



云化增强现实关键场景及技术 白皮书 (2020)

中国通信标准化协会

(2020年12月)

版权声明

本白皮书版权属于中国通信标准化协会，并受法律保护。
转载、摘编或利用其他方式使用本白皮书文字或者观点的，应
注明“来源：中国通信标准化协会”。违反本声明者，本协会
将追究其相关法律责任。

前 言

云化增强现实（Cloud AR）将 AR 的内容上云、渲染上云，降低终端要求，提升 AR 用户体验，是当前 AR 产业自主选择的规模化发展之路。5G 网络的商用提供了时延更低的网络，可满足云化 AR 实时交互的响应需求；云化增强现实凭借共享云端/边缘服务器上的强大计算和存储资源，可以有效降低用户体验 AR 业务的终端硬件成本；通过云平台对基础设施的升级维护，降低用户端设备更换和应用升级的成本。云 AR 平台的建设使得 AR 产业的生态更加丰富，需求得到匹配，从而促进云 AR 端到端的用户体验提升。

随着我国 5G 网络商用以及未来网络的发展，5G 移动网络与云 AR 业务的结合已成为重要趋势，将有力促进 AR 业务的大规模普适化应用。本白皮书基于 Cloud AR 产业背景及国内外标准化情况，介绍了 AR 云化在若干领域应用的关键场景，分析了 AR 云化的需求，提出了 Cloud AR 解决方案的总体技术架构和关键技术，梳理了 Cloud AR 终端、平台、网络三大层次的关键技术方案，旨在为我国产业构筑 Cloud AR 业务端到端能力提出标准化建议，推动我国 AR 产业更好、更快的发展。

本白皮书由中国通信标准化协会移动互联网应用和终端技术工作委员会（TC11）虚拟现实/增强现实（VR/AR）子工作组负责编写。主要编写单位包括：中国移动通信有限公司、华为技术有限公司、中国电信集团有限公司、中国信息通信研究院、深圳市腾讯计算机系统有限公司、北京视博云科技有限公司、中国联合网络通信有限公司、OPPO 广东移动通信有限公司、北京邮电大学、北京易智时代数字科技有限公司、上海商汤智能科技有限公司、中兴通讯股份有限公司。

编写人:杨蕾、聂蔚青、朱奇、张志远、周佳俊、刘永清、高扬、梁运恺、武娟、刘晓军、徐晓青、马治国、李志斌、牛长峰、蔡庆宇、邱倩琳、来航曼、乔秀全、欧中洪、刘毓、王红梅、姜翰青、蒋慧、李秋婷

目 录

一、云化增强现实产业概述.....	1
(一) 云化 AR 尚处于发展初期，潜在市场规模巨大.....	1
(二) AR 产业面临诸多挑战，云化 AR 技术提供新思路.....	2
(三) 产业界及标准化组织在 CLOUD AR 领域的探索，推动 CLOUD AR 行业发展.....	5
二、云化增强现实关键场景.....	7
(一) 通话.....	7
(二) 教育.....	9
(三) 游戏.....	12
(四) 地图.....	13
(五) 工业.....	15
(六) 展馆.....	18
(七) 党建.....	19
(八) 安防.....	19
三、云化 AR 总体技术方案.....	20
(一) 终端处理能力云化降低端侧设备成本.....	21
(二) 网络传输技术满足 AR 应用云化需求.....	22
1. 传输内容.....	23
2. 传输需求.....	25
(三) 云平台构建支撑多种云化 AR 应用.....	25

(四) 中心云能力下沉减轻云端压力.....	28
(五) 边缘云提供强计算能力.....	29
四、云化 AR 技术方案—增加典型应用描述.....	30
(一) 终端技术方案.....	31
1.主流 SDK.....	31
2. 光学显示方案.....	32
(二) 网络能力及传输技术方案.....	33
1.超低时延流化技术.....	34
2.实时端边协同.....	37
3.网络原生优化.....	38
4.数据传输内容.....	39
(三) 平台技术方案.....	40
1.AR 框架云化技术.....	40
2.多路并发承载技术.....	41
3.多应用并行技术.....	42
(四) 中心云能力及技术方案.....	43
(五) 边缘云能力及技术方案.....	44
1.实时建模.....	44
2.实时跟踪.....	45
3.模型实时驱动.....	46
4.SLAM 及高精度空间计算.....	47
5.对象识别及模型匹配.....	47

6.高精度模型实时渲染.....	48
五、总结及发展展望.....	53
参考文献.....	56
附录1 AR 计算端云分布建议.....	57
附录2 端侧无法承载的高阶 AR 计算说明.....	59
附录3 缩略语.....	62
附录4 标准化情况.....	64

一、云化增强现实产业概述

（一）云化 AR 尚处于发展初期，潜在市场规模巨大

AR（Augmented Reality）即增强现实，是指将真实世界信息和虚拟世界信息“无缝”集成的新技术，实现真实的物理环境和虚拟的信息（如 3D 模型、视频、文字等）实时地叠加到同一个画面或空间中。不同于 VR 中模拟或复制出来的完全虚拟的环境，AR 是利用虚拟信息对当下的现实世界进行补充，实现增强现实的效果。

随着 5G 正式商用落地，AR 将提供更多使用场景，云化 AR 场景将迎来更多机遇。据 2019 年 Gartner 技术成熟度曲线显示，Cloud AR 正处于触发期，未来的 5-10 年将会对商业、社会和民生，产生重大影响。据 Greenlight 预测，2020 年全球 XR 产业规模将超 2000 亿元，其中 AR 市场规模将超 400 亿元。普华永道预测，2030 年，VR 和 AR 将会为中国经济规模增加 1833 亿美元(约合人民币 1.288 万亿元)，等同于 GDP 增长 2.09%。AR 将提供最大的贡献，达到 1295 亿美元。

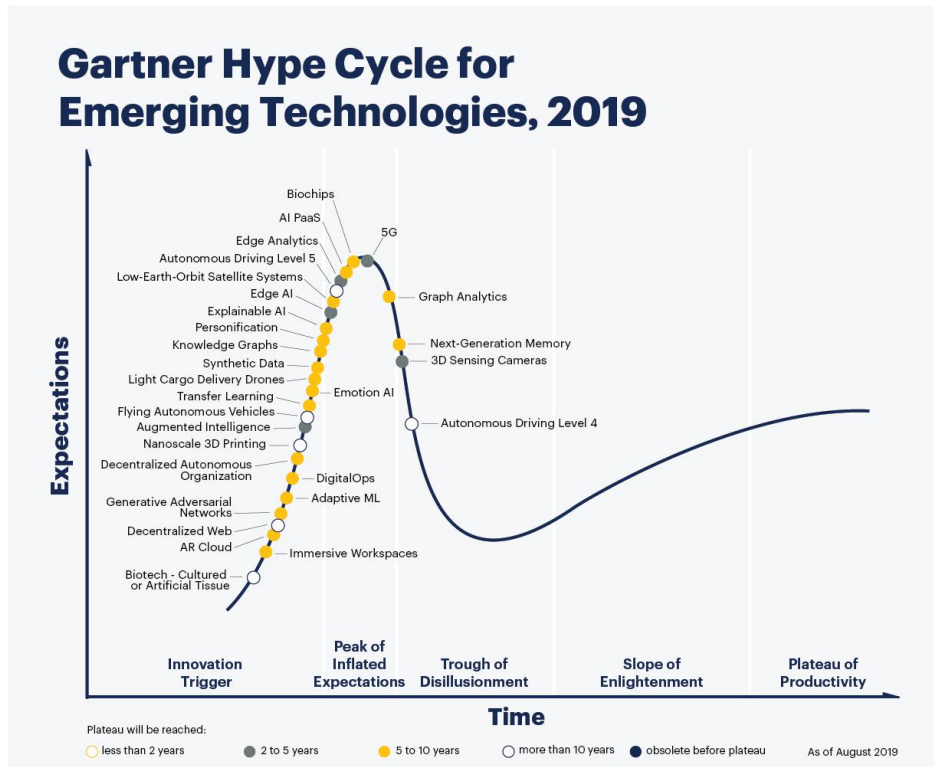


图 1. Gartner 成熟度曲线

（二）AR 产业面临诸多挑战，云化 AR 技术提供新思路

ETSI 在 2018 年 5 月份公布了一份针对 AR 用户开展的体验调研报告，针对以上用户体验发现 AR 产业的痛点：

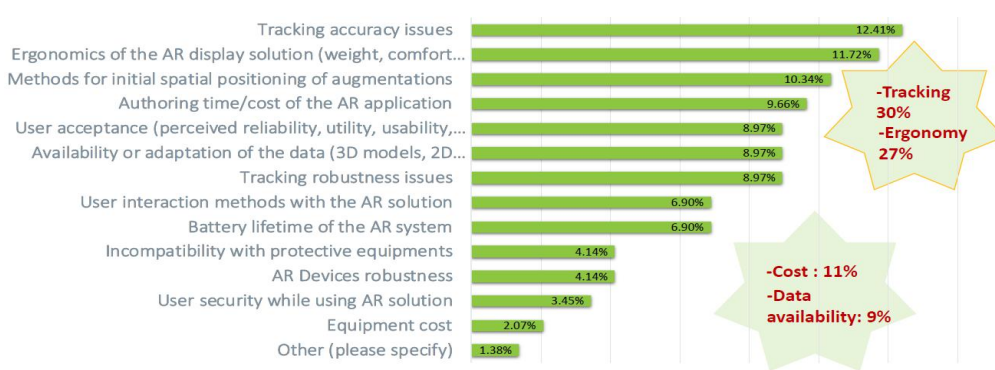


图 2. ETSI 在 2018 年开展的关于 AR 业务用户体验测试和调研结果

1). **端侧计算能力不足**：桌面（板级）NPU、GPU 算力分别为移动终端侧 20 倍、50-100 倍。与 App 类的 AR 应用相比，基于 Web 的 AR 应用端侧计算力更弱。而 AR 的核心关键追踪问题本质是计算问题，主要原因是传感器误差高、终端计算能力不足、追踪算法不成熟导致的。

目前 AR 终端具有一定的渲染能力，但只有桌面（板级）GPU 的性能的二十分之一。AR 应用与普通 3D 应用相比，要求较复杂的混合渲染能力，并融合图片、视频、3D 数字模型，动态模拟效果等，体验与 GPU 性能正相关。AR 终端具有一定的感知能力（推理），包括语义感知和基于 SLAM 的空间感知，但计算能力与板级 NPU 相比差异明显，边缘侧板级 NPU 进行 AR 应用感知是一种必然的选择。



图 3. 桌面（版级）GPU 与移动 GPU 渲染能力对比

2). **人体工程问题**：人体工程问题主要是由于设备相对笨重、交互方式误判率高导致的。目前 AR 终端的重量仍然过于沉重，用户无法在非工作场景自由佩戴，这也是 ToC 市场无法爆发的主要原因之一。



图 4. HoloLens2: 566g



图 5. MLO: 360g



图 6. 普通眼镜: 10~130g

3). **成本问题**: 当前阶段用户数量较低, 端到端成本无法控制、压缩。

4). **数据内容问题**: 模型精度仍待提升, 稳定叠加、遮挡关系实现难度高, 高精度模型需要更强的渲染能力, 端侧能力有限。

AR 模型随着存储点数或三角面数的增加, 精细度增强, 用户使用体验逐渐逼真。以真人建模为例, 精细化的 3D 人体模型需要 ~1,000,000 数量级的点云或 ~100,000 数量级的网格进行存储和传输。以目前行业压缩算法性能, 1,000,000 数量级的点云需约 190Mbps 的数据实时处理传输。

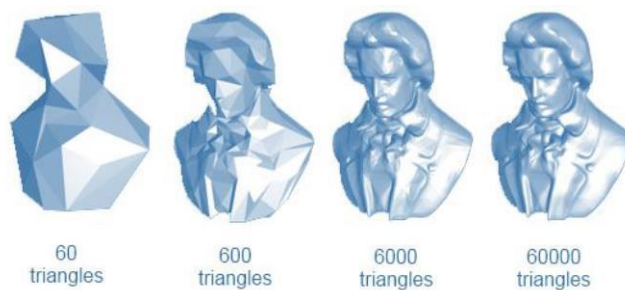


图 7. 建模三角面数示意图

实现虚拟物体的稳定叠加、复杂的遮挡关系, 都需要复杂的算法和大量的计算。消费级应用, 叠加精度要求低; 实物识别+叠加, 对精度提出更高要求; 虚拟图像在现实场景中穿梭, 需要考虑到更复杂的遮挡关系。



图 8. 消费级应用



图 9. 实物识别+叠加



图 10. 虚拟的皮卡丘在腿间穿梭

针对产业痛点问题，云化增强现实（Cloud AR）技术提供了新思路。产业界在 Cloud AR 领域探索，在终端推出 SDK 方案，解决了手机终端渲染能力低，功耗高等问题；云平台提供渲染能力，处理高精度传感器数据，提升模型精细度；传输侧提供流化传输方案，降低业务传输时延，同时 5G 网络的商用提供了时延更低的网络，可满足云 AR 交互响应需求。

（三）产业界及标准化组织在 Cloud AR 领域的探索，推动 Cloud AR 行业发展

近年来，国内厂商在网络架构、AR 开发平台、算法等方面都取得了不错的进展，并且在制造、医疗、旅游、营销、文化娱乐等领域的应用明显提升，典型案例不断涌现。国内各级政府陆续出台多项增强现实产业相关政策，并且多地政府纷纷建设增强现实/虚拟现实产业基地或启动规划，逐渐形成规模化的产业集群。

随着家庭千兆网络及 5G 网络的持续铺开，为 Cloud AR 业务提供宽阔通道，3GPP、GSMA 以及国内的标准化组织开始 Cloud AR 领域的标准化探索。

3GPP SA4 的技术报告 TR26.928, 提出了 XR 的端侧及边缘服务器的架构及相关技术模块, 包括端侧及边侧 XR 捕捉、传感器数据处理、场景生成、媒体解码、与 5G 相结合等技术。GSMA 发布了 AR/VR 白皮书, 其中提出了一下关键技术: App 的边缘部署、AR/VR 媒体实时编解码及传输、低延时动态 FOV、网络侧拼接等技术。ETSI 成立了增强现实框架 (ISG ARF) 新行业规范组, 开发具有相关组件和接口的增强现实 (AR) 框架, 以实现组件、业务和系统的互操作, 其提出了云化的 AR 系统及功能架构。国内标准化组织在国标和行标中纷纷定义了 AR 相关术语及技术架构, 推动了国内 AR 技术发展。

同时, 为促进 AR 技术和产业的快速发展, 各种增强现实相关的产业联盟或工作组相继成立。工业互联网产业联盟成立工业互联网 VR/AR 特设组, 开展 VR/AR 相关技术工业应用研究与推广, 发布了《工业虚拟(增强)现实应用场景白皮书(2019)》, 其中也提出了基于云化 AR/VR 的工业应用的技术现状和部署建议。中国增强现实核心技术产业联盟的成立旨在通过聚集各类资源形成创新生态, 推动中国 AR 产业核心技术的创新与发展。区域性的增强现实产业联盟, 如上海虚拟现实与增强现实产业联盟, 联合国内外产业链各个环节的企业, 共同建立适合中国市场的行业标准和产业规范, 规范和引导行业健康发展。

二、云化增强现实关键场景

随着 AR 技术发展以及终端设备的迭代，AR 应用越来越广泛。目前市场上的 AR 应用服务分为两类，一类为基于智能手机平台的 AR 服务，由于其成本低、用户基数大，已在工具类、电商类、游戏类 APP 中实现 2C 覆盖。另一类为基于 AR 眼镜的服务，主要在远程协助、物流/机房管理、军事领域实现 2B 商用。

AR 的应用服务需要根据场景和用户操作等信息进行展现，包括识别、跟踪、重建、融合、交互等典型应用。其中识别是指 AR 应用服务获取真实场景后，利用特定算法，对 AR 场景中真实世界的视频关键帧进行标示物特征点识别。跟踪是指识别之后 AR 应用需对视频帧的其他图片进行跟踪比对，寻找特征位置，完成对 AR 视频中标示物的识别跟踪。重建是指 AR 场景中虚拟场景的建立，目的是获取实际三维环境的三维数据，并根据应用的需要，利用获取的三维数据建立相应的虚拟模型。融合表示虚实融合显示，完成真实场景和虚拟对象信息的融合叠加，交互是将用户的交互操作输入，经过处理将交互的结果通过显示设备显示输出的过程，可以通过外接设备、特定标志以及徒手交互等方式实现。

随着上述服务场景及典型应用的升级，云化 AR 的需求逐渐显露，本章梳理了云化增强现实的关键场景。

（一）通话

Cloud AR 技术可为视频通话带来感知和交互升级，感知升级帮助用户更好地理解视频中的人、物、场景、语义，交互升级帮助用户更好地与视频中的人、物、场景联系起来，让视频通话不止于看。AR 与视频通话技术的紧密结合将开拓新的领域，产生巨大的市场前景，带来新的想象空间。具体的云化技术包括以下三个方面：

1).感知跟踪模型云端处理。 2D、3D 识别与跟踪技术，可以识别理解画面中的人物场景，为后续叠加特效、信息处理提供支持，衍生出通话特效、内容监管、可信通话身份等应用；场景语义分割技术，可帮助理解前景背景、场景语义，衍生出场景替换、特效处理、隐私保护等应用，解决传统视频通话形态可能存在的隐私泄露顾虑。

2).虚实环境融合云端渲染。 在现实世界视频画面的基础上渲染叠加真实、有用的虚拟世界。SLAM 技术可帮助虚拟物体空间定位，让虚拟的物体模型也知道自己到底“在哪儿”，理解真实环境中的高低起伏、障碍物、遮挡关系；通话协作技术通过场景进行精准定位，并发送给对方，将增强的信息放到场景的有效位置，实现协助、互动功能。

3).人机交互数据云端处理。 把人和视频中的人、物、场景联系起来，带来包括表情、手势、眼动、多模交互等新交互形态，手势/动作接挂电话、发送指令成为可能，也能帮助特殊人群跨越沟通障碍。

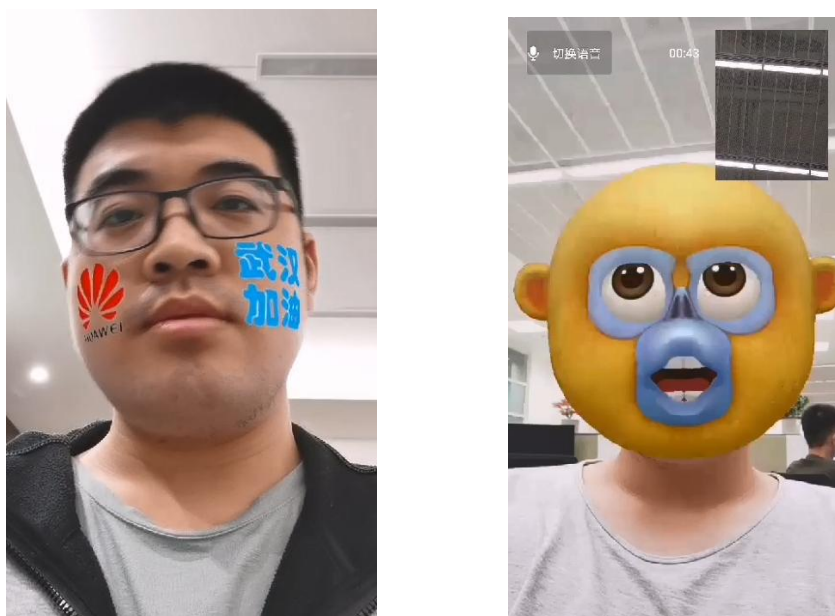


图 11. AR 通话模型实时驱动

（二）教育

AR 教育是结合 AR 技术，将传统教育中的静态图画等转化为动态可视化的三维声形互动内容，在增加学习的趣味性的同时，注重将复杂事物简单化，抽象事物具象化，隐含原理表象化。

将 AR 技术用于教育的发展历史较长，应用场景主要集中在 K-12 阶段的学科教育和科普教育之中。受硬件成本和普及程度影响，目前基于智能手机平台的 AR 教育应用占据市场主流，主要面向 C 端市场。通过 AR 技术所传授的知识则是自然科学多于人文科学，特别是对宇宙空间、生物环境的认知方面，可以起到重要的作用。



图 12. 智能手机 AR 科普读物

许多出版商把 AR 技术与儿童科普读物结合，把抽象难懂的科普知识变得生动形象，让文字信息与视觉、听觉信息同步，在阅读中也可以与之交互，使阅读成为一件轻松的事情。

在 AR 科普读物基础上，通过 AR 技术将丰富的地理知识融入一个地球仪，使地球仪从一个工具转变为一个载体，使其外延扩大。学生借助智能手机、平板电脑专用 APP 即可完成浏览，更加生动、形象地理解地理知识。



图 13 . AR 地球仪

AR 地球仪所涉及的知识涵盖地球结构、地质特点、洋流分布、天气现象、生物多样性、国家概况、旅游特色等多个方面，这些知识都不是一个传统地球仪所能够传递的，无论是对学生的课堂学习还是课外学习均能够起到帮助。

针对以上的 AR 教育应用场景，提出了“云端存储+本地渲染”的云化实现方式：

基于智能手机平台的 AR 教育应用普遍采用“云端存储+本地渲染”的实现方式，借助无线网络将三维模型等资源推送到本地，再由终端处理器完成渲染，其画质效果和时延受网络环境和手机终端质量影响较大。



图 14. 基于 AR 眼镜的教育内容

从上图可以明显看出，同样是展示地球，借助 AR 眼镜进行观看，其视觉效果和新奇程度都是手机 AR 所不能比拟的，有充分理由相信 AR 技术在学生对自然世界的认知上将会有所发展。但是 AR 眼镜的成像关键技术尚需进一步完善，对青少年视力影响也需进一步论证。除此之外，目前单台动辄万元的硬件价格也是一般学校和个人用户所不能承受的。

同时，由于 AR 眼镜的本地存储能力有限，承载内容有限，因而将内容存储于云端，并在中心侧或边缘侧完成渲染，借助 5G 网络传输到本地的方式就成为了未来的必然选择。由于这一方式降低了对终端设备的算力要求，设备成本也将大幅度降低。有理由相信，未来基于云+AR 眼镜的教育内容将进入校园和家庭。

（三）游戏

CloudAR 是统一多个智能终端虚实融合的一种云端增强现实能力。

针对云化 AR 游戏，主要是在云端提供 AR 相关信息存储及处理：云端储存的 AR 信息包括世界坐标系、特征点云 MAP 等信息，然后被智能终端检索，并融合现实与虚拟世界的信息。

AR 云的共享，要求多名玩家共享世界坐标系，每位玩家进入“共享”坐标系系统中，需要将自己的局部坐标系重定位到统一的全局地图中，并获得正确的位姿信息。接下来每一帧都会进行交互信息的同步。



图 15. Cloud AR 游戏

（四）地图

AR 地图是未来 5G 环境部署的一个关键场景，也是 AR 与 2C 业务场景下比较典型的代表之一。例如，华为发布的 AR 地图——河图（Cyberverse）旨在打造的一个地球级的、不断演进的，与现实无缝融合的

数字新世界。对于云化 AR 地图场景，需要多种云化技术进行计算、定位、渲染。

1).高精度空间计算能力云端处理。运用高精度空间计算能力将用户、空间与数字内容链接到了一起，通过 3D 高清地图的采集和生产即可完成物理世界的数字刻画，河图的 VPS 算法（视觉定位）、SLAM（同步定位与地图构建）等技术实现随时随地高精度的空间计算，无论是高楼林立的商业街区，还是传统定位导航无法胜任的室内环境，华为 AR 地图都能达到厘米级的定位能力和 1 度以内的定姿能力。在华为 AR 地图服务覆盖区域，用户能感受到所见即所需 3D 实景步行导引，既不用担心看不懂 2D 地图，也不用担心进入 GPS 信号弱的环境。



图 16. 华为 Cyberversion 地图

2).软硬件协同实现云端渲染。通过大量的软硬件协同优化，AR 地图能够实现高质量虚实遮挡、光照估计、逼真渲染，让虚拟物体看上去更加真实，与真实世界融为一体，带来沉浸式 AR 体验。在 AR 地图中，可以在任何地方稳定真实地放置 AR 内容，同时还会面向不同的用户提供个性

化的 AR 体验，将单一的物理世界变成个性化的数字世界。另一方面，针对 AR 地图的功能可以构建以地图为中心的生态服务，包括 3D 地图处理工具、3D 场景编辑工具、运营管理平台等工具或平台，帮助合作伙伴降低接入门槛。未来只需要一台手机或全景相机就可以在工具平台上完成新内容的编辑和注入，使用户共同构建 AR 地图的生态应用。

（五）工业

近些年来，工业 4.0 和智能制造的概念愈发火热，其中 AR 增强现实经常被提及，被认为是未来工业智造中不可忽视的一环，也是当前企业进行智能化转型的一把利器。

AR 在工业智能化的关键作用可以从 AR 在工业智能化应用的三个阶段进行阐述：

第一阶段：基于 AR 设备“第一视角，解放双手”优势实现基于远程音视频通讯和基础 AR 交互的远程指导，维修，培训。实现即需，即学，即会，即用。把互联网和信息技术的触角延伸到产业工人的身上，让工人初步享受到智能化和信息化带来的便捷。一方面解决了企业当前员工素质增次不齐的问题，并同步降低了工作中的错误率以及企业运营成本。另一方面培养员工能力，提升员工素质。

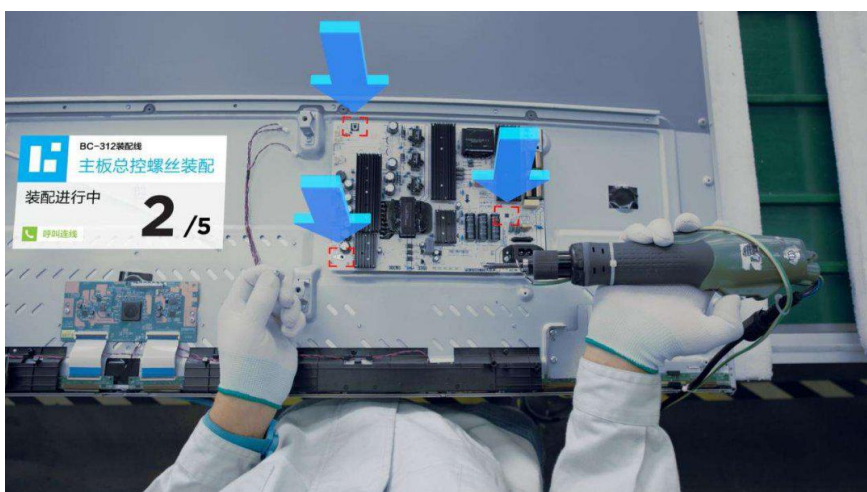


图 17. AR 远程指导

第二阶段：AR+物联网+数字孪生，数字孪生构建了一个对称的虚拟世界，而物联网的数据支持则能让这个世界鲜活起来，并与真实世界同步。从端侧数据采集到云端整理归纳再到端测实时反馈，形成一整套完善的数据应用链，在这个阶段中不再是单纯的人与人的交流，可以在任意地方获得来自云端的海量的数据支持。这个阶段里，AR应用已经初具智能化，信息获取速度更快、使用效率更高、数据更加透明，能够大幅提高工业设计和生产中效率。



图 18. 工厂 AR 数字孪生

第三阶段：AR+AI，该阶段在数据深度挖掘，和学习理解上有了巨大提升，这是AI赋予AR的强大力量，是真正智能化的体现。在这个阶段，AR系统可以自主进行分析，并通过对比以往的数据给出分析结果，指导人类完成后续工作。第三阶段的智能体现在智能识别与理解，智能分析与反馈，这对于精密化程度愈发苛刻的工业4.0时代而言，是提高生产效率，降低故障率的最佳方法。



图 19. AR+AI 应用

由此可见，在工业4.0的浪潮中，为快速满足新任务和生产活动的需求，AR应用有望成为下一个主流业务。工业AR的典型云化应用场景包含以下场景：

远程协作：通过AR远程协作系统，异地同事共享一个虚拟空间，精准回传第一视角视频，展示3D数字资产等，具体应用场景包括远程维修、异地设计研发等。

装配指导：通过将装配工作流程呈现在 AR 眼镜中，给予员工沉浸式的装配指导，提升装配效率。

设备巡检：用于工程师巡检产线和设备，基于产线上的设备数据的网联化，实现设备数据可视化可交互。

远程培训：利用 AR 技术实现最优资源和技术能力跨区域共享，将培训资料进行数字化处理，形成交互性强，沉浸式的学习资源。

仓储物流：通过 AR 和物联网融合技术，实现基于仓储的智能盘点，提升物料盘点效率和准确性。

（六）展馆

目前在各种会展中，每个参展企业受展台空间等因素限制，只能展现出非常有限的代表产品，现场参观的观众无法有效的获取更多信息，如：产品的性能、成功案例、与其他竞品的优势等。对于同属于展馆类别的博物馆亦是如此。博物馆蕴藏的历史、人文等丰富的信息，大多数游客只能走马观花，无法对文物等展品有更深入的理解。展馆场景对云化 AR 提出了云边协同的需求。

通过在中心云或 5G 边缘 MEC 上部署云化 AR 的智能识别和感知能力、虚实结合和 SLAM 跟踪定位能力平台，并配合部署在云端的中心管理平台，实现 AR 企业产品和 AR 博物馆等相关的 AR 识别类的应用。当智能终端对现实场景的参展产品拍照或摄像时，在画面上叠加产品每个细节的说明、生产工艺、性能参数、在实际环境中的表现、既往客户的评价等，使观众获得产品多维度、全范围的了解；对文物拍照或摄像，在画面

上叠加导览信息，如路线指引、互动讲解、文物修复、互动留念、历史重现等，提升用户参观博物馆的体验。

（七）党建

AR 党建，是将包括 VR 、 AR 、 3D 全息及动作捕捉和交互等在内的多项技术手段与丰富厚重的红色文化、新时代中国梦、以及十九大精神等内容相结合，通过新感官、新体验、新角度重新演绎，从而让人们更加深刻生动、身临其境的体验爱国主义历史精神和新时代建设成果，达到寓教于乐、严肃活泼的教育效果。党建场景对云化 AR 提出了云边协同的需求。

通过在中心云或者 5G 边缘 MEC 上部署云化 AR 的渲染能力、会话控制及媒体处理平台，协同云化 AR 的业务处理系统，并配合部署在云端的中心管理平台，实现高速率、低时延、高并发的数据传输，有效丰富党建学习的多样化教学手段、促进学习效果进一步提升。通过使用手机/PAD 等移动端，扫描现实景观、物品、二维码等，可以获得扩展的信息提示、影像叠加、情境再现等效果。通过 AR 互动，将枯燥的文字、图片等传统展示素材，赋予新的活力和更强的感染力，使学习或参观者留下更深刻的印象。使用动作捕捉识别和 AR 显示技术，与体验者的动作和影像互动，并投射在虚拟场景中，实现重温誓言，虚拟合影等功能，实现从外在到内在精神上与崇高信仰的高度融合、统一。

（八）安防

云化 AR 在安防中的应用包括公共安防、民用安防两个场景。

在公共安防领域中，利用云化 AR 提供的 AI 图像识别算力、数据挖掘与检索能力等，协助警方通过人脸图像识别系统进行身份核查；通过图像处理、文字识别等技术对车辆拍照的图像进行自动识别，同时在后台调取该车辆在公安的注册信息，确认车辆的状态。

在物业管理等民用安防领域中，除了上述基于人脸识别的身份核查、文字识别的车辆管控等以外，还可对物业内的主要楼宇或者建筑进行全向三维建模。当物业管理人员到达已经建模的楼宇附近时，使用 AR 眼镜对楼宇进行识别，通过调取后台该建筑内业主、配套设施的登记信息，可帮助物业管理人员快速获取巡防重点以及公共设施利用情况，提升服务效率。此外通过实时定位、可视化调度结合实时视频回传以及公网对讲，实现真正的可视化指挥调度集群对讲、巡更考勤。

三、云化 AR 总体技术方案

面对云化 AR 应用场景，梳理了云化 AR 总体方案。在平台侧，提供 AR 云化处理能力。将对时延敏感的能力下沉到边缘侧，包括 AR 数据处理、AR 渲染以及边缘分发能力，将业务系统部署在中心云上，主要负责用户管理、AR 数字资产管理以及边缘调度等业务相关服务。网络承载可以通过 4G/5G 网络或者在家中连接 WiFi 接入。

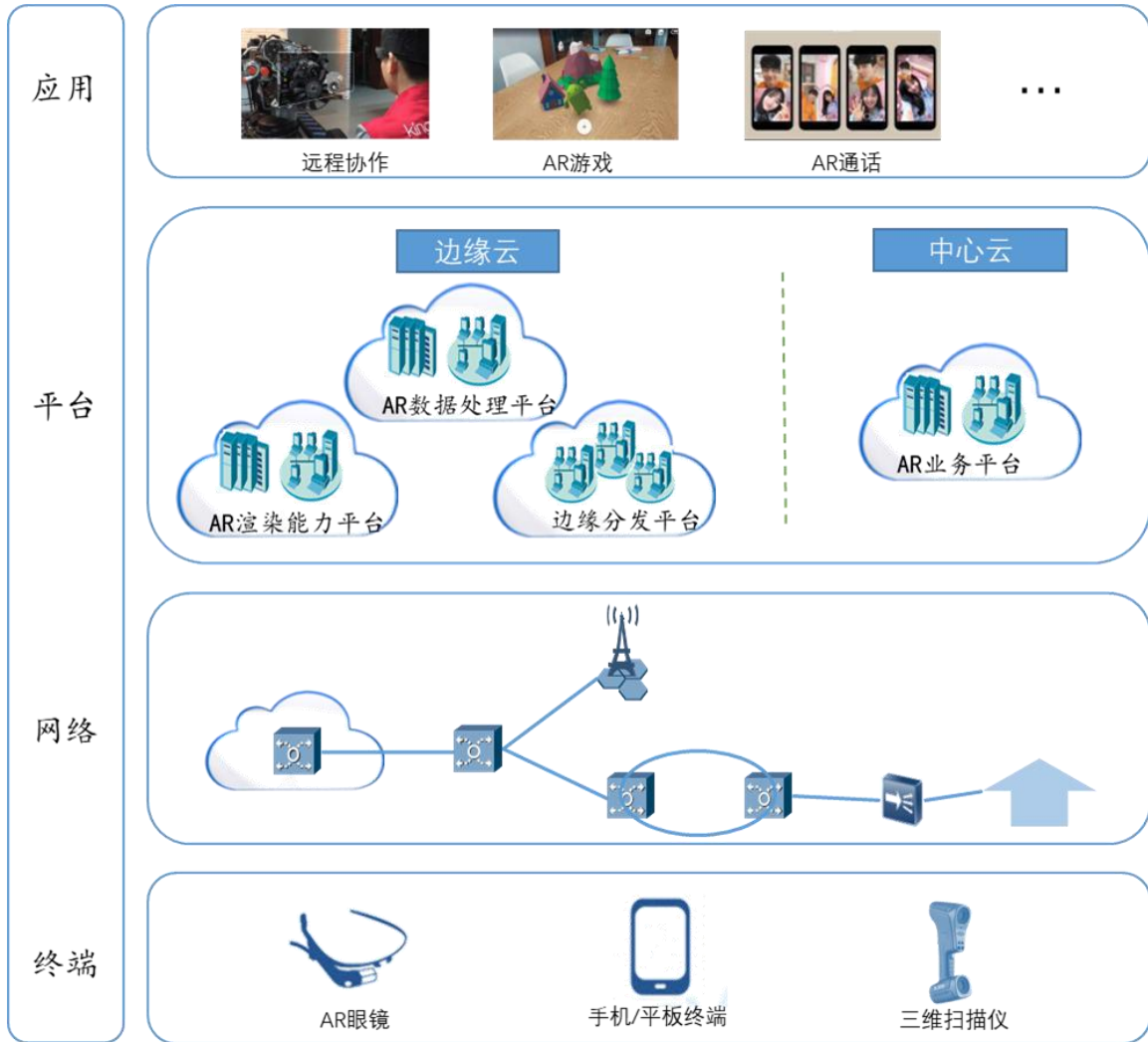


图 20. 云化 AR 总体架构

（一）终端处理能力云化降低端侧设备成本

当前的 AR 应用主要集中在手机端，主要原因在于一方面真正意义上的 AR 应用尚未普及，大众能够接触到的 AR 应用，主要集中在相对简单的 AR 滤镜类应用，以及小范围的 AR 展示等应用，对于终端性能要求不高。另一方面受限于 AR 眼镜终端的发展，目前的 AR 眼镜在性能、重量、

轻便性以及价格上，仍未被普通大众用户接受。AR 终端向轻便型过渡中，终端的计算能力将会大大降低，终端的成本也会随之降低。由于算力需求的骤减，对于电量的消耗也大幅减小，因此终端设备可以做得更加轻量化，或者更长的续航时间。

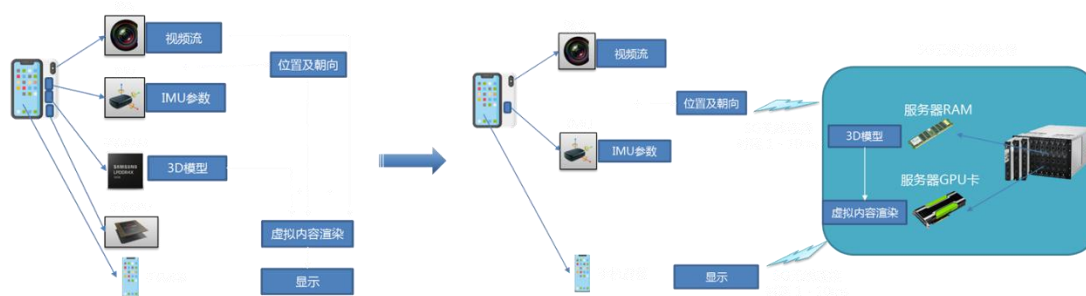


图 21. 终端计算上云

复杂度高的数据处理能力上云，在减低终端设备成本的同时，促进平台能力技术的发展。使得 AR 产业链中终端设备厂商更好的聚焦当前的设备解决方案。在 AR 终端处理能力上云的过程中，终端只负责必要的信息处理，例如：传感器信息采集、交互信息反馈、视频流解码显示等。

（二）网络传输技术满足 AR 应用云化需求

交互式 AR 应用将虚拟场景与现实场景相结合，传感器的实时定位，画面的实时渲染显示，是典型的计算密集型的低时延应用。

1). 5G 网络的商用以及家宽业务的发展，在网络上，保障了 AR 应用数据传输的稳定性及安全性。

2). 边缘计算的发展，促进云端计算和存储资源的进一步下沉，在 AR

内容传输及分发过程中，降低了数据传输时延。

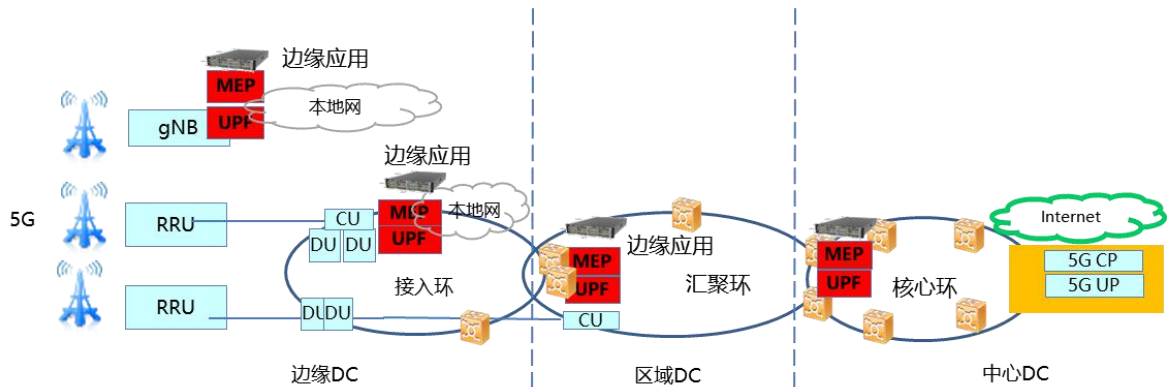


图 22. 边缘计算部署位置图

3). 优化的传输方案，有效的保证数据的流化传输方案，并行流化技术能够有效提高处理及传输效率，大幅度减少云化时延。

1. 传输内容

云化 AR 基于追踪计算、虚拟渲染上云的方式，为了达到近似或超越本地终端运行的效果，导致其对网络传输速率、传输时延和稳定性都具有较高的要求。结合 AR 服务的特征，其网络传输的主要信息包括：

1).终端向云端传输：本地化实景信息、终端测操控指令、交互同步信息等。

a) 本地化实景信息：构建 AR 服务的真实世界信息

- i. **实景视频：**即终端测将实际拍摄的视频上传到云端，并由云端完成针对实景视频实时位置追踪、三维重构和虚拟渲染等。
- ii. **实景逻辑：**即终端本地完成实景视频的追踪和三维重构后，将相应的实景逻辑上传到云端，云端仅处理有关虚拟渲染部分。

- iii. **位置信息：**即终端测只需传送与实景有关的位置信息（GPS/北斗等定位信息等），云端根据资源库已有的实景资源、实景逻辑，完成虚拟渲染。

注：上述三种情况，能够满足不同的 AR 应用场景需求，该部分相对云 VR 而言需要更大的上行带宽。

- b) **终端测操作指令：**包括各种触屏操控、手柄、手势、语音等操作方式，主要是针对虚拟景象的操作指令。其数据包一般为 KB 级，上行带宽要求较低，主要对网络时延要求较高。
- c) **交互同步信息：**是指多人 AR 场景中，彼此互动关联的逻辑和数据交换，包括云端和终端双向传输，其中终端向云端传输的是多人交互的逻辑信息。其主要对上行网络通道的时延具有要求。

因此，终端向云端传输过程中主要是对上行网络通道具有要求，尤其是实景视频实时上传的场景对上行带宽将带来相对云 VR 而言数百倍以上提升。

2).云端向终端传输：渲染处理结果、交互同步信息等。

- a) **渲染处理结果：**云端完成相关的处理后，根据不同的技术处理方法和相应的应用场景，向终端传输的相关虚景、虚实融合或虚景逻辑等：
 - i. **虚景：**云端将完成渲染的虚景，包括虚拟场景、虚拟角色、光影特效等下发到终端，并由终端按照三维重构逻辑完成虚实结合。
 - ii. **虚景逻辑：**云端只需在位置追踪、三维重构技术基础上，完成

虚景的逻辑计算，并将其下发到终端，由终端根据虚景逻辑完成虚景渲染和虚实融合。

注：其中云端下发虚景逻辑的场景，对终端的能力要求较高，一般应用于终端能力较强，且需要多终端超低时延互动的应用场景。

b) 交互同步信息：包括平台向终端发送的通知、多用户之间的交流互动等信息。

因此，云端对终端传输过程中，主要是对下行网络通道具有要求，尤其是渲染处理结果下发过程中，如果需要将虚实融合结果下发，将对下行网络带宽具有较高的要求。

2. 传输需求

针对不同的云化 AR 处理技术，云端和终端双向传输的内容具有较大差别，不过就当前终端能力普遍不足，为了更好地实现多人低时延互动的目的，采取终端测将实景实时上传，由云端完成位置追踪、三维重构、虚景渲染以及虚实融合逻辑，最终通过网络将完成 AR 全部处理的数据（虚景）下发到终端，终端测仅做最终的展示。

相对于云 VR 而言，云 AR 可能存在云端和终端双向的高清视频流，对于网络传输速率、时延和稳定性等要求更好，需要采用超低时延流化处理和实时端边协同等技术创新，降低其对网络传输的压力，提升用户体验。

（三）云平台构建支撑多种云化 AR 应用

云化 AR，不同于云 VR 强互动业务，例如云 VR 游戏等重度应用，更多是聚合众多创新小应用，这就造成了云化 AR 需要一个强有力的平台作为支撑。因此，云 AR 平台需要具备 AR 服务、内容聚合、服务分发、资源调度等能力。

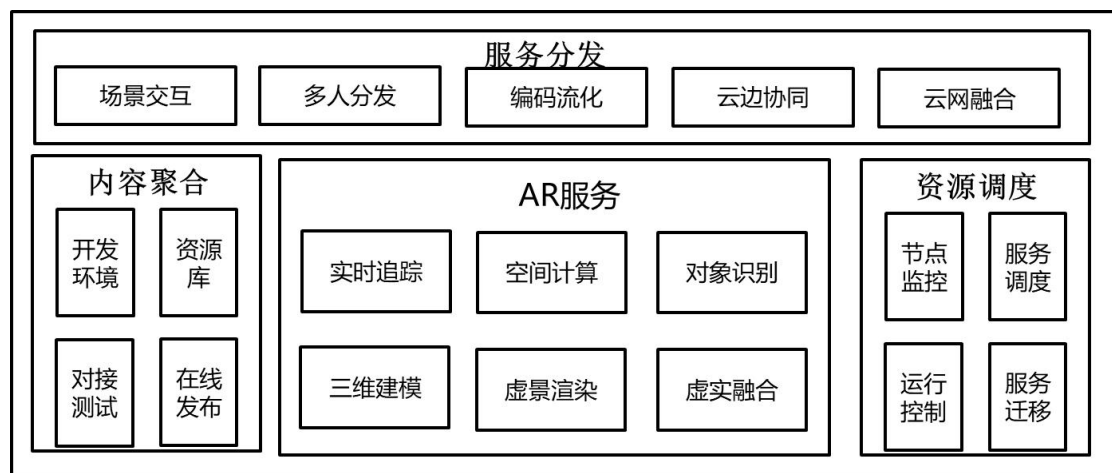


图 23. 云化 AR 平台架构

1).AR 服务：云 AR 平台的核心能力，主要提供实时追踪、空间计算、对象识别、三维建模、虚景渲染和虚实融合等功能：

- a) 实时追踪：是指能够根据终端测实时上行的实景视频，并结合终端传感器或景深镜头信息，确定需要进行 AR 处理部分进行距离、方位、景深等定位标识。
- b) 空间计算：针对实景部分内容，根据实时追踪结果，通过数字化确定实景中锚定物品之间的实际位置关系。
- c) 对象识别：针对实景中的各种物品进行有效识别处理，主要包括平台识别、空间确定和虚景关联物品锁定等。

- d) 三维建模：将完成数字化位置和物品识别的实景，完成虚拟的三维模型建立，为虚景的渲染奠定基础。
- e) 虚景渲染：基于实景的三维模型，按照虚拟景象逻辑，渲染出相应的虚拟场景、虚拟角色和光影特效等。
- f) 虚实融合：将实景和虚景完成相互位置关系、图层关系、交互关系等方面进行对应，实现融合输出逻辑关系。

2).资源调度：作为云 AR 平台的资源调度和管理部分，只要负责节点监控、服务调度、运行控制和服务迁移等功能。

- a) 节点监控：是指针对 AR 服务节点进行状态监控，包括资源占用、设备故障、安全隐患等相关内容。
- b) 服务调度：响应终端测请求，按照就近、空闲和能力有效等原则，将服务分发到对应的服务节点上的过程。
- c) 运行控制：按照统一的能力分配，针对请求服务，完成对应服务端上的并发承载运行、状况监控等。
- d) 服务迁移：当出现某一具体 AR 服务或服务节点故障时，能够实时将该服务重新开启或迁移到其他服务节点。

3).内容聚合：是指为了实现 AR 能力开放和应用开发，面向内容提供商和开发商，提供开发环境、资源库、对接测试和在线发布等功能。

- a) 开发环境：是指基于平台开发所需的开发软件、软件开发工具包、编程接口等，为开发者提供一站式 AR 应用开发服务。
- b) 资源库：是指支持在线开发的实景视频库、虚拟资源库等，方便开发者在线调试和开发操作。

c) 对接测试：开发者将其开发完成的 AR 应用，在云 AR 平台试运行，并与其本地连接进行功能测试等。

d) 在线发布：开发者将完成测试，并经过平台方认证通过后的应用，在云 AR 平台对外发布。

4).服务分发：该部分主要完成云 AR 服务的交互、网络传输、云边协同、云网融合等功能。

a) 场景交互：接收终端测的操控指令和传感器距离、方位信息，并完成针对 AR 虚景的操控转译，传递给 AR 服务部分，实现对用户操控的逻辑计算和场景渲染回应。

b) 多人分发：针对同一 AR 应用需要同时提供给多个终端用户的场景，可采用 CDN 等方式实现多人同步、高效分发。

c) 编码流化：是指将 AR 服务输出的画面、声音等，压缩编码成为视频流，并采用串流方式下发到终端测。

d) 云边协同：针对多人互动场景，云端完成相关的 AR 逻辑运算后，将渲染任务发送给靠近服务对象的边缘节点。该功能就是针对这一场景的管理、协调和资源调度。

e) 云网融合：实时采集云端和终端网络传输质量，并自适应网络状况调整云端输出的码流大小、云端协作分工等，以便实现最佳的用户体验效果。

（四）中心云能力下沉减轻云端压力

传统的 AR 应用，中心云的处理及分发压力，难以保证多终端低延时的业务应用需求。视频信息以及其他传感器采集的信息，需要上传到云端服务器，其中视频信息上传至视频服务中心，传感器数据上传到数据处理中心。当数据处理中心将 2D 数据转化为 3D 数据后，再将数据模型渲染到视频中，对视频信息进行编转码，以视频流的形式分发到客户端。

云化 AR 应用，AR 云平台数据处理能力及渲染能力的下沉，减轻了中心云的压力。中心云集中负责业务处理，对外提供业务能力，对接各种 AR 应用内容，为其进行分发；对内中心云统一用户管理，AR 终端设备管理，提供高效的运维服务，对边缘平台进行调度，优化 AR 内容分发链路。

（五）边缘云提供强计算能力

边缘计算在靠近数据源或用户的地方提供计算、存储等基础设施，并为边缘应用提供边缘能力服务和 IT 环境服务。相比于集中部署的云计算服务，边缘计算解决了时延过长、汇聚流量过大等问题，为实时性和带宽密集型业务提供了更好的支持。

AR 业务云化，将原本位于端侧的传感器数据处理、复杂场景生成、复杂渲染功能移动至边缘侧，边缘侧同时配合具备 2D 或 3D 编码、元数据生成能力。

1). **降低时延**：边缘实现实时位置注册，确保端侧呈现时延在 30ms 以内

2). **算力上浮**：3D 建模渲染计算从端侧上浮到边侧，降低端侧成本

3). 多终端数据共享：多个终端数据处理集中在边侧处理，为实现多人业务互动提供了条件



四、云化 AR 技术方案

图 24. 云化 AR 技术方案框图

云化 AR 技术方案的框图如图所示，主要包含终端、网络传输与平台等技术方案，用于支撑各个应用场景的实现。

（一）终端技术方案

1. 主流 SDK

端侧目前主流的形态为 AR 眼镜与手机，主要的呈现方式不同。底层包含 CPU、GPU 或 NPU 等支撑基础。依托不同的 OS（Android、IOS）操作系统完成终端与云端的数据交互。另外，通过调用 AREngine、ARCore 与 ARKit 等 SDK 完成各个数据的采集。下面针对一些支撑 AR 采集与呈现的主流 SDK 做具体的描述。

1). HUAWEI AR Engine

HUAWEI AR Engine 是一个用于在 Android 上构建增强现实应用的平台。HUAWEI AR Engine 通过垂直整合 AR 核心算法和海思麒麟芯片提供 AR 基础能力，目前提供了运动跟踪、平面检测、光照估计和命中检测、手势识别和指关节跟踪、人体姿态识别和骨骼跟踪、人体 Mask、图像跟踪、环境 Mesh、人脸表情跟踪和人脸 Mesh。

2). Google ARCore

ARCore 是 Google 的增强现实体验构建平台。ARCore 利用不同的 API 让用户的手机能够感知其环境、理解现实世界并与信息进行交互。目前提供了运动跟踪、环境理解、光照估计、锚点和可追踪对象、增强图像与共享体验等 AR 呈现。

3). Apple ARKit

Apple ARKit 是一个基于 IOS 操作系统的增强现实应用平台。它通过某种方式将二维或三维元素添加到设备摄像机的实时视图中,从而使这些元素看起来与现实世界融合。ARKit 结合了设备运动跟踪、相机场景捕获、高级场景处理和显示便利性,简化了构建 AR 体验的任务。

4). 商汤 SenseAR

SenseAR 平台是商汤科技基于机器视觉和深度学习算法与大量的数据积累,为手机、平板电脑等智能终端实现空间定位、虚实融合、人脸解锁、人脸支付、AR 特效、双摄虚化、智能相册、人像光效、手势识别等功能方案提供的一个整体式 AR 解决方案。SenseAR 平台包括 AR 底层驱动引擎、AR 内容创作工具链和 AR 硬件装置推荐,为不同行业的商家和用户提供了完备的 AR 服务。

该解决方案实现了跨平台支持(支持 Android、iOS、Linux、Windows、MacOS 等操作系统),提供一致的调用接口,并已作为底层能力支持 Unity AR Foundation。

2. 光学显示方案

光学显示技术是 AR 设备的核心,也是影响 AR 硬件设备体积和显示效果的最大因素,AR 显示从虚实结合的形式上可以分为视频透视和光学透视两种。视频透视在传统显示设备中呈现虚实结合的视频效果,其显示技术就是传统的 2D 或者 3D 显示技术。光学透视只呈现虚拟信息,虚拟信息与现实世界的叠加最终是在使用者的感官上达成的,即使用者看到虚

拟信息投射到眼睛的同时，从同一个方向还可以透过显示系统看到外部世界的信息，这就需要在传统的显示技术的基础上进行更多更复杂的光学设计，从而实现这种“半透半反”的效果。

光学透视 AR 设备的光学显示系统通常由微型显示屏和光学元件组成，微型显示技术目前主要以 LCoS（液晶附硅显示）、HTPS-TFT LCD（高温多晶硅薄膜晶体管液晶显示）、OLED（有机发光二极管显示）、AMOLED（有源矩阵有机发光二极管显示）为主，而未来的趋势逐渐向 DLP（数字光处理）、Micro LED（微型发光二极管显示）及 MEMS（LBS）（微型激光束扫描投影显示）技术发展。光学显示方案目前市场上主要包括棱镜、自由曲面、Birdbath、光波导等几种不同的技术方案，其中，光波导具有镜片式、全透明的优点，是消费级 AR 眼镜的主要解决方案。近年来，得益于微型显示技术的进步，AR 体验效果从清晰度、视场角、亮度等各方面都得到了提升。目前市场上面向 C 端的 AR 眼镜产品分辨率均达到了 720P 或 1080P，视场角基于不同的光学方案，最高可达到 120°。

综合来说，端侧的硬件、操作系统及 SDK 都能独立呈现出 AR 的效果，但由于算力与成本的限制，使其无法提供更高的效果。但是，利用端侧提供的各种采集的数据并依托云侧强大的算力可以为 AR 的应用提高另一个台阶。

（二）网络能力及传输技术方案

1. 超低时延流化技术

与本地 AR 不同，Cloud AR 在云端进行计算和渲染后，通过网络以实时通信（RTC）的方式传输给终端，端到端的处理时延直接决定了用户的 AR 体验，较长的云化时延将导致用户失去沉浸感，重则导致动晕症的不适。因此超低时延流化技术是用户能够良好体验 Cloud AR 的重要前提。在该技术中采用并行流化技术能够有效提高处理及传输效率，大幅度减少云化时延。另外，为了避免网络丢包，通过动态增加冗余 FEC 纠错包的方式能够在避免数据丢失的前提下降低带宽占用，进而提高用户的 AR 体验。

1) 并行流化技术

通常流程将编码、传输及解码串行处理，使得画面进行编码时，传输和解码区域处于空闲状态，利用率极低。编码、传输和解码并行处理，是降低系统时延的一种非常有效的方法。非实时场景，并行处理通常可以以帧为粒度进行，编码、传输和解码可以同时处理不同的连续帧。但对于实时场景，尤其是超低时延场景，以帧为粒度，会导致时延增加，因此需要更小的并行粒度。

1). 基于Slice的并行处理

为了满足 Cloud AR 端到端超低时延传输要求，可以采用以 Slice 为粒度编码、传输和解码并行操作，将各部分操作效率最大化，极大降低了端到端的时延。该技术将一帧画面分为多个 Slice 进行编码，编码器完成一

个 Slice 的编码后，将其发送给终端，终端在收到 Slice 后开始进行解码传输和解码流程同样进行如上操作，每当处理完一个区域画面后直接发送给下一级，使得各模块能够对一帧画面的不同区域进行并发处理，大幅节省了处理时间，降低了云化时延。基于 Slice 的并行处理方案示意图如下：

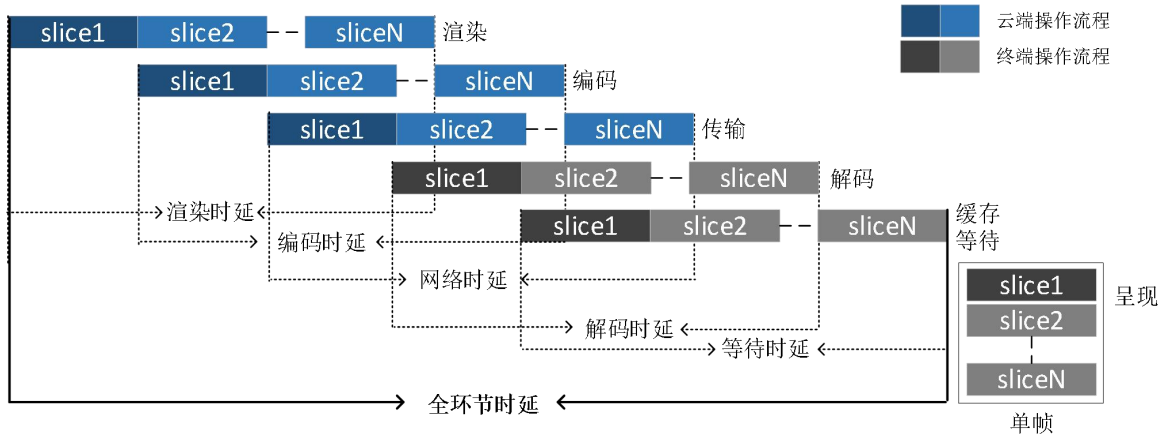


图 25. 基于 Slice 的并行处理方案

但这种方案存在的问题是要求终端能够支持该解码方式，因此为了考虑兼容性提出下文基于左右眼的并行处理方案，在无需对终端提出额外需求的基础上通过并行处理降低端到端时延。

2). 基于左右眼的并行处理

除了上述将一帧画面分为不同区域分别进行操作的技术外，还能够对左眼画面和右眼画面编码、传输、解码流程进行并发操作。通常流程是在编码器对左眼画面和右眼画面全部完成编码任务后再将图像发送给终端，而终端需要将左眼画面和右眼画面都接受完毕后才开始解码工作。而左右眼分别编解码技术是将左眼画面编码完成后直接将图像发送给终端再开始对右眼画面的编码操作。同样，终端在接收完成左眼图像后不必等待右

眼图像，能够直接对接收的图像进行解码。该技术将编码和解码操作并发执行，有效降低了端到端的时延。左右眼分别编解码技术示意图如下。基于左右眼的并行处理技术示意图如下所示：

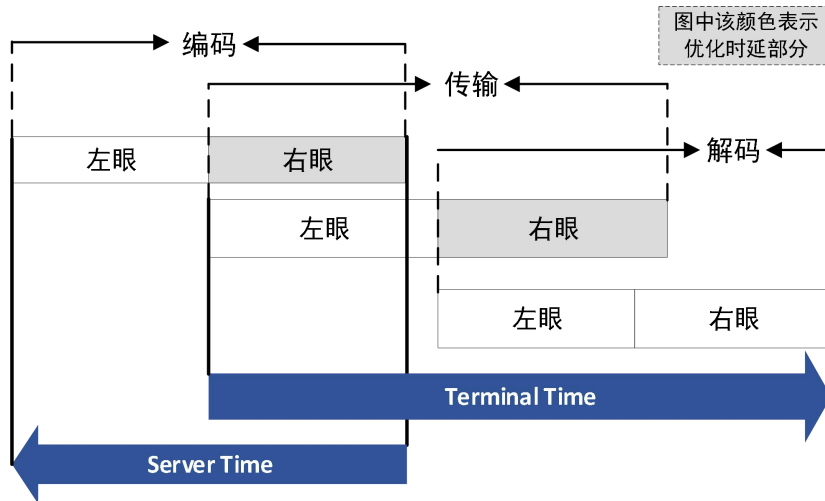


图 26. 左右眼分别编解码技术示意图

2) 低时延传输技术

在 Cloud AR 实际应用中，云化时延将直接影响到用户的体验，因此在入门体验阶段需要将云化时延控制在 50 毫秒以内，要想达到极致体验则需要将其控制在 30 毫秒以内。为了达到这一程度，超低时延传输技术成为 Cloud AR 不可或缺的关键技术方案。在数据传输过程中利用 FEC 丢包纠错方式减少重传，并根据网络状态动态调整 FEC 达到减少带宽占用的效果。在云端到终端的编码、传输和解码过程中采取编码\传输\解码并行操作或左右眼分别编解码技术，能够有效提高计算和传输效率，降低云化时延，提高用户 AR 体验。

1). 利用FEC减少重传

在云端到终端的信息传输过程中，对于产生的丢包情况主要有两种处理方式，ARQ 丢包重传方式和 FEC 丢包纠错方式。在采用 ARQ 方式时，会加大端到端时延，若网络发生拥塞丢包是，AQR 会加大重传数据量，导致拥塞更严重。而采用 FEC 方式牺牲少量的带宽添加冗余纠错包，当发生丢包时，能够通过纠错包恢复丢失的数据，可以容忍网络低于 $1e-6\sim 1e-5$ 的丢包，减少数据重传。

2). 动态FEC减少带宽占用

为了更进一步减少网络时延，使用动态 FEC 技术通过判断网络状态来动态调整 FEC 纠错包的数量，当网络状态好的情况下去掉不必要的纠错包，减少数据发送量，有效减少带宽占用，提高传输效率；当网络状态较差时，增加纠错包数量，能够有效避免网络丢包造成的数据损失。

2. 实时端边协同

云 AR 需要超低时延为保障，将服务节点前置到用户侧，例如 5G 接入网，基于边缘计算提升用户体验，是一种有效的解决方案。边缘节点最大的优点就是更加靠近用户，而且云 AR 在不采用云端实景预设方式情况下，需要依靠终端测实时追踪和定位能力，实时端边协同就成为了关键。

实时端边协同技术，主要包括追踪和定位协同技术、端边渲染协同技术、端边识别检测协同技术和终端托管协同等，实现融合边缘计算节点和终端能力，为用户提供极致的云 AR 体验。

1). **追踪和定位技术协同**：依托终端测的传感器和视觉系统，完成针对周边环境的扫描和追踪之后，将相应的信息极速传送到边缘节点，由边缘节点针对实景完成定位处理。

2). **端边渲染协同技术**：同时融合终端和边缘节点的计算能力，优化网络传输，在边缘节点完成 AR 相关的定位、三维建模、虚景逻辑运算等功能后，将对应的计算结果传输到终端，直接调用终端测的渲染能力完成虚景渲染以及虚实结合输出。基于该种处理技术，即能充分利用边缘节点计算能力，又可基于终端渲染能力实现虚实场景无法对接，剔除大量的数据或视频流交互，能够有效减少网络传输压力。

3). **端边识别检测协同技术**：依托终端的视觉系统和深度传感器，完成对二维及三维物体的识别后，将识别的信息传送到云端，由云端节点完成对实景的检测和匹配工作。

4). **终端托管协同技术**：基于边缘节点无限靠近终端的特点，实现边缘计算节点接管终端的各种传感器、视觉系统的方式，减少终端与边缘节点数据传输的回路，有效降低网络传输时延。

3. 网络原生优化

为了提供更加优质的网络传输服务，针对云 AR 业务，需要在承载网、核心网层面进行调优，减少网络传输路径，定向提升业务上下行带宽，以便满足云 AR 业务近似苛刻的网络传输要求。网络优化技术方面，包括 VPN 加速技术、基于 Radius 网络原生优化技术等。其中基于 Radius 网元

设备的网络原生优化技术，是指在网络认证服务器 Radius 上增加网络调优服务能力，通过其认证后，即可将对应的云 AR 业务打标，并通过与承载网上加速节点衔接，完成路径调优减少时延；同时通过临时或个性化调整上下行带宽，增加网络传输能力，其相对 VPN 等技术具有先天独到的优势。

Radius 网络原生优化技术，可以理解为在网络认证服务器上增加网络调优的边缘计算功能，无需终端侧安装客户端，降低了用户体验网络优化的门槛，并实现用户、云 AR 业务、网络的多重绑定和个性化网络优化。

4. 数据传输内容

1) Action 数据传输

Action 数据是指由用户或真实环境的动作或状态发生改变后，被端侧传感器采集并转换成高帧率的控制数据，这些控制数据依照特定的协议转换成相应的机器语言，发送至云侧指导或触发云上构建的虚拟环境的模型或场景发生动作或状态的改变。Action 数据桥接了真实场景与虚拟场景，完成一系列虚实动作的互动。Action 数据传输过程中包含了针对 Action 数据特有的抗丢包方法及编码协议。

2) 点云数据传输

点云数据可以表示一个采集的对象点的 RGB 颜色，灰度值，深度，分割结果。在 Cloud AR 系统中，点云数据被端侧的传感器所采集并传输至云端，为后续的 SLAM 及高精度空间计算提供了可靠的依据，点云数

据的精确程度决定了虚拟空间与真实空间的契合度。由于点云数据内容较大，为了节省传输带宽，传输时需要对点云数据进行编解码。

（三）平台技术方案

云 AR 平台需要依托 AR 技术框架云化、具备多路承载和多应用并行等能力，以便支持不同终端的服务请求。

1. AR 框架云化技术

实现将单机版 AR 运行框架云化，包括跟踪注册能力上云、对象识别能力上云、三维建模上云、虚景渲染能力上云等，并以相应的接口将云化 AR 框架开放给应用开发者。

1). **跟踪注册能力上云**：根据其实现技术手段不同，可分为实时终端协助上云和实景预设能力上云两种方式。

a) **实时终端协助上云**：即需要终端测实时定位跟踪能力，协助平台测完成实时追踪、空间计算和三维建模过程，包括终端测向云端传递实景视频+定位跟踪信息以及实景逻辑+定位信息两种不同的实现方式。

i. **终端提供实景视频+定位跟踪信息**：终端测仅提供拍摄的实景、各种传感器/机器视觉跟踪信息等；平台测根据终端上报的原始跟踪信息，完成实景位置锚定和空间计算。

ii. 终端提供实景逻辑+定位跟踪信息：即在终端测需完成实景相对位置分析，并将其结果上报到平台；平台测主要完成对应的实景相关空间位置标识。

b) 实景预设能力上云：是指云端平台预先完成实景的先期处理，包括针对实际场景的追踪、空间定位和三维建模。在实际运行过程中，终端仅需提供实景标识，例如地址坐标、物品标识等。

2). 对象识别能力上云：平台测具有相对终端数十倍的计算能力提升，采用深度学习等技术手段，相比终端具备更加突出的对象识别能力。该种能力完成上云后，能够提供更加精确化的服务体验。

3). 三维建模能力上云：在跟踪注册和对象识别的基础上，实现针对实际场景的数字化处理，包括空间距离/方位关系、平面范围/层次、光线角度/明亮等，为后续的虚景构建承载的基础。

4). 虚景渲染能力上云：在云端完成对应的虚拟场景逻辑关系计算、实景三维建模位置锚定、虚景元素（地图图素、人物贴图、光影特效等）调用等，最终实现虚拟场景渲染输出。

2. 多路并发承载技术

为了更好地实现平台测计算能力复用，需要基于单台服务器构建面向多用户的相互隔离的云 AR 运行环境，实现多路并发承载。其主要依托于服务器底层的 CPU、GPU 虚拟化技术。CPU 虚拟化方面，相对较为成熟，基于多核并行原理实现，即可实现有效的环境隔离；GPU 虚拟化方面，

技术实现上较为复杂，实现方式上包括 GPU 直通技术、虚拟机技术、GPU 共享技术等。

1). GPU 直通技术：即 GPU pass-through 直通，是将 PCI 插槽上的物理设备（这里特指 GPU 设备）绕过 OS 层，直接提供给了 VM 或并行运行环境。该种技术可以做到将一块物理显卡直接提供给某一 AR 应用使用。其主要优点是硬件性能具有完备性和独占性，同时存在资源无法复用、不支持无缝迁移、缺乏运行监控能力等问题。

2). 虚拟机技术：采用 GPU 切片技术，根据需要将 GPU 切分成不同规格的时间片，分配给多台虚拟机使用，并允许虚拟机直接访问 GPU 显存等核心资源，真正实现无缝迁移和 GPU 资源的复用。该种技术路线需要依托专属的虚拟机软件，通常情况下会产生对应的虚拟机授权费用。

3). GPU 共享技术：基于 GPU 切片技术原理，采用中间件模式采用基于 API 转发机制，实现专有图形库（GPU 显存），从并行运行环境转发到运行监控器中，完成并发环境运行过程。其主要优点是具备按照 AR 应用实现个性化 GPU 显存分配。因为采用 API 转发机制，该方法存在一定程度上的性能上损失和 OS 的兼容性问题。

3. 多应用并行技术

多应用并行，是以多路并发承载技术为基础，实现多种 AR 应用基于云端同台服务器运行的技术，需要多维度认证、资源分布式存储、应用灵活迁移等技术为基础。

1). **多维度认证技术**：构建用户、终端设备、应用权限的多维度认证体系，完成用户鉴权、终端注册和应用鉴权，为后续用户、终端和应用三者对应关系奠定基础。

2). **资源分布式存储**：相关资源包括 AR 资源库、用户 AR 应用历史存档等。为了确保不同应用跨服务器调度运行，需要采用分布式存储以及缓存机制，实现资源的快捷、高效载入。

3). **应用灵活迁移**：当出现运行应用的并行环境或者服务器故障时，支持应用在同台服务器快速重启或不同服务器间的无缝迁移。

（四）中心云能力及技术方案

1). 云 AR 内容中心

内容中心聚合来自 AR 内容商以及互联网 AR 内容库的原始 AR 内容。用专业的 AR 内容编辑软件制作和拼接处理，注入到内容源站供 AR 业务中心调用。云 AR 内容中心同时提供版权登记服务，原创 AR 内容可申请版权，保护 AR 制作团队权益。

2). 云 AR 业务中心

业务中心对 AR 内容转码、封装、打包、加密、流化，通过 CDN 分发。业务中心对 AR 业务进行管理，例如：业务订购、资源分配、生命周期管理等；同时，运营 AR 业务，挖掘 AR 内容的市场价值，例如：智能搜索、业务推荐、智能广告等。

3). 云 AR 管理中心

管理中心负责云平台的软硬件资源，为 AR 业务提供基础能力支撑。管理的硬件资源包括服务站点的 CPU、GPU、智能网卡、高速存储等；软件资源包括虚拟环境、5G 网络等。同时，对资源的安装部署、升级维护以及生命周期进行管理。云 AR 边缘云部署在地市及以下边缘机房，如 5G 本地核心机房，位置与周边区域的用户密度和站点容量相关。受制于机房空间、散热、成本等因素，边缘需要支持轻量化、分布式、多节点共部署。

（五）边缘云能力及技术方案

1. 实时建模

在 AR 场景中，如果建立的模型均为已经设定的或不与当前真实环境不契合的，会使用户体验的内容过于单调，若当前的真实环境的某个人或物可以被采集并实时的建模，最后在场景中被创建，可以提高虚实结合的互动性。在 Cloud AR 系统中，强大的计算能力为实时建模提供了基础。当端侧采集 AR 场景下真实人或物的特征信息，并传输至边缘云后，边缘云将根据这些特征在虚拟空间中实时建立相似的模型。使用户体可以与实时建立的模型产生互动。



Fig. 1. Reconstruction pipeline. 1) Generate a points cloud. 2) Extract a visibility-consistency mesh. 3) Refine the mesh with photo-consistency optimization and regularization.

图 27. 建筑场景三维实时建模示意图

2. 实时跟踪

在 Cloud AR 系统中，涉及到对某个真实的物体进行实时的跟踪，从而根据真实物体的状态的变化作用于虚拟模型等一系列操作。通常跟踪的真实对象是已选定的，在端侧完成实时采集的图像经过传输后到达边缘云侧，边缘云将通过跟踪算法对采集的内容进行检索，在发现对应的物体后，输出该物体的空间及其状态等信息。系统可以利用这些信息实时的作用于设定的 AR 模型标识，接着对已经改变的虚拟物体进行渲染。最后将生成的画面下发至端侧呈现。

3. 模型实时驱动

在 Cloud AR 场景中，为了实现虚拟物体与真实物体的交互，常采用模型实时驱动技术。一般为真实物体作用于虚拟物体，即利用传感器捕捉当前场景下真实的人或物的动作与状态的变化，并将采集到的信息生成 Action 数据，从而去改变虚拟模型的动作或状态，以此实现实际场景与虚拟场景之间的互动关系。典型的有 AR 人像模型的实时驱动。首先，传感器采集真实场景中人物的如抬手，闭眼或张嘴等动作，经过翻译后生成 Action 数据传输至边缘云，再对 AR 场景下对某一个虚拟的 AR 人像模型进行实时的驱动，最后渲染并将生成的画面下发至端侧呈现。



图 28. AR 通话模型实时驱动

4. SLAM 及高精度空间计算

SLAM（Simultaneous Localization and Mapping）意为同时定位与建图技术，在实际场景应用中，精确的定位依赖与当前地图的精度，而高精度地图的每个元素的构建都依赖于高精度的定位，定位与建图的相互依赖关系催生了 SLAM 技术。高健壮性、高精度、实时性的视觉 SLAM 对于 AR 场景具有非常重要的意义。在 Cloud AR 场景中，为了达到高逼真度的虚实结合的效果，虚拟物体在真实场景中的空间定位就依赖于 SLAM 技术，通过 SLAM 及高精度空间计算，使虚拟空间定位与真实空间定位高度契合，从而有效提高用户在 AR 体验中的沉浸感。定位越高，需要的算力越大。

5. 对象识别及模型匹配

在 Cloud AR 系统中，通常涉及到对真实物体的识别，并对识别的物体与虚拟模型相互匹配，最终实现叠加或是虚实对照的效果。端侧通过采集视频图像或点云数据等信息，实时的传输至边缘云侧，边缘云利用这些采集的图像信息在已训练的样本库中匹配，其匹配度达到某个阈值时，将该物体的识别结果输出。在模型匹配环节，可利用识别结果选定出对应的虚拟模型与之匹配，再将叠加的结果下发给端侧呈现给用户。



图 29. 汽车识别及模型匹配

6. 高精度模型实时渲染

Cloud AR 中，虚拟对象及时准确渲染和呈现是影响用户体验的关键因素。为了满足 Cloud AR 用户对沉浸感的要求，需要 Cloud AR 系统能提供对高精度实时渲染的支持。

实时渲染是一般应用于交互场景，需要根据用户的行为迅速的渲染出相应的画面。总的来讲，要求渲染帧率不小于 60fps，当前大部分 VR/AR 设备要求的渲染帧率为 75~90fps，有的甚至达到 120fps 以上。以 90fps 的渲染帧率为例，要求每帧渲染时间不大于 11.1ms。Cloud AR 属于强交互应用场景，对实时性渲染的要求很高。

1) 实时渲染

在 Cloud AR 系统中，因为网络和传输时延(含编解码时延)的存在，通常的方案很难满足 AR 对时延的要求。因此在 Cloud AR 方案中，在网

络和传输时延存在的情况下，如何解决因时延增加而导致用户体验下降的问题，是渲染环节的一个关键技术。采用云端和终端分离渲染的技术可以很好解决网络和传输时延导致体验下降的问题。

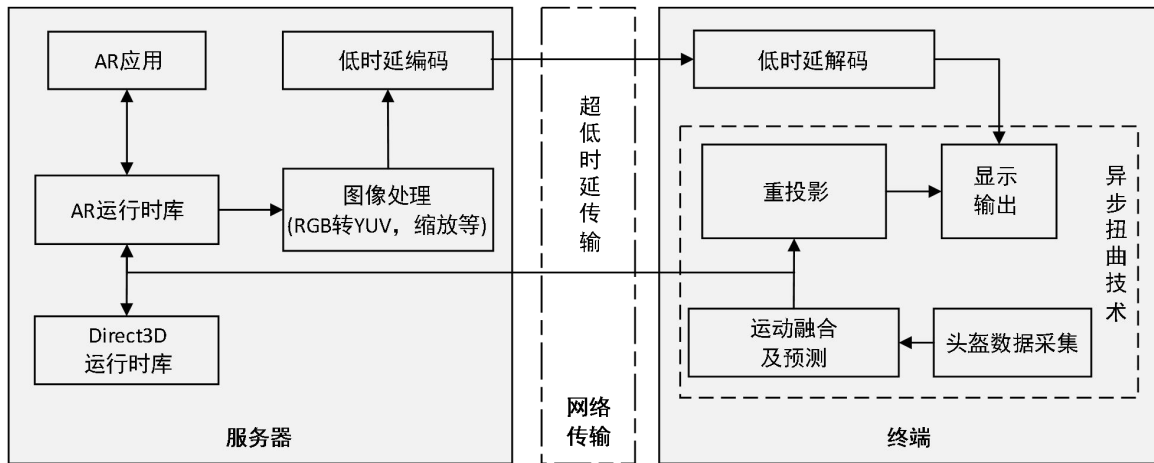


图 30. Cloud AR 功能框图

1). 云端和终端分离渲染：

对于 AR/VR 应用，为了解决动晕症问题，一般需要 MTP 时间控制在 20 毫秒以内，而对于 Cloud AR 来说这个挑战更为艰巨。云端和终端分离渲染方案是将渲染过程分为云端和终端两部分，云端应用根据上传的用户姿态渲染出游戏画面，游戏画面在下发到终端后，终端对该画面根据当时的姿态进行重投影。通过这种方式，最大限度的保证了最终展现给用户的画面与用户姿态接近，减少了 MTP 时延，提高用户的 AR 体验。重投影可分为旋转重投影与位置重投影两种技术。旋转重投影主要解决的是头部的旋转变换，而位置重投影则侧重于解决头部的位置变换。重投影操作

属于对图像的外插变换，利用已有的视频图像和姿态信息，通过对图像的变换操作产生与用户当前姿态一致的新图像。

在 Cloud AR 中运用该技术能够看到更显著的作用，使用该技术能够让 AR 终端直接利用已有的信息产生新的画面，不必经过云端而直接得到下一帧的图像，省去了此帧图像从服务端到终端的传输过程，在实际场景中该技术能够使 MTP 时延达到与非云端渲染相同水平。从而保证 Cloud AR 应用的用户体验。

在重投影技术的应用中，也有一些问题亟须采用新的技术解决。其中最主要的是画面扭曲失真现象和黑边问题。由于用户在使用 AR 过程中位置可能随时发生移动，当用户转移到新的角度观察虚拟对象，或绕过障碍物重新观察原本被遮挡的虚拟对象时，位置重投影则无法通过上一帧画面生成新的观察角度或遮挡部分的图像，此时只能使用虚拟对象周边的画面对缺失部分进行简单修复，用户将会看到错误画面。不过该问题如果在云化时延低、移动速度较慢时，画面扭曲失真的现象会很难察觉。因此，为了避免该问题的发生，建议将云化时延控制在 30~50 毫秒之间。

2) 高精度渲染

在 Cloud AR 中，为了带给用户沉浸式体验，也要求渲染的画面非常接近真实世界，这就对渲染的精度提出了要求。如图 2 所示，为了达到高精度高逼真的渲染效果，一方面与普通的 3D 应用类似，也要求 AR 应用能提供更精确的 3D 模型和纹理等，同时也要求显卡提供更高阶的渲染功能，比如光线跟踪等。另一方面，AR 应用对于渲染有自己独特的要

求，AR 应用将虚像与实景结合，并一起呈现给用户，要让用户保持好的体验效果，要求虚像与实际场景能够紧密结合并能够精确的融入到实景中。因此在 Cloud AR 系统中，需要提供一些重要的能力，以便虚像能基于实际场景来进行渲染。主要包括以下几个关键功能：空间锚点、光照信息、环境标识、环境 3D Mesh、语言识别和平面检测。

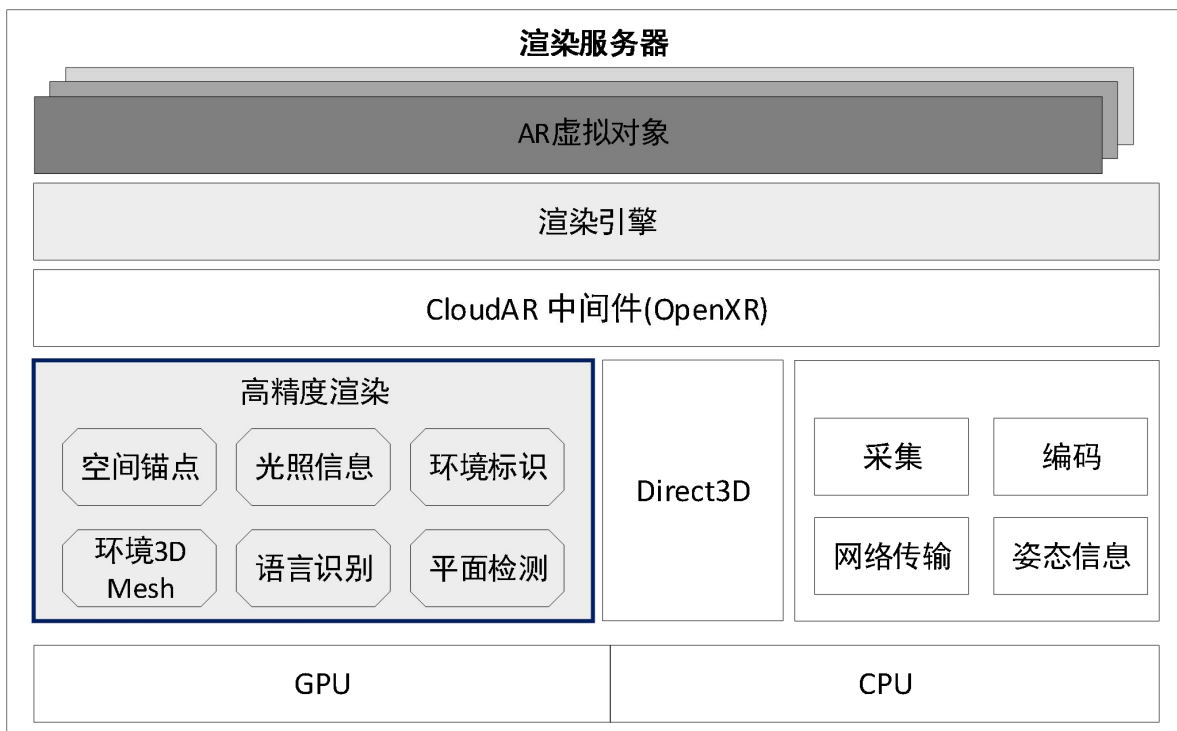


图 31. 高精度实时渲染结构图

1). 空间锚点:

空间锚点是标记虚拟对象位置的关键点，它能够根据环境进行实时调整。当终端学习了周围环境后，空间锚点能够调整自身的相对位置，确保锚点能够准确地停留在真实环境中对应的位置。通过将锚点与用户放置到

环境中的虚拟对象绑定,即使用户位置改变数米也能够保证虚拟对象位置保持稳定。

2). 光照信息:

AR 设备对于空间的定位依赖于光照情况,当光线过暗或过亮时可能导致环境检测出现误差。在室内环境中,灯泡交流电产生的脉冲也会导致某些帧过亮或过暗,当终端尝试对脉冲进行补偿时甚至会造成曝光过度。因此对于光照信息进行合理的处理能够使虚拟对象在真实环境中位置的定位更加精确。另外光照信息也将影响虚拟对象的渲染结果,包括光源方向对于虚拟对象阴影的影响,光照强度对于虚拟对象亮度的影响等。根据真实环境中的光照信息对虚拟对象进行渲染能够使其更好的与真实世界相融合。

3). 环境标识:

用户在使用 AR 时视野中会同时出现虚拟与真实环境,环境标识技术能够让虚拟对象以更和谐的方式融入到现实空间中。在实际应用中,环境场景一般较为复杂,场景内包括多种材质的物品,使用环境标识技术能够有效识别场景的材质与纹理,根据物体的透明度、光反射强度等特征,在物体对虚拟对象产生遮挡或接触时得到光线反馈。如虚拟物体被玻璃杯遮挡时会出现光线折射效果,而在镜子前则会出现反射,但在木质桌面上则只会出现阴影。

4). 环境3D Mesh:

空间曲面能够记录真实世界的表面情况，以三角形网络的形式附加到真实世界的空间表面，该网络称为环境 3D Mesh。环境 3D Mesh 能够确保虚拟对象能够准确放置到现实环境平面上。另外，现实环境中物体对于虚拟对象的遮挡能够使用户更真实的感受到虚拟与现实的结合。

5). 语言识别:

相较于 AR 终端常用的手势操作，通过语音命令来操作虚拟对象有显著的优点。语言操作能够最大限度的避免复杂的操作步骤，直观且易于用户入门操作；直接使用语言来操作让用户能够在繁忙时同时处理多项任务，更快捷的达到操作目的。因此在 AR 中使用语言识别技术可以有效提升用户的体验和操作效率。

6). 平面检测:

平面检测能够在真实场景中检测到水平平面和垂直平面的位置、边缘以及大小。平面检测能够使用户放置虚拟对象时更合理，正确的平面检测结果能够使虚拟对象紧贴垂直平面或水平平面，避免对象被放置到空中或穿到环境物体内部。

利用这些信息，应用可以渲染出更加精确的虚像。

五、总结及发展展望

本白皮书描述了云化 AR 的关键场景并提出了云化 AR 的关键技术。5G 网络的商用以及 AR 技术的发展，为云化 AR 应用带来了更多的可能性。

云化 AR 的计算上云、渲染上云为业务降低了业务时延及网络时延，给用户带来了良好的体验；同时云化 AR 平台的构建，使 AR 内容制作更便捷，提供更高质量的 AR 画面，为 AR 内容提供商和应用开发者提供了更多的选择，推动了 AR 产业生态的发展。

目前国家在积极推动数字化建设以及 AR 应用的普及，各地方政府及企业积极推动 AR 技术与特色产业相结合。在产业化推进的过程中，AR 云化发展应以需求为导向，通过在垂直领域的创新发展，促进云 AR 的行业应用普及。

5G 网络与云 AR 平台结合仍处于技术探索阶段，需要在更多的产业实践中探索并积累经验。现就云化 AR 技术与应用发展提出如下的建议：

加速和完善云化 AR 标准系列的制定和应用推广。建立云化 AR 业务的平台侧、网络侧和终端侧设备技术指标体系和测试方法，推进行业达成共识，指导和规范产业生态的良性发展。保障 AR 业务平台、网络设备和终端之间的互联互通、信息安全和设备兼容等。同时，积极开展云化 AR 系列标准在各垂直行业的应用推广，通过可依据、可执行的云化 AR 全要素技术指导和规范，为各行业应用创新提供有力保障，推进 AR 与各行业深度融合发展；

开展规模化应用试点，探索具备落地潜力的解决方案。紧抓 5G 时代机遇窗口期，以云化架构为引领，基于业务需求优化网络和算力资源调度，保障端到端业务质量，深化 AR 与各垂直行业应用的探索融合，实现产业级、网联式、规模性、差异化的应用普及。

AR 技术创新持续取得突破，不断推动产品升级，全球 AR 产业进入快速增长新阶段。随着 AR 终端舒适度的提升和成本下降，平台的软硬件能力提升和算力成本降低，端到端业务的用户体验的提升，必将使能各垂直行业，为人们的工作和生活带来更多信息化发展的便利。

参考文献

- [1] 《增强现实（AR）C 端应用白皮书》2019，太平洋未来科技等联合发布
- [2] 《关于 AR 业务用户体验测试和调研结果》2018，ETSI
- [3] Xiuquan Qiao, Pei Ren, Schahram Dustdar, Ling Liu, Huadong Ma, Jun-Liang Chen. Web AR: A Promising Future for Mobile Augmented Reality - - State of the Art, Challenges, and Insights. Proceedings of the IEEE, vol. 107, no. 4, April 2019, pp. 651-666.

附录 1 AR 计算端云分布建议

表 1 AR 计算端-云（边）按需分布说明

AR 计算端云按需分布			
AR 关键技术点	端侧	云（边）侧	备注
SLAM	√	√√	位姿、锚点
表情识别	√√	√√	
动作识别	√√	√√	
表情驱动	√	√√	
动作驱动	√	√√	端侧处理帧率受限
图像处理	√	√√	对象识别跟踪(端侧处理帧率受限)
渲染	√	√√	
高阶表情驱动	√	√√	更丰富的表情基+更高帧率
高阶动作驱动	×	√√√	实时滤波、运动美型
高阶图像处理	×	√√√	对象识别跟踪(高帧率)、图像分割
高阶渲染	×	√√√	精细度、SSS、光照等
虚实图像叠加	√√	√√	
操作交互 UI	√√√	×	
操作交互云服务	×	√√√	多人：标注、模型操作

图例	含义
√√√	必选
√√	推荐
√	可选
×	无法构建

附录 2 端侧无法承载的高阶 AR 计算说明

表 2 高阶动作驱动

高阶动作驱动		
对比内容	端侧	云（边）侧
实时滤波	×	√
运动美型	×	√
唇声 AI 对照	×	√
帧率	15-18fps	60fps
结论	仅适合在云端计算	

图例	含义
√	可构建
×	无法构建/尚未构建

表 3 高阶图像处理（以图像分割为例）

高阶图像处理		
对比内容	端侧	云（边）侧
示例		
描述	手部下方露出（普通背景更换为长城背景）	图像分割效果（普通背景更换为天空背景）
图像比较	端侧计算在局部会有大面积背景露出	云端计算可较好的实现分割效果
帧率	30fps	60fps
结论	仅适合在云端计算	

表 4 高阶渲染

高阶图像处理		
对比内容	端侧	云（边）侧
示例		
描述	全息通话端侧渲染场景	全息通话云（边）缘渲染场景
局部渲染	较差	较优
光照效果	较差	较优
帧率	12-15fps	60fps
结论	仅适合在云边计算	

附录 3 缩略语

表 5 缩略语

简称	全称	中文含义
AR	Augmented Reality	增强现实
Cloud AR	Cloud Virtual Reality	云化增强现实
VR	Virtual Reality	虚拟现实
XR	Extended Reality	扩展现实
NPU	Neural-network Processing Unit	网络处理器
GPU	Graphics Processing Unit	图形处理器
SLAM	Simultaneous Localization And Mapping	即时定位与地图构建
3GPP	3rd Generation Partnership Project	第三代合作组织
GSMA	Global System for Mobile communications Association	全球移动通信系统协会
FOV	Field of View	视场
2C	To Customer	面向个人
2B	To Business	面向企业
K-12	Kindergarten through twelfth grade	学前教育至高中教育
VPS	Visual Positioning Service	视觉定位服务
HMD	Head Mounted Display	头戴式显示
MEC	Mobile Edge Computing	移动边缘计算
LCoS	Liquid Crystal on Silicon	硅基液晶显示
HTPS-TFT	High Temperature Poly-Silicon Thin Film	高温多晶硅薄膜晶体管
LCD	Transistor Liquid Crystal Display	液晶显示

简称	全称	中文含义
OLED	Organic Light-Emitting Diode	有机发光二极管显示
AMOLED	Active-matrix Organic Light-emitting Diode	有源矩阵有机发光二极管显示
DLP	Digital Light Processing	数字光处理
Micro LED	Micro Light-Emitting Diode	微型发光二极管
MEMS(LBS)	Micro Electromechanical System (laser beam scanning)	微型激光束扫描投影显示
ARQ	Automatic Repeat-reQuest	自动重传请求
FEC	Forward Error Correction	前向纠错
MTP	Motion-to-Photons	运动响应时延

附录 4 标准化情况

表 6 标准化情况

序号	标准号	标准名称	状态
1.	GB/T 38247-2019	信息技术 增强现实 术语	发布
2.	GB/T 36341.1-2018	信息技术 形状建模信息表示 第 1 部分：体系结构和组件	发布
3.	GB/T 36341.2-2018	信息技术 形状建模信息表示 第 2 部分：特征约束	发布
4.	GB/T 36341.3-2018	信息技术 形状建模信息表示 第 3 部分：流式传输	发布
5.	GB/T 36341.4-2018	信息技术 形状建模信息表示 第 4 部分：存储格式	发布
6.	GB/T 28170.1-2011	信息技术 计算机图形和图像处理 可扩展三维组件（X3D） 第 1 部分：体系结构和基础组件	发布
7.	20171072-T-469	信息技术 手势交互系统 第 1 部分：通用技术要求	报批
8.	20171075-T-469	信息技术 手势交互系统 第 2 部分：系统外部接口	报批
9.	20132345-T-469	信息技术 计算机图形和图像处理 可扩展三维图形（X3D） 第 2 部分：场景访问接口	报批
10.	YD/T 3078-2016	移动增强现实业务能力总体技术要求	发布

中国通信标准化协会

地址：北京市海淀区花园北路 52 号

邮政编码：100191

联系电话：010-62302734

传真：010-62301849

网址：www.ccsa.org.cn

