

基于生成式模型的新型机器语义通信方法研究



Novel Generative Model-Based Machine-Type Semantic Communications

承楠/CHENG Nan¹, 王秀程/WANG Xiucheng¹,
沈学民/SHEN Xuemin²

(1. 西安电子科技大学, 中国 西安 710071;

2. 滑铁卢大学, 加拿大 滑铁卢 N2L3G1)

(1. Xidian University, Xi'an 710071, China;

2. University of Waterloo, Waterloo N2L3G1, Canada)

DOI: 10.12142/ZTETJ.202504007

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.tn.20250716.1341.007.html>

网络出版日期: 2025-07-16

收稿日期: 2025-06-10

摘要: 创新性地提出了一种基于隐扩散模型 (LDM) 的语义通信架构, 能够利用预训练生成模型实现无需重新训练的语义特征还原。借助 LDM 固有的去噪能力, 该架构在面对信道噪声扰动和分布外输入时, 依然能够保持稳定的重构性能。此外, 该架构支持灵活集成外部大模型资源, 显著提升了语义通信系统的演进能力。在大规模图像数据集上的实验结果表明, 该方法在低信噪比 (SNR) 条件下仍具有优异的图像恢复与语义保真能力, 尤其在基于学习的感知相似度等语义指标上显著优于现有主流方法。本研究为资源受限设备的机器语义通信提供了全新思路, 具有良好的实用性与推广潜力。

关键词: 语义通信; 隐扩散模型; 机器通信; 机器视觉; 信道鲁棒性

Abstract: A novel semantic communication architecture based on the Latent Diffusion Model (LDM) is proposed. This architecture leverages pre-trained generative models to enable semantic feature reconstruction without the need for task-specific retraining. By exploiting the intrinsic denoising capability of LDM, the proposed system maintains stable reconstruction performance under channel noise perturbations and out-of-distribution inputs. Additionally, the architecture supports the flexible integration of external large-scale models, significantly enhancing the evolutionary capability of semantic communication systems. Experimental results on large-scale image datasets demonstrate that our method achieves superior image recovery and semantic fidelity, even under low signal-to-noise ratio (SNR) conditions. In particular, it significantly outperforms existing mainstream approaches in terms of learning-based perceptual similarity metrics. This study offers a new perspective for deploying machine-type semantic communication on resource-constrained devices and exhibits strong potential for practical applications and large-scale deployment.

Keywords: semantic communication; Latent Diffusion Model; machine-type communication; machine vision; channel robustness

引用格式: 承楠, 王秀程, 沈学民. 基于生成式模型的新型机器语义通信方法研究 [J]. 中兴通讯技术, 2025, 31(4): 41-47. DOI: 10.12142/ZTETJ.202504007

Citation: CHENG N, WANG X C, SHEN X M. Novel generative model-based machine-type semantic communications [J]. ZTE technology journal, 2025, 31(4): 41-47. DOI: 10.12142/ZTETJ.202504007

随着通信系统从 5G 加速向 6G 演进, 物联网中各类监测设备数量激增, 传统以比特无差错传输为目标的通信方式面临严重的带宽瓶颈和高时延问题, 难以满足机器通信对实时性与能效的需求。以智慧城市为例, 海量部署的摄像头与传感器广泛应用于目标检测、人脸识别等高层视觉任务。然而, 由于边缘设备在算力与能耗方面的限制, 终端通常需将大量原始数据上传至云端处理, 这不仅导致通信拥

塞, 还可能造成服务质量明显下降^[1]。为应对这一挑战, 国际电信联盟提出从“传输比特”向“传递语义”的通信范式转型, 旨在支撑超高速率、低时延与高密度连接需求, 从而更好地服务大规模机器通信与人机融合等 6G 关键应用场景^[2]。语义通信通过显著压缩信息冗余、降低通信时延并增强数据安全性, 顺应了 6G “通信-感知-计算一体化”的发展趋势, 为构建高效、智能且可靠的下一代机器通信系统提供了新路径^[3]。

近年来, 得益于神经网络在特征表达与数据压缩方面的

基金项目: 国家重点研发计划项目 (2024YFB907500)

强大能力，基于深度学习的语义通信系统取得了显著进展^[4]。研究表明，神经网络能够通过自动提取数据中的语义特征，显著提升通信效率，甚至实现极低码率下的高保真重构^[5]。然而，当前主流的神经网络语义通信系统仍面临四大核心挑战：噪声泛化能力差、可扩展性不足、系统演化能力弱以及训练成本高。首先，现有系统通常依赖在特定信噪比下训练的神经网络自编码器，当实际信道环境与训练条件存在差异时，系统性能会明显下降^[4]。这一问题的根源在于判别式神经网络难以适应输入数据分布偏移（OOD）导致的特征变化^[6]。具体来说，信道噪声扰动会导致解码端接收到的特征分布与训练数据分布不一致，从而严重削弱系统对未知环境的鲁棒性。更为严峻的是，已有研究表明，在特征空间直接添加噪声将极大削弱神经网络的判别能力，而无线信道噪声正是以这种方式干扰语义特征传输，从理论上限制了现有结构的抗噪能力。其次，神经网络模型的可扩展性存在明显瓶颈。现有系统往往只能在与训练集分布高度一致的数据集上展现优异性能，但在实际应用中，数据分布经常发生漂移，或面临跨场景迁移需求，此时系统性能往往急剧下降^[7]。为提升泛化能力，部分研究尝试训练多个专用模型以适应不同场景。然而，这种方法不仅大幅增加模型存储与系统运维成本，还难以满足实际部署对灵活性和低资源消耗的要求。此外，系统演化能力不足是当前深度学习驱动语义通信面临的关键挑战之一。与计算机视觉等领域可通过大规模预训练模型实现跨任务迁移和持续学习不同，当前语义通信系统仍严重依赖端到端的专用训练流程，难以有效复用现有通用大模型资源，这限制了其在动态通信环境中实现自适应进化和迭代优化的潜力。最后，高昂的训练与部署成本也制约了语义通信的广泛应用。相比于传统通信方案，基于神经网络的端到端系统通常需要大量标注数据进行有监督训练，且需针对不同应用场景反复调整网络结构和优化参数，增加了系统开发、维护与部署的复杂度。

从统计学习角度看，当前语义通信系统面临的诸多挑战，其根源在于主流方法普遍采用判别式神经网络，缺乏对数据先验分布的建模能力。这导致在信道噪声引起的特征分布漂移条件下，系统难以实现鲁棒的语义信息恢复。相比之下，生成式人工智能（特别是扩散模型）因其对复杂数据分布的显式建模能力，能够从任意噪声干扰中恢复高质量语义特征，天然适用于应对无线通信中的信道不确定性和分布外输入等挑战^[8]。受此启发，本文提出基于隐扩散模型（LDM）^[9]的智能语义通信架构。该方法将信道噪声建模与语义特征生成过程有机结合，构建了从噪声“逆演化”到高维语义重构的端到端学习范式。具体而言，所提方法以大规

模预训练扩散模型为基础，将信道噪声视为高斯扰动，并通过扩散-去噪反演过程显式恢复原始语义特征，显著提升了系统对多样化信道条件和复杂数据分布的泛化与适应能力。同时，得益于预训练模型的知识迁移能力，所提系统在低样本甚至零样本条件下仍能实现鲁棒的语义传输，大幅降低了模型训练与系统部署的资源开销。此外，该架构支持灵活集成外部大模型资源，为语义通信系统的持续演化和任务自适应能力提供了坚实基础。

1 系统模型与问题描述

语义通信系统通常包括信源、联合信源信道编码（JSCC）模块、噪声信道、噪声抑制与解码模块以及信宿。其核心目标是通过有效提取、压缩、传输和还原数据中的语义信息，实现高效可靠的端到端传输。具体来说，信源 x 代表待传输的原始数据。该数据首先经过JSCC模块，将原始信息压缩为低维语义特征表示 z ，以显著降低数据冗余和带宽占用。随后， z 通过物理信道进行传输，在此过程中受到加性高斯白噪声（AWGN）等信道噪声影响。 y 为接收端实际收到的信号。在信宿侧，系统首先通过噪声抑制模块对 $\hat{z}$ 进行去噪与特征恢复，随后进入解码模块，以最大限度还原原始语义信息，输出信宿重构数据 $\hat{x}$ 。假设信源数据 $x \in R^m$ ，其通过联合编码器 $\mathcal{E}_\theta$ 映射为低维语义特征 $z = \mathcal{E}_\theta(x) \in R^k$ ， $k < m$ 。其通过AWGN信道后所接收到的信号 $\hat{y}$ 服从公式（1）：

$$y = z + n, n \sim \mathcal{N}(0, \sigma^2 I) \quad (1),$$

其中， n 为信道噪声， σ^2 为噪声功率。信宿端首先利用噪声抑制模块 $\mathcal{G}_\phi$ 对信号 y 进行处理，获得去噪后的特征 $\hat{z} = \mathcal{G}_\phi(y) \in R^k$ 。之后通过解码器 $\mathcal{D}_\psi$ 还原信源数据，得到重构输出 $\hat{x} = \mathcal{D}_\psi(\hat{z}) \in R^m$ 。语义通信系统的目标是 최소화信宿重构数据 $\hat{x}$ 与信源 x 之间的均方误差（MSE）。该优化目标可以表示为公式（2）：

$$\begin{aligned} \min_{\theta, \phi, \psi} \mathbb{E} [\|\hat{x} - x\|_2] &= \min_{\theta, \phi, \psi} \mathbb{E} [\|\mathcal{D}_\psi(\mathcal{G}_\phi(\mathcal{E}_\theta(x)) + n) - x\|_2] \\ \text{s.t. } n &\sim \mathcal{N}(0, \sigma^2 I) \end{aligned} \quad (2),$$

其中， θ 、 ϕ 、 ψ 分别表示编码器、噪声抑制器和解码器的可训练参数。该架构面临两个主要技术挑战：首先，由于 $k < m$ 导致的维度压缩，解码器 $\mathcal{D}_\psi$ 需要求解一个非满秩矩阵构成的方程组，这使得准确恢复信源数据具有本质困难。其次，无线信道的噪声干扰会导致接收端获取的语义特征产生扰动，难以直接还原为高保真原始数据。与显式数据（如图

像或音频)的噪声干扰不同,神经网络对潜空间语义特征向量的噪声扰动表现出更低的鲁棒性^[6]。这是因为潜空间特征通常是高度压缩且分布紧致的抽象表示,高斯扰动更容易导致其分布偏移,从而破坏语义结构的完整性。这种特性使得解码器难以有效捕捉并还原原始信息。因此,在语义通信场景中,如何在潜空间层面对噪声扰动进行精确建模与有效修复,成为决定系统解码质量的关键科学问题。该挑战的核心在于建立能够保持语义结构完整性的鲁棒特征表示方法,以及开发针对压缩特征空间的专用去噪算法。

2 基于去噪扩散模型的智能语义通信系统理论分析

2.1 扩散模型去噪理论基础

扩散模型(DM)最初作为一种高质量数据生成方法被提出,其核心机制是通过多步随机扰动(扩散过程)及其逆向过程(去噪过程),建立数据分布与噪声分布之间的映射^[10]。近年来,得益于对复杂数据分布的强大建模能力和优异的去噪性能,DM在信号处理领域展现出重要应用价值,特别适用于高斯噪声环境下的信号恢复任务。DM的工作原理是通过在数据空间中逐步注入高斯噪声,构建一个从真实数据分布到标准高斯分布的马尔可夫链。设原始信号为 x_0 ,经过 t 步高斯噪声扰动后得到 x_t ,该正向扩散过程可用随机微分方程(SDE)描述^[11]:

$$dx_t = f_t x_t dt + g_t dw_t \quad (3),$$

其中, f_t 和 g_t 是和扩散次数相关的漂移项和扩散项, w_t 是标准Wiener过程。对于高斯噪声扩散,通常设 $f_t x_t = h_t$, $g_t^2 = 1$,则该过程可简化为下述积分方程形式:

$$x_t = x_0 + \int_0^t h_t dt + \int_0^t dw_t \quad (4),$$

其中, $\int_0^1 h_t dt + x_0 = 0$ 。通常可令 $h_t = -x_0$,可以得到公式(5)^[12]:

$$x_t = (1-t)x_0 + \sqrt{t}\epsilon \quad (5),$$

其中, ϵ 是标准高斯变量。要实现去噪,需对上述扩散过程进行反向建模。逆扩散过程同样可以用SDE来描述,其目标是从含噪观测 x_t 逐步还原出干净信号 x_0 。逆过程可表述为:

$$dx_t = [f_t x_t - g_t^2 \nabla_{x_t} \log p_t(x_t) + g_t d\bar{w}_t] \quad (6),$$

其中, $g_t^2 \nabla_{x_t} \log p_t(x_t)$ 表示关于当前状态的概率密度梯度, $g_t d\bar{w}_t$ 为独立于正向过程的Wiener过程。在实际神经网络实

现中,常采用离散时间的迭代形式。对于一步逆扩散,逆过程可简化为:

$$x_{t-\Delta t} = x_t - \frac{\Delta t}{t} \epsilon + \sqrt{\Delta t(t-\Delta t)} \tilde{\epsilon} \quad (7)。$$

在扩散模型的数学框架中, ϵ 代表噪声预测项,而 $\tilde{\epsilon}$ 是为了平衡方差在去噪过程中引入的独立标准高斯噪声。基于通用逼近理论,一个具有足够容量且经过充分训练的神经网络能够逼近任意分布映射关系。这一理论保证使得扩散模型理论上具备去除任意强度高斯噪声的能力。DM的本质即“去噪扩散概率模型”(DDPM)^[10]。大量实验研究表明,该模型能够在不同噪声水平和多样化的数据分布条件下实现高质量去噪。具体而言,当预训练的扩散模型接收到受高斯噪声污染的信号时,只需根据噪声强度选择合适的扩散步数(即参数 t),即可通过逐步去噪过程恢复出高保真度的原始信号。

这一理论特性为扩散模型在通信系统中作为通用噪声抑制模块的应用奠定了理论基础。该模型具有3个显著优势:首先,其对不同强度的噪声均表现出良好的适应性;其次,能够有效建模复杂的数据分布特征;最后,在各种条件下都能保持稳定的去噪性能。这些特性使扩散模型特别适用于处理无线通信信道中的噪声干扰问题,为其作为通信系统噪声抑制模块的可行性提供了理论支撑。

2.2 扩散步骤与噪声信号之间的关联

传统基于神经网络的去噪方法存在明显的局限性,通常需要针对特定信噪比(SNR)分别训练专用模型。当面对OOD噪声水平时,这类方法往往会出现明显的退化。而DM在理论上具有对任意信噪比高斯噪声自适应去噪的能力。该能力的核心机制在于:DM能够根据接收信号中的噪声能量特征,动态确定最优的扩散步数 t 。这一自适应调整过程确保了模型去噪能力与实际信号噪声特性精确匹配,从而保证最优的去噪效果(如图1所示)。

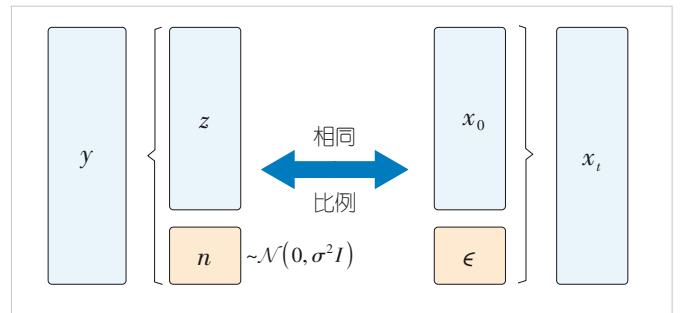


图1 扩散去噪比例示意图

具体而言，接收信号的SNR可通过公式（8）来定义：

$$\text{SNR} = \frac{\mathbb{E}[\|z\|_2^2]}{\mathbb{E}[\|n\|_2^2]} \quad (8)。$$

为了实现自适应去噪，需要确保实际接收信号的SNR与DM输入的 x_t 相匹配。为此，需使得二者满足：

$$\frac{\mathbb{E}[\|z\|_2^2]}{\mathbb{E}[\|n\|_2^2]} = \frac{\mathbb{E}[\|(1-t)x_0\|_2^2]}{\mathbb{E}[\|\sqrt{t}\epsilon\|_2^2]} \quad (9)。$$

假设信号发射功率 $\mathbb{E}[\|x_0\|_2^2] = \gamma$ ，那么对于任意 $\mathbb{E}[\|z\|_2^2]$ 和信道噪声强度 σ^2 有，当 t 满足公式（10）时，公式（9）成立：

$$t = \frac{2 + \frac{\mathbb{E}[\|z\|_2^2]}{\sigma^2\gamma} - \sqrt{\frac{\mathbb{E}[\|z\|_2^2]}{\sigma^2\gamma} \left(\frac{\mathbb{E}[\|z\|_2^2]}{\sigma^2\gamma} + 4 \right)}}{2} \quad (10)。$$

公式（10）可通过如下方法证明。首先对于 $\mathbb{E}[\|(1-t)x_0\|_2^2]$ 有：

$$\mathbb{E}[\|(1-t)x_0\|_2^2] = (1-t)^2 \mathbb{E}[\|x_0\|_2^2] = (1-t)^2 \gamma \quad (11)。$$

相似的有：

$$\mathbb{E}[\|\sqrt{t}\epsilon\|_2^2] = t \quad (12)。$$

因此，原方程可写作：

$$\frac{\mathbb{E}[\|z\|_2^2]}{\sigma^2} = \frac{(1-t)^2 \gamma}{t} \quad (13)。$$

上述关于 t 的一元二次方程。通过求根公式可以得到：

$$t = \frac{2 + \frac{\mathbb{E}[\|z\|_2^2]}{\sigma^2\gamma} - \sqrt{\frac{\mathbb{E}[\|z\|_2^2]}{\sigma^2\gamma} \left(\frac{\mathbb{E}[\|z\|_2^2]}{\sigma^2\gamma} + 4 \right)}}{2} \quad (14)。$$

如图2所示，实际通信系统中接收信号的统计分布往往与DM训练时的输入分布存在偏差。若直接将观测信号输入预训练的DM，可能引发OOD问题，导致去噪性能下降。为解决这一问题，需要通过对接收信号进行线性缩放处理，使其统计特性与DM扩散过程中的变量分布保持一致。这种预处理方法能够确保DM的去噪能力在不同信噪比条件下均能得到有效发挥。

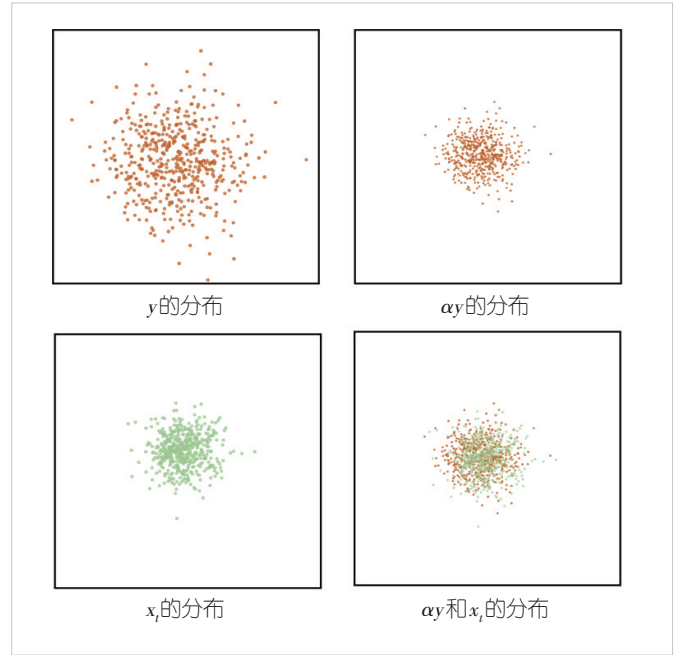


图2 数据分布示意图

通过缩放 y 至 αy ，可以让 αy 的分布与 x_t 一致，这要求 α 满足如下关系：

$$\alpha = \sqrt{\frac{(1-t)^2 \mathbb{E}[\|z\|_2^2] + t}{\mathbb{E}[\|z\|_2^2] + \sigma^2}} \quad (15)。$$

公式（15）的推导过程如下：

首先，期望 $\mathbb{E}[\|\alpha y\|_2^2]$ 可表示为：

$$\mathbb{E}[\|\alpha y\|_2^2] = \mathbb{E}[\|\alpha(z+n)\|_2^2] = \alpha^2 (\mathbb{E}[\|z\|_2^2] + \mathbb{E}[\|n\|_2^2]) \quad (16)。$$

同时， $\mathbb{E}[\|x_t\|_2^2]$ 的计算如下：

$$\mathbb{E}[\|x_t\|_2^2] = \mathbb{E}[\|(1-t)x_0 + \sqrt{t}\epsilon\|_2^2] = (1-t)^2 \mathbb{E}[\|x_0\|_2^2] + t \quad (17)。$$

通过求解 $\mathbb{E}[\|x_t\|_2^2] = \mathbb{E}[\|\alpha y\|_2^2]$ 即可得到 α 的表达。

2.3 基于LDM的语义通信架构

根据上述理论分析，针对语义通信系统在不同信噪比环境下易受分布外噪声干扰、模型泛化能力不足等问题，本文提出了一种以LDM为核心的端到端语义通信架构，如图3所示。该架构充分融合了变分自编码器（VAE）与DM的各自优势，实现了信源语义信息的高效压缩、任意信噪比下的鲁棒恢复及系统低成本可持续演化。在本架构中，信源数据 x

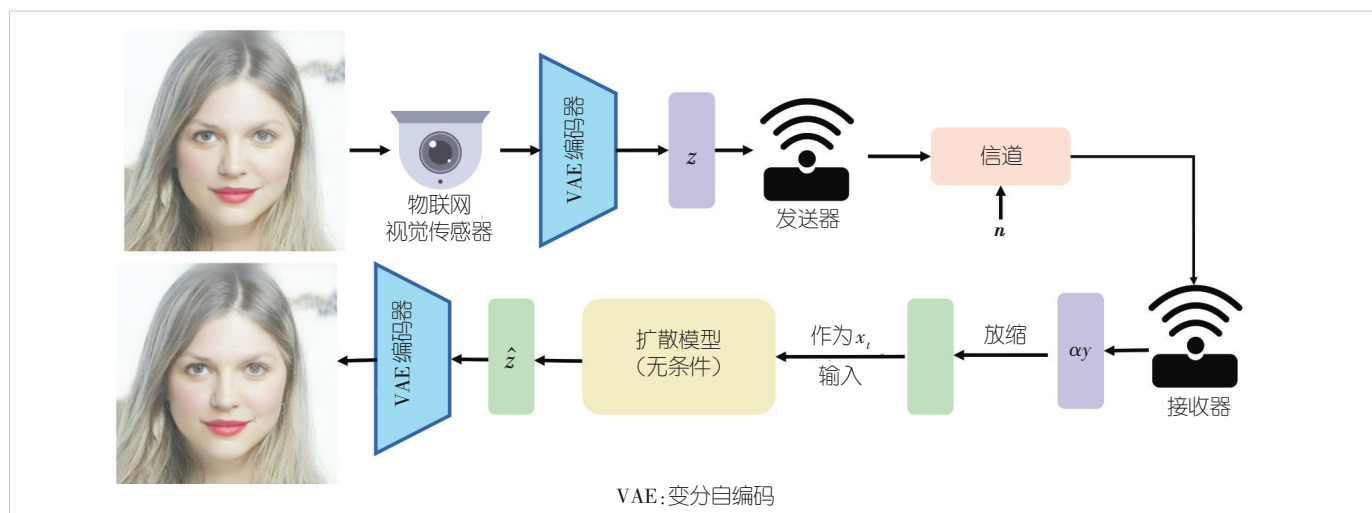


图3 基于隐扩散模型的语义通信系统结构图

首先通过VAE编码器得到隐变量 z ，实现数据语义层的紧凑表示，从而显著降低传输带宽和冗余信息。隐变量 z 经过加性白噪声信道（AWGN）传输后，接收端获得信号 y 。由于信道噪声的影响，隐变量在传输后与原始分布存在差异，直接进行恢复可能导致性能退化，尤其在信噪比剧烈变化时，传统神经网络方法往往难以适应这种分布漂移。

为此，本文引入了LDM作为信号的去噪主干模块。接收端根据当前观测信号的噪声方差 σ^2 ，首先自适应地确定扩散去噪的步数 t 。为了确保输入数据的分布与LDM训练阶段一致，我们对 y 施加了缩放变换 αy 。经过缩放和步数自适应调整后的接收信号与 t 一起输入预训练的LDM。通过扩散逆过程，我们实现了对高斯噪声的逐步抑制，最终获得去噪后的隐变量 $\hat{z}$ 。去噪后的 $\hat{z}$ 经过VAE解码器还原为重构数据 $\hat{x}$ ，从而实现信源到信宿的高保真语义信息恢复。

该架构的核心优势在于其对任意SNR条件下噪声的自适应处理能力，以及对预训练大模型能力的高效继承。首先，本文提出的LDM语义通信架构通过联合自适应的去噪步数与缩放因子机制，能够动态调节模型在不同SNR条件下的去噪策略，从而有效抑制接收语义特征中的残余噪声。该机制有效克服了传统神经网络架构在信道条件变化下容易产生OOD输入的问题，增强了系统在复杂信道环境下的鲁棒性与泛化能力。其次，本文的方法通过引入预训练的生成式大模型，有效利用其对数据先验分布的深层建模能力，实现对带噪语义特征的最大后验恢复，从而提升了带噪语义特征的恢复能力。相比之下，现有基于条件生成模型的方法在处理语义传输任务时，往往因条件编码器难以适配多样化噪声分布，导致通过信道传输后的语义特征形成OOD输入，严重削弱了生成模型的重构性能。而本文所采用的LDM将

噪声视为潜空间中的扰动变量，借助其训练过程中已学得完整数据分布信息，可直接实现跨信道条件的高质量语义还原，无须针对不同信噪比场景重新训练或微调。

3 实验设置与结果分析

3.1 实验设置

为验证所提出的基于LDM的语义通信方法在机器通信任务中的有效性，本文选取与人脸图像识别相关的语义传输场景作为代表应用，并基于大规模高质量人脸图像数据集CelebA-HQ开展仿真实验。该数据集包含多样化的人脸图像，具备丰富的身份语义与结构特征，能有效模拟物联网终端中摄像头采集的人体视觉信息。考虑到所提方法无须重新训练，直接采用了公开发布的预训练LDM，未针对特定信道或任务进行任何微调。

具体而言，原始图像首先通过VAE编码器映射至低维潜空间表示，随后在该潜空间中引入高斯信道噪声，以模拟无线信道的干扰过程。最后，通过LDM的反向扩散过程对受扰语义特征进行去噪恢复，并经VAE解码器还原图像或提取语义特征。为全面评估所提方法的鲁棒性与语义保真性能，本文设计与两种典型深度JSCC方法的对比实验，分别为：

- 1) Deep JSCC^[13]：基于卷积神经网络设计的端到端图像传输结构，具有较强的空间特征提取与压缩能力；
- 2) Deep JSCC-l++^[14]：融合Transformer架构的先进模型，利用注意力机制建模图像中的长程依赖关系，增强复杂信道下的语义保持能力。

上述基线方法均按照其原始论文在CelebA-HQ数据集

上进行复现和训练，并在不同信噪比条件下进行对比测试。为了兼顾像素精度与语义保真，本文采用了两类指标进行评估：1) 图像重建指标，包括峰值信噪比 (PSNR) 和结构相似性 (SSIM)；2) 语义保持指标，即学习感知相似度 (LPIPS)，这是一种基于深度神经网络感知特征的图像相似性度量指标，用于评估重建图像与参考图像在人类视觉感知下的相似程度。

3.2 结果分析

图4展示了在不同信噪比条件下各算法的图像恢复效果。从图4可以看出，所提出的LDM方法即使在完全无训练的条件下，仍能在低信噪比环境中实现较高质量的图像还原。如图5所示，在信噪比低于-7.5 dB的低信噪比场景下，所提出的方法还原的人脸图像在轮廓、结构及语义层面均表现出良好的保持能力，基本能够恢复关键的身份特征。相比之下，两种经典的深度JSCC方法在相同信道条件下的重建结果则出现模糊和失真，表现出明显的退化现象。这一现象可归因于它们依赖判别式训练，在面对OOD信道噪声时缺乏鲁棒性，导致语义信息无法有效还原。

图6展示了不同方法在多个性能指标下的定量比较，这些性能指标包括PSNR、SSIM与LPIPS。在LPIPS指标上，

所提出的方法显著优于所有对比方法，这表明该方法在保持人类感知语义一致性方面具有突出优势。这意味着，即使在高噪声或未知信道条件下，LDM所生成的图像在视觉语义上依然贴近原始图像，更加契合物联网设备中语义感知与识

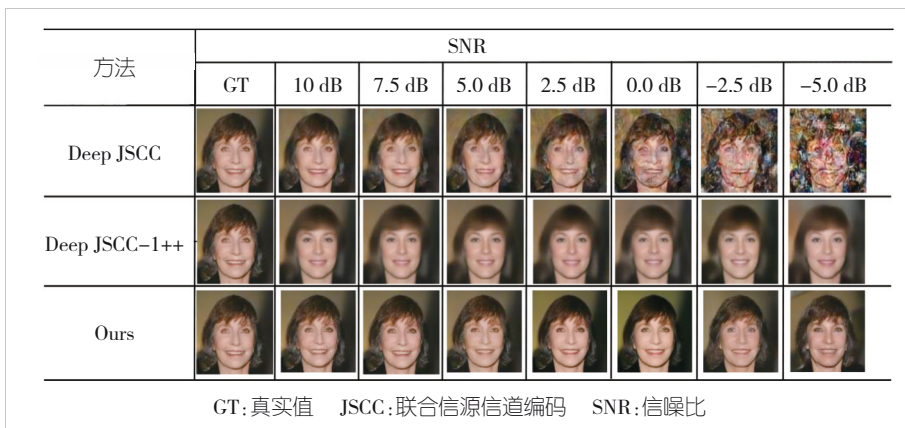


图4 不同信噪比下图像恢复效果对比图

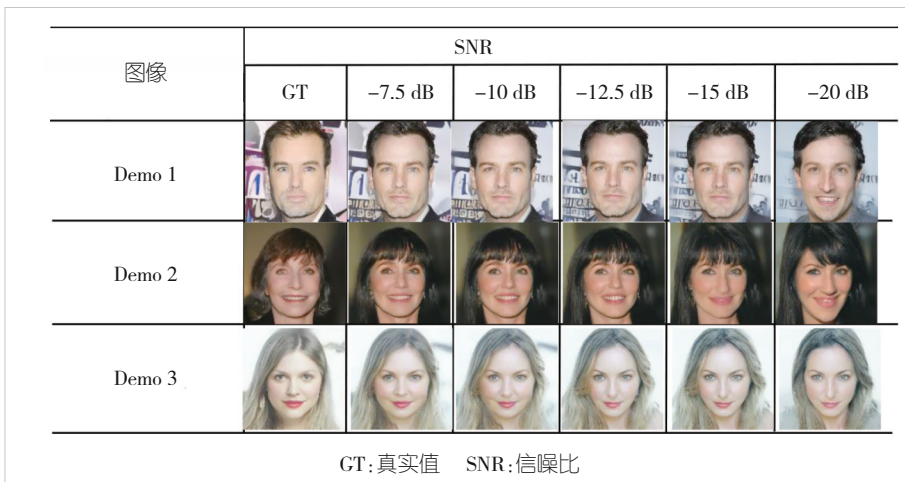


图5 本文所提出方法在低信噪比下图像恢复效果图

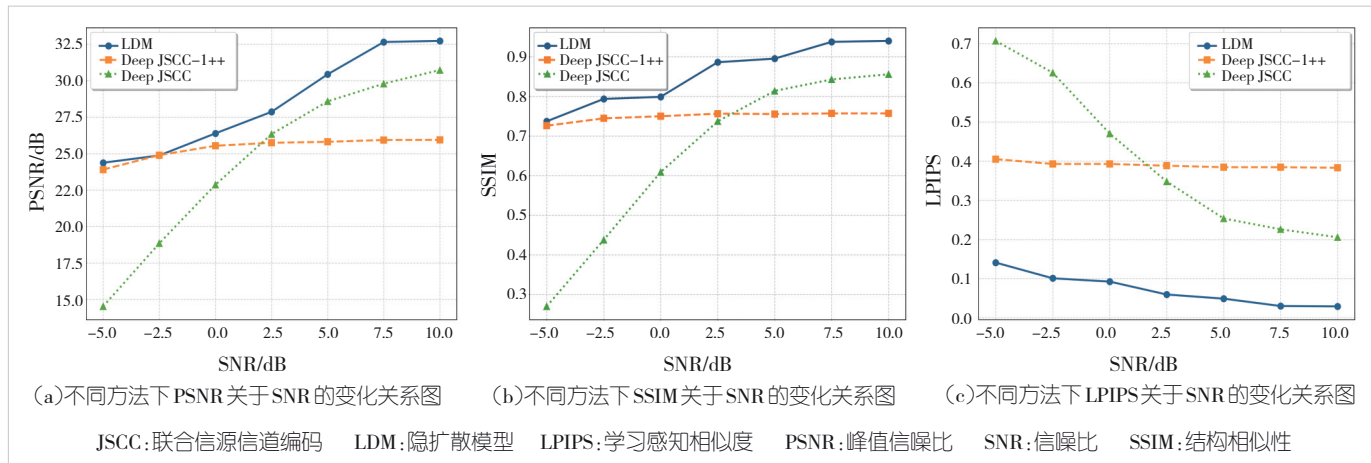


图6 不同信噪比下性能对比图

别任务的实际需求。综上所述,实验结果充分验证了本文方法在语义保真性、噪声鲁棒性及泛化能力方面的综合优势,为资源受限终端实现高效、可靠的语义通信提供了有力支撑。

4 结束语

本文针对6G机器通信场景中的语义通信需求,提出了一种基于LDM的语义通信新架构。该架构利用预训练的LDM对信道噪声进行建模与逆演化,无须重新训练即可鲁棒地还原语义信息,特别适用于资源受限的机器通信场景。在CelebA-HQ数据集上的仿真结果表明,所提方法即使在低信噪比条件下,依然具备出色的图像重建与语义保持能力,并且在LPIPS等感知指标上明显优于现有的典型深度通信模型,验证了其强大的泛化性与实用性。本方法允许直接采用预训练的LDM作为语义通信架构的主干,为后续智能语义通信系统的低成本、可持续演化发展提供了新的思路 and 理论基础。未来,我们将进一步研究模型轻量化与多模态扩展,以推动其在物联网与边缘智能中的实际部署。

参考文献

- [1] 林美玉. 5G网络赋能物联网安全[J]. 中兴通讯技术, 2022, 28(6): 70–74. DOI: 10.12142/ZTETJ.202206012
- [2] DANG S P, AMIN O, SHIHADA B, et al. What should 6G be? [J]. Nature electronics, 2020, 3(1): 20–29. DOI: 10.1038/s41928-019-0355-6
- [3] 廖俊淇, 魏昕, 周亮. 人工智能驱动的跨模态语义通信系统[J]. 中兴通讯技术, 2024, 30(S1): 33–39. DOI: 10.12142/ZTETJ.2024S1005
- [4] ZHANG P, XU W J, LIU Y M, et al. Intellicise wireless networks from semantic communications: a survey, research issues, and challenges [J]. IEEE communications surveys & tutorials, 2024, 27(3): 2051–2084. DOI: 10.1109/comst.2024.3443193
- [5] XU W J, ZHANG Y M, WANG F Y, et al. Semantic communication for the Internet of vehicles: a multiuser cooperative approach [J]. IEEE vehicular technology magazine, 18(1): 100–109. DOI: 10.1109/mvt.2022.3227723
- [6] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436–444
- [7] GETU T M, KADDOUM G, BENNIS M. Semantic communication: a survey on research landscape, challenges, and future directions [J]. Proceedings of the IEEE, 2024, 112(11): 1649–1685. DOI: 10.1109/jproc.2024.3520707
- [8] HUANG Y, HUANG J, LIU Y, et al. Diffusion model-based image editing: a survey [J]. IEEE transactions on pattern analysis and machine intelligence, 2025, 47(6): 4409–4437. DOI: 10.1109/tpami.2025.3541625

- [9] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 10684–10695. DOI: 10.1109/cvpr52688.2022.01042
- [10] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models [EB/OL]. [2025-05-28]. <https://dl.acm.org/doi/abs/10.5555/3495724.3496298>
- [11] SONG Y, DURKAN C, MURRAY I, et al. Maximum likelihood training of score-based diffusion models [EB/OL]. <https://arxiv.org/abs/2101.09258>
- [12] HUANG Y, QIN Z, LIN X, et al. Decoupled diffusion models: simultaneous image to zero and zero to noise [EB/OL]. [2023-06-23][2025-05-23]. <https://arxiv.org/abs/2306.13720>
- [13] BOURTSOULATZE E, BURTH KURKA D, GUNDUZ D. Deep joint source-channel coding for wireless image transmission [J]. IEEE transactions on cognitive communications and networking, 5(3): 567–579. DOI: 10.1109/tccn.2019.2919300
- [14] BIAN C, SHAO Y, GUNDUZ D. DeepJSCC-1++: robust and bandwidth-adaptive wireless image transmission [C]//2023 IEEE Global Communications Conference. IEEE, 2023: 3148–3154

作者简介



承楠, 西安电子科技大学教授; 主要研究方向为6G按需服务关键技术、空天地一体化网络、智能通信网络; 主持或参与国家重点研发计划项目、国家“863”项目、国家自然科学基金重点项目, 以及加拿大自然科学基金委员会重点基金项目等; 获 IEEE GLOBECOM 2014、IEEE ICC 2015、IEEE WCSP 2015、IEEE ICC 2019 最佳会议论文奖; 发表论文80余篇, 获授权美国专利9项。



王秀程, 西安电子科技大学在读博士研究生; 主要研究方向为智能无线网络技术、生成式人工智能。



沈学民, 加拿大滑铁卢大学教授, 中国工程院外籍院士, 加拿大工程院、工程研究院以及皇家科学院院士, IEEE Fellow, 无线通信网络领域谷歌学术引用次数最高的作者之一; 主要研究方向为空天地一体化网络、车联网、网络安全、人工智能等; 荣获44项国际权威期刊及顶级会议的最佳论文奖; 发表论文600余篇, 撰写专著27本。