Latent Diffusion-Based 3D Molecular Recovery from Vibrational Spectra

Wenjin Wu[1], Aleš Leonardis[1], Linjiang Chen[1,2,3], Jianbo jiao[1]

[1]School of Computer Science, University of Birmingham, United Kingdom
[2]School of Chemistry, University of Birmingham, United Kingdom
[3]State Key Laboratory of Precision and Intelligent Chemistry, University of Science and Technology of China, China

Project Page

I. Background

- Existing approaches for recovering molecular structures from IR spectra typically rely on 1D SMILES strings or 2D molecular graphs.
- However, no existing diffusion model has explored the distribution of 3D molecular geometries corresponding to a single IR spectrum.

II. Method

Problem Definition

For a molecule with N atoms, the molecular geometry is represented as $G = \langle \mathbf{x}, \mathbf{h} \rangle$,

- $\mathbf{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_N) \in \mathbb{R}^{N\times 3}$: 3D Cartesian coordinates of atoms,
- $\mathbf{h} = (\mathrm{h}_1, \ldots, \mathrm{h}_N) \in \mathbb{R}^{N\times 1}$: atomic types.

Goal: Learn a probabilistic model θ that recovers the conditional distribution of molecular geometries G given an IR spectrum S, *i.e.*, $p_\theta(G|S)$.

- In our setting, we assume $\mathbf{h}$ and N are known and focus on modelling the conditional distribution over atomic coordinates $\mathbf{x}$, *i.e.*, $p_\theta(\mathbf{x}|S, \mathbf{h})$.

Proposed Model: IR-GeoDiff

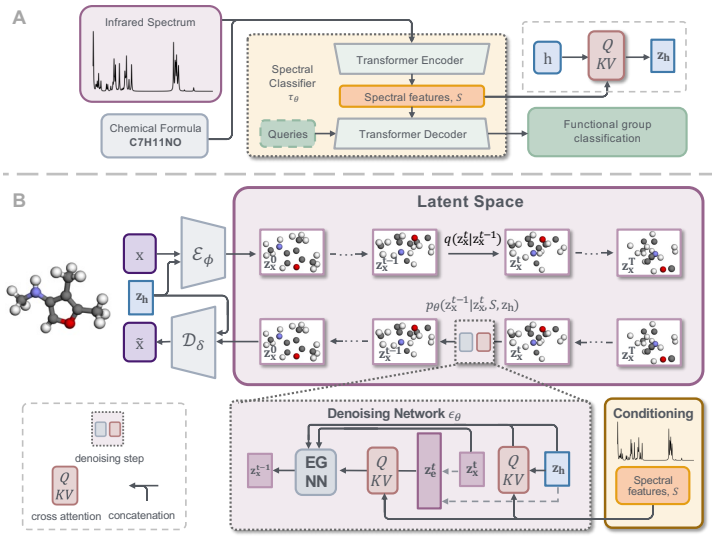

Fig 1. Overview of the proposed **IR-GeoDiff**. **A**: Spectral features S are extracted by a Transformer-based spectral classifier τ_θ. **B**: The encoder $\mathcal{E}_\phi$ maps molecular geometries into a latent representation $\mathbf{z_x}$, which is perturbed through a forward diffusion process and denoised by an equivariant network ϵ_θ conditioned on the spectral features which are injected into node and edge representations via cross-attention. Finally, the decoder $\mathcal{D}_\delta$ reconstructs the 3D geometry from the denoised latent representation.

Evaluation Metrics

Structural Similarity: Chemical graph similarity ($\mathrm{sim_g}$) is measured between the recovered and reference molecular structures, defined as the Tanimoto similarity between Morgan fingerprints computed using RDKit.

Spectral Similarity: Spectral Information Similarity (SIS)[1] is adopted to quantify the similarity between the input spectrum and the spectrum of the recovered molecule. SIS* is further defined to compute the SIS restricted to the functional group region (1,350-4,000 $\mathrm{cm^{-1}}$).

Dataset

We use the **QM9S**[2] dataset, which is derived from QM9 and contains approximately 130k small molecules composed of five atom types (H, C, N, O, and F) with up to nine heavy atoms. QM9S extends QM9 by providing multiple types of spectra. In particular, IR spectra are obtained by re-optimising molecular geometries and performing vibrational analysis using Gaussian 16. We randomly select 1,000 molecules as the test set, with the remaining molecules split into training and validation sets at a 95:5 ratio.

III. Results and Discussion

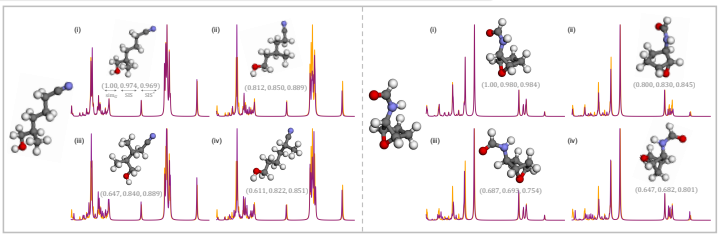

Fig 2. **Examples** of molecules sampled under IR spectral conditions. For spectra, orange curves: input spectra; purple curves: spectra of sampled molecules. For atoms, hydrogen: white , carbon: gray , oxygen: red , nitrogen: blue .

Attention-based Analysis

Both spectral-edge cross-attention module and atom-edge cross-attention module within the model are able to focus on spectral regions associated with characteristic functional groups. This behavior qualitatively resembles how chemists interpret IR spectra.

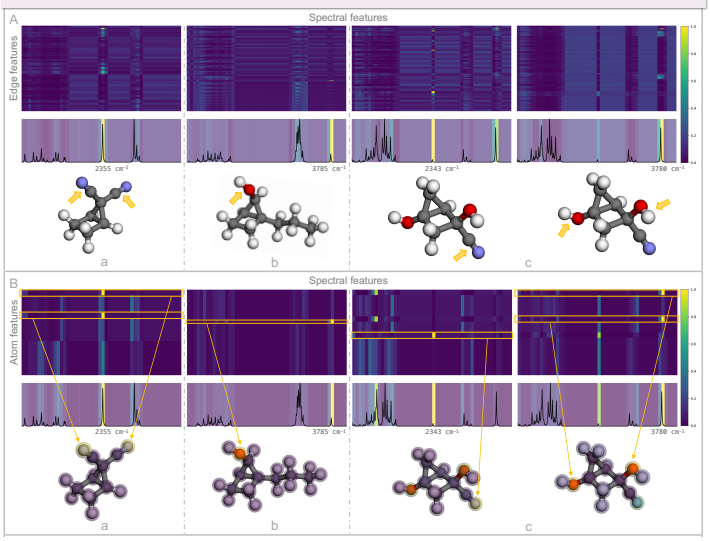

Fig 3. Visualisation of cross-attention between spectral features and molecular structure representations. **A**: Spectral–edge cross-attention highlighting functional-group-specific vibrational signatures. **B**: Spectral–atom cross-attention revealing associations between characteristic peaks and non-carbon heavy atoms.

Analysis on Exceptions

- A sharp density peak appears near the top-right corner, indicating a cluster of molecules that are near-exact matches to the reference.
- While graph similarity generally increases with SIS, the correlation between the two is moderate.
  High graph similarity but low SIS: conformational change
  High SIS but low graph similarity: carbon skeleton change

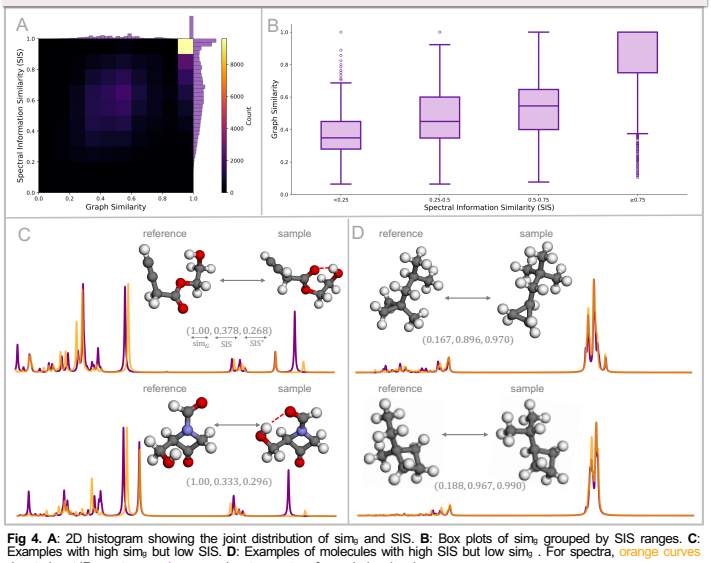

Fig 4. A: 2D histogram showing the joint distribution of $\mathrm{sim_g}$ and SIS. **B**: Box plots of $\mathrm{sim_g}$ grouped by SIS ranges. **C**: Examples with high $\mathrm{sim_g}$ but low SIS. **D**: Examples of molecules with high SIS but low $\mathrm{sim_g}$. For spectra, orange curves denote input IR spectra; purple curves denote spectra of sampled molecules.

1. McGill, C., Forsuelo, M., Guan, Y . and Green, W .H., 2021. Predicting infrared spectra with message passing neural networks. *Journal of Chemical Information and Modeling*, *61*(6), pp.2594-2609.
2. Zou, Z., Zhang, Y., Liang, L., Wei, M., Leng, J., Jiang, J., Luo, Y. and Hu, W., 2023. A deep learning model for predicting selected organic molecular spectra. *Nature Computational Science*, *3*(11), pp.957-964.