6G无线多模态通信技术

任 超 丁思颖 张晓奇 张海君*
(北京科技大学 北京 10083)

摘 要：该文综述了多模态通信作为一种能够同时交互多种模态形式的信息转移方式在不同应用场景下的重要性及其未来在6G无线通信技术中的发展前景。首先，将多模态通信分为3类，并探讨了其在这些领域中的关键作用。随后，针对6G无线通信系统可能面临的通信、感知、计算和存储资源限制以及跨域资源管理问题进行了深入剖析，指出未来的6G无线多模态通信将实现通感算存的深度融合和通信能力的提升。在多模态通信实现过程中，必须考虑多个环节，包括多发送端处理、传输技术和接收端处理等，以解决多模态语料库构建、多模态信息压缩、传输、干扰处理、降噪、对齐、融合和扩充等方面的挑战，以及资源管理问题。最后，强调了6G网络的跨域多模态信息转移、互补和协同的重要性，这将更好地整合和应用海量异构信息，以满足未来高速、低延迟、智能互联的通信需求。

关键词：多模态通信；多媒体信息；无线通信技术
中图分类号：TN92 **文献标识码**：A **文章编号**：1009-5896(2024)05-0001-14
DOI: 10.11999/JEIT231201

Wireless Multimodal Communications for 6G

REN Chao DING Siying ZHANG Xiaoqi ZHANG Haijun
(*Beijing University of Science and Technology*, *Beijing* 100083, *China*)

Abstract: An overview of multimodal communication as an important information transfer mode that can simultaneously interact with multiple modal forms in different application scenarios is proposed in this paper. The future development prospects of multimodal communication in 6G wireless communication technology is also discussed. Firstly, multimodal communication is classified into three categories, and its key roles in these fields are explored. Furthermore, a deep analysis is conducted on the communication, sensation, computation, and storage resource limitations, as well as cross-domain resource management issues that 6G wireless communication systems may face. It points out that future 6G wireless multimodal communication will achieve deep integration of communication perception, computation, and storage, as well as enhance communication capabilities. In the process of implementing multimodal communication, various aspects must be considered, including multi-transmitter processing, transmission technology, and receiver processing, in order to address challenges in multimodal corpus construction, multimodal information compression, transmission, interference handling, noise reduction, alignment, fusion, and expansion, as well as resource management issues. Finally, the importance of cross-domain multimodal information transfer, complementarity, and collaboration in the 6G network is emphasized. This will better integrate and apply a massive amount of heterogeneous information to meet the future communication demands of high-speed, low-latency, and intelligent interconnection.

Key words: Multimodal communication; Multimedia information; Wireless communications

收稿日期：2023-11-01；改回日期：2024-03-01；网络出版：2024-03-11
*通信作者： 张海君 haijunzhang@ieee.org
基金项目：国家自然科学基金(62201034, U22B2003, 62341103)，北京市自然科学基金(L212004, L212004-03)
Foundation Items: The National Natural Science Foundation of China (62201034, U22B2003, 62341103), Beijing Municipal Natural Science Foundation (L212004-03, L212004)

1 引言

作为拥有复杂、多元资源的物理-信息融合世界中的关键桥梁，未来无线通信需要加速演化以满足社会的多维需求。6G通信技术有望在未来实现全球网络覆盖，并提供更高的可靠性、更快的传输速度和更强的智慧互联能力，在智慧城市、智能医疗和应急通信等领域支撑高效、可靠的多传输媒介协同接入和跨域泛在服务。然而，6G无线通信系统在空天地一体、全息临场通信等复杂环境中可能受到设备能源携带、实时处理和通信性能的制约，导致云-边-端3侧的任务执行失败。6G智联网(Artificial Intelligence of Things, AIoT)设备数量的激增也对计算、感知和通信等分属于不同域的资源和服务提出了更高的资源管理要求，这促使6G无线网络需要加强在通感算存一体化方面的研究。因此未来6G通信，从一个新的跨域融合角度，通过共享和传递多种媒介获得的多种模态信息和通感算存跨域资源，更好地减轻云-边-端3侧、通感算存多域以及空天地无线传输中与任务无关的复杂负面影响。在这一背景下，6G无线网络中多模态通信的定义和范畴将得到不断完善和扩展，以涵盖通信、感知、计算和存储等广泛领域。6G作为人类社会与虚拟数字空间的通信桥梁，还应支持对世界的广泛感知和内生运算(包括人工智能、信息处理、和存储等)以最终实现物理世界和数字世界的高效互动和高度融合，即“自由连接的物理数字融合世界”愿景。例如，未来AIoT应用需要同时负责感知、传输、存储和处理数据，而6G无线网络将原生地支持通信、感知、计算和存储服务，从而形成构建未来自由连接的物理和数字世界的多模态数字基础能力。这种多模态数字基础能力对通信、感知、计算和存储的融合要求很高，可以视作多种跨域模态的融合，并通过无线技术在6G网络中更灵活地连接这些新型多模态信息和资源。

早期的多模态通信指的是能够同时交互多种媒体形式的信息转移、互补和协同的方式。这些传统的媒体模态包括文本、图像、音频和视频等。当前，这种以多媒体信息为基础的多模态感知技术，已经在医疗、交通和教育等领域得到了广泛应用。例如，在医疗领域中，结合感知到的图像和声音等多种模态可以更全面、准确地检测患者的身体状况；无人驾驶汽车可以通过结合摄像头、雷达等捕获的图像和视频来更准确地判断前方道路的情况和司机的驾驶疲劳度，并作出相应的调整；在线教育可以感知并综合视频、声音和文本等多种模态来评估学生的课堂状况，并调整教学节奏。此外，多模态识别技术也在实际应用中得到了广泛的推广，利用视频中的语音和视觉模态的相关性，可以完成多模态情感识别分析[1]；通过不同类型的传感器采集到包含文本、图像等多模态数据的云数据集，利用深度学习方法学习异质性深度特征，实现云数据的分类，更好地服务与遥感技术的应用[2]。多模态技术目前得到广泛应用的原因在于合理利用了“多”种形式的技术和信息，形成了技术、信息集群效应。例如，在灾害搜救中，仅依靠图像可能无法找到被废墟掩盖的伤员，但利用多模态感知的数据分析可以准确地定位援助位置、检测生命体征；在国家安全方面，单一的视觉模态可能无法找到隐藏的威胁，但多模态交互可以使无人机等侦查设备更精确地定位目标，并能更好地利用各种模态所收集的数据隐藏自身；聚焦个人生活，多模态信息融合分析提高了无人驾驶的安全性，丰富了社交媒体体验，增强了在线教育的多样性。

可见，通过多种来源于感知设备、存储器甚至算法导出的模态之间信息和资源可以通过多模态信息传输实现相互支持，从而更好的发挥不同模态的优势，深度计算让多模态“聚变”输出更丰富的结果，保障更准确地完成相关任务。站在通信的视角，多模态通信的内涵可以理解为：信源应尽可能多的寻找可用的信息和信息形式，并通过选择最合适的无线资源块将它们尽可能独立地传输给信宿，信宿则利用自身可用的资源协同处理这些独立信息，让整个通信系统将更好的支持海量、并发、面向全感官的未来任务。鉴于此，未来面向6G的无线多模态通信，将把多模态的概念扩充到更多资源域，在更多层面上更为深入地交叉多模态技术、计算机、人工智能技术和无线通信技术。

2 多模态通信的分类

可以根据不同的底层实现和应用场景将多模态通信进行分类，见图1。

2.1 以多媒体技术为基础的传统多模态通信技术

传统的计算机通信网络通常需要传输和处理多种不同的媒体形式，如文本、图像、视频、音频等。在连续情感识别和智慧城市等现代应用中，媒体形式往往被定义为多模态。例如，在构建多模态情感库的过程[3]中，音频和视觉可以共同作为多模态输入。近年来，多模态技术主要研究如何融合不同类型的数据，并以图像和文本作为主要的融合对象。神经网络的发展和应用，实现了对多种媒体形式的信息更好的特征提取和融合，例如在网站类型判断[4]中，可以结合网站上的文本内容和图像进行分类。随着越来越多种媒体形式被纳入处理范畴，传统的多模态通信开始加速发展。

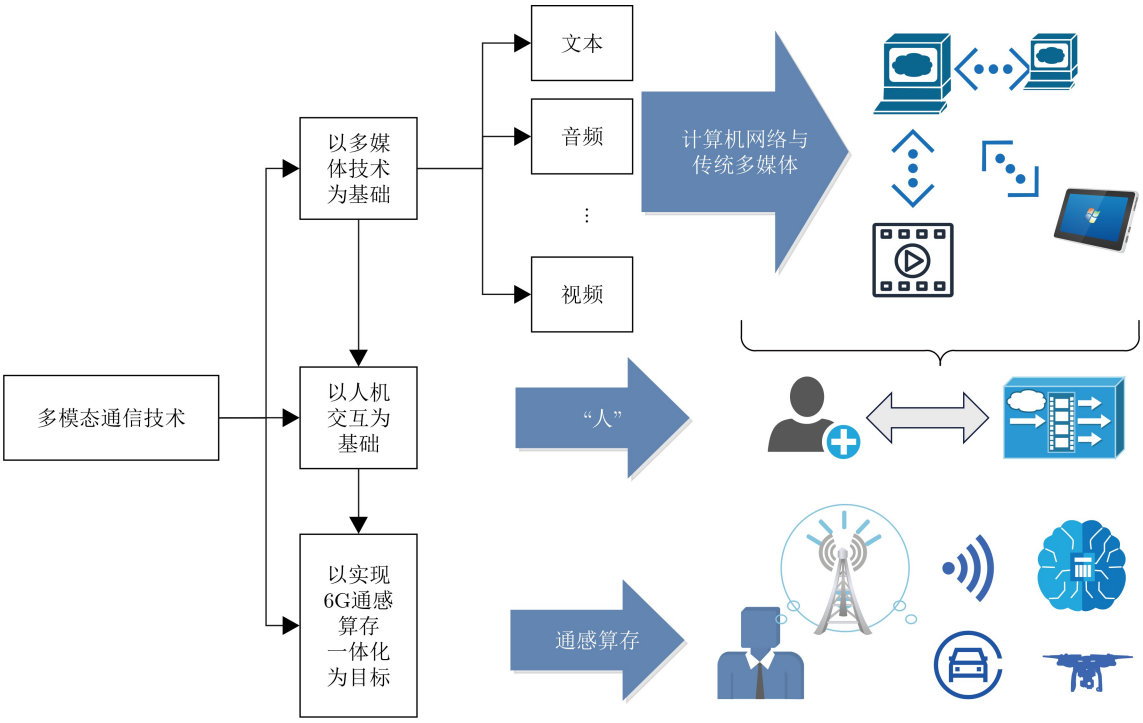

图 1 多模态通信分类

以多媒体技术为基础的多模态通信主要表现为一种以计算机网络或神经网络为中心的信息交互管理模块，实现系统信息获取、可视化呈现、多媒体分析识别、记录存档和信息交互等功能[5]。在电子参与领域，多模态通信通过多种线上媒体的方式促进公众在政府和公共事务中的参与合作，逐渐将文本交流从纸质空间转移到数字空间。尽管基于文本的交流(通过论坛、聊天、电子邮件和社交媒体等方式)易于使用和处理，但它无法有效传达场景或情感，也难于反映复杂的概念和想法[6]，因此需要更多模态来提供支持。同时，新兴的虚拟现实(Virtual Reality, VR)技术也为实现的多模态通信提供了一个整合多种主流多媒体的统一通信平台[6]。

以多媒体技术为基础的多模态通信在发展初期存在跨模态重要信息丢失的问题，美国空军研究实验室开发了一种以网络为中心的多模态通信监控套件，尽管这种多任务间通信可以跨不同可视化终端，但未对人的参与进行很好优化，因此常出现因操作员失误而丢失信息的问题。多模态信息恢复技术可以在跨模态交互过程中改善信息丢失，解决不同信息权重占比和多模态信息干扰等问题[7]。另一方面，多模态配准和对齐技术[8]，如基于Shearlet的模态鲁棒描述符(Shearlet-based Modality Robust Descriptor, SMRD)技术，可以对来自不同传感器或成像设备的多模态信息进行对齐、准确定位和匹配，降低跨模态信息丢失的发生率。在多模态通信交互中，可以进行多模态图像融合，解决不同模态图像之间的非线性强度变化、纹理和边缘信息不一致等跨模态通信问题，可以更好地处理所获得的数据，有利于与其他模态进行融合。对于图像这一模态，可以使用Neo4j来存储和可视化图数据，在多模态通信交互过程中快速检索和搜索相关资料[9]。在图像特征提取方面，可以应用基于VGG19的模型提取图像的关键信息，最大程度地减少无关信息的干扰[3]。对于文本这一模态，可以使用Google于2018年提出的基于Transformer架构的开源语言理解模型，即基于Bert的文本特征提取模型框架[3]，对大量文字进行筛选与提取，减小减轻通信的压力。在多媒体技术的交互进程中，当发送端获取到任意两种模态时，并需要联合处理时，就已经进入了以多媒体技术为基础的多模态通信领域；进而，系统通过不同类型多媒体数据内容上的关联，辅助利用所有数据来达成整体目标。因此，还可以使用多模态融合矩阵分解双线性池化检测模型[10]等端到端模型，或使用多模态的深度交互图神经网络模型[11]等技术。

2.2　以人机交互为基础的多模态相关通信技术

传统的基于多媒体技术的多模态通信受到领域本身的制约，其信息交互过程仅限于计算机系统甚至算法内部，在涉及到人为操作时也可能出现跨模态信息丢失的问题，而以人机交互为基础的多模态相关通信技术则将“人”作为新的参与者引入到信

息交互过程中，通过更大程度地将用户纳入通信过程，这种方式可以更好地满足用户对于多种模态的需求。随着远程操作成为人机交互的现代标准，混合主动的士兵-机器人团队使用多种通信方法，以流畅自然的方式实现信息共享。因此，在人机交互应用中，利用多模态收集到的信息可以更好地传递人的需求，使得人与人、人与机器之间的联系更紧密[12]。在基于人机交互的多模态通信中，现代计算和通信技术能够结合“人”的角色和需求，将抽象信息转化为可以通过全面处理文字、声音、图形、图像等信息进行感知、管理和交互[13]。

随着人机交互领域多模态通信需求的增长，相关服务亟需应用大量高效灵活的信息交互技术和多模态联合处理技术。在虚拟现实仿真中，某些任务可以通过多模态反馈信道将数据传输至接收端，让操作者能够快速在视觉和听觉上了解机器人所处位置和运行状态，从而通过多模态通信提升服务的准确性[14]。在多模态通信中使用的触摸界面、语音识别、人体检测、手势识别等识别方法，能够更好地实现机器与人的多模态通信，并丰富二者之间的互动[15]。在人与计算机的信息交互中，可以基于自主意识的虚拟空间来建立感知模型，通过搜索环境中各个模态的信息辅助计算机建模，达成高效的人-机器通信 [16]。当前智能感知模块的输入包括语音交互、眼动交互和身体交互等各种交互方式[17]，这些多模态信息的反馈能够对操作者的认知状态和交互意图提供全面的决策支持。在面向人机交互的多模态通信的过程中，接收端所收到的数据通常较为零散。为了方便对数据进行分析、管理和利用，需要将信息结构化处理，即将不同模态获取的信息合理地放置在最合适的位置，以使得所获得的信息结构有序，提供最佳性能的多模态、全感官服务。

2.3 以6G通感算存一体化为目标的新兴多模态通信技术

近些年，多模态通信从多媒体数据类型逐渐扩展——涵盖整个信息通信过程的信息和资源的异构、跨域类型。6G旨在全球覆盖中以极低延迟快速传输和处理信息，同时网络还需要附带更大的空间来存储信息，逐渐形成了以6G通感算存一体化为目标的新兴多模态通信技术。其中，通信域模态指调动通信媒介资源(如网络设备、无线频谱、光纤、声波等)，以传输指令、多媒体信息或其他任务的多模态相关信息；6G感知域模态主要指通过多种媒介资源(如摄像、声呐、雷达、无线接收机等)获取环境的多模态信息，包括图像、声音、视频等；计算和存储域模态指调动计算资源(如终端的处理器、硬盘、云-边设备等)，处理信息和认知环境，并生成相关结果的多模态信息。最终，6G无线新兴多模态通信的目标是形成跨域的泛在感知、综合计算、存储和智能通信的融合一体化无线网络。

在以感知域为核心的多模态通感融合中，需要加强感知域多模态数据的人工智能分析与计算，并与通信域形成资源和信息的一体化。在以计算域为核心的多模态算网融合中，需要通信域资源提升算力互通，进一步推动云网融合、算存融合、算力网络、算网云调度技术的发展，促进算力资源的充分均衡，最终实现感知、计算(含存储)、通信三者的跨域跨模态融合。进而，6G可以推动多模态任务和海量异构需求向网络边缘移动，减少数据传输到云端，降低延时，同时形成通感算存一体化的多模态数字基础能力，增强每种资源和信息的可用性，实现更灵活、自然的无线多模态通信。然而，目前以6G通感算存一体化为目标的新兴多模态通信在各个跨域模态之间深度融合的机理层相关研究较少。

一种通感算一体化网络架构[18]表现为将通信、感知与计算等功能融合，通过对业务需求和网络物理与数字空间状态进行感知，实现业务加载时的预配置和动态调整，合理分配计算任务，为6G无线多模态通信的整体架构提供了一个系统级思路。除此之外，现有研究更多集中于通感算存一体化为目标的新兴多模态通信的典型应用场景。

在6G通感一体化场景中，为将通信与感知的跨域多模态融合，可以通过研究通信和定位感知所涉及的无线资源的多模态共性，实现通信和位置感知的机理层融合，让位置感知的通信信号接收合一，并提高通信和感知域资源的利用率[19]。

在6G通算存一体化场景和人工智能即服务场景中，存在跨域资源的空时分布不均、多模态资源互相“兑换”困难以及不同模态权重分配等挑战性问题，因此为解决多模态通信系统中通感算存等资源的跨域管理问题，可以使用一种基于优先级的中间媒介变量，应用经济学思路来处理超密集AIoT中的计算、感知资源的资源管理问题[20]。

在6G空天地一体化场景中，一种基于6G无线多模态通信的无人机网络[21]旨在同时同频在无线通信媒介上实现无线供电、无线通信和定位感知功能，然而这些多模态无线信号可能仅仅适用于接收端的不同组件，并导致多模态干扰；为解决这类多模态干扰问题，可以使用多个分布式系统作为信号的多模态接收器，或者通过其他信道共享传输方式避免干扰。此外，在基于6G无线多模态通信的无人机网络中，还可以利用来源分离技术，将干扰信

号和所需信号进行分离[21]，或者通过任务定向的多模态非交集映射来提高任务特定信息的可靠性，解决任务执行中其他系统的干扰和任务执行可靠性的问题，无人机通过面向任务的多模态并发信息的多模态接收、最大比合并来获取任务信息的足够补充，从而提升了多模态分集阶[22]；在排除无关模态或其他干扰信息的影响后，可以运用一种基于机会任务空间的方法完成多模态信息进行对齐和交互[23]，让6G无线多模态通信更好的支撑空天地一体化场景下具体任务的执行。

3　6G无线多模态通信的使能技术

根据通信系统的执行流程，多模态通信的关键实现技术(如图2)可以分为发送端处理、传输技术以及接收端处理3个方面，进而为云-边-端3侧任务完成提供跨域多模态信息支持，在未来6G无线多模态通信中，这些技术将构建一个统一的系统，更高效地解决复杂多模态、复杂通信场景为6G无线网络带来的新挑战。

3.1　多模态信息发送端处理

3.1.1　多模态语料库的构建

构建多模态语料库涉及多种模态的数据收集、整合和标注，包括文本、图像和音视频等多种传统信息类型和面向6G通信的新型资源标识。多模态语料库通过提供更加丰富的信息、知识和语义帮助发送端准确地表达其意思，进而协助选择多种域的多模态媒介传输信息，降低产生误解和歧义的可能性，以更好地满足未来6G网络海量异构的用户需求。

在多模态通信中，接收到环境嘈杂的信号后需要进行筛选甄别。“In-The-Wild”多模态度语料库可用于不同语言的语料库记录[24]，构建这类多模态语料库需要存储一定的数据信息，以便在需要时快速支持多模态信息传输。鉴于此，可以应用Neo4j进行存储和可视化呈现[9]，并使用语音活性检测(Voice Activity Detector, VAD)技术进行多模态信号存储[25]。存储的过程中，可以使用Turbo编码[26]、里德-所罗门码(Reed Solomon,RS)码、1/2倍速卷积编码等[27]以及低密度奇偶校验码(Low Density Parity Check code, LDPC)和极化码(polar codes)等信道编码技术[28]对从多模态获得的信息进行编码，提高下一步传输的可靠性。在存储的基础上融入预处理等计算能够更好的发挥语料库的作用，减少在接收端处理数据、纠错和处理分集的压力，并可以使用Elan等软件进行多层次多模态的数据分析[29]。

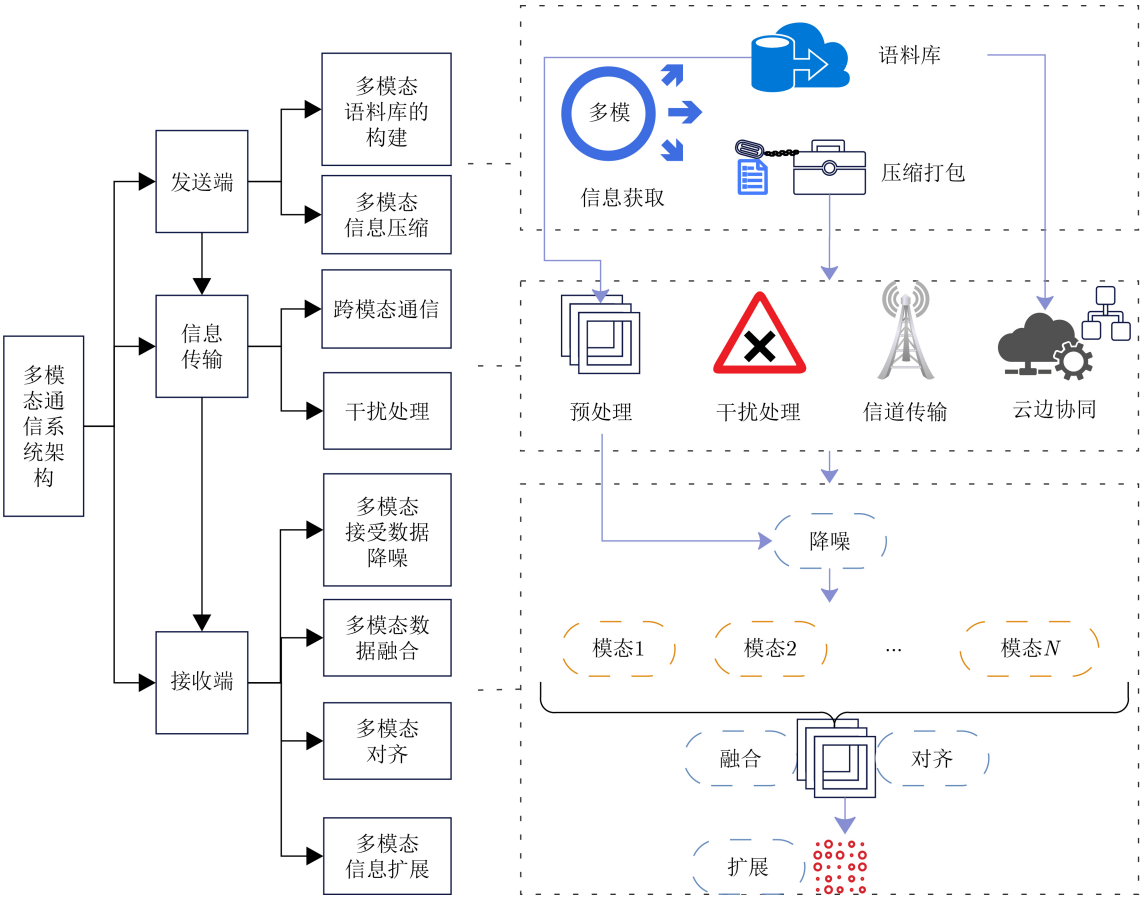

图 2 多模态通信系统的使能技术

目前国内外已经形成了一些高质量的多模态语料库，例如GEMEP, ACCORPUS, MOUD和MOSI等。这些语料库涵盖了不同表达情感、语义和知识的模态，例如文本、语音、面部表情、手势等，并且标注也日趋全面。然而，国内的多模态语料库研究仍相对较少，也存在一些如标注不规范、样本不足等的问题，尤其在面向6G通信的新数据、新场景、新问题，目前多模态语料库仍然有很大的研究空间，随着6G时代的到来，全球各地的用户连接更为紧密，多模态数据的传输速率和质量将大幅提升，人工智能的应用领域也越来越广泛，因此多模态语料库要能够包含更加丰富的跨文化数据资源，提供更高清晰度、更高保真度的数据，还要更加关注跨模态数据之间的集成与分析，以支持更复杂的多模态人工智能的应用。对此需要国内研究、研发人员积极参与构建新的语料库，并不断提高语料库的质量和完整度[30]。在这一过程中我们可以利用6G技术的高速度、低时延、高可靠性等特性，可以实时采集数据，在模态数据被采集后多模态语料库立即对其进行标注，之后进行不断的反馈与修改，以提升标注的一致性与准确性，另外利用6G的高速度和可靠性，可以实现远程协作和分布式标注，人们可以共享标注结果，从而加快标注进程。对于已采集到的数据，我们可以利用相关技术对其进行变换或者扩充，进而丰富数据样本，扩大多模态语料库的规模。可见，6G技术也为未来多模态语料库的发展提供了有力支撑。

3.1.2 多模态数据的压缩处理

在6G通感一体化获得海量信息以及空天地一体化通信场景下，使用跨域多模态对系统内外信息获取后，发送端存储将积累大量异构数据，为了减轻发送端和传输过程的双重负担，需要在发送端对多模态数据进行压缩。在压缩过程中，可以使用基于联合源-信道编码或压缩感知的方式[31]。在端到端的无线信道环境下，可以应用压缩感知技术对语音进行压缩[32]；图像可以利用深度学习技术学习图像的特征表示和压缩方法，在一定程度上提高压缩效率，同时保持图像质量；对于文字这一模态则可以应用Lempel-Ziv压缩算法，通过识别和存储重复的片段来实现压缩；视频则可以使用高效率视频编码(High Efficiency Video Coding, HEVC)的方式，通过更高效的编码方式，降低了视频传输的数据量，提高了视频质量，这些压缩技术减少了数据传输的冗余，大幅度减少对发送端存储和传输过程通信资源的占用；同时，可以降低处理复杂度、提升传输效率，并保障6G无线通信的可靠性、安全性和鲁棒性。尽管当前主流的压缩技术，均可以在6G无线多模态通信中找到适用的模态信息，然而针对无线通信架构，还可以通过信源-信道联合编码利用更多先验传输信息改进压缩算法，从而为6G无线多模态发送端技术应用提供更适配的选择。

3.2 多模态信息传输技术

在接收端处理好不同模态获得的数据后，系统需要将获得的多模态数据进行传输，使用6G无线通信的相关技术降低不同环境下的传输延迟，并克服传输过程中的多模态干扰、无线信道衰落等问题，保障接收端能高效、可靠地获得相应的信息和执行云-边-端3侧任务。

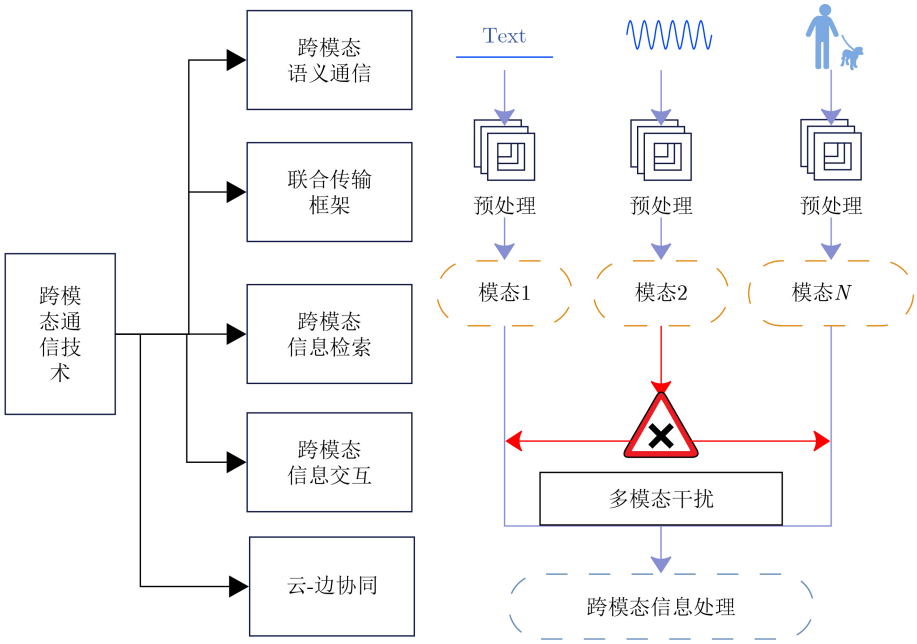

图 3 跨模态通信技术

3.2.1 跨模态通信技术

跨模态通信是指在不同的媒介或信号形式间进行信息交流的过程，其相关技术见图3。在现实生活中，人们需要通过不同的感官来感知和理解周围环境，如视频、音频、触觉等。同时，随着物联网、人工智能等技术的发展，不同类型的信息需要在不同的媒介间进行传递和交流；6G网络的通感算存一体化对跨域多模态融合提出了更高的协同处理和传输要求。因此，跨模态通信提供了多个通信渠道，可以充分利用不同类型的信息资源，在有效性方面，可以将不同模态信号相结合可以提供更全面的环境理解，提高信息的综合利用效率；在可靠性方面，通过单信息多模态并发提高信息传输的冗余度，提高了数据的可靠性和容错性，当一个模态出现问题，从其他模态仍然可以获取正确信息。

6G的应用场景广泛，涵盖全息通信、数字孪生、元宇宙和触觉互联网等，这些场景将为用户创造更沉浸的感官体验，涉及视觉、听觉和触觉等多种模态。其中，为保障多模态信息的最佳体验，可以采用基于深度学习的跨模态语义通信方法[33]，提升端到端的通信性能，同时降低人为因素的干扰。受网络复杂传输特性的制约，跨模态通信中应当调整功率，消除由延时引起的接收信号不连续性，并提高模态交叉后的整体性能，对此可以构建一个联合传输框架[34]。在6G整体性能的视角下，跨模态通信的资源分配问题不容忽视，新的多模态获取资源分配架构可以用来解决信道中无线感知资源分配、对等交互和用户保证等方面的问题。在跨模态信息检索过程中，可以使用对抗引导梯度估计哈希(Adversarial guided Gradient Estimation Hashing, AGEH)的跨模态哈希方法实现跨模态检索，有效解决不同模态间信息交互的语义鸿沟的问题[35]。在系统内部跨模态信息交互通信的过程中，通过将不同模态的数据表示在同一位数空间，自注意力机制可以很好地提取多模态数据之间的关联信息[36]。为在不同媒体类型之间进行信息关联，提供更大的数据库，可以直接测量交叉模态数据间的相似性，也可以通过整合局部相似性和设计通用特征向量和标签间的相似性，获得更全面的跨模态交互性能[37]。

当前跨模态通信架构主要基于云-边-端协同，在这里云指的是远程中央服务器，而边是指与用户接近的边缘设备，如手机和IoT设备[38]；可以整合云端-边缘更强大的计算、存储和网络通信等资源，这种基于云-边-端协同的跨模态通信方法将在算网深度融合、通算一体的6G无线网络中发挥更大的作用，例如，典型的音频-视觉-触觉重建(Audio-Visual-Haptic Reconstruction, AVHR)方法通过云-边缘深度学习模型，将跨模态数据转换为感知数据，从而为跨模态交互提供了一种更自然、更丰富的方式。该方法的跨模态通信支持在云-边-端主要分为两个阶段：第1阶段是云-边缘数据传输，主要是用于特征提取；第2阶段是共享语义学习和触觉信号生成，主要是用于重建触觉信号。在跨模态交互中，多种6G场景的复杂性带来了基于多模态信息的任务执行的复杂性，例如空天地一体化的无人机通信系统需要在复杂环境中获取多模态信息，并亟需相关多模态信息传输技术提高执行任务的成功率，通过云-边-无人机的协作可以增强系统在执行任务时对多模态信息的利用率，利用充沛的计算资源在云-边-无人机3侧更充分地发掘多模态信息的潜力，避免因任务的串行到达或实时执行引起的任务失败，进而提升整个多模态通信系统的任务执行速度和可靠性。具体地，在多模态通信传输过程中，利用感知域模态去采集大量异构信息，通过边缘计算节点或无人机直接传输到云端，运用计算域模态资源和网络中的计算媒介提升多模态信息处理效率、智能联合多域资源形成分集，灵活满足紧急救援等任务的实时决策需求、满足战场环境的高可靠和安全性需求等[22]。

3.2.2 多模态传输干扰处理技术

在多模态通信传输过程中，干扰无法避免。基于多模态信息协同的干扰处理技术，可以将跨域模态资源、媒介和信息在传输过程中协同合作、相互融合，以更全面、更有针对性地支持特定任务。在这个视角下，多模态干扰处理技术，也可以作为跨模态通信的一种实现形式，主要面临有效性即信息协同效率提升和可靠性即跨模态干扰处理两方面的挑战。

为提高无人机等通信系统在任务中获取信息的可靠性和执行效率，需要研究多模态通信传输中的多模态信息协同和干扰处理技术。最新的6G无线多模态通信系统[21]，考虑了位置相关联的信息获取、互相配合和干扰，利用多模态信息增强通信的信噪比，从多种模态中获取所有正确的任务信息，加强无人机多模态通信的可靠性[23]。由于6G空天地一体化多模态通信中会存在多种跨域传输模态(或媒介)的信号并存，何时采用哪种模态通信以及如何协同是解决问题的关键，通过多信道的传输建模，可以看到不同模态在3维运动空间中的存在的干扰域和可行域，进而辅助通信系统在灵活运动避免干扰，通过来源分离技术还可以更好地分离不同模态的干扰信号和所需模态的信号[21]。在6G无线

通信和感知两种模态融合的环节中，可以利用具有同源匹配的广义Prony方法(Generalized Prony with Homologous Matching, GPHM)和 多源恒模算法(Multi-Source Constant Modulus Algorithm, MSCMA)方法实现干扰消除和目标定位[19]，通过研究通信和定位感知所涉及的无线资源的多模态共性，让位置感知的通信信号接收合一，大幅度提高多模态通信的有效性[19]。

3.3 多模态信息接收端处理

面对接收到的海量信息，6G无线多模态信息接收端需要整合、处理来自不同模态的信息并且确保在时间和空间上的同步性，以提供一致的体验。此外，接收端通常还应具备解析和理解不同模态内容的能力，例如多模态识别、语义理解等，进而更好地为终端用户提供更丰富、交互性更强、友好的智能网联体验。因此，接收端应不仅能进行数据处理，还需要进一步提升通信质量，接收端使用技术如图4。

3.3.1 多模态接收数据降噪

接收端收到的多模态信息可能会受到各种干扰与噪声的影响，需要多模态躁降技术提高数据的质量和可用性。每种模态的数据传输都面临特定的噪声和干扰源，因此接收端需要匹配最佳的降噪方法和降噪程度，并针对任务需求进行个性化的处理。某些多模态应用需要快速响应，例如语音识别或视频会议，在接收端进行去噪可以更好地减小处理延时，以提高实时性能。6G系统致力于实现更高级别的感知和理解能力，以更全面地理解环境和用户行为，通过在多模态数据中进行降噪，可以更准确地提取有用的信息，同时也能减少数据传输中的错误，提供更加优质的通信环境，增强6G通信技术的可靠性。

针对于图像模态，可以采用基于滤波的去噪方法[10]以及密集残差去噪 U-Net (Dense Residual Denoising U-Net, DRDU-Net)算法[39]；彩色图像的降噪，可以运用基于转换假四通道图像的色彩滤波阵列(Color Filter Array, CFA)原始数据重建原始的彩色图像[40]。在多模态通信中，信号处理、图像处理、音频处理等均可以采用基于小波变换的阈值去噪方法[41]，结合该模态的领域知识，以数据驱动模型完善基于深度学习无监督随机去噪方法[42]和其他基于神经网络的方法，从而更有效和鲁棒地消除复杂结构细节噪声[43]。当多模态通信网络中出现未定义异常信号时，则可以采用基于大数据分析的通信网络异常信号去噪控制方法[44]。

在6G时代中，智能管理需要降噪技术减少数据中的干扰和误差，最大化的利用有效数据，最终实现精确管理。在无人驾驶这一领域中，通过降噪技术，可以提高传感器数据的质量，减少误判和错误决策，从而提高自动驾驶安全性和可靠性。

3.3.2 多模态对齐

多模态对齐是一种将来自不同感知模态的数据进行关联和整合的技术，通过发现不同模态描述内容的对应关系实现关联匹配，从而能够提高6G网络中智能系统环境的感知能力，对于智慧城市的建立和感知决策发挥着重要作用。不同的感知模态提供的信息类型不同，多模态对齐将这些信息整合在一起，提供更全面和更丰富的数据信息。

智能交通和城市管理、环境监测和资源管理等方面的应用中，6G系统中需要处理大规模时空数据，在时间空间对齐方面，可以运用规范化时间对齐(Canonical Time Warping, CTW)[45]将多个传感器的不同模态数据进行时间对准，并结合多种表征学习技术，实现对多模态数据的高效处理。在传统

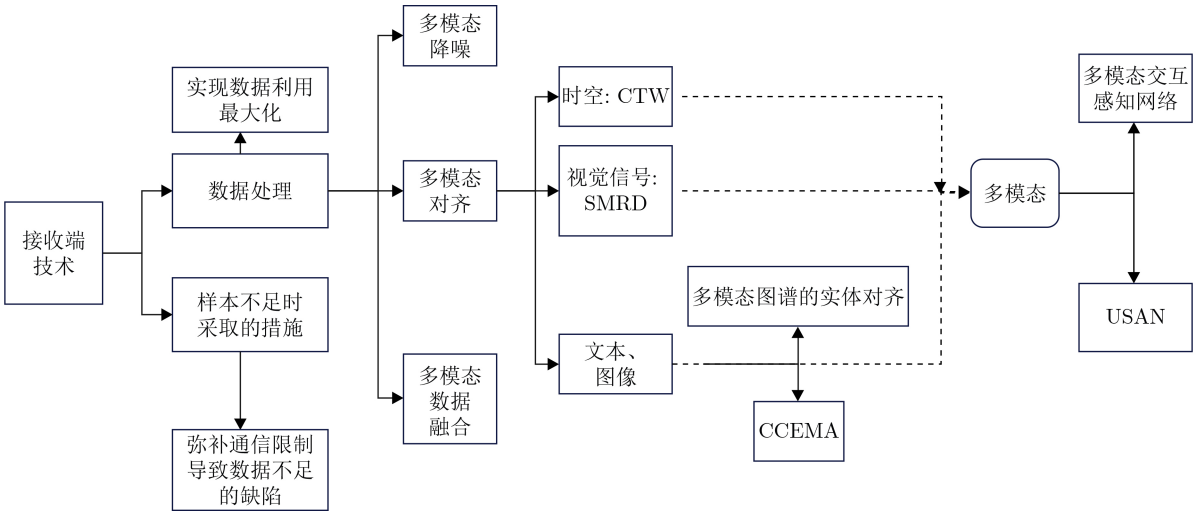

图 4 多模态通信接收端技术

的多模态通信中，视觉模态占有重要的比重，接收端需要将异构模态的视觉信号进行对齐和匹配，以便他们能够进行有效地比较、融合、分析，例如基于Shearlet的模态鲁棒描述符(Shearlet-based Modality Robust Descriptor, SMRD) [8]。对于文本和图像这两种模态的对齐，可以使用多模态知识图谱的实体对齐方法，这一方法在基于文本和图片这两种模态构建的数据集DB15K-FB15K和YAGO15K-FB15K上进行了大量工作[46]，除此之外还可以运用基于自监督的预训练语言模型中文跨模态实体对齐(Chinese Cross-Modal Entity Alignment, CCMEA)来判断图像实体和文本实体是否指代同一对象[47]。如果在接收端多个模态能够同时进行对齐，将大幅度提升整体的信息交互效率，多交互感知网络(Multi-Dynamic Aware Network, MultiDAN)可以同时提取多个模态之间的交互信息[48]，一阶段对齐网络(One-Stage Alignment Network, OSAN)则可以同时对多个不同模态的数据进行对齐操作[49]。模态对齐能够提高信息接收的丰富性，找到各模态数据之间的联系，让接收到的不同模态信息之间互为补充、相互关联，有助于提高多模态通信中信息的准确性。

3.3.3 多模态数据融合方法

多模态数据融合是将来自不同感知模态的信息整合到一起，以获得更全面、准确和丰富的数据表示的过程。一些接收端具有富裕的计算和处理资源，可以更好地适应不同类型的多模态数据。根据接收端的任务需求和环境条件进行数据融合，可以获得更好的整体性能。一些典型应用，如视频通话、VR交互，对数据传输和处理的延迟尤为敏感，在6G网络中，高速、低延迟的数据传输能力可以为多模态数据的实时传输提供充分保障，而多模态数据的融合由接收端完成则可以进一步减少延迟，提供更好的实时性能。一个模态所获取的信息可能是模糊的、片面的，多模态数据融合有利于各个模态之间互相补充或者相互增强，接收端针对不同类型的模态选择不同的处理技术和算法，可以提高各种任务的准确性和整体通信效率。

多模态特征表示将不同模态的数据有效地映射到一个共享表示空间，以便不同模态的信息可以在这个空间里进行比较和集成。图像识别、语音识别、语音信号转换、视频分析等多个领域都需要先进行多模态特征表示。从多个模态的数据源中探索有用信息时，可以使用多模态特征矢量判别框架(包括特征提取、特征融合和分类器建模3个过程)，将来自不同模态的特征融合到一个判别空间中进行特征提取，实现多模态特征表示[50]。在特征提取中，可以经过PCA操作来提取主要特征通道[51]，使用对声音信号抗干扰能力较强的梅尔倒频谱系数(Mel-Frequency Cepstral Coefficients, MFCC)，在低码率和高噪声环境下使用线性预测码，在低码率和高噪声条件下可以使用离散小波线性预测编码(Discrete Wavelet Linear Predictive Coding, DWLPC)[52]。在视听情感识别和文本图像识别任务，可以使用ILMMHA模型增强多视图特征表示学习，发挥其在信息挖掘方面的优势[53]。对于音频信息的频域特征，可以利用梅尔频率倒谱系数(MFCC)提取倒谱系数，短时傅里叶变换(Short Time Fourier Transform, STFT)可以将信号分帧，提取频谱特征，而线性预测编码(Linear Predictive Coding, LPC)使用线性预测模型对信号进行建模，提取线性预测系数作为特征。在6G无线多模态通信中，空天地一体化移动终端可以通过将面向任务的多模态交集映射到“任务空间”和“任务确定空间”来提高任务特定信息的可靠性，解决任务执行中多模态干扰等问题[23]。

接收端如果想从多个模态的数据中学习出低维度、稀疏、有判别性、可解释的特征表示，可以使用新的无监督多模式特征表示方法[54]，实现多个模态的特征表示。在多模态数据融合前可以先进行多模态数据检测识别，在多个感知模态的数据中进行模式识别和分类的任务，通过整合不同模态的信息来提高识别的准确性和性能，从而针对性地提高多模态融合效率。在特征提取后，可以利用基于BERT的文本特征提取模型和基于VGG19的图像特征提取模型，分别对两种模态的特征赋予不同的权重并进行特征融合分类，而为了提高识别的准确性和鲁棒性，过程中可以采用关于联合稀疏表示的方法[55]。多模态数据的融合需要将不同模态获得的信息关联到一起，在这里可以构建多模态深度神经网络(multimodal deep neural networks)[56]、多模态交互网络[57]，通过深度学习网络体系结构，可以处理多种不同类型的数据，并在这些数据之间进行交互和信息共享，这种交互网络可以完成多种感知模态数据的任务。在文本与图像模态的融合中，可以使用基于矩阵分解双线性池化的多模态融合检测模型[10] 或者利用多模态的深度交互图神经网络模型形成多模态交互图实现多模态信息在局部特征上的有效融合[11]。针对视觉和声音这两种模态时，可以利用多模态情感识别算法，通过融合视觉和语音的相互关联特征来提高识别性能。多模态数据融合不局限于两种模态[1]，还包括图像模型方法、基于核的方法和神经

网络方法；在非线性非高斯动态过程中模态故障的情况下，可以使用鲁棒的多模态动态数据融合，实现多种模态的数据如图像、文字、音频的动态信息交互，保持高稳健性[58]。在6G无线多模态通信中，无人机通过面向任务的多模态并发信息的多模态接收、最大比合并来获取任务信息的足够补充，从而提升多模态分集阶数[22]。

多模态数据融合的过程中面对丰富的信息可以赋予不同模态不同的权重，进行线性加权融合$D = W_1M_1 + W_2M_2 + \cdots + W_nM_n$，其中$D$为融合后获得的数据，$M_n$表示不同类型的模态，$W$代表每个模态的权重，加权融合可以更地的协调各个模态之间的信息，调取最有用的部分为任务服务通过分析异质性深度分析增强系统的性能。其中还可以[2]。多模态通信中情境可能会随时发生变化，例如噪声水平的改变等，合理地运用注意力机制可以根据情境的改变调整权重，确保系统对于变化具有适应性。如果能探索模态之间的互补和跨模态信息，可以根据信息内容更好地赋予权重进行融合。在多模态融合中还可以使用贝叶斯公式进行多模态独立成分分析，贝叶斯框架的概率模型在多模态融合中通过不同模态所收到的数据可以被视为条件概率，条件概率融合为一个综合的后验概率分布，以表示目标或事件的状态。

3.3.4 多模态信息扩充技术

在多模态通信的过程中，尽管可以通过模态间实现协同、模态内降噪，可以获取更多的可用信息，但是在6G复杂场景执行复杂任务仍然会存在信息量不足的情况，例如6G空天地一体化的无人机在林地卫星传输阻断、偏远基站直传覆盖不足等情况下需要通过自身感知域模态和已有不同模态信息获取任务执行关键信息。信息扩充技术，如小样本学习，能够在在数据总量不足的情况下充分利用当前资源，更好地理解和处理多模态通信中的信息，以获得复杂环境下更好的通信性能。

多模态通信中接收端面临的复杂环境和复杂任务具有多元性，利用元学习可以帮助系统更灵活地适应不同的任务，发现不同模态之间的关联，提高多模态系统的适用性。元学习可以让模型从之前学习到的任务中抽象出通用的知识并加以运用，从而提高模型对复杂环境的适应性[59]，应用在包括图像分类、视觉识别、自然语言处理、语音处理和元原型学习等各种任务中[60]。此外，还可以使用自适应学习知识网络 (Adaptive Learning Knowledge Networks, ALKN)学习新概念的知识，提升在样本不足的情况完成任务的可用信息总量[61]。在多种模态信息不均匀的情况下，基于迁移学习的小样本学习方法可以帮助系统在不同数据类型中进行有效的查询与检索，例如当文本信息量远大于图像时可以从文本中获取图像片段，因此迁移学习能够在多模态信息接收端发挥重要作用[62]。此外，主动学习在多模态通信中有很好的应用前景，可以帮助通信系统自动选择最有价值的多模态数据样本进行标记，减少处理跨域信息的时间和成本，通过在不同场景下主动去选择不同的数据，提高小样本数据情况下的信息可靠性，典型的技术有基于3维注意力机制和自监督预训练的少样本学习方法[63]。

4 未来6G无线多模态通信的研究展望

通过进一步研究以下关键方向，未来的6G无线多模态通信将实现更高效、可靠和智能化的通信，推动各领域的数字化和智能化发展。同时，多模态信息的发送、传输和接收使能技术的提高，也将全面提高跨域多模态的整合性、一致性和准确性，扩充6G无线多模态通信的应用维度。

(1) 在多模态方面，研究跨域多模态融合和深度融合：研究开发新的模态关联和匹配方法，提高模态间信息的综合利用效率。

(2) 在通信方面，加速6G通信技术的演化：提高多模态传输需要的无线通信能力，提高多模态所在媒介的通感算存融合能力，满足多维、高时延和高可靠性需求。

(3) 在数据技术方面，研究新型多模态数据的对齐和融合：解决多模态数据在时间和空间上的对齐问题，有效整合不同模态的信息，获得准确、全面的数据表示。

(4) 在资源管控方面，研究跨域多模态资源管理：设计合理的跨域模态资源管理方法，解决资源分布不均、模态间单位不同、资源交换评价困难和模态权重分配等问题。

5 结束语

本文通过对6G无线多模态通信技术的探讨，揭示了该领域面临的挑战、发展和解决方案。多模态通信在无线通信中具有重要的应用和发展潜力，可以更好地支持我们现在生活中不断增长繁杂的多模态信息传输。然而，要充分发挥多模态通信的优势，还需要解决跨域模态融合和深度融合、通信能力提升、多模态信息处理和应用等方面的技术难题。面对这些挑战，未来的研究可围绕数据对齐、融合和降噪方法，以及多模态信息的扩充技术展开，通过深入研究和技术创新，实现面向6G无线多模态通信的高效、可靠和智能化。希望本文对该

领域的研究提供一些启发，为构建自由连接的物理数字融合世界做出贡献。

参 考 文 献

[1] CHEN Guanghui and ZENG Xiaoping. Multi-modal emotion recognition by fusing correlation features of speech-visual[J]. *IEEE Signal Processing Letters*, 2021, 28: 533–537. doi: 10.1109/LSP.2021.3055755.

[2] LIU Shuang, DUAN Linlin, ZHANG Zhong, *et al*. Multimodal ground-based remote sensing cloud classification via learning heterogeneous deep features[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(11): 7790–7800. doi: 10.1109/TGRS.2020.2984265.

[3] CHEN Haifeng, JIANG Dongmei, and SAHLI H. Transformer encoder with multi-modal multi-head attention for continuous affect recognition[J]. *IEEE Transactions on Multimedia*, 2021, 23: 4171–4183. doi: 10.1109/TMM.2020.3037496.

[4] DENG Li, DU Xi and SHEN Ji zhong. Web page classification based on heterogeneous features and a combination of multiple classifiers[J]. *Frontiers of Information Technology & Electronic Engineering*, 2020, 21(217): 995–1004. doi: 10.1631/FITEE.1900240.

[5] FINOMORE V JR, POPIK D JR, CASTLE C JR, *et al*. Effects of a network-centric multi-modal communication tool on a communication monitoring task[J]. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2010, 54(25): 2125–2129. doi: 10.1177/154193121005402501.

[6] PORWOL L and OJO A. VR-participation: The feasibility of the virtual reality-driven multi-modal communication technology facilitating e-Participation[C]. The 11th International Conference on Theory and Practice of Electronic Governance, Galway, Ireland, 2018: 269–278. doi: 10.1145/3209415.3209515.

[7] 徐建博, 魏昕, 周亮. 面向跨模态通信的信息恢复技术[J]. 电子学报, 2022, 50(7): 1631–1642. doi: 10.12263/DZXB.20210945.

XU Jianbo, WEI Xin, and ZHOU Liang. Information recovery technology for cross-modal communications[J]. *Acta Electronica Sinica*, 2022, 50(7): 1631–1642. doi: 10.12263/DZXB.20210945.

[8] XIE Jiayu, JIN Xin, and CAO Hongkun. SMRD: A local feature descriptor for multi-modal image registration[C]. 2021 International Conference on Visual Communications and Image Processing (VCIP), Munich, Germany, 2021: 1–5. doi: 10.1109/VCIP53242.2021.9675401.

[9] GERMINIAN J F and TRICYA ESTERINA WIDAGDO ST. Utilizing postGIS extension to process spatial data stored in Neo4j database[C]. IEEE International Conference on Data and Software Engineering (ICoDSE), Toba, Indonesia, 2023: 250–255. doi: 10.1109/ICoDSE59534.2023.10291400.

[10] 王佩瑾, 闫志远, 容雪娥, 等. 数据受限条件下的多模态处理技术综述[J]. 中国图象图形学报, 2022, 27(10): 2803–2834.

WANG Peijin, YAN Zhiyuan, RONG Xuee, *et al*. Review of multimodal data processing techniques with limited data[J]. *Journal of Image and Graphics*, 2022, 27(10): 2803–2834.

[11] BHANDARI D and PAUL S. Perception net: A multimodal deep neural network for machine perception[C]. 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018: 1419–1426. doi: 10.1109/SSCI.2018.8628711.

[12] LACKEY S, BARBER D, REINERMAN L, *et al*. Defining next-generation multi-modal communication in human robot interaction[J]. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2011, 55(1): 461–464. doi: 10.1177/1071181311551095.

[13] AI Tianfang. Application of human-computer interaction technology in mobile interface design for digital media[C]. 2022 International Conference on Electronics and Devices, Computational Science (ICEDCS), Marseille, France, 2022: 152–155. doi: 10.1109/ICEDCS57360.2022.00040.

[14] SHU Beibei, SZIEBIG G, and PIETERS R. Architecture for safe human-robot collaboration: Multi-modal communication in virtual reality for efficient task execution[C]. 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), Vancouver, Canada, 2019: 2297–2302. doi: 10.1109/ISIE.2019.8781372.

[15] TODA Y and KUBOTA N. Computational intelligence for human-friendly robot partners based on multi-modal communication[C]. The 1st IEEE Global Conference on Consumer Electronics 2012, Tokyo, Japan, 2012: 309–313. doi: 10.1109/GCCE.2012.6379611.

[16] CHU Ziyan, WANG Jiacheng, JIANG Xinyue, *et al*. mind-vr: a utility approach of human-computer interaction in virtual space based on autonomous consciousness[C]. 2022 International Conference on Virtual Reality, Human-Computer Interaction and Artificial Intelligence (VRHCIAI), Changsha, China, 2022: 134–138. doi: 10.1109/VRHCIAI57205.2022.00030.

[17] WANG Dayan, QU Jue, WANG Wei, *et al*. The verification system for interface intelligent perception of human-computer interaction[C]. 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), Sanya, China, 2020: 43–48. doi: 10.1109/ICHCI51889.2020.00017.

[18] LIU Weiwei, PANG Yalong and LUAN Shenshen, *et al*.

Building a Spaceborne integrated high-performance processing and computing platform based on SpaceVPX. 2022 International Conference on Computing, Communication, Perception and Quantum Technology (CCPQT), Xiamen, China, 2022: 287-293. doi: 10.1109/CCPQT56151.2022.00056.

[19] REN Chao and LIU Luchuan. Toward full passive internet of things: Symbiotic localization and ambient backscatter communication[J]. *IEEE Internet of Things Journal*, 2023, 10(22): 19495–19506. doi: 10.1109/JIOT.2023.3262779.

[20] REN Chao, HE Zongrui, LI Yingqi, *et al*. Priority aggregation network with integrated computation and sensation for ultra dense artificial intelligence of things[J]. *IEEE Wireless Communications Letters*, 2024, 13(2): 270–273. doi: 10.1109/LWC.2023.3323421.

[21] REN Chao, LIU Luchuan, and ZHANG Haijun. Multimodal interference compatible passive UAV network based on location-aware flexibility[J]. *IEEE Wireless Communications Letters*, 2023, 12(4): 640–643. doi: 10.1109/LWC.2023.3237637.

[22] REN Chao, GONG Chao, and LIU Luchuan. Task-oriented multimodal communication based on cloud-edge-UAV collaboration[J]. *IEEE Internet of Things Journal*, 2024, 11(1): 125–136. doi: 10.1109/JIOT.2023.3295650.

[23] REN Chao, GONG Chao, CAO Difei, *et al*. Enhancing reliability in multimodal UAV communication based on opportunistic task space[J]. *IEEE Wireless Communications Letters*, 2024, 13(2): 284–287. doi: 10.1109/LWC.2023.3326130.

[24] KASHEVNIK A, LASHKOV I, AXYONOV A, *et al*. Multimodal corpus design for audio-visual speech recognition in vehicle cabin[J]. *IEEE Access*, 2021, 9: 34986–35003. doi: 10.1109/ACCESS.2021.3062752.

[25] KARPOV A, RONZHIN A, and KIPYATKOVA I. Designing a multimodal corpus of audio-visual speech using a high-speed camera[C]. 2012 IEEE 11th International Conference on Signal Processing, Beijing, China, 2012: 519–522. doi: 10.1109/ICoSP.2012.6491539.

[26] MONTORSI G and BENEDETTO S. Design of spatially coupled turbo product codes for optical communications[C]. 2021 11th International Symposium on Topics in Coding (ISTC), Montreal, Canada, 2021: 1–5. doi: 10.1109/ISTC49272.2021.9594120.

[27] ÇALIŞKAN E K, GÜLBAZ A, TENGIZLER B, *et al*. NDC-O-OFDM with channel coding for IM/DD communication systems[C]. 2022 30th Signal Processing and Communications Applications Conference (SIU), Safranbolu, Turkey, 2022: 1–4. doi: 10.1109/SIU55565.2022.9864710.

[28] KHAN M H and ZHANG Gongxuan. Evaluation of channel coding techniques for massive machine-type communication in 5G cellular network[C]. 2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP), Shanghai, China, 2020: 375–379. doi: 10.1109/ICICSP50920.2020.9232037.

[29] WANG Yue, SHI Qingwen, XU Li, *et al*. A study on the construction of traditional Chinese medicine multimodal corpus under the view of self-media[C]. 2022 International Conference on Computers, Information Processing and Advanced Education (CIPAE), Ottawa, Canada, 2022: 373–375. doi: 10.1109/CIPAE55637.2022.00084.

[30] LI Guangyan, FANG Tian, LIANG Li, *et al*. Current situation of multimodal corpora and the method of Tibetan corpus construction[C]. 2018 IEEE International Conference of Safety Produce Informatization (IICSPI), Chongqing, China, 2018: 449–453. doi: 10.1109/IICSPI.2018.8690378.

[31] YEH C H, LIN Y L, LI C C, *et al*. Compressed-and-forward: Compressive sensing for cooperative communication[C]. 2012 International Symposium on Intelligent Signal Processing and Communications Systems, Tamsui, China, 2012: 319–322. doi: 10.1109/ISPACS.2012.6473503.

[32] HANECHE H, BOUDRAA B, and OUAHABI A. Compressed sensing investigation in an end-to-end rayleigh communication system: Speech compression[C]. 2018 International Conference on Smart Communications in Network Technologies (SaCoNeT), El Oued, Algeria, 2018: 73–77. doi: 10.1109/SaCoNeT.2018.8585702.

[33] CHEN Mingkai, LIU Minghao, WANG Wenjun, *et al*. Cross-modal semantic communications in 6G[C]. 2023 IEEE/CIC International Conference on Communications in China (ICCC), Dalian, China, 2023: 1–6. doi: 10.1109/ICCC57788.2023.10233481.

[34] YANG Lianxin, WU Dan, and ZHOU Liang. Heterogeneous stream scheduling for cross-modal transmission[J]. *IEEE Transactions on Communications*, 2021, 69(9): 6037–6049. doi: 10.1109/TCOMM.2021.3086522.

[35] LU Kangkang, LIANG Meiyu, XUE Zhe, *et al*. Adversarial guided gradient estimation hashing for cross-modal retrieval[C]. 2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS), Chengdu, China, 2022: 109–113. doi: 10.1109/CCIS57298.2022.10016424.

[36] ZHAO Wenying, XU Qian, WANG Haoyu, *et al*. Cross-modal hashing for material surface properties fusion[C]. 2023 International Wireless Communications and Mobile Computing (IWCMC), Marrakesh, Morocco, 2023: 194–198. doi: 10.1109/IWCMC58020.2023.10183090.

[37] HUANG Yingying, WANG Quan, ZHANG Yipeng, *et al*. A unified perspective of multi-level cross-modal similarity for cross-modal retrieval[C]. 2022 5th International Conference on Information Communication and Signal Processing (ICICSP), Shenzhen, China, 2022: 466–471. doi: 10.1109/ICICSP55539.2022.10050678.

[38] WEI Xin, SHI Yingying, and ZHOU Liang. Haptic signal reconstruction for cross-modal communications[J]. *IEEE Transactions on Multimedia*, 2022, 24: 4514–4525. doi: 10.1109/TMM.2021.3119860.

[39] 王惠琴, 高大庆, 何永强, 等. GPR图像的数据集构建及其DRDU-Net去噪算法[J/OL]. 湖南大学学报: 自然科学版,http://kns.cnki.net/kcms/detail/43.1061.N.20231017.1626.004.html, 2023.
WANG Huiqin, GAO Daqing, HE Yongqiang, *et al*. DRDU-net based denoising algorithm for GPR image dataset[J/OL]. *Journal of Hunan University: Natural Sciences*,http://kns.cnki.net/kcms/detail/43.1061.N.20231017.1626.004.html, 2023.

[40] AKIYAMA H, TANAKA M, and OKUTOMI M. Pseudo four-channel image denoising for noisy CFA raw data[C]. 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, Canada, 2015: 4778–4782. doi: 10.1109/ICIP.2015.7351714.

[41] LI Jun, YU Menghong, and YUAN Wei. Application of wavelet new threshold denoising in ship data preprocessing and modeling[C]. 2020 Chinese Automation Congress (CAC), Shanghai, China, 2020: 1263–1267. doi: 10.1109/CAC51589.2020.9326954.

[42] WANG Feng, YANG Bo, WANG Yuqing, *et al*. Learning from noisy data: An unsupervised random denoising method for seismic data using model-based deep learning[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5913314. doi: 10.1109/TGRS.2022.3165037.

[43] TIAN Chongpeng, HONG Mei, LI Dongying, *et al*. Deep recurrent neural network for ground-penetrating radar signal denoising[C]. 2022 4th International Conference on Intelligent Information Processing (IIP), Guangzhou, China, 2022: 85-88. doi: 10.1109/IIP57348.2022.00024.

[44] GONG Jianhu. Denoising control method of abnormal signals in communication networks based on big data analysis[C]. 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Beihai, China, 2021: 329–334. doi: 10.1109/ICMTMA52658.2021.00077.

[45] LIU Yuchi, YAO Yue, WANG Zhengjie, *et al*. Generalized alignment for multimodal physiological signal learning[C]. 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019: 1–10. doi: 10.1109/IJCNN.2019.8852216.

[46] CHEN Liyi, LI Zhi and WANG Yijun, *et al*. MMEA: Entity alignment for multi-modal knowledge graph[C]. In Knowledge Science, Engineering and Management: 13th International Conference, KSEM 2020, Hangzhou, China, 2020: 134–147. doi: https://doi.org/10.1007/978-3-030-55130-8_12.

[47] 王欢, 宋丽娟, 杜方. 基于多模态知识图谱的中文跨模态实体对齐方法[J]. 计算机工程, 2023, 49(12): 88–95. doi: 10.19678/j.issn.1000-3428.0066938.
WANG Huan, SONG Lijuan, and DU Fang. Chinese cross-modal entity alignment method based on multi-modal knowledge graph[J]. *Computer Engineering*, 2023, 49(12): 88–95. doi: 10.19678/j.issn.1000-3428.0066938.

[48] 罗俊豪, 朱焱. 用于未对齐多模态语言序列情感分析的多交互感知网络[J]. 计算机应用, 2024, 44(1): 79–85. doi: 10.11772/j.issn.1001-9081.2023060815.
LUO Junhao and ZHU Yan. Multi-dynamic aware network for unaligned multimodal language sequence sentiment analysis[J]. *Journal of Computer Applications*, 2024, 44(1): 79–85. doi: 10.11772/j.issn.1001-9081.2023060815.

[49] LIU Ye, QIAO Lingfeng, LU Changchong, *et al*. OSAN: A one-stage alignment network to unify multimodal alignment and unsupervised domain adaptation[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, 2023: 3551–3560. doi: 10.1109/CVPR52729.2023.00346.

[50] GAO Lei and GUAN Ling. A discriminative vectorial framework for multi-modal feature representation[J]. *IEEE Transactions on Multimedia*, 2022, 24: 1503–1514. doi: 10.1109/TMM.2021.3066118.

[51] XUE Zhixiang, YU Xuchu, ZHANG Pengqiang, *et al*. Self-supervised feature learning and few-shot land cover classification for cross-modal remote sensing images[C]. IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022: 3600–3603. doi: 10.1109/IGARSS46834.2022.9884265.

[52] AL-HMOUZ R, DAQROUQ K, MORFEQ A, *et al*. Multimodal biometrics using multiple feature representations to speaker identification system[C]. 2015 International Conference on Information and Communication Technology Research (ICTRC), Abu Dhabi, United Arab Emirates, 2015: 314–317. doi: 10.1109/ICTRC.2015.7156485.

[53] GAO Lei and GUAN Ling. Interpretable learning-based multi-modal hashing analysis for multi-view feature representation learning[C]. 2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR), CA, USA, 2022: 47–52. doi: 10.1109/

MIPR54900.2022.00016.

[54] WANG Shiping and GUO Wenzhong. Sparse multigraph embedding for multimodal feature representation[J]. *IEEE Transactions on Multimedia*, 2017, 19(7): 1454–1466. doi: 10.1109/TMM.2017.2663324.

[55] SHEKHAR S, PATEL V M, NASRABADI N M, *et al*. Joint sparse representation for robust multimodal biometrics recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(1): 113–126. doi: 10.1109/TPAMI.2013.109.

[56] HOU Xiaofeng, XU Cheng, LIU Jiacheng, *et al*. Characterizing and understanding end-to-end multi-modal neural networks on GPUs[J]. *IEEE Computer Architecture Letters*, 2022, 21(2): 125–128. doi: 10.1109/LCA.2022.3215718.

[57] LI Shuzhen, ZHANG Tong, CHEN Bianna, *et al*. MIA-Net: Multi-modal interactive attention network for multi-modal affective analysis[J]. *IEEE Transactions on Affective Computing*, 2023, 14(4): 2796–2809. doi: 10.1109/TAFFC.2023.3259010.

[58] LIU Bin. Robust dynamic multi-modal data fusion: A model uncertainty perspective[J]. *IEEE Signal Processing Letters*, 2021, 28: 2107–2111. doi: 10.1109/LSP.2021.3117731.

[59] TIAN Pinzhuo, LI Wenbin, and GAO Yang. Consistent meta-regularization for better meta-knowledge in few-shot learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(12): 7277–7288. doi: 10.1109/TNNLS.2021.3084733.

[60] WANG Ruiqi, ZHANG Xuyao, and LIU Chenglin. Meta-prototypical learning for domain-agnostic few-shot recognition[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(11): 6990–6996. doi: 10.1109/TNNLS.2021.3083650.

[61] YAN Minghao. Adaptive learning knowledge networks for few-shot learning[J]. *IEEE Access*, 2019, 7: 119041–119051. doi: 10.1109/ACCESS.2019.2934694.

[62] WANG Kai. An overview of deep learning based small sample medical imaging classification[C]. 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML), Stanford, USA, 2021: 278–281. doi: 10.1109/CONF-SPML54095.2021.00060.

[63] LIANG Yong, CHEN Zetao, LIN Daoqian, *et al*. Three-dimension attention mechanism and self-supervised pretext task for augmenting few-shot learning[J]. *IEEE Access*, 2023, 11: 59428–59437. doi: 10.1109/ACCESS.2023.3285721.

任 超：男，讲师，研究方向为协作无线通信、空天地一体化通信、边缘计算和通感算一体化通信技术.

丁思颖：女，本科生，研究方向为智能通信技术.

张晓奇：女，博士生，研究方向为6G移动通信和人工智能技术.

张海君：男，教授，研究方向为6G移动通信、B5G行业应用、NTN网络、数字孪生和人工智能.

责任编辑：余 蓉