# Supplementary Information: An Image-enhanced Molecular Graph Representation Learning Framework

**Hongxin Xiang**[1,2], **Shuting Jin**[3], **Jun Xia**[4], **Man Zhou**[5], **Jianmin Wang**[6], **Li Zeng**[2], **Xiangxiang Zeng**[1,*]

[1]College of Computer Science and Electronic Engineering, Hunan University, Changsha, China
[2]Department of AIDD, Shanghai Yuyao Biotechnology Co., Ltd., Shanghai, China
[3]School of Computer Science & Technology, Wuhan University of Science and Technology, Wuhan, China
[4]School of Engineering, Westlake University, Hangzhou, China
[5]University of Science and Technology of China, Hefei, China
[6]The Interdisciplinary Graduate Program in Integrative Biotechnology, Yonsei University, Incheon, Korea
Correspondding author: xzeng@hnu.edu.cn

## A Pearson correlation between different modalities

In order to study the correlation between different modal data, we choose 2D graph, 3D graph (with conformation), 2D image (rendered by RDKit) and the proposed multi-view 3D image. Subsequently, we use EGNN for feature extraction of 2D graph and 3D graph and ResNet18 for feature extraction of 2D image and 3D image. These model is trained from scratch. Finally, we utilize atom predictor, bound predictor, geometry predictor and property predictor to predict on atoms $S^{atom}$, bonds $S^{bound}$, geometry $S^{geom}$, and chemical properties $S^{prop}$ (See Appendix F). The experimental data selects the first 10,000 items in the pre-training dataset. We split the training set, validation set, and test set in a ratio of 8:1:1 and select the best model based on the validation set. All runs used identical experimental settings. We use a batch size of 8, a learning rate of 0.005, a 100-dimensional hidden layer and perform 30 epochs for training.

As shown in Figure S1, the Pearson correlation coefficient of the prediction results of different models is shown. We use this coefficient to reflect the differences between different models. We find that EGNN has a significantly high correlation in the prediction of 2D graph and 3D graph, reaching 87%, indicating that there is a large amount of redundancy in the information between 2D graph and 3D graph, resulting in little cross-modal information compensation. In addition, We also find that ResNet(2D) and ResNet(3D) do not have that high similarity, only 19%, which shows that the information of 3D images is better than that of 2D images (the performance advantages in Table S1 can illustrate this point). It is worth noting that the Pearson correlation of 2D images and 3D images (19%) is still higher than that of EGNN(2D) with 10% correlation and EGNN(3D) with 10% correlation, indicating that the correlation between images is higher than the correlation between graphs. This inspires us to use more differentiated modalities to enhance the characterization of molecules.

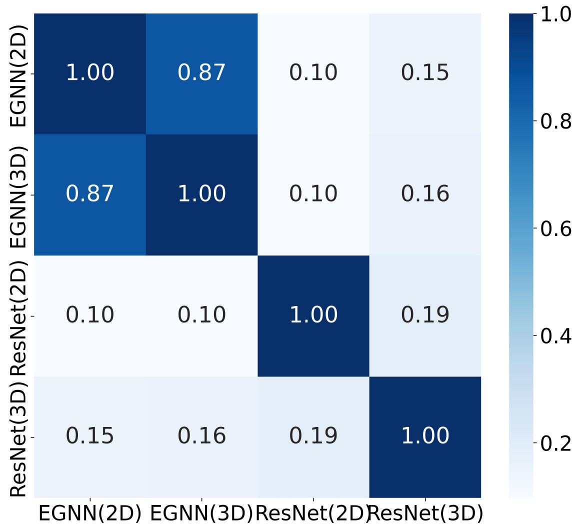


Figure S1: Pearson correlation coefficient between different models on 4 basic prior knowledge prediction task.

## B Prediction ability of fundamental prior knowledge

Table S1: The RMSE performance of different models on atoms $S^{atom}$, bonds $S^{bound}$, geometry $S^{geom}$, and chemical properties $S^{prop}$. The **Bold** indicates the best result and the underline indicates the second best result. $\Delta$ represents the relative performance improvement of ResNet (3D) compared to the second place.

| | use conformation? | geomery $S^{geom}$ | atom $S^{atom}$ | bound $S^{bound}$ | property $S^{prop}$ |
|---|---|---|---|---|---|
| GCN | × | 0.27359 | 1.276 | <u>3.109</u> | 62.304 |
| GIN | × | 0.27410 | 1.312 | 3.144 | 62.980 |
| EGNN (2D) | × | <u>0.26857</u> | 0.576 | 3.481 | 17.418 |
| EGNN (3D) | ✓ | **0.26856** | 0.579 | 3.481 | 16.684 |
| schnet (3D) | ✓ | 0.26869 | 0.785 | 3.612 | - |
| ResNet (2D) | × | 0.26875 | <u>0.496</u> | 3.119 | <u>12.469</u> |
| ResNet (3D) | ✓ | 0.26868 | **0.424** | **3.042** | **8.999** |
| $\Delta$ | - | $\downarrow$ 0.04% | $\uparrow$ 16.98% | $\uparrow$ 2.16% | $\uparrow$ 27.83% |

Here, we use the experimental settings in Appendix A and report the performance of different models on four types of prior knowledge ($S^{atom}$, bonds $S^{bound}$, geometry $S^{geom}$, and chemical properties $S^{prop}$). As shown in Table S1, we find that the proposed 3D image achieves the best RMSE performance with an maximum relative performance improvement of 27.83%, indicating that the proposed molecular representation has good discriminative ability with respect to the fundamental knowledge of molecules.

# C The visualization of molecular images

Here, we visualize molecular images obtained by three ways in Figure S2, including canvas-based technology, 3D CAD modeling technology, and physical microscopy-based technology.

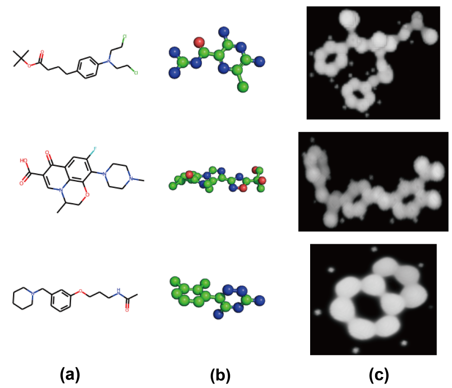


Figure S2: Three molecular rendering technologies. (a) Canvas-based technology. (b) 3D CAD modeling technology. (c) Physical microscopy-based technology, in which shows images from Cryo-EM.

# D The overall process of IEM

In order to clearly describe the proposed IEM framework, we show the overal process in Algorithm 2.

**Algorithm 1** The overall process of image-enhanced molecular graph representation learning framework (IEM).

**Data:** The molecular graphs $\mathcal{G}$, corresponding ground-truth labels $y$; the 2D images $\mathcal{V}^{2D}$ and the set of multi-view 3D images $\mathcal{V}^{3D}$, corresponding set of four priors of molecule $\mathcal{S} = (\mathcal{S}^{atom}, \mathcal{S}^{bound}, \mathcal{S}^{geom}, \mathcal{S}^{prop})$.

**Stage I: Pre-training teacher:** The single-view images $\mathcal{V}^{2D}$ and multi-view images $\mathcal{V}^{3D}$ are input into the 3D image encoder and 2D image encoder to extract features $\mathcal{F}^{2D}$ and $\mathcal{F}^{3D}$, respectively. Subsequently, $\mathcal{F}^{2D}$ and $\mathcal{F}^{3D}$ are aligned through unsupervised contrastive learning and forward propagated into 4 predictors to get 4 prediction logits $\mathrm{P}^{atom}(\mathcal{F}^{2D/3D})$, $\mathrm{P}^{bound}(\mathcal{F}^{2D/3D})$, $\mathrm{P}^{geom}(\mathcal{F}^{2D/3D})$ and $\mathrm{P}^{prop}(\mathcal{F}^{2D/3D})$ supervised by $\mathcal{S}$. After training the teacher, freeze the 2D/3D image encoder and these 4 predictors.

**Stage II: Image-enhanced distillation strategy:** The multi-view images and graphs are input to the frozen 3D image encoder (teacher) and GNN (student) to extract features $\mathcal{F}^{3D}$ and $\mathcal{F}^{g}$, respectively. These features are then forward-propagated into knowledge enhancer and task enhancer to obtain prediction logits $\mathcal{E}^{3D} = \{\mathrm{Enh}^{k}(\mathcal{F}^{3D}), \mathrm{Enh}^{t}(\mathcal{F}^{3D})\}$ and $\mathcal{E}^{g} = \{\mathrm{Enh}^{k}(\mathcal{F}^{g}), \mathrm{Enh}^{t}(\mathcal{F}^{g})\}$. The image-enhanced graph representations can be obtained through images as a supervision signals $\mathcal{E}^{3D}$ to guide $\mathcal{E}^{g}$.

**Stage III: Training and Inference:** During training, we utilize the 3D image encoder (teacher) and GNN (student) to extract features and update the parameters of the GNN using the loss obtained from the knowledge enhancer and the task enhancer. During inference, we only input graphs into GNN and get the corresponding prediction results.

# E Image rendering details

We use RDKit [Landrum, 2013] and PyMol [DeLano and others, 2002] to render 2D images and 3D multi-view images of molecules respectively. In detail, we render a 2D image for the molecule by calling *MolsToGridImage()* method in RDKit. To render 3D multi-view images, we use PyMol to execute these commands (*bg_color white;set stick_ball,on;set stick_ball_ratio,3.5;set stick_radius,0.15;set sphere_scale,0.2;set valence,1;set valence_mode,0; set valence_size, 0.1*) to render molecules in stick-ball mode and use these commands (*rotate x,0;rotate x,180;rotate y,180;rotate z,180*) to get 4 molecular images of different views respectively, where "rotate x,180" means rotating the image 180 degrees along the x-axis.

# F Fundamental prior knowledge used in pretraining teacher

Table S2: Details of 8 chemical properties.

| No. | Attributes | Description |
|---|---|---|
| 1 | molecular weight | the weight of molecule |
| 2 | MolLogP | Wildman-Crippen LogP value |
| 3 | MolMR | Wildman-Crippen MR value |
| 4 | BalabanJ | Balaban's J value for a molecule |
| 5 | NumHAcceptors | Number of Hydrogen Bond Acceptors |
| 6 | NumHDonors | Number of Hydrogen Bond Donors |
| 7 | NumValenceElectrons | The number of valence electrons the molecule has |
| 8 | TPSA | TPSA / total molecular surface area |

We carefully design fundamental prior knowledge inspired by the following two criteria: (1) Rich knowledge can pre-train an excellent teacher; (2) Knowledge with strong applicability can play a role in uncertain downstream tasks. Therefore, we consider the atoms $S^{atom}$, bonds $S^{bound}$, geometry $S^{geom}$, and chemical properties $S^{prop}$ of molecules.

- **Atom knowledge** $S^{atom} \in \mathbb{R}^{n^{atom}}$ counts the chemical element distribution of 19 types of atoms in molecules, including {C, N, O, F, S, Cl, Br, P, Si, B, Se, Ge, As, H, Ti, Ga, Ca, Mg, Zn}, where $n^{atom} = 19$.
- **Bound knowledge** $S^{bound} \in \mathbb{R}^{n^{bound}}$ counts the distribution of 4 types of bounds in molecules, including {single bound, aromatic bound, double bound, triple bound}, where $n^{bound} = 4$.
- **Geometry knowledge** $S^{geom} \in \mathbb{R}^{n^{geom}}$ counts the geometry distribution in molecules. In detail, given a molecule with $n$ atoms, we extract the 3D coordinates of each atom and normalize them. Then, we flatten these normalized three-dimensional coordinates into a one-dimensional vector of length $n \times 3$. Since the number of atoms in each molecule varies, we set the maximum dimension of $S^{geom}$ to $n^{geom} = 60$. If the molecule is below this dimension, it is padded with 0, and if it is above this dimension, it is truncated.
- **Chemical properties knowledge** $S^{prop} \in \mathbb{R}^{n^{prop}}$ counts the property distribution in molecules. Different from the properties in downstream molecular property prediction tasks, the properties here are basic attributes possessed by every molecule. We used a total of 8 attributes, including {molecular weight, MolLogP, MolMR, BalabanJ, NumHAcceptors, NumHDonors, NumValenceElectrons, TPSA}. See Table S2 for details.

# G Training details of teacher model

Assuming there are $n$ molecules, we first generate 2D images $\mathcal{V}^{2D} \in \mathbb{R}^{n\times 224\times 224\times 3}$ and 3D multi-view images $\mathcal{V}^{3D} \in \mathbb{R}^{n\times 4\times 224\times 224\times 3}$ for these molecules respectively. Then, we input $\mathcal{V}^{2D}$ and $\mathcal{V}^{3D}$ into 2D encoder $\text{Enc}^{2D}$ and 3D encoder $\text{Enc}^{3D}$ to extract molecular representation $\mathcal{F}^{2D} = \text{Enc}^{2D}(\mathcal{V}^{2D}) \in \mathbb{R}^{n\times 512}$ and $\mathcal{F}^{3D} = \text{Enc}^{3D}(\mathcal{V}^{3D}) \in \mathbb{R}^{n\times 512}$, respectively. Here, we concated 2D and 3D molecular representation to obtain $\mathcal{F}^{I} = \{\mathcal{F}^{2D}, \mathcal{F}^{3D}\} \in \mathbb{R}^{2\times n\times 512}$. Here, we calculate the loss $\mathcal{L}_{ICL}$. Next, we input $\mathcal{F}^{2D}$ and $\mathcal{F}^{3D}$ into 4 predictors ($\text{P}^{atom}$, $\text{P}^{bound}$, $\text{P}^{geom}$, $\text{P}^{prop}$) to get the corresponding logits $\{p^{2D}_{atom}, p^{2D}_{bound}, p^{2D}_{geom}, p^{2D}_{prop}\}$ and $\{p^{3D}_{atom}, p^{3D}_{bound}, p^{3D}_{geom}, p^{3D}_{prop}\}$. The 2D loss $\mathcal{L}_{2D}$ and 3D loss $\mathcal{L}_{3D}$ can be calculated by using ground-truth labels ($\mathcal{S}^{atom}, \mathcal{S}^{bound}, \mathcal{S}^{geom}, \mathcal{S}^{prop}$), repespectively. Finally, We perform backpropagation via $\mathcal{L}^{Teacher}_{Pretrain} = \mathcal{L}_{ICL} + \mathcal{L}_{2D} + \mathcal{L}_{3D}$. See Algorithm 2 for details of the pseudo-code.

# H Proof about the lower bound of the information increment $\mathcal{I}_{diff}$

We describe in detail the theoretical proof about the lower bound $\Omega$ of the information increment $\mathcal{I}_{diff}$. We define information increment as the increase in useful information of one feature compared to another feature. Let $\mathcal{I}_{diff} = \mathcal{I}^{IE} - \mathcal{I}^{g}$, which represents the knowledge gain after image guidance. Since the image-enhanced graph features $\mathcal{F}^{IE}$ are related to the graph encoder $\text{Enc}^{g}$ and the knowledge from the image teacher. Therefore, the information amount of $\mathcal{F}^{IE}$ can be formalized as $\mathcal{I}^{IE} = \mathcal{I}(\mathcal{F}^{IE}|\mathcal{V}, y;\text{Enc}^{g}, \gamma)$ given images $\mathcal{V}$ and the corresponding ground-truths label $y$. In the same way, the information amount of $\mathcal{F}^{g}$ only from the student can be expressed as $\mathcal{I}^{g} = \mathcal{I}(\mathcal{F}^{g}|\mathcal{G}, y;\text{Enc}^{g})$. In order to find the lower bound of $I_{diff}$, We take the 3D image encoder $\text{Enc}^{3D}$ as the teacher as an example and make the following derivation:

$$\mathcal{I}_{diff} = \mathcal{I}(\mathcal{F}^{3D}|\mathcal{V}, y; \text{Enc}^{3D}) - \mathcal{I}(\mathcal{F}^{g}|\mathcal{G}, y; \text{Enc}^{g}) + \mathcal{I}(\mathcal{F}^{IE}|\mathcal{G}, y; \text{Enc}^{g}, \gamma) - \mathcal{I}(\mathcal{F}^{3D}|\mathcal{V}, y; \text{Enc}^{3D}) \quad (1)$$

**Algorithm 2** The pseudo-code for training teacher model.

**Input:** $\text{Enc}^{2D}$, $\text{Enc}^{3D}$, $\text{P}^{atom}$, $\text{P}^{bound}$, $\text{P}^{geom}$, $\text{P}^{prop}$
**for** sampled minibatch $\mathcal{V}^{2D}$, $\mathcal{V}^{3D}$, $\mathcal{S}^{atom}$, $\mathcal{S}^{bound}$, $\mathcal{S}^{geom}$, $\mathcal{S}^{prop}$ **do**
$\mathcal{F}^{2D}, \mathcal{F}^{3D} = \text{Enc}^{2D}(\mathcal{V}^{2D}), \text{Enc}^{3D}(\mathcal{V}^{3D})$
$\mathcal{F}^{I} = \text{concat}(\mathcal{F}^{2D}, \mathcal{F}^{3D})$
$p^{2D}_{atom}, p^{2D}_{bound}, p^{2D}_{geom}, p^{2D}_{prop}$=$\text{P}^{atom}(\mathcal{F}^{2D}), \text{P}^{bound}(\mathcal{F}^{2D}), \text{P}^{geom}(\mathcal{F}^{2D}), \text{P}^{prop}(\mathcal{F}^{3D})$
$p^{3D}_{atom}, p^{3D}_{bound}, p^{3D}_{geom}, p^{3D}_{prop}$= $\text{P}^{atom}(\mathcal{F}^{3D}), \text{P}^{bound}(\mathcal{F}^{3D}), \text{P}^{geom}(\mathcal{F}^{3D}), \text{P}^{prop}(\mathcal{F}^{3D})$
$\text{Item}_{2D}$= $\{p^{2D}_{atom}, p^{2D}_{bound}, p^{2D}_{geom}, p^{2D}_{prop}, \mathcal{S}^{atom}, \mathcal{S}^{bound}, \mathcal{S}^{geom}, \mathcal{S}^{prop}\}$
$\text{Item}_{3D}$= $\{p^{3D}_{atom}, p^{3D}_{bound}, p^{3D}_{geom}, p^{3D}_{prop}, \mathcal{S}^{atom}, \mathcal{S}^{bound}, \mathcal{S}^{geom}, \mathcal{S}^{prop}\}$
$L^{Teacher}_{Pretrain} = L_{ICL}(\mathcal{F}^{I}) + L_{2D}(\text{Item}_{2D}) + L_{3D}(\text{Item}_{3D})$
update networks $\text{Enc}^{2D}$, $\text{Enc}^{3D}$, $\text{P}^{atom}$, $\text{P}^{bound}$, $\text{P}^{geom}$, $\text{P}^{prop}$ to minimize $L^{Teacher}_{Pretrain}$
**end for**
**return** $\text{Enc}^{2D}$, $\text{Enc}^{3D}$, $\text{P}^{atom}$, $\text{P}^{bound}$, $\text{P}^{geom}$, $\text{P}^{prop}$

where $\mathcal{I}(\mathcal{F}^{3D}|\mathcal{V}, y; \text{Enc}^{3D})$ is a constant because the parameters of $\text{Enc}^{3D}$ are freezed in the knowledge distillation stage. $\mathcal{I}(\mathcal{F}^{IE}|\mathcal{V}, y; \text{Enc}^{g}, \gamma)$ is an image-guided graph-based feature and it should be greater than 0 when negative transfer problem does not occur. Therefore, we can get the following inequality:

$$\mathcal{I}_{diff} \geq \mathcal{I}(\mathcal{F}^{3D}|\mathcal{V}, y; \text{Enc}^{3D}) - \mathcal{I}(\mathcal{F}^{g}|\mathcal{G}, y; \text{Enc}^{g}) \tag{2}$$

where $\mathcal{I}(\mathcal{F}^{3D}|\mathcal{V}, y; \text{Enc}^{3D}) - \mathcal{I}(\mathcal{F}^{g}|\mathcal{G}, y; \text{Enc}^{g})$ represents the information increment of the teacher relative to the student. Obviously, $\mathcal{I}_{Diff} \geq 0$ holds when the teacher is no worse than the student. Therefore, this provides a theoretical basis and inspires us to consider knowledgeable teachers and distillation strategies that are resistant to negative transfer in cross-modal knowledge distillation.

# I Datasets of downstream tasks

The ESOL, Lipo, Malaria [Gamo *et al.*, 2010] and CEP [Hachmann *et al.*, 2011] datasets are from GraphMVP [Liu *et al.*, 2021], other datasets are from MoleculeNet [Wu *et al.*, 2018]. The details of these datasets are listed below:

- BBBP (Blood-Brain Barrier Penetration) records the relationship between drugs and barrier permeability, which is important for assessing whether drugs are blocked the blood-brain barrier.
- Tox21 (Toxicology in the 21st Century) contains qualitative toxicity measurements of compounds on 12 different targets, including nuclear receptors and stress response pathways.
- ClinTox is a compound toxicity dataset that includes FDA-approved drugs and drugs that failed clinical trials for toxicity reasons.
- BACE (Beta-secretase) provides qualitative (binary marker) results of compounds on human $\beta$-secretase 1 (BACE-1) inhibitors.
- Sider (Side Effect Resource) is a dataset of marketed drugs and adverse drug reactions (ADRs) with 27 system organ classes.
- ToxCast is an extension of the Tox21 dataset, including toxicology data from a large compound library based on in vitro high-throughput screening.
- HIV provides compounds that inhibit HIV replication, which are derived the Drug Therapeutics Program (DTP) AIDS Antiviral Screen.
- ESOL records the water solubility of compounds, which is widely used to validate machine learning models that estimate solubility directly from molecular structure.
- Lipo (Lipophilicity) provides experimental results of a compound's octanol/water distribution coefficient (log$D$ at pH 7.4), which is an important characteristic of drug molecules that affects membrane permeability and solubility.
- Malaria measures the drug efficacy against the parasite that causes malaria.
- CEP dataset is a subset of the Havard Clean Energy Project (CEP).

# J Details of pretraining teacher

We summarize the training loss of the teacher in Figure S3, which were pretrained for approximately 450k steps. Obviously, the downward trend of all losses indicates that the teacher can learn the knowledge contained in the 5 pre-training tasks well.

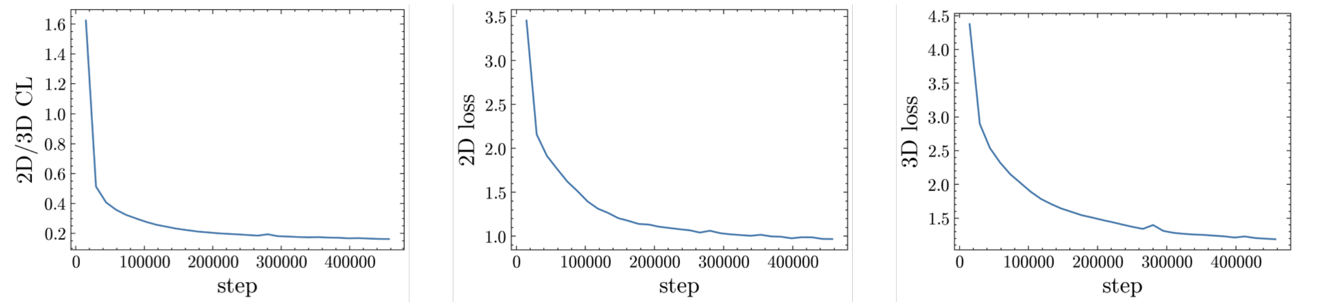


Figure S3: Training teacher loss details. 2D/3D CL represents the contrastive learning loss between 2D images and 3D images (Formula **??**). 2D loss and 3D loss respectively represent the loss of 2D image and 3D image on 4 prior knowledge (Formula **??** and Formula **??**).

## K Compare with more state-of-the-art methods

We provide more comparative methods in Table S3. Obviously, compared to these methods, our method achieves optimal performance.

Table S3: The ROC-AUC (%) performance of different methods on 8 classification datasets of molecular property prediction. We report the mean (standard deviation) ROC-AUC of 10 random seeds from 0 to 9 with scaffold splitting. The best and second best results are marked **bold** and underlined. IEM-baseline represents baseline equipped with IEM. $\Delta$ represents the absolute improvement percentage calculated by $\mathrm{AUC}_{\text{w/ IEM}} - \mathrm{AUC}_{\text{w/o IEM}}$.

| | Tox21 | ToxCast | Sider | ClinTox | MUV | HIV | BBBP | BACE | Average |
|---|---|---|---|---|---|---|---|---|---|
| #Molecules | 7831 | 8576 | 1427 | 1478 | 93087 | 41127 | 2039 | 1513 | - |
| #Task | 12 | 617 | 27 | 2 | 17 | 1 | 1 | 1 | - |
| InfoGraph [Sun *et al.*, 2020] | 73.3 (0.6) | 61.8(0.4) | 58.7(0.6) | 75.4(4.3) | 74.4(1.8) | 74.2(0.9) | 68.7(0.6) | 74.3(2.6) | 70.10 |
| GPT-GNN [Hu *et al.*, 2020b] | 74.9 (0.3) | 62.5(0.4) | 58.1(0.3) | 58.3(5.2) | 75.9(2.3) | 65.2(2.1) | 64.5(1.4) | 77.9(3.2) | 68.45 |
| ContextPred [Hu *et al.*, 2020a] | 73.6 (0.3) | 62.6(0.6) | 59.7(1.8) | 74.0(3.4) | 72.5(1.5) | 75.6(1.0) | 70.6(1.5) | 78.8(1.2) | 70.93 |
| GraphLoG [Xu *et al.*, 2021] | 75.0 (0.6) | 63.4(0.6) | 59.6(1.9) | 75.7(2.4) | 75.5(1.6) | 76.1(0.8) | 68.7(1.6) | 78.6(1.0) | 71.56 |
| G-Contextual [Rong *et al.*, 2020] | 75.0 (0.6) | 62.8(0.7) | 58.7(1.0) | 60.6(5.2) | 72.1(0.7) | 76.3(1.5) | 69.9(2.1) | 79.3(1.1) | 69.34 |
| G-Motif [Rong *et al.*, 2020] | 73.6 (0.7) | 62.3 (0.6) | 61.0(1.5) | 77.7(2.7) | 73.0(1.8) | 73.8(1.2) | 66.9(3.1) | 73.0(3.3) | 70.16 |
| AD-GCL [Suresh *et al.*, 2021] | 74.9 (0.4) | 63.4(0.7) | 61.5(0.9) | 77.2(2.7) | 76.3(1.4) | 76.7(1.2) | 70.7(0.3) | 76.6(1.5) | 72.16 |
| JOAO [You *et al.*, 2021] | 74.8(0.6) | 62.8(0.7) | 60.4(1.5) | 66.6(3.1) | 76.6(1.7) | 76.9(0.7) | 66.4(1.0) | 73.2(1.6) | 69.71 |
| SimGRACE [Xia *et al.*, 2022] | 74.4 (0.3) | 62.6(0.7) | 60.2 (0.9) | 75.5(2.0) | 75.4(1.3) | 75.0(0.6) | 71.2(1.1) | 74.9(2.0) | 71.15 |
| GraphCL [You *et al.*, 2020] | 75.1 (0.7) | 63.0(0.4) | 59.8(1.3) | 77.5(3.8) | 76.4(0.4) | 75.1(0.7) | 67.8(2.4) | 74.6(2.1) | 71.16 |
| GraphMAE [Hou *et al.*, 2022] | 75.2 (0.9) | 63.6(0.3) | 60.5(1.2) | 76.5(3.0) | 76.4(2.0) | 76.8(0.6) | 71.2(1.0) | 78.2(1.5) | 72.30 |
| 3D InfoMax [Stärk *et al.*, 2021] | 74.5(0.7) | 63.5(0.8) | 56.8(2.1) | 62.7(3.3) | 76.2(1.4) | 76.1(1.3) | 69.1(1.2) | 78.6(1.9) | 69.69 |
| MGSSL [Zhang *et al.*, 2021] | 75.2 (0.6) | 63.3 (0.5) | 61.6(1.0) | 77.1(4.5) | 77.6(0.4) | 75.8(0.4) | 68.8(0.6) | 78.8 (0.9) | 72.28 |
| AttrMask [Hu *et al.*, 2020a] | 75.1 (0.9) | 63.3 (0.6) | 60.5(0.9) | 73.5(4.3) | 75.8(1.0) | 75.3(1.5) | 65.2(1.4) | 77.8(1.8) | 70.81 |
| GIN [Xu *et al.*, 2018] | 74.3(0.9) | 61.5(0.8) | 57.3(1.2) | 57.2(4.1) | 71.6(2.8) | 75.2(2.0) | 66.7(1.8) | 69.6(5.5) | 66.68 |
| IEM-GIN | 74.5(0.4) | 62.5(0.8) | 59.1(1.7) | 62.6(4.1) | 77.7(2.9) | 77.9(1.3) | 69.3(1.9) | 77.7(3.5) | 70.16 |
| $\Delta$ | ↑ 0.2 | ↑ 1.0 | ↑ 1.8 | ↑ 5.4 | ↑ 6.1 | ↑ 2.7 | ↑ 2.6 | ↑ 8.1 | ↑ 3.5 |
| EdgePred [Hu *et al.*, 2020a] | 76.0(0.6) | 64.1(0.6) | 60.4(0.7) | 64.1(3.7) | 75.1(1.2) | 76.3(1.0) | 67.3(2.4) | 77.3(3.5) | 70.08 |
| IEM-EdgePred | 76.3(0.6) | 64.6(0.6) | 61.2(0.6) | 67.5(2.3) | 78.3(1.3) | 78.3(1.3) | 67.8(2.2) | **84.1(0.8)** | 72.26 |
| $\Delta$ | ↑ 0.3 | ↑ 0.5 | ↑ 0.8 | ↑ 3.4 | ↑ 3.2 | ↑ 2.0 | ↑ 0.5 | ↑ 6.8 | ↑ 2.2 |
| GraphMVP [Liu *et al.*, 2021] | 74.5(0.7) | 63.4(0.5) | 60.7(1.4) | 78.4(6.4) | 73.0(2.3) | 75.6(1.6) | 67.4(2.4) | 75.8(3.0) | 71.10 |
| IEM-GraphMVP | 75.9(0.7) | 64.4(0.6) | 61.9(1.7) | **80.8(3.1)** | 77.3(1.2) | **78.8(1.1)** | 68.7(1.0) | 83.3(1.4) | **73.89** |
| $\Delta$ | ↑ 1.4 | ↑ 1.0 | ↑ 1.2 | ↑ 2.4 | ↑ 4.3 | ↑ 3.2 | ↑ 1.3 | ↑ 7.5 | ↑ 2.8 |
| GraphMVP-C [Liu *et al.*, 2021] | 74.6(0.4) | 63.4(0.6) | 60.6(1.3) | 76.9(3.7) | 72.8(2.4) | 77.1(2.1) | 69.9(1.4) | 79.6(1.7) | 71.86 |
| IEM-GraphMVP-C | 75.6(0.6) | 64.8(0.5) | 62.0(0.9) | 79.2(2.9) | 77.0(1.7) | 78.2(1.0) | **71.4(1.4)** | 81.9(1.6) | 73.76 |
| $\Delta$ | ↑ 1.0 | ↑ 1.4 | ↑ 1.4 | ↑ 2.3 | ↑ 4.2 | ↑ 1.1 | ↑ 1.5 | ↑ 2.3 | ↑ 1.9 |
| Mole-BERT [Xia *et al.*, 2023] | 77.0(0.3) | 64.4(0.2) | 63.2(0.7) | 72.7(2.7) | 79.2(2.0) | 77.7(0.7) | 65.7(2.3) | 80.2(0.9) | 72.51 |
| IEM-Mole-BERT | **77.8(0.4)** | **65.6(0.3)** | **65.3(0.8)** | 72.2(1.4) | **79.7(1.8)** | **78.8(0.6)** | 68.1(1.0) | 83.0(0.9) | 73.81 |
| $\Delta$ | ↑ 0.8 | ↑ 1.2 | ↑ 2.1 | -0.5 | ↑ 0.5 | ↑ 1.1 | ↑ 2.4 | ↑ 2.8 | ↑ 1.3 |

## References


[DeLano and others, 2002] Warren L DeLano et al. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr*, 40(1):82–92, 2002.

[Gamo *et al.*, 2010] Francisco-Javier Gamo, Laura M Sanz, Jaume Vidal, Cristina De Cozar, Emilio Alvarez, Jose-Luis Lavandera, Dana E Vanderwall, Darren VS Green, Vinod Kumar, Samiul Hasan, et al. Thousands of chemical starting points for antimalarial lead identification. *Nature*, 465(7296):305–310, 2010.

[Hachmann *et al.*, 2011] Johannes Hachmann, Roberto Olivares-Amaya, Sule Atahan-Evrenk, Carlos Amador-Bedolla, Roel S Sánchez-Carrera, Aryeh Gold-Parker, Leslie Vogt, Anna M Brockway, and Alán Aspuru-Guzik. The harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. *The Journal of Physical Chemistry Letters*, 2(17):2241–2251, 2011.

[Hou *et al.*, 2022] Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. Graphmae: Self-supervised masked graph autoencoders. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 594–604, 2022.

[Hu *et al.*, 2020a] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. *ICLR*, 2020.

[Hu *et al.*, 2020b] 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*, pages 1857–1867, 2020.

[Landrum, 2013] Greg Landrum. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling, 2013.

[Liu *et al.*, 2021] Shengchao Liu, Hanchen Wang, Weiyang Liu, Joan Lasenby, Hongyu Guo, and Jian Tang. Pre-training molecular graph representation with 3d geometry. *arXiv preprint arXiv:2110.07728*, 2021.

[Rong *et al.*, 2020] Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. In *NeurIPS*, 2020.

[Stärk *et al.*, 2021] Hannes Stärk, Dominique Beaini, Gabriele Corso, Prudencio Tossou, Christian Dallago, Stephan Günnemann, and Pietro Lio. 3d infomax improves gnns for molecular property prediction. In *NeurIPS 2021 AI for Science Workshop*, 2021.

[Sun *et al.*, 2020] Fan-Yun Sun, Jordon Hoffman, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. In *International Conference on Learning Representations*, 2020.

[Suresh *et al.*, 2021] Susheel Suresh, Pan Li, Cong Hao, and Jennifer Neville. Adversarial graph augmentation to improve graph contrastive learning. *Advances in Neural Information Processing Systems*, 34:15920–15933, 2021.

[Wu *et al.*, 2018] 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.

[Xia *et al.*, 2022] Jun Xia, Lirong Wu, Jintao Chen, Bozhen Hu, and Stan Z Li. Simgrace: A simple framework for graph contrastive learning without data augmentation. In *Proceedings of the ACM Web Conference 2022*, pages 1070–1079, 2022.

[Xia *et al.*, 2023] 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.

[Xu *et al.*, 2018] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2018.

[Xu *et al.*, 2021] Minghao Xu, Hang Wang, Bingbing Ni, Hongyu Guo, and Jian Tang. Self-supervised graph-level representation learning with local and global structure. *arXiv preprint arXiv:2106.04113*, 2021.

[You *et al.*, 2020] 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.

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

[Zhang *et al.*, 2021] 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.