引用:王晓涵,田宽,邱义杨,等. 频谱约束潜空间扩散模型及其在多变量时序生成中的应用[J]. 聊城大学学报(自然科学版),2026,39(6):833-846.

Citation: Wang Xiaohan, Tian Kuan, Qiu Yiyang, Yu Xiaofeng, Hu Xinyue, Yang Xinru, Wang Yongkang, Luo Guifu, Yang Fang, Yu Xu. Spectral-constrained latent diffusion model and its application in mulitivariate time series generation[J]. Journal of Liaocheng University (Natural Science Edition),2026,39(6):833-846.

文章编号 1672-6634(2026)06-0833-14 **DOI** 10.19728/j.issn1672-6634.2026040018

频谱约束潜空间扩散模型及其在多变量时序生成中的应用

王晓涵[1,2,3],田宽[1,2,3],邱义杨[1,2,3],于晓枫[4],胡欣悦[1,2,3],杨新汝[1,2,3],王永康[1,2,3],罗桂富[5],杨芳[5],于旭[1,2,3]

(1. 中国石油大学(华东) 计算机科学与技术学院,山东 青岛 266580;2. 中国石油大学(华东) 青岛软件学院,山东 青岛 266580;3. 中国石油大学(华东) 山东省智能油气工业软件重点实验室,山东 青岛 266580;4. 山东极视角科技有限公司,山东 青岛 266000;5. 青岛华正信息技术股份有限公司,山东 青岛 266101)

摘要 现有扩散生成模型在处理多变量时间序列数据时,存在复杂时序特征提取难、频域与全局分布易失真以及推理低效等瓶颈;为解决上述问题,提出一种频谱约束潜空间扩散时序生成方法。首先,利用 1D Autoencoder 将高维时序数据压缩至潜空间以提取跨通道的相关性特征;然后,在潜空间中通过集成多头自注意力机制的 1D UNet 执行反向去噪,并引入频谱损失约束以保留数据的周期性结构;最后,结合加速采样器与分布校正、均值还原技术实现与真实分布的统计对齐。多数据集实验表明,SpecLDM-TS 多项指标显著优于基线,生成样本在统计特征与波形动态方面较好地逼近真实数据分布,验证了其在多变量时序生成任务中的有效性与优越性。

关键词 去噪扩散模型;时序数据生成;多变量时间序列;潜空间架构;频谱损失;数据增强

中图分类号 TP181 **文献标志码** A

Spectral-constrained latent diffusion model and its application in multivariate time series generation

WANG Xiaohan[1,2,3], TIAN Kuan[1,2,3], QIU Yiyang[1,2,3], YU Xiaofeng[4], HU Xinyue[1,2,3], YANG Xinru[1,2,3], WANG Yongkang[1,2,3], LUO Guifu[5], YANG Fang[5], YU Xu[1,2,3]

(1. College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China; 2. Qingdao Institute of Software, China University of Petroleum (East China), Qingdao 266580, China; 3. Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, China University of Petroleum (East China), Qingdao 266580, China; 4. Shandong Extreme Vision Technology Co., Ltd., Qingdao 266000, China; 5. Qingdao HuaZheng Information Co., Ltd, Qingdao 266101, China)

收稿日期:2026-04-28

基金项目:国家自然科学基金项目(62172249,62472441);青岛市自然科学基金项目(24-8-4-zrjj-3-jch);山东省重点研发计划(2024TSGC0195)资助

通信作者:于旭,男,汉,博士,教授,研究方向:数据挖掘与迁移学习,E-mail:yuxu0532@upc.edu.cn。

Abstract Existing diffusion generative models encounter bottlenecks such as difficulty in extracting complex temporal features, distortion of frequency and global distributions, and low inference efficiency when processing multivariate time series data. To address these issues, a Spectral-Constrained Latent Diffusion Model for Time Series Generation (SpecLDM-TS) is proposed. Firstly, a 1D Autoencoder is utilized to compress high-dimensional time series data into a latent space to extract cross-channel correlation features. Subsequently, reverse denoising is performed in the latent space via a 1D UNet integrated with a multi-head self-attention mechanism, while a spectral loss constraint is introduced to preserve the periodic structures of the data. Finally, an accelerated sampler is combined with distribution correction and mean restoration techniques to achieve statistical alignment with the real distribution. Experimental results on multiple datasets demonstrate that SpecLDM-TS significantly outperforms baseline methods across various metrics. The generated samples closely approximate the real data distribution in terms of statistical characteristics and waveform dynamics, validating the effectiveness and superiority of the model in multivariate time series generation tasks.

Keywords denoising diffusion models; time series generation; multivariate time series; latent space architecure; spectrum loss; data augmentation

0 引言

多变量时间序列(Multivariate Time Series, MTS)作为现代工业系统、能源互联网及智能监控中的核心数据形态,刻画了系统运行状态的动态演化过程以及不同变量之间的时空相关性[1]。与单变量时间序列相比,多变量序列不仅包含各维度自身的时间依赖关系,还蕴含变量之间的耦合关系与潜在交互机制,因此其建模难度显著提升。在实际应用场景中,如电力负荷监测、工业设备运行状态分析以及智能制造系统监控等,多变量时间序列数据通常呈现出高维度、强相关性以及非平稳性等特征,这对模型的表达能力和泛化能力提出了更高要求。对该类数据进行有效建模,是实现异常检测、趋势预测和智能决策的重要基础[2]。然而,在实际工程环境中,由于数据隐私保护政策限制、传感器设备故障以及极端异常工况样本稀缺等因素,高质量且具有代表性的时序数据往往难以获取,这已成为数据驱动模型性能提升的重要瓶颈[3]。因此,构建高保真的时序生成模型,通过合成具有真实统计特性的虚拟数据以实现数据增强,是缓解数据稀缺问题、提升模型鲁棒性的重要途径,也逐渐成为工业人工智能领域的研究热点[4]。

近年来,去噪扩散概率模型(Denoising Diffusion Probabilistic Model, DDPM)凭借稳定的训练过程和强大的分布建模能力,在图像生成和语音合成等领域取得了显著进展[5-6]。扩散模型通过构建逐步加噪与反向去噪的过程,使模型能够逐渐学习数据分布的细粒度结构,从而在生成质量与多样性之间取得良好平衡。与生成对抗网络(GANs)相比,扩散模型在生成样本的多样性与稳定性方面具有明显优势[7]。这一建模范式也逐渐被引入时间序列领域,例如,TimeGrad 通过自回归方式结合扩散过程,实现了对多变量时间序列的概率建模;CSDI 则利用条件扩散模型完成缺失值填补任务,这些工作验证了扩散模型在捕捉多变量依赖关系与不确定性建模方面的潜力[8-9]。为进一步提升高维数据建模效率,潜空间扩散模型(Latent Diffusion Model, LDM)通过先将数据映射到低维潜空间再进行生成,有效降低了计算复杂度并提升了生成质量[10-11]。

尽管潜空间扩散模型在低维任务中表现良好,但在高维多变量时序场景(如包含数百测量通道的电力负荷数据)中仍面临两方面挑战。其一是高维特征表征受限与推理效率瓶颈。面对具有数百测量通道的复杂工业数据,扩散模型不仅难以在原始高维空间捕获深层时空依赖,由于采样效率固有偏低,推理速度难以满足大规模数据合成需求[12]。虽然已有研究通过最大均值差异(MMD)等方法对扩散模型的推理过程进行了加速探索[1],但针对多变量时序特征表征与生成效率协同优化的研究仍较少。其二是全局分布与频域特征易失真。现有方法多采用时域均方误差(MSE)作为训练目标[13],容易导致模型在生成过程中忽略高频成分,从而使生成序列在频谱分布上偏离真实数据的周期性结构[14]。

针对上述问题,本文提出一种基于潜空间扩散模型的多变量时间序列生成方法 SpecLDM-TS(Spectral-

Constrained Latent Diffusion Model for Multivariate Time Series Generation)。整体框架由“编码-潜空间生成-解码”三阶段组成。在数据预处理阶段，引入去均值操作以消除直流分量对模型学习的干扰，并结合局部切片(Patching)策略对长序列进行分段处理，从而提升模型对局部结构的建模能力。同时，设计支持高维多通道输入的 1D Autoencoder，将原始序列高效压缩至潜空间，实现信息表达与计算效率之间的平衡。在潜空间生成阶段，以融合多头自注意力机制的 1D UNet 为核心网络，通过自注意力机制增强模型对长距离依赖关系与跨变量交互特征的捕捉能力。此外，引入快速傅里叶变换(FFT)构建频谱损失函数，通过在频域上约束生成信号的幅度分布，实现时域与频域的联合优化，从而有效缓解高频信息丢失问题。在采样阶段，结合跳步采样策略(如 DDIM)以降低生成所需的迭代步数，并引入基于统计量匹配的分布校正机制，对潜变量分布进行后处理，使生成数据在均值与方差等统计特性上更加接近真实数据，从而在提升推理效率的同时保证生成质量。

本文的主要贡献归纳如下：(1) 提出了一种频谱约束潜空间扩散时序生成模型(SpecLDM-TS)。通过 1D Autoencoder 与自注意力 1D UNet 的协同运作，将潜空间生成理念应用于高并发序列数据生成中，有效突破了复杂时序建模的计算限制，实现了对高维序列数据的高效建模。(2) 针对扩散模型在时序生成中易丢失高频特征的问题，显式引入 FFT 频谱振幅差值作为优化目标，缓解传统扩散模型中高频信息丢失的问题。(3) 在真实多变量数据集上进行了实验验证。结果表明，SpecLDM-TS 在分布一致性、重构误差及统计特征等方面均优于基线方法。

1　相关工作

1.1　基于扩散模型的时序数据生成

扩散模型是一类通过逐步向数据中注入噪声并学习反向去噪过程来拟合真实数据分布的生成模型。其核心包含两个马尔可夫过程：在前向过程中，模型从真实样本出发，按步添加高斯噪声，直至数据逐渐逼近标准高斯分布；在反向过程中，模型通过神经网络学习去噪策略，将随机噪声逐步还原为符合真实分布的生成样本[5,15]。由于这种逐步建模方式能够在不同噪声尺度下刻画数据特性，扩散模型在处理复杂高维数据时表现出较强的建模能力，近年来在生成建模领域取得了较为显著的进展[6]。与生成对抗网络(GAN)相比，扩散模型无需依赖生成器与判别器的对抗训练，从而避免了训练不稳定、模式崩溃等问题，能够更稳健、更完整地拟合真实数据分布[16]。与变分自编码器(VAE)相比，扩散模型无需对潜变量空间施加严格的分布假设，其渐进去噪机制能够保留更丰富的细节与结构信息，在生成质量、保真度和多样性方面通常更优[17]。

扩散模型在图像生成领域展现出显著效果后[5]，研究者逐渐将其引入时间序列建模领域，用于学习复杂时序数据的动态分布特征。现有研究从不同角度改进扩散模型以适应时序数据结构。例如，Rasul 等[8]提出的 TimeGrad，借助自回归结构与扩散模型相结合，对多变量时间序列分布展开建模，有效捕捉时间依赖关系。Yuan 等[13]提出的 Diffusion-TS，运用基于卷积的时序建模结构，增强对局部时间模式的捕捉能力。Zhou 等[18]提出了非自回归扩散时序模型，通过残差去噪网络提高长期依赖建模能力。

随着研究深入，扩散模型在时序领域逐渐从无条件生成扩展到条件生成与长序列建模。Tashiro 等[9]提出的 CSDI 将条件扩散用于时间序列补全任务，通过历史观测和缺失掩码引导去噪过程，从而提升了不完整序列的重建质量与全局一致性。Fons 等[19]进一步提出 LSCD，将 Lomb-Scargle 条件信息引入扩散框架，以增强不规则采样场景下的时间序列补全能力。在长时序建模与高效生成方面，Song 等[12]提出的 DDIM 通过确定性隐式反向过程实现少步采样，显著降低了扩散模型的生成迭代成本。Feng 等[20]提出的 LDT 则将潜空间扩散与 Transformer 结合，用于多变量时间序列预测，在压缩表示的同时提升了长序列建模效率。此外，Wang 等[21]提出 FreDF，从频域角度学习时序预测表示，说明频域建模有助于缓解长序列中的非平稳性与频率偏移问题。尽管这些工作在条件生成、采样效率和长序列建模方面取得了进展，但在高维多变量时序场景中，采样成本、长程依赖建模以及频率结构保真性仍然是亟待解决的问题。

尽管上述方法在时序数据生成方面有一定进展，但它们的优化目标大多是以均方误差(Mean Squared Error，MSE)或变分下界(Evidence Lower Bound，ELBO)为主[8,9,13]，对于生成序列和真实序列整体分布一致性的约束相对较少。已有研究表明，如果仅依赖时域去噪损失，模型通常难以充分保留时间序列的频率结

构、长期动态特征[22]。因此,在实际应用中容易出现统计特征偏移、频域特性失真以及长期依赖建模不足等问题,限制了扩散模型在高质量时序数据生成任务中的表现。

1.2 潜空间扩散模型

为了降低扩散模型在高维数据上的计算花销,潜空间扩散模型(Latent Diffusion Models, LDM)借助自编码器把原始数据映射到低维潜空间,在潜空间里执行扩散、去噪流程,最后经解码器恢复到数据空间。图1是潜空间扩散模型的计算流程图[10],这种机制能够降低计算复杂度,同时维持较高的生成质量。潜空间扩散模型最初应用于图像生成任务,并展现出良好的生成效果[10]。受此启发,研究者开始尝试将其拓展至视频生成与时序数据建模领域。

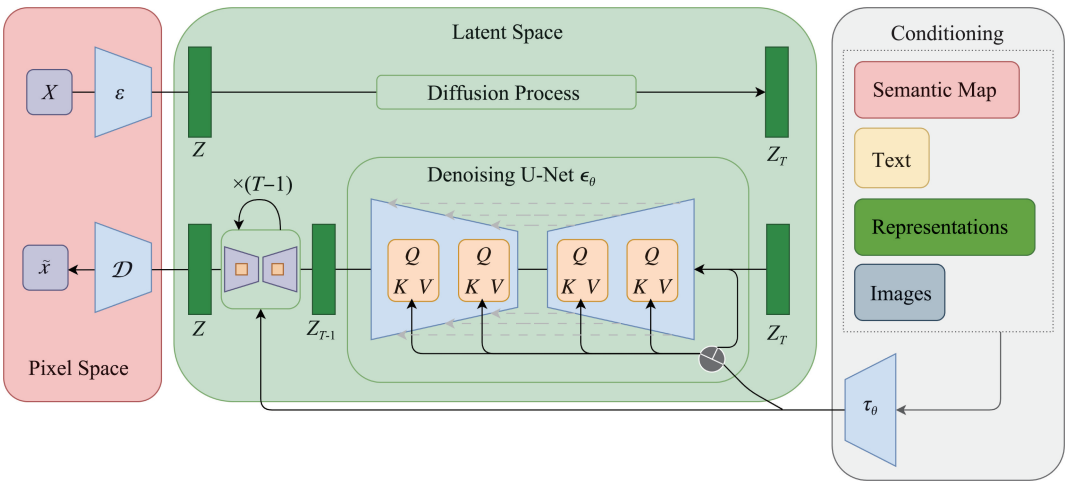


图1 潜空间扩散模型计算流程图

Fig. 1 Calculation flow of latent space diffusion model

在时序数据生成方面,大部分研究通常构建时序自编码器或变分自编码器以获得更为紧凑的潜空间表示,随后在潜空间中利用扩散模型学习序列的动态变化,以此提高生成效率并降低训练成本。针对时序数据,近期有研究把潜空间扩散模型迁移到该领域。例如,Feng 等[20]提出的 LDT(Latent Diffusion Transformer),利用统计感知的自编码器将高维多变量时序数据压缩至潜空间,并通过基于扩散的条件生成器进行预测,实现了潜空间中的时序建模。此外,Lin 等[16]在其综述中系统梳理了扩散模型在时间序列领域的应用现状,特别指出潜空间扩散模型在处理复杂多变量时序数据生成任务中展现出巨大潜力,并讨论了其未来发展方向。

上述工作验证了潜空间扩散模型在时序生成中的可行性,但在建模复杂多变量关系和控制潜空间分布方面仍存在不足。现有的潜空间时序扩散模型大多侧重于单变量或者低维时序表示,对于多变量间复杂相关结构的建模能力比较有限[23]。当推理步数比较少时,潜空间中的分布偏移问题越发突出,生成分布与真实分布之间的高阶统计匹配难以有效约束,进而对最终解码后的序列质量产生影响。

1.3 频谱约束与损失

时间序列数据通常具有明显的周期性和频率结构。若仅依赖时域损失函数训练生成模型,模型往往难以准确学习原始数据的频谱特征,导致生成序列在周期性和振荡模式方面与真实数据存在偏差。为解决这一问题,相关研究开始引入频谱约束,在频谱空间内让生成序列与真实序列对齐,以此增强模型对时间序列动态结构的建模能力[24-25]。常见的做法是对时间序列实施快速傅里叶变换(FFT),将时域信号转换到频谱空间,在频谱幅值或功率谱密度层面构建损失函数,从而约束生成序列的频率分布[19,21,26]。

频谱损失能够弥补时域均方误差在周期结构建模方面的不足,尤其适用于具有明显周期性或多尺度振荡特征的时间序列生成任务。已有研究通常将时域损失与频谱损失联合优化,以兼顾局部时间连续性与全局频率一致性,从而提升生成序列的真实性与稳定性[21,25,27]。例如,FreDF 通过频域表示学习缓解了非平稳时序中的频率偏移问题[21],CM-TFD 则通过时频解耦增强了多变量时序的频率表达能力[27]。

进一步地,频域思想也被引入扩散模型框架,用于使生成样本的频谱特征与真实数据保持一致。Fons

等[19]提出的 LSCD 将条件扩散与不规则采样场景下的频率信息结合，用于时间序列补全任务；Jeong 等[28]提出频率条件扩散方法，通过频率条件引导生成过程，提升了模型的频域可控性；Crabbé 等[29]则直接在频域中构建扩散过程，验证了频域表征在时间序列生成中的可行性。这些工作表明，频域信息已逐渐成为时序扩散建模的重要补充，但现有方法大多仍集中于原始时域空间或低维输入场景，尚缺乏将频谱约束与潜空间扩散统一结合、并面向高维多变量时间序列生成的系统研究[30]。

2　频谱约束潜空间扩散时序生成方法

本文提出一种频谱约束潜空间扩散时序生成方法（Spectral-Constrained Latent Diffusion Model for Multivariate Time Series Generation，SpecLDM-TS），如图 2 所示，采用了“编码-潜空间生成-解码”三阶段架构，将高维时序数据映射至低维潜空间，在该空间中完成扩散建模与分布优化，达到兼顾生成质量与计算效率的目标。

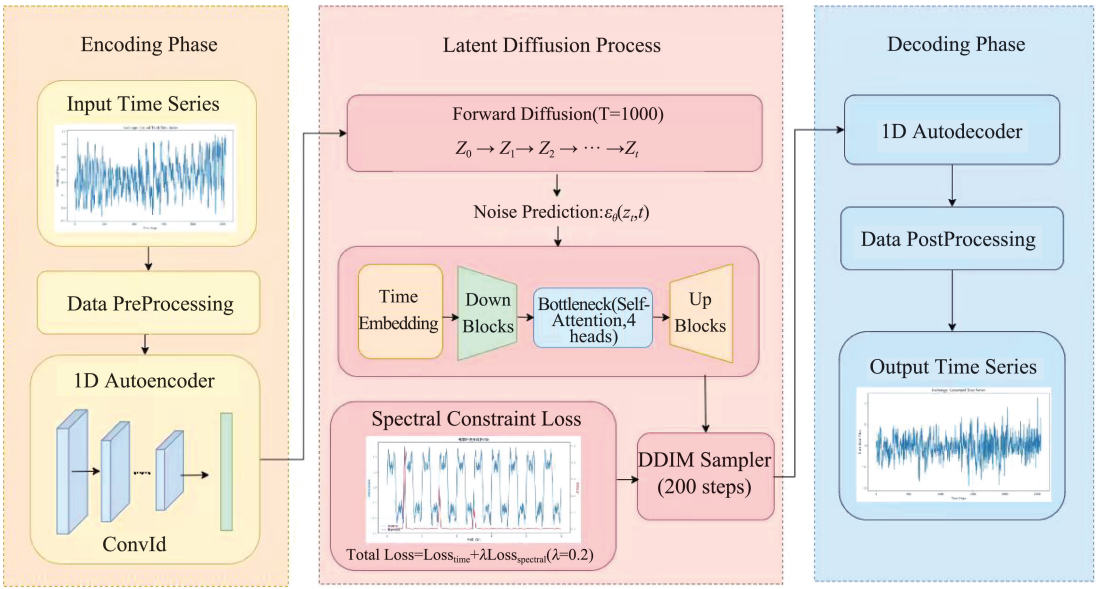


图 2　频谱约束潜空间扩散时序生成模型框架

Fig. 2　Spectral-constrained latent diffusion model for multivariate time series generation

2.1　自编码器

为降低扩散建模的复杂度并提升训练稳定性，SpecLDM-TS 采用一维自编码器对时序数据进行潜空间表示学习。编码器由 4 层一维卷积构成，通过逐级下采样提取时间依赖特征并压缩时间维度，将输入序列映射为低维潜变量

$$\boldsymbol{z} = E_{\text{ncdr}}(\boldsymbol{x}), \tag{1}$$

式中 $\boldsymbol{x}$ 表示原始多变量时序数据；$E_{\text{ncdr}}(\cdot)$ 表示一维卷积编码器；$\boldsymbol{z}$ 表示原始多变量时序数据的潜空间表示。在这一过程中，卷积核能够在局部时间窗口内提取短期动态变化模式，同时建模不同变量之间的相关性，从而实现对复杂时序结构的有效编码。相较于直接在原始空间进行建模，这种潜空间表示具有更紧凑的分布形式，并在一定程度上更接近高斯分布，使用该编码器将数据映射到潜空间可对结果产生积极作用。

解码器采用与编码器近似对称结构，由 3 个一维卷积层和 1 个输出卷积层组成，通过逐步上采样恢复时间分辨率，将潜变量映射回原始时序空间，实现对输入数据的重构 $\hat{\boldsymbol{x}} = D_{\text{ecdr}}(\hat{\boldsymbol{z}})$，式中 $D_{\text{ecdr}}(\cdot)$ 表示解码器；$\hat{\boldsymbol{x}}$ 表示生成的时间序列；$\hat{\boldsymbol{z}}$ 表示潜空间扩散模型生成的潜变量，将潜变量还原回原始数据空间，保证潜空间表示具有可逆性，为潜空间扩散模型生成提供可解释输出。

为保证潜空间表示的有效性与可逆性，模型通过最小化重构损失 $\mathcal{L}_{\text{AE}}$ 对自编码器进行训练

$$\mathcal{L}_{\text{AE}} = \frac{1}{N}\sum_{i=1}^{n} \| \boldsymbol{x}_i - \hat{\boldsymbol{x}}_i \|_2^2, \tag{2}$$

式中 $\boldsymbol{x}_i$ 表示第 i 个真实多变量时序样本，与式(1)中编码器输入 $\boldsymbol{x}$ 指代同一类对象；$\hat{\boldsymbol{x}}_i$ 表示生成序列；

$\|\cdot\|_2^2$ 表示 L2 误差项。使生成序列在数值和结构上尽可能逼近真实数据。该重构约束本质上限制了编码过程中的信息损失,避免潜空间发生信息塌缩,从而确保潜变量能够完整保留原始序列的关键特征。在处理多变量时间序列时,若直接在原始高维空间(如数百个通道)执行扩散模型的逐步去噪,不仅难以有效提取复杂时空依赖,且极易受冗余噪声干扰。为解决高维特征难以直接表征的痛点,本文采用的潜空间映射机制可通过变分推断(Variational Inference)中的证据下界(ELBO)得到严格的理论解释。假设真实数据的对数似然为 $\log p(\boldsymbol{x})$,引入潜变量 $\boldsymbol{z}$,其边际似然满足 $\log p(\boldsymbol{x}) \geqslant \mathbb{E}_{q_\varphi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\psi(\boldsymbol{x} \mid \boldsymbol{z})] - D_{\mathrm{KL}}(q_\varphi(\boldsymbol{z} \mid \boldsymbol{x}) \| p(\boldsymbol{z}))$,式中 $q_\varphi(\boldsymbol{z} \mid \boldsymbol{x})$ 对应于带有参数 φ 的编码过程,$p_\psi(\boldsymbol{x} \mid \boldsymbol{z})$ 对应于带有参数 ψ 的重构过程。最大化期望重构对数似然 $\mathbb{E}_{q_\varphi(\boldsymbol{z}|\boldsymbol{x})}[\cdot]$ 等价于最小化式(2)中的均方误差 $\mathcal{L}_{\mathrm{AE}}$。KL 散度项 $D_{\mathrm{KL}}(\cdot)$ 促使潜空间分布逼近标准高斯先验。因此,通过这种机制,编码器不仅在工程上滤除了跨通道的冗余维度,更在数学上迫使潜变量流形坍缩至良态分布,为后续扩散阶段提供了理论可解的基础空间。在训练策略上,SpecLDM-TS 采用分阶段优化方式,首先对自编码器进行独立预训练以获得稳定的潜空间表示,随后在扩散模型训练阶段冻结编码器参数,以避免潜空间分布在训练过程中发生漂移,从而显著提升整体模型的收敛稳定性与生成一致性。

2.2 潜空间扩散模型

在获得潜变量表示后,SpecLDM-TS 在潜空间中构建高斯扩散模型进行生成建模。扩散过程包括前向加噪与反向去噪两个阶段。在前向过程中,逐步添加高斯噪声将潜变量扰动至标准正态分布 $q(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}) = \mathcal{N}(\boldsymbol{z}_t; \sqrt{1-\beta_t}\boldsymbol{z}_{t-1}, \beta_t \boldsymbol{I})$,式中 $\boldsymbol{z}_t$ 表示第 t 步的潜变量;β_t 表示噪声调度系数;$\mathcal{N}$ 表示高斯分布;$\boldsymbol{I}$ 表示单位矩阵。逐步加噪的过程中,每一步都在保留一部分原有结构信息的同时,引入一定强度的高斯噪声,最终把潜变量逐步扰动到接近标准正态分布。从而为后续反向生成过程构造一个可学习的噪声演化路径。

在反向过程中,具体的反向条件分布形式为 $p_\theta(\boldsymbol{z}_{t-1} \mid \boldsymbol{z}_t) = \mathcal{N}(\boldsymbol{z}_{t-1}; \varepsilon_\theta(\boldsymbol{z}_t, t), \sum_\theta(\boldsymbol{z}_t, t))$,式中 $\varepsilon_\theta(\boldsymbol{z}_t, t)$ 表示神经网络预测的噪声;$\sum_\theta(\boldsymbol{z}_t, t)$ 表示协方差。通过该反向过程可逐步将高斯噪声映射回数据分布,实现从噪声到数据的生成过程。训练神经网络(UNet)预测噪声,从而逐步恢复数据结构。对应的优化目标为

$$\mathcal{L}_{\mathrm{diff}} = \mathbb{E}_{\boldsymbol{z},t,\varepsilon}[\|\varepsilon - \varepsilon_\theta(\boldsymbol{z}_t, t)\|^2], \tag{3}$$

式中 ε 表示标准高斯噪声,$\mathbb{E}_{\boldsymbol{z},t,\varepsilon}[\cdot]$ 表示对潜变量、时间步和噪声的期望。使训练网络准确估计每一步噪声,从而让模型学会如何从带噪潜变量中逐步恢复出原始数据结构。

去噪网络采用一维 UNet 结构,通过编码-解码路径与跳跃连接融合多尺度特征。同时,为弥补卷积结构在全局建模能力上的不足,SpecLDM-TS 在瓶颈层引入多头自注意力机制,从而增强模型对长距离时间依赖与周期模式的建模能力。

进一步地,为解决扩散模型在时序生成中可能出现的频谱失真与模式崩溃问题,SpecLDM-TS 引入频域约束损失 $\mathcal{L}_{\mathrm{freq}} = \|F(\boldsymbol{x}) - F(\hat{\boldsymbol{x}})\|^2$,式中 $\mathcal{F}(\cdot)$ 表示傅里叶变换。通过潜空间映射与自注意力机制,模型有效捕获了低维空间的长程依赖。然而,我们发现若仅依赖传统的时域 MSE 损失进行去噪,网络仍会受限于深度学习的频率偏置现象(优先拟合低频),这是引发生成波形过度平滑与频域特征失真的根本机制。

为弥补单一时域优化的这一致命不足,本文引入了上述频谱约束。根据数字信号处理中的帕塞瓦尔定理(Parseval's Theorem),对于时域误差残差 $\boldsymbol{e}_n = \boldsymbol{x}_n - \hat{\boldsymbol{x}}_n$ 及其傅里叶变换 $F_e(k)$,其能量满足等价关系 $\sum_{n=1}^{N} |\boldsymbol{e}_n|^2 = \frac{1}{N}\sum_{k=1}^{N} |F_e(k)|^2$。该定理表明,显式引入式(3)所示的频谱幅值损失,本质上是在损失泛函中增加了一个关于功率谱密度(PSD)的全局正交约束。该约束在频域积分尺度上强制生成样本的高频分量与真实样本对齐,从物理机制上阻断了高频能量在去噪迭代中的流失,以此缓解频域特征的失真问题。潜空间扩散模型的最终训练目标为 $\mathcal{L} = \mathcal{L}_{\mathrm{diff}} + \lambda \mathcal{L}_{\mathrm{freq}}$,式中 λ 表示权重系数,SpecLDM-TS 中设置为 0.1。该联合损失函数从时间域与频率域两个角度约束模型学习,从而显著提升生成序列的周期一致性与结构稳定性。

在推理阶段,采用去噪扩散隐式模型(DDIM)进行采样,以显著减少生成所需时间步数,同时保持较高的生成质量。在不改变训练目标函数的前提下,DDIM 通过对反向过程进行重新参数化,将原本基于马尔可

夫链的随机采样过程转化为确定性或半确定性的生成路径，从而显著减少采样步数并提高生成效率。

DDIM 首先根据当前时刻的潜变量估计原始潜变量 $\hat{\boldsymbol{z}}_0=\frac{\boldsymbol{z}_t-\sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\boldsymbol{z}_t,t)}{\sqrt{\bar{\alpha}_t}}$，式中 $\boldsymbol{z}_t$ 表示当前时刻潜变量；$\hat{\boldsymbol{z}}_0$ 表示估计的原始潜变量；$\bar{\alpha}_t$ 为噪声调度；t 表示给定的时间步。在此基础上，DDIM 的反向更新过程可表示为

$$\boldsymbol{z}_{t-1}=\sqrt{\bar{\alpha}_{t-1}}\hat{\boldsymbol{z}}_0+\sqrt{1-\bar{\alpha}_{t-1}}\epsilon_\theta(\boldsymbol{z}_t,t)。\tag{4}$$

该方法通过跳步采样策略可在较少时间步内完成从噪声到潜变量的映射，降低推理阶段的计算开销；尽管上述时频联合约束从机制上保证了生成质量，但传统的马尔可夫去噪过程仍面临推理极其耗时的效率瓶颈。为打破这一限制以满足大规模合成需求，本文采用式(4)的确定性跳步机制，其稳定性建立在连续时间概率流常微分方程(Probability Flow ODE)的基础之上。潜变量随时间 $\tau\in[0,1]$ 的反向演化轨迹可连续化表示为 $\frac{\mathrm{d}\bar{\boldsymbol{z}}_\tau}{\mathrm{d}\tau}=-\frac{1}{2}\frac{\mathrm{d}\log\alpha_\tau}{\mathrm{d}\tau}(\bar{\boldsymbol{z}}_\tau-\epsilon_\theta(\bar{\boldsymbol{z}}_\tau,\tau))$，式中标度化潜变量 $\bar{\boldsymbol{z}}_\tau=\boldsymbol{z}_\tau/\sqrt{\alpha_\tau}$。式(4)本质上是对上述目标 ODE 的一阶隐式欧拉(Implicit Euler)离散化求解。由于在扩散阶段冻结了自编码器参数，压缩后的潜空间支撑集表现出良好的利普希茨连续性，确保了该常微分方程具备唯一解。这使得确定性生成路径不仅截断了累积误差，更能以有界的离散化误差收敛至目标边缘概率测度，从而在大幅提升推理速度的同时保持时序的长期依赖结构。由此提升整体生成性能。

2.3　后处理

在生成阶段之后，为进一步提升生成数据的质量与分布一致性，SpecLDM-TS 设计了一系列后处理与优化机制。首先对生成潜变量进行统计量校正，使其均值与方差与真实数据保持一致 $\boldsymbol{z}'=\frac{\boldsymbol{z}-\boldsymbol{\mu}_g}{\boldsymbol{\sigma}_g}\cdot\boldsymbol{\sigma}_r+\boldsymbol{\mu}_r$，式中 $(\boldsymbol{\mu}_g,\boldsymbol{\sigma}_g)$ 和 $(\boldsymbol{\mu}_r,\boldsymbol{\sigma}_r)$ 分别表示生成数据与真实数据的统计量，$\boldsymbol{\mu},\boldsymbol{\sigma}$ 分别代表了均值和标准差。使其与真实数据分布的一阶与二阶统计特性保持一致。这一过程可以视为对生成分布的后验对齐，可有效缓解扩散模型在有限训练条件下可能产生的整体偏移问题。其次，针对基于局部片段生成时序数据可能带来的边界不连续问题，采用重叠区域加权融合策略对相邻子序列进行平滑拼接。在重叠区域内引入加权机制，有助于削弱不同片段之间的突变，从而保证生成序列在时间维度上的连续性与平滑性。为减少边界效应与拼接误差，采用重叠区域加权融合策略对相邻子序列进行平滑拼接，从而提升生成序列的连续性与平滑性。

通过上述后处理与优化步骤，生成数据在统计特性、频谱结构以及时间连续性等方面均能够较好地逼近真实数据分布。

3　实验

为了验证提出的基于“编码-潜空间生成-解码”三阶段扩散模型在多变量时序数据生成任务中的有效性，我们在多个真实的多元时间序列数据集上进行了全面的实验评估。

3.1　实验设置

3.1.1　*数据集介绍*。实验采用了 6 个广泛应用的公开多元时间序列数据集，具体包括：Electricity、Solar、Wind、Traffic、NN5 以及 Exchange。这些数据集涵盖了能源消耗、气象环境、交通流量以及宏观经济等多个领域的复杂动态规律，能够全面评估模型捕捉不同特征和长周期依赖关系的能力。以 Electricity 数据集为例，其记录了 321 个独立用户的连续电力消耗数据。数据的采样频率固定，为了评估模型对周期性规律的学习能力，本文把序列长度统一设定为 168 个时间步，即代表一周的连续行为。在数据划分方面，按照时间先后顺序把数据划分为训练集与测试集，训练集用于 Autoencoder 和扩散模型的联合训练，测试集用于提取真实数据的统计分布并与生成数据进行对比。

此外，为降低随机初始化和训练过程波动对实验结果的影响，本文对主要定量评估结果采用了 5 次独立重复实验后的平均值进行报告。具体而言，在保持数据划分、网络结构和训练流程一致的条件下，分别使用随机种子 42、123、456、789 和 1 024 独立运行实验，并对 MMD 与 KL 散度结果取平均，以更客观地反映模

型的整体性能。除定量结果外,本文还结合生成序列的时域、频域与周期性可视化分析,以及消融实验和超参数敏感性分析,从多个角度验证所提方法的有效性与稳定性。

3.1.2 *数据预处理策略*。针对多元时序数据中存在的量纲不一致以及直流分量干扰等问题,文章设计了多级数据预处理流水线。由于不同通道,例如不同用户或传感器的基线存在显著差异,对所有通道进行独立的最大最小值归一化,使数据严格约束在[0,1]范围内。此举有效避免了极端大值通道在计算损失函数时主导梯度更新。真实时序数据的频谱图中通常在 0 Hz 处存在较大的直流分量,即全局均值。为了使模型更加专注于学习交流分量,对切分后的单条样本进行去均值操作,去均值后的波形严格以 0 为中心,显著降低了模型拟合静态基线的难度。为降低长序列输入带来的计算复杂度并增强局部特征提取能力,将长序列切分为多个固定长度为 16 的 Patch,滑动步长设为 8,即相邻 Patch 之间存在 50%的重叠。这种重叠切分机制不仅保留了时间连续性,还有效消除了重构过程中的边界伪影。

3.1.3 *实验环境与超参数配置*。实验的整体架构分为 Autoencoder 训练与扩散模型训练两个解耦的阶段。编码器与解码器由多层 1D 卷积网络构成。以 Electricity 数据集为例,输入通道数为 321,模型将其特征压缩并映射至维度为 32 的低维潜空间中。采用 Adam 优化器,学习率设定为 10^{-3},批大小为 16,共训练 100 个 epochs。扩散网络主体采用融合了多头自注意力机制的 1D UNet。在前向扩散过程中,总扩散步数 T 定为 1 000。反向去噪过程同样采用 Adam 优化器,学习率为 10^{-4},批大小为 16,训练轮数为 500 epochs。

在推理阶段,为了加速生成过程,我们引入了 DDIM 采样器,将实际采样步数从 1 000 步大幅压缩至 200 步。采样温度参数设为 1.0。

3.1.4 *基线模型*。为全面且客观地评估 SpecLDM-TS 的多变量时序生成性能,实验选取了 5 种当前主流的深度生成模型作为对比基线。(1) Diffusion-TS[13]:该方法直接在原始数据空间中执行去噪过程。由于其不具备潜空间压缩机制,它被作为验证本文潜空间扩散在处理高并发、高维序列时计算效率与性能增益的核心参照。(2) TimeGAN[31]:作为时序生成的基准模型,它通过联合优化有监督和无监督损失,在嵌入空间中捕获时序特征。(3) GT-GAN[32]:一种融合图神经网络的先进对抗架构,专门针对多变量时序中复杂的变量间空间拓扑和通道耦合关系进行建模。(4) TimeVAE[33]:针对时序数据的趋势与季节性分解设计的 VAE 架构。其具备较好的训练稳定性,但在生成样本的高频保真度和多样性方面通常被视为一种平滑后的对比基准。(5) TimeVQVAE[34]:引入矢量量化机制,利用离散编码本增强对高维复杂特征的表达能力,是评估模型重构细节能力的关键参照。

3.1.5 *评估指标体系*。鉴于时间序列生成任务的特殊性,传统的点对点误差指标往往无法客观反映生成数据在整体概率分布上的质量。因此选取了最大均值差异 MMD 和 Kullback-Leibler 散度作为核心定量评估指标,以严格衡量真实数据分布与生成数据分布之间的相似性。最大均值差异是一种基于再生核希尔伯特空间的非参数统计检验方法,能够有效衡量两个高维分布之间的距离。本文采用径向基函数作为核函数,在具体计算上,该指标无需直接估计概率密度,而是通过分别计算真实样本集合内部的平均核相似度、生成样本集合内部的平均核相似度,并减去真实样本与生成样本之间的交叉平均核相似度,来综合量化两个分布在高维特征空间中的均值距离。MMD 值越接近 0,表明生成数据与真实数据在再生核空间中的特征映射越一致,分布拟合效果越好。KL 散度用于衡量生成分布近似真实分布时的信息损失量。在多元时序数据的统计评估中,本文提取特征空间的边缘分布,计算其相对熵。具体而言,该指标通过对真实数据的概率密度与生成数据概率密度比值的对数进行期望求值,来直观反映两个概率分布在整个样本空间上的差异程度。较低的 KL 散度说明扩散模型在反向去噪过程中成功还原了真实时序数据的底层概率密度函数。

为了避免仅依赖 MMD 和 KL 散度等分布指标而难以全面反映生成序列的动态结构保真性,本文进一步引入时域波形、频域幅值谱以及自相关函数三类可视化分析。具体而言,时域波形对比用于观察生成序列对局部波动模式、峰谷位置及整体变化趋势的保持能力;频域幅值谱对比用于检验生成样本在主频位置与能量分布上的一致性;自相关函数对比则用于衡量模型对周期结构及长短期依赖关系的保留程度。通过"定量指标+定性可视化"的联合评估方式,可以更全面地验证所提方法在多变量时序生成任务中的真实性与稳定性。

3.2 实验结果

为了系统评估本文提出的基于"编码-潜空间生成-解码"三阶段扩散模型 SpecLDM-TS 在多变量时序数

据生成任务中的有效性与泛化能力，本节将其与当前主流的 5 种深度生成模型 Diffusion-TS、TimeGAN、GT-GAN、TimeVAE、TimeVQVAE 进行了对比实验。我们将基于 DDIM 采样的生成结果与原始测试集进行统计对比，并分别从最大均值差异 MMD 和 Kullback-Leibler 散度两个核心维度展开定量评估。需要说明的是，表 1 和表 2 中的结果均为 5 次独立重复实验后的平均值，以尽可能降低单次实验随机性的影响。加粗数值代表最优结果，下划线数值代表次优结果。

3.2.1　*最大均值差异评估*。最大均值差异侧重于评估生成数据与真实数据在高维特征空间中的整体分布相似度，其值越小表明分布拟合效果越好。表 1 展示了不同生成模型在 6 个基准数据集上的 MMD 平均评估结果。

表 1　不同生成模型在各基准数据集上的最大均值差异评估结果

Tab. 1　Maximum mean discrepancy evaluation results of different generative models on various benchmark datasets

	Diffusion-TS	TimeGAN	GT-GAN	TimeVAE	TimeVQVAE	SpecLDM-TS
Electricity	**0.003**	0.367	0.254	0.577	0.152	<u>0.018</u>
Solar	<u>0.062</u>	0.628	0.578	0.353	0.437	**0.061**
Wind	<u>0.035</u>	0.213	0.170	0.170	0.131	**0.024**
Traffic	<u>0.111</u>	0.567	0.538	0.218	0.213	**0.068**
NN5	**0.186**	0.874	0.632	0.821	0.335	<u>0.331</u>
Exchange	<u>0.239</u>	0.530	0.475	0.543	0.342	**0.236**

从表 1 可以看出，SpecLDM-TS 在绝大多数基准测试场景中均展现了卓越的分布匹配能力。具体而言，在 Solar、Wind、Traffic 以及 Exchange 四个多变量耦合关系极其复杂的数据集上，SpecLDM-TS 的 MMD 指标均取得全场最优；在 Electricity 和 NN5 数据集上，虽以微弱差距落后于直接在数据空间建模的 Diffusion-TS 位列次优，但其总体误差依然保持在极低水平，峰值仅为 0.331，远低于 TimeGAN 等传统对抗类基线模型的平均误差。

值得注意的是，SpecLDM-TS 在包含高度非线性动态的 Traffic 数据集上优势尤为显著。面对传统对抗生成模型，如 TimeGAN 和 GT-GAN 显著的拟合劣势，5 种基线模型在该数据集上的 MMD 平均值高达 0.329，而 SpecLDM-TS 将其大幅降低至 0.068。这不仅较基线平均水平提升了约 79.3%，更比表现最好的基线模型进一步降低了 38.7%。上述定量结果有力地证明：通过 1D UNet 引入的潜空间桥接机制，模型有效避免了在高维原始空间直接去噪带来的训练不稳定问题，从而能够在压缩后的低维潜空间中精准捕获高维多元联合分布。

3.2.2　*Kullback-Leibler 散度评估*。KL 散度用于衡量生成分布近似真实分布时的信息损失量，是评估模型是否捕捉到真实概率密度函数且未发生模式崩溃的核心指标。表 2 展示了各模型在 KL 散度指标上的平均评估结果。

表 2　不同生成模型在各基准数据集上的 KL 散度评估结果

Tab. 2　KL divergence evaluation results of different generative models on various benchmark datasets

	Diffusion-TS	TimeGAN	GT-GAN	TimeVAE	TimeVQVAE	SpecLDM-TS
Electricity	0.315	0.488	0.407	0.734	<u>0.280</u>	**0.142**
Solar	<u>0.066</u>	0.612	0.120	0.260	0.865	**0.038**
Wind	0.548	1.924	<u>0.107</u>	0.484	0.483	**0.008**
Traffic	<u>0.120</u>	1.305	1.409	0.211	0.178	**0.009**
NN5	<u>0.165</u>	4.928	1.386	1.337	1.063	**0.018**
Exchange	<u>2.337</u>	8.861	7.201	10.404	5.052	**0.902**

由表 2 可知，SpecLDM-TS 展现出了显著的领先优势，不仅在全部 6 个基准数据集上取得全场最优，且相较于次优模型实现了倍数级的性能提升。特别是在高频波动的 Wind 数据集和具有长周期依赖的 Traffic 数据集中，SpecLDM-TS 的 KL 散度分别降至极低的 0.008 和 0.009，远优于其他基线模型。即便在处理缺乏明显季节性且噪声剧烈的 Exchange 宏观经济数据集时，面对 5 种基线模型出现的严重性能衰退，基线平

均 KL 散度高达 6.771，TimeVAE 甚至偏移至 10.404，SpecLDM-TS 仍能将 KL 散度稳定控制在 0.902，相较基线平均水平实现了约 86.6%的误差缩减。这一系列倍数级的性能跃升充分证明：SpecLDM-TS 在反向去噪过程中不仅拟合了数据的表层统计特征，更成功还原了真实时序数据的底层概率密度函数，在保持生成样本高多样性的同时，从根本上规避了 GAN 模型常见的模式崩溃与 VAE 模型的过度平滑现象。

为了更加直观地展现 SpecLDM-TS 在不同数据环境下的综合拟合稳定性，我们将模型在 6 个基准数据集上的 MMD 与 KL 散度评估结果进行了可视化对比，如图 3 所示。

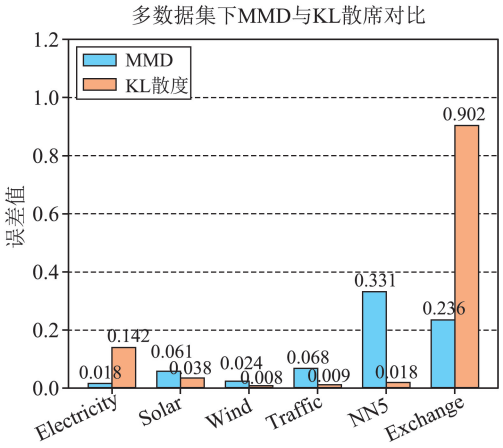


图 3 SpecLDM-TS 在各基准数据集上的 MMD 与 KL 散度综合对比

Fig. 3 Comprehensive comparison of MMD and KL divergence for SpecLDM-TS on various benchmark datasets

值得注意的是，在 NN5 和 Exchange 两个高难度宏观经济类数据集中，由于数据本身蕴含着强烈的非平稳性和剧烈波动，误差指标出现了一定程度的浮动，如 Exchange 的 KL 散度达到了 0.902，NN5 的 MMD 达到了 0.331。但正如前文定量的横向对比所示，这一误差水平依然大幅优于现有的其他基线模型，且 MMD 均严格控制在 0.8 之内。

3.2.3 生成序列的时域、频域与周期性可视化分析。仅通过 MMD 和 KL 散度等统计指标，虽然可以衡量生成分布与真实分布之间的整体差异，但仍难以直观反映生成样本在局部动态、频率结构以及周期性方面的保真程度。为此，本文进一步从时域、频域和周期性三个角度，对真实序列与 SpecLDM-TS 生成序列进行了可视化对比分析。

如图 4 所示，从时域波形角度来看，SpecLDM-TS 生成样本在整体变化趋势、局部波动幅值以及主要峰谷位置上均与真实序列较为接近。这说明所提方法不仅能够在统计意义上逼近真实分布，还能够较好地保留时间序列中的局部动态结构。相较于仅依赖时域点对点误差的生成方式，本文方法通过潜空间扩散建模与频谱约束的协同作用，有效减弱了波形过度平滑和局部振荡失真的现象。

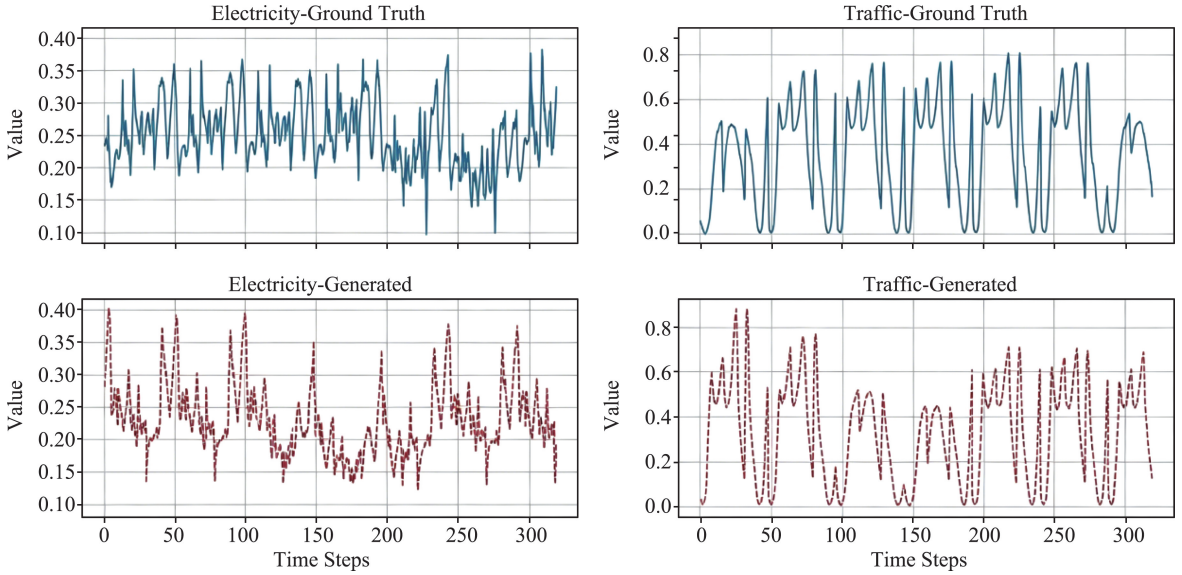


图 4 真实序列与生成序列的时域波形对比

Fig. 4 Time-domain waveform comparison between real and generated sequences

如图 5 所示，从频域结构来看，生成序列在主频位置上与真实数据基本保持一致，且主要频段上的能量分布具有较高相似性。这表明所引入的频谱损失能够有效约束扩散模型在反向生成过程中保留原始序列的频率结构，尤其有助于降低高频成分丢失和周期信息偏移的问题。对于具有明显周期性或多尺度振荡特征

的多变量时序数据而言，这种频域一致性是保证生成样本真实性的重要条件。

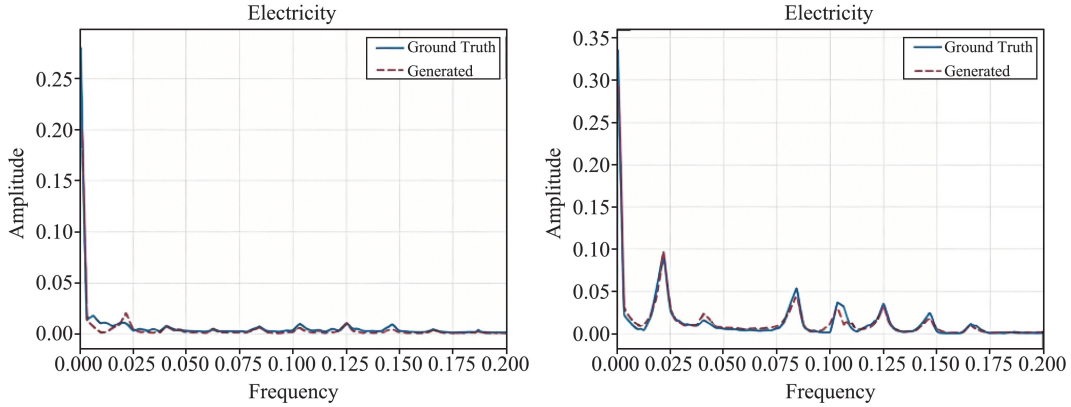


图 5　真实序列与生成序列的频域幅值谱对比

Fig. 5　Frequency-domain amplitude spectrum comparison between real and generated sequences

如图 6 所示，从周期性与长程依赖角度来看，生成序列在主要自相关峰的位置及整体衰减趋势上与真实序列较为一致，说明模型较好地恢复了原始数据中的周期重复模式与时序依赖关系。特别是在较大滞后步长下，生成样本仍能维持与真实样本相近的相关结构，这进一步验证了所提方法在周期模式建模方面的有效性。

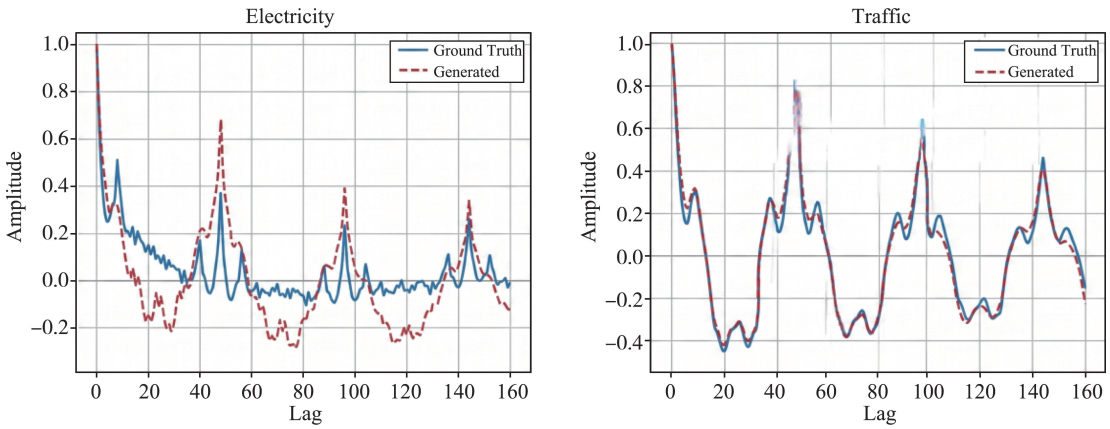


图 6　真实序列与生成序列的自相关函数对比

Fig. 6　Autocorrelation function comparison between real and generated sequences

综合图 3～5 可以看出，SpecLDM-TS 不仅在 MMD 和 KL 散度等定量指标上取得了较优结果，而且在时域波形、频域主频结构以及周期性依赖模式等方面均表现出较好的保真能力。这说明所提方法能够较为完整地学习真实多变量时序数据的动态演化规律，为后续数据增强与高保真样本合成提供了可靠支持。

3.3　消融实验

为了进一步分析 SpecLDM-TS 各组成模块对生成性能的具体贡献，本文在代表性数据集 Rain 上设计了消融实验，对模型中的自注意力机制、潜空间压缩结构以及统计校正模块分别进行移除，并与完整模型进行对比。实验结果如表 3 所示。各变体模型除被移除模块外，其余训练设置与完整模型保持一致，从而保证比较结果的公平性。

表 3　SpecLDM-TS 各关键模块的消融实验结果

Tab. 3　Ablation results of key components in SpecLDM-TS

方法	MSE	MAE	自相关相似度	MMD
完整模型	1.728 1	0.140 9	0.597 3	0.196 9
不使用自注意力	1.728 8	0.141 0	0.571 2	0.218 9
不使用潜空间压缩	1.747 1	0.190 1	0.575 5	0.200 5
不使用统计校正	1.728 9	0.141 3	0.578 9	0.240 7

由表 3 可以看出,完整模型在总体误差、周期结构保持能力和分布对齐之间表现出较为均衡的性能,说明本文构建的“潜空间扩散 + 频谱约束 + 加速采样 + 统计校正”框架整体上是有效的。在移除自注意力机制后,模型的 MSE、MAE 和 MMD 变化较小,但自相关相似度由 0.597 3 下降至 0.571 2。这表明自注意力机制对局部重构误差的影响较弱,但对长程依赖关系和周期结构建模仍具有一定积极作用。在不使用潜空间压缩时,模型的 MSE 由 1.728 1 上升至 1.747 1,MAE 由 0.140 9 上升至 0.190 1,说明直接在原始高维空间中建模会使点对点重构误差恶化。尽管该设置下 MMD 和自相关相似度出现一定波动,但综合误差结果来看,潜空间压缩仍有助于提升模型的紧凑性和表示能力。在移除统计校正模块后,模型的 MSE、MAE、MMD 和自相关相似度与完整模型相比变化均较小。该结果表明,在 Rain 数据集当前设置下,统计校正模块对最终分布对齐的影响相对有限,其贡献弱于潜空间压缩模块。综合来看,消融实验表明,不同模块对模型的作用并不完全相同:自注意力机制主要改善周期结构建模,潜空间压缩主要降低重构误差,而统计校正模块在当前数据集上的增益相对有限。上述结果进一步说明,SpecLDM-TS 的性能提升来自多个组成模块的协同作用,而不仅仅依赖单一损失项。

3.4 超参数敏感性分析

为进一步分析模型在不同参数配置下的性能变化规律,本文在代表性数据集 Exchange 上对频谱损失权重、潜空间维度以及 DDIM 采样步数进行了敏感性分析。通过比较不同参数设置下模型在重构误差、自相关相似度以及推理开销等指标上的变化,可以更清晰地揭示 SpecLDM-TS 在生成质量与计算效率之间的权衡关系。不同参数设置的综合对比结果如表 4 所示。

表 4 Exchange 数据集上不同超参数的敏感性分析结果

Tab. 4 Sensitivity analysis of different hyperparameters on the exchange dataset

参数类型	参数取值	MSE	MAE	自相关相似度	推理时间(s)
频谱损失权重 λ	0.1	5.099 8	0.504 6	0.874 9	4.33
	0.2	5.036 4	0.503 7	0.846 3	9.69
	0.3	5.132 8	0.513 5	0.854 3	9.73
潜空间维度	16	5.063 5	0.485 6	0.870 5	5.15
	32	5.099 8	0.504 6	0.874 9	4.19
	64	5.156 4	0.540 0	0.908 0	5.00
DDIM 采样步数	100	5.099 7	0.504 1	0.875 0	7.13
	200	5.109 8	0.504 7	0.874 4	12.49
	500	5.045 0	0.508 7	0.864 5	33.25

3.4.1 *频谱损失权重分析*。频谱损失权重λ 用于调节时域重构目标与频域约束目标之间的相对强度。较小的λ 难以充分引入频域约束,而过大的λ 则可能削弱模型对整体重构误差的优化效果。由表 4 第一部分可以看出,不同频谱损失权重会对模型性能产生明显影响。$\lambda=0.2$ 时取得了最优的 MSE 和 MAE 结果,说明适度增强频域约束有助于改善重构误差。

当λ 从 0.1 增加至 0.2 时,模型误差有所下降;但当λ 继续增大至 0.3 后,MSE 和 MAE 均出现回升,说明过强的频谱损失约束并不一定带来持续增益。另一方面,自相关相似度在$\lambda=0.1$ 时略高于其他设置,表明频域约束增强后,不同评价指标之间存在一定权衡关系。综合来看,$\lambda=0.2$ 在当前实验中表现出更均衡的综合性能,因此可视为较合理的默认设置。

3.4.2 *潜空间维度分析*。潜空间维度决定了自编码器对原始序列的压缩程度,同时影响扩散模型在潜空间中的表示容量。由表 4 第二部分可以看出,潜空间维度对不同指标的影响并不完全一致。16 维取得了最小的 MSE 和 MAE,32 维取得了最短的推理时间,而 64 维在自相关相似度上表现更优。这说明潜空间维度并不存在对所有指标统一适用的绝对最优值。

从综合表现来看,32 维在重构效果与推理效率之间表现出更好的平衡。相较于 64 维,32 维的 MSE、MAE 与推理时间更有优势;相较于 16 维,32 维的推理效率占优,且自相关相似度略好。因此,采用 32 维作

为统一配置更强调多指标均衡和实验设置的一致性。

3.4.3　DDIM 采样步数分析。DDIM 采样步数决定了扩散模型反向去噪过程的迭代次数。由表 4 可以看出，不同 DDIM 采样步数下误差类和相关性指标的差异总体较小，而对推理时间的影响十分显著。当步数由 100 增加至 500 时，虽然 MSE 出现微弱下降，由 5.099 7 降至 5.045 0，但 MAE 略有上升，由 0.504 1 升至 0.508 7，且自相关相似度略有下降，说明在当前实验中盲目增加采样步数并未带来一致的质量提升。

相比之下，推理时间随着采样步数增加而显著上升：由 100 步的 7.13 s 增加至 200 步的 12.49 s，再增加至 500 步的 33.25 s。这表明 DDIM 步数更适合作为调节生成效率的工程参数，而不是决定生成质量的主导因素。在当前数据集上 100 步已能够取得与 200 步和 500 步相近的生成效果，同时显著降低推理开销，进一步验证了 DDIM 加速采样对于提升 SpecLDM-TS 实用性的价值。

4　结论

为解决现有扩散生成模型，在处理多变量时间序列数据时存在采样效率低与频域特征丢失的不足，本文提出了基于潜空间扩散模型的多变量时序数据生成方法 SpecLDM-TS 验证了融合 1D Autoencoder、集成多头自注意力机制的 1D UNet 以及频谱损失约束的混合生成框架，用于实现高维时序数据的高效、高保真建模。通过理论分析、对比实验与消融实验，证明了该模型结合自编码器的降维压缩能力、UNet 与自注意力的长距离依赖建模能力，以及频谱损失在保留周期性物理结构方面的全局正则化优势，有效形成了兼顾生成质量与计算效率的时序建模体系。实验结果表明，在 Electricity、Solar、Wind、Traffic 等多个真实多变量数据集上，SpecLDM-TS 模型在最大均值差异和 KL 散度等评估指标上均显著优于 TimeGAN、Diffusion-TS 等基准模型，显示出其在拟合底层概率密度函数及避免模式崩溃方面的显著优势。综上所述，SpecLDM-TS 通过潜空间映射、频域与时域联合约束以及基于 DDIM 的跳步采样与后处理校正机制，为复杂多变量时序数据的数据增强与高保真生成提供了一种新颖的解决方案。

然而，该模型仍存在一定的局限性，该模型在拟合具有强非平稳性与剧烈波动的宏观经济数据时，其生成分布与真实分布的误差仍有所增加。未来工作将尝试结合图神经网络构建多变量通道间的空间拓扑关联矩阵，以进一步提升模型在极端非平稳环境下的特征提取能力与鲁棒性。

参　考　文　献

[1] Aiello E, Valsesia D, Magli E. Fast inference in denoising diffusion models via MMD finetuning[EB/OL]. arXiv preprint arXiv: arXiv: 2301.07969v1.

[2] Wu H X, Xu J H, Wang J M, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting[C]// Proceedings of Neural Information Processing Systems, 2021,34:22419-22430.

[3] Wen Q S, Sun L, Yang F, et al. Time series data augmentation for deep learning: a survey[C]//Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence International Joint Conferences on Artificial Intelligence Organization, 2021:4653-4660.

[4] Brophy E, Wang Z W, She Q, et al. Generative adversarial networks in time series: a systematic literature review[J]. ACM Computing Surveys, 2023, 55(10): 1-31.

[5] Ho J, Jain A N, Abbeel P. Denoising diffusion probabilistic models[EB/OL]. arXiv preprint arXiv: arXiv:2006.11239.

[6] Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis[C]//Proceedings of the Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates, 2021.

[7] Yang L, Zhang Z L, Song Y, et al. Diffusion models: a comprehensive survey of methods and applications[J]. ACM Computing Surveys, 2024, 56(4): 1-39.

[8] Rasul K, Seward C, Schuster I, et al. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting[EB/OL]. arXiv preprint arXiv: arXiv:2101.12072.

[9] Tashiro Y, Song J M, Song Y, et al. CSDI: conditional score-based diffusion models for probabilistic time series imputation[EB/OL]. arXiv preprint arXiv: arXiv: 2107.03502

[10] 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, 2022:10674-10685.

[11] Qian J, Xie B Y, Wan B, et al. TimeLDM: latent diffusion model for unconditional time series generation[EB/OL]. arXiv preprint arXiv: 2407.04211.

[12] Song J M, Meng C L, Ermon S. Denoising diffusion implicit models[EB/OL]. arXiv preprint arXiv: arXiv: 2010.02502.

[13] Yuan X Y, Qiao Y. Diffusion-TS: interpretable diffusion for general time series generation[EB/OL]. arXiv preprint arXiv: arXiv: 2403.01742.

[14] Abbahaddou Y, Aboussalah A M. A geometry-aware metric for mode collapse in time series generative models[EB/OL]. arXiv preprint arXiv: arXiv: 2412.11292v1.

[15] Song Y, Sohl-Dickstein J, Kingma D P, et al. Score-based generative modeling through stochastic differential equations[EB/OL]. arXiv preprint arXiv: arXiv: 2011.13456.

[16] Lin L Q, Li Z Z, Li R K, et al. Diffusion models for time-series applications: a survey[J]. Frontiers of Information Technology & Electronic Engineering, 2024, 25(1): 19-41.

[17] Sordo Z, Chagnon E, Hu Z X, et al. Synthetic scientific image generation with VAE, GAN, and diffusion model architectures[J]. Journal of Imaging, 2025,11(8): 252.

[18] Zhou W T, Kang Z, Tian L, et al. Non-autoregressive diffusion-based temporal point processes for continuous-time long-term event prediction[J]. Expert Systems with Applications, 2025, 267: 126210.

[19] Kong Z, Ping W, Huang J J, et al. DiffWave: a versatile diffusion model for audio synthesis[EB/OL]. arXiv preprint arXiv: arXiv: 2009.09761.

[20] Feng S B, Miao C Y, Zhang Z, et al. Latent diffusion transformer for probabilistic time series forecasting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024,38(11): 11979-11987.

[21] Gao S H, Koker T, Queen O, et al. UniTS: a unified multi-task time series model[EB/OL]. arXiv preprint arXiv: arXiv:2403.00131v2.

[22] Fan W, Yi K, Ye H T, et al. Deep frequency derivative learning for non-stationary time series forecasting[EB/OL]. arXiv preprint arXiv: arXiv: 2407.00502.

[23] Takaki S, Nakashika T, Wang X, et al. STFT spectral loss for training a neural speech waveform model[C]//Proceedings of ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing, 2019: 7065-7069.

[24] Fons E, Sztrajman A, El-Laham Y, et al. LSCD: lomb-scargle conditioned diffusion for time series imputation [EB/OL]. arXiv preprint arXiv: arXiv: 2506.17039.

[25] Ye X Y, Sui M C, Zeng M S, et al. Frequency domain discrimination and double loss function optimization of image generation for GANs [J]. The Journal of Supercomputing, 2025, 82(1):3.

[26] Wang H, Pan L C, Chen Z C, et al. FreDF: learning to forecast in the frequency domain[C]//Proceedings of International Conference on Learning Representations, 2024.

[27] Ning N W, Feng Y T, Li Z X, et al. CM-TFD: channel mask-based time-frequency decoupling for multivariate time series forecasting [J]. Knowledge-Based Systems, 2026,342: 115864.

[28] Jeong S, Sohn J, Jeon J, et al. Frequency-conditioned diffusion models for time series generation[C]//Proceedings of the ACM International Conference on Information and Knowledge Management ACM, 2025: 1114-1123.

[29] Crabbe J, Huynh N, Stanczuk J, et al. Time series diffusion in the frequency domain[C]//Proceedings of the 41st International Conference on Machine Learning,2024,235:9407-9438.

[30] Yang Y Y, Jin M, Wen H M, et al. A survey on diffusion models for time series and spatio-temporal data[J]. ACM Computing Surveys, 2026, 58(8): 1-39.

[31] Yoon J, Jarrett D, Van Der Schaar M. Time-series generative adversarial networks[C]//Proceedings of Neural Information Processing Systems,2019.

[32] Jeon J, Kim J, Song H, et al. GT-GAN: general purpose time series synthesis with generative adversarial networks[J]// Advances in Neural Information Processing Systems, 2022,35:36999-37010.

[33] Desai A, Freeman C, Wang Z H, et al. TimeVAE: a variational auto-encoder for multivariate time series generation[EB/OL]. arXiv preprint arXiv: arXiv: 2111.08095.

[34] Lee D, Malacarne S, Aune E. Vector quantized time series generation with a bidirectional prior model[C]//Proceedings of International Conference on Artificial Intelligence and Statistics, 2023.

(责任编辑:刘庆松)