# 面向资源受限电力物联网终端的语义安全通信方法

杨会峰[1]，尚立[1]，崔俊彬[1]，刘红艳[1]，王九成[1]，蔺鹏[2]
（1. 国网河北省电力有限公司信息通信分公司，河北 石家庄 050051；2. 北京万可信息技术有限公司，北京 100085）

**摘　要：**语义模型训练通常需要耗费大量能量和时间，阻碍了在资源受限的电力物联网终端上实施语义传输。为减轻终端的能耗和时耗，建立了一种新的语义通信架构。首先，将待传数据上传至机器学习即服务（MLaaS, machine learning as a service）平台；然后，在 MLaaS 平台完成语义模型训练并将模型参数回传给终端；最后，通过终端进行语义推理。然而，该架构存在 MLaaS 平台泄露语义模型参数导致语义信息被窃听的问题。因此，进一步设计了基于特征混淆的抗窃听方法以解决语义推理阶段的 MLaaS 平台安全通信问题。实验结果表明，所提方法在面对被动窃听者时是有效的，能够在保证合法终端图像恢复质量的同时，显著降低窃听者恢复图像的成功率。此外，还初步验证了特征混淆模块在终端上的计算开销和时延，结果显示该方法在资源受限的电力物联网终端上具有实际应用可行性。

**关键词：**电力物联网终端；语义通信；机器学习即服务；抗窃听
**中图分类号：**TN919.8
**文献标志码：**A
**doi：**10.11959/j.issn.2096−3750.2025.00455

# A semantic secure communication approach for resource-constrained power IoT terminals

YANG Huifeng[1], SHANG Li[1], CUI Junbin[1], LIU Hongyan[1], WANG Jiucheng[1], LIN Peng[2]
1. Information and Communication Branch, State Grid Hebei Elecrtic Power Co., Ltd., Shijiazhuang 050051, China
2. Beijing Waco Information Technology Co., Ltd., Beijing 100085, China

**Abstract:** The training of semantic models typically consumes a large amount of energy and time, which hinders the implementation of semantic transmission on power Internet of things (IoT) terminals with limited resources. To reduce the energy and time consumption at the terminal, a novel semantic communication architecture was proposed. Firstly, the data to be transmitted was uploaded to a machine learning as a service (MLaaS) platform. Then, the semantic model training was performed on the MLaaS platform, and the model parameters were sent back to the terminal. Finally, semantic reasoning was carried out through the terminal. However, this architecture faces the risk of the MLaaS platform leaking semantic model parameters, leading to the potential eavesdropping of semantic information. Therefore, an anti-eavesdropping method based on feature obfuscation was designed to address the security communication issue of the MLaaS platform during the semantic inference phase. Experimental results show that the proposed method is effective against passive eavesdroppers. The effectiveness of the proposed anti-eavesdropping method was demonstrated. Additionally, the computational overhead and latency of the data obfuscation module on terminals have been preliminarily verified, showing that the method is feasible for practical application on resource-constrained power IoT terminals.

**Key words:** power IoT terminal, semantic communication, machine learning as a service, anti-eavesdropping

**收稿日期：**2024−10−08；**修回日期：**2024−12−09
**通信作者：**蔺鹏，linpeng@vectinfo.com
**基金项目：**国网河北省电力有限公司科技项目（No. SGHEXT00GCJS2310140）
**Foundation Item:** The Science and Technology Project of State Grid Hebei Electric Power Co., Ltd. (No. SGHEXT00GCJS2310140)

## 0 引言

随着5G技术的广泛应用，研究人员开始探索下一代移动通信技术。未来6G时代将继续沿用5G技术并扩展至更多领域，旨在构建一个数字与现实世界无缝整合的未来，为广大用户提供更多样化和定制化的服务体验[1-4]。在这样的愿景指引下，电力物联网不仅追求更快的数据传输速率，而且朝着能够连接众多设备、支持广泛智能化应用场景的全新网络架构方向发展。这种创新架构将具备感知、学习、决策、自适应演进的智能能力，能够在自动电力巡检、配电自动化、虚拟现实等多样化电力场景中，有效支持包括机器之间、人与机器、人与人之间的多种先进通信模式。

通信系统虽然在比特级别的数据传输方面取得了巨大成功，但在未来高度互联和智能化的电力物联网中，简单地传输比特流已经无法满足日益增长的需求。电力通信网需要更深入地理解数据背后的意义和意图，以便更高效、更智能地传输信息。因此，语义通信[5-8]作为一种新兴的通信范式应运而生，它不仅关注比特的准确传输，更关注信息的真正含义和意图。语义通信能够显著提高频谱效率，并在低信噪比（SNR, signal-to-noise ratio）条件下表现出色，为电力物联网提供强大的技术支持。

近年来，人工智能和计算技术的发展为电力物联网在理解和处理语义信息上提供了可能。应用人工智能从数据中提炼对接收端任务有意义的信息，可极大减少不必要的数据传输，从而提升信息传递的效率和有效性[9-11]。这一进步为语义通信的实际应用铺平了道路，使得电力物联网不仅能传输数据，还能理解数据背后的含义，进而实现更加智能化和个性化的服务。在电力物联网环境中，语义通信的价值尤为突出。它可以通过减少冗余数据传输来降低能耗，这对于资源受限的电力物联网终端至关重要。此外，语义通信还可以增强系统的安全性和隐私保护，确保只有授权用户能够解析出传输内容的真实意义，防止敏感信息泄露。这不仅提高了系统的可靠性，也增强了用户的信任感，有助于推动电力物联网技术的普及和发展。

基于深度学习的语义通信，利用深度学习技术从原始数据中提取关键的语义信息进行传输，通常包括语义模型训练和语义推理两个阶段。尽管深度学习框架提供了这一能力，但语义模型训练过程需要较高的能耗和较长的时间，这阻碍了在资源受限的终端上实施语义通信[12]。一个有效的解决办法是将模型训练的计算任务卸载到机器学习即服务（MLaaS, machine learning as a service）平台[13]，借助MLaaS平台的强大算力执行复杂的计算，再将训练结果返回给终端。在语义通信的推理阶段，终端使用已完成训练的语义模型进行语义传输。

虽然使用MLaaS平台能够解决本地算力不足的问题，然而这种将数据传输给云服务提供商的行为也带来了安全性方面的顾虑。MLaaS平台可能存在内部人员泄露数据的风险，即MLaaS平台的内部员工可能出于各种原因（如经济利益、个人兴趣等）故意泄露训练数据或模型参数。一旦数据发生泄露，将会对语义通信产生严重的威胁：窃听者可以通过获取语义模型参数，解析出传输数据的真实含义，从而破坏通信的保密性；窃听者可以利用获取到的模型参数进行模型反演攻击，恢复出训练数据中的敏感信息，进一步扩大数据泄露的影响；数据泄露事件可能导致用户对电力物联网系统的信任度下降，影响系统的推广和应用。

考虑如图1所示的基于MLaaS平台的语义通信场景：在阶段1，终端A上传训练数据到MLaaS平台，然后MLaaS平台基于深度学习设计语义编解码器，模拟数据的语义无线传输过程完成语义模型训练；在阶段2，MLaaS平台将语义模型参数回传给终端A及其语义数据传输对象，如终端B；在阶段3，终端A在本地使用完成训练的语义编码器将数据编码为适合无线传输的形式，然后通过无线信道传输到终端B。终端B接收到数据后，使用完成训练的语义解码器恢复原始数据，并进行语义推理，实现A和B之间的语义无线传输。然而，由于无线信道的广播特性，窃听者可收到A发送的语义信息，如果其还能通过MLaaS平台获取语义编解码器参数，则将能窃听到A的数据。值得注意的是，这里的窃听者属于被动窃听者，仅限于接收信息，无法注入或修改信息。

本文的主要贡献如下。

1) 本文建立了一种新的语义通信架构。资源受限的电力物联网终端仅在本地执行语义推理，而将需要高能耗和计算量的语义模型训练任务卸载到MLaaS平台上完成。这种架构有效减轻了终端的计

算负担和能耗，提高了系统的整体效率。

2) 针对MLaaS平台存在的模型参数泄露风险，本文设计了一种基于随机排列的特征混淆技术，用于保护语义推理阶段的数据安全。具体而言，该技术通过对图像的行、列、通道进行随机排列，使得窃听者即使获取到数据也无法正确恢复原始图像。此外，合法终端在接收到混淆数据后，可以通过预先约定的逆变换方法恢复出原始图像，从而实现安全的语义传输。

3) 本文面向图像传输任务展开实验验证，结果表明合法终端在不同SNR条件下均能保持较高的峰值信噪比（PSNR, peak signal-to-noise ratio）、多尺度结构相似性指数（MS-SSIM, multi-scale structural similarity）和可学习感知图像块相似度（LPIPS, learned perceptual image patch similarity），而窃听者的PSNR、MS-SSIM和LPIPS显著低于合法终端。此外，本文还初步验证了特征混淆模块在终端上的计算开销和时延，结果显示该方法在资源受限的电力物联网终端上具有实际应用可行性。这些成果为电力物联网中的安全语义通信提供了新的解决方案，具有重要的理论和应用价值。

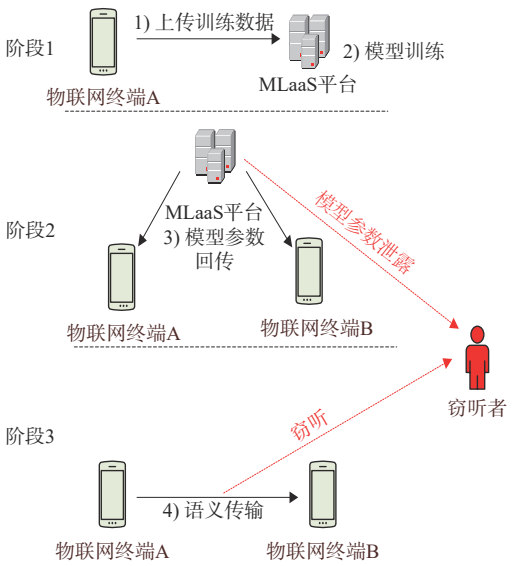



图1　基于MLaaS平台的语义通信场景

## 1　相关工作

### 1.1　基于深度学习的语义通信

2018年，Nariman Farsad首次提出联合信源信道编码（JSCC, joint source-channel coding），这是一项具有开创性的基于深度学习展开的语义通信研究，其利用了一个循环神经网络模型，在二进制擦除信道上进行文本的语义传递[14]。随后，JSCC方法在图像压缩和传输领域逐渐受到关注。与传统的分离型信源和信道编码方案相比，在低SNR环境下，JSCC展现出更优秀的图像恢复效果[15]。针对分布式信源编码的需求，文献[16]提出了基于图像的分布式JSCC方案。然而，上述方案均在一致的信道条件下进行训练和测试，若要使语义通信网络适应不同的SNR水平，则需要训练多个神经网络模型，这对于收发装置的存储容量提出了较高要求。为解决这一挑战，研究人员开发了一种可适应不同SNR的JSCC神经网络[17]。文献[18]采用JSCC架构在多径衰落信道上的无线图像传输，结合了双重注意力机制与正交频分复用技术，并通过利用估计的信道状态信息适应信道增益和噪声功率的变化。文献[19]首次提出了端到端的JSCC视频传输方案。文献[20]通过训练具备JSCC和正交信号调制能力的神经网络模型，从而在语义通信中实现多址接入。

### 1.2　深度学习中的安全问题及对抗策略

深度学习的各个阶段都面临大量安全性方面的问题[21]，在数据收集阶段，可能会存在数据污染、假数据等问题；在训练阶段，可能会存在数据投毒、语义后门等问题；在推理阶段，可能会存在被动窃听、对抗样本、成员推断攻击和模型反演攻击等问题。

1) 对于数据收集阶段的安全问题，研究人员常用匿名化、认证等方法进行防御[22]。

2) 对于训练阶段的安全问题，由于窃听对手可能在训练过程中提取部分训练数据，并通过模型反转攻击从面部识别系统中恢复的图像，研究人员提出了使用差分隐私作为隐私保护的方法[23]；还有通过向训练数据集添加随机噪声来隐藏敏感信息，以防止将训练数据泄露给不受信任的第三方。

3) 对于推理阶段的安全问题，研究人员使用同态加密的方式来保护用户数据。云服务器通过这种加密手段，只对用户数据的密文起作用，能够有效保护推理阶段的数据安全，但该方法有极高的通信开销，在资源受限的电力物联网终端上是不可行的。对于神经网络的中间特征存在隐私泄露的问题，文献[24]提出了一种基于对抗训练的隐私保护特征提取方法，通过在特征提取过程中施加隐私约

束，有效减少所提取的特征对隐私的泄露。然而，对抗训练需要重新设计或调整模型结构，增加了实现难度，并且对于已知模型参数的窃听者来说，仍然存在被绕过的风险。

尽管针对深度学习各个阶段的安全问题已经有许多有效策略研究，然而，窃听者如果能够完整地获取神经网络模型及其参数，就能绕过采用了对抗训练等抗窃听技术的网络保护措施，对推理阶段进行窃听。因此，如何在资源受限的环境下提供有效的抗窃听机制仍然是一个挑战。

## 2　基于MLaaS平台的语义通信

本文构建的语义通信模型，让基于JSCC的语义训练阶段的计算任务与语义知识库的存储在MLaaS平台执行，然后在语义通信的推理阶段，终端使用MLaaS平台获得的JSCC参数进行语义传输。

基于MLaaS平台的语义通信架构主要涉及数据上传、模型训练和语义推理3个阶段，下面对每个阶段的算法复杂度进行详细分析。

1) 数据上传阶段的时间复杂度为$O(D/R)$，其中$D$为数据大小，$R$为网络传输速率；存储复杂度为$O(D)$。

2) 模型训练阶段的时间复杂度为$O(NdPT)$，其中$N$为训练数据集大小，$d$为每个样本的特征维度，$P$为模型参数数量，$T$为训练迭代次数；内存复杂度为$O(Nd+P)$。

3) 语义推理阶段的时间复杂度为$O(Pd+M^2)$，其中$d$为输入数据维度，$M$为图像大小；内存复杂度为$O(P+M^2)$。

与已有的算法相比，本文提出的基于MLaaS平台的语义通信方法在复杂度方面具有明显优势，将模型训练任务卸载到MLaaS平台，大大减轻了终端设备的计算负担和能耗，使得资源受限的电力物联网终端能够高效地进行语义通信。然而，该方法也存在局限性，如模型训练阶段在数据集较大或模型复杂度较高时的训练时间可能会较长，存储开销也会增加。

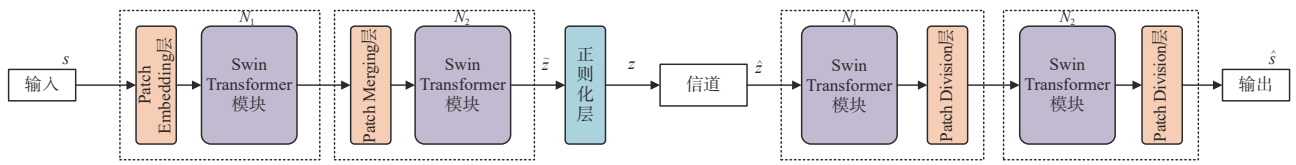
### 2.1　语义模型训练

图2展示了基于MLaaS平台的语义训练。MLaaS平台采用基于Swin Transformer[25-28]的JSCC语义训练模型，如图2(a)所示。MLaaS平台接收到图像数据$\boldsymbol{s}\in\mathbf{R}^n$，通过语义通信编码器模块将其映射为复值信道输入符号$\boldsymbol{z}\in\mathbf{C}^k$。如果将图像维度$n$记作源带宽，信道维度$k$记作信道带宽，则$k/n$被称为带宽压缩比。$N_1$、$N_2$代表不同的Swin Transformer模块个数。初始时，图像样本$\boldsymbol{s}\in\mathbf{R}^{H\times W\times 3}$（$H$和$W$分别表示图像的高和宽，3表示通道数）经Patch Embedding获得图像特征$\boldsymbol{z}^0\in\mathbf{R}^{H/2\times W/2\times C_1}$（$C_1$表示图像经过压缩后的通道数），之后图像特征进入Swin Transformer模块，Swin Transformer模块如图2(b)所示。

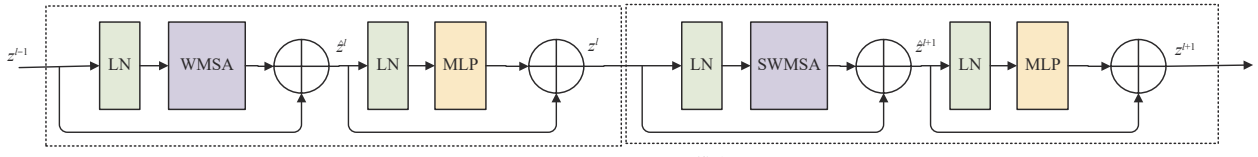
图2(b)中，图像初始进入Swin Transformer模块时，有$l-1=0$，其中$l$表示图像第$l$次进入Swin Transformer模块。Swin Transformer模块首先利用层归一化（LN, layer normalization）将$\boldsymbol{z}^{l-1}$归一化获得$\mathrm{LN}(\boldsymbol{z}^{l-1})$，接着通过窗口多头自注意力（WMSA, windows multihead self attention）模块，仅针对窗口内的元素计算自注意力，提高计算效率并同时增强模型捕捉局部特征和全局上下文的能力，之后再与$\boldsymbol{z}^{l-1}$相加，获得特征$\hat{\boldsymbol{z}}^l$

$$\hat{\boldsymbol{z}}^l=\mathrm{WMSA}\left(\mathrm{LN}\left(\boldsymbol{z}^{l-1}\right)\right)+\boldsymbol{z}^{l-1} \tag{1}$$

其中，WMSA( )为对括号内函数进行窗口多头自注




(a) 基于Swin Transformer的JSCC语义训练模型




(b) Swin Transformer 模块

图2　基于MLaaS平台的语义训练

意力处理，LN( )为对括号内函数进行归一化处理，$\boldsymbol{z}^{l-1}$为$\boldsymbol{z}$的$l-1$次方运算。

$\hat{\boldsymbol{z}}^{l}$通过LN和多层感知（MLP, multi-layer perceptron）处理，获得$\boldsymbol{z}^{l}$

$$\boldsymbol{z}^{l}=\mathrm{MLP}\left(\mathrm{LN}\left(\hat{\boldsymbol{z}}^{l}\right)\right) \tag{2}$$

其中，MLP( )为对括号内函数进行多层感知处理，$\boldsymbol{z}^{l}$为$\boldsymbol{z}$的$l$次方运算。

$\boldsymbol{z}^{l}$通过LN和偏移窗口多头自注意力（SWMSA, shifted window multihead self attention）处理，并与特征$\boldsymbol{z}^{l}$相加，获得$\hat{\boldsymbol{z}}^{l+1}$

$$\hat{\boldsymbol{z}}^{l+1}=\mathrm{SWMSA}\left(\mathrm{LN}\left(\boldsymbol{z}^{l}\right)\right)+\boldsymbol{z}^{l} \tag{3}$$

其中，SWMSA( )为对括号内函数进行偏移窗口多头自注意力处理，$\hat{\boldsymbol{z}}^{l+1}$为$\hat{\boldsymbol{z}}$的$l+1$次方运算。从而通过窗口移位，高效整合不同范围的图像信息，增强模型的理解能力。

最后通过LN和MLP模块处理，并与特征$\hat{\boldsymbol{z}}^{l+1}$相加，最终获得特征$\boldsymbol{z}^{l+1}$，完成了一个完整的Swin Transformer模块运算

$$\boldsymbol{z}^{l+1}=\mathrm{MLP}\left(\mathrm{LN}\left(\hat{\boldsymbol{z}}^{l+1}\right)\right)+\hat{\boldsymbol{z}}^{l+1} \tag{4}$$

模型在经过$N_1$个Swin Transformer模块处理后，进行补丁合并（Patch Merging）从而将图像压缩到$\mathbf{R}^{H/4\times W/4\times C_2}$（$C_2$表示图像再次经过压缩后的通道数）；然后，再通过$N_2$个Swin Transformer模块处理后得到$\tilde{\boldsymbol{z}}$。接着，让$\tilde{\boldsymbol{z}}$通过归一化层以约束发射功率，获得信道输入$\boldsymbol{z}$

$$\boldsymbol{z}=\sqrt{kP}\,\frac{\tilde{\boldsymbol{z}}}{\sqrt{\tilde{\boldsymbol{z}}^{*}\tilde{\boldsymbol{z}}}} \tag{5}$$

其中，$\tilde{\boldsymbol{z}}^{*}$为$\tilde{\boldsymbol{z}}$的共轭转置，$P$为平均发射功率。

基于Swin Transformer的语义JSCC模型训练，假设无线信道为加性高斯白噪声（AWGN, additive white Gaussian noise）信道，并建模为不可训练层，则信道输出为$\hat{\boldsymbol{z}}=\boldsymbol{z}+\boldsymbol{o}$，其中$\boldsymbol{o}\in\mathbf{R}^{H/4\times W/4\times C_2}$为AWGN向量。

语义解码器是语义编码器的对称架构，其接收到$\hat{\boldsymbol{z}}$后通过Swin Transformer模块和补丁切分（Patch Division）来翻转编码器执行的操作，从而将图像特征映射为原始图像的估计$\hat{\boldsymbol{s}}$。令$d\left(\boldsymbol{s},\hat{\boldsymbol{s}}\right)$作为损失函数来评价重建图像$\hat{\boldsymbol{s}}$与原始图像$\boldsymbol{s}$之间的损失

$$d\left(\boldsymbol{s},\hat{\boldsymbol{s}}\right)=\frac{1}{N}\sum_{i=1}^{N}d\left(\boldsymbol{s}_i,\hat{\boldsymbol{s}}_i\right) \tag{6}$$

其中，$\hat{\boldsymbol{s}}_i$为第$i$个重建图像，$\boldsymbol{s}_i$为第$i$个原始图像，$d\left(\boldsymbol{s}_i,\hat{\boldsymbol{s}}_i\right)$为$\hat{\boldsymbol{s}}_i$和$\boldsymbol{s}_i$的均方误差，$i\in\{1,2,\cdots,N\}$，$N$为图像样本总数。

使用Adam优化器更新模型参数，让原始图像$\boldsymbol{s}$与重建图像$\hat{\boldsymbol{s}}$之间的平均失真最小，从而得到语义编解码器模型的训练参数

$$\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right)=\arg\min E_{p(\boldsymbol{s},\hat{\boldsymbol{s}})}\left(d\left(\boldsymbol{s},\hat{\boldsymbol{s}}\right)\right) \tag{7}$$

其中，$p\left(\boldsymbol{s},\hat{\boldsymbol{s}}\right)$为原始图像$\boldsymbol{s}$与重建图像$\hat{\boldsymbol{s}}$的联合概率分布，$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$分别为构成语义通信的神经网络的权重与偏置参数，$E(\ )$表示期望运算。

### 2.2　语义推理阶段的安全问题

MLaaS平台完成语义模型训练后，将模型参数回传给电力物联网终端A及B；A在本地使用该语义模型完成A与B之间的语义无线传输，即进行语义推理。语义通信的特点在于传输经过解析的信息含义，而非原始数据本身，相对于传统的数据传输方式，这种方法在一定程度上增加了信息的不明确性，让未经授权的第三方更加难以解读信息，这自然形成了一种隐式的安全防护。因此，对于想要窃听的个体来说，除非他们确切知晓目标正在采用语义通信，并成功获得相关的网络模型和参数，否则很难实现有效窃听。

本文考虑一种极端情况，即MLaaS平台是一个恶意云提供商，窃听者能从中获取完整的语义JSCC模型及参数。在合法设备进行本地语义传输过程中，窃听者利用获得的解码器及参数，从接收到的语义信号$\hat{\boldsymbol{y}}_{\mathrm{e}}$中重建图像$\hat{\boldsymbol{s}}_{\mathrm{e}}$，即

$$\hat{\boldsymbol{s}}_{\mathrm{e}}=D_{\theta_{\mathrm{e}}}\left(\boldsymbol{\theta}_{\mathrm{e}},\hat{\boldsymbol{y}}_{\mathrm{e}}\right) \tag{8}$$

其中，$D_{\theta_{\mathrm{e}}}\left(\boldsymbol{\theta}_{\mathrm{e}},\cdot\right)$代表窃听者非法获得的解码器，$\boldsymbol{\theta}_{\mathrm{e}}$表示窃听者端语义解码器参数。

### 2.3　语义推理阶段的抗窃听策略

受到模型反转窃听攻击中随机排列和替换方法[8]的启发，本文采用了一种改进的随机排列方法。语义推理阶段的抗窃听方案如图3所示。在完成语义模型训练的MLaaS平台中，将语义编解码模型参数回传给终端的同时，窃听者也能通过不安全的应用程序接口（API, application program interface）、网络监听或内部交易方式从恶意MLaaS平台获取语义JSCC模型参数。在语义推理阶段，利用特征混淆技术对图像的行、列、通道进行随机排列，使得窃听者即使获取到数据也无法正确恢复原始图

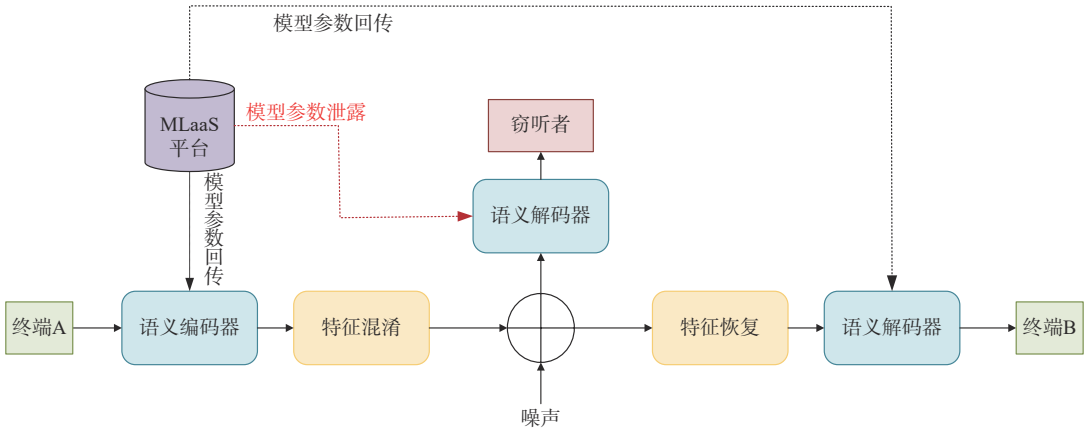



图3　语义推理阶段的抗窃听方案

像。此外，合法终端在接收到混淆数据后，可以通过预先约定的逆变换方法恢复出原始图像，从而实现安全的语义传输。

**2.3.1**　语义推理阶段合法设备的图像重建

在语义推理阶段，终端A向B传输图像数据$\boldsymbol{s}$，经过本地编码器模块生成待传输特征$\boldsymbol{y}$

$$\boldsymbol{y}=\Lambda_{\theta_a}\left(\boldsymbol{\theta}_a,\boldsymbol{s}\right) \tag{9}$$

其中，$\Lambda_{\theta_a}\left(\boldsymbol{\theta}_a,\cdot\right)$表示本地语义编码器输出，$\boldsymbol{\theta}_a$为本地语义编码器参数。

令$\boldsymbol{y}$通过特征混淆模块。特征混淆过程具体如下。

1)行混淆，如图4所示。对于传输的特征$\boldsymbol{y}$，首先沿着行维度对张量进行随机排列，将阵列$\left[y_0,y_1,\cdots,y_{h-1}\right]$的任一个随机排列定义为$\boldsymbol{u}$，其中$\left[0,1,\cdots,h-1\right]$的每个元素表示$\boldsymbol{y}$的行索引。在应用$\boldsymbol{u}$之后，得到行排列后的传输特征$\check{\boldsymbol{y}}$。

2)列混淆，如图5所示。定义$\boldsymbol{v}$为阵列$\left[\check{y}_0,\check{y}_1,\cdots,\check{y}_{w-1}\right]$的任一个随机排列，其中$\left[0,1,\cdots,w-1\right]$的每个元素表示$\check{\boldsymbol{y}}$的列索引。在应用$\boldsymbol{v}$之后，得到了行排列后的传输特征$\tilde{\check{\boldsymbol{y}}}$。

3)通道混淆，如图6所示。定义$\boldsymbol{q}$为阵列$\left[\tilde{\check{y}}_0,\tilde{\check{y}}_1,\cdots,\tilde{\check{y}}_{c-1}\right]$的任一个随机排列，其中$\left[0,1,\cdots,c-1\right]$的每个元素表示$\tilde{\check{\boldsymbol{y}}}$的通道索引。应用$\boldsymbol{q}$之后，得到传输语义特征

$$\underline{\tilde{\check{\boldsymbol{y}}}}=\boldsymbol{K}_e\left(\boldsymbol{\theta}_{\boldsymbol{K}_e},\Lambda_{\theta_a}\left(\boldsymbol{\theta}_a,\boldsymbol{s}\right)\right) \tag{10}$$

其中，$\boldsymbol{K}_e\left(\boldsymbol{\theta}_{\boldsymbol{K}_e},\cdot\right)$表示特征混淆加密器的输出，$\boldsymbol{\theta}_{\boldsymbol{K}_e}$表示加密参数，即对所述行、列、通道排列方案$\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{q}$使用A与B之间的共享密钥进行加密，将加密后的排列方案$\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{q}$与$\underline{\tilde{\check{\boldsymbol{y}}}}$一同传输。

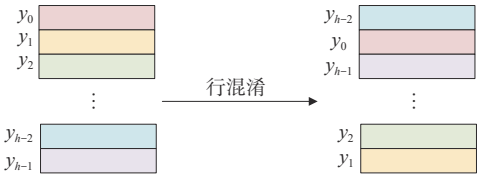



图4　行混淆

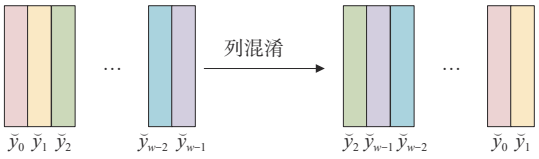



图5　列混淆

终端B接收到来自发射端的排列方案$\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{q}$与$\underline{\tilde{\check{\boldsymbol{y}}}}$，首先使用共享密钥对排列方案$\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{q}$进行

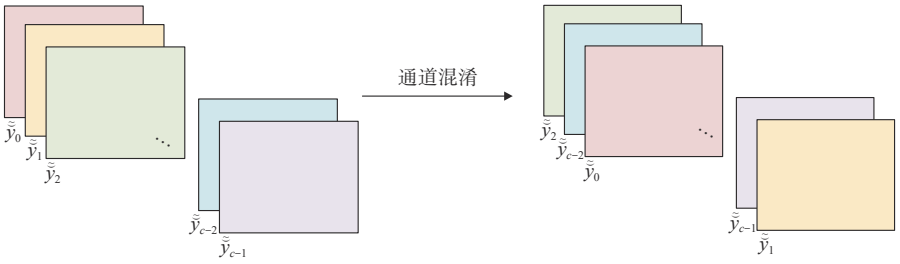



图6　通道混淆

解密，然后通过解码后的$\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{q}$从$\tilde{\underline{\boldsymbol{y}}}$中恢复被混淆的数据，最后通过语义解码器得到

$$\hat{\boldsymbol{s}} = D_{\theta_{\mathrm{b}}}\left(\boldsymbol{\theta}_{\mathrm{b}}, \boldsymbol{K}_{\mathrm{d}}\left(\boldsymbol{\theta}_{\boldsymbol{K}_{\mathrm{d}}}, \boldsymbol{K}_{\mathrm{e}}\left(\boldsymbol{\theta}_{\boldsymbol{K}_{\mathrm{e}}}, \Lambda_{\theta_{\mathrm{a}}}\left(\boldsymbol{\theta}_{\mathrm{a}}, \boldsymbol{s}\right)\right)\right)\right)\right) \tag{11}$$

其中，$\boldsymbol{K}_{\mathrm{d}}\left(\boldsymbol{\theta}_{\boldsymbol{K}_{\mathrm{d}}}, \cdot\right)$为解密器的输出，$\boldsymbol{\theta}_{\boldsymbol{K}_{\mathrm{e}}}$解密器参数，$D_{\theta_{\mathrm{b}}}\left(\boldsymbol{\theta}_{\mathrm{b}}, \cdot\right)$为终端B的语义解码器输出，$\boldsymbol{\theta}_{\mathrm{b}}$为语义解码器参数，$\hat{\boldsymbol{s}}$为B重建的图像。

#### 2.3.2　语义推理阶段窃听者的图像重建

窃听者通过恶意MLaaS平台获取训练好的语义通信神经网络参数，进而窃听终端A与B进行通信时的图像数据$\boldsymbol{s}$，由于窃听者并没有从MLaaS平台中得到特征混淆的相关内容，因此直接利用从MLaaS平台获得的解码器模块，试图从接收到的特征$\tilde{\underline{\boldsymbol{y}}}$中恢复图像

$$\hat{\boldsymbol{s}}_{\mathrm{e}} = D_{\theta_{\mathrm{e}}}\left(\boldsymbol{\theta}_{\mathrm{e}}, \boldsymbol{K}_{\mathrm{e}}\left(\boldsymbol{\theta}_{\boldsymbol{K}_{\mathrm{e}}}, \Lambda_{\theta_{\mathrm{a}}}\left(\boldsymbol{\theta}_{\mathrm{a}}, \boldsymbol{s}\right)\right)\right) \tag{12}$$

其中，$\hat{\boldsymbol{s}}_{\mathrm{e}}$为窃听者恢复的图像。

## 3　评价

本文使用CIFAR10图像数据集对神经网络进行训练。训练数据包括50 000张32像素×32像素分辨率的RGB图像，在CIFAR10数据集中的10 000张测试图像上进行评估，对比窃听者的窃听性能以及所提出的防御方法的有效性。为了测量图像质量，本文使用PSNR、MS-SSIM和LPIPS 3个指标进行评估，其中PSNR和MS-SSIM的值越高代表图像质量越好，LPIPS的值越低表示两张图像越相似，反之则差异越大。假设通信信道和窃听信道均为AWGN信道，SNR分别设置为1 dB、4 dB、7 dB、10 dB和13 dB，压缩比设置为1/16，$N_1$和$N_2$分别为2和4，$C_1$和$C_2$分别为128和256。使用Adam作为优化器，最初的学习率设置为$10^{-4}$。

图7展示了在SNR=13 dB下窃听者端与电力物联网终端的图像恢复质量。由图7可以观察到，窃听图像在视觉上是不可识别的，而终端能完好地恢复原始图像，证明了本文所提出的防御方法在防止窃听者窃听原始图像方面的有效性。

不同SNR下的PSNR和MS-SSIM分别如图8和图9所示。由图8、图9可以观察到：当窃听者恢复图片时，其图像的PSNR仅在10 dB以下，而MS-SSIM的值也不会超过0.1，这两项指标与正常通信情况下的性能相距甚远，后者通常能够保持PSNR

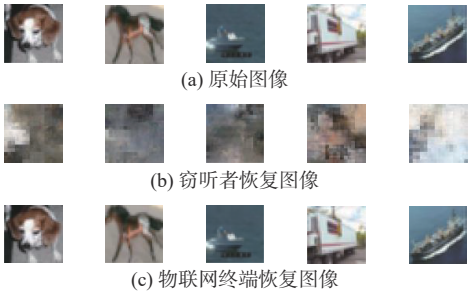



图7　在SNR=13 dB下窃听者端与电力物联网终端的图像恢复质量

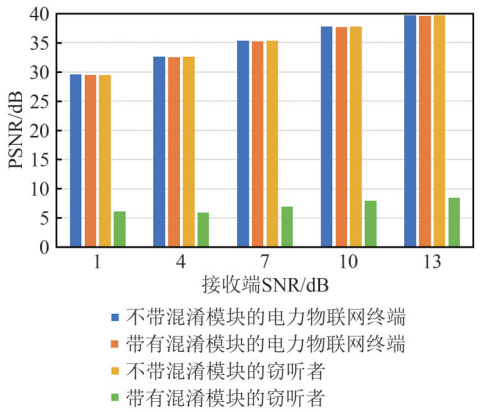



图8　不同SNR下的PSNR

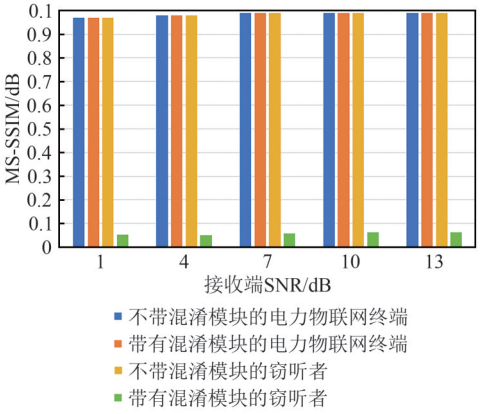



图9　不同SNR下的MS-SSIM

值在29 dB以上，以及MS-SSIM值在0.95以上。无论是PSNR还是MS-SSIM，引入特征混淆模块对图像的质量几乎没有负面影响。这一结果验证了关于所提出防御机制不会损害图像复原性能的初衷。特征混淆模块的引入造成窃听端恢复图像的PSNR指标最高下降81.72%，MS-SSIM指标最高下降94.54%，证明了所提出抗窃听方法的有效性。

不同SNR下窃听者端与电力物联网终端的LPIPS如图10所示。LPIPS依托于预训练的深度学习模型，通过提取和对比图像特征的距离来评估相似度，这种方式更能贴近人的感知效果。数据显示，

终端的LPIPS值在0.1以下，SNR达到13 dB时的值甚至降至0.005，几乎与原图无异。相对地，窃听者的LPIPS值普遍超过0.5，这意味着他们所能恢复的图像信息非常有限，即使在表现最佳的13 dB SNR条件下，也很难从接收到的图像中辨识出原始图像。这一结果充分说明本文提出的防御策略对阻止窃听者获取图像中有意义的信息极为有效。

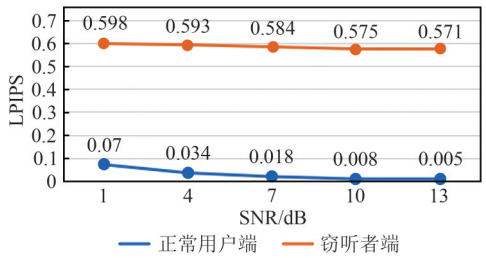



图10　不同SNR下窃听者端与电力物联网终端的LPIPS

为了评估所提出方法的优越性，本文将其与模型反转窃听攻击（MIEA, model inversion eavesdropping attack）[29]进行了对比，通过模型反转技术，利用窃听到的传输信号重建原始数据。该攻击分为白盒攻击和黑盒攻击两种类型，分别在攻击者了解和不了解模型参数的情况下进行。由于本文考虑的窃听者能从MLaaS平台获取完整的语义JSCC模型及参数，为了确保对比的公平性，特别与MIEA攻击中的白盒攻击进行了对比。窃听者恢复图像质量对比见表1。在相同条件下，即SNR为10 dB时，MIEA白盒攻击的窃听者能够成功重建高质量的图像，PSNR值和MS-SSIM值均高于本文所提方法的窃听者。具体来说，MIEA白盒攻击在SNR为10 dB时的PSNR为12.72 dB，MS-SSIM为0.07；而采用本文所提方法，在SNR为10 dB时的PSNR为7.92 dB，MS-SSIM为0.063。结果表明通过特征混淆技术，本文所提方法显著降低了窃听者恢复高质量图像的可能性，保护了数据的安全性。

**表1　　窃听者恢复图像质量对比**

| 方法 | PSNR/dB | MS-SSIM |
|---|---|---|
| MIEA | 12.72 | 0.07 |
| 本文 | 7.92 | 0.063 |

为了更全面地评估所提出的特征混淆技术在资源受限的电力物联网终端上的实际应用可行性，本文在实验部分增加了对所提方法在终端上的计算开销和时延的评估。实验设置如下。1) 终端设备配置：处理器为ARM Cortex-A72 @ 1.5 GHz，内存为1 GB RAM，操作系统为Linux。2) 测试数据集：图像数据集为CIFAR-10数据集，每个SNR条件下测试100张图像。3) 测试指标：计算开销，记录终端在进行特征混淆和特征恢复过程中的CPU使用率和内存占用；时延，测量终端从接收到混淆数据到完全恢复图像所需的时间。计算开销与平均时延的测试结果见表2。

**表2　　计算开销与平均时延测试结果**

| SNR/dB | 计算开销 | | 平均时延/ms |
|---|---|---|---|
| | CPU使用率 | 内存占用/MB | |
| 1 | 25.4% | 120.3 | 32.5 |
| 4 | 26.1% | 121.5 | 33.1 |
| 7 | 25.8% | 120.9 | 32.8 |
| 10 | 26.3% | 121.2 | 33.4 |
| 13 | 25.9% | 120.7 | 32.9 |

从表2中可以看出，无论SNR如何变化，终端在进行特征混淆和特征恢复过程中的CPU使用率和内存占用均保持在较低水平。具体来说，CPU使用率保持在25%左右，表明该方法对CPU资源的需求较低，不会显著影响终端的其他任务。内存占用在120 MB左右，在资源受限的终端上是可以接受的。以上结果表明所提出的特征混淆方法对终端的计算资源需求较小，适合在资源受限的电力物联网终端上应用。此外，终端从接收到混淆数据到完全恢复图像所需的平均时延在所有SNR条件下均保持在32.5~33.4 ms。这个时延在电力物联网终端的可接受范围内，说明所提出的特征混淆方法在实际应用中具有较好的实时性。由此可以得出，本文所提特征混淆方法在计算资源需求和时延方面均表现良好，能够在不影响终端性能的前提下，有效保护语义通信的安全性。

## 4　结束语

本文引入了一个建立在MLaaS平台之上的语义通信系统，并同时关注了该系统潜在的窃听风险。为了确保所设计的系统在保持电力物联网终端普适性和信息安全性的同时，还能准确完成语义传递任务，本文采用了特征混淆技术进行加密处理，以防止潜在的窃听者截获和理解传输中的语义内容。通过一系列的仿真实验，证实了引入特征混淆策略可以有效增强语义通信系统在隐私保护方面的性能。下一步笔者将对面向电力物联网终端的多模态语义

数据传输安全增强技术展开研究。

## 参考文献：


[1] WANG C X, YOU X H, GAO X Q, et al. On the road to 6G: visions, requirements, key technologies, and testbeds[J]. IEEE Communications Surveys & Tutorials, 2023, 25(2): 905-974.

[2] MAHMOOD N H, BERARDINELLI G, KHATIB E J, et al. A functional architecture for 6G special-purpose industrial IoT networks[J]. IEEE Transactions on Industrial Informatics, 2023, 19(3): 2530-2540.

[3] LI L L. A survey on intelligence-endogenous network: architecture and technologies for future 6G[J]. Intelligent and Converged Networks, 2024, 5(1): 53-67.

[4] FERRAG M A, FRIHA O, KANTARCI B, et al. Edge learning for 6G-enabled Internet of Things: a comprehensive survey of vulnerabilities, datasets, and defenses[J]. IEEE Communications Surveys & Tutorials, 2023, 25(4): 2654-2713.

[5] 秦志金, 赵荧荧, 李凡, 等. 多模态语义通信研究综述[J]. 通信学报, 2023, 44(5): 28-41.
QIN Z J, ZHAO T T, LI F, et al. Survey of research on multimodal semantic communication[J]. Journal on Communications, 2023, 44(5): 28-41.

[6] HUANG D L, GAO F F, TAO X M, et al. Toward semantic communications: deep learning-based image semantic coding[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(1): 55-71.

[7] YANG W T, DU H Y, LIEW Z Q, et al. Semantic communications for future Internet: fundamentals, applications, and challenges[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 213-250.

[8] HU Q Y, ZHANG G Y, QIN Z J, et al. Robust semantic communications with masked VQ-VAE enabled codebook[J]. IEEE Transactions on Wireless Communications, 2023, 22(12): 8707-8722.

[9] 张平, 牛凯, 姚圣时, 等. 面向未来的语义通信: 基本原理与实现方法[J]. 通信学报, 2023, 44(5): 1-14.
ZHANG P, NIU K, YAO S S, et al. Semantic communications for future: basic principle and implementation methodology[J]. Journal on Communications, 2023, 44(5): 1-14.

[10] 王碧舳, 罗倩雯, 卞志强, 等. 面向6G的语义通信系统[J]. 移动通信, 2023, 47(4): 2-6.
WANG B Z, LUO Q W, BIAN Z Q, et al. A semantic communication system for 6G networks[J]. Mobile Communications, 2023, 47(4): 2-6.

[11] XIE H Q, QIN Z J, LI G Y, et al. Deep learning based semantic communications: an initial investigation[C]//GLOBECOM 2020-2020 IEEE Global Communications Conference. Piscataway: IEEE Press, 2020: 1-6.

[12] XIE H Q, QIN Z J. A lite distributed semantic communication system for Internet of Things[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(1): 142-153.

[13] HUANG C Y, WANG J Z, CHEN H X, et al. zkMLaaS: a verifiable scheme for machine learning as a service[C]//Proceedings of the GLOBECOM 2022 - 2022 IEEE Global Communications Conference. Piscataway: IEEE Press, 2022: 5475-5480.

[14] FARSAD N, RAO M, GOLDSMITH A. Deep learning for joint source-channel coding of text[C]//Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2018: 2326-2330.

[15] 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.

[16] WANG S X, YANG K, DAI J C, et al. Distributed image transmission using deep joint source-channel coding[C]//Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2022: 5208-5212.

[17] 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.

[18] WU H T, SHAO Y L, MIKOLAJCZYK K, et al. Channel-adaptive wireless image transmission with OFDM[J]. IEEE Wireless Communications Letters, 2022, 11(11): 2400-2404.

[19] TUNG T Y, GÜNDÜZ D. DeepWiVe: deep-learning-aided wireless video transmission[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(9): 2570-2583.

[20] ZHANG W Y, BAI K Y, ZEADALLY S, et al. DeepMA: end-to-end deep multiple access for wireless image transmission in semantic communication[J]. IEEE Transactions on Cognitive Communications and Networking, 2024, 10(2): 387-402.

[21] DAI J X, FAN H, ZHAO Z X, et al. Secure resource allocation for integrated sensing and semantic communication system[C]//Proceedings of the 2024 IEEE International Conference on Communications Workshops (ICC Workshops). Piscataway: IEEE Press, 2024: 1225-1230.

[22] 谷勇浩, 郭振洋, 刘威歆. 匿名化隐私保护技术性能评估方法研究[J]. 信息安全研究, 2019, 5(4): 293-297.
GU Y H, GUO Z Y, LIU W X. Research on performance evaluation method of anonymization privacy preservation technologies[J]. Journal of Information Security Research, 2019, 5(4): 293-297.

[23] ABADI M, CHU A, GOODFELLOW I, et al. Deep learning with differential privacy[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2016: 308-318.

[24] CHEN Y H, YANG Q Q, SHI Z G, et al. The model inversion eavesdropping attack in semantic communication systems[C]//Proceedings of the GLOBECOM 2023-2023 IEEE Global Communications Conference. Piscataway: IEEE Press, 2023: 5171-5177.

[25] YANG K, WANG S X, DAI J C, et al. SwinJSCC: taming swin transformer for deep joint source-channel coding[J]. IEEE Transactions on Cognitive Communications and Networking, 2025, 11(1): 90-104.

[26] YANG K, WANG S X, DAI J C, et al. WITT: a wireless image

transmission transformer for semantic communications[C]//Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2023: 1-5.

[27] NGUYEN L X, KIM K, LIN TUN Y, et al. Optimizing multi-user semantic communication via transfer learning and knowledge distillation[J]. IEEE Communications Letters, 2025, 29(1): 90-94.

[28] YU L, WU S Y, GABBOUJ M. Multi-swin transformer based spatio-temporal information exploration for compressed video quality enhancement[J]. IEEE Signal Processing Letters, 2024, 31: 1880-1884.

[29] CHEN Y H, YANG Q Q, SHI Z G, et al. The model inversion eavesdropping attack in semantic communication systems[C]//Proceedings of the GLOBECOM 2023-2023 IEEE Global Communications Conference. Piscataway: IEEE Press, 2023: 5171-5177.

[作者简介]

**杨会峰**（1973– ），男，国网河北省电力有限公司信息通信分公司副总经理、高级工程师，主要研究方向为通信传输网络、交换网络、智能电网等。

**尚立**（1982– ），男，国网河北省电力有限公司信息通信分公司高级工程师，主要研究方向为数据传输网络、应急通信等。

**崔俊彬**（1989– ），男，国网河北省电力有限公司信息通信分公司工程师，主要研究方向为数据传输网络、交换网络等。

**刘红艳**（1989– ），女，国网河北省电力有限公司信息通信分公司工程师，主要研究方向为交换网络、数据通信网等。

**王九成**（1993– ），男，国网河北省电力有限公司信息通信分公司工程师，主要研究方向为数据通信网络、应急通信等。

**蔺鹏**（1987– ），男，北京万可信息技术有限公司总经理、工程师，主要研究方向为5G/6G网络管理与优化、网络智能管控、数据网络安全、智能电网通信网等。