Learning the Crystallographic Phase Problem

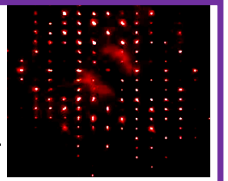
Sarah Scripps, Dr James Cumby, Dr Sohan Seth
s1904252@ed.ac.uk

Skills 4 Scientists

Sponsored by: optibrium, dotmatics knowledge solutions

X-ray diffraction and the phase problem:

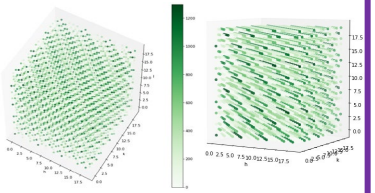
X-ray crystallography is the most widespread technique in determination of crystal structures.
Diffraction is a geometric phenomenon where electromagnetic radiation is scattered depending on the distances between objects.
The interference of the scattered waves and resulting diffraction pattern is determined by the phase difference between them.
In order to explicitly calculate the electron density map from the diffraction pattern the phases of the scattered rays must be known.
However, the phase difference cannot be cannot be measured.
This is known as the phase problem.



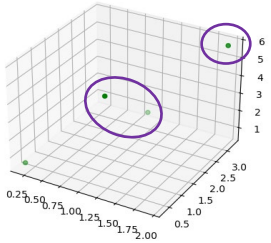
Diffraction pattern of a crystal[1]

1. G. Harburn, C.A. Taylor, T.R. Welberry. Atlas of Optical Transforms. G. Bell and Sons. London (1975).
2. Scikit-learn: Machine Learning in Python, Pedregosa, F. et al, Journal of Machine Learning Research, 12, 2825--2830, 2011
3. Çiçek, Özgün & Abdulkadir, Ahmed & Lienkamp, Soeren & Brox, Thomas & Ronneberger, Olaf. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.

Input Diffraction patterns

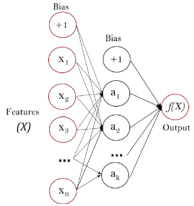


Target
Real space structure



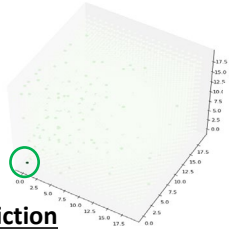
Sklearn MLP Regressor:

MLP Regressor is a neural network which implements a multi-layer perceptron and trains using backpropagation.



MLP architecture[2]

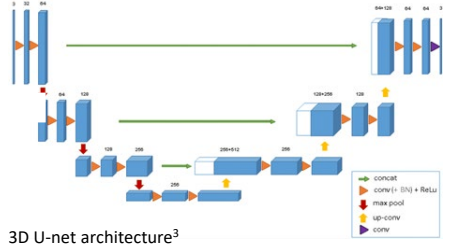
Prediction



- The predictions show some clustering of higher intensities, with the highest intensity peak at the origin.
- Origin peak is a constant fraction of the predicted total electron density for the structure.
- Predicted total electron density is several times greater (~ x2) than the true total.
- Current method to resolve the phase problem setting all phases to 0 also produces a peak at the origin.

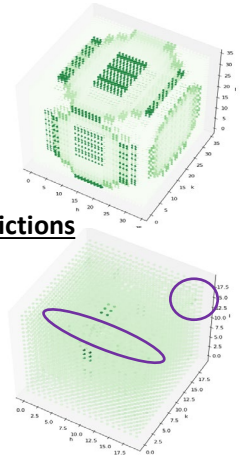
3d U-net:

3D U-net is a convolutional neural network, currently used in segmentation of biological images.



3D U-net architecture[3]

Predictions



- With small training sets, 26-54 structures.
- Predictions all the same octagon based pattern
- Predictions are highly symmetric, rather than random noise.
- Manually reducing the intensity of the visible octagon shows some clusters of higher intensity.
- Some clusters seem to be resolving around atomic coordinates.

THE UNIVERSITY OF EDINBURGH

AI³ Science Network+

PSDS PHYSICAL SCIENCES DATA-SCIENCE SERVICE

RSC INTEREST GROUP CHEMICAL INFORMATION AND COMPUTER APPLICATIONS GROUP