3D-GSRD: 3D Molecular Graph Auto-Encoder with Selective Re-mask Decoding

Chang Wu[1*], Zhiyuan Liu[2*], Wen Shu[3], Liang Wang[4], Yanchen Luo[1],
Wenqiang Lei[3], Yatao Bian[2], Junfeng Fang[2†], Xiang Wang[1†]

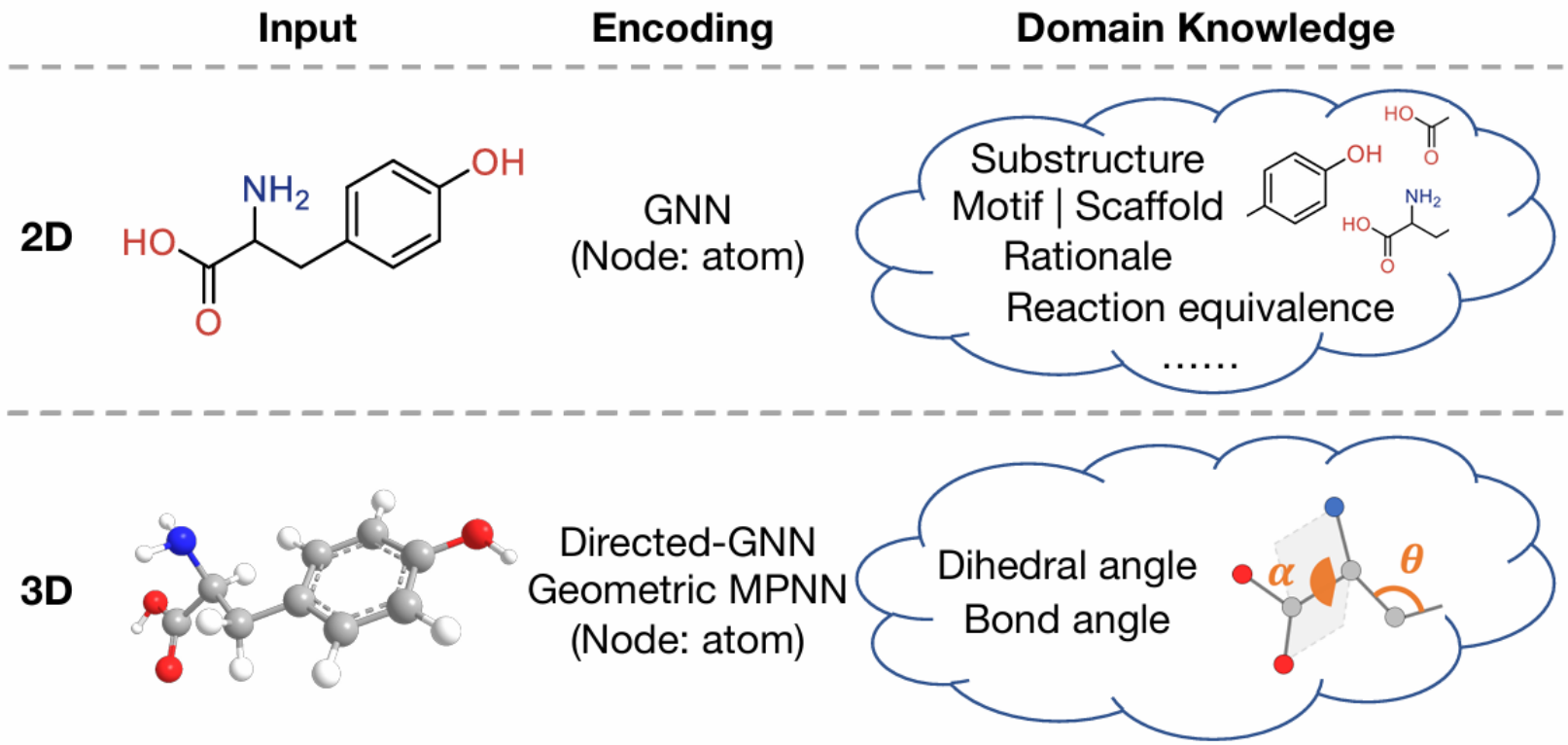
[1] University of Science and Technology of China, [2] National University of Singapore
[3] Sichuan University, [4] Institute of Automation, Chinese Academy of Sciences

❖ Outline

1. **Introduction**
2. Method
3. Analysis
4. Experiments
5. Conclusion

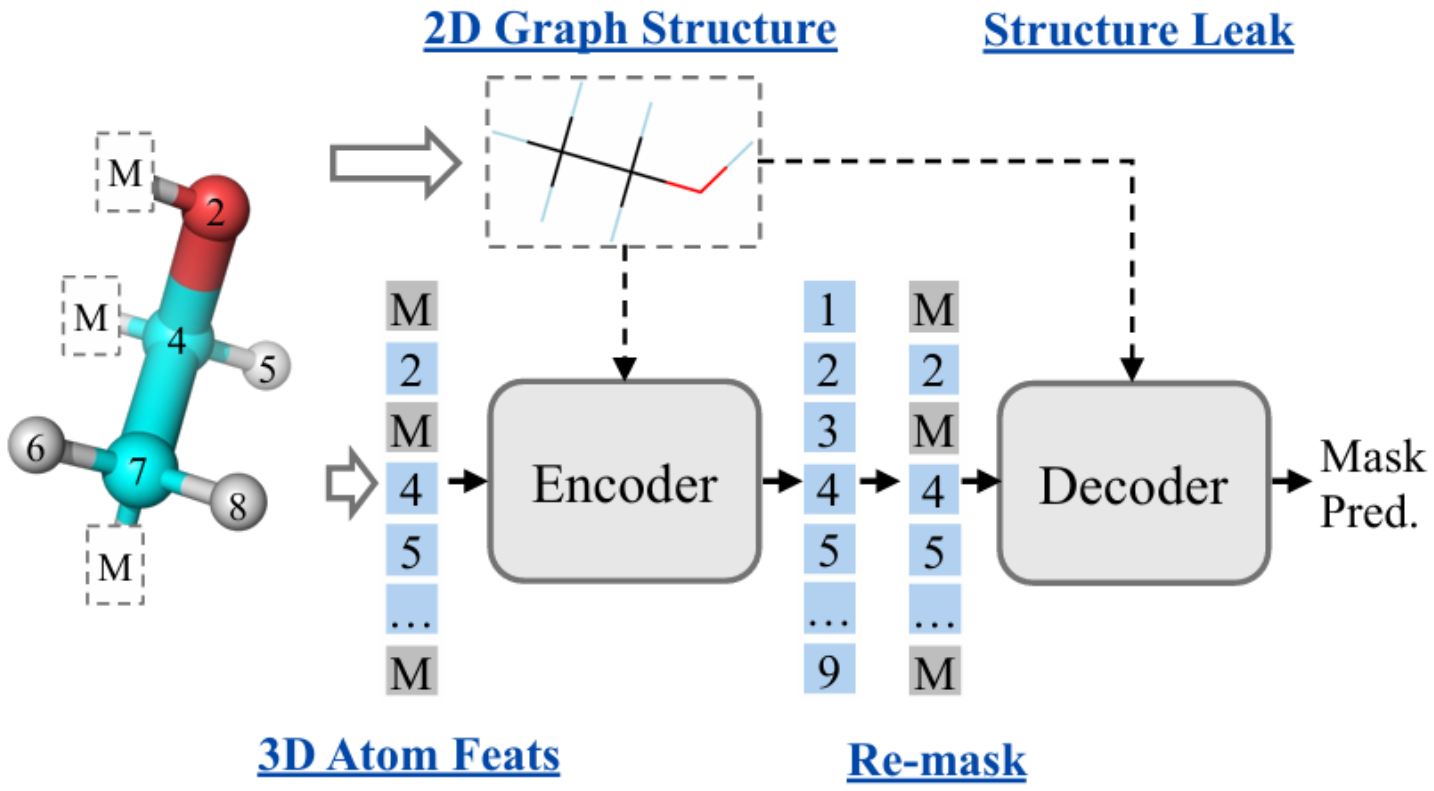
1.1 Molecular Representation Learning

- Molecular representation learning is fundamental to a wide range of downstream tasks, including de novo drug design, molecular dynamics simulation, and molecular property prediction, in which 3D structural information is critical.



1.2 Masked graph modeling

- Masked Graph Modeling typically consists of three components: (1) 3D graph masking; (2) A 3D graph encoder; (3) A 3D graph decoder.



1.2 3D Masked Graph Modeling

To adapt re-mask decoding for 3D MGM, we identify two challenges:

- Leaking 2D structure to decoder weakens encoder's MRL capability.
- Structure-independent decoding can prevent structure leakage, but hinders reconstruction of re-masked atoms.

To address the challenges, we introduce 3D Molecular Graph Auto-Encoder with Selective Re-maskDecoding (3D-GSRD).

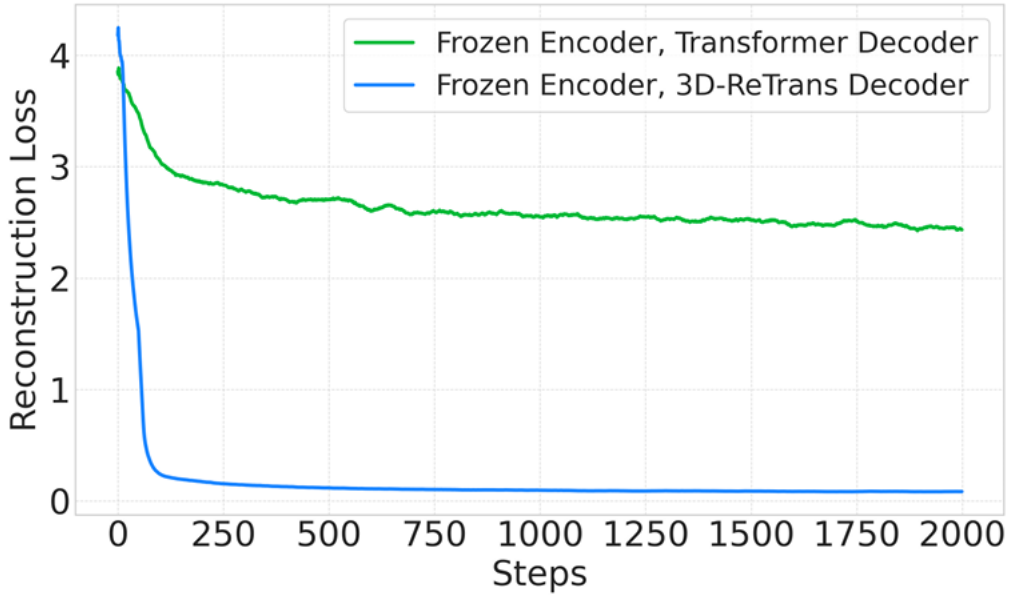


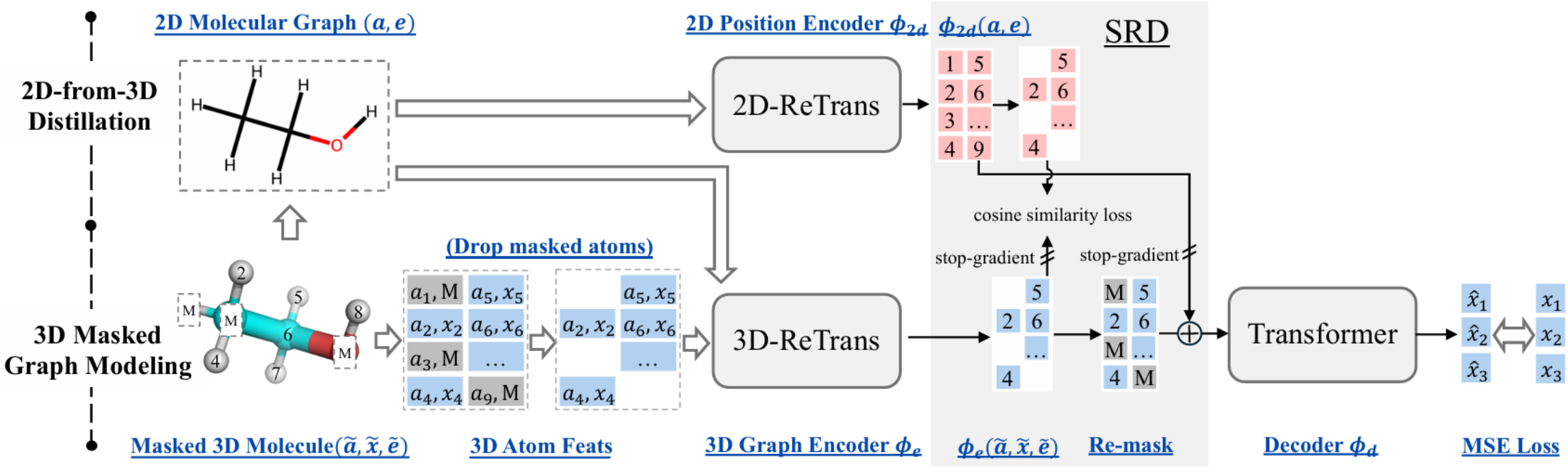
Figure 2: Reconstruction loss across two pre-training settings. We compare settings using frozen encoders (3D-ReTrans) and structure-independent (Transformer) versus structure-dependent (3D-ReTrans) decoders.

❖ Outline

1. Introduction
2. **Method**
3. Analysis
4. Experiments
5. Conclusion

2.1 Framework Overview: 3D-GSRD

- 3D-GSRD contains three key elements: (1) a 3D-ReTrans encoder; (2) the SRD that re-masks only 3D-relevant information from the encoder representations while preserving its 2D structure information via 2D-from-3D distillation; (3) a structure-independent decoder.



2.2 SRD: Selective Re-mask Decoding

- Re-mask Decoding with 2D Graph Position Encoder

$$\text{re-mask}(\mathbf{h}_i) = \begin{cases} \mathbf{m}_h, & \forall i \in \mathcal{V}_m, \\ \mathbf{h}_i, & \text{otherwise.} \end{cases} \qquad \text{SRD}(\mathbf{h}, \tilde{G}) = \text{re-mask}(\mathbf{h}) + \text{stop-grad}(\phi_{2d}(\mathbf{a}, \mathbf{e}))$$

- Building a 2D Graph Position Encoder without Structure Leakage via 2D-from-3D Distillation

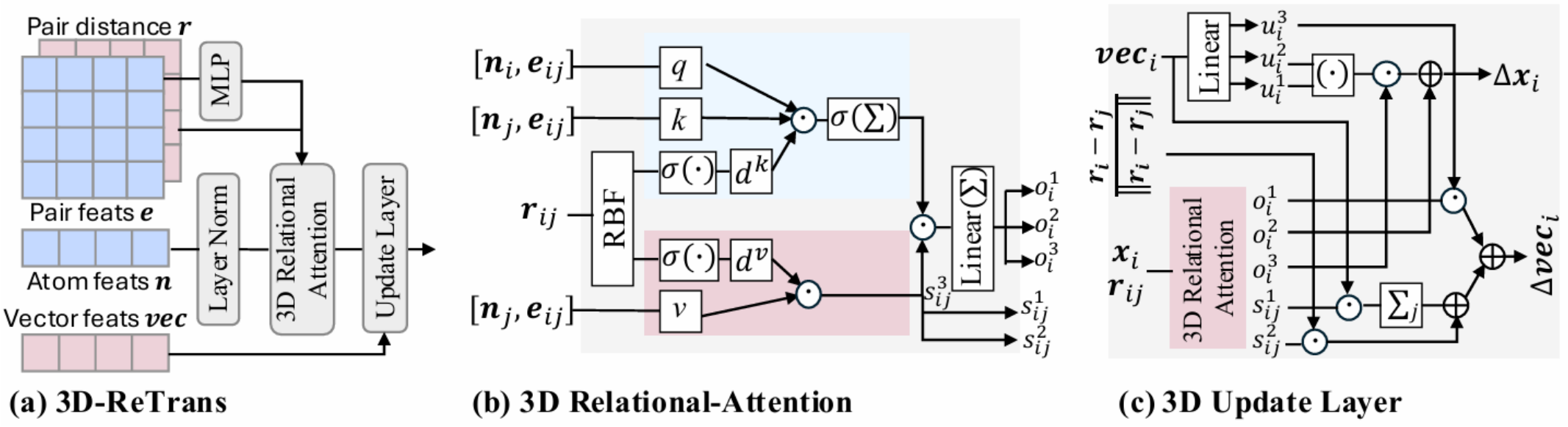
$$\mathcal{L}_{\text{distill}} = -\sum_{i \notin \mathcal{V}_m} \cos(\phi_{2d}(\mathbf{a}, \mathbf{e})_i, \text{stop-grad}(\phi_e(\mathbf{x}, \mathbf{a}, \mathbf{e})_i))$$

- Pretraining loss

$$\mathcal{L}_{\text{MGM}} = \sum_{i \in \mathcal{V}_m} \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|^2. \qquad \mathcal{L}_{\text{pretrain}} = \mathcal{L}_{\text{MGM}} + \mathcal{L}_{\text{distill}}$$

2.3 3D Relational-Transformer

- (a) 3D-ReTrans is constructed by stacking multiple 3D Relational-Attention and 3D Update Layers. (b) 3D Relational-Attention that processes both atom-wise and pair-wise features. (c) 3D Update Layer that includes a residual connection.



Core design: explicitly separate and jointly process two types of features

- scalar features: encode scalar information like atom types and distances
- vector features: capture directional geometric information

❖ Outline

1. Introduction
2. Method
3. **Analysis**
4. Experiments
5. Conclusion

3 Analysis

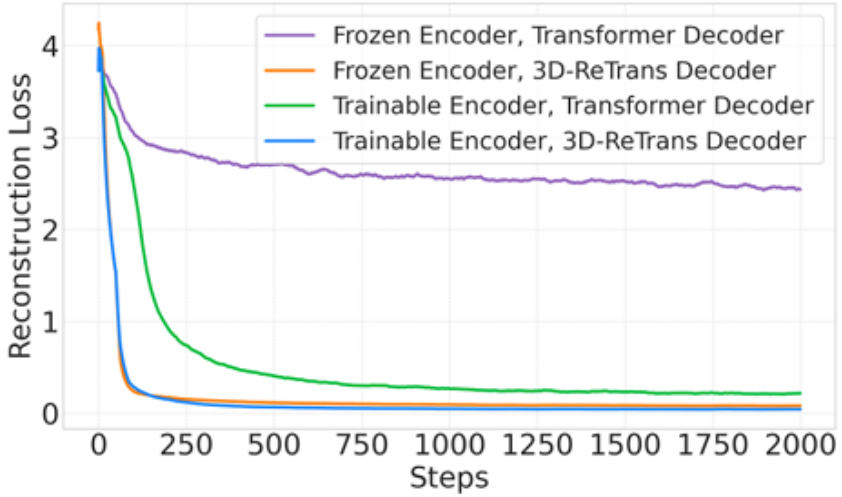


Figure 5: Reconstruction loss across four pretraining settings.

Table 1: Analyzing the decoder and SRD. Performance (MAE ↓) on MD17. The variant without SRD and $\mathcal{L}_{\text{distill}}$ corresponds to the model ablated without 2D-PE.

Decoder	SRD	$\mathcal{L}_{\text{distill}}$	Salicylic	Toluene	Uracil
Structure-dependent	✓	✓	0.0401	0.0291	0.0334
Structure-independent	✗	✗	0.0404	0.0292	0.0328
Structure-independent	✓	✗	0.0416	0.0293	0.0329
Structure-independent	✓	✓	**0.0387**	**0.0275**	**0.0315**

Analysis1. The structure-dependent decoder can diminish the encoder's role in MRL. The structure-independent decoder heavily relies on high-quality encoder representations.

Analysis 2. Structure-independent decoder improves downstream performance compared to structure-dependent decoder.

Analysis 3. 2D-PE and 2D-from-3D distillation boost downstream performance.

3 Analysis

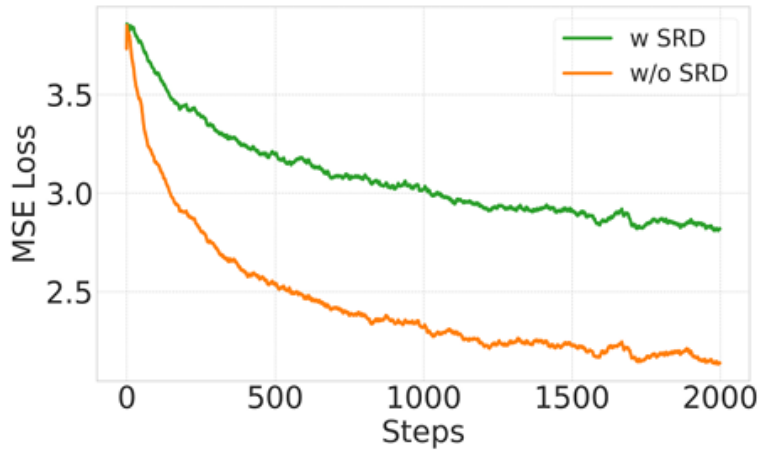


Figure 6: Probe encoder for masked atom coordinates when pretrained with/without SRD.

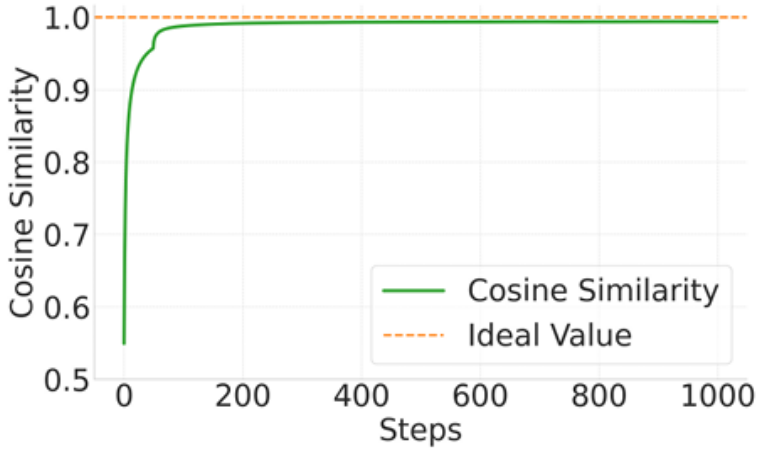


Figure 7: Reconstructing 2D-PE representation using 3D encoder representation.

Analysis 4. SRD prevents the encoder representation from containing information about 3D coordinates.

Analysis 5. 2D-PE produces 2D structural context without introducing information leakage.

❖ Outline

1. Introduction
2. Method
3. Analysis
4. **Experiments**
5. Conclusion

4.1 Results on MD17

Table 2: Performance (MAE ↓) on MD17 force prediction. The best results are **bold**. The second-best results are underline. Results marked with * are reproduced by us.

Models	Aspirin	Benzene	Ethanol	Malonaldehyde	Naphthalene	Salicylic	Toluene	Uracil
TorchMD-NET	0.1216	0.1479	0.0492	0.0695	0.0390	0.0655	0.0393	0.0484
3D-EMGP	0.1560	0.1648	0.0389	0.0737	0.0829	0.1187	0.0619	0.0773
3D-EMGP (TorchMD-NET)	0.1124	0.1417	0.0445	0.0618	0.0352	0.0586	0.0385	0.0477
Frad*	0.0825	**0.1355**	0.0432	0.0535	0.0431	0.0569	0.0433	0.0482
3D-ReTrans	0.0726	0.1619	0.0556	0.0659	0.0423	0.0523	0.0417	0.0427
3D-GSRD	**0.0583**	0.1435	**0.0355**	**0.0468**	**0.0266**	**0.0356**	**0.0274**	**0.0292**

- 3D-ReTrans achieves performance comparable to TorchMD-NET and surpasses 3D-EMGP on 7 of 8 molecules
- 3D-GSRD attains state-of-the art results on 7 of 8 molecules except Benzene, exceeding the strongest baseline (i.e.,Frad) by a large margin.

4.2 Results on QM9

Table 3: Performance (MAE ↓) on QM9. The best results are **bold**. The second-best results are underline. Results marked with * are reproduced by us.

Models	μ (D)	α (a_0^3)	$homo$ (meV)	$lumo$ (meV)	gap (meV)	$< R^2 >$ (a_0^2)	ZPVE (meV)	U_0 (meV)	U (meV)	H (meV)	G (meV)	C_v ($\frac{cal}{molK}$)
Uni-Mol2	0.089	0.305	-	-	-	5.26	-	-	-	-	-	0.144
SchNet	0.033	0.235	41.0	34.0	63.0	0.07	1.70	14.00	19.00	14.00	14.00	0.033
E(n)-GNN	0.029	0.071	29.0	25.0	48.0	0.11	1.55	11.00	12.00	12.00	12.00	0.031
DimeNet++	0.030	0.043	24.6	19.5	32.6	0.33	1.21	6.32	6.28	6.53	7.56	0.023
PaiNN	0.012	0.045	27.6	20.4	45.7	0.07	1.28	5.85	5.83	5.98	7.35	0.024
SphereNet	0.025	0.045	22.8	18.9	31.1	0.27	**1.12**	6.26	6.36	6.33	7.78	0.022
ComENet	0.025	0.045	23.1	19.8	32.4	0.259	1.20	6.59	6.82	6.86	7.98	0.024
TorchMD-NET	0.011	0.059	20.3	18.6	36.1	**0.033**	1.84	6.15	6.38	6.16	7.62	0.026
3D-ReTrans	0.016	0.055	22.0	17.8	38.0	0.341	1.85	6.18	6.36	6.51	7.89	0.029
Transformer-M	0.037	0.041	17.5	16.2	**27.4**	0.075	1.18	9.37	9.41	9.39	9.63	0.022
SE(3)-DDM	0.015	0.046	23.5	19.5	40.2	0.122	1.31	6.92	6.99	7.09	7.65	0.024
3D-EMGP	0.020	0.057	21.3	18.2	37.1	0.092	1.38	8.60	8.60	8.70	9.30	0.026
Coord	0.016	0.052	17.7	14.7	31.8	0.450	1.71	6.57	6.11	6.45	6.91	**0.020**
Frad*	0.012	0.045	**15.4**	**13.7**	30.6	0.428	1.56	15.88	14.67	14.87	13.52	0.023
SliDe*	0.015	0.050	18.7	16.2	28.8	0.606	1.78	10.05	10.79	11.34	11.80	0.025
Mol-AE*	0.152	0.434	-	-	-	6.962	-	-	-	-	-	0.215
Uni-GEM	0.019	0.060	20.9	16.7	34.5	-	-	-	-	-	-	0.023
3D-GSRD	**0.009**	**0.038**	18.0	14.5	31.1	0.047	1.38	**5.48**	**5.67**	**5.84**	**6.90**	**0.020**

4.3 Ablation Studies

Table 4: Ablation on 3D-ReTrans components. Performance (MAE ↓) on QM9.

Model Components	homo	lumo	zpve
Relational-Transformer	27.7	24.0	1.97
+ 3D Data Augmentation	24.6	23.2	1.92
+ 3D Relational-Attention	23.4	20.3	1.90
+ 3D Update Layer (3D-ReTrans)	**22.0**	**17.8**	**1.85**

Table 5: Analyzing SRD on the Relational-Transformer. Performance (MAE ↓) on MD17.

Decoder	SRD	$\mathcal{L}_{\text{distill}}$	Toluene	Uracil
Structure-dependent	✓	✓	0.1144	0.0813
Structure-independent	✗	✗	0.1250	0.0828
Structure-independent	✓	✗	0.0998	0.0843
Structure-independent	✓	✓	**0.0745**	**0.0733**

The results show that 3D Relational-Attention, the 3D Update Layer, and 3D data augmentation each enhance molecular property prediction, collectively boosting overall performance.

The results show that incorporating SRD consistently improves downstream performance, demonstrating its effectiveness as a general pretraining strategy applicable to diverse 3D graph encoder architectures.

❖ Outline

1. Introduction
2. Method
3. Analysis
4. Experiments
5. **Conclusion**

5 Conclusion

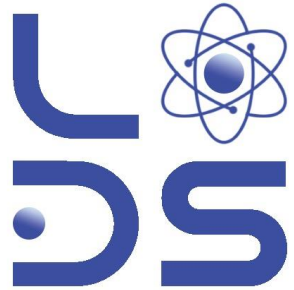
We introduce 3D-GSRD, a 3D MGM framework with three key components: (1) the Selective Re-mask Decoding that selectively re-masks 3D-relevant information while preserving 2D graph structures; (2) a structure-independent decoder that eliminates all structural information by relying solely on encoder representation; and (3) 3D-ReTrans as the 3D graph encoder for MRL. Our detailed analysis reveals the internal mechanisms of SRD and the structure-independent decoder. Extensive experiments demonstrate that 3D-GSRD significantly outperforms baselines on downstream datasets such as QM9 and MD17.


1958
University of Science and Technology of China

NUS
National University
of Singapore

1896
SICHUAN UNIVERSITY

CASIA

NEURAL INFORMATION
PROCESSING SYSTEMS


Thanks!