**Title:** DIFFUSION MODELS: A COMPREHENSIVE SURVEY OF METHODS AND APPLICATIONS

**Abstract:**

扩散模型这一深度生成模型，在密集理论基础的各种任务上表现很好，但是它有昂贵的采样过程和次优似然估计。最近的研究对提高扩散模型的性能有巨大热情。本文对现有扩散模型进行全面回顾，并将其分成三类。本文还阐述了扩散模型和其他五种生成模型之间的关系（即变分自编码器、生成对抗网络、标准化流、自回归模型和基于能量的模型），还对扩散模型的应用进行了描述：包括计算机视觉、自然语言处理、波形信号处理、多模态建模、分子图生成、时间序列建模和对抗性净化。此外，本文提出了关于这种生成模型发展的新视角。

**Intro：**

扩散模型是一种很有前途的算法。然而，原始的扩散模型仍然存在着采样过程缓慢的问题，通常需要数千步的评估才能绘制一个样本。而且，它在实现与其他基于似然的模型（例如自回归模型）竞争的对数似然方面一直都很困难。最近的研究从实际考虑或从理论上分析模型容量来提高扩散模型的性能。本文对扩散模型进行了第一次全面的回顾，阐明扩散模型的设计考虑因素和先进方法，介绍其在不同领域的应用，并指出未来研究的方向。如图1所示。

扩散模型这类模型由两部分构成，首先是正向过程，通过在多个尺度上添加噪声来逐步干扰数据分布，然后是反向过程，学习恢复数据结构。扩散模型的过程可以被视为随机微分方程（SDE）的离散化，其中正向过程和反向过程对应于正向SDE和反向SDE。因此，通过SDE对扩散模型进行分析可以提供密集的理论结果和模型改进，特别是在采样策略方面。基于这些观点，本文将扩散模型分为三类：采样过程增强（第3节）、似然最大化增强（第4节）和泛化能力增强（第5节）。在每个类别中，本文分析了离散时间设置和连续时间设置下的模型，这两种设置在经验和理论结果方面都很丰富。

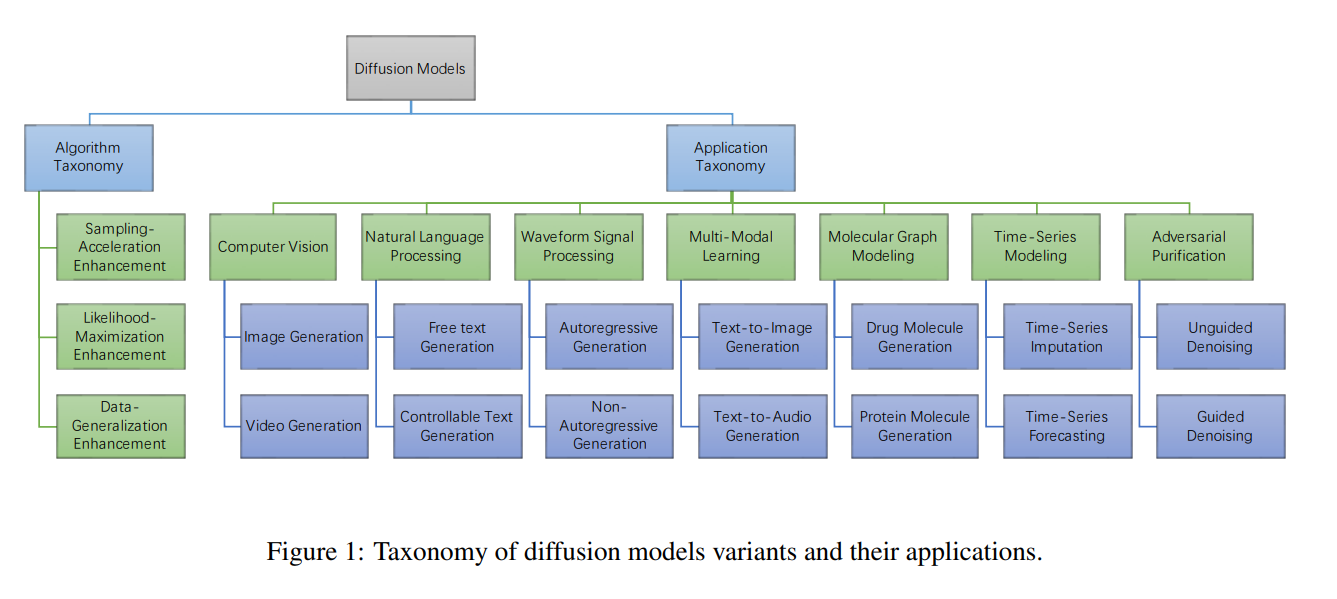
在对三种扩散模型进行分析后，本文介绍了其他五种常用的生成模型（第6节），即变分自编码器（VAE）、生成对抗网络（GAN）、标准化流（NF）、自回归模型（AR）和基于能量的模型（EBM）。由于扩散模型的良好性质，研究人员开始将扩散模型与这些传统的生成模型相结合。本文对这些组合工作进行了具体的介绍，并阐明了对原始生成模型的改进。然后，本文系统地介绍了扩散模型在广泛任务（第7节）中的应用，包括计算机视觉、自然语言处理、波形信号处理、多模态建模、分子图生成、时间序列建模和对抗性净化。对于每个任务介绍了利用扩散模型来处理问题的研究工作。在第8节中，本文提出了这个快速发展的领域的潜在研究方向，并在第9节中总结了这项调查。

本文的主要贡献有：

新分类法；综合评述；未来的研究方向

**Conclusion：**

在本文中，首先提出了一个新的、系统的扩散模型及其应用的分类法。具体来说，将有的扩散模型分为三类：采样加速、似然最大化和数据泛化，这是针对模型增强的不同方面进行的分类。此外，扩散模型的应用分为七类：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、时间序列建模和对抗性净化。最后，本文提出了关于扩散模型在算法和应用方面发展的新观点。

**Figure：**

扩散模型变体的分类及其应用

**Preliminaries of Diffusion Models：**

生成模型的一个核心问题是模型概率分布在灵活性和可处理性之间的权衡。扩散模型的基本思想是通过正向扩散过程系统地扰动数据分布的结构，然后通过学习反向扩散过程来恢复结构，从而得到高度灵活和易于处理的生成模型。

本节说明了生成过程在DDPM（A denoising diffusion probabilistic model）中是反向SDE的特殊离散化。

**Diffusion Models with Sampling-Acceleration Enhancement：**

尽管扩散模型在生成高质量和多样化的样本方面取得了巨大的成功，但它们的采样过程非常耗时，通常需要数千次离散化步骤来学习扩散过程，以达到所需的精度。

本节讨论了可以显著减少学习数量步骤的方法

Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12413–12422, 2022

因此，人们确实可以选择一个最大化VLB的最优轨迹。其他方法要么将正向过程推广为非马尔可夫过程，从而导致确定性生成过程，要么重新训练可以通过知识提炼走捷径的学生网络。

**Diffusion Models with Likelihood-Maximization Enhancement：**

去噪扩散概率模型（DDPM）与其他基于似然的模型相比，不具有竞争力的对数似然性，最近有多种方法可以增强对数似然性的最大化。由于直接计算对数似然的困难性，研究主要集中在设计和分析变分下界。

**Diffusion Models with Data-Generalization Enhancement：**

扩散模型假设数据在欧几里得空间上得到支持，即具有平坦几何结构的流形，并添加高斯噪声将不可避免地将数据转换为连续状态空间。各种研究都集中在解决这些限制上。通过首先将数据编码到潜在空间，然后将扩散模型应用于该空间，依赖于VAE框架，扩散模型可以应用于各种数据类型。这里的关键问题是如何设计损失函数来联合学习VAE和扩散模型，因为潜在变量的先验分布应该既灵活又可处理。

**Connections with Other Generative Models：**

生成模型是一个热门的研究领域，有许多应用。例如，它们可以用来生成高质量的图像，合成逼真的语音和音乐，推进半监督学习，识别对抗性示例，进行模仿学习，以及在强化学习中进行优化。在每个小节中，我们首先介绍其他五类重要的生成模型，并分析它们的优点和局限性。本文介绍扩散模型是如何与它们相联系的，并说明这些生成模型是如何通过结合扩散模型得到提升的。

1. Variational Autoencoder and Its Connections with Diffusion Models

变分自编码器学习一个编码器和一个解码器，将输入数据映射到连续潜在空间中的值。

DDPM可以被视为具有前缀编码器的分层马尔可夫VAE。正向过程表示编码器，反向过程表示解码器。此外，DDPM在多个层上共享解码器，并且所有潜在变量都与样本数据的大小相同。在连续时间设置中，优化扩散模型可以被视为训练一个无限的、深度的、分层的VAE，这证明了扩散模型可以被解释为分层VAE的连续极限的常见认知。

2. Generative Adversarial Network and Its Connections with Diffusion Models

生成对抗网络GANs主要由两个模型组成：生成器G和鉴别器D。这两个模型通常由神经网络构建，但也可以用任何形式的可微分系统实现，该系统将输入数据从一个空间映射到另一个空间。生成器G试图模拟真实例子的分布，并生成新的例子。鉴别器D通常是一个二元分类器，用于以尽可能高的精度从真实例子中识别生成的例子。

GAN的主要实际问题之一是训练过程的不稳定性，这主要是由于输入数据的分布和生成数据的分布之间的不重叠所导致的。一个实际的解决方案是向鉴别器输入中注入噪声。利用灵活的扩散模型，向鉴别器中注入噪声，其噪声调度是自适应的，由扩散模型确定。反过来，GAN可以帮助提高扩散模型的采样速度。慢采样是由去噪步骤中的高斯假设引起的，这只对小步长是合理的。因此，他们提出使用条件GAN对每个去噪步骤进行建模，从而允许更大的步长。

**A Collection of Applications with Diffusion Models：**

由于扩散模型的灵活性和强大性，它们最近被应用于许多现实世界的应用中。

1. Image Generation

1）图像超分辨率和修复：

图像超分辨率旨在从低分辨率（LR）图像中恢复高分辨率（HR）图像，而图像修复则是重建图像中缺失或损坏的区域。

超分辨率扩散（SRDiff）是第一个基于扩散的单图像超分辨率模型，通过数据似然的变分界限进行优化。

通过重复细化进行超分辨率（SR3）采用去噪扩散概率模型DDPM进行条件图像生成，并通过随机迭代去噪过程进行图像超分辨率。

LDM提出了潜在扩散模型，这是一种在不损失质量的情况下提高去噪扩散模型的训练和采样效率的有效方法。

RePaint设计了一种改进的去噪策略，通过重新采样迭代来更好地调节图像。

Palette提出了一个基于条件扩散模型的统一框架，并在四个具有挑战性的图像生成任务上评估了这个框架，例如着色、修复、未裁剪和JPEG恢复。

级联扩散模型（CDM）由级联的多个扩散模型组成，这些模型生成逐渐增加的分辨率的图像。CDM能够在类条件ImageNet[123]生成基准数据集上生成高质量的图像，无需任何来自辅助图像分类器的监督信息。

多速扩散（MSDiff）产生了一个条件多速扩散估计器（CMDE），这是一个条件得分的估计器，它结合了以前的条件得分估计方法。

2）semantic segmentation

语义分割是将图像中属于同一类别的部分聚集在一起。预训练可以提高语义分割模型的标签利用率，而预训练语义分割模型的另一种方法是生成建模。

最近研究表明，DDPM具有捕获对下游视觉任务有价值的高级语义信息的能力。

3.）Anomaly Detection.

异常检测是机器学习和计算机视觉中的一个关键且具有挑战性的问题。生成模型已经被证明拥有强大的异常检测机制，包括GAN、VAE和扩散模型。

AnoDDPM 提出了一种新的异常检测方法，该方法利用DDPM破坏输入图像并重建健康的图像近似。这种方法优于对抗性训练，因为它可以更好地对较小的数据集进行建模，具有先进的样本质量和更稳定的训练。

DDPM-CD 提出了一种新的方法，通过DDPM将大量无监督的遥感图像纳入训练过程中。它利用预训练的DDPM和应用来自扩散模型解码器的多尺度表示来进行遥感变化检测。它的目标是训练一个轻量级的变化检测分类器，以有效地检测精确的变化。

4）Point Cloud Completion and Generation.

3D点云是捕捉现实世界3D对象的关键形式。然而，由于部分观察和自我遮挡，现实世界中扫描的点云通常是不完整的。许多下游任务，包括3D重建、增强现实（AR）和场景理解通过推理缺失部分来恢复完整形状。

PVD将去噪扩散模型与3D形状的点体素表示相结合。受非平衡热力学中扩散过程的启发，将点云中的点视为与热浴相连的热力学系统中的粒子，这使得从原始分布向噪声分布的扩散。

点扩散优化（PDR）利用条件DDPM生成给定部分观察的粗糙补全，并在生成的点云和统一的地面真相之间构建1对1的点对点映射。然后，它优化均方误差损失以实现均匀生成。

2. Video Generation

最近的研究采用扩散模型来提高生成视频的质量。

灵活扩散模型（FDM）提出了一种基于DDPM的新视频生成框架，该框架在不同现实场景中生成长期的视频补全。它引入了一个生成模型，该模型可以在测试期间对视频帧的任意子集进行采样，条件是其他子集，并提出了一个为此目的设计的架构。

残差视频扩散（RVD）提出了一个自回归的、端到端优化的视频扩散模型。它通过用随机残差校正确定性的下一帧预测来连续生成未来的帧，该随机残差是由逆扩散过程生成的。

视频扩散模型（VDM）为时空视频扩展引入了一种条件采样方法。它优于以前提出的方法，并生成了时间长、分辨率更高的视频。

**Multi-Modal Learning：**

1. Text-to-Image Generation

文本到图像生成是从描述性文本生成相应图像的任务。

混合扩散利用预训练的DDPM和CLIP模型，首次提出了基于区域的图像编辑解决方案，该方案使用自然语言指导，适用于真实和多样化的图像。

unCLIP提出了一个两阶段模型，一个先验模型，可以生成基于CLIP的图像嵌入，条件是一个文本标题，以及一个基于扩散的解码器，可以生成基于图像嵌入的图像。

Imagen提出了一个具有更好语言理解和前所未有逼真度的文本到图像扩散模型。

GLIDE将引导扩散应用于文本条件图像合成。

VQ-Diffusion提出了一种用于文本到图像生成的向量量化扩散模型，它消除了单向偏差并避免了累积预测误差。

2. Text-to-Audio Generation

文本到音频生成是将普通语言文本转换为语音输出的任务。

Grad-TTS 提出了一种新型的文本到语音模型，该模型具有基于分数的解码器和扩散模型。它逐渐变换编码器预测的噪声，并通过单调对齐搜索的方法与文本输入进一步对齐。

Diffsound 提出了一种基于离散扩散模型的非自回归解码器，该解码器首先在每个单步中预测所有的梅尔频谱图标记，然后在后续步骤中细化预测的标记。

**Future Directions：**

1. Towards New Perspectives.

我们观察到离散扩散模型中仍然存在未解决的问题，这些问题在NLP中显示出实际价值。数据的离散性质使得用连续高斯噪声恢复损坏的数据非常困难。同样的问题也存在于其他数据类型，如图形。因此，需要新的方法和视角。在理论层面上，我们仍然需要检查扩散模型中的一些普遍相信的前提。例如，在实践中人们普遍认为前向过程会将数据转换为标准高斯噪声。然而，SDE的有限时间解不能忘记原始数据的分布。这些实践和理论之间的不匹配可以激发更好的模型设计。在实际层面上，由于扩散模型的灵活性，有许多经验方法需要进一步评估和分析它们的泛化能力。

2. Generalizing to More Applications.

扩散模型已被应用于七种不同类型的场景。然而，仍有一些场景尚未得到充分探索，例如文本到视听语音合成和视觉问题回答（VQA）。此外，我们可以明显地发现，大多数现有应用仅限于单一输入/输出或简单输入/输出。因此，研究人员使扩散模型能够处理复杂的输入并生成多个输出，并在现实场景中获得更好的性能，这是具有挑战性的。虽然扩散模型已经在一些研究领域得到了研究，如鲁棒学习、表示学习和强化学习，但仍有机会与更多的研究领域相结合。