

# 基于 Transformer 扩散模型的高动态通信信号生成<sup>①</sup>



丁泽全<sup>1</sup>, 魏祥麟<sup>2</sup>, 杨凌升<sup>1</sup>

<sup>1</sup>(南京信息工程大学 电子与信息工程学院, 南京 210044)

<sup>2</sup>(国防科技大学 第六十三研究所, 南京 210007)

通信作者: 魏祥麟, E-mail: [weixianglin@nudt.edu.cn](mailto:weixianglin@nudt.edu.cn)

**摘要:** 通信信号生成是小样本、少样本条件下智能通信系统设计与优化的重要支撑. 针对静态环境, 当前设计的信号生成方法具有较好效果, 但其固定生成机制难以有效捕捉动态场景中快速演变的细微时空特征, 导致生成信号的一致性与准确性下降. 针对该问题, 本文提出了一种基于 Transformer 扩散模型的高动态通信信号生成模型 TransDiffusion, 该模型将 Transformer 架构融入到扩散模型, 通过嵌入多层级解码器模块, 增强对动态特征的长程依赖建模能力, 并优化噪声预测网络以适应时变环境特性. 在自主构建的高动态仿真数据集 (模拟城市交通场景下的多目标运动与射频特征) 上的实验结果表明, 相较于 RF-Diffusion 模型, 所提模型的最大均值差异、均方误差、平均绝对误差分别降低了 85.13%、40.92%、30.62%, 短时傅里叶变换相似性以及功率谱相似性分别提升了 154.30% 和 5.28%, 优于其他几种基线模型.

**关键词:** 扩散模型; Transformer; 生成模型; 时间序列; 通信信号

引用格式: 丁泽全, 魏祥麟, 杨凌升. 基于 Transformer 扩散模型的高动态通信信号生成. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/10108.html>

## Transformer-based Diffusion Model for High-dynamic Communication Signal Generation

DING Ze-Quan<sup>1</sup>, WEI Xiang-Lin<sup>2</sup>, YANG Ling-Sheng<sup>1</sup>

<sup>1</sup>(School of Electronics & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China)

<sup>2</sup>(The 63rd Research Institute, National University of Defense Technology, Nanjing 210007, China)

**Abstract:** Communication signal generation is fundamental for designing and optimizing intelligent communication systems under few-shot conditions. Existing methods perform well in static environments but struggle to capture rapidly evolving spatio-temporal features in dynamic scenarios, leading to reduced consistency and accuracy. To address this, this study proposes a Transformer-based diffusion model, named TransDiffusion, for high-dynamic communication signal generation. TransDiffusion integrates Transformer architecture into the diffusion framework and enhances long-range dependency modeling through embedded multi-level spatio-temporal attention. The noise prediction network is also optimized for time-varying environmental characteristics. Experiments on a custom high-dynamic simulation dataset that emulates multi-target motion and RF signatures in urban traffic show that TransDiffusion significantly outperforms the RF-Diffusion baseline: *MME*, *MSE*, and *MAE* are reduced by 85.13%, 40.92%, and 30.62%, while *STFT-sim* and *Spectral-sim* increase by 154.30% and 5.28%, respectively. This demonstrates the effectiveness of the proposed method in reconstructing high-fidelity communication signals in dynamic scenarios.

**Key words:** diffusion model; Transformer; generative model; time series; communication signal

<sup>①</sup> 基金项目: 国家自然科学基金 (62571539)

收稿时间: 2025-09-06; 修改时间: 2025-10-10; 采用时间: 2025-10-29; csa 在线出版时间: 2026-01-15

随着无线通信向高频段、高动态场景加速扩展,数据驱动的智能通信系统设计与优化逐渐成为通信领域研究的新范式.其中,通信信号生成技术是支撑小样本、少样本<sup>[1]</sup>甚至零样本<sup>[2]</sup>场景下信号智能设计、识别与设计的重要基础.零样本场景下,新型通信协议验证缺乏实测数据,需依赖物理规律生成可信信号;少样本和小样本场景下,需通过数据增强提升模型泛化性.虽然AR模型(autoregressive model)<sup>[3]</sup>、GAN(generative adversarial network)<sup>[4]</sup>等生成方法可以利用大规模标注数据合成静态信号,却难以应对动态时变特性和物理规律约束的联合优化需求,导致生成信号与真实环境存在较大偏差.此外,这些方法依赖大规模标注数据训练,在数据稀缺条件下易出现过拟合或模式崩塌,导致生成信号与真实物理规律(比如电磁传播特性和多径效应)的一致性显著下降.近年来,扩散模型凭借其渐进式去噪机制与隐式概率建模优势,在通信信号生成中取得突破<sup>[5]</sup>,但其在高动态通信信号生成中的应用仍面临一些挑战.

现有基于扩散模型的通信信号生成方法,如RF-diffusion<sup>[6]</sup>、DiffWave<sup>[7]</sup>等,在静态信号如固定频段调制波形生成中表现良好,但是难以适应动态时变场景中信号的瞬时波动性与多尺度时空关联性.动态环境下的通信信号兼具强非平稳性和跨域耦合特征,要求生成模型能够同时捕捉局部快速变化和全局长程依赖.主流扩散框架采用的U-Net<sup>[8]</sup>或RNN<sup>[9]</sup>结构受限于局部卷积和循环操作的归纳偏置,对高频瞬态特征的捕捉仍存在不足,难以捕捉信号的瞬时波动性与多尺度时空关联性.

针对这些问题,本文设计一种基于Transformer扩散模型的高动态通信信号生成模型TransDiffusion,整体结构由卷积多层感知机与可学习位置编码等输入层模块开始,通过堆叠的编码器和多种注意力机制<sup>[10]</sup>以及金字塔池化进行多粒度的特征提取,再通过深度解码器完成时序成分的细粒度生成与重构.模型还在核心结构中引入了自适应归一化等动态调整模块与局部卷积操作,从而在捕捉局部突变的同时兼顾全局趋势,实现对高动态通信信号的精准建模<sup>[11]</sup>.本文主要工作和贡献包括以下3个方面.

(1) 问题提出建模.针对高动态通信信号生成场景,通过实验验证分析了传统生成方法与现有扩散模型无法捕捉信号的高频瞬态特征,并对高动态通信信号生

成问题进行了建模.

(2) 设计了基于Transformer的高动态通信信号生成模型,模型包含编码器、解码器、多尺度映射与重构等模块,在编码器(encoder)端编码时域波形和频域谱,并通过可学习位置编码实现跨模态对齐;金字塔池化与跨尺度注意力融合多分辨率特征,兼顾短时突变与长期趋势;局部和全局注意力结合卷积MLP,协同建模高频抖动与低频漂移;解码器(decoder)反向生成逐步去噪,重构时域与频域信号以保证连贯性和频谱保真度.

(3) 将TransDiffusion模型在自主构建的高动态仿真数据集上进行实验验证,实验结果显示,改进后的模型取得了优异的精度提升,相较于Denoising diffusion probabilistic models(DDPM)<sup>[12]</sup>、WaveGAN<sup>[13]</sup>、SwinLSTM<sup>[14]</sup>等基线模型,在平均绝对误差(MAE)和均方误差(MSE)等评价指标上均实现显著提升.

本文第1节介绍扩散模型在通信信号生成方面的相关进展.第2节对高动态通信信号问题进行陈述.第3节介绍算法的具体框架和模型结构.第4节介绍实验参数设置并对实验结果进行分析.第5节总结全文并对未来的工作进行展望.

## 1 相关工作

通信信号生成作为序列建模与预测中的关键任务之一,旨在从已有的通信信号序列数据中学习潜在的时序分布并生成与真实序列具有相似统计特性的新序列.与传统的自回归生成模型<sup>[15]</sup>相比,扩散模型通过正向与反向扩散过程在潜在空间进行建模,能够更灵活地捕捉序列内部的复杂非线性依赖关系,常见的扩散模型主要可以分为基于离散时间步和基于连续时间步的扩散方法.基于离散时间步的方法通常采用逐步腐蚀数据并学习反向还原的思路,如DDPM<sup>[12]</sup>、SGM<sup>[16]</sup>等;而基于连续时间步的方法则基于随机微分方程(SDE)<sup>[17]</sup>来构造噪声过程,通过积分求解的方式在扩散和反向扩散之间进行平滑转换,典型代表包括Score-based SDE、Variance exploding<sup>[18]</sup>、Variance preserving SDE<sup>[19]</sup>等.

在通信信号时间序列数据生成方面,早期研究更多集中于利用自回归或循环神经网络(RNN)来建模时序依赖,但对于高动态通信信号或多变量场景往往效果有限.近年来,随着扩散模型在语音合成、时间序列

插值、异常检测等任务中的成功应用,采用扩散式思路对通信信号时间序列进行建模成为一种新的趋势.例如, WaveGrad<sup>[20]</sup>与 DiffWave<sup>[7]</sup>等方法在语音生成与增强任务上表现优异,通过多层级噪声注入的方式使模型更容易学习语音信号的复杂细节; TimeGrad<sup>[21]</sup>与 CDCD<sup>[22]</sup>等方法则进一步将扩散模型拓展到通用的通信信号预测或生成领域,通过引入自回归框架或条件网络实现多步预测以及对非平稳时序的建模; CSDI<sup>[23]</sup>等研究也探索了在有缺失的数据场景中,通过扩散模

型实现对不完整通信信号的修复与插值.与此同时, RF-Diffusion 提出了一种针对无线通信信号特性的扩散模型架构,通过引入射频信号动态特征提取模块与自适应噪声调度策略,显著提升了在复杂信道环境下信号生成的时域连续性与频谱保真度,相比于传统生成对抗网络 (GAN)<sup>[24]</sup>或变分自编码器 (VAE)<sup>[25]</sup>等生成方法,扩散模型在训练的稳定性与生成的多样性方面具有一定优势,同时能够在高维分布中更精细地控制生成过程.具体对比如表 1 所示.

表 1 不同生成模型在通信信号生成中的性能对比

模型类型	生成质量	训练稳定性	计算效率	多尺度建模能力
扩散模型 (Transformer结构)	更高的时频一致性	更稳健	需多步迭代	强 (自注意力可捕获全局多尺度)
扩散模型 (U-Net结构)	时频保真	无模式坍塌	需多步迭代	强 (多级噪声输入)
GAN	易出现伪影	模式坍塌	单步生成	弱 (依赖对抗训练)
VAE	生成模糊	较高 (ELBO优化)	单步解码	中 (潜在空间约束)
自回归模型	误差累积	高 (精确密度估计)	复杂设计	弱 (结构限制)

尽管上述扩散模型在通信信号生成和预测中取得了显著进展,但在应对高动态、强时间相关性的通信信号时序数据时仍面临一些挑战.一方面,通信信号往往同时包含短期的快速突变与长期的趋势漂移等多重尺度变化,需要更具针对性的多尺度结构或注意力机制来捕捉;另一方面,现有方法在对不同时间步或特征维度的信息进行灵活建模时仍存在不足,难以兼顾整体趋势和局部突变等异质特征.此外,在面对动态性较强的通信信号时序数据时,扩散模型的生成精度和稳定性也有待进一步提升.

的瞬态突变特征使得传统的时频分析方法难以同时捕捉快速变化的瞬时频率与精细的频谱结构,正是这种多尺度耦合与瞬态尖峰,使得传统模型往往会平滑掉关键信息或难以同步建模整体趋势和局部激变,因此提取高动态通信信号特征对模型的频域精度和跨步依赖捕捉能力提出了更高要求.

## 2 问题建模与分析

### 2.1 问题建模

高动态信号与静态信号相比,具有更剧烈的时变特性、更弱的自相关性和更强的多尺度耦合现象.静态信号通常存在于传播环境稳定、收发端相对静止的场景,其信道参数在观测时间内保持相对恒定,莱斯因子较高且路径损耗变化缓慢.如图 1 所示为某静态通信信号,可以看出静态信号呈现出小幅、平稳的振荡特性,曲线相对平滑且周期性较为明显,相邻时间点之间具有强自相关性,这种稳定性使其易于通过常规傅里叶变换或线性滤波方法进行特征提取与分析.图 2 所示高动态信号则展现出强烈的时变特性,相邻采样点间的相关性显著降低,且伴随非线性瞬态振荡现象,整体表现出明显的非平稳性与多尺度耦合.此类信号

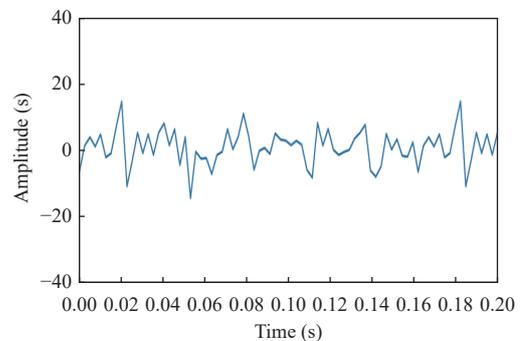


图 1 静态通信信号

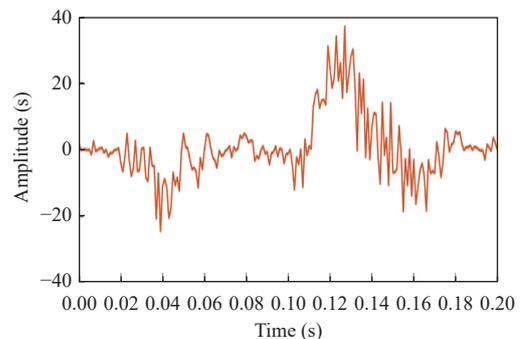


图 2 高动态通信信号

将通信信号对应的时间序列数据记为:

$$X_{1:\tau} = (x_1, \dots, x_\tau) \in \mathbf{R}^{\tau \times d} \quad (1)$$

其中,  $\tau$ 表示时间步长,  $d$ 为观测信号的维度. 给定包含  $N$ 个时间序列样本的数据集  $D_A = \{X_{1:\tau}^i\}_{i=1}^N$ , 信号生成的目标是通过生成函数  $G$ 将输入的高斯随机向量转换为高动态通信信号, 即:

$$\hat{X}_{1:\tau}^i = G(Z^i) \quad (2)$$

其中,  $Z^i = (z_1^i, \dots, z_T^i) \in \mathbf{R}^{\tau \times d \times T}$  为输入的高斯随机向量. 转换为与  $D_A$  中信号高度相似的输出, 其中  $T$  为总扩散步数. 根据式 (1) 和式 (2), 可以将通信信号生成问题转化为以下优化目标:

$$\min_G \frac{1}{N} \sum_{i=1}^N L(X_{1:\tau}^i, G(Z^i)) \quad (3)$$

其中,  $G$ 表示生成函数, 其作用是将输入的高斯随机向量  $Z^i = (z_1^i, \dots, z_T^i) \in \mathbf{R}^{\tau \times d \times T}$  转换为生成信号  $X_{1:\tau}^i$ ,  $T$  为总扩散步数,  $L$  为损失函数, 用于衡量生成信号与真实信号之间的差异. 这个优化目标旨在通过调整生成函数  $G$  的参数, 使生成的信号  $G(Z^i)$  与数据集  $D_A = \{X_{1:\tau}^i\}_{i=1}^N$  中的真实信号  $X_{1:\tau}^i$  尽可能相似. 通过最小化损失函数  $L$  可以使生成函数  $G$  学习到真实信号数据的分布特征, 从而生成与真实信号高度相似的高动态通信信号.

## 2.2 现有方法分析

在处理高动态时间序列数据时, 现有生成模型在精度上仍然存在一定的不足. 高动态数据通常表现为快速变化的局部特征和长期的趋势漂移, 传统的自回归模型和循环神经网络 (RNN) 很难在短时间窗口内捕捉到这些局部快速变化, 从而导致生成结果与真实序列之间的显著偏差. 尽管近年来扩散模型在时间序列预测与生成领域取得了良好进展, 但对于此类高动态场景, 其生成结果仍存在明显的偏差和不稳定性.

为了验证模型的检测性能, 图 3(a) 展示了高动态序列上的扩散模型的方法的生成结果对比. 在该场景下, 行人和车辆的速度都设置为随机, 行人的速度在 1–1.9 m/s, 车辆的行驶速度范围在 40–160 km/h, 其中蓝色为真实序列, 橙色为生成序列. 可以看出, 在局部剧烈变化或整体趋势发生较大波动时, 方法出现了与真实序列不匹配的情况, 模型往往难以同时兼顾短时刻的精准刻画与全程的平稳性. 图 3(b) 则是在行人速度为 0 m/s、车辆速度固定为 10 km/h 的场景下的序列

生成结果对比. 从整体上看, 蓝色的真实序列在该场景下振幅和波动都相对平稳, 橙色的生成序列整体幅值变化也不大且起伏趋于平滑, 说明模型在动态场景下的表现更不稳定, 幅度波动大; 而在静态场景下, 幅度逐渐下降并趋于平稳, 表现相对稳定.

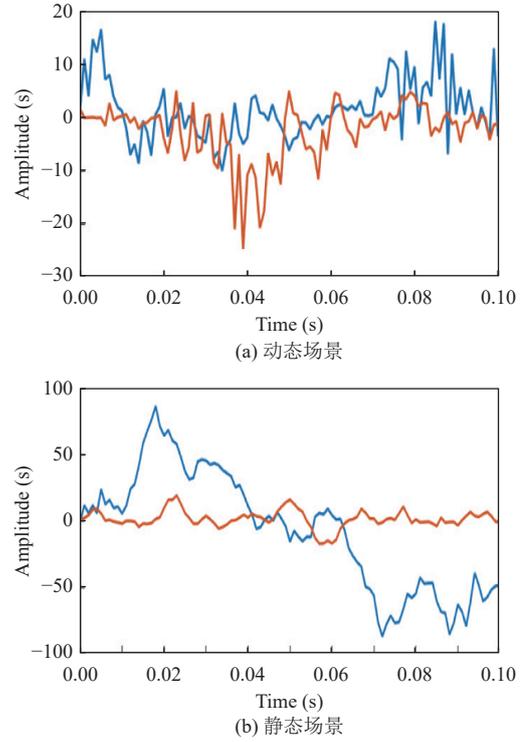


图 3 RF-Diffusion 生成 IQ 信号对比

图 4 所示的 5 项评价指标 (第 4.2 节中详细介绍) 揭示了 RF-Diffusion 在不同层面上的性能表现. 在动态时间序列上, 首先,  $MMD$  值高达 0.9652, 表明生成信号与真实信号在分布空间上仍存在明显的偏差; 其次,  $MSE=0.3245$  和  $MAE=0.5056$  说明模型在时域上能相对精准地重构幅度变化, 但仍有一定误差; 再次,  $STFT-sim$  仅为 0.3033, 暗示短时频谱结构未能得到有效保留; 而频谱相似度 0.4922 则反映出全局频谱特性虽有一定契合, 但仍存在不足. 而在静态时间序列上, 由于信号相对平稳, 模型在各项评价指标上都优于动态序列, 这主要源于静态信号本身的平稳性. 首先,  $MMD$  值从动态场景的 0.9652 降至 0.4442, 表明了生成样本在整体分布空间中与真实数据高度重合; 其次, 在时域误差方面,  $MSE$  从 0.3245 降至 0.1902、 $MAE$  从 0.5056 降至约 0.4609, 这意味着静态序列的振幅波动平滑均匀, 模型无需应对突发跳变或多目标叠加, 因而能更加

精准地重现每一时刻的样本值; *STFT-sim* 则从仅有的 0.3033 飙升至接近 0.9535, 说明模型几乎完美保留了局部时窗内的频谱结构; 功率谱相似性也从动态场景的 0.4922 提升至 0.9231, 反映出整体能量分布和频谱峰值位置都被准确捕获。

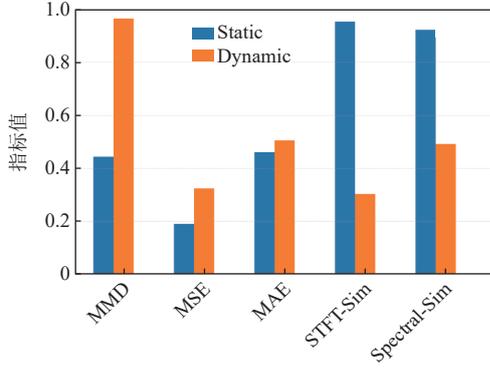


图4 模型评估性能

因此, 对于静态通信信号的合成, 模型能够凭借其低维、易学习的特征分布, 在分布拟合、时域重构和频域保真等多方面具有卓越表现; 但在动态场景中, 由于多普勒抖动、遮挡突变和多目标叠加引入了丰富的非平稳特征, 模型在分布匹配、时域重构和频谱保留等方面都表现欠佳。

### 2.3 DDPM

针对上述问题, 本文构建基于扩散模型的数学框架, 旨在实现动态特征的捕捉与时序依赖的建模, 如图5所示。首先介绍包含了两个核心过程的扩散模型: 前向过程和反向过程。在前向过程中, 数据分布中的样本  $x_0 \sim q(x)$  通过逐步加噪被转化为标准高斯噪声  $x_T \sim N(0, I)$ , 其状态转移由参数化的马尔可夫链定义:

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \quad (4)$$

其中,  $\beta_1, \dots, \beta_t, \dots, \beta_T$  是预定义的噪声调度参数。根据高斯分布的性质, 可以推导出:

$$q(x_t|x_0) = N(x_t; \sqrt{\bar{\alpha}_t}x_0(1-\bar{\alpha}_t), I) \quad (5)$$

其中,  $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ ,  $\alpha_t = 1-\beta_t$ 。通过采样  $\varepsilon \sim N(0, I)$  并使用重参数化技巧, 可以得到样本  $x_t = \sqrt{\bar{\alpha}_t}x_0 + (\sqrt{1-\bar{\alpha}_t})\varepsilon$ 。在一定条件下, 最终步骤的分布  $q(x_T)$  近似为标准高斯分布。

逆向生成过程是从纯噪声样本  $x_T$  开始, 逐步去噪重建  $x_{T-1}, x_{T-2}, \dots, x_0$ , 最终得到逼真的样本。逆向过程

被定义为条件概率分布  $p_\theta(x_{t-1}|x_t)$ , 通过 1 个神经网络进行学习, 以此来近似  $q(x_{t-1}|x_0)$  这一目标分布。为了学习  $p_\theta(x_{t-1}|x_t)$ , 只需训练模型输出  $\mu_\theta(x_t, t)$  来恢复生成  $x_t$  时所加入的噪声。

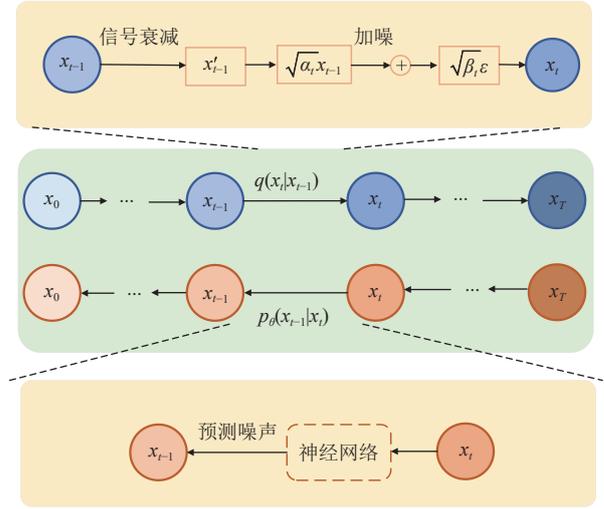


图5 扩散模型流程图

通过反向扩散过程对噪声  $x_T$  进行去噪等价于训练 1 个参数化代理近似器, 以估计所有时间步  $t$  下的均值  $\mu_\theta(x_t, t)$ , 采用加权均方误差损失训练该去噪模型:

$$L_{\text{MSE}}(x_0) = \sum_{t=1}^T E_{q(x_t|x_0)} \|\mu(x_t, x_0) - \mu_\theta(x_t, t)\|^2 \quad (6)$$

其中,  $\mu(x_t, x_0)$  是后验分布  $q(x_{t-1}|x_0, x_t)$  的均值。  $\mu_\theta(x_t, t)$  是神经网络预测的噪声, 训练目标是通过最小化预测噪声与实际噪声之间的差异来学习逆向扩散过程从而生成高相似度的样本。

### 2.4 核心思想

在本节中提出了本文方法的核心思想, 并解释了 TransDiffusion 如何优化噪声预测网络以适应时变环境。模型采用了编码器-解码器 Transformer 架构, 以增强其在时间序列数据中捕获全局相关性和模式的能力。这确保了在解码之前, 整个含噪序列都会被编码。本文在解码器部分新增了两个模块: 多尺度映射与重构模块 (multi-scale projector module) 和周期注意力模块 (periodic attention module)。解码器采用多层结构, 每个解码器包含 1 个 Transformer 块、1 个前馈神经网络以及上述 2 个新增模块。整体结构的详细描述见第 3.5 节。此处重点介绍这两个模块。通过对不同组件施加不同形式的约束, 并为各组件引入不同的归纳偏置, 使其能够

更容易学习特定的语义知识。

#### 2.4.1 捕捉随时间缓慢变化的趋势

多尺度映射与重构模块负责实现这一功能,旨在对缓慢变化的行为进行建模。为了生成合理的趋势分量,采用多项式回归器对趋势 $V^{\text{tr}}$ ,进行建模:

$$V_i^{\text{tr}} = \sum_{i=1}^D (C \cdot \text{Linear}(w_{i,t}^{\text{tr}}) + X_{i,t}^{\text{tr}}), C = [1, c, \dots, c^p] \quad (7)$$

其中,  $X_{i,t}^{\text{tr}}$  表示第  $i$  个解码器块的平均输出, “ $\cdot$ ” 表示张量乘法。缓慢变化的多项式空间  $C$  是向量  $c = [0, 1, 2, \dots, \tau - 1]^T / \tau$  的幂矩阵,  $p$  为较小的多项式阶数 (如  $p = 3$ ), 用于捕捉低频行为。

#### 2.4.2 捕捉时间序列中的周期性趋势

周期注意力模块用于捕捉时间序列中的周期性趋势和误差。我们在该模块中采用傅里叶基来建模时间序列的周期成分:

$$A_{i,t}(k) = F(w_{i,t}^{\text{per}})_k, \Phi_{i,t}(k) = \varphi F(w_{i,t}^{\text{per}})_k \quad (8)$$

为了提取信号中的主导周期模式,在频域中筛选出能量最强的  $K$  个频率分量。令  $\Omega = \{1, \dots, [\pi/2] + 1\}$  为有效频率索引集合。定义  $K_{i,t}$  为  $\Omega$  的一个子集, 该子集包含  $K$  个使得对应幅值之和最大的频率索引, 即:

$$K_{i,t} = \underset{k \subset \Omega, |S|=K}{\text{argmax}} \sum_{k \in S} A_{i,t}(k) \quad (9)$$

其中,  $|S|$  表示集合  $S$  的基数。通过该操作, 模型仅保留了对信号重构贡献最大的  $K$  个主频率。

$$P_{i,t} = \sum_{k=1}^K A_{k_{i,t}(k), i,t} \cos\left(\frac{2\pi f_{k_{i,t}(k)} C}{\tau} + \Phi_{k_{i,t}(k), i,t}\right) + \cos\left(\frac{2\pi \bar{f}_{k_{i,t}(k)} C}{\tau} + \bar{\Phi}_{k_{i,t}(k), i,t}\right) \quad (10)$$

周期注意力模块选择频域中幅值最大的基函数, 然后通过逆变换回到时域, 从而建模周期性趋势。最终可以得到重构信号:

$$\hat{x}^0(x_t, t, \theta) = V_t^{\text{tr}} + \sum_{i=1}^D P_{i,t} + R \quad (11)$$

其中,  $R$  表示最后一个解码器块的输出, 代表残余周期性和其他噪声。

#### 2.4.3 噪声预测网络的优化目标

扩散模型的基本优化目标如下: 在扩散框架中, 原始数据  $x_0$  被逐步扰动为高斯噪声  $x_T \sim N(0, I)$ 。网络的任务是学习反向过程  $p_{\theta}(x_{t-1}|x_t)$ , 即逐步对样本去噪。优化的目标是 minimize 预测均值  $\mu_{\theta}(x_t, t)$  与真实后验均值  $\mu(x_t, x_0)$  之间的差异。

由于我们在解码器中引入了两个额外的模块来分别捕捉时间序列中缓慢变化的趋势和周期模式, 噪声预测网络不仅输出对  $x_0$  的整体估计, 还将其分解为慢变趋势和周期趋势。这种设计增强了模型的稳定性。在训练过程中, 网络通过反向扩散过程不断迭代优化, 从而能够在每个时间步  $t$  预测去噪目标。

具体而言, 多尺度投影模块采用多项式基约束, 使其更容易捕捉低频和缓慢变化的信号; 同时, 周期注意力模块利用傅里叶基选择最显著的频率成分, 使噪声预测网络能够更有效地建模全局周期性。通过这种方式, 噪声预测网络被赋予了明确的归纳偏置和有针对性的优化目标, 从而实现更快的收敛速度和更高的稳定性。

## 3 TransDiffusion

### 3.1 TransDiffusion 模型设计

TransDiffusion 模型基于 Transformer 架构与扩散模型的融合, 旨在处理高动态通信信号的生成任务。整体结构如图 6 所示, 包括噪声注入过程、编码器 (encoder)、解码器 (decoder) 和噪声预测网络。输入信号首先通过噪声注入过程逐步添加高斯噪声, 形成噪声序列  $X_T$ , 然后从  $X_T$  逆向去噪生成输出信号。编码器负责提取时空特征, 解码器则通过多层次模块进行细粒度重构, 噪声预测网络优化时变噪声估计。TransDiffusion 模型的整体结构由 2 层编码器与 4 层解码器组成, 总共 6 层网络框架。模型输入序列  $X$  通过线性映射后进入隐层维度 32 的特征空间, 并在每一层中采用 2 个注意力头以平衡特征捕获能力与计算复杂度。每个解码块接收来自编码器的上下文特征以及扩散时间步嵌入并通过自注意力和交叉注意力机制进行特征交互。之后, 特征经由  $1 \times 1$  卷积进行通道投影并分为两路, 分别输入至多尺度映射模块和周期注意力模块。前者利用不同尺度的卷积核在时域上提取多层次的局部变化模式, 以增强对信号包络与趋势的感知; 后者在频域上通过稀疏傅里叶变换聚焦主要能量分量, 实现对信号周期性结构的建模。两路输出经门控加权融合后, 经过残差连接回到主干通道, 以保证稳定训练与信息保真。扩散过程的时间步数设置为 500, 采样步数为 50, 整个模型总参数量约为 8.549M。这样的配置在保证建模能力的同时兼顾计算效率, 能够在复杂无线信号生成任务中有效捕获时间-频率双域特征。

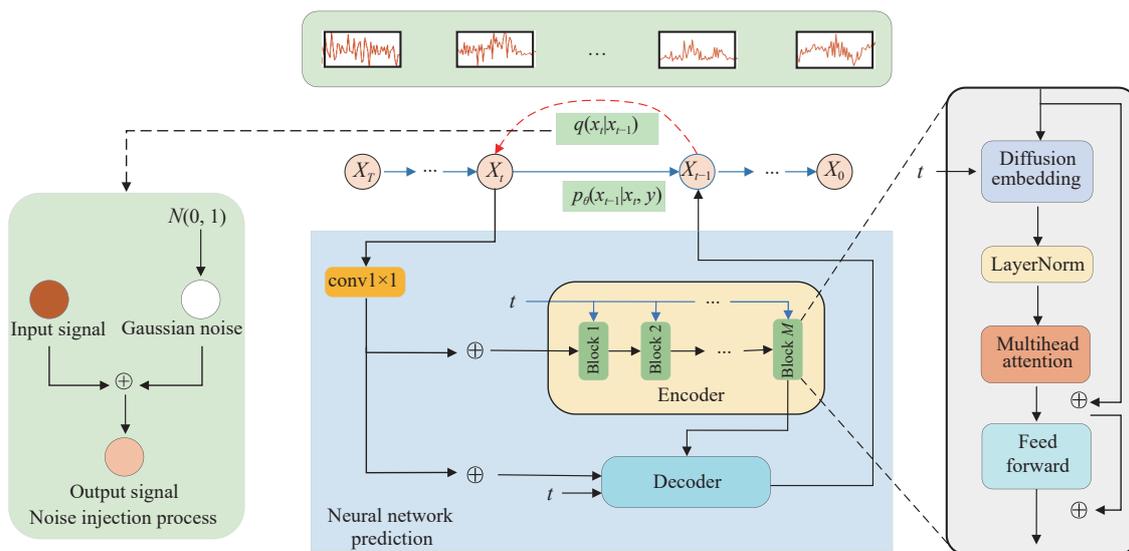


图6 基于Transformer的扩散模型生成网络整体架构

### 3.2 编码器

编码器逐层提取时间序列的多尺度特征表示. 第1层编码器对原始序列进行卷积嵌入, 将输入IQ信号映射到高维特征空间, 并加上可学习位置编码 (learnable positional encoding) 以注入时间位置信息. 随后, 编码器由多个层级的编码块. 在每个编码块中, 首先通过自适应层归一化 (AdaLayerNorm), 动态融合扩散步嵌入条件, 生成缩放因子与平移参数, 调节特征分布以适应不同噪声水平. 公式表达为:

$$Output = LayerNorm(x)e^{(1 + \gamma(t))} + \beta(t) \quad (12)$$

其中,  $\gamma(t)$ 和 $\beta(t)$ 由扩散步条件映射得到. 其尺度和平移由当前扩散步  $t$  的嵌入经过 MLP 映射得到. 这样, 网络层的分布在不同噪声水平下可调整, 确保模型在扩散过程各阶段均表现良好. 随后通过全局多头自注意力 (full attention) 层, 采用16头注意力机制, 计算所有时间步间的依赖权重, 捕获长程时序关联. 注意力矩阵通过缩放点积计算, 并应用 dropout (概率0.1) 防止过拟合. 在通过全局多头注意力后经过前馈神经网络 (MLP), 其包含两层线性变换与 GELU 激活函数, 隐层维度扩展4倍, 通过残差连接保留原始特征信息. 编码块末尾包含残差连接, 将输入特征与卷积和注意力变换后的特征相加, 提高梯度传递稳定性. 编码块结构如图7所示.

经过2层处理后, 输出包含多尺度语义的上下文特征矩阵. 这样的深层编码块逐步抽象特征, 浅层捕获

局部波动, 深层建模全局趋势, 如此迭代, 编码器获得一系列从细到粗的特征表示, 传递至解码器引导生成过程.

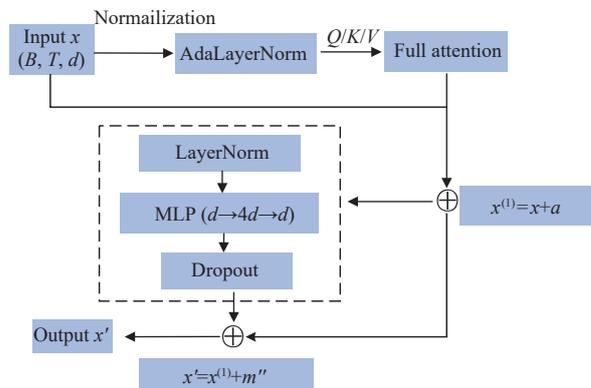


图7 编码块结构

### 3.3 多尺度映射与重构模块 (multi-scale projector module)

为了增强模型对非平稳趋势成分的建模能力, 本文设计了多尺度映射与重构模块, 专门处理输入序列中的长期趋势与局部结构变化, 如图8所示. 该模块以通道维度的序列片段为输入, 通过一系列多阶多尺度的线性与非线性映射, 提取不同时间尺度上的趋势特征. 具体而言, 输入序列首先通过多个并行的1D卷积核实现局部子序列的特征捕获, 随后将这些多尺度特征融合并通过残差重构路径回传至主干网络. 相比传统的多项式回归方法, multi-scale projector 模块能更灵

活地适配非线性趋势和突变点, 具备更强的泛化能力与局部建模能力, 有效提升序列重建精度.

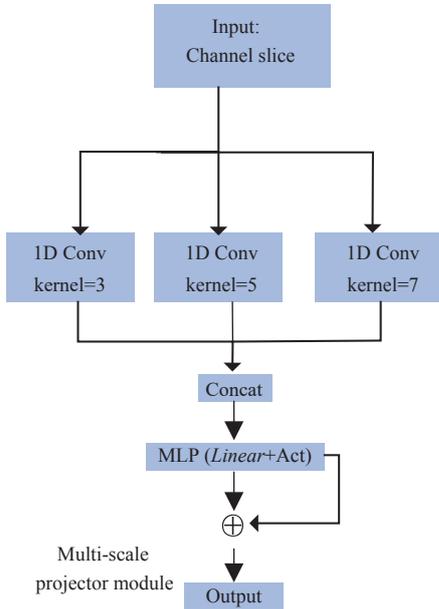


图 8 多尺度映射与重构模块

### 3.4 周期注意力模块 (periodic attention module)

周期性是许多时间序列中的关键结构特征, 传统方法常依赖傅里叶分析进行建模. 为此, 本文提出了周期注意力模块, 将频域建模与注意力机制结合, 用以提取并强化主导周期模式, 如图 9 所示. 该模块首先对输入序列执行快速傅里叶变换 (FFT), 在频域上获取复数谱图, 通过幅度排序筛选前  $K$  个显著频率分量. 然后, 利用反傅里叶变换将选频分量重构为时域周期信号, 并引入门控机制 (gating function) 对不同频段进行加权组合, 从而强化有效频段、抑制噪声干扰. 该模块可以适配不同周期结构的组合模式, 尤其适用于多重周期和非平稳周期的建模任务.

### 3.5 解码器

在本文提出的解码器结构中, 我们设计了 4 层堆叠的解码器块, 以充分挖掘和重建时间序列中的多尺度时序特征. 每层解码器块包含 5 个主要子模块, 按如下顺序依次处理输入序列: 首先进行自注意力子层 (self-attention layer), 结构与编码器一致, 采用自适应归一化层 (AdaLayerNorm) 加上全局注意力和残差连接的方式建模解码器内部的上下文依赖; 其次是交叉注意力子层 (cross-attention layer), 通过自适应归一化层和交叉注意力以及残差连接的结构, 实现对编码器信息的有效条件建模.

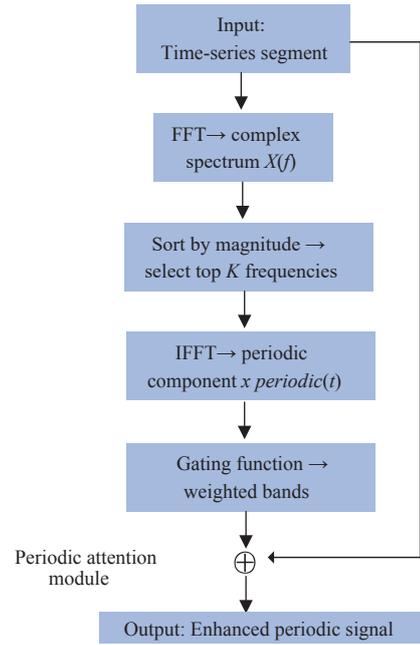


图 9 周期注意力模块

随后, 引入 1 个投影拆分模块, 对当前隐藏状态通过  $1 \times 1$  卷积进行通道映射与分解, 得到两个平行通道:

$$(x_1, x_2) \in \mathbb{R}^{B \times C} \times T \times \mathbb{R}^{B \times C \times T} \quad (13)$$

这两个平行通道分别用于后续的整体建模与局部建模. 我们在此基础上提出了两类专门设计的时序建模模块: 多尺度映射与重构模块用于建模输入序列的长期趋势与局部结构, 通过多阶多尺度映射与残差重构操作, 增强了局部时序模式的表达能力; 而周期注意力模块则在频域捕捉潜在的周期模式, 利用稀疏傅里叶变换选频、反投影与门控机制, 提取出对序列影响显著的主导周期成分. 这两个模块分别作用于  $x_1$  和  $x_2$ , 以建模趋势变化与周期波动等不同类型的时序动态. 解码块结构如图 10 所示.

在解码器块中, 多尺度映射与重构模块和周期注意力模块的协同通过并行分支处理和融合机制实现. 具体而言, 从交叉注意力子层后的隐藏状态经投影拆分模块 ( $1 \times 1$  卷积) 分解为两个平行通道  $x^{(1)}$  和  $x^{(2)}$  (如式 (12) 所示). 其中,  $x^{(1)}$  输入多尺度投影模块, 采用并行多核卷积 (核大小 3、5、7) 捕捉长期趋势与局部结构, 输出  $x_1$ , 增强对高动态信号的非平稳趋势建模;  $x^{(2)}$  输入周期注意力模块, 通过 FFT 选频和门控逆变换提取主导周期成分, 输出  $x_2$ , 聚焦频域谐波特征. 如图 10 所示, 两模块输出  $x_1$  和  $x_2$  通过残差加法融合:  $x' = x_1 + x_2$ , 随后经 MeanPool+Linear 处理, 进一步整合为

后续 cross attention 的输入. 该融合确保趋势 (时域多尺度) 和周期 (频域波动) 的互补, 避免单一模块的偏差. 从物理角度, 该机制对应高动态通信信号的耦合特

性: 多尺度模块处理运动轨迹诱发的非线性衰减 (如车辆多径效应), 周期模块捕捉射频频干扰 (如多普勒频移抖动), 协同提升生成信号的时频保真度和一致性.

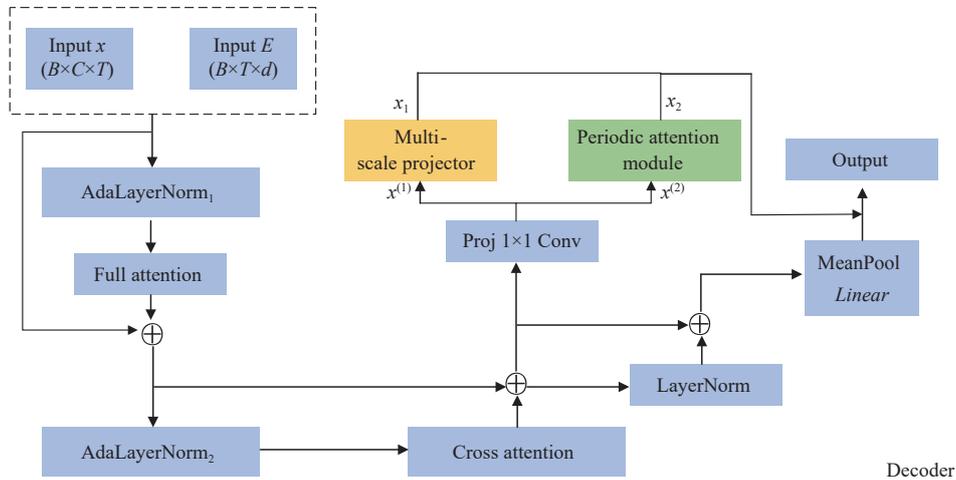


图 10 解码块结构示意图

这种设计通过在编码器和解码器结构中同时引入自适应噪声条件归一化、多尺度趋势映射与频域周期注意力, 在保留短时局部突变与长时全局依赖的同时, 又能够灵活捕捉主导周期性成分和非平稳突变, 使得模型在面向高动态、强时序相关的复杂信号生成与重建时, 既能适应快速变化的局部波动, 也能维护整体的时频一致性, 从而实现了对高动态信号的精确重构与生成.

#### 4 实验设计与结果分析

本节首先对 TransDiffusion 的实验结果进行分析. 随后, 开展对比实验和消融实验以验证其有效性. 最后, 通过参数敏感性实验研究动态信号对 TransDiffusion 性能的影响.

##### 4.1 数据集构建

为了验证模型的性能, 本文基于 Matlab 构建高动态仿真数据集, 以模拟城市交通场景下的多目标运动与射频特征的场景. 该数据集主要通过 3 个核心要素来构建动态场景. 首先是多目标运动建模: 包含 20 个行人 (随机游走模型, 速度 1.1–1.9 m/s) 和 8 辆车辆 (车道约束运动, 速度 40–160 km/h), 接收机固定于 (150 m, 0 m); 其次是物理场耦合: 计算 5.9 GHz 频段多普勒频移 (行人/车辆独立建模, 频移范围 $\pm 300$  Hz), 并引入动态遮挡效应 (车辆触发 -6 dB 衰减, 行人梯度衰减); 最后是信号动态性: 合成包含非平稳噪声 (引擎锯齿波

调制加语音正弦干扰) 的时变 IQ 信号, 采样率 1 000 Hz 持续 30 s. 具体数据集信息如表 2 所示.

表 2 数据集信息

场景配置	参数说明	参数值
信号接收环境	仿真总时长	30 s
	采样频率	1 000 Hz
	时空分辨率	1 ms
	接收机坐标	(150 m, 0 m) 道路中央
行人动态模型	运动个体数量	20个随机行人
	移动速度范围	1.1–1.9 m/s
	活动区域	200×200 m <sup>2</sup> 自由空间
车辆动态模型	运动模式	随机游走+边界反射
	运动车辆数量	8辆多车道车辆
	行驶速度范围	40–160 km/h
	道路配置	双向6车道 (宽度 15 m)
信号生成参数	运动模式	固定车道匀速运动
	载波中心频率	5.9 GHz (C-V2X标准频段)
	基带信号生成	AR(1)模型 ( $\alpha=0.98$ )
	噪声类型	I路: 引擎调制噪声; Q路: 人声干扰噪声
环境衰落模型	相位扰动	高斯白噪声
	车辆遮挡衰减	50 m范围内信号衰减75%
	行人遮挡衰减	20 m内渐变衰减
数据存储规格	基础路径损耗	固定衰减因子0.8
	辅助存储格式	CSV双通道
	数据维度	30k×2时间序列
	动态范围	多普勒频移

本研究数据集通过时域波形与动态轨迹两类图表呈现多维运动特征. 图 11 (幅度时域波形图) 为两个通

信号通道的幅度随时间变化的时域波形,其中蓝色和橙色曲线分别代表 I、Q 两路基带信号,时间范围为 0–30 s,纵轴为信号幅度.该时域波形具有宽带高动态范围特性,幅度随时间呈现出强烈的起伏变化,显示出典型的高动态通信信号特征.图 12 (动态场景图)直观展示了动态场景中发射机、接收机、车辆路径与行人轨迹的空间分布格局.其中,车辆在 Y 轴多个固定通道内沿 X 轴平稳前进,展现出典型的线性运动特性.相比之下,行人路径则显著聚集于 X 轴在 0–200 m 间的区域,并在空间上呈现出规则、非线性的扩散轨迹,部分路径存在短时大幅横向跳变,反映出行人行为的高

度随机性与不可预测性.

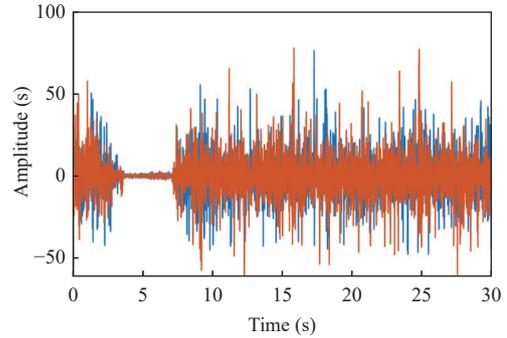


图 11 幅度时域波形图

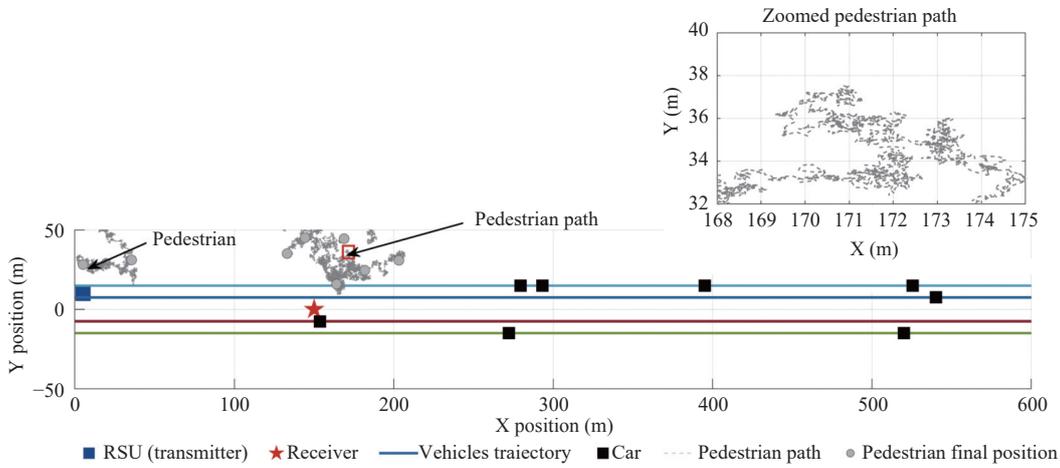


图 12 动态场景图

#### 4.2 评估指标

为全面衡量模型生成信号的质量,本文引入了 5 种性能评估指标:最大均值差异 (*MMD*)、均方误差 (*MSE*)、平均绝对误差 (*MAE*)、短时傅里叶变换相似性 (*STFT-sim*) 和功率谱相似性 (*Spectral-sim*).这些指标分别从分布匹配、数值偏差、时频保真度和频域能量分布这 4 个维度考察生成信号与真实信号之间的一致性.

首先,在信号分布匹配方面,最大均值差异通过核函数将信号映射到再生核希尔伯特空间,衡量生成信号与原始信号的整体分布差异,捕捉二者统计特性的全局偏离,避免局部样本对比局限;其次,在数值偏差方面,均方误差量化时域波形逐点平方误差,对幅度突变偏差尤为敏感,平均绝对误差衡量绝对误差平均值,对异常值更具鲁棒性,反映生成信号全局保真度;最后,在时频保真度与频域能量分布方面,短时傅里叶变换相似性通过归一化短时傅里叶能量相关性评估生成信

号局部时频动态特性匹配程度,功率谱相似性通过归一化功率谱相关性检验生成信号是否保留真实信号物理传播规律.这 5 项指标构成全面互补的评估框架,可系统验证扩散模型在静态和高动态通信信号生成中的性能差异和优势.具体指标定义如式 (14)–式 (18):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (14)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

其中,  $\hat{y}_i$  为预测的值,  $y_i$  为真实值,  $S_t$  为第  $t$  个时间窗的 STFT 矩阵.

$$STFT-sim = \frac{1}{T} \sum_{t=1}^T \frac{\|S_t(y_i)\|_F \cdot \|S_t(\hat{y}_i)\|_F}{\|S_t(y_i)\|_F \cdot \|S_t(\hat{y}_i)\|_F} \quad (16)$$

其中,  $S_t(y_i)$  和  $S_t(\hat{y}_i)$  表示信号  $y_i$  和  $\hat{y}_i$  在第  $t$  个时间窗的短时傅里叶变换 (STFT) 复数矩阵,  $\hat{y}_i$  为预测的值,  $y_i$  为真

实值,  $\odot$ 表示逐元素乘积,  $\|\cdot\|_F$ 为 Frobenius 范数.

$$Spectral-sim = 1 - \frac{\sum_{k=1}^N (P_{true}(f_k) - P_{gen}(f_k))^2}{\sum_{k=1}^N P_{true}^2(f_k)} \quad (17)$$

其中,  $P_{true}(f_k)$ 为真实信号在频率 $f_k$ 处的功率谱密度,  $P_{gen}(f_k)$ 为生成信号在频率 $f_k$ 处的功率谱密度.

$$MMD(P, Q; H) = \sup_{\|f\|_H \leq 1} (E_{x \sim P}[f(x)] - E_{y \sim Q}[f(y)]) \quad (18)$$

其中,  $f$ 是RKHS中的函数,  $H$ 是再生核希尔伯特空间,  $\|f\|_H \leq 1$ 表示函数 $f$ 的RKHS范数被约束在1以内.

我们将提出的 Transformer 扩散模型与多种基准模型进行比较, 这些模型包括传统时间序列分析方法与现代深度学习生成技术. 详细介绍如表3所示.

### 4.3 实验结果与分析

为了验证模型的有效性, 本文选取了一些主流的扩散模型算法进行实验比较, 实验结果如表4所示, 其中MMD值评估生成数据与真实数据的分布一致性. MMD值越小, 表示两者分布越接近, 由表4可知, 相较于RF-Diffusion<sup>[6]</sup>、AR<sup>[3]</sup>、MA<sup>[26]</sup>、ARMA<sup>[27]</sup>、GARCH<sup>[28]</sup>、WaveGAN<sup>[13]</sup>、TimeGAN<sup>[29]</sup>、GT-GAN<sup>[30]</sup>, 改进后的模型取得出色的实验结果. 相较于GAN类方

法, 比如WaveGAN, TransDiffusion在MMD、MSE和Spectral-sim指标上分别实现78.57%、77.18%和105.23%的显著提升, 证明其在分布对齐、时域精度和频域保真度方面的优越性. 尽管STFT-sim提升幅度相对较小(6.47%), 但结合其他指标可知, TransDiffusion在保留时频局部结构的同时, 避免了WaveGAN的频域能量失真问题; 相较于传统时序模型, 比如AR, MMD提升了79.38%, 表明TransDiffusion生成数据的分布更接近真实数据, 而AR模型因强假设限制(如线性)导致生成样本多样性不足. MSE下降157.58%, MAE下降58.37%, 反映了AR模型在简单波形拟合上具有优势, 但其生成模式单一. STFT-sim提升156.50%, Spectral-sim提升918.27%, 证明AR模型生成的信号频域能量分布与真实数据严重偏离, 而本文方法通过复杂建模显著改善了这一缺陷, AR模型虽在时域指标(MSE/MAE)上表现优异, 但其频域失真(Spectral-sim=0.0509)和分布偏移(MMD=0.6982)揭示了传统方法的局限性; 尽管AR模型在时域精度(MSE=0.0745)上表现优异, 但其生成的信号严重偏离真实数据的频域分布(Spectral-sim=0.0509)且缺乏多样性(MMD=0.6982).

表3 基准模型对比

模型名称	类型	功能描述	应用领域
RF-Diffusion <sup>[6]</sup>	深度学习	基于时间-频率扩散理论生成高质量时间序列RF数据, 捕捉时间、频率和复值域特征	RF信号生成
AR <sup>[3]</sup>	传统时间序列	当前值基于过去值的线性组合, 预测未来值	时间序列预测
MA <sup>[26]</sup>	传统时间序列	当前值基于过去误差项的线性组合, 建模残差依赖	时间序列分析
ARMA <sup>[27]</sup>	传统时间序列	结合AR和MA, 全面表示时间序列数据	时间序列建模
GARCH <sup>[28]</sup>	传统时间序列	建模具有波动聚集特征的时间序列, 捕捉动态方差	金融数据、通信信号波动
WaveGAN <sup>[13]</sup>	深度学习 (GAN)	使用小波变换生成时间序列, 捕捉特征分布和时间动态	时间序列生成
TimeGAN <sup>[29]</sup>	深度学习 (GAN)	结合无监督和监督方法, 生成保持时间动态的合成时间序列	时间序列生成

表4 不同模型算法的实验结果

指标	MMD↓	MSE↓	MAE↓	STFT-sim↑	Spectral-sim↑
TransDiffusion	0.1437	0.1918	0.3508	0.7713	0.5182
RF-Diffusion <sup>[6]</sup>	0.9652	0.3245	0.5056	0.3033	0.4922
SwinLSTM <sup>[14]</sup>	0.3749	0.2046	0.5122	0.5268	0.4356
AR <sup>[3]</sup>	0.6982	0.0745	0.2215	0.3007	0.0509
MA <sup>[26]</sup>	0.6982	0.0745	0.2215	0.3007	0.4211
ARMA <sup>[27]</sup>	0.6982	0.0745	0.2215	0.3008	0.0509
GARCH <sup>[28]</sup>	0.5961	0.0942	0.2315	0.7143	0.5095
WaveGAN <sup>[13]</sup>	0.6710	0.8405	0.8483	0.7244	0.2525
TimeGAN <sup>[29]</sup>	0.6802	0.0750	0.2221	0.5077	0.2691
GT-GAN <sup>[30]</sup>	0.4458	0.4162	0.5498	0.7731	0.5175

TransDiffusion通过引入扩散过程, 在显著提升分布匹配性(MMD下降79.38%)和频域保真度(Spectral-

sim上升918.27%)的同时, 仅适度牺牲时域精度(MSE上升157.58%). 这一权衡表明, TransDiffusion更适用

于需生成复杂时序模式的任务(如非平稳信号合成),而传统方法仅适合简单波形拟合场景;相较于扩散模型,比如 RF-Diffusion, TransDiffusion 在  $MMD$ 、 $MSE$  和  $STFT-sim$  指标上分别实现 85.09%、40.89% 和 154.14% 的显著提升,验证了其在分布对齐、时域精度及时频特征保留上的优势。尽管  $Spectral-sim$  提升有限(5.28%),但结合其他指标可知,TransDiffusion 在频域能量分布上仍具有优势。

为了更进一步验证本文提出的 TransDiffusion 模型在时间序列信号生成任务中的有效性,我们选取了基于循环神经网络(RNN)的代表性模型 SwinLSTM 作为对比模型。SwinLSTM 模型在传统 LSTM 网络的基础上,引入了 Swin Transformer 的局部窗口注意力机制,以增强模型在长序列依赖建模和局部时频模式提取方面的能力。该模型能够在捕捉时间动态的同时,利用窗口化的自注意力结构挖掘信号中的多尺度关联特征,兼具 RNN 的时间记忆性与 Transformer 的全局建模优势。在具体实现过程中, SwinLSTM 模型以时间序列为输入,通过滑动窗口划分为局部片段,先利用局部注意力机制提取每个片段的关键特征,再通过堆叠的 LSTM 层建模时间依赖关系。该结构结合了局部注意力的特征聚合能力与 LSTM 的时序记忆能力,能够同时捕捉信号的短时变化与长期动态。SwinLSTM 与 TransDiffusion 的实验结果如表 4 所示。

由表 4 可以看出, TransDiffusion 在  $MMD$ 、 $MSE$  和  $MAE$  指标上分别为 0.1437、0.1918 和 0.3508,均优于 SwinLSTM 的 0.3749、0.2046 和 0.5122,说明其生成数据在分布匹配性和时域精度上更接近真实信号;在  $STFT-sim$  和  $Spectral-sim$  指标上, TransDiffusion 分别为 0.7713 和 0.5182,相比 SwinLSTM 的 0.5268 和 0.4356 也有所提升,表明其在保留信号局部时频模式和频域能量分布方面具有优势。综合来看, TransDiffusion 不仅能够生成分布更加接近真实数据的多样化样本,还能在保持复杂时序模式的同时提高时域和频域的精度,相较于 SwinLSTM 展现出更强的综合性能和适应性。

#### 4.4 真实射频数据集性能评估

虽然自建数据集(Matlab 仿真,详见表 2)提供了丰富信息,但其在真实射频信号上的性能尚未完全验证。为了进一步评估 TransDiffusion 的泛化能力,我们在一个包含 LTE、Wi-Fi 和 DVB-T 的 IQ 样本的真实数据集上进行了评估,该数据集为比利时根特市 6

个不同地点的空口(over-the-air)采集:UZ、Reep、Rabot、Merelbeke、iGent 和 Gentbrugge。由于建筑密度、交通状况和环境噪声的差异,每个地点具有独特的信号传播特性。

在实验中,我们选择了 Merelbeke 地区的数据。该郊区环境建筑密度低、交通适中,其信号传播条件与本文描述的场景较为接近。经过预处理并转换为适合模型输入的格式后,共获得 8870 个样本。

表 5 展示了 TransDiffusion 在仿真数据集和真实 DEDICAT 数据集上的性能。正如预期, TransDiffusion 在仿真数据集上的性能更优,其中  $MMD$  为 0.1437,而真实数据集为 0.2154;  $MSE$  为 0.1918,而真实数据集为 0.2483。这是因为仿真环境相对理想化,多径效应和干扰被简化处理。在捕捉真实无线环境的 DEDICAT 数据集上,模型性能略有下降,如  $STFT-sim$  从 0.7713 降至 0.6328。尽管如此,模型仍保持了较高的频谱相似性(0.5182 对比 0.4851),表明 TransDiffusion 在实际射频频信号环境中仍具有良好的泛化能力和有效性。

表 5 TransDiffusion 在 DEDICAT 数据集上的性能

Dataset	$MMD\downarrow$	$MSE\downarrow$	$MAE\downarrow$	$STFT-sim\uparrow$	$Spectral-sim\uparrow$
DEDICAT	0.2154	0.2483	0.4127	0.6328	0.4851
Simulated dataset	0.1437	0.1918	0.3508	0.7713	0.5182

这些结果表明,尽管噪声、动态多径和干扰等真实环境因素会增加建模难度, TransDiffusion 依然展现出较强的泛化能力。未来工作将致力于进一步提升模型在复杂环境下的鲁棒性。

#### 4.5 消融实验

为了进一步验证 TransDiffusion 模型各模块对整体性能的贡献,本文进行了消融实验。分别移除了多尺度映射与重构模块和周期注意力模块,并将实验结果与完整的 TransDiffusion 模型进行对比。参与消融实验的模块位于图 10 所示解码器部分,包含两个消融模型:去除多尺度映射与重构模块模型和去除周期注意力模块模型。在自建的高动态数据集上设计并进行了 3 组消融实验:(1)去除多尺度映射与重构模块,仅保留周期注意力模块;(2)去除周期注意力模块,仅保留多尺度映射与重构模块;(3)同时去除这两个模块,其余网络结构保持不变。对于每个实验配置,我们使用  $MMD$ 、 $MSE$ 、 $MAE$ 、 $STFT-sim$  和  $Spectral-sim$  作为评价指标。实验结果如表 6 所示。

表 6 消融实验结果

模型配置	<i>MMD</i>	<i>MSE</i>	<i>MAE</i>	<i>STFT-sim</i>	<i>Spectral-sim</i>
完整模型	0.1437	0.1918	0.3508	0.7713	0.5182
去除多尺度映射与重构模块	0.5824	0.2543	0.4125	0.5839	0.4038
去除周期注意力模块	0.2748	0.1638	0.3692	0.4920	0.4973
同时去除多尺度映射与重构模块和周期注意力模块	0.7354	0.3086	0.8642	0.0349	0.1653

消融实验结果表明,多尺度映射与重构模块和周期注意力模块各自针对不同的重建难点发挥了关键作用.首先,移除多尺度映射与重构模块后,*MMD*从0.1437升至0.5824,*MSE*和*MAE*分别从0.1918、0.3508上升到0.2543、0.4125,*STFT-sim*由0.7713降至0.5839,*Spectral-sim*由0.5182降至0.4038.这说明,多尺度模块在捕捉时间序列中不同尺度的细粒度波动和长程趋势方面至关重要,缺失它会严重削弱模型对数据分布的对齐能力以及对时域和时-频局部结构的恢复精度.

其次,移除周期注意力模块时,*MMD*上升至0.2748,*STFT-sim*和*Spectral-sim*分别下降到0.4920和0.4973,表明周期模块对于提取信号的主导频率成分和维护频域一致性很重要.但是,此时*MSE*改善到0.1638,可能是因为去掉频域专注结构后,模型在时域回归任务上更专一地最小化平方误差,但总体*MAE*上升至0.3692,表明这种“时域优化”是以牺牲频域特征保真为代价的.

最后,当同时去除两个模块时,所有指标均退化到最差.*MMD*达到0.7354,*MSE*达到0.3086,*MAE*达到0.8642,*STFT-sim*降至0.0349,*Spectral-sim*降至0.1653.这一结果充分证明,多尺度映射与重构与周期注意力两个模块协同作用,不仅让模型能够兼顾长短期动态和局部细节,还能保障生成信号在频域上的完整性和连贯性,二者缺一都会导致模型在高动态信号重建任务中性能大幅下滑.

#### 4.6 参数影响

为了评估动态场景下不同参数对TransDiffusion性能的影响,我们进行了参数影响实验.在所有对比实验中,除正在变化的参数外,其他参数均保持如下基线设置:行人速度:1.5 m/s,车辆速度:80 km/h,遮挡半径:50 m,信号衰减:-6 dB,噪声类型:引擎调制+人声干扰.

##### 4.6.1 运动速度设置

我们首先研究动态运动条件对TransDiffusion性能的影响.具体而言,在行人速度为0、1.5、2.0、3.5和5 m/s及车辆速度为40、80、120和160 km/h的情况下进行测试,用于分析TransDiffusion在不同运动速

度下的预测能力.图13和图14分别绘制了在不同行人速度和车辆速度下获得的评价指标结果.

从图13和图14可以明显看出,随着行人速度的增加,TransDiffusion在分布对齐、时域重建和时频保真度方面表现出明显的下降.模型难以跟随快速的局部波动,导致分布漂移、时域误差增大,以及*STFT-sim*和*Spectral-sim*降低.总体而言,更高的行人速度显著增加了信号生成的难度,使各项指标性能下降.车辆速度也呈现类似规律,随着车辆速度的提高,*MMD*、*MSE*和*MAE*明显增加,而*STFT-sim*和*Spectral-sim*下降,说明更快的运动增加了信号的动态性,使建模更加困难.特别是当车辆速度超过120 km/h时,TransDiffusion输出的信号谱一致性显著下降.

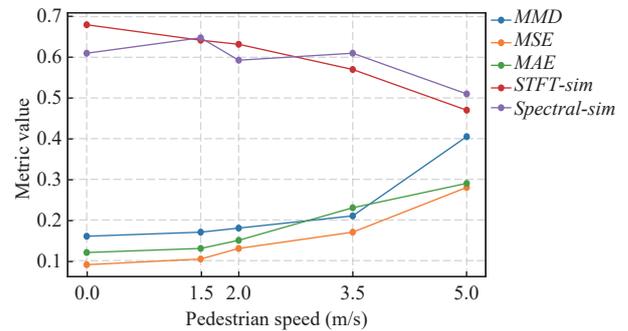


图 13 不同行人速度下的评估结果

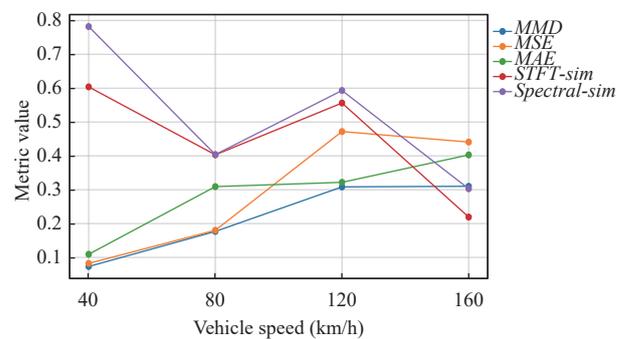


图 14 不同车速下的评估结果

##### 4.6.2 遮挡模型设置

遮挡模型通过改变障碍物半径和信号衰减来模拟多径传播和阻挡效应.我们在遮挡半径为15 m、30 m、

45 m、60 m 和 75 m, 以及信号衰减为-1 dB、-4 dB、-6 dB、-8 dB 和-10 dB 的条件下评估了模型的性能。

图 15 和图 16 分别绘制了在不同遮挡半径和衰减水平下获得的评价指标结果。从图 15 可以看出, 随着障碍物半径的增加, 生成信号与真实信号之间的分布差异略有上升, 时域误差整体增大, 并且在最大半径下重建精度明显下降。同时, *STFT-sim* 和 *Spectral-sim* 均显著下降, 说明更大范围的障碍物会引起多径失真和能量重分布, 从而抑制了模型在时域和频域的重建能力。从图 16 可以明显看出, 随着信号衰减加深, 模型的性能在时域和频域均呈非线性下降趋势: 分布差异增大, 表明适度衰减即可引起显著扰动; 尽管存在一定波动, 时域误差总体上升; 频域保真度稳步下降, 说明更强的衰减削弱了关键谱分量, 阻碍 TransDiffusion 重建信号频率结构的能力。

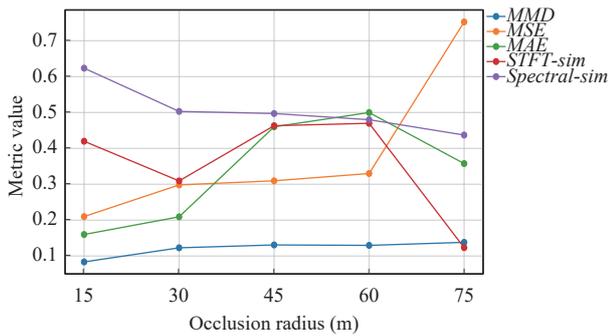


图 15 不同遮挡半径下的评估结果

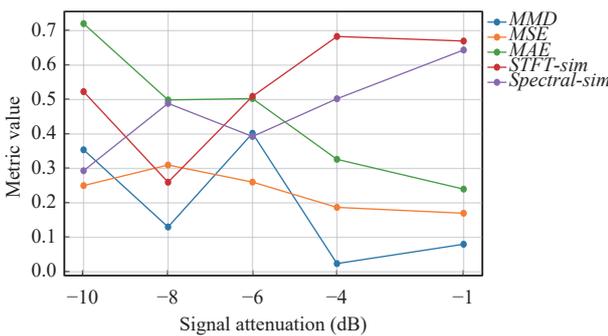


图 16 不同信号衰减条件下的评估结果

## 5 结论与展望

本文针对高动态时间序列生成任务中动态特征捕捉与时空关联建模的难题, 提出了一种基于 Transformer 架构的改进型扩散模型。通过构建多层级时空注意力机制与自适应噪声预测网络, 实现了对快速演

化时空特征的细粒度捕捉。模型设计中引入的多尺度映射与重构模块有效提升了非平稳趋势建模能力, 而周期注意力模块通过频域注意力机制强化了周期特征的提取精度。在自主构建的 Matlab 高动态仿真数据集上的实验表明, 相较于 DDPM、WaveGAN 等基线模型, 本文方法在 *MMD*、*MSE* 等核心指标上取得显著提升, 证明其在分布一致性、时域重建精度及时频特征保留方面的优越性。特别是通过多层深度解码器与多尺度残差重构路径的协同设计, 模型成功解决了传统方法在局部突变与全局趋势耦合建模中的矛盾, 为动态时序生成提供了新的技术路径。

本研究为高动态时间序列生成提供了新的解决方案, 但其物理可解释性与极端场景鲁棒性仍需持续改进。随着神经微分方程与物理引导生成技术的发展, 构建兼具数据驱动灵活性与物理规律约束性的生成模型将成为重要研究方向。

## 参考文献

- Hu QL, Zhang JQ, He YY. An embedding-based multi-modal few-shot learning method for network intrusion detection. Proceedings of the 21st International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). Chengdu: IEEE, 2024. 1-4.
- Pourpanah F, Abdar M, Luo YX, et al. A review of generalized zero-shot learning methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 4051-4070.
- Li X, Zhang L, Wang Z. Forecast methods for time series data: A survey. IEEE Access, 2021, 9: 90504-90518. [doi: 10.1109/ACCESS.2021.3091162]
- Huang ZZ, Zhang JP, Zhang Y, et al. DU-GAN: Generative adversarial networks with dual-domain U-Net-based discriminators for low-dose CT denoising. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 4500512.
- Letafati M, Ali S, Latva-aho M. Denoising diffusion probabilistic models for hardware-impaired communications. Proceedings of the 2024 IEEE Wireless Communications and Networking Conference (WCNC). Dubai: IEEE, 2024. 1-6.
- Chi GX, Yang Z, Wu CS, et al. RF-diffusion: Radio signal generation via time-frequency diffusion. Proceedings of the 30th Annual International Conference on Mobile Computing and Networking. Washington: ACM, 2024. 77-92.
- Kong Z, Ping W, Huang J, et al. DiffWave: A versatile

- diffusion model for audio synthesis. Proceedings of the 2021 International Conference on Learning Representations (ICLR). Vienna: OpenReview.net, 2021.
- 8 Zhang CY, Chen YF, Fan ZX, *et al.* TC-DiffRecon: Texture coordination MRI reconstruction method based on diffusion model modified MF-UNet method. Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging (ISBI), Athens: IEEE, 2024. 1–5.
  - 9 Jin L, Li S, Hu B. RNN models for dynamic matrix inversion: A control-theoretical perspective. IEEE Transactions on Industrial Informatics, 2018, 14(1): 189–199. [doi: [10.1109/TII.2017.2717079](https://doi.org/10.1109/TII.2017.2717079)]
  - 10 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
  - 11 Deng T, Ma ML, Liu QH, *et al.* High-precision carrier tracking algorithm for extremely weak and high-dynamic signals. Radio Science, 2021, 56(5): 1–14.
  - 12 Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. arXiv:2006.11239, 2020.
  - 13 Yamamoto R, Song E, Kim JM. Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram. Proceedings of the 45th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona: IEEE, 2020. 6199–6203.
  - 14 Tang S, Li C, Zhang P, *et al.* SwinLSTM: Improving spatiotemporal prediction accuracy using swin Transformer and LSTM. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023. 13424–13433.
  - 15 Chen X, Mishra N, Rohaninejad M, *et al.* PixelSNAIL: An improved autoregressive generative model. Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018. 864–872.
  - 16 Ma T, Zhu HB, Huang WJ, *et al.* Stereo image dense matching based on SGM constrained by feature matching. Proceedings of the 9th International Conference on Computer and Communications (ICCC). Chengdu: IEEE, 2023. 1911–1915.
  - 17 Kolářová E, Brančík L. Noise influenced transmission line model via partial stochastic differential equations. Proceedings of the 42nd International Conference on Telecommunications and Signal Processing (TSP). Budapest: IEEE, 2019. 492–495.
  - 18 Yang RF, Wang ZJ, Jiang B, *et al.* Leveraging drift to improve sample complexity of variance exploding diffusion models. Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2024. 3420.
  - 19 Wang YC, Ebara H. Variance-preserving stochastic differential equation algorithm for the traveling salesman problem. Proceedings of the 2024 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS). Dubrovnik: IEEE, 2024. 54–61.
  - 20 Chen N, Zhang Y, Zen H, *et al.* WaveGrad: Estimating gradients for waveform generation. Proceedings of the 2021 International Conference on Learning Representations (ICLR). Vienna: OpenReview.net, 2021.
  - 21 Rasul K, Seward C, Schuster I, *et al.* Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021. 8857–8868.
  - 22 Letafati M, Ali S, Latva-Aho M. conditional denoising diffusion probabilistic models for data reconstruction enhancement in wireless communications. IEEE Transactions on Machine Learning in Communications and Networking, 2025, 3: 133–146. [doi: [10.1109/TMLCN.2024.3522872](https://doi.org/10.1109/TMLCN.2024.3522872)]
  - 23 Tashiro Y, Song JM, Song Y, *et al.* CSDI: Conditional score-based diffusion models for probabilistic time series imputation. Proceedings of the 35th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2021. 1900.
  - 24 Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial networks. Communications of the ACM, 2020, 63(11): 139–144. [doi: [10.1145/3422622](https://doi.org/10.1145/3422622)]
  - 25 Kingma DP, Welling M. Auto-encoding variational bayes. Proceedings of the 2014 International Conference on Learning Representations (ICLR). Banff: OpenReview.net, 2014. 1–14.
  - 26 Kato H, Taniguchi M, Honda M. Statistical analysis for multiplicatively modulated nonlinear autoregressive model and its applications to electrophysiological signal analysis in humans. IEEE Transactions on Signal Processing, 2006, 54(9): 3414–3425. [doi: [10.1109/TSP.2006.877663](https://doi.org/10.1109/TSP.2006.877663)]
  - 27 Kashyap RL. Optimal choice of AR and MA parts in autoregressive moving average models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1982, PAMI-4(2): 99–104.
  - 28 Wang Y, Sznaiar M, Camps O, *et al.* Identification of a class of generalized autoregressive conditional heteroskedasticity

- (GARCH) models with applications to covariance propagation. Proceedings of the 54th IEEE Conference on Decision and Control (CDC). Osaka: IEEE, 2015. 795–800.
- 29 Öztürk C. Enhancing financial time-series analysis with timeGAN: A novel approach. Proceedings of the 9th International Conference on Computer Science and Engineering (UBMK). Antalya: IEEE, 2024. 447–450.
- 30 Jeon J, Kim J, Song H, *et al.* GT-GAN: General purpose time series synthesis with generative adversarial networks. Advances in Neural Information Processing Systems, 2022, 35: 36999–37010.

(校对责编: 李慧鑫)