中国人民大学
RENMIN UNIVERSITY OF CHINA

高瓴人工智能学院
Gaoling School of Artificial Intelligence

3D Infomax improves GNNs for Molecular Property Prediction

Hannes Stark, Dominique Beaini, Gabriele Corso, Prudencio Tossou,
Christian Dallago, Stephan Gunnemann, Pietro Lio

(ICML 2022)

Fanmeng Wang

2023-7-26

Outline

- Introduction
- Background
- Method: 3D Infomax
- Experiment
- Conclusion

- Introduction
- Background
- Method: 3D Infomax
- Experiment
- Conclusion

Introduction

- **Molecular property prediction** is one of the fastestgrowing applications of deep learning with critical real-world impacts.
- **GNNs** is the standard approach method for molecular property prediction.

Standard Approach:
Use GNNs with the molecular graph as the only input and ignore 3D based atomic interactions.

Predictions are fast but less accurate since 3D information cannot be leveraged.

- **3D structures** can improve the accuracy of molecular property prediction.

Explicit 3D Approach:
Employ classic (1) or learned (2) methods to compute 3D coordinates and use them as input to a 3D Graph Neural Network.

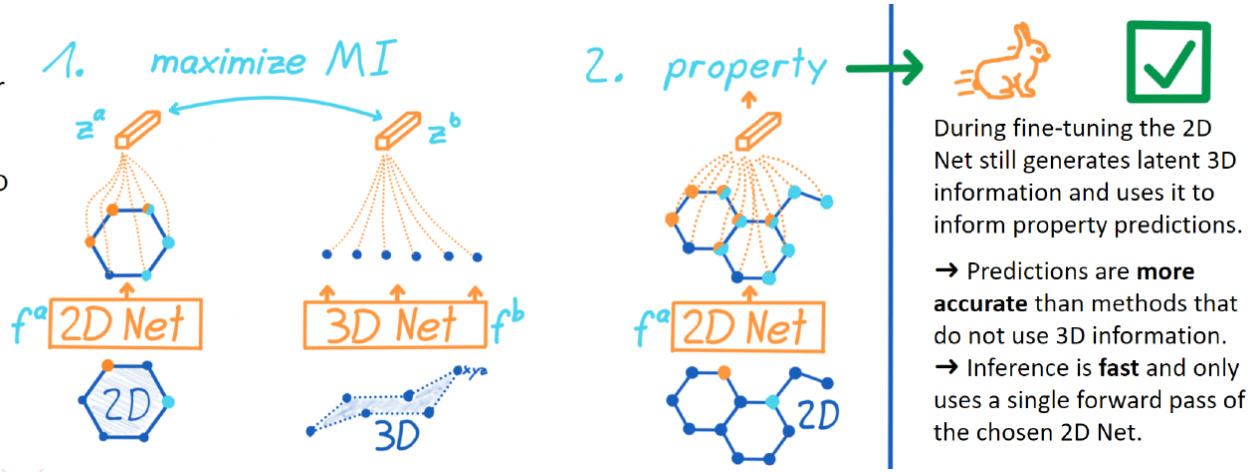
Accurate predictions but methods for generating coordinates are too slow for many real-world applications.

Introduction

- In this paper, we propose a pre-training strategy: **3D Infomax**
  - We use existing 3D molecular datasets to **pre-train** a model to reason about the geometry of molecules given only their 2D molecular graphs.
  - **Pre-train:** we maximizes the **mutual information** between learned 3D summary vectors and the representations of a graph neural network (GNN).
  - **Fine-tune:** for molecules with unknown geometry, the GNN is still able to produce **implicit 3D information** and uses it for downstream tasks.

Our 3D Infomax:
1. Pre-train 2D Net with the molecules for which 3D information is available and learn to generate implicit 3D information in latent representations.

2. Transfer weights of 2D Net and **fine-tune** for predicting molecular properties.



During fine-tuning the 2D Net still generates latent 3D information and uses it to inform property predictions.

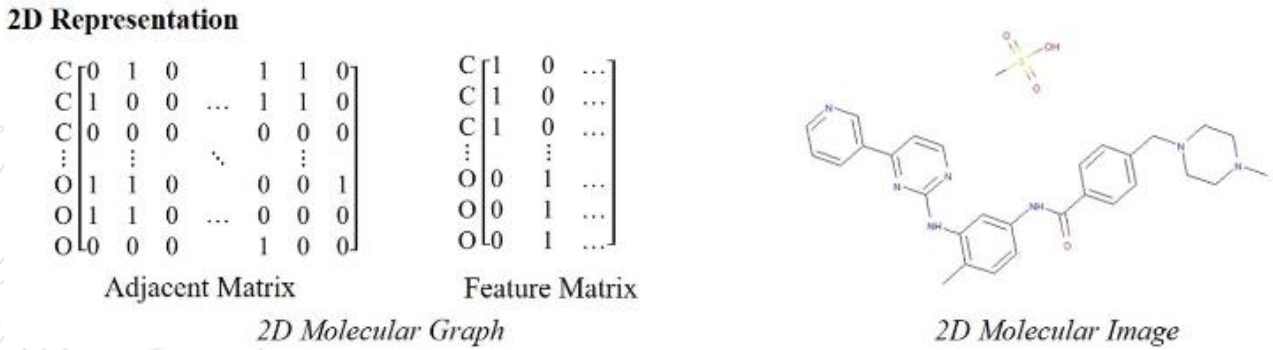
→ Predictions are **more accurate** than methods that do not use 3D information.
→ Inference is **fast** and only uses a single forward pass of the chosen 2D Net.

- Introduction
- Background
- Method: 3D Infomax
- Experiment
- Conclusion

Background

■ 2D Molecular Graphs

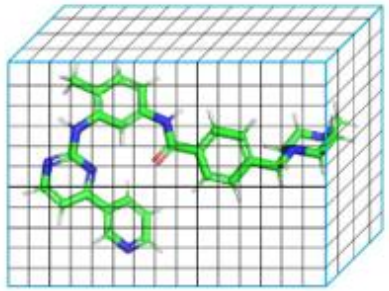
- A molecule's 2D information can be represented as a **graph** G = (V, E) with atoms V as **nodes**, and the **edges** E given by covalent bonds.
- The 2D information of edges could contain the **bond type** while nodes are attributed with features such as the **atomic number**, but no 3D coordinates.



Background

■ 3D Molecular Conformers

- Molecules are **dynamic 3D structures** that exist in different spatial conformations.
- For a single 2D molecular graph, there are **multiple 3D conformers** and they can exhibit different chemical properties.
- Multiple 3D conformers are represented as a set of point clouds $\{R^j\}_{j\epsilon\{1...c\}}$. Each point cloud $R = \{r_v\}_{v\in V}$ specifies the locations of all atoms V in the molecule.
- **Several tools** to compute conformers: RDKit's ETKDG algorithm && CREST



3D Molecular Grid

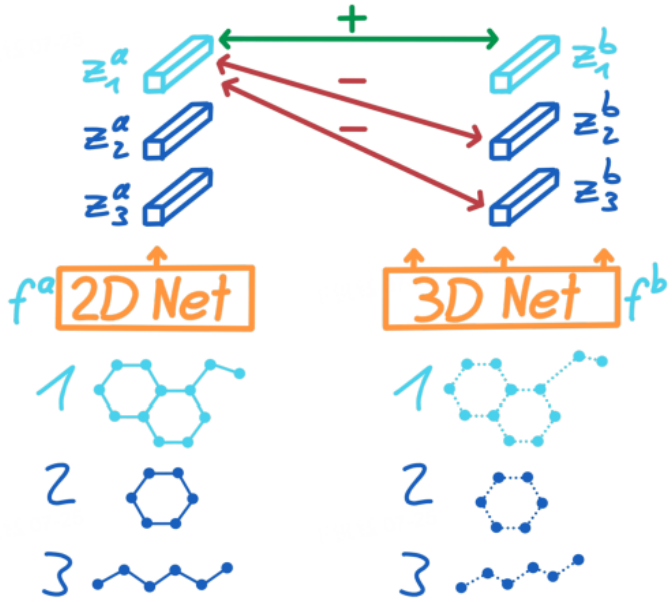
C	6.16	0.32	0.78
C	7.85	0.67	−0.90
C	7.12	−0.56	1.58
⋮		⋮	
O	−0.83	0.25	−1.04
O	−6.61	−0.26	−0.56
O	−1.83	−0.23	−0.11

3D Coordinates

Background

■ Self-Supervised Learning (SSL)

- SSL attempts to find supervision signals in **unlabelled data** to learn meaningful representations.
- **Contrastive learning** is a popular class of methods that learn representations by comparing the embeddings of similar and dissimilar inputs.



- Introduction
- Background
- Method: 3D Infomax
- Experiment
- Conclusion

Method: 3D Infomax

■ Method Design

- To achieve our goal of having a **2D GNN** that is able to reason about **3D geometry** from **only 2D inputs**, we pre-train our model using **contrastive learning**.
- We can use a **similar loss** to jointly optimize our models:

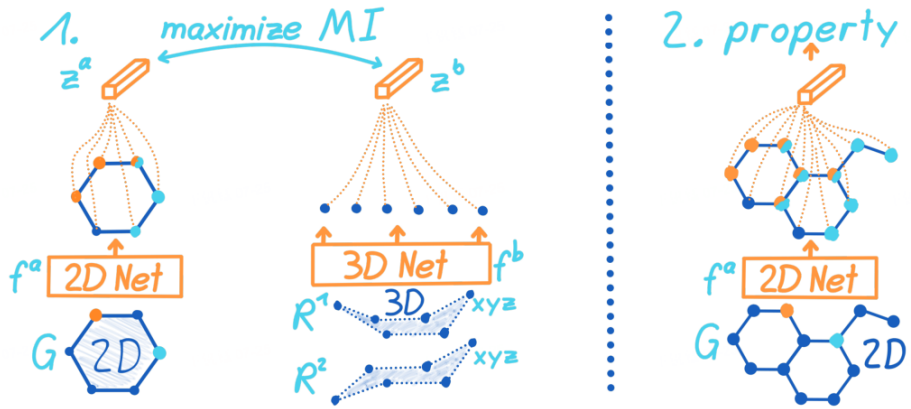


Figure 2. We first pre-train a 2D network f^a by maximizing the mutual information (MI) between its representation z^a of a molecular graph G and a 3D representation z^b produced from the molecules' conformers R^j. In step 2, the weights of f^a are transferred and fine-tuned to predict properties.

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[\log \frac{e^{sim(z_i^a, z_i^b)/\tau}}{\sum_{\substack{k=1 \\ k\neq i}}^{N} e^{sim(z_i^a, z_k^b)/\tau}}\right]$$

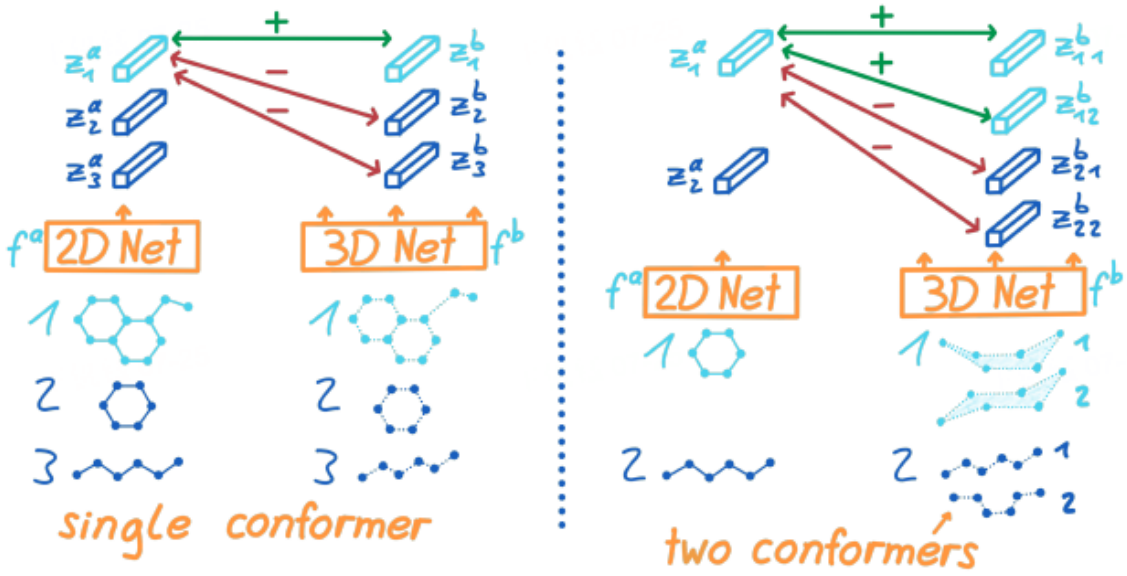
$$sim(z^a, z^b) = z^a \cdot z^b / (\|z^a\| \|z^b\|)$$

- **Contrastive distillation perspective**: the **student 2D network** learns from **the teacher 3D network** to produce **3D information**.

Method: 3D Infomax

■ Method Design

- For most molecules, there are **multiple low-energy stable conformers**.
- We found that leveraging **structural information from multiple conformers** provides significant benefits.
- We can sum over **the similarities of all conformers** to obtain the final loss:



$$\mathcal{L}^{multi3D} = -\frac{1}{N}\sum_{i=1}^{N}\left[\log \frac{\sum_{j=1}^{c} e^{sim(z_i^a, z_{i,j}^b)/\tau}}{\sum_{\substack{k=1 \\ k\neq i}}^{N}\sum_{j=1}^{c} e^{sim(z_i^a, z_{k,j}^b)/\tau}}\right]$$

Method: 3D Infomax

■ 3D Network

- The **2D Network** can be any GNN that one chooses for the downstream task.
- The 3D network takes **the 3D coordinates of the atoms** as input and produces an **SE(3) invariant representation** vector.
- The **initial edge representations** are given by the encoded pair distances fed through an initial feed-forward network.

$$\gamma(d_{uv}) = \Big(d_{uv}, \sin\big(\frac{d_{uv}}{2^0}\big), \cos\big(d_{uv}/2^0\big), \ldots, \sin\big(\frac{d_{uv}}{2^{F-1}}\big), \cos\big(\frac{d_{uv}}{}\big)\Big).$$

$$d_{uv}^0 = U_{init}(\gamma(d_{uv}))$$

- The **initial atom representations** are all set to the same learned vector with a standard normal.

Method: 3D Infomax

■ 3D Network

➢ Every layer updates **the edge and atom representations** and iteratively encodes 3D information into them as follows:

$$m_{uv} = U_{edge}([h_u^l \parallel h_v^l \parallel d_{uv}^l])$$

$$d_{uv}^{l+1} = d_{uv}^l + m_{uv}$$

$$h_u^{l+1} = U_h([h_u \parallel \sum_{\substack{v=1 \\ v \neq u}}^{n} m_{uv} * \sigma(U_{softedge}(m_{uv})]).$$

➢ **Note:** 3D network have no access to 2D information such as atom or bond features; otherwise, the empirical estimate of the mutual information could be increased.

- Introduction
- Background
- Method: 3D Infomax
- **Experiment**
- Conclusion

Experiment

■ Quantum Mechanical Properties

Table 1. MAE for QM9's properties. **3D Infomax** is tested with three different pre-training datasets and **GraphCL** uses a two times larger subset of GEOM-Drugs. **True 3D SMP** is a 3D GNN using ground truth 3D coordinates (hidden from other methods). Details on confidence intervals are in Appendix B. Colors indicate improvement (lower MAE) or worse performance compared to the randomly initialized (**Rand Init**) model.

		Pre-training baselines				Our 3D Infomax			RDKit	*True 3D*
Target	Rand Init	GraphCL	ProPred	DisPred	ConfGen	QM9	Drugs	QMugs	SMP	*SMP*
μ	0.4133±0.003	0.3937	0.3975	0.4626	0.3940	**0.3507**	**0.3512**	0.3668	0.4344	0.0726
α	0.3972±0.014	0.3295	0.3732	0.3570	0.4219	0.3268	0.2959	**0.2807**	0.3020	0.1542
HOMO	82.10±0.33	79.57	93.11	80.58	79.75	**68.96**	70.78	70.77	82.51	56.19
LUMO	85.72±1.62	80.81	99.84	84.93	79.16	**69.51**	71.38	78.10	80.36	43.58
GAP	123.08±3.98	120.08	131.99	116.21	110.72	**101.71**	102.59	103.85	114.24	85.10
R2	22.14±0.21	21.84	29.21	29.23	20.86	**17.39**	18.96	18.00	22.63	1.51
ZPVE	15.08±2.83	12.39	11.17	25.91	21.10	7.966	9.677	12.06	**5.18**	2.69
c_v	0.1670±0.004	0.1422	0.1795	0.1587	0.1555	0.1306	0.1409	**0.1208**	0.1419	0.0498

Table 2. The MAE for predicting GEOM-Drugs' properties. **3D Infomax** compared with **GraphCL** and no pre-training.

Method	Gibbs	$\langle E \rangle$
Rand Init	.2035	.1026
GraphCL	.1941	.0995
3D Infomax QM9	.1852	.0968
3D Infomax Drugs	**.1811**	**.0952**
3D Infomax QMugs	.1835	.0965

- **Significant improvements:** 3D Infomax provides significant improvements for a wide range of properties.
- **Great generalization from pre-training to fine-tuning:** Pre-training on GEOM-Drugs and QMugs can also leads to improvements of 19% and 18% respectively.

Experiment

■ Number of Conformers and pre-training Molecules

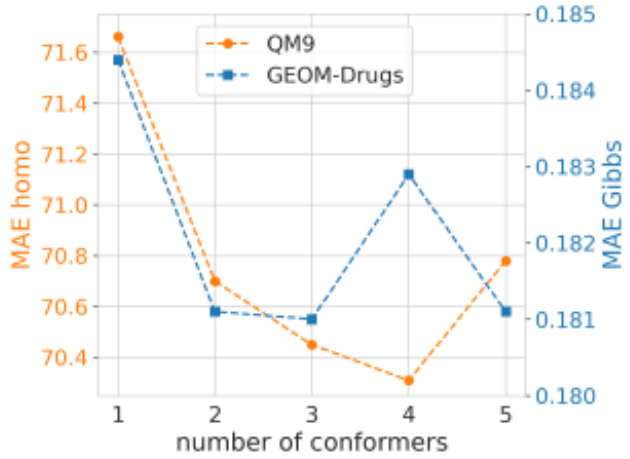


Figure 4. The MAE for the QM9's homo and GEOM-Drugs' Gibbs property when varying the number of GEOM-Drugs' conformers used during pre-training.

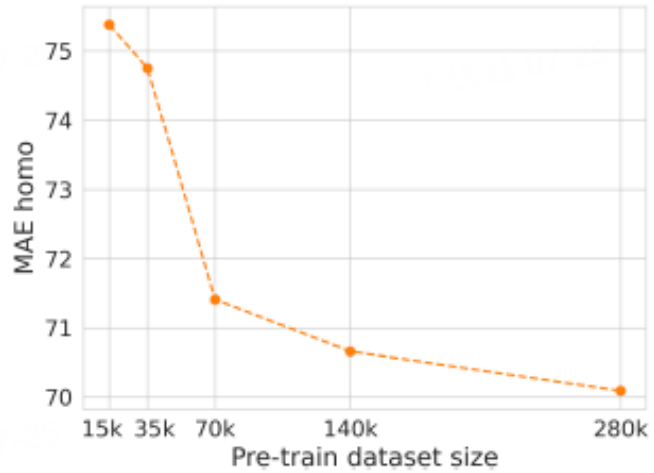


Figure 5. The MAE when using different numbers of molecules of GEOM-Drugs during pre-training.

- Using more than a single conformer can show great benefits.
- The performance improving as the size of the pre-training dataset increases. However, the returns are diminishing, and we cannot claim that even larger pre-training datasets are likely to drastically improve performance.

- Introduction
- Background
- Method: 3D Infomax
- **Experiment**
- Conclusion

Conclusion

- We propose a pre-training strategy, **3D Infomax**, that teaches a GNN to produce latent 3D information from 2D molecular graphs.
- The pretrained GNN can be later used during fine-tuning to improve molecular property predictions while **keeping the inference speed** of a standard GNN operating on 2D molecular graphs.
- Besides, the embedded 3D knowledge can be transferred across highly **different types of molecules** (e.g., from molecules with an average of 18 atoms to drug-like molecules with 44.4 atoms).
- Lastly, we observed that **using multiple molecular conformers** during pre-training provides valuable additional information to further improve downstream property predictions.

Thank You for listening!

Fanmeng Wang

2023-7-26