GeomGCL: Geometric Graph Contrastive Learning for Molecular Property Prediction

Shuangli Li[1, 2]*, Jingbo Zhou[2]†, Tong Xu[1], Dejing Dou[2]†, Hui Xiong[3]†

[1]School of Computer Science and Technology, University of Science and Technology of China
[2]Business Intelligence Lab, Baidu Research
[3]Artificial Intelligence Thrust, The Hong Kong University of Science and Technology
lsl1997@mail.ustc.edu.cn, {zhoujingbo, doudejing}@baidu.com, tongxu@ustc.edu.cn, xionghui@ust.hk

Abstract

Recently many efforts have been devoted to applying graph neural networks (GNNs) to molecular property prediction which is a fundamental task for computational drug and material discovery. One of major obstacles to hinder the successful prediction of molecular property by GNNs is the scarcity of labeled data. Though graph contrastive learning (GCL) methods have achieved extraordinary performance with insufficient labeled data, most focused on designing data augmentation schemes for general graphs. However, the fundamental property of a molecule could be altered with the augmentation method (like random perturbation) on molecular graphs. Whereas, the critical geometric information of molecules remains rarely explored under the current GNN and GCL architectures. To this end, we propose a novel graph contrastive learning method utilizing the geometry of the molecule across 2D and 3D views, which is named GeomGCL. Specifically, we first devise a dual-view geometric message passing network (GeomMPNN) to adaptively leverage the rich information of both 2D and 3D graphs of a molecule. The incorporation of geometric properties at different levels can greatly facilitate the molecular representation learning. Then a novel geometric graph contrastive scheme is designed to make both geometric views collaboratively supervise each other to improve the generalization ability of GeomMPNN. We evaluate GeomGCL on various downstream property prediction tasks via a finetune process. Experimental results on seven real-life molecular datasets demonstrate the effectiveness of our proposed GeomGCL against state-of-the-art baselines.

1   Introduction

The prediction of molecular property has been widely considered as one of the most significant tasks in computational drug and material discovery (Goh, Hodas, and Vishnu 2017; Wu et al. 2018; Chen et al. 2018). Accurately predicting the property can help to evaluate and select the appropriate chemical molecules with desired characteristics for many downstream applications (Xiong et al. 2019; Yang et al. 2019; Song et al. 2020; Chen et al. 2021).

With the remarkable success of graph neural networks (GNNs) in various graph-related tasks in recent years (Wu et al. 2020), a number of efforts have been made from different directions to design GNN models for molecular property prediction like (Yang et al. 2019; Danel et al. 2020; Maziarka et al. 2020; Song et al. 2020; Chen et al. 2021). The fundamental idea is to regard the topology of atoms and bonds as a graph, and translate each molecule to a representation vector with powerful GNN encoders, followed by the prediction model for the specific property.

Along the other line of development for GNNs, graph contrastive learning (GCL) methods (Wu et al. 2021) have shown promising performance in many applications when there is a lack of sufficient labeled data. The scarcity of labeled data is one of the major obstacles to hinder the prediction performance of GNN models (as well as other deep learning models) for molecular property prediction. For example, it usually requires a high cost to collect the labeled data for certain computational drug discovery task.

More special attention should be paid to developing GCL for molecular property prediction. Existing GCL methods usually adopt different data augmentation schemes for the graph, which may change the semantics of graphs across domains. Most of the current GCL methods on molecular graphs are still based on such data augmentation paradigm which could inevitably alter the natural structure of a molecule. For example, You et al. (2020) proposes to drop atoms, perturb edges and mask attributes to augment the data. However, since each of atoms has an effect on the molecular property, such random dropping and perturbation of atoms could destroy the structure of a molecule. Though some other methods like MoCL (Sun et al. 2021) adopt predefined sub-structures of the molecule to alleviate the problem of random corruption, such substitution rules still have a probability to violate the chemical principle.

Our insight is to design a new GCL method for molecular property prediction from different geometric views without corrupting the molecular structure. As shown in Figure 1, the molecules can be represented as two-dimensional (2D) and three-dimensional (3D) structural graphs. Although there are emerging geometric GNN models which can make use of multiple factors from the 2D chemical graph or 3D spatial graph (Maziarka et al. 2020; Klicpera, Groß, and

Günnemann 2020; Shui and Karypis 2020; Danel et al. 2020), all of them capture the geometric information from a single view for representation learning. In practice, the 2D view and 3D view of a molecule are generated based on different methods: the 2D view derives directly from the structural formula of a chemical compound, while the 3D view is usually estimated by conformation generation procedure (Hawkins 2017) using the tools like RDKit (Tosco, Stiefl, and Landrum 2014). Considering that molecular graphs in 2D and 3D views can provide chemical and geometric information at different levels and further complement each other, there is a demand to develop a new paradigm for the molecule-driven contrastive learning without changing chemical semantics. Therefore, such different views of one molecule provide a great opportunity for designing a unique graph contrastive learning scheme on molecular graphs.

To tackle the aforementioned challenges, we propose a novel geometric-enhanced graph contrastive learning model (GeomGCL) for molecular property prediction, which is equipped with the adaptive geometric message passing network (GeomMPNN) as well as a contrastive learning strategy to augment the 2D-3D geometric structure learning process. Firstly, we devise the dual-channel geometric learning procedure to adapt the graph aggregation to both views with leveraging distance and angle information at different granularity levels. Secondly, we further aim to make both views complement each other for better geometric comprehension. Here, we take the molecular geometry into consideration and propose the 2D-3D geometric contrastive scheme to bridge the knowledge gap between geometric structure modeling and graph representation learning without labels. The representative 3D spatial information can be distilled with the guidance of the stable 2D information, which provides the chemical semantics. The fusion of 2D and 3D graphs promotes the proposed GeomGCL extract more expressive representation for property prediction. To summarize, the main contributions of our work are as follows:

- To the best of our knowledge, we are among the first to develop the contrastive learning method for molecular graphs based on geometric views. By means of naturally contrasting the 2D and 3D view graphs of one molecule without any random augmentation process, our proposed GeomGCL makes best of the consistent and realistic structures for better representation learning.
- The GeomGCL employs the dual-channel message passing neural network, which adaptively captures both 2D and 3D geometric information of molecular graphs. The additional spatial regularizer further preserves the relative relation of geometry and improves the performance.
- The experiments on seven molecular datasets demonstrate that the proposed model outperforms the state-of-the-art GNN and graph contrastive learning methods.

2 Related Work

The research topic of this paper is associated with molecular representation learning based on graph neural networks, especially the promising geometric and contrastive learning on molecular graphs. We will briefly discuss these topics.
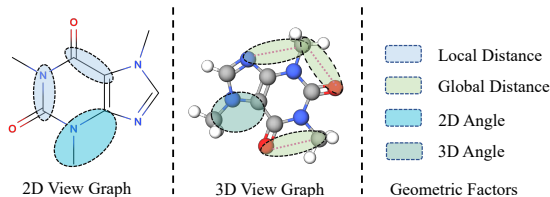

Figure 1: An illustrative example of geometric distances and angles in 2D view and 3D view graphs.

2.1 Molecular Representation Learning

As the important basis of property prediction, molecular representation learning has been a popular research area. The earlier feature-based methods learn the fixed representations from molecular descriptors or chemical fingerprints (Rogers and Hahn 2010), which ignore the graph structures and rely on the feature engineering. Recent years have witnessed the great advantage of graph neural networks (GNNs) in modeling graph data (Wu et al. 2020), much attention has been paid to applying GNN models to learn molecular graph representations. The graph convolution model (Duvenaud et al. 2015) is first introduced to encode molecular graphs based on atomic features. AttentiveFP (Xiong et al. 2019) adopts the graph attention network to make the best of graph structures and has become one of the state-of-the-art methods for molecular property prediction. More recently, incorporating bond features into the message passing neural networks (Gilmer et al. 2017) turns into a trend of learning better representations. DMPNN (Yang et al. 2019) is proposed to perform the edge-based message passing process over the edge-oriented directed graph, which obtains atom and bond embeddings concurrently. The communicative message passing models (Song et al. 2020; Chen et al. 2021) further extend this work through improving the node-edge interactive kernels as well as applying the transformer framework to capture long-range dependencies. Nevertheless, these approaches can not deal with the geometry of molecules and lack the ability of learning the non-local correlations among atomic nodes on the molecular graph.

2.2 Geometric Learning on Molecular Graphs

In the field of deep learning, geometry-based methods have shown prominent performance (Bronstein et al. 2017; Zhou et al. 2020; Li et al. 2020). Since molecules have geometric structures intrinsically, a few attempts have also been made to develop geometric graph learning models for the molecular graphs (Atz, Grisoni, and Schneider 2021). From the 2D view of the molecular graph, Recent works (Maziarka et al. 2020; Ying et al. 2021) are designed to encode interatomic distances with augmenting the attention mechanism in a transformer architecture. Meanwhile, there are some efforts to model the 3D molecular structures. SGCN (Danel et al. 2020) simply utilizes the 3D coordinates to apply the aggregation process. Such an intuitive method is sensitive to the coordinate systems, which leads to the poor performance of learning geometric graphs. Furthermore, several models that are invariant to translation and rotation of atom coordinates are proposed through designing the geometric

kernels (Klicpera, Groß, and Günnemann 2020) or strengthening the node-edge interactions with geometric information (Shui and Karypis 2020). However, these studies have shown not just strengths but also some limitations. Firstly, most of the efficient molecule-oriented geometric learning methods target at the quite small molecules for quantum property prediction. Secondly, none of these methods incorporate 2D and 3D geometric information simultaneously. To overcome these limitations, we propose to learn the 2D-3D geometric factors adaptively and synergistically.

2.3 Contrastive Learning on Molecular Graphs

Along the other line of development, graph contrastive learning methods (Wu et al. 2021) have their own advantages and have achieved extraordinary performance in many applications (You et al. 2020; Qiu et al. 2020; Wang et al. 2021). However, the existing models designed for molecular graphs receive little attention. InfoGraph (Sun et al. 2019) manages to maximize the mutual information between the representations of the graph and its substructures to guide the molecular representation learning. To alleviate the problem of random corruption on molecular graphs which may alter the chemical semantics, MoCL (Sun et al. 2021) adopts the domain knowledge-driven contrastive learning framework at both local- and global-level to preserve the semantics of graphs in the augmentation process. However, the learning ability of such model depends on the well-designed substitution rules. The deficiency of geometric information in graph contrastive learning limits the capability of effective molecular representation learning. To this end, we develop a novel contrastive learning model to integrate the geometry of molecules with chemical semantics by means of contrasting the 2D and 3D view graphs.

3 Preliminaries

Generally, a molecular graph with geometric information can be represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{C})$, where $\mathcal{V}$ and $\mathcal{E}$ denote the node (atom) set and edge (bond) set respectively. $\boldsymbol{C} \in \mathbb{R}^{|V| \times d}$ denotes the coordinate matrix for atoms, where $d \in \{2, 3\}$ is the spatial dimension. Given a molecule, the specific 2D and 3D view graphs are defined as following.

Definition 1 (2D View Graph.) *The 2D view graph is defined as $\mathcal{G}^{2d} = (\mathcal{V}, \mathcal{E}^{2d}, \boldsymbol{C}^{2d})$, where the edges in $\mathcal{E}^{2d}$ correspond to the primary covalent bonds in the molecule, and $\boldsymbol{C}_v^{2d} = \{x, y\}$ denotes the coordinate of atom v.*

Definition 2 (3D View Graph.) *Similarly, the 3D view graph can be represented as $\mathcal{G}^{3d} = (\mathcal{V}, \mathcal{E}^{3d}, \boldsymbol{C}^{3d})$. Note that the generated coordinate in three dimensional space is non-deterministic by means of the estimation algorithm, thus we repeatedly generate coordinates P times to tackle the randomness problem and $\boldsymbol{C}^{3d} = \frac{1}{P} \sum_{p=1}^{P} \boldsymbol{C}^{3d,p}$. The edge set $\mathcal{E}^{3d}$ is constructed based on the 3D spatial coordinates, which contains all edges whose distances are smaller than the cutoff threshold d_θ. It can be formulated as $\mathcal{E}^{3d} = \{e_{uv} | dist(\boldsymbol{C}_u^{3d}, \boldsymbol{C}_v^{3d}) < d_\theta\}$.*

Problem Statement. Given a molecule, we can construct the 2D view graph $\mathcal{G}^{2d}$ and 3D view graph $\mathcal{G}^{3d}$.

Let $\boldsymbol{X}$ be the atom feature matrix and $\boldsymbol{E}$ be the bond feature matrix, our goal is to train a geometric graph encoder $f(\mathcal{G}^{2d}, \mathcal{G}^{3d}, \boldsymbol{X}, \boldsymbol{E})$ to learn the molecular representation vector $\boldsymbol{h}$ without any label information. Then the well-trained model is utilized for various downstream property prediction tasks through the finetune process.

4 Model Framework

In this section, we present the proposed contrastive learning framework GeomGCL with leveraging dual geometric views of the molecular graph. As shown in Figure 2, after the derivation of 2D and 3D view graphs from the original molecule in SMILES format, GeomGCL equips with a dual-channel geometric message passing architecture (GeomMPNN) to learn the representations of both graphs adaptively. By contrast with the deterministic graph structure of 2D view, the 3D structure of molecule is always calculated through a stochastic process of 3D conformation generation. Consequently, while the 3D view graph contains more abundant geometric structure, such uncertain information is not always beneficial for molecular representation learning. To this end, we propose to bypass this challenge by adopting a geometric contrastive learning strategy across 2D and 3D views. This geometric-view supervision mutually makes both views complement each other. On the one hand, GeomMPNN can distill the valuable 3D structure under the guidance of 2D view. On the other hand, it helps to inject 3D geometric information for better 2D molecular representation learning. In the following sections, we use the bold letters ($\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{e}_{uv}$ or $\boldsymbol{a}_u$) to represent embeddings of the corresponding symbolic indicators ($l, r, \phi, \theta, e_{uv}$ or a_u).

4.1 Geometric Embedding

Since the primary 2D or 3D coordinates are changeable and inconsistent across different coordinate systems, we manage to calculate the definite geometric factors (i.e., angle and distance) and then utilize radial basis functions (RBF) to obtain dual-level geometric embeddings. As illustrated in Figure 1, the local distance l and 2D angle ϕ refer to the distance and angle based on the covalent bonds respectively, which carry the critical chemical information. In the 3D view graph, the global distance r further provides non-local correlations in a molecule, while the 3D angle θ indicates the spatial distribution of global connections. Following the previous work (Shui and Karypis 2020), we adopt several RBF layers to encode diverse geometric factors:

$$\boldsymbol{l} = RBF(l) = \underset{k=1}{\overset{K}{\frown}} \exp\big(-\beta_l(\exp(-l) - \mu_{l,k})^2\big) \quad (1)$$

$$\boldsymbol{r} = RBF(r) = \underset{k=1}{\overset{K}{\frown}} \exp\big(-\beta_r(\exp(-r) - \mu_{r,k})^2\big) \quad (2)$$

$$\boldsymbol{\phi} = RBF(\phi) = \underset{k=1}{\overset{K}{\frown}} \exp\big(-\beta_\phi(\phi - \mu_{\phi,k})^2\big) \quad (3)$$

$$\boldsymbol{\theta} = RBF(\theta) = \underset{k=1}{\overset{K}{\frown}} \exp\big(-\beta_\theta(\theta - \mu_{\theta,k})^2\big) \quad (4)$$

where $\frown$ is the concatenation operation over scalar values to form a K-dimensional geometric embedding. For local and global distances, the K central points $\{\mu_{*,k}\}$ are uniformly
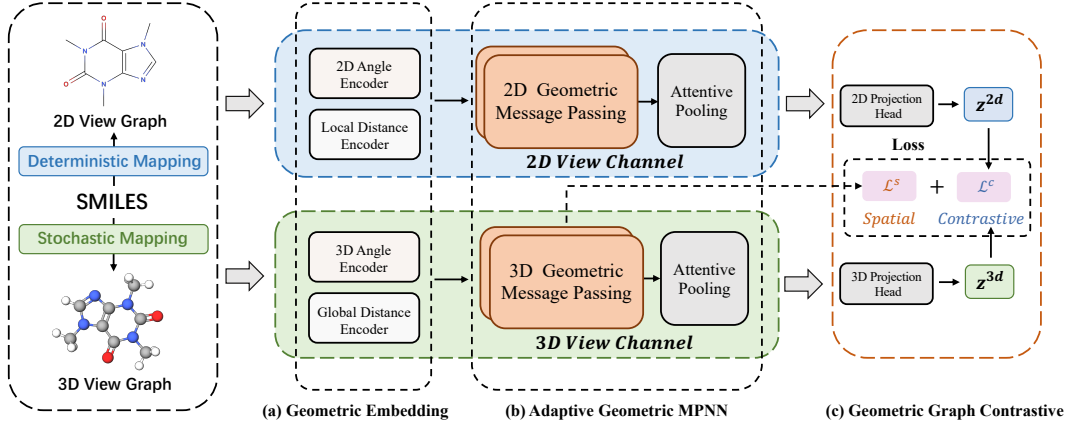
Figure 2: Illustration of geometric graph contrastive learning framework (GeomGCL) on molecular graphs.

selected between $\exp(-*)$ ($*$ is l or d) and 1, while $\beta_* = \left(\frac{2}{K}(1 - \exp(-*))\right)^{-2}$. For 2D and 3D angles, each $\mu_{*,k}$ is between 0 and π with $\beta_* = \left(\frac{2\pi}{K}\right)^{-2}$, where $*$ denotes ϕ or θ.

4.2 Adaptive Geometric Message Passing

As shown in Figure 3, we further design an adaptive message passing scheme (GeomMPNN) to learn the topological structures of molecules with geometric information in a Node-Edge interactive manner. On the whole, GeomMPNN consists of *Node→Edge*, *Edge→Edge* and *Edge→Node* three-stage message passing layers to iteratively update the node and edge embeddings, followed by a *Node→Graph* attentive pooling process. Both of the dual-channel networks generally follow such architecture and can adaptively learn the 2D and 3D geometric factors with fine-grained designs.

(i) Node→Edge Message Passing. Since only the existing bonds in a molecule have the initial edge features (i.e., bond features), the edge embedding should be firstly updated through aggregating the pairwise node embeddings with involving the associated features. To enrich the connection information from different aspects, we use MLP to integrate the chemical bond feature e_{uv}^0 and global distance embeddings r_{uv} for 2D- and 3D-edge embeddings respectively:

$$e_{uv}^{2d,t} = MLP(a_u^{2d,t-1} \parallel a_v^{2d,t-1} \parallel e_{uv}^0) \quad (5)$$

$$e_{uv}^{3d,t} = MLP(a_u^{3d,t-1} \parallel a_v^{3d,t-1} \parallel r_{uv}) \quad (6)$$

where $e_{uv}^{2d,t}$ and $e_{uv}^{3d,t}$ are edge embeddings at t-th layer, a_u is the node (atom) embedding, the superscript $2d$ or $3d$ indicates the view channel, and $\parallel$ represents the concatenation operator. Then the dual-channel edge embeddings can contain both geometric and chemical semantic information.

(ii) Edge→Edge Message Passing. Different from the general graph, both the 2D view and 3D view graphs of a molecule have the unique geometric attributes, which can significantly influence the specific property of the molecule. After the derivation of edge embeddings, GeomMPNN performs an edge→edge message passing process to perceive the geometric distribution in the molecule through 2D and 3D angle-aware aggregations. Considering that the neighbors in 2D view are more sparse by contrast with the neigh-

bors in 3D view, we develop the well-directed layers for both views. Specifically, for 2D view graph learning, the following angle-injected function is employed to update the 2D edge embedding:

$$e_{uv}^{2d,t} = \sum_{e_{wu} \in \mathcal{A}(e_{uv})} W_\phi^t \phi_{wuv} \odot (W_e^t e_{uv}^{2d,t}) \quad (7)$$

where $\mathcal{A}(e_{uv})$ denotes the set of neighboring edges of the edge e_{uv}, ϕ_{wuv} is the 2D angle embedding between the edge e_{wu} and e_{uv}, $\odot$ is the element-wise dot operation, W_ϕ^t and $W_e^{2d,t}$ are learnable parameters. For the 3D view graph, there are sufficient neighbors around each edge. Inspired by the recent work (Li et al. 2021), we divide the neighboring edges of each target edge into several angle domains $\mathcal{A}_1, ..., \mathcal{A}_n$ according to 3D spatial angle θ. Then we apply a hierarchical aggregation process among edges for 3D view graph learning, which consists of local and global stages:

$$e_{uv,i}^{3d,t} = \sum_{e_{wu} \in \mathcal{A}_i(e_{uv})} W_{\theta,i}^t \theta_{wuv} \odot (W_{e,i}^t e_{uv}^{3d,t}) \quad (8)$$

$$e_{uv}^{3d,t} = \mathop{\Big\|}_{i=1}^{n} \text{Pool}\left(\{e_{uv,i}^{3d,t} | 1 \le i \le n\}\right) \odot e_{uv,i}^{3d,t} \quad (9)$$

where $e_{uv,i}^{3d,t}$ is the aggregated 3D edge embedding at i-th angle domain through the local stage, $W_{e,i}$ and $W_{\theta,i}^t$ are trainable parameters, θ_{wuv} is the 3D angle embedding between the edge e_{wu} and e_{uv}. Pool means the max pooling function over all n local edge embeddings, which can generally extract the high-level spatial distribution information to strengthen the geometric structure learning.

(iii) Edge→Node Message Passing. After obtaining the angle-aware edge embeddings $e_{uv}^{2d,t}$ and $e_{uv}^{3d,t}$, we apply the edge→node message passing to fulfil the propagation process from the edge back to the node. The essential distance factor between nodes is well-considered via the similar adaptive scheme with the previous edge→edge component:

$$a_v^{2d,t} = \sum_{e_{uv} \in \mathcal{D}(a_v)} W_l^t l_{uv} \odot (W_a^t e_{uv}^{2d,t}) \quad (10)$$

$$a_v^{3d,t} = \mathop{\Big\|}_{i=1}^{m} \sum_{e_{uv} \in \mathcal{D}_i(a_v)} W_{r,i}^t r_{uv} \odot (W_{a,i}^t e_{uv}^{3d,t}) \quad (11)$$
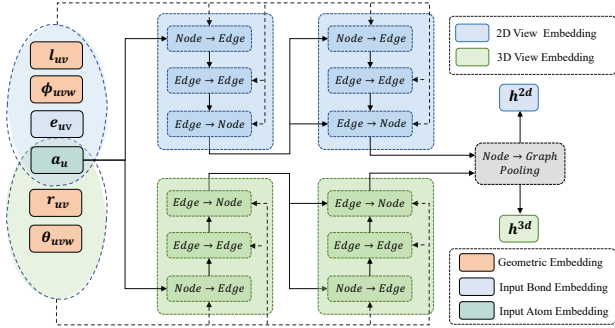
Figure 3: The architecture of GeomMPNN layers.

where W_l^t, W_a^t, $W_{r,i}^t$ and $W_{a,i}^t$ are learnable parameters, $\boldsymbol{l}_{uv}$ and $\boldsymbol{r}_{uv}$ are the geometric embeddings of local and global distances respectively, $\mathcal{D}(a_u)$ is the set of all neighboring edges for node a_u in 2D view, $\mathcal{D}_i(a_u)$ is the set of neighboring edges located in i-th distance domain among all m divided domains in 3D view.

(iv) Node→Graph Attentive Pooling. Since all geometric factors thoroughly enrich the node and edge embeddings in both 2D and 3D views via T stacked message passing layers, the final representations $\boldsymbol{a}_v^{2d,T}$ and $\boldsymbol{a}_v^{3d,T}$ can reflect the graph topology along with the molecular geometry. To further get the graph-level representation with identifying the important nodes, we follow (Xiong et al. 2019) and adopt the node→graph attentive pooling layer. For simplicity, we use $\boldsymbol{a}_v$ to represent $\boldsymbol{a}_v^{2d,T}$ or $\boldsymbol{a}_v^{3d,T}$ and use $\boldsymbol{h}$ to represent $\boldsymbol{h}^{2d}$ or $\boldsymbol{h}^{3d}$. The graph-level embedding $\boldsymbol{h}$ is updated iteratively through the attentive propagation process, which starts with the initial embedding $\boldsymbol{h}^0 = \sum_v \boldsymbol{a}_v$.

$$\alpha_v = LeakyRelu(\boldsymbol{q}^T \cdot [\boldsymbol{h}^t \parallel \boldsymbol{a}_v]) \tag{12}$$

$$\boldsymbol{g}^t = \sum_{a_v \in \mathcal{V}} \text{softmax}(\alpha_v) W_g^t \boldsymbol{a}_v \tag{13}$$

$$\boldsymbol{h}^{t+1} = \text{GRU}(\boldsymbol{h}^t, \boldsymbol{g}^t), \ t = 0, 1, .., T_g \tag{14}$$

where $\boldsymbol{q}$ is the learnable parameter, $\boldsymbol{g}^t$ is the global context message at t-th pooling layer, which aggregates the valuable nodes from the full set $\mathcal{V}$. After performing T_g pooling layers, the final graph representations $\boldsymbol{h}^{2d}$ for 2D view and $\boldsymbol{h}^{3d}$ for 3D view are acquired.

4.3 Geometric Contrastive Optimization

Despite the great progress made in the study of graph contrastive learning, the distinctive semantics and geometry of molecular graphs are always ignored. As a result, the conventional data augmentations may change the targeted molecular property as stated in (Sun et al. 2021), which emphasized the importance of domain knowledge to generate molecule variants. To tackle this challenge, we intend to enhance the molecular representation learning from the perspective of geometric contrast. As shown in Figure 2(c), the correlated 2D and 3D views can supervise each other without constructing additional fake samples. First, the 2D and 3D projection heads are adopted to map the representations

of two views into the space for contrastive learning:

$$\boldsymbol{z}_i^{2d} = MLP(\boldsymbol{h}_i^{2d}), \ \boldsymbol{z}_i^{3d} = MLP(\boldsymbol{h}_i^{3d}) \tag{15}$$

Our goal is to make the maximum consistency between 2D-3D positive pairs $\{\boldsymbol{z}_i^{2d}, \boldsymbol{z}_i^{3d}\}$ compared with negative pairs. Given one batch with N molecules, we have the following contrastive loss function under 2D-3D geometric views:

$$\mathcal{L}_i^c = \mathcal{L}_i^{2d,c} + \mathcal{L}_i^{3d,c}$$
$$= -\log \frac{e^{\langle \boldsymbol{z}_i^{2d}, \boldsymbol{z}_i^{3d} \rangle / \tau}}{\sum_{j=1}^N e^{\langle \boldsymbol{z}_i^{2d}, \boldsymbol{z}_j^{3d} \rangle / \tau}} - \log \frac{e^{\langle \boldsymbol{z}_i^{3d}, \boldsymbol{z}_i^{2d} \rangle / \tau}}{\sum_{j=1}^N e^{\langle \boldsymbol{z}_i^{3d}, \boldsymbol{z}_j^{2d} \rangle / \tau}} \tag{16}$$

where $\langle \cdot \rangle$ denotes the inner product to measure the similarity, τ is the scale parameter. Since $\boldsymbol{z}_i^{2d}$ and $\boldsymbol{z}_i^{3d}$ are two embeddings in different views from the same molecule, they are regarded as a positive pair while the remaining pairs in the batch are considered as negative pairs. Furthermore, to reflect the local spatial correlations across 3D geometric domains, we propose additional constraints as a spatial regularization technique. Since informative distance domains can mean different forces (such as covalent bond or van der Waals force), we focus on applying the regularization in angle domains. The key idea of this regularizer is to encourage the transformation matrices of adjacent angle domains to be similar to each other:

$$\mathcal{L}^s = \sum_{t=1}^T \sum_{i=1}^{n-1} \|W_{\theta,i+1}^t - W_{\theta,i}^t\|^2 \tag{17}$$

Finally, we combine the spatial and contrastive loss and arrive at the following objective function:

$$\mathcal{L}_{gcl} = \sum_{i=1}^M \mathcal{L}_i^c + \lambda \mathcal{L}^s \tag{18}$$

where M is the number of molecules in the dataset, λ is the trade-off parameter that controls the importance of spatial regularizer to better guide the representation learning.

4.4 Downstream Inference

When GeomGCL has been optimized through the geometric contrastive learning process, we utilize the well-trained 2D geometric MPNN $f_{2d}(\cdot)$ and 3D geometric MPNN $f_{3d}(\cdot)$ for downstream applications. The representations of two views are combined to predict the molecular properties via the finetune process. Formally, the prediction head can be written as follows:

$$\hat{y} = MLP\big(MLP\big(f_{2d}(\mathcal{G}^{2d})\big) + MLP\big(f_{3d}(\mathcal{G}^{3d})\big)\big) \tag{19}$$

For different tasks, we use the cross entropy loss function $\mathcal{L}_{ce}$ for classification loss $\mathcal{L}_{cls}$ and use the L1 loss function $\mathcal{L}_1$ for regression loss $\mathcal{L}_{reg}$. The spatial regularizer is also adopted for better performance.

$$\mathcal{L}_{cls} = \mathcal{L}_{ce}(\hat{y}, y) + \lambda_{cls}\mathcal{L}^s \tag{20}$$
$$\mathcal{L}_{reg} = \mathcal{L}_1(\hat{y}, y) + \lambda_{reg}\mathcal{L}^s \tag{21}$$

where λ_{cls} and λ_{reg} are trade-off parameters in the inference stage, $\hat{y}$ denotes the predicted value, and y is the measured true value of one specific molecular property.

Dataset	# Tasks	Task Type	# Molecules
ClinTox	2	Classification	1484
Sider	27	Classification	1427
Tox21	12	Classification	7831
ToxCast	617	Classification	8597
ESOL	1	Regression	1128
FreeSolv	1	Regression	643
Lipophilicity	1	Regression	4200

Table 1: Statistics of seven molecular datasets.

5 Experiments

In this section, we conduct experiments on seven well-known benchmark datasets to demonstrate the effectiveness of GeomGCL for molecular property prediction.

5.1 Experiment Settings

Datasets. To evaluate the performance of our proposed model with the existing molecular representation learning methods, we use seven molecular datasets from MoleculeNet (Wu et al. 2018) including ClinTox, Sider, Tox21 and ToxCast four physiology datasets for graph classification tasks, as well as ESOL, FreeSolv and Lipophilicity three physical chemistry datasets for graph regression tasks. The main statistics of datasets are summarized in Table 1.

Baselines. We compare the proposed GeomGCL with a variety of state-of-the-art baseline models for molecular property prediction, which includes molecular message passing-based methods, geometry learning-based GNN methods, and graph contrastive learning methods. The first group consists of three well-designed message passing neural networks. AttentiveFP (Xiong et al. 2019) adopts the graph attention network for molecular representation learning. DMPNN (Yang et al. 2019) and CoMPT (Chen et al. 2021) are message passing models with considering edge features in a node-edge interactive manner. Geometry learning-based models contain several GNN approaches. SGCN (Danel et al. 2020) directly encodes the atomic position information in the aggregation process. MAT (Maziarka et al. 2020) incorporates the local geometric distance into the graph-based transformer model. To comprehensively reflect the superiority of our proposed model, we also compare GeomGCL against with HMGNN (Shui and Karypis 2020) and DimeNet (Klicpera, Groß, and Günnemann 2020), both of which can learn the geometric distance and angle factors in 3D space for quantum property prediction. Besides, the recent graph contrastive models are compared to show the power of our proposed 2D-3D geometric contrastive learning strategy. InfoGraph (Sun et al. 2019) maximizes the mutual information between nodes and graphs, while MoCL (Sun et al. 2021) introduces the multi-level domain knowledge for molecular graphs in a well-designed contrastive learning framework.

Implementation Details. Following the previous works, we evaluate all methods through the k-fold cross-validation experiments, and we set k as 10 to report the robustly aver-

age experimental results. As recommended by the MoleculeNet benchmarks (Wu et al. 2018), we randomly split each dataset into training, validation, and testing set with a ratio of 0.8/0.1/0.1. The validation set is used for early stop and model selection. We use ROC-AUC and RMSE metrics for graph classification and graph regression tasks respectively.

The 3D structures of molecules are generated for $P = 50$ times through the stochastic optimization algorithm of Merck Molecular Force Field (MMFF), which is implemented in the RDKit package (Tosco, Stiefl, and Landrum 2014). For our model, We use Adam optimizer for model training with a learning rate of 1e-3. We set the batch size as 256 for contrastive learning and 32 for finetuning with the scale parameter $\tau = 0.5$. The hidden size of all models is set to 128. The cutoff distance d_θ is determined (4 Å or 5 Å) according to the size of the molecule on each dataset. We set the dimension K of geometric embedding as 64. The numbers of 3D angle domains and global distance domains are set to 4. The balancing hyper-parameter λ is set to 0.01 according to the performance on validation set. For baseline models, we tune parameters of each method based on recommended settings in the paper to ensure the best performance. As a general setting (Sun et al. 2019, 2021), our proposed GeomGCL and contrastive learning baselines are pretrained on molecular graphs of each dataset, and then we finetune the model for the downstream task on the same dataset.

5.2 Performance Evaluation

Overall Comparision. The performance results of evaluating each model for graph classification and regression tasks are presented in Table 2. As we can see, our model significantly outperforms all the baselines on both types of tasks. On the whole, we can observe that our proposed GeomGCL improves the performance over the best message passing baselines with 2.18% and 10.4% for classification and regression tasks, respectively.

Among all baseline approaches, the well-designed message passing models generally show the best performance, which indicates that the essential bond features of the molecule can provide chemical semantic information for molecular representation learning. Specifically, since DMPNN and CoMPT adopt the node-edge interactive scheme, they perform slightly better than AttentiveFP. As to geometry-based baseline models, MAT can take advantage of the geometric distance from the molecular graph and performs much better than SGCN, which directly encodes the 3D coordinates and can be easily affected by the coordinate systems. Although HMGNN and DimeNet can identify the distance and angle information, they learn the molecular embedding only based on the 3D geometric graph which might be noisy. Besides, these models are designed for quantum property prediction and may not be expert in modeling the larger molecules. By contrast, our GeomMPNN and GeomGCL are capable of learning from the stable 2D graph and the informative 3D geometric structure. For contrastive learning methods, MoCL achieves better results than InfoGraph, showing the significance of domain knowledge for developing the contrastive strategy without changing the chemical semantics. However, the failure of leveraging the

Model	Graph Classification (ROC-AUC) ↑					Graph Regression (RMSE) ↓			
	ClinTox	Sider	Tox21	ToxCast	Cls.Ave	ESOL	FreeSolv	Lipophilicity	Reg.Ave
AttentiveFP	0.808	0.605	0.835	0.743	0.748	0.578	1.034	0.602	0.738
DMPNN	0.886	0.637	0.848	0.743	0.779	0.647	1.092	0.591	0.777
CoMPT	0.877	0.626	0.836	0.755	0.774	0.589	1.103	0.590	0.761
SGCN	0.825	0.560	0.769	0.656	0.703	1.329	2.061	1.075	1.488
MAT	0.898	0.619	0.834	0.735	0.772	0.624	1.059	0.705	0.796
HMGNN	0.680	0.607	0.794	0.702	0.696	0.701	1.207	0.720	0.876
DimeNet	0.760	0.615	0.780	0.645	0.7000	0.633	0.978	0.614	0.742
InfoGraph	0.781	0.585	0.793	0.705	0.716	0.914	2.104	0.845	1.288
MoCL	0.739	0.629	0.824	0.718	0.727	0.934	1.478	0.742	1.051
GeomMPNN	0.900	0.638	0.838	0.743	0.780	**0.555**	0.913	0.578	0.682
GeomGCL	**0.919**	**0.648**	**0.850**	**0.763**	**0.796**	0.575	**0.866**	**0.541**	**0.661**

Table 2: Experimental results of our proposed GeomMPNN and GeomGCL along with all baselines on seven molecular graph datasets. Cls.Ave and Reg.Ave denote the average results of classification and regression tasks, respectively.
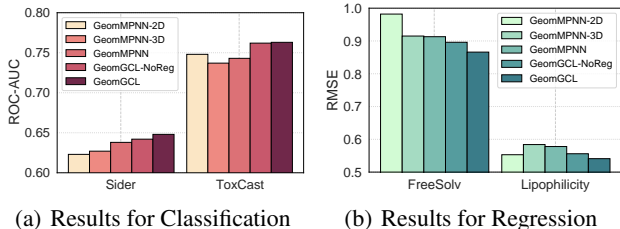

(a) Results for Classification   (b) Results for Regression

Figure 4: Evaluation of GeomGCL with its variants.

Figure 5: Analysis for the balancing parameter of $\mathcal{L}^s$.

critical geometric information limits their ability to model the molecule without labels, while our method can capture geometry-aware structural information by contrasting the 2D-3D geometric views. Therefore, GeomGCL is much effective for molecular representation learning and can accurately predict each targeted property.

Ablation Study. To further investigate the factors that influence the performance of the proposed GeomGCL framework, we conduct the ablation study on four benchmark datasets for classification and regression tasks with designing different variants of GeomGCL.

- **GeomMPNN-2D** only preserves the single-channel 2D geometric message passing layers.

- **GeomMPNN-3D** only preserves the single-channel 3D geometric message passing layers.

- **GeomMPNN** uses the dual-channel geometric message passing layers without contrastive learning.

- **GeomGCL-NoReg** removes the spatial regularizer $\mathcal{L}^s$.

As shown in Figure 4, GeomGCL achieves the best performance among all architectures, proving the necessity of learning the 2D-3D geometric structures contrastively and synergistically. To be specific, we can find that learning representations from a single 2D or 3D view can not always perform better than the other view across different datasets, which supports our hypothesis that only a geometric view of molecular graph is not sufficient. Therefore, decoupling 2D or 3D geometric message passing layers from GeomGCL yields a significant drop in performance. Additionally, the use of spatial regularizer when training the model can help GeomGCL to discriminate the relative correlations of differ-
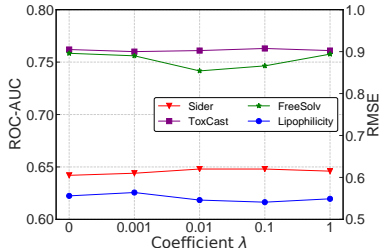
ent angle domains and then contributes to the performance improvements. What's more, GeomMPNN significantly performs worse than GeomGCL, which confirms that our geometric contrastive learning scheme is beneficial for molecular representation learning.

Parameter Analysis. Finally, we analyze the performance variation for GeomGCL by varying the coefficient λ to look deeper into the impact of the spatial regularizer. As depicted in Figure 5, we observe that the performance first tends to get better with incorporating more 3D angle domain information while training the model, and then begins to drop off slightly. The appropriate trade-off weight can assist the model in identifying the geometric factors and enhancing the representation learning. Overall, the performance of GeomGCL is stable and always better than baseline methods.

6 Conclusion

In this paper, we propose a novel geometric graph contrastive learning framework named GeomGCL for molecular representation learning, which builds the bridge between the geometric structure learning and the graph contrastive learning. Along this line, we design the dual-channel geometric message passing neural networks to sufficiently capture the distance and angle information under both 2D and 3D views. Then the appropriate geometry-based contrastive learning strategy is proposed to enhance the molecular representation learning with the spatial regularizer. The experimental results on the downstream property prediction tasks demonstrate the effectiveness of the proposed GeomGCL.

Acknowledgment

References

Atz, K.; Grisoni, F.; and Schneider, G. 2021. Geometric Deep Learning on Molecular Representations. *arXiv preprint arXiv:2107.12375*.

Bronstein, M. M.; Bruna, J.; LeCun, Y.; Szlam, A.; and Vandergheynst, P. 2017. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4): 18–42.

Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; and Blaschke, T. 2018. The rise of deep learning in drug discovery. *Drug discovery today*, 23(6): 1241–1250.

Chen, J.; Zheng, S.; Song, Y.; Rao, J.; and Yang, Y. 2021. Learning Attributed Graph Representation with Communicative Message Passing Transformer. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 2242–2248.

Danel, T.; Spurek, P.; Tabor, J.; Śmieja, M.; Struski, Ł.; Słowik, A.; and Maziarka, Ł. 2020. Spatial graph convolutional networks. In *International Conference on Neural Information Processing*, 668–675. Springer.

Duvenaud, D. K.; Maclaurin, D.; Iparraguirre, J.; Bombarell, R.; Hirzel, T.; Aspuru-Guzik, A.; and Adams, R. P. 2015. Convolutional Networks on Graphs for Learning Molecular Fingerprints. *Advances in Neural Information Processing Systems*, 28: 2224–2232.

Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; and Dahl, G. E. 2017. Neural message passing for quantum chemistry. In *International conference on machine learning*, 1263–1272. PMLR.

Goh, G. B.; Hodas, N. O.; and Vishnu, A. 2017. Deep learning for computational chemistry. *Journal of computational chemistry*, 38(16): 1291–1307.

Hawkins, P. C. 2017. Conformation generation: the state of the art. *Journal of Chemical Information and Modeling*, 57(8): 1747–1756.

Klicpera, J.; Groß, J.; and Günnemann, S. 2020. Directional Message Passing for Molecular Graphs. In *International Conference on Learning Representations*.

Li, S.; Zhou, J.; Xu, T.; Huang, L.; Wang, F.; Xiong, H.; Huang, W.; Dou, D.; and Xiong, H. 2021. Structure-aware Interactive Graph Neural Networks for the Prediction of Protein-Ligand Binding Affinity. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 975–985.

Li, S.; Zhou, J.; Xu, T.; Liu, H.; Lu, X.; and Xiong, H. 2020. Competitive analysis for points of interest. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1265–1274.

Maziarka, Ł.; Danel, T.; Mucha, S.; Rataj, K.; Tabor, J.; and Jastrzebski, S. 2020. Molecule Attention Transformer. *arXiv preprint arXiv:2002.08264*.

Qiu, J.; Chen, Q.; Dong, Y.; Zhang, J.; Yang, H.; Ding, M.; Wang, K.; and Tang, J. 2020. Gcc: Graph contrastive coding for graph neural network pre-training. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1150–1160.

Rogers, D.; and Hahn, M. 2010. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5): 742–754.

Shui, Z.; and Karypis, G. 2020. Heterogeneous molecular graph neural networks for predicting molecule properties. In *2020 IEEE International Conference on Data Mining*, 492–500. IEEE.

Song, Y.; Zheng, S.; Niu, Z.; Fu, Z.-H.; Lu, Y.; and Yang, Y. 2020. Communicative Representation Learning on Attributed Molecular Graphs. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, 2831–2838.

Sun, F.-Y.; Hoffman, J.; Verma, V.; and Tang, J. 2019. InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization. In *International Conference on Learning Representations*.

Sun, M.; Xing, J.; Wang, H.; Chen, B.; and Zhou, J. 2021. MoCL: Data-driven Molecular Fingerprint via Knowledge-aware Contrastive Learning from Molecular Graph. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 3585–3594.

Tosco, P.; Stiefl, N.; and Landrum, G. 2014. Bringing the MMFF force field to the RDKit: implementation and validation. *Journal of cheminformatics*, 6(1): 1–4.

Wang, X.; Liu, N.; Han, H.; and Shi, C. 2021. Self-Supervised Heterogeneous Graph Neural Network with Co-Contrastive Learning. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1726–1736.

Wu, L.; Lin, H.; Gao, Z.; Tan, C.; Li, S.; et al. 2021. Self-supervised on Graphs: Contrastive, Generative, or Predictive. *arXiv preprint arXiv:2105.07342*.

Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; and Philip, S. Y. 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1): 4–24.

Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; and Pande, V. 2018. MoleculeNet: a benchmark for molecular machine learning. *Chemical science*, 9(2): 513–530.

Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; et al. 2019. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of medicinal chemistry*, 63(16): 8749–8760.

Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.;

et al. 2019. Analyzing learned molecular representations for property prediction. *Journal of chemical information and modeling*, 59(8): 3370–3388.

Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; and Liu, T.-Y. 2021. Do Transformers Really Perform Badly for Graph Representation? In *Thirty-Fifth Conference on Neural Information Processing Systems*.

You, Y.; Chen, T.; Sui, Y.; Chen, T.; Wang, Z.; and Shen, Y. 2020. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 33: 5812–5823.

Zhou, J.; Li, S.; Huang, L.; Xiong, H.; Wang, F.; Xu, T.; Xiong, H.; and Dou, D. 2020. Distance-aware Molecule Graph Attention Network for Drug-Target Binding Affinity Prediction. *arXiv preprint arXiv:2012.09624*.