

## 面向语义信息直传的通信架构

石光明<sup>1,2</sup>, 杨旻曦<sup>1,3</sup>, 高大化<sup>1,3</sup>, 柴靖轩<sup>1,3</sup>

(1. 西安电子科技大学人工智能学院, 陕西 西安 710071; 2. 鹏城实验室, 广东 深圳 518055; 3. 琶洲实验室, 广东 广州 510555)

**摘要:** 随着聚焦提升带宽和频谱效率的传统通信发展模式渐入瓶颈, 越来越多的研究将智能通信的目标从语法层转向语义层, 通过感知并传输语义而非完整信号来节省带宽资源。信宿端智能体只需接收可理解语义中有信息部分即可。若能在信源端从语义角度中甄别出对信宿端有信息的部分加以传输, 将进一步降低带宽资源和信宿端语义信息处理的时间和功耗。为此, 首先探讨了智能体的语义信息处理和理解的过程, 并将信宿从信息中感知到的语义划分为冗余语义、难以理解语义(暗语义)和有信息的语义(语信); 接着, 提出了面向传输语义中有信息部分的通信范式——语信通信, 并将通信范式划分为语法、语义、语信、语用四层; 最后, 通过仿真实验验证了语信通信的可行性和有效性。这为下一代通信范式的发展提供了新思路和技术牵引。

**关键词:** 下一代通信; 语义通信; 语义; 信息; 通信范式分级

**中图分类号:** TN92

**文献标志码:** A

**DOI:** 10.11959/j.issn.1000-436x.2023098

## Communication framework for directed transmission of informative semantic

SHI Guangming<sup>1,2</sup>, YANG Minxi<sup>1,3</sup>, GAO Dahua<sup>1,3</sup>, CHAI Jingxuan<sup>1,3</sup>

1. School of Artificial Intelligence, Xidian University, Xi'an 710071, China

2. Peng Cheng Laboratory, Shenzhen 518055, China

3. Pazhou Lab, Guangzhou 510555, China

**Abstract:** As the traditional communication development model focusing on enhancing bandwidth and spectrum efficiency is getting bottlenecked, more and more research is shifting the goal of intelligent communication from the syntactic level to the semantic level to save bandwidth resources by sensing and transmitting semantics rather than the complete signal. For the receiver agent, it is enough to receive only the information part of the understandable semantics. If the informative parts from the semantics can be filtered for transmission, it will further reduce the bandwidth resources and the time and power consumption of semantic information processing at the sender-side. To this end, firstly, the process of semantic information processing and comprehension of an intelligent body was explored. Secondly, the semantics perceived by the receiver from the message were classified into redundant semantics, unintelligible semantics, and informative semantic. Then, a communication paradigm oriented to transmitting the informative part of the semantics, called informative communication, was proposed, and the communication paradigm was extended to include four layers: syntax, semantic, informative, and pragmatic. Finally, the feasibility and effectiveness of informative communication were verified through simulation. This provides new ideas and technical traction for the development of next-generation communication models.

**Keywords:** next-generation communication, semantic communication, semantic, information, communications paradigm hierarchy

收稿日期: 2023-02-05; 修回日期: 2023-05-05

通信作者: 高大化, dhgao@xidian.edu.cn

基金项目: 国家重点研发计划基金资助项目(No.2019YFA0706604); 国家自然科学基金资助项目(No.61976169, No.62293483); 广州市场景理解与智能交互重点实验室资助项目(No.202201000001)

**Foundation Items:** The National Key Research and Development Program of China (No.2019YFA0706604), The National Natural Science Foundation of China (No.61976169, No.62293483), The Guangzhou Key Laboratory of Scene Understanding and Intelligent Interaction Program (No.202201000001)

## 0 引言

在 Shannon<sup>[1]</sup>提出信息论后的几十年内, 通信领域致力于通过提升带宽和频谱效率增加传输速率, 研发了 1G 到 5G 的多代无线通信技术<sup>[2]</sup>, 同时占用的带宽与频谱越来越高、功耗不断增加。更进一步, 当前的频谱效率已逼近信息论给出的上限<sup>[3]</sup>, 限制了通信技术的进一步发展。针对这一现状, 随着通信终端算力和人工智能技术的发展, 通信技术开始由语法层向语义层转变<sup>[4]</sup>, 即语义通信<sup>[5-7]</sup>。语义通信以语义符号的形式表征、传输信息, 具有更高的信息压缩率、更强的抗噪声干扰能力, 并更易于信宿理解。

近年来, 随着深度学习技术的发展, 基于深度学习的语义编译码系统逐渐成为语义通信的一个热点研究课题。受到联合信源信道编码思想的启发, 很多研究借助深度学习对复杂信号(如图像等)的强大特征提取能力, 直接端到端地提取和编码信号中的语义。Xie 等<sup>[8]</sup>提出一种基于 Transformer 的语义通信系统 DeepSC 来对文本进行传输。基于 DeepSC 框架, Weng 等<sup>[9]</sup>提出了端到端语义通信系统 DeepSC-S, 实现了语义通信从文本模态到语音模态的拓展。Yoo 等<sup>[10]</sup>则针对图像提出一种基于 ViT (vision transformer) 的端到端语义通信系统。此外, 还有 Li 等<sup>[11]</sup>提出一种基于生成对抗网络(GAN)的语义编码方法, Huang 等<sup>[12]</sup>提出一种使用草图、分割图等表示图像的编码方法。此类方法可直接利用端到端的深度学习将语义转化为隐向量, 因而在信息压缩和抗噪声方面取得了优于传统方法的性能。现有基于端到端深度学习的语义编译码方法普遍使用峰值信噪比 (PSNR, peak signal-to-noise ratio) 和双语替换评测 (BLEU, bilingual evaluation understudy) 等传统通信的保真评估标准作为优化目标, 以实现消息的完全重构。无论是传统通信还是现有语义编译码, 它们都只需保证消息的准确传递, 而将信息处理的任务交给了信宿端的智能体。这意味着它们会将不含智能体所需语义信息的消息也通过信道传递。这一方面消耗了更多带宽资源, 另一方面可能导致智能体需要更多时间和功耗从包含冗余信息的消息中处理所需语义信息。众所周知, 通信的目的是传递信息。对于信宿端智能体而言, 其只需要消息中有信息的部分, 不需要事先已知和难以理解的部分, 若能从消息中甄别有

信息的部分加以传输, 将进一步降低带宽资源和信宿端语义信息处理的时间和功耗。

为了提升语义信息的传输效率, 首先需要确定语义信息的定义。1928 年, Hartley<sup>[13]</sup>提出“信息是指有新内容、新知识的消息”, 之后有更多不同的信息定义被提出。1949 年, Shannon<sup>[1]</sup>将信息定义为“信息是用以消除随机不确定性的东西”。1957 年, Richard 和 Ashby<sup>[14]</sup>提出“信息是集合的变异度”, 认为信息的本性在于事物本身具有变异度。1975 年, Longo<sup>[15]</sup>提出“信息是反映事物构成、关系和差别的东西, 它包含在事物的差异之中, 而不在事物的本身”。1988 年, 钟义信<sup>[16]</sup>将信息定义为“事物运动的状态以及它的状态改变方式, 是物质的一种属性”。这些信息定义都聚焦客观事物的变化与差异, 有意回避了信宿的主观感受。不可否认, 这在面向消息保真的传统通信中是正确与有效的。

在智能通信中, 信源和信宿都是智能体, 拥有比传统通信节点更强的计算能力和知识库驱动的记忆能力。此时, 讨论信息必须考虑主观的发送意图和理解需求。对于同一消息, 不同知识背景和处境的人感受到的信息很可能不同。为了研究如何根据智能体的主观理解程度来客观地划分语义, 本文首先提出实景虚景事物及信号、语义特征函数等概念来客观地探讨智能体的语义信息处理过程, 具体为: 面对客观的实景事物, 智能体先用传感器得到图像等实景信号; 再从信号中感知表征语义的实景语义特征函数; 然后根据实景语义特征函数建立起实景事物在信宿主观中对应的虚像, 即虚景事物; 智能体可进一步使用语言符号等虚景语义特征函数描述虚景事物, 并通过声带等器官将其转化为语音等虚景信号输出; 虚景信号也可以借助虚景语义特征函数参与虚景事物的构建。在此基础上, 本文发现信宿从消息中感知到的语义可根据其与信宿理解事物的符合程度分为 3 个部分, 分别为符合程度过高的冗余语义(即信宿事先已知的内容)、符合程度过低而难以理解的暗语义(即无法解码其含义的内容)和符合程度适中的有信息的语义(即接收前不知道、接收后能够理解的部分内容, 简称语信)。相比于传输全部语义, 传输不包含冗余和难理解部分的语信能够节省带宽和智能体处理的时间和功耗。正如语义通信通过语义编译码筛选信号中的内容是语法通信在语义层面上的扩展一样, 语

信通信面向理解对语义进行筛选处理，可看作语义通信在信息层面上的扩展。为此，本文提出根据目的将通信划分为 4 个层级，并设计了相应的框架。最后通过仿真实验验证了本文所提语信通信的可行性和有效性。

本文的主要贡献总结如下。

1) 提出实景虚景事物及信号、语义特征函数等概念，探讨了智能体的语义信息处理过程，为在语义的范畴下客观地讨论信宿的主观过程这一难题提供了一种解决思路。

2) 基于虚景事物探讨了消息中语义理解的过程，并将信宿从信息中感知到的语义划分为冗余语义、无法理解语义和有信息的语义三部分，揭示了在语义通信基础上进一步节省通信带宽和智能体信息处理所需时间和功耗的发展方向。

3) 提出面向传输语义中有信息的部分的通信范式，即语信通信，将通信范式的三层划分拓展为语法、语义、语信、语用四层，为下一代通信范式的发展开辟了新的赛道。

4) 通过仿真实验验证了本文所提语信通信的可行性和有效性，为下一代通信技术的落地应用提供了牵引。

## 1 语义信息处理和理解过程

Shannon 等基于符号概率建立的信息论假设了信源信宿无记忆且通信符号独立不相关。这决定了

在此基础上构建的传统通信范式只追求信号的无失真传输而不需要也无力关注信号中蕴含的语义。事实上，以 Shannon 为代表的学者们早就意识到提取并传输信号中的信息可以有效提升效率，并曾提出“通信有三层，分别是语法层、语义层和语用层”。在当时的条件下，由于缺乏对语义的客观刻画，也缺乏从信号中提取语义的机器，而且当时的通信几乎都是面向有完善语义信息处理能力的人类，因此传统通信技术不涉及语义处理。

经过人工智能技术十余年的高速发展，以工业机器人和智能手机为代表的通信节点可以拥有较强的计算和记忆能力，并催生出工业互联网、物联网等新一代通信应用场景。在这些场景中，一方面，通信节点的通信特点是信源具有较强的非线性信号处理能力和记忆能力，且可根据场景变化动态学习和更新；另一方面，通信需要传输的信息量大，信号模式多样。在新的应用场景下，若要有效发挥智能通信节点的信息处理能力实现面向语义信息的通信，首先需要对语义、信息等关键概念进行刻画。为此，本节提出实景虚景事物及信号、语义特征函数等概念来客观表示信宿的主观性，分别探讨语义信息的处理和理解过程，实现对语义、信息等关键概念的刻画。

### 1.1 语义信息处理过程

本文认为智能体的语义信息处理过程如图 1 所示。其具体过程如下。

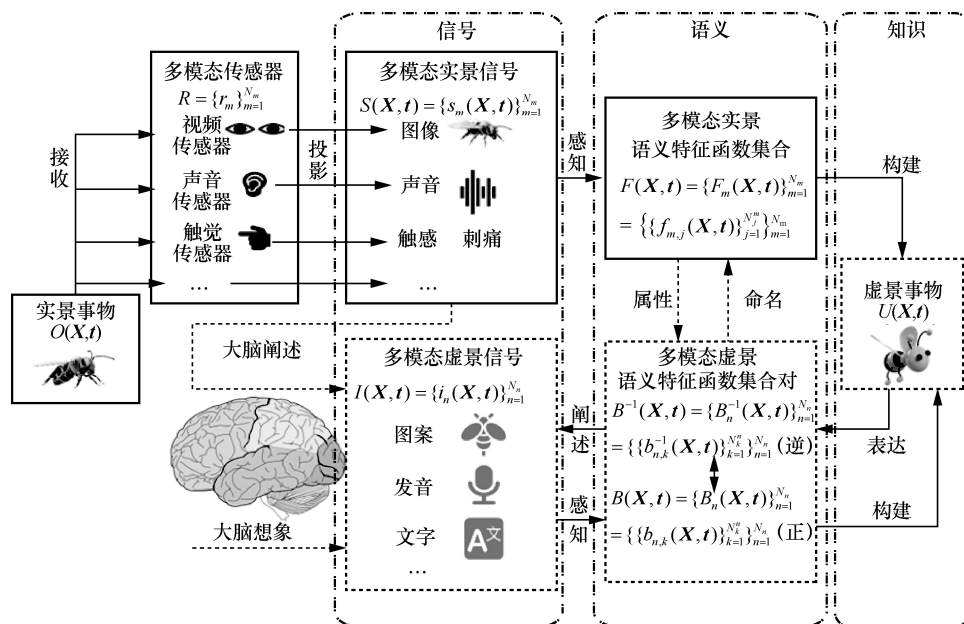


图 1 语义信息处理过程

1) 面对客观的实景事物, 智能体用传感器得到图像等实景信号。

几乎所有的通信场景针对的都是现实世界中客观存在的事物, 本文称这类客观存在的事物为实景事物。这些事物可能因时间或空间状态不同表现出不同的形态, 故本文将其表示为空间状态变量  $\mathbf{X}$  和时间状态变量  $t$  的函数  $O(\mathbf{X}, t)$ 。

智能体利用其  $N_m$  种模态的传感器  $R = \{r_m\}_{m=1}^{N_m}$ , 从实景事物中获取的多模态实景信号  $S(\mathbf{X}, t) = \{s_m(\mathbf{X}, t)\}_{m=1}^{N_m}$  也是实际客观且不以主观意志为转移的, 故本文称之为实景信号。智能体通过接收并投影信号获得对实景事物时空状态的了解。

2) 从信号感知表征语义的实景语义特征函数。

在语义表示方面的早期研究中, 受限于计算能力, 复杂信号的刻画主要依赖于手工特征的设计, 如尺度不变特征转换 (SIFT, scale-invariant feature transform)<sup>[17]</sup>和基于部件的模板匹配方法<sup>[18]</sup>。但是仅使用简单映射或变换难以跨越特征与语义间的鸿沟, 导致这类方法表示出来的语义泛化性差, 无法应用于多变的通信场景中。近年来, 学者利用神经网络提出了具有可解释性和更高灵活的语义结构信息表示方法。例如, 使用决策树或者多层感知机将语义概念组合的语义瓶颈模型<sup>[19]</sup>、使用深度学习多种典型部件的原型网络<sup>[20]</sup>, 以及基于语义先验构建树形结构<sup>[21]</sup>或者一组流形<sup>[22]</sup>。

本文沿用文献[23]中提出的多模态特征函数来表征语义。语义特征函数为智能体从信号中感知出语义所依据的信号特征, 是用虚景事物的特征信号标注或命名实景事物的特征信号, 且该标注需被群体所承认。智能体若要从一种模态的信号中感知一种语义就需要至少一个对应的语义特征函数。通过确定语义特征函数, 就可以相对客观地描述智能体语义感知这一主观过程。为了便于不同智能体之间进行通信, 尤其是有人类参与的通信, 应确保语义特征函数具有可解释性。本文沿用文献[23]中提出的选择有代表性的信号作为语义特征函数, 并通过信号匹配的方式实现语义感知。本文将  $N_m$  种模态的实景信号对应的语义特征函数称为多模态实景语义特征函数, 记为  $F(\mathbf{X}, t) = \{F_m(\mathbf{X}, t)\}_{m=1}^{N_m}$ , 其中, 每个模态包含了  $N_j^m$  个语义特征函数, 即  $F_m(\mathbf{X}, t) = \{f_{m,j}(\mathbf{X}, t)\}_{j=1}^{N_j^m}$ 。在模

态  $m$  下, 根据实景语义特征函数从实景信号中感知语义的过程记为

$$z(\mathbf{X}, t) = F_m(\mathbf{X}, t, s_m(\mathbf{X}, t)) \quad (1)$$

其中,  $z(\mathbf{X}, t)$  表示感知出的语义。语义是模态无关的, 不同模态的语义特征函数对应相同的语义。

3) 根据实景语义特征函数构建虚景事物。

本文将实景事物在智能体主观中对应的虚像称为虚景事物。虚景事物是智能体借助从信号中感知到的语义, 总结出的实景事物的规律。本文认为这样的规律就是知识的一种表征形式。在本文中, 知识狭义地指代虚景事物。本文将虚景事物表征为实景事物可感知出的语义  $z(\mathbf{X}, t)$  的概率模型, 并可用极大似然估计法表示为

$$U(\mathbf{X}, t)^* = \arg \max_U \mathcal{L}(U(\mathbf{X}, t) | \{z(\mathbf{X}, t)\}) \quad (2)$$

其中,  $U(\mathbf{X}, t)$  为虚景事物,  $\{z(\mathbf{X}, t)\}$  为构建虚景事物用到的语义集合,  $\mathcal{L}$  为似然函数,  $U(\mathbf{X}, t)^*$  为给定  $\{z(\mathbf{X}, t)\}$  后用极大似然估计求得的最优虚景事物。

4) 使用虚景语义特征函数描述虚景事物, 并转化为虚景信号输出。

与实景信号相对应, 本文将大脑阐述或想象出的信号称为虚景信号。虚景信号的本源是虚景事物, 智能体先用虚景语义特征函数描述虚景事物, 再通过声带等器官转化为虚景信号输出。虚景语义特征函数用于从虚景信号中感知语义, 可表示为

$$z(\mathbf{X}, t) = B_n(\mathbf{X}, t, i_n(\mathbf{X}, t)) \quad (3)$$

其中,  $i_n$  为虚景信号,  $B_n$  为虚景语义特征函数。不同于实景语义特征函数, 智能体在确定虚景语义特征函数时, 往往还会确定其逆函数, 用于根据语义输出虚景信号

$$i_n(\mathbf{X}, t) = B_n^{-1}(\mathbf{X}, t, z(\mathbf{X}, t)) \quad (4)$$

其中,  $B_n^{-1}$  为逆虚景语义特征函数。有了正、逆虚景语义特征函数, 智能体间才能实现双向通信。需要说明的是, 从虚景信号和实景信号感知来的语义具有同样的效力, 都可以参与构建虚景事物。

语义信息处理相关概念的符号和含义对照如表 1 所示。

表 1 语义信息处理相关概念的符号含义对照

符号	含义
$X$	空间状态变量
$t$	时间状态变量
$O(X, t)$	实景事物，即客观存在的事物
$r_m$	模态 $m$ 对应的传感器
$N_m$	信号模态总数
$R$	多模态传感器集合
$s_m(X, t)$	模态 $m$ 的实景信号，即利用 $r_m$ 从实景事物中获取的信号
$S(X, t)$	多模态实景信号集合
$f_{m,j}(X, t)$	模态 $m$ 的第 $j$ 个实景语义特征函数，其中语义特征函数定义见文献[23]
$N_j^m$	模态 $m$ 下实景语义特征函数总数
$F_m(X, t)$	模态 $m$ 下实景语义特征函数集合
$F(X, t)$	多模态实景语义特征函数集合
$z(X, t)$	信号蕴含的语义，其详细描述和刻画可见文献[23]
$U(X, t)$	虚景事物，实景事物在智能体主观中对应的虚像
$b_{n,k}(X, t)$	模态 $n$ 下第 $k$ 个虚景语义特征函数
$N_n$	模态 $n$ 下虚景语义特征函数总数
$B_n(X, t)$	模态 $n$ 下虚景语义特征函数集合
$B(X, t)$	多模态虚景语义特征函数集合
$i_n(X, t)$	模态 $n$ 的虚景信号
$I(X, t)$	多模态虚景信号集合
$B^{-1}(X, t)$	多模态逆虚景语义特征函数集合

1.2 语义信息理解过程

在语义信息处理的基础上，语义信息理解的过程如图 2 所示。其具体过程如下。

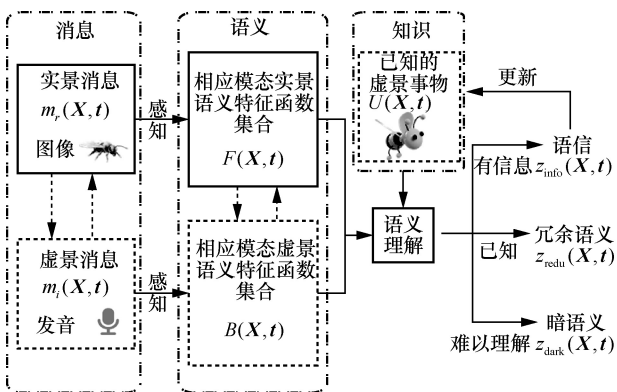


图 2 语义信息理解过程

1) 智能体接收到实景消息或虚景消息。

由于消息需要在信道上传输，因此在语义范畴下的消息是包含语义的信号。实景消息是实景信号；虚景消息是虚景信号。由于实景消息和虚景消息都可以携带语义，故智能体接收到实景消息或者虚景消息都可以进行语义信息理解。本文将实景消息和虚景消息分为记为  $m_r(X, t)$  和  $m_i(X, t)$ 。

2) 用相应的语义特征函数从消息中感知语义。

智能体使用与消息相应的实景语义特征函数  $F(X, t)$  或虚景语义特征函数  $B(X, t)$  从消息中感知语义。语义感知过程详见第 1.1 节中的 2) 和 4)。

3) 由已知虚景事物对感知所得语义进行理解。

利用概率模型表征的虚景事物，可计算感知所得语义的概率为

$$p = U(X, t, z(X, t)) \quad (5)$$

其中， $U(X, t)$  为智能体在接收到消息之前就已知虚景事物， $z(X, t)$  为智能体从消息中感知到的语义， $p$  为根据知识评估语义的概率。 $p$  反映了接收消息中所含语义与已知虚景事物之间的符合程度。本文认为可通过设定 2 个阈值  $\epsilon_{\text{known}}$  和  $\epsilon_{\text{dark}}$  (满足  $\epsilon_{\text{dark}} < \epsilon_{\text{known}}$ ) 将语义分为三类： $p$  较高，即  $p \geq \epsilon_{\text{known}}$ ，代表该语义与知识相符，可认为该语义为已知的，本文称其为冗余语义  $z_{\text{redu}}(X, t)$ ； $p$  较低，即  $p \leq \epsilon_{\text{dark}}$ ，代表该语义与知识相悖，可认为该语义为难以理解、难以接受的，本文称其为暗语义  $z_{\text{dark}}(X, t)$ ； $p$  适中，即  $\epsilon_{\text{dark}} < p < \epsilon_{\text{known}}$ ，则可认为该语义有信息，本文称其为语信，记为  $z_{\text{info}}(X, t)$ 。

4) 根据语信更新虚景事物。

智能体将语信作为语义样本使用极大似然估计法求得更新过的虚景事物，如

$$U' = \arg \max_U \mathcal{L}(U | \{z\} \cup \{z_{\text{info}}\}) \quad (6)$$

其中，本文为了简洁隐去了所有变量的时空状态  $(X, t)$ ， $U$  为虚景事物， $\{z\}$  为接收消息之前已知的语义集合， $\{z_{\text{info}}\}$  为接收到的消息中的语信集合， $U'$  为更新后的虚景事物。

语义信息理解相关概念的符号和含义对照如表 2 所示。

表 2 语义信息理解相关概念的符号含义对照

符号	含义
$m_r(X, t)$	实景消息
$m_i(X, t)$	虚景消息
$F(X, t)$	多模态实景语义特征函数集合
$B(X, t)$	多模态虚景语义特征函数集合
$p$	根据已知虚景事物评估语义的概率, 反映了接收消息中所含语义与已知虚景事物之间的符合程度
$\epsilon_{\text{known}}$	$p$ 的高阈值
$\epsilon_{\text{dark}}$	$p$ 的低阈值
$z_{\text{redu}}(X, t)$	冗余语义, 满足 $p \geq \epsilon_{\text{known}}$
$z_{\text{dark}}(X, t)$	暗语义, 满足 $p \leq \epsilon_{\text{dark}}$
$z_{\text{info}}(X, t)$	有信息的语义, 简称“语信”, 满足 $\epsilon_{\text{dark}} < p < \epsilon_{\text{known}}$
$U'(X, t)$	根据语信更新后的虚景事物

## 2 通信范式分级

本文提出将通信范式划分为从低到高的语法、语义、语信、语用这 4 个层级。通信范式的 4 个层级的框架如图 3 所示。其中加粗部分表示与上一级

通信范式不同的地方。通信 4 个层级不是独立的, 而是如图 4 所示的嵌套关系。语义通信在语法通信的基础上, 先基于语义感知对信号进行语义层面的编码, 再通过语法层面的信源编译码与传输实现语义传输。本文提出的语信通信则是在语义通信的基础上, 基于动态语义知识库从语义感知的结果中筛选出有信息的部分。语用通信则可在语信通信的基础上, 聚焦重现信号的下游应用。本节接下来将对语义通信和语信通信展开介绍。

### 2.1 语义通信

近年来, 利用智能通信节点实现面向语义的通信范式, 即语义通信, 逐渐成为研究热点。

语义通信的目的是通过保真地传输信号中的语义, 让信宿端获取信源端从信号中提取到的语义, 而不在于信号是否完全无失真, 即

$$\min \mathcal{E}(z, z') \quad (7)$$

其中,  $z$  为信源端从信号中感知的语义,  $z'$  为信宿端重现的语义。

相比于语法通信, 语义通信由于避免了信号保真传输, 转化为语义符号传输, 故传输相同语义时,

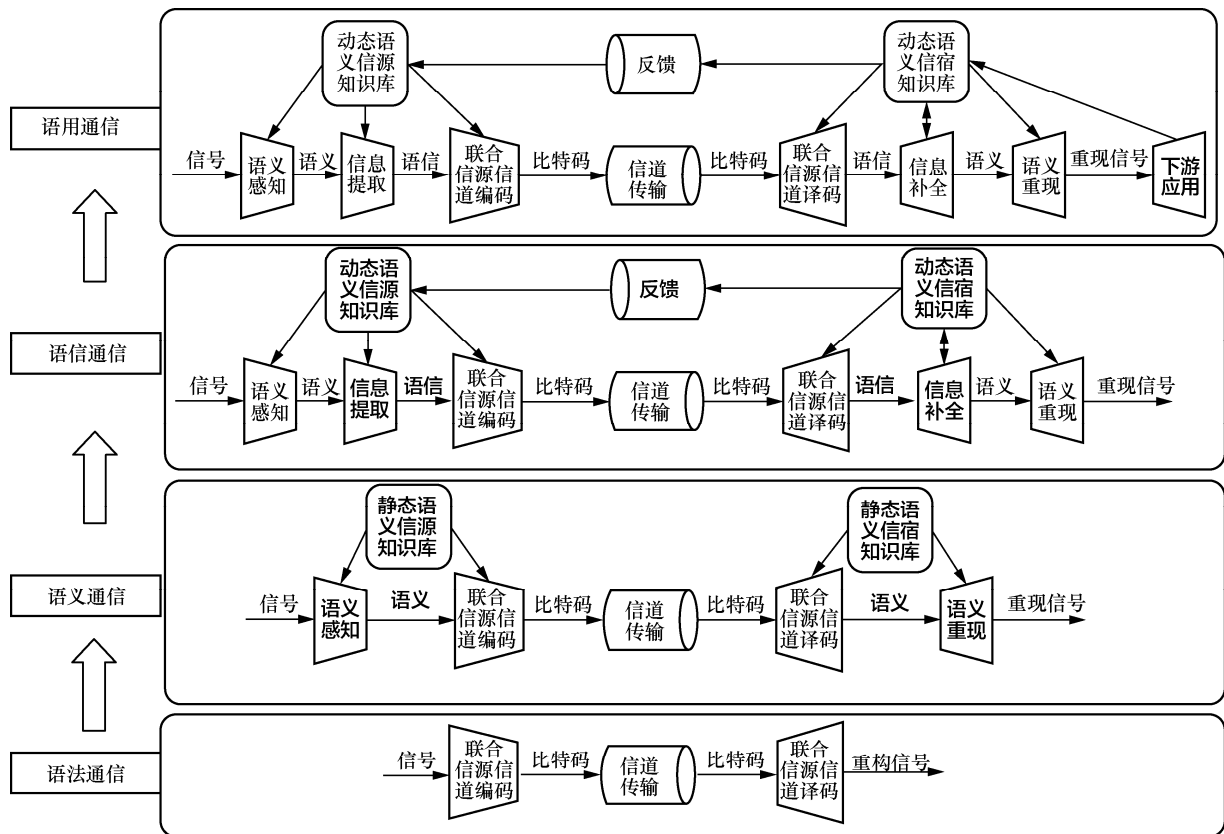


图 3 通信范式的 4 个层级的框架

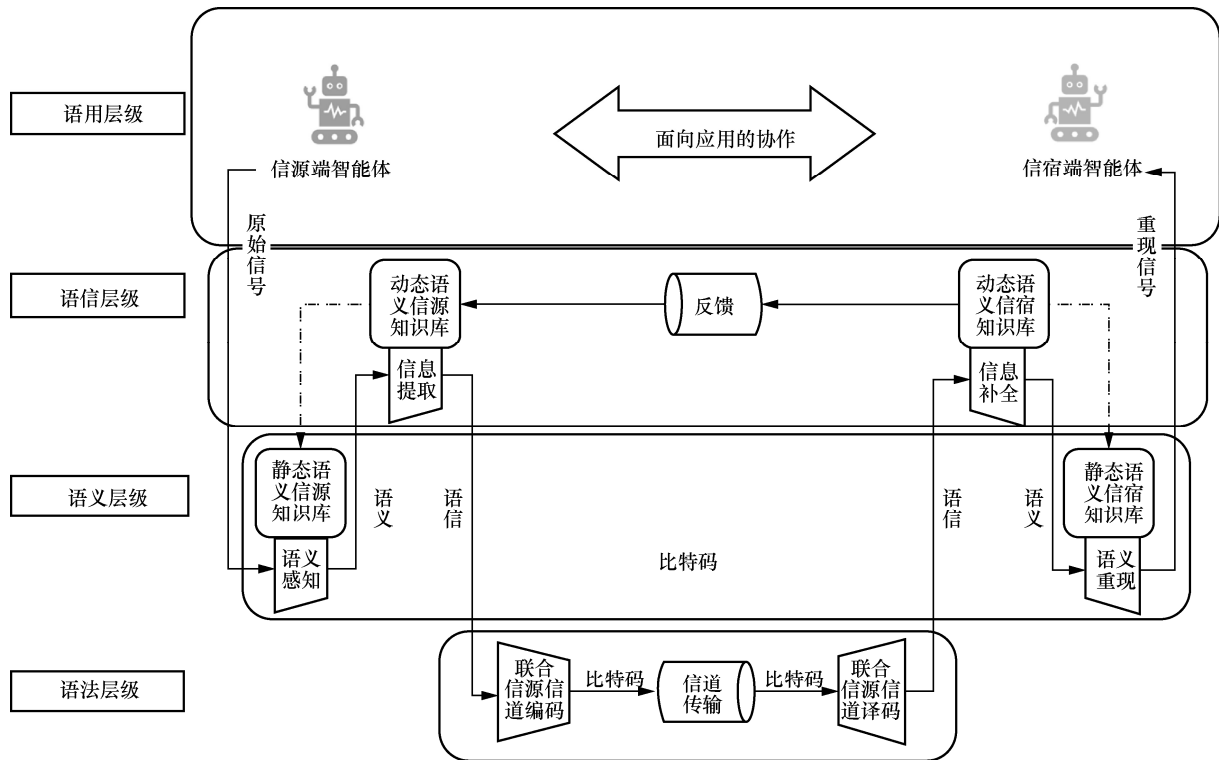


图 4 通信范式的 4 个层级之间的嵌套关系

语义通信所需通信带宽低于语法通信。

如图 3 所示，相比于语法通信，语义通信将信源编/译码改进为静态语义知识库指导下的语义感知/重现和信源信道编/译码。其中，静态语义知识库在通信前构建好，不会在通信过程中更新。在语义通信中，一般采用可优化神经网络实现的语义感知、信源信道编/译码和语义重现，而静态语义知识库则包含最优化的语义感知、语义重现、信源编码和信源译码的参数。使用静态语义知识库而非动态语义知识库的原因是对处理图像、视频等复杂信号的语义感知重现模型普遍存在模型参数量大、训练所需数据量大且存在灾难性遗忘问题，一旦作为本地知识库部署到各个通信节点，想要在通信中通过反馈进行更新的代价非常高。语义通信过程如算法 1 所示。

**算法 1** 语义通信过程

给定输入信号、静态语义信源和信宿知识库。

- 1) 根据静态语义信源知识库中的最优化参数，按照式(1)或式(3)从信号中感知语义；
- 2) 根据静态语义信源知识库中的最优化参数，使用编码器将信号编码为表达语义的比特码流；
- 3) 由信道将比特码从信源端输出至信宿端；
- 4) 根据静态语义信宿知识库中的最优化参数，

使用译码器从接收比特码恢复语义；

- 5) 根据静态语义信宿知识库中的最优化参数，按照式(4)根据语义重现信号。

**2.2 语信通信**

语法通信和语义通信实现了间接或直接地传输语义，其本质是信源端向信宿端分享所见所感。在人与人交流中，这种场景往往是为了闲聊以满足社交的心理需求。闲聊对机器间通信没有意义，更有意义的是传授知识。对于大型智能通信网络而言，由于并网早晚或分工不同，智能通信节点普遍存在知识差异，以传授知识为导向的通信有显著意义。语信通信是指传输信宿事先不知且能理解的信息。

为此，语信通信的目的是通过保真地传输有信息的语义（语信），让信宿端与信源端的知识差异尽可能缩小，即

$$\max D(U_s, U_r) - D(U_s, U_r') \quad (8)$$

其中， $D(\cdot, \cdot)$  为概率分布间的距离度量， $U_s$  为信源的知识， $U_r$  和  $U_r'$  分别为通信前后信宿的知识。

相比于语义通信，语信通信由于避免了传输语义中冗余部分和难理解的部分，语信通信所需信道带宽和信宿端进行语义信息处理的功耗和时间均低于语义通信。此外，语信通信需要进行知识库更

新，其对变化的通信场景和环境具有更强的适应能力。得益于静态语义知识库指导的语义感知和重现实现了信号与通用语义间的转化，相较于原始信号，语义的复杂度已经大大降低。在此基础上，只需轻量化模型，便可根据信源信宿双方的情况对通用语义进行信息提取与补全，且在算力和信道资源有限的情况下也能实现动态更新。

如图 3 所示，相比于语义通信，语信通信将静态语义知识库改进为动态语义知识库，并在语义感知后进行信息提取，在语义重现前进行信息补全，用来实现语义和语信间的转换。此外，语信通信还需要在信息补全过程利用接收到的语信更新动态语义信宿知识库，并引入带有编译码功能的反馈通道用于将信宿的语义理解程度反馈给信源。动态语义知识库除了需要包含静态最优化的语义感知、语义重现、信源编码和信源译码的参数，还需要包括动态的虚景事物用于提取和补全语信。其通信过程如算法 2 所示。

**算法 2 语信通信过程**

给定输入信号、动态语义信源和信宿知识库中的最优化参数，以及信源和信宿的初始虚景事物。

- 1) 根据动态语义信源知识库中的最优化参数，按照式(1)或式(3)从信号中感知语义；
- 2) 根据动态语义信源知识库中的虚景事物和式(5)表述的  $p$  提取语义中的语信；
- 3) 根据动态语义信源知识库中的最优化参数，使用编码器将信号编码为表达语信的比特码；
- 4) 由信道将比特码从信源端输出至信宿端；
- 5) 根据动态语义信宿知识库中的最优化参数，使用译码器从接收比特码恢复语信；
- 6) 按照式(6)根据接收语信更新动态语义信宿知识库中的虚景事物；

7) 根据虚景事物采样若干语义作为补全语义，从而实现信息补全得到语义；

8) 根据动态语义信宿知识库中的最优化参数，按照式(4)根据语义重现信号；

9) 信宿端将用于信息补全的其他语义通过带有编译码功能的反馈通道传递给信源端；

10) 信宿端判断补全语义是否为信息提取时的冗余语义，若是，则结束通信；反之，则在更新动态语义信源知识库后，将补全语义中缺失的部分传输给信宿端，直到反馈的补全语义完整为止。

**3 面向图像理解的语信通信仿真分析**

本文节基于面向图像理解的通信任务，通过实验验证本文所提语信通信的可行性和有效性。

**3.1 图像数据集**

本节实验采用视觉关系识别数据集 Visual Genome<sup>[24]</sup>，其中包含图像及其对应的由对象、属性和关系组成的语义场景图。本文采用了与文献[25]相同的数据集设置，同样筛选出其中频率最大的 150 类对象和 50 类关系展开训练和测试。

**3.2 语义相似度**

为了客观地度量人类视角下不同图像间的语义相似度，本文使用在 COCO 数据集<sup>[26]</sup>上训练好的模型从图像中生成文本描述<sup>[27]</sup>，再将文本在开源的 BERT<sup>[28]</sup>嵌入空间中的内积作为语义相似度。本文使用语信通信的原始图像与重现图像之间的语义相似度作为语信通信性能评估标准。

**3.3 语信通信架构实现细节**

根据语信通信架构，本文设计了如图 5 所示的面向图像理解的语信通信系统。其实现细节如下。

1) 语义感知

本文使用联合训练的神经网络模型作为式(1)

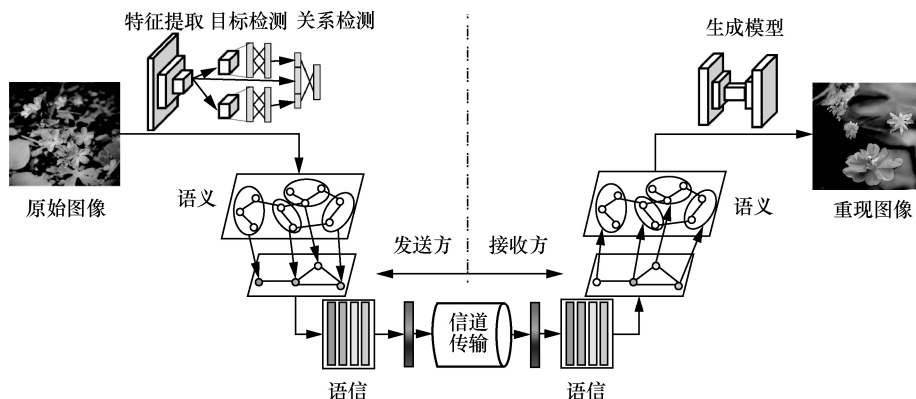


图 5 面向图像理解的语信通信系统

所述的实景语义特征函数，从图像模态的实景信号中感知实景语义。具体过程借鉴视觉关系检测网络 Scene Graph Benchmark<sup>[29]</sup>，首先利用主干网络从图像中提取特征，再依次通过目标检测和关系检测网络提取带属性的对象和对象间的关系，形成图结构表示的语义。语义感知部分的网络结构和参数设置同文献[29]，并需要先在 COCO 数据集<sup>[26]</sup>上预训练，再与整个通信系统进行联合微调。

## 2) 信息提取

针对图表示的语义，本文首先基于一层 16 通道的图卷积神经网络 (GCN, graph convolutional neural network)<sup>[30]</sup>对其进行节点近邻信息传递；接着，计算出图中各语义的注意力权重，并按照降序排序，得到排序后的权重  $\{w_i\}$ ；最后，筛选出其中前  $j$  个语义作为语信， $j$  满足  $\sum_{i=1}^j w_i \leq t, \sum_{i=1}^{j+1} w_i > t$ ，其中， $t$  为语义节点权重累计阈值，是界定语信的阈值超参数，本文随后将对语信阈值对语信通信性能的影响展开实验验证。

## 3) 编译码与传输

本文使用一层 16 维的线性层对语信进行编码后，通过模拟加性白高斯噪声信道<sup>[31]</sup>进行传输，并使用一层 16 维的线性层进行译码。

## 4) 信息补全

本文使用与信息提取中对称的网络补全信息。

## 5) 基于语义的图像生成

本文使用文献[32]介绍的图像生成模型实现式(4)所述的逆虚景语义特征函数，根据图结构语义中的带属性的对象及其之间的关系直接生成图像。图像生成模型需要先在 OpenImageV6 数据集<sup>[33]</sup>上预训练，再与整个通信系统联合微调。

## 3.4 信噪比对语信通信性能的影响

在语信阈值为 100%、不同信噪比的条件下语信通信前后图像的平均语义相似度如表 3 所示，部分可视化结果如图 6 所示。综合表 3 的平均语义相似度评估和对图 6 中图像的主观感受，虽然语信通信生成的图像与原始图像在像素上不一致，但是其中所包含语义在目标类别、位置等层面上是基本一致的，可在低带宽下(每张图片传输量低于 0.7 KB)完成面向分类、检测等图像理解的图像传输任务。在信噪比大于 20 dB 的情况下，噪声对语信的传输没有影响，平均语义相似度不变且结果图像中语义对象的类别、位置和大小均不变。当信噪比很低，

如为 0 时，噪声对语信传输影响很大，以至于生成图像中语义几乎完全失真。而当信噪比稍低，如为 10 dB 时，噪声对语信传输的影响仍然很大，但已经可以通过选择性地提取和补全信息传输相对重要的语义。

表 3 不同信噪比下语信通信性能对比

信噪比/dB	平均语义相似度
0	0.909
10	0.949
20	0.960
30	0.960

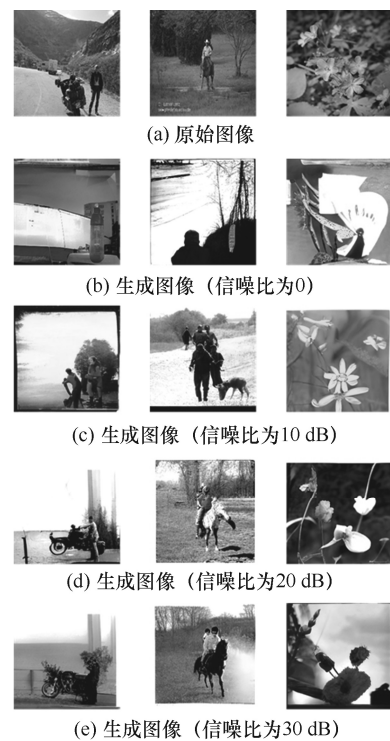


图 6 不同信噪比下语信通信结果与原始图像对照

## 3.5 语信阈值对语信通信性能的影响

语信通信会对面向理解从语义中提取有信息的部分，从而进一步提升对信道资源的利用程度。其中语信阈值的选择将影响语义保真度与传输数据量。在信噪比为 30 dB、不同语信阈值的条件下语信通信前后图像的平均语义相似度如表 4 所示，部分可视化结果如图 7 所示。其中注意力权重累计超过语言阈值且数量最少的一组语义将被选作语信。综合表 4 和图 7 可见，当语信阈值小于或等于 0.9 时，随着语信阈值的提升，平均语信占比也在提升，即信息提取操作保留的部分增加，平

均语义相似度也在提升；而当语信阈值大于 0.9 时，随着语信阈值的提升，平均语信占比也在提升，信息提取操作保留的部分增加，平均语义相似度却略微下降了。这说明语信通信中的信息提取与补全可以将冗余语义和暗语义去掉，不仅没有降低语义通信性能，反而通过信息筛选突出了语信，使性能略微提升。

表 4 不同语信阈值下语信通信性能对比

语信阈值	平均语信占比	平均语义相似度
0.30	24.92%	0.919
0.50	47.21%	0.941
0.70	63.93%	0.952
0.90	86.22%	0.962
0.99	100.00%	0.960



图 7 不同语信阈值下语信通信结果

### 3.6 知识动态更新过程中语信通信性能的变化

为了展现语信通信中知识动态更新对性能的影响，本文将基于语信通信系统进行多个阶段的传输任务，并在每个阶段结束后对知识进行更新。在第一个传输阶段前，在 COCO 数据集<sup>[26]</sup>对语义感知网络进行训练，在 OpenImageV6 数据集<sup>[33]</sup>上对图像生成模型进行训练，并在整个动态更新过程中保持参数固定，作为语信通信的基础知识。

每个传输阶段分为以下几部分：1) 固定所有参数进行传输，并记录性能；2) 将信源端感知得到的图结构语义无失真地发送给信宿端，作为语义真值；3) 信宿端根据语义真值和信息补全得到的语义间的误差按照式(6)对知识进行更新。具体操作为信宿端使用梯度下降法优化信息补全模型参数。信宿端通过无失真传输梯度的方式反馈传输给信源端，使其优化信息提取模型参数。完成基于反馈的知识动态更新。在信噪比为 5 dB 的条件下，语信通信前后图像的平均语义相似度随着知识更新（即累计训练样本数增加）的变化如表 5 所示，部分可视化结果如图 8 所示，原始图像如图 6 所示。综合表 5 的客观语义相似度评估和对图 8 中图像的主观感受可以看到，随着训练样本累计，知识不断更新，语信通信的性能持续提升。在累计训练样本数达到 1 000 的情况下提取的语信就已经能够比较全面地体现图像中的语义。

表 5 随着知识更新语信通信的性能变化

累计训练样本数	平均语义相似度
1	0.907
10	0.924
100	0.929
1 000	0.945
10 000	0.958



图 8 不同累计训练样本数下语信通信结果

## 4 语信通信与语义通信对比仿真分析

本节通过实验验证本文所提语信通信相比于语义通信能够进一步提升对带宽的利用程度。本节实验将直接对知识图谱表征的语义进行传输，消除语义感知与语义重现的影响。

### 4.1 语义样本集

本节实验使用 arXiv-GrQc3 数据集<sup>[34]</sup>作为语义样本集。该数据集将学者间的引用情况用图结构表示， $G=\{V,E\}$ 。其中， $V$  为节点，表示学者； $E$  为边，表示引用关系。对于其中包含的 5 242 个节点和 14 496 条边，本文随机取出 80%的边作为训练信息补全模型的训练集，剩下 20%的边作为测试集用于测试语信通信系统的性能。在训练和测试阶段分别使用训练集和测试集作为原始语义图。

### 4.2 语信通信架构实现细节

#### 1) 信息提取

本文使用根据拓扑先验知识从语义图中提取关键节点的算法进行信息提取。关键节点提取算法流程如算法 3 所示。

#### 算法 3 关键节点提取算法

输入 原始语义图  $G = \{V_0, E_0\}$ ，并初始化关键节点集合  $V = \emptyset$  和边集合  $E = E_0$

输出 关键节点集合  $V$

- ① 将  $G$  中所有的节点按度的降序排列，得到节点序列  $X = \{x_1, \dots, x_N\}$ ；
- ② 循环 for  $i = 1:1:N$  do:
- ③ 从  $E$  中将  $x_i$  的所有边取出构成集合  $E_i$ ， $E_i = E \cap \{(x_i, v) | v \in V_0\}$ ；
- ④ 从  $E$  中将  $E_i$  取出， $E \leftarrow E - E_i$ ；
- ⑤ 将  $x_i$  加入  $V$ ， $V \leftarrow V \cup \{x_i\}$ ；
- ⑥ 当所有的边都被取出时， $E = \emptyset$ ，提前停止循环；

#### 2) 信源编译器

信源编码器用于将语信转换为在信道中传输的比特码，信源译码器用于将接收到的比特码解码为语信。本文采用端到端的方法训练信源编译器，其过程如下。首先，通过一个嵌入函数将符号表示的语信映射到向量空间，得到其嵌入表示；然后，通过两层感知机将嵌入向量，得到比特码；接着，用两层感知机从接收比特码中预测符号表示的语信。该信源编译器的训练目标函数为交叉熵损

失函数，学习率设置为 0.001，训练信噪比为 5 dB。并设置网络结构参数：嵌入空间维度为 128；编码器中第一层感知机的输出维度为 32，第二层感知机的输出维度为 13；译码器中第一层和第二层感知机的输出维度分别为 128 和 5 242。

#### 3) 信息补全

由于本文在信源端通过关键节点提取算法获取了语信，相应地，信宿端将根据这些关键节点构成的语信恢复完整语义，即信息补全。在信息补全中，本文使用 GAN 的图嵌入模型 GraphGAN<sup>[35]</sup>将图节点映射到嵌入空间。在嵌入空间中，一对节点  $(u, v)$  之间存在边的可能性可以用它们对应的嵌入向量  $(\mathbf{u}, \mathbf{v})$  的内积评估，即  $P(\mathbf{u}, \mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle$ 。对于接收的语信  $V = \{v_i\}_{i=1}^N$ ，其中任意 2 个节点都可能存在边，即有  $\frac{N(N-1)}{2}$  条潜在的边。计算所有潜在边的可能性，选出其中前  $N_E$  个作为恢复的边， $N_E$  为原始语义图  $G$  中边的数目。

#### 4) 反馈

为了避免在反馈过程中传输过量数据，信宿端先对恢复出的语义图按照算法 3 进行信息提取，将关键实体集合  $V' = \{v'_i\}$  反馈给信源端。

信源端将  $V'$  与  $V$  作对比，找到其中缺少的部分  $W = C_V V'$ ，然后将  $G$  中所有包含  $W$  的边通过信道发送给信宿端。

### 4.3 仿真结果与分析

在训练集上对信道编译器进行训练后，在测试集进行测试，语信通信与语义通信性能对比如图 9 所示，其中，传输恢复比表示传输量与正确恢复量的比值。

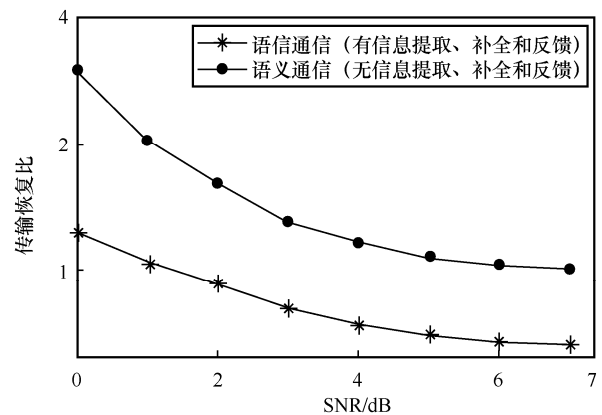


图9 语信通信与语义通信性能对比

语信通信包含信息提取、信道编译码、信息补全和反馈，而语义通信则仅包含信道编译码。由图 9 可见，不论信噪比如何变化，语信通信始终优于语义通信，且信噪比越低，传输恢复比的差距越大。值得注意的是，语义通信的传输恢复比在信噪比为 0 处超过了 3，随着信噪比增大而趋近 1，而语信通信在较低信噪比（略大于 1 dB）时，就突破了传输恢复比为 1 的限制。相比于语义通信，语信通信由于去掉了冗余语义而只传输有信息的语义，在信宿端信息补全的作用下，突破了传输恢复比为 1 的限制。

## 5 结束语

本文提出将语义划分为冗余语义、暗语义和语信三部分，以此提出了面向传输语义中有信息部分的通信范式——语信通信，将通信范式划分为语法、语义、语信、语用四层。语信通信得益于面向语义信息的信息甄别，减少了带宽需求以及语义信息处理的时间和功耗。

## 参考文献:

- [1] SHANNON C E, WEAVER W. The mathematical theory of communication[M]. Urbana: University of Illinois Press, 1949.
- [2] VORA L J. Evolution of mobile generation technology: 1G to 5G and review of upcoming wireless technology 5G[J]. International journal of modern trends in engineering and research, 2015, 2(10): 281-290.
- [3] ARIKAN E. Channel polarization: a method for constructing capacity-achieving codes[C]//Proceedings of 2008 IEEE International Symposium on Information Theory. Piscataway: IEEE Press, 2008: 1173-1177.
- [4] 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.
- [5] 石光明, 李莹玉, 谢雪梅. 语义通讯: 智能时代的产物[J]. 模式识别与人工智能, 2018, 31(1): 91-99.  
SHI G M, LI Y Y, XIE X M. Semantic communications: outcome of the intelligence era[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(1): 91-99.
- [6] SHI G M, GAO D H, SONG X D, et al. A new communication paradigm: from bit accuracy to semantic fidelity[J]. arXiv Preprint, arXiv: 2101.12649, 2021.
- [7] QIN Z, TAO X, LU J, et al. Semantic communications: principles and challenges[J]. arXiv Preprint, arXiv: 2201.01389, 2022.
- [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.
- [9] WENG Z Z, QIN Z J, LI G Y. Semantic communications for speech signals[C]//Proceedings of IEEE International Conference on Communications. Piscataway: IEEE Press, 2021: 1-6.
- [10] YOO H, JUNG T, DAI L, et al. Demo: real-time semantic communications with a vision transformer[J]. arXiv Preprint, arXiv: 2205.03886, 2022.
- [11] LI J G, JIA C M, ZHANG X F, et al. Cross modal compression: towards human-comprehensible semantic compression[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM Press, 2021: 4230-4238.
- [12] HUANG D L, TAO X M, GAO F F, et al. Deep learning-based image semantic coding for semantic communications[C]//Proceedings of IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2022: 1-6.
- [13] HARTLEY R V L. Transmission of information[J]. The Bell System Technical Journal, 1928, 7(3): 535-563.
- [14] RICHARDS A G, ASHBY W R. An introduction to cybernetics[J]. AIBS Bulletin, 1957, 7(1): 41.
- [15] LONGO G. Information theory new trends and open problems[M]. Vienna: Springer Vienna, 1975.
- [16] 钟义信. 信息科学原理[M]. 福州: 福建人民出版社, 1988.  
ZHONG Y X. Principles of information science[M]. Fuzhou: Fujian People's Publishing House, 1988.
- [17] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [18] CRANDALL D, FELZENSZWALB P, HUTTENLOCHER D. Spatial priors for part-based recognition using statistical models[C]//Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2005: 10-17.
- [19] LOSCH M, FRITZ M, SCHIELE B. Interpretability beyond classification output: semantic bottleneck networks[J]. arXiv Preprint, arXiv: 1907.10882, 2019.
- [20] JI Z L, ZOU X L, LIU X, et al. Neural information processing in hierarchical prototypical networks[C]//International Conference on Neural Information Processing. Berlin: Springer, 2018: 603-611.
- [21] SHI G M, ZHANG Z Q, GAO D H, et al. Knowledge-guided semantic computing network[J]. Neurocomputing, 2021, 426: 70-84.
- [22] SHI G M, YANG M X, GAO D H. A novel intrinsically explainable model with semantic manifolds established via transformed priors[J]. Knowledge-Based Systems, 2022, 252: 109386.
- [23] 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.  
SHI G M, GAO D H, YANG M X, et al. Semantic characterization and measurement of signals[J]. Acta Electronica Sinica, 2022, 50(9): 2068-2078.
- [24] KRISHNA R, ZHU Y K, GROTH O, et al. Visual genome: connecting language and vision using crowdsourced dense image annotations[J]. International Journal of Computer Vision, 2017, 123(1): 32-73.
- [25] XU D F, ZHU Y K, CHOY C B, et al. Scene graph generation by iterative message passing[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 3097-3106.

- [26] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//European Conference on Computer Vision. Berlin: Springer, 2014: 740-755.
- [27] KRAUSE J, JOHNSON J, KRISHNA R, et al. A hierarchical approach for generating descriptive image paragraphs[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 3337-3345.
- [28] DEVLIN J, CHANG M, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. arXiv Preprint, arXiv: 1810.04805, 2018.
- [29] TANG K H, NIU Y L, HUANG J Q, et al. Unbiased scene graph generation from biased training[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 3713-3722.
- [30] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint arXiv:1609.02907, 2016.
- [31] O'SHEA T, HOYDIS J. An introduction to deep learning for the physical layer[J]. IEEE Transactions on Cognitive Communications and Networking, 2017, 3(4): 563-575.
- [32] ESSER P, ROMBACH R, OMMER B. Taming transformers for high-resolution image synthesis[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 12868-12878.
- [33] KUZNETSOVA A, ROM H, ALLDRIN N, et al. The open images dataset v4[J]. International Journal of Computer Vision, 2020, 128(7): 1956-1981.
- [34] LESKOVEC J, KLEINBERG J, FALOUTSOS C. Graph evolution: densification and shrinking diameters[J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1): 1-42.
- [35] WANG H W, WANG J L, WANG J, et al. Learning graph representation with generative adversarial nets[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(8): 3090-3103.

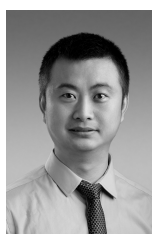
## [作者简介]



石光明（1965-），男，江西南昌人，博士，西安电子科技大学教授，主要研究方向为人工智能、语义通信等。



杨旻曦（1996-），男，四川成都人，西安电子科技大学博士生，主要研究方向为表征学习、计算机视觉、语义通信等。



高大化（1979-），男，河南开封人，博士，西安电子科技大学教授，主要研究方向为智能信息处理、智能感知等。



柴靖轩（1996-），男，河南郑州人，西安电子科技大学博士生，主要研究方向为知识图谱推理、语义通信、信息论等。