

---

# 6G 智简信息传输及使能技术白皮书

(2025 年)

中关村泛联移动通信技术  
创新应用研究院

编制单位：中关村泛联移动通信技术创新应用研究院、北京邮电大学、中信科移动通信技术股份有限公司、中国联合网络通信有限公司研究院、中国移动通信研究院

## 前　　言

随着 6G 网络的快速发展，通信系统正迈向智能化、差异化和多样化需求的新阶段。在此背景下，针对未来机器人等智能化终端泛用所导致的海量多模态数据（如图像、视频、点云等）传输需求，传统以精确到比特级信息传输为核心的编码调制方法逐渐接近香农极限，信息传输速率和系统性能的提升遭遇瓶颈。语义通信作为一种新兴范式，以人工智能技术为核心驱动，将通信范式从比特级跃迁至语义级，聚焦于信息的语义内容而非原始比特流，以突破传统通信香农极限，为未来 6G 网络多模态信息的智简传输提供了全新的技术路径。

因此，本白皮书聚焦 6G 智简信息传输及使能技术，旨在构建适用于语义通信的新型传输理论框架，设计端到端的自适应多模态语义通信系统，并研发动态匹配与更新的语义知识库技术，为语义通信网络的高效和可靠运行提供理论和模型基础，进而构建多模态业务适配的智简信息传输保障使能技术体系，支撑智简传输系统未来可能的工程落地。

# 目 录

1. 引言 .....	1
2. 智简信息传输总体框架.....	2
2.1 设计思路.....	2
2.2 技术框架设计.....	3
3. 新型智简传输理论.....	5
3.1 模分多址传输理论.....	6
3.1.1 MDMA 基本原理 .....	7
3.1.2 MDMA 理论性能分析 .....	8
3.1.3 MDMA 系统性能仿真对比 .....	10
3.2 准正交模分多址传输理论.....	11
3.2.1 0-MDMA 基本原理 .....	11
3.2.2 0-MDMA 理论性能分析 .....	11
3.2.3 0-MDMA 系统性能仿真对比 .....	14
4. 多模态智简传输技术.....	15
4.1 多模态自适应语义通信 .....	15
4.1.1 基于层的图像语义通信系统.....	15
4.1.2 面向点云的语义通信系统 (PCSC) .....	19
4.1.3 基于模分多址的视频语义通信系统 (MDVSC) .....	21
4.2 语义知识库动态匹配与更新 .....	22
5. 智简信息传输使能技术.....	25
5.1 智能新型网络传输与业务保障技术.....	25
5.2 多维资源感知与智能调度技术.....	29

6. 总结与未来展望.....	34
缩略语列表.....	36
参考文献.....	38

## 1. 引言

随着 6G 网络的快速发展，通信系统正迈向智能化、差异化和多样化需求的新阶段。在此背景下，针对未来机器人等智能化终端泛用所导致的海量多模态数据（如图像、视频、点云等）传输需求，传统以精确到比特级信息传输为核心的编码调制方法逐渐接近香农极限，信息传输速率和系统性能的提升遭遇瓶颈。传统通信范式已然难以适配 6G 网络日趋高效的传输需求，亟待开拓新的优化空间与维度，提高通信传输的信息量显得尤为重要。

智简信息传输应运而生，以达意传输为目标，不再是单纯追求越精确越好的数据搬运，而是通过人工智能驱动的语义表征技术，推动通信系统向更高层次的语义理解与表达跃迁。语义通信正是在这一技术需求的强烈牵引下出现，是先理解后传输的一种通信新兴范式，完美契合未来 6G 智简信息传输需求。语义通信以“智（人工智能）”作为核心技术驱动力，将传统通信精确到比特级的传输范式跃迁至语义级，聚焦于信息的语义内容而非原始比特流，从而实现传输的“简”，为 6G 网络信息的智简传输提供了全新的技术路径，为 6G 时代复杂场景下的通信需求开辟了突破物理层桎梏的技术方向，助力实现高效、智能、简约的通信愿景。

同时，6G 网络中以语义传输为代表的智简信息传输仍面临多重挑战。当前存在以下关键问题：首先，在基础理论层面，语义通信和语义传输理论处于构建并逐渐完善阶段，语义信息的提取与表征方法尚未统一，缺乏标准化的语义解析与编码规范；其次，在动态适应性方面，现有技术难以实现多源多端的语义知识库对齐，缺少有效的自适应同步机制；此外，在传输控制层面，面向多样化业务需求的智能精细化管控能力有待提升；最后，在系统融合方面，语义传输与传统语法传输的资源分配和业务传输保障策略尚未实现有效协同。

针对这些挑战，需要从理论方法和使能技术两个维度进行突破：一方面要构建新型智简传输理论框架，探索多模态自适应传输技术路径；另一方面需研究智能新型网络传输与业务保障技术，建立全局资源认知体系及智能调度机制，为语义与语法融合传输的差异化服务质量保障提供系统支撑。

因此，本白皮书聚焦 6G 智简信息传输及使能技术，旨在构建适用于语义通信的新型传输理论框架，设计端到端的自适应多模态语义通信系统，并研发动态匹配与更新的语义知识库技术，为语义通信网络的高效和可靠运行提供理论和模型基础；进而，研究更适合捕捉数据流特征的人工智能技术，实现多模态业务流的特征识别，为智简传输的精准保障提供前提；研究业务数据流的差异化服务质量保障技术，为智简传输的网络资源高效利用和精准适配用户体验提供基础；研究多维资源感知方法与智能调度技术，形成统一的资源表征图谱，并在此基础上智能调度配置智简网络资源，优化系统整体信息传输效率，提升用户体验，为自适应多模态语义通信的实现提供支撑。

## 2. 智简信息传输总体框架

### 2.1 设计思路

由于智简信息传输范式相较于传统通信范式发生了三个层面的核心转变：传统通信以保障比特级精确传输为目标，而智简传输聚焦于语义层面的信息理解与重构；传统信源编码采用基于概率模型的压缩算法，而智简传输通过 AI 驱动的信源信道联合编码实现场景适配；传统通信传输以单一信道容量为优化目标，而智简传输需构建多模态业务语义传输需求的按需、动态、弹性适配的差异化传输保障系统支撑能力。因此，智简信息传输总体框架设计需从理论突破、技术实现、系统支撑三个层面展开：

**第一，构建新型智简传输信息理论，为 6G 智简信息传输提供理论基础。**传统通信理论依托香农信息论，聚焦于确保信息传输的比特级精确性，位于信息传递的三个层次的第一层。然而，在 6G 智简信息传输中，信息传输需求向语义理解层跃迁，传统理论框架与语义通信的发展诉求产生根本性错位。语义通信作为第二层通信范式，其性能评估不再局限于比特级精确性，而需建立面向语义内涵的新型理论模型。

**第二，研发面向多模态的语义智能编码框架，为 6G 智简信息传输提供技术实现路径。**语义通信主要面向结构化信源的传输需求，而结构化数据（如视频/

图像/文本)已占据全网流量的 85%以上。需借助成熟的人工智能语义提取框架,综合考量通信信道状态和条件,生成适用于通信传输的最优编码,从而提升对通信资源的利用率,提升网络总体性能,为 6G 智简信息传输提供语义传输模型。此框架需解决两大技术难点:一是建立语义知识库动态匹配机制,确保收发端语义先验的一致性;二是设计语义知识库的更新迭代算法,解决预定知识库无法覆盖全场景通信需求的问题。

**第三,构建多模态业务适配的智简信息传输保障使能技术体系,支撑智简传输系统的工程落地。**面向 6G 网络承载的多模态业务,需构建三层使能技术体系:在业务保障层,开发基于 AI 的业务特征识别引擎,实现业务流数据模态的自动分类,设计 QoS 控制策略自生成机制,实现为不同类别和用户体验需求的业务提供优化的 QoS 策略选择,从而完成业务适配和传输策略控制;在资源调度层,动态获取网络资源与状态,结合 QoS 控制策略,应用基于 AI 的动态资源分配算法,实现弹性、动态的资源调度与配置优化,提升资源利用效率与服务保障能力;在智简传输层,构建多模态自适应语义通信传输处理机制,基于语义编解码和语义知识库动态匹配与更新,实现高效、鲁棒地传输不同数据形态(模态)的信息语义,并兼容语法传输和语义传输。缺少这些使能技术,语义通信系统将退化为面向单一业务的专用通道。因此,该技术体系是连接理论创新与工程实践的关键桥梁。

## 2.2 技术框架设计

基于以上设计思路，本白皮书依托于新型智简传输信息理论基础，在从智简传输、资源调度、业务保障维度对 6G 智简信息传输架构进行了设计，如

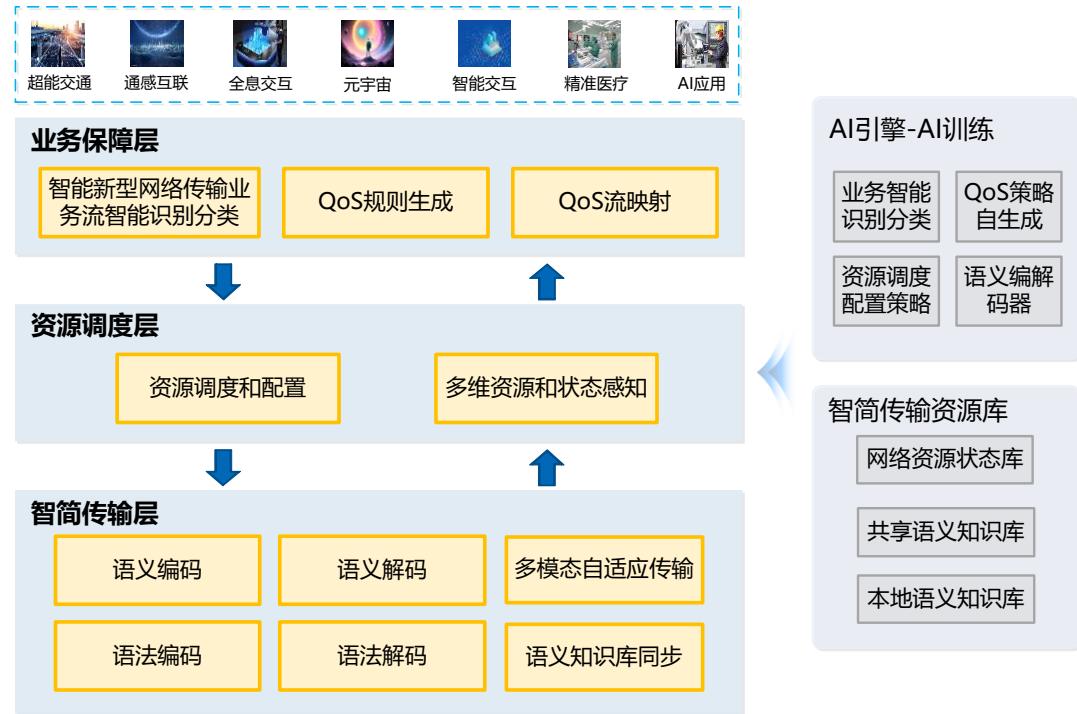


图 2-1 所示：

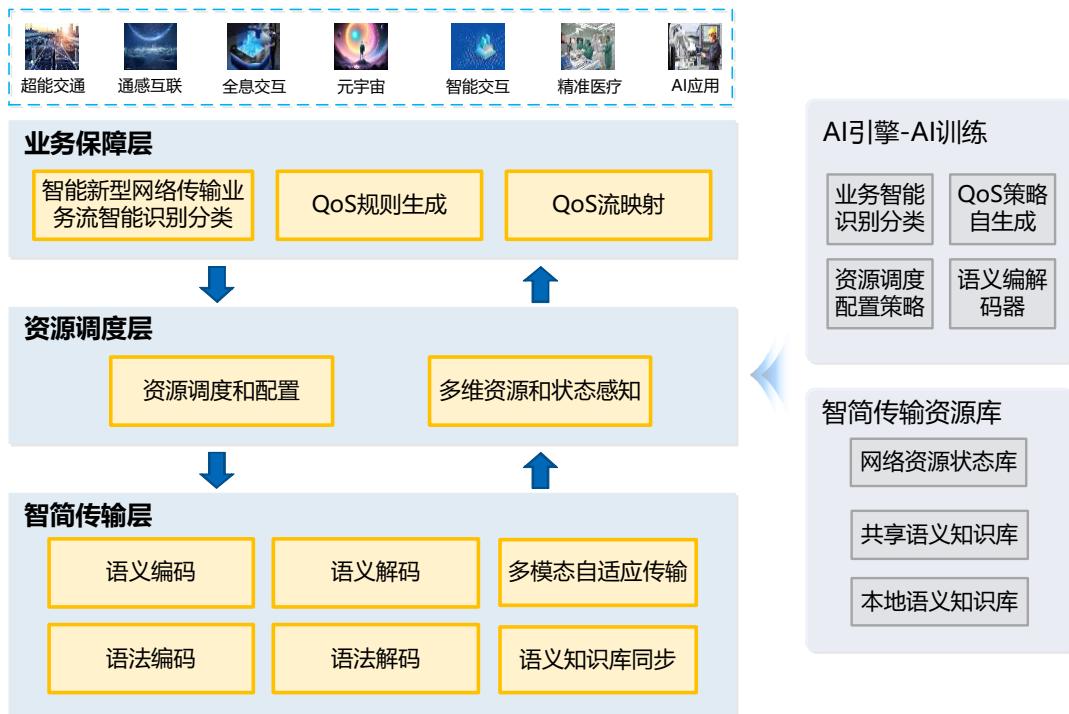


图 2-1 6G 智简信息传输总体架构

### (1) . 智简传输层

智简传输层是智简信息传输理论创新的关键基础能力层，通过智能、灵活的多模态自适应传输技术，高效鲁棒地传输理解不同模态的信息；建立高效的语义知识库动态匹配与更新机制，降低语义知识库更新成本，实现多任务/多信源高效协同适配。

### (2) . 资源调度层

资源调度层通过多维资源和状态的智能感知机制，获取和构建网络资源图谱，为资源调度决策提供数据支撑。通过资源的智能统一调度，显著提升了关键信息传输的可靠性，从而全面增强了网络的 QoS 保障能力。

### (3) . 业务保障层

业务保障层通过智能新型网络传输的业务流智能识别和分类，完成了多模态业务流的精准区分；传输业务保障技术，提供基于 AI 的 QoS 控制策略的生成功能，结合智能分类信息和用户体验需求确定差异化的 QoS 映射规则，用于传输过程中的 QoS 流映射，为服务质量保障奠定基础。

同时，AI 作为赋能智简信息传输的智慧内生引擎，提供智简信息传输架构所需的模型训练能力，包括业务智能识别分类、QoS 策略自生成、资源调度配置

策略和语义编解码器等模型供给。而智简传输资源库则包括语义知识库的管理和网络资源与状态的管理，是智简传输数据层面的基础设施。可通过共享语义知识库和本地语义知识库，为智简传输层语义知识库的动态匹配与更新提供数据支撑；网络资源与状态库则存储网络资源与状态，包括网络拓扑、网络资源与状态信息，可提供给 AI 引擎用于智简模型训练，以及提供全域资源状态给资源调度层作为资源调度和分配的依据。

通过智简传输信息理论和以上技术体系，可为 6G 智简信息传输的标准化和产业化奠定基础，进一步推动学术界和产业界协同努力，助力未来 6G 智简信息传输的可能的工程落地。

### 3. 新型智简传输理论

新型智简传输理论的核心，在于将通信的关注点从 Shannon-Weaver 模型的第一层（精确比特传输）提升至第二层（语义信息传递）。这一范式跃迁意味着，传统通信理论围绕比特恢复构建的基石——香农信息论、误码率（BER）指标及其理论极限——在面对图像、文本等富含语义的信源时，其适配性和有效性面临根本性挑战。语义通信追求的是信息含义的有效理解和传递，这需要突破传统框架，建立一套全新的性能指标和理论基础，以支撑 6G 所追求的“智简”传输愿景，即在复杂环境下实现更智能、更简约、更高效的信息交互。

在这一理论构建过程中，多用户语义通信支持构成了一个核心且极具挑战性的问题。现有的多址接入技术（如 TDMA、FDMA、SDMA、CDMA、NOMA）<sup>[1]</sup>均植根于对物理通信资源（时隙、频段、空间流、码字、功率）的划分与共享机制。然而在语义层面，这些现有技术便暴露根本局限性：它们无法有效感知和利用信源本身（如图像、文本）内在的语义结构。现实世界的信息往往包含共享的公共特征（如一张图片的背景、一段文本的主题）和个性化的私有特征（如图片中特定人物的细节、文本中用户关心的关键信息）。现有多址技术直接切割物理资源，却忽略了信源中蕴含的这种结构性语义信息。这导致了大量冗余信息（尤其是共享特征）被重复传输给多个用户，使得通信带宽效率低下；在低信噪比（SNR）等恶劣信道条件下，基于物理资源争夺的现有方案性能急剧恶化，难以保障基本的语义可达性；更关键的是，它们完全缺乏对用户个性化语义需求的建模与响应。

能力——每个用户可能只关心信息流中与其高度相关的特定语义片段，而现有机制对此束手无策。

因此，关键突破在于发展能在语义模型信息域本身进行资源分配和用户区分的新理论，这直接驱动了对模分多址(MDMA)<sup>[2]</sup>及其演进准正交模分多址(0-MDMA)<sup>[3]</sup>理论界限的研究。MDMA 不再局限于物理资源的“硬分割”，而是深入到信源的语义特征空间，根据其内在的结构(共享性、个性化)以及用户的具体语义需求，进行智能、动态的“软划分”与资源调配。0-MDMA 则更进一步，通过设计“准正交”的不同特征映射(语义模型)，来支持不同模态(如图像与文本)、速率和信号长度的异构数据源同时进行语义传输，且最大程度地控制模态间干扰(保障个性化特征传输的可靠性)。

### 3.1 模分多址传输理论



图 3-1 语义通信创建的模型信息空间

语义通信系统的新范式产生了一种新的分配资源，即模型信息空间，为新的多址技术——模型分多址(MDMA)的建立提供了理论基础。MDMA 是一种从更高信息维度利用信源语义域特征的新型多址方法。借助人工智能模型从信息源中提取语义特征，可以对用户的信源特征进行建模，为多用户语义信息的区分提供知识基础。如图 3-1 所示，结合传统物理资源和模型信息空间，信道容量可表示为：

$$C = TB \sum_{k=0}^n \log_2 \left( \mathbf{I} + \frac{P_k}{\sigma^2} \mathbf{H} \boldsymbol{\phi}_k \mathbf{H}^H \right), \quad (3-1)$$

其中， $T$ 、 $B$ 、 $P$ 、 $\mathbf{H}$ 、 $\boldsymbol{\Phi}$  分别表示时域资源、频域资源、功率域资源、空域资源和模型信息空间资源， $\mathbf{I}$  为单位矩阵， $\sigma^2$  为噪声方差。

### 3.1.1 MDMA 基本原理

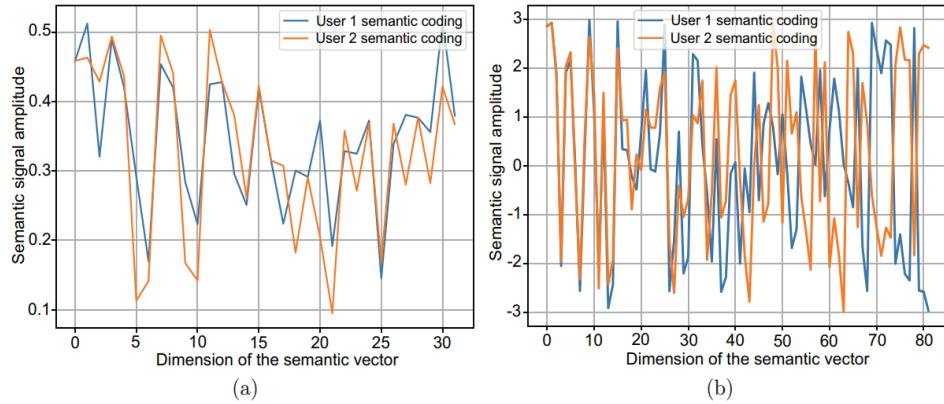


图 3-2 两个用户为 MNIST (a) 和 Open Image (b) 数据集提取的不同图像的语义信号<sup>[2]</sup>

提取语义信号时，信源被映射到高维语义模型空间。本节说明不同用户在该空间存在大量共性语义信息。下面将介绍用户共性与个性语义信息，并描述基于此的 MDMA 上下行流程。

实验使用图像语义通信系统 (MNIST/Open Image 数据集)。如图 3-2 所示，用户 1 和用户 2 的语义信号 (蓝线/橙线) 在不同图像上表现出显著相似性：(a) 不同数字图像信号幅度/相位仅有很小的变化；(b) 复杂图像信号在前 30 维高度相似，60-80 维幅度相似但极性相反。这表明用户间存在大量共性信息。

MDMA 技术旨在利用此共性信息避免冗余传输浪费带宽。它并非在信号空间区分用户，而是在模型信息空间挖掘共性与个性信息：共性信息在相同时频资源内传输复用，个性信息单独传输。相较于传统多址，MDMA 的增益主要源于模型信息空间中共性信息的重用。

图 3-3 描述了多个用户同时向基站传输语义信息的上行场景，首先基站匹配两个发起上行传输指令的用户，用户 1 和用户 2 分别提取各自的信源语义信息  $S_x$  和  $S_y$ ，然后用户 1 和用户 2 在时刻 1 (频率 1) 向基站发送共性语义信息  $S_{x_s} + S_{y_s}$ ，之后用户 1 和用户 2 在时刻 2 (频率 2) 向基站发送个性语义信息  $S_{x_p}$  和  $S_{y_p}$ 。基站接收到用户 1 和用户 2 发送的共性信息和个性信息后，恢复出各个用户的原始语义信号，并通过语义解码器恢复出信息来源。

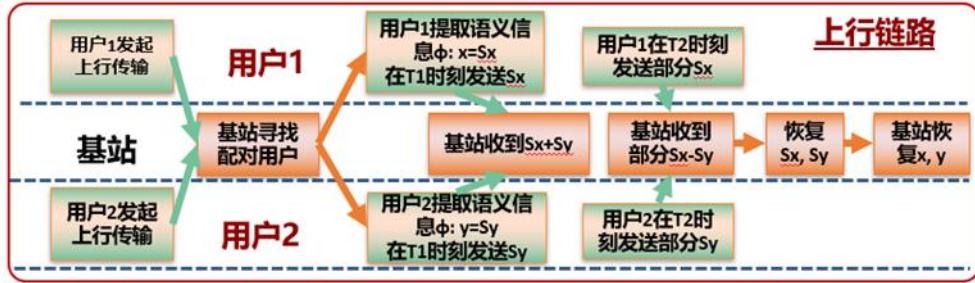


图 3-3 MDMA 上行链路示意图

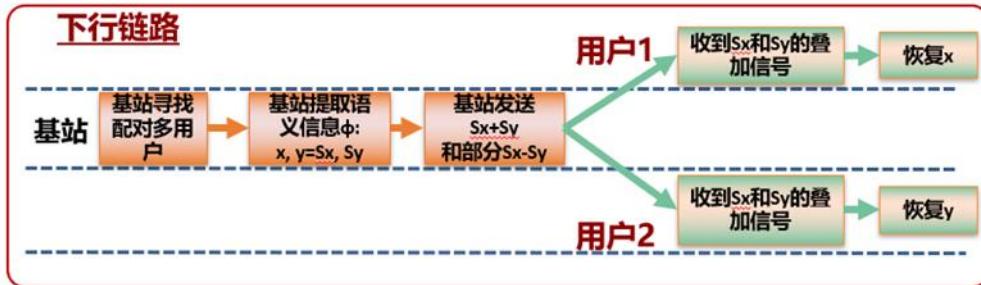


图 3-4 MDMA 下行链路示意图

下行链路设计如图 3-4 所示，允许基站同时将来自多个信源数据的语义信号分发给多个用户。首先，基站根据用户需求，搜索匹配的用户，并使用基于人工智能的语义模型提取语义信息。通过比较两个语义特征的相似性，可以提取共性信息  $S_{x_s}$  和  $S_{y_s}$  以及个性信息  $S_{x_p}$  和  $S_{y_p}$ 。携带共性信息  $S_{x_s}$  和  $S_{y_s}$  的信号被合并一起发送，而携带个性信息  $S_{x_p}$  和  $S_{y_p}$  的信号则分别发送给用户。因此，用户 1 和用户 2 接收到共享信息和它们自己的个性化信息，并且可以使用语义解码器恢复原始图像  $X$  和  $Y$ 。

### 3.1.2 MDMA 理论性能分析

#### (1) 上行 MDMA

考虑一个具有  $K$  用户 ( $K \geq 2$ ) 上行 MDMA 系统。每个用户  $i$  ( $1 \leq i \leq K$ ) 尝试将信源  $x_i$  发送到基站，使用基于模型的语义编码器将  $x_i$  映射到信道输入序列  $\phi_i(x_i)$ 。假设每个发送者  $i$  在一系列共享时频资源上以功率  $P_i$  发送其信号。基站接收到的信号可表示如下，

$$y = \sum_{i=1}^K \sqrt{P_i} \phi_i(x_i) + n, \quad (3-2)$$

其中  $n$  表示信道的高斯噪声。根据上节的分析，由于用户间存在共性语义信

息,因此用户使用的带宽和时间可以重叠。令 $T_i$ 和 $F_i$ 分别为用户 $i(1 \leq i \leq K)$ 占用的时间和带宽资源,令 $T_i = |T_i|$ 和 $B_i = |F_i|$ 表示用户 $i(1 \leq i \leq K)$ 对应的时间和带宽开销。这里定义一个度量标准,称为语义重叠率( $Sor$ )来表征时间或频域中的资源重用率,可以表示如下:

$$Sor = \frac{|T_1 \cap T_2 \cap \dots \cap T_K|}{|T_1| + |T_2| + \dots + |T_K|}, \text{ 或者 } Sor = \frac{|F_1 \cap F_2 \cap \dots \cap F_K|}{|F_1| + |F_2| + \dots + |F_K|}. \quad (3-3)$$

通过评估重用率 $Sor$ 与多用户信号检测性能之间的关系,可以显示上行链路MDMA方案与传统多址接入方法相比的资源利用能力。

此外,这里通过理论分析,解释了上行MDMA系统相较于传统上行MA方法具有性能优势的原因。显然,如公式(3-2)所示,上行MDMA的系统模型在形式上与上行功率域NOMA(PD-NOMA)的系统模型相似。因此,很容易证明上行MDMA系统可以像PD-NOMA方案一样达到MA信道可实现容量区域的极限。为了便于从理论上对上行MDMA系统与NOMA系统的性能进行比较,这里将信道传输速率 $R_{cc}$ (上限为多址接入系统的信道容量 $C$ )与信源编码速率 $R_{sc}$ (下限为信源 $X$ 的熵 $H(X)$ )之比设置为一个新的指标,称为适配度 $F = R_{cc}/R_{sc}$ ,用于表征信道可以为多址接入系统提供的服务能力,基于这一新度量,将容量区域扩展到上行链路MDMA和其他具有双用户多址信道的比较方案的可实现适配度区域。为了实现可实现可行区域的上限,假设基站检测两个用户的信号以达到可实现容量区域的极限,例如通过使用连续干扰消除(SIC)。 $C_i^u$ 表示用户 $i$ 可实现的信道容量,即

$$C_i^u = \frac{1}{2} \log_2 \left( 1 + \frac{P_i}{\sigma^2} \right), i = 1, 2, \quad (3-4)$$

$C_{12}^u$ 表示两用户多址接入信道的总的信道容量,

$$C_{12}^u = \frac{1}{2} \log_2 \left( 1 + \frac{P_1 + P_2}{\sigma^2} \right), \quad (3-5)$$

其中 $P_1$ 和 $P_2$ 表示用户1和用户2传输的信号能量。

如图3-5(左)所示,与上行NOMA方案的适配域相比,上行TDMA/FDMA方案的信道服务容量较低,因为它们的总传输速率无法达到总信道容量。由于多用户共性信息的重用,语义编码器通常表现出比传统编码器 $R_{min}^{NOMA}$ 更低的编码速率 $R_{min}^{MDMA}$ ,因此上行MDMA的适配域可以大于上行NOMA,

$$\frac{C_i^u}{R_{min}^{MDMA}} > \frac{C_i^u}{R_{min}^{NOMA}}, i=1,2, \quad (3-6)$$

因此上行 MDMA 可以提供更高的服务能力, 使得上行 MDMA 相较于上行 NOMA 获得性能优势。

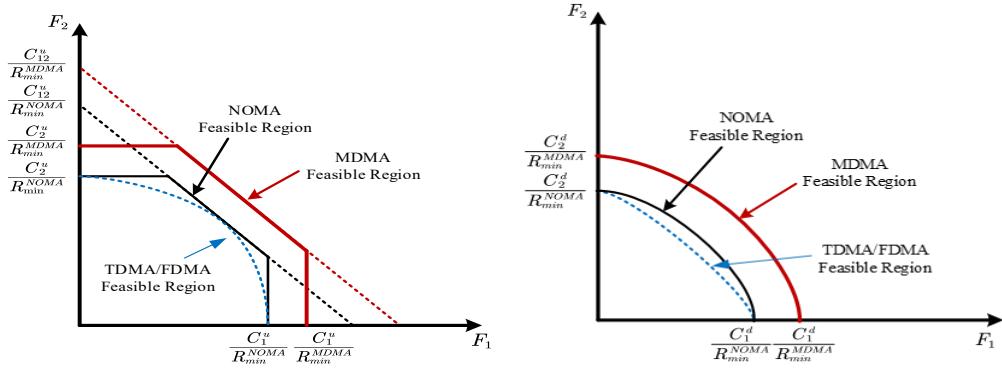


图 3-5 上行 (左) / 下行 (右) 链路 MDMA 和比较方案的可行区域<sup>[2]</sup>

## (2) 下行 MDMA

考虑具有高斯广播信道的  $K$  用户 ( $K \geq 2$ ) 下行 MDMA 系统, 基站尝试将  $K$  个信源  $x_1, x_2, \dots, x_K$  发送到其相应的用户, 类似地, 可以利用 *Sor* 来评估下行 MDMA 方案的资源节省能力。此外, 从理论分析的角度, 可以建立一个适配域来评估下行 MDMA 方案在双用户广播信道下所能提供的服务能力。在总发射功率为  $P$  时, 应直接检测功率因数  $\alpha$  较大的用户 1 的发射信号, 因此用户 1 和用户 2 的信道容量可以表示为,

$$C_1^d = \frac{1}{2} \log_2 \left( 1 + \frac{\alpha P}{(1-\alpha)P + \sigma_1^2} \right), \quad C_2^d = \frac{1}{2} \log_2 \left( 1 + \frac{(1-\alpha)P}{\sigma_1^2} \right), \quad (3-7)$$

和上行 MDMA 同理, 下行 MDMA 的适配域要大于下行 NOMA,

$$\frac{C_i^d}{R_{min}^{MDMA}} > \frac{C_i^d}{R_{min}^{NOMA}}, i=1,2, \quad (3-8)$$

由此得到与上行场景相同的结论, 如 3-5 (右) 所示, 下行 MDMA 能提供更高的服务能力, 因此下行 MDMA 可以比传统方案获得更大的性能增益。

### 3.1.3 MDMA 系统性能仿真对比

为展示 MDMA 在语义通信中的兼容性与优势, 这里将其与 JSCC、LSCI 结合, 并与传统系统对比。图 3-6 显示, 经功率带宽归一化后, 在 AWGN 信道下, 高信噪比时 MDMA+JSCC 性能略低于 NOMA+LDPC+BPSK 和 OFDM+LDPC+BPSK, 但在低信噪

比下优势显著。信噪比为-10dB时,MDMA+JSCC的PSNR达18dB,而NOMA+LDPC+BPSK仅11dB,MDMA+JSCC带宽优势至少7dB。

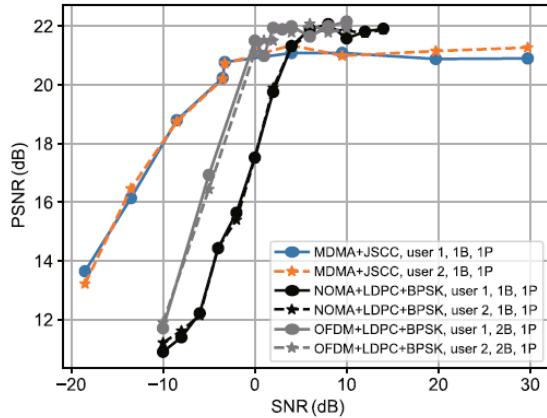


图 3-6 基于 JSCC 的 MDMA 语义通信系统与传统 NOMA 和 OFDM 结合 LDPC 编码系统的性能对比<sup>[2]</sup>

## 3.2 准正交模分多址传输理论

### 3.2.1 O-MDMA 基本原理

模型分多址 (MDMA) 利用语义空间共享信息减少传输量,但受限于同模态和一致信号长度<sup>[2]</sup>,用户增多时扩展性受限。为此,提出结合语义准正交信号和深度联合信源信道编码 (DeepJSCC) 的准正交模型分多址 (O-MDMA)<sup>[3]</sup>。O-MDMA 利用模型理解能力区分用户,将其他用户视作干扰,并利用 DeepJSCC 的抗干扰能力(从抗噪声扩展到抗多用户干扰),提升多用户性能。相比 MDMA, O-MDMA 支持不同模态(如图像与文本)、速率和信号长度的异构数据源。例如在低 SNR (-6 dB 到 6 dB) 的 AWGN 信道中,性能优于 NOMA 和 MDMA。

O-MDMA 的关键优势包括: 1) 利用语义准正交信号区分用户,不依赖传统物理资源分配,减少冗余提升带宽效率; 2) DeepJSCC 功率分配策略在低 SNR 下带来至少 5 dB 性能优势; 3) 支持多模态语义模型,通过协议控制和信号拼接适应不同维度信号,解决 MDMA 的扩展性限制。

### 3.2.2 O-MDMA 理论性能分析

#### (1) 语义准正交信号

如图 3-7 所示,被同一语义编码器转换的语义信号叠加传输,并由同一解码器解码出的信号发生重叠的现象,然而,对于不同结构和不同信源的语义编码器

和解码器模型，解码结果不受影响，一定程度上说明此类语义信号是近似准正交的。

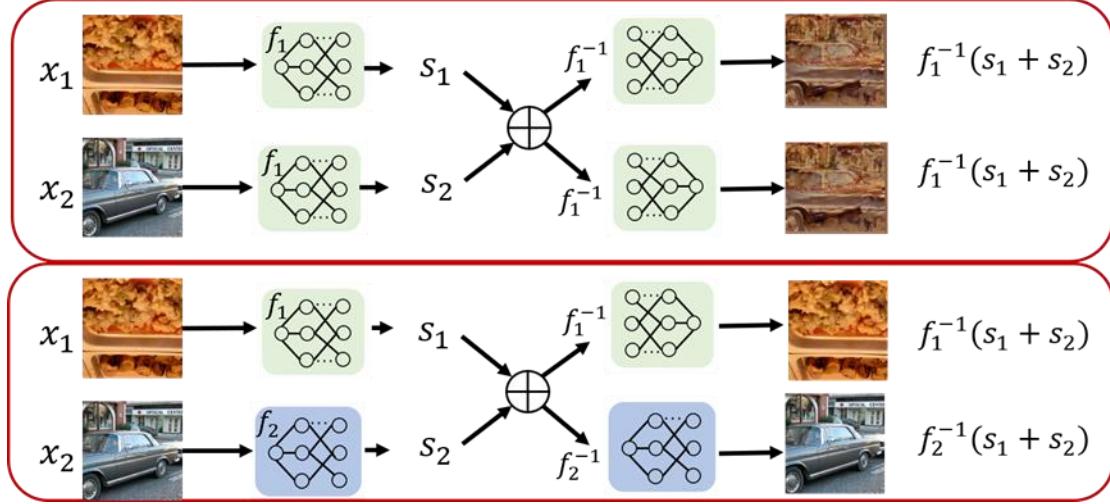


图 3-7 验证语义准正交信号实验框架

为了进一步说明来自相同模型和不同模型的语义信号干扰的影响，这里对重建图像的残差进行了统计分析。具体过程如下：将高斯噪声、不同模型生成的语义信号  $s_{j,x_3}$  和同一模型生成的语义信号  $s_{i,x_2}$  分别添加到不同模型提取的语义信号  $s_{i,x_1}$  中。随后，使用解码过程来计算重建图像和原始图像之间的残差分布，

$$\begin{cases} R_n = f_d(s_{i,x_1} + n; \theta_i) - f_d(s_{i,x_1}; \theta_i), \\ R_o = f_d(s_{i,x_1} + s_{j,x_3}; \theta_i) - f_d(s_{i,x_1}; \theta_i), \\ R_{no} = f_d(s_{i,x_1} + s_{i,x_2}; \theta_i) - f_d(s_{i,x_1}; \theta_i), \end{cases} \quad (3-9)$$

其中  $n$  表示高斯分布， $R_n$ 、 $R_o$  和  $R_{no}$  的分布如图 3-8 所示。不同语义模型产生的语义信号的干扰近似等于高斯噪声的干扰。它甚至可能比高斯噪声的影响更弱，这可以依靠基于 DeepJSCC 的语义模型消除干扰。然而，同一语义模型干扰产生的语义信号具有破坏性。

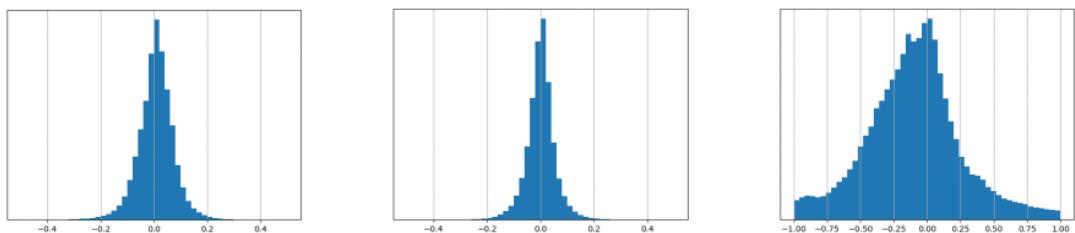


图 3-8  $R_n$  (左)、 $R_o$  (中) 和  $R_{no}$  (右) 的分布图<sup>[3]</sup>

## (2) 准正交 MDMA 上下行链路可行域分析

0-MDMA 的上行、下行链路设计和 1.1 章节的 MDMA 的上行、下行链路设计方

案相同。为了对 O-MDMA 提供更全面的理论解释, 这里推导了 O-MDMA 的可行区域  $F = R_{cc}/R_{sc}$ , 该可行区域结合了信道传输速率  $R_{cc}$  (上限受多址接入系统的信道容量  $C$  限制) 和源编码速率  $R_{sc}$  (下限受源 X 的熵  $H(X)$  限制)。当  $R_{sc}$  归一化为 1 时, 它表示传统的信道容量, 只考虑物理信道资源(时间、频率、空间和功率), 缺乏对与信息源相关的语义业务能力的考虑。在实践中, 模型信息空间的准正交性可以进一步减少物理信道上的带宽资源。因此, 可行性度量可以更准确地描述语义通信系统中的收益: 无论是增加物理资源还是减少业务资源。

经过推导, 上行链路可行域和总和可行域可以表示为,

$$F_{min}^{MAC,O-MDMA} = \frac{C_{min}^{MAX}}{R_{min}^{O-MDMA}} = \min\left(\frac{C_1^*}{R_{min}^{O-MDMA}}, \frac{C_2^*}{R_{min}^{O-MDMA}}\right), \quad (3-10)$$

$$F_{min}^{MAC,NOMA} = \frac{C_{min}^{MAX}}{R_{min}} = \min\left(\frac{C_1^*}{R_{min}}, \frac{C_2^*}{R_{min}}\right), \quad (3-11)$$

$$F^{MACSR,O-MDMA} = \frac{C^{MACSR}}{R_{min}^{O-MDMA}}, \quad (3-12)$$

$$F^{MACSR,NOMA} = \frac{C^{MACSR}}{R_{min}}. \quad (3-13)$$

基于上述等式, 图 3-9 展示了上行/下行链路中传统多址技术和 O-MDMA 的可行域。

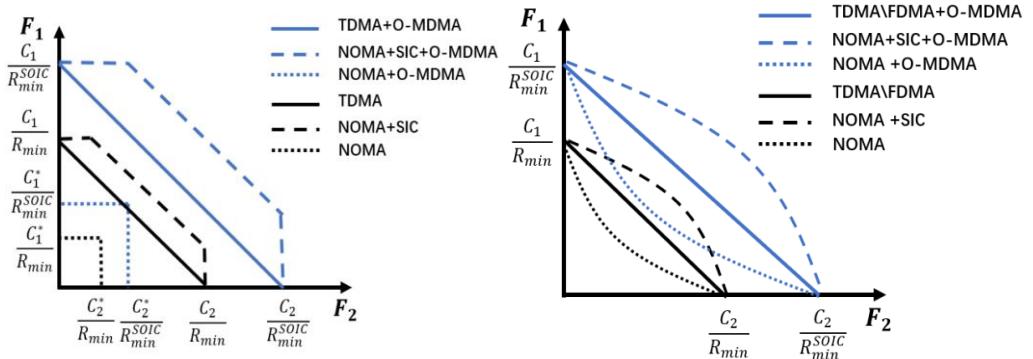


图 3-9 准正交 MDMA 上行/下行链路的可行域分析

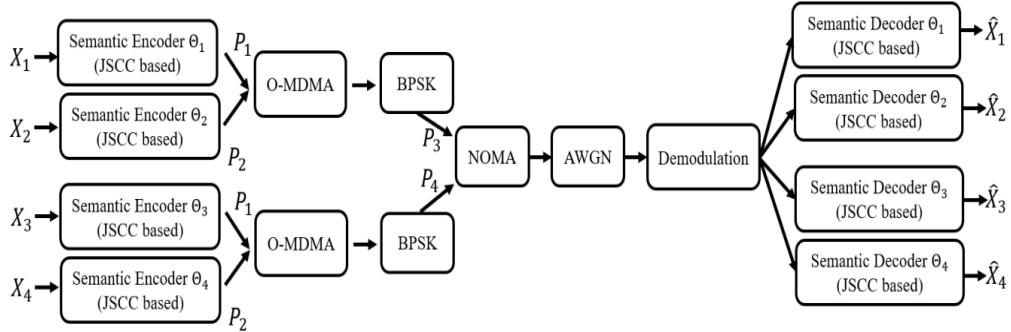


图 3-10 准正交 MDMA+NOMA 系统框架<sup>[3]</sup>

### 3.2.3 O-MDMA 系统性能仿真对比

本节展示 O-MDMA+NOMA 系统性能。如图 3-10 所示, 4 用户信号首先用 O-MDMA 叠加语义信号, 再用 NOMA 叠加传输, 接收端通过各自的语义解码器解码。实验设置  $\theta_1 = \theta_3$ ,  $\theta_2 = \theta_4$ ,  $\theta_1 \cap \theta_2 = \emptyset$  且四个模型性能相同。此外设置  $P_1 = P_2 = 0.5$ ,  $P_3 = 0.8$ ,  $P_4 = 0.2$ 。

采用 CBR=0.041 的 LSCI 模型(结构同参数异的准正交模型)作为语义模型。

图 3-11(a) 显示两用户 O-MDMA 与 NOMA 的性能: 由于 O-MDMA 功率分配均匀, 用户 1 和 2 性能相同; 在低信噪比(-6dB ~ 6dB)下, O-MDMA 明显优于 NOMA。O-MDMA 性能在 6dB 以上饱和(因 0dB 时已达模型最佳)。O-MDMA 支持不同模态语义模型, 图 3-11(b)展示了其用于 LSCI 和 DeepSC 传输, 相比 NOMA 具有显著优势。

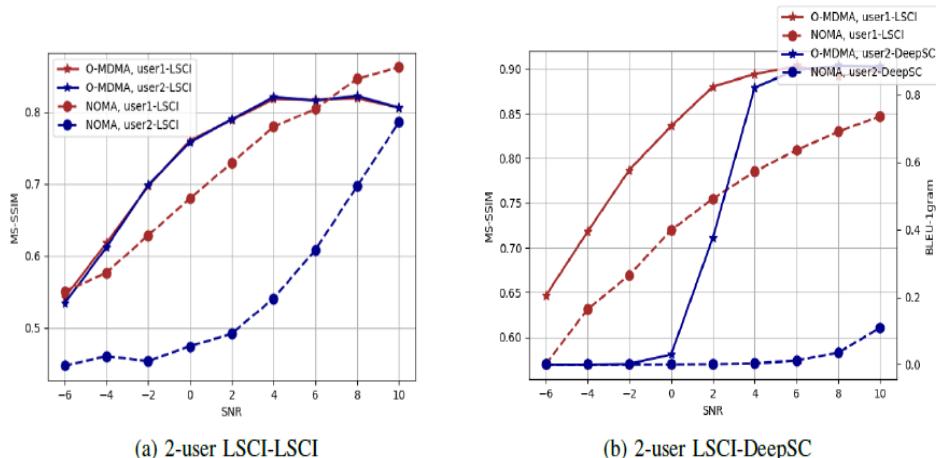


图 3-11 AWGN 信道中使用准正交 MDMA 和 NOMA 的 LSCI 和 DeepSC 性能<sup>[3]</sup>

## 4. 多模态智简传输技术

MDMA 与 0-MDMA 通过建立语义特征空间的多址传输方案, 为多用户场景提供了语义级资源分配与干扰管理的核心框架。这一理论的价值不仅限于多址接入, 其将信息解耦为可独立操作的语义组件(模态)并实现动态调度的思想为更广义的语义通信演进提供了关键技术支撑。在多模态自适应语义通信中, MDMA 的特征空间映射机制可应用于不同模态(图像/点云/视频等)的特征级资源优化, 解决低 SNR 鲁棒性、特征压缩与码率控制等挑战; 同时, 其模态重组逻辑可驱动跨源跨任务的特征复用。对于语义知识库动态匹配与更新, MDMA 的“按需模态分配”范式进一步演化为模型被解耦为可动态调度的语义模态组件(如通用特征提取器与任务专用模块), 结合 0-MDMA 的准正交资源分配原则, 实现边缘侧轻量、抗噪的增量更新与跨任务适配。至此, MDMA/0-MDMA 从多址传输方案升维为语义特征动态管理的普适性框架, 成为支撑多模态自适应传输与知识库演化的共同基石。

### 4.1 多模态自适应语义通信

多模态自适应语义通信的核心目标是高效鲁棒地传输理解不同模态的信息。6G 网络的多样化应用(如物联网、自动驾驶、XR)产生关键数据形态——图像、点云和视频, 其数据结构、语义特征、需求和传输挑战差异显著。单一通用方案难以兼顾, 需针对关键模态开发针对性方案: 图像需解决资源消耗、低 SNR 鲁棒性和低比特率语义保真度; 点云需应对庞大数据规模、信道干扰下的可靠性、特征高效利用与灵活码率控制; 视频需克服计算复杂度、时空特征高效捕获、复杂信道鲁棒性、精准码率控制与延迟抖动。研究专用系统旨在深入解决这些模态特有难题。

#### 4.1.1 基于层的图像语义通信系统

6G 和 AI 的快速发展推动语义通信成为下一代范式, 其聚焦语义内容而非比特传输, 显著提升了效率<sup>[4-5]</sup>。图像作为 6G 关键模态, 其语义通信面临核心挑战:

首先，传播用于语义提取的 AI 模型消耗高昂通信资源，大型模型参数在资源受限设备上的传输代价巨大，限制了灵活部署。其次，低信噪比下，传统基于无差错比特传输的系统易受“悬崖效应”影响，造成语义传输失效，难以满足 6G 低时延高可靠性要求。最后，现有图像压缩技术（如 JPEG、JPEG2000）在低比特率时普遍存在模糊和块效应问题，严重损害语义保真度，进而影响下游智能任务性能。这些挑战制约了图像语义通信在 6G 中的应用效率与实用性。

为应对挑战，本白皮书提出基于层的图像语义通信系统（LSCI）<sup>[6]</sup>，其引入语义切片模型（SeSM）实现高效的语义提取、传输和重建<sup>[19]</sup>。LSCI 的核心优势在于：

- 1) 降低模型传播消耗：分层设计的 SeSM 允许基于性能、信道状态和目标动态选择切片模型，增强基础模型输出而无需传输完整 AI 模型。例如，提升目标检测准确率从 95% 到 98% 仅需传输优化的 SeSM，显著减少流量，缓解设备部署压力。
- 2) 提升低 SNR 鲁棒性：结合信道编码切片模型与反馈机制，利用 JSCC 思想有效缓解“悬崖效应”。
- 3) 优化低比特率语义保真度：通过多尺度 GAN 进行语义压缩与重建，在低比特率下实现约 2% 比特率降低，比 JPEG2000 保留更清晰的纹理和语义细节。

#### （1）LSCI 整体结构

LSCI 由基础模型和增强模型组成，如图 4-1 所示。基础模型是一个基于多尺度的语义压缩系统 GAN 技术。增强模型的作用包括准确性和语义控制的改进。每一层增强模型的每一层都可以被视为本文中提到的语义切片模型。此外，还有信道编码的切片模型适用于抵抗不同的信道噪声。

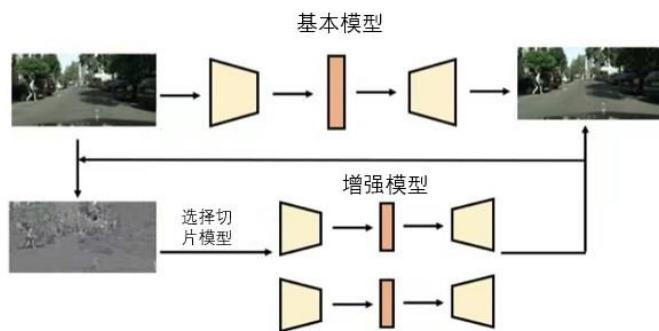


图 4-1 LSCI 的体系结构

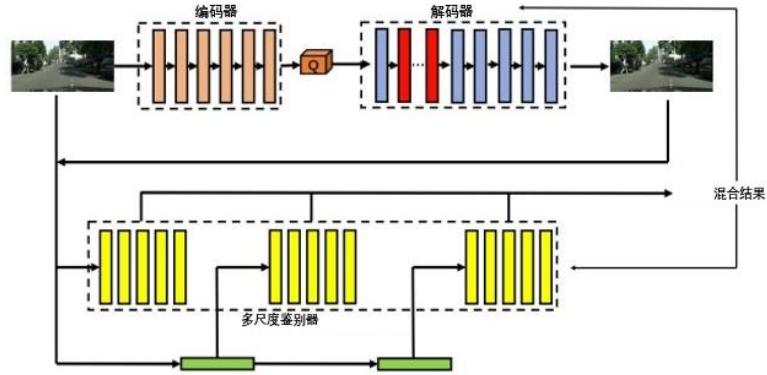


图 4-2 基本模型结构

基本模型实现了图像语义压缩和如图 4-2 所示, 基本模型以极低的比特率实现图像语义压缩和重建。基本模型由编码器、量化器和生成器组成。其中在发射器处提取图像  $x$  的基本语义特征  $w$  , 通过量化后在接收端进行语义恢复。量化采用最近邻量化,

$$w_i^q = \operatorname{argmin} | | w_i - c_j | | . \quad (4-1)$$

如系统结构所示, 使用多尺度鉴别器。下采样层提供具有不同分辨率的三幅图像, 然后将这些图像馈送到三个鉴别器中, 产生融合决策结果。鉴别器的优化目标是能够区分数据集和恢复的图像。鉴别器的优化目标是能够区分数据集  $X$  和恢复的图像,

$$\max \mathbb{E}[f(D(x))] + \mathbb{E}[g(D(x'))], \quad (4-2)$$

其中  $f$  和  $g$  是标量函数,  $x'$  是恢复的图像,  $D$  表示鉴别器, 这有助于最小化恢复图像  $x'$  的概率分布和  $x$  的概率分布之间的差异。

如图 4-3 所示, 增强切片模型用于提升图像关键区域的细节和精度。基本模型输出为有损压缩结果, 增强模型通过处理原始图像与重建图像间的残差信息 (受语义区域掩码和下采样控制) 来恢复精度, 精度提升幅度与所需传输带宽正相关。

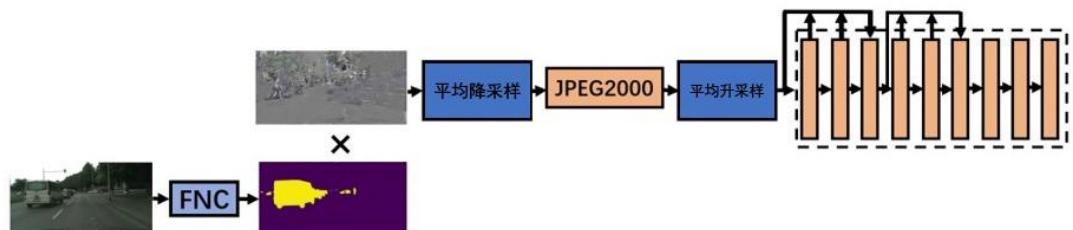


图 4-3 增强模型结构

## (2) 信道编码切片模型

现有联合信源信道编码 (JSCC) 方案通常绑定特定信道, 部署灵活性差。LSCI 提出独立训练的信道编码切片模型 (不更新语义源压缩模型), 结合反馈机制提升对不同信道的适应性, 如图 4-4 所示。体系结构的上半部分是信源编码部分, 在基本模型和增强切片模型中引入, 不需要更新。下半部为信道编码部分, 主要由信道切片编码器和信道切片解码器组成。

输入图像表示为  $x$ ,  $w^q$  在源编码之后获得,  $y$  由信道编码器编码。

$$z = h \cdot y + n, \quad (4-3)$$

其中,  $h$  表示信道增益,  $n$  表示独立的均匀分布的高斯噪声。考虑到噪声反馈的存在, 将包含信源信号特征和噪声的反馈信号的接收部分与语义特征拼接,

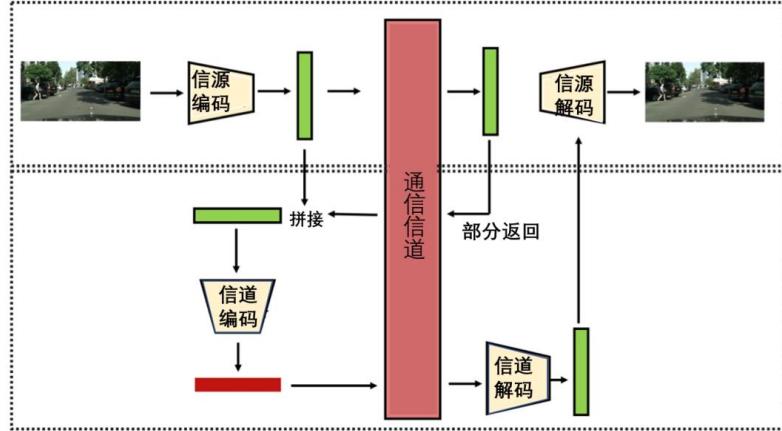


图 4-4 信道编码分层结构

$$w^f = \text{cat}[w^q, z]. \quad (4-4)$$

信道切片编码器  $C_e$  可以通过适用来自反馈信号的信道特性增加冗余以抵抗噪声, 接收机通过信道切片解码器  $C_d$  恢复信源语义信息,  $C_e$  和  $C_d$  的优化目标除了是最小化语义特征的距离, 还最小化重构图像和原图像之间的距离,

$$\min |w^q - \widehat{w}^q| + \gamma d(x, x'). \quad (4-5)$$

如图 4-5 所示, 相比传统方案 (JPEG2000+LDPC+4QAM), LSCI 在低信噪比 (SNR) 下显著缓解了“悬崖效应”。传统方案在 AWGN 信道约 3dB、瑞利信道约 6dB 处出现性能骤降。错误比特的随机性导致传统系统重建图像可见损伤。

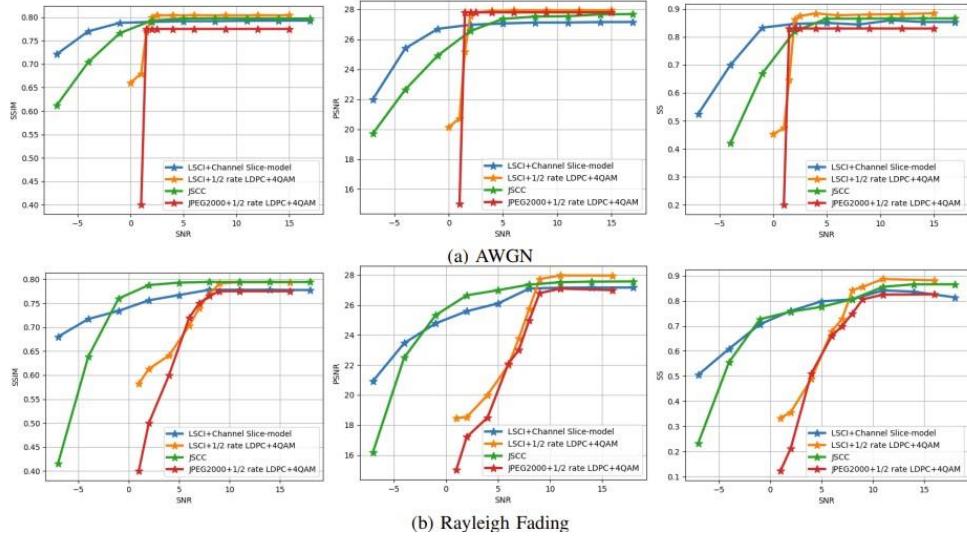


图 4-5 语义恢复不同信道性能的数据比较<sup>[6]</sup>

#### 4.1.2 面向点云的语义通信系统（PCSC）

点云作为 6G 关键应用（如自动驾驶、VR/AR<sup>[7]</sup>）中三维信息的重要载体，其高效语义传输面临挑战：庞大资源消耗（百万级点数导致高带宽/存储需求，难以满足资源受限设备实时性）；低信噪比鲁棒性差（动态信道干扰下分离编码易受“悬崖效应”，性能骤降）；语义保真度不足（传统压缩技术如 G-PCC 侧重几何信息，低比特率下易丢失关键语义信息，影响下游任务）；编码速率控制复杂（现有语义通信系统缺乏简易动态速率控制，调整常需重训或复杂模块）

为应对上述挑战，本白皮书提出点云语义通信系统（PCSC）<sup>[8]</sup>，1) 降低传输消耗：对编码语义向量基于重要性（如 VSV、VGSV）排序并动态丢弃低值向量，减少带宽需求。2) 提升抗噪鲁棒性：采用联合信源信道编码（JSCC）技术缓解“悬崖效应”。3) 增强语义保真度：基于 Voxception-ResNet 提取重建语义特征，低 CBR 下 PSNR D1/D2 指标优于 G-PCC 和 PCL，确保关键语义保真。4) 简化速率控制：引入无需重训的速率分配方法动态调整编码速率。

如图 4-6 所示，PCSC 工作流程：点云预处理分割成立方体（图 4-7）；JSCC 编码器映射至语义向量  $y$ ；速率分配模块排序并丢弃不重要信息；经变长编码和信道传输（含噪声损伤）；接收信号经零填充恢复长度；JSCC 解码器输出重建点云。系统为端到端对称网络联合训练。

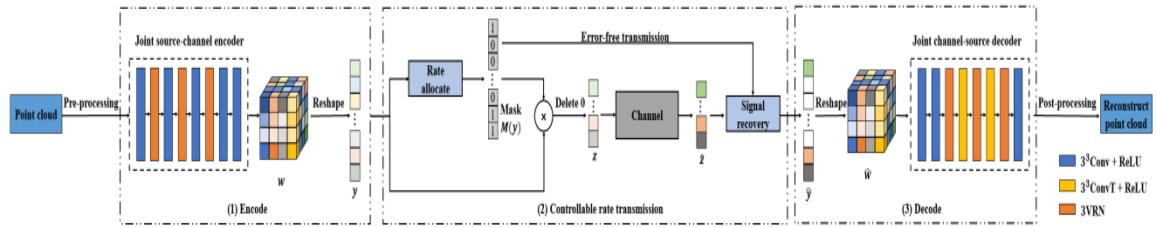


图 4-6 PCSC 的整体设计

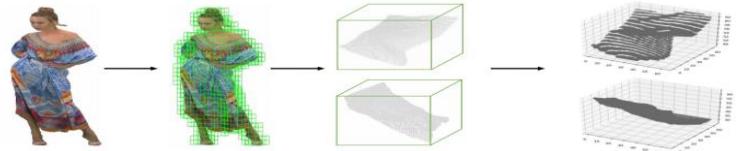


图 4-7 点云预处理图<sup>[8]</sup>

对于速率分配模块，根据 $y$ 的重要性排序和限制生成 0-1 掩码 $M(y)$ 。 $M(y)$ 与 $y$ 按元素相乘，0 值的结果将被丢弃以获得传输符号 $z$ 。缩短的符号 $z$ 通过无线信道传输，掩码 $M(y)$ 将无差错地传输到接收器。接收到 $\hat{z}$ 后，在信号恢复模块中，对 $\hat{z}$ 进行补零，以保证 $\hat{y}$ 的长度等于 $y$ 。其中，重要性排序的依据有：语义向量 $y$ 的值，语义向量 $y$ 的梯度，语义向量 $y$ 的值和梯度的乘积。

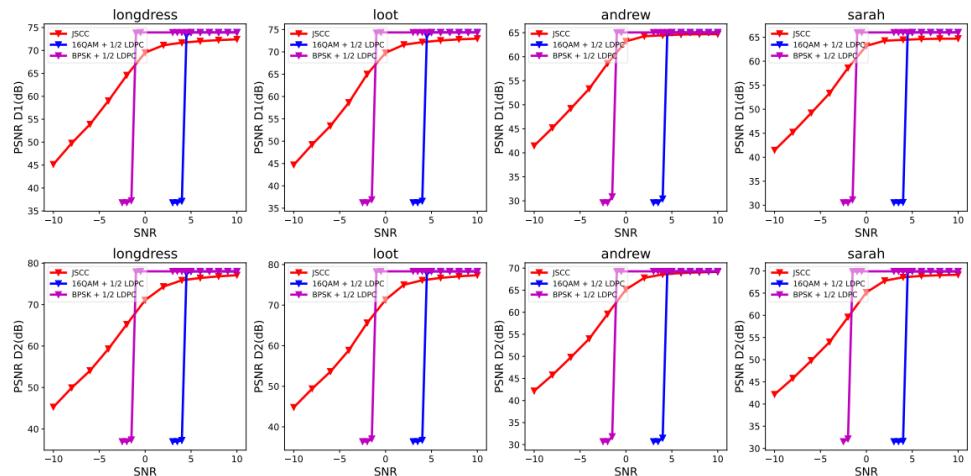


图 4-8 PCSC 在不同信噪比下的性能<sup>[8]</sup>

不同信噪比下的传输性能如图 4-8 所示。PCSC 在信噪比为 10dB 下训练一次，并在各种信噪比下进行测试。CBR 设置为 0.0625。比较了在 AWGN 信道中采用 JSCC 的 PCSC 与 1/2 LDPC+16QAM 和 1/2 LDPC+BPSK 的 PCSC 的差异。在高信噪比下，JSCC 和 LDPC 编码性能稳定，LDPC 编码性能略高于 JSCC 编码。随着信噪比的降低，采用 LDPC 编码的 PCSC 的性能急剧下降，表现为“悬崖效应”。

### 4.1.3 基于模分多址的视频语义通信系统（MDVSC）

视频流媒体的 6G 语义通信<sup>[9]</sup>面临多种挑战，1) 高计算复杂度：传统 MEMC 架构（光流估计等）在像素域操作，计算负担随分辨率/色深剧增，IoT 设备难以实时处理。2) 语义信息捕获低效：原始像素与需传输语义符号间存在鸿沟，语义特征表示效率低下。3) 通信鲁棒性不足：分离编码低 SNR 下“悬崖效应”显著；现有深度 JSCL 方案在复杂无线信道中仍需优化。4) 编码长度控制不精准：现有系统多依赖隐式学习，难以在固定 CBR 约束下精准适配数据量以充分利用信道容量，限制传输效率。5) 延迟抖动影响：无线信道动态性导致延迟抖动，缺乏实时信道状态自适应调整编码策略的有效机制。

本白皮书提出高效无线模型分割视频语义通信系统（MDVSC）<sup>[10]</sup>应对挑战，1) 轻量化设计：采用潜在空间变换和轻量化网络，降低计算负担，适合 IoT。2) 特征提取优化：通过公共特征提取器（CFE）分离并传输一次视频帧组（GOP）的共性特征（如背景纹理），减少冗余传输。3) 提升鲁棒性：利用深度 JSCL 克服“悬崖效应”。实验性能优于传统（H.264/H.265 + LDPC）。4) 精准码率控制：引入基于熵的语义重要性编码，通过熵值显式量化符号重要性并控制编码长度。实验表明近半数据可丢弃而不显著影响性能。5) 缓解延迟抖动：自适应调整编码长度匹配信道容量，确保 CBR 严格及低 SNR 下稳定传输。

#### （1）研究方法

受 MDMA 中共享与个性化语义特征的启发，提出无线模型划分视频语义通信系统（MDVSC）<sup>[17]</sup>，从连续视频帧提取共性语义特征与个性语义特征。为解决延迟抖动问题，提出基于熵的变长编码方案：引入熵值量化每个符号的重要性，据此显式控制编码器输出码长，提升压缩灵活性。

#### （2）整体框架

如图 4-9 所示，MDVSC 系统工作流程如下：输入图片组（GOP） $X_g$ 转化为潜在表示 $L_g$ ；JSCL 编码器提取语义特征图 $L_g$ 中提取出语义特征图 $Y_g$ ；然后经过公共特征提取器 CFE 输出该 GOP 的共性特征图 $W_{g_c}$ 及各帧个性特征图 $W_{g_i}$ 。利用熵模型估计符号熵值，据此对共性/个性特征图进行变长编码生成信道输入符号序列 $S_g$ 。变长编码旨在有限信道约束下最大化传输性能。

解码过程： $S_g$  经过信道后得到失真估计为  $\hat{S}_g$ ，经过 JSCC 解码器输出潜在表示  $\hat{Y}_g$ ，再经过隐空间反变换，映射到原始分布空间，直接恢复生成 GOP  $\hat{X}_g$ 。

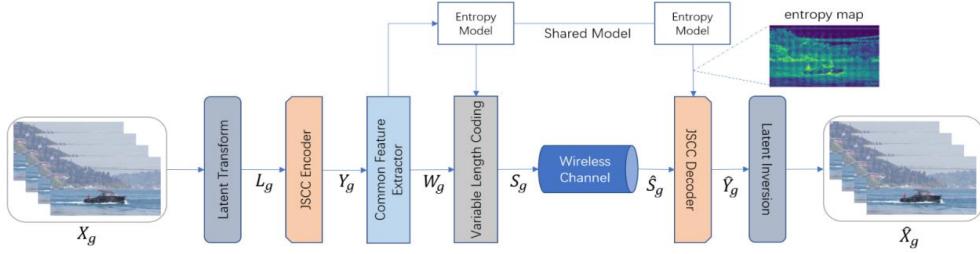


图 4-9 MDVSC 系统框架<sup>[10]</sup>

### (3) 性能分析

如图 4-10 所示，在信道状况变化的情况下，当信噪比波动时，MDVSC 能够较好地抵抗悬崖效应，保证视频内容的正常传输。

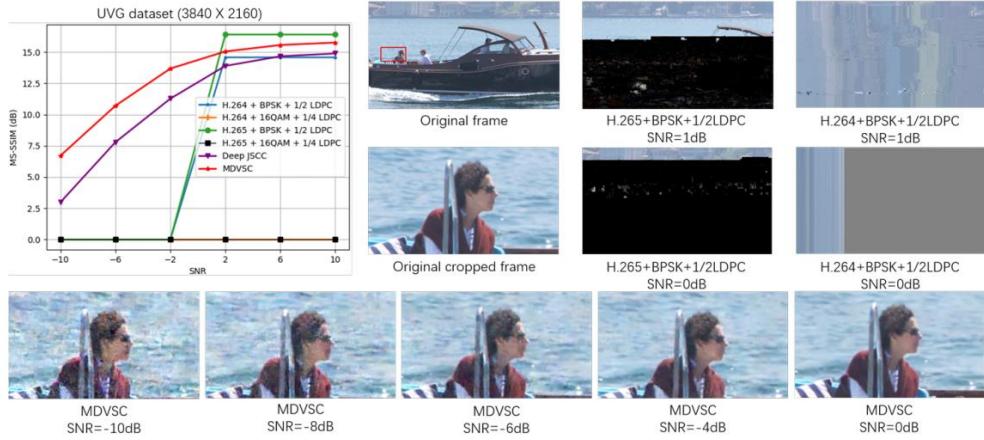


图 4-10 本方案在不同信噪比下的通信性能<sup>[10]</sup>

## 4.2 语义知识库动态匹配与更新

为解决不同模态语义编码模型演化所需的语义知识库的同步问题，需在资源受限的 6G 边缘节点实现自适应语义知识库更新，解决轻量、灵活获取与演化模型的难题<sup>[11-12]</sup>。直接部署大型通用模型或频繁全量传输参数以适应新信源/任务，将带来难以承受的算力、存储与带宽开销，且受限于信道噪声及模型架构僵化。因此，“基于模型组件的智简传输方案”成为关键，其需克服：边缘资源严苛限制、模型更新高成本低效、无线传输可靠性隐患、多任务/多信源高效协同适配。

为应对上述挑战，提出基于语义模型组份（SMC）的模型驱动语义通信系统

以应对，1) 模块化分发：将 AI 模型解耦为通用特征提取器与任务特定模块，仅传输所需增量组件（如增量/跨信源/跨任务 SMC），显著降低带宽与更新时间。2) 跨源跨任务适配：结合迁移/增量学习，复用通用特征提取器知识，扩展特定提取器适配新信源/任务。3) 抗噪鲁棒性：训练中引入噪声注入与正则化，提升参数抗信道噪声能力。4) 高效增量学习：冻结通用特征提取器并扩展特定提取器，支持动态扩展能力且避免灾难性遗忘。

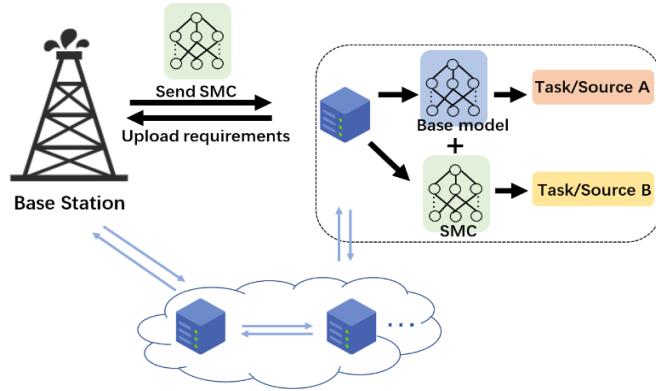


图 4-11 基于组份的语义通信系统，由基础模型和增强模型构成

### (1) 模型组份传输整体方案

在设想的系统中，边缘节点向基站发送语义模型更新需求（如适应新任务/信源）；基站响应需求训练并下发相应模型参数更新包(SMC)；边缘节点将 SMC 与基础模型集成以实现模型更新，获得处理新信源/任务的能力。

如图 4-11 所示，资源受限的边缘节点部署语义模型处理特定信源/任务。基站拥有强大计算/存储能力，负责训练 SMC；边缘节点保有基础模型和少量本地数据用于微调。传输过程包含三阶段，1) 上报需求：边缘节点上传由环境变化/决策产生的更新需求。2) 训练与分发：基站利用其算力、数据及基础模型训练所需 SMC，并将其下发至边缘节点。3) 模型集成：边缘节点将 SMC 集成到基础模型，使组合模型能处理新信源和任务。

### (2) 模型组份学习方法

在类别增量 SMC 训练中，基站可访问原始类别集合 $\{Y\}$ 、新增集合 $\{Y_{SM}\}$ 及其训练数据 $\{D\}$ 和 $\{D_{SM}\}$ 。边缘节点可采用排演策略，保存部分数据于内存 $\{M\}$ 用于微调。AI 模型被解耦为特征提取器与任务执行模块。模型迁移性主要受特征提取器影响：高层神经元更专注于原始任务，泛化能力弱，可能损害新任务性能。因此，需识别原始模型中可泛化至新任务的特征提取器部分并复用，扩展结构以

提升新任务的特征提取能力。增量 SMC 学习分为两阶段：

1) 广义特征提取器的学习：假设基站拥有原始任务和目标任务的模型，学习目标是找到一个能够从这两个模型中提取共享特征的特征提取器。该方法利用典型相关分析 (SVCCA) 计算两个不同任务模型对应网络层激活输出的最大相关系数  $\rho_1$ ，以此量化它们的表达相似度。通过观察  $\rho_1$  在网络深度上的变化（通常浅层更高），并设定一个阈值  $\rho T$ ，将网络的特征提取能力区分为广义特征提取器  $\Phi_G$ （负责共享的、通用的特征）和特殊特征提取器  $\Phi_S$ （负责任务特定的特征）。目标是最终学习或识别出这个通用的  $\Phi_G$ 。

2) 可扩展学习：

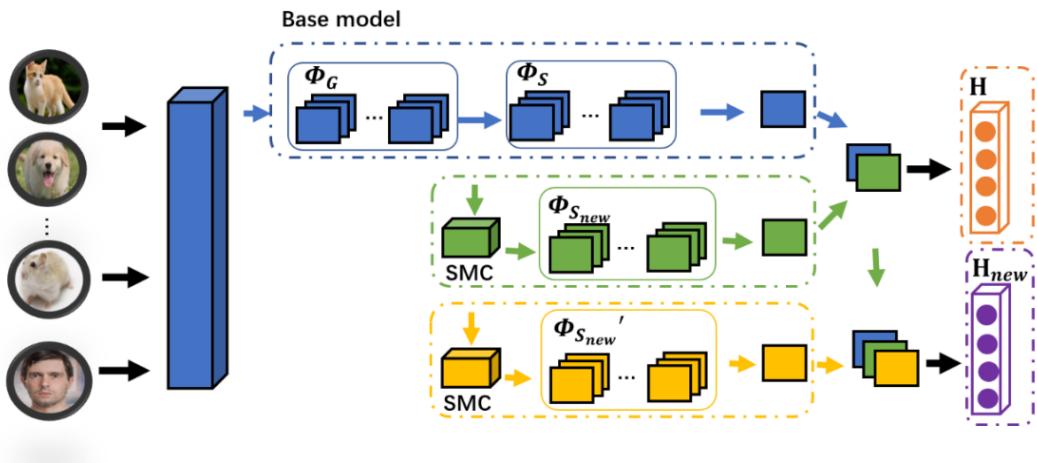


图 4-12 可扩展学习的模型架构

可扩展学习旨在扩展基础模型以适应新任务，同时保护旧任务性能。如图 4-12 所示，基础模型由特征提取器  $\Phi = [\Phi_G, \Phi_S]$  和任务执行模块  $H$  组成。学习新任务时：1) 重用  $\Phi_G$  提取通用特征；2) 冻结  $\Phi_G$  和  $\Phi_S$  的参数以防遗忘；3) 扩展一个新的特殊特征提取器  $\Phi_{S,new}$ ，其参数由  $\Phi_S$  初始化；4) 将  $\Phi_S$  作用于通用特征得到的旧特征与  $\Phi_{S,new}$  作用于通用特征得到的新特征拼接，形成更新后的特殊特征；5) 使用由  $H$  初始化的新任务执行模块  $H_{new}$  处理拼接特征以预测新任务。训练时，在最小化新任务的交叉熵损失基础上，增加一项特征语义距离约束：最大化新特征 ( $\Phi_{S,new}$  输出) 与旧特征 ( $\Phi_S$  输出) 的均值向量之间的距离（如 L2 距离），以迫使  $\Phi_{S,new}$  学习与旧任务不同的新特征，而非重复旧特征。跨信源/跨任务学习采用相同扩展架构，但因不影响旧任务，通常无需冻结和距离约束。

## 5. 智简信息传输使能技术

### 5.1 智能新型网络传输与业务保障技术

#### (1) . 业务特征智识别与智分类技术

业务流分类对于网络异常监控、业务差异化服务保障等具有重要意义，因此需要深入研究业务特征智识别与智分类技术。从网络架构的高层(网络层)出发，精准智能地识别出不同信源或不同模态的业务流和语义流类别，已然成为为各类业务数据流提供有针对性的 QoS 策略、合理安排资源分配、确定编解码等机制的关键前提条件。然而，随着通信与多媒体技术的持续发展，智简网络环境中传输的视频流、音频流、点云等业务在格式特征层面呈现多样化与复杂化趋势。进而通过语义表征得到的语义业务数据流将不仅包括单模态语义，还会涉及多模态语义，其中多模态语义信息在信源编码阶段普遍采用差异化编码策略，这对精准业务分类构成显著挑战。与此同时，互联网应用普遍采用自定义的私有协议对网络智能识别技术提出更高要求。

当前网络流量识别主要依赖基于端口和深度包检测的方法，通过匹配预定义规则实现分类。然而，在 6G 网络复杂环境下，不同模态业务特征差异显著，且语义与语法通信并存，单一规则难以精准识别业务流特性，无法满足网络高效管理需求，亟需更智能的智识别技术以支撑未来网络的精细化管理。

为满足智简网络传输需求并精准识别业务，可以从智简协议栈设计和人工智能技术应用两方面同步开展研究。智简协议栈可以更好地适配语义和语法并存的业务场景。另一方面，在不改变现有协议栈架构的基础上，还可结合人工智能算法，进一步实现对业务的精准识别与分类。

6G 智简网络的“简”体现在多个方面，包括网络架构和接口、网络协议、网络编码、网络平台等。其中，网络协议是贯穿其中的关键，并且在智简网络数据传输的业务类型识别和分类中作用关键，传输类型考虑多模态语义，进一步甄别语义数据流的重要程度。传统网络协议是面向连接设计，协议栈分割且冗余，包括控制面协议与用户面协议不同、协议功能分割复杂、数据处理过程复杂。6G 智简网络的协议将促使控制面协议栈与用户面协议栈融合统一，协议栈功能原子化，

针对不同类型数据业务需求灵活提供不同能力组合的效果。智简传输中数据流从全量冗余向语义表达的转变，协议栈同样去冗余。

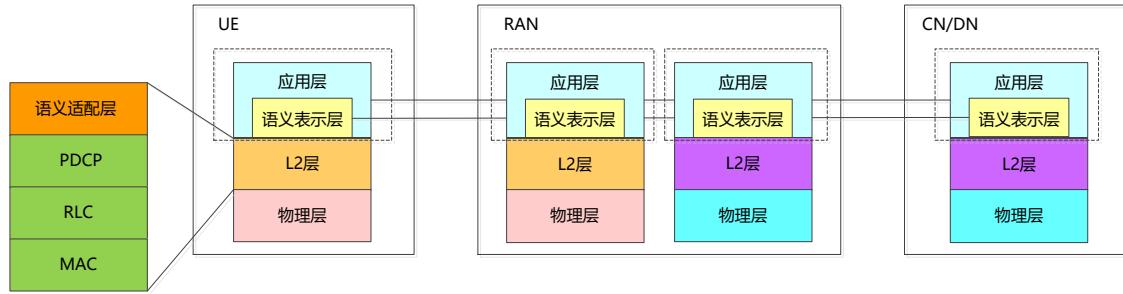


图 5-1 6G 智简网络协议栈

智简网络协议栈的设计按照分层重构的思想出发，采用多维度适配的创新思路。在应用层内新增语义表示层。对应用层的语义库、语义编码方式、语义数据流类型进行标识。重构空口 L2 层，将空口 L2 层协议栈功能面向语义信息特征重构设计，由语法通信中面向连接的设计思路，向兼顾面向任务的数据处理过程设计。增强 SDAP 层为语义适配层，面向语义信息空口传输特征，适配传输速率，面向任务调度资源。物理层处理机制面向语义数据特征进行重构，通过信源信道联合编码实现语法与语义信息流的差异化处理。该设计建议通过全栈层语义特征的显式标识与动态适配，构建具有强任务识别能力的下一代通信协议体系。

深度神经网络利用大量业务数据预先训练人工智能模型，能从原始数据中提炼深层次特征，学习到数据的复杂模式和特征，是赋能网络 AI 用例不可或缺的关键技术。针对业务流特征识别和分类，可选用更适合捕捉数据流特征的一维深度神经网络，实现多模态业务流的特征识别和分类方法。通过构建融合语法特征与语义特征的混合数据集，可训练具备双通道识别能力的业务流分类模型，实现基于语法特征的分类及语义重要性判断的双重识别功能。基于深度神经网络的智能分类方案具有三大技术优势：其一，突破传统规则依赖，无需人工设计特征规则即可自主学习；其二，适配语法通信与语义通信的混合场景，有效应对动态场景变化；其三，通过大数据量训练获得强泛化模型，可以得到更强泛化性的模型，从而在不同的通信技术和流量特征下都能有效分类，不需要针对特定环境进行调整，甚至能够应对加密流量带来的挑战。

进一步可在用户面功能内置基于一维残差网络的业务流识别 AI 模型，实现对实时传输的业务流进行样本采集和智能业务识别，随后将识别结果上报至核心网功能，制定相应的策略，从而为高需求业务提供专属保障。

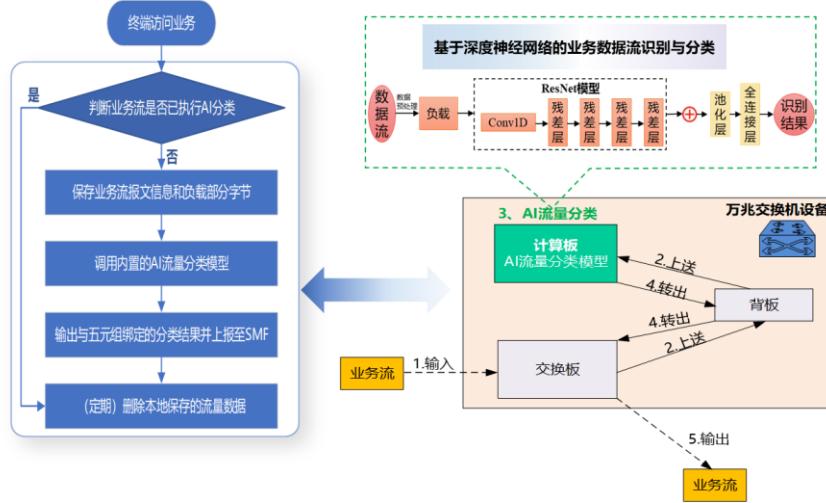


图 5-2 智能用户面功能业务流智能识别

## (2) . 业务数据流的差异化 QoS 控制技术

未来网络将涌现出海量业务，业务对差异化服务质量保障的需求日益凸显。针对语音、视频直播等时延敏感型业务，QoS 的策略制定是决定业务性能表现的关键因素。因此，服务质量保障是至关重要的，网络通过执行有保障的 QoS 机制，以确保通信质量，避免高时延和高丢包率问题。然而，现行的 5G 网络采用静态配置按照模板参数选择 QoS 策略，面对多元化的业务需求短板凸显，无法根据网络环境和用户需求提供灵活、多样化和个性化的 QoS 控制策略。构建动态 QoS 控制机制可通过实时感知用户体验与网络资源状态，实施粒度更精细的差异化保障策略。不仅能提高用户体验，还能优化网络资源的利用效率。特别适用于高优先级业务场景，能精准匹配关键用户的服务需求，为其构建定制化网络性能保障体系。

面向差异化 QoS 控制需求，可以采用一种结合生成对抗网络和强化学习的差异化 QoS 控制机制，实现依据输入的业务数据流特征、动态感知的网络性能分析结果、历史的 QoS 控制策略以及应用程序提供的 QoE 信息，即用户提出的 QoS 要求，灵活确定业务数据流的 QoS 策略。

- 基于生成对抗网络的 QoS 策略生成：实现 QoS 策略训练数据的生成，增加 QoS 样本数量。
- 基于强化学习的 QoE/QoS 映射：使用强化学习算法求解 QoS 和 QoE 的映射关系，使得在给定 QoE 要求时确定至少满足 QoE 的 QoS 策略。

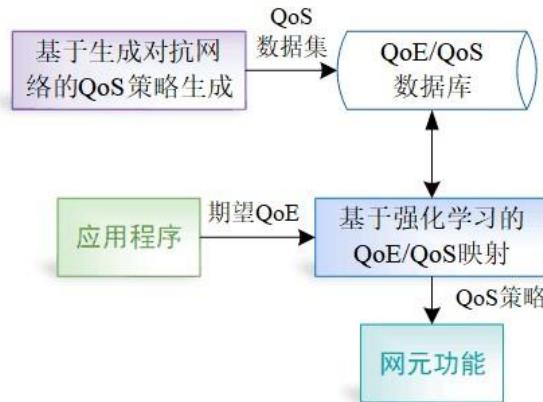


图 5-3 业务数据流的差异化 QoS 控制

利用基于生成式 AI 技术可以对标准 5QI 或 QoS 历史数据进行数据增广，从而生成更细粒度的多种 QoS 指标组合的数据集。这为利用强化学习深入学习 QoS 参数与业务体验之间的复杂关系提供支撑。

进一步地，采用强化学习精准适配用户体验，进行最佳 QoS 策略的选择，实现指定 QoE 业务流的 QoS 策略自生成。简要来说，采用强化学习的奖惩机制来优化 QoS 配置，将网络中收集的数据经过 AI 技术得到业务体验分析和网络性能分析结果作为奖励，而满足 QoS 需求所消耗的网络资源则作为惩罚，从而控制输出的 QoS 配置恰好满足业务体验，在减少网络资源开销的同时实现差异化的 QoS 保障目标。

智简网络可以监控当前采取的 QoS 策略是否能够满足业务需求，如果不能满足，可向 AI 分析功能请求生成辅助的 QoS 策略。AI 分析功能涵盖网络性能分析、QoS 可持续性分析及预期 UE 行为参数分析，分析结果反映网络状态和业务体验，辅助策略控制并增强 QoS。随后，策略功能依据 AI 分析辅助信息更新 QoS 策略。终端、基站和核心网执行更新的 QoS 策略后，网络状态和业务体验便会随之改变。进一步，AI 分析功能收集新 QoS 策略执行后网络中与分析需求相关的数据，将业务体验与网络性能分析结果反馈给策略生成功能，其利用强化学习思想，以分

析结果为奖励和状态反馈，进一步训练生成式 AI 模型生成优化策略并实施。经多次迭代，网络执行理想 QoS 策略，该策略存入历史库，相同业务发起时可依此初始化参数，减少迭代。

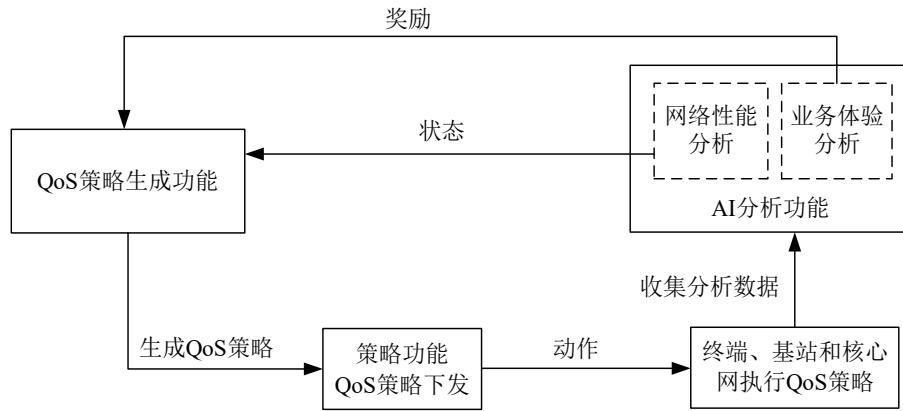


图 5-4 基于智能功能辅助的差异化 QoS 控制

此外，采用强化学习奖惩机制，当惩罚变量越小时，如果仍能满足业务体验需求，则反映了所用网络资源越少，从而达到控制输出的 QoS 配置恰好满足业务体验，在减少网络资源开销的同时实现差异化的 QoS 保障目标。

## 5.2 多维资源感知与智能调度技术

### (1) . 多维网络资源与状态智能感知技术

在 6G 智简网络传输体系中，网络节点的拓扑结构和网络的资源与状态是影响网络资源的调度与分配效率的关键因素。为实现网络节点资源与能力的全局优化利用，必须建立对网络环境的整体拓扑认知，并在此基础上获取网络的计算、存储、传输等资源与状态。

针对目前通信网络中，网络资源与状态的采集、分析和存储等功能比较分散，并未定义统一的资源与状态获取方案，难以形成对网络的及时、有效状态感知，无法满足新型网络架构对资源与状态智能感知、按需资源调度和分配等的需求问题。提出了多维融合网络统一资源图谱的生成与表征方法，并在网络中增加网络资源与状态信息感知功能，实现资源与状态的智能按需获取或自动上报。

对 6G 智简网络架构的基本网元/NF 进行了重构，增加内生的网络资源获取功能，结合具体的新型网络架构，来进行多维网络资源与状态的按需或自动获取，

并上报给资源与状态的管理和存储功能。6G 网络的标准化还在进行中，并未有确定的网络架构，作为网络资源感知核心功能的网络资源获取功能与 6G 智简网络架构的一种可能的结合方式如图 5-5 示：

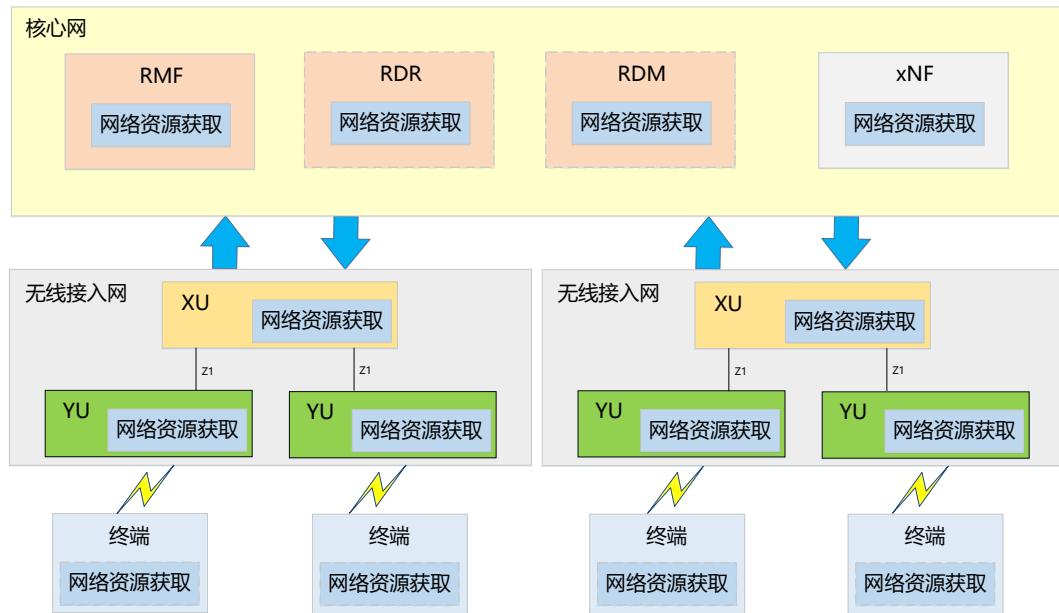


图 5-5 6G 智简网络架构中的网络资源获取功能

需要获取网络资源的网元或节点，内生具有资源状态感知和上报功能，此处体现为网络资源获取功能。

- 终端中网络资源获取是可选功能，如终端不能部署网络资源获取功能，则仍需通过传统方式来完成网络资源与状态的获取。
- 无线接入网的下一代基站（xNB）具有各种可能的逻辑形式，可以是 5G 的单体式 gNB 形式，也可以是 CU、DU 分离的 gNB 形式，还可以是类似 O-RAN 的 CU、DU、RU 分离形式，也可能进化为其它（例如：XU 和 YU）结构形式或演进为服务化架构。但无论基站以何种形式展现，在组成基站的逻辑单元内，都需内生网络资源获取功能。如无线接入网演进为服务化架构（例如：RAN 控制面服务化），为支持资源与状态感知，则 RAN 侧 NF 中也都内生具有网络资源获取功能。
- 核心网侧 NF 中内生具有网络资源获取功能，来完成资源与状态的获取。

6G 智简网络中的网络资源获取功能，可以是网元的可编程机制的一个运行实例，也可以是网元支持内生 AI 特性的一个模块或子功能，取决于具体的网络架构设计。

多维资源的感知结果，包含网络拓扑和资源与状态信息，通过网络资源管理功能保存到网络资源与状态库中。在前述的网络架构中，体现为核心网侧 NF 的网络资源管理功能（RMF）、网络资源与状态数据存储（RDR）和网络资源与状态数据管理（RDM）中。可选地，网络资源与状态的存储功能可以复用核心网的 UDR/UDM 实现；或者网络资源管理功能和网络资源与状态库还可能位于无线接入网和核心网都能访问的 OAM 中。资源感知结果在网络资源与状态库中的具体表征形式可能为结构化数据、图数据或者知识图谱。

在网络中增加网络资源与状态信息感知功能，生成和表征多维融合统一的资源图谱，可实现资源与状态的智能按需获取或自动上报。适用于多维网络资源与状态的智能感知，为 6G 智简信息传输中资源调度和分配等提供依据。

## （2）. 资源智能统一调度技术

6G 智简网络中，资源的智能统一调度，需要基于网络资源与状态感知的资源图谱，应用现代人工智能技术进行精准的资源调度和分配，实现资源的动态优化和网络性能提升。资源图谱中感知的资源，可能位于端到端的网络节点和传输链路。网络节点的传输资源，包括网络带宽、发射功率等。网络带宽状态包括理论带宽和可用带宽，发射功率状态包括最大功率和可用功率等。点对点的网络资源与状态可能包括传输速率、时延范围、抖动范围、最大误码率等。6G 智简网络中的语义通信部分，涉及语义库、语义特征提取和语义压缩等特有影响因素，这些对于资源的调度和分配策略既是制约条件，又可能是可以优化、改进的可利用的调控指标。

在移动通信数据传输过程中，无线资源的调度和分配，是影响多用户、多业务流的服务质量保障的关键因素。无线资源管理一般包括带宽分配、功率控制等。以基站和多终端间的语义通信场景下的无线资源的调度和分配为例，为了提高系统的业务 QoS 支撑能力，在当前系统状态下，基于 QoS 控制策略，合理分配资源给不同关键程度的语义信息，从而可以支持不同信源信息的差异化传输，在保证用户体验的前提下，提高系统对不同网络条件的适应能力。

在智简信息传输背景下，需要对业务 QoS 支撑能力进行解读。而现有移动通信中的业务 QoS 支撑能力并未有确切定义，也没有对业务 QoS 支撑能力提升制定明确对应的指标与标准。经过分析和论证，采用流比特速率、最大丢包率作为 QoS 支撑能力的衡量指标。当信息传输效用越高，同样网络条件下可传输的关键语义信息越多，可以有效减少因拥塞导致的丢包，降低业务丢包率，从而支持系统所能容忍更大的丢包率，支持流比特速率相比传统通信业务数据传输需求降低，相应可实现最多支持的业务流数提升，对系统 QoS 支撑能力越强。

方案针对多用户的业务数据流，在已分类和标记重要性的语义通信数据传输过程中，统筹考虑 UE 需求、UE 周围基站信号质量及基站负载情况，满足当前可用无线传输资源（可用功率、可用带宽）和用户体验限制的条件下，基于系统自生成的 QoS 控制策略，以最大化信息传输效用，兼顾用户公平性和资源利用率为目，采用多信息权重下基于 AI 的网络资源调度方法来确定无线资源调度和分配策略，从而提高系统重要语义信息的传输保障能力，支持容忍更大的丢包率，实现最多支持的业务流数提升，提高 QoS 支撑能力。

基于 AI 的网络资源调度算法，以采用强化学习算法为例，考虑到要优化调度和分配的无线资源中的带宽和功率均为连续变量，为有效处理连续动作空间，可以采用双延迟深度确定性策略梯度（TD3）算法，进行训练得到资源分配策略用于语义数据传输的无线资源分配过程。多信息权重下基于强化学习的网络资源调度方法可以对无线资源分配进行更细粒度的配置，例如：对传输带宽，可以根

据不同语义重要程度以物理资源块（PRB）级别进行分配。其算法概要框图如

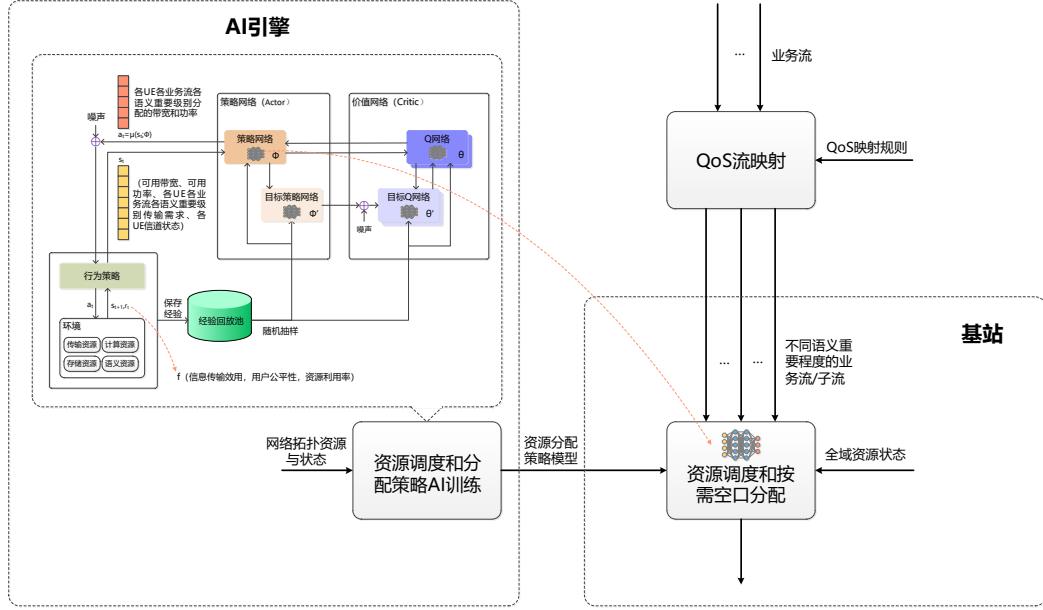


图 5-6 所示：

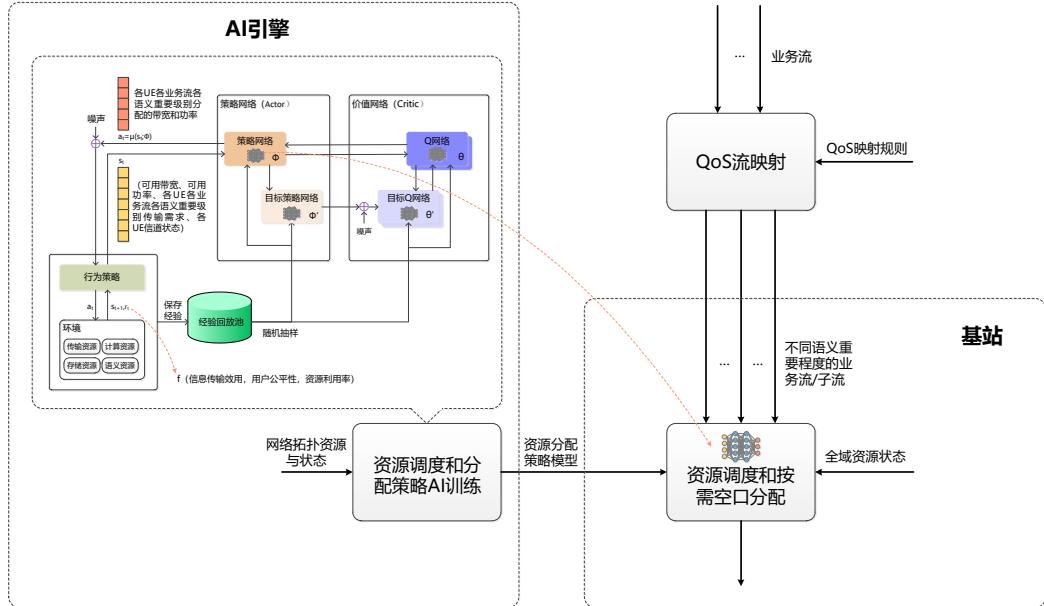


图 5-6 多信息权重下基于强化学习的网络资源调度

其中，强化学习算法需要考虑的要素有：

- 智能体：考虑到资源调动和分配在基站执行，将基站作为智能体；
- 状态空间：需要从基站或智简传输资源管理的网络与状态库获取可用无线资源和各 UE 传输需求，状态空间  $S$  需要考虑系统中当前的可用传输资源（基站可用带宽、可用功率），以及各 UE 各业务流不同语义重要级别的传输需求；

- 动作空间：需要考虑为各 UE 各业务流不同语义重要级别分配的传输资源（发射功率和带宽（PRB））；
- 奖励：当基站执行动作作为待传输数据分配新的无线资源后，会根据信息传输效用、用户公平性和资源利用率计算得到反馈的即时奖励；
- 算法优化目标：以累积的奖励最大化为优化目标；
- 策略函数：算法训练收敛后得到无线资源分配策略，即根据当前的网络状态、待传输的语义数据需求、QoS 控制策略等，可直接得到动作，即无线资源（发射功率、带宽）分配值。

基于强化学习的资源调度和按需空口配置方法，在系统方面实现资源的高效使用，业务方面实现 QoS 保障的提升，尤其适用于信道质量较差，语法通信难以传送数据时，而本方法仍能传输。在资源受限场景，可优先保证重要信息数据的传输，提高通信的效率和可靠性；在资源可控场景，可保留更多的背景和基础信息，提升用户体验。

## 6. 总结与未来展望

本白皮书聚焦 6G 智简信息传输及使能技术，围绕新型智简传输理论、多模态自适应语义通信、多模态业务适配的智简信息传输保障三大方向，系统构建了 6G 智简传输的理论框架、技术路径和使能技术体系，以突破传统信息传输技术的局限，为未来 6G 网络多模态数据智简传输奠定了理论与技术基础，展现出广阔的应用前景。

智简信息传输以达意传输为目标催生了语义通信这一关键技术。语义通信将推动通信系统向更高效、更精简、更智能的方向演进，但其引入也对现有网络架构与机制提出了新的挑战，未来研究可对以下关键问题深入探索：

- 1) 网络架构演进：支持语义与语法协同传输。语义知识库的构建、存储和同步，信源信道编码方式等的引入，将重塑智简信息传输架构。需深入探索语义与传统语法传输的共生路径，构建将语义通信技术框架融入现有网络的架构设计，为后续研究提供坚实支撑。

- 2) 协议栈重构: 支持跨层语义协同处理。语义通信不仅关注信源的内容, 又要兼顾信道, 传统协议栈的分层设计难以满足语义传输编码需求, 需重构协议栈, 打破传统分层界限, 促使各层能够协同处理语义信息, 实现跨层语义信息交互与动态优化。
- 3) 服务质量保障机制创新: 传统 QoS 机制无法适配语义传输特性, 需进一步探索建立兼顾语义优先级与比特公平性的统一服务质量评估与保障体系。
- 4) 端边云算力协同机制: 赋能语义通信的强计算需求。从原始信号中识别、提取和解释语义信息需要强大算力, 现有网络难以满足其对算力的大规模、高效率需求。需探索高效的算力、AI 资源共享方式, 全面支撑语义通信的实现。
- 5) 强化安全可信机制: 为了保障语义知识库的同步, 需要在收发两端建立模型更新共享的机制, 知识库的传递可能涉及到隐私安全问题, 需要评估并构建与之适配的数据一致性和安全可信机制。

由上可知, 要支持未来 6G 网络多模态数据智简传输, 需要传统网络架构发生根本性变革, 如支持 AI 新能力、网络协议栈功能可按需重构等。因此, 未来仍需要与学术界和产业界协同努力, 共同推动智简传输系统在未来 6G 网络的可行部署与应用落地。

## 缩略语列表

缩略语	英文全名	中文解释
5QI	5G QoS Identifier	5G业务质量标识
AI	Artificial Intelligence	人工智能
AWGN	Additive White Gaussian Noise	加性高斯白噪声
CN	Core Network	核心网
CU	Central Unit	中央单元
DeepJSCC	Deep Joint Source–Channel Coding	深度联合信源信道编码
DU	Distributed Unit	分布单元
GOP	Group Of Picture,	图片组
LSCI	Layer-based Semantic Communication system for Images	基于层的图像语义通信系统
MDMA	Model Division Multiple Access	模分多址
MDVSC	Model Division Video Semantic Communication	模型分割视频语义通信
MAC	Media Access Control	媒体访问控制
NOMA	Non-Orthogonal Multiple Access	非正交多址
NF	Network Function	网络功能
OAM	Operation, Administration, and Management	运行、管理和维护
O-MDMA	Orthogonal Model Division Multiple Access	准正交模分多址
O-RAN	Open Radio Access Network	开放无线接入网络
PCSC	Point Cloud-based Semantic Communication	点云语义通信
PDCP	Packet Data Convergence Protocol	分组数据汇聚层协议
PRB	Physical Resource Block	物理资源块
PSNR	Peak Signal-to-Noise Ratio	峰值信噪比
QoE	Quality of Experience	体验质量
QoS	Quality of Service	服务质量
RAN	Radio Access Network	无线接入网
RDM	Resource Data Management	资源与状态数据管理

RDR	Resource Data Repository	资源与状态数据存储
ResNet	Residual Neural Network	残差网络
RLC	Radio Link Control	无线链路控制
RMF	Resource Management Function	资源管理功能
SBA	Service Based Architecture	服务化架构
SDAP	Service Data Adaptation Protocol	服务数据适配协议
SeSM	Semantic Slice-Models	语义切片模型
SMC	Semantic Model Components	语义模型组份
TD3	Twin Delayed Deep Deterministic Policy Gradient	双延迟深度确定性策略 梯度
UDM	Unified Data Management	统一数据管理
UDR	Unified Data Repository	统一数据存储
UE	User Equipment	用户设备

## 参考文献

- [1] Mao Y, Dizdar O, Clerckx B, et al. Rate-Splitting Multiple Access: Fundamentals, Survey, and Future Research Trends[J]. *IEEE Communications Surveys & Tutorials*, 2022, 24(4): 2073–2126.
- [2] Zhang P, XU X, Dong C, et al. Model division multiple access for semantic communications[J]. *Frontiers of Information Technology & Electronic Engineering*, 2023, 24: 801–812.
- [3] Liang H, Liu K, Liu X, et al. Orthogonal Model Division Multiple Access[J]. *IEEE Transactions on Wireless Communications*, 2023, Major Revision.
- [4] Richardson I E G. H. 264 MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia[M]. Hoboken, NJ, USA: Wiley, 2004.
- [5] Huang D, Tao X, Gao F, et al. Deep learning-based image semantic coding for semantic communications[C]//2021 IEEE Global Communications Conference (GLOBECOM) (GLOBECOM), Madrid, Spain: IEEE, 2021: 1–6.
- [6] Dong C, Liang H, Xu X, et al. Semantic Communication System Based on Semantic Slice Models Propagation[J]. *IEEE Journal on Selected Areas in Communications*, 2023, 41(1): 202–213.
- [7] Schwarz S, Preda M, Baroncini V, et al. Emerging mpeg standards for point cloud compression[J]. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2018, 9(1): 133 – 148.
- [8] X. Liu, H. Liang, Z. Bao, C. Dong and X. Xu. A Semantic Communication System for Point Cloud[J]. *IEEE Transactions on Vehicular Technology*, 2025, 74(1): 894–910.