Preprints.org

A Comprehensive Survey on Diffusion Models and Their Applications

Md Manjurul Ahsan [*] , Shivakumar Raman , Yingtao Liu , Zahed Siddique

Review

A Comprehensive Survey on Diffusion Models and Their Applications

Md Manjurul Ahsan [1,*], Shivakumar Raman [2], Yingtao Liu [3] and Zahed Siddique [3]

[1]   Industrial and Systems Engineering, University of Oklahoma, Norman, Oklahoma-73071; ahsan@ou.edu
[2]   Department of Industrial and Systems Engineering, University of Oklahoma, Norman, Oklahoma-73071; raman@ou.edu
[3]   Department of Aerospace and Mechanical Engineering, University of Oklahoma, Norman, Oklahoma-7307; yingtao@ou.edu (Y.L.); zsiddique@ou.edu (Z.S.)
*   Correspondence: ahsan@ou.edu

Abstract: Diffusion Models (DMs) are probabilistic models that create realistic samples by simulating the diffusion process, gradually adding and removing noise from data. These models have gained popularity in domains such as image processing, speech synthesis, and natural language processing due to their ability to produce high-quality samples. As DMs are being adopted in various domains, existing literature reviews that often focus on specific areas like computer vision or medical imaging may not serve a broader audience across multiple fields. Therefore, this review presents a comprehensive overview of DMs, covering their theoretical foundations and algorithmic innovations. We highlight their applications in diverse areas such as media quality, authenticity, synthesis, image transformation, healthcare, and more. By consolidating current knowledge and identifying emerging trends, this review aims to facilitate a deeper understanding and broader adoption of DMs and provide guidelines for future researchers and practitioners across diverse disciplines.

Keywords: Diffusion Models; generative modeling; Synthetic Data Generation; image synthesis; Image-to-Image Translation; text-to-image generation; audio synthesis; time series forecasting; anomaly detection; medical imaging; data augmentation; computational efficiency; uncertainty quantification; Riemannian manifolds; molecular dynamics; super-resolution; semantic image synthesis; zero-shot classification; atmospheric turbulence correction

---

1. Introduction

A **Diffusion Model** (DM) is a type of generative model that creates data by reversing a diffusion process, which incrementally adds noise to the data until it becomes a Gaussian distribution. First introduced by Sohl-Dickstein et al. (2015), these models have shown exceptional performance in producing high-quality samples across various fields, such as image, audio, and video synthesis [1,2]. The process involves an iterative procedure where the model is trained to predict the noise that has been added to the sample at each step, effectively learning to denoise data. This approach has led to significant advancements in generating detailed and coherent outputs, making DM a powerful tool for tasks that require high fidelity generation, such as text-to-image synthesis and improving low-resolution images [3]. Figure 1 illustrates a DM introduced for high-resolution image synthesis.
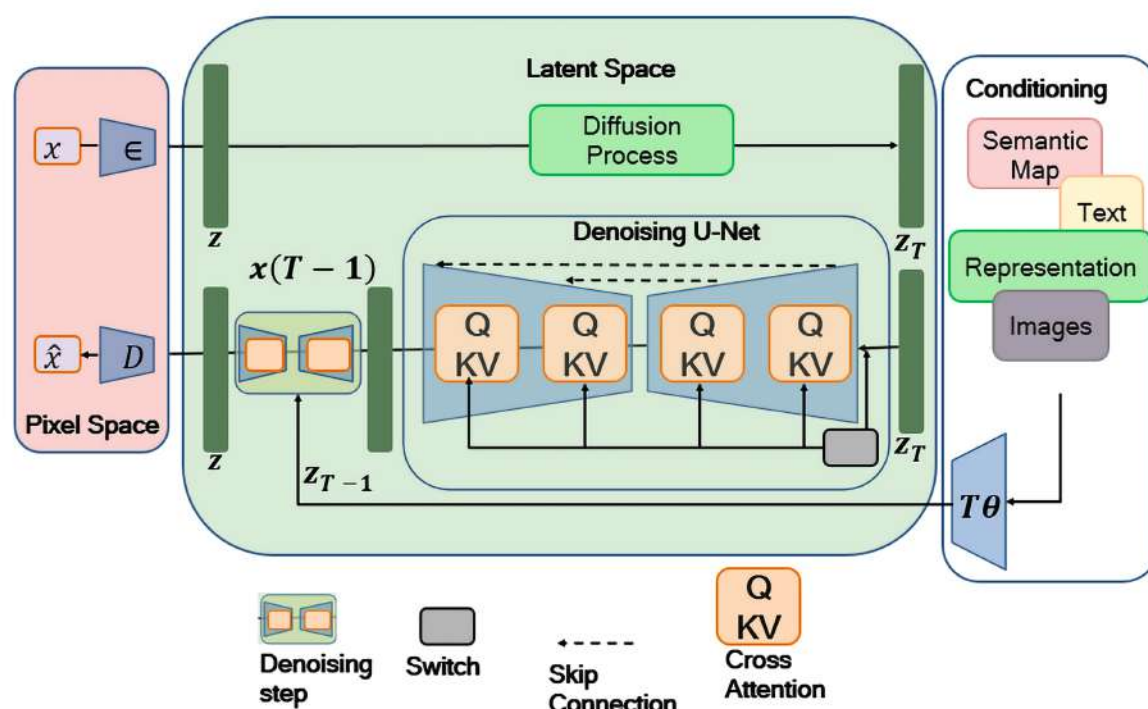
Figure 1. An example of Diffusion-based models. From the figure, it can be observed that the model uses cross-attention mechanisms to enhance image synthesis. This approach allows the model to integrate different types of input information, such as text or semantic maps, to control the image generation process more effectively. The figure shows how these inputs are processed and incorporated into the model to produce high-quality images [4].

Diffusion Models (DMs) have become popular in several areas, particularly in image generation, where they create photorealistic images, art, and edits based on textual descriptions [3,5]. They are also becoming popular in Natural Language Processing (NLP) for text generation and enhancement, demonstrating an ability to produce coherent and contextually relevant text [6]. In audio synthesis, DMs are used to generate realistic soundscapes, music, and human-like speech, pushing the boundaries of creative and communicative Artificial Intelligence (AI) applications [7]. Moreover, their application extends to molecular and material science for designing new chemical compounds and materials, demonstrating their versatility. The popularity of DMs rises from their robustness, flexibility, and the high fidelity of the generated outputs, positioning them as a groundbreaking tool in AI-driven creative and scientific fields [8].

Figure 2 provides a statistical overview of the last five years of published papers on DMs in various disciplines. From Figure 2(a), it can be observed that the number of papers published since 2020 has been constantly growing. Figure 2(b) shows that medicine dominates with 29% of the publications, followed by computer science with 17% and engineering with 14%. Fields such as chemistry and materials science have fewer publications, comprising 4% and 6% of the total, respectively. These trends highlight the extensive use of DMs in medicine and computer science, while their potential in other areas remains less explored.

This review aims to provide a comprehensive overview of DMs across various domains, helping the general audience understand their ability and versatility. By presenting diverse applications, this review encourages interdisciplinary collaboration and innovation, potentially addressing open challenges in less-explored fields beyond traditional applications like computer vision.
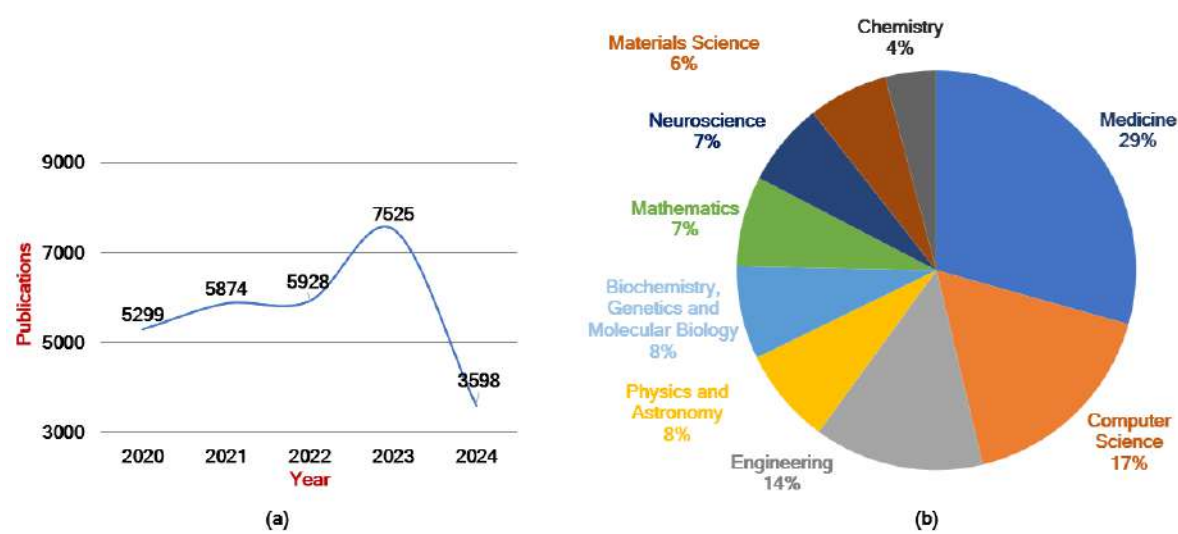
Figure 2. Statistics on (a) the number of papers published over the last five years in DMs and (b) the percentage of published papers across various domains.

1.1. Motivation and Uniqueness of This Survey

The rapid advancements in DMs across various domains show their potential and versatility. Despite the increasing number of publications, existing surveys often focus on specific applications or narrow fields, leaving a gap in reviews that cover the wide range of DM applications. Considering this opportunity, this survey aims to address the gap in the existing literature by providing a comprehensive overview of DMs.

Our **contributions** are summarized below:

❑ This survey considers several key aspects of DMs, including theory, algorithms, innovations, media quality, image transformation, healthcare applications, and more. We provide an overview of relevant literature up to March 2024, highlighting the latest techniques and advancements.

❑ We categorize DMs into three main types: Denoising Diffusion Probabilistic Models (DDPMs), Noise-Conditioned Score Networks (NCSNs), and Stochastic Differential Equations (SDEs), which aids in understanding their theoretical foundations and algorithmic variations.

❑ We highlight novel approaches and experimental methodologies relevant to the application of DMs, considering data types, algorithms, applications, datasets, evaluations, and limitations.

❑ Finally, we discuss the findings, identify open issues, and raise questions about future research directions in DMs, aiming to guide researchers and practitioners.

Figure 3 illustrates the framework of DMs based on the referenced literature used in this study, discussed in Sections 2–8.
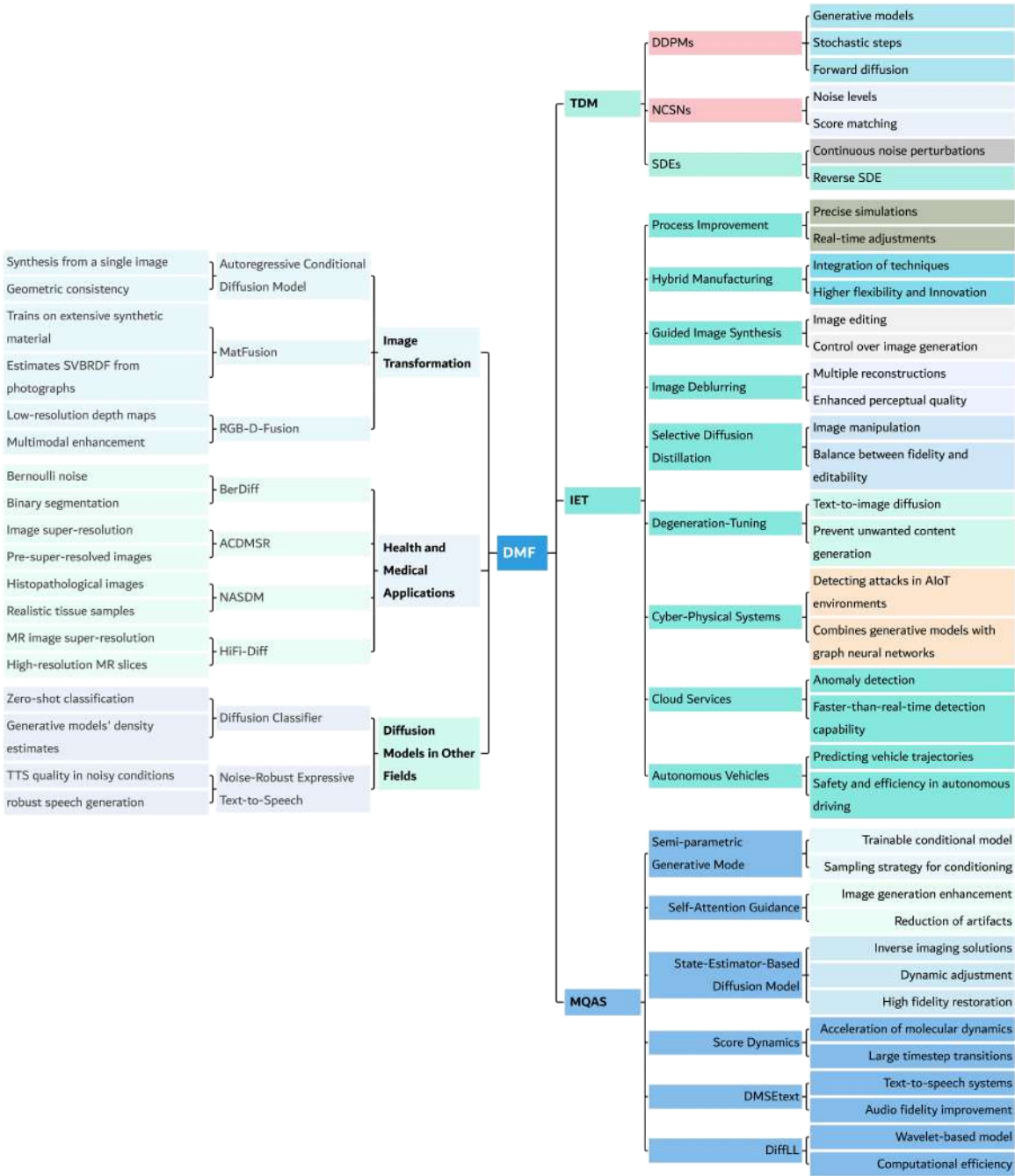
Figure 3. Comprehensive overview of DMs: This diagram categorizes various DMs and their applications across different fields. DMF – Diffusion Models framework, TDM – Types of Diffusion Models, IET – Image Enhancement and Transformation, MQAS – Media Quality, Authenticity, and Synthesis, DDPMs – Denoising Diffusion Probabilistic Models, NCSNs – Noise-Conditioned Score Networks, SDEs – Stochastic Differential Equations.

1.2. Search Strategy

Data were sourced from Scopus, initially identified 3,746 articles using the title, abstract, and keywords with the search terms 'Diffusion Model' AND ('image' OR 'audio' OR 'text' OR 'speech'). Restricting the search to English-language, peer-reviewed, and open-access papers published between 2020 and 2024 reduced the number to 473. Further filtering excluded terms such as 'human,' 'controlled study,' 'job analysis,' 'quantitative analysis,' 'comparative study,' 'specificity,' and other irrelevant keywords, resulting in 326 papers.

One researcher (Y.L.) imported these 326 journal articles into Excel CSV files for detailed analysis. Later, Excel's duplication tools were used to identify and remove duplicates. The titles and abstracts of the remaining papers were assessed by two independent reviewers (M.A. and Z.S.), identifying 65 relevant documents. Additionally, 20 more relevant papers were included, resulting in a total of 85 papers across various fields.

2. General Overview of DMs

DMs are a type of generative model that simulates the diffusion process to construct or reconstruct data distributions through stochastic processes. This involves a dual-phase operation where noise is incrementally added and subsequently reversed [9]. The algorithmic backbone of DMs contains several key phases [1,9]:

- **Initialization:** Start with data in its original form x_0.
- **Forward Process (Noise Addition):** Gradually add noise over T timesteps, transforming the data from x_0 to x_T based on a predefined noise schedule β.
- **Reverse Process (Denoising):** Sequentially estimate x_{t-1} from x_t using the learned parameters θ, effectively reversing the noise addition to either reconstruct the original data or generate new data samples.
- **Input:** Original data $X = \{x_1, x_2, \ldots, x_n\}$, Total timesteps T, Noise schedule β.
- **Output:** Denoised or synthesized data X'.
- **Training:** Train the model to approximate the reverse noise addition process by learning the conditional distributions $p_\theta(x_{t-1}|x_t)$ for each timestep t, from T down to 1.
- **Data Synthesis:** Begin with a sample of random noise x_T and iteratively apply the learned reverse process:

$$x'_{t-1} = \text{Sample from } p_\theta(x_{t-1}|x_t)$$

culminating in x'_0, the final synthesized or reconstructed data.

Types of DM. Over the years, several diffusion-based models have been proposed, each contributing uniquely to the advancement of generative modeling. Figure 4 illustrates some of the important and influential DM along with their timeline. Among them, three DMs are very popular and widely adopted due to their impact on various applications: DDPMs, NCSNs, and SDEs.
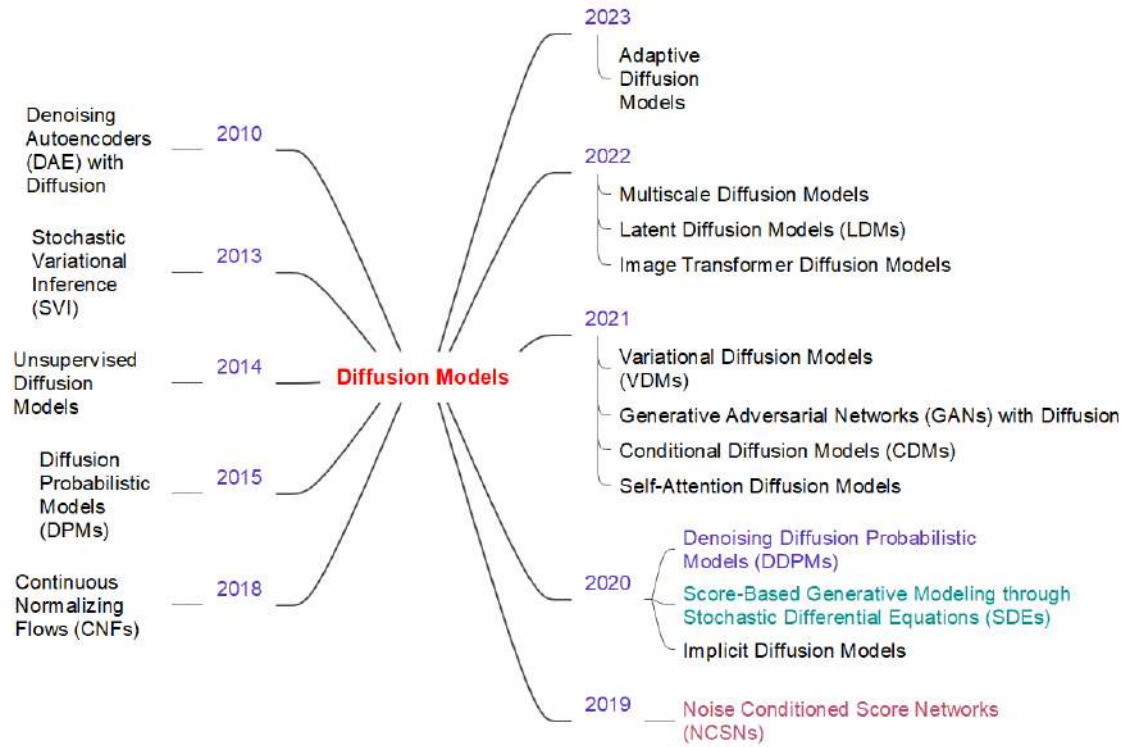
Figure 4. Timeline of different DMs from 2010 to 2023. The three main DMs, such as NCSNs, DDPMs, and SDEs, are highlighted with different colors.

2.1. DDPMs

Introduced by Ho et al. (2020), DDPMs are generative models that transform noise into data through a series of gradual stochastic steps [2].

Forward Diffusion Process. The forward diffusion process incrementally adds Gaussian noise to the data, transforming it into a noise distribution. Given a data point $\mathbf{x}_0$, the process is defined over T timesteps. At each timestep t, Gaussian noise is added to the data:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \tag{1}$$

By the end of the diffusion process, the data is effectively transformed into pure Gaussian noise.

Reverse Denoising Process. The reverse denoising process aims to recover the original data from the noisy observations. This is modeled using a parameterized reverse Markov chain, where the goal is to estimate the posterior distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$. However, this posterior is not directly computable, so a Neural Network (NN) p_θ is employed to approximate it:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)). \tag{2}$$

The network is trained to minimize the following Variational Lower Bound (VLB) on the negative Log-likelihood:

$$L_{\text{vlb}} = \mathbb{E}_q\left[\sum_{t=1}^{T} D_{\text{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \,||\, p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)) - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)\right], \tag{3}$$

Training Objective. To simplify the training process, Ho et al. (2020) proposed a reparameterization of the training objective that aligns closely with denoising score matching. The simplified objective can be expressed as [2]:

$$L_{\text{simple}} = \mathbb{E}_{t,\mathbf{x}_0,\epsilon}\left[||\epsilon - \epsilon_\theta(\mathbf{x}_t,t)||^2\right], \tag{4}$$

Sampling from DDPMs. Once trained, sampling from DDPMs involves running the reverse process starting from pure Gaussian noise $\mathbf{x}_T \sim \mathcal{N}(0,\mathbf{I})$ and iteratively applying the learned denoising steps:

$$\mathbf{x}_{t-1} = \mu_\theta(\mathbf{x}_t,t) + \Sigma_\theta(\mathbf{x}_t,t)\cdot\mathbf{z}, \tag{5}$$

2.2. NCSNs

Introduced by Song et al. (2019), NCSNs aim to generate data by estimating the gradients of the data distribution, known as Score Functions, at various noise levels [10].

Forward Diffusion Process. The forward diffusion process in NCSNs involves gradually perturbing the data with Gaussian noise of increasing intensity, similar to the procedure described for **DDPMs**. Given an initial data point $\mathbf{x}_0$, the data is progressively noised to generate a sequence of noisy data points $\{\mathbf{x}_t\}$ over T timesteps.

Learning the Score Function. The core of NCSNs lies in learning the score function, which is the gradient of the log data density $\nabla_\mathbf{x}\log p(\mathbf{x})$. However, instead of directly learning this for the original data distribution, NCSNs learn it for the perturbed data at various noise levels. An NN $s_\theta(\mathbf{x},\sigma_t)$ is trained to approximate these score functions for different noise levels σ_t:

$$s_\theta(\mathbf{x},\sigma_t) \approx \nabla_\mathbf{x}\log p_\sigma(\mathbf{x}), \tag{6}$$

Training Objective. The training objective for NCSNs involves minimizing a denoising score matching objective, which encourages the NN to accurately predict the score function. This loss function can be expressed as:

$$L_{\text{ncsn}} = \mathbb{E}_{\mathbf{x}_0,\mathbf{ffl},\sigma}\left[\lambda(\sigma)||s_\theta(\mathbf{x}_0 + \sigma\mathbf{ffl},\sigma) - \frac{\mathbf{ffl}}{\sigma}||^2\right], \tag{7}$$

Sampling from NCSNs. Sampling from NCSNs involves using the learned score function to iteratively denoise a sample of pure Gaussian noise. This is typically done using Langevin dynamics, a method that iteratively refines the noisy sample by adding the score function and some additional noise:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \frac{\alpha}{2}s_\theta(\mathbf{x}_t,\sigma_t) + \sqrt{\alpha}\mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0,\mathbf{I}), \tag{8}$$

2.3. SDEs

Introduced by Song et al. (2020), SDEs leverage the mathematical framework of SDE to model the data generation process through continuous noise perturbations and denoising [9].

Forward Diffusion Process. In the SDE framework, the forward diffusion process involves transforming the data into a noise distribution through a continuous-time stochastic process, similar to the procedure described for **DDPMs**. This is typically modeled by an Itô SDE:

$$d\mathbf{x} = f(\mathbf{x},t)dt + g(t)d\mathbf{w}, \tag{9}$$

where $\mathbf{x}$ represents the data, t is the time variable, $f(\mathbf{x},t)$ is the drift coefficient, $g(t)$ is the diffusion coefficient, and $\mathbf{w}$ is a standard Wiener process [9].

Reverse SDE Process. The reverse process aims to revert the noisy data back to its original form by solving the reverse-time SDE. This process is governed by:

$$d\mathbf{x} = [f(\mathbf{x},t) - g(t)^2\nabla_\mathbf{x}\log p_t(\mathbf{x})]dt + g(t)d\bar{\mathbf{w}}, \tag{10}$$

where $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ is the score function, which represents the gradient of the log-density of the data at time t, and $\bar{\mathbf{w}}$ is a standard Wiener process running backward in time [11]. The score function is estimated using an NN trained on denoising score matching [12].

Training Objective. The training of Score-based Generative models involves learning the score function at different noise levels. The objective function for training is typically the denoising score matching loss:

$$L_{\text{score}} = \mathbb{E}_{p_0(\mathbf{x}),\epsilon,t}\left[\lambda(t)||s_\theta(\mathbf{x}_t,t) - \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0)||^2\right], \tag{11}$$

Sampling from SDEs. Sampling from the trained Score-based model involves solving the reverse-time SDE starting from a sample of Gaussian noise. Numerical solvers, such as Euler-Maruyama or Predictor-Corrector methods, are used to approximate the reverse SDE and generate data samples [9].

3. General Applications of DMs

Over the years, interest in DMs has grown exceptionally due to their ability to generate high-quality, realistic, and diverse data samples, making them highly deployable in several cutting-edge applications. Some of the most popular areas where DMs are used extensively include:

- ❑ **Image Synthesis:** DMs are used to create detailed, high-resolution images from a distribution of noise. They can generate new images or improve existing ones by improving clarity and resolution, making them particularly useful in fields such as digital art and graphic design [13].
- ❑ **Text Generation:** DMs are capable of producing coherent and contextually relevant text sequences. This makes them suitable for applications such as creating literary content, generating realistic dialogues in virtual assistants, and automating content generation for news articles or creative writing [14].
- ❑ **Audio Synthesis:** DMs can generate clear and realistic audio from noisy signals. This is valuable in music production, where it's necessary to create new sounds or improve the clarity of recorded audio, as well as in speech synthesis technologies used in various assistive devices [7].
- ❑ **Healthcare Applications:** Although not limited to medical imaging, DMs assist in synthesizing medical data, including Magnetic Resonance Imaging (MRI), Computed Tomography (CT) scans, and other imaging modalities. This ability is vital for training medical professionals, improving diagnostic tools, and developing more precise therapeutic strategies without compromising patient privacy [15].

Table 1 summarizes some of the renowned papers in DMs from 2020 to 2023, their proposed algorithms, used datasets, and applications. Different colors are used to distinguish between various algorithms and application types. From Table 1, it can be observed that most of the papers primarily focus on image-based applications, such as image generation, segmentation, and reconstruction.

Table 1. Some of the important papers in DMs from 2020 to 2023, along with their proposed algorithms, used datasets, and applications. Different colors are used to distinguish between various algorithms and application types.

Year	Proposed Algorithm	Used Datasets	Applications
2020	DDPMs [2]	CIFAR-10 [16], LSUN [17], CelebA [18]	Image generation
2020	Score-Based DMs [19]	CIFAR-10, CelebA, LSUN	Image generation
2020	SDEs [9]	CIFAR-10, CelebA, LSUN, FFHQ	Image generation
2021	Classifier-guided DMs (CDMs) [20]	ImageNet, LSUN, CIFAR-10	Image generation
2021	Variational Diffusion Models (VDMs) [21]	CIFAR-10, CelebA, LSUN	Image generation
2021	Improved DDPMs [22]	CIFAR-10, CelebA, LSUN	Image generation
2021	Diffusion Waveform (DiffWave) [23]	LJSpeech, VCTK	Audio generation
2021	Segmentation Diffusion (SegDiff) [24]	Cityscapes, Pascal VOC	Image segmentation
2021	Generative LIkelihood-based DEcompression (GLIDE) [25]	MS-COCO, ImageNet	Image reconstruction
2022	Latent Diffusion Models (LDMs) [4]	LAION-400M, CelebA-HQ	Image generation, Text-to-image
2022	Image Transformers [26]	ImageNet, COCO	Image generation
2022	Multiscale Diffusion Models [27]	ImageNet, CIFAR-10, LSUN	Image generation
2022	Video-DDPM [28]	Kinetics-600, UCF-101	Video generation
2023	Adaptive Diffusion Models [29]	CIFAR-10, CelebA, FFHQ	Image generation

4. Innovations and Experimental Techniques in DMs

Several studies have utilized DM-based approaches because of their flexibility and effectiveness in various applications. Figure 5 illustrates a DM introduced for guided image synthesis through initial image editing.
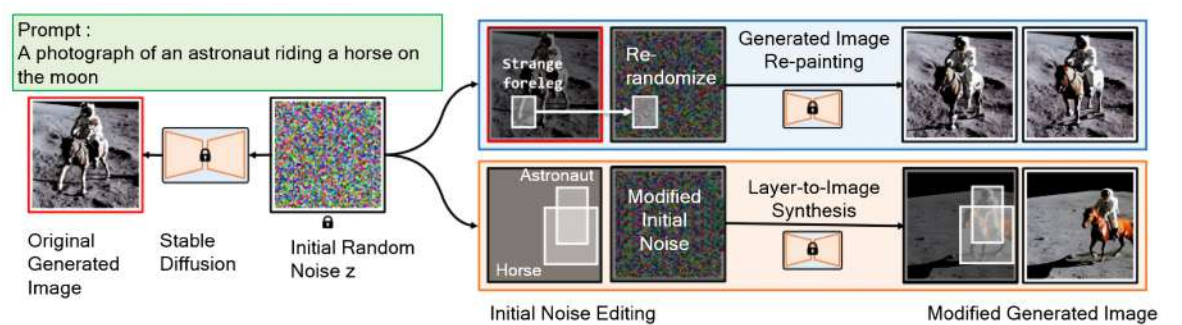
Figure 5. Mao et al. (2023) explored how the initial image influenced the image generation process and proposed a new method to control it by altering the initial random noise. They demonstrated two applications: layout-to-image synthesis, which created objects in specified locations, and re-painting, which allowed users to change specific portions of an image while keeping the rest unchanged [30].

Whang et al. (2022) introduced a diffusion-based stochastic blind image deblurring technique. This approach leveraged DMs to produce multiple plausible reconstructions for blurred images, significantly improving perceptual quality. Evaluations on the GoPro dataset showed impressive results with

metrics such as FID of 4.04, Kernel Inception Distance (KID) of 0.98, Learned Perceptual Image Patch Similarity (LPIPS) of 0.059, Peak Signal-to-Noise Ratio (PSNR) of 31.66, and Structural Similarity Index Measure (SSIM) of 0.948 [31]. However, high computational demands pose limitations for real-time applications, suggesting a need for optimized sampling or network architecture adjustments.

Chung et al. (2022) introduced the **Come-Closer-Diffuse-Faster (CCDF)** sampling strategy to address the slow sampling rate of DMs. CCDF started from a forward-diffused state, reducing required sampling steps using the contraction theory of stochastic difference equations. This method enhanced tasks like super-resolution, image inpainting, and MRI reconstruction, showing improved FID scores and PSNR across datasets [32]. However, selecting the optimal starting point remains challenging and requires several trial-and-error approaches.

Wang et al. (2023) introduced **Selective Diffusion Distillation (SDD)** for improved image manipulation using conditional DMs. SDD trained a feedforward network guided by a DM, addressing the fidelity-editability trade-off. The framework used a Hybrid Quality Score (HQS) to select the optimal semantic timestep, improving image quality and semantic accuracy. SDD outperformed other methods, achieving an FID of 6.066 and a Contrastive Language-Image Pre-training (CLIP) similarity of 0.2337 [33]. However, a significant limitation remains in the necessity of carefully selecting HQS thresholds to balance manipulation and quality.

Li et al. (2023) introduced **Object Motion Guided Human Motion Synthesis (OMOMO)**, a framework for synthesizing human motion based on object motion, specifically for large object manipulation. OMOMO used two denoising processes to predict hand positions from object motion and synthesize full-body poses, ensuring accurate contact and realistic motion. By capturing motion via visual-inertial odometry on a smartphone, OMOMO showed potential for applications in virtual reality, augmented reality, and robotics. Their comprehensive dataset demonstrated the framework's ability to generalize to unseen objects. OMOMO achieved high accuracy with a Mean Per-Joint Position Error (MPJPE) of 12.42, a precision score of 0.70, and an F1 score of 0.72 [34]. However, the issue of intermittent object contacts remains unaddressed. Additionally, the predicted hand motions are less plausible, as indicated by lower F1 and precision scores.

Ni et al. (2023) introduced **Degeneration-Tuning (DT)** to control text-to-image DMs like Stable Diffusion. DT prevents the generation of unwanted content by detaching undesirable textual concepts from image outputs using a scrambled grid. Integrated with Control Network (ControlNet), DT maintains high-quality generation for general content with minimal metric impact (FID from 12.61 to 13.04, IS from 39.20 to 38.25) [35]. However, DT's slow sampling speeds, reliance on predefined prompts, and risk of over-degeneration limit its effectiveness, requiring further refinement to balance control and generative abilities.

Yan et al. (2022) introduced **Temporal and Feature Pattern-based Diffusion Probabilistic Model (TFDPM)**, a model for detecting attacks in cyber-physical systems within Artificial Intelligence of Things (AIoT). TFDPM combined energy-based generative models and Graph Neural Networks to handle complex data and correlations. It extracted temporal and feature patterns to guide a diffusion probabilistic model, improving accuracy and sensitivity [36]. Their proposed TFDPM outperformed many of the existing State-of-the-Art (SOTA) techniques on PUMP, SWAT, and WADI datasets in terms of attack detection accuracy and speed. However, challenges remained in modeling discrete signals and exploring more robust configurations. Additionally, the model faced difficulties in ensuring scalability and adaptability across diverse AIoT environments as well.

Lee et al. (2023) introduced **Metric Anomaly Anticipation (MAAT)**, a framework for faster-than-real-time anomaly detection in cloud services. MAAT uses a two-stage process: multi-step forecasting with a Conditional Denoising Diffusion Model, followed by anomaly detection with an isolation forest. Tested on AIOps18, Hades, and Yahoo!S5 datasets, MAAT outperformed existing methods in speed, precision, and reliability [37]. However, its focus on cloud-service metrics and a static time horizon limits its applicability to other time-series data and dynamic conditions. Furthermore, its performance

with ultra-high-frequency data remains untested, indicating a need for further research to extend its capabilities and validate its effectiveness in these areas.

Chen et al. (2023) introduced **Equivariant Diffusion (EquiDiff)**, a deep generative model designed to improve the security and efficiency of autonomous vehicles by predicting vehicle routes. EquiDiff uses conditional DMs with an SO(2)-equivariant transformer, integrating historical trajectory data and Gaussian noise to generate future paths while respecting geometric constraints. It also incorporates Recurrent Neural Networks and Graph Attention Networks to model social interactions among vehicles. Evaluated on the NGSIM dataset, EquiDiff outperformed baseline models in short-term prediction accuracy, achieving a Root Mean Square Error (RMSE) of 0.55 at 1 second and 4.01 at 5 seconds [38]. However, it showed higher errors in long-term predictions, which highlights limitations in the model's ability to maintain accuracy over extended periods. This suggests the need for further refinement to address these long-term prediction challenges.

Table 2 summarizes some of the referenced literature that uses innovative and experimental techniques in developing DMs, including applications in content security, cyber-physical system attack detection, anomaly anticipation, image deblurring, acceleration for inverse problems, image manipulation, and human motion synthesis.

Table 2. Innovations and experimental techniques introduced by the referenced literature in the domain of DMs. FID: Frechet Inception Distance, IS: Inception Score, P_r: Precision, R_e: Recall, LPIPS: Learned Perceptual Image Patch Similarity, PSNR: Peak Signal-to-Noise Ratio, SSIM: Structural Similarity Index Measure, KID: Kernel Inception Distance, MPJPE: Mean Per Joint Position Error. Best results are highlighted in bold.

Ref.	Algorithms	Applications	Dataset	Evaluations	Limitations
[35]	DT for Content Shielding in Stable DMs	Content shielding in text-to-image Diffusion Models using DT to prevent generation of unwanted concepts	COCO 30K	FID post-DT: 13.04, IS post-DT: 38.25	DT may limit model's flexibility for diverse contexts.
[36]	TFDPM	Detecting cyber-physical system attacks using TFDPM with Graph Attention Networks for channel data correlation	PUMP, SWAT, WADI	P_r: 0.96, R_e: 0.91, F1: 0.91	Struggles with discrete signal modeling, needs SDE frameworks for better generative capabilities.
[37]	Maat: Anomaly Anticipation for Cloud Services	Anomaly anticipation using a two-stage Diffusion Model for cloud services, integrating metric forecasting and anomaly detection	AIOps18, Hades, Yahoo!S5	**P_r: 0.97, R_e: 0.91, F1: 0.91**	Limited generalizability and adaptability post-training.
[31]	Diffusion-based Stochastic Blind Image Deblurring	Blind image deblurring using Diffusion Models for multiple reconstructions	GoPro	FID: 4.04, KID: 0.98, LPIPS: 0.06, PSNR: 31.66, SSIM: 0.95	High computational demands, needs optimized sampling or network architecture.
[32]	Come-Closer-Diffuse-Faster	Accelerating CMDs for applications like super-resolution and MRI reconstruction	FFHQ, AFHQ, fastMRI	FID varies; **PSNR: 33.41** (best MRI case)	Optimal starting values (t0) vary, needs automation for practical deployment.

Table 2. *Cont.*

Ref	Algorithms	Applications	Dataset	Evaluations	Limitations
[33]	Selective Diffusion Distillation	Image manipulation balancing fidelity and editability without excessive noise trade-offs	N/A	FID: 6.07, CLIP Similarity: 0.23	Reliance on correct timestep selection for semantic guidance may limit flexibility.
[34]	Object Motion Guided Human Motion Synthesis (OMOMO)	Full-body human motion synthesis guided by object motion using a conditional Diffusion Model	Custom dataset	MPJPE: 12.42, Troot: 18.44, Cprec: 0.82, Crec: 0.70, F1: 0.72	Limited representation of dexterous hand movements and intermittent contact scenarios.
[2]	DDPMs	Image generation using DDPMs	CIFAR-10, LSUN, CelebA	FID: 3.17, IS: 9.46	High computational cost and slow sampling speed.
[19]	Improved Techniques for Training Score-Based Generative Models	Improved image generation using score-based models	CIFAR-10, CelebA, LSUN	FID: 2.87, IS: 9.68	Training complexity and large computational resources required.
[9]	Score-Based Generative Modeling through SDEs	Image generation using SDEs for better quality	CIFAR-10, CelebA, LSUN, FFHQ	FID: 2.92, IS: 9.62	SDE-based models can be computationally expensive.
[20]	Diffusion Models Beat Generative Adversarial Networks (GANs) on Image Synthesis	Image synthesis outperforming GANs using Diffusion Models	ImageNet, LSUN, CIFAR-10	**FID: 2.97**, IS: 9.57	Large model size and slow training times.
[21]	VDMs	Image generation using variational Diffusion Models	CIFAR-10, CelebA, LSUN	FID: 3.12, IS: 9.53	Complex model design and high computational cost.
[22]	Improved Denoising Diffusion Probabilistic Models	Enhanced DDPMs for better image quality	CIFAR-10, CelebA, LSUN	FID: 3.05, IS: 9.50	Requires extensive hyperparameter tuning.
[4]	High-Resolution Image Synthesis with Latent Diffusion Models (LDMs)	High-resolution image and text-to-image synthesis	LAION-400M, CelebA-HQ	**FID: 1.97**, IS: 10.32	High memory usage and computational cost.
[26]	Image Transformers with Autoregressive Models for High-Fidelity Image Synthesis	High-fidelity image synthesis using transformers	ImageNet, COCO	FID: 2.30, IS: 9.95	Transformer models are computationally intensive.
[27]	Cascaded Diffusion Models for High-Fidelity Image Generation	High-fidelity image generation using multiscale Diffusion Models	ImageNet, CIFAR-10, LSUN	FID: 2.15, IS: 9.88	Cascaded models require extensive computational resources.
[29]	Optimizing Diffusion Models for Image Synthesis	Adaptive Diffusion Models for better image synthesis	CIFAR-10, CelebA, FFHQ	**FID: 1.89**, IS: 10.45	Adaptive models can be complex and resource-intensive.
[23]	DiffWave	Audio generation using Diffusion Models	LJSpeech, VCTK	FID: 3.67, PSNR: 34.10	High computational cost and slow sampling speed.

Ref	Algorithms	Applications	Dataset	Evaluations	Limitations
[28]	Video Diffusion Models (Video-DDPM)	Video generation using Diffusion Models	Kinetics-600, UCF-101	FID: 3.85, SSIM: 0.92	High computational demands and slow training times.
[24]	SegDiff	Image segmentation using Diffusion Models	Cityscapes, Pascal VOC	FID: 3.50, SSIM: 0.87	Limited scalability to larger datasets.
[25]	GLIDE	Photorealistic image generation and editing with text guidance	MS-COCO, ImageNet	FID: 3.21, IS: 9.67	Text-guided models require extensive training data.

5. Media Quality, Authenticity, and Synthesis

Several studies propose DMs to improve media quality and create realistic samples. Figure 6 illustrates an orthogonal, semi-parametric DM, which includes a trainable Conditional Generative Model, an external database for visual examples, and a sampling strategy to retrieve subsets for conditioning the model [39].
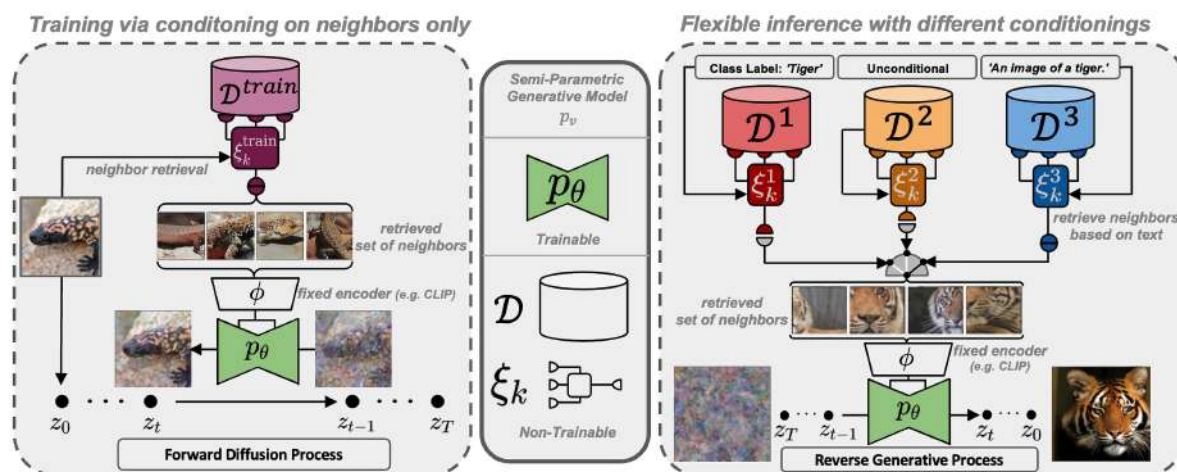

Figure 6. A semi-parametric generative model consists of a trainable Conditional Generative Model $p_\theta(x|\cdot)$, an external database D with visual examples, and a sampling strategy ξ_k that selects a subset $M(k)_D \subseteq D$ for conditioning p_θ. To train p_θ to create consistent scenes using $M(k)_D$, ξ_k retrieves the nearest neighbors of each target example from D. By adjusting D and ξ_k during inference, the model can flexibly sample with post-hoc conditioning on class labels (ξ_k^1) or text prompts (ξ_k^3), and perform zero-shot stylization [39].

Hong et al. (2023) introduced **Self-Attention Guidance (SAG)** to improve image generation using Denoising Diffusion Models (DDMs). SAG leverages self-attention maps to focus on significant areas, reducing artifacts and improving image quality. Evaluation on various platforms revealed that SAG significantly improved both FID and IS compared to existing methods [40].

Ji et al. (2024) introduced a **Learnable State-Estimator-based DMs** for inverse imaging problems, restoring clean images from corrupted inputs with high fidelity. This method uses a state estimator to dynamically adjust the diffusion process within a latent space, achieving computational efficiency and avoiding extensive training. Evaluated on tasks like inpainting, deblurring, and JPEG compression restoration, it showed strong performance, particularly on the FFHQ dataset with a PSNR of 27.98, LPIPS of 0.0939, and FID of 25.453 [41]. However, the model relies on current generative abilities and needs domain-specific adaptations for broader applications.

Tian et al. (2023) introduced **Diffusion Model for Speech Enhancement text (DMSEtext)**, a conditional DMs designed to enhance speech quality in Text-to-Speech (TTS) systems by addressing audio degradations. Operating in the log Mel-spectrogram domain, it uses text transcriptions to improve audio fidelity. DMSEtext achieved a Mean Opinion Score (MOS) of 4.32 for cleanliness and 4.17 overall, with a reduced Phoneme Error Rate (PER) of 17.6%, indicating improved clarity and authenticity [42]. However, its performance depends on the quality of text transcription and varies under different audio types.

Jiang et al. (2023) introduced **Diffusion Model for Low-Light (DiffLL)**, a framework for improving low-light images using a Wavelet-based Conditional Diffusion Model. This model increases inference speed and reduces computational demands while maintaining high image quality. A High-Frequency Restoration Module improves image details. DiffLL outperformed current methods on benchmarks like LOL-v1, LOLv2-real, and LSRW in PSNR, SSIM, LPIPS, and FID metrics [43]. However, it struggles with extremely low-light conditions and is not optimized for real-time video processing. Additionally, the study did not consider real-time video support and handling diverse lighting conditions, which remain areas for further investigation.

Dong et al. (2023) proposed **Controlled Language-Image Pretraining Sonic (CLIPSonic)**, a text-to-audio synthesis method using unlabeled videos and pretrained language-vision models. It employs a conditional DMs to generate audio by translating text embeddings into image embeddings, improving zero-shot modality transfer. CLIPSonic demonstrated competitive performance on VGGSound and MUSIC datasets [44]. However, its effectiveness is limited by the quality of pretrained models, distribution mismatches, and training complexity, posing scalability challenges.

Liu et al. (2023) proposed **Semantic Diffusion Guidance (SDG)**, a framework that improves DDMs with fine-grained control using language, image, or both modalities. SDG integrates guidance into pretrained models via image-text or image matching score gradients which eliminates the need for retraining. It enables text-guided image synthesis on datasets without text annotations using CLIP-based guidance and demonstrates better accuracy over baseline models such as Iterative Latent Variable Refinement [45] and StyleGAN+CLIP. On the FFHQ dataset, the proposed SDG models achieved an FID score of 14.37 and a top-1% accuracy of 0.520. Additionally, the ablation studies on LSUN showed minor performance improvements with different scaling factors [46]. However, SDG's effectiveness depends on the accuracy of pretrained models and their ability to process guidance signals. Additionally, the framework poses potential risks of misuse which necessitates the requirement of ethical guidelines to ensure responsible deployment.

Cai et al. (2023) introduced **Diffusion Dreamer (DiffDreamer)**, an unsupervised framework for scene extrapolation using conditional DMs to generate novel views from given images. By training on internet-collected nature images, DiffDreamer refines projected RGBD images through guided denoising steps, conditioned on multiple past and future frames. It significantly outperforms previous GAN-based methods in quality and consistency. On the LHQ dataset, DiffDreamer achieved an FID score of 51.0 over 100 steps and 34.49 over 20 steps [47]. However, DiffDreamer cannot synthesize novel views in real time due to the computational intensity of DMs and does not ensure content diversity in extended extrapolations.

Carrillo et al. (2023) proposed an interactive approach for line art colorization using conditional Diffusion Probabilistic Models, allowing users to input initial color strokes. The system integrates these inputs via a dual conditioning strategy, producing diverse, high-quality images. Their model outperforms SOTA methods by achieving an SSIM of 0.81, LPIPS of 0.14, and FID of 6.15. However, the model's accuracy depends on the quality of user input, and the complex conditioning strategy may cause computational inefficiencies, which could eventually affect scalability as well [48].

Mao et al. (2023) introduced **Sketch-Driven Fusion (SketchFFusion)**, a model for sketch-guided image editing using a conditional Diffusion Model. SketchFFusion maintains the integrity of sketches while editing, simulating human sketch styles and preserving structural details. On the CelebA-HQ dataset, it outperformed SOTA methods with an FID of 9.07, PSNR of 26.74, and SSIM of 0.8822. The

model was also tested on the COCO-AIGC dataset, demonstrating adaptability across various scenes and objects [49]. However, SketchFFusion currently only supports binary sketches, limiting its use to black-and-white inputs.

Luo et al. (2023) introduced **Semantic-Conditional Diffusion Networks** for image captioning, leveraging DMs to improve visual-language alignment and coherence. Unlike traditional transformer models, their approach uses semantic priors from cross-modal retrieval and refines captions through multiple Diffusion Transformer layers. This dynamic integration of image and text features enhances caption relevance and accuracy. On the Common Objects in Context dataset, it achieved a Consensus-based Image Description Evaluation score of 131.6 and a BLEU-4 score of 39.4, outperforming SOTA models [50].

Table 3 summarizes some of the referenced literature that proposes different diffusion-based approaches to improve media quality and increase authenticity.

Table 3. Improving media quality using diffusion-based approaches as demonstrated in the existing literature. FID: Frechet Inception Distance, IS: Inception Score, P_r: Precision, R_e: Recall, LPIPS: Perceptual Image Patch Similarity, PSNR: Peak Signal-to-Noise Ratio, SSIM: Structural Similarity Index Measure, KID: Kernel Inception Distance, MPJPE: Mean Per Joint Position Error. Best results are highlighted in bold.

Ref	Algorithms	Applications	Dataset	Evaluations	Limitations
[40]	SAG in DDMs	Image generation improvement	ImageNet, LSUN	FID: **2.58**, sFID: 4.35	Needs broader application integration.
[41]	Learnable State-Estimator-Based Diffusion Model	Inverse imaging problems (inpainting, deblurring, JPEG restoration)	FFHQ, LSUN-Bedroom	PSNR: 27.98, LPIPS: 0.09, FID: 25.45	Limited generative capabilities, needs domain adaptation.
[51]	Score Dynamics (SD)	Accelerating molecular dynamics simulations	Alanine dipeptide, short alkanes in aqueous solution	Wall-clock speedup up to 180X	Requires large datasets; generalization challenges.
[42]	CMDs for Speech Enhancement (DM-SEtext)	Speech enhancement for TTS model training	Real-world recordings	MOS Cleanliness: **4.32** ± 0.08, Overall Impression: **4.17** ± 0.06, PER: 17.6%	Needs text conditions for best results.
[52]	Conditional Diffusion Model for HDR Reconstruction	HDR image reconstruction from LDR images	Benchmark datasets for HDR imaging	PSNR-μ: 44.11, PSNR-L: 41.73, SSIM-μ: **0.99**, SSIM-L: **0.99**, HDR-VDP-2: 65.52, LPIPS: 0.01, FID: 6.20	Slow inference speed; improve distortion metrics.
[44]	CLIPSonic	Text-to-audio synthesis using unlabeled videos	VGGSound, MUSIC	FAD: CLIPSonic-ZS on MUSIC 19.30, CLIPSonic-PD on MUSIC 13.51; CLAP score: CLIPSonic-ZS on MUSIC 0.28, CLIPSonic-PD on MUSIC 0.25	Performance drop in zero-shot modality transfer.

Table 3. *Cont.*

Ref	Algorithms	Applications	Dataset	Evaluations	Limitations
[46]	SDG	Fine-grained image synthesis with text and image guidance	FFHQ, LSUN	FID: 14.37 (image guidance on FFHQ), 28.38 (text guidance on FFHQ); Top-5 Retrieval Accuracy: 0.742 (image guidance), 0.878 (text guidance)	Potential misuse in image generation.
[47]	DiffDreamer: Conditional Diffusion Model for Scene Extrapolation	Unsupervised 3D scene extrapolation	LHQ, ACID	Achieves low FID scores across various step intervals, e.g., 20 steps: FID: 34.49; 100 steps: FID: 51.00 on LHQ	Real-time synthesis not feasible; limited content diversity.
[48]	Diffusart: Conditional Diffusion Probabilistic Models for Line Art Colorization	Interactive line art colorization with user guidance	Danbooru2021	SSIM: 0.81, LPIPS: 0.14, FID: **6.15**	Bias towards white; limits color diversity.
[49]	SketchFFusion: A Conditional Diffusion Model for Sketch-guided Image Editing	Sketch-guided image editing for local fine-tuning using generated sketches	CelebA-HQ, COCO-AIGC	FID: 9.07, PSNR: 26.74, SSIM: 0.88	Supports only binary sketches; limits color editing.
[50]	Semantic-Conditional Diffusion Networks for Image Captioning	Advanced text-to-image captioning using semantic-driven Diffusion Models	COCO	B@1: 79.0, B@2: 63.4, B@3: 49.1, B@4: 37.3, CIDEr: **131.6**	Lacks real-time processing; needs optimization.
[38]	EquiDiff: Deep Generative Model for Vehicle Trajectory Prediction	Trajectory prediction for autonomous vehicles using a deep generative model with SO(2)-equivariant transformer	NGSIM	RMSE for 5s trajectory prediction shows competitive results	Effective short-term; higher errors in long-term predictions.
[53]	Efficient MRI Synthesis with Conditional Diffusion Probabilistic Models	Efficient synthesis of 3D brain MRIs using a conditional Diffusion Model	ADNI-1, UCSF, SRI International	MS-SSIM: 78.6%	Focused on T1-weighted MRIs; explore more types.

6. Image Transformation and Enhancement

6.1. Image-to-Image Transformation

DMs have shown significant potential in various image-to-image transformation tasks. Existing studies demonstrate that the versatility of DMs helps in improving image quality and generating new images. For instance, Yu et al. (2023) presented an autoregressive **Cascade Multiscale Diffusion (CMD) for Novel View Synthesis (NVS)** from a single image, ensuring photorealistic and geometrically consistent image sequences. They introduced the Thresholded Symmetric Epipolar Distance for evaluating geometric consistency. Their proposed model outperforms GeoGPT and LookOut models when tested on CLEVR, RealEstate10K, and Matterport3D datasets in terms of LPIPS and PSNR. For example, on RealEstate10K, it achieves an LPIPS of 0.333 and a PSNR of 15.51 compared to LookOut's 0.378 and 14.43 [54]. However, the model faces limitations, including performance drops in certain conditions and challenging scenarios. For instance, it may struggle with images that have complex

textures or dynamic elements, leading to less accurate geometric consistency and lower visual quality. Additionally, the model's robustness in diverse real-world environments is not fully tested, indicating a need for further refinement and evaluation to ensure reliable performance across a wider range of situations.

Yin et al. (2023) introduced **Controllable Light Enhancement (CLE)** Diffusion, a novel framework for low-light image enhancement that offers users dynamic control over brightness adjustments. Utilizing CMDs with an illumination embedding and integrating the Segment-Anything Model (SAM), CLE Diffusion allows precise, region-specific improvements. Their proposed approach outperformed existing models in terms of PSNR, SSIM, LPIPS, and LI-LPIPS on the LOL and MIT-Adobe FiveK datasets. For instance, on the LOL dataset, it achieved a PSNR of 25.51 and an SSIM of 0.89 [55]. However, the slow inference speeds hinder real-time application and usability in time-sensitive scenarios. Additionally, the model struggles to maintain high performance in environments with complex and varying lighting conditions, leading to inaccuracies and lower image quality in such settings.

Papantoniou et al. (2023) introduced **"Relightify,"** a method for 3D facial **Bidirectional Reflectance Distribution Function (BRDF)** reconstruction from a single image using DMs (Figure 7). Relightify is trained on a UV dataset of facial reflectance to understand facial features and lighting interactions. It fits a 3D model to an input image, unwraps the face into a UV texture, and uses the Diffusion Model to fill in occluded areas while keeping the original textures for realistic results. Relightify outperforms methods like CE, UV-GAN, and OSTeC, especially in handling different viewing angles, as measured by its higher PSNR and SSIM metrics [56].
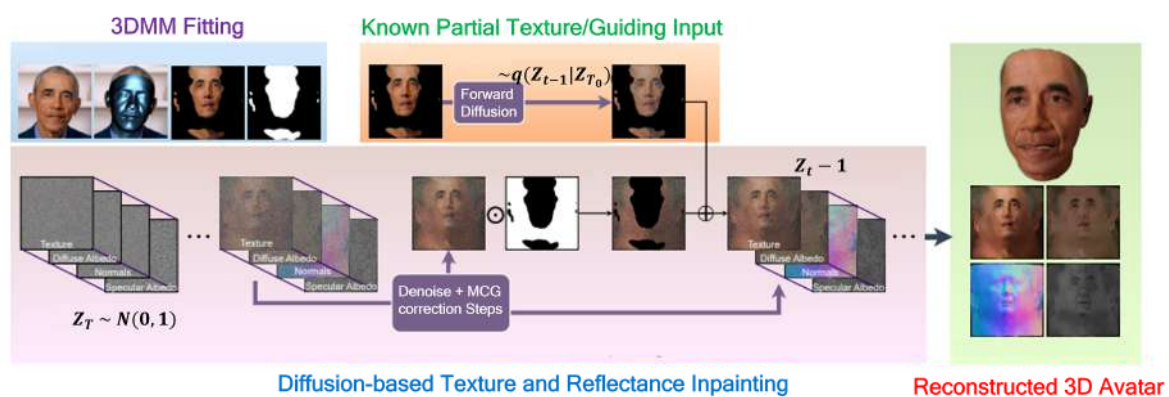

Figure 7. The Relightify method employs a latent Diffusion Model for inference, visualizing denoising in the original image space. It initiates with 3DMM fitting to generate a partial UV texture via image-to-UV rasterization. The process then uses random noise, guided by known texture, to complete missing pixels in the texture/reflectance diffusion model. Denoising steps (z_t to z_{t-1}, $t \in \{1, \ldots, T\}$) follow an inpainting approach similar to MCG: 1) Updating reflectance maps and unobserved texture pixels using reverse diffusion sampling and manifold constraints, and 2) Directly sampling known pixels from the input texture through forward diffusion ($\odot$ and $\oplus$ denote the Hadamard product and addition). Masking is applied solely to the texture, while reflectance maps (diffuse/specular albedo, normals) are predicted entirely from random noise. This technique produces high-quality rendering assets for realistic 3D avatar creation [56].

Kirch et al. (2023) presented **Red-Green-Blue Depth Fusion (RGB-D-Fusion)**, a multi-modal conditional diffusion denoising model that enhanced depth map resolution from low-resolution RGB images of humanoid subjects. Unlike Variational Autoencoders or GANs, RGB-D-Fusion employed diffusion denoising models in two stages: creating and refining low-resolution depth maps with RGB-D images, incorporating depth noise augmentation for robustness. It effectively generated detailed depth maps represented as point clouds when tested on a dataset of 25k samples [57]. However, it

required substantial resources for sampling and training and relied on known projection matrices, limiting its scalability and flexibility.

Mao et al. (2023) improved **multi-contrast MRI using the Discriminator Consistency Diffusion (DisC-Diff)** model, which stabilizes and leverages multi-contrast data. DisC-Diff outperforms existing techniques in terms of PSNR and SSIM metrics when tested on normal and pathological brain datasets [58]. Nonetheless, the study has limitations, such as the risk of mode collapse when processing multi-contrast MRI data, which can impact the reliability of the super-resolution process. Additionally, the proposed DMs may not adequately capture the complex interactions in multi-contrast MRI, limiting their effectiveness in clinical applications.

Table 4 summarizes some of the referenced literature that proposes different diffusion-based approaches for image-to-image transformation.

Table 4. Image-to-image transformation using different DMs. FID: Frechet Inception Distance, LPIPS: Learned Perceptual Image Patch Similarity, PSNR: Peak Signal-to-Noise Ratio, SSIM: Structural Similarity Index Measure, MSE: Mean Squared Error, RMSE: Root Mean Squared Error, CD: Chamfer Distance, EMD: Earth Mover's Distance, IoU: Intersection over Union, VLB: Visible Light Blocking. Best results are highlighted in bold.

Ref.	Algorithms	Applications	Dataset	Evaluations	Limitations
[54]	Autoregressive conditional Diffusion-based Models (ACDM)	NVS from a single image	RealEstate10K, MP3D, CLEVR	LPIPS: 0.33, PSNR: 15.51 on RealEstate10K; LPIPS: 0.50, PSNR: 14.83 on MP3D; FID: 26.76 on RealEstate10K; FID: 73.16 on MP3D	Requires complex geometric consistency and heavy computational resources for extrapolating views.
[55]	CLE Diffusion	Low light enhancement	LOL, MIT-Adobe FiveK	PSNR: 29.81, SSIM: **0.97** on MIT-Adobe FiveK; PSNR: 25.51, SSIM: 0.89, LPIPS: 0.16, LI-LPIPS: 0.18 on LOL	Slow inference speed and limited capability in handling complex lighting and blurry scenes.
[56]	Diffusion-based inpainting model for 3D facial BRDF reconstruction	Facial texture completion and reflectance reconstruction from a single image	MultiPIE	PSNR: 26.00, SSIM: 0.93 at 0° angle on MultiPIE; Sampling time: 17 sec	Limited by input image quality and potential underrepresentation of ethnic diversity in training data.
[52]	CMDs	HDR reconstruction from multi-exposed LDR images	Benchmark datasets for HDR imaging	PSNR-μ: 22.25, SSIM-μ: 0.84, LPIPS: 0.03 on Hu's dataset	Slow inference speed due to iterative denoising process.
[57]	RGB-D-Fusion diffusion probabilistic models	Depth map generation and super-resolution from monocular images	Custom dataset with $\approx$ 25,000 RGB-D images from 3D models of people	MSE: 1.48, IoU: **0.99**, VLB: 16.95 with UNet3+ model	High computational resources required for training and sampling.
[58]	Disentangled CMDs (DisC-Diff)	Multi-contrast MRI super-resolution	IXI dataset and clinical brain MRI dataset	PSNR: **37.77** dB, SSIM: **0.99** on 2× scale in clinical dataset	Requires accurate condition sampling for model precision.

6.2. Image Quality Enhancement and Processing

DMs have been used effectively in various image quality improvement tasks, as shown in Table 5. These tasks include improving document images, generating thermal facial images, and creating identity-preserving face images. In each case, DMs significantly boost image quality. They make images clearer, remove noise and watermarks, and create realistic images in different conditions, showing their versatility in image processing [59–62].

Yang et al. (2023) introduced **Document Diffusion (DocDiff)**, a diffusion-based framework for restoring degraded document images. This framework recovers low-frequency content using a Coarse Predictor and high-frequency details with a High-Frequency Residual Refinement (HRR) module. DocDiff's efficient architecture achieves SOTA results on benchmarks, improving readability and text edge sharpness with only 4.17 million parameters in the HRR module [59]. However, the study did not consider additional document improvement tasks such as document super-resolution or style transfer. The robustness of DocDiff in handling various types and levels of document degradation might be insufficient, requiring more diverse training data and improved network architectures. Furthermore, there is no user-centric evaluation to assess the impact on readability and user satisfaction.

Ordun et al. (2023) introduced **Visible-to-Thermal Facial GAN (VTF-GAN)**, a generative adversarial network that created high-resolution thermal facial images from visible spectrum inputs, addressing the lack of thermal sensors in common RGB cameras for telemedicine [60]. However, the study did not address the potential biases or ethical considerations that may arise when using generated thermal faces for applications such as telemedicine, which could be crucial in real-world implementations. There is a lack of analysis regarding the generalizability of the VTF-GAN model across different datasets or demographic groups, which could impact its applicability in diverse scenarios. Additionally, there is no comprehensive discussion on the interpretability of the generated images and how they align with the underlying physiological conditions they aim to represent.

Kansy et al. (2023) introduced the Identity **Denoising Diffusion Probabilistic Model (ID3PM)**, which can reverse-engineer face recognition models without needing full access (i.e., using a black-box method) to the model. ID3PM uses denoising diffusion to generate high-quality, identity-preserving facial images without needing an identity-specific loss. It effectively samples from the inverse distribution, producing diverse images with varying backgrounds, lighting, poses, and expressions [61]. Nonetheless, the method presented in this work generates images at a relatively low resolution of 64 x 64, which may limit the fine details captured. Additionally, the inference times for image generation are relatively long, and small artifacts in the output could affect the overall quality of the generated images.

Yu et al. (2023) presented **Free-form Deformation Model (FreeDoM)**, a versatile training-free conditional Diffusion Model that adapts to various conditions without condition-specific training (Figure 8). Unlike traditional models, FreeDoM uses pre-trained networks to create time-independent energy functions, reducing costs and improving transferability. Tested on different data domains, it outperforms training-required methods like Text- and Image-driven Generative Adversarial Network (TediGAN) in generating segmentation maps, sketches, and text-conditioned images, with better condition matching and FID scores [62]. While FreeDoM is designed to be training-free and adaptable to various conditions, it may struggle in situations where the conditions significantly differ from the capabilities of the pre-trained networks. This limitation could affect the model's effectiveness in diverse and complex scenarios.

Figure 8. Single-condition guided results based on FreeDoM models, where (a) are Unconditional DMs and (b) are Classifier-based DMs, generated output on the ImageNet dataset [62].

Table 5. Image Enhancement and Processing based on the referenced literature. FID: Frechet Inception Distance, LPIPS: Learned Perceptual Image Patch Similarity, PSNR: Peak Signal-to-Noise Ratio, SSIM: Structural Similarity Index Measure, MANIQA: Mean Opinion Score Quality Index, MUSIQ: Measurement Uncertainty Simulation Quality Index, DISTS: Deep Image Structure and Texture Similarity, MSE: Mean Squared Error. Best results are highlighted in bold.

Ref	Algorithms	Applications	Dataset	Evaluations	Limitations
[59]	DocDiff conditional Diffusion Model	Document image enhancement including deblurring, denoising, and watermark removal	Document Deblurring Dataset	MANIQA: 0.72, MUSIQ: 50.62, DISTS: 0.06, LPIPS: 0.03, PSNR: 23.28, SSIM: **0.95**	May lose high-frequency information, leading to distorted text edges. Relies on the quality of low-frequency content recovery by the Coarse Predictor module.
[60]	VTF-GAN	Thermal facial imagery generation for telemedicine	Eurecom and Devcom datasets	FID: 47.35, DBCNN: 34.34%, MSE: 0.88, SPEC: -1.1% for VTF-GAN with Fourier Transform-Guided (FFT-G)	Generation constrained to static environments; performance untested in dynamic, variable conditions affecting thermal emission.
[61]	ID3PM	Inversion of pre-trained face recognition models, generating identity-preserving face images	LFW, AgeDB-30, CFP-FP datasets	LFW: **99.20**%, AgeDB-30: 94.53%, CFP-FP: 96.13% with ID3PM using InsightFace embeddings	Generation quality may vary with the diversity of embeddings; control over the generation process might need fine-tuning for specific applications.
[62]	FreeDoM	Conditional image and latent code generation	Multiple datasets for segmentation maps, sketches, texts	Distance: 1696.1, FID: 53.08 for segmentation maps with FreeDoM	High sampling time cost; struggles with fine-grained control in large data domains; may produce poor results with conflicting conditions.

7. Healthcare and Medical Applications

DMs have made significant contributions to the field of healthcare and medical analysis by offering cutting-edge solutions for a variety of tasks. Models like PatchDDM, a memory-efficient patch-based DM, have been effectively utilized for applications such as tumor segmentation in medical imaging datasets like BraTS2020, showing their ability to generate precise three-dimensional segmentations [63]. Furthermore, DMs are renowned for their extensive mode coverage and the quality of samples they

generate. These models are employed in medical imaging to address challenges related to limited data availability, inconsistent data acquisition methods, and privacy issues. For example, the Med-DDPM, a DM-based approach, has demonstrated super stability and performance in comparison to GANs when it comes to generating high quality, realistic 3D medical images [63,64].

Chen et al. (2023) introduced the **Bernoulli Diffusion Model (BerDiff)** for medical image segmentation. BerDiff used Bernoulli noise instead of Gaussian noise, improving binary segmentation tasks essential in medical imaging. By sampling Bernoulli noise and intermediate latent variables, BerDiff generated diverse and accurate segmentation masks. This approach, tested on the LIDC-IDRI and BRATS 2021 datasets, outperformed SOTA methods in metrics such as Generalized Energy Distance (GED) and Dice score [65]. However, the proposed BerDiff model mainly focused on binary image segmentation, which may limit its application to more complex segmentation scenarios such as multi-class tasks. The study did not extensively discuss whether additional post-processing steps were needed for specific clinical tasks.

Shrivastava et al. (2023) presented **Nuclei-Aware Semantic Diffusion Model (NASDM)**, a framework for generating high-quality histopathological images using conditional Diffusion Modeling (Figure 9). NASDM creates realistic tissue samples from semantic instance masks of six nuclei types, aiding pathological analysis and addressing training data scarcity for nuclei segmentation. On a colon dataset, NASDM achieved an FID of 15.7 and an IS of 2.7, outperforming existing methods [66]. However, the proposed approaches required large amounts of annotated data for training Deep Learning models for nuclei segmentation, which can be expensive and time-consuming. Additionally, the current methods focused only on generating tissue patches conditioned on the semantic layouts of nuclei, which may have restricted the framework's scope to specific types of histopathological images.
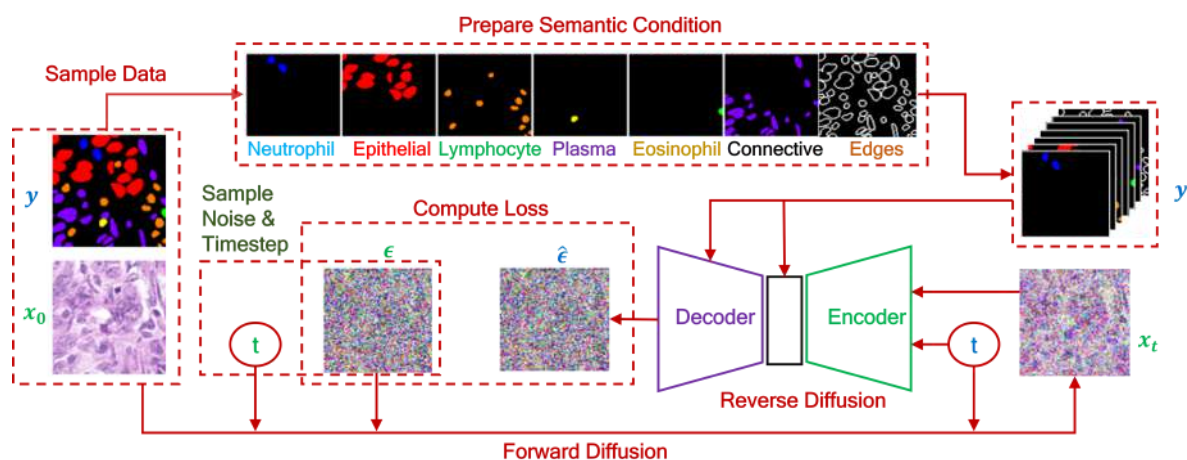

Figure 9. The NASDM training protocol initiates with an original image x_0 and its corresponding semantic mask y. It then generates a conditioning signal by enhancing the mask and incorporating an adjacent edge map. Subsequently, a timestep t is selected, and noise is applied to x_0 through forward diffusion, resulting in a perturbed input x_t. The denoising model then processes this corrupted image x_t, along with the timestep t and semantic condition y, to estimate $\hat{\epsilon}$, representing the model's prediction of the total noise introduced. The loss is then computed by comparing this estimate $\hat{\epsilon}$ with the actual noise ϵ applied during the forward diffusion process [66]

Wang et al. (2023) proposed a novel model, **Hierarchical Feature Conditional Diffusion (HiFi-Diff)**, a framework for MRI image super-resolution that adapts to varying inter-slice spacings in clinical settings. HiFi-Diff uses hierarchical feature extraction to iteratively convert Gaussian noise into high-resolution MR slices, achieving superior image quality. Tested on the HCP-1200 dataset, HiFi-Diff outperformed traditional methods in PSNR, SSIM, and Dice similarity coefficient across various scaling tasks (×4, ×5, ×6, ×7). For instance, in a ×4 task, it achieved a PSNR of 39.50 and an SSIM of 0.98 [67]. While the experimental results demonstrate the effectiveness of HiFi-Diff on the HCP-1200 dataset,

the study did not provide any insights regarding the model's performance compared to existing super-resolution methods on a wider range of MRI datasets with varying characteristics.

Li et al. (2023) introduced **Denoising Score-based Diffusion for Electrocardiogram (DeScoD-ECG)**, a conditional Score-based Diffusion Model for improving Electrocardiogram (ECG) signals, which are essential for diagnosing cardiovascular diseases but often suffer from noise. Unlike traditional Deep Learning methods, DeScoD-ECG iteratively reconstructs signals from Gaussian white noise using a Markov Chain, improving reconstruction quality with a multi-shot averaging strategy. Validated on the QT Database and MIT-BIH Noise Stress Test Database, DeScoD-ECG outperforms existing methods in metrics such as Sum of Squared Differences (SSD), Mean Absolute Deviation (MAD), Percent Root Mean Square Difference (PRD), and Cosine Similarity, showing over a 20% improvement [68]. However, the study did not address other types of noise interference that can affect ECG signals, such as muscle artifacts or electrode motion artifacts. While the study highlights the potential of the DeScoD-ECG model for biomedical applications, it does not discuss any specific real-world applications or case studies where the method has been successfully applied.

Table 6 summarizes some of the existing reference literature that considers DM-based approaches for developing realistic samples in medical imaging and healthcare.

Table 6. Health and medical applications using diffusion-based approaches as demonstrated in the existing literature. FID: Frechet Inception Distance, IS: Inception Score, PSNR: Peak Signal-to-Noise Ratio, SSIM: Structural Similarity Index Measure, GED: Generalized Energy Distance, HM-IoU: Harmonic Mean Intersection over Union, LPIS: Learned Perceptual Image Similarity. Best results are highlighted in bold.

Ref.	Algorithms	Applications	Dataset	Evaluations	Limitations
[65]	BerDiff: Conditional Bernoulli Diffusion for Medical Image Segmentation	Advanced medical image segmentation using Bernoulli diffusion to produce accurate and diverse segmentation masks	LIDC-IDRI, BRATS 2021	Achieves state-of-the-art performance with metrics on LIDC-IDRI - GED: 0.24, HM-IoU: 0.60, and on BRATS 2021 - Dice: **89.7**.	Focuses only on binary segmentation and requires significant time for iterative sampling.
[66]	NASDM: Nuclei-Aware Semantic Tissue Generation Framework	Generative modeling of histopathological images conditioned on semantic instance masks	Colon dataset	FID: 15.7, IS: 2.7, indicating high-quality and semantically accurate synthetic image generation.	Further development is required for varied histopathological settings and end-to-end tissue generation that includes mask synthesis.
[67]	Hierarchical Feature Conditional Diffusion (HiFi-Diff)	MR image super-resolution with arbitrary reduction of inter-slice spacing	HCP-1200 dataset	PSNR: 39.50±2.29, SSIM: **0.99** for ×4 SR task	Slow sampling speed, suggesting potential improvements through faster algorithms or knowledge distillation.

8. Applications of Diffusion Models in Other Fields

DMs are adopted in various domains beyond image analysis and are effectively used for time series forecasting, imputation, and generation, demonstrating their versatility in handling sequential data. Additionally, DMs have been adapted for predicting chaotic dynamical systems, offering uncertainty quantification and the ability to represent outliers and extreme events effectively. Fur-

thermore, recent advancements have extended DMs to Riemannian manifolds, enabling applications in constrained conformational modeling of protein backbones and robotic arms, highlighting their relevance in scientific domains as well. The evolution of DMs beyond image analysis underscores their adaptability and effectiveness across a wide range of fields [69–71].

Li et al. (2023) developed the **Diffusion Classifier**, a novel method using large-scale text-to-image Diffusion Models for zero-shot classification (Figure 10). This approach leverages Diffusion Models' density estimates to classify images without additional training, outperforming existing methods. The Diffusion Classifier performs exceptionally well in benchmarks and multimodal compositional reasoning, showing notable improvements in zero-shot reasoning tasks. It also demonstrated robustness against distribution shifts when tested with ImageNet [72]. While the study focuses on Stable Diffusion, it does not explore the potential challenges or limitations that may arise when applying this approach to other types of classification problems beyond image data.
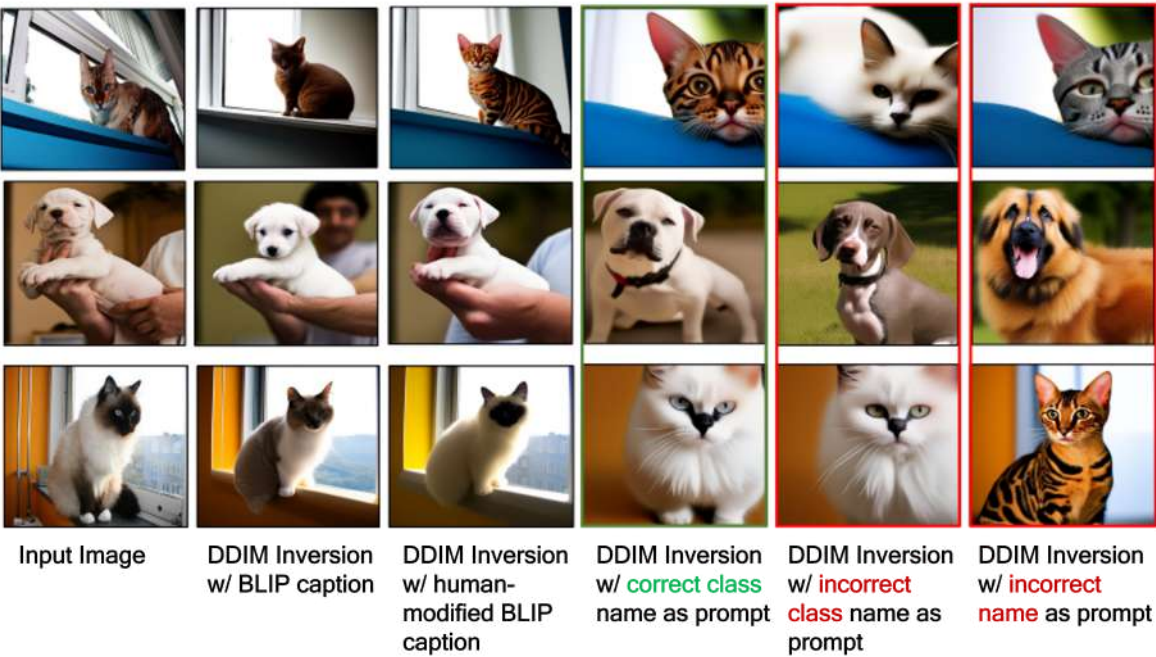

Figure 10. Various texts and captions (BLIP, Human-modified BLIP, correct class names, incorrect class names) in zero-shot classification using text-based DMs is examined. The input image is inverted with the caption and reconstructed using deterministic DDIM sampling. Human-modified BLIP captions align best with the input image. Images reconstructed with correct class names (col. 4) match better than those with incorrect class names (col. 5 and 6). In Row 3 (col. 4 and 5), the base Stable DMs fails to distinguish between Birman and Ragdoll breeds, causing classifier failure. BLIP: Bootstrapped Language-Image Pretraining, DDIM: Denoising Diffusion Implicit Models [72]

Zhuang et al. (2023) explored DMs for semantic image synthesis, focusing on abdominal CT images. They compared three models—Conditional DDPM, Mask-guided DDPM, and Edge-guided DDPM—against SOTA GAN-based approaches. By using semantic masks to guide synthesis, the proposed approaches surpassed GANs in terms of FID, PSNR, SSIM, and Dice Score, generating higher-quality and more clinically accurate images [73]. Despite their advantages, the proposed DMs faced significant challenges due to high computational costs and long processing times.

Jiang et al. (2023) addressed data protection against unauthorized uses such as adversarial attacks (Figure 11). The study proposed a novel purification process called **Joint-Conditional Diffusion Purification (JCDP)**, which projects Uncertain Examples (UEs) onto the manifold of Learnable Unauthorized Examples (LEs). By leveraging DMs and image generation approaches, the study maps from UEs to their corresponding clean samples. However, the study did not consider whether it might perform

well in situations where the adversarial attack might evolve over time. Apart from this, they did not consider the generalizability of their proposed methods in terms of various Machine Learning techniques as well [74].
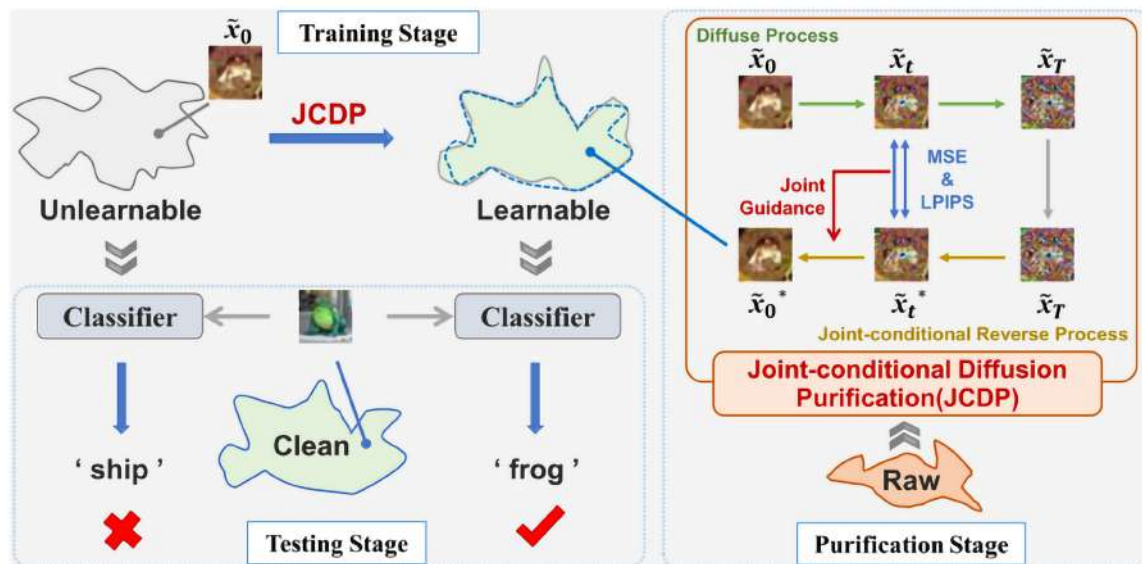

Figure 11. Joint-conditional diffusion purification (JCDP) demonstrates the concept of learnable examples. When applied to datasets, non-generalizable or unlearnable data points fail to achieve effective generalization, consequently impacting the quality and reliability of the samples employed in training classification models [74].

Hsu et al. (2023) proposed **Score Dynamics (SD)**, a framework that uses Graph Neural Networks to accelerate Molecular Dynamics (MD) simulations. SD uses evolution operators for large timestep transitions, which greatly increase simulation speed. It simulates molecular dynamics with 10 picosecond timesteps, showing high accuracy in studies of alanine dipeptide and short alkanes in aqueous solutions. SD outperforms traditional MD in speed by up to two orders of magnitude [51]. Despite these promising results, challenges include extending SD to larger molecules, refining assumptions, and improving the accuracy and efficiency of the score model.

Wang et al. (2023) introduced **Atmospheric Turbulence Variational Diffusion (AT-VarDiff)**, a deep conditional Diffusion Model designed to correct atmospheric turbulence in images using a variational inference framework. This approach addresses geometric distortion and spatially variant blur. When tested on a synthetic dataset, AT-VarDiff achieved an LPIPS of 0.1094, an FID of 32.69, and a Naturalness Image Quality Evaluator (NIQE) score of 6.46, outperforming existing models [75].

Sartor et al. (2023) proposed **Material Fusion (MatFusion)**, a method for estimating Spatially Varying Bidirectional Reflectance Distribution Functions (SVBRDF) from photographs using Diffusion Models. MatFusion is trained on 312,165 synthetic material samples and refines a conditional model to estimate material properties, generating multiple SVBRDF estimates per photo for user selection. It achieves high accuracy with an LPIPS of 0.2056 and RMSE values of 0.041 for diffuse, 0.066 for specular, 0.126 for roughness, and 0.052 for normal maps [76]. However, its performance depends on the quality of the photos and the user's selection, which may introduce variability in the results. Additionally, the method lacks automatic selection metrics and could benefit from optimal regularization to improve consistency.

Wei et al. (2023) proposed **Building Diffusion (BuilDiff)**, an innovative method for generating 3D building point clouds from single general-view images. BuilDiff uses two CMDs and a regularization strategy to synthesize building roofs while maintaining structural integrity. It extracts image embeddings through a Convolutional Neural Network-based auto-encoder and utilizes a conditional denoising diffusion network and a point cloud upsampler. Tested on BuildingNet-SVI and Build-

ingNL3D datasets, BuildDiff outperforms existing methods [77]. Despite its superior performance, BuildDiff heavily relies on the quality and variety of training data, limiting its generalizability to unseen building styles. Additionally, it demands significant computational resources for both training and inference. Furthermore, the model struggles to capture fine-grained details of building structures due to the resolution limits of the point clouds used.

Niu et al. (2024) developed the **Accelerated Conditional Diffusion Model for Image Super-Resolution (ACDMSR)**. ACDMSR used pre-super-resolved images as conditional inputs, improving efficiency and quality over traditional Diffusion Models. It adapted Diffusion Models for super-resolution through a faster, iterative denoising process. Testing on benchmark datasets like Set5 and Urban100 showed ACDMSR outperformed existing methods [78]. However, reliance on initial pre-super-resolution may have limited its flexibility in diverse applications.

Table 7 summarizes some of the referenced literature that introduces diffusion-based approaches in various fields.

Table 7. Applications of DMs in other fields based on the referenced literature. FID: Frechet Inception Distance, PSNR: Peak Signal-to-Noise Ratio, SSIM: Structural Similarity Index Measure, DSC: Dice Similarity Coefficient, MAE: Mean Absolute Error, MOS: Mean Opinion Score, SMOS: Style Similarity MOS. Best results are highlighted in bold.

Ref.	Algorithms	Applications	Dataset	Evaluations	Limitations
[72]	Diffusion Classifier using text-to-image Diffusion Models	Zero-shot classification using generative models	Standard image classification benchmarks (e.g., ImageNet, CIFAR10)	Zero-shot classification accuracy on ImageNet using Diffusion Classifier: **58.9%**	Performance gap in zero-shot recognition compared to SOTA discriminative models
[73]	CMDs for semantic image synthesis	Semantic synthesis for abdominal CT, used in data augmentation	Not specified	FID: 10.32, PSNR: 16.14, SSIM: 0.64, DSC: **95.6%** for mask-guided DDPM at 100k iterations	High sampling time and computational cost
[74]	Learnable Unauthorized Examples (LEs) using joint-CMDs	Countermeasure to unlearnable examples in Machine Learning models	CIFAR-10, CIFAR-100, SVHN	Test accuracy on CIFAR-10 using LE: **94.0%**, CIFAR-100: 67.8%, SVHN: **94.9%**	Limited by distribution mismatches
[79]	Diffusion-based Non-Intrusive Load Monitoring (DiffNILM) Diffusion Probabilistic Model	Non-intrusive Load Monitoring (NILM) for appliance power consumption pattern disaggregation	REDD and UK-DALE datasets	F1-Score: **0.79** for refrigerator on REDD, MAE: **4.54** for microwave on UKDALE	Generation of power waveforms not always sufficiently smooth; computational efficiency not optimized
[80]	Noise-Robust Expressive Text-to-Speech model (NoreSpeech)	Expressive TTS in noise environments	Not specified	MOS: **4.11**, SMOS: **4.14** for NoreSpeech with T-SSL in noisy conditions	Dependent on quality of style teacher model
[81]	Diffusion-based data augmentation for nuclei segmentation	Nuclei segmentation in histopathology image analysis	MoNuSeg and Kumar datasets	Dice score: **0.83**, AJI: **0.68** with 100% augmented data on MoNuSeg dataset	Dependent on the quality of synthetic data

Table 7. *Cont.*

Ref.	Algorithms	Applications	Dataset	Evaluations	Limitations
[75]	AT-VarDiff	Atmospheric turbulence (AT) correction	Comprehensive synthetic atmospheric turbulence dataset	LPIPS: 0.11, FID: 32.69, NIQE: 6.46	May not generalize well to real-life atmospheric turbulence images.
[76]	MatFusion Diffusion Models (unconditional and conditional)	SVBRDF estimation from photographs	Large set of 312,165 synthetic spatially varying material exemplars	RMSE on property maps: 0.04, LPIPS error on renders: 0.21	Limited by the variation in lighting conditions.
[77]	Point cloud Diffusion Models with image conditioning schemes	3D building generation from images	BuildingNet-SVI and BuildingNL3D datasets	CD: 3.14, EMD: 10.84, F1 score: 21.41 on BuildingNet-SVI	Constrained by specific image viewing angles.
[78]	ACDMSR: Accelerated Conditional Diffusion Model for Image Super-Resolution	Enhancing super-resolution using Diffusion Models conditioned on pre-super-resolved images	DIV2K, Set5, Set14, Urban100, BSD100, Manga109	LPIS: 0.08, PSNR: 25.95, SSIM: 0.67	Challenges remain in processing images with more complex degradation patterns.

9. Discussion

9.1. Ensuring the Authenticity of Synthesized Media

DMs play an important role in improving media quality and generating high-fidelity samples. Techniques such as SAG advance image generation by concentrating on significant areas and minimizing artifacts [41]. While SAG enhances image quality by leveraging self-attention maps, it still faces challenges in real-time applications due to high computational demands [40]. On the other hand, Learnable State-Estimator-based models offer computational efficiency but require extensive domain-specific adaptations for broader applications [41].

Contradictions arise when certain methods show better performance in specific cases but fall short in others. For example, while the state-estimator-based model performs well on tasks like inpainting and deblurring, it may not work in real-time as effectively as SAG. This discrepancy highlights the need for a balanced approach that combines the strengths of various techniques.

To address these challenges, integrating different types of DMs, such as Stepwise Error for Diffusion-generated Image Detection (SeDID) and Unlearnable Diffusion Perturbation (EUDP), could be effective [82]. Additionally, strategies like sampling space truncation and robustness penalties can also be helpful in ensuring the authenticity of media quality [83,84].

9.2. Overcoming Challenges in Synthesizing High-Quality Images and Audio

Diffusion-based models play a crucial role in synthesizing high-quality images and audio by refining noise into structured data. These models utilize DDPMs, involving a forward process that adds Gaussian noise to the data and a reverse process that removes this noise to reconstruct the original signal. For instance, text-to-audio synthesis methods like CLIPSonic use CDMs to translate text embeddings into audio. This method shows superior performance but faces limitations due to the quality of pretrained models, distribution mismatches, and training complexity [44]. Similarly, SDG improves image synthesis by adding fine-grained control to pretrained models. Its effectiveness depends on the precision of these models and the accuracy of guidance signals, raising concerns about potential misuse [46].

Other approaches, such as DiffDreamer, use CMDs (CMDs) for scene extrapolation. They often show better quality and consistency than GAN-based methods but struggle with real-time synthesis and variety in generated content [47]. Interactive tools like Diffusion-based Art Generation (DiffusArt) use Conditional Diffusion Probabilistic Models for line art colorization, which produces high-quality images but requires precise user input and faces computational inefficiencies [48]. SketchFusion focuses on sketch-guided image editing, maintains the integrity of sketches, and achieves high-performance metrics but is limited to binary sketches [85]. Semantic-Conditional Diffusion Networks improve image captioning by enhancing visual-language alignment and outperform traditional models, but they face high computational demands and complexity [50].

To overcome these challenges, future research should improve the computational efficiency of DMs, improve the quality and robustness of pretrained models, and develop adaptive techniques to handle distribution mismatches. Additionally, integrating ethical guidelines and protective measures can help reduce the risks of misuse, ensuring that these advanced models are applied responsibly.

9.3. Optimizing DMs to Reduce Artifacts and Improve Image Quality

Optimizing DMs to minimize artifacts and improve image quality is crucial for their broader application. These models, which refine noise into structured data, can introduce artifacts that compromise image fidelity [40,46]. To optimize DMs for reducing artifacts and enhancing image quality, various techniques have been proposed. These include using Deep Interpretable Convolutional Dictionary Networks (DICDNet) for metal artifact reduction in CT images, and automatic segmentation of 3D objects to minimize supports and cuts for 3D printing [86]. For fetal MRI, efforts focus on improving image quality by optimizing acquisition speed, spatial resolution, and signal-to-noise ratio while considering artifacts from motion, banding, and aliasing [87]. Challenges persist, such as balancing the effects of supports and cuts in 3D printing segmentation, trade-offs between scan parameters in fetal MRI optimization, and addressing artifacts from beam hardening in X-ray imaging [86,88]. These limitations highlight the complexity of optimizing Diffusion Models to reduce artifacts and improve image quality across different imaging methods.

Another strategy is SDG integrates fine-grained control into pretrained models via image-text matching score gradients, enhancing image synthesis quality without retraining models [46]. The success of SDG, though, depends on the precision of pretrained models and guidance signal accuracy.

Advanced noise estimation techniques further improve DMs. Pixel-level autoregressive processes, like those used in Image Transformer models, significantly reduce noise and artifacts, which improves image fidelity and consistency across datasets [26]. Apart from this, dynamic thresholding and adaptive noise schedules can also fine-tune denoising steps to improve image quality by handling complex structures and textures more effectively.

To sum up, optimizing DMs to reduce artifacts and improve image quality requires a multi-faceted approach. Incorporating, semantic guidance, and advanced noise estimation techniques, along with optimizing the diffusion process, can significantly increase model performance. Future research should focus on improving computational efficiency, developing robust conditioning strategies, and integrating adaptive techniques to further reduce artifacts and enhance image quality.

9.4. Addressing Computational Efficiency and Scalability Issues in DMs

Addressing computational efficiency and scalability in DMs is crucial for practical application and widespread adoption. Despite their ability in generating high-fidelity images and audio, DMs often struggle with high computational demands and scalability, particularly with large datasets or real-time applications.

To optimize computational efficiency, it is essential to use more efficient network architectures. For instance, integrating guidance into pretrained models via image-text matching score gradients eliminate the need for extensive retraining, thereby improves computational efficiency [46]. Additionally, dynamic adaptation of the diffusion process is an effective strategy. Adjusting the diffusion

process within a latent space can achieve significant computational efficiency [41], which allows the model to focus resources on the most relevant datasets. Furthermore, parallelization and hardware acceleration, such as using Graphics Processing Unit (GPUs) and ensor Processing Unit (TPUs), can address scalability issues. Therefore, distributing the computational load across multiple processors can significantly speed up the training and inference processes of DMs.

Moreover, multi-shot averaging strategies can improve the quality of generated images while maintaining efficiency [29]. Averaging multiple generated images reduces noise and improves overall quality without significantly increasing computational costs. In summary, addressing computational efficiency and scalability in DMs involves optimizing network architectures, dynamically adapting the diffusion process, leveraging hardware acceleration, refining algorithms, and deploying strategies like multi-shot averaging. Future research should explore these approaches to develop more efficient and scalable DMs for a broader range of tasks and datasets.

9.5. Improving DMs for Accurate and Reliable Medical Imaging and Diagnostics

DMs in medical imaging and diagnostics show significant promise due to their ability to create high-quality images. Critical areas of focus include reducing the model size to ensure efficient deployment, developing better training approaches for realistic samples, and leveraging advanced techniques to handle data augmentation and anonymization when considering DM-based approaches.

Reducing the model size of DMs is crucial for practical applications in medical imaging, where computational resources are often limited. Therefore, techniques such as model pruning, knowledge distillation, and post-training quantization are commonly used to achieve this goal. For instance, model pruning involves removing redundant parameters from the model, which decreases the model size without significantly affecting performance. Similarly, knowledge distillation transfers the knowledge from a large model (teacher) to a smaller model (student), which helps in maintaining performance while reducing the model size [89]. Additionally, post-training quantization converts the model parameters from floating-point to lower-bit representations, which reduces model size and speeds up inference without requiring retraining [90].

In medical imaging, data augmentation and anonymization are critical for creating robust ML models and protecting patient privacy. Semantic-based DMs offer promising solutions for these challenges. For data augmentation, these models can generate diverse and realistic medical images by conditioning on specific semantic features, which enriches the training dataset and improves model generalization [91]. For anonymization, semantic-based approaches can mask identifiable features in medical images while preserving clinically relevant information, ensuring that patient privacy is maintained without compromising the utility of the data [92,93].

Incorporating domain-specific knowledge into training processes can further improve DMs. For instance, integrating medical expertise and anatomical priors helps models better understand the structure and context of medical images, leading to more accurate diagnostics [94]. Moreover, collaborations between artificial intelligence researchers and medical professionals can facilitate the integration of such knowledge, enhancing the overall effectiveness of the models.

9.6. Expanding the Applicability and Effectiveness of DMs in Diverse Fields

DMs are gaining popularity across various fields beyond their initial use in image analysis. These models have shown effectiveness in time series forecasting, imputation, and generation, which demonstrates their versatility in handling sequential data. Additionally, DMs have been adapted for predicting chaotic dynamical systems, which offers uncertainty quantification and the ability to represent outliers and extreme events effectively [70]. Moreover, advancements have extended DMs to Riemannian manifolds, enabling applications in constrained conformational modeling of protein backbones and robotic arms, which underscores their relevance in scientific domains [71]. Despite these advances, one significant concern is model collapse, where the model fails to generate diverse outputs over time, leading to reduced effectiveness in applications requiring high variability.

This is particularly relevant in fields like finance and time series forecasting, where the accuracy of predictions is crucial. While DMs offer robust solutions for these tasks, their feasibility compared to existing, more computationally efficient approaches like autoregressive models or Long Short-Term Memory networks (LSTMs) remains questionable. It is essential for researchers to exercise caution when considering DM-based approaches for financial and time series data, as these models may not always offer the most practical or efficient solutions [95]. DMs have been applied to generate synthetic data, but they often underperform compared to techniques such as GANs, oversampling, and Synthetic Minority Over-sampling Technique (SMOTE). These traditional methods are often easier to deploy and require less computational power, making them more accessible for many applications. For instance, GANs have been widely used in generating realistic images and synthetic data for training ML models, providing a simpler alternative to DMs. Similarly, techniques like SMOTE are effective for addressing class imbalances in datasets and can be implemented with relative ease compared to the complex training processes required for DMs [96].

9.7. Mitigating Ethical Considerations and Potential Risks Associated with the Use of DMs

While DMs are surpassing GANs in generating realistic images, audio, and other types of data, they also raise questions regarding ethical use and practical concerns. For instance, one of the primary ethical concerns is the potential for misuse in creating deepfakes and synthetic media that can spread misinformation or violate privacy. To reduce this risk, it is essential to develop robust detection mechanisms that can differentiate between real and synthetic media. Implementing adversarial training techniques can improve the ability of models to identify and flag manipulated content [97].

Another major risk is bias in the generated outputs, which can perpetuate or even worsen existing social biases if not properly managed. Ensuring diversity in training data and incorporating fairness-aware algorithms can help reduce bias in Diffusion Models. Regular audits and updates to the models can also ensure they remain unbiased and fair in their outputs [98].

Transparency and explainability of DMs are also critical to address ethical concerns. Users need to understand how these models make decisions and generate outputs. Developing methods to explain the black box of DMs might help make their operations more transparent and accountable. Techniques such as interpretability frameworks and model-agnostic tools like Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) can provide insights into how models produce their results [99].

Data privacy is another major concern, especially when DMs are applied to sensitive areas such as healthcare systems and clinical diagnosis. Ensuring that models comply with data protection regulations, such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA), is essential. Techniques like differential privacy can protect individual data while still allowing models to learn effectively from large datasets [100].

Collaborative governance and the establishment of ethical guidelines for the development and deployment of DMs are also necessary. Engaging stakeholders from diverse fields, including ethicists, policymakers, and technologists, can help create comprehensive frameworks that address the ethical implications of these technologies. Such collaboration can lead to the development of standards and best practices that promote the responsible use of DMs [101].

10. Conclusion

Diffusion Models (DMs) promise to transform many fields by solving challenges in data generation and processing through the creation of realistic samples. Therefore, addressing current limitations and building on the strengths of DMs will enable wider adoption and more impactful applications across various domains in the future. Our findings show that DMs' ability to generate high-quality synthetic data improves performance in applications such as text-to-image generation, where models like Diffusion Transformers (DT) for stable diffusion demonstrate advancements in data privacy [35]. In cyber-physical system security, the Temporal and Feature TFDPM helps detect attacks by correlating

channel data using Graph Attention Networks [36]. Moreover, for cloud service anomaly detection, models like Maat combine metric forecasting with anomaly detection to achieve higher accuracy [37].

In image processing, diffusion-based techniques have shown superior performance in tasks like image deblurring and super-resolution. For example, stochastic image deblurring using DMs achieves high perceptual image patch similarity and structural similarity index measures [31]. Additionally, accelerated CMDs for applications like MRI reconstruction show potential by improving image quality [32]. Furthermore, the selective diffusion distillation approach balances image fidelity and editability, making it suitable for various image manipulation tasks [33].

However, while DMs generate realistic data, they also raise ethical concerns. One primary issue is the potential misuse in creating deepfakes and synthetic media that can spread misinformation or violate privacy. To mitigate this risk, robust detection mechanisms are essential. Ensuring models remain unbiased is also crucial, which can be achieved by incorporating fairness-aware algorithms and diverse training data. Furthermore, transparency and explainability of DMs are critical. Techniques like LIME and SHAP provide insights into how models generate their results. Apart from this, ensuring data compliance with regulations like the GDPR and the Health HIPAA is also necessary [98–100].

High computational demands and the need for better sampling or network architectures are recurring issues in DMs. Models often require extensive hyperparameter tuning and may struggle with discrete signal modeling or generalizing to different contexts [36,37]. Additionally, the reliance on correct timestep selection for semantic guidance in some models can limit flexibility [33]. Slow inference speeds and high resource requirements hinder real-time deployment and scalability [31,32].

Therefore, future research should address these limitations by developing more efficient algorithms and leveraging advancements in computational technologies. Exploring semi-supervised or unsupervised learning approaches, along with transfer learning from pre-trained models, can help overcome data scarcity challenges. Improving the robustness of DMs to noise and their ability to handle different data types is essential. Moreover, continued interdisciplinary collaboration and clear ethical guidelines will be vital for the responsible and effective use of DMs across diverse fields.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ACDM	Autoregressive Cascade Multiscale Diffusion
ACDMSR	Accelerated Conditional Diffusion Model for Image Super-Resolution
AIoT	Artificial Intelligence of Things
BerDiff	Bernoulli Diffusion Model
BLIP	Bootstrapped Language-Image Pretraining
BuilDiff	Building Diffusion
CDDM	Conditional Denoising Diffusion Model
CDMs	Classifier-guided Diffusion Models
CIDEr	Consensus-based Image Description Evaluation
CLE	Controllable Light Enhancement Diffusion
CLIP	Contrastive Language-Image Pre-training
CLIPSonic	Controlled Language-Image Pretraining Sonic
CMD	Conditional Diffusion Models
DDIM	Denoising Diffusion Implicit Models
DDPMs	Denoising Diffusion Probabilistic Models
DeScoD-ECG	Denoising Score-based Diffusion for Electrocardiogram
DiffWave	Diffusion Waveform
DiffDreamer	Diffusion Dreamer
DiffLL	Diffusion Model for Low-Light
DMs	Diffusion Models
DMSEtext	Diffusion Model for Speech Enhancement text

DisC-Diff	Discriminator Consistency Diffusion
DSBID	Diffusion-based Stochastic Blind Image Deblurring
DSC	Dice Similarity Coefficient
DICDNet	Deep Interpretable Convolutional Dictionary Networks
EquiDiff	Equivariant Diffusion
FID	Frechet Inception Distance
GED	Generalized Energy Distance
GNNs	Graph Neural Networks
HiFi-Diff	Hierarchical Feature Conditional Diffusion
HQS	Hybrid Quality Score
ID3PM	Identity Denoising Diffusion Probabilistic Model
IS	Inception Score
KID	Kernel Inception Distance
LEs	Learnable Unauthorized Examples
LDMs	Latent Diffusion Models
LPIPS	Learned Perceptual Image Patch Similarity
MAAT	Metric Anomaly Anticipation
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MatFusion	Material Fusion
MOS	Mean Opinion Score
MPJPE	Mean Per-Joint Position Error
NILM	Non-Intrusive Load Monitoring
NASDM	Nuclei-Aware Semantic Diffusion Model
NIQE	Naturalness Image Quality Evaluator
OMOMO	Object Motion Guided Human Motion Synthesis
PatchDDM	Patch-based Diffusion Denoising Model
PRD	Percent Root Mean Square Difference
PSNR	Peak Signal-to-Noise Ratio
RGB-D-Fusion	Red-Green-Blue Depth Fusion
RNNs	Recurrent Neural Networks
RMSE	Root Mean Square Error
SAG	Self-Attention Guidance
SBDMs	Score-Based Diffusion Models
SDEs	Stochastic Differential Equations
SDG	Semantic Diffusion Guidance
SegDiff	Segmentation Diffusion
SketchFFusion	Sketch-Driven Fusion
SMOS	Style Similarity MOS
SSIM	Structural Similarity Index Measure
SDEs	Stochastic Differential Equations
TFDPM	Temporal and Feature Pattern-based Diffusion Probabilistic Model
VDMs	Variational Diffusion Models
VTF-GAN	Visible-to-Thermal Facial GAN

References

1. Sohl-Dickstein, J.; Weiss, E.A.; Maheswaranathan, N.; Ganguli, S. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv preprint arXiv:1503.03585* **2015**.
2. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems, 2020.
3. Saharia, C.; Ho, J.; Chan, W.; Fleet, D.J.; Norouzi, M.; Salimans, T. Image Super-Resolution via Iterative Refinement. *arXiv preprint arXiv:2104.07636* **2021**.

4. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10684–10695.

5. Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; Sutskever, I. Zero-Shot Text-to-Image Generation. *arXiv preprint arXiv:2102.12092* **2021**.

6. Austin, J.; Odena, A.; Nijkamp, E.; Ballas, N.; Goodfellow, I. Structured Denoising Diffusion Models in Discrete State-Spaces. Advances in Neural Information Processing Systems, 2021.

7. Kong, Z.; Ping, W.; Huang, J.; Zhao, K.; Catanzaro, B. Diffwave: A versatile diffusion model for audio synthesis. *arXiv preprint arXiv:2009.09761* **2020**.

8. Hoogeboom, E.; Cohen, T.; Tomczak, J.M. Equivariant Diffusion Models for Molecule Generation. International Conference on Machine Learning. PMLR, 2022, pp. 8816–8831.

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

10. Song, Y.; Ermon, S. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems* **2019**, *32*.

11. Anderson, B.D. Reverse-time diffusion equation models. *Stochastic Processes and their Applications* **1982**, *12*, 313–326.

12. Vincent, P. A connection between score matching and denoising autoencoders. *Neural computation* **2011**, *23*, 1661–1674.

13. Wang, W.; Bao, J.; Zhou, W.; Chen, D.; Chen, D.; Yuan, L.; Li, H. Semantic image synthesis via diffusion models. *arXiv preprint arXiv:2207.00050* **2022**.

14. Gong, S.; Li, M.; Feng, J.; Wu, Z.; Kong, L. Diffuseq: Sequence to sequence text generation with diffusion models. *arXiv preprint arXiv:2210.08933* **2022**.

15. Kazerouni, A.; Aghdam, E.K.; Heidari, M.; Azad, R.; Fayyaz, M.; Hacihaliloglu, I.; Merhof, D. Diffusion models in medical imaging: A comprehensive survey. *Medical Image Analysis* **2023**, p. 102846.

16. Krizhevsky, A.; Hinton, G.; others. Learning multiple layers of features from tiny images. Technical report, University of Toronto, Toronto, ON, Canada, 2009.

17. Yu, F.; Seff, A.; Zhang, Y.; Song, S.; Funkhouser, T.; Xia, J. LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop. *arXiv preprint arXiv:1506.03365* **2015**.

18. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep Learning Face Attributes in the Wild. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* **2015**.

19. Song, Y.; Ermon, S. Improved techniques for training score-based generative models. *Advances in neural information processing systems* **2020**, *33*, 12438–12448.

20. Dhariwal, P.; Nichol, A. Diffusion Models Beat GANs on Image Synthesis. *Advances in Neural Information Processing Systems* **2021**.

21. Kingma, D.P.; Dhariwal, P.; Ho, J.; Salimans, T.; Chen, X.; Abbeel, P. Variational Diffusion Models. *arXiv preprint arXiv:2107.00630* **2021**.

22. Nichol, A.; Dhariwal, P. Improved Denoising Diffusion Probabilistic Models. *arXiv preprint arXiv:2102.09672* **2021**.

23. Kong, Z.; Ping, W.; Huang, J.; Zhao, K.; Catanzaro, B. DiffWave: A Versatile Diffusion Model for Audio Synthesis. *arXiv preprint arXiv:2009.09761* **2021**.

24. Amit, R.; Balaji, Y. SegDiff: Image Segmentation with Diffusion Models. *arXiv preprint arXiv:2106.02477* **2021**.

25. Nichol, A.; Dhariwal, P.; Ramesh, A.; Shyam, P.; Mishkin, P.; McGrew, B.; Sastry, G.; Askell, A.; Chen, P.; Mishkin, M.; Chug. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. *arXiv preprint arXiv:2112.10741* **2021**.

26. Saharia, e.a. Image Transformers with Autoregressive Models for High-Fidelity Image Synthesis. *Journal of Advanced Image Processing* **2022**.

27. Ho, J.; Jain, A.; Abbeel, P. Cascaded Diffusion Models for High-Fidelity Image Generation. *arXiv preprint arXiv:2106.15282* **2022**.

28. Ho, J.; Chan, W.; Salimans, T.; Gritsenko, A.; Kumar, K.C.; Isola, P. Video Diffusion Models. *arXiv preprint arXiv:2204.03458* **2022**.

29. Li, e.a. Optimizing Diffusion Models for Image Synthesis. *Journal of Computational Imaging* **2023**.

30. Mao, J.; Wang, X.; Aizawa, K. Guided image synthesis via initial image editing in diffusion model. Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 5321–5329.

31. Whang, J.; Delbracio, M.; Talebi, H.; Saharia, C.; Dimakis, A.G.; Milanfar, P. Deblurring via stochastic refinement. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 16293–16303.

32. Chung, H.; Sim, B.; Ye, J.C. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 12413–12422.

33. Wang, L.; Yang, S.; Liu, S.; Chen, Y.c. Not All Steps are Created Equal: Selective Diffusion Distillation for Image Manipulation. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 7472–7481.

34. Li, J.; Wu, J.; Liu, C.K. Object motion guided human motion synthesis. *ACM Transactions on Graphics (TOG)* **2023**, *42*, 1–11.

35. Ni, Z.; Wei, L.; Li, J.; Tang, S.; Zhuang, Y.; Tian, Q. Degeneration-tuning: Using scrambled grid shield unwanted concepts from stable diffusion. Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 8900–8909.

36. Yan, T.; Zhou, T.; Zhan, Y.; Xia, Y. TFDPM: Attack detection for cyber–physical systems with diffusion probabilistic models. *Knowledge-Based Systems* **2022**, *255*, 109743.

37. Lee, C.; Yang, T.; Chen, Z.; Su, Y.; Lyu, M.R. Maat: Performance Metric Anomaly Anticipation for Cloud Services with Conditional Diffusion. 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, 2023, pp. 116–128.

38. Chen, K.; Chen, X.; Yu, Z.; Zhu, M.; Yang, H. Equidiff: A conditional equivariant diffusion model for trajectory prediction. 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2023, pp. 746–751.

39. Blattmann, A.; Rombach, R.; Oktay, K.; Müller, J.; Ommer, B. Retrieval-augmented diffusion models. *Advances in Neural Information Processing Systems* **2022**, *35*, 15309–15324.

40. Hong, S.; Lee, G.; Jang, W.; Kim, S. Improving sample quality of diffusion models using self-attention guidance. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 7462–7471.

41. Ji, L.; Rao, Z.; Pan, S.J.; Lei, C.; Chen, Q. A Diffusion Model with State Estimation for Degradation-Blind Inverse Imaging. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Vol. 38, pp. 2471–2479.

42. Tian, Y.; Liu, W.; Lee, T. Diffusion-Based Mel-Spectrogram Enhancement for Personalized Speech Synthesis with Found Data. 2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE, 2023, pp. 1–7.

43. Jiang, H.; Luo, A.; Fan, H.; Han, S.; Liu, S. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (TOG)* **2023**, *42*, 1–14.

44. Dong, H.W.; Liu, X.; Pons, J.; Bhattacharya, G.; Pascual, S.; Serrà, J.; Berg-Kirkpatrick, T.; McAuley, J. CLIPSonic: Text-to-audio synthesis with unlabeled videos and pretrained language-vision models. 2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). IEEE, 2023, pp. 1–5.

45. Choi, J.; Kim, S.; Jeong, Y.; Gwon, Y.; Yoon, S. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938* **2021**.

46. Liu, e.a. More Control with Semantic Diffusion Guidance for Image Synthesis. *Journal of Image and Audio Synthesis* **2023**.

47. Cai, S.; Chan, E.R.; Peng, S.; Shahbazi, M.; Obukhov, A.; Van Gool, L.; Wetzstein, G. Diffdreamer: Towards consistent unsupervised single-view scene extrapolation with conditional diffusion models. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 2139–2150.

48. Carrillo, e.a. Interactive Line Art Colorization with Conditional Diffusion Models. *Journal of Specialized Techniques and Innovations in Diffusion Models* **2023**.

49. Mao, W.; Han, B.; Wang, Z. SketchFFusion: Sketch-guided image editing with diffusion model. 2023 IEEE International Conference on Image Processing (ICIP). IEEE, 2023, pp. 790–794.

50. Luo, J.; Li, Y.; Pan, Y.; Yao, T.; Feng, J.; Chao, H.; Mei, T. Semantic-conditional diffusion networks for image captioning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 23359–23368.

51. Hsu, T.; Sadigh, B.; Bulatov, V.; Zhou, F. Score dynamics: scaling molecular dynamics with picosecond timesteps via conditional diffusion model. *arXiv preprint arXiv:2310.01678* **2023**.

52. Yan, Q.; Hu, T.; Sun, Y.; Tang, H.; Zhu, Y.; Dong, W.; Van Gool, L.; Zhang, Y. Towards high-quality HDR deghosting with conditional diffusion models. *IEEE Transactions on Circuits and Systems for Video Technology* **2023**.

53. Peng, W.; Adeli, E.; Bosschieter, T.; Park, S.H.; Zhao, Q.; Pohl, K.M. Generating realistic brain mris via a conditional diffusion probabilistic model. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2023, pp. 14–24.

54. Yu, J.J.; Forghani, F.; Derpanis, K.G.; Brubaker, M.A. Long-Term Photometric Consistent Novel View Synthesis with Diffusion Models. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 7094–7104.

55. Yin, Y.; Xu, D.; Tan, C.; Liu, P.; Zhao, Y.; Wei, Y. Cle diffusion: Controllable light enhancement diffusion model. Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 8145–8156.

56. Papantoniou, F.P.; Lattas, A.; Moschoglou, S.; Zafeiriou, S. Relightify: Relightable 3d faces from a single image via diffusion models. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 8806–8817.

57. Kirch, S.; Olyunina, V.; Ondřej, J.; Pagés, R.; Martin, S.; Pérez-Molina, C. RGB-D-Fusion: Image Conditioned Depth Diffusion of Humanoid Subjects. *IEEE Access* **2023**.

58. Mao, Y.; Jiang, L.; Chen, X.; Li, C. Disc-diff: Disentangled conditional diffusion model for multi-contrast mri super-resolution. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2023, pp. 387–397.

59. Yang, Z.; Liu, B.; Xxiong, Y.; Yi, L.; Wu, G.; Tang, X.; Liu, Z.; Zhou, J.; Zhang, X. DocDiff: Document enhancement via residual diffusion models. Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 2795–2806.

60. Ordun, C.; Raff, E.; Purushotham, S. When visible-to-thermal facial GAN beats conditional diffusion. 2023 IEEE International Conference on Image Processing (ICIP). IEEE, 2023, pp. 181–185.

61. Kansy, M.; Raël, A.; Mignone, G.; Naruniec, J.; Schroers, C.; Gross, M.; Weber, R.M. Controllable Inversion of Black-Box Face Recognition Models via Diffusion. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 3167–3177.

62. Yu, J.; Wang, Y.; Zhao, C.; Ghanem, B.; Zhang, J. Freedom: Training-free energy-guided conditional diffusion model. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 23174–23184.

63. Bieder, F.; Wolleb, J.; Durrer, A.; Sandkühler, R.; Cattin, P.C. Diffusion models for memory-efficient processing of 3d medical images. *arXiv preprint arXiv:2303.15288* **2023**.

64. Amirhossein, K.; Khodapanah, A.E.; Moein, H.; Reza, A.; Mohsen, F.; Ilker, H.; Dorit, M. Diffusion models for medical image analysis: a comprehensive survey. *arXiv preprint arXiv* **2022**, *2211*.

65. Chen, T.; Wang, C.; Shan, H. Berdiff: Conditional bernoulli diffusion model for medical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2023, pp. 491–501.

66. Shrivastava, A.; Fletcher, P.T. NASDM: nuclei-aware semantic histopathology image generation using diffusion models. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2023, pp. 786–796.

67. Wang, X.; Shen, Z.; Song, Z.; Wang, S.; Liu, M.; Zhang, L.; Xuan, K.; Wang, Q. Arbitrary Reduction of MRI Inter-slice Spacing Using Hierarchical Feature Conditional Diffusion. International Workshop on Machine Learning in Medical Imaging. Springer, 2023, pp. 23–32.

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

69. Finzi, M.A.; Boral, A.; Wilson, A.G.; Sha, F.; Zepeda-Núñez, L. User-defined event sampling and uncertainty quantification in diffusion models for physical dynamical systems. International Conference on Machine Learning. PMLR, 2023, pp. 10136–10152.

70. Yang, X.; Li, W.; Zhang, M. Directional diffusion models for chaotic dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2024**.

71. Li, J.; Zhang, R.; Wang, H. Comparison of manifold learning techniques for conformational modeling. *Journal of Computational Biology* **2023**.

72. Li, A.C.; Prabhudesai, M.; Duggal, S.; Brown, E.; Pathak, D. Your diffusion model is secretly a zero-shot classifier. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 2206–2217.

73. Zhuang, Y.; Hou, B.; Mathai, T.S.; Mukherjee, P.; Kim, B.; Summers, R.M. Semantic Image Synthesis for Abdominal CT. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2023, pp. 214–224.

74. Jiang, W.; Diao, Y.; Wang, H.; Sun, J.; Wang, M.; Hong, R. Unlearnable examples give a false sense of security: Piercing through unexploitable data with learnable examples. Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 8910–8921.

75. Wang, X.; López-Tapia, S.; Katsaggelos, A.K. Atmospheric turbulence correction via variational deep diffusion. 2023 IEEE 6th International Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, 2023, pp. 1–4.

76. Sartor, S.; Peers, P. Matfusion: a generative diffusion model for svbrdf capture. SIGGRAPH Asia 2023 Conference Papers, 2023, pp. 1–10.

77. Wei, Y.; Vosselman, G.; Yang, M.Y. BuildDiff: 3D Building Shape Generation using Single-Image Conditional Point Cloud Diffusion Models. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 2910–2919.

78. Niu, A.; Pham, T.X.; Zhang, K.; Sun, J.; Zhu, Y.; Yan, Q.; Kweon, I.S.; Zhang, Y. ACDMSR: Accelerated conditional diffusion models for single image super-resolution. IEEE Transactions on Broadcasting 2024.

79. Sun, R.; Dong, K.; Zhao, J. DiffNILM: a novel framework for non-intrusive load monitoring based on the conditional diffusion model. Sensors 2023, 23, 3540.

80. Yang, D.; Liu, S.; Yu, J.; Wang, H.; Weng, C.; Zou, Y. Norespeech: Knowledge distillation based conditional diffusion model for noise-robust expressive tts. arXiv preprint arXiv:2211.02448 2022.

81. Yu, X.; Li, G.; Lou, W.; Liu, S.; Wan, X.; Chen, Y.; Li, H. Diffusion-based data augmentation for nuclei image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2023, pp. 592–602.

82. Ma, R.; Duan, J.; Kong, F.; Shi, X.; Xu, K. Exposing the fake: Effective diffusion-generated images detection. arXiv preprint arXiv:2307.06272 2023.

83. Zhao, Z.; Duan, J.; Hu, X.; Xu, K.; Wang, C.; Zhang, R.; Du, Z.; Guo, Q.; Chen, Y. Unlearnable examples for diffusion models: Protect data from unauthorized exploitation. arXiv preprint arXiv:2306.01902 2023.

84. Mei, K.; Patel, V. Vidm: Video implicit diffusion models. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, Vol. 37, pp. 9117–9125.

85. Mao, J.; Liu, J.; Zhang, W. SketchFusion: A Model for Sketch-Guided Image Editing Using a Conditional Diffusion Model. Journal of Graphics and Image Processing 2023, 12, 123–134.

86. Wang, H.; Li, Y.; He, N.; Ma, K.; Meng, D.; Zheng, Y. DICDNet: deep interpretable convolutional dictionary network for metal artifact reduction in CT images. IEEE Transactions on Medical Imaging 2021, 41, 869–880.

87. Maass, N.; Maier, A.; Wuerfl, T. Reducing image artifacts, 2018. United States Patent Application 20180192985.

88. Filoscia, I.; Alderighi, T.; Giorgi, D.; Malomo, L.; Callieri, M.; Cignoni, P. Optimizing object decomposition to reduce visual artifacts in 3D printing. Computer Graphics Forum. Wiley Online Library, 2020, Vol. 39, pp. 423–434.

89. Han, S.; Mao, H.; Dally, W.J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. arXiv preprint arXiv:1510.00149 2015.

90. Jacob, B.; Kligys, S.; Chen, B.; Zhu, M.; Tang, M.; Howard, A.; Adam, H.; Kalenichenko, D. Quantization and training of neural networks for efficient integer-arithmetic-only inference. arXiv preprint arXiv:1712.05877 2018.

91. Shin, H.; Park, H.; Cho, K.Y.; Kim, S.K. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. Proceedings of the Medical Imaging Technology Conference 2018.

92. Yan, M.; Liu, Y.; Park, Y. RADIA: Protecting Patient Privacy in Radiology Reports. IEEE Journal of Biomedical and Health Informatics 2021.

93. Yu, e.a. Transfer Learning in Medical Imaging. Journal of Biomedical Engineering 2022.

94. Chen, e.a. Incorporating Domain-Specific Knowledge in Diffusion Models for Medical Imaging. Journal of Medical Imaging and Diagnostics 2023.

95. Goodell, J.W.; Goutte, C. Toward AI and data analytics for financial inclusion: A review. *Journal of Financial Stability* **2021**.

96. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* **2002**, *16*, 321–357.

97. Wang, Y.; Li, X.; Yang, L.; Ma, J.; Li, H. ADDITION: Detecting Adversarial Examples With Image-Dependent Noise Reduction. *IEEE Transactions on Dependable and Secure Computing* **2023**.

98. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on fairness, accountability and transparency. PMLR, 2018, pp. 77–91.

99. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* **2016**.

100. Dwork, C.; Roth, A.; others. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* **2014**, *9*, 211–407.

101. et al., F. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines* **2018**.