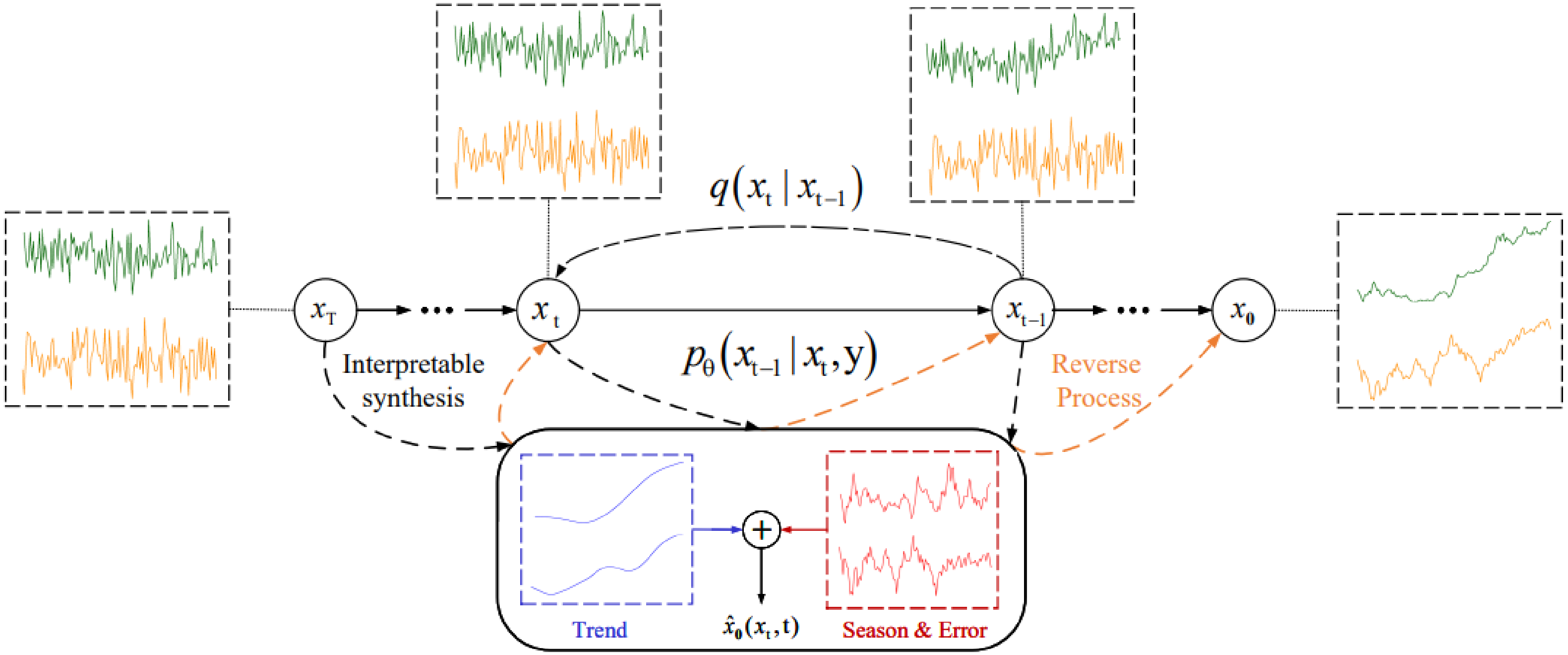
2024 ICLR

Diffusion-TS

Interpretable Diffusion for General Time Series Generation



2024 ICLR

Diffusion-TS

Interpretable Diffusion for General Time Series Generation

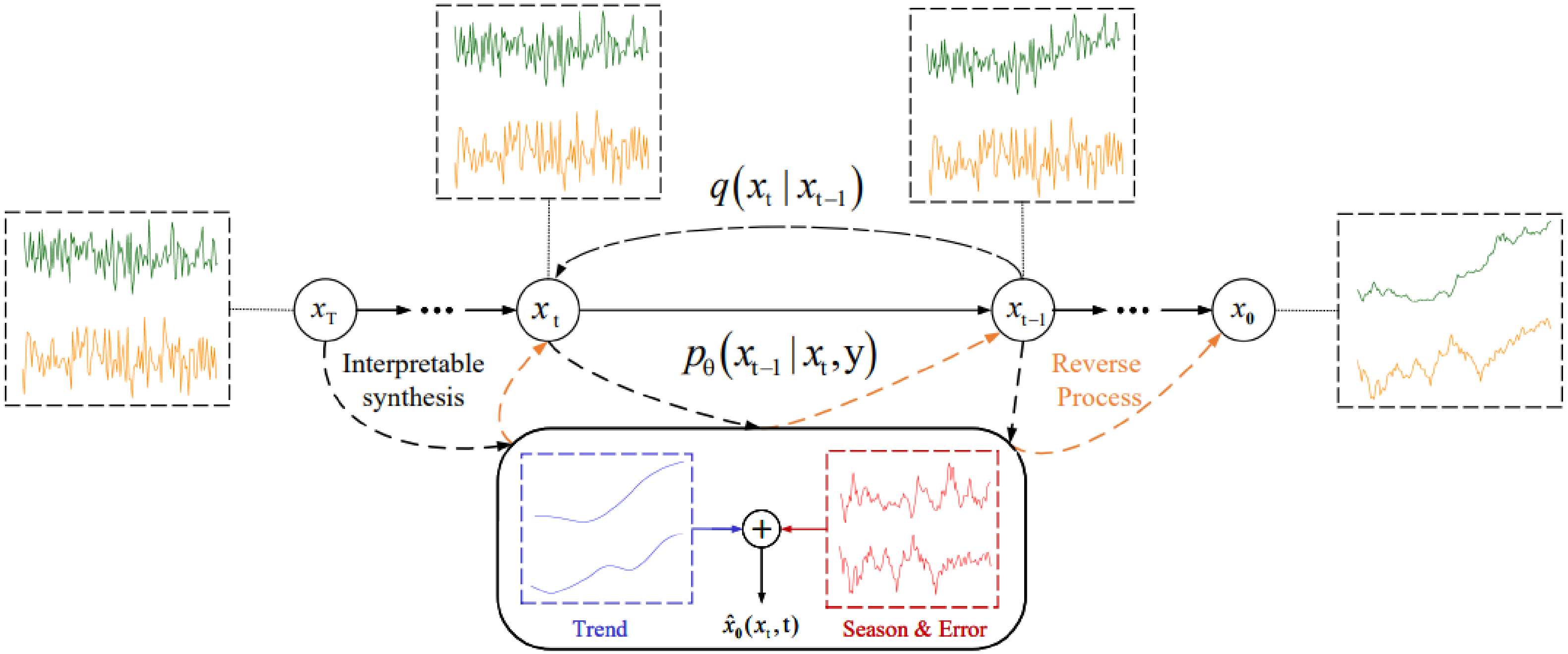
24.2.27

Presented by Yyyq

01

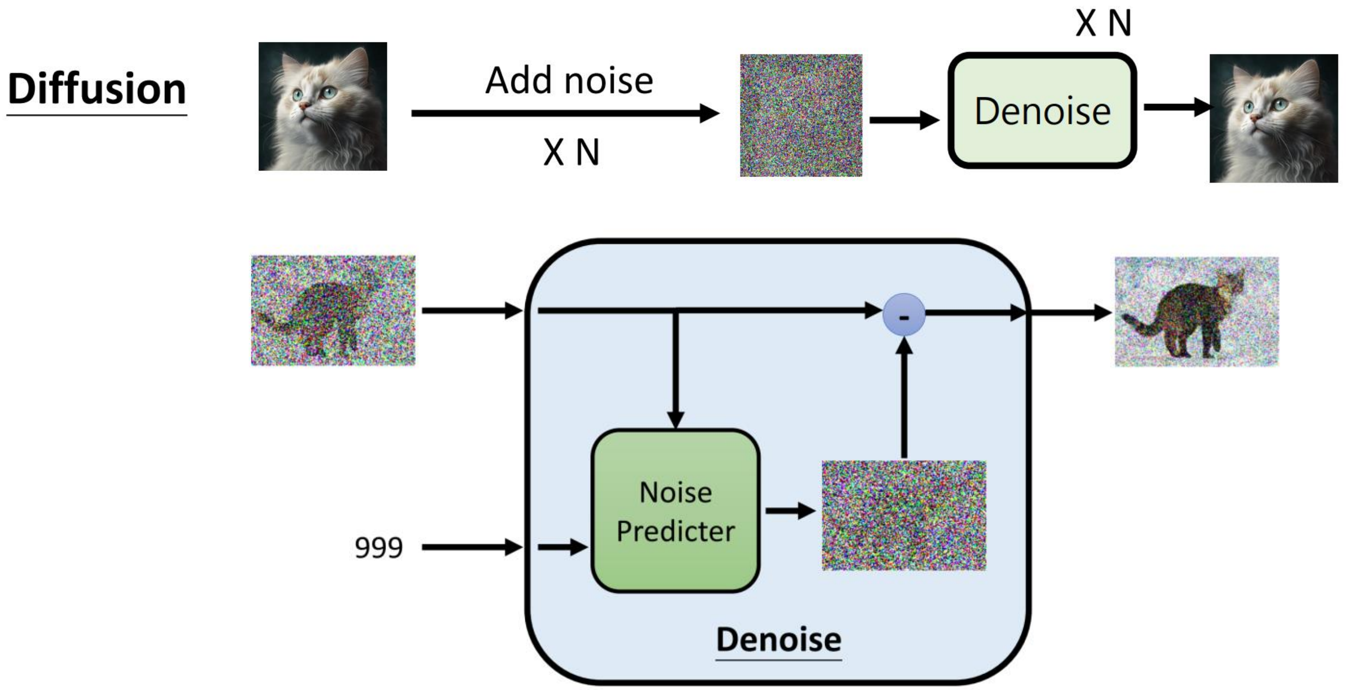
主要内容

- 扩散模型（Diffusion-TS）
- 时序生成（Time Series Generation） ➡ 条件生成，适应不同的生成任务（插补、预测）
- 可解释性（Interpretable Diffusion） ➡ 结合趋势分解，学习有意义的时间属性



02 创新点

- 在Diffusion Model（DDPM）框架中，重新设计<u>去噪网络</u>
- 去噪网络基于transformer（encoder-decoder），引入<u>可解释的分解结构</u>
- 训练一个无条件的模型，兼容不同的有条件任务



03 问题陈述

τ：时间步

N：不同的时序信号

T：扩散步

① $X_{1:\tau} = (x_1, \ldots, x_\tau) \in \mathbb{R}^{\tau \times d}, \quad D = \{X^i_{1:\tau}\}^N_{i=1}$

无条件生成：$\hat{X}^i_{1:\tau} = G(Z_i), \quad Z_i = (z^i_1, \ldots, z^i_t) \in \mathbb{R}^{\tau \times d \times T}$

function　　高斯向量

任务目标：通过 G 将 Z 映射到与 D 中的 X 最相似的信号

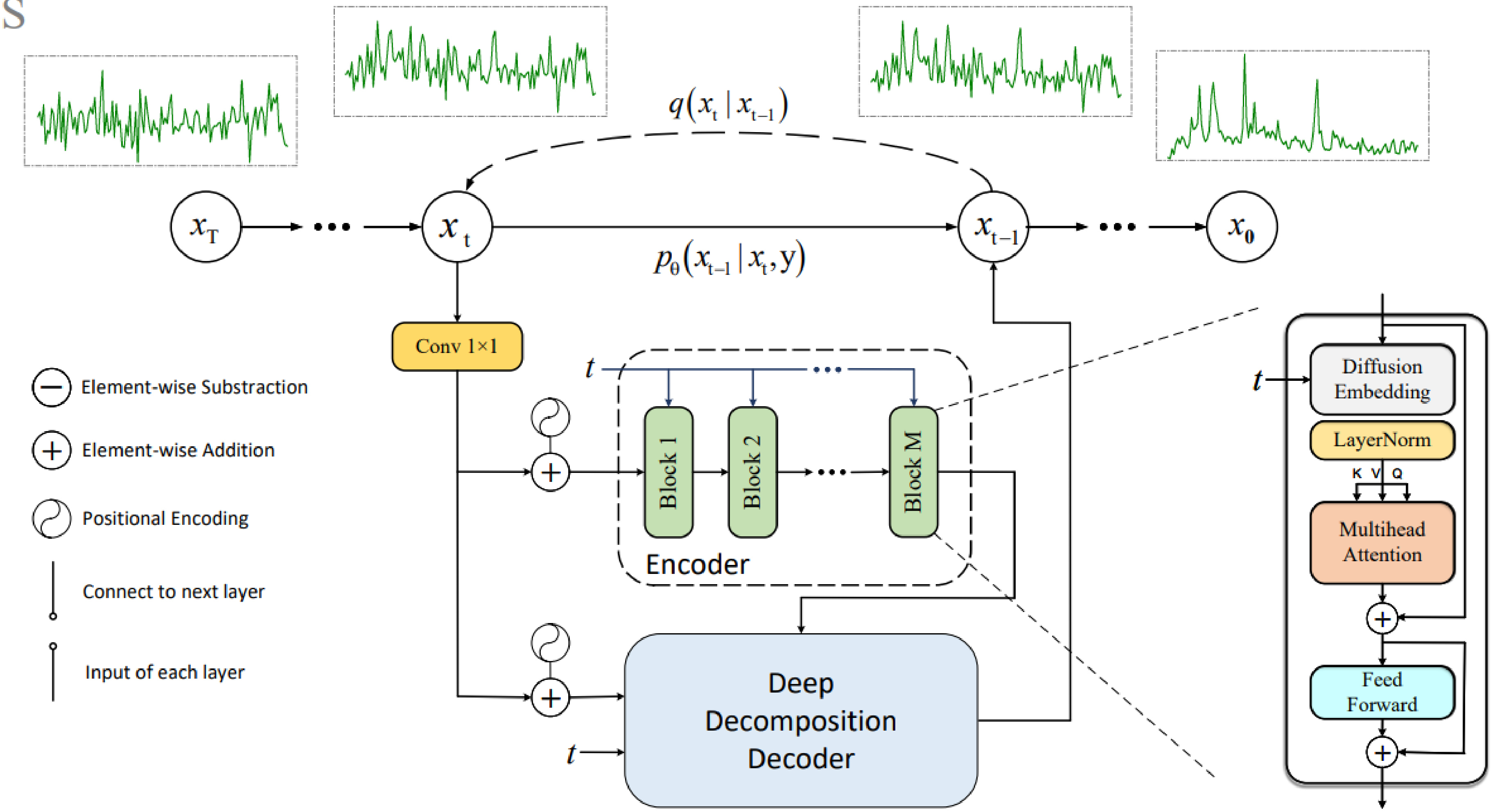
② $$x_j = \zeta_j + \sum_{i=1}^{m} s_{i,j} + e_j, \quad j = 0, 1, \ldots, \tau - 1,$$

时序 = 趋势分量 + 多季节分量 + 剩余分量

③ 条件生成：$p(.|y)$

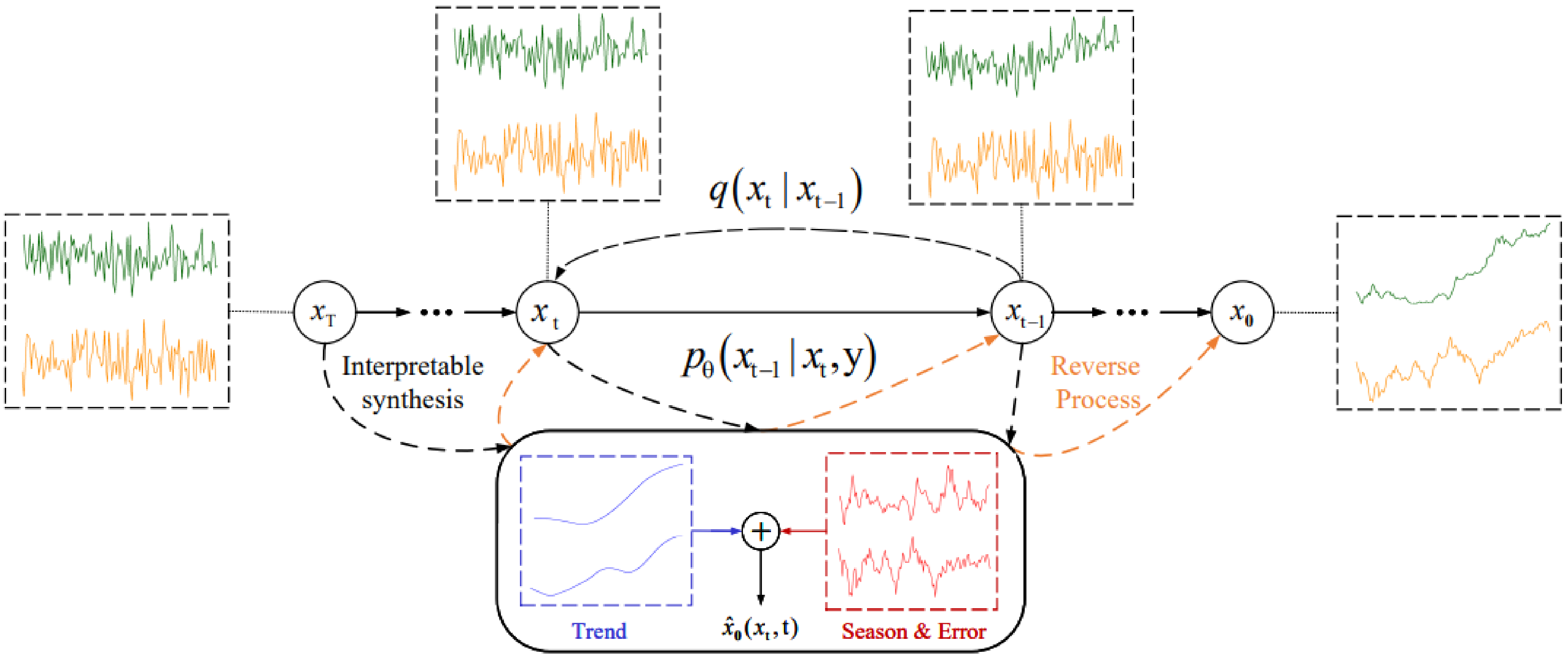
04 算法描述：整体结构

Diffusion-TS



04 算法描述：diffusion network

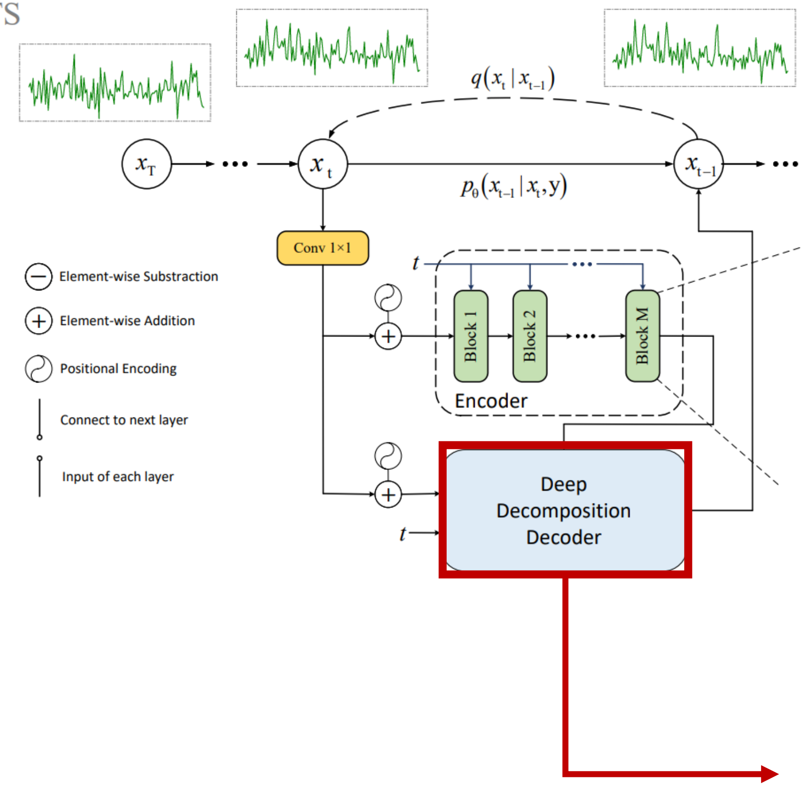
- 正向扩散：$x_0 \rightarrow x_T$，高斯噪声
- 反向去噪：clean x_T，简化为学习一个近似器 $\mu_\theta(x_t, t)$



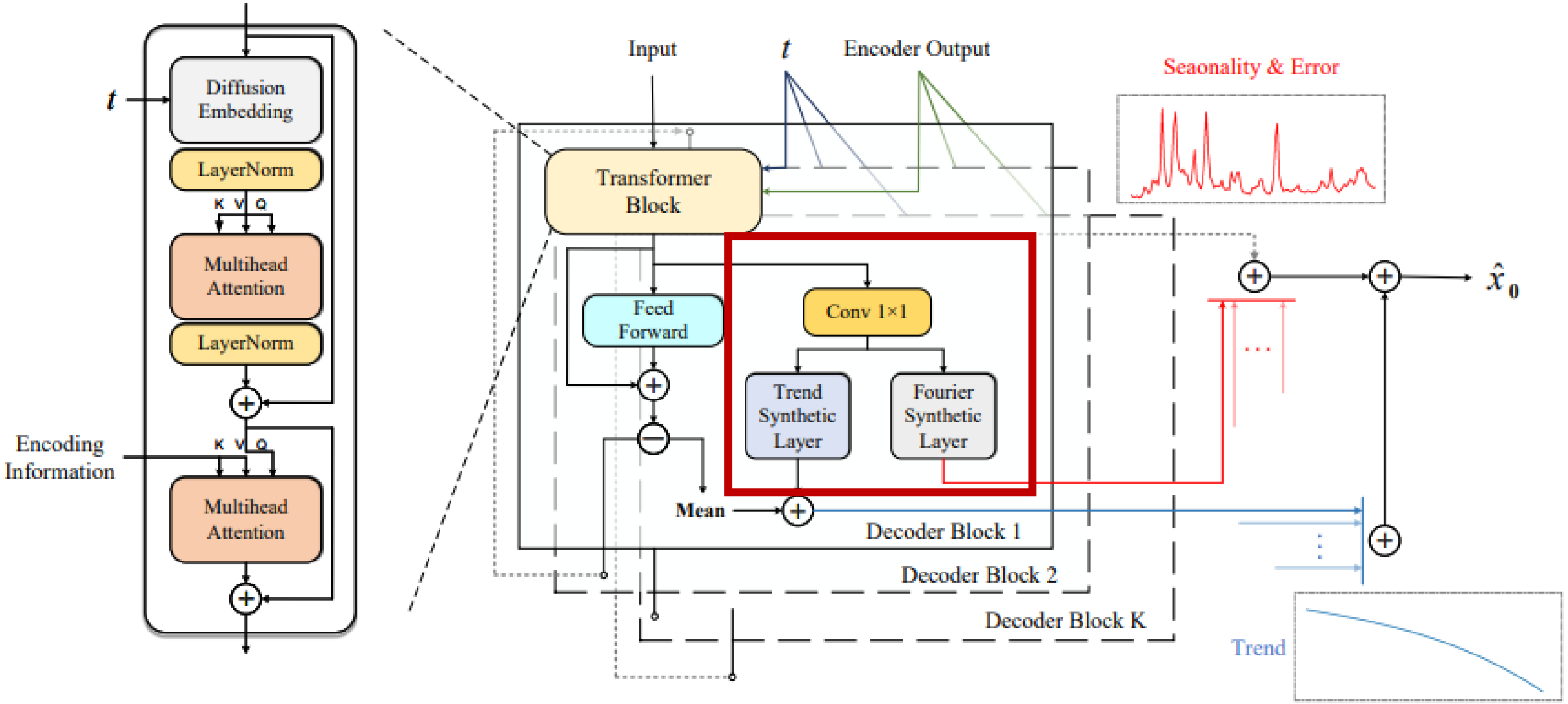
04

算法描述：deep decomposition decoder

Diffusion-TS

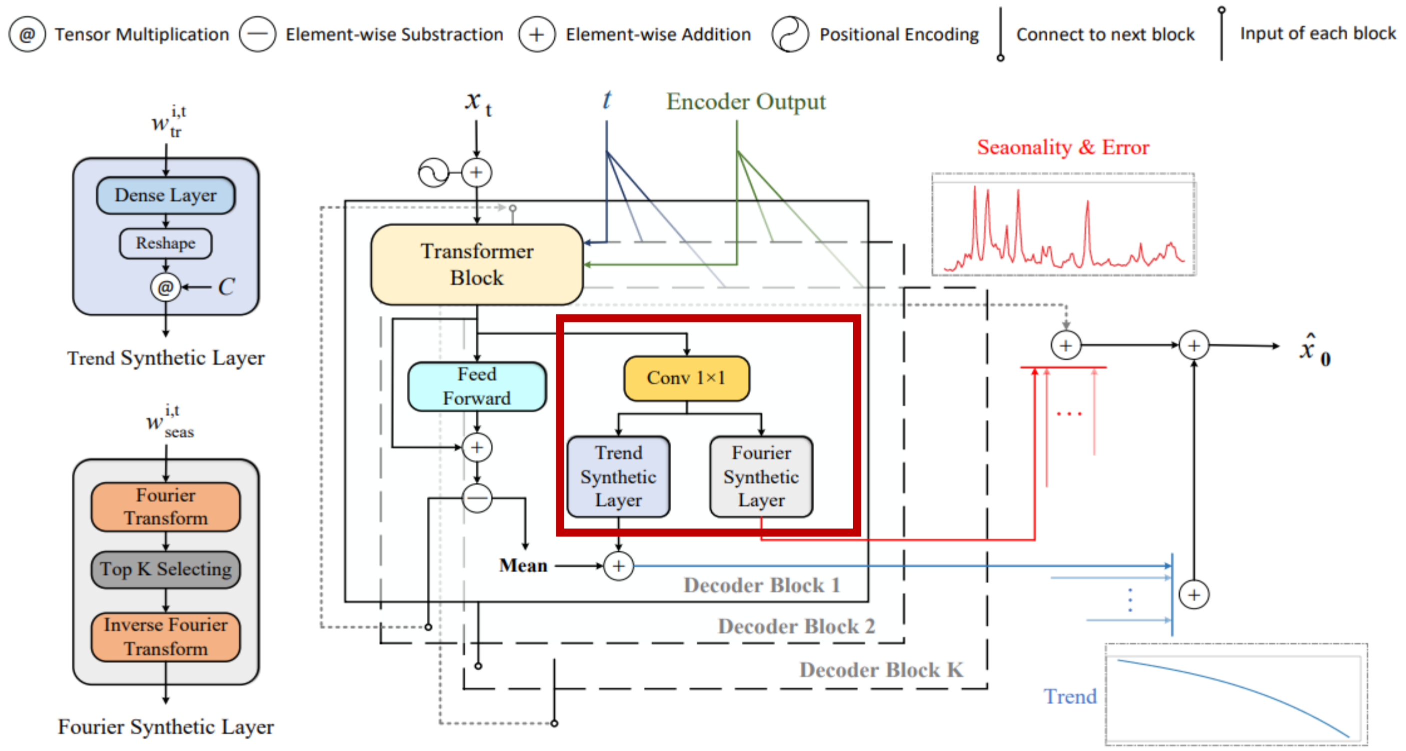


- Transformer：增强捕获全局相关性的能力
- Encoder：在解码之前对整个噪声序列的信息进行编码
- Decoder：将解码器更新为深度分解架构

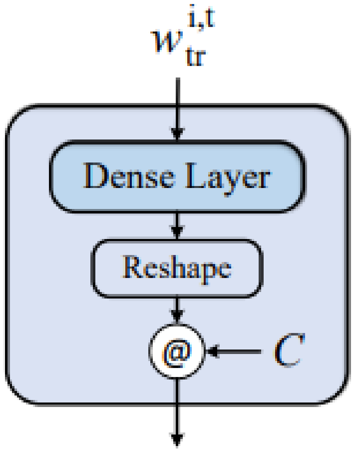


04 算法描述：deep decomposition decoder

- Interpretable layers
  - 分解表示（disentangle representation）：即分解为trend，seasonality，error/remainder
  - 合成任务（synthesis task）：即趋势合成、季节和误差合成
  - 输入：$w^{i,t}_{(\cdot)}$ ，i 表示decoder block序号



算法描述：deep decomposition decoder



Trend Synthetic Layer

- Interpretable layers：Trend Synthesis
  - 趋势分量即数据的平滑均值
  - 用多项式回归器

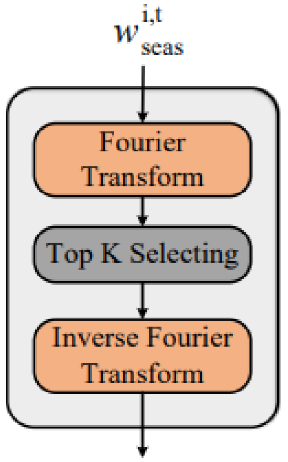
$$V_{tr}^{t} = \sum_{i=1}^{\cancel{N}\,K} \left(C \cdot \mathrm{Linear}(w_{tr}^{i,t}) + \underset{\text{mean}}{\mathcal{X}_{tr}^{i,t}}\right),$$

$$\mathbf{C} = [1, c, \ldots, c^{p}]$$

$$c = [0, 1, 2, \ldots, \tau-2, \tau-1]^{T}/\tau,$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & \frac{1}{\tau} & \frac{1}{\tau^2} & \frac{1}{\tau^3} \\ 1 & \frac{2}{\tau} & \frac{4}{\tau^2} & \frac{8}{\tau^3} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \frac{\tau-1}{\tau} & \frac{(\tau-1)^2}{\tau^2} & \frac{(\tau-1)^3}{\tau^3} \end{bmatrix}$$

04 算法描述：deep decomposition decoder

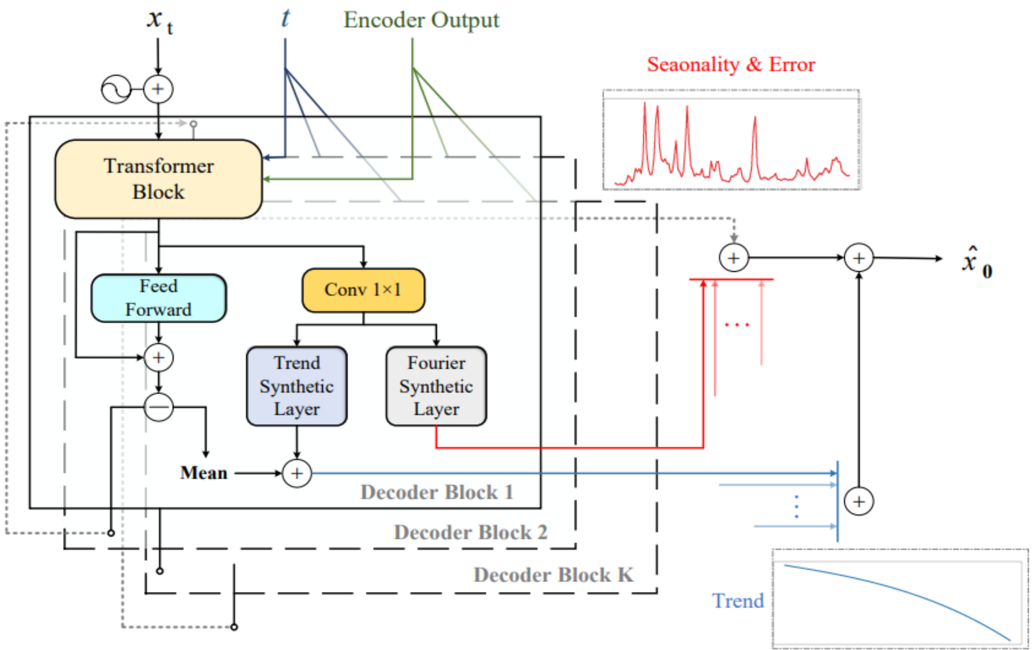


Fourier Synthetic Layer

- Interpretable layers：Seasonality & Error Synthesis
  - 从噪声输入中识别季节模式趋势分量
  - 在频域中选择振幅 TopK 个基，使用傅里叶级数来表示季节性成分

$$S_{i,t} = \sum_{k=1}^{K} A_{i,t}^{\kappa_{i,t}^{(k)}} \left[\cos(2\pi f_{\kappa_{i,t}^{(k)}} \tau c + \Phi_{i,t}^{\kappa_{i,t}^{(k)}}) + \cos(2\pi \bar{f}_{\kappa_{i,t}^{(k)}} \tau c + \bar{\Phi}_{i,t}^{\kappa_{i,t}^{(k)}}) \right]$$

- $$\hat{x}_0(x_t, t, \theta) = V_{tr}^t + \sum_{i=1}^{K} S_{i,t} + R,$$



04 算法描述：Fourier-based training objective

➢ 基于傅里叶的损失项有利于精确重建时间序列信号

$$\mathcal{L}_{simple} = \mathbb{E}_{t,x_0}\left[w_t\|x_0 - \hat{x}_0(x_t,t,\theta)\|^2\right], \quad w_t = \frac{\lambda\alpha_t(1-\bar{\alpha}_t)}{\beta_t^2},$$

⬇

$$\mathcal{L}_\theta = \mathbb{E}_{t,x_0}\left[w_t\left[\lambda_1\|x_0 - \hat{x}_0(x_t,t,\theta)\|^2 + \lambda_2\|\underline{\mathcal{FFT}(x_0)} - \underline{\mathcal{FFT}(\hat{x}_0(x_t,t,\theta))}\|^2\right]\right]$$

*Fons E, Sztrajman A, El-Laham Y, et al. HyperTime: Implicit Neural Representations for Time Series[C]//NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research. 2022.

04 算法描述：conditional generation

- Diffusion-TS的条件扩展
  - 预训练扩散模型：生成与数据集中的样本相似的合成样本。
  - 训练分类器
  - 梯度引导的调节：调节预训练的扩散模型，使其能够生成符合特定任务要求的合成样本。
- 目标：$p(x_{0:T}|y) = \prod_{t=1}^{T} p(x_{t-1}|x_t, y)$
  - 给定条件部分 $\mathbf{x}_\mathrm{a}$ 和生成部分 $\mathbf{x}_b$ 的情况下
  - 对生成部分 $\mathbf{x}_b$ 进行梯度更新，尽可能减小条件部分 $\mathbf{x}_\mathrm{a}$ 与模型预测的条件部分 $\hat{\boldsymbol{x}}_a$ 之间的差异
  - 即通过梯度更新来引导样本生成过程。

*Fons E, Sztrajman A, El-Laham Y, et al. HyperTime: Implicit Neural Representations for Time Series[C]//NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research. 2022.

05 实验评估

- 评估任务：无条件生成，有条件生成
- 基线：
  - 时序生成：TimeVAE (2021)、Diffwave (2021)、 TimeGAN (2019) 、Cot-GAN (2020)
  - 条件任务：CSDI (2021)
- 评估指标：

① Discriminative score：|accuracy − 0.5| （原始数据和合成数据的相似性）

② Predictive score：MAE

③ Context-FID score：计算本地上下文的时间序列表示之间的差异

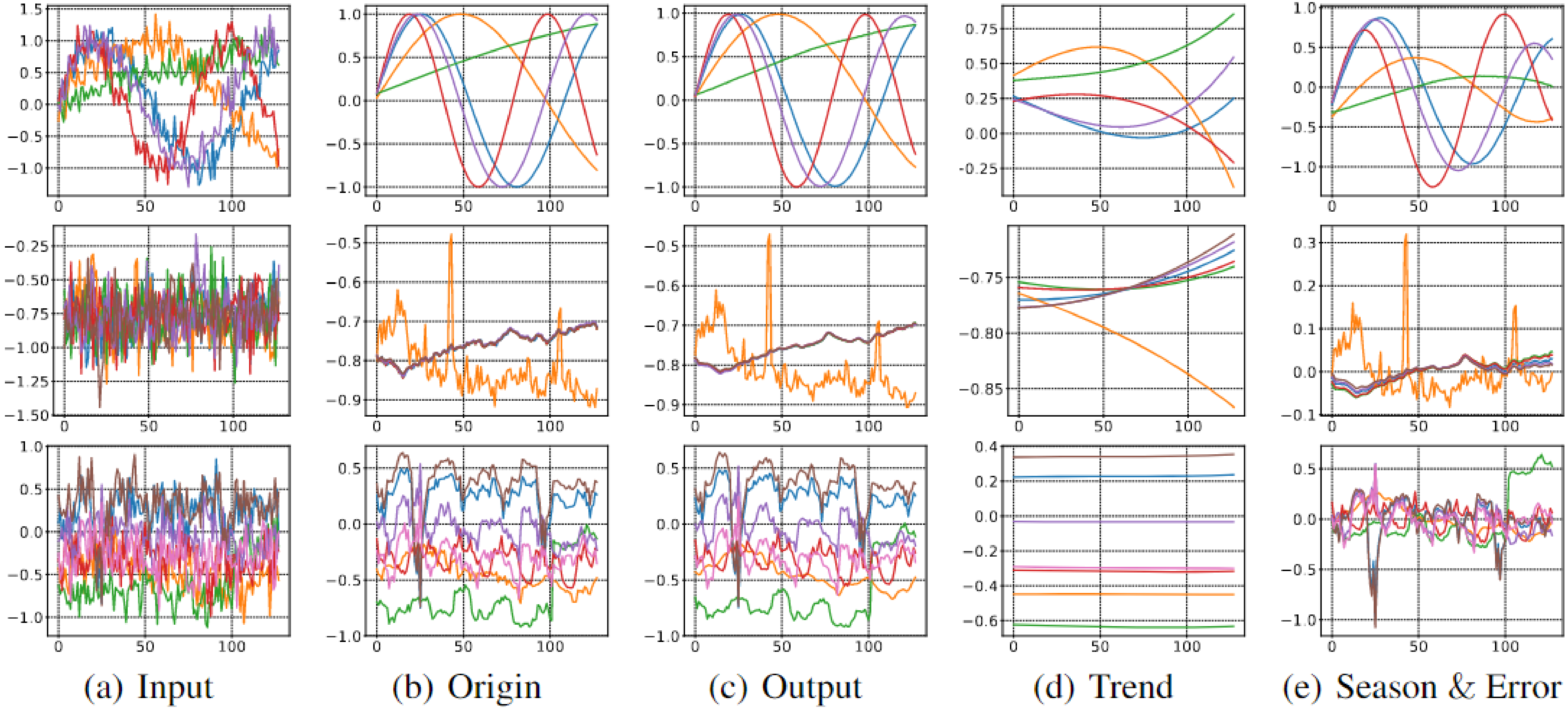
④ Correlational score：评估时间依赖性

05 实验评估

➢ 数据集

	描述	时间步	特征维度
Stocks	Google 股票价格数据	1 day	6
ETTh	变压器油温	15 min	7
Energy	能源使用预测		28
fMRI	血样水平（模拟）		50（实验选择其中1个）
Sines	每个特征独立，且具有不同的频率和相位		5
MuJoCo	多元物理（模拟）		14

05 实验评估：可解释性实验结果



05

实验评估：无条件时序生成

24-length time series generation

Table 1: Results on Multiple Time-Series Datasets (Bold indicates best performance).

Metric	Methods	Sines	Stocks	ETTh	MuJoCo	Energy	fMRI
Context-FID Score (Lower the Better)	Diffusion-TS	**0.006±.000**	0.147±.025	**0.116±.010**	**0.013±.001**	**0.089±.024**	**0.112±.010**
	TimeGAN	0.101±.014	**0.103±.013**	0.300±.013	0.563±.052	0.767±.103	1.292±.218
	TimeVAE	0.307±.060	0.215±.035	0.805±.186	0.251±.015	1.631±.142	14.449±.969
	Diffwave	0.014±.002	0.232±.032	0.873±.061	0.393±.041	1.031±.131	0.244±.018
	Cot-GAN	1.337±.068	0.408±.086	0.980±.071	1.094±.079	1.039±.028	7.813±.550
Correlational Score (Lower the Better)	Diffusion-TS	**0.015±.004**	**0.004±.001**	**0.049±.008**	**0.193±.027**	**0.856±.147**	**1.356±.021**
	TimeGAN	0.045±.010	0.063±.005	0.210±.006	0.886±.039	4.010±.104	23.502±.039
	TimeVAE	0.131±.010	0.095±.008	0.111±020	0.388±.041	1.688±.226	17.296±.526
	Diffwave	0.022±.005	0.030±.020	0.175±.006	0.579±.018	5.001±.154	3.927±.049
	Cot-GAN	0.049±.010	0.087±.004	0.249±.009	1.042±.007	3.164±.061	26.824±.449
Discriminative Score (Lower the Better)	Diffusion-TS	**0.006±.007**	**0.067±.015**	**0.061±.009**	**0.008±.002**	**0.122±.003**	**0.195±.031**
	TimeGAN	0.011±.008	0.102±.021	0.114±.055	0.238±.068	0.236±.012	0.484±.042
	TimeVAE	0.041±.044	0.145±.120	0.209±.058	0.230±.102	0.499±.000	0.476±.044
	Diffwave	0.017±.008	0.232±.061	0.190±.008	0.203±.096	0.493±.004	0.402±.029
	Cot-GAN	0.254±.137	0.230±.016	0.325±.099	0.426±.022	0.498±.002	0.492±.018
Predictive Score (Lower the Better)	Diffusion-TS	**0.093±.000**	**0.036±.000**	**0.119±.002**	**0.007±.000**	**0.250±.000**	**0.099±.000**
	TimeGAN	0.093±.019	0.038±.001	0.124±.001	0.025±.003	0.273±.004	0.126±.002
	TimeVAE	0.093±.000	0.039±.000	0.126±.004	0.012±.002	0.292±.000	0.113±.003
	Diffwave	0.093±.000	0.047±.000	0.130±.001	0.013±.000	0.251±.000	0.101±.000
	Cot-GAN	0.100±.000	0.047±.001	0.129±.000	0.068±.009	0.259±.000	0.185±.003
	Original	0.094±.001	0.036±.001	0.121±.005	0.007±.001	0.250±.003	0.090±.001

05

实验评估：无条件时序生成

长期的时序生成：(64, 128, 256)

Table 3: Detailed Results of Long-term Time-series Generation. (Bold indicates best performance).

Dataset	Metric	Length	Diffusion-TS	TimeGAN	TimeVAE	Diffwave	Cot-GAN
ETTh	Context-FID Score (Lower the Better)	64	**0.631±.058**	1.130±.102	0.827±.146	1.543±.153	3.008±.277
		128	**0.787±.062**	1.553±.169	1.062±.134	2.354±.170	2.639±.427
		256	**0.423±.038**	5.872±.208	0.826±.093	2.899±.289	4.075±.894
	Correlational Score (Lower the Better)	64	0.082±.005	0.483±.019	**0.067±.006**	0.186±.008	0.271±.007
		128	0.088±.005	0.188±.006	**0.054±.007**	0.203±.006	0.176±.006
		256	0.064±.007	0.522±.013	**0.046±.007**	0.199±.003	0.222±.010
	Discriminative Score (Lower the Better)	64	**0.106±.048**	0.227±.078	0.171±.142	0.254±.074	0.296±.348
		128	**0.144±.060**	0.188±.074	0.154±.087	0.274±.047	0.451±.080
		256	**0.060±.030**	0.442±.056	0.178±.076	0.304±.068	0.461±.010
	Predictive Score (Lower the Better)	64	**0.116±.000**	0.132±.008	0.118±.004	0.133±.008	0.135±.003
		128	**0.110±.003**	0.153±.014	0.113±.005	0.129±.003	0.126±.001
		256	**0.109±.013**	0.220±.008	0.110±.027	0.132±.001	0.129±.000
Energy	Context-FID Score (Lower the Better)	64	**0.135±.017**	1.230±.070	2.662±.087	2.697±.418	1.824±.144
		128	**0.087±.019**	2.535±.372	3.125±.106	5.552±.528	1.822±.271
		256	**0.126±.024**	5.032±.831	3.768±.998	5.572±.584	2.533±.467
	Correlational Score (Lower the Better)	64	**0.672±.035**	3.668±.106	1.653±.208	6.847±.083	3.319±.062
		128	**0.451±.079**	4.790±.116	1.820±.329	6.663±.112	3.713±.055
		256	**0.361±.092**	4.487±.214	1.279±.114	5.690±.102	3.739±.089
	Discriminative Score (Lower the Better)	64	**0.078±.021**	0.498±.001	0.499±.000	0.497±.004	0.499±.001
		128	**0.143±.075**	0.499±.001	0.499±.000	0.499±.001	0.499±.001
		256	**0.290±.123**	0.499±.000	0.499±.000	0.499±.000	0.498±.004
	Predictive Score (Lower the Better)	64	**0.249±.000**	0.291±.003	0.302±.001	0.252±.001	0.262±.002
		128	**0.247±.001**	0.303±.002	0.318±.000	0.252±.000	0.269±.002
		256	**0.245±.001**	0.351±.004	0.353±.003	0.251±.000	0.275±.004

05 实验评估：无条件时序生成

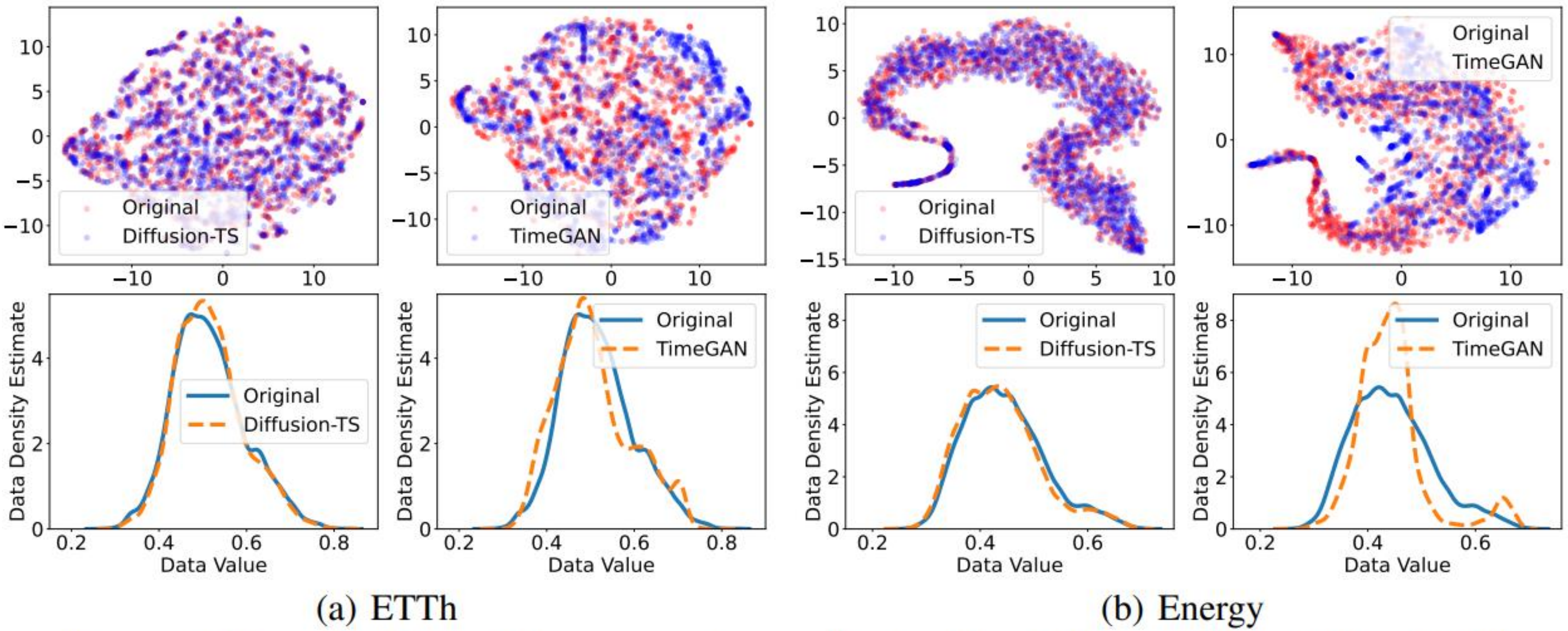
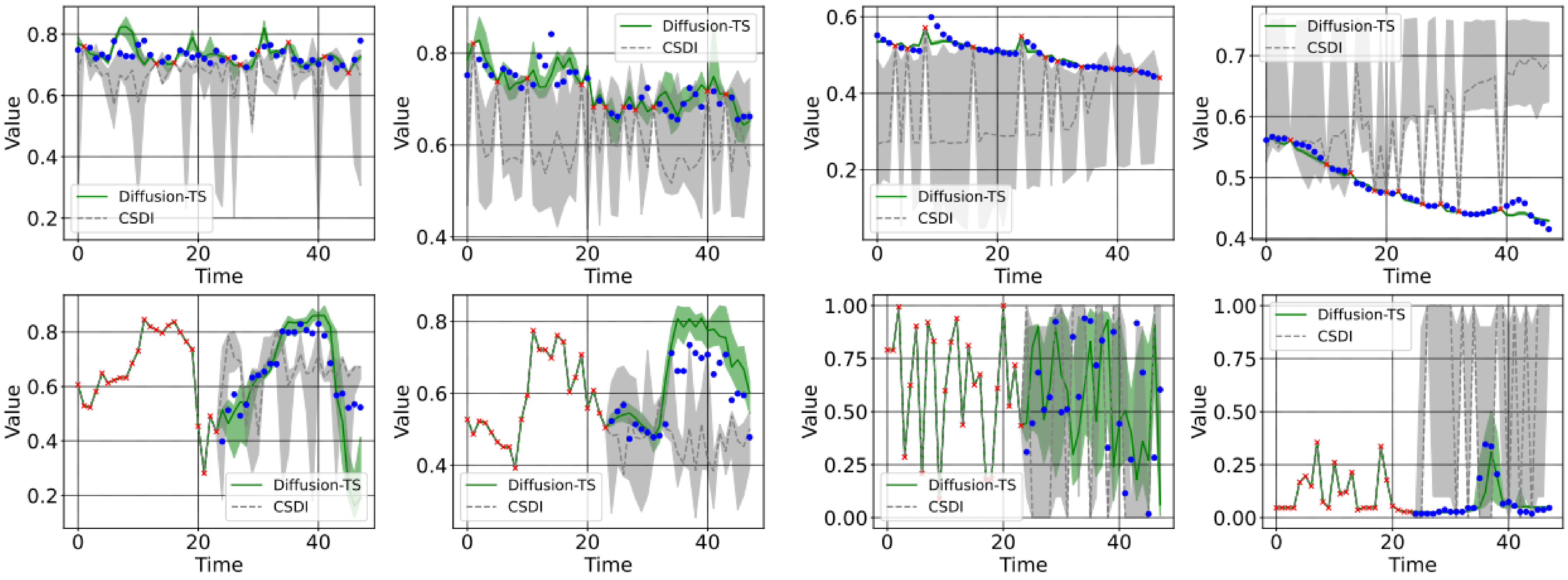


Figure 4: Visualizations of the time series synthesized by Diffusion-TS and TimeGAN.

05

实验评估：条件时序生成

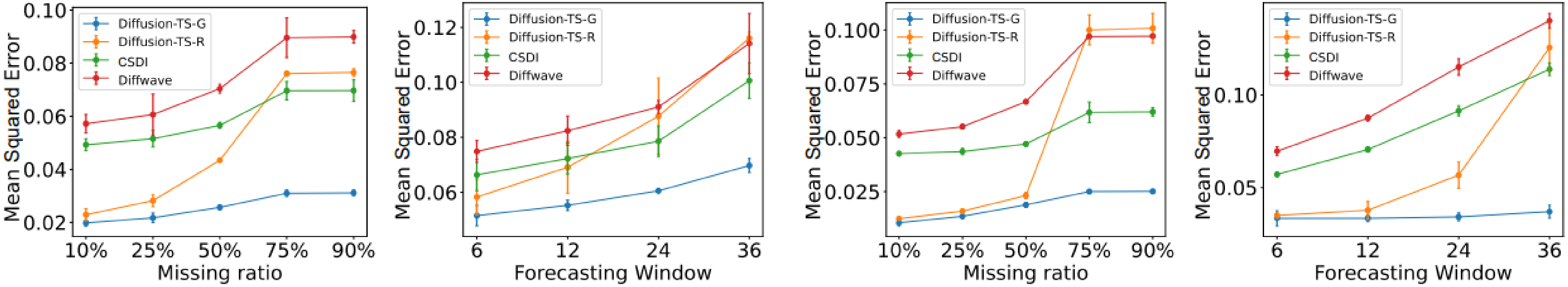
插补和预测



(a) ETTh (b) Energy

05 实验评估：条件时序生成

高缺失率



(a) ETTh　　(b) Energy

Diffusion-TS-R 和 DiffusionTS-G：

季节趋势分解可以在一定程度上帮助模型理解数据的结构和模式，并在数据缺失率较低时表现良好。

但是，当数据缺失率很高且模型缺乏足够的约束时，重构引导的采样策略可能会在填充缺失值时带来更好的效果，因为它可以提供额外的信息来引导模型生成更准确的数据。

05 实验评估：消融实验

Table 2: Ablation study for model architecture and options. (Bold indicates best performance).

Metric	Method	Sines	Stocks	ETTh	Mujoco	Energy	fMRI
Discriminative Score (Lower the Better)	Diffusion-TS	**0.006±.007**	**0.067±.015**	**0.061±.009**	**0.008±.002**	**0.122±.003**	0.195±.031
	w/o FFT	0.007±.006	0.127±.019	0.096±.007	0.010±.002	0.135±.004	0.207±.072
	w/o Interpretability	0.009±.006	0.101±.096	0.071±.010	0.021±.014	0.125±.003	0.267±.034
	w/o Transformer	0.010±.007	0.104±.024	0.082±.006	0.039±.014	0.324±.015	**0.123±.064**
	ϵ-prediction	0.040±.011	0.131±.014	0.099±.010	0.023±.006	0.197±.001	0.168±.030
Predictive Score (Lower the Better)	Diffusion-TS	**0.093±.000**	**0.036±.000**	**0.119±.002**	**0.007±.000**	**0.250±.000**	**0.099±.000**
	w/o FFT	**0.093±.000**	0.038±.000	0.121±.004	0.008±.001	**0.250±.000**	0.100±.001
	w/o Interpretability	**0.093±.000**	0.037±.000	**0.119±.008**	0.008±.001	0.251±.000	0.101±.000
	w/o Transformer	**0.093±.000**	**0.036±.000**	0.126±.004	0.008±.000	0.319±.006	**0.099±.000**
	ϵ-prediction	0.097±.000	0.039±.000	0.120±.002	0.008±.001	0.251±.000	0.100±.000
	Original	0.094±.001	0.036±.001	0.121±.005	0.007±.001	0.250±.003	0.090±.001

05

实验评估：实验效率（Nvidia 3090 GPU）

Table 8: Hyperparameters, training details, and compute resources used for each model

Parameter	Sines	Stocks	ETTh	MuJoCo	Energy	fMRI
basic dimension	64	64	64	64	96	96
attention heads	4	4	4	4	4	4
attention head dimension	16	16	16	16	24	24
encoder layers	1	2	3	3	4	4
decoder layers	2	2	2	2	3	4
batch size	128	64	128	128	64	64
timesteps / sampling steps	500	500	500	1000	1000	1000
training steps	12000	10000	18000	14000	25000	15000
model size	232,177	291,318	350,459	357,214	1,135,144	1,382,290
training time	17min	15min	31min	25min	60min	44min
sampling time (every 2000)	23s	26s	31s	50s	65s	72s

谢谢观看

MANY THANKS !

23.2.27