6G沉浸式通信与AI原生网络融合：愿景、关键技术与挑战

黄亚坤[1]，苏洋[1]，王春生[2]，陆平[3]，谢大雄[3]，乔秀全[1]
（1.北京邮电大学网络与交换技术全国重点实验室，北京 100876；2.中电万维信息技术有限责任公司，甘肃 兰州 730030；
3.移动网络和移动多媒体技术国家重点实验室，广东 深圳 518057）

摘要：沉浸式通信是6G面向虚实融合与智能协同的重要业务形态，其多模态高保真、亚毫秒级交互与时空一致性要求，使传统以比特可靠传输为中心的网络优化难以直接满足体验约束。现有标准化工作主要给出服务级需求与能力目标，相关综述多围绕单点技术展开，缺少将用户侧物理体验质量（QoPE）与网络KPI及AI机制关联起来的系统梳理。结合标准化进展与代表性研究，提出了QoPE-KPI的双维层级化需求框架，归纳内容处理、传输、计算与控制的一体化关键技术体系，重点总结语义通信与生成式重构、AI-native空口与资源管理、端边云协同渲染、多模态同步与QoPE反馈，以及AI Agent协同交互等方向，并讨论上行受限、端到端时延与抖动、空天地异构、语义鲁棒性与隐私安全等问题及可能的演进方向。

关键词：沉浸式通信；AI原生网络；物理体验质量；语义通信；边缘智能
中图分类号：TP37；TP393
文献标志码：A
doi：10.11959/j.issn.1000−436x.TXXB260039

Integration of 6G immersive communications and AI-native networks: vision, key technologies, and challenges

Huang Yakun[1], Su Yang[1], Wang Chunsheng[2], Lu Ping[3], Xie Daxiong[3], Qiao Xiuquan[1]
1. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
2. China Telecom Wanwei Information Technology Co., Ltd., Lanzhou 730030, China
3. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518057, China

Abstract: Immersive communication is a key 6G service paradigm for enabling cyber-physical integration and intelligent collaboration. Its demands for multimodal high fidelity, sub-millisecond interaction, and spatio-temporal consistency made traditional network optimization focused on reliable bit delivery inadequate for meeting user experience requirements. Current standardization mainly specifies service requirements and capability targets, while most existing surveys examine individual technologies without systematically linking user-side quality of physical experience (QoPE) to network KPI and AI mechanisms. Drawing on recent standardization progress and representative studies, a dual-dimensional hierarchical framework connecting QoPE-KPI was proposed, and presented a unified technical system spanning content processing, transmission, computing, and control. It highlighted semantic communication and generative reconstruction, AI-native air interface and resource management, end-edge-cloud collaborative rendering, multimodal synchronization and QoPE feedback, and AI-agent-based collaborative interaction. Issues such as uplink constraints, end-to-end latency and jitter, heterogeneous space-air-ground networks, semantic robustness, and privacy and security, and outlines future research directions were also discussed.

Key words: immersive communication, AI-native network, QoPE, semantic communication, edge intelligence

收稿日期：2026−01−15；修回日期：2026−03−11
通信作者：乔秀全，qiaoxq@bupt.edu.cn
基金项目：国家自然科学基金资助项目（No.62571055, No.62431015）
Foundation Items：The National Natural Science Foundation of China (No.62571055, No.62431015)

0　引言

随着5G规模部署和应用推进，移动通信研究正由现有系统能力增强转向第六代移动通信系统（6G）体系设计[1]。面向虚实融合、智能协同和全域感知等需求，6G将在网络架构与资源组织中引入更强的智能化能力[2]。人工智能（artificial intelligence, AI）不再局限于应用侧附加功能，而是逐步进入网络侧感知、控制和运维链路，形成以数据驱动的状态感知、决策与策略更新能力，即AI-native网络[3-4]。沉浸式通信被视为6G的重要应用形态之一，其业务已经由早期的以多媒体增强为主，演进为以全息通信、数字孪生和多主体交互为特征的沉浸式协作场景。不同于传统视频业务，沉浸式通信需对高分辨率视觉、空间音频、触觉反馈和环境感知等多模态数据进行实时协同传输与同步控制，呈现多模态融合、强交互和高并发特征，并对峰值吞吐、端到端时延、抖动及时空一致性提出更高要求[5-6]。在物理层容量受限且端侧算力、能耗受约束条件下，单纯以比特可靠传输为中心的网络设计已难以满足上述需求，成为制约沉浸式业务持续提升的关键问题[7]。

针对上述需求，标准化组织已开展相关研究。3GPP SA1（3rd generation partnership project, service and system aspects working group 1）及无线电接入网（radio access network, RAN）等工作组发布的6G用例、服务需求与能力目标研究报告[8-9]，已将沉浸式通信、虚拟世界/数字孪生和AI Agent纳入讨论，并提出高速率、低时延、高可靠和智能化等目标。此类研究多以服务需求和网络关键绩效指标（key performance indicator, KPI）为主线，但沉浸式业务的可用性还受多模态同步、交互一致性和人体生理反馈等因素影响。因此，一方面，如何建立物理体验质量（quality of physical experience, QoPE）与网络侧指标之间可操作和可验证的映射关系，并据此设计AI-native网络仍缺乏系统梳理。另一方面，在物理层容量受限且端侧算力、能耗受约束的条件下，AI-native网络成为支撑沉浸式业务的一类重要方向。与应用侧叠加智能模块不同，AI-native网络强调在网络侧贯通数据获取、推理决策与执行反馈，以支持面向体验或任务目标的在线调整[10-11]。同时，语义通信和任务导向通信通过提取任务相关语义表征，并结合知识推理或重构，缓解沉浸式场景下的多模态传输压力[12-14]。

AI-native网络与跨层自适应等方向虽已显示出可行性，但将沉浸式业务需求转化为可部署和可验证的协同方案仍面临若干关键问题。首先，舒适度、多模态一致性等体验要求难以由现有KPI直接表征，QoPE的观测、估计与验证尚未统一；其次，业务状态表征与网络控制入口缺乏一致性约定，既制约跨层方法复用，也不利于与标准化成果对齐；再次，端到端链路涵盖编解码、传输和渲染等多个环节，尾部时延与抖动难以给出可验证的上界及异常工况处置策略，体验稳定性缺乏保障。此外，跨域数据共享还受隐私合规、数据治理成本和责任边界限制。上述因素使沉浸式业务的主要挑战由平均带宽保障转向交互时延、抖动控制和多模态一致性的稳定满足。因此，AI-native网络融合并非简单的功能叠加，而是围绕QoPE目标（或阈值要求）明确业务侧可观测信息与网络侧可控参数的对应关系，形成跨域策略、执行与反馈更新机制，从而支撑面向体验的协同优化。

尽管6G沉浸式通信、AI赋能网络和语义通信等方向已取得一定进展，但现有综述仍存在以下不足：多数工作以网络KPI或媒体编解码为主线，对QoPE的定义、度量及其与KPI映射关系的讨论不足；关于AI与沉浸式通信的梳理多聚焦于单点技术，缺少贯穿内容处理、传输、计算和控制各环节的分析；对标准化进展的引用和分析较为分散，难以直接支撑工程落地与跨层协同设计。表1给出了本文方法与代表性综述在覆盖范围等方面的对比。本文结合标准化进展与代表性研究，对6G沉浸式通信开展综述，主要贡献如下。

（1）提出了QoPE-KPI层级化需求框架，给出QoPE的可操作定义，并补充可测信号的采样频率与统计窗口、噪声及时滞处理、个体差异校准，以及由QoPE阈值反推KPI预算和策略边界的方法。

（2）从内容处理、传输、计算和控制4个环节梳理关键技术，围绕语义通信与生成式重构、AI-native空口与资源管理、端边云协同渲染、多模态同步保障和AI Agent协同交互，总结研究进展、代表性方法及待解问题。

（3）从工程应用角度讨论上行受限、端到端时延与抖动控制、空天地异构、语义鲁棒性及隐私安全等挑战，并讨论未来技术演进与标准化方向。

表 1　代表性综述工作对比

代表性综述	主要侧重点	覆盖范围	与本文区别
文献[6]	6G沉浸式通信总体进展	沉浸式场景与需求概览	缺少QoPE定义与QoPE-KPI对应，AI-native机制讨论少
文献[7]	AI-native 6G综述	AI-native架构与技术	偏AI-native总体架构，沉浸式渲染/同步/体验约束覆盖范围有限
文献[14]	语义通信技术综述	语义理论/框架/应用	偏通用语义通信，未覆盖沉浸式多模态与QoPE约束
文献[15]	沉浸式视频编码综述	编码与压缩技术	偏编码压缩，缺少RAN/CN协同、语义传输与QoPE约束组织
本文方法	AI与沉浸式通信融合的系统图谱	需求、技术体系及挑战/方向	QoPE-KPI、端到端技术梳理、标准化关联/挑战

1　沉浸式通信的发展脉络与演进趋势

沉浸式通信随多媒体编码、人机交互和网络能力演进，主要经历4个阶段（如图1所示）：以视听质量提升为主的基础感知阶段，随扩展现实（extended reality, XR）发展形成的多模态交互阶段[16]，基于全息通信与体积视频表示的三维真实重构阶段[17-18]，以及面向6G、侧重语义压缩与智能协同阶段[14]。这一演进表明，系统优化目标正由比特传输能力转向交互效用与沉浸体验一致性。

1.1　高清视频与多媒体通信阶段：平面感知增强

在4G/5G发展早期，沉浸感主要依赖更高的视觉分辨率和更丰富的声场呈现。随着H.265/HEVC及后续编码格式普及，移动视频由标清向4K/8K超高清演进，并加速终端普及[19]。网络侧通过自适应码率（adaptive bit rate, ABR）和分层内容分发网络（content delivery network, CDN）部署，降低了无线带宽波动引发的卡顿和中断风险[20-21]。此阶段业务形态仍以单向二维视频流为主，用户交互有限，相应的网络优化重点主要在于提升链路与物理层频谱效率，以满足更高吞吐需求，如更高阶调制和多输入多输出（multiple-input multiple-output, MIMO）结构演进等[22-24]。由于内容仍以二维视频呈现为主，系统对空间自由度、深度信息和语义结构的处理需求尚不突出，主要体现为平面视听质量增强。

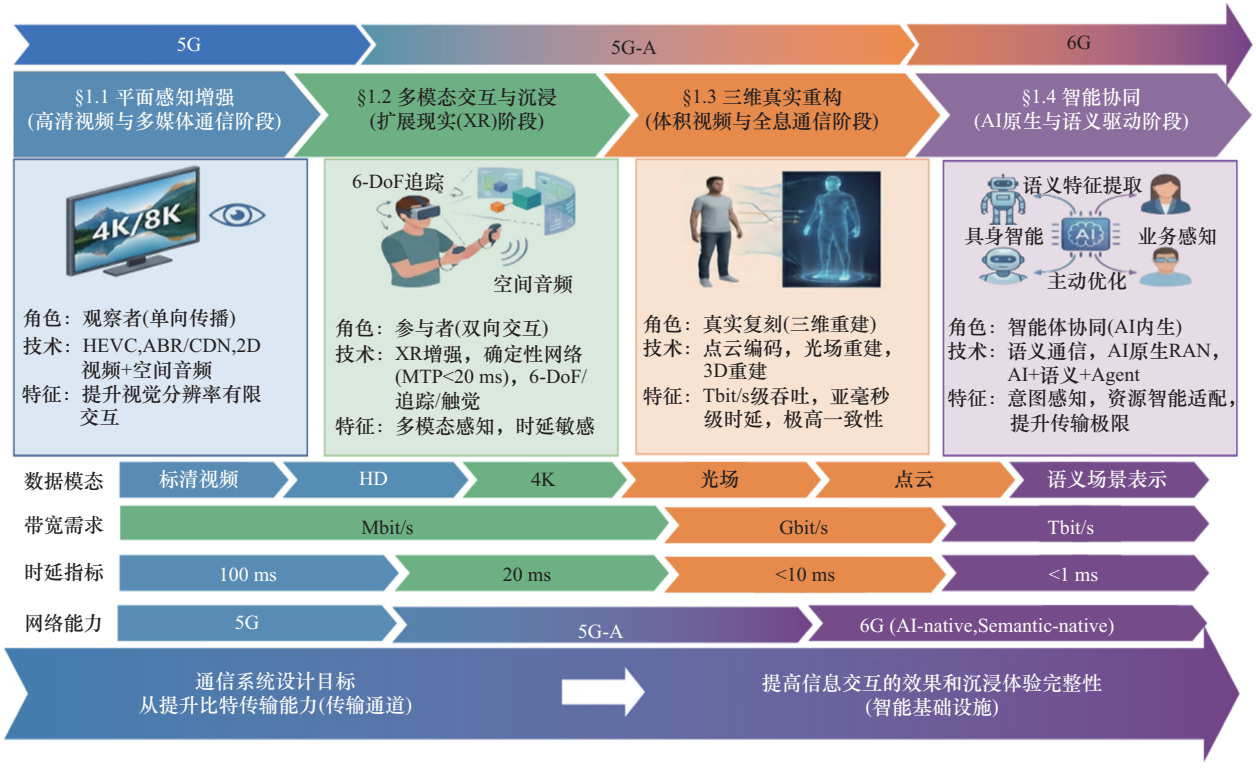

图1　6G沉浸式通信技术、需求与网络能力演进

1.2　扩展现实(XR)阶段:多模态交互与沉浸

随着头显设备和体感交互技术成熟，沉浸式通信进入以XR为代表的多模态交互阶段。XR引入六自由度追踪、双目立体渲染和空间音频处理，使用户由观看转向虚拟环境中的交互参与。不同于二维视频，该阶段的网络约束不再局限于下行吞吐，而是对上行姿态和视点更新的频率及时效性也提出更高要求。为降低晕动症风险，运动到光子（motion-to-photon, MTP）时延通常需控制在20 ms以内，因此端到端时延与抖动需满足更严格且可预测的约束[25-26]。针对XR需求，3GPP在Rel-18中引入了XR增强机制，包括应用感知资源调度、协议数据单元（protocol data unit, PDU）Set级服务质量（quality of service, QoS）控制、低功耗策略和多模态数据同步等[27]。同时，SA4的沉浸式语音与音频服务标准也给出了空间音频编解码与传输保障规范[28]。总体来看，XR阶段推动网络优化重点由带宽和可靠性保障，进一步转向时延、抖动和多模态同步控制。

1.3　体积视频与全息通信阶段:三维真实重构

在6G场景下，体积视频与全息通信推动沉浸式表示由视听增强/交互进一步走向三维真实重构[15]。该阶段以点云编码与光场重建为基础，典型技术包括运动图像专家组（motion picture experts group, MPEG）基于视频的点云压缩（video-based point cloud compression, V-PCC）与基于几何的点云压缩（geometry-based point cloud compression, G-PCC）等[29-30]，目标是在远端实现对物体与环境的高精度三维表示与重构。相较于XR，全息业务不仅数据规模显著增加，同时对空间一致性与多视角同步提出更严格约束。已有研究指出，实时全息交互可能需要Tbit/s量级峰值吞吐与亚毫秒级端到端时延[31-32]，超出现有5G体系的常规能力范围。进一步，全息数字孪生可在虚拟空间中实时映射物理实体，为工业控制与远程医疗等场景提供更高保真度的远程操作环境。

1.4　AI原生与语义驱动阶段:智能协同

面向2030年及6G后续演进，沉浸式通信在智能协同与语义驱动视角下受到更多关注。一方面，沉浸式业务产生的大规模多模态数据可为具身智能、数字人和多主体协同提供高维观测与训练数据；另一方面，网络侧智能能力的引入使系统能够更细粒度地感知和预测沉浸式业务状态，从而支持更主动的资源调节。在标准化方面，3GPP的6G用例与服务需求研究已将沉浸式通信、虚拟世界/数字孪生和AI Agent纳入讨论，并提出面向体验与任务的能力保障需求[8]。为缓解沉浸式业务的数据传输压力，任务导向和语义驱动的特征提取与传输受到关注，并被认为可结合XR、工业控制等场景，降低带宽占用并改善交互时效[12-14]。在学术研究方面，一类工作面向语义驱动的全息传输，仅传递关键结构信息，并在边缘侧利用生成模型重构细节，提高传输效率[33-35]；另一类工作关注AI赋能的资源管理，采用深度强化学习等方法结合AI-native RAN设计，将感知、预测与决策纳入接入侧调度流程，以改进资源分配与能效[36-37]。

2　面向QoPE-KPI映射的6G沉浸式通信层级化需求体系

沉浸式通信对6G网络的吞吐、时延与一致性保障提出了更高要求。为将应用侧需求转化为可评估的系统能力，3GPP正从业务应用与无线接入两条主线开展需求研究：一方面，6G用例与服务需求研究报告给出了沉浸式通信等典型用例及服务级要求[8]；另一方面，面向无线接入网的场景与需求研究从RAN视角讨论关键能力目标与相关假设[9]。但在当前表述体系下，服务级需求与网络KPI之间仍难直接对齐：沉浸式体验的可用性往往由多模态同步、交互一致性与人体生理反馈等因素共同决定，而这些因素难以用传统KPI进行充分刻画。基于上述标准化进展，本节提出了面向QoPE-KPI对应关系的层级化需求体系，为后续AI-native机制设计提供可操作的指标表述与约束条件。

2.1　业务视角:从多媒体增强迈向全息协同

作为3GPP面向6G第一阶段的需求研究成果，典型用例与服务需求研究报告将沉浸式通信、数字孪生与虚拟世界交互，以及AI Agent协同纳入典型用例与服务需求范围[8]。相较于5G时代以XR为代表的沉浸式业务，6G沉浸式通信的差异主要体现在交互形态与协作对象两方面：（1）交互以视听为主扩展为多模态交互，全息远程呈现除高保真三维视觉外，还涉及触觉、动态感知与环境传感等数据的同步；（2）协作由人机交互进一步扩展到人机物多主体协同，AI Agent的引入使网络除承载用户媒

体数据外，还需支持数字人、机器人与物理环境之间的大规模状态同步与控制信令交互。相应地，服务需求不再仅以带宽等单一指标刻画，而更强调体验质量与任务效果相关约束，并要求网络在动态条件下能够维持多模态一致性与交互时效性。

2.2　系统视角：确定性性能与多维能力边界

为满足沉浸式通信等业务对无线接入的要求，RAN 侧的 6G 场景与需求研究报告与 ITU-R IMT-2030 框架中提出的能力目标在方向上具有一致性[9,38]，但相关工作仍处于研究阶段，尚未形成类似 IMT-2020 最低技术性能要求那样可直接用于验收的量化指标体系。沉浸式通信网络侧的关键约束主要体现在上、下行峰值需求与非对称性，端到端时延与抖动的上界控制，以及多模态同步、定位与算网协同等跨域能力要求。由此可见，仅对传统 KPI 进行简单扩展难以覆盖体验层约束，有必要建立体验度量（quality of experience, QoE）和 QoPE 与可控 KPI 之间的可操作对应关系。

2.3　沉浸式通信的关键需求体系归纳

如图 2 所示，本文将 6G 沉浸式通信需求归纳为由体验度量、能力约束和运行机制构成的层级化需求体系。不同于仅在业务与应用层罗列的网络指标，本文在上层引入以 QoPE 为核心的体验度量：业务与应用层给出临场感、同步度和舒适性等交互目标，系统侧通过 QoPE/QoE 模型量化估计用户侧物理、行为和主观反馈，并将其转化为网络侧可实施的约束与目标。同时，建议采用“快变量+慢变量”的双时间尺度观测口径：MTP 时延、帧间抖动、跨模态偏差和渲染队列等快变量按每帧或 10~20 ms 滑动窗口采集，位姿/视点等状态按 60~120 Hz 采集，主观评分和任务完成质量等慢变量按会话段或任务段统计；生理信号根据设备能力采用 10~100 Hz 的 PPG/皮电采样或更高频 EEG 采样，并在统一时间戳下对齐。在线估计中，可通过滤波、指数加权移动平均平滑和异常值剔除抑制突发尖峰与测量噪声；对于网络变化到生理反应之间的时滞，可利用互相关或时序模型进行因果对齐；对于个体差异，可通过静息基线、分用户标准化和在线自适应参数进行校准。

表 2 进一步将 QoPE 维度转为可执行的 KPI 预算与控制边界。设网络与系统关键量 $k=\left[d,j,e_{\text{sync}},p_{\text{loss}},r_{\text{retx}},d_{\text{comp}},e_{\text{pos}}\right]^{\mathrm{T}}$，其中，$d$ 表示端到端时延，j 表示时延抖动，e_{sync} 表示跨模态同步误差，p_{loss} 表示丢包率，r_{retx} 表示重传相关开销，d_{comp} 表示计算时延，e_{pos} 表示定位误差。记 QoPE 预测值 $\hat{q}=f\left(k,c,\theta_u\right)$，其中，$c$ 表示场景上下文信息，θ_u 表示个体参数。给定业务侧 QoPE 阈值为 $q_{\min}$，将满足 $\hat{q}\geqslant q_{\min}$ 的可行集定义为网络预算集合。在运行点附近，可采用局部线性形式：

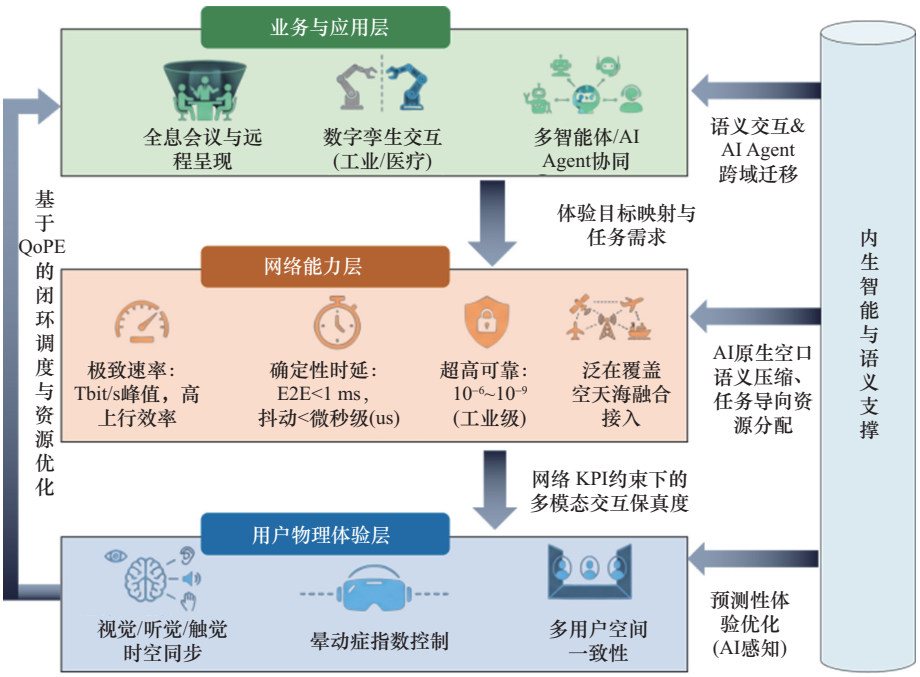

图2　面向 QoPE-KPI 映射的 6G 沉浸式通信层级化需求体系

表2　　QoPE维度、测量口径与KPI预算/策略边界映射示例

QoPE维度	可测信号/建议采样方式	噪声、滞后与个体差异处理	KPI预算/策略边界
舒适性/晕眩	头动/视点60~120 Hz；MTP/帧间抖动按帧或10~20 ms窗口；PPG/皮电10~100 Hz；不适评分按30~60 s会话段	统一时戳；中值滤波去尖峰；互相关估计生理滞后；静息基线归一化并在线更新敏感度	单次重传开销超过剩余时延裕量时，禁用重传，改用FEC、遮挡补偿或降分辨率
模态同步一致性	视–听–触偏差按事件时戳或5~10 ms窗口；音视频到达偏差按帧；控制事件实时记录	公共时基；迟到样本剔除；因果对齐估计网络–渲染滞后；初始校准偏移修正用户差异	预计超阈值时优先保证时序一致，并降低分辨率或触觉刷新；仅在补偿后仍满足阈值时允许缓存拉伸
交互真实感/任务达成	成功率/任务时间按任务段；注视、点击、操作日志采样频率30~120 Hz；渲染队列/推理时延按10~20 ms滑动窗口采集	日志去抖与异常合并；主观评分采用分用户z-score或分层模型；任务反应滞后按累积效应建模	资源紧张时优先保障关键对象或语义层，非关键细节分级丢弃
空间一致性/定位稳定	位姿/锚点误差60~120 Hz；SLAM/协同地图更新时间10~50 Hz；共享坐标漂移按会话段统计	卡尔曼/因子图滤波抑噪；握持/佩戴偏差作为初始化偏置；区分瞬时抖动与累计漂移	边缘地图更新超界时切换本地锚定或弱共享模式，并提高定位报文优先级

$$\hat{q} \approx q_0 - \sum_i w_i\left(k_i - k_{i,0}\right) \tag{1}$$

则得到：

$$\sum_i w_i \Delta k_i \leqslant q_0 - q_{\min} \tag{2}$$

其中，w_i表示第i个指标对QoPE的局部敏感度系数，Δk_i表示相对参考运行点的扰动量。该计算式表示当其他因素固定时，可直接反推出抖动上界、同步误差上界、定位误差上界等单项预算。当多项因素同时波动时，则按局部敏感度系数w_i在传输、计算与渲染之间分配总裕量。例如，可将端到端时延预算分解为：

$$D_{\text{e2e}} = D_{\text{ul}} + D_{\text{ran}} + D_{\text{edge}} + D_{\text{render}} + D_{\text{display}} \leqslant D_{\text{th}}\left(q_{\min}\right) \tag{3}$$

其中，D_{ul}、D_{ran}、D_{edge}、D_{render}和D_{display}分别表示上行传输、无线接入、边缘处理、渲染和显示环节的时延预算。同时，将跨模态同步约束定义为$e_{\text{sync}} \leqslant e_{\text{th}}\left(q_{\min}\right)$，并把丢包/重传策略的触发边界定义为：若一次重传预计带来的额外时延超过当前剩余时延裕量，则优先采用前向纠错（forward error correction, FEC）、错误隐藏、视场降级或语义补全；仅当重传仍在预算内时，才允许有限次自动重传请求（automatic repeat request, ARQ）或混合自动重传请求（hybrid automatic repeat request, HARQ）。

3　AI与沉浸式通信融合关键网络技术体系

6G沉浸式通信对网络在超高带宽、低时延和多模态一致性保障方面提出同时满足的要求。与此同时，物理层频谱效率提升空间受容量上限约束，仅依赖带宽扩展或协议层叠加难以覆盖沉浸式业务对吞吐、时延与一致性约束的综合需求。因此，研究普遍关注在网络侧引入AI能力，将感知、传输、计算与控制纳入统一运行流程，以支持面向体验与任务的在线调节[39-40]。本文从端到端视角归纳相关关键技术，如表3所示：沉浸式内容生成与理解、面向任务与体验的语义通信新范式、AI驱动的确定性空口与资源管理、协同渲染与边缘智能体系、多模态同步与QoPE保障，以及AI Agent驱动的沉浸式协同交互。

3.1　沉浸式内容生成与理解

全息、点云和多视点视频等沉浸式媒体数据维度高，处理链路计算开销大。随着深度生成模型发展，内容表示与压缩正由基于离散采样的显式表示，转向以隐式表征为核心的参数化表示。以神经辐射场（neural radiance field, NeRF）[41]和3D高斯溅射（3D Gaussian splatting, 3DGS）[42]为例，这类方法可在稀疏视角输入下实现高保真场景重建。以网络参数或高斯核为载体的表示方式，一方面降低了采集与建模对密集视角数据的依赖，另一方面也为“低码率–高质量”的联合权衡提供了基础。基于这一变化，语义理解可进一步指导传输侧资源分配：借助视觉Transformer等模型对人物、物体和环境结构的解析结果，系统可按重要性差异化处理，对关键区域采用高保真编码，对非关键区域在边缘侧进行生成式重建，从而在保证关键感知要素的前提下降低传输负载（如图3所示）。此外，

表 3 AI 与沉浸式通信融合的关键网络技术

关键技术	关键目标	代表性机制	开放问题
沉浸式内容生成与理解	降低采编开销，提升重构质量	NeRF/3DGS 表征；显著性驱动分级编码	实时性/端侧轻量化/跨场景泛化
面向任务与体验的语义通信新范式	以任务成功替代比特无误	语义特征提取；共享知识；生成式重构	语义度量/率失真理论/鲁棒性
AI 驱动的确定性空口与资源管理	保障时延/抖动，支持 QoPE 导向调度	流量预测；深度强化学习调度；闭环控制	可验证安全/跨域泛化/样本效率
协同渲染与边缘智能体系	缓解终端算力/能耗，提升交互流畅	FoV 预测；任务卸载；边缘渲染/推理	多模态协同/一致性维护/算网编排
多模态同步与 QoPE 保障	跨模态时空一致，降低晕动/不适	同步校准；生物反馈 QoPE 映射；预测补偿	体验可测性/隐私合规/闭环稳定性
AI Agent 驱动的沉浸式协同交互	支持任务级协同与自治编排	端/边/网智能体；意图映射	互操作标准/信任治理/责任边界

3GPP SA4 制定的沉浸式语音与音频业务规范，也为空间/沉浸式音频的编解码与端到端传输提供了标准化依据[28]。

3.2 面向任务与体验的语义通信新范式

与以比特可靠恢复为目标的经典通信不同，语义通信并不将逐比特无误重构作为唯一准则，而以任务相关信息的有效传递为优化目标。在香农框架下，源编码与信道编码通常围绕码率、误码率和与信号重构相关的通用失真度量进行设计。语义通信则引入任务/语义失真度量 $D_{\text{sem}}(\cdot)$，允许对任务无关的信息进行有损压缩，甚至不传输，从而在带宽受限条件下提高任务完成质量或体验可用性。

对沉浸式业务而言，上述“传输代价–语义/体验失真”权衡尤为突出：业务目标不仅是信号重构精度，还包括交互一致性、关键事件理解与任务达成等指标[43-44]。语义通信与任务导向通信提供了一种面向任务的传输组织方式（如图 3 所示）：发送端利用模型提取与任务高度相关的语义表征（如关键对象结构、动作参数与交互事件），接收端结合共享知识与生成模型进行重构或推理，从而降低对原始信号逐比特恢复的依赖。在沉浸式业务场景中，将其抽象为语义编码、信道传输、语义解码与生成式重构流程。设源端多模态输入为 x，任务与体验目标为 τ，语义编码器提取 $m = S_\theta(x,\tau)$，经信道传输得到 $\bar{m}$，接收端结合共享知识 K 与生成模型输出 $\bar{x} = G_\varphi(\bar{m},K)$。相应地，采用联合目标进行建模，如最小化 $R(m) + \lambda D_{\text{task}}(\tau,x,\bar{x})$，其中 R 为语义表征的传输代价，D_{task} 刻画任务失败或体验失真，λ 为权衡系数。

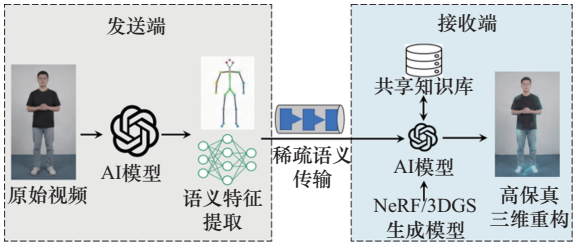

图 3 基于语义通信与生成式重构的传输框架

3.3 AI 驱动的确定性空口与资源管理

沉浸式业务对无线接入网的压力主要来自峰值速率需求、多用户并发下的调度复杂度，以及对时延和抖动上界的严格约束。在这类强时变场景中，固定规则或启发式调度方法难以稳定满足体验约束，因此研究常采用 AI 的预测与策略学习方法，如深度强化学习和在线学习，以支持自适应调度与资源编排[37,45-46]。需要指出的是，联邦学习并非具体调度算法，而是在隐私约束或数据不可出域条件下训练预测或策略模型的协同机制：各基站和边缘节点基于本地数据更新模型，仅交换并聚合参数或梯度，从而在不传输原始数据的前提下提升模型泛化性与适用性。本文将 AI-native 空口资源智能调度体系归纳为 4 个方面，如图 4 所示，并进一步讨论相应的关键优化问题。

（1）业务感知的自适应调度与预测性链路优化。在 3GPP Rel-18 XR 框架下，基站侧可利用在线学习方法刻画 XR/全息业务的流量与交互特征，如视点变化和交互强度等[47-48]。基于时序预测模型，网络可提前估计短期带宽需求，并据此进行细粒度资源预留与链路参数调整，以减弱突发流量对时延和抖动约束的影响。

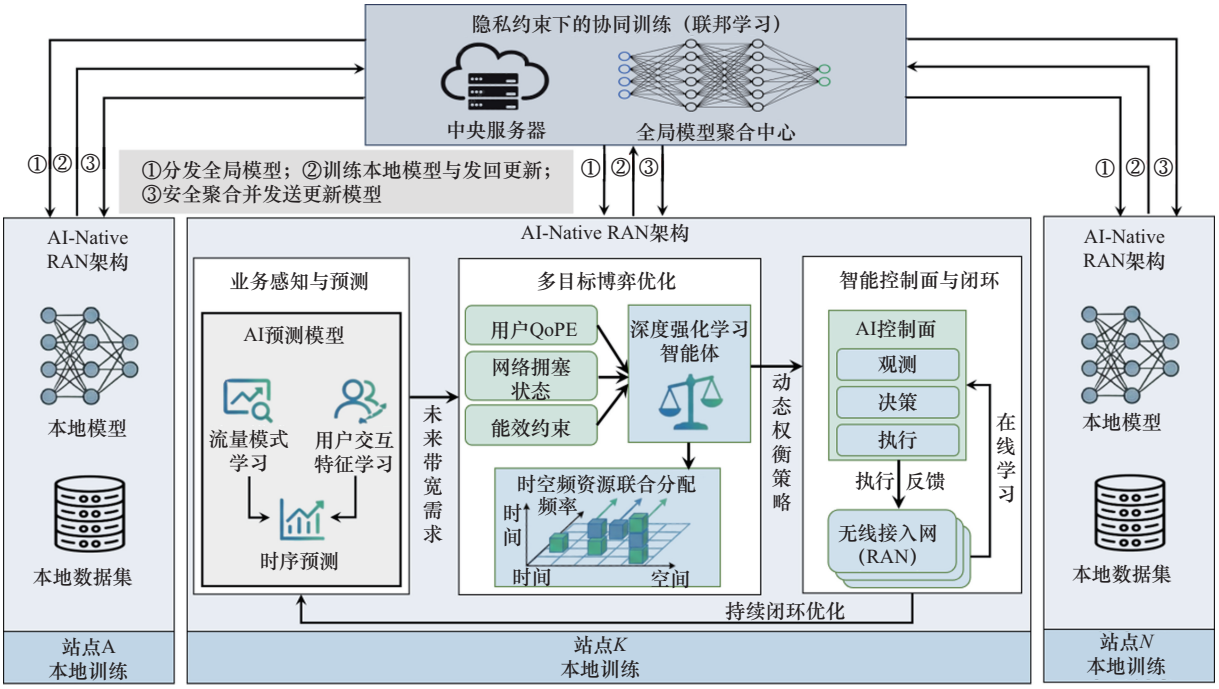

图4　AI-native空口资源智能调度体系

（2）多目标下的时空频资源联合优化。由于频谱、功率、时间和空间域资源相互制约，已有研究采用深度强化学习方法构建联合优化策略，在QoE、拥塞状态和能效等目标之间进行权衡[37,49]。这类方法仍以研究探索为主，其效果依赖场景假设和训练稳定性，但为高维资源配置提供了可行思路。

（3）引入智能控制面的AI-native RAN架构。部分工作提出在RAN中引入独立的数据与智能控制功能，将优化过程组织为观测、决策、执行和反馈的在线更新闭环[50-51]。在该架构下，可持续采集会话级指标，如帧率、重传率及必要的体验反馈，并通过在线学习方法调整调度优先级与切片资源配置，使资源管理更贴近体验约束。

（4）隐私约束下的分布式模型/策略协同训练。在跨站点和跨运营域的资源管理中，流量日志及QoPE相关数据受隐私与合规限制，难以集中处理。联邦学习可用于训练流量预测、QoPE映射或调度策略模型，各站点在本地更新参数，周期性上传模型更新，由聚合端（核心网或可信边缘）完成聚合并下发。该方式不需要传输原始数据，有助于提升跨场景适用性，但仍需结合安全聚合、差分隐私等机制降低参数泄露风险。

3.4　协同渲染与边缘智能体系

沉浸式通信，尤其是XR与全息交互，对图形渲染和推理计算提出了较高算力需求，而移动终端受功耗与散热限制，难以长期承担全部计算负载。因此，边缘渲染与端边云协同计算架构成为常见实现路径。该架构通常采用分级计算：终端侧负责感知输入采集与显示呈现，边缘侧负责低时延渲染和部分推理任务，云端负责复杂模型训练及全局场景数据处理。在此基础上，可通过资源调度在不同节点间分配计算负载，以满足时延与能耗约束。需要指出的是，协同渲染对象不局限于视觉渲染，也包括空间音频渲染以及触觉/力反馈等低时延交互链路。

以视觉渲染及相关推理为主的协同渲染方法可归纳为两类。（1）视点感知的自适应渲染：利用学习模型对头部姿态和眼动轨迹进行短时预测，边缘侧可提前预取视场（field of view, FoV）内容，抵消视觉渲染与传输时延[52-53]。结合注视点渲染，可对预测注视区域维持较高分辨率与刷新率，对外围区域降低采样密度，从而在满足体验约束的同时减少计算和带宽开销。（2）任务卸载与资源编排：在信道状态和算力资源随时间变化的条件下，可采用深度强化学习等方法决策任务划分与卸载位置，并综合终端电量、链路抖动和业务优先级等因素，在端边云之间选择满足时延与能耗约束的执行

方案[54-55]。

3.5 多模态同步与 QoPE 保障

沉浸式通信的体验质量不再由单一媒体流信噪比决定，更取决于视、听、触等多模态信息在时间与空间上的一致性。即使 MTP 时延仅有毫秒级偏差，或模态间存在不同步，也可能引发眩晕并降低交互稳定性。因此，需要引入以 QoPE 为核心的体验度量，并将其作为网络与终端协同控制的约束或目标。在观测窗口内，设沉浸式业务状态向量为 $z_t=\left[k_t,c_t,u_t\right]$，其中，$k_t$ 表示网络侧关键指标，如时延、抖动、同步偏差和丢包；c_t 表示业务上下文特征，如位姿预测误差、视场角变化率和渲染链路状态；u_t 表示用户生理与行为观测，如心率变异性、眼动轨迹和交互日志。QoPE 可定义为从多源状态到体验效用的非线性映射 $q_t=f\left(z_t,\theta_u\right)$，其中，$\theta_u$ 用于描述个体差异，如晕动敏感度、注意负荷阈值和佩戴偏置。在具体实现上，k_t 与渲染链路状态可以按每帧或 10~20 ms 窗口更新，位姿/视点根据 60~120 Hz 更新，生理信号则按传感器能力进行 10~100 Hz 轻量采样或更高频 EEG 采样，并在统一时戳下对齐。为抑制噪声与时滞影响，可先进行中值滤波、异常值剔除和滑动加权平均平滑和迟到样本剔除，再利用互相关或带时延嵌入的时序模型估计网络事件到主观/生理反应的滞后。对不同用户，则通过静息基线、分用户 z-score 或分层贝叶斯参数进行归一化与在线更新。由此，可将跨模态时序不一致与空间几何误差对体验质量的影响纳入统一刻画，使评估对象由网络 KPI 扩展为体验效用。为将 QoPE 用于在线控制，可构建可微预测模型 $\hat{q}_t$，利用时序模型刻画网络 KPI 波动与生理反应之间的时滞和累积效应，并通过注意力等机制识别不同场景下的主导因素。模型训练可基于主观评分、生理指标和行为日志构成的弱监督标签，采用回归与排序损失联合优化，以提升跨用户一致性。在线运行时，$\hat{q}_t$ 既可作为控制反馈，也可作为预算分解与降级触发判据，用于指导空口资源调度、边缘卸载和渲染参数调整，在满足 QoPE 阈值约束的前提下兼顾资源消耗与性能目标。

图 5 给出了多模态同步的 QoPE 控制流程，涉及 3 类关键机制。（1）时域同步与抖动补偿：在异构网络条件下，端到端时延具有随机波动特征。尽管现有 XR 架构已支持时延分解，但对随机抖动的抑制仍受限[56]。针对这一问题，可基于编解码、传输与渲染等环节的时延统计与瞬时状态建立预测模型，并据此进行前瞻式调度与缓冲控制，如提前安排关键帧、动态调整去抖动缓冲深度，并在必要时对不同模态流进行更细粒度的时间校准，以降低主观感知层面的不同步。（2）空域对齐与参考系一致性：空间一致性依赖定位与姿态估计精度。结合 3GPP Rel-18 引入的定位增强能力（如载波相位定位）以及学习驱动的姿态估计与视觉同步定位与地图构建（simultaneous localization and mapping, SLAM）等方法，网络与终端可在交互过程中维持一致的空间参考系[57]，从而提高虚拟对象锚定精度并降低多人协作下的空间漂移。（3）QoPE 评价与生理反馈利用：不同于主要关注清晰度与流畅度的传统 QoE 指标，QoPE 更关注晕动症相关反应、

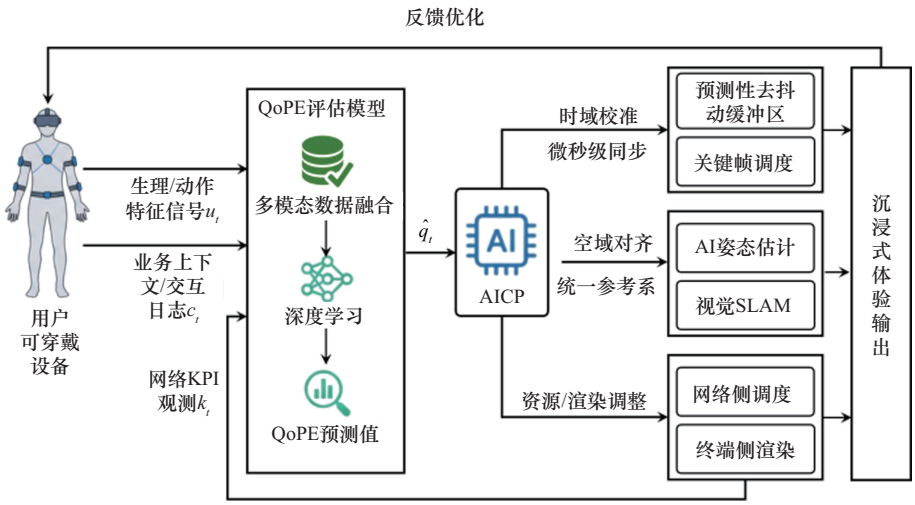

图 5 多模态同步的 QoPE 控制流程

交互一致性与认知负荷等生理与心理反馈[58]。在AI-native网络架构下，可融合可穿戴传感数据与交互日志，学习QoPE映射模型，将体验侧反馈转化为可用于策略调整的度量指标，并据此对渲染参数与资源调度进行更新。

3.6　AI Agent驱动的沉浸式协同交互

相较于以人际通信为主的既有移动通信系统，6G的连接对象与交互形态进一步扩展到人机物协同场景。在沉浸式业务场景中，用户、虚拟代理、数字人与实体机器人之间的交互，对网络与终端的感知、推理和决策能力提出了更高要求。为支撑这类能力，互联网工程任务组（Internet engineering task force, IETF）在自治网络体系中提出自治服务代理（autonomic service agent, ASA），用于在节点侧执行感知、分析和控制等功能[59]。同时，3GPP在网络管理与编排领域推进可管理AI/M能力框架及生命周期管理接口（如TS 28.105），为端边云协同提供了接口基础[60]。

结合3GPP对沉浸式通信、虚拟世界和AI赋能方向的讨论[8,39]，本文将AI Agent视为连接应用意图、网络能力与计算资源的协调实体，其作用主要体现在3个方面：在任务级通信组织中，AI Agent根据任务状态与交互阶段决定信息传输的时机和精度，如仅在决策关键时刻触发高精度全息数据传输，从而在资源受限条件下优先保障任务相关信息；在跨层意图表达与资源编排中，AI Agent将会议、训练等上层任务模式转化为切片与QoS约束，并驱动RAN和核心网侧资源调整；在交互体验支持中，AI Agent利用边缘侧推理能力辅助自然语言和多模态交互处理，降低端侧计算压力，并缓解链路负载波动对体验的影响。

然而，沉浸式业务状态空间通常高维且强时变，控制链路又跨越RAN调度、切片管理和多接入边缘计算（multi-access edge computing, MEC）渲染等多个环节，单靠固定规则往往难以在资源波动下稳定满足QoPE阈值要求。因此，引入AI Agent并非对现有平台的简单叠加，而是为跨域决策与不确定条件下的协同控制提供实现方式。尤其在应用侧与网络侧分别优化时，两类控制行为可能相互干扰，如应用侧为稳帧增大缓冲，而网络侧因资源竞争导致抖动尾部增大时，体验指标可能出现明显的波动。跨域协同的目标是在统一约束下协调体验策略与资源策略，减少二者相互抵消或放大的情况。如图6所示，本文将相关实现归纳为两类接口和一项保障机制：意图与体验约束接口，由沉浸式应用侧提供任务类型、交互强度、QoPE阈值及可退化

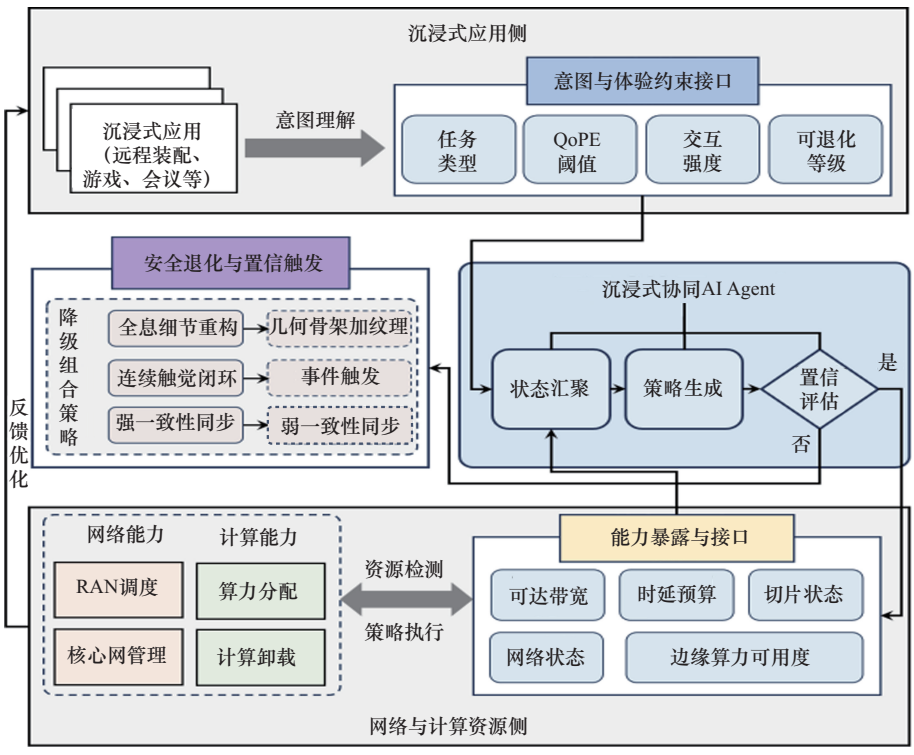

图6　沉浸式业务与网络协同的QoPE闭环与AI Agent接口

等级；网络能力与执行接口，由网络侧提供可达带宽、时延预算和边缘算力可用度，并反馈策略执行后的关键指标；降级与安全保障机制，当预测不确定性增大或尾部时延持续上升时，触发预设降级组合，如将全息细节重构降为几何骨架纹理表示、将强一致同步切换为弱一致同步模式，以维持服务连续性与体验下界。通过上述接口与机制，AI Agent可在体验约束与资源条件之间建立可执行的对应关系，使系统在极端网络条件下保持可预测的服务行为。

4 AI与沉浸式通信融合的挑战

前文给出了AI与沉浸式通信融合的框架，但应用实现中仍面临若干约束。沉浸式业务对资源和交互稳定性要求较高，部分指标已接近现有空口与协议机制的能力边界。同时，AI方法在推理时延、误差分布和可解释性方面的特征，并不能满足网络侧对可预测服务的需求。在频谱与能耗受限条件下，如何在吞吐、时延/抖动和可靠性之间实现可控权衡，仍是6G体系设计的关键问题。

4.1 上行容量、终端能耗与计算开销约束

与传统视频业务相比，沉浸式通信显著增加了上行容量。在6-DoF XR、体积视频和全息交互等场景中，终端需持续回传位姿、环境点云和多模态传感信息，上行由辅助链路转为容量瓶颈。在多视角或体积编码条件下，上行流量甚至可能接近或超过下行流量，限制端到端能力上限。同时，终端受能耗与热设计约束，感知、渲染和通信并发运行易造成持续热积累。计算卸载虽可降低端侧能耗，却可能增加上行占用并引入额外传输时延，使瓶颈在算力、带宽和时延之间迁移。任务导向的语义提取有助于减少回传数据，但在线特征提取与模型更新也会带来额外的计算和通信开销。总体而言，在不同场景与资源条件下，如何在上行容量、终端能耗和计算开销之间实现可调节的系统级折中，潜在方向包括：以语义或事件驱动减少冗余回传；端侧采用轻量特征提取，边缘侧进行重构或补全；结合能耗感知卸载与上行功率控制，实现通信与计算联合节能，并在RAN侧探索上行增强与非对称空口设计，以提升上行能力。

4.2 确定性时延约束与AI推理不确定性

沉浸式业务对端到端时延与抖动上界要求严格，MTP时延通常需控制在20 ms以内，更强交互形态还要求进一步压缩。一方面，在端边云协同渲染中，异步时间/空间扭曲可在一定程度上掩盖时延累积，但会增加系统复杂度，并提高对反馈精度与稳定性的依赖。另一方面，视点预测、资源调度等环节引入学习模型后，会带来推理开销与尾部时延。若缺少对最坏时延与异常工况的约束，模型的不确定性可能抵消协议层优化收益，并放大体验波动。因此，更关键的是把推理、调度与渲染统一纳入端到端时延预算，并与QoPE阈值联动。除平均时延外，还应显式给出抖动剩余裕量、同步剩余裕量和重传剩余裕量等指标。当剩余裕量小于一次重传、一次跨域迁移或一次大模型推理的预计开销时，系统应直接触发降级（如切换轻量模型、关闭非关键模态增强或降低重构等级），而不是继续等待额外补偿。在工程实现上，可将不确定性估计与降级触发结合，利用时效性网络（time-sensitive networking, TSN）的预算分解与队列控制，同时通过近端加速、模型压缩、算力预留与限次重传策略降低推理尾部时延。

4.3 空天地一体化下的链路异构与非平稳性

空天地一体化网络可借助非地面节点扩展覆盖，但也带来更强的链路异构性和时间非平稳性。非地面网络（non-terrestrial network, NTN）的传播时延、拓扑变化与强交互业务的低时延要求存在天然张力，即便采用预测与缓存，也难为高交互全息业务提供一致的实时性保障。更重要的是，地面蜂窝与NTN之间的跨域切换不仅是链路层问题，还会影响上层会话连续性，切换判决与执行滞后都可能导致体验断层。尽管轨迹预测、多连接和多路径传输可提升链路选择能力，但在多接入、多时延尺度并存的环境下，如何组织跨域切换、同步与状态一致性机制以维持稳定体验，仍是开放问题。在此背景下，更稳妥的体系设计是分层保障：地面侧优先承载强交互业务的确定性需求，NTN侧提供覆盖补充，并通过预测切换、前瞻缓存和多连接等机制减轻跨域切换对会话一致性的冲击；必要时再引入体验层降级，使跨域过程中的服务行为保持可预期。

4.4 语义歧义与鲁棒性问题

语义通信与AI方法有助于降低传输负载并支持任务导向处理，但也带来了语义歧义、分布偏移和可解释性不足等问题。多模态语义信息的度量与理论基础仍在发展，工程具体实现对数据分布变化

较为敏感，在高可靠场景下可能引发语义失真或误判。深度模型的黑盒特性还会增加互操作与标准化难度，并提高故障定位成本。考虑到沉浸式体验与人体感知直接相关，语义层不稳定更易被用户放大感知。因此，仍需建立可度量和可校准的语义质量指标与鲁棒性机制。围绕鲁棒性与可解释性开展研究：一是构建任务相关的语义失真度量，并引入分布外检测与鲁棒优化；二是通过语义空间对齐和元数据规范提升互操作性，并结合对抗防护降低恶意扰动引发的语义误判风险。

4.5　安全、隐私与信任

沉浸式通信与AI融合扩大了攻击面。沉浸式应用涉及生物特征、行为轨迹与环境空间信息，隐私风险高于传统移动互联网；语义特征虽然被压缩，但仍可能泄露高维行为信息；生成式模型带来内容伪造与篡改风险，同时模型本身也可能遭受对抗样本等攻击。因此，安全机制需要覆盖模型训练、语义传输与终端呈现等环节，不宜仅在单点加固，而应覆盖协同训练、语义传输与终端呈现等链路。现有研究较常采用的组合包括：协同训练阶段的安全聚合与差分隐私，执行阶段的可信执行环境与远程证明，以及内容侧的水印/指纹与可追溯治理，用于提升内容可信度与责任界定能力。

4.6　协议演进与跨层标准化

引入AI与语义处理后，部分功能需要跨层协同，这与传统分层解耦的协议设计思路存在张力。虽然ITU-R与3GPP已开展6G研究，但语义接口、AI能力调用与QoPE指标体系仍在讨论中。同时，算力投入与收益分配、模型治理与计费模式尚未明确，影响工程推进。在标准化推进上，更现实的路径是接口优先、渐进演进：先在不破坏现有协议栈的前提下明确语义元数据、模型参数交换以及AI控制与数据功能之间的必要接口，再通过跨组织协作与工程验证逐步固化约束与一致假设，从而降低生态碎片化风险。

5　未来发展方向与演进趋势

5.1　构建沉浸原生的6G网络架构

面向沉浸式业务，6G网络架构与评价体系需由“传输指标主导”扩展到“体验约束可表达、可约束、可验证”。除速率、时延和可靠性外，多模态同步、空间一致性和身体体验相关指标也将成为重要补充，并可能逐步纳入3GPP与IMT-2030的能力讨论框架。在机制层面，媒体编码、渲染处理和网络协议需具备更明确的协同接口，如面向沉浸式业务的跨层信息交互，支持基于终端姿态、场景复杂度和渲染状态等反馈的在线调整。图7将该类系统抽象为三层：体验/任务层，以QoPE、QoE和任务成功率为目标；语义/模型层，以语义表征、共享知识库和生成式模型为载体，支持“按任务传输”；算网控制层，以RAN、核心网和MEC为基础，通过AI控制面（artificial intelligence control plane, AICP）与AI数据面（artificial intelligence data plane, AIDP）提供可编排的策略执行和数据支

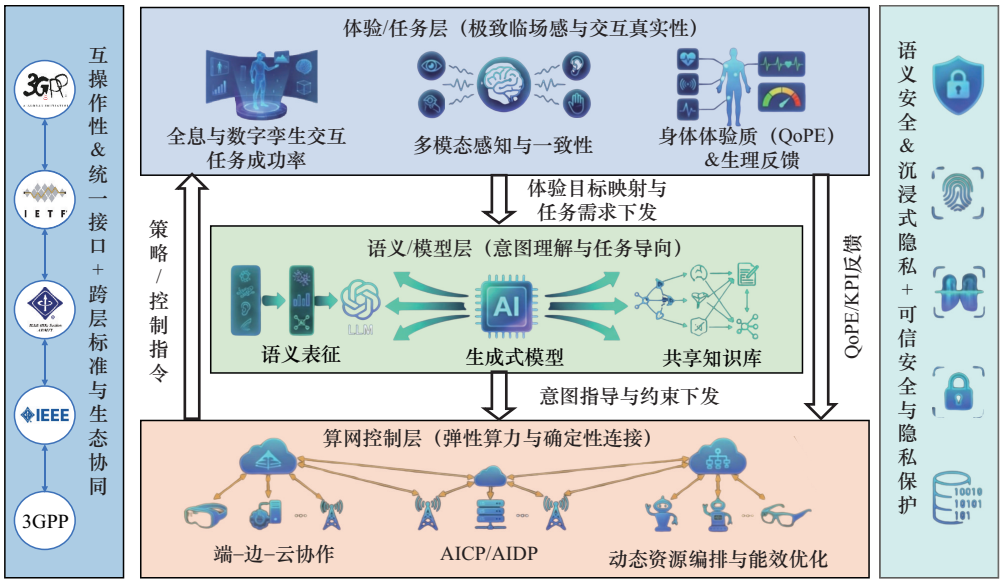

图7　沉浸式通信三层范式

撑。三层之间通过QoPE映射与反馈建立约束传递关系，使体验目标能够对应可实施的网络约束与策略。

5.2 面向沉浸式体验的AI-native网络演进

在沉浸式通信场景中，AI-native网络不仅用于局部预测，更需要覆盖状态感知、策略生成与跨域协调等功能。在架构上，可通过引入AICP承担全局状态汇聚、资源编排、语义相关决策与跨域协调。同时，由AIDP支持语义特征、模型参数与推理任务在网络侧的传输、缓存与调度，使模型与推理能力以“可调度资源”的形式进入网络能力集合。面向大规模会话的动态波动，核心网与边缘侧算网资源需要具备弹性伸缩与按需调度能力，以减少峰值配置带来的闲置，并降低拥塞状态下的失效风险。

5.3 语义通信与生成式AI的融合演进

沉浸式场景中的语义通信将与多模态大语言模型（multimodal large language model, MLLM）及生成式模型更紧密结合。MLLM可用于场景、动作与意图理解，支持语义特征的自适应抽取与任务导向编码。结合NeRF、3DGS与扩散模型等生成式技术，可在传输低维语义信息的同时，在边缘侧重建高保真内容，以降低比特传输负担。该路线能否工程化，依赖语义侧信息的规范化与可对齐：背景知识、模型版本与场景先验需要具备一致的描述与更新机制。对工业控制、医疗协作等高可靠场景，还需要可解释、可验证的语义度量体系，以界定“语义误差”对任务与体验的影响边界。

5.4 超分布式边缘智能与协同渲染体系

随着沉浸式业务向多终端、多屏与多主体协作发展，边缘侧将承担更多与业务相关的实时处理功能，而不仅是通用卸载节点。系统可采用分层明确的端边云协同：终端侧负责轻量渲染与局部更新，边缘侧承担主渲染、推理与场景管理，云端侧负责模型训练与全局同步。分布式场景带来的难点在于时空与物理一致性（如光照、遮挡与碰撞），需要结合物理约束建模与多视角融合算法，并在多人协作场景中提供场景同步、时延补偿与交互状态管理。考虑到边缘能耗与资源受限，分辨率、刷新率与编码参数的在线编排可用于在体验约束下控制能耗，使端边协同具备可持续运行能力。

5.5 多模态同步与QoPE指标体系

多用户协作与全息交互推动体验评价由QoS/QoE指标进一步扩展至更直接的体验约束。为此，有必要建立覆盖视听触同步、运动反馈一致性、多用户空间对齐及晕动症相关指标的量化体系，并结合生理信号、行为日志和网络参数，形成可复现的QoPE估计与校准方法。在此基础上，QoPE可作为编解码、资源调度和渲染参数调整的约束条件或反馈变量，用于维持体验下界及其一致性。随着XR与全息应用的规模化发展，QoPE有望与QoS和QoE并列成为重要评估指标，但其前提是采样频率、统计窗口、去噪与时滞补偿、分用户校准，以及由阈值反推KPI预算等口径逐步实现标准化。

5.6 网络–应用–智能体协同生态与标准化演进

AI与沉浸式通信的协同还将受到接口互操作和产业分工方式的共同影响。在标准化方面，IETF关于AI Agent协议的草案，以及3GPP研究报告中对6G用例与协同需求的讨论，均指向网络侧智能体能力的接口化与可互操作描述[8,33]。同时，全息与体积视频涉及编码、传输和渲染等链路，要求3GPP、MPEG、电气电子工程师学会和IETF等组织在接口、元数据和兼容性方面加强协同，以降低端到端互操作成本。在产业生态方面，算力平台、模型分发和会话接口需形成更清晰的协作边界，并同步完善沉浸式隐私与语义安全规范，以应对生物特征、行为轨迹和环境空间数据带来的安全风险。

6 结束语

本文围绕6G沉浸式通信与AI-native网络协同开展综述。首先，基于标准化进展与代表性研究，本文提出了QoPE-KPI双维层级化需求框架，给出了QoPE的形式化定义，并补充了QoPE可测信号的采样频率与统计窗口、噪声及时滞处理、个体差异校准，以及由QoPE阈值反推时延、抖动、同步误差和丢包/重传策略预算的方法，为将体验约束纳入网络设计提供更具可操作性的指标表述。进一步，本文从内容处理、传输、计算和控制4个环节梳理了关键技术与研究进展，涵盖语义通信与生成式重构、AI-native资源管理、端边云协同渲染、多模态同步与QoPE反馈机制，以及AI Agent协同交互等方向，并归纳尚待解决的问题。最后，本文讨论了工程实现中的主要约束，包括上行受限、确定性时延与模型不确定性、空天地异构、语义鲁棒

性、隐私安全和标准化协同等，并据此提出了值得关注的技术与标准化议题。本文旨在为后续QoPE导向方法的评估与对比提供更明确的问题边界、测量口径和预算分解思路，并为系统设计与标准化讨论中的AI-native能力引入和跨层接口定义提供参考。

参考文献：

[1] Saad W, Bennis M, Chen M Z. A vision of 6G wireless systems: applications, trends, technologies, and open research problems[J]. IEEE Network, 2020, 34(3): 134-142.

[2] Giordani M, Polese M, Mezzavilla M, et al. Toward 6G networks: use cases and technologies[J]. IEEE Communications Magazine, 2020, 58(3): 55-61.

[3] Strinati E C, Barbarossa S. 6G networks: beyond Shannon towards semantic and goal-oriented communications[J]. Computer Networks, 2021, 190: 107930.

[4] Cui Q M, You X H, Wei N, et al. Overview of AI and communication for 6G network: fundamentals, challenges, and future research opportunities[J]. Science China Information Sciences, 2025, 68(7): 171301.

[5] Zhou C H, Hu S S, Gao J, et al. User-centric immersive communications in 6G: a data-oriented framework via digital twin[J]. IEEE Wireless Communications, 2025, 32(3): 122-129.

[6] Shen X M, Gao J, Li M S, et al. Toward immersive communications in 6G[J]. Frontiers in Computer Science, 2023, 4: 1068478.

[7] Ogenyi F C, Ugwu C N, Ugwu O P. A comprehensive review of AI-native 6G: integrating semantic communications, reconfigurable intelligent surfaces, and edge intelligence for next-generation connectivity[J]. Frontiers in Communications and Networks, 2025, 6: 1655410.

[8] 3GPP. TR 22.870 V0.5.0 (Release 20): study on 6G use cases and service requirements[R]. [2026-01-15].

[9] 3GPP. TR 38.914 V0.3.0 (Release 20): study on 6G scenarios and requirements[R]. [2026-01-15].

[10] Yan X Q, Wang J F, Zhang X R, et al. Data plane design for AI-native 6G networks[J]. Huawei Research, 2024(7): 73-82.

[11] Wang X Y, Lu L, Li Q, et al. A task-driven design approach for 6G AI-native architecture[J]. Engineering, 2026, 56: 87-103.

[12] Zhang P, Shi G M, Cui S G, et al. Semantic communications: theories, technologies and applications[J]. China Communications, 2024, 21(7): iii-vii.

[13] 刘传宏, 郭彩丽, 杨洋, 等. 人工智能物联网中面向智能任务的语义通信方法[J]. 通信学报, 2021, 42(11): 97-108.
Liu C H, Guo C L, Yang Y, et al. Intelligent task-oriented semantic communication method in artificial intelligence of things[J]. Journal on Communications, 2021, 42(11): 97-108.

[14] Liu Y T, Wang X J, Ning Z L, et al. A survey on semantic communications: technologies, solutions, applications and challenges[J]. Digital Communications and Networks, 2024, 10(3): 528-545.

[15] 曾焕强, 孔庆玮, 陈婧, 等. 沉浸式视频编码技术综述[J]. 电子与信息学报, 2024, 46(2): 602-614.
Zeng H Q, Kong Q W, Chen J, et al. Overview of immersive video coding[J]. Journal of Electronics & Information Technology, 2024, 46(2): 602-614.

[16] Next G Alliance. Multi-sensory extended reality (XR) in 6G[R]. Next G Alliance White Paper, 2023-03-10.

[17] 侯文军, 白冰, 杨本植. 6G网络下的全息技术发展及业务趋势[J]. 电信科学, 2021, 37(11): 1-10.
Hou W J, Bai B, Yang B Z. Research on the development and business trend of holographic technology under 6G network[J]. Telecommunications Science, 2021, 37(11): 1-10.

[18] Zhang H J, Zhang Z Y, Liu X N, et al. Integrated sensing and communication for 6G holographic digital twins[J]. IEEE Wireless Communications, 2025, 32(2): 104-112.

[19] Sullivan G J, Ohm J R, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.

[20] Chung J M. Content delivery network (CDN) technology[M]//Emerging Metaverse XR and Video Multimedia Technologies: Modern Streaming and Multimedia Systems and Applications. Berkeley, CA: Apress, 2023: : 251-277.

[21] Subhan F E, Yaqoob A, Muntean C H, et al. A survey on artificial intelligence techniques for improved rich media content delivery in a 5G and beyond network slicing context[J]. IEEE Communications Surveys & Tutorials, 2025, 27(2): 1427-1487.

[22] Ngo H Q, Interdonato G, Larsson E G, et al. Ultradense cell-free massive MIMO for 6G: technical overview and open questions[J]. Proceedings of the IEEE, 2024, 112(7): 805-831.

[23] Islam N, Shin S. Deep learning in physical layer: review on data driven end-to-end communication systems and their enabling semantic applications[J]. IEEE Open Journal of the Communications Society, 2024, 5: 4207-4240.

[24] 李彦, 万征. 深度强化学习在边缘视频传输优化中的应用综述[J]. 计算机工程与应用, 2025, 61(4): 43-58.
Li Y, Wan Z. Survey on applications of deep reinforcement learning in edge video transmission optimization[J]. Computer Engineering and Applications, 2025, 61(4): 43-58.

[25] Pohl S, Becher A, Grauschopf T, et al. Neural network 3D body pose tracking and prediction for motion-to-photon latency compensation in distributed virtual reality[C]//Artificial Neural Networks and Machine Learning-ICANN 2019: Image Processing. Berlin: Springer, 2019: 429-442.

[26] 周一青, 孙布勒, 齐彦丽, 等. 面向通信与计算融合的5G移动增强/虚拟现实[J]. 电信科学, 2018, 34(8): 19-33.
Zhou Y Q, Sun B L, Qi Y L, et al. Mobile AR/VR in 5G based on convergence of communication and computing[J]. Telecommunications Science, 2018, 34(8): 19-33.

[27] Gül S, Ahsan S, Paris S, et al. PDU set-based QoS handling in 3GPP: release 18 overview and future directions[J]. IEEE Communications Standards Magazine, 2025, 9(3): 104-110.

[28] Multrus M, Bruhn S, Torres J, et al. Immersive voice and audio ser-

vices (IVAS) codec: the new 3GPP standard for immersive communication[C]//Proceedings of the AES 157th Convention. 2024:1-13.

[29] 元辉, 丁丹丹, 张伟, 等. 三维点云编码通信进展[J]. 中国图象图形学报, 2026, doi: cjig.cn/zh/article/146832480/.
Yuan H, Ding D D, Zhang W, et al. Research progress of 3D point cloud coding and communication[J]. Journal of Image and Graphics, 2026, doi: cjig.cn/zh/article/146832480/.

[30] Liu H, Yuan H, Liu Q, et al. A comprehensive study and comparison of core technologies for MPEG 3-D point cloud compression[J]. IEEE Transactions on Broadcasting, 2020, 66(3): 701-717.

[31] 赵亚军, 郁光辉, 徐汉青. 6G移动通信网络: 愿景、挑战与关键技术[J]. 中国科学: 信息科学, 2019, 49(8): 963-987.
Zhao Y J, Yu G H, Xu H Q. 6G mobile communication networks: vision, challenges, and key technologies[J]. Scientia Sinica (Informationis), 2019, 49(8): 963-987.

[32] Hamidouche W, Bariah L, Debbah M. Immersive media and massive twinning: advancing toward the metaverse[J]. IEEE Communications Magazine, 2024, 62(7): 20-32.

[33] Stephan E, Schott R, Lopez D, et al. AI agent protocols for 6G systems [EB]. (2025-10-20)[2026-01-15].

[34] 张平, 牛凯, 姚圣时, 等. 面向未来的语义通信: 基本原理与实现方法[J]. 通信学报, 2023, 44(5): 1-14.
Zhang P, Niu K, Yao S S, et al. Semantic communications for future: basic principle and implementation methodology[J]. Journal on Communications, 2023, 44(5): 1-14.

[35] Dong C, Liang H T, Xu X D, et al. Semantic communication system based on semantic slice models propagation[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(1): 202-213.

[36] Ali S, Saad W, Rajatheva N, et al. 6G white paper on machine learning in wireless communication networks[PP]. V1. (2020-04-28)[2026-01-15]. arXiv: arXiv.2004.13875.

[37] Naderializadeh N, Sydir J J, Simsek M, et al. Resource management in wireless networks via multi-agent deep reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2021, 20(6): 3507-3523.

[38] ITU-R. Recommendation ITU-R M.2160-0 (11/2023): framework and overall objectives of the future development of IMT for 2030 and beyond[R]. International Telecommunication Union, 2023-11-05.

[39] Letaief K B, Chen W, Shi Y M, et al. The roadmap to 6G: AI empowered wireless networks[J]. IEEE Communications Magazine, 2019, 57(8): 84-90.

[40] Zhang C Y, Patras P, Haddadi H. Deep learning in mobile and wireless networking: a survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(3): 2224-2287.

[41] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: representing scenes as neural radiance fields for view synthesis[J]. Communications of the ACM, 2022, 65(1): 99-106.

[42] Kerbl B, Kopanas G, Leimkuehler T, et al. 3D Gaussian splatting for real-time radiance field rendering[J]. ACM Transactions on Graphics, 2023, 42(4): 1-14.

[43] Niu K, Dai J C, Yao S S, et al. A paradigm shift toward semantic communications[J]. IEEE Communications Magazine, 2022, 60(11): 113-119.

[44] Zhang P, Xu W J, Gao H, et al. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks[J]. Engineering, 2022, 8: 60-73.

[45] Niknam S, Dhillon H S, Reed J H. Federated learning for wireless communications: motivation, opportunities, and challenges[J]. IEEE Communications Magazine, 2020, 58(6): 46-51.

[46] Luong N C, Hoang D T, Gong S M, et al. Applications of deep reinforcement learning in communications and networking: a survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3133-3174.

[47] Chao F Y, Ozcinar C, Smolic A. Transformer-based long-term viewport prediction in 360° video: scanpath is all you need[C]//Proceedings of the 2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP). Piscataway: IEEE Press, 2021: 1-6.

[48] Xu X, Tan X B, Wang S Y, et al. Multi-features fusion based viewport prediction with GNN for 360-degree video streaming[C]//Proceedings of the 2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom). Piscataway: IEEE Press, 2023: 57-64.

[49] 李可, 马赛, 戴朋林, 等. 基于多目标深度强化学习的车车通信无线资源分配算法[J]. 计算机研究与发展, 2024, 61(9): 2229-2245.
Li K, Ma S, Dai P L, et al. Wireless resource allocation algorithm based on multi-objective deep reinforcement learning for vehicle-to-vehicle communications[J]. Journal of Computer Research and Development, 2024, 61(9): 2229-2245.

[50] Shah S D A, Nezami Z, Hafeez M, et al. The interplay of AI-and-RAN: dynamic resource allocation for converged 6G platform[PP]. V1. (2025-03-10)[2026-01-15]. arXiv: arXiv.2503.07420.

[51] Li N, Sun Q, Wang L H, et al. Towards AI-native RAN: an operator's perspective of 6G day 1 standardization[PP]. V1. (2025-07-11)[2026-01-15]. arXiv: arXiv.2507.08403.

[52] 李英群, 胡啸, 徐翔, 等. 三维场景注视点渲染深度学习方法综述[J]. 中国图象图形学报, 2024, 29(10): 2955-2978.
Li Y Q, Hu X, Xu X, et al. Deep learning-based foveated rendering in 3D space: a review[J]. Journal of Image and Graphics, 2024, 29(10): 2955-2978.

[53] Hu X, Wu X L, Ma M C, et al. Efficient VR rendering: survey on foveated, stereo, cloud, and low-power rendering techniques[J]. Virtual Reality & Intelligent Hardware, 2025, 7(5): 421-452.

[54] 邝祝芳, 陈清林, 李林峰, 等. 基于深度强化学习的多用户边缘计算任务卸载调度与资源分配算法[J]. 计算机学报, 2022, 45(4): 812-824.
Kuang Z F, Chen Q L, Li L F, et al. Multi-user edge computing task offloading scheduling and resource allocation based on deep reinforcement learning[J]. Chinese Journal of Computers, 2022, 45(4): 812-824.

[55] Liu X, Chai Z Y, Cheng Y Y, et al. Evolutionary multi-objective deep reinforcement learning for task offloading in industrial Internet of things[J]. IEEE Transactions on Network and Service Management, 2025, 22(5): 5074-5089.

[56] Ran X K, Slocum C, Tsai Y Z, et al. Multi-user augmented reality with communication efficient and spatially consistent virtual objects[C]//Proceedings of the 16th International Conference on Emerging Net-

working Experiments and Technologies. New York: ACM Press, 2020: 386-398.

[57] Chen J S, Ramakrishnan K K, Dhakazl A, et al. Networked architectures for localization-based multi-user augmented reality[J]. IEEE Communications Magazine, 2023, 61(12): 104-110.

[58] Liu M T, Yang B H, Xu M D, et al. Exploring quantitative assessment of cybersickness in virtual reality using EEG signals and a CNN-ECA-LSTM network[J]. Displays, 2024, 81: 102602.

[59] Behringer M H, Pritikin M, Bjarnason S, et al. Autonomic networking: definitions and design goals[EB]. (2015-06-01)[2026-01-15].

[60] ETSI. ETSI TS 128 105 V18.4.0: 5G; Management and orchestration; Artificial intelligence/machine learning (AI/ML) management (3GPP TS 28.105 version 18.4.0 release 18)[S]. (2024-07-30)[2026-01-15].

[作者简介]



黄亚坤（1992– ），男，博士，北京邮电大学副教授，主要研究方向为沉浸式通信、3D多媒体通信、智能AR/VR服务。



苏洋（1991– ），男，北京邮电大学博士生，主要研究方向为沉浸式通信、多媒体通信。



王春生（1979– ），男，中电万维信息技术有限责任公司高级工程师，主要研究方向为空间计算、沉浸式通信、点云流和AR/VR智能服务。



陆平（1971– ），男，博士，移动网络和移动多媒体技术国家重点实验室教授级高级工程师，主要研究方向为云计算、大数据、增强现实和基于多媒体服务的技术。



谢大雄（1963– ），男，博士，移动网络和移动多媒体技术国家重点实验室教授级高级工程师，主要研究方向为云计算、大数据、增强现实和基于多媒体服务的技术。



乔秀全（1978– ），男，博士，北京邮电大学教授、博士生导师，主要研究方向为6G沉浸式通信、元宇宙、空间计算和增强/虚拟现实技术。