3DMol-Net: Learn 3D Molecular Representation using Adaptive Graph Convolutional Network Based on Rotation Invariance

Chunyan Li, Wei Wei, Jin Li, Junfeng Yao, Xiangxiang Zeng, Zhihan Lv

***Abstract*—Studying the deep learning-based molecular representation has great significance on predicting molecular property, promoted the development of drug screening and new drug discovery, and improving human well-being for avoiding illnesses. It is essential to learn the characterization of drug for various downstream tasks, such as molecular property prediction. In particular, the 3D structure features of molecules play an important role in biochemical function and activity prediction. The 3D characteristics of molecules largely determine the properties of the drug and the binding characteristics of the target. However, most current methods merely rely on 1D or 2D properties while ignoring the 3D topological structure, thereby degrading the performance of molecular inferring. In this paper, we propose 3DMol-Net to enhance the molecular representation, considering both the topology and rotation invariance (RI) of the 3D molecular structure. Specifically, we construct a molecular graph with soft relations related to the spatial arrangement of the 3D coordinates to learn 3D topology of arbitrary graph structure and employ an adaptive graph convolutional network to predict molecular properties and biochemical activities. Comparing with current graph-based methods, 3DMol-Net demonstrates superior performance in terms of both regression and classification tasks. Further verification of RI and visualization also show better robustness and representation capacity of our model.**

***Index Terms*—Molecular Representation, Graph Convolutional Network, Rotation Invariance, Adaptive, Property Prediction**

I. Introduction

This work was supported in part by the scientific research fund of Yunnan Provincial Department of Education (No.2020J0362), the National Natural Science Foundation of China (No. 62072388 and No. 61902203), the Industry Guidance Project Foundation of Science Technology Bureau of Fujian Province in 2020 (No.2020H0047), the Natural Science Foundation of the Science Technology Bureau of Fujian Province in 2019 (No. 2019J01601), the Science Technology Bureau Project of Fujian Province in 2019 (No. 2019C0021) and the Key Research and Development Plan - Major Scientific and Technological Innovation Projects of ShanDong Province (No. 2019JZZY020101).

Chunyan Li (lchy0316@gmail.com) is with the School of Informatics, Xiamen University, Xiamen 361005, China and also is with Yunnan Minzu University, Kunming 650500, China.

Wei Wei (weiwei@xaut.edu.cn) is with the School of Computer Science and Engineering, Xian University of Technology, Xian 710048, China.

Jin Li (lijin@ynu.edu.cn) is with the School of Software, Yunnan University, Kunming 650091, China.

Junfeng Yao (yao0010@xmu.edu.cn) is with the School of Informatics, Xiamen University, Xiamen 361005, China.

Xiangxiang Zeng (xzeng@hnu.edu.cn) is with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China.

Zhihan Lv (lvzhihan@gmail.com) is with the School of Data Science and Software Engineering, Qingdao University, China.

Manuscript received April 20, 2021; revised August 1, 2021. Corresponding author: Junfeng Yao (email: yao0010@xmu.edu.cn) and Xiangxiang Zeng (email: xzeng@hnu.edu.cn).

ACCURATE representation of molecules is the basis of various downstream applications, such as molecular property prediction and drug design. Finding chemicals with good pharmacological, toxicological and pharmacokinetic properties remains a huge challenge in the field of drug discovery, that will take an important step for human health. Therefore, molecular property prediction is a crucial task. Traditionally, molecular property prediction is based on the experience of a chemist or pharmacist and relies on time-consuming simulations and experiments to obtain the molecular characteristics. In recent decades, neural-network-inspired models have shown strong potentials for computational biology and chemistry, including molecular modeling [1], protein-ligand activity classification [2], [3], molecular property prediction [4]–[7], biological interaction networks evaluation [8]–[11], and so on. Such models showed superior predictive performance compared with traditional computation methods that are time-consuming and costly. By means of molecular modeling and chemical property prediction, deep learning methods have greatly accelerated the speed of new drug research and development, and saved significant resources [12], [13].

In general, the predictive accuracy of neural network models for various properties prediction tasks depends on the molecular representation. MoleculeNet [5] shows that learnable representation is powerful tool for molecular machine learning. Most of the previous work focused on 1D sequence-based manner [14], 2D molecular fingerprint [15] and 2D graph [16], [17], or a mixture of 1D and 2D methods [4]. However, existing deep neural networks for ligand-based models take 1D or 2D features as input, that limit the extraction of features to compositions of pre-specified molecular structures defined during the process of feature descriptor. Hence, the ability for extracting arbitrary features is eliminated [18]. Almost all interactions among molecules are carried out in 3D space, such as protein-ligand interaction, affinity binding and poses finding, a more accurate prediction is required to learn 3D structure-based features [19]. From a microscopic view, these interactions are associated with their 3D coordinate. At present, there are already some applications [20], [21] attempting to explore molecular conformation space [22], which is shown that 3D structure-based models successfully predict new active molecules or binding sites, which is unpredictable by 2D-based models. However, the molecular 3D representation methods have encountered fatal limitations, such as large amount of parameters and memory [1], [2],

[23], low predictive performances [22] and lack of robustness caused by unconsidered 3D invariant features [24].

To address these limitations for effective molecular representation and considering the remarkable achievements mentioned above, in this paper, we proposed a novel architecture 3DMol-Net, considering the 3D structure and its rotation invariance (RI). Such invariance is an important problem in 3D geometric modeling and point cloud processing [25]. The 3D coordinates are determined by coordinate systems and different coordinate systems determine different 3D coordinates. To eliminate the influence of different coordinate systems, the proposed 3DMol-Net is based on RI. Specifically, 3DMol-Net is an adaptive graph convolutional model for 3D molecular representation that can extract soft relationships among nodes and construct 3D graph Laplacian in 3D space. Then, we proposed the rotation invariance mapping (RIM) to learn 3D characteristics of molecules based on 3D graph Laplacian. Extensive experiments demonstrated that the proposed model achieved superior performance on four commonly molecule datasets compared with current state-of-the-art models. In summary, the key contributions to this paper are as follows:

- We propose a soft relation-based 3D graph Laplacian method, an efficient method to obtain 3D spatial relationships between atoms in molecules.
- We propose a rotation invariance mapping method, based on that an adaptive graph convolutional model 3DMol-Net is constructed to extract RI features. 3DMol-Net allows arbitrary graph size as input, and is a powerful model for processing non-Euclidean data with graph structure.
- Extensive experiments are conducted on different datasets to demonstrate the effectiveness of our method. Our model uses fewer training parameters and obtains better performance.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents our approaches. Section 4 analyzes and discusses the experimental results. Section 5 concludes this paper.

II. Related work

A. Molecular representation methods

Simplified Molecular Input Line Entry System (SMILES) strings are in form of a single line text (1D) made up of molecular notations for representing molecules. Extended Connectivity Fingerprint (ECFP) [15] is a widely used molecular fingerprint to construct quantitative structure-activity relationship (QSAR) model of predicting activity of compounds for 2D molecular representation. Graph-based neural networks, such as graph convolutional network (GCN), can directly process graph topology and vertex attribute information, extract features from molecular structure and improve performance of prediction [26]–[28]. The above works focused on 1D and 2D molecular representation methods. In addition, the extended 3D fingerprint representation of molecular conformers (E3FP) [22] was proposed, which applied the logic of ECFP and was a better predictor of similarity in bioactivity. MaSIF [23], based on geometric deep learning, was proposed to capture fingerprints for specific biomolecular interactions. Voxel-based models [2], [6] have been introduced into molecular representation for bioactivity prediction and protein-ligand interaction. But, the computational cost for voxel-based models is very high because of the large and intensive memory requirement, accompanied by the exponential growth of the number of parameters. 3DGCN [24], [29] is based on the bond direction between atoms, scalar features, vector features and mutual transformation are fully considered in the process of 3D convolution operation. However, the relative direction changes with different coordinate systems. When the coordinate system is changed, the relative direction between atoms is also changed that is not an invariant. Additionally, some conventional machine learning-based models [30]–[33] are implemented to learn molecular representation and predict properties. However, most of them use the SMILES sequences of molecules for extracting feature without considering the 3D structural feature of molecules.

B. Graph convolutional neural network (GCNN)

Graph convolution operator in spectral domain [34] has a high computational complexity due to the eigendecomposition of graph Laplacian matrix $\mathcal{L}$, therefore, K-order polynomials are proposed to approximate the original matrix decomposition [34]–[36]. Classic GCN [37] focuses on the first-order approximation. Each convolution layer only considers the direct neighborhood. A large receptive field can be obtained by stacking several layers.

C. Rotation-invariant 3D features

The traditional convolution neural network (CNN) can learn RI through data augmentation (rotation, scaling, etc.) and max-pooling mechanism. 3D RI is still not resolved in traditional CNN. Recent works started to explore the 3D RI using atom-centered symmetry functions [38] and point cloud applications [25]. Previous works, such as PointNet and PointNet++, applied symmetric functions and transformation networks to obtain a global representation. However, they inadequately considered the local structure features of point clouds. On the contrary, 3DTI-net [25] fully considers the local structure of point clouds, but does not capture the long-range dependence between points. Isometric transformation-invariant neural network based on graph [39] has also been proposed and applied to spherical image classification in the field of computer vision [40].

III. Methods

The key idea behind 3DMol-Net is that we focus on 3D structure and RI to obtain the spatial representation of molecules. Figure 1 illustrates the design of our framework, which has two main components: (i) 3D graph Laplacian with soft relations and (ii) rotation-invariant mapping. Given a preprocessed training dataset $\{\mathcal{S}_i\}_{i=1}^{M}$ of M molecules with different number of atoms, which have been transformed from the original SMILES to conformers with 3D structure (Figure 1 (a)), we first calculate the soft relation $\mathcal{F}_{sr}^{i}$ of sample $\mathcal{S}_i$, then How???

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3089162, IEEE Journal of Biomedical and Health Informatics

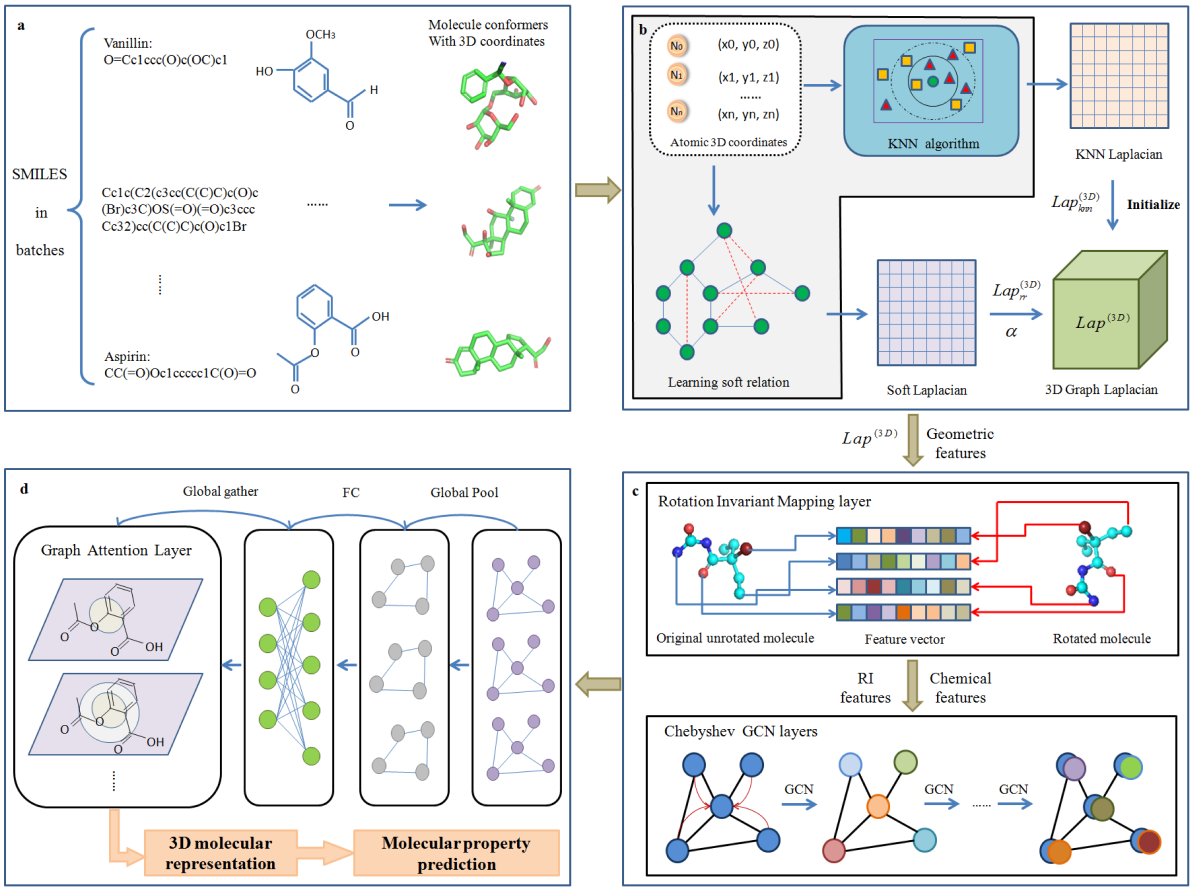

Fig. 1. Molecular adaptive graph representation architecture 3DMol-net. (a) Preprocessing the datasets with the format of SMILES to generate 3D molecular conformers. (b) Constructing the 3D graph Laplacian composed of KNN Laplacian and soft Laplacian. (c) Illustrating of the RIM and Chebyshev GCN layers. (d) Obtaining graph-level 3D representation via virtual node-based attention mechanism.

present 3D graph Laplacian Lap^i for $\mathcal{S}_i$ initialized by KNN Laplacian Lap^i_{knn} and optimized by soft relation Laplacian Lap^i_{sr} (Figure 1 (b)). Then, we feed Lap^i to RIM layer to obtain the RI feature, which is fed to Chebyshev GCN to obtain atomic representation (Figure 1 (c)). We further take the output of Chebyshev GCN as input to the virtual node-based attention module for obtaining the graph-level feature for predicting various tasks (Figure 1 (d)).

We first present the soft relations (Section 3.1) then construct 3D graph Laplacian to learn 3D topology of graph structure (Section 3.2). We finally propose RIM to obtain the rotation-invariant 3D structural features (Section 3.3).

A. Learning 3D soft relation for graph update

The original GCN [37] always has an edge between the associated nodes in a graph. If no relationship occurs between nodes, then no edge exists by default. In molecular structure, we always pay attention to whether a bond exists among atoms. If two atoms are associated, then a bond exists; otherwise, no bond exists. The above relationship between atoms is called a hard relation $\Phi_h \in \{0, 1\}$. However, it has a great limitation to use 0 and 1 to completely describe the relation between atoms. It is essential to find the deeper relationship between atoms in 3D space. Here, we call the relation as soft relation $\Phi_s \in [0, 1]$. An adaptive graph network needs to be built to mine Φ_s. The learning of Φ_s is of great significance to extract the deep atomic features for obtaining comprehensive molecular representations due to the complex interaction between atoms in 3D molecule.

We design a multilayer perceptron (MLP) network to learn the edge relationship and mine the soft relation. Let $edge$ be the input matrix of MLP. Each element g_{ij} of $edge$ is formed by concatenating the feature vectors of two different nodes i and j. The g_{ij} is also regarded as a feature of the edge between nodes i and j. Similarly the g_{ji} is the edge feature between node j and i. The feature vector for each atom here refers to the spatial coordinates in 3D space. Let $\hat{\mathcal{F}_{sr}}$ is the predictive soft relation matrix via MLP, and $\hat{f_{ij}}$ is the element of $\hat{\mathcal{F}_{sr}}$ between nodes i and j. To ensure the symmetry of prediction results, let $f_{ij} = \frac{\hat{f_{ij}}+\hat{f_{ji}}}{2}$. The final soft relation matrix $\mathcal{F}_{sr}$ can be obtained. It is a symmetric matrix, which is further transformed into a normalized Laplacian matrix $Lap^{(3D)}_{sr}$ as follows.

$$Lap^{(3D)}_{sr} = I_n - {D_{sr}}^{-\frac{1}{2}} \mathcal{F}_{sr} {D_{sr}}^{-\frac{1}{2}} \tag{1}$$

where D_{sr} is the degree matrix of $\mathcal{F}_{sr}$. In this way, a soft relation Laplacian matrix $Lap^{(3D)}_{sr}$ in 3D space is obtained, that is abbreviated as soft Laplacian in the following. Figure.2 shows three relations between atoms in this molecular diagram.

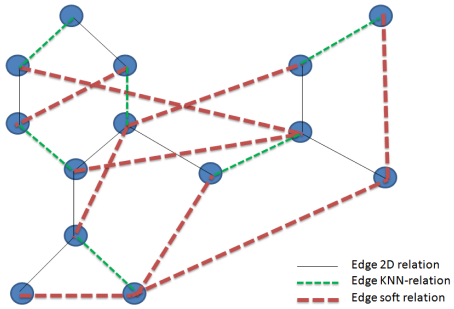

Fig. 2. Relationships between atoms in a 3D molecule. The relations based on 2D molecular graph are recorded as "Edge 2D relation", which are represented using black lines. The relations obtained using k-nearest neighbors algorithm are illustrated as "Edge KNN-relation", which are represented using green dotted lines. The red dotted lines are the relations of "Edge soft relation".

B. Constructing 3D graph Laplacian

The Laplacian matrix of 2D graphs can be easily obtained using the adjacency matrix of graph. However, it is difficult to obtain the Laplacian matrix accurately based on 3D spatial topology. Let $X = \{x_1, x_2, x_3, ...\} \in \mathcal{R}^3$ be a set of 3D points, then K-nearest neighbors (KNNs) of each point are found using kd-tree search algorithm. We suppose that the neighbouring points with large distance are less correlated. On this basis, the edge weight between points can be computed as below:

$$e_{i,j} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2 - \mu}{\sigma^2}}, & if \quad j \in \mathcal{N}_i \\ 0, & otherwise \end{cases} \tag{2}$$

where $\mathcal{N}_i$ is a KNN set of point i, μ is the mean, σ is the variance, and $e_{i,j}$ is the edge weight based on KNN between points i and j. An adjacency matrix obtained using the KNN algorithm is recorded as $\mathcal{F}_{knn}$. Its Laplacian is set to $Lap^{(3D)}_{knn}$, which can be calculated as

$$Lap^{(3D)}_{knn} = I_n - {D_{knn}}^{-\frac{1}{2}} \mathcal{F}_{knn} {D_{knn}}^{-\frac{1}{2}} \tag{3}$$

where D_{knn} is the degree matrix of $\mathcal{F}_{knn}$. Let $Lap^{(3D)}$ is the 3D graph Laplacian matrix, that is what we finally need to obtain. The $Lap^{(3D)}$ is first initialized by the $Lap^{(3D)}_{knn}$ Laplacian, and then further optimized by $Lap^{(3D)}_{rr}$ learned through neural network. This process is considered to be the process of constructing an adaptive graph, shown as Figure 1 (b). Inspired by literature [26], the optimal 3D graph Laplacian matrix $Lap^{(3D)}$ can be defined as follows:

$$Lap^{(3D)} = Lap^{(3D)}_{knn} + \alpha Lap^{(3D)}_{sr} \tag{4}$$

where $Lap^{(3D)}_{sr}$ is the soft Laplacian, and parameter α is a trainable parameter vector. The influence of $Lap^{(3D)}_{sr}$ on the final graph topology is controlled using α. $Lap^{(3D)}$ can extract structure characteristics of atoms in 3D molecules. The performance of downstream tasks can be improved by using $Lap^{(3D)}$ to propagate graph signal and update graph. $Lap^{(3D)}$ is abbreviated as 3D graph Laplacian.

C. Building 3DMol-Net architecture with RI

A graph convolution network is used to extract the features. Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be a graph with initial node features f_{x_v} for $v \in \mathcal{V}$ and edge features $f_{e_{uv}}$ for $(u, v) \in \mathcal{E}$. Graph network can obtain the representation h_v for every node via learning the graph structures. The m-th iteration of message passing can be defined as :

$$h_v^{(m)} = CO^{(m)}(h_v^{(m-1)}, AG^{(m)}(\left\{(h_v^{(m-1)}, h_u^{(m-1)}, f_{e_{uv}}) : u \in \mathcal{N}(v)\right\})) \tag{5}$$

where $h_v^{(m)}$ is the representation vector of node v at the m-th layer, $f_{e_{uv}}$ is the embedding of the edge between node u and node v, $\mathcal{N}(v)$ is the neighbor set of node v, $h_v^{(0)} = f_{x_v}$. $CO(\cdot)$ and $AG(\cdot)$ are the functions parameterized using neural networks. $AG(\cdot)$ is generally an aggregate function and $CO(\cdot)$ is an update function. In consideration of the high computational complexity of Laplacian eigendecomposition in spectral domain, Chebyshev GCN [34], [36] with K-order approximation can effectively reduce the computational complexity and obtain neighbor node information with K-hop, which is important for capturing complex relationships, especially long-range dependence. It also can effectively control the receptive field size. Therefore, We use Chebyshev GCN to capture the geometric information and extract abstract topological features from high-dimensional features transformed via RI. Then, the final convolution in the form of matrix can be expressed as:

$$Y(X) = \sum_{k=0}^{K-1} \theta_k^{'} T_k(\hat{\mathcal{L}}) X \tag{6}$$

where $T_k(\hat{\mathcal{L}}) = 2\hat{\mathcal{L}} T_{k-1}(\hat{\mathcal{L}}) - T_{k-2}(\hat{\mathcal{L}})$, $T_0(\hat{\mathcal{L}}) = 1$, $T_1(\hat{\mathcal{L}}) = \hat{\mathcal{L}}$ and $\hat{\mathcal{L}} = \frac{2}{\lambda_{max}} \mathcal{L} - I_n$, $\mathcal{L}$ is the normalized graph Laplacian. $T_k(\hat{\mathcal{L}})$ is a set of Chebyshev basis. $T_k(\hat{\mathcal{L}})X$ is a projection of the input features to the Chebyshev basis. $\theta^{'} = \left\{\theta_0^{'}, \theta_1^{'}, ..., \theta_{k-1}^{'}\right\}$ are learnable weights shared across nodes. Here, $\hat{\mathcal{L}} = Lap^{(3D)}$, that is 3D graph Laplacian obtained from the previous section. Algorithm 1 describes the RI-based adaptive graph convolution process for 3D molecular representation.

Traditional convolution is mainly applied to Euclidean structure data, such as images, which can guarantee convolution properties such as translation invariance and rotation invariance by using data augmented from different perspectives. The invariance of transformation on non-Euclidean data is still an urgent problem to be solved, especially in dealing with the task of geometric transformation. RI is a prerequisite that needs to be satisfied for learning the 3D topological structure of atoms effectively in molecules. The problem definition of rotation-invariant mapping (RIM) is given as:

Definition 1 (RIM). *Let F be a mapping function that learns 3D rotation-invariant features. X and $\bar{X}$ are the features before and after the mapping. $\bar{X} = F(X)$. Let R be an arbitrary rotation matrix. RIM can be described as: $\bar{X} = F(RX) = F(X), X \in \mathcal{R}^{N \times G}$ with N nodes and G attributes. Then $F(X)$ is a rotation-invariant mapping.*

In this paper, we use 3D coordinates as the feature for each atom, hence, $X \in \mathcal{R}^{N \times 3}$. After F mapping, the original

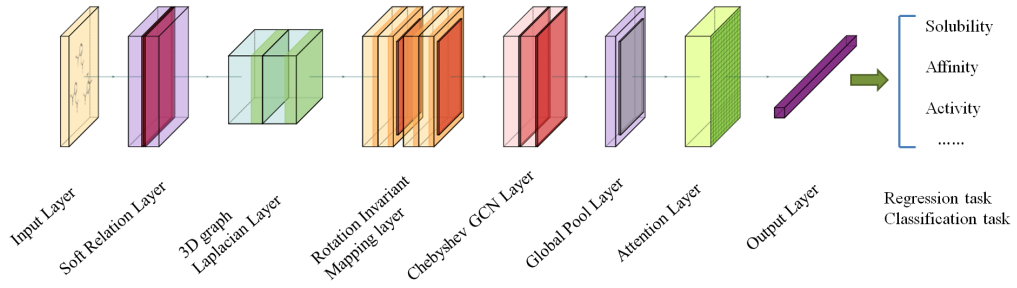

Fig. 3. The framework of 3DMol-Net. The RIM and Chebyshev GCN layer can extract 3D features of each node in a molecule. The global molecular representation can be obtained using attention.

Algorithm 1 RI-based adaptive graph convolution algorithm for 3D molecular representation

Input: dateset $\mathcal{S}$, 3D coordinates vector V_{coord}, multilayer perceptron network MLP, chemical features Att_{atom}, edge original feature $edge$, KNN algorithm KNN, concatenate function $Concatenate$, RIM function RIM, Chebyshev GCN $ChebyshevGCN$, virtual node-based attention $Virtual\_node\_Attention$.

Parameter: Chebyshev order k, layer size l, convolutional feature dimensions d.

Output: 3D molecular representation.

$\mathcal{F}_{sr} \leftarrow MLP(edge)$ ▷ Satisfy symmetry
$\mathcal{F}_{knn} \leftarrow KNN(\mathcal{S})$ ▷ Equ (2)
$Lap_{sr}^{(3D)} \leftarrow I_n - {D_{sr}}^{-\frac{1}{2}} \mathcal{F}_{sr} {D_{sr}}^{-\frac{1}{2}}$ ▷ Equ(1)
$Lap_{knn}^{(3D)} \leftarrow I_n - {D_{knn}}^{-\frac{1}{2}} \mathcal{F}_{knn} {D_{knn}}^{-\frac{1}{2}}$ ▷ Equ(3)
$Lap^{(3D)} \leftarrow Lap_{knn}^{(3D)} + \alpha Lap_{sr}^{(3D)}$ ▷ Equ(4)
$F_{ri\_geom} \leftarrow RIM(Lap^{(3D)}, V_{coord})$ ▷ Equ (7)
$F_{ri} \leftarrow Concatenate\ (F_{ri\_geom}, Att_{atom})$
while $n = 1 : l$ **do**
    $X_{atom}^{(n)} \leftarrow ChebyshevGCN(F_{ri}, Lap^{(3D)}, k, d)$ ▷ Equ (6)
    $F_{ri} \leftarrow X_{atom}^{(n)}$
end while
$F_{mol} \leftarrow Virtual\_node\_Attention(X_{atom}^{(l)})$ ▷ Equ (8, 9)
return F_{mol}.

molecular 3D topology features are transformed into high-dimensional feature $\bar{X}$. Motivated by transform-invariant geometry features [25] in 3D point clouds, computer vision [40], and signal processing [41], we propose RI based on soft relation and 3D graph Laplacian for molecular representation.

Lemma 1. *Let $\mathcal{L}$ be a Laplacian matrix of adjacency matrix A and $\widetilde{X} = \mathcal{T}(\mathcal{L})X = h_0 IX + h_1 \mathcal{L}X + ... + h_{K-1}\mathcal{L}^{K-1}X$. Let $\mathcal{F}_j(X) = \|(\mathcal{T}(\mathcal{L})X)_j\|_2^2, j \in 0, ..., K-1$, where $\|\cdot\|_2$ indicates $(\mathcal{T}(\mathcal{L})X)_j$ norm of each row and $\{h_0, h_1, ..., h_{K-1}\}$ is a group of filter coefficient, then $\mathcal{F}_j(X)$ is rotation and translation invariant [25].*

Due to $\mathcal{L} = I_n - \mathbf{D}^{-\frac{1}{2}}\boldsymbol{A}\mathbf{D}^{-\frac{1}{2}}$, then $\|\mathcal{L}X\|$ indicates the Euclidean distance from the center point to the weighted average point. $\|\cdot\|_2^2$ is a rigorously rotation-invariant operator. Theorem 1 is proposed as below:

Theorem 1. *For rotation- and translation-invariant $\mathcal{F}_j(X)$, defined in Lemma 1, let*

$$\widetilde{\mathcal{F}}_j(X) = \left\| (\mathcal{T}(Lap^{(3D)})X)_j \right\|_2^2, j \in 0, ..., K-1. \quad (7)$$

where $Lap^{(3D)}$ is the 3D graph Laplacian, and $\|\cdot\|_2$ indicates $(T(Lap^{(3D)})X)_j$ norm of each row. $\widetilde{\mathcal{F}}_j(X)$ is rotation and translation invariant. $\Psi_{ChebyGCN}(\widetilde{\mathcal{F}}_j(X))$ is also rotational and translation invariant, where $\Psi_{ChebyGCN}$ is the Chebyshev GCN mapping, that is the result of implementing Chebyshev graph convolutional layer.

The proof of Theorem 1 is included in the supplementary material with reference to the proof in [25]. 3DMol-Net is an adaptive GCN to learn 3D molecular representation based on RI. The overview of 3DMol-Net is shown in Figure 3. The 3D spatial coordinates are input to $\widetilde{\mathcal{F}}_j(X)$ to achieve the final rotation-invariant feature. Then, the RI feature and chemical feature (such as atomic symbol, degree, valence) are concatenated together as a feature vector that is input to Chebyshev GCN (Figure 1 (c)). For obtaining the graph-

level features, an attention is implemented (Figure 1 (d)), that considers each molecule as a virtual node connected to every atom in the molecule by using a special edge type. The process is performed on stacked attention layers for molecule-level embedding. The attention weight and context can be formulated as:

$$att_{mn} = \frac{exp(LeakyRelu(W \cdot [h_m, h_n]))}{\sum_{n \in N(m)} exp(LeakyRelu(W \cdot [h_m, h_n]))} \quad (8)$$

$$context_m = Elu(\sum_{n \in N(m)} att_{mn} \cdot W \cdot h_n) \quad (9)$$

where m is the target node, n is the neighbor atom, and h_m, h_n represent the state vectors of node m and n, respectively. W is a trainbale weight matrix. $LeakeRelu$ and Elu are nonlinear activation functions. As a result, a graph-level representation can be achieved.

IV. Experiments

Extensive experiments have been conducted using different datasets to evaluate the performance of our proposed model.

A. Experiment Setups

Benchmark Datasets and Preprocessing. We evaluate the proposed 3DMol-Net on the basis of several commonly molecule datasets recommended by MoleculeNet [5]. We choose regression task and classification task datasets to evaluate 3DMol-Net, including the physical chemistry datasets (ESOL, Lipophilicity) and biophysics bioactivity datasets (HIV, BACE). The details of datasets are shown in Table I, the metrics is recommended by MoleculeNet.

- ESOL. ESOL dataset [42] contains the water solubility (log solubility in moles per liter) data of 1,128 compounds. This dataset is based on the SMILES structure and is often used to train models that need to estimate water solubility from chemical structures. However, this dataset does not include the 3D coordinates of atoms. The ESOL dataset can be used for regression task in neural networks.
- Lipophilicity. Lipophilicity [43] is curated from ChEMBL database, containing 4,200 experimental results of octanol/water distribution coefficient (logD at pH 7.4). The task is of high significance to drug development due to the importance of lipophilicity in membrane permeability and solubility.
- HIV. HIV dataset [44] is introduced by the Drug Therapeutics Program AIDS Antiviral Screen. This organization tested 41,127 compounds to obtain their capability of inhibiting HIV replication. The compounds are divided into three categories: confirmed inactive(CI), confirmed moderately (CM) active and confirmed active(CA). In the classification task, the last two categories always are merged into one label as active (label = 1), and the first category CI is treated as one label as inactive (label = 0). This task is a typical dual-classification problem.
- BACE. BACE dataset [45] provides quantitative (IC_{50}) and qualitative (label) binding results for a set of inhibitors of human β-secretase 1 and contains 1,522 compounds with their 2D structures and binary labels.

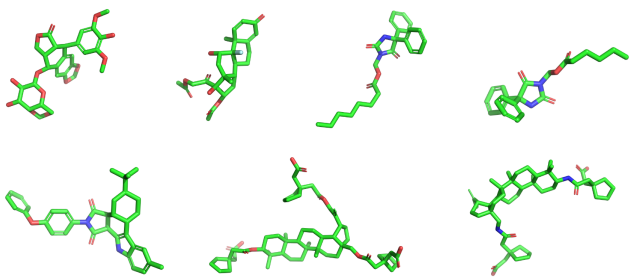

Fig. 4. Visualization of molecular 3D conformers that have been preprocessed with 3D structure. Each atom in the molecules has X, Y, Z coordinates.

We generate molecular conformers following the previous method [46] by using RDKit [47], which is also adopted in E3FP [22]. For each molecule, the number of conformations varies with the number of molecular rotational bonds. We minimize conformers with the Universal Force Field [48] in RDKit, then select a conformer with a minimum energy for each molecule, that contains the 3D atomic coordinates and bonds information. We normalize the coordinates to prepare for splitting data, which will be fed to a deep neural network in batches. Figure 4 shows the 3D conformers generated from our preprocessed datasets.

TABLE I
STATISTICS OF DATASETS.

Dataset	Data Type	Task Type	No.	Metric
ESOL	SMILES	Regression	1,128	MAE,RMSE
Lipophilicity	SMILES	Regression	4,200	MAE,RMSE
HIV	SMILES	Classification	41,127	AUROC,AUPRC
BACE	SMILES	Classification	1,522	AUROC,AUPRC

Baselines. We focus on the models based on graph neural networks, especially for processing molecular 3D topology structures. We compare the proposed method with the following graph-based models.

- Message Passing Neural Network (MPNN). MPNN is a generalized framework [27] for graph-based model. The prediction process consists of two parts: message passing and readout phases. The former aims to extract the abstract features of a graph, and the latter is to obtain a graph-level representation for predicting various tasks.
- Directed MPNN (D-MPNN). Inspired by structure2vec [49], D-MPNN [50] is introduced by adopting a message passing paradigm based on directed edges rather than nodes. It is a variant of the generic MPNN architecture.
- Weave. It is a molecular graph convolutional model [17], which extracts atom, bond, and graph distance information to form molecule-level representations and shows competitive performance.
- Graph convolutional models (GC). GC are neural graph fingerprints, proposed in [16], which extract molecular features on the basis of circular fingerprints for arbitrary-sized graphs. They are regarded as the implementation of circular ECFPs by using a graph neural network.

- 3DGCN. 3DGCN receives 3D molecular information from the augmented information of bond direction based original GCN model, proposed in [29]. It provides a critical step toward 3D research in chemistry.
- 3DTI-Net. 3DTI-Net achieves transform invariance by learning 3D features on the basis of a local graph representation for a point cloud task [25]. It is partially rewritten to adapt to molecular datasets because it is used as a strong baseline for comparison.
- 3DMol-Net (w/o RI). It is 3DMol-Net without RI for evaluating the validity of 3D RI in 3DMol-Net.
- 3DMol-Net. It is the method presented in this paper. 3DMol-Net pays considerable attention to the study of molecular characterization in 3D space. It attempts to explore the molecular conformation space and provides an insight into the topological study of molecules.

Implementation Details. 3DMol-Net is trained on the basis of Keras framework, with Tensorflow [51] as the backend and Adam [52] algorithm as the optimizer to apply gradient back-propagation. All datasets are split with the ratio of 80%, 10% and 10% for training, validation and test, respectively. The number of neighbors K in KNN algorithm is set to 5, the learning rate is 0.001 and learning rate decay is set using the ReduceLROnPlateau method ($factor = 0.9, patience = 10, min_lr = 0.0005$) in Keras. The Chebyshev order is set to 3 with 2 graph convolutional layers and 128 feature dimensions. Following [5], we use Mean Squared Error (MSE) as model training loss, Root Mean Squard Error (RMSE) and Mean Absolute Error (MAE) as test loss for regression. As for classification, cross entropy loss is applied to optimize model parameters, the area under the receiver operating characteristics curve (AUROC) and area under precision-recall curve (AUPRC) are used to evaluate the overall performance. All experiments are carried out on Ubuntu 16 with Intel® Core™ i9-7920X CPU @ 2.90GHz, 2 TITAN Xp GPUs with total 24GB memory.

B. Performance Comparison

Performance in Graph Regressions. Solubility and lipophilicity are basic physical chemistry properties. They are very important properties for understanding how molecules interact with solvents. Table II compares 3DMol-Net performances to other state-of-the-art graph-based models. The averaged RMSE values for ESOL dataset in validate set and test set are 0.589 log ($mol\ L^{-1}$) and 0.618 log ($mol\ L^{-1}$), that is slightly lower than MPNN model with the RMSE of 0.550 in validation set and 0.580 in test set. However our model almost reaches the consistent performance with MPNN, and 3DMol-Net uses fewer parameters in the training process. In addition, we obtains the best performance on Lipophilicity dataset. Another metrics results are shown in Table IV, which is a performance comparison based on the 3D graph model using MAE for ESOL, Lipophilicity datasets.

Performance in Graph Classifications. Table III shows the predictive AUROC values of biophysics bioactivity for HIV and BACE datasets. Compared with other 2D or 3D-based graph models, 3DMol-Net obtains the best performance. Since the proportion of positive and negative samples in HIV is seriously unbalanced, we randomly sample negative samples with a positive and negative sample ratio of 1:1. Table IV shows the predictive AUPRC value for HIV, BACE datasets.

In summary, 3DMol-Net shows competitive predictive performance in both regression tasks and classification tasks with fewer trainable parameters, which speeds up the convergence of the model and reduces the training time. It also further proves the importance and effectiveness of spatial topology in the accurate prediction of molecular properties.

C. Ablation Study

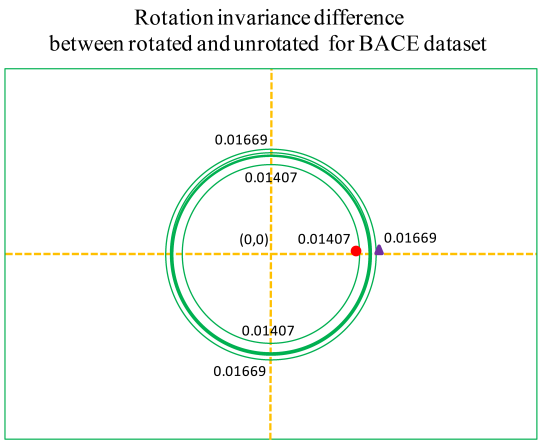

Fig. 6. RI difference in the values of the model evaluated at test set and the model evaluated at the same test set but rotated for BACE dataset. The rotated angles are $-120°$, $-60°$, $60°$, $120°$, $180°$, $0°$ and random rotation. The original unrotated point is set to the coordinate origin, that is the rotated $0°$ point.

We conducted ablation experiments on the four datasets to check if RI influence the performance of proposed 3DMol-Net framework. As shown in Table II, III and IV, 3DMol-Net with the RI shows the better performances than 3DMol-Net (w/o RI). It can be seen that RI is indispensable for 3DMol-Net architecture to improve performance. The experiment verifies the rotation invariance of 3DMol-Net as follows.

Verify RI of 3DMol-net. RI is a necessary property in the study of 3D geometric features. Current deep neural networks for learning 3D geometric features are vulnerable to 3D rotation. Here, we verify RI of 3DMol-net. Rotating molecular conformers of test set, then evaluating the rotated conformers using 3DMol-net model which has been trained using normal molecular conformers. We choose random rotation and rotate different degrees around X, Y, Z axes respectively. Figure 5 shows the verification results of RI for ESOL and HIV, respectively. Fig. 6 demonstrates RI for BACE in another way, that shows the difference in the values of the model evaluated at test set and the model evaluated at the same test set but rotated. The x-axis is AUROC, the y-axis is AUPRC. The original unrotated point is set to the coordinate origin. Then, the distances between the rotated points and the original unrotated point are used as the radii to draw circles. The red dot and purple triangle represent the lower and upper

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3089162, IEEE Journal of Biomedical and Health Informatics

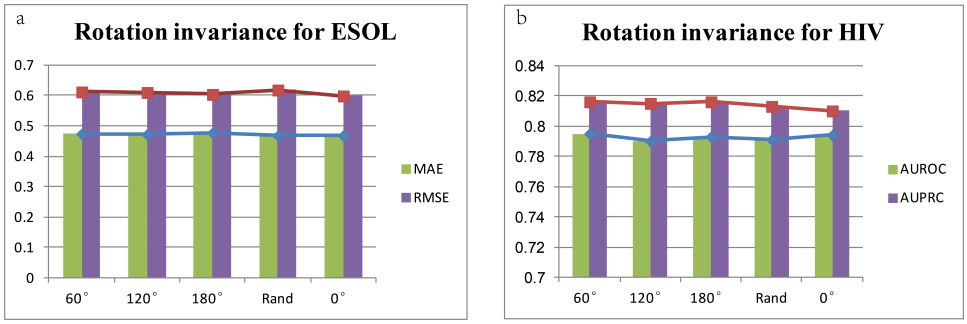

Fig. 5. RI for (a) ESOL dataset and (b) HIV dataset. The rotated angles of 3D molecular conformers are 60°, 120°, 180° and *Rand*, respectively. *Rand* represents random rotation. 0° represents no rotation. The maximum difference in MAE and RMSE for ESOL is 0.0092 and 0.0195, respectively. The maximum difference in AUROC and AUPRC for HIV is 0.005 and 0.006, respectively.

TABLE II
PREDICTIVE PERFORMANCES FOR ESOL, LIPOPHILICITY DATASETS WITH THE MEAN AND STANDARD DEVIATION OF RMSE USING A 10-FOLD CROSS VALIDATION. THE NUMBER OF TRAINABLE PARAMETERS IS ALSO SHOWN IN THE LAST COLUMN OF THE TABLE. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD. THE SMALLER THE RMSE, THE BETTER.

Dataset	Model type	Model	RMSE		Number of parameters
			Validation	Test	
ESOL	2D graph-based models	MPNN	**0.550±0.020**	**0.580±0.030**	0.34M
		Weave	0.570±0.040	0.610±0.070	0.38M
		GC	1.050±0.150	0.970±0.010	0.40M
	3D graph-based models	3DGCN	0.572±0.059	0.605±0.046	0.39M
		3DTI-Net	0.705±0.079	0.723±0.061	0.25M
		3DMol-Net (w/o RI)	0.662±0.072	0.728±0.058	0.23M
		3DMol-Net	0.589±0.045	0.618±0.047	**0.23M**
Lipophilicity	2D graph-based models	MPNN	0.757±0.030	0.719±0.031	0.34M
		Weave	0.734±0.011	0.715±0.035	0.38M
		GC	0.678±0.040	0.655±0.036	0.40M
	3D graph-based models	3DGCN	0.658±0.047	0.665±0.040	0.39M
		3DTI-Net	0.701±0.043	0.693±0.032	0.25M
		3DMol-Net (w/o RI)	0.651±0.032	0.698±0.039	0.23M
		3DMol-Net	**0.643±0.035**	**0.651±0.027**	**0.23M**

TABLE III
PREDICTIVE PERFORMANCES FOR HIV, BACE DATASETS WITH THE MEAN AND STANDARD DEVIATION OF AUROC USING A 10-FOLD CROSS VALIDATION. THE NUMBER OF TRAINABLE PARAMETERS IS ALSO SHOWN IN THE LAST COLUMN OF THE TABLE. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD. THE LARGER THE AUROC, THE BETTER.

Dataset	Model type	Model	AUROC		Number of parameters
			Validation	Test	
HIV	2D graph-based models	D-MPNN	0.806±0.016	0.776±0.008	0.35M
		Weave	0.742±0.040	0.703±0.039	0.38M
		GC	0.792±0.014	0.763±0.016	0.40M
	3D graph-based models	3DGCN	0.807±0.019	0.793±0.019	0.39M
		3DTI-Net	0.741±0.058	0.728±0.047	0.25M
		3DMol-Net (w/o RI)	0.748±0.026	0.723±0.018	0.23M
		3DMol-Net	**0.810±0.021**	**0.794±0.016**	**0.23M**
BACE	2D graph-based models	D-MPNN	0.865±0.041	0.838±0.056	0.35M
		Weave	0.638±0.014	0.806±0.002	0.38M
		GC	0.627±0.015	0.783±0.014	0.40M
	3D graph-based models	3DGCN	0.886±0.021	0.857±0.036	0.39M
		3DTI-Net	0.798±0.033	0.804±0.046	0.25M
		3DMol-Net (w/o RI)	0.808±0.025	0.791±0.028	0.23M
		3DMol-Net	**0.897±0.029**	**0.865±0.033**	**0.23M**

TABLE IV
PREDICTIVE PERFORMANCES BASED ON 3D GRAPH MODELS FOR ESOL, LIPOPHILICITY DATASETS WITH THE MAE (THE SMALLER THE VALUE, THE BETTER) AND FOR HIV, BACE DATASETS WITH THE AUPRC (THE LARGER THE VALUE, THE BETTER). THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Dataset	Metric	3D graph-based models			
		3DGCN	3DTI-Net	3DMol-Net (w/o RI)	3DMol-Net
ESOL	Validation MAE	**0.453±0.043**	0.523±0.058	0.536±0.055	0.464±0.038
	Test MAE	**0.475±0.040**	0.561±0.041	0.569±0.050	0.481±0.044
Lipophilicity	Validation MAE	0.496±0.033	0.535±0.045	0.550±0.039	**0.493±0.028**
	Test MAE	0.531±0.027	0.579±0.036	0.586±0.031	**0.523±0.026**
HIV	Validation AUPRC	0.408±0.036	0.758±0.043	0.714±0.039	**0.812±0.032**
	Test AUPRC	0.384±0.030	0.755±0.046	0.746±0.029	**0.801±0.035**
BACE	Validation AUPRC	**0.850±0.043**	0.808±0.038	0.785±0.031	**0.850±0.035**
	Test AUPRC	0.816±0.037	0.784±0.046	0.776±0.050	**0.822±0.043**

bounds of the rotation invariance difference, respectively. Obviously, 3DMol-net shows the consistent performance as before rotation, and the prediction accuracy is not reduced by the rotation of 3D conformers, although it has never been trained by a group of rotated 3D molecules. The consistent performance clearly shows that 3DMol-net architecture has the property of rotation invariance and can learn topological information of 3D molecules well.

D. Visualize the atomic contribution

Furthermore, the contribution of atoms to molecule-level feature for ESOL dataset is visualized by Figure 7, that is represented by colormap. The shown molecules are randomly selected from the test set. The hydrophilic groups play a more important role in predicting water solubility for ESOL. They have a greater contribution to molecular feature, that plays a decisive role in the final prediction. In addition, The most relevant group in predicting the activity of inhibiting HIV replication shows more contribution to graph-level representation. From another perspective, it shows that the proposed model has good interpretability. More visualizations can be found in supplementary material.

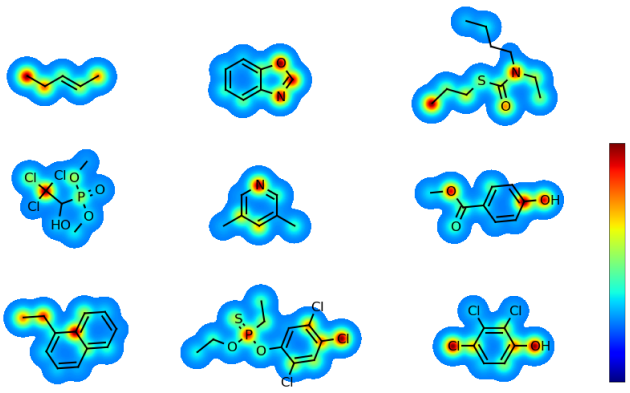

Fig. 7. Visualizing the contribution of atoms to molecule-level feature for example molecules chosen randomly from ESOL dataset. It is shown using jet colormap with the range of 0.25 to 1.0. The intensity represented by the colors in the colorbar gradually becomes stronger from bottom to top.

V. CONCLUSION

Molecules exist as 3D structures in the real world. The interactions among molecules are carried out in 3D space, which is greatly affected by molecular conformation and spatial position. However, in the field of deep learning chemistry, limited research is available on the 3D spatial information of molecules. In fact, the 3D structure of molecules plays a decisive role in biochemical function, which will further promote the development of the field of drug research for improving human well-being.

In this study, we propose the 3DMol-net architecture and prove its RI of 3D coordinates on the basis of graph convolutional neural network. Then, we construct adaptive graph topology that is learned from the invisible 3D spatial residual relationship between atoms, and construct a 3D graph Laplacian. Chebyshev GCN is used to predict molecular properties accurately from the 3D molecular structure. Our model is trained on several datasets in the Biochemistry field. The results show that 3DMol-net has competitive performance compared with the state-of-the-art deep learning graph models used in chemistry.

The research on the 3D spatial information of molecules is rarely compared with the research on 2D molecular information. We have attempted to provide an insight into the research of 3D molecules, which will play a positive role in the discovery and development of modern drugs. However, many tasks, such as a molecular docking task based on 3D molecular conformation and the interactions between molecules and proteins, need to be studied further. Our next work will continue to focus on the field of 3D molecular research. Determining how to represent molecules in a non-Euclidean space and predict the interaction between molecules and proteins accurately will be a considerably interesting topic.

APPENDIX

A. Method

Theorem 1. For rotation- and translation-invariant $\mathcal{F}_j(X)$, let

$$\widetilde{\mathcal{F}}_j(X) = \left\| (\mathcal{T}(Lap^{(3D)})X)_j \right\|_2^2, j \in 0, ..., K-1. \quad (10)$$

where $Lap^{(3D)}$ is the 3D graph Laplacian, and $\|\cdot\|_2$ indicates $\mathcal{T}(Lap^{(3D)})X)_j$ norm of each row. $\widetilde{\mathcal{F}}_j(X)$ is rotation and

translation invariant. $\Psi_{ChebyGCN}(\widetilde{\mathcal{F}}_j(X))$ is also rotational and translation invariant, where $\Psi_{ChebyGCN}$ is the Chebyshev GCN mapping, that is the result of implementing Chebyshev graph convolutional layer.

Proof. For any rotational matrix R,

$$\begin{aligned}\widetilde{\mathcal{F}}_j(XR) &= \left\|(\mathcal{T}(Lap^{(3D)})XR)_j\right\|_2^2 \\ &= (\mathcal{T}(Lap^{(3D)}))_j XRR^T X^T \mathcal{T}(Lap^{(3D)}))_j^T \\ &= (\mathcal{T}(Lap^{(3D)}))_j XX^T \mathcal{T}(Lap^{(3D)}))_j^T \\ &= \left\|(\mathcal{T}(Lap^{(3D)})X)_j\right\|_2^2 \\ &= \widetilde{\mathcal{F}}_j(X)\end{aligned} \tag{11}$$

where $X \in \mathcal{R}^{N\times 3}$, $RR^T = R^T R = I_3$. Thus, $\widetilde{\mathcal{F}}_j(X)$ is rotation invariant. The j-th output channel of rotation-invariant feature mapping can be indicated as $\widetilde{X}'_j = \sum_{t=0}^{K} \theta_{tj} \left\|(Lap^{(3D)})^t X\right\|_2^2 + b_j$, $\widetilde{X}'_j \in \mathcal{R}^{N\times NF}$, $j \in 0, 1, ..., NF-1$, where NF is the number of feature map and K is the polynomial order. From Definition 1, graph isometric transformation g, $(\mathcal{J}_g\widetilde{\mathcal{F}}_j)(X) = \widetilde{\mathcal{F}}_j(XR) = \widetilde{\mathcal{F}}_j(X)$, $\widetilde{\mathcal{F}}_j(X)$ is equivariant to graph isometric transformations. From Lemma 1, the Chebyshev polynomial filter is equivariant to the graph isometric transformation. Accordingly, the graph convolutional layer $\Psi_{ChebyGCN}(\widetilde{\mathcal{F}}_j(X))$ is equivariant to a graph isometric transformation. Therefore, $\Psi_{ChebyGCN}(\widetilde{\mathcal{F}}_j(X))$ is rotation invariant. Translation invariant can be easily obtained through re-centering mapping. □

B. Visualize the atomic contribution

Visualize the atomic contribution to molecular feature for Lipophilicity and BACE datasets, which are shown in Fig. 8.

REFERENCES

[1] J. Sieg, F. Flachsenberg, and M. Rarey, "In need of bias control: Evaluating chemical data for machine learning in structure-based virtual screening," *Chemical Information and Modeling*, vol. 59, no. 3, pp. 947–961, 2019.

[2] M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, and D. R. Koes, "Protein-ligand scoring with convolutional neural networks," *Chemical Information and Modeling*, vol. 57, no. 4, 2017.

[3] J. Lim, S. Ryu, K. Park, Y. J. Choe, and W. Y. Kim, "Predicting drug-target interaction using a novel graph neural network with 3D structure-embedded graph representation," *Chemical Information and Modeling*, vol. 59, no. 9, pp. 3981–3988, 2019.

[4] A. Paul, D. Jha, R. Al-Bahrani, W. keng Liao, A. Choudhary, and A. Agrawal, "Chemixnet: Mixed DNN architectures for predicting chemical properties using multiple molecular representations," in *Nips Workshop on Machine Learning for Molecules & Materials*, 2018.

[5] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing, and V. S. Pande, "Moleculenet: A benchmark for molecular machine learning," *Chemical Science*, vol. 9, no. 2, pp. 513–530, 2018. [Online]. Available: https://doi.org/10.1039/C7SC02664A.

[6] C. Li, J. Wang, Z. Niu, J. Yao, and X. Zeng, "A spatial-temporal gated attention module for molecular property prediction based on molecular geometry," *Briefings in Bioinformatics*, 2021. [Online]. Available: https://doi.org/10.1093/bib/bbab078.

[7] Y. Song, S. Zheng, Z. Niu, Z. Fu, Y. Lu, and Y. Yang, "Communicative representation learning on attributed molecular graphs," in *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 2020.

[8] S. Jin, X. Zeng, F. Xia, W. Huang, and X. Liu, "Application of deep learning methods in biological networks," *Briefings in Bioinformatics*, vol. 22, no. 2, 2021.

[9] X. Chen, D. Xie, L. Wang, Q. Zhao, Z. H. You, and H. Liu, "BNPMDA: Bipartite network projection for MiRNA-Disease association prediction," *Bioinformatics*, vol. 34, no. 18, pp. 3178–3186, 2018.

[10] C. Li, H. Liu, Q. Hu, J. Que, and J. Yao, "A novel computational model for predicting microRNA-Disease associations based on heterogeneous graph convolutional networks," *Cells*, vol. 8, no. 9, p. 977, 2019.

[11] J. Wang, J. Li, K. Yue, L. Wang, Y. Ma, and Q. Li, "NMCMDA: neural multicategory mirna - disease association prediction," *Briefings in Bioinformatics*, 2021. [Online]. Available: https://doi.org/10.1093/bib/bbab074.

[12] A. C. Mater and M. L. Coote, "Deep learning in chemistry," *Chemical Information and Modeling*, vol. 59, no. 6, pp. 2545–2559, 2019.

[13] X. Zeng, S. Zhu, X. Liu, Y. Zhou, and F. Cheng, "deepDR: A network-based deep learning approach to in silico drug repositioning," *Bioinformatics*, vol. 35, no. 24, pp. 5191–5198, 2019.

[14] X. Lin, Z. Quan, Z. Wang, H. Huang, and X. Zeng, "A novel molecular representation with BiGRU neural networks for learning atom," *Briefings in Bioinformatics*, Nov 2019, bbz125. [Online]. Available: https://doi.org/10.1093/bib/bbz125.

[15] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *Chemical Information and Modeling*, vol. 50, no. 5, pp. 742–754, 2010.

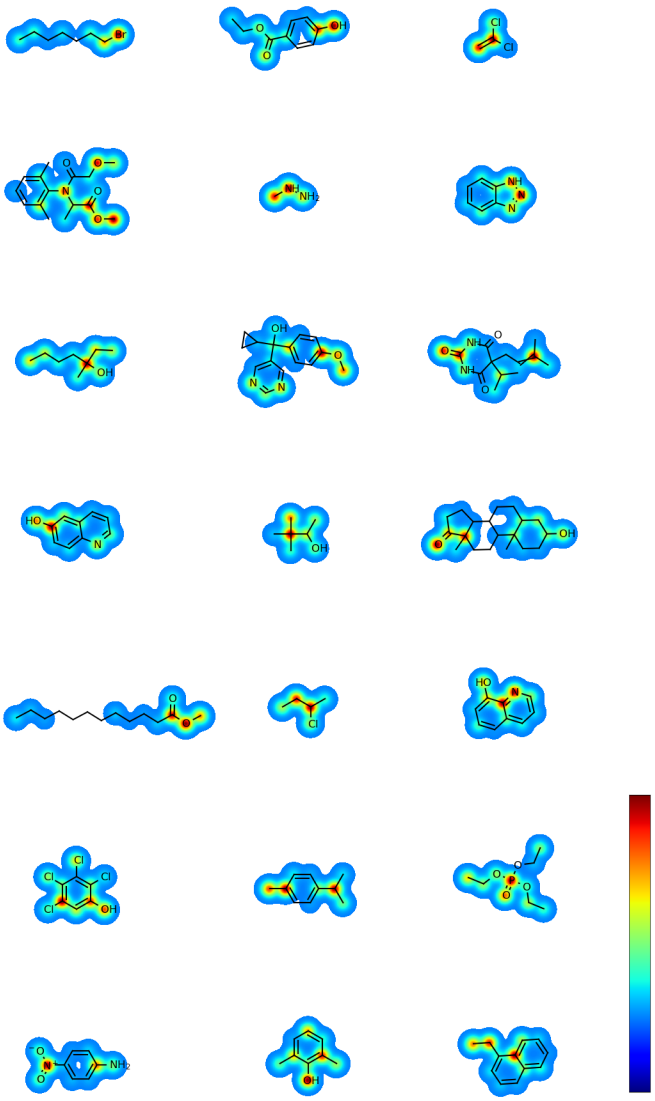

Fig. 8. Visualizing the contribution of atoms to molecule-level feature for example molecules chosen randomly from ESOL dataset. It is shown using jet colormap with the range of 0.25 to 1.0. The intensity represented by the colors in the colorbar gradually becomes stronger from bottom to top.

[16] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," in *Proc. NIPS*, 2015, pp. 2224–2232. [Online]. Available: http://papers.nips.cc/paper/5954-convolutional-networks-on-graphs-for-learning-molecular-fingerprints.pdf.

[17] S. Kearnes, K. Mccloskey, M. Berndl, V. Pande, and P. Riley, "Molecular graph convolutions: moving beyond fingerprints," *Computer-Aided Molecular Design*, vol. 30, no. 8, pp. 595–608, 2016.

[18] I. Wallach, M. Dzamba, and A. Heifets, "Atomnet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery," *ArXiv*, 2015. [Online]. Available: https://arxiv.org/abs/1510.02855.

[19] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, "Schnet: a continuous-filter convolutional neural network for modeling quantum interactions," in *Proceedings of the 31st International Conference on Neural Information Processing (NIPS)*, 2017.

[20] Y. Li, J. Pei, and L. Lai, "Learning to design drug-like molecules in three-dimensional space using deep generative models," *ArXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2104.08474.

[21] S. Axelrod and R. Gómez-Bombarelli, "Molecular machine learning with conformer ensembles," *ArXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2012.08452.

[22] Axen, Seth, D., Roth, Bryan, L., Huang, Xi-Ping, Keiser, and M. and, "A simple representation of three-dimensional molecular structure," *Medicinal Chemistry*, vol. 60, no. 17, pp. 7393–7409, 2017.

[23] P. Gainza, F. Sverrisson, F. Monti, E. Rodolà, D. Boscaini, M. M. Bronstein, and B. E. Correia, "Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning," *Nature Methods*, 2020.

[24] H. Cho and I. S. Choi, "Enhanced deep-learning prediction of molecular properties via augmentation of bond topology," *ChemMedChem*, vol. 14, no. 17, pp. 1604–1609, 2019.

[25] G. Pan, J. Wang, R. Ying, and P. Liu, "3DTI-Net: Learn inner transform invariant 3D geometry features using dynamic GCN," *ArXiv*, 2018. [Online]. Available: https://arxiv.org/abs/1812.06254.

[26] R. Li, S. Wang, F. Zhu, and J. Huang, "Adaptive graph convolutional neural networks," in *The 32nd AAAI on Artificial Intelligence*, 2018. [Online]. Available: http://https://arxiv.org/abs/1801.03226.

[27] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," *ArXiv*, 2017. [Online]. Available: https://arxiv.org/abs/1704.01212.

[28] Z. Hao, C.Lu, Z.Huang, H. Wang, Z. Hu, Q. Liu, E.Chen, and C. Lee, "ASGN: An active semi-supervised graph neural network for molecular property prediction," in *Proceeding of the 26th ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2020.

[29] H. Cho and I. S. Choi, "Three-dimensionally embedded graph convolutional network (3DGCN) for molecule interpretation," *ArXiv*, 2018. [Online]. Available: https://arxiv.org/abs/1811.09794.

[30] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *The Annals of Statistics*, 2000.

[31] S. J. Swamidass, C. A. Azencott, T. W. Lin, H. Gramajo, S. C. Tsai, and P. Baldi, "Influence relevance voting: An accurate and interpretable virtual high throughput screening method," *Journal of Chemical Information and Modeling*, vol. 49, pp. 756–766, 2009.

[32] B. Ramsundar, S. Kearnes, P. Riley, D. Webster, D. Konerding, and V. Pande, "Massively multitask networks for drug discovery," *ArXiv*, 2015. [Online]. Available: https://arxiv.org/abs/1502.02072.

[33] J. P. Ebejer, G. M. Morris, and C. M. Deane, "Is multitask deep learning practical for pharma?" *Journal of Chemical Information and Modeling*, vol. 57, no. 8, pp. 2068–2076, 2017.

[34] J. Bruna, W. Zaremba, A. Szlam, and Y. Lecun, "Spectral networks and locally connected networks on graphs," in *Proc. ICLR*, 2014. [Online]. Available: http://arxiv.org/abs/1312.6203.

[35] G. R. M, D. K. Hammond, and P. Vandergheynst, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[36] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. NIPS*, 2016.

[37] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *ArXiv*, 2016. [Online]. Available: https://arxiv.org/abs/1609.02907.

[38] Behler and Jörg, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *Chemical Physics*, vol. 134, 2011. [Online]. Available: https://doi.org/10.1063/1.3553717.

[39] R. Khasanova and P. Frossard, "Isometric transformation invariant graph-based deep neural network," *ArXiv*, 2018. [Online]. Available: https://arxiv.org/abs/1808.07366.

[40] Q. Yang, C. Li, W. Dai, J. Zou, G. Qi, and H. Xiong, "Rotation equivariant graph convolutional network for spherical image classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 4303–4312.

[41] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[42] J. S. Delaney, "ESOL: estimating aqueous solubility directly from molecular structure," *Chemical Information and Computer Sciences*, vol. 44, no. 3, p. 1000, 2004.

[43] M. Wenlock and N. Tomkinson. Experimental in vitro DMPK and physicochemical data on a set of publicly disclosed compounds. [Online]. Available: https://doi.org/10.6019/chembl3301361.

[44] Zaharevd. (2004) Aids antiviral screen data. [Online]. Available: https://wiki.nci.nih.gov/display/NCIDTPdata/AIDS+Antiviral+Screen+Data

[45] Denny, Rajiah, Aldrin, Pande, Vijay, Subramanian, Govindan, Ramsundar, and Bharath, "Computational modeling of beta-secretase 1 (bace-1) inhibitors using ligand based approaches," *Chemical Information and Modeling*, 2016.

[46] J. P. Ebejer, G. M. Morris, and C. M. Deane, "Freely available conformer generation methods: How good are they?" *Chemical Information and Modeling*, vol. 52, no. 5, pp. 1146–1158, 2012.

[47] G. Landrum. (2010) RDKit: Open-source cheminformatics. [Online]. Available: http://www.rdkit.org

[48] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. G. III, and W. M. Skiff, "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations," *American Chemical Society*, vol. 114, no. 25, pp. 10 024–10 035, 1992.

[49] H. Dai, B. Dai, and L. Song, "Discriminative embeddings of latent variable models for structured data," in *Proc. ICML*, 2016.

[50] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen, and R. Barzilay, "Analyzing learned molecular representations for property prediction," *Chemical Information and Modeling*, vol. 59, no. 8, pp. 3370–3388, 2019.

[51] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, ser. OSDI 2016. USA: USENIX Association, 2016, pp. 265–283.

[52] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, Dec 2014.