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。

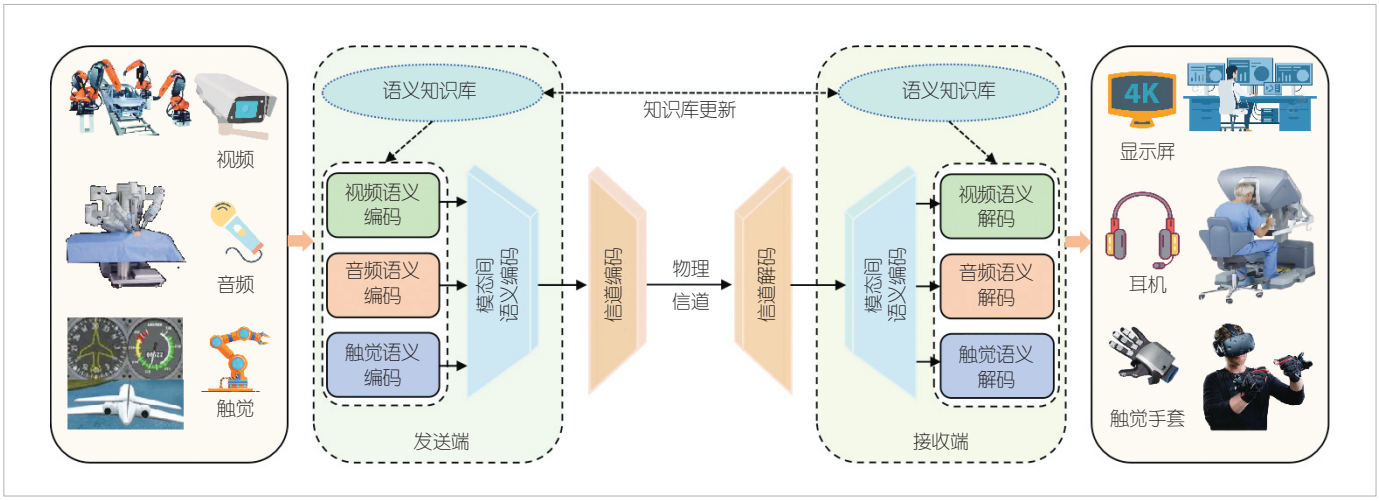
2 系统架构与关键技术

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：

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

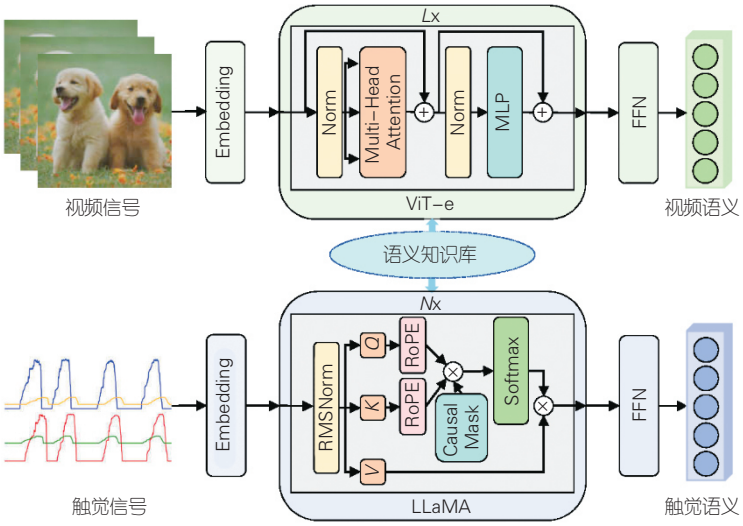
其中，$H(\hat{W}_{vah},\hat{W}_{\Delta v})$、$H(\hat{W}_{vah},\hat{W}_{\Delta a})$、$H(\hat{W}_{vah},\hat{W}_{\Delta \text{h}})$分别表示接收的语义关联与3个模态残留语义的联合语义熵，$\hat{W}_{\text{v}}$、$\hat{W}_{\text{a}}$、$\hat{W}_h$分别经过模态间语义解码后的视频、语音、触觉模态语义特征，$H(\hat{W}_v)$、$H(\hat{W}_a)$、$H(\hat{W}_{\text{h}})$分别表示解码后的各模态语义熵，l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。

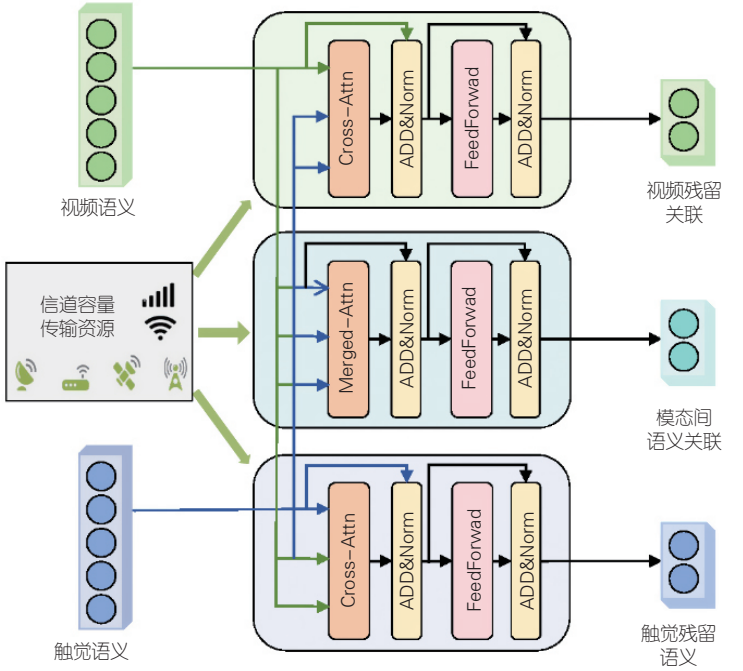
2）模态间语义编码：将视频语义特征和触觉语义特征作为输入，进一步挖掘提炼二者间的潜在关联，以获得视频—触觉语义关联以及视频残留语义和触觉残留语义。在现有研究工作中，文献[11]通过手动标注语义关系矩阵获得潜在的语义关联，文献[12]采用基于注意力机制网络获得视频和触觉模态间潜在的语义关联。鉴于上述分析，本文认为采用基于Cross-Attention的Transformer结构[15]和基于Merged-Attention的Transformer结构[16-17]可以提取视频-触觉语义关联，以及视频残留语义和触觉残留语义，如图3所示。具体而言，这两种Transformer结构的核心目标是从大量语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



FFN：前馈神经网络　MLP：多层感知机　RMS：均方根　RoPE：旋转位置嵌入

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：

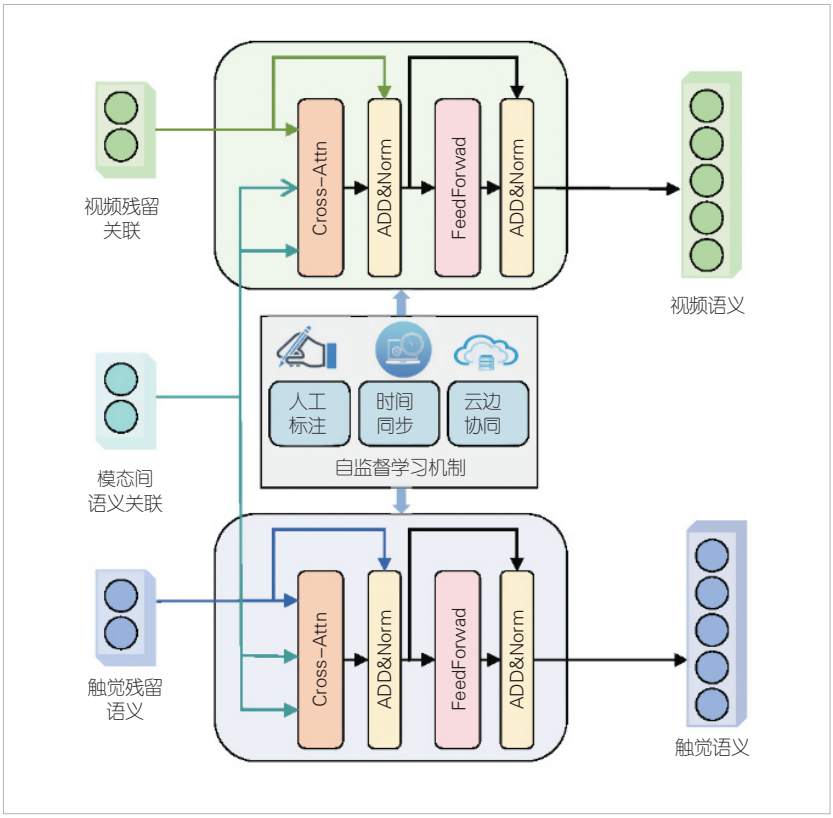
1）远程教育。在疫情期间，远程教育得到极大关注。

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

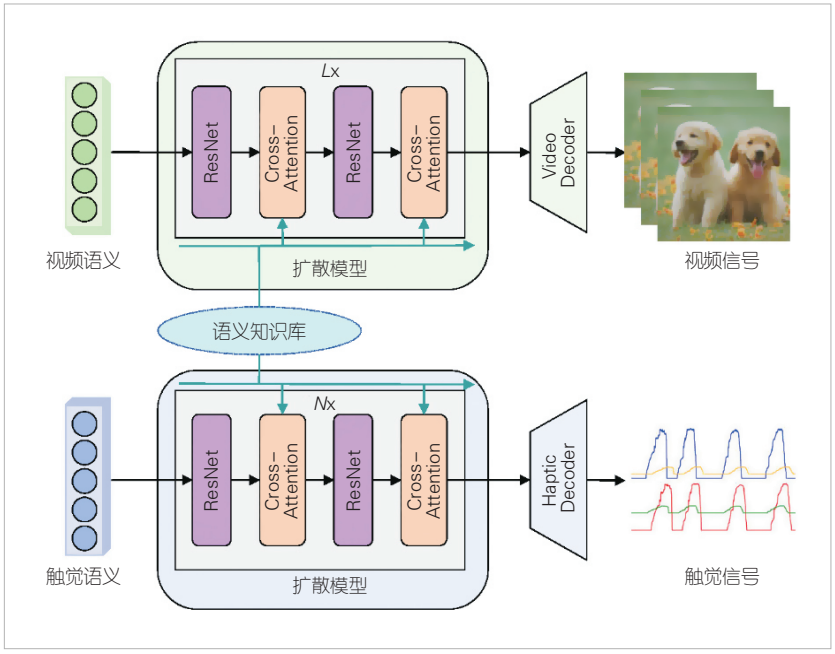
3.2 存在的挑战

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。