

References

30

- [] Bent, B. 2024. Semantic Approach to Quantifying the Consistency of Diffusion Model Image Generation. *arXiv:2404.08799*.
- [] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Networks. *arXiv:1406.2661*.
- [] Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. *arXiv:2006.11239*.
- [] Hu, X.; Jin, Y.; Liang, J.; Liu, J.; Luo, R.; Li, M.; and Peng, T. 2023. Diffusion Model for Image Generation - A Survey. In *2023 2nd International Conference on Artificial Intelligence, Human-Computer Interaction and Robotics (AI-HCIR)*, 416–424.
- [] Huang, Y.; Huang, J.; Liu, Y.; Yan, M.; Lv, J.; Liu, J.; Xiong, W.; Zhang, H.; Chen, S.; and Cao, L. 2024. Diffusion Model-Based Image Editing: A Survey. *arXiv:2402.17525*.
- [] Hyvärinen, A. 2005. Estimation of Non-Normalized Statistical Models by Score Matching. *Journal of Machine Learning Research*, 6(24): 695–709.
- [] Kingma, D. P.; and Welling, M. 2022. Auto-Encoding Variational Bayes. *arXiv:1312.6114*.
- [] Lam, M. W. Y.; Wang, J.; Su, D.; and Yu, D. 2022. BDDM: Bilateral Denoising Diffusion Models for Fast and High-Quality Speech Synthesis. *arXiv:2203.13508*.
- [] Li, X.; Ren, Y.; Jin, X.; Lan, C.; Wang, X.; Zeng, W.; Wang, X.; and Chen, Z. 2023a. Diffusion Models for Image Restoration and Enhancement – A Comprehensive Survey. *arXiv:2308.09388*.
- [] Li, Y.; Zhou, K.; Zhao, W. X.; and Wen, J.-R. 2023b. Diffusion Models for Non-autoregressive
- [] Moser, B. B.; Shanbhag, A. S.; Raue, F.; Frolov, S.; Palacio, S.; and Dengel, A. 2024. Diffusion Models, Image Super-Resolution, and Everything: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, 1–21.
- [] Pan, Z.; Zhuang, B.; Huang, D.-A.; Nie, W.; Yu, Z.; Xiao, C.; Cai, J.; and Anandkumar, A. 2024. T-Stitch: Accelerating Sampling in Pre-Trained Diffusion Models with Trajectory Stitching. *arXiv:2402.14167*.
- [] Popov, V.; Vovk, I.; Gogoryan, V.; Sadekova, T.; and Kudinov, M. 2021. Grad-TTS: A Diffusion Probabilistic Model for Text-to-Speech. *arXiv:2105.06337*.
- [] Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; and Chen, M. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv:2204.06125*.
- [] Richter, J.; Welker, S.; Lemercier, J.-M.; Lay, B.; and Gerkmann, T. 2023. Speech Enhancement and Dereverberation with Diffusion-based Generative Models. *arXiv:2208.05830*.
- [] Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. *arXiv:2112.10752*.
- [] Sohl-Dickstein, J.; Weiss, E. A.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv:1503.03585*.
- [] Song, J.; Meng, C.; and Ermon, S. 2022. Denoising Diffusion Implicit Models. *arXiv:2010.02502*.
- [] Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2021. Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv:2011.13456*.
- [] Wang, Z.; Li, D.; and Jiang, R. 2024. Diffusion Models in 3D Vision: A Survey. *arXiv:2410.04738*.
- [] Wang, Z.; Lin, J.; Qian, Y.; Huang, Y.; Tian, S.; Chai, B.; Deng, J.; Yang, Q.; Du, L.; Chen,

□ Zou, H.-L.; Kim, Z. M.; and Kang, D. 2023. A Survey of Diffusion Models in Natural Language Processing.

□ 曾庆宇; and 王师琦. 2021. 扩散噪声下协方差矩阵重构的语音分离与降噪. 声学学报, 46(5): 775–784.

/TemplateVersion (2025.1)

A simple survey of Diffusion Model

Shuanggang East Street, Qingshan Lake District
Nanchang City, Jiangxi Province, China Postal Code: 330044

Abstract

扩散模型 (Diffusion Models) 是近年来在深度学习领域迅速崛起的一类生成模型。与传统的生成对抗网络 (GANs) 和变分自编码器 (VAEs) 不同, 扩散模型通过模拟数据从噪声到真实数据的逐步转化过程, 在多个领域取得了突破性进展。本文粗浅地总结了扩散模型的基本原理、技术发展历程、主要应用领域以及面临的挑战。通过分析扩散模型的数学基础及其与传统模型的区别, 探讨了其在各个应用场景中的优势与创新, 并对未来的研究方向进行了展望。研究表明, 扩散模型在生成质量、训练稳定性和应用灵活性等方面具有显著优势, 有望成为下一代深度生成模型的重要发展方向。

1 引言

深度学习在过去十年间取得了革命性的突破, 特别是在计算机视觉、自然语言处理、语音合成等领域。生成模型作为深度学习的重要分支, 致力于生成与真实数据分布相似的数据。传统的生成模型, 如生成对抗网络 (GANs) 和变分自编码器 (VAEs), 在生成质量上取得了优异的成绩。然而, 这些模型往往存在训练不稳定、模式崩溃、生成多样性不足等挑战。

扩散模型作为一种新兴的生成模型范式, 通过设计巧妙的正向扩散和反向去噪过程, 成功解决了传统生成模型的诸多问题。自 2015 年由 Sohl-Dickstein 等人首次提出以来, 扩散模型在理论基础和实际应用两个方面都取得了长足的进展。特别是在图像生成领域, 扩散模型展现出了超越 GANs[2] 和 VAEs[7] 的潜力, 成为目前最受关注的生成模型之一。随着技术

的不断进步, 扩散模型的应用范围已经从最初的图像生成扩展到语音合成、文本生成、分子设计等诸多领域。

本文将从以下几个方面系统地介绍扩散模型的研究进展: 第二部分介绍扩散模型的基本原理和数学描述; 第三部分详细探讨扩散模型的发展历史和最新的技术突破; 第四部分深入分析扩散模型在各个领域的具体应用及其创新; 第五部分讨论当前面临的主要挑战和未来可能的研究方向。

2 扩散模型基础

2.1 扩散模型的基本原理

扩散模型 (Diffusion Models) 是通过一系列可逆的噪声添加与去噪过程来生成数据。正向扩散过程将真实数据逐步添加噪声, 直到数据变成纯噪声。反向去噪过程则是通过训练一个去噪网络来学习如何从噪声中逐步恢复到原始数据。具体来说, 扩散模型的正向过程由以下公式描述:

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

其中, x_t 表示第 t 步的样本, β_t 是一个控制噪声添加速率的参数, $\mathcal{N}()$ 表示正态分布。

反向过程的目标是学习如何从噪声中逐步恢复数据, 通常通过最小化以下损失函数来训练去噪网络:

$$L(\theta) = E_{q(x_0, \dots, x_T)} [\|\epsilon_\theta(x_t, t) - \epsilon\|^2]$$

其中, $\epsilon_\theta(x_t, t)$ 是去噪网络的输出, ϵ 是真实的噪声。

2.2 与传统深度学习模型的比较

与 GANs 相比, 扩散模型避免了对抗训练带来的不稳定性。GANs 需要同时训练生成器和判别器, 这种对抗训练机制容易导致模式崩溃和训练不稳定。相比之下, 扩散模型通过优化简单的去噪目标函数进行训练, 过程更加稳定可控。与 VAEs 相比, 扩散模型不需要对复杂的后验分布进行显式建模。VAEs 通过变分推断来学习数据的潜在表示, 这个过程往往会引入表示能力的瓶颈。扩散模型则直接在数据空间中进行建模, 避免了潜在空间表示的限制。

3 扩散模型的发展历程

3.1 早期探索阶段 (2015-2019)

扩散模型在 2015 年由 Sohl-Dickstein 等人 [21] 提出, 他们创新性地将非平衡热力学中的朗之万动力学引入深度学习, 为扩散模型奠定了理论框架。在这个阶段, 研究者们主要致力于构建理论基础。概率流常微分方程 (ODE) [22] 的引入为模型提供了一种新的数学描述方式, 使得对扩散过程的理解更加深入。得分匹配 (Score Matching) 方法 [6] 的应用则为模型的训练提供了一种有效的手段。然而, 尽管这些早期工作在理论上具有重要意义, 但在实际应用效果方面, 扩散模型还难以与当时已经较为成熟的 GANs 等模型相媲美。这一时期的研究为后续的发展奠定了坚实的基础。一方面, 理论框架的建立为后续的改进和优化提供了方向; 另一方面, 对概率流常微分方程和得分匹配方法的探索, 也为扩散模型的进一步发展积累了经验。

3.2 快速发展阶段 (2020-2021)

2020 年, Ho 等人 [3] 提出了 DDPM (Denoising Diffusion Probabilistic Models), 这一成果被认为是扩散模型发展的一个重要里程碑。DDPM 通过简化训练目标和改进网络架构, 显著提升了扩散模型的生成质量。它使得扩散模型在图像生成等任务上的表现有了质的飞跃, 开始引起广泛关注。紧接着, Song 等人提出了基于得分的生成模型 (Score-based Generative Models) [23]。该模型将连续时间扩散过程与神经网络相结合, 进一步推动了扩散模型的理论

发展。这一阶段的研究不仅在理论上深化了对扩散模型的理解, 而且在实际应用中也展示出了巨大的潜力。

3.3 应用突破阶段 (2022 至今)

近两年来, 扩散模型在实际应用方面取得了突破性进展。DALL-E 2[18]、Stable Diffusion[20] 等模型展示了扩散模型在文本引导图像生成方面的巨大潜力。这些模型能够根据用户提供的文本描述生成高质量的图像, 为创意产业和设计领域带来了新的机遇。研究者们还提出了各种加速采样的方法, 如 DDIM (Denoising Diffusion Implicit Models) [3]、采样技巧优化等 [26]。这些方法大大提高了模型的实用性, 使得扩散模型在实际应用中的效率得到了显著提升。此外, 扩散模型在其他领域也开始展现出广泛的应用前景。例如, 在视频生成领域 [14], 基于扩散模型的视频生成技术正在迅速发展, 为视频制作带来了新的创意和可能性。

总的来说, 扩散模型的发展历程充满了创新和突破。从早期的理论探索到快速发展阶段的重要成果, 再到如今在实际应用中的突破性进展, 扩散模型已经成为人工智能领域的一个重要研究方向。随着技术的不断进步, 相信扩散模型在未来还将在更多领域发挥重要作用。

3.4 主要挑战

扩散模型在早期探索阶段面临着诸多理论基础构建的挑战

- 数据复杂性带来的挑战:

在早期阶段, 扩散模型需要处理复杂的数据结构和特征。不同类型的数据可能具有不同的分布和特性, 这使得构建适用于多种数据类型的理论基础变得困难。例如, 图像数据 [15]、文本数据和音频数据在特征表示和分布上存在很大差异, 如何设计一种通用的理论框架能够有效地处理这些不同类型的数据是一个重大挑战。同时, 高维数据的处理也是一个难题, 高维空间中的数据分布往往非常复杂, 难以准确地建模和分析。这就要求在理论基础构建中找到有效的方法来降低数据维

度，同时保留关键信息，以便更好地进行扩散模型的训练和应用。

- 模型可解释性的挑战

早期的扩散模型往往缺乏可解释性 [1]，这使得人们难以理解模型的决策过程和预测结果。对于实际应用来说，可解释性是非常重要的，因为它可以帮助用户信任模型并做出合理的决策。例如，在医疗领域，如果一个扩散模型用于疾病诊断，医生需要了解模型是如何做出诊断的，以便对结果进行评估和验证。然而，早期的扩散模型通常是黑箱模型，难以解释其内部的工作机制。为了解决这个问题，研究人员需要在理论基础构建中注重模型的可解释性，开发出能够解释模型决策过程的方法和技术。

- 计算效率的挑战

扩散模型的计算复杂度通常较高，尤其是在处理大规模数据时。在早期阶段，计算资源有限，如何提高模型的计算效率是一个关键问题。例如，在训练扩散模型时，需要进行大量的迭代计算，这可能需要很长时间才能完成。此外，在模型的应用阶段，也需要快速地对新数据进行处理和预测。因此，在理论基础构建中，需要考虑如何设计高效的算法和计算框架，以提高扩散模型的计算效率 [12]。

- 参数调整 and 优化的挑战

扩散模型通常具有大量的参数需要调整和优化。在早期阶段，缺乏有效的参数调整方法和优化策略，这使得模型的性能难以达到最佳状态。例如，不同的参数设置可能会导致模型在不同的数据集上表现出很大的差异，如何找到最优的参数组合是一个难题。此外，参数的调整和优化也需要大量的计算资源和时间，这进一步增加了理论基础构建的难度。

- 模型稳定性的挑战

早期的扩散模型在训练和应用过程中可能会出现不稳定的情况。例如，模型可能会陷入局部最优解，或者在训练过程中出现梯度消失或爆炸的问题。这些不稳定因素会影响模型的性能和可靠性，

使得理论基础构建更加困难。为了解决这个问题，研究人员需要在理论基础构建中考虑模型的稳定性，开发出能够提高模型稳定性的方法和技术 [20]。

综上所述，扩散模型在早期探索阶段面临着数据复杂性、模型可解释性、计算效率、参数调整和优化以及模型稳定性等多方面的挑战。这些挑战需要研究人员在理论基础构建中加以解决，以推动扩散模型的发展和应用。

4 扩散模型的应用

4.1 图像生成

在图像生成领域，扩散模型已经成为最具竞争力的方法之一。主要应用包括：

- 高清图像生成

通过级联扩散模型或自适应注意力机制，扩散模型能够生成分辨率高达 2048×2048 的逼真图像 [9]。除了级联扩散模型和自适应注意力机制外，还有一些创新的方法进一步提升了扩散模型在高清图像生成方面的能力。例如，采用多尺度特征融合策略，将不同层次的特征进行融合，从而更好地捕捉图像的细节和全局结构。同时，一些研究通过引入对抗训练机制，增强生成图像的真实性和清晰度。相关研究中提到 [4]，扩散模型在图像生成领域的应用包括无条件生成、条件生成、文本到图像翻译等多个方面，其中高清图像生成是其重要应用之一。扩散模型的出色图像生成能力使其成为当前图像生成领域的研究热点。

- 文本引导图像生成

结合大规模语言模型，扩散模型能够准确理解复杂的文本描述并生成对应的图像。代表性工作如 DALL-E 2 [18] 和 Stable Diffusion [20] 不仅能生成高质量图像，还具备优秀的文本理解能力。此外，多模态融合也是一个重要的研究方向，将文本、图像和其他模态信息进行融合，以生成更加丰富和准确的图像。如 [25] 中提出的 DiffX 模型，是一种新颖的扩散模型，用于一般布局引导的跨模态生成，通过引入联合模态嵌入器 (JME)，增强

了布局和文本条件之间的交互，展示了在跨模态“RGB+X”图像生成方面的强大能力。

- 图像编辑与修复

扩散模型在图像修复、风格迁移、局部编辑等任务上展现出独特优势。通过巧妙设计的条件生成机制，模型能够保持图像的整体结构同时实现局部细节的精确修改。[5]。同时，一些研究通过结合深度学习中的其他技术，如生成对抗网络（GAN），进一步提升图像编辑的效果。此外，为了满足不同用户的需求，一些方法还提供了交互式的图像编辑界面，使用户能够更加直观地进行图像编辑。[4] 中也对图像编辑和修复技术进行了深入讨论，指出扩散模型在这些任务上具有很大的潜力。

4.2 语音合成

在语音领域，扩散模型为解决传统方法面临的问题提供了新思路：

- 高质量语音合成

扩散模型在高质量语音合成方面的优势不仅在于生成自然、富有表现力的语音，还在于其能够处理不同的语音风格和情感。一些研究通过引入情感标签或风格向量，使模型能够生成具有特定情感或风格的语音。同时，为了提高语音合成的效率，一些方法采用了并行化的计算策略，加快了模型的训练和推理速度。例如 [8] 中提出的双边去噪扩散模型（BDDM），能够生成高保真的音频样本，并且在采样速度上具有很大的优势。[17]。

- 语音转换

在语音转换任务中，扩散模型可以通过学习不同说话人的语音特征，实现说话人身份转换。同时，还可以通过控制模型的输入参数，实现情感风格转换等任务。一些研究还探索了多模态信息在语音转换中的应用，如结合图像或文本信息，使语音转换更加自然和准确。[8] 中的 BDDM 模型也可以应用于语音转换任务，通过调整模型的参数和输入条件，实现不同类型的语音转换。

- 语音增强

扩散模型在语音增强任务中的性能优势主要体现

在对噪声的抑制和语音信号的恢复上。一些研究通过设计专门的噪声模型，使扩散模型能够更好地适应不同类型的噪声环境。同时，为了提高语音增强的效果，一些方法还结合了其他信号处理技术，如谱减法、维纳滤波等。[30] 中提出的用于语音分离及降噪的空间协方差模型及参数估计方法，将扩散噪声视为独立声源，利用由导向矢量重构的空间协方差矩阵建模目标声源的空间特性，为语音增强提供了一种有效的方法。[19]

4.3 自然语言处理

扩散模型在文本生成和处理方面也取得了重要进展：

- 文本生成

扩散模型在文本生成方面的应用不仅仅局限于生成连贯且符合语法的文本，还可以通过引入主题模型或情感分析等技术，使生成的文本具有特定的主题或情感倾向。同时，为了提高文本生成的多样性和创造性，一些方法采用了随机采样或蒙特卡洛方法，增加生成文本的随机性。[28] 此外，多语言文本生成也是一个重要的研究方向，扩散模型可以通过学习不同语言的语法和语义特征，实现多语言文本的生成。[29] 中提到扩散模型在自然语言处理中有多种应用，如自然语言生成、情感分析、话题建模等，其中文本生成是其重要应用之一。

- 机器翻译

扩散模型为神经机器翻译提供了新的建模范式，特别是在处理低资源语言对时显示出优势。一些研究通过引入多模态信息，如图像、音频等，提高机器翻译的准确性和鲁棒性。同时，为了提高模型的训练效率，一些方法采用了分布式训练或模型压缩技术，减少模型的训练时间和存储空间。[29] 中对扩散模型在机器翻译中的应用进行了讨论，指出其在处理低资源语言对时的优势。

- 文本改写

在文本改写任务中，扩散模型可以通过学习不同的文本风格和表达方式，实现文本风格转换、简

化复杂表达等任务。[10] 一些研究通过引入对抗训练机制，提高文本改写的质量和多样性。同时，为了满足不同用户的需求，一些方法还提供了交互式的文本改写界面，使用户能够更加直观地进行文本改写。[29] 中也提到了扩散模型在文本改写方面的应用潜力。

4.4 跨模态生成

扩散模型在处理跨模态任务时展现出独特优势：

- 视频生成
通过时空扩散模型，实现了从文本描述或静态图像生成连贯视频的能力。[14] 一些研究通过引入注意力机制或递归神经网络，提高视频生成的质量和连贯性。同时，为了提高视频生成的效率，一些方法采用了并行化的计算策略，加快了模型的训练和推理速度。此外，多模态融合也是视频生成的一个重要研究方向，将文本、图像、音频等多种模态信息进行融合，以生成更加丰富和准确的视频。[25] 中的 DiffX 模型在跨模态 “RGB+X” 图像生成方面的成功，为视频生成提供了新的思路和方法。
- 3D 内容生成
结合新视角合成技术，扩散模型能够从单一图像重建 3D 场景或物体。[24] 一些研究通过引入深度学习中的其他技术，如卷积神经网络（CNN）或生成对抗网络（GAN），提高 3D 内容生成的质量和准确性。同时，为了满足不同用户的需求，一些方法还提供了交互式的 3D 内容生成界面，使用户能够更加直观地进行 3D 内容创作。[4] 中对扩散模型在 3D 内容生成方面的应用进行了讨论，指出其在该领域的巨大潜力。
- 音乐生成
通过建模音频信号的时频表示，实现了高质量的音乐生成和编辑。[27] 一些研究通过引入音乐理论知识或情感分析技术，使生成的音乐更加符合人类的听觉感受和情感需求。同时，为了提高音乐生成的效率，一些方法采用了并行化的计算策略，加快了模型的训练和推理速度。此外，多模态

融合也是音乐生成的一个重要研究方向，将文本、图像等其他模态信息与音频信号进行融合，以生成更加丰富和多样化的音乐。[11] 中虽然主要讨论的是扩散模型在时间序列的应用，但也为音乐生成等领域提供了一些参考和启示。

5 挑战与展望

5.1 当前面临的主要挑战

- 计算效率
尽管已有多种加速方法，但扩散模型的采样过程仍然相对耗时，这限制了其在实时应用中的使用。扩散模型通常需要大量的计算资源来进行采样，尤其是在处理高分辨率图像或大规模数据集时。这是因为扩散模型的采样过程是一个逐步去噪的过程，需要进行多次迭代才能得到高质量的生成结果。例如，在图像生成任务中，扩散模型可能需要数百次甚至数千次的迭代才能生成清晰的图像。这种计算成本使得扩散模型在实时应用中面临巨大的挑战，例如视频生成、实时交互等场景。
- 训练稳定性
虽然比生成对抗网络（GANs）更稳定，但在处理复杂数据分布时仍存在训练不稳定的问题。扩散模型通过逐步添加噪声并学习去噪过程来生成数据，这种方法在一定程度上避免了 GANs 中常见的模式崩溃和训练不稳定问题。然而，当处理复杂的数据分布时，扩散模型仍然可能出现训练不稳定的情况。例如，在处理具有多模态分布的数据时，扩散模型可能会陷入局部最优解，导致生成的结果不够多样化。此外，扩散模型的训练过程也受到超参数的影响，不同的超参数设置可能会导致不同的训练结果，这也增加了训练的难度。[16]
- 模型理论
对扩散过程的理论理解仍然不够深入，特别是在最优采样策略的设计方面。目前，虽然已经有一些关于扩散模型的理论研究，但对于扩散过程的理解仍然不够深入。例如，对于扩散模型中噪声的添加和去除过程，以及如何选择最优的采样策

略等问题，仍然没有得到很好的解决。这使得在设计和优化扩散模型时缺乏理论指导，往往需要通过大量的实验来进行调整。[16]

5.2 未来研究方向

- 改进采样效率

研究新的采样算法和加速技术，减少生成过程所需的计算资源。为了提高扩散模型的采样效率，可以从多个方面进行研究。一方面，可以探索新的采样算法，例如基于重要性采样、马尔可夫链蒙特卡洛（MCMC）等方法的采样算法，以提高采样的效率和质量。另一方面，可以利用硬件加速技术，如 GPU、TPU 等，来加速扩散模型的采样过程。此外，还可以通过模型压缩、知识蒸馏等技术来减少模型的计算量，提高采样效率。

- 增强控制能力

开发更精确的条件控制机制，使模型能够更好地响应用户意图。为了使扩散模型能够更好地响应用户意图，可以开发更精确的条件控制机制。例如，可以通过引入条件变量、注意力机制等方法，使模型能够根据用户提供的条件信息生成特定的输出。[16] 此外，还可以探索多模态条件控制机制，使模型能够同时根据图像、文本等多种模态的信息进行生成，提高模型的灵活性和通用性。

- 理论突破

深化对扩散过程的理论理解，为模型设计提供更好的指导。为了进一步提高扩散模型的性能，需要深化对扩散过程的理论理解。这包括对扩散模型的数学性质、收敛性、稳定性等方面的研究，以及对最优采样策略、模型结构设计等问题的探索。通过理论研究，可以为扩散模型的设计和优化提供更好的指导，提高模型的性能和泛化能力。

- 扩展应用范围

探索在新领域的应用可能，如药物分子设计、蛋白质结构预测等。扩散模型具有强大的生成能力和灵活性，可以应用于多个领域。例如，在药物分子设计领域，可以利用扩散模型生成具有特定性质的药物分子结构，为药物研发提供新的思路和

方法[?]。在蛋白质结构预测领域，可以利用扩散模型学习蛋白质的结构分布，预测未知蛋白质的结构，为生命科学研究提供帮助。此外，扩散模型还可以应用于语音合成、视频生成等领域，具有广阔的应用前景。[13]

6 结论

扩散模型作为一种新兴的生成模型范式，通过其独特的渐进式生成机制，在多个领域展现出了强大的潜力。其良好的训练稳定性、优秀的生成质量以及灵活的应用能力，使其成为深度学习领域最具前景的研究方向之一。虽然目前仍面临一些挑战，但随着理论研究的深入和技术的不断创新，扩散模型有望在更多领域发挥重要作用，推动生成模型技术的进一步发展。

References

- [1] Bent, B. 2024. Semantic Approach to Quantifying the Consistency of Diffusion Model Image Generation. arXiv:2404.08799.
- [2] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Networks. arXiv:1406.2661.
- [3] Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. arXiv:2006.11239.
- [4] Hu, X.; Jin, Y.; Liang, J.; Liu, J.; Luo, R.; Li, M.; and Peng, T. 2023. Diffusion Model for Image Generation - A Survey. In *2023 2nd International Conference on Artificial Intelligence, Human-Computer Interaction and Robotics (AIHCIR)*, 416–424.
- [5] Huang, Y.; Huang, J.; Liu, Y.; Yan, M.; Lv, J.; Liu, J.; Xiong, W.; Zhang, H.; Chen, S.; and Cao, L. 2024. Diffusion Model-Based Image Editing: A Survey. arXiv:2402.17525.
- [6] Hyvärinen, A. 2005. Estimation of Non-Normalized Statistical Models by Score Match-

- ing. *Journal of Machine Learning Research*, 6(24): 695–709.
- [7] Kingma, D. P.; and Welling, M. 2022. Auto-Encoding Variational Bayes. arXiv:1312.6114.
 - [8] Lam, M. W. Y.; Wang, J.; Su, D.; and Yu, D. 2022. BDDM: Bilateral Denoising Diffusion Models for Fast and High-Quality Speech Synthesis. arXiv:2203.13508.
 - [9] Li, X.; Ren, Y.; Jin, X.; Lan, C.; Wang, X.; Zeng, W.; Wang, X.; and Chen, Z. 2023. Diffusion Models for Image Restoration and Enhancement – A Comprehensive Survey. arXiv:2308.09388.
 - [10] Li, Y.; Zhou, K.; Zhao, W. X.; and Wen, J.-R. 2023. Diffusion Models for Non-autoregressive Text Generation: A Survey. arXiv:2303.06574.
 - [11] Lin, L.; Li, Z.; Li, R.; Li, X.; and Gao, J. 2024. Diffusion models for time-series applications: a survey. *Frontiers of Information Technology & Electronic Engineering*, 25(1): 19–41.
 - [12] Luhman, T.; and Luhman, E. 2022. Improving Diffusion Model Efficiency Through Patching. arXiv:2207.04316.
 - [13] Luo, C. 2022. Understanding Diffusion Models: A Unified Perspective. arXiv:2208.11970.
 - [14] Melnik, A.; Ljubljanac, M.; Lu, C.; Yan, Q.; Ren, W.; and Ritter, H. 2024. Video Diffusion Models: A Survey. arXiv:2405.03150.
 - [15] Moser, B. B.; Shanbhag, A. S.; Raue, F.; Frolov, S.; Palacio, S.; and Dengel, A. 2024. Diffusion Models, Image Super-Resolution, and Everything: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, 1–21.
 - [16] Pan, Z.; Zhuang, B.; Huang, D.-A.; Nie, W.; Yu, Z.; Xiao, C.; Cai, J.; and Anandkumar, A. 2024. T-Stitch: Accelerating Sampling in Pre-Trained Diffusion Models with Trajectory Stitching. arXiv:2402.14167.
 - [17] Popov, V.; Vovk, I.; Gogoryan, V.; Sadekova, T.; and Kudinov, M. 2021. Grad-TTS: A Diffusion Probabilistic Model for Text-to-Speech. arXiv:2105.06337.
 - [18] Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; and Chen, M. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv:2204.06125.
 - [19] Richter, J.; Welker, S.; Lemercier, J.-M.; Lay, B.; and Gerkmann, T. 2023. Speech Enhancement and Dereverberation with Diffusion-based Generative Models. arXiv:2208.05830.
 - [20] Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752.
 - [21] Sohl-Dickstein, J.; Weiss, E. A.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. arXiv:1503.03585.
 - [22] Song, J.; Meng, C.; and Ermon, S. 2022. Denoising Diffusion Implicit Models. arXiv:2010.02502.
 - [23] Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2021. Score-Based Generative Modeling through Stochastic Differential Equations. arXiv:2011.13456.
 - [24] Wang, Z.; Li, D.; and Jiang, R. 2024. Diffusion Models in 3D Vision: A Survey. arXiv:2410.04738.
 - [25] Wang, Z.; Lin, J.; Qian, Y.; Huang, Y.; Tian, S.; Chai, B.; Deng, J.; Yang, Q.; Du, L.; Chen, C.; and Huang, K. 2024. DiffX: Guide Your Layout to Cross-Modal Generative Modeling. arXiv:2407.15488.
 - [26] Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Zhang, W.; Cui, B.; and Yang, M.-H. 2024. Diffusion Models: A Compre-

hensive Survey of Methods and Applications.
arXiv:2209.00796.

- [27] Zhang, C.; Zhang, C.; Zheng, S.; Zhang, M.; Qamar, M.; Bae, S.-H.; and Kweon, I. S. 2023. A Survey on Audio Diffusion Models: Text To Speech Synthesis and Enhancement in Generative AI. arXiv:2303.13336.
- [28] Zhu, Y.; and Zhao, Y. 2023. Diffusion Models in NLP: A Survey. arXiv:2303.07576.
- [29] Zou, H.-L.; Kim, Z. M.; and Kang, D. 2023. A Survey of Diffusion Models in Natural Language Processing.
- [30] 曾庆宁; and 王师琦. 2021. 扩散噪声下协方差矩阵重构的语音分离与降噪. 声学学报, 46(5): 775–784.