中文核心期刊　中国科技核心期刊　中国核心学术期刊
第三届国家期刊奖百种重点期刊　信息通信领域产学研合作特色期刊

ISSN 1009-6868
CN 34-1228/TN

中兴通讯技术

ZTE TECHNOLOGY JOURNAL

http://tech.zte.com.cn

第 30 卷 · 2024 年 7 月 · 增刊 1

专题：网媒融合

（封面图片详解见 P88）

ISSN 1009-6868
9 771009 686243 99>

《中兴通讯技术》第 9 届编辑委员会成员名单

目　次

中兴通讯技术（ZHONGXING TONGXUN JISHU）
第 30 卷　总第 178 期　2024 年 7 月　增刊 1

中文核心期刊 中国科技核心期刊 第三届国家期刊奖百种重点期刊 信息通信领域产学研合作特色期刊 中国知网、万方数据、重庆维普等数据库收录期刊 1995 年创刊

热点专题 ▶

网媒融合

企业视界 ▶

《中兴通讯技术》2024 年热点专题名称及策划人

1. 下一代多址技术
北京交通大学教授 艾渤
北京交通大学教授 陈为

2. 网络大模型
中国电信 IP 领域首席专家 唐宏
中兴通讯无线首席架构师 熊先奎

3. 6G 多天线技术
东南大学首席教授 金石
北京交通大学教授 章嘉懿
东南大学副研究员 韩瑜

4. 6G 无线系统技术
中国信息通信研究院副院长 王志勤
中国移动研究院院长 黄宇红
东南大学教授 王东明

5. 卫星通信技术
哈尔滨工业大学（深圳）教授 张钦宇

6. 数据通信新技术
中国电信研究院教授级高工 解冲锋
中国联通研究院首席科学家 唐雄燕

CONTENTS

ZTE TECHNOLOGY JOURNAL
Jul. 2024 Vol. 30 S1

Special Topic ▶

Synergy of Network and Media

Enterprise View ▶

期刊基本参数：CN 34-1228/TN*1995*b*16*90*zh*P*¥20.00*6500*12*2024-07

网媒融合专题导读

专题策划人

谢大雄

丁文华

沉浸式通信是IMT-2030（6G）列出的6G重要应用场景之一，已成为科技和产业革命争夺的制高点。沉浸式应用包含以数字孪生工厂、虚实融合生活、多模态大模型等为代表的新兴媒体服务，具有强交互、沉浸式、智能化的特点，这些特点要求网络及媒体服务需要提供超高通信能力、超强算力、智能算法处理能力。在经典的网媒分离范式下，通信、计算、算法三大能力存在瓶颈，难以支撑沉浸式通信发展。因此，移动网络和移动多媒体技术国家重点实验室提出网媒融合这一沉浸式通信下的变革性技术，旨在突破“管道化”传统思维，集成“通感智算存”统一调度能力，研究网络传输与媒体产生、呈现的融合，实现全局动态最优，最终跨越三大性能瓶颈。网媒融合体系架构具有3个方面特点：一是网媒底层能力的融合，即网媒通信、感知、智能、计算、存储能力统一度量与调度；二是网媒上层架构的融合，即新兴媒体服务和网络资源实时双向感知与匹配，人工智能（AI）赋能网媒间内容的生成、传输与呈现；三是网媒间贯穿信息服务全生命周期的一体化优化设计，实现全局动态最优。

网媒融合研究当前面临诸多挑战。网媒融合基础理论如何突破？网媒融合一体化架构如何设计？网媒融合的新兴媒体内容生产，以及网络感知、编码、计算、传输的关键技术如何突破？网媒融合的示范应用如何落地？为此，本期以网媒融合为主题，共收录了11篇文章，针对网媒融合中的关键技术开展讨论。

本期开篇是丁文华院士团队撰写的《元宇宙初探：概念内涵、技术体系及发展建议》，从元宇宙的必备要素、技术支撑、应用场景3个方面进行深入阐述分析，并从技术突破、规则制定、产业布局、人才培养等方面对元宇宙的未来发展提出相关建议。《面向边缘智能的通信计算一体化研究》一文围绕通信网与算力网的网媒融合，对信道衰落和噪声可能会带来的聚合失真等问题开展了研究。《语义编码与经典信道编码融合研究》对基于联合信源信道编码的语义通信系统进行了理论分析，研究了后续将语义通信应用于经典通信框架的基础方法。《人工智能驱动的跨模态语义通信系统》提出了基于人工智能的跨模态语义通信系统架构、核心思想、关键技术、实践应用以及存在的挑战。《具身智能机器人技术》提出了一种智能制造中的具身智能机器人技术，介绍了网媒融合下的一种应用场景。《用于混合现实的三维场景生成技术》介绍了近年来三维场景生成的各项技术方法，以及混合现实场景下三维场景生成的现状，并对其发展趋势进行了分析与展望。《基于流式路径追踪的实时真实感渲染技术》从图形处理器（GPU）的线程调度和内存访问两个角度出发，提出了一种基于流式路径追踪的实时真实感渲染方

DOI：10.12142/ZTETJ.2024S1001
收稿日期：2024-06-18

案。《基于深度生成模型的视觉模式表示与编码》介绍了概念图像编码、概念视频编码、跨模态语义编码等生成式编码方法，总结了各智能视频编码技术的发展趋势与挑战。《从2B到4B——电信行业与垂直行业的供需协同倍增发展》从无线专网、5G 2B（To Business）的发展趋势与挑战出发，提出向4B（For Business）转变来重塑垂直行业5G发展体系。《3D IC系统架构概览》介绍了网媒融合芯片设计中，芯片3D架构在性能、功耗等方面的优势，分析了3D架构在物理实现、封装测试、工艺能力等方面的挑战。《XR网业协同技术》提出了网媒融合下面向扩展现实（XR）业务的网业协同技术，需要构建基于XR业务的低时延、大带宽和高可靠的广义确定性网络。

本期作者主要来自对网媒融合领域有深入研究的知名高校、企业，从关键理论、技术挑战等方面介绍了网媒融合最新研究成果。希望本期内容能为读者提供有益的启示与帮助，在此对所有作者的大力支持和审稿专家的辛勤指导表示由衷的感谢！

策 划 人 简 介

谢大雄，中兴通讯股份有限公司监事长、移动网络和移动多媒体技术国家重点实验室主任，教授级高工，中国发明协会会士、国家级领军人才，享受国务院特殊津贴，2023年担任工业和信息化部通信科技委员会副主任，是《国家中长期科学和技术发展规划纲要（2006–2020年）》“新一代宽带无线移动通信网”重大专项论证委员会委员、国家“973计划”和“863计划”项目带头人；2002年、2010年先后获得国家科技进步奖二等奖2项，2017年获得国家技术发明奖1项，2002年获得首届深圳市市长奖。

丁文华，中国工程院信息与电子工程学部院士，现任深圳大学电子与信息工程学院院长，曾任中央电视台总工程师，中国电视台网络制播领域的技术创新带头人；主要从事广播电视技术、多媒体信息处理和计算机网络工程应用研究；曾获得国家科技进步奖一等奖1项、省部级科技进步奖突出贡献奖2项、省部级科技进步奖一等奖13项，2007年成为中国首位被亚广联（ABU）授予“亚太地区广播工业杰出贡献奖”的技术专家，并荣获“何梁何利基金科学与技术创新奖”“国家有突出贡献中青年专家”“王选科学技术杰出人才奖”等荣誉称号。

元宇宙初探：概念内涵、技术体系及发展建议

Metaverse: Concept, Architecture, and Suggestions

冯大权/FENG Daquan[1,2]，张胜利/ZHANG Shengli[1]，吕星月/LYU Xingyue[1,2]，王振中/WANG Zhenzhong[3]
（1. 深圳大学电子与信息工程学院，中国 深圳 518060；
2. 深圳大学数字创意研究中心，中国 深圳 518060；
3. 中央广播电视总台技术局，中国 北京 100038）
(1. College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China;
2. Digital Creative Research Center, Shenzhen University, Shenzhen 518060, China;
3. Technical Management Center, China Media Group, Beijing 100038, China)

DOI：10.12142/ZTETJ.2024S1002
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20240724.1017.002.html
网络出版日期：2024-07-24
收稿日期：2023-12-05

摘要：人工智能、虚拟现实、数字孪生等数字技术的快速发展，使得人类的生活空间不再仅仅局限于现实物理世界，以数字内容、数据算法构成的数字虚拟空间逐渐成为人们生活中不可或缺的一部分。“Metaverse（元宇宙）”被认为是未来虚实相融的终极形态，但由于概念转译问题以及缺乏广泛的学术共识，目前“元宇宙”依然是众说纷纭的状态，公众对这一概念难以形成准确清晰的认识。基于此，通过梳理数字生成技术的发展进程，从中国数字化进程和国情出发提出“复合宇宙”概念，从而促进相关部门、从业者、用户更好地理解这一新兴技术，引导产业的健康发展。在阐述复合宇宙的概念内涵的基础上，对元宇宙的必备要素、技术支撑、应用场景这3个方面进行深入阐述分析。最后，对元宇宙的未来发展提出建议，包括技术突破、规则制定、产业布局、人才培养等方面。

关键词：元宇宙；复合宇宙；技术体系；发展建议

Abstract: With the rapid development of digital technologies such as artificial intelligence, virtual reality, and digital twins, human living space is no longer limited to the real physical world. The digital virtual space composed of digital content and data algorithms has gradually become an indispensable part of people´s lives. The Metaverse is considered the ultimate form of the fusion of virtuality and reality in the future, but due to issues with conceptual translation and a lack of broad academic consensus, the Metaverse is still in a state of divergent opinions, making it difficult for the public to form an accurate and clear understanding of this concept. The development process of digital generation technology is reviewed and the concept of "complex universe" is proposed from the perspective of China´s digitalization process and national conditions, in order to promote relevant departments, practitioners, and users to better understand this emerging technology and guide the healthy development of the industry. Based on the concept and connotation of complex universe, an in-depth analysis of essential elements, technical support, and application scenarios of the metaverse is provided. Finally, relevant suggestions are put forward for the future development of Metaverse, including technological breakthroughs, rule formulation, industrial layout, and talent cultivation.

Keywords: Metaverse; complex universe; technology system; development suggestion

引用格式：冯大权, 张胜利, 吕星月, 等. 元宇宙初探：概念内涵、技术体系及发展建议 [J]. 中兴通讯技术, 2024, 30(S1): 3-15. DOI: 10.12142/ZTETJ.2024S1002
Citation: FENG D Q, ZHANG S L, LYU X Y, et al. Metaverse: concept, architecture, and suggestions [J]. ZTE technology journal, 2024, 30(S1): 3-15. DOI: 10.12142/ZTETJ.2024S1002

1 元宇宙的技术演进与认知挑战

随着数字技术的迅速发展，同时伴随着几何物体、自然场景、工业场景等研究在计算机图像领域中的不断推进，到20世纪末21世纪初，利用计算机图形技术构建现实中难以实现的场景即数字场景技术，在影视数字内容制作[1-2]、模拟飞行训练[3-4]、城市场景三维重建[5-6]等领域得到广泛应用。

步入21世纪后，数字生成技术越发成熟，以数字场景技术及其应用为起点和基础，陆续衍生发展出虚拟现实（VR）、增强现实（AR）、混合现实（MR）、数字孪生等技术及相应产品，如表1所示。其中，VR技术面向纯粹的虚拟世界，具体是指利用360°拍摄，通过头戴式显示器

（HMD）、VR Glass等终端设备为用户提供包括视觉、听觉、体感等在内的全方位沉浸式主观体验。与VR不同，AR强调在真实世界中增加虚拟信息，实现数字化的文字、图像、视频、三维模型等叠加入现实世界，形成增强型主观体验。从AR开始，数字生成技术不仅面向纯粹虚拟，而且不断追求虚实共融和连通[7]。1994年，文献[8]总结了VR和AR的关系，并进一步提出了MR概念。MR是虚拟现实技术的进一步发展，其基于光场技术，将虚拟现实与增强现实融合到一起，在呈现效果上使虚拟对象看起来是现实世界的一部分，在学界[8]和业界[9]都陆续获得了关注。最近，扩展现实（XR）概念又被提出，它由VR、AR、MR共同构成，将数字对象作为现实世界的代理并受之驱动，力图将现实世界与虚拟世界融为一体并相互作用。

数字孪生是虚拟世界对现实世界的完整镜像，强调数字虚拟空间中映射真实世界的物体。数字孪生包含物理产品、虚拟产品及二者之间的联系，在2010年由美国国家航空航天局（NASA）正式提出。德国的“工业4.0”则进一步推动了数字孪生的发展[10]。中国学者指出，数字孪生是综合运用感知、计算、建模等信息技术，对真实物理空间进行多物理量、多尺度、多概率的仿真并在虚拟空间中完成映射[11]，认为其将给制造业带来革命性变化，并广泛服务于智能设计、智能加工、智能装配和智能服务[12]。

当前，人工智能、虚拟现实、云计算、区块链、5G、产业互联网以及数字孪生等数字技术，加速推进数字生成技术、数字内容创作方式、数字场景互动模式发生变革，并不断将虚拟世界和现实世界的融合发展推升至新的高度。近两年，曾经在科幻小说《Snow Crash》中描绘的人们可以通过各自虚拟分身（Avatar）进行语言、行为等交互行为的虚拟世界[13]，开始被各方力量推动从而实现从概念到现实的飞跃。这一虚拟世界被原著作者STEPHENSON命名为“Metaverse”，中国通常翻译为“元宇宙”。2021年以来，众多美国高科技巨头企业包括Roblox、Meta、英伟达、微软等，纷纷推出自己的元宇宙发展计划，在全球范围掀起了元宇宙产业发展的浪潮。

元宇宙的发展热潮迅速在全球范围内攀升，并从最初的社交、游戏、娱乐领域，逐渐拓展至公共治理、军事、文化等多个领域。2021年11月，首尔市政府制定了全球第一个中长期元宇宙政策文件，宣布建立“元宇宙首尔”（Meta-verse Seoul）平台[14]。该平台计划打造虚拟市长办公室、虚拟旅游区等，用于改善市政管理、优化公共治理。同年12月，日本多家相关公司成立元宇宙的业界团体“日本元宇宙协会”（The Japan Metaverse Association），该协会通过组织多种形式的交流活动促进元宇宙技术和相关服务的普及，提升相关商业环境和用户保护体制的健全性[15]。2022年6月，美国国防初创公司Red 6发布基于AR技术创建的机载战术增强现实系统（ATARS），该系统使美国战斗机飞行员能够与包括中国和俄罗斯战机飞行员在内的虚拟对手进行空中格斗练习[16]。Red 6宣称ATARS系统的目标是要创建一个“军事元宇宙”（Military Metaverse），通过这一系统将世界各地的作战人员连接起来，使他们能够实时训练。

随着5G、大数据、人工智能、区块链等数字技术的兴起，互联网应用拓展至人们生活及工作的方方面面，工业、交通、金融、文旅、教育、医疗等传统产业与互联网不断融合，数字经济正快速成为继农业经济、工业经济之后主要的经济形态。2022年1月，中共中央国务院印发了《“十四五”数字经济发展规划》（以下简称《规划》）。《规划》从优化升级数字基础设施、充分发挥数据要素作用、大力推进产业数字化、持续提升公共服务数字化水平、健全完善数字经济治理体系、着力强化数字经济安全体系、有效拓展数字经济国际合作8个方面对中国“十四五”时期数字经济发展做出了总体的部署[17]。2022年10月，工业和信息化部、教育部、文化和旅游部、国家广播电视总局、国家体育总局等五部门联合发布《虚拟现实与行业应用融合发展行动计划（2022—2026年）》，强调虚拟现实是新一代信息技术的重要前沿方向，是数字经济的重大前瞻领域，将深刻改变人类的生产生活方式。

2021年后元宇宙的热度持续攀升，中国一些省市陆续出台针对元宇宙相关技术产业的重要战略规划。2021年12

▼表1 数字生成技术相关概念的演进历程

概念	定义
数字场景	利用数字技术和计算机图形技术构建现实中难以实现的场景，根据真人和场景的互动关系再将真人嵌入数字场景中
虚拟现实	利用360°拍摄，通过HMD、VR Glass等终端设备为用户提供包括视觉、听觉、体感等在内的全方位沉浸式主观体验
增强现实	在真实世界中增加虚拟信息，实现数字化的文字、图像、视频、三维模型等叠加入现实世界，形成增强型主观体验
混合现实	基于光场技术实现，融合了增强现实与虚拟现实，使虚拟对象看起来是现实世界的一部分，将虚拟对象与现实世界融为一体
数字孪生	在数字虚拟空间中映射真实世界的物体，并通过数据方式对虚拟对象驱动，从而形成客观信息镜像模型

HMD：头盔显示器 VR：虚拟现实

月印发的《上海市电子信息制造业发展“十四五”规划》中提到，上海要前瞻部署量子计算、第三代半导体、6G通信和元宇宙等领域[18]。2022年1月，浙江省数字经济发展领导小组办公室发布的《关于浙江省未来产业先导区建设的指导意见》中明确，元宇宙与人工智能、区块链、第三代半导体等产业并列，是浙江省到2023年重点未来产业先导区的布局领域之一[19]。2022年4月，《广州市黄埔区、广州开发区促进元宇宙创新发展办法》发布，政策涵盖技术创新、应用示范、知识产权保护等10个方面，明确工业元宇宙、数字虚拟人、数字艺术品交易等体现元宇宙发展趋势的重点培养领域[20]。2022年8月，北京市通州区人民政府与市科委等机构联合发布了《北京城市副中心元宇宙创新发展行动计划（2022—2024年）》的通知，力争在未来3年将城市副中心打造成为以文旅内容为特色的元宇宙应用示范区，打造元宇宙生态链企业，建成典型应用场景项目，制定相关标准，同时形成元宇宙与文化、旅游、商业、城市服务等各领域虚实融合发展“1+*N*”产业空间体系，以助力北京建设数字经济标杆城市[21]。

与此同时，相关专业机构、院校、行业组织等也开始对元宇宙进行布局。2022年6月，“十四五”数字孪生黄河建设方案通过水利部审查，拟构建与物理黄河同步仿真运行、虚实交互、迭代优化的自主产权数字仿真平台[22]。2022年8月，中关村互联网教育创新中心等单位发布《元宇宙教育共识》，阐述了通过元宇宙来革新未来学习理念、学习环境，赋能未来学习与教育新范式，提高教育生产力[23]。2022年9月，世界人工智能大会举办“AI医疗与元宇宙论坛”举办，提出利用元宇宙技术来构建多元医疗场景和多重服务来重塑下一代数字医疗[24]。

中国高科技企业方面，字节跳动通过抖音、今日头条、飞书等产品，建立起了全球化的内容流量入口，同时通过入股VR硬件厂商Pico，打通了“设备+内容+平台”的生态闭环。腾讯以投资、合作等形式在VR、AR硬件开发方面持续布局，2012年收购元宇宙世界标杆企业Epic Games 40%的股份，2020年2月参投“元宇宙第一股”Roblox，2022年6月提出“超级数字场景”[25]概念并成立XR部门。腾讯不断在视频、影视、文学、音乐等泛文娱领域延伸内容产品布局，形成了“社交+内容+娱乐”的初步元宇宙版图[26]。此外，许多其他中国公司也陆续投入到相关建设实践中去，如百度“希壤”元宇宙项目、华为云虚拟数字人“云笙”、阿里巴巴云游戏服务品牌“元境”、网易伏羲旗下沉浸式活动平台“瑶台”等。

然而，目前对Metaverse概念的讨论仍处于众说纷纭的现状，而且关于未来它能够呈现出的形态，似乎没人能给出一个明确的描述。表2整理了目前国际知名企业及组织对Metaverse的定义。其中，Roblox从游戏社区的角度将元宇宙定义为永久在线、人人共享的3D虚拟游戏空间；Meta从社交平台的角度将元宇宙界定为一个聚焦于社交连结的3D虚拟世界网络，是融合虚拟现实技术并用专属的硬件设备打造一个具有超强沉浸感的社交平台；NVIDIA则从算力技术平台的角度推出Omniverse（工业元宇宙）平台旨在打造“工程师的元宇宙”；Epic从渲染引擎的角度将元宇宙视为一场前所未有的大规模参与的实时3D媒介；微软从办公平台的角度强调元宇宙帮助人们在数字化环境中相聚并进行舒适的协作。能够看到不同机构所设想的Metaverse空间性质是由该机构的功能性质或服务对象所决定，例如Meta、Epic、微软、英伟达分别探索构建虚实共生的社交空间、经济空间、工作空间以及生产空间。此外，在名称转译方面，虽然中国网络界、信息技术人士和新闻与传播学界将Metaverse通常译成元宇宙，但也有学者进行词源分析后质疑这一中文译名的准确性[27]。

▼表2 国际知名企业或组织对Metaverse概念认知

企业或组织	定义
Roblox（罗布乐思游戏社区）	Metaverse是一个将所有人关联起来的永久性的共享的3D虚拟世界，每个人都拥有自己的数字身份 Metaverse具备了身份、朋友、沉浸感、低延迟、多样性、随地性、经济系统、文明[28]
Meta（Facebook于2021年10月28日更名为Meta）	Metaverse是一组虚拟空间，在这里你可以与不同物理空间的其他人一起创建和探索，开展工作、娱乐、学习、购物、创作等活动[29]
NVIDIA（英伟达人工智能计算公司）	Metaverse是一个或多个通过Omniverse的连接来实现的共享的虚拟世界，具有交互性、沉浸性和协作性[30]
Epic	元宇宙将是一场前所未有的大规模参与的实时3D媒介，带有公平的经济系统，所有创作者可以参与、赚钱并获得奖励
Microsoft（微软公司）	Metaverse使跨越物理世界和数字世界的共享体验得以实现，帮助人们在数字化环境中相聚，更舒适地使用化身，并促进来自世界各地的创造性合作[31]

总的来看，由于概念转译不统一、学术共识缺乏等问题，目前元宇宙依然处于众说纷纭的状态，未形成统一概念。这同时也导致公众对该词的理解缺乏整体清晰的认识。元宇宙的未来设想是打造一个虚实融合的世界，这一虚实融合的世界包含4个层面的内容，即现实世界、模拟现实世界的虚拟世界、构建创新的虚拟世界、虚实融合的世界[32]。为促进新一代信息技术的融合创新发展和应用落地，引导相关产业的健康发展，帮助政策制定者和相关行业的从业者了解并参与领域建设。本文从中国数字化发展进程和国情出发，提出更便于理解的概念“复合宇宙”[1]，翻译为“Complex Universe”，混合词为“Comverse”。本文通过这一概念来解释元宇宙，进而对元宇宙的必备要素、技术体系、应用场景作出系统性的阐述。

2 复合宇宙：元宇宙的内涵解读

2.1 概念特征

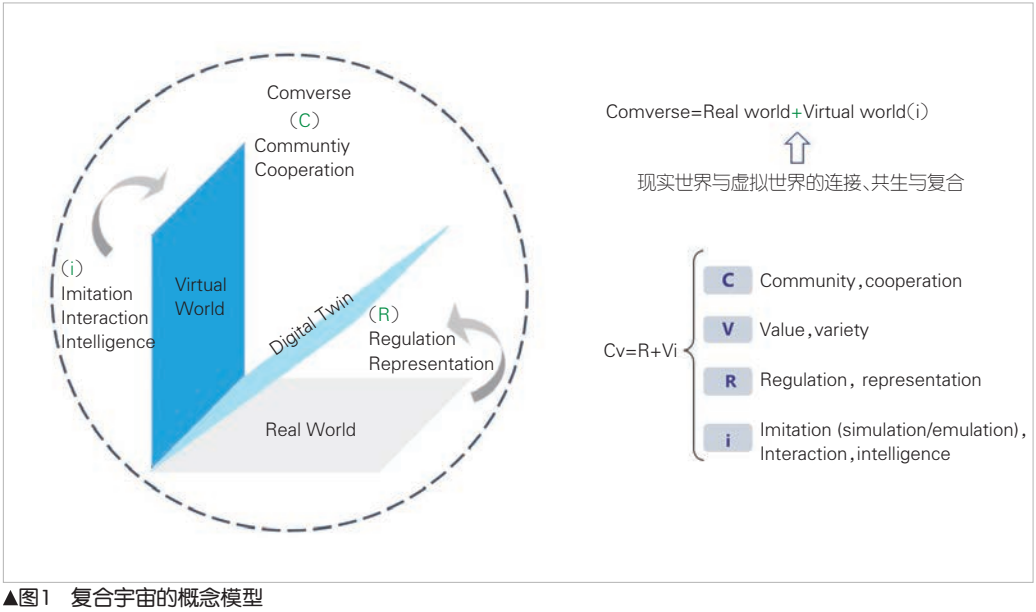
虚拟世界和现实世界能够通过介质（数字载体）实现相互转换、相互影响以及相互拓展。未来虚拟世界和现实世界不是平行的关系，而是正交的关系。复合宇宙是对元宇宙内涵的一种解读和拓展，它更加全面、深入地描述了虚拟世界与现实世界融合的终极形态，为理解和探索未来世界提供了新的视角和思路。复合宇宙本质是虚拟、现实两个世界复合而成，因此借助复数虚实坐标这种形象的表达来描述复合宇宙虚实并存的形态特征，如图1所示。具体来说，复合宇宙（Comverse）是由“Com”和“Verse”两个单词组成，缩写为“Cv”。其中，“Com”表示复合的“Complex”，Verse指宇宙“Universe”。复合宇宙概念公式“Cv = R + Vi”中，“R”表示现实世界“Real world”，具备物质实体；“V”表示虚拟世界“Virtual world”，具备数字内容与数据编码；i代表一系列技术，包括Imitation（Simulation/Emulation）、Interaction、Intelligence等；“+”符号强调复合宇宙虚实世界的连接、超越、创新等复合关系的全域性特征。此外，复合宇宙还具备以下属性：

1）社区（Community）与社会性。复合宇宙可以视为各类实体及其相关关系的集合，底层结构是主体参与、互动以及身体感官浸入，在虚拟空间中互动并形成现实社区、仿现实虚拟社区、超现实虚拟社区。虚拟社区既包括工作环境、学习环境、生活环境、娱乐环境等数字，又包括成员们的数字化身。

2）协作（Cooperation）与共享。在复合宇宙社区中，用户、场景、工具、功能之间实现连接与交互，实现多方共建与互通共享。任何现实中的人都可通过数字工具实现在虚拟世界中的身份构建，继而参与到协同网络之中，在不同领域的数字环境中实现创作、交易、社交、娱乐、收益等功能。

3）规则（Regulation）与有序性。复合宇宙通过制定运行规则，处理人与人（真实人与数字人、数字人与数字人的对照关系及伦理关系等）、人与物（时间关系、空间关系等）、物与物（现实物品与虚拟物品、虚拟物品与虚拟物品间的映射关系及价值关系等）之间的关系问题，以维持可控性与稳定性。其中，人与人之间的规则包括数字人身份管理与建立限制、虚拟社区伦理道德与行为规范，人与物之间的规则体现于主体意识对存在本体的映射与转换，物与物之间的规则体现在所有权申明（数字资产）、等价交换原则与一般等价物（数字货币）确立。

4）呈现（Representation）与表达。传统数字内



▲图1 复合宇宙的概念模型

1 本文根据丁文华院士2021年11月提出的复合宇宙概念整理成文。

容呈现方式主要通过照片、视频等方式进行视听方面的呈现，本质上是将内容记录之后进行情景再现。目前，数字内容呈现方式主要通过4K/8K超高清、VR技术实现逼真还原，视听呈现内容在感知延迟、清晰度、自由度、观看体验方面都得到了极大提升。在复合宇宙时代，虚拟与现实的边界将逐渐模糊，数字内容生成方式将会迎来革新。届时同一对象将具备更多类型的呈现形式，包括VR（面向纯粹虚拟世界）、AR（以现实世界为主提供增强信息）、MR（以现实世界为主或以虚拟世界为主）、Digital Twin（虚拟世界对现实世界的完整镜像）、EVR（以虚拟世界为主，数字对象作为现实世界的代理并受之驱动）。

5）仿效（Imitation）模拟及仿真。复合宇宙具备仿真的拟态空间和沉浸的体验感。在游戏电竞与娱乐影视领域，复合宇宙中可具备数字人仿真性、参与上的交互性、VR沉浸式体验、多用户社交、个人数字形象可定制等特点。在工业生产领域，复合宇宙提供现实世界的数字化映射及驱动数据，可以在数字世界中搭建真实工厂的数字场景，进行工业化生产的模拟、规划、设计及测试等工作，实现在与真实工厂环境一样的数字孪生工厂中的生产作业。在医疗健康领域，复合宇宙中可形成人类身体及器官的数字仿生体，并与真人关联同步，辅助健康管理和疾病看护，实现对身体数据的实时诊断。

6）交互与相互作用（Interaction）。复合宇宙是虚拟世界与现实世界融合的终极形态，必然会回到现实社会产生反作用并形成推动现实社会发展与演化的内生动力和外部条件。复合宇宙是虚实共生的，物理世界与数字世界是相互构造的。人们可以随时随地切换身份穿梭于虚拟世界和现实世界，在虚拟时空节点中工作、学习、娱乐，交易所形成的数字产品。

7）智能化（Intelligence）。人工智能为复合宇宙中的数字内容创作和数字人创建提供基础理论保障和前沿技术支撑，赋能应用场景扩建与用户内容生产。围绕2D+/3D高精数字人建模、驱动，渲染及评测方法等关键技术，通过在传统制作流程中全面融入AI技术，预期在数字人模型和材质自动化生成、基于语音的表情和嘴唇动画、数字人实时渲染、数字环境引入、数字人与数字环境融合等方面获得关键技术突破，极大提升数字人的逼真度和制造效率。智能化的目标之一是使虚拟数字人达到具备高感知能力且能够实现自我认知甚至进化的智能驱动。

8）价值（Value）。复合宇宙具有价值属性，是处理人与人、人与物、物与物之间经济关系的具体表征。复合宇宙的价值属性具体表现于类似于现实世界的经济系统和货币交易体系，包括数字市场、数字产品、数字消费、数字货币、数字资产等要素。基于区块链所形成的价值互联网，未来复合宇宙平台可实现从信息互联到价值互联、从用户生成内容到AI生成内容等转变[33]。用户既是消费者、体验者，又是生产者、创作者。这将提升用户的创作热情和参与感，推动建立全新的知识经济模式。

9）Variety（多样性）。虚拟世界具有超越现实的多样性，表现在多重的内容生产形式、传播形式、呈现方式，以及广泛的用户群体、丰富的应用场景和新业态。各行各业的用户都能在复合宇宙社区中创造内容，因此平台产生内容的方式将由专业生产内容（PGC）向用户生产内容（UGC）、人工智能生成内容（AIGC）转变，内容创作也将具备多种形态。未来复合宇宙在实现最广泛用户连接的同时，基于庞大而多样的用户群体，新的消费需求将不断被开发，催发出新业态、新场景、新服务。复合宇宙是一个动态演进的复合体，其边界将不断扩大。当社会数字化程度深化时，复合宇宙的边界和囊括范围将不断扩大。

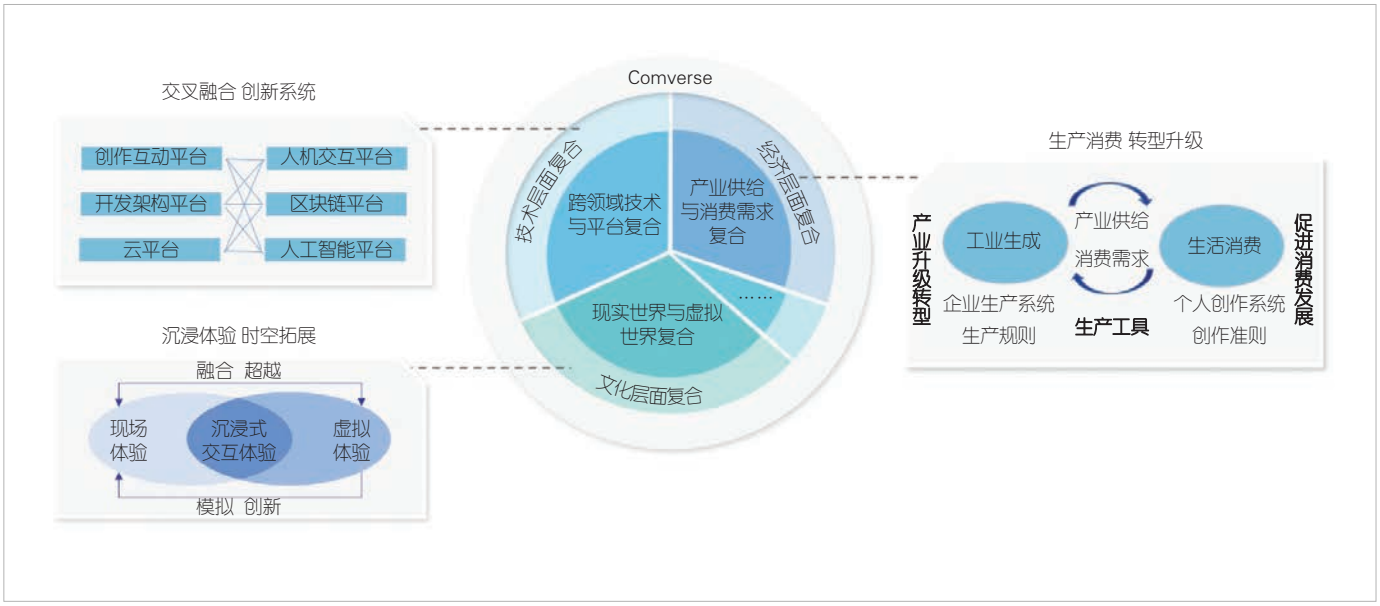
以上复合宇宙九大属性可归纳为“Cv = R + Vi”的形象表达，如图1所示，其中C代表Community、Cooperation等，V代表Value、Variety，R代表Regulation、Representation，i代表Imitation（Simulation/Emulation）、Interaction、Intelligence等。

2.2 多层次复合

复合宇宙虚实融合的生态，催生出人们在虚拟世界的第二身份、全新生活方式等新的社会维度，为人们提供另一维度下的全新生活。虚拟世界具备和现实世界的相同特征，包括身份、娱乐社交、商业、社会治理等，以及对虚拟世界的感知能力。除了总体虚实层面的融合，复合宇宙中的“复合”也体现在技术、经济、文化等各个方面，如图2所示。

一是技术层面的复合。复合宇宙是整合了多种技术而产生的虚实相融的社会形态。技术层面的复合是指跨领域技术与平台的复合。通过技术层面的复合搭建需求为导向的高效平台，如云平台、人工智能平台、人机交互平台、数字人生产平台等。通过各种技术和平台的交叉融合，一方面可以解决虚拟世界内容生产问题，满足复合宇宙快速扩充需要，另一方面可以改变行业生产方式，降低进入门槛，从而颠覆产业格局。

二是经济层面的复合。经济层面的复合是指产业供给与消费需求的复合。供需层面的复合可促进产业升级转型和消费发展，畅通供需大循环。在消费端，复合宇宙通过整合各项数字技术，给我们带来全新的社交、游戏、购物体验，催



▲图2 复合宇宙在技术、供需、虚实层面的复合

生新服务、新业态。在产业供给端，复合宇宙为工业生产带来了高效的技术支撑和生产平台，推动供应链领域数实融生的生产范式革新以及全链条的优化范式革新。例如，在智能制造领域，宝马集团借助英伟达的Omniverse平台，首次实现了整座工厂生产线的仿真，通过数字化工厂提升了灵活性和精确性，总体实现了30%的效率提升。

三是文化层面的复合。文化层面是以虚拟世界和现实世界的文化活动为内容载体。例如，演出、旅游、电竞、体育赛事、图书馆、博物馆等，通过沉浸式交互体验实现在欣赏、观看、阅读情景中现实体验和虚拟体验的复合，达到了虚实融合、时空拓展的文化传播与共享的目的。未来复合宇宙时代的数字内容生成与呈现方式将引来革新，实现虚拟体验与现场体验深度融合，进而通过虚拟世界与现实世界互作用机制更新规则、应用、产品，影响至人们的社交模式、教育模式、娱乐方式、创作与传播手段等社会生活文化领域各方面。

3 元宇宙的必备要素

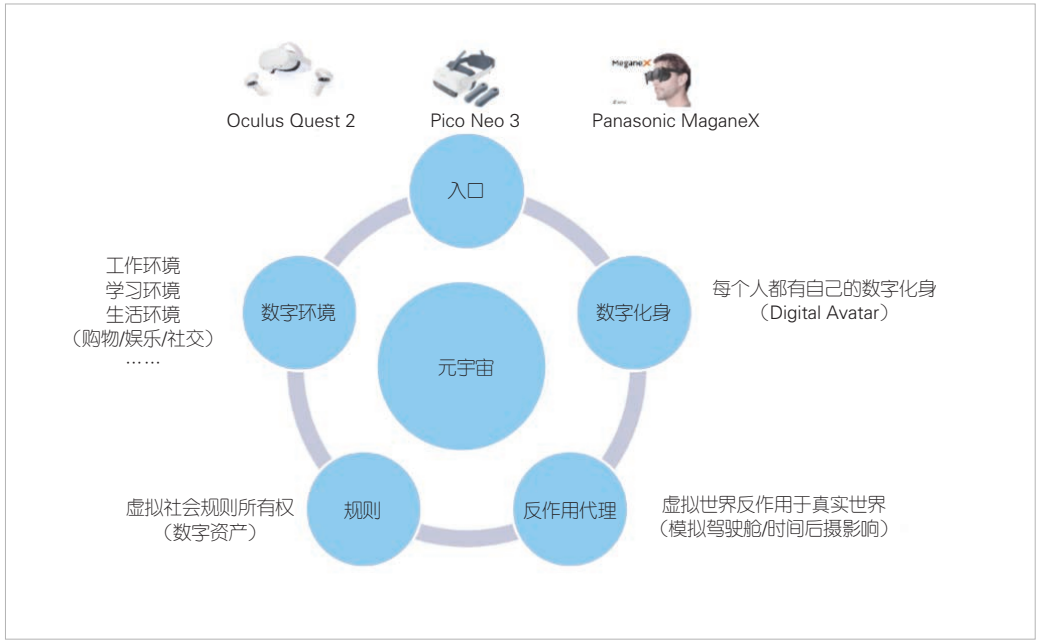
元宇宙是虚拟世界和现实世界融合的终极形态。虚拟世界可以是根据现实世界模拟出来的，也可以是创新出来的虚拟人、物、环境等，即在现实世界中没有与之对应的实际人、物、环境。元宇宙的虚实融合包含现实世界、模拟现实世界的虚拟世界、构建创新的虚拟世界、虚实融合的世界4个层面。要实现元宇宙的虚实融合需要具备5个必备要素：虚拟世界入口、数字环境、数字化身和虚拟世界规则、反作用代理，如图3所示。

3.1 虚拟世界入口

用于人机交互的硬件设备是元宇宙的入口。目前比较成熟的技术有VR和AR，而脑机接口则是一个未来开拓方向。脑机接口通过采集分析人类大脑产生的脑信号来直接操控外部设备，有助于提高人类与外界交互的效率，同时提供味觉、嗅觉、触觉等多感官体验。现在的应用大多通过用户的使用习惯或者大数据库来进行识别或提供，一旦脑机接口成熟后，用户可以通过意念来传输对外界的信息交互。未来这将是在元宇宙交互中不可或缺的一环。需要说明的是，在未来人机交互需要多种交互显示技术相互结合才能让用户获得更好的体验。

3.2 数字环境

数字环境是构建元宇宙的基础架构，它通过数字孪生、AI等计算机技术模拟物理世界的自然环境，或者创造物理世界不存在的虚拟环境。目前已有一些应用案例，例如孪生城市、数字工厂等。孪生城市通过可视化决策系统将城市各部门海量信息资源进行整合共享，覆盖城市管理的重点关注领域，通过数据可视化构建一系列业务决策模型，实现对当前状态的评估、对已发生问题的诊断和对未来趋势的预判，提高城市运营管理水平。数字工厂通过数据可视化对工业厂房、生产线、设备等管理要素进行三维仿真展示，集成视频监控、设备运行监测及其他传感器设备实时上传的监测数据，对生产流程、设备运行状态进行实时监测，真实再现生产流程、设备运转过程等，为设备的研制、改进、定型、维护和效能评估等，提供有效、准确的决策依据。



▲图3　元宇宙五大必备要素

3.3 数字人

数字人是元宇宙的主要活动体，人们可通过数字身份自由地参与到未来元宇宙的运行中，人们可以在其中进行学习、工作、创造、娱乐、社交、交易等各种活动。通过3D建模、渲染、人工智能、NPL等技术创建一个数字化的虚拟人，使其具备社交、互动以及记忆等属性。

3.4 虚拟世界规则

伴随着技术发展和应用场景的不断成熟，未来元宇宙将演化成为一个超大规模、极致开放、动态优化的复杂系统。这一系统将由多个领域的建设者共同构建，涵盖了网络空间、硬件终端、各类厂商和广大用户，保障虚拟现实应用场景的广泛连接，并展现为超大型数字应用生态的外在形式。元宇宙的本质是达到虚实融合的世界体系。由于其开放性和复杂性，现有物理世界的规则难以满足未来元宇宙中各种生态应用的需求，因此需要在目前真实物理世界规则的基础上再构建一套科学高效的数字规则体系，以保证元宇宙的可持续发展。

3.5 反作用代理

元宇宙是虚实共生的，现实世界与虚拟世界是共同构造的，其中必然存在相互作用机制。元宇宙中两个世界的互作用表现为：不是现实世界单向作用于虚拟世界，而是现实世界中的事物可以在虚拟世界中找到映照的同时，虚拟事物同样可以反作用于现实世界并产生现实影响。在虚拟时空节点中工作、训练、学习、娱乐、交易所形成的数字产品、模拟结果，一部分还会传回现实世界进而对现实世界产生影响。例如，服务机器人可以看成是虚拟数字人在现实世界的反向代理，模拟飞行器和模拟生产甚至能够发挥时间的后摄作用，以达到虚拟世界对现实世界的实质影响，避免或降低一些灾难性事件的影响。

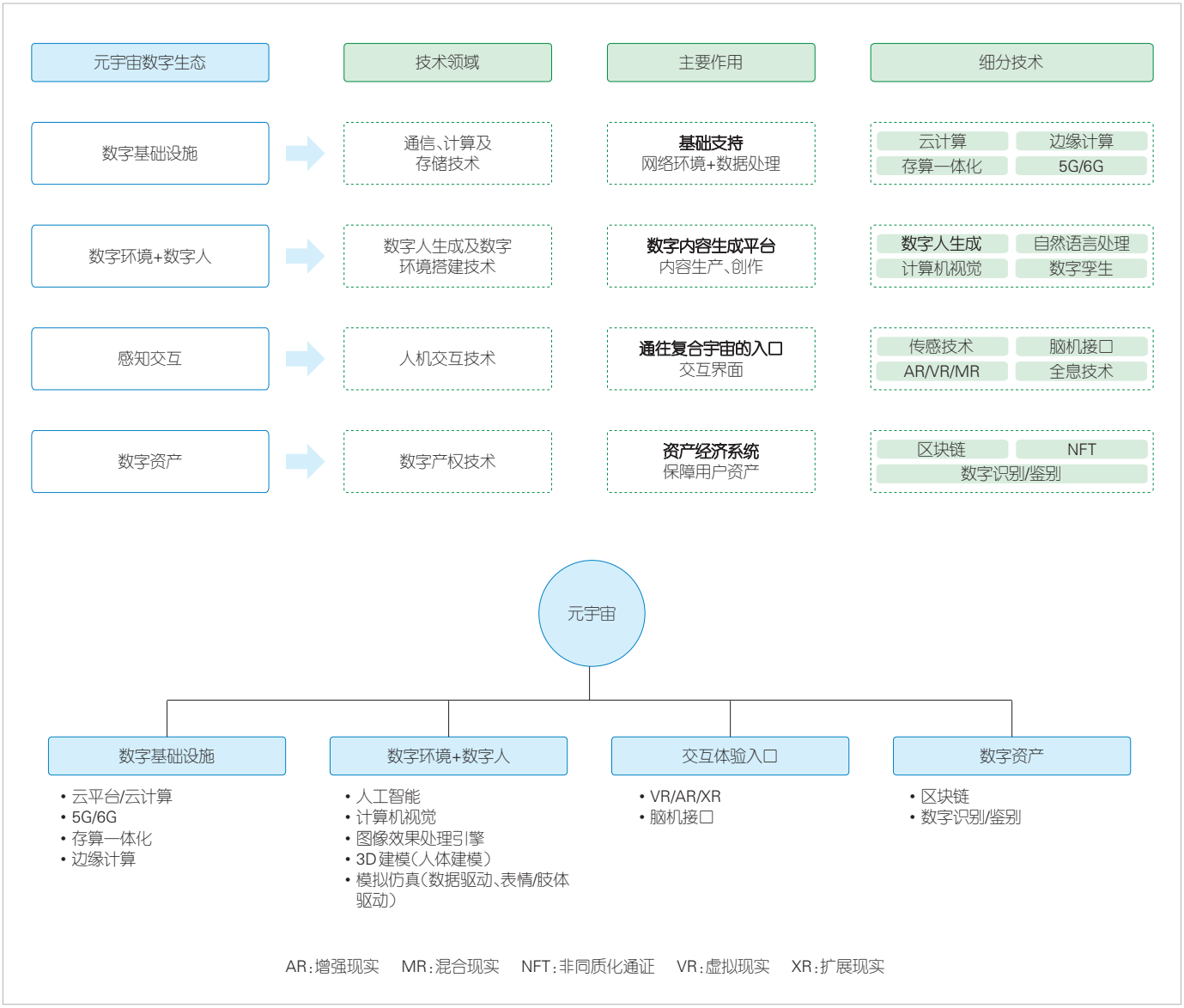
4 元宇宙支撑技术

打造元宇宙生态，在数字环境中建立媲美于真实世界的虚拟世界，需要依托下一代移动网络提供全新的与之相适配的高性能支撑，解决数字对象和数字环境内容在生产、传输和交互各层面的关键技术问题，实现高效高质的数字对象与数字环境的适配和驱动。此外，还需要制定适用于元宇宙的规则体系并建立规则引擎，确保虚实世界有效融合、符合伦理、有序运行。下面我们将围绕构建元宇宙的数字生态体系，介绍相关关键基础支撑技术，如图4所示。

4.1 数字基础设施

元宇宙中的虚拟世界并不是完全“虚拟的”，而是基于人的智力与体力劳动构建的数字对象和数字环境，是运行在底层各种物理资源之上形成的用于生产生活、可呈现可参与的信息复合体。人–人、人–物和人–信息之间的全新交互方式将会带来爆炸性的数据量。这些新场景和新需求的实现都离不开强大的通信支撑。动态环境建模、实时动作捕捉、实时3D图形生成、多元数据处理、实时定位跟踪等数字环境的实现都需要海量的算力支持。因此，包括通信和算力等基础设施的数字基建是构建元宇宙的基础，支撑着元宇宙内部的泛在连接和高效配置。通信基建支撑元宇宙运行，覆盖了5G/6G移动通信、光纤通信、Wi–Fi 6、工业互联网等多种类型系统设施，为元宇宙的互联和交互提供保障。算力基建作为元宇宙数字化进程中的算力底座，依托数据中心、智能计算中心等实现大数据、云计算、边缘计算、存算一体等能力支撑。

未来元宇宙不仅支持个人沉浸式娱乐消费型应用，还将



▲图4 元宇宙关键支撑技术

广泛应用于交通、工业、医疗、国防等生产型应用。不同应用场景下，各业务的可靠性、速率、延迟、隐私保护、移动性管理等服务需求都可能有很大的不同。因此，如何利用有限的网络资源，通过弹性灵活的资源管理调配，为不同业务的差异化需求提供按需的服务并提高元宇宙的经济价值，将是元宇宙生态建设中面临的一个重要挑战。未来需要结合计算集群管理、网络切片、分布式数据存储等技术，充分调动各方算力、带宽、存储资源，建立有确定性网络服务能力的资源管理体系，提供差异化的业务体验，满足多维度的服务需求。根据不同业务的服务质量需求，设计与之相适应的高效接入控制与资源匹配机制，并针对传输信道时变、用户业务请求动态变化来设计智能化的通信、存储、计算资源配置机制，实现资源与业务的灵活、智能、动态匹配，从而为元宇宙的生态应用提供可靠的差异化服务。

4.2 数字环境与数字人

元宇宙需要通过提供海量的数字人、数字孪生体等数字对象来构建沉浸式的虚拟世界，同时需要建立合理、可操作、可执行的规则体系，实现物理世界与虚拟数字世界之间的实时交互。然而，目前数字环境、数字人等数字内容生成技术存在生成周期长、效率低、成本高、进入门槛高等痛点。现有的人工智能模型通常比较大，需要大量的算力，这对资源受限的移动设备是一大挑战。因此，未来需要使用高效生产工具（如3D建模引擎、渲染引擎等）解决虚拟世界

中人与物的快速生成问题，降低普通用户的参入门槛。此外，目前人工智能模型算法大都是一个“黑盒子”，缺乏可解释性，元宇宙中的开发人员、虚拟世界设计师和用户都无法理解人工智能的决策过程。因此，需要借助可解释性人工智能，提高用户的信任度、模型可审核性和操作效率，同时降低法律风险和安全威胁，保证用户的可靠体验。

与此同时，元宇宙虚实融合的新型社会形态，其多样性、复杂性远超现有真实世界。现有的规则体系无法支撑元宇宙的健康持续发展。需要研发通用的数字规则引擎来建立合理的可操作可执行的相应业务规则体系，推动自动化工作流程并形成自我进化机制。

4.3 感知交互

感知交互技术是物理世界与数字世界进行交流互动的桥梁。未来，在感知交互技术的支持下，我们与虚拟世界的交流互动将更加高效便捷。目前，元宇宙的终端呈现大多依赖VR/AR/XR等入口技术，对于用户和数字世界的交互问题，多是采用定位、语音、手柄等单一交互方式，力度、温度、味道等其他感官的交互体验尚需完善，需要结合不同维度的感知和接口技术提出更具便捷性、普适性的交互方法，并实现多要素融合的高级情感交互，在虚拟世界中实现与真实世界等同的交互体验。此外，对于脑机接口技术，目前还处于早期研究阶段，脑信号处理的原理机制、用户的感觉调控还需要进一步明确。另外，一些安全问题需要亟待解决，例如：脑机接口植入人脑时可能带来的健康风险、使用脑机接口可能带来的隐私信息泄露问题、外界通过脑机接口对使用者的大脑进行干涉产生的安全问题等。

4.4 数字资产

元宇宙中虚拟世界都是基于数字技术完成的。所有的数字对象、数字环境，以及他们的状态变化等都以数据比特的形式记录于系统设备上，因此，需要建立安全、高效的数字资产管理系统，来充分调动各方智力、算力、带宽、存储资源，并提供良好的利益分配机制，解决传统互联网产业里贡献者利益分配不合理、网络资源利用不充分等弊端。

元宇宙中数字资产可分为虚拟原生资产和虚实共生资产。虚拟原生是虚拟世界自身创造或者经济循环而产生的资产，例如虚拟世界工具、虚拟人以及游戏中售卖的道具等。虚实共生是虚拟世界和现实世界映射而产生的数字资产，是物理世界实物资产的数字化、三维化体现，例如虚拟住宅、办公楼、生活生产工具等。构建一套支撑元宇宙运作的数字资产系统离不开多种技术的融合，相关技术主要包括区块链技术和数字身份识别技术。

区块链是支撑元宇宙经济体系的技术基础。去中心化思想的引入，赋予了元宇宙从根本上颠覆当前现实世界中社会生产关系和协作方式的潜力。非同质化通证（NFT）、去中心化自治组织（DAO）、智能合约、DeFi等区块链技术，将激发创作经济时代的发展，催生海量的创新内容。区块链技术将有效打造元宇宙去中心化的结算平台和价值传递机制，保障价值归属与流转，实现元宇宙经济系统运行的稳定、高效、透明和确定性。

数字身份可以将真实信息以数字代码的方式展现，从而对个体进行可识别刻画，以便对这个真实身份持有者的实时行为信息进行绑定、查询和验证。在元宇宙中，数字身份识别技术可以将物理身份与数字身份相匹配。数字身份识别的广泛应用可有效保障隐私数据安全，降低客户接入成本，提高互联网的经济价值。

5 元宇宙的应用场景

从全球市场增长来看，元宇宙产业具有广阔的市场价值。根据Insider在2022年7月发布的数据，2024—2026年全球元宇宙的市场预测约在2 400亿～8 000亿美元之间。2022年11月，工业和信息化部、教育部、文化和旅游部、国家广播电视总局和国家体育总局联合发布《虚拟现实与行业应用融合发展行动计划（2022—2026年）》（简称为《行动计划》），明确到2026年，中国虚拟现实产业总体规模（含相关硬件、软件、应用等）将超过3 500亿元，虚拟现实终端销量将超过2 500万台，建成10个产业公共服务平台。《行动计划》中提出的10类“虚拟现实+”规模化应用试点，具有市场需求、政策支持和产业基础等落地优势。总体来看，元宇宙的应用场景归纳为To C端（直接面向个体消费者提供相关的产品服务）和To B端（面向企业或特定用户群体提供相关的产品服务）两类。To C端的应用包括文化旅游、演艺娱乐、体育健康、教育培训、残障辅助等；To B端的应用包括工业生产、融合媒体、商贸创意、智慧城市、安全应急等。

5.1 文化旅游

元宇宙时代的文化旅游将提供生动的线上游览与观赏体验。数字与文化在元宇宙时代的融合创新，实现了文化形态从以内容化传播到互动化体验的升级，在元宇宙中构建的虚实联动的场景，让文化内容展现得更丰富，使文化传播与共享更便捷。一方面，借助地理信息技术、虚拟现实技术、人机交互技术，打造数字IP景区空间（如虚拟公园、线上文

化展馆、数字街区等）和沉浸体验产品（如旅行规划体验、线上导游导览、文物古迹复原等），让用户在身临其境的同时可以与虚拟空间中的角色和展品进行互动；另一方面，元宇宙中进行数字“云游”能够使景区景点、文化藏品、展会作品获得传统线下展览无法相比拟的曝光度，让优秀传统文化和旅游资源得到更广泛的传播和推广。

5.2 演艺娱乐

现阶段的娱乐方式大多为旅游、影视、游戏等，在元宇宙时代这些娱乐场景将采用互动平台、全息投影、体感交互等技术打造沉浸体验。元宇宙的应用将发展成三维层面上的应用，不再拘泥于现有的手机、PC上的二维平面式的呈现关系。在AR、VR设备帮助下搭建常态化的虚拟现实线上演播摄制播出环境、新型互动表演区，观众可以进入数字环境中体验，既可以以客观视角观看，也可以以主观视角参与。未来游戏体验也会朝着更加沉浸式方向发展。通过可穿戴的全身触觉传感衣，游戏中的摔倒、受伤可以通过网络传导到触觉感受器上，让我们有类似的切身感受。游戏平台也正在由传统的“游戏+社交”向外扩展，融入文化、娱乐、商业等多重要素。

5.3 体育健康

聚焦“大体育，大健康”发展需求，围绕日常健身、专业训练、运动休闲等不同情景，在元宇宙时代，不论是体育用品、运动设施，还是健身软件、健康管理平台，都将实现虚拟现实终端及内容兼容适配，打造线上线下相结合的数字化、智能化、沉浸化的新型体育运动解决方案。元宇宙将提供治疗方案可视化平台，用于病情诊断、手术辅助、病人沟通等多个方面，有效提高医疗效率。同时，元宇宙可形成人类身体及器官的数字仿生体，并与真人关联同步，通过对虚拟身体数据的模拟分析，得到最佳的医疗与健身方案。在医疗资源紧张的环境下，虚拟世界的智能医生将发挥重要作用。对于极高难度手术，患者的身体数据完全仿真，医生可在虚拟场景中进行模拟练习。这不仅保证医疗安全，还能够降低医疗成本。

5.4 教育培训

元宇宙技术能够通过对场景空间的再造与重现，使师生通过数字化身参与课堂，在虚拟教学场所（包括虚拟现实课堂、教研室、实验室与虚拟仿真实训基地）中实现无障碍地进行实时交流，教学将因为虚拟世界的无限可能性而变得更加生动有趣。例如，学生可以通过VR设备进入细胞、器官甚至月球表面，看到物质的变化、分离和重组。互动结合多样化的数字对象，能够实现与复杂现象与抽象概念的互动实操，可以打破物理世界中的器具限制，实现面向仿真环境的实践教学，特别是在人体解剖、手术模拟、化工实验等领域更能极大程度上降低实验损耗并起到防范突发事件的作用[34]。

5.5 残障辅助

所谓的信息无障碍是指，无论健全人还是残疾人、无论年轻人还是老年人都能够从信息技术中获益，任何人在任何情况下都能平等地、方便地、无障碍地获取和利用信息[35]。在元宇宙时代，虚实融合的生态将催生出人们在虚拟世界的第二身份以及全新生活方式，能够为残障弱势人群提供健全的生活保障，满足老弱病残等弱势群体的出行需求、文娱需求、就业需求等。例如在满足出行需求方面，开发出行辅助技术，研制并推广一批适配残障弱势人群的虚拟现实设备，使老年人和残疾人重新获得出行的能力，通过化身进行虚拟出行，并配置远程出行监控。

5.6 工业生产

元宇宙提供现实世界的数字化映射及驱动数据，可以进行工业化生产的模拟、规划、设计及测试等工作，改变现有“计算机仿真–模拟验收”的生产方式，达到提质、增效、降本的效果。元宇宙在工业上的应用则比数字孪生更具潜力空间，更加重视虚拟空间和现实空间的协同联动，例如：支持多人协作和模拟仿真的虚拟现实开放式服务平台，支撑产品设计、加工制造、远程运维等环节的数字化和智能化转型。其所反映的虚拟世界既具有现实世界的映射，又具有现实世界中尚未实现甚至无法实现的体验与交互，实现虚拟操作指导现实工业[36]。例如，现实世界的生产流程可以先在虚拟世界进行全方位的模拟和仿真，从而找到最优的生产顺序与投料时间和方式。

5.7 融合媒体

元宇宙时代的数字内容生成与呈现方式将引来革新，在新闻报道、体育赛事、影视动画等融合媒体内容制作领域，超高清、虚拟现实技术的进一步发展将使视听呈现内容在清晰度、自由度、观看体验等方面实现大幅提升。通过运用虚拟现实全景摄像机、三维扫描仪、声场麦克风等设备，探索新型导演叙事、虚拟拍摄技术，实现虚拟体验与现场体验深度融合。在人与人之间的社交活动中，将产生基于虚拟化身等新形式的互动社交新业态。通过真人驱动的数字化身，无论身处何地，人们都可以与好友、同事进行“面对面”沟通。此外，经过赋能的数字化身可以独立开展社交活动，后

台驱动的真实人可通过“主观视角+客观视角”观看。数字化身可为社交障碍人群提供社交引导。

5.8 商贸创意

随着中国经济进入高质量发展阶段，民众对高品质教育、医疗、养老、家庭服务、文化娱乐等诸多方面的需求将持续增加。与此同时，新一代信息技术促进制造业和服务业在产业链上的融合，人工智能技术引领服务业革新，它们共同催生更多定制性、创意性、互动性的服务，打造商贸新业态。在智慧家装、虚拟看房、大型会展、时尚创意等商贸领域，将产生商贸活动体验新模式。用户在外卖点餐时可直接线上进入餐厅，如同现实前台点餐一样，体验餐厅服务，同时餐品将通过无人机在第一时间送至家中。

5.9 智慧城市

智慧城市或称数字城市，是一个虚实交互的城市空间概念。虚拟数字城市再现了城市实体空间，实体城市生活同时为虚拟数字空间提供需求。城市管理者对于城市基础信息和状态的掌握，受限于数据收集和统计的时间。面对管理方面存在的延时性和不准确性等问题，现实的物理城市空间将在元宇宙中实现实时化、精细化、动态化的数字重现与改造，有效支撑城市规划、空间治理、公共服务等多项智能城市应用，打造虚实融合的餐饮、购物、休闲体验，优化资源配置和空间利用，推动城市智慧商圈建立，赋能城市经济发展。

5.10 安全应急

元宇宙技术中虚拟现实的仿真应用及其可视化延伸，促进安全应急防范系统和应急演练的数字化转型，开展沉浸式虚拟演练，加快构建满足矿山安全、危化品安全、自然灾害防治需求的智慧警务与应急管理体系。例如，学习游泳时可以通过模拟训练进行安全教学，仅通过大脑获得溺水的信号来教会学员如何避免溺水，而不必亲身经历危险的溺水。除了应急演习和安全教育教学之外，元宇宙技术还将在军工生产、士兵训练、军事教育等国防军事方面产生重要影响。

6 未来发展建议

6.1 锚定关键基础技术，共同支撑概念落地

元宇宙通过数字对象和数字环境的生成、传输、呈现等技术，在数字环境中建立起媲美于真实世界的虚拟世界，实现虚实双向协同与互动，具有广阔的应用前景，但实现从概念到落地还需要依靠多项基础技术的共同支撑。在构建数字环境方面，由于数字环境存在领域众多、待建模型数量巨大等难点，未来数字人和数字环境生成技术及平台需要同时兼顾效率与成本等。在数字人生成方面，则需要重点解决现实世界中的以数字人进入虚拟世界的建模、驱动、渲染等技术问题，提升数字人制作的逼真度和制造效率。此外，我们还需要确立虚拟世界规则，包括虚拟世界中人与人、人与物的相互协同关系，以及虚拟世界反作用真实世界的准则。

6.2 建立高效生产平台，技术赋能开发应用

元宇宙的繁荣有赖于多方的共建、共享和共治。构建数字内容是元宇宙发展的基石，但目前存在生成周期长、效率低、成本高、大规模应用难以实现的发展痛点，阻碍了元宇宙的快速扩充。因此，未来需要建立高效生产平台和工具集，解决虚拟世界中人与物的快速生成问题，降低普通用户的参入门槛。Epic公司开发了虚幻引擎（UE），形成一套完整的开发工具。面向任何使用实时技术工作的用户，从设计可视化和电影式体验，到制作PC、主机、移动设备、VR和AR平台上的高品质游戏，该套工具适用于游戏、建筑、汽车与运输、广播与实况节目、影视、模拟仿真等领域[37]。虚幻引擎和Quixel的合作后，通过Quixel网站提供的工具，可访问Megascans资产库，获取数千种来自真实世界的3D资产。全新赋能平台具有颠覆产业的能力，因此中国应抓住时机，尽早布局研究建立高效的生产平台和工具集，攻克决定行业走向的共性关键技术，打造开放并举的元宇宙体系。

6.3 探索制定协同规则，维持虚拟世界秩序

元宇宙正处于发展初期，制定通用的标准体系和互操作接口规范，是建立开放互联、兼容并举的多元元宇宙生态的基础条件。2022年5月25日在瑞士达沃斯举行的年会上，世界经济论坛宣布了一项新倡议“定义和建设Metaverse”[38]。2022年6月，Meta、微软、英伟达、华为、Epic Games、Adobe等33家企业与组织作为创始成员，成立了元宇宙标准论坛[39]。论坛主要围绕开放元宇宙所需的互操作性标准即互通性、互兼容性以及验证互通性，共同探索和推动相关标准的制定。随着现实世界和虚拟世界越发深度融合，虚实互动的场景与方式将越来越多样，互动效果也将越来越真实，相应的监管问题将成为元宇宙持续、健康发展的关键议题。

与现实世界类似，元宇宙涵盖各种民事活动和商业活动，诸如产权侵犯、商业纠纷，甚至诈骗问题是无法避免的，需要借鉴实体世界的规章制度，建立相应伦理法规和商业规则。注重保护个体权利、维护社区稳定、保障国家总体安全，亟需在现行法律法规基础上优化并制定虚拟世界的协

同规则。通过制定身份管理规则、伦理法律规范、物品权属与交易规则等一系列规则，处理人与人、人与物、物与物之间的各种关系，以维持系统可控性与稳定性。布局包括伦理、经济、信息安全、社区共识等各方面在内的治理规则，推动虚拟世界和现实世界的规则相接轨，助推产业向上、向善发展。以技术促发展，科技发展成果全民共享。

6.4 拥抱重要历史变革，科学布局新兴产业

如今元宇宙相关概念大热，全球科技巨头企业争相在该领域布局，核心目的是打造以企业自身为主的产业生态并制定有利规则[40]。韩国、日本等也希望在全球虚拟空间中积累先发优势，成立了各类联盟，并打造国家级虚实融合发展平台。中国不少地方政府和企业积极布局相关产业，但整体上在布局基础硬件和底层操作系统、制定虚拟世界规则、聚焦以数字技术服务实体经济发展方面关注较少。当前正处在虚拟世界与现实世界融合发展的起步阶段，中国应提前谋划，站在元宇宙的战略高地，通过政策良性引导产业健康发展，避免资本炒作概念，支持企业布局核心软硬件和颠覆性技术，加强信息通信基础设施和先进算力的保障，改变过去在信息技术领域长期受制于人的被动局面。最后，坚持以实体经济为本发展元宇宙产业，避免脱离实体经济发展，加速实体产业与数字经济的融合，更好解决工业生产和社会生活中智慧工厂、智慧交通、智慧城市等各种场景中的实际问题。

6.5 立足时代风口，探索培养复合型人才机制

数字经济的快速发展对新兴产业人才的数量和质量都提出了更高的要求。在虚实深入融合成为趋势的当下，数字化人才供给如何跟上时代步伐，是元宇宙发展的一大挑战，更是人才教育领域的重要思考方向。当前尚未存在一套成熟的元宇宙人才培养模式，元宇宙探索建设亟需复合型人才队伍。对于高校和科研院所来说，当务之急是完善自身建设，培养一批搭建元宇宙全产业链工程师。然后由这些元宇宙工程师对其他专业进行升级，主要包括搭建元宇宙虚拟空间、通过虚拟实训为企业培养人才。作为人才培养基地，高校和科研院只有理顺了培养方案、课程设置、平台设备等基本事项，才能更好地抓住机会。

参考文献

[1] HUEBSCHMAN M, MUNJULURI B, GARNER H. Dynamic holographic 3-D image projection [J]. Optics express, 2003, 11 (5): 437-445. DOI: 10.1364/oe.11.000437

[2] HANJALIC A, LAGENDIJK R L, BIEMOND J. Automated high-level movie segmentation for advanced video-retrieval systems [J]. IEEE transactions on circuits and systems for video technology, 1999, 9(4): 580-588. DOI: 10.1109/76.767124

[3] JONES R M, LAIRD J, NIELSEN P, et al. Automated intelligent pilots for combat flight simulation [J]. AI magazine. 1999, 20(1): 27-41

[4] HESS R A, MALSBURY T. Closed-loop assessment of flight simulator fidelity [J]. Journal of guidance, control, and dynamics, 1991, 14(1): 191-197. DOI: 10.2514/3.20621

[5] BAILLARD C, MAîTRE H. 3-D reconstruction of urban scenes from aerial stereo imagery: a focusing strategy [J]. Computer vision and image understanding, 1999, 76(3): 244-258. DOI: 10.1006/cviu.1999.0793

[6] JAYNES C, RISEMAN E, HANSON A. Recognition and reconstruction of buildings from multiple aerial images [J]. Computer vision and image understanding, 2003, 90(1): 68-98. DOI: 10.1016/s1077-3142(03)00027-4

[7] 王同聚. 虚拟和增强现实(VR/AR)技术在教学中的应用与前景展望 J]. 数字教育, 2017, 3(1): 1-10

[8] Mixed Reality Laboratory. About us [EB/OL]. [2022-07-20]. https://www. nottingham. ac. uk/research/groups/mixedrealitylab/about.aspx

[9] MagicLeap. 关于 magicleap 介绍 [EB/OL]. (2020-10-23) [2022-07-20]. http://www.magicleap.org.cn/

[10] NEGRI E, FUMAGALLI L, MACCHI M. A review of the roles of digital twin in CPS-based production systems [M]//CRESPO MÁRQUEZ A, MACCHI M, PARLIKAD A. Value Based and Intelligent Asset Management. Cham: Springer, 2020: 291-307.10.1007/978-3-030-20704-5_13

[11] 河北日报. 数字孪生将带来制造业变革 [EB/OL]. (2020-09-30) [2022-07-24]. http://hbrb. hebnews. cn/pc/paper/c/202009/30/content_56999.html

[12] 中国政府网. 拥有"人"的能力，数字化时代制造业将带来哪些新形态？ [EB/OL]. (2020-10-14) [2022-07-24] http://www. gov. cn/xinwen/2020-10/14/content_5551358.htm

[13] STEPHENSON N. Snow crash [M]. 郭泽译, 四川科学技术出版社, 2009

[14] Smart City Korea. Seoul metropolitan government builds its own 'Metaverse Platform' for the first time in a new concept public service [EB/OL]. (2021-11-03) [2024-02-26]. https://smartcity.go.kr/en/

[15] The Japan Metaverse Association .About us [EB/OL]. [2022-06-031]. https://japanmeta.org/en/en_about/

[16] MELNICK K. Red 6 just completed its first AR training mission with aircraft [EB/OL]. (2022-06-07) [2024-02-26]. https://vrscout.com/news/red-6-just-completed-its-first-ar-training-mission-with-aircraft/

[17] 中共中央国务院. "十四五"数字经济发展规划 [EB/OL]. (2022-01-12) [2022-6-30]. http://www. gov. cn/zhengce/zhengceku/2022-01/12/content_5667817.htm

[18] 上海市经济和信息化委员会. 上海市电子信息产业发展"十四五"规划 [EB/OL]. (2021-12-30) [2022-07-01]. http://www. sheitc. sh. gov. cn/cyfz/20211230/99677f56ada245ac834e12bb3dd214a9.html

[19] 浙江省民营经济研究中心. 省数字经济发展领导小组办公室印发《关于浙江省未来产业先导区建设的指导意见》[EB/OL]. (2022-01-05)[2022-07-01]. http://www.myjjzx.cn/cj/view.php?aid=484

[20] 广州市黄埔区人民政府. 广州市黄埔区 广州开发区促进元宇宙创新发展办法 [EB/OL]. (2022-04-06) [2022-07-01]. http://www.hp.gov. cn/gzjg/qt/gzshpqrcjlfwzx/hprczc/rencaizc/zchb/content/post_8337694.html

[21] 北京市通州区人民政府. 关于印发《北京城市副中心元宇宙创新发展行动计划（2022-2024 年）》的通知（通政发〔2022〕13 号） [EB/OL]. (2022-08-24) [2022-08-26]. http://www. bjtzh. gov. cn/bjtz/

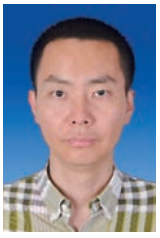
jdhy/202208/1612429.shtml
[22] 水利部信息中心．“十四五”数字孪生黄河建设方案通过水利部审查 [EB/OL]. [2022−08−26] http://xxzx. mwr. gov. cn/zxpd/gdcz/202206/t20220623_1581441.html
[23] 北京海淀．全国首份《元宇宙教育共识》在2022第二届元宇宙教育前沿峰会上发布 [EB/OL]. [2022−08−25]. https://baijiahao.baidu.com/s?id=1741235129158696331&wfr=spider&for=pc
[24] 亿欧网．共话AI医疗与元宇宙，影像与康复的新价值高地在何方？[EB/OL]. [2022−09−05]. https://baijiahao. baidu. com/s? id=1743043125699356356&wfr=spider&for=pc
[25] 央广网．腾讯公司高级副总裁马晓轶：游戏正成为一个“超级数字场景”[EB/OL]. (2022−06−28) [2022−07−26]. http://tech. cnr. cn/techph/20220628/t20220628_525886199.shtml
[26] iNFTnews. 一文读懂腾讯的元宇宙生态布局 [EB/OL]. (2022−07−18)[2024−02−26]. https://inftnews.com/57794/
[27] 刘建明．科技大国“元宇宙”研究观点述评 [J]. 中国广播电视学刊，2022(6): 11−17
[28] LIN X. Metaverse: what? why? when? [EB/OL]. [2022−07−04]. https://www.solactive.com/metaverse−what−why−when/
[29] Meta. Building the metaverse responsibly [EB/OL]. (2021−09−27) [2022−07−03]. https://about. fb. com/news/2021/09/building−the−metaverse−responsibly/
[30] NVIDIA. What is the metaverse? [EB/OL]. (2021−08−10)[2022−07−04]. https://blogs. nvidia. com/blog/2021/08/10/what−is−the−metaverse/
[31] The Official Microsoft Blog. Microsoft cloud at ignite 2021: metaverse, AI and hyperconnectivity in a hybrid world [EB/OL]. [2022−07−04]. https://blogs.microsoft.com/
[32] 德勤中国．元宇宙综观：愿景、技术和应对 [EB/OL]. [2022−07−24]. https://www2. deloitte. com/cn/zh/pages/technology−media−and−telecommunications/articles/metaverse−report.html
[33] 梅夏英，曹建峰．从信息互联到价值互联：元宇宙中知识经济的模式变革与治理重构 [J]. 图书与情报，2021(6): 69−74
[34] 张欣．“元宇宙”将对教育产生什么影响 [N]. 中国教育报，2022−01−03(2)
[35] 中国残疾人联合会“无障碍声明”[EB/OL]. [2022−10−24]. https://www.cdpf.org.cn/wzasm/index.htm
[36] 赛迪智库．工业元宇宙：展望智能制造的未来形态 [J]. 河南科技，2022, 41(9): 1−3
[37] Epic. Engine−5 [EB/OL]. [2024−02−26]. https://www.unrealengine.com/zh−CN/unreal−engine−5
[38] World Economic Forum. New initiative to build an equitable, interoperable and safe metaverse [EB/OL]. [2022−7−2]. https://www. weforum. org/press/2022/05/new−initiative−to−build−an−equitable−interoperable−and−safe−metaverse
[39] Metaverse Standards Forum. 领先的标准组织和公司联合起来推动开放的元宇宙互操作性 [EB/OL]. [2024−02−26]. https://metaverse−standards.org/
[40] 邢相烨．元宇宙时代如何把握发展契机？[J]. 国资报告，2022(2): 108−110

作 者 简 介

冯大权，深圳大学副教授、博士生导师，广东省哲学社会科学重点实验室（文化数字化与文化创新发展）副主任；主要研究领域为无线边缘网络、沉浸式通信、数字创意技术。

张胜利，深圳大学教授、博士生导师，深圳大学电子与信息工程学院副院长；主要研究领域为区块链技术、物理层网络编码。

吕星月，深圳大学数字创意研究中心助理研究员；主要研究领域为信息咨询与分析、数字人文。



王振中，中央广播电视总台技术局正高级工程师，总台技术局技术规划研究部副主任；主要研究领域为媒体融合传播、4K/8K 超高清制播。

面向边缘智能的通信计算一体化研究

Integrated Communication and Computation for Edge Intelligence

江炳青/JIANG Bingqing[1]，杜军/DU Jun[1]，
王劲涛/WANG Jintao[1]，牟林/MU Lin[2,3]
（1. 清华大学电子工程系，中国 北京 100084；
2. 中兴通讯股份有限公司，中国 深圳 518057；
3. 移动网络和移动多媒体技术国家重点实验室，中国 深圳 518055）
(1. Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;
2. ZTE Corporation, Shenzhen 518057, China;
3. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China)

DOI：10.12142/ZTETJ.2024S1003
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20240724.1044.006.html
网络出版日期：2024-07-25
收稿日期：2023-12-06

摘要：为了进一步提高无线数据聚合效率，空中计算技术通过利用无线信道波形叠加特性允许模型更新信息在空中“一次性”完成聚合，实现通信网与算力网的“网媒融合”。然而在这过程中，信道衰落和噪声可能会带来聚合失真。此外，更新数据的质量以及边缘设备的传输能耗也可能影响模型聚合以及收敛效率。为此提出了基于空中计算的联邦学习系统，并针对其存在的信道干扰、高效数据传输和数据失真问题建立动态设备调度机制，在满足接收端信噪比条件下选择适当数量质量较高的设备参与模型训练。该机制利用梯度重要性、信道条件和传输能耗衡量设备质量并保留累积未被选择设备的梯度以加速收敛。基于李雅普诺夫优化理论进行问题建模和求解，仿真结果表明该机制具有较高训练精度和较快收敛速度，同时针对不同噪声功率具有一定鲁棒性。

关键词：空中计算；联邦学习；设备调度；设备质量；鲁棒性

Abstract: Over-the-air computation (AirComp) technology leverages the waveform superposition characteristics of wireless channels to further enhance the efficiency of wireless data aggregation, enabling model update information to be aggregated “in one shot”. This achieves a convergence of communication networks and computational power networks, exemplifying the concept of “network and computation fusion”. However, channel fading and noise may introduce aggregation distortion during this process. Additionally, the quality of update information and the transmission energy consumption of edge devices can impact model aggregation and convergence efficiency. Therefore, we establish an AirComp enabled federated learning system and propose a dynamic device scheduling mechanism to address issues related to channel interference, efficient data transmission, and data distortion. Specifically, an appropriate number of higher-quality devices are selected to participate in model training while satisfying receiving signal-to-noise ratio conditions. It utilizes gradient importance, channel conditions, and transmission energy consumption to assess device quality and retains and accumulates gradients from unselected devices to accelerate convergence. The problem is modeled and solved based on the Lyapunov optimization theory. Simulation results demonstrate that this mechanism achieves higher training accuracy, faster convergence speed, and a certain level of robustness against varying noise power levels.

Keywords: over-the-air computation; federated learning; device scheduling; device quality; robustness

引用格式：江炳青，杜军，王劲涛，等. 面向边缘智能的通信计算一体化研究 [J]. 中兴通讯技术，2024, 30(S1): 16-23. DOI: 10.12142/ZTETJ.2024S1003
Citation: JIANG B Q, DU J, WANG J T, et al. Integrated communication and computation for edge intelligence [J]. ZTE technology journal, 2024, 30(S1): 16-23. DOI: 10.12142/ZTETJ.2024S1003

随着移动互联网由“万物互联”发展为“万物智联”，智能化已经成为新的主要需求和发展趋势。这一巨大变革推动了新的智能化应用，同时对于超可靠低延迟通信、能量效率、智能与安全也提出了更加严格的要求[1]。在智能化引领发展的阶段中，人工智能和机器学习技术被广泛应用于移动互联网领域及其发展中。得益于特有的普适性、自主性以及迭代优化等特性，人工智能和机器学习技术通过数据处理得到更加严谨、稳固的模型和推演结果以提供先进的计

基金项目：国家自然科学基金项目（U23A20281、61971257）

算能力，支持越来越多的智能应用[2]。

传统机器学习范式通常采用依赖于中央云服务器的集中式学习方式，然而智能应用的爆炸式增长以及网络边缘产生的海量数据使得计算密集型任务在数据传输与模型计算方面产生较大时延与带宽负担。此外，由于集中式学习需要聚集所有的原始数据，尽管可以获得性能更优的训练效果，但这一过程中可能会遭到恶意攻击或者窃听，产生了隐私安全问题。随着隐私安全意识的逐渐加强，这一缺点对于部分应用会造成致命的影响，例如金融、医疗等行业中隐私高度敏感的应用。

受益于边缘设备日渐增长的存储与计算能力，分布式学习架构的提出解决了隐私安全和带宽资源问题。中央服务器将数据限制在各个边缘设备处进行模型训练，并将最终的模型汇聚到服务器进行汇总和存储[3]。分布式学习架构实现了将任务分布到不同边缘设备进行训练，避免了原始数据的上传，有效保护了数据安全。但是由于不同边缘设备仅拥有部分数据且彼此之间缺乏数据或模型方面的“沟通”，系统内数据无法交汇融合从而导致训练模型缺乏全局性和泛化能力，这一问题也被称为“数据孤岛”问题[4]。

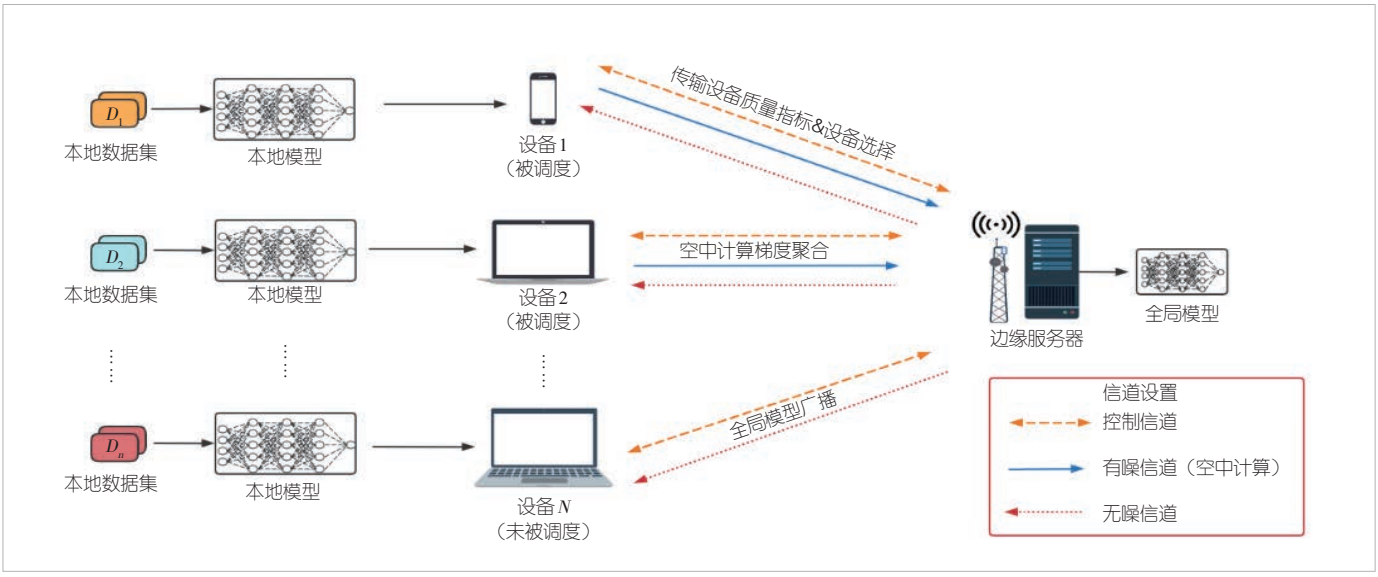
针对以上问题，由谷歌提出的联邦学习框架使得智能边缘协同模型训练成为现实[5]。在不共享边缘设备本地数据集的情况下，通过边缘服务器进行协作训练[6-7]。典型的联邦学习是通过通信与计算相分离的方式实现的。具体来说，每个边缘设备都使用边缘服务器广播的当前全局模型进行本地模型训练，通过运行随机梯度下降算法计算本地模型参数；本地训练结束后，服务器聚合边缘设备上传的本地模型参数及其梯度，并利用聚合值的平均值进行全局模型更新。服务器将得到的全局模型再次广播给边缘设备以开始下一轮训练，这一过程是较为经典的联邦平均算法。

然而在通信与计算相分离的方式下，尽管本地模型的数据量相比于数据本身而言已经小得多，但向无线信道上传本地模型参数或其梯度仍然需要消耗大量频谱资源，造成了“通信瓶颈”问题；另一方面，服务器仅需要本地模型参数的加权平均值而非其本身来训练全局模型，而传统的模型聚合利用编解码辅助准确接收模型参数，这会导致不必要的通信和计算资源浪费。因此，将通信网与算力网进行“网媒融合”，实现通信计算一体化是十分必要的。受此启发，空中计算的引入为“通信瓶颈”问题提供了有效的解决方案，且已被文献[8]证明是一种利用无线信道叠加特性计算和训练的有效通信方法。通过实现通信与计算一体化融合，空中计算可以很好地满足联邦学习系统的高通信效率和低延迟需求，辅助联邦学习在无需精准接收模型参数的情况下实现快速、高效的无线数据聚合。

本文将研究基于空中计算的联邦学习系统，其中在边缘服务器的协同下，边缘设备协作训练并高效通信以完成全局训练任务。本文从收敛性分析着手，综合考虑不同边缘设备间对于全局模型的贡献程度差异，设计了动态设备调度策略以追求高模型精度、能量效率以及通信效率。

1 系统模型

如图1所示，基于空中计算的联邦学习系统由边缘服务器和N个边缘设备组成，设备本地数据集为D_n并采用小批



▲图1 基于空中计算的联邦学习系统架构图

量($D_{n,t}^m \subseteq D_n$)随机梯度下降算法计算本地模型梯度。本文定义一次全局模型更新为一个通信轮次，索引表示为$t \in \{1,\cdots,T\}$。由于通信资源限制每轮只有部分设备上传信息，S_t为被选设备集合，$|S_t| = \sum_{n=1}^{N} a_{n,t}$，其中$a_{n,t} \in \{0,1\}$为设备$n$是否被选择的二元变量。

1.1 联邦学习模型

定义设备n的局部损失函数为$f_n(\boldsymbol{w}) = \frac{1}{|D_n|}\sum_{(s_i,q_i)\in D_n} l(\boldsymbol{w}; s_i, q_i)$，其中$l(\boldsymbol{w}; s_i, q_i)$为数据样本$(s_i,q_i)$的损失函数，则衡量全局模型对整体数据集平均拟合度的全局损失函数为

$$f(\boldsymbol{w}) = \frac{1}{\left|\bigcup_{n\in\mathbb{N}} D_n\right|}\sum_{n=1}^{N}|D_n| f_n(\boldsymbol{w})。 \tag{1}$$

联邦学习旨在寻找最佳模型参数$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} E[f(\boldsymbol{w}_T)]$。应用伸缩和法则，优化问题化为

$$P_1: \min \sum_{t=1}^{T} E\left[f(\boldsymbol{w}_t)\right] - E\left[f(\boldsymbol{w}_{t-1})\right], \tag{2a}$$

$$\text{s.t.} \quad \frac{1}{T}\sum_{t=1}^{T} a_{n,t} E_{n,t} \le \bar{E}_n, \tag{2b}$$

$$a_{n,t} \in \{0,1\}, \tag{2c}$$

其中，$\bar{E}_n$为设备n的平均传输能量约束。由于神经网络模型参数无法预测，$E\left[f(\boldsymbol{w}_t)\right] - E\left[f(\boldsymbol{w}_{t-1})\right]$难以写出闭式形式，因此后文将通过收敛性分析得到其闭式上界进行近似简化。

1.2 空中计算传输聚合模型

假设信道增益遵循独立同分布准静态瑞利衰落，$h_{n,t} \sim CN(0,1)$，本文提出基于信道反转法的功率控制以抵抗信道衰落[9]。设备n的发射功率为

$$P_{n,t} = \sigma_t \frac{h_{n,t}^H}{|h_{n,t}|^2}, \tag{3}$$

其中功率扩展因子σ_t决定服务器接收端信噪比。尽管$h_{n,t}$会引入相关复数操作，但发射功率的计算是不受影响的。令设备发射信号$\boldsymbol{x}_{n,t} = \boldsymbol{\Phi}(\boldsymbol{g}_{n,t})$，$\boldsymbol{\Phi}()$为预处理操作以确保$\boldsymbol{x}_{n,t}$具有零均值和方差$P_{n,t}$以便能量控制[9]，则聚合信号为

$$\boldsymbol{y}_t = \sum_{n\in S_t} h_{n,t}\boldsymbol{x}_{n,t} + \boldsymbol{\varepsilon}_t, \tag{4}$$

其中$\boldsymbol{\varepsilon}_t$为0均值方差$\sigma_0^2$的高斯噪声。服务器完成后处理后更新全局模型：

$$\boldsymbol{w}_t = \boldsymbol{w}_{t-1} - \eta_t\left(\frac{\sum_{n\in S_t}\boldsymbol{g}_{n,t}}{|S_t|} + \hat{\boldsymbol{\varepsilon}}_t\right), \tag{5}$$

其中η_t为学习速率，$\hat{\boldsymbol{\varepsilon}}_t = \phi^{-1}\left(\frac{\boldsymbol{\varepsilon}_t}{\sigma_t|S_t|}\right)$。根据公式（4），接收端信噪比$\gamma_s = \frac{|h_{n,t}|^2 E\left[\left|\left|\boldsymbol{x}_{n,t}\right|\right|_2^2\right]}{\sigma_0^2}$，令$\gamma_{thr}$为预设信噪比阈值，要求服务器处满足$\gamma_s \ge \gamma_{thr}$，则${\sigma_t}^2$为

$${\sigma_t}^2 = \gamma_{thr}\sigma_0^2。 \tag{6}$$

2 基于梯度与信道感知的动态设备调度机制

2.1 动态残差反馈机制

由于设备选择机制的存在，未被选择设备的本地更新信息无法用于全局模型更新，从而会导致“数据失真”问题。目前大多数研究中未被选择设备的本地更新信息直接被丢弃，这种粗糙的操作可能会导致训练偏差，甚至改变梯度的方向和大小并进一步影响模型训练的稳定性和准确性[10]。因此，有必要保留未被选择设备的训练信息，并在数据和信道质量都满足条件而被调度时联合累积的梯度一同传输至服务器。为此，本文提出了一种动态残差反馈方案，该方案允许边缘设备传输过去未发送的累积局部梯度，而未被选择的设备在本地保存本轮模型梯度并等待下一次传输。定义被选择的边缘设备n在第t轮的本地更新信息$\tilde{\boldsymbol{g}}_{n,t}$为当前通信轮次的梯度向量$\boldsymbol{g}_{n,t}$与累积残差$\boldsymbol{r}_{n,t}$的组合，即

$$\tilde{\boldsymbol{g}}_{n,t} = \boldsymbol{g}_{n,t} + \boldsymbol{r}_{n,t}, \tag{7a}$$

$$\boldsymbol{r}_{n,t} = \begin{cases} 0, n \in S_{t-1}, t \ge 2 \\ \xi\boldsymbol{g}_{n,t-1}, n \notin S_{t-1}, t \ge 2 \end{cases}, \tag{7b}$$

其中，$0 \le \xi \le 1$表示累积残差对于当前训练的重要程度，$\boldsymbol{r}_{n,1} = 0$。

2.2 基于收敛性分析的问题转化

本文首先引入以下假设以利于分析收敛性[11]：

假设1（l–smooth）：损失函数$f_1(\boldsymbol{w}), \cdots, f_N(\boldsymbol{w})$具有$L$–光滑性，即对于$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{R}^{L_g}$以及$\forall n \in N, \exists l < \infty$使得损失函数满足

$$f_n(\boldsymbol{y}) \ge f_n(\boldsymbol{x}) + \nabla f_n(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) + \frac{l}{2}\left|\left|\boldsymbol{y} - \boldsymbol{x}\right|\right|^2。 \tag{8}$$

假设2（μ–strongly）：损失函数$f_1(\boldsymbol{w}), \cdots, f_N(\boldsymbol{w})$具有$\mu$–鲁棒性，即对于$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{R}^{L_g}$以及$\forall n \in N, \exists \mu < \infty$使得损

失函数满足

$$f_n(\boldsymbol{y}) \leqslant f_n(\boldsymbol{x}) + \nabla f_n(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}) + \frac{\mu}{2}\left\|\boldsymbol{y}-\boldsymbol{x}\right\|^2_2。 \tag{9}$$

假设3（无偏估计&方差有界）：本地随机梯度具有无偏估计性且方差有界，即$\forall n \in \mathbb{N}, \forall t \in T$,对随机数据抽样取期望后满足

$$E\left[\boldsymbol{g}_{n,t}\right]|_{D^m_{n,t} \subseteq D_n} = \boldsymbol{g}_t, \tag{10a}$$

$$E\left[\left\|\boldsymbol{g}_{n,t}-\boldsymbol{g}_t\right\|^2_2\right]|_{(s_i,q_i)\in D_n} \leqslant G^2, \tag{10b}$$

其中，G^2表示局部梯度的方差上界且可以认为是常数。基于以上假设，本文首先在以下引理1中推导了单轮平均累积残差的期望用于进一步的收敛性分析。

引理1：给定第t轮的全局模型梯度$\boldsymbol{g}_t$和本地随机梯度的方差上界G^2，单轮平均累积残差的上界为

$$\left[\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{r}_{n,t}\right] \leqslant \frac{P_{r_t}\left(G^2+\left\|\boldsymbol{g}_t\right\|^2_2\right)}{1-P_{r_t}}, \tag{11}$$

其中P_{r_t}表示第t轮中设备未被选择的平均概率，$0<P_{r_t}<1, \forall t \in \{0,\cdots,T-1\}$。相关证明见文献[12]附录。

随后基于假设和引理1，本文通过推导$E\left[f(\boldsymbol{w}_t)\right]-E\left[f(\boldsymbol{w}_{t-1})\right]$的上界进行单轮收敛性分析，此处综合考虑了发射功率、信道噪声、平均累积残差上界以及随机梯度来进行分析，分析结果如引理2。

引理2：给定第t轮的调度设备集合S_t和小批量大小$D^m_{n,t}$，通信的收敛速度由公式（12）给出。

$$E\left[f(\boldsymbol{w}_t)\right]-E\left[f(\boldsymbol{w}_{t-1})\right] \leqslant \frac{l\eta_t^2}{2}\left\|\boldsymbol{g}_t\right\|^2_2 + \frac{l\eta_t^2+\eta_t}{2}m^2 + \frac{l\eta_t^2+\eta_t}{2}\frac{P_{r_t}\left(G^2+\left\|\boldsymbol{g}_t\right\|^2_2\right)}{1-P_{r_t}} + \frac{l\eta_t^2}{2}\left[\frac{\delta^2}{\gamma_{thr}\left|S_t\right|^2}+\frac{G^2}{\left|S_t\right|\left|D^m_{n,t}\right|}\right], \tag{12}$$

其中，m是全局梯度的估计平均值，δ^2是全局梯度的估计方差值。证明见文献[12]附录。

观察引理2可知，每轮所选择的设备数量，即$|S_t| = \sum_{n\in S_t}a_{n,t}$，在抵抗信道噪声影响方面起到了至关重要的作用。因此，高效的设备调度机制需要根据每一轮中的信道噪声水平来优化设备选择策略$a_{n,t}$。与最大化平均每轮所选择的设备数量不同，引理2提供了一个更加合理的优化方向，从收敛速率着手，最小化$E\left[f(\boldsymbol{w}_t)\right]-E\left[f(\boldsymbol{w}_{t-1})\right]$的上界以抵抗信道扰动的影响。此外，由于本文使用单轮收敛速率的上界进行近似得到闭式表达式，使得操作可执行化。对引理2做进一步观察，$\boldsymbol{g}_t$是定义在整个数据集上的全局梯度，这表明对于固定数据集，$\boldsymbol{g}_t$也是固定的。同时引理2中公式（3）–（11）的前三项与调度策略无关，可以视作常数对待。因此，忽略常数项$l\eta_t^2/2$，令

$$U_t \triangleq \frac{\delta^2}{\gamma_{thr}\left|S_t\right|^2}+\frac{G^2}{\left|S_t\right|\left|D^m_{n,t}\right|}。 \tag{13}$$

进而问题P_1可以优化转化为P_2进行求解

$$P_2 \min \quad \sum_{t=1}^{T}U_t,$$

$$\text{s.t.} \quad \frac{1}{T}\sum_{t=1}^{T}a_{n,t}E_{n,t} \leqslant \bar{E}_n,$$

$$a_{n,t} \in \{0,1\}。 \tag{14}$$

2.3 设备质量指标设计

在联邦平均算法以及当前大多数研究中，服务器随机选择边缘设备参与全局模型训练，忽略了不同边缘设备之间的差异。然而，上传更新变化不大的局部模型参数梯度对全局模型性能提升十分有限，因此降低了传输资源的利用效率。此外，传输所带来的能量消耗以及信道条件都会影响模型聚合的效率和准确性。综合考虑上述因素，本小节旨在设计量化合适的设备质量指标以测量不同边缘设备对全局模型的潜在影响。

本地更新信息重要性由其l_2–范数$\|\tilde{\boldsymbol{g}}_t\|_2^2$度量并定义为数据状态信息，较大数据状态信息可提供较大模型变化[13]；本文用$|h_{n,t}|$衡量信道条件并定义为信道状态信息，信道状态信息越大信道条件越好，避免拖后腿问题[9]。之后归一化数据与信道状态信息$v_{DSI_{n,t}} = \frac{\left\|\tilde{\boldsymbol{g}}_t\right\|^2_2}{g_{\max}}$，$v_{CSI_{n,t}} = \frac{|h_{n,t}|}{h_{\max}}$以统一范围并定义设备重要性，即：

$$V_{n,t} = \rho_1 v_{DSI_{n,t}} + \rho_2 v_{CSI_{n,t}}, \rho_1+\rho_2 = 1。 \tag{15}$$

由于90%能耗来自于传输过程，因此选择机制需要考虑设备传输能耗$E_{n,t} = \sigma_t^2/\|h_{n,t}\|^2$以追求高能效[14]。综合考虑能耗损失和设备重要性增益，定义设备质量指标以衡量不同设备对全局模型的潜在影响：

$$I_{n,t} = \lambda_V V_{n,t} - \lambda_E E_{n,t}, \lambda_E+\lambda_V = 1。 \tag{16}$$

设备能本地计算$I_{n,t}$并传输到服务器，服务器每轮选择具有较大$I_{n,t}$的设备参与训练。

2.4 基于李雅普诺夫漂移优化算法问题建模和动态设备调度算法

基于上述分析，所选设备数量和设备质量是调度机制中的两大关键要素，因此本文旨在设计一种设备调度机制以决定所选设备数量减少噪声干扰，并确保在有限资源下所被选择设备质量尽可能高以充分利用其训练性能。基于李雅普诺夫漂移惩罚算法的设备选择优化问题建立为

$$P_3\ \min_{a_{n,t}} \alpha U_t - \sum_{n=1}^{N} a_{n,t} I_{n,t},$$

约束条件公式(2b)和(2c)， (17)

其中，α是平衡收敛上界可调整项U_t和设备质量$I_{n,t}$之和的李雅普诺夫因子。此外，参数α的选择还需要考虑平均能量约束（2b）是否满足。

算法1给出了基于李雅普诺夫优化理论的求解算法，算法复杂度为O(NlogN)。尽管传输$I_{n,t}$带来额外通信开销，但相比于传输梯度几乎可忽略不计。此外，$I_{n,t}$的定义十分灵活，算法具有很强的适应性。

算法1：基于梯度和信道条件感知的动态设备调度机制

输入：$\boldsymbol{\lambda}_E, \boldsymbol{\lambda}_V, \boldsymbol{\alpha}, \boldsymbol{\gamma}_{thr}$.
输出：设备调度决策$\boldsymbol{a}_{n,t}$
初始化$\boldsymbol{w}_0$
for t=1——t=T do
  服务器广播全局模型参数$\boldsymbol{w}_{t-1}$
  for n=1——n=N do
    本地模型训练计算$\boldsymbol{g}_{n,t}$
    计算$\boldsymbol{I}_{n,t}$并通过控制信道传输到服务器
  end for
  服务器降序排序$\boldsymbol{I}_{n,t}$
  for k=1——k=N do
    服务器计算相应惩罚项$\boldsymbol{p}_t(\boldsymbol{k}) = \boldsymbol{\alpha U}_t - \sum_{n=1}^{k} \boldsymbol{I}_{n,t}$
  end for
  $\boldsymbol{k}^* = \arg\min \boldsymbol{p}_t(\boldsymbol{k})$
for n=1——n=N do
  if $\boldsymbol{I}_{n,t} > \boldsymbol{I}_{k^*,t}$ then
    $\boldsymbol{a}_{n,t} = 1$
  else
    $\boldsymbol{a}_{n,t} = 0$
  end if
  end for
 for n=1——n=N do
  根据式(7)更新本地残差
  if $\boldsymbol{a}_{n,t} = 1$ then
    根据式(3)设置传输功率并通过空中计算向服务器传输$\tilde{\boldsymbol{g}}_{n,t}$
  end if
 end for
 服务器根据式(5)更新全局模型
end for

3 仿真结果与分析

在仿真中，本文研究基于空中计算的联邦学习系统，该系统由一个边缘服务器和N=100个边缘设备组成。本文假设每个边缘设备到边缘服务器的无线信道遵循独立同分布准静态瑞利衰落，建模为$h_{n,t} \sim CN(0,1)$，且边缘服务器和边缘设备都可以观察到准确的信道增益。发射功率根据公式（6）进行计算，其中信道噪声的方差σ_0^2分别在{0.5,1,3}中取值以表示不同的信道条件，通过调整不同σ_0^2的大小验证所提出的设备调度机制的鲁棒性。本文设置平均能量阈值为$\bar{E}_n$=1 ($\forall n \in \mathbb{N}$)。在联邦学习训练设置中，本文通过使用两个5×5卷积层的卷积神经网络、一个带有512个线性整流单元(ReLu）激活的全连接层以及一个SoftMax输出层来执行图像分类的学习任务[5]。此外，本文使用了MNIST数据集，其中设置了60 000个被标记的训练数据样本和10 000个测试数据样本，同时考虑了独立同分布以及非独立同分布数据分布。联邦学习的超参数设置为：动量优化器为0.5，本地模型训练轮次为2，总共训练T=50轮，本地小批量的大小为$|D_{n,t}^m|$=10，学习速率η_t=0.01。此外，本文在预训练中将G^2设置为所有局部梯度方差的最大值[11]。李雅普诺夫因子设置为α=5 000，并令ρ_1=ρ_2=0.5，λ_E=λ_V=0.5以保证公平性[10]。

本文研究系统将与以下基线算法进行性能对比。1）理想基线：使用所有无噪梯度平均值更新全局模型以提供基准精度；2）基线1[9]：$|h_{n,t}| \geqslant |h|_{th} = 1$被调度；3）基线2[13]：首选$|h_{n,t}|$最大的$k_c = 50$个设备，再选$\left|\left|g_t\right|\right|_2^2$最大的$k = 20$个设备；4）基线3[10]：设备根据公式（18）决定是否被调度。

$$P_n^t = \begin{cases} \left(\dfrac{|h_{n,t}|}{|h_{n,t}|^2 + c}\right)^2, \dfrac{\left|\left|\tilde{\boldsymbol{g}}_t\right|\right|_2^2 |h_{n,t}|^2}{|h_{n,t}|^2 + c} \geqslant cP_{\text{on}} \\ 0, \text{else} \end{cases}, \quad (18)$$

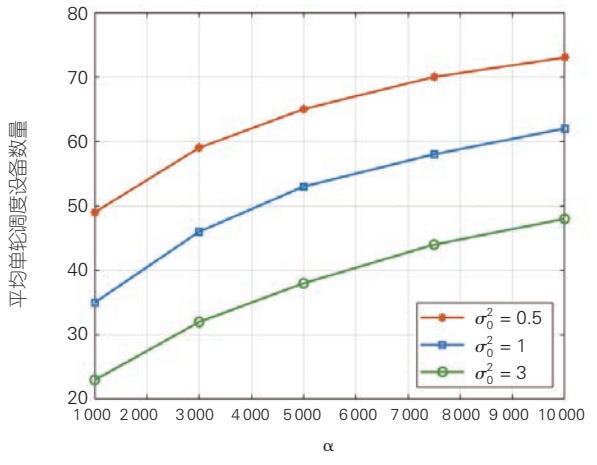
其中，功率消耗代价$c = 1$，设备激活功率$P_{\text{on}} = 4$。

3.1 李雅普诺夫因子α的影响

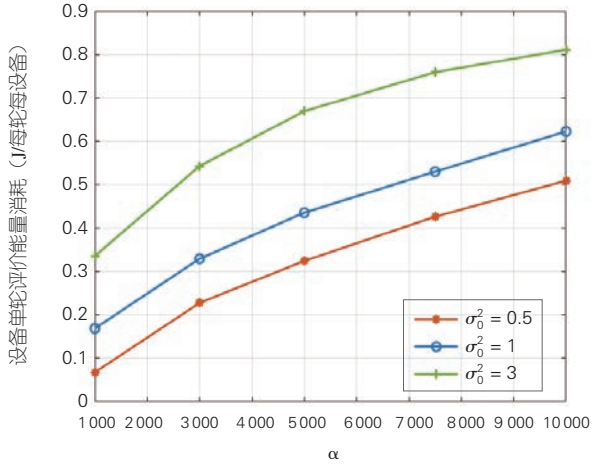
本文首先分析了李雅普诺夫因子α对于调度设备数量和单轮设备平均传输能耗的影响，仿真结果如图2所示。图2（a）表明，调度设备数量随着α的增加而增加，这是因为根据设备调度问题的优化目标、公式（13）以及$I_{n,t}$的定义，边缘服务器倾向于选择更多设备以减少噪声的影响，但是相应代价是增加系统的传输能耗，如图2（b）所示。具体的，根据优化目标可知，增加α意味着服务器更倾向于通过增加调度设备数量以降低收敛上界可调整项U_t，而适当降低由$I_{n,t}$定义的设备质量权重。另一方面，随着α的降低，在可选设备数量降低的条件下，服务器更倾向于选择那些具有“良好”设备质量（即$I_{n,t}$值较大）的设备以抵抗信道噪声影响。

3.2 算法性能对比

图3为算法性能比较图，结果表明得益于动态残差反

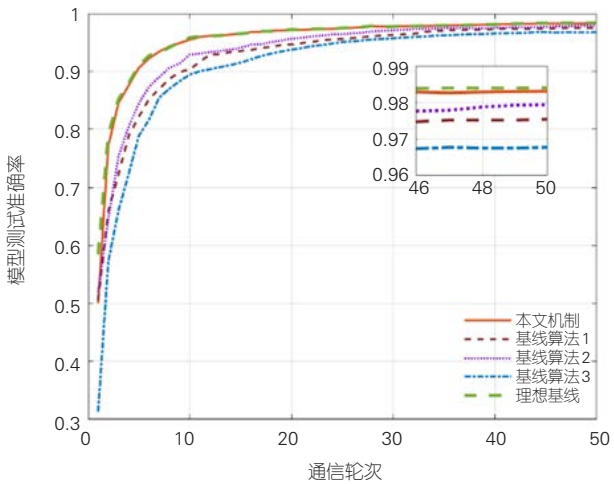


（a）调度设备数量

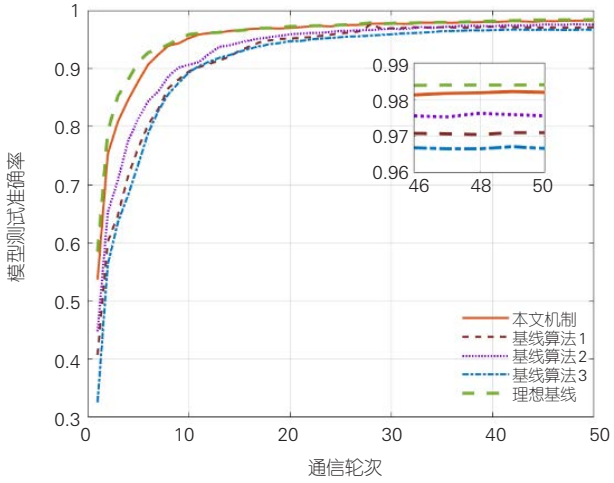


（b）平均能量消耗

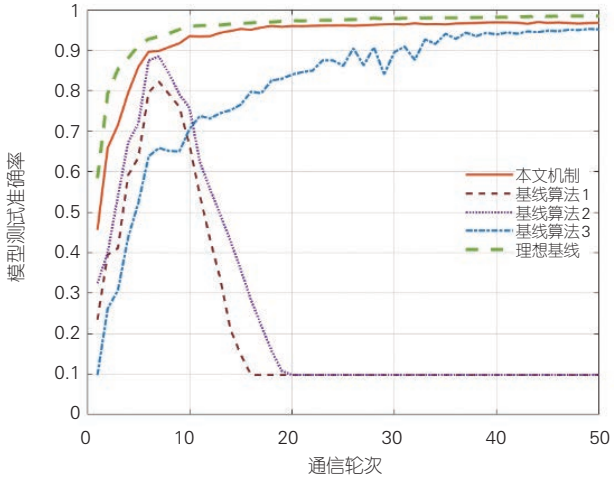
▲图2 不同α和σ_0^2下的调度设备数量、平均能量消耗



（a）σ_0^2 = 0.5



（b）σ_0^2 = 1



（b）σ_0^2 = 3

▲图3 不同信道条件下的模型测试精度

馈、基于信道反转法的功率控制以及基于梯度和信道感知的动态设备调度，该机制实现了最高的测试精度和最快的收敛速度，并针对不同信道条件下表现出一定鲁棒性。当信道条件相对较好时（$\sigma_0^2 = 0.5$ 和 $\sigma_0^2 = 1$）被测模型均可达到理想精度。然而当 $\sigma_0^2 = 3$ 时，基线 1 和 2 性能迅速恶化，因为它们均没有综合考虑信道状态与数据状态，在信道逐渐恶化时无法提供较好更新质量导致性能崩溃；基线 3 虽然综合考虑了信道状态和数据状态，但其传输功率无法有效适应信道变化，没有很好地消除信道衰落影响，此外基线 3 由于未考虑不同噪声功率下的动态调度设置而缺乏一定鲁棒性。综上所述，与基线算法相比，本文所提出的设备调度机制在不同信道条件下表现出最佳的训练精度和鲁棒性，并且能够达到接近基准算法所示的最佳测试精度。

3.3 非独立同分布数据分布性能测试

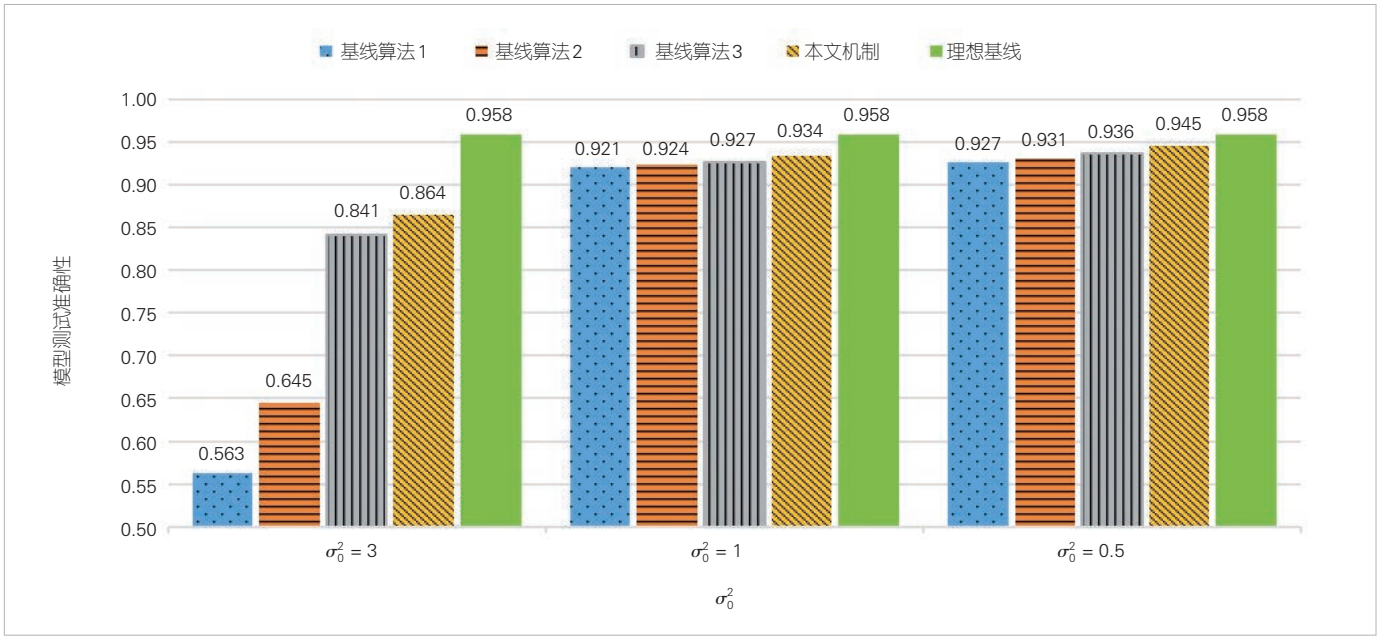
如图 4 所示，在非独立同分布数据分布下，噪声对于精度的影响比独立同分布数据分布下更大。但当 σ_0^2 增大时，受益于适应不同噪声功率的动态调度和辅助全面利用数据集的累积残差，该机制仍然优于基线算法。

4 总结

本文研究了基于空中计算的联邦学习系统，该系统利用多址接入信道的叠加特性完成了高效数据聚合，实现了通信网与算力网的一体融合。此外，综合考虑本地模型更新信息重要性、信道条件以及能耗水平，本文设计了基于梯度和信道感知的动态设备调度机制，在每轮训练中服务器将选择合适数量、具有较高设备质量的边缘设备参与全局模型更新过程。在该机制中，还使用了基于信道反转法的功率控制方案以抵抗信道衰落和噪声的影响。为了提高全局模型更新的效率和准确性，本文还提出了基于动态累积残差反馈的梯度传输机制。最后，为了在线求解设备调度优化问题，本文提出了基于李雅普诺夫漂移优化理论的求解算法并进行了收敛性分析。通过与不同基线算法比较，仿真结果验证了本文所提出的调度机制能够有效提高模型测试精度和收敛速度，并且针对不同信道条件下具有一定鲁棒性。

参考文献

[1] DU J, JIANG C X, WANG J, et al. Machine learning for 6G wireless networks: carrying forward enhanced bandwidth, massive access, and ultrareliable/low-latency service [J]. IEEE vehicular technology magazine, 2020, 15(4): 122-134. DOI: 10.1109/MVT.2020.3019650
[2] 贺倩. 人工智能技术在移动互联网发展中的应用 [J]. 信息通信技术与政策, 2017(2): 1-4
[3] 杨强，童咏昕，王晏晟, 等. 群体智能中的联邦学习算法综述 [J]. 智能科学与技术学报, 2022, 4(1): 29-44. DOI: 10.11959/j.issn.2096-6652.202218
[4] LIU Y, MA Z, YAN Z, et al. Privacy-preserving federated k-means for proactive caching in next generation cellular networks [J]. Information sciences, 2020, 521: 14-31. DOI: 10.1016/j.ins.2020.02.042
[5] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-



▲图 4 非独立同分布数据分布下不同信道条件模型精度对比

efficient learning of deep networks from decentralized data [EB/OL]. (2017-02-28) [2024-05-16]. https://arxiv. org/abs/1602.05629
[6] ZHU Z Q, WAN S, FAN P Y, et al. Federated multiagent actor - critic learning for age sensitive mobile-edge computing [J]. IEEE Internet of Things journal, 2022, 9(2): 1053 - 1067. DOI: 10.1109/JIOT.2021.3078514
[7] KHAN L U, SAAD W, HAN Z, et al. Federated learning for Internet of Things: recent advances, taxonomy, and open challenges [J]. IEEE communications surveys & tutorials, 2021, 23(3): 1759-1799. DOI: 10.1109/COMST.2021.3090430
[8] NAZER B, GASTPAR M. Computation over multiple-access channels [J]. IEEE transactions on information theory, 2007, 53 (10): 3498-3516. DOI: 10.1109/TIT.2007.904785
[9] ZHU G X, WANG Y, HUANG K B. Broadband analog aggregation for low-latency federated edge learning [J]. IEEE transactions on wireless communications, 2020, 19(1): 491-506. DOI: 10.1109/TWC.2019.2946245
[10] SU L Q, LAU V K N. Data and channel-adaptive sensor scheduling for federated edge learning via over-the-air gradient aggregation [J]. IEEE Internet of things journal, 2022, 9(3): 1640 - 1654. DOI: 10.1109/JIOT.2021.3096570
[11] SUN Y X, ZHOU S, NIU Z S, et al. Dynamic scheduling for over-the-air federated edge learning with energy constraints [J]. IEEE journal on selected areas in communications, 2022, 40(1): 227-242. DOI: 10.1109/JSAC.2021.3126078
[12] DU J, JIANG B Q, JIANG C X, et al. Gradient and channel aware dynamic scheduling for over-the-air computation in federated edge learning systems [J]. IEEE journal on selected areas in communications, 2023, 41(4): 1035 - 1050. DOI: 10.1109/JSAC.2023.3242727
[13] AMIRI M M, GÜNDÜZ D, KULKARNI S R, et al. Convergence of update aware device scheduling for federated learning at the wireless edge [J]. IEEE transactions on wireless communications, 2021, 20(6): 3643-3658. DOI: 10.1109/TWC.2021.3052681
[14] TAÏK A, MLIKA Z, CHERKAOUI S. Data-aware device scheduling for federated edge learning [J]. IEEE transactions on cognitive communications and networking, 2022, 8(1): 408-421. DOI: 10.1109/TCCN.2021.3100574

作 者 简 介

江炳青，清华大学电子工程系在读硕士研究生；主要研究领域为无线通信与人工智能。

杜军，清华大学电子工程系助理研究员、硕士生导师；研究方向为异构网络智能协同与组网优化；入选中国科协青年人才托举工程，曾获中国电子学会科学技术奖技术发明一等奖；发表论文80余篇，授权国家发明专利10余项，出版专著6部。

王劲涛，清华大学电子工程系教授、博士生导师；研究方向为数字多媒体广播、无线光异构融合通信、AI增强的智能通信信号处理技术；曾获国家科技进步奖一等奖、日内瓦国际发明展金奖、北京市科学技术奖一等奖；发表学术论文180余篇，授权国家发明专利50余项，出版专著5部。

牟林，中兴通讯移动网络和移动多媒体技术国家重点实验室高级研发总监、技术规划与行业趋势资深专家；研究方向为移动网络和移动多媒体。

语义编码与经典信道编码融合研究

Research on Fusion of Semantic Coding and Classical Channel Coding

向际鹰/XIANG Jiying[1,2]，段向阳/DUAN Xiangyang[1,2]，冯雨龙/FENG Yulong[1,2]
（1. 中兴通讯股份有限公司，中国 深圳 518057；
2. 移动网络和移动多媒体技术国家重点实验室，中国 深圳 518055）
(1. ZTE Corporation, Shenzhen 518057, China;
2. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China)

DOI：10.12142/ZTETJ.2024S1004
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20240724.1035.004.html
网络出版日期：2024-07-24
收稿日期：2023-11-25

摘要：目前的语义通信研究尚未阐明语义编码与经典信道编码之间的关系、语义编码在现有通信框架中的可行性，以及影响语义编码的关键因素等。对基于联合信源信道编码的语义通信系统进行了理论分析，设计了语义编码与经典信道编码的融合实验，展示了语义编码的潜在优势，探索了语义编码与经典信道编码之间的关系，研究了后续将语义通信应用于经典通信框架的基础方法。

关键词：人工智能；语义通信；经典信道编码；联合信源信道编码；融合实验

Abstract: Current research on semantic communication has not yet elucidated the relationship between semantic coding and classical channel coding, the feasibility of semantic coding within existing communication frameworks, and the critical factors influencing semantic coding. In this paper, we conduct theoretical analysis of the semantic communication system based on joint source-channel coding and perform the integrated experiment between semantic coding and traditional channel coding. These efforts partially reveal the potential advantages of semantic communication, establish the relationship between semantic coding and classical channel coding, and lay the foundation for the future application of semantic communication within classical communication frameworks.

Keywords: artificial intelligence; semantic communication; classical channel coding; joint source-channel coding; integration experiments

引用格式：向际鹰，段向阳，冯雨龙．语义编码与经典信道编码融合研究 [J]. 中兴通讯技术，2024, 30(S1): 24-32. DOI: 10.12142/ZTETJ.2024S1004
Citation: XIANG J Y, DUAN X Y, FENG Y L. Research on fusion of semantic coding and classical channel coding [J]. ZTE technology journal, 2024, 30(S1): 24-32. DOI: 10.12142/ZTETJ.2024S1004

1948年，SHANNON采用“熵”的概念，对通信过程进行了数学建模，并一直沿用至今[1]。但是，他同时也强调“These semantic aspects of communication are irrelevant to the engineering problem”。因此，后续该论文再版时，WEAVER重新定义了通信的概念[2]，并从3个层面对其进行了阐述：

Level A：通信符号能多精确地传输？

Level B：传输符号能多准确地传达所期望的含义？

Level C：接收含义能多有效地以期望的方式影响行为？

人们现在将这3个层面的问题归纳为：语法问题、语义问题和语用问题。其中，第1个问题在信息论中得到了很好的解决，但后续2个问题引发了对通信定义的进一步讨论和思考，并催生出语义通信的概念。CARNAP和BAR-HILLEL率先用逻辑概率代替统计概率来描述基于命题的语义信息[3]。他们认为，句子为真的逻辑概率越高，其语义信息含量就越低。随后，FLORIDI等也分别从各种自然语言属性，包括语言真性[4-5]、模糊性[6]、随机性[7]和物理属性[8]等，研究了语义测量指标。然而，由于自然语言的复杂性以及缺少有效的数学统计工具[9]，到目前为止，业界依然没有一个公认的度量方法来衡量语义。因此，语义通信中，信道到底扮演什么样的角色，还缺少类似SHANNON三大定理那样的理论基础。

近年来，人工智能（AI）的发展给信息的存储和使用带来了革命性变化。尤其是一系列大型语言模型，彻底改变了

基金项目：国家重点研发计划项目（2020YFB1807202）

信息的传统使用方式[10-11]，信息技术新时代的大门即将被打开。AI的快速发展对于传统通信行业而言既是挑战，也是机遇。一方面，AI应用的兴起必将带来巨量的数据传输，可能会使目前的通信网络不堪重负；另一方面，人们发现人工神经网络在自然语言和语法逻辑建模方面十分有效，是最佳的语义衡量器和提取器。AI的不断发展也促进了语义通信的进一步发展，包括以传输图像为主的深度信源信道联合编码（DeepJSCC）系统[12]、以传输文本为主的深度学习语义通信（DeepSC）系统[13]，以及后续以此为基础的一系列语义通信系统的演变[14-16]。

语义通信基于人工神经网络（ANN）强大的非线性拟合能力，意图在语义维度上对信源实现进一步数据压缩。语义通信系统的基础架构一般为生成对抗网络（GAN）[12]或者自编码器（AE）[13]。语义通信系统的编译码器采用端到端训练方式优化，在训练过程中无法像经典通信那样严格区分信源编码模块与信道编码模块。这样的训练过程也决定了语义通信系统需要基于联合信源信道编码（JSCC）方式实现[17]。但是，一方面，作为联合编译码器的ANN难以解释，无法直接建模说明语义通信系统到底能否从联合编码中获得增益。尤其是目前的语义通信系统实现均与信道没有显性相关关系，只是在语义编译码器训练过程中将信道视为其中的一层。因此，还需要进一步实验来确认其优于分离编码方式。另一方面，SHANNON在理论上证明了无限复杂度、无限码长，以及在无失真条件下信源信道分离编码（SSCC）系统的最优性[1]。但如果上述任意一项条件被破坏，分离实现的最优性是否依然成立，就值得商榷。更进一步地说，JSCC在这种情况下，理论上是否能比SSCC方式产生增益，还需要进一步阐明。

为了解决上述问题，本文主要研究如下：

1）针对语义通信JSCC方式，理论分析其在实际应用场景中可以达到最优，但此时SSCC方式不一定与其等价。

2）针对语义编码与经典信道编码进行融合实验，在一定程度上证明基于JSCC方式的语义通信，相比基于SSCC的语义通信具有一定增益。

3）针对影响语义编码与经典信道编码融合系统性能的因素进行深入分析，为后续将语义通信融入经典通信框架，在实际通信场景中的部署奠定基础。

1 信源信道联合编码增益的理论分析

SHANNON在其著名文章[1]中提出，当不考虑复杂度时，SSCC可以达到与JSCC一样的效果。但是，实际通信系统并不严格满足上述两者等价的前提条件。实际通信场景中，信源往往采用失真编码方式，信道不是离散无记忆信道。此外，根据实际通信系统使用情况，还需要考虑有限码长、多用户、时延要求等条件。这就使得理论证明不适用于实际情形，基于SHANNON理论搭建的通信系统在实际使用场景中无法达到最优的情况。

1.1 分离编码最优性

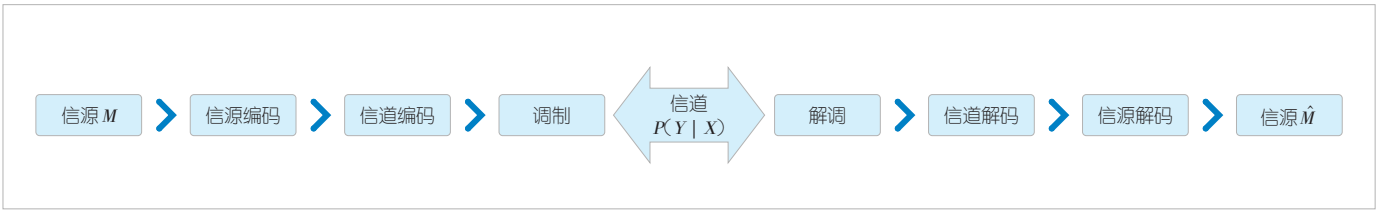
图1是基于SHANNON无差错传输理论的通信系统，其中M表示信源发出消息，$\hat{M}$表示信宿接收消息，X表示信道输入，Y表示信道输出，$P(Y|X)$表示信道转移概率分布，或称其为信道条件分布。

经典通信系统基于信源编码定理、信道编码定理和分离定理[9]，其通信过程如图1所示。如果M表示信源，$P(M)$表示信源分布，$H(M)$表示信源的熵，R_s表示信源编码速率，当不考虑信源符号所代表的具体含义，并且认为其属于典型集时，可得到无损信源编码定理：

定理1：对于给定的信源M和编码速率R_s，若$R_s > H(M)$，则R_s是可达的；若$R_s < H(M)$，则R_s是不可达的。

该定理表明，对于给定的离散无记忆信源，如果信源编码速率R_s超过信源熵，则存在编码方法，只需要对典型序列进行标号，当编码码字无限长时，能够使得译码错误率为任意小。但在实际使用过程中，信源往往不是离散无记忆信源（尤其针对图像），码字长度也不能趋于无限长，这是因为通信系统对译码复杂度或时延是有要求的。

对于信道，假设给定容量为C的离散无记忆信道$\{X, P(Y|X), Y\}$，其中X表示信道输入，Y表示信道输出，X与



▲图1 基于SHANNON无差错传输理论的通信系统

Y构成联合典型集，$P(Y|X)$为信道条件分布，此时存在以下信道编码定理：

定理2：若信道编码速率$R_C < C$，则速率R_C是可达的。即对于任意信道编码速率$R_C < C$，存在一个$(2^{nR_C},n)$的码字序列，当$n \to \infty$时，其误码率$P_e^{(n)} \to 0$；反之，若$(2^{nR_C},n)$码字序列的误码率$P_e^{(n)} \to 0$，必有$R_C < C$。

该定理表明，当所采用信道编码的编码速率小于信道容量时，可以借助编码的方法使得信道译码错误率趋近于0，而且该错误率会随着码长的增加按照指数规律下降。但是，在实际通信过程中，往往无法保证信道是离散无记忆的，更无法保证信道编码可以无限长，信道译码可以不计复杂度。

需要注意的是，上述无损信源编码定理不依赖于信道，信道编码定理也不依赖于信源分布，因此根据SHANNON所述信源信道编码定理，即可得到分离编码最优的结果。

定理3：如果$m_1,m_2,\cdots,m_n$为有限字母表上满足渐进均分性和$H(M) < C$的随机过程，则存在一个信源信道编码使得误差概率$P_e^{(n)} \to 0$。反之，对于任意平稳随机过程，如果$H(M) > C$，那么误差概率远离0，从而不可能以任意低的误差概率通过信道发送这个信源。

定理3结合无损信源编码定理与信道编码定理，最终得到“如果对原始信源M进行传输，JSCC与SSCC是等价的”这个结论，即分离定理。但是，上述SHANNON通信模型是很理想的抽象，虽高度概括了各类通信系统的本质，但远远不能刻画人类面临的实际通信问题：一方面，如果将人类语言作为信源，由于其具有非平稳性，不是各态历经的，而且语言还具有模糊性，并非一定是确定的，因此概率论中没有有效的工具能够处理如此复杂的过程；另一方面，如果将人作为信宿，其发出消息和接收消息的空间是不同的，不同人还具有不同的消息判决策略，因此这些空间和策略是动态变化的。事实上，SHANNON也曾试图采用概率方法对英语做近似表述[18]，计算相应的语言熵，但其效能有限，未能获得进一步发展。综上所述，经典的信息理论远远无法刻画所有的通信过程[9]。

1.2 数据失真情况下的联合编码最优性

语义通信就是为了解决上述困境而诞生的。本文认为，语义通信允许数据失真，但是可以采用建立先验信息，即知识库，使得语义能够保真传输的通信方式[17]。这是因为，语义通信更关注人，或者具有“智能”的机器对信息的理解与感受，而并不聚焦在数据是否无失真传输。当信源编码采用失真编码时，1.1节中基于无损信源编码定理与信道编码定理无依赖条件，所得到的分离定理并不成立，此时描述信源编码过程的是率失真函数。本文用$d(m,\hat{m})$来表示失真函数，其期望表示平均失真$D = E[d(m,\hat{m})]$。如果用P_D表示满足平均失真D的任意概率分布，$I(M;\hat{M})$表示发出消息与接收消息的互信息，那么率失真理论的主要定理为[19]：

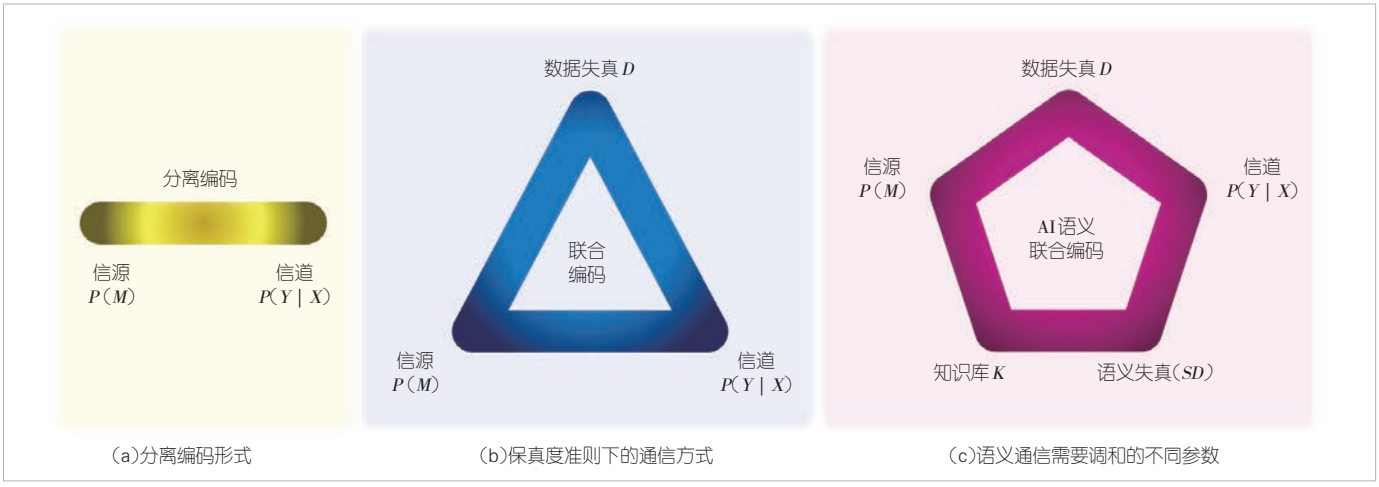
定理4：对一个分布为$P(M)$的独立同分布信源M，假设其信源译码后为$\hat{M}$，如果失真函数为$d(m,\hat{m})$，则率失真函数$R(D)$等于信息率–失真函数$R^{(I)}(D)$，即$R(D) = R^{(I)}(D) = \min\limits_{P(\hat{m}|m) \in P_D} I(M;\hat{M})$是在平均失真$D$下可达的最小信息速率。反之，对任意满足平均失真$D$的率失真码，必然有$R \geqslant R^{(I)}(D)$。

当采用率失真函数描述信源编码过程，采用信道编码定理描述信道编码过程时，两者是否具有依赖性，就成为JSCC是否比SSCC更有优势的关键。实际上，SHANNON在提出率失真理论的同时，也对两者进行了对比[19]，发现此时的JSCC依旧可以达到最优，但是并没有对JSCC达到最优的条件以及SSCC是否依然与其等价，做进一步分析。后来经过人们进一步研究，对第一个问题做出了部分回答[20]：

定理5：对于给定的信源分布$P(M)$，信道条件分布$P(Y|X)$，以及相应的单字母编码方式(f,g)，如果$I(M;\hat{M}) > 0$，当且仅当失真函数满足$d(m;\hat{m}) = -c\log_2 p(m|\hat{m}) + d_0(m)$，其中$c > 0$，并且$d_0$为$m$的函数时，通信系统能够达到最优传输效果。

图2形象地表示了SHANNON三大定理描述的分离编码形式、保真度准则下的通信方式，以及语义通信需要调和的不同参数。

对于无损信源编码，定理1、定理2与定理3共同指出，可以通过编码的方式对信源分布与信道条件分布进行调和，并且可以在信源端与信宿端分离编码，效果与联合编码等价，不会影响其最优性；对于有损信源编码，即保真度准则下的通信，定理3、定理4与定理5共同指出，在存在一定数据失真D的情况下，JSCC方式依然可以对信源分布与信道条件分布进行调和，并且达到最优性能。但是，失真函数需要满足固定形式，该形式与信源分布与信道条件分布均有关。进一步地，在目前工程上可以实现的情况下，在实际通信过程中获得最优码率的JSCC方式是可实现的。但是，SSCC方式由于无法建立码率的解析表达式，因此是次优的实现方案。语义通信需要建立先验知识库K，是在先验知识库K的条件下建立的最优编码方案。此外，或许还可以根据通信的目的，产生一定的语义失真SD，因而其信源压缩或许可以更进一步，调和信源分布与信道条件分布的编码或许



(a)分离编码形式　(b)保真度准则下的通信方式　(c)语义通信需要调和的不同参数

▲图2　在通信方式向语义演进过程中,不同编码方式需要考虑的因素

也更加容易设计，从而产生增益。但是，由于目前没有统一的语义通信理论，上述关于语义通信的结论并未获得严格证明，因此需要进行实验验证。

2 针对联合信源信道编码方式的语义通信系统实验

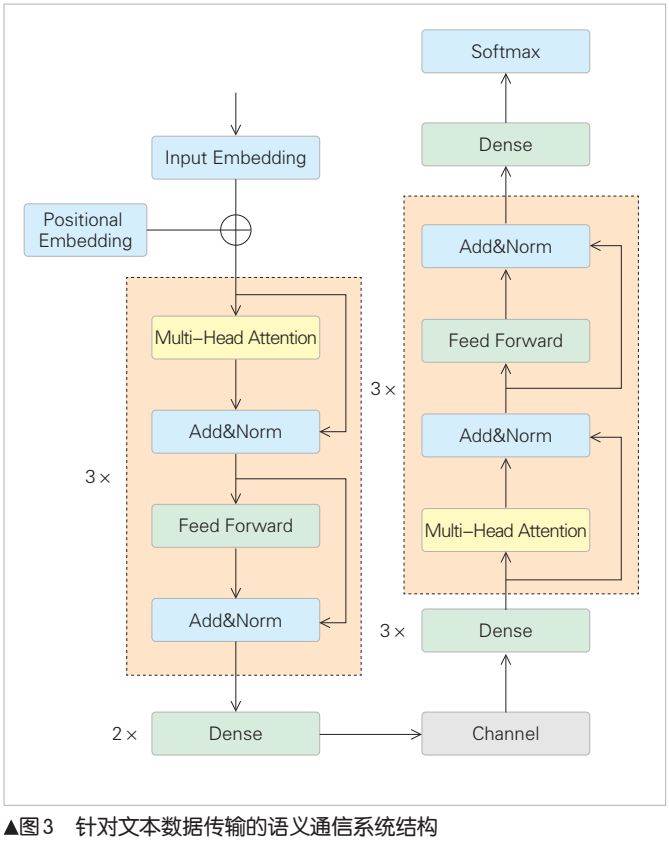
为了验证基于JSCC方式的语义通信效果，本文设计了一套基于文本传输的语义通信系统，并与经典的SSCC方式进行比较。

2.1 语义通信系统框架与传输基准

图3为针对文本数据传输的语义通信系统结构示意图。该系统采用Transformer Encoder作为语义编码器，采用Dense层控制信道输入和输出数据维度，采用Transformer Decoder作为语义译码器。具体参数见表1。

本文采用的实验数据为“欧洲议会数据集”中的语料数据，该数据集总共包括200多万条英文句子。考虑到训练语义通信系统的复杂性和硬件设施的限制，首先筛选其中长度为30个左右英文单词的句子，从而得到73 536条句子，再将其中10%作为测试集，剩余句子作为训练集，进而对语义编译码器的参数进行优化。训练阶段采用Adma优化器，学习率设置为10^{-4}，epoch设置为100，batch大小设置为64，在一块Tesla V100的图形处理器（GPU）上进行训练。测试阶段让系统在各自测试条件下分别运行，并采用贪婪搜索方式进行译码。

由于构成语义编译码器的ANN一般采用端到端联合训练，这会使得梯度经过信道回传，从而影响语义编译码器的参数优化。本文认为这是一种JSCC的语义通信系统，简称JSCC语义通信。但是，如果作为语义编译码器的ANN脱离信道训练，只参与信源编译码，这样的系统我们认为是一种SSCC的语义通信系统。在SSCC语义通信系统中，信源编码方式采用语义编译码器实现，仍然采用端到端联合训练，与JSCC不同之处在于训练时不加信道的影响。但是，在测试过程中，需要采用信道编码对抗信道噪声，信道编码仍然采用经典的编码方式。本文采用的是新空口（NR）低密度奇偶校验码（LDPC）编码，简称为两者的联合为SSCC语义通



▲图3　针对文本数据传输的语义通信系统结构

▼表1 针对文本数据传输的语义通信系统参数配置

模型	层	参数
语义编码器	Position Encoding	512
	Dropout	$p = 0.1$
	Transformer Encoder × 3	128（8 heads）
	Dense+ReLU	256
	Dense	16
	Power Normalization	$x/\Sigma x^2$
信道	AWGN	SNR：−8~16 dB
语义译码器	Dense+ReLU	128
	Dense+ReLU	512
	Dense	128
	Transformer Decoder × 3	128（8 heads）
	Dense	3 780
	Softmax	Greed search

AWGN: 加性高斯白噪声 SNR: 信噪比

信。此外，如果JSCC语义通信与经典信道编码联合实现，则将其称为融合经典信道编码的语义通信系统，简称为融合语义通信。

本文采用双语评估研究指标（BLEU）作为衡量通信效果的指标。该指标在自然语言处理中常被用来衡量翻译效果，在语义通信系统中则常被用来衡量文本传输效果，其计算公式如下：

$$\mathrm{BLEU} = \mathrm{BP} \cdot \exp\left(\sum_{n=1}^{N} \omega_n \log P_n\right), \tag{1}$$

$$\mathrm{BP} = \begin{cases} 1 & \text{如果} c > r \\ e^{1-r/c} & \text{如果} c \leqslant r, \end{cases} \tag{2}$$

其中，BP为长度惩罚因子，r为目标句子长度，c为要传输的句子长度，ω_n为n元词的权重（在本次实验中选择为$\omega_1 = 1$，即只统计一元词），P_n为传输信息中n元词在接收信息中出现的概率：

$$P_n = \frac{\sum_k \min\left(C_k(\hat{M}), C_k(M)\right)}{\sum_k \min\left(C_k(\hat{M})\right)}, \tag{3}$$

其中，C_k表示第k个n元词出现的频次。

2.2 结合经典信道编码的传输实验

为了保证对比充分，本文设计了3组实验，每组实验又分为3个小组。各组实验配置如表2所示。

每一组实验分为3个小组，总共开展9个实验。第一组实验针对原始语义通信系统进行测试，第二组实验针对量化后的语义通信系统进行测试，第三组实验针对融合NR LDPC编码的语义通信系统进行测试。实验结果如图4所示。

第一组实验采用原始的语义通信系统进行传输。该组实验分为3小组，其中实验1–1不通过信道训练语义编译码器，同时也不通过信道测试，该结果为直接进行语义编译码的数据压缩效果，作为实验对比基准。实验1–2不通过信道训练语义编译码器，但是通过信道测试，该结果为直接采用语义压缩后的传输效果。实验1–3通过信道训练语义编译码器，同时通过信道测试，该结果为JSCC语义通信传输效果。

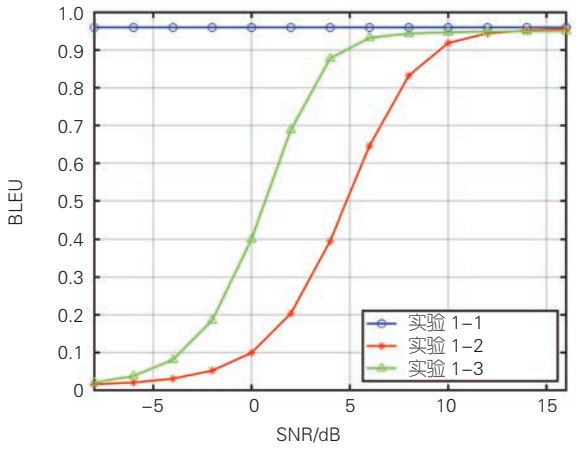
第二组实验采用量化后的语义通信系统。由于信道编码需要对01序列实现，因此需要先将语义编码的输出做量化，这里采用均匀量化方法。同时，为了保证每次实验使用的信道次数相同，第二组实验将每一个语义编码量化为2 bit，并通过16QAM调制后传输。其中，实验2–1不通过信道训练，也不通过信道测试，该结果为量化后语义编译码的数据压缩效果。实验2–2不通过信道训练，但是通过信道测试，展示了量化后语义压缩的传输效果。实验2–3通过信道训练，同时也通过信道测试，展示了量化后JSCC语义通信传输效果。

第三组实验采用与信道编码融合的语义通信系统。为了

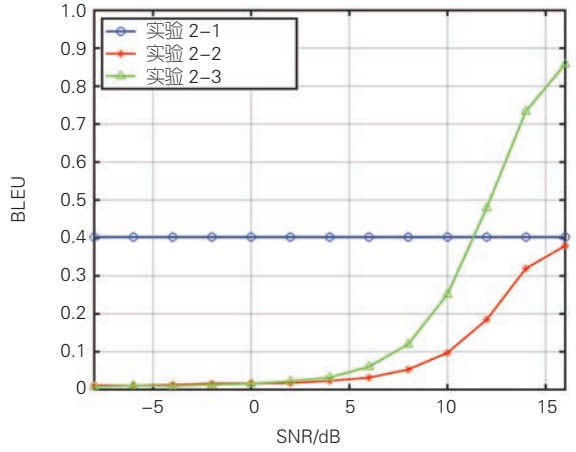
▼表2 传输实验的参数配置

实验序号		语义通信系统是否通过信道训练	量化比特数	调制	NR LDPC码率	是否通过信道测试
第一组	1–1	否	无	无	无	否
	1–2	否	无	无	无	是
	1–3	是	无	无	无	是
第二组	2–1	否	2	16QAM	无	否
	2–2	否	2	16QAM	无	是
	2–3	是	2	16QAM	无	是
第三组	3–1	否	2	64QAM	2/3	否
	3–2	否	2	64QAM	2/3	是
	3–3	是	2	64QAM	2/3	是

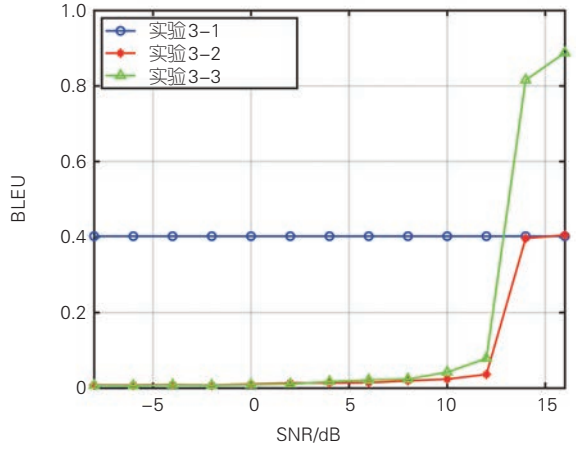
LDPC：低密度奇偶校验码 NR：新空口 QAM：正交振幅调制



（a）第一组实验结果



（b）第二组实验结果



（c）第三组实验结果

BLEU：双语评估研究指标　SNR：信噪比

▲图4　结合经典信道编码的传输实验结果

保证所有实验使用的信道次数相同，这里采用码率为2/3的NR LDPC编码，码字长度为3 072，对应的语义编码量化为2 bit，调制方式为64QAM。其中，实验3–1不通过信道训练，也不通过信道测试，该结果是与信道编码融合的语义编译码数据压缩效果。实验3–2不采用信道训练，但是采用信道测试，该结果是与信道编码融合的语义压缩后传输效果，即SSCC语义通信传输效果。实验3–3采用信道训练，同时采用信道测试，该结果是与信道编码融合后的JSCC语义通信传输效果，即融合语义通信传输效果。

图4（a）中，每组实验结果分别用不同颜色的曲线与标记表示。观察蓝线圆圈标记的实验1–1结果可知，如果直接进行语义压缩然后恢复，其BLEU能够达到0.959 9，可以认为该数值为本文中的语义通信系统能够达到的最好结果。对比绿线三角标记的实验1–2与红线星标记的实验1–3，发现通过信道传输时采用JSCC语义通信系统，比直接进行语义压缩后传输好4 dB左右。

图4（b）展示了实验2–1、2–2和2–3的结果。观察蓝线圆圈标记的实验2–1结果可知，如果直接对语义编码进行量化调制，然后解调恢复，其BLEU只能达到0.404 3，这说明2 bit量化操作产生的量化噪声极大地影响了语义压缩效果。对比绿线三角标记的实验2–2与红线星标记的实验2–3结果，发现经过量化后的JSCC语义编译码器不仅优于不通过信道训练的语义编译码器，而且优于实验2–1的结果。实验2–3的结果优于实验2–2，这是因为前者训练时考虑了信道，而后者没有任何保护。实验2–3的性能能够突破实验2–1的限制，一方面可能是因为JSCC语义通信系统能够更好地调和信源分布、信源失真、信道条件分布之间的关系（此时的信道噪声将帮助语义编码器对信源进行失真处理，而语义译码器则是基于失真数据尽可能恢复出原始数据的语义信息，信道噪声不再是需要克服的壁障，反而成为克服量化噪声的有利因素）；另一方面，量化噪声也是噪声，因此通过信道训练的JSCC语义通信具有克服一部分量化噪声的能力，最终导致实验2–3传输效果好于实验2–1。出现该现象的原因可能与本文使用的语义编码器结构和训练设置有关。这与我们在第1.2节中的理论分析结果一致。

图4（c）展示了实验3–1、3–2和3–3的结果。观察蓝线圆圈标记的实验3–1结果可知，如果直接进行语义压缩、量化调制以及信道编码，其BLEU同样达到0.404 3。这说明融入信道编码不会影响语义通信压缩性能。对比绿线三角标记的实验3–2与红线星标记的实验3–3结果可知，发现融合语义通信性能优于SSCC语义通信。该结果充分说明了JSCC方式带来了超越经典信道编码的传输性能优势，而这部分增

益很可能来自于在先验信息加持下，即语义编译码器参数优化后，对信源失真和信道条件分布的调和。此外，实验3–3的结果能超过实验3–1的上限，同样说明采用JSCC形式的语义编译码器能够克服一部分量化负增益。

3 影响融合系统性能因素分析

上述实验效果还受到量化位数、调制方式、LDPC码率等参数的影响。本节在信道使用次数固定的情况下，调整了这3个参数（见表3），分别观察其带来的影响。需要注意的是，下述每组实验均对通过信道训练的语义通信系统与不通过信道训练的语义通信系统进行了测试。

由于信道使用次数被固定（为语义编码长度的一半），因此确定量化位数与调制方式后，相当于信源需要传输的比特数目与信道能够传输的比特数目均被确定，此时LDPC所能取到的最大码率就是前者与后者的比值。因此，如表3所示，能够选择LDPC码率的只有第2、4、5组实验，其他3组实验主要体现了量化噪声和调制方式对语义通信系统的影响。

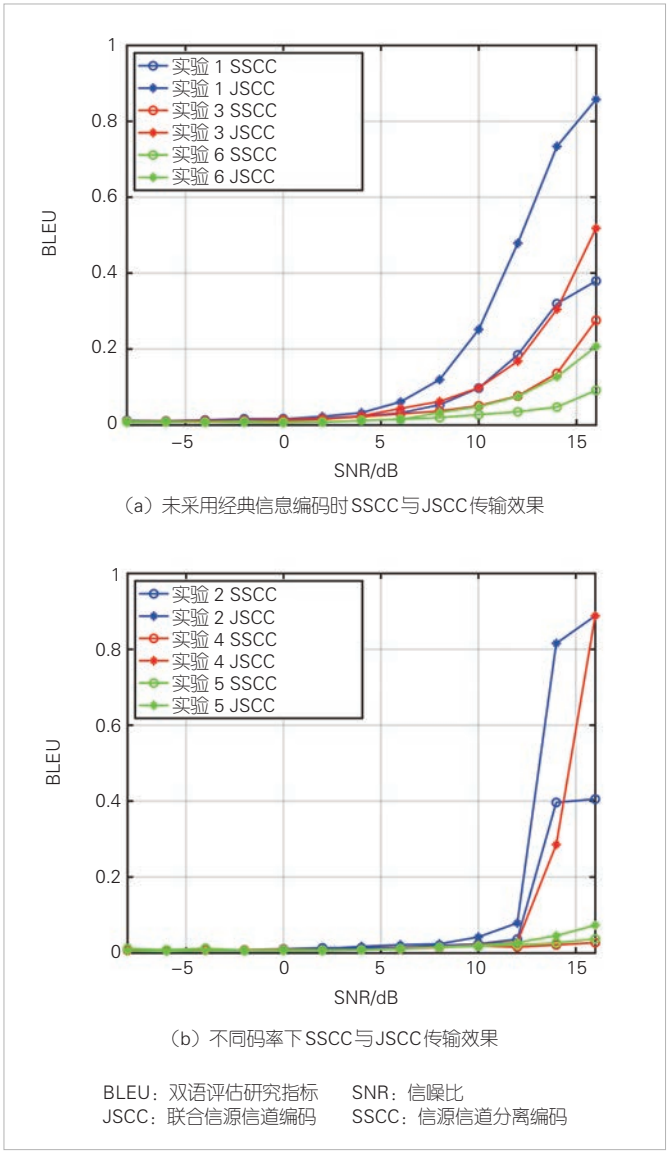
图5（a）展示了表3中第1、3、6组的实验结果，这3组实验均没有采用经典信道编码，每组实验分别使用不通过信道训练的语义通信系统（圆标的SSCC）与通过信道训练的语义通信系统（星标的JSCC）实现。图5（b）展示了表3中第2、4、5组的实验结果，这3组实验采用不同码率的NR LDPC编码与不通过信道训练的语义通信系统（圆标的SSCC）及通过信道训练的语义通信系统（星标的JSCC）融合实现。

图5（a）实验不做LDPC编码，直接反映量化和调制对传输效果带来的影响。蓝色线表示第1组实验结果，通过信道训练的语义通信系统传输效果使用标记星的线条表示，不通过信道训练的语义通信系统使用标记圈的线条表示。红线与绿线分别表示第3组与第6组实验，其中线条标记含义与第一组实验一致。实验结果显示，当量化与调制参数相同

▼表3　在信道使用次数不变的情况下，量化位数、调制方式与LDPC码率的参数配置

实验序号	量化比特数	调制	NR LDPC码率
1	2	16QAM	1
2	2	64QAM	2/3
3	3	64QAM	1
4	2	256QAM	1/2
5	3	256QAM	3/4
6	4	256QAM	1

LDPC：低密度奇偶校验码　NR：新空口　QAM：正交振幅调制



（a）未采用经典信息编码时SSCC与JSCC传输效果

（b）不同码率下SSCC与JSCC传输效果

BLEU：双语评估研究指标　SNR：信噪比
JSCC：联合信源信道编码　SSCC：信源信道分离编码

▲图5　不同条件下SSCC与联合信源信道编码(JSCC)的传输效果

时，通过信道训练的语义通信系统传输效果较好。这同样说明通过信道训练的语义编译码器能够抵消一部分量化带来的负增益。此外，当使用的语义通信系统相同时，量化比特数越大，恢复的语义编码越准确，调制阶数越高，接收符号受到信道影响越大。实验结果表明：量化比特数增大带来的增益不足以抵消调制阶数升高带来的负增益。

图5（b）展示了通过信道训练的语义通信系统和不通过信道训练的语义通信系统，与NR LDPC信道编码融合的传输效果。本文将前者简称为融合语义通信系统，后者简称为SSCC语义通信系统。其中，蓝线表示第2组实验结果，融合语义通信系统传输效果使用标记星的线条表示，SSCC语义通信系统传输效果使用标记圈的线条表示。红线与绿线

分别表示第4组与第5组实验，其线条标记含义与第2组一致。实验结果表明：融合语义通信系统传输效果好于SSCC语义通信系统。这与第1.2节的理论分析结果一致，即通过信道训练的语义编译码器，能够在知识库（优化参数）的加持下，更好地调和信源分布与信道条件分布，实现更优传输。此外，当使用的语义通信系统相同时，信道编码码率下降，克服信道噪声能力就越强，但是受到调制的极大影响，其带来的增益也不足以抵消调制阶数带来的负增益。这一点与经典通信并不一样。

综上所述，将语义编码与经典信道编码融合实现时，采用通过信道训练的语义通信系统更有优势，而且采用低阶调制才不会使得性能显著下降，另外量化位数太少所带来的负增益，也能够在一定程度上被克服。需要注意的是，图5(a)显示，使用量化位数为2、调制方式为16QAM的结果是最好的，因此该组参数也被选择进行第2节中的实验。图5(b)显示，选择量化位数为2、调制方式为64QAM、LDPC码率为2/3的结果是最好的，因此该组参数也被选择进行第2节中的实验。此外，针对本文中的语义通信系统参数配置为最佳，但不代表针对所有语义通信系统均是如此。

4 结束语

本文首先对基于JSCC方式的语义通信理论进行了详细分析。在不满足理想条件应用场景中，JSCC方式在理论上依然可以达到最优，而SSCC方式在工程实现上可能无法达到最优。此外，语义通信基于ANN对语义进行建模与提取，往往需要端到端联合训练，因此语义通信天然具有JSCC架构。进一步来说，语义通信聚焦于通信目的，具有先验知识库，数据是否失真并不影响通信过程，因此在JSCC基础上还可以进一步对信源数据进行压缩，同时也降低了信源分布与信道条件分布之间的调和难度。虽然目前语义通信理论尚不完善，但这让人们看到了打破经典通信壁垒的曙光。

其次，本文针对JSCC语义通信系统、SSCC语义通信系统，以及融合语义通信系统进行了传输效果实验对比。文中的三组实验结果显示，JSCC语义通信系统能够利用信道噪声的影响，增强语义编译码的能力，从而使得JSCC语义通信系统更容易与经典量化、调制、信道编码等方法进行融合实现。此外，实验结果还表明，影响融合最关键的因素是量化和调制，加入信道编码对语义通信系统的影响几乎可以忽略不计。

最后，本文针对语义编码与经典信道编码融合系统，进一步探索了量化位数、调制方式、信道编码码率对系统整体传输效果的影响。实验发现，影响融合系统性能最大的因素是调制阶数，即调制阶数不能太高，否则会造成严重失真。量化位数也会影响语义通信系统性能，但是其中一部分负增益可以被JSCC语义通信系统增益抵消。信道编码码率不会对融合系统的性能造成较大影响，码率较高或者较低产生的增益都无法抵消调制阶数带来的负增益。

由于目前语义通信发展处于初期，大量理论和技术实现还没有达成业界共识，本文中的实验和结论还需要做进一步延伸扩展。首先，在理论上，还未能严格证明语义通信相比经典通信具有增益，甚至关于语义通信的基础概念、基础组成模块尚有争论，还需要进一步研究。其次，文中的实验在参数选择方面相对固定。这是由于原始的语义通信系统输出语义编码长度无法调控，为保证所有实验使用的信道次数相同，导致可选择的量化位数、调制方式、信道编码码率等受到了较多限制。因此，后续还需要引入实现语义通信的其他网络框架，或者针对其他数据模态进行相应实验，以扩大实验结果的适用范围。此外，基于本文研究，后续如果要将经典编码与语义编码完全融合，还需要进一步研究经典编码过程如何进行梯度计算和回传，以便将其也纳入训练过程，或许能够进一步提升融合系统性能。最后，文中虽然采用了NR LDPC编码方式，但并没有将5G标准中的交织码、循环冗余校验等引入实验，因此对经典通信而言，可能还未发挥其最大潜力。未来还需要进一步探索其他经典通信方式中一些对语义通信系统融入有帮助的模块。

致谢

感谢中兴通讯股份有限公司算法部许进、胡留军、郁光辉、梁楚龙对本研究的帮助！

参考文献

[1] SHANNON C E. A mathematical theory of communication [J]. The bell system technical journal, 1948, 27(3): 379–423. DOI: 10.1002/j.1538–7305.1948.tb01338.x
[2] WEAVER W. Recent contributions to the mathematical theory of communication [J]. ETC: a review of general semantics, 1953, 74: 136–157
[3] BAR–HILLEL Y, CARNAP R. Semantic information [J]. The British journal for the philosophy of science, 1953, 4(14): 147–157. DOI: 10.1093/bjps/iv.14.147
[4] FLORIDI L. Outline of a theory of strongly semantic information [J]. Minds and machines, 2004, 14(2): 197–221. DOI: 10.1023/B:MIND.0000021684.50925.c9
[5] FLORIDI L. Is semantic information meaningful data? [J]. Philosophy and phenomenological research, 2005, 70(2): 351–370. DOI: 10.1111/j.1933–1592.2005.tb00531.x
[6] D’ALFONSO S. On quantifying semantic information [J]. Information, 2011, 2(1): 61–101. DOI: 10.3390/info2010061
[7] BAO J, BASU P, DEAN M K, et al. Towards a theory of semantic

communication [C]//Proceedings of IEEE Network Science Workshop. IEEE, 2011: 110–117. DOI: 10.1109/NSW.2011.6004632
[8] KOLCHINSKY A, WOLPERT D H. Semantic information, autonomous agency and non–equilibrium statistical physics [J]. Interface focus, 2018, 8(6): 20180041. DOI: 10.1098/rsfs.2018.0041
[9] WANG Y, LI H. Information theory and coding theory [M]. Beijing: Higher Education Press, 2013: 338
[10] LIN T Y, WANG Y X, LIU X Y, et al. A survey of transformers [J]. AI open, 2022, 3: 111–132. DOI: 10.1016/j.aiopen.2022.10.001
[11] ZHAO W X, ZHOU K, LI J Y, et al. A survey of large language models [EB/OL]. (2023–03–31) [2024–05–20]. http://arxiv. org/abs/2303.18223
[12] BOURTSOULATZE E, BURTH KURKA D, GÜNDÜZ D. Deep joint source–channel coding for wireless image transmission [J]. IEEE transactions on cognitive communications and networking, 2019, 5(3): 567–579. DOI: 10.1109/TCCN.2019.2919300
[13] XIE H Q, QIN Z J, LI G Y, et al. Deep learning enabled semantic communication systems [J]. IEEE transactions on signal processing, 2021, 69: 2663–2675. DOI: 10.1109/TSP.2021.3071210
[14] ZHOU Q Y, LI R P, ZHAO Z F, et al. Semantic communication with adaptive universal transformer [J]. IEEE wireless communications letters, 2022, 11(3): 453–457. DOI: 10.1109/LWC.2021.3132067
[15] DAI J C, WANG S X, TAN K L, et al. Nonlinear transform source–channel coding for semantic communications [J]. IEEE journal on selected areas in communications, 2022, 40(8): 2300–2316. DOI: 10.1109/JSAC.2022.3180802
[16] JIANG P W, WEN C K, JIN S, et al. Wireless semantic communications for video conferencing [J]. IEEE journal on selected areas in communications, 2023, 41(1): 230–244. DOI: 10.1109/JSAC.2022.3221968
[17] FENG Y L, XU J, LIANG C L, et al. Decoupling source and semantic encoding: an implementation study [J]. Electronics, 2023, 12(13): 2755. DOI: 10.3390/electronics12132755
[18] SHANNON C E. The redundancy of English [EB/OL]. [2024–05–20]. https://jontalle.web.engr.illinois.edu/uploads/537.F18/Papers/Shannon50b.pdf
[19] SHANNON C E. Coding theorems for a discrete source with a fidelity CriterionInstitute of radio engineers [EB/OL]. [2024–05–20]. https://ieeexplore.ieee.org/document/5311476
[20] GASTPAR M, RIMOLDI B, VETTERLI M. To code, or not to code: lossy source–channel communication revisited [J]. IEEE transactions on information theory, 2003, 49(5): 1147–1158. DOI: 10.1109/TIT.2003.810631

作 者 简 介

向际鹰，中兴通讯股份有限公司首席科学家；先后从事3G、4G、5G、B5G和6G相关研发工作；曾获国家科技进步奖特等奖、二等奖、技术发明奖等，并先后获得中国通信产业技术贡献人物、中华杰出工程师等称号。

段向阳，中兴通讯股份有限公司副总裁，正高级工程师，兼国家重大专项专家组成员；负责中兴通讯无线系统关键技术规划与预研工作，拥有超过20年的移动通信关键技术和产品研发经验；曾获中国电子学会科技进步奖一等奖、陕西省科技进步奖一等奖、深圳市科技进步奖一等奖。

冯雨龙，中兴通讯股份有限公司算法工程师；主要研究领域为语义通信、人工智能、机器学习；已发表论文6篇。

人工智能驱动的跨模态语义通信系统

Artificial Intelligence-Driven Cross-Modal Semantic Communication System

廖俊淇/LIAO Junqi，魏昕/WEI Xin，周亮/ZHOU Liang
（南京邮电大学，中国 南京 210003）
(Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

DOI：10.12142/ZTETJ.2024S1005
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20240724.1130.012.html
网络出版日期：2024-07-24
收稿日期：2023-12-10

摘要：概述了跨模态语义通信的相关研究背景，具体包括语义通信面临的两大挑战、跨模态通信的核心思想，以及跨模态语义通信具有的优势与存在的研究空白。针对跨模态语义通信尚存在的研究空白，在人工智能技术的驱动下，提出跨模态语义通信系统架构，详细介绍了跨模态语义通信的核心思想、关键技术，以及实践落地中需要考虑的重要因素，探讨了跨模态语义通信系统的应用场景以及存在的挑战。

关键词：跨模态语义通信；人工智能；语义关联；语义知识库

Abstract: The research background of cross-modal semantic communications is summarized, including the two major challenges faced by semantic communications, the core concepts of cross-modal communications, as well as the advantages and existing research gaps in cross-modal semantic communications. To address these gaps, a system architecture of cross-modal semantic communications driven by artificial intelligence technology is proposed. The core ideas, key technologies, and important factors to consider in the practical implementation of cross-modal semantic communications are introduced in detail. Additionally, the application scenarios and existing challenges of the cross-modal semantic communications are explored.

Keywords: cross-modal semantic communications; artificial intelligence; semantic correlation; semantic knowledge base

引用格式：廖俊淇, 魏昕, 周亮. 人工智能驱动的跨模态语义通信系统 [J]. 中兴通讯技术, 2024, 30(S1): 33-39. DOI: 10.12142/ZTETJ.2024S1005
Citation: LIAO J Q, WEI X, ZHOU L. Artificial intelligence-driven cross-modal semantic communication system [J]. ZTE technology journal, 2024, 30(S1): 33-39. DOI: 10.12142/ZTETJ.2024S1005

克劳德·香农的通信理论将通信系统分为3个层级，分别是语法、语义、语用[1]。传统通信系统属于语法层级，其目标是准确传输海量信息比特或符号。而作为第二层级的通信范式，近些年备受关注的语义通信只传输信息背后蕴含的语义。由于语义的数据量远小于符号，因而语义通信可望大幅减少通信系统以及网络的传输负担，提升传输和处理效率。语用层级从通信的目的出发，涉及信息发送者的意图、接收者的理解以及信息在特定环境中所产生的效果。与此同时，随着多模态服务的不断发展，跨模态通信技术通过深入挖掘并利用模态间的语义相关性，在模态之间进行信息交互或转换，实现了多模态信号的协同传输与处理。在此背景下，将语义通信和跨模态通信结合形成的跨模态语义通信[2]，在语义层级上进行模态间语义信息的交互或转换，可望进一步适应有限的通信与网络资源，保障用户的沉浸式体验。然而，对于跨模态语义通信的研究，在核心思想、关键技术、实践应用等方面都存在很多空白。基于此，本文在人工智能技术的驱动下，进一步探究跨模态语义通信系统。

基金项目：国家自然科学基金项目（62231017、62071254）

1 跨模态语义通信研究背景

1.1 语义通信

当前，语义通信可以进一步分为单模态语义通信和多模态语义通信。单模态语义通信主要聚焦于从文本、图像、语音、视频等其中某个模态提炼语义并进行传输，实现文本分析、图像重建、机器翻译等任务[3-5]。而多模态语义通信主要聚焦于文本和图像双模态语义信息的传输与处理[6]。

然而，当前语义通信系统的发展仍面临着两大挑战[2]：多义性和模糊性。多义性指的是发送端在没有足够背景知识

的前提下，难以准确提炼源信号所传达的含义。例如，对于“包袱很重”这句话，无法确定是指物理上的包裹还是心理上的思想负担。模糊性指的是由于传输过程中的语义噪声所导致的语义失真，使得接收端难以准确地恢复源信号的真实语义。例如，即使发送端提取了“苹果”的视觉语义特征，如形状、颜色和纹理特征，但由于语义噪声，接收端恢复出的可能是“梨”。

1.2 跨模态通信

为了支撑以音频、视频、触觉为代表的新型多媒体业务，跨模态通信应运而生[7-8]。跨模态通信旨在探索不同模态之间的潜在相关性，从而构建能够协同传输和综合处理音频、视觉和触觉信号的架构，以实现高效的音频、视频、触觉信号的传输与处理。在发送端，不同模态的信号相互协助进行压缩，以减少冗余信息的传输；在接收端，通过融合不同模态之间的相关特征来重构完整的信号，从而保障多模态服务质量，提升用户体验。

1.3 跨模态语义通信

为了解决语义通信存在多义性和模糊性两大挑战，文献[2]尝试将跨模态通信引入语义通信，首次提出跨模态语义通信的概念。跨模态语义通信充分发挥了语义通信和跨模态通信二者的优势，可望进一步满足以音频、视频、触觉为代表的新型多媒体业务对于低时延、高可靠、大容量的传输需求。然而，对于跨模态语义通信的研究，目前仍存在很多空白，例如：核心思想尚未明晰、具有可实现性的系统架构以及关键技术尚未形成、实践落地以及应用场景较少。这些仍制约着跨模态语义通信理论发展和落地应用。

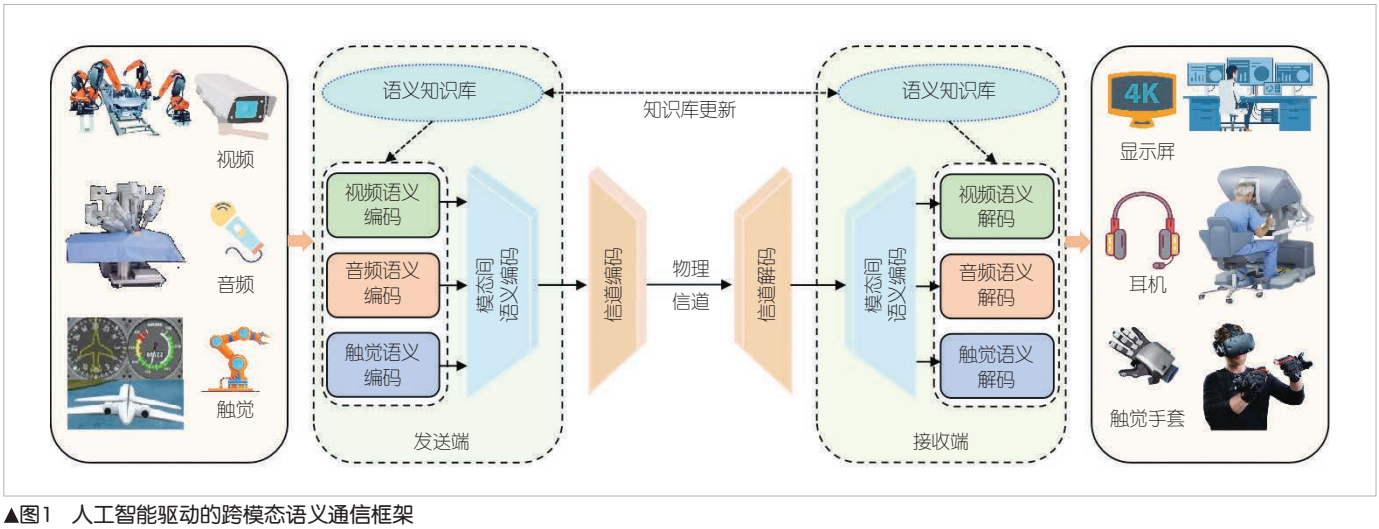
2 系统架构与关键技术

2.1 系统架构

考虑到人工智能技术的蓬勃发展以及对通信系统的加持，本文在文献[9]提出的跨模态通信框架基础上，提出人工智能驱动的跨模态语义通信架构，如图1所示。该架构由5个主要功能模块组成：模态内语义编码器、模态间语义编码器、模态内语义解码器、模态间语义解码器、语义知识库。文献[9]中信源编码被分为模态内语义编码器和模态间语义编码器，信源解码被分为模态内语义解码器和模态间语义解码器。其中，位于发送端的模态内语义编码器分别从各个模态的源信号中提取相应的语义特征；模态间语义编码器在各个模态语义特征的基础上，提炼得到模态间语义关联，并且基于该语义关联，进一步压缩各模态的语义特征（也称残留语义特征）。而后，模态间语义关联以及各模态残留语义通过物理信道，由发送端传输至接收端。在接收端，首先通过模态间语义解码器，由模态间语义关联以及各模态残留语义特征，恢复出各模态语义特征；模态内语义解码器再将各模态语义特征恢复出各模态信号。此外，语义知识库位于发送端和接收端，分别为模态内编码和模态内解码模块提供必要的背景知识。需要说明的是，与文献[2]相比，本文进一步将跨模态编解码过程分为模态内语义编解码和模态间语义编解码两个子过程，从而更加有效地压缩传输数据量和融合各模态语义特征。这样做的目的是让接收端正确理解发送端试图表达的语义信息，并尽可能准确地恢复源信号。

2.2 核心思想

对于语义通信而言，无论是单模态语义通信还是多模态



▲图1 人工智能驱动的跨模态语义通信框架

语义通信，其核心目标是从各模态信号内捕获其试图表达的“含义”，以实现有效的信息传输与接收[3-6]。这一含义可称为“模态内语义”。而对于跨模态通信而言，其核心目标是利用音频、视频、触觉信号之间的潜在相关性来实现多模态信息的高效传输与接收。这一潜在相关性可称为“模态间语义”。基于上述分析，本文认为，跨模态语义通信的核心思想正是将传统语义通信中的“模态内语义”与跨模态通信中的“模态间语义”相结合，充分利用二者优势实现高效的信息传输与接收。

传统语义通信和跨模态通信已经建立了信息理论。例如：文献[10]提出了单模态语义通信的基础理论，定义了语义信道、语义噪声、语义熵和语义信道容量的概念；文献[11]定义了跨模态通信中跨模态编码的语义熵和率失真理论。然而，跨模态语义通信理论尚未建立。基于此，参考传统语义通信以及跨模态通信，并从图1所构建的框架出发，本文认为发送端总体目标函数可定义为：

$$F_{\text{encode}} = I(S_v;\ W_{\Delta v}, W_{vah}) + I(S_a;\ W_{\Delta a}, W_{vah}) + I(S_{\text{h}};\ W_{\Delta \text{h}}, W_{vah}) + \mu \cdot \psi(I_c;\ W_{vah}, W_{\Delta v}, W_{\Delta a}, W_{\Delta \text{h}}, \delta), \quad (1)$$

其中，S_{v}、S_{a}、S_h分别表示经过模态内语义编码得到的视频、语音、触觉语义特征，$W_{\Delta \text{v}}$、$W_{\Delta a}$、$W_{\Delta h}$分别表示经过模态间语义编码后得到的各模态残留语义，W_{vah}表示模态间语义关联，I表示3个模态的模态内语义与残留语义、模态间语义关联的互信息量，I_{c}表示信道容量，δ表示模态间语义关联表征范围，ψ表示对信道容量与模态间语义关联和残留语义的约束，μ表示控制系数。在编码时，互信息量的数值越大表示语义关联程度越大，这意味着可以更大程度地压缩传输数据量。同时，ψ项表示将传输数据速率与信道容量相适应：当信道资源宽裕时，可减小语义压缩率以提高传输数据速率；当信道资源受限时，增大语义压缩率以降低传输数据速率。ψ项保证了可以在不超过信道容量的前提下，最大化传输的语义信息量。最终通过最大化目标函数F_{encode}指导发送端模态内语义编码和模态间语义编码的设计和优化。

接收端总体目标函数可定义为：

$$F_{\text{decode}} = H(\hat{W}_{vah}, \hat{W}_{\Delta v}) - H(\hat{W}_v) + H(\hat{W}_{vah}, \hat{W}_{\Delta a}) - H(\hat{W}_a) + H(\hat{W}_{vah}, \hat{W}_{\Delta h}) - H(\hat{W}_h) + \lambda \cdot d(\hat{W}_v, \hat{W}_a, \hat{W}_h;\ l), \quad (2)$$

其中，$H(\hat{W}_{vah},\hat{W}_{\Delta v})$、$H(\hat{W}_{vah},\hat{W}_{\Delta a})$、$H(\hat{W}_{vah},\hat{W}_{\Delta \text{h}})$分别表示接收的语义关联与3个模态残留语义的联合语义熵，$\hat{W}_{\text{v}}$、$\hat{W}_{\text{a}}$、$\hat{W}_h$分别经过模态间语义解码后的视频、语音、触觉模态语义特征，$H(\hat{W}_v)$、$H(\hat{W}_a)$、$H(\hat{W}_{\text{h}})$分别表示解码后的各模态语义熵，l表示公共语义标签，d表示语义判别器，λ表示控制系数。在模态间解码时，应该最小化各个模态的联合语义熵与模态内语义熵的差值，以实现模态内的语义恢复。d项用于判别3个模态的语义是否一致，以提升语义恢复质量。最终通过最小化目标函数F_{decode}指导接收端模态内语义解码器和模态间语义解码器的设计和优化。

2.3 关键技术

1）模态内语义编码：分别将各模态原始信号作为该模块的输入，以提取对应的语义特征。鉴于不同模态信号的特点，需要设计不同类型的模态内语义编码器。以视频和触觉信号传输与恢复为例，对于视频信号，可以使用卷积神经网络来提取语义特征；对于触觉信号，由于其具有序列性质，则可以使用循环神经网络来捕获语义信息[11]。此外，人工智能大模型在计算机视觉、自然语言处理等领域取得了突破性进展。本文认为人工智能大模型可以成为有效的模态内语义编码器。例如，PaLI[13]采用ViT-e模型在视频理解任务中表现出显著优势；LLaMA模型[14]在自然语言处理方面性能卓越，同样适用于处理时间序列信号。因此，ViT-e和LLaMA的注意力模块可以分别用作视频语义编码器和触觉语义编码器，如图2所示。该方案充分利用了大模型所具备的强大的语义表征能力，可以实现更加精确的语义信息提取。

2）模态间语义编码：将视频语义特征和触觉语义特征作为输入，进一步挖掘提炼二者间的潜在关联，以获得视频—触觉语义关联以及视频残留语义和触觉残留语义。在现有研究工作中，文献[11]通过手动标注语义关系矩阵获得潜在的语义关联，文献[12]采用基于注意力机制网络获得视频和触觉模态间潜在的语义关联。鉴于上述分析，本文认为采用基于Cross-Attention的Transformer结构[15]和基于Merged-Attention的Transformer结构[16-17]可以提取视频-触觉语义关联，以及视频残留语义和触觉残留语义，如图3所示。具体而言，这两种Transformer结构的核心目标是从大量语义信息中筛选出最关键部分，因而可以有效建立视频-触觉模态间的潜在关联。此外，基于视频-触觉语义关联，并且充分考虑有限的信道容量和传输资源，通过优化公式（1）中的目标函数，得到视频残留语义和触觉残留语义。

3）模态间语义解码：模态间语义解码的主要任务是将视频-触觉语义关联以及视频残留语义和触觉残留语义解码为原始的视频语义和触觉语义。考虑到传输过程中的语义噪声容易引起语义失真而导致产生语义模糊性，在模态间语义解码时引入一个基于Cross-Attention结构[15]的融合模块，在Transformer模型的加持以及自监督学习机制的引导下，分别将视频残留语义和触觉残留语义与模态间关联语义进行有机

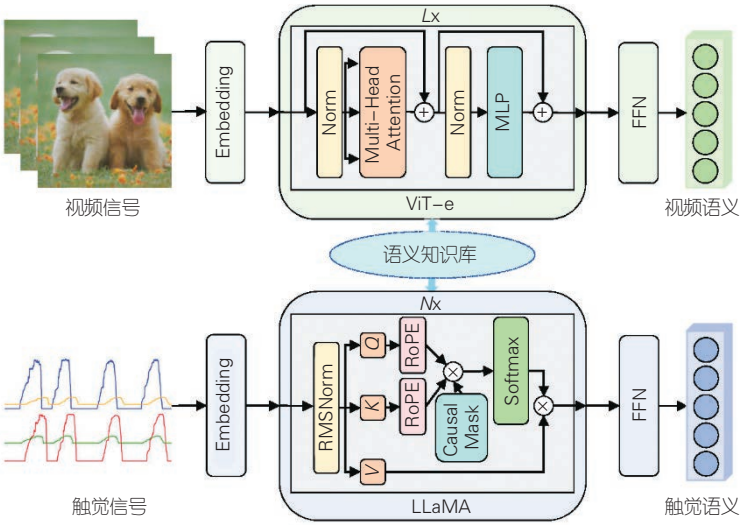
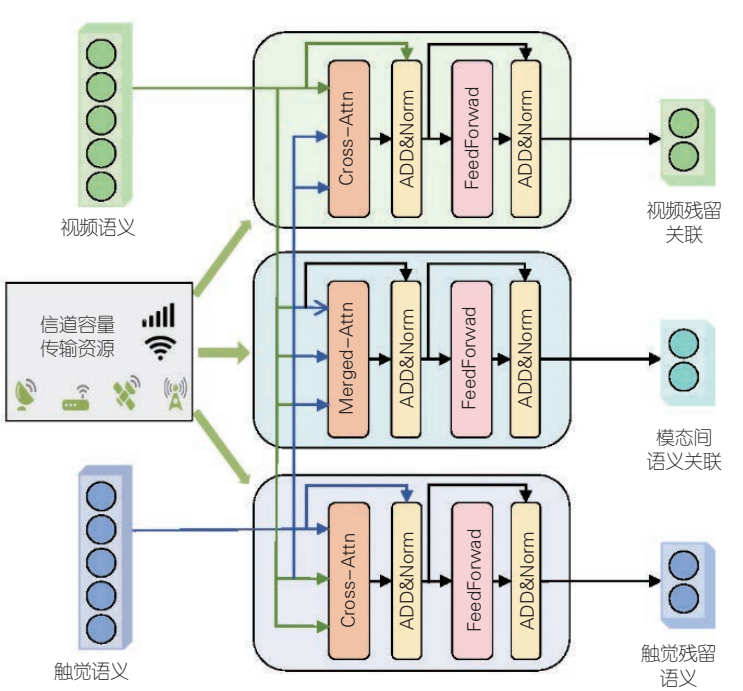


图2 模态内语义编码器



▲图3 模态间语义编码器

融合，以确保视频语义特征、触觉语义特征的恢复完整性，如图4所示。需要注意的是，这里的自监督学习机制可以基于人工标注，也可以利用触觉和视频流中的同步时间戳，或者利用来自云服务器的引导并通过云边协同等手段实现。优化公式（2）中的目标函数可恢复出视频语义特征和触觉语义特征。

4）模态内语义解码：该模块在语义库提供的相关背景知识引导下，分别将视频语义特征、触觉语义特征恢复为视频信号、触觉信号。现有研究方案中主要采用生成对抗网络方法实现该过程，如图5所示。扩散模型[18]已成功用于视频生成与恢复。基于上述分析，基于扩散模型可望更好地实现模态内语义解码。具体而言，搭建两个基于扩散模型的模态内语义解码器，分别将视频特征语义和触觉特征语义作为输入，并且利用知识蒸馏、迁移学习等技术，将语义知识库中的背景知识融入扩散模型，从而生成期望的视频信号以及触觉信号。

5）语义知识库：语义知识库分别为模态内语义编码和模态内语义解码提供了必要的背景知识。在编码阶段，基于相关背景知识系统刊能有效提取语义特征。在解码阶段，结合相关背景知识，系统可弥补语义失真和重建完整的源信号。需要强调的是，作为一种知识存储结构，跨模态语义通信中的语义知识库包括了对海量实体以及实体间关系的直观描述。得益于生成式人工智能大模型的成功应用，本文认为基于大模型的语义知识库可应用于语义通信系统，可以从大规模的语料库训练得到。一方面，利用其所蕴含的“世界知识”，可以准确提取各模态语义特征，并将这些特征隐式地存储在大模型的参数和权重中。另一方面，将其部署在现有的云边端网络架构之中，随着新信息的出现，在执行语义知识库更新时只需在边缘节点进行局部微调即可，从而最大程度地降低发送端和接收端的语义知识库的同步成本。

2.4 实践落地

此外，本文介绍几种现有的语义通信和跨模态通信平台，它们的特点和优缺点具体如表1所示。

3 应用与挑战

3.1 应用场景

基于上述分析，本文认为跨模态语义通信系统的应用场景包括如下方面：

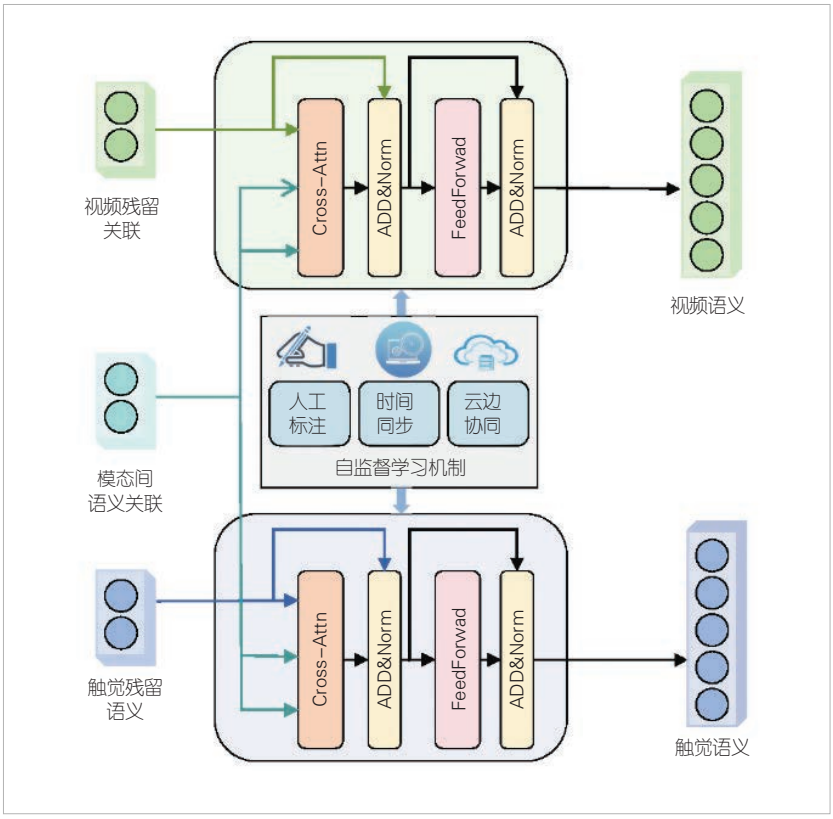
1）远程教育。在疫情期间，远程教育得到极大关注。

然而，大规模线上教学需要占用大量通信与网络资源。因此，可以将跨模态语义通信应用于远程教育，通过传输数据量较少的语义特征缓解通信压力，并通过融合多个模态的媒体流以增强学习效果。特别是对于网络资源受限的边远地区，跨模态语义通信是一个非常有前景的解决方案。

2）军事远程救护。在远程救护中，通过在战场端采集伤员的视频和触觉信号，传输到远端的救护中心，通过远程操控及时救护伤员。然而，在军事场景中，通信容易受到电磁干扰，带宽往往在kB/s级别，难以传输符号级别的多媒体信号。因此，可以利用跨模态语义通信系统通过传输语义特征，能够以较小的带宽完成传输任务。

3）远程康复训练。在现有的远程诊断基础上引入触觉感官信息，有助于医生更全面地了解患者病情。然而，实时传输多模态流需要大量带宽，这会对网络造成压力。通过利用跨模态语义通信系统来传输这些多媒体流，并在背景知识的辅助下进行语义压缩和重建，提升现有康复训练质量和医患满意度。

4）远程工业操控。远程工业操控利用远程技术和自动化系统来监控、操作和控制工业设备、过程和系统，可以提高工业领域的效率、安全性和可持续性。然而，远程工业中大量的传感器需要传输海量数据，将跨模态语义通信系统应用于远程工业操控，可实现视频和触觉信号的高效传输和精确处理，有效提高控制器的交互效果。



▲图4　模态间语义解码器

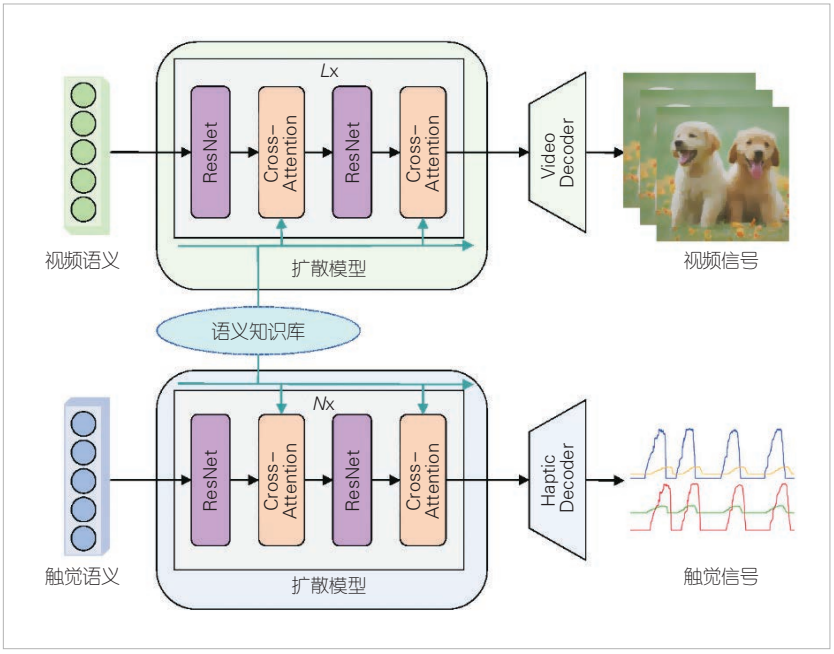
3.2 存在的挑战

作为一种新型的通信范式，跨模态语义通信可望在6G时代的多媒体业务中发挥更大的支撑作用。在未来跨模态语义通信仍存在较多技术挑战，具体如下：

首先，跨模态语义通信不是语义通信和跨

▼表1　语义通信和跨模态通信平台

平台名称	特点	优点	缺点
用于表面缺陷检测任务的语义通信原型[3]	由摄像机、边缘服务器、一组通用软件无线电设备(USRP)和天线组成，基于用户数据报协议(UDP)传输	可用于热轧钢条表面缺陷检测，取代主观和重复的人工检测过程	仅使用视频模态信号来检测缺陷，其精确度受限
面向任务的实时移动语义通信系统原型[5]	收发端用户由树莓派、Wi-Fi模块和显示屏组成；能实现语义编解码和特征选择，通过Wi-Fi模块实现传输	提高对语义信息歧义的鲁棒性，只选择与任务相关的语义信息进行传输，进一步降低通信成本	仅考虑视频模态的单任务语义通信，无法面向通用任务；没有考虑语义传输过程中的数据安全性问题
用于文本图像查询的多用户语义通信系统[6]	两个单天线用户作为发送端，一个多天线用户作为接收端；把语义特征转成了复数值然后通过信道	对于图像传输，显著降低了传输符号的数量和计算复杂度，可节省图像的传输和处理时间	对于文本传输，需要传输更多的符号，稍微牺牲文本的传输时间
针灸技能训练的虚拟交互平台[8]	该平台包括3个组成部分：生动的触觉渲染、增强现实技术处理和一个技能评估子系统	方便实施远程教学，特别是那些需要实际操作或实验的课程	基于符号级别的跨模态传输方案，其传输数据量仍很大；其互动性仍受限制，仍难以完全模拟真实的针灸操作体验
视觉触觉人机交互系统[8]	由机械手臂、基于直线伺服驱动的远程人机交互触觉感知手套和Kinect相机组成	利用视频信号补偿触觉信号损伤，利用跨模态信号重构技术，可以进一步提高人机交互的可靠性	基于符号级别的传输以及跨模态信号重构时，可能引入延迟，可能难以满足超低时延要求



▲图5 模态内语义解码器

模态通信的简单叠加，因此，如何由二者的信息理论出发，将其有机融合，深化并完善适合跨模态语义通信自身特点的信息熵理论，是需要进一步探讨的问题。

其次，本文提出的跨模态语义通信架构及其关键技术是把模态内语义编解码和模态间语义编解码分开考虑的，虽然其具有很好的可解释性，但效率仍相对较低。因此，如何将模态内与模态间语义编解码以及语义传输联合优化，进一步提升通信效率，值得深入研究。

最后，在语义编解码以及传输过程中，内外部攻击以及语义知识库的访问和共享会带来信息安全问题。因此，如何保护传输过程中的信息隐私泄露和语义知识库的安全，也是跨模态语义通信发展所面临的关键挑战。

4 结束语

本文深入探讨了人工智能驱动的跨模态语义通信系统，对跨模态语义通信的相关背景进行了概述，构建了跨模态语义通信的架构，并且明晰了跨模态语义通信的核心思想、关键技术以及实践落地所需要重点考虑的因素。跨模态语义通信将在6G中扮演重要角色，但也面临一些技术挑战。未来将继续深入研究跨模态语义通信的信息熵理论，为融合更多感知模态提供理论指导；联合优化跨模态语义编解码和语义传输，提升端到端传输的效率；探索可靠的语义传输安全机制，保护传输过程中的信息泄露和语义知识库的安全。

参考文献

[1] SHANNON C E, WEAVER W. The mathematical theory of communication [M]. Urbana: University of Illinois Press, 1949
[2] LI A, WEI X, WU D, et al. Cross-modal semantic communications [J]. IEEE wireless communications, 2022, 29(6): 144-151. DOI: 10.1109/MWC.008.2200180
[3] YANG Y, GUO C L, LIU F F, et al. Semantic communications with AI tasks [EB/OL]. [2023-06-05]. http://arxiv.org/abs/2109.14170
[4] FENG Y L, XU J, LIANG C L, et al. Decoupling source and semantic encoding: an implementation study [J]. Electronics, 2023, 12 (13): 2755. DOI: 10.3390/electronics12132755
[5] MA S, QIAO W N, WU Y L, et al. Task-oriented explainable semantic communications [J]. IEEE transactions on wireless communications, 2023, 22(12): 9248-9262. DOI: 10.1109/TWC.2023.3269444
[6] XIE H Q, QIN Z J, LI G Y. Task-oriented multi-user semantic communications for VQA [J]. IEEE wireless communications letters, 2022, 11(3): 553-557. DOI: 10.1109/LWC.2021.3136045
[7] ZHOU L, WU D, CHEN J X, et al. Cross-modal collaborative communications [J]. IEEE wireless communications, 2020, 27(2): 112-117. DOI: 10.1109/MWC.001.1900201
[8] WEI X, WU D, ZHOU L, et al. Cross-modal communication technology: A survey [J]. Fundamental research, 2023. DOI: 10.1016/j.fmre.2023.08.00
[9] WEI X, ZHANG M, ZHOU L. Cross-modal transmission strategy [J]. IEEE transactions on circuits and systems for video technology, 2022, 32(6): 3991-4003. DOI: 10.1109/TCSVT.2021.3105130
[10] BAO J, BASU P, DEAN M K, et al. Towards a theory of semantic communication [C]//Proceedings of IEEE Network Science Workshop. IEEE, 2011: 110-117. DOI: 10.1109/NSW.2011.6004632
[11] YUAN Z, KANG B, WEI X, et al. Exploring the benefits of cross-modal coding [J]. IEEE transactions on circuits and systems for video technology, 2022, 32(12): 8781-8794. DOI: 10.1109/TCSVT.2022.3196586
[12] ALAMEH M, ABBASS Y, IBRAHIM A, et al. Touch modality classification using recurrent neural networks [J]. IEEE sensors journal, 2021, 21(8): 9983-9993. DOI: 10.1109/JSEN.2021.3055565
[13] CHEN X, WANG X, CHANGPINYO S, et al. PaLI: A Jointly-scaled multilingual language-image model [EB/OL]. (2022-09-14) [2023-06-05]. https://arxiv.org/abs/2209.06794
[14] TOUVRON H, LAVRIL T, IZACARD G, et al. Llama: open and efficient foundation language models [EB/OL]. [2023-03-27]. https://arxiv.org/abs/2302.13971
[15] TAN H, and BANSAL M. LXMERT: learning cross modality encoder representations from transformers [EB/OL]. (2019-08-20) [2022-10-03]. https://arxiv.org/abs/1908.07490
[16] SU W, ZHU X, CAO Y, et al. VL-BERT: Pre-training of generic visual-linguistic representations [EB/OL]. (2019-08-22) [2022-02-18]. https://arxiv.org/abs/1908.08530
[17] DOU Z Y, XU Y C, GAN Z, et al. An empirical study of training end-to-end vision-and-language transformers [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 18145-18155. DOI: 10.1109/CVPR52688.2022.01763

[18] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 10674-10685. DOI: 10.1109/CVPR52688.2022.01042

作 者 简 介

廖俊淇，南京邮电大学在读博士研究生；主要研究方向为多媒体通信、多媒体大数据分析与处理。

魏昕，南京邮电大学教授、博士生导师；主要研究方向为多媒体通信与信息处理、教育信息化；主持多项国家自然科学基金以及产学研合作项目；发表学术论文70余篇；出版英文学术专著2部，获得授权中国发明专利30余项、美国发明专利2项，其中8项已实现成果转化。

周亮，南京邮电大学副校长、教授、博士生导师，教育部宽带无线通信与传感网技术重点实验室主任；主要研究领域为多媒体通信；先后获教育部“长江学者奖励计划”特聘教授、中共中央组织部“海外高层次青年专家”等荣誉称号，获国家自然科学基金委员会“优秀青年基金”资助；作为项目负责人主持多项国家级重点科技攻关项目。

具身智能机器人技术

Embodied Intelligent Robotics

邵宏/SHAO Hong，谢大雄/XIE Daxiong
（中兴通讯股份有限公司，中国 深圳 518057）
(ZTE Corporation, Shenzhen 518057, China)

DOI：10.12142/ZTETJ.2024S1006
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20240724.1458.018.html
网络出版日期：2024-07-25
收稿日期：2023-11-25

摘要：提出了一种用于智能制造的具身智能机器人技术。提出的具身智能机器人通过主动感知环境、自主学习和自主决策来执行拟人化任务，其“大脑”是以强化学习为核心的智能体，感知部分具有双目视觉、空间六维力感知和本体状态感知等多模态感知能力；通过“感知-行动”反馈环，构建了一个控制周期为1 ms的机器人实时控制系统。在中兴通讯的5G智能制造工厂中部署了具身智能机器人进行实际生产，用来取代工人插拔5G小站产品的RJ45插头和光模块。实践表明，具身智能可打通全自动化生产线的最后断点，是一项在智能制造中有广阔应用前景的机器人技术。

关键词：具身智能；强化学习；机器人；多模态感知

Abstract: An embodied intelligent robot technology in industrial intelligent manufacturing is proposed. The robots can learn the policy, make decisions, and act autonomously to complete anthropomorphic tasks through active perception and environmental interaction. The embodied intelligent robot uses a reinforcement learning agent as its “brain” with a real-time control system of 1 ms control cycle with a multi-modal "perception-action" feedback loop, which includes stereo vision, spatial six-dimensional force sensor, and robot proprioception. The system was subsequently deployed in our 5G intelligent manufacturing factory to replace workers, to plug and unplug network RJ45 crystal heads and optical modules for 5G small station manufacture, which eliminated our last breakpoint of automation. The practice shows that embodied intelligence is a promising robot technology in the future manufacturing industry.

Keywords: embodied AI; reinforcement learning; robotic; multimodal perception

引用格式：邵宏, 谢大雄. 具身智能机器人技术 [J]. 中兴通讯技术, 2024, 30(S1): 40-44. DOI: 10.12142/ZTETJ.2024S1006
Citation: SHAO H, XIE D X. Embodied intelligent robotics [J]. ZTE technology journal, 2024, 30(S1): 40-44. DOI: 10.12142/ZTETJ.2024S1006

在生产制造领域，随着自动化程度的提升，越来越多的人工工位由工业机器人和自动化站点来替代，但仍有一些操作无法采用传统工业机器人那种按预定轨迹执行任务的方式完成，如线缆插拔、光模块安装、气密堵头装配、脆弱物体安装等。因为器件的受力情况和运动轨迹在操作过程中需要实时变化，所以这些操作不能由传统机器人按预定轨迹执行，仍然需要人工操作，是全自动化生产线中的众多“断点”。要实现完全无人化的“黑灯工厂”，我们需要一种全新的具有感知、决策和控制能力的智能机器人技术，即具身智能（Embodied AI）机器人来完成这类柔性装配任务[1]，打通自动化生产线的“断点”。

当前业界对具身智能机器人尚没有统一或确切的定义，一般理解为有身体并支持物理交互的智能体，比如人形机器人。具身智能机器人的核心是在“智”而不在“形”，特别是在工业场景中，机器人可以是各种构形的，其核心是可以像人一样主动感知和决策的“大脑”，使它们能够与环境交互以完成拟人任务。上海交通大学卢策吾教授提出，具身智能不仅仅是具有物理身体，而且是具有与人一样的身体体验的能力。

目前的机器人智能技术广泛采用有监督学习方式，例如用机器视觉引导的智能机器人系统，利用深度神经网络完成目标识别、分割等任务。有监督学习的训练是在带有Ground Truth标签的数据集上进行的，是一种旁观标签学习方式，并非通过主动体验进行学习，因此很难通过这种学习方式发展出未来的具身智能。发展具身智能的关键路径是在与环境的交互中实现主动感知与行动的反馈环。虽然以具身智能机器人第一视角得到的数据不够稳定，但这种反馈环可以帮助机器人解决更多真实问题，特别是完成柔性装配任务。这更类似强化学习（RL）的感知和行动过程[2]。

1 系统模型

为了建立“感知-行动”反馈环模型，本文采用强化学习来构建具身智能机器人的系统。感知部分结合了双目视觉

摄像头获取的视觉信息、机器人末端空间力反馈、机器人本体状态数据等，作为机器人智能体对环境的观察变量。行动部分是由智能体周期性地对机器人输出实时控制指令。

本文将具身智能机器人控制系统简化为一个无限步数的部分可观测马尔可夫决策过程（POMDP）[3-4]，这个过程可参数化描述为元组$(\mathcal{O},\mathcal{A},p,r,\gamma)$。其中$\mathcal{O}$为多模态感知状态空间，$\mathcal{A}$为动作空间，观测状态转移概率$Pr(o_t'|o_{\leqslant t},a_t)$为当前动作$a_t$执行后从包括$t$时刻及以前过往观测感知状态迁移到下一观测感知状态$o_t'$的概率，$r:\mathcal{O}\times\mathcal{A}\rightarrow\mathbb{R}$为当前的观测感知和动作的奖励函数，而$\gamma$为折扣系数。

考虑到机器人动作时，由双目视觉摄像头视觉信息、机器人末端空间力反馈和机器人本体状态数据等组成的观测感知，可看作一系列相关联的连续过程。我们可以将其联合作为t时刻的状态变量s_t构成状态空间$\mathcal{S}$，其中：

$$s_t=\{o_t,o_{t-1},o_{t-2},\cdots\}。\tag{1}$$

这样我们就将POMDP过程转化为标准的马尔可夫决策过程（MDP）[5]，可以用元组$(\mathcal{S},\mathcal{A},p,r,\gamma)$描述。这时其状态转移概率$p$和奖励函数$r_t$分别为：

$$p=Pr(s_t'|s_t,a_t),\tag{2}$$

$$r_t=r(s_t,a_t)。\tag{3}$$

此时控制系统的目标就是找到最佳决策$\pi(a_t|s_t)$使得累计折扣奖励最大化：

$$J(\pi)=\max_\pi[\mathbb{E}[\sum_{t=1}^{\infty}\gamma^t r_t|a_t\sim\pi(\cdot|s_t),\ s_t'\sim p(\cdot|s_t,a_t),s_1\sim p(\cdot)])]。\tag{4}$$

最佳决策可以采用基于神经网络的RL算法通过环境交互来训练。为了得到较好的环境适应能力，通常采用无模型(model-free)的RL算法来优化策略函数π。对于实际智能制造环境中的机器人来说，由于在实际环境中进行强化学习训练的成本比较高，因此需要高效利用样本。比较好的方式是采用Off-policy的RL算法而不是On-policy算法。Off-policy的RL方法，如Double Q-learning[6]深度确定策略梯度（DDPG）[7]和SAC（Soft Actor-Critic）[8]等算法，是用带有随机性的行为策略对环境进行探索获取学习样本，然后使用目标策略在行为策略收集交互样本数据上进行训练，最终找到最优策略。

本文提出的具身智能机器人的系统框架如图1所示。

这里采用在真实机器人控制任务中性能优秀、表现稳定的SAC算法。其在优化策略以获取更高累计收益的同时，也会最大化策略的熵，其价值评判（Critic）函数$Q_\theta(s_t,a_t)$和动作策略（Actor）函数$\pi_\theta(a_t|s_t)$分别由不同参数θ的神经网络模型来拟合，并使用温度控制系数α优化寻找经γ折扣后的最大熵来训练模型参数。

在训练时，s_t是从机器人与环境交互产生的过往数据中抽取出的，而a_t是从当前的策略中采样得来。SAC算法的动作策略函数$\pi_\theta(a_t|s_t)$输出是一个关于动作的分布，用tanh-高斯分布的均值μ_θ和标准差σ_θ来控制，而执行确切的动作时，则对均值和标准差的高斯分布进行一次采样，将采样结果a_t作为策略的决策动作：

$$a_t=\tanh(\mu_\theta(s_t)+\sigma_\theta(s_t)\epsilon),\epsilon\sim\mathcal{N}(0,1)。\tag{5}$$

实际任务的复杂性和控制的鲁棒性对于具身智能机器人仍然是一个很大挑战，所以本文在RL模型之外，增加了预设的动作轨迹规划a_p并和RL算法输出的a_t叠加，从而减少π搜索的空间,叠加动作a_u我们可以用下式来描述:

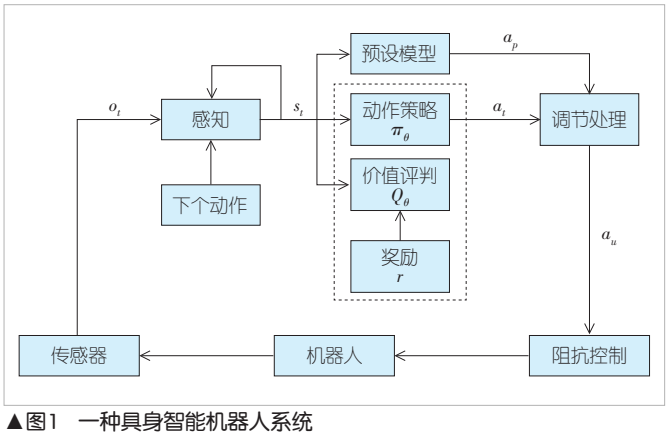
$$a_u=(1-\beta)a_p+\beta a_t,\ \beta\in[0,1],\tag{6}$$

其中，β为模型调节参数。极端情况下，β取值为0时，系统完全工作在基于预先模型的规划动作状态；当取值为1时，系统完全工作在无模型的强化学习状态。

2 多模态感知神经网络结构

通常来说，具身智能机器人对环境的多模态感知会包含视觉、触觉、声音和对自然语言的理解等，但对于制造业中能够完成柔性装配任务的具身智能机器人来说，除了机器人本体姿态感知外，视觉和空间力感知也是最基本的观测变量。这是因为视觉感知可以提供环境和操作工件空间状态信息，空间力感知可以反馈机器人在装配接触过程末端的受力状态。

相较于较低维度的机器人本体状态信息，来自双目视觉摄像头的视觉信息具有很高的维度。因此，在进行多模态感



▲图1　一种具身智能机器人系统

知设计时，我们需要对低维度张量和高维度张量采用不同的处理方法。深度神经网络能够很好地进行高维度的表征学习，但模型训练比较困难，难以在强化学习中直接训练，对此可以在强化学习之前采用自监督学习或自编码器训练一个预训练网络[9]。

我们设计了一种多模态感知神经网络，如图2所示。感知内容主要包括3个部分：双目视觉感知、末端六维力感知、本体状态感知。我们采用带有关节力矩感知器件的7轴机器人作为本体，用装于机器人腕部的双目视觉摄像头作为视觉传感器，结合力反馈和机器人本体状态等低维度信息，以感知更丰富、更深入的周围环境信息。

根据公式（1），将输入传感器感知按照时间序列以8帧$(o_t,o_{t+1},\cdots,o_{t+7})$堆叠后，作为输入的状态变量。

在低维度信息的输入处理部分，本体状态信息自编码器从本体的状态信号中提取出机器人末端工具的位置信息和四元数姿态共7维信息，形成7×8的时间序列帧，然后通过5层因果卷积网络[10]转换为12维特征向量。力反馈自编码器对末端六维力反馈进行处理，读取最近的8帧末端6维F/T力矩，形成8×6时间序列，通过一个4层因果卷积网络[7]转换为12维特征向量。具体实现时，为了训练本体状态信息自编码器和力反馈自编码器，我们用转置反卷积构建了解码器，用采集的数据对两个自编码器进行了预训练。

在高维度的视觉信息处理部分，我们将双目视觉摄像头的图像信息截取为84×84×3的图像块，按8帧进行堆叠，然后输入到4层的卷积神经网络（CNN），再经过1个线性连接层感知机（MLP），最终输出50维的特征向量。为了加快收敛，在训练时参考DrQ算法[11]对图像进行了移动增强处理。

最终上述两部分向量会拼接融合为一个向量作为多模态表征，即动作策略网络和价值评判网络的状态输入维度。动作策略网络输出动作维度为6维的末端笛卡尔坐标变化量，经过控制器生成阻抗控制的动作轨迹，实时控制机器人动作。

3 控制器设计

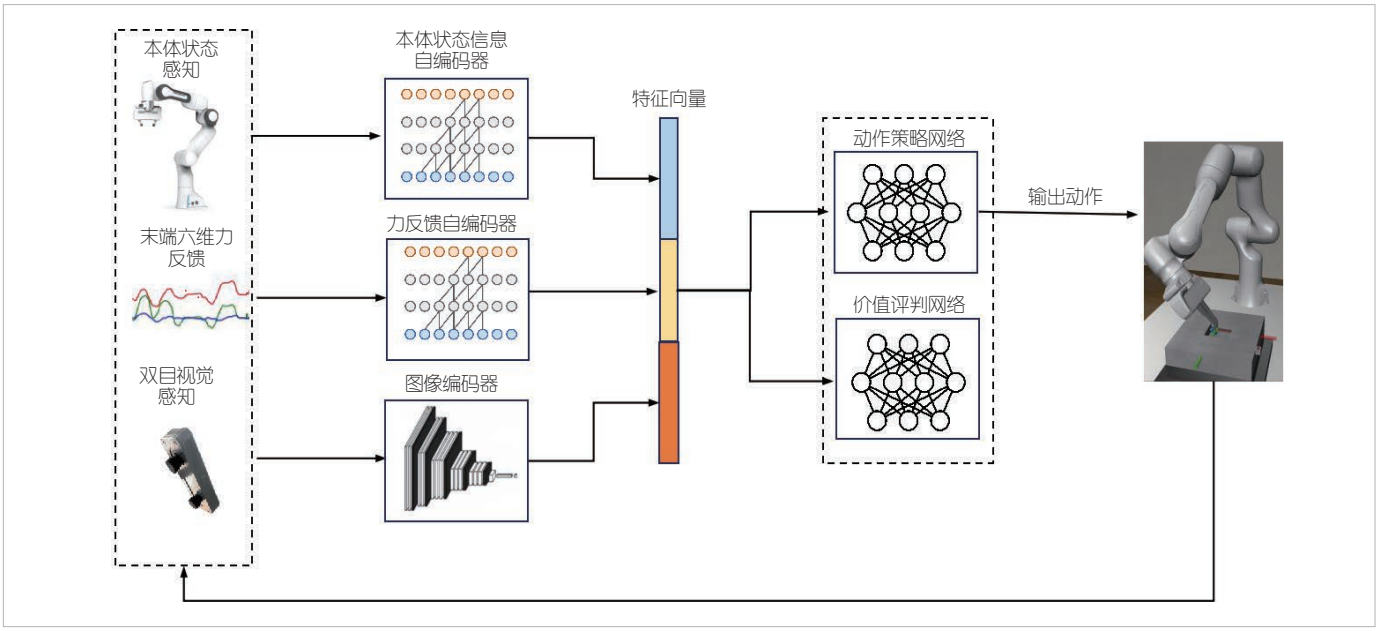
当周期性的多模态感知信息输入到智能体后，智能体向控制器输出20 Hz的末端笛卡尔坐标增量信号a_t，然后由控制器生成1 kHz的机器人关节直接控制力矩τ_u，即控制器将较低带宽的控制策略输出信号转换为高带宽的机器人控制指令，如图3所示。本文设计了阻抗式PID控制器，它以1 kHz频率实时获取机器人当前的末端位置x_t，用x_t+a_t计算出末端期望位置P_d，然后用1 ms的周期做线性插值到X_d，最后阻抗式PID控制器输出控制力矩τ_u：

$$\tau_u = J^T(q)\Lambda\left[P_d - K_p(X - X_d) - K_v(V - V_d)\right], \tag{7}$$

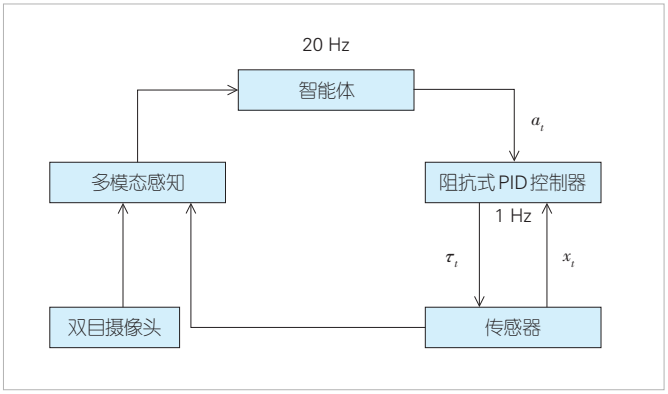
其中，J为实时读取关节角为q的机器人动力学模型和雅可比矩阵，Λ为惯量矩阵，X为机器人末端当前位置，V为空间速度。K_p和K_v分别为可以调节的刚性和阻尼系数。

4 系统仿真与真实环境测试

为了对具身智能机器人系统进行了验证，特别是评估其能否完成5G小站自动插拔线缆等柔性任务，我们搭建了仿



▲图2 多模态感知神经网络设计



▲图3　机器人控制器设计

真环境，使用了Franka七自由度机器人。该机器人具有内置的关节力矩传感器，可以实时给出末端的空间六维力感知数据和机器人的本体姿态数据。

4.1 奖励设计

在插拔5G小站RJ45插头任务时，系统主要依赖根据RJ45插头是否成功插入插口设置稀疏奖励，但为了提高学习效率，也加入密集奖励部分，我们设计了下述奖励函数：

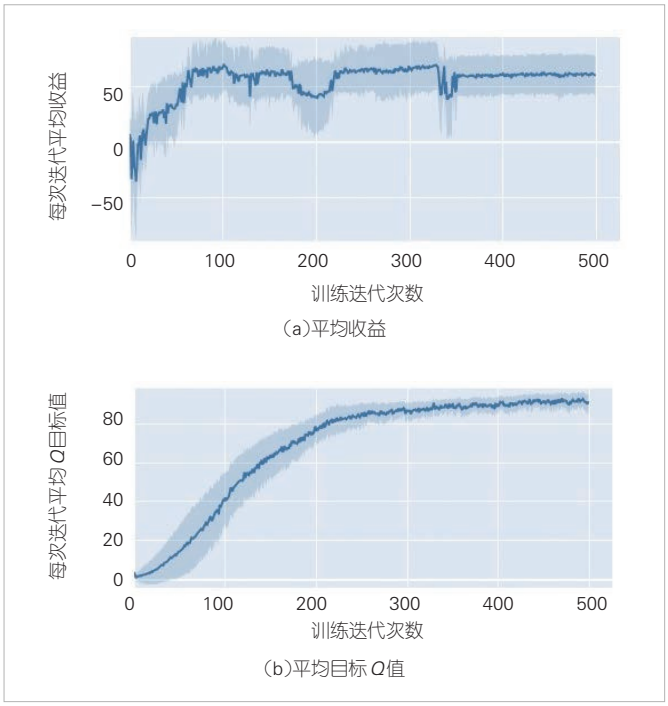
$$r(\mathrm{s})=\begin{cases}\lambda_1\left(c_a-l_x\right), & \text{当 } l_x \geqslant d \text{ 时}\\ \lambda_2\left(c_b-l_x\right), & \text{当 } l_x<d \text{ 且末端六维力超过阈值时}\\ 200, & \text{当Yolo网络判断已经插入成功时}\end{cases}, \quad (8)$$

其中，在机器人工具坐标中插头的初始位姿与插入后目标位姿的相对位姿为$(l_x,l_y,l_z,0,0,0)$，d为RJ45插口的深度，式中c_a和c_b为常数，λ_1和λ_2为缩放比例系数。我们根据收到的机器人末端的空间六维力是否超过阈值，来判断操作过程中插头是否成功插入插口。

为了准确判断是否插入成功，我们另外预训练了一个单独的用于判别是否插入成功的目标识别Yolo网络[12]，并采集双目摄像头的视觉信息，做了有监督的预训练。后续在真实的机器人部署环境中，为了增加生产可靠性，我们还对已插入的插头增加了回拉操作，确认插头已经可靠地插入后再给出稀疏奖励信号。

4.2 仿真模拟

仿真模拟中，动作策略网络和价值评判网络的训练采用Adam优化器，批大小设置为128，SAC的软目标更新率设置为0.005。我们训练了500次迭代（epochs），每次迭代2 500步，采用随机种子进行了10次实验，取平均收益和平均目标Q值，如图4所示。仿真结果表明，系统已经能够很好地完成自动插拔任务。

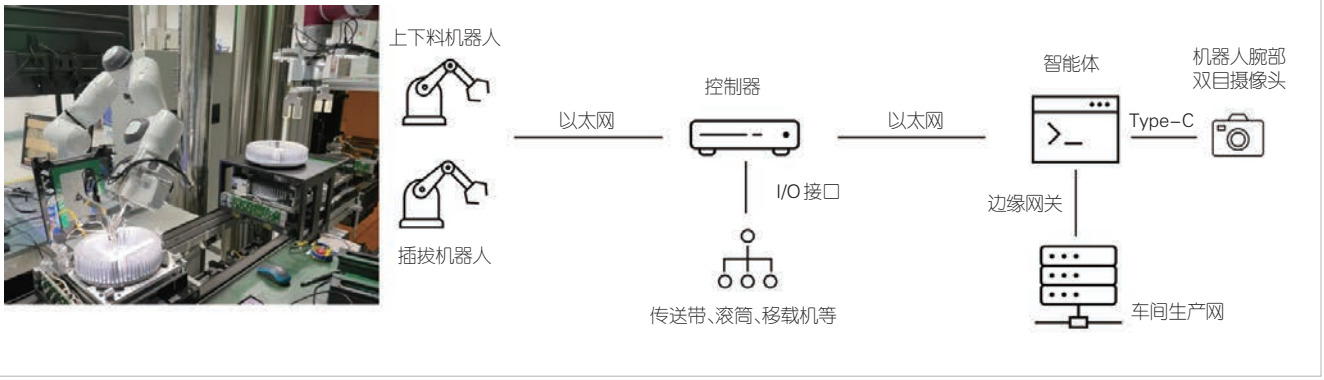


▲图4　仿真强化学习训练结果

4.3 真实生产环境测试

我们将训练后的模型迁移到5G小站生产的器件插拔自动化工位上进行应用和试验。自动插拔工位由两台协作机器人组成，其中一台是与仿真实验相同的Franka机器人，负责网线、光模块的插拔，另外一台机器人则负责自动上下料。自动化插拔工位和后序的环岛测试台的可编程逻辑控制器（PLC）对接以完成整机自动化测试。工位的输入输出（I/O）部分还需要对接控制滚筒和移载机等设备，自动呼叫自动导引车（AGV）接送料等，如图5所示。测试时，为了达到实时性要求，机器人控制器采用打上实时（RT）补丁的Linux操作系统，机器人和控制器之间采用用户数据报协议（UDP）通信，其控制指令的周期为1 ms,并且其要求往返时延（RTT）小于300 μs。超过这个时延的实时数据包将被丢弃。时延连续超过一定阈值，机器人将会停止动作并告警。为了保证机器人控制的稳定性，现场网络必须满足低时延要求，因此测试中采用了低时延的以太网交换机。

在真实生产环境中，RJ45插头插拔和光模块安装任务由对应强化学习模型2个实例化的策略网络来完成，采样和训练分成两个线程并行处理。每次任务操作固定为500步采样，并在每次采样后都进行训练。经过两天的实际训练，机器人能够可靠地自动完成RJ45插头和光模块的插入操作。通过调整上述公式（6）中的参数β可以很好地将有模型和无模型的控制有效结合：在末端工具到达接触范围之前的动



▲图5　具身机器人在实际生产环境中的部署

作中设定β的值为0，到达接触的范围时，将β设定为小于1的一个合适的值，这样就能可靠地批量作业。

5 结束语

具身智能机器人是数字世界融入现实世界的载体，将会成为未来的主流机器人技术之一。通过增加智能体大脑，传统的机器人升级为能够在与环境的互动中进行主动感知和学习的具身智能机器人，并通过获得完成任务后的奖励，在生产制造中不断学习来完成拟人化的任务。本文中的具身智能机器人的设计和应用实践表明，这一技术能够有效打通传统自动化生产中的“断点”，大大提高生产机器人的智能化程度。

参考文献

[1] LIAN W Z, KELCH T, HOLZ D, et al. Benchmarking off-the-shelf solutions to robotic assembly tasks [EB/OL]. [2024-01-15]. http://arxiv.org/abs/2103.05140
[2] SUTTON R S, BARTO A G. Reinforcement learning: an introduction, 2nd edition [M]. London: The MIT Press, 2015
[3] KAELBLING L P, LITTMAN M L, CASSANDRA A R. Planning and acting in partially observable stochastic domains [J]. Artificial intelligence, 1998, 101(1): 99-134
[4] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning [EB/OL]. [2024-01-15]. http://arxiv.org/abs/1312.5602
[5] KOSTRIKOV I, YARATS D, FERGUS R. Image augmentation is all you need: regularizing deep reinforcement learning from pixels [EB/OL]. [2024-01-15]. http://arxiv.org/abs/2004.13649
[6] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning [EB/OL]. [2024-01-15]. http://arxiv.org/abs/1509.06461
[7] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning [EB/OL]. [2024-01-15]. http://arxiv.org/abs/1509.02971
[8] HAARBOJA T, ZHOU A, HARTIKAINEN K, et al. Soft actor-critic algorithms and applications [EB/OL]. [2024-01-15]. https://arxiv.org/pdf/1812.05905
[9] LEE M A, ZHU Y K, ZACHARES P, et al. Making sense of vision and touch: learning multimodal representations for contact-rich tasks [J]. IEEE transactions on robotics, 2020, 36(3): 582-596. DOI: 10.1109/TRO.2019.2959445
[10] OORD A, DIELEMAN S, ZEN H, et al. Wavenet: a generative model for raw audio [EB/OL]. [2024-01-15]. https://arxiv. org/abs/1609.03499
[11] KOSTRIKOV I, YARATS D, FERGUS R. Image augmentation is all you need: regularizing deep reinforcement learning from pixels [EB/OL]. [2024-01-15]. http://arxiv.org/abs/2004.13649
[12] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [EB/OL]. [2024-01-15]. http://arxiv.org/abs/1506.02640

作　者　简　介

邵宏，中兴通讯股份有限公司架构算法专家、移动网络和移动多媒体技术国家重点实验室项目经理，现任中国计算机学会（CCF）智能汽车分会执行常委，是ITU-T、IEEE和IETF资深会员；主要从事具身智能机器人的研究工作；获工业和信息化部ITU-T文稿贡献奖、广东省优秀专利奖、深圳科学技术奖。

谢大雄，中兴通讯股份有限公司监事长、移动网络和移动多媒体技术国家重点实验室主任，教授级高工，中国发明协会会士、国家级领军人才，享受国务院特殊津贴，2023年担任工业和信息化部通信科技术委员会副主任，是《国家中长期科学和技术发展规划纲要（2006-2020年）》“新一代宽带无线移动通信网”重大专项论证委员会委员、国家“973计划”和“863计划”项目带头人；2002年、2010年先后获得国家科技进步奖二等奖2项，2017年获得国家技术发明奖1项，2002年获得首届深圳市市长奖。

用于混合现实的三维场景生成技术

3D Scene Generation for Mixed Reality

江海燕/JIANG Haiyan[1]，东野啸诺/DONGYE Xiaonuo[1]，王涌天/WANG Yongtian[1,2]
（1. 北京市混合现实与新型显示工程技术研究中心，北京理工大学光电学院，中国 北京 100081；
2. 北理工郑州智能科技研究院，中国 郑州 450000）
(1. Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China;
2. Zhengzhou Academy of Intelligent Technology, Zhengzhou 450000, China)

DOI：10.12142/ZTETJ.2024S1007
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20240724.1130.014.html
网络出版日期：2024-07-25
收稿日期：2023-11-26

摘要：在混合现实系统中，三维场景作为虚拟空间的关键构成要素，其高效生成方法一直是本领域的研究热点。人工智能辅助内容生成技术的发展，为该问题的解决提供了新的思路。综述性的归纳与总结了近年来三维场景生成的各项技术方法，以及混合现实场景下三维场景生成的现状，并对其发展趋势进行了分析与展望。

关键词：三维场景生成；混合现实；人工智能

Abstract: Mixed reality, as a typical form of a multimodal digital information system, aims to seamlessly merge virtual information with real-world information. It is one of the key technologies for the next-generation Internet. In mixed reality systems, the efficient generation of three-dimensional scenes, as the core element of virtual space, has been a key research focus in this field. In recent years, the development of artificial intelligence-assisted content generation method has introduced novel approaches to addressing this issue. This paper provides a synthesis and summary of various methods for three-dimensional scene generation in recent years and offers an analysis and outlook on their development trends.

Keywords: 3D scene generation; mixed reality; AI

引用格式：江海燕，东野啸诺，王涌天．用于混合现实的三维场景生成技术 [J]. 中兴通讯技术，2024, 30(S1): 43-53. DOI: 10.12142/ZTETJ.2024S1007
Citation: JIANG H Y, DONGYE X N, WANG Y T. 3D scene generation for mixed reality [J]. ZTE technology journal, 2024, 30(S1): 43-53. DOI: 10.12142/ZTETJ.2024S1007

混合现实技术致力于实现真实世界、虚拟世界和参与者三者之间的无缝融合，其最终目的是实现自然逼真的虚实融合人机交互。该技术既解决了虚拟现实中用户因无法看到真实环境导致行动受限的问题，也通过叠加虚拟信息的方式扩展了物理世界的边界，在医疗康复、教学培训、航空航天、娱乐休闲等领域具有广阔的应用前景。

三维场景的生成是混合现实场景实现高沉浸、自由交互的前提，也是该方向的研究重点。近年来，随着头戴显示器、立体投影等硬件设备的不断成熟，使得人们对三维场景生成的需求不断提升。同时，人工智能技术的发展，尤其是智能生成技术的快速发展，为这一领域注入了新的活力。

目前，三维场景自动生成的技术方法主要包括自回归神经网络、语法过程建模、图推理、自注意力模型等。随着多模态生成模型的发展，还可以通过自然语言条件约束、扩散模型来控制场景的生成。

具体到混合现实环境中，由于用户所处的物理环境与虚拟环境相互影响，三维场景的生成不仅需要考虑虚拟场景的需求，也要考虑用户以及周围物理环境的因素。针对用户因素，一些方法会通过用户参与的交互过程来控制场景的生成，实现更符合用户意图的混合现实环境。这类人在环的半自动生成方法允许用户实时控制和选择虚拟物体，具有很好的可控性。然而，由于其人在环的特性，场景生成速度较为缓慢。针对物理环境因素，一些方法通过实时动态监测物理环境，并将提取到的物理信息用于虚拟环境的生成，实现具有物理环境特征的实时三维场景生成。

基金项目：国家自然科学基金重点项目（62332003）；长沙市2022年科技重大专项（kh2301019）

1 通用三维场景生成技术

三维场景生成通常使用计算机图形学技术和计算机视觉技术，自动生成逼真的三维室内外环境。这些场景可用于虚拟现实、游戏开发、电影制作等多个领域。

从整体思路上来说，当前的主要研究倾向于将预先创建好的单体模型通过某种方式实现自动布局以构成所需要的三维场景。从方法上来说，三维场景生成的主要技术手段包括自回归神经网络、语法过程建模、图推理、自注意力模型以及扩散模型等。在这些技术手段之上，设计者可以通过加入场景生成的先验知识，对生成的三维模型进行风格化创建，实现条件场景的生成。

1.1 基于自回归神经网络的方法

基于自回归神经网络的方法常用于生成具有高真实感的室内场景。这种方法一般会利用网络来学习输入数据中物体出现及相互关联关系的概率分布，然后用其生成与训练数据相似的场景。

2018年，WANG等[1]提出了基于图像的自回归深度卷积网络的室内场景合成方法，实现从零开始迭代生成包含多个物体的房间，如图1所示。该方法仅将房间矩形轮廓作为输入，采用基于正交自上而下视图表示，通过卷积神经网络实现单个物体的添加。但在该方法中，一个模型只能针对特定房间类型（如卧室、客厅、办公室），进行生成，并且忽略了房间的层次关系、对象之间的功能关系、物体的大小等。此外，该方法是基于局部进行推理，难以用于推理物体在全局坐标中的位置。

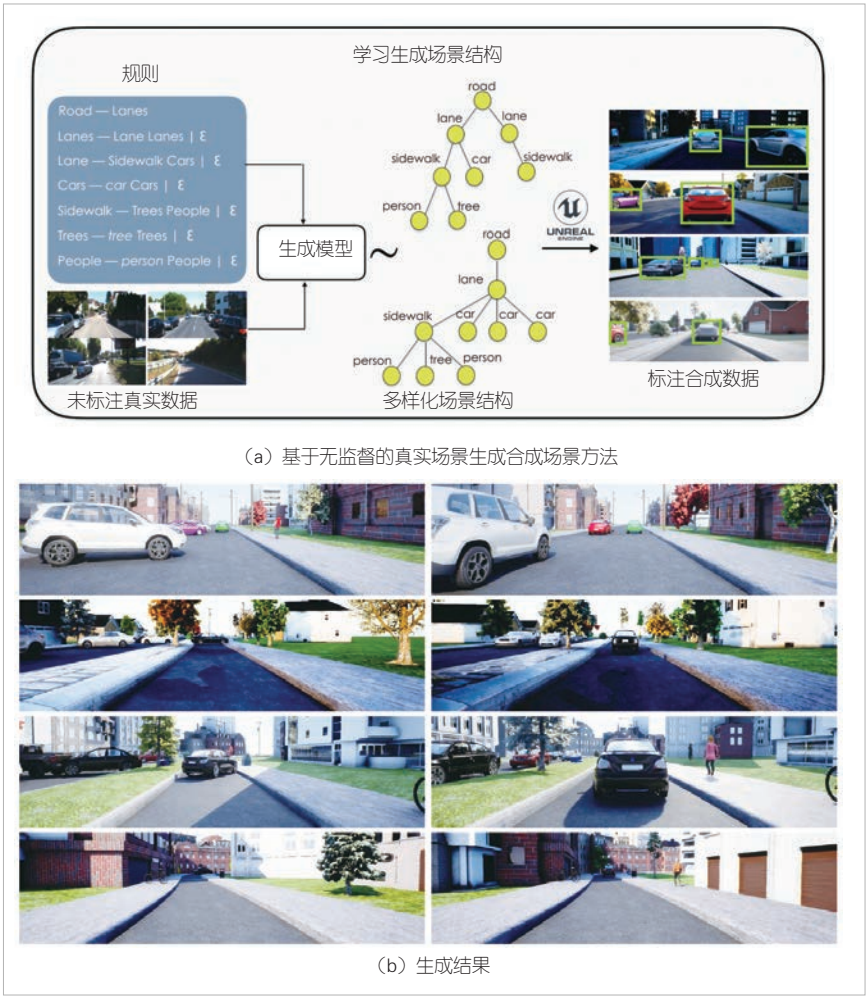
基于上述工作，2019年，RITCHIE等[2]采用自上而下的平面图像作为输入，并加入房间的几何信息（如天花板、墙等），使用自回归深度卷积神经网络实现快速的场景生成。相比于上述方法，该方法通过使用单独的神经网络模块预测对象的类别、位置、方向和大小，实现了部分场景的自动补全以及完整场景的合成，并且生成单个场景的平均速度达到1.858 s，而上述方法生成单个场景耗时约240 s。但同上述方法一样，该方法仍然忽略了场景的层次结构，并且难以生成具有风格一致性的场景。

1.2 基于语法的过程建模方法

基于语法的过程建模方法可用于生成具有较为明确层次感的场景,通过从给定的概率场景语法图中采样，逐步添加或者修改物体属性从而实现场景的生成。

2019年，KAR等相继提出了Meta-Sim[3]以及Meta-sim2[4]算法。Meta-Sim将场景解析为概率场景语法，通过神经网络从语法中采样场景结构，并实现场景中物体位置、姿态以及其他属性的修改，实现更符合真实关联关系的场景生成，如图2所示。Meta-Sim算法可以灵活地调整合成内容的结构和外观。基于这个特性，该算法可以通过优化生成更符合下游任务的场景数据集。但是由于Meta-Sim依赖语法获取场景结构，其可以生成的场景仍然受到限制。基于Meta-Sim以非监督学习的方式实现过程建模的场景生成，Meta-sim2比较真实空间合生成场景的特征空间离散度，通过强化学习方式学习给定的概率场景语法顺序采样规则，得到更符合真实场景分布的生成场景。

▲图1 基于自回归神经网络场景生成方法示意图[1]



▲图2 基于语法的过程建模方法示意图[4]

2020年，PURKAIT等提出SG-VAE算法[5]。该算法通过基于语法的自编码器，学习不同对象类别的形状和位置等参数，从而实现紧凑而准确的场景布局。SG-VAE算法从训练数据中，提取“物体是否可以同时存在”的信息，并由此进行推理。随后该算法将推理后形成的生成规范用于自动构建语法信息，并通过增强解析树来表示场景，以确保生成的场景始终符合正确的语义信息。基于语法的过程建模方法也可以用于超大规模的场景生成任务中。早在2001年，PARISH和MÜLLER就提出了一个模拟城市的系统[7]。该系统将土地划分为地块，为各地块分配的建筑物创建适当的几何形状，并可以连接各个地块生成一个公路网络和街道系统。基于生成的三维几何信息，该方法在几何信息上添加额外的纹理，以赋予建筑物更多的细节。生成城市的过程中，利用图像地图作为输入数据，控制道路和建筑的分布和形状。在道路网络生成环节中，该系统分别使用高速公路和街道进行区域划分。在建筑生成方面，该系统使用另一个参数的随机化系统来生成建筑的几何信息。每个建筑都是由一个任意的地面轮廓经过变换和挤压而成，形成摩天大楼、商业建筑和住宅房屋等不同种类的建筑，并分别由分区规则和图像地图来控制其生成。2006年，PARISH等[6]提出了一种形状语法，用于生成具有高视觉质量和几何细节的建筑外壳。该方法不仅可以高效地创建大规模的城市模型，还可以表达复杂的屋顶和立面结构。此外，该方法还分析了从简单的体积形状生成复杂的建筑外壳所面临的问题，并提出了两种重要的机制：遮挡查询和吸附查询，用以处理形状之间的相互作用和冲突。

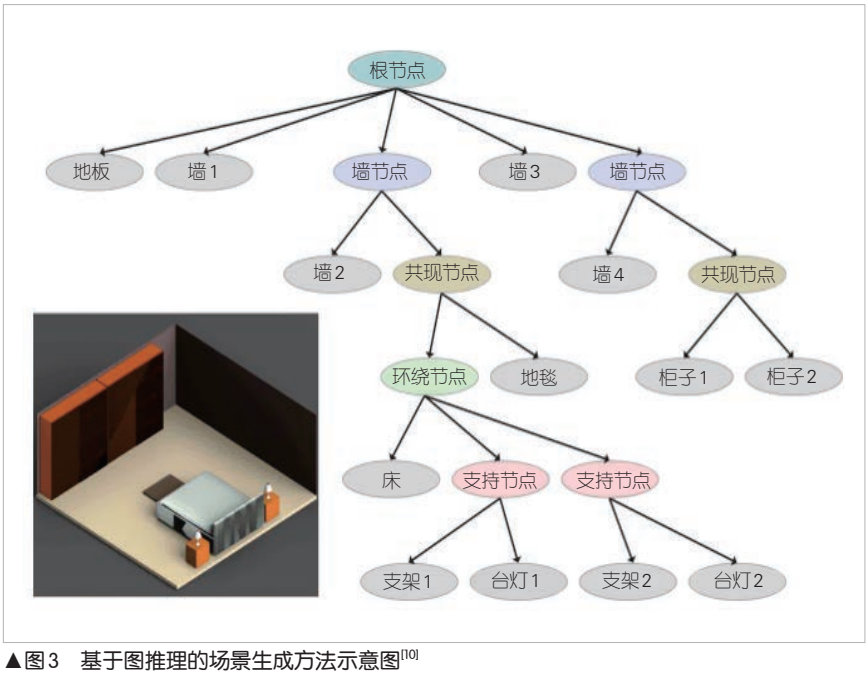
1.3 基于图推理的方法

基于图推理的方法利用图结构来表示和推理室内场景中物体之间的关系。图结构不仅可以描述物体的类别、位置、朝向等属性，也可以描述物体之间的相对距离、方向、对齐、支撑等关系。基于图推理的方法可以根据输入的房间形状和大小，或者参考样例场景，生成符合物体约束关系的室内场景。根据场景输入是否存在参考样例，可以将室内场景分为无样例生成算法和有样例生成算法。无样例算法通常从大规模数据中总结规则来生成场景，而有样例算法则基于文本、图像等输入，要求生成场景与输入在一定程度上匹配，属于有条件的场景生成任务。

无样例生成算法中，对象之间的排列常由向量表示，而这种抽象表示容易忽略几何细节，因此德克萨斯大学的ZHANG等[8]于2020年提出一种三维对象排列表示方法。三维对象排列表示基于对象的大小和形状属性对对象的位置和方向进行建模。通过与投影二维图像表示组合来训练三维场景生成器，同时兼具了二维场景生成和三维场景生成的优点。此外，该表示可以使用基于数据驱动的方法，通过分析数据集中对象的“共现”来提取先验。但是，高频率的“共现”并不一定代表强空间关系。鉴于上述问题，来自清华大学的研究团队[9]通过完全空间随机性测试来衡量对象之间空间关联的强度，并基于能够准确表示离散布局模式的样本提取复杂先验。该方法通过将输入对象划分为不相交的组，然后基于豪斯多夫度量进行布局优化，最终实现算法加速和置信度增强。

此外，无样例生成算法中更为常用的方法是对图形结构的编码。2019年，LI等[10]基于空间关系为房间中的每个家具构建树形的层次结构；对给定的室内场景，对象和对象组分别由叶节点和内部节点表示，将空间接近的物体视作相关的对象，并将对象间关系分类为支撑、环绕和共存三种。该方法首先将场景空间重构为一个具有各种对象关系和相应场景层次的物理模型。该研究以一个卧室为例，通过支撑关系将两对床头柜和台灯独立融合，然后通过环绕关系与床融合，如图3所示。利用上述结构，他们训练了一个变分递归自编码器，该编码器在编码阶段执行场景对象分组，在解码期间执行场景生成。同样地，WANG等[11]将面向对象和面向空间的范式结合在一起，提出了新的布局概念框架。该框架通过关系图表示场景的规划，将对象编码视为节点，将对象之间的空间–语义关系编码作为边。在规划阶段，采用深度图卷积生成模型对关系图进行综合；在实例化阶段，基于图像的卷积网络模块被用来指导搜索过程，以与关系图一致的方式将对象放置到场景中。

在有样例生成算法上，2020年LUO等[12]提出了一种三维场景布局网络，将变分自编码器与图卷积网络相结合，该网络能够基于相同输入的图结构生成不同的场景。在测试时，从学习分布中采样潜在编码，并与场景图一起发送给解码器以生成场景布局。在训练过程中，编码器将地面真实场景布局和场景图转换成一个分布，从中采样和解码潜在编码。而KESHAVARZI等[13]另辟蹊径，尝试将虚拟环境中的物体迁移到真实场景中，提出了一种基于场景图的生成式新框架。该框架允许从现有的场景中预测虚拟物体的位置和方向，从而实现基于环境感知的场景增强。ZHOU等[14]设计了一种神经信息传递方法，以预测给定场景中特定位置对象类型的概率分布。该方法将场景建模为图模型，其中节点表示场景中的现有对象，边表示对象之间的结构关系，通过图结构学习对象间的关系。



▲图3 基于图推理的场景生成方法示意图[10]

此外，以设计者输入的文字提示作为条件，可以生成具有特定语义信息的三维场景。斯坦福大学的研究团队[15]在该方向上进行了一系列的早期探索。他们首先将设计者提示的文本信息在符号化的语义空间进行解析，再通过解析出的语义信息生成相应的三维场景。这一工作包括收集带有自然语言描述的三维场景数据集，对场景生成文本进行语义解析，学习先验语义推断三维物体的隐含空间约束和通过深度相机学习人–物体间几何分布的概率模型等。他们早期的工作主要是将文本信息解析成三维对象和场景之间的符号化连接关系，并且通过数据集的收集获取不同场景类型中物体出现的统计数据作为隐性约束，从而生成可信的三维场景。在该团队的研究成果基础之上，更多的研究人员对语言条件驱动的三维场景生成方法进行了探索。例如，西蒙弗雷泽大学的研究团队提出了一种由语言驱动的三维场景建模方法[16]。该方法首先从三维场景数据库中通过用户的语言输入进行子场景检索，随后将检索到的子场景与当前环境合成新的三维场景。在每次用户编辑时，输入的语句都会转化为语义场景图，使用图对齐的方法从三维场景数据库中检索合适的子场景。检索到子场景后，根据输入文本和场景上下文，用附加对象对场景进行增强。然后，将增强后的子场景与当前场景进行语义对齐。最后，将增强的子场景拼接到当前场景中，合成一个新的场景。

除了利用设计者给出的条件进行三维场景生成，另一种方法是从人体的姿势、动作、移动进行场景生成。这种方法通过捕捉人体与环境物体交互过程的动作，估计场景内物体的位置、种类等内容，并联合推理。这种方式与自顶向下的设计逻辑相反，在给定大量的三维人体运动数据的基础上，生成与运动学信息相符的三维场景。为了采集人体的接触信息和运动信息，研究人员首先通过深度相机、惯性传感器或光学动作捕捉装置对三维人体运动数据进行获取。这些数据中蕴含着丰富的人与环境的交互信息，因此在场景生成上具有

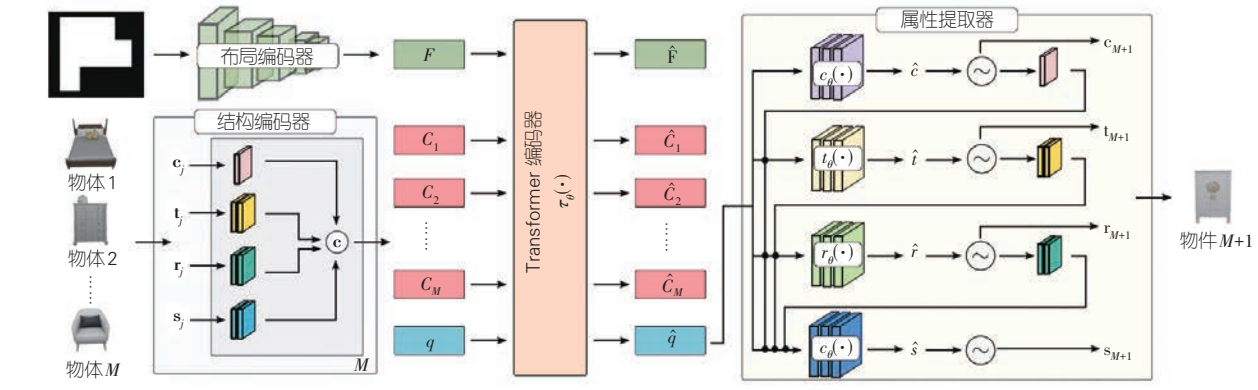
高度的可行性与合理性[17]。在获得了大量的用户姿势和动作信息之后，该方法在采集到的数据集上进行预测，估计被接触物体的种类，并使用空间、图来表示室内场景的空间属性，根据场景内物体的交互功能进行编码，形成马尔可夫随机场。随后，该方法从室内数据集场景中学习其数据分布，使用蒙特卡洛方法对马尔可夫链上的结点进行采样，形成三维场景。这种方法考虑到了物体的交互功能，因此在满足了视觉准确性的同时，在室内场景布局的功能性和自然性上具有优势。

1.4 基于自注意力模型的方法

自注意力机制允许模型在处理一个序列时考虑序列中每个元素与其他所有元素的关系。在三维场景生成中，使用自注意力模型可以更好的学习物体之间的关系，并将这些关系编码为排列表示。

PASCHALIDOU 等[18]将场景生成视为无序的对象集生成，提出了一个自动化生成三维室内场景的方法。该方法可以根据房间类型和形状，在不需要人为添加规则或关系图注释情况下，基于自注意力模型以自回归的方式生成具有逼真外观和3D一致性的家具布局，如图4所示。此外，该方法为用户提供对象约束，允许用户将任意类别和数目的家具固定在场景中特定位置。此后，PARA 等[19]针对该方法进行改进，提出了一种新的编码器–解码器架构的家具布局生成方式。该方法使用自回归布局生成器生成具有任意条件信息的布局，不仅允许仅输入对象的单个属性，而且可以对生成场景的细粒度进行控制，进一步增加了生成的逼真度和灵活度。在对象和场景的生成序列上，慕尼黑工业大学的WANG 等[20]提出了一个基于数据驱动的室内场景生成方法。该方法将场景生成问题转化为对象及其属性的序列生成问题，通过隐式学习对象关系，并使用自注意力模型解码器的交叉注意机制来构建条件模型，进而避免了生成结果对手工注释的依赖。

除此之外，该技术还可用于条件场景生成，通过人体在场景中的运动进行场景合成。自注意力机制模型输入可以仅依赖于场景中的三维人体姿态轨迹[21]。该方法所述的系统包括两个模块：接触预测模块和场景合成模块。接触预测模块利用现有的人–物体交互数据集来学习从人体到接触对象语义标签的映射，通过结合时域的上下文信息来增强标签预测在时间上的一致性。在生成估计的语义接触点后，场景合成模块首先根据语义可供性和物理可供性搜索适合接触点的对象，然后借鉴人体的运动推测出交互物体，用与人类没有接触的其他物体填充场景。2023 年，YI 等[22]提出了基于“人体运动”作为输入的三维室内场景生成方法。该方法基于人类与环境的交互，包含了场景对象的位置信息这一特性来推



（a）生成方法流程图

（b）3个不同平面图生成的场景结果

▲图4 基于自注意力模型的场景生成方法示意图[18]

断室内场景，并在此基础上使用自回归的自注意力模型结构，在给定人类运动序列的情况下，预测与人类接触的家具，从而生成合理的室内场景。

1.5 基于扩散模型的方法

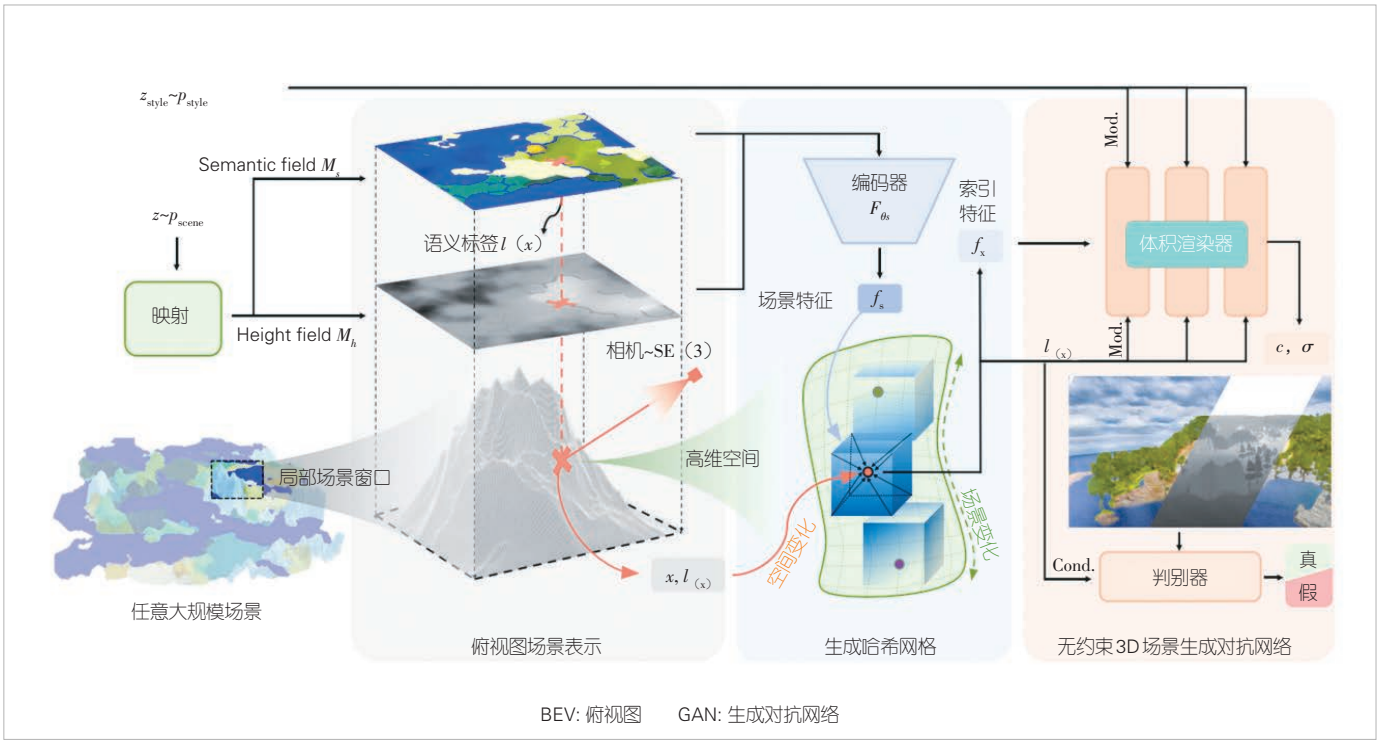
基于扩散模型的方法是一种利用随机微分方程来平滑地扰乱数据分布，将原始数据分布转化到已知的先验分布，然后通过学习逆向随机微分方程来从先验分布生成新的数据的方法。

基于扩散模型的场景生成方法主要通过随机噪声合成大规模的三维场景。最近出现的SceneDreamer就是一个典型的从随机噪声中生成三维场景的模型[23]，能够合成大规模的三维场景，并获得逼真的渲染效果。如图5所示，SceneDreamer提出了高效而富有表现力的鸟瞰视图场景表示方法、新颖的生成式神经哈希网格和基于风格的体积渲染器；给定从哈希网格中采样的潜在特征，渲染器即可通过风格调制的体积渲染来将其生成为逼真的三维视图。在此基础之上，CityDreamer模型[24]则可用于生成无限组合的三维城市场景。该模型将建筑实例和其他背景对象的生成分开，以处理生成建筑的多样性，并在此基础上构建了OSM数据集和Google Earth数据集，以提高生成的三维城市在其布局和外观的真实性。CityDreamer生成城市场景的方法可以分为三步：首先使用无边界城市布局生成技术生成城市场景，随后生成城市背景和建筑实例，最后将城市布局与建筑示例进行图像融合。

在风格化条件场景生成方面，设计者通过一个现有的场景和一个风格文本提示生成一个与文本相符、几何形状和纹理一致的新三维场景。此类方法首先生成三维场景的纹理，然后对网格纹理和几何图形进行联合优化，从网格中心开始，使用扩散模型更新未改变的区域，以确保生成的纹理与场景风格一致。

2023年，香港科技大学的研究团队将场景的位置布局和外观生成两阶段进行了拆分[25]。在位置布局阶段，采用了基于文本的条件扩散模型。该方法允许用户对生成的场景进行灵活的编辑，以生成高可信度、高纹理的优质三维场景。

在产业界方面，苹果公司[26]通过将真实三维场景的序列图像映射到一个完全分离的辐射场中进行潜在编码，之后通过大量的可变换视图在这些潜在编码上学习生成模型。该生成模型在训练时可引入图片、文本提示等不同的条件变量，以生成和这些条件变量一致的辐射场。Meta公司提出了一种基于语义信息的一致性风格室内场景生成方法[27]。该方法主要训练一个自回归模型，在推理过程中，将场景内已生成的物体作为条件，在每一步输出一个包括新物体及其位置信息的预测，以此实现风格一致且类别多样的场景。



BEV: 俯视图　　GAN: 生成对抗网络

▲图5　基于扩散模型的场景生成技术示意图[23]

2 混合现实环境中的三维场景生成方法

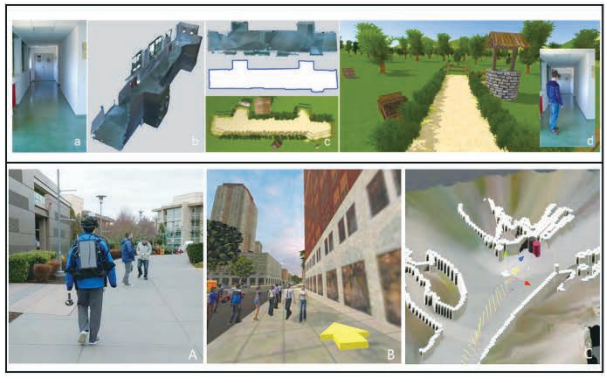
混合现实环境中，用户所处的实时环境与虚拟环境往往处于相互对应、相互映射的关系条件下。混合现实环境中的三维场景生成技术需要考虑真实环境的实时变化，通过环境感知将提取的真实环境信息用于虚拟场景的生成。这种情况下，往往需要通过虚实融合的手段，将真实场景中的物体映射为虚拟物体，并实现虚实配准。当前，完成这一任务有两种主要途径（如图6所示），一种是通过人在环对物体进行选择，实现物体间的映射；另一种方法是通过环境识别，自动将环境内物体进行映射。

2.1 基于人在环的场景生成方法

人在环场景生成方法采用半自动的方式，通过设计者在三维环境中实时编辑和操作虚拟物体，生成具有高合理性和高质量的三维场景。

早期的研究工作主要通过人为定义的配对方式，将真实物体替换为虚拟对象。例如，通过设计者的交互选择实现场景内的物体替换：系统向用户呈现一系列可以选择的物体，随后用户通过三维控制器点选所需的虚拟物体，并将虚拟物体和真实对象进行匹配，进而实现人在环的场景生成。

（a）人在环场景生成方法示意图[28]



（b）自动场景生成方法示意图[29]

▲图6　混合现实环境中的两种三维场景生成技术

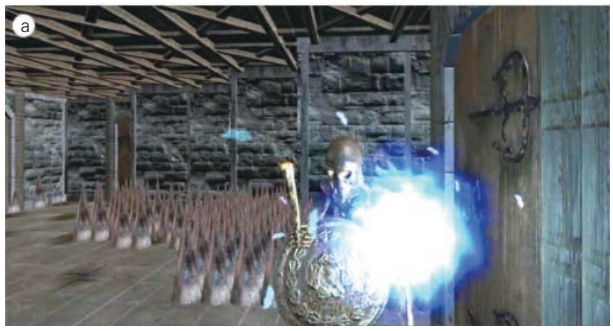
2021年，普渡大学的研究团队提出了一种端到端的系统设计框架：VRFromX[28]。该框架允许用户从真实世界的扫描中创作交互式三维场景。首先，用户使用手持式3D扫描仪扫描真实世界场景，形成点云图像；然后通过人工智能辅助在模型库中检索，并基于用户的交互选择，将点云对象替换为相应的虚拟模型；随后定义虚拟对象的功能和虚拟对象间的逻辑联系。该方法所设计的虚拟场景可以允许用户在场景中进行培训、交互等任务。

2.2 自动场景生成方法

人在环的场景生成方法在场景生成方面依赖于设计者的多次参与，生成效率比较低。为此，研究人员们希望探索一种基于环境感知的自动场景生成方法，实现从真实环境到虚拟环境的自动映射。

2017年，麻省理工学院的团队提出了一种以真实环境为输入，自动生成高沉浸感虚拟场景的系统[29]。该系统首先捕捉室内三维场景，检测家具和墙壁等障碍物，并绘制可行走区域地图。随后，系统将检测到的物体与虚拟对象配对，通过真实的物体向用户提供触觉反馈。该方法允许用户在任意大小和形状的室内空间中自动生成虚拟世界。在此基础上，他们又进行了系统优化[30]，允许用户通过自身虚拟代理与现实世界的物体进行交互，从而获得完整的触觉反馈，并将这些触觉反馈融入到三维场景的语义生成中。

此外，谷歌公司提出了一种通过真实环境生成虚拟环境的方法[31]。该方法将三维场景生成问题转化为一个同时满足多个约束的优化问题，约束条件包括场景几何信息、场景内物体、语义信息和物理约束。该系统首先将物体逐个生成的流程看作是在一个马尔可夫链上进行蒙特卡罗采样，然后将采样到的物体放置在一个二维俯视图上，以推断虚拟世界的环境布局。然后，使用满足几何、语义、物理等条件约束的多种物体和角色对模型进行填充。最后，将这些约束联合并使用一组离散变换，基于半定松弛的最新技术进行全局优化。微软公司则提出了一个可以在室外移动使用的虚拟现实系统[32]，如图7所示。该系统允许用户在真实世界中自由移动，并通过头戴显示设备完全沉浸在大型虚拟环境中。该系统首先预设了一个虚拟的环境，并通过视觉引导用户在虚拟世界中行走。在用户行走的同时，系统的跟踪器融合了GPS定位、光学跟踪和深度摄像机的画面，在现实环境中定位用



（a）实时动态生成虚拟现实场景

（b）对应用户真实环境

▲图7 自动场景映射生成技术叙事对应示意图[32]

户位置，并将用户重定向到目的地。同时，将场景实时感知到的用户行进路径上的障碍物映射为虚拟对象，并显示到用户头戴显示中。此后，微软公司的另一项自动生成虚拟场景的工作[33]以深度相机进行环境的识别和分割，实时检测并提取可行走区域或障碍物，并将可行走区域映射为实例化的预创建虚拟房间，障碍物则映射为阻挡用户前进的物体，进而提供长时间、高沉浸感的用户交互体验。

3 结论与展望

目前，虽然已经使用包括变分自编码器、生成对抗网络、自回归神经网络、图模型以及扩散模型等手段，面向混合现实的三维场景生成技术仍然需要较多的人工参与以及人为标注工作。以真实场景为输入，自动映射为风格化的虚拟场景在混合现实环境中仍然是一大挑战。

未来的研究工作可以专注于以下4个方面：1）专用数据集的构建问题：基于人工智能的生成方法虽然降低了场景生成的难度，增加了场景的丰富度，但目前仍然缺乏足够的数据集。而且目前较多的数据集是基于室内场景构建的，缺乏室外场景数据集，尤其是具有细节的室外物体数据集。2）场景物体多样性问题：生成方法大多基于已有虚拟物体模型生成布局，导致场景中的虚拟物体相对单一。未来可以利用人工智能技术直接生成差异化的虚拟物体。3）用户参与性问题：后续算法需要进一步简化用户的控制难度，同时提升用户对生成场景细节的精准控制，最终在两者间达到更好的平衡。4）环境交互问题：虽然已有方法考虑了部分环境的交互性，但大多局限于障碍物的识别。未来应更多地考虑用户与真实环境中不同物体的交互，将更多的物体功能融合到混合现实环境中。

参考文献

[1] RITCHIE D, WANG K, LIN Y. Fast and flexible indoor scene synthesis via deep convolutional generative models [C]//Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 6175 – 6183. DOI: 10.1109/CVPR.2019.00634

[2] WANG K, SAVVA M, CHANG A X, et al. Deep convolutional priors for indoor scene synthesis [J]. ACM transactions on graphics, 2018, 37(4): 1 – 14. DOI: 10.1145/3197517.3201362

[3] KAR A, PRAKASH A, LIU M Y, et al. Meta-sim: learning to generate synthetic datasets [C]//Proc. IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 4550 – 4559. DOI: 10.1109/ICCV.2019.00465

[4] DEVARANJAN J, KAR A, FIDLER S. Meta-Sim2: unsupervised learning of scene structure for synthetic data generation [C]// Computer Vision ECCV 2020. Springer, 2020: 715 – 733. DOI: 10.1007/978-3-030-58520-4_42

[5] PURKAIT P, ZACH C, REID I. SG-VAE: scene grammar variational autoencoder to generate new indoor scenes [C]//Computer Vision ECCV 2020. Springer, 2020: 155 – 171. DOI: 10.1007/978-3-030-58586-0_10

[6] MÜLLER P, WONKA P, HAEGLER S, et al. Procedural modeling of buildings [M]//ACM SIGGRAPH 2006 Papers. 2006: 614–623

[7] PARISH Y I H, MÜLLER P. Procedural modeling of cities [C]//Proc. 28th annual conference on Computer graphics and interactive techniques. ACM, 2001: 301 – 308. DOI: 10.1145/383259.383292

[8] ZHANG Z W, YANG Z P, MA C Y, et al. Deep generative modeling for scene synthesis via hybrid representations [J]. ACM transactions on graphics, 2020, 39(2): 1 – 21. DOI: 10.1145/3381866

[9] ZHANG S H, ZHANG S K, XIE W Y, et al. Fast 3D indoor scene synthesis with discrete and exact layout pattern extraction [EB/OL].(2020-02-05)[2024-03-15]. http://arxiv.org/abs/2002.00328

[10] LI M Y, PATIL A G, XU K, et al. GRAINS: generative recursive autoencoders for indoor scenes [J]. ACM transactions on graphics, 2019, 38(2): 1 – 16. DOI: 10.1145/3303766

[11] WANG K, LIN Y A, WEISSMANN B, et al. Planit: planning and instantiating indoor scenes with relation graph and spatial prior networks [J]. ACM transactions on graphics, 2019, 38(4): 1 – 15. DOI: 10.1145/3306346.3322941

[12] LUO A, ZHANG Z T, WU J J, et al. End-to-end optimization of scene layout [C]//Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 3753 – 3762. DOI: 10.1109/CVPR42600.2020.00381

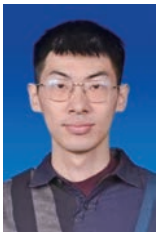
[13] KESHAVARZI M, PARIKH A, ZHAI X Y, et al. SceneGen: generative contextual scene augmentation using scene graph priors [EB/OL]. (2020-09-30)[2024-03-15]. http://arxiv.org/abs/2009.12395

[14] ZHOU Y, WHILE Z, KALOGERAKIS E. SceneGraphNet: neural message passing for 3D indoor scene augmentation [C]//Proc.

IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 7383 – 7391. DOI: 10.1109/ICCV.2019.00748
[15] CHANG A, SAVVA M, MANNING C D. Learning spatial knowledge for text to 3D scene generation [C]//Proc. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, 2014: 2028 – 2038. DOI: 10.3115/v1/d14–1217
[16] MA R, PATIL A G, FISHER M, et al. Language–driven synthesis of 3D scenes from scene databases [J]. ACM transactions on graphics, 2018, 37(6): 1 – 16. DOI: 10.1145/3272127.3275035
[17] QI S Y, ZHU Y X, HUANG S Y, et al. Human–centric indoor scene synthesis using stochastic grammar [C]//Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 5899 – 5908. DOI: 10.1109/CVPR.2018.00618
[18] PASCHALIDOU D, KAR A, SHUGRINA M, et al. ATISS: autoregressive transformers for indoor scene synthesis [EB/OL]. (2021–10–07)[2024–03–20]. http://arxiv.org/abs/2110.03675
[19] PARA W R, GUERRERO P, MITRA N, et al. COFS: controllable furniture layout synthesis [C]//Proc. Special Interest Group on Computer Graphics and Interactive Techniques Conference. ACM, 2023: 1 – 11. DOI: 10.1145/3588432.3591561
[20] WANG X P, YESHWANTH C, NIEßNER M. SceneFormer: indoor scene generation with transformers [C]//Proc. International Conference on 3D Vision (3DV). IEEE, 2021: 106 – 115. DOI: 10.1109/3DV53792.2021.00021
[21] YE S F, WANG Y X, LI J M, et al. Scene synthesis from human motion [C]//Proc. SIGGRAPH Asia 2022 Conference. ACM, 2022: 1 – 9. DOI: 10.1145/3550469.3555426
[22] YI H W, HUANG C H P, TRIPATHI S, et al. MIME: human–aware 3D scene generation [C]//Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2023: 12965 – 12976. DOI: 10.1109/CVPR52729.2023.01246
[23] CHEN Z X, WANG G C, LIU Z W. SceneDreamer: unbounded 3D scene generation from 2D image collections [EB/OL]. (2023–12–07)[2024–03–20]. http://arxiv.org/abs/2302.01330
[24] XIE H Z, CHEN Z X, HONG F Z, et al. CityDreamer: compositional generative model of unbounded 3D cities [EB/OL]. (2023–09–01)[2024–03–20]. https://arxiv.org/abs/2309.00610
[25] FANG C, DONG Y, LUO K M, et al. Ctrl–room: controllable text–to–3D room meshes generation with layout constraints [EB/OL]. (2023–10–05)[2024–03–20]. http://arxiv.org/abs/2310.03602
[26] BAUTISTA M A, GUO P S, ABNAR S, et al. Gaudi: a neural architect for immersive 3D scene generation [EB/OL]. (2023–07–27)[2024–03–20]. https://arxiv.org/abs/2207.13751
[27] XIONG W H, OĞUZ B, GUPTA A, et al. Simple local attentions remain competitive for long–context tasks [EB/OL]. (2021–12–14)[2024–03–20]. https://arxiv.org/abs/2112.07210
[28] IPSITA A, LI H, DUAN R L, et al. VRFromX: from scanned reality to interactive virtual experience with human–in–the–loop [C]// Proc. Extended Abstracts of 2021 CHI Conference on Human Factors in Computing Systems. ACM, 2021: 1 – 7. DOI: 10.1145/3411763.3451747
[29] SRA M, GARRIDO–JURADO S, MAES P. Oasis: procedurally generated social virtual spaces from 3D scanned real spaces [J]. IEEE transactions on visualization and computer graphics, 2018, 24(12): 3174 – 3187. DOI: 10.1109/TVCG.2017.2762691
[30] SRA M, GARRIDO–JURADO S, SCHMANDT C, et al. Procedurally generated virtual reality from 3D reconstructed physical space [C]//Proc. 22nd ACM Conference on Virtual Reality Software and Technology. ACM, 2016: 191 – 200. DOI: 10.1145/2993369.2993372
[31] SHAPIRA L, FREEDMAN D. Reality skins: creating immersive and tactile virtual environments [C]//Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2016: 115 – 124. DOI: 10.1109/ISMAR.2016.23
[32] CHENG L P, OFEK E, HOLZ C, et al. VRoamer: generating on–the–fly VR experiences while walking inside large, unknown real–world building environments [C]//Proc. IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2019: 359 – 366. DOI: 10.1109/VR.2019.8798074
[33] YANG J J, HOLZ C, OFEK E, et al. DreamWalker: Substituting real–world walking experiences with a virtual reality [C]//Proc. 32nd Annual ACM Symposium on User Interface Software and Technology. ACM, 2019: 1093 – 1107. DOI: 10.1145/3332165.3347875

作 者 简 介

江海燕，北京理工大学在读博士研究生；研究方向为混合现实、人机交互与人工智能；曾获第七届中国国际“互联网+”大学生创新创业大赛银奖；发表论文20篇，申请专利18项（已授权7项）。



东野啸诺，北京理工大学在读博士研究生；主要研究方向为虚拟现实、具身智能、多模态人机交互等。

王涌天，北京理工大学教授、博士生导师，北京市混合现实与新型显示工程技术研究中心主任，“长江学者”、国家杰出青年科学基金获得者；长期在技术光学、虚拟现实和增强现实领域从事教学和科研工作，主要研究方向为光学系统设计和CAD、新型三维显示、虚拟现实和增强现实、医学图像处理等；获得国家技术发明奖和国家科技进步奖各1项，省部级和国家一级学会/协会科技奖励10余项；出版专著4部，发表论文320余篇，授权发明专利200余项，主持制定虚拟现实和增强现实领域首批国家标准6项。

基于流式路径追踪的实时真实感渲染技术

Streaming Path Tracing for Real-Time Realistic Rendering Technology

王宸/WANG Chen，过洁/GUO Jie，郭延文/GUO Yanwen
（南京大学，中国 南京 210023）
(Nanjing University, Nanjing 210023, China)

DOI：10.12142/ZTETJ.2024S1008
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.tn.20240724.1131.016.html
网络出版日期：2024-07-24
收稿日期：2023-11-20

摘要：从图形处理器（GPU）的线程调度和内存访问两个角度出发，提出了一种基于流式路径追踪的实时真实感渲染方案。它将传统的路径追踪算法分解成多个独立的逻辑模块，使得算法的实现更加贴合GPU硬件的调度模式，并且使用数组结构体（SoA）风格的内存布局重新排列数据，使得算法在GPU中运行可以减少对动态随机存取存储器（DRAM）的访问次数，从而提升算法的运行性能。该方案对GPU友好，相比于无优化的路径追踪算法，在GPU中的运行时间降低了83%~88%。

关键词：实时真实感渲染；路径追踪；GPU友好；流式；SoA

Abstract: A streaming path tracing for real-time realistic rendering by examining the graphics processing unit (GPU) thread scheduling and memory access is presented. Traditional path tracing algorithms are decomposed into several independent logic modules, making the implementation more in line with GPU hardware´s scheduling pattern. Furthermore, using a structure of arrays (SoA) style memory layout rearranges data to reduce the number of accesses to dynamic random access memory (DRAM) when algorithms run on the GPU, thereby improving performance. The proposed scheme is GPU-friendly, showing a decrease in the runtime by approximately 83% to 88% on the GPU compared to the unoptimized path tracing algorithm.

Keywords: real-time realistic rendering; path tracing; GPU-friendly; streaming; SoA

引用格式：王宸，过洁，郭延文．基于流式路径追踪的实时真实感渲染技术 [J]. 中兴通讯技术，2024，30(S1)：54-59. DOI：10.12142/ZTETJ.2024S1008
Citation: WANG C, GUO J, GUO Y W. Streaming path tracing for real-time realistic rendering technology [J]. ZTE technology journal, 2024, 30(S1): 54-59. DOI: 10.12142/ZTETJ.2024S1008

光线追踪技术是一种用来产生真实感光照效果的图形渲染技术。这种技术由于需要极高的计算代价，通常被用于离线的图形渲染任务中，如：高端的视觉特效或三维动画制作。在传统的实时渲染任务中，三维场景通常使用光栅化方法进行渲染，将三维场景转换为二维图像，但这种方法很难准确模拟光线在真实世界中复杂的反射、折射以及散射现象。而光线追踪是一类用于追踪光线在三维场景中传播的方法，由于其直接模拟物理世界中光线传播的现象，因此能够捕捉到更为真实的光照效果。

近年来，随着显卡技术的更新迭代，通用图形处理器（GPGPU）[1]发展迅速，并且诞生了针对光线追踪加速优化的硬件单元RT Core。得益于此，部分光线追踪技术已经可以在游戏和其他实时应用中落地使用，如：基于光线追踪的环境光遮蔽、基于光线追踪的阴影等，一定程度上提高了实时渲染结果的真实感。尽管如此，离线渲染中的基于物理的无偏光线追踪渲染算法，例如：路径追踪[2]、路径指引[3]等算法依然难以在图形处理器（GPU）实现中达到实时的要求。

本文中，我们将结合GPGPU编程的特点，提出一种GPU友好型的路径追踪实现方法。通过重新调度算法子任务，以减少GPU线程的等待时间，加快线程的内存访问速度，从而达到在实时渲染的要求下仍能保证高真实感的目的。相比于无优化的路径追踪算法，该方案在GPU中的运行时间降低了83%~88%。

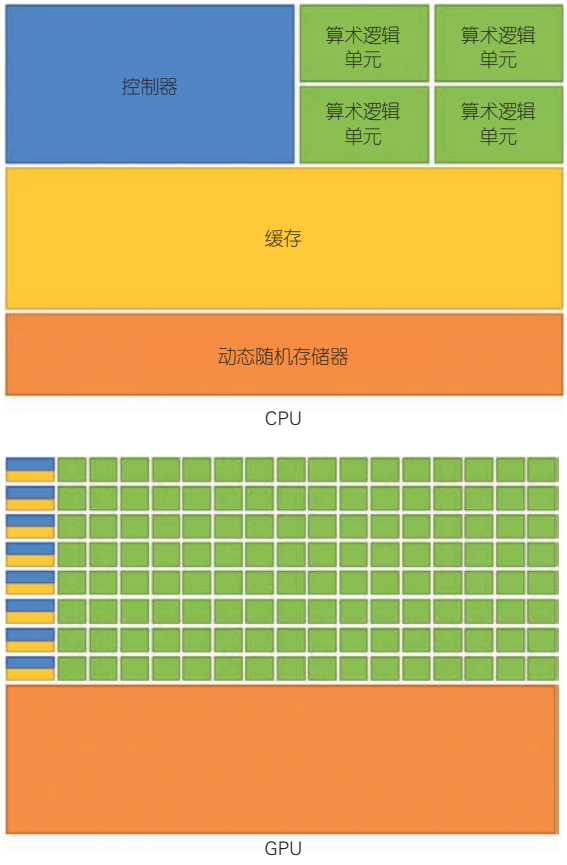
基金项目：国家自然科学基金项目（61972194、62032011）；江苏省自然科学基金项目（BK20211147）

1 GPGPU编程特点分析

现代显卡编程接口的设计，如统一计算架构（CUDA）、开放计算语言（OpenCL）以及图形管线中的Computer Shader，便于用户对显卡进行通用编程，而不是局限于传统的光栅化图形管线。与面向中央处理器（CPU）的编程类似，这些编程接口提供了对显存的读写操作，并为并行编程提供了原子操作与同步功能，使得通用的计算任务可以相对方便地移植到GPU中实现。同时，GPGPU编程需要考虑到GPU硬件架构与CPU的区别。如图1所示，GPU比CPU拥有更多的算术逻辑单元和更弱的控制器和缓存机制，导致GPU中的线程执行模型和内存访问模式与CPU有本质的区别。

1.1 GPU的线程执行模型

GPU采用单指令多线程（SIMT）的执行模型，用于组织调度成千上万个线程。由于图形渲染任务中，不同像素的绘制过程通常可以独立执行，很容易地按照像素将渲染任务分割成很多独立的任务并发执行，即可以为每个像素分配独立的线程，同时使用不同的数据执行相同的绘制指令。



CPU：中央处理器　GPU：图形处理器

▲图1　CPU和GPU硬件架构对比图

在SIMT执行模型中，线程会以一定数量（通常，英伟达公司发布的显卡数量为32，AMD公司发布的显卡数量为64）被打包成一组线程束（Warp）。同一组线程束的线程执行相同的指令。当同一组线程束内的线程需要执行不同的分支时，同一分支的线程会继续执行，而不同的分支线程会进入等待状态，从而造成系统性能损失。

与传统基于光栅化的图形渲染算法不同的是，基于物理的路径追踪算法控制分支较多，例如：由于场景复杂的遮挡关系，不同像素的光线弹射次数并不一致，并且不同的材质模型需要选择不同的分支计算。这使得传统的算法实现形式在GPU线程中会存在大量的分支等待时间，大大影响了算法的整体性能。

1.2 GPU的内存访问模式

现代GPU具有高带宽的内存系统，而这种高内存带宽通常是以产生相对较长的延迟为代价的[4]。内存访问延迟一般是指指令访问内存到获取返回结果之间的时间间隔。而在GPU中，线程执行一次全局内存访问需要消耗几百个时钟周期。因此，考虑线程束的内存访问延迟是十分重要的。

另一方面，GPU在设计中允许同一时钟周期内，容纳比可并发执行线程数更多的线程，以便一组线程在等待内存访问时，可以调度另一组线程执行，从而隐藏了内存访问时存在的延时情况。延迟隐藏的能力与线程的资源使用情况有关，如寄存器的使用数量。当计算内核使用的寄存器数量越多时，GPU中可以同时执行的线程数就越少，延迟隐藏的能力就越差。

2 GPU友好的实时真实感渲染方案

基于上述GPGPU编程的特点，本文针对蒙特卡洛路径追踪算法使用流式结构，设计了一套GPU友好的算法实现方案，使得算法的运行更贴合单指令多线程的执行模型，提高了线程资源的利用率，同时优化了内存布局以减少全局内存访问的次数，从而减少内存访问带来的延时问题。

2.1 传统路径追踪

传统路径追踪算法通过追踪光线在场景中传播的路径来计算全局光照效果。算法规定每一个像素从相机出发向场景内发射光线，经过若干次反射、折射或者散射，若最终击中光源，则评估该路径对像素颜色值的贡献。物理世界中认为光线可以弹射无数次，直到能量被完全吸收，而算法中通常光线的弹射次数（路径深度）由参数指定，作为算法终止的

条件之一。

在图2中，场景由康奈尔盒子、斯坦福兔子和一个实心小球组成。其中，地面是导体材质，小球为玻璃材质，其余均为漫反射材质。不难看出，随着光线弹射次数的增加，由路径追踪算法渲染的结果中，复杂材质的反射、透射现象逐渐接近真实世界的效果。

传统路径追踪算法在GPU中实现时，如PURCELL等[5]在早期可编程图形硬件上实现的光线追踪算法，将算法的条件分支和循环整体在一个内核程序里实现。这种方法通常被称为大内核方法（Mega-kernel）。在这种形式下，GPU线程的调度是以一个像素的路径追踪流程为单位的。然而，复杂的算法结构会导致内核存在大量的控制分支和内存访问需求。同时，由于渲染场景的复杂性，不同像素发射的光线可能会击中不同类型的材质，并且有些像素发出的光线在采样的过程中可能会提前终止，而另一些像素发出的光线会在场景中经过多次反射或折射，因此光线的路径通常是完全不同的。这使得大内核路径追踪方法中，由SIMT方式调度的GPU线程之间存在大量的互相等待时间。同时，高分辨率的渲染任务通常需要绘制百万甚至千万数量级别的像素，而消费级GPU的计算单元数通常不到一万。这使得同时调度的线程束中，先执行完成的线程束仍在占用线程资源，而等待的任务无法第一时间获得线程的使用权，使得算法效率低下。

2.2 流式路径追踪

为了减少线程资源的浪费，我们选择将整体算法按流程划分为多个指令数少且独立的小内核程序，使得数据可以按照上一个内核程序的执行结果分流到不同的内核程序中继续执行，从数据流的角度减少了算法程序在GPU中执行的条件分支。根据蒙特卡洛路径追踪算法的流程，我们将其分解成以下7个小内核程序：

1）路径初始化：根据相机参数为每个像素生成一条初始光线。

2）求交测试：该内核需要调用光追硬件的求交指令进行光线求交步骤，记录当前光线是否与场景有交点，以及交点的局部信息，如：几何信息和材质信息。场景光线的求交测试结果是路径追踪流程中产生大量分支的原因之一。当光线与场景有交点时，光线会继续追踪下去；而当光线与场景无交点时，该条路径会被终止追踪。流式方法会将产生这两种结果的路径分流到不同的内核中，从而提前释放将要终止路径的线程资源。

3）光线相交处理：与场景有交点的光路会被分流到该内核继续进行计算。若当前光线的方向由出射点的双向散射分布函数（BSDF）采样所得，需要根据BSDF采样的权重计算能量贡献。同时，需要对光源进行采样，得到光路可能的发射方向，并且记录光源的采样信息，同时生成一条阴影测试光线，用于判断光源是否被遮挡。

4）阴影测试：该内核用于在采样光源时，判断被采样的光源光线是否被物体遮挡。该操作同样需要调用光追硬件的求交指令。虽然硬件光追的求交速度很快，但相对来说仍然是算法整体耗时的主要来源，因此尽量避免不必要的求交步骤是很重要的。由于光源采样的概率可能为0，即可能采样到失效样本，并且不需要对该样本计算贡献，因此，采样到失效样本的路径不需要进行阴影测试，可以在光源采样步骤中先筛选出有效样本，再将其统一调度到该内核中调用求交指令。

5）光源贡献计算：通过阴影测试的路径会被分流到该内核，进而继续计算光源的贡献值。

6）材质BSDF采样：需要进行材质BSDF采样步骤的路径与需要进行光线相交处理的路径相同，经过阴影测试和光源贡献计算后的数据流会在该内核汇合。根据当前交点的材质BSDF采样出当前路径的下一次光线弹射方向，同时，由于采样存在失效情况，如：BSDF值为0或样本概率为0，可以丢弃样本失效的路径，以减少无效的线程资源占用。

▲图2 不同光线弹射次数的路径追踪渲染结果图

7）光线丢失处理：求交测试失败的路径会被分流到该内核，即当前光路与场景中没有交点，需要中断该条路径。在该内核中，通常需要计算环境光照对光路能量的贡献。

各个内核的执行流程如图3所示。初始化阶段以像素为单位生成初始光线。当初始光线进入光线追踪循环后，在CPU端，调度程序按照前一个内核的数据分流执行计算内核。分流的计算内核可以并发执行，而合流的内核需要等待所有上一步内核的计算完成。在光源贡献计算结束且光线追踪的弹射次数达到规定的数值后，即可停止光线追踪，将结果记录到内存后进行后处理并输出。此外，当计算资源足够多时，即可以调度的GPU线程数远大于像素数时，可以流水线式执行图3所示流程，进一步提高并发度。

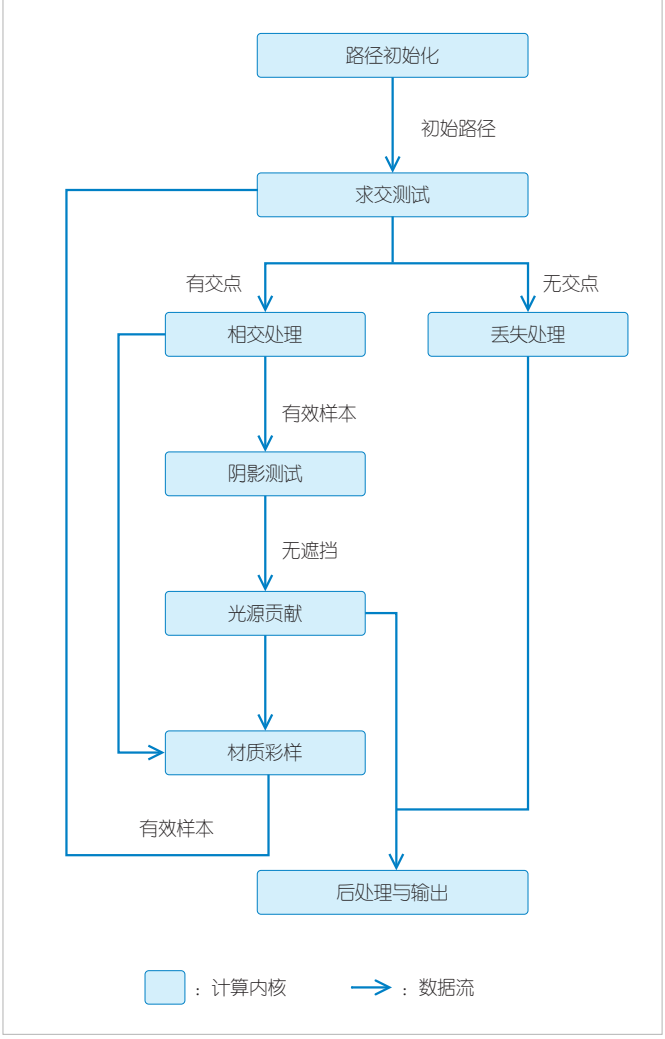
2.3 SoA的数据内存布局

路径追踪算法过程具有大量的全局内存访问操作，而数据在内存中的布局直接影响了GPU中线程的内存访问速度。这是因为在GPU硬件的设计中，全局内存是以二进制存储在动态随机存取存储器（DRAM）单元中的。相对计算单元的时钟周期来说，DRAM的数据访问速度非常慢。

为了优化对DRAM中数据的访问速度，GPU引入了内存合并技术。若一个线程束内所有线程访问的全局内存是连续的，硬件会将这些访问指令合并成对一个连续内存的访问，通过减少内存业务来提高访问效率。

基于内存合并的规则，我们在实现路径追踪算法时，使用数组结构体[6]（SoA）的内存布局组织数据。如图4所示，若一个对象类型含有3个成员变量X、Y、Z，在面向对象风格的编程模式中，该类型的数组会按照结构体数组（AoS）的形式存储，数据会按照成员的声明顺序排列。这导致在线程束中线程访问不同对象的同一个成员时，需要查询不连续的内存，进而无法使用内存合并技术。反之，使用SoA的内存布局使得同一线程束内的线程访问的是连续内存。

此外，SoA的内存布局还能更好地利用GPU的带宽，因为线程中执行的指令未必需要访问对象的所有成员。SoA的内存布局使得其在访问对象的部分成员时，另一部分成员不需要载入共享内存或者寄存器，从而避免了无效的带宽浪费。

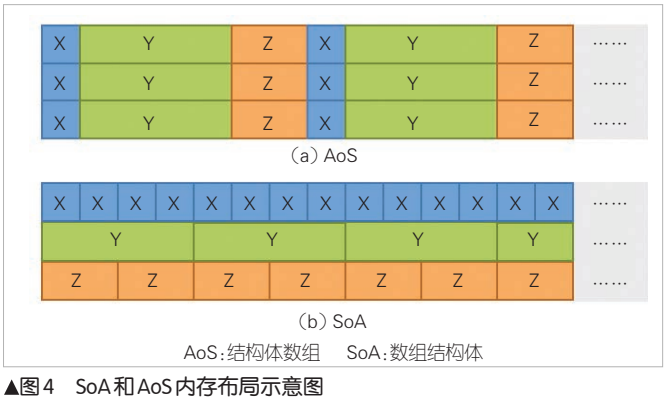


▲图3　流式路径追踪流程图

3 实验分析和效果对比

我们基于硬件光追API OptiX[7]分别实现了传统大内核的路径追踪方法、流式路径追踪方法，以及结合SoA优化内存布局的流式路径追踪方法，并且在英伟达RTX-3090显卡（24 GB）上完成性能的对比测试。

我们在Rendering Resources网站[8]上选取了11个测试场景，分别是：bathroom、bathroom2、bedroom、classroom、kitchen、living-room、living-room2、living-room3、staircase、staircase2和veach-ajar。其中，部分场景渲染结果如图5所示。



▲图4　SoA和AoS内存布局示意图

▲图5 部分测试场景渲染结果图

为了验证流式路径追踪方法和SoA内存布局在GPU上运行的优势，我们固定光线的弹射次数（深度）和质量（单像素的采样数量），在不同场景中分别使用3种方法进行渲染测试。在这11个场景上，为保证渲染效果的真实感，同时平衡渲染耗时，我们设置光线的最大弹射次数为3、单帧的采样数为1。测试得到渲染的平均帧率如表1所示。其中，流式路径追踪算法的实时帧率是大内核方式帧率的2～3倍，使用SoA内存布局的方法使系统性能有一定的提升。

若不考虑实时帧率，设置光线深度为64、单像素采样数为1 024的测试结果如图6所示。不难发现，SoA内存布局在光线弹射次数更多时，对算法性能的优化更明显，能够得到相比于无内存优化方法1.5～3.5倍的性能提升，总的优化提升达到了5.8～8.6倍。

▼表1 测试场景的平均渲染帧数

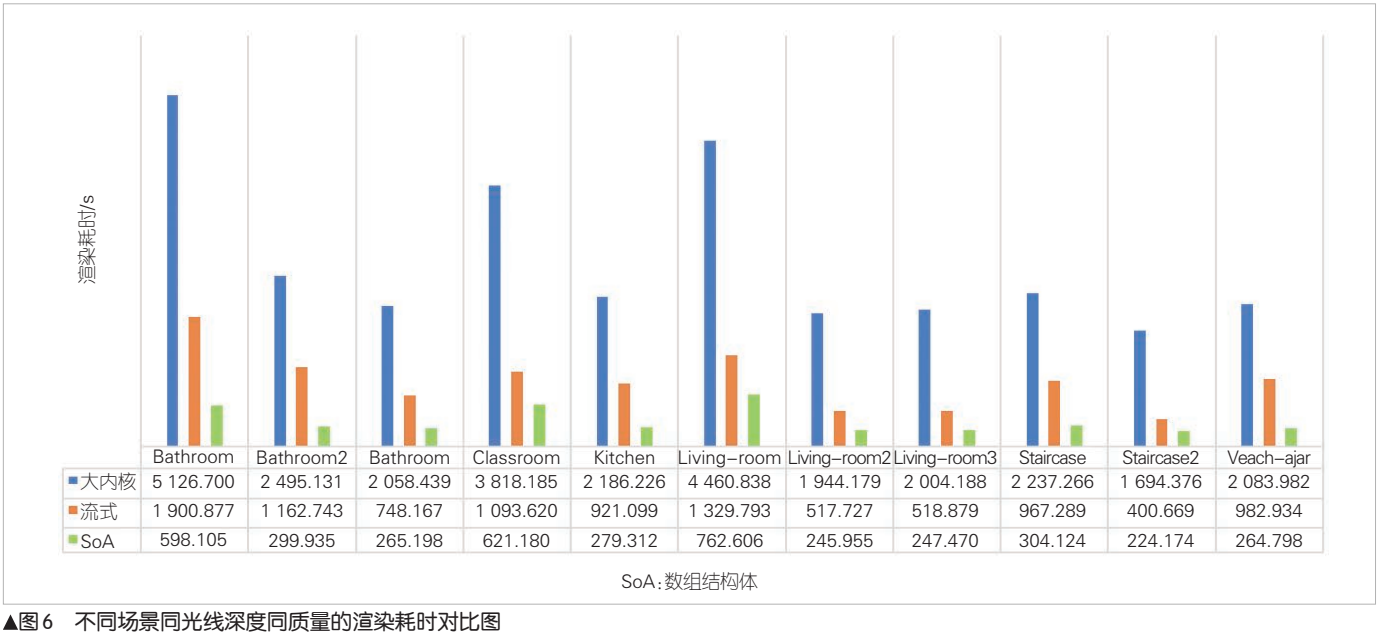
场景	面片数	分辨率	平均帧率(FPS)		
			大内核	流式	流式SoA
Bathroom	1 551 922	1 920×1 080	9	30	38
Bathroom2	3 731 807	1 280×720	22	60	83
Bedroom	4 475 258	1 280×720	24	63	84
Classroom	311 496	1 280×720	9	30	33
Kitchen	4 324 591	1 280×720	23	65	95
Living-room	419 697	1 280×720	8	24	26
Living-room2	1 779 233	1 280×720	23	59	90
Living-room3	2 358 624	1 280×720	21	45	73
Staircase	787 987	720×1 280	25	60	85
Staircase2	92 765	1 024×1 024	20	59	90
Veach-ajar	1 148 068	1 280×720	30	70	110

FPS：每秒帧数 SoA：数组结构体

另外，我们改变光线的弹射次数，比较了3种方法的渲染性能。图7展示了在living-room3场景下，3种方法在不同光线深度时渲染所需的时间。可以看出，深度越深流式方法的性能提升越明显。这在加速离线渲染的场景中具有很大的应用价值。

4 总结与展望

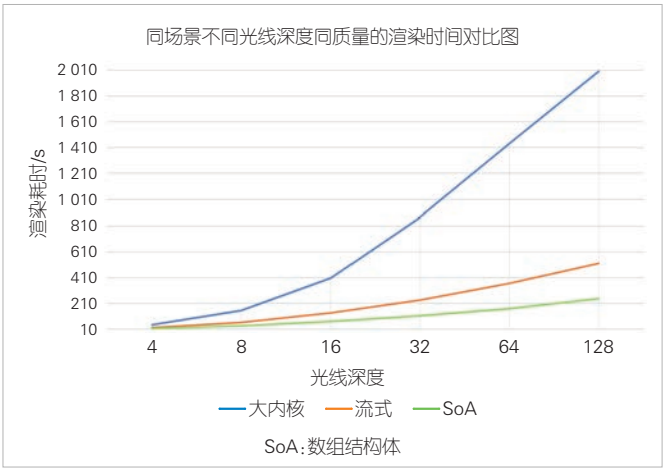
本文中，我们从现代GPU硬件的线程调度和内存访问两个角度出发，探索了GPU友好的路径追踪实现方案，提出了流式路径追踪方法，并使用SoA的内存布局优化了算法运行时的内存访问。通过控制光线的弹射次数和采样数，该方法在保证了渲染真实感的同时也满足了实时渲染的性能要



	Bathroom	Bathroom2	Bathroom	Classroom	Kitchen	Living-room	Living-room2	Living-room3	Staircase	Staircase2	Veach-ajar
大内核	5 126.700	2 495.131	2 058.439	3 818.185	2 186.226	4 460.838	1 944.179	2 004.188	2 237.266	1 694.376	2 083.982
流式	1 900.877	1 162.743	748.167	1 093.620	921.099	1 329.793	517.727	518.879	967.289	400.669	982.934
SoA	598.105	299.935	265.198	621.180	279.312	762.606	245.955	247.470	304.124	224.174	264.798

SoA：数组结构体

▲图6 不同场景同光线深度同质量的渲染耗时对比图



▲图7 同场景不同光线深度同质量的渲染时间折线图

求，在测试场景中相比于未优化的大内核方法减少了超过80%的渲染耗时。

随着现代GPU硬件的快速发展，使用硬件光追已然成为未来的发展方向，而传统的离线渲染方法，如路径追踪、路径指引等，势必会逐步迁移到实时渲染领域，以提高实时渲染的真实感。同时，这种GPGPU的编程思想可以扩展到其他算法中，有助于现有算法在实际生产中落地应用。

参考文献

[1] GHORPADE J, PARANDE J, KULKARNI M, et al. GPGPU processing in CUDA architecture [EB/OL]. [2024-02-25]. http://arxiv.org/abs/1202.4347
[2] KAJIYA J T. The rendering equation [C]//Proceedings of the 13th annual conference on Computer graphics and interactive techniques. ACM, 1986: 143-150.. DOI: 10.1145/15922.15902
[3] MÜLLER T, GROSS M, NOVÁK J. Practical path guiding for efficient light-transport simulation [J]. Computer graphics forum, 2017, 36(4): 91-100. DOI: 10.1111/cgf.13227
[4] LAINE S, KARRAS T, AILA T M. Megakernels considered harmful: wavefront path tracing on GPUs [C]//Proceedings of the 5th High-Performance Graphics Conference. ACM, 2013: 137 - 143. DOI: 10.1145/2492045.2492060
[5] PURCELL T J, BUCK I, MARK W R, et al. Ray tracing on programmable graphics hardware [J]. ACM transactions on graphics, 21(3): 703-712. DOI: 10.1145/566654.566640
[6] MICIKEVICIUS P. GPU performance analysis and optimization [EB/OL]. [2024-02-25]. https://on-demand. gputechconf. com/gtc/2012/presentations/S0514-GTC2012-GPU-Performance-Analysis.pdf
[7] PARKER S G, BIGLER J, DIETRICH A, et al. OptiX: a general purpose ray tracing engine [J]. ACM transactions on graphics, 29 (4): 66. DOI: 10.1145/1778765.1778803
[8] BITTERLI B. Rendering resources [EB/OL]. [2024-02-25]. https://benedikt-bitterli.me/resources/

作 者 简 介

王宸，南京大学在读硕士研究生；主要研究领域为计算机图形学。

过洁（通信作者），南京大学副研究员；主要研究领域为计算机图形学和虚拟现实技术；先后主持和参加基金项目20余项；已发表论文70余篇。

郭延文，南京大学教授；主要研究领域为计算机图形学和三维视觉；先后主持和参加基金项目20余项；已发表论文100余篇。

基于深度生成模型的视觉模式表示与编码

Visual Pattern Representation and Coding Based on Deep Generative Models

郭怡琳/GUO Yilin[1]，常建慧/CHANG Jianhui[2]，黄成/HUANG Cheng[3]，马思伟/MA Siwei[2,4]
（1. 北京大学深圳研究生院，中国 深圳 518055；
2. 北京大学，中国 北京 100871；
3. 中兴通讯股份有限公司，中国 深圳 518057；
4. 鹏城实验室，中国 深圳 518057）
(1. Peking University Shenzhen Graduate School, Shenzhen 518055, China;
2. Peking University, Beijing 100871, China;
3. ZTE Corporation, Shenzhen 518057, China;
4. Pengcheng Laboratory, Shenzhen 518057, China)

DOI：10.12142/ZTETJ.2024S1009
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20240726.1421.002.html
网络出版日期：2024-07-29
收稿日期：2023-11-25

摘要：认为早期智能编码方法的性能受限于手工设计的方案，当前基于神经网络的编码方法可解释性不足，不利于后续面向人机视觉的分析与交互。受生成模型的启发，生成式编码方法通过构建生成模型来实现图像和视频的压缩和合成，获得可解释的紧凑视觉表示并生成符合图像先验分布的高视觉质量内容。其中概念图像编码与概念视频编码利用生成模型强大的样本生成能力与紧凑层次视觉表示模型，实现了编码性能更优的图像与视频编码；跨模态语义编码对图像与文本域进行跨模态转换与编码，保持可解释的同时实现上千倍的超高压缩比与令人满意的重构结果。

关键词：智能视频编码；生成式编码；跨模态压缩；概念编码

Abstract: The performance of early intelligent encoding methods was limited by manually designed solutions, while current neural network-based encoding methods lack interpretability, which hinders subsequent analysis and interaction between humans and machine vision. Inspired by generative models, the generative encoding methods aim to achieve compression and synthesis of images and videos by constructing efficient generative models, obtaining interpretable compact visual representations, and synthesizing high-quality visual content that conforms to the prior distribution of images. Among them, conceptual image encoding and conceptual video encoding leverage the powerful sample generation capability and compact hierarchical visual representation models of generative models, resulting in superior encoding performance for images and videos. Cross-modal semantic coding, on the other hand, enables cross-modal transformation and coding between the image and text domains while maintaining interpretability, achieving ultra-high compression ratios of thousands of times and satisfactory reconstruction results.

Keywords: intelligent video encoding; generative encoding; cross-modal compression; conceptual coding

引用格式：郭怡琳, 常建慧, 黄成, 等. 基于深度生成模型的视觉模式表示与编码 [J]. 中兴通讯技术, 2024, 30(S1): 60-66. DOI: 10.12142/ZTETJ.2024S1009
Citation: GUO Y L, CHANG J H, HUANG C, et al. Visual pattern representation and coding based on deep generative models [J]. ZTE technology journal, 2024, 30(S1): 60-66. DOI: 10.12142/ZTETJ.2024S1009

智能视频编码技术可以追溯到20世纪80年代末[1]，旨在使用知识和语义来设计紧凑的视觉内容表示模型和方法，以便在不同的粒度级别（如块、网格、区域和对象）上对视觉信息进行结构化描述。然而，早期的研究主要针对特定信号源，并基于专家知识进行人工设计编码方案，导致编码性能与应用范围受限。随着移动设备、监控摄像头和其他视频采集设备的大量增加，视频数据量显著增加。在大数据时代，图像和视频处理需要更高效和智能的编码技术。在过去几年中，神经网络在图像和视频理解、处理和压缩等多个

基金项目：国家自然科学基金项目（62025101）；鹏城实验室重大攻关项目（PCL2024A02）

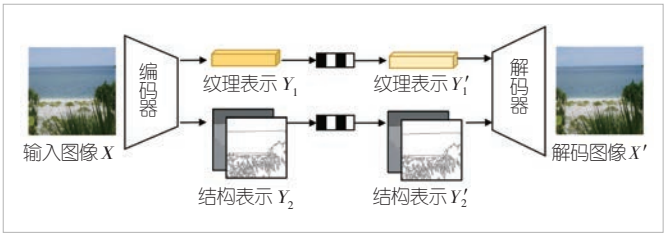
领域展示了巨大的潜力。在压缩任务中，神经网络中的参数可以基于大量图像视频样本进行训练，通过非线性变换将输入的像素数据映射到更紧凑的潜在表示[2]。此外，通过熵估计模型获得的可微码率约束，协同失真约束能够实现自适应率失真优化，能够有效减轻模型对手动设计模块的依赖。得益于这些优良的特性，基于深度学习的编码可以在图像和视频编码领域发挥出更大的潜力。但现有基于深度学习的编码方法将信号数据压缩为不可解释的紧凑潜在表示，这种机制不仅对面向人类视觉的分析与交互不够友好，也无法高效协助下游的机器分析任务。

近期生成模型的快速发展为图像和视频压缩的研究带来了新的启示。生成模型可以对联合概率分布进行建模，从统计的角度表示数据分布情况，并通过控制采样过程生成符合目标分布的样本。生成对抗网络[3]（GAN）、变分自编码器[4]（VAE）和深度扩散模型[5]（DPM）等生成模型在多模态交互和可控生成视觉内容方面展现出强大的能力。本文认为，新兴的生成式编码方法可以通过构建生成模型来实现图像和视频的压缩和合成，通过学习图像和视频数据的先验分布和跨域映射来挖掘样本的潜在特征和数据间的相关性，获得可解释的紧凑视觉表示，从而帮助人类理解和机器分析。此外，在数据信号严重损失的情况下，生成式编码允许生成符合图像先验分布和人眼视觉感知的纹理，从而在带宽受限的条件下保持高质量的视觉重建。作为新一代智能编码的关键技术，当前生成式编码的研究致力于构建视觉内容的概念表示模型，将图像、视频等内容编码成可解释的高层结构、纹理和语义表示，提出了概念图像编码[6–8]、概念视频编码[9]和跨模态语义编码[10]等方法,能够保持可解释性的同时在上千倍的超高压缩比下获得主观感知良好的重构结果。

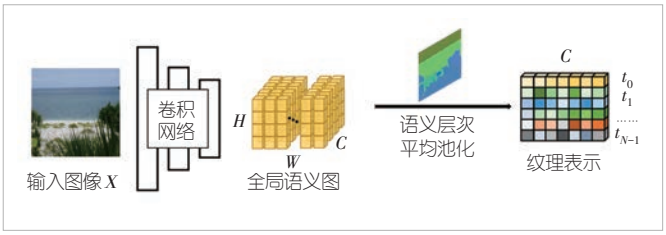
1 概念图像编码

人类视觉系统（HVS）[11]通过处理多方面信息并将其整合为抽象的高级概念（如结构、纹理和语义）来感知视觉内容，从而形成后续认知过程的基础[12]。受HVS启发，概念编码旨在将图像编码成紧凑的、高维的、可解释的表示，以获得高视觉质量的重构与更高效、更便于分析的压缩架构。在解码端，多层解码表征融合并以深度生成的方式合成目标图像。概念编码面临的主要挑战包括如何实现高效的表征分离，以及如何设计有效的生成模型以实现高视觉质量的重建。GREGOR 等[13]引入了带注意力的卷积深度递归输入器（DRAW）[14]，扩展了VAE[4]，使用RNN作为编码器和解码器，将图像转换成一系列越来越细节的表示。然而，其学习到的图像表征的可解释性仍然不足，且仅适用于小分辨率的数据集。神经视频压缩也受到类似的限制，典型的视频压缩方法[15]与图像压缩方法[13]共享相同的VAE架构，并将原始序列转换为低维表示。然而，学习到的视频表征的可解释性仍然缺乏探索。因此，我们在压缩过程中引入结构信息或高级语义信息等可解释表征，以增强图像和视频表征的可解释性。

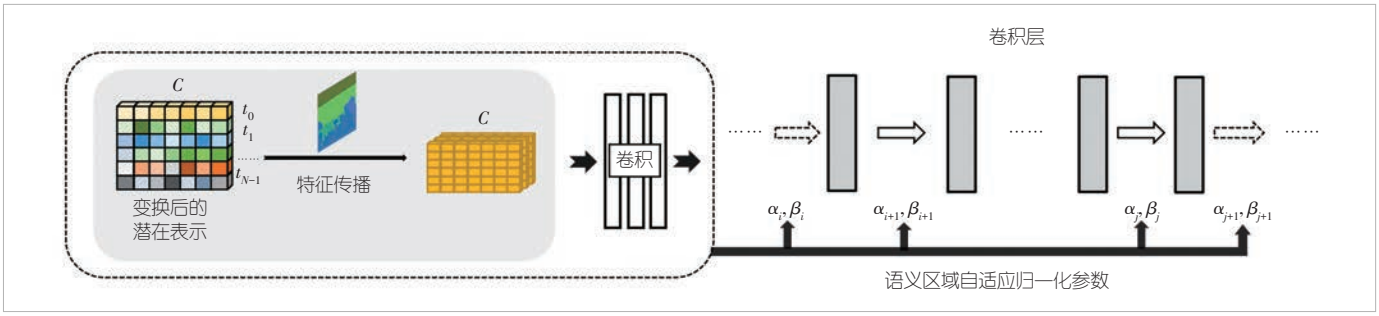
概念编码的突破性进展在于将图像编码为互补的两层视觉概念表示[6–7]。如图1所示，图像被分解为结构和纹理表示两部分，其中纹理表示通过图2展示的纹理建模过程得到，在解码端由图3所示的模块实现图像合成。图4展示了该研究中提出的特征域空间中结构和纹理表征分离。结构层由边



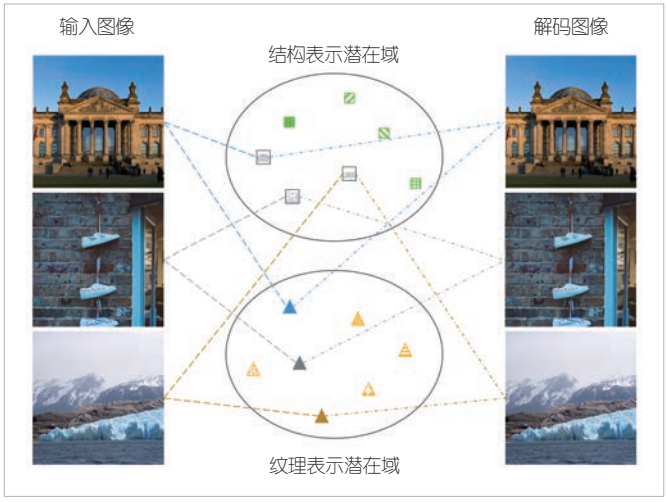
▲图1 典型的概念图像编码框架



▲图2 典型的概念图像编码纹理建模过程



▲图3 典型的概念图像编码合成过程



▲图4 典型的概念图像编码图示

缘图表示，纹理层由变分自编码器提取的低维潜在变量表示。为了从压缩的分层特征中重建原始图像，分层融合生成模型被设计出来[8]，其中纹理层和结构层通过自适应实例归一化（AdaIN）融合引导生成过程。大量实验表明[7]我们提出的概念压缩框架可实现在极低的比特率（<0.1 bpp）下保持较高的视觉重构质量，并通过结构编辑与风格变换以及关键点检测实验证明了概念编码在内容控制和分析任务上的优势。然而，仅使用一组隐变量对整个图像的复杂纹理进行建模是一项巨大的挑战；此外，如何为视觉表征建立有效的熵模型并联合率失真优化，尚未进行过探讨。文献[8]提出了用于语义先验引导的纹理建模，探索了语义粒度纹理表示建模与压缩，以实现高质量的图像合成和可观的编码效率。此外，文献[8]还提出了一种跨通道熵模型，用于联合纹理表示压缩和重建优化。文献[16]进一步引入了结构建模，提出了一种一致性对比学习方法，通过将纹理表示空间与源像素空间对齐来优化纹理表示空间，从而获得更高的压缩性能。与特定应用领域中最先进的多功能视频编码（VVC）相比，文献[8,16]中提出的方法以超低比特率（<0.1 bpp）实现了卓越的视觉重构质量。

由于概念编码方法追求在极低的比特率下取得视觉上令人信服的重构结果，在评价时除了用户打分之外，通常会选择可学习感知图像块相似度（LPIPS）指标[17]作为量化的感知失真度量指标。在之前建立的基准[18]中，该指标已被证明与人类视觉感知而非信号保真度高度相关。为了进行性能比较，表1列出了VVC、典型的端到端学习图像编码方法（E2E）[19]、分层概念编码方法（LCIC）[7]和语义先验模型（SPM）[8]在FFHQ[20]和ADE20K[21]户外测试集中低比特率范围内的率失真性能。结果表明，与基于信号的压缩方法相比，

▼表1 VVC、E2E[19]以及概念编码方法LCIC[7]和SPM[8]在FFHQ和ADE20K户外测试集上的定量结果（LPIPS为失真指标）

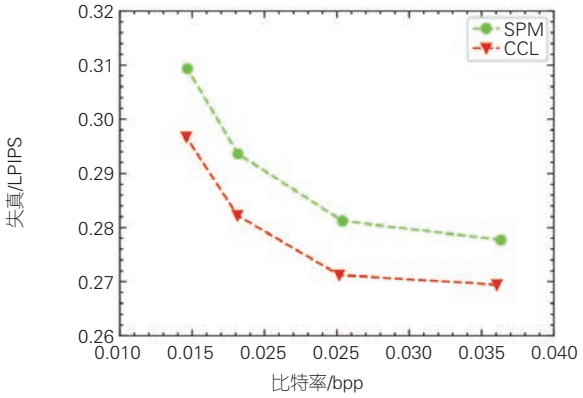
性能	数据集			
	FFHQ		ADE20K	
	比特率/（bpp）	LPIPS（10^{-2}↓）	比特率/（bpp）	LPIPS（10^{-2}↓）
VVC	0.045	36.900	0.035	58.300
	0.067	29.600	0.040	56.300
	0.075	27.400	0.047	54.000
	0.095	23.100	0.055	51.700
E2E[13]	0.039	33.400	0.016	62.800
	0.067	26.300	0.026	60.200
	0.071	25.600	0.035	53.300
	0.092	24.600	0.052	49.800
LCIC[8]	0.046	27.900	0.036	54.300
	0.055	26.800	0.046	52.000
	0.064	26.100	0.053	50.900
	0.074	25.900	0.061	50.300
SPM[9]	0.049	25.100	0.015	31.000
	0.063	24.100	0.018	29.400
	0.079	23.400	0.025	28.100
	0.110	23.100	0.036	27.800

E2E：端到端可学习图像编码方法
LCIC：分层概念编码方法
LPIPS：可学习感知图像块相似度
SPM：语义先验模型
VVC：多功能视频编码

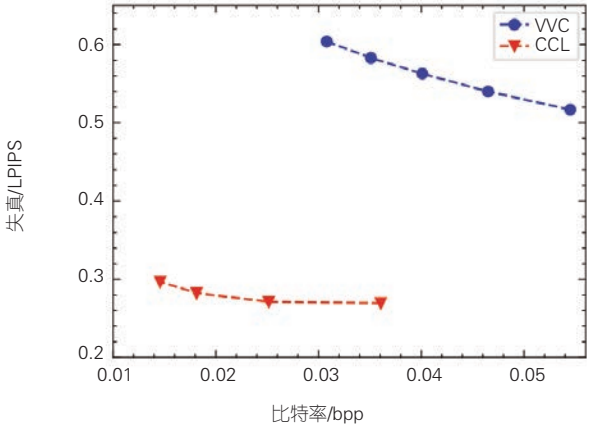
概念编码方法能够在特定领域以极低的比特率获得更高的视觉重构结果。此外，与FFHQ相比，LCIC在更具挑战性的ADE20K内容上表现一般，而FFHQ包含大量面部语义区域。相比之下，SPM在具有不同语义区域和纹理的挑战性场景中实现了重建质量的显著提高，验证了所提出的语义先验建模机制的有效性。此外，在ADE20K室外测试集的LPIPS指标上，VVC、SPM[8]和最新研究成果CCL[16]的率失真曲线如图5所示。对比结果验证了应用一致性对比学习方法所带来的重建质量的提高。大量实验表明，与之前的工作相比，概念图像编码在高效视觉表征学习、高效图像压缩（<0.1 bpp）、更好的视觉重建质量以及智能视觉应用（如控制和分析）方面都具有一定优势。

2 概念视频压缩

由于深度生成模型的强大功能，许多方法[15]将视频序列映射为潜在表示，并通过生成网络实现低比特率压缩重构。KONUKO等[22]基于FOMM等图像动画模型[23]，开发了一种用于视频会议的生成式压缩框架。WANG等[24]也提出了一种用于视频会议的说话人头部合成模型，通过自适应地从输入视频中提取三维关键点，实现了与H.264/AVC[25]相当的视频质



（a）SPM vs CCL



（b）VVC vs CCL

CCL：基于一致性对比学习的概念编码方法　SPM：基于语义先验建模的概念编码方法
LPIPS：可学习感知图像块相似度　VVC：多功能视频编码

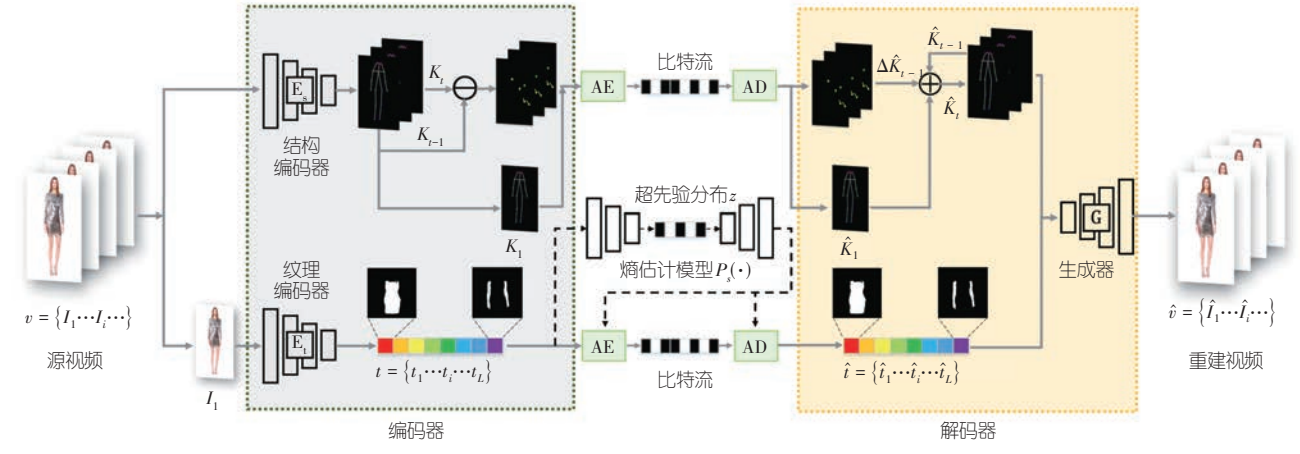
▲图5　SPM[8]、CCL[16]和VVC的率失真曲线

量，而带宽只有后者的十分之一。然而，在极高的压缩比（例如1 000倍）下实现高视觉质量的视频压缩框架仍待研究。

受近年来概念图像压缩的启发，DHVC[9]首次尝试利用分离的视觉表征进行极低码率的人体视频压缩，提出了概念视频压缩方法。在编码端，输入视频序列被分解为结构和纹理表示进行高效压缩。结构表示采用预先训练好的结构编码器来估计每个帧的人体姿势关键点，与传统视频编解码器中的运动矢量类似，计算每个关键点坐标的位移，作为两个帧之间运动信息的特征。为了节省比特率，编码时只传输第一帧的结构表示和后续帧的运动特征。纹理方面，纹理编码器将第一帧提取为语义级纹理表示，用于表示输入视频序列的纹理信息。为了确保所有帧的纹理一致性，引入对比学习[26]来优化纹理表示。在解码端，结构表示通过迭代进行重建，生成器根据纹理表示和结构表示还原视频。最后，引入纹理表示的熵估计模型，与对比学习相结合，建立率失真优化的端到端视频压缩框架，实现更有效的码率节省和更好的重建效果。

如图6所示，人体的主要结构信息可以通过人体姿势关键点有效地表示出来。结构编码器Es采用预先训练好的姿态估计器[27]，提取每帧图像的结构信息作为紧凑的结构表示。纹理编码器Et将图像帧提取为纹理表示。为了更好地捕捉每帧图像的纹理细节，分解组件编码（DCE）模块[28]被用来嵌入语义感知纹理表示。

为了确保同一视频中所有帧的纹理一致性，文献[26]在训练纹理编码器Et时引入了对比学习。选择同一视频中的帧作为正样本，其他视频的帧被视为负样本。此外，该框架



AD：算数解码器　AE：算数编码器

▲图6　概念视频压缩框架

还提出了基于语义层面的对比学习，并利用公式（1）计算语义级信息归一化交叉熵（infoNCE）损失[29]：

$$L_{cst}=-\sum_{i=1}^{L}\log\frac{\exp(t_i\cdot t_i^{-}/\tau)}{\sum_{j=1}^{Q}\exp(t_i\cdot t_{ij}^{-}/\tau)},\tag{1}$$

其中，t_i、t_i^{+}、t_i^{-}、τ、L和Q分别表示输入帧、同一视频中的另一帧、不同视频中的其他帧、温度参数、图像语义区域数量和负样本集的长度。这种技术使编码器既能利用正样本对(t,t^{+})的相似性，又能利用负样本对(t,t^{-})的不相似性。按照MoCo[26]，使用一个队列来存储先前输入帧的负样本t_i^{-}。这样，该模块就能在小批次下高效地进行对比学习。

表2列出了在Fashion数据集和TaichiHD数据集上测试LPIPS和DISTS指标的平均结果。其中，VVC和ArtAni方法压缩的比特率略高于DHVC[9]，但DHVC[9]使用框架仍在超低比特率下达到LPIPS和DISTS分数最低的性能，优于其他压缩框架。此外，表2中的定量结果进一步验证了将对比学习与压缩技术相结合可获得更好的视觉质量。总的来说，由于DHVC[9]将纹理和结构表征分离开来，因此与之前的方法相比，DHVC[9]实现了更高的视觉质量，从而产生了更清晰的结果，保留了更多细节，如面部特征和复杂背景纹理等。

▼表2 视频压缩方法比较（其中，分数越低代表视觉质量越好，w/o c.表示使用的模型不使用对比学习技术）

	Fashion[30]		Taichi[23]	
	LPIPS↓	DISTS↓	LPIPS↓	DISTS↓
VVC	0.268 7	0.300 9	0.314 7	0.270 9
ArtAni	0.177 7	0.227 3	0.301 1	0.249 1
Proposed (w/o c.)	0.110 9	0.168 7	0.215 3	0.220 6
Proposed	0.102 8	0.160 4	0.202 8	0.198 7

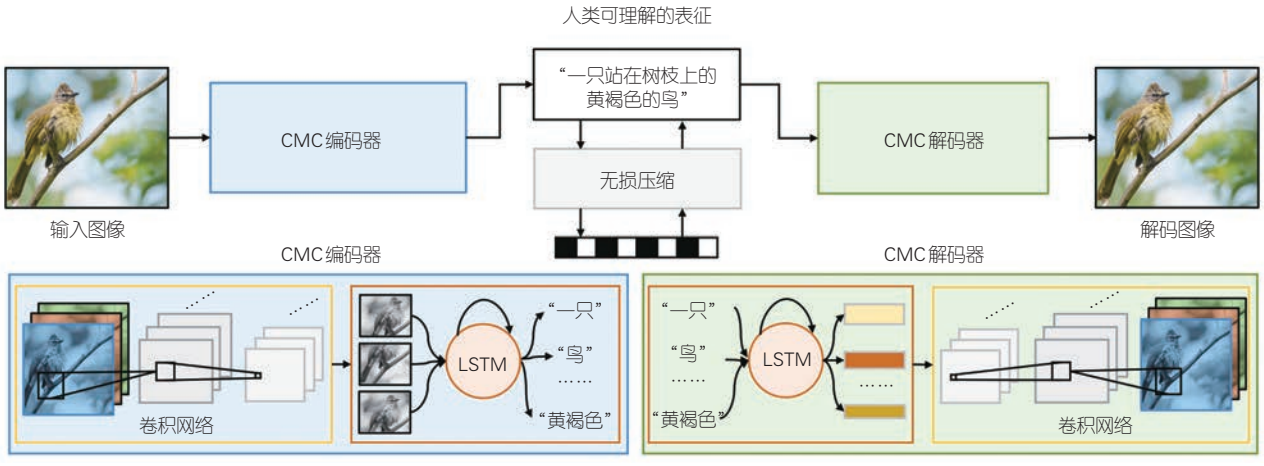
DISTS：深度图像结构与纹理相似度　　VVC：多功能视频编码
LPIPS：可学习感知图像块相似度

3 跨模态语义编码

概念压缩框架将视觉内容编码为从深度神经网络中提取的潜在变量等表示形式，这些表示形式对人类来说不易理解。人类可理解的表示形式，如文本、草图、语义图和属性等，在各种应用中具有重要意义，比如语义监测和以人为中心的应用。语义监测旨在监测语义信息，例如身份识别、人流量或车流量，而不是原始信号或潜在变量。以人为中心的应用旨在直接向人类用户传达视觉数据中蕴含的人类可理解信息。因此，我们提出跨模态压缩（CMC）[10]，旨在以超高压缩比将高度冗余的视觉数据转化为紧凑的、人类可理解的表示形式。

如图7所示，CMC框架由4个子模块组成：CMC编码器、CMC解码器、压缩域编码器和压缩域解码器。压缩过程也包括4个步骤：首先，CMC编码器将原始信号压缩成紧凑的、人类可理解的图像表示；其次，压缩域编码器以无损方式将图像表示编码为比特流；然后，压缩域解码器以无损方式从比特流中重建图像表示；最后，CMC解码器以语义一致的方式从图像表示中重建信号。CMC框架通过训练找到一个紧凑的压缩域优化了比特率，并在CMC编码器和解码器中保留语义以减小失真。

随着图像字幕技术[31]和文本引导图像生成技术[32]的发展，从图像生成高质量文本和从文本生成高质量图像变得更加可行。由此，我们建立了一个高效的图像–文本–图像的



CMC：跨模态压缩　　LSTM：长短期记忆网络

▲图7 跨模态压缩框架示意图

CMC范式，将图像压缩到文本域，而文本域具有通用性、普遍性和人类可理解性的特征。具体来说，采用经典的CNN-RNN模型[31]作为CMC编码器将图像压缩为文本。其中，CNN以图像为输入，用于提取图像特征，并将提取到的图像特征输入RNN以自动渐进的方式生成文本。哈夫曼编码[33]可用作组合域压缩的编码器/解码器，以无损方式减少文本的静态冗余。使用文本到图像的生成方面表现出色的AttnGAN[32]作为CMC解码器，用于从文本重建图像。通过在多个数据集上进行大量实验，CMC框架的有效性得到了验证。该模型以超高压缩比（4 000~7 000倍）获得了令人鼓舞的重构结果，显示出比广泛使用的JPEG基线[34]更好的压缩性能。

4 结束语

智能视频编码发展快速，未来或可开发出更先进的模型，进一步提高视觉信号的编码和表示效率。概念编码及跨模态编码技术提高了对使用压缩数据的下游任务的支持能力，开发了具有高度可解释性的潜在表征。通过使用这种表征，未来有可能开发交互式编码技术，从而实现一系列新颖的应用，如内容编辑和身临其境的交互，使其能够提供超越传统视频压缩方法的多功能特性和功能，为压缩领域带来了新的机遇。然而，智能视频编码领域也面临许多新的挑战，包括：1）数据安全性问题。智能视频编码从涉及信号信息的网络中提取的潜在表示可用于重建整个视频流，但这种表征没有加密，存在敏感信息泄露的风险。因此，可信和稳健的编码网络设计在实际应用中发挥着核心作用。2）模型泛化能力问题。考虑智能视频编码的落地实施和部署，关键是确定智能视频编码如何在实际应用领域得到应用，寻找到智能视频编解码器可以满足多样化需求的应用场景。例如，一些为户外场景训练的智能视频编解码器可能并不适合用于编码面部图像。但在实际应用中，采用多个模型进行场景适应是不切实际的。此外，推动智能视频编码标准与其他媒体数据标准相协调，将有助于智能视频编码在实际领域（例如移动设备上的短视频和沉浸式媒体应用）上应用的进程。3）标准制定问题。为了使符合智能视频编码标准的终端和系统能够无歧义地解码潜在表征，有必要通过定义适当的规则并给其分配相应语法元素来实现标准化。在系统层面，结构、语义和文本表征应由兼容的结构、语义或文本解码器正确解析。同时，符合智能视频编码标准的网络应能理解和处理智能模型层面的潜在表征含义。然而，如何可视化或分析高度紧凑的潜在表示比特流，评估现有智能视频编解码器的语义一致性是一项巨大的挑战。一些专业应用可能还需要对潜在表示域进行限制或扩展。如何支持特定领域的扩展和专业化，同时确保无歧义的一致性验证是一个关键问题，需要继续研究和探索。

参考文献

[1] NETRAVALI A N, STULLER J A. Motion-compensated transform coding [EB/OL]. [2023-12-12]. https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1979.tb02277.x

[2] MA S W, ZHANG X F, JIA C M, et al. Image and video compression with neural networks: A review [J]. IEEE transactions on circuits and systems for video technology, 2020, 30(6): 1683 - 1698. DOI: 10.1109/TCSVT.2019.2910119

[3] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//The 27th International Conference on Neural Information Processing Systems. ACM, 2014: 2672 - 2680

[4] KINGMA D P, WELLING M. Auto-encoding variational Bayes [EB/OL]. (2023-10-18) [2024-05-01]. http://arxiv.org/abs/1312.6114

[5] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models [C]//The 34th International Conference on Neural Information Processing Systems. ACM, 2020: 6840 - 6851

[6] CHANG J H, MAO Q, ZHAO Z H, et al. Layered conceptual image compression via deep semantic synthesis [C]//IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 694 - 698. DOI: 10.1109/ICIP.2019.8803805

[7] CHANG J H, ZHAO Z H, JIA C M, et al. Conceptual compression via deep structure and texture synthesis [J]. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society, 2022, 31: 2809 - 2823. DOI: 10.1109/TIP.2022.3159477

[8] CHANG J H, ZHAO Z H, YANG L B, et al. Thousand to one: semantic prior modeling for conceptual coding [C]//IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2021: 1 - 6. DOI: 10.1109/ICME51207.2021.9428366

[9] WANG R F, MAO Q, WANG S Q, et al. Disentangled visual representations for extreme human body video compression [C]//IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2022: 1 - 6. DOI: 10.1109/ICME52920.2022.9859831

[10] LI J G, JIA C M, ZHANG X F, et al. Cross modal compression: towards human-comprehensible semantic compression [C]//The 29th ACM International Conference on Multimedia. ACM, 2021: 10.1145/3474085.3475558. DOI: 10.1145/3474085.3475558

[11] KRÜGER N, JANSSEN P, KALKAN S, et al. Deep hierarchies in the primate visual cortex: what can we learn for computer vision? [J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(8): 1847 - 1871. DOI: 10.1109/TPAMI.2012.272

[12] ZHANG Y Z, HAN K, WORTH R, et al. Connecting concepts in the brain by mapping cortical representations of semantic relations. [EB/OL]. (2023-10-18) [2024-05-01]. https://www.nature.com/articles/s41467-020-15804-w

[13] GREGOR K, BESSE F, JIMENEZ REZENDE D, et al. Towards Conceptual Compression [C]//The 30th International Conference on Neural Information Processing Systems. ACM, 2016: 3556 - 3564

[14] GREGOR K, DANIHELKA I, GRAVES A, et al. DRAW: a recurrent neural network for image generation [C]//The 32nd International Conference on Machine Learning. PMLR, 2015: 1462 - 1471

[15] LOMBARDO S, HAN J, SCHROERS C, et al. Deep generative video compression [C]//The 33rd International Conference on Neural Information Processing Systems. ACM, 2019: 9287 - 9298

[16] CHANG J H, ZHANG J, XU Y M, et al. Consistency-contrast learning for conceptual coding [C]//The 30th ACM International Conference on Multimedia. ACM, 2022: 2681 - 2690
[17] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 586 - 595. DOI: 10.1109/CVPR.2018.00068
[18] LI Y, WANG S Q, ZHANG X F. Quality assessment of end-to-end learned image compression [C]//The 29th ACM International Conference on Multimedia. ACM, 2021: 4297 - 4305
[19] MINNEN D, BALLÉ J, TODERICI G D. Joint autoregressive and hierarchical priors for learned image compression [C]//The 32nd International Conference on Neural Information Processing Systems. ACM, 2018: 10794 - 10803
[20] KARRAS T, LAINE S, AILA T. A style-based generator architecture for generative adversarial networks [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 4396 - 4405. DOI: 10.1109/CVPR.2019.00453
[21] ZHOU B L, ZHAO H, PUIG X, et al. Scene parsing through ADE20K dataset [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 5122 - 5130. DOI: 10.1109/CVPR.2017.544
[22] KONUKO G, VALENZISE G, LATHUILIÈRE S. Ultra-low bitrate video conferencing using deep image animation [C]//Proceedings of ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021: 4210 - 4214. DOI: 10.1109/ICASSP39728.2021.9414731
[23] SIAROHIN A, LATHUILIÈRE S, TULYAKOV S, et al. First order motion model for image animation [C]//The 33rd International Conference on Neural Information Processing Systems. ACM, 2019: 7137 - 7147
[24] WANG T C, MALLYA A, LIU M Y. One-shot free-view neural talking-head synthesis for video conferencing [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2021: 10034 - 10044. DOI: 10.1109/CVPR46437.2021.00991
[25] RICHARDSON I. H.264 and MPEG-4 Video compression: video coding for next-generation multimedia [M]. New York, USA: John Wiley & Sons, Ltd, 2003
[26] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 9726 - 9735. DOI: 10.1109/CVPR42600.2020.00975
[27] CAO Z, HIDALGO G, SIMON T, et al. OpenPose: realtime multi-person 2D Pose estimation using part affinity fields [J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 43(1): 172 - 182. doi: 10.1109/TPAMI.2019.2929257
[28] MEN Y F, MAO Y M, JIANG Y N, et al. Controllable person image synthesis with attribute-decomposed GAN [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 5083 - 5092. DOI: 10.1109/CVPR42600.2020.00513
[29] OORD A VD, LI Y, VINYALS O. Representation learning with contrastive predictive coding [EB/OL]. (2019-07-22) [2023-10-18]. http://arxiv.org/abs/1807.03748
[30] ZABLOTSKAIA P, SIAROHIN A, ZHAO B, et al. DwNet: dense warp-based network for pose-guided human video generation [EB/OL]. (2019-10-21) [2023-10-18]. http://arxiv. org/abs/1910.09139
[31] VINYALS O, TOSHEV A, BENGIO S, et al. Show and tell: a neural image caption generator [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015: 3156 - 3164. DOI: 10.1109/CVPR.2015.7298935
[32] XU T, ZHANG P C, HUANG Q Y, et al. AttnGAN: fine-grained text to image generation with attentional generative adversarial networks [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 1316 - 1324. DOI: 10.1109/CVPR.2018.00143
[33] HUFFMAN D A. A method for the construction of minimum-redundancy codes [J]. Resonance, 2006, 11(2): 91 - 99. DOI: 10.1007/BF02837279
[34] WALLACE G K. Overview of the JPEG (ISO/CCITT) still image compression standard [J]. Proceedings of SPIE. The International Society for Optical Engineering, 1990: 1244. DOI: 10.1117/12.19537

作 者 简 介

郭怡琳，北京大学在读硕士研究生；主要研究领域为图像生成与图像压缩。

常建慧，北京大学在读博士研究生；主要研究领域为图像生成与图像压缩；在国际高水平会议及期刊上发表论文10余篇。

黄成，中兴通讯股份有限公司资深系统架构师、新一代视频编码标准预研项目经理，现任CCSA TC1 WG3信源编码组副组长、AVS系统组联合组长、TC28 SC29信标委多媒体分委会委员；主要研究方向为视频视觉编码与智能媒体业务系统。

马思伟，北京大学博雅特聘教授、视频与视觉技术国家工程研究中心副主任、国家杰出青年科学基金获得者、IEEE Fellow，并担任AVS标准视频组长，是“863计划”、国家重点研发计划首席专家；主要研究方向为视频编码与处理。

从2B到4B——电信行业与垂直行业的供需协同倍增发展

From 2B to 4B—Supply-Demand Synergy and Value-Multiplying Development of Telecom Industry and Vertical Industries

钟章队/ZHONG Zhangdui[1,2,3]，官科/GUAN Ke[1,2,4]，丁建文/DING Jianwen[1,2,3]，陈姝/CHEN Shu[5]
（1. 北京交通大学电子信息工程学院，中国 北京 100044；
2. 北京交通大学宽带移动信息通信铁路行业重点实验室，中国 北京 100044；
3. 轨道交通安全协同创新中心，中国 北京 100044；
4. 智慧高铁系统前沿科学中心，中国 北京 100044；
5. 佳讯飞鸿（北京）智能科技研究院有限公司，中国 北京 100044）
(1. School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing 100044, China;
2. Key Laboratory of Railway Industry of Broadband Mobile Information Communications, Beijing Jiaotong University, Beijing 100044, China;
3. Collaborative Innovation Center of Railway Traffic Safety, Beijing 100044, China;
4. Frontiers Science Center for Smart High-Speed Railway System, Beijing 100044, China;
5. Jiaxun Feihong Intelligent Technology Institute, Beijing 100044, China)

DOI：10.12142/ZTETJ.2024S1010
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20240723.1647.006.html
网络出版日期：2024-07-25
收稿日期：2023-12-10

摘要：为更好地发展下一代移动通信技术，加快5G/6G建设，需将发力点由“面向企业”的2B（To Business）向“为了企业”的4B（For Business）转变。从2B到4B，是从供给侧主导向需求侧主导的转变，是从“供给外生赋能”向“供需内生协同”本质的转变。发展5G/6G公专网，需要树立全生命周期的可持续发展理念，深度理解目标企业的核心诉求，充分发挥企业的主体作用。应由垂直行业主导公专网应用的标准制定与生态建设，设计可复制、可定义的商业模式，从顶层设计开始，将数字技术融入到垂直行业数字化转型之中，实现5G/6G公专网发展从2B到4B的转变，创造电信行业与垂直行业供需协同、价值倍增的可持续发展模式。

关键词：5G；6G；2B；4B；垂直行业；公专网

Abstract: In order to better develop the next generation mobile communication technology and accelerate the construction of 5G/6G, the focus needs to be shifted from 2B (To Business) to 4B (For Business). From 2B to 4B, it is a change from supply-side dominance to demand-side dominance, and from "exogenous empowerment of supply" to "endogenous synergy between supply and demand". To develop 5G/6G public-private network, it is necessary to establish the concept of sustainable development over the entire life cycle, deeply understand the core demands of the target enterprises, give full play to the main role of enterprises. Vertical industries should lead the standard setting and ecological construction of public-private network applications, and design replicable and definable business models. Starting from the top-level design, the digital technology is integrated into the digital transformation of vertical industries, realizing the transformation of the 5G/6G public-private network development paradigm from 2B to 4B, and creating a sustainable development model of supply-demand synergy and value multiplication for both telecommunication industry and vertical industries.

Keywords: 5G; 6G; to business; for business; vertical industry; public-private network

引用格式：钟章队, 官科, 丁建文, 等. 从2B到4B——电信行业与垂直行业的供需协同倍增发展 [J]. 中兴通讯技术, 2024, 30(S1): 67-75. DOI: 10.12142/ZTETJ.2024S1010
Citation: ZHONG Z D, GUAN K, DING J W, et al. From 2B to 4B—supply-demand synergy and value-multiplying development of the telecom industry and vertical industries [J]. ZTE technology journal, 2024, 30(S1): 67-75. DOI: 10.12142/ZTETJ.2024S1010

1 无线专网发展历史与趋势

2019年6月5G商用牌照发放，中国用4年时间建成了全球规模最大的5G网络，5G融入到了超六成的国民经济大类中。2023年6月，国际电信联盟（ITU）大会达成了6G愿景共识，标志着6G开启新阶段[1]。如何使5G、6G更好地与实体经济融合，驱动经济、政府和社会的数字化转型，成为全社会共同关心的话题。

在利用5G赋能千行万业的时代背景下，由电信网络运营商建设，供公共用户使用的无线公网难以适应行业市场多样化、特殊性的需求[2]。为满足不同行业内部的组织管理、

基金项目：中央高校基本科研业务费项目（2022JBQY004、2022JBXT001）；国家自然科学基金项目（62271043、62371033、62171021）；中国国家铁路集团有限公司科技研究开发计划项目（N2023G055）；教育部基金项目（8091B032123）

安全生产、调度指挥等需要，亟需依托网元虚拟化、功能服务化和编排智能化的技术保障，建设无线专网[1]。无线专网通信在集群调度、应急通信、即时通信等方面有着独特优势，广泛应用于国家安全、交通管理、建筑施工、机械制造等国民经济重要领域。

1.1 无线专网的发展历程

至今，公众移动通信系统经历了从1G到5G的发展，无线专网也伴随着移动通信的技术发展经历了1G到5G的更迭，技术标准制式多样。

第1代无线专网以模拟对讲和专用陆地集群移动通信（MPT-1327）为代表，用于公安、铁路、民航、政务等相关行业的语音通话、数据传输、调度指挥等业务，主要代表性的频段包括150 MHz、350 MHz、400 MHz、450 MHz等。

第2代无线专网包括铁路数字移动通信系统（GSM-R）、泛欧集群无线电（TETRA）、集成数字增强型网络（iDEN）、警用数字集群（PDT）、Project 25（P25）、数字移动无线电（DMR）、数字专用移动无线电（dPMR）等多种制式。GSM-R主要应用于铁路，提供调度通信、列控安全数据传输等业务。TETRA和iDEN主要用于城市轨道交通、公共安全、民航等。PDT为国产制式集群，主要用于公安、应急、无线政务等，提供集群调度通信业务，频段与TETRA和iDEN相同。DMR/dPMR主要用于商业和工业用户。

第3代无线专网包括全球开放式集群架构（GoTa）、多载波无线信息本地环路（McWiLL）等制式。GoTa基于时分同步码分多址（TD-SCDMA）技术进行定制开发，增加了图像传输，用于公共事业和社会服务、交通运输等行业。McWiLL是信威通信研发的移动宽带无线接入系统，主要用于油田、应急通信等场景。

第4代无线专网包括宽带集群通信（B-TrunC）、城市轨道交通车地综合通信系统（LTE-M）和铁路宽带移动通信系统（LTE-R）等制式。B-TrunC是基于分时长期演进（TD-LTE）的“LTE数字传输+集群语音通信”专网宽带集群系统标准，用于矿山、政务、机场、港口、电力、石油、矿山等。LTE-M是针对城市轨道交通需求设计的TD-LTE系统，调度语音采用B-TrunC制式，基于通信的列车控制系统（CBTC）和运行监控等业务采用标准LTE。LTE-R是针对铁路系统设计，调度语音早期采用公网对讲（PoC）制式，后期演进至关键任务服务（MCX）制式。

第5代无线专网包括5G独立专网、5G公专网等，此时5G面向企业（To Business, 2B）、垂直行业等概念开始出现。第3代合作伙伴计划（3GPP）定义了两种5G专网部署模式：公网专用和独立部署。公网专用是垂直行业可通过与运营商5G公网共享无线接入网（RAN），或者共享RAN和核心网控制面，或者端到端共享5G公网的方式来部署5G专网；独立部署是垂直行业独立部署从基站到核心网的整个5G网络，与运营商5G公网完全隔离[3]。

1.2 国际5G专网专用频谱划分情况

以德国、法国、瑞典、英国、日本、韩国等为代表的工业发达国家十分重视5G独立专网的发展。

截至2023年5月，德国联邦网络局在3 700～3 800 MHz频率范围内发放了321份5G本地网络频谱许可，服务对象包括奥迪、大众、宝马、空客、德铁、赫希曼、西门子等众多国际知名企业[4-5]，覆盖汽车工业、研究开发、物流自动化、道路运输、铁路运输等多个行业。截至2022年10月，法国共计核发13张3.8～4.0 GHz频段100 MHz本地5G专网实验执照，包含智慧医疗、智慧工厂与智慧城市等应用，目前取得5G专网实验执照的垂直行业单位包括法国原子能和替代能源委员会（CEA）、能源管理商Schneider Electric、法国电力集团（EDF）、史特拉斯堡大学附设医学研究中心、法国铁路公司（SNCF）等[6-7]。根据英国监管机构报告，自2019年推出共享频率接入许可计划以来，已发放了1 600余个许可。目前在3.8～4.2 GHz频段约有500个许可[6]。2021年11月，瑞典监管机构在3.5 GHz频段规划了40 MHz用于5G专网，制定了简要的管理办法并收取低额的频率使用费[6]。截至2022年11月，日本监管机构总务省（MIC）已向超126家机构发放了149张区域5G频率许可证，申请主体包括制造企业、电视公司、IT服务/系统开发商、政府机构、教育科研机构等，积极推动区域5G专网的行业应用[4,8]。韩国科学技术情报通信部（MSIT）于2021年1月26日发布《5G专网政策方案》，当时初步规划开放28 GHz（28.9～29.5 GHz）频段供5G专网使用。其后，MSIT于2021年6月29日发布《5G专网频率供应计划》，增加开放4.7 GHz（4.72～4.82 GHz）频段供5G专网使用[9]。截至2023年3月，MSIT已向10余个申请单位发放了专网频率许可。2020年11月，欧洲电子通讯委员会（ECC）发布ECC决定(20)22号文件[10]，为铁路移动无线电业务协调分配874.4～880 MHz、919.4～925 MHz及1 900～1 910 MHz专用频段，实现不同国家间使用统一频段的铁路无线通信系统进行跨境运输。这一铁路专网频段的分配极大地推动了未来铁路移动通信系统（FRMCS）的标准化制定、技术攻关、装备研制与生态建设。

表1针对2023年10月之前的国际5G专网专用情况行了

▼表1 国际5G专网专用频谱划分现状

国家	频率	行业/应用/申请主体现状
德国	3.7～3.8 GHz	覆盖汽车工业、研究与开发、物流自动化、道路与运输等多个行业
法国	3.8～4.0 GHz	包含智慧医疗、智慧工厂与智慧城市等应用
英国	3.8～4.2 GHz	已发放1 600余个许可，在3.8～4.2 GHz频段约有500个许可
瑞典	3.5 GHz	制定了简单的管理办法并收取非常低的频率使用费
日本	4.6～4.9 GHz、28.2～29.1 GHz	申请主体包括制造企业、电视公司、IT服务/系统开发商、政府机构、教育科研机构等
韩国	4.72～4.82 GHz、28.9～29.5 GHz	覆盖水电、航空、医疗、智能管理、交通运输等多个行业

注：上述发达国家对垂直行业专网频谱授权的成功实施，为中国5G专网频谱划分提供了借鉴和参考。

归纳整理。由表1可知，5G专网专用频谱主要集中在6 GHz以下（Sub-6 GHz），应用行业主要分布在工业制造、IT服务、科研教育等方面。

上述发达国家对垂直行业专网频谱授权的成功实施，为中国5G专网频谱划分提供了借鉴和参考。

1.3 中国5G行业专用频谱需求及相关政策

1）新频率划分的顶层设计和意义

2023年5月23日，工信部第62号令发布新版《中华人民共和国无线电频率划分规定》（简称《划分规定》），自2023年7月1日起施行。《划分规定》首次将6 GHz上半段（U6G）的6 425～7 125 MHz共700 MHz全部或部分频段划分用于5G/6G系统。国际电信联盟2023年世界无线电通信大会（WRC-23）还将进一步讨论决定5 925～6 425 MHz共500 MHz频段的最终分配。同时，中国还新增了24.25 GHz～27.5 GHz、37 GHz～43.5 GHz、66 GHz～71 GHz共14.75 GHz带宽的毫米波频段用于发展5G/6G。频率的确定是产业链起步的重要标志，能够支持5G/6G移动通信长远发展，同时保障各行业对频谱资源的中长期需求[11]。

长期以来，中国无线专网发展保持着对无线电频率的旺盛需求，支撑着国民经济、军队、国防、政务等建设和发展。特别是十八大以来，数字中国、数字经济、新基建、网络强国、海洋强国、制造强国、交通强国等一系列政策规划和战略文件，释放出对无线专网的大量刚需。企业数字化转型加快进行，迫切需要5G连接，以提供大带宽、低时延、高可靠等确定性保障[12]。2022年9月工信部下发了《5G全连接工厂建设指南》，指出“十四五”时期，要在全国重点行业领域，推动万家企业开展5G全连接工厂建设，建成1 000个分类分级、特色鲜明的工厂，打造100个标杆工厂，推动5G融合应用纵深发展。2023年中国发布《数字中国建设整体布局规划》，提出建设数字中国是数字时代推进中国式现代化的重要引擎，优化升级数字基础设施，大力推进产业数字化转型。5G行业专网被视为赋能千行万业、促进行业信息化、数字化、网络化、智能化转型的锚点，同时也是实现5G发展重点从民用消费领域转向以安全生产为主要目标的产业互联网领域的关键举措[13]，是行业高质量发展的必经之路。

2）5G的行业专用频谱需求及其划分

无线专网发展到今天，5G公网专用和5G专网专用是发展5G行业应用的两种重要模式。不同的部署模式适用于不同的场景，也决定了网络最终的能力和不同的生态。

当前，工业制造、钢铁、港口、矿山、电力、铁路、城市轨道交通等行业均对5G应用提出了需求[13]，垂直行业可根据自身特点，选择适合的5G部署模式。部分垂直行业由于安全生产作业、保障运输安全的红线，要求移动通信网络具有至少99.999%的高可靠性、双网覆盖的高冗余性、7×24小时的高可用性、网络及业务定制开发和运营维护的自主可控性、避免数据泄露和恶意攻击的高安全性等。5G公网专用无法为上述特定垂直行业提供定制化、确定性的网络服务，很难根据垂直行业管理和运行需求实时响应，无法与公众网络物理隔离，无法承担网络故障和服务不到位、响应不及时给垂直行业带来的安全作业风险和生命财产损失。上述垂直行业必须自建5G专网才能满足要求。自建5G专网需要国家给垂直行业分配5G专网频率。工信部在2022年11月给商飞发放了一张企业5G专网的频率许可，频率范围为5 925～6 125 MHz和24.75～25.15 GHz。该频段为中高频段，适用于工厂、园区、港口等局域场景，覆盖距离较短，站址密集，投资巨大，无法满足交通等行业对于广域覆盖场景的需求。

以铁路行业为例，随着智能铁路新业务需求的不断涌现和公网2G（GSM）退网带来的GSM-R产业链、生态链快速萎缩[14]，为确保铁路专用移动通信的可持续发展，国铁集团于2020年发布《关于加快推进5G技术铁路应用发展的实施意见》和《铁路5G技术应用科技攻关三年行动计划》，提出到2023年完成铁路5G专网关键技术攻关和主要专用设备研制，开展安全保障、出行服务等领域急需业务试验验证和试

用考核，完成5G专网主要技术标准制定。2023年9月，工信部向国铁集团批复了铁路5G专用移动通信系统（5G-R）试验频率（上行1 965～1 975 MHz，下行2 155～2 165 MHz），支持国铁集团开展5G-R系统外场技术试验，这将快速推动铁路通信技术升级换代，引领中国铁路朝向高质量数智化发展，为实现《数字铁路规划》中提到的2027年铁路数字化水平大幅提升、2035年铁路数字化转型全面完成的目标[15]奠定了基础。

截至2023年10月，针对中国已发布及正在申请的部分专网专用频段进行了统计，如表2所示。其中，中国交通运输行业专网专频的应用需求大。

3）对新时代行业无线专网发展的建议

为了更好地促进行业无线专网的生态建设和发展，满足铁路、城市轨道交通、民航、政务、公安等特殊行业对5G专网高可靠性、高安全性、高可用性、自主可控等需求，应尽快研究和布局中国无线专网的专用频谱划分，解决行业专网发展的“卡脖子”难题。

2 5G 2B发展现状与挑战

随着5G商用进程不断加速，行业应用与5G融合已成产业变革大势。从5G开始，技术与需求的底层逻辑开始变化：从服务个人转变为赋能千行百业。构建国家新一代信息基础设施、赋能垂直行业数字化转型升级，成为国家战略。由电信行业供给侧主导，以数字技术赋能产业，从供给侧对应用场景进行外向开拓的5G 2B业务发展如火如荼。本章将阐述中国5G 2B的发展现状、5G公专网商业模式、5G 2B可持续发展面临的挑战和存在的困难，并为支撑5G 2B可持续发展提出相关建议。

2.1 5G 2B发展现状

2021年，中国发布《5G“扬帆”计划》，大力推进5G行业应用。2022年9月，工信部发布《5G全连接工厂建设指南》，定义了工厂级、车间级、产线级三级应用场景，指引着5G应用向核心生产环境的规模化和纵深发展。产业界参与热度高涨，涌现出众多优秀创新案例，如中兴通讯作为5G设备龙头制造商，不仅帮助垂直行业客户打造全连接工厂，还将位于南京滨江基地的5G生产线打造为“5G全连接智慧工厂”，提出“用5G制造5G”的智能制造理念。2023年6月底，中国垂直行业虚拟专网的数量超过1.6万个，应用案例数超5万个。5G在工业、智慧城市、教育等行业的应用快速增长，成为信息化、智能化转型升级的重要数字底座。

2.2 5G公专网需求侧商业模式

从5G公专网需求侧的角度来看，主要有3种商业模式，其各自的工作方式、适用场景和特点在表3中进行了总结。

一是自建自维（SBSO）：需求企业拥有5G专用频谱，自行投资、建设和发展5G专用网络，自行日常维护和运营

▼表2　中国专网专用频谱划分现状

行业	行业专网	专频频段
民航	5G AeroMACS	5 091～5 150 MHz
制造	商飞5G制造专网试用频段	5 925～6 125 MHz、24.75～25.15 GHz
铁路	铁路专网：GSM-R、5G-R	GSM-R专用频段为885～889 MHz、930～934 MHz；5G-R试验频段为1 965～1 975 MHz、2 155～2 165 MHz
公路	公路专网	5 905～5 925 MHz
城轨	LTE-M	1 785～1 805 MHz
电力	电力专网	223～226 MHz、229～233 MHz
政府	政务专网	1 447～1 467 MHz
U6G频段	/	6 425～7 125 MHz（CHN45）

AeroMACS：机场场面宽带移动通信系统　GSM-R：铁路数字移动通信系统　5G-R：铁路新一代移动通信系统

▼表3　5G公专网需求侧商业模式

模式	工作方式	适用场景及特点
自建自维（SBSO）	需求企业拥有5G专用频谱，自行投资、建设和发展、运营5G专用网络	适用于对网络控制力要求较高、有足够资金和技术实力的企业，能够实现更高的自主权和灵活性
代建代维（BOT）	需求企业委托专业的网络运营商或服务提供商进行5G公专网的建设和运营	适用于希望将网络建设和运营风险外包给专业服务商的企业，以降低初始投资和运营成本
共建共维（CBCO）	供需双方共同投资和建设5G专用网络	适用于多个企业在相同地区或行业内共同需要5G专网的情况，通过共同投资和共享资源，提高网络的覆盖范围和性能

网络，根据自身需求和价值来定制、优化网络。此模式适用于对网络控制力要求较高、有足够资金和技术实力的企业，能够实现更高的自主权和灵活性。

二是代建代维（BOT）：需求企业委托专业的网络运营商或服务提供商进行5G公专网的建设和运营。后者负责投资建设、运营维护和网络的可持续发展，根据需要可以在合同期满后将网络的所有权和运营权转移给委托企业。代建代维适用于希望将网络建设和运营风险外包给专业服务商的企业，以降低初始投资和运营成本。

三是共建共维（CBCO）：CBCO是指供需双方共同投资和建设5G专用网络，并共享网络资源和基础设施，需求方可以是多个企业，供给方可以包括多个投资商。供需双方通过合作和资源共享，共同承担网络建设和维护的成本，以实现资源的最优利用和协同运营。共建共维适用于多个企业在相同地区或行业内共同需要5G专网的情况，通过共同投资和共享资源，提高网络的覆盖范围和性能。

2.3 5G公专网供给侧商业模式

从5G公专网供给侧的角度来看，尽管三大运营商对5G公专网部署方式命名不同，但彼此间架构相似，主要有3种模式，其各自的技术方案、应用场景和服务模式在表4中进行了总结。

一是广域专网:（公网共用）：与2C网络完全共享，通过网络切片建立专用链路，形成全国或区域广域覆盖。

二是局域专网（公网专用）：通过用户平面功能（UPF）下沉和移动边缘计算（MEC）部署，实现本地流量卸载、边缘数据处理，形成局域开放园区、热点地区的覆盖场景。

三是物理专网（专网专用）：在模式二的基础上，核心网、承载网、基站、频率专建专享，形成封闭区域、热点地区的专门覆盖场景。

2.4 5G 2B可持续发展

1）5G 2B发展面临的挑战

5G 2B发展时间较短、市场规模还不够大，垂直行业需求多样、差异性大、对产品要求多样化，端到端应用集成和维护成本较高，商业模式尚处于探索阶段，规模化复制和拓展仍然任重道远。5G 2B全生命周期的可持续发展主要面临以下挑战：

一是网络建设：缺少整体建设和应用规划，网络建设和终端成本高，网络定制化要求高，缺乏标准体系支撑。

二是业务运用：对垂直行业业务的可靠承载缺少有效验证，垂直行业对5G网络不可知、不可测，垂直行业用户难以评估5G公专网服务质量和业务服务质量。

三是管理维护：不同垂直行业的需求不同，有些期待免维护、少维护，有些期望能够深入介入运维，实时掌握网络运行状态，网络发生故障时能够及时倒换至冗余备用网络，对网络及业务故障的原因要做到“件件分析、定位准确、及时解决”。

四是信息安全：不同垂直行业对安全的级别要求千差万别，有些采用逻辑切片就可以满足要求，有些要求2B和2C网络实现物理隔离。

2）5G 2B发展存在的困难

从根本上讲，5G 2B的发展需要电信行业和垂直行业协同创新合作。实际上，双方在需求、目标、组织、管理等方面存在较大差异，不同行业经济基础不同，导致供需不匹配，具体体现在以下3个方面：

一是缺乏共识和应用标准：当前，5G公专网缺乏应用标准规范，网络和服务定制化水平不足，运营商缺少行业专业知识和经验，核心、无线、承载网络冗余不足，冗余倒换时间长、业务中断长，上行业务带宽不够，网络服务质量和业务服务质量监测、运维响应时间等需求在不同行业千差万别，运营商响应与行业需求不匹配。上述问题制约5G 2B发展和应用。

二是理念和管理模式差异大：数字资产管理模式不同，部分垂直行业需要全生命周期管理和维护，公专网的运维平台能力开放程度不够。此外，数字产业和垂直行业“步调不一致”。前者技术进步快、迭代快、换代快，1G到6G，10年一代，软硬件版本更新快；后者需求牵引、价值导向，步步为营，要求全生命周期管理和维护。前者开放，后者

▼表4　5G公专网供给侧商业模式

模式	广域专网(公网共用)	局域专网(公网专用)	物理专网(专网专用)
技术方案	与公网完全共享，通过5QI/网络切片建立专用链路	与公网部分共享，通过UPF下沉、MEC部署等实现本地流量卸载、边缘数据处理	在UPF下沉的基础上，基站、频率专建专享，构建专用无线网络、SA核心网
应用场景	广域覆盖场景	局域开放园区、热点地区场景	局域封闭区域、热点地区场景
服务模式	优享模式(中国移动)、虚拟专网(中国联通)、致远模式(中国电信)	专享模式(中国移动)、混合专网(中国联通)、比邻模式(中国电信)	尊享模式(中国移动)、独立专网(中国联通)、如翼模式(中国电信)

5QI：5G QoS特性　MEC：移动边缘计算　SA：独立组网　UPF：用户平面功能

保密。

三是垂直行业的主动性、主导性没有充分释放：针对技术复杂的5G公专网，垂直行业缺少专业人才，积极性不高，没有将需求充分地梳理和总结并提供给电信营商。

3）5G 2B可发展之道

5G 2B的可持续发展，既不能“千行一面”，也不能“千行千面”，而是要“精准匹配，高质量发展”。5G公专网的本质是输出基于5G网络的数字技术能力，解决垂直行业的痛点问题，为垂直行业提供发展动能，创造价值，重塑生产要素和经济结构。具体有3点发展建议：

一是电信运营商、设备提供商要理解垂直行业数字化转型的顶层设计，深谙垂直行业的发展痛点和价值导向，提供垂直行业真需求的真落地解决方案。

二是垂直行业主导，协同构建基于5G公专网开展垂直行业应用的技术标准体系，网络安全技术体系，以及跨行业的网络、管控、运用、维护一体化平台。

三是协同创新高可靠多模、多功能的公专融合终端技术，跨行业、多域数据联合分析与挖掘技术，网络、应用和安全一体化的融合管控技术。

5G公专网是垂直行业数智化转型的关键基础设施，通过“5G公专网+人工智能+垂直行业应用”的深度融合汇通，让5G公专网深入生产，让垂直行业场景质变升级，加速推进垂直行业数字化转型、智能化升级，助力经济、社会高质量发展。

3 重塑垂直行业5G发展体系

为使5G更好的赋能垂直行业，实现垂直行业数字化转型发展，应将发展的内在逻辑从“以数字技术赋能产业，供给侧对应用场景外向开拓”转变为“从产业角度出发凝练技术特征，需求侧对技术反向定义”，即从2B（To Business）向4B（For Business）进行转变，重塑垂直行业5G发展体系。

3.1 数字化转型发展

党的十九大以来，中共中央、国务院，相关部门和各省市频频释放政策信号，特别是党的二十大报告提出了中国式现代化。2019年7月中央政治局会议“加强推进信息网络等新型基础设施建设”，随后国家发展改革委、工信部明确新基建范围，相关部门和各省市出台新基建行动计划，加快5G/6G、人工智能、工业互联网、物联网、数据中心等建设速度。2020年8月国务院国资委印发《关于加快推进国有企业数字化转型工作的通知》，相关部门频频释放政策信号，国企数字化转型持续加速、有序推进。2023年2月27日中共中央、国务院印发了《数字中国建设整体布局规划》，数字中国战略将引领国家实现经济体系现代化、科技创新现代化、国家治理体系与治理能力现代化和人的现代化，探索出一条具有中国特色的中国式数字化发展新道路。2023版国家机构改革，新组建国家数据局，中央部署要协调推进数据基础制度建设，统筹推进数字中国、数字经济、数字社会规划和建设。2022年，中国数字经济规模超过50万亿元，到2032年，将超过100万亿元。数字经济是国家发展的重要新引擎，数据上升为核心要素。

行业的每一项数字化转型活动，都应围绕组织的价值效益来展开。数字化转型在根本上是要推动行业组织价值体系的优化、创新和重构，不断创造新赛道、新模式。通过顶层设计，打造行业的数字化生存和发展四方面能力：一要适应环境的快速变化，深化应用新一代信息技术，建立、提升、重构内外部平台能力；二要赋能业务加速创新转型，构建竞争合作新优势；三要改造提升传统动能形成新动能，形成不断创造新价值、实现新发展的能力；四要基于智能技术的赋能作用，获取多样化发展效率，形成共创、共建、共享开放能力，应对日益个性化、动态化、协同化的需求，构建市场核心竞争力。

5G公专网是垂直行业数字化转型的基石。当前，5G 2B公专网的发展方式由供给侧主导，其本质是“供给外生赋能、实现新增长”。为实现数字化转型升级，应将发展方式转变为由需求侧主导，本质为“供需内生协同、实现转型升级”的5G 4B。

3.2 5G/6G公专网发展

在政策的引领下驱动垂直行业实施数字化转型发展，行业数字化转型与智能化升级需要通过顶层设计，实施战略驱动，最终构建发展能力。面向垂直行业需求，将发展理念由2B转变为4B，实现主导力量、内在逻辑和本质的全面革新，准确把握5G/6G公专网发展路线，树立全生命周期的可持续发展理念。如图1所示，全生命周期需要包含需求定义、规划设计、实施部署、运维管理和更新迭代五大阶段。各阶段均需要运营商和垂直行业供需双方形成契约，紧密合作与协调，确保网络能够满足垂直行业需求，产业链能够可持续发展。

1）发展基础

公专网发展的基础是确立好运营商和垂直行业用户之间商业模式。运营商与垂直行业之间的契约服务关系，在对等性、业务复杂性、网络建设模式、服务模式、收益模式等方

面更具有特殊性。设计可复制、可定义的5G/6G公专网商业模式需要考虑以下几个关键要素：

一是面向行业需求有针对性地设计商业模式。不同垂直行业具有不同的需求和商业模式，有些垂直行业采用公网共用、定制服务质量（QoS）采购流量包形式，有些垂直行业需要公网专用、部署专用UPF，有些垂直行业需要专网专用、部署专用频段基站和专用UPF，甚至专用核心网，需要针对性地设计商业模式。

二是确定网络部署模式。明确网络部署的出资模式和资产归属（运营商出资、垂直行业出资、运营商与垂直行业按照比例共同出资），定义网络部署过程中运营商与垂直行业的分工界面。

三是确定收益模式。设计清晰的收益模式，包括收入来源、价格策略和付费方式。不断优化和改进收费模式、成本回收期限和盈利模式，如设备租费模式、流量收费模式、运维服务收费模式等，以适应不同垂直行业的需求，使垂直行业“用得起、用得好、用得放心、用得省心”，实现双赢。

四是建立长期合作伙伴关系。运营商、设备供应商、解决方案提供商与垂直行业之间建立长期的合作伙伴关系，共同推动公专网商业模式的实施和发展。通过合作伙伴的资源和技术能力，增强商业模式的可复制性和可定义性。

2）发展关键

公专网发展的关键是由垂直行业主导，协同运营商，从国家标准、行业标准、团体标准、企业标准等多个层面，构建5G/6G公专网垂直行业应用的技术标准体系。标准体系的构建需要遵循“统一高效、保障安全、健全体系、适应发展、强化基础、鼓励创新、统筹设计、突出重点”的原则，为装备技术、工程建设、运营维护等应用领域提供依据，形成谱系化。通过应用标准体系建设，实现以下目标：

一是促进互操作性。通过制定统一的技术标准和协议，不同供货商、设备和应用之间的互联互通，确保业务、功能、性能满足垂直行业需求。

二是提升网络安全性。加强网络安全防护和管理，确保数据传输的安全性和隐私保护，减少网络攻击和数据泄露的风险，提升5G/6G公专网的安全性。

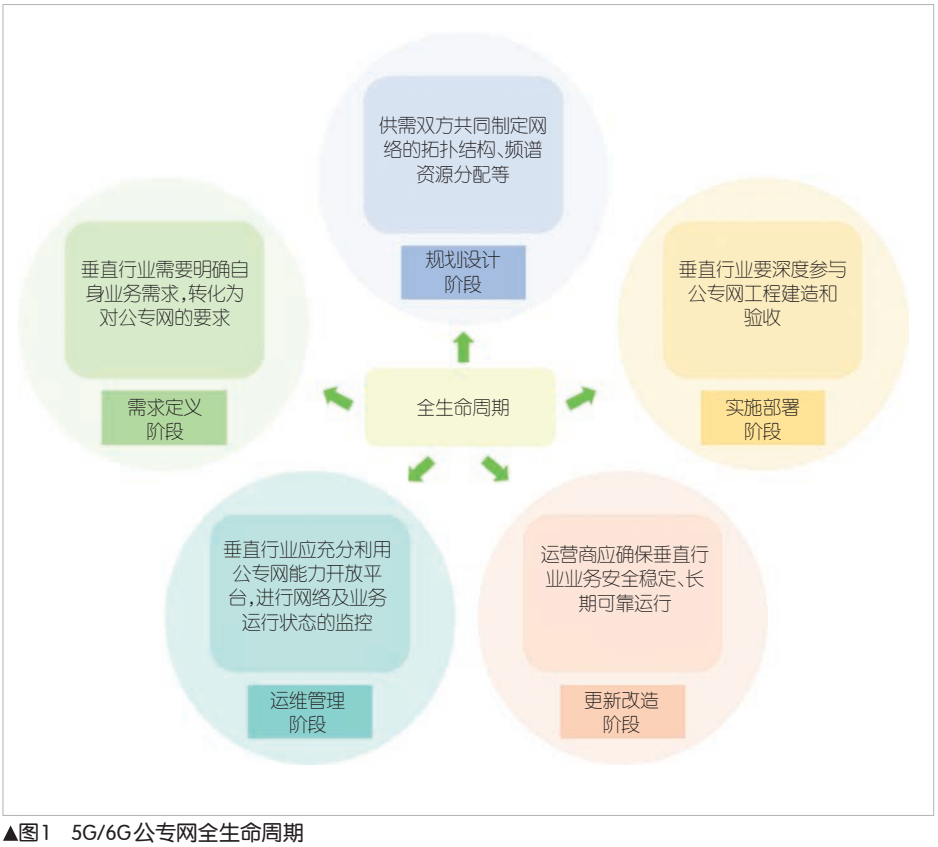
三是保证服务质量和用户体验。通过制定标准化的服务质量指标和用户体验指标，可以统一运营商和垂直行业对于高质量服务的认知，增强垂直行业用户对于5G/6G公专网的接受度和认可度。

四是降低成本和提高效率。通过制定标准化的技术规范，可以减少自定义开发和集成的工作量，降低数字化建设的时间和成本，激发市场创新和发展，推动5G/6G公专网的成熟和普及。

五是促进产业合作和创新。标准化可以提供一个共同的平台和语言，运营商可以深入理解垂直行业需求和发展趋势，参与技术研发和创新应用，推动垂直行业和运营商之间的合作和共享。

3）发展核心

公专网发展的核心是打造好创新链、产业链。理论创新要聚焦行业的数字化转型、价值导向、精品网络的规划；技术与标准创新要聚焦业务需求、业务承载、安全保障，云边端协同智能；关键装备创新要聚焦管控平台、日常运维监测、定制终端；系统集成创新是公专网建设的重点，包括工程实施策略、装备与管理体系协同、试验与验收方法、应用集成管控办法等。针对上



▲图1　5G/6G公专网全生命周期

述4个方面的创新需求，其具体建设目标及相应技术方案如图2所示。

4）发展保障

公专网发展的保障是做好生态建设。

一要深度理解行业的需求，建立广泛良好的合作伙伴关系。深入了解不同垂直行业或企业的需求和挑战，包括业务流程、数据管理、安全性要求等方面。通过与行业相关的研究和交流，洞察行业的数字化转型和智能化升级的趋势，为生态建设提供指导。建立广泛的合作伙伴关系，包括运营商、设备供应商、解决方案集成商、行业协会等。与各方合作，共同开发适用于垂直行业或企业的5G/6G公专网解决方案。通过合作伙伴的资源和技术能力，实现生态系统的丰富和互补。

二要标准和规范引领，在3GPP的源头纳入垂直行业的需求，贯通系统网络的全生命周期建设。参与制定行业标准和规范（例如未来铁路移动通信系统FRMCS），以推动5G/6G公专网在企业中的应用。标准化能够提升互操作性和互联互通性，促进不同解决方案和设备的无缝集成和互联。

三要政产学研用协同创新，实现创新链、产业链、供应链的协调、可持续发展。鼓励创新和研发活动，包括技术创新、应用创新和商业模式创新。通过支持创新企业和创业团队，培育新兴技术和解决方案，为垂直行业或企业的数字化转型和智能化升级提供新的驱动力。为垂直行业或企业提供相关的培训和支持，帮助其了解和掌握5G/6G公专网的应用和技术。培训内容可以包括网络架构、安全性、数据管理、智能化应用等方面，以提升用户对公专网的认知和能力。通过建立应用示范项目，展示5G/6G公专网在垂直行业或企业中的应用案例和效果。通过推广成功案例，向更多的行业和企业展示公专网的潜力和价值，推动其采纳和应用。建立监测和反馈机制，持续跟踪垂直行业或企业的需求和反馈。通过收集用户反馈和市场信息，及时调整和优化生态建设的策略和方向，以不断提升公专网的适应性和价值。

4 结束语

无线专网是企业或行业数字化转型、智能化升级的基石，是行业融合基础设施的底座。长期以来，中国无线专网发展保持着对无线电频率的旺盛需求，频率的确定是产业链起步的重要标志，也是保障千行百业高质量发展的关键要素。5G是2B的起跑线，对于发展5G 2B的初心，国家、行业、企业有不同的视角。然而，当前5G公专网发展面临较大挑战和困难，为打破5G公专网的发展瓶颈，应将发展模式由2B转变为4B。就未来5G/6G公专网发展而言，要树立全生命周期可持续发展理念，明确发展的基础是设计可复制、可定义的公专网商业模式；发展的关键是建设由垂直行业主导，与运营商协同，从多个层面出发的公专网应用标准体系；发展的核心是打造理论、技术与标准、关键装备及系统集成的创新链和产业链；发展的保障是做好公专网生态建设。在构建国家新一代信息基础设施的战略背景下，以行业政策引领为动能，供需协同、双赢倍增为动力，实现电信行业无线专网从2B到4B的发展范式转变，赋能垂直行业数字化转型升级。

理论创新	技术与标准创新	关键装备创新	系统集成创新
1）面向垂直行业数字化业务和需求的新型网络架构 创新组网方案、高可靠保障方案、安全保障体系，突破安全与效率相互制约的瓶颈 2）公专网系统安全模型与失效影响 构建网络安全风险源辨识方法和风险源场景库，建立单点失效、多点失效的影响模型及对应的处理机制 3）基于自主化射线跟踪（RT）技术的公专网规划与优化 构建BIM+GTS融合数据库，提出机理模型与数据驱动相融合的5G/6G公专网无线网络规划优化方法，增强无线网络效能	1）面向垂直行业数字化业务需求的公专网技术标准体系 构建装备技术、工程建设、运输服务等公专网技术标准体系 2）基于5G/6G公专网的业务承载技术 基于能力匹配模型、覆盖模型和射线跟踪仿真技术的规划 3）基于5G/6G公专网的网络安全保障技术 技术方案和安全评测方法，自主化加密算法、用户隐私你护机制和数据完整性保护机制 4）垂直行业5G/6G公专网边缘计算与应用技术 “云-管-边-端”协同方案：数据恢复、主备/双活、业务切换及自愈灾备技术	1）5G/6G公专网的一体化管控平台 创新网络跨行业“共管共维”模式，研发“网络、终端与应用”二位一体管控平台 2）基于5G/6G公专网的通信系统 研发基于3GPP MCX国际标准的5G/6G公专网通信系统，并成功应用 3）面向垂直行业智慧业务的一体化智能综合监测系统 创新性实现网络服务质量和业务运行质量的有效评估和预警 4）公专网成套专用终端装备 研制谱系化、简统化多模终端装备，解决垂直行业装备多样等问题	1）提出符合垂直行业数字业务需求的公专网实施策略 “搭建实验室测试平台——试验段试验工程——试验线示范工程”三步走策略 2）跨垂直行业和电信行业的装备与管理体系协同创新 联结“产、学、研、用”全链条，实现协同创新 3）面向垂直行业公专网和业务应用的测试试验方法 建立基于通信云高效的测试试验体系 4）应用集成创新 支撑垂直行业自动驾驶、可视化无线调度、基础设施监测等应用创新

▲图2　5G/6G公专网创新需求及其相应技术方案[16-18]

致谢

本研究中的调研工作和图表制作工作得到了北京交通大学在读硕士研究生单馨漪、郭梓烨、乔琬淇、徐航和佳讯飞鸿智能科技研究院龙志勇、刘艳兵、柴文宇、李莉的帮助。在此向他们表示感谢！

参考文献

[1] 程锦霞, 邓伟, 翁玮文, 等. 面向6G的天地一体无线网络技术研究[J]. 无线电通信技术, 2023, 49(5): 1-7
[2] 雒艳. 5G行业专网建设模式探索 [J]. 价值工程, 2023,42(24): 94-98
[3] 祝啄升, 魏长水, 张骁. 5G公网铁路专用网络架构及安全部署方案[J]. 铁道通信信号, 2023, 59(1): 13-18
[4] 汪卫国, 于青民. 国际5G专网应用发展态势 [J]. 通信世界, 2023, (10): 24-27
[5] 5G private networks [EB/OL]. [2022-11-18]. https://5gobservatory.eu/5g-private-networks/
[6] 部分欧洲国家5G专网发展动态 [EB/OL]. [2022-11-18]. https://www.c114.com.cn/wireless/2935/a1215820.html
[7] Platform for businesses and manufacturers trials in the 3.8-4.0 GHz band: arcep delivers an initial assessment [EB/OL]. [2022-12-12]. https://en. arcep. fr/news/press-releases/view/n/5g-121022.html
[8] 刘琪, 潘峰, 姜博. 日本区域5G专网发展分析及思考 [J]. 信息通信技术与政策, 2022, 2022(6): 75-79
[9] Establishment of fifth-generation (5G) private network policy plan [EB/OL]. [2020-01-25]. https://en. arcep. fr/news/press-releases/view/n/5g-121022.html
[10] European Electronic Communications Committee. Harmonised use of the paired frequency bands 874.4-880.0 MHz and 919.4-925.0 MHz and of the unpaired frequency band 1900-1910 MHz for Railway Mobile Radio (RMR) [R]. 2020
[11] 工信部产业政策与法规司. 法管理频谱资源 促进资源合理有效使用 [N]. 中国电子报, 2023-07-04(2)
[12] 庞明慧, 台鑫, 吕崇玉, 等. 面向5G无人机通信场景的传播路径概率预测模型 [J]. 电波科学学报, 2023, 38(1): 54-62
[13] 艾渤. "面向垂直行业场景的5G及B5G电波传播与无线信道研究"专题前言 [J]. 电波科学学报, 2023, 38(1): 1
[14] AI B, MOLISCH A F, RUPP M, et al. 5G key technologies for smart railways [J] Proceedings of the IEEE, 2020, 108(6): 856-893
[15]《数字铁路规划》印发：到2035年铁路数字化转型全面完成 [EB/OL]. [2023-09-11]. https://baijiahao. baidu. com/s? id=1776734061887064368&wfr=spider&for=pc
[16] HE D, GUAN K, YAN D, et al. Physics and AI-based digital twin of multi-spectrum propagation characteristics for communication and sensing in 6G and beyond [J]. IEEE journal on selected areas in communications, 2023, 41(11): 3461-3473
[17] 钟章队, 官科, 陈为, 等. 铁路新一代移动通信的挑战与思考 [J]. 中兴通讯技术, 2021, 27(4): 44-50. DOI: 10.12142/ZTETJ.202104009
[18] YAN W, SHU Q, GAO P. Security risk prevention and control deployment for 5G private industrial networks [J] China communications, 2021,18(9): 167-174

作 者 简 介

钟章队，北京交通大学教授、博士生导师，教育部"面向高速铁路控制的无线移动通信系统研究"创新团队带头人，宽带移动信息通信铁路行业重点实验室主任；从事无线通信与宽带移动通信、计算机通信与信息技术等研究与教学；1994年提出基于GSM-R技术建设中国铁路数字移动通信网络，奠定高速铁路CTCS3级列控系统发展基础；完成100多项科研项目，研究成果广泛应用于青藏铁路、大秦重载运输铁路、客运专线、高速铁路等工程建设；获国家科技进步奖一等奖1项，省部级科技特等奖1项、一等奖3项、二等奖5项，中国图书优秀学术著作一等奖1项，中国高等学校十大科技进展1项，中国研究生教育成果奖二等奖1项，中国电子学会优秀博士学位论文指导导师奖；1998年获铁道部有突出贡献的中青年科技专家称号，1999年享受国务院政府特殊津贴，2004年获茅以升科学技术奖（铁道科技奖），2007年获第八届詹天佑铁道科学技术奖贡献奖，2010年获得第十届詹天佑铁道科学技术成就奖。

官科，北京交通大学教授、博士生导师，先进轨道交通自主运行全国重点实验室攻关团队成员，宽带信息通信铁路行业重点实验室副主任，太赫兹通信标准《IEEE 802.15.3d-2017》的信道模型主创者，《IEEE Vehicular Technology Magazine》《电波科学学报》等期刊的编委；研究领域为5G、毫米波/太赫兹以及智能轨道交通电波传播与无线信道；获德国洪堡基金会外国科学家研究基金资助，并获国际无线电科学联盟（URSI）青年科学家奖、中国铁道学会科学技术奖一等奖、教育部高等学校科学研究优秀成果奖二等奖。

丁建文，北京交通大学研究员、博士生导师，宽带移动信息通信铁路行业重点实验室副主任，国家铁路局铁路科技标准规划专家；长期从事宽带移动通信、轨道交通专用移动通信等领域研究；主持和参与国家级、省部级和企业科研项目100余项，主持和参编铁道行业标准、企标及标准性技术文件40项，获詹天佑铁道科学技术青年奖1项，中国铁道学会科学技术奖一等奖4项、二等奖4项，国家铁路局重大科技成果入库9项，成果应用于铁路移动通信网络规划、设备研制、测试试验、工程设计、优化与验收；发表学术论文80余篇，编写著作9部。

陈姝，佳讯飞鸿智能科技研究院院长、宽带移动信息通信铁路行业重点实验室副主任；主要从事轨道交通宽带移动通信、云计算、大数据技术研究；参与多项铁路标准、行业白皮书编制，主持了公司铁路通信云、铁路5G专用网关、铁路信号设备PHM系统等多项新产品的孵化工作。

3D IC系统架构概述

An Overview of 3D IC System Architecture

陈昊/CHEN Hao[1,2]，谢业磊/XIE Yelei[1,2]，
庞健/PANG Jian[1,2]，欧阳可青/OUYANG Keqing[1,2,3]
（1. 移动网络和移动多媒体技术国家重点实验室，中国 深圳 518055；
2. 深圳市中兴微电子技术有限公司，中国 深圳 518081；
3. 射频异质异构集成全国重点实验室，中国 深圳 518061）
(1. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China;
2. Sanechips Technology Co., Ltd, Shenzhen 518081, China;
3. State Key Laboratory of Radio Frequency Heterogeneous Integration, Shenzhen 518061, China)

DOI：10.12142/ZTETJ.2024S1011
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20240909.1731.002.html
网络出版日期：2024-09-09
收稿日期：2023-11-25

摘要：随着芯片制造工艺接近物理极限，使用多Die堆叠的三维集成电路（3D IC）已经成为延续摩尔定律的最佳途径之一。利用3D IC将芯片垂直堆叠集成，可以极大程度降低互联长度，提升互联带宽。详细介绍了一些常见的3D IC系统架构方案，说明了使用不同3D架构对于整体芯片系统在性能、功耗等方面的优势，也列举了在物理实现、封装测试、工艺能力等方面的挑战。最后综述了一些业内使用3D IC的典型产品，并介绍了这些产品的系统架构、典型参数、适用领域，以及使用3D IC后给产品带来的竞争力提升情况。针对业界现状，认为应该把握机遇，不惧挑战，实现弯道超车。

关键词：三维集成电路；三维堆叠芯片；三维片上系统；存储堆叠逻辑；逻辑堆叠逻辑

Abstract: As the chip manufacturing process approaches its physical limits, multi-die stacking 3D integrated circuit (IC) technology has emerged as a promising approach to sustain Moore's law. Integrating chips vertically with 3D IC can significantly reduce interconnection length and improve interconnection bandwidth. This paper provides a detailed overview of common 3D IC system architecture solutions and discusses the advantages of using different 3D architectures in terms of performance, power, and area. It also outlines the challenges related to physical implementation, packaging, testing, and process capability. This paper summarizes some typical commercial products that utilize 3D IC technology and introduces their system architecture, typical parameters, applicable fields, and competitiveness improvement. Considering the current industry landscape, the paper suggests that China should comprehensively assess the current situation, capitalize on opportunities, confront challenges without fear, and strive for leadership in this domain.

Keywords: 3D IC; 3D stack integrated circuit; 3D system on chip; memory on logic; logic on logic

引用格式：陈昊, 谢业磊, 庞健, 等. 3D IC系统架构概述 [J]. 中兴通讯技术, 2024, 30(S1): 76-83. DOI: 10.12142/ZTETJ.2024S1011
Citation: CHEN H, XIE Y L, PANG J, et al. An overview of 3D IC system architecture [J]. ZTE technology journal, 2024, 30(S1): 76-83. DOI: 10.12142/ZTETJ.2024S1011

1985年，著名物理学家诺贝尔奖获得者理查德·费曼在日本发表“未来的计算机器”演讲时就预言，未来芯片发展的一大方向就是通过扩展平面芯片到三维层面来提升系统性能。2006年，三星电子CEO黄昌圭博士在国际电子器件年会（IDEM）上发表主题演讲“硅半导体业界新范式”时提出，电子技术新时代即将来临，可以将内存、逻辑、传感器、处理器等不同器件集合在一起的三维集成技术是电子技术新时代的核心。[1]

1 3D IC分类概览

2009年的国际半导体技术路线图（ITRS）从互联的不同层面对三维集成电路进行了规范。在板级的互联称三维封装，使用传统封装工艺，互联尺寸大，密度低，工艺简单，如封装上封装（PoP）、封装内封装（PiP）、多芯片模组（MCM）等。在封装级的互联称三维晶圆级封装，互联使用Bump-Pad和重布线层（RDL）工艺，如TSMC的CoWoS、InFO和Intel的EMIB等，又称封装内系统（SiP）。在片内的互联分为3种：三维芯片堆叠（3D SIC）为功能模块级堆叠，互联尺寸在10 μm左右，互联密度达一万个/mm^2；三维片上系统（3D SOC）为电路单元级堆叠，互联尺寸在1 μm左右，互联密度达一百万个/mm^2；三维集成电路（3D IC）为晶体管级三维堆叠，互联尺寸达100 nm级，互联密度高

达一亿个/mm^2，又称单片三维集成电路（Monolithic 3D IC）或顺序3D IC。从板级到封装级，再到片内高密度互联，3D IC工艺在不断成熟，三维互联也在微缩化。[2–3]

集成电路产业的发展与工艺技术演进密不可分。3D早期集中在板级和封装级互联，如PoP、PiP、SiP、2.5D等，使用打线或焊球进行片间互联，有源芯片间互联密度较低，互联速率、互联功耗、信号完整性等较差，需要芯片到芯片间互联模块（D2D IP）进行协议转换和驱动增强保证信号质量，而这会导致面积、功耗、延迟劣化。这些早期技术主要目的是突破工艺制造光罩尺寸限制，提升可拓展性，因此只能称为三维封装，而不是3D IC。现在3D IC已发展到3D SIC和3D SOC阶段，属芯片级或晶圆级互联，互联密度高，传输延迟低，可同步直连无需协议转换，晶体管间垂直互联对延迟、功耗开销小，与2D片内直联相容度高，弥合了晶圆制造和封装工艺间的鸿沟，让芯片互联走向新纪元。

2 3D IC带来价值

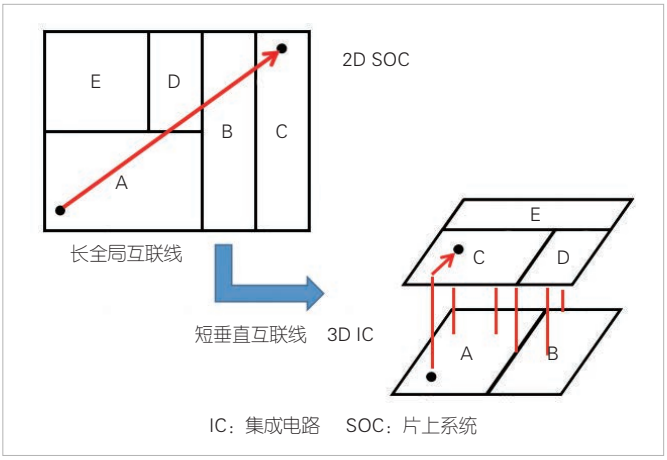
三维堆叠集成带来价值总体体现在4个层面：互联、内存、小型化和异构集成。

首先是互联层面。几十年来半导体先进工艺的发展带来了器件的微缩，可以在单位面积内集成更多器件，降低了器件延迟与功耗，但同时也带来了单位面积内绕线密度增加与栅极厚度减小的问题。芯片电路延迟由器件单元延迟和器件间互联绕线延迟两部分组成。绕线密度增加导致线宽线距降低，增加了互联RC延迟与互联功耗，器件互联线延迟与功耗占总延迟与总功耗的比例越来越高。另外，芯片功能复杂度提升增大了单芯片尺寸，故全局互联信号和时钟无法享受工艺微缩红利带来的延迟降低。此外，随着单位面积内器件集成度提高，更多互联绕线资源需求引起后道金属绕线密度提升，但受到制造工艺、应力、电源信号完整性等限制，后道金属密度和层数不能无限提高，绕线资源紧张导致拥塞与互联带宽的压缩。如图1所示，3D IC可通过模块切分并垂直堆叠，降低模块间全局互联长度，减少互联线延迟与互联信号和时钟网络上引入的功耗与面积，同时垂直相比水平互联拓宽了绕线资源，提升了模块间可容许的互联带宽。

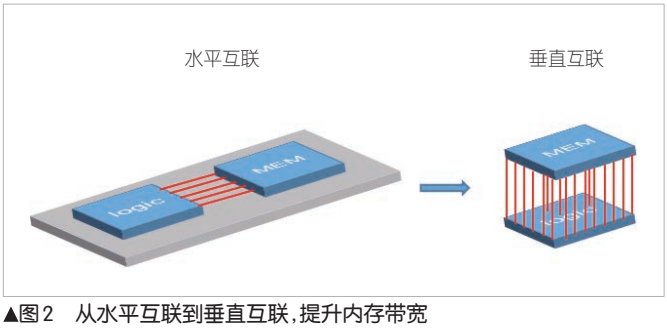
其次是内存层面。工艺技术、电路设计、系统架构的不断优化驱动处理器性能和主存容量成指数级提升。主存带宽每2年提升1.4～1.6倍，无法跟上处理器性能每2年提升约3倍的步伐，成为限制整体系统性能的短板，这就是内存墙[4]。在计算机体系架构中，使用静态随机存储器（SRAM）替代动态随机存储器（DRAM），采用高速缓存来构建多级内存层级，可以尽可能缓解内存墙影响。在先进工艺世代，SRAM微缩难度相比逻辑单元更大，其性能、功耗、面积等指标也滞后于算术逻辑单元（ALU）。另外，多级内存架构在物理上越靠近ALU的缓存，层级越低、容量越小、对带宽需求越高，多级缓存失效带来的延时惩罚就越严重。通过3D切分，将缓存、主存从原有的2D芯片互联中独立出来，与逻辑运算芯片垂直堆叠，物理上提升逻辑到内存互联带宽，打破2D场景下缓存容量限制，缓解内存墙问题。

再次是小型化。芯片制造工艺跟不上用户需求增长的步伐，为了满足更复杂功能实现与更高性能要求，在高性能计算领域使用多核多线程并行处理方式来构建超级计算机，在物联网领域集成计算和智能处理核心来构建边缘AI处理器。核数增多与功能集成导致面积扩增，会触及光罩极限，带来良率与成本劣化。2.5D水平互联可以解决单芯片切分问题，但无法满足带宽和封装尺寸要求。通过扩展Z维度，实现芯片垂直互联，可以提高集成密度，降低芯片外形尺寸，如图2所示。3D IC互联工艺还会对硅晶圆背面进行减薄，总体厚度相比于封装级三维堆叠显著降低。另外，利用堆叠前后的多步测试手段和使用芯片堆叠晶圆（D2W）或芯片堆叠芯片（D2D）工艺，可在提高单晶片良率的同时提升封装后芯片的良率和产率。

最后是异构集成。摩尔定律预测，每16～24个月，集



▲图1　从2D到3D，降低互联长度



▲图2　从水平互联到垂直互联，提升内存带宽

成电路上晶体管速度、集成度可以提升1倍。但随着器件尺寸不断接近物理极限，工艺制造时非理想效应凸显，这一步伐逐渐趋缓。不仅如此，数字电路、模拟电路放缓速率不同，模拟与内存电路晶体管速度在栅长22 nm时已经趋于饱和。数字电路尺寸微缩主要体现在垂直堆叠栅极构建多Fin结构。目前，N3节点鳍式场效应晶体管（FinFET）栅长约16 nm，继续降低通道长度会导致漏电流过大难以关断。综上可知，如果仍将数字、模拟、内存等使用同一种工艺制造在同一颗2D芯片上，将不利于整个系统的成本、面积和性能的优化[5]。使用不同工艺制造不同类型电路利用3D IC进行垂直互联，可以在满足性能的前提下更好地利用工艺甜点（Sweet Node），实现性能、成本和良率的最优化。不仅如此，在超越摩尔定律（Beyond CMOS）的范畴，一些新架构、新器件、新材料也可以利用3D IC进行异构集成。其中，三维图像传感器（3D CIS）就是典型的异构集成案例。此外，该技术还可以拓展到存内计算、射频、光互联等领域。

根据具体应用，3D IC价值主要面向两大类场景：一类是高端、行业旗舰级应用，一类是中低端、用户消费级应用。在高端应用方面，3D IC可带来高性能与高功能集成度。3D IC可以缩短连线长度，提升通信带宽，异构集成多种小芯片来构建封装内系统，可用于人工智能模型、大型科学计算、生命科学等领域。在中低端应用方面，3D IC可实现设备小型化，降低互联功耗，能够更好地利用工艺甜点降低总成本等。因此，3D IC在桌面级与手持应用、可穿戴式设备和万物互联等领域也有较高实用价值。

3 3D IC系统架构

3D IC系统架构，可按顶层（Top die）和底层（Bottom die）堆叠功能类型进行区分，主要有内存堆叠内存（MOM）、内存堆叠逻辑（MOL）、逻辑堆叠逻辑（LOL）。

MOM架构形式可将多颗DRAM堆叠成为内存芯粒（Memory Chiplet）的形式，提升系统内存容量和内存带宽，如高带宽内存（HBM）和混合内存立方体（HMC）；也可SRAM堆叠成为缓存堆叠缓存（Cache on Cache）并集成到SOC内，做成可拓展系统级缓存（SLC），提升缓存容量，如AMD的3D V-cache；还可设计新型3D SRAM，通过拆分SRAM单元阵列优化访问延迟，如宾夕法尼亚州立大学（PSU）的谢源教授研究团队用字线和位线拆分构建3D SRAM[6]。

MOL架构形式可将SRAM高速缓存（L1/L2 cache）和逻辑单元堆叠，做成逻辑上缓存形式，如将中央处理器（CPU）中的L1/L2缓存拆分和ALU堆叠。比利时微电子中心（IMEC）的DRAGOMIR等使用openSPARC-T1核进行3D缓存堆叠，相比2D可以提升11.4%的核频率[7]。佐治亚理工大学（GT）和ARM的ZHU等基于CMN600和N1核进行了SLC和CPU堆叠，可以提升17%的单周期指令数（IPC）[8]；也可将DRAM主存（Main Memory）和SOC堆叠，用来替代L2/L3缓存，做成近存计算（PNM）架构，例如：圣母大学与惠普实验室的CHEN等建立CACTI-3DD模型分析了3D DRAM的优势[9]，阿里达摩院的NIU等用逻辑和DRAM三维堆叠实现了一款高带宽高能效AI处理器[10]。

LOL架构形式有以下几种：1）SOC on IP堆叠，将SOC中的外围接口IP放于底层，接口IP多为模拟模块，对计算性能要求不高，在低工艺节点实现，采用三维堆叠异质集成，降低尺寸与综合成本。2）Core on NOC堆叠（NOC指片上路由网络），把计算核心与NOC进行3D拆分堆叠，降低计算核心间网络路由物理距离。该方法对逻辑切分挑战较大，设计较复杂。3）Core on Core堆叠，将不同计算或处理核心相互堆叠。但Core一般是发热大户，可能导致散热方面问题。IMEC的CHEN等分析了3D CPU堆叠的散热相比2D结温会提高140%[11]。4）三维片上网络（3D NOC）的形式，这是一种新型片上互联方式。相比于传统的2D NOC，3D NOC可以显著降低平均网络访问延迟和时钟周期数，常见的有基于Tree结构和基于Mesh结构两种。ARM的FEERO和PANDE分析了3D NOC相比于2D NOC在带宽、功耗、吞吐量方面都有较大优势，结合NOC交换延迟的降低可以更好地提升性能[12]。5）处理器内部流水线级3D堆叠，3D NOC和流水线堆叠需要芯片互联架构层面做出适配性改动，对硅通孔（TSV）和芯片键合互联工艺微缩也提出较大挑战。

3D IC系统架构选择主要受需求侧和实现侧的影响。需求侧是动力和源泉，影响架构的可能方向；实现侧是限制与约束，影响架构的可行方向。两者的集中体现是3D IC系统设计。

4 3D IC系统设计

自2000年以来，美国国防部高级研究计划局先进微电子研究委员会（DARPA）就开始资助3D IC和异构集成相关项目。2000年，麻省理工学院（MIT）的RAIF教授等利用Rent定律理论论证使用3D IC可以显著降低互联延迟和芯片面积，但实现的最关键因素是垂直层间过孔的高密度互联，多层堆叠的性能瓶颈主要在于过孔的禁空区（KOZ）大小[13]。这为3D互联工艺微缩指明了方向。

3D IC研究早期由于工艺技术难以跟上架构发展，研究集中在架构的可行性论证和收益方面。2004年，Intel的BLACK等对深度流水线处理器——iA32进行了3D切分架构的尝试，通过三维布局规划，可以节约处理器流水线中的RC延迟，如时钟时延、浮点处理时延、寄存器堆访问时延等，可以降低约25%的流水线级数。综合来说，处理器级的3D实现可以提升15%性能的同时优化约15%的功耗[14]。

2006年，佐治亚理工学院的PUTTASWAMY等使用3D方式实现了256输入的物理寄存器堆，两层堆叠可以在优化58.5%的能耗同时达成24.1%的延迟优化，四层堆叠可以在优化58.2%的能耗同时达成36%的延迟优化。此外，他们还提出了寄存器堆的3种3D切分方案：寄存器切分、比特位切分、访问端口切分。3D切分堆叠带来的优势不局限于处理器微架构，对于不同的处理器配置，处理器中的关键延迟部件不同。不同的微架构需要调整3D堆叠策略，以达成最佳的时延降低[15]。寄存器堆是线延迟主导，因此借助3D实现来降低访问延迟和简化控制数据流来构建高性能微处理器将成为可能。

2009年，MIT开发的3D工艺逐渐可以支持处理器微架构在硅上进行原型实现。宾夕法尼亚州立大学谢源课题组使用MIT Lincoln Labs的180 nm 3D FDSOI工艺制造了两种基本计算单元：3D加法器和3D乘法器。相比于2D实现，Kogge-Stone型3D加法器通过将12 bit计算带宽提升为72 bit，可以降低10.6%～34.3%的延迟和11.0%～46.1%的能耗，32×32的Wallace-Tree型3D乘法器可以降低14.4%延迟和6.8%能耗[16]。2010年，他们发布了一个用于H.264编解码的3D流处理器，该处理器可以提供多个内存通道，内存控制器和并行访问策略经过了重新设计，来充分利用3D DRAM的内存带宽提升[17]。这表明，在设计3D计算机系统架构时，应考虑如何将系统架构设计和3D堆叠集成带来的好处结合起来，例如：可以重新设计缓存层级和片上互联方案来充分利用3D带来的优势。

2009年，北卡罗莱纳州立大学（NSCU）FRANZON研究组报告了一种用于合成孔径雷达（SAR）的快速傅里叶变换（FFT）处理器。SAR FFT处理器架构有大量全共享全互联内存和计算单元，通过将一个大内存拆分成很多小的内存并实现内存逻辑垂直堆叠，可以提高计算单元访问内存的并行度，并降低FFT处理器60.3%的内存能耗。相比于2D实现，SAR FFT处理器可降低53%平均线长，提升24.6%计算频率，使内存带宽提升到原来的8.55倍，总硅面积降低25.3%[18]。

后来，人们看到了3D工艺制造上的局限性，开始在系统层面探索多核片上系统和3D IC之间的相容性。2012年，佐治亚理工学院SUNG课题组设计了一种基于3D堆叠内存的多核并行处理器架构——3D MAPS。这种架构的顶层是64个通用计算核心，底层是对应每个核心的4kB SRAM共256 kB。晶圆制造使用了格芯（GF）的130 nm技术，3D封装使用了安靠（Amkor）的Tezzaron TSV工艺。在3D MAPS单核设计方面，设计者详细考虑了流水线深度、寄存器堆容量、发射带宽、计算单元、指令集等的设计，并充分优化了内核微架构来满足3D IC技术提供的大内存带宽的优势。在多核间互联和核与内存互联方面，他们对核间路由的同步与协调和内存分块做了精细调整，让访存容量和核心处理能力相匹配，并可以让单个核心分别控制4个内存分区。除此之外，3D MAPS还引入了可测试性设计，并定制扫描链和测试控制器进行功能和性能基准测试。基准测试内容包括最大内存带宽、周期指令数和功耗等。基于Median Filter标准，3D MAPS可利用的内存带宽为63.8 GB/s，可达理论最大值的89.99%，这证明了微架构调整和3D IC结构之间的适配度[19-20]。

2021年，法国原子能委员会电子与信息技术实验室（CEA-Leti）与格勒诺布尔大学共同发布了一款基于有源中介层的6片芯粒组成的96核3D堆叠处理器——IntAct。有源中介层不同于2.5D情况下的中介层，除提供硅桥互联和电容电感等被动元件外，IntAct的有源中介层中还集成了电源管理模块SCVR、分布式3D插入式路由、传感器、串行解串器等模拟IP，还有用于可测试的设计与端口等。IntAct使用意法半导体（ST）的FDSOI 28 nm工艺堆叠在65 nm工艺的有源中介层上，可以达成分布式互联、3D异构集成、电源管理动态调压调频、芯粒重用、成本优化等诸多目标[21]。

2022年，苏黎世联邦理工（ETH）LUCA课题组提出了一种开源众核SOC架构——MemPool，其中高达256可编程核心集群可以共享大容量L1缓存，可以满足低延迟、低功耗和高吞吐量要求。基于MemPool，佐治亚理工学院和IMEC的研究人员结合3D IC技术，提出增强型MemPool-3D，有效将MemPool的体系架构设计和3D IC的优势结合起来，解决了常规MemPool在2D IC场景下的路由拥塞和全局传播延迟等问题[22]。3D IC系统设计汇总如表1所示。

可以看到，3D IC系统设计主要由架构设计、功能切分、工艺实现3个方面来决定。

架构设计方面，传统2D IC情况下仅有单个Die，通过2D的前道工艺（FEOL）和后道工艺（BEOL）来实现。2D的系统架构参数，如核数、主频、流水级别、缓存级别、缓存容量、访存带宽等结合2D IC的物理实现工艺做了最优

▼表1 3D IC系统设计汇总

	DAC–2009（NCSU）	3D IC–2009（PSU）	3D IC–2010（PSU）	ISSCC–2012（GT）	JSSC–2021（CEA–Leti）	DATE–2022（ETH,IMEC）
应用	3D SAR FFT pro - cessor	3D微架构–加法器，乘法器	3D DRAM H.264流处理器	3D MAPS–并行计算	IntAct–3D SOC	3D MemPool–多核并行计算
芯片工艺	MIT LL 180 nm 3DFDSOI	MIT LL 180 nm 3DFDSOI	GF 130 nm Tezzaron 3D	GF 130 nm Tezzaron 3D	ST 28 nm 65 nm Active inter - poser	28 nm(prototype)
3D拆分架构	3 Layers–MOL	3 Layers–LOL	5 Layers–LOL/MOM/MOL	3 Layers–MOM/MOL	2 Layers–LOL	2 Layers–MOL
3D pitch	3.9 μm	2.65 μm	4 μm	5 μm/2.5 μm	20 μm/40 μm	1 μm
3D堆叠架构	F2F/F2B	F2F/F2B	F2F/F2B	F2F/F2B	F2F	F2F
优势	53% 线长降低 24.6% 频率提升 8.55倍内存带宽提升 25.3% Si面积降低	10.6% ~ 34.3% 延迟降低 11.0% ~ 46.1% 能耗降低	并行存储访问 支持8个独立内存通道	64个通用计算核 比2D成本降低3% 63.8 GB/s内存带宽	集成片上电源调制器 可拓展式3D+2.5D集成 算力220 Gops超过 7 nm同类芯片	256核共享L1缓存 9.1% 性能提升 15% 能耗降低

DAC：国际设计自动化会议
DATE：欧洲设计自动化与测试学术会议
DRAM：动态随机存储器
F2B：面对背堆叠
F2F：面对面堆叠
FDSOI：全耗尽绝缘体上硅工艺
FFT：快速傅里叶变换
GF：格芯
IC：集成电路
ISSCC：国际固态电路会议
JSSC：IEEE固态电路杂志
LOL：逻辑堆叠逻辑
MOL：内存堆叠逻辑
MOM：内存堆叠内存
SAR：合成孔径雷达
SOC：片上系统

化。过渡到3D情形，晶体管可以实现垂直堆叠并互联，增加了物理实现自由度。如果仍选择之前2D IC对应的系统参数与指标，则难以达成3D IC情况下的最优结果。因此，针对3D IC，要面向3D IC的工艺特点和价值取向，进行3D IC系统架构重构，充分利用3D IC优势。架构重构取决于具体产品的架构形式、做3D IC的目的、解决的关键问题，以及要突破的产品瓶颈。如果关注SOC多核性能，如延迟、吞吐量、核负载均匀性等，就需要考虑多核访存容量、带宽、路由性能等瓶颈，拓展缓存容量和带宽，进行NOC架构级修改或NOC节点配置修改。如果要提升具体IP单核性能，则要考虑IP核内限制性能提升的因素，如寄存器堆配置、流水线深度、L1/L2带宽、物理可实现性等。这些都可以利用3D IC来改善，具体需要结合架构与微架构修改和底层代码实现。如果要进行内存容量或带宽拓展，或将片外缓存拿到片内，或降低内存层次，实现近存架构，那么可采取MOM或MOL，并结合架构和代码具体情况将内存从2D SOC中切分出来，或将扩展内存挂载到原有SOC上。总体来说，考虑延迟敏感性，对互联带宽、内存容量有要求的场景，可以通过功能切分，在顶层和底层上实现垂直互联[23]。

功能切分与工艺可实现性需要同步考虑。哪些模块需要放到哪颗Die上？具体放到Die上的什么位置？解决这些问题的方法称为3D IC的功能切分。能够实现功能切分的层级和3D IC工艺微缩程度息息相关。其中，最重要的工艺指标是Die间互联（主要是TSV和混合键合）尺寸、间距、电气特性，这决定了Die间互联的带宽、速率、功耗。尺寸间距越小，电气性能越好，3D互联带宽、速率越高，功耗就越低。相对来说，只要打破带宽和速率的限制，3D IC架构实现就更灵活，可以满足的架构方式就更多样化。从2D SOC到3D SOC，不同的3D IC方案按功能切分层级进行分类。切分和堆叠的最小单元在芯片设计中的层次高低被定义为3D IC的切分粒度。基于SoC架构级别的切分是粗粒度的切分，互联密度低，工艺要求低，对应MOL主存堆叠逻辑和LOL中的Core on Core或SOC on IP形式。细一级的是基于IP功能模块的切分，对应MOM、MOL缓存堆叠逻辑和LOL中Core on NOC或3D NOC形式；再细一级的是功能模块内部逻辑级的切分，对应LOL中流水线级别拆分。最细粒度的拆分是门级的切分，通过LOL在总线位宽上拆分或网表内部基本单元级的拆分，需要依赖于超高密度互联的Monolithic 3D IC工艺[24]。可以看到，切分每细化一个级别，就会对3D工艺微缩提出更高要求，带宽可以做得更高，因此可以容许架构设计做出更大的改动。

我们看到，从2D到2.5D、3D SIC，再到3D SOC、Monolithic 3D IC，3D工艺的演进是架构革新的助推器，现阶段业内3D工艺普遍可以支持3D架构的第一、第二级别——架构级和功能级。也就是说，是3D互联工艺限制了Die间互联的性能和带宽，同时也限制了具体3D IC实现时的系统架构与切分方案的自由度。反之，更精细化的3D架构设计与切分方案也对3D工艺演进提出了更多的要求。随着3D

互联工艺的演进，3D架构设计与切分方案也可以更加多样化。

除此之外，3D IC系统设计也需要综合考量价值与开销，使整个芯片系统性能、功耗、面积、成本（PPAC）达到最优。

3D IC中最显著的问题是散热问题。芯片的三维堆叠带来功耗密度的多层叠加，但3D IC相比2D却没有引入额外的散热通路，与2D相比，3D的结温会更高，需要更全面的产热、散热系统解决方案。热源端需要在系统设计和逻辑实现时考虑热因素，采用低功耗的设计方法（如多电压域、时钟电源关断等）降低系统积热，需要在功能切分和物理实现时使用热感知的布局技术，并在设计签核时考虑热带来的时序和电源恶化的影响。散热端需要考虑功能切分与物理设计时尽可能使高发热和热敏感模块离散热器更近，添加额外的散热通路如Dummy die和Dummy TSV等辅助散热，还可以结合一些先进散热技术如微通道液冷、金刚石散热等作为芯片级冷却解决方案。

除了散热之外，3D IC的工艺成熟度也备受关注。其中，芯片键合与硅通孔是3D IC最重要的两项技术。二者作为片间垂直互连的实际存在形式，从特征尺寸、电气特性、工艺良率3个方面决定了3D IC性能与质量。Bonding和TSV特征尺寸越小，互联密度越高，3D IC可以提供的片间带宽就越大。选取电气性能优异的互联材料，可以降低3D IC互联延迟与功耗。持续优化工艺细节和工艺流程，如改善晶圆键合时的平整度和对准问题，控制键合焊盘和工艺温度，改善铜凸问题等，对于提高3D IC整体良率、降低3D IC制造成本有重要意义。

此外，还需要考虑3D IC的成本问题。3D IC涉及多颗芯片堆叠，在成本核算方面更加复杂，包括设计成本、制造成本、封装成本、测试成本，以及不同供应商的物流成本。3D IC设计复杂度比2D更高，需要考虑的设计维度更多，但可以通过将高性能模块做成多种或多代产品可复用的3D Chiplet形式，以降低设计和制造成本。在制造层面，3D可以堆叠集成多种工艺，降低芯片面积，提升良率，但同时也需要加工更多光罩（Mask layer），提高一次性工程成本（NRE）。3D封装工艺对产线控制要求更高，良率曲线还在爬升，当前阶段比2D封装成本有显著提升。3D需要嵌入更多的测试与修复设计，尽可能保证筛片良率。此外，采用新的测试针卡、机台、测试用例等也会提高3D测试成本。

除前述之外，多物理场耦合、物理设计实现、多工艺签核、可测试性设计与测试流程等问题对3D IC整体可实现性、性能、成本折中也有重要影响。3D IC芯片系统设计应综合考虑功能、性能、可靠性、成本等多种因素，确定最终系统的参数选项，根据具体设计情况，迭代达成最佳实现目标。

5 3D IC产品概览

最近几年，产业界已推出较为成熟的3D IC产品。最早应用3D IC技术的产品是3D CIS和MOM的HBM。但由于3D CIS架构相对简单、应用领域单一，同理HBM不能脱离逻辑芯片独立功能存在。

通用3D IC产品在2019年产生。Intel发布了业内第一款商用3D IC的低功耗小尺寸SOC芯片——Lakefield处理器。这款芯片使用逻辑堆叠逻辑SOC on IP异质堆叠形式，顶层使用Intel 10 nm工艺的计算芯粒（Chiplet），包含CPU、GPU、图像处理器（IPU）和一些显示引擎。底层使用Intel 22 nm低功耗工艺的IO芯粒，包含通用串行总线（USB）、串行外设接口（SPI）等一些常用接口IP。顶层和底层之间使用Intel 3D Foveros Microbump互联工艺。3D同步互联速率可达500 MHz，3D互联功耗为0.2 pj/bit。Lakefield凭借低待机功耗和小尺寸，成为当时最小的桌面级CPU处理器，应用于Samsung Galaxy Book S、Lenovo X1 Fold等超薄笔记本中[25]。

2022年，Intel发布号称芯片航空母舰的Ponto Vecchio（PVC），这款芯片从SOC架构设计到实现工艺都代表Intel当前最先进的芯片水平。PVC结合了3D IC和2.5D技术优势，共有63个芯粒，其中47个是功能芯粒，另外16个芯粒用于支撑与散热，共由超过1 000亿个晶体管构成。PVC有两层堆叠，采用了混合SOC on IP和MOL 3D架构，顶层有16个计算芯粒和8个内存芯粒，内含128个Xe核和120 MB扩展L3缓存，底层有2个大的基础互联芯粒，内含片上集成电源模块（FIVR）、内存控制器、PCIe、CXL、L2缓存等。3D异步互联速率达2.97 GHz，互联功耗0.2 pj/bit。PVC是同时采用多家供应商不同工艺芯粒进行3D IC堆叠的第一款大规模商用产品，其计算芯粒为TSMC N5，内存芯粒与基础芯粒为Intel N7。这表明3D IC可在紧耦合互联层面弥合多家供应商界限，打造全新商业模式。PVC应用在阿贡实验室定制的面向人工智能与科学计算的超算服务器，预期算力可达2 ExaFlops[26]。

AMD的3D IC商用化脚步紧随Intel之后。2022年，AMD发布第一款使用3D V-cache技术的产品，底层是计算芯粒，内含8个CPU核和32 MB的L3缓存，顶层是内存芯粒，内含64 MB L3缓存。3D V-cache采用TSMC最新3D SoIC技术，利用间距9 μm的高密度TSV与混合键合（HB）进行互联。3D V-cache架构本质属于Cache on Cache的

MOM，其中3D堆叠互联区域只有L3缓存。3D V-cache目前有两代，第一代的内存芯粒与计算芯粒都是TSMC N7，第二代把计算芯粒升级为N5，复用之前N7的内存芯粒堆叠。3D V-cache同步登录AMD服务器与桌面处理器产品线。相较于2D版本，3D版本服务器性能可提升66%[27]。

2023年，AMD发布了新一代3D IC处理器架构——Instinct MI300。MI300的3D IC架构与PVC类似，也是两层堆叠SOC on IP和MOL混合3D架构。顶层是TSMC N5工艺计算芯粒，针对不同的产品线有两种分型：MI300A由3个Zen4架构CPU芯粒和6个CDNA3架构GPU芯粒组成混合架构；MI300X把3个CPU芯粒替换成2个GPU芯粒，共8颗GPU芯粒组成纯GPU架构。底层是4个TSMC N6工艺的基础芯粒，内含集成输入和输出（I/O）接口、路由仲裁与拓展缓存等模块。MI300主要面向大语言模型与高性能计算，用于劳伦斯-利弗莫尔国家安全实验室El Capitan超级计算机，算力可达2 ExaFlops[28]。

2022年，人工智能芯片公司Graphcore推出创新架构智能处理器——Bow IPU，使用TSMC 3D SoIC晶圆堆叠晶圆（WOW）方式，顶层是人工智能逻辑芯粒，底层是集成了深硅刻蚀电容（DTC）的电源管理芯粒，可以改善电源完整性并获取更高能耗效率[29]。

Meta公司在2024年国际固态电路会议（ISSCC 2024）上报告了一款用于穿戴式虚拟现实应用的3D IC原型芯片，基于TSMC SoIC W2W F2F堆叠，顶层是扩展SRAM与传感器模块，底层是芯片启动与控制系统。通过三维堆叠，这一产品至少可节约55%的内存访问功耗，提升40%的系统性能，超小的体积也为边缘式AI应用拓展了兼容性。这是3D IC在消费级应用方面的重大创新[30]。

在2022年ISSCC上，阿里达摩院发布了一款完全自主制造的基于3D IC的近存计算人工智能样片。该产品采用了3D WOW形式HB堆叠，顶层是25 nm工艺的DRAM芯粒，共36 Gbit DRAM，底层是55 nm工艺的逻辑芯粒，由神经网络引擎（NE）、匹配引擎（ME）等组成。虽然仅采用了远低于业界先进的工艺节点，但是3D跨Die互联速率可达150 MHz，3D互联功耗仅为0.88 pj/bit，并且吞吐量、带宽、能耗效率等相比其他采用更先进工艺的2D和2.5D芯片也有较大优势[10]。

6 机遇与挑战

可以看到，最近几年，3D IC芯片如井喷般涌现，根本原因是摩尔定律推进困难，先进工艺红利不再，存储墙问题凸显。伴随着3D制造工艺尤其是TSV和HB工艺的逐渐成熟，物理尺寸可与BEOL尺寸相比。此外，高性能计算、人工智能、大模型、数字孪生等场景应用对芯片大算力、大内存、大带宽的需求进一步加大。目前看来，3D IC的主要优势在于通过互联和内存性能提升产品高度，通过功耗与面积优化拓宽产品广度。随着工艺优化和产品迭代带来的成本与风险降低，3D IC也必将步入大规模商用化的道路。我们应该把握机遇，勇于做科技的领跑者。

致谢

感谢深圳市中兴微电子技术有限公司高级工程师武辰飞、李乐琪、黄彤彤、高静丽对本研究的帮助！

参考文献

[1] GARROU P, BROWN C, RAMM P. Handbook of 3D integration, volume 1: technology and applications of 3D integrated circuits [M]. New York: John Wiley & Sons, 2011
[2] LAU J H. Chiplet design and heterogeneous integration packaging [M]. 2023
[3] LI Y, GOYAL D. 3D microelectronic packaging: from architectures to applications [M]. Second edition. Singapore: Springer, 2021
[4] GHOLAMI A, YAO Z, KIM S, et al. AI and memory wall [EB/OL]. (2024-03-21)[2024-08-08]. https://arxiv.org/pdf/2403.14123
[5] NAUTA B. 1.2 racing down the slopes of Moore's law [C]// Proceedings of IEEE International Solid-State Circuits Conference (ISSCC). IEEE, 2024: 16-23. DOI: 10.1109/ISSCC49657.2024.10454417
[6] TSAI Y F, XIE Y, VIJAYKRISHNAN N, et al. Three-dimensional cache design exploration using 3DCacti [C]//Proceedings of International Conference on Computer Design. IEEE, 2005: 519-524. DOI: 10.1109/ICCD.2005.108
[7] NAEIM M, YANG H Q, CHEN P H, et al. Design enablement of 3-dies stacked 3D-ICs using fine-pitch hybrid-bonding and TSVs [C]//Proceedings of IEEE International 3D Systems Integration Conference (3DIC). IEEE, 2023: 1-4. DOI: 10.1109/3DIC57175.2023.10155075
[8] ZHU L J, TA T, LIU R, et al. Power delivery and thermal-aware arm-based multi-tier 3D architecture [C]//Proceedings of IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED). IEEE, 2021: 1-6. DOI: 10.1109/ISLPED52811.2021.9502481
[9] CHEN K, LI S, MURALIMANOHAR N, et al. CACTI-3DD: architecture-level modeling for 3D die-stacked DRAM main memory [C]//Proceedings of Design, Automation & Test in Europe Conference & Exhibition (DATE). IEEE, 2012: 33-38
[10] NIU D M, LI S C, WANG Y H, et al. 184QPS/W 64Mb/mm^{2}3D logic-to-DRAM hybrid bonding with process-near-memory engine for recommendation system [C]//Proceedings of IEEE International Solid-State Circuits Conference (ISSCC). IEEE, 2022: 1-3. DOI: 10.1109/ISSCC42614.2022.9731694
[11] CHEN R, LOFRANO M, MIRABELLI G, et al. Power, performance, area and thermal analysis of 2D and 3D ICs at A14 node designed with back-side power delivery network [C]// Proceedings of International Electron Devices Meeting (IEDM). IEEE, 2022: 23.4.1-23.4.4. DOI: 10.1109/IEDM45625.2022.10019349
[12] FEERO B S, PANDE P P. Networks-on-chip in a three-dimensional environment: a performance evaluation [J]. IEEE

transactions on computers, 2009, 58(1), 32–45
[13] RAHMAN A, REIF R. System–level performance evaluation of three–dimensional integrated circuits [J]. IEEE transactions on very large scale integration (VLSI) systems, 2000, 8(6): 671–678. DOI: 10.1109/92.902261
[14] BLACK B, NELSON D W, WEBB C, et al. 3D processing technology and its impact on iA32 microprocessors [C]// Proceedings of IEEE International Conference on Computer Design: VLSI in Computers and Processors, 2004. ICCD 2004. Proceedings. IEEE, 2004: 316–318. DOI: 10.1109/ICCD.2004.1347939
[15] PUTTASWAMY K, LOH G H. Implementing register files for high–performance microprocessors in a die–stacked (3D) technology [C]//Proceedings of IEEE Computer Society Annual Symposium on Emerging VLSI Technologies and Architectures (ISVLSI'06). IEEE, 2006: 6. DOI: 10.1109/ISVLSI.2006.56
[16] OUYANG J, SUN G, CHEN Y, et al. Arithmetic unit design using 180nm TSV–based 3D stacking technology [C]//Proceedings of IEEE International Conference on 3D System Integration. IEEE, 2009: 1–4. DOI: 10.1109/3DIC.2009.5306565
[17] ZHANG T, WANG K, FENG Y, et al. A 3D SoC design for H.264 application with on–chip DRAM stacking [C]//Proceedings of IEEE International 3D Systems Integration Conference (3DIC). IEEE, 2010: 1–6. DOI: 10.1109/3DIC.2010.5751446
[18] THOROLFSSON T, GONSALVES K, FRANZON P D. Design automation for a 3DIC FFT processor for synthetic aperture radar: a case study [C]//Proceedings of 46th ACM/IEEE Design Automation Conference. IEEE, 2009: 51–56
[19] KIM D H, ATHIKULWONGSE K, HEALY M, et al. 3D–MAPS: 3D massively parallel processor with stacked memory [C]// Proceedings of IEEE International Solid–State Circuits Conference. IEEE, 2012: 188–190. DOI: 10.1109/ISSCC.2012.6176969
[20] KIM D H, ATHIKULWONGSE K, HEALY M B, et al. Design and analysis of 3D–MAPS (3D massively parallel processor with stacked memory) [J]. IEEE transactions on computers, 2015, 64 (1): 112–125. DOI: 10.1109/TC.2013.192
[21] VIVET P, GUTHMULLER E, THONNART Y, et al. IntAct: a 96–core processor with six chiplets 3D–stacked on an active interposer with distributed interconnects and integrated power management [J]. IEEE journal of solid–state circuits, 2021, 56(1): 79–97. DOI: 10.1109/JSSC.2020.3036341
[22] CAVALCANTE M, AGNESINA A, RIEDEL S, et al. MemPool–3D: boosting performance and efficiency of shared–L1 memory many–core clusters with 3D integration [C]//Proceedings of Design, Automation & Test in Europe Conference & Exhibition (DATE). IEEE, 2022: 394–399
[23] LOH G H, XIE Y, BLACK B. Processor design in 3D die–stacking technologies [J]. IEEE micro, 2007, 27(3): 31–48. DOI: 10.1109/MM.2007.59
[24] Batude P, Ernst T, Arcamone J, et al. 3–D sequential integration: a key enabling technology for heterogeneous co–integration of new function with CMOS [J]. IEEE journal on emerging and selected topics in circuits and systems, 2012, 2(4): 714–722
[25] GOMES W, KHUSHU S, INGERLY D B, et al. 8.1 lakefield and mobility compute: a 3D stacked 10 nm and 22FFL hybrid processor system in 12 × 12 mm^2, 1 mm package–on–package [C]//Proceedings of IEEE International Solid–State Circuits Conference – (ISSCC). IEEE, 2020: 144–146. DOI: 10.1109/ISSCC19947.2020.9062957
[26] GOMES W, KOKER A, STOVER P, et al. Ponte vecchio: a multi–tile 3D stacked processor for exascale computing [C]// Proceedings of IEEE International Solid–State Circuits Conference (ISSCC). IEEE, 2022: 42–44. DOI: 10.1109/ISSCC42614.2022.9731673
[27] BURD T, LI W, PISTOLE J, et al. Zen3: the AMD 2nd–generation 7nm x86–64 microprocessor core [C]//Proceedings of IEEE International Solid–State Circuits Conference (ISSCC). IEEE, 2022: 1–3. DOI: 10.1109/ISSCC42614.2022.9731678
[28] SMITH A, LOH G H, SCHULTE M J, et al. Realizing the AMD exascale heterogeneous processor vision: industry product [C]// Proceedings of ACM/IEEE 51st Annual International Symposium on Computer Architecture (ISCA). IEEE, 2024: 876–889. DOI: 10.1109/ISCA59077.2024.00068
[29] MOORE S. Graphcore uses TSMC 3D chip tech to speed AI by 40% [EB/OL]. (2022–03–03) [2024–08–08]. https://spectrum.ieee.org/graphcore–ai–processor
[30] WU T F, LIU H C, SUMBUL H E, et al. 11.2 A 3D integrated Prototype System–on–Chip for Augmented Reality Applications Using Face–to–Face Wafer Bonded 7 nm Logic at <2 μm Pitch with up to 40% Energy Reduction at Iso–Area Footprint [C]// Proceedings of IEEE International Solid–State Circuits Conference (ISSCC). IEEE, 2024: 210–212. DOI: 10.1109/ISSCC49657.2024.10454529

作 者 简 介

陈昊，移动网络和移动多媒体技术国家重点实验室研究员、深圳市中兴微电子技术有限公司后端设计部后端设计工程师；主要研究方向为三维集成电路物理实现，负责先进技术规划和研发工作。

谢业磊，移动网络和移动多媒体技术国家重点实验室研究员、深圳市中兴微电子技术有限公司封测工程部封装设计专家；拥有9年以上封装设计研究经验，负责多个先进封装预研项目的开发工作。

庞健，移动网络和移动多媒体技术国家重点实验室研究员、深圳市中兴微电子技术有限公司先进封装技术总工；负责封装技术规划和研发，完成多款先进封装交付。

欧阳可青，深圳市中兴微电子技术有限公司副总经理、IC平台研发中心主任，并担任射频异质异构集成全国重点实验室副主任、移动网络和移动多媒体技术国家重点实验室研究中心主任；长期从事复杂SOC芯片的设计方法学研究工作，在先进工艺、数模混合设计、2.5D/3D先进封装、高可靠性设计等领域取得多项关键技术突破。

XR网业协同技术

Network and Service Collaboration Technology Based on XR

李娜/LI Na[1,2]，张诗壮/ZHANG Shizhuang[1,2]，
程义超/CHENG Yichao[1,2]

（1. 移动网络和移动多媒体技术国家重点实验室，中国 深圳 518055；
2. 中兴通讯股份有限公司，中国 深圳 518057）
(1. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China;
2. ZTE Corporation, Shenzhen 518057, China)

DOI：10.12142/ZTETJ.2024S1012
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.tn.20240724.1107.008.html
网络出版日期：2024-07-24
收稿日期：2023-11-20

摘要：分析扩展现实（XR）业务在业务网络双向感知、业务保障和业务评估方面面临的挑战，根据XR业务特征和网络要求，研究并提出面向XR业务的网业协同技术，构建基于XR业务的低时延、大带宽和高可靠的广义确定性网络。

关键词：扩展现实；业务评估；业务保障；感知；网业协同；元宇宙

Abstract: The challenges of bidirectional perception of the network and service, service assurance and service evaluation for extended reality (XR) services are analyzed. Based on the characteristics of XR services and network requirements, the network service coordination technologies for XR services are proposed, and the XR service-based Detnet with low latency, large bandwidth and high reliability are constructed.

Keywords: XR; service evaluation; service assurance; perception; network service coordination; Metaverse

引用格式：李娜, 张诗壮, 程义超. XR网业协同技术 [J]. 中兴通讯技术, 2024, 30(S1): 84-90. DOI: 10.12142/ZTETJ.2024S1012
Citation: LI N, ZHANG S Z, CHENG Y C. Network and service collaboration technology based on XR [J]. ZTE technology journal, 2024, 30(S1): 84-90. DOI: 10.12142/ZTETJ.2024S1012

1 XR业务概述

扩展现实（XR）是不同类型现实的总称，指的是计算机技术和可穿戴设备产生的真实和虚拟融合以及可人机交互的环境，包含虚拟现实（VR）、增强现实（AR）、混合现实（MR）以及其他沉浸式技术。

VR是以渲染的视觉和听觉为主导，通过计算机模拟虚拟环境给人以环境沉浸感，可让人完全沉浸在虚拟环境中，通常需要佩戴视听设备。AR是实时根据现实世界的位置和角度，并叠加相应的虚拟图像和三维技术，即在真实空间叠加虚拟物体，把虚拟信息映射在现实环境中，但不能与真实环境交互。MR指增强型的AR，是虚拟与现实的混合体，可将现实世界数字化，并与虚拟世界融合产生新世界，虚拟物体和现实世界的对象在新世界共存并实时交互[2]。

MR的终极形态是元宇宙（Metaverse），元宇宙是下一代沉浸式互联网，是超越虚拟与现实的终极愿景，意在创造一个平行于现实世界的人造虚拟空间，承载用户社交娱乐、创作展示、经济交易等一切活动。因其高沉浸感和完全的同步性，逐步与现实世界融合、互相延伸拓展，最终达成“超越”虚拟与现实的“元宇宙”，为人类社会拓宽无限的生活空间。

XR在各种领域都有良好的发展前景，XR可穿戴设备作为下一代沉浸式个人智能计算终端，将重构“人-物-场”的连接关系，创造新的生态入口。表1列举了XR技术在不同领域、不同行业的部分应用场景。政策上，XR同样被寄予厚望，“十四五”规划将“虚拟现实和增强现实”产业列为数字经济发展的七大重点产业之一。中国信息通信研究院

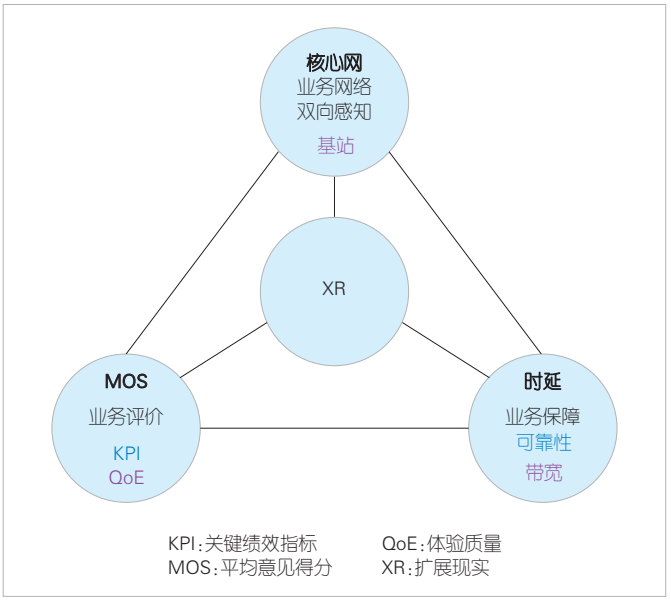
▼表1 XR应用场景分类

应用领域	应用场景	场景方向
娱乐领域	云XR游戏、VR直播、VR电影	ToB、ToC
教育领域	虚拟实验，模拟训练	ToB、ToC
医学领域	远程医疗、VR医学成像、康复训练	ToB、ToC
文旅领域	VR博物馆、AR导览系统、春晚舞台	ToB、ToC
电商领域	MR试装、虚拟商店、商品展示	ToB、ToC
电视转播领域	AR体育赛事转播、虚拟角色	ToC
军事领域	VR模拟战斗、作战协同、心理适应性训练	ToB
建筑领域	MR建筑图、设计展示与沟通	ToB、ToC
生活领域	AR测距、VR社交	ToC
航天领域	VR虚拟训练、模拟维修、科普教育	ToB
工业领域	VR工程解决方案设计	ToB

AR：增强现实　ToB：面向企业　VR：虚拟现实
MR：混合现实　ToC：面向个人　XR：扩展现实

发布的《虚拟（增强）现实白皮书》预测[3]：2022—2025年AR/VR设备总量增长较为缓慢，2026年后随终端成熟度提升，数量开始快速增长；AR终端前期增长较快，预计2024年设备数量超过VR终端；2030年蜂窝网络激活率约为60%，蜂窝联网AR/VR设备规模达到8 130万以上；预计2040年蜂窝网络激活率达到95%，蜂窝联网AR/VR设备规模达到6.57亿，普及率约为49.6%。

XR涉及终端、应用、网络、平台等，其中网络侧为平台侧和终端侧构建低时延、大带宽和高可靠的传输通道，XR的网络架构如图1所示[1]。高清XR业务的分辨率达24K，全景传输带宽要求达8 Gbit/s，基于视场角(FoV)的传输带宽要求达3 Gbit/s，时延如动作到画面时延(MTP)要求在10ms以内，可靠性如分组丢失率控制在10^{-7}以内[2]，这些都对网络的带宽、时延和可靠性提出了新的挑战。为了满足XR业务诉求，网业协同技术的研究刻不容缓。



▲图2　XR业务面临的挑战

2 XR网业协同技术

XR业务具有准周期性（可能存在抖动）和实时交互的特征，对时延有更严苛的要求；具有超高清分辨率和高帧率的特征，对带宽有更大的诉求。为了满足用户极致流畅体验如无卡顿花屏等，XR对可靠性提出了更高的要求。

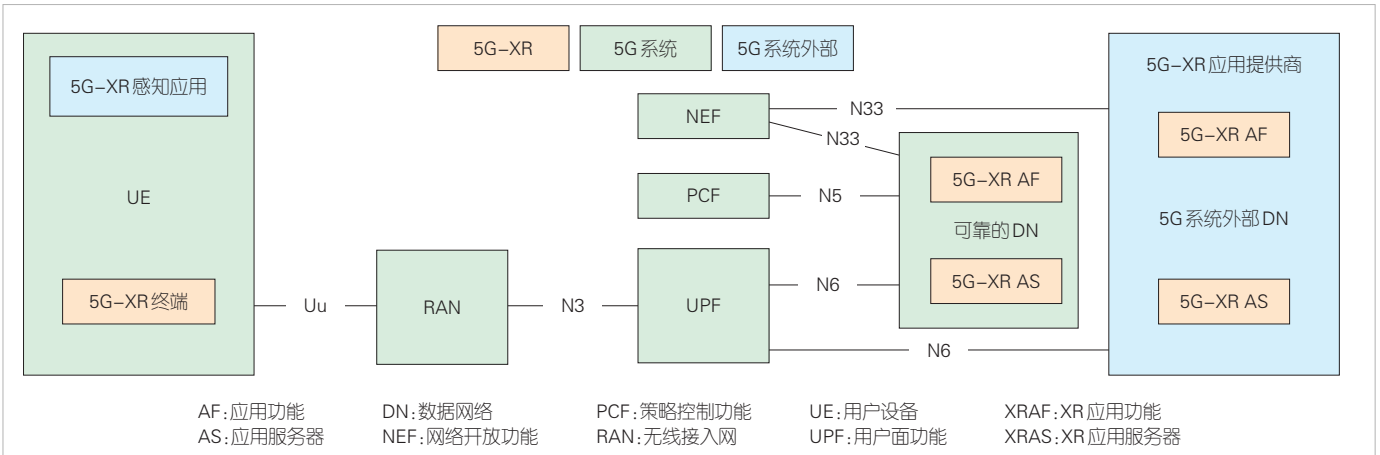
具有高可靠、低时延、大带宽特征的XR业务，在业务网络双向感知、业务保障、和业务评估等方面还有很大的改进空间，如图2所示。需要结合XR业务特征和网络要求，研究面向XR业务的网业协同技术，增强业务与网络的双向感知，实现网随业动、业由网生，构建基于XR业务的广义确定性网络。

2.1 业务网络双向感知

帧是音视频中非常重要的概念，包括I帧、P帧和B帧。I帧包含一幅完整的图像信息，属于帧内编码图像，在解码时不需要参考其他帧图像。P帧是帧间编码帧，利用之前的I帧或P帧进行预测编码。B帧是双向预测编码图像帧，利用之前和（或）之后的I帧或P帧进行双向预测编码，B帧不作为参考帧。XR业务以协议数据单元集（PDU Set）为粒度承载数据，PDU Set等价于帧（如视频帧、音频帧等）。

2.1.1 网络感知业务

网络侧需要感知帧信息，并对XR业务实施帧级保障。第三代合作伙伴计划（3GPP）定义了承载XR业务的5G服务质量标识（5QI），多个PDU Set可映射到同一个5QI，其中包括5QI80和5QI87–5QI90。5QI80业务类型为非保证比特率业务（NGBR），建议承载AR业务，包时延预算（PDB）



▲图1　XR在5G系统的基础网络架构

为10 ms，丢包率（PER）为10^{-6}；5QI87–5QI90业务类型为保证比特率业务（GBR），建议承载交互和视觉渲染服务等，PDB范围为5~20 ms，PER范围为10^{-4}~10^{-3}。目前网络只支持基于5QI和网络切片（可建立多个5QI）识别XR业务，无法满足XR业务的帧级要求，也无法区分帧进行差异化保障，在相同的保障策略下，优先级低的数据帧如P帧占据高优先级数据帧如I帧的调度机会和资源，降低用户体验和容量。

未来，基站需要基于标准接口与核心网进行信息交互，这对核心网提出了新的挑战。核心网需要支持携带上下行XR业务特征、PDU Set信息和部分应用层信息，用于辅助基站对XR业务进行特定的保障。

核心网通过实时传送协议（RTP）报头扩展标记PDU Set，核心网可提供业务识别信息包括多PDU Set间同步感知信息和PDU结束信息、数据突发结束信息、PDU Set重要性、PDU Set序列号、PDU Set中的PDU序列号、PDU Set大小等PDU Set信息和PDU Set错误率（PSER）、PDU Set时延预算（PSDB）、PDU Set集成处理信息（PSIHI）等PDU Set服务质量（QoS）参数信息和QoS流上下行业务的周期、突发到达时间、存活时间、抖动等时间敏感通信辅助信息（TSCAI）[4]，以及应用层信息包括体验质量（QoE）测量信息，如缓存信息、吞吐量等。网络侧根据感知的PDU Set标记等信息进行帧调度和保障，PDU结束信息确定PDU Set边界，网络侧根据边界进行基于PSDB的调度；结合TSCAI信息和PDU Set大小，网络侧进行调度授权；PDU Set重要性表明帧的重要性，网络侧以此确认优先级并结合PSIHI进行丢帧等。

2.1.2 业务感知网络

业务侧需要感知网络状态，根据网络状态调整其清晰度/帧率和业务达到时间等。网络侧可将网络拥塞状态/推荐的业务速率和业务偏移时间等信息通知业务侧，业务侧通过调整清晰度/帧率和业务达到时间进而调整业务突发速率，减少拥塞。

例如：在摄像头十分密集的港口多座岸桥场景中，I帧携带视频画面全部信息，数据量较大。当多个网络摄像机同时向网络平台发送I帧，多个I帧在同一时刻传输，发生I帧碰撞，导致瞬时网络流量峰值超过网络传输能力，网络传输链路拥塞，视频出现延迟、卡顿以及丢包等问题。目前可通过网络侧适当延长空口处理时延，错开I帧，达到降低视频回传业务对空口峰值速率的要求。未来，业务侧需要实时感知网络能力，拥塞时调整业务特征，错开I帧到达时间，减少碰撞，降低网络负担，保障用户体验。

2.2 业务保障

对具有高可靠、低时延、大带宽的XR业务，从时延、带宽、可靠性、终端节能的维度采取相应的业务保障技术，增强网络和业务的协同，为用户提供沉浸式体验。

2.2.1 带宽保障

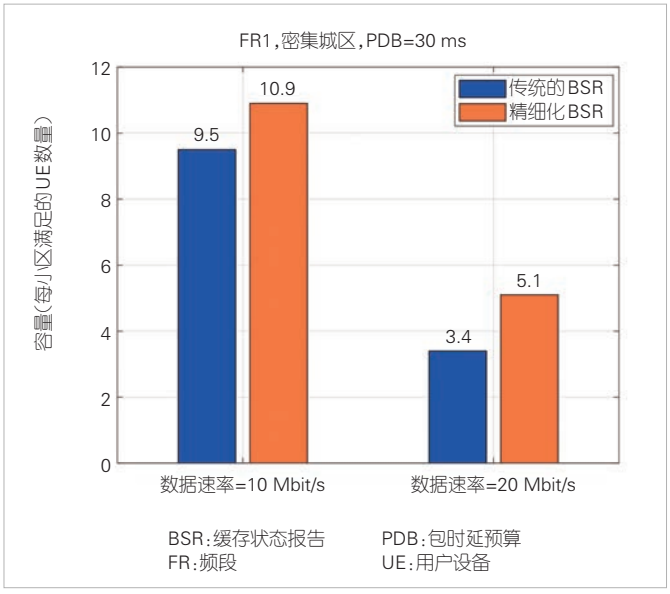
1）缓存状态报告（BSR）增强

上行动态调度BSR增强包括对BSR上报的数据量精度的增强。我们最早在Rel–17标准讨论中提供相关的仿真分析[5]，发现针对上行XR视频业务的数据量大小，当前协议下的BSR上报的数量和真实数据量差距较大，导致基站为数据包分配的资源冗余，造成资源利用效率下降，从而降低系统容量。对此，我们在Rel–18标准讨论中提出了增强BSR上报精度的解决方案，为XR业务的数据量范围专门设计了一个BSR的表格[6]，通过线性插值/指数插值的方式，使BSR上报精度提高，从而提高系统容量。图3给出了BSR精度对系统容量的影响分析[5]。

2）载波聚合（CA）

为满足XR业务的大带宽诉求，网络通过CA技术将多个载波聚合在一起，终端通过多个载波传输数据，增加上下行传输带宽，满足带宽保障。

100 M带宽下，2.5 ms双周期帧结构与2.5 ms双周期帧结构带内CA（intra–band CA），上下行流量相对于2.5 ms双周期单载波增益为100%；5 ms单周期帧结构和2.5 ms双周期帧结构intra–band CA，上下行流量相对于5 ms单周期帧结构单载波增益分别为150%和83%。



▲图3　上行XR视频业务的BSR精度对系统容量的影响

3）双链接（DC）

终端使用两个及以上基站的无线资源，同时发送和接收多个载波的数据，提升用户吞吐量，降低切换时延。

4）1D3U帧结构

为了应对具有上行大带宽特征的XR业务，网络侧采用1D3U的帧结构，保障上行带宽，满足用户体验。在相同配置下，1D3U帧结构上行极限流量可达766 Mbit/s，是2.5 ms双周期帧结构上行极限流量的2倍，是5 ms单周期帧结构上行极限流量的3倍。

5）智能优先比特速率（Smart PBR）

XR业务的速率是动态变化的，同时XR业务的带宽保障要求比较高，固定的PBR配置无法满足带宽要求。对此，网络自适应学习XR业务速率，并将其配置给XR业务专载，满足带宽诉求。实测终端灌包，开启Smart PBR功能的终端流量更接近实际XR业务流量，避免对XR的过度保障，节约了网络资源，提升整体系统容量。

2.2.2 时延保障

1）增强的上行免调度（eCG）

针对XR上行视频业务的数据量大，数据包大小随时间随机变化，具有周期性强、时延要求高等特点。我们在Rel-18标准讨论中最早提出了增强的上行免调度，主要包括两个特征[7]：第一，基站支持单周期内多个资源配置；第二，终端支持未使用资源的上报，基站进行动态释放。对于第一个特征，eCG能够减少上行动态调度的信令交互所产生的时延，极大地降低了上行调度的时延。对于第二个特征，由于XR上行视频业务的数据包的数据量随时间变化，预配置的传输资源可能存在浪费的情况，对于这种情况，中兴通讯提出了资源的回收和重调度，即引入了未使用传输时机上行控制信息（UTO-UCI），使UE可以上报配置但未使用的资源给基站，基站根据UE上报的资源使用情况，对未使用资源进行回收并重新调度。中兴提案中的仿真结果最早证明[7]，其提出的上行多传输时机调度，在时延要求高的上行XR视频传输场景下，容量有非常显著的提高。此方案得到各家公司仿真证明并认可，已被5G标准采纳。图4为上行多传输时机免调度机制仿真结果[8]。

2）CA载波分流

基于CA场景，网络根据业务特征对多个业务进行差异化分流，当用户存在低时延，大带宽业务的情况时，引导低时延业务在低时延载波进行调度，确保时延体验；对于视频类大带宽业务，引流到有利于大速率的载波上，满足带宽诉求。

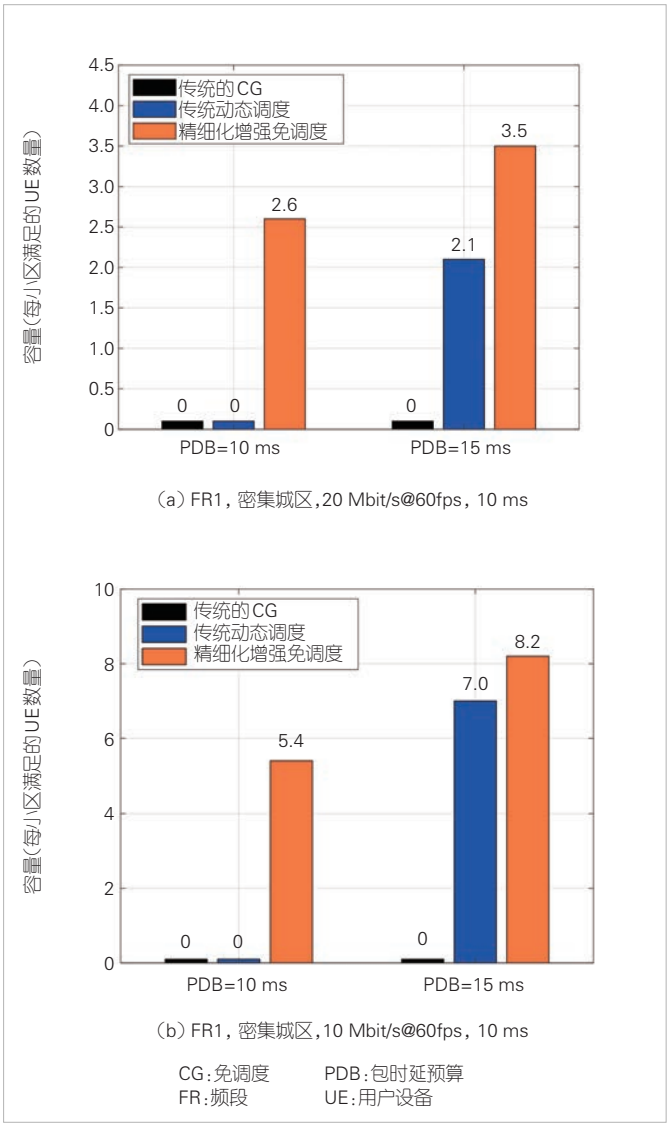
3）时分双工（TDD）CA互补帧结构

网络侧根据业务的低时延和大带宽诉求，在两个TDD载波上配置互补的上下行时隙配比，通过跨载波调度，使得终端在任何时刻都有上行传输机会和下行传输机会，为XR业务提供带宽能力提升和极致的空口时延。根据实测结果，CA互补帧结构场景下，端到端最小时延为3.4 ms，平均时延为4.3 ms，相比中移单载波降低30%。

4）双活动协议栈切换（DAPS HO）

在DAPS HO中，UE在收到切换命令后同时保持同源基站以及目标基站的连接，直到收到来自网络侧的指示释放和源基站的连接[9]。由于在切换时UE依然可以和源基站通信，减少切换中断时延，极致可达0 ms切换中断时延，满足XR业务对移动性的要求。

5）PSDB调度



▲图4 上行多传输时机免调度机制仿真结果

XR业务以PDU Set为单位，时延要求需要在PSDB范围之内。网络侧以业务的时延要求PSDB和PDU Set的时延余量为权重进行资源分配，提升XR业务时延满足度。

6）DS帧结构

网络侧采用时域分割的方式，将时域划分为多个时隙，每个时隙包含多个符号，通过灵活配置时隙和符号，选择合适的传输模式，为XR业务提供低时延和高可靠的通信。

7）能力开放

能力开放是网络感知业务需求和特征，基于相对确定性的业务特征，提升无线空口时延、带宽、可靠性的确定性保障能力。同时，网络抽象成能力服务开放给业务，通过智能和闭环化运作，达到XR行业客户要求的服务水平协议（SLA）保障，促进网络和业务的跨层协同优化[10]。通过能力开放，在DS帧结构下，可达到端到端时延10 ms和可靠性99.999%的SLA保障。

8）动态分离渲染

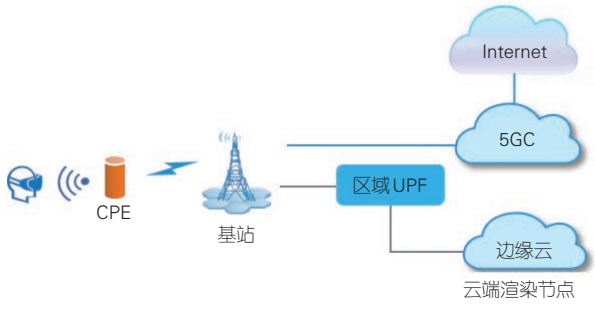
根据业务需求、服务器算力和负荷等信息，系统动态选择渲染节点，如通过5.5G的高带宽低时延，将终端的渲染能力放到边缘云，实现了高效快速的渲染服务，通过云网边端协同支持XR业务，如图5所示。

9）异步时间扭曲（ATW）

ATW是一种生成中间帧的技术，当XR业务不能保持足够帧率的时候，终端通过ATW产生中间帧，弥补XR运行中因为渲染延时而丢失的大部分帧，从而有效减少画面时延和抖动，在消耗很少的计算资源前提下大幅提升图像的连贯性，通过端网业协同，提升用户体验。

10）低延迟、低丢包、可扩展吞吐量（L4S）

XR应用服务器根据网络情况实时调整业务速率以适应网络条件，确保用户的期望体验[11]。当网络发生拥塞时，使用显示拥塞通知（ECN）比特位来标记数据流，并通过ECN比特位标记与应用层交互，应用层基于ECN比特位的反馈触发速率自适应，如图6所示。应用层业务和网络之间进行信息交互，达到降低时延、减少拥塞和保证用户体验的效果。



5GC：5G核心网　CPE：客户终端设备　UPF：用户面功能

▲图5　动态分离渲染

2.2.3 可靠性保障

1）CA PDCP Duplication

CA PDCP Duplication采取冗余传输的方式，在两个载波上传输同一份数据，提高无线传输的可靠性、降低无线传输的时延，从而提升用户体验。经过实测，相对于普通承载，CA PDCP Duplication承载达到可靠性提升70%，端到端时延提升17%的效果。

2）基于PSIHI的帧传输

PDU Set集成处理信息（PSIHI）指示接收侧的应用层是否需要PDU Set的所有PDU来使用PDU Set，当PSIHI指示为True，网络侧需要保障PDU Set的完整传输，通过丢帧，避免XR业务错误传播，保障PDU Set传输的可靠性。

当PSIHI指示为False，如果关键帧I帧传输错误，丢弃I帧剩余报文及其后的所有P帧，直到下一个I帧，如图7（a）所示。若P帧编码不是参考帧，P帧传输错误，丢弃P帧剩余报文；若P帧是参考帧，则P帧传输错误时，丢弃当前P帧剩余报文及其后的所有P帧，直至下一个I帧，如图7（b）所示。

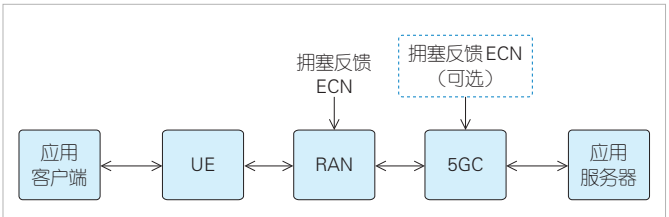
当PSIHI指示为True，如果关键帧I帧传输错误，丢弃当前I帧及其后的所有P帧，直到下一个I帧。若P帧编码不是参考帧，P帧传输错误，丢弃当前P帧；若P帧是参考帧，则P帧传输错误时，丢弃当前P帧及其后的所有P帧，直至下一个I帧。

3）拥塞丢帧

网络侧识别到拥塞时根据PDU Set Importance进行丢帧操作，丢弃不重要的帧，减少拥塞，降低时延、提高可靠性。

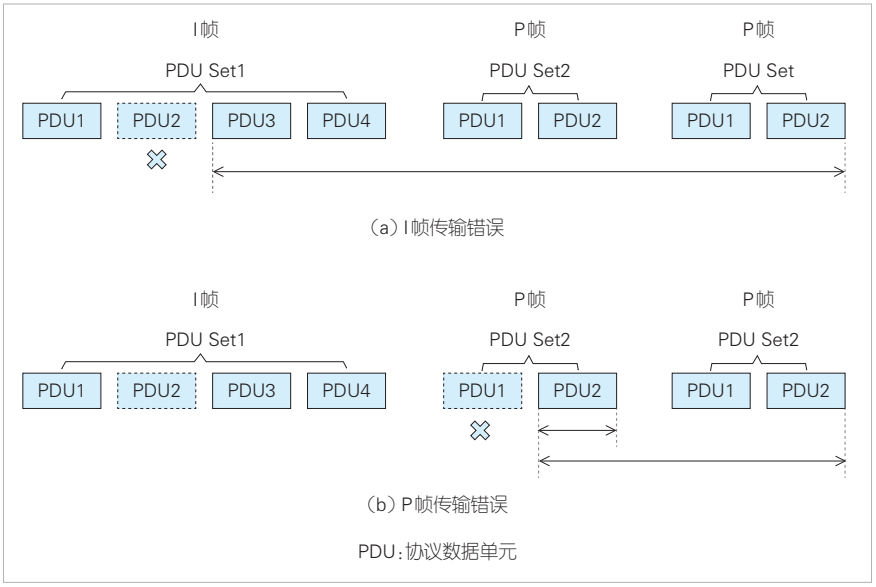
2.2.4 终端节能

3GPP考虑了在保障业务传输的前提下实现终端节能的方案。其终端节能方案主要根据XR业务的特点进行设计。



5GC：5G核心网　ECN：显示拥塞通知　RAN：无线接入网

▲图6　低延迟、低丢包、可扩展吞吐量（L4S）机制



▲图7 PDU Set集成处理信息指示为False时,I帧传输错误和P帧传输错误的情况

XR业务有准周期性（可能存在抖动）特性，且XR业务主要为视频业务，其周期与每秒发送的帧数有关，因此多数视频业务的周期为非整数。在终端节能方案中，非连续接收（DRX）是常用的，节能效果较好的一个方案。但是DRX的周期在Release 17以及之前的版本中都是整数值，这与XR业务的非整数周期特性不匹配，这种不匹配会造成XR业务与DRX激活时间窗的错位，导致XR业务的调度时延增加，影响业务传输。这个问题最早由我们提出[12]，在后续的XR标准讨论中，各公司针对所述问题考虑对DRX进行增强，最终确定了配置非整数周期的DRX增强方案。非整数周期的DRX可以和XR业务配置相同的周期，使得XR业务的到达时间可以和DRX激活时间窗对齐，从而保障XR业务的有效传输。

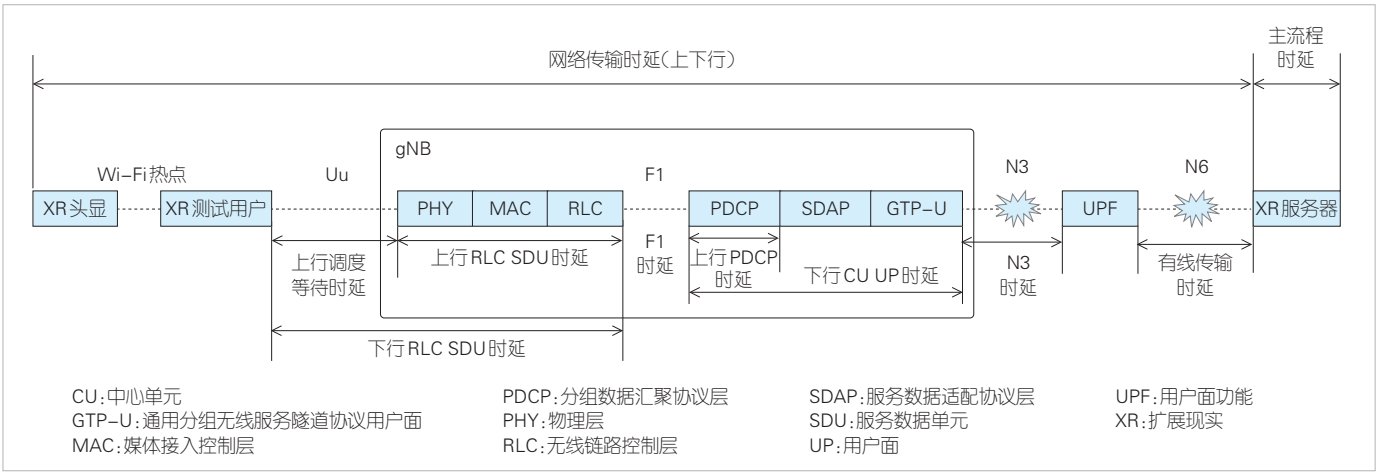
2.3 业务评估

3GPP定义了影响XR业务QoE的测量量，包括码率、帧率、FoV、自由度（DoF）、分辨率、刷新率、解码能力等[13]，但是这些测量量主要分布于终端和服务器，网络侧大多无法获取。现有评价标准方法适合基于渐进式下载流和基于可靠传输的自适应流媒体，对于拆分计算和渲染的XR业务，沉浸度要求更高，需要新的评估方式以指导保障策略的选择和优化。

XR业务质量包含卡顿、花屏、黑边、音视频不同步等[14]。影响XR业务质量的主要是从用户操作经过终端传感器获取到姿势动作信息到下一帧画面清晰地显示在头显期间的端到端交互响应时延、丢包、抖动和速率。由于端到端部分指标无法获取，网络侧目前可通过分解全流程中各段时延、丢包、抖动等指标，配合针对性测试，研究网络侧指标与业务质量存在的关系，通过网络侧指标间接评价业务质量。

针对XR业务提出基于无参考（NR）的业务体验客观评估方法，该方法以XR系统的测量量和网络关键绩效指标（KPI）等参数作为入参，通过AI学习和数学拟合，获得XR业务体验评分。网络KPI指标包含帧级网络传输时延、帧级网络传输时延分布、帧级抖动、帧级速率和帧级丢包等；其中，网络传输时延包含多段，如图8所示。帧级指标的统计强依赖核心网对PDU Set级QoS参数的支持和网络侧对业务帧级特征的识别。

扩展现实平均意见得分标识(XRMI)为XR业务的平均意见得分（MOS）值，反映了小区中XR业务质量综合体验。



▲图8 网络传输时延示意图

基于呈现质量（如卡顿、花屏、黑边等）、交互质量（如交互响应、抖动、音视频同步等）、图像质量（如码率、帧率、分辨率等）维度构建XR评估模型。不同业务场景的XRMI关联指标不同，如XR全景视频、影院等弱交互场景和XR实时游戏等强交互场景对交互指标的关注度存在差异，需要根据不同的业务场景进行XRMI评估，XRMI与主观感受的映射关系见表2。

3 总结与展望

本文介绍了沉浸式XR和元宇宙的含义、网元结构、应用场景，着重分析XR业务在业务网络双向感知、业务保障和业务评估方面的挑战，研究和提出了满足XR业务大带宽、低时延和高可靠诉求的网络业务双向感知协同技术。

未来期望进一步研究网络和业务的深度融合技术，包括XR业务的多流协同、多模态同步、核心网/应用层与基站的交互机制等；结合体验满足度的定量研究和测试结果，形成XR业务体验质量评估的分档标准，推动沉浸式XR的发展，促进元宇宙从概念走向现实。

随着XR、5G、网业深度融合、云计算等技术成熟度的提升，元宇宙的正向循环将逐步打通，底层技术推动应用迭代，市场需求提升反哺底层技术持续进步。元宇宙将真正改变我们与时空互动的方式，对社会和个人带来广阔价值空间。

▼表2　XRMI与主观感受的映射关系

级别	XRMI分值	评价标准
优	5	音视频互动体验流畅，音视频以及交互操作不存在任何可感知的卡顿、花屏、黑屏、声音不连续等，存在感、沉浸感极强，达到了类似于真实世界的效果，无晕眩现象，不可察觉
良	4	音视频互动体验尚可，偶尔存在很轻微的卡顿、花屏、或声音不连续等，存在感、沉浸感尚可，达到逼近真实世界的效果，偶尔存在轻微晕眩现象，可感知但不令人讨厌
中	3	音视频互动体验一般，存在少量可感知的卡顿、花屏、声音不连续等，其余时间正常，存在感、沉浸感一般，少量时间无法达到真实世界的效果或存在晕眩现象，无大量连续卡顿花屏现象，有点讨厌
差	2	音视频互动体验差，存在大量连续卡顿、花屏、声音不连续等，其余时间正常，存在感、沉浸感差，大量时间无法达到真实世界的效果或存在晕眩现象，很讨厌
劣	1	音视频体验非常差，绝大部分时间存在可感知的卡顿、花屏、声音不连续等，存在感、沉浸感非常差，体验完全背离真实世界，大部分时间存在晕眩现象，非常讨厌，不想继续体验，也不推荐其他人体验

XRMI：扩展现实平均意见得分标识

致谢

感谢中兴通讯股份有限公司黄俊、王美英、徐俊、沙秀斌、王新台、戴博等专家对本研究的帮助！

参考文献

[1] 3GPP. Extended reality (XR) in 5G: TR 26.928 version 16.1.0 [S]. 2021
[2] 唐宏. 5G XR 技术与应用 [M]. 北京: 人民邮电出版社, 2021
[3] 中国信息通信研究院. 虚拟(增强)显示白皮书 [R]. 2019
[4] 3GPP. System architecture for the 5G system (5GS): TR 23.501 [S]. 2020
[5] ZTE. Performance evaluation results for XR [Z]. 2021
[6] ZTE. BSR enhancements for XR [Z]. 2023
[7] ZTE. Discussion on XR specific capacity enhancements techniques [Z]. 2022
[8] ZTE. XR specific capacity enhancements [Z]. 2022
[9] 3GPP. NR and NG-RAN overall description: TR 38.300 [S]. 2018
[10] 中国移动通信有限公司研究院. 无线云网融合智慧服务白皮书2.0 [R]. 2021
[11] 3GPP. Study on XR (extended reality) and media services: TR 23.700 [S]. 2016
[12] ZTE. Considerations on XR specific enhancements [Z]. 2021
[13] 3GPP. QoE parameters and metrics relevant to the virtual reality (VR) user experience: TR 26.929 [S]. 2020
[14] 中国移动. XR网络技术体系白皮书 [R]. 2023

作　者　简　介

李娜，中兴通讯股份有限公司无线算法工程师、移动网络和移动多媒体技术国家重点实验室研究员；研究方向为移动网络和沉浸式XR。

张诗壮，中兴通讯股份有限公司首席专家、移动网络和移动多媒体技术国家重点实验室未来无线和边缘网络架构方向学术带头人；主持完成信源编码、统一移动网络硬件平台、5G大容量基带池等重大项目，为中国移动通信系统关键技术创新做出了重要贡献；获得国家科学技术进步奖特等奖1项。

程义超，中兴通讯股份有限公司技术规划部副部长、移动网络和移动多媒体技术国家重点实验室办公室主任、移动网络高级项目经理、资深研发总工；主要从事5G、6G无线通信系统的技术预研工作，拥有15年以上移动网络研发经验。

中兴通讯技术杂志社
促进产学研合作青年专家委员会

刊物相关信息

投稿须知　投稿平台　过刊下载　论文索引与引用指南

中兴通讯技术

(ZHONGXING TONGXUN JISHU)

办刊宗旨：

以人为本，荟萃通信技术领域精英
迎接挑战，把握世界通信技术动态
立即行动，求解通信发展疑难课题
励精图治，促进民族信息产业崛起

双月刊　1995 年创刊
第 30 卷　总第 178 期
2024 年 7 月　增刊 1

主管：安徽出版集团有限责任公司
主办：时代出版传媒股份有限公司
　　深圳航天广宇工业有限公司
出版：安徽科学技术出版社
编辑、发行：中兴通讯技术杂志社

《中兴通讯技术》编辑部
地址：合肥市金寨路 329 号凯旋大厦 1201 室
邮编：230061
网址：tech.zte.com.cn
投稿平台：tech.zte.com.cn/submission
电子信箱：magazine@zte.com.cn
电话：（0551）65533356

发行方式：自办发行
印刷：合肥添彩包装有限公司
出版日期：2024 年 7 月 29 日
中国标准连续出版物号：ISSN 1009-6868
CN 34-1228/TN
增刊备案号：341228202401
定价：每册 20.00 元