信息通信领域产学研合作特色期刊　十佳皖刊
第三届国家期刊奖百种重点期刊　中国科技核心期刊

ISSN 1009-6868
CN 34-1228/TN

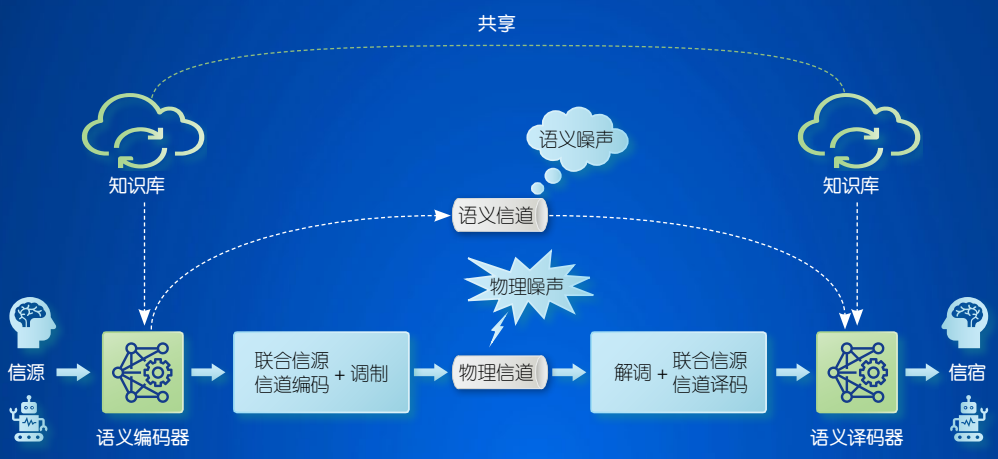
中兴通讯技术

ZTE TECHNOLOGY JOURNAL

http://tech.zte.com.cn

2023 年 4 月 · 第 2 期

专题：语义通信



（封面图片详解见 P63）

ISSN 1009-6868
9 771009 686236 04>

《中兴通讯技术》第 9 届编辑委员会成员名单

顾　问　**侯为贵**（中兴通讯股份有限公司创始人）　**钟义信**（北京邮电大学教授）
陈锡生（南京邮电大学教授）　**糜正琨**（南京邮电大学教授）

主　任　**陆建华**（中国科学院院士）

副主任　**李自学**（中兴通讯股份有限公司董事长）　**李建东**（西安电子科技大学教授）

编　委（按姓名拼音排序）

陈建平　上海交通大学教授
陈前斌　重庆邮电大学教授、副校长
段晓东　中国移动研究院副院长
葛建华　西安电子科技大学教授
管海兵　上海交通大学教授
郭　庆　哈尔滨工业大学教授
洪　伟　东南大学教授
黄宇红　中国移动研究院院长
纪越峰　北京邮电大学教授
江　涛　华中科技大学教授
蒋林涛　中国信息通信研究院科技委主任
金　石　东南大学首席教授、副校长
李尔平　浙江大学教授
李红滨　北京大学教授
李厚强　中国科学技术大学教授
李建东　西安电子科技大学教授
李乐民　中国工程院院士
李融林　华南理工大学教授
李自学　中兴通讯股份有限公司董事长
林晓东　中兴通讯股份有限公司副总裁
刘　健　中兴通讯股份有限公司高级副总裁
刘建伟　北京航空航天大学教授
隆克平　北京科技大学教授
陆建华　中国科学院院士
马建国　浙江大学教授
毛军发　中国科学院院士
孟洛明　北京邮电大学教授
石光明　鹏城实验室副主任
孙知信　南京邮电大学教授
谈振辉　北京交通大学教授
唐　宏　中国电信 IP 领域首席专家
唐雄燕　中国联通研究院副院长
陶小峰　北京邮电大学教授
王文博　北京邮电大学教授、副校长
王文东　北京邮电大学教授
王喜瑜　中兴通讯股份有限公司执行副总裁
王　翔　中兴通讯股份有限公司高级副总裁
王耀南　中国工程院院士
卫　国　中国科学技术大学教授
吴春明　浙江大学教授
邬贺铨　中国工程院院士
向际鹰　中兴通讯股份有限公司首席科学家
肖　甫　南京邮电大学教授、副校长
解冲锋　中国电信研究院教授级高工
徐安士　北京大学教授
徐子阳　中兴通讯股份有限公司总裁
续合元　中国信息通信研究院副总工
薛向阳　复旦大学教授
薛一波　清华大学教授
杨义先　北京邮电大学教授
叶　茂　电子科技大学教授
易芝玲　中国移动研究院首席科学家
张宏科　中国工程院院士
张　平　中国工程院院士
张钦宇　哈尔滨工业大学（深圳）教授、副校长
张　卫　复旦大学教授
张云勇　中国联通云南分公司总经理
赵慧玲　工业和信息化部通信科技委专职常委
郑纬民　中国工程院院士
钟章队　北京交通大学教授
周　亮　南京邮电大学教授
朱近康　中国科学技术大学教授
祝宁华　中国科学院院士

目　次

中兴通讯技术（ZHONGXING TONGXUN JISHU）
总第169期　第29卷　第2期　2023年4月

信息通信领域产学研合作特色期刊　第三届国家期刊奖百种重点期刊　中国科技核心期刊　工信部优秀科技期刊　十佳皖刊　中国五大文献数据库收录期刊　1995年创刊

热点专题 ▶

语义通信

01　专题导读 …… 陆建华，陶晓明
02　信源信道联合的新范式：语义通信 …… 吕守晔，戴金晟，张平
09　语义信息论的回顾与展望 …… 辛港涛，樊平毅
13　基于信息论的语义通信：理论与挑战 …… 施雨轩，吴泳澎，张文军
19　面向语义通信的语义知识库综述 …… 孙亚萍，崔曙光，张平
24　基于CSI反馈的语义图像传输 …… 姜培文，韩瑜，金石，李潇
29　基于深度联合信源信道编码的CSI反馈技术 …… 许佳龙，陈为，艾渤
34　语音信源的语义编码传输方法研究 …… 牛凯，姚圣时，戴金晟
40　智简语义通信的链路设计及关键技术研究 …… 孙梦颖，熊华超，王怡宁，韩书君，许晓东
46　语义通信性能评估体系及指标 …… 郑远，王凤玉，许文俊
54　基于深度学习的图像语义通信系统 …… 张振国，杨倩倩，贺诗波
62　基于语义重要度的不等错误保护数据传输机制 …… 郭帅帅，李树静

专家论坛 ▶

67　2022年度十大天线技术进展 …… 张跃平

企业视界 ▶

72　通感一体化关键技术与应用 …… 李萍，郭晓江

技术广角 ▶

79　神经辐射场加速技术综述 …… 郑清芳

《中兴通讯技术》2023年热点专题名称及策划人

1. 面向云网安全的新型防护技术
中国电信研究院教授级高工　解冲锋
北京邮电大学教授　杨义先

2. 语义通信
中国科学院院士　陆建华
清华大学教授　陶晓明

3. 数字孪生技术
重庆邮电大学教授　陈前斌

4. 算力网络和东数西算
工业和信息化部通信科技委
专职常委　赵慧玲

5. 6G网络技术
北京邮电大学教授　王文东

6. 面向双碳的新一代无线通信网络
华中科技大学教授　葛晓虎
西安电子科技大学教授　李建东

MAIN CONTENTS

ZTE TECHNOLOGY JOURNAL
Vol. 29 No. 2 Apr. 2023

Special Topic ▶ **Semantic Communications**

01 Editorial ····· LU Jianhua, TAO Xiaoming

02 New Paradigm of Joint Source and Channel Processing: Semantic Communications ····· LYU Shouye, DAI Jincheng, ZHANG Ping

09 Semantic Information Theory: Past and Future ····· XIN Gangtao, FAN Pingyi

13 Semantic Communications from Information Theoretical Perspective: Theory and Challenges ····· SHI Yuxuan, WU Yongpeng, ZHANG Wenjun

19 Survey on Semantic Knowledge Base for Semantic Communications ····· SUN Yaping, CUI Shuguang, ZHANG Ping

24 Semantic Image Transmission Based on CSI Feedback ····· JIANG Peiwen, HAN Yu, JIN Shi, LI Xiao

29 Deep Joint Source-Channel Coding Based CSI Feedback ····· XU Jialong, CHEN Wei, AI Bo

34 Semantic Coded Speech Transmission ····· NIU Kai, YAO Shengshi, DAI Jincheng

40 Link Design and Key Technologies of Intellicise Semantic Communication System ····· SUN Mengying, XIONG Huachao, WANG Yining, HAN Shujun, XU Xiaodong

46 Performance Evaluation Systems and Metrics for Semantic Communications ····· ZHENG Yuan, WANG Fengyu, XU Wenjun

54 Deep Learning-Based Image Semantic Communication System ····· ZHANG Zhenguo, YANG Qianqian, HE Shibo

62 Semantic Importance-Aware Data Transmission with Unequal Error Protection ····· GUO Shuaishuai, LI Shujing

Expert Forum ▶

67 Top Ten Antenna Technology Advances in 2022 ····· ZHANG Yueping

Enterprise View ▶

72 Key Technologies and Application of ISAC ····· LI Ping, GUO Xiaojiang

Reseach Paper ▶

79 Survey of Neural Radiance Field Acceleration Technologies ····· ZHENG Qingfang

期刊基本参数：CN 34-1228/TN*1995*b*16*86*zh*P*¥20.00*6500*15*2023-04

敬告读者

本刊享有所发表文章的版权，包括英文版、电子版、网络版和优先数字出版版权，所支付的稿酬已经包含上述各版本的费用。
未经本刊许可，不得以任何形式全文转载本刊内容；如部分引用本刊内容，须注明该内容出自本刊。

语义通信专题导读

专题策划人

陆建华，清华大学电子工程系教授，中国科学院院士，国家自然科学基金委员会党组成员、副主任，IEEE Fellow，中国电子学会会士，中国通信学会会士，中国人工智能学会会士，《中国通信（英文版）》主编，中国电子学会副理事长，国务院学位委员会信息与通信工程学科评议组召集人；主要从事空间网络、宽带无线通信、信息与信号处理等研究；取得的多项创新科技成果在国家绕月探测、大容量微波接力系统等工程中获得重要应用；受邀在多个全球学术会议上做学术报告，获国家自然科学二等奖、国家技术发明二等奖，以及“中国卫星应用杰出贡献奖”等奖项，被授予“探月工程嫦娥二号任务突出贡献者”称号；发表主要学术论文200余篇，获国家发明专利80余项。

陶晓明，清华大学电子工程系教授、博士生导师，国家杰出青年科学基金获得者，中国电子学会青年女科学家俱乐部第一届理事会轮值主席，中国电子学会通信分会副主任，中国通信学会通信理论与信号处理专业委员会副主任，中国图象图形学学会图像视频通信专业委员会副主任，《IEEE Transactions on Wireless Communications》《China Communications》《Pattern Recognition》编辑；主要从事无线多媒体通信、多媒体计算通信、语义通信等研究；取得的多项创新科技成果在国家平安乡村、智慧城市等工程中获得重要应用；受邀在IEEE Globecom等旗舰国际会议上做专题报告，获北京市科技进步一等奖、教育部科学技术进步奖一等奖、中国人工智能学会技术发明一等奖，以及中国青年科技奖、中国青年女科学家奖、科学探索奖等奖项；发表主要学术论文近200篇，获国家发明专利50余项。

现有通信系统主要解决消息符号的传输问题，难以满足未来多模态业务的多样化高效传输需求。语义通信通过对信息语义的提取、编码和传输，可大幅提升通信效率。随着新一代智能通信技术的蓬勃发展，语义通信在多智能体协作、人机交互、城市协同安防、消费电子等领域展现出广阔的应用前景，受到了全球学术界和工业界的广泛关注。

本专题主要针对语义通信所面临的关键问题和核心挑战展开讨论并提出新的研究思路，共收录论文11篇，从语义通信框架、语义信息论、语义知识库、面向多模态信源的语义–信道联合编码、语义通信性能评估等方面介绍相关研究成果。《信源信道联合的新范式：语义通信》提出了一种语义通信编码传输新框架；《语义信息论的回顾与展望》系统总结了语义通信中语义熵、语义率失真和语义信道容量的内涵和要义；《基于信息论的语义通信：理论与挑战》探讨了语义通信的理论基础，介绍了基于语义感知的信源–信道联合编码模型；《面向语义通信的语义知识库综述》梳理了语义知识库的研究现状，分析了面向语义传输的语义知识库研究面临的一系列难题；《基于CSI反馈的语义图像传输》提出了一种基于信道状态信息反馈的图像语义分割编码方案，基于子信道信噪比对所需传输的关键特征信息进行分割和编码，并根据任务需求保护相关语义特征；《基于深度联合信源信道编码CSI反馈技术》提出了一种基于信源–信道联合编码的信道状态信息反馈方法，实现对信道噪声的自适应；《语音信源的语义编码传输方法研究》总结了语音语义编码传输方法；《智简语义通信的链路设计及关键技术研究》阐述了智简语义通信的需求和挑战，探讨了智简语义通信系统的链路结构和关键技术；《语义通信性能评估体系及指标》分析研究了泛化的语义通信性能评估体系，为不同任务场景、不同模态信息下语义通信模型之间的性能对比研究提供指导；《基于深度学习的图像语义通信系统》提出了一种端到端语义特征提取和重建方法，具有较高的重建精度；《基于语义重要度的不等错误保护数据传输机制》提出了一种信源信道分离的语义通信模型，进一步提升无线通信的可靠性与有效性。

本期专题的论文作者来自知名高校与科研机构，是语义通信领域的先知先觉者。希望本期专题的内容能为有兴趣了解语义通信的读者提供有益的借鉴与启示，并在此对所有作者和专家的大力支持与指导表示由衷的感谢！

DOI：10.12142/ZTETJ.202302001
收稿日期：2023-04-10

信源信道联合的新范式：语义通信

New Paradigm of Joint Source and Channel Processing: Semantic Communications

吕守晔/LYU Shouye[1]，戴金晟/DAI Jincheng[1]，
张平/ZHANG Ping[2]
（1. 北京邮电大学泛网无线通信教育部重点实验室，中国 北京 100876；
2. 北京邮电大学网络与交换技术国家重点实验室，中国 北京 100876）
(1. The Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China;
2. The State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

DOI：10.12142/ZTETJ.202302002
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230412.1103.002.html
网络出版日期：2023-04-12
收稿日期：2023-01-25

摘要：从信源信道联合处理的角度，总结了编码传输技术的发展历程，进而引入基于非线性处理的语义通信编码传输新技术框架。该框架采用“非线性变换”和“非线性编码”两个新机制，可有效提取信源的语义特征，并与信道传输实现精准匹配，从而显著提升系统传输效率，实现通信传输从经典技术层到现代语义层的突破。

关键词：语义通信；信源信道联合；非线性变换；非线性编码；深度学习

Abstract: From the perspective of joint processing of sources and channels, the development process of coding transmission technologies is summarized, and a new technical framework of semantic communication coding and transmission based on nonlinear processing is introduced. This framework introduces two new mechanisms—nonlinear transform and nonlinear coding, which can effectively extract the semantic characteristics of a source and achieve accurate matching with channel transmission, thus significantly improving the system transmission efficiency and realizing the breakthrough of communication transmission from the classical technology layer to the modern semantic layer.

Keywords: semantic communication; joint source and channel; nonlinear transform; nonlinear coding; deep learning

过去70年，现代通信已经从香农的理论概念发展为高效实用的成熟系统。随着移动通信的快速发展，人们逐渐形成一种思维定势：系统瓶颈在于信道容量，只要增加容量就可解决大部分问题。据此，在经典通信系统中，人们采取用复杂度换取性能的方式，主要通过模块堆叠和技术密集化来解决链路传输能力不足的问题。然而，随着技术的快速进步，该模式面临着重重阻力。具体而言，经典通信系统以信源信道分离编码（SSCC）为核心，采用模块化设计，将信源与信道分别进行优化。经过长时间的探索，信源与信道编码技术已分别逼近各自的理论极限。性能的进一步提升使得复杂度剧增且收效甚微，难以可持续发展。线性处理的方式限制了各模块性能的提升。此外，模块的最优并不等同于整体性能的最优，模块化的设计将带来性能损耗。

根据著名信息学家W. WEAVER[1]的观点，当前通信系统仅停留在技术层，未涉及更高的语义层面。得益于人工智能技术与算力技术，通信系统朝着智能内生的方向不断发展。从比特到语义的范式转变，将从根本上改变当前通信系统的现状，因此人们对语义通信的研究正当时。张平院士对语义信息的特征进行了分析，指出语义信息可以用高维空间的特征参量进行表征[2]。在此基础上，张平院士提出“智简”理念[3]，以及由信源信道联合语义处理得到的新特征驱动的语义通信框架。牛凯等深入研究了从经典通信到语义通信的范式转变，提出语义通信系统的基本框架[4-5]。秦志金等探索了基于深度学习的语义通信理论、框架[6]，提出任务驱动的多用户语义通信系统[7]。戴金晟等将信源语义信息处理与通信传输技术结合，提出由语义引导的信源信道编码统一框架[8]。

本文总结了信源信道联合编码（JSCC）的发展历程，介绍了基于非线性变换JSCC的语义通信系统框架。语义通信系统引入了“非线性变换”和“非线性编码”两个新机制，将通信范式升级到语义层面，使系统能够以更低的数据传输要求提供相同的端到端服务质量。考虑到噪声等条件对实际性能的影响，语义通信系统围绕着“有限带宽”和“有

基金项目：国家自然科学基金（62293481、92067202、62071058、62001049）

损传输”两大关键特性进行设计。语义通信系统框架中主要包含两类范式：

1）直接编码：以降低系统失真为优化目标，通过非线性处理得到信源表征，并将其信源压缩为定长编码，实现固定速率的JSCC。

2）变换编码：以端到端率失真函数为优化目标，通过对非线性变换提取的数据紧致语义表征进行变分熵建模，准确估计数据的语义熵分布，从而实现变速率的JSCC。

1 JSCC的发展历程

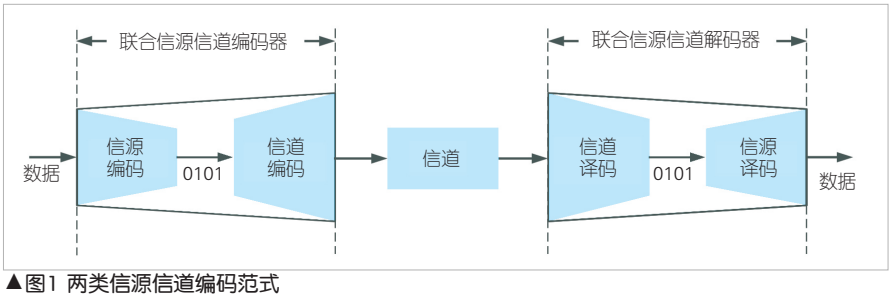
在信源信道编码定理中，香农证明：当$R(D)<C$时，存在一种信源信道编码方式使得端到端传输的失真$\leqslant D$；反之，不存在信源信道编码使端到端传输失真$\leqslant D$。根据该定理，香农给出了如图1所示的两类信源信道编码范式：

1）SSCC：将信源数据压缩为最有效表征，再将该表征编码为适合传输的符号；

2）JSCC：利用合适的编码器直接将信源数据映射为适合信道传输的符号。

理论上，当信源和信道均达到渐进均分特性（AEP）时，最优设计的信源编码与信道编码的组合等价于最优设计的信源信道联合编码，这是著名的“信源信道分离定理”。“分离定理”为后续通信系统的工程研究提供了重要指导，在理论上证明了信源压缩技术和信道传输技术分离优化的可实现性，极大降低了工程设计的复杂度。

然而，还存在着一些似乎违反“分离定理”的情况。以中文文本的传输过程为例，经典通信系统先通过最高效的信源编码得到原始文本的表征，再将所得表征通过先进的信道编码压缩后进行传输。变长码在信道编码中已广泛应用，因此当信道条件恶劣导致信道译码发生差错时，信源译码将会出现严重错误，这导致原始文本语义内容无法被理解。这种现象被称作“悬崖效应”。反之，若通过信道直接传输原始文本，我们可能会失去部分字符，但依然可以大致理解文本的含义。这种现象说明，由于避免了译码过程中的差错传播效应，在未经压缩和信道编码的条件下直接传输信源数据，反而可以较好地保留信源语义信息。其根本原因是信源的结构冗余，天然适配了信道条件。这一观察启发了人们对于信源信道联合编码技术的研究。这些复杂度和码长有限的实际场景，难以满足“分离定理”中码长不受限的假设，无法实现AEP特性。



▲图1 两类信源信道编码范式

从端到端的过程出发，研究者们围绕提升频谱利用率这一核心目标，对JSCC进行设计。联合编码充分利用端到端传输过程中的信源冗余，提升传输系统整体性能。由于历史局限，香农在1948年提出的经典信息论中[9]并未对联合编码技术的设计给出指导。很长一段时间，JSCC的研究仅停留在理论概念层面。

图2为JSCC技术的发展脉络，清晰地展示了各时期的关键技术。联合编码的范式完成了从基于显式概率模型的经典方案向面向语义的现代方案转变，实现了从线性处理到非线性处理的进步。经典方案以高效信道编码为蓝本，采用显式概率模型进行编码器的设计。受限于解析求解能力，经典方案依赖线性处理方式进行编码，无法关注到更深层次的特别是非线性的信源信息。通过人工神经网络的引入，现代方案弥补了经典方案在技术上的不足，具备了非线性处理能力。通过非线性变换，现代方案得到原始数据在语义隐空间的紧致表征，从而实现了对信源语义信息的提取与挖掘。

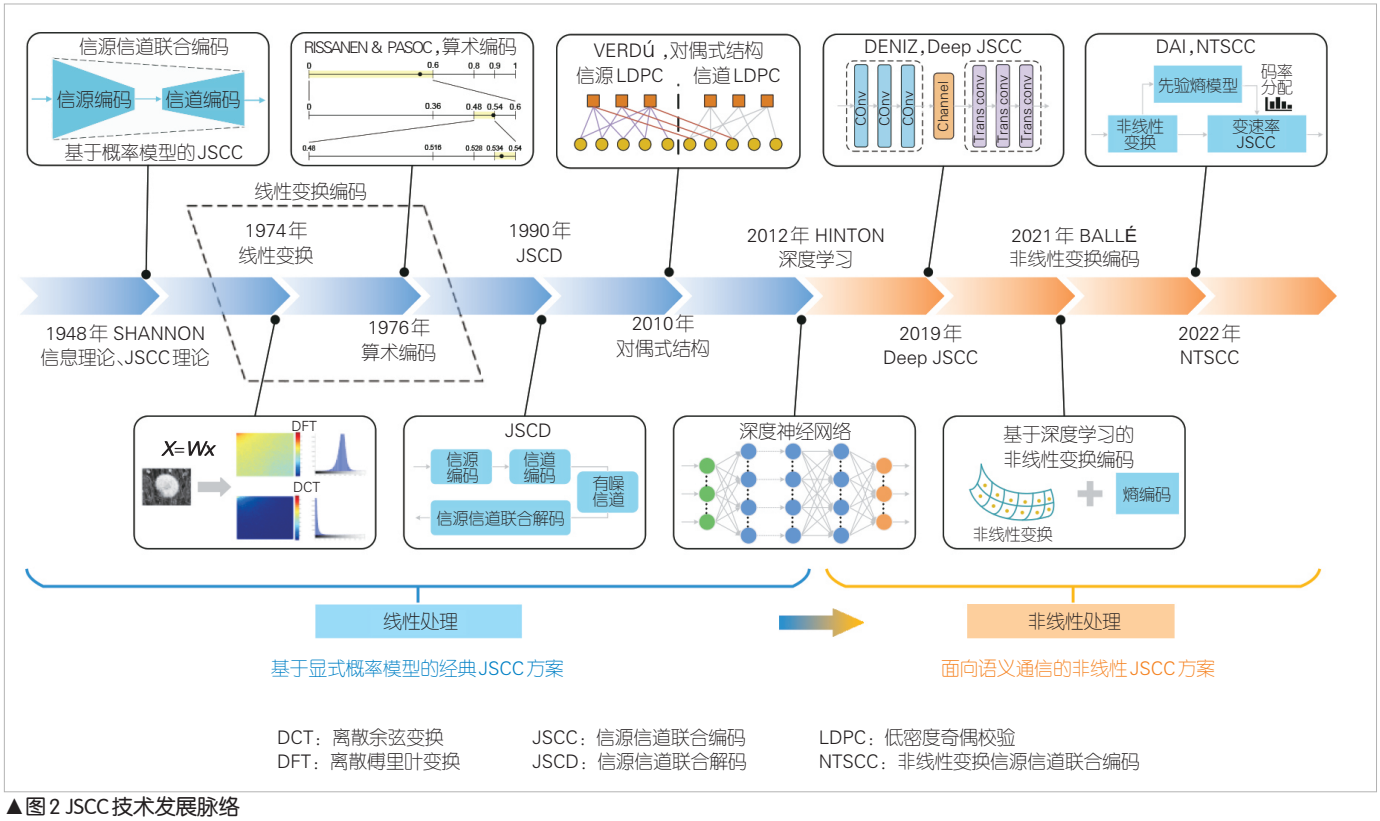
2 经典JSCC

由于技术限制，针对早期的经典通信系统，人们难以设计出性能优异的联合编码框架。20世纪90年代，因具备较高的抗干扰、抗衰落能力，以Turbo码、低密度奇偶校验码（LDPC）为代表的高效信道编码得到了学术界和工业界的广泛关注。研究者对高效信道编码进行了不断探索，从而将香农的概念设想变为实际的经典联合编码方案。本节回顾了经典信源信道联合编码中两条重要分支：信源信道联合解码与对偶式结构，并对各技术路线进行了总结。

2.1 信源信道联合解码

联合解码仍采用分离编码中模块化的设计理念，通过联合优化解码过程的方式，提升整体系统性能。在实现上，联合解码采用经典信源解码（如基于上下文的自适应算术解码）与高效信道编码级联的组合形式。

早期的联合解码使用隐马尔可夫信源与因子图结合的形式进行联合解码[10-12]。面向结构简单的信源，联合解码的方式取得了不错的性能增益。受限



▲图2 JSCC技术发展脉络

于隐马尔可夫信源的表达能力，后续研究将更加聚焦于实际信源的应用，譬如JPEG 2000[13-14]、语音信源[15]等。随着信源复杂度的提升，联合解码无法取得理想的性能增益，且会引起较大的复杂度提升。此外，联合解码并未跳出分离编码的框架，经典信源解码同高效信道编码之间并不总是匹配的，无法实现理想的端到端优化。

联合解码是早期人们对联合编码的一次尝试。联合解码没有实现真正的联合编码结构，而是对分离编码的完善和补充，因此并未摆脱分离编码所带来的影响。基于分离编码的设计结构限制了联合解码性能的进一步提升。

2.2 对偶式结构

2010年，VERDÚ等提出双低密度奇偶校验码（DLDPC）模型[16]，利用对偶式结构首次真正实现了端到端编码传输框架。

LDPC的编码过程可以看作是一个利用生成矩阵$\boldsymbol{G}$，将信源编码后的序列$\boldsymbol{b}$编码为待传输符号序列$\boldsymbol{c}$的线性映射。解码过程利用校验矩阵$\boldsymbol{H}$，将接收到的符号序列$\hat{\boldsymbol{c}}$译为序列$\hat{\boldsymbol{b}}$。从码长变化的角度，生成矩阵$\boldsymbol{G}$将短码编为了长码，校验矩阵$\boldsymbol{H}$将长码编为了短码，二者互为对偶式处理。信源编码将长码编为短码，从效果上看，与LDPC码的译码过程一致。

基于这一思想，对偶式结构创新性地使用信道编码中的对偶式处理代替经典通信系统中的信源编码，利用校验矩阵实现了信源压缩。在对偶式结构中，信源与信道编码具备统一的模型结构，能够被耦合为一个模块，这解决了联合解码信源与信道解码器不匹配的问题。

基于高效信道编码器，JIN L. Q.等还提出了信源信道联合极化（DPolar）方案[17]。通过实验对比，对偶式结构取得了更加优异的性能增益，但并未超过最先进的分离编码方案。如前文所述，经典联合编码的基础是仅具有线性处理能力的高效信道编码器，特征提取仅限于线性层面。线性处理方式导致性能提升空间受限，通信模块亟待向非线性处理方式转型，以寻求赋能通信系统长期可持续发展的新范式。

3 面向语义通信的JSCC

在经典编码的方法中，不论是基于分离式的信源信道编码，还是基于高效信道编码的经典JSCC，其编码策略都注重信源的结构状态，依靠显式概率模型进行设计。在面向人类感知和智能类机器任务时，经典编码缺少对语义信息的提取和处理，难以提升编码增益。语义编码传输将表征学习、信源编码、信道编码三者进行匹配融合，根据优化目标进行

端到端设计，实现数据高保真传输。其中，表征学习是对信源语义特征的有效提取，是语义编码传输过程中最关键的一步。信源信道联合编码实现对语义特征高效、鲁棒传输。整个链路的各个模块使用非线性设计处理方式，进一步提高语义提取和编码保护能力。

3.1 直接编码传输

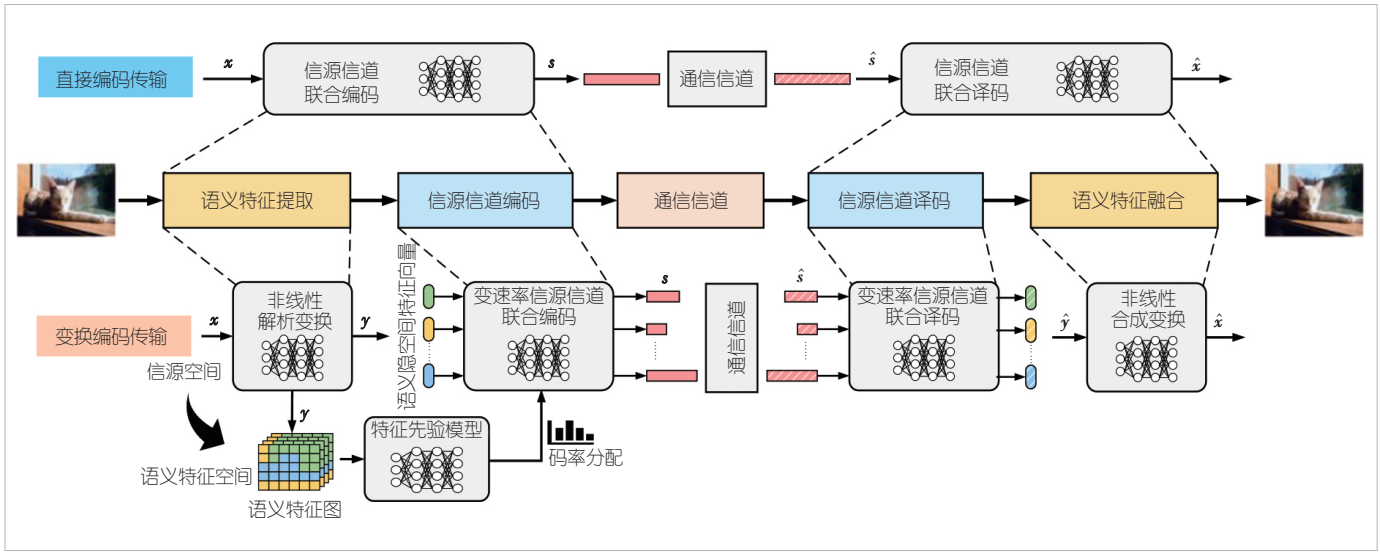
语义编码传输的一种方法是直接编码传输，如图3所示。这类编码方法在发送端将语义特征提取、信源编码和信道编码封装为编码器模块，在接收端将信源信道解码和语义特征融合封装为译码器模块，从而实现端到端的一体设计。编码器和译码器模块均可部署为深度神经网络（DNN），以便学习信源空间到编码信号空间的非线性映射。这种方法普遍被称为Deep JSCC[18–20]，本质上属于自编码器（AE）架构。考虑到对信道传输失真的影响，神经网络中间包含一个不可训练的噪声引入层，从而提高了编解码器对抗信道噪声、衰落等不利因素的能力。

基于香农信息论的思想，Deep JSCC编码实际上是高维信源矢量到低维编码传输矢量的映射。以图像信源为例，编码器f_e将输入图像$\boldsymbol{x}\in\mathbb{R}^m$，通过参数化函数$f_e(\cdot;\boldsymbol{\varphi}_f)$，映射到复值信道输入$\boldsymbol{s}\in\mathbb{R}^k$。其中，通常有$k<m$，称$\rho=k/m$为信道带宽比（CBR）。编码操作之后，JSCC序列$\boldsymbol{s}$被直接送入噪声信道，信道向传输序列引入随机误差，记为$W(\cdot;\boldsymbol{\nu})$，信道参数封装在$\boldsymbol{\nu}$中。接收序列为$\hat{\boldsymbol{s}}=W(\boldsymbol{s};\boldsymbol{\nu})$，转移概率为$p_{\hat{s}|s}$。以典型的加性高斯白噪声（AWGN）信道为例，$\hat{\boldsymbol{s}}=W(\boldsymbol{s};\boldsymbol{\sigma}_n)=\boldsymbol{s}+\boldsymbol{n}$，其中$\boldsymbol{n}\sim\mathcal{N}(0,\sigma^2\boldsymbol{I}_k)$，$\sigma^2$代表平均噪声功率。解码器$f_d$将信道输出序列$\hat{\boldsymbol{s}}$通过函数$f_d(\cdot;\boldsymbol{\phi}_f)$重建为信源矢量$\hat{\boldsymbol{x}}$。编码和解码函数通过最小化端到端失真$d(\boldsymbol{x},\hat{\boldsymbol{x}})$进行联合优化。

Deep JSCC通过自编码器框架对信源数据进行降维，这在本质上属于定长编码。在经过训练之后，编码和解码函数的编码码长k对于给定的信源样本$\boldsymbol{x}$是确定的。这种方式可以避免变长编码在译码时遇到的差错传播问题，防止在信道条件发生突变时，Deep JSCC系统发生“悬崖效应”。已有的实验结果表明，以Deep JSCC为代表的直接编码方式仅对极低分辨率图像（32×32，CIFAR10数据集）产生明显的编码增益，超出信道容量可达条件下的经典分离编码方案（特别是在低信噪比条件下）。这种性能优势的核心原因是基于深度学习的非线性编码方式可以灵活适配各种信源分布。与之相对的是，Deep JSCC仅考虑训练样本分布和失真，没有考虑信源内部在语义复杂度方面的差异。这导致编码效率低，在编码长度k、信道信噪比（SNR）逐渐增大时会表现出性能饱和。受限于单一编解码器处理能力，随着信源维度的增加，Deep JSCC的性能将会退化，甚至弱于经典分离编码方案。目前业界主流的图像、视频编码压缩方法（例如JPEG、BPG（H.265）[21]等），均针对高分辨率图像、视频设计，不面向低维度数据优化。因此，直接编码传输所带来的性能增益不具有普遍性。

3.2 变换编码传输

虽然降维可以被看作是一种简单的压缩形式，但并不等同于压缩。因为压缩的目标是在发送方和接收方共享的先验



▲图3 面向语义通信的信源信道联合编码框架

概率模型（熵模型）下尽可能降低表示的熵，而不仅仅是降维。语义编码传输的另一种方法是变换编码传输。这种方法首先将信源矢量$\boldsymbol{x}$通过非线性变换，映射到语义空间上的隐表征$\boldsymbol{y}$，之后对$\boldsymbol{y}$进行信源信道联合编码传输。由于非线性变换的解相关作用，语义隐空间上$\boldsymbol{y}$的各维之间相关性减弱，从而便于进行语义熵估计。熵估计的结果可用于指导变速率的信源信道联合编码，从而提高系统整体编码传输效率。

3.2.1 传输系统框架

如图3所示，语义非线性变换编码兼具非线性变换和JSCC两者的优势，因此被称为非线性变换信源信道联合编码（NTSCC）[22]。在编码器端，信源矢量$\boldsymbol{x}$先通过非线性解析变换$g_a(\cdot;\boldsymbol{\varphi}_g)$，产生深层语义特征图$\boldsymbol{y}$。之后，$\boldsymbol{y}$被送入信源信道联合编码器$f_e(\cdot;\boldsymbol{\varphi}_f)$，得到信道输入矢量$\boldsymbol{s}$。以AWGN信道为例，$\boldsymbol{s}$直接经过信道，通过信道传输方程$W(\cdot;\boldsymbol{\sigma}_n)$，转换为信道输出矢量$\hat{\boldsymbol{s}}$。译码器端执行相反的操作：$\hat{\boldsymbol{s}}$经过信源信道联合解码器$f_d(\cdot;\boldsymbol{\theta}_f)$首先恢复为语义特征图$\hat{\boldsymbol{y}}$，之后合成变换$g_s(\cdot;\boldsymbol{\theta}_g)$，将$\hat{\boldsymbol{y}}$重建为信源输入$\hat{\boldsymbol{x}}$。整个过程如公式（1）：

$$\boldsymbol{x} \xrightarrow{g_a(\cdot;\boldsymbol{\varphi}_g)} \boldsymbol{y} \xrightarrow{f_e(\cdot;\boldsymbol{\varphi}_f)} \boldsymbol{s} \xrightarrow{W(\cdot;\boldsymbol{\sigma}_n)} \hat{\boldsymbol{s}} \xrightarrow{f_d(\cdot;\boldsymbol{\theta}_f)} \hat{\boldsymbol{y}} \xrightarrow{g_s(\cdot;\boldsymbol{\theta}_g)} \hat{\boldsymbol{x}}。 \tag{1}$$

不同于Deep JSCC的定长编码方案，非线性变换使得语义特征$\boldsymbol{y}$的变分熵估计成为可能。此处，使用DNN构建的参数化方程$p_y(\cdot;\boldsymbol{\varphi})$来构建语义熵模型，计算得到$\boldsymbol{y}$每一个维度$y_i$值的概率。编码器$f_e$进而根据语义熵大小$-\log p_y(\cdot;\boldsymbol{\varphi})$对每个$y_i$分配编码码率，编码获得长度不等的$s_i$。因此，信道输入$\boldsymbol{s}$的总维度是依据传输数据语义内容复杂程度动态变化的，这使NTSCC实现“变速率编码传输”。系统的优化目标可以定义为：

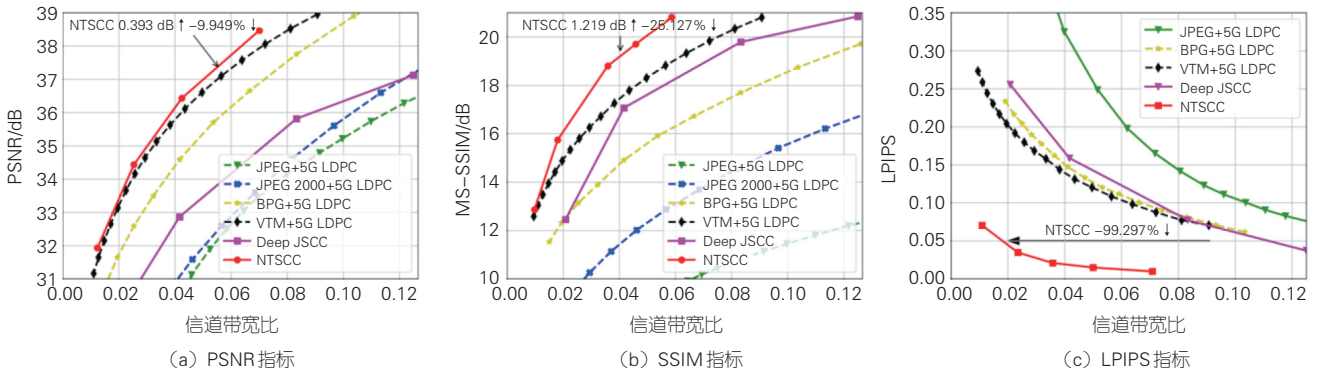
$$L_{RD} = R + \lambda D。 \tag{2}$$

公式（2）中，λ表示控制速率和失真两者权衡的超参数，R表示总信道传输速率（可定义为信道宽带比CBR，CBR=信道传输符号数/信源维度，信道传输符号数为所有s_i长度总合），D表示端到端失真，可采用不同的度量指标，比如衡量图像质量的峰值信噪比（PSNR）、多尺度结构相似度（MS-SSIM）等。为了实现语义通信中人类感知优化的目标，需要提高收端重构信源$\hat{\boldsymbol{x}}$的全局主观感知质量。我们可以在公式（2）中引入感知失真评价项，比如引入基于DNN的可学习感知图像块相似度（LPIPS）[23]和生成对抗网络的鉴别器打分度量，从而能够更好地提高主观上的端到端的传输性能。

3.2.2 系统性能

为了评估语义变换编码传输系统的性能，针对不同分辨率的开源图像数据集CIFAR10（32×32像素）、Kodak（768×512像素）和CLIC2021（2 048×1 890像素），我们在AWGN信道下使用NTSCC传输系统进行传输测试，并与Deep JSCC方案和经典分离式信源信道编码方案比较。分离式方案采用了以JPEG、JPEG 2000、BPG（H.256视频编码标准的帧内图像编码方案）和VTM（H.266视频编码标准的测试模型）为代表的图像信源编码方案，并结合实际应用的5G标准LDPC码，形成通信传输系统（分别记为“JPEG + 5G LDPC”“JPEG 2000 + 5G LDPC”“BPG + 5G LDPC”和“VTM + 5G LDPC”）。NTSCC选择从Open Images数据集中随机采样的500 000张图片作为数据集。在训练过程中，这些图片被随机裁剪成大小为256×256像素块。性能评价指标采用应用广泛的像素级度量指标（例如：PSNR和MS-SSIM）和最近兴起的基于深度学习的感知度量指标LPIPS。其中，更高的PSNR/MS-SSIM指数意味着更好的传输表现，更低的LPIPS代表更小的损失。

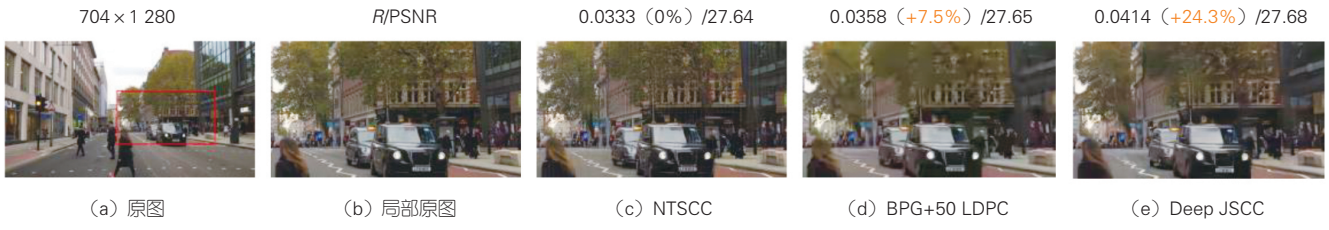
以2021年发布的包含高分辨率图像数据的具有挑战性的可学习图像压缩数据集（CLIC2021）为例，图4的实验结果表明，在信噪比为10 dB的AWGN信道上，NTSCC的失真表现为在各个信道传输速率上均优于Deep JSCC，并且两者的性能差距会随着图像分辨率和传输速率的增大而逐渐增大。特别地，在更贴近人类感知的MS-SSIM和LPIPS指标下，分离式方案普遍差于信源信道联合的语义通信方案。在PNSR和MS-SSIM指标下，图4展示了NTSCC相较于“VTM + 5G LDPC”方案的性能优势。如图4（a）、图4（b）中的箭头指示，公式的第1项表示在同等传输速率下的平均失真增益，第2项表示在同等失真下的平均传输速率节省。两个指标的计算采用了文献[24]和[25]中提出的性能评价公式。图4（c）中的语义通信模型以感知指标LPIPS进行优化，箭头上方的数字代表平均速率节省。显然，感知优化的NTSCC在性能上远优于其他方案。为了直观地展示出NTSCC的传输效果，图5选择了城市街景视频序列的传输来进行可视化结果对比。我们从这些结果中发现，即使是在相似的PSNR指标下，使用NTSCC传输的图像也能够显示出更好的视觉质量。



（a）PSNR指标　（b）SSIM指标　（c）LPIPS指标

BPG：更好的可移植图形　LDPC：低密度奇偶校验码　NTSCC：非线性变换信源信道联合编码
JPEG：联合图像专家组　LPIPS：可学习感知图像块相似度　PSNR：峰值信噪比
JSCC：信源信道联合编码　MS−SSIM：多尺度结构相似度　VTM：通用视频编码测试模型

▲图4　10 dB的加性高斯白噪声(AWGN)信道条件下，各个度量指标随信道带宽比的变化(CLIC2021)



（a）原图　（b）局部原图　（c）NTSCC　（d）BPG+50 LDPC　（e）Deep JSCC

注：（a）和（b）为原图及其局部图像，（c）—（e）显示了在SNR = 0 dB的AWGN信道上通过使用不同的传输方案重建的图像，橙色数字表示与NTSCC相比带宽成本增加的百分比

AWGN：加性高斯白噪声　JSCC：信源信道联合编码　NTSCC：非线性变换信源信道联合编码　SNR：信噪比
BPG：更好的可移植图形　LDPC：低密度奇偶校验码　RSNR：峰值信噪比

▲图5　城市街景视频序列传输的可视化效果对比

4 结束语

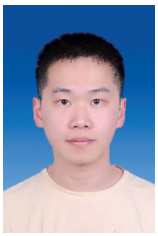
本文围绕着线性处理到非线性处理的变革，介绍了信源信道联合编码发展历程中的关键技术。我们总结了基于非线性变换信源信道联合编码的语义通信系统框架，引入了“非线性变换”和“非线性编码”两个新机制。利用这两个机制，语义通信系统框架的系统传输性能够得到有效的提升。

参考文献

[1] SHANNON C E, WEAVER W. Recent contributions to the mathematical theory of communication [M]. Champaign−Urbana: The University of Illinois Press, 1964
[2] ZHANG P, XU W, GAO H, et al. Toward wisdom−evolutionary and primitive−concise 6G: a new paradigm of semantic communication networks [J]. Engineering, 2022, 8(1): 60−73
[3] ZHANG P, XU X, DONG C, et al. Intellicise communication system: model−driven semantic communications [J]. The journal of China universities of posts and telecommunications, 2022, 29(1): 2−12. DOI: 10.19682/j.cnki.1005−8885.2022.2002
[4] NIU K, DAI J C, YAO S S, et al. A paradigm shift toward semantic communications [J]. IEEE communications magazine, 2022, 60(11): 113−119. DOI: 10.1109/MCOM.001.2200099
[5] 牛凯, 戴金晟, 张平, 等. 面向6G的语义通信 [J]. 移动通信, 2021, 45(4): 85−90. DOI: 10.3969/j.issn.1006−1010.2021.04.014
[6] QIN Z J, TAO X M, LU J H, et al. Semantic communications: principles and challenges [EB/OL]. (2021−12−30) [2023−01−20]. https://arxiv.org/abs/2201.01389
[7] XIE H Q, QIN Z J, TAO X M, et al. Task−oriented multi−user semantic communications [J]. IEEE journal on selected areas in communications, 2022, 40(9): 2584−2597. DOI: 10.1109/JSAC.2022.3191326
[8] DAI J C, ZHANG P, NIU K, et al. Communication beyond transmitting bits: semantics−guided source and channel coding [J]. IEEE wireless communications, 2022(99): 1−8. DOI: 10.1109/MWC.017.2100705
[9] 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
[10] GARCIA−FRIAS J, VILLASENOR J D. Combining hidden Markov source models and parallel concatenated codes [J]. IEEE communications letters, 1997, 1(4): 111−113. DOI: 10.1109/4234.602600
[11] GUYADER A, FABRE E, GUILLEMOT C, et al. Joint source−channel turbo decoding of entropy−coded sources [J]. IEEE journal on selected areas in communications, 2001, 19(9): 1680−1696. DOI: 10.1109/49.947033
[12] ZHU G C, ALAJAJI F. Joint source−channel turbo coding for binary Markov sources [J]. IEEE transactions on wireless communications, 2006, 5(5): 1065−1075. DOI: 10.1109/TWC.2006.1633359
[13] PAN X, CUHADAR A, BANIHASHEMI A H. Combined source and channel coding with JPEG2000 and rate−compatible low−density parity−check codes [J]. IEEE transactions on signal processing, 2006, 54(3): 1160−1164. DOI: 10.1109/TSP.2005.863032
[14] PU L L, WU Z Y, BILGIN A, et al. LDPC−based iterative joint source−channel decoding for JPEG2000 [J]. IEEE transactions on image processing, 2007, 16(2): 577−581. DOI: 10.1109/TIP.2006.888329

[15] WANG Y, QIN M H, NARAYANAN K R, et al. Joint source-channel decoding of polar codes for language-based sources [C]//Proceedings of 2016 IEEE Global Communications Conference (GLOBECOM). IEEE, 2016: 1-6. DOI: 10.1109/GLOCOM.2016.7841934
[16] FRESIA M, PERÉZ-CRUZ F, POOR H V, et al. Joint source and channel coding [J]. IEEE signal processing magazine, 2010, 27(6): 104-113. DOI: 10.1109/MSP.2010.938080
[17] JIN L Q, YANG H W. Joint source-channel polarization with side information [J]. IEEE access, 2018, 6: 7340-7349. DOI: 10.1109/ACCESS.2017.2788887
[18] FARSAD N, RAO M, GOLDSMITH A. Deep learning for joint source-channel coding of text [C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 2326-2330. DOI: 10.1109/ICASSP.2018.8461983
[19] CHOI K, TATWAWADI K, GROVER A, et al. Neural joint source-channel coding [EB/OL]. (2019-05-14) [2023-01-20]. https://arxiv. org/abs/1811.07557
[20] BOURTSOULATZE E, BURTH KURKA D, GÜNDÜZ D. Deep joint source-channel coding for wireless image transmission [C]//Proceedings of IEEE Transactions on Cognitive Communications and Networking. IEEE, 2019: 567-579. DOI: 10.1109/TCCN.2019.2919300
[21] BPG image format [EB/OL]. (2018-04-21) [2023-01-22]. https://bellard.org/bpg/
[22] 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
[23] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 586-595. DOI: 10.1109/CVPR.2018.00068
[24] BJONTEGAARD G. Calculation of average PSNR differences between RD-curves: VCEG-M33 [S]. 2001
[25] BJONTEGAARD G. Improvements of the BD-PSNR model: InITU-T SG16/Q6 [S]. 2008

作 者 简 介



吕守晔，北京邮电大学在读硕士研究生；研究方向为信源信道联合编码及语义通信。

戴金晟，北京邮电大学副教授、博士生导师；从事信息通信、人工智能相关的科研和教学工作，研究方向为智能语义通信、信源信道编码；入选“中国科协青年人才托举工程”名单，曾获中国电子学会科学技术奖自然科学一等奖；发表论文60余篇，授权国家发明专利23项，出版专著1本。

张平，北京邮电大学教授、博士生导师，中国工程院院士；现从事信息通信相关的科研和教学工作，研究方向为无线通信；曾获得国家科学技术进步奖特等奖1项、国家科学技术进步奖一等奖1项、国家技术发明奖二等奖2项、国家科学技术进步奖二等奖2项，曾获光华工程科技奖、何梁何利基金科学与技术进步奖，带领的团队入选首批“全国高校黄大年式教师团队”。

语义信息论的回顾与展望

Semantic Information Theory: Past and Future

辛港涛/XIN Gangtao，樊平毅/FAN Pingyi
（清华大学，中国 北京 100084）
（Tsinghua University，Beijing 100084，China）

DOI：10.12142/ZTETJ.202302003
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1249.004.html
网络出版日期：2023-04-11
收稿日期：2023-01-27

摘要：近年来，语义通信受到业界的广泛关注，但一套完整且有效的语义信息论框架仍未完全建立。将语义信息论分为语义熵、语义率失真、语义信道容量3个部分，分别探讨语义信息的度量问题、语义编码与失真问题、最大语义通信量问题。同时，从信息理论角度出发，系统回顾和梳理了语义信息论的相关发展，对未来语义信息论发展中可能存在的开放性问题进行了讨论。认为目前语义通信的发展处于初级阶段，仍然有很多问题尚未解决。

关键词：语义信息论；语义通信；信息论；6G

Abstract: In recent years, semantic communications has received wide attention from academia and industry, but a complete and effective framework of semantic information theory has not been fully established. The semantic information theory is divided into three parts: semantic entropy, semantic rate-distortion, and semantic channel capacity, and the measurement of semantic information, semantic coding and distortion, and the maximum semantic traffic are then discussed. At the same time, the related work in semantic information theory is sorted out, and some open problems with high probabilities about the future of semantic information theory are discussed. It is believed that the development of semantic communications is in its initial stage, and there are still many unsolved problems.

Keywords: semantic information theory; semantic communication; information theory; 6G

近年来，随着规模化无线通信的快速发展与智能处理需求的快速增多，各种基于无线通信技术的新兴智能业务得到迅速发展，这给通信技术带来了新的挑战。一方面，这些新兴业务（如工业互联网、虚拟/增强/混合现实、元宇宙、全息通信）的成功高度依赖于在海量数据集上对大型学习模型的训练。此类应用承载的大量流量可能会使现存的网络容量趋于饱和。因此，通信基础设施需要融入智能，确保在必要的时间，以合理的速率，传输所需的流量。另一方面，这些新兴业务需要极低的端到端时延，因此通信技术需要考虑流量的相关性和紧迫性，以最快、最可靠的方式提取和交付与任务相关的信息。这将导致通信网络架构从单纯追求高速率传输向追求智能化过渡[1]。

语义通信是一种全新的通信架构，它将用户对信息的需求及任务语义融入通信过程，有望大幅度提高通信实施效率，提升用户的体验，并从根本上解决基于数据的传统通信协议中存在的跨系统、跨协议、跨网络等技术难题[2]。C. E. SHANNON 和 W. WEAVER 在 1949 年发表的论文中提出了广义通信的3个层次的问题[3]，分别为：

1）技术层：传输的符号是否准确？

2）语义层：传输的符号是否准确表达了任务预期含义？

3）有效层：接收的信息含义能否按照预期方式影响用户的行为？

从技术层到语义层，通信的目标从符号的准确传输转变为语义的有效交换。经典的香农信息论建立在概率与统计的基础上，忽略了具体任务的语义表示方式，从技术层解决了两个基本问题：临界数据压缩的值和临界通信传输速率的值。语义通信的数学理论和语义的数学表征可以归结为语义信息论的问题。尽管目前语义信息论没有公认且统一的理论框架，但近些年来，学术界和工业界对语义信息论的研究越来越多。2021 年，华为公司提出的后香农时代信息通信技术（ICT）领域的十大挑战问题[4]，将语义信息论列为基础理论的首要问题。本文中，我们关注语义信息论的核心概念，介绍语义熵、语义率失真和语义信道容量方面的进展。

1 语义熵

语义信息的度量问题是语义信息理论的基础。熵是对随机变量不确定性的度量，语义熵是对语义不确定度或信息量大小的度量。到目前为止，对语义熵直观且通用的数学描述

基金项目：国家重点研发计划 [2021YFA1000500（4）]

仍然是一个具有挑战性的难题：一方面语义基本内涵不易定义从而难以度量；另一方面人们对语义的产生机理和过程并不明确[5-7]。

1）语义熵起源于前香农时代，与自然语言的分析相关[8]。针对语言任务，1952年，R. CARNAP和Y. BAR-HILLEL[9]提出了语义熵的概念，用逻辑概率来衡量一个句子所包含的语义信息量，即：

$$H_s(e) = -\log m(e), \tag{1}$$

其中，$m(e)$是指事件e在所有可能情况下为真的概率，也就是事件e的逻辑概率。可以看出，一个句子为真的逻辑概率越大，包含的信息量就越小。但是，这产生了一个悖论，即任何事实与其自身矛盾时会具有无穷大的信息量，如$H_s(e \wedge \neg e)$（表示“事件e与事件非e”发生的概率）为无穷大。2004年，L. FLORIDI[10]提出强语义信息理论，用与真实事件之间的距离来表示语义信息量的大小，解决了这一悖论。2011年，S. D'AFLONSO[11]基于真理相似性对语义信息进行了定量描述。L. FLORID和S. D'AFLONSO的方法衡量的是某一事件相对于参考事件的信息，信息值总是在0～1。然而，这些度量都依赖于参考事件的存在。从本质上讲，他们的工作提供了两个句子之间语义相似度的度量，而不是语义不确定性或信息量大小的测量。2011年，遵循R. CARNAP的定义，J. BAO、P. BASU等[12]采用命题逻辑拓展了$m(e)$的表示，进一步拓展了语言任务中语义熵的含义。

2）针对智能任务，I. D. MELAMED [13]提出了一种测量文本中单词的语义熵的方法。具体而言，对于一个词w，语义熵可以被表示为：

$$H_s(w) = H(T|w) + N(w) = -\sum_{t \in T} p(t|w)\log p(t|w) + p(NULL|w)\log F(w), \tag{2}$$

其中，$H(T|w)$代表了翻译不一致性，表示一个词被翻译成不同的方式所带来的不确定性，T表示目标词的集合，$N(w)$表示w的空链接的贡献，表示从一种语言翻译到另一种语言时遇到困难的可能性，$F(w)$是w的频率。对于分类任务，LIU X. D.、JIA W. J.等[14]引入匹配度和隶属度的概念来定义语义熵。隶属度是模糊集理论中的一个难以解析的概念，一般是根据经验来给定。定义ς为语义概念，$\mu_\varsigma(x)$为每个$x \in X$的隶属度。对于某个类别C_j，匹配度$D_j(\varsigma)$定义为：

$$D_j(\varsigma) = \frac{\sum_{x \in X_{C_j}} \mu_\varsigma(x)}{\sum_{x \in X} \mu_\varsigma(x)}。 \tag{3}$$

对于类别C_j，其语义熵定义为$H_S(\varsigma) = -D_j(\varsigma)\log D_j(\varsigma)$。将$X$中所有类的语义熵相加，可以得到集合$X$上的整体语义熵。

A. CHATTOPADHYAY等[15]提出一种量化任务相关的语义熵，他们将语义熵定义为关于数据X的语义查询的最小数量，其答案足以预测任务V，可以表示为：

$$H_S(X;\ V) \triangleq \min_E \mathrm{E}_S\left[\left|Code_Q^E(X)\right|\right],$$
$$\text{s.t. } P\left(v|Code_Q^E(x)\right) = P(v|x),\ \forall x, v, \tag{4}$$

其中，$Code_Q^E(X)$表示使用语义编码器E从x中提取的查询向量。

3）语义是对信号的理解，知识在语义的产生过程中充当了语义编码和表示的角色，发挥了重要作用。J. CHOI等[16]基于逻辑概率，从知识库的角度研究一个句子的语义熵。具体而言，记知识库为K，用$p[K\vdash q]$表示句子q相对于知识库K正确的概率，简化为$p_q = p[K\vdash q]$，则q相对于K的语义熵为：

$$H_S(q) = -p_q \log p_q + \left(1 - p_q\right)\log\left(1 - p_q\right)。 \tag{5}$$

众所周知，扩展模式可以把简单的事物组合成复杂的系统，甚至产生智能。在人类的语言系统中，主谓宾、定状补的成分组成了句子，可以表达出单个单词所不可能具有的含义。受此启发，XIN G. T.和FAN P. Y.[17]认为智能语义应当是可以扩展的，语义熵的表示形式要体现知识扩展的理念。在语义的扩展中，知识层面发生了碰撞。语义正是在知识碰撞作用下，与信号作用的产物。例如，Apple Inc.属于商业公司类别，thirteen属于数字类别，但是它们的碰撞可能会产生一个单词——iPhone，属于移动通信产品类别。用X_1和X_2表示信号，K_A^1和K_A^2表示两个知识实例，则语义熵可以表示为$H_S(X_1 \oplus X_2, K_A^1 \odot K_A^2)$，其中$\oplus$表示扩展，$\odot$表示碰撞。

4）不同于逻辑概率或统计概率模型，针对不同任务和背景，语义熵的推导形式也各不相同。A. KOLCHINSKY与D. H. WOLPERT[18]将语义信息定义为描述系统与其所处环境之间关系的语法信息。M. KOUNTOURIS和N. PAPPAS[19]使用Rényi熵对语义信息进行度量。N. J. VENHUIZEN等[20]基于背景知识的语言理解模型推导出语义熵。C. LU[21]引入广义信息论，基于贝叶斯公式、逻辑概率和模糊集合等理论对语义信息进行数学度量描述。

2 语义率失真

在香农信息论中，率失真理论解决了失真情况下的编码

问题。若信源为$X\sim p(x)$，编码后的表示为$\hat{X}\sim p(\hat{x})$，率失真度量为$d(x,\hat{x})$，则率失真函数为：

$$R(D)=\min_{p(x|\hat{x})} I(X;\hat{X})。\tag{6}$$

如果$R>R(D)$，则存在编码方法使得平均失真趋向于D。如果$R<R(D)$，则这样的码序列不存在。

在语义通信中，在语义层面进行信息提取与编码表示，有助于进一步压缩语义信息的冗余，提取最重要的语义特征，从而提高语义传输的有效性。J. LIU 等[22]将语义率失真定义为：

$$R(D_s,D_a)=\min I(X;\hat{S},\hat{X}),$$

$$\text{s.t. } E\hat{d}_s(X,\hat{S})\leqslant D_s, E\hat{d}_a(X,\hat{X})\leqslant D_a,\tag{7}$$

其中，S和$\hat{S}$为发送方和接收方对语义的理解，X和$\hat{X}$则分别是语义表示。D_s表示语义失真，D_a为表示信号失真，分别代表语义编码和信道噪声所造成的失真。通过上述简化，这个问题变为带有两个失真度量的香农率失真问题。基于此，GUO T. 等[23]从边信息的角度，对两个用户的语义率失真做了进一步的推广，可表示为：

$$R(D_1,D_2,D_s)=\min I(X_1,X_2;\hat{X}_1,\hat{X}_2,\hat{S}|Y),\tag{8}$$

其中，X_1、X_2分别是两个用户的语义表示，Y是边信息。2022 年，P. A. STAVROU 和 M. KOUNTOURIS[24]进一步研究了在汉明失真度量下的特征。

语义率失真理论直接对应编码技术，有效的语义编码工作具有高复杂性和挑战性。目前，用双失真度量的方式来表示语义编码是较普遍的做法。Y. SHAO 等[25]用语义失真L_U和语义成本D_U来定义语义编码的可达区域。语义失真反映了接收方和发送方在语义理解上的差异，语义成本则代表了信息的简洁性或可理解性，它可以是相应比特序列的长度。语义成本定义可表述为：若存在一个语义编码方案U使得$D_U=D$，$L_U=L$，则称失真–成本区域(L,D)是可达的。P. AGHELI[26]等对多用户的语义编码进行了研究，发送端向具有两个不同目标的接收端传输语义，用户根据最大化两个接收端的语义感知效用函数的加权和，分配所选数据包到达的最佳实际码字长度。Y. XIAO 等[27]提出了策略语义通信的率失真理论。该将博弈论模型与率失真理论相结合，以表征语义编码器和解码器之间的信息交互对通信失真性能的影响。

3 语义信道容量

信道容量是信息论的中心问题，也是信息论中最著名的成果。对于一个输入为X、输出为Y的离散无记忆信道，信道容量被定义为：

$$C=\max_{p(x)} I(X;Y)。\tag{9}$$

信道容量指出了在传输信息时可达速率的上界。所有小于信道容量C的码率都可在无差错条件下可达，而所有大于信道容量的码率是不可无差错达到的。也就是说，对任意的码率$R<C$，存在误差概率趋于0的一个$(2^{nR},n)$码序列；反之，如果码率$R>C$，那么误差概率将远离0。

语义信道容量的相关研究工作较少。T. OKAMOTO[28]认为语义信道容量可以表示为通过语义信道传输的语义信息的最大速率或者最大语义通信量与通信数据大小的比率。一方面，在该速率下，可以通过语义信道发送语义信息，并在语义信道的输出端恢复信息；另一方面，在信道上的语义数据信息传输速率不可能超过信道的语义容量。

2011，J. BAO 等[12]对语义信道容量做出如下定义：

$$C_s=\sup_{P(Z|X)}\{I(X;V)-H(Z|X)+\overline{H_S(V)}\}。\tag{10}$$

对任意$\epsilon>0$和$R<C_s$，存在一种编码策略，使得语义错误的最大概率小于ϵ。其中，X是信源，V是传输任务，Z是语义表示，$I(X;V)$是信源X和传输任务V之间的互信息，$H(Z|X)$表示编码的语义不确定度，$\overline{H_S(V)}$是任务V的接收消息的平均逻辑信息。

语义信道容量可以高于或低于香农信道容量$I(X;V)$，这取决于语义编码策略和接收机解释所收到消息的能力。若$\overline{H_S(V)}-H(Z|X)>0$，则意味着接收者可以处理语义歧义，接收者可以用收到的语句回答问题，语义压缩可以实现更高的传输速率。在这种情况下，C_S高于香农容量。若$\overline{H_S(V)}-H(Z|X)<0$，接收者不能解决原本存在的语义歧义，接收者无法根据收到的语句回答问题，C_S低于香农容量。

4 语义信息论展望

如果说目前语义通信处于幼儿期，那么语义信息论则处于婴儿期，仍然有大量开放性问题尚未解决，例如：

1）语义、语义熵、语义编码器是否存在通用的数学表示形式？还是可以分类表示？

2）语义通信的基本性能极限是什么？由于语义通信旨在让传输的符号精确地传达内容含义，因此语义通信的基本性能极限难以用一个具体的容量值描述，应由技术层和语义层所共同决定，其形式应体现多维特点。涉及的极限问题包括但不限于网络容量、最优编码、数据传输速率对语义传输速率的影响等，可能还需要兼顾物理层通信速率与语义解码

器表达能力之间的均衡特征。

3）基于深度学习的语义通信系统的理论框架应如何发展？目前深度学习被广泛用于信源信道联合编码和面向任务的通信结构设计，在智能处理服务中逐渐显现其重要性。探索基于深度学习的语义通信的理论框架并给出理论指导，包括其合适的性能指标与语义相似度度量等是非常必要的。

5 结束语

语义通信作为一种新的通信结构，改变了传统的数据传输范式，有望为大型智能处理服务带来新气象。然而，目前的语义通信研究刚刚起步，有大量的工作值得进一步研究，如知识共享框架、资源调度与分配、语义噪声建模与度量等。本文从语义通信的理论角度出发，依次回顾了语义熵、语义率失真和语义信道容量等相关概念，梳理了语义信息论的相关发展，同时对未来语义信息论存在的开放性问题进行了简单讨论。希望这些问题能引起专家学者的关注，助力他们为语义信息论的发展做出更多的贡献。

参考文献

[1] GÜNDÜZ D, QIN Z J, AGUERRI I E, et al. Beyond transmitting bits: context, semantics, and task-oriented communications [J]. IEEE journal on selected areas in communications, 2023, 41(1): 5-41. DOI: 10.1109/JSAC.2022.3223408

[2] 石光明, 肖泳, 李莹玉, 等. 面向万物智联的语义通信网络 [J]. 物联网学报, 2021, 5(2): 26-36. DOI: 10.11959/j.issn.2096-3750.2021.00209

[3] SHANNON C E, WEAVER W, BURKS A. The mathematical theory of communication [J]. The bell system technical journal, 1948: 379-428

[4] 徐文伟, 张弓, 白铂, 等. 后香农时代ICT领域的十大挑战问题 [J]. 中国科学: 数学, 2021, 51(7): 1095-1138. DOI: 10.1360/SSM-2021-0013

[5] 石光明, 高大化, 杨旻曦, 等. 信号的语义刻画与度量 [J]. 电子学报, 2022, 50(9): 2068-2078. DOI: 10.12263/DZXB.20210238

[6] 刘传宏, 郭彩丽, 杨洋, 等. 面向智能任务的语义通信: 理论、技术和挑战 [J]. 通信学报, 2022, 43(6): 41-57. DOI: 10.11959/j.issn.1000-436x.2022117

[7] QIN Z J, TAO X M, LU J H, et al. Semantic communications: principles and challenges [EB/OL]. [2023-01-22]. https://arxiv.org/abs/2201.01389

[8] ZHANG P, XU W, GAO H, et al. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks [J]. Engineering, 2022(8): 60-73. DOI: 10.1016/j.eng.2021.11.003

[9] CARNAP R, BAR-HILLEL Y. An outline of a theory of semantic information [EB/OL]. (1952-10-27) [2023-01-12]. https://dspace. mit. edu/handle/1721.1/4821

[10] 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

[11] D'ALFONSO S. On quantifying semantic information [J]. Information, 2011, 2(1): 61-101. DOI: 10.3390/info2010061

[12] BAO J, BASU P, DEAN M K, et al. Towards a theory of semantic communication [C]//Proceedings of 2011 IEEE Network Science Workshop. IEEE, 2011: 110-117. DOI: 10.1109/NSW.2011.6004632

[13] MELAMED I D. Measuring semantic entropy [EB/OL]. [2022-12-28]. https://www. semanticscholar. org/paper/Measuring-Semantic-Entropy-Melamed/4e107ca440ad32ccf279545a3b44228af587da42

[14] LIU X D, JIA W J, LIU W Q, et al. AFSSE: an interpretable classifier with axiomatic fuzzy set and semantic entropy [J]. IEEE transactions on fuzzy systems, 2020, 28(11): 2825-2840. DOI: 10.1109/TFUZZ.2019.2945239

[15] CHATTOPADHYAY A, HAEFFELE B D, GEMAN D, et al. Quantifying task complexity through generalized information measures [EB/OL]. (2021-01-01) [2022-01-20]. https://paperswithcode. com/paper/quantifying-task-complexity-through

[16] CHOI J, LOKE S W, PARK J. A unified view on semantic information and communication: a probabilistic logic approach [C]//Proceedings of 2022 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2022: 705-710. DOI: 10.1109/ICCWorkshops53468.2022.9814642

[17] XIN G T, FAN P Y. EXK-SC: a semantic communication model based on information framework expansion and knowledge collision [J]. Entropy (Basel, Switzerland), 2022, 24(12): 1842. DOI: 10.3390/e24121842

[18] KOLCHINSKY A, WOLPERT D H. Semantic information, autonomous agency and non-equilibrium statistical physics [J]. Interface focus, 2018, 8(6): 41. DOI: 10.1098/rsfs.2018.0041

[19] KOUNTOURIS M, PAPPAS N. Semantics-empowered communication for networked intelligent systems [J]. IEEE communications magazine, 2021, 59(6): 96-102. DOI: 10.1109/MCOM.001.2000604

[20] VENHUIZEN N J, CROCKER M W, BROUWER H. Semantic entropy in language comprehension [J]. Entropy, 2019, 21(12): 1159. DOI: 10.3390/e21121159

[21] LU C. From Bayesian inference to logical Bayesian inference: a new mathematical frame for semantic communication and machine learning [EB/OL]. [2023-01-26]. https://link. springer. com/chapter/10.1007/978-3-030-01313-4_2

[22] LIU J K, ZHANG W Y, POOR H V. A rate-distortion framework for characterizing semantic information [C]//Proceedings of 2021 IEEE International Symposium on Information Theory (ISIT). IEEE, 2021: 2894-2899. DOI: 10.1109/ISIT45174.2021.9518240

[23] GUO T, WANG Y, HAN J, et al. Semantic compression with side information: a rate-distortion perspective [EB/OL]. [2023-01-25]. https://arxiv.org/abs/2208.06094

[24] STAVROU P A, KOUNTOURIS M. A rate distortion approach to goal-oriented communication [C]//Proceedings of 2022 IEEE International Symposium on Information Theory (ISIT). IEEE, 2022: 590-595. DOI: 10.1109/ISIT50566.2022.9834593

[25] SHAO Y, CAO Q, GUNDUZ D. A theory of semantic communication [EB/OL]. [2023-01-26]. https://arxiv.org/abs/2212.01485

[26] AGHELI P, PAPPAS N, KOUNTOURIS M. Semantic source coding for two users with heterogeneous goals [C]//Proceedings of GLOBECOM 2022 - 2022 IEEE Global Communications Conference. IEEE, 2023: 4983-4988. DOI: 10.1109/GLOBECOM48099.2022.10001230

[27] XIAO Y, ZHANG X, LI Y Y, et al. Rate-distortion theory for strategic semantic communication [C]//Proceedings of 2022 IEEE Information Theory Workshop (ITW). IEEE, 2022: 279-284. DOI: 10.1109/ITW54588.2022.9965825

[28] OKAMOTO T. A unified paradigm of organized complexity and semantic information theory [EB/OL]. [2023-01-26]. https://arxiv.org/abs/1608.00941

作 者 简 介

辛港涛，清华大学在读博士研究生；主要研究领域为信息论、语义通信、图像编码。

樊平毅，清华大学教授；主要研究领域为6G无线通信网络与机器学习、语义信息论与广义信息论、大数据处理理论以及智能网络与系统检测；承担多个国家“973”“863”项目、移动专项、重点研发计划、国家自然基金和国际合作项目；已发表论文200余篇，出版学术著作4部，申请国家发明专利30余项、国际专利5项。

基于信息论的语义通信：理论与挑战

Semantic Communications from Information Theoretical Perspective: Theory and Challenges

施雨轩/SHI Yuxuan，吴泳澎/WU Yongpeng，
张文军/ZHANG Wenjun
（上海交通大学，中国 上海 200240）
(Shanghai Jiao Tong University, Shanghai 200240, China)

DOI：10.12142/ZTETJ.202302004
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1527.024.html
网络出版日期：2023-04-12
收稿日期：2023-02-23

摘要：对语义通信的发展历史与最新研究进行了概述，并基于经典香农信息论探寻了语义通信的数学建模。详细介绍了语义感知的信源信道联合编码模型及其渐进性能分析，具体包括速率失真区域、错误概率的内外界等，指出了该模型的退化情况与现有工程框架的关系。探讨了语义通信在实际应用中存在的潜在挑战与开放性问题。认为语义通信作为将用户需求与信息含义融入通信传输的全新架构，有望成为未来万物智联网络的新型基础范式。

关键词：无线通信；语义通信；香农信息论；信源信道联合编码；间接信源编码

Abstract: The development history and latest research of semantic communications are summarized, and the mathematical modeling of semantic communications based on the classical Shannon information theory is explored. The joint source channel encoding model of semantic perception and the progressive performance analysis, including the rate distortion region, and the inner outside of the error probability, are introduced in detail. The close connection between the degradation of the model and the existing engineering framework is pointed out. The potential challenges and open questions of semantic communications are then discussed. It is believed that semantic communications, as a new architecture integrating user needs and information meaning into communication transmission, is expected to become a new basic paradigm of all things intelligent connection network in the future.

Keywords: wireless communication; semantic communication; Shannon information theory; joint source-channel coding; indirect source coding

2019年底，5G移动通信网络的正式商用为人、机、物的互联开创了崭新的通信范式。人们对下一代移动通信网络的愿景，不仅体现在通信、感知、计算泛在融合的硬性需求中，还体现在人、机、物、灵（genie）互相联结的超大规模架构中[1]。“万物智联”的通信范式将通信领域的边界从物理世界拓宽到数字世界，实现高效、智能与大规模的传输网络。与此同时，新一代通信网络也给现有的网络运营与设备铺设带来极大挑战。据国际电信联盟（ITU）预测：全球移动数据流量的年增长速率将会在2030年达到55%，数据总量将会提升至现在的100倍，远超5G的体量。此外，人们多维度、多途径的通信需求会带来数据模态的多样化（语音、图像、视频）、任务需求的多样化、智能业务的多样化，这些都会成为现有架构的难点。为了解决这些难点，人们将关注点从“如何精确恢复传输的符号”转移到“如何精确恢复传输的语义”[2]。语义层传输即语义通信，是一种关注传输符号中蕴含信息的通信范式。这种通信范式能利用语义关联性去除冗余数据，并且在接收端通过背景知识库来恢复语义，以执行特定任务。需传输的抽象语义概念的数据量通常远小于精确的特征数据量，因此基于该范式的通信可以极大地提升传输效率。

语义通信的首次提出可以追溯到C. E. SHANNON与W. WEAVER的开创性的著作《The Mathematical Theory of Communication》中[2]。逻辑概率在经典信息论中对统计概率的替换，是信息论先贤们对语义信息度量的理解[3]。语义熵、语义信道容量[4]、语义反馈[5]以及背景知识[6]等相关概念相继引入，极大地推动了语义传输的理论研究。然而这些工作主要基于逻辑概率的测度，难以拓展到文本以外的应用场景。近年来，深度学习及其应用（如自然语言处理、语音识别和计算机视觉）的最新进展为多模态语义通信的实现提供

基金项目：国家重点研发计划（2018YFB1801102）；国家自然科学基金项目（62122052/62071289）；中央高校基本科研业务费专项资金项目；高等学校学科创新引智计划（BP0719010）；上海市自然科学基金项目（22DZ2229005）

了可能，同时信源信道联合编码（JSCC）[7]被认为是一种语义通信的可达方案，用于捕获和传输语义特征。基于以上方法，一系列为多模态数据[8-10]传输而开发的语义通信框架引起业界广泛的关注。上述的语义通信方案在已有的架构上取得不错的表现，但由于缺乏语义表示的通用数学模型，语义通信的理论建模与从信息论视角的解读仍然是一个开放性问题。

基于以上讨论，本文从理论模型与可实现框架的综合视角来解读语义通信，并提出一个通用的语义通信理论模型。通过将语义通信过程建模为一个考虑间接信源的JSCC问题，我们可以利用经典信息论工具来分析该模型的性能极限，并作为现实编码方案/神经网络的设计依据。在该模型中，间接信源被视为不可观测的语义信息，直接信源则被视为可被捕获的外部信息，联合信源信道编码方案用于描述间接信源编码与语义感知的信道编码的结合。在给定有限/无限块长的设定上，基于模型对错误概率、率失真函数、编码方案进行分析，对现实语义通信框架的设计有着指导意义。

1 语义通信研究进展

现阶段语义通信的工作大致分为两部分：语义信息的理论工作，如度量定义、数学建模等；基于机器学习方法的语义提取与通信框架的相关工作。本节中，围绕这两个方面，我们重点介绍语义信息与语义通信的发展历史与研究进展。

1.1 无线网络面临的挑战与语义通信的发展动机

1949 年，C. E. SHANNON 与 W. WEAVER 在《The Mathematical Theory of Communication》中正式提出了“语义层面的信息传输”的相关概念[2]。其中，W. WEAVER 将通信分为3个层级的问题：技术、语义与效率。作为旨在有噪信道上追求精确比特传输的范式，技术层级的通信已服务于面向内容交付的无线网络长达半个世纪之久。其明确的性能极限早已由香农理论给出，并被现有技术不断逼近。然而人们的通信需求却似乎永无极限：5G、6G的不断迭代、通信需求的爆炸式增长、下游任务的多样化与智能化，都是现有无线网络亟需解决的难题。作为能够解决以上问题的未来基础通信范式之一，语义通信是一种不再关注精准比特恢复而是面向任务交付的新型范式。如图1所示，针对图像识别/分类任务，图中的孩童与摩托车被抽象为依赖于下游任务的语义概念，随后编码器对语义进行编码并通过信道进行传输。在接收端通过背景知识库恢复出来的并非是原有的精确像素。但利用恢复得到的语义依然可以轻松地完成图像识别或分类任务。需要注意的是，语义通信在很大程度上依赖于译码端的背景知识，因此可能会出现无碍于任务执行的比特层面的误差，例如收发两端摩托车的样式等。由于抽象出语义信息的数量级远小于传统通信，因此语义通信可以极大地提升通信效率。

1.2 语义信息的度量定义与数学建模

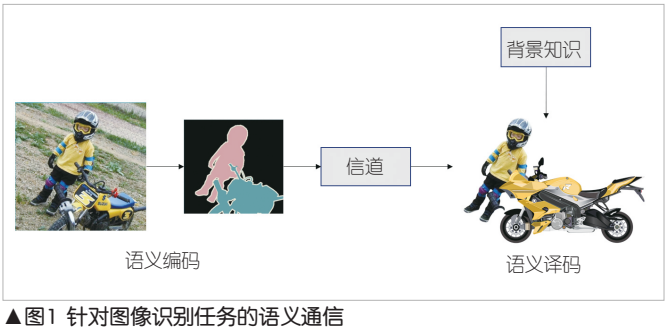
对语义信息进行刻画是语义通信中不可或缺的环节。不同语义信息的度量与定义，会对数据的语义提取、编码方案以及下游任务的执行等，产生重要的影响。

1.2.1 语义信息刻画及度量

在前香农时代，人们更多关注的是基于自然语言的信息度量。1953 年，Y. BAR-HILLEL 和 R. CARNAP[3]审视了 W. WEAVER 的工作，发现了其中遗漏的语义问题，并基于逻辑概率的测度初步定义了语义信息。所谓逻辑概率，即事件为“真”的概率越大，则该事件的熵值就越小。直到2004年，针对这种概率测度，L. FLORIDI[6]发现了“Carnap-Bar-Hillel矛盾”，即基于逻辑概率定义的熵会导致任意事件与其补事件交集的混乱度变得无穷大。同时，他提出应该用“距离”这种非负的度量来衡量语义信息的信息量。此后，学者们更为关注语义信息的本质，而不再拘泥于自然语言层面的问题。2011 年 J. BAO[4]在《Towards a Theory of Semantic Communication》一文中拓展语义相关的通信因素，阐明了语义噪声和语义信道的概念，并提出了一个语义通信框架来最小化语义错误。在此框架中，语义信道的容量刻画为：

$$C_S = \sup_{P(X|W)} \left\{ I(X;Y) - H(W|X) + \overline{H_S(Y)} \right\}, \tag{1}$$

其中，条件熵 $H(W|X)$ 为码字中包含的语义不确定性，例如码字“苹果”蕴含了“水果”与“品牌”两种意思；$\overline{H_S(Y)}$ 为接收到码字基于逻辑概率的平均信息量。这篇启发性的文章促进了更多相关理论的产生。2020年，基于Rényi熵，M. KOUNTOURIS 和 N. PAPPAS[11]提出了通信系统中语义感知的信息度量方式；2021年，A. CHATTOPADHYAY[12]等为了量

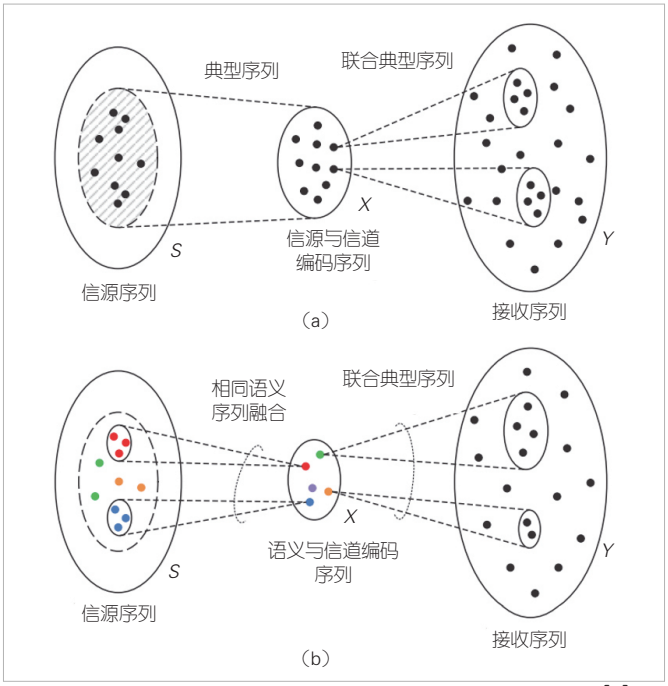


▲图1 针对图像识别任务的语义通信

化语义任务的复杂度，提出了用特定任务中最少需要回答的问题数量而非最少比特数（即熵）来刻画复杂度。

1.2.2 语义信息的理论建模

随着语义信息理论的发展，人们开始基于经典的香农信息论来研究语义建模。2021年，LIU J. K.等[13]基于经典的间接信源编码模型来刻画语义感知的率失真函数。其中，语义信息被建模为间接信源，表征信息被建模为直接信源。与传统信源编码不同，该模型同时对语义信源与表征信息的重建有失真度的要求。2022年，ZHANG P.[1]提出用“语义基”（Seb）的概念作为语义信息的表征框架，这样能够更好地描述发送数据的应用意图与不同模态；在文献[14]中，NIU K.等从联合典型的角度考虑了经典编码方案与语义编码方案的异同。与经典编码中信源序列S与码字序列X映射的唯一性不同，语义编码中不同的信源序列有概率映射到同一个码字上。这是由于针对特定任务，不同的信源序列可能拥有相近的“语义”，即在语义空间毗邻。图2[14]很好地解释了语义编码能够拥有更小码本空间的原因。由于背景知识库的存在，即使带有歧义的映射也能完成特定任务的无损传输，这也是语义通信的优势所在。文献[15]将语义信息与神经网络的隐含层变量联系起来，根据信息瓶颈理论将语义损失刻画到神经网络的损失函数中，以提高图像的传输表现。其基本原理是基于变分自编码器（VAE）中的证据下界，构建辅助分布用于近似较难计算的语义分布。



▲图2 经典编码传输与语义编码传输的对比(从序列典型性角度)[14]

1.2.3 基于机器学习方法的语义提取与通信框架

现代语义通信致力于建立完整、统一的数学表述。在此过程中，涌现大量语义感知的工程可实现框架[8-11]。接下来，我们分别从数据提取与通信设计两个角度对这些工作进行介绍。

1）多模态数据的语义提取

基于深度学习的语义提取模块，使得语义通信突破文本的桎梏，广泛活跃于语音、图片、视频等多模态的数据传输场景中。WENG Z. Z.等[8]提出了适用于动态信道的语音传输框架，其中利用自注意力机制的语义感知网络可以最小化语音传输中的语义错误；针对图片的传输，E. BOURTSOULATZE[7]等首次提出了基于深度学习的信源信道联合编码（D-JSCC）方案。该方案相当于在卷积自编码器中插入了一层不可训练的信道参数层，很快被证明对语义信息的提取有较好的帮助。例如，在JIANG P. W.等的工作[9]中，基于Transformer的D-JSCC模块被用于视频会议的语义传输，其任务需求是在恢复端恢复出表情保真的人脸信息。语义信息被理解为视频流帧间的残差信息，在译码端根据初始帧进行恢复。这种基于语义的视频处理方式能够较大地减少通信开销，并能确保下游任务一定的精确度。

2）语义感知的通信设计

语义信息能够对多模态数据语义进行处理，还可以与经典通信技术相结合。基于语义的混合自动重传请求（SR-HARQ），在文献[9]中用于克服信道状态的动态变化，并获得了优于传统HARQ的性能表现；DAI J. C.[10]等利用基于高斯噪声的非线性变换实现自适应语义通信框架，并可根据不同信道情况进行速率分配，以达到在无线信道中较好的传输效果。此外，应用于物联网（IoT）的云网络架构场景与后5G大规模连接场景中的相关框架也有优异的表现。

3）本地背景知识库设计

在语义通信的场景中，一个优异的本地知识库是收发端进行语义提取和生成的重要因素。关于本地背景知识库的工作目前仍处于探索阶段。XIA L.[16]等从理论角度讨论了本地知识库匹配问题对语义通信中用户联合与带宽分配的影响。S. KADAM[17]等针对不同场景进行定制化的背景知识库的收集和获取，也就是说收发端仅在该案例下完成对知识库的利用。S. KADAM等在工作中考虑足球赛事解说的场景并构建语义知识库，显著提高传输效率。此外，聊天生成式预训练转换器（ChatGPT）[18]使用参数量巨大的模型和海量的优质原始数据，预训练出一个能够胜任各种场景、各种问题的语义交互知识库，使通用人工智能技术不再遥不可及。

业界的相关研究都验证了语义通信在6G智能网络应用

中的巨大潜力。因此，语义通信被视为实现“万物智联”愿景的重要助力。

2 语义通信理论建模

基于相关工作，我们不难发现，通用、完整的理论建模仍是语义通信领域的一大空白。为了对已有的语义通信框架提供适配的理论模型，本节中我们将着重讨论语义通信理论的数学建模。我们将点对点刻画语义信息的信源编码模型，拓展到JSCC模型，并给出特定场景的一些初步结论。

2.1 语义感知的信源编码模型

文献[10]中，LIU J. K.提出刻画语义信息的信源编码模型。如图3所示[13]，信源的语义特征被刻画为不可见的信源S，S服从分布P_S且至信源可观测外部特征X的转移概率为$P_{X|S}$。与传统间接信源编码仅要求恢复不可见信源$\hat{S}$有所不同，译码端要求同时恢复语义重建与比特重建，失真测度函数$d_s(s,\hat{s})$与$d_x(x,\hat{x})$分别对应要求的失真度约束。基于该模型，其速率–失真函数被表述为：

$$R(D_S,D_X)=\min_{P_{\hat{S},\hat{X}|X}} I(X;\hat{S},\hat{X}),$$
$$\text{s.t.}\quad \mathbb{E}_{P_{\hat{X},X}}\left[d_x(X,\hat{X})\right]\leqslant D_X,$$
$$\mathbb{E}_{P_{\hat{S},S}}\left[d_s(S,\hat{S})\right]\leqslant D_S。\tag{2}$$

在特定情形下（如S与X联合高斯分布且失真测度采用均方误差），基于该模型的速率失真函数可通过线性最小均方误差估计（LMMSE）的线性特性以及代理（surrogate）②失真测度获得。

2.2 语义感知的边信息编码模型

在实际应用场景中，上述的间接信源编码模型可以拓展为具有边信息的场景模型。

图4展示了具有边信息的语义信源编码模型。该模型在编码器与解码器端存在的背景知识库被建模为两个有限样本空间的随机变量U与V，即收发端在处理语义序列时利用两边的边信息进行编解码。考虑到背景知识的编码模型更贴近实际的语义通信模型，其速率–失真度函数$R^S(D_S,D_X)$可通过求解拓展至长马尔科夫链的Wyner–Ziv问题来获得。基于该模型的退化情形如下：

1）U与V都可用且U=V，$R^S(D_S,D_X)$退化至语义感知的条件率失真函数为

$$R_{X|V}(D_S,D_X)=\min_{P_{\hat{S},\hat{X}|X,V}} I(X;\hat{S},\hat{X}|V);\tag{3}$$

2）U与V都不可用，$R^S(D_S,D_X)$退化至$R(D_S,D_X)$；

3）U不可用，V可用，我们引入不同于码字W的辅助随机变量Q，$\hat{X}=\hat{X}(Q,V)$，$\hat{S}=\hat{S}(Q,V)$，可以得到：

$$R^S(D_S,D_X)=\min_{P_{Q|X,V}} I(X;Q|V)。\tag{4}$$

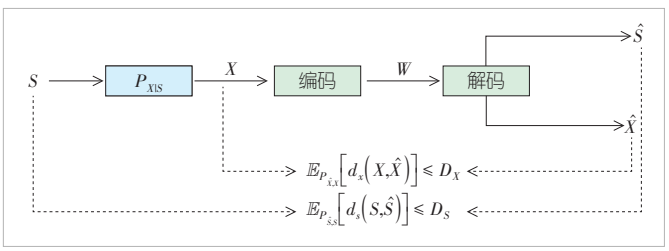
公式（4）的失真度约束同语义感知的信源编码模型（公式（2））。对于具有边信息的任意信源模型，其率失真函数的闭式表达仍是一个开放问题。我们将该点对点的信源编码模型拓展至通信问题中，并提出一个泛用性较高的语义感知的JSCC模型。

2.3 语义感知的信源信道联合编码模型

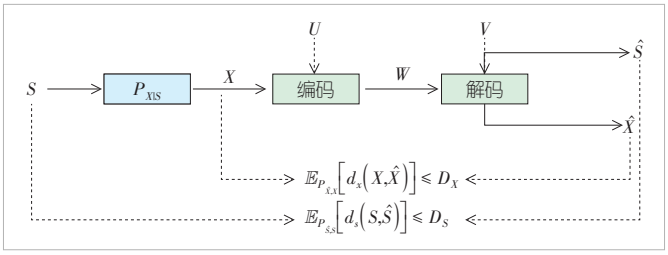
在文献[19]中，我们引入了语义感知的JSCC模型，具体如图5。对可观测信源序列X^k进行JSCC编码后，可以得到码字Y^n，再经信道后得到码字Z^n，之后译码为语义恢复$\hat{S}$与观测恢复$\hat{X}$。其中，编码方案要求语义恢复与观测恢复的失真程度在给定阈值D_S与D_X范围内。与信源压缩模型不同的是，引入信道模型后，速率失真函数不再具备较高的参考价值。因此，我们定义错误事件与错误概率：

$$\mathcal{E}(s^k,x^k)=\{z^n:d_s(s^k,\psi_S{}^k(z^n))>D_S\ or\ d_x(x^k,\psi_X{}^k(z^n))>D_X\},$$
$$\mathbb{P}(\mathcal{E})=\sum_{x^k}P_{x^k}(s^k)\sum_{s^k}P_{s^k|x^k}(s^k|x^k)\sum_{z^n\in\mathcal{E}(s^k,x^k)}P_{z^n|y^n}(z^n|\varphi^n(x^k))。\tag{5}$$

基于上述语义感知的JSCC模型的错误概率的内外界能够反映语义与观测失真度约束对通信译码错误概率的影响，具有较高的研究价值。考虑到码字块长$k\to\infty$的渐进情况，



▲图3 刻画语义信息的信源编码模型[13]



▲图4 具有边信息的语义信源编码模型

② 在联合高斯情形下，对S的失真约束可以退化为对X的失真约束。

错误概率内外界的具体推导由Csiszar的单约束JSCC发展为两约束JSCC：在长马尔科夫链的情形下，错误概率的逆定理即球填充外界（Sphere-packing Bound），可由最优列表译码得到；错误概率的正定理则可以通过构造考虑语义序列的高维随机箱（即随机编码）得到。值得一提的是，可达界可以通过删除较差码字得到更好的随机码本。该错误概率模型同样可以通过特定信源信道分布如信源联合高斯、信道瑞丽衰落、信道多输入多输出（MIMO）等得到更具体的表达式，并进行解析或数值求解。此外，除了错误概率，该模型可以推导错误指数（反应误码率下降的指数级速度），从收敛性角度指导损失函数的计算；推导可行速率区域，通过角点指导现实方案的设计等。

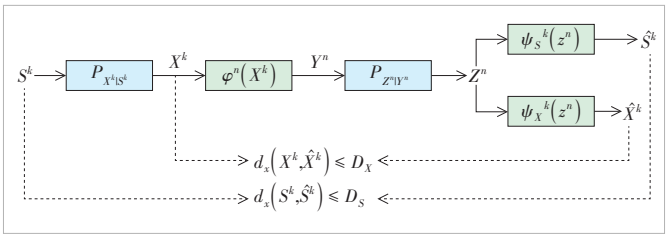
该JSCC模型很大程度上适配了现有的语义通信工作模型，我们以现有的工程可实现框架为例：一方面，在文献[15]中，图片序列X^k蕴含语义特征S^k，经由JSCC编码后得到序列Y^n，并经过信道传输接收得到加扰序列Z^n，在译码端由$\psi_X{}^k(z^n)$恢复为图片像素$\hat{X}$，从而完成重建任务，再由$\psi_S{}^k(z^n)$恢复为语义特征进行分类任务，即为同时执行图像恢复与下游任务的语义通信框架[②]。另一方面，将观测失真度约束取消即$D_X \to \infty$，对应文献[9]中恢复视频信号的单任务语义通信框架；将语义失真度取消即$D_S \to \infty$，对应以精准恢复比特为目标的传统通信框架；将JSCC拆解为独立的两个编码器，$\varphi^n(\cdot)=\varphi_1{}^m(\cdot)\varphi_2{}^n(\cdot)$，即采用分离信源信道编码（SSCC）的语义通信架构，见文献[12]。

3 挑战与问题

基于以上的介绍与讨论，本节将会从理论建模和实际应用两个角度，介绍语义通信在无线网络应用中的潜在挑战与开放性问题。

3.1 语义通信理论通用数学建模

现有语义信息的研究通常集中于测度定义与理论建模。本文中，我们介绍了一个具有研究价值的JSCC模型，但其点对点的建模特性对现实场景的通信设计仍有影响，例如：在多用户场景、云/边缘场景、考虑边信息（即收发两端的语义知识库）场景中，语义通信的理论建模将会发生改变。因此，如何将语义通信在现实通用的场景中合理建模是一个挑战。首先，语义信息的具有隐藏信源的特性，因此在CEO问题（间接多用户信源编码模型中的率失真问题）、Wyner-Ziv问题、多址接入（Multiple-Access Channel）、广播（Broadcast Channel）问题中引入间接信源是一种解决方案，但可达速率的边界、错误概率等的推导仍不清楚；其次，我们研究的是基于无限块长的编码模型，渐进等分性、联合典型工具对有限块长的性能损失研究是无法适用的。



▲图5 语义感知的信源信道联合编码模型[19]

3.2 语义传输与通信协议设计

大多数的语义通信框架集中于物理层的设计，语义与通信协议层结合的设计涉及较少。这是由于现有的下层协议设计大多是为了满足上层不同类型的需求，这会导致协议设计需要较高的泛用性。当语义通信执行特定下游任务时，现有的协议设计将会带来大量的物理层资源消耗。这会使得再好的物理层设计（如语义提取模块、信道CSI估计算法）都存在不必要的性能瓶颈。因此，语义觉知能力与现有通信协议的结合是一个亟待解决的问题。

3.3 跨模态语义融合传输架构设计

现有的语义传输更关注端到端的单一模态通信问题，例如：发送图片、接受图片、执行图片相关的下游任务。可以预见的是，在6G的智能网络范畴内，跨模态的通信模式是一种必然，即发送端需要同时处理文本与视频或语音与图片等模态组合，随后在接收端执行多媒体的语义任务。对于这类需求，现有的框架大都无法满足，或是只能做到简单组合。因此对于宽泛的应用需求，如何满足跨模态的语义通信仍不明确。

3.4 本地背景知识库的设计与更新

对于语义通信的范式而言，本地知识库是完成低码率传输的一个重要保障。也就是说，语义通信的性能在很大程度上取决于收发两端的背景知识的重叠率与完备程度。除了背景设计之外，语句的含义或者知识库会随着时代的发展改变或迭代，因此需要周期性地更新以达到预期的语义通信性

② k指块长，即k张图片对应k个语义特征，且k张图片同时编码得到n长的码字。图片本身维度、语义特征序列的维度与块长没有关系。在现实语义通信场景中，我们通常需要X的维度大于Y的维度。

能。针对收发两端本地知识库的重叠比率、涵盖范围等，已有研究进行了初步的建模，但仍然存在泛化性较差、耗费资源过多等问题。因此如何高效地对本地背景知识库进行设计与更新是一个重大挑战。

4 结束语

本文中，我们对语义通信的研究现状与前景进行了讨论，并基于理论模型与机器学习展开了分析。首先，从刻画语义信息的间接信源编码模型出发，将间接信源视为不可观测的语义信息，直接信源视为可被捕获的外部信息，并介绍了两个具有研究价值的语义信源编码模型。接着，我们将其中点对点的信源编码模型拓展至语义感知的JSCC模型中，并推导了初步性能边界。我们认为该模型对现有语义通信框架有较高的适配度，其性能极限分析对现有工作也有着有较大的指导意义。最后，指出现有语义通信框架面临的挑战与开放性问题。语义通信是继物理层通信之后的6G无线网络新兴范式，在“万物智联”愿景时代中必然占有一席之地。

参考文献

[1] ZHANG P, XU W J, GAO H, et al. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks [J]. Engineering, 2022, 8: 60-73. DOI: 10.1016/j.eng.2021.11.003

[2] SHANNON C E, WEAVER W. The mathematical theory of communication [M]. Urbana: University of Illinois Press, 1949

[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] BAO J, BASU P, DEAN M K, et al. Towards a theory of semantic communication [C]//Proceedings of 2011 IEEE Network Science Workshop. IEEE, 2011: 110-117. DOI: 10.1109/NSW.2011.6004632

[5] JUBA B. Universal semantic communication [M]. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2011

[6] 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

[7] 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

[8] WENG Z Z, QIN Z J. Semantic communication systems for speech transmission [J]. IEEE journal on selected areas in communications, 2021, 39(8): 2434-2444, DOI: 10.1109/JSAC.2021.3087240

[9] 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

[10] DAI J C, WANG S X, YANG K, et al. Adaptive semantic communications: overfitting the source and channel for profit [EB/OL]. (2022-11-08). [2023-01-20]. https://arxiv.org/abs/2211.04339

[11] KOUNTOURIS M, PAPPAS N. Semantics-empowered communication for networked intelligent systems [J]. IEEE communications magazine, 2021, 59(6): 96-102. DOI: 10.1109/MCOM.001.2000604

[12] CHATTOPADHYAY A, HAEFFELE B D, GERMAN D, et al. Quantifying task complexity through generalized information measures [EB/OL]. (2020-09-28) [2023-01-28]. https://openreview.net/pdf?id=vcKVhY7AZqK, 2020

[13] LIU J K, SHAO S, ZHANG W Y, et al. An indirect rate-distortion characterization for semantic sources: general model and the case of Gaussian observation [J]. IEEE transactions on communications, 2022, 70 (9): 5946-5959. DOI: 10.1109/TCOMM.2022.3194978

[14] NIU K, DAI J C, YAO S S, et al. A paradigm shift toward semantic communications [J]. IEEE communication magazine, 2022, 60(11): 113-119, DOI: 10.1109/MCOM.001.2200099

[15] TONG W J, LIU F F, SUN Z F, et al. Image semantic communications: an extended rate-distortion theory based scheme [C]//Proceedings of 2022 IEEE Globecom Workshops (GC Wkshps). IEEE, 2023: 1723-1728. DOI: 10.1109/GCWkshps56602.2022.10008733

[16] XIA L, SUN Y, NIYATUO D, et al. Wireless semantic communication: A networking perspective [EB/OL]. (2022-12-29) [2023-01-28]. https://arxiv.org/abs/2212.14142

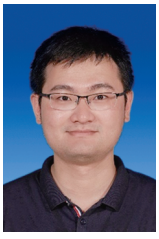
[17] KADAM S, KIM D I. Knowledge-aware semantic communication system design and data allocation [EB/OL]. (2022-12-30) [2023-01-28]. https://arxiv.org/abs/2301.034568

[18] ABDULLAH M, MADAIN A, JARARWEH Y. ChatGPT: fundamentals, applications and social impacts [C]//2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS). IEEE, 2022: 1-8. DOI: 10.1109/SNAMS58071.2022.10062688

[19] SHI Y X, SHAO S, WU Y P, et al. Excess distortion exponent analysis for semantic-aware MIMO communication systems [J]. IEEE transactions on wireless communications, 2023, (99): 1. DOI: 10.1109/TWC.2023.3238463

作　者　简　介

施雨轩，上海交通大学网络空间安全学院在读博士研究生；主要研究方向为信息论与编码、语义通信。



吴泳澎，上海交通大学教授、IEEE Senior Member，IEEE ICC、Globecom、VTC等国际著名会议技术委员会主席及委员；研究方向为空时无线通信理论与关键技术；主持国家自然科学基金、国家重点研发子课题，负责中兴通讯、国家电网等的多个科研项目；获国家自然基金委优秀青年基金、IEEE通信学会亚太区杰出青年研究学者奖、季寒冰青年教师奖、中国科协“青年托举人才计划”和2014德国洪堡学者等荣誉，博士论文《多用户MIMO传输理论方法》获首届中国通信学会“全国信息通信领域优秀博士学位论文”，并获多个学术会议最佳论文奖；发表论文100余篇。

张文军，上海交通大学教授、数字电视国家工程研究中心首席科学家、未来媒体网络协同创新中心主任；主要研究方向为视频编码与无线传输、多媒体语义分析与广播/宽带网络融合；是2006年发布的中国DTTB标准（DTMB）的主要贡献者；发表论文130余篇，拥有专利238项。

面向语义通信的语义知识库综述

Survey on Semantic Knowledge Base for Semantic Communications

孙亚萍/SUN Yaping[1]，崔曙光/CUI Shuguang[1,2,3]，张平/ZHANG Ping[1,4]

（1. 鹏城实验室，中国 深圳 518055；
2. 香港中文大学（深圳），中国 深圳 518055；
3. 香港中文大学（深圳）未来智联网络研究院，中国 深圳 518000；
4. 北京邮电大学泛网无线通信教育部重点实验室，中国 北京 100876）
(1. Peng Cheng Laboratory, Shenzhen 518055, China;
2. The Chinese University of Hong Kong (Shenzhen), Shenzhen 518055, China;
3. Future Network of Intelligence Institute (FNii), The Chinese University of Hong Kong (Shenzhen), Shenzhen 518000, China;
4. The Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China)

DOI：10.12142/ZTETJ.202302005
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230412.1134.004.html
网络出版日期：2023-04-12
收稿日期：2023-01-25

摘要：语义知识库是一种可为数据信息提供相关语义知识描述的、结构化的且具备记忆能力的知识网络模型，是语义通信的关键使能技术之一。首先，归纳分析计算机领域语义知识库研究现状，说明知识库在语义信息提取等方面的关键作用；然后，梳理无线通信领域中信源、信道、任务语义知识库的研究现状，揭示语义知识库在语义传输效率提升方面的潜能；最后，分别从多层级构建、动态智能演进、多智能体协同更新3个方面分析了语义传输中语义知识库面临的挑战。对于如何深度融合人工智能与通信技术，创新性地提出多层级语义知识库框架，认为一个跨模态、跨任务、跨环境的知识库的构建是高效语义传输的重点研究方向。

关键词：语义传输；语义知识库；语义信道联合编码；多层级构建；动态演进；多智能体协同

Abstract: The semantic knowledge base is a structured and memorable knowledge network model that can provide relevant semantic knowledge descriptions of data information, and it is one of the key enabling technologies for semantic communications. Firstly, the research status of the semantic knowledge base in the computer field is analyzed, explaining the key role of the knowledge base in semantic information extraction. Then, the research status of the semantic knowledge base for sources, channels, and tasks in the field of wireless communications is sorted out, and the potential of the semantic knowledge base in improving semantic transmission efficiency is revealed. Finally, the new research challenges brought by semantic knowledge base in semantic transmission are analyzed from three aspects: multi-level construction, dynamic intelligent evolution, and multi-agent collaborative update. The multi-level semantic knowledge base framework is innovatively proposed for deeply integrating artificial intelligence and communication technologies. It is believed that building a cross-modal, cross-task, and cross-environment knowledge base is the key research direction of efficient semantic transmission.

Keywords: semantic communication; semantic knowledge base; semantic-channel joint coding; multi-level construction; dynamic evolution; multi-agent collaboration

随着5G的不断成熟和商用，在构建“万物智联、数字孪生”的人类社会这个总体愿景的驱动下，6G将通过人工智能与通信技术的深度融合，实现人、机、物、智能体的智慧互联，有望成为支撑扩展现实（XR）、工业互联网、智慧城市等典型场景的核心基础设施[1]。与传统语法通信主要关注比特数据的准确传输不同[2]，人、机、物、智能体交互重点关注语义信息的准确传递。因此，传统语法通信系统的设计无法满足未来6G通信需求，而语义通信能够有效降低传输负载，提高任务服务质量[3]，已成为6G的关键技术之一。

在传统通信中，基于人工设计的编码映射，编码将信源符号映射到传统码流。映射函数的确立基于经验性的设计与构建。在语义通信中，基于人工智能（AI）的编码映射，语义编码将信源符号映射到语义码流。映射函数的确立基于数据与模型双驱动的学习与搜索。语义知识库定义了高效搜索空间，规范了搜索路径。因此，语义知识库是语义通信的外挂和效率的加持器。

语义知识库是一种可为数据信息提供相关语义知识描述的、结构化的且具备记忆能力的知识网络模型。面向语义通

基金项目：国家自然科学基金（62293482）；河套深港科技创新合作区深圳园区项目（HZQB-KCZYZ-2021067）；国家重点研发计划（2018YFB1800800）；深圳杰出人才计划（202002）；广东省引进创新创业团队项目（2017ZT07X152）；广东省人才项目（2019CX01X104）；广东省未来智联网络重点实验室项目（2022B1212010001）；深圳市大数据与人工智能重点项目（ZDSYS20170725140905 5）；鹏城实验室宽带通信部重大项目

信的语义知识库可分为信源、信道、任务知识库，分别为信源数据（如文本、图片、视频）、信道传输环境（如传输中障碍物和散射体位置与形状信息、智能反射面位置信息与配置矩阵），以及任务需求（如图片分类、三维重建、语义分割）提供多层级语义知识表征，从而支撑端到端语义通信的高效实现。如图1所示，基于语义知识库进行语义信道联合编解码设计，实现语义信息的准确传递，是当前语义通信的研究热点。

具体而言，在端到端语义通信中，发送端基于信源、信道以及任务知识库，获取信源数据多层级语义知识描述、传输环境的语义推断与估计，以及下游任务的语义需求，从而进行语义信道联合编码。接收端基于本地语义知识库，对接收到的信息进行知识检索与理解，完成语义信道联合解码，从而实现任务驱动的语义通信。

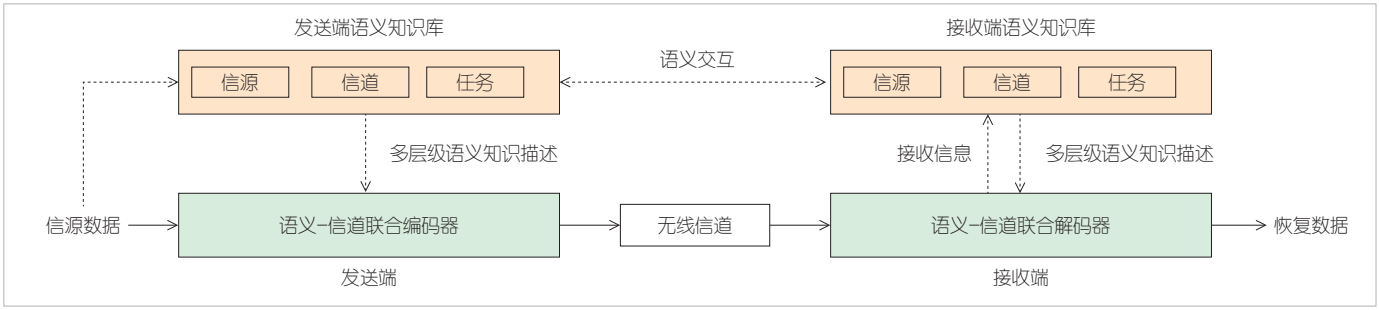
以图像的语义传输为例，假设发送端获得一张“斑马”图片，则可基于本地知识库，获取该图片的多层级语义知识向量。面向图像分类任务的最高级语义知识可表示为“斑马”；面向语义推断任务的中级语义知识包括斑马的属性描述“颜色：黑白”“轮廓：马”“条纹：有”等；面向图片传输任务的低级语义知识包括该图片的像素级特征向量。除此之外，信道知识库对传输环境的描述也会影响各层级语义知识表征维度。基于语义信道弹性编码方法，发送端首先传输最高级语义知识“斑马”。若接收端知识库里存储“斑马”相关语义特征描述，则任务完成；否则，则传输中级语义知识。若接收端知识库可解析“斑马”语义属性描述，则语义交互成功；否则，则发送低级语义知识，即原始图片信息，供接收端进行图片理解。

由此可见，语义知识库对于语义信道编解码方案的设计至关重要，可以提升语义通信传输效率以及智能任务的服务准确度。语义知识库的概念自提出以来在不断地发展和完善。语义知识库早期的工作主要基于计算机领域的知识图谱，以描述各实体概念及其相互关系为目标。近年来，随着语义通信重获关注，语义知识库也呈现出新的发展契机。

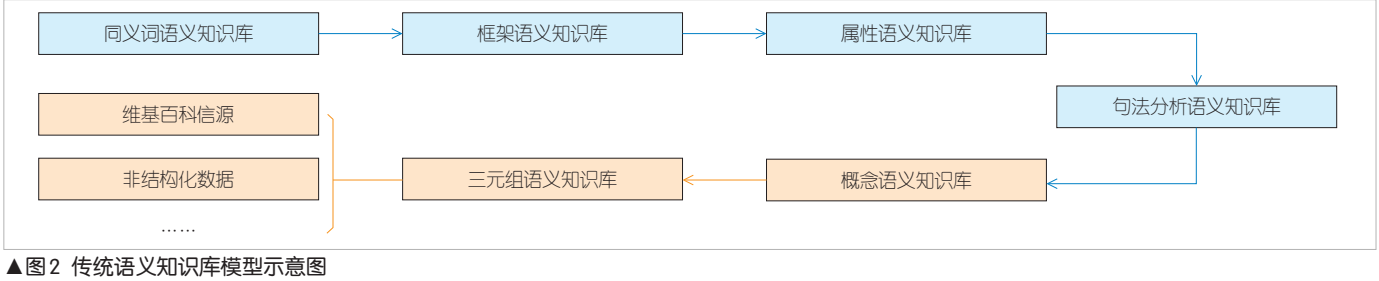
1 计算机领域知识库研究现状

传统语义知识库主要应用于计算机领域，如图2所示。最早的语义知识库可以追溯到WordNet[4]。WordNet的基本单元是同义词集合，即每个集合中的元素相互之间构成同义关系。WordNet使用不同的关系来构造语义知识库（包括上下位关系、反义关系、整体–部分关系等），并且基于这些关系构成了语义知识表征。随后FrameNet[5]引入了“框架”的概念，使得语义资料库可以灵活地表达更加复杂的语义关系。接下来，综合语言知识库（ILD）[6]为词语引入了属性的概念，通过建立词语之间的“实体–用途”的关系，丰富了语义知识库的表达。

由于手工建立语义知识库的效率非常低，MindNet[7]首次使用句法分析器自动分析英语词典释义文本，从而构建语义知识库。HowNet[8]作为首个汉语语义知识库，以概念本身以及概念与概念之间所具有的关系为基础，构建了语义知识库。ConceptNet[9]以实体、关系和属性三元组的形式进行语义知识库的构造，并以该形式完成了多代更新。DBPedia[10]通过爬取维基百科数据，完成了语义知识库的构建。



▲图1 端到端语义通信框架图



▲图2 传统语义知识库模型示意图

NELL[11]使用数据挖掘的方式从非结构化的数据中构建语义知识库，实现了基于更广泛文本信息来源的语义知识库自动化构造。

另一些研究则通过引入图像、程序源码等其他的模态数据源来进行语义关系的提取，丰富了知识库的数据来源。文献[12]将图像信息作为实体的一个额外特征，提出了一个新型大规模数据集，提升了知识库的表征学习能力。文献[13]通过分析深度学习及其对应的源代码，提出了一个融合文本、图像、程序源码等多模态的知识库，用于文献的挖掘与检索。文献[14]利用文章中的文本信息，提取实体及实体之间的关系，并且通过检测文章中的图像，构造了图像子知识库，从而利用文本与图像之间的特征匹配关系，解决了图像配文问题。

语义知识库的构建涉及典型场景的数据集构建、语义知识提取与建模。在计算机视觉领域，ImageNet[15]与COCO[16]数据集的提出为物体检测与图像分割任务提供了大规模的标注数据，促进了该领域的快速发展。卷积神经网络[17]与变形神经网络（Transformer）[18]为图像与自然语言等数据提供了强大的语义提取技术支撑。针对语义知识库构建过程中数据构建和语义知识提取方法设计，我们构建了大规模视觉语义信息数据集[19]，并且分别面向图像处理[20]、语义分割[21]、三维场景理解[22]与重建[23]、点云数据采样与恢复[24]等智能任务，提出了语义信息提取方法。

综上所述，基于常见数据模态（如文本、图像等）信息的语义知识库构建方法已得到充分的发展。然而，现有语义知识库仅可简单地描述各实体概念及其之间的相互关系，无法满足语义通信对信源、任务、信道等数据信息的多层级语义知识描述的需求。

2 通信系统知识库研究现状

语义传输中涉及的智能体包括发送端不同模态信源智能体、信道中不同传输环境智能体，以及接收端不同任务智能体知识库。本节中，我们将分别对语义传输中信源、信道以及任务语义知识库研究现状进行介绍。

2.1 信源知识库

目前，通信系统中信源知识库的构建主要有3类方法。第1类方法基于知识图谱构建知识库。针对文本的传输，文献[25]使用描述语义信息的三元组（包含头部实体、关系、尾部实体）来构建语义知识图谱，并且将此作为收发端语义知识库，以指导文本传输中的语义编解码。基于该语义知识库，文献[26]提取文本信源中所包含的语义三元组集合，并且度量各三元组的语义重要性。基于此，根据信道状态的好坏，文献[26]还提出智能化地选择所发送的三元组集合，以保证收发端信息的语义相似度。针对语音传输，文献[27]提出基于知识图谱的多层级结构的语义知识库基础模型，以及包含语义表达和语义符号抽象两个步骤的语义知识库构建方法。数值结果表明，基于该知识库的语义传输框架可在保证语义保真度的同时，降低带宽开销。针对图数据传输，鹏城实验室石光明教授研究团队提出了由显式语义、隐式语义以及与用户相关的知识推理机制组成的多层语义表征方法，并且基于模仿学习对接收端用户的语义推理机制进行训练，从而与发送端推理机制保持一致，降低了传输负载[28]。除此之外，文献[28]提出了语义知识库使能的异构网络中协同推理机制。

第2类方法以带标签的训练数据集作为知识库。上海交通大学陶梅霞教授研究团队提出，当需要传输的数据信息与训练数据集的统计特性分布不同时，可利用迁移学习中的领域自适应技术，降低两者间分布的差异性，并且动态地更新语义信道编解码方案[29]。该方法的有效性在图像传输任务中得到了验证。

第3类方法将基于深度学习模型提取的特征向量作为语义知识库。清华大学秦志金教授研究团队定义了一组有限离散语义基向量集合为语义知识库，并且对语义编解码以及语义知识库构建进行了端到端联合训练[31]。数值结果显示语义知识库可提高语义通信对语义噪声的鲁棒性。

2.2 信道知识库

现有信道环境知识库可分为特定于站点的数据库与特定于位置信息的知识库两大类。具体而言，特定于站点的数据库旨在提供准确的物理环境地图信息，主要包括三维城市地图[32]、无线电环境地图[33–34]等。然而，该类设计需要运行复杂度较高的算法（如射线跟踪算法），计算与存储资源开销较大。为了降低计算与存储资源的开销，特定于位置信息的知识库不再保留发送端与接收端相关活动信息，重点关注与信道特性相关的知识描述（如信道增益、阴影、入射角等）。该类设计主要包括信道增益地图[35]、信道路径地图[36]、波束索引映射[37]等。然而，该类设计主要局限于特定传输环境下信道知识的构建，在多变环境的自适应性以及泛化能力方面有待于进一步提高。

除此之外，针对收发端传播环境动态多变特性，北京邮电大学张建华教授研究团队提出基于环境特征、环境图表示等定义传播环境语义特征，用以辅助波束预测等任务完成[38]。数值仿真结果显示，在节省87%时间开销下，信道评

估与最大功率散射体检测任务准确度分别可达0.92与0.90。

2.3 任务知识库

现有任务知识库的构建主要为与任务相关的特征向量集合。文献[30]提出了面向图像分类任务的语义传输系统。该系统首先利用带类别标签的图像数据集来预训练一个图像分类网络，接着量化分类网络所提取的特征图与物体类别信息之间的相关性，并且以此特征图与类别相关性作为语义知识库。传输特征统计特性大大降低了带宽开销。文献[39]面向多任务需求多模态数据源，构建了由离散码本组成的跨任务共享的语义知识库，并且与收发端语义信道编解码器进行联合训练。

3 面向语义传输的语义知识库

实现语义知识库对语义通信的高效支撑仍有3方面的挑战：

1）多层级语义知识库构建。针对典型场景（如XR、工业互联网、智慧城市等）中复杂多变的信源、任务、信道环境等数据信息，如何高效构建多层级语义知识库至关重要。现有基于知识图谱与特征统计特性等构建的知识库仅考虑文本、图片等信息，无法满足语义通信的多层级语义表征需求。因此，亟需研究多层级语义知识库的构建理论与方法，以此形成统一的方法论，从而指导典型场景中语义知识库的高效构建。

2）语义知识库动态演进。语义通信的高效实现同样离不开收发端之间知识库的高度匹配，以及知识库对信源、任务与信道等数据信息的时变自适应性。然而，端到端语义传输中交互信息往往冗余且易受无线信道噪声与干扰的影响。现有知识库更新方法尚未考虑这一问题，无法直接应用于语义通信中知识库的动态演进。因此，亟需设计语义传输中知识库动态更新方法，揭示知识库动态演进机理。

3）多智能体知识库协同更新。语义通信呈现发送端信源多模态化、接收端任务需求多样化以及信道环境智能多变等特性。然而，现有知识库协同机制同样仅针对文本、图片等信息，场景单一，并且尚未考虑语义传输中跨智能体的信息缺失、重叠以及隐私保护等，无法满足跨模态语义融合、跨任务语义理解以及跨环境语义传输需求。因此，如何高效协同更新多智能体（包括多模态信源、多任务、以及多环境信道）语义知识库是另一个亟待解决的问题。

4 结束语

基于知识图谱、带标签的训练数据集，特征统计特性的语义知识库已被应用于端到端语义通信中，并取得了一定的成果。然而，现有研究工作面临着构建理论与方法匮乏、动态演进机理不明确、多智能体协同更新策略不清晰等问题。如何深度融合人工智能与通信技术，创新性地提出多层级语义知识库框架。文章认为打造跨模态、跨任务、跨环境的知识库是高效语义传输新的研究主题与挑战。

致谢

本研究得到了香港中文大学（深圳）陈冠英教授、韩晓光教授、许杰教授的大力支持与帮助，在此表示感谢！

参考文献

[1] IMT-2030（6G）推进组. 6G总体愿景与潜在关键技术白皮书 [EB/OL]. [2023-01-17]. https://www. xdyanbao. com/doc/gde13jnbhh? bd_vid=8625646446254978981
[2] 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
[3] MATTI L, KARI L, FEDERICO C, et al. Key drivers and research challenges for 6G ubiquitous wireless intelligence [EB/OL]. [2023-01-17]. https://elib.dlr.de/133477/
[4] MILLER G A. WordNet [J]. Communications of the ACM, 1995, 38(11): 39-41. DOI: 10.1145/219717.219748
[5] BAKER C F, FILLMORE C J, LOWE J B. The Berkeley FrameNet project [EB/OL]. (1998-08-10) [2023-01-16]. https://www. semanticscholar. org/paper/The-Berkeley-FrameNet-Project-Baker-Fillmore/547f23597f9ec8a93f66cedaa6fbfb73960426b1
[6] KRUYT J G. The integrated language database of 8th-21st-century Dutch [EB/OL]. [2023-01-20]. https://www. researchgate. net/publication/228714920_The_Integrated_Language_Database_of_8th-21st-Century_Dutch
[7] RICHARDSON S D, DOLAN W B, VANDERWENDE L. MindNet: acquiring and structuring semantic information from text [EB/OL]. [2023-01-20]. https://www. researchgate. net/publication/230876760_MindNet_Acquiring_and_Structuring_Semantic_Information_from_Text
[8] DONG Z D, DONG Q. HowNet - a hybrid language and knowledge resource [C]//Proceedings of International Conference on Natural Language Processing and Knowledge Engineering. IEEE, 2004: 820-824. DOI: 10.1109/NLPKE.2003.1276017
[9] LIU H, SINGH P. ConceptNet—a practical commonsense reasoning tool-kit [J]. BT technology journal, 2004, 22(4): 211-226. DOI: 10.1023/B:BTTJ.0000047600.45421.6d
[10] AUER S, BIZER C, KOBILAROV G, et al. DBpedia: a nucleus for a web of open data [M]. The semantic web. Berlin: Springer, 2007: 722-735. DOI: 10.1007/978-3-540-76298-0_52
[11] MITCHELL T, FREDKIN E. Never-ending language learning [C]// Proceedings of 2014 IEEE International Conference on Big Data (Big Data). IEEE, 2015: 1. DOI: 10.1109/BigData.2014.7004203
[12] MOUSSELLY S H, BOTSCHEN T, GUREVYCH I, et al. A multimodal translation-based approach for knowledge graph representation learning [EB/OL]. [2023-01-19]. https://aclanthology.org/S18-2027/
[13] KANNAN A V, FRADKIN D, AKROTIRIANAKIS I, et al. Multimodal knowledge graph for deep learning papers and code [C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. ACM, 2020: 3417-3420. DOI: 10.1145/3340531.3417439
[14] ZHAO W, HU Y, WANG H, et al. Boosting entity-aware image captioning with multi-modal knowledge graph [EB/OL]. (2021-06-26) [2023-01-22]. https://arxiv.org/abs/2107.11970
[15] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database [C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248-255. DOI: 10.1109/CVPR.2009.5206848
[16] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects

in context [M]// Computer Vision – ECCV 2014. Cham: Springer International Publishing, 2014: 740–755. DOI: 10.1007/978–3–319–10602–1_48
[17] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84–90. DOI: 10.1145/3065386
[18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2023–01–19]. https://arxiv.org/abs/1706.03762
[19] ZHU H, CAO Y, JIN H, et al. Deep fashion3D: a dataset and benchmark for 3D garment reconstruction from single images [EB/OL]. [2023–01–19]. https://arxiv.org/abs/2003.12753
[20] ZHANG S Y, CUI S G, DING Z. Hypergraph–based image processing [C]// Proceedings of 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020: 216–220. DOI: 10.1109/ICIP40778.2020.9190874
[21] QIU L T, XIONG Z Y, WANG X H, et al. ETHSeg: an amodel instance segmentation network and a real–world dataset for X–ray waste inspection [C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 2273–2282. DOI: 10.1109/CVPR52688.2022.00232
[22] YUAN Z H, YAN X, LIAO Y H, et al. X–Trans2Cap: cross–modal knowledge transfer using transformer for 3D dense captioning [C]// Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 8553–8563. DOI: 10.1109/CVPR52688.2022.00837
[23] HAN X G, ZHANG Z X, DU D, et al. Deep reinforcement learning of volume–guided progressive view inpainting for 3D point scene completion from a single depth image [C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 234–243. DOI: 10.1109/CVPR.2019.00032
[24] ZHAO W, YAN X, GAO J, et al. PointLIE: locally invertible embedding for point cloud sampling and recovery [EB/OL]. [2023–01–20]. https://arxiv.org/abs/2104.14769
[25] ZHOU F, LI Y, ZHANG X, et al. Cognitive semantic communication systems driven by knowledge graph [EB/OL]. [2023–01–18]. https://arxiv.org/abs/2202.11958
[26] JIANG S, LIU Y, ZHANG Y, et al. Reliable semantic communication system enabled by knowledge graph [J]. Entropy, 2022, 24(6): 846. DOI: 10.3390/e24060846
[27] SHI G, GAO D, SONG X, et al. A new communication paradigm: from bit accuracy to semantic fidelity [EB/OL]. [2023–01–20]. https://arxiv.org/abs/2101.12649
[28] XIAO Y, SUN Z, SHI G, et al. Imitation learning–based implicit semantic–aware communication networks: multi–layer representation and collaborative reasoning [EB/OL]. [2023–01–20]. https://arxiv. org/abs/2210.16118
[29] ZHANG H, SHAO S, TAO M, et al. Deep learning–enabled semantic communication systems with task–unaware transmitter and dynamic data [EB/OL]. (2022–10–28)[2023–01–18]. https://arxiv.org/abs/2210.16118
[30] YANG Y, GUO C, LIU F, et al. Semantic communications with AI tasks [EB/OL]. (2021–09–29)[2023–01–16]. https://arxiv.org/abs/2109.14170
[31] HU Q, ZHANG G, QIN Z, et al. Robust semantic communications against semantic noise [EB/OL].(2022–02–07)[2023–01–16]. https://arxiv.org/abs/2202.03338
[32] ESRAFILIAN O, GANGULA R, GESBERT D. 3D–map assisted UAV trajectory design under cellular connectivity constraints [C]//Proceedings of ICC 2020 – 2020 IEEE International Conference on Communications (ICC). IEEE, 2020: 1–6. DOI: 10.1109/ICC40277.2020.9149190
[33] BI S Z, LYU J B, DING Z, et al. Engineering radio maps for wireless resource management [J]. IEEE wireless communications, 2019, 26(2): 133–141. DOI: 10.1109/MWC.2019.1800146
[34] CHEN J T, YATNALLI U, GESBERT D. Learning radio maps for UAV–aided wireless networks: a segmented regression approach [C]//Proceedings of 2017 IEEE International Conference on Communications (ICC). IEEE, 2017: 1–6. DOI: 10.1109/ICC.2017.7997333
[35] DALL´ANESE E, KIM S J, GIANNAKIS G B. Channel gain map tracking via distributed kriging [J]. IEEE transactions on vehicular technology, 2011, 60(3): 1205–1211. DOI: 10.1109/TVT.2011.2113195
[36] WU D, ZENG Y, JIN S, et al. Environment–aware and training–free beam alignment for mmWave massive MIMO via channel knowledge map [C]// Proceedings of 2021 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2021: 1–7. DOI: 10.1109/ICCWorkshops50388.2021.9473871
[37] VA V, CHOI J, SHIMIZU T, et al. Inverse multipath fingerprinting for millimeter wave V2I beam alignment [J]. IEEE transactions on vehicular technology, 2018, 67(5): 4042–4058. DOI: 10.1109/TVT.2017.2787627
[38] SUN Y T, ZHANG J H, YU L, et al. How to define the propagation environment semantics and its application in scatterer–based beam prediction [EB/OL]. (2023–01–18) [2023–01–23]. https://ieeexplore. ieee.org/document/10020175
[39] ZHANG G, HU Q, QIN Z, et al. A unified multi–task semantic communication system for multimodal data [EB/OL]. (2022–09–16)[2023–01–20]. https://arxiv.org/abs/2209.07689

作 者 简 介

孙亚萍，鹏城实验室助理研究员；主要研究方向为语义通信和移动边缘计算。

崔曙光，香港中文大学（深圳）校长学勤讲座教授、未来智联网络研究院院长，国家重点研发计划首席科学家，全球高被引学者，IEEE Fellow，深圳市杰出人才培养计划首批入选人，港中深–京东集团人工智能联合实验室主任，深圳市大数据研究院常务副院长，IEEE 无线技术委员会主席，2014 年入选 IEEE 通信协会杰出讲师、汤森路透全球高被引科学家、ScienceWatch 全球最具影响力科学家；研究方向为数据驱动、AI 赋能的大规模系统控制和资源管理；获得 IEEE 信号处理协会 2012 年最佳论文奖、IEEE ICC 2020 年最佳论文奖、IEEE ICIP 最佳论文列表、IEEE GLOBECOM 最佳论文奖、中国 ICT 2020 创新应用奖、Chinagraph 首个图形开源数据集奖、中国电子学会自然科学一等奖、中国通信学会技术发明一等奖；发表论文 280 余篇。

张平，北京邮电大学教授、博士生导师，中国工程院院士；现从事信息通信相关的科研和教学工作，研究方向为无线通信；曾获得国家科学技术进步奖特等奖 1 项、国家科学技术进步奖一等奖 1 项、国家技术发明奖二等奖 2 项、国家科学技术进步奖二等奖 2 项，曾获光华工程科技奖、何梁何利基金科学与技术进步奖，带领的团队入选首批“全国高校黄大年式教师团队”。

基于CSI反馈的语义图像传输

Semantic Image Transmission Based on CSI Feedback

姜培文/JIANG Peiwen，韩瑜/HAN Yu，
金石/JIN Shi，李潇/LI Xiao

（东南大学移动通信国家重点实验室，中国 南京 210096）
(National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China)

DOI：10.12142/ZTETJ.202302006
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1503.020.html
网络出版日期：2023-04-11
收稿日期：2023-01-25

摘要：语义通信通过构建传输背景知识库、感知传输内容，大大节省传输带宽，提升传输质量。该方面的研究逐步与无线通信物理层融合，从而构建完整的语义传输系统，但现存的端到端语义编码设计无法感知不断变化的无线信道。提出了基于信道状态信息反馈的图像语义分割编码（SS-CSI）方案，根据不同子信道上的信噪比（SNR）对传输的关键特征信息进行分割和编码，并根据实际分类任务需求保护相关语义特征。仿真测试表明，SS-CSI根据实际信道环境，在低SNR下大幅提升了物体传输质量和分类性能，在高信噪比上则进一步传输更多背景信息，提升了整体的图像均方误差性能。

关键词：语义分割；无线通信；图像传输

Abstract: Semantic communications greatly saves transmission bandwidth and improves transmission quality by constructing transmission background knowledge base and sensing transmission content. At present, this research is gradually integrated with the wireless communication physical layer to build a complete semantic transmission system. However, the existing end-to-end semantic coding design is unable to perceive the changing wireless channel. An image semantic segmentation coding based on CSI feedback (SS-CSI) is proposed. According to the signal-to-noise ratio (SNR) on different sub-channels, the key feature information transmitted is segmented and coded. And the relevant semantic features are protected according to the actual classification requirements. Simulation results show that SS-CSI greatly improves the object transmission quality and classification performance at low SNR, and further transmits more background information at high SNR, which improves the overall image mean square error performance.

Keywords: semantic segmentation; wireless communication; image transmission

智能通信系统在提升用户体验的同时，也服务于大量的终端。由于新冠疫情大流行，移动工作和在线会议成为人类生活中必不可少的一部分。据相关机构统计，移动网络传输流量与大流行爆发前相比增加了60%以上。为了应对难以承受的负载，一些服务提供商（如YouTube等），只能选择在高峰时间降低视频质量。此外，自动驾驶汽车依赖于数千个传感器，同时又与其他车辆通信，从而产生大量的数据。这些传输服务于特定的任务。因此，在智能化进程中通信系统往往需要通过对文本、语音、图像、视频等多模态数据的理解和传输，来完成远程操控、辅助驾驶等特定任务[1-3]。基于经典的香农编码范式[4]，信源编码压缩传输内容并对内容进行解耦，以生成独立分布的码字。信道编码通过对码字增加冗余来对抗无线信道带来的传输差错。因此，传统的编解码模块通常遵循分而治之的设计原则。但是这种设计方式在全新的应用场景下仍然存在问题：一方面，诸如在线会议、自动驾驶等低延迟场景的码长受限，最优编码方式难以获得；另一方面，特定任务下的传输特征具有较强的内容相关性，新的深度学习方法（如人脸重构[5]）无须使用额外码字即可完成错误修复。语义通信作为智能通信的关键技术之一，通过构建传输背景知识库[6]、感知传输内容,极大地节省传输带宽,提升传输质量,成为极具潜力的研究方向。

传统的物理模块都是独立优化的，不适用于基于语义的通信系统。例如，调制和信号检测被设计为最小化误码率（BER）。信道估计是最小化估计信道与真实信道之间的均方误差。事实上，对于语义通信，不同传输比特具有不同的重要性，并且这些传输比特也呈现出一定的语义相关性。文献[7]利用句子相似度形成一种新的调制方法，这种方法的符号在星座图上的位置并不是等间隔分布的。在实际应用中，峰均功率比也是一大挑战，这是因为高峰均功率比对硬件设备要求很高。在文献[8]中，峰均功率比被认为是一个额外的损失函数，语义网络被训练后能同时实现最小化语义损失和峰均功率比。这意味着峰均功率比的降低应该与语义指标相匹配。在不同的信道环境下，混合重传请求（HARQ）是保证传输成功的关键。传统的HARQ主要基于用于前向纠错

的信道编码和用于错误检测的循环冗余校验码。文献[9–10]引入基于语义的联合信源信道编码设计HARQ，大幅提高了传输性能。信道状态信息（CSI）反馈中的数据隐藏方法[11]也是一种潜在的深度学习方法，用于消除CSI反馈中的传输负载。

上述语义方案引入基于知识库的编解码网络、任务导向的语义损失函数等，大幅提升了系统的性能。但这些方案需要借助端到端的训练，才能实现发射端和接收端对知识库的共享，以及发射端对任务需求的感知。因此，语义通信方法通常训练在可导的高斯白噪声信道或者比特翻转信道上等，无法充分利用实际的信道特征，难以实现对语义码字灵活的保护。本文讨论了语义通信方法在无线通信系统中的应用，提出了基于语义分割方法的图像传输系统，并利用反馈的无线信道特征来指导编码网络对传输内容的保护。通过比较不同方法在分类准确率和均方误差上的性能，所提系统可以灵活适应变化的信道，在低信噪比下优先保护影响分类准确率的物体，在高信噪比下进一步保护背景信息，从而实现在不同的信噪比上都能取得比其他方法更好的分类准确率和均方误差性能。

1 系统模型

为了传输信源信息，假设存在一张图片$\boldsymbol{p}$，传统发射机首先对其进行信源编码$S(\cdot)$和信道编码$C(\cdot)$，从而减少传输冗余，对抗变化的无线信道。经过编码后的码字表示为：

$$\boldsymbol{s} = C(S(\boldsymbol{p}))。\tag{1}$$

对于传统的正交频分复用（OFDM）系统来说，编码码字$\boldsymbol{s}$会被调制到K个子载波上，即$\boldsymbol{s} = [s_1, s_2, \cdots, s_K]$。在通过无线信道后，接收到的信号可以表示为：

$$\boldsymbol{y} = \boldsymbol{h} \cdot \boldsymbol{s} + \boldsymbol{z},\tag{2}$$

其中，$\boldsymbol{h}$是信道的频率响应，$\cdot$是哈达玛积，$\boldsymbol{z}$是均值为0、方差为$\boldsymbol{\sigma}^2$的加性高斯白噪声。对于频率选择性信道来说，$\boldsymbol{h}$可以被看作K个平坦的子信道，每个子信道具有不同幅度的信道频率响应，即$\boldsymbol{h} = [h_1, h_2, \cdots, h_K]$。考虑到不同子信道的频率响应，接收的码字可以表示为：

$$\hat{\boldsymbol{s}} = \left[\frac{y_1}{\hat{h}_1}, \cdots, \frac{y_K}{\hat{h}_K}\right]。\tag{3}$$

因此，整体的信噪比（SNR）可以表示为：

$$\mathrm{SNR} = \frac{\sum_{k=1}^{K} \| h_k \cdot s_k \|^2}{K\sigma^2}。\tag{4}$$

在接收端恢复的图片为：

$$\hat{\boldsymbol{p}} = S^{-1}(C^{-1}(\hat{\boldsymbol{s}}))，\tag{5}$$

其中，$S^{-1}(\cdot)$和$C^{-1}(\cdot)$分别代表信源译码和信道解码。

传统OFDM系统的传输目标是使传输BER达到最小，但在复杂的信道环境和海量的移动设备等情况下，无线通信带宽是有限的。语义通信方法能够感知传输内容，分割提取语义特征，基于任务需求进行端到端训练，从而大幅度降低传输带宽，提升传输质量。现有的语义方法大多是在可导的高斯白噪声信道上训练获得的，而实际OFDM系统下的信道是乘性的、不可导的，因此端到端训练的语义通信方法无法充分利用信道特征，传输性能受限。对此，我们利用CSI反馈，灵活调整编码方法。

2 网络结构和传输方案

本节中，我们提出基于语义分割的图像传输方法。首先，根据语义分割内容实现对信源信息的编码和解码，以及对传输噪声的对抗；接着，考虑无线信道和CSI反馈对网络设计的影响，并根据不同子载波上的SNR，实现对语义分割后的特征保护。

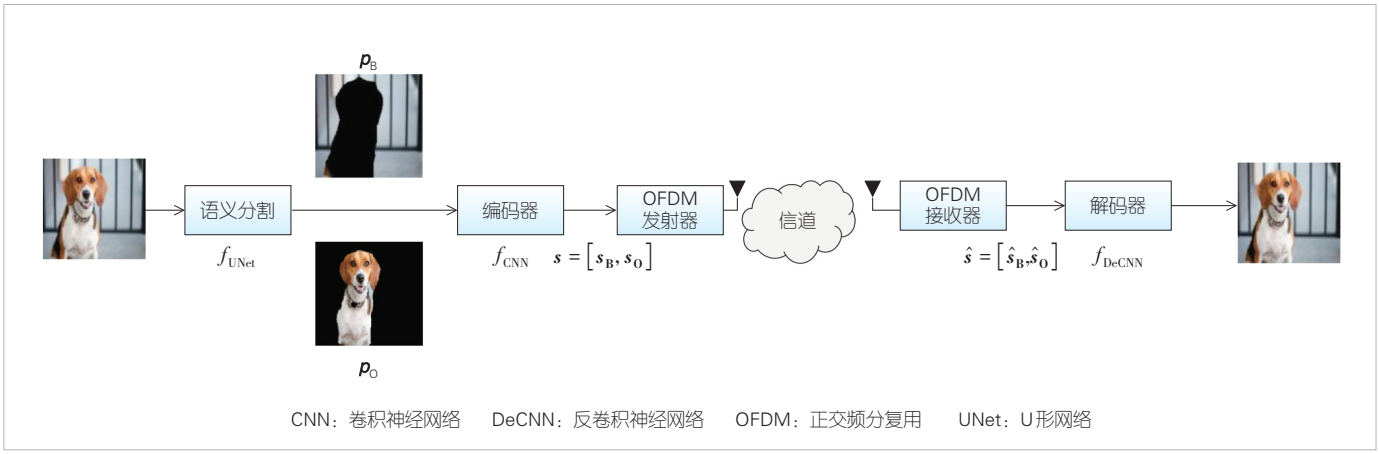
区别于传统的信源编码，语义信源编码语义分割将源数据划分为不同的语义部分，每个语义部分通常具有不同的重要程度。在特定的场景中，分割依赖于一个训练好的网络[12]，其蕴含着在这个训练场景下的语义知识库。比如：一幅图像可以分为背景和物体，其中背景通常较少受到用户关注且对物体分类识别性能影响不大。语义分割后，信道编码可以保护信源的不同部分，并在不同信道条件下保护最重要的语义部分。

针对图片传输的问题，本文采用基于UNet[13]的语义分割方法。如图1所示，图片被分割为背景和物体两类，分别表示为$\boldsymbol{p}_{\mathrm{B}}$和$\boldsymbol{p}_{\mathrm{O}}$。对于分割后的两部分待传图片信息，分别采用卷积编码网络，将其编码为码字$\boldsymbol{s}$，即：

$$[\boldsymbol{p}_{\mathrm{B}}, \boldsymbol{p}_{\mathrm{O}}] = f_{\mathrm{UNet}}(\boldsymbol{p}),\tag{6}$$

$$\boldsymbol{s} = [\boldsymbol{s}_{\mathrm{B}}, \boldsymbol{s}_{\mathrm{O}}] = f_{\mathrm{CNN}}([\boldsymbol{p}_{\mathrm{B}}, \boldsymbol{p}_{\mathrm{O}}]),\tag{7}$$

其中，语义分割网络表示为$f_{\mathrm{UNet}}(\cdot)$，$\boldsymbol{p}_{\mathrm{B}}$和$\boldsymbol{p}_{\mathrm{O}}$分别为分割后的背景信息和物体图片信息，$\boldsymbol{s}_{\mathrm{B}}$和$\boldsymbol{s}_{\mathrm{O}}$分别为背景信息和图片信息被卷积网络$f_{\mathrm{CNN}}(\cdot)$编码后得到的码字。卷积网络输入维度为（160,160,3），输出维度为（1,1,16），输出激活函数为Tanh，即$\boldsymbol{s}_{\mathrm{B}}$、$\boldsymbol{s}_{\mathrm{O}}$分别编码出16个–1~1的实数。该网络包含两个中间层，中间层均由卷积和下采样组成，其中第1个中



▲图1 基于语义分割的图片传输系统

间层输出维度为（40,40,64），第2个中间层输出维度为（10,10,128），激活函数为Relu。在该码字通过上述OFDM传输系统后，码字的估计$\hat{s}=[\hat{s}_B,\hat{s}_O]$，该码字的估计可以通过反卷积网络$f_{DeCNN}(\cdot)$恢复成原图片，即：

$$\hat{p}=f_{DeCNN}([\hat{s}_B,\hat{s}_O]), \tag{8}$$

其中，反卷积网络输入为（1,1,32），其中间层尺寸与卷积编码网络一一对应。

在进行端到端训练时，由于实际信道影响不可导，假设从s到$\hat{s}$的变化符合高斯白噪声。图像训练通常采用均方误差（MSE）损失函数来恢复每个像素点的信息，但在训练时为了引导网络去保护对任务更重要的信息，会引入基于分类识别任务的损失函数。此损失函数依赖预训练好的视觉几何组（VGG）分类网络$f_{VGG}(\cdot)$，并计算接收图像分类结果和发送图像分类结果的交叉熵（CE）损失$L_{CE}(\cdot)$。训练的损失函数可以写为：

$$L=L_{MSE}(\hat{p},p)+0.1L_{CE}(f_{VGG}(\hat{p}),p)。 \tag{9}$$

在获得CSI反馈后，计算出K个子信道上的传输SNR值，即可获得每个子信道的传输质量。因此，我们把语义分割和信道编码与CSI反馈相结合，以更加有效地保护传输特征。对于上述训练好的编码方式，可采用一层全连接网络将其分配在不同的子信道上，并且通过端到端训练得到合理的分配方式。具体方法如下：

$$s=[s_1,s_2,\cdots,s_K]=\mathrm{Tanh}(W[s_B,s_O]), \tag{10}$$

其中，W为学习到的乘性可训练参数分配方式。由于每个信道的SNR值已知，可以按照从高到低的顺序依次将$[s_1,s_2,\cdots,s_K]$放置在不同的子信道上，即：

$$y=[y_1,\cdots,y_K]=[h_{max},\cdots,h_{min}]\cdot[s_1,s_2,\cdots,s_K]+z。 \tag{11}$$

因此，s_1总是传输在SNR最高的子载波上，s_K则传输在SNR最差的子载波上。基于2.1节中已经训练好的模型参数，保持其他参数不变，只训练W，则训练损失与2.1一致。以上网络训练均采用Adam优化器，学习率设置为0.002，训练1 000轮。

由于采用不同子信道上的传输SNR来训练参数W，并且考虑实际信道模型和信道估计方法对不同子载波上SNR的影响，待传输的语义特征可以更加有效地适应信道环境的变化。网络更倾向于将分割后的语义特征$[s_B,s_O]$中的重要信息放置在较高SNR的子信道上，从而在低SNR下优先保护分类准确率。

3 仿真与测试

3.1 系统参数

传输一张图片会占用两个OFDM符号，每个OFDM符号占用32个子载波。其中，第1个OFDM符号为已知导频，用于信道估计并反馈各子载波上的信道响应；第2个OFDM符号携带图片信息。32个子载波上的复数由64个实数组成，即发送信号s为64个实数。在语义分割下，s_B、s_O各占用一半子载波数信道模型采用多径信道，在[0,4,10]采样点处有3条延时扩展路径，功率分布[0,–5,–10] dB。测试数据采用视觉对象类别（VOC）中的动物分类数据集。我们使用物种识别准确率来衡量物体传输的质量，并采用MSE衡量整体图片像素信息传输的质量。

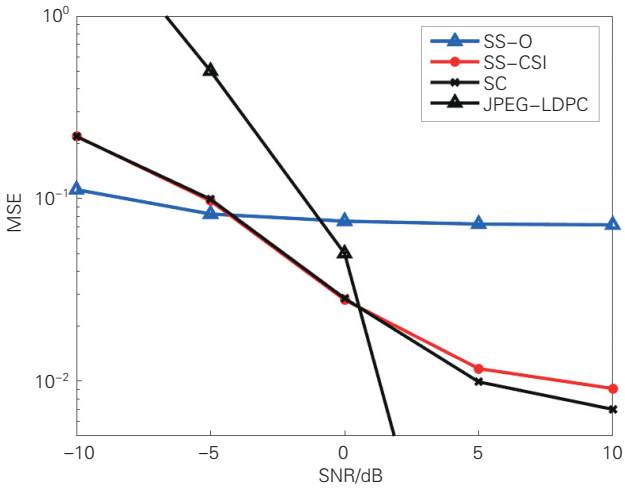
4种比较方法为：只传输物体部分的图像语义分割编码方案（SS–O）、基于信道状态信息反馈的图像语义分割编码（SS–CSI）、传输全部内容的图像语义分割编码（SC），以及

传统的JPEG信源编码加上低密度奇偶校验码（LDPC）传输，其中低密度奇偶校验码码率为64/255。

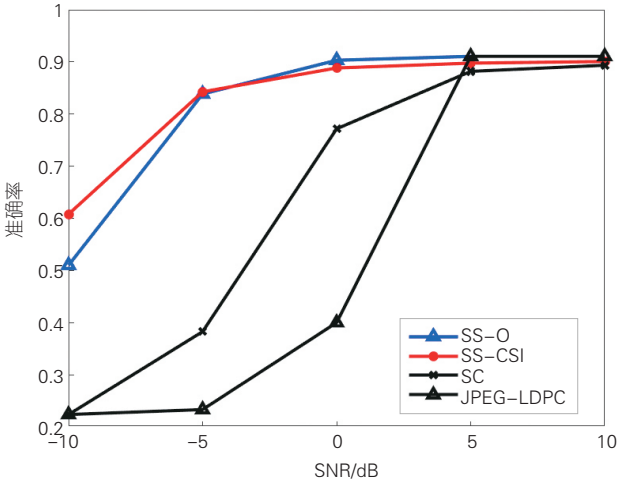
3.2 测试结果

如图2所示，我们比较了不同方法下的图片传输分类准确率和MSE性能。分类准确率反映了图片中物体特征的还原度，MSE性能反映了图片的像素信息在传输中的损失程度。

在分类准确率性能方面，SS-O的所有32子载波均用于传输语义分割后的物体部分，因此它在较低SNR（例如-5 ~ 5 dB）下也拥有很好的分类准确率。但是过低的SNR会使得SS-O的传输差错增大，例如：在-10 ~ -5 dB下，SS-O的分类准确率反而不如SS-CSI。SS-CSI根据反馈的每个子载波上的SNR来分配不同的语义特征。实际上，由于在训练时考虑了分类准确率，SS-CSI通常把代表物体的语义特征s_0分配到SNR更高的子载波上，因此在SNR为-10 ~ -5 dB时，部分较差信道带来的差错对SS-CSI影响更小，分类准确率更高。当SNR为0 ~ 10 dB时，由于大部分子信道条件都较好，而SS-O用于传输物体的子载波数大于SS-CSI，此时SS-O的分类准确率性能要略好于SS-CSI。SC中物体和背景占用的子载波数目相同且不用CSI反馈来辅助分配这些待传输的码字，因此在SNR较差的情况下，SC的分类准确率不如SS-O和SS-CSI。在SNR较高时，JPEG-LDPC的分类准确率达到分类网络的上限。但随着SNR的降低，传输差错超出LDPC编码的纠错能力，JPEG-LDPC分类准确率性能快速下降。



（a）不同信噪比下的均方误差性能比较



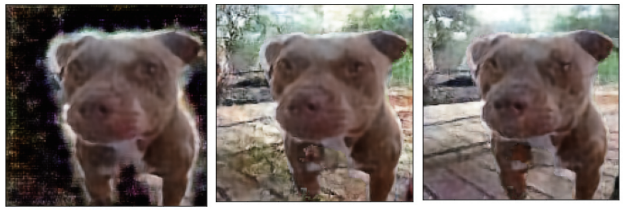
（b）不同信噪比下的物种识别准确率性能比较

JPEG：联合图像专家组
LDPC：低密度奇偶校验码
MSE：均方误差
SC：传输全部内容的图像语义分割编码
SNR：信噪比
SS-CSI：基于信道状态信息反馈的图像语义分割编码
SS-O：只传输物体部分的图像语义分割编码

▲图2 基于语义分割的图片传输系统性能比较

在MSE性能方面，SS-O完全舍弃了背景部分的传输方法，以便于在SNR很低时可以更好地保护物体部分的MSE性能。然而，在信道条件较好的情况下，SS-O也无法恢复背景部分，使得MSE几乎不会随着SNR的上升而降低。在SNR较低的情况下，SS-CSI和SC的MSE性能相当。CSI反馈会使得SS-CSI总倾向于将更好的子信道用于物体部分的传输，但在SNR较高的情况下，增加占比较大的背景部分会进一步降低MSE，导致SS-CSI在SNR较高的情况下略差于SC。传统的JPEG-LDPC方法在SNR较高时，使传输错误均被纠出，此时MSE几乎为0。同样地，当SNR降低导致传输错误超出其纠错能力后，MSE性能也会迅速降低。

在SNR=0 dB时，不同传输方法对图像的恢复结果如图3所示。其中，SS-O可将所有传输带宽用于恢复无图部分的像素信息。虽然物体部分恢复得最清晰，但背景却被完全舍弃。与SS-O不同，本文提出的方法SS-CSI在尽可能恢复物



（a）SS-O方法接收的图像　（b）SS-CSI方法接收的图像　（c）SC方法接收的图像

SC：传输全部内容的图像语义分割编码
SS-CSI：基于信道状态信息反馈的图像语义分割编码
SS-O：只传输物体部分的图像语义分割编码

▲图3 信噪比为0 dB时3种方法的实例图

体信息的同时，利用SNR较低的信道，传输一部分背景信息。传统的语义传输方法SC不考虑信道条件，物体和背景均有不同程度的损失。从图3可以看出，SS-CSI能够根据CSI反馈的信息对不同传输部分进行灵活保护，即在MSE性能和分类准确率两种损失函数的指导下，网络学习到优先降低物体部分的MSE，在提高分类性能的同时对背景进行保护。通过这样的方法，网络既可以在高SNR下取得很好的MSE性能，又可以在低SNR下取得很好的分类性能。

4 结束语

本文中，我们介绍了语义通信系统在提升传输质量、节省传输带宽方面的优势；考虑到典型的OFDM无线通信系统以及无线信道的特性，阐述了语义信源信道编码在无线通信中的融合应用；设计了基于语义分割的无线图片传输方案，即通过对不同重要性的传输部分进行分割（如背景和物体信息），实现对重要语义特征的保护。考虑到CSI反馈在无线通信中的重要性，SS-CSI方法使用反馈的CSI信息指导网络对不同传输部分进行保护，并使用具体的任务指导网络训练。此外，我们比较了SS-CSI和CSI反馈对分类准确率和MSE的影响，并通过具体实例证明了分割和CSI反馈可以对接收图片的重要部分进行更好地保护。

参考文献

[1] GUNDUZ D, QIN Z, AGUERRI I E, et al. Beyond transmitting bits: context, semantics, and task-oriented communications [EB/OL]. [2023-01-25]. https://arxiv.org/abs/2207.09353

[2] LU K, ZHOU Q Y, LI R P, et al. Rethinking modern communication from semantic coding to semantic communication [J]. IEEE wireless communications, 2022,30(1): 1-13. DOI: 10.1109/MWC.013.2100642

[3] SHI G M, XIAO Y, LI Y Y, et al. From semantic communication to semantic-aware networking: model, architecture, and open problems [J]. IEEE communications magazine, 2021, 59(8): 44-50. DOI: 10.1109/MCOM.001.2001239

[4] SHANNON C E, WEAVER W, WIENER N. The mathematical theory of communication [J]. Physics today, 1950, 3(9): 31-32. DOI: 10.1063/1.3067010

[5] LI Y J, LIU S F, YANG J M, et al. Generative face completion [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 5892-5900. DOI: 10.1109/CVPR.2017.624

[6] BAO J, BASU P, DEAN M K, et al. Towards a theory of semantic communication [C]//Proceedings of 2011 IEEE Network Science Workshop. IEEE, 2011: 110-117. DOI: 10.1109/NSW.2011.6004632

[7] GUO S S, WANG Y H, ZHANG P. Signal shaping for semantic communication systems with A few message candidates [EB/OL]. [2023-01-25]. https://arxiv.org/abs/2202.02072

[8] SHAO Y L, GUNDUZ D. Semantic communications with discrete-time analog transmission: a PAPR perspective [EB/OL]. [2023-01-25]. https://arxiv.org/abs/2208.08342

[9] JIANG P W, WEN C K, JIN S, et al. Deep source-channel coding for sentence semantic transmission with HARQ [J]. IEEE transactions on communications, 2022, 70(8): 5225-5240. DOI: 10.1109/TCOMM.2022.3180997

[10] ZHOU Q Y, LI R P, ZHAO Z F, et al. Adaptive bit rate control in semantic communication with incremental knowledge-based HARQ [EB/OL]. [2023-01-25]. https://arxiv.org/abs/2203.06634

[11] GUO J J, WEN C K, JIN S. Eliminating CSI feedback overhead via deep learning-based data hiding [J]. IEEE journal on selected areas in communications, 2022, 40(8): 2267-2281. DOI: 10.1109/JSAC.2022.3180806

[12] GARCIA-GARCIA A, ORTS-ESCOLANO S, OPREA S, et al. A review on deep learning techniques applied to semantic segmentation [EB/OL]. [2023-01-25]. https://arxiv.org/abs/1704.06857

[13] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241. DOI: 10.1007/978-3-319-24574-4_28

作 者 简 介

姜培文，东南大学移动通信国家重点实验室在读博士研究生；研究方向为智能通信物理层关键技术、语义通信等。

韩瑜，东南大学移动通信国家重点实验室副教授；主要研究方向为大规模MIMO无线传输理论与关键技术、智能超表面辅助移动通信等。

金石，东南大学移动通信国家重点实验室教授、博士生导师，教育部“长江学者奖励计划”特聘教授，国家自然科学基金杰出青年科学基金获得者，国家“万人计划”科技创新领军人才；主要研究方向为移动通信理论与关键技术、物联网理论与关键技术，以及人工智能在无线通信中的应用等；发表学术论文400余篇，获授权国际/国家发明专利50余项，出版专著2部、教材1部。

李潇，东南大学移动通信国家重点实验室教授、博士生导师；主要研究方向为智能超表面辅助的移动通信、人工智能在无线通信中的应用等。

基于深度联合信源信道编码的CSI反馈技术

Deep Joint Source-Channel Coding Based CSI Feedback

许佳龙 /XU Jialong，陈为/CHEN Wei，艾渤/AI Bo
（北京交通大学，中国 北京100044）
(Beijing Jiaotong University, Beijing 100044, China)

DOI：10.12142/ZTETJ.202302007
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1649.026.html
网络出版日期：2023-04-12
收稿日期：2023-01-16

摘要：提出了一种基于深度联合信源信道编码的信道状态信息（CSI）反馈方法。该方法使用非线性编码对原始的CSI信息进行降维，之后使用多层网络生成信道输入符号，并利用注意力机制实现了针对信道噪声的自适应功能。与现有CSI压缩反馈方法相比，得益于信源和信道的深度联合编码，该方法可以在有限的带宽下获得更好的预编码任务的性能。此外，所提的方法可使用近似量化方法将复信道输入符号转换为有限的星座点符号，能够与现代移动通信系统有效兼容。

关键词：联合信源信道编码；深度学习；CSI反馈；预编码

Abstract: A deep joint source-channel coding based channel state information (CSI) feedback method is proposed. The proposed method uses nonlinear coding to reduce the dimensionality of the original CSI information and a multi-layer network to generate channel input symbols, and employs an attention mechanism to realize the adaption for channel noise. Compared with existing compression-based CSI feedback methods, the proposed method can obtain better performance of the precoding task under limited bandwidth, which benefits from deep joint source-channel coding. Moreover, the proposed method can employ approximation quantization to convert infinite channel input symbols to finite quantization constellation symbols, which is effectively compatible with modern mobile communication systems.

Keywords: joint source-channel coding; deep learning; CSI feedback; precoding

为服务于全息通信和元宇宙等具有海量通信吞吐需求的新兴智能应用，6G通信系统应具备超大带宽和超高频谱效率[1-4]。随着毫米波以上频段的使用，无线通信的可用带宽进一步增加。然而，在毫米波和太赫兹频率下，传播损耗将变得更加严重，这会导致更高的通信功耗或更小的小区覆盖半径。为提高通信频谱效率，多输入多输出（MIMO）技术和大规模MIMO技术已经成为4G和5G的关键技术。超大规模MIMO亦有望成为6G的关键技术[5]。

典型的大规模MIMO场景为具有大量天线的基站同时服务多个用户设备。为了充分利用MIMO系统的多天线优势，基站侧需要掌握瞬时下行信道状态信息（CSI）。在时分双工模式下，基站通过用户发送的导频信号估计上行CSI，并利用信道互易性推断出下行CSI。在频分双工模式下，由于上下行链路工作在不同的频率上，不再满足信道互易性，因此需要通过三步交互来获取下行CSI：首先基站向用户发送导频信号，然后用户根据导频信号估计下行CSI，最后用户将估计的下行CSI反馈给基站。然而，这种CSI反馈机制不可避免地占用了部分通信上行资源，挤占了原本可用于数据传输的上行资源块。

为了减少反馈开销，5G采用基于码本的方法，使用反馈码本索引的方式代替反馈下行CSI的全信息的方式[6]。然而，随着对通信吞吐量需求的持续提升，天线数量不断增加，码本空间急剧扩大，反馈开销急速提升。通过利用信道特性对下行CSI进行压缩是解决这一问题的关键。由于信道在变换域的近似稀疏特性，基于压缩感知的方法可对稀疏的信道表示进行压缩，以减少反馈开销[7]。然而，在实际场景下，信道在变换域并不能满足严格稀疏的条件，降低了压缩感知方法的性能。此外，现有基于压缩感知的重建算法中的迭代过程通常非常耗时，难以在计算受限或严苛的时延要求场景下应用。

近年来，随着深度学习技术在计算机视觉和自然语言处理的卓越表现，深度学习被引入对CSI信息的压缩中[8-10]。这些工作被视为基于分离源信道编码方案的通信系统的信源

基金项目：国家自然科学基金（62122012、62221001、61961130391、62171380）；北京市自然科学基金（L202019、L211012）

编码模块。该方案通常假设信道编码模块和调制模块能够保证完美传输，即可以根据反馈信道质量自适应调整调制编码方式，成功传输所有的反馈码字。然而，基于分离信源信道编码的CSI反馈有如下缺点：首先，理论上，在有限码长下分离信源信道编码方案已被证明不如联合源信道编码方案[11]。其次，分离信源信道编码方案在真实无线场景中具有“悬崖效应”[12]。如果实际反馈信道条件比预期的信道条件差，并且超出所采用的信道编码方案的能力，则CSI的重建质量会急剧下降。在这种情况下，基站无法使用恢复的CSI进行后续处理。然而，在实际信道条件变得比预期信道条件更差的情况下，联合信源信道编码方案也能提供平稳的性能下降，这使得恢复的CSI对于基站的后续执行过程仍然有价值。最后，混合自动请求重传技术尽管可以补偿由信道条件不匹配引起的信道译码错误，但增加了额外的反馈开销，并为CSI反馈任务带来了延迟问题。因此，有必要研究基于深度联合信源信道编码的CSI反馈方案，以减少上行反馈资源占用，解决当前CSI反馈任务中的潜在问题。

1 CSI反馈方案设计

1.1 CSI反馈系统模型

本文考虑频分双工模式下的大规模多天线正交频分复用（OFDM）系统。其中，基站侧部署N_t根天线，用户侧部署单天线，上行链路和下行链路均使用N_c个子载波。假设完美的上行CSI和下行CSI可以分别在基站侧和用户侧通过基于导频的训练获得。基于深度联合信源信道编码的CSI反馈网络设计如图1所示。下行CSI和上行CSI分别表示为$\boldsymbol{H}_d \in \mathbb{C}^{N_c \times N_t}$、$\boldsymbol{H}_u \in \mathbb{C}^{N_c \times N_t}$。位于用户侧的深度联合信源信道编码器将下行CSI信息$\boldsymbol{H}_d$和已知的上行链路的信噪比信息μ编码为复向量$\boldsymbol{s}$：

$$\boldsymbol{s} = f_{\varphi}(\boldsymbol{H}_d, \mu) \in \mathbb{C}^k, \tag{1}$$

其中，φ表示深度联合信源信道编码器的参数集。复向量$\boldsymbol{s} = [s_1, s_2, \cdots, s_k]^T$被OFDM映射模块映射至$k$个子载波上，其中，$s_i$是承载在第$i$个子载波上的符号。假设一个子载波上的一个符号的平均功率为P，对编码向量$\boldsymbol{s}$进行功率归一化，即$1/k\,\mathbb{E}(\boldsymbol{s}\boldsymbol{s}^*) = P$，以满足用户侧的发射功率约束。

基站接收到的第i个子载波上的反馈信号可表示为：

$$\boldsymbol{y}_i = \boldsymbol{h}_u^i s_i + z_i, \tag{2}$$

其中，复向量$\boldsymbol{y}_i \in \mathbb{C}^{N_t}$表示在基站侧$N_t$根天线第$i$个子载波上接收到的符号，复向量$\boldsymbol{h}_u^i \in \mathbb{C}^{N_t}$表示在第$i$个子载波上的上行信道，复向量$z_i \in \mathbb{C}^{N_t}$为噪声向量。噪声向量$z_i$的分布为$\mathcal{CN}(0, \sigma^2 \boldsymbol{I})$，其中，$\sigma^2$表示噪声功率。

在基站侧天线接收到s_i的多个带噪声的副本后，执行最大比合并算法以获取在基站侧恢复的$\hat{s}_i$。该过程可表示为：

$$\hat{s}_i = \boldsymbol{w}_i^* \boldsymbol{y}_i, \tag{3}$$

其中，合并向量$\boldsymbol{w}_i$为$\boldsymbol{h}_u^i / \left\| \boldsymbol{h}_u^i \right\|_2$。随后，各子载波上的接收符号合并为复向量$\hat{\boldsymbol{s}} = [\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_k]^T$。这里，复向量$\hat{\boldsymbol{s}}$表示基站侧对信道输入复向量$\boldsymbol{s}$的重构。随后，基站侧使用深度联合信源信道译码器将复向量$\hat{\boldsymbol{s}}$和信噪比信息μ映射为恢复的CSI信息$\hat{\boldsymbol{H}}_d$：

$$\hat{\boldsymbol{H}}_d = f_{\phi}(\hat{\boldsymbol{s}}, \mu) \in \mathbb{C}^{N_c \times N_t}, \tag{4}$$

其中，ϕ表示深度联合信源信道译码器的参数集。

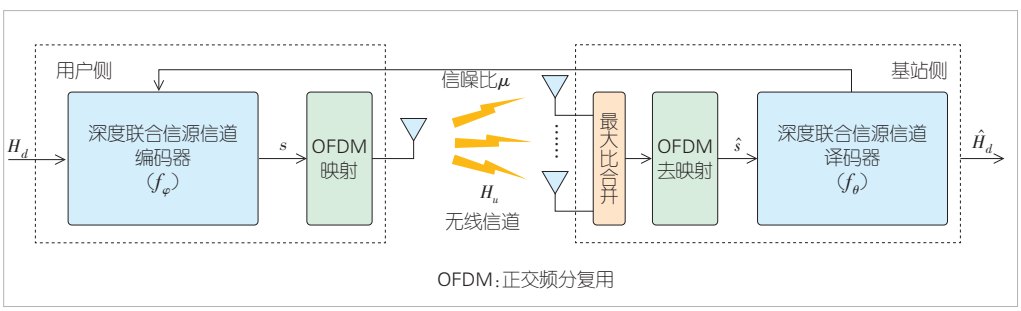
基于深度联合信源信道编码的CSI反馈，通过在确定的信道带宽k下最小化空间频率域的CSI失真，来优化参数集$\Theta = \{\varphi, \phi\}$。

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}}\, \mathbb{E}_{p(\mu)} \frac{1}{T} \sum_{i=1}^{T} \left\| \boldsymbol{H}_d^{(i)} - \hat{\boldsymbol{H}}_d^{(i)} \right\|_2^2, \tag{5}$$

其中，Θ^*表示最优参数集，$p(\mu)$表示信噪比的概率分布函数，$\boldsymbol{H}_d^{(i)}$表示在空间频率域的第i个训练集的样本，$\hat{\boldsymbol{H}}_d^{(i)}$表示基站侧在空间频率域重构的第$i$个训练集的样本，$T$表示训练数据集的大小。

1.2 CSI反馈网络

基于深度联合信源信道编码的CSI反馈网络如图2所示，分为深度联合信源信道编码网络和深度联合信源信道译码网络。在用户侧，现有CSI压缩反馈使用截断的2D-DFT变换



▲图1 基于深度联合信源信道编码的信道状态信息反馈系统

可初步减小CSI的反馈量。受此启发，深度联合信源信道编码网络首先使用3层卷积网络对原始的空频域CSI进行降维，之后基于文献[9]提出的网络，使用卷积层和全连接层对降维后的CSI信息进行特征提取。压缩后的$2k$维实数向量两两结合，重组为k维复向量，在进行功率归一化约束后，承载至OFDM子载波进行发射。在基站侧，深度联合信源信道译码网络从恢复的复向量中提取实部和虚部，并将其重组为实数向量，之后依次输入至全连接层、卷积层和残差块中，以恢复降维后的CSI信息；随后，使用3层转置卷积网络对降维后的CSI信息进行非线性变换，恢复原始的空频域CSI信息。由于传统的神经网络无法针对不同的信道条件进行动态调整，本文使用文献[13]中提出的即插即用的注意力特征（AF）模块为深度联合信源信道编码网络提供信道自适应功能。文献[14]对基于深度联合信源信道编码的CSI反馈网络的设计原则进行了详细的介绍。

1.3 评估指标

现有CSI反馈任务通常首选归一化均方误差（NMSE）对CSI反馈精度进行评估。然而，在实际通信系统中，基站在接收到CSI的反馈信息后通常会做进一步处理，以适配后续通信任务，如预编码任务。因此，对CSI反馈任务性能的评估应当更加关注后续任务的执行性能。为此，本文选取余弦相似度指标进行评估：

$$\rho = \mathbb{E}\left\{\frac{1}{N_c}\sum_{i=1}^{N_c}\frac{|\hat{\boldsymbol{h}}_d^{i\,H}\boldsymbol{h}_d^i|}{\|\hat{\boldsymbol{h}}_d^i\|_2\|\boldsymbol{h}_d^i\|_2}\right\}。\tag{6}$$

当基站侧使用$\hat{\boldsymbol{h}}_d^i/\|\hat{\boldsymbol{h}}_d^i\|_2, i=1,2,\cdots,N_c$作为预编码向量时，余弦相似度指标可以表示基站侧收到CSI反馈后执行预编码任务时的任务性能。

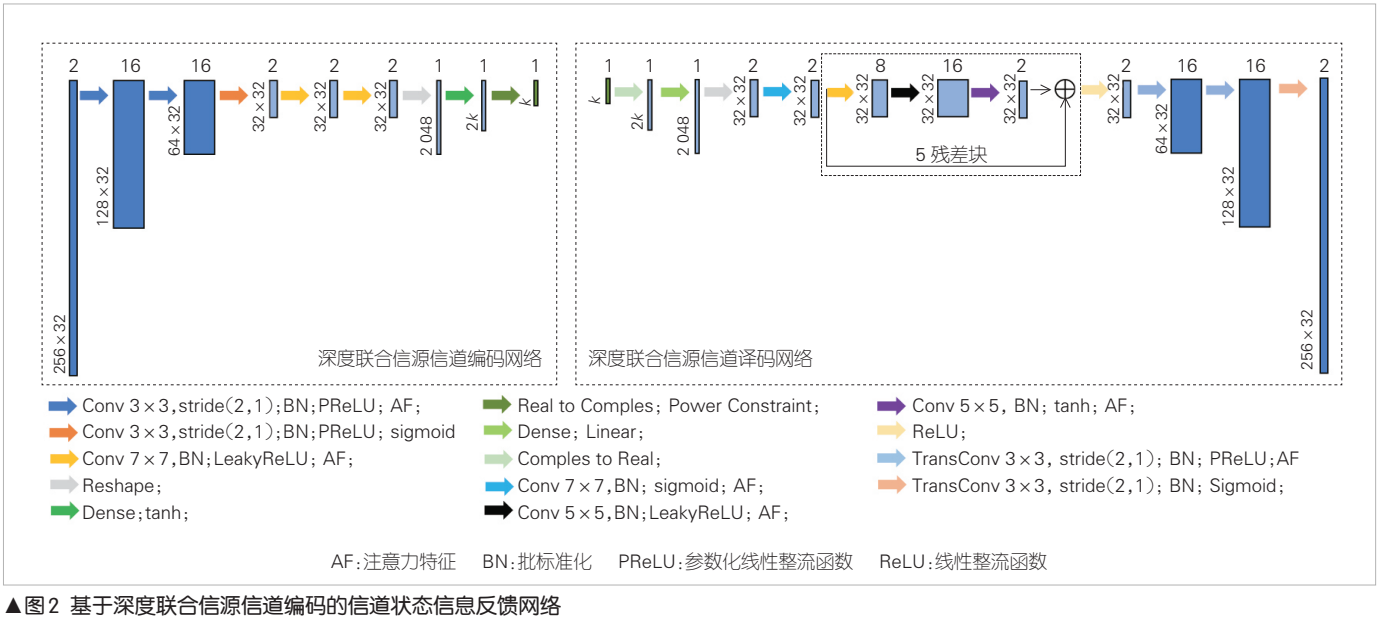
2 实验与结果分析

2.1 实验环境与数据集

本文使用TensorFlow及其高阶API Keras进行下述实验。根据第3代合作伙伴计划（3GPP）的TR 38.901[15]，上行链路CSI和下行链路CSI可由QuaDRiGa[16]生成。本文创建了一个开放的室内场景，其中下行链路的中心频率为5.2 GHz，上行链路的中心频率为5.4 GHz。在开放室内场景中，信道包含20个簇，每个簇包含15条子径。基站位于一个20 m×20 m大小的正方形区域中心。具有半波长天线空间的均匀线性阵列部署在基站侧。基站侧的天线数量为$N_t=32$，用户侧的天线数量为$N_r=1$。基站侧和用户侧均使用全向天线。基站的高度为3 m，用户的高度为1.5 m。上行链路和下行链路均使用$N_c=256$个子载波。CSI的训练集、验证集和测试集分别包含100 000、30 000、20 000个样本对。一个样本对包含一个下行链路CSI的样本和一个上行链路CSI的样本。

2.2 对比实验方案

本文采用分离信源信道编码方案作为对比方案。具体而言，我们使用文献[9]提出的CSINet+作为信源编码并使用5G上行控制信息（UCI）传输作为信道编码的方案。UCI的调



▲图2 基于深度联合信源信道编码的信道状态信息反馈网络

制方案包括二进制相移键控（BPSK）、4符号正交幅度调制（4 QAM）、16 QAM、64 QAM和256 QAM。为与本文所提的深度联合信源信道编码网络进行公平的比较，分离信源信道编码方案需要调整信源编码码率、信道编码码率及调制方式，以匹配反馈带宽。

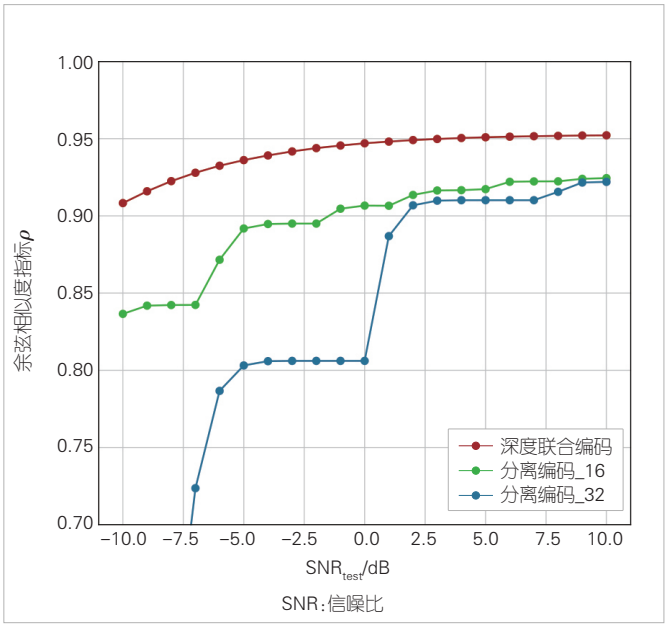
2.3 仿真结果分析

对于本文提出的基于深度联合信源信道编码的CSI反馈方法，在训练阶段，批大小被设置为200。Adam优化器首先以10^{-3}的学习率初始化。当损失函数在20个周期内均未下降时，学习率将衰减为原来的一半。学习率的下限设置为10^{-4}。为了使网络收敛，网络的训练周期被设置为500。

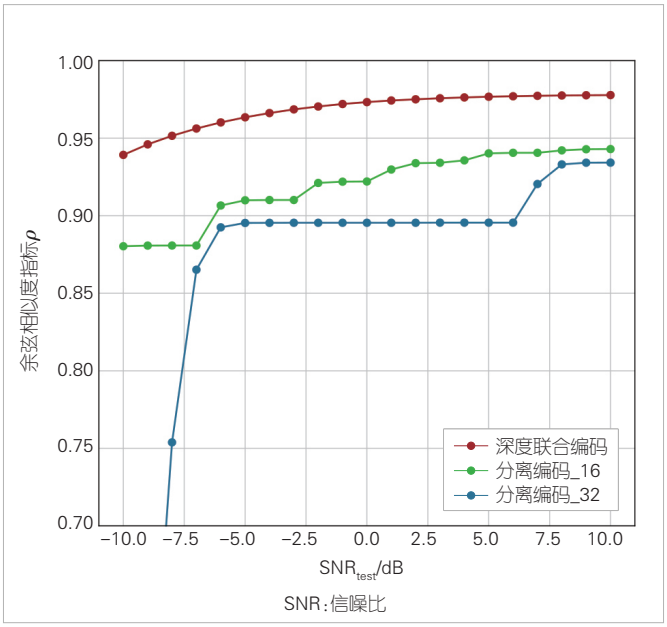
图3比较了在反馈带宽$k=16$时，基于深度联合信源信道编码的CSI反馈和基于分离编码的CSI反馈的余弦相似度性能。其中，分离编码_16、分离编码_32分别表示使用CSINet+编码网络的输出维度分别为16和32时分离编码的性能。当$\mathrm{SNR}_{\mathrm{test}}\in[-10,10]$ dB时，深度联合编码的性能比分离编码_16的性能高出0.03～0.07；当$\mathrm{SNR}_{\mathrm{test}}\in[-7,10]$ dB时，比分离编码_32的性能高出0.03～0.2，如图4所示。因此深度联合编码的性能优势在低信噪比下更加显著。

图4比较了在反馈带宽$k=32$时，基于深度联合信源信道编码的CSI反馈和基于分离编码的CSI反馈的余弦相似度性能。当$\mathrm{SNR}_{\mathrm{test}}\in[-10,10]$ dB时，深度联合编码的性能比分离编码_32的性能高出0.035～0.058；当$\mathrm{SNR}_{\mathrm{test}}\in[-7,10]$ dB时，比分离编码_64的性能高出0.044～0.091。当反馈带宽从$k=16$增长到$k=32$时，深度联合编码的性能仍然优于分离编码的性能。

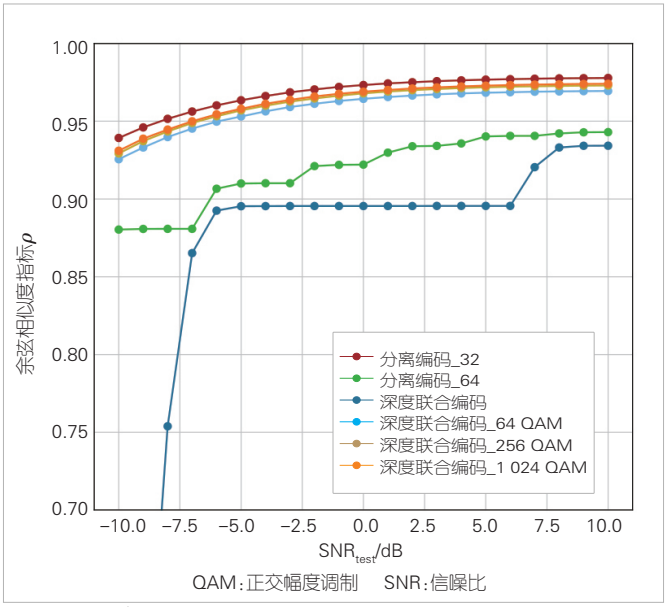
上述实验假设深度联合编码的发送端可以发送任何复数值。然而，现有的移动通信系统使用数字调制方式进行发射，其发射值为固定的离散星座点。本实验将使用近似量化的方法，将深度联合编码器编码后的复数符号映射至距离最近的离散星座点后，再进行发射，以使深度联合编码方法与现有移动通信系统兼容。图5分别展示了在反馈带宽$k=32$的条件下将深度联合编码方法分别量化为64 QAM、256 QAM和1 024 QAM时的性能。可以看到，随着可量化星座点数量的不断减少，深度联合编码的性能不断下降。例如，在$\mathrm{SNR}_{\mathrm{test}}\in[-10,10]$ dB时，量化为1 024 QAM的深度联合编码的性能比未使用星座点量化的深度联合编码的性能低约0.008；量化为256 QAM的深度联合编码的性能与量化为1 024 QAM的深度联合编码的性能近似；量化为64 QAM的深度联合编码的性能则比量化为256 QAM的深度联合编码的性能低约0.005。即便深度联合编码使用量化星座点发射带来了一定程度的性能损失，然而，量化等级最低的深度联合编码_64 QAM的性能仍然优于分离编码的性能。需要说明的是，本实验仅使用了最基本的量化方法，若能够将星座点量化过程考虑到训练过程中，或者针对量化星座点的位置进行专门的设计，量化后的性能损失可进一步降低。



▲图3 在反馈带宽$k=16$时，基于深度联合信源信道编码的信道状态信息(CSI)反馈和基于分离编码的CSI反馈性能对比



▲图4 在反馈带宽$k=32$时，基于深度联合信源信道编码的信道状态信息(CSI)反馈和基于分离编码的CSI反馈性能对比



▲图5 在反馈带宽$k = 32$时，基于深度联合信源信道编码的信道状态信息(CSI)反馈的数字星座点性能对比

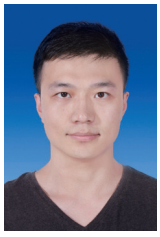
3 结束语

本文提出了一种基于深度联合信源信道编码的CSI反馈方法。该方法使用非线性变换网络对原始CSI信息进行初步降维后，使用深度学习网络对CSI信息进行特征提取，并使用注意力特征模块实现了对信道的自适应功能。不同于现有方法使用CSI恢复精度作为性能评估指标，针对CSI反馈的后续预编码任务，本文使用余弦相似度指标对基站侧使用CSI反馈进行预编码的性能进行了评估。相比于分离信源信道编码方法，本文提出的深度联合信源信道编码方法能够有效提升任务性能。此外，为使所提出的方法有效兼容现有移动通信系统，本文使用近似量化将编码后的复数符号映射至距离最近的数字星座点。实验结果显示，量化后的深度联合信源信道编码方法虽然会导致一定程度的性能下降，但其性能仍然远优于基于分离信源信道编码的CSI反馈性能。

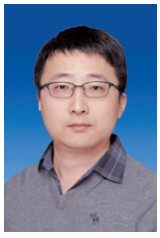
参考文献

[1] 葛建华，李靖. 对6G关键技术发展的思考 [J]. 中兴通讯技术, 2021, 27(2): 45−48. DOI:10.12142/ZTETJ.202102010
[2] 王海明, 陈祎祎. 6G：跨频段协同通信 [J]. 中兴通讯技术, 2022, 27(4): 42−43. DOI:10.12142/ZTETJ.202204009
[3] 谢峰, 王菲, 刘汉超. 面向6G的多频段智能融合组网 [J]. 中兴通讯技术, 2022, 27(4): 25−30. DOI:10.12142/ZTETJ.202204006
[4] 韩书君, 董晴, 许晓东. 面向6G全场景的多频段协同覆盖扩展技术 [J]. 中兴通讯技术, 2022, 27(4): 31−35. DOI:10.12142/ZTETJ.202204007
[5] 张平, 牛凯, 田辉, 等. 6G移动通信技术展望 [J]. 通信学报, 2019, 40(1): 141−148. DOI: 10.11959/j.issn.1000−436x.2019022
[6] 3GPP. Physical layer procedures for data: 3GPP TS 38.214 [S]. 2020
[7] QIN Z J, FAN J C, LIU Y W, et al. Sparse representation for wireless communications: a compressive sensing approach [J]. IEEE signal processing magazine, 2018, 35(3): 40−58. DOI: 10.1109/MSP.2018.2789521
[8] WEN C K, SHIH W T, JIN S. Deep learning for massive MIMO CSI feedback [J]. IEEE wireless communications letters, 2018, 7(5): 748−751. DOI: 10.1109/LWC.2018.2818160
[9] GUO J J, WEN C K, JIN S, et al. Convolutional neural network−based multiple−rate compressive sensing for massive MIMO CSI feedback: design, simulation, and analysis [J]. IEEE transactions on wireless communications, 2020, 19(4): 2827−2840. DOI: 10.1109/TWC.2020.2968430
[10] MASHHADI M B, YANG Q Q, GÜNDÜZ D. Distributed deep convolutional compression for massive MIMO CSI feedback [J]. IEEE transactions on wireless communications, 2021, 20(4): 2621−2633. DOI: 10.1109/TWC.2020.3043502
[11] KOSTINA V, VERDÚ S. Lossy joint source−channel coding in the finite blocklength regime [C]//Proceedings of 2012 IEEE International Symposium on Information Theory Proceedings. IEEE, 2012: 1553−1557. DOI: 10.1109/ISIT.2012.6283533
[12] SKOGLUND M, PHAMDO N, ALAJAJI F. Hybrid digital−analog source−channel coding for bandwidth compression/expansion [J]. IEEE transactions on information theory, 2006, 52(8): 3757−3763. DOI: 10.1109/TIT.2006.878212
[13] XU J L, AI B, CHEN W, et al. Wireless image transmission using deep source channel coding with attention modules [J]. IEEE transactions on circuits and systems for video technology, 2022, 32(4): 2315−2328. DOI: 10.1109/TCSVT.2021.3082521
[14] XU J L, AI B, WANG N, et al. Deep joint source−channel coding for CSI feedback: an end−to−end approach [J]. IEEE journal on selected areas in communications, 2023, 41(1): 260−273. DOI: 10.1109/JSAC.2022.3221963
[15] JAECKEL S, RASCHKOWSKI L, BÖRNER K, et al. QuaDRiGa−quasi deterministic radio channel gene rator, user manual and documentation [R]. 2021
[16] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz: 3GPP 38.901 V16.1.0 [R]. 2020

作 者 简 介



许佳龙，北京交通大学在读博士研究生；主要研究方向为基于深度学习的无线编码；发表论文4篇。



陈为，北京交通大学教授、博士生导师，英国剑桥大学博士、博士后；近年来主要研究兴趣包括智慧无线通信、高速移动通信等，主持国家自然科学基金委优秀青年基金等项目；发表论文100余篇。

艾渤，北京交通大学教授、博士生导师、电子信息工程学院院长，轨道交通控制与安全国家重点实验室常务副主任，IEEE/IET Fellow，IEEE VTS杰出讲师，国家自然科学基金委创新群体负责人，国家杰青、优青、牛顿高级学者基金获得者，国家6G总体组专家，工信部5G AIA综合轨道交通工作组组长，中国电子学会智慧交通信息工程分会副主任兼秘书长，入选2020、2021、2022年全球前2%顶尖科学家；主要从事宽带移动通信和轨道交通专用移动通信的研究；发表IEEE期刊论文160余篇。

语音信源的语义编码传输方法研究

Semantic Coded Speech Transmission

牛凯/NIU Kai，姚圣时/YAO Shengshi，戴金晟/DAI Jincheng
（北京邮电大学泛网无线通信教育部重点实验室，中国 北京 100876）
(The Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China)

DOI：10.12142/ZTETJ.202302008
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.tn.20230411.1437.016.html
网络出版日期：2023-04-12
收稿日期：2023-01-25

摘要：作为一种新兴的通信范式，语义通信在有效提升端到端传输性能方面展现出巨大的潜力。语音信源的语义编码传输研究方法分为两大类：基于信号波形的语音语义编码传输和生成式语音语义编码传输。在基于信号波形的语音语义编码传输中，现有方案的语义信息无法度量，编码效率低。基于非线性变换的语音语义信源编码方案通过对语音的语义特征进行变分建模，有效衡量语义特征内容复杂度，并引入信源信道联合编码，使语义编码传输更加高效可靠。针对生成式语音语义编码传输方法，分析了其优势、挑战，以及研究前景。

关键词：语音编码；语义编码传输；语义通信

Abstract: As an emerging communication paradigm, semantic communications has shown great potential in effectively boosting end-to-end transmission performance. The problem of semantic coded speech transmission is investigated, which can be divided into two main categories: waveform-based and generative semantic speech coded transmission methods. In waveform-based semantic speech coding and transmission, existing solutions cannot quantify semantic information effectively, resulting in low efficiency. The proposed speech semantic coding scheme based on nonlinear transform measures the complexity of semantic features through variational modeling and introduces joint source-channel coding, making semantic coded transmission more efficient and reliable. The advantages, challenges and future research prospects of generative semantic speech coded transmission are summarized.

Keywords: speech coding; semantic coded transmission; semantic communication

过去70年，通信技术取得了突飞猛进的发展。基于经典信息理论[1]，先进的编码技术包括信源编码、信道编码，已经逼近了信源熵/率失真函数以及信道容量。然而，经典信息论只研究语法信息，致力于信息传输的有效性和可靠性。从认识论观点看，信息分为3个层次：语法、语义和语用[1]。C. E. SHANNON和W. WEAVER[2]指出了通信的另一层面，即通信的语义问题以及有效性问题。得益于人工智能技术的发展，之前未能被解决的信息的语义及语用层次重新得到了研究人员的关注。北京邮电大学张平院士深入分析语义信息特征，提出语义基（Seb）模型[2]，指出语义信息可以用Seb进行表征。张平院士提出“智简（Intellicise）”理念[4]，进一步提出模型驱动的语义通信框架，实现通信系统由传输比特演进为传输经语义处理得到的新特征，例如Seb等。牛凯等深入研究了从经典通信到语义通信的范式转变，提出了语义通信系统的基本框架[5-7]。其他团队也对语义通信开展了多项研究工作：石光明等提出了语义通信的基本框架[8]，秦志金等针对语义通信系统探讨了深度学习赋能的语义通信理论、框架和系统模型[9]，北京交通大学、帝国理工大学团队分析了信源信道联合编码在语义通信中的应用[10]。本文总结并回顾已有的面向语音信源的语义编码传输方法，并指出未来的研究方向。

语音编码一直是移动通信中的基础问题，其目的是获得语音信号的紧凑表征，从而实现高效传输。波形编码和参量编码是传统语音编码的两大主流方法。波形编码旨在编码端对波形信号进行数字化，并在解码端尽可能重构波形，以脉冲编码调制（PCM）、差分脉冲编码调制（DPCM）和自适应差分脉冲编码调制（ADPCM）[11]为代表。通过传输声学模型的系数，参数编码以低比特率成本提供了可比较的重建质量。线性预测编码（LPC）引入了信源-滤波器模型，它的参数通过线性来预测，并根据语音样本来确定。码激励线性预测（CELP）[12]是一种结合波形编码和LPC模型的混合编码方法，适用于窄频带和中频带语音编码。Opus是近几年性能较好的开源音频编码器[13]，支持动态可调速率以及从窄带到全带宽的音频编码。其中，语音编码采用了基于LPC的SILK和能量约束重叠变换（CELT）编码器。

基金项目：国家自然科学基金（92067202、62293481、62071058）

现有的通信系统采用传统的音频编码器级联信道编码的方法，但其编码信息与语义无关，且易出现差错传播的情况。从语义信息传输的角度来设计传输方法，可以在语义空间内纠正信道传输引起的偏差。这种方法的针对性更强，最终可以提升端到端的编码传输效率。基于神经网络的语音编码方法已经有了初步研究，引入线性预测和基于变分自编码器的矢量量化（VQ-VAE）[14]，可以将语音特征以低速率传输。这些方法对语音信号的语义编码传输起到了指导作用。传统语音评价指标的不可微的性质导致无法直接用于语义编码模型优化，因此寻找合适的语音信号的语义传输的感知优化函数至关重要。

1 语义通信系统简介

语义通信系统的模型[5]如图1所示，参照C. E. SHANNON和W. WEAVER的思想[2]，该模型分为Level A技术级通信与Level B语义级通信两个层级，具体由信源、语义知识库、语义发射机、信号发射机、信道、信号接收机、语义接收机、信宿8个部分组成。

其中，语义知识库包括信道环境特征、信源及任务特征，可以对语义发射机/接收机、信号发射机/接收机起到指导作用。语义编码传输问题即如何在语义知识库的指导下进行语义发射机/接收机和信号发射机/接收机的编译码设计问题。

语音信源的语义传输任务通常分为两大类：1）面向听觉的语音信号传输；2）传递内涵文本信息的语音信号传输。两类任务的语义编码优化设计准则也不同：面向语音文本内容的传输以误词率、文本重合度为语义传输性能的主要评价指标；面向人的听觉的语义传输主要考虑的是人的主观感受。针对如何提高人的主观感受（或称为感知编码），研究者提出了客观语音质量评估（PESQ）[15]、ViSQOL[16]等语音感知质量评价指标。在语音质量的主观评价方面，除平均主观意见分（MOS）外，常见的还有多激励隐藏参考基准测试（MUSHRA）[17]方法。

2 面向语音的语义编码传输方法

语音信源的语义编码传输一般包括两类方法：基于波形的语义编码传输和生成式的语义编码传输。

2.1 基于波形的语义编码传输

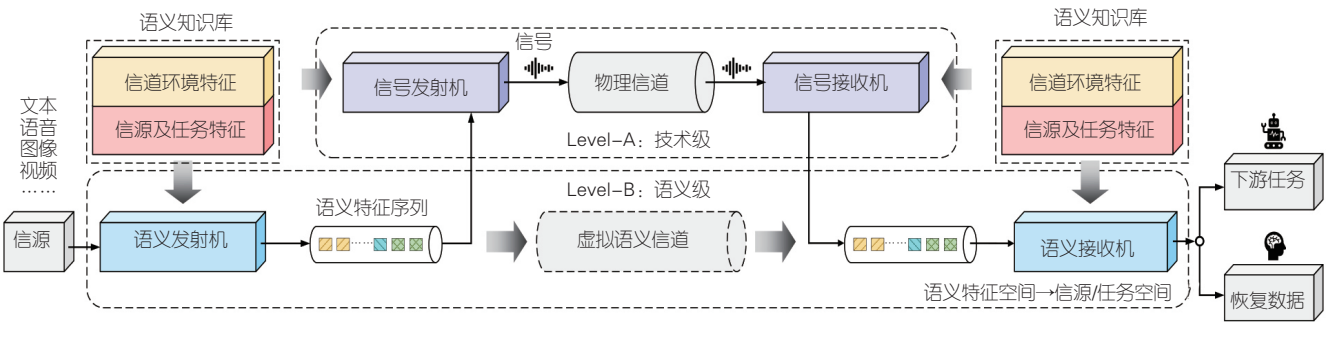
基于时域波形信号的语义编码传输方法直接从语音波形信号中提取语义特征并映射为信道中的传输符号。此类方法属于信源信道联合编码范畴。利用深度神经网络的信源信道联合编码，重建质量随信噪比的降低而柔和降低，不存在悬崖效应。QIN Z. J.等针对语音信源，构建了一种面向语音信源的深度语义通信网络（DeepSC-S）[18]，提高了语音信号的重建质量。该方法主要利用卷积神经网络与“压缩-激励”模块提取语音信源语义特征，实现从信源数据到信道中传输符号的映射。该编码过程与信道条件有关，达到信源信道联合编码的效果。

语音信源向量$\boldsymbol{s}\in\mathbb{R}^m$经过编码神经网络$f_e$映射，得到语义编码向量$\boldsymbol{x}\in\mathbb{R}^k$，满足编码约束$\boldsymbol{x}=f_e(\boldsymbol{s};\phi_e)$。其中，$\phi_f$为DeepSC-S编码器模型参数。经过加性高斯白噪声（AWGN）信道后，解码器输入语义向量为$\boldsymbol{y}=\boldsymbol{x}+\boldsymbol{n}$。其中，$\boldsymbol{n}\sim\mathcal{N}(0,\sigma_n^2\mathbf{I}_k)$是独立同分布的高斯噪声向量，$\sigma_n^2$是噪声方差，$\mathbf{I}_k$是单位阵。接收端译码器由卷积神经网络构成，经过解码，得到输出向量$\hat{\boldsymbol{x}}=f_d(\boldsymbol{y};\phi_d)$。最后，整个模型联合进行端到端联合训练，损失函数为：

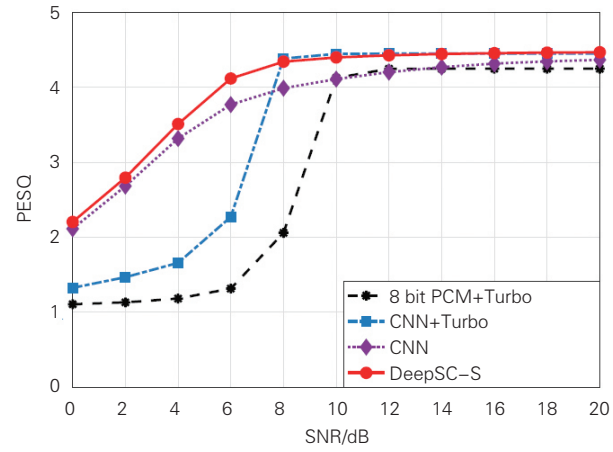
$$(\phi_e,\phi_d)=\arg\min_{\phi_e,\phi_d}\mathbb{E}_{s\sim p(s)}\mathbb{E}_{\hat{s}\sim p(\hat{s}|s)}\left[d_s(\boldsymbol{s},\hat{\boldsymbol{s}})\right],\tag{1}$$

其中，$d_s(\boldsymbol{s},\hat{\boldsymbol{s}})$为信源和信宿语音信号的失真，通常采用均方误差（MSE）表示。

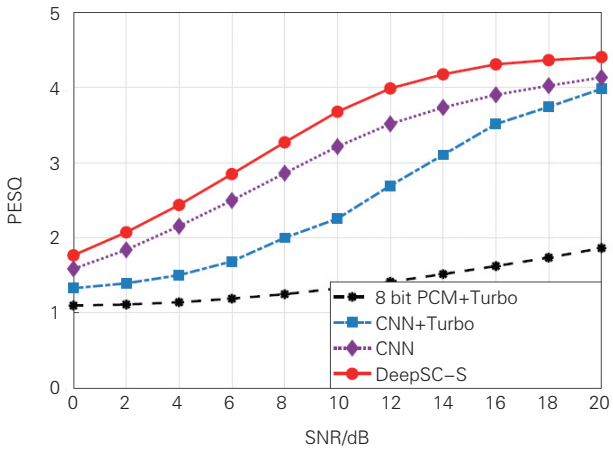
DeepSC-S的PESQ-SNR曲线如图2所示，在不同信道中



▲图1 语义通信的系统模型



（a）AWGN信道



（b）瑞利衰落信道

AWGN：加性高斯白噪声
CNN：卷积神经网络
DeepSC–S：面向语音信源的深度语义通信网络
PCM：脉冲编码调制
PESQ：客观语音质量评估
SNR：信噪比

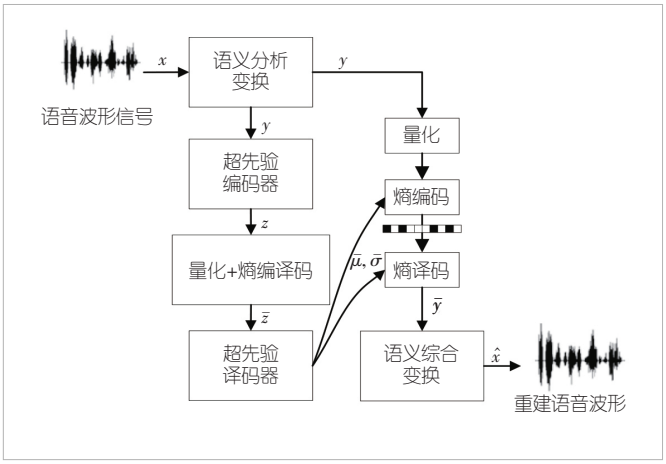
▲图2 DeepSC–S的PESQ–SNR曲线[18]

的相同信噪比下，DeepSC–S（红色）的性能优于传统分离式编码PCM+Turbo码的性能。

然而，DeepSC–S对不同帧语音信号的编码速率是相同的，且提取的语义特征不明确，无法区分不同语音帧上的内容差异、语义差异，即无法从细粒度上刻画语义信息量的多少，因此难以获得接近于熵编码的压缩增益。这导致传输带宽较高，且编码效率较低。

为了进一步提高语音信号的语义传输效率，我们借鉴了非线性变换编码的思想[19–21]，设计了基于非线性变换的语音语义信源编码方案[22]和语义编码传输方案[23]。下面我们简要介绍两种方案的基本原理。

基于信源变换编码的流程如图3所示。首先，语音经过分帧预加重等预处理后再经波形分析变换就可得到语义特征



▲图3 基于非线性变换的语音信源语义编码流程图

向量$\boldsymbol{y}=g_{a,\phi_g}(\boldsymbol{x})$，其中$\phi_g$为波形分析网络的模型参数。其次，一对超先验编译码器将语义特征向量$\boldsymbol{y}$变分建模为简单的高斯分布，从而使用算数编码器根据概率值对语义向量$\boldsymbol{y}$进行熵编码。在收端，波形综合变换网络将恢复的语义特征向量$\bar{\boldsymbol{y}}$重构为语音波形信号$\hat{\boldsymbol{x}}=g_{s,\psi_g}(\bar{\boldsymbol{y}})$，其中$\psi_g$为波形综合网络的模型参数。

在端到端优化时，为解决量化不可导的问题，引入了取值范围为$\left(-\frac{1}{2},\frac{1}{2}\right)$的均匀噪声，得到语义特征向量的近似量化形式$\tilde{\boldsymbol{y}}$，$\tilde{\boldsymbol{y}}=y+U\left(-\frac{1}{2},\frac{1}{2}\right)$。类似地，超先验向量$\boldsymbol{z}$也可近似量化为$\tilde{\boldsymbol{z}}$，从而得到了熵模型：

$$p\left(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{z}};\ \psi_h\right)=\prod_i\left(\mathcal{N}\left(\mu_i,\sigma_i^2\right)*\mathcal{U}\left(-\frac{1}{2},\ +\frac{1}{2}\right)\right), \tag{2}$$

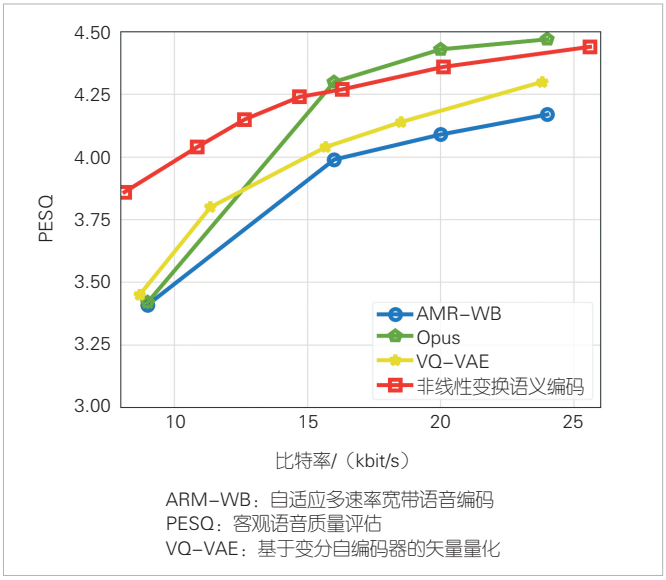
其中，$\boldsymbol{z}=h_{a,\phi_h}(\boldsymbol{y})$，$\phi_h$为超先验编码器的参数集合，$\mathcal{N}(\mu,\sigma^2)$表示均值为$\mu$，方差为$\sigma^2$的高斯分布，$*$表示卷积。

变换编码的优化问题是一个有损信源编码问题。若语音语义传输是面向人主观听觉感知质量的，那么优化语音的主观感知质量比波形失真更加重要。总的来说，该问题可以转变为最小化真实的联合后验概率与实际概率密度分布的KL散度，即：

$$\mathbb{E}_{\boldsymbol{x}}D_{\mathrm{KL}}\left[q_{\tilde{\boldsymbol{y}},\tilde{\boldsymbol{z}}|\boldsymbol{x}}\|p_{\tilde{\boldsymbol{y}},\tilde{\boldsymbol{z}}|\boldsymbol{x}}\right]=$$
$$\mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\tilde{\boldsymbol{y}},\tilde{\boldsymbol{z}}\sim q_{\tilde{\boldsymbol{y}},\tilde{\boldsymbol{z}}|\boldsymbol{x}}}\left[-\log p\left(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{z}}\right)-\log p\left(\boldsymbol{z}\right)-\log p\left(\boldsymbol{x}|\tilde{\boldsymbol{y}}\right)+\log q\left(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{z}}|\boldsymbol{x}\right)\right], \tag{3}$$

其中，等式后大括号内第1、2项刻画了$\boldsymbol{y}$和$\boldsymbol{z}$的编码速率，第3项刻画了失真度量，第4项为求均值后为常数。

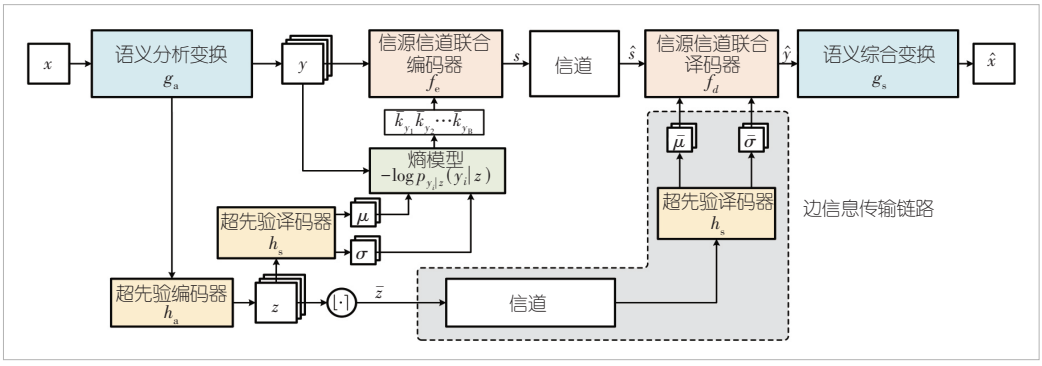
基于非线性变换的语义编码的率失真性能曲线如图4所示。我们可以发现，在低速率时，相较于自适应多速率宽带



▲图4 基于非线性变换的语音语义编码PESQ分数(取值为1.0～4.5)

语音编码（AMR-WB）[23]以及基于变分自编码器的矢量量化（VQ-VAE）神经网络编码，基于非线性变换的语义编码的感知质量优势明显；在高速率时，PESQ也与Opus性能相近。

在非线性变换语义信源编码的基础上，语音信源的语义-信源信道联合变换编码不用对语义特征向量量化并熵编码，而是将其映射为信道传输符号，其流程如图5所示。一方面，语义特征向量$\boldsymbol{y}$需要通过超先验编解码器将其变分建模为高斯分布。此步骤是为了用语义特征向量的概率分布及概率值来衡量该语音帧的内容复杂度，从而决定编码所需要的符号数。另一方面，利用Transformer[24]网络构成的信源信道联合编码器，可以将N帧的语义特征向量$\boldsymbol{y}^1,\cdots,\mathbf{y}^N$映射为信道中传输符号序列$\boldsymbol{s}=f_{e,\phi_f}\left(\boldsymbol{y}^1,\cdots,\mathbf{y}^N\right)$。其中，$\phi_f$为信源信道联合编码器的参数集合。每个语义特征向量最终的编码长度由熵模型给出。



▲图5 基于非线性变换的语音语义信源信道联合编码系统结构

对于传输问题，其优化函数转变为端到端的率失真优化问题，速率此时定义为传输带宽，训练的损失函数如公式（4）：

$$L=\mathbb{E}_{\mathbf{u}\sim p(\mathbf{u})}\left\{\lambda\left[\underbrace{-\eta\log P\left(\bar{s}|\bar{z}\right)}_{k_s}-\underbrace{\frac{\log P\left(\bar{z}\right)}{C_z}}_{k_z}\right]+d_s(u,v)\right\},\tag{4}$$

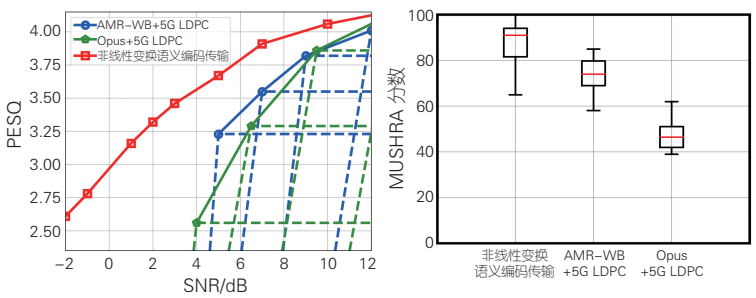
其中，$s=g_a\left(u;\phi_g\right)$，$z=h_a\left(s;\phi_h\right)$，$\bar{s}=Q(\mathbf{s})$与$\bar{z}=Q(z)$表示均匀量化，$v=g_s\left(\mathbf{r};\theta_g\right)$，$k_s$是语义隐向量$\boldsymbol{s}$的维度，$k_z$是边信息$\mathbf{z}$的维度，$C_z$是相应的信道容量，参数$\eta$控制语义信息编码速率，拉格朗日因子$\lambda$控制发送信息速率$k_s+k_z$与端到端的语义失真$d_s$之间的折中关系。

图5中基于Transformer网络的信源信道联合编/译码器包括4层8头注意力层，维度为1 024。语音数据来自于TIMIT（英文语音数据集）[25]，采样率为16 kHz。图6给出了不同信道下基于非线性变换的语义编码传输方法的仿真结果。图6（a）为AWGN信道在传输带宽为4 kHz的条件下，不同编码传输方法在PESQ评估指标下的端到端率失真曲线图以及主观得分箱线图；图6（b）为COST2100衰落信道[26]在传输带宽10 kHz、平均信噪比为2 dB的条件下，不同编码方案的PESQ评估指标下的端到端率失真曲线图及MUSHRA主观得分箱线图。其中，红色线为非线性变换语义编码传输方法的曲线，蓝色线和绿色线为传统分离式编码方法曲线。我们可以发现，在相同主观/客观感知质量指标下，相较于传统语音编码的AMR-WB、Opus编码器级联5G LDPC信道编码，以及DeepSC-S，基于变换编码的语音语义传输所需带宽更少，且在衰落信道中也表现出更加出色的鲁棒性。

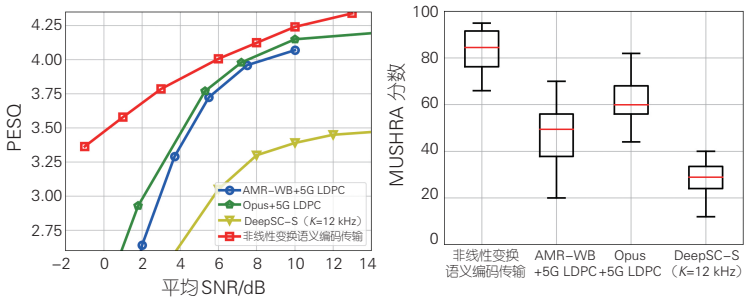
2.2 生成式语音语义编码传输

生成式语音语义编码传输的研究工作目前还较少，HAN T. X.等研究了上述所说的两种语音语义传输的范式[27]。在语音传输范式中，HAN T. X.等将卷积神经网络级联双向长短期记忆网络（LSTM）作为语义编码器，对语音帧的40组频率滤波器系数进行语义特征提取。同时，他们还借鉴了FastSpeech2[28]的做法，通过对齐模块预测每帧对应的子单词，并计算每个音

素的音高、功率等声学特征。在接收侧，文本特征、声学特征向量经对齐级联，由文本-语音综合器恢复语音频谱，最后声码器根据频谱重新生成语音波形，生成效果如图7所示。重建语音的MOS在AWGN信道中的各个信噪比下都优于DeepSC-S。



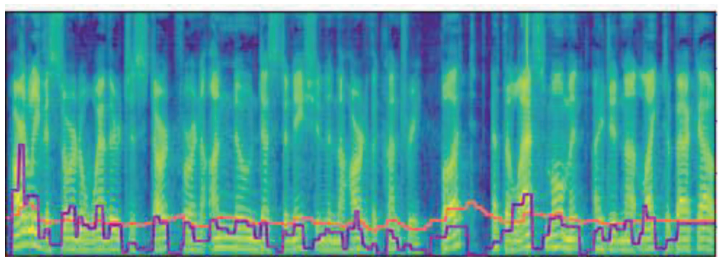
（a）AWGN信道（带宽为4 kHz）



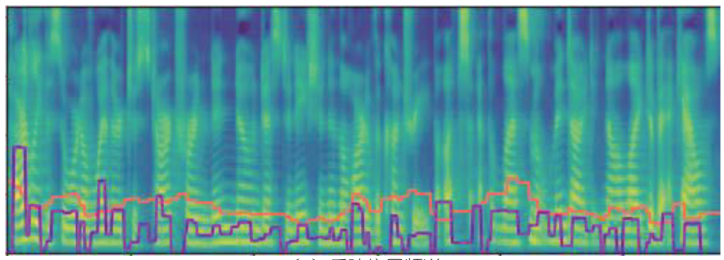
（b）COST2100信道（带宽为10 kHz、信噪比为2 dB）

AMR-WB：自适应多速率宽带语音编码
AWGN：加性高斯白噪声
DeepSC-S：面向语音信源的深度语义通信网络
LDPC：低密度奇偶校验码
MUSHRA：多激励隐藏参考基准测试
PESQ：客观语音质量评估
SNR：信噪比

▲图6 基于非线性变换的语音语义编码传输重建语音质量性能



（a）原始语音信号频谱



（b）重建信号频谱

注：橙色线为音高，紫色线为功率

▲图7 生成式语音语义编码传输效果[27]

生成式语音语义编码传输的优势在于所需带宽少，无须将语音波形信号的全部信息编码，利用人的听觉特性用少量的滤波器组在编码前就可将高采样率的波形信号转化为低采样率的声学特征信号，这样可以降低数据量。但根据数据处理定理，对语音信号的处理会导致部分信息的损失。如果接收侧关注编码语音特征以外的信息，则会导致语义传输的效果变差。另外，生成式的语义编码传输方法通过生成模型来重建语音波形，生成过程的复杂度、编译码器模型的训练难度都将给生成式语义编码传输方法的设计和实际应用带来挑战。

近两年，生成模型正在快速发展，人工智能生成内容（AIGC）的质量及自然度逼近人类。可以预见，基于生成式的语音编码传输也将对语音信号的语义编码设计带来重要影响。

3 结束语

本文中，我们对语音信源的语义编码传输方法进行了整理与总结：首先回顾了现有的传统语音编码技术，分析了语音语义编码传输的优化目标，具体而言，在面向人的感知体验的情境下，优化目标主要为提高人的感知质量；而在面向文本任务的情境下，优化目标主要为提升文本识别正确率。其次，我们将现有的语音语义编码传输方法分为两大类：一类为直接对语音信号样值序列进行编码，一类为对传输语音信号的声学特征进行编码和传输，在接收侧通过生成式声码器重建语音。最后，我们展望了未来生成模型应用于语音语义编码传输的优势及挑战。

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] BURKS A, SHANNON C, WEAVER W. The mathematical theory of communication [M]. USA: The University of Illinois Press, 1951
[3] ZHANG P, XU W, GAO H. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks [J]. Engineering, 2022, 8(1): 60-73. DOI: 10.1016/j.eng.2021.11.003
[4] ZHANG P, XU X D, DONG C, et al. Intellicise communication system: model-driven semantic communications [J]. The journal of China universities of posts and telecommunications, 2022, 29(1): 2-12. DOI: 10.19682/j.cnki.1005-8885.2022.2002
[5] NIU K, DAI J C, YAO S S, et al. A paradigm shift toward semantic communications [J]. IEEE communications magazine, 2022, 60(11): 113-119. DOI: 10.1109/MCOM.001.2200099

[6] DAI J C, ZHANG P, NIU K, et al. Communication beyond transmitting bits: semantics-guided source and channel coding [J]. IEEE wireless communications, 2022, (99): 1-8. DOI: 10.1109/MWC.017.2100705
[7] 牛凯, 戴金晟, 张平, 等. 面向6G的语义通信 [J]. 移动通信, 2021, 45(4): 85-90. DOI: 10.3969/j.issn.1006-1010.2021.04.014
[8] SHI G, XIAO Y, LI Y, et al. From semantic communication to semantic-aware networking: model, architecture, and open problems [J]. IEEE communications magazine, 2021, 59(8): 44-50. DOI: 10.1109/MCOM.001.2001239
[9] 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
[10] XU J L, TUNG T Y, AI B, et al. Deep joint source-channel coding for semantic communications [EB/OL]. [2023-01-20]. https://arxiv. org/abs/2211.08747
[11] VAUDREUIL G, PARSONS G W. Toll quality voice - 32 kbit/s adaptive differential pulse code modulation (ADPCM) MIME sub-type registration [EB/OL]. (2004-06-01) [2023-01-23]. https://www. semanticscholar. org/paper/Toll-Quality-Voice-32-kbit-s-Adaptive-Differential-Vaudreuil-Parsons/9236c8615a9b26f7c30a50c61178b7fb45b0652e
[12] BCKSTRM T. Speech coding: with code-excited linear prediction [M]. Germany: Springer, 2017
[13] VALIN J ,VOS K, TERRIBERRY T. Definition of the Opus audio codec [EB/OL]. (2012-09-01)[2023-01-25]. https://www.semanticscholar.org/paper/Definition-of-the-Opus-Audio-Codec-Valin-Vos/ce6784cb8deaf3a05d3395e2624c02e8796a30e5
[14] OORD D V A, VINYALS O, KAVUKCUOGLU K, et al. Neural discrete representation learning [EB/OL]. (2017-11-02)[2023-01-18]. https://arxiv.org/abs/1711.00937
[15] RIX A W, BEERENDS J G, HOLLIER M P, et al. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs [C]//Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal. IEEE, 2002: 749-752. DOI: 10.1109/ICASSP.2001.941023
[16] CHINEN M, LIM F S C, SKOGLUND J, et al. ViSQOL v3: an open source production ready objective speech and audio metric [C]//Proceedings of 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2020: 1-6. DOI: 10.1109/QoMEX48832.2020.9123150
[17] SCHOEFFLER M. Towards the next generation of web-based experiments: a case study assessing basic audio quality following the ITU-R recommendation BS. 1534 (MUSHRA) [EB/OL]. [2023-01-22]. https://www. semanticscholar. org/paper/Towards-the-Next-Generation-of-Web-based-%3A-A-Case-.-Schoeffler/908924842e5ec23f9dda290fdc1276606c4abbbd
[18] WENG Z Z, QIN Z J. Semantic communication systems for speech transmission [J]. IEEE journal on selected areas in communications, 2021, 39(8): 2434-2444. DOI: 10.1109/JSAC.2021.3087240
[19] BALLÉ J, CHOU P A, MINNEN D, et al. Nonlinear transform coding [J]. IEEE journal of selected topics in signal processing, 2020, 15(2): 339-353. DOI: 10.1109/JSTSP.2020.3034501
[20] 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
[21] YAO S S, XIAO Z X, WANG S X, et al. Variational speech waveform compression to catalyze semantic communications [EB/OL]. [2023-01-22]. https://arxiv.org/abs/2212.05294
[22] XIAO Z X , YAO S S, DAI J C, et al. Wireless deep speech semantic transmission [EB/OL]. (2022-11-04) [2023-01-18]. https://deepai. org/publication/wireless-deep-speech-semantic-transmission
[23] BESSETTE B, SALAMI R, LEFEBVRE R, et al. The adaptive multirate wideband speech codec (AMR-WB) [J]. IEEE transactions on speech and audio processing, 2002, 10(8): 620-636. DOI: 10.1109/TSA.2002.804299
[24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2023-01-18]. https://arxiv.org/abs/1706.03762
[25] GAROFOLO, JOHN S. TIMIT acoustic phonetic continuous speech corpus [EB/OL]. [2023-01-18]. https://catalog.ldc.upenn.edu/LDC93S1
[26] LIU L F, OESTGES C, POUTANEN J, et al. The COST 2100 MIMO channel model [J]. IEEE wireless communications, 2012, 19(6): 92-99. DOI: 10.1109/MWC.2012.6393523
[27] HAN T X, YANG Q Q, SHI Z G, et al. Semantic-preserved communication system for highly efficient speech transmission [EB/OL]. (2022-05-25) [2023-01-18]. https://arxiv.org/abs/2205.12727v1
[28] REN Y, HU C X, TAN X, et al. Fastspeech 2: fast and high-quality end-to-end text to speech [EB/OL]. (2022-11-04)[2023-01-18]. https://arxiv.org/abs/2006.04558

作 者 简 介

牛凯，北京邮电大学教授、中国电子学会信息论分会副主任委员；主要研究方向为信息论与极化码、智能信号处理；所提极化码高性能编译码算法成为5G标准主流方案，获中国电子学会科技奖自然科学一等奖；发表论文200余篇，申请国家发明专利80项。

姚圣时，北京邮电大学在读博士研究生；主要研究方向为信息论、信源信道编码和语义通信。

戴金晟，北京邮电大学副教授、博士生导师；从事信息通信、人工智能相关的科研和教学工作，研究方向为智能语义通信、信源信道编码；入选“中国科协青年人才托举工程”名单，曾获中国电子学会科学技术奖自然科学一等奖；发表论文60余篇，授权国家发明专利23项，出版专著1本。

智简语义通信的链路设计及关键技术研究

Link Design and Key Technologies of Intellicise Semantic Communication System

孙梦颖/SUN Mengying，熊华超/XIONG Huachao，
王怡宁/WANG Yining，韩书君/HAN Shujun，
许晓东/XU Xiaodong

（北京邮电大学，中国 北京 100876）
(Beijing University of Posts and Telecommunications, Beijing 100876, China)

DOI：10.12142/ZTETJ.202302009
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1438.018.html
网络出版日期：2023-04-11
收稿日期：2023-01-28

摘要：智简语义通信是一种模型驱动的语义通信新范式，融合了人工智能与通信技术，实现通信对象间高效的语义交互。提出了智简语义通信系统的链路结构和关键技术，从链路结构、语义模型增强、模型传输3个角度，实现语义通信系统的整体性能增强。提出了4种智简语义通信的关键技术，为未来6G赋能多种垂直行业和新场景应用提供了参考：语义知识图谱增强的智简通信技术通过增加语义知识这一信息维度，提升了语义知识恢复准确度和传输效率；语义知识图谱的云-边-端协同预缓存技术可以实现语义知识图谱的高效获取，辅助语义恢复性能；模型传输与部署技术可以实现模型、网络资源与终端能力的有效适配；语义模型传输与恢复级联过程的部署成功率为大规模语义模型传输及资源部署提供理论依据。

关键词：智简语义通信；模型处理；模型传输；语义知识库

Abstract: Intellicise semantic communications is a new paradigm of model-driven semantic communication, which integrates artificial intelligence and communication technologies to realize efficient semantic interaction between communication objects. In general, the link structure and key technologies of the intellicise semantic communication system are proposed from three perspectives: link structure, semantic model enhancement, and model transmission, and the overall performance of the semantic communication system is enhanced. Four key technologies of intellicise semantic communications are proposed, which provide a reference for the future 6G empowerment of various vertical industries and new scene applications: a semantic knowledge graph-enhanced intellicise communication technology improves the accuracy of semantic knowledge recovery and the transmission efficiency by increasing the information dimension of semantic knowledge; the cloud-edge-device collaborative pre-cache technology of semantic knowledge graph can realize the efficient acquisition of semantic knowledge graph and assist semantic recovery; model transmission and deployment technology can achieve the effective adaptation of the models, network resources and terminal capabilities; the deployment success probability of the semantic model related to the transmission-recovery cascade process is analyzed, which provides a theoretical basis for large-scale semantic model transmission and resource deployment.

Keywords: intellicise semantic communication; model processing; model transmission; semantic knowledge base

随着传输能力的不断增强，移动通信系统以系统复杂度换取性能增益的堆叠式发展将无法满足日益增长的网络智能化和服务多样化的业务需求[1]。因此，6G亟需发展面向未来多类型应用的移动通信新范式。

语义通信是一种全新的通信范式，通过将任务和需求与信息传输有机融合，大幅度提升通信效率，改善用户的业务体验。依托智能与通信技术、网络技术等的深度融合，通过语义特征提取、语义信息传输以及语义信息恢复，语义通信可实现信息的高效准确传达和精准控制[2]。语义通信以节点智能为基础，赋予节点类人感知能力，实现达意传输，满足人-机-物之间的敏捷交互。同时，语义通信也将赋能智慧车联网、元宇宙、智能制造、多维感官互联、新兴多媒体、智慧城市等新场景、新业态[3-4]，如图1所示。

1 智简语义通信系统

1953年，W. WEAVER[5]提出了通信传输的3个层次，即语法层、语义层和语用层。其中，语法层解决的是符号级通

基金项目：国家重点研发计划（2020YFB1806900）；国家自然科学基金（62201079、61871045）

▲图1 语义通信赋能新场景、新业态

信中信息如何精准传输的问题，语义层关注的是所传递的符号信息如何准确地传达所需的含义，语用层针对的是所接收的语义信息是否能够以最佳的方式被利用。这3个层次之间相互重叠又相互关联。从1G到5G，学术界和工业界聚焦于语法信息层次，以期增强通信系统的传输性能。近年来，面向语义层的语义通信成为学术界关注的焦点，也成为6G极具竞争力的关键技术[6]。

1.1 智简语义通信系统的概念

基于上文所提的通信三大层次，智简语义通信系统着眼于语义传输，以人工智能和信息论等多学科的科学理论为指导，实现面向需求和任务的通信。智简语义通信系统深挖语法、语义与语用之间的转换关系，将需求信息进行特征提取后转化为语义信息，通过编码，再转化成语法信息。借助深度神经网络的刻画能力，智简语义通信系统对信源信息的需求及任务进行准确地识别和特征提取，并在接收端进行原始信息重建，实现精准高效的传输。同时，智简语义通信也进行了基础理论的创新：以香农信息论为基础，拓展了语义熵和系统熵[7]的概念，从理论上度量语义信息量。在网络传输层面上，“节点极智、链路极柔、网络极简”将是智简通信的主要特征，这有助于网络性能的全面提升。通过柔性简约的通信链路和按需编排的资源配置，网络中的智能节点将自主执行数据感知、处理、应用和传输等功能，实现网络由繁变简、多维资源的高效利用。智简语义通信系统将颠覆原有的通信模式，赋予信源到信宿包括感知能力、通信能力和理解能力在内的更多新能力。

1.2 智简语义通信系统的需求和挑战

智简语义通信系统旨在建立一套全新的通信体系，通过海量的语义模型和语义知识图谱对智能性进行支撑，实现人–机–物及智能体之间的智能交互。语义模型训练及语义知识图谱的构建将消耗巨大的时间与计算资源，因此，提升模型的训练效率，降低模型的训练成本，实现模型在网络中高效的传输和部署，将是智简语义通信的重要基础，也是所面临的关键挑战之一。

未来6G网络的信息多样性将得到进一步拓展，存在文本、语音、图像、触觉等多模态的信息形式，这给语义特征提取与建模带来巨大挑战。在面向需求和任务的通信中，发送端需要对接收端的需求或任务进行精准预测，以保障对重要信息的准确表达，而这进一步增加了语义信息的提取难度。多类型的信息具有不同的数据格式，如何对语义信息进行一致化建模，实现不同模态间的语义信息的高效转换，也将是一大挑战。

语义信息的传输与准确恢复是智简语义通信系统的重要过程。原始信息的准确传达，不仅需要语义信息的可靠传输，也需要语义解码模型的准确适配。在这种双重的语义信息恢复条件制约下，如何实现语义信息及语义模型的高效准确的传输也是智简语义通信系统的一大挑战。

智简语义通信是建立在节点智能性的基础上的。在当前阶段，模型的体量无法适配于节点受限的处理能力和网络资源。在保障模型性能的前提下，如何将大型模型转换为轻量级的小型模型是语义通信的迫切需求。不同节点的能力也存在不均衡性，因此如何根据不同节点的需求和能力部署相应的模型，以提升语义恢复准确度，并降低服务响应时延，也是智简语义通信的一个关键挑战。

1.3 智简语义通信系统的链路结构

利用深度神经网络技术，人们对文字、图像、视频以及多模态业务的语义通信技术进行了探索[8-11]。在此基础上，业界提出了基于人工智能模型的以网络智能决策和模型传输为特征的智简语义通信系统，如图2所示。在智简语义通信系统中，收发端分别部署语义编码模块和译码模块，模块对应的语义模型分别用于提取和恢复语义信息。边缘或云端生成语义模型，并根据收发端的请求，部署适配的语义模型。同时，收发端将语义模型存储在各自的语义模型库中。发送端将原始信息输入到语义提取与表征模块，得到语义信息，并通过联合的语义编码和信道编码将语义信息转化为比特数据，再进行传输。接收端对接收到的比特数据进行联合信道译码、语义译码，以及语义信息恢复重建，恢复出原始信息。由于上下文、通信环境等背景因素会影响语义信息的恢复，语义译码模块对背景因素带来的误差可进行补偿。

为了保障模型的适配性，需要不断对模型进行优化和更新。迁移学习是一种高效的方式，只需要对部分语义编码器和译码器的神经网络层进行重新设计和训练，重用预训练模型中某些网络层的权值，就可以减少模型再训练的时间、资

源等消耗。智简通信系统结合“残差传播”理念，对智简模型进行切分传输，即信源端根据模型切分技术将智简模型分为多个切片。这些切片在多路进行传输，并于信宿端重新组建。模型切分传输的方式一方面平衡无线通信网络负载，减少网络拥塞，提升模型并行传播速率；另一方面促进不同节点进行模型共享，利于后续模型的传输。模型压缩技术在保障信息恢复准确性的前提下，可有效地提升网络传输效率。

2 知识库增强的智简语义通信系统设计

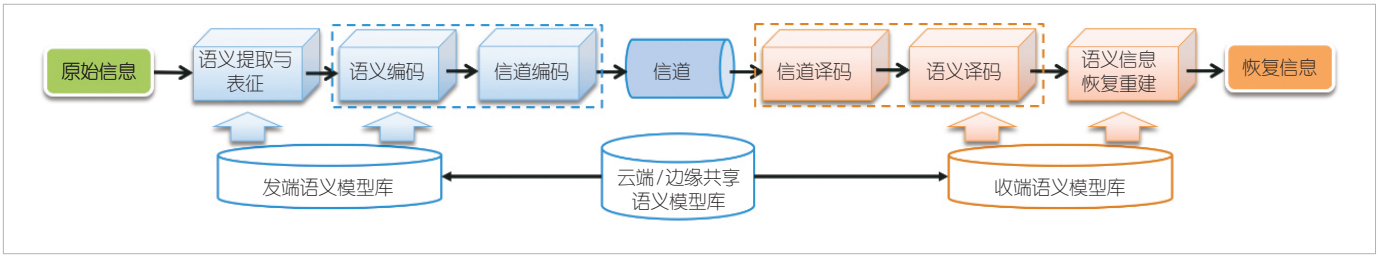
知识图谱是一种语义表示，可以对现实世界中的实体、概念、属性以及它们之间的关系进行建模，具有极强的灵活性和表达能力。语义知识库是由诸多的语义知识图谱组成的。在上一节所提的智简语义通信链路架构的基础上，利用知识图谱的灵活精简表征的功能来部署语义知识库可有效地增强语义通信性能[12]。

2.1 知识库增强的智简语义通信关键技术

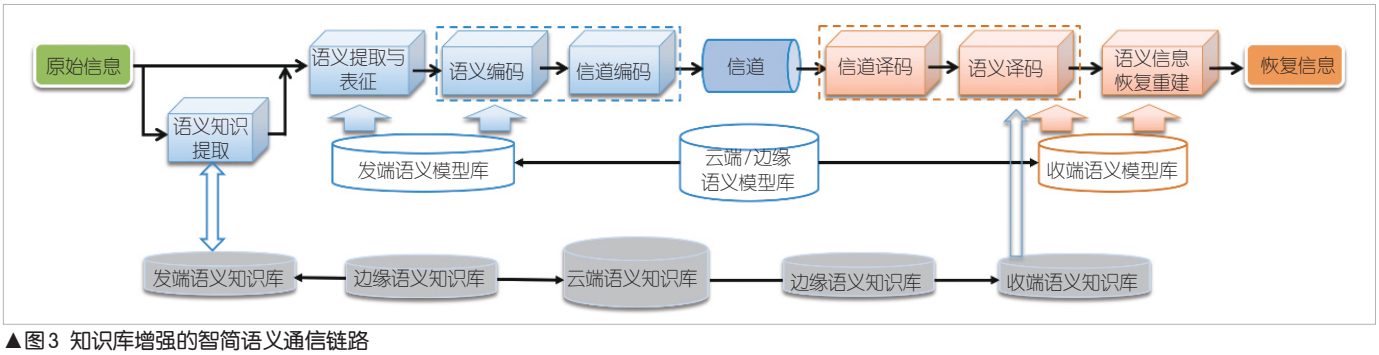
当前，针对文本、语音、图像等多种信息类型的语义通信的研究工作已取得诸多进展，如基于深度学习的语义通信（Deep-SC）[8-9]和基于深度学习的信源信道联合编码（Deep-JSCC）[10-11]等。大部分的语义编译码是基于Transformer模型来实现计的。以Transformer模型为基础，智简语义通信系统引入了语义知识图谱以增强语义通信的性能，构建出如图3所示的基于知识增强的智简语义通信系统链路结构。

语义知识是从原始信息的语义知识中提取得到的。语义提取和表征编码器将语义知识和原始信息进行联合语义提取。所提取出来的语义信息将会根据重要性进行排序，重要性程度较高的语义信息将会以更可靠的方式进行联合语义信道编码。接收端利用相匹配的语义知识图谱来增强语义译码，辅助恢复信息。语义知识图谱需进行云-边-端同步，以保障收发端能够基于相匹配的语义知识图谱来增强语义提取与恢复。为了使接收端能够从语义知识库中快速查找到相应的语义知识图谱，我们在发送端的信道导频之前增加了语义导频，以使接收端基于语义导频快速查找到相应的语义知识图谱，增强语义恢复效率和准确性。从语义提取模块输出的数据格式为浮点型，我们需对浮点型数据进行量化和数学变换以降低编码冗余。根据信道状态及语义信息重要程度，我们对提取的语义数据进行可变码长的方式进行编码，在保障语义传输准确度的前提下，提升传输效率。

性能评估也是语义通信系统设计中的重要一环。在面向文本的语义通信中，误字率（WER）[13]和双语替换评价（BLEU）[14]是两种经典的文本语义信息恢复准确度的评估指标。基于语义知识增强的智简语义通信系统已被验证能够有效地提升语义信息恢复准确性能和效率。此外，针对图像传输，峰值信噪比（PSNR）和结构相似指数（SSIM）是常用的信息相似度评估的方法[15]。针对语音传输，音频信号转为文本过程采用WER来评估，转成的文本之间采用Transformer模型的双向编码器（BERT）方法[16]来评估。如何高效地构建多种模态信源信息的知识图谱，提升语义知识图谱的更新效率，降低传输冗余，仍然是智简语义通信设计中亟待



▲图2 智简语义通信系统链路结构



▲图3 知识库增强的智简语义通信链路

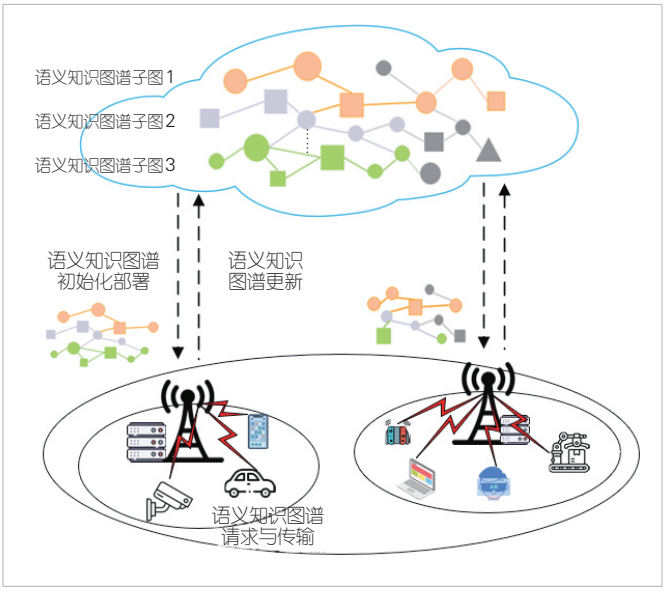
解决的问题。

2.2 智简语义通信中基于缓存增强的语义知识库构建方法

如2.1节所述，智简语义通信系统需在云–边–端协同部署语义知识库，以提升接收端的语义信息恢复准确度。语义知识图谱的生成往往需要大量的计算和存储资源，在云端生成知识图谱是一种行之有效的方式。然而，受到多维资源的制约，云与用户之间的多跳传输，会给语义知识库的传输时延、传输成功率和语义恢复准确度带来较大的影响。每个用户所需的语义知识图谱繁杂且具有时变性，因此边缘服务器或用户进行语义知识库预缓存是一种必然趋势，可有效降低语义知识库获取时延，具体的云–边–端协同语义知识图谱的预缓存机制如图4所示。基于大数据分析和人工智能技术，根据用户侧的历史请求构建知识图谱的预测模型，动态调整和定期更新语义知识库，是语义知识图谱高效谱缓存和快速获取的关键[17]。采用数学拟合的方式，得出针对特定应用场景的语义知识图谱的恢复准确度与知识图谱大小之间的数学表达，这对定量地缓存语义知识图谱来说具有重要意义。如何利用用户偏好和能力，实现语义知识图谱的分布式协同存储，也是进一步改善智简语义通信系统性能的有效方式之一。

3 智简语义通信中的模型传输关键技术

面向人–机–物智联化应用，节点极智是必然的发展趋势。节点利用其智能性进行网络感知与推理。智简语义通信需要通信对象配置相同的语义知识库和语义模型。语义模型



▲图4 云–边–端协同语义知识图谱缓存机制

的训练和传输部署也是未来网络的一项基本业务。模型的训练精度和更新频率将影响智能节点语义感知性能，轻量化且适配的语义模型部署是智简语义通信系统的基石。

3.1 语义模型处理方法

当前比较成熟的语义模型往往体量巨大，以谷歌推出的Transformer模型为例，其参数高达1.6万亿个，训练时间和成本以及模型运行所耗费的计算资源都是终端设备所无法承受的。因此，需要针对特定的应用和业务，并考虑终端设备的处理能力，对模型进行压缩、剪枝[18]、量化、蒸馏[19]、切片等处理。处理后的轻量级模型将被分发至边缘服务器及终端设备进行存储。终端设备利用部署的模型，处理感知数据，执行推理应用。

接下来，我们会对这几种模型处理技术进行简要介绍。

1）模型剪枝是通过去除两个神经元之间多余的或不重要的连接，或直接去除不需要的神经元及其相关分支，减小模型规模。

2）模型量化主要是通过减小权重参数，在服务质量（QoS）容忍范围内降低模型的精度。这种方式也是通过减小模型规模以节省缓存空间，加速模型训练，并减少传输的比特数据量。

3）模型蒸馏是基于大规模的教师模型，在满足QoS的情况下，训练一个轻量级的学生模型。该方法可以将预先训练的复杂模型的知识和泛化能力转移到一个结构更简单的网络上。具体来说，该方法利用教师模型的输出作为监督信号来训练学生模型，实现一个具有相似知识的简单网络。

4）模型选择是根据终端的QoS要求选择合适的模型，进一步优化整个系统的延迟、能效、能耗和决策精度。该方法是将相同功能的不同尺寸和准确度的模型全部存储下来，以按需分配，但难免造成缓存冗余。

5）语义切片技术是将模型进行分层化设计（包括基础模型和增强切片），基础模型以较低的通信和计算开销独立完成语义通信，也可以根据QoS要求、信道条件和传输目标选择合适的模型切片来增强基础模型的性能。

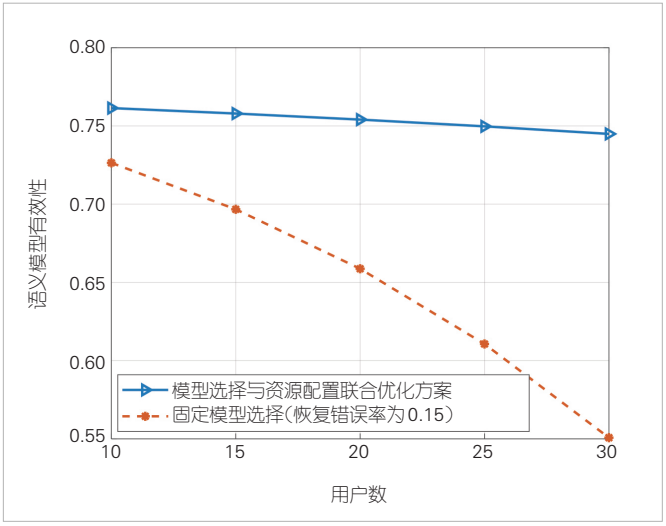
6）模型处理是智简语义通信系统实现模型有效传输及部署的关键，值得我们进行深入探索。

3.2 语义模型传输方法

面向智简语义通信的模型传输方案旨在将面向数据认知的智能能力引入数据生成节点和服务生成节点，对模型和网络资源进行端到端编排，并充分利用网络中的分布式计算资源和数据，在架构层面实现通信连接和智能能力的融合。语

义模型的训练过程是网络中的单个设备无法承担的，往往需要特定的训练中心来执行。网络针对特定的任务进行模型的定制化处理，并基于用户的请求，将语义模型传输给各个终端设备。基于云–边–端3层网络架构，语义模型可实现多级分布式的缓存和管理，边缘及终端也可基于本地信息进行模型的微调，并将微调信息反馈到网络，实现全局的更新。边缘和设备根据模型流行度及用户偏好提前缓存部分语义模型，终端设备之间也可以共享语义模型，这将实现模型及时获取和缓存效率提升。联邦学习是进行模型分布式训练及更新的有效方式之一，设备之间通过参数交互进行共同的模型优化，保障了数据的隐私和安全。

模型选择是适配网络通信能力和节点算力的重要方式，即对相同功能且不同大小和精度的模型进行选择并传输。为了验证模型选择及模型传输在语义通信中的优势，我们构建了一个模型选择与通信资源配置的联合优化方案。该方案提出了一个新的评估指标，即模型有效性，其定义为模型传输成功率与语义恢复准确率的乘积。其中，传输成功率被定义为模型在容忍时延内传输至接收端的概率。在仿真验证中，主要参数设置为：对于相同功能的模型集合，对应的恢复错误率为{0.125, 0.150, 0.175, 0.100, 0.250, 0.300}，相应的模型大小分别为{82.274 6, 46.671 9, 28.899 3, 19.079 4, 9.532 9, 5.407 8} Mbit，上述参数由针对图像传输的语义模型训练所得。如图5所示，在给定通信资源的情况下，与固定模型大小和准确度的方案对比，模型选择和通信资源联合配置可有效地提升语义模型的有效性。随着用户数增多，单用户获取的资源减少，所提方案可有效平衡模型大小与准确度。上述研究的论文将于后续发表。为了实现模型驱动的智简语义通信系统设计，一方面需要模型预处理，使其适配于终端设备；另一方面也需要针对模型传输进行模型选择和资源配置的联合优化，以增强总体网络性能。



▲图5 语义模型有效性随用户数的变化曲线

4 结束语

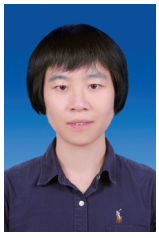
随着元宇宙、智慧车联网、新媒体等新场景和新业务的出现，智简语义通信对6G关键服务起到重要的支撑作用。业界也正从不同的角度探索语义通信基础理论及关键技术，并设计原型系统，以实现多类型信息的高效语义传输。依托人工智能与模型传输技术，智简语义通信将实现网元之间高效敏捷交互，有望成为6G的颠覆性技术之一。

参考文献

[1] 张平, 陶运铮, 张治. 5G若干关键技术评述 [J]. 通信学报, 2016, 37(7): 15–29. DOI: 10.11959/j.issn.1000–436x.2016130
[2] 石光明, 肖泳, 李莹玉, 等. 面向万物智联的语义通信网络 [J]. 物联网学报, 2021, 5(2): 26–36. DOI: 10.11959/j.issn.2096–3750.2021.00209
[3] 牛凯, 戴金晟, 张平, 等. 面向6G的语义通信 [J]. 移动通信, 2021, 45(4): 85–90. DOI: 10.3969/j.issn.1006–1010.2021.04.014
[4] 徐英姿, 刘原, 时梦然, 等. 语义在通信中的应用综述 [J]. 电信科学, 2022, 38(S1): 43–59. DOI: 10.11959/j.issn.1000–0801.2022107
[5] WEAVER W. Recent contributions to the mathematical theory of communication [J]. ETC: a review of general semantics, 1953, 10(4): 261–281
[6] ZHANG P, XU W J, GAO H, et al. Toward wisdom–evolutionary and primitive–concise 6G: a new paradigm of semantic communication networks [J]. Engineering, 2022, 8: 60–73. DOI: 10.1016/j.eng.2021.11.003
[7] 张平, 许文俊, 王凤玉, 等. 智简空天地一体化网络 [J]. 无线电通信技术, 2022, 48(3): 381–384. DOI: 10.3969/j.issn.1003–3114.2022.03.001
[8] 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
[9] WENG Z Z, QIN Z J. Semantic communication systems for speech transmission [J]. IEEE journal on selected areas in communications, 2021, 39(8): 2434–2444. DOI: 10.1109/JSAC.2021.3087240
[10] 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
[11] WANG S, DAI J, LIANG Z, et al. Wireless deep video semantic transmission [EB/OL]. [2023–01–22]. https://arxiv.org/abs/2205.13129
[12] XU X, XIONG H, WANG Y, et al. Knowledge–enhanced semantic communication system with OFDM transmissions [EB/OL]. (2022–07–01) [2023–01–23]. https://www. sciengine. com/SCIS/doi/10.1007/s11432–022–3624–4
[13] FARSAD N, RAO M, GOLDSMITH A. Deep learning for joint source–channel coding of text [C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 2326–2330. DOI: 10.1109/ICASSP.2018.8461983
[14] PAPINENI K, ROUKOS S, WARD T, et al. BLEU: a method for automatic evaluation of machine translation [C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. ACM, 2002: 311–318. DOI: 10.3115/1073083.1073135
[15] SARA U, AKTER M, UDDIN M S. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study [J]. Journal of computer and communications, 2019, 7(3): 8–18. DOI: 10.4236/jcc.2019.73002
[16] DEVLIN J, CHANG M, LEE K, et al. BERT: pre–training of deep bidirectional transformers for language understanding [EB/OL]. (2018–10–11)[2023–01–22]. https://arxiv.org/abs/1810.04805
[17] CHE Y, XIONG H C, HAN S J, et al. Cache–enabled knowledge base construction strategy in semantic communications [C]//Proceedings of 2022 IEEE Globecom Workshops (GC Wkshps). IEEE, 2023: 1507–1512. DOI: 10.1109/GCWkshps56602.2022.10008494

[18] HAN S, POOL J, TRAN J, et al. Learning both weights and connections for efficient neural networks [C]//Proceedings of the 28th International Conference on Neural Information Processing Systems – Volume 1. ACM, 2015: 1135–1143. DOI: 10.5555/2969239.2969366
[19] WANG C, CHEN D F, MEI J P, et al. SemCKD: semantic calibration for cross-layer knowledge distillation [J]. IEEE transactions on knowledge and data engineering, 2022, 99: 1. DOI: 10.1109/TKDE.2022.3171571

作　者　简　介

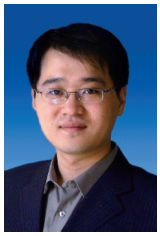


孙梦颖，北京邮电大学博士后；主要研究方向为面向6G的语义通信与边缘智能技术；发表论文10余篇，申请专利10余项。

熊华超，北京邮电大学在读硕士研究生；主要研究方向为智简语义通信及组网技术、6G原型系统搭建；发表论文3篇，申请专利2项。

王怡宁，北京邮电大学在读博士研究生；主要研究方向为智简语义通信及组网技术；发表论文1篇，申请专利3项。

韩书君，北京邮电大学博士后；主要研究方向为面向B5G/6G的新一代移动通信理论与技术；主持和参加项目10余项，获得1项科研成果奖；发表论文20余篇，申请专利60余项。



许晓东，北京邮电大学教授；主要研究方向为无线组网及覆盖扩展技术；作为负责人完成多项国家级项目，获北京市科学技术奖二等奖、中国通信学会科学技术奖一等奖，出版的专著获得“中国出版政府奖”；近5年发表重要期刊论文50余篇，申请专利80余项。

语义通信性能评估体系及指标

Performance Evaluation Systems and Metrics for Semantic Communications

郑远/ZHENG Yuan，王凤玉/WANG Fengyu，
许文俊/XU Wenjun

（北京邮电大学，中国 北京100876）
(Beijing University of Posts and Telecommunications, Beijing 100876, China)

DOI：10.12142/ZTETJ.202302010
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1340.008.html
网络出版日期：2023-04-11
收稿日期：2023-02-20

摘要：鉴于语义通信系统缺乏统一的、具泛化价值的性能评估体系，对现有研究中评估方法的使用情况与不足进行分析，并提出语义通信效率指标E_{sc}和语义通信效用指标U_{sc}。其中，指标E_{sc}用于衡量通信系统在给定通信资源下单位时间开销内的任务完成准确性，指标U_{sc}用于衡量通信系统在给定通信资源下与任务性能上限的接近程度。与现有评估指标相比，这两种指标具有更好的泛用性，可为不同任务场景、模态信息下的语义通信模型横向性能对比提供指导。此外，以语义图像重建与语音重建任务为例，分别搭建端到端语义通信仿真模型，并基于所提E_{sc}与U_{sc}指标对仿真模型性能进行评估。

关键词：语义通信；人工智能；性能评估指标

Abstract: The absence of a unified and generalized performance evaluation system brings challenges to the realization of semantic communication systems. The usages and shortcomings of evaluation methods in existing research are analyzed. Two evaluation metrics, namely the semantic communication efficiency metric E_{sc} and the semantic communication utility metric U_{sc}, are proposed. E_{sc} measures the accuracy of task completion within unit time cost of a communication system under certain resources. U_{sc} measures how close a communication system is to the upper limit of task performance under certain resources. Compared with the existing evaluation methods, the two proposed metrics are of better generality, which can provide guidance for the performance comparison of semantic communication models under different scenarios and information modalities. In addition, the semantic image reconstruction and speech reconstruction are taken as examples to build end-to-end semantic communication simulation models respectively, and the performance of models based on the proposed E_{sc} and U_{sc} metrics is evaluated.

Keywords: semantic communication; artificial intelligence; performance evaluation metric

从20世纪80年代后期1G的诞生到如今5G商业化进程的稳步展开，移动通信经历了近半个世纪的高速发展。得益于高峰值速率、低传输时延与高连接密度，5G系统支持包含虚拟/增强现实、智能驾驶、智慧城市等在内的多样化垂直应用[1]。随着全球新一轮科技与产业革命的加速发展，信息通信技术将进一步重构人与人、人与物、物与物之间的复杂联系，为工业、医疗、交通、教育、金融、娱乐、零售等千行百业带来革新。当前通信系统主要基于香农信息论研究设计。从1G到5G，通信系统虽然在业务形式、服务对象、网络架构和承载资源等方面进行了技术变革，但都依赖于增加信息传输的物理维度、通过技术堆叠处理来逼近香农信息论极限。

随着通信系统与各垂直行业领域的紧密结合，面向个人、行业等用户的智能需求将被进一步挖掘，这对通信系统准确传递信息含义提出更高要求。而以香农信息论为基础的传统通信系统并不关注信息中承载的含义，只关注每个传输比特的正确接收。这导致不必要的通信资源耗费，难以满足未来通信持续发展的需求。语义通信通过交叉融合人工智能与通信技术，深度挖掘信息本身的语义维度，引入语义层次的信息，关注传输信息内容而非编码符号，更加满足未来通信需求。通过充分利用语义层面信息的高度抽象、智能简约等特性，语义通信将变革经典香农信息论框架，有望突破经典通信系统的传输瓶颈，形成智能化的新型通信体系[2]。

当前，语义通信尚处于研究初期，构建统一的、具有泛化价值的性能评估体系对语义通信的发展至关重要。本文围绕评估体系在语义通信系统设计实现中的作用展开探讨，并在分析现有通信系统评估指标的基础上提出具有泛化性的语

基金项目：国家自然科学基金（62293485）；中央高校基本科研业务费专项资金（2022RC18）

义通信评估体系，为语义通信的发展奠定基础。

1 语义通信性能评估体系需求

当前，语义通信尚处于研究初期，语义通信理论在代表性的人–机、机–机等智能交互场景中快速发展，关键性成果不断涌现，系统性能增益更加显著。然而，由于缺乏统一且具有泛化价值的性能评估体系，语义通信研究成果的可理解性有所欠缺，成果间的横向对比难以实现。

一般而言，准确性与时效性是评价通信系统性能的两个重要指标。现有语义通信相关研究所采用的评估体系主要由传统通信系统评估体系演化而来，或由下游语义任务评估体系迁移而来。对于由传统通信系统评估体系演化而来的情况，相应的准确性评估指标常采用误比特率，时效性评估指标常采用比特传输速率。这类评估指标的问题在于无法有效反映通信收发端语义信息传递的能力。对于由下游语义任务评估体系迁移而来的情况，由于系统模型多采用通信模块与下游语义任务模块组合确定结构、端到端训练固定参数的方式，这类评估指标具有与通信任务场景、信息模态高度关联的特征。如面向语义重建的语义通信系统普遍采用原始信息与重建信息的误差函数作为准确性度量，采用压缩比的函数作为时效性度量；而面向文本翻译、图像分割等语义任务的语义通信系统则普遍将模型输出与监督标签的差异作为准确性度量。

上述各评估指标的定义不同，取值范围存在显著差异，导致诸多研究的系统模型之间无法横向贯通。即使针对相同评估指标，由于信道、干扰、信源语义特征等差异性存在，不同场景下的语义传输方案也无法进行有效对比。统一且具泛化价值的语义通信评估体系，从整体视角对通信系统性能进行直观的评估，不仅能验证系统的合理性，还可以为系统的优化改进提供参考，进而为系统演进提供方向。语义通信评估体系相关研究亟待开展，以便为语义通信的稳步发展夯实基础。

2 现有语义通信性能评估指标

语义通信性能评估指标与具体通信任务场景、信息模态高度关联。其中，信息重建任务广泛出现在以人类作为最终接收用户的应用场景中，如音视频通话、多媒体文件云备份等，该类任务要求在收端将信息恢复至发端信息模态，并尽量保证信息的准确，性能评估指标基于发送信息与重建信息的差异进行定义；非信息重建任务多出现于人–机、机–机交互场景中，如智能驾驶、工业物联网等，该类任务要求所发信息的关键内容能够被理解和应用，性能评估指标基于所输出信息与相应监督标签之间的差异定义。对于信息模态而言，人类针对文本、图像/视频、语音等模态信息采用不同的重点捕捉方式，各模态信息的失真无法简单通过欧氏距离等方式统一度量。各模态信息所衍生的后续任务存在很大差异，文本情感分析、智能问答、图像分割、语音识别等任务均需要定义各自的性能评估指标。

基于重建任务与其他代表性人工智能任务（例如文本分类、图像目标检测、语音识别）的划分，本节对文本、图像/视频、语音3种常见模态信息传输模型的性能评估指标进行分析总结，为后续构建统一、具有泛化性的语义通信性能评估体系奠定基础。

2.1 面向重建任务的评估指标

2.1.1 文本信息重建

文本信息重建的现有常用衡量指标主要为双语替换评测分数[3]（BLEU），该指标最初被用于文本翻译的质量评估。BLEU基于加权n元模型（n – gram）精确度进行定义，其具体形式为：

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

其中，$\mathrm{BP}=\begin{cases}1 & if\ l_{\hat{X}}>l_X\\ e^{1-\frac{l_X}{l_{\hat{X}}}} & if\ l_{\hat{X}}\leqslant l_X\end{cases}$为针对重建文本语句长度过短的惩罚项，$l_{\hat{X}}$、$l_X$分别为重建文本与原始文本语句的长度，$w_n$为归一化权重，$N=4$为经验取值。

$$p_n=\frac{\sum_{\mathrm{n-gram}\in\hat{X}}\mathrm{Count}_{\mathrm{clip}}(n-\mathrm{gram})}{\sum_{\mathrm{n-gram}'\in\hat{X}'}\mathrm{Count}(n-\mathrm{gram}')}, \tag{2}$$

其中，p_n表示n–gram精确度，$\mathrm{Count}_{\mathrm{clip}}(n-\mathrm{gram})$、$\mathrm{Count}(n-\mathrm{gram}')$分别表示$n$长词组在原始文本$X$与重建文本$\hat{X}$中的出现次数。BLEU的物理意义是衡量重建文本的所有n长词组中同时出现在原始文本中的数量占比。尽管作为最为常用的文本信息重建衡量指标，BLEU仍无法实现对于词汇语义的理解[4]，存在未考虑到词汇语料的召回率[5]等不足。

2.1.2 图像/视频信息重建

现有图像/视频信息重建的常用衡量指标包括峰值信噪比（PSNR）与多尺度结构相似性[6]（MS – SSIM）。PSNR定义为均方误差（MSE）的对数函数。对于动态范围N的图像/视频帧，PSNR计算如下：

$$\mathrm{PSNR} = 10\lg\frac{N^2}{\mathrm{MSE}(X,\hat{X})}, \tag{3}$$

其中，$\mathrm{MSE}\left(X,\hat{X}\right)=\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X_{ij}-\hat{X}_{ij}\right)^2/(HW)$ 表示尺寸为 $H\times W$ 的原始与重建图像/视频帧的均方误差。PSNR 是对图像/视频帧像素级重建误差的精确反映，因定义简单且易于计算，在早期工作[7]中被广泛应用。但由于未考虑人眼生理特性等因素，PSNR 的评估准确性不如 MS－SSIM。

MS－SSIM 将人眼生理特性纳入考虑：人眼在衡量两幅图的相似性时，对于两幅图的局部结构差异更为敏感。故一方面，MS－SSIM 的运算基于一定尺寸的图像块；另一方面，MS－SSIM 通过分别定义亮度、对比度、结构对比函数，将 3 类信息拆分，赋予它们不同的权重并独立进行对比。相较于 PSNR，MS－SSIM 的评估结果更符合人眼感知。MS－SSIM 的计算方式为：

$$\mathrm{MS-SSIM}\left(X,\hat{X}\right)=\left[L_M\left(X,\hat{X}\right)\right]^{\alpha M}\prod_1^M\left[C_J\left(X_M,\hat{X}_M\right)\right]^{\beta_J}\left[S_J\left(X,\hat{X}_M\right)\right]^{\gamma_J}, \tag{4}$$

其中，降采样次数 M、亮度权重 αM、对比度权重 β_J、结构对比权重 γ_J 为超参数，其常用取值为：$M=5,\beta_1=\gamma_1=0.044\,8,\beta_2=\gamma_2=0.285\,6,\beta_3=\gamma_3=0.300\,1,\beta_4=\gamma_4=0.236\,3,\alpha_5=\beta_5=\gamma_5=0.133\,3$。$X_M$、$\hat{X}_M$ 分别为原图及重构图像/视频帧的 M 级降采样结果。

$$L_M\left(X_M,\hat{X}_M\right)=\frac{2\mu_{X_M}\mu_{\hat{X}_M}+C_1}{\mu_{X_M}^2+\mu_{\hat{X}_M}^2+C_1}, \tag{5}$$

$$C_J\left(X_M,\hat{X}_M\right)=\frac{2\sigma_{X_M}\sigma_{\hat{X}_M}+C_2}{\sigma_{X_M}^2+\sigma_{\hat{X}_M}^2+C_2}, \tag{6}$$

$$S_J\left(X_M,\hat{X}_M\right)=\frac{\sigma_{X_M\hat{X}_M}+C_3}{\sigma_{X_M}\sigma_{\hat{X}_M}+C_3}, \tag{7}$$

其中，$L_M\left(X_M,\hat{X}_M\right)$、$C_J\left(X_M,\hat{X}_M\right)$、$S_J\left(X_M,\hat{X}_M\right)$ 分别为指标的亮度对比因子、对比度因子、结构对比因子，μ_{X_M}、$\mu_{\hat{X}_M}$、σ_{X_M}、$\sigma_{\hat{X}_M}$ 分别为 X_M、$\hat{X}_M$ 的均值与标准差，$\sigma_{X_M\hat{X}_M}$ 为两者的协方差。

$C_1=(k_1N)^2$、$C_2=(k_2N)^2$、$C_3=C_2/2$ 为固定参数，用于防止除法运算问题的产生。N 为图像/视频帧的动态范围，$k_1=0.01$、$k_2=0.03$ 为经验取值。

近年的相关工作[8-9]多采用 PSNR 与 MS－SSIM 指标并行的方式。此外，文献[8]还采用了对数形式的 MS－SSIM 指标，以应对 MS－SSIM 取值范围过小的问题。

2.1.3 语音信息重建

语音信息重建的现有衡量指标包括语音质量感知评估[10]（PESQ）与短时客观可懂度[11]（STOI）等。早期语音质量衡量主要通过主观打分方式获得平均意见值（MOS）。PESQ 是用于模拟 MOS 打分的音频质量评价算法，其定义为：

$$\mathrm{PESQ}=4.5-0.1d_{\mathrm{SYM}}-0.0309d_{\mathrm{ASYM}}, \tag{8}$$

其中，d_{SYM} 与 d_{ASYM} 分别为算法求得的语音信息对称与非对称干扰，反映重建模型的预测精度和概括能力。PESQ 是对重建语音信号可理解程度的衡量，侧重人的主观感受，在早期的无线通信系统中广泛使用。

同样被广泛使用的指标还包括 STOI。STOI 是对重建语音信息失真程度的客观度量，其取值与基于人工智能的语音识别等下游任务性能相关。具体而言，首先定义 $\hat{x}(k,m)$ 为第 m 帧语音信号的第 k 个离散傅里叶变换（DFT）波瓣，计算第 m 帧语音信号的第 j 个 1/3 倍频程范数 $X_j(m)$。

$$X_j(m)=\sqrt{\sum_{k=k_1(j)}^{k_2(j)-1}\left|\hat{x}(k,m)\right|^2}, \tag{9}$$

其中，$k_1(j)$、$k_2(j)$ 为定义的取值边界。同理，可定义重建语音信号的相应范数 $Y_j(m)$，计算 $Y_j(m)$ 的归一化形式 $Y_j(m)'$：

$$Y_j(m)'=\max\left(\min\left(\alpha Y_j(m),\ X_j(m)+10^{-\frac{\beta}{20}}X_j(m)\right),\ X_j(m)-10^{-\frac{\beta}{20}}X_j(m)\right), \tag{10}$$

其中，$\beta\in\{-\infty,-30,-20,-15,-10\}$ 与 $N\in\{20,30,40,50,60\}$ 为可选参数，$\alpha=\left(\frac{\sum_n X_j(n)^2}{\sum_n Y_j(n)^2}\right)^{1/2}$ 为功率归一化参数，$n\in\mathcal{M}=\{(m-N+1),\cdots,m\}$。

之后，定义第 m 帧语音信号的第 j 个 1/3 倍频程损失 $d_j(m)$：

$$d_j(m)=\frac{\sum_n\left(X_j(n)-\frac{1}{N}\sum_l X_j(l)\right)\left(\left(Y_j'(n)-\frac{1}{N}\sum_l Y_j'(l)\right)\right)}{\sqrt{\sum_n\left(X_j(n)-\frac{1}{N}\sum_l X_j(l)\right)^2\sum_n\left(\left(Y_j'(n)-\frac{1}{N}\sum_l Y_j'(l)\right)^2\right)}},\ l\in\mathcal{M}。 \tag{11}$$

STOI = $\frac{1}{JM}\sum_{j,m} d_j(m)$ 为 $d_j(m)$ 的均值，其中 M、J 分别表示语音信号的总帧数与1/3倍频程数。现有工作[12]多同时采用PESQ与STOI，以获得对于所提模型更为全面的性能认知。

2.2 面向其他人工智能任务的评估指标

2.2.1 文本分类

文本分类是自然语言处理领域的典型任务，其目的在于通过人工智能为文本信息按照一定的标准进行标记，可细分为情感分析、主题分类、问答任务、意图识别等多种应用场景。以单标签二分类的情感分析任务为例，现有评价指标主要为分类准确率与F分数[13]（F－score），具体计算逻辑如下：

按照各样本的分类结果与真实标签，可将各样本的分类结果分为真阳性（TP）、真阴性（TN）、假阳性（FP）、假阴性（FN），如表1混淆矩阵所示。在此基础上，准确率 $ACCURACY = \frac{TP + TN}{TP + FP + TN + FN}$ 定义为经模型正确分类的样本占总样本的比例，是对任务性能最简单直观的度量。在实际应用场景中，FN与FP的代价可能是不同的，准确率指标无法对其进行细化评估，而F分数弥补了这方面的不足。

$F - score = \frac{(1+\beta^2)\cdot precision \cdot recall}{\beta^2 \cdot precision + recall}$ 是基于所有样本的分类精确率 $precision = \frac{TP}{TP + FP}$ 与召回率 $recall = \frac{TP}{TP + FN}$ 的调和平均数定义的度量，超参数β赋予了指标为精确率与召回率针对性加权的能力。在现今的相关研究中，$F1 - score = \frac{2 \cdot precision \cdot recall}{precision + recall}, (\beta = 1)$ 是常用的一种形式。对于多分类任务，marco－F1、weighted－F1等变体指标被广泛应用[14]。当数据集的分布不均衡时，F分数仍存在主要类别的评估值偏高[15]等弊端。相应的改进方法有基于接收者操作特征曲线（ROC）的度量等。

2.2.2 图像目标检测

图像目标检测的目的在于通过人工智能判断出图像中物体的类别与相应位置。图像目标检测任务的现有评价指标主要为平均精确度均值（mAP），其定义为各类目标查准率（P）–查全率（R）曲线下方面积的均值。

具体而言，首先，根据目标检测结果 B_p 的置信度，以及 B_p 与真实结果 B_{gt} 的重合程度（可用两者的交并比IoU度量，$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}$），可将检测结果分为如表2混淆矩阵所示的TP、TN、FP、FN；其次，由检测顺序中TP、FP数量的累加情况计算查准率与查全率的累加值：

$$P = \frac{TP}{TP + FP}, \tag{12}$$

$$R = \frac{TP}{TP + FN}; \tag{13}$$

然后，对于各类目标样本，以查全率为横坐标，以查准率为纵坐标，绘制 P–R 曲线并计算精确度均值（AP）；最后，由AP加权平均可得mAP，即 $mAP = \frac{\sum_1^M N_M AP_M}{\sum_1^M N_M}$，$M$、$N_M$ 分别为目标类别数量与该类别样本数量。

目前，基于上述定义的mAP指标在现有工作[16]中已被广泛使用。值得注意的是，在如图像检索等实例级图像任务中，存在另一种基于整图匹配结果定义的mAP指标（已应用在文献[17]的端到端车辆重识别系统中）。两者的运算逻辑一致，仅在混淆矩阵定义上存在差异。

2.2.3 语音识别

语音识别的目的在于通过人工智能实现语音信息到文本信息的模态转换过程。语音识别任务的现有评价指标主要为准确率 $ACCURACY$。准确率可基于字错误率（WER）定义，即：

$$ACCURACY = 1 - WER, \tag{14}$$

其中，$WER = \frac{N_{sub} + N_{ins} + N_{del}}{N}$，$N_{sub}$、$N_{ins}$、$N_{del}$ 分别为识别文本到达参考文本所需的单词替换、插入与删除操作的次数，N 为参考文本的总单词数。ACCURACY已在文献[18]等工作中获得应用。

▼表1 二分类任务混淆矩阵

	分类结果为真	分类结果为假
真实标签为真	TP	FN
真实标签为假	FP	TN

FN：假阴性　FP：假阳性　TN：真阴性　TP：真阳性

▼表2 目标检测任务混淆矩阵

	B_p 置信度大于阈值	B_p 置信度小于阈值
B_{gt} 与 B_p 的IoU大于阈值	TP	FN
任意 B_{gt} 与 B_p 的IoU均小于阈值	FP	TN

FN：假阴性　IoU：交并比　TP：真阳性
FP：假阳性　TN：真阴性

整体而言，语义通信评价指标存在以下问题：

1）现有评价指标普遍从具体的重建任务/人工智能任务直接迁移而来，侧重于反映任务完成的准确性维度，忽视了时效性维度。语义通信系统追求的是在有限带宽资源消耗与时延下的任务完成准确性，而非不计成本地逼近准确性上限。

2）不同任务场景/模态信息下的评估指标不互通，性能评估结果不易理解。在语义通信过程中，受通信意图变化的影响，系统的评估指标会不断发生变化，造成通信质量评估混乱。

3）各评价指标的取值范围、量纲不一致，给性能的横向对比带来困难。如上文中所提到的PESQ、PSNR指标与文本分类和语音识别准确率指标存在明显差异。这种差异主要表现在取值范围与物理意义方面。

综上所述，对于语义通信系统的评价指标，一方面，应在选取应用任务导向的准确性指标的同时，引入时效性指标以进行制约；另一方面，应筛选物理意义相似的指标以组成指标集合，并设计指标的标准化逻辑统一取值范围，使在不同任务场景下构建的语义通信系统之间能够进行横向性能对比。对于包含不同通信意图的语义通信系统，它们的具体实现需要与场景、应用紧密结合。从整体视角进行直接的性能评估，有助于系统之间横向贯通，促进系统的迭代演进。

3 语义通信性能评估新角度

为解决现有语义通信性能评估指标存在的问题，我们提出语义通信效率指标E_{sc}与语义通信效用指标U_{sc}，并以语义图像重建与语音重建任务为例，搭建端到端语义通信仿真系统，基于所提指标对仿真系统性能进行评估。

3.1 新评估指标定义

首先，我们定义标准化准确性指标$\overline{ACC}\in[0,1]$与标准化时效性指标$\overline{TIM}\in[1,+\infty)$：

$$\overline{ACC(\gamma)}=\min\left\{\frac{ACC(\gamma)-ACC_{\min}(\gamma)}{ACC_{\text{th}}(\gamma)},1\right\},\tag{15}$$

$$\overline{TIM(\gamma)}=\max\left\{\frac{TIM(\gamma)}{TIM_{\text{th}}(\gamma)},1\right\},\tag{16}$$

其中，γ表示给定的通信资源条件，包含信噪比（SNR）与带宽等参数；$ACC\in\{g(\text{BLEU,MS}-\text{SSIM,TOP}-n\ \text{ACCURACY,mAP},\cdots)\}$、$TIM\in\{f(\text{计算时延,传输时延,端到端时延},\cdots)\}$分别为语义任务相关的准确性与时效性评价指标，$g(\cdot)$与$f(\cdot)$分别为相应的变换函数；$ACC_{\min}$与$ACC_{\text{th}}$分别为特定语义任务的准确性下界与优化边界，$TIM_{\text{th}}$为特定语义任务时效性优化边界。这里，超越边界的准确性或时效性指标的取值不存在或不具备进一步优化的需求。

标准化操作旨在规范不同评估指标的取值尺度，剔除语义任务难易度差异对语义通信系统性能评估产生的影响。进一步地，基于上述标准化指标，我们定义语义通信效率指标E_{sc}：

$$E_{sc}(\gamma)=\frac{\overline{ACC(\gamma)}}{\overline{TIM(\gamma)}}。\tag{17}$$

语义通信效率指标$E_{sc}\in[0,1]$旨在衡量通信系统在给定通信资源下、在单位时间开销内的任务完成准确性。当通信系统的准确性指标增大或时效性指标减小时，E_{sc}会增大，即E_{sc}越接近1，系统的语义通信效率越高。

语义通信效用指标U_{sc}为：

$$U_{sc}(\gamma)=\frac{\lambda}{1+\lambda}\cdot\overline{ACC(\gamma)}+\frac{1}{1+\lambda}\cdot\frac{1}{\overline{TIM(\gamma)}},\tag{18}$$

其中，λ表示加权因子，用于权衡准确性需求和时效性需求的比重。语义通信效用指标$U_{sc}\in[0,1]$旨在衡量通信系统在给定通信资源下，对于任务性能上限的接近程度。当系统侧重于任务完成准确性或时效性时，U_{sc}的前项或后项会增大，但另一项会相应减小。U_{sc}越接近1，系统在准确性与时效性两方面的综合效用表现越高。

3.2 新性能指标分析与评估

本文首先以图像重建任务为例，选取车联网领域的传感器扩展应用作为具体验证场景，基于上述评估指标进行仿真分析。传感器扩展应用要求车辆、行人、交通设施等通信节点之间能够实现传感器所采集的图像/视频信息的实时交互，对数据传输的准确性、时效性均有较高要求。

针对上述应用场景，准确性指标ACC转换为对数形式MS－SSIM，时效性指标TIM为传输时延，变换函数$g(\cdot)$与$f(\cdot)$均选取恒等函数，即：

$$ACC(\gamma)=MS-SSIM(\gamma),\tag{19}$$

$$TIM(\gamma)=\text{传输时延}(\gamma)。\tag{20}$$

根据第3代合作伙伴计划（3GPP）制定的《5G NR Rel-16 V2X车联网标准》，准确性边界、时效性边界与通信参数

如表3所示。

进一步地，我们搭建包含图像压缩重建、信道编码与调制模块在内的端到端通信仿真模型，基于选定指标进行验证。图像压缩重建模块基于文献[19]的网络结构实现，并分别采用ImageNet数据集、Cityscapes数据集进行训练和验证。图像压缩重建模块在训练过程中采用了4组不同的率失真参数 $\alpha = \{256, 512, 1\,024, 2\,048\}$。采用更高的 α 参数训练的模型会更倾向于降低图像压缩比以换取更高的重建质量。模型采用低密度奇偶校验码（LDPC）进行信道编码，并根据信噪比条件将LDPC码率控制在1/5～8/9之间。调制方式采用二进制相移键控（BPSK）。

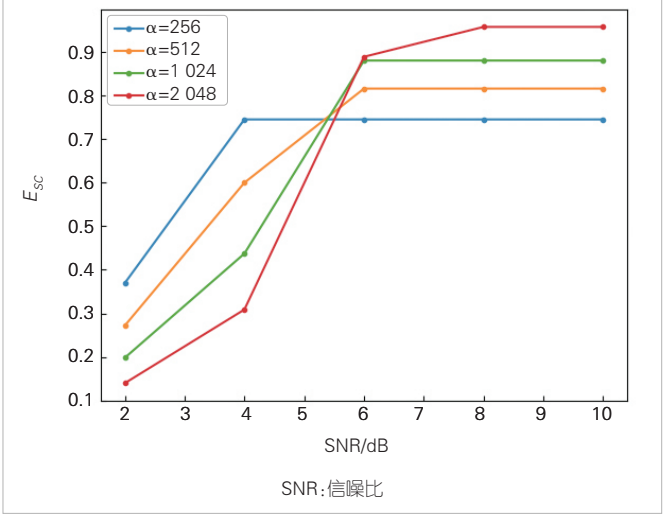
图1给出了语义通信效率 E_{sc} 与通信系统信噪比的关系曲线。随着信噪比的上升，包含不同参数的语义任务模型均表现出更高的语义通信效率。此外，采用高 α 参数训练的模型对信噪比的变化更为敏感，在低信噪比条件下表现出明显的性能劣化，同时在高信噪比条件下具有更高的性能上限。这是由于此类模型在低信噪比条件下产生了过多的编码开销以保护信息不被噪声破坏，从而无法满足任务的时效性需求；在高信噪比条件下，通信所需的编码冗余显著减少。此类模型能够在时效性不超出阈值的前提下，实现更高的任务准确性。

图2给出了 $\lambda = 5$ 条件下语义通信效用 U_{sc} 与通信系统信噪比的关系曲线。考虑到实际语义任务场景中存在对准确性或时效性更为侧重的情况，参数 λ 使效用指标更为真实地反映语义任务需求。本部分以准确性需求导向的任务为例。随着信噪比的上升，4组模型的语义通信效用均表现出与直觉吻合的上升趋势。即使在低信噪比条件下，高 α 参数模型表现出较高的准确性与较低的时效性，其语义通信效率劣于低 α 参数模型。但由于任务对于准确性需求的侧重，高 α 参数模型具有更高的语义通信效用。图3与表4分别展示了典型图像经语义图像重构模型仿真传输后的可视化结果与具体性能参数。高 α 参数模型表现为更出色的图像重建质量（红绿灯的色彩更准确，建筑物、地砖的细节更清晰）与更高的图像传输时延开销，在较好的信道条件下，其传输时延未超出阈值，故高 α 参数模型具有更高的语义通信效率与语义通信效用。

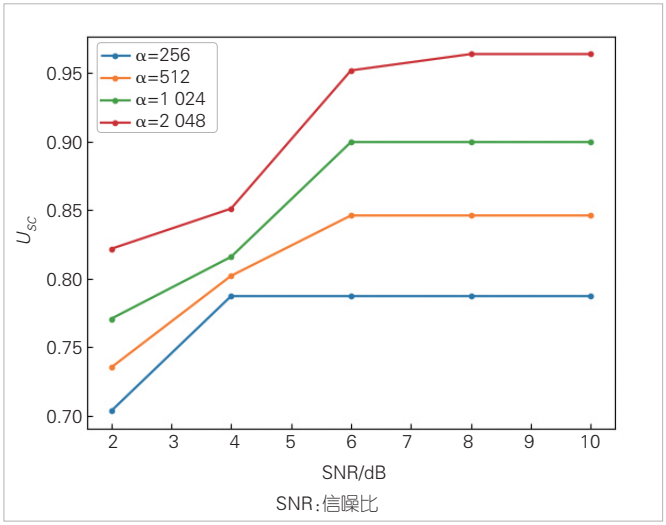
▼表3 通信模型仿真参数

	图像重建任务	语音重建任务
准确性下界 $ACC_{\min}$	0	1
准确性优化边界 ACC_{th}	0.99	4.5
时效性优化边界 TIM_{th}	100 ms	与所传输语音信息时长相等
通信参数 γ	信噪比2～10 dB，带宽3 MHz	信噪比2～10 dB，带宽10 kHz
信道编码与调制方式	LDPC+BPSK	LDPC+BPSK

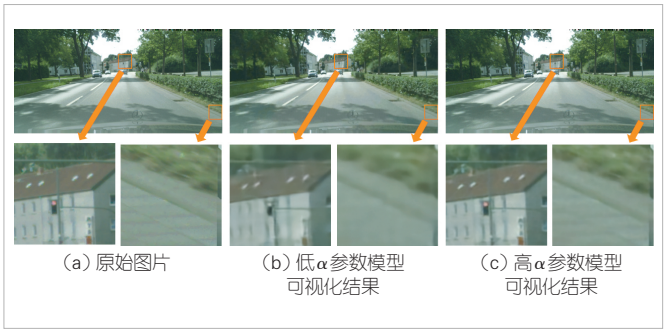
LDPC：低密度奇偶校验码　BPSK：二进制相移键控



▲图1 图像重建任务语义通信效率 E_{sc} 与信噪比的关系曲线



▲图2 图像重建任务语义通信效用 U_{sc} 与信噪比的关系曲线（$\lambda = 5$）



▲图3 通信模型可视化结果

▼表4 通信模型性能参数

性能参数	低α参数模型	高α参数模型
信噪比SNR	10 dB	10 dB
率失真参数α	256	2 048
MS－SSIM	0.961 2	0.986 3
传输时延	20.77 ms	73.20 ms
语义通信效率E_{sc}	0.705 6	0.931 6
语义通信效用U_{sc}	0.754 7	0.943 0

MS-SSIM：多尺度结构相似性　SNR：信噪比

接着我们以语音信息重建任务为例，构建应用自适应多速率宽带（AMR－WB）语音编码的高清语音通信模型，并基于上述评估指标进行仿真分析。准确性指标ACC转换为宽带语音质量感知评估（WB－PESQ），时效性指标TIM为传输时延，变换函数$g(\cdot)$与$f(\cdot)$均选取恒等函数，即：

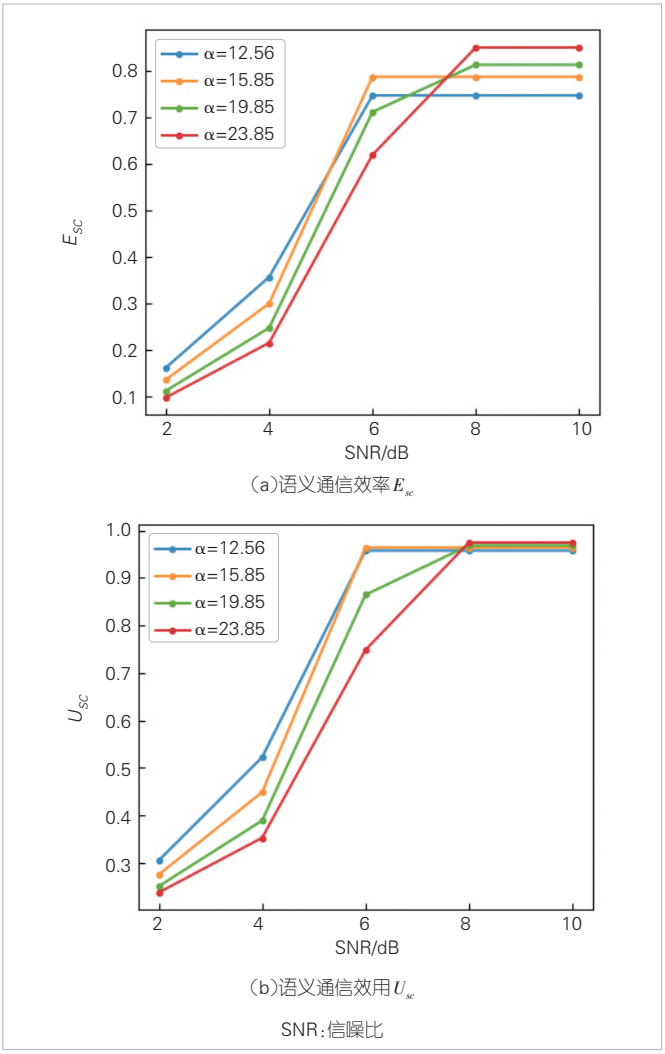
$$ACC(\gamma)=\mathrm{WB}-\mathrm{PESQ}(\gamma), \tag{21}$$

$$TIM(\gamma)=\text{传输时延}(\gamma)。 \tag{22}$$

同时，准确性边界、时效性边界与通信参数如表3所示。

本文采用日本电报电话公司（NTT）宽带语音数据集[20]的美式英语子集进行验证，同时为AMR－WB编码设定4组不同的速率参数$\alpha=\{12.65,15.85,19.85,23.85\,(\mathrm{kbit/s})\}$。采用更高码率的模型会更倾向于降低压缩比以获得更高的重建质量。模型同样采用LDPC信道编码与BPSK调制方式，并根据信噪比条件将LDPC的码率控制在1/5～8/9之间。

图4给出了$\lambda=1/5$条件下语义通信效率E_{sc}与效用U_{sc}随通信系统信噪比的变化趋势曲线。与面向图像重建任务的语义通信系统相似，随着信噪比的上升，效率与效用指标均表现出上升趋势。同时，在语义效用U_{sc}方面，低速率编码方案具有更显著的优势，能够反映出系统对时效性需求的侧重。在上述面向图像与语音重建任务的通信模型仿真中，除带宽之外的通信参数都是一致的。语义通信效率指标与效用指标的评估结果表现出统一性，为两者的横向性能对比提供指导。带宽参数的设定与信息模态相关，它所产生的影响可通过对变换函数进一步设计来消除。



▲图4 语音重建任务语义通信效率E_{sc}与效用U_{sc}随信噪比的关系曲线（$\lambda=1/5$）

4 结束语

通过融合人工智能与通信技术，充分利用语义层面信息的高度抽象、智能简约等特性，语义通信有望形成突破经典通信系统传输瓶颈的智能新型通信体系。由于语义通信尚处于研究初期，统一且具泛化价值的性能评估体系的缺失，阻碍了关键性成果间的横向贯通与对比。本文通过分析语义通信评估痛点，提出语义通信效率指标E_{sc}与语义通信效用指标U_{sc}。该指标具有更好的泛用性，且为不同任务场景、模态信息下的语义通信模型的横向性能对比提供指导，进而促进语义通信的快速发展。

参考文献

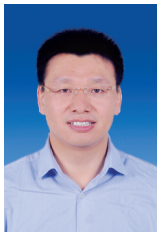
[1] 张平, 张建华, 戚琦, 等. Ubiquitous-X: 构建未来6G网络 [J]. 中国科学: 信息科学, 2020, 50(6): 913-930
[2] ZHANG P, XU W J, GAO H, et al. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks [J]. Engineering, 2022, 8(1): 60-73. DOI: 10.1016/j.eng.2021.11.003
[3] PAPINENI K, ROUKOS S, WARD T, et al. BLEU: a method for automatic evaluation of machine translation [C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. New York: ACM, 2002: 311－318. DOI: 10.3115/1073083.1073135
[4] 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

[5] BANERJEE S, LAVIE A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments[C]//Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization. 2005: 65–72.
[6] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment [C]//Proceedings of Thrity–Seventh Asilomar Conference on Signals, Systems & Computers. IEEE, 2004: 1398–1402. DOI: 10.1109/ACSSC.2003.1292216
[7] JIANG J. Image compression with neural networks – A survey [J]. Signal processing: image communication, 1999, 14(9): 737–760. DOI: 10.1016/S0923–5965(98)00041–1
[8] MINNEN D, BALLÉ J, TODERICI G. Joint autoregressive and hierarchical priors for learned image compression [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. ACM, 2018: 10794 – 10803. DOI: 10.5555/3327546.3327736
[9] LU G, OUYANG W L, XU D, et al. DVC: an end–to–end deep video compression framework [C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 10998–11007. DOI: 10.1109/CVPR.2019.01126
[10] RIX A W, BEERENDS J G, HOLLIER M P, et al. Perceptual evaluation of speech quality (PESQ)–a new method for speech quality assessment of telephone networks and codecs [C]//Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221). IEEE, 2002: 749–752. DOI: 10.1109/ICASSP.2001.941023
[11] TAAL C H, HENDRIKS R C, HEUSDENS R, et al. A short–time objective intelligibility measure for time–frequency weighted noisy speech [C]// Proceedings of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2010: 4214–4217. DOI: 10.1109/ICASSP.2010.5495701
[12] FU S W, WANG T W, TSAO Y, et al. End–to–end waveform utterance enhancement for direct evaluation metrics optimization by fully convolutional neural networks [J]. IEEE/ACM transactions on audio, speech, and language processing, 2018, 26(9): 1570–1584. DOI: 10.1109/TASLP.2018.2821903
[13] CHINCHOR N. MUC–4 evaluation metrics[C]//Conference on Message Understanding. Association for Computational Linguistics, 1992.
[14] MA Y K, PENG H Y, CAMBRIA E. Targeted aspect–based sentiment analysis via embedding commonsense knowledge into an attentive LSTM [C]//Proceedings of the AAAI conference on artificial intelligence. ACM, 2018, 32(1): 5876 – 5883
[15] BEKKAR M, DJEMAA H, ALITOUCHE T A. Evaluation measures for models assessment over imbalanced data sets [J]. Journal of information engineering and applications, 2013, 3: 27–38
[16] REN S Q, HE K M, GIRSHICK R, et al. Faster R–CNN: towards real–time object detection with region proposal networks [J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137–1149. DOI: 10.1109/TPAMI.2016.2577031
[17] XU W J, ZHANG Y M, WANG F Y, et al. Semantic communication for Internet of vehicles: a multi–user cooperative approach [EB/OL]. [2023–02–16]. https://arxiv.org/abs/2212.03037
[18] CHIU C, RAFFEL C. Monotonic chunkwise attention [EB/OL]. [2023–02–16]. https://arxiv.org/abs/1712.05382
[19] BALLE J, MINNEN D, SINGH S, et al. Variational image compression with a scale hyperprior [EB/OL]. [2023–02–16]. https://arxiv.org/abs/2212.03037
[20] NTT Advanced Technology Corporation (NTT–AT). Multi–lingual speech database for telephonometry [DB]. 1994

作 者 简 介

郑远，北京邮电大学在读博士研究生；主要研究领域为语义通信。

王凤玉，北京邮电大学讲师；主要研究方向为语义通信、无线人工智能、通感一体化等。



许文俊，北京邮电大学教授、鹏城实验室兼职研究员；主要研究领域为语义通信、B5G/6G智能通信网络；曾获国家技术发明奖二等奖、教育部技术发明奖一等奖、IEEE ICC 2019最佳论文奖等奖项；发表论文130余篇。

基于深度学习的图像语义通信系统

Deep Learning-Based Image Semantic Communication System

张振国/ZHANG Zhenguo，杨倩倩/YANG Qianqian，
贺诗波/HE Shibo
（浙江大学，中国 杭州 310058）
(Zhejiang University, Hangzhou 310058, China)

DOI：10.12142/ZTETJ.202302011
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1400.010.html
网络出版日期：2023-04-11
收稿日期：2023-02-06

摘要：语义通信是一种新颖的通信方式，可通过传输数据的语义信息提高带宽效率。提出一种用于无线图像传输的系统。该系统基于深度学习技术开发并以端到端（E2E）的方式进行训练。利用深度学习实现语义特征的提取和重建，在发送端提取信源信息不同类型和不同形式的语义特征，在接收端融合各类型语义特征进行目标语义恢复。仿真结果表明，与基准模型相比，所提模型在信道环境恶劣的情况下，具有更好的重建精度。

关键词：端到端通信；多级语义通信；图像压缩；图像传输

Abstract: Semantic communications is a novel form of communication that improves bandwidth efficiency by transmitting semantic information about data. A system for wireless image transmission is introduced, which is developed based on deep learning techniques and trained by an end-to-end (E2E) approach. Deep learning is used to extract and reconstruct semantic features, extract different types and forms of semantic features of source information at the sending end, and fuse various types of semantic features at the receiving end for target semantic recovery. The simulation results show that compared with the benchmark model, the proposed model has better reconstruction accuracy under the bad channel environment.

Keywords: end-to-end communication; multi-level semantic communication; image transmission

当前，中国正大力发展信息产业。无线通信技术的快速发展为智慧城市、高清视频、自动驾驶、远程医疗等带来产业变革。基于物联网的智能感知网络迅速发展，在提供便利的同时也带来了庞大的无线通信数据[1]。数据通信方式不再受限于人-人通信，而是转变为以目标为导向的通信方式[2]。随着人工智能的快速发展，基于深度学习的通信为克服传统通信困难提供新思路。基于深度学习的架构在通信系统中取得了显著成果，达到甚至超过传统方案的性能。基于香农定理的1G到5G技术已经无法满足当前的通信需求，基于深度学习的语义通信为下一代无线通信技术带来更多可能[3-5]。深度学习已被用于优化基于分离模块设计的传统通信，如信源编解码器、信道编解码器和调制解调模块[6-8]。文献[9]建立端到端（E2E）的信源信道联合编码（JSCC）通信系统，该系统能够有效应对传统通信系统中的瓶颈问题。E2E通信系统的发射器和接收器由深度神经网络（DNN）组成。编码器学习输入数据的特征向量，并通过无线信道将复数符号发送到解码器进行目标重建。

与传统通信相比，语义通信侧重于传递信息的含义，而不是符号的精确传输。区别于传统通信系统利用误码率（BER）或符号错误率（SER）评估通信结果，语义通信系统通过最小化输入和重建信息之间的语义损失，来恢复接收器处的信源信息[9]。现有的语义通信系统联合设计发射器和接收器，以实现更好的传输效率和鲁棒性[10]。深度学习在语义通信模型中得到广泛的应用，具有良好的特征提取和学习能力，能够对信源包含的语义信息进行提取和传输[11]。相较于数字通信通信模型，基于深度学习的通信模型不会出现“悬崖效应”。当前，语义通信主要基于深度学习进行开发和探索，并且已经取得一定的效果[12]。基于深度学习的语义通信系统显示出巨大的潜力，能够有效传输不同类型的信息。随着物联网设备的大量部署，以目标为导向的通信方式[2]（如人-机、机-机）不断涌现，这给边缘设备带来巨大的通

基金项目：国家自然科学基金（62201505）；浙江大学-新加坡科技设计大学创新、设计与创业联盟项目（202102）；中央大学基础研究基金（2021FZZX001-20）

信压力。语义通信只传输目标需要的信息，大大减少数据通信量，提高通信效率。语义通信将成为下一代物联网无线通信技术的重要组成部分。

信道噪声干扰是影响无线通信系统性能的主要因素之一，因此提高通信系统应对噪声环境的鲁棒性是传统通信和语义通信的共同目标。数字通信方案通过增加信道编码量来提高系统的抗噪能力，使通信量急剧增加。当前，基于深度学习的通信系统通过DNN缓解噪声对系统的干扰，同时平衡系统的通信量。文献[13]提出了一种用于通用无线信道的基于深度学习的端到端通信系统，其中信源编码、信道编码、调制解调等传统模块被DNN所取代。该端到端系统以数据驱动的方式成功地利用各种相关性，获得较好的结果。文献[14]提出了一种基于强化学习的方法，在不知道信道传递函数或信道状态信息（CSI）的情况下优化发送端DNN。在实际系统中，由于存在信令和检测方案的损伤、硬件缺陷、变化的信道条件等情况，经过训练的网络所得到的信道与用于训练网络的信道显著不同。文献[15]使用随机扰动方法设计了一个无信道模型的端到端通信框架。该模型在真实信道中训练基于深度学习的通信系统，不需要对信道模型进行任何假设。文献[16]开发了一种基于条件对抗生成网络（StyleGAN）的方法，用于构建端到端通信系统，其中StyleGAN用于构建信道效应模型。该模型将与导频数据相对应的接收信号作为调节信息的一部分，在接收机处获得信道状态信息，并将其用于信号检测。

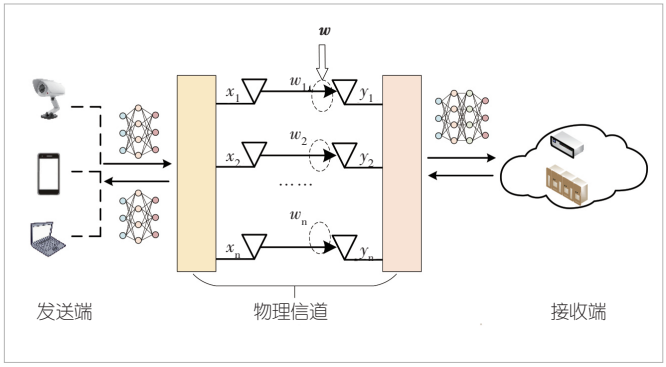
近年来，研究人员提出了多种基于深度学习的语义通信系统，旨在提高在噪声信道上自然语言传输的性能。文献[17]提出了一种用于自然语言的JSCC系统，通过最小化E2E失真，实现比传统方法更好的通信性能。文献[18]提出了两种方案来处理有限的数据速率问题：自适应传输方案和广义数据表示方案。其中，自适应传输方案可以在不同信道条件下以均方误差（MSE）约束最大化数据速率；广义数据表示方案取代了one-hot表示，以获得更高的数据速率。文献[19]则提出了一种新型语义通信系统，通过信道估计辅助训练实现文本重建。与文本语义通信系统相比，语音信号难以提取和表示基本的语义特征。研究人员使用多种方法来开发用于语音信号传输的语义通信系统，文献[20]提出了一种基于深度学习的语音通信系统。该系统通过联合训练语义和信道编解码器，学习和提取语音特征，同时利用注意力机制减轻实际通信场景中的信道失真和衰减，使系统获得更好的通信性能。文献[21]采用波向量（wav2vec）来提取音频语义特征，并通过强化学习（RL）提高特征提取的准确性。

针对图像信息的压缩和传输问题，基于深度学习的图像压缩技术已经成为研究热点。一系列的图像无线信号通信系统相继产生。文献[22]提出了一种基于深度学习的无线图像传输系统，实现基于E2E的JSCC，使用峰值信噪比（PSNR）和结构相似性指数（SSIM）测量重建图像的质量。在JSCC的基础上，另一种具有信道反馈的图像重建方案DeepJSCC-f，通过接收来自接收器的信道反馈，进一步提高图像重建精度[23]。文献[24]提出了基于正交频分复用（OFDM）数据路径的JSCC方案，用于多路径衰落信道的无线图像传输。该通信方案通过整合专家知识实现了更好的性能。文献[25]开展了物联网（IoT）设备用于图像传输的研究，通过将两个DNN的联合传输识别方案部署在设备端，在识别精度上比传统方案更好。文献[26]提出了联合特征压缩和传输系统，以处理边缘服务器上有限的计算资源。该方案不仅提高了E2E的可靠性，而且降低了计算复杂度。此外，M. JANKOWSKI等提出了基于自编码器严格约束的设备边缘通信系统，在有限的计算能力下实现了更好的分类精度[27]。文献[28]提出了一种新的航空图像传输范式，在无人机端部署一个轻量级模型，用于感知图像和信道条件的语义传输模块，在接收端通过计算能力更强的基站对收到的信息进一步处理，提高分类精度。文献[29]提出了基于深度学习的用于传输单模态和多模态数据的多用户语义通信系统，通过在编码器和解码器层之间添加连接来融合多模态数据，实现图像检索和视觉问答等功能。然而，当前仍缺乏对于图像语义的评估准则，导致语义通信系统的图像恢复仍需要使用传统的图像重建准则来衡量系统性能。

语义通信作为下一代通信技术的重要组成部分，为新一代的通信变革提供新方法。目前，语义通信发展仍处于探索阶段。作为下一代通信技术的重要组成部分，语义需要理论和技术的双重发展。目前，语义通信在数学理论证明方面仍未得到普遍认可，在技术方面深度学习的机理和可解释性仍然有待研究。本文是在技术层面对图像语义通信的一次探索，通过简单的通信环境假设，验证通信系统的可行性；通过基于深度学习的图像信息挖掘，探索信息内容的语义性对信息传递的影响。实验结果表明，本文所提出的语义通信系统具有较好的鲁棒性，特别是在有限带宽条件下更具优势。这表明高级语义信息在图像传输中具有优势。

1 语义系统模型

图像语义通信模型的一般框架如图1所示。与传统通信系统相同，语义通信系统包含发送端、无线信道和接收端3个部分。两者的主要区别在于语义通信系统所采用的编解码方式。语义通信系统将信源信道的编解码作为一个整体。信



▲图1 图像语义通信系统

道编解码在应对噪声的同时对信源编码信息的重要程度进行衡量，为其中重要的信息分配更多的符号位以进行编码。针对信源中的语义内容，语义编码对有含义的信息进行编码表达，其过程实际就是对信源中语义概念的高度抽象与压缩。为提高图像无线通信的高效性和准确性，本文提出一种面向语义的图像通信系统，该系统通过联合信源语义编解码和信道编解码的方式，提高系统的通信效率和鲁棒性。

1.1 编码器

编码器由两部分构成：语义特征提取器和联合语义信道编码器。具体过程为：首先，编码器的输入图像S由归一化层预处理，使得每个元素都在[0,1]范围内；其次，通过多个基于神经网络的多级语义特征提取器提取输入图像的不同语义特征；最后，利用联合语义信道编码器将语义特征编码为符号，通过物理通道传输到接收器，传输向量x：

$$x = \mathrm{T}_{\beta}(\mathrm{T}_{\alpha}(S)), \tag{1}$$

其中，$\mathrm{T}_{\alpha}(\cdot)$为多级语义特征提取网络，网络参数为$\alpha$；$\mathrm{T}_{\beta}(\cdot)$为联合语义信道编码器，网络参数为$\beta$。

本文所提方案在一个广泛使用的物理信道——加性白高斯噪声（AWGN）信道上进行模型测试。则解码器上接收到的信号y为：

$$y = x + w, \tag{2}$$

其中，w为圆对称高斯分布的独立同分布向量，$w \sim \mathcal{CN}(0,\sigma^2\boldsymbol{I})$，$\sigma^2$为信道的平均噪声功率，$\boldsymbol{I}$为单位矩阵。

1.2 解码器

解码器也由两部分组成：联合语义信道解码器和图像重建模块。联合语义信道解码器减轻信号在AWGN信道的噪声干扰，并恢复多级语义特征。图像重建模块融合不同层次的语义信息并重建目标图像。反归一化层将每个元素重新缩放为图像像素值（0～255）。联合语义信道解码器和图像重建模块的参数为ξ和η，接收器重建图像为：

$$\hat{S} = \mathrm{R}_{\eta}(\mathrm{R}_{\xi}(y)), \tag{3}$$

其中，$\mathrm{R}_{\xi}(\cdot)$和$\mathrm{R}_{\eta}(\cdot)$分别为联合语义信道编码器和图像重建模块，y为从信道接收到的信号。

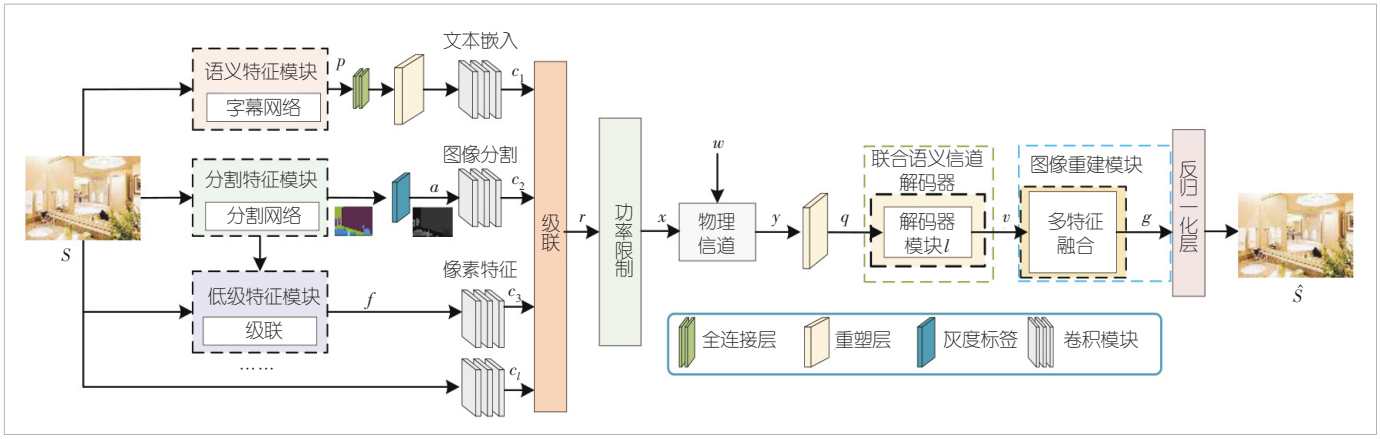
图像语义通信系统的目标是最小化输入图像S和重建图像$\hat{S}$之间的平均误差，均方差（MSE）用来评估S和$\hat{S}$之间的失真：

$$\mathcal{L} = \frac{1}{N}\sum_{k=1}^{N} d(S_k, \hat{S}_k), \tag{4}$$

其中，$d(S,\hat{S}) = \frac{1}{n}\|S - \hat{S}\|^2$为均方误差分布，$N$为样本数。

2 多级图像语义通信模型

本文提出一种基于深度学习的无线图像传输语义通信系统，如图2所示。其中，多级语义特征提取器用于提取不同级别的语义特征。高级语义信息包含图像的抽象性和通用性



▲图2 图像语义通信系统的整体架构

指标，低级语义信息包含图像的局部细节语义信息[30]。该系统通过基于深度学习的特征提取器提取信源特征，并通过与语义信道的联合训练给不同的信息赋予不同的权重。语义信道编码器和解码器联合在接收器处成功恢复这些语义特征，并通过图像重建模块对多级语义信息进行融合并重构目标图像。

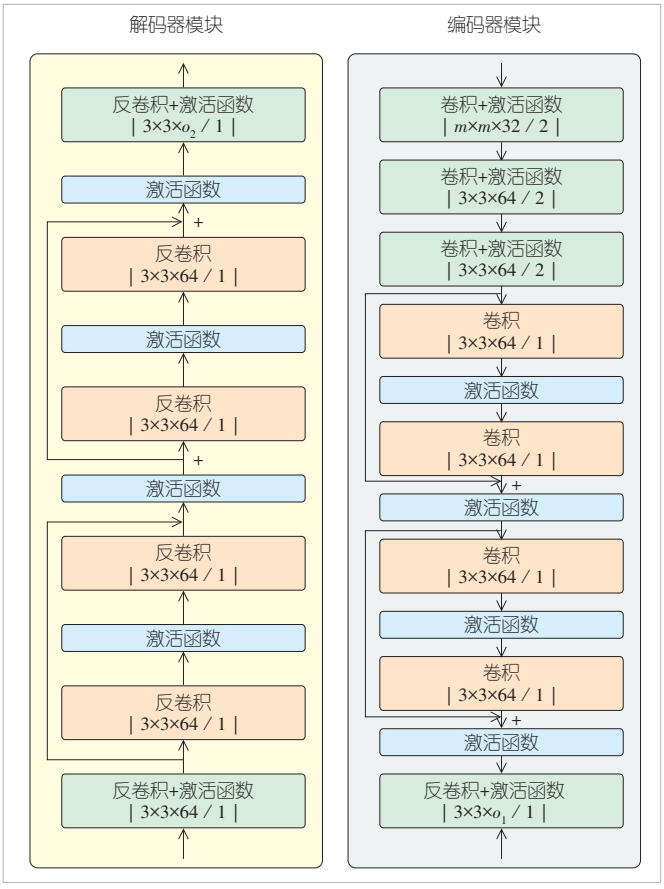
2.1 多级语义特征提取器

对于输入图像$S \in \Re^{b\times h\times w\times 3}$，首先通过归一化层将像素值映射到[0,1]范围，其中b、h和w分别为图像的批数量、高和宽，3表示图像对应的图像通道数。一个批次的图像数据会被送入多级语义特征提取器。该提取器包含3个模块：语义特征模块、分割特征模块和低级特征模块。不同模块分别提取不同层级和不同形式的图像语义信息。语义特征模块包含一个预训练的图像字幕模型，该模型由ResNet-152模型[31]和长短期记忆网络（LSTM）[32]层组成，用于提取图像文本形式的高级语义信息$p \in \Re^{b\times 2\cdot h\cdot w/t^2\times 1}$，其中$t$为下采样因子。文本形式语义特征是带有图像纹理信息的文本嵌入信息。为了方便理解，本文称这些特征为“文本形式”。分割特征模块通过预训练的图像分割[33]模型获取图像分割形式的高级语义信息$a \in \Re^{b\times h\times w\times 1}$。低级特征模块由两部分组成：联合部分和直取部分。其中，联合部分通过concatenate按通道维度将归一化图像和分割特征相结合，并输出$f \in \Re^{b\times h\times w\times 1}$；直取部分为直接输入归一化后的图像特征。低级特征模块主要用于图像细节特征的提取，为高级语义信息做细节补充，实现更丰富的信息重构。超参数l控制低级特征传输信息量，该参数决定通信量和目标信息的重建质量。

2.2 联合语义信道编解码

通过多级语义特征提取器可获得4种类型的目标语义信息特征：文本形式特征p、图像分割特征a、联合特征f和图像归一化低级特征。语义特征p首先通过全连接层进行维度压缩，进而通过整形层和编码模块输出$c_1 \in \Re^{b\times \frac{h}{t}\times \frac{w}{t}\times 1}$，其中×1表示维度信息。语义特征$a$和$f$分别输入两个独立的编码器模块，并输出不同的特征维度信息$c_2 \in \Re^{b\times \frac{h}{t}\times \frac{w}{t}\times 1}$和$c_3 \in \Re^{b\times \frac{h}{t}\times \frac{w}{t}\times 3}$。$c_3$和$c_l$除了最后一层的输出通道数不同外，其他结构均相同。编码器模型c_l获取图像细节信息特征$c_l \in \Re^{b\times \frac{h}{t}\times \frac{w}{t}\times l}$。联合层将获取的不同内容的特征$c_1,\cdots,c_l$进行联合和压缩，并输出$r \in \Re^{b\times \frac{h}{t}\times \frac{w}{t}\times e}, e=1,2,\cdots,3\times l-4$；进而通过功率归一化层生成$k$个数传输符号$x$，同时这些符号的发射功率低于给定值，其中$x \in \mathrm{E}^{b\times k}$，$k=\frac{h}{t}\times\frac{w}{t}\times e$。信道带宽压缩比为$k/n=\frac{3\times t^2}{g}$，$g=1,\cdots,3\times l-4$，该值可根据信道环境自适应调整。高级语义信息是目标重建的增强信息。当$e=1$时，联合语义信道编码输出为$c_3 \in \Re^{b\times \frac{h}{t}\times \frac{w}{t}\times 1}$。

接收端的重塑层将接收到的信号重组为$q_i, i=1,\cdots,l$个语义特征，其尺寸为$b\times\frac{h}{t}\times\frac{w}{t}\times(3\times l-4)$。当连接接收到的符号的实部和虚部时，最后一个维度加倍，如$q_1 \in \Re^{b\times \frac{h}{t}\times \frac{w}{t}\times 2}$。同样，$q_i$使用第$3i-1$和$3i+1$元素的最后一个维度，$q_i \in \Re^{b\times \frac{h}{t}\times \frac{w}{t}\times 6}$。每个$q_i, i=1,\cdots,l$都输入到解码器模块，其具有相同的卷积核尺寸（$m$=3），如图3所示。每个解码器的输出语义特征为$d_i \in \Re^{b\times \frac{2h}{t}\times \frac{2w}{t}\times o}, i=1,\cdots,l$，其中$o$为输出通道数。卷积神经网络的架构由$m\times m\times o/st$参数构



▲图3 编码器模块、解码器模块结构示意图

成，其中m、o和st分别是卷积核尺寸、通道输出数量和步长。每个卷积层之后为广义的归一化变换层（包括广义除数归一化（GDN）和逆GDN（IGDN））。

2.3 图像重建模块

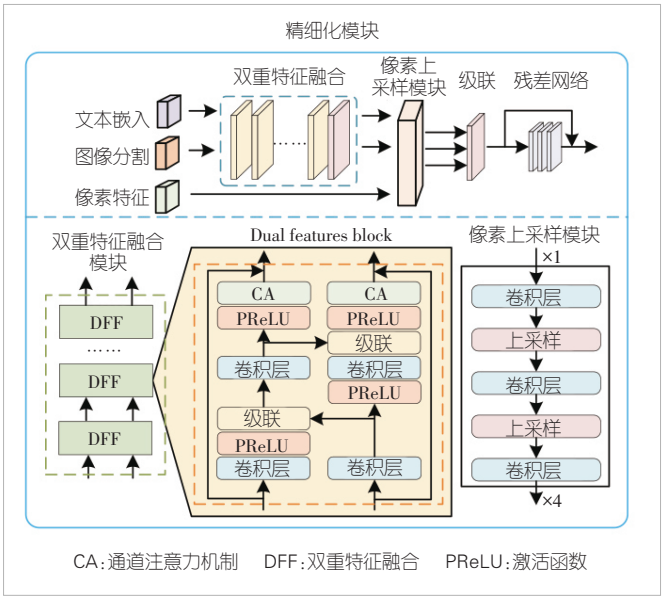
图像重建模块需要融合不同形式和内容的语义特征，完成不同语义内容之间的相互补充，通过注意力机制深度挖掘融合信息，进而将融合特征重建为目标图像。首先借助双特征融合模块对两种形式的高级语义特征进行融合，即双特征融合模块通过交叉结构和通道注意力机制（CA）学习输入特征；然后通过像素上采样模块对特征信息升维，其中像素上采样模块由卷积层和像素上采样层构成；最后将相同维度的高级语义信息和低级细节补充信息进行级联操作，通过残差网络对融合后的信息进行提取并重建目标图像。残差网络由反卷积层和PReLU激活函数构成（最后一层为sigmoid激活函数），其网络结构如图4所示。在图像重建模块中，不同形式特征生成的粗糙图像含有不同的成分，例如比较平滑的低频信息和充满边缘、纹理的高频信息。同时，卷积层的每个过滤器都包含一个局部感受野，其输出无法利用局部信息之外的上下文信息。因此，通过通道注意力机制改变特征权重，能够提高重要信息的权重占比。通道注意力机制的网络结构如图5所示。

3 仿真结果与分析

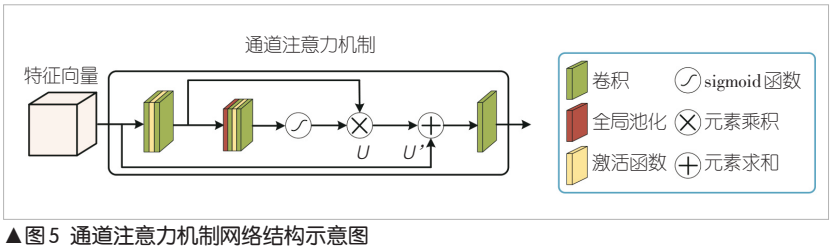
3.1 仿真参数

根据设计模型特点，需要使用具有多类型标签的数据集对模型进行训练。本文使用MSCOCO[34]和ADE20K[35]数据集训练和评估所提出的系统。MSCOCO数据集包含123 287张图像（82 783张用于训练，40 504张用于测试），每张图像都包含5个不同的字幕标题。ADE20K数据集包含27 574张图像、150个语义标签，图像高度和宽度至少为512像素。训练模型时，每个图像都被裁剪为固定大小：h = 128，w = 128。对于文本形式特征，本文使用带有文本标签的MSCOCO数据集来训练语义特征模块CaptionNet[24-25]，并使用带有分割标签的ADE20K数据集来训练系统的其余部分（冻结语义特征模块）。最后，本文在Kodak[36]图像数据集上测试所提出的图像无线通信系统。Kodak数据集总共包含24张固定尺寸768×512的图像。

所提模型在Pytorch[37]中实现。并使用Adam算法[38]进行优化。学习率设置为0.000 1，批次大小设置为32，下采样因子t = 8。本文采用现有的基于深度学习的方法（DeepJSCC和Aided Deep-JSCC[39]），以及传统的基于分离的数字传输方案（JPEG）作为基准进行比较，并使用PSNR和SSIM指标来评估性能。PSNR表示信号最大可能功率和影响其表示精度的破坏性噪声功率的比值，$\mathrm{PSNR} = 10\log_{10}\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\mathrm{dB}$，其中$d(x,y)$为均值误差，MAX表示图像点颜色的最大数值。SSIM使用3个标准度量图像：亮度、对比度和结构。$\mathrm{SSIM} = \frac{(2\mu_s\mu_{\hat{s}} + C_1)(2\sigma_{s\hat{s}} + C_2)}{(\mu_s^2 + \mu_{\hat{s}}^2 + C_1)(\sigma_s^2 + \sigma_{\hat{s}}^2 + C_2)}$，其中$\mu$为亮度，$\sigma$为对比度，$C_1$和$C_2$为常数。Aided Deep-JSCC是一种用于无线传感器网络的分布式语义通信方案。其中，每个传感器观察并编码一个公共图像，并通过无线信道独自将其发送给接收器。接收器融合来自不同传感器的信息并重建原始图像。该方案类似于本文所提方案，即从原始图像中提取多个版本的语义信息，并将其发送给接收者以用于信源信息的重建。根据香农分离定理，传输速率必须降低信道容量，通过物理信道传输的信源信息需要满足$R \leqslant \frac{k}{n}\log_2(1 + SNR)$，其中$R$为信源信



▲图4 图像重建模块结构示意图



▲图5 通道注意力机制网络结构示意图

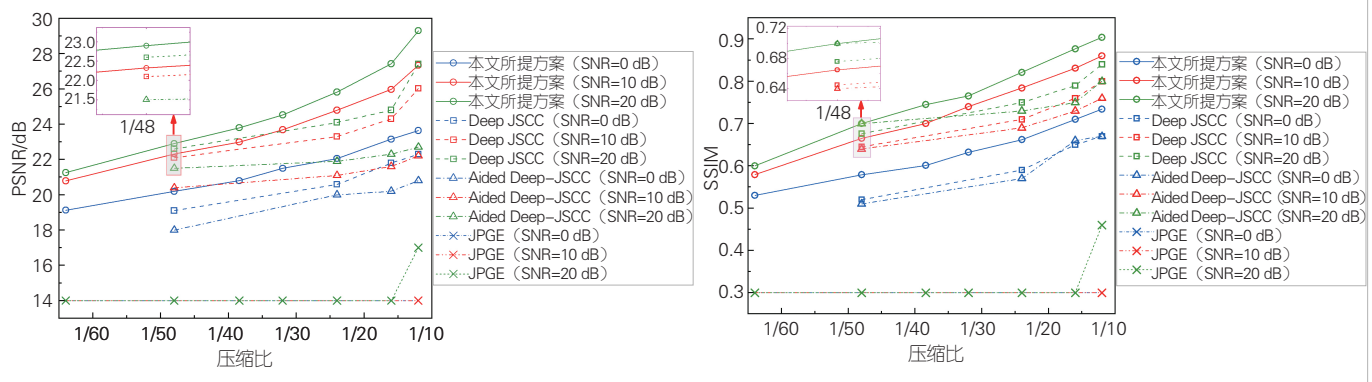
息的传输速率，$\log_2(1+SNR)$为信道容量，k为信道维度，n为图像维度。信源信息在物理信道上可靠传输的最大速率为R_{max}。传统的图像压缩方案具有最小压缩码率R_{min}，这是接收端重建目标图像的极限。如果$R_{min} > R_{max}$，则接收器无法重建输入图像。$SNR = 10\lg\frac{P_s}{P_n}$为信噪比，其中$P_s$为信号功率，$P_n$为噪声功率。

3.2 结果分析

图6对比了不同压缩比下图像重建PSNR的性能表现。可以看出，本文所提算法优于其他深度学习方案和传统方案。同时，基于深度学习的图像通信系统不会因“悬崖效应”（信道条件低于某个阈值，接收器无法恢复传输的图像）而遭受大幅的性能下降。数字传输方案在较差的信道环境中（SNR<10 dB）和低压缩比（k/n < 1/10）条件下引发系统崩溃，而基于深度学习的系统仍然可以完成目标信息的重建。本文所提方案中的联合语义信道编解码方案能够均衡分配信源信息编码和抗噪编码字符数，使得在信号压缩和重建过程中能够更好地表达语义信息。在低压缩比信道情况较好的环境中，本文所提方案性能与DeepJSCC模型相近。这是由于：在低压缩比下，发射端没有足够的带宽来传达更详细的图像细节信息，导致系统重建目标在细节信息的刻画上不足。随着压缩比的增大，本文所提算法性能与其他方案之间的差距逐渐拉大。这进一步体现出所提模型在抗噪和信息恢复上的优势。

图6同时展示了不同方法在SSIM评估标准下的性能比较。SSIM从图像结构方面反映了原始图像和重建图像之间的相似性。本文所提方案在低压缩比及较差的信道条件下（例如，SNR=0）的性能显著优于其他方案。Aided Deep-JSCC系统将SSIM和MSE作为损失函数，在SSIM指标上具有更好的性能。本文所提方法的图像重建模块对高级语义信息进行深度挖掘，并利用残差网络对融合后的多级信息进行学习，进而提高图像重建质量；在低压缩比信道环境较好的情况下，依然能够达到与Aided Deep-JSCC方案相同的性能（本文所提方案中并未使用SSIM作为损失函数）。在高压缩比情况下，Aided Deep-JSCC的重建质量最差。这是由于：Aided Deep-JSCC方案中的每个传感器都独立于图像进行编码，在不同编码器之间引入了语义信息冗余，造成不必要的带宽浪费。而本文所提出的方案以联合的方式提取多级语义信息，强制不同编码器获得的语义信息不同，相对提高了不同压缩比在较差物理信道情况下（SNR<10 dB）SSIM的评分，进一步证明了所提方案在带宽有限、信道条件恶劣情况下的优势。

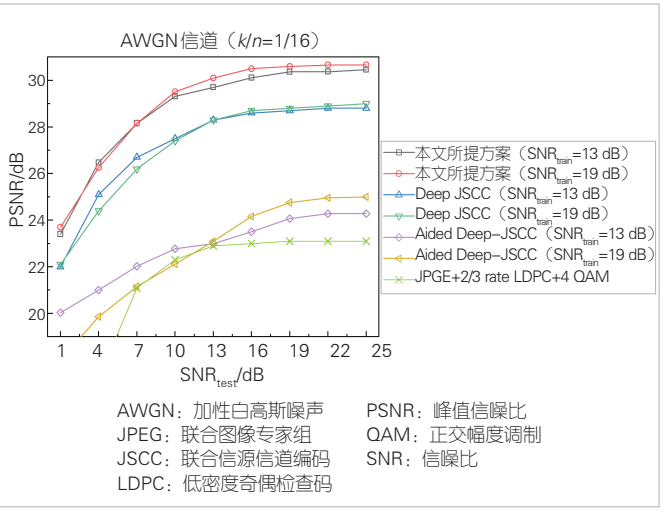
图7比较了本文所提方案和基准方法在Kodak数据集上的PSNR性能表现。在AWGN信道环境下压缩比k/n设置为1/16，数字通信方案信源编解码采用JPEG，信道编解码为低密度奇偶检查码（LDPC），使用4符号正交幅度调制（QAM）数字调制方案。如图7所示，本文所提方案优于其他基于深度学习的方法，表现出多级语义信息在无线图像传输上的优势。同时，采用数字通信方案的图像传输质量远低于基于深度学习的方法。造成该现象的主要原因是传统方案对信道变化引起的信道错误率比较敏感。由于可用于压缩的位数是固定的，当信道环境达到一定阈值后，数字通信方案图像重建质量不会再提高。在$SNR_{test} < SNR_{train}$的情况下，本文所提系统不会受到数字传输方案中的“悬崖效应”的影响。相反，本文所提方案随着信噪比值的降低，性能平滑地



（a）不同方法在PSNR评估标准下的性能　（b）不同方法在SSIM评估标准下的性能比较

AWGN：加性白高斯噪声　JPEG：联合图像专家组　JSCC：联合信源信道编码　PSNR：峰值信噪比　SNR：信噪比　SSIM：结构相似性指数

▲图6 AWGN信道上具有不同压缩比的PSNR和SSIM性能比较



▲图7 不同图像无线传输方案在Kodak数据集上PSNR的性能比较

下降，其能够平衡压缩比和鲁棒性之间的重要层度。如果模型以高SNR值训练，则系统的PSNR性能主要由带宽压缩比决定，反之亦然。本文所提方法的PSNR性能，比Deep JSCC高1.37～1.96 dB，比Aided Deep-JSCC高3.37～7.39 dB，比传统方法高6.8～9.06 dB。相较于Aided Deep-JSCC的指标波动性，本文所提方法在面对不同信道环境时具有更平滑的性能表现。如图8所示，相比于静态信道环境，本文所提方案的图像重建质量仍具有较好表现，这说明网络具有估计信道状态的能力。

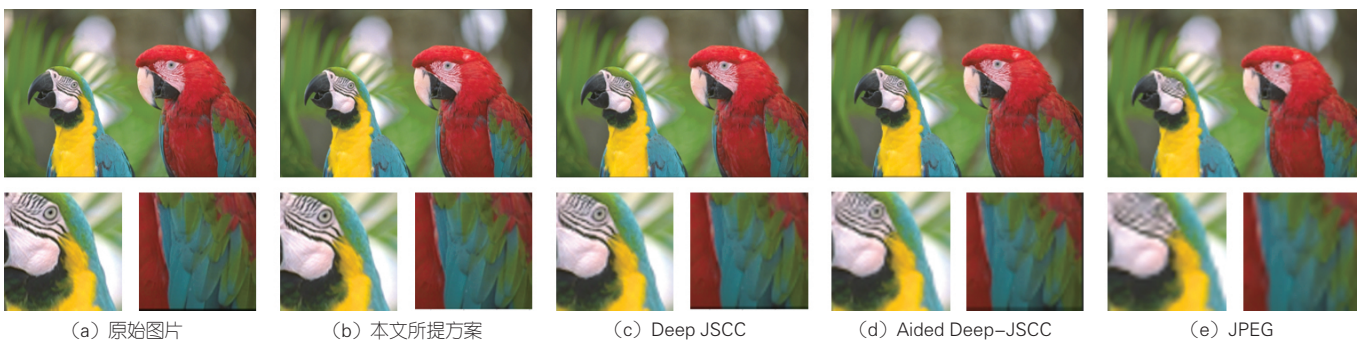
4 结束语

本文提出了一种基于深度学习的无线图像传输语义通信系统，与其他基于深度学习和基于分离的数字传输方案相比，性能表现优异。仿真结果表明，所提出的语义通信系统的有效性和鲁棒性均优于其他方案。

随着无线通信技术的快速发展，未来6G高效、智能的通信方式，将给工业生产和大众生活带来颠覆性的变革。语义通信作为新一代通信方式，将助力6G发展。然而，有关语义通信的关键技术和理论基础仍需要进一步发展和完善。本文仅作为语义通信的一次尝试和探索。语义通信技术的发展还需要更多的专家学者共同推进。

参考文献

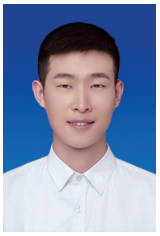
[1] 徐英姿, 刘原, 时梦然, 等. 语义在通信中的应用综述 [J]. 电信科学, 2022, 38(S1): 43-59. DOI: 10.11959/j.issn.1000-0801.2022107
[2] 刘传宏, 郭彩丽, 杨洋, 等. 面向智能任务的语义通信: 理论、技术和挑战 [J]. 通信学报, 2022, 43(6): 41-57. DOI: 10.11959/j.issn.1000-436x.2022117
[3] CALVANESE STRINATI E, BARBAROSSA S. 6G networks: beyond Shannon towards semantic and goal-oriented communications [J]. Computer networks, 2021, 190: 107930. DOI: 10.1016/j.comnet.2021.107930
[4] STRINATI E C, BELOT D, FALEMPIN A, et al. Toward 6G: from new hardware design to wireless semantic and goal-oriented communication paradigms [C]//Proceedings of ESSCIRC 2021 – IEEE 47th European Solid State Circuits Conference (ESSCIRC). IEEE, 2021: 275-282. DOI: 10.1109/ESSCIRC53450.2021.9567793
[5] WANG C X, HUANG J, WANG H M, et al. 6G oriented wireless communication channel characteristics analysis and modeling [EB/OL]. [2023-01-10]. https://arxiv.org/abs/2007.13958
[6] UMAIR M M, MILAN S, GIUSEPPE V, et al. Improved deep distributed light field coding [J]. IEEE open journal of circuits and systems, 2021, 2: 325-337
[7] YE H, LI G Y, JUANG B H. Power of deep learning for channel estimation and signal detection in OFDM systems [J]. IEEE wireless communications letters, 2017, 7(1): 114-117. DOI: 10.1109/LWC.2017.2757490
[8] WANG Y, LIU M, YANG J, et al. Data-driven deep learning for automatic modulation recognition in cognitive radios [J]. IEEE transactions on vehicular technology, 2019, 68(4): 4074-4077. DOI: 10.1109/TVT.2019.2900460
[9] TSE D, VISWANATH P. Fundamentals of wireless communication [M]. Cambridge: Cambridge University Press, 2005
[10] BASU P, BAO J, DEAN M, et al. Preserving quality of information by using semantic relationships [J]. Pervasive and mobile computing, 2014, 11: 188-202. DOI: 10.1016/j.pmcj.2013.07.013
[11] 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
[12] YE H, LI G Y, JUANG B H. Deep learning based end-to-end wireless communication systems without pilots [J]. IEEE transactions on cognitive communications and networking, 2021, 7(3): 702-714. DOI: 10.1109/TCCN.2021.3061464
[13] YE H, LIANG L, LI G Y, et al. Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels [J]. IEEE transactions on wireless communications, 2020, 19(5): 3133-3143. DOI: 10.1109/TWC.2020.2970707
[14] AOUDIA F A, HOYDIS J. Model-free training of end-to-end communication systems [J]. IEEE journal on selected areas in communications, 2019, 37(11):



▲图8 在SNR=13 dB和压缩比为1/16的情况下不同方法的可视化比较

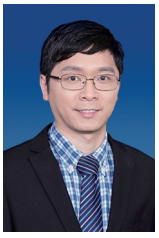
2503–2516. DOI: 10.1109/JSAC.2019.2933891
[15] RAJ V, KALYANI S. Backpropagating through the air: deep learning at physical layer without channel models [J]. IEEE communications letters, 2018, 22(11): 2278–2281. DOI: 10.1109/LCOMM.2018.2868103
[16] YE H, LI G Y, JUANG B H F, et al. Channel agnostic end–to–end learning based communication systems with conditional GAN [C]//Proceedings of 2018 IEEE Globecom Workshops (GC Wkshps). IEEE, 2019: 1–5. DOI: 10.1109/GLOCOMW.2018.8644250
[17] FARSAD N, RAO M, GOLDSMITH A. Deep learning for joint source–channel coding of text [C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 2326–2330. DOI: 10.1109/ICASSP.2018.8461983
[18] CHEN X, CHENG J L, ZHANG Z C, et al. Data–rate driven transmission strategies for deep learning–based communication systems [J]. IEEE transactions on communications, 2020, 68(4): 2129–2142. DOI: 10.1109/TCOMM.2020.2968314
[19] 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
[20] WENG Z Z, QIN Z J. Semantic communication systems for speech transmission [J]. IEEE journal on selected areas in communications, 2021, 39(8): 2434–2444. DOI: 10.1109/JSAC.2021.3087240
[21] TONG H N, YANG Z H, WANG S H, et al. Federated learning for audio semantic communication [J]. Frontiers in communications and networks, 2021, 2: 43. DOI: 10.3389/frcmn.2021.734402
[22] 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
[23] KURKA D B, GÜNDÜZ D. DeepJSCC–f: deep joint source–channel coding of images with feedback [J]. IEEE journal on selected areas in information theory, 2020, 1(1): 178–193. DOI: 10.1109/JSAIT.2020.2987203
[24] YANG M Y, BIAN C H, KIM H S. Deep joint source channel coding for wireless image transmission with OFDM [C]//Proceedings of ICC 2021 – IEEE International Conference on Communications. IEEE, 2021: 1–6. DOI: 10.1109/ICC42927.2021.9500996
[25] LEE C H, LIN J W, CHEN P H, et al. Deep learning–constructed joint transmission–recognition for Internet of Things [J]. IEEE access, 2019, 7: 76547–76561. DOI: 10.1109/ACCESS.2019.2920929
[26] JANKOWSKI M, GÜNDÜZ D, MIKOLAJCZYK K. Wireless image retrieval at the edge [J]. IEEE journal on selected areas in communications, 2021, 39(1): 89–100. DOI: 10.1109/JSAC.2020.3036955
[27] JANKOWSKI M, GÜNDÜZ D, MIKOLAJCZYK K. Joint device–edge inference over wireless links with pruning [C]//Proceedings of 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2020: 1–5. DOI: 10.1109/SPAWC48557.2020.9154306
[28] KANG X, SONG B, GUO J, et al. Task–oriented image transmission for scene classification in unmanned aerial systems [J]. IEEE transactions on communications, 2022, 70(8): 5181–5192. DOI: 10.1109/TCOMM.2022.3182325
[29] XIE H Q, QIN Z J, TAO X M, et al. Task–oriented multi–user semantic communications [J]. IEEE journal on selected areas in communications, 2022, 40(9): 2584–2597. DOI: 10.1109/JSAC.2022.3191326
[30] LIU X X, XU Q Y. Adaptive attention–based high–level semantic introduction for image caption [J]. ACM transactions on multimedia computing, communications, and applications, 2020, 16(4): 1 – 22. DOI: 10.1145/3409388
[31] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 770–778. DOI: 10.1109/CVPR.2016.90
[32] YU Y, SI X S, HU C H, et al. A review of recurrent neural networks: LSTM cells and network architectures [J]. Neural computation, 2019, 31(7): 1235–1270. DOI: 10.1162/neco_a_01199[PubMed]
[33] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 6230–6239. DOI: 10.1109/CVPR.2017.660
[34] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context [EB/OL]. [2023–01–10]. https://arxiv.org/abs/1405.0312
[35] ZHOU B L, ZHAO H, PUIG X, et al. Scene parsing through ADE20K dataset [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 5122–5130. DOI: 10.1109/CVPR.2017.544
[36] FRANZEN R. Kodak lossless true color image suite [EB/OL]. [2023–01–10]. https://r0k.us/graphics/kodak/
[37] PASZKE A, GROSS S, MASSA F, et al. Pytorch: an imperative style, high–performance deep learning library [EB/OL]. [2023–01–10]. https://arxiv.org/abs/1912.01703
[38] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. [2023–01–10]. https://arxiv.org/abs/1412.6980
[39] YAN J T, HUANG J H, HUANG C. Deep learning aided joint source–channel coding for wireless networks [C]//Proceedings of 2021 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2021: 805–810. DOI: 10.1109/ICCC52777.2021.9580301

作 者 简 介



张振国，浙江大学在读博士研究生；主要研究方向为基于深度学习的语义通信技术。

杨倩倩，浙江大学百人计划研究员、IMT–2030无线AI推进组成员，IEEE VTC 2022、IEEE ICC 2023等旗舰会议的技术委员会成员/分会主席，《IEEE Transactions on Wireless Communications》《IEEE Journal on Selected Areas in Communications》等刊物的审稿人；主要研究领域为无线通信、信息理论、语义通信；获中国科技产业化促进会科技创新奖一等奖等奖项；发表论文40余篇。



贺诗波，浙江大学控制科学与工程学院长聘教授，工业信息物理融合系统省部共建协同创新中心主任、工程师学院双专副院长、工业控制研究所副所长，2014年入选国家青年人才计划、浙江大学百人计划，IEEE高级会员、IET Fellow；获得教育部自然科学奖一等奖、IEEE通信学会亚太区杰出青年研究学者奖、中国自动化学会首届优秀博士学位论文奖，并获IEEE iThings、IEEE RFID–TA、IEEE WCNC、IEEE PIMRC最佳论文奖等奖项。

基于语义重要度的不等错误保护数据传输机制

Semantic Importance-Aware Data Transmission with Unequal Error Protection

郭帅帅/GUO Shuaishuai，李树静/LI Shujing
（山东大学，中国 济南 250062）
(Shandong University, Jinan 250062, China)

DOI：10.12142/ZTETJ.202302012
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1241.002.html
网络出版日期：2023-04-11
收稿日期：2023-02-16

摘要：语义通信是一种智能、简约的通信新范式。从系统组成和理论基础两方面，对语义通信和传统通信进行对比，指出其关键区别在于背景知识的引入。将基于信源信道联合设计的语义通信（JSCC-SC）分为面向数据重建和面向任务执行两类，并对其网络架构进行比较。提出一种信源信道分离设计的语义通信系统。该系统通过训练好的神经网络识别数据的语义重要度，并基于语义重要度构造不等错误保护（UEP）传输机制。通过仿真实验证明了其能够提升重要数据的传输可靠性，即降低重要数据传输的误码率，在保持复杂性的同时，进一步提升无线通信的可靠性与有效性。

关键词：语义通信；语义信息论；6G；智能通信

Abstract: Semantic communications is a new intelligent and simple communication paradigm. Semantic communications are compared with traditional communications in terms of system composition and theoretical basis. It is pointed out that the key difference lies in the introduction of background knowledge. The existing semantic communications based on the joint source channel coding (JSCC-SC) research is classified into two categories: data-reconstruction-oriented and task-execution-oriented. Then their network architectures are compared. A data importance-aware semantic communications system with separated source and channel coding is proposed. The system learns the semantic importance of data through a pre-trained neural network, and constructs a transmission mechanism with unequal error protection (UEP) based on the semantic importance. Simulation shows that it can improve the transmission reliability of important data, that is, reduce the bit error rate of important data transmission, and further improve the reliability and effectiveness of wireless communication while maintaining complexity.

Keywords: semantic communication; semantic information theory; 6G; intelligent communication

随着人工智能的发展，各种新型通信对象大量涌现。通信网络不仅要支撑人与人（H2H）之间的交互，还要支撑人与智能机器（H2SM）的交互和智能机器与智能机器（SM2SM）的交互。星地融合的全覆盖和机器-机器意图通信或将成为6G使能的创新应用[1]。然而，传统通信系统中信息压缩和传输的方法已经逼近香农信息论极限。语义通信有望打破这一瓶颈，成为6G的关键技术[2]。语义通信关注数据中蕴含的语义，结合背景知识提取与任务相关的语义，其通信目标是接收端在语义层面上恢复信息。现有关于语义通信的研究大多基于信源信道联合设计的思路，在仿真中表现出优越的性能，但在神经网络训练耗时和通信安全等方面仍存在一些问题[3]。本文从信源信道分离设计的角度出发，提出了一种基于语义重要度构造不等错误保护的传输机制。

1 传统通信与语义通信

传统通信以香农经典信息论为指导，按照技术堆叠的思路发展，系统复杂度和对资源的消耗都不断提高。语义通信拓展香农信息论，挖掘信息的语义维度，实现信息的高度抽象表征和智能简约传输。本节将从系统组成和理论基础两方面讨论传统通信和语义通信之间的异同。

1.1 系统组成

传统通信系统模型如图1所示，在信源到信宿之间采用

基金项目：国家自然科学基金（62171262）；山东省重大科技创新工程（2020CXGC010109）；山东省自然科学基金（ZR2021YQ47）；山东省泰山人才工程（tsqn201909043）

信源、信道分离编解码。信源编码把经过采样和量化的模拟信号变换成数字脉冲，信道编码按照一定的协议对信息进行纠、检错编码，以弥补由高斯白噪声（AWGN）、信道衰落等引起的信息错误。信源、信道译码是编码过程的逆过程。

信源信道联合设计的语义通信系统模型如图2所示，该模型除采用信源信道联合编码外，还引入了语义编解码功能块。语义编码器和语义解码器都会使用神经网络。语义编码器从要传输的数据中提取出与任务相关的语义信息，并进行语义编码，降低传输数据的维数；语义解码器接收语义信息，并根据背景知识进行语义恢复。在这个过程中存在两种噪声，即物理信道中的物理噪声和语义信道中的语义噪声。背景知识库不匹配，或者传输内容存在歧义，都是语义噪声的来源[3]。语义噪声的存在可能使信源和信宿之间产生误解，导致语义失真。建立知识库可以解决这一问题。语义通信系统是一种基于知识的系统[3]。背景知识库的建立十分耗时，它在训练的过程中通过对环境的感知学习不断更新，但发送端和接收端所处环境与学习过程不同会导致两者的知识库不匹配[4]。通过收发端知识库共享可以减少这种不匹配，提高语义通信的准确性。但在实际应用中，出于对通信开销和隐私保护等因素的考虑，背景知识完全共享难以实现。

信源信道分离设计的语义通信系统模型如图3所示。信源信道的编解码和调制解调可使用经典的技术。添加一个训练好的神经网络可使系统根据背景知识对数据进行学习。引入神经网络是为了识别数据的语义重要度，以便对具有不同语义重要度的数据设置差异化的传输机制：重要度高的语义信息使用比特保护能力强的传输机制，重要度低的语义信息使用比特保护能力弱的传输机制。

1.2 经典信息论与语义信息论

信息论是建立在概率论的基础上的，香农在文献[5]中引入了信息熵的概念，以比特为单位衡量信息量，为通信过程建立数学模型，这奠定了通信的理论基础。对于特定的传输任务，如何衡量语义信息量，目前理论研究尚不完备。

1）信息熵与语义熵

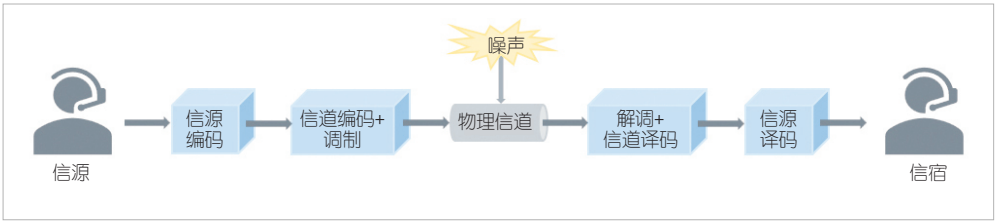
经典信息论用信息熵$H(X)$表示无损传输的信源压缩上限。对于给定的信源X，其信息熵可以表示为：

$$H(X)=-\sum_{x\in X}p(x)\log p(x)。\tag{1}$$

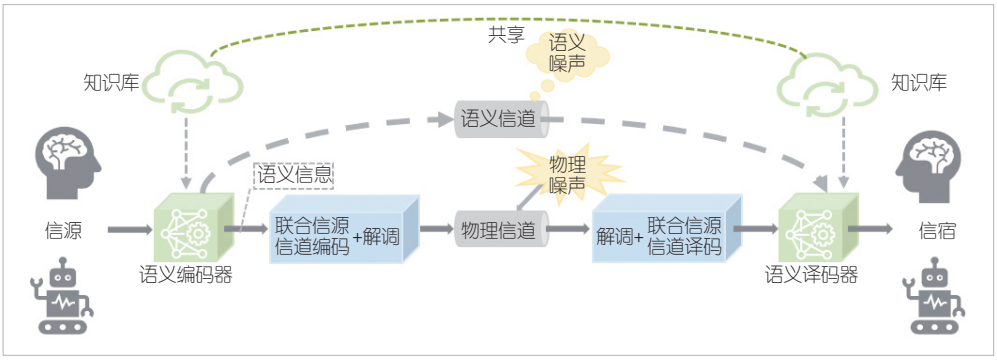
信息熵的计算只考虑原始数据，而语义熵的计算还需要考虑语义。例如文献[6]从逻辑概率角度出发，将香农的通信理论从语法层面扩展到语义层面，定义了句子x的语义熵：

$$H(x)=-\log_2\left(P(x)\right),\tag{2}$$

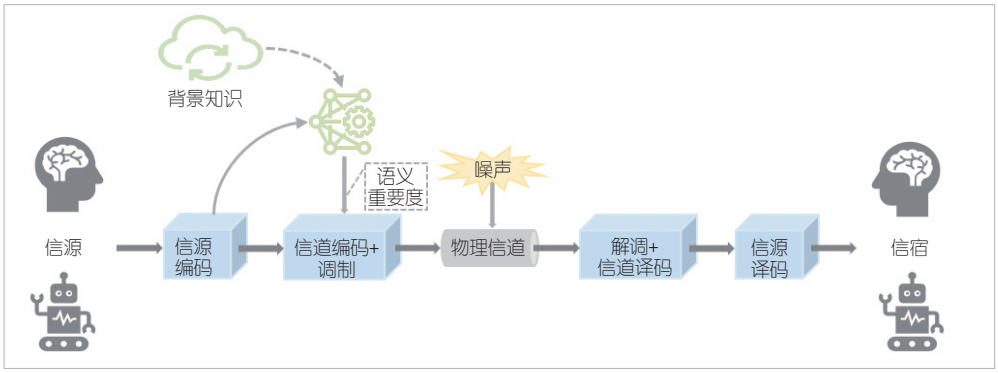
其中，$P(x)$是句子的逻辑概率。然而，这种定义存在逻辑



▲图1 传统通信系统模型



▲图2 信源信道联合设计的语义通信系统模型



▲图3 信源信道分离设计的语义通信系统模型

悖论问题，且不具有普适性。从本质上讲，语义信息只是原始数据有损压缩后的信息。对此，我们先定义语义失真测度，然后依据失真测度推导率失真定理，再根据信息熵是无失真编码码率的下界，将语义熵描述为语义失真最小的编码码率的下界。这种定义方式具有普适性且与传统信息论逻辑自洽，有利于理论及其支撑下的编码技术的平滑演进。

2） 传统信道容量与语义信道容量

信道传输的信息量可以由信道输入 X 和信道输出 Y 间的互信息表示，互信息表达式为：

$$I(X;Y)=H(X)-H(X|Y), \tag{3}$$

其中，条件熵 $H(X|Y)$ 表示信道传输中信息量的损失。高斯信道容量取上述互信息的最大值，计算公式为：

$$C=B\log\left(1+\frac{P}{N_0B}\right), \tag{4}$$

其中，B 表示系统带宽，P 表示输入信号平均功率限制，N_0 代表噪声功率的谱密度。传统通信中信道容量的大小与传输的内容无关，而语义信道容量既与通信系统本身有关，也与传输任务有关。文献[6] 还开发了语义信道容量的定理。假设 X 代表信道输入，Z 是从中提取的语义特征，Y 为信道的输出，离散无记忆信道的语义信道容量可以表示为：

$$C_s=\sup_{p(X|Z)}\left\{I(X;Y)-H(X|Z)+\overline{H_S(Y)}\right\}, \tag{5}$$

其中，$I(X;Y)$ 表示语法信道的性能，$C=\sup\left\{I(X;Y)\right\}$。$H(X|Z)$ 代表语义编码后的歧义，表示语义编码器的性能。$\overline{H_S(Y)}$ 代表接收消息的平均逻辑信息，其值越大表示接收者对接收到的消息的解释能力越强。在语法信道容量的基础上，语义信道容量的大小取决于 $H(X|Z)$ 和 $\overline{H_S(Y)}$。

3） 率失真与语义率失真

率失真理论也叫有损信源编码定理。对于给定的最大平均失真 D^*，率失真函数 $R(D^*)$ 给出传输比特率的下限[7]：

$$R(D^*)=\min_{D\leqslant D^*}I(X;Y), \tag{6}$$

其中，失真 D 与 X 和 Y 的距离有关。不考虑外部信息，语义率失真要同时考虑传输的失真和语义层面的失真。文献[8] 中给出了一种计算公式：

$$R(D_s,D_a)=\min I\left(Z;\hat{X},\hat{Z}\right), \tag{7}$$

其中，D_s 是信源 X 和恢复信息 $\hat{X}$ 之间的语义失真，D_a 是信道噪声引起的语义表征 Z 和接收端接收到的语义表征 $\hat{Z}$ 之间的失真。

一些研究人员引入信息瓶颈理论来进行率失真权衡。文献[9] 设计了一种新的损失函数：

$$L=\underbrace{I(Z;X)}_{\text{压缩}}-\underbrace{(1+\alpha)I\left(Z;\hat{Z}\right)}_{\text{互信息}}+\underbrace{\beta KL\left(X,\hat{Z}\right)}_{\text{推理}}, \tag{8}$$

其中，α 和 β 是调节互信息项和推理项的权重的参数。压缩项表示 X 所需的平均位数。推理项是编码器处的 X 和接收器捕获的 $\hat{Z}$ 之间的 KL 散度，表示接收端推理性能。在文献[10] 中，作者利用信息瓶颈公式化了边缘推理系统的率失真权衡，损失函数如下：

$$\mathcal{L}_{IB}(\phi)=\underbrace{-I\left(\hat{Z},V\right)}_{\text{失真}}+\beta\underbrace{I\left(\hat{Z},X\right)}_{\text{压缩}}, \tag{9}$$

其中，V 是推理任务，失真项表示推理结果的不确定性，压缩项表示给定 X 条件下的 $\hat{Z}$ 中保留的信息。

2 语义通信系统架构

语义通信系统的架构有信源信道联合设计和信源信道分离设计两种。语义通信的研究工作主要分为两类：一是以数据重建为目标，如文本图像的高质量传输；另一个是以执行任务为目标，如图像的分类、识别、分割和视觉问答等。现有工作主要集中于联合信源信道编码的语义通信系统。本章节按照数据重建导向和任务执行导向的分类方式对相关研究进行总结，并提出一种信源信道分离设计的方案。

2.1 信源信道联合设计的语义通信

1） 面向数据重建的语义通信

通信系统传输的数据类型主要有图像、文本和语音 3 种。面向数据重建的通信任务目标是在接收端恢复信源数据。对于此类任务，语义通信系统借助先验背景知识库，仅对提取的语义信息进行编码传输，而不是传输全部原始数据。大量仿真实验表明，语义通信系统具有比传统通信系统更高的鲁棒性和更低的通信开销。

文献[11] 提出了传输高分辨率图像的方案——基于深度学习的联合信源信道编码（Deep JSCC），将图像像素值直接映射为复值信道输入符号。编码器和解码器通过两个卷积神经网络进行联合训练。该方案分别在 CIFAR-10 和 Kodak 图像数据集中进行训练和测试。实验显示，在峰值信噪比和结构相似度方面，Deep JSCC 方案性能比传统的信源信道分离方案更优，在低信噪比区域的优势尤其显著，且在时变信道上表现出更高的鲁棒性。

文献[12]提出一种用于文本传输的方案，将互信息作为

损失函数的一部分，通过恢复句子的含义不仅使系统容量达到最大，还可以使语义错误变得最少。文献[12]还定义了名为句子相似性的新指标，以描述两个句子在语义信息方面的相似水平。该指标的计算公式如公式（10）所示，其中B_Φ代表BERT（一个巨大的用于提取语义信息的预训练模型）。

$$\mathrm{match}(\hat{s},s)=\frac{B_\Phi(s)\cdot B_\Phi(\hat{s})^T}{\left\|B_\Phi(s)\right\|\left\|B_\Phi(\hat{s})\right\|}。\tag{10}$$

文献[13]设计了用于语音传输的语义通信系统（DeepSC-S），采用了联合信源信道编码的方式，尤其使用了压缩和激励（SE）网络来提取重要语音语义信息，并在电话系统和多媒体传输系统中验证了模型的适应性。结果表明，在两种情况下的信号失真比和语音失真感知评估等语音信号指标都优于传统通信。

2） 面向任务执行的语义通信

在面向任务执行的语义通信系统中，发送端要在知识库的支持下提取充分且尽量少的语义特征，并且忽略无关信息，使接收端能够根据接收到的语义特征有效地执行任务，同时最大程度地减少通信开销，提高资源利用率。面向数据的通信能够满足H2H通信的需求，而H2SM通信和SM2SM通信需要更加智能、高效的面向任务执行的通信。

文献[14]研究了多用户语义通信，并将图像检索、机器翻译和视觉问答（VQA）3个智能任务作为传输目标，基于Transformer提出了不同的模型。用于VQA任务的系统模型在结合图像检索和机器翻译外，还加入了一种新的用于图像-文本信息融合的语义解码器网络。

文献[10]和[15]设计了一种用于边缘推理的语义通信系统，将特征提取器和联合信源信道编码器布置在边缘设备，利用信息瓶颈原理提取推理任务的相关特征，以低延迟进行准确边缘推理。对于动态信道条件下的图像分类任务，文献[10]提出了一种可变长度变分特征编码的端到端架构，可根据信道条件调整特征编码的激活维度数。该架构在延迟和准确率方面的性能良好。为弥补该设备感知的不足，文献[15]提出了用于多设备协作边缘计算的语义通信方案（VDDIB-SR），并开发了一种选择性重传机制，以识别多个边缘设备的编码特征中的冗余。在多视图图像分类和多视图对象识别任务上进行的实验表明，VDDIB-SR与现有方法相比，实现了更高的分类和识别精度，同时降低了通信开销和延迟。

2.2 信源信道分离设计的语义通信

当传输任务或通信环境改变时，信源信道联合设计的语义通信系统需要对整个网络进行重新训练。这是一个非常复杂且耗时的过程。此外，信道中直接传输的语义特征容易受到攻击，存在隐私/秘密泄露等风险。现有的语法通信是基于信源信道分离定理设计的，且分离设计的通信技术已经非常成熟，这些都有助于语义通信系统信源信道的分离设计。

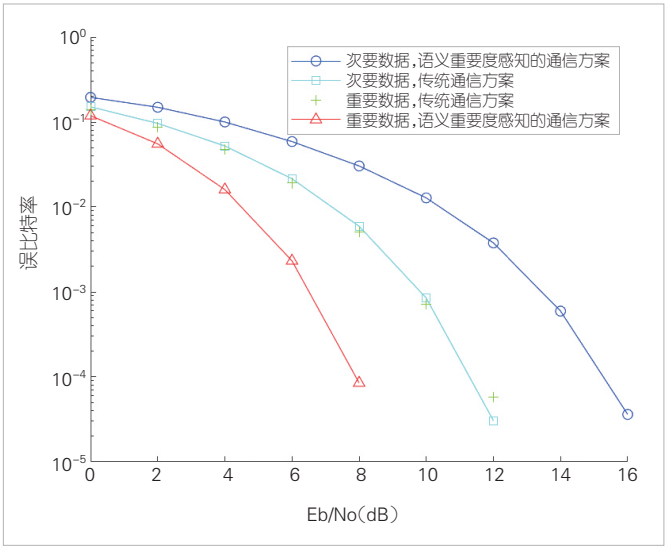
文献[16]提出了一种语义通信信号成形（SSSC）方法，以最小化语义损失。语义损失通过来自BERT模型的预训练双向编码器测量。该方法将信号集优化问题转化为受功率约束的矢量优化，能够合理设计星座图。因为考虑了每个信号点的语义，所以文中设计的信号星座是不规则的。语义相似度越高的信号星座点越接近，留出的空间就越多，有利于放置其他信号星座点。这里我们将所提出的SSSC与二进制相移键控和正交相移键控等设计相比较。仿真结果表明，SSSC在减少语义损失方面能够带来显著性能增益。

本文提出了一种基于语义重要度构造不等错误传输机制的方案。基于背景知识，该方案在信源编码时对语义按重要度进行分类，根据重要度设置不同的误码率限制并采用不同的传输方案（包括信道编码等），为重要度高的语义信息构建强保护传输的传输机制，为重要度低的语义信息设计弱保护传输机制，用较少的通信资源实现良好的通信性能。

系统模型如公式（11）所示。对于要传输的数据X，根据背景知识将其语义重要度从低到高划分为n类，原始数据X也分为n组。对应地，传输机制也分为n个等级，保护性能的强弱与等级高低一致。

$$X\rightarrow\left\{\begin{array}{l}X_1\rightarrow\widehat{X_1}\\X_2\rightarrow\widehat{X_2}\\\cdots\rightarrow\cdots\\X_n\rightarrow\widehat{X_n}\end{array}\right\}\rightarrow\hat{X}\quad。\tag{11}$$

我们在AWGN信道下用Mnist数据集进行仿真实验，使用神经网络识别图片上的数字。不失一般性，我们假设语义重要度与数值大小一致（9对应的重要度最高，0对应的重要度最低），并与不区分语义重要度的传统通信方法进行对比。图4给出了在实现相同的误比特率（BER）条件下，信源信道分离的语义通信系统与传统通信系统BER性能对比。仿真中，重要数据对应数字9，共1 000张图片，采用（7，4）汉明码+4符号正交幅度调制（4 QAM）;次要数据对应数字0，共1 000张图片，采用（7，4）汉明码+64 QAM。为了公平对比，我们采用传统通信方案传输这2 000张图片，不进行重要度区分，均采用（7，4）汉明码+16 QAM调制。仿真结果显示，在BER=10^{-3}时，我们所提的语义重要度感知的通信方法比传统通信方案节省了约3 dB的功率。这一结果表明，区分语义重要度能够更好地保障重要数据的传输，



▲图4 信源信道分离设计的语义通信与传统通信方案的性能对比

从而节约通信资源。

3 语义通信的开放问题

1）近年来，深度学习取得了巨大的进步，其在语义通信中的应用弥补了概率统计等理论基础的不足，使语义通信再次成为研究热点。但深度学习的可解释性差，语义信息度量、语义传输容量等基础理论问题仍未突破，语义通信的研究仍然缺乏一种通用的科学理论指导。

2）现有语法通信是基于信源信道分离定理而设计的，而目前语义通信系统的研究几乎全部采用信源信道联合设计的方法。通信网络中任何一个节点都有可能遭受攻击，而语义特征的提取是基于背景知识库的。按照理想的设计思路，在背景知识未知的情况下，系统中传输的信息是无意义的。也就是说，语义特征提取与语义编码的过程为语义通信提供了天然的加密层。实际上这种加密是否有效还有待考证。

4 结束语

传统的语法通信已经逼近香农定理的极限，在原有思路上提升性能只能依靠技术的堆叠，而这也带来了系统复杂度和能耗的提高。语义通信提供了一条新的发展思路，旨在构建“智能、简约、达意”的通信系统，且其在理论和技术方面都具有光明的发展前景和广阔的研究空间。

参考文献

[1] CALVANESE STRINATI E, BARBAROSSA S. 6G networks: beyond Shannon towards semantic and goal-oriented communications [J]. Computer networks, 2021, 190: 107930. DOI: 10.1016/j.comnet.2021.107930
[2] 牛凯, 戴金晟, 张平, 等. 面向6G的语义通信 [J]. 移动通信, 2021, 45(4): 85-90. DOI: 10.3969/j.issn.1006-1010.2021.04.014
[3] LUO X W, CHEN H H, GUO Q. Semantic communications: overview, open issues, and future research directions [J]. IEEE wireless communications, 2022, 29(1): 210-219. DOI: 10.1109/MWC.101.2100269
[4] 石光明, 肖泳, 李莹玉, 等. 面向万物智联的语义通信网络 [J]. 物联网学报, 2021, 5(2): 26-36. DOI: 10.11959/j.issn.2096-3750.2021.00209
[5] SHANNON C E, WEAVER W. The mathematical theory of communication [M]. Urbana: University of Illinois Press, 1949
[6] BAO J, BASU P, DEAN M K, et al. Towards a theory of semantic communication [C]//Proceedings of 2011 IEEE Network Science Workshop. IEEE, 2011: 110-117. DOI: 10.1109/NSW.2011.6004632
[7] QIN Z J, TAO X M, LU J H, et al. Semantic communications: principles and challenges [EB/OL]. [2023-02-16]. https://arxiv.org/abs/2201.01389
[8] LIU J K, ZHANG W Y, POOR H V. A rate-distortion framework for characterizing semantic information [C]//Proceedings of 2021 IEEE International Symposium on Information Theory (ISIT). IEEE, 2021: 2894-2899. DOI: 10.1109/ISIT45174.2021.9518240
[9] SANA M, STRINATI E C. Learning semantics: an opportunity for effective 6G communications [C]//Proceedings of 2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC). IEEE, 2022: 631-636. DOI: 10.1109/CCNC49033.2022.9700645
[10] SHAO J W, MAO Y Y, ZHANG J. Learning task-oriented communication for edge inference: an information bottleneck approach [J]. IEEE journal on selected areas in communications, 2022, 40(1): 197-211. DOI: 10.1109/JSAC.2021.3126087
[11] 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
[12] 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
[13] PARK S, SIMEONE O, KANG J. End-to-end fast training of communication links without a channel model via online meta-learning [C]//Proceedings of 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2020: 1-5. DOI: 10.1109/SPAWC48557.2020.9154322
[14] XIE H Q, QIN Z J, TAO X M, et al. Task-oriented multi-user semantic communications [J]. IEEE journal on selected areas in communications, 2022, 40(9): 2584-2597. DOI: 10.1109/JSAC.2022.3191326
[15] SHAO J W, MAO Y Y, ZHANG J. Task-oriented communication for multidevice cooperative edge inference [J]. IEEE transactions on wireless communications, 2023, 22(1): 73-87. DOI: 10.1109/TWC.2022.3191118
[16] GUO S S, WANG Y H, ZHANGZ P. Signal shaping for semantic communication systems with A few message candidates [C]//Proceedings of 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall). IEEE, 2023: 1-5. DOI: 10.1109/VTC2022-Fall57202.2022.10012981

作 者 简 介

郭帅帅，山东大学教授；主要研究领域为通信系统性能分析、优化设计与应用创新，智能通信与网络，通信感知一体化设计等；主持和参与多项国家级和省部级项目；发表论文50余篇。

李树静，山东大学在读硕士研究生；主要研究方向为语义通信。

2022年度十大天线技术进展

Top Ten Antenna Technology Advances in 2022

张跃平/ZHANG Yueping
（南洋理工大学，新加坡 639798）
(Nanyang Technological University, Singapore 639798, Singapore)

DOI：10.12142/ZTETJ.202302013
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1321.006.html
网络出版日期：2023-04-11
收稿日期：2023-01-16

摘要：“十大天线技术进展”旨在遴选、记录天线领域的年度标志性成果。在广泛征集中国天线产学研界多位专家的意见后，整理出了2022年度十大天线技术进展，其中有九大天线技术来自中国。分别介绍了这些天线技术的技术特点和优势。认为在天线开发应用方面，中国后来居上。指出天线强则国家强，国家强则天线强。

关键词：无线输能天线；大规模多输入多输出天线；太空卫星阵列天线；滤波天线；封装天线

Abstract: The "Top Ten Antenna Technology Advances" aims to select and record annual landmark achievements in the antenna field. After extensively soliciting the opinions of multiple experts from the antenna industry, university, and research community in China, the top ten antenna technology advances in 2022 are selected, including nine antenna technologies from China. The technical characteristics and advantages of these technologies are analyzed separately. It is believed that in terms of antenna development and application, China has come from behind. It is pointed out that if the antenna technology is strong, the country will be strong, and if the country is strong, the antenna technology will be strong.

Keywords: wireless power transmission antenna; massive multi-input multi-output antenna; space satellite array antenna; filter antenna; encapsulated antenna

天线随处可见，与人们的生活息息相关。它镶嵌在手机内，提供全新的、高品质用户体验；它安装在汽车上，为安全保驾护航。

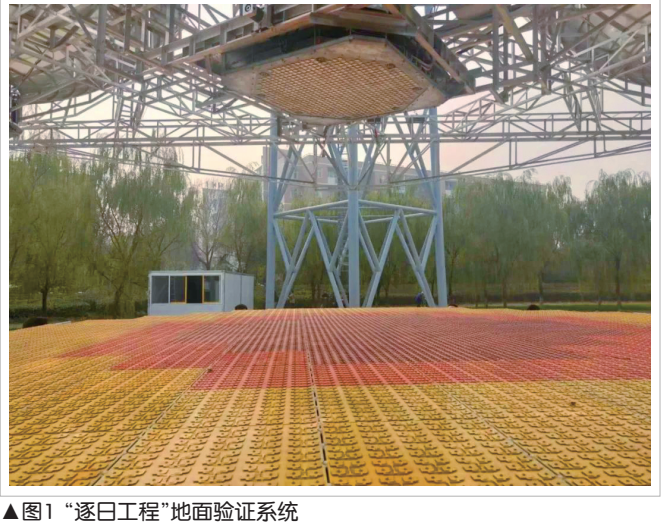
天线也会架设在丛山峻岭中，仰望星空，探索着宇宙的起源。天线更会遨游在太空中，俯瞰浩瀚的海洋与苍茫大地，用电波“激励”着你我他。

天线的应用层出不穷。新应用、新需求为天线研发者带来新的挑战。每当挑战得到克服，天线技术就会取得进步。

基于公开发表的信息，在广泛征集中国天线产学研界多位专家的意见后，本文整理出2022年度十大天线技术进展，供大家参考。

1 “逐日工程”微波发射天线、接收与整流天线

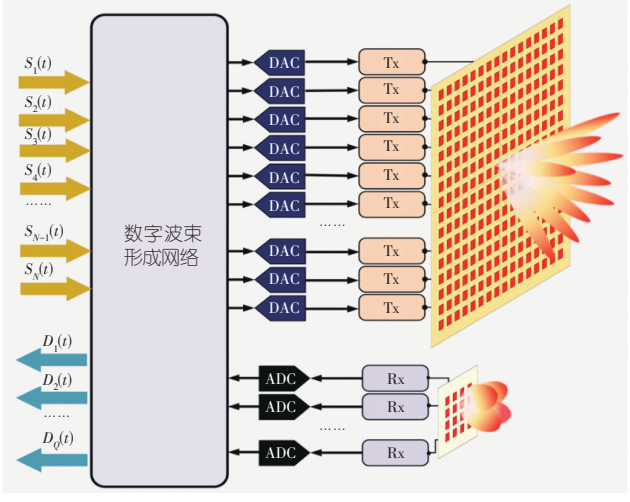
由西安电子科技大学段宝岩院士牵头研制的全链路、全系统空间太阳能电站地面验证系统，于2022年6月15日通过验收，如图1所示。该验证系统主要包括五大子系统：欧米伽聚光与光电转换、电力传输与管理、微波发射天线、接收与整流天线、控制与测量。其工作原理是：首先根据太阳高度角确定聚光镜需要倾斜的角度，在接收到聚光镜反射的太阳光后，位于聚光镜中心的光伏电池阵，将其转化为直流电能。随后，通过电源管理模块，4个聚光系统转换得到的电能汇聚到中间发射天线。经过振荡器和放大器等模块，电能被进一步转化为微波。微波以无线传输的形式被发射到接收天线。最后，接收天线将微波整流，再次转换成直流电，供给负载。



▲图1 “逐日工程”地面验证系统

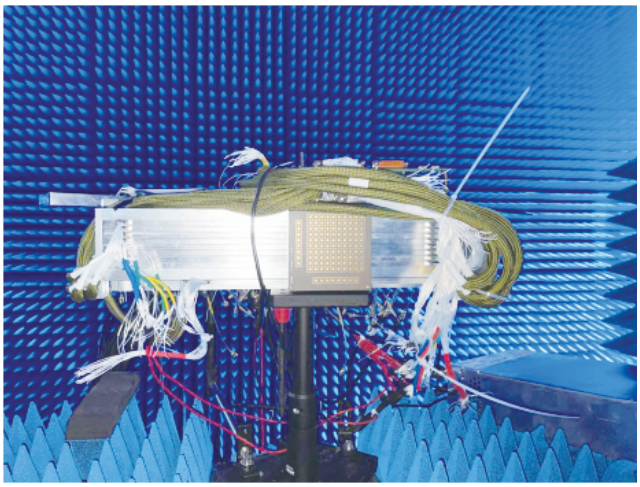
2 非对称全数字大规模MIMO有源阵列天线

众所周知，1G、2G移动通信基站采用的是扇形波束单天线技术，即单输入单输出（SISO）架构。3G、4G开始采用多天线技术，即多输入多输出（MIMO）架构。而5G则采用基于混合多波束有源阵列天线的大规模MIMO技术，在频分、时分、码分的基础上继续挖掘空间资源，以有效提升系统容量。大规模MIMO技术的核心是多波束有源阵列天线。大规模全数字多波束阵列天线架构是最佳方案，但面临系统复杂度、功耗、成本、海量数据实时处理等一系列瓶颈问题。因此，5G退而求其次采用了基于相控子阵的混合多波束方案。为了进一步大幅提升系统容量，同时解决上述瓶颈问题，面向6G，东南大学洪伟教授团队于2019年提出了“非对称全数字大规模MIMO阵列天线”的概念，并获得国家重点研发计划项目“非对称毫米波亚毫米波大规模MIMO关键技术研究及系统验证”（2020YFB1804900）的支持，如图2所示。3年来，非对称大规模MIMO阵列系统架构、天线阵列最佳拓扑结构、非互易信道特性与建模、阵列通道的校准与自校准、通道非线性行为建模与线性化，以及多通道收发芯片和系统实验验证等均取得重要进展。研究者们研制了国际上首套毫米波非对称64T16R全数字大规模MIMO阵列原理样机，如图3所示。该样机常被作为对比目标。此外，研究者们还研制了国际上首套毫米波全数字64T64R大规模MIMO阵列样机，如图4所示。该样机支持20个波束/数据流，总数据吞吐率达到50 Gbit/s。图5展示了支持全数字多波束阵列的毫米波4通道发射/接收芯片。相关研究论文已发表在《IEEE Transactions on Antennas and Propagation》《IEEE Transactions on Microwave Theory and Techniques》《IEEE Journal of Solid-State Circuits》《中国科学：信息科学》等权威期刊上，其中有2篇入选ESI高被引论文，1篇获IEEE MTT-S微波奖。



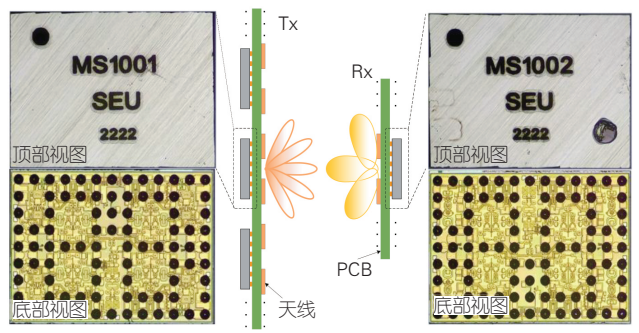
ADC：模数转换　DAC：数模转换　Rx：接收器　Tx：发射器

▲图2 非对称全数字大规模多输入多输出阵列天线架构



▲图3 64T16R非对称全数字大规模多输入多输出阵列天线

▲图4 64T16R非对称全数字多波束阵列天线原理样机与动态性能测试



PCB：印刷电路板　Rx：接受器　SEU：东南大学　Tx：发射器

▲图5 支持全数字多波束阵列的毫米波4通道发射/接收芯片

3 蓝色行者3商用移动通信阵列

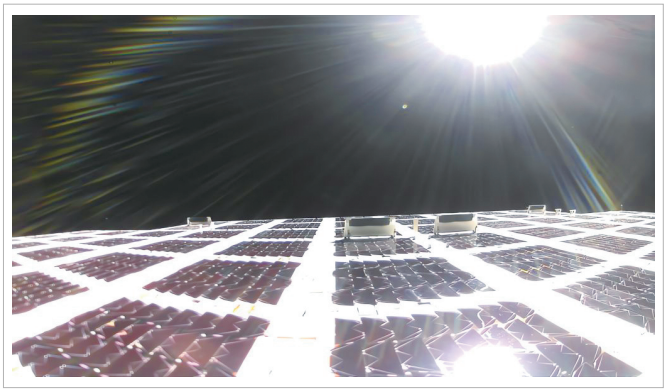
美国AST太空蓝色行者3是部署在近地轨道上的最大商业通信阵列，旨在通过第3代合作伙伴计划（3GPP）标准频率以5G速度直接与蜂窝设备通信。卫星阵列天线已经于2022年11月14日在太空展开（如图6所示），面积为64 m^2。这一设计对于支持天基蜂窝宽带网络至关重要。该卫星预计在地球表面拥有超过776 996 km^2的视野。

4 大规模分布孔径深空探测天线

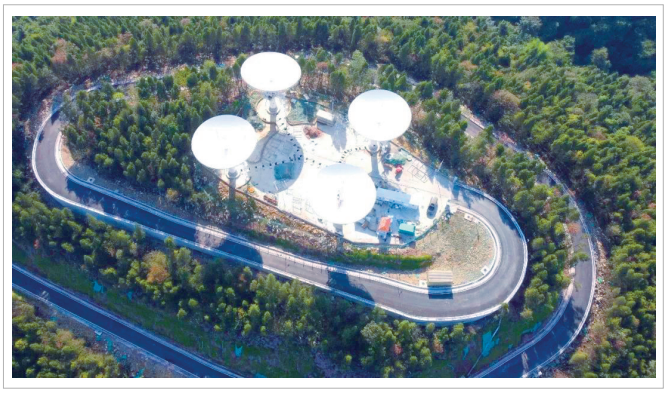
近日，由北京理工大学牵头建设的“中国复眼”成功开机观测。与世界著名的“中国天眼”不同，它的大天线由很多小天线合成，就像昆虫的眼睛一样，因此得名“中国复眼”，意为“中国复兴之眼”。整个系统的建设包括3各阶段，其中，第1阶段工程建设已完成（如图7所示），第2阶段工程建设已经启动。建成后“中国复眼”可实现超远程探测，探测距离可达1.5亿千米。

5 跨频段共口径大规模天线

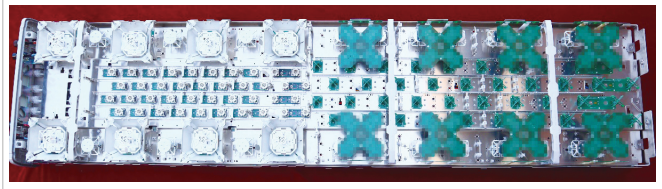
华南理工大学章秀银教授课题组对跨频段共口径大规模天线阵列关键技术进行了研究与验证，如图8所示。在异频共口径天线阵列中，不同频段阵子密集分布导致互耦严重、端口隔离差。此外，处在上方的低频天线单元会遮挡下方的高频天线单元，导致高频辐射方向图产生畸变。针对该问题，课题组提出了基于集成滤波天线的隔离度提升技术，利用集成滤波天线的带外抑制特性来降低不同频段的互耦，提高隔离度。此外，课题组还提出了基于空间滤波天线的方向图畸变改善技术，设计了具有空间滤波特性的低频天线，使其下方的高频天线辐射的电磁波可以透过低频天线，从而改善高频天线的辐射方向图。基于滤波天线技术设计系列多频共口径基站阵列天线，实现了3 GHz/4 GHz/5 GHz不同频段天线的一体化集成。



▲图6 太空蓝色行者64 m^2阵列天线



▲图7 大规模分布孔径深空探测雷达一期工程天线



▲图8 跨频段共口径大规模天线阵列

6 毫米波片上天线测量系统

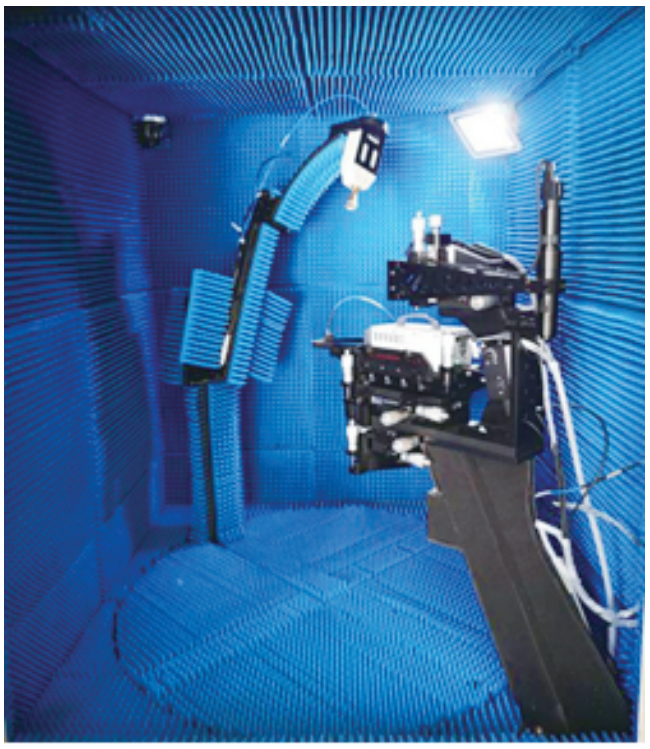
电科思仪毫米波片上天线测量系统的频率覆盖范围为8～110 GHz（可扩展至500 GHz）。该系统具有频率覆盖范围宽、动态范围大、馈电形式灵活、结构紧凑、测试参数全面等优点。系统采用通用化、模块化设计思想，以电科思仪高性能矢量网络分析仪为核心，由频率扩展模块、高精度转台、片上天线馈电平台、微波暗箱等单元组成，如图9所示。该系统具有方向图、副瓣电平、增益、轴比、驻波等多种参数测量功能，可用于片上天线、集成天线、封装天线等非常规馈电形式的微型天线电性能参数测试分析。

7 自动化微型相控阵天线测试系统

成都天锐星通科技有限公司自动化微型相控阵天线测试系统，频率覆盖范围为10～110 GHz，占地面积不足5 m^2，如图10所示。系统采用多功能实时控制模块、同步信号调理模块和宽带双极化馈源，实现了0.5 m口径相控阵天线的全指标一键高效测量。结合大数据分析技术，该系统可快速实现产品的基本性能分析以及一致性和相关性分析。同时产品配备了自动装卸和进出料模块，自动化程度高，易于产线部署，可用于相控阵天线的大规模生产。

8 大规模阵列的封装天线技术

未来移动通信将通过极致连接为人类社会提供全无线、



▲图9 毫米波片上天线测量系统

▲图10 自动化微型相控阵天线测试系统

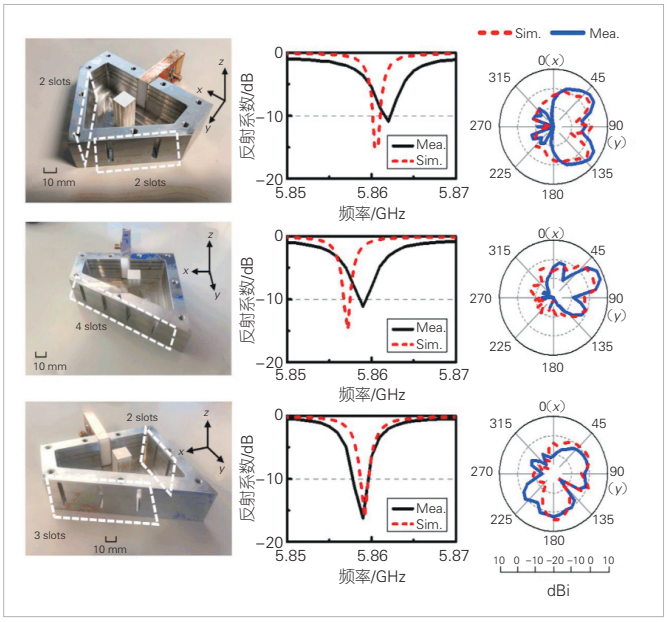
以人为中心的沉浸式体验，真正开启智能时代。华为研究团队在毫米波70 GHz频段，成功研制了超低功耗、超高吞吐、超低时延的短距通信原型样机，实现了超过10 Gbit/s的吞吐率和亚毫秒级的时延，并实时演示了4K虚拟现实（VR）业务，如图11所示。此样机具有突出的短距传输性能。传输

▲图11 华为70 GHz频段短距通信原型样机与封装天线

速率是通用串行总线（USB）等有线通信方式的数倍，且整机功耗低于560 mW。该原型样机采用了高效锗硅大规模阵列。不规则阵列天线可封装在手机规格的低温共烧陶瓷模块中，使微型的封装天线也可以应用在可穿戴设备中。

9 近零介电常数媒质构造几何无关天线

清华大学电子工程系李越副教授课题组使用近零介电常数（ENZ）媒质构造几何无关的新型天线，使天线的工作频率不再与天线的几何形状有关，实现了天线空间辐射特性和工作频率的独立设计，如图12所示。该课题组于2022年6月22日将相关成果发表在《Nature Communications》上[1]。该成果为天线设计提供了全新的灵活性，在柔性电子、智能感知与无线通信等领域具有潜在应用价值。



▲图12 几何无关天线[1]

10 太赫兹16元环形耦合振荡器-辐射器阵列

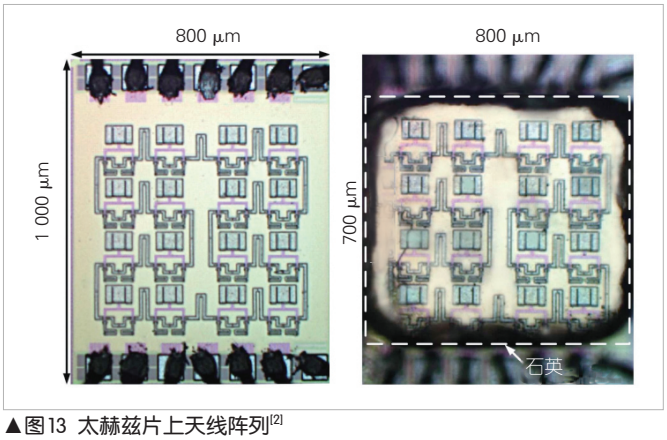
香港城市大学太赫兹及毫米波国家重点实验室陈志豪教授团队使用65 nm的互补金属氧化物半导体（CMOS）在0.8 mm^2的总面积上实现了16元环形耦合振荡器-辐射器阵列，如图13所示。最大辐射功率-2.8 dBm是在472 GHz的条件下测得的。这种设计实现了目前最高的直流太赫兹转换效率和频率调谐范围分别为0.12%和4.2%。

11 结束语

天线作为“器”的发明与发展，先后经历了创世纪阶段、大跃进阶段、添新丁阶段。然而，中国在这些阶段基本上是缺席者。

在天线作为“术”与“用”的发展中，中国后来居上。这可以从中国天线产业居世界首位、学界发表论文全球第一得到佐证。

2022年度十大天线技术进展中，有九大天线技术是由中国天线科技工作者完成的。我们不难得出这样的结论：天线强则国家强，国家强则天线强！



▲图13 太赫兹片上天线阵列[2]

参考文献

[1] LI H, ZHOU Z H, HE Y J, et al. Geometry-independent antenna based on Epsilon-near-zero medium [J]. Nature communications, 2022, 13(1): 1-8. DOI: 10.1038/s41467-022-31013-z

[2] GAO L, CHAN C H. A 0.47 THz ring scalable coupled oscillator - radiator array with miniature patch antennas [J]. IEEE transactions on microwave theory and techniques, 2022, 70(8): 3964-3974. DOI: 10.1109/TMTT.2022.3186329

作 者 简 介

张跃平，无线电电子学家、新加坡南洋理工大学教授、2010级IEEE Fellow、2012年度IEEE天线与传播学会谢昆诺夫论文奖获得者、2020年度克劳斯天线奖获得者、2022年度IEEE电子封装学会杰出技术成就奖获得者，曾任香港大学客座教授、太原理工大学教授、IEEE天线与传播学会杰出讲师、IEEE天线与传播汇刊副编辑；曾多次在IEEE（美国）、CIE（中国）、IET（英国）、EurAAP（欧洲）、IEICE（日本）、KIEES（韩国）组织的旗舰会议上发表演讲；2020年获欧洲天线与传播学会大会组委会赠送的、为纪念奥斯特发现电磁现象的牌匾，2021年获香港中文大学电子工程系建系50周年杰出校友奖，曾多次获国际学术会议最佳论文奖；发表论文300余篇，拥有7项美国发明专利。

通感一体化关键技术与应用

Key Technologies and Application of ISAC

李萍/LI Ping，郭晓江/GUO Xiaojiang
（中兴通讯股份有限公司，中国 深圳 518057）
（ZTE Corporation，Shenzhen 518057，China）

DOI：10.12142/ZTETJ.202302014
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1401.012.html
网络出版日期：2023-04-11
收稿日期：2023-01-16

摘要：通信网络在感知领域具有巨大的应用空间。指出通信系统用于感知的能力边界，包括感知距离、感知分辨率、感知精度等关键技术指标。基于正交频分复用（OFDM）波形的感知系统模型，提出一种利用匹配滤波进行感知信号参数估计的方法。提出的高分辨率感知信号参数估计方法，不仅能够解决感知过程中角度估计的低分辨率问题，还能大大降低计算复杂度。认为在所能获得的感知信息和服务基础之上，通感一体化（ISAC）系统将更好地服务于未来智慧低空、智慧交通、智慧生活等。

关键词：ISAC；感知性能；参数估计；高分辨率

Abstract: Communication networks have great application space in the field of sensing. The capability boundaries of sensing based on communication systems are put forward, including key technical indicators such as sensing distance, sensing resolution, and sensingaccuracy. Based on the sensing system model of orthogonal frequency-division multiplexing (OFDM) waveforms, a parameters estimation method based on matched filters is proposed. The proposed parameters estimation method with high resolution only solves the low resolution problem of angle estimation in the sensing process, but also greatly reduces computational complexity of angle estimation. It is believed that on the basis of available perceptual information and services, the integrated sensing and communications (ISAC) system will better serve intelligent low altitude, intelligent transportation, intelligent life, etc. in the future.

Keywords: ISAC; sensing performance; parameter estimation; high resolution

通信系统和感知系统一直以来被作为两个独立的系统进行研究，但是这两个系统又有很多相似之处。通信感知一体化通过空口及协议联合设计、频谱资源共享、软硬件设备共享等手段，实现通信、感知、计算功能的一体化设计，达到通信功能与感知功能的融合共生。通信感知一体化在进行信息传递的同时，还能通过分析无线电波的直射、反射、散射等特性，对目标或环境信息等进行定位、测距、测速、成像、检测、识别、环境重构等[1]。无线通信感知一体化不仅开辟了全新的业务，还具有超越传统移动通信网络连接的潜力，因此得到了产业界的广泛关注。学术界和工业界普遍认为，通感一体化将成为5G-A和6G无线通信系统最核心的特性[2-5]。

通信系统主要采用正交频分复用（OFDM）波形，而传统雷达感知系统主要采用线性调频（LFM）波形。通信感知一体化系统的目标是实现一体化波形。早期M. ROBERTON等先获得独立的通信波形和雷达波形，再采用叠加的方式实现一体化波形[6]。随后产生了最小频移键控（MSK）和线性调频（LFM）信号结合的MSK-LFM、OFDM和LFM相结合OFDM-LFM、多输入多输出正交频分复用（MIMO-OFDM）等一体化波形设计方法[7-10]。通感一体化采用OFDM波形或者扩展的OFDM波形来实现感知功能的趋势越来越明显。

通信系统一般采用基站A发送-基站B接收的工作模式，而传统的雷达感知系统往往采用基站A发送-基站A接收的工作模式。相比于基站A发送-基站B接收的工作模式，基站A发送-基站A接收可以显著降低组网的难度，减少波束扫描开销。高频段尤其是毫米波和太赫兹等波段对基站A发送-基站A接收模式的需求更加迫切。传统基于基站A发送-基站A接收工作模式的感知方案主要包括两大类：1）连续波体制方案。该方案的优点是没有近距离盲区，缺点是对收发隔离度的要求非常高，一般应用于毫米波及以上波段；2）脉冲体制方案，该方案的优点是可以完全复用通信天线，获得全阵面的增益，缺点是存在近距离盲区，一般应用于X波段及以下波段。通感一体化有源天线单元（AAU）技术往往需要融合空域自干扰消除、射频域自干扰消除、中频域自干扰对消和数字域自干扰消除等先进技术解决方案[10-11]。

针对通信感知一体化信号处理，现有文献主要研究了距离维处理、多普勒维处理、空间维处理等方法，以实现测距、测速、测角等多种功能[12]。L. GABBIEL等针对OFDM

相位编码信号提出了一种新的脉冲压缩方法，该方法基于FFT且等效于匹配滤波方法[13]。DUAN J. Q.提出了针对OFDM信号的几种多普勒处理方法，以便更好地估计目标的径向速度[14]。刘永军等采用子空间投影方法实现对目标的距离与多普勒的超分辨估计[15]。这些一体化信号处理方法大多采用一个或者几个维度进行滤波匹配，忽略了感知系统中各个参数整体性能的提升。经过理论分析，我们认为通感一体化系统的性能主要受限于空间维度，基于通感一体化波形提出多维参数联合和级联的估计方法，最后提出通感一体化高分辨率信号处理方法。所提方法可以有效均衡各个维度参数估计的性能，大大降低计算复杂度。

1 通感一体化关键技术

1.1 感知距离方程

通感一体化系统对目标的感知大致包括如下4个过程：1）发射机发射无线电波；2）无线电波照射到目标后进行反射；3）经过目标的反射回波被接收机接收；4）接收机对回波信号进行处理，检测是否存在目标，并对目标的参数进行估计。参考雷达方程[16-17]，感知系统接收端采用匹配滤波算法处理后的信噪比（SNR）一般可以表示为：

$$\mathrm{SNR}=\frac{P_tG_tG_r\lambda^2\sigma N_{\mathrm{sym}}^{\gamma}\tau}{(4\pi)^3R^4kTFL}, \tag{1}$$

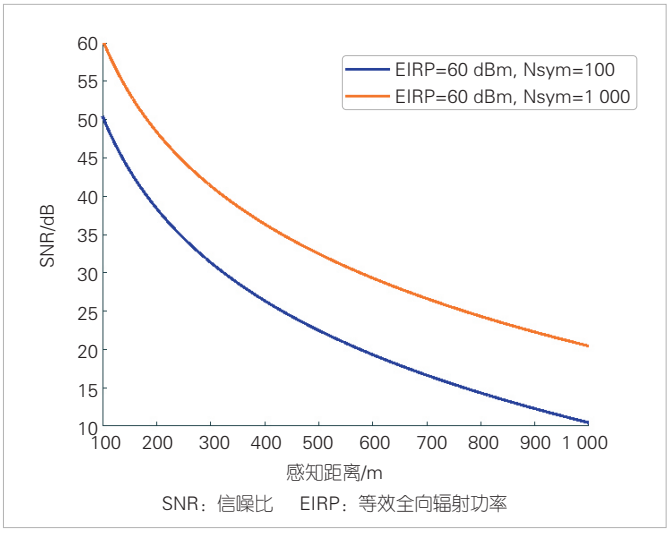
其中，SNR表示经接收机经过处理后目标的信噪比，P_t表示无线信号的平均发射功率，G_t和G_r分别表示发射天线和接收天线的增益，λ表示系统载波波长，σ表示目标的雷达散射截面积（RCS），τ表示一个感知脉冲或者符号的发射信号时长，N_{sym}表示系统波束驻留目标期间接收的感知脉冲或者符号的个数；γ表示雷达接收机进行脉冲或者符号积累的效率（$\gamma\leqslant 1$），R表示目标到天线的距离。k表示玻尔兹曼常数，取1.38×10^{-23} J/K。T表示标准温度，一般可取290 K。F表示系统噪声系数，L表示整个系统的损耗。

对于商用毫米波基站来说，其发射波束的等效全向辐射功率（EIRP）等价于P_tG_t。假设接收天线增益G_r = 24 dB，目标的雷达散射截面积σ = 0.02 m^2，1个用于感知的OFDM符号长度τ = 8.3 μs，26 GHz载波频率对应的载波波长λ = 0.011 5 m，系统的噪声系数和损耗假设为FL = 8 dB。根据雷达方程公式得出的不同感知距离下接收端的SNR如图1所示。可以看出，当EIPR为60 dBm、N_{sym}=100、σ = 0.02 m^{2}时，目标距离基站1 km时接收端处理后的SNR>10 dB。也就是说，现行商用5G毫米波基站对σ = 0.02 m^{2}的目标的感知距离大于1 km。

1.2 感知性能指标

无线感知性能可以从分辨率、精度等方面来描述。感知的分辨率表示可以被系统区分的两个相邻目标的最接近程度，主要包括距离分辨率、角度分辨率和速度分辨率。其中，距离分辨率取决于信号带宽，角度分辨率取决于天线的半功率波束宽度，速度分辨率取决于感知信号相参积累时间。通常情况下，上述3个维度都无法被分辨的情况是很少的。感知的精度表示测量值和真实值的误差，通常用均方根误差来表示，主要包括距离精度、角度精度和速度精度等。感知精度除了和信号带宽、天线波束宽度、信号相参积累时间有关外，还和回波信号的SNR相关。感知分辨率和精度的相关理论值[18]如表1所示。

其中，c表示光速，B表示信号带宽，D表示天线阵列孔径，N_{sym}表示感知信号脉冲或者符号个数，T_r表示感知脉



▲图1 感知距离和接收端SNR关系

▼表1 感知分辨率和精度

参数名称	取值
距离分辨率	$\Delta R=\frac{c}{2B}$
角度分辨率	$\theta_{3\,\mathrm{dB}}\approx\frac{0.886\lambda}{D}$
速度分辨率	$\Delta v=\frac{\lambda}{2N_{\mathrm{sym}}T_r}$
距离精度	$\sigma_R=\frac{c}{2B\sqrt{2SNR}}$
角度精度	$\sigma_R=\frac{\theta_{3\,\mathrm{dB}}}{1.6\sqrt{2SNR}}$
速度精度	$\sigma_R=\frac{\lambda}{2N_{\mathrm{sym}}T_r\sqrt{2SNR}}$

冲或者和符号的重复周期。从表1可以得出：距离分辨率可以通过信号带宽获取，并且随着通信系统带宽的不断增加，距离分辨率已经不是感知的短板；速度分辨率可以通过增加感知信号时长获取，这在感知系统中也是相对容易实现的；而角度分辨率则需要增加阵列孔径长度，通常来说这是不现实的。对于阵元间距为半波长、阵列结构为8 × 8的均匀线阵来说，孔径长度$D = 8 \times \lambda/2 = 4\lambda$，其角度分辨率（3 dB波束宽度）$\theta_{3\,\mathrm{dB}} \approx 12.7^{\circ}$。这样的波束宽度在很多应用场景下难以满足分辨率要求，比如智慧交通场景中距离和速度相同而角度差异较小的行人和车辆的分辨，因此超高分辨的角度估计算法一直是热门研究领域。

1.3 一体化感知算法

在未来5G-A和6G的通感一体化系统中，采用通感一体化基站、一体化OFDM波形进行感知是未来的重要演进方向。本文给出了一种基于OFDM波形的感知系统模型和感知算法处理方案。

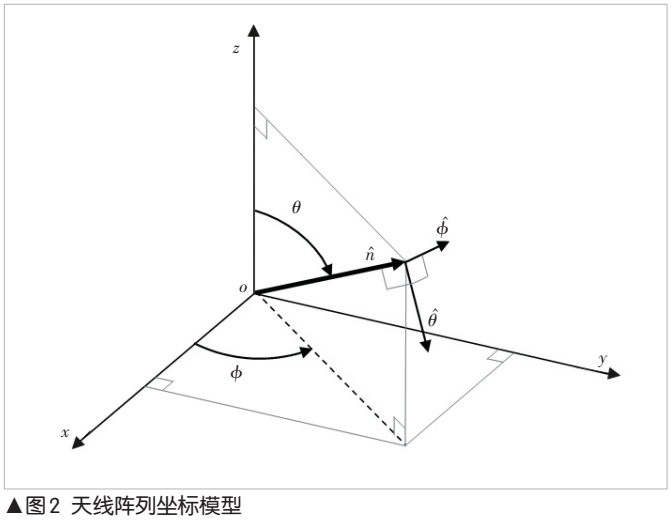
1.3.1 系统模型

假设用于感知的基站（BS）或者用户设备（UE）的天线阵列的坐标模型如图2所示，其中θ为天顶角，φ为方位角。$\theta = 0^{\circ}$表示天顶方向，$\theta = 90^{0}$表示水平方向，该面阵位于yoz平面上。

假设均匀面阵的横向天线阵子数为M，纵向天线阵子数为N，横向和纵向的阵元间距分别为d_H和d_V。那么该阵列的导向矢量可以表示为：

$$\boldsymbol{a}(\theta,\varphi) = \boldsymbol{a}_Z(\theta) \otimes \boldsymbol{a}_Y(\theta,\varphi), \tag{2}$$

其中，



▲图2 天线阵列坐标模型

$$\boldsymbol{a}_Y(\theta,\varphi) = \left[1 \quad e^{j2\pi\frac{d_H \sin\theta\sin\varphi}{\lambda}} \quad \cdots \quad e^{j2\pi\frac{(M-1)d_H \sin\theta\sin\varphi}{\lambda}}\right]^T, \tag{3}$$

$$\boldsymbol{a}_Z(\theta) = \left[1 \quad e^{j2\pi\frac{d_V \cos\theta}{\lambda}} \quad \cdots \quad e^{j2\pi\frac{(N-1)d_V \cos\theta}{\lambda}}\right]^T。 \tag{4}$$

对此，我们可以将$\sin\theta\sin\varphi$视为沿y轴的空间频率，将$\cos\theta$视为沿z轴的空间频率。

对于接收端来说，如果采用OFDM信号用于感知，那么去掉母码后，第(n,m)个天线的第l个OFDM符号第k个子载波的频域信道可以表示为：

$$\mathbf{H}(k,l,n,m) = \sum_{n_{\mathrm{tar}}=0}^{N_{\mathrm{tar}}-1} h_{\mathrm{tar}} \cdot e^{-j2\pi k\Delta f\tau_{\mathrm{tar}}} \cdot e^{j2\pi f_{d,\mathrm{tar}} lT_r} \cdot e^{j2\pi\frac{nd_V\cos\theta_{\mathrm{tar}}}{\lambda}} \cdot e^{j2\pi\frac{md_H\sin\theta_{\mathrm{tar}}\sin\varphi_{\mathrm{tar}}}{\lambda}} + \boldsymbol{n}(k,l,n,m), \tag{5}$$

其中，N_{tar}表示目标个数，h_{tar}是一个和目标散射特性相关的复数，τ_{tar}表示目标时延，T_r表示用于感知的相邻OFDM符号之间的间隔，$\boldsymbol{n}$表示噪声。公式（5）中的相位第1项表示由目标距离产生的时延导致的相位偏移，第2项表示由目标Doppler运动产生的相位偏移，第3项和第4项表示由天线位置产生的相位偏移。

1.3.2 OFDM信号匹配滤波感知算法

假设OFDM信号有K个有效子载波，定义一个和距离（时延）相关的矢量：

$$\boldsymbol{s}_r(\tau) = \left[1 \quad e^{j2\pi k\Delta f\tau} \quad \cdots \quad e^{j2\pi(K-1)\Delta f\tau}\right]^T。 \tag{6}$$

假设采用L个OFDM符号进行感知，定义一个和速度（Doppler）相关的矢量：

$$\boldsymbol{s}_v(f_d) = \left[1 \quad e^{j2\pi f_d T_r} \quad \cdots \quad e^{j2\pi f_d(L-1)T_r}\right]^T。 \tag{7}$$

采用OFDM进行感知时，回波信号的频域$\mathbf{H}(k,l,n,m)$为$K \times L \times N \times M$的矩阵，将其拉直可得$vec(\mathbf{H})$。$vec(\mathbf{H})$是一个列向量，其维度为$KLNM \times 1$。

在高斯白噪声条件下，采用匹配滤波能够获取最佳SNR，从而提高目标的检测能力。采用匹配滤波估计目标的参数$(\tau,f_d,\theta,\varphi)$，具体表达式为：

$$\mathbf{P}_{\text{match filter}}(\tau,f_d,\theta,\varphi) = \sum_{k=0}^{K-1}\sum_{l=0}^{L-1}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left[e^{j2\pi k\Delta f\tau} \cdot e^{-j2\pi f_d lT_r} \cdot e^{-j2\pi\frac{md_H\sin\theta\sin\varphi}{\lambda}} \cdot e^{-j2\pi\frac{nd_V\cos\theta}{\lambda}} \cdot \mathbf{H}(k,l,n,m)\right] = \left[\boldsymbol{a}_Y(\theta,\varphi) \otimes \boldsymbol{a}_Z(\theta) \otimes \boldsymbol{s}_v(f_d) \otimes \boldsymbol{s}_r(\tau)\right]^H \cdot vec(\mathbf{H})。 \tag{8}$$

显然，公式（8）是一个关于$(\tau,f_d,\theta,\varphi)$的四维匹配滤波

过程，搜索次数至少为$KLNM$，共需要$(KLNM)^2$次复乘法和$(KLNM)^2$次复加法，复杂度非常高，不适合工程应用。

在损失一定匹配滤波性能的基础上，将上述$(\tau,f_d,\theta,\varphi)$的联合匹配滤波改为级联匹配滤波，即对每个参数分别搜索，将四维搜索问题转换成4种一维搜索问题，从而大大降低计算复杂度，这也是目前雷达系统通用的处理手段。通感一体化回波信号的感知级联处理流程如图3所示，在三维角度空间中的空间维处理包括方位维处理和俯仰维处理。

距离维搜索需要进行LNM次搜索，单次搜索的式子可以表示为：

$$\mathbf{P}_{\tau}(\tau,l,n,m)=\sum_{k=0}^{K-1}e^{j2\pi k\Delta f\tau}\cdot\mathbf{H}(k,l,n,m)=\boldsymbol{s}_r^H(\tau)\cdot vec\left(\mathbf{H}(:,l,n,m)\right)。\tag{9}$$

速度维搜索需要进行KNM次搜索，单次搜索的式子可以表示为：

$$\mathbf{P}_{\tau,f_d}(\tau,f_d,n,m)=\sum_{l=0}^{L-1}e^{-j2\pi f_d lT_r}\cdot\mathbf{P}_{\tau}(\tau,l,n,m)=\boldsymbol{s}_v^H(f_d)\cdot vec\left(\mathbf{P}_{\tau}(\tau,:,n,m)\right)。\tag{10}$$

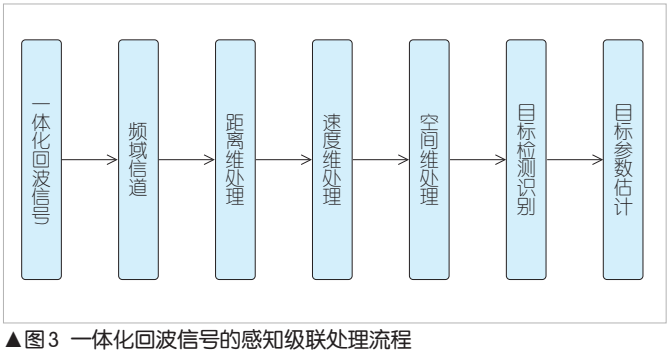
俯仰维搜索需要进行KLM次搜索，单次搜索的式子可以表示为：

$$\mathbf{P}_{\tau,f_d,\theta}(\tau,f_d,\theta,m)=\sum_{n=0}^{N-1}e^{-j2\pi\frac{nd_v\cos\theta}{\lambda}}\cdot\mathbf{P}_{\tau,f_d}(\tau,f_d,n,m)=\boldsymbol{a}_Z^H(\theta)\cdot vec\left(\mathbf{P}_{\tau,f_d}(\tau,f_d,:,m)\right)。\tag{11}$$

方位维搜索需要进行KLN次搜索，单次搜索的式子可以表示为：

$$\mathbf{P}_{\tau,f_d,\theta,\varphi}(\tau,f_d,\theta,\varphi)=\sum_{m=0}^{M-1}e^{-j2\pi\frac{md_H\sin\theta\sin\varphi}{\lambda}}\cdot\mathbf{P}_{\tau,f_d,\theta}(\tau,f_d,n,m)=\boldsymbol{a}_Y^H(\theta,\varphi)\cdot vec\left(\mathbf{P}_{\tau,f_d,\theta}(\tau,f_d,\theta,:)\right)。\tag{12}$$

上述4种一维搜索的检索顺序可以任意调换，不影响最终结果。每个维度搜索的复杂度均为$KLNM$次复乘法和$KLNM$次复加法。4个维度级联搜索后共需要$4KLNM$次复乘法和$4KLNM$次复加法，复杂度降低到原来的$4/(KLNM)$。



▲图3 一体化回波信号的感知级联处理流程

1.3.3 OFDM信号空时频级联感知算法

匹配滤波感知算法实现简单，复杂度相对较低，但是分辨率较低，在很多情况下难以满足应用需求。将超分辨算法引入OFDM信号的感知，可以大大提高系统分辨率。常用的超分辨算法有Burg最大熵算法、Capon最小方差法、多重信号分类（MUSIC）算法[19]、基于旋转不变技术的信号参数估计（ESPRIT）算法[20]、传播算子（PM）算法[21]等。其中，MUSIC算法是超分辨谱估计理论中的一个标志性常用算法，它开启了特征结构类算法的新时代。MUSIC算法对自相关矩阵进行特征分解，以得到相应的信号子空间与噪声子空间，并利用这两个子空间的正交性进行信号参数的估计。

要获得OFDM波形的超分辨感知，需要先计算信道$\mathbf{H}(k,l,n,m)$的协方差矩阵$\mathbf{R}_{HH}$：

$$\mathbf{R}_{HH}=E\left\{vec(\mathbf{H})\cdot\left(vec(\mathbf{H})\right)^H\right\},\tag{13}$$

其中，$E\{\}$表示对多个样本求期望。接着对$\mathbf{R}_{HH}$进行特征分解：

$$\mathbf{R}_{HH}=\mathbf{U}_s\boldsymbol{\Sigma}_s\mathbf{U}_s^H+\mathbf{U}_n\boldsymbol{\Sigma}_n\mathbf{U}_n^H,\tag{14}$$

其中，$\mathbf{U}_s$和$\mathbf{U}_n$分别表示回波信号的信号子空间和噪声子空间。那么用于感知目标参数$(\tau,f_d,\theta,\varphi)$估计的MUSIC谱可以表示为：

$$P_{\mathrm{MUSIC}}(\tau,f_d,\theta,\varphi)=\frac{1}{\boldsymbol{a}^H(\tau,f_d,\theta,\varphi)\mathbf{U}_n\mathbf{U}_n^H\boldsymbol{a}(\tau,f_d,\theta,\varphi)},\tag{15}$$

其中，$\boldsymbol{a}(\tau,f_d,\theta,\varphi)=\boldsymbol{a}_Y(\theta,\varphi)\otimes\boldsymbol{a}_Z(\theta)\otimes\boldsymbol{s}_v(f_d)\otimes\boldsymbol{s}_r(\tau)$。上述MUSIC谱包含了方位–俯仰–距离–速度4个维度，进行特征分解和谱峰搜索的复杂度非常高，难以实现工程应用。为了得出一种工程可实现的算法，将上述联合估计转换成距离–速度–空间的级联处理，即距离维搜索和速度维搜索采用常规匹配滤波算法，方位和俯仰维角度估计采用超分辨算法。这主要是由于：对基站来说，在距离维和速度维获取高分辨率相对容易，而在空间维度获取高分辨率常规算法需要很大的天线孔径。

假设在(τ,f_d)处存在目标，那么采用MUSIC算法进行角度估计时，需要计算的协方差矩阵可以表示为：

$$\mathbf{R}_{\tau,f_d}=vec\left(\mathbf{P}_{\tau,f_d}(\tau,f_d,:,:)\right)\cdot\left(vec\left(\mathbf{P}_{\tau,f_d}(\tau,f_d,:,:)\right)\right)^H。\tag{16}$$

由于进行了距离维-速度维处理后，在(τ,f_d)处的目标变成了单次快拍（往往只有一个样本），且在该点处不同目标的距离和速度相同。也就是说，哪怕是获得多个样本，不同目标之间也是相干的。协方差矩阵$\mathbf{R}_{\tau f_d}$的秩为1。进行特征分解后噪声子空间被信号子空间污染，最终使得MUSIC算法的性能急速恶化，甚至还不如常规匹配滤波算法。为了解决这个问题，空间平滑算法[22]被引入以恢复MUSIC的超分辨能力。空间平滑算法的核心思想是将整个阵列分成若干阵列结构相同的子阵。假设第p个子阵的协方差矩阵为$\mathbf{R}_{p,\tau f_d}$，那么第p个子阵的距离维和速度维匹配滤波后的结果可以表示为：

$$\mathbf{P}_p(\tau,f_d)=vec\left(\mathbf{P}(\tau,f_d,p{:}p+N_{\text{sub}}-1,p{:}p+M_{\text{sub}}-1)\right), \quad (17)$$

其中，$N_{\text{sub}}=N+1-P$，$M_{\text{sub}}=M+1-P$，那么第p个子阵的协方差矩阵$\mathbf{R}_{p,\tau f_d}$可以表示为：

$$\mathbf{R}_{p,\tau,f_d}=\mathbf{P}_p(\tau,f_d)\cdot\mathbf{P}_p^H(\tau,f_d)。 \quad (18)$$

采用前向空间平滑后的协方差矩阵为：

$$\hat{\mathbf{R}}_{\tau,f_d}=\frac{1}{P}\sum_{p=1}^{P}\mathbf{R}_{p,\tau,f_d}。 \quad (19)$$

然后对$\hat{\mathbf{R}}_{\tau f_d}$进行特征分解，获得噪声子空间$\hat{\mathbf{U}}_n$。利用子阵的导向矢量$\hat{\boldsymbol{a}}(\theta,\varphi)=\hat{\boldsymbol{a}}_Y(\theta,\varphi)\otimes\hat{\boldsymbol{a}}_Z(\theta)$进行空间谱搜索，最终获得的MUSCI谱可以表示为：

$$P_{\text{MUSIC}}(\theta,\varphi)=\frac{1}{\hat{\boldsymbol{a}}^H(\theta,\varphi)\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\hat{\boldsymbol{a}}(\theta,\varphi)}。 \quad (20)$$

虽然将四维的MUSIC谱简化成二维的MUSIC谱可大大降低复杂度，但是仍然需要二维谱峰搜索，复杂度仍然非常高。为了进一步降低复杂度，这里引入root-MUSIC算法[23]，该算法可以避免谱峰搜索，从而大大降低复杂度。

Root-MUSIC算法需要先定义一个多项式：

$$f(z)=\mathbf{p}^H(z)\mathbf{U}_n\mathbf{U}_n^H\mathbf{p}(z), \quad (21)$$

其中，$\mathbf{p}(z)=[1\quad z\quad\cdots\quad z^{M-1}]^T$。定义当$z=\exp(j\omega)$时，多项式的根正好位于单位圆上，只要获得上式的根即可获得目标的角度信息。显然上述求根公式只适合均匀线阵。当采用一个垂直维度的均匀线阵进行角度估计时，假设$\hat{z}_i$为第i个位于单元圆上的根，那么垂直维度的角度可以通过公式（22）估计：

$$\theta_i=\arccos\left(\frac{\lambda}{2\pi d_V}\arg\{\hat{z}_i\}\right)。 \quad (22)$$

对基站来说，当天线阵为二维阵时，需要将二维阵转换成两个一维阵，然后分别采用root-MUSIC算法获取目标的方位角和俯仰角，这里不再赘述。

2 通感一体化应用

从通信感知一体化的初衷来看，通信感知一体化系统基于获得的感知信息，可提供定位、成像、虚拟环境重构等服务。此外，这些服务还可以用于提升移动通信的自身性能。在所能获得的感知信息和服务基础之上，通信感知一体化系统将更好地服务于未来智慧低空、智慧交通、智慧生活等。

2.1 智慧低空

近年来，中国民用无人机市场规模快速增长。2020—2024年，中国民用无人机产业年复合增长率超过50%。2020年民航局批准13个无人机试验基地。工业无人机市场规模年增长率超过50%。2025年无人机将实现规模商用。对无人机进行有效探测和管控，是产业规模发展的前提。无人机探测是近年来的重要研究方向。未来3~5年，无人机监控和防御的市场规模超过330亿元[24]。

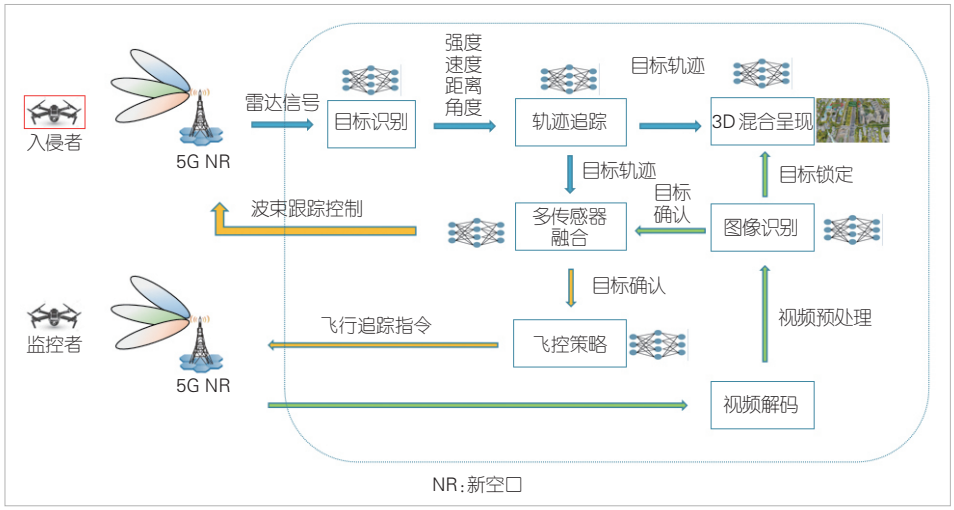
传统的低空雷达方案面临3个困境。1）部署成本高：需要在敏感空域重新部署低空警戒雷达；2）频谱资源短缺：目前雷达频段一般主要是24 GHz和77 GHz；3）感知范围受限：大多数雷达的EIRP受限导致最远感知距离有限。通感一体化方案助力智慧低空发展，在智慧低空中的主要应用包括：无人机入侵检测、无人机监管和避障、无人机路径管理等。

相比于传统的低空雷达方案，通感一体化无人机感知方案具有3个明显的优势：1）部署成本低，可在传统基站上部署感知功能；2）频谱可以共享，即通信和感知共享频谱，提高了频谱利用率；3）感知范围广，例如单站感知距离超过1 km，依赖5G基站可实现大范围连续覆盖。

2022年8月，中国联通携手中兴通讯在上海完成低空无人机通感算控一体化验证测试。该测试以低空园区安防应用场景为背景，采用24 GHz毫米波单AAU实现通信和感知信号的发送与接收。低空无人机的感知精度达到亚米级，探测距离超过1 km，通信和感知过程中性能良好稳定。2022年11月，在IMT-2020（5G）推进组的指导下，中兴通讯使用4.9 GHz低频5G商用基站，实现了室外无人机超过1 400 m的感知距离。智慧低空场景的通感一体化平台如图4所示。

2.2 智慧交通

通感一体化在智慧交通中的应用目前聚焦于车辆信息统



▲图4 智慧低空场景的通感一体化平台

计、行人入侵检测、辅助无人驾驶等方面。当前主流的智慧交通方案可以分为两大类：单车智能方案和车路协同或者通感一体方案。

单车智能方案主要采用汽车雷达和摄像头等，但存在以下缺陷：1）覆盖有盲区：单车感知覆盖区域成扇形，有盲点区域；2）瞬时误差大：单车由于移动速度快，只掌握瞬时的感知信息，目标虚警漏警概率大；3）遮挡影响大：容易产生前车遮挡，无法检测被遮挡区域状况；4）感知距离短，且不能全天候感知：恶劣天气感知范围大大缩减，无灯光感知效果也受限。

相比而言，通感一体方案在以下方面具有明显优势：1）自动驾驶需要车路协同，通感一体与单车雷达是互补和协同关系；2）通感一体化系统不需部署新系统，具有无缝、全天候、低时延等优势，可满足控制辅助、服务车联网等应用需求。

2022年11月，在IMT-2020（5G）推进组指导下，中兴通讯采用24 GHz毫米波通感一体基站，在上海城市开放道路（双向六车道）完成了车辆和行人感知测试，实现了超过800 m的感知距离和车道级的感知精度。随后，中兴通讯使用4.9 GHz低频5G商用基站，完成了室外无人机、车辆和行人感知测试验证。验证结果显示，在一定环境下室外能够同时实现无人机、车辆和行人多目标移动轨迹感知，初步验证了低频感知在低空安防、交通安全管理等的应用可行性。

2.3 智慧生活

通感一体化在智慧生活方面的应用目前主要聚焦于呼吸监测、入侵检测、健身检测、手势检测、天气检测等方面。传统的智慧生活检测方案存在诸多限制，例如：1）需要检测仪或者可穿戴设备进行接触式检测；2）需要额外安装摄像头、传感器、探测器等，部署成本较高；3）感知范围往往较短；4）视频类传统检测方案涉及个人敏感信息。

相比而言，通感一体化方案在智慧生活方面具有诸多优势：1）采用无线信号，可以进行全天候非接触式检测和识别；2）在基站上部署感知功能，不需要额外部署新系统，具有无缝、广覆盖等优势；3）不涉及个人敏感信息等。

2022年11月，在IMT-2020（5G）推进组指导下，中兴通讯使用4.9 GHz低频5G商用基站，在室内场景下开展了视距和非视距呼吸感知测试。其中，在视距场景下呼吸监测能力与商用呼吸监测仪表相当，可助力健康管理等感知应用。

3 结束语

通感一体化开辟了全新的业务方向，赋予无线网络感知物理世界的能力，既充分满足多维感官的交融互通需求，又有效支撑通信能力的广域拓展，具有超越传统移动通信网络连接的巨大潜力。本文概述了5G-A和6G演进过程中通感一体化的应用方向，给出了基于现有通信基站的感知能力边界，以及通感一体化系统参数估计的匹配滤波算法，最后提出了一种高分辨率、低复杂度的感知信号处理算法，有助于推动通感一体化技术从理论走向实际应用。

致谢

本研究得到中兴通讯股份有限公司耿鹏、崔亦军、赵志勇、窦建武、吴建军等专家的帮助，在此对他们表示感谢！

参考文献

[1] LIU Y J, LIAO G S, XU J W, et al. Adaptive OFDM integrated radar and communications waveform design based on information theory [J]. IEEE communications letters, 2017, 21(10): 2174-2177. DOI: 10.1109/LCOMM.2017.2723890
[2] 刘凡, 袁伟杰, 原进宏, 等. 雷达通信频谱共享及一体化: 综述与展望 [J]. 雷达学报, 2021, 10(3): 467-484. DOI: 10.12000/JR20113
[3] 潘成康. 通信感知一体化技术思考 [J]. 中兴通讯技术, 2022, 27(5): 53-56. DOI: 10.12142/ZTETJ.202205010
[4] 江甲沫, 韩凯峰, 徐晓燕. 6G通信感知一体化系统的性能指标 [J]. 中兴通讯技术, 2022, 27(5): 39-45. DOI: 10.12142/ZTETJ.202205008
[5] 李晓阳, 周梓钦, 贡毅. 通信感知计算一体化波束赋形设计 [J]. 中兴通讯技术, 2022, 27(5): 23-28. DOI: 10.12142/ZTETJ.202205006
[6] ROBERTON M, BROWN E R. Integrated radar and communications based on chirped spread-spectrum techniques [C]//Proceedings of IEEE MTT-S

International Microwave Symposium Digest. IEEE, 2003: 611–614. DOI: 10.1109/MWSYM.2003.1211013
[7] CHEN X B, WANG X M, XU S F, et al. A novel radar waveform compatible with communication [C]//Proceedings of 2011 International Conference on Computational Problem–Solving (ICCP). IEEE, 2011: 177–181. DOI: 10.1109/ICCPS.2011.6092272
[8] ZHANG W K, LIU Z P. Design and implementation of modulator of a novel rader waveform compatible with communication [C]//Proceedings of 2013 International Workshop on Microwave and Millimeter Wave Circuits and System Technology. IEEE, 2014: 357–360. DOI: 10.1109/MMWCST.2013.6814653
[9] LIU Y J, LIAO G S, YANG Z W, et al. Design of integrated radar and communication system based on MIMO–OFDM waveform [J]. Journal of systems engineering and electronics, 2017, 28(4): 669–680. DOI: 10.21629/JSEE.2017.04.06
[10] IMT–2030 (6G)推进组. 新型双工技术研究报告 [R]. 2022
[11] GUO X J, GAO Y S, WANG K Z, et al. Transmitter leakage canceling for LFMCW SAR [C]//Proceedings of 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016: 5047–5050. DOI: 10.1109/IGARSS.2016.7730316
[12] 肖博, 霍凯, 刘永祥. 雷达通信一体化研究现状与发展趋势 [J]. 电子与信息学报, 2019, 41(3): 739–750. DOI: 10.11999/JEIT180515
[13] LELLOUCH G, MISHRA A, INGGS M. Impact of the Doppler modulation on the range and Doppler processing in OFDM radar [C]//Proceedings of 2014 IEEE Radar Conference. IEEE, 2014: 803–808. DOI: 10.1109/RADAR.2014.6875700
[14] DUAN J Q. Multicarrier coherent pulse shaping for radar and corresponding signal processing [C]//Proceedings of 2007 8th International Conference on Electronic Measurement and Instruments. IEEE, 2007: 3–843. DOI: 10.1109/ICEMI.2007.4351049
[15] 刘永军, 廖桂生, 杨志伟, 等. 一种超分辨 OFDM 雷达通信一体化设计方法 [J]. 电子与信息学报, 2016, 38(2): 425–433. DOI: 10.11999/JEIT150320
[16] 丁鹭飞. 雷达原理 [M]. 西安: 西北电讯工程学院出版社, 1984
[17] 斯科尼克(Merrill I.Skolnik). 雷达手册 [M]. 南京电子技术研究所译. 北京: 电子工业出版社, 2010
[18] IMT–2030 (6G)推进组. 通信感知一体化技术研究报告 [R]. 2022
[19] SCHMIDT R. Multiple emitter location and signal parameter estimation [J]. IEEE transactions on antennas and propagation, 1986, 34(3): 276–280. DOI: 10.1109/TAP.1986.1143830
[20] ROY R, PAULRAJ A, KAILATH T. ESPRIT: a subspace rotation approach to estimation of parameters of cisoids in noise [J]. IEEE transactions on acoustics, speech, and signal processing, 1986, 34(5): 1340–1342. DOI: 10.1109/TASSP.1986.1164935
[21] MARCOS S, BENIDIR M. On a high resolution array processing method non–based on the eigenanalysis approach [C]//Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'90). IEEE, 1990: 2955–2958
[22] SHAN T J, KAILATH T. Adaptive beamforming for coherent signals and interference [J]. IEEE transactions on acoustics, speech, and signal processing, 1985, 33(3): 527–536. DOI: 10.1109/TASSP.1985.1164583
[23] RAO B D, HARI K V S. Performance analysis of root–music [J]. IEEE transactions on acoustics, speech, and signal processing, 1989, 37(12): 1939–1949. DOI: 10.1109/29.45540
[24] 中国移动. 5G–Advanced无线网络能力与产业进展白皮书 [R]. 2022

作 者 简 介

李萍，中兴通讯股份有限公司资深算法系统工程师；主要研究方向包括移动通信关键数字信号处理技术、多天线技术、系统性能优化与技术、智能超表面技术、组网与网络规划技术等；负责多项国家级重大课题，申请专利近100项，发表论文10篇。

郭晓江，中兴通讯股份有限公司资深算法专家；主要研究领域为无线通信信号处理技术、雷达信号处理技术、通感一体化等；发表论文10余篇。

神经辐射场加速技术综述

Survey of Neural Radiance Field Acceleration Technologies

郑清芳/ZHENG Qingfang[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.202302015
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1433.014.html
网络出版日期：2023-04-11
收稿日期：2023-02-26

摘要：神经辐射场（NeRF）技术可以从2D图像中学习场景的3D隐式模型，并合成出高清逼真新视角图像。该技术有着良好的应用前景，受到业界广泛关注。针对NeRF技术运算缓慢的问题，近两年业界研究者提出了各种加速技术。对现有加速技术进行了综述，分类梳理并分析了速度提升背后的技术机理和工程技巧，同时讨论了未来加速技术演进的方向。本研究有助于激发更高效算法的产生，从而推进NeRF技术在视觉内容生成及其他领域的应用。

关键词：NeRF；神经渲染；视点合成；体渲染

Abstract: The neural radiance field (NeRF) technology, which can learn the 3D implicit representation of a scene from a set of 2D images and synthesize high-resolution and photo-realistic images of novel views, has aroused extensive research interest due to its vast application potential. In order to solve NeRF's problem of slow running speed, various acceleration technologies have been proposed in recent two years. We review current acceleration technologies by categorizing and analyzing their technical mechanism and engineering skills. We also discuss directions for further acceleration. Our work will contribute to inspiring the invention of more efficient algorithms and promote NeRF's application in multiple fields including visual content generation and beyond.

Keywords: NeRF; neural rendering; view synthesis; volume rendering

过去10年，视频相关技术[1-5]领域最深刻的变革发生在内容分析方面。自从2012年AlexNet[6]在ImageNet大规模图像识别挑战（ILSVRC）竞赛中夺冠以来，基于深度学习的计算机视觉技术突飞猛进，将内容分析的准确率提升至前所未有的水平，并催生出巨大的市场应用规模。以人脸识别为代表的各项视频内容分析技术走出实验室，服务于千行百业。

未来10年，同样激动人心的突破有望发生在视觉内容生成方面。简单便捷地从2D视频/图像集中合成出崭新视角的视频/图像，甚至重建出物体及场景的3D模型，并且画质达到照片级的逼真度和清晰度，一直是内容生成方面的长期技术研究课题[7]。2020年美国加州大学伯克利分校的研究团队提出神经辐射场技术（NeRF）[8]。NeRF因其创新的方法及突出的效果，吸引了业界的广泛关注，成为视图合成/3D重建领域新的技术框架。自从发表后近两年的时间里，该NeRF论文被引用超过1 000次。同时业界研究者对NeRF技术进行了大量的改进，并将其应用领域扩展到视频编辑[9]、数据压缩[10]、虚拟人[11]、城市建模[12]、地图构建[13]等诸多方面。

NeRF技术的一个显著缺点是模型训练及图像渲染的速度极慢。在Nvidia的高端显卡上，训练一个场景的模型需耗时1~2 d，而从模型中渲染出一幅800×800分辨率的图像需耗时超过20 s。运算速度方面的不足阻碍了NeRF技术在实际应用中的部署。可喜的是，经过业界研究者近两年的努力，渲染速度提升超过10 000倍[14]，训练速度提升超过300倍[15]。

针对NeRF的各种加速技术，本文梳理并总结了速度提升的技术机理和工程技巧，并分析各项技术之间互相结合以达到复合加速效果的可能性，从而有助于激发更高效算法的产生，进一步推进NeRF技术在内容生成及其他领域的应用。

1 相关研究工作

文献[16]和文献[17]分别对2021年3月之前的NeRF相关技术做了综述。文献[16]针对促使NeRF出现的各种技术和NeRF出现后的各种改进性技术这两个主题，提供了注释性的参考文献，但不涉及对各技术的详细说明。文献[17]将相关技术大致分为两大类：第1类对NeRF表示方法的理论性

质和不足进行分析，并提出优化策略，包括对合成精度、绘制效率以及对模型泛用性的优化；第2类则以NeRF的框架为基础对算法进行扩展和延伸，使其能够解决更加复杂的问题。文献[16]和文献[17]促使更多的研究者对NeRF进行研究，但也因其成文时间较早，无法涵盖对2021年3月以后NeRF的许多重要进展的总结。

文献[18]综述了神经渲染技术的整体发展。神经渲染技术广义上是指所有利用神经网络产生新的视觉内容的技术，而NeRF仅是其中的一个子领域，侧重于合成出新的视角的视觉内容。文献[18]重点介绍了将经典渲染与可学习3D表示相结合的高级神经渲染方法，尽管提及了许多NeRF相关的文献，但本质上不是针对NeRF的综述。

在本文的撰写过程中，加拿大滑铁卢大学的研究者在Arxiv.org上展示了预印本[19]，全面介绍了过去两年业界提出的各种NeRF改进，以及NeRF技术在各种计算机视觉任务中的应用。与文献[19]不同，本文着眼于运算速度的提升，对各种加速技术进行分类，阐释技术背后的机理和工程技巧，展现NeRF发表以来的技术演进脉络，以期为相关研究者提供有益参考。

2 NeRF技术简介

对于给定的三维场景，任意位置的外观取决于具体位置和观测角度。场景表现出的颜色与光照条件相关，导致从不同角度观察同一位置时颜色也会出现变化。NeRF是一个描述三维场景的函数$(r,g,b,\sigma)=F_\Theta(x,y,z,\theta,\varphi)$，其中$F_\Theta$用多层感知机（MLP）来具体表示。输入位置信息$(x,y,z)$和观测角度$(\theta,\varphi)$后，该函数输出该位置的体密度$\sigma$和在对应观测角度的颜色值$rgb$。在基于NeRF的场景表示基础上，可以采用经典体渲染方法渲染出不同视角的新图像。具体地，对于图像中任意像素，沿着观测角度的光线r采样N个点$X_i(i=1,\cdots,N)$，对每个采样点先根据F_Θ计算出σ_i和rgb_i，然后根据以下公式计算出最终的颜色值：

$$\hat{rgb}=\sum_{i=1}^{N} rgb_i w_i, \tag{1}$$

$$w_i=T_i\alpha_i, \tag{2}$$

$$T_i=\exp\left(-\sum_{j=1}^{i-1}\sigma_j\delta_j\right), \tag{3}$$

$$\alpha_i=1-\exp(-\sigma_i\delta_i), \tag{4}$$

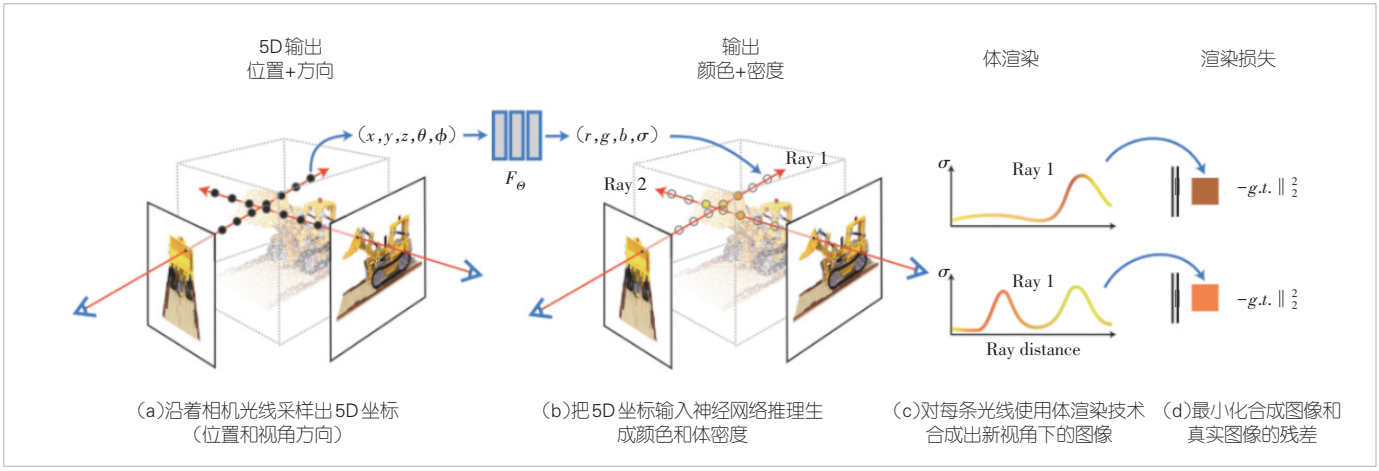
其中，δ_i代表在光线r上的采样间隔。

为了训练出F_Θ对应的MLP具体参数，对于给定的场景，采用不同位姿的摄像头拍摄得到n幅图像，利用梯度下降的方法，通过最小化预测图像I_p与真值图像I_c之间的误差对F_Θ进行拟合，即$\min\sum_{i=1}^{n}|I_p-I_c|^2$。

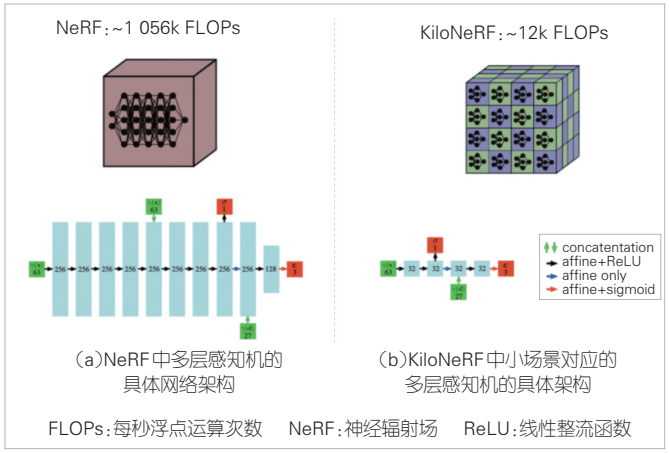
图1给出了NeRF算法的流程。在文献[8]中，作者为了得到含有更多高频信息的输出图像，对输入MLP的位置和视角参数进行了高阶编码操作$\gamma(.)$。另外，为了提升运行速度，作者采用先粗放后精细的策略提高采样的效率：首先在光线方向上均匀采样64个点，然后由这64点的密度值估计出密度分布函数，再对高密度的区域采样128点。

NeRF的运行速度十分缓慢，主要有两个原因：

1）巨大的计算量。例如：在渲染一幅分辨率为800×800的图像时，NeRF为了计算出每个像素的最终颜色值，需在光线行进的方向上采样192（64+128）个点，并进行256次MLP推理。这意味着渲染该图像总共需要800×800×256=163 840 000次MLP推理。MLP的网络架构如图2（a）所示。每次推理计算需要超过100万次浮点运算。总体而



▲图1 神经辐射场算法流程[8]



▲图2 NeRF与KiloNeRF的比较[21]

言，整个过程需要超过100T次浮点计算。

2）低效的实现方式。在通常针对Nvidia图形处理器（GPU）优化的深度学习函数库中，MLP是逐层计算的。每一层用一个核函数来实现具体的计算，并向GPU全局显存写入计算结果，而下一层在计算时又需从全局显存读取该计算结果并将其作为本层的输入。在目前的GPU芯片架构中，全局显存的数据读写速度远小于计算速度。频繁的数据读写严重制约了GPU的实际工作性能。

3 加速技术介绍

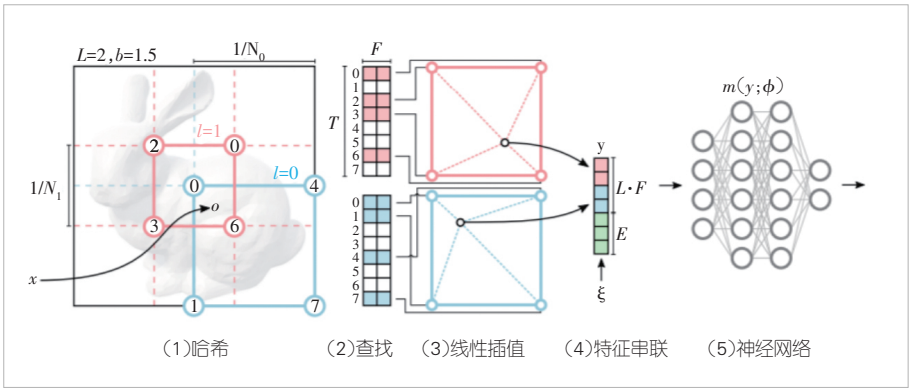
鉴于NeRF巨大的潜在应用前景，针对当前NeRF十分低效的运行速度，近两年来研究者们提出了一系列加速技术。由上一节的分析可知，对于给定分辨率的图像，NeRF的实际运行速度受到每道光线中的MLP的计算复杂度、采样点数量、硬件的技术特性等因素的综合影响。本文从采用小型化MLP、减少采样点数量、缓存中间计算结果以及充分利用硬件特性4个方面对现有技术进行分类和介绍。

3.1 采用小型化MLP

在NeRF的实现中，图像中每个像素的色彩值的计算都需上百次的MLP推理，因此减轻MLP的计算复杂度是非常有必要的。如果采用更小型化的MLP，不论是减少其深度还是宽度，都会导致模型表征能力的下降，损害最终输出图像的视觉质量。因而，此类方法的关键在于采用何种策略能够确保小型化MLP的最终输出不会影响最终的视觉质量。

DeRF[20]和KiloNeRF[21]采用了分而治之的策略：用更小的MLP来表示目标场景的一部分而非整个场景。DeRF把整个场景划分为不规则的互相独立的16个Voronoi单元。相比于NeRF中的MLP，每个Voronoi单元对应的MLP深度不变但宽度减半，因此计算量只有原来的1/4。在渲染时，由于每条光线上的每个采样点只计算与它相对应的MLP，因此整体速度可以达到原始NeRF的1.7倍。KiloNeRF采用沿坐标轴均匀分解场景的方法，最多可以把场景分为16×16×16=4 096个小场景。每个小场景对应的MLP的具体架构如图2（b）所示。该架构仅有4个隐含层且每层只有32个通道，其计算量为NeRF中MLP的1/87。相比于原始的NeRF，KiloNeRF的总体渲染速度可以达到3个数量级的加速倍速。为了保证视觉质量，在KiloNet训练过程中采用知识蒸馏的方式，使KiloNeRF的输出与NeRF的输出相一致。

Instant NeRF[22]的核心思路是：既然MLP的最终输出值取决于MLP自身的参数和输入的特征，那么小型化MLP表征能力的减弱可以通过增强输入特征的表征能力来弥补。Instant NeRF中的MLP由两个分别包含1个及2个隐含层且每层都为64个通道的小型MLP串联组成。不同于NeRF中的位置编码，Instant NeRF对输入参数采取多分辨哈希编码方式：输入参数在某个分辨率中经过哈希后对应一个特征向量，把输入参数在所有分辨率中对应的特征向量串联起来形成最终的特征向量。Instant NeRF不但加速了渲染过程，在Nvidia RTX 3090 GPU上能够以60 fps的速度输出1 920×1 080的图片，而且解决了NeRF模型训练慢的问题，将NeRF训练速度提高了60倍。实验结果表明，在最快的情况下，Instant NeRF模型的训练时间只需要5 s。图3以2D空间场景为例解释了Instant NeRF的计算过程。该计算过程包括5个步骤：1）对于给定的输入坐标x，在不同的分辨率中分别找到周围的体素；2）在哈希表中查询不同分辨率的体素所对应的特征向量；3）根据x在各自体素中的相对位置，插值计算



▲图3 Instant NeRF的计算过程[22]

出 x 在不同分辨中的特征向量；4）将在各分辨率上的特征向量串联，形成最终的特征向量；5）将特征向量输入神经网络进行推理计算

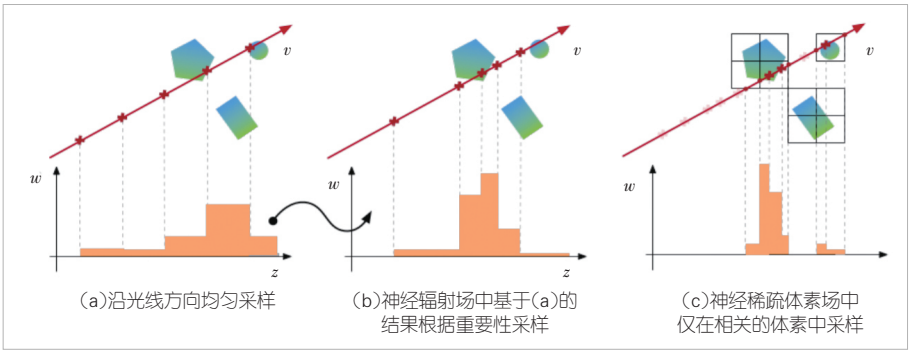
3.2 减少采样点数量

尽管 NeRF 已经采用了层次化的采样策略来避免对整条光线进行密集采样，但是仍然需要固定的 192 个采样点。事实上，由于目标场景通常无法完全充满整个三维空间，必然有某些采样点落在目标场景之外。另外，某些采样点在视角方向上被完全遮挡，使得这些采样点对最终的计算结果并无帮助。因此，更合理的采样策略应该可以避免把计算资源浪费在这些采样点上。

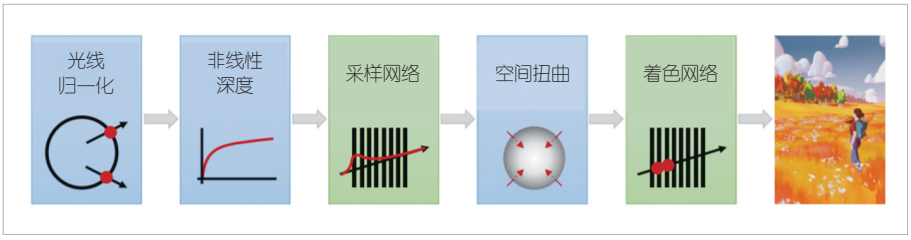
文献[23]引入了一种用于快速和高质量自由视角渲染的新神经场景表示方法：神经稀疏体素场（NSVF）。NSVF 定义了一组由稀疏体素八叉树组织的体素有界隐式场，以对每个体素中的局部属性进行建模，并为体素的每个顶点分配一个特征。体素内部具体位置的特征通过对体素 8 个顶点处的特征进行插值计算。在渲染过程中，需要对每条光线进行轴对齐边界框相交（AABB）测试，即比较从光线原点到体素的 6 个边界平面中的距离，检查光线是否与体素相交。对于不相交的空体素，可以直接跳过，从而实现 10 倍以上的渲染加速。图 4 比较了 NSVF 与 NeRF 的不同采样策略。因为 NSVF 渲染过程是完全可微的，所以可以通过将渲染的输出结果与一组目标图像进行比较，然后进行反向传播来实现端到端优化。监督训练 NSVF 的过程采用了渐进式的策略，使得不包含场景信息的稀疏体素会被修剪掉，以允许网络专注于具有场景内容的体积区域的隐函数学习。文献[23]中的实验表明，只需 1 万~10 万个稀疏体素就能够实现复杂场景的逼真渲染。

尽管 TermiNeRF[24]、NeuSample[25]、DONeRF[26]的具体做法不同，但背后的思想却是类似的：在训练的过程中，联合训练 NeRF 和一个采样神经网络，而在渲染过程中，仅需对每条光线推理一次采样神经网络，即可得到所需的全部采样点位置。以加速效果最好的 DONeRF 为例，为了在不影响图像质量的前提下大幅减少每条光线所需的采样点数量，文献[26]的作者引入真实深度信息，只考虑物体表面周围的重要采样点。DONeRF 由一个着色网络和一个采样网络组成。其中，着色网络使用类似 NeRF 的光线行进累积法来输出颜色值，而采样网络则通过将空间沿光线离散化并预测沿光线的采样概率，来预测每条光线上的多个潜在采样对象。为了消除输入的模糊性，光线被转换到一个统一的空间中。作者使用非线性采样来追踪接近的区域，并在采样网络和着色网络之间，对局部采样进行扭曲，以使着色网络的高频预测被引导到前景上。图 5 展示了 DONeRF 的计算过程。实验结果表明，DONeRF 只用 4 个采样点就取得了与 NeRF 相似的图像质量，渲染速度可实现 20 ~ 48 倍的提升。

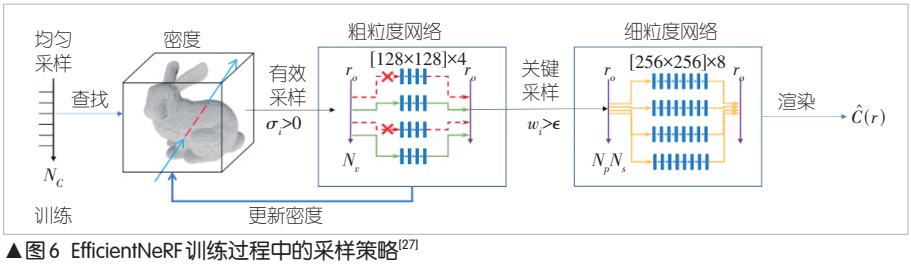
如图 6 所示，EfficientNeRF[27]在训练时采用了与 NeRF 相似的先粗放后精细的采样策略，但是在粗放采样阶段只计算体密度 $\sigma > 0$ 的有效样本，在精细采样阶段只计算 $w > 0.000\,1$ 的关键样本以及与其临近的另外 4 个关键样本，整体的训练时间减少了 88%。每个位置对应的 σ 值被初始化为非零值并存储在 V_σ 中，在后续的每次训练迭代中根据 $V_\sigma^i =$



▲图 4 神经稀疏体素场和神经辐射场的不同采样策略对比[23]



▲图 5 DONeRF 的计算过程[26]



▲图 6 EfficientNeRF 训练过程中的采样策略[27]

$(1-\beta)V_{\sigma}^{i-1}+\beta\sigma(x)$进行更新。其中，$\beta\in(0,1)$是控制更新率的参数，$\sigma(x)$是本次迭代中得到的体密度值。$w$根据公式（2）—（4）计算。

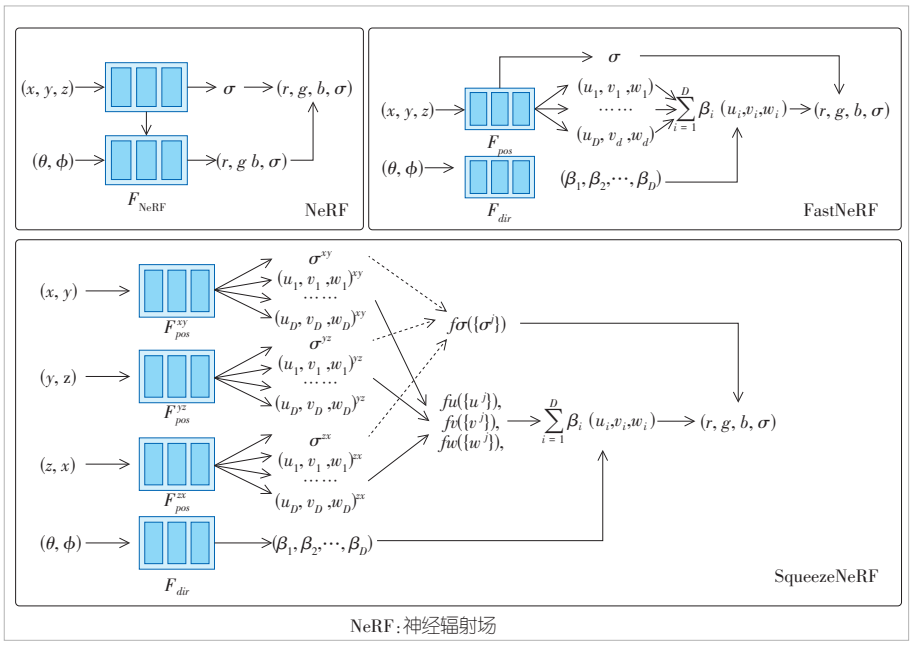
3.3 缓存中间计算结果

通过预先计算并将计算结果缓存起来，在后续使用时直接获取该结果，是常见的加速方法。NeRF本质上是将5维的输入参数（3维的位置参数+2维的视角参数）映射到4维向量的函数$(r,g,b,\sigma)=F_{\Theta}(x,y,z,\theta,\varphi)$。对此，最简单方法是直接将5维输入空间离散化，并将对应的计算结果全部缓存。但该方法因为所需的内存容量太大而不具可行性：即使假设输入参数每个维度的分辨率都是512，所需的内存也超过150 TB。

如图7所示，FastNeRF[28]和SqueezeNeRF[29]都利用了函数分解的思想。FastNeRF将NeRF分解为两个函数：第1个函数$(\sigma,u,v,w)=F_{\text{pos}}(x,y,z)$将位置参数映射到体密度$\sigma$和辐射图$(u,v,w)$；第2个函数$\beta=F_{\text{dir}}(\theta,\varphi)$则将视角参数映射到与辐射图对应的权重向量$\beta$，其中$u$、$v$、$w$、$\beta$都是D维向量。$F_{\text{dir}}$和$F_{\text{pos}}$的全部结果被预先计算并缓存。渲染时先根据位置和视角参数从缓存中分别查询得到权重向量和深度辐射图，然后通过计算二者的内积即可获得颜色值：$(r,g,b)=\sum_{i=1}^{D}\beta_i\,(u_i,v_i,w_i)$。该计算量远小于NeRF中的一次MLP推理，因此FastNeRF的渲染速度比NeRF提升了约3 000倍。假设k和l分别表示位置和视角方向的分辨率，缓存F_{pos}和F_{dir}所需的内存空间复杂度分别为$O(Dk^3)$和$O(Dl^2)$，且$O(Dk^3)$和$O(Dl^2)$远小于缓存NeRF所需的内存空间复杂度$O(k^3l^2)$。在通常设置中，$k=l=1\,024$，$D=8$，FastNeRF所需的最大缓存空间大约为54 GB。为了能够运行在内存更小的嵌入式设备上，如图7所示，SqueezeNeRF在FastNeRF的基础上将函数F_{pos}进一步分解为3个函数并缓存其结果，所需内存空间的复杂度也从$O(Dk^3)$降到$O(Dk^2)$。

在文献[30]中，作者首次提出NeRF-SH模型$(\sigma,k)=F_{\text{SH}}(x,y,z)$，将空间位置映射为体密度$\sigma$和球谐系数$k=(k_l^m)_{l:0\leqslant l\leqslant l_{\max}}^{m:-l\leqslant m\leqslant l}$。NeRF-SH的训练过程、渲染过程与NeRF类似。作者采用稀疏八叉树表示3维场景，在叶子节点上存储各体素位置对应的体密度σ和球谐系数k。该位置在方向(θ,φ)的颜色值为$(r,g,b)=S(\sum_{l=0}^{m}\sum_{m=-l}^{l}k_l^mY_l^m(\theta,\varphi))$，其中，$Y_l^m(\theta,\varphi)$是与方向$(\theta,\varphi)$相对应的球谐系数，$S(\mathrm{x})=(1+\exp(-x))^{-1}$。与NeRF相比，PlenOctree技术不仅可以将渲染速度提升3 000多倍，还能将训练速度提升约4倍。在一般的场景中，PlenOctree需要的内存空间通常不超过5 GB。EfficientNeRF[27]采用不同的数据结构来组织缓存数据，用一个两层的树NerfTree代替PlenOctree中的稀疏八叉树，实现了更快的缓存查询速度。

根据图形学知识，物体的颜色可以表示为漫反射色与镜面反射色两者之和，其中镜面反射色与相机观测角度有关。文献[31]定义一个函数$(\sigma,\mathrm{c}_{\text{diffuse}},v_{\text{specular}})=F(x,y,z)$，场景中的每个位置都对应着体密度$\sigma$、漫反射色$c_{\text{diffuse}}$，以及与镜面反射色有关的4维特征向量$v_{\text{specular}}$。这些数据经预先计算后被缓存在稀疏神经辐射网格（SNeRG）中。在渲染时，每条光线上采样点通过查询缓存直接获取对应的σ、c_{diffuse}和v_{specular}，这些采样点的c_{diffuse}和v_{specular}分别根据σ值加权累积得到C_{diffuse}和V_{specular}。V_{specular}只需经过一次MLP推理就可得到累积的镜面反射色C_{specular}，则最终像素的颜色为$rgb=C_{\text{diffuse}}+C_{\text{specular}}$。利用该技术在AMD Radeon Pro 5500M GPU上渲染Synthetic 360°图像时，速度可达到84 fps。SNeRG中的数组内容可以通过便携式网络图形（PNG）、联合图像专家组（JPEG）等算法被压缩到平均90 MB以内。



NeRF：神经辐射场

▲图7 NeRF、FastNeRF、SqueezeNeRR 3种神经网络架构的对比

3.4 充分利用硬件特性

在实际部署中，算法总是运行在特定芯片上的。提升算法的运行速度通常意味着必须高效地利用芯片中的并行处理能力和内存/缓存资源。

在Nvidia GPU上，Instant NeRF和KiloNeRF取得显著加速效果的重要原因之一在于：小型化的MLP能够更充分利用GPU芯片的技术特性。例如，与NeRF中的MLP相比，KiloNeRF中MLP的理论计算量只有1/87，而实际加速效果却高出1 000倍。在常规的统一计算设备架构（CUDA）深度学习函数库中，MLP逐层用一个核函数计算，需要在GPU的全局显存中读写中间计算结果。对于计算小型化MLP，数据访问的时间远大于数据计算的时间。KiloNeRF和Instant NeRF都使用CUDA重新编写一个核函数，并在其中完成MLP的所有计算，从而省去中间计算结果的数据搬运操作，减少频繁启停核函数的时间开销。

为了在移动设备上部署NeRF，Google推出MobileNeRF[14]。MobileNeRF充分利用了标准GPU光栅化管道的并行性。测试结果表明，在输出图像视觉质量相当的前提下，MobileNeRF能够比SNeRG快10倍，相当于比原始NeRF快了10 000倍以上。为了适配GPU的光栅化管道，MobileNeRF采用与原始NeRF不同的训练过程和表征方法，用带有纹理的多边形来表征每个场景模型。其中，多边形大致沿着场景表面排布，纹理图中存储特征向量和离散的不透明度。渲染时MobileNeRF先利用带Z-buffering的经典多边形光栅化管道为每个像素生成特征向量，然后将特征向量传递给OpenGL着色语言（GLSL）片段着色器，并在其中运行小型化MLP，生成每个像素的色彩值。此外，MobileNeRF的GPU显存利用率也高于SNeRG，在运行过程中前者占用的GPU显存约为后者的1/5。

基于现场可编程逻辑门阵列（FPGA），上海科技大学开发了首个针对NeRF渲染算法的定制化芯片ICARUS[32]。ICARUS的架构由定制的全光核组成，其中每个全光核集成了位置编码单元（PEU）、MLP引擎和体渲染单元（VRU）。当采用40 nm互补金属氧化物半导体（CMOS）工艺且工作在300 MHz时，单个全光核仅占7.59 mm^2面积，功耗为309.8 mW，能效比GPU高146倍。ICARUS的高效性能主要得益于以下3个方面：

1）使用经过量化的定点数模型，尤其对于对复杂度最高的MLP计算，使用移位累加等近似算法。

2）全光核内部完成NeRF的全部计算过程。当芯片加载经过训练的NeRF网络模型参数后，只要输入观察位置与视角，即可输出对应像素的最终色彩值，无须在片外存储中间计算结果，从而消除了各运算单元内部、单元之间的数据存储和搬运操作。

3）每个像素的计算过程和结果完全独立，控制逻辑大大简化，可以方便地通过增加全光核数量来实现并行加速。

4 总结与讨论

NeRF技术可以从不同视角的2D图像集中学习并建立3D场景的隐含模型，并渲染出崭新视角的图像。不仅如此，新图像的视觉效果能够达到非常逼真的程度。自从2020年第1篇关于NeRF的论文发表以来，NeRF技术为视角合成乃至3D重建领域带来新的研究思路。在两年左右的时间里，该技术引起了业界广泛关注，并得到了突飞猛进的发展。在未来，NeRF技术将为视觉内容生成领域带来巨大变革，如同当前深度卷积网络技术为视觉内容分析领域带来的变革一样，在虚拟现实（VR）/增强现实（AR）及未来的元宇宙时代起到关键作用。

为了解决由NeRF技术运行速度缓慢导致的实际部署难的问题，研究者们已经提出各种加速技术。本文介绍了NeRF的技术原理，并分析该技术运行缓慢的原因：在获得每个像素的最终颜色值时，整体运行速度取决于MLP的计算复杂度、每道光线沿线的采样点数量等综合因素。本文相应地从采用小型化MLP、减少采样点数量、缓存中间计算结果以及充分利用硬件特性4个方面对现有技术进行综述，介绍了各技术的加速原理和实现方法，希望可以帮助相关研究者快速了解本领域的技术现状及演进脉络。

另外，NeRF相关技术仍在快速发展中，同时实际应用场景仍需要更加高效的加速技术。展望未来的技术发展，我们认为应重点关注以下几个研究方向：

1） 复合加速效果

必须指出的是，尽管本文从技术原理的角度做了正交分类，并在各分类中列举了代表性的工作，但所提及的诸多具体技术都综合利用了多种加速原理。例如，EfficientNeRF在训练阶段减少采样点数量，而在渲染阶段缓存计算结果；Instant NeRF和KiloNeRF都采用小型化MLP，并针对特定GPU架构优化MLP的推理速度。我们推测，通过结合额外的加速原理，现有的方法可以实现更高的加速倍数，例如：KiloNeRF可以进一步与DONeRF相结合，减少采样点数量，进一步提高渲染速度；Instant NeRF可以在训练阶段结合EfficientNeRF中的采样策略，并采用DS-NeRF[33]中在损失函数里增加深度信息约束的做法，来加快训练过程收敛。对此，我们希望本文的分析能够启发感兴趣的研究者设计出更加高效的算法。

2）训练加速和渲染加速

NeRF技术的特点是：针对每一个静态场景都需要训练一个模型，然后从模型中渲染出所需的图像。鉴于原始NeRF的训练渲染过程都十分缓慢，为了在实际应用中使用NeRF技术，加速训练过程和渲染过程都十分必要。前述加速方法中有些只适用于渲染过程，甚至是以牺牲训练速度为代价的，例如：在KiloNeRF和MobileNeRF之类的采用小型化MLP的方法中，为了保证最终模型的输出质量，需要先训练出NeRF中的MLP模型，再通过知识蒸馏的方式，训练出更小型化的MLP。

在诸如电商货品展示的应用场景中，可以通过两阶段的过程来综合利用上述两类加速方法：先离线使用某种加速方法训练出NeRF模型，并进一步转换成更高效的表达形式，然后在线展示过程中采用另外一种加速方法渲染出图片。而对于诸如3D视频通信的端到端实时应用而言，往往需要同时加速模型训练和渲染过程。在本文提及的方法中，PlenOctree、EfficientNeRF和Instant-NeRF能同时加速训练过程和渲染过程，但训练过程的加速比远小于渲染过程，训练速度远小于30 fps。

3）专用加速芯片

算法与芯片的发展总是相辅相成、互相促进的。当某种算法被广泛采用时，通常研究者会为之设计专用的加速芯片，在性能、成本、功耗等方面实现更佳的匹配，从而进一步推广算法的应用。在关于AlexNet的论文发表之后大约两年的时间，中科院计算所设计出第一款深度卷积网络加速原型芯片DIANNAO[34]。该芯片将速度提升近120倍，从此拉开波澜壮阔的人工智能（AI）计算芯片产业化序幕；在有关NeRF的论文发表之后大约两年的时间，上海科技大学设计出第一款NeRF渲染加速芯片ICARUS，使能效提升近140倍。ICARUS是否会同样在芯片产业风起云涌？现有GPU的技术特性并不完全适配神经渲染的计算流程。类似Mobile-NeRF的技术通过复杂的转化过程后，可以更加高效地利用现有GPU的并行能力，从而能够运行在移动设备中的嵌入式GPU上。我们十分期待业界共同努力，持续创新，研发出神经渲染专用加速芯片产品，并创造出巨大的市场应用空间，使得在各种设备上便捷、快速、经济地渲染出逼真高清的视觉内容成为现实。

参考文献

[1] 张嘉琪, 雷萌, 马思伟. AVS3视频编码关键技术及应用 [J]. 中兴通讯技术, 2021, 27(1): 10–16. DOI:10.12142/ZTETJ.202101004

[2] 杨文哲, 徐迈, 白琳. 视频质量增强模型加速算法 [J]. 中兴通讯技术, 2021, 27(1): 21–26. DOI:10.12142/ZTETJ.202101006

[3] 高宸, 李勇, 金德鹏. 基于图神经网络的视频推荐系统 [J]. 中兴通讯技术, 2021, 27(1): 27–32. DOI:10.12142/ZTETJ.202101007

[4] 吕达, 郑清芳. 构建智能实时网络, 使能5G视频业务繁荣 [J]. 中兴通讯技术, 2021, 27(1): 60–67. DOI:10.12142/ZTETJ.202101013

[5] GAO N Z, YU Y F, HUA X H, et al. A content-aware bitrate selection method using multi-step prediction for 360-degree video streaming [J]. ZTE Communications, 2022, 20(4): 96–109. DOI: 10.12142/ZTECOM.202204012

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84–90. DOI: 10.1145/3065386

[7] CHAN S C, SHUM H Y, NG K T. Image-based rendering and synthesis [J]. IEEE signal processing magazine, 2007, 24(99): 22–33. DOI: 10.1109/msp.2007.4317461

[8] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. NeRF: representing scenes as neural radiance fields for view synthesis [M]//Computer Vision – ECCV 2020. Cham: Springer International Publishing, 2020: 405–421. DOI: 10.1007/978-3-030-58452-8_24

[9] ZHANG J K, LIU X H, YE X Y, et al. Editable free-viewpoint video using a layered neural representation [J]. ACM transactions on graphics, 2021, 40 (4): 1 – 18. DOI: 10.1145/3450626.3459756

[10] ISIK B. Neural 3D scene compression via model compression [EB/OL]. [2023-02-25]. https://arxiv.org/abs/2105.03120

[11] WENG C, CURLESS B, SRINIVASAN P P, et al. HumanNeRF: free-viewpoint rendering of moving people from monocular video [C]// Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 16189–16199. DOI: 10.1109/CVPR52688.2022.01573

[12] TANCIK M, CASSER V, YAN X C, et al. Block-NeRF: scalable large scene neural view synthesis [EB/OL]. [2023-02-25]. https://arxiv. org/abs/2202.05263

[13] ZHU Z H, PENG S Y, LARSSON V, et al. NICE-SLAM: neural implicit scalable encoding for SLAM [EB/OL]. [2023-02-25]. https://arxiv.org/abs/2112.12130

[14] CHEN Z Q, FUNKHOUSER T, HEDMAN P, et al. MobileNeRF: exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures [EB/OL]. [2023-02-25]. https://arxiv. org/abs/2208.00277

[15] ZHANG Q, BAEK S H, RUSINKIEWICZ S, et al. Differentiable point-based radiance fields for efficient view synthesis [EB/OL]. [2023-02-25]. https://arxiv.org/abs/2205.14330

[16] DELLAERT F, YEN-CHEN L. Neural volume rendering: NeRF and beyond [EB/OL]. [2023-02-25]. https://arxiv.org/abs/2101.05204

[17] 常远, 盖孟. 基于神经辐射场的视点合成算法综述 [J]. 图学学报, 2021, 42(3): 376–384. DOI: 10.11996/JG.j.2095-302X.2021030376

[18] TEWARI A, THIES J, MILDENHALL B, et al. Advances in neural rendering [EB/OL]. [2023-02-25]. https://doi.org/10.48550/arXiv.2111.05849

[19] GAO K, GAO Y N, HE H J, et al. NeRF: neural radiance field in 3D vision, A comprehensive review [EB/OL]. [2023-02-25]. https://arxiv. org/abs/2210.00379

[20] REBAIN D, JIANG W, YAZDANI S, et al. DeRF: decomposed radiance fields [EB/OL]. [2023-02-25]. https://arxiv.org/abs/2011.12490

[21] REISER C, PENG S Y, LIAO Y Y, et al. KiloNeRF: speeding up neural radiance fields with thousands of tiny MLPs [C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2022: 14315–14325. DOI: 10.1109/ICCV48922.2021.01407

[22] MÜLLER T, EVANS A, SCHIED C, et al. Instant neural graphics primitives with a multiresolution hash encoding [J]. ACM transactions on graphics, 2022, 41(4): 1 – 15. DOI: 10.1145/3528223.3530127

[23] LIU L J, GU J T, LIN K Z, et al. Neural sparse voxel fields [EB/OL]. [2023-02-25]. https://arxiv.org/abs/2007.11571

[24] PIALA M, CLARK R. TermiNeRF: ray termination prediction for efficient neural rendering [C]//Proceedings of 2021 International Conference on 3D Vision (3DV). IEEE, 2022: 1106–1114. DOI: 10.1109/3DV53792.2021.00118

[25] FANG J M, XIE L X, WANG X G, et al. NeuSample: neural sample field for efficient view synthesis [EB/OL]. [2023-02-25]. https://arxiv. org/abs/2111.15552

[26] NEFF T, STADLBAUER P, PARGER M, et al. DONeRF: towards real-time rendering of compact neural radiance fields using depth oracle networks [J]. Computer graphics forum, 2021, 40(4): 45–59. DOI: 10.1111/cgf.14340

[27] HU T, LIU S, CHEN Y L, et al. EfficientNeRF – efficient neural radiance fields [C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 12892–12901. DOI: 10.1109/CVPR52688.2022.01256
[28] GARBIN S J, KOWALSKI M, JOHNSON M, et al. FastNeRF: high–fidelity neural rendering at 200FPS [C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2022: 14326–14335. DOI: 10.1109/ICCV48922.2021.01408
[29] WADHWANI K, KOJIMA T. SqueezeNeRF: further factorized FastNeRF for memory–efficient inference [EB/OL]. [2023–02–25]. https://arxiv. org/abs/2204.02585
[30] YU A, LI R L, TANCIK M, et al. PlenOctrees for real–time rendering of neural radiance fields [C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2022: 5732–5741. DOI: 10.1109/ICCV48922.2021.00570
[31] HEDMAN P, SRINIVASAN P P, MILDENHALL B, et al. Baking neural radiance fields for real–time view synthesis [EB/OL]. [2023–02–25]. https://arxiv.org/abs/2103.14645
[32] RAO C L, YU H J, WAN H C, et al. ICARUS: a specialized architecture for neural radiance fields rendering [EB/OL]. [2023–02–25]. https://arxiv. org/abs/2203.01414
[33] DENG K L, LIU A, ZHU J Y, et al. Depth–supervised NeRF: fewer views and faster training for free [EB/OL]. [2023–02–25]. https://arxiv. org/abs/2107.02791
[34] CHEN T S, DU Z D, SUN N H, et al. DianNao: a small–footprint high–throughput accelerator for ubiquitous machine–learning [C]//Proceedings of the 19th international conference on Architectural support for programming languages and operating systems. ACM, 2014: 42(1): 269 – 284. DOI: 10.1145/2541940.2541967

作　者　简　介

郑清芳，中兴通讯股份有限公司云视频首席科学家；主要从事视频领域的技术规划及研发工作，主要研究兴趣包括沉浸式视频内容生成、低时延视频传输、视频编解码、视频智能分析等。

中兴通讯技术杂志社
促进产学研合作青年专家委员会

主　任　陈　为（北京交通大学）

副主任　秦晓琦（北京邮电大学）　卢　丹（中兴通讯技术杂志社）

委　员　（按姓名拼音排序）

曹　进　西安电子科技大学
陈　力　中国科学技术大学
陈琪美　武汉大学
陈舒怡　哈尔滨工业大学
陈　为　北京交通大学
官　科　北京交通大学
韩凯峰　中国信息通信研究院
何　姿　南京理工大学
胡　杰　电子科技大学
黄　晨　紫金山实验室
李　昂　西安交通大学
刘春森　复旦大学
刘　凡　南方科技大学
刘俊宇　西安电子科技大学
卢　丹　中兴通讯技术杂志社
陆游游　清华大学
宁兆龙　重庆邮电大学
祁　亮　上海交通大学
秦晓琦　北京邮电大学
秦志金　清华大学
史颖欢　南京大学
王景璟　北京航空航天大学
王兴刚　华中科技大学
王勇强　天津大学
温淼文　华南理工大学
吴泳澎　上海交通大学
夏文超　南京邮电大学
徐梦炜　北京邮电大学
徐天衡　中国科学院上海高等研究院
杨川川　北京大学
尹海帆　华中科技大学
于季弘　北京理工大学
张　娇　北京邮电大学
张宇超　北京邮电大学
章嘉懿　北京交通大学
赵昱达　浙江大学
周　伊　西南交通大学
朱秉诚　东南大学

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

产业顾问（按姓名拼音排序）：

段向阳、高　音、胡留军、华新海、刘新阳、
陆　平、史伟强、屠要峰、王会涛、熊先奎、
赵亚军、赵志勇、朱晓光

双月刊　1995 年创刊　总第 169 期
2023 年 4 月　第 29 卷 第 2 期

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

总编辑：王喜瑜
主编：蒋贤骏
执行主编：黄新明
编辑部主任：卢丹
责任编辑：徐烨
编辑：杨广西、朱莉、任溪溪
设计排版：徐莹
发行：王萍萍
编务：王坤

《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：2023 年 4 月 15 日
中国标准连续出版物号：ISSN 1009-6868
CN 34-1228/TN
定价：每册 20.00 元