和历史研究国际会议（ICLR 2023）。

O-gnn：将环先验纳入分子建模*

[1]**Jinhua Zhu,**[1] **Kehan Wu,**[1] **Bohan Wang,**[2] **Yingce Xia,**[3] **Shufang Xie,**[2] **Qi Meng、**
[2]**吴丽君，**[2] **秦涛，**[1] **周文刚，**[1] **李后强，**[2] **刘铁岩**[1] 中国科学技术大学，[2] 微软
研究院 AI4Science[3] 中国人民大学高岭人工智能学院
[1]{teslazhu, wu 2018}@mail.ustc.edu.cn,[1] bhwangfy@gmail.com,[1]
{zhwg,lihq}@ustc.edu.cn,[3] shufangxie@ruc.edu.cn,[2] {yingce.xia,
meq, lijuwu, taoqin, tyliu}@microsoft.com

摘要

至少含有一个环的环状化合物在药物设计中发挥着重要作用。尽管近年来利用图神经网络（GNN）建立分子模型取得了巨大成功，但很少有模型明确考虑到化合物中的环，从而限制了模型的表现力。在这项工作中，我们设计了一种新的图神经网络变体--环增强图神经网络（-GNN），除了对化合物中的原子和键进行建模外，还对环进行了明确建模。在 -GNN 中，每个环由一个潜在向量表示，该潜在向量对原子和化学键的表示有贡献，并通过原子和化学键的表示进行迭代更新。理论分析表明，-GNN 只需一层就能区分位于不同环上的两个同构子图，而传统的图卷积神经网络需要多层才能区分，这表明 -GNN 更具表现力。通过实验，-GNN 在 **11 个**公开数据集上表现出了良好的性能。特别是，它在 PCQM4Mv1 基准（优于之前的 KDDCup 冠军解决方案）和 DrugBank 上的药物相互作用预测任务上取得了最先进的验证结果。此外，在分子性质预测和逆合成预测任务中，-GNN 的表现优于强基线（无建模环）。代码发布于 https://github.com/O-GNN/O-GNN。

1 导言

环状化合物是指体系中至少有一个环的分子，自然存在于化学空间中。根据我们从广泛使用的化学库 PubChem（Kim 等人，2019 年）中对 1.09 亿个化合物的统计，90% 以上的化合物至少有一个环。这些环可能很小/很简单（*如*苯是一个六元碳环，而戊唑是一个五元氮环），也可能很大/很复杂（*如*图 1 所示的分子）。

环在药物发现中非常重要，例如：（1）环可以降低分子的灵活性，减少与目标蛋白质相互作用时的不确定性，并将分子锁定在其生物活性构象上（Sun 等人，2012 年）。(2）大环化合物通常具有 12 个原子以上的环，在抗生素设计（Venugopal 和 Johnson，2011 年）和多肽药物设计（Bhardwaj 等人，2022 年）中发挥重要作用。

最近，深度神经网络，尤其是图神经网络（GNN）（Kipf & Welling, 2017; Hamilton et

al.GNN 将图作为输入，不同节点的信息沿边传递。GNN 在科学发现方面取得了巨大成功：（1）Stokes 等人（2020 年）训练 GNN 预测大肠杆菌的生长抑制，发现卤素是一种广谱杀菌抗生素。(2) Shan 等人（2022 年）利用 GNN 对蛋白质之间的相互作用进行建模，最终获得了 SARS-CoV-2 的可能抗体。此外，GNN 还被广泛应用于药物性质预测（Rong 等人，2020 年）、药物-靶标相互作用建模（Torng 和 Altman，2019 年）、逆合成（retrosynthesis

*这项工作是朱金华、吴克勤和王博涵在微软研究院 AI4Science 实习时完成的。通讯作者：：Yingce Xia.

(Chen & Jung，2021 年）等。然而，上述工作都没有明确地将环信息建模到 GNN 中。从应用的角度来看，它们错过了任务的一个重要特征。Loukas (2020) 从机器学习的角度指出，当网络宽度和高度的乘积不够大时，现有的基于消息传递的 GNN 无法正确捕捉环信息（见 Loukas (2020) 中的表 1）。因此，传统的 GNN 无法很好地利用化合物中的环状信息。

为了解决这个问题，我们在这项工作中提出了一种新的模型--环增强 GNN（简称-GNN），它明确地模拟了化合物中的环信息。环代表分子中的环，发音为 "O"。一般来说，-GNN 堆叠了 L 层、每一层依次更新边缘表示、我们主要使用自注意层来进行自适应信息传递，并使用前馈层来引入非线性表征。我们主要使用自注意层进行自适应信息传递，并使用前馈层为表征引入非线性。

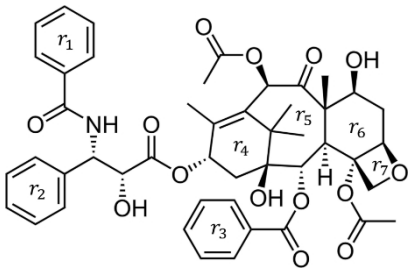


图 1：紫杉醇，一种具有 7 个单环的化合物。Kampan 等人（2015 年）总结说，完整的紫杉烷环（即 r_4 , r_5 , r_6 ）和一个四元环氧乙烷侧环（即 r_7 ）对诱导细胞毒性活性。

我们首先通过理论分析证明-GNN 的优势。-GNN仅用一层就能区分位于不同环上的两个同构子图（示例见图 2）。相反，如果我们将环建模组件从-而 O-GNN 则需要多个层才能实现这种可区分性（详细分析见第 2.3 节）。这些结果表明，O-GNN 比传统的图卷积网络没有明确的环建模。

然后，我们在三个任务的 11 个数据集上进行了实验，包括分子特性预测、药物-药物预测、药物-药物预测和药物-药物预测。

相互作用预测和逆合成：

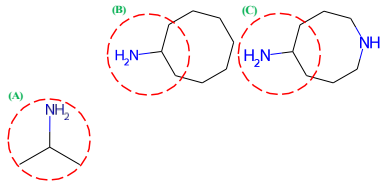


图 2：理论结果示例。红圈中的三个子结构是同构的。第二和第三个子结构位于不同的环（一个环辛烷和一个氮杂环辛烷）上。常规 GNN 需要多层结构来区分这三个子结构，而 -GNN 由于环的代表关系只需要一层结构。

(1) 在分子性质预测方面，我们首先在 PCQM4Mv1 上进行实验，即预测分子的 HOMO-LUMO 间隙。我们的方法在验证集上优于 KDDCup 的冠军解（Shi 等人，2022 年）（注意测试集标签不可用）。接下来，我们在来自 MoleculeNet（Wu 等人，2018 年）的六个数据集上验证了 -GNN，该数据集用于预测分子的若干制药相关特性。结果表明，-GNN 优于相应的无环 GNN 基线。最后，我们在FS-Mol（Stan- ley et al.(2) 对于药物相互作用预测，即预测两种药物是否相互作用

另外，我们按照之前的设置（Nyamabo 等人，2021 年；Li 等人，2022 年）在 DrugBank 上测试了 -GNN，并取得了最先进的结果。(3) 在回溯合成方面，我们将 -GNN 应用于

和历史研究国际会议（ICLR 2023）。

LocalRetro (Chen & Jung, 2021)，这是一种基于 GNN 的强大回溯合成方法。在 USPTO-50k 上，我们的方法显著提高了准确率。

2 方法

2.1 符号和前言

让 G = (V, E) 表示分子图，其中 V 和 E 是节点/原子和边/键的集合。[1].让 R 表示 G 中的环集合。定义 $V = \{v, v_{12}, - - -, v_{|V|}\}$ 和

[1]在上下文明确的情况下，我们在本文中交替使用节点/原子和边/键。

和历史研究国际会议（ICLR 2023）。

$E = \{e_{ij}\}$，其中 v_i 是第 i 个原子，e_{ij} 是连接 v_i 和 v_j 的键。上下文清楚时，我们用 i 表示原子 v_i，用 $e(v_i, v_j)$ 表示边 e_{ij}。让 $\mathcal{N}(i) = \{j \mid e_{ij} \in E\}$ 表示原子 i 的邻域，即 $\mathcal{N}(i) = \{j \mid e_{ij} \in E\}$。定义 $R = \{r_1, r_2, \ldots, r_{|R|}\}$，其中每个 r_i 都是一个简单环。单环不包含任何环状结构。例如，对于图 3 中的分子，我们标出了两个简单环（r_1 和 r_2）。环 (1, 2, 3, 4, 5, 6, 7, 8, 9, 1) 不是简单环。让 $R(v_i)$ 和 $R(e_{ij})$ 表示原子 v_i 或键 e_{ij} 所在的环，V(r) 和 E(r) 表示环 r 上的所有原子和键。例如，在图 3 中，$R(v_4) = \{r_1, r_2\}$，而 $R(v_3) = r_2$。 $R(e_{49}) = \{r_1, r_2\}$，而 $R(e_{78}) = r_1$。 $V(r_1) = \{v_4, v, v, v_{567}, v_8, v_9\}$，而 $E(r_1) = \{e, e, e, e, e_{455667788994}\}$，e。

图神经网络（GNN）通常由多个相同的 GNN 层堆叠而成。每个 GNN 层都由一个聚合函数和一个更新函数组成、

$$h_i' = \text{Update}(h_i, \text{Aggregate}(h_j \mid j \in \mathcal{N}(i))), \tag{1}$$

其中 h_i 是原子 i 的表示，h' 是其更新后的表示。不同的 GNN 有不同的聚合函数和更新函数。附录 D 概述了相关细节。

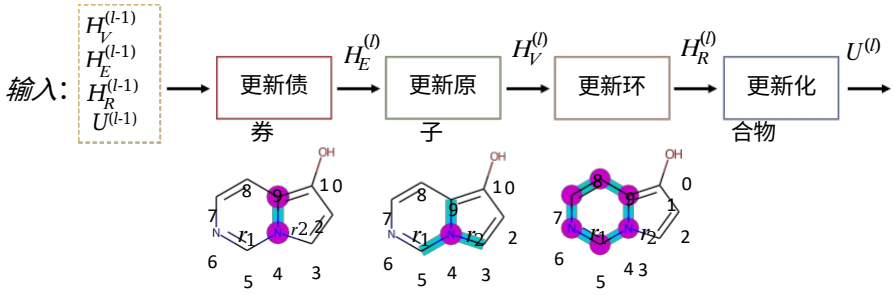


图 3：我们的方法的工作流程。$H^{(l)}$，$H_E^{(l)}$，$H_V^{(l)}$ 和 $U_R^{(l)}$ 表示收集到的表征。键、原子、环和第 l 层的整体化合物。

2.2 型号

我们的模型由具有不同参数的 L 个相同层组成。每一层的结构如下如图 3 所示。让 $h_i^{(l)}$、$h_{ij}^{(l)}$ 和 $h_r^{(l)}$ 分别表示原子 v、键 e 和键 f 的输出表示。h_i h_{ij} 分别表示第 l 层的环 r。让 $U^{(l)}$ 表示第 l 层的化合物表示。我们通过可学习嵌入层初始化 $h^{(0)}$，该嵌入层表示其原子类型、手性、度数、形式电荷、杂化类型等。同样，我们通过可学习嵌入层对 $h^{(0)}$ 进行初始化，可学习嵌入层表示其键类型、立体异构类型以及键是否共轭。

然后，我们通过连接节点嵌入和边缘嵌入来初始化 $h^{(0)}$，再用非线性层对其进行转换。最后，我们用可学习的嵌入来初始化化合物表示。在每一层中，我们依次更新节点、键、环和化合物的表示。我们将经常使用 MLP( )（一种具有一个隐藏层的多层感知网络）来构建我们的模型。MLP 的输入被串联成一个长向量，并由网络进行处理。

(1) *更新键的表示*：通过连接的原子、键所属的环和上一层的化合物表示更新键的表示：

$$h_{ij}^{(l)} = h_{ij}^{(l-1)} + \text{MLP}\left(h_{ij}^{(l-1)}, h_i^{(l-1)}, h_j^{(l-1)}, \sum_{r \in R(e_{ij})} \frac{h_r^{(l-1)}}{|R(e_{ij})|}, U^{(l-1)}\right), \tag{2}$$

(2) *更新原子表征*：我们使用注意力模型将键代表聚合到集中原子中。数学方法

和历史研究国际会议（ICLR 2023）。

$$
\begin{aligned}
\bar{h}_i^{(l)} &= \sum_{j \in N(i)} \alpha_{ij} W_v \, \mathrm{concat}(h_{ij}^{(l)}, h_j^{(l-1)}); \\
\alpha_j &\propto \exp(\boldsymbol{a}^{\top} \, \mathrm{LeakyReLU}(W h_i^{(l-1)} + W_k \, \mathrm{concat}(h_j^{(l-1)}, h_{ij}^{(l)}))); \\
h_i^{(l)} &= h_i^{(l-1)} + \mathrm{MLP}\left(h_i^{(l-1)}, \bar{h}_i^{(l)}, \frac{1}{|R(v_i)|} \sum_{r \in R(v_i)} h_r^{(l-1)}, U^{(l-1)}\right).
\end{aligned}
\tag{3}
$$

和历史研究国际会议（ICLR 2023）。

在公式（3）中，W 表示要学习的参数，concat 表示将输入向量连接成一个长向量。

(3) *更新环状表征*：使用 MLP 网络更新环状表征：

$$h_r^{(l)} = h_r^{(l-1)} + \mathrm{MLP}\left(h_r^{(l-1)}, \sum_{v_i \in V(r)} h_i^{(l)}, \sum_{e_{ij} \in E(r)} h_{ij}^{(l)}, U^{(l-1)}\right) \quad (4)$$

(4) *更新复合表示法*:

$$U^{(l)} = U^{(l-1)} + \mathrm{MLP}\left(\frac{1}{|V|}\sum_{i=1} h_i^{(l)}, \frac{1}{|E|}\sum_{i,j} h_{ij}^{(l)}, \frac{1}{|R|}\sum_{r \in R} h_r^{(l)}, U^{(l-1)}\right), \quad (5)$$

堆叠 L 层 O-GNN层后，我们通过一个简单的平均池化层得到图表示，即 $h_G = \frac{1}{|V|}\sum h_{|V|}^{(L)}$，可用于图分类任务。对于节点分类我们可以在 $h^{(L)}$ 上添加一个分类头。

2.3 理论分析

在本节中，我们将比较标准 GNN（无环表示）和 O-GNN 之间的可区分性。除了第 2.1 节中定义的符号外，我们还将图 G = (V, E) 的有值版本定义为三元组 $\mathrm{VALUE}_f(G) = (V, E, f)$，其中 f 是存储特征信息的映射，并将节点或边映射到相应的输入特征（例如 256 维表示）。我们称 f 为 G 上的特征映射。

定义 1（k-邻接节点）。*对于分子图* G = (V, E) *和两个节点* u, v ∈ V，*如果* G *中存在一条连接* u *和* v *的长度不大于* k *的路径，我们就说* u *是* v *的* k *邻域。更正式地说，当且仅当存在一组节点* $\{v, v_{01}, \ldots, v_t\} \subset V$，*使得* $t \le k$，$v_0 = v$，$v_t = u$，*并且对于任意* $i \in \{0, \ldots, t-1\}$，$v_{i+1} \in N(v_i)$ *时，u 才是 v 的* k 邻域。

我们在此强调，v 是一个 0 邻接节点（因此也是一个 k 邻接节点，与任何 $k \ge 0$）的本身。

定义 2（k 邻域子图）。*对于分子图* G = (V, E) *和* G *中的节点* v，*我们将* v *的* k *邻域子图定义为由所有* v *的* k 邻域 *节点组成的子图。更正式地说，我们稍微滥用一下符号，将* v 的 k *邻域子图表示为* $G(v, k) \triangleq (V(v, k), E(v, k))$，*其中*

$$V(v, k) \triangleq \{u \in V : u \text{ 是 } v \text{ 的 } k \text{ 邻节点}\},\ E(v, k) \triangleq \{e(v_1, v_2) \in E : v_1, v_2 \in V(v, k)\}。$$

定义 3（等价图）。*对于两个有值图* $\mathrm{VALUE}_{f_1}(G_1) = (V_1, E_1, f_1)$ *和* $\mathrm{VALUE}_{f_2}(G_2) = (V_2, E_2, f_2)$，*我们说在以下情况下它们是等价的：(i).* G_1 *和* G_2 *是同构的，即存在一个一一对应的映射* $P : V_1 \to V_2$，*从而保留了边；(ii).*P *还保留了边的值和节点的值，即* ∀u, v ∈ G、1

$$e(u, v) \in E_1 \Leftrightarrow e(P(u), P(v)) \in E、2$$
$$f_1(u) = f_2(P(u)),\ f_1(v) = f_2(P(v)),\ f_1(e(u, v)) = f_2(e(P(u), P(v))).$$

有了上述准备工作，我们现在就可以定义图形特征提取器及其判别能力了。

定义 4（图特征提取器及其判别能力）。*如果一个映射* Φ *能将一个有值图* $\mathrm{VALUE}_f(G)$ 映射*到* G *上的一个新特征映射* $\tilde{f}$，*我们就称该映射为图特征提取器。我们还允许将* Φ *参数化为* Φ_θ，*并称* Φ_θ *为参数化图特征提取器。*

θ*对于参数化图特征提取器* Φ_θ，*如果对于任意有值图* (G，f) *和* G *中任意两个节点* u、

和历史研究国际会议（ICLR 2023）。

v，*如果* u *和* v 的*有值* k *邻域子图（即*G（u，k），f）*和*（G（v，k），f））*等价，则存在* $\theta^{\star}$ *，使得* $\Phi_{\theta^{\star}}((G, f))(u) = \Phi_{\theta^{\star}}((G, f))(v)$ *。在这种情况下，我们也可以说* $\Phi_{\theta^{\star}}$ *可以区分* u *和* v*。*

和历史研究国际会议（ICLR 2023）。

我们指出，由公式 (2, 3, 4, 5) 定义的 $\{h\}^{(l)}_i \cup \{h\}^{(l)}_{i,j}$ 是一个参数化的特征提取器、因此，以上给出了 O-GNN 判别能力的正式定义。

下一个命题表明，如果没有环表示，OGNN 至少需要 $k+1$ 层对 k 邻域子图的判别能力。

命题 1. *如果没有环状呈现，层数不超过 k 的 O 对 k 邻域子图不具有判别能力。*

需要注意的是，命题 1 可以很容易地扩展到传统的图卷积神经网络，因为传统的图卷积神经网络只汇聚 1 邻节点的信息。然后我们将证明，在环状表示法下， ，只有一层的 O-GNN 具有判别能力。

命题 2. *如果 u 和 v 位于不同的环上，只有一层的 O-GNN 可以将它们区分开来。*

由于篇幅所限，证明将放在附录 B 中。从命题 1 和 2 中我们可以看出，O-GNN 比不建模环的常规 GNN 更具表现力。常规 GNN 至少需要 k 层才能区分不同环上的两个同构 k 邻域子图，而 O-GNN 只需要一层就能做到（见图 2 中的示例）。将 O-GNN 与具有相同层数的普通 GNN 相比，环演示建模会不断增加参数的百分比（与 k 无关）。然而，普通 GNN 可能需要 k 层才能达到对 k 邻域子图的判别能力。当 k 为 $\{6, 8, 10, 12\}$ 时，因此，O-GNN 的参数效率要高得多。更多讨论见附录 C.5。

3 实验

为了验证我们方法的有效性，我们在以下三个任务中测试了 O-GNN：分子性质预测、药物相互作用预测和逆合成。前两个任务是图分类任务，第三个任务是节点/链接预测任务。

3.1 应用于分子特性预测

数据集。我们在此应用中使用了三个数据集：

(1) PCQM4Mv1 数据集的 HOMO-LUMO 能隙预测（Hu 等人，2021 年）。输入是一个二维分子图，目标是其 HOMO-LUMO 能隙，这是量子化学中的一个基本分子性质。PCQM4Mv1 有 3045360 和 380670 个训练和验证数据（没有测试标签）。这些属性是通过密度函数理论获得的。

(2) MoleculeNet 数据集的分子性质预测（Wu 等人，2018 年）。这是一个关于小分子药物性质预测的数据集。我们选择了六个分子性质预测任务（包括 BBBP、Tox21、ClinTox、HIV、BACE 和 SIDER），数据量从 1.5k 到 41k。

(3) 对 FS-Mol 数据集（Stanley 等人，2021 年）进行分子特性预测。FS-Mol 数据集由从 ChEMBL27（`https://www.ebi.ac.uk/ chembl/`）中提取的 5120 个独立检测项目组成。每项检测平均有 94 对分子特性对。

训练配置。对于 PCQM4Mv1，我们设置层数为 12，隐藏维数为 256，这是在训练集上通过交叉验证方法选择的。对于 FS-Mol，层数为 6，隐藏维数为 256。MoleculeNet 的候选层数和隐藏维数分别为 4、6、8、12 和 $\{128, 256\}$。在 FS-Mol 和 MoleculetNet 上，超参数是根据验证性能选择的。我们在以下平台上训练所有这些任务

和历史研究国际会议（ICLR 2023）。

一个 GPU。优化器是 AdamW（Loshchilov & Hutter，2019 年）。更多详细参数见附录 A 表 5。

PCQM4Mv1 的结果 PCQM4Mv1 的结果见表 1。我们将 GNN 与以下基线进行了比较：(1) 有/无虚拟节点的传统 GCN/GIN（用 "vn "标记）。结果来自 Hu 等人 (2021)；(2) ConfDSS (Liu 等人，2021)，它以低成本构象集为条件预测量子特性；(3) Two-branch Transformer (Xia 等人，2021)，它有一个回归头和一个分类头，这两个头相互学习；(4) Graphormer（Ying 等人，2021 年；Shi 等人，2022 年），PCQM4Mv1 的冠军解决方案。

方法	MAE (↓)
GCN	0.1684
GCN + vn	0.1510
GIN	0.1536
GIN + vn	0.1396
ConfDSS	0.1278
双支变压器	0.1237
$\text{Graphormer}_{\text{base}}$	0.1193
$\text{Graphormer}_{\text{large}}$	0.1231
O-GNN （我们的）	0.1148

表 1：PCQM4Mv1 的验证 MAE.

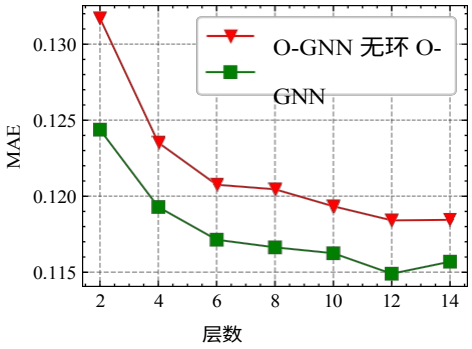
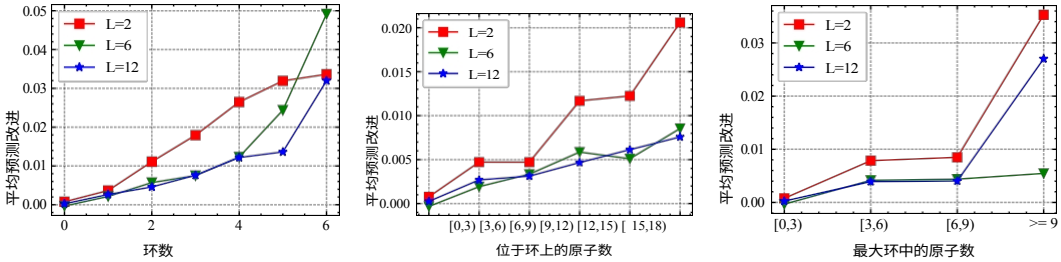


图 4：MAE 随层数变化。



(a) 环数 (b) 位于环上的原子数。 (c) 环的最大尺寸。

图 5：在多个环特性上的性能改进。

PCQM4Mv1 的所有者没有公布测试集的标签，因此我们只能比较验证集的 结果。评估指标是平均绝对误差（MAE）。从表 1 中可以看出，在强基线模型中，-GNN 的结果最好，这说明了我们方法的有效性。此外，GIN vn、ConfDSS 和 Graphormer 都没有明确使用环信息，今后我们将把 O-GNN 与强方法结合起来。

为了研究环状信息的重要性，我们研究了-GNN 的一种变体，即去掉- G N N 中的环状建模部分，并将这种变体命名为"-GNN w/o ring"。具体来说，它是通过删除式（4）和式（2,3,5）中的所有 h_r's来实现的。我们对从 2 层到 14 层的"-GNN "和"-GNN（无环）"。结果如图 4 所示。我们可以看到通过利用环信息，无论层数多少，性能都得到了提升。此外，我们还发现，6 层的 -GNN 与 12 层的 -GNN （无环）性能相当，这显示了在 GNN 中建立环模型的巨大威力。我们还发现，就参数数量而言，-GNN 优于 "无环 -GNN"（见图 10）。值得注意的是，14 层 -GNN 的验证 MAE 与 12 层 -GNN 相比略有下降。请注意，在 Graphormer（Shi 等人，2022 年）中也观察到了这一现象，即更大的模型并不总能带来更好的验证结果。我们将在未来探索如何训练更深层次的模型。

在 PCQM4Mv1 上，我们还研究了与几个环属性相关的平均性能改进。性能改进定义为 ϵ_2/ϵ_1，其中 ϵ_1 和 ϵ_2 分别表示"-GNN（无环）"和"-GNN（无环）"的验证 MAE。环属性包括 (i) 分子中的环数；

(ii) 位于环上的原子数；(iii) 最大环上的原子数。我们对不同层数（L = 2、6、12）的网络

和历史研究国际会议（ICLR 2023）。

进行了实验。结果见图 5。我们可以得出结论：总体而言，随着环数、最大环大小和环上原子数的增加，-GNN 与不对环建模的变体相比取得了更大的改进。更多分析见附录 C.4。

关于 MoleculeNet 的结果 对于 MoleculeNet，我们与预训练和非训练方法进行了比较。对于非预处理方法，我们与以下基线进行了比较：(i) 带虚拟节点的 GCN（Kipf & Welling，2017 年）；(ii) 带虚拟节点的 GIN（Xu 等人，2018 年）；(iii) 不使用环信息的 O-GNN（表示为 "O-GNN w/o ring"）。对于预训练方法，我们选择

数据集 # 分子	BBBP 2039	Tox21 7831	ClinTox 1478	HIV 41127	BACE 1513	SIDER 1478
(Hu 等人，2020 年)	71.2 ± 0.9	74.2 ± 0.8	73.7 ± 4.0	75.8 ± 1.1	78.6 ± 1.4	60.4 ± 0.6
G-Contextual (Liu et al., 2022)	70.3 ± 1.6	75.2 ± 0.3	59.9 ± 8.2	75.9 ± 0.9	79.2 ± 0.3	58.4 ± 0.6
G-Motif （Liu 等人，2022 年）	66.4 ± 3.4	73.2 ± 0.8	77.8 ± 2.0	73.8 ± 1.4	73.4 ± 4.0	60.6 ± 1.1
GraphMVP （Liu 等人，2022 年）	72.4 ± 1.6	75.9 ± 0.5	79.1 ± 2.8	77.0 ± 1.2	81.2 ± 0.9	63.9 ± 1.2
GCN + vn	72.7 ± 1.3	75.0 ± 0.4	92.0 ± 1.1	78.8 ± 1.1	80.0 ± 0.8	62.9 ± 1.3
GIN + vn	71.7 ± 0.6	74.8 ± 0.6	89.4 ± 3.2	79.3 ± 1.0	82.0 ± 1.0	60.8 ± 0.8
无环 O-GNN	74.5 ± 1.4	75.2 ± 0.9	90.2 ± 2.1	80.5 ± 1.0	84.2 ± 1.5	65.5 ± 1.6
O-GNN （我们的）	76.4 ± 0.4	75.7 ± 0.7	94.3 ± 1.6	81.3 ± 1.2	85.8 ± 1.0	66.2 ± 1.2

表 2：不同方法在 MoleculeNet 基准的 6 个二元分类任务上的测试 ROC-AUC (%) 性能。训练集、验证集和测试集由 DeepChem 提供。每个实验独立运行三次。报告了平均值和标准推导值。

几种有代表性的基于图的方法：(i) Hu 等人（2020 年）提出在图上预测屏蔽属性，并保持子图与其相邻图之间的一致性；
(ii) G-Contextual、Motif 是（Rong 等人，2020 年）的变体，由 Liu 等人（2022 年）提供。(iii) GraphMVP（Liu 等人，2022 年）是二维分子与其三维构象之间的联合预训练。(Hu 等人，2020)、G- Contextual、Motif 和 GraphMVP 的结果均摘自 Liu 等人 (2022)，因为 Liu 等人 (2022) 与我们使用相同的基于支架的拆分方法。

结果见表 2。我们可以看到(i) -GNN 优于传统网络架构，如带有虚拟节点的 GIN 和 GCN，这证明了我们的新架构的有效性；(ii) 与 G-Contextual、Motif、GraphMVP（Liu 等人，2022 年）和 Hu 等人（2020 年）这些预训练方法相比，-GNN 仍然优于这些方法。(关于预训练方法的更多讨论见附录 C.5 表 11）这表明 -GNN 具有巨大的潜力，我们将在未来把它与预训练方法结合起来。(iii) 对比-GNN 和-GNN（无环），六项任务的平均改进幅度为 1.6。
这显示了在分子特性预测中使用环信息的优势。

对 FS-Mol 的研究结果斯坦利等人（2021 年）证实，原型--与 MAML（Finn 等人，2017 年）、多任务学习（MT）和随机森林（RF）等其他方法相比，古典网络（PN）在 FS-Mol 上的表现最好。Stanley 等人（2021 年）使用类似于 Transformer 的残差网络来进行少镜头分类。我们将该骨干网络替换为"-GNN "和"-GNN w/o ring"，其他部分保持不变。按照 Stanley 等人（2021 年）的方法，报告了不同支持集大小（用 $_{u,support}$ 表示）的结果。一个支持集由几个输入标签对的示例组成，用于训练模型。评估指标是 ΔAUPRC，即 AUPRC（精确度-召回曲线下的面积）与该查询集中有效化合物比率之间的差值。越高 ΔAUPRC 分数表明模型的分类性能更好。

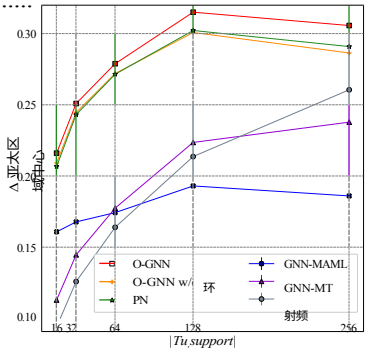


图 6：FS-Mol 的结果。

和历史研究国际会议（ICLR 2023）。

结果见图 6。我们报告了不同任务在不同支持大小下的平均值和标准推导值。我们有以下观察结果：(i) 通过使用 -GNN 作为原型网络的骨干模型，不同支持集大小的结果都得到了提升。

(ii) 当支持集规模较大时，改进效果更为显著。当 $_{u,support}$ = 128 和 256 时，改进幅度分别为 0.014 和 0.016。当支持集大小减小到 16/32/64 时，它们的改进幅度都在 0.008 左右。今后，我们将进一步改进有限数据量下的结果。

3.2 应用于 DDI 预测

药物相互作用（DDI）预测是指预测两种药物同时服用的治疗效果，如增加某些副作用的风险，或增强同时服用两种药物的效果。我们

和历史研究国际会议（ICLR 2023）。

分类任务的输入是两个药物分子和一种相互作用（如抑制），输出为 0 或 1，表示这两种药物是否具有这种特定的相互作用。继 Nyamabo 等人（2022 年）和 Li 等人（2022 年）之后，我们在 DrugBank 数据集（Wishart 等人，2018 年）的归纳设置上开展工作，该数据集有 1 706 种药物、86 种相互作用类型和 191 808 个三联体。为了测试该模型的泛化能力，我们在以下两种情况下进行了实验

在 S1 设置中，测试集中的两种药物都没有出现在训练集中；在 S2 设置中，一种药物出现在训练集中，另一种则没有。注意到测试集中的药物对没有出现在训练集中。因此，DrugBank 数据按药物的可见性分为训练集和测试集，并离线生成负样本。我们直接使用 Nyamabo 等人（2021；2022）提供的数据，其中先保留 20% 的药物作为未见过的药物来制定测试集，其余 80% 的药物用于创建训练集。

	S2 设置（1 种已知药物 + 1 种未知药物）				S1 设置（2 种未知药物）			
方法	**ACC**	**AUROC**	**AP**	**F1**	**ACC**	**AUROC**	**AP**	**F1**
GAT-DDI （Nyamabo 等人，2021 年）	69.83	77.29	75.79	73.01	62.63	70.92	73.01	45.81
MHCADDI (Deac et al., 2019)	70.58	77.84	76.16	72.74	65.40	73.43	75.03	54.12
MR-GNN (Xu et al., 2019)	74.67	83.15	83.81	69.88	66.50	72.53	71.06	67.21
SSI-DDI （Nyamabo 等人，2021 年）	76.38	84.23	84.94	73.54	66.31	72.75	71.61	68.68
GMPNN （Nyamabo 等人，2022 年）	77.72	84.84	84.87	78.29	68.57	74.96	75.44	65.32
MSAN-GCN （Zhu 等人，2022 年）	77.81	85.74	-	76.48	69.17	76.12	-	67.10
MSN-DDI （Li 等人，2022 年）	81.92	91.01	91.09	80.18	73.42	81.79	81.82	70.34
无环 O-GNN	87.72	94.51	95.28	85.91	75.47	83.83	85.58	65.59
O-GNN （我们的）	88.47	95.87	96.51	86.91	76.81	87.64	88.70	70.81

表 3：DrugBank 上的药物相互作用预测结果。

为了预测两种药物之间的相互作用，我们使用一个 6 层 -GNN 来提取两种药物的特征。具体来说，对于每种药物，我们将最后一层输出的节点表示平均值作为药物特征。我们将两种药物特征连接在一起，然后乘以交互嵌入进行预测。具体参数见附录 A 表 6。

结果见表 3。 -在准确度（ACC）、接收者操作特征下面积（AUROC）、平均精度（AP）和 F1 分数方面，GNN 明显优于之前的基线。以前的研究大多使用 GCN、GIN 或 GAT 骨架，并侧重于设计全面的交互模块（Nyamabo 等人，2021 年；Li 等人，2022 年）。通过使用先进的 -GNN 主干网，我们可以在不设计复杂交互模块的情况下显著改善结果。这表明了我们方法的有效性。

3.3 应用于逆合成

逆合成是预测给定产物的反应物。各种 GNN 已被应用于这项任务。例如，GLN（Dai 等人，2019 年）使用 GNN 预测候选反应模板和反应物的分布。GraphRetro（Somnath 等人，2021 年）和 G2G（Shi 等人，2020 年）使用 GNN 预测在何处断键以及如何添加片段以完成合成。为了证明我们的 能力，我们将我们的方法与 LocalRetro（Chen & Jung，2021 年）相结合后者是目前最好的基于图的逆合成模型（不使用预训练）。

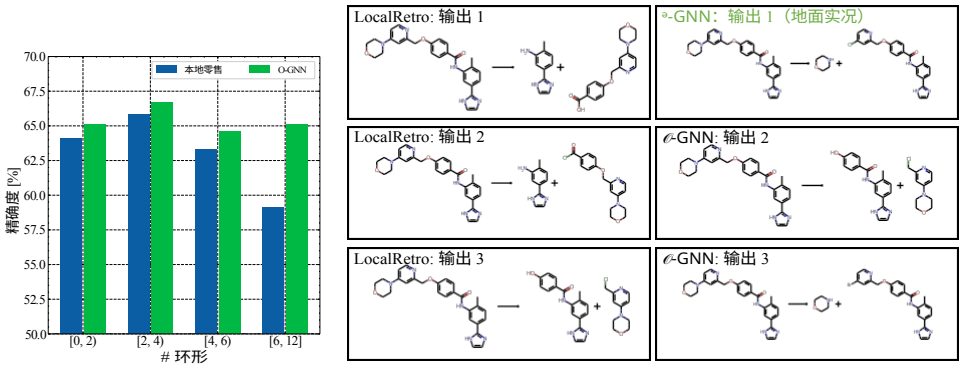
和历史研究国际会议（ICLR 2023）。

LocalRetro 使用 GNN 预测每个原子和每个键的可能模板，并根据概率对预测模板进行排序。顶部模板将通过 RDKit（Landrum 等人，2016 年）应用于对应的原子或键，生成反应物。Chen & Jung（2021 年）使用 MPNN（Gilmer 等人，2017a）进行预测，我们则用 GNN 取代 MPNN。我们在 USPTO-50k 数据集（Coley et al.
行动按照 Chen 和 Jung（2021 年）的方法，我们将数据集划分为 45k 训练集、5k 验证集和 45k 训练集。
和 5k 测试集。评价指标是前 k 级准确率，其中 k = 1、3、5、10、50。结果汇总于表 4。
我们可以看到，GNN 比没有环信息的基线预测反应更准确。特别是当反应类型已知时，我们
前 1 名的精度提高了 1.8 个点，前 3 名的精度提高了 1.6 个点。这些结果表明了环状结构建模的重要性和我们方法的有效性。

不同环数的性能。为了研究不同环数分子的预测性能，我们按照产品分子中的环数对 USPTO-50k 测试集进行分组，并计算每组的前 1 级准确率。更具体地说，我们将

方法	反应类型未知 前 1 名	前三名	前五名	前 10 名	前 50 名	反应类型已知 前 1 名	前三名	前五名	前十名	前 50 名
G2G	48.9	67.6	72.5	75.5	-	61.0	81.3	86.0	88.7	-
GLN	52.5	69.0	75.6	83.7	92.4	64.2	79.1	85.2	90.0	93.2
GraphRetro	53.7	68.3	72.2	75.5	-	63.9	81.5	85.2	88.1	-
本地零售	53.4	77.5	85.9	92.4	97.7	63.9	86.1	92.4	96.3	97.9
O-GNN （我们的）	54.1	77.7	86.0	92.5	98.2	65.7	87.7	93.4	96.9	98.3

表 4：反应类型已知/未知的美国专利商标局 50k 数据集的结果。



(a) 与 #环有关的精确度。　　(b) 具有复杂环的分子案例研究。

图 7：-GNN 对逆合成任务的研究。(a) 根据产品分子中的环数计算的前 1 级准确率。(b) 具有五个环的产品分子的一步逆合成预测。第一个 O-GNN 输出与地面实况相同（绿色标记）。

将测试集分为四组，环号分别为 [0，2)、[2，4)、[4，6)、[6，12]，这四组分别有 808、2347、1617、235 个反应。结果见图 7(a)，其中蓝色条代表 LocalRetro 基线，绿色条代表 -GNN。结果显示，-GNN 在所有组别上都有更好的准确性，分别提高了 0.99、0.85、1.30 和 5.96。总的来说，当分子中的环越多时，改进幅度就越大。特别是当一个组中至少有 6 个环（即最后一列）时，-GNN 会提高对以下情况的准确度 5.96 分，表明我们的方法可以更好地利用环形结构。

案例研究。图 7(b)显示了对一个有 5 个环的产物分子的预测示例。左侧面板中的反应是 LocalRetro 基线预测的前 3 个反应，右侧面板中的反应是 -GNN 预测的前 3 个反应。我们的方法在第一次输出中成功预测出了正确的反应物（绿色标记），而基线系统却未能给出正确的预测。更重要的是，基线系统甚至无法识别需要改变的正确化学键。这些结果表明，环结构建模对于准确预测反应至关重要， O-GNN 是一种有效的逆合成算法。

4　结论和今后的工作

在这项工作中，我们提出了一种用于分子建模的新模型--环增强 GNN（简称-GNN）。我们明确地将环表示纳入 GNN，并与原子和键表示共同更新。我们提供了 -GNN 的理论分析，并证明通过使用

和历史研究国际会议（ICLR 2023）。

O-与不使用环状表征的变体相比，-GNN 的节点表征更容易区分。我们在分子性质预测、药物相互作用（DDI）预测和O逆合成方面进行了实验。-在这些任务中，-GNN 的表现优于强基线，并在 PCQM4Mv1 和 DDI 预测的验证性能方面取得了最先进的结果。在今后的工作中，首先，我们将结合预训练来获得更强的 -GNN。其次，当训练数据非常有限时（如支持集大小为 16 或更少时），我们需要进一步改进我们的模型。第三，如何有效地识别和纳入结构更复杂的表征是另一个有趣的探索方向。第四，我们将把我们的模型应用到更多的实际场景中，比如大环天然产物的合成和生成。

和历史研究国际会议（ICLR 2023）。

致谢

这项工作部分得到了国家自然科学基金委员会（NSFC）第 61836011 号合同的资助，部分得到了中央高校基础研究基金（WK3490000007）的资助。

参考资料

Ravichandra Addanki、Peter Battaglia、David Budden、Andreea Deac、Jonathan Godwin、Thomas Keck、Wai Lok Sibon Li、Alvaro Sanchez-Gonzalez、Jacklynn Stott、Shantanu Thakoor 和 Petar Velicˇkovic´.大规模图表示学习与深度gnns和自我监督。*arXiv preprint arXiv:2107.09422*, 2021.

Gaurav Bhardwaj、Jacob O'Connor、Stephen Rettie、Yen-Hua Huang、Theresa A. Ramelot、Vikram Khipple Mulligan、Gizem Gokce Alpkilic、Jonathan Palmer、Asim K. Bera、Matthew J. Bick、Maddalena Di Piazza、Xinting Li、Parisa Hosseinzadeh、Timothy W. Craven、Roberto Tejero、Anna Lauko、Ryan Choi、Calina Glynn、Linlin Dong、Robert Griffin、Wesley C. van Voorhis、Jose Rodriguez、L.Craven, Roberto Tejero, Anna Lauko, Ryan Choi, Calina Glynn, Linlin Dong, Robert Griffin, Wesley C. van Voorhis, Jose Rodriguez, Lance Stewart, Gaetano T. Montelione, David Craik, and David Baker.膜穿越大环的精确从头设计。*细胞》*，185（19）：3520- 3532.e26，2022。ISSN 0092-8674. DOI: https://doi.org/10.1016/j.cell.2022.07.019.URL https：//www.sciencedirect.com/science/article/pii/S0092867422009229.

Shaked Brody, Uri Alon, and Eran Yahav.图注意力网络有多专注？ *arXiv preprint arXiv:2105.14491*, 2021.

Shuan Chen 和 Yousung Jung.利用局部反应性和全局注意力的深度逆合成反应预测。doi: 10.1021/jacsau.1c00246.

Connor W. Coley、Luke Rogers、William H. Green 和 Klavs F. Jensen。基于分子相似性的计算机辅助合成。*ACS Central Science*, 3(12):1237-1245, Decem- ber 2017. ISSN 2374-7943, 2374-7951. doi: 10.1021/acscentsci.7b00355. URL https：//pubs.acs.org/doi/10.1021/acscentsci.7b00355.

Weilin Cong, Morteza Ramezani, and Mehrdad Mahdavi.论训练图卷积网络时深度的可证明优势。见 A. Beygelzimer、Y. Dauphin、P. Liang 和 J. Wortman Vaughan（编），《*神经信息处理系统进展》*，2021 年。 URL https：//openreview.net/forum?id=r-oRRT-ElX.

戴汉军、李成涛、康纳-科利、戴波、宋乐。用条件图逻辑网络进行逆合成预测。*神经信息处理系统进展》*，第 8870-8880 页，2019 年。

Andreea Deac, Yu-Hsiang Huang, Petar Velicˇkovic´, Pietro Lio`, and Jian Tang.用图共注意力预

和历史研究国际会议（ICLR 2023）。

测药物不良反应。*arXiv preprint arXiv:1905.00534*, 2019.

Jo¨rg Degen, Christof Wegscheid-Gerlach, Andrea Zaliani, and Matthias Rarey.论编译和使用 "类药物 "化学片段空间的艺术。*ChemMedChem：ChemMedChem: Chemistry Enabling Drug Discovery*, 3(10):1503-1507, 2008.

方晓敏、刘力航、雷洁琼、何东龙、张善卓、周静波、王帆、吴华、王海峰。用于性质预测的几何增强型分子表征学习。*自然机器智能*》，4（2）：127-134，2022。

Chelsea Finn、Pieter Abbeel 和 Sergey Levine.用于深度网络快速适应的模型识别元学习。In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp.PMLR，2017 年 8 月 6-11 日。

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl.量子化学的神经信息传递》，2017a。URL https://arxiv.org/abs/1704.01212.

和历史研究国际会议（ICLR 2023）。

Justin Gilmer、Samuel S Schoenholz、Patrick F Riley、Oriol Vinyals 和 George E Dahl。量子化学的神经信息传递。*国际机器学习会议*，第 1263-1272 页。PMLR，2017b。

Will Hamilton, Zhitao Ying, and Jure Leskovec.大型图上的归纳表示学习In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Gar- nett (eds.), *Advances in Neural Information Processing Systems, volume* 30.Curran Asso- ciates, Inc., 2017a。URL https://proceedings.neurips.cc/paper/2017/file/5dd9db5e033da9c6fb5ba83c7a7ebea9-Paper.pdf.

Will Hamilton, Zhitao Ying, and Jure Leskovec.大型图上的归纳表示学习 *神经信息处理系统进展》*，30，2017b。

胡伟华、刘博文、约瑟夫-戈麦斯、马林卡-齐特尼克、梁珀西、维杰-潘德和尤雷-莱斯科维奇。预训练图神经网络的策略。2020 年 *国际学习表征会议*。URL https://openreview.net/forum?id= HJlWWJSFDH.

Weihua Hu、Matthias Fey、Hongyu Ren、Maho Nakata、Yuxiao Dong 和 Jure Leskovec。Ogb-lsc：*arXiv preprint arXiv:2103.09430*, 2021.

Nirmala Chandralega Kampan、Mutsa Tatenda Madondo、Orla M McNally、Michael Quinn 和 Magdalena Plebanski。紫杉醇及其在卵巢癌治疗中不断发展的作用》。*Biomed Res. Int.*, 2015:413076, June 2015.

Sunghwan Kim、Jie Chen、Tiejun Cheng、Asta Gindulyte、Jia He、Siqian He、Qingliang Li、Ben- jamin A Shoemaker、Paul A Thiessen、Bo Yu、Leonid Zaslavsky、Jian Zhang 和 Evan E Bolton。PubChem 2019 update: Improved access to chemical data.*Nucleic Acids Res.*, 47(D1):D1102- D1109, January 2019.

Thomas N Kipf 和 Max Welling.使用图卷积网络进行半监督分类。*ICLR*, 2017.

Greg Landrum 等人，Rdkit：开源化学信息学软件，2016.*URL http://www. rdkit. org/, https://github. com/rdkit/rdkit*, 149:150, 2016.

李国豪、熊晨鑫、阿里-塔贝特和伯纳德-加内姆。Deepergcn：2020 年训练更深层 GCN 所需的一切。

Junying Li, Deng Cai, and Xiaofei He.为药物发现学习图级表示》，*arXiv preprint arXiv:1709.03741*, 2017.

Zimeng Li, Shichao Zhu, Bin Shao, Tie-Yan Liu, Xiangxiang Zeng, and Tong Wang.用于药物相互作用预测的多视图子结构学习，2022.URL https://arxiv.org/abs/2203.14513.

Meng Liu, Cong Fu, Xuan Zhang, Limei Wang, Yaochen Xie, Hao Yuan, Youzhi Luo, Zhao Xu, Shenglong Xu, and Shuiwang Ji.通过更深的二维和三维图网络进行快速量子特性预测》，2021 年。URL https://ogb.stanford.edu/paper/kddcup2021/pcqm4m_DIVE.pdf。

刘胜超、王汉臣、刘伟阳、Joan Lasenby、郭宏宇和唐健。用三维几何预训练分子图表示。

国际学习表征会议，2022 年。URL https://openreview.net/forum?id=xQUe1pOKPam.

伊利亚-洛希洛夫和弗兰克-胡特解耦权重衰减正则化。*国际学习表征会议*，2019。URL https://openreview.net/forum?id= Bkg6RiCqY7.

安德烈亚斯-卢卡斯图神经网络无法学习的东西：深度与宽度。*国际学习表征会议*，2020 年。URL https://openreview.net/forum?id= B1l2bp4YwS.

Arnold K Nyamabo、Hui Yu 和 Jian-Yu Shi。Ssi-Ddi：用于药物相互作用预测的亚结构-亚结构相互作用。*生物信息学简报*，2021 年。

Arnold K Nyamabo, Hui Yu, Zun Liu, and Jian-Yu Shi.利用可学习的尺寸自适应分子子结构预测药物间相互作用。*生物信息学简报*，2022年。

Trang Pham, Truyen Tran, Hoa Dam, and Svetha Venkatesh.通过虚拟节点深度学习进行图分类》，*arXiv preprint arXiv:1708.04357*, 2017.

于荣、边亚涛、徐汀阳、谢伟阳、魏颖、黄文兵、黄俊洲。大规模分子数据的自监督图变换器。*神经信息处理系统进展》*，33:12559-12571，2020.

单思思、罗世彤、杨子清、洪俊贤、苏玉峰、丁帆、付丽丽、李晨宇、陈鹏、马建珠、史璇玲、张琦、Bonnie Berger、张琳琪和彭健。以深度学习为指导优化具有广泛中和作用的人类抗sars-cov-2变体抗体。doi: 10.1073/ pnas.2122954119.

Chence Shi, Minkai Xu, Hongyu Guo, Ming Zhang, and Jian Tang.用于逆合成预测的图到图框架。*第 37 届机器学习国际会议论文集》*，《*机器学习研究论文集》第* 119 卷，第 8818-8827 页。PMLR，2020 年 7 月 13-18 日。

Yu Shi, Shuxin Zheng, Guolin Ke, Yifei Shen, Jiacheng You, Jiyan He, Shengjie Luo, Chang Liu, Di He, and Tie-Yan Liu.在大规模分子建模数据集上对 graphormer 进行基准测试，2022 年。URL https://arxiv.org/abs/2203.04810.

Vignesh Ram Somnath、Charlotte Bunne、Connor Coley、Andreas Krause 和 Regina Barzilay。为逆合成预测学习图模型。见 M. Ranzato、A. Beygelzimer、Y. Dauphin、P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems, volume* 34, pp.Curran Associates, Inc.，2021 年。

Megan Stanley、John F Bronskill、Krzysztof Maziarz、Hubert Misztela、Jessica Lanini、Marwin Segler、Nadine Schneider 和 Marc Brockschmidt。 FS-mol：分子的少量学习数据集。*第三十五届神经信息处理系统会议数据集和基准轨道（第二轮）*，2021 年。URL https://openreview.net/forum?id= 701FtuyLlAd。

Jonathan M. Stokes、Kevin Yang、Kyle Swanson、Wengong Jin、Andres Cubillos-Ruiz、Nina M. Donghia、 Craig R. MacNair、 Shawn French、 Lindsey A. Carfrae、 Zohar Bloom-Ackermann、 Vic- toria M. Tran、 Anush Chiappino-Pepe、 Ahmed H. Badran、 Ian W. Andrews、 Emma J. Chory、 George M. Church、 Eric D. Brown、 Tommi S. Jaakkola、Regina Barzilay 和 James J. Collins。Jaakkola, Regina Barzilay, and James J. Collins.抗生素发现的深度学习方法。*细胞》*，180（4）：688-702.e13，2020。ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2020.01.021.URL https://www.sciencedirect.com/science/article/pii/S0092867420301021.

和历史研究国际会议（ICLR 2023）。

Hongmao Sun, Gregory Tawa, and Anders Wallqvist.脚手架跳转方法的分类。*Drug Discov.Today*, 17(7-8):310-324, April 2012.

孙若曦、戴汉军和于蘅玮。GNN 预训练有助于分子表征吗？见 Alice H. Oh、Alekh Agarwal、Danielle Belgrave 和 Kyunghyun Cho（编），《*神经信息处理系统进展》*，2022 年。URL https://openreview.net/forum? id=uytgM9N0vlR.

Wen Torng 和 Russ B. Altman.用于预测药物靶点相互作用的图卷积神经网络。DOI: 10.1021/ACS.JCIM.9B00628.

Petar Velicˇkovic´, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio.图注意网络》，*arXiv preprint arXiv:1710.10903*, 2017.

和历史研究国际会议（ICLR 2023）。

Anilrudh A. Venugopal 和 Stuart Johnson.菲达霉素：获准用于治疗艰难梭菌感染的新型大环抗生素。*临床感染性疾病*》，54（4）：568-574，2011 年 12 月。ISSN 1058-4838。DOI：10.1093/CID/CIR830。URL https://doi.org/10.1093/cid/ cir830.

David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, et al. Drugbank 5.0：2018 年药物数据库重大更新。*核酸研究*》，46（D1）：D1074-D1082，2018.

Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande.Moleculenet：分子机器学习的基准。DOI: 10.1039/C7SC02664A. URL http：//dx.doi.org/10.1039/C7SC02664A.

Yingce Xia, Jinhua Zhu, Lijun Wu, Yang Fan, Shufang Xie, Yutai Hou, and Tao Qin.当跨前遇上图神经网络 2021.URL https://ogb.stanford.edu/paper/kddcup2021/pcqm4m_GNNLearner.pdf.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka.图神经网络有多强大？ *arXiv preprint arXiv:1810.00826*, 2018.

Nuo Xu, Pinghui Wang, Long Chen, Jing Tao, and Junzhou Zhao.Mr-gnn：用于预测结构化实体相互作用的多分辨率和双图神经网络。*arXiv preprint arXiv:1905.09558*, 2019.

应承萱、蔡天乐、罗胜杰、郑淑欣、柯国林、何迪、沈彦明、刘铁岩。变换器在图表示方面真的表现糟糕吗？*神经信息处理系统进展*》，第 34 卷，第 28877-28888 页，2021 年。

Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee.用于分子性质预测的基于动机的图自我监督学习。*神经信息处理系统进展*》，34:15870-15882，2021 年。

Xinyu Zhu, Yongliang Shen, and Weiming Lu.用于药物相互作用预测的分子亚结构感知网络。*CIKM*，2022年。

A 详细的实验配置

表 5、表 6 和表 7 分别汇总了分子性质预测、药物相互作用预测和逆合成的超参数。

	PCQM4Mv1	FS-Mol	MoleculeNet
层数	12		{64, 6, 8, 12}
隐藏维度	256	256	{128, 256}
优化器	AdamW	AdamW	AdamW 辍学
学习率	0. 0.0003	{00.0, 0.1, 0.} {0.0001, 0.0002, 0.0003}	{.0, 0.1, 0.2, 0.3, 0.5} {0.00005,0.0001,0.0002,0.0005}

和历史研究国际会议（ICLR 2023）。

训练步骤	300 次迭代	10000 次迭代 50	{,100 次迭代
批次大小	512	16	{32,64}
重量衰减	0.1	{0.01, 0.1}	0.01
学习率衰减	余弦	余弦	线性

表 5：分子性质预测任务的超参数详情。

和历史研究国际会议（ICLR 2023）。

层数	6
隐藏维度	512
优化器	亚当
{辍学率}	0.2，0.5
学习率	0.0003
训练步骤	100 个历元
批量大小	128
重量衰减	0.01
学习率衰减	余弦

表 6：药物相互作用预测的超参数详情。

层数	6
隐藏维度	512
优化器	AdamW
Dropout	0.1
学习率	0.0003
训练	步骤 200 个历元
批量大小	64
重量衰减	0.1
学习率衰减	余弦

表 7：用于逆合成的详细超参数。

B 两个命题的证明

*命题 1 的证明*我们首先明确写下 O

GNN 的无环变体，以此开始证明。

具体来说，债券表示法如下

$$h_{ij}^{(l)} = h_{ij}^{(l-1)} + \texttt{MLP}(h_i^{(l-1)}, h_j^{(l-1)}, h_{ij}^{(l-1)}, U^{(l-1)}).$$

原子表示法如下

$$\begin{aligned} h_i^{(l)} &= \sum_{j \in N(i)} \alpha_{ij} W_v \ \texttt{concat}(h_j^{(l-1)}, h_{ij}^{(l-1)}); \\ \alpha_{ij} &\propto \exp(a^\top \ \texttt{LeakyReLU}(W_q h_i^{(l-1)} + W_k \ \texttt{concat}(h_j^{(l-1)}, h_{ij}^{(l)}))); \\ h_i^{(L)} &= h_i^{(L-1)} + \texttt{MLP}\left(h_i^{(L-1)}, \bar{h}_i^{(L)}, U^{(L-1)}\right). \end{aligned} \tag{6}$$

复合表示法如下

$$U^{(l)} = U^{(l-1)} + \texttt{MLP}\left(\frac{1}{|V|}\sum_{i=1}^{|V|} h_i^{(l)}, \frac{1}{|E|}\sum_{i,j} h_{ij}^{(l)}, U^{(l-1)}\right). \tag{7}$$

根据上述符号，命题 1 可以转化为下面的主张：

声称。对于任意有值图（G，f）和 G 中的任意两个节点 v_a , v_b ，如果有值 k 邻域

和历史研究国际会议（ICLR 2023）。

u 和 v 的子图（即（$G(v_a, k)$，f）和（$G(v_b, k)$，f））是等价的，因此 $h^l = h^l_a{}_b$ 。任何 $l \in \{1, - -, k\}$ 都成立，与参数无关。

我们将 $(G(v_a, k), f)$ 和 $(G(v_b, k), f)$ 之间的等价映射记为 P。如果 $P(v_i) = v_j$，我们将略微滥用符号，让 $P(i) = j$。

具体来说，我们将证明，对于任意 $l \in \{0, 1, - - -, k\}$，我们有 $h^l_{c1} = h^l_{P(c1)}$ 对于任意 $v_{c1} \in V(v_a, k - l)$，且 $h^l_{c1c2} = h^l_{P(c1)P(c2)}$ 对于任意 $v_{c_1}, v_{c_2} \in V(v_a, k - l)$。

和历史研究国际会议（ICLR 2023）。

基本情况：对于 $l = 0$，根据 f 的定义，我们有 $f(v_{c_1}) = f(P(v_{c_1}))$ 对于每一个 $v_{c_1} \in V(v_a, k)$ 且 $f(v_{c_1}, v_{c_2}) = f(P(v_{c_1}), P(v_{c_2}))$ 对于每个 $v_{c_1}, v_{c_2} \in V(v_a, k)$。紧随其后的是作为 $h^0_{c_1} = f(v_{c_1}), h^0_{P(c_1)} = f(P(v_{c_1})), h^0_{c_1c_2} = f(v_{c_1}, v_{c_2})$，而 $h^0_{P(c_1)P(c_2)} = f(P(c_1), P(c_2))$.

归纳步骤：假设在 $l = i \in \{0, \ldots, k-1\}$ 条件下，主张为真。那么，对于 $l = i + 1$，我们有这样的结论：对于每一个 $v_{c_1}, v_{c_2} \in V(v_a, k-l)$、

$$\begin{aligned} h^{(l)}_{c_1c_2} &= h^{(l-1)}_{c_1c_2} + \mathrm{MLP}(h^{(l-1)}_{c_1}, h^{(l-1)}_{c_2}, h^{(l-1)}_{c_1c_2}, U^{(l-1)}) \\ &\overset{(\star)}{=} h^{(l-1)}_{P(c_1)P(c_2)} + \mathrm{MLP}(h^{(l-1)}_{P(c_1)}, h^{(l-1)}_{P(c_2)}, h^{(l-1)}_{P(c_1)P(c_2)}, U^{(l-1)}) \\ &= h^{(l)}_{P(c_1)P(c_2)}, \end{aligned} \tag{8}$$

其中式$(\star)$是由于归纳假设，因为 $v_{c_1}, v_{c_2} \in V(v_a, k-l) \subset V(v_a, k-(l-1))$.同样，对于每个 $v_{c_1} \in V(v_a, k-l)$、

$$\begin{aligned} \bar{h}^{(l)}_{c_1} &= \sum_{j \in N(c_1)} \alpha W_{jv}\, \mathrm{concat}(h^{(l)}_{c_1j}, h_j)^{(l-1)} \\ &\overset{(\circ)}{=} \sum_{j \in N(c_1)} \alpha W_{jv}\, \mathrm{concat}(h^{(l)}_{P(c_1)P(j)}, h^{(l-1)}_{P(j)}) \\ &\overset{(\diamond)}{=} \sum_{j \in N(c_1)} \alpha W_{P(j)v}\, \mathrm{concat}(h^{(l)}_{P(c_1)P(j)}, h^{(l-1)}_{P(j)}) \\ &= \sum_{j \in N(c_1)} \alpha W_{jv}\, \mathrm{concat}(h)^{(l)}_{P(c_1)j}, h^{(l-1)}_{j}) \\ &\bar{h}^{(l)}_{P(c_1)}, \end{aligned}$$

式中（$\circ$）是由于归纳假设和式 (8)。式 $(\diamond)$ 是由于

$$\begin{aligned} \alpha_j &\propto \exp(\boldsymbol{a}^\top\, \mathrm{LeakyReLU}(W h_q^{(l-1)} + W_k\, \mathrm{concat}(h_j^{(l-1)}, h_{c_1}^{(l)}))) \\ &= \exp(\boldsymbol{a}^\top\, \mathrm{LeakyReLU}(W h_{P(c_1)}^{(l-1)} + W_k\, \mathrm{concat}(h^{(l-1)}_{P(j)}, h^{(l)}_{P(c_1)P(j)}))), \end{aligned}$$

因此，对于任意 $j \in N(c_1)$ 而言，

$\alpha_j = \alpha_{P(j)}$。于是我们有

$$\begin{aligned} h^{(L)}_{c_1} &= h^{(L-1)}_{c_1} + \mathrm{MLP}(h^{(L-1)}_{c_1}, \bar{h}^{(L)}_{c_1}, U^{(L-1)}) \\ &= h^{(l-1)}_{P(c_1)} + \mathrm{MLP}(h^{(l-1)}_{P(c_1)}, \bar{h}^{(l)}_{P(c_1)}, U^{(l-1)}) \\ &= h^{(l)}_{P(c_1)}. \end{aligned}$$

因此，对于 $l = i + 1$，权利要求成立，归纳权利要求的证明完成。因此，对于每一个 $l \in \{0, \ldots, k\}$，本命题都是正确的。

对于每个 $l \in \{0, \ldots, k\}$ $u \in G(v_a, 0) \subset G(v_a, k-l)$。因此，我们有 $h^{(l)}_a = h^{(l)}_{P(a)} = h^{(l)}_b$、证明就完成了。□

和历史研究国际会议（ICLR 2023）。

*命题 2 的证明*对于两个等价子图（$G(v_a, k), f$）和（$G(v_b, k), f$），如果 v_a 和 v_b 位于不同的环上，我们有

$$\frac{1}{|R(v_a)|}\sum_{r\in R(v_a)} h_r^{(0)} = \frac{1}{|R(v_b)|}\sum_{r\in R(v_b)} h_r^{(0)}.$$

和历史研究国际会议（ICLR 2023）。

因此，存在一种 MLP 选择，使得

$$\begin{aligned} h_a^{(1)(0)} &= h_a + \mathrm{MLP}\left(h_a^{(0)}, \bar{h}_a^{(1)}, \frac{1}{|R(v_a)|}\sum_{r\in R(v_a)} h_r^{(0)}, U^{(0)}\right) \\ &= h_b^{(0)} + \mathrm{MLP}\left(h_b^{(0)}, h_b^{-(1)}, \frac{1}{|R(v_b)|}\sum_{r\in R(v_b)} h_r^{(0)}, U^{(0)}\right) \\ &= h_b^{(1)}. \end{aligned}$$

证明完成。 □

C 更多消融研究

C.1 节点表征集合与复合表征

我们探讨了使用平均池化 $h_G = \frac{1}{|V|}\sum_{i=1}^{|V|} h_i^{(L)}$ 和复合池化之间的区别。表示 $U^{(L)}$ 进行分类。我们尝试了两个具有不同层数（$L = 6$ 和 12）的网络。我们在 PCQM4Mv1 数据集上进行了实验。验证平均绝对误差（MAE）见表 8。我们可以看到，使用平均节点池比使用

复合表示。这与在 GIN 中使用虚拟节点的发现是一致的（Hu 等人，2021 年）。虚拟节点可被视为一种复合表示，它与图中的所有节点相连。在使用虚拟节点时，通常的做法是使用节点表示的平均值或总和来表示图。具体实现可参考 https://github.com/ snap-stanford/ogb/blob/1c875697fdb20ab452b2c11cf8bfa2c0e88b5ad3/ examples/lsc/pcqm4m/gnn.py#L60。

	$L = 6$	$L = 12$
平均节点集合	0.1171	0.1149
复合表示法	0.1196	0.1167

表 8：使用平均节点表示法与复合表示法的比较

C.2 认真汇总各环的信息

在公式(4)中，我们将原子表征的总和集合、键表征的总和集合和化合物表征集合起来更新环表征。另一种解决方案是使用注意力模型来汇总原子和键的表示。我们对更新环状表征的变体进行了如下研究：

$$h_r^{(l)} = h_r^{(l-1)} + \mathrm{MLP}\left(h_r^{(l-1)}, \sum_{v_i\in V(r)} \alpha_i h_i^{(l)}, U^{(l)}, \sum_{e_{ij}\in E(r)} \beta_{ij} h_{ij}^{(l)}\right) \tag{9}$$

在公式(9)中

$$\alpha_i^{(l)} \propto \exp\left(W_{q1} h_r^{(l-1)} + W_{k1} h_i^{(l)}\right) \text{ 和 } \beta_{ij}^{(l)} \propto \exp\left(W_{q2} h_r^{(l-1)} + W_{k2} h_{ij}^{(l)}\right), \tag{10}$$

其中四个 W 是要学习的参数。结果见表 9。我们可以看到，虽然我们的方法很简单，但它

和历史研究国际会议（ICLR 2023）。

能有效地利用环信息，并优于这种基于注意力的变体。

C.3 O-GNN与BRICS

我们的方法中使用的环形表示法可被视为一种特殊的图案。也许有人会问，其他类型的新图案是否会有所帮助。为了了解效果，我们使用 BRICS 模型（Degen 等人，2008 年）将分子分解成碎片。BRICS 设计了 16 条断键规则，可与一组化学反应相匹配。式（2,3,4,5）中的环表示被替换为

和历史研究国际会议（ICLR 2023）。

	L = 6	L = 12
O-GNN	0.1171	0.1149
O-GNN with attention models when updating ring representations	0.1179	0.1160

表 9：我们的方法与使用注意力模型更新环形代表的比较

	L = 2	L = 4	L = 6	L = 8	L = 12
O-GNN	0.1247	0.1201	0.1181	0.1172	0.1155
-不带圆环的 GNN	0.1325	0.1243	0.1222	0.1221	0.1204
金砖国家	0.1294	0.1239	0.1219	0.1208	0.1193

表 10：使用简单环（即我们的方法）与使用基于金砖五国的碎片的比较。

这些图案表示法。其余部分保持不变。我们在 PCQM4Mv1 数据集上进行了实验，结果如表 10 所示。由于时间和计算资源的限制，所有模型都进行了 200 次训练。

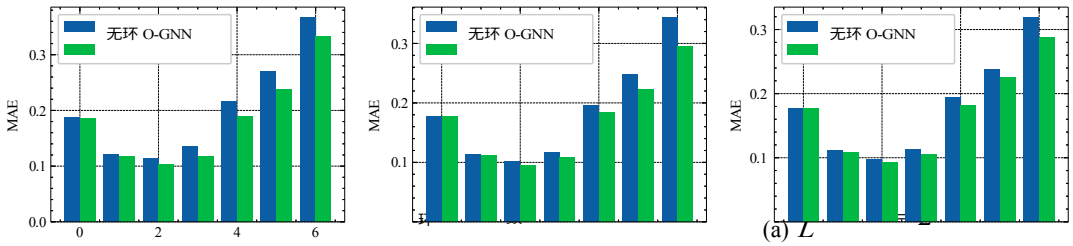
从表 10 中我们可以得出以下结论(1) 使用简单的环形表示法比使用 BRICS 获得更好的结果；(2) 总体而言，使用 BRICS 比不使用任何环形或基于环形的碎片信息的变体更好。我们将继续探索更多的分割方法。

C.4 O-GNN 与无环 O-GNN 的更多比较

作为对图 5 中分子环数 MAE 分析的补充，我们还在图 8 中报告了 -GNN 和变体"-GNN 无环 "的预测误差（即平均绝对误差，MAE）。我们可以看到，当分子没有环时，这两种方法的表现类似。当环的数量从 1 个增加到 6 个时，平均绝对误差增大，而 O-GNN 始终优于 "O-GNN（无环）"变体。

C.5 补充讨论

关于过度平滑 有人可能会好奇，既然我们构建了一个 12 层的网络，它是否会遭受过度平滑呢？实际上，Cong 等人（2021 年）指出："*过度平滑在实践中并不一定会发生，更深层次的模型具有可证明的表现力，能以线性收敛率收敛到全局最优，只要训练得当，就能达到非常高的训练精度*。(语出（Cong 等人，2021 年），表达准确）。此外，Li 等人（2020 年）和 Addanki 等人（2021 年）都成功地训练了 50 多层网络。我们的方法如下



(a) L

和历史研究国际会议（ICLR 2023）。

(b) $L = 6$ (c) $L = 12$

图 8：按不同属性分类的预测 MAE。X 轴表示环的数量，Y 轴表示验证集上的平均绝对误差 (MAE)。

和历史研究国际会议（ICLR 2023）。

Addanki 等人，2021 年）的结构，因此我们认为我们的模型不存在过度平滑的问题。

k 邻接*建模*：如果我们要明确使用 k 邻接信息，可能需要额外的模块来处理它们，例如

$$\texttt{net}_1\,(1\text{ 个邻居节点}) + \texttt{net}_2\,(2\text{ 个邻居节点}) + \text{- - - } + \texttt{net}_k\,(\text{k 个邻居节点})。 \tag{11}$$

为确保表达能力，我们通常不共享参数。因此，参数是普通 GNN 的 k 倍。-GNN 会不断增加参数的百分比（与 k 无关）。当 k 较大时，-GNN 的参数效率会更高。另一方面，最佳 k* 并不容易确定。例如，在 DrugBank 中，环的最大尺寸从 3（如 DB00658）到 53（如 DB05034）不等。很难确定哪一个 k 是最好的。

*关于不变约束*在 中，原子、键和环的特征是不变的。具体来说，原子和化学键的特征与它们的类型、相关电子数、邻域数等有关（详情请参考 `https://github.com/O-GNN/O-GNN/blob/5b70a4f9dc9a5f87a0171eea1e9cecde30489eb8/ogb/utils/features.py#L2`）。环表示法通过原子和键表示法获得（请参考公式(4)），这些表示法也是不变的。变体特征（如坐标）不进行编码。

收敛速度比较：PCQM4Mv1 的验证 MAE 曲线如图 9 所示。报告了 6 层 -GNN（带/不带环）和 12 层 -GNN（带/不带环）的结果。我们可以看到

(1) 通过对 6 层 O-GNN 进行 175 次训练，结果与对 12 层 "无环 O-GNN" 进行 275 次训练的结果几乎相同；

(2) 通过对 12 层 O-GNN 进行 75 次训练，结果与对 12 层 "无环 O-GNN" 进行 275 次训练的结果几乎相同。

这些结果表明 O-GNN 具有更好的收敛速度。

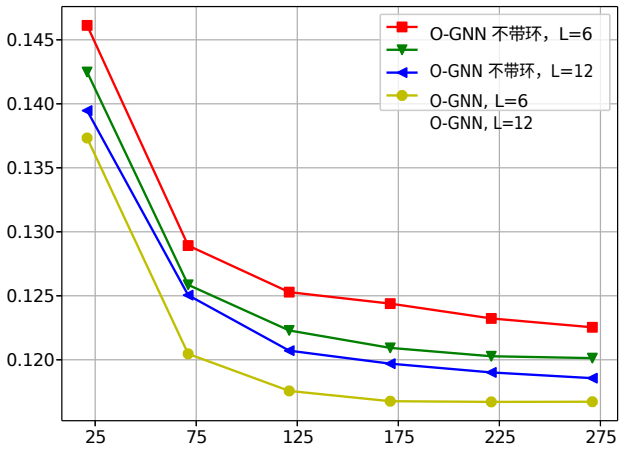


图 9："-GNN "和"-GNN（无环）"的收敛速度比较。 O

*不同参数数之间的比较*图 4 显示了"-GNN "和"-GNN（无环）"随层数变化的验证 MAE。我们还将验证 MAE
与参数数量的关系见图 10。我们可以观察到，当与参数数保持一致时，O-GNN 仍然优于无建

和历史研究国际会议（ICLR 2023）。

模环的变体。

*MoleculeNet 上的预训练基线。*表 11 总结了 MoleculeNet 上的预训练基线。Sun 等人（2022 年）的研究表明，不同的数据拆分方法会导致明显不同的结果。我们按照惯例使用基于支架的拆分方法，并引用了 Fang 等人（2022 年）的 Rong 等人（2020 年）的结果。[O] 请注意，-GNN 的结果并没有在未标记的分子上进行预训练。我们可以看到，在平均得分方面，我们的方法

和历史研究国际会议（ICLR 2023）。

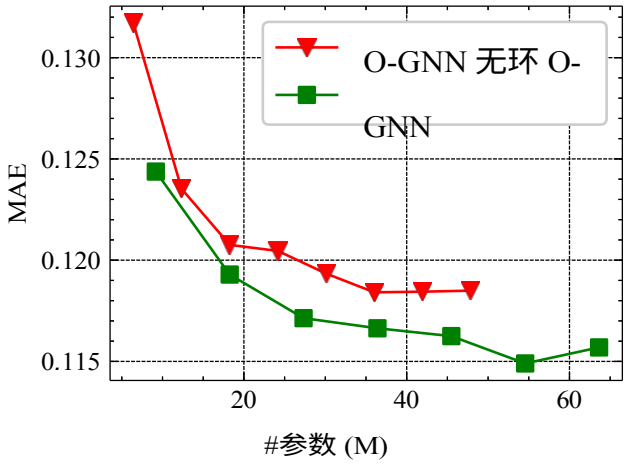


图 10：PCQM4Mv1 的验证 MAE 随参数数量的变化。

与那些强基线相当，这证明了我们方法的有效性。今后，我们将把我们的方法与预训练相结合。

数据集 # 分子	BBBP 2039	Tox21 7831	ClinTox 1478	HIV 41127	BACE 1513	SIDER 1478	平均值
(Hu 等人，2020 年)	71.2 ± 0.9	74.2 ± 0.8	73.7 ± 4.0	75.8 ± 1.1	78.6 ± 1.4	60.4 ± 0.6	72.3
G-Contextual (Liu et al., 2022)	70.3 ± 1.6	75.2 ± 0.3	59.9 ± 8.2	75.9 ± 0.9	79.2 ± 0.3	58.4 ± 0.6	69.8
G-Motif （Liu 等人，2022 年）	66.4 ± 3.4	73.2 ± 0.8	77.8 ± 2.0	73.8 ± 1.4	73.4 ± 4.0	60.6 ± 1.1	70.9
GraphMVP （Liu 等人，2022 年）	72.4 ± 1.6	75.9 ± 0.5	79.1 ± 2.8	77.0 ± 1.2	81.2 ± 0.9	63.9 ± 1.2	74.9
MGSSL（Zhang 等人，2021 年）	70.5 ± 1.1	76.5 ± 0.3	80.7 ± 2.1	79.5 ± 1.1	79.7 ± 0.8	61.8 ± 0.8	74.8
GROVERbase （Rong 等人，2020 年）	70.0 ± 0.1	74.3 ± 0.1	81.2 ± 3.0	62.5 ± 0.9	82.6 ± 0.7	64.8 ± 0.6	72.6
GROVERlarge (Rong et al., 2020)	69.5 ± 0.1	73.5 ± 0.1	76.2 ± 3.7	68.2 ± 1.1	81.0 ± 1.4	65.4 ± 0.1	72.3
创业板指数（Fang 等人，2022 年）	72.4 ± 0.4	78.1 ± 0.1	90.1 ± 1.3	80.6 ± 0.9	85.6 ± 1.1	67.2 ± 0.4	79.0
O-GNN （我们的）	76.4 ± 0.4	75.7 ± 0.7	94.3 ± 1.6	81.3 ± 1.2	85.8 ± 1.0	66.2 ± 1.2	80.0

表 11：MoleculeNet 上的预训练基线。

D 相关工作总结

GCN（Kipf & Welling，2017 年）根据邻接矩阵和度矩阵聚合邻居信息，然后通过线性变换和非线性激活层更新聚合信息。GraphSAGE（Hamilton等人，2017b）通过元素平均法聚合邻居信息。GAT（Velic˘kovic´等人，2017 年）在 GNN 中引入了注意力机制，通过它可以自适应地聚合邻居的表征。Brody 等人（2021 年）提出了 GATv2，以更具表现力的方式改进注意力机制。徐
等人（2018）开发了一种简单的聚合函数，它涉及一个ϵ参数和多层外显子（MLP），可以证明它与 Weisfeiler-Lehman 图同构检验一样强大。此外，Gilmer 等人（2017b）；Li 等人（2017）；Pham 等人（2017）提议增强图
虚拟节点捕捉图的全局信息。虚拟节点连接图中的所有其他节点，并在训练过程中共同更

和历史研究国际会议（ICLR 2023）。

新。它的有效性在一系列图分类任务中得到了验证。

然而，这些工作并没有在图神经网络中明确使用环 R。作为对这些工作的补充，我们考虑了如何将环信息（节点和边信息之外的另一个重要组成部分）纳入分子建模。这些先进的汇总和更新功能也适用于我们的工作。