The Journal of Open Source Software

mlgidGUI - an annotation program for 2D scattering data

Constantin Völter[1*], **Vladimir Starostin**[2*], **Mikhail Romodin**[1], **Ekaterina Kneschaurek**[1], **Dmitry Lapkin**[1], **Alexander Hinderhofer**[1], and **Frank Schreiber**[1]

1 Institute of Applied Physics – University of Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany **2** Cluster of Excellence "Machine Learning for Science", University of Tübingen, Maria-von-Linden-Str. 6, 72076 Tübingen, Germany * These authors contributed equally.

DOI: 10.21105/joss.08499

Software

- Review
- Repository
- Archive

Editor: Lucy Whalley

Reviewers:

- @stuartcampbell
- @kstenio

Submitted: 17 March 2025
Published: 27 October 2025

Summary

We introduce mlgidGUI, a graphical user interface for the analysis and labeling of 2D scattering data with radial symmetry. It is designed to create grazing-incidence wide-angle X-ray scattering (GIWAXS; or GID, grazing-incidence diffraction) datasets for machine learning (ML) training and testing purposes (Starostin, Munteanu, et al., 2022; Völter et al., 2025). Moreover, it enables the visualization and modification of the outputs produced by such ML models. mlgidGUI is designed for 2D scattering images with radial symmetry and provides tools to streamline the manual annotation of Bragg peaks. It supports the conversion to polar coordinates based on experimental settings, such as beam center position and Q-scale adjustments. mlgidGUI enables direct annotation with rings instead of conventional rectangular boxes. Given the low signal-to-noise ratio of some peaks, the software provides customized contrast settings to enhance the visibility. Additionally, an integrated crystallographic toolkit enables peak position simulations, aiding in the identification of even the weakest Bragg peaks. The export as PASCAL VOC dataset enables the integration to ML pipelines. A ML-based peak detection (Starostin, Pithan, et al., 2022) can provide preliminary peaks to accelerate the annotation.

Statement of need

Grazing-incidence wide-angle X-ray scattering (GIWAXS) is a powerful technique for characterizing crystalline structures on surfaces. It is instrumental in a wide range of applications, including organic photovoltaics and semiconductors (Banerjee et al., 2020; Feidenhans'l, 1989). Typical GIWAXS measurements at synchrotron facilities generate hundreds of thousands of diffraction images per day. While recent ML tools aid in analyzing these vast datasets, proper benchmarking on annotated datasets is crucial for ensuring robust scientific applications.

In our recent paper (Völter et al., 2025), we developed a methodology for benchmarking GIWAXS data and published the first annotated dataset. Here, we open-source the software used for annotating this data to further support efforts in improving, standardizing, and automating GIWAXS analysis. With mlgidGUI, we support researchers in producing GIWAXS datasets with less effort and higher-quality annotations. Figure 1 shows the appearance of mlgidGUI:

The Journal of Open Source Software

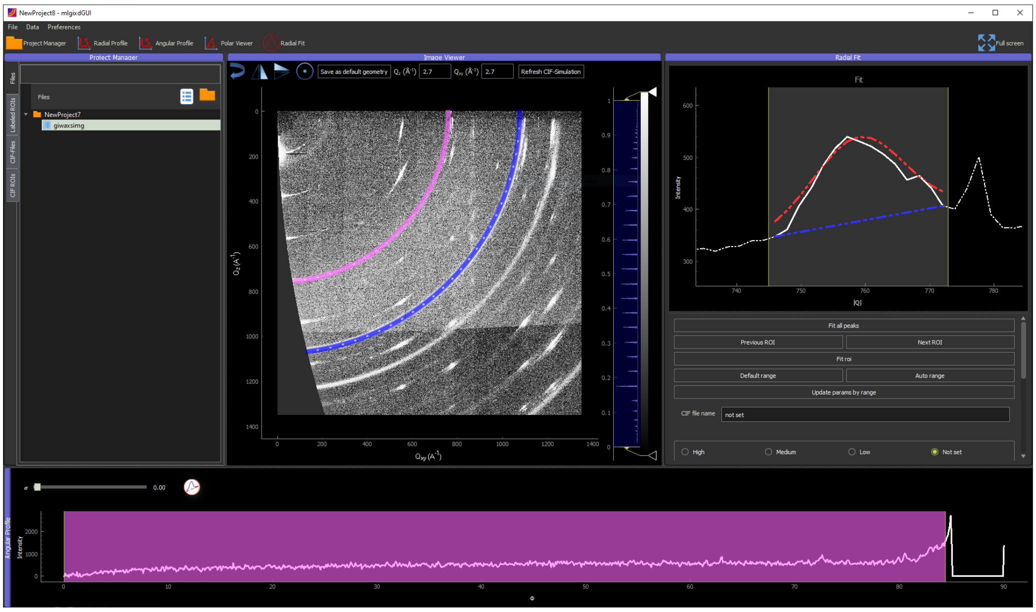

Figure 1: Appearance of the mlgidGUI: Project Manager on the left, Angular Profile at the bottom, Image Viewer in the center, and Radial Fit widget on the right.

General workflow

mlgidGUI is intended for GIWAXS images which are already converted to reciprocal space. Optionally, the converted image can be analyzed beforehand by an automated peak detection (Starostin, Munteanu, et al., 2022) to accelerate the annotation process. The contrast correction is enabled by default to aid in finding peaks with low intensities. Next, the user has to set the geometry settings to ensure the correct beam center position and Q-range. Then, diffraction peaks are added manually with optional preferred crystallographic orientation. Each peak can be fitted with a chosen profile and background to ensure accurate position, intensity, and width. According to the proposed methodology in (Völter et al., 2025), each peak is assigned with a confidence label - low, medium, or high. Additionally, an integrated crystallographic toolkit can be used to superimpose the measured data with the simulated diffraction pattern of a provided structure. These simulated peaks help the user in identifying weak peaks. The intermediate result of the labeling process is saved automatically. When an image is fully labeled, the resulting annotation can be saved as a new HDF5 file, added to an existsing HDF5 dataset or exported as PASCAL VOC dataset. Figure 2 outlines the entire workflow: Raw detector data is first converted and optionally undergoes peak detection before it is visualized and annotated with the mlgidGUI. The resulting datasets from mlgidGUI can be used for training or validation purposes of the machine learning model.

The Journal of Open Source Software

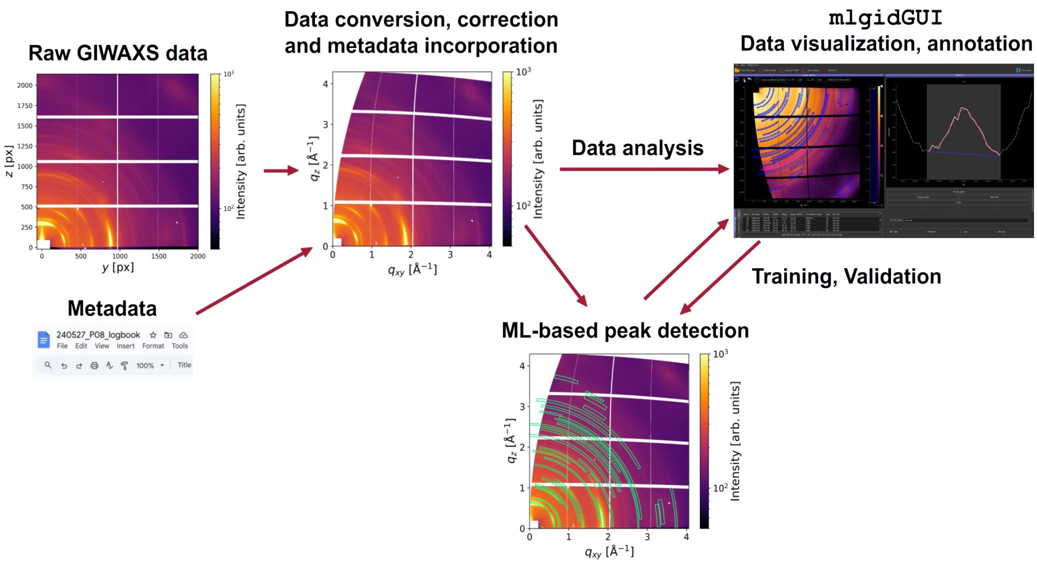

Figure 2: End-to- end workflow from raw detector output through data conversion followed by visualization and annotation for the ML-based peak detection

Related work

Several software packages for the analysis of scattering data are available, (Filik et al., 2017; Hammersley, 2016; Jiang, 2015) but have limitations. Some software is designed only for 1D data, and some are challenging to use or are not optimized for labelling large batches of data. Most importantly, these tools lack compatibility with ML integration, as they do not support the conversion to polar coordinates or the export of datasets in ML-compatible formats. Conversely, object detection annotation software for ML exist (CVAT.ai Corporation, 2023; Dutta & Zisserman, 2019; Tzutalin, 2015; Wada, n.d.), but lack specific adaptations for GIWAXS data. For example, they do not allow labelling in reciprocal space and have no crystallographic toolkit to detect weak peaks.

Author contributions

The authors have contributed to this work in accordance with the CRediT (Contributor Roles Taxonomy) author statement:

- **Constantin Völter**: Conceptualization, Methodology, Software, Validation, Writing - Original Draft
- **Vladimir Starostin**: Conceptualization, Methodology, Software, Validation, Writing - Review & Editing
- **Mikhail Romodin**: Conceptualization, Software, Data Curation, Writing - Review & Editing
- **Ekaterina Kneschaurek**: Conceptualization, Validation, Resources, Data Curation, Writing - Review & Editing
- **Dmitry Lapkin**: Conceptualization, Validation, Resources, Data Curation, Writing - Review & Editing, Supervision, Project administration
- **Alexander Hinderhofer**: Conceptualization, Validation, Resources, Data Curation, Writing - Review & Editing, Supervision, Project administration
- **Frank Schreiber**: Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645. F. Schreiber is a member of the Machine Learning Cluster of Excellence, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645. Funded by the BMBF and the DAPHNE4NFDI.

References

Banerjee, R., Kowarik, S., & Schreiber, F. (2020). Grazing Incidence X-ray Scattering Techniques to Study Growth Mechanism of Organic Films. In *Advanced Characterization of Nanostructured Materials: Vols. Volume 21* (pp. 49–95). WORLD SCIENTIFIC. https://doi.org/10.1142/9789811231513_0002

CVAT.ai Corporation. (2023). *Computer Vision Annotation Tool (CVAT)* (Version 2.25.0). https://github.com/cvat-ai/cvat

Dutta, A., & Zisserman, A. (2019). The VIA annotation software for images, audio and video. *Proceedings of the 27th ACM International Conference on Multimedia*. https://doi.org/10.1145/3343031.3350535

Feidenhans'l, R. (1989). Surface structure determination by X-ray diffraction. *Surface Science Reports*, *10*(3), 105–188. https://doi.org/10.1016/0167-5729(89)90002-2

Filik, J., Ashton, A. W., Chang, P. C. Y., Chater, P. A., Day, S. J., Drakopoulos, M., Gerring, M. W., Hart, M. L., Magdysyuk, O. V., Michalik, S., Smith, A., Tang, C. C., Terrill, N. J., Wharmby, M. T., & Wilhelm, H. (2017). Processing two-dimensional X-ray diffraction and small-angle scattering data in DAWN 2. *Journal of Applied Crystallography*, *50*(3), 959–966. https://doi.org/10.1107/S1600576717004708

Hammersley, A. P. (2016). FIT2D: A multi-purpose data reduction, analysis and visualization program. *Journal of Applied Crystallography*, *49*(2), 646–652. https://doi.org/10.1107/S1600576716000455

Jiang, Z. (2015). GIXSGUI: A MATLAB toolbox for grazing-incidence X-ray scattering data visualization and reduction, and indexing of buried three-dimensional periodic nanostructured films. *Journal of Applied Crystallography*, *48*(3), 917–926. https://doi.org/10.1107/S1600576715004434

Starostin, V., Munteanu, V., Greco, A., Kneschaurek, E., Pleli, A., Bertram, F., Gerlach, A., Hinderhofer, A., & Schreiber, F. (2022). Tracking perovskite crystallization via deep learning-based feature detection on 2D X-ray scattering data. *Npj Computational Materials*, *8*(1), 1–9. https://doi.org/10.1038/s41524-022-00778-8

Starostin, V., Pithan, L., Greco, A., Munteanu, V., Gerlach, A., Hinderhofer, A., & Schreiber, F. (2022). End-to-End Deep Learning Pipeline for Real-Time Processing of Surface Scattering Data at Synchrotron Facilities. *Synchrotron Radiation News*, *35*, 1–7. https://doi.org/10.1080/08940886.2022.2112499

Tzutalin. (2015). *LabelImg*. Free Software: MIT License. https://github.com/tzutalin/labelImg

Völter, C., Starostin, V., Lapkin, D., Munteanu, V., Romodin, M., Hylinski, M., Gerlach, A., Hinderhofer, A., & Schreiber, F. (2025). *Benchmarking deep learning for automated peak detection on GIWAXS data*. *58*. https://doi.org/10.1107/S1600576725000974

Wada, K. (n.d.). *Labelme: Image Polygonal Annotation with Python*. https://doi.org/10.

5281/zenodo.5711226