IJCAI JEJU 2024

International Joint Conference
on Artificial Intelligence
Jeju 03.08.24 - 09.08.24

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,*]

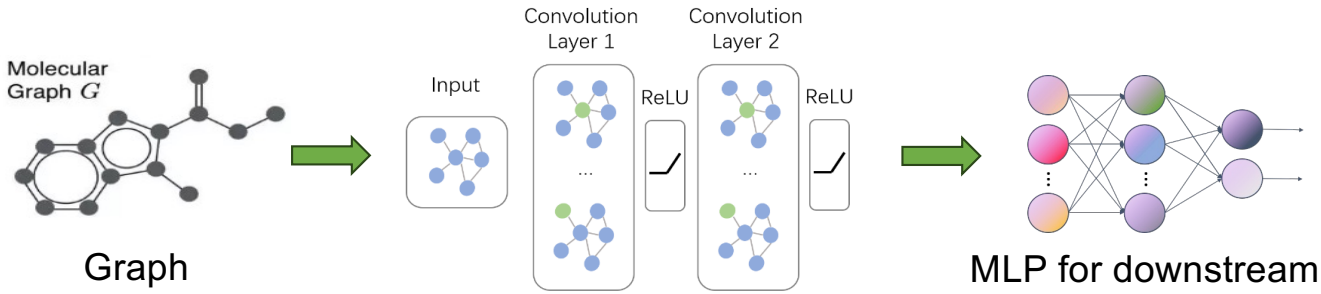
Presenter: Hongxin Xiang

[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
[*] Corresponding author

Paper ID: 3261

Challenges in Molecular Representation Learning

Limited by a single modality: The paradigm of learning from a single modality (e.g., molecular graph) gradually encounters the bottleneck of limited representation capabilities.

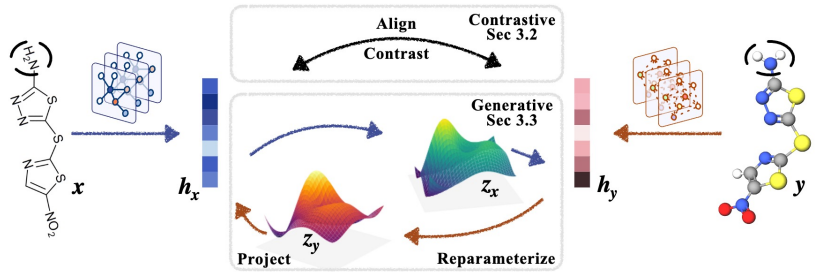


Graph

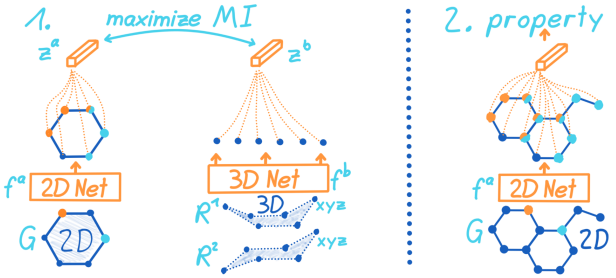
GNN or its variants

MLP for downstream prediction

Therefore, some researchers proposed multi-modality learning methods between 2D graph and 3D graph to generalize molecular representation, such as GraphMVP and 3D Infomax.



GraphMVP[1]



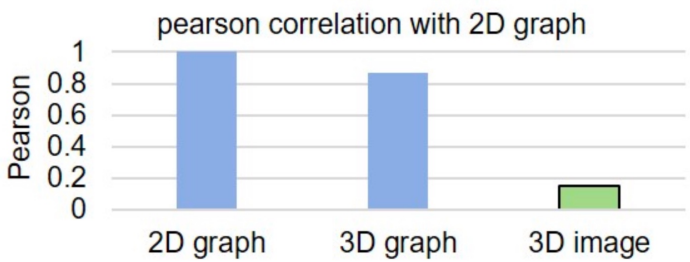
3D Infomax[2]

[1] Liu S, Wang H, Liu W, et al. Pre-training Molecular Graph Representation with 3D Geometry[C]//International Conference on Learning Representations.
[2] Stärk H, Beaini D, Corso G, et al. 3d infomax improves gnns for molecular property prediction[C]//International Conference on Machine Learning. PMLR, 2022: 20479-20502.

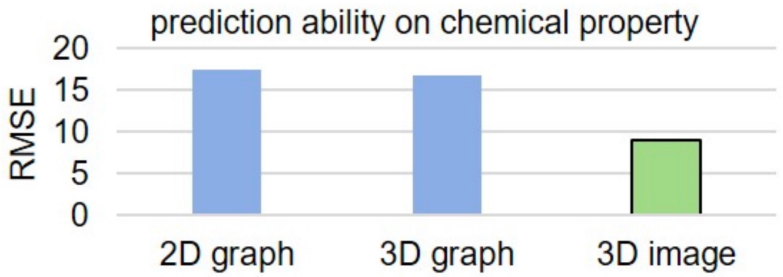
Challenges in Molecular Representation Learning

Multimodal fusion has limited improvements:

- **Similar modalities and encoding ways**: As shown in the left figure below, the 3D graph and the 2D graph has a high Pearson similarity.
- **Weak feature extraction ability**, resulting in insufficient complementary information between modalities. As shown in the right figure below, 2D graph and 3D graph are limited in understanding the 8 basic attributes of molecules (such as molecular weight, LogP, etc.).



Correlation coefficients between different molecular representations and 2D graphs



Average RMSE performance of different molecular representations on 8 basic chemical attributes

We find that:

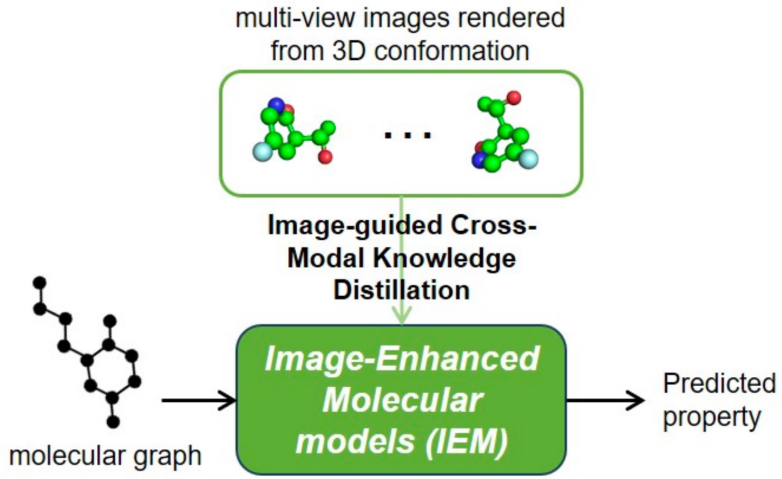
- **3D image has high Pearson similarity with 2D graph**, indicating that 3D image can provide more information to 2D graph;
- **3D image achieves good performance on 8 basic attributes of molecules**, which can help 2D graph to better understand the basic attributes of molecules.

Challenges in Molecular Representation Learning

Exploiting the rich information in molecular images to enhance representation learning of molecular graphs:

✗ **Multi-modal fusion learning between graph and image**: requires additional computational costs in the training and inference stages

✓ **Knowledge distillation:** describe the process of information transfer as how to use a knowledgeable teacher (image) to teach an excellent student (graph), which only introduces the prior of the image into the graph-based model during the training phase without modifying any baseline model.

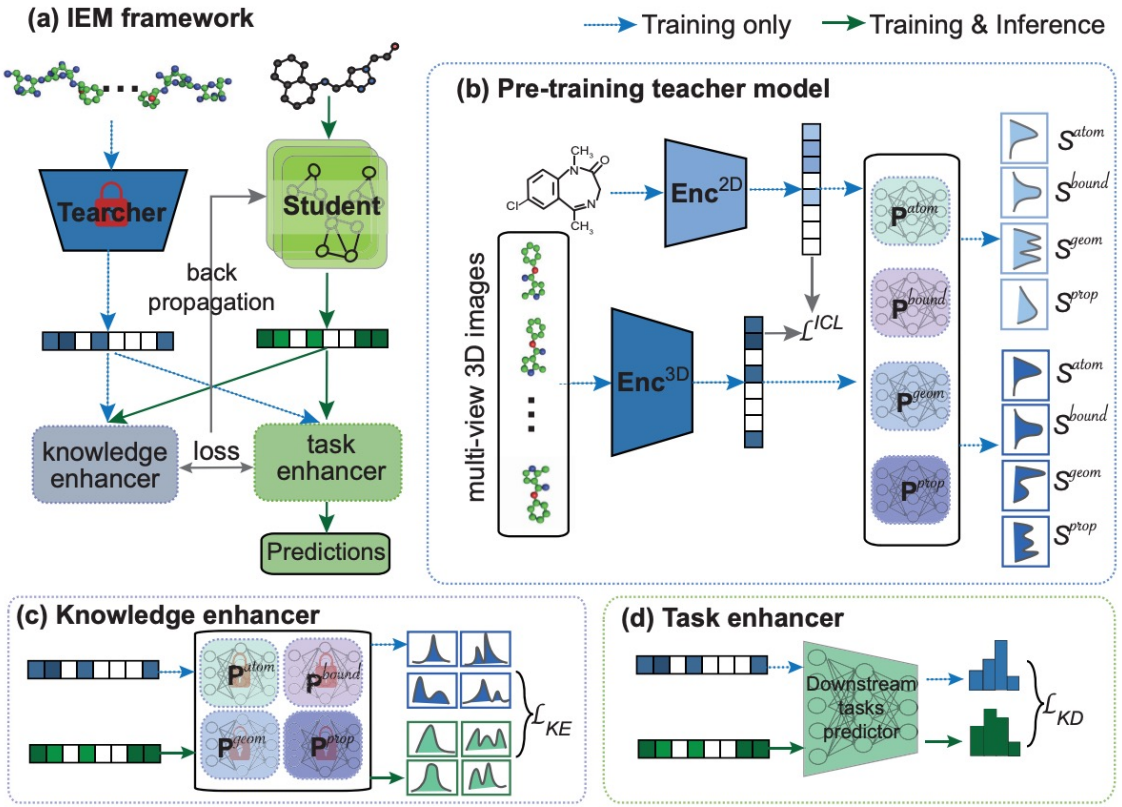


An Image-enhanced Molecular Graph Representation Learning Framework (IEM)

Therefore, we proposed an **Image-enhanced Molecular Graph Representation Learning Framework**, called IEM.

There are two key design principles:

- Knowledgeable teacher model.
- Effective distillation strategy.



An Image-enhanced Molecular Graph Representation Learning Framework (IEM)

Knowledgeable molecular image-based teacher model

We considered four different types of knowledge, as follows:

- **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.

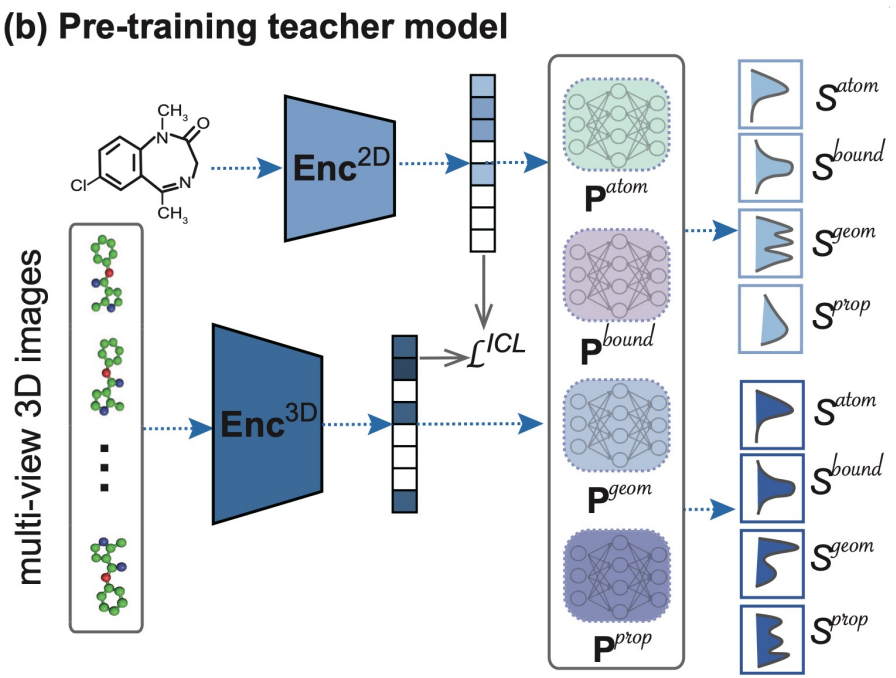
An Image-enhanced Molecular Graph Representation Learning Framework (IEM)

Knowledgeable molecular image-based teacher model

Pretraining data: 2 million molecular conformations.

5 pre-training strategies to enhance the representational power of the teacher model:

- ICL: contrastive learning between 2D and 3D images
- ADP: atom distribution pre- diction task
- BDP: bound distribution prediction task
- GDP: geometry distribution prediction task
- PDP: property distribution prediction task

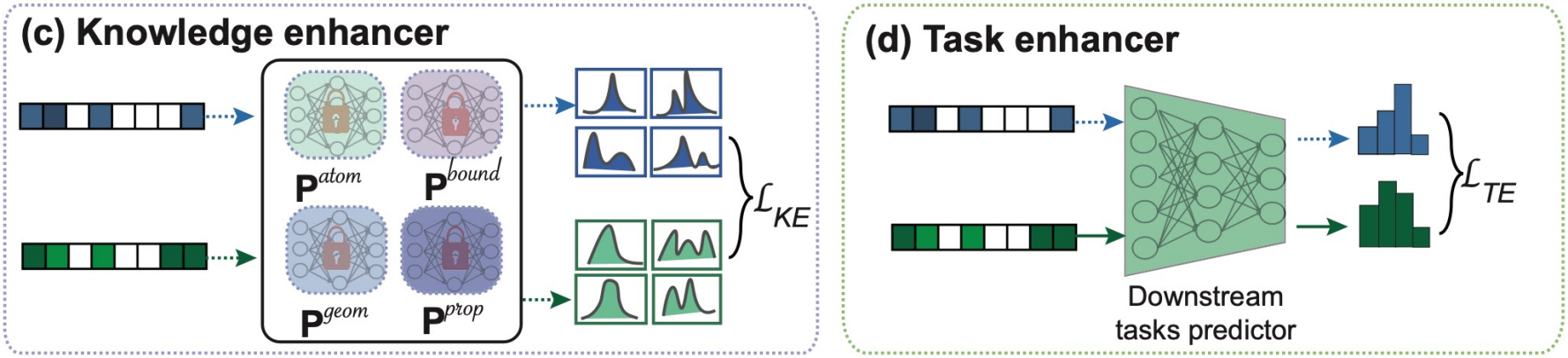


An Image-enhanced Molecular Graph Representation Learning Framework (IEM)

Effective distillation strategy

2 enhancers to align graph and image in logit space to avoid modality gaps in feature space:

- **Knowledge enhancer** is used to transfer 4 basic knowledge (atom, bound, geometry, and chemical property) from image to graph.
- **Task enhancer** is used to transfer knowledge related to downstream tasks from images to graph.



An Image-enhanced Molecular Graph Representation Learning Framework (IEM)

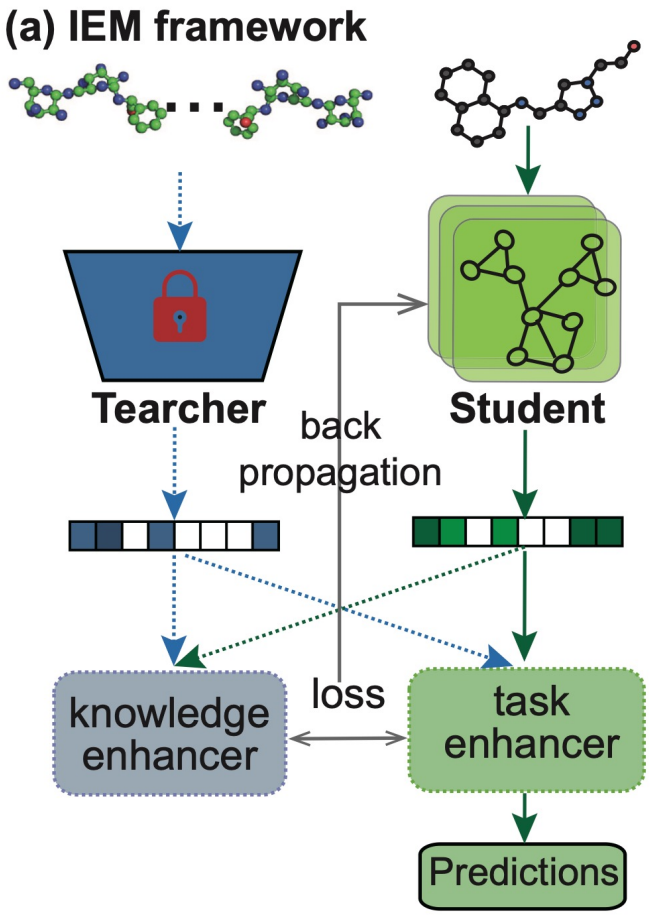
Training and inference

In the knowledge distillation stage, we freeze the teacher model and use the knowledge enhancer and task enhancer to transfer the teacher's knowledge to the student. In the subsequent reasoning stage, we no longer need the teacher model, only the student model is needed to complete the reasoning.

The loss function of distillation stage:

$$\mathcal{L}_{total} = \lambda_{KE}\mathcal{L}_{KE} + \lambda_{TE}\mathcal{L}_{TE} + \mathcal{L}_{T}$$

- $\mathcal{L}_{KE}$ is distillation loss from knowledge enhancer.
- $\mathcal{L}_{\mathrm{TE}}$ is distillation loss from task enhancer.
- $\mathcal{L}_{\mathrm{T}}$ is the loss from the student model on the downstream task, such as the cross entropy loss.
- λ_{KE} and λ_{TE} are the balance coefficient.



Datasets and Settings

- **Datasets:** 8 classification datasets and 4 regression datasets from molecular property prediction task.
- **Splitting:** All datasets are divided into training set, validation set and test set according to 8:1:1 with scaffold split.
- **Evaluation Metric:** ROC-AUC for classification tasks and RMSE for regression tasks.
- We report the mean (standard deviation) performance of 10 random seeds from 0 to 9.

Results

The performance on 8 classification datasets and 4 regression datasets.

- IEM consistently improves performance across all baselines.

	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	-
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
Δ	↑ 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
Δ	↑ 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**
Δ	↑ 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
Δ	↑ 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
Δ	↑ 0.8	↑ 1.2	↑ 2.1	-0.5	↑ 0.5	↑ 1.1	↑ 2.4	↑ 2.8	↑ 1.3

Table 1: 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. Δ represents the absolute improvement percentage calculated by $\text{AUC}_{\text{w/ IEM}} - \text{AUC}_{\text{w/o IEM}}$.

	ESOL	Lipo	Malaria	CEP
#Molecules	1,128	4,200	9,999	29,978
#Task	1	1	1	1
GIN	1.472(0.038)	0.832(0.025)	1.113(0.011)	1.340(0.018)
IEM-GIN	1.346(0.045)	0.817(0.019)	1.084(0.003)	1.329(0.021)
Δ	↑ 8.56%	↑ 1.80%	↑ 2.61%	↑ 0.82%
EdgePred	1.367(0.041)	0.778(0.013)	1.110(0.011)	1.362(0.025)
IEM-EdgePred	1.350(0.027)	0.769(0.006)	1.088(0.005)	1.345(0.016)
Δ	↑ 1.24%	↑ 1.16%	↑ 1.98%	↑ 1.25%
GraphMVP	1.322(0.062)	0.773(0.016)	1.128(0.019)	1.308(0.024)
IEM-GraphMVP	1.281(0.044)	0.754(0.015)	1.089(0.005)	**1.294**(0.020)
Δ	↑ 3.10%	↑ 2.46%	↑ 3.46%	↑ 1.07%
GraphMVP-C	1.333(0.055)	0.768(0.013)	1.114(0.008)	1.304(0.020)
IEM-GraphMVP-C	1.274(0.037)	0.761(0.017)	1.090(0.004)	1.296(0.012)
Δ	↑ 4.43%	↑ 0.91%	↑ 2.15%	↑ 0.61%
MoleBERT	1.115(0.017)	0.727(0.006)	1.137(0.021)	1.350(0.015)
IEM-MoleBERT	1.090(0.031)	**0.716**(0.003)	**1.080**(0.003)	1.343(0.013)
Δ	↑ 2.24%	↑ 1.51%	↑ 5.01%	↑ 0.52%
GraphMVP-F	1.094(0.037)	0.724(0.009)	1.106(0.013)	1.397(0.040)
IEM-GraphMVP-F	**1.067**(0.039)	**0.716**(0.010)	1.093(0.012)	1.392(0.026)
Δ	↑ 2.47%	↑ 1.10%	↑ 1.18%	↑ 0.36%

Table 2: The RMSE performance on 4 regression datasets of molecular property prediction. We report the mean (standard deviation) RMSE of 10 random seeds from 0 to 9 with scaffold splitting. IEM-baseline represents baseline equipped with IEM. Δ represents the relative improvement percentage calculated by $(1 - \frac{\text{w/o IEM}}{\text{w/ IEM}}) \times 100$.

Results

- **IEM can improve the performance of different GNN architectures**

	GCN	GIN	GAT	GraphSAGE
w/o IEM	66.88	66.68	66.53	66.99
w/ IEM	69.81	70.16	69.76	69.61
Δ	$\uparrow$ 4.39%	$\uparrow$ 5.23%	$\uparrow$ 4.87%	$\uparrow$ 3.92%

Table 3: The average ROC-AUC (%) performance on 8 classification datasets with different GNN architectures. w/o means baseline without IEM and w/ means baseline with IEM. Δ represents the relative improvement percentage calculated by $(1 - \frac{\text{w/o IEM}}{\text{w/ IEM}}) \times 100$.

- **IEM is compatible with conformation-free molecular images, which improves the performance of EdgePred and GraphMVP by using 2D images.**

Image rendering		Method	
Image type	Rendering strategy	EdgePred	GraphMVP
×	×	70.08	71.1
2D	RDKit	72.21 ($\uparrow$ 3.04%)	73.34 ($\uparrow$ 3.15%)
2D	PyMol	72.00 ($\uparrow$ 2.74%)	73.41 ($\uparrow$ 3.25%)
3D	PyMol	**72.26** ($\uparrow$ 3.11%)	**73.89** ($\uparrow$ 3.92%)

Table 4: The average ROC-AUC (%) performance on 8 classification datasets with different image rendering methods. The number in bracket indicates the percentage of absolute performance improvement compared to the baseline without IEM.

Results

Ablation study on different image size

- The more images used, the more obvious the performance improvement
- When using only 5% of the image data, it can still achieve a good performance improvement, showing the efficiency of IEM.

	image size					
	0%	5%	10%	20%	50%	100%
IEM	71.10	72.20	72.26	72.95	73.38	73.89
Δ	-	↑ 1.55%	↑ 1.64%	↑ 2.60%	↑ 3.20%	↑ 3.92%

Table 5: The average ROC-AUC (%) performance on 8 classification datasets with different number of images. The image size represents the proportion of image samples used. We use GraphMVP as baseline model. Δ represents the relative improvement percentage.

Results

Ablation study on two enhancers KE and TE

- KE and TE can consistently improve the performance of different GNN architectures
- The improvement of KE is larger than that of TE, indicating that atomic, bond, geometric, and property knowledge is more effective.
- By combining both enhancers, even more performance gains can be achieved.

Enhancer		Method			
KE	TE	GCN	GIN	GAT	GraphSAGE
×	×	66.88	66.68	66.53	66.99
×	✓	68.07 (1.19)	68.16 (1.48)	68.48 (1.95)	68.44 (1.45)
✓	×	68.26 (1.38)	68.60 (1.92)	68.59 (2.06)	68.58 (1.59)
✓	✓	69.81 (2.93)	70.16 (3.48)	69.76 (3.23)	69.61 (2.62)

Table 6: Ablation results on knowledge enhancer (KE) and task enhancer (TE). The average ROC-AUC (%) performance on 8 classification datasets is reported. The number in bracket indicates the absolute performance improvement compared to the baseline without KE and TE.

Acknowledge

This work was supported by the National Natural Science Foundation of China (grant nos. 62122025, U22A2037, 62250028), Postgraduate Scientific Research Innovation Project of Hunan Province (grant no. CX20220380).

Special thanks to Shanghai Yuyao Biotechnology Co., Ltd. for scientific cooperation and DrugAI for professional guidance.

Code