MolSpectra: Pre-training 3D Molecular Representation with Multi-modal Energy Spectra

Liang Wang[1,2] Shaozhen Liu[1] Yu Rong[3,†] Deli Zhao[3] Qiang Liu[1,2,†] Shu Wu[1,2] Liang Wang[1,2]

[1]NLPR, MAIS, CASIA [2]Sch. AI, UCAS [3]DAMO Academy, Alibaba Group [†]Correspondence

TL;DR

Existing methods for learning 3D molecular representations focus on modeling molecular energy states from classical mechanics, overlooking quantum mechanical effects such as discrete energy levels. We propose `MolSpectra`, which leverages multi-modal molecular spectra for pre-training, thereby integrating quantum mechanical knowledge into molecular representations.

Background

I. Denoising as learning a force field.

Denoising has emerged as a prominent pre-training objective in 3D molecular representation learning. This approach is physically interpretable due to its proven equivalence to learning the molecular force field.

$$\mathcal{L}_{\text{Denoising}}(\mathcal{M}) = \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{x}_0)p(\boldsymbol{x}_0)}\|\text{GNN}_\theta(\boldsymbol{x}) - (\boldsymbol{x} - \boldsymbol{x}_0)\|^2 \simeq \mathbb{E}_{p(\boldsymbol{x})}\|\text{GNN}_\theta(\boldsymbol{x}) - (-\nabla_{\boldsymbol{x}} E(\boldsymbol{x}))\|^2,$$

Essentially, it reveals that *establishing the relationship between 3D geometries and the energy states is an effective pathway to learn 3D molecular representations.*

II. Prior works only model classical mechanics by denoising, overlooking the energy level structures from a quantum mechanical perspective, which can be measured using molecular spectra.

III. Different types of molecular spectra reveal distinct energy level structures. For instance, IR spectra reveal vibrational energy levels, whereas UV-Vis spectra reveal electronic energy levels.

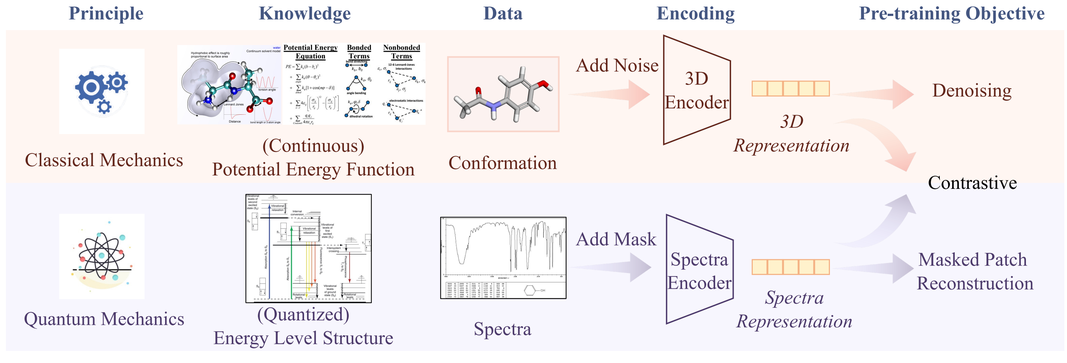
Main Contributions

I. We propose `MolSpectra`, introducing molecular spectra into molecular representation learning for the first time.

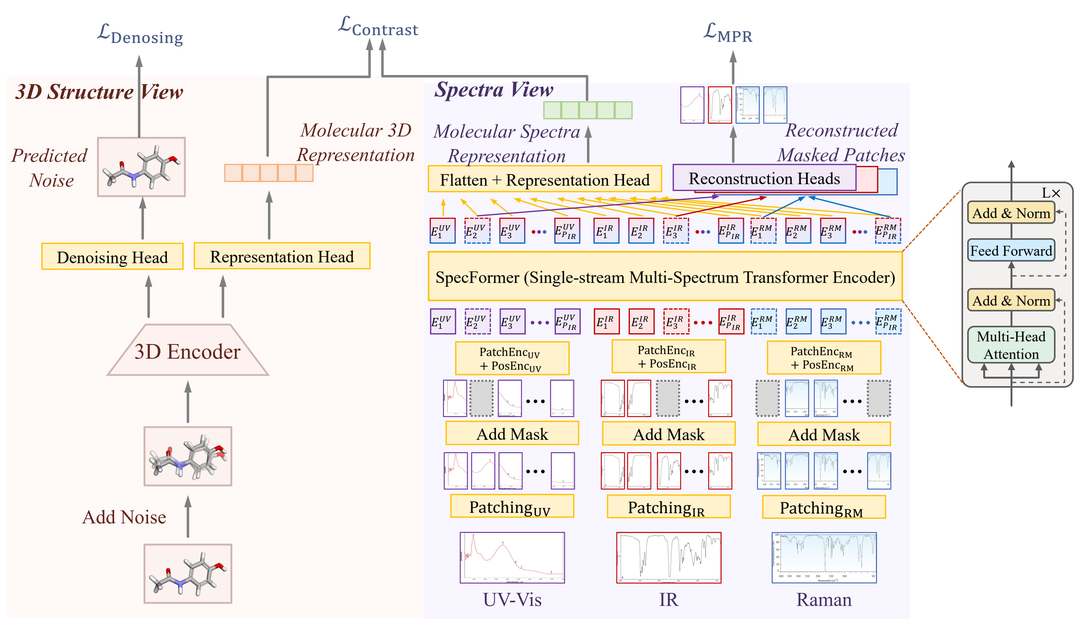
II. We propose `SpecFormer` as an expressive multi-spectrum encoder, along with the masked patches reconstruction objective for spectral representation learning.

MolSpectra

I. Conceptual view of `MolSpectra`, leveraging both conformation and spectra for pre-training.



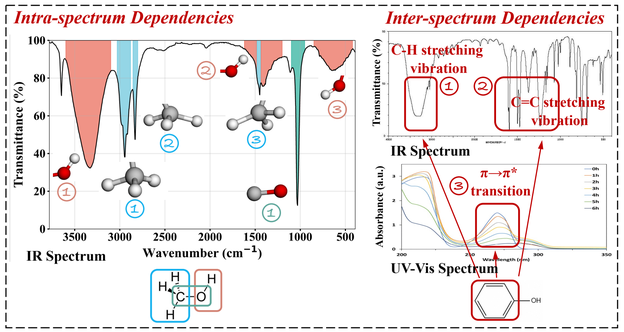
II. Overview of the `MolSpectra` pre-training framework.



`MolSpectra` comprises three sub-objectives: the denoising objective and the masked patches reconstruction objective guide the representation learning of the 3D and spectral modalities respectively, and the contrastive objective aligns the representations of both modalities.

MolSpectra (Cont.)

III. Illustration of intra-spectrum (left) and inter-spectrum (right) dependencies.



Experiments

I. Performance (MAE↓) on benchmark datasets.

	μ (D)	homo (meV)	lumo (meV)	gap (meV)	U_0 (meV)	U (meV)	H (meV)	G (meV)
EGNN	0.029	29.0	25.0	48.0	11.00	12.00	12.00	12.00
PaiNN	0.012	27.6	20.4	45.7	5.85	5.83	5.98	7.35
SphereNet	0.025	22.8	18.9	31.1	6.26	6.36	6.33	7.78
TorchMD-Net	0.011	20.3	17.5	36.1	6.15	6.38	6.16	7.62
Transformer-M	0.037	17.5	16.2	27.4	9.37	9.41	9.39	9.63
SE(3)-DDM	0.015	23.5	19.5	40.2	6.92	6.99	7.09	7.65
3D-EMGP	0.020	21.3	18.2	37.1	8.60	8.60	8.70	9.30
Coord	0.016	17.7	14.7	31.8	6.57	6.11	6.45	6.91
MolSpectra	**0.011**	**15.5**	**13.1**	**26.8**	**5.67**	**5.45**	**5.87**	**6.18**

II. Ablation of spectral modalities.

UV-Vis	IR	Raman	homo	lumo	gap
✓	✓	✓	15.5	13.1	26.8
-	✓	✓	15.8	13.3	27.1
✓	-	✓	16.6	14.1	28.9
✓	✓	-	16.1	13.9	28.3

Each spectral modality contributes differently, with the IR spectrum having the largest contribution.