



浙江工商大學

# 硕士学位论文

论文题目：基于扩散模型的三维药物分子设计框架

作者姓名：徐璨

学科专业：统计学

研究方向：数理统计

指导教师：王伟刚

提交日期：2024 年 1 月

**Dissertation Submitted to Zhejiang Gongshang University  
for Master's Degree of Science**

**A Diffusion-based 3D Molecule Generative Framework**

**Author:** Can Xu

**Major:** Statistics

**Supervisor:** Prof. Weigang Wang



**Jan. 2024**

**School of Statistics and Mathematics**

**Zhejiang Gongshang University**

**Hangzhou, 310018, P. R. China**

---

## 摘 要

去噪扩散模型在多个研究领域显示出巨大潜力。现有的基于扩散生成模型在全新的三维分子生成任务中面临两个主要挑战。由于分子中的大部分重原子能够通过单键与多个原子相连，使用成对距离来建模分子几何结构是不足够的。因此，第一个挑战是提出一个有效的神经网络作为去噪核，能够捕捉复杂的原子间关系并学习高质量的特征。此外，由于图的离散性质，当前主流的基于扩散分子生成模型严重依赖预设的规则，并以间接的方式生成边。故第二个挑战是将分子生成的过程与扩散的学习过程相结合，有效地准确预测键的存在。本文认为扩散过程中分子构象的迭代更新方式与分子动力学一致，故提出一种名为几何辅助的分子扩散 (Geometric-facilitated Molecular Diffusion / GFMDiff) 的新型分子生成框架。针对第一个挑战，我们引入了一种双轨 Transformer 网络 (Dual-track Transformer Network / DTN)，以全面挖掘全局空间关系并学习高质量的表示，从而提高对特征和几何结构的准确预测能力。至于第二个挑战，我们设计了几何辅助损失 (Geometric-facilitated Loss / GFLoss)，该损失在训练期间干预键的形成，而不是直接将边嵌入隐变量空间。本文在多个基准数据集上开展了全面实验表明 GFMDiff 性能达到该领域时下最优的水平。

关键词: 扩散模型; 分子生成; 分子学习; 图神经网络; 几何神经网络

---

## Abstract

Denoising diffusion models have shown great potential in multiple research areas. Existing diffusion-based generative methods on *de novo* 3D molecule generation face two major challenges. Since majority heavy atoms in molecules allow connections to multiple atoms through single bonds, modeling molecule geometries using pair-wise distance is insufficient. Therefore, the first one involves proposing an effective neural network as the denoising kernel that is capable to capture complex interatomic relationships and learn high-quality features. Due to the discrete nature of graphs, mainstream diffusion-based methods for molecules heavily rely on predefined rules and generate edges in an indirect manner. The second one involves accommodating molecule generation to the learning process of diffusion and accurately predicting the existence of bonds effectively. In our research, we view the iterative way of updating molecule conformations in diffusion process is consistent with molecular dynamics and introduce a novel molecule generation method named Geometric-Facilitated Molecular Diffusion (GFMDiff). For the first challenge, we introduce a Dual-track Transformer Network (DTN) to fully excavate global spatial relationships and learn high quality representations which contribute to accurate predictions of features and geometries. As for the second challenge, we design Geometric-facilitated Loss (GFLoss) which intervenes the formation of bonds during the training period, instead of directly embedding edges into the latent space. Comprehensive experiments on current benchmarks demonstrate the superiority of GFMDiff.

KEYWORDS: Diffusion models; Molecule generation; Molecular learning; Graph neural networks; Geometry neural Networks

---

# 目录

---

# 第1章 引言

## 1.1 选题背景与研究意义

### 1.1.1 选题背景

除图像、视频，音频与自然语言处理等领域，AI 技术的快速发展也带动相关交叉学科的发展。AI4Science 近年在计算生物、计算化学、材料设计、计算天文，计算育种等都有广泛应用，相关 AI 技术的应用能够大幅加速相关科学研究进展。在计算制药领域，近年来相关 AI 技术在药物性质预测，生成，开发，实验等领域的运用不仅加速相关研究的进展，也能够降低相关研究的研发成本。

基于 AI 的生成模型近十年来也被广泛研究，他们包括变分自编码器 (Variational autoencoders / VAEs)<sup>[1]</sup>，生成对抗模型 (Generative adversarial networks / GANs)<sup>[2]</sup>，流形模型 (Normalizing flows / NFs)<sup>[3-4]</sup>，自回归 (Autoregressive models / ARs)<sup>[5]</sup> 与扩散模型 (Diffusion)<sup>[6-7]</sup> 等。相关方法在图像，文字等方面也有了许多成功应用。

深度学习在分子化学领域近年来也有着成功的应用。以分子学习为例，化学分子常以简化分子线性输入规范字符串 (Simplified molecular-input line-entry system / SMILES)<sup>[8]</sup> 存储，每一个分子式对应一个 SMILES 字符串。随着早期深度学习方法，如卷积神经网络 (Convolutional neural networks / CNNs)<sup>[9]</sup> 和循环神经网络 (Recurrent neural networks / RNNs)<sup>[10-12]</sup> 的发展，一些研究试图运用这些深度学习算法，对以字符串形式存在的分子式进行学习，以获得预测特定原子或是分子整体的性质的能力。随着图神经网络 (Graph neural networks / GNNs)<sup>[13-15]</sup> 的出现，其对非结构化数据的建模能力和对节点间拓扑关系学习的能力被证明十分优异。分子作为自然界中天然存在的图结构，原子和键对应着图中的节点和边，这为分子学习提供了新的思路与方法。从最早的图卷积神经网络开始，相关研究者致力于提出新的图学习算法，以提升对分子图学习的性能。随着相关化学模拟技术的发展，让三维分子建模成为可能。这也在拓扑结构信息以外，提供了更丰富的几何构型信息，这也驱动着相关研究拓展至几何图神经网络上。分子三维构象允许研究者对分子进行更准确的研究，同时也推动更多复杂任务的出现，包括分子生成，Ligand 生成，Protacs 生成，药物亲和力预测，蛋白质预测等等。

### 1.1.2 研究意义

人工智能在智能计算相关研究中开始扮演越来越重要的角色，相关模型在药物发现、药物属性预测等应用中已经展现出良好的性能和极大的潜力。由于深度学习技术应用具备为药物研发的多阶段降本增效的潜力，在 2022 年，AI 制药赛道相关企业融资总金额达百亿美元。在这一赛道竞逐的有国内互联网巨头如百度百图生科、华为 EIHealth、腾讯云深智药，及初创企业晶泰科技，剂泰医药，星药科技等。相关成果已经展现出深度学习在该领域的强大性能和广阔前景。

## 1.2 文献综述

### 1.2.1 基于深度学习的分子学习

分子最早被表示为简化分子线性输入规范字符串 (SMILES)<sup>[8]</sup>，随着早期深度学习模型卷积神经网络 (CNN) 和循环神经网络 (RNN) 的发展，相关模型利用 CNNs 和 RNNs 对分子性质做出学习。Hirohara 等<sup>[9]</sup>的研究提出使用 CNN 对分子级别的特征和分子基团性质进行有效学习。由于 RNNs 在早期自然语言护理任务上有良好表现，Bjerrum<sup>[10]</sup>提出 LSTM-QSAR 模型用于学习分子性质。

伴随图神经网络 (GNNs) 的发展，由于分子的形式天然的属于图结构，一些研究开始使用 GNNs 进行分子学习。CGCNN<sup>[16]</sup>将图卷积神经网络引入到分子性质学习，用以模拟并替代复杂的 DFT 计算。Xiong 等<sup>[17]</sup>提出 Attentive FP，一个结合注意力机制的图神经网络实现分子性质的有效学习与预测。GraSeq<sup>[18]</sup>提出在运用图神经网络学习拓扑结构同时，利用双向 LSTM 模型对 SMILES 分子式进行学习，通过两个通道联合预测分子性质。伴随相关技术的发展，研究者可以不再拘泥于原子拓扑结构，进而实现对分子三维结构的建模与学习。鉴于某一分子对应大量的同分异构体，而不同构象对应属性不尽相同，因此对三维构象的有效学习是十分必要的。EGNN<sup>[19]</sup>在流行的图网络基础上，保留最基本的几何信息，即原子间距离，其简单的设计也成为了一些。在分子预训练框架 GEM<sup>[20]</sup>中，研究者提出 GeoGNN 图网络，将键长视作原子节点图的边特征，又对原子键构图，并将键角作为原子键图的边特征，通过在两个网络上的信息传递实现分子局部空间几何性质学习。Schütt 等<sup>[21]</sup>在继承 GNNs 的信息传递范式的同时，将原子间距离用径向基函数 (Radial basis function / RBF) 建模后融入边特征，使算法对三维几何信息的有效学习的同时保证了等变性要求。SphereNet<sup>[22]</sup>提出了基于球坐标系的信息传递范式 SMP。通过一系列参考原子或键的规则，SMP 在保证对原子

对距离，键角和键扭转角这三个空间几何信息完整提取的同时，避免计算复杂度的爆炸式增长。ComENet<sup>[23]</sup>在 ShpereNet 的基础上，简化了空间几何信息的提取范式，在保证利用完整空间信息的前提下，以损失部分精度为代价，大幅度提升计算速度。

伴随着 Transformer<sup>[24]</sup> 相关研究在图像与文本领域的兴起，相关研究<sup>[25-27]</sup> 也利用 Transformer 对 SMILES 分子式进行学习。随着 GTN<sup>[28]</sup> 将 Transformer 引入图学习，越来越多的研究也将多头注意力机制用于分子图学习领域。大规模分子图预训练框架 GROVER<sup>[29]</sup> 中，分子学习内核运用了 GTransformer，同时学习分子中的原子与键的节点嵌入 (embedding) 或边嵌入。在分子预训练框架 MPG<sup>[30]</sup> 中提出的图学习内核 MolGNet 放弃了对边嵌入的学习，仅利用多头注意力学习节点嵌入，结果证明了该图学习算法的有效性。

### 1.2.2 基于深度生成模型的分子设计

自深度学习研究兴起以来，深度生成模型一直是研究者重点研究的对象。作画、翻译、对话，渲染等应用能够直接服务于广大用户。主流的深度生成模型有变分自编码器 (Variational autoencoders / VAEs)<sup>[1]</sup>，生成对抗模型 (Generative adversarial networks / GANs)<sup>[2]</sup>，流形模型 (Normalizing flows / NFs)<sup>[3-4]</sup>，自回归 (Autoregressive models / ARs)<sup>[5]</sup> 与扩散模型 (Diffusion)<sup>[6-7]</sup> 等。

与图学习的演进过程相似，基于深度生成模型的分子设计也经历了从二维图结构到三维几何构象的演进。主流分子设计任务具体又可以被细分为：分子生成，分子优化，构象生成，蛋白质配体生成，蛋白质生成，蛋白降解靶向嵌合体生成。

分子生成的任务就是根据给定分子数据，使模型具备凭空生成全新且有效的药物分子图或三维结构。考虑到复杂药物分子主要由官能团等子结构组成，JT-VAE<sup>[31]</sup> 基于 VAE 生成树结构骨架，而后利用树结构骨架逐步生成分子图结构。GraphVAE<sup>[32]</sup> 是早期的基于 VAE 的图生成研究，为避免离散化结构的线性表示的相关障碍，使其中解码器直接输出预设的最大概率的全连接图。基于 NF 模型，MoFlow<sup>[33]</sup> 将隐式表征逐步映射到条件流过程中，模型首先生成连接原子的键，随后通过图条件流生成原子，并最终组成有效的分子图。

现有的基于扩散的分子生成模型较少，此领域最早的研究为 EDM<sup>[34]</sup>，其基于 EGNN<sup>[19]</sup> 的去噪过程内核，通过生成点云，再根据预设定的化学性质生成连接边的键。MDM<sup>[35]</sup> 在此基础上，提出了全新的去噪内核，通过两个 SchNet<sup>[21]</sup> 分别对局部节点和全局节点特征进行学习。由于这种扩散模型更擅长在连续样本空间上的学习，故主流方法并不直接预测分子图中边的存在性，针对这一问题，DiGress<sup>[36]</sup> 和



MiDi<sup>[37]</sup> 提出将图邻接矩阵引入扩散过程，并相应将马尔可夫状态转移矩阵引入噪声序列而非传统的高斯噪声。GCDM<sup>[38]</sup> 针对对空间几何信息提取不足的问题，引入 ColfNet<sup>[39]</sup> 相似的空间几何学习范式，实现了对几何信息的充分学习。

### 1.2.3 扩散模型

在图学习领域中，一个图可以被表示为一个元组  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ ，其中包含了一个节点集合  $\mathcal{V}$  和一个边集合  $\mathcal{E}$ 。在训练阶段，模型学习图在扩散过程中的概率分布，并被用于后续的采样阶段，以迭代地方式生成新的图。基于扩散的生成模型因其在多个领域的生成任务中表现出色而受到广泛关注，例如计算机视觉<sup>[40-43]</sup>，自然语言处理<sup>[44-46]</sup>，以及其他各种跨学科任务等<sup>[47-49]</sup>。

人体骨架的表示形式与分子类似，都是表示为被边连接起的点云。不同之处在于生成人体骨架动作时，不需要对边的存在性进行预测。基于扩散模型，MoFusion<sup>[50]</sup> 在人体动作生成中使用 U-Net<sup>[51]</sup> 作为扩散模型中去噪核心的架构。MoDi<sup>[52]</sup> 利用结构感知神经滤波器和 3D 卷积，实现对每个关节的精确控制。

除了在分子生成和人体动作生成上的应用外，许多研究工作致力于将扩散过程用于其他形式的图数据生成。SaGess<sup>[53]</sup> 通过引入广义分治框架来增强 DiGress<sup>[36]</sup>。对于 2D 图生成，EDP-GNN<sup>[54]</sup> 将 Score SDEs 结合到离散图邻接矩阵生成中。在此基础上，GraphGDP<sup>[55]</sup> 提出了名为 Position-enhanced 图得分网络的去噪核心，以实现位置信息的进一步利用。GSMD<sup>[56]</sup> 不仅仅在边属性上进行采样，还利用了节点特征和图谱空间上的 Score SDE 高效地生成图，这在通用数据集和分子数据集上得到了验证。受 VAEs 启发，NVDiff<sup>[57]</sup> 采用 VGAE 结构，首先对图在隐变量空间上的特征进行采样，然后将其解码为节点或边特征。SLD<sup>[58]</sup> 以类似的方式进行图生成任务。GraphARM<sup>[59]</sup> 通过引入节点吸收扩散过程，将自回归模型与扩散模型结合起来。EDGE<sup>[60]</sup> 将扩散模型引入大图生成，扩散过程中逐渐移除边，直到图为空。该模型还通过仅关注部分节点来避免生成过多的边。专为双曲图设计的 HGDM<sup>[61]</sup> 在提取双曲嵌入的复杂几何特征方面表现出优良的性能。

## 1.3 创新点

近年来，深度生成模型，特别是基于扩散模型的生成模型<sup>[62-64]</sup> 在各个生成式人工智能研究领域中都取得了广泛且重大的进展。与生成方法的发展趋势相一致，分子发现领域的主流方法已经从之前的生成模型转变为基于扩散的模型，并从设计 2D 图形转变为 3D 构象。然而，3D 分子生成面临两个主要挑战。第一个挑

战是在预测准确稳定的分子构象方面，而另一个挑战则是充分利用几何信息以促进离散图结构的生成。在本文中，我们提出了几何促进的分子扩散（Geometric-facilitated Molecular Diffusion / GFMDiff），这是一种解决上述挑战的全新的 3D 分子生成方法。GFMDiff 能够生成准确的 3D 几何构型，同时解决图自身离散性带来的诸多问题。

作为扩散模型中广泛采用的范式，去噪扩散概率模型（Denoising Diffusion Probabilistic Models / DDPMs<sup>[62,65]</sup>）在各种生成任务中表现出色。通过在前向过程中逐渐添加高斯噪声，将输入数据转换为预定义的噪声分布，并在采样阶段迭代去噪，最终得到生成结果。与变分自编码器（VAEs）<sup>[1]</sup>和生成对抗网络（GANs）<sup>[2]</sup>等端到端方法相比，这种在每个时间步训练扩散模型的方法在准确性、效率和训练难度方面表现更好。扩散模型在计算化学<sup>[34,66]</sup>和生物学<sup>[67-68]</sup>等领域的应用已经显示出卓越的性能。在分子生成任务的背景下，逐步调整分子构象与分子动力学的核心原理高度一致。

全新的分子生成是分子生成领域一个重要的任务，该任务要求生成有效、新颖且结构稳定的分子。为了解决生成的 3D 构象所要求的等变性条件，一些基于扩散的方法<sup>[34-35]</sup>通过原子间距离间接对分子进行建模，这直接反映了原子间相互作用力的强度。然而，早期的方法<sup>[34]</sup>没有解决多个原子之间的复杂原子间关系。而 MDM<sup>[35]</sup>只是简单地使用阈值区分化学键和原子间力所造成的影响，而不考虑具体的原子和键的类型。最近的研究<sup>[69]</sup>表明，键角之于分子学习与原子对间相互距离同样重要，但只有少数方法充分利用了空间信息。此外，由于分子扩散方法只作用于点云，传统的图卷积无法区分不同原子的重要性。鉴于这些挑战，我们设计了一种新颖的双轨分子学习框架，命名为双轨 Transformer 网络（Dual-track Transformer Network / DTN）。通过集成全局 Transformer 架构，本文将 DTN 作为扩散模型的去噪核函数，实现对空间几何信息的充分学习。

鉴于扩散模型在连续数据上的出色性能，大多数分子图生成模型采用了在笛卡尔坐标和特征上采用扩散和去噪方法，然后基于预定义的规则生成分子图，而不是直接通过模型预测键的存在。间接生成图形的方式可能导致生成样本的稳定性和有效性有所下降。为了使扩散模型适用于分子的多模态数据，一些研究<sup>[36-37,54]</sup>在扩散和去噪过程中引入了邻接矩阵。然而，将图和边嵌入模型中会导致计算成本的增加。在我们的研究中，我们将扩散模型视为一个过程，即在每个时间步根据局部多体原子间关系逐步更新原子信息。准确的特征学习有助于对分子构型进行精确预测。为了预测准确的分子构象，我们设计了一种在训练过程中减少嵌入和局部几何之间差距的方法。在本文中，我们通过精心设计的损失函数几何促进的

---

损失函数 (Geometric-facilitate Loss / GFLoss), 在训练过程中积极的干预模型学习, 促使模型生成稳定且合理的化学键。

在本文中, 我们提出了用于 3D 分子生成的几何促进的分子扩散 (Geometric-facilitated Molecular Diffusion / GFMDiff) 框架。与先前的方法主要基于原子对距离学习原子特征不同, 本文成功地将三元几何信息与原子对距离有效地结合到分子学习中。本领域大多数研究在生成式直接生成点云, 并根据预设规则完成 3D 图结构的搭建。然而这种方法存在两个主要问题。首先, 间接的图生成方式导致样本的稳定性和有效性下降。其次, 传统的图卷积不足以区分局部和全局信息。为了解决第一个约束, 本文设计了一个精巧的几何促进的损失函数 (Geometric-facilitate Loss / GFLoss), 在训练阶段主动引导键的形成。至于第二个约束, 我们引入了双轨 Transformer 网络 (Dual-track Transformer Network / DTN), 这是一个基于全局 Transformer 的神经网络, 以促进全面的几何学习和局部特征学习。最后在试验阶段, 本文进一步在分子性质预测任务上检验了本文提出的 DTN 神经网络的有效性。总而言之, 本文的创新点如下:

- 本文提出的 GFMDiff 框架及其中的 DTN 网络, 能够综合且全面的利用空间信息, 以捕捉原子之间的多体相互作用, 这对生成有效且稳定的全新分子至关重要。
- 为了解决图形的离散性问题, 引入了一个精心设计的 GFLoss, 以促进键的形成, 高效的处理由图内在离散性与扩散模型不兼容带来的挑战。
- 提出了 DTN 作为全局图卷积的替代方案, 可以有效地捕捉全局和局部信息, 并在后续分子性质预测任务中进一步验证其有效性。

## 1.4 基本框架

本文可分为四章, 每章的主要内容如下:

第一章为引言, 本文在该部分介绍了分子生成的研究背景及意义, 相关研究领域的研究现状及本文的主要创新点。

第二章中, 本文在该部分介绍了分子图与扩散模型等预备知识。

第三章中, 本文在该部分介绍了本文提出的 GFMDiff 框架。

第四章中, 本文在该部分介绍了本文的 GFMDiff 模型分子生成任务中的表现及其中去噪核心 DTN 在分子学习任务上的表现。

第五章中, 本文在该部分总结了本文研究的内容, 并对未来发展做出展望。

## 第2章 分子图学习与基于扩散模型的图生成

### 2.1 分子图学习

在图学习中通常有两类监督任务：节点分类/回归和图分类/回归。对于图学习，一个分子可以被抽象为一个图  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ ，其中  $|\mathcal{V}| = n$  表示节点（原子）的集合， $|\mathcal{E}| = m$  表示分子中的边（化学键）的集合。我们用  $e_i$  表示节点  $i$  的特征，用  $e_{ij}$  表示边  $(i, j)$  的特征。

在节点分类或回归任务中，每个节点  $i$  都有一个标签或目标  $y_i$ 。任务目标是通过学习，预测未见节点的标签。这个任务可被应用于许多应用，比如识别分子中的功能团或预测各个原子的性质。

此外，在图分类或回归任务中，给定向一组图  $\{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_N\}$  和对应的标签或目标  $\{y_1, \dots, y_N\}$ 。此时任务目标是根据图的结构和节点边的特征，预测给定图的标签或目标。这个任务被广泛应用于化合物分类或基于结构预测分子性质等。

在这两类任务中，目标是利用监督学习技术训练模型，有效地捕捉图的结构与相关标签/目标之间的关系。现有的图学习方法主要有基于图神经网络的模型和基于 Transformer 架构的模型。

图神经网络 (Graph Neural Networks / GNNs)，近来在知识图谱、社交网络和药物发现等各个领域引起了广泛的关注。GNNs 的核心操作在于将图中节点或边之间的特征进行传递（也称为邻居聚合）。消息传递操作通过聚合节点  $i$  的邻居节点和边的隐式特征来迭代更新节点  $i$  自身的隐式特征  $e_i$ 。一般来说，消息传递过程包含多轮迭代，每轮迭代可以被视为对更远距离邻居的消息聚合。假设有  $L$  轮迭代，第  $l$  轮迭代会将目标节点的 1 跳邻居特征注入目标节点的隐式特征。第  $l$  轮迭代中，消息的传递与聚合可被表示为：

$$m_j^{(l)} = \text{MSG}^{(l)}(e_j^{(l-1)}), j \in \mathcal{N}_i, \quad (2.1)$$

$$e_i^{(l)} = \text{AGG}^{(l)}(\{m_j^{(l)}, j \in \mathcal{N}_i\}, m_i^{(l)}), \quad (2.2)$$

其中  $m_j^{(l)}$  为聚合后的消息，MSG 为消息聚合函数，主流的选择包括多层感知机 (MLP)，可学习权重。 $e_i^{(l)}$  为聚合后的目标节点特征，AGG 为聚合函数，主流的方式有平均值池化、最大池化和图注意力机制等。对每一次消息传递，MSG 和 AGG 函数都可能存在可训练的参数，这些参数在每轮迭代中一般是共享的。经过  $L$  次

消息传递与聚合后，最后一次迭代后的隐式特征被视为模型预测的节点嵌入，即  $e_i^{(L)}$ 。最后，根据具体任务需要，运用相应的 READOUT 函数即可得到节点信息或是全图级别的信息。

近年来，伴随 Transformer 模型的出现，一些研究也开始将该架构引入图学习中。多头注意机制是 Transformer 的核心模块，它将多个注意力层堆叠在一起，实现并行运算。一个注意力层的核心要素一般为一组查询、键、值组成，记为  $(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ 。通过查询和键的点积，经过 softmax 函数归一化后即可获得注意力概率，具体可表示为：

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q} \cdot \mathbf{K}^T}{\sqrt{d}}\right) \cdot \mathbf{V}, \quad (2.3)$$

其中  $d$  为注意力头数。多头注意力机制已经在广泛的应用中证明了其优异的性能。在图学习领域，基于多头注意力机制的 Transformer 机制不依赖于图神经网络在每一层之间进行消息传递的范式，能够对全局信息有更全面的建模。但是相应的，与图神经网络相比，Transformer 模型对算力也有更高的要求。

## 2.2 等变性要求

等变性是涉及到几何空间的深度学习模型需要满足的基础条件。设  $T_g : X \rightarrow X$  是在  $g \in G$  上对  $X$  的变换集合。我们称一个函数  $\phi : X \rightarrow Y$  在  $g$  上具有等变性，如果存在一个变换  $S_g : Y \rightarrow Y$ ，使得下式成立：

$$\phi(T_g(x)) = S_g(\phi(x)) \quad (2.4)$$

假设  $\phi(\cdot)$  是一个非线性函数， $x = (x_1, \dots, x_N) \in \mathbb{R}^{N \times d}$  是在  $d$  维空间中的点云输入集合， $\phi(x) = y \in \mathbb{R}^{N \times d}$  是经变换后的点云集合， $T_g(x) = x + g$  和  $S_g(y) = y + g$  分别是针对输入集合和输出集合的平移变换，如果两个变化具备等变性，那么输入和输出也是等变的。如果变换  $\phi : X \rightarrow Y$  对平移变换具有等变性，则先对输入集合进行平移  $T_g(x)$ ，然后对其应用函数  $\phi(T_g(x))$ ，得到的结果与先对原函数进行变换得到  $y = \phi(x)$ ，然后对输出进行等效平移  $T_g(y)$  后的结果相同，满足方程  $\phi(x + g) = \phi(x) + g$ 。在本工作中，我们探讨了以下三种粒子集合等变性：

- 1. 平移等变性：将输入集合在  $g \in \mathbb{R}^d$  上进行平移，结果等价于对输出进行平移。令该变换  $x + g$  表示为  $(x_1 + g, \dots, x_N + g)$ ，则有  $y + g = \phi(x + g)$ 。

- 2. 旋转和反射等变性：对于任意正交矩阵  $Q \in \mathbb{R}^{d \times d}$ ，对输入做旋转或反射变化等价于对输出结果做旋转或反射变化  $Qy = \phi(Qx)$ 。
- 3. 置换等变性：对输入进行置换会等效地对输出进行相同的置换。 $P(y) = \phi(P(x))$ ，其中  $P$  是行索引上的一种置换。

等变性要求在高维空间内的深度学习中是一项需要满足的基本条件。对于在  $n$  维空间上同时满足平移，旋转，和置换不变性的变换集合，可记为  $E(n)$ 。

## 2.3 基于扩散模型的图生成

扩散模型在生成模型中引起了相当大的关注。通过学习逆过程的去噪核，这些模型能够揭示噪声样本的潜在分布。在给定一段数据的情况下，正向过程被视为马尔可夫链，通过学习可学习参数控制噪声强度，逐渐向数据添加高斯噪声  $T$  次。在生成过程中，模型将噪声还原回真实数据的原始分布。

### 2.3.1 扩散过程

设  $G_t (t = 0, 1, \dots, T)$  表示分子几何信息的分布， $\beta_t \in (0, 1), t = 0, 1, \dots, T$  表示马尔可夫链的噪声方差序列。因此我们可以得到几何信息在  $t$  或  $T$  时刻在后验分布：

$$q(G_{1:T}|G_0) = \prod_{t=1}^T q(G_t|G_{t-1}), \quad (2.5)$$

$$q(G_t|G_{t-1}) = \mathcal{N}(G_t; \sqrt{1 - \beta_t}G_{t-1}, \beta_t I). \quad (2.6)$$

随着时间步  $t$  的增加，扩散过程逐步将更多的噪声添加到原始数据分布，具体表现为噪声方差  $\beta_t$  从 0 平滑地过渡到 1。设  $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s = \prod_{s=1}^t (1 - \beta_s)$ ，则任意  $t$  时刻样本数据的分布可以表示为：

$$q(G_t|G_0) = \mathcal{N}(G_t; \sqrt{\bar{\alpha}_t}G_0, (1 - \bar{\alpha}_t)I). \quad (2.7)$$

经过足够多的时间步  $T$ ，初始分布被转化成白噪声，即  $G_t \sim \mathcal{N}(G_t; 0, I)$ 。在此基础上，去噪过程会将白噪声反向还原成不带噪声的样本分布。

## 去噪过程

在去噪过程（也称采样过程）中，模型通过学习接近真实的逆过程  $q(G_{t-1}|G_t)$  的马尔可夫核函数  $p_\theta(G_{0:T-1}|G_T) = \prod_{t=1}^T p_\theta(G_{t-1}|G_t)$  来重新构建原始的几何信息。每个时间步中学习到的去噪分布为：

$$p_\theta(G_{t-1}|G_t) = \mathcal{N}(G_{t-1}; \mu_\theta(G_t, t), \sigma_t^2 I), \quad (2.8)$$

其中  $\mu_\theta(G_t, t)$  是可训练的神经网络，被用于近似估计均值， $\sigma_t^2 = \frac{(\beta_t - \beta_{t-1})\beta_{t-1}}{(1 - \beta_{t-1})\beta_t}$  是预定义的方差。在采样过程开始时， $p_\theta(G_T)$  采样自标准高斯分布中，然后通过迭代的去噪过程逐步优化参数化的神经网络，并最终还原得到初始时刻的几何信息，即采样结果。

理论上，神经网络的目标函数采用数据的对数似然函数的变分下界形式：

$$\log p(G) \geq \mathcal{L}_{base} + \sum_{t=0}^T \mathcal{L}_t, \quad (2.9)$$

$$\mathcal{L}_{base} = -KL(q(G_T|G_0)|p(G_T)), \quad (2.10)$$

$$\mathcal{L}_t = KL(q(G_{t-1}|G_t)|p(G_{t-1}|G_t)). \quad (2.11)$$

然而，研究发现，神经网络预测的高斯噪声  $\epsilon$  能够更容易的用于训练。因此，在实际优化神经网络时，神经网络常用的目标函数  $\mathcal{L}_t$  <sup>[70]</sup> 形式为：

$$\mathcal{L}_t = E_{\epsilon_t \sim \mathcal{N}(0, I)} \left[ \frac{1}{2} \left( 1 - \frac{\text{SNR}(t-1)}{\text{SNR}(t)} \right) \|\epsilon_t - \hat{\epsilon}_t\|^2 \right]. \quad (2.12)$$

### 第3章 几何促进的 3D 分子图生成 (GFMDiff)

在本节中，本文将着重介绍 3D 分子生成的模型框架，具体包括本文提出的  $E(n)$  等变去噪内核，几何信息促进的损失函数，扩散及去噪过程和优化目标。本文提出的 3D 分子生成整体框架由图 3-1所示。

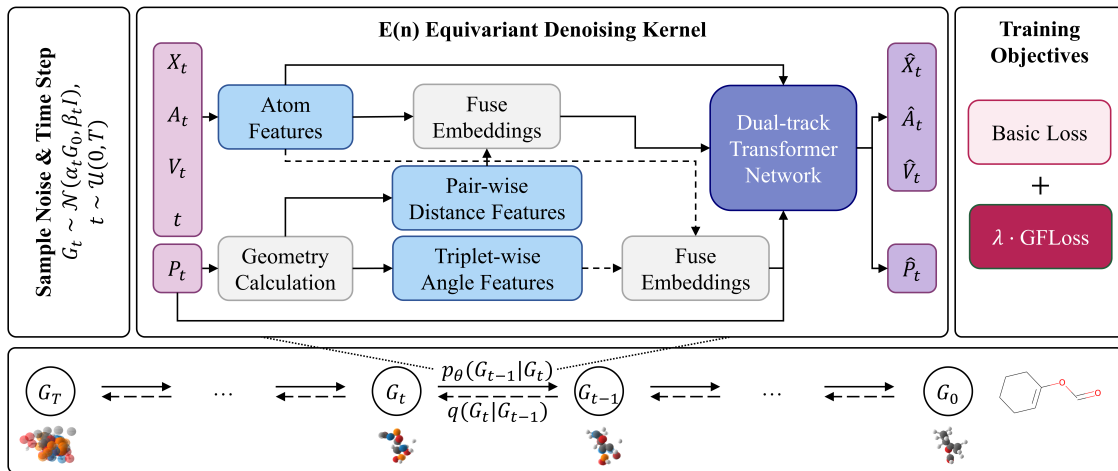


图 3-1: GFMDiff 模型框架示意图

对于每个训练样本，模型的输入为随机选取的时间步  $t$  和采样自对应时刻噪声分布的样本。通过几何计算及编码，得到的原子特征，原子对距离特征，三元组角度特征将被用于  $E(n)$  等变的去噪内核 DTN 中的分子学习，并得到更新后的样本在  $t - 1$  时刻的分布，即对应的原子特征和位置信息。

#### 3.1 双轨 Transformer 网络 (DTN)

在这个小节中，我们将详细介绍作为 GFMDiff 的  $E(n)$  等变去噪内核的双轨 Transformer 网络 (Dual-track Transformer Network / DTN)。DTN 被设计用于有效捕捉原子之间的关系和原子特征。由于三维分子几何具有旋转、平移、反射和排列等不变性质，使得去噪核满足这些性质是很重要的。所提出的 DTN 不仅是  $E(n)$  等变的，还能充分利用空间信息。

在我们提出的方法中，我们将具有总原子数  $N$  的输入分子视为  $G = (P, X, A, V)$ ，其中  $P = (p_1, p_2, \dots, p_N) \in \mathbb{R}^{N \times 3}$  表示原子坐标， $X = (x_1, x_2, \dots, x_N) \in \mathbb{R}^{N \times nf}$  表示原子编号的独热编码， $A = (a_1, a_2, \dots, a_N) \in \mathbb{R}^N$  表示原子编号， $V = (v_1, v_2, \dots, v_N) \in \mathbb{R}^N$  表示原子的价数。为了确保等变性，DTN 利用原子对距离信息和三元组角度信息捕捉几何信息。原子  $i$  和  $j$  之间的欧几里得距离反映了原子间相互作用的强



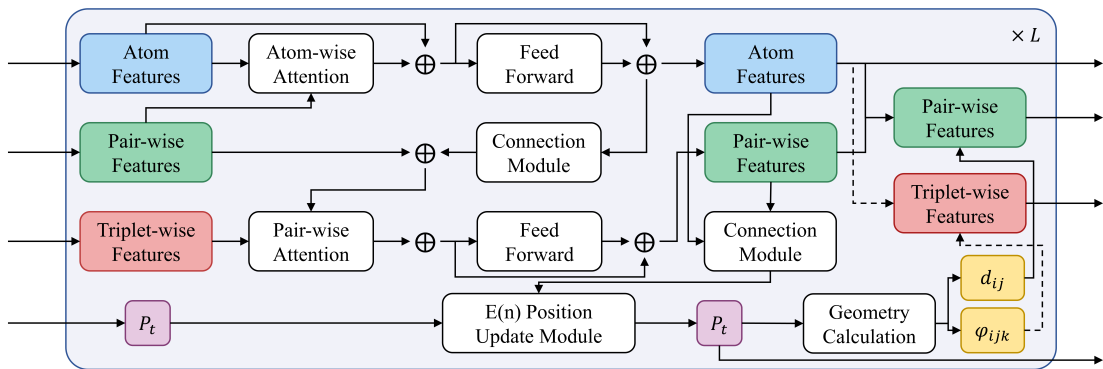


图 3-2: DTN 去噪内核结构示意图

度，可以通过以下公式获得：

$$d_{ij} = \|p_i - p_j\|_2. \quad (3.1)$$

在分子中，除了氢原子，其他多数原子能够形成超过一个单键，故存在大量原子间的多体复杂关系。因此仅使用原子对距离是不足以充分提取空间几何信息。本文提出进一步使用以下公式计算三元角度：

$$\varphi_{ijk} = \arccos\left(\frac{(p_i - p_j) \times (p_i - p_k)}{\|p_i - p_j\|_2 \times \|p_i - p_k\|_2}\right). \quad (3.2)$$

经过上述几何计算，本文通过径向基神经网络（Radial Basis Function / RBF）编码得到可以被用作神经网络学习的原子对距离特征和三元组角度特征：

$$e_{ij} = \text{Linear}(\text{RBF}(d_{ij}), e_i, e_j), \quad (3.3)$$

$$e_{ijk} = \text{Linear}(\text{RBF}(\varphi_{ijk}), e_i, e_j, e_k), \quad (3.4)$$

其中  $e_i = \text{Embedding}(x_i, a_i, v_i)$  是原子  $i$  的节点嵌入，由原子序数和化合价决定。这些特征随后被输入到  $L$  层的 DTN 中。

DTN 的每一层由以下组件组成：原子对轨道、对-三元轨道和连接模块。原子对轨道模拟了原子之间的相互作用力对目标原子的影响，而对-三元轨道模型则模拟了潜在键角对边的影响。连接模块作为两个轨道之间的桥梁，将原子特征注入到成对特征中，以促进更好的表示学习。

原子对轨道涉及预测其他原子和原子之间相互作用力对目标原子的影响。该

轨道以原子嵌入  $e_i$  和对嵌入  $e_{ij}$  作为输入：

$$e_i = \text{LayerNorm}(e_i), e_{ij} = \text{LayerNorm}(e_{ij}), \quad (3.5)$$

$$\mathbf{Q}_i = \text{Linear}(e_i), \mathbf{K}_i = \text{Linear}(e_i) + \text{Linear}(e_{ij}), \quad (3.6)$$

$$a_i = \text{Dropout}(\text{softmax} \frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{d_h}}), \quad (3.7)$$

$$\mathbf{V}_i = \text{Linear}(e_{ij}) + \text{Linear}(e_i) + \text{Linear}(e_j), \quad (3.8)$$

$$\hat{e}_i = \text{Linear}(a_i \mathbf{V}_i^T), \quad (3.9)$$

其中  $d_h$  是头的数量。原子嵌入首先通过添加原子对轨道的预测来更新，然后传递给前馈网络。在每一层中，原子吸收其他原子和相应的原子对的聚合表示。

类似地，对-三元轨道预测了复杂几何亚结构对原子间相互作用力的影响。

$$e_{ij} = \text{LayerNorm}(e_{ij}), \quad (3.10)$$

$$e_{ijk} = \text{LayerNorm}(e_{ijk}), \quad (3.11)$$

$$\mathbf{Q}_{ij} = \text{Linear}(e_{ij}), \quad (3.12)$$

$$\mathbf{K}_{ij} = \text{Linear}(e_{ij}) + \text{Linear}(e_{ijk}), \quad (3.13)$$

$$a_{ij} = \text{Dropout}(\text{softmax} \frac{\mathbf{Q}_{ij} \mathbf{K}_{ij}^T}{\sqrt{d_h}}), \quad (3.14)$$

$$\mathbf{V}_{ij} = \text{Linear}(e_{ij}) + \text{Linear}(e_{ijk}), \quad (3.15)$$

$$\hat{e}_{ij} = \text{Linear}(a_{ij} \mathbf{V}_{ij}^T). \quad (3.16)$$

值得注意的是，三元嵌入  $e_{ijk}$  在 Transformer 结构中不会得到更新，因为这会显著增加计算资源的需求。它们只会在原子坐标更新时得到更新。

连接模块的作用是将原子嵌入融合到对嵌入中。对于对嵌入  $e_{ij}$ ，它同时吸收来自连接模块的原子特征信息和来自对-三元轨道的局部空间信息。

$$e_{ij} = \text{LayerNorm}(e_{ij} + \text{Linear}(\text{Linear}(e_i) \otimes \text{Linear}(e_j))) \quad (3.17)$$

在更新坐标的方法上，我们遵循 EDM<sup>[34]</sup> 和 MDM<sup>[35]</sup> 的相关设计。由于原子坐标得到了更新，原子对和原子三元组的嵌入也将得到更新：

$$e_{ij} = \text{Linear}[\text{Linear}(\text{RBF}(\hat{d}_{ij}), \hat{e}_i, \hat{e}_j)], \quad (3.18)$$

$$e_{ijk} = \text{Linear}(\text{RBF}(\hat{\varphi}_{ijk}), e_{ijk}). \quad (3.19)$$

### 3.2 几何信息促进的损失函数 (GFLoss)

预测化学键的存在是分子图生成中的基本且不可或缺的任务。与以往的研究完全依赖于预设规则生成边不同，我们提出在训练过程中积极干预化学键的形成，通过设计一种精细的训练目标函数，命名为几何促进损失 (Geometric-facilitated Loss / GFLoss)。这个损失函数的目的是引导模型生成既具有有效的拓扑结构，又具有稳定构象的分子。在分子生成中，我们认为原子的价是一种非常重要的辅助特征类型。因此，在上述提到的分子学习网络 DTN 中，原子的价被作为原子特征的一部分进行了整合。这使得模型能够学习和利用原子的价信息。

根据预定义的规则，具有适当距离的原子对被认为是由化学键连接的。对于单键、双键或三键，某些原子之间存在典型距离。如果一对原子之间的距离在某个范围内，这两个原子被认为是对应类型的键连接的。假设预定义的距离和边界为  $\mathbf{D} \in \mathbb{R}^{nf \times nf \times 3}$  和  $\mathbf{M} \in \mathbb{R}^3$ ，其中 3 表示键的类型数量。基于 DTN 的输出  $\hat{\mathbf{C}}_t = (\hat{P}_t, \hat{X}_t, \hat{A}_t, \hat{V}_t)$ ，我们首先使用 softmax 函数预测原子类型的概率：

$$\mathbf{p}_t(\hat{X}_{\text{atom}}) = \text{softmax}(\hat{X}_t) \in \mathbb{R}^{N \times nf}, \quad (3.20)$$

其中  $\hat{X}_t$  在此处表示维度为  $nf$  的独热编码格式的预测原子类型。原子对的类型概率为

$$\mathbf{p}_t(\hat{X}_{\text{pair}}) = \mathbf{p}_t(\hat{X}_{\text{atom}}) \cdot \mathbf{p}_t(\hat{X}_{\text{atom}}) \in \mathbb{R}^{N \times N \times nf \times nf}. \quad (3.21)$$

利用预测的原子坐标  $\hat{P}_t$ ，可以得到原子对之间的距离矩阵  $\mathbf{d}_t \in \mathbb{R}^{N \times N}$ ，并将其扩展为方便操作的  $\mathbb{R}^{N \times N \times nf \times nf \times 3}$ 。然后，原子对之间的距离与典型的键距离之间的边界  $\mathbf{m}_t$  计算如下：

$$\mathbf{m}_t = \mathbf{d}_t - (\mathbf{D} + \mathbf{M}) \in \mathbb{R}^{N \times N \times nf \times nf \times 3}. \quad (3.22)$$

以原子  $i$  和  $j$  为例，假设它们被认为是碳原子的概率大于零，如果边界  $\mathbf{m}_t(i, j, \text{C}, \text{C}, :)$  中的任何元素小于零，则表示原子  $i$  和  $j$  之间存在一条键。键的具体类型由边界  $\mathbf{m}_t(i, j, \text{C}, \text{C}, :)$  中最小值的索引确定。如果  $\arg \min(\mathbf{m}_t(i, j, \text{C}, \text{C}, :))$  为 1，则它们由一条单键连接。如果  $\arg \min(\mathbf{m}_t(i, j, \text{C}, \text{C}, :))$  为 3，则它们由一条三键连接。表示键的布尔矩阵记为  $\text{ISBOND}_t \in \mathbb{R}^{N \times N \times nf \times nf}$ 。

一旦我们获得了原子对的类型概率和键的存在情况，就可以估计原子的可能价：

$$\hat{V}_{\text{pred}}(t) = \text{sum}(\mathbf{p}_t(\hat{X}_{\text{pair}}) \odot \text{ISBOND}_t) \in \mathbb{R}^N. \quad (3.23)$$

由于输入数据受到不同水平噪声的影响，GFLoss 被定义为预测价  $V_{\text{pred}}$  与真实价  $V$  之间的均方误差：

$$\mathcal{L}_{GF}(t) = \|\alpha_t(\hat{V}_{\text{pred}}(t) - V_t)\|^2, \quad (3.24)$$

其中  $\alpha_t$  是扩散过程中噪声数据中真实数据的水平。

### 3.3 扩散及去噪过程

在提出的 GFMDiff 的前向过程中，位置、原子序数和原子价逐渐被噪声先验分布所污染。通过具有预定义方差调度  $\beta_t \in (0, 1), t = 0, 1, \dots, T$  的马尔可夫链将真实构象  $G_0 = (P_0, X_0, A_0, V_0)$  进行转换。后验分布定义为：

$$q(G_{1:T}|G_0) = \prod_{t=1}^T q(G_t|G_{t-1}), \quad (3.25)$$

$$q(G_t|G_{t-1}) = \mathcal{N}_p(p_t; \sqrt{\alpha_t}p_{t-1}, \beta_t I). \quad (3.26)$$

$$\mathcal{N}_h(h_t; \sqrt{\alpha_t}h_{t-1}, \beta_t I), \quad (3.27)$$

其中  $h_t = \text{concat}(x_t, a_t, v_t)$  表示原子特征，为了方便起见。为了满足等变性的要求，在将原子坐标引入扩散之前，它们必须先转换为线性子空间，其中质心为零<sup>[34,71-72]</sup>。因此，噪声分布和上述后验分布都受限于相同的线性子空间。虽然原子特征对于变换是不变的，但它们在整個前向和后向过程中与位置按相同的比例进行转换。

在去噪过程中，我们使用上述的去噪核函数来逼近每个时间步的分子构象，

$$p_\theta(G_{t-1}|G_t) = \mathcal{N}(G_{t-1}; \mu_\theta(G_t, t), \sigma_t^2 I), \quad (3.28)$$

$$\hat{G}_t = G_t/\alpha_t - \hat{\epsilon}_t * \sigma_t/\alpha_t, \quad (3.29)$$

其中  $\hat{\epsilon}_t$  是参数化神经网络的输出。

---

### 3.4 目标函数

对于 DDPMs, 典型的目标函数是数据对数似然的变分下界。在前人的方法基础上, 我们将目标函数与 GFLoss 相结合:

$$\mathcal{L}_t = E_{\epsilon_t \sim \mathcal{N}(0, I)} [\frac{1}{2} \omega(t) \|\epsilon_t - \hat{\epsilon}_t\|^2 + \lambda \|\alpha_t (\hat{V}_{pred}(t) - V_t)\|^2], \quad (3.30)$$

其中  $\omega(t) = (1 - \text{SNR}(t)/\text{SNR}(t-1))$ 。

## 第4章 实验结果及分析

### 4.1 分子生成

#### 4.1.1 实验设置

在本节中,我们报告了 GFMDiff 在两个主流数据集 GEOM-QM9<sup>[73]</sup> 和 GEOM-Drugs<sup>[74]</sup> 上的性能。结果表明,我们的方法在多个方面显著优于一些最先进的模型。

为了进行全面的比较,我们在两个分子生成的基准数据集上进行了实验:GEOM-QM9<sup>[73]</sup> 和 GEOM-Drugs<sup>[74]</sup>。GEOM-QM9 数据集包含超过 130K 个分子及其对应的构象,其中平均每个分子有 18 个原子,包括氢原子。GEOM-Drugs 是一个规模较大的数据集,包括的分子数量和平均分子数量都较多。它包含超过 450K 个分子和 37M 个构象,其中平均分子大小为 44。

#### 4.1.2 在 GEOM-QM9 数据集上的全新三维分子生成

表 4-1: Performance comparison on QM9

Method	NLL↓	Atom Stable (%) ↑	Mol Stable (%) ↑	Valid (%) ↑	Unique-Valid (%) ↑
E-NF	-59.7	85.0	4.9	40.2	39.4
G-SchNet	N/A	95.7	68.1	85.5	80.3
EDM	-110.7±1.5	98.7±0.1	82.0±0.4	91.9±0.5	90.7±0.6
Bridge	N/A	98.7±0.1	81.8±0.2	N/A	N/A
Bridge+Force	N/A	<u>98.8±0.1</u>	84.6±0.3	N/A	N/A
GCDM	<b>-171.0±0.2</b>	98.7±0.0	85.7±0.4	94.8±0.2	93.3±0.0
GFMDiff w/o tri	-123.1±0.4	98.7±0.1	85.9±0.2	94.9±0.2	94.2±0.2
GFMDiff w/o GFLoss	-127.5±0.4	98.7±0.0	<u>86.5±0.1</u>	<u>95.2±0.0</u>	<u>94.5±0.0</u>
GFMDiff	<u>-128.0±0.2</u>	<b>98.9±0.0</b>	<b>87.7±0.2</b>	<b>96.3±0.3</b>	<b>95.1±0.2</b>
Data		99.0	95.2	97.7	97.7

表 4-2: Performance comparison on QM9

Task	$\alpha$	$\Delta\varepsilon$	$\varepsilon_{\text{HOMO}}$	$\varepsilon_{\text{LUMO}}$	$\mu$	$C_v$
Units	Bohr <sup>3</sup>	meV	meV	meV	D	$\frac{\text{cal}}{\text{mol}}$ K
Naive (Upper-bound)	9.01	1470	645	1457	1.616	6.857
# Atom	3.86	866	426	813	1.053	1.971
EDM	2.76	655	356	584	1.111	1.101
GCDM	1.97	602	344	479	0.844	0.689
GFMDiff	<b>1.74</b>	<b>558</b>	<b>321</b>	<b>430</b>	<b>0.728</b>	<b>0.593</b>
QM9 (Lower-bound)	0.10	64	39	36	0.043	0.040

表 4-3: Performance comparison on Drugs

Type	Method	Atom Stable (%) $\uparrow$	Mol Stable (%) $\uparrow$
Normalizing flow	E-NF	75.0	0
DDPM	EDM	81.3	0.0
	Bridge	81.0 $\pm$ 0.7	0.0
	Bridge+Force	82.4 $\pm$ 0.8	0.0
	GCDM	86.4 $\pm$ 0.2	3.7 $\pm$ 0.3
Ours	GFMDiff	<b>86.5<math>\pm</math>0.2</b>	<b>3.9<math>\pm</math>0.2</b>
Data		86.5	2.8

#### 4.1.3 在 GEOM-QM9 数据集上的条件三维分子生成

#### 4.1.4 在 GEOM-Drugs 数据集上的三维分子生成

### 4.2 分子性质预测

#### 4.2.1 实验设置

#### 4.2.2 在 GEOM-QM9 数据集上的分子性质预测

表 4-4: Performance comparison on QM9

Task	$\alpha$	$\Delta\varepsilon$	$\varepsilon_{\text{HOMO}}$	$\varepsilon_{\text{LUMO}}$	$\mu$	$C_v$
Units	Bohr <sup>3</sup>	meV	meV	meV	D	$\frac{\text{cal}}{\text{mol}}$ K
Naive (Upper-bound)	9.01	1470	645	1457	1.616	6.857
# Atom	3.86	866	426	813	1.053	1.971
EDM	2.76	655	356	584	1.111	1.101
GCDM	1.97	602	344	479	0.844	0.689
GeoDiff						
QM9 (Lower-bound)	0.10	64	39	36	0.043	0.040

#### 4.2.3 在 OC20 数据集上的分子性质预测

---

## 第5章 结论与展望



---

## 参考文献

- [1] KINGMA D P, WELLING M. Auto-encoding variational bayes[A]. 2013. arXiv: 1312.6114.
- [2] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//GHAHRAMANI Z, WELLING M, CORTES C, et al. Advances in Neural Information Processing Systems: Vol. 27. Curran Associates, Inc., 2014.
- [3] DINH L, KRUEGER D, BENGIO Y. Nice: Non-linear independent components estimation. arxiv e-prints, 2014[C]//Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego. 2015.
- [4] DINH L, SOHL-DICKSTEIN J, BENGIO S. Density estimation using real NVP[C]//International Conference on Learning Representations. 2017.
- [5] VAN DEN OORD A, KALCHBRENNER N, KAVUKCUOGLU K. Pixel recurrent neural networks[C]//BALCAN M F, WEINBERGER K Q. Proceedings of Machine Learning Research: Vol. 48 Proceedings of The 33rd International Conference on Machine Learning. New York, New York, USA: PMLR, 2016: 1747-1756.
- [6] SOHL-DICKSTEIN J, WEISS E, MAHESWARANATHAN N, et al. Deep unsupervised learning using nonequilibrium thermodynamics[C]//BACH F, BLEI D. Proceedings of Machine Learning Research: Vol. 37 Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR, 2015: 2256-2265.
- [7] SONG Y, ERMON S. Generative modeling by estimating gradients of the data distribution[C]//WALLACH H, LAROCHELLE H, BEYGELZIMER A, et al. Advances in Neural Information Processing Systems: Vol. 32. Curran Associates, Inc., 2019.
- [8] WEININGER D. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules[J/OL]. Journal of Chemical

- 
- Information and Computer Sciences, 1988, 28(1): 31-36. DOI: 10.1021/ci00057a005.
- [9] HIROHARA M, SAITO Y, KODA Y, et al. Convolutional neural network based on smiles representation of compounds for detecting chemical motif [J/OL]. BMC bioinformatics, 2018, 19: 83-94. DOI: 10.1186/s12859-018-2523-5.
- [10] BJERRUM E J. Smiles enumeration as data augmentation for neural network modeling of molecules[A]. 2017. arXiv: 1703.07076.
- [11] LIU S, ALNAMMI M, ERICKSEN S S, et al. Practical model selection for prospective virtual screening[J/OL]. Journal of Chemical Information and Modeling, 2019, 59(1): 282-293. DOI: 10.1021/acs.jcim.8b00363.
- [12] HUANG K, FU T, GLASS L M, et al. Deeppurpose: a deep learning library for drug-target interaction prediction[J/OL]. Bioinformatics, 2020, 36(22-23): 5545-5547. DOI: 10.1093/bioinformatics/btaa1005.
- [13] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[C]//International Conference on Learning Representations. 2017.
- [14] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C]//GUYON I, LUXBURG U V, BENGIO S, et al. Advances in Neural Information Processing Systems: Vol. 30. Curran Associates, Inc., 2017.
- [15] XU K, HU W, LESKOVEC J, et al. How powerful are graph neural networks? [C]//International Conference on Learning Representations. 2019.
- [16] XIE T, GROSSMAN J C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties[J/OL]. Phys. Rev. Lett., 2018, 120: 145301. DOI: 10.1103/PhysRevLett.120.145301.
- [17] XIONG Z, WANG D, LIU X, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism[J/OL].

- 
- Journal of Medicinal Chemistry, 2020, 63(16): 8749-8760. DOI: 10.1021/acs.jmedchem.9b00959.
- [18] GUO Z, YU W, ZHANG C, et al. Graseq: Graph and sequence fusion learning for molecular property prediction[C/OL]//CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York, NY, USA: Association for Computing Machinery, 2020: 435–443. DOI: 10.1145/3340531.3411981.
- [19] SATORRAS V G, HOOGEBOOM E, WELLING M. E(n) equivariant graph neural networks[C]//MEILA M, ZHANG T. Proceedings of Machine Learning Research: Vol. 139 Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021: 9323-9332.
- [20] FANG X, LIU L, LEI J, et al. Geometry-enhanced molecular representation learning for property prediction[J]. Nature Machine Intelligence, 2022, 4(2): 127-134.
- [21] SCHÜTT K, KINDERMANS P J, SAUCEDA FELIX H E, et al. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions[C]//GUYON I, LUXBURG U V, BENGIO S, et al. Advances in Neural Information Processing Systems: Vol. 30. Curran Associates, Inc., 2017.
- [22] LIU Y, WANG L, LIU M, et al. Spherical message passing for 3d molecular graphs[C]//International Conference on Learning Representations. 2022.
- [23] WANG L, LIU Y, LIN Y, et al. ComENet: Towards complete and efficient message passing for 3d molecular graphs[C]//OH A H, AGARWAL A, BELGRAVE D, et al. Advances in Neural Information Processing Systems. 2022.
- [24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems: Vol. 30. Curran Associates, Inc., 2017.
- [25] HONDA S, SHI S, UEDA H R. Smiles transformer: Pre-trained molecular fingerprint for low data drug discovery[A]. 2019. arXiv: 1911.04738.

- 
- [26] WANG S, GUO Y, WANG Y, et al. Smiles-bert: Large scale unsupervised pre-training for molecular property prediction[C/OL]//BCB '19: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. New York, NY, USA: Association for Computing Machinery, 2019: 429–436. DOI: 10.1145/3307339.3342186.
- [27] CHITHRANANDA S, GRAND G, RAMSUNDAR B. Chemberta: Large-scale self-supervised pretraining for molecular property prediction[A]. 2020. arXiv: 2010.09885.
- [28] YUN S, JEONG M, KIM R, et al. Graph transformer networks[C]//WALLACH H, LAROCHELLE H, BEYGELZIMER A, et al. Advances in Neural Information Processing Systems: Vol. 32. Curran Associates, Inc., 2019.
- [29] RONG Y, BIAN Y, XU T, et al. Self-supervised graph transformer on large-scale molecular data[C]//Advances in Neural Information Processing Systems: Vol. 33. Curran Associates, Inc., 2020: 12559-12571.
- [30] LI P, WANG J, QIAO Y, et al. An effective self-supervised framework for learning expressive molecular global representations to drug discovery[J/OL]. Briefings in Bioinformatics, 2021, 22(6). DOI: 10.1093/bib/bbab109.
- [31] JIN W, BARZILAY R, JAAKKOLA T. Junction tree variational autoencoder for molecular graph generation[C]//DY J, KRAUSE A. Proceedings of Machine Learning Research: Vol. 80 Proceedings of the 35th International Conference on Machine Learning. PMLR, 2018: 2323-2332.
- [32] SIMONOVSKY M, KOMODAKIS N. Graphvae: Towards generation of small graphs using variational autoencoders[C]//Artificial Neural Networks and Machine Learning – ICANN 2018. Cham: Springer International Publishing, 2018: 412-422.
- [33] ZANG C, WANG F. Moflow: An invertible flow model for generating molecular graphs[C/OL]//KDD '20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York,

- 
- NY, USA: Association for Computing Machinery, 2020: 617–626. DOI: 10.1145/3394486.3403104.
- [34] HOOGEBOOM E, SATORRAS V G, VIGNAC C, et al. Equivariant diffusion for molecule generation in 3D[C]//Proceedings of Machine Learning Research: Vol. 162 Proceedings of the 39th International Conference on Machine Learning. PMLR, 2022: 8867-8887.
- [35] HUANG L, ZHANG H, XU T, et al. Mdm: Molecular diffusion model for 3d molecule generation[C/OL]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 37. 2023: 5105-5112. <https://ojs.aaai.org/index.php/AAAI/article/view/25639>. DOI: 10.1609/aaai.v37i4.25639.
- [36] VIGNAC C, KRAWCZUK I, SIRAUDIN A, et al. Digress: Discrete denoising diffusion for graph generation[C]//International Conference on Learning Representations. 2023.
- [37] VIGNAC C, OSMAN N, TONI L, et al. Midi: Mixed graph and 3d denoising diffusion for molecule generation[C]//ICLR 2023 - Machine Learning for Drug Discovery workshop. 2023.
- [38] MOREHEAD A, CHENG J. Geometry-complete diffusion for 3d molecule generation[C]//ICLR 2023 - Machine Learning for Drug Discovery workshop. 2023.
- [39] DU W, ZHANG H, DU Y, et al. SE(3) equivariant graph neural networks with complete local frames[C]//Proceedings of Machine Learning Research: Vol. 162 Proceedings of the 39th International Conference on Machine Learning. PMLR, 2022: 5583-5608.
- [40] AVRAHAM O, LISCHINSKI D, FRIED O. Blended diffusion for text-driven editing of natural images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 18208-18218.
- [41] HO J, SAHARIA C, CHAN W, et al. Cascaded diffusion models for high fidelity image generation[J]. Journal of Machine Learning Research, 2022, 23 (47): 1-33.

- 
- [42] CAI R, YANG G, AVERBUCH-ELOR H, et al. Learning gradient fields for shape generation[C]//VEDALDI A, BISCHOF H, BROX T, et al. Computer Vision – ECCV 2020. Cham: Springer International Publishing, 2020: 364-381.
  - [43] LUO S, HU W. Score-based point cloud denoising[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021: 4583-4592.
  - [44] AUSTIN J, JOHNSON D D, HO J, et al. Structured denoising diffusion models in discrete state-spaces[C]//RANZATO M, BEYGELZIMER A, DAUPHIN Y, et al. Advances in Neural Information Processing Systems: Vol. 34. Curran Associates, Inc., 2021: 17981-17993.
  - [45] HOOGEBOOM E, NIELSEN D, JAINI P, et al. Argmax flows and multinomial diffusion: Learning categorical distributions[C]//RANZATO M, BEYGELZIMER A, DAUPHIN Y, et al. Advances in Neural Information Processing Systems: Vol. 34. Curran Associates, Inc., 2021: 12454-12465.
  - [46] SAVINOV N, CHUNG J, BINKOWSKI M, et al. Step-unrolled denoising autoencoders for text generation[C]//International Conference on Learning Representations. 2022.
  - [47] XIE T, FU X, GANEA O E, et al. Crystal diffusion variational autoencoder for periodic material generation[C]//International Conference on Learning Representations. 2022.
  - [48] SHABANI M A, HOSSEINI S, FURUKAWA Y. Housediffusion: Vector floor-plan generation via a diffusion model with discrete and continuous denoising [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023: 5466-5475.
  - [49] LEI J, DENG C, SHEN B, et al. Nap: Neural 3d articulation prior[A/OL]. 2023. arXiv: 2305.16315.
  - [50] DABRAL R, MUGHAL M H, GOLYANIK V, et al. Mofusion: A framework for denoising-diffusion-based motion synthesis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023: 9760-9770.

- 
- [51] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C/OL]//NAVAB N, HORNEGGER J, WELLS W M, et al. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241. DOI: 10.1007/978-3-319-24574-4\_28.
- [52] RAAB S, LEIBOVITCH I, LI P, et al. Modi: Unconditional motion synthesis from diverse data[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023: 13873-13883.
- [53] LIMNIOS S, SELVARAJ P, CUCURINGU M, et al. Sagess: Sampling graph denoising diffusion model for scalable graph generation[A/OL]. 2023. arXiv: 2306.16827.
- [54] NIU C, SONG Y, SONG J, et al. Permutation invariant graph generation via score-based generative modeling[C]//CHIAPPA S, CALANDRA R. Proceedings of Machine Learning Research: Vol. 108 Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics. PMLR, 2020: 4474-4484.
- [55] HUANG H, SUN L, DU B, et al. Graphgdp: Generative diffusion processes for permutation invariant graph generation[A/OL]. 2022. arXiv: 2212.01842.
- [56] LUO T, MO Z, PAN S J. Fast graph generation via spectral diffusion[A/OL]. 2022. arXiv: 2211.08892.
- [57] CHEN X, LI Y, ZHANG A, et al. Nvdif: Graph generation through the diffusion of node vectors[A/OL]. 2022. arXiv: 2211.10794.
- [58] YANG L, ZHANG Z, ZHANG W, et al. Score-based graph generative modeling with self-guided latent diffusion[Z]. 2023.
- [59] KONG L, CUI J, SUN H, et al. Autoregressive diffusion model for graph generation[Z]. 2023.
- [60] CHEN X, HE J, HAN X, et al. Efficient and degree-guided graph generation via discrete diffusion modeling[A/OL]. 2023. arXiv: 2305.04111.

- 
- [61] WEN L, WEI X. Hyperbolic graph diffusion model for molecule generation [A/OL]. 2023. arXiv: 2306.07618.
- [62] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[C/OL]// LAROCHELLE H, RANZATO M, HADSELL R, et al. Advances in Neural Information Processing Systems: Vol. 33. Curran Associates, Inc., 2020: 6840-6851. [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf).
- [63] SONG Y, ERMON S. Generative modeling by estimating gradients of the data distribution[C/OL]//WALLACH H, LAROCHELLE H, BEYGELZIMER A, et al. Advances in Neural Information Processing Systems: Vol. 32. Curran Associates, Inc., 2019. [https://proceedings.neurips.cc/paper\\_files/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf).
- [64] SONG Y, SOHL-DICKSTEIN J, KINGMA D P, et al. Score-based generative modeling through stochastic differential equations[C/OL]//International Conference on Learning Representations. 2021. <https://openreview.net/forum?id=PxTIG12RRHS>.
- [65] SOHL-DICKSTEIN J, WEISS E, MAHESWARANATHAN N, et al. Deep unsupervised learning using nonequilibrium thermodynamics[C/OL]//BACH F, BLEI D. Proceedings of Machine Learning Research: Vol. 37 Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR, 2015: 2256-2265. <https://proceedings.mlr.press/v37/sohl-dickstein15.html>.
- [66] GUAN J, QIAN W W, PENG X, et al. 3d equivariant diffusion for target-aware molecule generation and affinity prediction[C/OL]//The Eleventh International Conference on Learning Representations. 2023. <https://openreview.net/forum?id=kJqXEPXMsE0>.
- [67] LUO S, SU Y, PENG X, et al. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures[C/OL]//OH A H, AGARWAL A, BELGRAVE D, et al. Advances in Neural Information Processing Systems. 2022. <https://openreview.net/forum?id=jSorGn2Tjg>.



- 
- [68] SHI C, WANG C, LU J, et al. Protein sequence and structure co-design with equivariant translation[C/OL]//The Eleventh International Conference on Learning Representations. 2023. <https://openreview.net/forum?id=pRCMXcfdihq>.
- [69] YUAN Z, ZHANG Y, TAN C, et al. Molecular geometry-aware transformer for accurate 3d atomic system modeling[A/OL]. 2023. arXiv: 2302.00855.
- [70] KINGMA D, SALIMANS T, POOLE B, et al. Variational diffusion models [C]//RANZATO M, BEYGELZIMER A, DAUPHIN Y, et al. Advances in Neural Information Processing Systems: Vol. 34. Curran Associates, Inc., 2021: 21696-21707.
- [71] GARCIA SATORRAS V, HOOGEBOOM E, FUCHS F, et al. E(n) equivariant normalizing flows[C]//RANZATO M, BEYGELZIMER A, DAUPHIN Y, et al. Advances in Neural Information Processing Systems: Vol. 34. Curran Associates, Inc., 2021: 4181-4192.
- [72] XU M, YU L, SONG Y, et al. Geodiff: A geometric diffusion model for molecular conformation generation[C]//International Conference on Learning Representations. 2022.
- [73] RAMAKRISHNAN R, DRAL P O, RUPP M, et al. Quantum chemistry structures and properties of 134 kilo molecules[J]. Scientific data, 2014, 1(1): 1-7.
- [74] AXELROD S, GOMEZ-BOMBARELLI R. Geom, energy-annotated molecular conformations for property prediction and molecular generation[J]. Scientific Data, 2022, 9(1): 185.

---

## 硕士研究生期间的科研成果

论文:

**XU C**, ZHANG Y, WANG W, DONG L. Pursuit and evasion strategy of a differential game based on deep reinforcement learning[J]. Frontiers in Bioengineering and Biotechnology, 2022, 10: 827408.

ZHANG Y, **XU C**, WU X, ZHANG Y, DONG L, WANG W. LFGCF: Light folksonomy graph collaborative filtering for tag-aware recommendation[J]. Expert Systems with Applications, 2022, Under Review.

**XU C**, ZHANG Y, CHEN H, DONG L, WANG W. A fairness-aware graph contrastive learning recommender framework for social tagging systems[J]. Information Sciences, 2023: 119064.

CHEN H, **XU C**, ZHENG L, ZHANG Q, LIN X. Diffusion-based graph generative methods[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, Under Review.

**XU C**, WANG H, WANG W, CHEN H. Geometric-facilitated Denoising Diffusion Model for 3D Molecule Generation[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Under Review.

课题:

面向分布式异构计算系统内存池化关键技术, 国家重点研发计划之先进计算与新兴软件。

竞赛:

“华为杯”第十八届中国研究生数学建模竞赛, 三等奖, 排名: 1/3。

第五届全国应用统计专业学位研究生案例大赛, 三等奖, 排名: 1/3。

OGB-LSC @NeurIPS 2022 (PCQM4Mv2 Track), NO.11, 排名: 1/5。

第十一届“泰迪杯”数据挖掘挑战赛, 三等奖, 排名: 1/3。

---

## 致谢