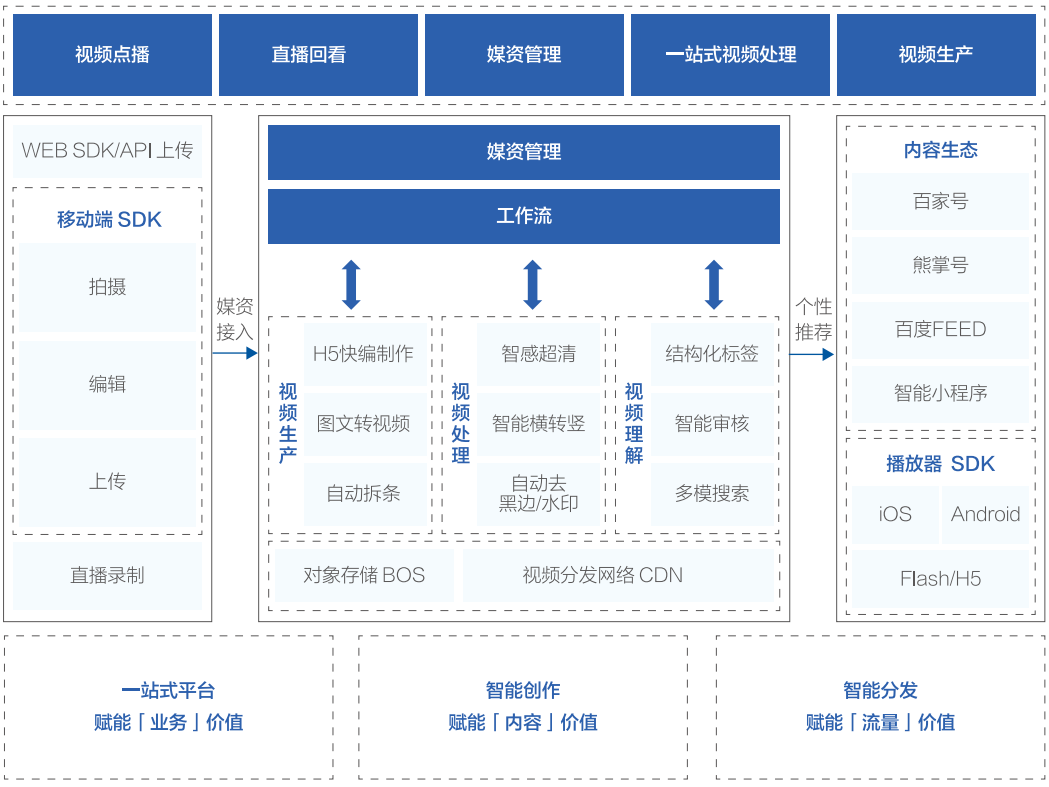
智能视频云 3.0 中的“云”，通过端边云一体、异构融合的基础架构提供先进的云计算能力；“智”则以业界领先的 AI 能力完成对视频语音、图像、文字等多模态信息的智能分析、理解、生成。同时，基于云智一体的基础架构，智能视频云 3.0 打造了面向不同场景应用的视频双平台——视频创作分发平台和视联网感知平台，为企业提供一站式和平台化的技术服务。

- **视频创作分发平台**

视频创作分发平台面向泛互联网、泛媒体等行业，聚焦业务流和媒资管理，覆盖视频的接入、生产、处理、审核分析和个性化推荐等全流程，实现媒资管理和工作流的智能化改造。通过媒资结构化处理和多摸精准搜索，媒资利用率可提升 40%，盘活媒资，最大化发挥内容价值。

相对于传统采编发的管理软件，视频创作分发平台利用 AI 能力对数据进行深度理解分析，加快数据流动，盘活媒资数据利用率；并为业务流提供可视化控制台，用户可以通过简单灵活的视频编排配置流程，轻松应对业务的个性化需求，适配移动化、多终端、多形态的业务场景。

VideoWorks 作为百度智能云的音视频统一入口，提供了音视频上传存储、生产处理、媒资管理、分发播放等覆盖视频全生命周期的 PaaS 服务，主要适用于视频点播分发、视频生产处理、媒资智能管理三大应用场景。



（图3.2.3-1 视频创作分发平台）

- **视联网感知平台**

视联网感知平台面向传统产业，核心价值是对泛视频设备和数据流实现统一的连接、分析和管理，打造面向生产 / 控制 / 治理场景的通用基础平台，为企业生产、园区管理、城市治理等场景提供平台化支撑。该平台特点如下：

开放性。广泛的设备接入兼容性，支持 10 种 + 主流视频协议和多厂商设备；具备高扩展的处理和分析能力，涵盖 20 种 + 媒体处理能力和可扩展性，为业务生态提供统一和开放兼容的平台化服务。

全面智能化。依托 200 种 +AI 视觉感知能力，可对视频数据进行结构化解析，满足千亿级特征数据规模的毫秒级查询，实现对视频内容的高效语义化理解和获取；依托百度强大的知识图谱能力，与应用融合，实现对感知内容的知识化处理，提升视频应用的认知能力。

端边云融合。融合边缘，提供最优的云智一体架构，使资源、数据、感知能力在端边云充分融合，提升管理便捷性和秒级告警能力，降低网络成本。

（图3.2.3-2 视联网感知平台）

3.3 智能视频云 3.0 中的“云”能力

3.3.1 视频分发基础设施CDN

- **1000+ 大节点，100T+ 带宽储备**

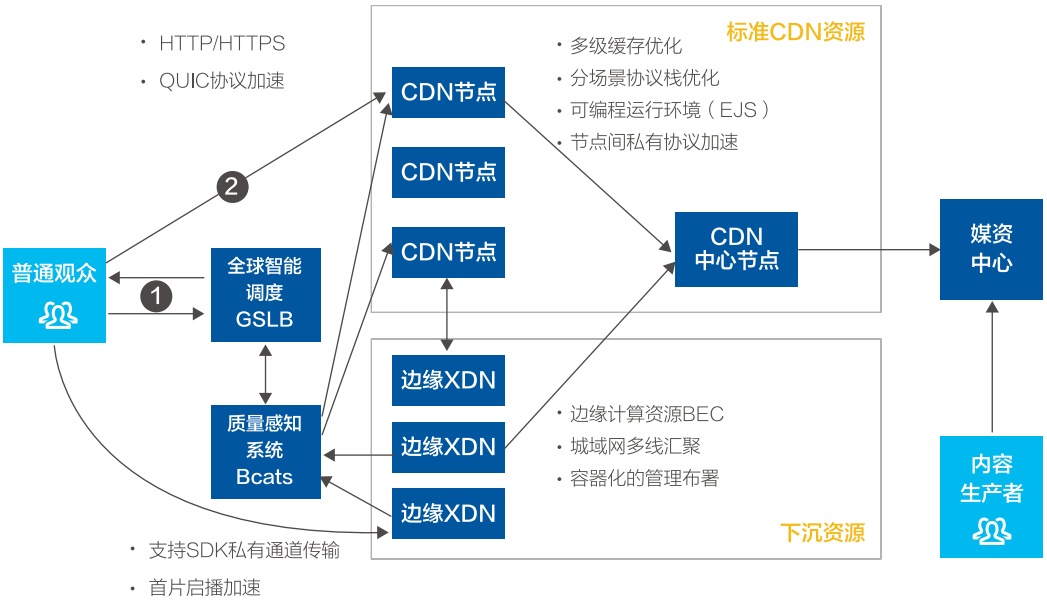
智能视频云 3.0 的视频分发基础设施 CDN，具有灵活的资源扩展能力，在已有的 1000+ 标准节点基础上，资源的使用能力持续下沉，可以与边缘计算的资源深度结合，做到流量和算力的双加速，还将一些视频编解码能力下沉至边缘节点。

- **全球智能调度 + 低时延感知**

智能视频 3.0 的视频分发基础设施，拥有全球智能调度系统和低时延质量感知系统，能做出近乎实时的智能调度，实现就近 CDN 节点的内容调度和快速触达。同时，通过对中间协议和多级缓存的优化，大大提升了视频内容在不同场景要求的访问体验水平。

- **系统可编程与分场景协议调优**

智能视频 3.0 的视频分发系统具有可编程特性，这意味系统可实现自生长，从而更好的应对视频应用的多样化场景需求，提升系统的敏捷开发水平，支撑起整个智能视频云 3.0 包括智能编码、智感超清等在内的几十种产品能力。



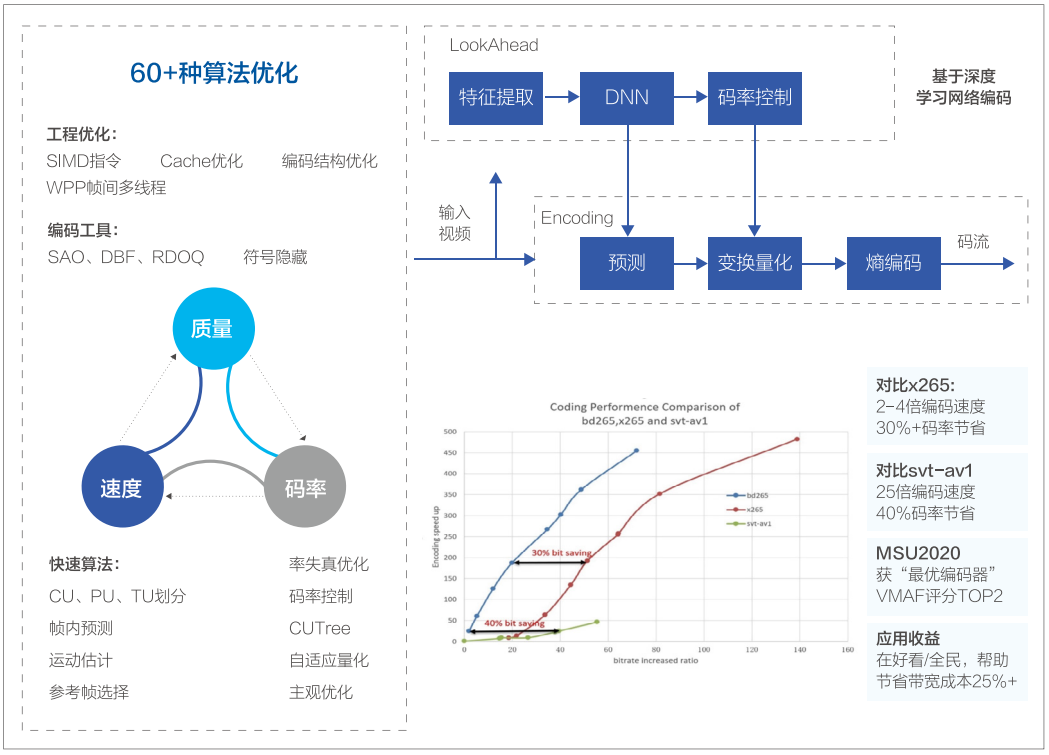
（图3.3.1-1 视频分发基础设施CDN）

3.3.2 智能编码

为帮助视频生产者在低成本与视频极致体验间取得最优解，百度智能云提出了 BD265 智能编码方案。BD265 智能编码技术利用深度学习网络提取视频特征，生成自适应编码参数，降低码率的同时提升视频的主观体验。BD265 研发了 60 多种算法，在视频质量、编码速度和编码码率间取得一个优化解；采用了多层并行构架包括 SIMD 指令并行、WPP 帧内并行、帧间并行等，提高编码速度。质量方面，BD265 采用自适应量化、码率分配、主观优化来提升视频质量。

相比于业界通用的开源编解码技术 X265，BD265 在视频编码速度上提升了 2-4 倍，码率节省 30%，带宽成本节省 25%+。以好看视频为例， BD265 在实际应用中节省带宽成本 25%+，同步提升了视频画质。可以说，BD265 智能编解码技术在视频内容的质量、码率和速度之间取得了最优解。

在全球顶尖的 MSU2020 世界视频编码器大赛中，BD265 获评“最优编码器”。目前，这套智能编解码技术已经广泛应到点播 / 直播等视频应用场景中，得到用户的高度认可。



智能编码BD265，同画质降低码率30%+，提速2x+

（图3.3.2-1 智能编码BD265）

3.3.3 智感超清

为了匹配超高清时代的视频体验极致追求，百度智能云打造了“智感超清”产品。核心价值就是利用去噪、去划痕、上色、画质增强、超分辨率等手段进行老片修新，实现对视频画质的提升和超高清内容的转换。智感超清背后涉及的技术众多，具体如下：

- **超分辨率**

基于 U 型神经网络框架，提出一整套级联提升网络框架，可对视频空间分辨率进行从粗到细地提升。

- **自适应去燥**

视频数据中的噪声模式难以人工模拟，通过自研的无监督去噪算法，可以摆脱有监督神经网络训练对成对噪声与干净图像的依赖；模型还设计了可用户交互的控制参数，支持用户调节适应不同噪声。

- **画质增强**

细节纹理增强，采用深度残差卷积神经网络，对视频内容进行特征提取与重建，对画面的饱和度与原画质视频的压缩失真进行细节还原；感兴趣区域（ROI）增强与编码，通过 ROI 区域的实时检测与跟踪技术，对 ROI 区域进行局部画质增强，同时编码时码率分配进行 ROI 区域优化，保证 ROI 区域视觉主观质量的提升。

- **SDR 转 HDR**

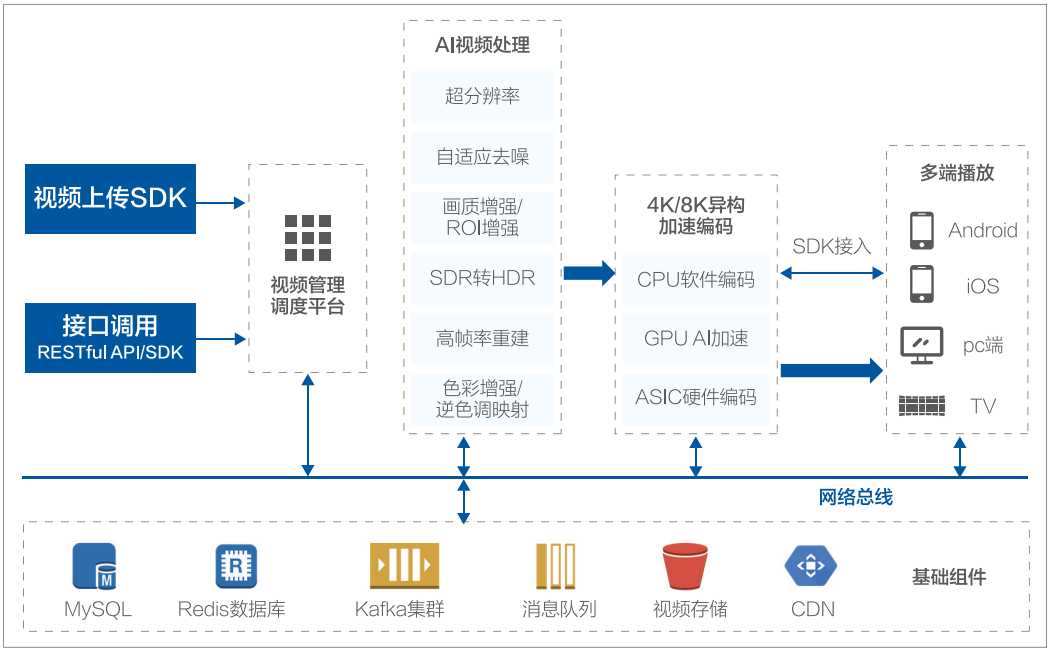
基于全局与局部信息融合的门控残差网络单元，构建端到端的 SDR 视频源到符合 HDR 标准视频信号的上变换网络模型，高效地将传统 8bit 位深 BT.709 色域空间的信号，上变换到 BT.2020 色域空间下 10bit 及以上位深的信号。

- **高帧率重制**

基于 MEMC（运动估计与运动补偿）的高帧率提升技术，通过对运动方向预判，据此进行插帧，提升画面流畅度与观屏舒适度。

- **超高清编码**

采用异构计算 4K 编码技术，基于 CPU+GPU+ASIC 的异构加速计算框架，对视频的处理和编码进行加速处理。



- **超分辨率：**
  U 型神经网络框架，级联提升网络框架，从粗到细提升
- **自适应去燥：**
  无监督去噪声算法
  可交互控制参数，支持用户调节
- **高帧率重制：**
  基于 MEMC 高帧率提升技术
- **画质增强：**
  深度残差神经网络，内容特征提取与重建
  饱和度与压缩失真还原
  ROI 区域增强与编码
- **超高清编码：**
  采用 CPU+GPU+ASIC 的异构加速计算框架
  支持国产昆仑 K200 芯片 AI 推理

智感超清，大幅提升画质，加速超高清发展

（图3.3.3-1 智感超清技术）

百度智能云自研的智感超清技术，在 CVPR、ECCV 等全球大赛上均取得了好成绩。同时，百度智能云也在牵头《智能视频处理系统通用技术规范》产业标准的组织工作，致力于制订出普惠大众的通用标准技术。

目前，智感超清技术已经广泛应用在电影频道、得物 APP 等泛媒体与泛互联网场景中。帮助电影频道完成了众多老旧影片的修复换新工作，在提供更好用户主观体验的同时，减少了人工逐帧修复的成本，提升处理效率 10 倍 +。

（图3.3.3-2 国际比赛奖项及第三方权威认证）

3.3.4 实时音视频通讯

2020 年整个世界遭遇了“新冠”黑天鹅事件，在疫情期间，人们的工作与生活中迫切需要解决物理距离的问题，实时音视频通讯正是在这个背景下获得了高速发展。比如视频会议、百度无人驾驶，都是通过云端对视频应用进行管理，实现低延时的互动体验。智能视频云 3.0 中的“实时音视频通讯”产品具有支持即时通讯、Android/iOS/Windows/Web/ 小程序等跨平台互通、屏幕分享、旁路直播及云端录制等特点。

“实时音视频通讯”产品具备抗弱网、端到端延时性能监控、高质量音视频通话等特点，具体如下：

- **弱网支持**

自研抗弱网算法，根据上下行网络链路恶化程度因子，调整 FEC、NACK、码控、超清权重选择因子，进行抗弱网抖动算法，使视频抗弱网可达 60%、音频可达 80%，弱网环境下保证高质量通信。

- **高性能**

通过优化后的混沌神经网络算法规划最短路径、实现动态平衡路由选择，将端到端延时稳定收敛到 200ms 左右，并在单个节点故障时，自动重建路由，保障 Server 的可用性。

- **WHIP 协议转换**

  解决了信令交互复杂、房间管理业务量大、数据库操作频繁等问题，可支撑每个房间在线人数达百万级别。

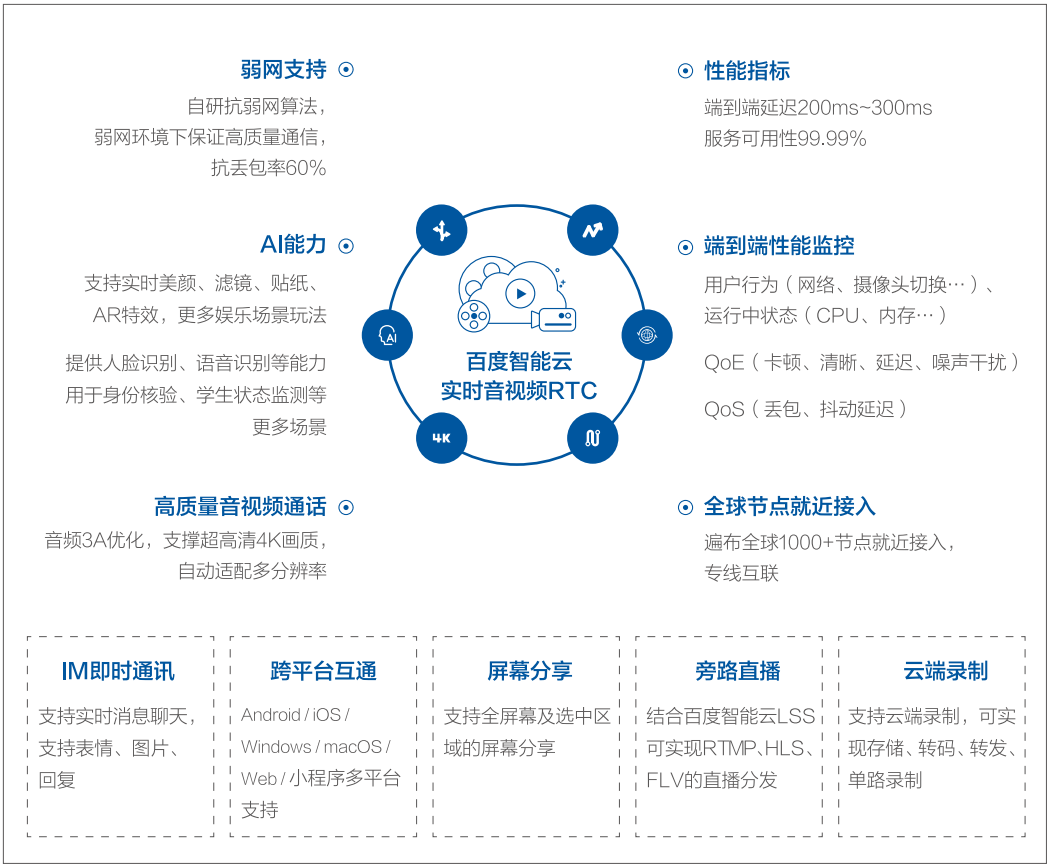
- **首屏秒开处理**

  关键帧缓存算法，保证 800ms 左右拿到首帧并渲染到屏幕，大幅提升用户体验。

- **AI 能力**

  支持实时美颜、滤镜、贴纸、AR 特效等娱乐场景玩法; 提供人脸识别、语音识别等技术，用于身份核验、学生状态监测等更多场景。

- **端到端性能监控 –** 端到端实时监控设备运行状态、QoE、QoS 以及用户行为。
- **高质量音视频通话 –** 音频 3A 优化，支撑超高清 4K 画质，自动适配多分辨率。
- **全球节点就近接入 –** 遍布全球 1000+ 节点，就近接入、专线互联。



（图3.3.4-1 实时音视频通讯）

3.3.5 低延时直播

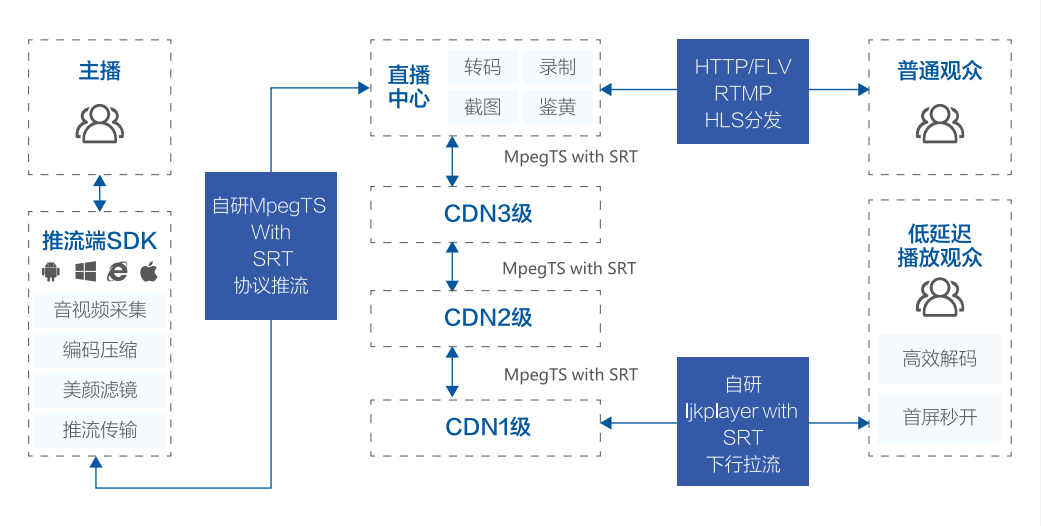
实时音视频通讯实现了线上超低延时的互动体验，但面对千万级并发量，如何应对？“低延时直播”就是满足此类场景的视频产品，其背后核心是百度智能云自研的通讯格式封装和协议栈优化技术。

直播包括采编码、推流、分发、接受、解码和渲染等环节，要在直播中实现时延的突破、分发和接收的时延变动等需求，主要取决于链路时延的抖动、协议栈的优化以及 CDN 节点资源等情况。

“智能视频云 3.0”低延时直播产品采用全链路 SRT 协议技术传输方案，用户从推流、播放及 CDN 内部传输等都采用 SRT 协议和 MPegTS 封装，从而将网络丢包的可能性降到最低，把延迟做到最小。

在用户通过 SRT 推流到 CDN 后，CDN 内部支持 SRT 流转成主流格式的直播协议，用户也可以使用http-flv、hls、rtmp等格式完成播放、录制、转码、截图等操作。相比传统的rtmp和flv存在5-10s的延迟，SRT 有效把时延控制在 1s 以内。

基于 SRT 协议的低延迟方案不仅时延低，结合 BD265 后，在同等清晰度下，带宽可节省 30%。



卓越抗弱网能力：自研 mpegts 数据格式封装、SRT 协议推流，将网络丢包的可能性降到最低，改善弱网环境推流服务品质。

首屏秒开：CDN 1 级先将编码参数信息发给播放器，播放器解析到编码信息，待下个关键帧到达后即可快速播放。

协议兼容：实现多种视频流协议的输入与输出， UDP 和 RTMP 流能够顺利融入 SRT 系统，视频转发更灵活性。

规模并发：与 Http/flv 支持的并发规模一致，支持千万级规模并发。

（图3.3.5-1 低延时直播）

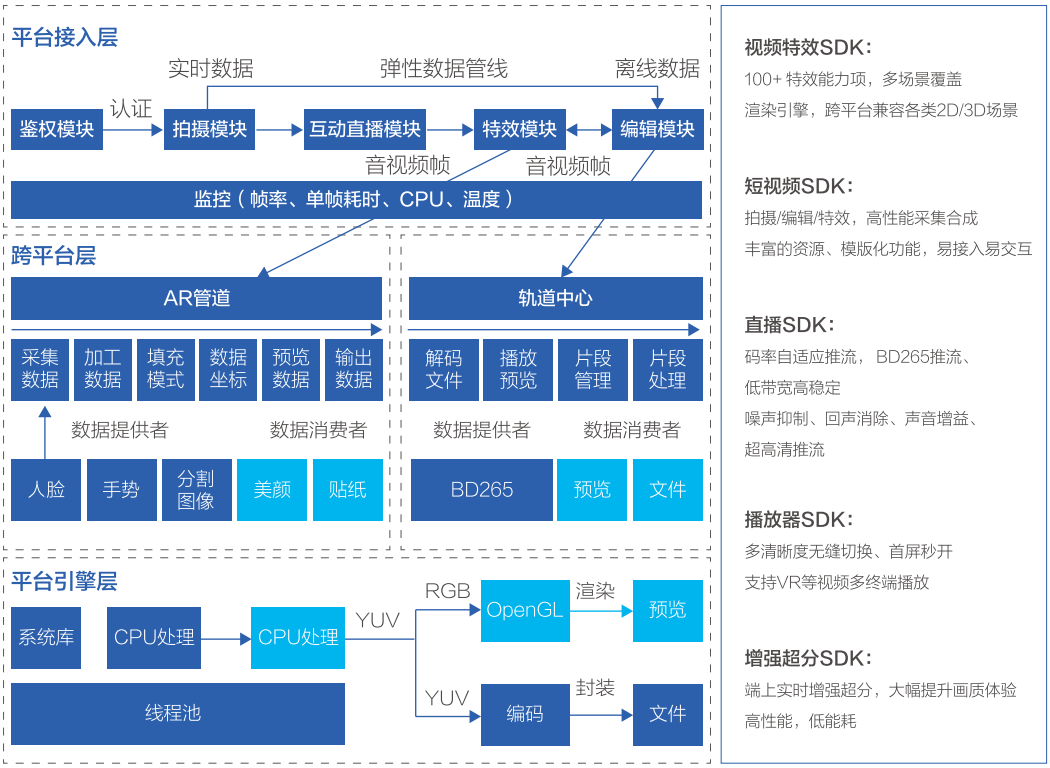
3.3.6 智能视频SDK矩阵

在全民拍摄的泛直播时代下，轻量化美颜、智能剪辑、低码率高品质推流、沉浸式播放体验等成为音视频终端业务不可或缺的基础能力。因此，百度智能云面向不同业务，提供丰富的 SDK 解决方案，支持业务发展。受益于终端算力的不断释放，使得终端设备运行如 PaddleLite 的人脸算法、语音增强算法、视觉增强算法、BD265 编码算法及全景渲染算法等百度自研算法变为现实。在这些底层算法的加持下，终端能力得到大幅提升。

智能视频终端框架正在经历从模块化向智能化的演进历程。平台接入层通过模块化解决多 SDK 复用和数据管线组合的问题，提高内部协同效率，提升产品按需交付的灵活度；辅以横贯模块间的数据监控，运行质量可一目了然。

跨平台层重点解决多终端数据处理一致性体验的问题。基于生产消费模型，将自研算法抽象成独立算子（如人脸算法），经过管线中不同处理节点，最终将数据转化为业务效果（如美颜等）。

针对终端数据的高效率处理，离不开平台引擎层提供的持续动力。终端并行计算技术，让 CPU 与 GPU 在时间和空间上得到更好的协同。



（图3.3.6-1 智能视频SDK矩阵）

3.4 智能视频 3.0 的“智”能力

回顾百度智能视频云 3.0 的全景架构图可以发现，智能视频技术在其中起着承上启下的重要作用，智能理解分析、智能视频生产与智能交互生成等关键 AI 技术充当着平台化与场景化创新应用的技术基石。

3.4.1 智能理解分析

- **视频中的单模态理解**

视频理解分析时常会遇到两个挑战：一是需要处理的视频中对象太多，比如各种环境下拍摄的人脸，处理方法也不尽相同；二是视频样本素材少，无法支撑起一个有效的训练模型，这在工业场景尤为明显。挑战带来的问题是，需要针对每个场景都单独建模，而素材稀少又造成模型训练效果不理想。

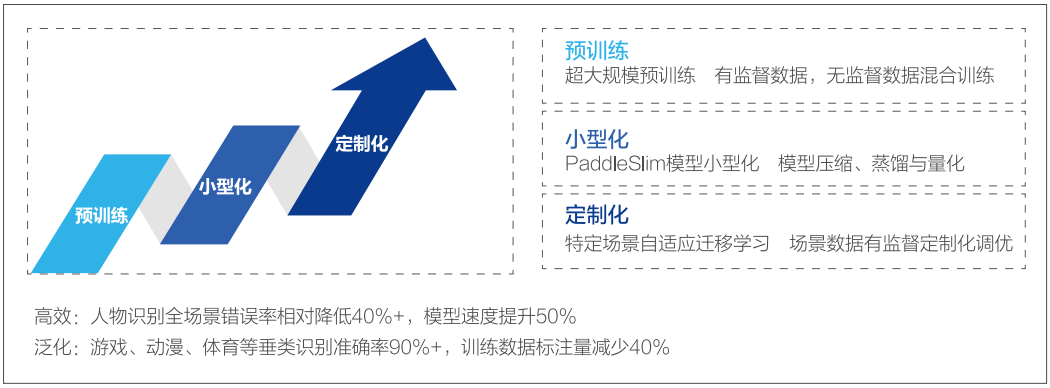
基于以上挑战，百度智能云提出了一套“预训练 – 小型化 – 定制化”的研发流程：

- **预训练**

以视频中的人物理解为例，通过百度资源库获取海量视频场景的人脸数据，依托百度飞桨平台提供的大规模训练框架，基于所有场景图像、视频等有监督数据和无监督数据进行混合训练，打造预训练模型。由于该训练框架的训练 ID 数可达千万级，因而训练出的模型具有强大的鉴别性能。基于百度海量数据的人脸识别预训练模型，人物识别全场景错误率相对降低 40%+，模型速度提升 50%。

- **小型化与定制化**

模型训练只是手段，应用才是目标。由于设备端规格多种多样，计算能力千差万别，边缘侧适配性成了关键。基于百度飞桨 PaddleSlim 套件，可以对模型进行搜索压缩与量化，将大模型蒸馏为小模型。缩小后的模型，既具备大模型的知识结构，又有通用性特点，结合特定场景的少量样本，即可进行自适应迁移学习，实现完美适配。相比于传统模型训练，基于 PaddleSlim 套件的模型训练，速度提升了两倍。而在游戏、动漫、体育等垂类场景中，识别准确率可达 90%+，训练数据标注量减少 40%。



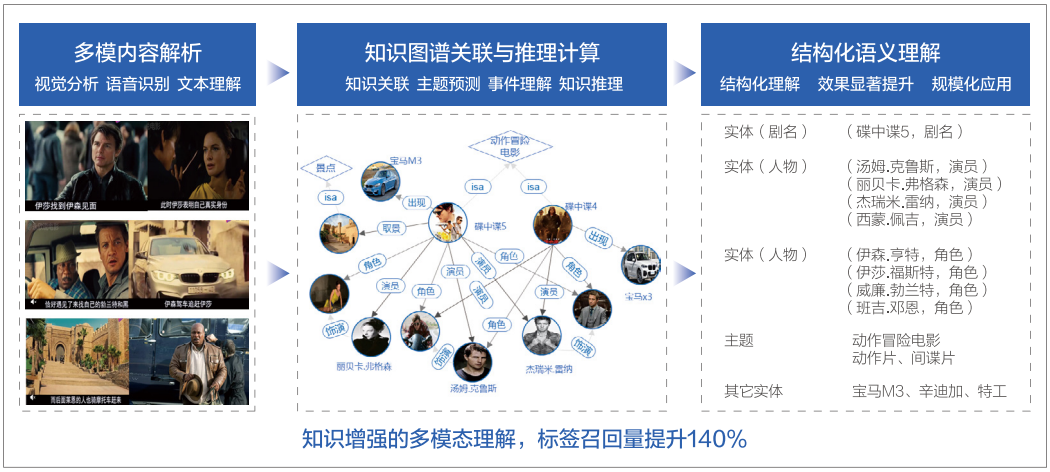
（图3.4.1-1 视频模态理解框架重磅升级）

- **知识增强的视频理解**

基于知识的视频语义理解技术，充分利用丰富、全面的知识图谱提升视频理解效果。该技术通过对视觉、语音、文本的多模态内容进行解析与融合，基于知识关联技术建立与视频理解知识图谱的连接，通过对背景知识的计算与推理，实现对视频的深度语义理解。

知识增强的视频理解技术，突破了传统的基于感知的视频内容分析的局限性，实现对视频的深度结构化语义理解，大幅提升视频理解效果，标签召回量提升 140%。目前基于知识的视频理解技术，已在百度的搜索、推荐、好看视频等多个重要业务中得到大规模应用，取得了显著效果。

智能理解分析综合应用的一个典型案例就是视频智能编目。面对海量信息，广电行业遇到的一大挑战就是在保证准确率的前提下，提升编目效率。视频智能编目通过对视频的多模态理解，算法能自动识别一系列的编目项，编目平均准确率超过 90%，相比于人工编目，效率能提升 6 至 8 倍。一段 30 分钟时长的新闻，人工编目需 4 小时，而智能编目只需要 30-40 分钟。



（图3.4.1-2 知识增强的视频理解）

3.4.2 智能视频生产

- **视觉模态创新技术**

旨在将多种内容载体向视频形式自动转换，提升传播效率。目前百度智能云可提供多种视频生产方式，如图文 / 图表 / 图集静态图片等转视频、长视频拆短视频以及 Video-in（内容智能嵌入）等。

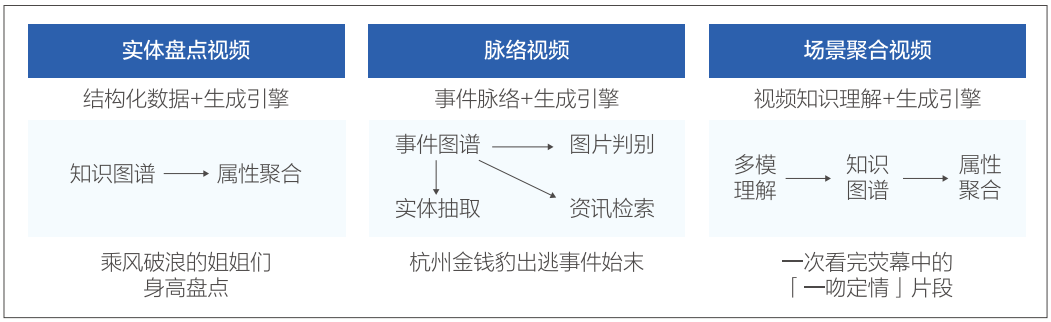
- **知识增强的视频生产**

基于知识图谱的视频生产技术是百度智能云视频生产技术的差异化亮点，主要包括三种：

实体盘点视频：借助知识图谱的海量事实，生产关于实体属性、关系的各种知识类、趣味性的盘点视频。

脉络视频：基于知识图谱中一项核心技术能力——多轮问答技术，可实现事件监测和事件论元抽取能力；通过事件识别、聚类、归一等技术总体构建事件脉络生成技术，达到对客观事件的动态理解。基于事件脉络能力，可生成关于热点事件来龙去脉的报道视频。

场景聚合视频：通过对影视剧的多模语义理解，能自动将某一类场景镜头进行聚合，呈现作品的精彩片段。



（图3.4.2-1 知识增强的视频生产）

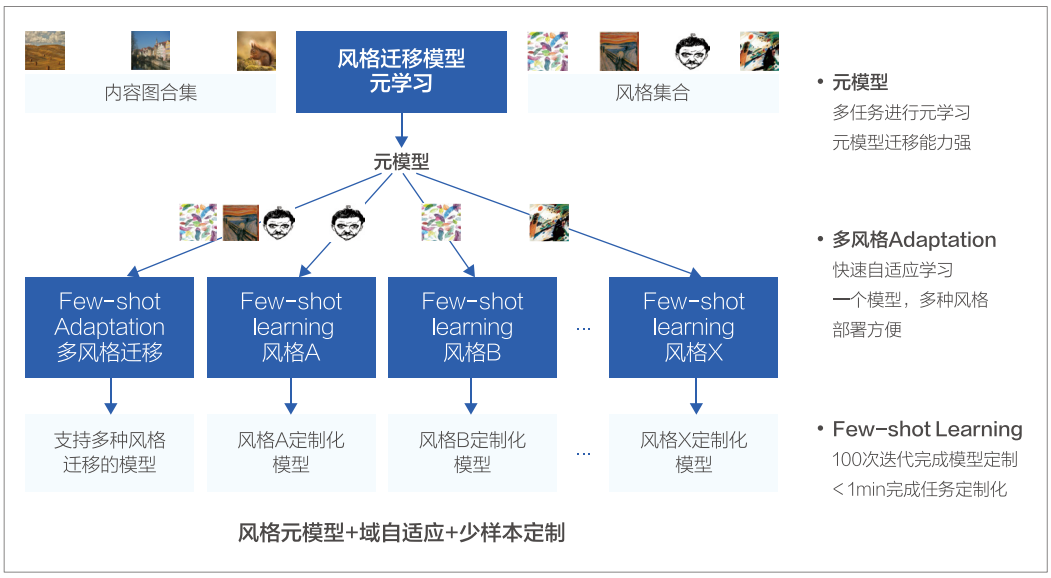
3.4.3 视频交互生成

智能视频云 3.0 拥有丰富的技术能力，全新构建了交互与生成 AI 能力矩阵，从 AR 特效技术到新型驱动特效、风格特效和多模态生成，全面支持各类泛娱乐场景的视频交互与生成需求。特别是数字人技术，融合了众多业界领先的语音、视觉和自然语言理解等方面的 AI 技术，是百度 AI 技术的集中呈现。

（图3.4.3-1 交互生成矩阵）

- **视觉模态的任意风格化**

视觉模态的任意风格化有两大挑战：一是形成任意风格之前都要先学习一个元模型；二是要在有限样本的条件下进行模型迁移。智能视频云 3.0 可以完美解决这些问题，基于元学习、多风格 Adaptation 技术、Few-shot Learning 等技术，使元模型具备多任务学习和快速自适应学习能力，实现元模型的快速迁移，从而可根据需求快速定制不同风格的高质量视频。



（图3.4.3-2 视觉模态的风格化）

- **数字人技术**

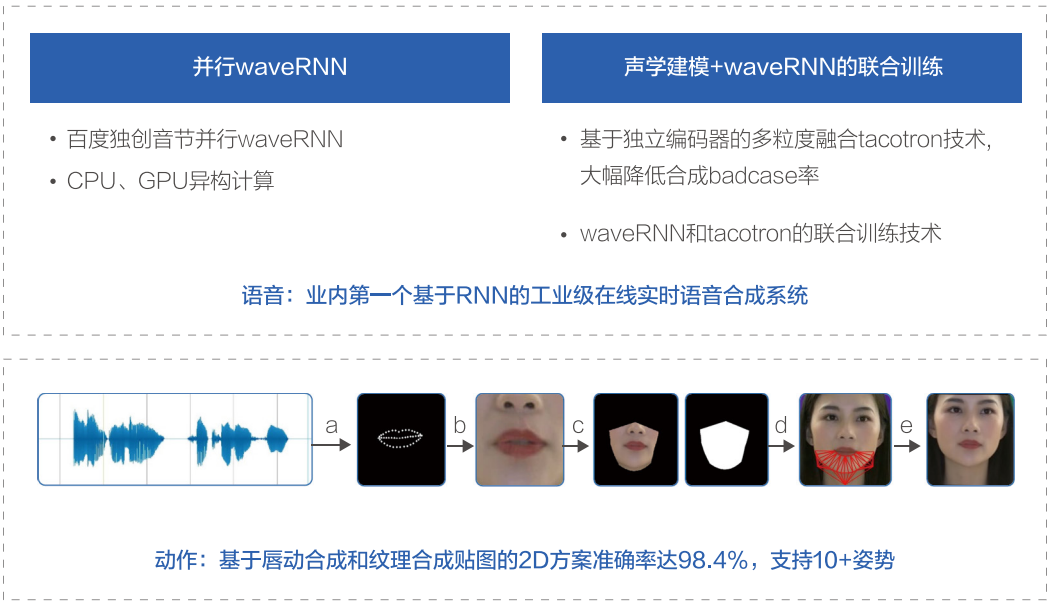
区别于传统的文字、语音等对话服务，数字人技术以更加拟人或还原真人的形式，提供更有温度的服务。智能视频云 3.0 的多模态技术，将虚拟数字人形象与语音合成、唇动生成等技术相结合，低成本制作形象丰富灵动、线上线下多端适配的数字人。并可通过智能对话引擎直达用户需求，更快速、直接的解决用户问题或辅助用户决策。领先的数字人技术需要解决四大挑战：

形象丰富

目前数字人的形象分为 2D 卡通、2D 真人影像、3D 卡通、3D 高精建模等 4 种。其中应用最广泛的是 3D 高精建模人像，一方面可高度还原真人，实现风格定制，另一方面可避免肖像权问题。其次广泛应用的是 2D 真人影像，制作成本相对比较低，但需要模特肖像授权，一般应用在替代真人的严肃场景。“3D 高精建模”分为重新构建和微调两种制作方式，模型构建需要大量真人原型数据采集，基础人像构建完成后则需要针对面部、皮肤、发型服装等元素进行材质贴图方面的调整。“2D 真人影像”采集绿幕影像数据，通过唇部抠图、打点切分等数据处理步骤，完成模型的构建。

交互自然

百度智能云拥有业界首个基于 RNN 的工业级在线实时语音合成技术，让数字人的语音合成更拟人、更自然、更动听；同时融合了下巴、面部唇动合成技术，通过高精度的纹理合成，让语音驱动的唇动自然度更高。语音和唇动的匹配准确率可达 98.4%，几乎可达到完美的口型声音一致。



（图3.4.3-3 数字人的四大挑战-交互自然）

制作高效

百度智能云数字人技术可以基于 5 分钟的视频物料就能定制一套自有 IP 的真人形象。同时，还提供快速复制多 IP 真人形象的技术，将单张人像迁移到虚拟人，就能生成多个不同 IP 的真人形象。

（图3.4.3-4 数字人的四大挑战-制作高效）

环境普适

目前数字人的应用载体较为广泛，除了PC端和手持移动端设备以外，百度智能云还自研了智能云屏、智能云席等数字人载体。为更好模拟真人交互，云屏、云席都集成了用来检测人脸和唇动的定制摄像头及拾音麦克风，可更好的识别主说话人，解决“鸡尾酒会问题”。云屏高度 170 厘米左右，1:1 还原真人大小，可批量在网点铺设。今年两会期间，百度智能云与央视网合作推出虚拟数字人小 C，策划了“两会 C+ 真探”的虚拟记者采访项目，一时受到市场广泛关注。

（图3.4.3-5 数字人的四大挑战-环境普适）

智能视频云3.0的
三大行业应用场景 >>

为充分发挥云智技术的价值，百度智能云深度到参与视频相关产业的智能化升级中，基于云智一体的智能视频云 3.0 平台，为泛互联网、泛媒体和泛产业等领域，提供产业类智能视频应用解决方案，覆盖互动娱乐、内容生产、智能分析、远程实时通讯、生产管理和安全管理等场景。

4.1 云智一体的视频技术推动泛媒体产业升级

在全面视频化的时代，政策驱动传统媒体向融合、智慧、移动、高清化转型，传统媒体正在向自主可控、互动、智能型媒体发展。百度智能视频云 3.0 以及智能视频全链条 AI 技术，为泛媒体视频的发展注入生命力，助力产业全面升级。

4.1.1 泛媒体产业格局：应对新挑战、把握新动能

- **新挑战**

竞争加剧。当前流量红利已经见顶，用户运营转向存量运营，媒体行业对增量用户的争夺日趋白热化。

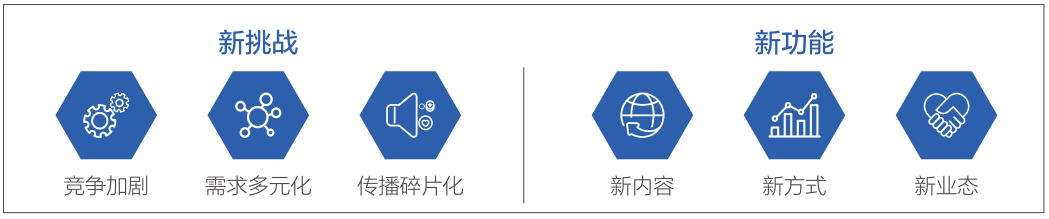
需求多元化与传播碎片化。用户的个性化需求增长迅速，多元化趋势显著；同时，随着传播渠道的形式与数量直线增加，中心化传播逐渐弱化，传播碎片化加剧。

- **新动能**

新内容。当前，直播、短视频等新媒体内容成为主流，AR/VR/AMR 等新内容供给机制跃跃欲试。

新方式。车载、家庭等多屏交互终端持续涌现，更有数字人等新交互技术加持。

新业态。面对新的业务模式，媒体 + 政务 / 服务 / 商务等泛媒体组合出现了新的业务机会，内容逐渐成为一种新产品业态。比如金融媒体，基于专业的内容生产，为银行、证券等金融机构提供垂类内容服务。



（图4.1.1-1 泛媒体产业的挑战与动能）

4.1.2 泛媒体产业演进的三条增长曲线

泛媒体产业存在着用户驱动、科技引领的三条增长曲线，不同的媒体形态和媒体阶段，都在这三条曲线上发展。

- **信息化增长曲线**

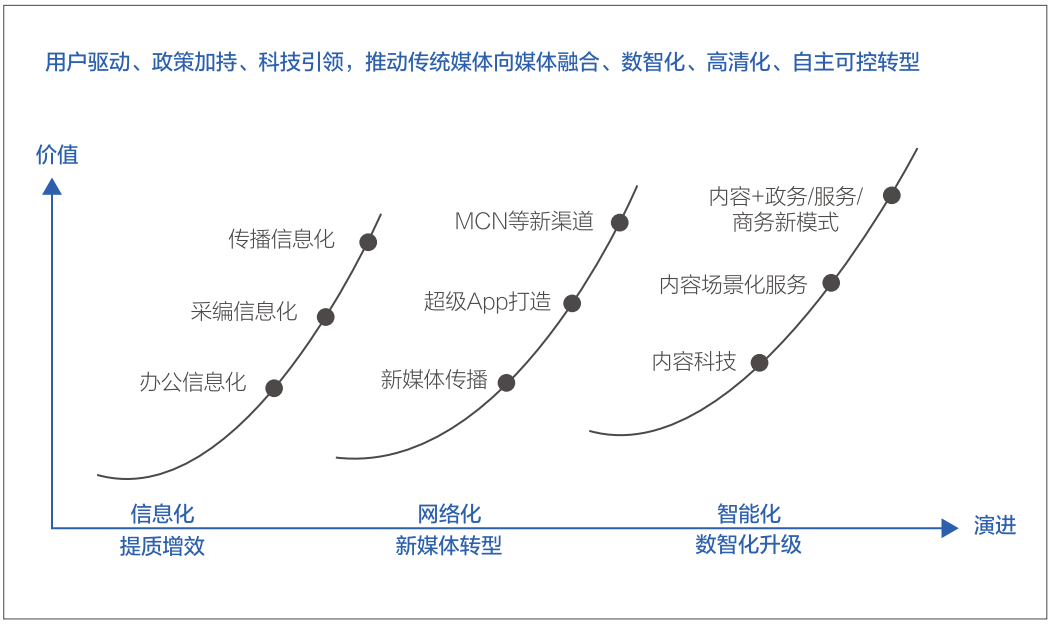
为实现提质增效的目标，传统广电媒体正在做办公、采编、传播等环节的信息化转型。

- **网络化增长曲线**

完成信息化升级的媒体机构，开始朝着网络化新媒体矩阵方向发展，广电、报业等诸多机构都在谋划自己的新媒体矩阵，一些省市广电传媒希望打造出超级 APP，并尝试通过 MCN 渠道来聚合更多内容。

- **智能化增长曲线**

人工智能引领的产业转型，产生了很多新的业务形态，既有内容 + 政务 / 服务 / 商务的新业务模式探索，也催生了更多内容的场景化服务。一些内容科技类企业把内容的全链条生产和全渠道分发作为立足点，为内容的供给方和消费方提供技术服务。



（图4.1.2-1 泛媒体产业增长曲线）

4.1.3 智能媒体解决方案全景图

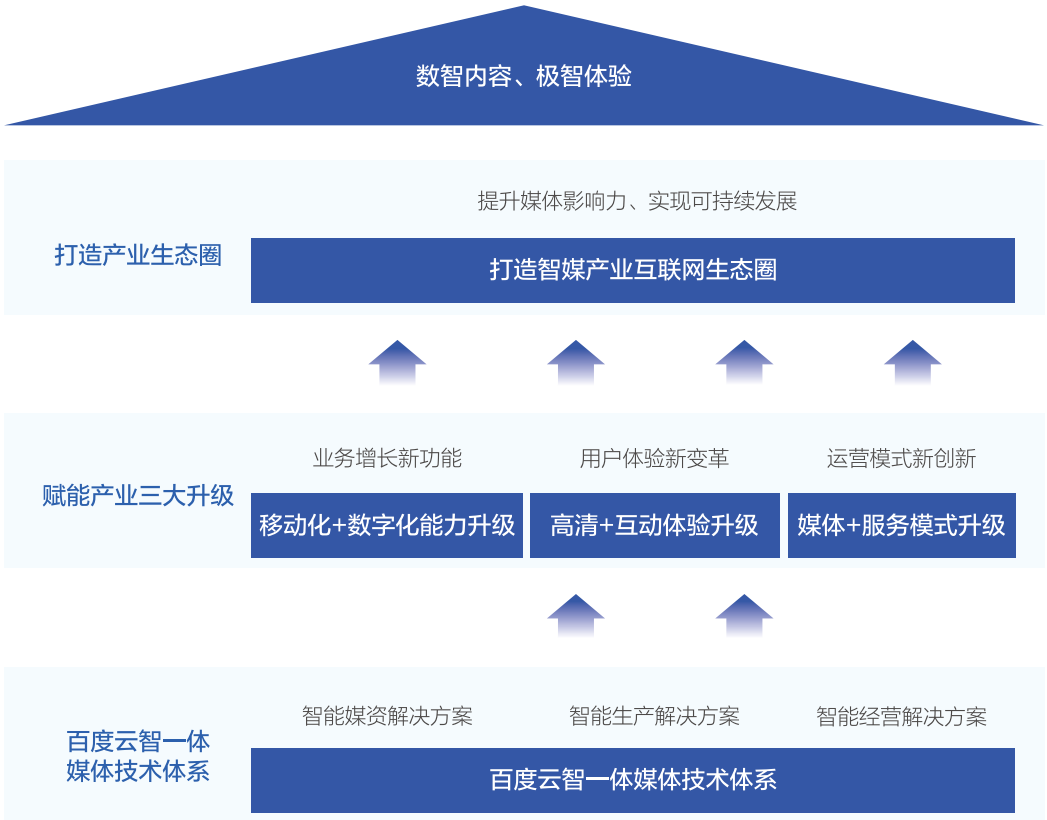
- **定位及价值主张**

百度智能云依托云智一体的视频云技术，致力于助力媒体机构的数智内容生产、打造极致体验。智能媒体解决方案的定位如下：

云智一体的媒体技术体系。百度智能云可为媒体机构提供包括智能媒资管理、智能生产创作和智慧经营等在内的云智一体技术方案，实现从内容生产到消费的全闭环技术体系。

赋能产业三大升级。通过移动化、数字化升级，助力媒体业务获得新的增长空间；打造高清互动的体验，实现用户体验的新变革；推动媒体与服务模式升级，为媒体创新与探索更多元化、交互性更强的商业机会提供技术支持。

打造产业生态圈。百度智能云希望通过跟央媒、省市广电等媒体机构以及头部学校合作，探索产学研结合，提升媒体影响力，打造新型主流媒体、融媒体的大发展格局，实现媒体产业的可持续发展。



（图4.1.3-1 智能媒体解决方案的定位及价值主张）

- **全景图解析**

云智基础设施。提供核心的视频云能力及多种智能化场景技术，围绕各类业务场景，打造媒体中台体系。

媒体 AI 中台。通过智能内容汇聚与生产、智能媒资一体化，全面覆盖媒体内容的策、采、编、审、发等全流程。

媒体知识中台。基于多元知识图谱，实现内容资产的知识化、资产化，将媒体自产和采购的海量内容进行标签索引，形成结构化的内容资源库。

媒体经营中台。所有媒体的内容生产都是为了触达受众，实现传播或者商业变现。百度有着丰富的多元化内容经营与分发经验，助力媒体打造智能经营中台体系。

行业方案体系。在云智基础设施和媒体中台技术的加持下，构建智能媒资管理、智能内容生产及智能经营等行业级方案体系，覆盖策采编审发等全流程，实现台网端的全渠道触达。同时，还可结合广电、网络、金融、文旅和体育等不同泛媒体领域的个性化场景，提供定制化解决方案。

百度媒体生态体系。百度是一家有 20 多年经验的信息服务类、搜索服务类企业，有着多年的媒体内容分发处理经验，以及大量内容数据渠道和生态资源，能够联合内容合作伙伴，实现对整个泛媒体进行全链条、全渠道的智能化赋能。

新媒体智能分发与运营。基于百度知识图谱，基于用户数据洞察和知识图谱化的内容资源，实现对用户和内容的智能化理解，形成用户画像和内容库，达到更加个性化、精准化的内容分发。同时，利用百度多年的内容流量生态运营经验，结合人工智能技术，达到整个产业从内容生产到消费的全闭环，实现一体化运营。

（图4.1.3-2 泛媒体产品和解决方案全景图）

4.1.4 泛媒体客户案例解析

- **广电传媒**

人民日报“创作大脑”。

当前，广电传媒机构都在响应国家政策号召，打造新型主流媒体，实现融媒体转型。百度智能云携手人民日报推出“创作大脑”：一是通过 AI 技术提供智能写作、AI 打标、在线快编等多样的智能化生产工具，覆盖人民日报的内容生产全场景，提升效率；二是对人民日报的海量信息，进行知识化和图谱化处理，对整体存量资源进行标签化索引，内容资源实现知识化和资产化，更好适用在媒体订阅、新闻数据追踪、热门检索和原创分析等生产分发场景中，提供最新热门线索及覆盖最广的原创查重等数据服务；三是实现流量端的整合，百度的百家号、手机百度等媒体，融合人民日报的人民号和超级 APP，实现人民日报的公私流量整合，提升整体传播效率。“创作大脑”重点功能如下:

直播剪辑工具：满足了新媒体视频生产中对新闻会议、活动直播中精彩、重要片段的“边播边剪”生产需求。可在直播会议、活动录制过程中，1 分钟内完成短视频内容生产，提升生产与发布效率，已经成为媒体高效制作短视频的重要工具。

新闻转视频工具：支持直接粘贴百家号、人民号、头条号、企鹅号、微信号网页链接等方式，自动生成视频；支持上传 word 文件自动生成视频。主要应用在一些需要及时发布的场景中（有新闻通讯稿但来不及做成视频）。只需把过往新闻稿件以 Word 方式或人民号生成的图文信息，上传链接或文档至创作大脑，就可以直接生成视频，并发布在各种短视频平台，实现最快速的新闻视频化传播。

视频横转竖工具：该功能在多终端多渠道进行内容发布时，尤为重要。利用编码和人工智能技术对视频画面进行处理，在确保最小压缩比的情况下实现横竖转换，便于传播和二次加工。同时对画面主体人物进行智能识别，保证主要人物始终处于画面中。

视频生产提效：视频自动生成字幕文件、视频片段转 GIF、视频大小自动适配压缩等功能，能够实现视频素材的快速处理，一次生产、多终端发布。

智能媒体资源库：对剪辑素材文件进行多维 AI 分析，智能提取文件的标签、画面文字、语音及关键字等，提高文件的多次利用率。

智能云盘：快捷剪辑出视频关键片段，高效生产短视频；一键分享文件链接，更有加密、删除分享、有效期设置，保证文件安全传播。

央视网“五智”人工智能编辑部。

央视网有海量的视频内容存储，百度智能云从其早期媒资库开始做标签索引，对超过 200 万小时的视频做媒资化处理，基于媒体中台打造了覆盖策采编审发全流程的央视网人工智能编辑部实体与线上业务的全流程逻辑。基于这套平台，央视网开发了诸如“帮你找”等创新应用，方便采编人员、受众都能快速通过央视网找到想要的视频素材。同时，今年两会期间，百度智能云与央视网合作推出虚拟数字人小 C，组织了“两会 C+ 真探”的虚拟记者采访项目，收到市场广泛关注。

携手打造人民日报创作大脑

（图4.1.4-1 广电传媒案例）

- **金融行业：垂类信息和技术的智能化服务**

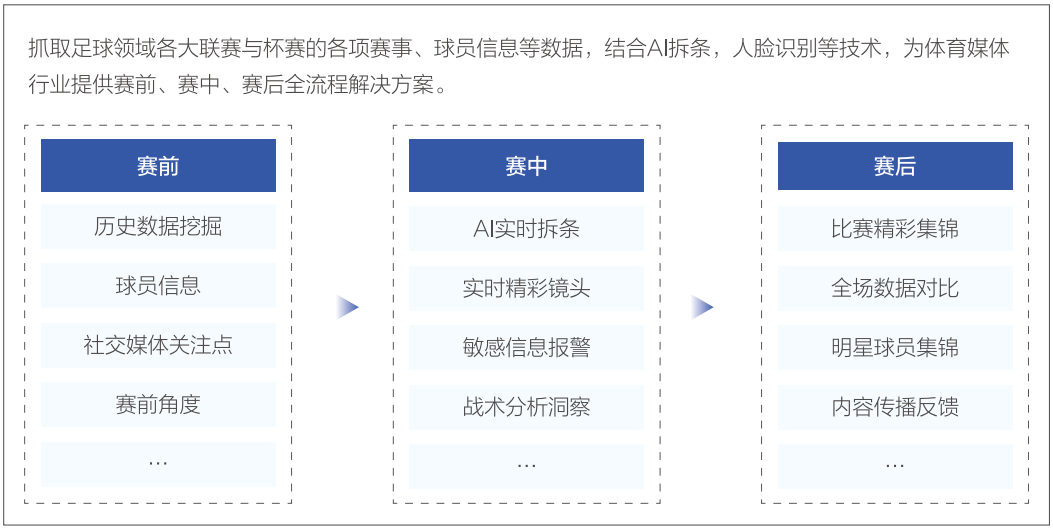
银行、保险、证券等金融机构，需要大量专业化内容。百度智能云赋能垂类金融内容供应商，追踪机构、行业等信息，对内容进行数据处理、标签化和索引化，提供可定制化的金融舆情信息和可视化展现服务，结合机器自动生产技术，快速生产金融机构所需的内容。

- **文旅行业：改善服务质量、提升用户体验**

文旅行业有很多存量视听资源，并还在不断生产。百度智能云联合重点景区、博物馆、爱国主义教育基地等机构，将人文与 AI 科技相结合，把景区影像资产数字化，基于 AI 知识图谱做内容结构化连接，为消费者提供 AI 导游、AI 引流、AI 客服以及安全隐患预警等服务，改善了服务质量和游览体验，增加了营收能力，打造了智慧文旅新基建标杆。

- **体育行业：全场景、全流程智能化**

百度智能云抓取足球领域各大联赛与杯赛的赛事与球员信息，结合 AI 拆条、人脸识别等技术，为体育媒体机构提供从赛前的历史数据分析、知识图谱化检索，球员信息梳理等服务，到赛中的 AI 实时拆条、实时精彩镜头、敏感信息报警、战术分析洞察等服务，再到赛后的赛事集锦、数据对比等服务。

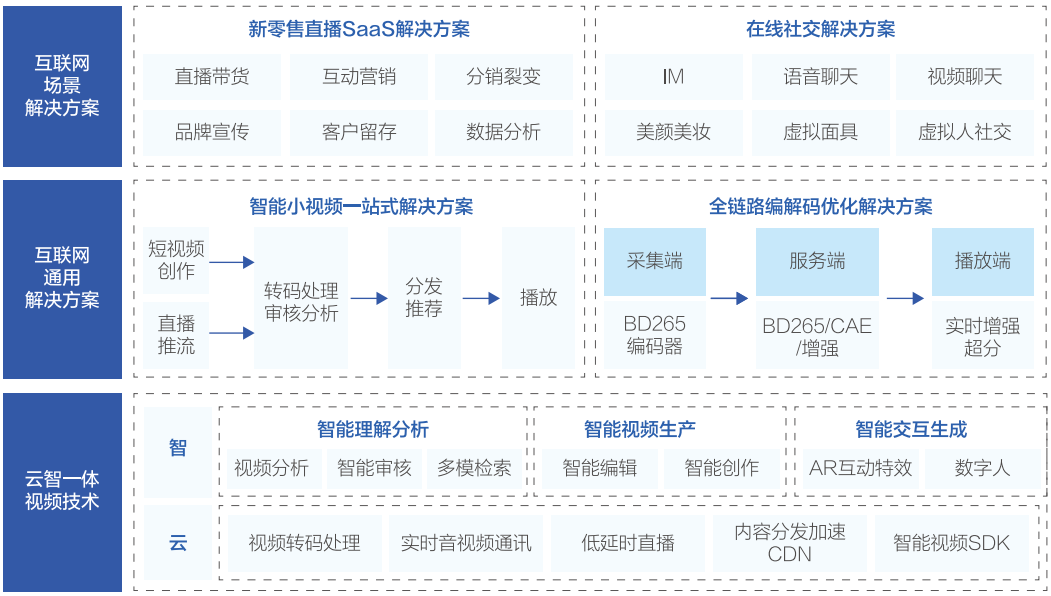


（图4.1.4-2 体育行业案例）

4.2 云智一体的视频技术推动泛互联网产业升级

内容和用户是互联网企业增长的核心要素。百度智能云助力优质内容生产者，深入垂类赛道，用 AI 技术降低码率、提升画质，以低成本分发的方式，精准触达到目标受众群，形成裂变增长。

云智一体的智能视频云 3.0 平台可为互联网企业提供四大视频解决方案，帮助企业提升内容生产效率，拉动创新内容的产量，优化视频质量，提升用户体验，实现降本增效。



四大解决方案，带来视频体验与效率升级

（图4.2-1 泛互联网的四大视频解决方案）

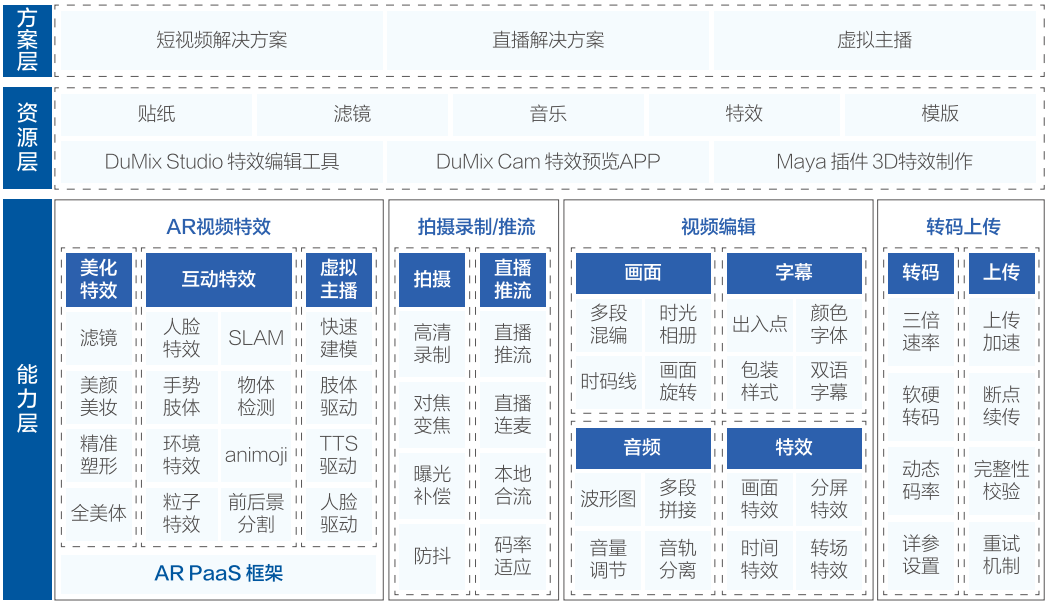
4.2.1 智能小视频解决方案：快速上线视频业务

智能小视频一站式解决方案，提供全套智能视频拍摄工具，在拍摄器上提供 100+ 的 AR 技术拍摄编辑玩法。通过一站式工作流，直接将拍摄视频传输至存储系统，进行媒资管理、视频处理与审核、以及发布与分发加速。同时，APP 样本及服务端部分代码开放给用户，可根据自身需求定制开发，快速打造视频应用。

智能小视频解决方案，帮助快速上线视频业务

（图4.2.1-1 智能小视频解决方案）

以其中的智能视频拍摄器为例。它为短视频及相关垂直行业的 UGC/PGC 场景，提供移动端采编传的视频内容创作工具 SDK，内置美颜美体、人脸识别、肢体动作识别及场景特效等 100 多种技术能力，可对视频做双语字幕等多种编辑操作，为直播场景提供移动端推流工具，支持单向直播和互动连麦。

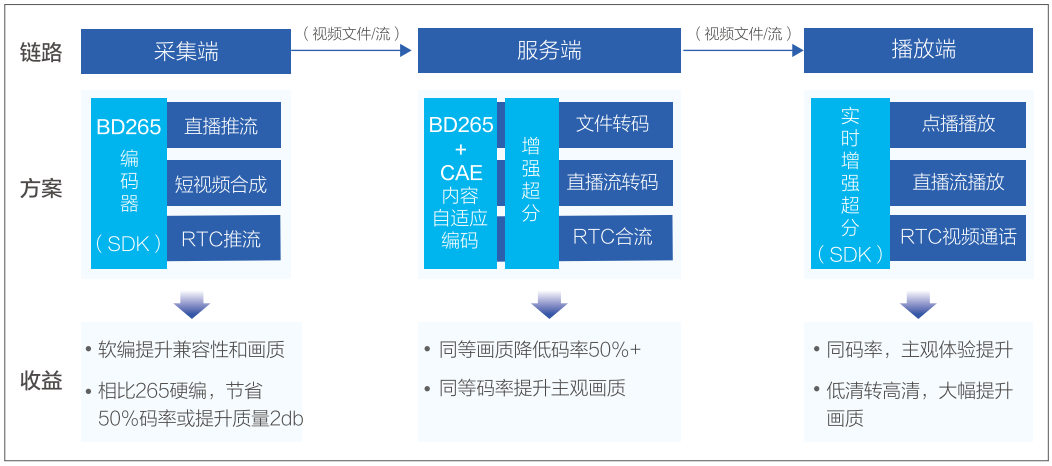


智能视频拍摄器，移动端视频内容创作利器

（图4.2.1-2 智能视频拍摄器）

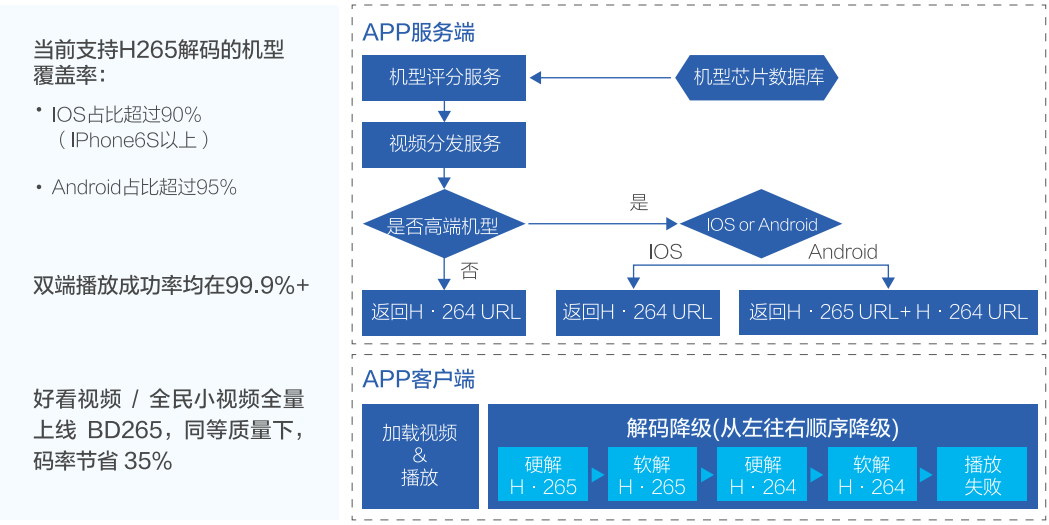
4.2.2 全链路编解码优化方案：降低成本、提升画质

全链路编解码优化解决方案，基于采集端的 BD265 方案，兼容软硬编码能力，相比 H265 编码，节省 50% 的码率或提升质量 2db。在服务端，结合 BD265 及智感超清 AI 能力，提升画质并降低码率，节省传输带宽和存储成本。在播放端，提供实时视频增强 SDK 技术，同码率情况下，大幅提升视频画质，提升主观体验。



（图4.2.2-1 全链路编解码方案）

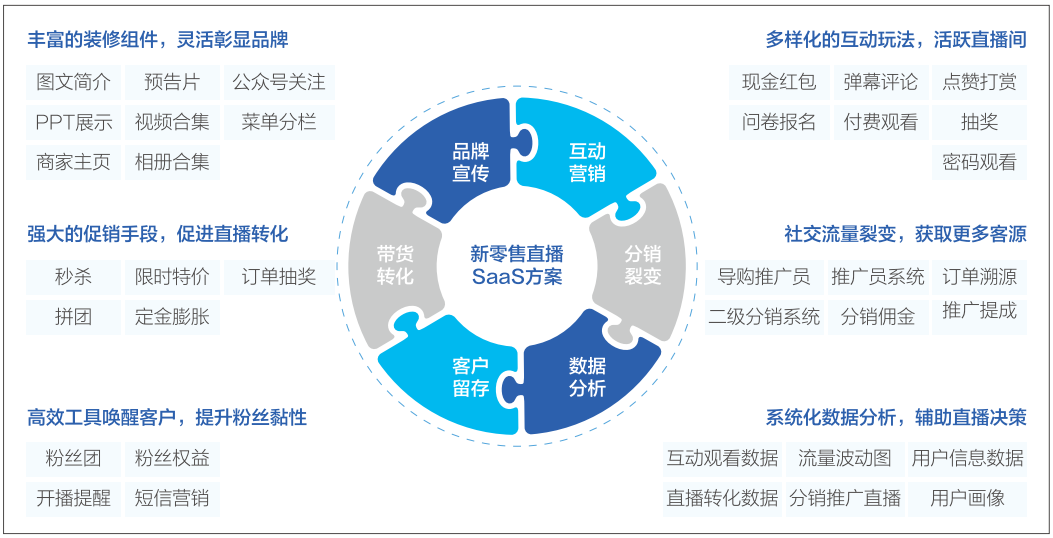
以 BD265 终端解码策略为例。支持 H265 解码的机型覆盖率超过 90%+，手机端普及 H265 的条件非常成熟。而 BD265 手机端解码产品在具备与 H265 同等条件下，相比 H265 码率节省 35%，通知支持降级解码，确保 99.9%+ 的解码成功率，保证整个终端播放的成功率，提升用户体验。



（图4.2.2-2 BD265终端解码策略）

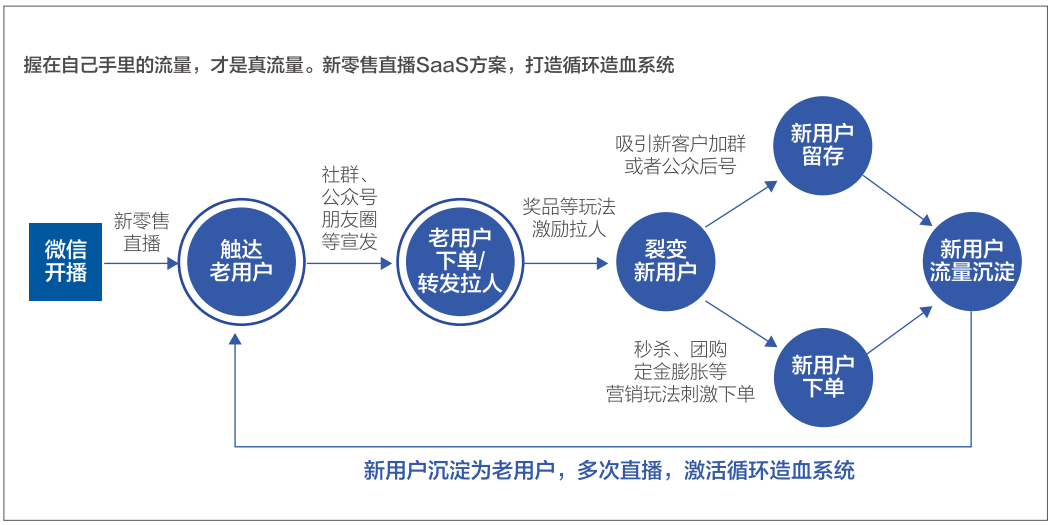
4.2.3 新零售直播SaaS解决方案：助力商家裂变获客、流量变现

新零售的直播 SaaS 解决方案，帮助网红主播、实体门店、家电家居、房地产、汽车、品牌厂商、微商电商等开启新零售导购直播。通过一键式导购直播能力，提升用户触达率，实现线上营销变现和裂变引流，获得更多的转化；提供分销系统，实现订单追踪，保证用户的归属和转化；具有数据分析能力，帮助用户做出直播决策。



（图4.2.3-1 新零售直播SaaS解决方案）

以私域流量运营为例。百度智能云帮助用户构建私域流量池，通过直播方案，先找到用户的种子客户，转化为老客户或潜在客户；再通过激励措施，让这些老客户去分享带动新用户进来，留存新用户并形成转化，最终形成自生长造血系统运营。



（图4.2.3-2 私域流量运营）

4.2.4 在线社交解决方案

在线社交解决方案支持 IM 聊天、实时音视频通话以及虚拟 AI 技术等三位一体的技术能力组合，可帮助用户打造文字、语音、视频等社交属性内容，并可使用动画制作功能，让用户去打造独一无二的虚拟形象，以真人实时驱动，构造定制化 IP。

（图4.2.4-1 在线社交解决方案）

4.2.5 泛互联网客户案例解析

- **智感超清帮助“得物 APP”大幅提升视频画质体验**

得物 App 是新一代潮流网购社区，聚集了新、潮、酷、炫的各类商品，也是各类潮流品牌发售和运营的首选阵地。用户痛点是 PGC/UGC 上传的短视频画质清晰度偏低，影响用户播放体验。百度智能云为低于 720P 的视频超分至 720P。基于智感超清解决方案，PGC/UGC 上传的短视频全量上线，视频画质大幅提升，用户观感体验获得好评。

- **为“好看视频 / 全民小视频”提供全链路处理分发服务**

好看视频，是一个为用户提供海量优质短视频内容的专业聚合平台。全民小视频 App 覆盖多种类型的小视频，用户可以发布小视频分享、记录生活，支持大眼瘦脸美颜和各种贴纸等功能。两款视频应用都是全面基于视频全链路处理分发平台来构建的，从视频拍摄、音视频处理、理解分析等都是基于整个平台来构建的，大幅提升视频画质，提升主观体验，节省用户的传输带宽和存储成本。

- **赋能快手为全球用户提高优质的点播 / 直播体验**

快手是国内最大短 / 直播视频平台之一，百度智能云为快手提供 BOS、CDN、LSS 等云服务，为快手全球数亿用户提供海量视频随时随地快速访问能力，让用户体验好的直播感受，真正实现了端边云一体的使用体验。百度智能云 BOS 结合遍布全球的 CDN 节点，为快手提供海量的图片和短视频处理需求加速服务。百度智能云为快手提供一站式直播 LSS 解决方案，智能优化链路，为受众带来高清、流畅、低延时的游戏直播，以及性能稳定的短视频、点播回看。

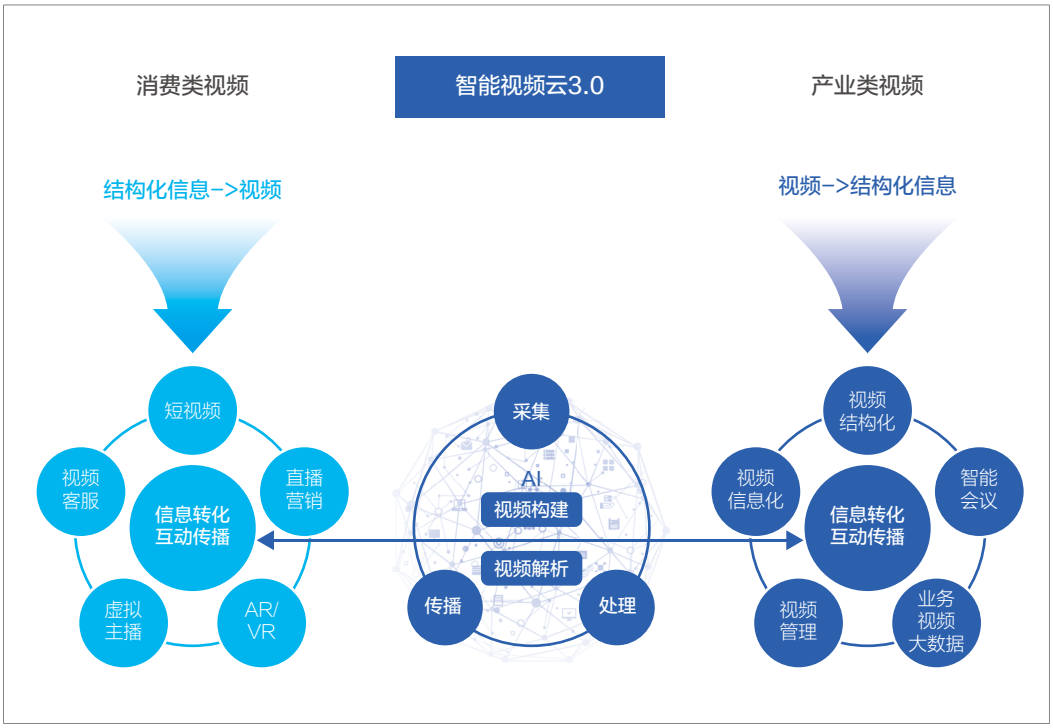
4.3 云智一体的视频技术推动泛产业的视频场景升级

百度智能视频云基于端边云一体的基础架构，以“平台 + 数据”为主导，满足泛产业的业务系统对接和数据平台化等需求。

4.3.1 产业类视频与消费类视频的区别

视频作为人类最直接、最容易感知世界的一种方式，广泛应用在日常生活和工作中，最喜闻乐见的视频场景主要是在传媒业、娱乐业等大众消费领域，此类视频应用场景实际上是将一些结果化信息、知识信息转化为多媒体形式，便于受众体验，提升与受众的互动传播性。

而对于产业类视频来说，主要是服务于业务场景。因此，信息的转化过程恰恰相反，是通过视频解析技术等，将视频进行结构化处理，把其中的知识信息以结构化形式提取出来，为企业业务提供决策依据。



（图4.3.1-1 产业类视频与消费类视频的区别）

4.3.2 视联网发展的三大阶段

- **视频管理阶段**

该阶段以硬件厂商为主导，提供简单的视频管理应用，服务场景主要是视频存管看等基础问题，服务厂商主要是传统设备商。这个阶段虽然能满足企业的视频管理等基础需求，但缺乏智能化手段，主要还是人工来完成流程操作，因而相对来说整体管理效率较为低下。

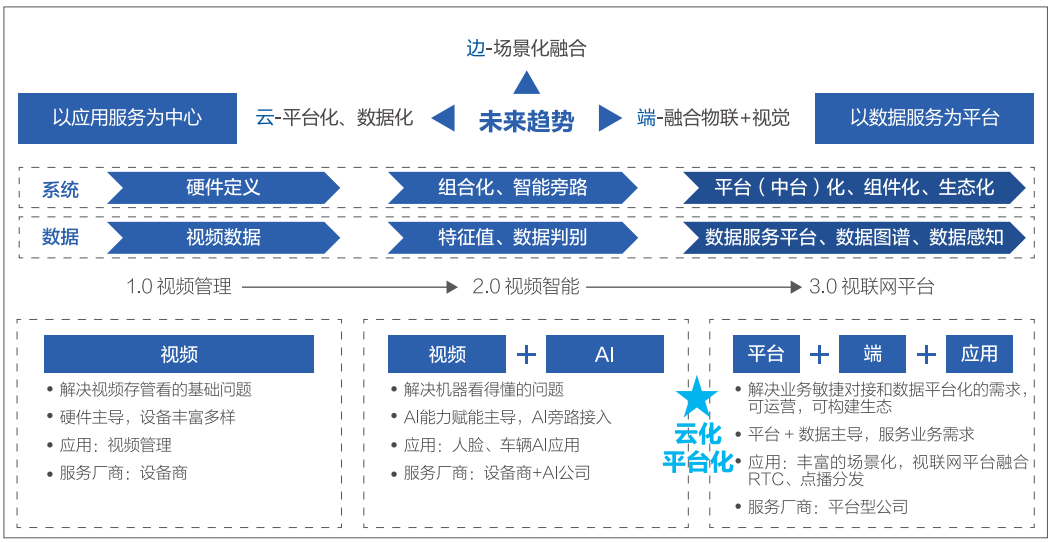
- **视频智能阶段**

随着 AI 技术的发展，AI 厂商通过旁路方式，把 AI 应用赋能到视频管理平台上，通过智能化手段提升视频理解能力，解决了视频“能看懂”的问题。这个阶段主要是 AI 能力主导，采用人脸、人体、车辆等 AI 基础应用来实现智能化改造。

- **视联网平台阶段**

视联网发展到上一阶段，虽然很多需求通过 AI 应用得到了满足，但这种组合化、智能旁路的思路，也带来了应用碎片化的问题，企业或政府要想针对不断涌现的业务应用需求做长期、可持续的规划与投入，难度较大。因此，为了满足业务发展过程中不断涌现的各类应用和场景需求，企业需要打造具备长期规划和投入属性的平台，这便进入第三阶段——视频网平台阶段。在该阶段，平台端与应用端充分融合，形成一个可与应用和业务快速对接、提供丰富数据服务的能力平台。企业、政府都可以利用该平台做长期规划和价值投资，将整体业务都跑在平台上，构建丰富的场景化应用体系。

综合来看，泛产业场景视联网的未来发展趋势：一是以应用服务为中心，更加贴近业务，业务需求也会水涨船高；二是以数据服务为平台，满足各种业务发展和决策的数据需求；三是在云侧实现平台化和数据化，在边侧依托 AI 技术实现场景化融合，在端侧融合物联网和视联网，实现更多数据的汇聚与处理。



（图4.3.2-1 泛产业视联网发展趋势）

4.3.3 百度智能云视联网感知平台

针对视联网平台第三发展阶段，百度智能云推出了视联网感知平台解决方案，该平台属于闭环性架构。

接入侧是完全开放模式，可兼容业内 10 余种的主流协议，市场上的主流终端都可以便捷的接入到平台。另外，还给厂商开放一些百度自研的 SDK 能力，其可快速和平台对接，实现统一管理。百度智能云还推出了度目等一体化智能终端，在端侧实现 AI 能力落地。

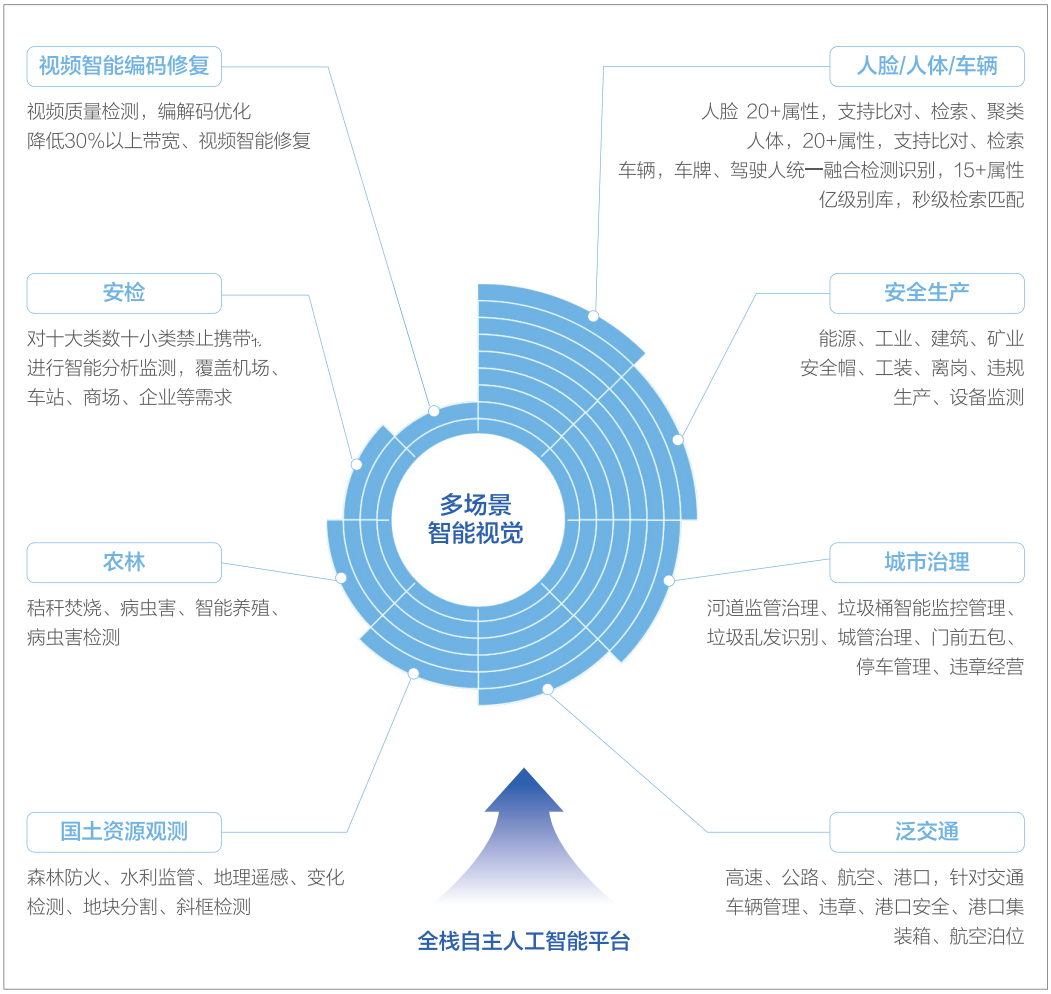
视频接入之后，视频管理平台可以提供 20 多种视频处理能力，而在应用商城上有 200 种 + AI 能力，可实现对多种视频场景的覆盖和解析。解析出来的结构化数据，会沉淀到 AI 平台——EasyDL，再利用这些数据产生丰富的、满足业务场景化需求的 AI 算子。算子发布到大数据平台，可为上层业务应用进行赋能。

（图4.3.3-1 视联网感知平台架构）

4.3.4 视联网感知平台架构的三大特点

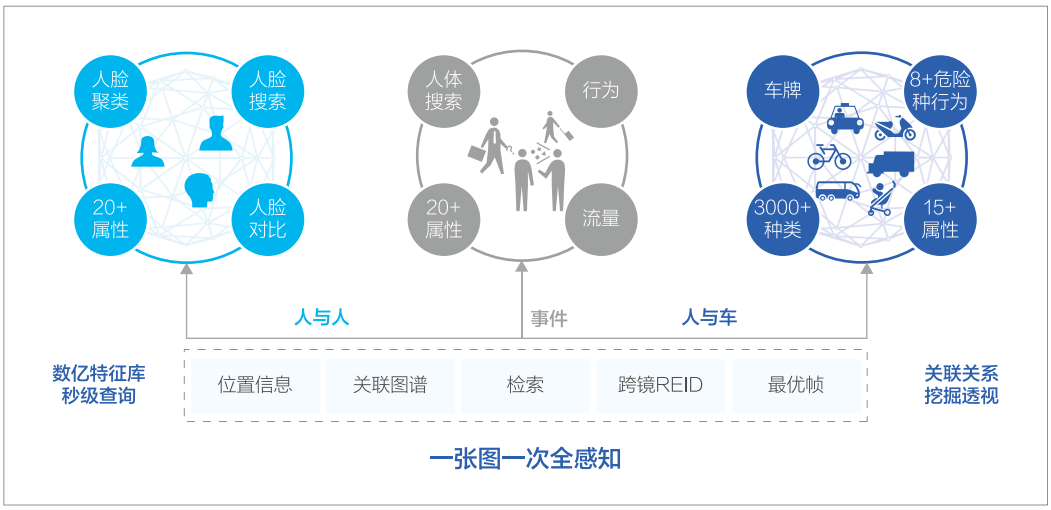
- **特点一：智能**

覆盖多行业和场景的智能视觉分析能力。视联网感知平台依托百度自研的全栈式 AI 平台能力，可为多种视频场景提供智能化视觉应用。常见如针对人脸 / 人体 / 车辆，有丰富的特征值信息，并可提供库查询，还具备视频修复能力，提升整体视频的理解能力。目前，应用已覆盖交通、能源、城市治理和农林等各类主流的业务场景。



（图4.3.4-1 视联网感知平台覆盖多行业和场景智能视觉分析能力）

全目标感知服务。相比于单一算子只能提供事件解析，不能提供更丰富数据的局限性，百度智能云视联网感知平台可提供全目标感知服务，基于一张图即可解析出人脸、人体、车辆以及其他关联性信息，支持深度关系数据的挖掘。比如在疫情期间，快速追踪到可能的密接者并获取关联信息，支持疫情防控的复杂性决策，这便是全目标感知服务的应用之一。



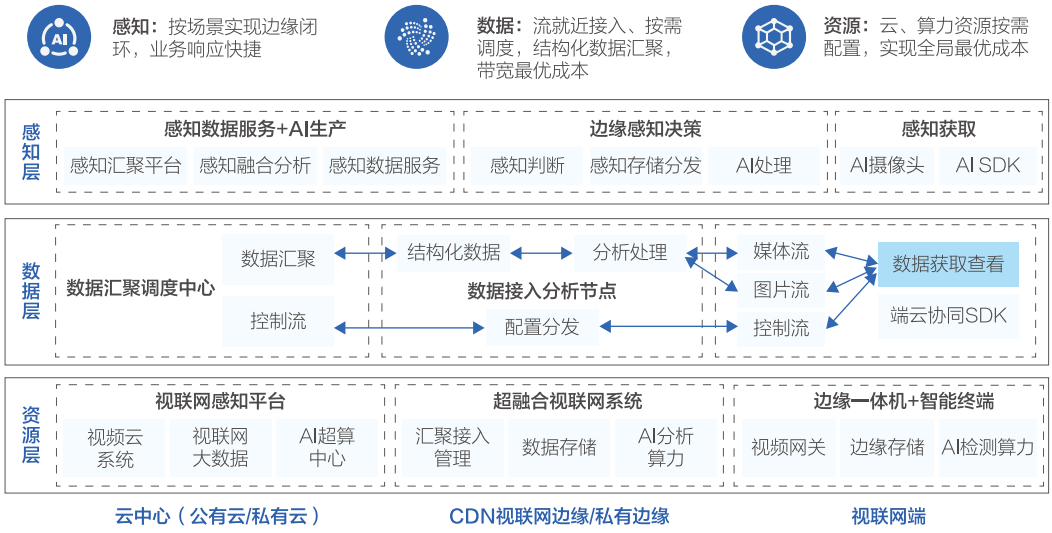
（图4.3.4-2 全目标感知服务）

高质量快捷闭环的算子生产应用。在一些产业视频场景中，很多需求和应用较为复杂，通用算法无法完全满足这些场景化需求。而基于百度 AI 生产能力的视联网感知平台，可以高质量的、快捷闭环的产生算子。将获取的数据在 EasyDL 平台上进行模型训练，整个过程可零代码方式快速实现。再将训练出的算子发布到视联网感知平台，通过应用的开发和集成，实现在硬件上的快速部署。另外，应用发布之后，可通过模型评测机制，实现数据和算子质量的评测。

（图4.3.4-3 高质量快捷闭环的算子生产）

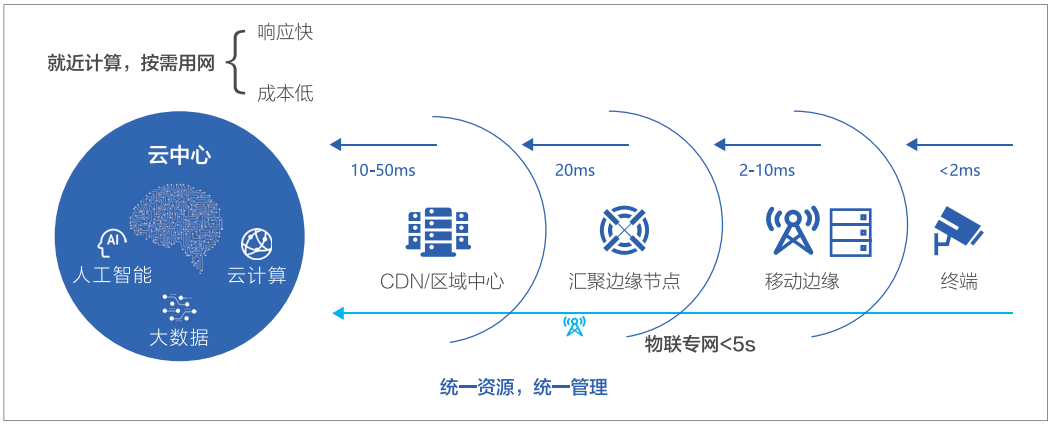
- **端边云融合**

视联网感知平台在资源层、数据层、感知层等三方面进行了全面的端边云融合。在端侧部署了算力和算子，快捷获取数据。在边缘侧按场景实现边缘闭环，业务响应快捷。在云侧将关键信息和数据传送到云端，实现数据的汇聚与解析，为业务提供更加云化和平台化的服务，打造更多的业务场景应用。



（图4.3.4-4 视联网感知平台的端边云融合架构）

首推融合边缘架构，实现全域最优成本全接入。视联网的构建相对复杂，网络和物理位置都非常分散，从而也导致网络和存储成本较高。百度智能云在业内首推“融合边缘架构”的概念，实现在整个终端移动端上汇集边缘、CDN 边缘、云边缘等，实现整体资源的统一管理，实现视频内容的就近处理，快速响应，同时节省带宽，拉低整体成本。



首推融合边缘架构，实现全域最优成本全接入

（图4.3.4-5 融合边缘架构）

- **开放平台能力**

开放端云协同：开放整个端云协同 SDK 与 AI SDK，设备厂商可便捷的配置终端设备，实现终端统一管理、快速注册，实现低延迟的高质量连接和 AI 应用端云协同。在公网环境下，可达到百 ms 级延迟，SDK 兼容多平台。

开放平台模块：视联网平台模块化开放服务，应用实现快速对接构建，并利用 PaaS 能力快速迭代，无需关注底层资源和运维能力，提升 SaaS 开发效率，解耦应用对平台资源的配置管理。

开放数据服务：通过 AI 能力构建数据关系数据引擎，提供丰富的数据服务 API 和模式，AI 开发平台实现协同，提升 AI 应用质量和场景化落地。

（图4.3.4-6开放平台能力）

4.3.5 产业类客户案例解析

点军区的“点军”二字，来源于三国，关羽在此练兵而得名。这里距离三峡大坝 47 公里，区域面积 530 平方公里，包括城区、乡镇、以及农村等三种生态的综合性区域。2019~2020 年是黑天鹅事件频发的 2 年，从疫情防控到环境保护、乡村振兴和产业升级，点军区政府正面对着史无前例的变化性需求和挑战：

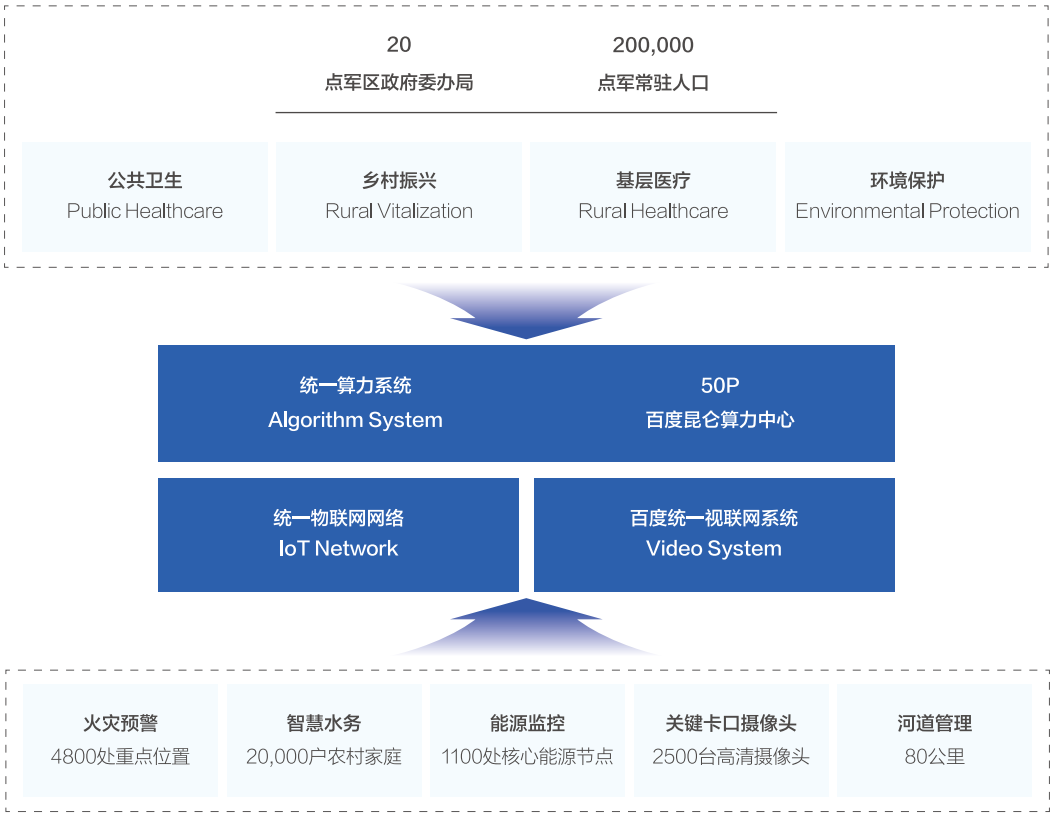
公共卫生：如何实现更高效、更智能化的疫情管控?

乡村振兴：面向乡村振兴的国家战略，该如何落实？该如何让偏远地区的百姓有安全感、获得感和幸福感?

基础医疗：如何让偏远地区基础医疗更加普及?

环境保护：如何让环境保护变得更加智能化?

面对这些挑战，宜昌点军区联合百度智能云，融合视联网平台、算力平台、物联网平台、大数据平台的全域数字化服务，打造区域性运营平台。用 50P 算力在视联网平台和物联网平台上构建丰富的业务场景应用，助力城市智能产业区域的社会治理和经济发展。



（图4.3.5-1 点军区平台落地实践）

“50P 算力 +1 个视联网平台 +1 个物联网平台 +N 个应用”的区域性运营平台系统，正在被点军区 20 个政府部门使用，服务于 20 万常住人口。平台满足了点军区政府对城市治理和民生服务的各类需求，避免了重复建设，极大降低了政府投入成本。取得的阶段成效举例如下：

- **公共卫生：**实现对疫情等突发性事件的及时响应，比如疫情爆发城市封闭，第一时间找到被滞留在高速路上的高危运输车辆；通过跟踪药品销售以及用药人群的分布趋势来预测流行病。
- **乡村振兴：**很多偏远地区消防车无法立即到达现场，意外火灾的处置是一个难题。点军区在这些地区架设物联网专网和传感器预警系统，不断加强偏远地区意外火灾防范能力，达到异常情况下秒级预警，激活邻里救助。为了让孤寡老人生活更安心，通过物联网专网和智能床垫，对老人夜间的异常健康情况进行监测，为老人提供健康保障。为了让百姓生活更便捷，通过物联网专网实现在线缴纳水费、电费。
- **基础医疗：**使用人工智能眼底监测设备，深入农村地区卫生所，对慢性病患者进行早期筛查，让村民能够尽早治疗，过上更健康的生活。
- **环境保护：**长江生态大保护实行十年禁渔禁捕，通过大数据构建非法捕捞行为的人工智能模型，同步结合人脸识别，锁定非法捕捞群体。
- **城市治理：**通过人工智能视觉系统，跟踪建筑材料运输过程，从源头打击违章建筑。

更多行业案例参考

- **媒体类**

人民日报“创作大脑”

人民日报 RENMIN RIBAO

央视网“五智”人工智能编辑部

CCTV 央视网 .com

- **视频类**

爱奇艺“云+AI”视频内容平台

iQIYI 爱奇艺

好看视频全链路处理分发服务

好看视频

- **直播类**

触手直播游戏视频识别模型

触手

虎牙游戏直播平台

虎牙直播 HUYA

- **交通物流类**

一汽物流无人机整车盘点服务

一汽物流有限公司 FAW LOGISTICS CO.,LTD.

欧了出行车联网视频云服务

欧了出行 OLESHARING

- **教育类**

沪江互联网学习直播服务

沪江 学习，成为更好的自己

威创儿童成长平台

vtron 威创儿童成长平台

- **工业类**

精研科技外观缺陷视觉检测

Gian 精研科技

新石器无人车

NEOLIX 新石器无人车

关于本报告

本报告内容来源于百度智能云重磅打造的 “云智技术论坛”系列活动第二期。5 月 13 日，第二场活动以《云智一体的全场景智能视频前沿技术》为主题在北京召开。百度多位资深技术和产品专家，全面介绍了百度最前沿的视频云和 AI 能力矩阵，并邀请合作伙伴共同呈现了面向互联网、媒体、产业等场景的云智一体智能视频方案与应用案例。

2021 年，“云智技术论坛”系列活动还将陆续推出其他主题，全方位诠释百度智能云“云智一体”的独有优势在各场景各行业的助力应用，敬请关注。

《云智一体的全场景智能视频前沿技术》

精彩回顾

百度智能云微信公众号

百度 AI 微信公众号

百度智能云
微信公众号

百度 AI
微信公众号