

AR 边缘云白皮书技术 概览





目录

序言

① AR 云的定义和重要作用	2
② AR 技术核心技术框架	9
③ AR 云重要核心技术和近期发展	13
④ AR 云 @5G 中的核心技术路径	18
⑤ AR 边缘云核心技术发展预测	23
后记：AR 边缘云 @5G 趋势把握	27

序言

“

2019 年，AR 及其云服务能力和融合构架进入了百舸争流的快速发展阶段。各 ICT，人工智能以及互联网的领军企业纷纷发布了令人瞩目的新技术并拓展其基于云的服务模式，如结合 AR 技术的 3D 高精度地图服务——Cybersavers 和基于视觉定位系统 Live View 的 Maps AR 导航功能。线下物理世界和线上数字世界之间的融合和相互激励已逐步呈现。

同时 5G 时代的到来和商用化部署的加速，加速了数字信息流转以及物联规模的发展，并以更加安全和灵活的方式保护了数字资产和隐私信息，为物理和数字世界的融合与启动以视觉自然交互为基础的人物交互机制 (HTI: Human Things Interaction) 开辟了高速公路。

机遇和挑战并存，如何积极和理性的看待当前相关技术的发展红利和机遇，同时厘清需要快速解决的核心技术问题，是时代赋予数字视频和机器视觉相关领域专家和从业者的重任，也是真正开启物理和数字世界融合之门的前奏。

”

01 AR云的 定义和重要作用



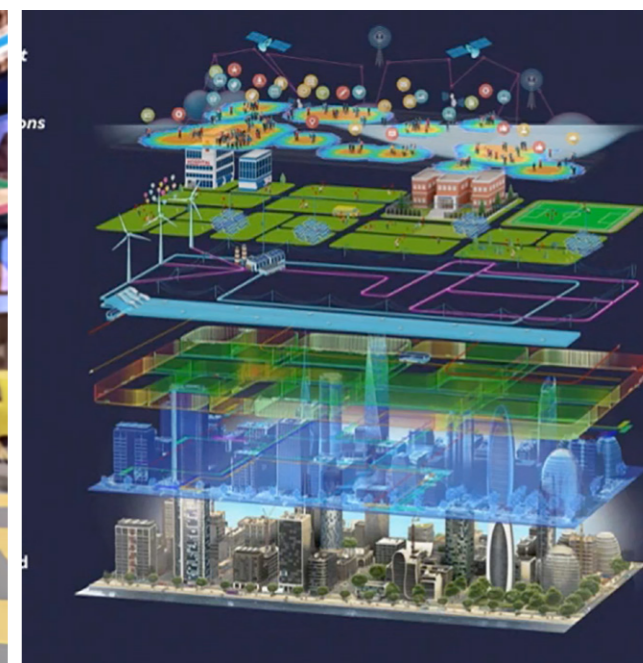
“

增强现实（AR）技术，融合了视觉感知为主的空间语义感知和以混合内容渲染为主的可视化内容生成技术，并通过新兴近眼显示为核心的可穿戴设备系统工程推动人和物理世界的自然交互，实现了“所见即所知”以及抛弃键盘、鼠标和显示器的信息传递高速公路。

作为一种革命性人机交互方式（HTI）的撬动支点，AR 集合了大量现代技术进步和成果，包括以机器视觉和SLAM 为核心的人工智能技术，以及以云部署和大数据融合为基础的信息后台技术，和以光波导为当前主要应用模式的近眼显示技术，同时最近大量商业化部署的 5G 技术也成为其连接技术的主要支柱。

结合 AR 这个系统技术集合，行业的数字知识资产的融合和基于视觉触发的高速信息交互，为智能制造、数字孪生以及新兴的智慧城市服务提供了更加广阔的想象空间。

”



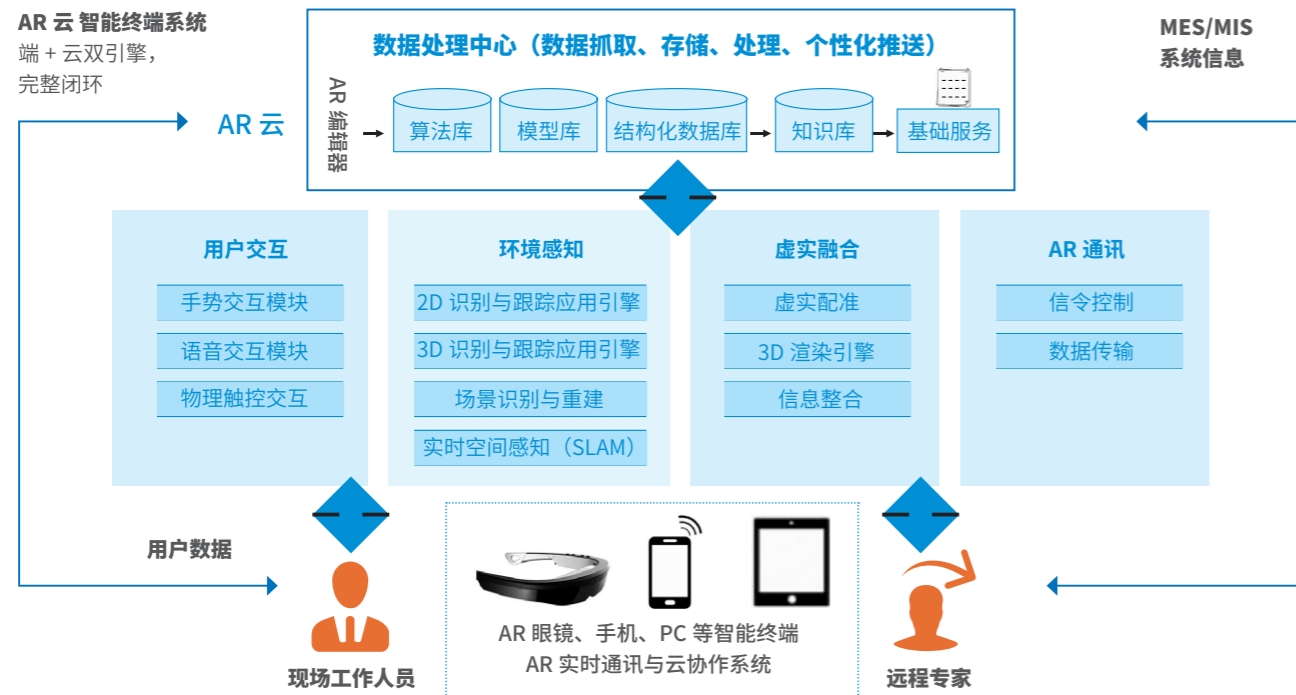


图 1-1 AR 云及其系统功能

智能制造中的AR云实践

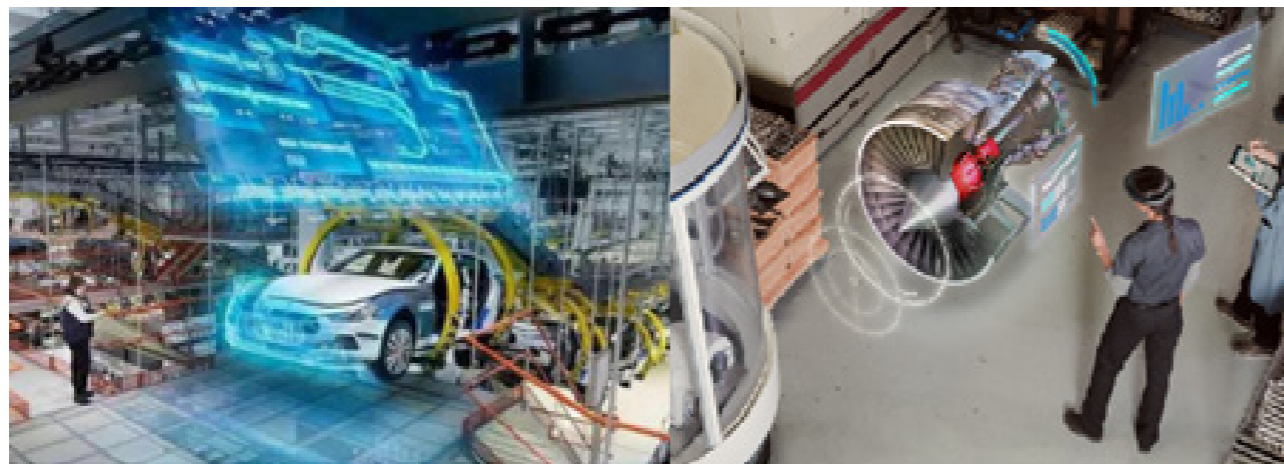


图 1-2 智能制造 AR 云应用场景



现阶段，在智能制造趋势下工业企业纷纷开始了探索新型信息化提升的路径，以实现增效降本，增强企业竞争力的目的。其中工业制造过程也是一个应用场景复杂，且垂直领域特点显著的实施场景。



首先：制造场景存在

- 1 生产场景环境复杂，如有大量移动的设备、人员和车辆；
- 2 设备多样，无论是外形、内部专业结构、或电气和电路连接拓扑；
- 3 设备运行环境差异大，可能存在高温、高危、高电磁的运行工况；
- 4 涉及的协作工段和流程多，对于精细化和柔性管理的要求高。

其次：作为制造环节中最重要的因素 -- 人

无论从需要掌握的专业技能还是需要关注的综合领域信息都呈现出愈来愈快速发展的趋势，同时高速运行的制造过程也要求生产人员更加专注的关注快速变化生产和实施过程。

在继互联网技术赋能的数字化信息制造过程演进后，核心要素人和设施设备的“人”和“物”的交互（革命性人机交互方式（HTI））和信息流加速就成为对智能制造向下一步生产力提升演进的重要途径。

当前利用 AR 技术与一体机设备这种可远程、实时性、主动式、多视角融合、多场景交互、解放双手的信息呈现形式，辅助产业工人的生产工作和培训，把互联网和信息技术延伸到一线员工的工作当中已被很多工业巨头积极实践，如通用和 GE 等，取得了大幅提升工作人员信息获取与转化率从而提高劳动效率的积极结果，即实现了——“增效降本，增强企业竞争力”。



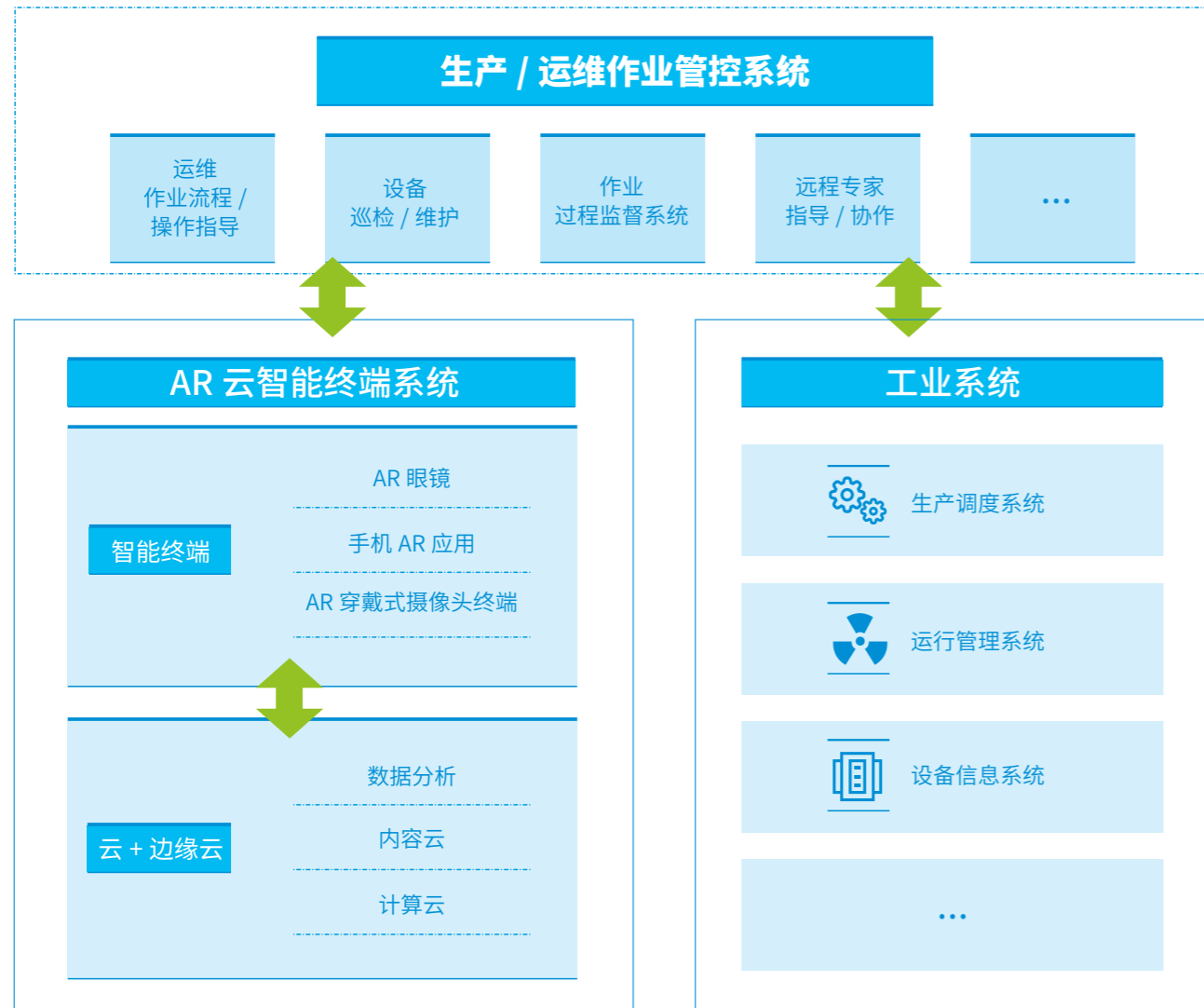


图 1-3 AR 云辅助智能制造系统架构图

目前，比较完整的 AR 云辅助智能制造系统如图 1-3 所示，包含了利用 AR 增强现实技术、AI 人工智能技术、IOT 物联网、云计算、5G 通讯技术以及 5G+MEC 边缘云技术的融合，以云 AR 智能终端系统作为连接现场作业人员、设备、运维中心和非现场人员的智能化工具，实现跨地域数据资源共享和生产过程监督管控以及非现场专家的远程指导。

在以上场景中，辅助 AR 设施需要最大程度具备灵活性和轻便性，以便生产现场的工作高效开展。云 AR 智能运维中心部署在云端，提供 AR 应用服务和交互以及数据知识库沉淀与智能化分析服务，通过在边缘云计算中心执行 AI/AR 算法引擎，提供算法模型的微服务管理、在线部署服务、模型运行监控和资源调度管理，从而提高生产效率，实现业务闭环。

AR云的核心功能框架

云 AR 智能终端系统作为连接现场作业人员与设备以及运维中心和非现场人员的智能化工具，是人物交互方式（HTI）和信息流加速的核心路径。同时它要求最大程度具备灵活性和轻便性，并具备足够的续航时间。这就为整个 AR 云构架提出了一个核心演进的方向，如图 1-4 所示。

其中，在 5G 通讯等核心技术的发展支撑下，AR 智能终端由应用内容存储、感知、数字化内容渲染、用户交互和显示综合能力本地化部署，逐渐向仅包含显示和用户交互的轻量化

能力部署转化，以云服务模式强化信息流融合和加速，以及深化数据服务的层次，同时提升灵活性和轻便性。

同时，在环境复杂、设备多样的场景下，前端完成视频内容内关注设备对象的空间识别以及场景识别以构建人物交互（HTI）通道，需要非常复杂的机器视觉算法的支撑，有效的利用云端算力也方便了人物交互的广泛利用和应用场景的多样性。

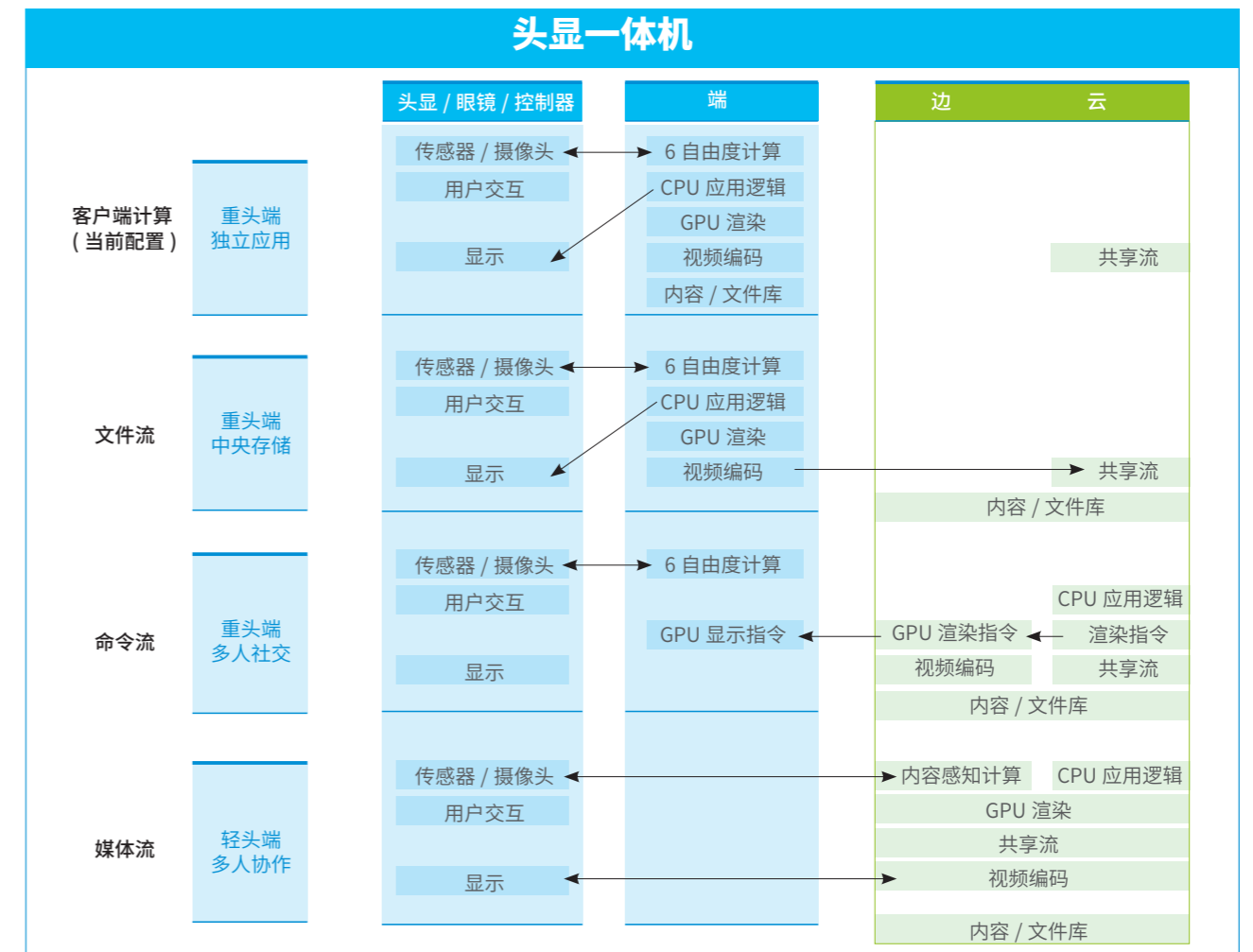


图 1-4 AR 云构架演进框图

AR云的发展趋势-AR边缘云

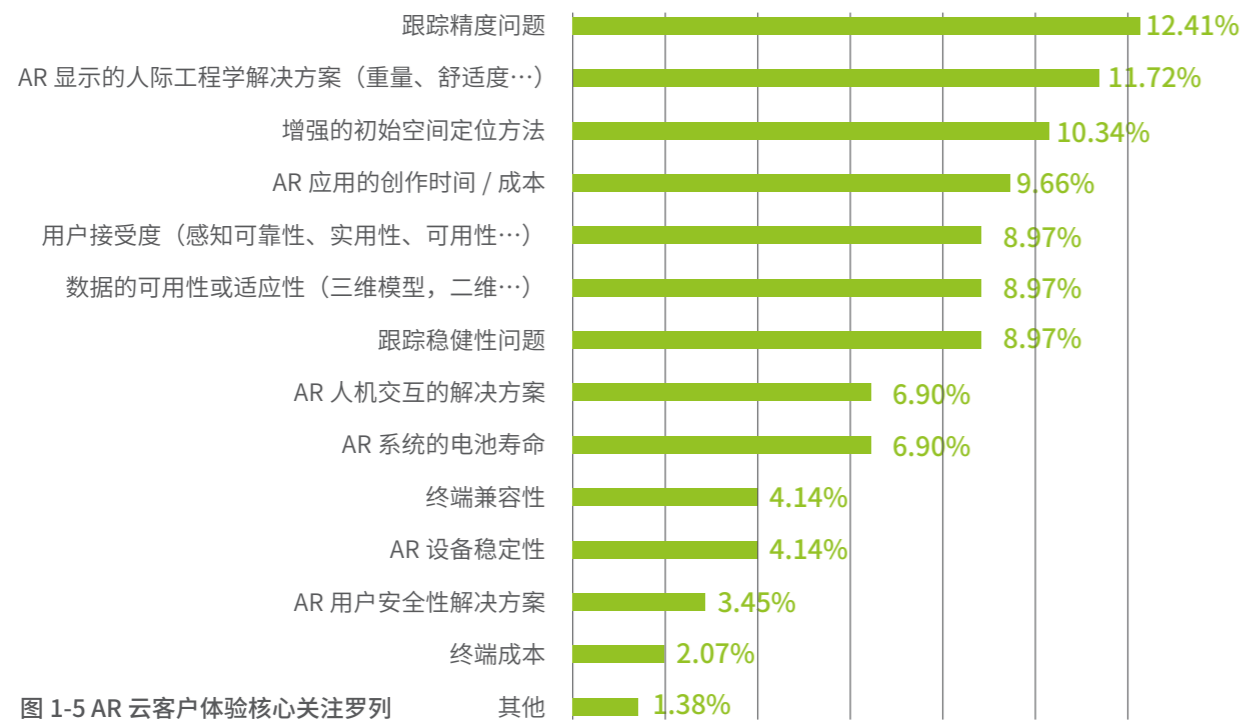


图 1-5 AR 云客户体验核心关注罗列

AR 及其云服务应用领域广阔，在其推广和普及的过程中用户体验关注点成为推动产业进步和技术变革的核心动力和指标。参照 PTC 的调研报告《State of Industrial Innovation》（工业创新发展状况）系列报告 68% 的工业企业正在加速增强现实的采用，同时结合 GSMA 相关工业场景的调研报告，如图 1-5 所示，影响客户的项目部署和规模推广的核心问题主要源自三个领域：

- 1 识别和跟踪定位（三维注册）这一核心环节和技术基础的整体体验；
- 2 人机交互的部件组成和整体，如交互器件 AR 显示终端的体验感受和整体交互方案体验感受；
- 3 和使用及投资成本相关的主要问题，如终端成本。

结合 1.2 章节的核心框架发展趋势，云端分析功能和数字信息融合功能的边缘化部署，成为提高领域 1) 和领域 3) 的重要手段，同时视频内容关注设备对象的空间识别以及场景识别的适用范围的扩展（即 AR 技术本身的提升）为平摊设备的使用成本和提升收益，以及提升识别和跟踪定位的体验就显得异常重要。而脑机接口的引入，为加速语义信息流转和更自然的人机交互提供了想象空间。

02 AR 技术核心技术框架



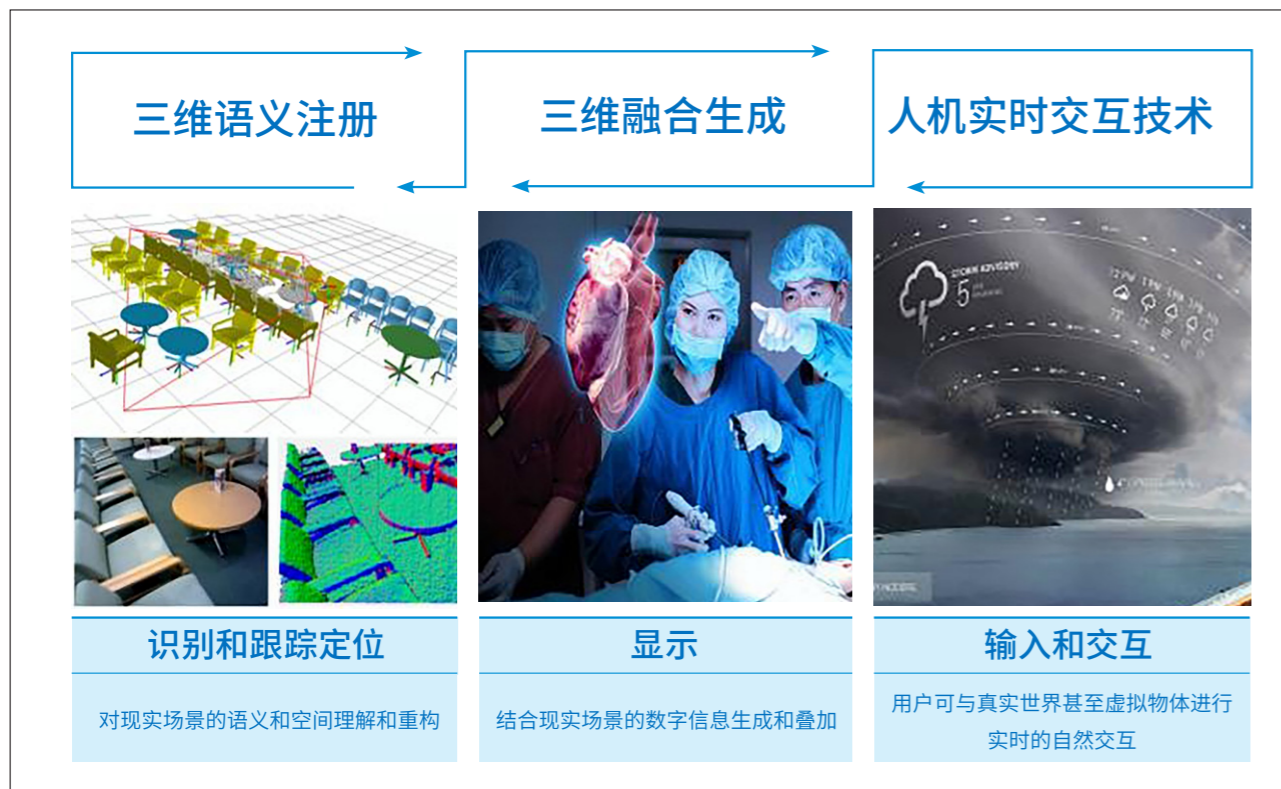


图 2-1 AR 核心技术系统图

作为人物交互 (HTI) 通道的核心技术体系, AR 其实就是需要解决识别并跟踪定位需要交互和已经关注的现实场景目标, 并根据手势或语音等交互方式完成意图的传递, 以实现最终结合现实场景目标信息叠加的显示。

随着应用场景的进一步拓展, 以及应用场合的不断丰富, 包括多人的长效体验, AR 的核心技术体系, 特别是考虑到云端处理, 也逐渐丰满和清晰。在垂直应用领域可以清晰可见的 AR 核心技术体系包括:

三维语义注册: 对现实场景的语义和空间理解和重构;

三维融合生成: 完成结合现实场景的数字信息生成和叠加;

人机实时交互技术: 用户可与真实世界甚至虚拟物体进行实时的自然交互。

三维语义注册: 对现实场景的语义和空间理解和重构



三维语义注册, 包含了对于实景的三维结构恢复和重建, 以及虚拟环境的实时三维注册技术两个部分。



- ① 其中三维结构的恢复和重建常常用来建立实景与观察者视野 (相机) 的坐标转换关系, 为虚拟物体与现实空间的三维注册奠定基础。其包含了多视图几何原理的遵循, 以及图像的特征点的提取和匹配、运动恢复结构、三维几何重建、表面纹理重建等核心技术。近年日益普及的深度传感器对于高效和准确的完成三维重建存在积极作用, 但传感器本身精度的差异化, 以及当前技术所能达到的精度范围, 还需要探索有效方法以实现最终场景精确的稠密重建。
- ② 其中虚拟环境的三维注册技术常常用来实现让虚实环境共享同一空间坐标系, 以实现将虚拟景物嵌入其中的目的。其包含了同时定位与地图构建 (SLAM), 平面标志物的检测和识别与跟踪, 以及三维实物的检测、识别与跟踪等核心技术。目前三维实物的跟踪已逐步进入实用化的阶段, 但在工业界的大规模应用尚有距离。

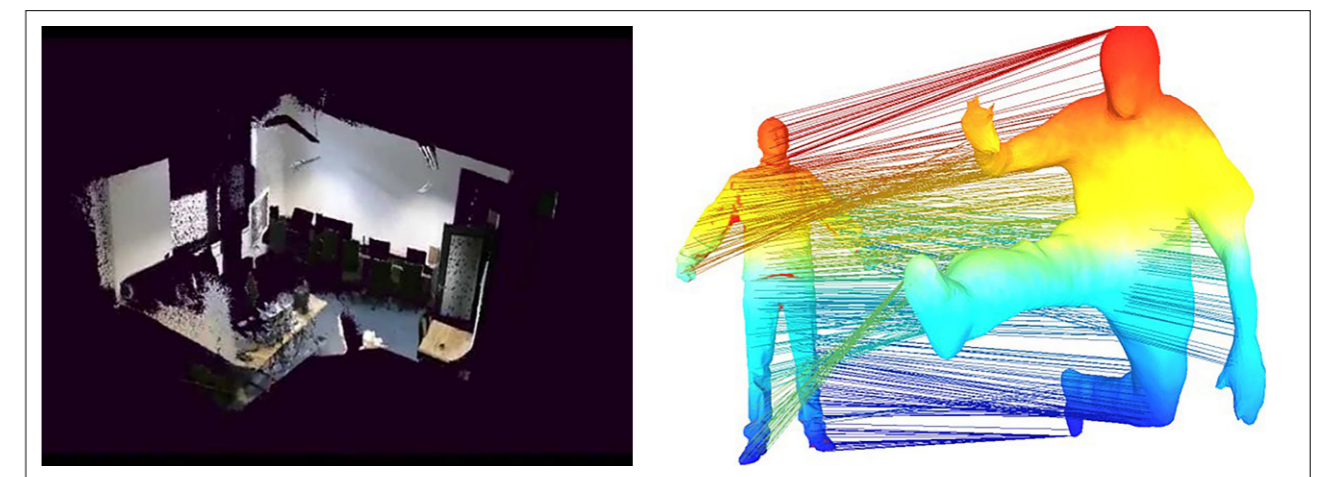


图 2-2 AR—三维语义注册结果示意图

三维融合生成：结合现实场景的数字信息生成和叠

“

三维融合生成，包含了虚拟物体的绘制、现实环境的光照估计、虚实视觉融合和信息与现实环境的融合显示技术；同时空间增强实现技术也与其密切相关。

其中空间增强实现技术涉及空间增强实现的具体方式，包含光学穿透方式、以及光场方式和多投影呈现方式，同时也包括投影画面的几何校正和颜色校正技术等核心技术。

随着 AR 服务从功能倾向逐步向体验倾向的转移，三维融合生成的技术重要性就显得愈加重要，通过准确实现融合光源和周围环境的虚拟数字化对象的形状、位姿、材质和纹理，真实感将跃然眼前。

”

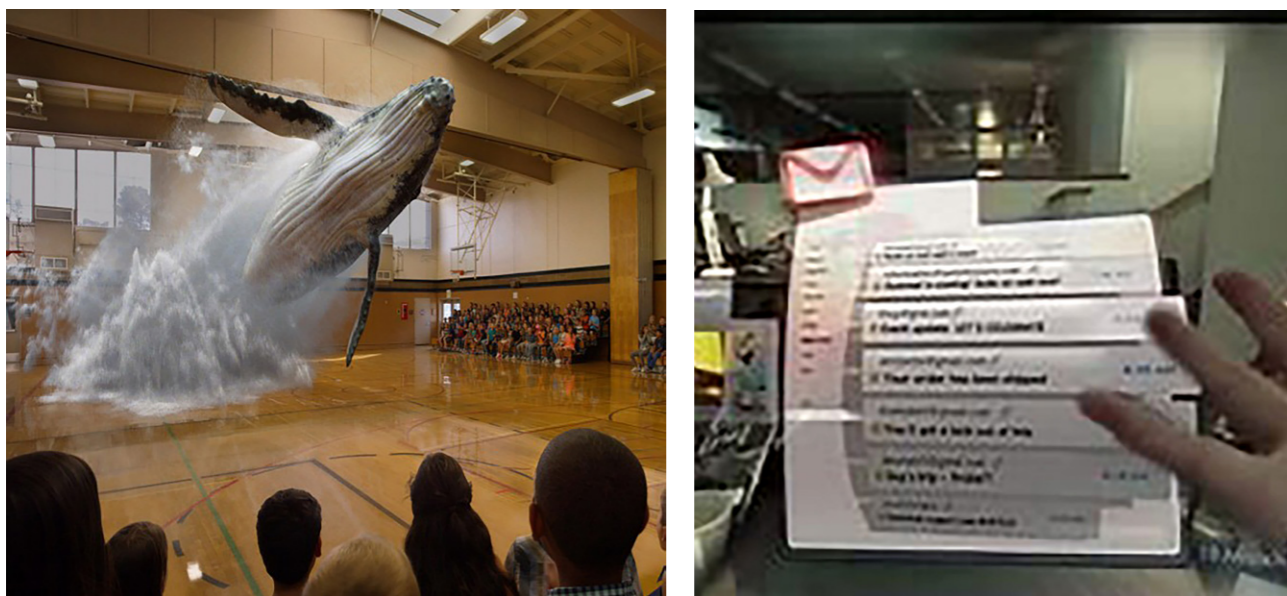


图 2-3 AR—三维融合生成结果示意图

03 AR云重要核心技术和近期发展



基于SLAM的空间及3D结构语



在 AR 的核心技术图谱上，3D 空间感知和 3D 对象的空间结构的感知一直占据了非常重要的地位和位置，传统基于激光的方法和基于光学的 SLAM 方法比较偏重于单纯空间结构的感知和分割，对于内容的语义和上下文场景的感知，还处于比较初级的阶段。由于缺少语义信息，它们不能直接用于构建更高级的人物交互（HTI）通道。

同时，深度学习已极大地提高了识别性能，但是这种识别大部分限于图像平面中的输出，或者在最佳情况下使用 3D 边界框完成一定意义上的语义空间结构表示，但这会使机器人或者准确数字信息叠加很难根据这些略显粗糙的输出进行操作。

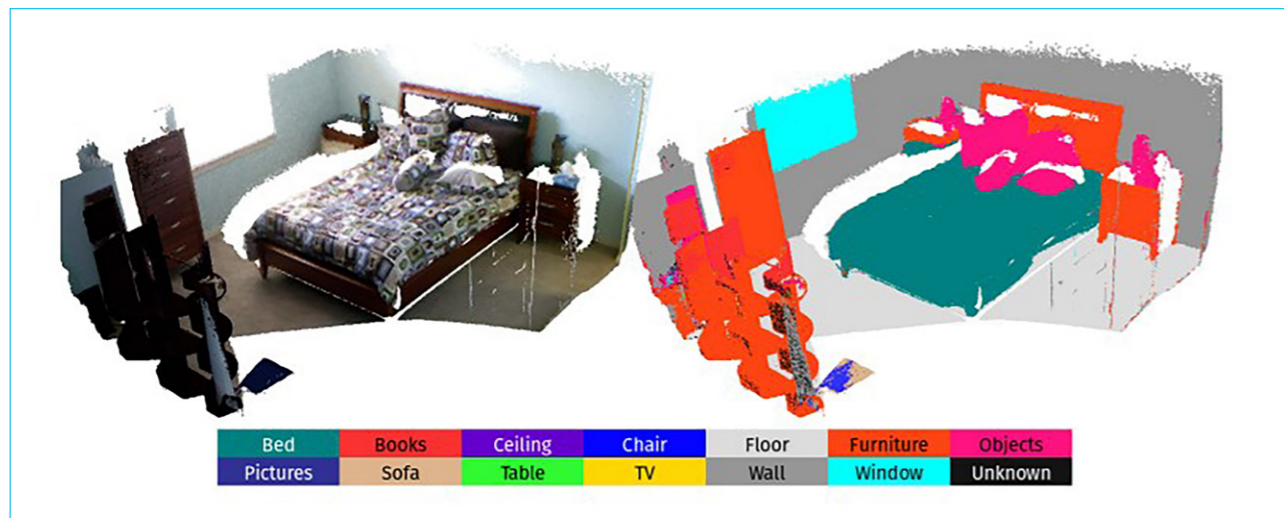


图 3-1 深度学习的 3D 空间语义感知结果图

将学到的知识和语义与 3D 重建相集成是解决这个问题的有希望的途径。例如，近年来提出的语义 3D 重建技术共同优化了场景的 3D 结构和语义，并且语义 SLAM 方法向所估计的 3D 结构添加了语义注释。

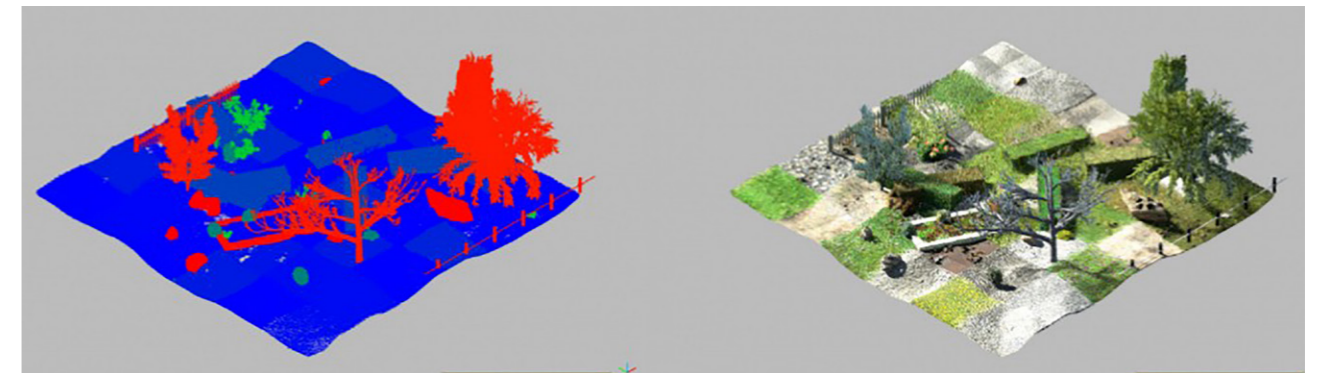


图 3-2 基于语义 3D 重建的 3D 结构语义感知

目前最新的语义 3D 重建和语义 SLAM 的挑战在于使用大型综合数据集所探寻的 3D 重建技术与识别和学习相结合的新方法，进而实现在 3D 重建技术中使用语义信息来改善密集匹配过程，并更有效的实现语义分割和 3D 技术之间的反馈环。

这也必将为构建更高级的人物交互（HTI）通道和泛化 AR 云服务场景开辟无比宽阔的发展之路。

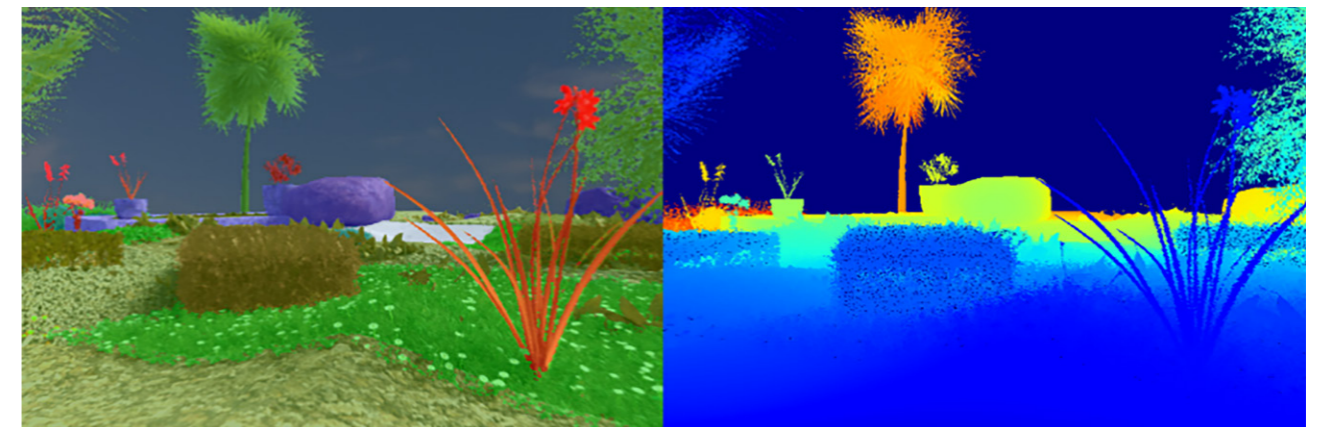
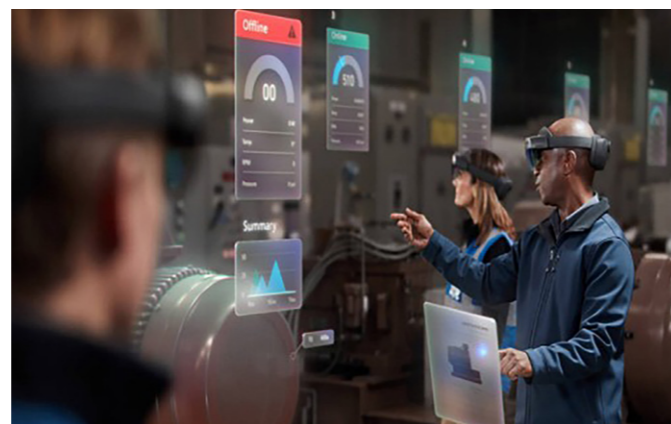


图 3-3 最新的语义 3D 重建和语义 SLAM 目标结果

基于空间地图的POI注册和同步



图 3-4 基于空间地图的 POI 注册和混合现实显示效果



“

和传统 AR 云服务相比，基于地理位置感知和兴趣点标注技术自 2018 年已经成为各个巨头推广和布局的重头，除了谷歌基于 GoogleMap 的 Live View，以及微软作为 Microsoft 扩展推出的云 AR 平台，称为 Spatial Anchor（空间锚点）。

基于空间地图的 POI 注册和同步技术的核心是地理位置强化的锚点技术（Anchor），——“空间锚点代表着系统应随时间推移跟踪的一个重点”。同时也是相关数据构建、同步、检索、管理等综合系统技术的突破，这也包含了相关 SLAM 语义感知以及基于神经网络的大规模特征比对技术的进步。

起初锚点只是让虚拟物体在 AR 场景中看起来待在原地不动，起源于依赖于在现实世界的记录中检测到的关键点及其描述符。随着强大的环境理解功能的演进，同时辅助 3D 高精度地图功能，空间计算功能，基于空间地图的 POI 注册和同步技术，就为数字孪生，以及实现真实世界和物理世界的无缝集成提供了可能。

基于空间位置的数字信息簇管理和搜索技术

“

如何基于将视觉感知转化为高精度 LBS（Location Based Service）的可视化信息的管理锚点和检索锚点是发展 AR 作为人物交互核心手段和发展相关业务的重要因素。这需要将相关数字信息归类和管理为 1: 在哪儿；2: 是谁的一部分；3: 能开展什么功能，并上传什么数据。

”

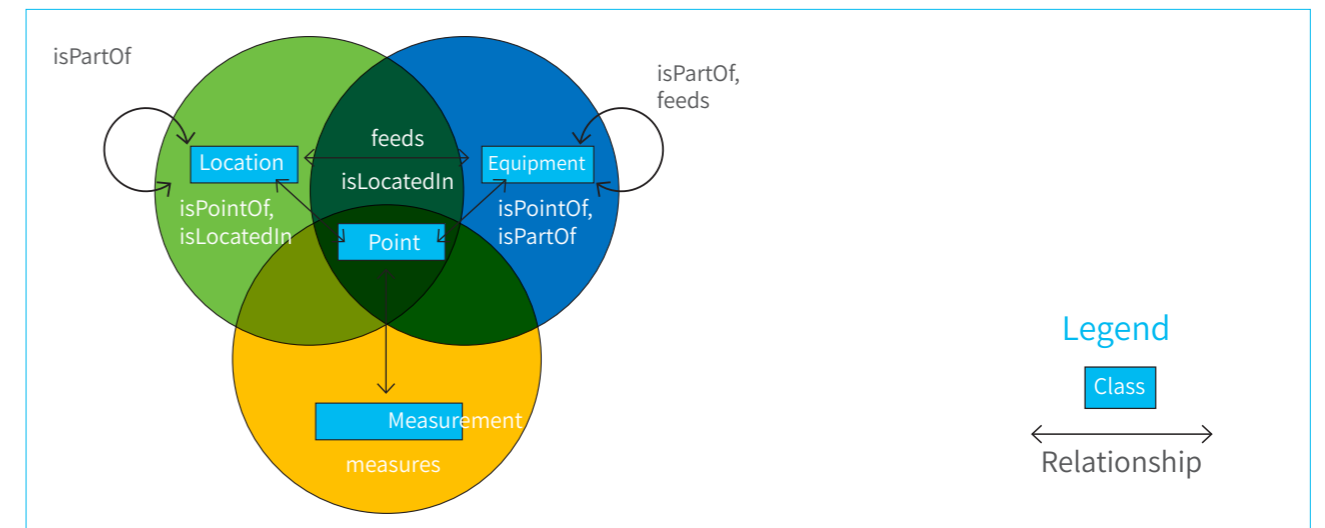


图 3-5 基于位置的关联数据模型

加州理工伯克利（Berkeley）分校电气和计算机学院 David E. Culler 教授在智能建筑的基于位置关联数据管理的尝试，“Brick”，就非常具有借鉴意义。

“Brick”源于资源描述框架（RDF: Resource Description Framework）。其通过以一组被称作三元组的主题预测对象表示知识。所有基于位置的关联系统信息在“Brick”中通过三元组中主体实体与另一些实体对象关系的抽象表示，即图中的定向边缘，予以表征。最终整个数据体系的管理体现为，由关系（定向边缘）相连的实体（节点）的知识图表，并通过 RDF 格式存储和使用，以及后续可通过使用 SPARQL 数学来遍历和查询这些知识。

基于以上知识体系，所有关联信息包含了物理世界目标的

- ① 分类标签；
- ② 地理位置；
- ③ 关联关系和关联目标；
- ④ 隶属和组成关系机器对应对象；
- ⑤ 提供什么检测数据或可开展何种控制动作。

04 AR云 @5G 中的核心技术路径

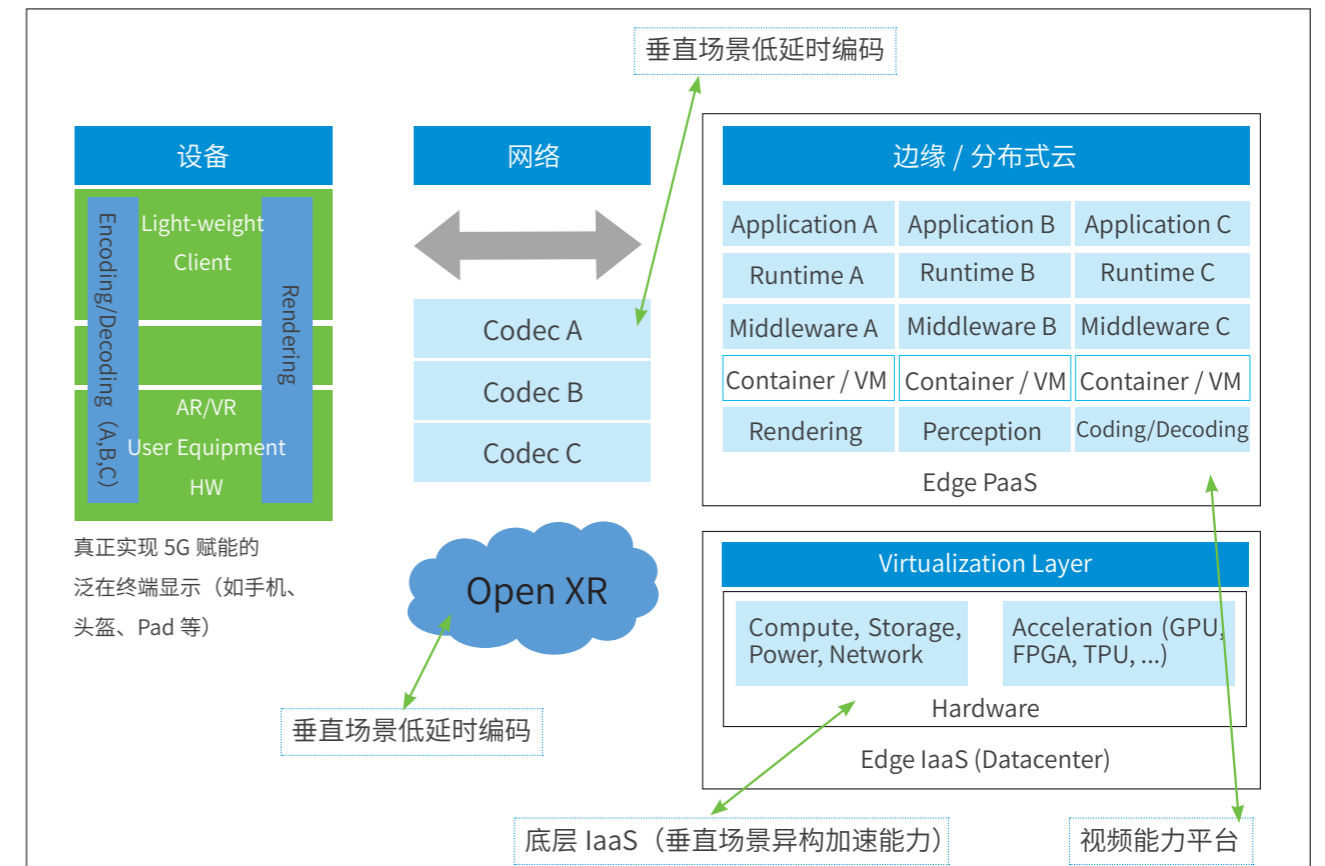


图 4-1 5G 云 AR 系统架构图

5G 作为通讯赋能行业的一项重要技术变革，对于 AR 能力和服务的云部署，不但带来了宽裕的上下行流量通道，同时超低的空口延时和安全灵活的数据传输帧组织方式，为基于核心接入的移动边缘计算框架提供了广阔的可能。

同时，5G 为核心和入口的庞大设备群落，也为 AR 服务的受众覆盖和效益辐射提供了不可比拟的广阔空间，也使得 5G+AR 推动的轻薄和便携 AR 服务成为苹果 (Apple)、谷歌 (Google) 乃至亚马逊 (Amazon) 等巨头着力投入和开拓的领域。

然而，机遇和挑战并存，移动边缘智能视频服务能力 (参照 GSMA Cloud AR/VR 白皮书) 并非是简单 5G 并整合已有技术的堆砌，其中切实实现 5G 为核心和入口的庞大设备群落的体验一致，以及云能力赋能终端提供更有吸引力客户体验 (如延时和清晰度等指标) 和经济模式就显得尤为重要。

结合行业的密切沟通和积极探索，当前三个技术领域的突破就显得意义非凡，如下篇章将对其分别进行详细阐述。

低延时编码



云能力赋能终端提供更有吸引力客户体验（如延时和清晰度等指标）在当前技术构架下，是依托 5G 空口低延时，以及高带宽完成待感知现实场景关键数据上传和显示结果下传，通过云端强大的服务能力完成更高体验质量的可视化融合展示效果，其中延时的体验因素就成为考察以上技术方案的核心和突破要素。

传统视频编码标准 H263~H265，以及即将投入使用的 H266 都是以压缩数据传输带宽为核心，对于图像质量以及传输延时都有一定的缺失。同时由于存在编码上传的流程，云处理带来的渲染融合和感知的功耗降低和本地编解码实现带来的功耗提升就成为影响前端设备续航能力的重要约束。

为此近期，联合图像专家组（JPEG），推出了一种新的开源低能量压缩解决方案，JPEG XS，以方便通过 5G 等网络以低延迟发送更高质量的图像。

JPEG XS 核心编码系统使用离散小波变换（DWT），然后采用 4 个连续系数组为基本单位实现实际系数幅度的熵编码。该标准定义可以提供场景和对象感知所需要的内容高清无损压缩，同时编码具有非常低的复杂度和内存资源消耗，并可以提供各种程度的并行度，方便 FPGA，ASIC，CPU 和 GPU 等各种平台上高效实现，并提供当前最低的编码延迟。

视听设备制造商压缩、图像处理和安全解决方案的创新技术提供商，IntoPix（比利时），即采用创新的视觉无损，轻量级和基于行的夹层压缩技术 TICO（SMPTE RDD35），实现了 4K60P 实时编码和解码 仅需 96 微秒的可用 JPEG XS 实现。



5G确定性网络



AR 服务需要数字信息和物理场景准确叠加，并能准确伴随关注点变化和场景变化。终端本地服务可以轻松保障这种延时和同步，而云能力赋能终端就面临着原有 IP 网络技术对于传输控制这方面的缺失（存在着大量的拥塞崩溃、数据分组时延抖动等问题）。

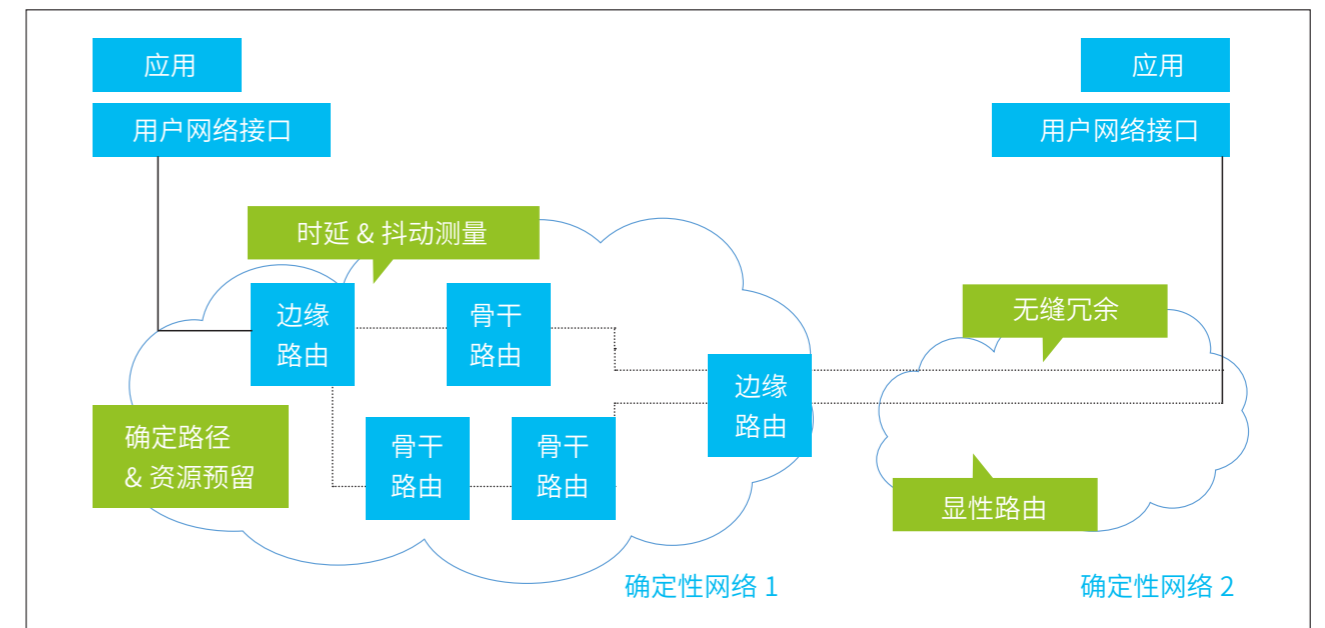


图 4-2 确定性网络架构

2019 年未来网络发展大会上，有关重量级专家提出了 5G 确定性网络（5G Deterministic Networking），以原生云、动态智能网络切片和超性能异构 MEC 为技术内核，打造一个有确定范围时延、丢包和时延抖动参数的确定性网络，保障极致用户体验。随即，5G 确定性网络产业联盟在刘韵洁院士的倡导下在 MWC19 成立。

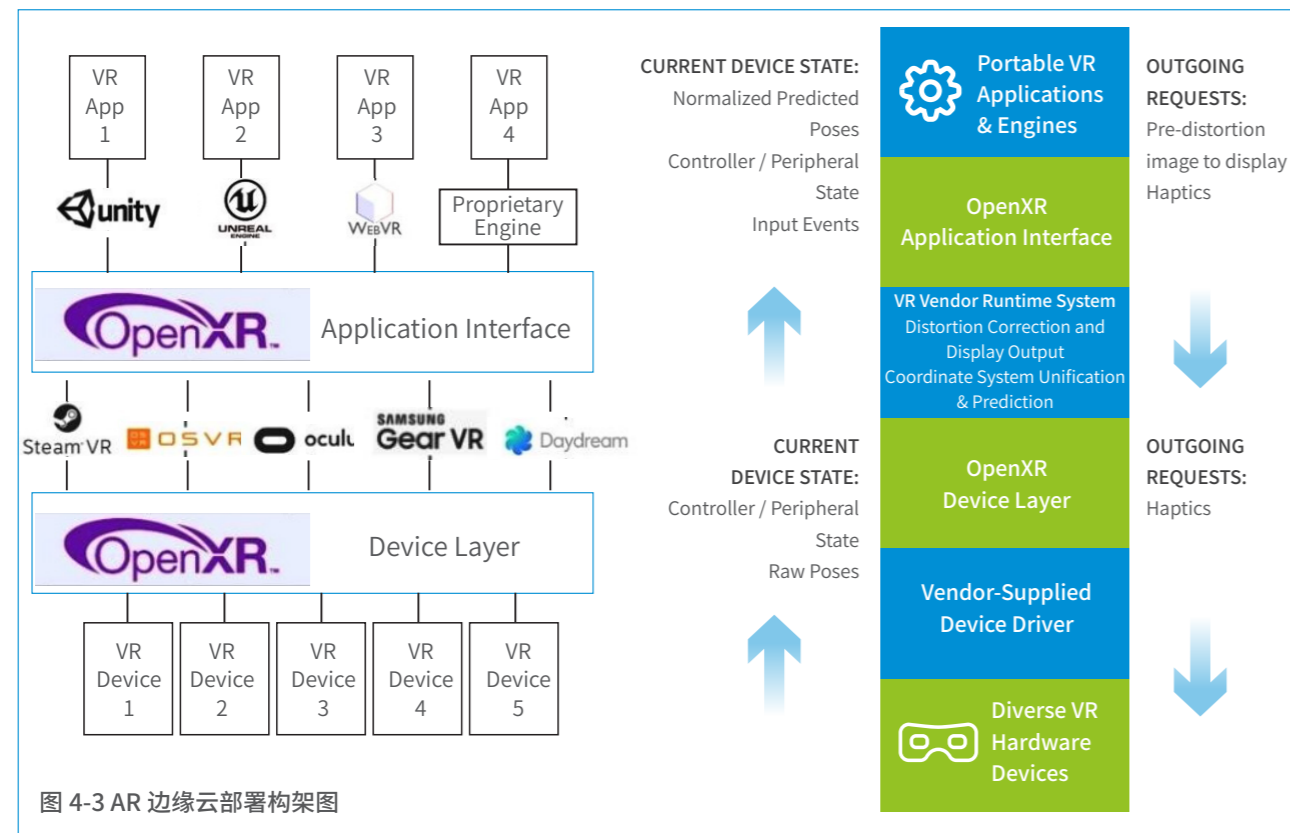
如图 4-2 所示，确定性网络（DetNet, Deterministic Networking），一项帮助实现 IP 网络从“尽力而为（best-effort）”到“准时、准确、快速”，控制并降低端到端时延的技术。5G 确定性网络对于确定性网络，是一个重要延伸和突破，真正解决从点对点的硬件互连转向无线互联，提高部署灵活性的诉求。

一致性能力集成接口

实现 5G 为核心和入口的庞大设备群落的体验一致，在之前面临着设备能力，渲染引擎能力接口和开发环境调用接口碎片化的问题。

Khronos 最新为 VR 和 AR 应用程序和设备创建一个开放且免版税的标准，OpenXR，并发布了标准版本 1.0 相关文档、接口定义和参考代码。

OpenXR 由两个主要层组成：OpenXR 设备插件扩展和更高级别的 OpenXR API。包含一组抽象，可允许 XR 硬件平台供应商公开 Runtime 运行时系统的功能。通过在应用程序生命周期，渲染，追踪，帧定时和输入中采用一组通用的对象和函数，软件开发者的应用程序只需最少量的移植工作即可兼容多个 XR 系统。



作为设备运行的基础操作系统的两大平台，UWP (Universal Windows Platform) 和 Linux，其对应维护组织微软和 Collabora 分别发布了支持对应平台的 OpenXR 1.0 Runtime，并为设备供应商提供针对平台的基本构件。其中特别是 GNU/Linux 的第一个 OpenXR™运行时，Monado，对 5G 云 AR 开放架构就显得异常重要。

05 AR边缘云核心技术发展预测



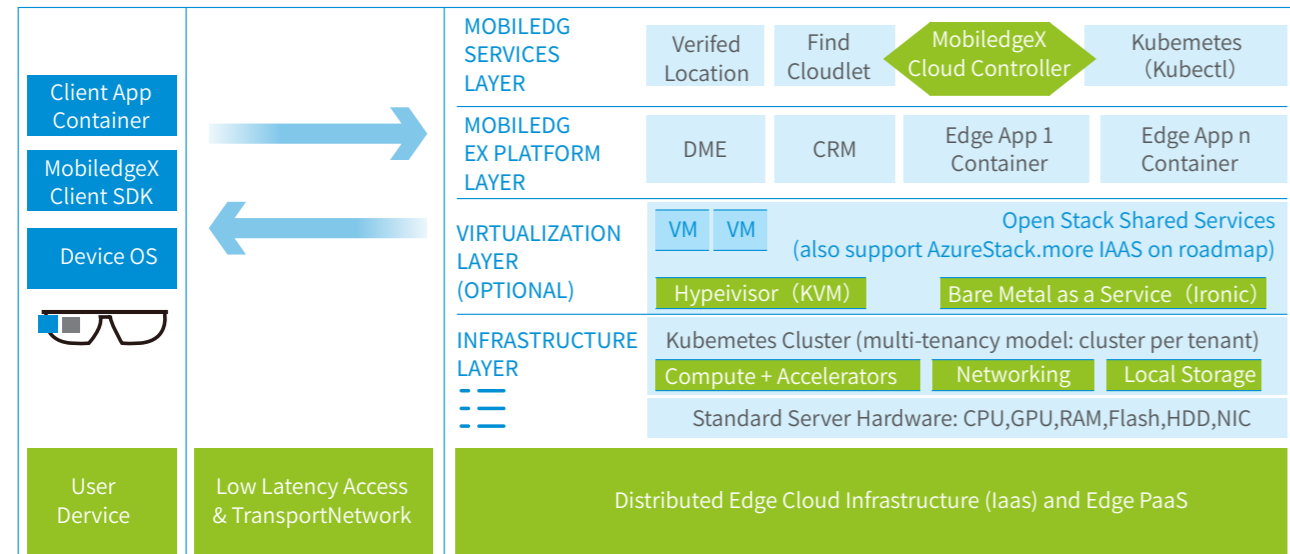


图 5-1 AR 边缘云部署架构图

“VM” Vs “Container” —— 虚拟机和容器的战争

将 AR 云的各种能力组件和应用逻辑通过虚拟机或者容器的方式在边缘云承载，并通过低延时传输完成被感知现实场景的传输和合成显示内容的推送是当前 AR 边缘云的标准实践模式。其中如何构建弹性的部署方式，又能提供对于开发生态比较好的包容和可用性，成为构架发展的重要影响因素和技术栈发展的关键挑战。

目前通过应用开发生态的主流工具如 Unity3D (2019 当前开发市场的 2/3 占比)，对于第三方能力部署组件仅能支持 Android, UWP, IOS 等有限模式。发展方向“Android Emulator on Docker”就成为 K8S+Docker 支持者在着力推进的一个方向。

同时业界刚刚兴起的“AnBox: Android in Box”和“Podman: The next generation of Linux container tools”能否借助 AnBox 对 Android 生态和发布的快速无缝对齐，以及可以和本机原生媲美的运行速度，同时结合 Podman 实现更加安全和更可靠和更高效的弹性部署可能是业界新锐团队可以突破的一个方向。

除了生态的推动和影响，容器化的硬件设备虚拟化（如 NVIDIA 的 nvDocker 和 AMD 的 ROCm 等），是连接应用逻辑和特定加速硬件的桥梁。也是决定性性价比的核心竞争力。

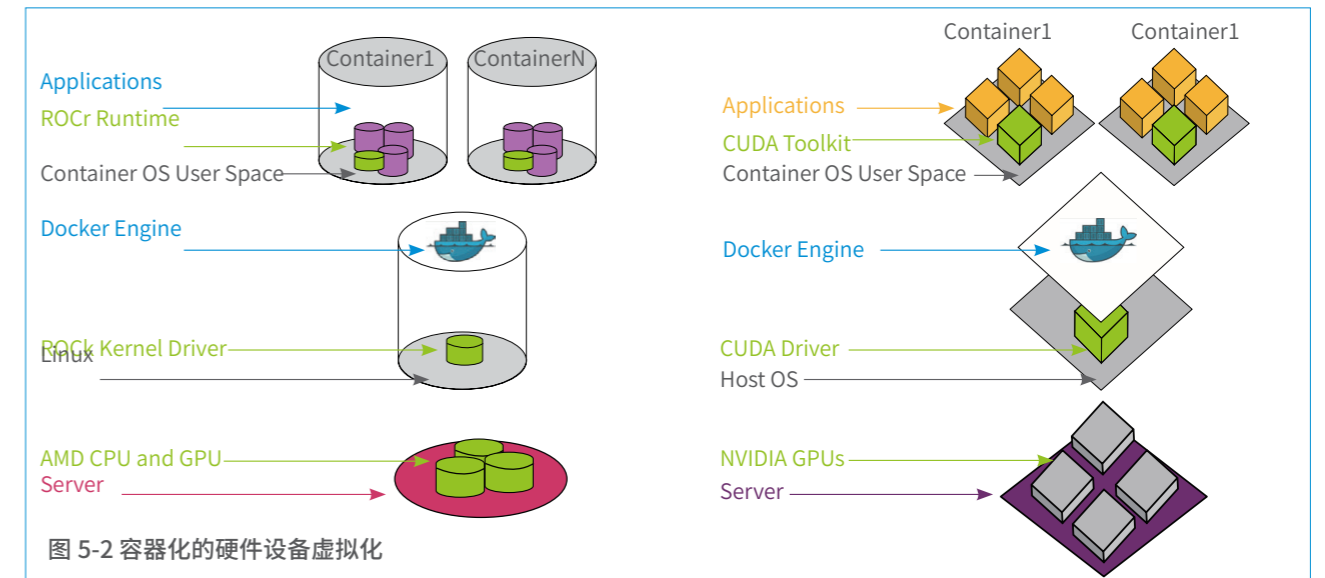
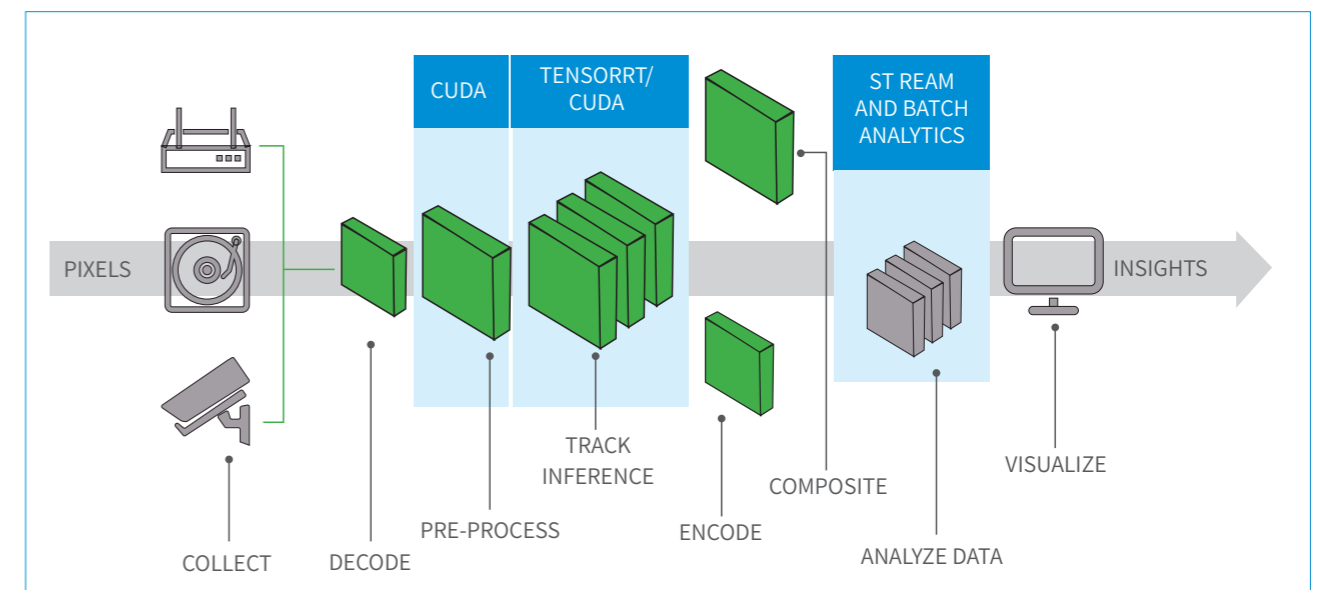


图 5-2 容器化的硬件设备虚拟化

而且，随着视觉感知的服务内容和传感器形态和模式的不断发展，加速芯片的发展也呈现了百花齐放百家争鸣的态势，包括了阿里、寒武纪、谷歌等知名企业。如何将硬件设备的感知能力、渲染能力和低延时编码能力能有效在硬件层完成有机整合和充分的可编程灵活就显得异常重要。

业界可见的包括了基于 RDMA 的英伟达的 DeepStream 框架以及中兴通讯宣称的 uSmartIN 包含的一种基于异构计算的任务处理方法及软硬件框架系统。如何在任务导向的系统设计以及对于软硬件整合构架设计中体现出独特点和差异点可能就是业界新锐团队后期实现自身突破的一个重要举措。



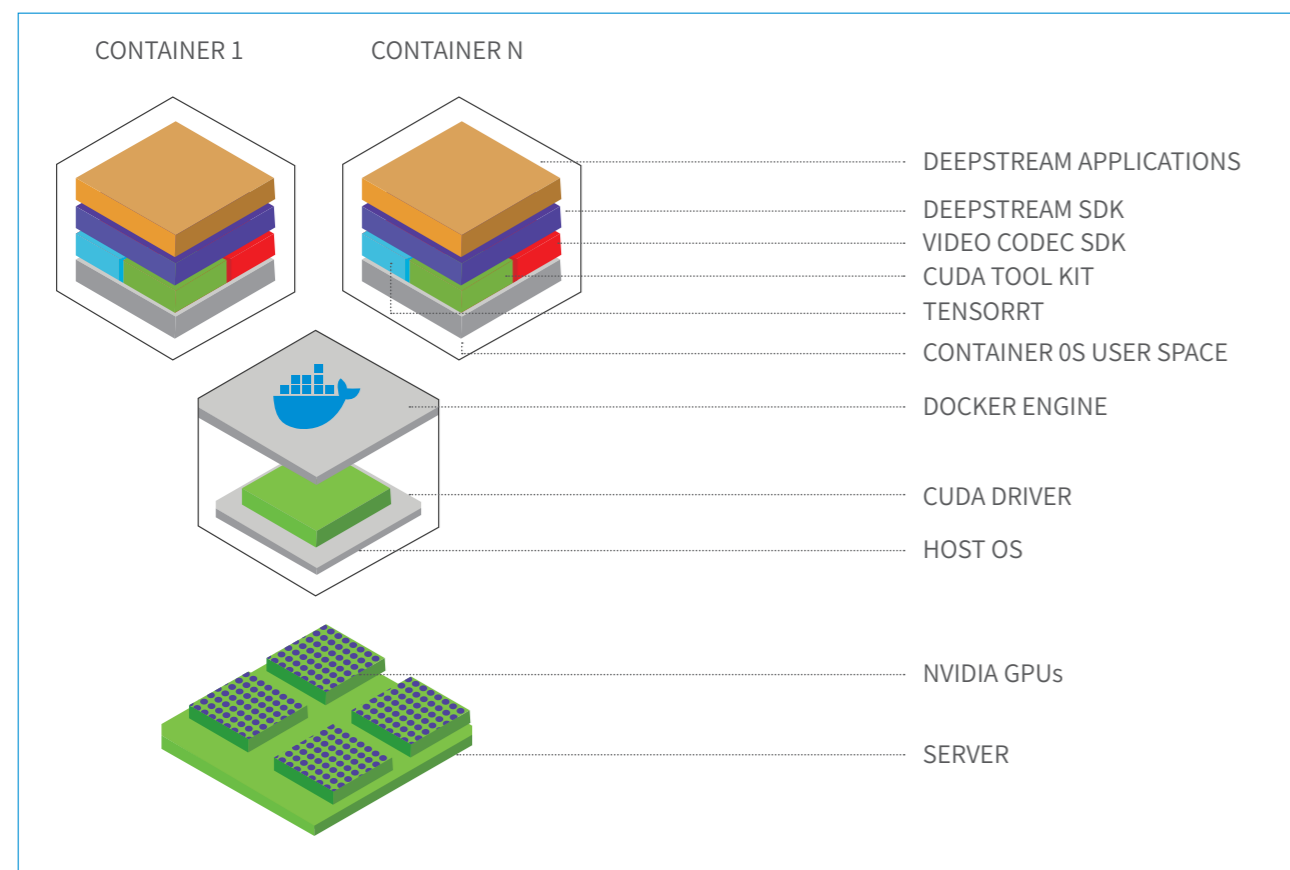


图 5-3 容器化的硬件设备虚拟化

后记：AR 边缘云@5G趋势把握



“产业趋势决定行业发展的方向，生态结构决定技术突破的价值”：

5G的大规模部署，随之而来的终端和模组成本的下行，必将放大其作为视觉主导的人物交互（HTI）中重要传输通道的作用。同时边缘计算的产业模式也必将使AR从垂直领域信息辅助，向知识和信息汇集呈现和新一代物联网的入口发展，其垂直领域的接口和数据转换，以及数据可视化呈现的创新方式都将为AR应用的繁荣和普及范围扩大带来更加广阔的前景。作为和人交互的入口，其在垂直领域的应用价值和垂直领域固有的生态结构密不可分，如何突显信息交互加速和便捷，以及信息可视化的重要作用就成为推广AR行业应用需要切实考虑的突破点。

“盈利模型决定技术框架的走向，客户体验驱动产品价值变现”：

AR系统部署的实际构架，除云—边—端整体构架趋向成熟外，其实际算力和功能的分布依然随着AR功能和应用的进一步充实还存在巨大的机遇。其中，泛在显示的概念和新兴近眼显示技术的进步，以及手机和AR头盔结合的应用，都将减低AR前端的客户进入门槛，推动产业繁荣，这也为前端模组和芯片的新锐厂商提供了差异化的切入机会，使得渲染、感知等单一功能优异者也有了赢得市场的机遇。同时客户体验是决定产品接受度的重要因素，为此无论是佩戴的舒适性，还是充分考虑使用场景和客户习惯将成为影响各个行业参与者成长和市场开拓的重要驱动。