MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
12 March 2026

REVISED
2 June 2026

ACCEPTED FOR PUBLICATION
17 June 2026

PUBLISHED
10 July 2026

PAPER

Automated peak detection in grazing-incidence diffraction using transformers

Constantin Völter, Ainur Abukaev, Ekaterina Kneschaurek, Mikhail Romodin, Oleg Gorobtsov, Alexander Hinderhofer*, Dmitry Lapkin and Frank Schreiber*

Institute of Applied Physics, University of Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany
* Authors to whom any correspondence should be addressed.

E-mail: alexander.hinderhofer@uni-tuebingen.de, frank.schreiber@uni-tuebingen.de, constantin.voelter@uni-tuebingen.de, ainur.abukaev@uni-tuebingen.de, ekaterina.kneschaurek@uni-tuebingen.de, mikhail.romodin@uni-tuebingen.de, oleg.gorobtsov@uni-tuebingen.de and dmitry.lapkin@uni-tuebingen.de

Keywords: Bragg peak detection, GIWAXS, deep learning, transformer

Supplementary material for this article is available online

Abstract

The adoption of grazing-incidence wide-angle x-ray scattering (GIWAXS) has expanded rapidly, as it has proven invaluable for extracting a wealth of structural information from thin films and surfaces on the atomic and molecular scale. The amount of scattering data collected in such experiments creates new challenges associated with relatively slow conventional data analysis due to the need for human input. Existing machine learning-based Bragg-peak detection approaches have demonstrated the potential of automation, but significant challenges such as weak peaks, peaks on top of Debye–Scherrer rings, and difficulties in differentiating closely spaced peaks remain. We introduce a powerful model for GIWAXS peak detection based on a visual transformer that achieves a performance strongly surpassing existing models. We evaluate our new approach using a diverse set of labelled, experimental GIWAXS patterns from perovskite and organic thin films. Notably, we demonstrate for the first time the reliable detection of Bragg peaks superimposed on other peaks and Debye–Scherrer rings, a significant improvement in detecting low-intensity peaks and a substantially stronger overall performance. In addition, we evaluate both pre- and postprocessing latency, confirming that the overall analysis speed is sufficient for real-time applications. With its ability to deliver a fast and high-performance analysis, our approach enables new types of GIWAXS measurements, including closed-loop experiments, that were previously limited by analysis bottlenecks. The studied model could potentially also be applied to other datasets with a large dynamic range and weak signal-to-noise ratio.

1. Introduction

Grazing-incidence wide-angle x-ray scattering (GIWAXS) is a key method for characterising surface and thin film structures at the atomic and molecular level [1, 2]. A major advantage of GIWAXS over many other methods is that it allows for non-invasive *in-situ* and *in-operando* studies on thin films with high temporal resolution, even in specific environments [3–8]. Its adoption has expanded rapidly, driven by major technological leaps in synchrotron facilities, including more brilliant x-ray sources and detectors with higher temporal and spatial resolution [9, 10]. GIWAXS measurements allow to extract structural information from thin films and surfaces, including crystal structure, crystalline domain orientation, crystallinity, and phase composition [8, 11–14]. Here, accurate peak detection is essential, since this information is encoded in the positions, widths, and intensities of Bragg peaks in the 2D scattering pattern. Conversely, inaccurate detections such as missed peaks, false positives, or merged closely spaced features, propagates through the entire analysis pipeline and compromises the reliability of all downstream results. The peak detection is complicated by the inherent properties of GIWAXS data. The data simultaneously contains sharp Bragg peaks from textured crystalline domains and Debye–Scherrer rings from

randomly oriented polycrystalline phases, both superimposed on strong substrate and ambient background scattering. Features span a large dynamic range, and vary strongly in size from only a few pixels to Debye–Scherrer rings spanning the whole azimuthal extent.

Machine learning (ML) has been applied to grazing-incidence scattering data along several directions, it provides a powerful solution, thanks to its capacity for parallel data processing and ability to extract new conceptual insights from large datasets [15–24]. Early work employed ML to improve the efficiency of synchrotron beam time through autonomous, Bayesian optimisation-driven GIXD mapping of thin-film libraries [25]. Separately, convolutional neural networks were shown to successfully classify thin-film morphologies and scattering regimes directly from GIWAXS and GISAXS images [15, 18, 26]. Beyond classification, ML has also been applied to data reconstruction tasks: Chavez *et al* [27] demonstrated that neural networks, such as U-Nets and mixed-scale dense networks can reconstruct detector gaps in experimental grazing-incidence scattering data. While these approaches showed that learned representations can capture relevant structural information from 2D scattering patterns, they address classification and reconstruction rather than the precise localisation of individual scattering features. The more demanding task of automated peak detection in GIWAXS was first addressed by Starostin *et al* [28], who introduced a modified Faster R-CNN architecture adapted to the GIWAXS geometry. For this approach, the tracking of perovskite crystallisation in time-resolved GIWAXS data was demonstrated. A subsequent benchmarking study [29] evaluated this model against a classical region-growing baseline, confirming substantial performance gains from the ML-based approach, particularly for adjacent and low-intensity features. The Bragg peak detection of Starostin *et al* [28] is integrated into the `mlgid` pipeline [30], which provides a fully automated workflow from raw detector images to structure identification (figure 1). Despite this progress, several challenges remain unresolved. Non-maximum suppression (NMS) is a structural necessity in region-based detectors such as Faster R-CNN, and is applied at multiple stages to suppress detections close to higher-confidence ones. This is especially problematic in GIWAXS, where Bragg peaks frequently lie on top of Debye–Scherrer rings, and where multiple peaks from structurally complex or multiphase samples may overlap. Additionally, the detection of low-intensity peaks, which blend into the scattering background or are obscured by the contrast correction, remains an unsolved issue [29]. These limitations directly affect the quality of all downstream analyses, motivating the development of a more capable detection model. In this work, we address these limitations by employing a transformer-based object detection model, DINO-DETR [31], adapted to the properties of GIWAXS data.

A transformer-based model, such as DETR [33], offers several advantages over Faster RCNN: It simplifies the object detection pipeline by removing manually parameterised components, such as anchor boxes and region proposals. Unlike previous models, DETR uses bipartite matching and views object detection as a set prediction task. While DETR does not require NMS by design, we optionally apply a NMS step as post-processing to reduce duplicate detections on GIWAXS data, as described in section 3.

We employ DINO-DETR [31], an advancement of DETR that achieves higher average precision (*AP*) through multiple modifications: It uses features from multiple scales to better detect small objects; the contrastive denoising training adds negative samples to stabilise the training; the deformable attention increases the computational efficiency; and the mixed query selection helps to better initialise the queries (figure 3). Since GIWAXS features are elongated in the azimuthal direction, we modified the standard square attention windows to be elongated in the vertical dimension as described in section 2.2. To isolate the effect of our modification, we conduct an ablation study by evaluating three additional variants with different backbones: the standard Swin-L backbone with unmodified square attention windows, the standard ResNet-50 [34] backbone, and the modified ResNet-50 from [28]. We trained all variants on simulated data and evaluated them on experimental GIWAXS patterns. The resulting peak detection model outperforms the existing Faster R-CNN-based approach. Our work, including the model, is published on GitHub and integrated into the `mlgid` [30] pipeline for direct use by the research community.

2. Method

2.1. Data preparation

GIWAXS patterns contain objects that differ significantly from those in benchmark datasets such as Microsoft COCO: Common Objects in Context [35]. Due to the symmetry, in most cases, the scattering features in the measured patterns represent arcs (i.e. Debye–Scherrer rings) or their segments (i.e. Bragg peaks) of increased intensity (figure 1(b)). The data is brought into a shape better suited for object detection by converting it into quasi-polar coordinates, which align better with the rectangular convolutions and attention windows used for our model. In this representation, the horizontal axis corresponds to the radial dimension and the vertical axis represents the azimuthal angle as described further

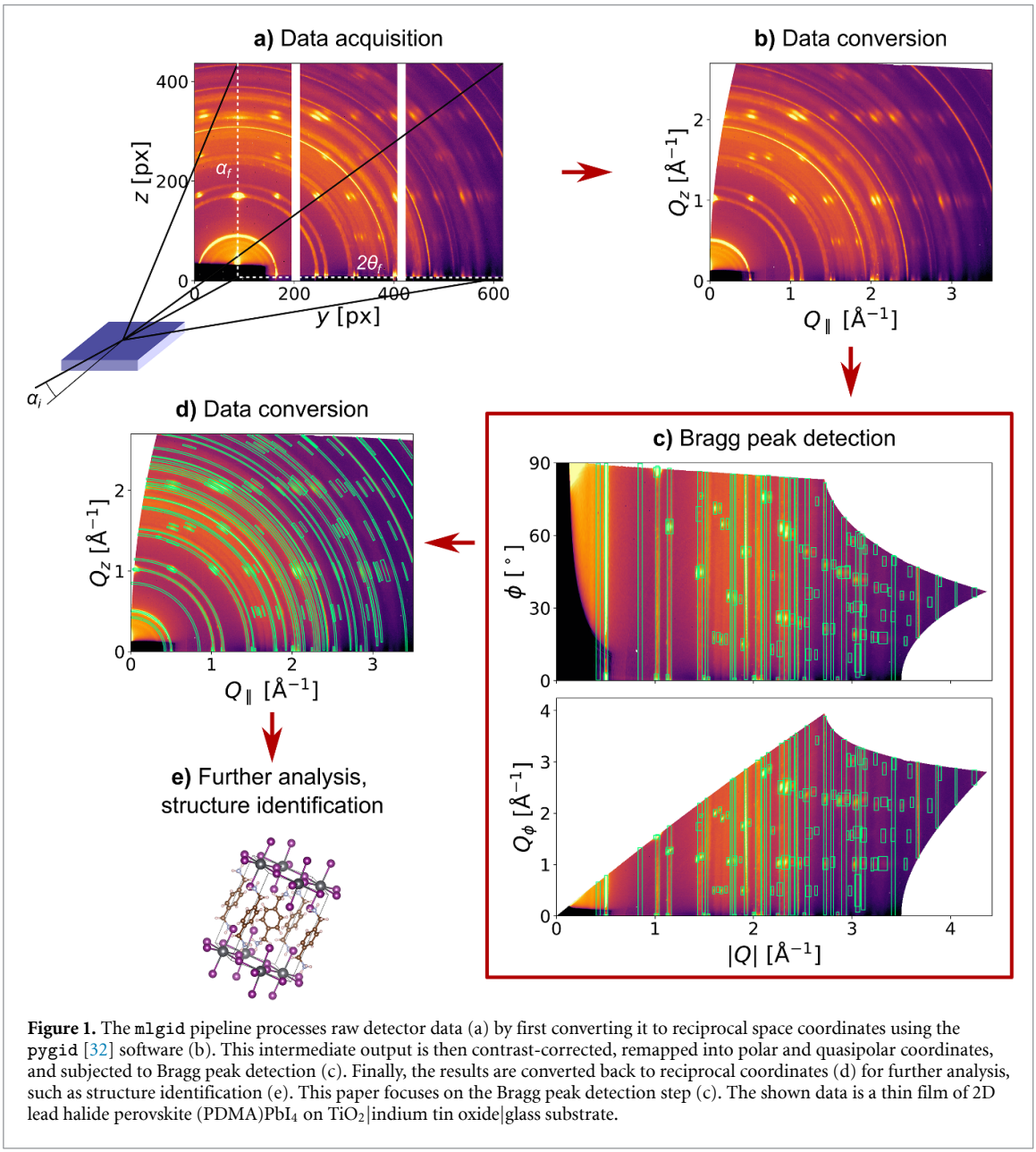

Figure 1. The `mlgid` pipeline processes raw detector data (a) by first converting it to reciprocal space coordinates using the `pygid` [32] software (b). This intermediate output is then contrast-corrected, remapped into polar and quasipolar coordinates, and subjected to Bragg peak detection (c). Finally, the results are converted back to reciprocal coordinates (d) for further analysis, such as structure identification (e). This paper focuses on the Bragg peak detection step (c). The shown data is a thin film of 2D lead halide perovskite (PDMA)PbI_4 on TiO_2|indium tin oxide|glass substrate.

in section 2.1.2. In these coordinates, the features become oriented along the vertical (azimuthal) direction (figure 1(c)), which makes them compatible with the rectangular bounding boxes.

2.1.1. Data simulation

To overcome the scarcity of labelled experimental data for training, we utilised a simulation built upon the PyTorch-GPU implementation described in [28]. The simulation covers a wide range of realistic GIWAXS patterns, sampling 2–400 objects per image, including scattering background, diverse artefacts and Bragg peaks superimposed on Debye–Scherrer rings. The patterns are simulated in either polar or quasipolar coordinates with the size of 1024×512 pixels ($W \times H$, width corresponds to the radial dimension, and height to the azimuthal one). The 2D-Gaussian-shaped Bragg peaks and 1D-Gaussian-shaped Debye–Scherrer rings are created in arbitrary positions, offering the advantage of not being bound to a material. The radial peak widths range from 1 to 5 pixels, while the radial ring widths vary between 2 and 15 pixels. Azimuthal positions extend from 0 to 110% of the image height to account for off-detector features, with azimuthal extents of peaks spanning 1–10 pixels and ring extents in the azimuthal direction of 50–512 pixels. The ground-truth bounding boxes are defined at the generated peak/ring positions with their sizes equal to the corresponding full-width-half-maximum (FWHM) widths of the Gaussians. Ring and peak intensities are randomised within the ranges of 4%–100% and 10%–100% of the maximum intensity, respectively. A variety of different noise types is added to the images. With

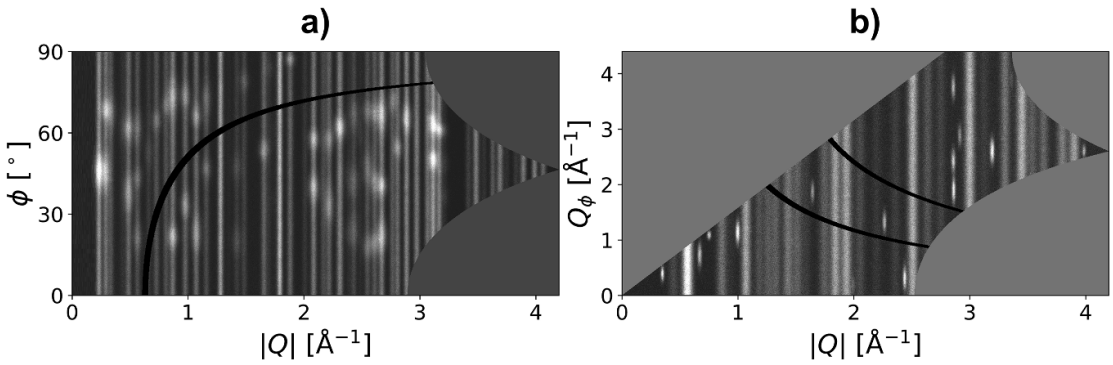

Figure 2. (a, b) Exemplary simulated patterns used for training: (a) in the polar geometry with prominent peaks atop of rings, (b) in the quasipolar geometry with a simulated glass background. Both examples show peak intensity variations similar to those observed in experimental patterns, and incorporate histogram equalisation.

a probability of 50%, the Poisson noise emulating the detector counting noise is added to the images. Hot pixels are added to 20% of images with a per-pixel probability of 0.001. Additionally, with a probability of 50% each, the images are modulated with Perlin noise and augmented with substrate scattering features, salt-and-pepper noise and detector gaps. Contrast adjustments and normalisation are performed using logarithmic scaling, histogram equalisation and a smoothing operation. The smoothing step applies a 3×3 low-pass convolution kernel that introduces mild spatial averaging while largely preserving the original intensity distribution. Images also undergo random horizontal or vertical flips, each applied with a probability of 0.3. Exemplary simulated patterns produced by this procedure in polar and quasipolar coordinates, are shown in figures 2(a) and (b), respectively. The background region outside the patterns is assigned a constant intensity value that is randomly sampled from the [0,1] interval for each individual pattern. All simulated images have a size of 1024×512 pixels ($W \times H$) and a single colour channel.

2.1.2. Experimental data conversion and annotation

The raw experimental data should first be converted from the detector pixel coordinates (figure 1(a)) into physically meaningful cylindrical coordinates ($Q_{||}$, Q_z) in reciprocal space (figure 1(b)). There exist various scripts for this data reduction step [36–39]. For the present work, we employed the `pygid` package [32] due to its high performance and seamless integration into the `mlgid` pipeline. At this stage, the solid angle and polarisation corrections, and background subtraction were applied. We then converted the patterns into polar ($|Q|$, ϕ) (figure 1(c)) and quasipolar ($|Q|$, Q_ϕ) (figure 1(c)) reciprocal coordinates defined as:

$$|Q| = \left(Q_z^2 + Q_{||}^2\right)^{1/2}, \quad \phi = \arctan\left(\frac{Q_z}{Q_{||}}\right), \quad Q_\phi = |Q| \arctan\left(\frac{Q_z}{Q_{||}}\right). \tag{1}$$

For the evaluation of the performance of the proposed peak detection approach, the experimental dataset should be manually labelled. However, GIWAXS patterns can contain hundreds of Bragg peaks with varying intensities that may be hard to distinguish from the underlying background signal. Together, these factors make the annotation a complex and time-consuming task that requires expert knowledge [38, 40]. Therefore, we used the `mlgidGUI` software [41] as an efficient approach for labelling and fitting of all scattering features. Each Bragg peak or Debye–Scherrer ring is fitted with a Gaussian profile on top of a linear background. The labelled bounding box widths are determined by the FWHM widths of the fitted Gaussian function.

2.1.3. Histogram equalisation

The intensity ranges and distributions of experimental images vary depending on the exposure time, the material being examined, the detector used, and other measurement conditions [8]. In order to ensure robust performance of the vision-based object-detection models, it is essential to normalise the intensity values to a standardised scale between 0 and 1 [42]. Additionally, we applied histogram equalisation to increase the image contrast and therefore the low-intensity peaks. Our preprocessing pipeline is as follows. We clip 0.1% of pixels with the highest intensity values and 5% of pixels with the lowest values. Then we normalise the remaining values to a [0, 1] interval, and finally apply histogram equalisation to enhance the contrast.

2.2. Architectural adaptations

We transformed both the simulated and experimental data into quasi-polar coordinates which align better with the rectangular convolutions and attention windows used for our model. However, even in these coordinates, the features are remarkably different and diverse from those in standard datasets. For instance, the Debye–Scherrer rings are elongated in the azimuthal (vertical) direction and span over the whole image, while some of the Bragg peaks span only a few pixels. The features also have substantially different intensities and can overlap in the horizontal direction. Therefore, a good performance on public benchmark datasets does not necessarily translate to an effective analysis of GIWAXS data. To account for these features, we modified both the ResNet-50 and the Swin-L backbones of DINO-DETR.

For the ResNet-50, we relied on the modifications presented in [28], which show strong performance through the use of feature maps elongated in the vertical direction. Specifically, the feature map dimensions were adjusted to (128×64×128), (256×32×128), (256×32×128), and (256×16×128) ($C \times W \times H$) for the input patterns with the size of 1024×512 pixels ($W \times H$).

The Swin-L backbone is a hierarchical vision transformer that constructs feature representations through shifted windows. Unlike traditional convolutional networks, Swin-L leverages a multi-stage architecture where image patches are progressively merged, forming a hierarchical representation that captures both fine-grained local details and high-level semantic information. We adapted the attention windows to shapes elongated in the vertical direction, as shown in figure 3, increasing the window dimensions from (12×12) tokens to (12×24) tokens ($W \times H$). Given the model's four-scale design and a token size of 4×4 pixels, these modified windows translate into effective receptive field sizes of (48×96), (96×192), (192×384), and (384×768) pixels ($W \times H$) across the different stages.

DINO-DETR uses a contrastive denoising approach to stabilise the training and accelerate convergence. Here, noise is added to the labels during the training process. One of the noise parameters ($\lambda_{\text{dn–xy}}$) affects the lateral shift of the labels, another ($\lambda_{\text{dn–wh}}$) affects the change in the label sizes. Since Bragg peaks are typically elongated in the azimuthal direction and have a smaller size in the radial direction than objects in the COCO datasets, we found that different parameters are needed for GIWAXS data. Therefore, we chose for $\lambda_{\text{dn–xy}}$ the value 0.05 (compared to 0.2) and 0.8 for ($\lambda_{\text{dn–wh}}$) (compared to 0.2). We trained all models with an initial learning rate of 10^{-5}, a batch size of 2, and 1000 simulated images per epoch. Training proceeded for 350 epochs, reducing the learning rate to 10^{-6} after 280 epochs. We trained all variants on simulated data as described in section 2.1.1. All training hyperparameters and architectural settings, including deviations from the standard DINO-DETR configuration are summarised in table B1 in the supplementary information.

3. Results

To evaluate the performance of our models, we used the metrics described in [29], including the Average Precision based on the overlap in the radial direction (AP_r). All evaluations are performed on a standard workstation with an Intel Core i5-9600K@3.70 GHz and an NVIDIA RTX 2080 Ti GPU with 11 GB of GDDR6 memory. We tested our peak detection solution on two representative datasets, both derived from materials commonly studied in GIWAXS experiments: one containing data from hybrid perovskite films collected during various crystallisation and annealing steps, and the other containing data from organic thin films [43].

3.1. Performance on perovskite data

The perovskite dataset was initially described in [29]. It includes GIWAXS patterns collected *in-situ* during spin-coating and annealing of 2D and 3D perovskite thin films as detailed in [44]. These films feature a range of different compositions, including various cations, anions, and spacer molecules. The GIWAXS patterns capture scattering features from the final perovskite structures, intermediate products, and precursors. All samples were prepared on substrates of indium tin oxide or fluorine-doped tin oxide, often supplemented with titanium oxide layers. The dataset consists of 45 expert-labelled images containing 1815 objects to detect. Full details of the dataset composition, substrates, and beamline configurations are provided in the supplementary information (section A and table A1). Table 1 compares the DINO-DETR model with four different backbones against the Faster R-CNN baseline of [28]. Among all evaluated configurations, our proposed Swin-L backbone with elongated attention windows of size 12×24 achieves the highest performance, reaching an AP_r of 75.5. The three reference configurations, the standard Swin-L with square 12×12 windows, the standard ResNet-50, and the modified ResNet-50 of [28] all fall substantially short of this result. For completeness, the modified Faster R-CNN from

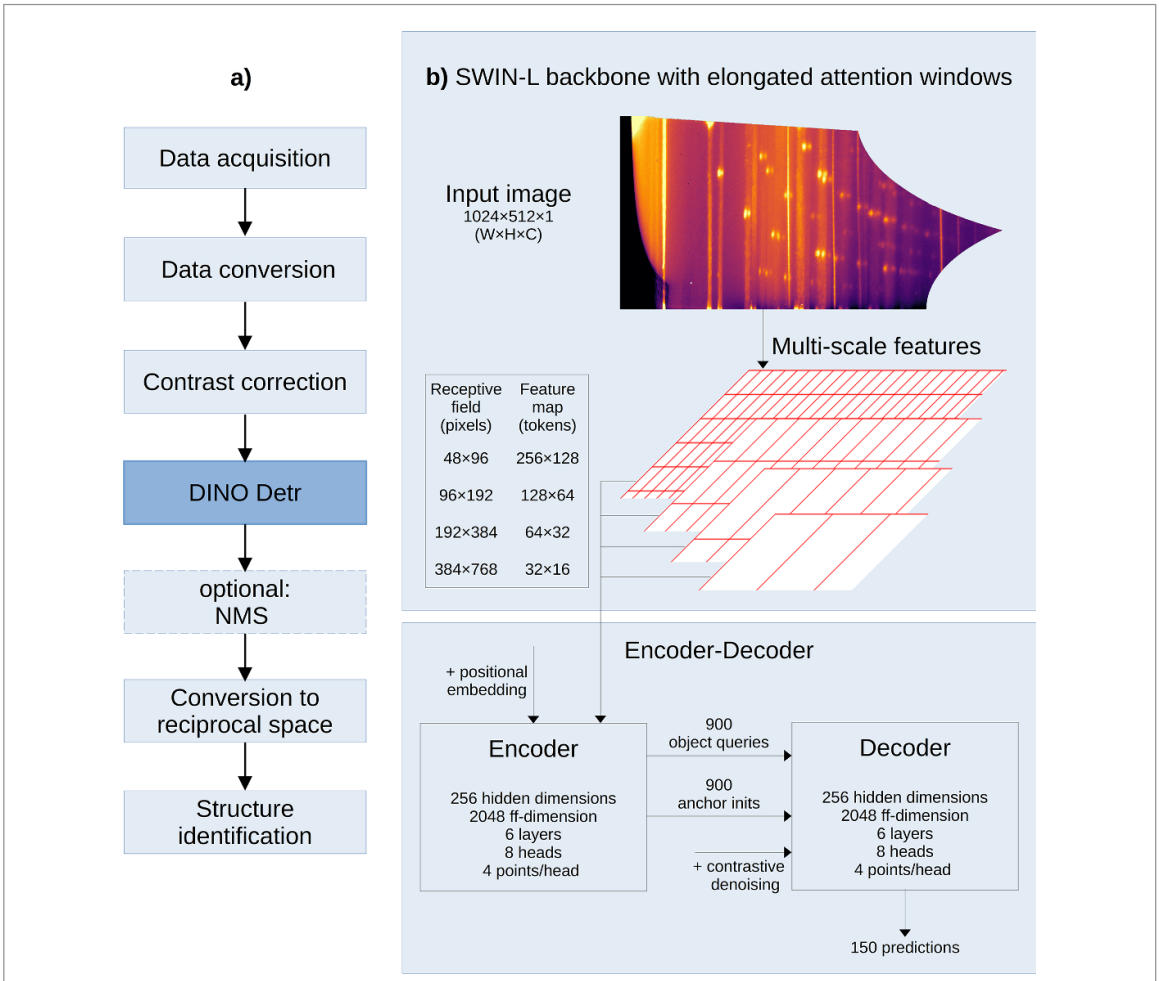

Figure 3. Overview of the DINO-DETR-based peak detection model and its integration into the `mlgid` pipeline. (a) The full analysis workflow, from raw data acquisition and conversion using `pygid` [32] through contrast correction and DINO-DETR for peak detection, to optional post-processing via NMS, conversion to reciprocal space, and structure identification. (b) Architecture of the DINO-DETR model. The input image (1024×512×1 pixels, $W \times H \times C$) is fed into a modified Swin-L backbone with elongated attention windows, producing multi-scale feature maps at four receptive field sizes. The resulting features are passed to a transformer encoder–decoder with 256 hidden dimensions, 2048 feed-forward dimensions, 6 layers, 8 heads, and 4 deformable attention points per head. The decoder operates on 900 object queries initialised from 900 anchor points and is trained with contrastive denoising, which adds negative samples to stabilise the training. NMS is applied as an optional post-processing step; its effect on detection performance is evaluated in table 1. After post-processing, the model outputs 150 peak and ring predictions per image.

Table 1. Comparison of DINO-DETR variants based on #parameters, latency, AP_r, and AP_r without NMS. The standard Swin-L backbone is used as is with a patch size of (4×4) and a window size of (12×12). The modified Swin-L backbone uses a patch size of (4×4) and a window size of (12×24), as described in section 2.2. The results for the modified Faster R-CNN are taken from [28]. All results were obtained for an input size of 512×1024 pixels and a batch size of 1. Bold values indicate the best result.

Model	#Parameters	Latency (ms)	AP_r	AP_r without NMS
DINO + SWIN-L 12×24 windows	218 M	191	**75.5**	**67.3**
DINO + SWIN-L 12×12 windows	217 M	157	74.4	52.3
DINO + ResNet-50	49 M	149	72.7	56.5
DINO + ResNet-50 from [28]	26 M	186	64.1	35.4
ModifiedFaster R-CNN [28]	23 M	149	69.9	—

[28] is also included, DINO-DETR with our modified Swin-L backbone yields a significant gain in AP_r at the cost of only a 50 ms increase in latency.

DETR-based models can be operated without NMS [45]. While our method does not completely remove NMS, it is applied just once during post-processing with a maximum IoU of 0.3. This approach effectively removes duplicate detections while preserving most of the Bragg peaks on the Debye–Scherrer rings as shown in figure 4(e). We also tested our method without NMS. The results can be found in the

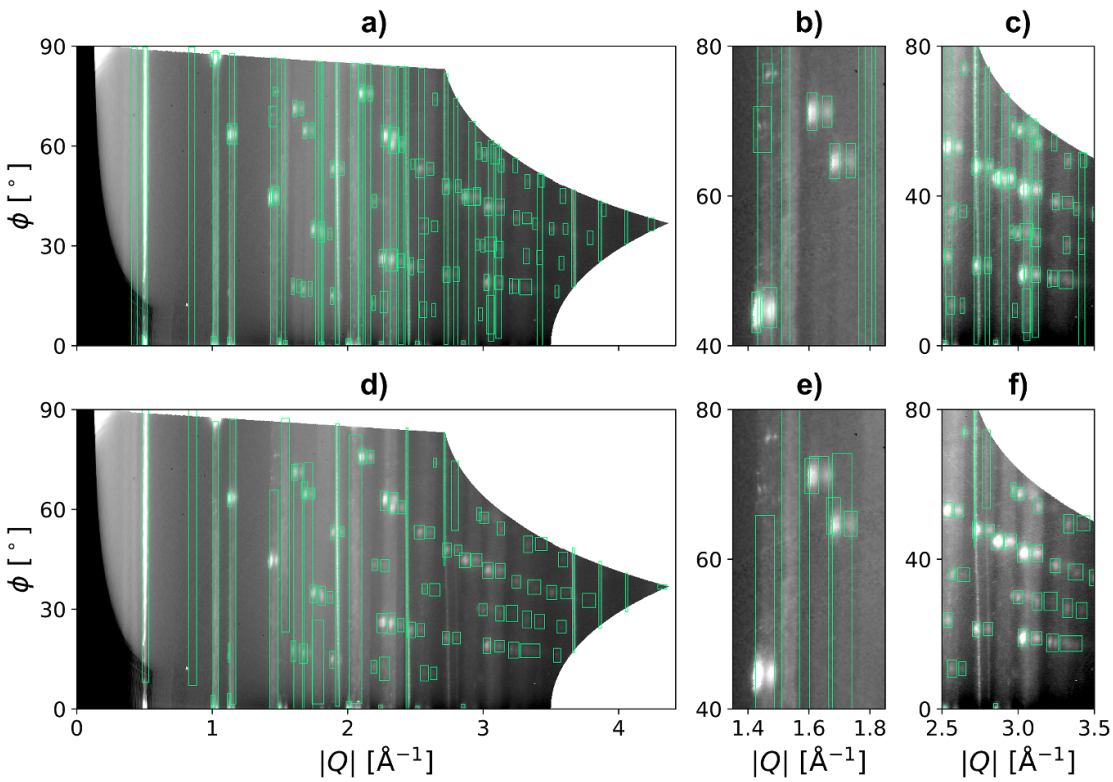

Figure 4. An example of Bragg peak detection results on the identical data shown in figure 1: (a)–(c) DINO + modified Swin-L and (d)–(f) modified Faster R-CNN. Panels (b)–(c) and (e)–(f) show enlarged parts of the pattern in (a) and (d) with the most prominent differences in the resulting detected peaks.

Table 2. Labelling criteria for the three confidence levels.

Confidence level	Criteria
High	– Bright rings, bright peaks, the brighter peak if two are close to each other. – If a Bragg peak is on top of a Debye–Scherrer ring, the underlying ring gets the label high.
Medium	– Medium intensity peaks and rings, medium intensity peaks near the missing wedge. – Medium intensity peaks if they span over a detector gap.
Low	– Low intensity peaks and rings, peaks near the missing wedge, peaks over a detector gap. – The less bright peaks if two are close to each other.

right column of table 1. While this was a functioning solution, it did not improve the performance compared to the Faster R-CNN method. However, omitting NMS could be useful in scenarios where many peaks significantly overlap. This is particularly important for samples with feature-rich GIWAXS patterns due to the presence of thin film peaks on top of substrate rings and/or the coexistence of multiple phases caused by the structural complexity of the studied materials.

While high-intensity Bragg peaks are generally easy to detect, those with lower intensity often blend into the scattering background. Additionally, peaks overlapping with Debye–Scherrer rings can be hard to detect. To reflect the varying detection difficulty for Bragg peaks, we categorised them into three levels of complexity based on their visual ambiguity and proximity to nearby peaks. Table 2 shows the exact criteria used to select the three levels.

We evaluated both the modified Faster R-CNN and DINO with the modified Swin-L backbone and calculated the AP_r for each confidence level, as shown in table 3. We observe a clear advantage of the transformer model in the medium- and low-confidence categories.

As shown in figure 2, we used two different polar conversions. They both have their advantages. While the quasipolar conversion on the left in figure 2 keeps the aspect ratio of all features the same and thereby does not distort the size of pixels, the one on the right increases the size of small features

Table 3. Average precision for different confidence levels and geometries. Bold values indicate the best result in each category.

Model	AP_r high confidence	AP_r medium confidence	AP_r low confidence	AP_r polar geometry	AP_r quasipolar geometry
DINO with modified Swin-L	**89.5**	**81.7**	**58.0**	**75.5**	**75.0**
Modified Faster R-CNN	89.1	77.4	46.3	69.9	67.6

Table 4. Average precision for the high, medium and low confidence levels of the organic dataset. Bold values indicate the best result in each category.

Model	AP_r high	AP_r medium	AP_r low	Overall AP_r
DINO with modified Swin-L	**65.5**	**50.9**	**26.4**	**49.3**
Modified Faster R-CNN	54.1	34.4	19.1	38.3

for low $|Q|$ values and is therefore beneficial for detecting small peaks. We trained the models on both geometries and noticed a better performance on the polar geometry, as shown in table 3.

3.2. Performance on organic films

To demonstrate that our approach generalises beyond perovskite systems, we also evaluated it on a dataset measured from small-molecule organic thin films. The dataset includes both well-oriented samples with sharp Bragg peaks — such as *p*-quaterphenyl (4P), dibenzotetrathiafulvalene (DBTTF), diindenoperylene (DIP), pentacene (PEN), and zinc phthalocyanine (ZnPc) – and 3D-powder-like samples, including buckminsterfullerene (C_{60}) and hexaazatriphenylenehexacarbonitrile (HATCN), which exhibit Debye–Scherrer rings with almost uniform distributions of scattering intensity along the azimuthal direction. The thin films were fabricated by thermal evaporation in a vacuum chamber on silicon substrates. Details of the deposition technique are described in [46]. The GIWAXS measurements were conducted at the ID10-SURF beamline of the European Synchrotron Radiation Facility (France) using an EIGER2 CdTe 4M detector and an x-ray beam of wavelength $\lambda = 0.62$ Å at different incidence angles ranging from 0.06° to 0.12°. The frame with the highest Bragg peak intensity (acquired slightly below the critical angle of the substrate) was used for further analysis.

The datasets of 4P, DBTTF, DIP, and PEN contain several hundred diffraction peaks, whereas the powder-like patterns comprised up to 20 rings, yielding a total of 684 features across 8 images. Full details of the dataset composition, substrates, and beamline configurations are provided in the supplementary information (section A and table A2). Closely spaced and weak peaks make the detection task inherently more challenging, resulting in a considerably lower AP_r than in the perovskite dataset. As observed for the perovskite dataset, the transformer-based DINO-DETR model again shows superior performance, achieving an AP_r of 49.3, substantially higher than the 38.3 obtained with Faster R-CNN, as summarised in table 4. An example of the corresponding detections is provided in figure 5, with DINO-DETR showing a notably larger number of correctly identified peaks.

3.3. Real-time application

For real-time applications, especially in closed-loop experiments, the speed of inference, preprocessing, postprocessing, and hardware requirements are just as critical as the AP_r. To provide a realistic evaluation of the real-time capabilities, we measured the required pre- and post-processing time. Our preprocessing pipeline involves converting raw detector data to polar coordinates, followed by a contrast correction. The post-processing consists of NMS and the application of a minimum confidence threshold. We report the corresponding latencies in table 5, we consider the resulting latency of 279.5 ms fast enough for real-time experiments.

The CPU-based implementation relies on pygid [32] and NumPy [47], whereas the GPU-based implementation leverages CuPy [48], a NumPy-compatible library that enables efficient array operations on CUDA and AMD ROCm-enabled devices. Note that the preprocessing latency is only relevant for *in-situ* applications. For *ex-situ* applications, the pre-processing steps can be buffered or performed in parallel, so a GPU implementation of the preprocessing steps does not provide a significant advantage.

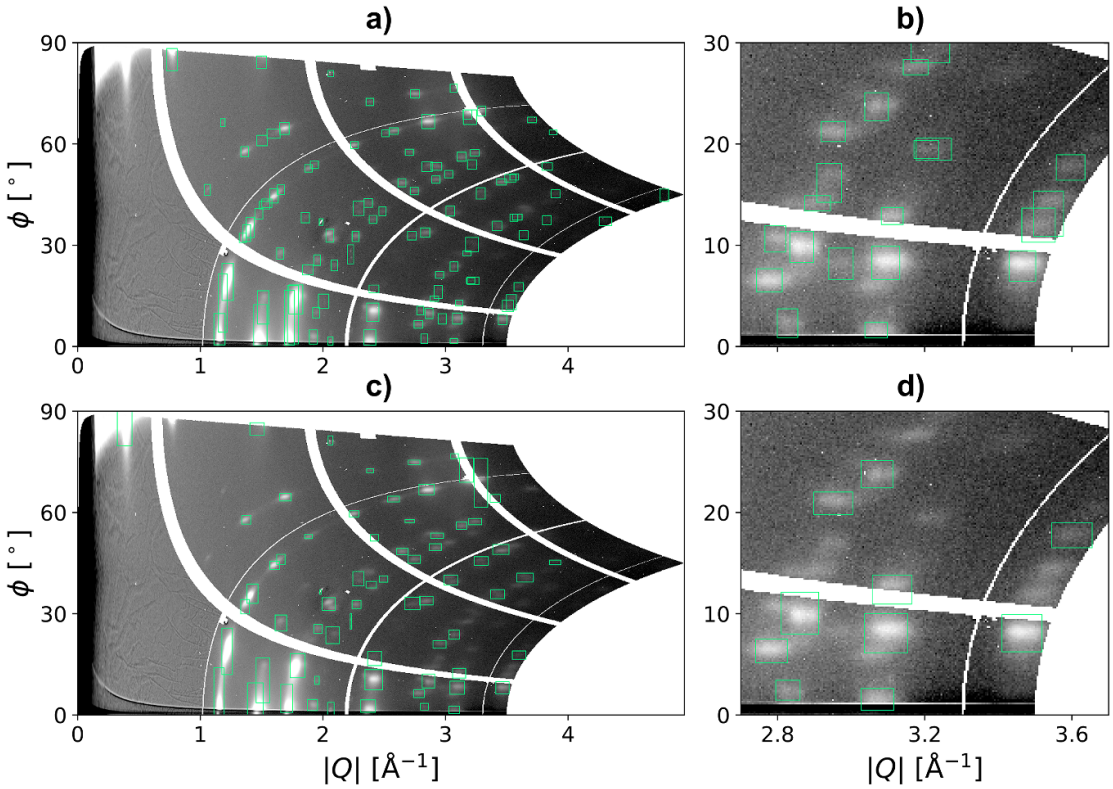

Figure 5. An example of Bragg peak detection results on a GIWAXS pattern from an organic thin film: (a)–(c) DINO + modified Swin-L and (d)–(f) modified Faster R-CNN. Panels (b) and (d) show enlarged parts of the pattern in (a) and (c) with the most prominent differences in the resulting detected peaks.

Table 5. Latency of pre- and post-processing steps as well as inference.

Step	Latency CPU (ms)	Latency GPU (ms)
Generation of the remapping grid for detector images to polar coordinates $(\|Q\|, \phi)$ [a,b]	80	—
Remapping of detector images to polar coordinates $(\|Q\|, \phi)$	3	—
Contrast correction	5	2
Sum pre-processing	88	85
Inference[ic]	—	191
Post-processing	.5	.5

[a] This step needs to be performed only once for all raw images acquired under identical experimental conditions.
[b] This step uses `pygid`, which currently only provides a CPU implementation.
[c] DINO-DETR is currently available only with CUDA support.

4. Discussion and conclusion

We achieved a significant performance improvement in Bragg peak detection for GIWAXS data by using a transformer-based object detection model. Our results on two experimental datasets of diverse compositions show that this approach outperforms existing methods in identifying weaker peaks, peaks superimposed on rings and closely spaced peaks. The elongated attention windows of the modified Swin-L backbone (12×24 tokens in $W \times H$) provide a physically motivated inductive bias. They allow the model to simultaneously attend to the azimuthally extended profile of Debye–Scherrer rings and to smaller peaks. Bipartite matching is beneficial for detecting closely spaced and overlapping peaks. Each learned query is forced by the Hungarian matching loss to attend to a unique object, which means the model is trained to disentangle peaks even when they are close to each other. This is in contrast to NMS-based detectors, which merge detections above an IoU threshold and therefore fail when two detected peaks produce bounding boxes with high overlap. Although we optionally apply a NMS step as post-processing to prune duplicate predictions (section 3), this differs fundamentally from the structural role NMS plays in region-based detectors. The deformable attention, with its 4 learned sampling points

per head, further supports this. Rather than integrating over a fixed region, the model can place sampling points selectively at the intensity maxima of individual peaks within a crowded region, allowing it to resolve their positions independently. The improvement in AP_r for medium- and low-confidence peaks (tables 3 and 4) is consistent with this interpretation, as these categories include peaks on rings and closely spaced peaks. The ability to train transformer models end-to-end is an additional strength, enabling joint training with downstream components of the analysis pipeline. For example, the peak detection model could be combined with a structure matching algorithm for automated identification and tracking of perovskite formation, as demonstrated in Starostin *et al* [28]. Looking further ahead, the attention mechanism allows peak tracking across time frames in *in-situ* experiments. Rather than relying on graph optimisation or regression, transformer attention layers can correlate features between frames [49, 50], enabling joint detection and tracking in a single model. To maximise accessibility and usability for research scientists, we have integrated DINO with modified Swin-L into the `mlgid` pipeline (figure 1). With a much stronger peak detection, the entire `mlgid` pipeline profits, leading to more robust results and higher confidence in subsequent analysis steps (e.g. structure determination). The studied model could potentially also be applied to other datasets with large dynamic range and weak signal-to-noise ratio.

Acknowledgments

We acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the provision of experimental facilities. Parts of this research were carried out at the P08 beamline at PETRA III. We would like to thank Dr Florian Bertram for assistance during the experiments. Beamtimes were allocated for proposals I-20190761 and I-20200254. This research was supported in part through the Maxwell computational resources operated at DESY. We also acknowledge the European Synchrotron Radiation Facility (ESRF) for provision of synchrotron radiation facilities at the ID10-SURF beamline under proposal number SC-5641, and we would like to thank Dr Oleg Konovalov for assistance and support in using the beamline. We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for support through the Cluster of Excellence 'Machine Learning—New Perspectives for Science' (EXC 2064/1, Project No. 390727645) and through Grant No. 460248799 to DAPHNE4NFDI. The research was part of the VIPR project (ErUM-DATA Project No. 05D23VT1) funded by the German Federal Ministry for Science and Education (BMBF). The authors acknowledge the OSCARS project, which has received funding from the European Commission's Horizon Europe Research and Innovation program under Grant Agreement No. 101129751.

Data availability statement

Data produced at the ID10-SURF beamline at ESRF (Grenoble, France) are available at https://doi.org/10.15151/ESRF-ES-1940867632 [51].

Supplementary data 1 available at https://doi.org/10.1088/2632-2153/ae7ec8/data1.

Author contributions

Constantin Völter 0009-0001-2690-4468
Conceptualization (equal), Methodology (equal), Software (equal), Writing – original draft (equal)

Ainur Abukaev 0000-0002-7164-5688
Data curation (equal), Writing – review & editing (equal)

Ekaterina Kneschaurek 0000-0002-0375-3695
Data curation (equal), Writing – review & editing (equal)

Mikhail Romodin 0009-0005-4073-9859
Data curation (equal), Writing – review & editing (equal)

Oleg Gorobtsov 0000-0002-0057-1968
Validation (equal), Writing – review & editing (equal)

Alexander Hinderhofer 0000-0001-8152-6386

Conceptualization (equal), Data curation (equal), Supervision (equal), Writing – review & editing (equal)

Dmitry Lapkin 0000-0003-0680-8740
Conceptualization (equal), Data curation (equal), Writing – original draft (equal)

Frank Schreiber 0000-0003-3659-6718
Supervision (equal), Writing – review & editing (equal)

References

[1] Feidenhans'l R 1989 Surf. structure determination by x-ray diffraction *Surf. Sci. Rep.* **10** 105–88
[2] Werzer O, Kowarik S, Gasser F, Jiang Z, Strzalka J, Nicklin C and Resel R 2024 X-ray diffraction under grazing incidence conditions *Nat. Rev. Methods Primers* **4** 15
[3] Fenter P, Schreiber F, Zhou L, Eisenberger P and Forrest S R 1997 In situ studies of morphology, strain and growth modes of a molecular organic thin film *Phys. Rev.* B **56** 3046–53
[4] Ritley K A, Krause B, Schreiber F, and Dosch H 2001 A portable ultrahigh vacuum organic molecular beam deposition system for in situ x-ray diffraction measurements *Rev. Sci. Instrum.* **72** 1453–7
[5] Kowarik S, Gerlach A, Sellner S, Schreiber F, Cavalcanti L and Konovalov O 2006 Real-time observation of structural and orientational transitions during growth of organic thin films *Phys. Rev. Lett.* **96** 125504
[6] Qin M, Chan P F and Lu X 2021 A systematic review of metal halide perovskite crystallization and film formation mechanism unveiled by in situ GIWAXS *Adv. Mater.* **33** 2105290
[7] Wang J, Wang W, Chen Y, Song L and Huang W 2021 Growth and degradation kinetics of organic-inorganic hybrid perovskite films determined by in situ grazing-incidence x-ray scattering techniques *Small Methods* **5** 2100829
[8] Steele J A *et al* 2023 How to GIWAXS: grazing incidence wide angle x-ray scattering applied to metal halide perovskite thin films *Adv. Energy Mater.* **13** 2300760
[9] Wang C, Steiner U and Sepe A 2018 Synchrotron big data science *Small* **14** 1802291
[10] Pithan L *et al* 2023 Closing the loop: autonomous experiments enabled by machine-learning-based online data analysis in synchrotron beamline environments *J. Synchrotron Rad.* **30** 1064–75
[11] Kavadiya S, Strzalka J, Niedzwiedzki D M and Biswas P 2019 Crystal reorientation in methylammonium lead iodide perovskite thin film with thermal annealing *J. Mater. Chem.* A **7** 12790–9
[12] Mahmood A and Wang J 2020 A review of grazing incidence small- and wide-angle x-ray scattering techniques for exploring the film morphology of organic solar cells *Sol. RRL* **4** 2000337
[13] Smilgies D-M 2025 Probing functional thin films with grazing incidence x-ray scattering: the power of indexing *Crystals* **15** 63
[14] Bommel S, Kleppmann N, Weber C, Spranger H, Schäfer P, Novak J, Roth S V, Schreiber F, Klapp S H L and Kowarik S 2014 Unravelling the multilayer growth of the fullerene C60 in real time *Nat. Commun.* **5** 5388
[15] Wang B, Yager K, Yu D and Hoai M 2017 X-ray scattering image classification using deep learning *2017 IEEE Winter Conf. Applications of Computer Vision (WACV)* pp 697–704
[16] Hexemer A, Zwart P H, McReynolds D, Green A, and Chavez Esparza T A 2021 MLExchange v1 (https://doi.org/10.11578/dc.20210805.4)
[17] Chen Z *et al* 2021 Machine learning on neutron and x-ray scattering and spectroscopies *Chem. Phys. Rev.* **2** 031301
[18] Van Herck W, Fisher J and Ganeva M 2021 Deep learning for x-ray or neutron scattering under grazing-incidence: extraction of distributions *Mater. Res. Express* **8** 045015
[19] Choudhary A, Fox G C and Hey T 2023 *Artificial Intelligence For Science: A Deep Learning Revolution* (World Scientific Publishing Company)
[20] Hinderhofer A, Greco A, Starostin V, Munteanu V, Pithan L, Gerlach A and Schreiber F 2023 Machine learning for scattering data: strategies, perspectives and applications to surf. scattering *J. Appl. Cryst.* **56** 3–11
[21] Johnson H M, Gusev F, Dull J T, Seo Y, Priestley R D, Isayev O and Rand B P 2024 Discovery of crystallizable organic semiconductors with machine learning *J. Am. Chem. Soc.* **146** 21583–90
[22] Munteanu V, Starostin V, Greco A, Pithan L, Gerlach A, Hinderhofer A, Kowarik S and Schreiber F 2024 Neural network analysis of neutron and x-ray reflectivity data incorporating prior knowledge *J. Appl. Cryst.* **57** 456–69
[23] Yang Z, Cui J, Yang L, Lu X, Suo X, Cui X, Chen H and Xing H 2025 Powder x-ray diffraction patterns based deep learning for the prediction of adsorption in metal-organic frameworks *Sep. Purif. Technol.* **372** 133390
[24] Chavez T *et al* 2025 A machine-learning-driven data labeling pipeline for scientific analysis in *MLExchange J. Appl. Cryst.* **58** 731–45
[25] Maruyama S, Ouchi K, Koganezawa T and Matsumoto Y 2020 High-throughput and autonomous grazing incidence x-ray diffraction mapping of organic combinatorial thin-film library driven by machine learning *ACS Comb. Sci.* **22** 348–55
[26] Liu S, Melton C N, Venkatakrishnan S, Pandolfi R J, Freychet G, Kumar D, Tang H, Hexemer A and Ushizima D M 2019 Convolutional neural networks for grazing incidence x-ray scattering patterns: thin film structure identification *MRS Commun.* **9** 586–92
[27] Chavez T, Roberts E J, Zwart P H and Hexemer A 2022 A comparison of deep-learning-based inpainting techniques for experimental x-ray scattering *Appl. Crystallogr.* **55** 1277–88
[28] Starostin V, Munteanu V, Greco A, Kneschaurek E, Pleli A, Bertram F, Gerlach A, Hinderhofer A and Schreiber F 2022 Tracking perovskite crystallization via deep learning-based feature detection on 2D x-ray scattering data *npj Comput. Mater.* **8** 101
[29] Völter C, Starostin V, Lapkin D, Munteanu V, Romodin M, Hylinski M, Gerlach A, Hinderhofer A and Schreiber F 2025 Benchmarking deep learning for automated peak detection on GIWAXS data *J. Appl. Cryst.* **58** 513–22
[30] mlgid: Automated GIWAXS analysis pipeline. (available at: https://github.com/mlgid-project) (Accessed 17 November 2025)
[31] Zhang H, Li F, Liu S, Zhang L, Su H, Zhu J, Ni L M, Shum H-Y 2022 DINO: DETR with improved DeNoising anchor boxes for end-to-end object detection (arXiv:2203.03605)
[32] Abukaev A, Völter C, Romodin M, Schwartzkopff S, Bertram F, Konovalov O, Hinderhofer A, Lapkin D and Schreiber F 2025 Pygid – a Python package for fast data reduction in grazing-incidence diffraction (GID) *J. Appl. Crystallogr.* **59** 263–75

[33] Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S 2020 End-to-end object detection with transformers *Computer Vision – Eccv 2020* vol 12346, ed A Vedaldi, H Bischof, T Brox and J-M Frahm (Springer) pp 213–29
[34] He K, Zhang X, Ren S, Sun J 2015 Deep residual learning for image recognition (arXiv:1512.03385)
[35] Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick C L, Dollár P, 2015 Microsoft COCO: common objects in context (arXiv:1405.0312)
[36] Kieffer J and Karkoulis D 2013 PyFAI, a versatile library for azimuthal regrouping *J. Phys.: Conf. Ser.* **425** 202012
[37] 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 *J. Appl. Cryst.* **48** 917–26
[38] Savikhin V, Steinrück H-G, Liang R-Z, Collins B A, Oosterhout S D, Beaujuge P M and Toney M F 2020 GIWAXS-SIIRkit: scattering intensity, indexing and refraction calculation toolkit for grazing-incidence wide-angle x-ray scattering of organic materials *J. Appl. Cryst.* **53** 1108–29
[39] Reus M A, Reb L K, Kosbahn D P, Roth S V and Müller-Buschbaum P 2024 INSIGHT: in situ heuristic tool for the efficient reduction of grazing-incidence x-ray scattering data *J. Appl. Cryst.* **57** 509–28
[40] Gasser F, Simbrunner J, Huck M, Moser A, Steinrück H-G and Resel R 2025 Intensity corrections for grazing-incidence x-ray diffraction of thin films using static area detectors *J. Appl. Cryst.* **58** 96–106
[41] Völter C, Starostin V, Romodin M, Kneschaurek E, Lapkin D, Hinderhofer A and Schreiber F 2025 mlgidGUI - an annotation program for 2D scattering data *J. Open Source Softw.* **10** 8499
[42] Zhao Z-Q, Zheng P, Xu S-T and Wu X 2019 Object detection with deep learning: a review *IEEE Trans. Neural Netw. Learn. Syst.* **30** 3212–32
[43] Völter C, Starostin V, Hinderhofer A, Lapkin D, Romodin M, Kneschaurek E, and Schreiber F 2026 GIWAXS data and corresponding fits for machine learning (https://doi.org/10.5281/zenodo.18385166)
[44] Kneschaurek E, Hinderhofer A, Hofferberth B, Scheffczyk N, Pithan L, Zimmermann P, Merten L, Bertram F and Schreiber F 2023 Compact sample environment for in situ x-ray scattering during spin-coating *Rev. Sci. Instrum.* **94** 063901
[45] Zhu X, Su W, Lu L, Li B, Wang X, and Dai J 2021 Deformable DETR: deformable transformers for end-to-end object detection (arXiv:2010.04159)
[46] Lapkin D, Nasro R, Hagara J, Hofferberth B, Hinderhofer A, Gerlach A and Schreiber F 2025 Vacuum chamber for deposition of gradient thin films: toward high-throughput structure–property correlative studies *Rev. Sci. Instrum.* **96** 053905
[47] Harris C R *et al* 2020 Array programming with NumPy *Nature* **585** 357–62
[48] Okuta R, Unno Y, Nishino D, Hido S and Loomis C 2017 CuPy: a NumPy-compatible library for NVIDIA GPU calculations *Nips* (http://learningsys.org/nips17/assets/papers/paper_16.pdf)
[49] Meinhardt T, Kirillov A, Leal-Taixe L, Feichtenhofer C 2022 TrackFormer: multi-object tracking with transformers (arXiv:2101.02702)
[50] Xu Y, Ban Y, Delorme G, Gan C, Rus D and Alameda-Pineda X 2023 TransCenter: transformers with dense representations for multiple-object tracking *IEEE Trans. Pattern Anal. Mach. Intell.* **45** 7820–35
[51] Abukaev A, Hylinski M, Lapkin D and Nasro R 2024 Towards automatized high-throughput studies of structure-property correlations in gradient two-component organic thin films *ESRF (European Synchrotron Radiation Facility) data repository* (https://doi.org/10.15151/ESRF-ES-1940867632)