arXiv:2303.02216v2 [cs.LG] 6 Jul 2023

基于非平衡分子的去噪预训练：实现准确且可迁移的神经势

Yuyang Wang, †,‡ Changwen Xu, † Zijie Li, † 和 Amir Barati Farimani*,†,‡,¶,§

†卡内基梅隆大学机械工程系，宾夕法尼亚州匹兹堡*15213*，美国

‡卡内基梅隆大学机器学习系，宾夕法尼亚州匹兹堡*15213*，美国

¶卡内基梅隆大学材料科学与工程系，宾夕法尼亚州匹兹堡*15213*，美国§卡内基梅隆大学化学工程系，宾夕法尼亚州匹兹堡*15213*，美国

电子邮箱：barati@cmu.edu

摘要

等变图神经网络（GNNs）的最新进展使深度学习能够用于开发快速代理模型，以替代计算成本高昂的 从头算量子力学（QM）方法，进行分子势能预测。然而，使用图神经网络构建准确且具有可迁移性的势能模型仍然具有挑战性，因为数据受到量子力学方法高昂计算成本和理论水平的极大限制，尤其是对于大型且复杂的分子系统。在本工作中，我们提出对非平衡分子构象进行去噪预训练，以实现更准确且更具可迁移性的图神经网络势能预测。具体而言，我们使用随机噪声扰动采样得到的非平衡构象的原子坐标，并对图神经网络进行预训练，使其对受扰动的分子构象进行去噪，从而恢复原始坐标。在多个基准测试上进行的严格实验表明，预训练

显著提高了神经网络势能的准确性。此外，我们表明，所提出的预训练方法与模型无关，因为它能够提升不同不变和等变图神经网络的性能。值得注意的是，我们在小分子上预训练的模型展现出显著的可迁移性：在多种分子系统上进行微调时，模型性能均得到提升，这些系统包括含有不同元素的体系、带电分子、生物分子以及更大型的体系。这些结果凸显了利用去噪预训练方法为复杂分子系统构建更具泛化能力的神经势的潜力。

1 引言

开发高效且可迁移的分子势，在对分子系统进行准确模拟方面发挥着关键作用。然而，准确且可迁移的从头算 量子力学（QM）方法，如 Hartree-Fock方法[1–3] 和密度泛函理论（DFT），[4,5] 计算成本高昂，限制了对大型复杂系统的模拟。另一方面，常见的力场基于量子力学计算或实验数据拟合而成，计算效率高，但通常只适用于特定系统。研究人员已致力于开发适用于特定应用的经验力场，[6] 包括小型有机分子、[7] 生物分子[8,9] 和材料。[10] 然而，这些经过精心筛选的方法可能无法模拟涉及显著极化和多体相互作用的复杂系统。此外，它们也难以迁移到不同系统，因为经验力场是针对不同系统的量子力学或实验数据进行拟合的。[11–13] 如何设计准确且可迁移的从头算量子力学方法近似，已成为现代计算化学面临的主要挑战之一。

机器学习（ML）已成为一种强大的工具，可通过拟合计算化学数据来学习原子间势或力场，有望以更快的计算速度达到接近量子力学的精度。[14,15] 机器学习势需要满足能量守恒和旋转平移不变性等物理约束，[16–18] 为此，人们已提出多种方法。

来满足这些约束。例如，梯度域机器学习（GDML）直接预测力场，并采用基于核的方法来保证能量的保守性。[19]对称 GDML（sGDML）在 GDML 的基础上纳入了空间群对称性和动态非刚性对称性。[20] 高斯近似势（GAP）将能量分解为各原子中心环境的贡献之和，[21,22] 此外，人们也研究了用于描述局部原子贡献的不同核函数或描述符。[23,24]

神经势使用深度神经网络（DNNs）对分子几何结构与势能面（PES）之间的关系进行建模。[25] DNNs 基于分子系统的几何描述符构建，用于模拟势能。最初使用原子间距离和键角，但此类描述符无法体现输入中原子顺序的置换不变性。[26,27] 高维神经网络势（HDNNP）[28] 应运而生，它将分子系统分解为局部原子环境。这些原子环境通过由二体项和三体项组成的原子中心对称函数（ACSFs）[29] 进行编码。此后，ANI[30–32] 和 TensorMol[33]对 HDNNP 进行了扩展。然而，这类模型的可迁移性有限，也就是说，针对新的化学元素重新参数化需要重新训练整个模型。值得注意的是，描述符的规模会随着元素数量的增加而呈平方级增长。

另一方面，端到端神经网络势直接基于笛卡尔坐标和元素类型构建，而非基于手工设计的描述符。基于消息传递[34] 的图神经网络（GNN）用于从分子结构中学习表征，其中节点表示原子，边表示原子间相互作用。通过在图神经网络学习到的表征之上构建多层感知机，来对势能进行建模。图神经网络通过聚合邻近原子的原子信息并更新节点，学习化学和物理相互作用。SchNet 实现连续卷积滤波器来建模原子间距离[35]，DimeNet 则在消息传递中引入键角，以建模三体相互作用。[36] 其他研究尝试引入化学领域知识，以开发性能更好的神经势。[37–39] 然而，这些模型中的消息传递操作

无法编码方向信息。在神经势中，我们关注相对于三维刚体变换的等变性，即神经网络与任意三维旋转、平移和/或反射可交换。第一类等变图神经网络基于群论中的不可约表示，包括 Tensor field network（TFN）、[40] Cormorant、[41] SE(3)-Transformer、[42] SEGNN[43] 等。这些模型利用 Clebsch–Gordan系数和球谐基，以及可学习的径向神经网络来构建等变性。第二类等变图神经网络则依赖于对向量特征进行线性运算（即缩放、线性组合、点积和向量积）。[44–47] PaiNN[48] 和 TorchMD-Net[49] 是这类等变图神经网络的两个例子。PaiNN 提出分别追踪标量特征和向量特征，并设计线性运算来传递二者之间的信息，从而保持等变性。TorchMD-Net 则在此类架构中引入了多头自注意力机制。[50] 鉴于等变图神经网络在神经网络势基准测试中表现出色，我们的工作聚焦于这类方法。[49]

为提升图神经网络在分子性质预测中的性能，研究人员已探索自监督学习（SSL）方法[51,52] 。首先通过 SSL 对图神经网络进行预训练，以学习具有表达能力的分子表征，然后在下游预测任务上进行微调。预测式 SSL 方法通过从部分观测或经过扰动的样本中恢复原始实例来开展学习，包括掩码属性预测、[53] 上下文预测、[54] 基序树生成。[55] 此外，GEM[56] 提出通过预测三维位置信息（包括原子间距离、键长和键角）来预训练图神经网络。另一方面，对比学习旨在通过将正实例与负实例进行对比来学习表征，已广泛应用于化学科学领域的图神经网络。[57–60] MolCLR[61] 通过随机掩码原子和边属性，生成分子图的对比实例。近期研究也探索了对比训练中的多层次子图。[62,63]GraphMVP[64]和三维信息最大化[65]通过对以二维拓扑图和三维几何结构表示的分子进行对比，将三维信息融入二维图

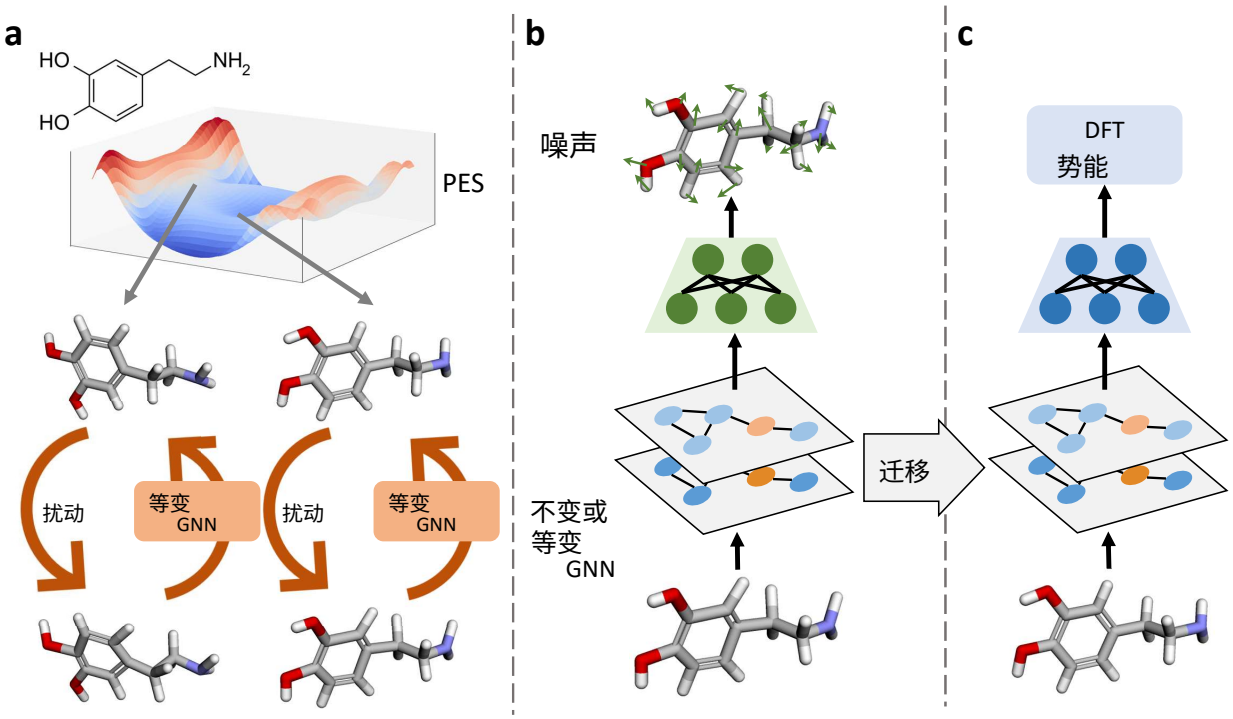


图 1：在非平衡分子上预训练不变和等变图神经网络以进行势能预测的框架。(a) 对不同的分子构象进行采样，并添加随机噪声以扰动这些构象。训练图神经网络，使其从扰动后的构象中恢复原始构象。(b) 在预训练期间，将扰动构象输入图神经网络以预测额外噪声。(3) 将预训练后的图神经网络迁移并与多层感知机头一起微调，以预测通过密度泛函理论计算得到的势能。

和三维几何结构。值得注意的是，大多数自监督学习研究都忽略了三维信息。尽管少数自监督学习方法纳入了三维信息，但它们并未基于等变图神经网络构建，这极大限制了其在精准神经势能预测中的应用。近期，一些研究提出将去噪作为预训练图神经网络的自监督学习方法。[66–68] 通过预测在平衡态下添加到原子坐标中的噪声，训练图神经网络学习特定的伪力场。这种方法已在 QM9 等量子力学性质预测基准上展现出有效性。[69] 然而，该方法仅利用处于平衡态的分子，这远不足以满足精准神经势的需求，因为势能预测需要在模拟中评估非平衡分子结构。总体而言，自监督学习在神经势能预测方面尚未得到充分研究。

尽管神经势已得到广泛研究，但其准确性和可迁移

性仍然有限，这是因为获取训练数据依赖昂贵的量子力学计算。尤其是对于大型且复杂的分子系统，收集准确且充足的量子力学数据极具挑战，甚至可能无法实现。这引出了两个问题：1）我们能否利用自监督学习，通过现有数据提高神经势的准确性；2）我们能否利用自监督学习，开发出可迁移性更强的神经势，充分利用相对丰富的分子势数据（例如小分子），并将预训练模型应用于数据有限的大型复杂分子系统？为应对这些挑战，我们提出了一种基于非平衡分子的自监督学习预训练策略，以提高神经势的准确性和可迁移性（图 1(a)）。该策略通过向非平衡态分子系统的原子坐标添加随机噪声，并以自监督学习的方式训练图神经网络预测这些人工噪声（图 1(b)）。随后，我们在多个具有挑战性的分子势预测基准上对预训练模型进行微调（图 1(c)）。严格的实验表明，我们提出的利用非平衡分子的预训练方法能够显著提高神经势的准确性，并且适用于多种不变或等变图神经网络架构，包括 SchNet、SE(3)-Transformer、EGNN 和 TorchMD-Net。值得注意的是，在小分子上预训练的图神经网络，在多种分子系统上进行微调后，势能预测能力显著提升；这些系统包括含有不同元素的系统、带电分子、生物分子以及更大的系统。这种去噪预训练方法有望促进用于替代昂贵量子力学方法的准确且可迁移代理模型的开发，并推动神经势在复杂系统中的应用。

2 方法

2.1 图神经网络

分子系统可以建模为图 $\mathcal{G} = (V, \mathcal{A})$，其中 V 表示系统中的 N 原子，$\mathcal{A}$ 表示原子之间的相互作用（例如化学键）。为了学习

分子图的表示，现代图神经网络采用消息传递策略。[34]设 $h_i^{(t)} \in \mathbb{R}^F$ 为图神经网络第 t 层原子 i 的 F 维特征， a_{ij}表示原子 i 和 j之间的相互作用。对于原子 i，来自其相邻原子 $j \in \mathcal{N}_i$ 的消息 $m_{ij}^{(t)}$ 通过函数 $f_m^{(t)}$进行聚合（公式 1）。所有 m_{ij} 相加，作为传递给节点 i 的消息；随后，更新函数 $f_u^{(t)}$利用邻居聚合后的消息及其原始特征更新原子特征（公式 2）。

$$m_{ij}^{(t)} = f_m^{(t)}(h_i^{(t)}, h_j^{(t)}, a_{ij}), \tag{1}$$

$$h_i^{(t+1)} = f_u^{(t)}(h_i^{(t)}, \sum_{j \in \mathcal{N}_i} m_{ij}^{(t)}). \tag{2}$$

2.2 等变图神经网络

在本研究中，我们探讨了用于分子势能预测的图神经网络，这类预测需要原子的三维位置信息。为此，我们引入了等变图神经网络，通过扩展消息传递函数来保持分子系统的物理对称性。令$T_g: \mathcal{X} \to \mathcal{X}$ 表示抽象群 $g \in G$ 在输入空间 X上的一组变换，令 $\phi: \mathcal{X} \to \mathcal{Y}$ 为将输入映射到输出空间 $\mathcal{Y}$的函数。若存在变换 $S_g: \mathcal{Y} \to \mathcal{Y}$，使得 $\phi(T_g(x)) = S_g(\phi(x))$, $\forall g \in G$ 且 $x \in \mathcal{X}$，则称函数 ϕ 关于 g 是等变的。E(n)-等变性表示关于欧氏群 E(n)的等变性，该群包含 n 维欧氏空间中的所有平移、旋转和反射；而 SE(n)-等变性仅满足平移和旋转等变性。[70] 等变图神经网络将等变性作为分子建模的归纳偏置，并在许多能量相关性质预测任务中展现出优异性能。

设计等变消息传递操作的一种策略是基于群论中的不可约表示。[41,42] TFN[40] 引入克莱布什–戈登系数、径向神经网络和球谐函数作为构建模块，用于构成SE(3)-等变消息传递

，如公式 3 和 4 所示（为简化符号，我们忽略层的上标）。

$$m_{ij} = \sum_{k \geq 0} W^{\ell k}(r_{ji}) h_j^k, \tag{3}$$

$$W^{\ell k}(x) = \sum_{J=|k-\ell|}^{k+\ell} \varphi_J^{\ell k}(\|x\|) \sum_{m=-J}^{J} Y_{Jm}(x/\|x\|) Q_{Jm}^{\ell k}, \tag{4}$$

其中，$r_{ij} = x_i - x_j$ 是两个原子 i 和 j的笛卡尔坐标 x_i 与 x_j之间的方向向量，h_j^k 是节点 j的类型为 k 的特征，而 $W^{\ell k}: \mathbb{R}^3 \to \mathbb{R}^{(2\ell+1)\times(2k+1)}$是将类型为 k 的特征映射为类型为 ℓ 的特征的权重核。$W^{\ell k}$ 被分解为不可学习的克莱布什-戈尔丹矩阵 $Q_{Jm}^{\ell k}$ 与球谐函数 $Y_J: \mathbb{R}^3 \to \mathbb{R}^{2J+1}$ 的线性组合，并通过可学习的径向网络 φ 实现，如公式 4 所示。由于克莱布什-戈尔丹张量积和球谐函数的实现，这类 SE(3)等变图神经网络通常计算开销较大，且可学习参数的数量有限。[18]

构建等变性的另一种方法是在消息传递过程中，仅对向量特征应用线性运算（即缩放、线性组合、点积和向量积）。[45,48] 根据这一思路，构建 E(3)等变运算的一种直接方法是，除标量特征 h_i 外，还为每个节点保留一个向量特征 $v_i \in \mathbb{R}^{3\times F}$。因此，向量特征的消息传递函数和更新函数分别如公式 5 和 6 所示。

$$m_{ij}^{(t)} = f_m^{(t)}(h_i^{(t)}, h_j^{(t)}, v_i^{(t)}, v_j^{(t)}, \|r_{ij}\|, a_{ij}), \tag{5}$$

$$v_i^{(t+1)} = v_i^{(t)} + \sum_{j \in \mathcal{N}_i} r_{ij} f_v^{(t)}(m_{ij}^{(t)}), \tag{6}$$

其中，f_v 将消息 m_{ij} 映射为标量，并在每一层 t中将每个原子的向量特征更新为方向向量 r_{ij} 的线性组合。需要注意的是，为了保持 f_m 中的等变性，消息函数只能对向量特征应用线性运算

。通常，会将距离原子 i 不超过截断距离 d_{cut} 的原子纳入邻居列表 $\mathcal{N}_i$。与不可约表示相比，这种策略为等变图神经网络提供了更灵活的实现方式。

经过 T 个等变消息传递层后，可以对所有节点特征进行全局池化，以获得整个分子的表示。在我们的工作中，我们对所有节点特征求和作为池化操作。[71] 随后将池化后的表示输入多层感知机，以预测分子势能：$\hat{}E = \texttt{MLP}(\sum_i h_i^{(T)})$。

此外，不变性意味着对输入进行变换（例如旋转和平移）不会改变输出，因此 $\phi(T_g(x)) = \phi(x)$，$\forall g \in G$ 和 $x \in \mathcal{X}$。不变性是一种特殊的等变性，其中 S_g 对于 $\forall g \in G$ 是恒等映射。构建不变图神经网络比构建等变图神经网络更直接。只需将公式 1 替换为公式 5，即可得到 E(3) 不变图神经网络，因为在消息传递过程中，几何信息仅以原子之间距离的形式嵌入。

在本研究中，我们实现了一种不变图神经网络（即 SchNet[35]）和三种等变图神经网络模型（即 SE(3)-Transformer、[42] EGNN、[70] 和 TorchMD-Net[49]），以验证所提出预训练方法的泛化能力。SchNet采用连续滤波器卷积，其中节点特征与仅取决于原子间距离的权重矩阵进行逐元素乘法。SE(3)-Transformer 属于第一类等变图神经网络，它通过在公式 3 中引入标量自注意力[50]项 α_{ij} 对 TFN 进行扩展。输入特征按照公式 4 中的可学习权重核映射为键和查询，以保持 SE(3)等变性；α_{ij} 则通过键和查询的归一化点积计算得出。另一方面，EGNN 和 TorchMD-Net 采用第二种等变图神经网络策略，即对向量特征进行线性运算。EGNN 直接将原子的坐标作为向量特征进行操作。它将公式 5 中的 f_m 实现为节点特征、边特征和距离的拼接，再通过多层感知机计算消息。随后，在每一层中，根据消息加权线性组合原子间方向向量，以更新原子坐标。TorchMD-Net

除节点特征外，还显式地对向量特征进行建模，并在 f_m中针对标量特征和原子间距离构建自注意力。自注意力项被用作原子间方向向量的标量项，以更新向量特征；而标量特征则通过向量特征的点积进行更新。这些模型涵盖了种类广泛的不变/等变图神经网络，并在多个与力场或能量相关的分子基准测试中取得了有竞争力的表现。[48,49]

2.3 非平衡分子去噪

该预训练策略基于预测添加到采样分子构象中的人工噪声。分子的一个构象记为 V 和 X，其中 V编码原子信息，而 $X \in \mathbb{R}^{N\times 3}$ 包含分子中所有 N原子的笛卡尔坐标。对于图神经网络模型，原子间相互作用 $\mathcal{A}$ 可直接由 X 推导得到。从高斯分布 $\mathcal{N}(0,\ \sigma I)$中采样的随机噪声 $\mathcal{E} \in \mathbb{R}^{N\times 3}$ 会添加到所有原子的位置，以扰动分子构象。扰动后的构象记为 $X = X + \mathcal{E}$。在预训练过程中，图神经网络模型经过训练以预测额外噪声，目标函数如公式 7 所示。

$$\mathbb{E}_{p(\hat{X};V)}\left[\|\phi_\theta(V,\hat{X}) - \mathcal{E}\|_2^2\right], \tag{7}$$

其中，ϕ_θ 表示由 θ 参数化的不变/等变图神经网络，$p(X;V)$用于衡量给定原子时扰动分子构象的概率分布。

这种自监督去噪策略与在扰动态学习伪力场有关。[66,72,73] 在这项工作中，我们将这一概念从平衡分子构象扩展到非平衡构象，这对分子模拟至关重要。对于一个包含 N 个原子、编码为 V的给定分子，分子构象 X 的概率为 $p(X;\ V)\ \sim \exp\left(-E(X;V)\right)$，遵循玻尔兹曼分布，其中$E(X;V)$是势能。构象中每个原子的力场为

$-\nabla_X E(X;\ V)\ =\ -\nabla_X \log\ p(X;\ V)$。分子构象 X 可以通过分子动力学模拟、正则模采样、扭转采样等方法采样。这些方法能够提供平衡态附近多样且符合物理和化学规律的非平衡构象，有助于理解分子系统的能量学。[30] 需要注意的是，在本文中，我们将非平衡态定义为不处于能量极小值的分子构象，这与统计力学中的术语不同。通过向每个原子坐标添加随机噪声，可以得到不切实际的构象 X ，其能量高于采样得到的构象 X。因此，我们假设$p(\hat{}X;\ V)$可由以X为中心的高斯分布 $q(\hat{}X;\ X,\ V)\ =\mathcal{N}(X,\ \sigma I_{3N})$近似。根据这一假设，对于给定方差 σ ，力场与扰动噪声成正比，如式 8 所示。

$$\nabla_{\hat{X}} \log p(\hat{X};V) \approx \nabla_{\hat{X}} \log q(\hat{X};X,V) \propto \frac{\hat{X}-X}{\sigma^2} \propto \mathcal{E}. \tag{8}$$

因此，假设采样构象周围原子位置的概率服从高斯分布，那么训练图神经网络以匹配扰动噪声，就等同于学习一个伪力场。在我们的案例中，用高斯分布近似时，玻尔兹曼分布中的温度 T 被视为一个虚构项，这意味着 T 不会在去噪预训练中显式纳入。这样的简化有助于利用不包含温度信息的正则模态采样或扭转采样数据。还值得注意的是，分子系统的玻尔兹曼分布中的温度项 T ，与预训练中随机噪声高斯分布的最优标准差 σ 的选择有关。探究如何通过数据和预训练策略将 σ与 T 联系起来，超出了本研究的范围，但可能是一个值得探索的方向。

需要指出的是，尽管高斯分布可以很好地近似局部极小值附近状态的分布，但对于高能量状态可能并不适用。这一假设与预训练数据集以及标准差的选择有关

σ 高斯噪声添加到原子位置时所采用的标准差。预训练分子构象应当符合物理规律，不应出现过度扭曲。因此，向这些构象添加噪声会产生能量更高的状态，从而确保去噪预训练过程能够使图神经网络学会一种评分函数，进而得到能量更低的构象。有关预训练数据集的详情，请参见第 3.1 节。此外，高斯噪声的标准差（σ）会影响预训练效果。添加过多噪声会使构象扭曲得过于严重，以至于去噪组件难以准确恢复能量较低的原始构象。另一方面，如果扰动过小，分子能量的变化可能微不足道，图神经网络也难以在预训练中学到有意义的表征。有关 σ如何影响神经网络势能性能的详细研究，请参见第 3.6 节。

2.4 噪声预测

为了根据受扰动的分子构象预测噪声，图神经网络模型需要输出向量，而非单个标量。对于缺乏显式向量特征建模的 SchNet 和 SE(3)-Transformer，模型首先输出预测能量 E（一个标量）。然后计算预测能量相对于受扰动原子坐标 X 的负梯度，以评估人工噪声，即 $\mathcal{E} = -\nabla_X E$。另一方面，EGNN 和 TorchMD-Net 会直接跟踪可用于预测噪声的向量特征。在 EGNN 中，噪声评估为 $\hat{\mathcal{E}} = X^{(T)} - \hat{X}$，其中 $X^{(T)}$是经过 T 个消息传递层更新后的坐标。在 TorchMD-Net 中，我们采用门控等变模块[48]，将经过 T层后的更新向量特征 $v^{(T)} \in \mathbb{R}^{N\times 3\times F}$ 和节点特征 $h^{(T)} \in \mathbb{R}^{N\times F}$ 映射为噪声 $\hat{\mathcal{E}} \in \mathbb{R}^{N\times 3}$，使得 $\hat{\mathcal{E}} = \texttt{GatedEquivariant}(v^{(T)}, h^{(T)})$，同时仍保持等变性。

表 1：数据集摘要，包括分子数量、构象数量、元素数量、每个分子中的原子数量、分子类型，以及各数据集是否用于预训练（PT）和微调（FT）。

数据集	分子数量	构象数量	元素数量	原子数量	分子类型	用途
ANI-1[30]	57,462	24,687,809	4	2∼26	小分子	PT 与 FT
ANI-1x[74]	63,865	5,496,771	4	2∼63	小分子	PT 与 FT
ISO17[75]	129	645,000	3	19	C_7O_2H 的异构体 [10]	FT
SPICE[76]	19,238	1,132,808	15	3∼50	小分子，二聚体，二肽、溶剂化氨基酸	FT
MD22[77]	7	223,422	4	42∼370	蛋白质、脂质、碳水化合物，核酸、超分子	FT

3 结果与讨论

3.1 数据集

为评估去噪预训练策略在分子势能预测上的表现，我们研究了五个数据集，其中包含各种具有密度泛函理论计算能量的非平衡分子构象，如表 1 所示。ANI-1[30] 通过正则模采样，从 GDB-17,[69] 中选取一个包含超过 2400 万个构象的子集，并计算密度泛函理论总能量。ANI-1x[74] 通过主动学习算法获得了超过 500 万个新构象，从而扩展了 ANI-1。[31]

除 ANI-1 和 ANI-1x 外，我们还研究了其他用于分子能量预测的数据集。ISO17[75] 选取了固定原子组成（即 $C_7O_2H_{10}$）的分子异构体的从头算 分子动力学（从头算分子动力学）轨迹，其中每个分子包含 5,000 个构象，并具有通过密度泛函理论计算的能量，总共纳入了 129 个分子。与此前仅研究小型有机分子的数据集不同，MD22[77] 和 SPICE[76] 涵盖了更多种类的分子。具体而言，MD22 包含蛋白质、碳水化合物、核酸和超分子（即巴基球捕获器和纳米管）的从头算分子动力学轨迹。MD22 中的大多数分子系统所含原子数多于 ANI-1 和 ANI-1x，更多关于 MD22 的详情可参见补充信息 S1。SPICE 涵盖了 110 多万个构象，由不同的分子系统组成

其中包括小分子、二聚体、二肽和氨基酸。其他数据集仅包含 H、C、N 和 O，而 SPICE 还包含带有卤素和金属的分子，使元素种类总数达到 15 种。它还包含带电分子和极性分子，进一步拓宽了其覆盖的化学空间。

在本研究中，我们将 ANI-1 和 ANI-1x 合并作为预训练数据集，因为它们包含具有不同构象的多种小型有机分子。预训练时，每个分子的所有构象按 95%/5% 的比例划分为训练集和验证集。微调时，我们使用包括 ANI-1 和 ANI-1x 在内的所有数据集，对势能预测进行基准测试。借此，我们考察在小分子上预训练的不变或等变图神经网络能否泛化到其他不同的分子系统。微调期间，对于 ANI-1、ANI-1x、MD22 和 SPICE，我们根据每个分子的构象，将数据集按 80%/10%/10% 的比例划分为训练集、验证集和测试集。此外，对于 ISO17，我们遵循原始文献[75] 中报告的划分策略。

3.2 实验设置

在预训练期间，每个不变或等变图神经网络训练 5 个轮次，最高学习率为 2×10^{-4}，权重衰减为零。所有模型均在 ANI-1 和 ANI-1x 的组合数据上进行预训练，并分别在各数据集上微调。预训练时，我们采用 AdamW 优化器[78] ，批次大小为 256，并使用线性学习率预热与余弦衰减[79] 。微调期间，模型在 ANI-1 和 ANI-1x 上训练 10 个轮次，在 SPICE 和 ISO17 上训练 50 个轮次。由于 MD22 中各分子的原子数量差异很大，我们针对其中的每种分子采用不同的实验设置。作为预训练模型的对照，我们还按照与相应去噪预训练模型相同的设置，在各数据集上从头训练图神经网络。详细的微调设置见补充信息 S2。微调时，迁移预训练消息传递层的参数。首先对消息传递层的原子特征 h^K 求和，再将其输入随机初始化的多层感知机，以预测每个

表 2：在 ANI-1 和 ANI-1x 上使用及不使用去噪预训练时，不同不变和等变图神经网络的性能。

		ANI-1		ANI-1x	
模型	预训练	RMSE（千卡/摩尔）	MAE (千卡/摩尔)	RMSE (千卡/摩尔)	MAE (千卡/摩尔)
SchNet		2.03	1.33	9.29	5.95
SchNet	✓	1.50	1.00	5.53	3.80
SE(3)-Transformer		5.82	4.14	17.67	11.89
SE(3)-Transformer	✓	5.12	3.64	13.76	9.50
EGNN		2.35	1.63	7.35	5.30
EGNN	✓	1.66	1.14	4.94	3.49
TorchMD-Net		0.60	0.39	2.27	1.50
TorchMD-Net	✓	0.49	0.32	1.54	1.01
平均提升		-21.5%	-21.2%	-31.9%	-30.9%

分子构象的能量。预训练层和多层感知机头均针对势能预测进行微调。各图神经网络的超参数依据原始文献中的建议设置。本文所实现的图神经网络的更多详情见补充信息 S3。为评估神经势的性能，我们报告其相对于密度泛函理论势能的均方根误差（RMSE）和平均绝对误差（MAE）。

3.3 神经势能预测

为考察去噪预训练的性能，我们在 ANI-1 和 ANI-1x 的组合数据上对不变图神经网络模型和等变图神经网络模型进行预训练，并在各个数据集上分别微调这些模型。表 2 比较了在 ANI-1 和 ANI-1x 上进行预训练与不进行预训练时的分子势能预测结果。

我们的实验表明，所提出的去噪预训练方法显著提升了图神经网络在分子势能预测方面的准确度。与未预训练的对应模型相比，预训练模型在 ANI-1 上的均方根误差和平均绝对误差平均分别降低了 21.5% 和 21.2%，在 ANI-1x 上则平均分别降低了 31.9% 和 30.9%。此外，结果还验证了去噪预训练

方法与模型无关，因为它显著提升了四种不同图神经网络在两个神经势数据集上的性能，包括 SchNet、SE(3)-Transformer、EGNN 和 TorchMD-Net。例如，不变性 SchNet，以及等变 EGNN 和 TorchMD-Net，在 ANI-1x 上的均方根误差和平均绝对误差均提升了 32% 以上。尤其是，去噪预训练提升了计算成本较低的图神经网络模型（如 SchNet 和 EGNN）的性能。在 ANI-1 和 ANI-1x 两个基准上，EGNN 从预训练中获益最多。这表明，去噪预训练能够提升简单、成本低廉的图神经网络模型的性能，从而有望减少设计复杂等变架构所需的工作。在将神经势应用于真实世界模拟时，这一点至关重要，因为势能计算的效率非常重要。各图神经网络模型计算效率的详情见补充信息 S4。此外，关于增加训练轮次进行微调的研究见补充信息 S5。这些结果表明，在分子势数据集上对不变和等变图神经网络进行去噪预训练，可直接提升预测准确度。

3.4 可迁移性

上一节的结果表明，当下游任务与预训练数据覆盖相似的化学空间（即小分子）时，去噪预训练能有效提升神经势的准确性。然而，真实世界模拟往往涉及规模更大、结构更复杂的分子系统，使用昂贵的 从头算 方法获取此类系统的能量颇具挑战。因此，若能将基于小分子（且训练数据充足）预训练的图神经网络模型迁移到规模更大、结构更复杂的分子系统（且训练数据有限），将大有裨益。为此，我们在另外三个分子势基准（即 ISO 17,[75] SPICE,[76] 和 MD22[77]）上，对预训练的四种图神经网络进行微调，这些网络既包括不变网络，也包括等变网络；这三个基准涵盖了不同于预训练数据集的分子系统。

图2a和图2b比较了采用与未采用去噪

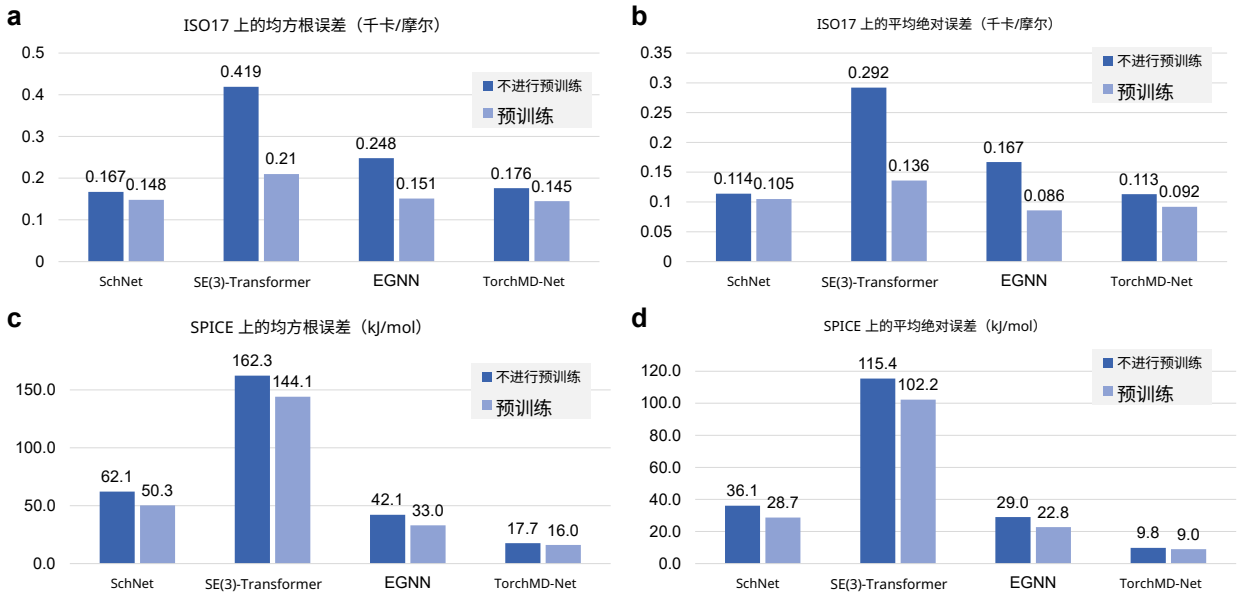


图2：（a）不同图神经网络模型在 ISO17 上的测试均方根误差，（b）不同图神经网络模型在 ISO17 上的测试平均绝对误差，（c）不同图神经网络模型在 SPICE 上的测试均方根误差，以及（d）不同图神经网络模型在 SPICE 上的测试平均绝对误差。

预训练的图神经网络模型在包含 $C_7O_2H_{10}$ 异构体的 ISO17 上的性能。结果表明，采用均方根误差和平均绝对误差进行评估时，预训练图神经网络模型的预测准确度均更高。尤其是，预训练使 SE(3)-Transformer 的性能显著提升了约 50%。结果还表明，预训练模型能够迁移至包含固定原子的特定分子系统。

图2c和2d展示了预训练和非预训练图神经网络在 SPICE 上的性能。SPICE 包含预训练数据集中未涵盖的卤素和金属元素，以及带电分子和极性分子，因此构成了一个具有挑战性的基准。尽管如此，预训练图神经网络的性能仍优于从头开始训练的模型。平均而言，预训练图神经网络在 SPICE 上的均方根误差和平均绝对误差均降低了15.3%。这表明，即使面对不同的元素和静电相互作用，去噪预训练也有助于提升图神经网络模型的可迁移性。

最后，我们在 MD22 上对预训练模型进行微调。MD22 由更大的分子系统组成，其中原子数最多可达 370 个。这为验证在小分子上预训练的图神经网络模型的可迁移性提供了合适的基准。四种图神经网络模型在每种分子数据上的测试平均绝对误差如图3所示。

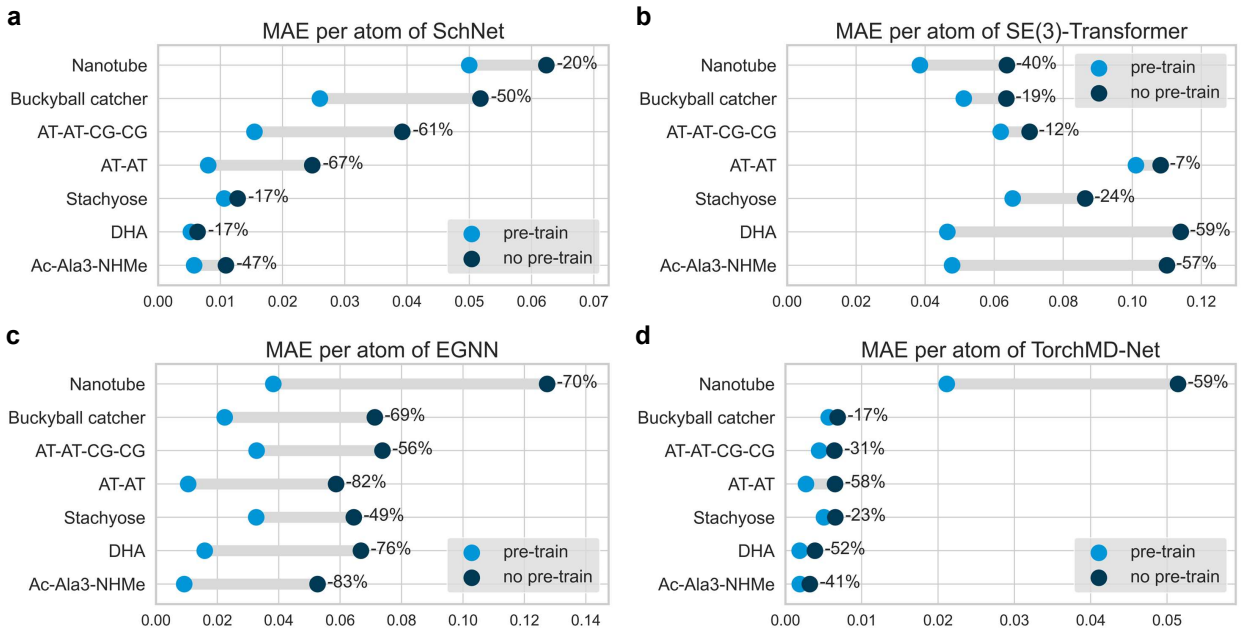


图3：在 MD22 上，(a) SchNet、(b) SE(3)-Transformer、(c) EGNN 和 (d) TorchMD-Net 使用和不使用去噪预训练时的每原子测试平均绝对误差（千卡/摩尔）。

对于 MD22 中所有不同的分子，预训练都提升了神经势能预测的性能。需要注意的是，与其他基准测试相比，MD22 包含的数据量有限，尤其是对于巴基球捕获器（6,102 条数据）和双壁纳米管（5,032 条数据）等大分子。经过在巴基球捕获器和双壁纳米管上微调的预训练 SchNet、SE(3)-Transformer、EGNN 和 TorchMD-Net，平均分别提升了 39.3% 和 47.3%。此外，在四种图神经网络中，EGNN 达到的提升最为显著。与 SchNet 和 TorchMD-Net 相比，EGNN 不包含交互层和更新层等复杂的架构设计。预训练技术大幅提升了简单等变图神经网络模型的表达能力，并取得了具有竞争力的性能。这些结果进一步验证了去噪预训练的有效性，即在小分子上预训练的图神经网络能够迁移到大型复杂系统中，用于分子势能预测。关于微调过程中神经势能预测不确定性的进一步研究，详见补充信息 S6。

总体而言，我们的实验表明，所提出的预训练方法能够提升图神经网络模型在多种分子系统中的准确性和可迁移性，使其成为一种有前景的方法，可在训练数据有限的情况下预测复杂分子系统的能量。

3.5 数据效率

为进一步评估去噪预训练对分子势能预测的益处，我们使用不同规模的数据集训练图神经网络。如图4所示，我们比较了经过预训练和未经预训练的 EGNN 在 ANI-1x 以及 MD22 中的 AT-AT、Ac-Ala3-NHMe 和巴基球捕获器数据集上的表现，数据量各不相同。具体而言，将每个数据集划分为训练集、验证集和测试集后，从训练集中抽取由 {5%、20%、50%、100%}数据组成的子集。结果表明，使用相同数量训练数据时，经过预训练的图神经网络表现优于未经预训练的图神经网络。例如，仅使用训练集总量的 5% 时，经过预训练的 EGNN 平均绝对误差为 10.04，远低于未经预训练模型的 39.37。其表现甚至优于使用多出 10 倍训练数据训练的未经预训练 EGNN。此外，如图4d所示，无论是否经过预训练，EGNN 在巴基球捕获器训练数据仅为 5%（不足 250 个样本）时表现都很差。然而，即使增加巴基球捕获器数据，未经预训练的 EGNN 的准确率也几乎没有提升，因为训练数据总量仍不足 5,000。使用量子力学方法为如此庞大而复杂的系统收集分子势能数据，成本可能非常高。另一方面，随着输入数据量增加，经过预训练的 EGNN 表现不断提升。使用全部训练数据时，经过预训练的 EGNN 的平均绝对误差比不进行预训练低 68% 以上。由此可见，在实现具有竞争力的势预测性能方面，预训练图神经网络模型比从头训练的模型更节省数据。在训练数据有限的情况下，将神经势应用于大型复杂分子系统时，这一优势尤为重要。

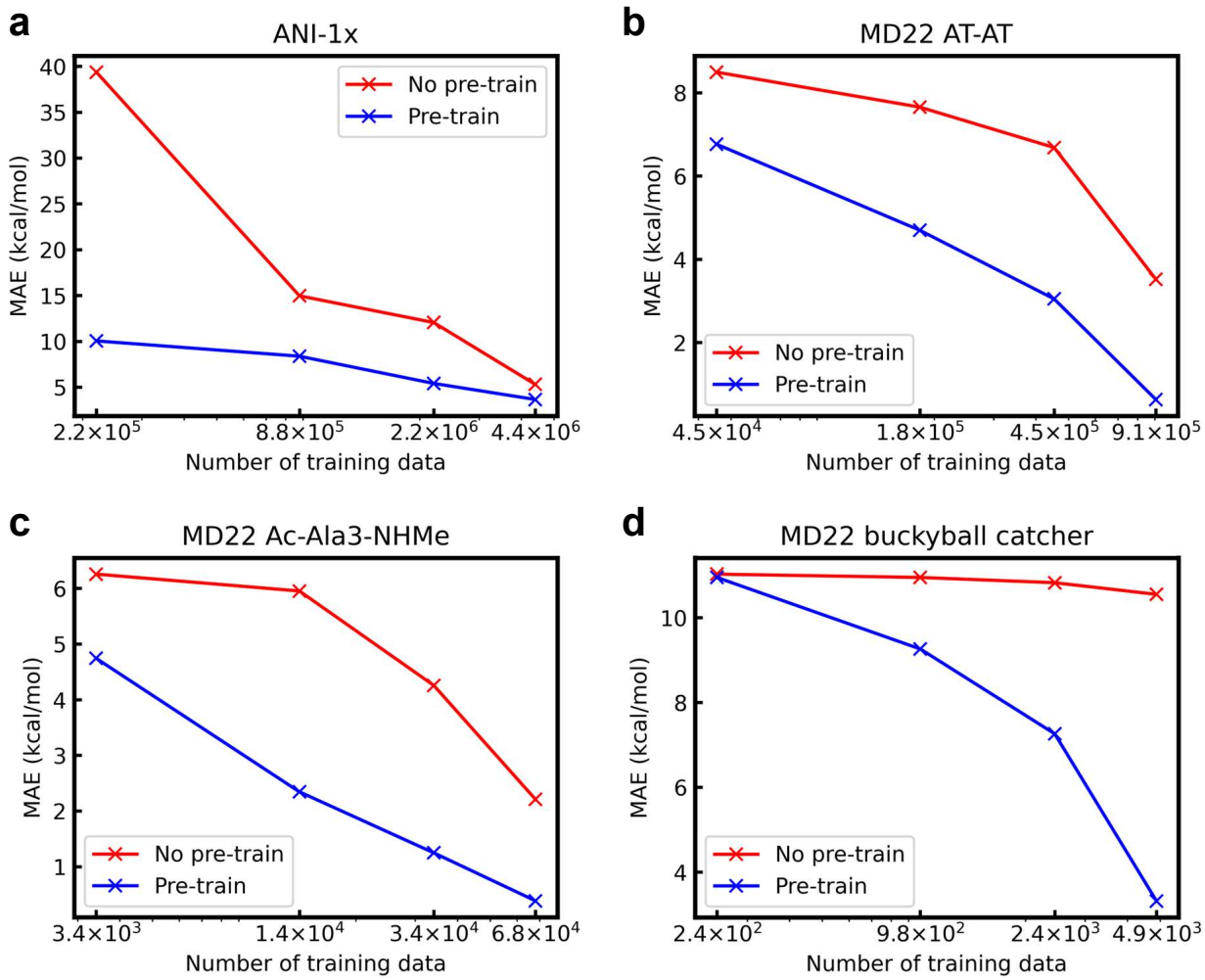


图4：EGNN 在 (a) ANI-1x、(b) MD22 AT-AT、(c) MD22 Ac-Ala3-NHMe 和 (d) MD22 巴基球捕获器数据集上，采用不同训练数据规模时的测试平均绝对误差。

3.6 噪声选择

用于扰动分子构象的噪声高斯分布标准差 σ 会影响预训练等变图神经网络的表现。扰动噪声过小，模型可能无法预测；噪声过大，则可能破坏分子构象，导致模型无法学习对分子模拟有用的信息。为选择合适的 σ，我们枚举了不同的 σ 值 $\{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0\}$ Å。图5展示了经过预训练的 TorchMD-Net 在 ANI-1x 上使用不同噪声尺度时的表现。作为对比，我们将未经预训练的 TorchMD-Net 纳入 $\sigma = 0.0$ Å。如图所示，在 RMSE 和

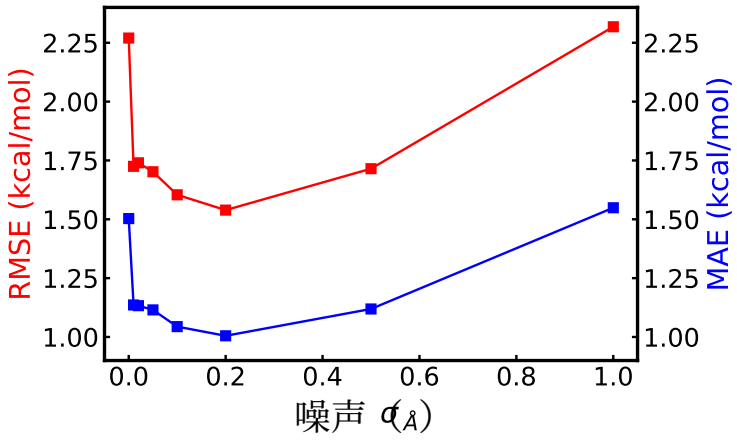


图5：预训练噪声的标准差 σ 对神经势性能的影响。展示了 TorchMD-Net 在 ANI-1x 上的测试均方根误差和平均绝对误差。

预测能量的平均绝对误差。此外，使用 0.01 至 0.5 范围内的其他噪声尺度进行预训练，也能有效提升神经势能预测。然而，当施加较大的噪声时（例如，$\sigma = 1.0\,^{\circ}\mathrm{A}$），预训练会损害势能预测，因为扰动构象远离采样构象的局部极小值。这可能是因为较大的噪声破坏了原始构象，违背了第2.3节中的假设，导致图神经网络无法学习与势能相关的有意义表征。基于实验结果，我们在其他实验中选择 $\sigma = 0.2$ 。

4 结论

总而言之，我们提出的针对非平衡分子构象上的不变且等变图神经网络（GNN）的去噪预训练方法，能够实现更准确且更具可迁移性的神经势能预测。我们在多个基准测试中开展的严格实验表明，预训练显著提高了神经势的准确性。此外，通过去噪在小分子上预训练的 GNN，在多种分子系统中展现出更强的可迁移性和数据效率，这些系统包括不同元素、极性分子、生物分子以及更大的系统。这种可迁移性对于在更大、更复杂的系统上构建神经势能模型尤为重要，因为在这些系统中，

通常很难获得充足的数据。值得注意的是，我们的预训练方法具有与模型无关的特性，这一点已由其在多种不同的不变且等变图神经网络模型上的性能提升所证实，包括 SchNet、SE(3)-Transformer、EGNN 和 TorchMD-Net。因此，我们提出的去噪预训练方法为提升神经势能预测开辟了道路，并在分子模拟的更广泛应用中展现出巨大潜力。

致谢

作者感谢卡内基梅隆大学机械工程系提供启动经费，以支持这项工作。

数据与代码的可用性

本研究使用的代码和数据可在 GitHub 代码库中找到：
`https://github.com/yuyangw/Denoise-Pretrain-ML-Potential`。

补充信息

MD22 概述、图神经网络的详细实现、微调设置详情、各图神经网络的计算效率、对微调轮次的探究以及微调不确定性均包含在补充信息中。 `https://pubs.acs.org/doi/suppl/10.1021/acs.jctc.3c00289/suppl_file/ct3c00289_si_001.pdf`

参考文献

(1) Monkhorst, H. J. 使用耦合簇方法计算性质。《国际量子化学杂志》**1977**，*12*，421–432。

(2) Hirata; Podeszwa; Tobita; Bartlett, R. J. 用于扩展体系的耦合簇单激发与双激发。《化学物理学杂志》**2004**，*120*，2581–2592。

(3) Cremer, D. Møller–Plesset微扰理论：从小分子方法到适用于数千个原子的方法。《*Wiley*跨学科评论：计算分子科学》**2011**，*1*，509–530。

(4) Levy, M. 电子密度、一阶密度矩阵和自然自旋轨道的通用变分泛函及v-可表示性问题的解决。《美国国家科学院院刊》**1979**，*76*，6062–6065。

(5) Thanthiriwatte, K. S.；Hohenstein, E. G.；Burns, L. A.；Sherrill, C. D. 评估密度泛函理论和DFT-D方法描述氢键相互作用距离依赖性的性能。*Journal of Chemical Theory and Computation* **2011**，*7*，88–96。

(6) Mark, P.；Nilsson, L. 298 K下TIP3P、SPC和SPC/E水模型的结构与动力学。《物理化学杂志*A*辑》**2001**，*105*，9954–9960。

(7) Halgren, T. A. Merck分子力场。I. MMFF94 的基础、形式、范围、参数化与性能。《计算化学杂志》**1996年**, *17*, 490–519.

(8) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. Amber生物分子模拟软件包概述。《*Wiley*跨学科评论：计算分子科学》**2013年**, *3*, 198–210.

(9) Huang, J.; MacKerell Jr, A. D. CHARMM36 全原子加和蛋白质力场：基于与核磁共振数据的比较进行验证。《计算化学杂志》**2013年**, *34*,2135–2145.

(10) Sun, H. COMPASS：一种针对凝聚相应用优化的从头算力场

概述，并详细介绍烷烃和苯化合物。物理化学杂志 *B* **1998**，*102*，7338–7364。

(11) Vitalini, F.; Mey, A. S.; No´e, F.; Keller, B. G. 力场的动态性质。《化学物理学杂志》**2015**，*142*，02B611 1。

(12) Harrison, J. A.; Schall, J. D.; Maskey, S.; Mikulski, P. T.; Knippenberg, M. T.; Mor- row, B. H. 原子尺度计算材料研究中使用的力场和分子间势综述。应用物理评论**2018年**，*5*，031104。

(13) Unke, O. T.; Koner, D.; Patra, S.; K¨aser, S.; Meuwly, M. 用于分子模拟的高维势能面：从经验方法到机器学习。机器学习：科学与技术**2020**，*1*，013001。

(14) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. 面向分子与材料科学的机器学习。《自然》**2018年**，*559*，547–555。

(15) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Sch¨utt, K. T.; Tkatchenko, A.; M¨uller, K.-R. 机器学习力场。《化学评论》**2021年**,*121*, 10142–10186。

(16) Li, Z.; Meidani, K.; Yadav, P.; Barati Farimani, A. 图神经网络加速分子动力学。《化学物理学杂志》**2022年**, *156*, 144103。

(17) Fu, X.; Wu, Z.; Wang, W.; Xie, T.; Keten, S.; Gomez-Bombarelli, R.; Jaakkola, T. 仅有力还不够：分子模拟中机器学习力场的基准与批判性评估。*arXiv* 预印本 *arXiv:2210.07237* **2022年**,

(18) Atz, K.; Grisoni, F.; Schneider, G. 基于分子表征的几何深度学习。《自然·机器智能》**2021年**，*3*，1023–1032。

(19) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Sch¨utt, K. T.; M¨uller, K.-R.

准确能量守恒分子力场的机器学习。《科学进展》**2017年**，*3*，e1603015。

(20) Chmiela, S.; Sauceda, H. E.; M¨uller, K.-R.; Tkatchenko, A. 朝向使用机器学习力场进行精确的分子动力学模拟。《自然-通讯》**2018年**，*9*，1–10。

(21) Bart´ok, A. P.; Payne, M. C.; Kondor, R.; Cs´anyi, G. 高斯近似势：无需电子即可达到量子力学的精度。*Physical Review Letters* **2010年**，*104*，136403。

(22) Bart´ok, A. P.; De, S.; Poelking, C.; Bernstein,N.; Kermode, J. R.; Cs´anyi, G.; Ceriotti, M. 机器学习统一了材料和分子的建模。《科学进展》**2017年**，*3*，e1701816。

(23) Bart´ok, A. P.; Kondor, R.; Cs´anyi, G. 化学环境的表征。《物理评论*B*辑》 **2013年**，*87*，184115。

(24) Grisafi, A.; Wilkins, D. M.; Cs´anyi, G.; Ceriotti, M. 面向原子系统张量性质的对称性适配机器学习。*Physical Review Letters* **2018年**，*120*，036002。

(25) Fedik, N.; Zubatyuk, R.; Kulichenko, M.; Lubbers, N.; Smith, J. S.; Nebgen, B.; Messerly, R.; Li,Y.W.; Boldyrev, A. I.;Barros, K., 等。将机器学习扩展到原子间势之外，以预测分子性质。《自然综述：化学》 **2022年**，*6*，653–672。

(26) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. 势能面的神经网络模型。《化学物理学杂志》 **1995年**，*103*，4129–4137。

(27) Brown, D. F.; Gibbs, M. N.; Clary, D. C. 结合从头算、神经网络和扩散蒙特卡罗：处理弱束缚分子的高效方法。《化学物理学杂志》**1996年**，*105*，7597–7604。

(28) Behler, J.; Parrinello, M. 高维势能面的广义神经网络表示。*Physical Review Letters*2007，*98*，146401。

(29) Behler, J. 用于构建高维神经网络势的原子中心对称函数。《化学物理学杂志》2011，*134*，074106。

(30) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1：一种可扩展的神经网络势，具有密度泛函理论精度，同时计算成本与力场相当。*Chemical Science*2017年，*8*，3192–3203。

(31) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. 少即是多：利用主动学习对化学空间进行采样。《化学物理学杂志》2018年，*148*，241733。

(32) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. 将 ANI 深度学习分子势的适用范围扩展至硫和卤素。*Journal of Chemical Theory and Computation* 2020，*16*，4192– 4202。

(33) Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. TensorMol-0.1 模型化学：一种融合长程物理的神经网络。*Chemical Science* 2018年，*9*，2261–2269。

(34) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. 用于量子化学的神经消息传递。国际机器学习会议。2017年；第1263–1272页。

(35) Sch¨utt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; M¨uller, K.-R. SchNet——一种用于分子和材料的深度学习架构。《化学物理学杂志》 2018年，*148*，241722。

(36) Gasteiger, J.; Groß, J.; G¨unnemann, S. 面向分子图的定向消息传递. International Conference on Learning Representations. 2020.

(37) Unke, O. T.; Meuwly, M. PhysNet：一种用于预测能量、力、偶极矩和部分电荷的神经网络。*Journal of chemical theory and computation* **2019**,*15*, 3678–3693.

(38) Lubbers, N.; Smith, J. S.; Barros, K. 使用深度神经网络对分子能量进行层次建模。《化学物理学杂志》**2018年**，*148*，241715。

(39) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. 使用分子中原子神经网络准确且可迁移地进行化学性质多任务预测。《科学进展》**2019**，*5*，eaav6490。

(40) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. 张量场网络：面向三维点云的旋转和平移等变神经网络。*arXiv* 预印本 *arXiv:1802.08219***2018年**,

(41) Anderson, B.; Hy, T. S.; Kondor, R. Cormorant：协变分子神经网络。神经信息处理系统进展**2019**，*32*。

(42) Fuchs, F.; Worrall, D.; Fischer, V.; Welling, M. SE(3)变换器：三维旋转平移等变注意力网络。神经信息处理系统进展**2020**, *33*, 1970–1981.

(43) Brandstetter, J.; Hesselink, R.; van der Pol, E.; Bekkers, E. J.; Welling,M. 几何量与物理量改进了E(3)等变消息传递。国际学习表征会议。2022年。

(44) Jing, B.; Eismann, S.; Suriana, P.; Townshend, R. J. L.; Dror, R. 利用几何向量感知器从蛋白质结构中学习。国际学习表征会议。2021年。

(45) Villar, S.; Hogg, D. W.; Storey-Fisher, K.; Yao, W.; Blum-Smith, B. 标量是

普适的：等变机器学习，其结构类似经典物理。神经信息处理系统进展。2021年。

(46) Gasteiger, J.; Becker, F.; G¨unnemann, S. Gemnet：适用于分子的通用方向图神经网络。神经信息处理系统进展**2021年**，*34*，6790–6802。

(47) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Moli- nari, N.; Smidt, T. E.; Kozinsky, B. 面向数据高效且准确的原子间势的 E (3) 等变图神经网络。《自然-通讯》**2022年**，*13*，1–11。

(48) Sch¨utt, K.; Unke, O.; Gastegger, M. 用于预测张量性质和分子光谱的等变消息传递。国际机器学习会议。2021年；第 9377–9388 页。

(49) Th¨olke,P.; De Fabritiis, G. TorchMD-NET：用于基于神经网络的分子势的等变 Transformer。*arXiv* 预印本 *arXiv:2202.02541***2022年**,

(50) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, ˙L.; Polosukhin, I. 注意力机制就是你所需要的一切。神经信息处理系统进展**2017年**, *30*.

(51) Hadsell, R.; Chopra, S.; LeCun, Y. 通过学习不变映射进行降维。2006年IEEE计算机学会计算机视觉与模式识别会议 (CVPR'06)。2006年; 第1735–1742页。

(52) Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. 一种用于视觉表征对比学习的简单框架。国际机器学习会议。2020；第1597–1607页。

(53) Hu*，W.；Liu*，B.；Gomes，J.；Zitnik，M.；Liang，P.；Pande，V.；Leskovec，J. 图神经网络预训练策略。国际学习表征会议。2020。

(54) Rong, Y.; Bian,Y.; Xu, T.; Xie, W.; Wei, Y.; Huang, W.; Huang,J. 基于大规模分子数据的自监督图 Transformer。《神经信息处理系统进展》2020年，33，12559–12571。

(55) Zhang, Z.; Liu, Q.; Wang, H.; Lu, C.; Lee, C.-K. 用于分子性质预测的基于基序的图自监督学习。《神经信息处理系统进展》2021年，34，15870–15882。

(56) Fang, X.; Liu, L.; Lei, J.; He, D.; Zhang, S.; Zhou, J.; Wang, F.; Wu, H.; Wang, H. 用于性质预测的几何增强分子表征学习。《自然·机器智能》2022年，4，127–134。

(57) Liu, Y.; Jin, M.; Pan, S.; Zhou, C.; Zheng, Y.; Xia, F.; Yu, P. 图自监督学习：综述。《IEEE 知识与数据工程汇刊》2022年,

(58) Krishnan, R.; Rajpurkar, P.; Topol, E. J. 医学与医疗保健中的自监督学习。《自然·生物医学工程》2022年,1–7.

(59) Magar, R.; Wang, Y.; Barati Farimani, A. 晶体孪生：用于晶体材料性质预测的自监督学习。《npj计算材料》2022年, 8, 1–8.

(60) Cao, Z.; Magar, R.; Wang, Y.; Farimani, A. B. MOFormer：用于金属有机框架性质预测的自监督 Transformer 模型。arXiv预印本 arXiv:2210.14188 2022年,

(61) Wang, Y.; Wang, J.; Cao, Z.; Barati Farimani, A. 通过图神经网络进行表征的分子对比学习。《自然·机器智能》**2022年**，*4*，279–287。

(62) Zhang, S.; Hu, Z.; Subramonian, A.; Sun, Y. 基于基元的图表示对比学习。*arXiv* 预印本 *arXiv:2012.12533*2020,

(63) Wang, Y.; Magar, R.; Liang, C.; Barati Farimani, A. 通过错误负样本缓解和分解片段对比改进分子对比学习。《化学信息与建模杂志》**2022年**,

(64) Liu, S.; Wang, H.; Liu, W.; Lasenby, J.; Guo, H.; Tang, J. 利用三维几何对分子图表示进行预训练。国际机器学习与表示会议。2022年。

(65) St¨ark, H.; Beaini, D.; Corso, G.; Tossou, P.; Dallago, C.; G¨unnemann, S.; Li`o, P. 三维信息最大化提升图神经网络的分子性质预测能力。国际机器学习会议。2022年；第 20479–20502 页。

(66) Zaidi, S.; Schaarschmidt, M.; Martens, J.; Kim, H.; Teh, Y. W.; Sanchez-Gonzalez, A.; Battaglia, P.; Pascanu, R.; Godwin, J. 通过去噪进行预训练以预测分子性质。*arXiv* 预印本 *arXiv:2206.00133* **2022年**,

(67) Liu, S.; Guo, H.; Tang, J. 基于 SE(3) 不变去噪距离匹配的分子几何预训练。*arXiv* 预印本 *arXiv:2206.13602* **2022年**,

(68) Zhou, G.; Gao, Z.; Ding, Q.; Zheng, H.; Xu, H.; Wei, Z.; Zhang, L.; Ke, G. Uni-Mol：通用三维分子表示学习框架。《第十一届国际学习表示会议》。2023年。

(69) Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J.-L. 化学宇宙数据库 GDB-17 中 1660 亿种有机小分子的枚举。《化学信息与建模杂志》**2012年**, *52*, 2864–2875。

(70) Satorras, V. G.; Hoogeboom, E.; Welling, M. E (n) 等变图神经网络。《国际机器学习会议》。2021年；第 9323–9332 页。

(71) Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. 图神经网络有多强大？《国际学习表征会议》。2019年。

(72) Xie, T.; Fu, X.; Ganea, O.-E.; Barzilay, R.; Jaakkola, T. S. 用于周期性材料生成的晶体扩散变分自编码器。国际学习表征会议。2022年。

(73) Arts, M.; Satorras, V. G.; Huang, C.-W.; Zuegner, D.; Federici, M.; Clementi, C.; No´e, F.; Pinsler, R.; Berg, R. v. d. 一举两得：用于粗粒化分子动力学的扩散模型和力场。*arXiv* 预印本 *arXiv:2302.00600* **2023**,

(74) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. ANI-1ccx 和 ANI-1x 数据集：分子的耦合簇和密度泛函理论性质。*Scientific Data* **2020**, *7*, 1–10。

(75) Sch¨utt, K.; Kindermans, P.-J.; Sauceda Felix, H. E.; Chmiela, S.; Tkatchenko, A.; M¨uller, K.-R. Schnet：一种用于模拟量子相互作用的连续滤波卷积神经网络。神经信息处理系统进展 **2017年**, *30*。

(76) Eastman, P.; Behara, P. K.; Dotson, D. L.; Galvelis, R.; Herr, J. E.; Horton, J. T.; Mao, Y.; Chodera, J. D.; Pritchard, B. P.; Wang, Y. 等。SPICE：用于训练机器学习势能的类药分子和肽数据集。*ScientificData* **2023**, *10*, 1–11。

(77) Chmiela, S.; Vassilev-Galindo, V.; Unke, O. T.; Kabylda, A.; Sauceda, H. E.; Tkatchenko, A.; M¨uller, K.-R. 面向含数百个原子的分子的精准全局机器学习力场。《科学进展》**2023**, *9*, eadf0873。

(78) Loshchilov, I.; Hutter, F. 解耦权重衰减正则化。*arXiv* 预印本 *arXiv:1711.05101* **2017年**,

(79) Loshchilov, I.; Hutter, F. SGDR：带热重启的随机梯度下降。国际学习表征会议。2017年。