FRAGMENT-BASED MULTI-VIEW MOLECULAR CONTRASTIVE LEARNING

Seojin Kim[1*], **Jaehyun Nam**[1*], **Junsu Kim**[1], **Hankook Lee**[2], **Sungsoo Ahn**[3], **Jinwoo Shin**[1]
[1]Korea Advanced Institute of Science and Technology (KAIST) [2]LG AI Research
[3]Pohang University of Science and Technology (POSTECH)
{osikjs,jaehyun.nam,junsu.kim,jinwoos}@kaist.ac.kr
hankook.lee@lgresearch.ai, sungsoo.ahn@postech.ac.kr

ABSTRACT

Molecular representation learning is a fundamental task for AI-based drug design and discovery. Self-supervised contrastive learning on molecular graphs, which aims to learn good representations via semantic-preserving transformations, is an attractive framework for this task. However, it is relatively under-explored to design such transformations for molecules under consideration of their chemical semantics. In this paper, we consider *fragmentation* which decomposes a molecule into a set of chemically meaningful fragments (e.g., functional groups) as the semantic-preserving transformation. Here, we also utilize the 3D geometric views of molecules as another source of such transformation. Based on these molecule-specialized semantic-preserving transformations, we propose ***Frag**ment-based multi-view molecular **C**ontrastive **L**earning (**FragCL**)*, an effective framework that learns chemically meaningful molecular representations. Extensive experiments demonstrate that our framework outperforms prior molecular representation learning methods across various molecular property prediction tasks.

1 INTRODUCTION

Obtaining discriminative representations of molecules is a long-standing research problem in chemistry (Morgan, 1965). Such a task is critical for many applications, such as drug discovery (Capecchi et al., 2020) and material design (Gómez-Bombarelli et al., 2018), since it is a fundamental building block for various downstream tasks, e.g., molecular property prediction (Duvenaud et al., 2015) and molecule generation (Mahmood et al., 2021). Over the past decades, researchers have focused on handcrafting the molecular representation which encodes the presence of chemically informative substructures, e.g., functional groups, in a molecule (Rogers & Hahn, 2010; Capecchi et al., 2020).

Recently, graph neural networks (GNNs, Kipf & Welling, 2017) have gained much attention as a framework to learn the molecular graph representation due to its remarkable performance in learning to predict chemical properties (Wu et al., 2018). However, they often suffer from overfitting when the number of labeled training samples is insufficient (Rong et al., 2020b). To resolve this, researchers have investigated self-supervised learning that generates supervisory signals without labels to utilize a huge amount of unlabeled molecules (Rong et al., 2020a; Zhou et al., 2022).

A notable approach in this line of work is contrastive learning, which learns a discriminative representation by maximizing the agreement of representations of "similar" positive views while minimizing the agreement of "dissimilar" negative views (Chen et al., 2020). It has widely demonstrated its effectiveness for representation learning not only for molecules (You et al., 2020; Wang et al., 2021; 2022), but also for other domains, e.g., image (Chen et al., 2020; He et al., 2020), video (Pan et al., 2021), language (Wu et al., 2020), and speech (Chung et al., 2021). Here, the common challenge for learning good representations is how to construct effective positive and negative views.

Contribution. Our key idea is to utilize *fragmentation* that decomposes a molecule into a set of chemically meaningful fragments (i.e., substructures) such as functional groups. In particular, we use Breaking of Retrosynthetically Interesting Chemical Substructures (BRICS, Degen et al., 2008)

[*]These authors contributed equally.

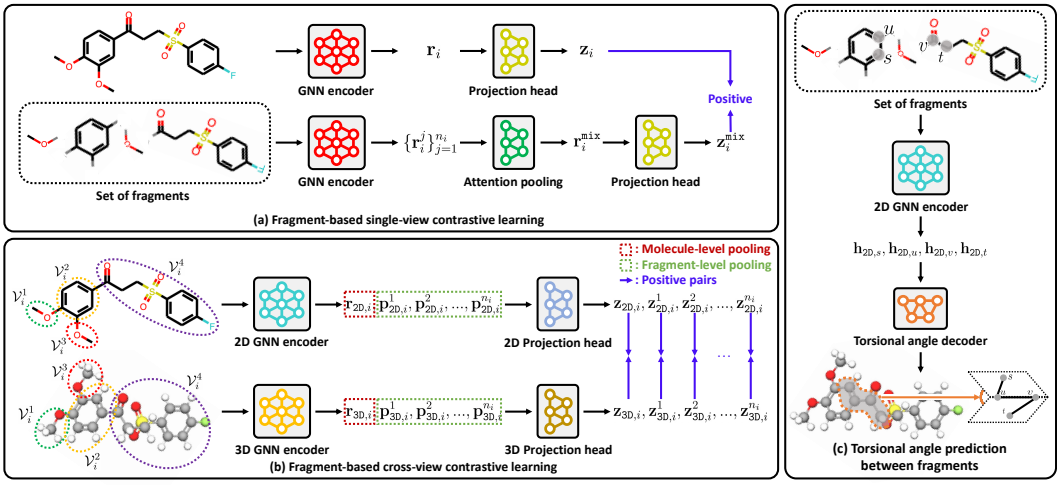


Figure 1: An overview of *Fragment-based multi-view molecular Contrastive Learning (FragCL)*. (a) A set of fragments is regarded as a positive view of a molecule. (b) Each molecule and fragment of 2D and 3D molecular graphs are considered as a positive pair. (c) 3D contextual information can be learned by predicting the torsional angle.

decomposition as a semantic-preserving transformation. We also utilize the 3D geometry (i.e., 3D atom positions) as another semantic-preserving view of 2D molecule and fragment graphs.[1] This is beneficial to learn better molecular representations because such 3D information is useful to predict various chemical properties of molecules such as polarizability (Anslyn & Dougherty, 2006). Furthermore, to exploit explicit 3D geometric information (e.g., energy surface), we suggest solving the torsional angle prediction task between adjacent fragments.

2 PRELIMINARIES

Problem Setup: Multi-view MRL. To consider a wide range of downstream tasks, we focus on learning a graph neural network (GNN) for 2D molecular graphs $f_{\texttt{2D}} : \mathcal{M}_{\texttt{2D}} \to \mathsf{R}^d$ where $\mathcal{M}_{\texttt{2D}}$ is the 2D molecular graph space. To be specific, we (i) pretrain a 2D molecule GNN $f_{\texttt{2D}}$ using an unlabeled set of molecules $\mathcal{D}_{\texttt{u}} \subseteq \mathcal{M}$ containing both 2D and 3D information, and then (ii) fine-tune $f_{\texttt{2D}}$ on various downstream tasks without 3D information, i.e., each task has a dataset $\mathcal{D} \subseteq \mathcal{M}_{\texttt{2D}} \times \mathcal{Y}$ where $\mathcal{Y}$ is the label space. Therefore, it is important to inject not only 2D topological information, but also 3D geometric information into the 2D molecule GNN $f_{\texttt{2D}}$ during pretraining. We remark that this *multi-view* pretraining setup has been recently investigated (Stärk et al., 2022; Liu et al., 2022).

Contrastive Learning. Generally speaking, contrastive learning aims to learn discriminative representations by attracting positive views while repelling negative views on the representation space, e.g., see Chen et al. (2020). A common practice for generating positive views is to utilize *semantic-preserving transformations*. Let $(\mathbf{x}, \mathbf{x}^+)$ be a positive pair generated by the transformations and $(\mathbf{x}, \mathbf{x}^-)$ be a negative pair obtained from different instances in a mini-batch. If $\mathbf{z}$, $\mathbf{z}^+$, and $\mathbf{z}^-$ are the representations of $\mathbf{x}, \mathbf{x}^+$, and $\mathbf{x}^-$, respectively, then the contrastive learning objective $\mathcal{L}_{\texttt{CL}}$ can be written as follows (Chen et al., 2020; You et al., 2020):

$$\mathcal{L}_{\texttt{CL}}(\mathbf{z}, \mathbf{z}^+, \{\mathbf{z}^-\}) = -\log \frac{\exp(\text{sim}(\mathbf{z}, \mathbf{z}^+)/\tau)}{\sum_{\mathbf{z}^-} \exp(\text{sim}(\mathbf{z}, \mathbf{z}^-)/\tau)}, \tag{1}$$

where $\text{sim}(\mathbf{z}, \widetilde{\mathbf{z}}) = \mathbf{z}^\top \widetilde{\mathbf{z}} / \|\mathbf{z}\|_2 \|\widetilde{\mathbf{z}}\|_2$ and τ is a temperature-scaling hyperparameter. The set $\{\mathbf{z}^-\}$ may include the positive $\mathbf{z}^+$ depending on the choice of objectives, e.g., NT-Xent (Chen et al., 2020).

[1] A molecule can be represented by (a) a 2D topological graph (V, E) of nodes V and edges E or (b) a 3D geometric graph (V, R) of nodes V and 3D coordinates R.

3 FRAGCL: FRAGMENT-BASED MULTI-VIEW MOLECULAR CONTRASTIVE LEARNING

Our framework crucially relies on the molecular fragmentation which decomposes a molecule into a set of chemically meaningful fragments (i.e., substructures). In this paper, we mainly use BRICS decomposition (Degen et al., 2008), which is designed to preserve most chemically informative substructures (Liu et al., 2017).

3.1 FRAGMENT-BASED SINGLE-VIEW CONTRASTIVE LEARNING

We first introduce our contrastive learning objective based on molecular fragmentation. Specifically, given a training batch $\{M_i\}_{i=1}^n$, we consider $(M_i, \{M_i^j\}_{j=1}^{n_i})$ as a positive pair (i.e., they share the same chemical semantics) where n_i is the number of fragments of the molecule M_i. To aggregate representations of the set of fragments $\{M_i^j\}_{j=1}^{n_i}$, we use the attention pooling mechanism (Li et al., 2016). Formally, the representation for the set $\{M_i^j\}_{j=1}^{n_i}$ is obtained as follows:

$$\mathbf{r}_i^{\mathtt{mix}} := \sum_{j=1}^{n_i} \frac{\exp(\mathbf{a}^\top \mathbf{r}_i^j + b)}{\sum_{k=1}^{n_i} \exp(\mathbf{a}^\top \mathbf{r}_i^k + b)} \cdot \mathbf{r}_i^j,$$

where $\mathbf{r}_i^j := f(M_i^j)$ is the representation for each fragment obtained by a molecule GNN f, $\mathbf{a} \in \mathsf{R}^d$ and $b \in \mathsf{R}$ are learnable parameters. Similarly, we compute the molecular representation $\mathbf{r}_i = f(M_i)$ for the whole structure. Then, we separately optimize the 2D-GNN $f_{\mathtt{2D}}$ and the 3D-GNN $f_{\mathtt{3D}}$ along with projection heads $g_{\mathtt{2D}} : \mathsf{R}^d \to \mathsf{R}^d$ and $g_{\mathtt{3D}} : \mathsf{R}^d \to \mathsf{R}^d$ by the following contrastive objective with the fragment-based positive pairs:

$$\mathcal{L}_{\mathtt{single}} := \frac{1}{n} \sum_{i=1}^{n} \left(\mathcal{L}_{\mathtt{CL}}(\mathbf{z}_{\mathtt{2D};i}, \mathbf{z}_{\mathtt{2D};i}^{\mathtt{mix}}, \{\mathbf{z}_{\mathtt{2D};j}^{\mathtt{mix}}\}_{j \neq i}) + \mathcal{L}_{\mathtt{CL}}(\mathbf{z}_{\mathtt{3D};i}, \mathbf{z}_{\mathtt{3D};i}^{\mathtt{mix}}, \{\mathbf{z}_{\mathtt{3D};j}^{\mathtt{mix}}\}_{j \neq i}) \right), \tag{2}$$

where $\mathbf{z}_i$ and $\mathbf{z}_i^{\mathtt{mix}}$ denote latent representations projected by g from $\mathbf{r}_i$ and $\mathbf{r}_i^{\mathtt{mix}}$, respectively.

3.2 FRAGMENT-BASED CROSS-VIEW CONTRASTIVE LEARNING

We here consider $(M_{\mathtt{2D}}, M_{\mathtt{3D}})$ as a positive pair. Then, the molecule-level contrastive objective can be written as follows:

$$\mathcal{L}_{\mathtt{cross;mol}} := \frac{1}{2n} \sum_{i=1}^{n} \left(\mathcal{L}_{\mathtt{CL}}(\mathbf{z}_{\mathtt{2D};i}, \mathbf{z}_{\mathtt{3D};i}, \{\mathbf{z}_{\mathtt{3D};j}\}_{j=1}^n) + \mathcal{L}_{\mathtt{CL}}(\mathbf{z}_{\mathtt{3D};i}, \mathbf{z}_{\mathtt{2D};i}, \{\mathbf{z}_{\mathtt{2D};j}\}_{j=1}^n) \right). \tag{3}$$

This objective is widely investigated in molecular representation learning (Stärk et al., 2022; Liu et al., 2022). However, modeling the cross-view contrastive objective based solely on the similarity of molecule-level representations may lack capturing fragment-level information (i.e., chemical property at a finer level). Therefore, we suggest *fragment-level cross-view contrastive learning* in what follows.

We consider $(M_{\mathtt{2D}}^j, M_{\mathtt{3D}}^j)$ as a fragment-level positive pair where $\{M^j\}$ is the set of fragments of a molecule M. To be specific, we compute the j-th fragment representation $\mathbf{p}_i^j$ of a molecule M_i via fragment-wise pooling by $\mathbf{p}_i^j := \frac{1}{|V_i^j|} \sum_{v \in V_i^j} \mathbf{h}_{v;i}$, where $\{\mathbf{h}_{v;i}\}_{v \in V}$ are the last-layer node representations of the whole molecular structure M_i. We then compute latent fragment representations by a projector g, e.g., $\mathbf{z}_{\mathtt{2D};i}^j = g_{\mathtt{2D}}(\mathbf{p}_{\mathtt{2D};i}^j)$. Using these representations, we compute the average of fragment-wise similarities $s_{i;j}$ between molecules M_i and M_j:

$$s_{i;i} := \frac{1}{n_i} \sum_{k=1}^{n_i} \mathrm{sim}(\mathbf{z}_{\mathtt{2D};i}^k, \mathbf{z}_{\mathtt{3D};i}^k), \quad s_{i;j}^{\mathtt{2D\,(or\,3D)}} := \frac{1}{n_i} \sum_{k=1}^{n_i} \max_{1 \leq l \leq n_j} \mathrm{sim}(\mathbf{z}_{\mathtt{2D\,(or\,3D)};i}^k, \mathbf{z}_{\mathtt{3D\,(or\,2D)};j}^l),$$

where n_i is the number of fragments of the molecule M_i. By introducing cross-view objective with 3D information, our framework can effectively discriminate a pair of different molecules whose fragments are the same, e.g., o-xylene and p-xylene have one phenyl and two methyl fragments.

Table 1: Test ROC-AUC score on the MoleculeNet downstream molecular property classification benchmarks. We report mean and standard deviation over 3 different seeds. We mark the best mean score and scores within one standard deviation of the best mean score to be bold. We denote the scores obtained from Liu et al. (2022) with (*). Otherwise, we reproduce scores under the same setup. Scores obtained through fine-tuning of the officially provided checkpoints are denoted by (†).[2]

Methods	BBBP	Tox21	ToxCast	Sider	Clintox	MUV	HIV	Bace	Avg.
-	65.4±2.4	**74.9**±0.8	61.6±1.2	58.0±2.4	58.8±5.5	71.0±2.5	**75.3**±0.5	72.6±4.9	67.2
Pretrained with 50k 2D molecular graphs of GEOM and fine-tuned on 2D molecular graphs of MoleculeNet									
EdgePred* Hamilton et al. (2017)	64.5±3.1	74.5±0.4	60.8±0.5	56.7±0.1	55.8±6.2	73.3±1.6	**75.1**±0.8	64.6±4.7	65.6
AttrMask* Hu et al. (2020a)	70.2±0.5	74.2±0.8	62.5±0.4	60.4±0.6	68.6±9.6	73.9±1.3	74.3±1.3	77.2±1.4	70.2
GPT-GNN* Hu et al. (2020b)	64.5±1.1	**75.3**±0.5	62.2±0.1	57.5±4.2	57.8±3.1	**76.1**±2.3	**75.1**±0.2	77.6±0.5	68.3
Infomax* Sun et al. (2019)	69.2±0.8	73.0±0.7	62.0±0.3	59.2±0.2	75.1±5.0	74.0±1.5	74.5±1.8	73.9±2.5	70.1
ContextPred* Hu et al. (2020a)	**71.2**±0.9	73.3±0.5	62.8±0.3	59.3±1.4	73.7±4.0	72.5±2.2	**75.8**±1.1	78.6±1.4	70.9
GraphLoG* Xu et al. (2021)	67.8±1.7	73.0±0.3	62.2±0.4	57.4±2.3	62.0±1.8	73.1±1.7	73.4±0.6	78.8±0.7	68.5
G-Contextual* Rong et al. (2020a)	70.3±1.6	**75.2**±0.3	62.6±0.3	58.4±0.6	59.9±8.2	72.3±0.9	**75.9**±0.9	79.2±0.3	69.2
G-Motif* Rong et al. (2020a)	66.4±3.4	73.2±0.8	62.6±0.5	60.6±1.1	77.8±2.0	73.3±2.0	73.8±1.4	73.4±4.0	70.1
GraphCL* You et al. (2020)	67.5±3.3	**75.0**±0.3	62.8±0.2	60.1±1.3	78.9±4.2	**77.1**±1.0	**75.0**±0.4	68.7±7.8	70.1
JOAO* You et al. (2021)	66.0±0.6	74.4±0.7	62.7±0.6	60.7±1.0	66.3±3.9	**77.0**±2.2	**76.6**±0.5	72.9±2.0	70.6
JOAOv2 You et al. (2021)	67.2±3.6	**75.0**±0.7	63.5±0.3	60.6±0.4	77.1±3.9	73.4±3.4	**77.7**±1.1	71.7±0.5	69.6
MGSSL Zhang et al. (2021)	67.3±0.9	74.5±0.2	63.6±0.4	58.4±0.2	75.4±3.8	73.9±1.4	**77.2**±2.5	76.2±1.3	70.8
MolCLR Wang et al. (2021)	67.6±0.6	74.4±1.3	62.9±0.2	58.7±1.1	57.9±3.0	70.8±2.8	**75.4**±1.2	74.6±3.5	67.8
D-SLA Kim et al. (2022)	69.6±2.4	73.7±0.7	63.3±0.2	59.2±2.0	60.5±1.0	75.3±0.6	**75.8**±0.9	**81.2**±2.5	69.8
Pretrained with 50k 2D and 3D molecular graphs of GEOM and fine-tuned on 2D molecular graphs of MoleculeNet									
3D-InfoMax Stärk et al. (2022)	67.9±1.2	**75.3**±0.3	**64.6**±0.4	59.6±0.7	89.7±0.5	**76.7**±0.6	73.4±1.2	79.9±0.9	73.4
GraphMVP† Liu et al. (2022)	69.6±0.2	**75.6**±0.7	63.7±0.3	61.3±0.6	89.0±1.4	75.7±1.0	**75.1**±0.3	**80.9**±1.3	73.9
GraphMVP-G† Liu et al. (2022)	70.1±0.7	**75.3**±0.9	64.2±0.9	61.0±0.5	89.4±1.5	**77.7**±1.6	**75.3**±0.8	80.2±1.5	74.1
GraphMVP-C† Liu et al. (2022)	69.6±1.4	74.6±0.1	64.1±0.2	**63.0**±0.1	88.7±2.6	73.9±1.7	**74.7**±2.0	**81.3**±0.7	73.7
FragCL (Ours)	**71.4**±0.4	**75.2**±0.7	**65.1**±0.8	61.0±0.6	**95.2**±1.0	**77.6**±1.0	**76.3**±0.4	**82.3**±1.6	**75.5**

Finally, we formulate our fragment-level cross-view contrastive objective as follows:

$$\mathcal{L}_{\texttt{cross;frag}} := -\frac{1}{2n}\sum_{i=1}^{n}\left(\log\frac{e^{s_{i;i}=}}{e^{s_{i;i}=}+\sum_{j\neq i}e^{s^{2D}_{i;j}=}}+\log\frac{e^{s_{i;i}=}}{e^{s_{i;i}=}+\sum_{j\neq i}e^{s^{3D}_{i;j}=}}\right). \tag{4}$$

To sum up, our cross-view objective is as follows:

$$\mathcal{L}_{\texttt{cross}} := \frac{1}{2}\left(\mathcal{L}_{\texttt{cross;mol}}+\mathcal{L}_{\texttt{cross;frag}}\right).$$

3.3 Torsional Angle Prediction between Fragments

We define the torsional angle prediction task for each fragmented bond: for a 2D molecule $M_{2D;i}$ and a fragmented bond $(u, v) \in E_{2D;i}$, we randomly select non-hydrogen atoms s and t adjacent to u and v, respectively, and compute the torsional angle y of the quartet (s, u, v, t) on the molecule M_i. If $\mathcal{T}$ is a collection of the tasks for all fragments, our loss function can be written as follows:

$$\mathcal{L}_{\texttt{tor}} := \frac{1}{|\mathcal{T}|}\sum_{(i;s;u;v;t;y)\in\mathcal{T}}\mathcal{L}_{\texttt{CE}}(\hat{y}_i(s,u,v,t),y), \tag{5}$$

where $\mathcal{L}_{\texttt{CE}}$ is the cross-entropy loss, y is the binned label for the angle, and $\hat{y}_i(s,u,v,t) := g_{\texttt{tor}}([\mathbf{h}_{2D;a;i}]_{a\in\{s;u;v;t\}})$ is the prediction from the concatenation of node representations of atoms (s, u, v, t) of the molecule $M_{2D;i}$ using a multi-layer perceptron (MLP) $g_{\texttt{tor}}(\cdot)$.

3.4 Overall Training Objective

From the discussion in,Section 3.1, 3.2, and 3.3, we propose our training loss function by:

$$\mathcal{L}_{\texttt{FragCL}} := \mathcal{L}_{\texttt{single}}+\mathcal{L}_{\texttt{cross}}+\mathcal{L}_{\texttt{tor}}. \tag{6}$$

Note that τ is the only hyperparameter that is newly proposed by our framework. We set $\tau = 0.1$ in Eq. (1) and (4) following You et al. (2020).

[2]GraphMVP (Liu et al., 2022) pretrains with explicit hydrogens, but fine-tunes without explicit hydrogens. We report fine-tuning results with explicit hydrogens from official checkpoints. Thus, our reported average value is slightly higher than the original paper.

Table 2: Test MAE score on the QM9 downstream quantum property regression benchmarks. For ours and all baselines, we employ GIN (Xu et al., 2019) as the 2D-GNN architecture and pretrain with entire 310k molecules from the GEOM dataset (Axelrod & Gomez-Bombarelli, 2022). We mark the best score bold.

Methods	ZPVE ↓	μ ↓	α ↓	C_v ↓	LUMO ↓	HOMO ↓	ε_{gap} ↓	R^2 ↓	U_0 ↓	U_{298} ↓	H_{298} ↓	G_{298} ↓
-	43.7	0.059	0.400	0.144	80.5	89.4	171.0	3.27	62.9	61.8	57.0	48.1
Pretrained on 310k 2D and 3D molecular graphs of GEOM and fine-tuned on 2D molecular graphs of QM9												
3D-Infomax Stärk et al. (2022)	27.0	0.051	0.355	0.126	63.4	55.2	103.8	2.99	**38.8**	45.6	41.0	40.8
GraphMVP-G Liu et al. (2022)	24.1	0.051	0.367	0.123	59.1	53.8	100.4	2.97	39.9	44.2	41.0	40.3
FragCL (Ours)	**24.0**	**0.049**	**0.353**	**0.121**	**57.1**	**51.8**	**97.1**	**2.90**	39.2	**42.9**	**40.3**	**40.0**

Table 3: Test MAE score of semi-supervised learning on the QM9 downstream quantum property regression benchmarks. We employ GIN (Xu et al., 2019) as the 2D-GNN architecture and pretrain with 110k QM9 training dataset. Then we fine-tune across different label fraction of QM9 training dataset. We mark the best score bold.

Methods	ZPVE ↓			LUMO ↓			HOMO ↓			U_0 ↓		
Label Fraction (%)	20	50	100	20	50	100	20	50	100	20	50	100
-	111.0	87.1	43.7	236.0	140.6	80.5	233.6	128.1	89.4	165.5	82.8	62.9
Pretrained on 110k 2D and 3D molecular graphs of QM9 and fine-tuned on 2D molecular graphs of QM9												
3D-Infomax Stärk et al. (2022)	87.2	42.8	24.4	215.0	98.4	57.9	181.0	102.4	57.7	148.2	75.0	42.1
GraphMVP-G Liu et al. (2022)	85.4	42.8	24.4	214.3	99.7	59.7	177.3	100.0	56.9	145.7	74.5	42.2
FragCL (Ours)	**83.7**	**39.4**	**22.2**	**202.2**	**97.8**	**54.6**	**172.9**	**91.0**	**48.4**	**138.7**	**71.8**	**38.0**

4 EXPERIMENTS

In our experiments, FragCL achieves the best performance in downstream molecular property prediction tasks. The results on the molecule retrieval results can be found in Appendix G, and the ablation study can be found in Appendix H.

Experimental setup. For pretraining, we consider the GEOM (Axelrod & Gomez-Bombarelli, 2022) and the QM9 (Ramakrishnan et al., 2014) datasets, which consist of 2D and 3D paired molecular graphs. We consider (a) transfer learning on the binary classification tasks from MoleculeNet benchmark (Wu et al., 2018), and (b) transfer learning and semi-supervised learning on the regression tasks using QM9 (Ramakrishnan et al., 2014). Details can be found in Appendix B and E.

MoleculeNet classification task. As reported in Table 1, FragCL achieves the best average test ROC-AUC score when transferred to MoleculeNet (Wu et al., 2018) downstream tasks after pretrained with 50k molecules from the GEOM (Axelrod & Gomez-Bombarelli, 2022) dataset. To be specific, FragCL improves the best average ROC-AUC score baseline, GraphMVP-G (Liu et al., 2022), by 74.1 $\rightarrow$ 75.5, achieving the state-of-the-art performance on 7 out of 8 downstream tasks. We emphasize that the improvement of FragCL is consistent over downstream tasks. For example, GraphMVP-C (Liu et al., 2022) achieves the best performance on Sider, while it fails to generalize on Tox21, resulting in even lower ROC-AUC score compared to the model with no pretraining. On the other hand, FragCL shows the best average performance with no such failure case, i.e., FragCL learns well-generalizable representations over a wide range of downstream tasks.

QM9 regression task. Table 2 and 3 show the overall results of transfer learning and semi-supervised learning on the QM9 (Ramakrishnan et al., 2014) regression benchmarks, respectively. For transfer learning (Table 2), we pretrain with 310k molecules from the GEOM (Axelrod & Gomez-Bombarelli, 2022) dataset. FragCL outperforms the baselines, achieving the best performances on 11 out of 12 downstream tasks. We emphasize that FragCL outperforms the baselines when transferred to both MoleculeNet and QM9 downstream tasks. For semi-supervised learning (Table 3), FragCL achieves the best performances over all tasks and label fractions. In particular, FragCL shows superior performance even in the fully supervised learning scenario (i.e., 100% label fraction), e.g., 89.4 $\rightarrow$ 48.4 for HOMO. This implies that FragCL indeed finds "good initialization" of GNN and show its wide applicability. More results for semi-supervised learning can be found in Appendix F.

REFERENCES

Amir Hosein Khas Ahmadi. *Memory-based graph networks*. PhD thesis, University of Toronto (Canada), 2020.

Eric V Anslyn and Dennis A Dougherty. *Modern physical organic chemistry*. University science books, 2006.

Simon Axelrod and Rafael Gomez-Bombarelli. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):1–14, 2022.

Alice Capecchi, Daniel Probst, and Jean-Louis Reymond. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. *Journal of cheminformatics*, 12(1):1–15, 2020.

Lei Chen, Wei-Ming Zeng, Yu-Dong Cai, Kai-Yan Feng, and Kuo-Chen Chou. Predicting anatomical therapeutic chemical (atc) classification of drugs by integrating chemical-chemical interactions and similarities. *PloS one*, 7(4):e35254, 2012.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

Yu-An Chung, Yu Zhang, Wei Han, Chung-Cheng Chiu, James Qin, Ruoming Pang, and Yonghui Wu. W2v-bert: Combining contrastive learning and masked language modeling for self-supervised speech pre-training. In *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 244–250. IEEE, 2021.

Jörg Degen, Christof Wegscheid-Gerlach, Andrea Zaliani, and Matthias Rarey. On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem: Chemistry Enabling Drug Discovery*, 3(10):1503–1507, 2008.

David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. *Advances in neural information processing systems*, 28, 2015.

Xiaomin Fang, Lihang Liu, Jieqiong Lei, Donglong He, Shanzhuo Zhang, Jingbo Zhou, Fan Wang, Hua Wu, and Haifeng Wang. Geometry-enhanced molecular representation learning for property prediction. *Nature Machine Intelligence*, 4(2):127–134, 2022.

Yin Fang, Haihong Yang, Xiang Zhuang, Xin Shao, Xiaohui Fan, and Huajun Chen. Knowledge-aware contrastive molecular graph learning. *arXiv preprint arXiv:2103.13047*, 2021.

Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

William L Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*, 2017.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9729–9738, 2020.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *International Conference on Learning Representations*, 2020a.

Ziniu Hu, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, and Yizhou Sun. Gpt-gnn: Generative pre-training of graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1857–1867, 2020b.

Rui Jiao, Jiaqi Han, Wenbing Huang, Yu Rong, and Yang Liu. Energy-motivated equivariant pretraining for 3d molecular graphs. *arXiv preprint arXiv:2207.08824*, 2022.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. In *International conference on machine learning*, pp. 2323–2332. PMLR, 2018.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Hierarchical generation of molecular graphs using structural motifs. In *International conference on machine learning*, pp. 4839–4848. PMLR, 2020.

Dongki Kim, Jinheon Baek, and Sung Ju Hwang. Graph self-supervised learning with accurate discrepancy learning. *arXiv preprint arXiv:2202.02989*, 2022.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. In *International Conference on Learning Representations*, 2016.

Shengchao Liu, Hanchen Wang, Weiyang Liu, Joan Lasenby, Hongyu Guo, and Jian Tang. Pre-training molecular graph representation with 3d geometry. In *International Conference on Learning Representations*, 2022.

Tairan Liu, Misagh Naderi, Chris Alvin, Supratik Mukhopadhyay, and Michal Brylinski. Break down in order to build up: decomposing small molecules for fragment-based drug design with e molfrag. *Journal of chemical information and modeling*, 57(4):627–631, 2017.

Yi Liu, Limei Wang, Meng Liu, Xuan Zhang, Bora Oztekin, and Shuiwang Ji. Spherical message passing for 3d graph networks. *arXiv preprint arXiv:2102.05013*, 2021.

Shengjie Luo, Tianlang Chen, Yixian Xu, Shuxin Zheng, Tie-Yan Liu, Liwei Wang, and Di He. One transformer can understand both 2d & 3d molecular data. *arXiv preprint arXiv:2210.01765*, 2022.

Omar Mahmood, Elman Mansimov, Richard Bonneau, and Kyunghyun Cho. Masked graph modeling for molecule generation. *Nature communications*, 12(1):1–12, 2021.

Krzysztof Maziarz, Henry Jackson-Flux, Pashmina Cameron, Finton Sirockin, Nadine Schneider, Nikolaus Stiefl, Marwin Segler, and Marc Brockschmidt. Learning to extend molecular scaffolds with structural motifs. *arXiv preprint arXiv:2103.03864*, 2021.

Harry L Morgan. The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *Journal of chemical documentation*, 5(2):107–113, 1965.

Tian Pan, Yibing Song, Tianyu Yang, Wenhao Jiang, and Wei Liu. Videomoco: Contrastive video representation learning with temporally adversarial examples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11205–11214, 2021.

Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

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

Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems*, 33:12559–12571, 2020a.

Yu Rong, Wenbing Huang, Tingyang Xu, and Junzhou Huang. Dropedge: Towards deep graph convolutional networks on node classification. In *International Conference on Learning Representations*, 2020b.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems*, 30, 2017.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

Janice G Smith. *Organic chemistry*. McGraw-Hill, 2008.

Hannes Stärk, Dominique Beaini, Gabriele Corso, Prudencio Tossou, Christian Dallago, Stephan Günnemann, and Pietro Liò. 3d infomax improves gnns for molecular property prediction. In *International Conference on Machine Learning*, pp. 20479–20502. PMLR, 2022.

Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. *arXiv preprint arXiv:1908.01000*, 2019.

Mengying Sun, Jing Xing, Huijun Wang, Bin Chen, and Jiayu Zhou. 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*, pp. 3585–3594, 2021.

Yuyang Wang, Jianren Wang, Zhonglin Cao, and Amir Barati Farimani. Molclr: Molecular contrastive learning of representations via graph neural networks. *arXiv preprint arXiv:2102.10056*, 2021.

Yuyang Wang, Rishikesh Magar, Chen Liang, and Amir Barati Farimani. Improving molecular contrastive learning via faulty negative mitigation and decomposed fragment contrast. *Journal of Chemical Information and Modeling*, 2022.

Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.

Zhuofeng Wu, Sinong Wang, Jiatao Gu, Madian Khabsa, Fei Sun, and Hao Ma. Clear: Contrastive learning for sentence representation. *arXiv preprint arXiv:2012.15466*, 2020.

Jun Xia, Chengshuai Zhao, Bozhen Hu, Zhangyang Gao, Cheng Tan, Yue Liu, Siyuan Li, and Stan Z. Li. Mole-BERT: Rethinking pre-training graph neural networks for molecules. In *The Eleventh International Conference on Learning Representations*, 2023.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

Minghao Xu, Hang Wang, Bingbing Ni, Hongyu Guo, and Jian Tang. Self-supervised graph-level representation learning with local and global structure. In *International Conference on Machine Learning*, pp. 11548–11558. PMLR, 2021.

Lewei Yao, Runhui Huang, Lu Hou, Guansong Lu, Minzhe Niu, Hang Xu, Xiaodan Liang, Zhenguo Li, Xin Jiang, and Chunjing Xu. FILIP: Fine-grained interactive language-image pre-training. In *International Conference on Learning Representations*, 2022.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 33: 5812–5823, 2020.

Yuning You, Tianlong Chen, Yang Shen, and Zhangyang Wang. Graph contrastive learning automated. In *International Conference on Machine Learning*, pp. 12121–12132. PMLR, 2021.

Sheheryar Zaidi, Michael Schaarschmidt, James Martens, Hyunjik Kim, Yee Whye Teh, Alvaro Sanchez-Gonzalez, Peter Battaglia, Razvan Pascanu, and Jonathan Godwin. Pre-training via denoising for molecular property prediction. *arXiv preprint arXiv:2206.00133*, 2022.

Shichang Zhang, Ziniu Hu, Arjun Subramonian, and Yizhou Sun. Motif-driven contrastive learning of graph representations. *arXiv preprint arXiv:2012.12533*, 2020.

Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34, 2021.

Lingxiao Zhao, Louis Härtel, Neil Shah, and Leman Akoglu. A practical, progressively-expressive gnn. *arXiv preprint arXiv:2210.09521*, 2022.

Gengmo Zhou, Zhifeng Gao, Qiankun Ding, Hang Zheng, Hongteng Xu, Zhewei Wei, Linfeng Zhang, and Guolin Ke. Uni-mol: A universal 3d molecular representation learning framework. 2022.

Jinhua Zhu, Yingce Xia, Tao Qin, Wengang Zhou, Houqiang Li, and Tie-Yan Liu. Dual-view molecule pre-training. *arXiv preprint arXiv:2106.10234*, 2021a.

Jinhua Zhu, Yingce Xia, Lijun Wu, Shufang Xie, Tao Qin, Wengang Zhou, Houqiang Li, and Tie-Yan Liu. Unified 2d and 3d pre-training of molecular representations. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2626–2636, 2022.

Yanqiao Zhu, Yichen Xu, Hejie Cui, Carl Yang, Qiang Liu, and Shu Wu. Structure-aware hard negative mining for heterogeneous graph contrastive learning. *arXiv preprint arXiv:2108.13886*, 2021b.

A Related Work

Multi-view molecular representation learning. Recent works have incorporated *multiple views* of a molecule (e.g., 2D topology and 3D geometry) into molecular representation learning (MRL) frameworks (Zhu et al., 2021a; Fang et al., 2022; Stärk et al., 2022; Liu et al., 2022). In particular, training 2D-GNNs with multi-view MRL has gained much attention to alleviate the large cost to obtain 3D geometry of molecules (Stärk et al., 2022; Liu et al., 2022). However, they focus on molecule-level objectives, which could lack capturing the local semantics (Yao et al., 2022). In this work, we develop a fragment-based multi-view MRL framework to incorporate fine-grained cross-view interactions. Moreover, since our method is architecture-agnostic, it would be an interesting future direction to incorporate our framework into Transformer-based approaches (Zhu et al., 2022; Luo et al., 2022).

Single-view molecular representation learning. One of the single-view (i.e., 2D topological or 3D geometric graph) molecular representation learning techniques is predictive pretext tasks. For example, those methods reconstruct the corrupted input as pre-defined pretext tasks (Hamilton et al., 2017; Hu et al., 2020a; Rong et al., 2020a; Zhang et al., 2021; Zhou et al., 2022; Jiao et al., 2022; Zaidi et al., 2022). Another large portion of technique is contrastive learning. For example, You et al. (2020; 2021); Wang et al. (2021); Zhang et al. (2020) utilize augmentation schemes to produce a positive view of molecular graphs, and Fang et al. (2021); Sun et al. (2021); Wang et al. (2022) mitigate the effect of semantically similar molecules in the negative samples (Zhu et al., 2021b).

Molecular fragmentation. Recent advancements in the field of molecule generation (Maziarz et al., 2021; Jin et al., 2018; 2020) have recognized the significance of semantically important substructures, also known as fragments, in determining the properties of molecules. This approach aligns with the chemical principle that the properties of a molecule are primarily determined by its important substructures, rather than atom-level features (Smith, 2008).

Recently, substructures of molecules has also been considered in molecular contrastive learning. For example, You et al. (2020); Wang et al. (2021); Zhang et al. (2020) construct a positive view of a molecule as its single substructure (i.e., subgraph) and Wang et al. (2022) repels representations of fragments from intra- and inter- molecule substructures. Compared to these prior works, we utilize fragmentation as a *semantic-preserving transformation*, considering the set of fragments as a positive view of a molecule.

B EXPERIMENTAL DETAILS

Self-supervised pretraining details. We follow the training setup considered in GraphMVP (Liu et al., 2022): Specifically, we use a batch size of 256 and no weight decay. Also, we set the temperature τ as 0.1 for overall experiments. We use $\{\texttt{Nodedrop}, \texttt{Attrmask}, \texttt{identity}\}$ randomly, i.e., $\frac{1}{3}$ probability for each fragment and the original 2D molecular graphs, and Gaussian noise $\mathcal{N}(0, I)$ to each coordinate of 3D molecular graphs. When `Nodedrop` or `Attrmask` is used, we drop/mask the portion of 0.1 vertices from the total vertices. For self-supervised pretraining, we train for 100 epochs using Adam optimizer (Kingma & Ba, 2014) with a learning rate of 0.001 and no dropout. For transfer learning to the QM9 (Ramakrishnan et al., 2014) dataset, we train with 310k entire unlabeled molecules from GEOM for 50 epochs. For semi-supervised leraning for the QM9 dataset, we train with 110k training molecules (without labels) from QM9 for 50 epochs. Our code is based on open-source codes of GraphMVP[3].

For FragCL trained only with single view objective and other reproduced 2D baselines, we exclude explicit hydrogens in molecular graph, following the common frameworks of (You et al., 2020; 2021) for 2D molecular graphs. For FragCL, 3D-InfoMax, GraphMVP, GraphMVP-C, and GraphMVP-G we include explicit hydrogens into molecular graph, following (Liu et al., 2022) that utilizes the 3D coordinates of hydrogen atoms provided in GEOM dataset (Axelrod & Gomez-Bombarelli, 2022). For torsional angle prediction task, we use 2-layer MLP for $g_{\texttt{tor}}$ and we construct the quartet of atoms (s, u, v, t) for the fragmented bond (u, v) so that s, t are non-hydrogen atoms, and the binning of y splits 0 to 2π into 18 uniform bins. In terms of time-complexity, FragCL takes almost the same amount of training cost as GraphMVP (Liu et al., 2022).

Evaluation on MoleculeNet downstream tasks. Following the baselines, we use *scaffold split* (Chen et al., 2012), which splits the molecules based on their substructures. We use the split ratio train:validation:test = 80:10:10 for each downstream task dataset to evaluate the performance. For the consistency of the input graphs in pretraining and fine-tuning, we exclude implicit hydrogen atoms of molecules in fine-tuning dataset for single-view pretrained FragCL and other reproduced 2D baselines and we include implicit hydrogen atoms of molecules in fine-tuning dataset for FragCL, 3D-InfoMax, GraphMVP, GraphMVP-C, and GraphMVP-G. Experimental detail follows GraphMVP (Liu et al., 2022); we fine-tune a pretrained 2D GNN with an initialized linear projection layer for 100 epochs with Adam optimizer and a learning rate of 0.001, and dropout probability of 0.5. Our results are calculated by the test ROC-AUC score of the epoch with the best validation ROC-AUC score. Besides the ROC-AUC score of individual downstream tasks, we also report the average ROC-AUC score across downstream datasets.

Evaluation on QM9 downstream tasks. Following (Liu et al., 2021), we split the molecules in the QM9 (Ramakrishnan et al., 2014) dataset into 110,000 molecules for training, 10,000 molecules for validation, and 10,831 molecules for test. Our result is calculated by the test MAE score of the epoch with the best validation MAE score. We fine-tune a pretrained 2D GNN with an initialized 2-layer MLP for 1,000 epochs with Adam optimizer and StepLR scheduler with decay ratio 0.5, and initial learning rate 5e-4.

Hardwares. We use a single NVIDIA GeForce RTX 3090 GPU with 36 CPU cores (Intel(R) Core(TM) i9-10980XE CPU @ 3.00GHz) for self-supervised pretraining, and a single NVIDIA GeForce RTX 2080 Ti GPU with 40 CPU cores (Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz) for fine-tuning.

[3] `https://github.com/chao1224/GraphMVP`

C BASELINES DETAILS

We compare our method with an extensive list of baseline methods in the literature of graph representation learning:

- *No pretraining* trains a model from scratch for downstream task.
- *EdgePred* (Hamilton et al., 2017) uses edge-reconstruction as a pretext task.
- *AttrMask* (Hu et al., 2020a) train GNN encoder by recovering the vertex features from the masked vertex features.
- *AttrMask* (Hu et al., 2020a) learns representation by recovering the vertex features after masking them.
- *GPT-GNN* (Hu et al., 2020b) uses the graph generation task as a pretext task.
- *Infomax* (Sun et al., 2019) maximizes mutual information between global representations (i.e., graph representations) and local representations (i.e. path representation).
- *ContextPred* (Hu et al., 2020a) learns representation by predicting surrounding subgraph of specific node edge.
- *GraphLoG* (Xu et al., 2021) discriminates graph and subgraph pairs from their opposing pairs to preserve local similarity between various graphs, which leads to the embedding alignment of correlated graphs.
- *G-Contextual* (Rong et al., 2020a) learns representations by randomly masking local subgraphs of target nodes (or edges) and predicting these contextual properties from node embeddings.
- *G-Motif* (Rong et al., 2020a) predicts the occurrence of the semantic motifs extracted by using chemical prior.
- *GraphCL* (You et al., 2020) is a generic graph contrastive learning method based on their graph-agnostic augmentation schemes, which do not use any molecule-specific knowledge.
- *JOAO* (You et al., 2021) proposes min-max optimization processes to learn optimal data augmentation strategies dynamically from a pre-fixed candidate set of augmentations.
- *MGSSL* (Zhang et al., 2021) introduces a generative self-supervised objective to reconstruct a motif-tree.
- *MolCLR* (Wang et al., 2021) performs a contrastive learning with NT-Xent (Chen et al., 2020), constructing positive views of a molecule by proposed molecule augmentation schemes.
- *D-SLA* (Kim et al., 2022) extracts graph representations by learning the exact discrepancy between the original graph and the augmented graphs.
- *3D-InfoMax* (Stärk et al., 2022) proposes to consider 2D topological molecule graph and 3D geometric molecule graph from the same molecule as a positive view of each other.
- *GraphMVP, GraphMVP-G,* and *GraphMVP-C* (Liu et al., 2022) regard 2D and 3D molecular graphs as a positive pair, and propose feature reconstruction of each view as a generative task.

D GRAPH NEURAL NETWORKS

In general, a molecule $M \in \mathcal{M}$ can be represented by an attributed graph $M = (V, E, A, B, R)$ where V is a set of nodes associated with atom features $A \in \mathsf{R}^{|V| \times d_{\text{atom}}}$ (e.g., atomic numbers), $E \subseteq V \times V$ is a set of edges associated with bond features $B \in \mathsf{R}^{|E| \times d_{\text{bond}}}$ (e.g., bond types), and $R \in \mathsf{R}^{|V| \times 3}$ is an array of 3D atom positions. Conventionally, $M_{\texttt{2D}} = (V, E, A, B)$ and $M_{\texttt{3D}} = (V, A, R)$ are referred to 2D topological and 3D geometric molecular graphs, respectively (Stärk et al., 2022; Liu et al., 2022). It is worth noting that obtaining accurate 3D geometric information R is very expensive due to iterative quantum computations and thus many real-world applications often suffer from the lack of such 3D information (Liu et al., 2022). We employ 5-layer graph isomorphism network (GIN) (Xu et al., 2019) as 2D-GNN $f_{\texttt{2D}}$ and 6-layer SchNet (Schütt et al., 2017) as 3D-GNN $f_{\texttt{3D}}$. We use mean pooling as readout function of both $f_{\texttt{2D}}$ and $f_{\texttt{3D}}$. The configuration is drawn from GraphMVP (Liu et al., 2022) for a fair comparison.

2D molecule GNN $f_{\texttt{2D}} : \mathcal{M}_{\texttt{2D}} \to \mathsf{R}^d$**.** For any 2D molecule $M_{\texttt{2D}} = (V, E, A, B) \in \mathcal{M}_{\texttt{2D}}$, graph neural networks for 2D molecules (2D-GNNs in short) compute molecular representations by applying (a) iterative neighborhood aggregation (also known as message passing) to acquire node-level representations based on the graph (V, E) and then (b) a readout function (e.g., mean pooling) to create graph-level representations at the final layer. Formally, node- and graph- level representations of L-layer 2D-GNN are as follows:

$$\mathbf{h}_v^{(\ell)} := \texttt{MP}(\mathbf{h}_v^{(\ell-1)}, \{\mathbf{h}_u^{(\ell-1)}, B_{uv}\}_{u \in \mathcal{N}(v)}),\ \ell \in [L],$$
$$f_{\texttt{2D}}(M) := f_{\texttt{2D}}(M_{\texttt{2D}}) = \texttt{Readout}(\{\mathbf{h}_v^{(L)}\}_{v \in V}),$$

where $\texttt{MP}(\cdot)$ is a message passing layer, $\texttt{Readout}(\cdot)$ is a readout function, $\mathbf{h}_v^{(0)} = A_v$ is the atom feature for a node v, B_{uv} is the bond feature for an edge $(u, v) \in E$, and $\mathcal{N}(v)$ is the set of adjacent nodes of v. In a decade, there have been developed a number of message passing layers and readout functions (Kipf & Welling, 2017; Xu et al., 2019; Ahmadi, 2020; Zhao et al., 2022). In this work, we mainly use the graph isomorphism network (GIN) architecture (Xu et al., 2019) following the standard MRL setup (Hu et al., 2020a).

Graph Isomorphism Network (GIN). We provide a detailed description of architecture of graph isomorphism network (GIN) (Xu et al., 2019), which we mainly consider as the feature extractor $f_{\texttt{2D}}(\cdot)$ in this paper. Particularly, GIN learns representation $\mathbf{h}_v^{(\ell)}$ by:

$$\mathbf{h}_v^{(\ell)} = \texttt{MLP}^{(\ell)}\big(\mathbf{h}_v^{(\ell-1)} + \sum_{u \in \mathcal{N}(v)} \big(\mathbf{h}_u^{(\ell-1)} + \mathbf{e}_{uv}^{(\ell-1)}\big)\big), \tag{7}$$

where $\mathbf{e}_{uv}^{(\ell-1)}$ is the embedding corresponding to the attribute of edge $\{u, v\} \in \mathcal{E}$.

3D molecule GNN $f_{\texttt{3D}} : \mathcal{M}_{\texttt{3D}} \to \mathsf{R}^d$**.** For any 3D molecule $M_{\texttt{3D}} = (V, A, R) \in \mathcal{M}_{\texttt{3D}}$, graph neural networks for 3D molecules (3D-GNNs in short) compute molecular representations by applying (a) iterative geometric interactions through distances and angles between nodes (i.e., atoms) to acquire node-level representations based on the 3D geometry R and then (b) a readout function to create graph-level representation at the final layer. Formally, node- and graph- level representations of L-layer 3D-GNN are as follows:

$$\mathbf{h}_v^{(\ell)} := \texttt{IB}(\mathbf{h}_v^{(\ell-1)}, R_v, \{\mathbf{h}_u^{(\ell-1)}, R_u\}_{u \in V \setminus \{v\}}), \ell \in [L],$$
$$f_{\texttt{3D}}(M) := f_{\texttt{3D}}(M_{\texttt{3D}}) = \texttt{Readout}(\{\mathbf{h}_v^{(L)}\}_{v \in V}),$$

where $\texttt{IB}(\cdot)$ is an interaction block, $\texttt{Readout}(\cdot)$ is a readout function, $\mathbf{h}_v^{(0)} = A_v$ and R_v is the atom feature and the 3D position for a node v, respectively. A number of interaction layers has been developed to encode geometric features of molecules (Schütt et al., 2017; Liu et al., 2021; Schütt et al., 2021). In this work, we mainly use the SchNet architecture (Schütt et al., 2017) following the setup of Liu et al. (2022).

SchNet. We consider SchNet (Schütt et al., 2017), which is a strong 3D graph neural network under fair comparison (Liu et al., 2022) as our $f_{\texttt{3D}}(\cdot)$ in this paper. Particularly, SchNet learns representation $\mathbf{h}_v^{(\ell)} = \texttt{MLP}^{(\ell)}$ by:

$$\mathbf{h}_v^{(\grave{})} = \texttt{MLP}^{(\grave{})}\Big(\sum_{u \in V}\Big(\quad (\mathbf{h}_u^{(\grave{}-1)}, \mathbf{r}_v, \mathbf{r}_u)\Big)\Big), \tag{8}$$

where is the continuous-filter convolution layer and $\mathbf{r}_v$ is the 3D position of the vertex v.

E DOWNSTREAM DATASET DETAILS

We perform transfer-learning on 8 benchmark binary classification datasets from MoleculeNet (Wu et al., 2018). More information on downstream tasks is described in Table 4.

- *BBBP* contains data on whether the compound is permeable to the blood-brain barrier.
- *Tox21* measures the toxicity of a compound and was used in the 2014 Tox21 Data Challenge.
- *ToxCast* includes multiple toxicity annotations of compounds collected after performing high-throughput screening tests.
- *Sider* refers to side effect resources, i.e., data on the marketed drugs and their side effects.
- *Clintox* is a dataset of comparison results between drugs approved through the FDA and drugs removed because of toxicity during clinical trials.
- *MUV* is a validation dataset of virtual screening technology. Specifically, it is subsampled in the PubChem BioAssay using refined nearest neighborhood analysis.
- *HIV* consists of data about capability to prevent HIV replication.
- *Bace* is collected dataset of compounds that could prevent (BACE-1).

Table 4: MoleculeNet downstream classification dataset statistics

Dataset	BBBP	Tox21	ToxCast	Sider	Clintox	MUV	HIV	Bace
Number of molecules	2,039	7,831	8,575	1,427	1,478	93,087	41,127	1,513
Number of tasks	1	12	617	27	2	17	1	1
Avg. Node	24.06	18.57	18.78	33.64	26.15	24.23	25.51	34.08
Avg. Degree	51.90	38.58	38.52	70.71	55.76	52.55	54.93	73.71

We also perform transfer-learning on 12 benchmark regression tasks from QM9 (Ramakrishnan et al., 2014). More information on downstream tasks is described in Table 5 .

Table 5: QM9 downstream regression tasks

Task	Summary	Unit
ZPVE	Zero point vibrational energy	meV
μ	Dipole moment	D
α	Isotropic polarizability	${a_0}^3$
C_v	Heat capacity at $298.15K$	$\text{cal}/\text{mol} \cdot \text{K}$
LUMO	Lowest unoccupied molecular orbital energy	meV
HOMO	Highest occupied molecular orbital energy	meV
ε_{gap}	Gap between HOMO and LUMO	meV
R^2	Electronic spatial extent	${a_0}^2$
U_0	Internal energy at $0K$	meV
U_{298}	Internal energy at $0K$	meV
H_{298}	Enthalpy at $0K$	meV
G_{298}	Gibbs energy at $0K$	meV

F DETAILED RESULTS ON QM9

Table 6: Comparison of test MAE score of semi-supervised learning on the QM9 downstream quantum property regression benchmarks. We pretrain GIN (Xu et al., 2019) as the 2D-GNN architecture with 110k QM9 training set and fine-tune on 10% subset of QM9 training set. We mark the best score bold.

Methods	ZPVE ↓	μ ↓	α ↓	C_v ↓	LUMO ↓	HOMO ↓	ε_{gap} ↓	R^2 ↓	U_0 ↓	U_{298} ↓	H_{298} ↓	G_{298} ↓
-	173.1	0.339	2.67	0.882	415.5	340.7	680.8	20.6	278.0	301.3	299.9	274.1
Pretrained on 110k 2D and 3D molecular graphs of QM9 and fine-tuned on 10% 2D molecular graphs of QM9												
3D-Infomax Stärk et al. (2022)	166.7	0.325	2.59	0.878	395.3	332.7	672.7	20.4	257.5	284.1	283.9	249.4
GraphMVP-G Liu et al. (2022)	152.6	0.324	2.58	0.872	388.3	325.8	662.7	19.9	255.4	281.4	271.7	245.3
FragCL (Ours)	**151.5**	**0.322**	**2.51**	**0.869**	**381.0**	**321.2**	**650.5**	**19.8**	**252.9**	**279.4**	**269.1**	**243.6**

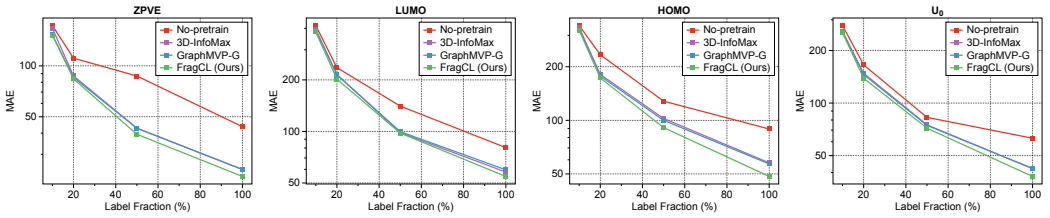
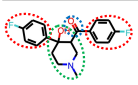
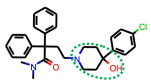
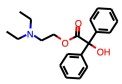
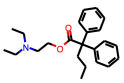
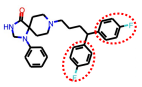
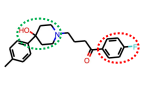
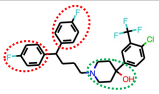
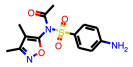
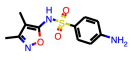


Figure 2: Comparison of test MAE score of semi-supervised learning with different fraction of labeled samples on QM9 downstream quantum property regression benchmarks. We pretrain GIN (Xu et al., 2019) as the 2D-GNN architecture with 110k molecules from QM9 pretraining dataset.

In this section, we provided detailed results for semi-supervisd learning on the QM9 (Ramakrishnan et al., 2014) dataset. Figure 2 shows the test MAE score across different label fractions after pretrained with the QM9 training dataset. We choose 4 downstream tasks which yields the highest performance gap after pretraining compared to non-pretraining (we exclude $\varepsilon_{gap} := |\text{HOMO} - \text{LUMO}|$ since we already include HOMO and LUMO). As visualized, FragCL consistently outperforms the considered baselines. Table 6 shows the results for all 12 downstream tasks of QM9 when fine-tuned with 10% of training data. For all downstream tasks, FragCL achieves the best performance.

G MOLECULE RETRIEVAL

Table 7: Retrieved molecules by searching the three most closest molecules from the Tox21 dataset to the query molecule in terms of similiarty in representation space from pretrained models of GraphMVP-G and FragCL (Ours) with the GEOM dataset (Axelrod & Gomez-Bombarelli, 2022). We mark fragments by BRICS as dotted lines.

Query	GraphMVP-G (Liu et al., 2022)			**FragCL (Ours)**		
						
						

We further perform molecule retrieval task for qualitative analysis. Using pretrained models by FragCL and GraphMVP-G (Liu et al., 2022), we calculate the cosine similarity of representations between the query molecule and the molecules in the Tox21 dataset. In Table 7, three molecules most similar to the query molecule are presented. While GraphMVP-G does not find molecules with similar fragments to the query molecule, FragCL effectively retrieves molecules with common fragments (indicated by dotted lines in Table 7) in the query molecule.

H Ablation Study

H.1 Main ablation

Table 8: Average ROC-AUC score with differnt posivie view construction strategy across 8 downstream tasks in MoleculeNet.

Positive view construction	Fragmentation strategy	Avg.
Nodedrop, Subgraph (You et al., 2020)	-	73.4
A set of fragments (Ours)	Random bond deletion	73.5
	Random non-ring bond deletion	74.0
	BRICS decomposition (Ours)	**75.5**

Table 9: Effectiveness of each objective as measured on the average ROC-AUC score across 8 downstream tasks in MoleculeNet.

Pretraining data	Cross-view interaction: Molecule-level	Fragment-level	Torsion-level	Avg.
Single-view (2D)	-	-	-	72.4
Multi-view (2D&3D)	X	-	-	74.7
	X	X	-	75.1
	X	X	X	**75.5**

Fragment-based positive view construction. In Table 8, we investigate how our positive view construction strategy is effective. We first compare our strategy with the alternative: an augmented molecular graph (i.e., random subgraph) as a positive view (You et al., 2020). We observe that deleting random bonds for positive-view construction does not improve the performance (73.4 $\rightarrow$ 73.5), since important substructures of molecules (e.g., aromatic ring) can be easily broken by random deletion of bonds, which could lead to significant change in chemical properties. Preventing such ring deformation increases overall performance by 73.5 $\rightarrow$ 74.0. BRICS decomposition further incorporates chemical prior to obtain *semantic-preserved fragments*, boosting the performance by 74.0 $\rightarrow$ 75.5. The result implies that considering chemically informative substructures is a key component of our framework. We provide detailed results in Appendix H.2.

Effectiveness of multi-view pretraining. In Table 9, we evaluate how each objective in our total loss $\mathcal{L}_{\texttt{FragCL}}$ affects performance. We observe that molecule-level cross-view contrastive learning ($\mathcal{L}_{\texttt{cross;mol}}$; Eq. (3)) between 2D and 3D molecular views improves the overall performance by 72.4 $\rightarrow$ 74.7. Introducing fragment-level cross-view contrastive learning ($\mathcal{L}_{\texttt{cross;frag}}$; Eq. (4)) further boosts the performance by 74.7 $\rightarrow$ 75.1, capturing fine-grained semantics of molecules. Torsional angle prediction ($\mathcal{L}_{\texttt{tor}}$; Eq. (5)) further improves the performance by 75.1 $\rightarrow$ 75.5 by directly injecting the information of 3D geometric view into 2D-GNN. These results confirm that FragCL effectively utilizes both 2D and 3D fragmented views for multi-view pretraining. Notably, ours with only single-view contrastive (2D) learning outperforms Mole-BERT (Xia et al., 2023), which is the prior state-of-the-art pretraining method on 2D molecule data. Detailed results can be found in Appendix H.2.

Table 10: Average number of atoms in MoleculeNet dataset. Hydrogen is included.

Avg. # of atoms	BBBP	Tox21	ToxCast	Sider	Clintox	MUV	HIV	Bace
Train	43.6	33.0	33.0	57.8	**49.6**	42.6	45.3	63.5
Test	52.1	49.8	52.3	102.0	**43.0**	44.8	45.5	67.9

Table 11: Improvement of FragCL compared to GraphMVP-G.

Gap b/w FragCL and GraphMVP-G	BBBP	BACE	MUV	HIV
Full test set	1.3	1.0	0.1	1.0
$<$ Avg. # of atoms in training set	**2.3**	**1.1**	**2.9**	**1.5**

Effectiveness on Clintox. In Table 1, the improvement of our method is the largest in Clintox dataset. In Table 10, ClinTox is unique in that it has a higher average number of atoms in the training set compared to the test set. Furthermore, we observe that the performance gap between FragCL and GraphMVP-G (the strongest baseline) tends to widen as the number of atoms in test molecules decreases. To illustrate this point, Table 11 below compares the performance gap when evaluated on the full test set and on the molecules in the test set with fewer atoms than the average in the training set. We consider the downstream tasks with $|$Avg. # of atoms in the training set molecules $-$ Avg. # of atoms in the test set molecules$| < 10$, due to the stability of evaluation. FragCL's enhanced performance on smaller molecules can be explained by its capacity to learn fine-grained molecular features through fragmentation.

H.2 Ablation details

Table 12: Comparison of positive view construction strategies for multi-view molecular contrastive learning framework. We report the test ROC-AUC score on the MoleculeNet downstream property classification benchmarks. We pretrain GIN (Xu et al., 2019) as the 2D-GNN architecture with 50k molecules from the GEOM dataset (Axelrod & Gomez-Bombarelli, 2022), following Liu et al. (2022). We report mean and standard deviation over 3 different seeds. We bold the best average score.

Positive view construction	Fragmentation strategy	BBBP	Tox21	ToxCast	Sider	Clintox	MUV	HIV	Bace	Avg.
Nodedrop, Subgraph	-	69.3±1.4	75.0±0.4	63.7±0.4	60.4±1.4	88.3±0.6	76.2±1.9	76.2±1.5	78.3±0.4	73.4
A set of fragments (Ours)	Random bond deletion	69.3±1.0	73.8±0.9	63.9±0.5	59.9±1.2	91.4±2.3	76.8±0.7	74.6±3.1	78.3±2.5	73.5
	Random non-ring bond deletion	69.5±0.9	73.7±0.2	64.0±0.1	60.5±0.5	93.2±1.5	77.3±2.5	75.2±0.9	78.8±0.4	74.0
	BRICS decomposition (Ours)	**71.4**±0.4	**75.2**±0.7	**65.1**±0.8	**61.0**±0.6	**95.2**±1.0	**77.6**±1.0	**76.3**±0.4	**82.3**±1.6	**75.5**

In Table 12 we provide a full result of Table 8 in Section H. We conduct an ablation study on regarding the set of fragments as a positive view of a molecule. Again, we emphasize that the result implies that considering chemically informative structures is a key component of FragCL.

Table 13: Ablation of components for multi-view molecular contrastive learning framework. We report the test ROC-AUC score on the MoleculeNet downstream property classification benchmarks. We pretrain GIN (Xu et al., 2019) as the 2D-GNN architecture with 50k molecules from the GEOM dataset (Axelrod & Gomez-Bombarelli, 2022), following Liu et al. (2022). We report mean and standard deviation over 3 different seeds. We mark the best mean score to be bold.

Pretraining data	Multi-view interaction: Molecule-level	Fragment-level	Torsion-level	BBBP	Tox21	ToxCast	Sider	Clintox	MUV	HIV	Bace	Avg.
Single-view (2D)	-	-	-	71.0±0.3	75.3±0.8	62.8±0.4	60.3±1.1	79.1±2.2	74.1±0.5	75.9±1.2	80.7±1.3	72.4
Multi-view (2D & 3D)	X	-	-	68.2±0.6	**75.6**±1.5	64.6±0.2	60.8±0.8	94.9±0.8	**77.7**±1.2	**76.3**±0.5	79.5±0.3	74.7
	X	X	-	71.0±0.8	75.3±0.9	64.4±0.3	**61.6**±2.6	95.1±1.5	76.4±1.6	76.2±0.7	80.9±2.6	75.1
	X	X	X	**71.4**±0.4	75.2±0.7	**65.1**±0.8	61.0±0.6	**95.2**±1.0	77.6±1.0	**76.3**±0.4	**82.3**±1.6	**75.5**

In Table 13, we provide a full result of Table 9 in Section H. We validate that each components of FragCL has an individual effect in improving the performance of multi-view pretraining.