

## How machine learning can supplement traditional quality indicators - and the human eye: A case study

Kristopher Nolte<sup>1</sup>, Yunyun Gao<sup>1</sup>, Sabrina Stäß<sup>1</sup>, Philip Kollmannsberger<sup>2</sup>, Andrea Thorn\*<sup>1</sup>

<sup>1</sup>*Institut für Nanostruktur und Festkörperphysik, Universität Hamburg, Luruper Chaussee 149 (Bldg. 610 - HARBOR), Germany,*

<sup>2</sup>*CCTB, Julius-Maximilians-Universität Würzburg, Germany*

\**andrea.thorn@uni-hamburg.de*

Detecting the ice diffraction artifacts in single-crystal datasets can be very difficult once the data have been integrated, scaled and merged. Automatic tools are available in CTRUNCATE [1], phenix.xtriage [2] and AUSPEX [3]. Recently, the AUSPEX icfinder score was improved by Moreau and colleagues [4]. Automatic recognition of these artifacts would be highly beneficial as macromolecular structure determination can be negatively impacted or even completely hindered by ice diffraction, but remains difficult.

In 2017, we have shown that inspection of plots of merged intensities against resolution permit an easy identification of ice ring contamination in integrated data sets - by eye. However, this approach could be matched by automatic routines. This has led us to attempt identification using convolutional neural networks, which are exceptionally suited to classification of multi-dimensional arrays because they can retain spatial information of the input.

Here, we present our results to employ convolutional neural networks to detect ice artefacts in processed macromolecular diffraction data, resulting in a new automatic detection called “Helcaraxe”. which outperforms previous indicators. We will also discuss the scope this may offer for the structural biology community to tap into the vast amount of data the field has accumulated in 50 years of deposition to the Protein Data Bank.

- [1] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP 4 suite and current developments. *Acta Cryst.* D67, 235–242. <https://doi.org/10.1107/S0907444910045749>
- [2] Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* D66, 213–221. <https://doi.org/10.1107/S0907444909052925>
- [3] Thorn, A., Parkhurst, J., Emsley, P., Nicholls, R. A., Vollmar, M., Evans, G., & Murshudov, G. N. (2017). AUSPEX: A graphical tool for X-ray diffraction data analysis. *Acta Cryst.* D73, 729–737. <https://doi.org/10.1107/S205979831700969X>
- [4] Moreau, D. W., Atakisi, H., & Thorne, R. E. (2021). Ice in biomolecular cryocrystallography. *Acta Cryst.* D77, 540–554. <https://doi.org/10.1107/S2059798321001170>

**Keywords:** machine learning, quality indicators, data processing, ice rings, X-ray data