A Survey on Diffusion Models for Time Series and Spatio-Temporal Data

Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang†, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, Jiang Bian, Shirui Pan, Qingsong Wen†

Abstract—The study of time series is crucial for understanding trends and anomalies over time, enabling predictive insights across various sectors. Spatio-temporal data, on the other hand, is vital for analyzing phenomena in both space and time, providing a dynamic perspective on complex system interactions. Recently, diffusion models have seen widespread application in time series and spatio-temporal data mining. Not only do they enhance the generative and inferential capabilities for sequential and temporal data, but they also extend to other downstream tasks. In this survey, we comprehensively and thoroughly review the use of diffusion models in time series and spatio-temporal data, categorizing them by model category, task type, data modality, and practical application domain. In detail, we categorize diffusion models into unconditioned and conditioned types and discuss time series and spatio-temporal data separately. Unconditioned models, which operate unsupervised, are subdivided into probability-based and score-based models, serving predictive and generative tasks such as forecasting, anomaly detection, classification, and imputation. Conditioned models, on the other hand, utilize extra information to enhance performance and are similarly divided for both predictive and generative tasks. Our survey extensively covers their application in various fields, including healthcare, recommendation, climate, energy, audio, and transportation, providing a foundational understanding of how these models analyze and generate data. Through this structured overview, we aim to provide researchers and practitioners with a comprehensive understanding of diffusion models for time series and spatio-temporal data analysis, aiming to direct future innovations and applications by addressing traditional challenges and exploring innovative solutions within the diffusion model framework.

Index Terms—survey, diffusion model, time series, spatio-temporal data, generative model, temporal data, DDPM, Score SDE.

1 Introduction

Diffusion models represent a family of probabilistic generative models that undergo optimization through a two-step process involving the injection and subsequent removal of noise across a set of training samples. This process comprises a forward phase, referred to as ***diffusion***, and a reverse phase, known as ***denoising***. By training the model to remove the noise added during the diffusion process, the model learns to generate effective data samples during inference that align closely with the distribution of the training data [2], [189].

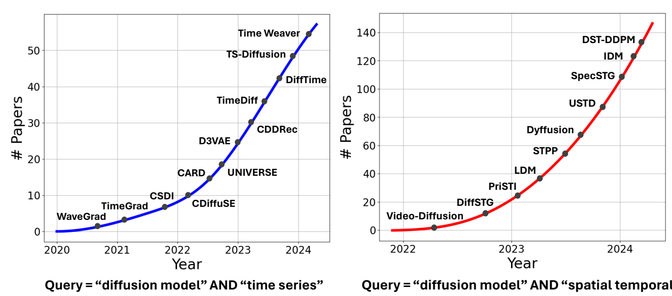

Fig. 1: Trends in the cumulative number of papers related to diffusion models for time series and spatio-temporal data.

In recent years, diffusion models have risen to prominence and significantly influenced various domains, including computer vision (CV) [2], [8], [190], [191], [192], natural language processing (NLP) [172], [193], [194], [195], and general multimodal learning [112], [196], [197], [198]. This challenges the long-time supremacy of generative adversarial networks (GANs) [115], [187]. Within these areas, diffusion models have demonstrated remarkable capabilities in applications such as text-to-image [112], [199], instance segmentation [200], [201], 3D shape generation [202], [203], molecule design [204], [205], [292], and audio generation [91], [206]. Remarkably, diffusion models have also gained popularity as a non-autoregressive alternative for tasks conventionally dominated by autoregressive methods [189]. Recently, the

- *Yiyuan Yang is with the Department of Computer Science, University of Oxford (yiyuan.yang@cs.ox.ac.uk).*
- *Ming Jin and Shirui Pan are with the School of Information and Communication Technology, Griffith University (mingjinedu@gmail.com; s.pan@griffith.edu.au).*
- *Haomin Wen is with Beijing Jiaotong University and the Hong Kong University of Science and Technology (Guangzhou) (wenhaomin@bjtu.edu.cn).*
- *Chaoli Zhang is with the School of Computer Science and Technology, Zhejiang Normal University (chaolizcl@zjnu.edu.cn).*
- *Yuxuan Liang is with INTR & DSA Thrust, Hong Kong University of Science and Technology (Guangzhou) (yuxliang@outlook.com).*
- *Lintao Ma is with Ant Group (lintao.mlt@antgroup.com).*
- *Yi Wang is with The University of Hong Kong (yiwang@eee.hku.hk).*
- *Chenghao Liu is with Salesforce Research (chenghao.liu@salesforce.com).*
- *Bin Yang is with East China Normal University (byang@dase.ecnu.edu.cn).*
- *Zenglin Xu is with Fudan University (zenglin@gmail.com).*
- *Jiang Bian is with Microsoft Research Asia (jiang.bian@microsoft.com).*
- *Qingsong Wen is with Squirrel Ai Learning (qingsongedu@gmail.com).*

Repo: https://github.com/yyysjz1997/Awesome-TimeSeries-SpatioTemporal-Diffusion-Model
†*Corresponding authors: Yuxuan Liang and Qingsong Wen.*
Version date: June 12, 2024.

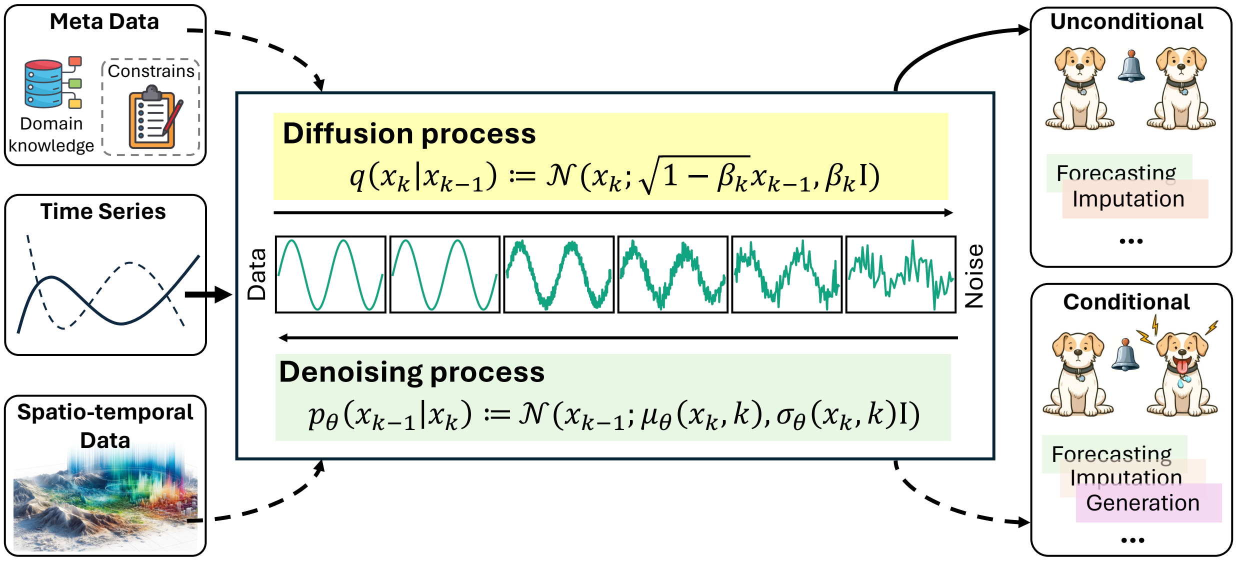

Fig. 2: An overview of diffusion models for time series and spatio-temporal data analysis. In diffusion process, x_k and x_{k-1} denote the results after adding noise at step k and $k-1$, respectively. This process can be represented by the size of the controlling steps $\beta_k \in (0, 1)$, the identity matrix $\mathbf{I}$, and a Gaussian distribution $\mathcal{N}(x; \mu, \sigma)$ of x with the mean μ and the covariance σ. During the denoising process, the model attempts to iteratively learn the data distribution by modelling the distribution $p_\theta(x_{k-1}|x_k)$. The functions $\mu_\theta(\cdot)$ and variance $\sigma_\theta(\cdot)$ are the model learnable parameters.

introduction of OpenAI Sora [87] marks the advent of diffusion models in modeling the physical world embedded within the spacetime continuum, highlighting their critical importance. In addition, AlphaFold 3 [292] proposed by Google DeepMind uses diffusion models to generate 3D atomic coordinates and predict biomolecular structures like proteins, DNA, and RNA.

Temporal data, which primarily includes time series and spatio-temporal data, encapsulates the dynamics of the vast majority of real-world systems [76]. These forms of temporal data have been extensively studied and are recognized as crucial for numerous applications [77], [207], [208]. However, deriving universal dynamic laws in the physical world from various data modalities remains a significant challenge within the field. Recently, the area of time series and spatio-temporal modeling has experienced a substantial shift from sensory intelligence towards general intelligence [209]. This shift is characterized by the emergence of unified foundation models (FMs) that possess versatile temporal data analytical capabilities [76], [209], challenging the supremacy of domain-specific models. Diffusion models have achieved state-of-the-art results on many modalities, including images, speech, and video [210]. Benefiting from the vast and diverse available data in these fields, diffusion models often serve as generative FMs alongside large language models (LLMs) or other foundation models, facilitating rapid development in these areas [7], [190]. In recent years, there has also been an increasing number of diffusion models crafted for modeling time series and spatio-temporal data (Fig. 1). Also, we have become aware of an increasing number of attempts to use diffusion models for temporal modeling (see Tab. 1). Observing the success of diffusion models, an intriguing question arises: ***what kind of sparks will emerge from the intersection of time series/spatio-temporal data analysis and diffusion models?***

Time series and spatio-temporal data analysis fundamentally rely on a profound understanding of their inherent temporal dynamics, wherein primary tasks predominantly focus on the generative capabilities of backbone models, such as forecasting [13], [86], [253], imputation [56], [69], [262] and generation [182], [198]. These analyses center on generating temporal data samples for specific purposes in ***conditional*** or ***unconditional*** manners. Having witnessed the recent development of time series and spatio-temporal foundation models [76], [254], whether built upon LLMs or trained from scratch, their success can be attributed to the ability to estimate the distribution of training samples where effective data representations can be drawn. In this regard, diffusion models emerge as a powerful generative framework that enables (1) the modeling of complex patterns within temporal data and (2) the support of a wide range of downstream tasks, as depicted in Fig. 2 .

To generate valid data samples for specific tasks, time series and spatio-temporal diffusion models usually operate in an unconditional manner without the need for supervision signals. Given the partially-observed nature of real-world applications [255], conditional diffusion models have emerged. They leverage data labels (e.g., instructions, metadata, or exogenous variables) to regulate the generation process, thereby enabling effective cross-modal prompting that leads to more tailored and improved outcomes [182]. We present a roadmap in Fig. 3 . By training on large-scale temporal data, diffusion models effectively fill the gap of time series/spatio-temporal data generation and exhibit significant potential in solving the puzzle of next-generation, LLM-empowered temporal data-centric agents [209], [256].

Despite the promising prospects and rapid advancement of diffusion models in handling time series and spatio-

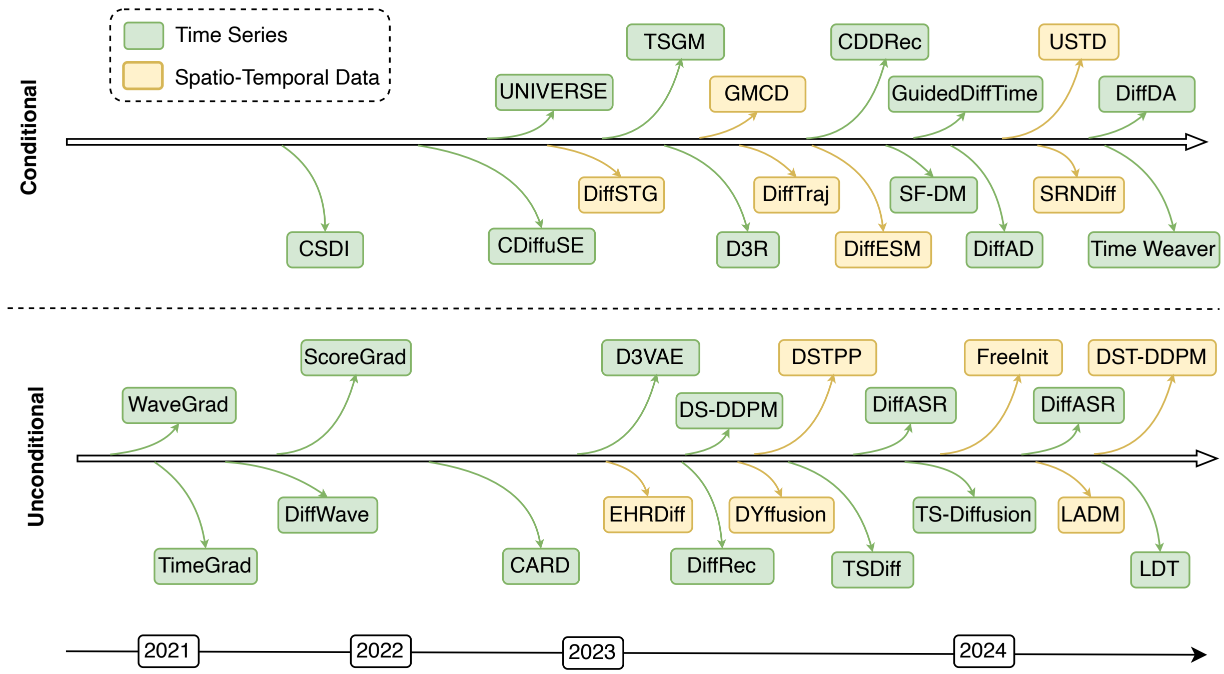

Fig. 3: Representative diffusion models for time series and spatio-temporal data in recent years.

temporal data, there has been a conspicuous lack of systematic analysis of this model family in the existing literature. This article aims to bridge this gap by providing a forward-looking review that elucidates both the 'why' and the 'how' —– detailing the reasons diffusion models are suited for these data modalities and unveiling the mechanisms through which they confer advantages. In this survey, we offer a detailed categorization, engage in thorough reviews, and identify burgeoning trends within this rapidly evolving landscape. Our main contributions are summarized as follows:

- **Comprehensive and up-to-date review.** We present a comprehensive, up-to-date, and forward-looking review of diffusion models for time series and spatio-temporal data. Our survey highlights the suitability of diffusion models for these data modalities and discusses the benefits they confer. By covering both a broad spectrum of the field and the specifics of individual methods, we furnish readers with a deep insight into this subject area.
- **Unified and structured categorization.** We introduce a clear and organized framework for categorizing the existing literature into two main types: unconditional and conditional diffusion models, focusing on time series and spatio-temporal data that span both predictive and generative tasks. This categorization offers the reader a coherent roadmap of the topic from multiple perspectives.
- **Insights into emerging advances.** We discuss cutting-edge techniques in both unconditional and conditional diffusion models, focusing on time series and spatio-temporal data. Our coverage includes the latest techniques and emerging trends such as multimodal conditional generation.
- **Summary of challenges and future directions.** We identify key challenges faced in the current research landscape and highlight several promising directions for future exploration.

The remainder of this paper is structured as follows: Sec. 2 provides a comprehensive background on diffusion models, detailing their development, theoretical foundations, and various implementations. Sec. 3 presents a structured overview and categorization of diffusion models applied to time series and spatio-temporal data, setting the stage for a deeper exploration of model perspectives in Sec. 4 , which discusses both standard and advanced diffusion models. Sec. 5 focuses on task perspectives, examining how diffusion models tackle forecasting, generation, imputation, anomaly detection, and more. Sec. 6 discusses data perspectives, highlighting challenges and solutions specific to time series and spatio-temporal data. Sec. 7 explores the application of diffusion models across various domains, such as healthcare, traffic, and energy, demonstrating their broad utility. Finally, Sec. 8 concludes the paper with an outlook on future opportunities and summarizing remarks.

2 Background

This paper primarily reviews recent improvements in using diffusion models to solve time series and spatio-temporal data challenges. In this section, we will first define time series and spatio-temporal data, as well as their corresponding tasks in various fields. Then, we will introduce the history of diffusion model and its advantages. Finally, some different kinds of diffusion models and their variants based on theoretical formula derivations and comparisons with other generative models will be presented.

2.1 Overview of Time Series and Spatio-Temporal Data

Temporal data, particularly time series and spatio-temporal data, are important data structures for a wide range of

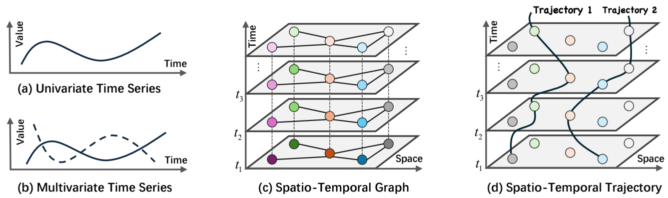

Fig. 4: Illustrations of time series and spatio-temporal data.

real-world applications [76]. A time series is defined as a sequential arrangement of data points, categorized by their temporal order. These sequences may be univariate, involving a single variable over time, or multivariate, incorporating multiple variables. For instance, daily air quality measurements in a city constitute a univariate time series, while combining daily temperature and humidity readings generates a multivariate series. In our discussions, we employ bold uppercase letters (e.g., $\mathbf{X}$) to represent matrices, bold lowercase (e.g., $\mathbf{x}$) for vectors, calligraphic uppercase (e.g., $\mathcal{X}$) for sets, and standard lowercase (e.g., x) for scalars.

We base our formal definitions of time series on those provided in [76], [77]. Specifically, a univariate time series (Fig. 4 (a)), denoted as $\mathbf{x} = (x_1, x_2, \ldots, x_T) \in \mathbb{R}^T$, consists of a sequence of T data points arranged chronologically, where each $x_t \in \mathbb{R}$ represents the series' value at time t. Conversely, a multivariate time series (Fig. 4 (b)), represented by $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T) \in \mathbb{R}^{T \times D}$, encompasses a sequence of T data points also in temporal sequence but across D feature channels, with $\mathbf{x}_t \in \mathbb{R}^D (1 \leq t \leq T)$ indicating the series' values at time t across D different channels. For a comprehensive exploration of time series, we direct the reader to [77].

Spatio-temporal data, in contrast, encompasses sequences of data points characterized by both their temporal and spatial dimensions. This type of data integrates the aspect of time, as seen in time series, with the additional dimension of space, capturing the complex dynamics of phenomena as they unfold over time and across different locations. From this perspective, multivariate time series may also be considered as a form of spatio-temporal data. Such data is instrumental in fields ranging from geography and meteorology to urban planning and environmental monitoring, where understanding the interplay between spatial patterns and temporal evolution is crucial.

In practical applications, spatio-temporal data refers to a collection of observations where each data point is defined by its position in space and time, encapsulating a diverse range of data structures such as graphs, trajectories, and even videos, as noted in [76].

For instance, a spatio-temporal graph (Fig. 4 (c)) representing urban traffic flow [286], [287] over time can be understood as spatio-temporal data where each node represents a specific location with specific attributes and edges are weighted by traffic volume, which changes over time. Likewise, trajectory data [281], [288] (Fig. 4 (d)) captures the movement of objects through space over time, including their paths, speeds, and changes in direction. Such data is vital for applications in transportation studies, wildlife tracking, and mobile network optimization, where analyzing the patterns of movement and predicting future locations based on historical data are of paramount importance. On this basis, these and other spatio-temporal constructs can be systematically characterized and analyzed.

Building on the definitions provided, we now proceed to succinctly introduce the representative tasks associated with each data category [76].

- **Time Series Analysis.** The analysis of time series with diffusion models encompasses four primary tasks: *forecasting*, *generation*, *anomaly detection*, and *imputation*. Forecasting focuses on predicting future values within a time series, which can be subdivided into short-term and long-term forecasts based on the temporal scope of the predictions. Generation involves creating new time series based on the statistical properties of a given dataset, serving as a way to simulate possible scenarios or enhance data diversity for training models. Anomaly detection, a specialized form of classification, aims to distinguish atypical series from normal ones. Imputation addresses the challenge of filling in missing values within a series, which is crucial for maintaining the integrity and utility of time series.
- **Spatio-Temporal Data Analysis.** Spatio-temporal data analysis, while encompassing tasks similar to those in time series analysis, often applies these methodologies within specific application scenarios. For instance, forecasting may focus on traffic flow [43] or air quality [78], utilizing historical data patterns to predict future conditions. Generation tasks might involve creating synthetic trajectories, offering privacy-compliant alternatives to original datasets by replacing sensitive information with generated, anonymized data [81], [290]. Anomaly detection becomes particularly crucial in scenarios such as vehicle trajectory analysis, where deviations from generated normative patterns may indicate unusual or suspicious behaviors [83]. Additionally, spatio-temporal imputation plays a vital role in addressing missing values in multivariate time series [69], ensuring comprehensive and accurate datasets for further analysis. More discussion is in Sec. 5 .

2.2 Why Diffusion Model and Its History

Diffusion models are a class of probability-based generative models. They are called after the mathematical process of diffusion, which is commonly used to describe phenomena such as particle movement in a gas or liquid [1]. In detail, the concept of diffusion models first appeared in statistical physics, used to describe the process of particles moving from areas of high concentration to areas of low concentration [10]. Early diffusion models were primarily concerned with accurately simulating the random diffusion behavior in the generation process.

One of the key breakthroughs in diffusion models occurred in 2015, when researchers proposed a method that combines variational inference to effectively train these models [1]. Since then, the field has experienced rapid development, especially in the area of high-resolution image generation [7]. Since 2020, diffusion models have begun to show their potential in more fields, such as text2image, music generation, and speech synthesis [32], [90], [112]. These advances are due to the optimization of model structures, improvements in training methods, and increased computational resources.

In addition to the field of application, in terms of theory, researchers began to explore how to generate data with specific features by controlling the reverse process. This study direction finally resulted in the development of diffusion models capable of producing high-quality, multidimensional data. In updated computer algorithms, it can be represented as the process of gradually modifying the data distribution, progressively injecting noise until it matches the target distribution, in order to obtain high-quality and realistic synthetic data samples [2].

Recently, more and more researchers and engineers are now focusing their perspectives on the diffusion model, and it has become one of the first choices for generating models [115]. Diffusion models excel at generating *high-quality*, *complex sequences*, including time series and spatial-temporal data, with detailed coherence by gradually removing noise. They offer *strong control* over generations, allowing fine-tuning based on the conditions [114]. These models are *flexible* and *adaptable* across various data types and modalities, *robust* against errors with a gradual noise reduction mechanism, and capable of exploring data diversity for *creative outputs*. Moreover, they can be *integrated with other model types*, like autoencoders, to enhance generation quality and control [115].

2.3 Typical Diffusion Models

Typically, the training process includes two steps: the ***forward process (diffusion)*** and the ***reverse process (denoising)***. Diffusion models start with a noise distribution, which is gradually altered through a series of steps. In the forward process, the model incrementally adds noise to the original data over multiple steps until the data turns into pure random noise. This process is usually Markovian, meaning that each step depends only on the preceding one. Then the reverse process takes place, involving learning to remove the noise from the data, essentially reversing the forward process. By training the model to remove the noise added during the diffusion process, the model learns to generate samples from the same distribution as the training data. The entire training process involves optimizing the model to denoise effectively. This is typically done using a loss function that encourages the model to produce samples that are close to the true data distribution [3].

The currently common frameworks for diffusion models include denoised diffusion probabilistic models (DDPMs) [1], [2], score-based stochastic differential equations (Score SDEs) [4], [6], conditional diffusion models [7], [8], [9], etc. Following, we will introduce the subclasses of the diffusion model through theoretical formula derivation.

2.3.1 Denoised Diffusion Probabilistic Models (DDPMs)

Denoised diffusion probabilistic models are built around a well-defined probabilistic process via dual Markov chains that consist of two parts: a diffusion (or forward) process that gradually transforms the data into noise with predetermined noise, such as Gaussian noise, and a denoising (or reverse) process that attempts to recover the original data by deep neural networks.

Forward (Diffusion) Process. Given a data distribution $q(\mathbf{x})$ and sample an initial clean data $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ from it. The subsequent forward diffusion process incrementally adulterates the initial data distribution by superimposing Gaussian noise, and finally progresses towards convergence with the standard Gaussian distribution. In the diffusion process up to step K, a sequence of distributed latent data $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K$ materializes. The diffusion process can be defined as a Markov chain transforms $\mathbf{x}_{k-1}$ to $\mathbf{x}_k$ with a diffusion transition kernel:

$$q(\mathbf{x}_k|\mathbf{x}_{k-1}) := \mathcal{N}(\mathbf{x}_k; \sqrt{1-\beta_k}\mathbf{x}_{k-1}, \beta_k \mathbf{I}), \tag{1}$$

for $\forall k \in \{1, \cdots, K\}$ with the size of the controlling steps $\beta_k \in (0, 1)$, the identity matrix $\mathbf{I}$, and a Gaussian distribution $\mathcal{N}(\mathbf{x}; \mu, \sigma)$ of $\mathbf{x}$ with the mean μ and the covariance σ. According to the properties of the Gaussian kernel, it is feasible to get $\mathbf{x}_k$ directly from $\mathbf{x}_0$ by equation 1, and collect noise samples straight from the original input $\mathbf{x}_0$ for any step, that is,

$$q(\mathbf{x}_k|\mathbf{x}_0) := \prod_{k=1}^{K} q(\mathbf{x}_k|\mathbf{x}_{k-1}) := \mathcal{N}(\mathbf{x}_k; \sqrt{\bar{\alpha_k}}\mathbf{x}_0, \sqrt{1-\bar{\alpha_k}}\mathbf{I}), \tag{2}$$

$$\text{where } \alpha_k := 1 - \beta_k, \text{ and } \bar{\alpha_k} := \prod_{i=1}^{K} \alpha_i. \tag{3}$$

Therefore, $\mathbf{x}_k = \sqrt{\bar{\alpha_k}}\mathbf{x}_0 + \sqrt{1-\bar{\alpha_k}}\epsilon$ with Gaussian noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$. Typically, it is designed $\bar{\alpha_k} \approx 0$, s.t., $q(\mathbf{x}_k) := \int q(\mathbf{x}_K|\mathbf{x}_0)q(\mathbf{x}_0)\mathbf{d}\mathbf{x}_0 \approx \mathcal{N}(\mathbf{x}_k; \mathbf{0}, \mathbf{I})$, i.e., the backward chain can begin with any Gaussian noise. Overall, the forward process gradually injects noise into the data until all structures have disappeared.

Reverse (Denoising) Process. The reverse process performs the denoising task at each step with a series of Markov chains until the damaged original data is reconstructed. Specifically, the series of reverse Markov chains start with a distribution $p(\mathbf{x}_K) = \mathcal{N}(\mathbf{x}_K; \mathbf{0}, \mathbf{I})$ and a learnable kernel $p_\theta(\mathbf{x}_{k-1}|\mathbf{x}_k)$ to generate $p_\theta(\mathbf{x}_0)$. The learnable Gaussian transition kernels p_θ can be represented as:

$$p_\theta(\mathbf{x}_{k-1}|\mathbf{x}_k) := \mathcal{N}(\mathbf{x}_{k-1}; \mu_\theta(\mathbf{x}_k, k), \sigma_\theta(\mathbf{x}_k, k)\mathbf{I}), \tag{4}$$

where the mean $\mu_\theta(\cdot)$ and variance $\sigma_\theta(\cdot)$ are the model learnable parameters. The model tries to learn the data distribution by the model distribution $p_\theta(\mathbf{x}_0)$ during the reverse denoising process.

Training. In order to approximate the real data distribution, the diffusion model is trained to minimize variational constraints on the negative log-likelihood (NLL):

$$\begin{aligned}\mathbb{E}\left[-\log p_\theta\left(\mathrm{x}_0\right)\right] &\leq \mathbb{E}_q\left[-\log \frac{p_\theta\left(\mathrm{x}_{0:K}\right)}{q\left(\mathrm{x}_{1:K} \mid \mathrm{x}_0\right)}\right] \\ &= \mathbb{E}_q\left[-\log p\left(\mathrm{x}_K\right) - \sum_{k \geq 1} \log \frac{p_\theta\left(\mathrm{x}_{k-1} \mid \mathrm{x}_k\right)}{q\left(\mathrm{x}_k \mid \mathrm{x}_{k-1}\right)}\right] \\ &=: L.\end{aligned} \tag{5}$$

It is equivalent to Kullback–Leibler divergence (KL divergence) format as mentioned in the DDPM paper [2] with three parts: the prior loss L_K, the divergence of the forwarding step and the corresponding reversing step L_{k-1}, and the reconstruction loss L_0:

$$
\begin{aligned}
L := & \mathbb{E}_q[\underbrace{D_{\mathrm{KL}}\left(q\left(\mathrm{x}_K \mid \mathrm{x}_0\right) \| p\left(\mathrm{x}_K\right)\right)}_{L_K} \\
& +\sum_{k>1} \underbrace{D_{\mathrm{KL}}\left(q\left(\mathrm{x}_{k-1} \mid \mathrm{x}_k, \mathrm{x}_0\right) \| p_\theta\left(\mathrm{x}_{k-1} \mid \mathrm{x}_k\right)\right)}_{L_{k-1}} \\
& \underbrace{-\log p_\theta\left(\mathrm{x}_0 \mid \mathrm{x}_1\right)}_{L_0}].
\end{aligned} \tag{6}
$$

Especially, to minimize the NLL, we can only train the divergence loss between two steps L_{k-1}, and using Baye's rule to parameterize the posterior $q\left(\mathrm{x}_{k-1} \mid \mathrm{x}_k, \mathrm{x}_0\right)$, that is:

$$
q\left(\mathrm{x}_{k-1} \mid \mathrm{x}_k, \mathrm{x}_0\right)=\mathcal{N}\left(\mathrm{x}_{k-1} ; \tilde{\boldsymbol{\mu}}_k\left(\mathrm{x}_k, \mathrm{x}_0\right), \tilde{\beta}_k I\right), \tag{7}
$$

$$
\tilde{\mu}_t(\mathbf{x}_k, \mathbf{x}_0) := \frac{\sqrt{\bar{\alpha}_{k-1}}\beta_k}{1-\bar{\alpha}_k}\mathbf{x}_0 + \frac{\sqrt{\alpha_k}(1-\bar{\alpha}_{k-1})}{1-\bar{\alpha}_k}\mathbf{x}_k \ , \tag{8}
$$

$$
\tilde{\beta}_k := \frac{1-\bar{\alpha}_{k-1}}{1-\bar{\alpha}_k}\beta_k. \tag{9}
$$

where α_k is $1-\beta_k$ and $\bar{\alpha}_k$ indicates $\prod_{k=1}^{K}\alpha_k$. L_{k-1} can be equated to the expected value of the ℓ_2-loss between the two mean coefficients:

$$
L_{k-1}=\mathbb{E}_q\left[\frac{1}{2 \sigma_k^2}\left\|\tilde{\boldsymbol{\mu}}_k\left(\mathrm{x}_k, \mathrm{x}_0\right)-\boldsymbol{\mu}_\theta\left(\mathrm{x}_k, k\right)\right\|^2\right]+C. \tag{10}
$$

Ho et al. [2] emphasize that, rather than parameterizing the mean $\mu_\theta(\mathrm{x}_k, k)$, predicting the noise vector at each time step in the forward process by parameterizing $\epsilon_\theta(\mathrm{x}_k, k)$ for simplification:

$$
\mathbb{E}_{k \sim \mathcal{U}(1,K), \mathbf{x}_0 \sim q(\mathbf{x}_0), \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})}\left[\lambda(k)\left\|\epsilon-\epsilon_\theta(\mathbf{x}_k, k)\right\|^2\right], \tag{11}
$$

where $\lambda(k)=\frac{{\beta_k}^2}{2{\sigma_K}^2\alpha_k(1-\bar{\alpha}_k)}$ is a weight that changes noise scale, and ϵ_θ is a model for Gaussian noise prediction. After training using the above loss function, ϵ_θ will be used in the reverse process of ancestor sampling.

Inference (Sampling). Given the noisy data x_K and starting with step K denoising, the final time series is generated through the equation:

$$
\begin{aligned}
p_\theta(\mathbf{x}_{k-1}|\mathbf{x}_k) &= \mathcal{N}(\mathbf{x}_{k-1}; \mu_\theta(\mathbf{x}_k, k), \sigma_\theta^2(\mathbf{x}_k, k)\mathbf{I}) \\
&\sim \frac{1}{\sqrt{\alpha_k}}(\mathbf{x}_k - \frac{\beta_k}{\sqrt{1-\overline{\alpha_k}}}\epsilon_\theta(\mathbf{x}_k, k)) + \sigma_\theta(\mathbf{x}_k, k)z,
\end{aligned} \tag{12}
$$

where $z \sim \mathcal{N}(0, \mathbf{I})$, also $\beta_k \approx \sigma_\theta^2(\mathbf{x}_k, k)$ in practice.

2.3.2 *Score SDE Formulation*

DDPM achieved a set of discrete steps in the forward processing, so it has some limitations about training designs. Score SDE further generalizes DDPM's discrete system to a continuous framework based on the stochastical differential equation [6]. Here we use T instead of the step size k in DDPM.

Forward Process. The corresponding continuous diffusion process can be represented using Itô SDE [291], including a mean shift and a Brownian motion (standard Wiener process) as follows:

$$
\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}, t \in [0, T], \tag{13}
$$

where $\mathbf{f}(\cdot, t)$ represents the drift coefficient for the stochastic process $\mathbf{x}(t)$, and $g(\cdot)$ is the diffusion coefficient linked with the Brownian motion $\mathbf{w}$.

Similar with DDPM, $\mathbf{x_0}$ and $\mathbf{x}_T$ represent sequence from the clean distribution $p_0 = \mathcal{N}(\mathbf{x}_0; \mathbf{0}, \mathbf{I})$ and the standard Gaussian distribution $p_T = \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I})$, respectively. And the corresponding SDE is:

$$
\mathrm{d}\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}\ dt + \sqrt{\beta(t)}\mathrm{d}\mathbf{w}. \tag{14}
$$

Reverse Process. The new samples can be synthesized from the known prior distribution p_T by solving the reverse-time SDE [174]:

$$
\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}, \tag{15}
$$

Where $\bar{\mathbf{w}}$ is a Brownian motion with reversed time flows [148]. The solution to the reverse-time SDE is approximated by a time-dependent neural network $s_\theta(\mathbf{x}, t)$ to a score function $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$. The solution of the forward SDE equation is that:

$$
L := \mathbb{E}_t\{\lambda(t)\mathbb{E}_{\mathbf{x}_0}\mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\|s_{\boldsymbol{\theta}}(\mathbf{x}_t t) - \nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t|\mathbf{x}_0)\|_2^2]\}, \tag{16}
$$

where $\mathbf{x}_0$ is sampled from distribution p_0 and $\lambda(t)$ is the positive weighting function. This method circumvents the direct approximation of the computationally infeasible score function by estimating the transition probability that follows a Gaussian distribution during the forward diffusion process [6].

Besides, there are some simplified explanations of reverse-time SDE solvers. Using these techniques, samples can be generated after training using various methods.

- Euler-Maruyama (EM) Method [147]: Solves the reverse-time SDE through a simple discretization technique, replacing $\mathrm{d}\mathbf{x}$ with $\triangle\mathbf{t}$ and $\mathrm{d}\bar{\mathbf{w}}$ with Gaussian noise z.
- Prediction-Correction (PC) Method: Operates in a sequential manner, alternating between predictor and corrector steps. The predictor can use any numerical solver, such as the EM method, for the reverse-time SDE, while the corrector can be any score-based Markov Chain Monte Carlo (MCMC) method.
- Probability Flow ODE Method [6]: Reformulates the forward SDE into an ODE that maintains the same marginal probability density p_t as the SDE. Sampling by

solving this reverse-time ODE is equivalent to solving the time-reversed SDE. There are also some advanced ODE solvers to speed up the process [145], [146].

2.3.3 Conditional Diffusion Models

In the previous sections, we discuss the DDPM and Score SDE from an unconditional view. They generate data without any explicit conditions or guidance. The model learns to produce outputs from the learned distribution of the input data. The general diffusion models are capable of generating data samples not just from an unconditional distribution p_0, but also from a conditional distribution $p_0(\mathbf{x}|c)$ when given a condition c. This condition could be class labels or features related to the input data $\mathbf{x}$ [7]. During training, the score network $s_\theta(\mathbf{x}, t, c)$ takes condition c as an input. Additionally, there are specific sampling algorithms designed for conditional generation, such as label-based conditions [8], label-free conditions [9], and further distillation-based guidance [8], self guidance [150], [171], textual based guidance [149], [172], graph-based guidance [173], physical based guidance [163], task-based guidance [165], etc. These conditional mechanisms are more conducive to the generation of application-specific fields by using the control of other information to generate results [115].

In detail, sampling under labels and classifiers' conditions involves using gradient guidance at each step, which typically requires an additional classifier with encoder architecture (e.g., U-Net [151] and Transformer [152]) to generate condition gradients for specific labels [8]. These labels are flexible, and can be textual or categorical, binary, or based on extracted features [8], [146], [153], [154], [155], [157], [158], [167], [169], [201]. Correspondingly, sampling under unlabeled conditions relies solely on self-information for guidance [150], [159], [164]. Compared to the high accuracy of the labeled conditional diffusion model, the unlabeled one has advantages in generating innovative and diverse data. Therefore, unlabelled models are more suitable for exploratory and creative application scenarios [160], [161]. Furthermore, there are also more conditional methods being proposed, which use information about the data itself [164], other modalities [153], [263], other representations [163], and other knowledge [168] as conditions to guide the diffusion model for generation. Currently, the condition-based diffusion model is also the most common method in various application scenarios due to its highly specific and controlled outputs [162], [199].

2.3.4 Improvements with Diffusion Model and Its Variants

While diffusion models have generally yielded satisfactory results in generation and various tasks, practical applications reveal certain limitations. These include the slow iterative sampling process and the computational complexity arising from high-dimensional input, which affects efficiency. Furthermore, there are concerns regarding their generalization performance across different distributions and the challenges in integrating them with other generative models. Next, we will explore the variants of diffusion models from efficiency and performance improvement, as shown in Fig. 5, highlighting the modifications and optimizations they bring in comparison to original diffusion models.

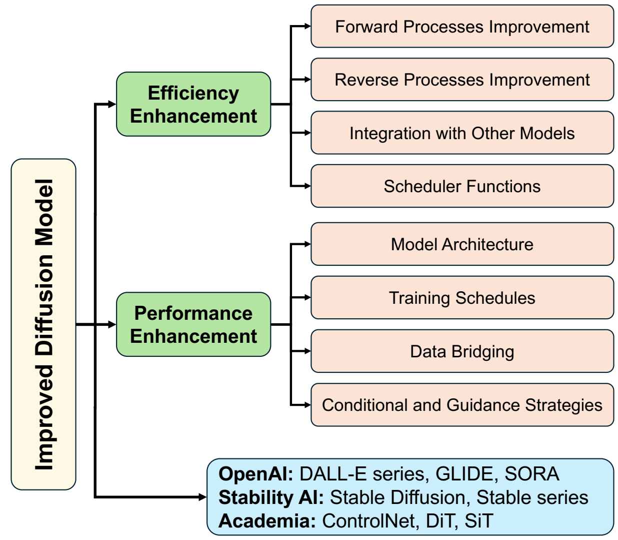

Fig. 5: Categorization of improvements with diffusion model and its variants.

Efficiency Enhancement. (1) **Forward Processes Improvement** shifted towards leveraging other physical phenomena to enhance model efficiency and robustness. For example, inspired by electric field dynamics, PFGM [211] and its extension PFGM++ [212] [213] were proposed. They are guided distributions along electric field lines and employing augmented dimensions for improved performance. Moreover, innovations like Cold Diffusion explore the use of image transformations as a forward process [214]. Also, some improved Gaussian perturbation kernel methods were presented [215]. (2) **Reverse Processes Improvement** usually reduces the number of generation steps or uses lightweight models to enhance efficiency. Training-free sampling methods offer a paradigm where the acceleration of the sampling process does not necessitate retraining and optimize the trajectory from noise to data distributions, such as ODE-based methods DDIM [12], its extensions gDDIM [216], PNDM [145], EDM [215], DEIS [217], DPM-Solver [146], etc, and SDE-based methods through advanced techniques like restart sampling [218], [219]. On the other hand, some methods apply knowledge distillation to create an efficient and smaller network by transferring insights from large, intricate teacher models to simpler student ones [220], [221]. (3) **Integration with Other Models** is a common trick to enhance efficiency, such as combining with VAE [224], [225] and GAN [222], [223]. Additionally, there are methods that use the latent space as an input for the diffusion model, which can improve efficiency by reducing the input dimensionality [7], [225], [227], [228]. (4) **Scheduler Functions** improve efficiency by optimizing the reverse process, enabling faster convergence and reducing the number of required iterations. Common improved schedulers and their algorithms include CMS [221], DDIM [12], IDDPM [3], DEIS [217], DPM-Solvers [146], [229], Euler and Heun scheduler [215], LCM [230], RePaintScheduler [231], TCD [232], UniPC [233], and VQD [234].

Performance Enhancement. (1) **Model Architecture**. Traditional DDPM uses the UNet architecture for its efficiency in reducing computational costs by leveraging a detailed feature space. UNet-based Innovations included normalization [220], [236], and different attentions with position encoding [217], [235], [237]. Recently, to improve performance, transformer-based architecture is being used more often [152]. Reverse noise prediction incorporates adaptations to accommodate temporal and conditional inputs. Common methods include ViT [239], SwinTransformer [240], [241], DiT [242], and SiT [243]. (2) **Training Schedules** involve advanced decoding strategies, mainly classified into optimizing the diffusion stages and innovative projection methods. Optimization strategies focus on adjusting the diffusion steps to fine-tune the model performance [222], [224]. Meanwhile, projection techniques investigate various diffusion processes, like using linear distortions and different kernels, to increase model versatility [244], [245]. Moreover, innovations in diffusion models tweak noise adjustment for better output and quicker alignment [3], [235], [246]. Also, there are some optimization methods focusing on loss and matching [247], [248]. (3) **Data Bridging** is designed to address the limitations of generating arbitrary Gaussian distributions with complex distributions. Research has innovated with SDE/ODE principles to bridge gaps between distributions. Techniques like α-blending create deterministic paths, using diffusion models for Gaussian-related cases [249]. Rectified Flow and other methods introduce optimizations and explore ODE creation between distributions [226], [250]. Additionally, leveraging the Schrödinger Bridge concept or Gaussian distributions as intermediaries offers new avenues in distribution transportation [251], [252]. (4) **Conditional and Guidance Strategies** are crucial for diffusion models, directing generation and enhancing output relevance and quality in response to specific conditions. Related works have been discussed in Sec. 2.3.3.

Currently, research institutions, such as OpenAI (DALL-E series, GLIDE, SORA) and Stability AI (Stable Diffusion), have launched numerous outstanding diffusion models, which have exploded in popularity in both academic research and applications. For instance, Stable Diffusion is a state-of-the-art diffusion model known for its ability to generate high-quality, detailed images from textual descriptions, offering a powerful tool for creative and generative tasks [7]. ControlNet is a framework designed to enhance control over the attributes and structure of generated images, enabling precise manipulation of visual elements in generative models [199]. The consistency model achieves state-of-the-art performance without the slow iterative process of traditional diffusion models and offers capabilities like zero-shot editing and efficient training options [221].

2.4 Diffusion Model vs. Other Generative Models

In addition to the diffusion model, there are many classical generative models, we take the most widely used variational autoencoders (VAEs), generative adversarial networks (GANs), and flow-based generative models as examples to introduce them and analyse their advantages, disadvantages and differences with the diffusion model. The workflow of these models is shown in Fig. 6.

VAE is a probabilistic generative model that encodes input data into a latent space z and decodes from that latent

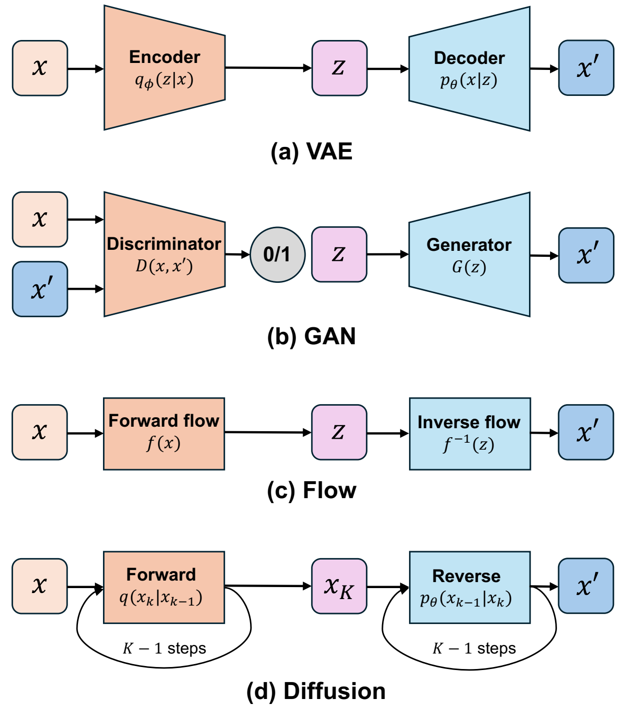

Fig. 6: Different types of generative models. (a) VAE is a probabilistic generative model based on encoding data into a latent space and decoding from this space for generation. (b) GAN generates new data similar to the training data by pitting two neural networks, a generator, and a discriminator, against each other in a game-like scenario. (c) Flow-based model generates data using invertible transformations, enabling precise computation of the probability density function of the data. (d) The diffusion model learns the underlying distribution of data by gradually introducing and then removing noise, enabling the generation of high-quality and diverse data samples.

space to generate data [186], as shown in Fig. 6(a). VAEs aim to maximize the lower bound of the log-likelihood of the data, known as the Evidence Lower Bound (ELBO).

$$\mathsf{ELBO} = \mathbb{E}_{q_\phi(z|x)}[\mathsf{log} p_\theta(x|z)] + \mathsf{KL}[q_\phi(z|x)||p_\theta(z)], \quad (17)$$

where $q_\phi(z|x)$ is the posterior distribution of the latent space output by the encoder and $p_\theta(x|z)$ is the conditional distribution of the data given the latent variable z.

GAN consists of a generator $G(z)$ and a discriminator $D(x, x')$, as shown in Fig. 6(b). The generator tries to produce samples as close to the real data x as possible, while the discriminator tries to distinguish between real data x and generated data x'. The training of GANs involves a zero-sum game, continually optimizing both the generator and the discriminator [187]. The objective function can be summarised as follows:

$$\begin{aligned} \mathsf{min}_G \mathsf{max}_D V(D, G) = \mathbb{E}_{x \sim p_{data(x)}}[\mathsf{log} D(x, x')] \\ + \mathbb{E}_{z \sim p_z(z)}[\mathsf{log}(1 - D((G(z)))]. \end{aligned} \quad (18)$$

Flow-based generative model (normalizing flows) transform data into a simpler distribution (e.g., Gaussian) through a sequence of reversible transformations, as shown in Fig. 6(c), ensuring precise likelihood evaluation due to the transformations being invertible [188]. The flow-based transformation formula is:

$$z = f(x); \quad x' = f^{-1}(z), \tag{19}$$

where x and z are the points in the data space and latent space, respectively. f is an invertible function. The log-likelihood of the transformation is given by:

$$\log p(x) = \log p_z(f(x)) + \log \left| \det(\frac{df}{dx}) \right|. \tag{20}$$

Actually, different generative models each have their own unique strengths and limitations, making them suitable for specific applications. VAE is appreciated for its simplicity, stability, and clear theoretical underpinnings, yet it tends to produce lower-quality data and has limited expressiveness in its latent spaces. GAN is renowned for having powerful generative capabilities but is notoriously difficult to train and prone to instability and mode collapse. The flow-based model provides precise likelihood estimation and facilitates high-quality generation, but it demands significant computational resources and involves complex model designs. Diffusion models excel in generating high-quality and detailed output, offering flexibility and solid probabilistic foundations, but suffer from long training times and high computational costs. However, many variants of diffusion models and improvements have mitigated these issues, as described in Sec. 2.3.4. Currently, the diffusion model is the most popular generative model.

3 Overview and Categorization

This section presents an overview and classification of diffusion models for addressing challenges in time series and spatio-temporal data analysis. Our survey organizes the discussion along four primary dimensions: categories of diffusion models, types of tasks, data modalities, and practical applications. A comprehensive summary of notable related works is depicted in Fig. 7 . We categorize the existing literature into two primary groups: ***unconditioned*** and ***conditioned*** diffusion models, focusing on time series and spatio-temporal data.

In the unconditioned category, diffusion models operate in an unsupervised manner to generate data samples without the need for supervision signals. This setting represents the foundational approach for analyzing time series and spatio-temporal data. Within this category, literature can be further divided into ***probability-based*** and ***score-based*** diffusion models. Examples include denoised diffusion probabilistic models (DDPMs) [2] and score-based stochastic differential equations (Score SDEs) [4], [6], as introduced in Sec. 2 . Research in this category is broadly organized into two task groups: ***predictive*** and ***generative*** tasks. Predictive tasks typically involve forecasting and anomaly detection, leveraging historical data and patterns to anticipate current and/or future events. Generative tasks, conversely, focus on identifying patterns within extensive datasets to generate new content, such as time series imputation and augmentation. Methods are developed for both primary data modalities: ***time series*** and ***spatio-temporal data***, catering to a wide range of applications across various sectors, including healthcare, energy, climate, traffic, and more.

In the conditioned category, diffusion models are tailored for conditioned analysis of time series and spatio-temporal data. Empirical studies have shown that conditional generative models, which utilize data labels, are easier to train and yield superior performance compared to their unconditional counterparts [75]. In this context, labels (a.k.a. conditions) often derive from various sources, such as extracted short-term trends [34] and urban flow maps [35], to enhance model inferences. This category embraces both probability-based and score-based diffusion models for predictive and generative tasks, offering a fresher perspective on leveraging diffusion models to tackle practical challenges in time series and spatio-temporal data analysis under specific constraints.

Building on the foundational understanding of model categories, task types, data modalities, and application domains, we delve deeper into the exploration of diffusion models for time series and spatio-temporal data analysis across the following sections. Each section is designed to unpack the complexities and nuances inherent in the application of diffusion models, providing a comprehensive overview from multiple perspectives. In Sec. 4 , we explore diffusion model landscapes, highlighting distinctions between unconditioned and conditioned approaches and their implications. Sec. 5 analyzes tasks from predictive and generative viewpoints, detailing specific functions such as forecasting, generation, anomaly detection, and data imputation. Sec. 6 examines data modalities, differentiating between time series and spatio-temporal data to outline model challenges and applicability. Lastly, Sec. 7 extends the discussion to application fields, demonstrating diffusion models' utility across sectors like healthcare, traffic, sequential recommendation, climate, energy, and audio. This structured exploration aims to equip readers with an in-depth understanding of the potential and current state of diffusion models for addressing complex time series and spatio-temporal data challenges.

4 Model Perspective

In this section, we will analyze how to use diffusion models for time series and spatio-temporal data from a model perspective. Specifically, we will focus on standard diffusion models (DDPM and score SDE) and improved diffusion models (conditional diffusion model, LDM, DDIM, and others).

4.1 Standard Diffusion Model

Standard diffusion models include probability-based model, i.e., DDPM and score-based model, score SDE, which have been described in detail in Sec. 2.3.1 and Sec. 2.3.2, respectively. Currently, diffusion models based on these two are also the most common methods for time series and spatio-temporal data analysis.

4.1.1 Probability-Based Model

Based on the probabilistic-based standard model, we mainly introduce DDPM there, which uses a ***discrete*** framework,

TABLE 1: Summary and main papers of the diffusion models for time series and spatio-temporal data modeling. Red indicates univariate time series, blue-violet is multivariate time series, and yellow shows spatio-temporal data.

Method	Data	Model	Task	Application	Institute	Venue	Year
WaveGrad [20]	Univariate	DDPM	Generation	Audio	Google	ICLR	2021
DiffWave [91]	Univariate	DDPM	Generation	Audio	UCSD	ICLR	2021
D-Va [47]	Univariate	DDPM	Forecasting	Finance	NUS	CIKM	2023
DiffLoad [28]	Univariate	DDPM	Forecasting	Electricity	HKU	ArXiv	2023
TDSTF [289]	Univariate	DDPM	Forecasting	Healthcare	U of A	ArXiv	2023
DiffuASR [23]	Univariate	DDPM	Generation	Recommendation	XJTU	CIKM	2023
DiffRec [132]	Univariate	DDPM	Generation	Recommendation	NUS	SIGIR	2023
RecFusion [130]	Univariate	DDPM	Generation	Recommendation	UvA	ArXiv	2023
DiffRec [134]	Univariate	DDPM	Generation	Recommendation	SCU	ArXiv	2023
DiffuRec [95]	Univariate	DDPM	Generation	Recommendation	WHU	ArXiv	2023
PDRec [135]	Univariate	DDPM	Generation	Recommendation	SDU	ArXiv	2024
TimeGrad [13]	Multivariate	DDPM	Forecasting	General	Zalando	ICML	2021
CARD [98]	Multivariate	DDPM	Classification	General	UT-Austin	NeurIPS	2022
BVAE [86]	Multivariate	DDPM	Forecasting	General	Baidu	NeurIPS	2023
TSDiff [164]	Multivariate	DDPM	Forecasting	General	Amazon	NeurIPS	2023
Kuo et al. [19]	Multivariate	DDPM	Generation	Healthcare	UNSW	ArXiv	2023
TS-Diffusion [278]	Multivariate	DDPM	Generation	General	Cambridge	ArXiv	2023
SSSD [275]	Multivariate	DDPM	Imputation	General	Oldenburg	TMLR	2023
DA-TASWDM [276]	Multivariate	DDPM	Imputation	Healthcare	HKBU	CIKM	2023
Pintilie et al. [259]	Multivariate	DDPM	Anomaly Detection	General	Bucharest	ICDM	2023
DDMT [99]	Multivariate	DDPM	Anomaly Detection	General	FNU	ArXiv	2023
D3A-TS [101]	Multivariate	DDPM	Imputation&Generation	General	Seville	ArXiv	2023
TSDM [103]	Multivariate	DDPM	Generation	Environment	HIT	ArXiv	2023
TimeDDPM [102]	Multivariate	DDPM	Generation	Environment	ZUST	IEEE Sens. J.	2023
DiffEEG [18]	Multivariate	DDPM	Forecasting	Healthcare	USTC	ArXiv	2023
Diff-E [89]	Multivariate	DDPM	Classification	Healthcare	Korea University	Interspeech	2023
Tosato et al. [119]	Multivariate	DDPM	Classification	Healthcare	Tilburg	Synapsium	2023
DiffCharge [96]	Multivariate	DDPM	Generation	Electricity	HKUST	ArXiv	2023
Yang et al. [238]	Multivariate	DDPM	Imputation&Anomaly Detection	AIOps	Microsoft	ESEC/FSE	2023
AnoDDPM [46]	Multivariate	DDPM	Anomaly Detection	General	Beihang	IEEE Sens. J.	2024
DiffShape [277]	Multivariate	DDPM	Classification	General	SCUT	AAAI	2024
STPP [70]	Spatio-temporal	DDPM	Forecasting	General Event	THU	KDD	2023
ERDiff [25]	Spatio-temporal	DDPM	Alignment	General	GIT	NeurIPS	2023
DVGNN [59]	Spatio-temporal	DDPM	Forecasting	Transportation	Halmstad	ArXiv	2023
Yun et al. [69]	Spatio-temporal	DDPM	Imputation	General	KAIST	OpenReview	2023
DiffTAD [83]	Spatio-temporal	DDPM	Anomaly Detection	Transportation	Xidian	KNOWL-BASED SYST	2024
SpecSTG [88]	Spatio-temporal	DDPM	Forecasting	Transportation	USYD	ArXiv	2024
DST-DDPM [78]	Spatio-temporal	DDPM	Forecasting	Environment	HKUST	ENVIRON RES	2024
Westny et al. [80]	Spatio-temporal	DDPM	Forecasting	Transportation	LiU	ArXiv	2024
UTD-PTP [79]	Spatio-temporal	DDPM	Forecasting	Transportation	BIT	IEEE Sens. J.	2024
SGMSE [284]	Univariate	Score-based	Generation	Audio	UHH	Interspeech	2022
DeScoD-ECG [57]	Univariate	Score-based	Denoising	Healthcare	U of A	IEEE J BIOMED HEALTH	2023
ScoreGrad [71]	Multivariate	Score-based	Forecasting	General	BIT	ArXiv	2021
Bilo et al. [104]	Multivariate	Score-based	Forecasting	General	TUM	ICML	2023
StoRM [105]	Multivariate	Score-based	Denoising&Generation	Audio	UHH	IEEE-ACM T AUDIO SPE	2023
Lay et al. [178]	Multivariate	Score-based	Denoising&Generation	Audio	UHH	Interspeech	2023
Risk-sensitive SDE [280]	Multivariate	Score-based	Generation	General	Cambridge	ArXiv	2024
TimeADDM [282]	Multivariate	Score-based	Anomaly Detection	General	HFUT	ICASSP	2024
Sasdim [72]	Spatio-temporal	Score-based	Imputation	General	CSU	ArXiv	2023
Dyffusion [285]	Spatio-temporal	Score-based	Forecasting	General	UCSD	NeurIPS	2023
Lu et al. [175]	Univariate	Conditional	Generation	Audio	CMU	ICASSP	2022
Wang et al. [142]	Univariate	Conditional	Generation	Electricity	U of Macau	ArXiv	2023
PulseDiff [144]	Univariate	Conditional	Imputation	Healthcare	IC	ArXiv	2023
DCDR [131]	Univariate	Conditional	Generation	Recommendation	Kuaishou	ArXiv	2023
Wang et al. [36]	Univariate	Conditional	Generation	Recommendation	UIC	ArXiv	2023
GiffCF [38]	Univariate	Conditional	Generation	Recommendation	USTC	ArXiv	2023
DR-DiffuSE [176]	Univariate	Conditional	Generation	Audio	UESTC	AAAI	2023
Dose [177]	Univariate	Conditional	Generation	Audio	UESTC	NeurIPS	2023
CRA-DIFFUSE [179]	Univariate	Conditional	Generation	Audio	XJU	ICME	2023
DiffsFormer [52]	Univariate	Conditional	Imputation&Generation	Finance	USTC	ArXiv	2024
DreamRec [167]	Univariate	Conditional	Generation	Recommendation	USTC	NeurIPS	2024
CSDI [56]	Multivariate	Conditional	Imputation	General	Stanford	NeurIPS	2021
SF-DM [169]	Multivariate	Conditional	Classification	Manufacture	Aalto	UBICOMP	2023
TimeDiff [34]	Multivariate	Conditional	Forecasting	General	HKUST	ICML	2023
DiffAD [260]	Multivariate	Conditional	Anomaly Detection	General	HENU	KDD	2023
D^{3}R [261]	Multivariate	Conditional	Anomaly Detection	General	BUPT	NeurIPS	2023
CLDM [107]	Multivariate	Conditional	Generation	Energy	NCEPU	IEEE T SUSTAIN ENERG	2023
MIDM [262]	Multivariate	Conditional	Imputation	General	USTC	KDD	2023
MEDiC [110]	Multivariate	Conditional	Imputation	Healthcare	IIT	NeurIPS	2023
VGCDM [263]	Multivariate	Conditional	Anomaly Detection	Electricity	XJTU	ArXiv	2023
DiffPLF [143]	Multivariate	Conditional	Forecasting	Electricity	HKUST(GZ)	ArXiv	2024
Time Weaver [182]	Multivariate	Conditional	Generation	General	UT-Austin	ArXiv	2024
Fu et al. [266]	Multivariate	Conditional	Generation	Electricity	NUS	ArXiv	2024
ImDiffusion [257]	Multivariate	Conditional	Anomaly Detection	General	PKU	VLDB	2024
Wang [264]	Multivariate	Conditional	Forecasting	Business	USTC	ICASSP	2024
DiffDA [26]	Multivariate	Conditional	Forecasting	Climate	ETH Zurich	ArXiv	2024
BioDiffusion [51]	Multivariate	Conditional	Generation	Healthcare	TSU	ArXiv	2024
DiffSTOCK [41]	Multivariate	Conditional	Forecasting	Finance	Purdue	ICASSP	2024
RF-Diffusion [44]	Multivariate	Conditional	Generation	Network	THU	MobiCom	2024
Klein et al. [42]	Multivariate	Conditional	Generation	Healthcare	Radboud	ArXiv	2024
DiffSTG [43]	Spatio-temporal	Conditional	Forecasting	General	BJTU	SIGSPATIAL	2023
PriSTI [111]	Spatio-temporal	Conditional	Imputation	General	Beihang	ICDE	2023
DiffUFlow [35]	Spatio-temporal	Conditional	Forecasting	Transportation	CSU	CIKM	2023
DiffTraj [81]	Spatio-temporal	Conditional	Generation	Transportation	SUSTech	NeurIPS	2023
ControlTraj [290]	Spatio-temporal	Conditional	Generation	Transportation	HKUST(GZ)	arXiv	2024
USTD [170]	Spatio-temporal	Conditional	Forecasting	General	NUS	ArXiv	2023
IDM [184]	Spatio-temporal	Conditional	Forecasting	Transportation	ZJU	ArXiv	2024
Diff-RNTraj [17]	Spatio-temporal	Conditional	Generation	Transportation	BJTU	ArXiv	2024
Evans et al. [185]	Univariate	LDM	Generation	Audio	Stability AI	ArXiv	2024
LDT [253]	Multivariate	LDM	Forecasting	General	NTU	AAAI	2024
Aristimunha et al. [128]	Multivariate	LDM	Generation	Healthcare	UPSaclay	NeurIPS	2023
TSDM [269]	Multivariate	DDIM	Imputation&Anomaly Detection	Electricity	HUST	ArXiv	2023
LADM [265]	Spatio-temporal	LDM	Forecasting	Transportation	XJU	IEEE T INSTRUM MEAS	2024

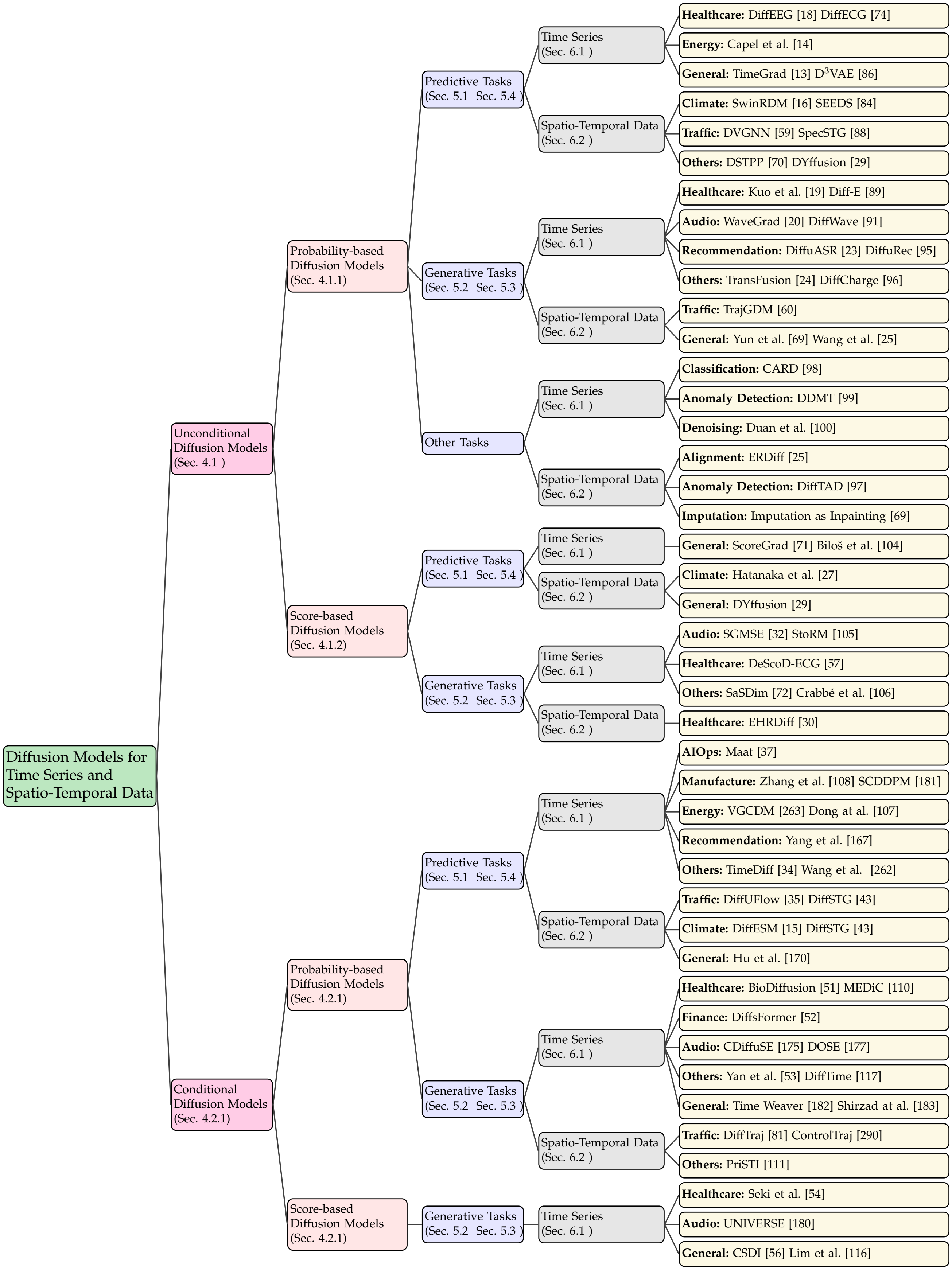

Fig. 7: A comprehensive taxonomy of diffusion models for time series and spatio-temporal data, categorized according to methodologies (i.e., unconditional vs. conditional), tasks (e.g., predictive versus generative), data types, and applications.

decomposing the diffusion process into a fixed number of steps. This method adds noise to the data at discrete intervals and then learns to reverse this process, generating data from noise. Both the training and inference process of the model are conducted on these discrete steps, making the process essentially stepwise and iterative.

TimeGrad is among the most classic DDPM-based time series forecasting methods, identified as a score-matching model that, through rigorous validation, has been proven effective in real-world datasets comprising thousands of interconnected dimensions [13]. D3VAE proposes a bidirectional variational auto-encoder that integrates diffusion, denoising, and disentanglement to enhance time series data without introducing extraneous uncertainties. This model marries multi-scale denoising score matching with bi-directional variational autoencoding for forecasting tasks [86]. In a similar vein, TSDiff leverages implicit probability densities to iteratively refine the base forecasts, thereby addressing three distinct tasks: forecasting, refining, and generating synthetic data [164]. For the universal challenge of anomaly detection, Pintilie et al. have proposed two multivariate time series anomaly detection algorithms based on diffusion models, securing leading outcomes [259]. Concurrently, D^3R tackles temporal anomaly detection drifts through decomposition and reconstruction, employing data-time mix-attention for dynamic decomposition alongside diffusion models for end-to-end training, thus overcoming the limitations imposed by local sliding windows and unstable data scenarios [261]. CARD presents a denoising diffusion-based generative model alongside a pretrained conditional mean estimator to serve both classification and regression tasks [98]. TS-Diffusion is specifically designed for complex sequences marked by irregular sampling, missingness, and extensive temporal feature dimensions, introducing a comprehensive model consisting of an ODE encoder, a representational learning module, and an ODE decoder to handle such intricate time series [278]. Furthermore, WaveGrad [20], DiffWave [91], and DiffuASR [23] have each successfully applied DDPM to waveforms, audio generation, and sequence recommendation, respectively.

On the other hand, in the field of spatio-temporal data, significant advances have been made with DDPM predominantly starting from 2023, indicating an exciting opportunity for more applications and theoretical breakthroughs. Yun et al. have put forward "Imputation as Inpainting," which integrates an unconditional diffusion model based on graph neural networks to first forecast complete spatio-temporal data. This approach adjusts the generation process by sampling in unobserved areas using the information from observed data, thus estimating missing values in spatio-temporal data [69]. For traffic and specifically trajectory forecasting, the adoption of graph-based DDPM methods is becoming more common. For example, MID++ leverages a GNN-based ALEncoder to extract spatio-temporal features and improves the training process with an important sampling strategy, leading to better trajectory forecasting [279]. In a similar vein, DiffTAD incorporates a Transformer-based decoupled time and space encoder to model spatial interactions among vehicles and conducts anomaly detection by evaluating the differences between query trajectories and their reconstructions [83]. Moreover, SpecSTG tackles the challenges of insufficient representation of spatial network features and the inability to detect unexpected fluctuations in future observations in traffic flow forecasting. It achieves this by converting the learning process to the spectral domain, generating Fourier representations of future time series that carry spatial information [88].

4.1.2 Score-Based Model

Unlike the discrete DDPM, score-based SDE models represent diffusion and the reverse process in a ***continuous*** form using stochastic differential equations, thereby covering continuous time. This approach allows for a theoretically more flexible and in-depth treatment of the diffusion process, capable of generating samples at any point in time, rather than being confined to fixed steps.

ScoreGrad and TimeGrad share similar goals and can be described as SDE-based TimeGrad models. They expand the diffusion process into a continuous spectrum and employ time series feature extraction modules along with a condition-based stochastic differential equation for score matching to facilitate forecasting [71]. This structure finds parallels in the work of Biloš et al. [104]. Additionally, recent studies by Crabbé et al. delve into representing time series data in the frequency domain using SDEs, showcasing how denoising score matching methods can facilitate diffusion modeling within the frequency domain based on differing time-frequency spaces [106]. TimeADDM introduces a score-based diffusion model for unsupervised anomaly detection in multivariate time series, applying diffusion steps to representations that encapsulate global time correlations through recurrent embeddings and designing a suite of reconstruction strategies to compute anomaly scores at various diffusion intensities [282]. Moreover, Li et al. extensively discuss the impact of noise samples in SDEs on data quality and the corresponding robustness of models [280].

From an application standpoint, score-based diffusion models have shown great promise in continuous data fields such as healthcare and acoustic data. EHRDiff [30] and DeScoD-ECG [57] leverage score-based diffusion models for realistic EHR synthesis and ECG data generation, respectively, incorporating algorithms or structures pertinent to the features of medical data, thus improving applicability in real-world settings. In audio-related applications, the focus has been on speech enhancement and dereverberation, with initial proposals by Welker et al. [284] and Richter et al. [32] for employing score-based diffusion models for speech enhancement, transforming time domain speech data into time-frequency information, and treating them as image input for models. StoRM introduces a stochastic resampling method, utilizing forecasting from a predictive model as guidance for subsequent diffusion steps, achieving higher-quality samples under complex conditions with high signal-to-noise ratios using fewer diffusion steps and more streamlined models [105]. Similarly, Lay et al. emphasize improving model efficiency and optimizing model parameters, proposing a Brownian bridge-based forward process to reconcile the gap between the forward process's end distribution and the prior distribution used in inference for the reverse process [178].

In the realm of spatio-temporal data, methods remain scarce, with Sasdim and DYffusion standing out. Sasdim, an adaptive noise scaling diffusion model, effectively performs spatial-temporal data imputation by capturing dynamic spatio-temporal dependencies through novel loss functions and global spatio-temporal convolution modules [72]. DYffusion, a dynamics-informed diffusion model for spatio-temporal data forecasting, distinguishes itself from probabilistic models by integrating temporal dynamics directly into the diffusion steps and training a stochastic, time-conditioned interpolator alongside a predictor network to simulate the forward and reverse processes of the standard diffusion model, thus striking a balance between performance and efficiency [285].

4.2 Improved/Advanced Diffusion Model

We have already discussed the improvement methods of diffusion model in Sec. 2.3.3 and Sec. 2.3.4. In the following, we will analyze the methods from the more frequently used ones.

4.2.1 Conditional Diffusion Model

Compared to standard approaches, conditional diffusion models are able to use given ***conditional information*** (e.g., different representations, different modalities, etc.) to steer the generative process and produce high-quality outputs that are tightly correlated with conditions [34]. More condition-based models are used for high-quality generation tasks.

CSDI and MIDM use conditional diffusion models for time series imputation [56], [262]. They utilize score-based and probability-based diffusion models conditioned on observed values and through explicit imputation training. They can leverage the correlations among observed values to further enhance performance. Similar to CSDI, DiffAD and ImDiffusion employ a similar method based on conditional weight-incremental diffusion to enhance the imputation performance of missing values and are applied for time series anomaly detection [257], [260]. These approaches preserve the information of observed values and significantly improve the generation quality for stable anomaly detection. On the other hand, there are models that utilize different guiding information [266]. For example, [169] adjusts the diffusion model based on statistical information such as mean, standard deviation, Z-scores, and skewness, thereby synthesizing sensor data. [177] proposes two different conditional enhancement techniques to enable the model to adaptively consider conditional information a priori, thus performing the speech enhancement task. MEDiC [110] introduces a class-conditional DDPM approach to generate synthetic EEG embeddings. VGCDM [263] employs a pulse voltage-guided diffusion model along with a cross-attention mechanism for more efficient generation of electrical signals. Meanwhile, Wang et al. leverage multimodal information (images and text) as generative conditions to enhance the quality of real-time sales forecasts [264]. DiffShape [277] introduces a self-supervised diffusion learning mechanism that uses real sub-sequences as conditions. By leveraging a large amount of unlabeled data, it enhances the similarity between learned shapelets and actual sub-sequences for classification.

For spatio-temporal data, most models adopt graph structures for representation, but the method of using conditions is similar to that of time series data [85], [184]. PriSTI draws inspiration from CSDI of conditioning on the observed value, employing a conditional feature extraction module based on linear interpolation for generating missing data [111]. DiffTraj estimates noise levels accurately using various external factors, such as the region of the trip and departure time [81]. ControlTraj further extends DiffTraj with the constraint of road network structures [290]. Additionally, many works guide their models using graph structures or feature graphs as conditions. For instance, DiffSTG [43] and USTD [170] use historical graph signals and graph structures as conditions, combining the spatio-temporal learning capabilities of GNNs and the uncertainty measurement of diffusion models to enhance the performance of forecasting tasks. DiffUFlow [35] utilizes extracted spatio-temporal feature graphs overlaid on coarse-grained flow graphs as a condition to guide the reverse process.

4.2.2 Latent Diffusion Model

The latent diffusion model (LDM) performs the diffusion process in a lower-dimensional latent space, which allows more ***efficient*** generation or other tasks. It allows the model to handle more complex data distributions, while reducing computational resource consumption and maintaining output quality.

LDCast utilizes LDM for near-term precipitation forecasting, enhancing training stability, optimizing computational demands, and accurately representing the uncertainty of forecasting compared to GAN [93]. CLDM presents a conditional latent diffusion model based on latent space mapping, which breaks down the generative task into deterministic forecasting and the generation of predictive error scenarios, thus increasing efficiency within the latent space [107]. Similarly, Aristimunha et al. suggest the use of LDM for generating sleep EEG signals, where the latent feature maps outputted by encoders serve as input to the diffusion model, with the generated results obtained through the corresponding decoder [128]. In a similar manner, LADM proposes a method for spatio-temporal trajectory forecasting, where VAE acts as a generator and DDPM as a refiner [265]. Furthermore, Feng et al. develop a latent diffusion transformer for time series forecasting, aimed at compressing multivariate time stamp patterns into succinct latent representations and efficiently generating authentic multivariate time stamp values in a continuous latent space [253].

4.2.3 Other Variants of Diffusion Models

There are also other methods that incorporate the different models above or use other more advanced diffusion models. For example, to accelerate generation, MedDiff for the first time utilizes DDIM with a new sampling strategy to generate high-dimensional large-scale electronic medical records [48]. Specifically, it incorporates a classifier to guide the sampling process, assigning high probabilities to the data with the correct labels. DCM is a diffusion-based causal inference model accelerated by DDIM, that more accurately captures counterfactual distributions under unmeasured confounding factors [267]. Asperti et al. propose GED, which uses conditional DDIM integrated with additional information and a post-processing network for high-performance weather forecasting [268].

Besides, some algorithms have also been proposed that use a two-stage or a diffusion model to fuse with other models. TSDM adopts an improved two-stage diffusion model for identifying and reconstructing measurements with various uncertainties for power system measurement recovery, where the first stage includes a classifier-guided conditional anomaly detection component, and the second stage involves a diffusion-based measurement imputation component [269]. Mueller proposes using attention-enhanced condition-guided DDIM to tackle sample imbalance and data scarcity issues, along with simple DDPM for signal denoising, thereby synthesizing effective machine fault data for efficient anomaly detection [270]. Similarly, Wong et al. utilize DDIM for load domain adaptation, further enhanced with CNN for bearing fault anomaly detection [272].

5 Task Perspective

In this section, we will explore the use of diffusion models in different tasks, such as forecasting, generation, imputation, and anomaly detection, also highlighting their effectiveness in complex time series and spatio-temporal data analysis across various domains.

5.1 Forecasting

The field of time series forecasting has seen significant advancements with the incorporation of diffusion models. TimeGrad [13] and D3VAE [86] both employ diffusion probabilistic models to enhance forecasting, with TimeGrad focusing on autoregressive techniques (i.e., RNN) for probabilistic forecasting and D3VAE introducing a bidirectional variational auto-encoder that includes diffusion and denoising processes. This generative approach is further extended in the work by D-Va [47], which addresses the stochastic nature of stock price data through a deep hierarchical VAE combined with diffusion probabilistic techniques. Meanwhile, the study presented in [104] takes a different approach by modeling temporal data as continuous functions, allowing for the handling of irregularly sampled data. Building upon the diffusion model, TimeDiff [34] introduces novel conditioning mechanisms, future mixup, and autoregressive initialization, to improve time series forecasting. Lastly, the research in [164] explores task-agnostic unconditional diffusion models, proposing TSDiff, which employs a self-guidance mechanism for versatile time series applications.

For spatio-temporal data, various diffusion model-based approaches have been proposed to tackle complex forecasting problems. DiffSTG [43] is the first work that generalizes denoising diffusion probabilistic models to spatio-temporal graphs, aiming to model complex spatio-temporal dependencies and intrinsic uncertainties within spatio-temporal graph data for better forecasting. DYffusion [285] introduces a framework that trains a stochastic, time-conditioned interpolator and a forecaster network to perform multi-step and long-range probabilistic forecasting for spatio-temporal data. On the other hand, DiffUFlow [35] focuses on urban data, which aims to address the challenge of fine-grained flow inference. DSTPP [70] provides a novel parameterization for spatio-temporal point processes. Meanwhile, SwinRDM [16] and DOT [65] demonstrate the adaptability of diffusion models in improving the quality of weather forecasts and travel time estimations, showcasing the wide applicability of these models across different domains and tasks. These works highlight the growing impact and potential of diffusion models in advancing spatio-temporal data analysis.

5.2 Generation

Inspired by the powerful ability of the diffusion model in high-dimensional distribution learning, an intuitive usage is applying the learned diffusion model for data generation. Currently, a variety of diffusion models have been proposed for generating audio data. WaveGrad [20] is a pioneering work that utilizes gradient-based sampling to generate high-fidelity audio waveforms, marking a significant advancement in audio synthesis. Similarly, DiffWave [91] employs a non-autoregressive approach to achieve efficient and high-quality raw audio synthesis through a Markov chain process. Both WaveGrad and DiffWave are part of the same research lineage, leveraging the power of diffusion models to create complex waveforms from simple noise distributions. DOSE [177] takes a different approach by focusing on the speech enhancement task, which introduces a model-agnostic method that integrates condition information into the diffusion process.

Besides, diffusion model is also used in sequential recommendation. DiffuASR [23] proposes a diffusion-based sequence generation framework to address the data sparsity and long-tail user problems in sequential recommendation systems. It introduces a sequential U-Net designed for discrete sequence generation tasks and utilizes two guide strategies to assimilate preferences between generated and original sequences. DreamRec [167] employs a Transformer encoder to create guidance representations as the condition in the diffusion process.

In the realm of spatial-temporal data, diffusion model is adopted to generate trajectory data. For instance, DiffTraj [81], [290] proposes the first effort that generates high-quality human trajectory with an unconditioned model, with the motivation to protect privacy. Meanwhile, [65] generates trajectories in the format of grids, by a condition model guided by the origin-destination information.

5.3 Imputation

In the domain of time series and spatial-temporal data analysis, imputation refers to generating the unobserved data conditioned on the given observed data. CSDI [56] proposes a score-based diffusion model that probabilistically imputes missing time series and spatio-temporal data. MIDM [262] redefines the evidence lower bound (ELBO) for conditional diffusion models tailored for multivariate time series imputation, which ensures the consistency of observed and missing values. PriSTI [111] introduces a conditional diffusion framework specifically designed for spatio-temporal data imputation, using global context priors and geographic relationships to tackle scenarios with high missing data rates due to sensor failures. More recently, TabCSDI [33] and MissDiff [113] leverage diffusion models for imputing missing values in tabular data, each addressing distinct aspects of this complex challenge. TabCSDI manages categorical and numerical data effectively, providing a tailored solution to diverse data types common in tabular datasets. On the other hand, MissDiff enhances the diffusion model's training process by introducing a novel masking technique for regression loss, which ensures consistent learning of data distributions and robustness against different missing data scenarios.

5.4 Anomaly Detection

Anomaly detection aims to identify anomalies in the given time series or spatial-temporal data, which is quite practical and crucial for many real-world applications. A majority of works have been proposed for general time series anomaly detection. Initially, DiffAD [260] and ImDiffusion [257] both explored the synergy of imputation techniques with diffusion models for time series anomaly detection, enhancing the robustness of anomaly detection processes by accurately modeling complex dependencies. Concurrently, [259] and [258] use similar diffusion-based methodologies to address anomaly detection but apply different enhancements to improve performance and computational efficiency. Meanwhile, [261] and DDMT [99] tackled the specific challenges of instability and noise, implementing advanced diffusion reconstruction techniques to maintain accuracy in dynamic environments.

Besides the above methods designed for general time series anomaly detection. There are other works that apply advanced diffusion techniques across various domains. For instance, Maat [37] anticipates performance metric anomalies in cloud services using a conditional denoising diffusion model to forecast metrics and detect anomalies. [270] enhances data synthesis for machine fault diagnosis using an attention mechanism within a conditional diffusion model framework. Diffusion-UDA [271] proposes a diffusion-based method for unsupervised domain adaptation in submersible fault diagnosis, leveraging diffusion processes to adapt domains for effective fault recognition.

6 Data Perspective

In this section, we analyze current work from a data perspective, including time series and spatio-temporal data. We focus on introducing how diffusion models capture the unique properties of different data modalities.

6.1 Time Series Data

Time series analysis is a critical issue explored across various real-world scenarios, including retail sales forecasting, filling in missing data in economic time series, identifying anomalies in industrial maintenance, and categorizing time series from different domains.

6.1.1 Univariate Time Series

Univariate time series is characterized by having only one variable of interest observed over a period of time. Data in this category include ECG signal, audio, electricity load, etc. The complex sequential patterns (e.g., the trend and the periodicity) are the core data property in univariate time series, which pose great challenges in developing relevant diffusion models.

Diffusion Model for Univariate Time Series. Diffusion models for univariate time series are primarily developed to model the uncertainty present in the data, facilitating tasks such as probabilistic forecasting or data generation. An essential element within the diffusion model is the denoising network, which determines the extent of noise removal at each stage. In the context of univariate time series, the denoising network commonly employs the 1-D CNN to identify sequential patterns in the input data. For instance, both WaveGrad [20] and DiffWave [91] incorporate a 1-D CNN in the denoising network to extract local sequential features from audio sequences. Furthermore, some studies opt to transform the univariate time series into the frequency domain initially to enhance the capture of long-term sequential correlations from a global standpoint. This transformation converts the time series from 1-D to 2-D data, enabling the use of a 2-D CNN in the denoising network to capture correlations within the frequency domain, such as [47] for the stock price forecasting and DiffLoad [28] for electricity load forecasting.

6.1.2 Multivariate Time Series

Multivariate time series is characterized by having multiple variables of interest observed over the same period of time. Instead of the sequential patterns in each variate, multivariate time series also leverage the interdependencies among different variables to capture more comprehensive information downstream tasks.

Diffusion Models for Multivaritate Time Series. Beyond approaches that focus on univariate time series, there have been efforts towards multivariate time series. Multivariate time series is naturally a 2-D data similar to the image data, where the signal of each variate can be considered as one channel in the image. To this end, diffusion models for multivariate time series usually adopt the vanilla Unet as the denoising net, which is a common practice in diffusion models for images.

For example, [5] first transforms demultiplex, denoising, and interpolation into image-to-image transformation tasks, then leverages the diffusion model for different tasks. In the demultiplex task, the diffusion model learns to remove the multiples without removing primary energy. In the denoising and interpolation task, the diffusion model learns to eliminate undesired uncorrelated noise and recover the missing values, while preserving the inherent characteristics of the data. MIDM [262] proposes a novel multivariate imputation diffusion model that incorporates correlations between observed and missing values. By re-deriving the ELBO of the conditional diffusion model, MIDM ensures consistency between observed and missing data, thus leading to improved imputation accuracy. Diff-E [109] further utilizes the diffusion model for representation learning, which combines DDPMs with a conditional autoencoder to enhance the decoding performance of speech-related EEG signals.

6.2 Spatio-Temporal Data

In real-world systems, a myriad of elements interact with each other both spatially and temporally, resulting in a spatio-temporal composition. Spatio-temporal data (STD) is the de facto most popular data structure for injecting such structural information into the formulation of practical problems. In this section, we introduce developments of diffusion models for spatial-temporal data, mainly with two data modalities: (1) the Spatio-Temporal Graph (STG) and (2) the Spatio-Temporal Trajectory (STT). Unlike diffusion models for time series data, existing works in STD need to model the temporal dependencies as well as the spatial dependencies, making the task even more challenging.

6.2.1 Spatio-Temporal Graph

The spatio-temporal graph is generated by sensors at different places in a period of time, where the correlation of those sensors is typically described as a graph. We categorize diffusion models for spatial-temporal graph into domain-oriented and domain-agnostic works.

Most of the existing works fall into the category of domain-oriented spatial-temporal graph diffusion models. They typically leverage the powerful distribution learning abilities of diffusion models for spatial-temporal data mining tasks in specific domains, such as traffic and climate. In this research trend, the majority of works come from the traffic domain. DiffUFlow [35] represents pioneering efforts as traffic diffusion models, which convert the fine-grained urban flow inference as a denoising diffusion process. SpecSTG [88] advances the field by conducting the diffusion process in the spectral space projected by the STG, resulting in faster inference speeds. In a similar vein, [59] uses a diffusion model to infer the link probability and reconstruct causal graphs in the decoder stage adaptively for the STG forecasting task. Furthermore, [65] solves the origin-destination travel time estimation problem with a two-step process, where the first step predicts the possible travel route with a diffusion model conditioned on the given origin-destination pair. Beyond the traffic domain, there have been initiatives towards the climate. A notable example is [78], which quantifies the uncertainty of air quality forecasting based on the spatial-temporal diffusion model. Similarly, SRNDiff [85] conducts the short-term rainfall nowcasting through the conditional diffusion model.

In the second group, researchers focus on developing domain-agnostic models, which can achieve promising performance across a variety of domains. For instance, DiffSTG represents a pioneering effort by introducing a unified diffusion framework for multiple STG tasks such as forecasting and imputation. Concurrently, PriSTI [111] and [69] study the spatial-temporal imputation task while treating the task of recovering unobserved values as a denoising process conditioned on the observed values. Similarly, DYffusion [29] models the temporal dynamics directly within diffusion steps, leading to a stochastic, time conditioned forecasting network. Another noteworthy contribution is [170] under the domain-agnostic and task-agnostic setting, which aims to harness the power of diffusion models, and unify diffusion models for probabilistic spatio-temproal graph learning.

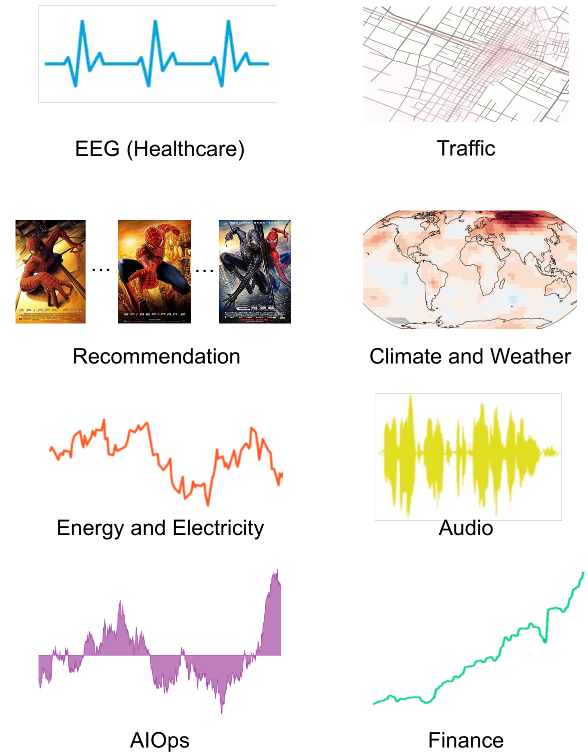

Fig. 8: Examples of time series and spatio-temporal data for different application scenarios.

6.2.2 Spatio-Temporal Trajectory

Spatial-temporal trajectory is a sequence of locations ordered by time that describe the movements of an object in a geographical space. Analysis of spatial-temporal modality is particularly crucial to discovering the mobility patterns of moving objects, which serves as the foundation for many downstream tasks, such as POI recommendation and next-location forecasting. In terms of diffusion models for trajectory data, most works are developed to solve the trajectory generation task. They leverage the ability of the diffusion model to learn high-dimensional data distribution while injecting spatial-temporal correlation into the diffusion process. As exemplified by DiffTraj [81], which generates the GPS trajectory with the diffusion probabilistic model in an unconditioned manner. Similarly, [60] utilizes the diffusion model for generating high-quality human mobility data. More recently, Diff-RNTraj [17] further generates the trajectory constrained by the road-network. Beyond generation, there is a growing interest in applying diffusion models for trajectory forecasting tasks, and [66] and [184] exemplify this trend.

7 Application Perspective

In this section, we summarize and discuss the most important applications of diffusion models in time series and spatio-temporal data, including healthcare, smart city, recommendation, climate and weather, energy and electricity, audio, video and so on. The toy signal examples of each application are shown in Fig. 8 .

In detail, we will discuss the models in different applications, aiming to show light on the model design of practical

TABLE 2: Summary of dataset resources in different application domains.

Applications	Dataset	Description	Timeframe	Source	Citations
Healthcare	Seed V dataset	From 16 participants in three emotions test sessions through video presentation.	Each test is around 50 minutes.	[118]	[119]
	QT Database	105 fifteen-minute excerpts of two-channel ECG Holter recordings. Examples of each morphology were included in this subset of annotated beats; at least 30 beats in each record, 3622 beats in all, were manually annotated in the database.	105*15 minutes	[120]	[57]
	MIT-BIH NST dataset	12 half-hour ECG recordings and 3 half-hour recordings of noise typical in ambulatory ECG recordings.	15 half hour	[121]	[57]
	PTB-XL dataset	Collection of clinical 12-lead ECG data comprising 21,837 records from 18,885 patients.	Each 10 seconds length.	[122]	[123], [124]
	MIT-BIH dataset	Healthy with 90,589 ob- servations, atrial premature with 2,779 observations, ventricular malfunction with 7,236 observations, fusion of ventricular and normal with 803 observations and a last class of 8,039 unclassified observations.	48 half-hour excerpts of two-channel ambulatory ECG recordings.	[125]	[124]
	Alzheimer's Disease EEG dataset	88 participants, categorized into three groups: 36 individuals diagnosed with Alzheimer's disease(AD group), 23 diagnosed with Frontotemporal Dementia (FTD group), and 29 healthy subjects (CN group).	The duration of the disease was measured in months and the median value was 25 with IQR range (Q1-Q3) being 24 - 28.5 months.	[126]	[110]
Traffic	Cab Trajectory datasets	Real World Trajectory data of Chengdu, Harbin and Xi'an City from Didi Chuxing, around 5.6 million trajectory number.	Starting from November 1, 2016, to November 30, 2016.	[81]	[17], [65], [81]
	PeMS08	The traffic data are aggregated into every 5-minute interval from the raw data collected by the Caltrans Performance Measurement System.	From July to August in 2016, containing 1979 detectors on 8 roads.	[82]	[59], [62], [63]
	MAAD-Highway dataset	A dataset for multi-agent anomaly detection based on the OpenAI Gym MultiCarRacing-v0 environment.	The sequences are sub-sampled to 10 Hz with a segment length of T = 1.5 seconds.	[64]	[66], [67]
	LargeST	Traffic flow data of California, including a total number of 8,600 sensors over road networks.	Starting from January 1, 2017, to December 31, 2021.	[283]	[283]
Recommendation	Amazon-book	This dataset contains product reviews and metadata from Amazon, including 142.8 million reviews.	May 1996 - July 2014.	\	[132], [135]
	Yelp	908,915 tips by 1,987,897 users Over 1.2 million business attributes like hours, parking, availability, and ambience Aggregated check-ins over time for each of the 131,930 businesses	\	\	[132]
	ML-IM	1 million ratings from 6000 users on 4000 movies.	\	[133]	[95], [132], [134]
	Amazon-beauty	Over 2 Million+ customer reviews and ratings of Beauty related products sold on their website.	\	\	[36], [95], [134]
	Amazon-toys	10,000 toy products on Amazon.com	\	\	[36], [95], [134], [135]
Climate and Weather	ERA5	The fifth generation ECMWF atmospheric reanalysis of the global climate covering the period from January 1940 to present. ERA5 provides hourly estimates of a large number of atmospheric, land and oceanic climate variables. The data cover the Earth on a 31km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80km. ERA5 includes information about uncertainties for all variables at reduced spatial and temporal resolutions.	1940 to present.	[136]	[16], [27], [84], [137]
	Data from IPSL-CM5A ESM	A full earth system climate.	\	[138]	[15]
	GOES West data	Highresolution atmospheric measurements over the Pacific Ocean in near real-time with the explicit goal of improving weather forecasting capabilities.	\	\	[27]
Energy and Electricity	GEFCom2014 Load Forecasting Data	\	7 years of matching load and temperature data	[139]	[28], [107]
	ACN-Data	Collect detailed data about each charging session in the system.	\	[140]	[96]
	Low Carbon London Project Data	Contains electricity consumption data of 5198 customers from November 2011 to February 2014 with a time granularity of 30 minutes.	November 2011 to February 2014	[141]	[142]
	Low Carbon London Project Data	Contains electricity consumption data of 5198 customers from November 2011 to February 2014 with a time granularity of 30 minutes.	November 2011 to February 2014	[141]	[142]
Audio	LJ Speech dataset	13,100 short audio clips of a single speaker reading passages from 7 non-fiction books.	The texts were published between 1884 and 1964, and are in the public domain. The audio was recorded in 2016-17 by the LibriVox project and is also in the public domain.	[68]	[20], [21], [22], [91]
	SC09 dataset	31,158 training utterances (8.7 hours in total) by 2,032 speakers.	\	[273]	[91], [274]

applications. Besides, the main datasets in each application domains are shown in Tab. 2 , offering convenience for future research.

7.1 Healthcare

In recent years, diffusion models have emerged as a powerful class of generative models with significant implications for time series in healthcare. These models, known for their ability to generate high-fidelity samples through a process of gradual refinement, have found diverse applications ranging from the synthesis of electronic health records (EHRs) to the enhancement of biomedical signal analysis. The following are some main topics in this area:.

- **EEG Signal Synthesis and Enhancement:** [74] introduces a generalized probabilistic diffusion model tailored for the synthesis of electrocardiogram (ECG) signals, aiming to support cardiac health research and diagnostic training with realistic data generation. [128] demonstrate the potential of diffusion models to create synthetic EEG signals, facilitating sleep disorder studies and neurological research. [57], [123] focus on enhancing the quality of ECG signals, addressing issues like baseline wander and noise, and generating conditional ECG signals based on specific patient states.
- **Healthcare Data Augmentation and Synthesis:** [18] highlights the use of diffusion models to augment datasets for seizure forecasting, thereby improving the robustness and accuracy of predictive models. [30], [31] explore the synthesis of realistic EHRs, tackling the challenges of data privacy and scarcity by providing synthetic datasets for research and model training.
- **Diagnostic and Predictive Analytics:** [50] leverages generative models to enhance the detection of autism, showcasing the broad applicability of these models in diagnosing and understanding complex conditions. In [49], diffusion models are applied to forecast critical physiological parameters in ICUs, demonstrating their potential to save lives by predicting adverse events before they occur.
- **Novel Methodologies and Techniques:** [58] exemplifies the versatility of diffusion models in handling multiple tasks simultaneously, such as predicting mortality in critically ill patients, showcasing the potential for comprehensive care management systems.

7.2 Traffic

There are also many works for traffic applications using diffusion models. The following are some of them. DiffUFlow [35] employs a denoising diffusion model to enable the model to capture the stochastic nature of urban flows and generate accurate and fine-grained forecasting. DiffTraj [81] presents a groundbreaking approach to generating GPS trajectories using a spatial-temporal diffusion probabilistic model. The model works by reconstructing geographic trajectories from white noise through a reverse trajectory denoising process, effectively turning random noise into meaningful trajectory data that reflects real-world movement patterns. DVGNN architecture [59] comprises an encoder component that learns latent node embeddings through GCN layers and a decoder component that treats the relationships of latent states as a diffusion process subject to stochastic differential equations. This allows for the inference of internal causal relationships among neighbor nodes and the formation of dynamic causal graphs. [61] addresses the growing risk of collisions between space objects by developing a diffusion model that predicts the positional uncertainty of objects during close encounters. [83] introduces a model that utilizes denoising diffusion probabilistic models for identifying anomalies in vehicle trajectories, aimed at improving the detection of unusual patterns in vehicle movements. SpecSTG [88] aims to efficiently handle the complexities of spatio-temporal data, offering a probabilistic approach to forecasting traffic flows. This method is distinguished by its speed and accuracy in generating forecasts and is a promising solution for real-time traffic management and planning applications.

7.3 Sequential Recommendation

In the scenario of sequential recommendation, diffusion models are also widely applied. [23] and [132] both explore the application of diffusion models in enhancing sequential recommendation systems, yet they focus on different aspects and methodologies within this domain. That is, [23] aims to bridge the gap between discrete item identities and the continuous nature of the data generated by diffusion models, while [132] focuses on predicting users' future interaction probabilities by corrupting and then denoising their interaction histories and address challenges specific, such as high resource costs and temporal shifts in user preferences. DCDR [131] is a discrete forward process with tractable posteriors and a conditional reverse process tailored for sequence generation. RecFusion [130] introduces a binomial diffusion process tailored for one-dimensional data, such as time series of user interactions, showcasing its effectiveness in sequential recommendation scenarios. DiffuRec [95] proposes a specific diffusion model framework designed for sequential recommendation, focusing on efficiently capturing and predicting evolving user interests.

Diffusion models have shown significant promise in improving sequential recommendation systems. Their ability to simulate the spread of information and preferences offers a nuanced method for predicting user behavior. Future research could explore hybrid models, scalability, and real-time adaptation, further enhancing the relevance and personalization of recommendations.

7.4 Climate and Weather

Diffusion models, initially developed for image generation, have found novel applications in weather forecasting due to their ability to generate high-resolution, realistic outputs. These models work by gradually refining a signal from a random noise distribution towards a desired output, making them well-suited for predicting complex atmospheric phenomena. [93] focuses on precipitation nowcasting with an emphasis on accurate uncertainty quantification, leveraging latent space representations to efficiently model and predict rainfall intensity and distribution. [94] presents a wind resolution-enhancing model that improves the detail and accuracy of wind speed forecasts by refining coarse forecasting to higher resolutions. SwinRDM [16] integrates

Swin Recurrent Neural Networks (SwinRNN) with diffusion models for state-of-the-art weather forecasting, achieving high resolution and quality by capturing both spatial and temporal dynamics. DiffESM [15] utilizes diffusion models for the conditional emulation of Earth System Models (ESMs), offering a computationally efficient alternative to traditional simulation techniques with improved accuracy. [27] addresses the challenge of high-resolution solar forecasting, showcasing the model's ability to accurately predict solar irradiance variations. DiffAD [26] introduces a diffusion model-based approach for weather-scale data assimilation, enhancing the integration of observational data into forecast models for improved accuracy.

7.5 Energy and Electricity

Diffusion models are now being adapted to tackle challenges in energy systems, offering novel solutions for probabilistic forecasting, signal synthesis, and system security. DiffLoad [28] is a novel approach to load forecasting that leverages diffusion models to quantify uncertainties effectively. By incorporating uncertainty quantification, DiffLoad offers a more reliable forecast, which is crucial for grid stability and operational planning. [142] explores how conditional diffusion models can synthesize load profiles for individual electricity customers, enhancing the personalization of demand-side management strategies. DiffPLF [143] addresses the challenges for load forecasting by providing probabilistic forecasts of EV charging loads, facilitating better grid management and infrastructure planning. [14] discusses the application of denoising diffusion probabilistic models in energy forecasting, emphasizing the model's ability to handle uncertainty and provide probabilistic forecasts, which are essential for integrating renewable energy sources. [107] demonstrates how conditional latent diffusion models can be used to generate realistic short-term wind power scenarios, aiding in system operation and planning.

7.6 Audio

Diffusion models have also found promising applications in generating waveforms, synthesizing audio, music generation and enhancing speech, offering significant improvements in quality, realism, and control over the generated or enhanced audio.

- **Waveform Generation and Audio Synthesis** WaveGrad [20] introduces a gradient-based approach to generate high-fidelity waveforms, demonstrating the potential of diffusion models in audio synthesis without the need for autoregressive models. DiffWave [91] extends the capabilities of diffusion models in audio synthesis, showcasing their versatility across different audio synthesis tasks, including voice and music.
- **Music Generation** [92] explores the fusion of diffusion models and GANs for generating symbolic music, allowing for emotion-driven control over the generation process. DiffuseRoll [90] highlights the application of diffusion models in multi-track music generation, providing nuanced control over various musical attributes.
- **Speech Enhancement** [32] introduces a diffusion-based generative approach to simultaneously enhance speech and reduce reverberation, showcasing significant improvements over traditional methods. [180] focuses on the universal applicability of diffusion models and demonstrates effectiveness in enhancing speech across a wide range of conditions. [178] proposes a method to reduce the prior mismatch in stochastic differential equations, enhancing model performance in speech tasks. CRA-DIFFUSE [179] introduces a pre-denoising step in the time-frequency domain to improve cross-domain speech enhancement, illustrating the potential for methodological innovations within diffusion model applications.

The application of diffusion models in audio processing has opened new avenues for research and development, offering novel solutions to longstanding challenges. As the field progresses, future research may focus on improving model efficiency, reducing computational demands, and exploring untapped applications within audio processing.

7.7 Others

Besides the applications discussed above, there are also some other applications of diffusion models. The following are some of them:

- **AIOps**: Cloud computing's reliability is paramount, yet it faces challenges like performance anomalies, unpredictable network traffic, and incomplete data. Diffusion models have emerged as a powerful tool to tackle these issues, offering new approaches for predictive maintenance, realistic traffic simulation, and enhanced failure forecasting. Maat [37] applies conditional diffusion models to anticipate performance metric anomalies in cloud services, demonstrating the potential for early anomaly detection, potentially reducing downtime, and improving service reliability. NetDiffus [39] applies diffusion models to time-series imaging for generating realistic network traffic patterns, showing how simulated traffic can support network planning, testing, and anomaly detection, enhancing network management and security. [45] investigates diffusion models for imputing missing data in time series, aiming to improve cloud failure forecasting accuracy, highlighting the effectiveness of diffusion models in handling data gaps, leading to better predictive outcomes for cloud service failures.
- **Finance**: The complexity and stochastic nature of financial markets make them an ideal candidate for the application of diffusion models, which can capture non-linearities and intricate patterns in data. Recent advancements have highlighted the potential of diffusion models in enhancing stock price forecasting, generating realistic financial tabular data, and augmenting stock factor data for improved investment strategies. [47] presents a novel approach to predict stock prices using a Diffusion Variational Autoencoder, aiming to better capture the stochastic nature of the market, highlighting the model's ability to handle market volatility. FinDiff [40] focuses on generating synthetic financial tabular data to address the scarcity of publicly available financial datasets. Synthetic data generated by FinDiff closely mimics real financial datasets, potentially aiding in model training and regulatory compliance without compromising sensitive information. DiffSTOCK [41]

employs diffusion models for probabilistic relational reasoning in stock market forecasting, aiming to capture the complex interdependencies between different market factors.

8 Outlook and Future Opportunities

In this section, we point out some future research directions of diffusion models for time series and spatio-temporal data that are worthy of further investigation.

8.1 Scalability and Efficiency

The computational complexity of diffusion models presents challenges for their application in resource-constrained or real-time environments. Therefore, enhancing their scalability and efficiency is pivotal for their deployment in real-world scenarios. Future directions can explore lighter and faster versions of diffusion models that significantly reduce computational demands while maintaining performance. Besides, further efforts include model compression, parallel computing, and efficient sampling strategies optimized for diffusion models adopted in time series and spatio-temporal data.

8.2 Robustness and Generalization

Data challenges like noise, missing data, anomalies, and distribution shifts often exist in real-world time series and spatio-temporal data. It is crucial to investigate and enhance the robustness of diffusion models against these data challenges. Therefore, enhancing the model's generalization capabilities across different datasets and scenarios is also an interesting direction to expand applicability and reliability in various domains. Furthermore, research could also focus on developing frameworks that adapt to new data characteristics or changing environments dynamically without human intervention.

8.3 Prior Knowledge Guided Generation

The generation process of time series and spatio-temporal data should adhere to unique constraints; for instance, generated trajectories should propagate over road networks, population migration data should conform to societal evolutionary patterns, and the spread of fires should adhere to thermodynamic principles. Most of the existing diffusion models, while capable of generating corresponding time series or spatio-temporal data based on some useful conditions, still lack adequate consideration for such prior knowledge in practice.

8.4 Multimodal Data Fusion

In complex real-world scenarios, time series and spatio-temporal data are often accompanied by other data types, such as textual and visual information. Exploring the fusion of multimodal data sources within diffusion models could significantly boost performance. This is particularly useful in domains like finance and healthcare, where integrating diverse data sources can lead to more comprehensive and accurate analyses. Future research could develop novel architectures that more effectively merge these diverse data streams, enhancing predictive performance and contextual understanding for multimodal time series and spatio-temporal data.

8.5 Integration of LLMs and Diffusion Models

The integration of LLMs and diffusion models for time series and spatio-temporal data analysis offers promising potential to advance the understanding of complex systems and improve decision-making. Specifically, leveraging the natural language understanding capabilities of LLMs can enhance temporal reasoning and offer a more holistic view of complex systems. Future research could include developing combined models that utilize the generative capabilities of diffusion models along with the rich semantic and syntactic processing of LLMs, potentially opening new avenues for automated reasoning and decision systems.

9 Conclusion

In this survey, we presented a comprehensive overview of the advancements and applications of diffusion models in the context of time series and spatio-temporal data analysis. We categorized diffusion models into unconditioned and conditioned types, each offering distinct advantages and challenges. Furthermore, we examined the various tasks associated with these models, including forecasting, generation, imputation, and anomaly detection. Additionally, we explored different application scenarios and provided insights into future opportunities and directions in this research field. It is our hope that this survey will contribute to the advancement of research in the area of diffusion models for time series and spatio-temporal data analysis.

REFERENCES

[1] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International conference on machine learning*. PMLR, 2015, pp. 2256–2265.

[2] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[3] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8162–8171.

[4] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in neural information processing systems*, vol. 32, 2019.

[5] R. Durall, A. Ghanim, M. R. Fernandez, N. Ettrich, and J. Keuper, "Deep diffusion models for seismic processing," *Computers & Geosciences*, vol. 177, p. 105377, 2023.

[6] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.

[7] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[8] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[9] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.

[10] C. Jarzynski, "Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach," *Physical Review E*, vol. 56, no. 5, p. 5018, 1997.

[11] E. Hoogeboom, D. Nielsen, P. Jaini, P. Forré, and M. Welling, "Argmax flows and multinomial diffusion: Learning categorical distributions," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12 454–12 465, 2021.

[12] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.

[13] K. Rasul, C. Seward, I. Schuster, and R. Vollgraf, "Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8857–8868.

[14] E. H. Capel and J. Dumas, "Denoising diffusion probabilistic models for probabilistic energy forecasting," in *2023 IEEE Belgrade PowerTech*. IEEE, 2023, pp. 1–6.

[15] S. Bassetti, B. Hutchinson, C. Tebaldi, and B. Kravitz, "Diffesm: Conditional emulation of earth system models with diffusion models," *arXiv preprint arXiv:2304.11699*, 2023.

[16] L. Chen, F. Du, Y. Hu, Z. Wang, and F. Wang, "Swinrdm: integrate swinrnn with diffusion model towards high-resolution and high-quality weather forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 1, 2023, pp. 322–330.

[17] T. Wei, Y. Lin, S. Guo, Y. Lin, Y. Huang, C. Xiang, Y. Bai, M. Ya, and H. Wan, "Diff-rntraj: A structure-aware diffusion model for road network-constrained trajectory generation," *arXiv preprint arXiv:2402.07369*, 2024.

[18] K. Shu, Y. Zhao, L. Wu, A. Liu, R. Qian, and X. Chen, "Data augmentation for seizure prediction with generative diffusion model," *arXiv preprint arXiv:2306.08256*, 2023.

[19] I. Nicholas, H. Kuo, F. Garcia, A. Sonnerborg, M. Bohm, R. Kaiser, M. Zazzi, L. Jorm, and S. Barbieri, "Synthetic health-related longitudinal data with mixed-type variables generated using diffusion models," in *NeurIPS 2023 Workshop on Synthetic Data Generation with Generative AI*, 2023.

[20] N. Chen, Y. Zhang, H. Zen, R. J. Weiss, M. Norouzi, and W. Chan, "Wavegrad: Estimating gradients for waveform generation," *arXiv preprint arXiv:2009.00713*, 2020.

[21] K. Lakhotia, E. Kharitonov, W.-N. Hsu, Y. Adi, A. Polyak, B. Bolte, T.-A. Nguyen, J. Copet, A. Baevski, A. Mohamed *et al.*, "On generative spoken language modeling from raw audio," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 1336–1354, 2021.

[22] A. H. Liu, W.-N. Hsu, M. Auli, and A. Baevski, "Towards end-to-end unsupervised speech recognition," in *2022 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2023, pp. 221–228.

[23] Q. Liu, F. Yan, X. Zhao, Z. Du, H. Guo, R. Tang, and F. Tian, "Diffusion augmentation for sequential recommendation," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 1576–1586.

[24] M. F. Sikder, R. Ramachandranpillai, and F. Heintz, "Transfusion: generating long, high fidelity time series using diffusion models with transformers," *arXiv preprint arXiv:2307.12667*, 2023.

[25] Y. Wang, Z. Wu, C. Li, and A. Wu, "Extraction and recovery of spatio-temporal structure in latent dynamics alignment with diffusion model," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[26] L. Huang, L. Gianinazzi, Y. Yu, P. D. Dueben, and T. Hoefler, "Diffda: a diffusion model for weather-scale data assimilation," *arXiv preprint arXiv:2401.05932*, 2024.

[27] Y. Hatanaka, Y. Glaser, G. Galgon, G. Torri, and P. Sadowski, "Diffusion models for high-resolution solar forecasts," *arXiv preprint arXiv:2302.00170*, 2023.

[28] Z. Wang, Q. Wen, C. Zhang, L. Sun, and Y. Wang, "Diffload: uncertainty quantification in load forecasting with diffusion model," *arXiv preprint arXiv:2306.01001*, 2023.

[29] S. R. Cachay, B. Zhao, H. James, and R. Yu, "Dyffusion: A dynamics-informed diffusion model for spatiotemporal forecasting," *arXiv preprint arXiv:2306.01984*, 2023.

[30] H. Yuan, S. Zhou, and S. Yu, "Ehrdiff: Exploring realistic ehr synthesis with diffusion models," *arXiv preprint arXiv:2303.05656*, 2023.

[31] T. Ceritli, G. O. Ghosheh, V. K. Chauhan, T. Zhu, A. P. Creagh, and D. A. Clifton, "Synthesizing mixed-type electronic health records using diffusion models," *arXiv preprint arXiv:2302.14679*, 2023.

[32] J. Richter, S. Welker, J.-M. Lemercier, B. Lay, and T. Gerkmann, "Speech enhancement and dereverberation with diffusion-based generative models," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.

[33] S. Zheng and N. Charoenphakdee, "Diffusion models for missing value imputation in tabular data," *arXiv preprint arXiv:2210.17128*, 2022.

[34] L. Shen and J. Kwok, "Non-autoregressive conditional diffusion models for time series prediction," *arXiv preprint arXiv:2306.05043*, 2023.

[35] Y. Zheng, L. Zhong, S. Wang, Y. Yang, W. Gu, J. Zhang, and J. Wang, "Diffuflow: Robust fine-grained urban flow inference with denoising diffusion model," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 3505–3513.

[36] Y. Wang, Z. Liu, L. Yang, and P. S. Yu, "Conditional denoising diffusion for sequential recommendation," *arXiv preprint arXiv:2304.11433*, 2023.

[37] C. Lee, T. Yang, Z. Chen, Y. Su, and M. R. Lyu, "Maat: Performance metric anomaly anticipation for cloud services with conditional diffusion," in *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 2023, pp. 116–128.

[38] Y. Zhu, C. Wang, and H. Xiong, "Towards graph-aware diffusion modeling for collaborative filtering," *arXiv preprint arXiv:2311.08744*, 2023.

[39] N. Sivaroopan, D. Bandara, C. Madarasingha, G. Jourjon, A. Jayasumana, and K. Thilakarathna, "Netdiffus: Network traffic generation by diffusion models through time-series imaging," *arXiv preprint arXiv:2310.04429*, 2023.

[40] T. Sattarov, M. Schreyer, and D. Borth, "Findiff: Diffusion models for financial tabular data generation," in *Proceedings of the Fourth ACM International Conference on AI in Finance*, 2023, pp. 64–72.

[41] D. Daiya, M. Yadav, and H. S. Rao, "Diffstock: Probabilistic relational stock market predictions using diffusion models," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 7335–7339.

[42] G. Klein, P. Guetschel, G. Silvestri, and M. Tangermann, "Synthesizing eeg signals from event-related potential paradigms with conditional diffusion models," *arXiv preprint arXiv:2403.18486*, 2024.

[43] H. Wen, Y. Lin, Y. Xia, H. Wan, Q. Wen, R. Zimmermann, and Y. Liang, "Diffstg: Probabilistic spatio-temporal graph forecasting with denoising diffusion models," in *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, 2023, pp. 1–12.

[44] G. Chi, Z. Yang, C. Wu, J. Xu, Y. Gao, Y. Liu, and T. X. Han, "Rf-diffusion: Radio signal generation via time-frequency diffusion," *arXiv preprint arXiv:2404.09140*, 2024.

[45] F. Yang, W. Yin, L. Wang, T. Li, P. Zhao, B. Liu, P. Wang, B. Qiao, Y. Liu, M. Björkman *et al.*, "Diffusion-based time series data imputation for cloud failure prediction at microsoft 365," in *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2023, pp. 2050–2055.

[46] J. Sui, J. Yu, Y. Song, and J. Zhang, "Anomaly detection for telemetry time series using a denoising diffusion probabilistic model," *IEEE Sensors Journal*, 2024.

[47] K. J. Koa, Y. Ma, R. Ng, and T.-S. Chua, "Diffusion variational autoencoder for tackling stochasticity in multi-step regression stock price prediction," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 1087–1096.

[48] H. He, S. Zhao, Y. Xi, and J. C. Ho, "Meddiff: Generating electronic health records using accelerated denoising diffusion model," *arXiv preprint arXiv:2302.04355*, 2023.

[49] P. Chang, H. Li, S. F. Quan, S. Lu, S.-F. Wung, J. Roveda, and A. Li, "A transformer-based diffusion probabilistic model for heart rate and blood pressure forecasting in intensive care unit," *Computer Methods and Programs in Biomedicine*, p. 108060, 2024.

[50] Y. Li, I. Y. Liao, N. Zhong, F. Toshihiro, Y. Wang, and S. Wang, "Generative ai enables the detection of autism using eeg signals," in *Chinese Conference on Biometric Recognition*. Springer, 2023, pp. 375–384.

[51] X. Li, M. Sakevych, G. Atkinson, and V. Metsis, "Biodiffusion: A versatile diffusion model for biomedical signal synthesis," *arXiv preprint arXiv:2401.10282*, 2024.

[52] Y. Gao, H. Chen, X. Wang, Z. Wang, X. Wang, J. Gao, and B. Ding, "Diffsformer: A diffusion transformer on stock factor augmentation," *arXiv preprint arXiv:2402.06656*, 2024.

[53] J. Yan, P. Li, and Y. Huang, "A short-term wind power scenario generation method based on conditional diffusion model," in *2023 IEEE Sustainable Power and Energy Conference (iSPEC)*. IEEE, 2023, pp. 1–6.

[54] M. Seki, Y.-Z. Zhang, and S. Imoto, "Imputing time-series microbiome abundance profiles with diffusion model," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 914–919.

[55] D. Scassola, S. Saccani, G. Carbone, and L. Bortolussi, "Conditioning score-based generative models by neuro-symbolic constraints," *arXiv preprint arXiv:2308.16534*, 2023.

[56] Y. Tashiro, J. Song, Y. Song, and S. Ermon, "Csdi: Conditional score-based diffusion models for probabilistic time series imputation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 24 804–24 816, 2021.

[57] H. Li, G. Ditzler, J. Roveda, and A. Li, "Descod-ecg: Deep score-based diffusion model for ecg baseline wander and noise removal," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[58] W. Zhao, Z. Chen, P. Xie, J. Liu, S. Hou, L. Xu, Y. Qiu, D. Wu, J. Xiao, and K. He, "Multi-task oriented diffusion model for mortality prediction in shock patients with incomplete data," *Information Fusion*, vol. 105, p. 102207, 2024.

[59] G. Liang, P. Tiwari, S. Nowaczyk, S. Byttner, and F. Alonso-Fernandez, "Dynamic causal explanation based diffusion-variational graph neural network for spatio-temporal forecasting," *arXiv preprint arXiv:2305.09703*, 2023.

[60] C. Chu, H. Zhang, P. Wang, and F. Lu, "Simulating human mobility with a trajectory generation framework based on diffusion model," *International Journal of Geographical Information Science*, pp. 1–32, 2024.

[61] M. Guimarães, C. Soares, and C. Manfletti, "Predicting the position uncertainty at the time of closest approach with diffusion models," *arXiv preprint arXiv:2311.05417*, 2023.

[62] J. Jiang, C. Han, W. X. Zhao, and J. Wang, "Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 4, 2023, pp. 4365–4373.

[63] Y. Y. Choi, M. Lee, S. W. Park, S. Lee, and J. Ko, "A gated mlp architecture for learning topological dependencies in spatio-temporal graphs," *arXiv preprint arXiv:2401.15894*, 2024.

[64] J. Wiederer, A. Bouazizi, M. Troina, U. Kressel, and V. Belagiannis, "Anomaly detection in multi-agent trajectories for automated driving," in *Conference on Robot Learning*. PMLR, 2022, pp. 1223–1233.

[65] Y. Lin, H. Wan, J. Hu, S. Guo, B. Yang, Y. Lin, and C. S. Jensen, "Origin-destination travel time oracle for map-based services," *Proceedings of the ACM on Management of Data*, vol. 1, no. 3, pp. 1–27, 2023.

[66] Y. Yao, Y. Liu, X. Dai, S. Chen, and Y. Lv, "A graph-based scene encoder for vehicle trajectory prediction using the diffusion model," in *2023 International Annual Conference on Complex Systems and Intelligent Science (CSIS-IAC)*, 2023, pp. 981–986.

[67] I. Kotseruba and J. K. Tsotsos, "Data limitations for modeling top-down effects on drivers' attention," *arXiv preprint arXiv:2404.08749*, 2024.

[68] K. Ito and L. Johnson, "The lj speech dataset," https://keithito.com/LJ-Speech-Dataset/, 2017.

[69] T. Yun, H. Jung, and J. Son, "Imputation as inpainting: Diffusion models for spatiotemporal data imputation," 2023.

[70] Y. Yuan, J. Ding, C. Shao, D. Jin, and Y. Li, "Spatio-temporal diffusion point processes," *arXiv preprint arXiv:2305.12403*, 2023.

[71] T. Yan, H. Zhang, T. Zhou, Y. Zhan, and Y. Xia, "Scoregrad: Multivariate probabilistic time series forecasting with continuous energy-based generative models," *arXiv preprint arXiv:2106.10121*, 2021.

[72] S. Zhang, S. Wang, X. Tan, R. Liu, J. Zhang, and J. Wang, "sasdim: self-adaptive noise scaling diffusion model for spatial time series imputation," *arXiv preprint arXiv:2309.01988*, 2023.

[73] Y. Cheng, K. Yamashita, J. Follum, and N. Yu, "Adversarial purification for data-driven power system event classifiers with diffusion models," *arXiv preprint arXiv:2311.07110*, 2023.

[74] N. Neifar, A. Ben-Hamadou, A. Mdhaffar, and M. Jmaiel, "Diffecg: A generalized probabilistic diffusion model for ecg signals synthesis," *arXiv preprint arXiv:2306.01875*, 2023.

[75] F. Bao, C. Li, J. Sun, and J. Zhu, "Why are conditional generative models better than unconditional ones?" *arXiv preprint arXiv:2212.00362*, 2022.

[76] M. Jin, Q. Wen, Y. Liang, C. Zhang, S. Xue, X. Wang, J. Zhang, Y. Wang, H. Chen, X. Li *et al.*, "Large models for time series and spatio-temporal data: A survey and outlook," *arXiv preprint arXiv:2310.10196*, 2023.

[77] M. Jin, H. Y. Koh, Q. Wen, D. Zambon, C. Alippi, G. I. Webb, I. King, and S. Pan, "A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection," *arXiv preprint arXiv:2307.03759*, 2023.

[78] K. Chen, G. Li, H. Li, Y. Wang, W. Wang, Q. Liu, and H. Wang, "Quantifying uncertainty: Air quality forecasting based on dynamic spatial-temporal denoising diffusion probabilistic model," *Environmental Research*, p. 118438, 2024.

[79] Y. Tang, H. He, Y. Wang, and Y. Wu, "Utilizing a diffusion model for pedestrian trajectory prediction in semi-open autonomous driving environments," *IEEE Sensors Journal*, pp. 1–1, 2024.

[80] T. Westny, B. Olofsson, and E. Frisk, "Diffusion-based environment-aware trajectory prediction," *arXiv preprint arXiv:2403.11643*, 2024.

[81] Y. Zhu, Y. Ye, S. Zhang, X. Zhao, and J. Yu, "Difftraj: Generating gps trajectory with diffusion probabilistic model," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[82] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 922–929.

[83] C. Li, G. Feng, Y. Li, R. Liu, Q. Miao, and L. Chang, "Difftad: Denoising diffusion probabilistic models for vehicle trajectory anomaly detection," *Knowledge-Based Systems*, vol. 286, p. 111387, 2024.

[84] L. Li, R. Carver, I. Lopez-Gomez, F. Sha, and J. Anderson, "Seeds: Emulation of weather forecast ensembles with diffusion models," *arXiv preprint arXiv:2306.14066*, 2023.

[85] X. Ling, C. Li, F. Qin, P. Yang, and Y. Huang, "Srndiff: Short-term rainfall nowcasting with condition diffusion model," *arXiv preprint arXiv:2402.13737*, 2024.

[86] Y. Li, X. Lu, Y. Wang, and D. Dou, "Generative time series forecasting with diffusion, denoise, and disentanglement," *Advances in Neural Information Processing Systems*, vol. 35, pp. 23 009–23 022, 2022.

[87] OpenAI Sora, "Video generation models as world simulators," https://openai.com/research/video-generation-models-as-world-simulators, accessed: 2024-04-29.

[88] L. Lin, D. Shi, A. Han, and J. Gao, "Specstg: A fast spectral diffusion framework for probabilistic spatio-temporal traffic forecasting," *arXiv preprint arXiv:2401.08119*, 2024.

[89] S. Kim, Y.-E. Lee, S.-H. Lee, and S.-W. Lee, "Diff-e: Diffusion-based learning for decoding imagined speech eeg," *arXiv preprint arXiv:2307.14389*, 2023.

[90] H. Wang, "Diffuseroll: Multi-track multi-category music generation based on diffusion model," *arXiv preprint arXiv:2303.07794*, 2023.

[91] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro, "Diffwave: A versatile diffusion model for audio synthesis," *arXiv preprint arXiv:2009.09761*, 2020.

[92] J. Zhang, G. Fazekas, and C. Saitis, "Fast diffusion gan model for symbolic music generation controlled by emotions," *arXiv preprint arXiv:2310.14040*, 2023.

[93] J. Leinonen, U. Hamann, D. Nerini, U. Germann, and G. Franch, "Latent diffusion models for generative precipitation nowcasting with accurate uncertainty quantification," *arXiv preprint arXiv:2304.12891*, 2023.

[94] R. Kurinchi-Vendhan, "Wirediff: a wind resolution-enhancing diffusion model."

[95] Z. Li, A. Sun, and C. Li, "Diffurec: A diffusion model for sequential recommendation," *arXiv preprint arXiv:2304.00686*, 2023.

[96] S. Li, H. Xiong, and Y. Chen, "Diffcharge: Generating ev charging scenarios via a denoising diffusion model," *IEEE Transactions on Smart Grid*, 2024.

[97] C. Li, G. Feng, Y. Li, R. Liu, Q. Miao, and L. Chang, "Difftad: Denoising diffusion probabilistic models for vehicle trajectory anomaly detection," *Knowledge-Based Systems*, vol. 286, p. 111387, 2024.

[98] X. Han, H. Zheng, and M. Zhou, "Card: Classification and regression diffusion models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 18 100–18 115, 2022.

[99] C. Yang, T. Wang, and X. Yan, "Ddmt: Denoising diffusion mask transformer models for multivariate time series anomaly detection," *arXiv preprint arXiv:2310.08800*, 2023.

[100] Y. Duan, J. Zhou, Z. Wang, Y.-C. Chang, Y.-K. Wang, and C.-T. Lin, "Domain-specific denoising diffusion probabilistic models for brain dynamics," *arXiv preprint arXiv:2305.04200*, 2023.

[101] D. Solis-Martin, J. Galan-Paez, and J. Borrego-Diaz, "D3a-ts: Denoising-driven data augmentation in time series," *arXiv preprint arXiv:2312.05550*, 2023.

[102] Y. Dai, C. Yang, K. Liu, A. Liu, and Y. Liu, "Timeddpm: Time series augmentation strategy for industrial soft sensing," *IEEE Sensors Journal*, 2023.

[103] H. Yi, L. Hou, Y. Jin, and N. A. Saeed, "Time series diffusion method: A denoising diffusion probabilistic model for vibration signal generation," *arXiv preprint arXiv:2312.07981*, 2023.

[104] M. Biloš, K. Rasul, A. Schneider, Y. Nevmyvaka, and S. Günnemann, "Modeling temporal data as continuous functions with process diffusion," *arXiv preprint arXiv:2211.02590*, 2022.

[105] J.-M. Lemercier, J. Richter, S. Welker, and T. Gerkmann, "Storm: A diffusion-based stochastic regeneration model for speech enhancement and dereverberation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.

[106] J. Crabbé, N. Huynh, J. Stanczuk, and M. van der Schaar, "Time series diffusion in the frequency domain," *arXiv preprint arXiv:2402.05933*, 2024.

[107] X. Dong, Z. Mao, Y. Sun, and X. Xu, "Short-term wind power scenario generation based on conditional latent diffusion models," *IEEE Transactions on Sustainable Energy*, 2023.

[108] Z. Zhang, H. Yang, J. Chen, and Z. Yin, "Multi-scale conditional diffusion model for deposited droplet volume measurement in inkjet printing manufacturing," *Journal of Manufacturing Systems*, vol. 71, pp. 595–608, 2023.

[109] S. Kim, S.-H. Lee, Y.-E. Lee, J.-W. Lee, J.-H. Park, and S.-W. Lee, "Brain-driven representation learning based on diffusion model," *arXiv preprint arXiv:2311.07925*, 2023.

[110] G. Sharma, A. Dhall, and R. Subramanian, "Medic: Mitigating eeg data scarcity via class-conditioned diffusion model," in *Deep Generative Models for Health Workshop NeurIPS 2023*, 2023.

[111] M. Liu, H. Huang, H. Feng, L. Sun, B. Du, and Y. Fu, "Pristi: A conditional diffusion framework for spatiotemporal imputation," *arXiv preprint arXiv:2302.09746*, 2023.

[112] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. Gontijo Lopes, B. Karagol Ayan, T. Salimans *et al.*, "Photorealistic text-to-image diffusion models with deep language understanding," *Advances in Neural Information Processing Systems*, vol. 35, pp. 36 479–36 494, 2022.

[113] Y. Ouyang, L. Xie, C. Li, and G. Cheng, "Missdiff: Training diffusion models on tabular data with missing values," in *ICML 2023 Workshop on Structured Probabilistic Inference {\&} Generative Modeling*, 2023.

[114] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[115] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang, "Diffusion models: A comprehensive survey of methods and applications," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–39, 2023.

[116] H. Lim, M. Kim, S. Park, and N. Park, "Regular time-series generation using sgm," *arXiv preprint arXiv:2301.08518*, 2023.

[117] A. Coletta, S. Gopalakrishnan, D. Borrajo, and S. Vyetrenko, "On the constrained time-series generation problem," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[118] W. Liu, J.-L. Qiu, W.-L. Zheng, and B.-L. Lu, "Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 715–729, 2021.

[119] G. Tosato, C. M. Dalbagno, and F. Fumagalli, "Eeg synthetic data generation using probabilistic diffusion models," *arXiv preprint arXiv:2303.06068*, 2023.

[120] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg," in *Computers in cardiology 1997*. IEEE, 1997, pp. 673–676.

[121] G. B. Moody and R. G. Mark, "The mit-bih arrhythmia database on cd-rom and software for use with it," in *[1990] Proceedings Computers in Cardiology*. IEEE, 1990, pp. 185–188.

[122] R. Bousseljot and D. Kreiseler, "Waveform recognition with 10,000 ecgs," in *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*. IEEE, 2000, pp. 331–334.

[123] J. M. L. Alcaraz and N. Strodthoff, "Diffusion-based conditional ecg generation with structured state space models," *Computers in Biology and Medicine*, p. 107115, 2023.

[124] J. F. Núñez, J. Arjona, A. Tormos, D. García, and J. Béjar, "Applying generative models and transfer learning to physiological data classification," in *Artificial Intelligence Research and Development*. IOS Press, 2023, pp. 28–37.

[125] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE engineering in medicine and biology magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[126] A. Miltiadous, K. D. Tzimourta, T. Afrantou, P. Ioannidis, N. Grigoriadis, D. G. Tsalikakis, P. Angelidis, M. G. Tsipouras, E. Glavas, N. Giannakeas *et al.*, "A dataset of scalp eeg recordings of alzheimer's disease, frontotemporal dementia and healthy subjects from routine eeg," *Data*, vol. 8, no. 6, p. 95, 2023.

[127] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185–1194, 2000.

[128] B. Aristimunha, R. Y. de Camargo, S. Chevallier, O. Lucena, A. G. Thomas, M. J. Cardoso, W. H. L. Pinaya, and J. Dafflon, "Synthetic sleep eeg signal generation using latent diffusion models," in *Deep Generative Models for Health Workshop NeurIPS 2023*, 2023.

[129] G.-Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mobley, and S. Redline, "The national sleep research resource: towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1351–1358, 2018.

[130] G. Bénédict, O. Jeunen, S. Papa, S. Bhargav, D. Odijk, and M. de Rijke, "Recfusion: A binomial diffusion process for 1d data for recommendation," *arXiv preprint arXiv:2306.08947*, 2023.

[131] X. Lin, X. Chen, C. Wang, H. Shu, L. Song, B. Li *et al.*, "Discrete conditional diffusion for reranking in recommendation," *arXiv preprint arXiv:2308.06982*, 2023.

[132] W. Wang, Y. Xu, F. Feng, X. Lin, X. He, and T.-S. Chua, "Diffusion recommender model," *arXiv preprint arXiv:2304.04971*, 2023.

[133] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *Acm transactions on interactive intelligent systems (tiis)*, vol. 5, no. 4, pp. 1–19, 2015.

[134] H. Du, H. Yuan, Z. Huang, P. Zhao, and X. Zhou, "Sequential recommendation with diffusion models," *arXiv preprint arXiv:2304.04541*, 2023.

[135] H. Ma, R. Xie, L. Meng, X. Chen, X. Zhang, L. Lin, and Z. Kang, "Plug-in diffusion model for sequential recommendation," *arXiv preprint arXiv:2401.02913*, 2024.

[136] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers *et al.*, "The era5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[137] L. Chuyao, X. Li, and Y. Ye, "Pfst-lstm: A spatiotemporal lstm model with pseudoflow prediction for precipitation nowcasting," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. PP, no. 99, pp. 1–1, 2020.

[138] P. Sepulchre, A. Caubel, J.-B. Ladant, L. Bopp, O. Boucher, P. Braconnot, P. Brockmann, A. Cozic, Y. Donnadieu, J.-L. Dufresne *et al.*, "Ipsl-cm5a2–an earth system model designed for multi-millennial climate simulations," *Geoscientific Model Development*, vol. 13, no. 7, pp. 3011–3053, 2020.

[139] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," pp. 896–913, 2016.

[140] Z. J. Lee, T. Li, and S. H. Low, "Acn-data: Analysis and applications of an open ev charging dataset," in *Proceedings of the tenth ACM international conference on future energy systems*, 2019, pp. 139–149.

[141] J. R. Schofield, R. Carmichael, S. Tindemans, M. Bilton, M. Woolf, G. Strbac *et al.*, "Low carbon london project: Data from the dynamic time-of-use electricity pricing trial, 2013," *uK Data Service, SN*, vol. 7857, no. 2015, pp. 7857–7851, 2015.

[142] Z. Wang and H. Zhang, "Customized load profiles synthesis for electricity customers based on conditional diffusion models," *arXiv preprint arXiv:2304.12076*, 2023.

[143] S. Li, H. Xiong, and Y. Chen, "Diffplf: A conditional diffusion model for probabilistic forecasting of ev charging load," *arXiv preprint arXiv:2402.13548*, 2024.

[144] A. Jenkins, Z. Chen, F. S. Ng, and D. Mandic, "Improving diffusion models for ecg imputation with an augmented template prior," *arXiv preprint arXiv:2310.15742*, 2023.

[145] L. Liu, Y. Ren, Z. Lin, and Z. Zhao, "Pseudo numerical methods for diffusion models on manifolds," *arXiv preprint arXiv:2202.09778*, 2022.

[146] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, "Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps," *Advances in Neural Information Processing Systems*, vol. 35, pp. 5775–5787, 2022.

[147] X. Mao, "The truncated euler–maruyama method for stochastic differential equations," *Journal of Computational and Applied Mathematics*, vol. 290, pp. 370–384, 2015.

[148] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[149] M. Le, A. Vyas, B. Shi, B. Karrer, L. Sari, R. Moritz, M. Williamson, V. Manohar, Y. Adi, J. Mahadeokar *et al.*, "Voicebox: Text-guided multilingual universal speech generation at scale," *Advances in neural information processing systems*, vol. 36, 2024.

[150] D. Epstein, A. Jabri, B. Poole, A. Efros, and A. Holynski, "Diffusion self-guidance for controllable image generation," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[151] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[152] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[153] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," *arXiv preprint arXiv:2112.10741*, 2021.

[154] C. Meng, R. Rombach, R. Gao, D. Kingma, S. Ermon, J. Ho, and T. Salimans, "On distillation of guided diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 297–14 306.

[155] M. Hu, Y. Wang, T.-J. Cham, J. Yang, and P. N. Suganthan, "Global context with discrete diffusion in vector quantised modelling for image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 502–11 511.

[156] J. Wolleb, F. Bieder, R. Sandkühler, and P. C. Cattin, "Diffusion models for medical anomaly detection," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2022, pp. 35–45.

[157] S. Chen, P. Sun, Y. Song, and P. Luo, "Diffusiondet: Diffusion model for object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19 830–19 843.

[158] D. Baranchuk, I. Rubachev, A. Voynov, V. Khrulkov, and A. Babenko, "Label-efficient semantic segmentation with diffusion models," *arXiv preprint arXiv:2112.03126*, 2021.

[159] C.-H. Chao, W.-F. Sun, B.-W. Cheng, and C.-Y. Lee, "Quasi-conservative score-based generative models," *arXiv preprint arXiv:2209.12753*, 2022.

[160] H. Chung, B. Sim, and J. C. Ye, "Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 413–12 422.

[161] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, "Ilvr: Conditioning method for denoising diffusion probabilistic models," *arXiv preprint arXiv:2108.02938*, 2021.

[162] A. Bansal, H.-M. Chu, A. Schwarzschild, S. Sengupta, M. Goldblum, J. Geiping, and T. Goldstein, "Universal guidance for diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 843–852.

[163] Y. Yuan, J. Song, U. Iqbal, A. Vahdat, and J. Kautz, "Physdiff: Physics-guided human motion diffusion model," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 16 010–16 021.

[164] M. Kollovieh, A. F. Ansari, M. Bohlke-Schneider, J. Zschiegner, H. Wang, and Y. B. Wang, "Predict, refine, synthesize: Self-guiding diffusion models for probabilistic time series forecasting," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[165] S. An, H. Lee, J. Jo, S. Lee, and S. J. Hwang, "Diffusionnag: Task-guided neural architecture generation with diffusion models," *arXiv preprint arXiv:2305.16943*, 2023.

[166] N. Torenvliet and J. Zelek, "Evaluating diffusion models for the automation of ultrasonic nondestructive evaluation data analysis," *Algorithms*, vol. 17, no. 4, p. 167, 2024.

[167] Z. Yang, J. Wu, Z. Wang, X. Wang, Y. Yuan, and X. He, "Generate what you prefer: Reshaping sequential recommendation via guided diffusion," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[168] P. Shao, J. Feng, J. Lu, P. Zhang, and C. Zou, "Data-driven and knowledge-guided denoising diffusion model for flood forecasting," *Expert Systems with Applications*, vol. 244, p. 122908, 2024.

[169] S. Zuo, V. F. Rey, S. Suh, S. Sigg, and P. Lukowicz, "Unsupervised statistical feature-guided diffusion model for sensor-based human activity recognition," *arXiv preprint arXiv:2306.05285*, 2023.

[170] J. Hu, X. Liu, Z. Fan, Y. Liang, and R. Zimmermann, "Towards unifying diffusion models for probabilistic spatio-temporal graph learning," *arXiv preprint arXiv:2310.17360*, 2023.

[171] L. Yang, Z. Zhang, W. Zhang, and S. Hong, "Score-based graph generative modeling with self-guided latent diffusion," 2022.

[172] S. Gong, M. Li, J. Feng, Z. Wu, and L. Kong, "Diffuseq: Sequence to sequence text generation with diffusion models," *arXiv preprint arXiv:2210.08933*, 2022.

[173] A. Schneuing, Y. Du, C. Harris, A. Jamasb, I. Igashov, W. Du, T. Blundell, P. Lió, C. Gomes, M. Welling *et al.*, "Structure-based drug design with equivariant diffusion models," *arXiv preprint arXiv:2210.13695*, 2022.

[174] B. D. Anderson, "Reverse-time diffusion equation models," *Stochastic Processes and their Applications*, vol. 12, no. 3, pp. 313–326, 1982.

[175] Y.-J. Lu, Z.-Q. Wang, S. Watanabe, A. Richard, C. Yu, and Y. Tsao, "Conditional diffusion probabilistic model for speech enhancement," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 7402–7406.

[176] W. Tai, F. Zhou, G. Trajcevski, and T. Zhong, "Revisiting denoising diffusion probabilistic models for speech enhancement: Condition collapse, efficiency and refinement," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 13 627–13 635.

[177] W. Tai, Y. Lei, F. Zhou, G. Trajcevski, and T. Zhong, "Dose: Diffusion dropout with adaptive prior for speech enhancement," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[178] B. Lay, S. Welker, J. Richter, and T. Gerkmann, "Reducing the prior mismatch of stochastic differential equations for diffusion-based speech enhancement," *arXiv preprint arXiv:2302.14748*, 2023.

[179] Z. Qiu, Y. Guo, M. Fu, H. Huang, Y. Hu, L. He, and F. Sun, "Cradiffuse: Improved cross-domain speech enhancement based on diffusion model with tf domain pre-denoising," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2023, pp. 1709–1714.
[180] J. Serrà, S. Pascual, J. Pons, R. O. Araz, and D. Scaini, "Universal speech enhancement with score-based diffusion," *arXiv preprint arXiv:2206.03065*, 2022.
[181] H. Chen, X. Dong, Y. Kong, Z. Chen, S. Zheng, X. Hu, and X. Zhao, "Online prediction of mechanical and electrical quality in ultrasonic metal welding using time series generation and deep learning," *Engineering Failure Analysis*, p. 108162, 2024.
[182] S. S. Narasimhan, S. Agarwal, O. Akcin, S. Sanghavi, and S. Chinchali, "Time weaver: A conditional time series generation model," *arXiv preprint arXiv:2403.02682*, 2024.
[183] H. Shirzad, R. Deng, H. Zhao, and F. Tung, "Conditional diffusion models as self-supervised learning backbone for irregular time series," in *ICLR 2024 Workshop on Learning from Time Series For Health*, 2024.
[184] C. Liu, S. He, H. Liu, and J. Chen, "Intention-aware denoising diffusion model for trajectory prediction," *arXiv preprint arXiv:2403.09190*, 2024.
[185] Z. Evans, J. D. Parker, C. Carr, Z. Zukowski, J. Taylor, and J. Pons, "Long-form music generation with latent diffusion," *arXiv preprint arXiv:2404.10301*, 2024.
[186] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
[187] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
[188] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *International conference on machine learning*. PMLR, 2015, pp. 1530–1538.
[189] Z. Chang, G. A. Koulieris, and H. P. Shum, "On the design fundamentals of diffusion models: A survey," *arXiv preprint arXiv:2306.04542*, 2023.
[190] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, and D. J. Fleet, "Video diffusion models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 8633–8646, 2022.
[191] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi, "Palette: Image-to-image diffusion models," in *ACM SIGGRAPH 2022 conference proceedings*, 2022, pp. 1–10.
[192] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 4, pp. 4713–4726, 2022.
[193] J. Austin, D. D. Johnson, J. Ho, D. Tarlow, and R. Van Den Berg, "Structured denoising diffusion models in discrete state-spaces," *Advances in Neural Information Processing Systems*, vol. 34, pp. 17 981–17 993, 2021.
[194] J. Lovelace, V. Kishore, C. Wan, E. Shekhtman, and K. Q. Weinberger, "Latent diffusion for language generation," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
[195] X. Li, J. Thickstun, I. Gulrajani, P. S. Liang, and T. B. Hashimoto, "Diffusion-lm improves controllable text generation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 4328–4343, 2022.
[196] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125*, vol. 1, no. 2, p. 3, 2022.
[197] D. Yang, J. Yu, H. Wang, W. Wang, C. Weng, Y. Zou, and D. Yu, "Diffsound: Discrete diffusion model for text-to-sound generation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.
[198] M. Zhang, Z. Cai, L. Pan, F. Hong, X. Guo, L. Yang, and Z. Liu, "Motiondiffuse: Text-driven human motion generation with diffusion model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
[199] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3836–3847.
[200] T. Amit, T. Shaharbany, E. Nachmani, and L. Wolf, "Segdiff: Image segmentation with diffusion probabilistic models," *arXiv preprint arXiv:2112.00390*, 2021.
[201] J. Wolleb, R. Sandkühler, F. Bieder, P. Valmaggia, and P. C. Cattin, "Diffusion models for implicit image segmentation ensembles," in *International Conference on Medical Imaging with Deep Learning*. PMLR, 2022, pp. 1336–1348.
[202] G. Nam, M. Khlifi, A. Rodriguez, A. Tono, L. Zhou, and P. Guerrero, "3d-ldm: Neural implicit 3d shape generation with latent diffusion models," *arXiv preprint arXiv:2212.00842*, 2022.
[203] A. Vahdat, F. Williams, Z. Gojcic, O. Litany, S. Fidler, K. Kreis *et al.*, "Lion: Latent point diffusion models for 3d shape generation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 10 021–10 039, 2022.
[204] M. Xu, L. Yu, Y. Song, C. Shi, S. Ermon, and J. Tang, "Geodiff: A geometric diffusion model for molecular conformation generation," *arXiv preprint arXiv:2203.02923*, 2022.
[205] Z. Guo, J. Liu, Y. Wang, M. Chen, D. Wang, D. Xu, and J. Cheng, "Diffusion models in bioinformatics and computational biology," *Nature Reviews Bioengineering*, pp. 1–19, 2023.
[206] R. Huang, M. W. Lam, J. Wang, D. Su, D. Yu, Y. Ren, and Z. Zhao, "Fastdiff: A fast conditional diffusion model for high-quality speech synthesis," *arXiv preprint arXiv:2204.09934*, 2022.
[207] H. Harutyunyan, H. Khachatrian, D. C. Kale, G. Ver Steeg, and A. Galstyan, "Multitask learning and benchmarking with clinical time series data," *Scientific data*, vol. 6, no. 1, p. 96, 2019.
[208] G. Jin, Y. Liang, Y. Fang, Z. Shao, J. Huang, J. Zhang, and Y. Zheng, "Spatio-temporal graph neural networks for predictive learning in urban computing: A survey," *IEEE Transactions on Knowledge and Data Engineering*, 2023.
[209] M. Jin, Y. Zhang, W. Chen, K. Zhang, Y. Liang, B. Yang, J. Wang, S. Pan, and Q. Wen, "Position paper: What can large language models tell us about time series analysis," *arXiv preprint arXiv:2402.02713*, 2024.
[210] S. Khanna, P. Liu, L. Zhou, C. Meng, R. Rombach, M. Burke, D. Lobell, and S. Ermon, "Diffusionsat: A generative foundation model for satellite imagery," *arXiv preprint arXiv:2312.03606*, 2023.
[211] Y. Xu, Z. Liu, M. Tegmark, and T. Jaakkola, "Poisson flow generative models," *ArXiv*, vol. abs/2209.11178, 2022.
[212] Y. Xu, Z. Liu, Y. Tian, S. Tong, M. Tegmark, and T. Jaakkola, "Pfgm++: Unlocking the potential of physics-inspired generative models," *ArXiv*, vol. abs/2302.04265, 2023.
[213] Z. Liu, D. Luo, Y. Xu, T. Jaakkola, and M. Tegmark, "Genphys: From physical processes to generative models," *ArXiv*, vol. abs/2304.02637, 2023.
[214] A. Bansal, E. Borgnia, H.-M. Chu, J. S. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein, "Cold diffusion: Inverting arbitrary image transforms without noise," *arXiv:2208.09392*, 2022.
[215] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the design space of diffusion-based generative models," *arXiv:2206.00364*, 2022.
[216] Q. Zhang, M. Tao, and Y. Chen, "gddim: Generalized denoising diffusion implicit models," *arXiv preprint arXiv:2206.05564*, 2022.
[217] Q. Zhang and Y. Chen, "Fast sampling of diffusion models with exponential integrator," *arXiv preprint arXiv:2204.13902*, 2022.
[218] A. Jolicoeur-Martineau, K. Li, R. Piché-Taillefer, T. Kachman, and I. Mitliagkas, "Gotta go fast when generating data with score-based models," *arXiv preprint arXiv:2105.14080*, 2021.
[219] Y. Xu, M. Deng, X. Cheng, Y. Tian, Z. Liu, and T. Jaakkola, "Restart sampling for improving generative processes," *Advances in Neural Information Processing Systems*, vol. 36, pp. 76 806–76 838, 2023.
[220] D. Berthelot, A. Autef, J. Lin, D. A. Yap, S. Zhai, S. Hu, D. Zheng, W. Talbott, and E. Gu, "Tract: Denoising diffusion models with transitive closure time-distillation," *arXiv preprint arXiv:2303.04248*, 2023.
[221] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, "Consistency models," *arXiv preprint arXiv:2303.01469*, 2023.
[222] H. Zheng, P. He, W. Chen, and M. Zhou, "Truncated diffusion probabilistic models," *arXiv preprint arXiv:2202.09671*, vol. 1, no. 3.1, p. 2, 2022.
[223] D. Watson, J. Ho, M. Norouzi, and W. Chan, "Learning to efficiently sample from diffusion probabilistic models," *arXiv preprint arXiv:2106.03802*, 2021.
[224] Z. Lyu, X. Xu, C. Yang, D. Lin, and B. Dai, "Accelerating diffusion models via early stop of the diffusion process," *arXiv preprint arXiv:2205.12524*, 2022.
[225] K. Pandey, A. Mukherjee, P. Rai, and A. Kumar, "Diffusevae: Efficient, controllable and high-fidelity generation from low-dimensional latents," *arXiv preprint arXiv:2201.00308*, 2022.

[226] X. Liu, C. Gong, and Q. Liu, “Flow straight and fast: Learning to generate and transfer data with rectified flow,” *arXiv preprint arXiv:2209.03003*, 2022.
[227] A. Vahdat, K. Kreis, and J. Kautz, “Score-based generative modeling in latent space,” *Advances in neural information processing systems*, vol. 34, pp. 11 287–11 302, 2021.
[228] H. Zhang, R. Feng, Z. Yang, L. Huang, Y. Liu, Y. Zhang, Y. Shen, D. Zhao, J. Zhou, and F. Cheng, “Dimensionality-varying diffusion process,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 307–14 316.
[229] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, “Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models,” *arXiv preprint arXiv:2211.01095*, 2022.
[230] S. Luo, Y. Tan, L. Huang, J. Li, and H. Zhao, “Latent consistency models: Synthesizing high-resolution images with few-step inference,” *arXiv preprint arXiv:2310.04378*, 2023.
[231] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, “Repaint: Inpainting using denoising diffusion probabilistic models,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 461–11 471.
[232] J. Zheng, M. Hu, Z. Fan, C. Wang, C. Ding, D. Tao, and T.-J. Cham, “Trajectory consistency distillation,” *arXiv preprint arXiv:2402.19159*, 2024.
[233] W. Zhao, L. Bai, Y. Rao, J. Zhou, and J. Lu, “Unipc: A unified predictor-corrector framework for fast sampling of diffusion models,” *Advances in Neural Information Processing Systems*, vol. 36, 2024.
[234] S. Gu, D. Chen, J. Bao, F. Wen, B. Zhang, D. Chen, L. Yuan, and B. Guo, “Vector quantized diffusion model for text-to-image synthesis,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 696–10 706.
[235] D. Kingma, T. Salimans, B. Poole, and J. Ho, “Variational diffusion models,” *Advances in neural information processing systems*, vol. 34, pp. 21 696–21 707, 2021.
[236] B. Kawar, M. Elad, S. Ermon, and J. Song, “Denoising diffusion restoration models,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 23 593–23 606, 2022.
[237] S. Rissanen, M. Heinonen, and A. Solin, “Generative modelling with inverse heat dissipation,” *arXiv preprint arXiv:2206.13397*, 2022.
[238] R. Yang, P. Srivastava, and S. Mandt, “Diffusion probabilistic modeling for video generation,” *Entropy*, vol. 25, no. 10, p. 1469, 2023.
[239] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
[240] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong *et al.*, “Swin transformer v2: Scaling up capacity and resolution,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12 009–12 019.
[241] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.
[242] W. Peebles and S. Xie, “Scalable diffusion models with transformers,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4195–4205.
[243] N. Ma, M. Goldstein, M. S. Albergo, N. M. Boffi, E. Vanden-Eijnden, and S. Xie, “Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers,” *arXiv preprint arXiv:2401.08740*, 2024.
[244] E. Hoogeboom and T. Salimans, “Blurring diffusion models,” *arXiv preprint arXiv:2209.05557*, 2022.
[245] G. Daras, M. Delbracio, H. Talebi, A. G. Dimakis, and P. Milanfar, “Soft diffusion: Score matching for general corruptions,” *arXiv preprint arXiv:2209.05442*, 2022.
[246] Z. Kong and W. Ping, “On fast sampling of diffusion probabilistic models,” *arXiv preprint arXiv:2106.00132*, 2021.
[247] Y. Song, C. Durkan, I. Murray, and S. Ermon, “Maximum likelihood training of score-based diffusion models,” *Advances in neural information processing systems*, vol. 34, pp. 1415–1428, 2021.
[248] C.-W. Huang, J. H. Lim, and A. C. Courville, “A variational perspective on diffusion-based generative models and score matching,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 22 863–22 876, 2021.
[249] E. Heitz, L. Belcour, and T. Chambon, “Iterative α-(de) blending: A minimalist deterministic diffusion model,” in *ACM SIGGRAPH 2023 Conference Proceedings*, 2023, pp. 1–8.
[250] M. S. Albergo and E. Vanden-Eijnden, “Building normalizing flows with stochastic interpolants,” *arXiv preprint arXiv:2209.15571*, 2022.
[251] X. Su, J. Song, C. Meng, and S. Ermon, “Dual diffusion implicit bridges for image-to-image translation,” *arXiv preprint arXiv:2203.08382*, 2022.
[252] G.-H. Liu, A. Vahdat, D.-A. Huang, E. A. Theodorou, W. Nie, and A. Anandkumar, “I^{2}sb: Image-to-image schr\"odinger bridge,” *arXiv preprint arXiv:2302.05872*, 2023.
[253] S. Feng, C. Miao, Z. Zhang, and P. Zhao, “Latent diffusion transformer for probabilistic time series forecasting,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 11, 2024, pp. 11 979–11 987.
[254] Y. Liang, H. Wen, Y. Nie, Y. Jiang, M. Jin, D. Song, S. Pan, and Q. Wen, “Foundation models for time series analysis: A tutorial and survey,” *arXiv preprint arXiv:2403.14735*, 2024.
[255] Y. Wang, H. Wu, J. Dong, Y. Liu, Y. Qiu, H. Zhang, J. Wang, and M. Long, “Timexer: Empowering transformers for time series forecasting with exogenous variables,” *arXiv preprint arXiv:2402.19072*, 2024.
[256] S. Wu, H. Fei, L. Qu, W. Ji, and T.-S. Chua, “Next-gpt: Any-to-any multimodal llm,” *arXiv preprint arXiv:2309.05519*, 2023.
[257] Y. Chen, C. Zhang, M. Ma, Y. Liu, R. Ding, B. Li, S. He, S. Rajmohan, Q. Lin, and D. Zhang, “Imdiffusion: Imputed diffusion models for multivariate time series anomaly detection,” *arXiv preprint arXiv:2307.00754*, 2023.
[258] V. Livernoche, V. Jain, Y. Hezaveh, and S. Ravanbakhsh, “On diffusion modeling for anomaly detection,” *arXiv preprint arXiv:2305.18593*, 2023.
[259] I. Pintilie, A. Manolache, and F. Brad, “Time series anomaly detection using diffusion-based models,” in *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2023, pp. 570–578.
[260] C. Xiao, Z. Gou, W. Tai, K. Zhang, and F. Zhou, “Imputation-based time-series anomaly detection with conditional weight-incremental diffusion models,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 2742–2751.
[261] C. Wang, Z. Zhuang, Q. Qi, J. Wang, X. Wang, H. Sun, and J. Liao, “Drift doesn’t matter: Dynamic decomposition with diffusion reconstruction for unstable multivariate time series anomaly detection,” *Advances in Neural Information Processing Systems*, vol. 36, 2024.
[262] X. Wang, H. Zhang, P. Wang, Y. Zhang, B. Wang, Z. Zhou, and Y. Wang, “An observed value consistent diffusion model for imputing missing values in multivariate time series,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 2409–2418.
[263] X. Liu, J. Chen, J. Xie, and Y. Chang, “Generating hsr bogie vibration signals via pulse voltage-guided conditional diffusion model,” *arXiv preprint arXiv:2311.00496*, 2023.
[264] L. Wang, “Multi-modality conditional diffusion model for time series forecasting of live sales volume,” in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 2675–2679.
[265] K. Lv, L. Yuan, and X. Ni, “Learning autoencoder diffusion models of pedestrian group relationships for multimodal trajectory prediction,” *IEEE Transactions on Instrumentation and Measurement*, 2024.
[266] C. Fu, H. Kazmi, M. Quintana, and C. Miller, “Creating synthetic energy meter data using conditional diffusion and building metadata,” *arXiv preprint arXiv:2404.00525*, 2024.
[267] T. Shimizu, “Diffusion model in causal inference with unmeasured confounders,” in *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2023, pp. 683–688.
[268] A. Asperti, F. Merizzi, A. Paparella, G. Pedrazzi, M. Angelinelli, and S. Colamonaco, “Precipitation nowcasting with generative diffusion models,” *arXiv preprint arXiv:2308.06733*, 2023.
[269] J. Pei, J. Wang, D. Shi, and P. Wang, “Improved efficient two-stage denoising diffusion power system measurement recovery against false data injection attacks and data losses,” *arXiv preprint arXiv:2312.04346*, 2023.
[270] P. N. Mueller, “Attention-enhanced conditional-diffusion-based data synthesis for data augmentation in machine fault diagnosis,” *Engineering Applications of Artificial Intelligence*, vol. 131, p. 107696, 2024.

[271] P. Zhao, X. Wang, Y. Zhang, Y. Li, H. Wang, and Y. Yang, "Diffusion-uda: Diffusion-based unsupervised domain adaptation for submersible fault diagnosis," *Electronics Letters*, vol. 60, no. 3, p. e13122, 2024.

[272] T. Y. Wong, M. H. Lim, W. K. Ngui, and M. S. Leong, "Denoising diffusion implicit model for bearing fault diagnosis under different working loads," in *ITM Web of Conferences*, vol. 63. EDP Sciences, 2024, p. 01025.

[273] K. Goel, A. Gu, C. Donahue, and C. Ré, "It's raw! audio generation with state-space models," *arXiv preprint arXiv:2202.09729*, 2022.

[274] C. Donahue, J. McAuley, and M. Puckette, "Adversarial audio synthesis," in *International Conference on Learning Representations*, 2019.

[275] J. M. L. Alcaraz and N. Strodthoff, "Diffusion-based time series imputation and forecasting with structured state space models," *arXiv preprint arXiv:2208.09399*, 2022.

[276] J. Xu, F. Lyu, and P. C. Yuen, "Density-aware temporal attentive step-wise diffusion model for medical time series imputation," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 2836–2845.

[277] Z. Liu, W. Pei, D. Lan, and Q. Ma, "Diffusion language-shapelets for semi-supervised time-series classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 13, 2024, pp. 14 079–14 087.

[278] Y. Li, "Ts-diffusion: Generating highly complex time series with diffusion models," *arXiv preprint arXiv:2311.03303*, 2023.

[279] Y. Yao, Y. Liu, X. Dai, S. Chen, and Y. Lv, "A graph-based scene encoder for vehicle trajectory prediction using the diffusion model," in *2023 International Annual Conference on Complex Systems and Intelligent Science (CSIS-IAC)*. IEEE, 2023, pp. 981–986.

[280] Y. Li, M. R. Luyten, and M. van der Schaar, "Risk-sensitive diffusion: Learning the underlying distribution from noisy samples," *arXiv preprint arXiv:2402.02081*, 2024.

[281] Y. Liang, K. Ouyang, H. Yan, Y. Wang, Z. Tong, and R. Zimmermann, "Modeling trajectories with neural ordinary differential equations." in *IJCAI*, 2021, pp. 1498–1504.

[282] R. Hu, X. Yuan, Y. Qiao, B. Zhang, and P. Zhao, "Unsupervised anomaly detection for multivariate time series using diffusion model," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 9606–9610.

[283] X. Liu, Y. Xia, Y. Liang, J. Hu, Y. Wang, L. Bai, C. Huang, Z. Liu, B. Hooi, and R. Zimmermann, "Largest: A benchmark dataset for large-scale traffic forecasting," in *Advances in Neural Information Processing Systems*, 2023.

[284] S. Welker, J. Richter, and T. Gerkmann, "Speech enhancement with score-based generative models in the complex stft domain," *arXiv preprint arXiv:2203.17004*, 2022.

[285] S. Rühling Cachay, B. Zhao, H. Joren, and R. Yu, "Dyffusion: A dynamics-informed diffusion model for spatiotemporal forecasting," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[286] J. Sun, J. Zhang, Q. Li, X. Yi, Y. Liang, and Y. Zheng, "Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 5, pp. 2348–2359, 2020.

[287] Y. Liang, K. Ouyang, J. Sun, Y. Wang, J. Zhang, Y. Zheng, D. Rosenblum, and R. Zimmermann, "Fine-grained urban flow prediction," in *Proceedings of the Web Conference 2021*, 2021, pp. 1833–1845.

[288] W. Chen, Y. Liang, Y. Zhu, Y. Chang, K. Luo, H. Wen, L. Li, Y. Yu, Q. Wen, C. Chen *et al.*, "Deep learning for trajectory data management and mining: A survey and beyond," *arXiv preprint arXiv:2403.14151*, 2024.

[289] P. Chang, H. Li, S. F. Quan, S. Lu, S.-F. Wung, J. Roveda, and A. Li, "Tdstf: Transformer-based diffusion probabilistic model for sparse time series forecasting," *arXiv preprint arXiv:2301.06625*, 2023.

[290] Y. Zhu, J. J. Yu, X. Zhao, Q. Liu, Y. Ye, W. Chen, Z. Zhang, X. Wei, and Y. Liang, "Controltraj: Controllable trajectory generation with topology-constrained diffusion model," *arXiv preprint arXiv:2404.15380*, 2024.

[291] K. Ito, K. Itô, K. Itô, J. Mathématicien, K. Itô, and J. Mathematician, *On stochastic differential equations*. American Mathematical Society New York, 1951, vol. 4.

[292] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick *et al.*, "Accurate structure prediction of biomolecular interactions with alphafold 3," *Nature*, pp. 1–3, 2024.