Hyperspectral Unmixing for Raman Spectroscopy via Physics-Constrained Autoencoders

Preprint

Codebase

Dimitar Georgiev — *Imperial College London*

Álvaro Fernández-Galiana — *Imperial College London*; *Present: University of Oxford*

Simon Vilms Pedersen — *Imperial College London*; *Present: University of Southern Denmark*

Georgios Papadopoulos — *Imperial College London*

Ruoxiao Xie — *Imperial College London*; *Present: University of Liverpool*

Molly M. Stevens — *University of Oxford, Imperial College London*

Mauricio Barahona — *Imperial College London*

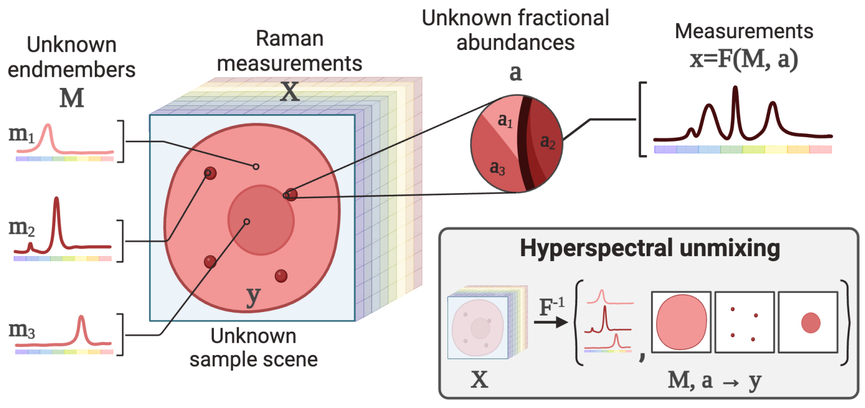
TL;DR

We propose **a framework for Raman spectroscopy unmixing** based on autoencoders. We develop a library of autoencoder models and perform a **systematic validation** on a range of synthetic and experimental datasets against standard methods for unmixing. Our results show that autoencoders consistently provide **more accurate, robust and efficient unmixing**.

Background

Raman spectroscopy is a powerful optical modality that facilitates non-destructive, label-free molecular characterisation via the analysis of inelastic scattering of light from matter.

Problem: Many important applications entail the analysis of complex mixtures of molecular species coexisting and interacting at micro- and nanoscales.

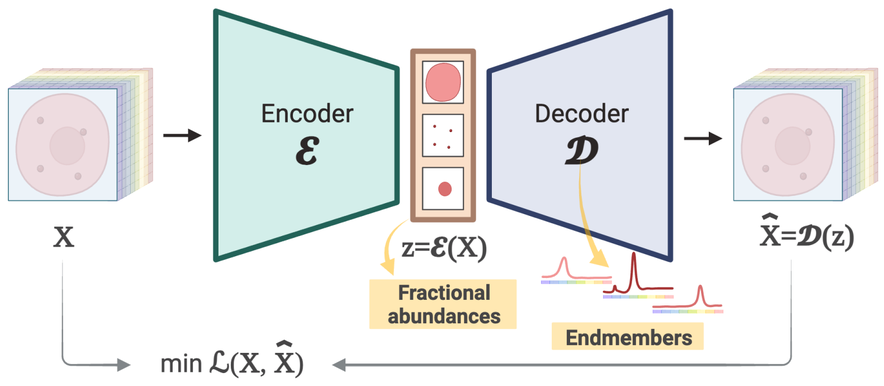


Hyperspectral unmixing aims to resolve mixed signals by identifying the individual components present (*endmembers*) and quantifying their proportions (*fractional abundances*).

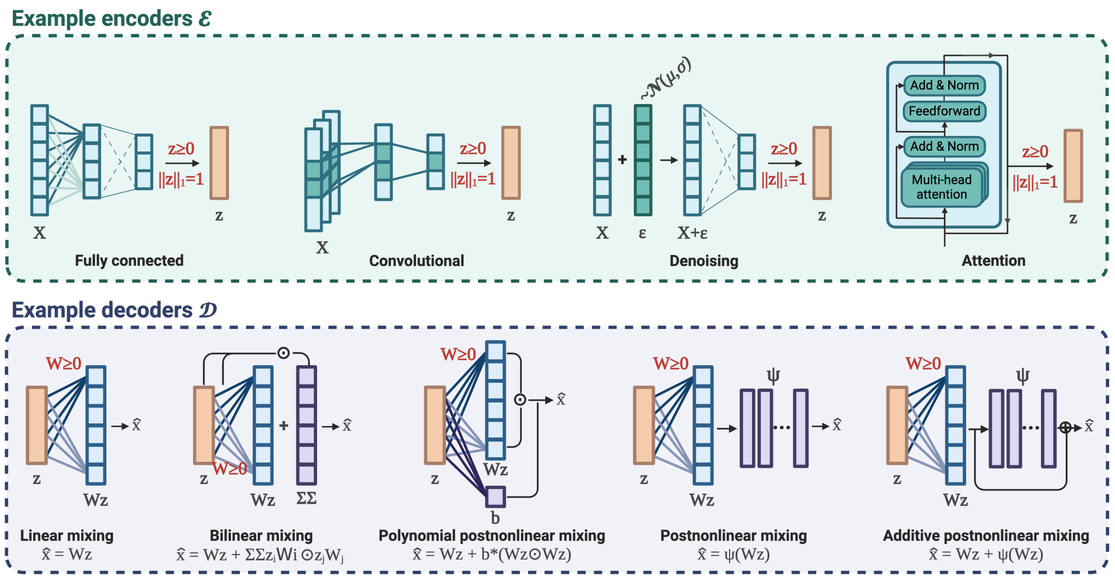
Standard methods for unmixing, such as N-FINDR[1] and Vertex Component Analysis (VCA)[2] for endmember identification, and Non-negative Least Squares (NNLS)[3] and Fully Constrained Least Squares (FCLS)[4] for abundance estimation, have many limitations: restricted to linear unmixing (i.e. X=Ma); lack robustness to data artefacts abundant in experimental Raman data (e.g. dark noise, baseline variations, cosmic spikes); rely on additional assumptions (e.g. endmembers present as 'pure pixels' in data); and can become computationally demanding.

Raman unmixing autoencoders

We approach **Raman unmixing as a self-supervised autoencoder (AE) learning problem**. During training, the decoder learns endmember signatures, while the encoder learns how to derive abundances.



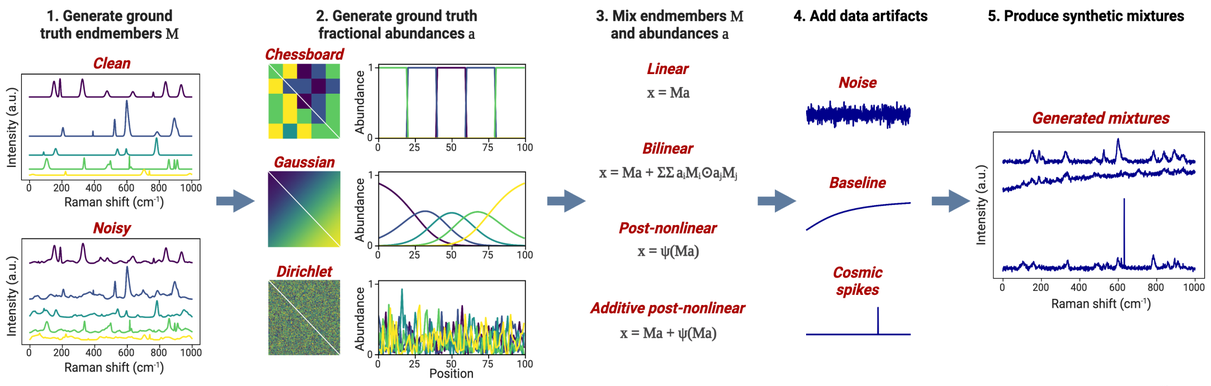
Encoders can incorporate concepts from representation learning to improve feature extraction and abundance estimation. **Decoders** can be structured to model different linear and non-linear mixing models. **Physics constraints** can be built into the model architecture (highlighted in red).



Study objective: Conduct a systematic validation of the approach against conventional methods for unmixing on a range of synthetic and experimental Raman spectroscopy datasets.

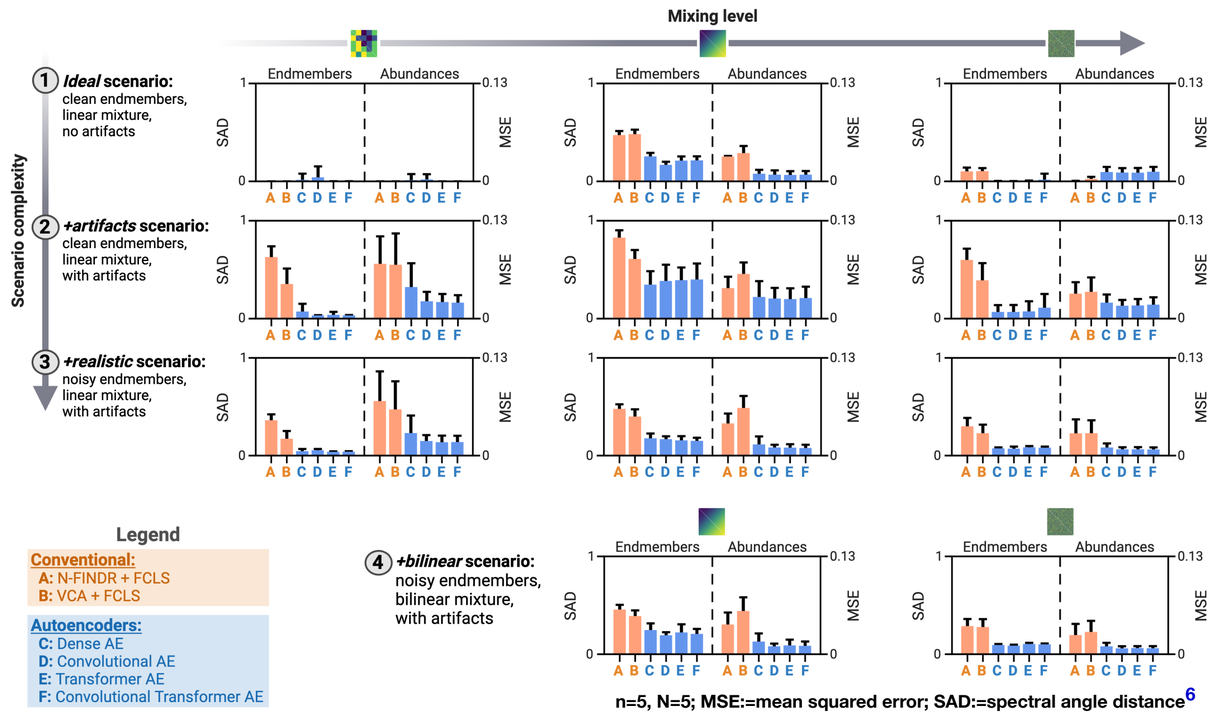
Synthetic Raman mixture generator

Data generation: We developed a Raman mixture generator (available within RamanSPy[5]), which allows us to create synthetic datasets of variable complexity with known ground truth. This enables us to quantitatively benchmark the performance of different methods for unmixing.



Validation on synthetic Raman mixtures

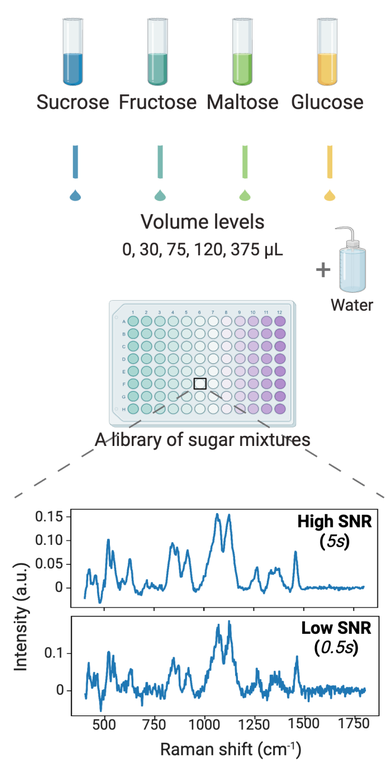
Benchmark results: Using our generator, we created synthetic datasets of increasing complexity. We tested 4 autoencoder models developed in-house, and showed that they consistently outperform conventional methods for unmixing across virtually all datasets and mixture settings.



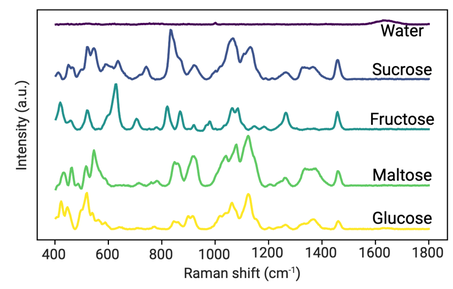
n=5, N=5; MSE:=mean squared error; SAD:=spectral angle distance[6]

Validation on experimental data from sugar mixtures

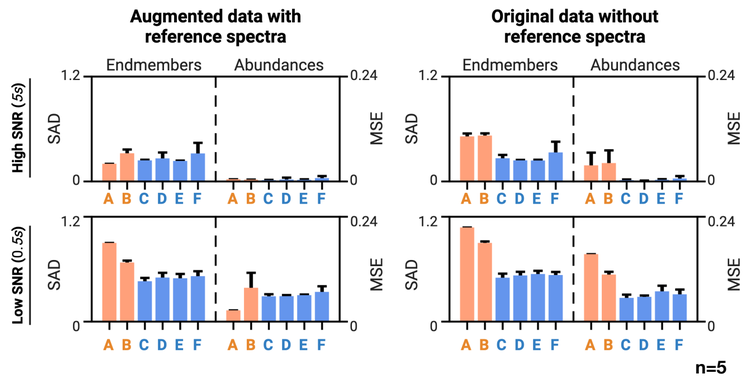
Data: We acquired high and low signal-to-noise ratio (SNR) measurements from a library of 240 sugar mixtures.



Ground truth: Abundances were calculated based on the experimental concentrations, and endmember signatures were derived from reference spectra from pure solutions.



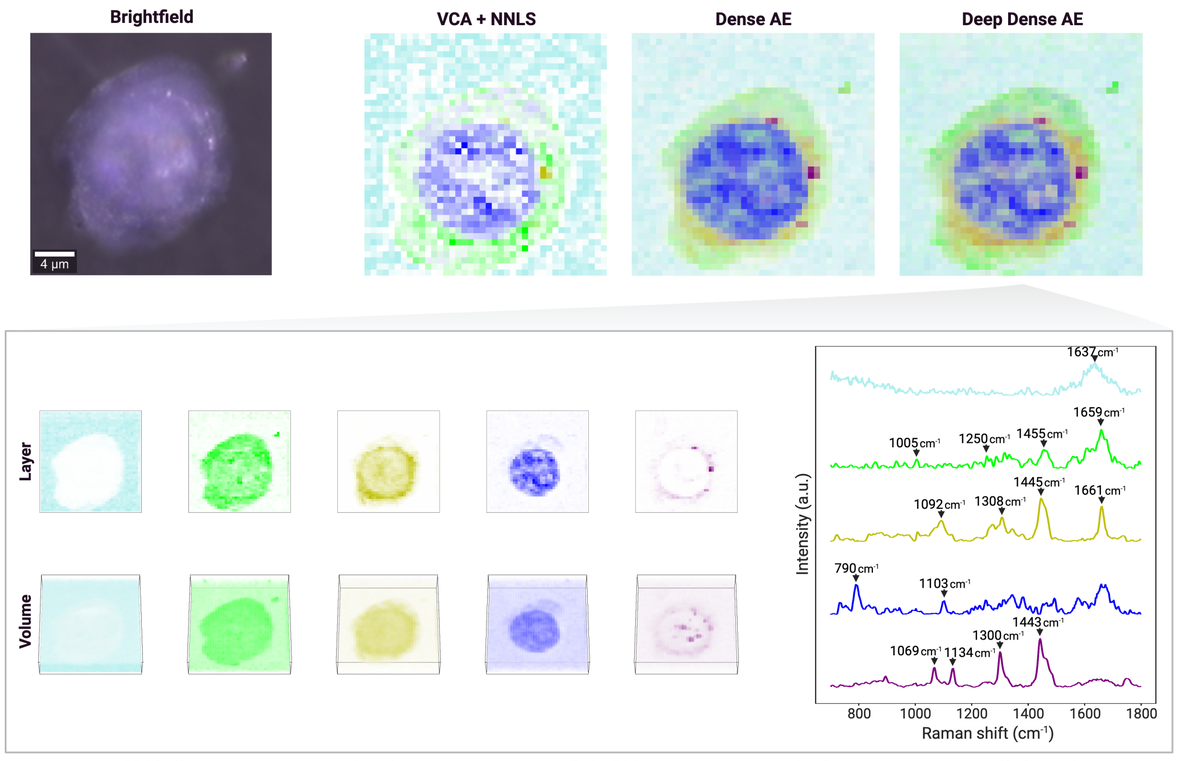
Benchmark results: Our autoencoder models achieved more precise unmixing, especially in scenarios where there are no 'pure pixels' and/or high levels of noise.



Volumetric Raman imaging of a cell

Data: Volumetric Raman imaging scan of a human leukaemia monocytic cell[7].

Results: Our autoencoder models provided improved biochemical characterisation, identifying deoxyribonucleic acid (DNA), proteins, triglycerides (TAGs), phospholipids (PLPs) and cholesterol esters.



Acknowledgements

D.G. and G.P. are supported by UK Research and Innovation [UKRI Centre for Doctoral Training in AI for Healthcare grant number EP/S023283/1]. A.F.G. acknowledges support from the Schmidt Science Fellows, in partnership with the Rhodes Trust. S.V.P. acknowledges support from the Independent Research Fund Denmark (0170-00011B). R.X. and M.M.S. acknowledge support from the Engineering and Physical Sciences Research Council (EP/P00114/1 and EP/T020792/1). M.M.S. acknowledges support from the Royal Academy of Engineering Chair in Emerging Technologies award (CiET2021\94) and the Bio Innovation Institute. M.B. acknowledges support from the Engineering and Physical Sciences Research Council (EP/N014529/1, funding the EPSRC Centre for Mathematics of Precision Healthcare at Imperial, and EP/T027258/1). Figures were assembled in BioRender.

References

[1] Winter ME. *Imaging spectrometry V* **1999**, 3753, 266–275.
[2] Nascimento JM, Dias JM. *IEEE Transactions on Geoscience and Remote Sensing* **2005**, 43, 898–910.
[3] Lawson CL, Hanson RJ. *Solving least squares problems.* Society for Industrial and Applied Mathematics, **1995**.
[4] Heinz DC, Chang C-I. *IEEE Transactions on Geoscience and Remote Sensing* **2001**, 39, 529–545.
[5] Georgiev D, Pedersen SV, Xie R, Fernández-Galiana A, Stevens MM, Barahona M. *Analytical Chemistry* **2024**, 96, 8492–8500.
[6] Kruse FA, Lefkoff AB, Boardman YJ, Heidebrecht KB, Shapiro AT, Barloon PJ, Goetz AF. *Remote Sensing of Environment* **1993**, 44, 145–163.
[7] Kallepitis C, Bergholt MS, Mazo MM, Leonardo V, Skaalure SC, Maynard SA, Stevens MM. *Nature Communications* **2017**, 8, 14843.

IMPERIAL

FHealth

CMPH Centre for Mathematics of Precision Healthcare

THE STEVENS GROUP

UNIVERSITY OF OXFORD