

智简通信系统及关键技术白皮书

中关村泛联移动通信技术创新应用研究院、北京邮电大学

全球B5G/6G创新论坛

中关村论坛

2023年5月30日

目 录

第一章：前言	3
第二章：智简通信愿景和驱动力	4
第三章：智简通信系统	6
1. 信息的利用层次	6
2. 智简通信系统架构	7
2.1智简通信系统链路结构	8
2.2智简通信网络架构	9
3. 智简通信协议	11
3.1智简通信协议架构	11
3.2智简模型切分传输技术	12
3.3智简通信路由技术	14
4. 智简通信系统指标体系	16
4.1文本对象	16
4.2图像和视频对象	17
4.3语音对象	18
4.4其他指标	18
第四章：智简通信关键技术	20
1. 信源信道联合编解码技术	20
2. 模型切分技术	21
3. 模型组建与更新机制	23
4. 智简内生安全技术	25
第五章 未来发展展望	28

第一章：前言

第六代移动通信系统（6G）关键技术的研究尚处于初期阶段，在国际形势变化及双循环战略大背景下，中关村泛联移动通信技术创新应用研究院与北京邮电大学等聚焦国家中长期发展的战略需求，开展了面向 6G 的“智慧内生”、“原生简约”的移动通信系统研究，目标是短期内能够缓解高端芯片“卡脖子”难题，长期实现理论原始创新和新范式生态构建，赋能移动通信可持续发展。

《智简通信系统及关键技术白皮书》针对如何实现“智慧内生”和“原生简约”的 6G 网络愿景，提出了智简通信系统这一新的通信理论和系统构建方案，围绕“节点极智、链路极柔、网络极简”的愿景，从智简通信系统架构、协议设计、指标体系以及关键技术等多方面进行了阐述，并对智简通信系统及关键技术未来演进进行了展望。

本白皮书是中关村泛联移动通信技术创新应用研究院与北京邮电大学联合对未来智简通信系统的初步技术展望，后续将持续迭代与优化，希望为学术界与业界技术发展提供参考。

本白皮书版权属于中关村泛联移动通信技术创新应用研究院与北京邮电大学所有并受法律保护，未经授权，任何单位或个人不得复制或拷贝本白皮书中部分或全部内容。引用本白皮书中的文字、数据、图片或者观点时，应注明来源。

第二章：智简通信愿景和驱动力

信息通信技术和人工智能技术（AI）的发展与融合推动着人类社会向数字智能新阶段演进。第五代（the fifth generation mobile communication system, 5G）移动通信网络通过有机地融合泛在通信、计算与控制（ubiquitous communication mobile communication system, computing, and control, UC³）能力，开启了人机物互联的新范式 [1]。而第六代（the sixth generation, 6G）移动通信网络则将对物理世界的人、物、事件等所有要素数字化打造一个与之对应的数字化“虚拟世界”，并在物理世界和数字世界之间提供即时、高效和智能的超连接来重塑世界，从而开启物理维度上的实体世界和信息维度上的数字世界共生共存、虚实交融的移动通信新征程 [2]。

目前，全球范围内 6G 的研究仍处于基本概念与系统框架的确立、关键技术的辨识发现及认证的起步阶段。随着 6G 白皮书的相继发布，业界对于 6G 的发展概念与推进路径越来越清晰，如 6G 的五大指标：沉浸化、智慧化、全域化三大趋势和沉浸式扩展现实、全息通信、感官互联、智慧交互、通信感知、普惠智能、数字孪生、全域覆盖八大应用场景等 [3]。移动通信新技术的发展和 6G 新需求的拉动成为 6G 移动通信发展的主要驱动力，6G 系统将具有如下特点：

1. 多模态信息的高效通信需求：在多媒体业务迅猛发展的数字时代，6G 移动网络中图像、视频等多媒体内容将空前繁荣，6G 业务所占数据量必将大幅增加且其信源格式将更为多样化[4]。此外，6G 业务中人机物与其对应的数字化世界的交互需要协同对多模态的信息资源以不同的表现形式，在移动网络上实现传输和处理，满足 6G 业务场景虚实结合、实时交互的新特点[5]。然而，自从香农奠基信息论以来，移动通信技术的发展已经逐步逼近通信理论极限，需要对现有的信息论进行扩展，构建语义智能表征方案和传输机制，突破通信链路的传输极限，应对未来 6G 更高的需求。

2. 网络自主认知与智能重塑的需求：6G 泛在互联网的特性使得 6G 移动通信网络呈现出高度复杂、动态变化的特征。依赖精准数学模型和理想先验假设的传统移动通信网络，很难适用于未来复杂多变的 6G 场景，急需由传统的“精准数学建模”的网络构建理念向到“新型关系建模”转变，构建具有自我学习、自

我适应、自我演进等原生智能能力的 6G 泛在无线网络，实现 6G 网络的智简构建、重塑和演化 [5]。

3. 高效、简洁的网络形态需求：6G 业务服务对象维度的进一步增加、网络智能节点间的复杂交互、信息流间的动态交换及高维信息空间中的动态变化，使得资源堆叠来提升网络能力的传统网络的复杂度迅速提高。为了突破这一壁垒，6G 需要进一步向智慧内生、原生简约的无线通信网络演进。此外，以香农理论为基础的传统分析方法难以对智能化网络进行量化分析，亟待立一整套机理分析、信息建模与网络优化理论，从整体优化与系统设计观点出发，构建以系统熵为度量指标的网络有序度分析方法，为未来信息与通信研究奠定理论基础。

综上所述，在面向 6G 的演进过程中，其全新的应用场景、极致个性化的性能需求以及复杂动态变化的网络特性将共同驱动原生智能与通信的进一步融合，智简通信系统及其关键技术将发挥重要作用，促进 6G 网络最终形成“智慧内生”和“原生简约”的发展愿景。

第三章：智简通信系统

1. 信息的利用层次

为应对未来万物智联和行业数字化的发展需求，面向 6G 的网络演进需突破移动网络的原有架构体制，6G 移动通信系统不从源头重新思考通信的本质问题，实现理论的突破和技术的变革，智简通信系统应运而生。智简通信系统是张平院士团队提出的新的表达方式[6]，代表“智慧内生、原生简约”。具体地，智简通信是指网络中的节点将成为具备智能的新型节点，而网络本身的协议结构将趋向于极简，通过原生智能、认知重塑等特性支撑网络，围绕不同通信对象构建有针对性的智能服务生态，最终达到网络“由智生简、以简促智”的自演进、自优化、自平衡的状态[5]。

首先，智简通信系统突破经典信息论只研究语法信息在研究范畴、研究层次与研究维度方面的局限，扩展信息研究的层次，从语法信息深入到语义信息。如图 1 所示，香农和威沃指出通信应该包含语法、语义和语用三个层次的内容，即通信符号如何准确地传输，传输的符号如何精确地表达期望的含义，传递意义如何有效地影响期望的行为 [7]。与语法通信不同，语义通信不要求比特级的无差错传输而是以数据的信息含义在通信双方之间准确传递为目标。通过结合人工智能、类脑智能等技术，语义通信可以预先理解业务需求与环境，对信源的语义特征进行理解、提取及传输，同时保证信宿可以理解收到的信源语义特征，从而成功恢复基于语义的信源信息 [2, 5]。通过对语义特征的准确提取与高效传输，语义通信将大幅度减少 6G 全新应用场景中视频、图片等大带宽业务的传输带宽需求，从而大幅度提高通信效率、改进用户体验。

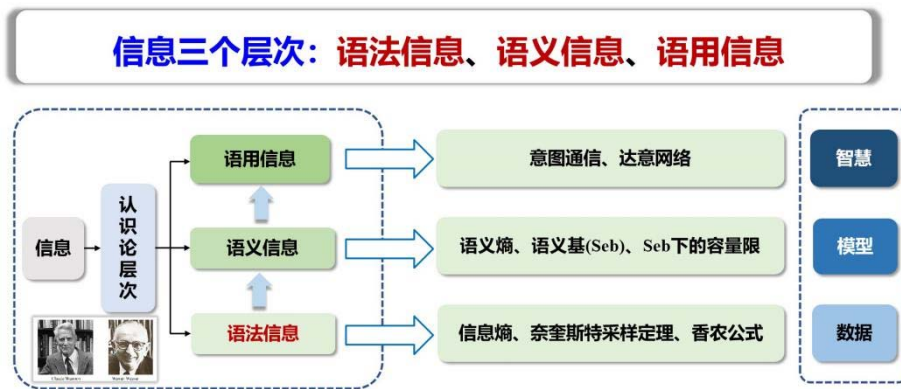


图 1：信息的三个层次：语法、语义、语用；对应数据、模型和智慧的传递

其次，智简语义通信模式使得 6G 通信从信号的相互传递向为意图的相互交流转变。随着通信设备的智能化水平和认知能力不断增强，以语义通信为基础的智能体语言将会促进智能体之间的高效合作、自主交互、意图理解，并灵活、迅速地构建通信与网络，为构建未来“达意网络”奠定了基础。此外，智简通信技术正向着与人工智能、脑信息处理机制结合的方向进一步发展，利用深度神经网络的多层结构以及层次化抽象机制与人脑信息处理的层次化抽象机制的共通性，研究脑皮层来启发全新的机器智能，并有可能从根本上解决基于语法的传统通信系统中存在的跨系统、跨协议、跨网络、跨人机涉及的不兼容和难互通等问题。

最后，针对 6G 网络通信对象间的复杂交互、信息流间的动态交换及高维信息空间中的动态变化的特征，智简通信系统将持续地探究和利用新的智能本原，探索广义信息论等基础理论，主动地以系统熵减为全局优化目标，自适应地重塑信息系统的核心模型，最终实现网络自身的智慧内生、原生简约。

2. 智简通信系统架构

在传统通信系统基础上，智简通信系统将进一步考虑了信源信道的联合处理，在传统信源编码基础上创新性的增加了新维度的特征挖掘，实现了通信系统由传统传输比特演进为传输“模型”。其次，智简通信提出通过“模型传输”实现“智慧互联”的重要概念。此外，基于人工智能技术与信息论及通信技术，智简通信系统建立全新的通信系统收发信机模块，最终实现节点极智、链路极柔、网络极简的新型通信系统及网络。下面，我们将分别介绍智简通信系统的链路和网络总体架构。

2.1 智简通信系统链路结构

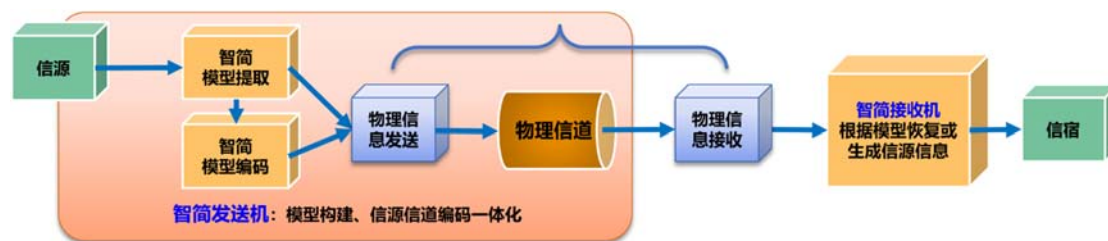


图 2: 智简通信系统链路结构

(1) 智简发射机设计

如图 2 所示，模型驱动的智能通信系统可以构建在传统物理层之上，发送端采用了新型的智简发送机，通过智简模型提取及智简模型编码从信源信息中提取语义等信息，并将提取的语义信息和用于提取语义信息的智简模型联合发送到物理信息发送模块中，从而实现模型构建、信源信道编码一体化能力。在智简发送机中，智能通信系统将考虑物理环境、频谱环境、电磁环境等都因素，对信源、环境、模型中的信息进行提取和整合产生信源信道一体化编码序列，以促进语义信息在无线信道中的传输。其中，基于一体化设计的信源信道联合编码机制将语义推理与物理层通信相结合，可以实现语义层面的纠错，从而保障了智能通信系统端到端的传输性能。智能接收机根据接收到的语义信息及模型，实现信源信息的提取与恢复。

具体地，如图 3 所示，智能发射机将构建统一的智能语义提取与编码模块，能够根据不同的信息类型，选择相应的语义模型，进而将信息分割成不同的语义基，再将综合信源语义、环境因素以及模型的信道编码序列发送至物理信道进行传播。根据信源冗余特性，信源端可以采用不同结构的深度学习网络模型。例如，时序以及文本信源可以采用循环神经网络（Recurrent Neural Network, RNN）模型、图像信源可以采用卷积神经网络（Convolutional Neural Network, CNN）模型、图数据源采用图卷积神经网络（Graph Convolutional Neural Network, GCN）模型。

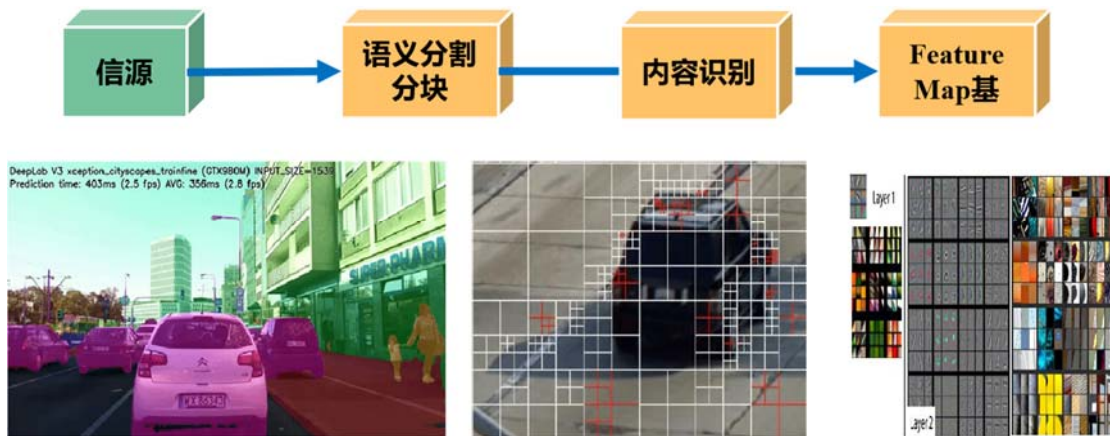


图 3：智简通信发射端信息处理

若信源具有多模态或异构性，则语义提取编码时还需要对多源数据进行语义综合。此外，智简通信系统也可以基于类脑智能的快速发展，通过借鉴类脑神经感知技术，借鉴人脑对多模态信息理解的神经机制，构建多模态协同编解码模型，以实现多媒体语义信息的精确提取、高效传输与恢复。

(2) 智简接收机设计

在接收端，信道输出的信号首先被送入环境联合设计模块以消除模型、环境等对语义恢复的影响。然后，信宿将从环境联合设计模块中输出的信息和接收端处的模型将被送入语义解码和恢复模块，以实现对接收的智简语义信息重建 [8]。

在智简通信系统中，信息在语义层面的分析、编码、解码和重组过程均依赖模型的驱动，且收发端可以通过传输模型的方式实现收发端的知识库共享和智慧的流动。若接收端没有相匹配的语义解码与恢复模型，接收端可以请求发送端或网络节点传输最新的智能模型。此外，智简通信系统通过压缩模型，或者只发送部分模型，即模型切片等方式，实现数据传输量的进一步降低。

2.2 智简通信网络架构

为满足 6G 全新的应用场景、极致个性化的性能需求以及复杂动态变化的网络特性带来的需求，智简通信系统采用以下三层架构：

- 1) 大型服务器：位于第一层的大型服务器可以部署在核心网、骨干网或者

是网络中具有强计算能力的设备上，如云计算服务器，cloudlet 小云片等。大型服务器将基于庞大的数据库和极强的模型训练能力持续产生最高效、稳健、可靠的模型。通过构建适当的激励机制，大型节点将与第二层和第三层的节点进行模型交互。

2) 普通节点：位于第二层的普通节点包括部署在接入网的具备计算能力的专用计算服务器、卫星或个人计算机、智能手机、智能车辆、智能船只、无人机等网络中具有较强计算能力的设备等。普通节点的功能在于基于迁移学习、联邦学习、对比学习等人工智能技术，进行中小型智能模型的联合训练与更新。例如，第二层的节点可以充分利用从不同大型服务器获取的模型，联合训练一个新的模型，以获得更高的性能。此外，普通节点可以采用知识蒸馏等技术压缩模型体积，使人工智能模型可以在有限带宽的无线信道内传输。

3) 部署节点：位于第三层的部署节点指网络中计算资源、能量、存储资源等受限的设备，如可穿戴设备、嵌入设备和传感器等。部署节点可以用于智能感知和收集数据和基于获取的模型完成通信和人工智能任务，但没有进行模型训练的资源与计算能力。

基于以上三层架构，人工智能模型可以在智简通信网络中多个位置进行训练，迁移，传播。具体地，大型服务器和普通节点可以利用本地存储的数据直接训练模型；也可以利用本地存储的模型，结合本地存储的数据，训练生成新的模型。大型服务器与其他大型服务器和部分普通节点之间可采用有线连接进行通信，普通节点与其他普通节点之间可采用有线连接或无线连接进行通信，部署节点与其他部署节点或普通节点之间可采用无线连接进行通信。

结合智简链路及网络架构，智简通信过程可由模型的需求者或者模型的生成者发起。

1) 由模型需求者发起的模型传播：智简节点可以周期性地（如为实现自身模型演进，订阅模型更新）或非周期性地（如大型服务器和普通节点需要其他模型来生成新的模型，或发现自身模型过期时）向智简网络对所需的一个或多个模型的发起传播请求

2) 由模型生成者发起的模型传播：当大型服务器或普通节点生成新的模型

后，可以向特定结点发送模型传播请求，要求其接收该新模型，使结点之间的模型保持统一。

当智简网络在接收模型传播请求后，将运行模型传播算法，制定最优传输策略，确定传输路径将所需模型或模型切片传输至目的节点。值得注意的是，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 N 个切片在多路进行传输，并于信宿端重新组建。其模型切分传输的方式一方面平衡了无线通信网络负载，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。

3. 智简通信系统协议

3.1 智简通信系统协议架构

智简通信系统基于现有网络通信基础架构，并进一步结合 6G 业务需求和 6G 移动通信特点，制定了智简通信双方完成通信或服务所需遵循的规则和约定，即智简通信系统协议。

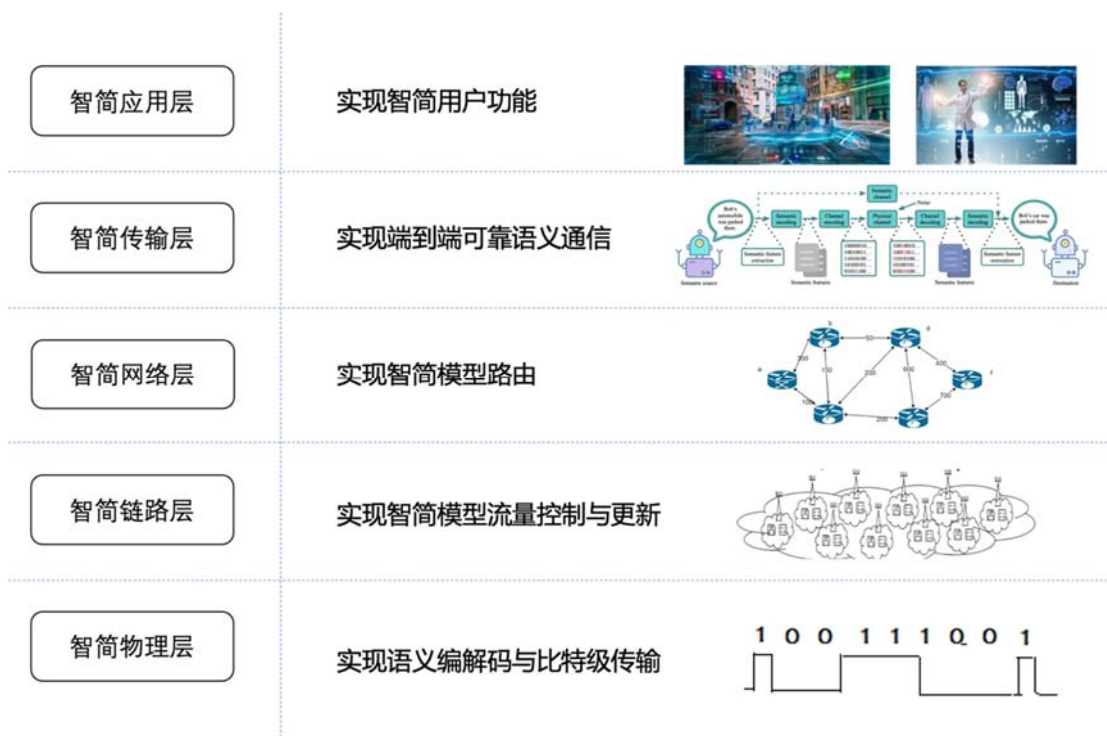


图 4：智简通信系统协议架构

如图 4 所示，智简通信系统协议整体架构分为智简物理层、智简链路层、智简网络层、智简传输层与智简应用层。区别于传统物理层，智简物理层除了完成传统比特信息的传递以外，还将实现语义提取恢复与语义编解码，结合信源信道联合编码完成智简物理层功能。智简链路层主要负责实现智简模型流量控制与更新。智简网络层主要负责动态规划智简模型传播路由，根据一定的协议将智简模型按照传输路径进行传播。智简传输层则负责实现端到端的可靠语义通信，保证语义通信与智简模型的正确传输。智简应用层则负责实现智简用户的功能，包括基于模型向用户提供相应的智简服务，以满足用户的应用需求。

3.2 智简模型切分传输技术

模型作为智简通信信息传输的核心方式，可以存储在智简通信的服务器等云端和用户侧，模型的高效传播是实现智简通信“智慧流动”的必要保证。为了满足 6G 全新的应用场景和个性化的极致性能需求，智简通信系统需要为 6G 用户提供“定制化”地模型，并制定高效的模型传输机制。进一步地，智简通信系统将结合“残差传播”理念，对智简模型进行切分传输，一方面平衡无线通信系统负载，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。

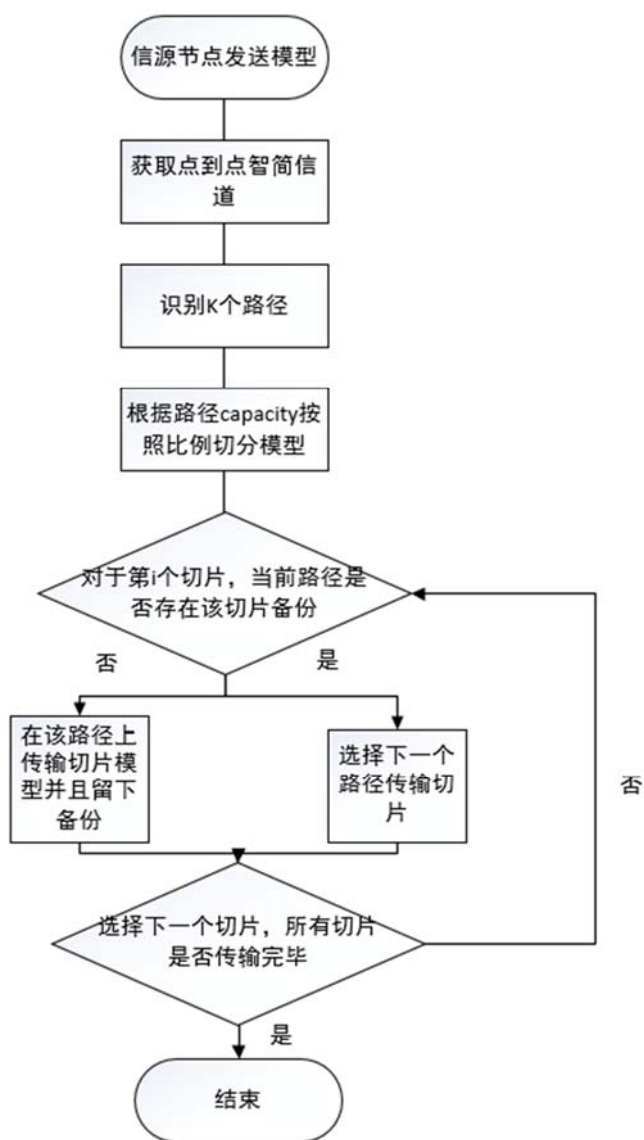


图 5 模型切分传输示意图

具体地，如图 5 所示，在智简网络中，当信源节点需要与信宿节点传输模型时，可以按照横向、纵向、或者功能切片的方式将模型切分为多份，并将切片在多条路径上传输。模型切片通过无线信道到达信宿端后，将重新组建成所需模型。此外，传播路径上每个智简节点可以根据业务需求保留切片备份，以组建、更新本地模型。智简通信系统模型切分传输机制一方面提高了模型的传输效率，一方面促进了模型共享。其具体的模型切分方式和模型组建方式将在第三章智简通信关键技术详细介绍。

3.3 智简通信路由技术

基于模型切分传输理念，智简通信将进一步开展智简通信路由技术的研究，即智简网络模型残差传播方法。残差网络最初被用于图像识别领域，通过使用跳跃连接的内部残差块，缓解了深度神经网络中增加网络深度带来的梯度消失问题。受残差网络的设计理念启发，智简通信系统基于模型切分传输理念，提出了模型差分传输技术，即对于需要传输的模型切片，智简网络中各个传输结点实际只需要传输所需模型与本地节点所持有模型的差值部分，形成模型残差传播，不仅降低了信宿结点获取所需模型切片的时延，同时也减少了端到端通信过程中的数据冗余，提高了整体网络的资源利用率。

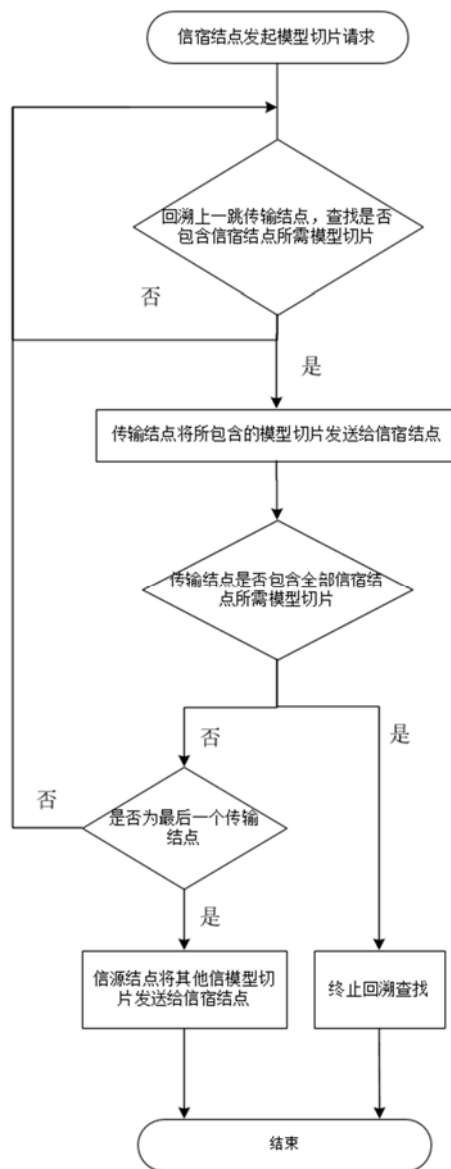


图 6 模型残差传播算法框图

如图 6 所示，在智简网络中，假设信源结点 A 与信宿节点 B 间有一条确定的路由通路，该通路经过的各个传输结点处有数量、种类不尽相同的模型切片，当信宿结点 B 想要获取自己所缺少的一些模型切片时，并非直接向拥有全部模型切片的信源结点 A 发起请求进行传输，而是根据传输路径，从信宿结点 B 的上一跳传输结点开始，逐一询问已知路由中各个传输结点。倘若上一跳节点没有所需的模型切片，则跳过该结点接着追溯到上一跳结点进行查找。倘若上一跳结点拥有 B 所需的部分你模型切片，则该传输结点向 B 发送自己所包含的 B 所需的模型切片，同时接着追溯到上一跳结点查找剩余的 B 所需的模型切片。倘若传

输结点包含全部 B 所需的模型切片，则该传输结点向 B 发送其所需的全部模型切片，同时终止回溯查找过程；倘若全部传输结点查找完毕，仍有 B 所需的模型切片未找到，则由信源结点 A 将剩余的模型切片发送至 B。

4. 智简通信系统指标体系

传统通信系统的主要任务是实现比特级或符号级的信息准确传输，因此，比特误码率（BER）或符号误码率（SER）作为其主要性能指标。而智简通信系统关注的是传输的符号如何精确地传达所需的含义，因此智简通信系统从语义的实际意义出发，提出全新的语义度量指标，用于评估语义信息恢复的准确度。此外，智简通信系统具有模型传输的全新特性，即语义信息处理的人工智能模型也通过物理信道进行传输。因此，智简通信系统的性能不仅取决于语义准确度，还在于接收方模型的性能，如所需耗费的计算、存储资源以及模型的普适性、可靠性等。在下面的内容中，我们将讨论可以智简通信系统中使用的候选指标。

首先，分别针对文本、语音、图像和视频信息讨论智简通信的评估准则：

4.1 文本对象

对于文本对象，其智简通信性能可以通过自然语言处理中最常用的文本生成度量来评估，如双语评估代理（Bilingual evaluation understudy, BLEU）等。BLEU 是一种用于评估机器翻译质量的算法，它使用一种改进的精度形式，将输出文本与语言专家翻译的多个参考句子进行比较。BLEU 的输出结果通常在 0-1 之间，BLEU 结果得分越高，说明其翻译质量越好。在该系统中，恢复的文本和原始文本可以作为引用和输出文本进行自然语言处理。BLEU 度量可以用来度量文本之间的语义错误，BLEU 精度值的计算公式如下：

$$BLEU = BP * \exp\left(\sum_{n=1}^N \omega_n p_n\right)$$

在上述公式中，通过采用最大长度的为 N 的第 n 个元素，来计算第 n 个元素的精确度，具体表示为

$$p_n = \frac{\sum_i \sum_k \min(h_k(c_i), \max_{j \in m} h_k(s_{ij}))}{\sum_i \sum_k \min(h_k(c_i))}$$

其中 c_i 表示产生的每个句子， s_{ij} 表示对于句子 c_i 的第 j 个参考句子。 $h_k(c_i)$ 和 $h_k(s_{ij})$ 分别表示在句子 c_i and s_{ij} 关于第 k 个元素的相关数目。 ω_n 是一个正的权重值，其总和为1，表示分配给每个修正精确度的权重。 BP 表示一个简单的惩罚，以避免翻译过短。当 $l_c > l_s$ 时， BP 设置为1，否则，设置为 $\exp(1-l_s/l_c)$ ，其中， l_c 和 l_s 分别表示产生的文本中和传输文本中的单词的统计量。

4.2 图像和视频对象

对于图像和视频的多媒体对象，其性能可以通过计算机视觉中用于图像生成和视频生成的最常用指标来评估，例如峰值信噪比 (**peak-signal-to-noise ratio, PSNR**)。对于给定所提出系统中的传输图像 f 和恢复图像 g ，尺寸为 $M \times N$ 的灰度图像的 PSNR 定义为

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE(f, g)} \right)$$

其中， $MSE(\cdot)$ 表示均方误差值，具体计算为

$$\sum_{i=1}^M \sum_{j=1}^N (f_{ij} - g_{ij})^2 / M * N$$

$PSNR$ 的值越小，表示传输中造成较高的语义错误。

除了像素级的相似度测量之外，通过使用感知距离来测量两幅图像的相似程度，与人类判断相一致的度量可以用于评估恢复的图像和视频的质量，例如结构相似度指数测量 (**structural similarity index measure, SSIM**)、学习感知相似度 (**learned perceptual similarity, LPIPS**) 度量和Fréchet 视频距离 (**Fréchet Video Distance, FVD**)。以SSIM 度量为例，SSIM度量通过引入人类视觉系统 (**Human Visual System, HVS**) 的质量感知概念来评估两幅图像之间的相似度。SSIM将图像失真建模为相关损失、亮度失真和对比度失真的组合：

$$SSIM(f, g) = l(f, g) * c(f, g) * s(f, g)$$

其中， $l(f, g) = (2\mu_f \mu_g + C_1) / (\mu_f^2 + \mu_g^2 + C_1)$ ，表示两个图像亮度 μ_f 和 μ_g 的接近度。

$c(f, g) = (2\sigma_f \sigma_g + C_2) / (\sigma_f^2 + \sigma_g^2 + C_2)$ 表示两个图像明暗对比度的接近度。

$s(f, g) = (\sigma_{fg} + C_3) / (\sigma_f \sigma_g + C_3)$ 通过两幅图像之间的相关系数来衡量表示两幅图像之间的结构比较， σ_{fg} 、 C_1 、 C_2 和 C_3 表示用于避免零分母的正的常数。

4.3 语音对象

对于语音多媒体信息，其性能可以通过语音质量以及生成信号与原始信号之间相似性的指标来评估，例如信号失真比（signal-to-distortion ratio, SDR）、语音失真的感知评估（perceptual evaluation of speech distortion, PESQ）。在这些指标中，SDR 通常被认为是对声音质量的总体衡量，计算为发送语音和接收语音信号之间的误差表示为：

$$m_{SDR} = 10 \log_{10} \left(\frac{\|s_{\text{target}}\|^2}{\|e_{\text{interf}} + e_{\text{noise}} + e_{\text{artif}}\|^2} \right)$$

其中， s_{target} 表示由传输者发送的信源， e_{interf} 、 e_{noise} 和 e_{artif} 分别表示干扰、噪声和人造影像的误差项。

除了SDR指标，PESQ也被视为评估各种条件下语音质量的一种有效方式，例如背景噪声、模拟滤波和可变延迟等条件下，计算公式为：

$$PESQ = 4.5 - 0.1d_{SYM} - 0.0309d_{ASYM}$$

其中， d_{SYM} 和 d_{ASYM} 分别表示对称和非对称扰动。

4.4 其他指标

此外，智简通信系统将进一步考虑智简模型与用户需求的关系，进一步讨论与模型性能、模型所占存储资源与运行模型所消耗的计算资源相关的智简通信指标。

由于模型传播的存在，所提出的通信系统可以根据服务需求和质量等不同条件接收和集成来自其他网络节点的深度学习模型。因此，在性能评估方面，语义信息的传输要注意语义恢复的信息是否能够满足预期。语义服务质量指标由下式给出：

$$SS = \frac{ST(\hat{S})}{ST(S)}$$

其中， S 表示发送端未处理的信息， \hat{S} 表示在接收端通过语义信息恢复所得到的信息， $ST(\cdot)$ 表示信息在执行后续任务中的性能。使用Sigmoid等函数将 $ST(\cdot)$ 的结果映射到 $[0,1]$ 。选择后续任务的指标应该是性能越好，指标越高。如果 SS 等于1，则语义信息完全恢复。如果 SS 等于0，则表示语义信息无法恢复。

此外，人工智能模型的时间复杂性和空间复杂性对语义通信系统的性能也至关重要。其中时间复杂度决定了模型的训练与预测时间。如果复杂度太高，模型则需要长时间的训练和预测推理过程，即其实时性无法保障。模型的时间复杂度可以用浮点计算量（FLOPs）衡量，以卷积神经网络（CNN）为例，其时间复杂性定义为

$$\text{Time} \sim O\left(\sum_{l=1}^D M_l^2 * K_l^2 * C_{l-1} * C_l\right)$$

其中， D 表示神经网络所具有的卷积层数，即网络深度， l 表示第 l 个卷积层， M_l 表示第 l 个卷积核输出特征图的边长， K_l 表示第 l 个卷积核的边长。 C_l 表示第 l 个卷积层的输出通道数。

空间复杂度由模型的参数量决定。如果模型的参数越多，则需要大量的训练数据，空间复杂度定义为

$$\text{Space} \sim O\left(\sum_{l=1}^D K_l^2 * C_{l-1} * C_l + \sum_{l=1}^D M_l^2 * C_l\right)$$

其中， $\sum_{l=1}^D K_l^2 * C_{l-1} * C_l$ 代表参数总量，其只与卷积核的尺寸 K ，通道数 C 与层数 D 有关，而与输入数据大小无关。 $\sum_{l=1}^D M_l^2 * C_l$ 表示输出特征图的空间占比，即为空间尺寸 M^2 与通道数 C 的乘积。

第四章：智简通信关键技术

1. 信源信道联合编解码技术

为真正实现“节点极智、链路极柔、网络极简”的智简通信目标，面向6G业务多模态信息的语义编解码的性能与效率至关重要。近年来，多媒体业务在图像编解码技术、视频编解码技术等方面均有了突出进展，可以实现不同信源模态业务的高效压缩与恢复。目前智简通信系统研究团队在此基础上，进一步研究针对不同服务质量需求的多媒体业务的语义信息模型和时变的信道条件，进行低算法复杂度和实现难度的信源信道联合编码技术。

此外，为了实现6G多模态信息的高效传输与语义恢复应用，保障语义信息高效准确传输，智简通信系统设计信源信道联合编码模块，利用人工智能技术对传输环境进行感知建模，以及对调制编码信道进行信道估计，基于AI模型，根据当前信道传输信息，对需要传输的数据进行信源信道编码，对抗信道状态的时变特性所产生的差错传输，以改善信源信道联合编码效率并提升语义恢复准确度。

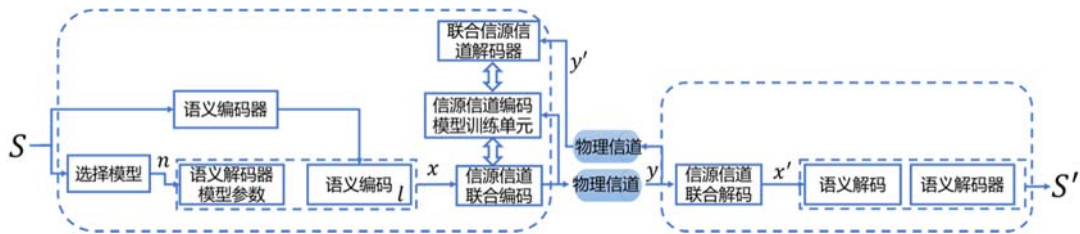


图7. 信源语义编码传输与接收复原流程

具体地，信源语义编码传输与接收复原流程如图7所示。在发送端，将信源信息输入分类器模型，检测出信源信息类别，基于信源的类别，结合多模态语义信息编码技术，对语义信息进行提取，将语义信息、传输环境状态信息输入到信源信道联合编码模块，基于训练好的深度学习模型，对上述数据信息进行信源信道联合编码，并发送至物理信道，进行比特数据传输。在接收端，构建一个相应的智简接收机模型。相应地，在接收端配置一个信源信道联合解码模块和一个语义解码模块。首先，基于训练好的深度学习模型，基于当前的信道状态信息，解码接收到的比特数据信息，得到相应的语义编码信息、关键特征提取信息及语义信息解码器模型参数，并将语义编码信息输入到语义解码器模块，基于语义解码

器模块参数，解码出相应的语义信息和关键特征提取信息，并通过语义信息与关键特征提取信息进行校验，有助于准确复原的语义信息。若校验信息通过，则表明信源信息得到正确传输，否则，利用关键特征信息进行修复，或者触发全部信源信息重新发送，或者部分信源信息重新发送。当解码出全部的分割的语义信息，再根据分割方式进行整体语义信息复原。

2. 模型切分技术

为了满足6G全新的应用场景和个性化的极致性能需求，智简通信系统需要为6G用户提供“定制化”地模型，并制定高效的模型传输机制，进一步地，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，一方面平衡无线通信系统负载，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。

在人工智能领域，常见的神经网络的模型并行方法有：横向按层划分和纵向跨层划分。其中横向按层划分指的是将神经网络每两层间的连边参数、激活函数值和误差传播值存储于一个工作结点。前馈时，每一个节点要先向前一个结点索要一层的数据，经过自身的计算后，再把数据存到自身的末层上，供下一层发出索要数据的申请。后馈时，每个结点要先向后一层发出申请，索要后一层传下来的误差，经过训练后将误差保存在自身结点的首层，供前一个结点索要误差值。相反地，纵向跨层划分把每一层的神经网络参数均等地划成若干份，每个结点存一份，相当于每个结点存了一整个神经网络。每个结点除了存自己那一部分神经元的的数据以外，还要存指向该神经元的边。当训练或推导时，若需要其他神经元的值（推导时），则需要向其他节点发出索要数据的申请；若需要将误差传给其他神经元（训练时），则需要向其他节点发出写入误差的申请。在基于分割网络与联邦学习的模型联合训练过程中，横向切分和纵向切分的模型并行方法，使得联邦学习过程可以在多个节点并行完成，使得训练过程可以在自身计算、存储等资源的限制条件下，完成高效的模型传输与训练。

考虑到在智简网络中传播的模型功能间的相似性，智简通信基于传统模型并行方法，进一步创造性地提出将AI模型在功能和性能上进行颗粒化处理，形成智简模型的精度切片和功能切片，并在网络中传输，一方面用于补齐已有模型的功

能或性能，另一方面可以在满足网络中AI任务的功能和性能需求的同时，节省模型传输开销，更好地发挥智简网络的优势[9]。

当智简用户为完成某个人工智能任务，需要利用其它模型改进其自身积累的模型时，可向中央模型管理器发起请求，其中包括用户当前原始模型的功能、性能和版本编号以及用户对新模型功能、性能和版本（可以用于区分模型体积等）的需求信息。用户可能需要更高性能，如精度更高、速度更快的模型，也可能需要具有更多功能的模型，如可识别的种类更多的分类模型等。

如图8所示，对MNIST手写数字识别模型增加功能切片，使其除了数字0-9之外能够识别字母，也可能对模型的版本、体积等属性有特定要求，如用户需要体积更小或版本更高的模型。这些模型增量都可以在模型切片中体现。

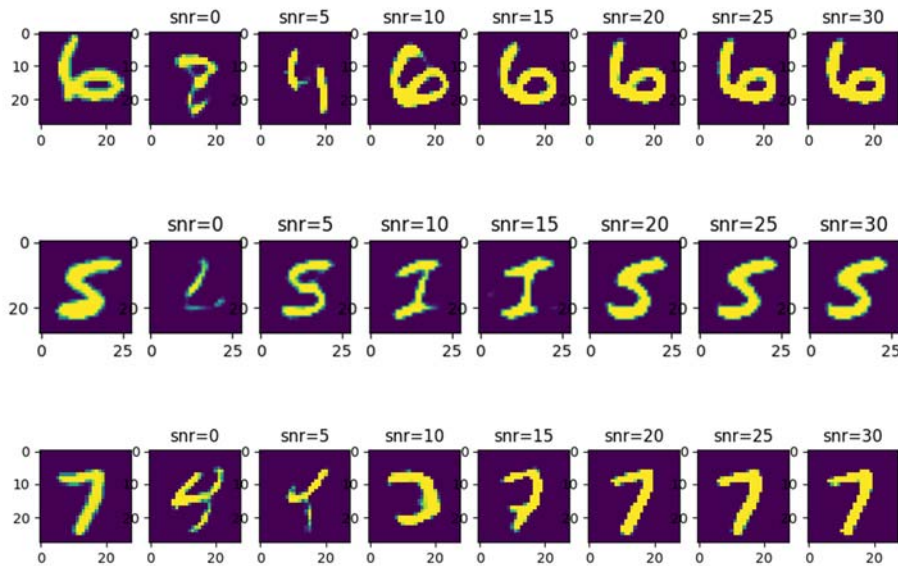


图8 MNIST手写数字数据集信源智简传输效果

如图9所示，使用 Open Image 数据集在智简通信系统中传输后再进行图像重建，在信噪比更高的情况下具有更好的图像重建效果。也可以根据模型、版本、属性有特定要求，如用户需要体积更小版本的模型等，进行相应的智简模型增量的部署。



图9 Open Image图片数据集信源智简传输效果

智简通信系统具备的中央模型管理器记录有网络中每个模型的功能、性能、版本编号，以及存储有该模型的节点编号。中央模型管理器根据请求查询到满足用户要求的模型及其存储者，并将用户的模型请求转发给目标模型存储者，当存储者同意协助该用户进行模型改进后，根据原始模型的功能、性能和版本编号以及用户对新模型功能、性能和版本的需求信息对存储者本地存储的满足用户需求的模型进行颗粒化处理，生成刚好可以补齐该用户功能、性能及版本需求的个性化模型增量切片，最后将该模型切片发送给用户。这里的模型增量切片体现的可以是模型的性能增量、功能增量或者版本差异，也可以是三者的组合。

3. 模型组建与更新机制

为解决通信系统日益复杂化带来的诸多问题和满足用户极致个性化的应用需求，智简通信以模型为核心，实现节点极智、网络极简。因此，如何组建、训练出满足未来用户需求和网络资源限制的模型是智简通信需要解决的关键问题。

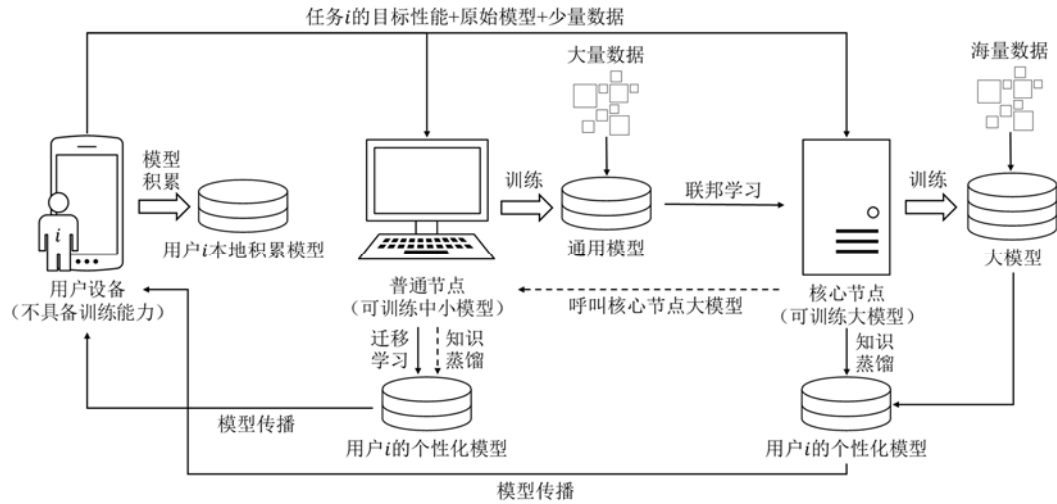


图 10 智简模型组建与更新机制

如图10所示，基于第三章提出的智简通信网络架构，即部署节点层/用户设备层、普通节点层和大型服务器/核心节点层，根据用户目标性能需求和资源限制，本地节点可以通过自身模型积累、呼叫周边节点模型、呼叫远距离大模型三个渠道组建自身模型：

1) 不同的智简用户设备/部署节点具有不同的目标性能，并且可以通过评价指标识别自身积累的已有模型和目标性能的差距，通过和周边节点（包括普通节点和核心节点）的交互，将本地模型和少量数据传输至周边普通节点或核心节点进行迁移学习或知识蒸馏，组建成满足目标性能要求的个性化模型。

2) 普通节点存储有大量数据和通用模型，且具备训练中小模型的能力，可以进行迁移学习或知识蒸馏来完成模型的更新，或作为客户端参与核心节点的联邦学习过程。普通节点可以和用户设备及核心节点进行交互，当用户在普通节点处完成迁移学习后尚未满足目标性能，可以呼叫远距离核心节点发送大模型进行知识蒸馏，将大模型的能力迁移到中小模型上，从而达到用户的目标性能要求。

3) 核心节点存储有海量数据和参数量繁多的大模型，可以进行迁移学习或知识蒸馏来完成用户模型的进化，或作为服务端进行联邦学习，完成对本地大模型及周边节点模型的更新。核心节点可与普通节点及周边用户设备进行交互，将本地大模型传输至周边节点，或接收目标性能要求较高的周边用户的数据和原始模型，进行迁移学习或知识蒸馏，进而大幅提升周边用户的模型性能。

三层网络结构之间的具体交互过程如下所示：

1) 用户设备*i*向周边普通节点或核心节点发送任务*i*的目标性能，根据本地模型与目标性能的差距以及传输距离选定与之进行交互的有效节点，当用户*i*周边存在核心节点且任务*i*的目标性能较高时，优先与核心节点进行交互；当用户*i*周边只有普通节点时，选择距离最近的普通节点进行交互。确定目标节点后，用户*i*向其发送任务*i*的目标性能、原始模型和少量数据，节点可对原始模型进行迁移学习或知识蒸馏，生成用户*i*的个性化模型。用户*i*得到用于完成任务*i*的个性化模型后，该模型可积累在用户设备本地，作为后续任务的原始模型。

2) 普通节点存储有大量数据，可用于训练得到中小规模的通用模型。普通节点接收到来自用户*i*的模型组建请求后，利用训练好的本地通用模型和用户*i*发来的少量数据对任务*i*的原始模型进行迁移学习。此时若满足用户*i*的目标性能要求，则将迁移学习得到的模型传输给用户*i*；若尚未满足目标性能要求，普通节点则呼叫核心节点传输大模型，进行知识蒸馏，再将得到的模型传输给用户*i*。

3) 核心节点存储有海量数据，可用于训练得到大规模、高性能的模型。核心节点接收到用户*i*的模型组建请求后，利用训练好的本地大模型和用户*i*发来的少量数据对任务*i*的原始模型进行知识蒸馏。这是因为核心节点处模型规模和原始模型规模差距较大，需要利用知识蒸馏将大模型优良的性能集成在轻量化的小模型上。达到任务*i*的目标性能后，核心节点将知识蒸馏得到的模型传输给用户*i*。

4. 智简内生安全技术

针对未来移动通信网络面向连接、异构共存、智能互联的特点，在支持无处不在的通信的同时，多种类型的设备与多种形态的网络相互连接，任何节点和网络都有可能成为攻击的突破口对6G网络进行渗透，安全的短板效应更加突出，6G网络的新架构、新应用、新技术的出现势必会引入新的安全威胁。然而，传统移动通信系统在设计之初把安全作为一种独立的技术，依靠“补丁式”“外挂式”，实现无线通信的安全防护，难以确保6G自身对安全内嵌的需求。因此，智简通信需要从无线网络的共性和语义通信本质出发，从通信架构设计之初就考虑语义通信内嵌安全缺陷，探讨智简网络中可能存在的新的攻击方式与攻击路径，具体

包括：

1) 模型联合训练阶段的隐私泄露风险：尽管智简通信系统可以通过结合联邦学习、知识蒸馏等技术，在充分利用智简通信系统分布式数据、提高语义通信模型性能的同时，保障各节点数据隐私和模型联合训练过程中的数据安全。然而，已有研究表明攻击者仍可以从模型参数中推断出原始数据的某些特性，如推断某数据是否被用于训练，甚至直接获取一部分或全部的训练数据。在智简通信系统背景下，攻击者可以进一步从“语义级”发动更深层次攻击。例如，在人工智能领域，基于梯度信息的隐私泄露攻击的成功率常常受到数据复杂度、联合训练数据批量大小、联合训练用户参与数等的限制。然而，在智简通信系统中，攻击者可以通过对数据的语义先验信息进行建模，从而大幅度地提高攻击成功率和攻击的时效性。

2) 模型联合训练阶段的安全风险：攻击者可以通过参与模型联合训练过程毒化语义通信模型，或通过训练过程中植入特定的后门触发器，使模型在特定的触发条件下产生误导性的输出结果，从而危害智简通信系统的安全。值得注意的是，在人工智能领域，攻击者通常需要对数据进行手工筛选或通过添加触发器的方式，来组建毒化数据集以毒化语义通信模型。而在智简通信系统中，攻击者可以通过对数据（甚至模型）在“语义级”进行直接处理，从而大幅度地提升攻击效率。

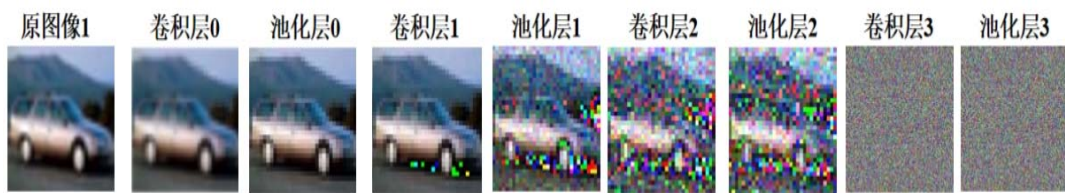


图11：基于不同层提取出的语义信息重构出的原始信息

3) 语义信息传输阶段的隐私泄露风险：如图11所示，在智简通信系统中，关键语义信息的提取、传输与恢复需要综合考虑用户需求、节点计算能力和网络资源等因素，同时也需要考虑语义信息传输过程中的隐私泄露风险。尽管，智简通信系统通过传输关键语义信息，而非原始信息，可以在一定程度上提高智简通信过程的安全性。然而，攻击者仍可以基于模型和关键语义信息推测出原始信息，从而威胁智简通信系统的保密性。因此，在模型切割和传输过程中，需要对模型切割层数进行合理地划分，并采取适当的加密和隐私保护措施，例如使用差分隐

私技术、同态加密技术等，以保护数据隐私和安全。

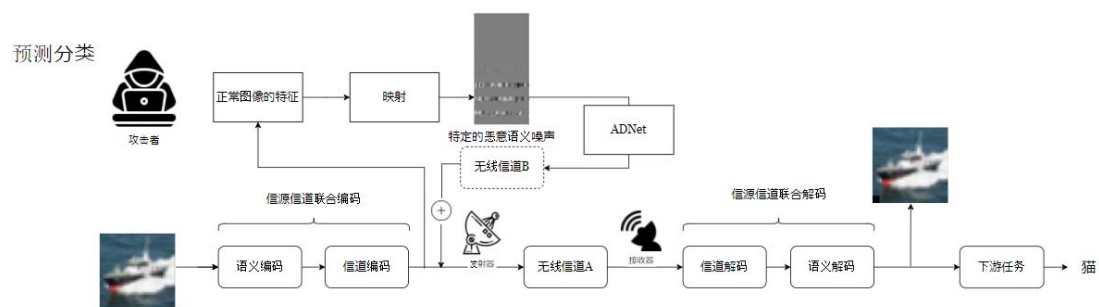


图12 在无线信道中对语义信息添加“恶意语义噪声”

4) 语义信息传输阶段的安全风险：如图12所示，在智简通信系统中，攻击者可以通过在无线信道中添加“恶意语义噪声”，从而影响正常信息在智简通信系统接收端的恢复和下游智能任务的执行。与模型联合训练过程的模型毒化攻击相比，在无线信道中添加恶意语义噪声将以一种更加“隐形”的方式，危害智简通信系统安全，需要结合智简通信系统特性设计更加可靠和安全的恶意噪声检测机制和信息完整性校验方法，保障语义信息的安全传输。

第五章 未来发展展望

在国际形势变化及双循环战略大背景下，全球6G竞赛已经全面拉开帷幕，中关村泛联移动通信技术创新应用研究院与北京邮电大学提出了智简通信系统，从系统架构、协议设计、指标体系以及关键技术等多方面变革了传统移动通信系统技术体系，也将从语义通信和模型传递等方面影响未来的产业生态，以加快全社会数字转型和数字经济的高质量发展。

将在智简通信系统现有的网络架构与关键使能技术的基础上，进一步开展智简通信系统一体化跨层设计研究，为包含多模态信源数据的多媒体业务提供更好的通信支撑，完善智简通信系统协议，支持模型按需、动态、智能的切分、组建与传输，并形成一整套模型追踪、溯源与审计机制，满足6G网络智慧内生、多维感知、数字孪生、安全内生等需求。

进一步，为厘清我国移动通信发展堵点，形成新形势下移动通信“双循环”发展格局，智简通信系统将进一步探索扩展基于经典信息论所建立的传统通信产业生态，推动引领性基础理论的进一步发展，紧抓万物智联的新机遇，拓宽通信产业新赛道，突破与垂直行业融合的技术、模式等多重梗阻，推进“智简”6G标准化布局，构建有中国特色和核心竞争力的“智简”技术体系和创新生态，助力我国在移动通信领域持续保持全球领先地位。

参考文献:

- [1] Zhang P, Zhang J, Qi Q, Hu Z, Nie G, Niu K, et al. Ubiquitous-X: constructing the future 6G networks[J]. *Sci Sin Inform* 2020; 50(6): 913-30.
- [2] Ping Zhang, Wenjun Xu, Hui Gao, Kai Niu, Xiaodong Xu, Xiaoqi Qin, Caixia Yuan, Zhijin Qin, Haitao Zhao, Jibo Wei, Fangwei Zhang. Toward Wisdom-Evolutionary and Primitive-Concise 6G: A New Paradigm of Semantic Communication Networks[J]. *Engineering*, 2022, 8(1): 60-73.
- [3] IMT-2030(6G)推进组, 《6G 总体愿景与潜在关键技术》白皮书[M], 2021.6.
- [4] Liu S. *Advanced hybrid information processing*[M]. Springer International Publishing, 2021.
- [5] 张平, 许晓东, 韩书君, 等. 智简无线网络赋能行业应用[J]. *北京邮电大学学报*, 2020(6): 9.
- [6] Ping Z, Xiaodong X, Chen D, et al. Intellicise communication system: model-driven semantic communications[J]. *The Journal of China Universities of Posts and Telecommunications*, 2022, 29(1): 2.
- [7] Weaver W. Recent contributions to the mathematical theory of communication. *ETC: A Review of General Semantics*[J], 1953, 10: 261 - 281
- [8] Xu Xiaodong, Xiong Huachao, Wang Yining, Han Shujun, Wang Bizhu, Zhang Ping. Knowledge-Enhanced Semantic Communication System with OFDM Transmissions[J]. *Sci China Inf Sci*, Nov. 18, 2022.
- [9] Chen Dong, Haotai Liang, Xiaodong Xu, Shujun Han, Bizhu Wang, Ping Zhang, Semantic Communication System Based on Semantic Slice Models Propagation[J], *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 202-213, Jan. 2023.
- [10] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu and G. Y. Li, Robust Semantic Communications with Masked VQ-VAE Enabled Codebook, *IEEE Transactions on Wireless Communications*, doi: 10.1109/TWC.2023.3265201.