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)

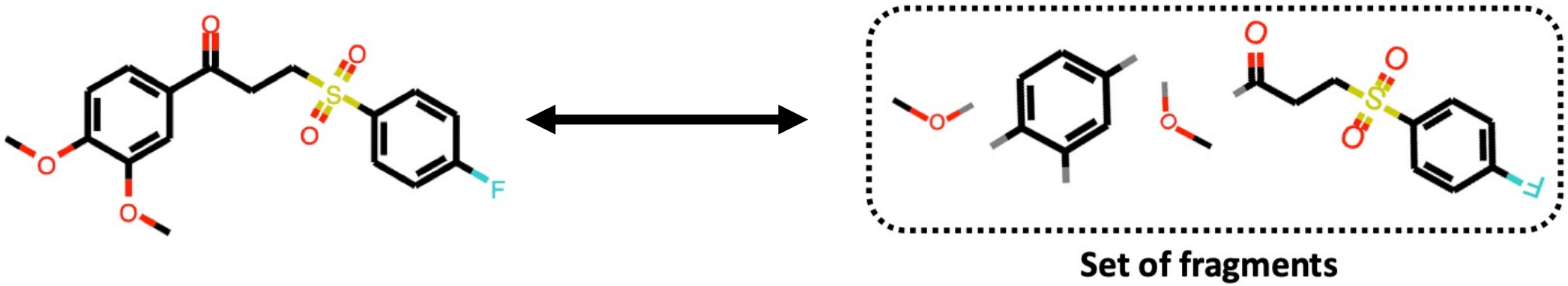
Importance of Fragment-based Molecular Representation Learning

Fragment-based molecular representation learning is important in chemical application:

- **Drug-likeness of molecules**: High labeling costs
- **Functional groups**: Correlation between substructures and molecular property

- **FragCL**: We propose **a novel framework for fragment-based molecular representation learning**.

💡 Superior performance compared to prior methods by capturing fragment-level features.



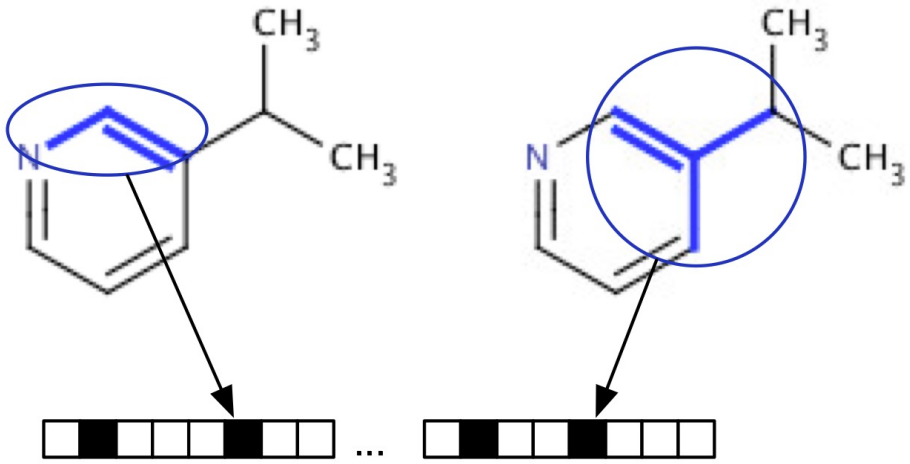
Motivation: Why fragment-level information is critical?

Properties of a molecule can be inferred from its fragments

- For example, a fluorobenzene fragment indicates the label-1 for BBBP dataset with 97.0 %
- Chemists tried to incorporate fragment-wise information into fingerprint [Morgan et al., 1965] representation
- However, it is non-learnable and not appropriate for ML framework

BBBP (76.4 % are label-1)	Ratio of label-1 (%)		
Top-3 to label-1	c1ccc(F)cc1 97.0	N1CCNCC1 90.7	N1CCCCC1 89.9
Top-3 to label-0	C(=O)O 17.9	S 34.0	N 56.9

HIV (3.5% are label-1)	Ratio of label-1 (%)	
Top-3 to label-1	C1CC(N=[N+]=[N-])CO1 78.2	S(=O)(=O)c1cc(C)c(Cl)cc1S 49.2
Top-3 to label-0	N1CCNCC1 0.4	S(=O)(=O)c1ccc(C)cc1 0.6



[Stark et al., 2022] 3D Infoamx improves GNNs for Molecular Property Prediction, ICML 2022
[Morgan et al., 1965] The Generation of a Unique Machine Description for Chemical Structures, Journal of Chemical Documentation 1965

Tackling Limited Label Issues: Multi-view Self-supervised Learning

Approaches for molecular representation learning: Multi-view self-supervised learning

- 3D molecular geometry is critical for predicting molecular properties, but costly to obtain in downstream tasks
- Pretrain 2D molecular GNN with the aid of 3D information (e.g., 3D-Infomax [Stark et al., 2022], GraphMVP [Liu et al., 2022])
- Assume we have 2D/3D pretraining dataset and 2D downstream dataset, following the practical scenario
- Beneficial than utilizing only 2D molecules (e.g., MGSSL [Zhang et al., 2021], MolCLR [Wang et al., 2022]) in pretraining

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
MGSSL	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	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
3D-InfoMax	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-G	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

[Stark et al., 2022] 3D Infoamx improves GNNs for Molecular Property Prediction, ICML 2022
[Liu et al., 2022] Pre-training Molecular Graph Representation with 3D Geometry, ICLR 2022
[Zhang et al., 2022] Motif-based Graph Self-Supervised Learning for Molecular Property Prediction, NeurIPS 2021
[Wang et al., 2022] MolCLR: Molecular Contrastive Learning of Representations via Graph Neural Networks, Nature Machine Intelligence 2022

Utilizing Fragments in Multi-view Self-supervised Learning?

Fragment-based multi-view contrastive learning: Inject fragment-level information into contrastive learning

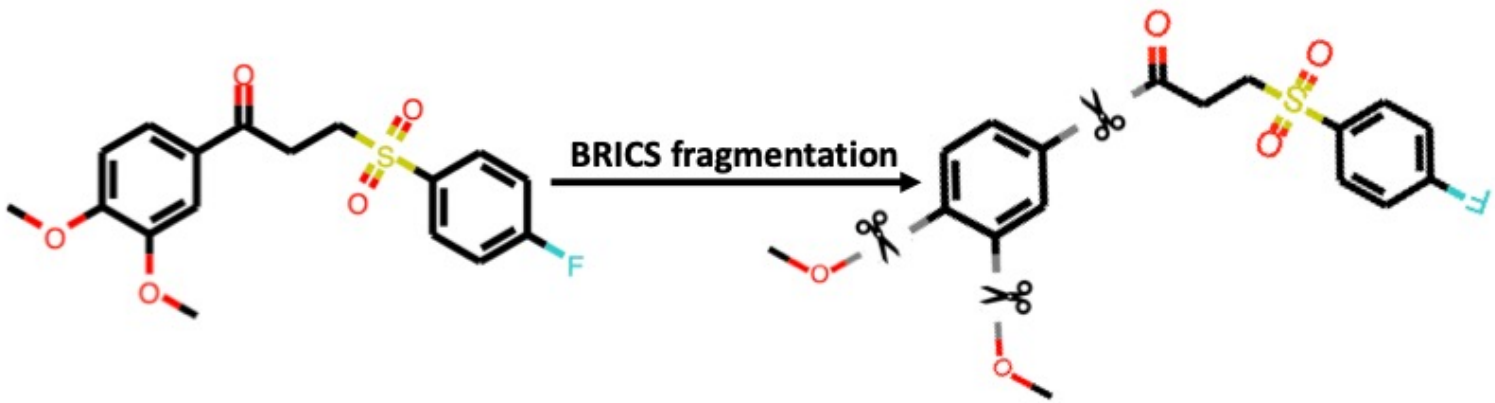
- Generating positive pairs and negative pairs is key to pretrain a neural network via contrastive learning.

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

🤔 How should we generate effective positive/negative pairs for fragment-based contrastive learning?

💡 **Idea 1:** We regard a set of fragments as a positive view of the original molecule.

- A molecule can be viewed as a set of its meaningful fragments; we use BRICS decomposition [Degen et al., 2008]



[Degen et al., 2008] On the Art of Compiling and Using 'Drug-like' Chemical Fragment Spaces, ChemMedChem, 2008

Utilizing Fragments in Multi-view Self-supervised Learning?

Fragment-based multi-view contrastive learning: Inject fragment-level information into contrastive learning

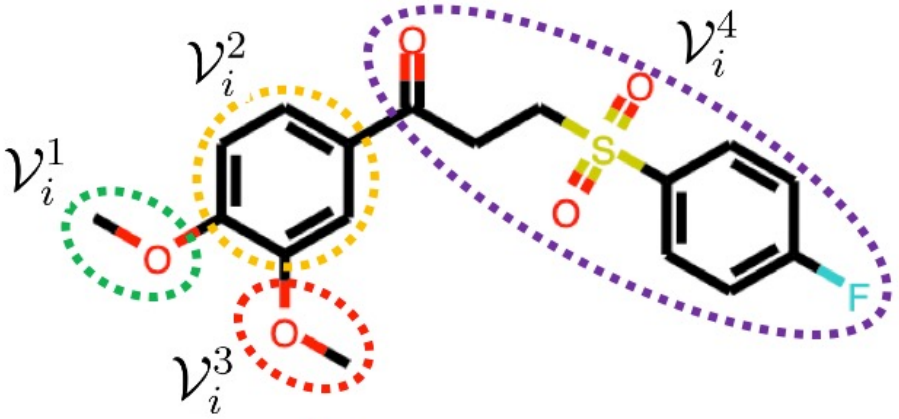
- Generating positive pairs and negative pairs is key to pretrain a neural network via contrastive learning.

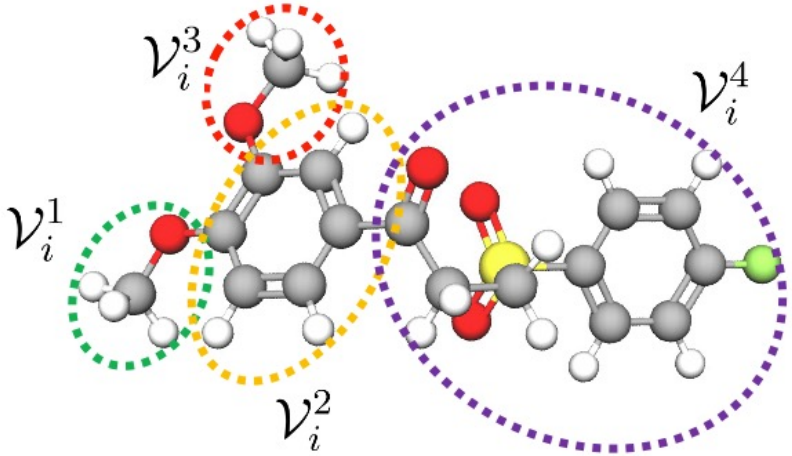
$$\mathcal{L}_{\mathrm{CL}}(\mathbf{z}, \mathbf{z}^+, \{\mathbf{z}^-\}) = -\log \frac{\exp(\mathrm{sim}(\mathbf{z}, \mathbf{z}^+)/\tau)}{\sum_{\mathbf{z}^-} \exp(\mathrm{sim}(\mathbf{z}, \mathbf{z}^-)/\tau)}$$

🤔 How should we generate effective positive/negative pairs for fragment-based contrastive learning?

💡 **Idea 2:** We regard the corresponding fragments in the 2D and 3D molecule as a positive view.

- Corresponding fragments represent exactly the same entity.





Proposed Framework: FragCL

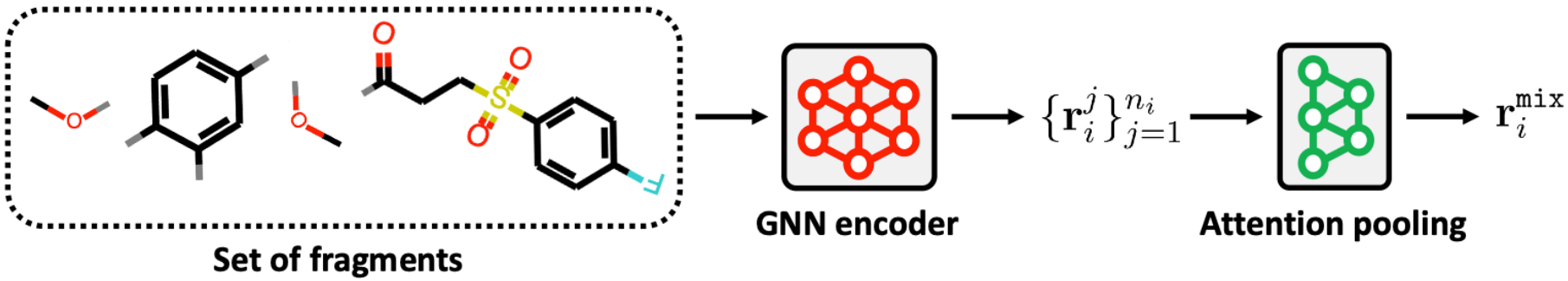
Self-supervised molecular pretraining with Fragment-based multi-view Contrastive Learning (FragCL)

- Generate fragment-based positive/negative views from a given unlabeled molecules.
- Apply multi-view contrastive learning framework to learn generalizable representations.

🤔 How can we make the representation for a set of fragments?

💡 Utilize attention pooling to aggregate fragment-wise representation!

- The encoder automatically learns how to mix fragment-level information

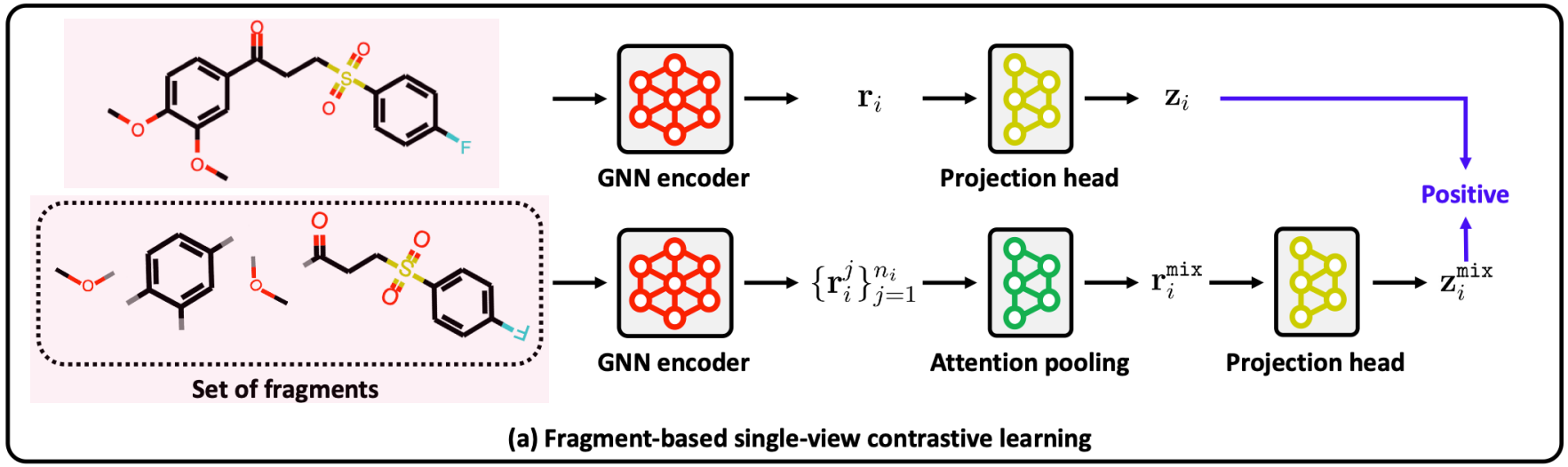


Fragment-based Contrastive Learning (FragCL): Single-view Objective

View construction strategy for single-view molecules within 2D (or 3D) molecules

- Positive views: A molecule and its set of fragments $(M_i, \{M_i^j\})$
- Negative views: A molecule and a set of fragments from another molecule $(M_i, \{M_k^l\})$ with $k \neq i$
- Where M_i denotes the i-the molecule in a mini-batch and M_i^j denotes the j-th fragment of i-th molecule

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



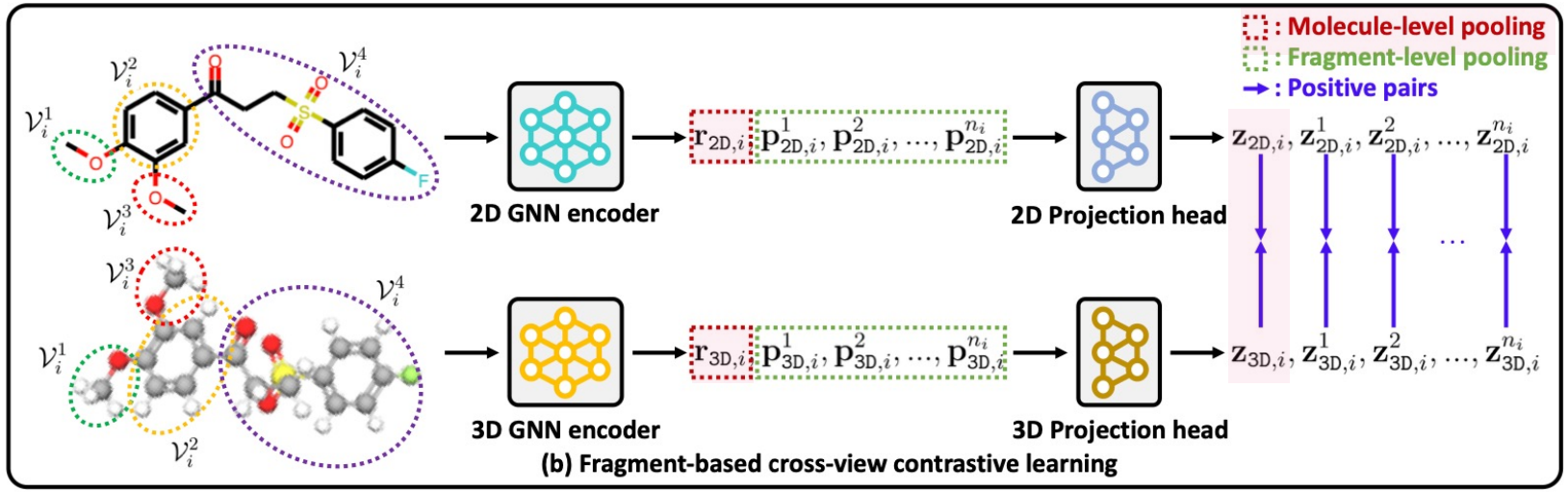
(a) Fragment-based single-view contrastive learning

Fragment-based Contrastive Learning (FragCL): Cross-view Objective

View construction strategy for cross-view molecules between 2D and 3D molecules: Molecule-level

- Positive views: Different view of the same molecule $(M_{i,\mathrm{2D}}, M_{i,\mathrm{3D}})$
- Negative views: Different view of different molecules $(M_{i,\mathrm{2D}}, M_{j,\mathrm{3D}})$ with $i \neq j$
- Where $M_{i,\mathrm{2D\ (or\ 3D)}}$ denotes the i-the molecule in 2D (or 3D) view; borrowed from [Stark et al., 2022], [Liu et al., 2022]

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



[Stark et al., 2022] 3D Infoamx improves GNNs for Molecular Property Prediction, ICML 2022

[Liu et al., 2022] Pre-training Molecular Graph Representation with 3D Geometry, ICLR 2022

Fragment-based Contrastive Learning (FragCL): Cross-view Objective

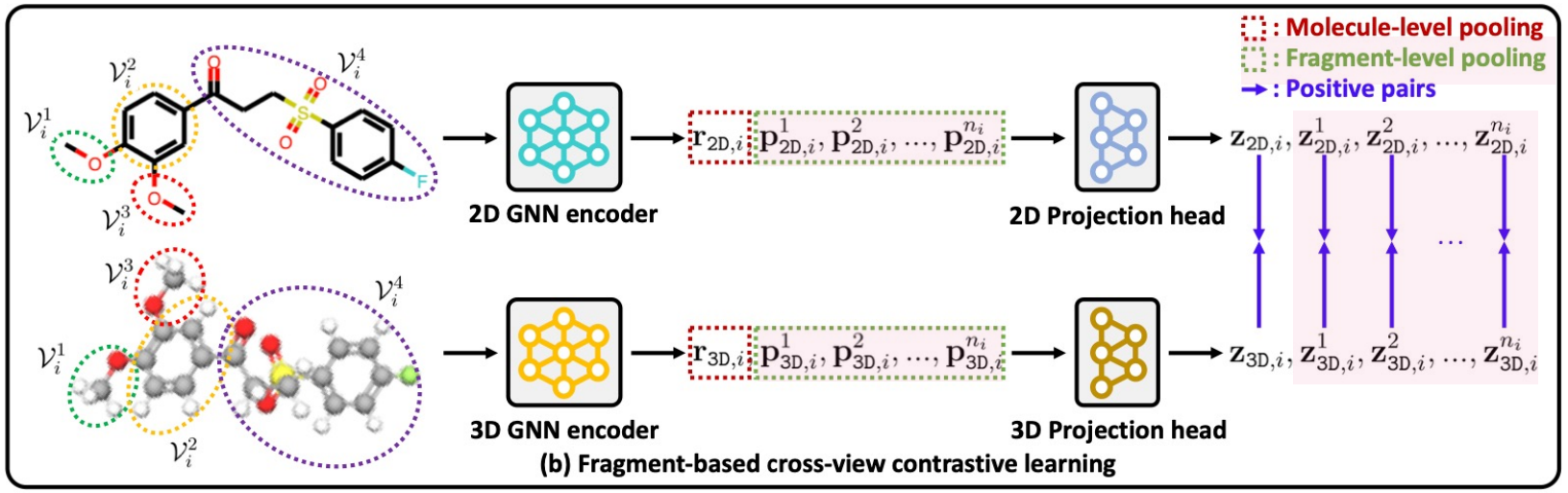
View construction strategy for cross-view molecules between 2D and 3D molecules: Fragment-level

- Positive views: Different view of the same fragment $(M^{j}_{i,\mathrm{2D}}, M^{j}_{i,\mathrm{3D}})$
- Negative views: Different view of fragments from different molecules $(M^{k}_{i,\mathrm{2D}}, M^{l}_{j,\mathrm{3D}})$ with $i \neq j$
- Where $M^{j}_{i,\mathrm{2D\ (or\ 3D)}}$ denotes the j-th fragment of i-the molecule in 2D (or 3D) view

$$\mathcal{L}_{\mathtt{cross,frag}} := -\frac{1}{2n}\sum_{i=1}^{n}\left(\log\frac{e^{s_{i,i}/\tau}}{e^{s_{i,i}/\tau}+\sum_{j\neq i}e^{s^{\mathrm{2D}}_{i,j}/\tau}}+\log\frac{e^{s_{i,i}/\tau}}{e^{s_{i,i}/\tau}+\sum_{j\neq i}e^{s^{\mathrm{3D}}_{i,j}/\tau}}\right)$$

$$s_{i,i} := \frac{1}{n_i}\sum_{k=1}^{n_i}\mathrm{sim}(\mathbf{z}^{k}_{\mathrm{2D},i}, \mathbf{z}^{k}_{\mathrm{3D},i})$$

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



Fragment-based Contrastive Learning (FragCL): Overall Framework

Additional objective to inject 3D information to 2D GNN: Torsional angle reconstruction

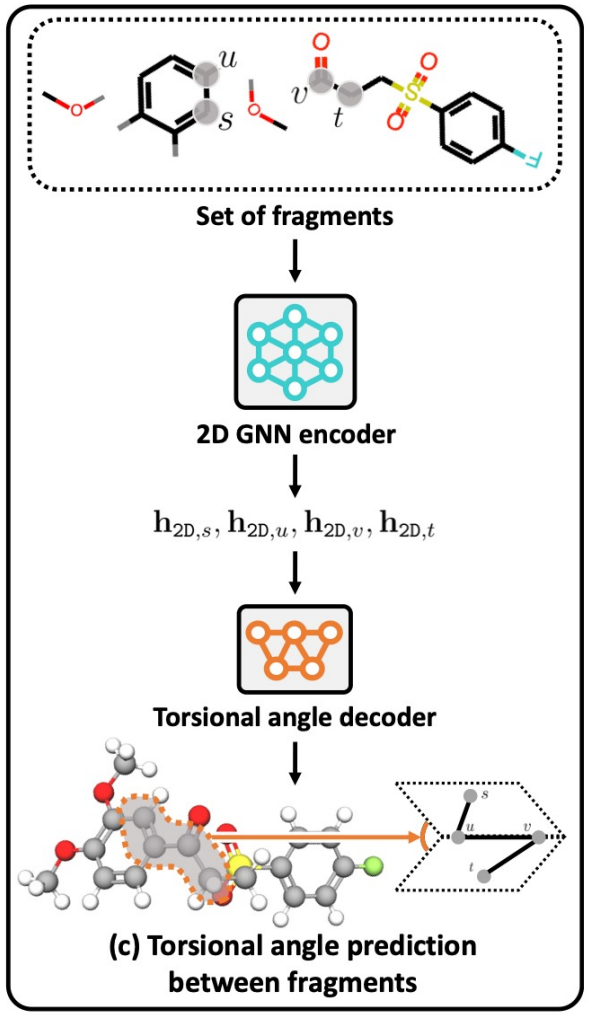
- Recap: Our main objective is to pretrain an effective 2D GNN
- 3D information (e.g., energy surface) is encoded in torsional angle
- Pretext task to predict the torsional angle between fragments $\{M^k, M^l\}$

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

Overall Framework

- Jointly train with proposed single-view, cross-view, and torsion loss

$$\mathcal{L}_{\texttt{FragCL}} \coloneqq \mathcal{L}_{\texttt{single}} + \mathcal{L}_{\texttt{cross}} + \mathcal{L}_{\texttt{tor}}$$



Experiments: FragCL is effective for various downstream tasks

We report the transfer-learning performance of pretrained 2D GNN on MoleculeNet downstream tasks

- We compare to 2D-only pretraining methods and 2D/3D multi-view pretraining methods.
- Best mean ROC-AUC score and scores within on standard deviation of the best mean score is marked bold.

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									
AttrMask	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
ContextPred	**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
G-Motif	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	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	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
MGSSL	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	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	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	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	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	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	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**

Experiments: FragCL is effective for various downstream tasks

We report the semi-supervised learning performance of pretrained 2D GNN on QM9 downstream tasks

- We compare to 2D/3D multi-view pretraining methods and the best MAE score is marked bold.
- Label fraction is the ratio of labels used for semi-supervised learning setup.

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	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	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**

Results with full label access (label fraction is 100%)

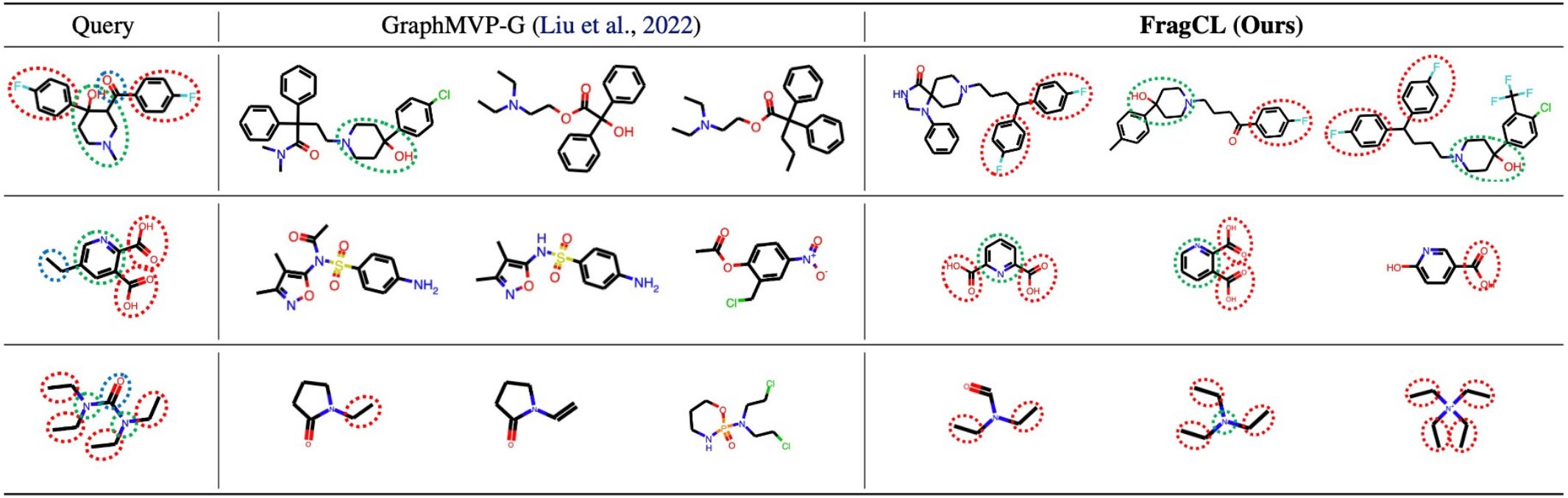
Methods	ZPVE ↓			LUMO ↓			HOMO ↓			U_0 ↓		
Label Fraction (%)	20	50	100	20	50	100	20	50	100	20	50	100
-	111.0	87.1	43.7	236.0	140.6	80.5	233.6	128.1	89.4	165.5	82.8	62.9
Pretrained on 110k 2D and 3D molecular graphs of QM9 and fine-tuned on 2D molecular graphs of QM9												
3D-Infomax	87.2	42.8	24.4	215.0	98.4	57.9	181.0	102.4	57.7	148.2	75.0	42.1
GraphMVP-G	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**

Results on various label fraction

Experiments: FragCL is effective for molecule-retrieval

FragCL framework can effectively recognize the fragment-wise structure of molecule

- For a query molecule, we report the top-3 similar molecules in the representation space.
- We mark the common fragments as the same-colored dotted lines.



Experiments: Ablation study on FragCL

Ablation study shows the effectiveness of each component in FragCL

- We show the efficacy of BRICS decomposition as a decomposition strategy to obtain positive view.
- Each multi-view interaction has its own benefits to improve overall performance.

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**

Ablation study on positive view construction & fragmentation strategy

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)	✓	-	-	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
	✓	✓	-	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
	✓	✓	✓	**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**

Albation study on multi-view interactions

FragCL: Simple & Effective Framework for Molecular Pretraining

Summary: We propose a simple yet effective framework for molecular contrastive learning.

We propose FragCL = Fragment-based multi-view Contrastive Learning for molecular self-supervised learning

1. Construct fragment-based positive/negative views for molecular contrastive learning
2. Transfer learning and semi supervised learning: Utilize unlabeled molecule to find good initialization of GNN
3. Molecule retrieval: Can be used to find semantically similar molecules