Integrating Frequency Domain into Machine Learning for Enhanced Molecular Excitation Spectra Prediction

M. Benaissa[1*], S. Tricot[1], K. Hatada[2], D. Sébilleau[1]

[1]*Université de Rennes, CNRS, IPR (Institut de Physique de Rennes) - UMR 6251 F-35000, Rennes, 35000, France.*

[2]*Department of Physics, University of Toyama, Toyama 930-8555, Gofuku 3190, Japan*

*mohammed.benaissa@univ-rennes.fr

Spectroscopy plays a pivotal role in unraveling the mysteries of the physical world, from material characterization to the discovery of novel phenomena. However, conventional methods like absorption, emission, and Raman spectroscopy can be time-consuming and resource-intensive, limiting their widespread adoption. While theoretical approaches based on quantum mechanics offer deeper insights, they often require significant computational power [1]. This abstract proposes a transformative AI-driven framework to predict excited state properties, particularly molecular excitation spectra, directly from molecular structures, overcoming the time-cost limitations of traditional methods.

Existing machine learning models for predicting spectra often directly learn spectral curves[1,2]. We propose a novel approach inspired by computer vision research, incorporating the frequency domain into the training process for enhanced predictive power [3–6]. Our model utilizes convolutional neural networks (CNNs) specifically tailored to handle the 2D vector output of Fourier transforms, allowing it to effectively learn the spatial patterns within the spectra.

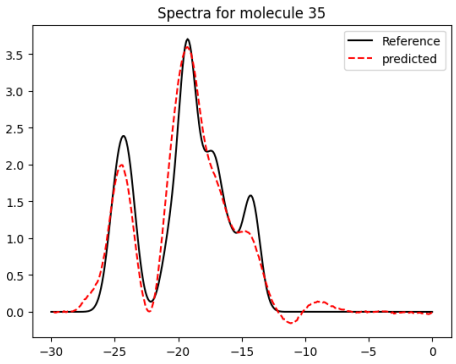

Figure 1. Comparison between ML model predicted and ground truth for molecular excitation spectra. spectra.

Employing molecular coordinates and charges as input features, the model is trained on a curated dataset of 7,000 molecules represented by Coulomb matrices [1]. A 64%-16%-20% split for training, validation, and testing, respectively, ensures robust model evaluation. On the test set, the model achieves a low Mean Squared Error (MSE) of 7.62, demonstrating accurate spectral predictions. This innovative AI-driven framework holds immense potential to improve spectroscopy, offering a data-driven paradigm for faster, more accurate analyses and accelerating scientific discovery.

References

[1] K. Ghosh, A. Stuke, M. Todorović, P.B. Jørgensen, M.N. Schmidt, A. Vehtari, P. Rinke, Advanced Science 6 (2019) 1801367.

[2] P. Mishra, D. Passos, F. Marini, J. Xu, J.M. Amigo, A.A. Gowen, J.J. Jansen, A. Biancolillo, J.M. Roger, D.N. Rutledge, A. Nordon, TrAC Trends in Analytical Chemistry 157 (2022) 116804.

[3] Z. Chen, H. Lin, Z. Qiang, J. Ou, Y. Wu, in: 2022 IEEE 8th International Conference on Computer and Communications (ICCC), IEEE, Chengdu, China, 2022, pp. 1958–1963.

[4] J.A. Stuchi, M.A. Angeloni, R.F. Pereira, L. Boccato, G. Folego, P.V.S. Prado, R.R.F. Attux, in: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, Tokyo, 2017, pp. 1–6.

[5] T. Watanabe, D.F. Wolf, Applied Soft Computing 112 (2021) 107851.

[6] J.A. Stuchi, L. Boccato, R. Attux, (2020).