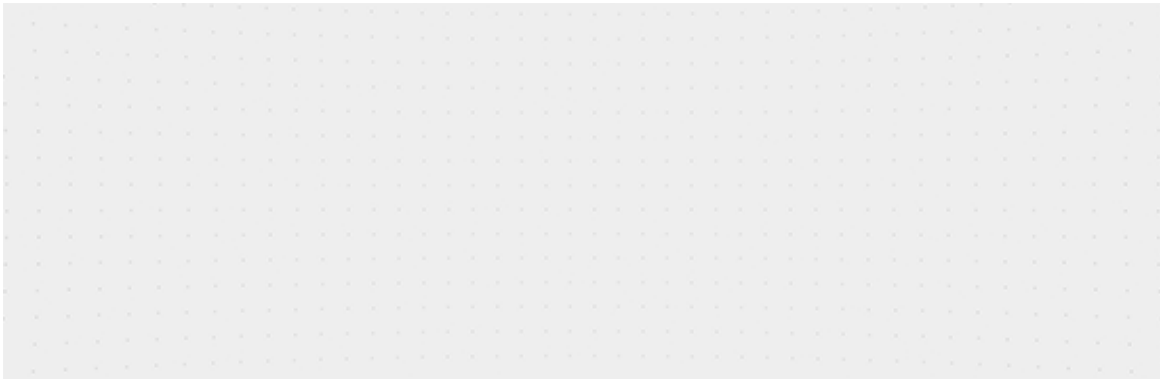


# 首篇扩散模型的综述与Github论文分类汇总

CV开发者都爱看的 极市平台 2022-09-13 22:00:36 发表于广东 手机阅读 𠄎

↑ 点击蓝字 关注极市平台



作者 | 杨灵

来源 | 机器之心

编辑 | 极市平台

极市导读

本文首次对现有的扩散生成模型（diffusion model）进行了全面的总结分析，还在Github分类汇总了相关论文。>>加入极市CV技术交流群，走在计算机视觉的最前沿

本综述（Diffusion Models: A Comprehensive Survey of Methods and Applications）来自加州大学 & Google Research 的 Ming-Hsuan Yang、北京大学崔斌实验室以及 CMU、UCLA、蒙特利尔 Mila 研究院等众研究团队，首次对现有的扩散生成模型（diffusion model）进行了全面的总结分析，从 diffusion model 算法细化分类、和其他五大生成模型的关联以及在七大领域中的应用等方面展开，最后提出了 diffusion model 的现有 limitation 和未来的发展方向。

# Diffusion Models: A Comprehensive Survey of Methods and Applications

LING YANG\*, Peking University, China

ZHILONG ZHANG\*, Peking University, China

SHENDA HONG, Peking University, China

RUNSHENG XU, University of California, Los Angeles, USA

YUE ZHAO, Carnegie Mellon University, USA

YINGXIA SHAO, Beijing University of Posts and Telecommunications, China

WENTAO ZHANG, Mila - Québec AI Institute, HEC Montréal, Canada

MING-HSUAN YANG, University of California at Merced, USA

BIN CUI, Peking University, China

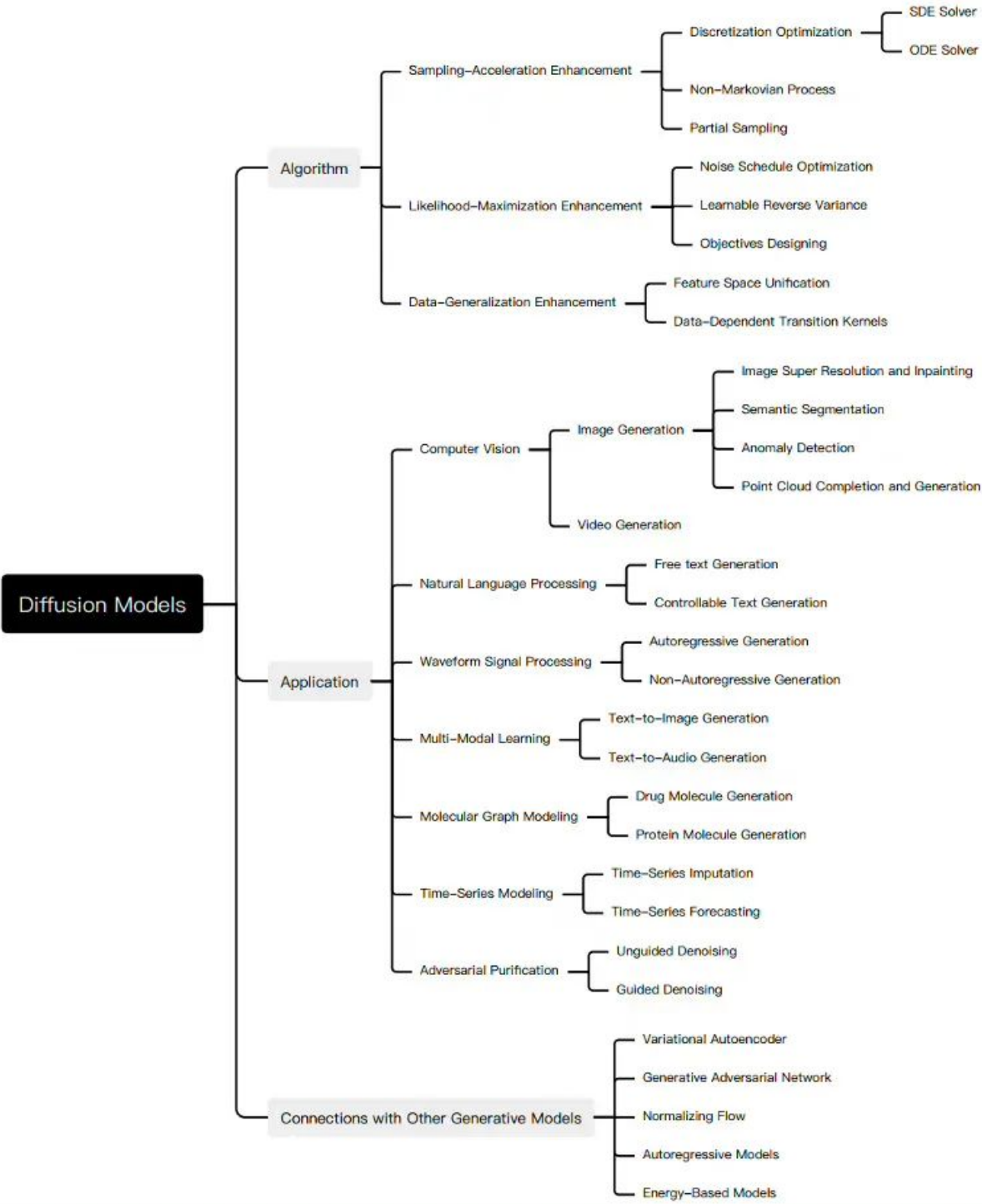
文章链接: <https://arxiv.org/abs/2209.00796>

论文分类汇总 github 链接: <https://github.com/YangLing0818/Diffusion-Models-Papers-Survey-Taxonomy>

## 介绍

扩散模型 (diffusion models) 是深度生成模型中新的 SOTA。扩散模型在图片生成任务中超越了原 SOTA: GAN, 并且在诸多应用领域都有出色的表现, 如计算机视觉, NLP、波形信号处理、多模态建模、分子图建模、时间序列建模、对抗性净化等。此外, 扩散模型与其他研究领域有着密切的联系, 如稳健学习、表示学习、强化学习。然而, 原始的扩散模型也有缺点, 它的采样速度慢, 通常需要数千个评估步骤才能抽取一个样本; 它的最大似然估计无法和基于似然的模型相比; 它泛化到各种数据类型的能力较差。如今很多研究已经从实际应用的角度解决上述限制做出了许多努力, 或从理论角度对模型能力进行了分析。

然而, 现在缺乏对扩散模型从算法到应用的最新进展的系统回顾。为了反映这一快速发展领域的进展, 我们对扩散模型进行了首个全面综述。我们设想我们的工作将阐明扩散模型的设计考虑和先进方法, 展示其在不同领域的应用, 并指出未来的研究方向。此综述的概要如下图所示:



尽管 diffusion model 在各类任务中都有着优秀的表现，它仍还有自己的缺点，并有诸多研究对 diffusion model 进行了改善。为了系统地阐明 diffusion model 的研究进展，我们总结了原始扩散模型的三个主要缺点，采样速度慢，最大化似然差、数据泛化能力弱，并提出将的 diffusion models 改进研究分为对应的三类：采样速度提升、最大似然增强和数据泛化增强。我们首先说明改善的动机，再根据方法的特性将每个改进方向的研究进一步细化分类，从而清楚的展现方法之间的联系与区别。在此我们仅选取部分重要方法为例，我们的工作中对每类方法都做了详细的介绍，内容如图所示：

Table 2. Three types of diffusion models are listed with corresponding articles and years, under continuous and discrete settings.

Category	Methods	Article	Year	Time Setting
Sampling Acceleration	Discretization Optimization	Song et al. [192]	2020	Continuous
		Jolicoeur et al. [96]	2021	Continuous
		Chuang et al. [33]	2022	Continuous
		Dockhorn et al. [49]	2022	Continuous
		Watson et al. [209]	2022	Continuous
		Karras et al. [99]	2022	Continuous
		Lu et al. [128]	2022	Continuous
		Zhang et al. [240]	2022	Continuous
		Bortoli et al. [42]	2022	Continuous
		Liu et al. [124]	2022	Continuous
	Non-Markovian Process	Zhang et al. [241]	2022	Continuous
		Liu et al. [124]	2022	Continuous
		Song et al. [185]	2020	Discrete
		Watson et al. [208]	2021	Discrete
	Partial Sampling	Salimans et al. [175]	2021	Discrete
		Lyu et al. [139]	2021	Discrete
		Zheng et al. [246]	2021	Discrete
		Luhman et al. [130]	2022	Discrete
Likelihood Maximization	Objectives Designing	Song et al.[187]	2021	Continuous
		Huang et al. [87]	2021	Continuous
		Chen et al. [27]	2021	Continuous
		Lu et al. [127]	2022	Continuous
	Noise Schedule Optimization	Nichol et al. [146]	2021	Discrete
		Kingma et al. [106]	2021	Discrete
Data Generalization	Learnable Reverse Variance	Bao et al.[10]	2022	Discrete
		Nichol et al. [146]	2021	Discrete
	Feature Space Unification	Vahdat et al. [200]	2021	Discrete
		Bortoli et al. [19]	2022	Continuous
		Liu et al. [124]	2022	Continuous
		Wehenkel et al. [210]	2021	Discrete
	Data-Dependent Transition Kernels	Niu et al. [150]	2020	Discrete
		Austin et al. [7]	2021	Discrete
		Xie et al. [220]	2022	Discrete
		Gu et al. [71]	2022	Discrete
		Xu et al. [223]	2022	Discrete

在分析完三类扩散模型后，我们将介绍其他的五种生成模型 GAN，VAE，Autoregressive model, Normalizing flow, Energy-based model。考虑到扩散模型的优良性质，研究者们已经根据其特性将 diffusion model 与其他生成模型结合，所以为了进一步展现 diffusion model 的特点和改进工作，我们详细地介绍了 diffusion model 和其他生成模型的结合的工作并阐明了在原始生成模型上的改进之处。Diffusion model 在诸多领域都有着优异的表现，并且考虑到不同领域的应用中 diffusion model 产生了不同的变形，我们系统地介绍了 diffusion model 的应用研究，其中包含如下领域：计算机视觉，NLP、波形信号处理、多模态建模、分子图建模、时间序列建模、对抗性净化。对于每个任务，我们定义了该任务并介绍利用扩散模型处理任务的工作，我们将本项工作的主要贡献总结如下：

- **新的分类方法：**我们对扩散模型和其应用提出了一种新的、系统的分类法。具体的我们将模型分为三类：采样速度增强、最大似然估计增强、数据泛化增强。进一步地，我们将扩散模型的应用分为七类：计算机视觉，NLP、波形信号处理、多模态建模、分子图建模、时间序列建模、对抗性净化。

- **全面的回顾**：我们首次全面地概述了现代扩散模型及其应用。我们展示了每种扩散模型的主要改进，和原始模型进行了必要的比较，并总结了相应的论文。对于扩散模型的每种类型的应用，我们展示了扩散模型要解决的主要问题，并说明它们如何解决这些问题。
- **未来研究方向**：我们对未来研究提出了开放型问题，并对扩散模型在算法和应用方面的未来发展提供了一些建议。

## 扩散模型基础

生成式建模的一个核心问题是模型的灵活性和可计算性之间的权衡。扩散模型的基本思想是正向扩散过程来系统地扰动数据中的分布，然后通过学习反向扩散过程恢复数据的分布，这样就产生了一个高度灵活且易于计算的生成模型。

### A. Denoising Diffusion Probabilistic Models (DDPM)

一个 DDPM 由两个参数化马尔可夫链组成，并使用变分推断以在有限时间后生成与原始数据分布一致的样本。前向链的作用是扰动数据，它根据预先设计的噪声进度向数据逐渐加入高斯噪声，直到数据的分布趋于先验分布，即标准高斯分布。反向链从给定的先验开始并使用参数化的高斯转换核，学习逐步恢复原数据分布。用  $x_0 \sim q(x_0)$  表示原始数据及其分布，则前向链的分布是可由下式表达：

$$q(\mathbf{x}_1, \dots, \mathbf{x}_T | \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1})$$

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

这说明前向链是马尔可夫过程， $\mathbf{x}_t$  是加入  $t$  步噪音后的样本， $\beta_t$  是事先给定的控制噪声进度的参数。当  $\prod_t 1 - \beta_t$  趋于 1 时， $\mathbf{x}_T$  可以近似认为服从标准高斯分布。当  $\beta_t$  很小时，逆向过程的转移核可以近似认为也是高斯的：

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t),$$

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

我们可以将变分下界作为损失函数进行学习：



$$\begin{aligned}
\mathbb{E}[-\log p_{\theta}(\mathbf{x}_0)] &\leq \mathbb{E}_q \left[ -\log \frac{p_{\theta}(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \right] \\
&= \mathbb{E}_q \left[ -\log p(\mathbf{x}_T) - \sum_{t \geq 1} \log \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} \right] \\
&= -L_{\text{VLB}}
\end{aligned}$$

## B.Score-Based Generative Models (SGM)

上述 DDPM 可以视作 SGM 的离散形式。SGM 构造一个随机微分方程 (SDE) 来平滑的扰动数据分布，将原始数据分布转化到已知的先验分布：

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

和一个相应的逆向 SDE，来将先验分布变换回原始数据分布：

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log q_t(\mathbf{x})dt + g(t)d\bar{\mathbf{w}}$$

因此，要逆转扩散过程并生成数据，我们需要的唯一信息就是在每个时间点的分数函数。利用 score-matching 的技巧我们可以通过如下损失函数来学习分数函数：

$$\mathbb{E}_{t, \mathbf{x}_0, \mathbf{x}_t} \left[ \lambda(t) \left\| \mathbf{s}_{\theta}(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log q_{0t}(\mathbf{x}_t | \mathbf{x}_0) \right\|^2 \right]$$

对两种方法的进一步介绍和两者关系的介绍请参见我们的文章。

原始扩散模型的三个主要缺点，采样速度慢，最大化似然差、数据泛化能力弱。最近许多研究都在解决这些缺点，因此我们将改进的扩散模型分为三类：采样速度提升、最大似然增强和数据泛化增强。在接下来的三、四、五节我们将对这三类模型进行详细的介绍。

## 采样加速方法

在应用时，为了让新样本的质量达到最佳，扩散模型往往需要进行成千上万步计算来获取一个新样本。这限制了 diffusion model 的实际应用价值，因为在实际应用时，我们往往需要产生大量的新样本，来为下一步处理提供材料。研究者们为提高 diffusion model 采样速度上进行

了大量的研究。我们对这些研究进行了详细的阐述。我们将其细化分类为三种方法：Discretization Optimization, Non-Markovian Process, Partial Sampling。

**A.Discretization Optimization** 方法优化求解 diffusion SDE 的方法。因为现实中求解复杂 SDE 只能使用离散解来逼近真正的解，所以该类方法试图优化 SDE 的离散化方法，在保证样本质量的同时减少离散步数。SGM 提出了一个通用的方法来求解逆向过程，即对前向和后向过程采取相同的离散方法。如果给定了前向 SDE 的离散方式：

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{f}_i(\mathbf{x}_i) + \mathbf{g}_i \mathbf{z}_i, \quad i = 0, 1, \dots, N-1$$

那么我们就可以以相同的方式离散化逆向 SDE：

$$\mathbf{x}_i = \mathbf{x}_{i+1} - \mathbf{f}_{i+1}(\mathbf{x}_{i+1}) + \mathbf{g}_{i+1} \mathbf{g}_{i+1}^T \mathbf{s}_{\theta^*}(\mathbf{x}_{i+1}, i+1) + \mathbf{g}_{i+1} \mathbf{z}_i,$$

这种方法比朴素 DDPM 效果略好一点。进一步，SGM 向 SDE 求解器中加入了一个矫正器，从而让每一步生成的样本都有正确的分布。在求解的每一步，求解器给出一个样本后，矫正器都使用马尔可夫链蒙特卡罗方法来矫正刚生成的样本的分布。实验表明向求解器中加入矫正器比直接增加求解器的步数效率更高。

**B.Non-Markovian Process** 方法突破了原有 Markovian Process 的限制，其逆过程的每一步可以依赖更多以往的样本来进行预测新样本，所以在步长较大时也能做出较好的预测，从而加速采样过程。其中主要的工作 DDIM，不再假设前向过程是马尔可夫过程，而是服从如下分布：

$$q(\mathbf{x}_1, \dots, \mathbf{x}_T | \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_0)$$

$$q_{\sigma}(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1} | \tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0), \sigma_t^2 \mathbf{I})$$

$$\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) = \sqrt{\alpha_{t-1}} \mathbf{x}_0 + \sqrt{\beta_{n-1} - \sigma_t^2} \cdot \frac{\mathbf{x}_n - \sqrt{\alpha_t} \mathbf{x}_0}{\sqrt{\beta_t}}.$$

DDIM 的采样过程可以视为离散化的神经微分方程，其采样过程更高效，并且支持样本的内插。进一步的研究发现 DDIM 可以视作流形上扩散模型 PNDM 的特例。

**C.Partial Sampling** 方法通过在 generation process 中忽略一部分的时间节点，而只使用剩下的时间节点来生成样本，直接减少了采样时间。例如，Progressive Distillation 从训练好的

扩散模型中蒸馏出效率更高的扩散模型。对于训练好的一个扩散模型，Progressive Distillation 会从新训练一个扩散模型，使新的扩散模型的一步对应于训练好的扩散模型的两步，这样新模型就可以省去老模型一半的采样过程。具体算法如下：

---

## Algorithm 2 Progressive distillation

---

**Require:** Trained teacher model  $\hat{\mathbf{x}}_\eta(\mathbf{z}_t)$   
**Require:** Data set  $\mathcal{D}$   
**Require:** Loss weight function  $w()$   
**Require:** Student sampling steps  $N$   
**for**  $K$  iterations **do**  
     $\theta \leftarrow \eta$   $\triangleright$  Init student from teacher  
    **while** not converged **do**  
         $\mathbf{x} \sim \mathcal{D}$   
         $t = i/N, i \sim \text{Cat}[1, 2, \dots, N]$   
         $\epsilon \sim N(0, I)$   
         $\mathbf{z}_t = \alpha_t \mathbf{x} + \sigma_t \epsilon$   
        # 2 steps of DDIM with teacher  
         $t' = t - 0.5/N, t'' = t - 1/N$   
         $\mathbf{z}_{t'} = \alpha_{t'} \hat{\mathbf{x}}_\eta(\mathbf{z}_t) + \frac{\sigma_{t'}}{\sigma_t} (\mathbf{z}_t - \alpha_t \hat{\mathbf{x}}_\eta(\mathbf{z}_t))$   
         $\mathbf{z}_{t''} = \alpha_{t''} \hat{\mathbf{x}}_\eta(\mathbf{z}_{t'}) + \frac{\sigma_{t''}}{\sigma_{t'}} (\mathbf{z}_{t'} - \alpha_{t'} \hat{\mathbf{x}}_\eta(\mathbf{z}_{t'}))$   
         $\tilde{\mathbf{x}} = \frac{\mathbf{z}_{t''} - (\sigma_{t''}/\sigma_t) \mathbf{z}_t}{\alpha_{t''} - (\sigma_{t''}/\sigma_t) \alpha_t}$   $\triangleright$  Teacher  $\hat{\mathbf{x}}$  target  
         $\lambda_t = \log[\alpha_t^2 / \sigma_t^2]$   
         $L_\theta = w(\lambda_t) \|\tilde{\mathbf{x}} - \hat{\mathbf{x}}_\theta(\mathbf{z}_t)\|_2^2$   
         $\theta \leftarrow \theta - \gamma \nabla_\theta L_\theta$   
    **end while**  
     $\eta \leftarrow \theta$   $\triangleright$  Student becomes next teacher  
     $N \leftarrow N/2$   $\triangleright$  Halve number of sampling steps  
**end for**

---

不断循环这个蒸馏过程就能让采样步骤指数级下降。



扩散模型在最大似然估计的表现差于基于似然函数的生成模型，但最大化似然估计在诸多应用场景都有重要意义，比如图片压缩，半监督学习，对抗性净化。由于对数似然难以直接计算，研究主要集中在优化和分析变分下界（VLB）。我们对提高扩散模型最大似然估计的模型进行了详细的阐述。我们将其细化分类为三类方法：Objectives Designing, Noise Schedule Optimization, Learnable Reverse Variance。

**A.Objectives Designing** 方法利用扩散 SDE 推导出生成数据的对数似然与分数函数匹配的损失函数的关系。这样通过适当设计损失函数，就可以最大化 VLB 和对数似然。Song et al. 证明了可以设计损失函数的权重函数，使得 plug-in reverse SDE 生成样本的似然函数值小于等于损失函数值，即损失函数是似然函数的上界。分数函数拟合的损失函数如下：

$$\mathcal{J}_{\text{DSM}}(\boldsymbol{\theta}; \lambda(\cdot)) := \frac{1}{2} \int_0^T \mathbb{E}_{p(\mathbf{x})p_{0t}(\mathbf{x}'|\mathbf{x})} [\lambda(t) \|\nabla_{\mathbf{x}'} \log p_{0t}(\mathbf{x}' | \mathbf{x}) - \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}', t)\|_2^2] dt.$$

我们只需将权重函数  $\lambda(t)$  设为扩散系数  $g(t)$  即可让损失函数成为似然函数的 VLB，即：

$$-\mathbb{E}_{p(\mathbf{x})} [\log p_{\boldsymbol{\theta}}^{\text{SDE}}(\mathbf{x})] \leq \mathcal{J}_{\text{SM}}(\boldsymbol{\theta}; g(\cdot)^2) + C_1$$

**B.Noise Schedule Optimization** 通过设计或学习前向过程的噪声进度来增大 VLB。VDM 证明了当离散步数接近无穷时，损失函数完全由信噪比函数 SNR (t) 的端点决定：

$$\mathcal{L}_{\infty}(\mathbf{x}) = \frac{1}{2} \mathbb{E}_{\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})} \int_{\text{SNR}_{\min}}^{\text{SNR}_{\max}} \|\mathbf{x} - \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(\mathbf{z}_v, v)\|_2^2 dv,$$

那么在离散步数接近无穷时，可以通过学习信噪比函数 SNR (t) 的端点最优化 VLB，而通过学习信噪比函数中间部分的函数值来实现模型其他方面的改进。

**C.Learnable Reverse Variance** 方法学习反向过程的方差，从而较少拟合误差，可以有效地最大化 VLB。Analytic-DPM 证明，在 DDPM 和 DDIM 中存在反向过程中的最优期望和方差：

**Theorem 1.** (Score representation of the optimal solution to Eq. (4), proof in Appendix A.2)

The optimal solution  $\boldsymbol{\mu}_n^*(\mathbf{x}_n)$  and  $\sigma_n^{*2}$  to Eq. (4) are

$$\begin{aligned} \boldsymbol{\mu}_n^*(\mathbf{x}_n) &= \tilde{\boldsymbol{\mu}}_n \left( \mathbf{x}_n, \frac{1}{\sqrt{\bar{\alpha}_n}} (\mathbf{x}_n + \bar{\beta}_n \nabla_{\mathbf{x}_n} \log q_n(\mathbf{x}_n)) \right), \\ \sigma_n^{*2} &= \lambda_n^2 + \left( \sqrt{\frac{\bar{\beta}_n}{\alpha_n}} - \sqrt{\bar{\beta}_{n-1} - \lambda_n^2} \right)^2 \left( 1 - \bar{\beta}_n \mathbb{E}_{q_n(\mathbf{x}_n)} \frac{\|\nabla_{\mathbf{x}_n} \log q_n(\mathbf{x}_n)\|_2^2}{d} \right), \end{aligned}$$

使用上述公式和训练好的分数函数，在给定前向过程的情况下，最优的 VLB 可以近似达到。

## 数据泛化增强

扩散模型假设数据存在于欧几里得空间，即具有平面几何形状的流形，并添加高斯噪声将不可避免地数据转换为连续状态空间，所以扩散模型最初只能处理图片等连续性数据，直接应用离散数据或其他数据类型的效果较差。这限制了扩散模型的应用场景。数个研究工作将扩散模型推广到适用于其他数据类型的模型，我们对这些方法进行了详细地阐释。我们将其细化分类为两类方法：Feature Space Unification, Data-Dependent Transition Kernels。

**A.Feature Space Unification** 方法将数据转化到统一形式的 latent space，然后在 latent space 上进行扩散。LSGM 提出将数据通过 VAE 框架先转换到连续的 latent space 上后再在其上进行扩散。这个方法的难点在于如何同时训练 VAE 和扩散模型。LSGM 表明由于潜在先验是 intractable 的，分数匹配损失不再适用。LSGM 直接使用 VAE 中传统的损失函数 ELBO 作为损失函数，并导出了 ELBO 和分数匹配的关系：

$$CE(q(\mathbf{z}_0|\mathbf{x})||p(\mathbf{z}_0)) = \mathbb{E}_{t \sim \mathcal{U}[0,1]} \left[ \frac{g(t)^2}{2} \mathbb{E}_{q(\mathbf{z}_t, \mathbf{z}_0|\mathbf{x})} \left[ \|\nabla_{\mathbf{z}_t} \log q(\mathbf{z}_t|\mathbf{z}_0) - \nabla_{\mathbf{z}_t} \log p(\mathbf{z}_t)\|_2^2 \right] \right]$$

该式在忽略常数的意义下成立。通过参数化扩散过程中样本的分数函数，LSGM 可以高效的学习和优化 ELBO。

**B.Data-Dependent Transition Kernels** 方法根据数据类型的特点设计 diffusion process 中的 transition kernels，使扩散模型可以直接应用于特定的数据类型。D3PM 为离散型数据设计了 transition kernel，可以设为 lazy random-walk, absorbing state 等。GEODIFF 为 3D 分子图数据设计了平移 - 旋转不变的图神经网络，并且证明了具有不变性的初分布和 transition kernel 可以导出具有不变性的边缘分布。假设  $T$  是一个平移 - 旋转变换，如：

$$p(\mathbf{x}_T) = p(\mathcal{T}(\mathbf{x}_T)), p(\mathbf{x}_{t-1}|\mathbf{x}_t) = p(\mathcal{T}(\mathbf{x}_{t-1})|\mathcal{T}(\mathbf{x}_t)),$$

那么生成的样本分布也有平移 - 旋转不变性：

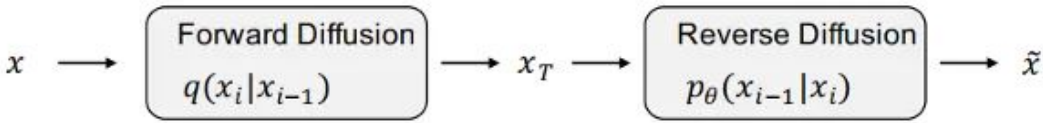
$$p_0(\mathbf{x}) = p_0(\mathcal{T}(\mathbf{x}))$$

## 和其他生成模型的联系

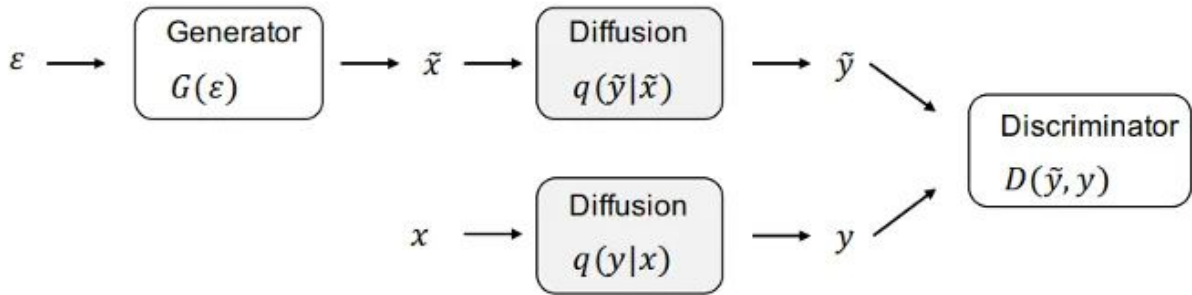
在下面的每个小节中，我们首先介绍其他五类重要的生成模型，并分析它们的优势和局限性。然后我们介绍了扩散模型是如何与它们联系起来的，并说明通过结合扩散模型来改进这些生成

模型。VAE，GAN，Autoregressive model, Normalizing flow, Energy-based model 和扩散模型的联系如下图所示：

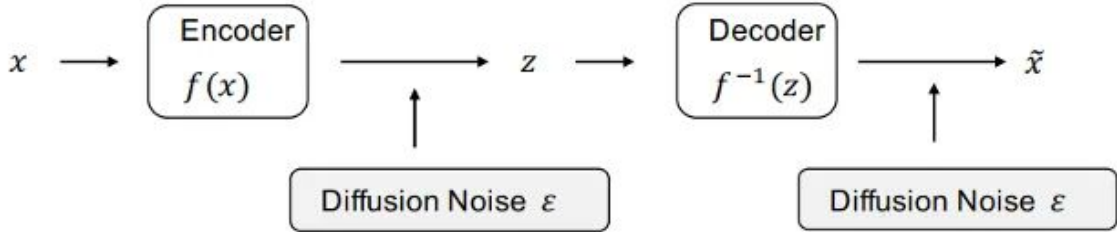
**Variational AutoEncoder**



**Generative Adversarial Network**



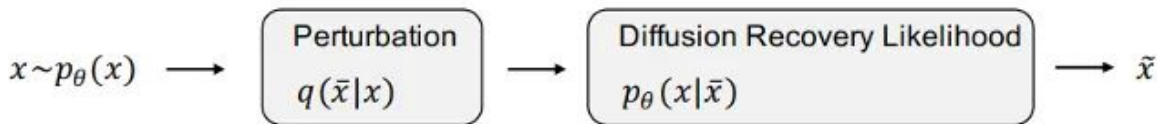
**Normalizing Flow**



**Autoregressive Model**



**Energy-based Model**



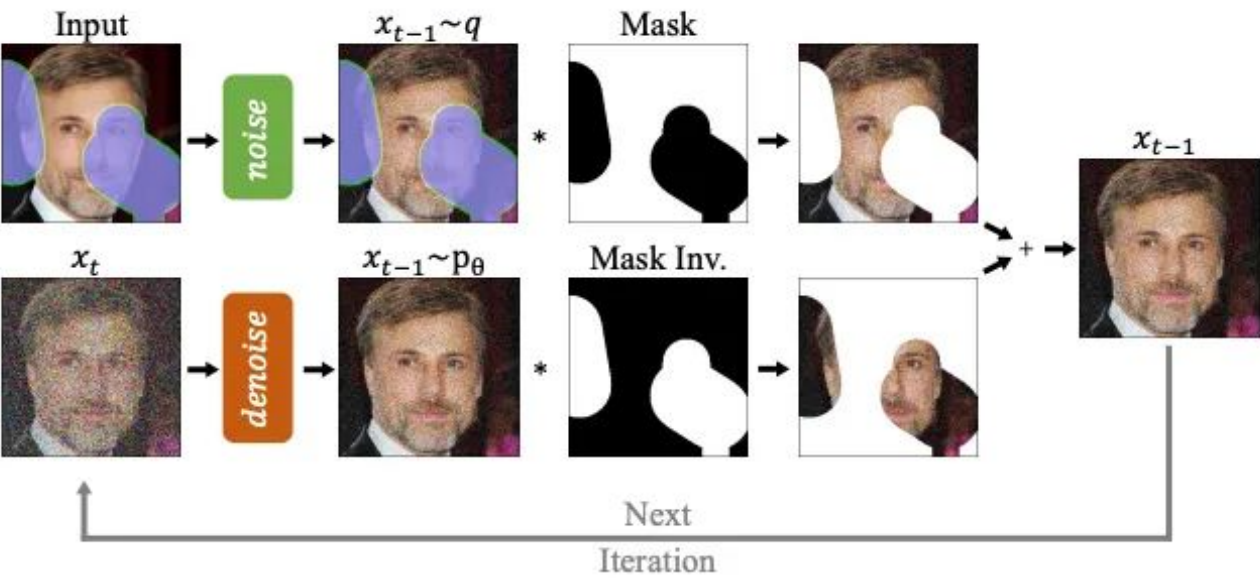
- 1. DDPM 可以视作层次马尔可夫 VAE（hierarchical Markovian VAE）。但 DDPM 和一般的 VAE 也有区别。DDPM 作为 VAE，它的 encoder 和 decoder 都服从高斯分布、有马尔科夫行；其隐变量的维数和数据维数相同；decoder 的所有层都共用一个神经网络。
- 2. DDPM 可以帮助 GAN 解决训练不稳定的问题。因为数据是在高维空间中的低维流形中，所以 GAN 生成数据的分布和真实数据的分布重合度低，导致训练不稳定。扩散模

型提供了一个系统地增加噪音的过程，通过扩散模型向生成的数据和真实数据添加噪音，然后将加入噪音的数据送入判别器，这样可以高效地解决 GAN 无法训练、训练不稳定的问题。

3. Normalizing flow 通过双射函数将数据转换到先验分布，这样的作法限制了 Normalizing flow 的表达能力，导致应用效果较差。类比扩散模型向 encoder 中加入噪声，可以增加 Normalizing flow 的表达能力，而从另一个视角看，这样的做法是将扩散模型推广到前向过程也可学习的模型。
4. Autoregressive model 在需要保证数据有一定的结构，这导致设计和参数化自回归模型非常困难。扩散模型的训练启发了自回归模型的训练，通过特定的训练方式避免了设计的困难。
5. Energy-based model 直接对原始数据的分布建模，但直接建模导致学习和采样都比较困难。通过使用扩散恢复似然，模型可以先对样本加入微小的噪声，再从有略微噪声的样本分布来推断原始样本的分布，使的学习和采样过程更简单和稳定。

扩散模型的应用

在本节中，我们分别介绍了扩散模型在计算机视觉、自然语言处理、波形信号处理、多模态学习、分子图生成、时间序列以及对抗学习等七大应用方向中的应用，并对每类应用中的方法进行了细分并解析。例如在计算机视觉中可以用 diffusion model 进行图像补全修复（RePaint）：

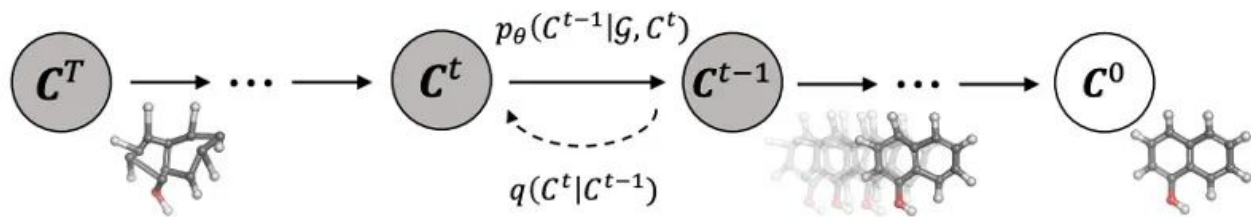


在多模态任务中可以用 diffusion model 进行文本到图像的生成（GLIDE）：





还可以在分子图生成中用 diffusion model 进行药物分子和蛋白质分子的生成（GeoDiff）：



应用分类汇总见表：

Primary Category	Secondary Category	Article
Computer Vision	Image Super Resolution and Inpainting	[4],[128],[129],[130],[131],[132],[133],[134],[135],[136],[137]
	Semantic Segmentation	[5],[138]
	Anomaly Detection	[139],[140]
	Point Cloud Completion and Generation	[141],[142],[143],[144]
	Video Generation	[145],[146],[147],[148]
Natural Language Processing	Natural Language Processing	[7],[149],[150]
Waveform Signal Processing	Waveform Signal Processing	[8],[9]
Multi-Modal Learning	Text-to-Image Generation	[10],[151],[11],[152],[58],[153],[154]
	Text-to-Audio Generation	[155],[156],[157],[158],[159],[160],[161]
Molecular Graph Modeling	Molecular Graph Modeling	[13],[14],[16],[59],[162],[163],[164],[165]
Time Series Modeling	Time Series Imputation	[166],[167],[168]
	Time Series Forecasting	[169],[167]
Adversarial Purification	Adversarial Purification	[170],[171],[18],[172],[173],[174]

未来研究方向

1. 应用假设再检验。我们需要检查我们在应用中普遍接受的假设。例如，实践中普遍认为扩散模型的前向过程会将数据转换为标准高斯分布，但事实并非如此，更多的前向扩散步骤会使最终的样本分布与标准高斯分布更接近，与采样过程一致；但更多的前向扩散步骤也会使估计分数函数更加困难。理论的条件很难获得，因此在实践中操作中会导致理论和实践的不匹配。我们应该意识到这种情况并设计适当的扩散模型。
2. 从离散时间到连续时间。由于扩散模型的灵活性，许多经验方法可以通过进一步分析得到加强。通过将离散时间的模型转化到对应的连续时间模型，然后再设计更多、更好的离散方法，这样的研究思路有前景。



3. 新的生成过程。扩散模型通过两种主要方法生成样本：一是离散化反向扩散 SDE，然后通过离散的反向 SDE 生成样本；另一个是使用逆过程中马尔可夫性质对样本逐步去噪。然而，对于一些任务，在实践中很难应用这些方法来生成样本。因此，需要进一步研究新的生成过程和视角。
4. 泛化到更复杂的场景和更多的研究领域。虽然目前 diffusion model 已经应用到多个场景中，但是大多数局限于单输入单输出的场景，将来可以考虑将其应用到更复杂的场景，比如 text-to-audiovisual speech synthesis。也可以考虑和更多的研究领域相结合。

公众号后台回复“**开学**”获取CVPR、ECCV、valse等资源汇总～

ACCV 2022

国际细粒度图像分析挑战赛

ACCV 2022 Fine-grained Image Analysis Challenge

🕒 2022.09.08-2022.11.08





标注数据集开放下载

国际型赛事

220G

.....

01 赛 题

网络监督的细粒度识别

02 赛 题

大规模细粒度哈希检索

评审嘉宾



魏秀参

南京理工大学

计算机科学与工程学院教授



王 鹏

University of Wollongong

计算机与信息技术学院助理教授



王 欢

南京理工大学

计算机科学与工程学院副教授



即刻扫码报名

主办单位



南京理工大学



University of Wollongong



技术支持：极市平台

https://mp.weixin.qq.com/s?\_\_biz=MzI5MDUyMDIxNA==&mid=2247621314&idx=4&sn=f04bc091a077d0bedda4ecfc59e6dab5&chksm=ec1df73bdbb6a7e... 15/18

**极市平台**

为计算机视觉开发者提供全流程算法开发训练平台，以及大咖技术分享、社区交流、竞赛...

848篇原创内容

公众号

△点击卡片关注极市平台，获取最新CV干货

## 极市干货

**算法竞赛：国际赛事证书，220G数据集开放下载！ACCV2022国际细粒度图像分析挑战赛开赛！**

**技术综述：浅聊对比学习（Contrastive Learning） | 深度学习图像分类任务中那些不得不看的11个tricks总结**

**极视角动态：极视角与华为联合发布基于昇腾AI的「AICE赋能行业解决方案」 | 算法误报怎么办？自训练工具使得算法迭代效率提升50%！**

# CV技术社群邀请函 #

△长按添加极市小助手

添加极市小助手微信（ID：cvmart2）

备注：姓名-学校/公司-研究方向-城市（如：小极-北大-目标检测-深圳）

即可申请加入极市目标检测/图像分割/工业检测/人脸/医学影像/3D/SLAM/自动驾驶/超分辨率/姿态估计/ReID/GAN/图像增强/OCR/视频理解等技术交流群

极市&深大CV技术交流群已创建，欢迎深大校友加入，在群内自由交流学术心得，分享学术讯息，共建良好的技术交流氛围。



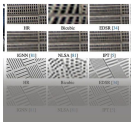
点击阅读原文进入CV社区  
收获更多技术干货

阅读原文

喜欢此内容的人还喜欢

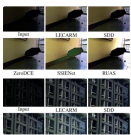
ICCV 2023 | 南开程明明团队提出适用于SR任务的新颖注意力机制（已开源）

极市平台



ICCV23 | 将隐式神经表征用于低光增强，北大张健团队提出NeRCo

极市平台



YOLOv5帮助母猪产仔？南京农业大学研发母猪产仔检测模型并部署到Jetson Nano开发板



